버전 16

JMP 설명서 라이브러리

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는 것이 아니라 새로운 눈을 가지는 데 있다."
Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

All JMP® 16 books are included in this document.

The correct bibliographic citation for this document is as follows: SAS Institute Inc. 2020–2021. JMP® 16 설명서 라이브러리. Cary, NC: SAS Institute Inc.

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

March 2021

August 2021

버전 16

JMP 살펴보기

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는
것이 아니라 새로운 눈을 가지는 데 있다 ."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.0

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 살펴보기 16. Cary, NC: SAS Institute Inc.

JMP 살펴보기® 16

Copyright © 2020-2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

2021년 3월

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP를 처음 사용하든 오랫동안 사용해왔든, JMP와 관련해서 배워야 하는 것은 언제나 있습니다.

JMP.com 을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다.

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP 를 사용하는 방법을 보여주는 성공 사례
- JMP 직원의 팁 , 힌트 및 사례가 포함된 블로그
- 다른 사용자와 JMP 에 대해 토론할 수 있는 포럼

https://www.jmp.com/getstarted

목차
JMP 살펴보기

설명서 소개 9

JMP 그래프 갤러리 11

1 JMP 알아보기 31

설명서 및 추가 리소스

JMP 설명서에 사용되는 서식 규칙 33
JMP 도움말 34
JMP 설명서 라이브러리 34
JMP 학습을 위한 추가 리소스 40
JMP 자습서 40
샘플 데이터 테이블 40
통계 및 JSL 용어에 대해 알아보기 41
JMP 팁 및 힌트 알아보기 41
JMP 툴팁 41
JMP 사용자 커뮤니티 41
무료 온라인 Statistical Thinking 교육 과정 42
새로운 사용자를 위한 JMP 입문 키트 42
Statistics Knowledge 포털 42
JMP 교육 과정 42
사용자가 작성한 JMP 설명서 42
JMP 시작하기 창 42
JMP 기술 지원 43

2 JMP 소개 45

기본 개념

JMP 필수 개념 47
JMP 시작하기 47
JMP 시작 48
샘플 데이터 사용 49
데이터 테이블 이해 51
JMP 워크플로우 이해 52
1 단계 : JMP 플랫폼 시작 및 결과 보기 53
2 단계 : JMP 보고서에서 상자 그림 제거 54
3 단계 : 추가 JMP 출력 요청 55

4 단계 : JMP 플랫폼 결과에서의 상호 작용 56
JMP 와 Excel 의 차이점 57
데이터 테이블의 구조 57
JMP 의 계산식 58
JMP 분석 및 그래프 생성 59

3 데이터 작업 61
그래프 생성 및 분석을 위한 데이터 준비

JMP 로 데이터 가져오기 63
데이터 테이블에 데이터 복사 및 붙여넣기 63
데이터 테이블로 데이터 가져오기 63
데이터 테이블에 데이터 입력 66
Excel 에서 JMP 로 데이터 전송 68
데이터 테이블 작업 69
데이터 테이블에서 데이터 편집 70
데이터 테이블에서 값 선택 , 선택 취소 및 찾기 72
데이터 테이블에서 열 정보 보기 또는 변경 75
계산식으로 값 계산의 예 77
보고서 데이터 필터링의 예 78
데이터 재구성의 예 80
요약 통계량 보기 80
부분집합 생성의 예 85
예 : 데이터 테이블 결합 87
데이터 정렬 88

4 데이터 시각화 91
일반적인 그래프

단변량 그래프의 단일 변수 분석 93
연속형 변수에 히스토그램 사용 93
범주형 변수에 막대 차트 사용 96
다중 변수 비교 98
산점도를 사용하여 여러 변수 비교 99
산점도 행렬을 사용하여 여러 변수 비교 102
병렬 상자 그림을 사용하여 여러 변수 비교 105
그래프 빌더를 사용하여 여러 변수 비교 107
버블 그림을 사용하여 여러 변수 비교 113
중첩 그림을 사용하여 여러 변수 비교 118
계량형 차트를 사용하여 여러 변수 비교 122

5 데이터 분석 127
분포 , 관계 및 모형

내용 소개 129
데이터 그래프 생성의 중요성 129

모델링 유형 이해 132
예 : 모델링 유형 결과 보기 133
모델링 유형 변경 134
분포 분석 136
연속형 변수의 분포 137
범주형 변수의 분포 139
관계 분석 142
하나의 예측 변수가 있는 회귀 사용 143
한 변수에 대한 평균 비교 148
비율 비교 151
여러 변수의 평균 비교 153
다중 예측 변수가 있는 회귀 사용 157

6 전체를 보는 시각 163
여러 플랫폼에서 데이터 탐색

재미있는 사실 : 연결된 분석 165
여러 플랫폼에서 데이터 탐색 165
분포 플랫폼에서 분포 분석 165
다변량 플랫폼에서 패턴 및 관계 분석 169
군집화 플랫폼에서 유사 값 분석 172

7 작업 저장 및 공유 177
결과 저장 및 다시 생성

프로젝트 작업 179
새 프로젝트 생성 180
프로젝트에서 파일 열기 180
프로젝트에서 파일 재배열 180
프로젝트 저장 182
프로젝트 작업 영역 182
프로젝트 책갈피 183
프로젝트 내용 184
프로젝트의 최근 사용한 파일 184
프로젝트 로그 185
파일을 프로젝트로 이동 185
프로젝트 내의 열기 스크립트 작성 186
새 프로젝트 생성의 예 186
플랫폼 결과를 저널로 저장 189
예 : 저널 생성 189
저널에 분석 추가 190
스크립트 저장 및 실행 191
예 : 스크립트 저장 및 실행 191
스크립트 및 JSL 192

보고서를 대화식 HTML 로 저장 192
대화식 HTML 의 데이터 포함 193
예 : 대화식 HTML 생성 194
보고서를 대화식 HTML 로 공유 195
보고서를 PowerPoint 프레젠테이션으로 저장 199
대시보드 생성 199
예 : 창 결합 200
예 : 두 개의 보고서가 포함된 대시보드 생성 201

8 특수 기능 203

자동 분석 업데이트 및 SAS Integration

분석 및 그래프 자동 업데이트 205
예 : 자동 재계산 사용 205
환경 설정 변경 209
예 : 환경 설정 변경 209
JMP 와 SAS 통합 212
예 : SAS 코드 생성 212
예 : SAS 코드 전송 213

A 기술 공지사항 215

설명서 소개

JMP 살펴보기에서는 JMP 소프트웨어와 관련한 일반적인 내용을 소개합니다. 이 설명서에서는 사용자가 JMP에 대한 지식이 없다고 가정합니다. 이 설명서에서는 분석가, 연구원, 학생, 교수 또는 통계학자 등을 대상으로 JMP의 사용자 인터페이스 및 기능에 대한 일반적인 개요를 제공합니다.

이 설명서에서는 다음과 같은 정보를 소개합니다.

- JMP 시작
- JMP 창의 구조
- 데이터 준비 및 조작
- 대화식 그래프를 기반으로 한 데이터 파악
- 간단한 분석을 통해 그래프 보강
- JMP 및 특수 기능 사용자 정의
- 결과 공유

이 설명서에는 6 개의 장이 있습니다. 각 장에서는 해당 장에서 제시된 개념을 보강하는 예를 소개합니다. 모든 통계적 개념은 소개 수준으로 제공됩니다. JMP 살펴보기에 사용된 샘플 데이터는 소프트웨어에 포함되어 있습니다. 다음은 각 장에 대한 설명입니다.

- 2 장, "JMP 소개" 에서는 JMP 응용 프로그램의 개요를 제공합니다. 콘텐츠가 구성된 방식과 소프트웨어를 탐색하는 방법을 안내합니다.
- 3 장, " 데이터 작업 " 에서는 다양한 소스에서 데이터를 가져와 분석을 위해 준비하는 방법을 설명합니다. 데이터 조작 도구에 대한 개요도 제공합니다.
- 4 장, " 데이터 시각화 " 에서는 데이터를 시각화하고 이해하기 위해 사용할 수 있는 그래프 및 차트에 대해 설명합니다. 단일 변수가 포함된 간단한 분석에서 여러 변수 간의 관계를 볼 수 있게 해 주는 다중 변수 그래프에 이르기까지 다양한 예제가 있습니다.
- 5 장, " 데이터 분석 " 에서는 일반적으로 사용되는 여러 가지 분석 기법을 설명합니다. 통계적 방법을 사용하지 않아도 되는 간단한 기법에서부터 통계 지식이 유용하게 사용되는 고급 기법에 이르기까지 다양합니다.
- 6 장, " 전체를 보는 시각 " 에서는 여러 플랫폼에서 분포, 패턴 및 유사한 값을 분석하는 방법을 보여 줍니다.
- 7 장, " 작업 저장 및 공유 " 에서는 PowerPoint 프레젠테이션 및 대화식 HTML 에서 JMP 를 사용하지 않는 사람들과 작업을 공유하는 방법을 설명합니다. 분석을 스크립트로 저장하고 JMP 사용자를 위해 저널 및 프로젝트에 작업을 저장하는 방법도 다룹니다.

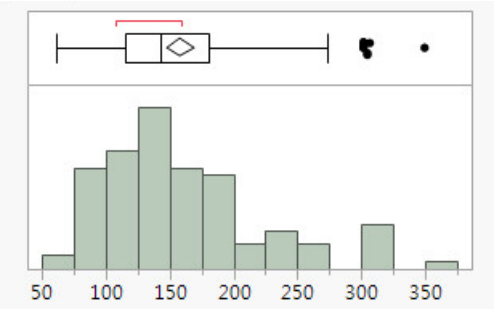
- 8 장 , " 특수 기능 " 에서는 데이터가 변경될 때 그래프 및 분석을 자동으로 업데이트하는 방법 , 환경 설정을 사용하여 보고서를 사용자 정의하는 방법 , JMP 가 SAS 와 상호 작용하는 방법에 대해 설명합니다 .

이 설명서를 검토한 후에는 JMP에서 편안하게 데이터를 탐색하고 작업할 수 있습니다.

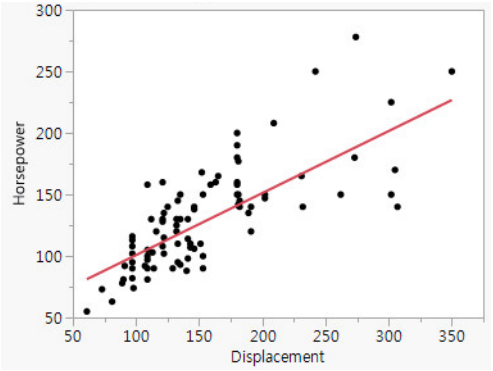
JMP는 Windows 및 macOS 운영 체제 모두에서 사용할 수 있습니다. 그러나 이 설명서의 내용은 Windows 운영 체제를 기반으로 합니다.

JMP 그래프 갤러리
다양한 그래프 및 플랫폼

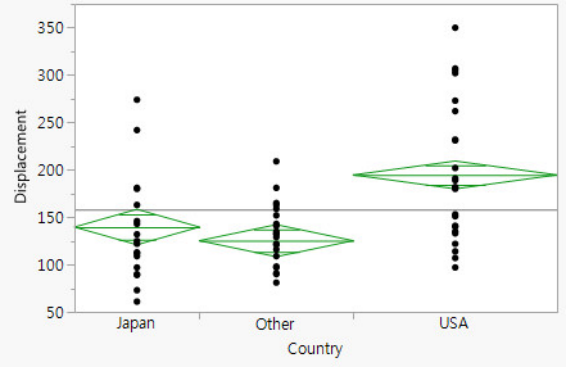
다음은 JMP로 생성할 수 있는 여러 가지 그래프의 그림입니다. 각 그림의 라벨에는 해당 그림을 생성하는 데 사용되는 플랫폼이 지정되어 있습니다. 플랫폼과 이러한 그래프 및 기타 그래프에 대한 자세한 내용은 **도움말 > 사용 설명서** 메뉴의 문서에서 확인하십시오.



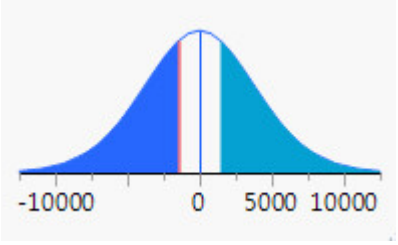
히스토그램
분석 > 분포



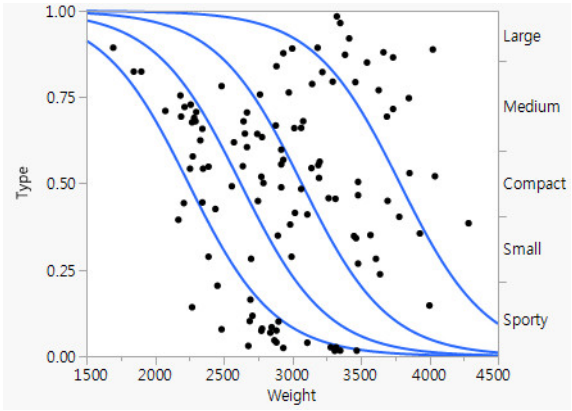
이변량
분석 > X 로 Y 적합



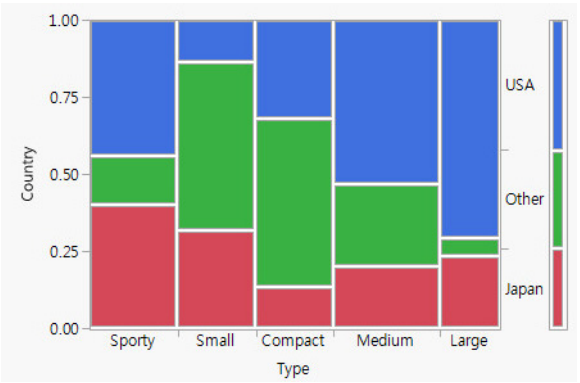
일원 분석
분석 > X 로 Y 적합



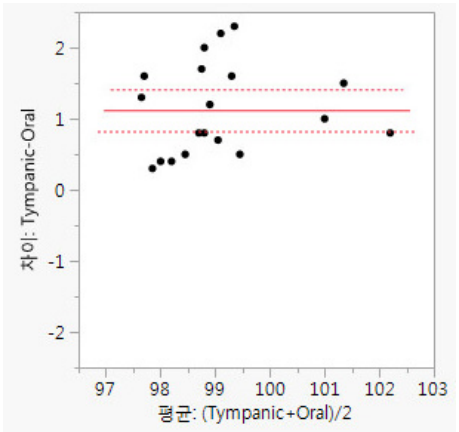
일원 t- 검정
분석 > X 로 Y 적합



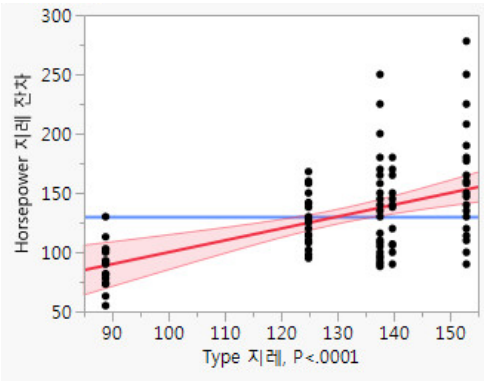
로지스틱
분석 > X 로 Y 적합



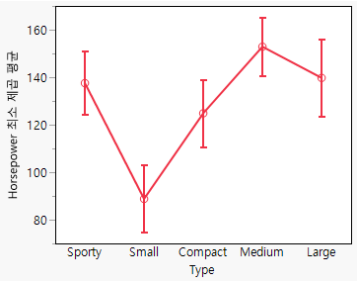
모자이크 그림
분석 > X 로 Y 적합



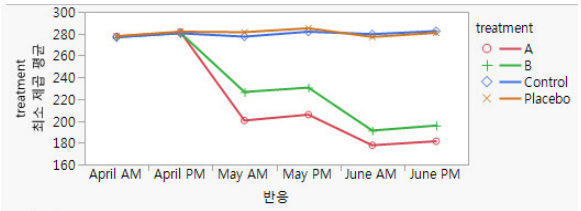
매칭 쌍
분석 > 전문 모형 > 매칭 쌍



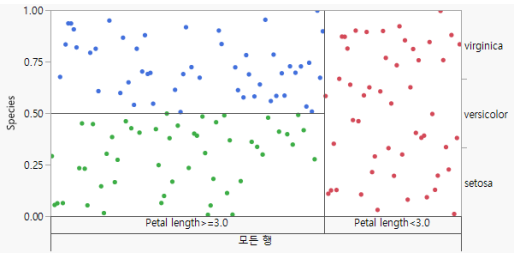
레버리지 그림
분석 > 모형 적합



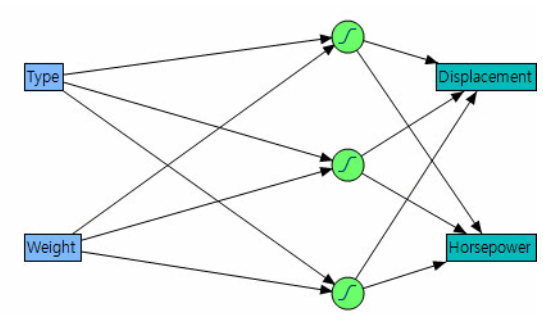
최소 제곱 평균 그림
분석 > 모형 적합



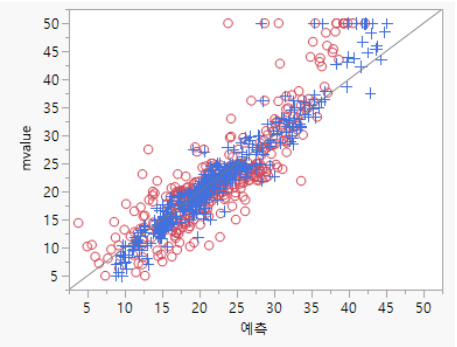
MANOVA
분석 > 모형 적합



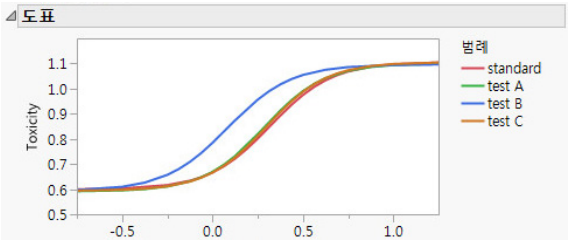
파티션
분석 > 예측 모형 > 파티션



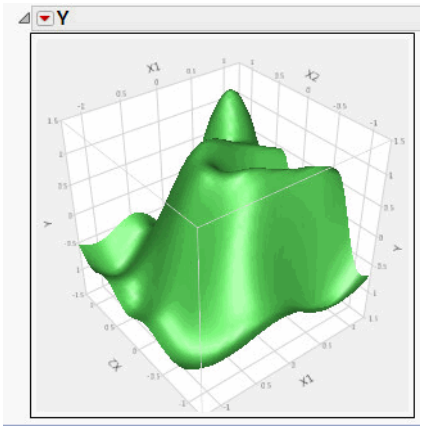
신경망 다이어그램
분석 > 예측 모형 > 신경망



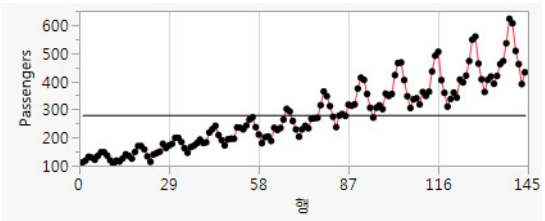
실제값 대 예측값
분석 > 예측 모형 > 모형 비교



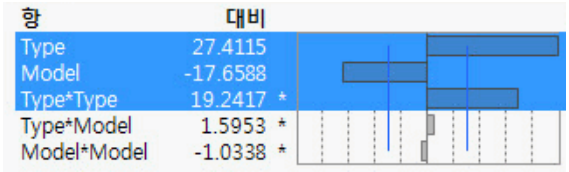
비선형 적합
분석 > 전문 모형 > 비선형



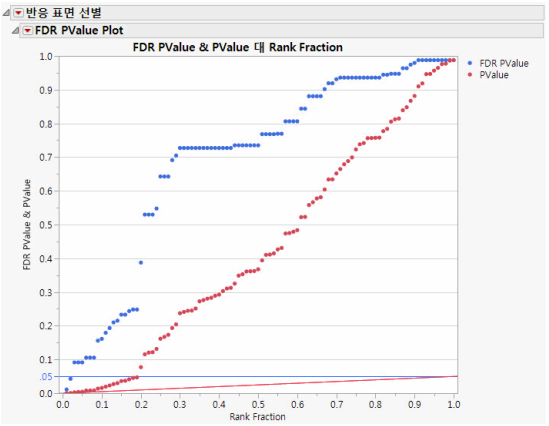
표면 프로파일러
분석 > 전문 모형 > 가우스 과정



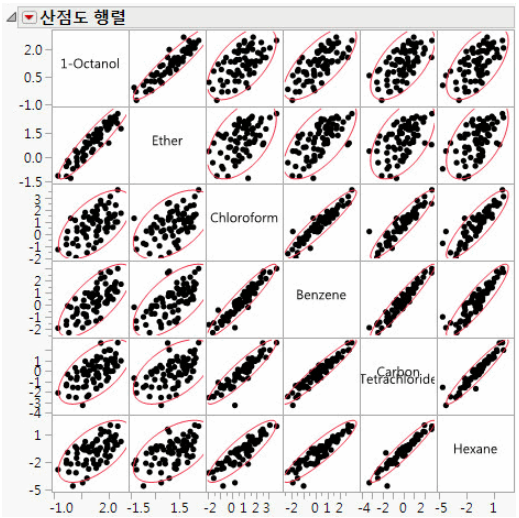
시계열
분석 > 전문 모형 > 시계열



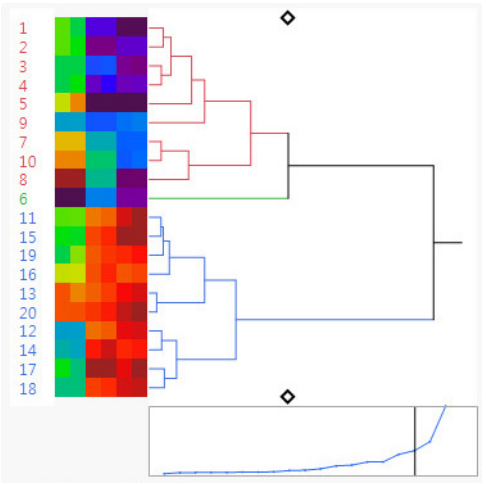
선별
분석 > 전문 모형 > 특수 DOE 모형 > 2 수준 선별 적합



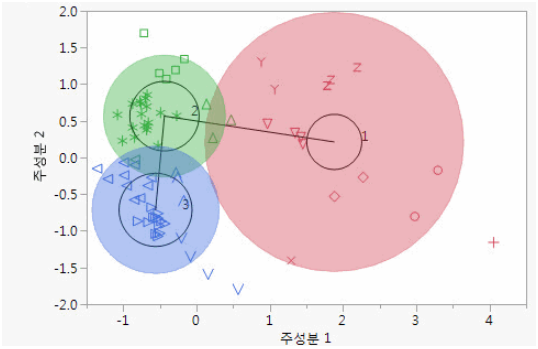
FDR p 값 그림
분석 > 선별 > 반응 변수 선별



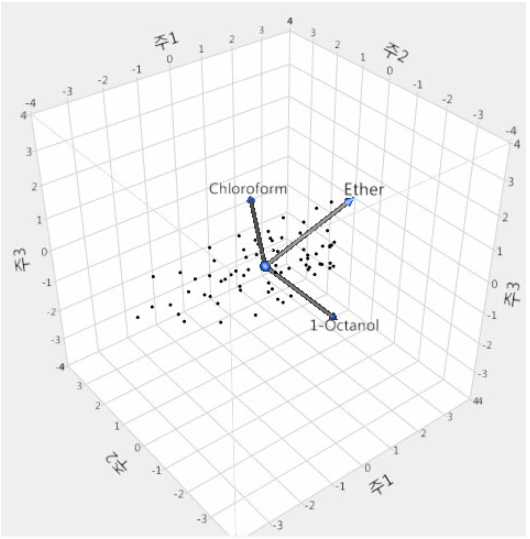
산점도 행렬
분석 > 다변량 방법 > 다변량



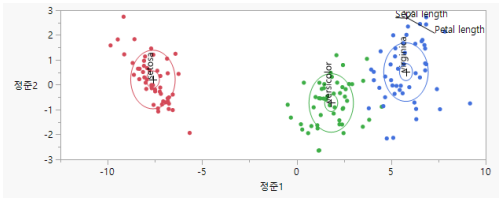
덴드로그램
분석 > 군집화 > 계층적 군집화



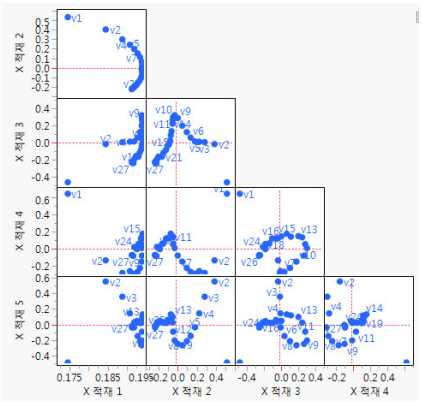
자기 조직화 지도
분석 > 군집화 > K 평균 군집화



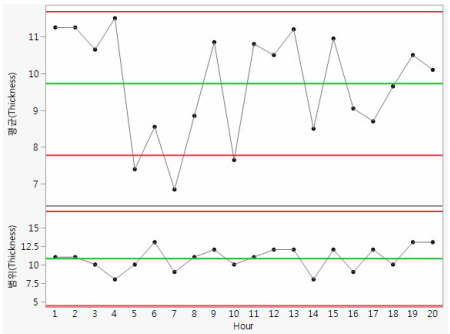
주성분
분석 > 다변량 방법 > 주성분



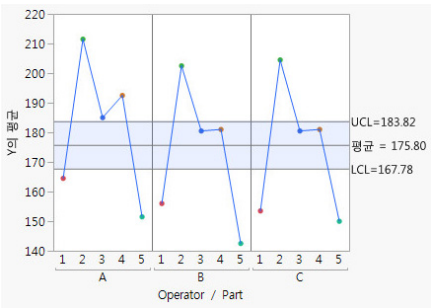
정준 그림
분석 > 다변량 방법 > 판별



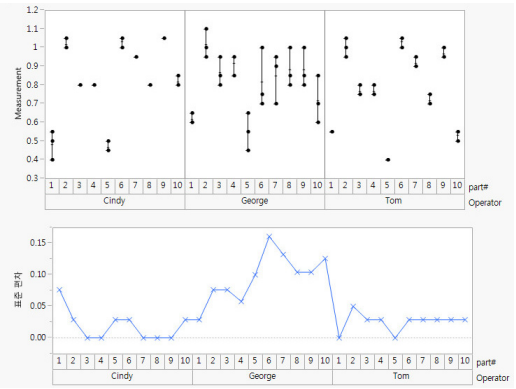
적재 그림
분석 > 다변량 방법 > 부분 최소 제곱



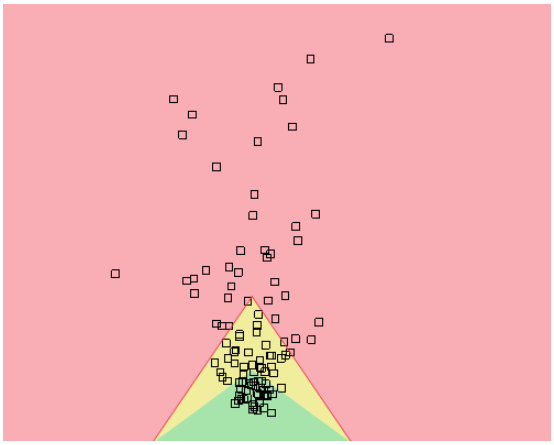
Xbar-R 차트
분석 > 품질 및 공정 > 관리도 빌더



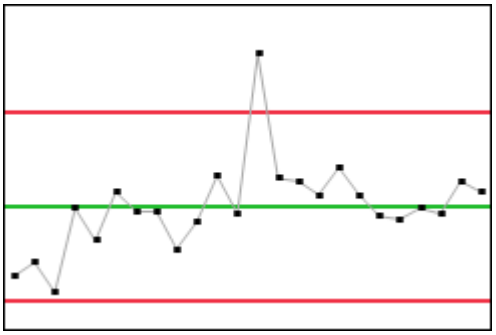
평균 차트
분석 > 품질 및 공정 > 측정 시스템 분석



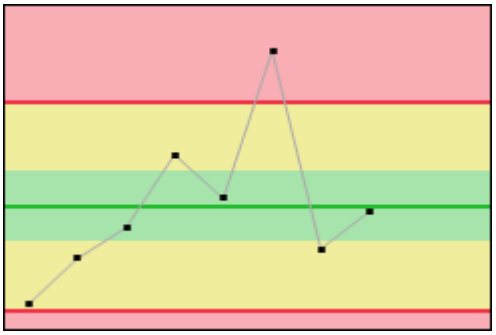
계량형 차트
분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트



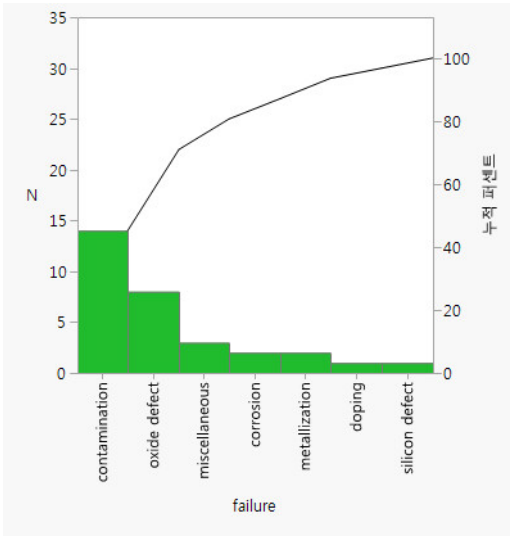
목표 그림
분석 > 품질 및 공정 > 공정 능력



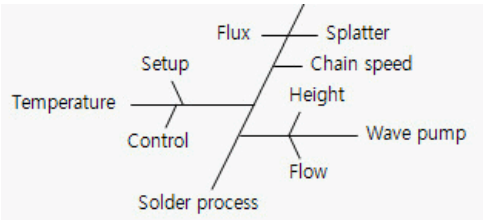
개별 측정값 차트
이동 범위 차트
분석 > 품질 및 공정 > 관리도 > IR



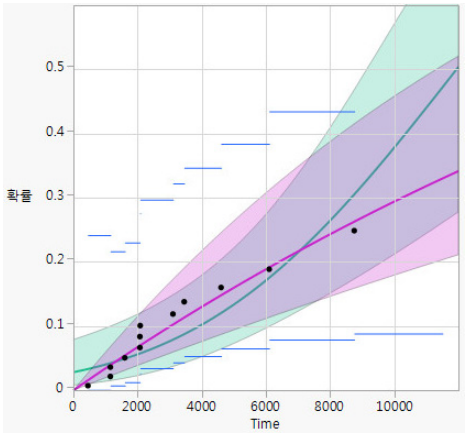
Xbar 차트
분석 > 품질 및 공정 > 관리도 > Xbar



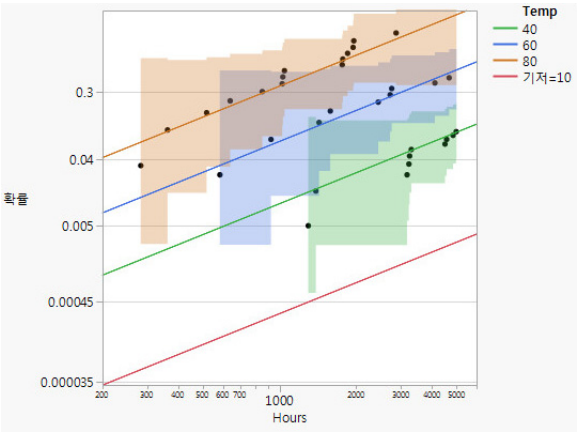
파레토도
분석 > 품질 및 공정 > 파레토도



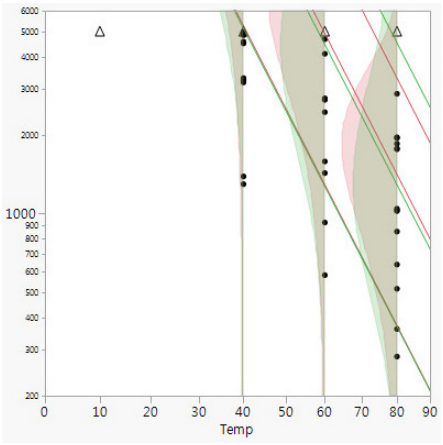
Ishikawa 차트
Fishbone 차트
분석 > 품질 및 공정 > 다이어그램



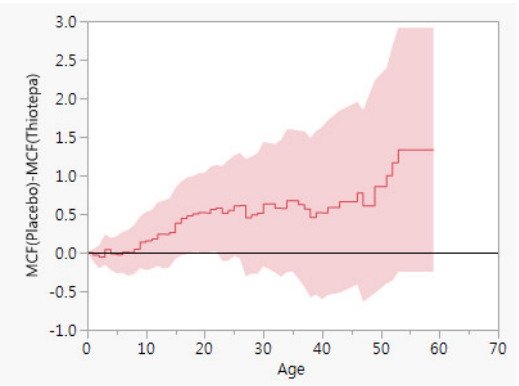
분포 비교
분석 > 신뢰성 및 생존 > 수명 분포



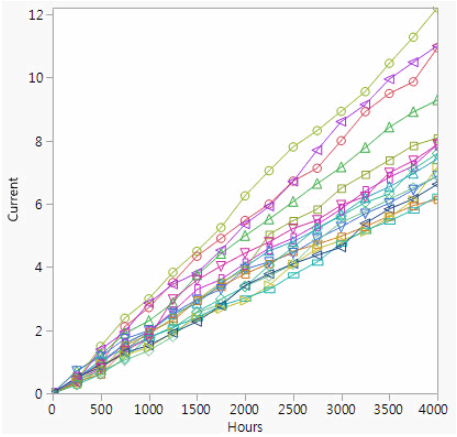
비모수 중첩
분석 > 신뢰성 및 생존 > 수명 분포 적합



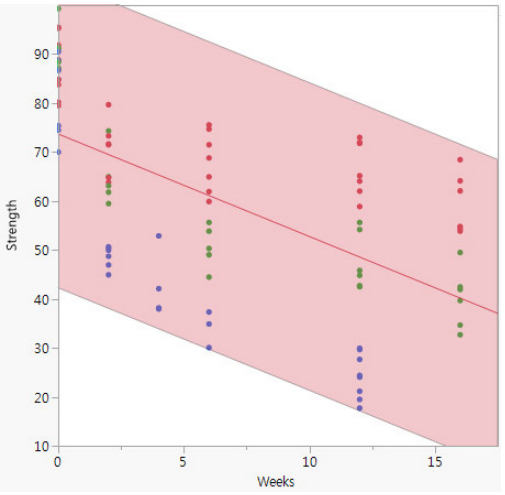
산점도
분석 > 신뢰성 및 생존 > 수명 분포 적합



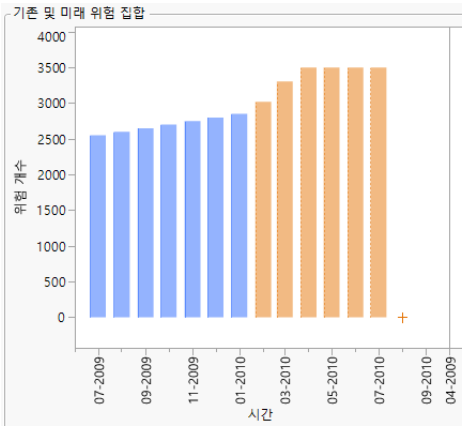
MCF 그림
분석 > 신뢰성 및 생존 > 재발 분석



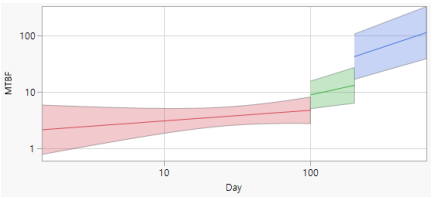
중첩
분석 > 신뢰성 및 생존 > 열화



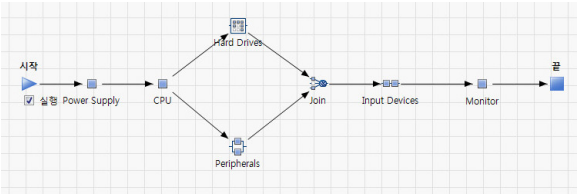
예측 구간
분석 > 신뢰성 및 생존 > 파괴 열화



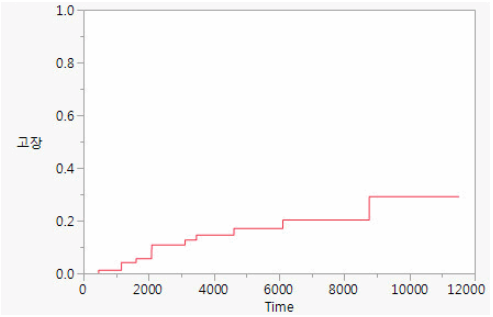
예측
분석 > 신뢰성 및 생존 > 신뢰도 예측



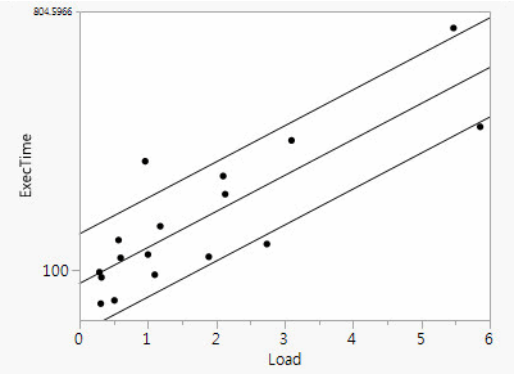
구간적 Weibull NHPP
분석 > 신뢰성 및 생존 > 신뢰도 성장



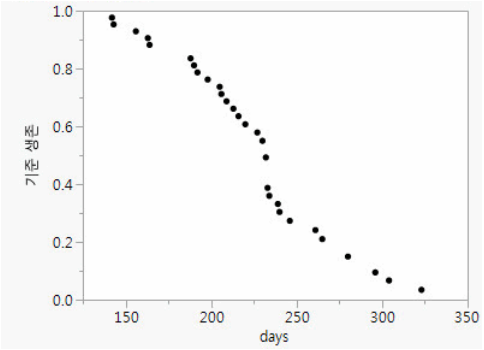
신뢰도 블록 다이어그램
분석 > 신뢰성 및 생존 > 신뢰도 블록 다이어그램



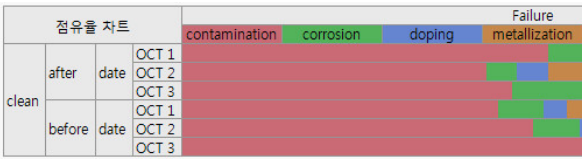
고장 그림
분석 > 신뢰성 및 생존 > 생존



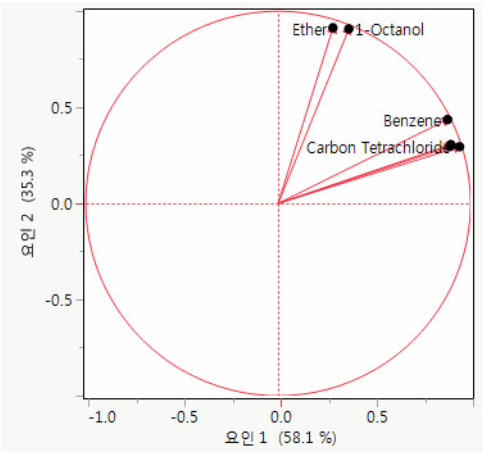
생존 분위수
분석 > 신뢰성 및 생존 > 모수 생존 모형 적합



기준 생존율
분석 > 신뢰성 및 생존 > 비례 위험 모형 적합

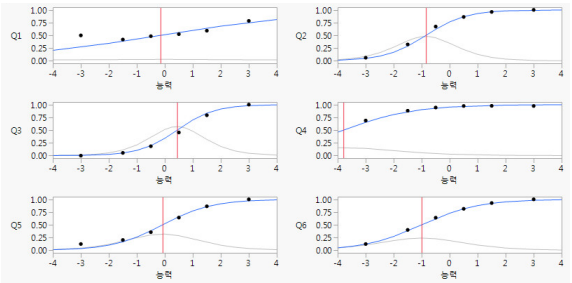


혼합물 프로파일러
분석 > 소비자 조사 > 범주형

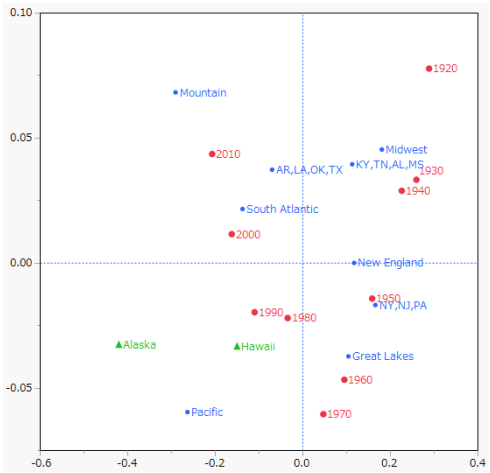


요인 적재 그림
분석 > 다변량 방법 > 요인 분석

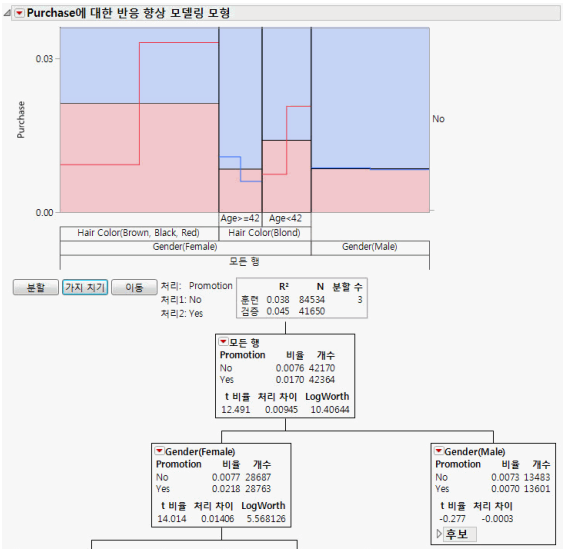
예측 프로파일
분석 > 소비자 조사 > 선택



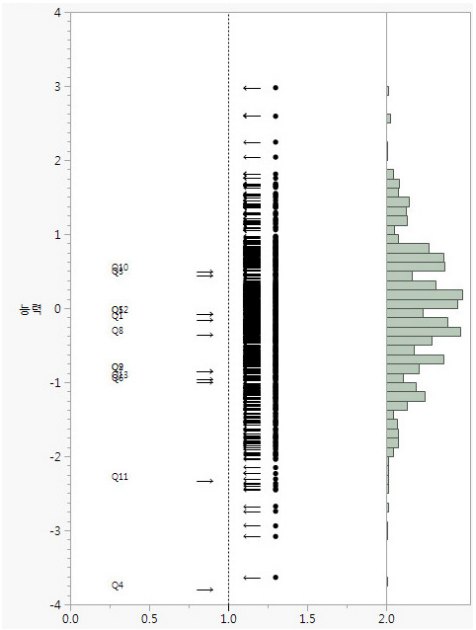
특성 곡선
분석 > 다변량 방법 > 항목 분석



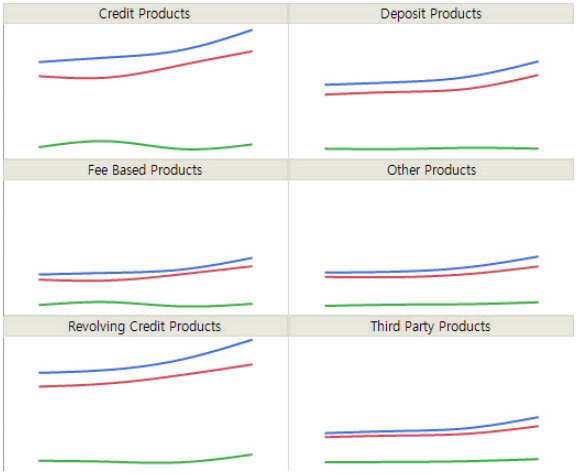
다중 대응 분석
분석 > 다변량 방법 > 다중 대응 분석



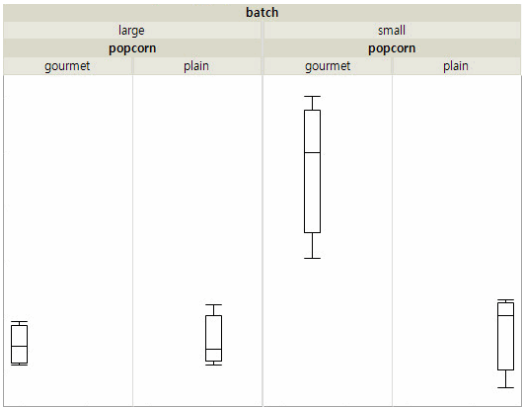
Uplift 모형
분석 > 소비자 조사 > Uplift



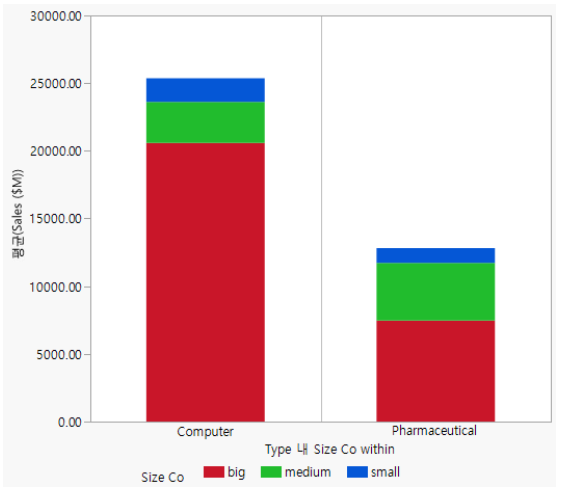
쌍대 그림
분석 > 다변량 방법 > 항목 분석



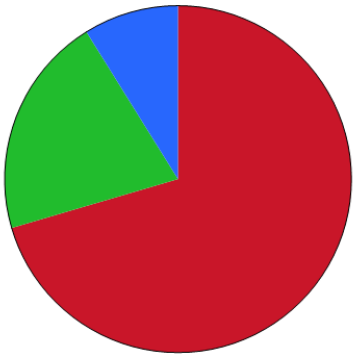
선 그래프
그래프 > 그래프 빌더



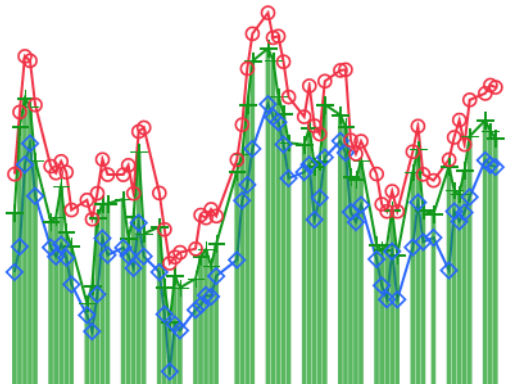
상자 그림
그래프 > 그래프 빌더



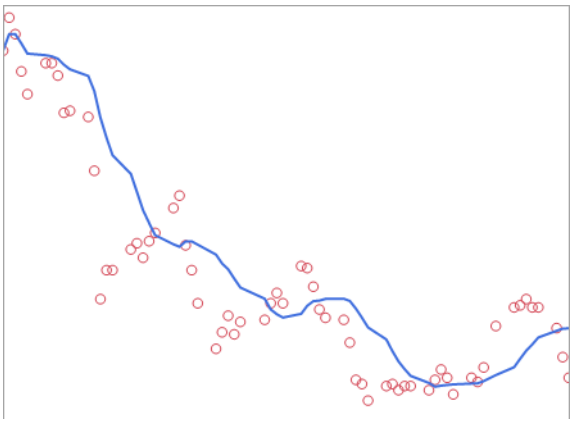
누적 막대 차트
그래프 > 그래프 빌더



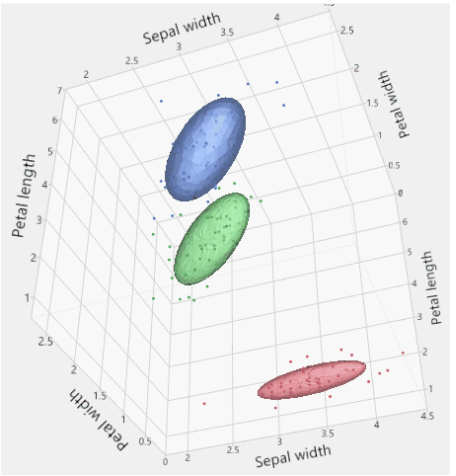
파이 차트
그래프 > 그래프 빌더



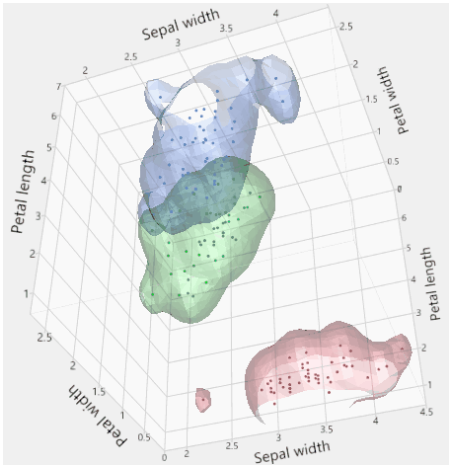
바늘 및 선 차트
그래프 > 그래프 빌더



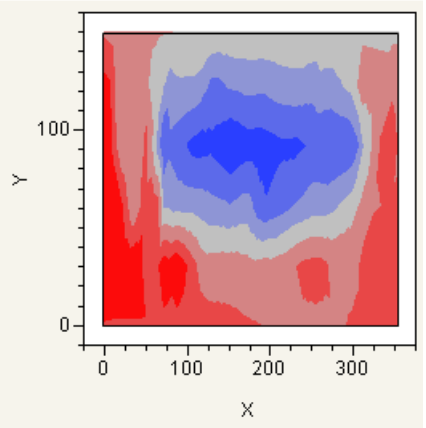
평활기
그래프 > 그래프 빌더



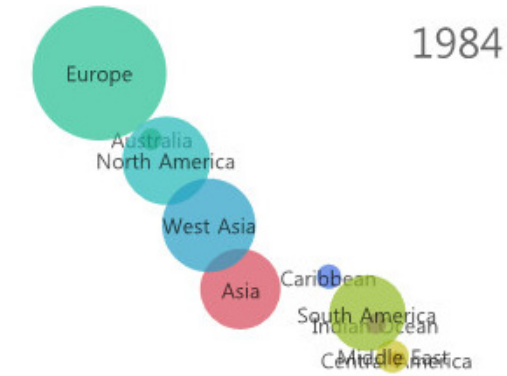
3 차원 산점도
그래프 > 3D 산점도



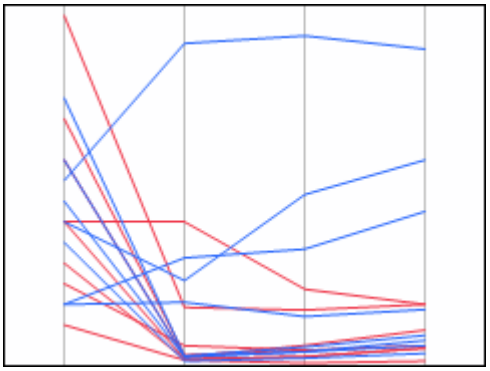
3 차원 산점도
그래프 > 3D 산점도



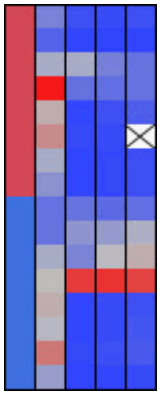
등고선 그림
그래프 > 그래프 빌더



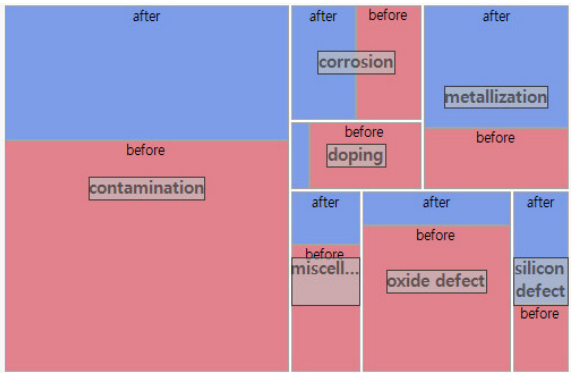
버블 그림
그래프 > 버블 그림



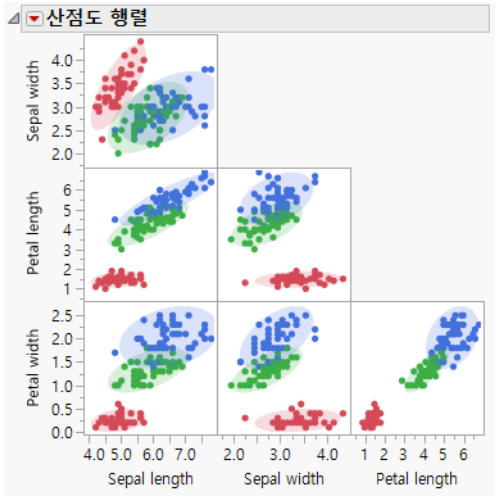
평행 그림
그래프 > 그래프 빌더



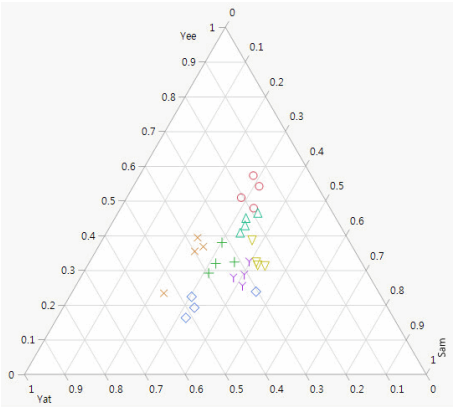
셀 그림
그래프 > 셀 그림



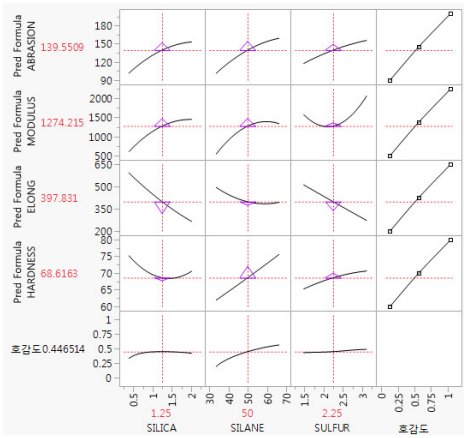
트리맵
그래프 > 그래프 빌더



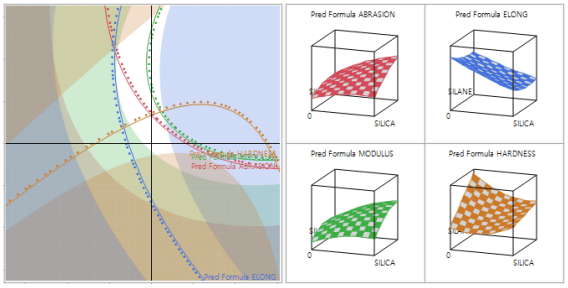
산점도 행렬
그래프 > 산점도 행렬



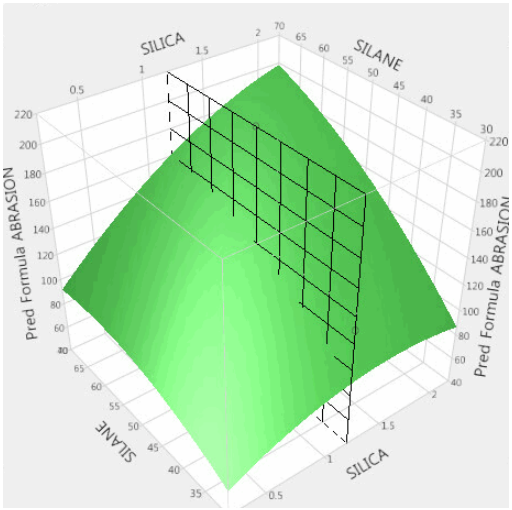
삼원 그림
그래프 > 삼원 그림



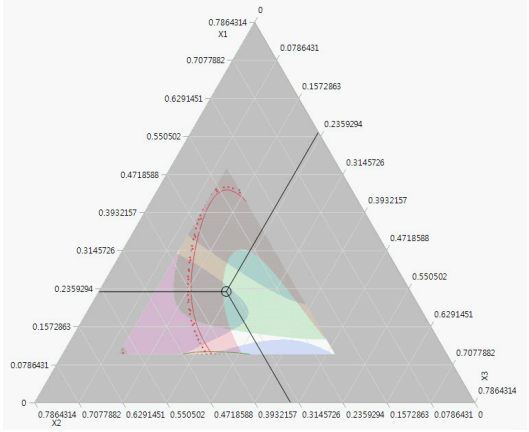
예측 프로파일러
그래프 > 프로파일러



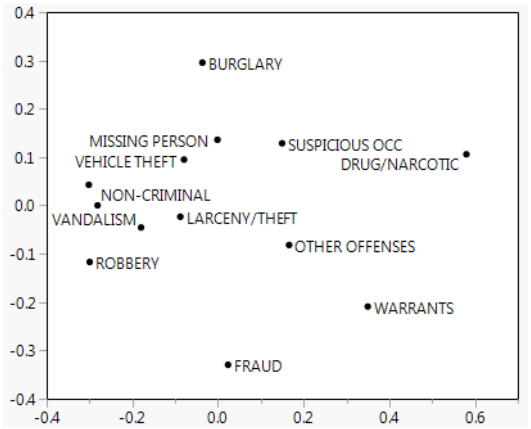
등고선 프로파일러
그래프 > 등고선 프로파일러



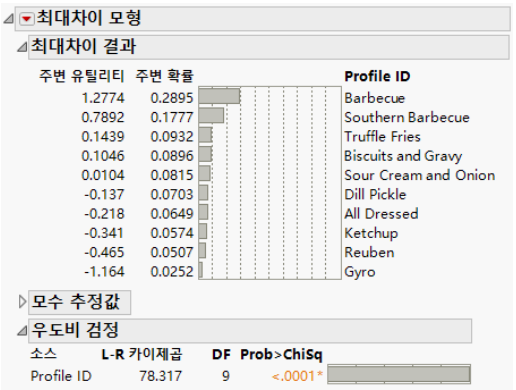
표면 그림
그래프 > 표면 그림



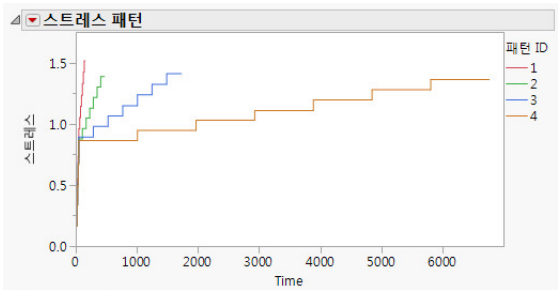
혼합물 프로파일러
그래프 > 혼합물 프로파일러



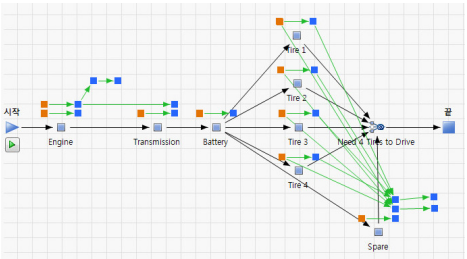
다차원 척도법
분석 > 다변량 방법 > 다차원 척도법



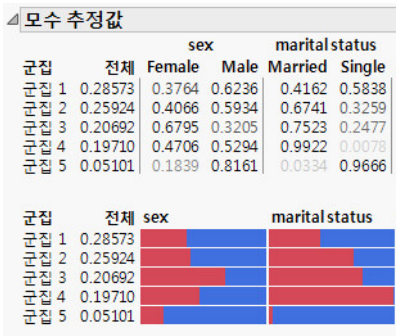
최대차이
분석 > 소비자 조사 > 최대차이



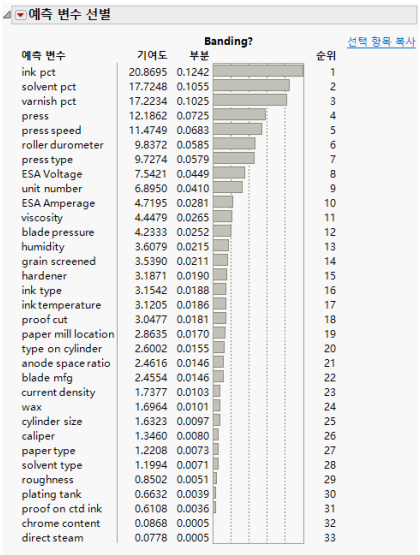
스트레스 패턴 그림
분석 > 신뢰성 및 생존 > 누적 손상



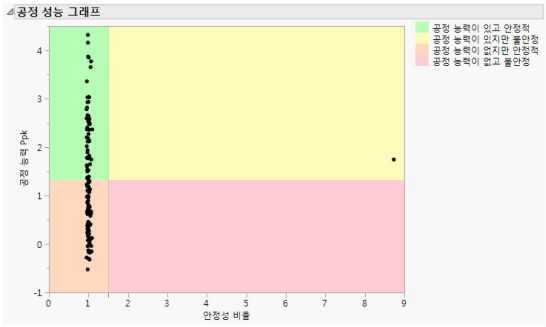
수리 가능 시스템 시뮬레이션
분석 > 신뢰성 및 생존 > 수리 가능 시스템 시뮬레이션



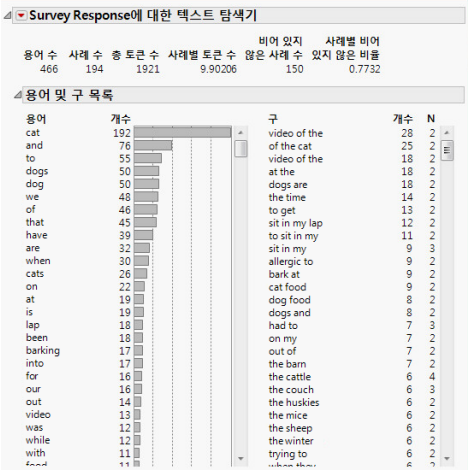
잠재 계층 분석
분석 > 군집화 > 잠재 계층 분석



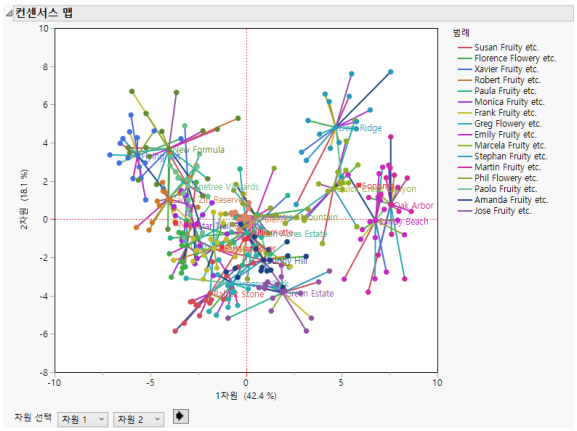
예측 변수 선별
분석 > 선별 > 예측 변수 선별



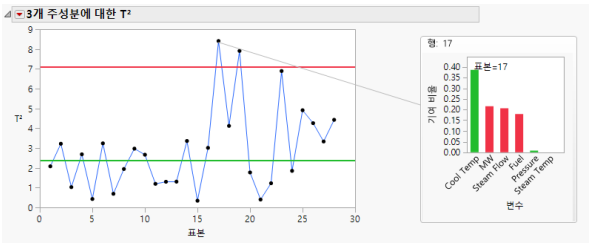
공정 변수 선별
분석 > 선별 > 공정 변수 선별



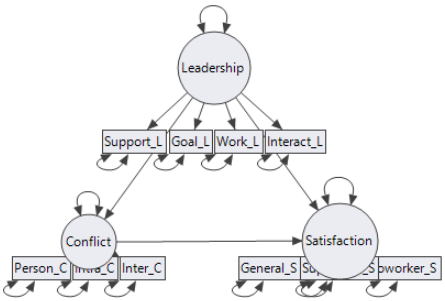
텍스트 탐색기
분석 > 텍스트 탐색기



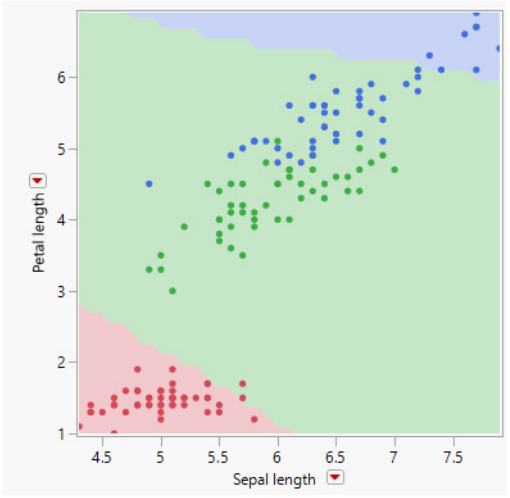
다중 요인 분석
분석 > 소비자 조사 > 다중 요인 분석



모형 기반 다변량 관리도
분석 > 품질 및 공정 > 모형 기반 다변량 관리도



구조 방정식 모형
분석 > 다변량 방법 > 구조 방정식 모형



서포트 벡터 머신
분석 > 예측 모형 > 서포트 벡터 머신

1 장

JMP 알아보기

설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 3
JMP 도움말 4
JMP 설명서 라이브러리 4
JMP 학습을 위한 추가 리소스 10
JMP 자습서 10
샘플 데이터 테이블 10
통계 및 JSL 용어에 대해 알아보기 11
JMP 팁 및 힌트 알아보기 11
JMP 툴팁 11
JMP 사용자 커뮤니티 11
무료 온라인 Statistical Thinking 교육 과정 12
새로운 사용자를 위한 JMP 입문 키트 12
Statistics Knowledge 포털 12
JMP 교육 과정 12
사용자가 작성한 JMP 설명서 12
JMP 시작하기 창 12
JMP 기술 지원 13

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 Lucida Sans Typewriter 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다 . 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다 .

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고 : JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다 . 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 "JMP 설명서 라이브러리 "(4 페이지) 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다 . 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다 . JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오 .

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오 .	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다 . 또한 결과를 공유하는 방법도 알아볼 수 있습니다 .
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다 .	데이터 가져오기 , 열 특성 수정 , 데이터 정렬 , SAS 연결 등을 비롯하여 JMP 의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다 .

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
Fitting Linear Models	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
Quality and Process Methods	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- " 새로운 사용자를 위한 JMP 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

새로운 사용자를 위한 JMP 입문 키트

새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기 " 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기

" 창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

2 장

JMP 소개
기본 개념

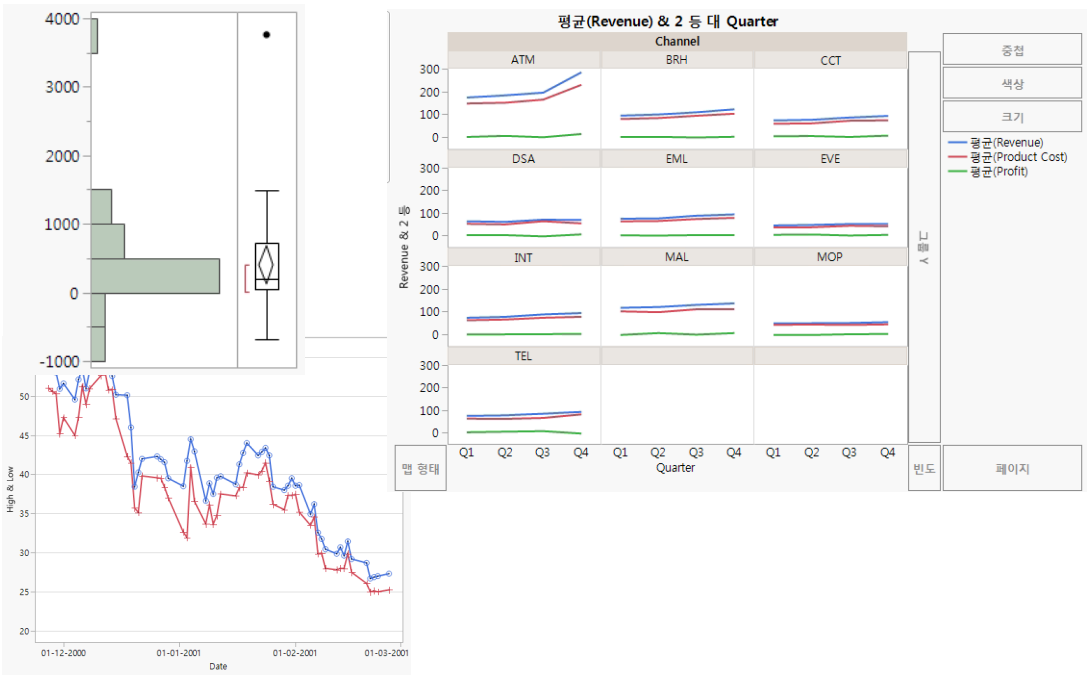
JMP(점프라고 발음)는 강력한 대화식 데이터 시각화 및 통계 분석 도구입니다. JMP에서 데이터 테이블, 그래프, 차트 및 보고서를 기반으로 분석을 수행하고 데이터와 상호 작용하면서 데이터에 대해 자세히 알아볼 수 있습니다.

JMP를 통해 연구자는 광범위한 통계 분석을 수행하고 모형을 생성할 수 있습니다. JMP는 데이터의 경향과 패턴을 신속하게 파악하려는 비즈니스 분석가에게도 유용합니다. JMP를 사용하면 통계 전문가가 아니라도 데이터에서 정보를 얻을 수 있습니다.

예를 들어 JMP를 사용하여 다음을 수행할 수 있습니다.

- 대화식 그래프 및 차트를 생성하여 데이터 탐색 및 관계 발견
- 한 번에 여러 변수에서 변동 패턴 발견
- 많은 양의 데이터 탐색 및 요약
- 미래를 예측할 수 있는 강력한 통계 모형 개발

그림 2.1 JMP 보고서의 예



목차

JMP 필수 개념 47
JMP 시작하기 47
JMP 시작 48
샘플 데이터 사용 49
데이터 테이블 이해 51
JMP 워크플로우 이해 52
1단계: JMP 플랫폼 시작 및 결과 보기 53
2단계: JMP 보고서에서 상자 그림 제거 54
3단계: 추가 JMP 출력 요청 55
4단계: JMP 플랫폼 결과에서의 상호 작용 56
JMP 와 Excel 의 차이점 57
데이터 테이블의 구조 57
JMP의 계산식 58
JMP 분석 및 그래프 생성 59

JMP 필수 개념

JMP 를 사용하기 전에 다음 개념을 잘 알아 두어야 합니다 .

- JMP 데이터 테이블을 사용하여 데이터 입력 , 확인 , 편집 및 조작
- **분석** , **그래프** 또는 **DOE** 메뉴에서 **플랫폼**선택 플랫폼에는 데이터를 분석하고 그래프로 작업하는 데 사용되는 대화식 창이 있습니다 .
- 플랫폼에서는 다음 창을 사용합니다 .
  - 시작 창 - 분석을 설정하고 실행합니다 .
  - 보고서 창 - 분석 결과를 보여 줍니다 .
- 보고서 창에는 일반적으로 다음 항목이 포함됩니다 .
  - 일부 유형의 그래프 (산점도 또는 차트 등)
  - 표시 아이콘 을 사용하여 표시하거나 숨길 수 있는 보고서
  - 빨간색 삼각형 메뉴 내에 있는 플랫폼 옵션

JMP 시작하기

JMP 의 일반적인 워크플로우는 간단합니다 .

1. 데이터를 JMP 로 가져옵니다 .
2. 플랫폼을 선택하고 시작 창을 완료합니다 .
3. 결과를 탐색하고 데이터가 제시하는 정보를 발견합니다 .

이 워크플로우는 "JMP 워크플로우 이해"(52 페이지)에서 자세히 설명합니다 .

JMP에서는 일반적으로 그래프를 사용하여 개별 변수 및 변수 간의 관계를 시각화하면서 작업을 시작합니다 . 그래프를 사용하면 이러한 정보를 쉽게 확인하고 더욱 심도 깊은 질문을 파악할 수 있습니다 . 그런 다음 분석 플랫폼을 사용하여 문제를 보다 심층적으로 탐구하고 해결책을 찾습니다 .

- " 데이터 작업 " 장 (61 페이지) 에서는 데이터를 JMP 로 가져오는 방법을 보여 줍니다 .
- " 데이터 시각화 " 장 (91 페이지) 에서는 JMP 에서 제공하는 유용한 그래프 중 일부를 사용하여 데이터를 보다 자세히 분석하는 방법을 보여 줍니다 .
- " 데이터 분석 " 장 (127 페이지) 에서는 일부 분석 플랫폼을 사용하는 방법을 보여 줍니다 .
- " 전체를 보는 시각 " 장 (163 페이지) 에서는 여러 플랫폼에서 분포 , 패턴 및 유사한 값을 분석하는 방법을 보여 줍니다 .

각 장에서는 예를 통해 방법을 안내합니다 . 이 장의 다음 섹션에서는 JMP에서 작업하는 데이터 테이블과 일반적인 개념에 대해 설명합니다 .

JMP 시작

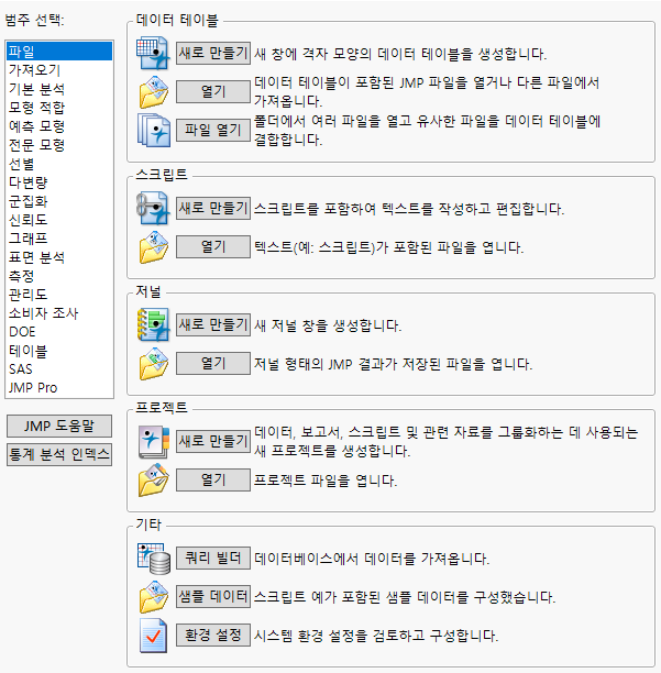
다음 두 가지 방법으로 JMP 를 시작하십시오 .

- 일반적으로 바탕 화면에 있는 JMP 아이콘을 두 번 클릭합니다 . 이렇게 하면 JMP 가 시작되지만 기존 JMP 파일은 열리지 않습니다 .
- 기존 JMP 파일을 두 번 클릭합니다 . 이렇게 하면 JMP 가 시작되고 파일이 열립니다 .

Windows 에서는 JMP 의 첫 화면에 "오늘의 유익한 정보" 창과 홈 창이 포함되고, macOS 에서는 "오늘의 유익한 정보", "JMP 시작하기" 및 홈 창이 처음에 나타납니다.

"JMP 시작하기" 창에서는 범주를 사용하여 작업 및 플랫폼을 분류합니다.

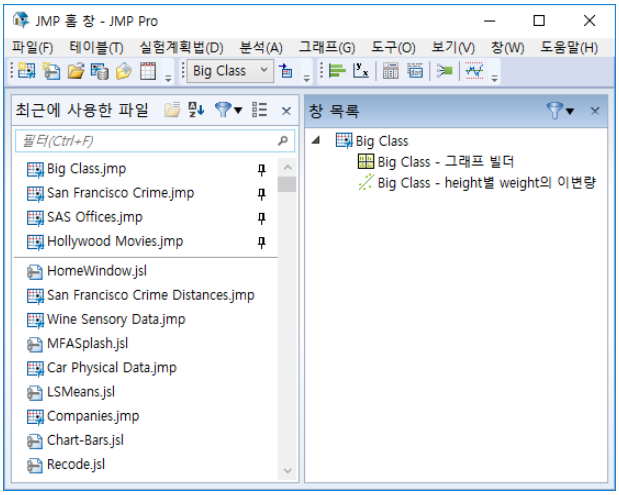
그림 2.2 JMP 시작하기



왼쪽에는 범주 목록이 있습니다. 범주를 클릭하면 해당 범주와 관련된 기능 및 명령을 볼 수 있습니다. "JMP 시작하기" 에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다.

홈 창에서는 JMP 에서 파일을 구성하고 액세스할 수 있도록 도와줍니다.

그림 2.3 Windows 의 홈 창



Windows 에서 홈 창을 열려면 **보기 > 홈 창**을 선택합니다 . macOS 에서는 **창 > JMP 홈**을 선택합니다 . 홈 창에는 다음에 대한 링크가 있습니다 .

- 현재 열려 있는 데이터 테이블 및 보고서 창
- 최근에 연 파일

홈 창에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

거의 모든 JMP 창에는 메뉴 표시줄과 도구 모음이 있습니다 . 대부분의 JMP 기능은 다음 세 가지 방법으로 찾을 수 있습니다 .

- 메뉴 표시줄 사용
- 도구 모음 버튼 사용
- "JMP 시작하기 " 창의 버튼 사용

메뉴 표시줄 및 도구 모음

메뉴 표시줄과 도구 모음은 많은 창에서 숨겨져 있습니다 . 표시하려면 창 제목 표시줄 아래에 있는 파란색 막대를 커서로 가리킵니다 . "JMP 시작하기 " 창 , 홈 창 및 모든 데이터 테이블의 메뉴는 항상 표시됩니다 .

샘플 데이터 사용

JMP 살펴보기와 기타 JMP 설명서의 예에서는 샘플 데이터 테이블을 사용합니다 . Windows 에서 샘플 데이터의 기본 위치는 다음과 같습니다 .

C:/Program Files/SAS/JMP/16/Samples/Data

C:/Program Files/SAS/JMPPro/16/Samples/Data

샘플 데이터 인덱스는 데이터 테이블을 범주별로 그룹화합니다. 표시 아이콘을 클릭하여 해당 범주의 데이터 테이블 목록을 표시한 후 링크를 클릭하여 데이터 테이블을 엽니다.

macOS 샘플 데이터는 /Library/Application Support/JMP/16/Samples/Data에 설치되어 있습니다.

JMP 샘플 데이터 테이블 열기

1. **도움말** 메뉴에서 **샘플 데이터**를 선택합니다.
2. **JMP 살펴보기에 사용된 데이터 테이블** 옆에 있는 표시 아이콘을 클릭하여 목록을 엽니다.
3. 데이터 테이블의 이름을 클릭하여 JMP 살펴보기의 예에서 사용합니다.

샘플 가져오기 데이터

다른 응용 프로그램의 파일을 사용하여 데이터를 JMP로 가져오는 방법을 알아봅니다.

Windows에서 샘플 가져오기 데이터의 기본 위치는 다음과 같습니다.

C:/Program Files/SAS/JMP/16/Samples/Import Data

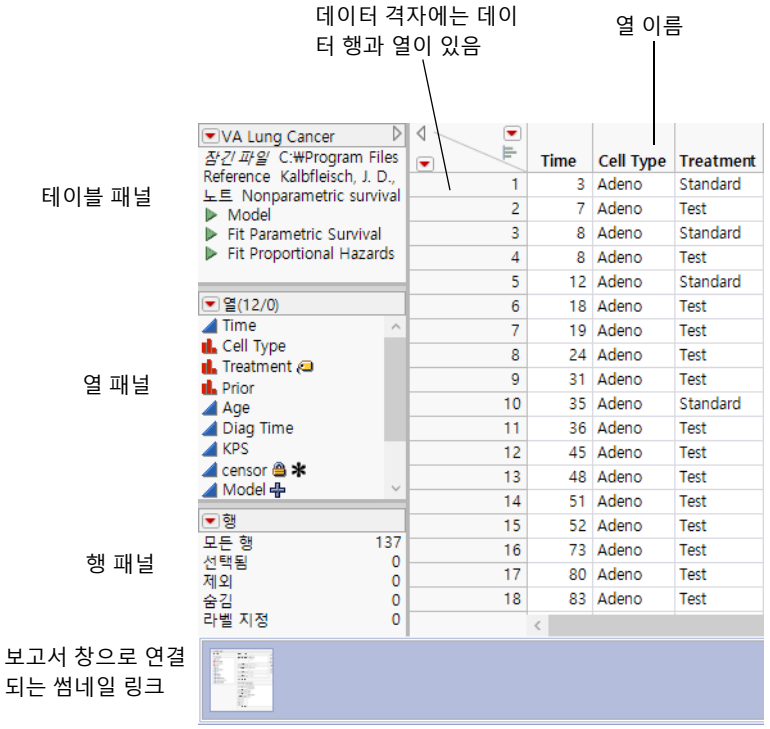
C:/Program Files/SAS/JMPPro/16/Samples/Import Data

데이터 테이블 이해

JMP 데이터 테이블은 행과 열로 구성된 데이터 모음입니다 . 데이터 테이블에는 노트 , 변수 및 스크립트와 같은 기타 정보가 포함될 수도 있습니다 . 이러한 보충 항목은 다음 장에서 설명합니다 .

"VA Lung Cancer" 데이터 테이블을 열고 여기에 설명된 데이터 테이블을 살펴보십시오 .

그림 2.4 데이터 테이블



데이터 테이블에는 다음과 같은 부분이 있습니다 .

데이터 격자 데이터 격자에는 행과 열로 정렬된 데이터가 있습니다 . 일반적으로 데이터 격자의 각 행은 관측값이며 열 (변수라고도 함) 은 관측값에 대한 정보를 제공합니다 . 그림 2.4 에서 각 행은 시험 대상에 해당하며 12 개의 정보 열이 있습니다 . 데이터 격자에는 12 개 열이 모두 표시될 수 없지만 " 열 " 패널에는 모든 열이 표시됩니다 . 각 시험 대상에 대한 정보로는 시간 , 세포 유형 , 치료법 등이 포함됩니다 . 각 열에는 머리글 또는 이름이 있습니다 . 이 이름은 테이블의 전체 행 수에 포함되지 않습니다 .

테이블 패널 테이블 패널에는 테이블 변수 또는 테이블 스크립트가 포함될 수 있습니다 . 그림 2.4 에는 자동으로 분석을 다시 생성할 수 있는 **Model** 이라는 저장된 스크립트가 하나 있습니다 . 또한 이 테이블에는 데이터에 대한 정보가 들어 있는 "Notes" 라는 변수가 있습니다 . 테이블 변수와 테이블 스크립트는 다음 장에서 설명합니다 .

열 패널 "열" 패널에는 열의 총 수, 열이 선택되었는지 여부 및 모든 열의 이름순 목록이 표시됩니다. 괄호 안의 숫자 (12/0) 는 12 개의 열이 있고 선택된 열이 없음을 보여 줍니다. 각 열 이름 왼쪽에 있는 아이콘은 해당 열의 모델링 유형을 표시합니다. 모델링 유형에 대해서는 "데이터 분석" 장의 "모델링 유형 이해"(132 페이지) 에서 설명합니다. 오른쪽 아이콘은 열에 할당된 속성을 표시합니다. 이러한 아이콘에 대한 자세한 내용은 "데이터 작업" 장의 "데이터 테이블에서 열 정보 보기 또는 변경"(75 페이지) 에서 확인하십시오.

행 패널 "행" 패널에는 데이터 테이블의 총 행 수와 선택된 행, 제외된 행, 숨겨진 행 또는 라벨이 지정된 행의 개수가 표시됩니다. 그림 2.4 의 데이터 테이블에는 137 개의 행이 있습니다.

보고서 창으로 연결되는 썸네일 링크 이 영역에는 데이터 테이블을 기반으로 하는 모든 보고서의 썸네일이 표시됩니다. 썸네일을 커서로 가리키면 보고서 창의 더 큰 미리보기를 볼 수 있습니다. 썸네일을 두 번 클릭하면 해당 보고서 창이 맨 앞에 표시됩니다.

행 및 열 추가, 데이터 입력 및 데이터 편집을 포함한 데이터 격자에서의 상호 작용에 대해서는 "데이터 작업" 장(61 페이지)에서 설명합니다. 여러 데이터 테이블을 열면 각 테이블이 별도의 창에 나타납니다.

JMP 데이터 테이블과 Excel 스프레드시트의 차이점에 대한 자세한 내용은 "JMP 와 Excel 의 차이점"(57 페이지)에서 확인하십시오.

JMP 워크플로우 이해

데이터가 데이터 테이블에 있으면 그래프 또는 그림을 생성하고 분석을 수행할 수 있습니다. 모든 기능은 플랫폼에 있으며 플랫폼은 주로 **분석** 또는 **그래프** 메뉴에 있습니다. 이를 플랫폼이라고 하는 것은 해당 기능이 단순한 정적 결과만 산출하는 것이 아니기 때문입니다. 플랫폼 결과는 보고서 창에 표시되고 대개 대화식이며 데이터 테이블과 연결되고 서로 간에도 연결됩니다.

분석 및 **그래프** 메뉴의 플랫폼은 다양한 분석 기능과 데이터 탐색 도구를 제공합니다.

그래프 또는 분석을 생성하는 일반적인 단계는 다음과 같습니다.

1. 데이터 테이블을 엽니다.
2. "그래프" 또는 "분석" 메뉴에서 플랫폼을 선택합니다.
3. 플랫폼 시작 창을 완료하여 분석을 설정합니다.
4. **확인**을 클릭하여 그래프 및 통계 분석이 포함된 보고서 창을 생성합니다.
5. 보고서 옵션을 사용하여 보고서를 사용자 정의합니다.
6. 결과를 저장하고, 내보내고, 다른 사람과 공유합니다.

이러한 개념에 대해서는 뒷부분 장에서 자세히 설명합니다.

다음 예에서는 네 단계로 간단한 분석을 수행하고 사용자 정의하는 방법을 보여 줍니다. 이 예에서는 Companies.jmp 파일의 샘플 데이터 테이블을 사용하여 Profits ($M) 변수의 기본 분석을 보여 줍니다.

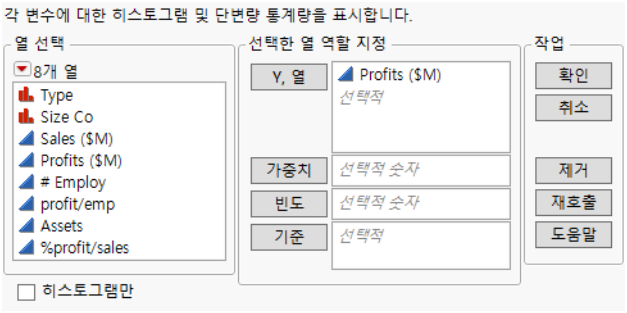
1 단계 : JMP 플랫폼 시작 및 결과 보기

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > 분포**를 선택하여 분포 시작 창을 엽니다 .
3. " 열 선택 " 상자에서 Profits ($M) 를 선택하고 **Y, 열** 버튼을 클릭합니다 .

   Profits ($M) 변수가 **Y, 열** 역할에 나타납니다 .

   변수를 할당하는 또 다른 방법은 " 열 선택 " 상자에서 열을 클릭하여 역할 상자로 드래그하는 것입니다 .

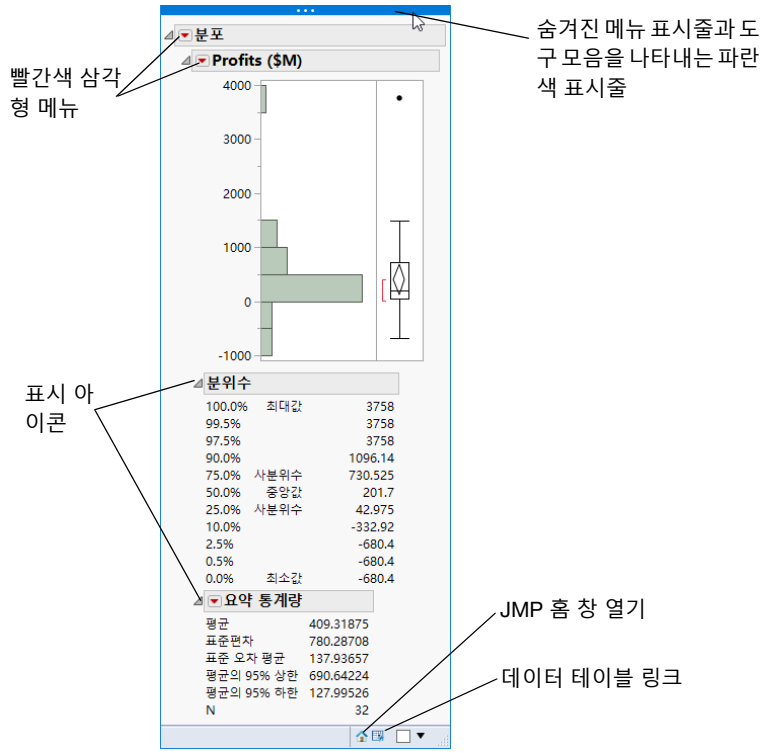
그림 2.5 Profits ($M) 할당



4. **확인**을 클릭합니다 .

   " 분포 " 보고서 창이 나타납니다 .

그림 2.6 Windows 의 분포 보고서 창



이 보고서 창에는 기본 그림 또는 그래프와 예비 분석 보고서가 포함됩니다. 결과는 개요 형식으로 표시되며 표시 아이콘을 클릭하여 보고서를 표시하거나 숨길 수 있습니다.

빨간색 삼각형 메뉴에는 언제든지 추가 그래프 및 분석을 요청할 수 있는 옵션과 명령이 포함되어 있습니다 .

- Windows에서 창 상단의 파란색 막대를 커서로 가리키면 메뉴 표시줄과 도구 모음이 표시됩니다 .
- Windows에서 오른쪽 하단에 있는 데이터 테이블 버튼을 클릭하면 이 보고서를 생성하는 데 사용된 데이터 테이블이 표시됩니다 . macOS 에서는 보고서 창 오른쪽 상단에 있는 **데이터 테이블 표시** 버튼을 클릭합니다 .
- Windows 에서 오른쪽 하단에 있는 **JMP 홈 창** 버튼을 클릭하면 홈 창이 표시됩니다 . macOS 에서는 **창 > JMP 홈**을 선택합니다 .

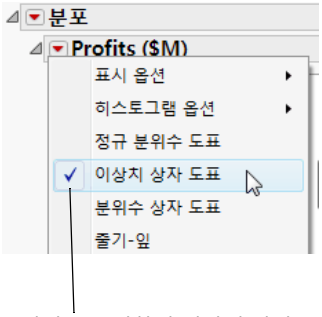
2 단계 : JMP 보고서에서 상자 그림 제거

앞서 생성한 " 분포 " 보고서를 계속 사용합니다 .

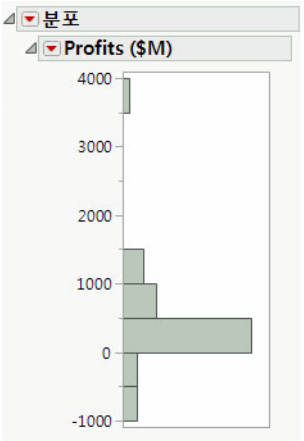
1. Profits ($M) 옆의 빨간색 삼각형을 클릭하여 보고서 옵션 메뉴를 표시합니다 .

2. **이상치 상자 그림**을 선택 취소하여 이 옵션을 해제합니다 .
   이상치 상자 그림이 보고서 창에서 제거됩니다 .

그림 2.7 이상치 상자 그림 제거



여기를 클릭하여 이상치 상자 그림을 제거합니다 .

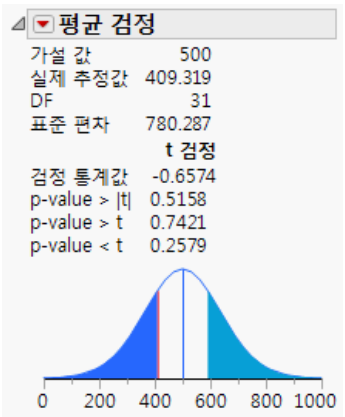


3 단계 : 추가 JMP 출력 요청

"2 단계 : JMP 보고서에서 상자 그림 제거 " 에서 생성한 같은 보고서 창을 계속 사용합니다 .

1. Profits ($M) 옆의 빨간색 삼각형을 클릭하고 **평균 검정**을 선택합니다 .
   " 평균 검정 " 창이 나타납니다 .
2. **가설 평균 지정** 상자에 "500" 을 입력합니다 .
3. **확인**을 클릭합니다 .
   평균에 대한 검정이 보고서 창에 추가됩니다 .

그림 2.8 평균 검정



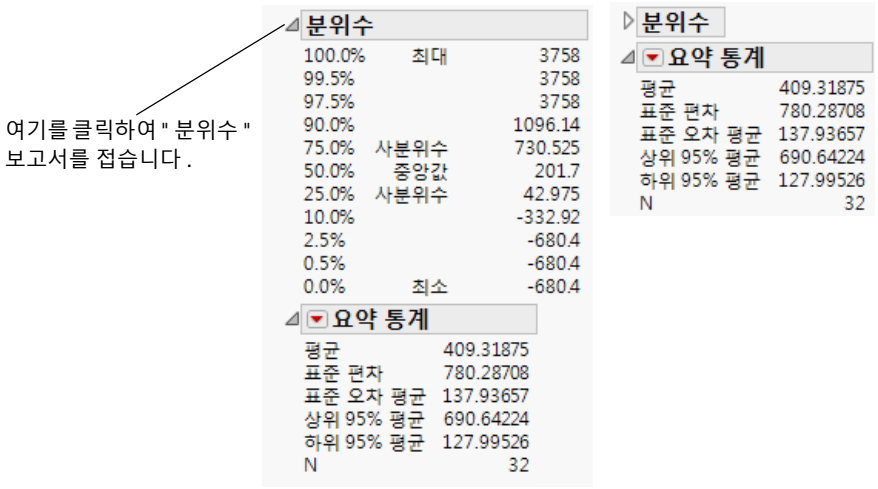
4 단계 : JMP 플랫폼 결과에서의 상호 작용

모든 플랫폼은 예를 들어 다음과 같은 대화식 결과를 생성합니다 .

- 보고서를 표시하거나 숨길 수 있습니다 .
- 용도에 맞게 그래프 및 통계 상세 정보를 추가하거나 제거할 수 있습니다 .
- 플랫폼 결과는 데이터 테이블과 연결되고 서로 간에도 연결됩니다 .

예를 들어 **분위수** 보고서를 닫으려면 **분위수** 옆의 표시 아이콘을 클릭합니다.

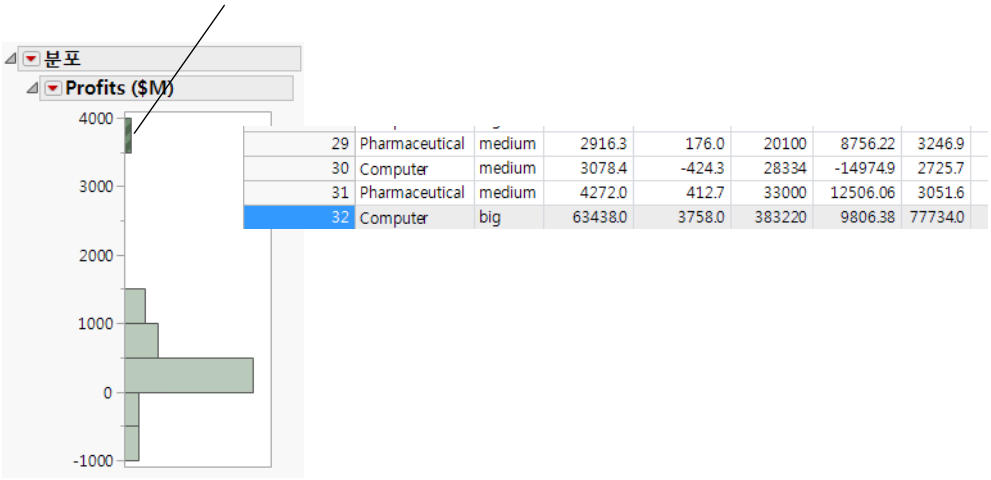
그림 2.9 분위수 보고서 닫기



플랫폼 결과는 데이터 테이블과 연결됩니다. 그림 2.10의 히스토그램에서는 한 회사 그룹이 다른 회사보다 훨씬 높은 수익을 올리고 있음을 보여 줍니다. 이 그룹을 빠르게 식별하려면 해당 히스토그램 막대를 클릭합니다. 데이터 테이블에서 해당 행이 선택됩니다.

그림 2.10 플랫폼 결과와 데이터 테이블 간의 연결

막대를 클릭하여 해당 행을 선택합니다.



이때 그룹에는 단 하나의 회사만 포함되며 해당 행 하나만 선택됩니다.

JMP 와 Excel 의 차이점

JMP는 데이터 테이블을 사용하는 통계 분석 프로그램입니다. Excel은 스프레드시트 응용 프로그램입니다. 데이터 테이블과 스프레드시트는 구조가 서로 다릅니다.

- " 데이터 테이블의 구조 "
- "JMP 의 계산식 "
- "JMP 분석 및 그래프 생성 "

데이터 테이블의 구조

데이터 테이블에는 고정된 행과 열이 있는 반면 스프레드시트는 셀 기반입니다. 스프레드시트에서는 모든 셀에 데이터, 머리글 또는 계산식을 포함할 수 있습니다. 데이터 테이블에서는 구조에 따라 분석용 데이터가 구성됩니다. 이 구조는 JMP 분석 및 그래프 플랫폼에 사용됩니다.

열 머리글 열 이름이 열 머리글입니다.

열 열에는 데이터가 포함되고 하나의 데이터 유형이 할당됩니다. 기본 열은 숫자 또는 문자 유형입니다. 열에 문자 및 숫자 데이터가 모두 포함되어 있을 때 전체 열의 데이터 유형은 문자이며 숫자는 문자 데이터로 처리됩니다. JMP 에는 이미지 등의 항목을 캡처하기 위한 특수한 열 유형도 있습니다. JMP 에서는 열의 데이터 유형에 따라 분석 옵션과 결과가 결정됩니다. 데이터 유형에 대한 자세한 내용은 " 데이터 분석 " 장의 " 모델링 유형 이해 "(132 페이지) 에서 확인하십시오.

행 행에는 관측값이 포함됩니다 . 행에 관측값이 없으면 해당 셀은 빈 채로 있습니다 . JMP 에서 점은 결측 숫자 값을 나타내고 공백은 결측 문자 값을 나타냅니다 .

JMP 데이터 테이블에 대한 자세한 내용은 "데이터 테이블 이해"(51 페이지) 에서 확인하십시오. JMP 열 특성에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

JMP 데이터 테이블은 Excel에서처럼 통합 문서 형식으로 구성할 수 없습니다. 각 JMP 데이터 테이블이 개별 파일이 되고 개별 창에 표시됩니다. 여러 테이블을 결합하는 방법은 JMP 사용의 에서 확인하십시오. JMP 테이블 및 출력을 구성하는 방법은 "작업 저장 및 공유" 장의 "스크립트 저장 및 실행"(191 페이지) 에서 확인하십시오.

팁 : 단일 분석에 두 개 이상의 테이블에 있는 데이터를 사용하려면 가상 결합 기능을 사용하십시오 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

JMP 의 계산식

스프레드시트에서는 계산식이 단일 셀에 적용되며 통합 문서의 다른 탭에 있는 셀을 비롯하여 스프레드시트의 모든 셀에 있는 데이터를 활용할 수 있습니다. 데이터 테이블에서는 계산식이 특정 열 전체에 적용되며 데이터 테이블에 있는 다른 열의 데이터를 사용할 수 있습니다. 열의 각 행에는 해당 행의 데이터를 기반으로 동일한 계산이 적용됩니다.

예를 들어 그림 2.11 에 표시된 것처럼 단순 합계를 계산하는 데이터 테이블이 있을 수 있습니다. height + weight 열에는 계산식이 있습니다. 이 계산식은 데이터 테이블의 각 행에 대해 height 값과 weight 값을 더합니다.

그림 2.11 계산식 열이 있는 데이터 테이블

	name	age	sex	height	weight	height + weight
1	KATIE	12	F	59	95	154
2	LOUISE	12	F	61	123	184
3	JANE	12	F	55	74	129
4	JACLYN	12	F	66	145	211
5	LILLIE	12	F	52	64	116
6	TIM	12	M	60	84	144
7	JAMES	12	M	61	128	189

JMP 계산식에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

팁 : 기본적인 열 요약 통계량이 필요하면 분포 플랫폼을 사용하십시오 . 자세한 내용은 기본 분석의 에서 확인하십시오 .

JMP 분석 및 그래프 생성

JMP에서는 데이터 분석 시 플랫폼을 사용합니다. 분석을 시작하려면 "분석" 메뉴로 이동합니다. 플랫폼 시작 창에서 분석할 변수를 선택하며, 분석 결과는 데이터 테이블 창이 아닌 별도의 보고서 창에 표시됩니다. 이는 분석 결과가 스프레드시트에 삽입되는 Excel과 다른 점입니다.

그래프 생성 옵션은 "그래프" 메뉴에 있습니다. 그래프 빌더를 사용하면 매우 간편하게 그래프 생성을 시작할 수 있습니다. 그래프 빌더에서 열을 드래그하여 놓고 빠르게 그래프를 생성하여 데이터를 탐색할 수 있습니다. 그래프 빌더에 대한 자세한 내용은 Essential Graphing의 에서 확인하십시오.

3 장

데이터 작업

그래프 생성 및 분석을 위한 데이터 준비

데이터를 그래프로 나타내거나 분석하려면 먼저 데이터 테이블에 적절한 형식으로 데이터가 준비되어야 합니다. 이 장에서는 다음을 포함하여 몇 가지 기본 데이터 관리 작업을 보여 줍니다.

- 새 데이터 테이블 생성
- 기존 데이터 테이블 열기
- 다른 응용 프로그램에서 JMP 로 데이터 가져오기
- 데이터 관리

그림 3.1 데이터 테이블의 예

Companies
잠긴 파일 C:₩Program File
노트 Selected Data on the

열(8/0)
Type
Size Co
Sales ($M)
Profits ($M)
Employ
profit/emp
Assets
%profit/sales

행
모든 행 32
선택됨 0
제외 0
숨김 0
라벨 지정 0

	Type	Size Co	Sales ($M)	Profits ($M)	# Employ	profit/emp	Assets	%profit/sales
1	Computer	small	855.1	31.0	7523	4120.70	615.2	3.63
2	Pharmaceutical	big	5453.5	859.8	40929	21007.11	4851.6	15.77
3	Computer	small	2153.7	153.0	8200	18658.54	2233.7	7.10
4	Pharmaceutical	big	6747.0	1102.2	50816	21690.02	5681.5	16.34
5	Computer	small	5284.0	454.0	12068	37620.15	2743.9	8.59
6	Pharmaceutical	big	9422.0	747.0	54100	13807.76	8497.0	7.93
7	Computer	small	2876.1	333.3	9500	35084.21	2090.4	11.59
8	Computer	small	709.3	41.4	5000	8280.00	468.1	5.84
9	Computer	small	2952.1	-680.4	18000	-378000	1860.7	-23.05
10	Computer	small	784.7	89.0	4708	18903.99	955.8	11.34
11	Computer	small	1324.3	-119.7	13740	-8711.79	1040.2	-9.04
12	Pharmaceutical	medium	4175.6	939.5	28200	33315.60	5848.0	22.50
13	Computer	big	11899.0	829.0	95000	8726.32	10075.0	6.97
14	Computer	small	873.6	79.5	8200	9695.12	808.0	9.10
15	Pharmaceutical	big	9844.0	1082.0	83100	13020.46	7919.0	10.99
16	Pharmaceutical	small	969.2	227.4	3418	66530.13	784.0	23.46
17	Pharmaceutical	medium	6698.4	1495.4	34400	43470.93	6756.7	22.32
18	Computer	big	5956.0	412.0	56000	7357.14	4500.0	6.92
19	Pharmaceutical	big	5903.7	681.1	42100	16178.15	8324.8	11.54
20	Computer	medium	2959.3	252.8	31404	8049.93	5611.1	8.54
21	Pharmaceutical	small	1198.3	86.5	8527	10144.25	1791.7	7.22

목차

JMP 로 데이터 가져오기 63
데이터 테이블에 데이터 복사 및 붙여넣기 63
데이터 테이블로 데이터 가져오기 63
데이터 테이블에 데이터 입력 66
Excel에서 JMP로 데이터 전송 68
데이터 테이블 작업 69
데이터 테이블에서 데이터 편집 70
데이터 테이블에서 값 선택, 선택 취소 및 찾기 72
데이터 테이블에서 열 정보 보기 또는 변경 75
계산식으로 값 계산의 예 77
보고서 데이터 필터링의 예 78
데이터 재구성의 예 80
요약 통계량 보기 80
부분집합 생성의 예 85
예: 데이터 테이블 결합 87
데이터 정렬 88

JMP 로 데이터 가져오기

JMP에서 제공하는 다양한 방법을 사용하여 JMP로 데이터를 가져올 수 있습니다.

- 다른 응용 프로그램의 데이터를 복사하여 붙여넣는 방법은 " 데이터 테이블에 데이터 복사 및 붙여넣기 "(63 페이지) 에서 확인하십시오 .
- 다른 응용 프로그램에서 데이터를 가져오는 방법은 " 데이터 테이블로 데이터 가져오기 "(63 페이지) 에서 확인하십시오 .
- 데이터를 직접 데이터 테이블에 입력하는 방법은 " 데이터 테이블에 데이터 입력 "(66 페이지) 에서 확인하십시오 .

데이터베이스에서 JMP로 데이터를 가져올 수도 있습니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

이 장에서는 JMP와 함께 설치되는 샘플 데이터 테이블과 샘플 가져오기 데이터를 사용합니다. 이러한 파일을 찾는 방법은 "JMP 소개" 장의 " 샘플 데이터 사용"(49 페이지) 에서 확인하십시오.

데이터 테이블에 데이터 복사 및 붙여넣기

Microsoft Excel 등의 다른 응용 프로그램이나 텍스트 파일에서 데이터를 복사하여 붙여넣는 방법으로 데이터를 JMP 로 이동할 수 있습니다 .

1. Microsoft Excel 에서 VA Lung Cancer.xlsx 파일을 엽니다 . 이 파일은 샘플 가져오기 데이터 폴더에 있습니다 .
2. 열 이름을 포함하여 모든 행과 열을 선택합니다 . 12 개의 열과 138 개의 행이 있습니다 .
3. 선택한 데이터를 복사합니다 .
4. JMP 에서 **파일 > 새로 만들기 > 데이터 테이블**을 선택하여 빈 테이블을 생성합니다 .
5. **편집 > 열 이름과 함께 붙여넣기**를 선택하여 데이터와 열 머리글을 붙여넣습니다 .

   JMP 에 붙여넣는 데이터에 열 이름이 없는 경우에는 **편집 > 붙여넣기**를 사용할 수 있습니다 .

데이터 테이블로 데이터 가져오기

Microsoft Excel, SAS 등의 다른 응용 프로그램이나 텍스트 파일에서 데이터를 가져오는 방법으로 데이터를 JMP 데이터 테이블로 이동할 수 있습니다 . 데이터를 가져오는 기본적인 단계는 다음과 같습니다 .

1. **파일 > 열기**를 선택합니다 .
2. 파일 위치로 이동합니다 .
3. 파일이 " 데이터 파일 열기 " 창에 표시되지 않으면 **파일 유형** 메뉴에서 올바른 파일 유형을 선택합니다 .
4. **열기**를 클릭합니다 .

예 : Microsoft Excel 파일 가져오기

1. **파일 > 열기**를 선택합니다 .
2. Samples/Import Data 폴더로 이동합니다 .
3. Team Results.xlsx 를 선택합니다 .

   데이터가 시작되는 행과 열에 유의하십시오 . 이 스프레드시트에는 두 개의 워크시트도 포함되어 있습니다 . 이 예에서는 "Ungrouped Team Results" 워크시트를 가져옵니다 .
4. **열기**를 클릭합니다 .

   스프레드시트가 "Excel 가져오기 마법사 " 에서 열리며 가져오기 옵션과 함께 데이터 미리보기가 나타납니다 .

   스프레드시트의 첫 번째 행에 있는 텍스트는 열 머리글입니다 . 하지만 스프레드시트의 3 행에 있는 텍스트를 열 머리글로 변환하려고 합니다 .
5. **열 머리글이 시작되는 행** 옆에 3 을 입력하고 Enter 키를 누릅니다 . 열 머리글이 데이터 미리보기에서 업데이트됩니다 . 데이터의 첫 번째 행에 대한 값이 4 로 업데이트됩니다 .
6. 이 워크시트에 대한 설정만 저장합니다 .
   - 창의 왼쪽 하단에서 **모든 워크시트에 사용**을 선택 취소합니다 .
   - 창의 오른쪽 상단에서 **Ungrouped Team Results** 를 선택합니다 .
7. **가져오기**를 클릭하여 스프레드시트를 지정한 대로 변환합니다 .

Excel 파일을 가져올 때 JMP 는 열 머리글이 있는지 여부와 열 이름이 1 행에 있는지 여부를 예측합니다 . 다음과 같은 경우에는 복사 및 붙여넣기 방법을 사용하는 것이 좋습니다 .

- 열 이름이 1 행 이외의 행에 있는 경우
- 파일에 열 이름이 없고 데이터가 1 행에서 시작하지 않는 경우
- 파일에 열 이름이 있고 데이터가 2 행에서 시작하지 않는 경우

Excel 파일 가져오기에 대한 자세한 내용은 JMP 사용의 "데이터 테이블에 데이터 복사 및 붙여넣기"(63 페이지) 및 에서 확인하십시오 .

예 : 텍스트 파일 가져오기

텍스트 파일을 가져오는 한 가지 방법은 JMP 에서 데이터 형식을 가정하고 데이터를 데이터 테이블에 배치하도록 하는 것입니다 . 이 방법을 사용할 경우 사용자가 환경 설정에서 지정한 설정이 사용됩니다 . 텍스트 가져오기 환경 설정을 지정하는 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

텍스트 파일을 가져오는 또 다른 방법은 텍스트 미리보기 창에서 데이터를 가져온 후 데이터 테이블이 표시되는 방식을 확인하고 조정하는 것입니다 . 다음 예에서는 텍스트 가져오기 미리보기 창을 사용하는 방법을 보여 줍니다 .

1. **파일 > 열기**를 선택합니다 .
2. Samples/Import Data 폴더로 이동합니다 .

3. Animals_line3.txt 를 선택합니다 .
4. " 열기 " 창 하단에서 **데이터 (미리보기 사용)** 를 선택합니다 .
5. **열기**를 클릭합니다 .

그림 3.2 초기 미리보기 창

이 텍스트 파일은 첫 번째 행에 제목이 있고 세 번째 행에 열 이름이 있으며 네 번째 행에서 데이터가 시작됩니다 . 이 파일을 JMP 에서 직접 열 때는 "Animals Data" 행이 첫 번째 열 이름이 되고 이후의 모든 열 이름과 데이터가 동기화되지 않습니다 . 미리보기 창을 사용하여 파일을 열기 전에 설정을 조정하고 조정 내용이 최종 데이터 테이블에 어떻게 영향을 미치는지 확인할 수 있습니다 .

6. **열 이름이 포함된 파일 . 행 번호** 필드에 3 을 입력합니다 .
7. **데이터가 시작되는 행** 필드에 4 를 입력합니다 .
8. **다음**을 클릭합니다 .

   두 번째 창에서는 가져오기에서 제외할 열을 지정하고 열의 데이터 모형을 변경할 수 있습니다 . 이 예에서는 기본 설정을 사용합니다 .
9. **가져오기**를 클릭합니다 .

새 테이블에는 species, subject, miles 및 season 이라는 열이 있습니다. species 및 season 열은 문자 데이터입니다. subject 및 miles 열은 연속형 숫자 데이터입니다.

팁 : 여러 개의 텍스트 파일을 한 번에 가져와 데이터 테이블을 생성할 수 있습니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

데이터 테이블에 데이터 입력

데이터 테이블에 직접 데이터를 입력할 수 있습니다. 다음 예에서는 몇 개월 동안 수집된 데이터를 데이터 테이블에 입력하는 방법을 보여 줍니다.

시나리오

표 3.1에서는 새로운 혈압약을 조사한 연구 자료를 보여 줍니다. 각 개인의 혈압을 6개월 동안 측정했습니다. 대조군 및 위약군과 함께 두 개의 투약 그룹("300mg" 및 "450mg")을 사용했습니다. 다음 데이터는 각 그룹의 평균 혈압을 보여 줍니다.

표 3.1 혈압 데이터

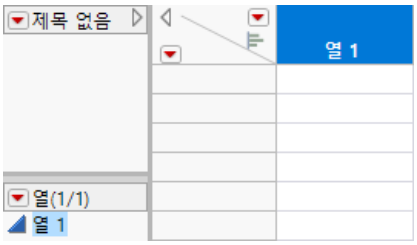
월	Control	Placebo	300mg	450mg
March	165	163	166	168
April	162	159	165	163
May	164	158	161	153
June	162	161	158	151
July	166	158	160	148
August	163	158	157	150

새 데이터 테이블에 데이터 입력

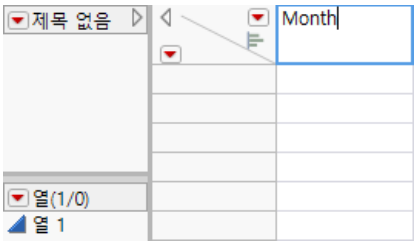
1. **파일 > 새로 만들기 > 데이터 테이블**을 선택하여 빈 데이터 테이블을 생성합니다 .
   새 데이터 테이블에는 하나의 열이 있고 행은 없습니다 .
2. 열 이름을 선택하고 이름을 Month 로 변경합니다 .

참고 : 열 이름을 바꾸려면 열 이름을 두 번 클릭하거나 , 열을 선택하고 Enter 키를 눌러도 됩니다 .

그림 3.3 열 이름 입력



열을 한 번 클릭하여 선택합니다 .

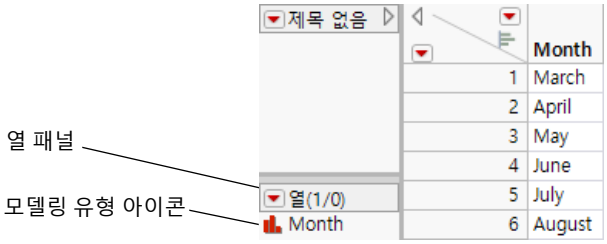


그런 다음 "Month" 를 입력합니다 .

3. **행 > 행 추가**를 선택합니다 .

   " 행 추가 " 창이 나타납니다 .
4. 6 개의 행을 추가하려고 하므로 "6" 을 입력합니다 .
5. **확인**을 클릭합니다 . 6 개의 빈 행이 데이터 테이블에 추가됩니다 .
6. 셀을 클릭하고 입력하는 방법으로 Month 정보를 입력합니다 .

그림 3.4 완료된 Month 열



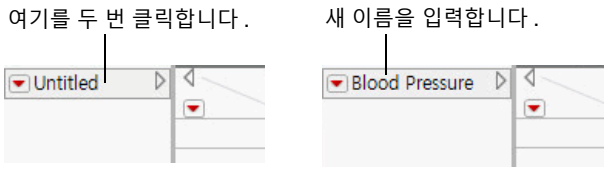
" 열 " 패널에서 열 이름 왼쪽에 있는 모델링 유형 아이콘을 확인합니다 . 이전에는 연속형이었던 Month 가 이제 명목형임을 반영하여 변경되었습니다 . 그림 3.3 의 "Column 1" 과 그림 3.4 의 "Month" 에 표시된 모델링 유형을 비교해 보십시오 . 이 차이점은 중요하므로 " 데이터 테이블에서 열 정보 보기 또는 변경 "(75 페이지) 에서 자세히 설명합니다 .

7. Control 열을 추가하기 위해 "Month" 열의 오른쪽에 있는 공간을 두 번 클릭합니다 .
8. 이름을 Control 로 변경합니다 .
9. 표 3.1 에 표시된 대로 Control 데이터를 입력합니다 . 데이터 테이블은 이제 6 개의 행과 2 개의 열로 구성되어 있습니다 .
10. 계속해서 표 3.1 에 표시된 대로 열을 추가하고 데이터를 입력하여 6 개의 행과 5 개의 열을 포함하는 최종 데이터 테이블을 생성합니다 .

데이터 테이블 이름 변경

1. 테이블 패널에서 데이터 테이블 이름 (" 제목 없음 ") 을 두 번 클릭합니다 .
2. 새 이름 ("Blood Pressure") 을 입력합니다 .

그림 3.5 데이터 테이블 이름 변경



Excel 에서 JMP 로 데이터 전송

Excel 용 JMP 추가기능을 사용하여 Excel 스프레드시트를 JMP 의 다음 항목으로 전송할 수 있습니다 .

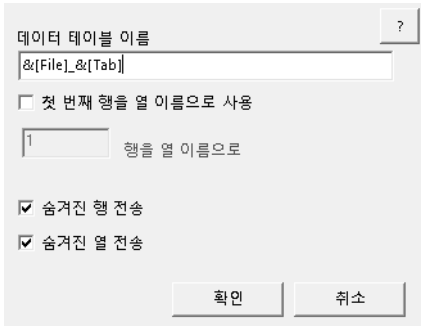
- 데이터 테이블
- 그래프 빌더
- 분포 플랫폼
- X 로 Y 적합 플랫폼
- 모형 적합 플랫폼
- 시계열 플랫폼
- 관리도 플랫폼

Excel 에서 JMP 추가기능 환경 설정 구성

JMP 추가기능 환경 설정을 구성하려면 다음을 수행하십시오 .

1. Excel 에서 **JMP > 환경 설정**을 선택합니다 .
   "JMP 환경 설정 " 창이 나타납니다 .

그림 3.6 JMP 추가기능 환경 설정



2. 기본 **데이터 테이블 이름** (파일 이름 _ 워크시트 이름) 을 그대로 사용하거나 이름을 입력합니다 .
3. 워크시트의 첫 번째 행에 열 머리글이 있으면 **첫 번째 행을 열 이름으로 사용**을 선택합니다 .

4. 첫 번째 행을 열 머리글로 사용하도록 선택한 경우 사용되는 행 수를 입력합니다 .
5. 워크시트에 있는 숨겨진 행을 JMP 데이터 테이블에 포함하려면 **숨겨진 행 전송**을 선택합니다 .
6. 워크시트에 있는 숨겨진 열을 JMP 데이터 테이블에 포함하려면 **숨겨진 열 전송**을 선택합니다 .
7. **확인**을 클릭하여 환경 설정을 저장합니다 .

JMP 로 이동

Excel 워크시트를 JMP 로 전송하려면 다음을 수행하십시오 .

1. Excel 파일을 엽니다 .
2. 전송할 워크시트를 선택합니다 .
3. **JMP** 를 선택한 후 다음 중에서 JMP 대상을 선택합니다 .
   - 데이터 테이블
   - 그래프 빌더
   - 분포 플랫폼
   - X 로 Y 적합 플랫폼
   - 모형 적합 플랫폼
   - 시계열 플랫폼
   - 관리도 플랫폼

   Excel 워크시트가 JMP에서 데이터 테이블로 열리고 선택한 플랫폼의 시작 창이 나타납니다.

데이터 테이블 작업

이 섹션에서는 데이터 테이블 작업에 대한 기본 개념을 설명합니다 .

- " 데이터 테이블에서 데이터 편집 "
- " 데이터 테이블에서 값 선택 , 선택 취소 및 찾기 "
- " 데이터 테이블에서 열 정보 보기 또는 변경 "
- " 계산식으로 값 계산의 예 "
- " 보고서 데이터 필터링의 예 "

팁 : " 일반 " 환경 설정에서 **자동 저장 타임아웃** 값을 설정하여 열려 있는 데이터 테이블을 지정된 시간 (분) 마다 자동으로 저장하는 것이 좋습니다 . 이 자동 저장 값은 저널 , 스크립트 , 프로젝트 및 보고서에도 적용됩니다 .

데이터 테이블에서 데이터 편집

한 번에 몇 개의 셀 또는 전체 열에 대해 데이터를 입력하거나 변경할 수 있습니다 . 이 섹션에는 다음과 같은 정보가 포함되어 있습니다 .

- " 데이터 테이블 셀의 값 변경 "
- " 재코딩 "
- " 패턴 있는 데이터 생성 "

데이터 테이블 셀의 값 변경

값을 변경하려면 셀을 선택하고 변경 내용을 입력합니다. 셀을 두 번 클릭하여 편집할 수도 있습니다.

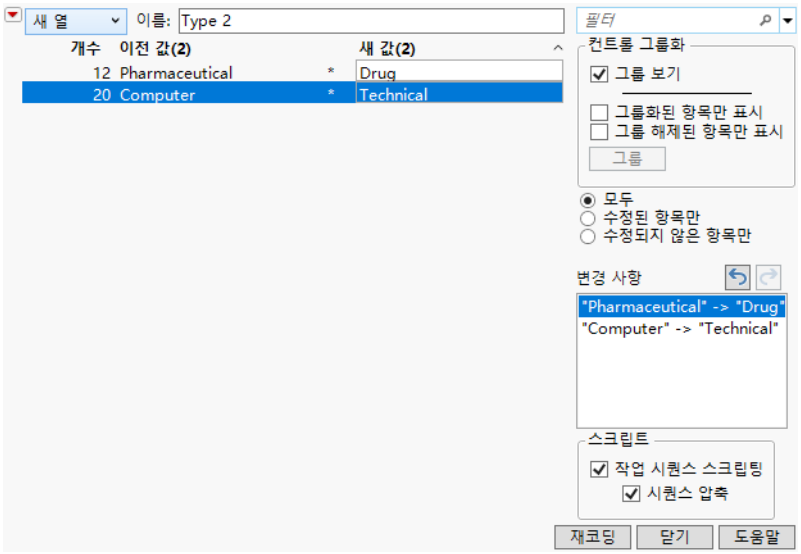
참고 : 셀을 두 번 클릭하는 것은 셀을 선택하는 것과 같지 않습니다 . 한 번 클릭하면 셀이 선택됩니다 . 한 번에 둘 이상의 셀을 선택할 수 있으며 선택한 셀에서 특정 작업을 수행할 수 있습니다 . 두 번 클릭해야 셀을 편집할 수 있습니다 . 행 , 열 및 셀 선택에 대한 자세한 내용은 " 데이터 테이블에서 값 선택 , 선택 취소 및 찾기 "(72 페이지) 에서 확인하십시오 .

재코딩

" 재코딩 " 도구를 사용하여 열의 모든 값을 한 번에 변경할 수 있습니다 . 예를 들어 컴퓨터 회사와 제약 회사의 매출을 비교하는 데 관심이 있다고 가정해 보겠습니다 . 현재 회사 라벨은 "Computer" 와 "Pharmaceutical" 입니다 . 이를 "Technical" 및 "Drug" 로 변경하려고 합니다 . 32 개 행의 모든 데이터를 검토하고 모든 값을 변경하는 것은 지루하고 비효율적이며 오류가 발생하기 쉽습니다 . 더 많은 데이터 행이 있을 때 특히 그렇습니다 . 재코딩이 더 좋은 방법입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. Type 열의 머리글을 한 번 클릭하여 해당 열을 선택합니다 .
3. **열 > 재코딩**을 선택합니다 .
4. " 재코딩 " 창의 " 새 값 " 열에서 "Computer" 행에 "Technical" 을 입력하고 "Pharmaceutical" 행에 "Drug" 를 입력합니다 .
5. " 새 열 " 목록에서 **현재 위치**를 선택합니다 .
6. **재코딩**을 클릭합니다 .

그림 3.7 재코딩 창



모든 셀이 자동으로 새 값으로 업데이트됩니다.

패턴 있는 데이터 생성

"채우기" 옵션을 사용하여 열을 패턴 있는 데이터로 채울 수 있습니다. "채우기" 옵션은 데이터 테이블이 커서 각 행에 값을 입력하는 것이 번거로울 때 특히 유용합니다.

예 : 패턴으로 열 채우기

1. 새 열을 추가합니다 .
2. 첫 번째 셀에는 "1" 을 , 두 번째 셀에는 "2" 를 , 세 번째 셀에는 "3" 을 입력합니다 .
3. 세 개의 셀을 선택하고 선택한 셀의 아무 곳이나 마우스 오른쪽 버튼으로 클릭하여 메뉴를 표시합니다 .
4. **채우기 > 테이블 끝까지 시퀀스 반복**을 선택합니다 .

   열의 나머지가 해당 시퀀스로 채워집니다 (1, 2, 3, 1, 2, 3, ...).

시퀀스를 반복하지 않고 패턴을 계속하려면 (1, 2, 3, 4, 5, 6, ...) **테이블 끝까지 시퀀스 계속**을 선택합니다. 이 명령을 사용하여 1, 1, 1, 2, 2, 2, 3, 3, 3, ... 같은 패턴을 생성할 수도 있습니다.

"채우기" 옵션은 간단한 산술 및 기하 시퀀스를 인식할 수 있습니다. 문자 데이터의 경우 "채우기" 옵션은 단순히 값을 반복합니다.

데이터 테이블에서 값 선택 , 선택 취소 및 찾기

데이터 테이블 내에서 행 , 열 또는 셀을 선택할 수 있습니다 . 예를 들어 기존 데이터 테이블의 부분집합을 생성하려면 먼저 테이블에서 부분집합을 만들려는 부분을 선택해야 합니다 . 또한 행을 선택하면 그래프에서 데이터 점이 쉽게 눈에 띄게 할 수 있습니다 . 수동으로 행과 열을 클릭하는 방법으로 선택하거나 , 특정 검색 기준을 충족하는 행을 선택하십시오 . 이 섹션에는 다음과 같은 정보가 포함되어 있습니다 .

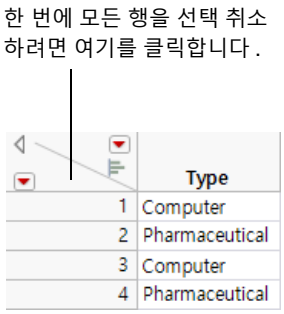
- " 행 선택 및 선택 취소 "(72 페이지)
- " 열 선택 및 선택 취소 "(73 페이지)
- " 셀 선택 및 선택 취소 "(74 페이지)
- " 값 검색 "(74 페이지)

행 선택 및 선택 취소

표 3.2 행 선택 및 선택 취소

작업	작업
한 번에 하나씩 행 선택	행 번호를 클릭합니다.
인접한 여러 행 선택	클릭하고 드래그하여 행 번호를 선택합니다. 또는 시작 행을 선택하고 Shift 키를 누른 후 마지막 행 번호를 클릭합니다.
인접하지 않은 여러 행 선택	첫 번째 행을 선택하고 Ctrl 키를 누른 후 다른 행 번호를 클릭합니다.
한 번에 하나씩 행 선택 취소	Ctrl 키를 누르고 행 번호를 클릭합니다.
모든 행 선택 취소	테이블의 왼쪽 상단에서 아래쪽 삼각형 공간을 클릭합니다(그림 3.8).

그림 3.8 행 선택 취소

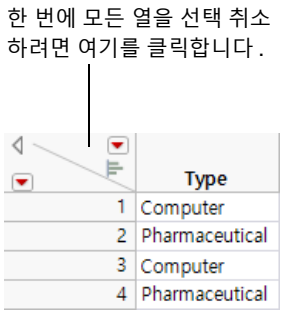


열 선택 및 선택 취소

표 3.3 열 선택 및 선택 취소

작업	작업
한 번에 하나씩 열 선택	열 머리글을 클릭합니다.
인접한 여러 열 선택	클릭하고 드래그하여 여러 열 머리글을 선택합니다. 또는 시작 열을 선택하고 Shift 키를 누른 후 마지막 머리글을 클릭합니다.
인접하지 않은 여러 열 선택	첫 번째 열을 선택하고 Ctrl 키를 누른 후 다른 열 머리글을 클릭합니다.
한 번에 하나씩 열 선택 취소	Ctrl 키를 누르고 열 머리글을 클릭합니다.
모든 열 선택 취소	테이블의 왼쪽 상단에서 위쪽 삼각형 공간을 클릭합니다(그림 3.9).

그림 3.9 열 선택 취소



셀 선택 및 선택 취소

표 3.4 셀 선택 및 선택 취소

작업	작업
한 번에 하나씩 셀 선택	각 셀을 개별적으로 클릭합니다.
인접한 여러 셀 선택	클릭하고 드래그하여 여러 셀을 선택합니다. 또는 시작 셀을 선택하고 Shift 키를 누른 후 마지막 셀을 클릭합니다.
인접하지 않은 여러 셀 선택	첫 번째 셀을 선택하고 Ctrl 키를 누른 후 다른 셀을 클릭합니다.
셀 선택 취소	테이블의 왼쪽 상단에서 위쪽 및 아래쪽 삼각형 공간을 클릭합니다.

값 검색

수천 또는 수만 개의 행이 있는 데이터 테이블에서는 테이블을 스크롤하여 특정 셀을 찾는 것이 어려울 수 있습니다. 특정 정보를 찾으려면 검색 기능을 사용하십시오. 데이터가 검색 기준과 매칭되면 해당 셀이 선택되고 창에 해당 셀이 표시되도록 데이터 격자가 스크롤됩니다. 예를 들어 Companies.jmp 데이터 테이블에는 총 매출이 11,899 달러인 회사에 대한 정보가 있습니다. 해당 셀을 찾으려면 검색 기능을 사용하십시오.

예 : 값 검색

1. **편집 > 검색 > 찾기**를 선택하여 검색 창을 시작합니다.
2. **찾을 내용** 상자에 "11899" 를 입력합니다.

3. **찾기**를 클릭합니다 . 11,899 가 포함된 첫 번째 셀이 검색되고 선택됩니다 .

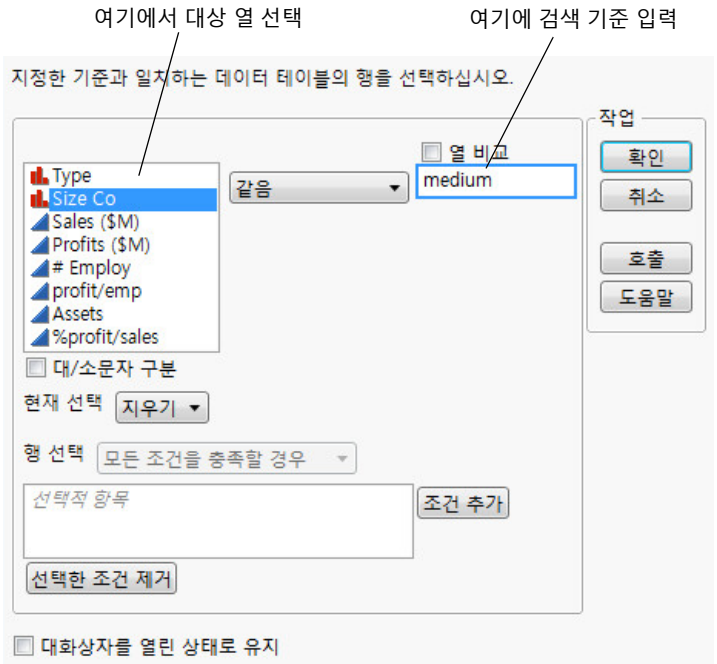
여러 셀이 검색 기준을 충족하면 **찾기**를 다시 클릭하여 검색어와 매칭되는 다음 셀을 찾습니다.

기준과 매칭되는 여러 행을 한 번에 검색할 수도 있습니다.

예 : 중간 규모 회사에 해당하는 모든 행 선택

1. **행 > 행 선택 > 선택 조건**을 선택하여 **행 선택하기** 창을 엽니다 .
2. 왼쪽의 열 목록 상자에서 Size Co 를 선택합니다 .
3. 오른쪽의 텍스트 상자에 "medium" 을 입력합니다 .
4. **확인**을 클릭합니다 .

그림 3.10 행 선택하기 창



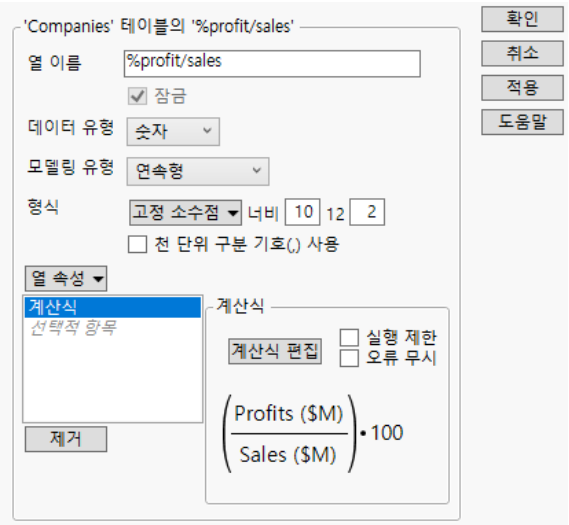
Size Co가 "medium"인 행이 모두 선택됩니다. 7개가 있습니다.

데이터 테이블에서 열 정보 보기 또는 변경

데이터 테이블 열에 대한 정보는 열의 데이터로 제한되지 않습니다. 데이터 유형, 모델링 유형, 형식 및 계산식도 설정할 수 있습니다.

열 특성을 보거나 변경하려면 열 머리글을 두 번 클릭합니다. 또는 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다. "열 정보" 창이 나타납니다.

그림 3.11 열 정보 창



열 이름 열 이름을 입력하거나 변경합니다. 두 열의 이름이 같을 수 없습니다.

데이터 유형 다음 데이터 유형 중 하나를 선택합니다.

숫자 열 값을 숫자로 지정합니다.

문자 열 값을 문자 또는 기호와 같이 숫자가 아닌 값으로 지정합니다.

행 상태 열 값을 행 상태로 지정합니다. 이에 대해서는 자세한 설명이 필요합니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

모델링 유형 모델링 유형은 값이 분석에서 사용되는 방식을 정의합니다. 다음 모델링 유형 중 하나를 선택합니다.

연속형 숫자 유형에만 해당됩니다.

순서형 숫자 또는 문자 값으로, 순서가 있는 범주입니다.

명목형 숫자 또는 문자 값이지만 순서가 없습니다.

형식 숫자 값의 형식을 선택합니다. 문자 데이터에는 이 옵션을 사용할 수 없습니다. 다음은 가장 일반적인 형식 중 몇 가지입니다.

최적 최적의 표시 형식이 자동으로 선택됩니다.

고정 소수점 표시되는 소수 자릿수를 지정합니다.

날짜 날짜 값의 구문을 지정합니다.

시간 시간 값의 구문을 지정합니다.

통화 통화 값에 사용되는 통화 유형과 소수 자릿수를 지정합니다.

열 특성 계산식, 노트 및 값 순서와 같은 특수한 열 특성을 설정합니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

잠금 열의 값을 변경할 수 없도록 열을 잠급니다.

계산식으로 값 계산의 예

계산식 편집기를 사용하여 계산된 값이 포함된 열을 생성할 수 있습니다.

시나리오

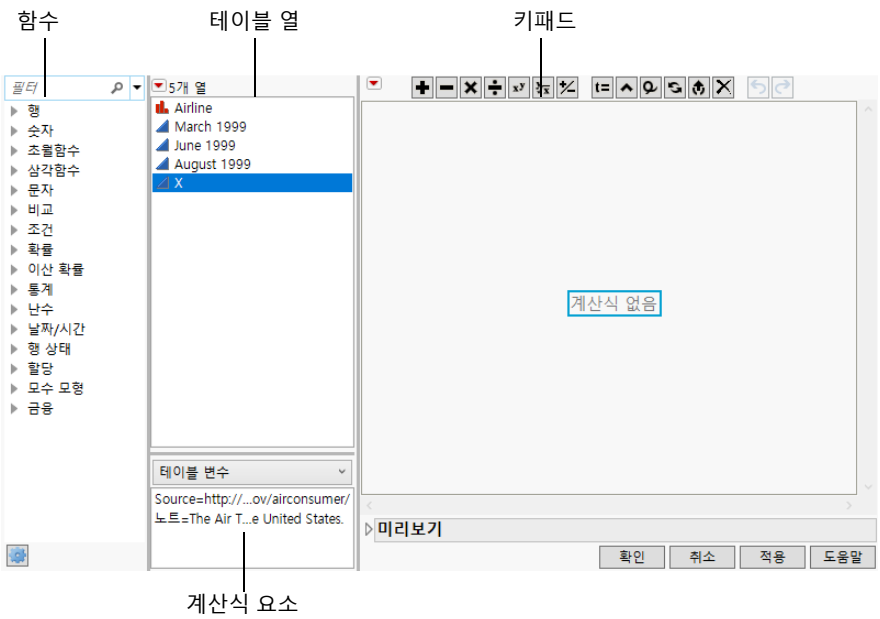
샘플 데이터 테이블 On-Time Arrivals.jmp는 여러 항공사의 정시 도착률을 나타냅니다. 수집된 데이터는 1999년 3월, 6월, 8월의 데이터입니다.

계산식 생성

각 항공사의 평균 정시 도착률을 포함하는 새 열을 생성하려고 한다고 가정해 보겠습니다.

1. 새 열을 추가합니다.
2. 새 열의 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택합니다. 계산식 편집기 창이 나타납니다.

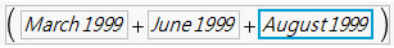
그림 3.12 계산식 편집기



각 항공사의 평균 정시 도착률에 대한 계산식을 생성합니다 .

3. 열 목록에서 March 1999 를 선택합니다 .
4. 키패드에서 ✚ 버튼을 클릭합니다 .
5. June 1999 를 선택한 후 다시 ✚ 기호를 클릭합니다 .
6. August 1999 를 선택합니다 .

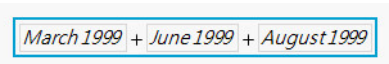
그림 3.13 모든 월의 합계



현재는 "August 1999" 만 선택되어 있습니다 (파란색 상자로 둘러싸임).

7. 전체 계산식을 둘러싼 상자를 클릭합니다 .

그림 3.14 전체 계산식이 선택됩니다 .



8. ➗ 버튼을 클릭합니다 .
9. 분모 상자에 "3" 을 입력한 다음 계산식의 바깥쪽 공백을 클릭합니다 .

그림 3.15 완료된 계산식

10. **확인**을 클릭합니다 .

새 열에 평균이 포함됩니다 .

계산식 편집기에는 많은 기본 제공 산술 함수와 통계 함수가 있습니다. 예를 들어 평균 정시 도착률을 계산하는 또 다른 방법은 통계 함수 목록에서 **Mean** 함수를 사용하는 것입니다. 계산식 편집기의 모든 함수에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

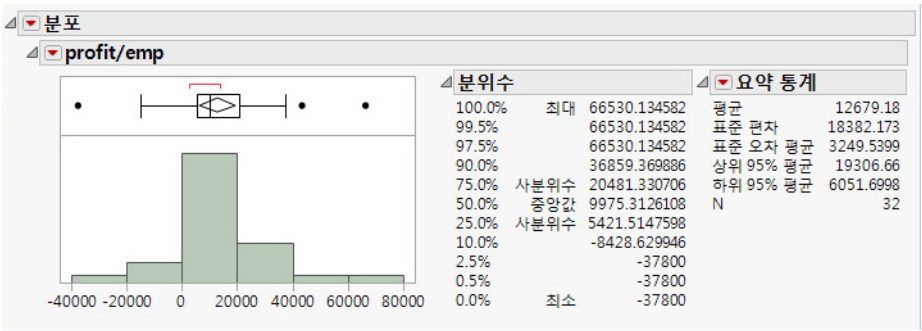
보고서 데이터 필터링의 예

데이터 필터를 사용하여 복잡한 데이터 부분집합을 대화식으로 선택하거나 , 이러한 부분집합을 그림에서 숨기거나 , 분석에서 제외할 수 있습니다 . 한 예로 컴퓨터 회사와 제약 회사의 직원당 수익을 살펴보겠습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. profit/emp 를 선택하고 **Y, 열**을 클릭합니다 .

4. **확인**을 클릭합니다 .
5. "profit/emp" 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 가로 레이아웃**을 선택합니다 .

그림 3.16 profit/emp 의 분포



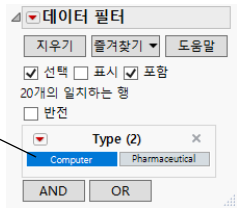
6. " 분포 " 의 빨간색 삼각형에서 **다시 실행 > 자동 재계산**을 선택하여 자동 재계산 기능을 설정합니다 .

   이 옵션을 설정하면 점을 숨기거나 제외하는 등의 변경 작업을 수행할 때마다 보고서 창이 자동으로 업데이트됩니다 .
7. 데이터 테이블에서 **행 > 데이터 필터**를 선택합니다 .
8. Type 을 선택하고 **추가**를 클릭합니다 .
9. " 선택 " 체크박스가 선택되어 있는지 확인합니다 .
10. 분포 결과에서 제약 회사를 제외하고 컴퓨터 회사만 포함하려면 데이터 필터 창에서 **Computer** 상자를 클릭합니다 .

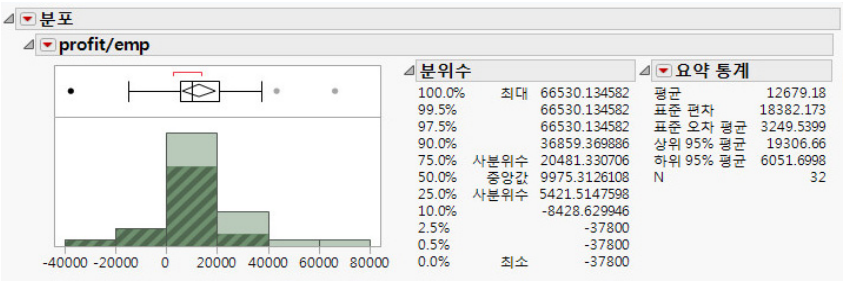
    분포 결과가 업데이트되어 컴퓨터 회사만 포함됩니다 .

그림 3.17 컴퓨터 회사에 대한 필터

분포 결과에 컴퓨터 회사만 포함하려면 "Computer" 상자를 클릭합니다 .



그래프 및 통계량 보고서에 선택된 행만 자동으로 반영됩니다 .



반대로 제약 회사만 포함되도록 분포 결과를 변경하려면 " 데이터 필터 " 창에서 **Pharmaceutical** 상자를 클릭합니다.

데이터 재구성의 예

테이블 메뉴의 명령과 **분석** 메뉴의 " 테이블 생성 " 명령은 그래프 생성 및 분석에 필요한 형식으로 데이터 테이블을 요약하고 조작합니다 . 이 섹션에서는 다음 5 가지 명령에 대해 설명합니다 .

요약 데이터를 설명하는 요약 통계량을 포함하는 테이블을 생성합니다 .

테이블 생성 요약 통계량을 생성하기 위한 드래그앤드롭 작업 공간을 제공합니다 .

부분집합 데이터의 부분집합을 포함하는 테이블을 생성합니다 .

결합 두 데이터 테이블의 데이터를 하나의 새 데이터 테이블로 결합합니다 .

정렬 하나 이상의 열을 기준으로 데이터를 정렬합니다 .

이러한 명령과 기타 " 테이블 " 메뉴 명령에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

요약 통계량 보기

합 및 평균과 같은 요약 통계량은 데이터에 대한 유용한 정보를 즉시 제공합니다. 예를 들어 32개 회사의 개별 연간 수익만 봐서는 소규모, 중간 규모, 대규모 기업의 수익을 비교하기가 어렵습니다. 요약에서는 이러한 정보를 즉시 보여 줍니다.

요약 테이블을 생성하려면 **요약** 또는 **테이블 생성** 명령을 사용합니다. **요약** 명령을 사용하면 새 데이터 테이블이 생성됩니다. 다른 데이터 테이블과 마찬가지로 요약 테이블에서도 분석을 수

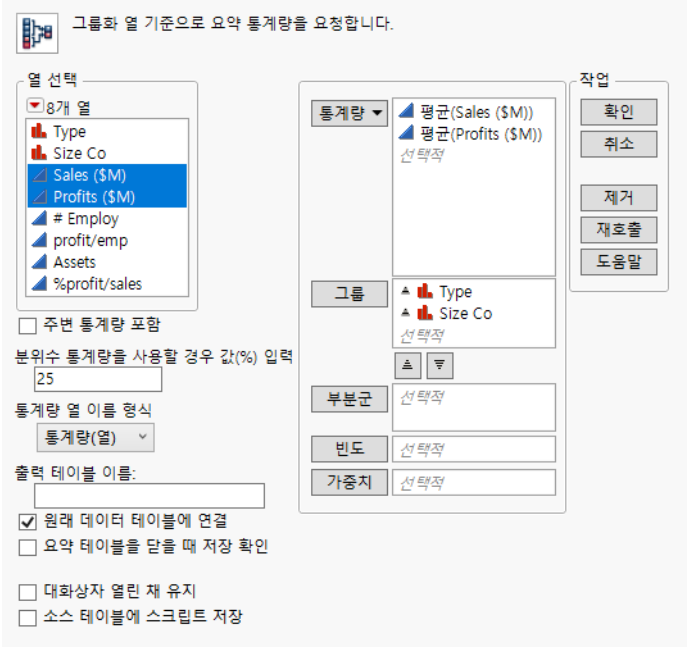
행하고 그래프를 생성할 수 있습니다. **테이블 생성** 명령을 사용하면 요약 데이터 테이블이 있는 보고서 창이 생성됩니다. "테이블 생성" 보고서에서 테이블을 생성할 수도 있습니다.

요약 테이블 예제

요약 테이블에는 그룹화 변수의 각 수준에 대한 통계량이 포함됩니다. 예를 들어 컴퓨터 및 제약 회사의 재무 데이터를 살펴보십시오. 회사 유형 및 규모의 각 조합에 대해 매출 평균 및 수익 평균을 계산한다고 가정해 보겠습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다.
2. **테이블 > 요약**을 선택합니다.
3. Type 및 Size Co 를 선택하고 **그룹**을 클릭합니다.
4. Sales ($M) 및 Profits ($M) 를 선택하고 **통계량 > 평균**을 클릭합니다.

그림 3.18 완료된 요약 창



5. **확인**을 클릭합니다.

Type 및 Size Co의 각 조합에 대해 Sales ($M) 평균과 Profit ($M) 평균이 계산됩니다.

그림 3.19 요약 테이블

(Type, Size Co) 기...
소스
열(5/0)
Type
Size Co
행 수
평균(Sales ($M))
평균(Profits ($M))

	Type	Size Co	행 수	평균(Sales ($M))	평균(Profits ($M))
1	Computer	big	4	20597.48	1089.93
2	Computer	medium	2	3018.85	-85.75
3	Computer	small	14	1758.06	44.94
4	Pharmaceutical	big	5	7474.04	894.42
5	Pharmaceutical	medium	5	4261.06	698.98
6	Pharmaceutical	small	2	1083.75	156.95

요약 테이블의 내용은 다음과 같습니다 .

- 각 그룹화 변수 (이 예의 경우 Type 및 Size Co) 에 대한 열이 있습니다 .
- **행 수** 열에는 그룹화 변수의 각 조합에 해당하는 원래 테이블의 행 수가 표시됩니다 . 예를 들어 원래 데이터 테이블에는 소규모 컴퓨터 회사에 해당하는 14 개 행이 있습니다 .
- 요청된 각 요약 통계량에 대한 열이 있습니다 . 이 예에는 Sales ($M) 의 평균 열과 Profits ($M) 의 평균 열이 있습니다 .

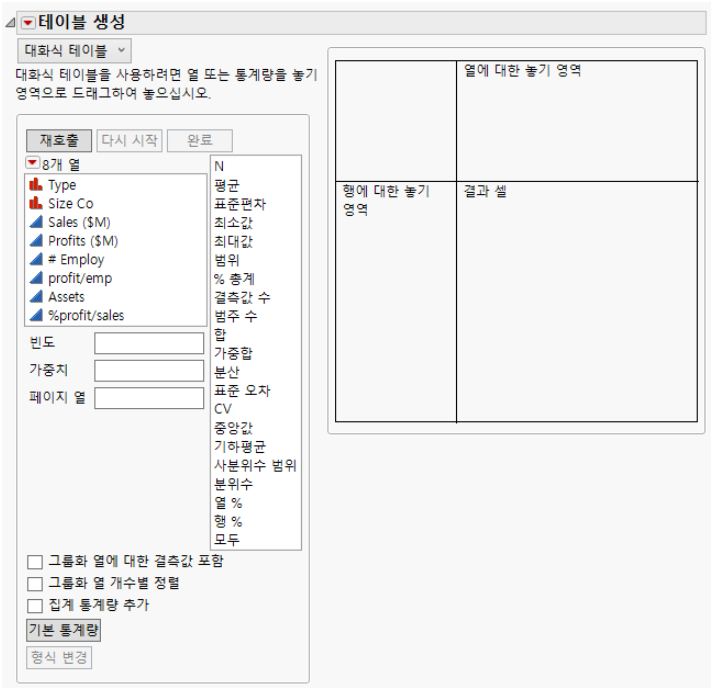
요약 테이블은 소스 테이블에 연결됩니다. 요약 테이블에서 행을 선택하면 소스 테이블에서도 해당 행이 선택됩니다.

테이블 생성 예제

" 테이블 생성 " 명령을 사용하여 열을 작업 공간으로 드래그하면 그룹화 변수의 각 조합에 대한 요약 통계량을 생성할 수 있습니다 . 이 예에서는 방금 " 요약 " 명령을 사용하여 생성한 것과 동일한 요약 정보를 " 테이블 생성 " 명령으로 생성하는 방법을 보여 줍니다 .

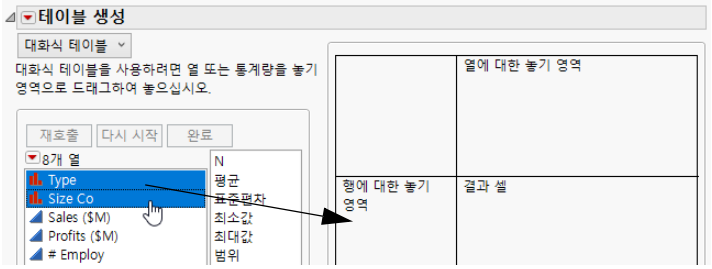
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .

그림 3.20 테이블 생성 작업 공간



3. Type 과 Size Co 를 모두 선택합니다 .
4. 선택한 항목을 **행에 대한 놓기 영역**으로 드래그하여 놓습니다 .

그림 3.21 열을 행 영역으로 드래그



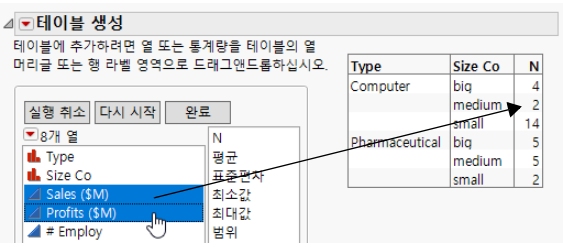
5. 머리글을 마우스 오른쪽 버튼으로 클릭하고 **그룹화 열 내포**를 선택합니다 .
   초기 테이블은 그룹당 행 수를 보여 줍니다 .

그림 3.22 초기 테이블

Type	Size Co	개수
Computer	big	4
	medium	2
	small	14
Pharmaceutical	big	5
	medium	5
	small	2

6. Sales ($M) 및 Profits ($M) 를 모두 선택한 후 테이블의 **개수**로 드래그하여 놓습니다 .

그림 3.23 매출 및 수익 추가



이제 테이블에는 그룹당 Sales ($M) 합과 Profits ($M) 합이 표시됩니다 .

그림 3.24 합 테이블

		Sales ($M)	Profits ($M)
Type	Size Co	합계	합계
Computer	big	82389.9	4359.7
	medium	6037.7	-171.5
	small	24612.8	629.1
Pharmaceutical	big	37370.2	4472.1
	medium	21305.3	3494.9
	small	2167.5	313.9

7. 마지막 단계는 합을 평균으로 변경하는 것입니다 . 둘 중 하나의 **합**을 마우스 오른쪽 버튼으로 클릭하고 **통계량** > **평균**을 선택합니다 .

그림 3.25 최종 테이블

		Sales ($M)	Profits ($M)
Type	Size Co	평균	평균
Computer	big	20597.48	1089.9
	medium	3018.85	-85.75
	small	1758.06	44.94
Pharmaceutical	big	7474.04	894.42
	medium	4261.06	698.98
	small	1083.75	156.95

이 평균은 요약 명령을 사용하여 구한 것과 동일합니다. 그림 3.25 와 그림 3.19 를 비교하십시오.

부분집합 생성의 예

데이터 테이블의 일부만 자세히 보기 위해 부분집합을 생성할 수 있습니다. 예를 들어 대규모, 중간 규모, 소규모 컴퓨터 회사와 제약 회사의 매출과 수익을 이미 비교했다고 가정해 보겠습니다. 이제 중간 규모 회사의 매출과 수익만 살펴보려고 합니다.

부분집합 생성 과정은 두 단계로 이루어집니다. 먼저 대상 데이터를 선택한 후 해당 데이터를 새 테이블로 추출합니다.

부분집합 명령으로 부분집합 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .

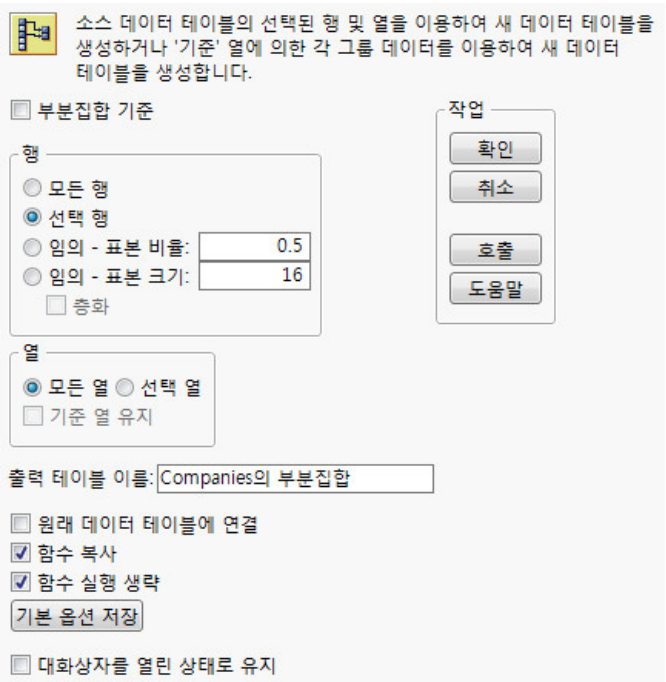
부분집합을 생성할 행 및 열 선택

2. **행 > 행 선택 > 선택 조건**을 선택합니다 .
3. 왼쪽의 열 목록 상자에서 Size Co 를 선택합니다 .
4. 텍스트 입력 상자에 "medium" 을 입력합니다 .
5. **확인**을 클릭합니다 .
6. Ctrl 키를 누르고 Type, Sales ($M) 및 Profits ($M) 열을 선택합니다 .

부분집합 테이블 생성

7. **테이블 > 부분집합**을 선택하여 " 부분집합 " 창을 시작합니다 .

그림 3.26 부분집합 창



8. 선택한 열로만 부분집합을 생성하려면 **선택 열**을 선택합니다. 추가 옵션을 선택하여 부분집합 테이블을 추가로 사용자 정의할 수도 있습니다.
9. **확인**을 클릭합니다.

결과 부분집합 데이터 테이블에는 7개의 행과 3개의 열이 있습니다. "부분집합" 명령에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

분포 플랫폼으로 부분집합 생성

부분집합을 생성하는 또 다른 방법은 플랫폼 결과와 데이터 테이블 간의 연결을 사용하는 것입니다.

예 : 분포 명령을 사용하여 부분집합 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. Type 을 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "Computer" 를 나타내는 히스토그램 막대를 두 번 클릭하여 컴퓨터 회사의 부분집합 테이블을 생성합니다.

주의 : 이 방법을 사용하면 **연결된** 부분집합 테이블이 생성됩니다 . 즉 , 부분집합 테이블의 데이터를 변경하면 소스 테이블에서도 해당 값이 변경됩니다 .

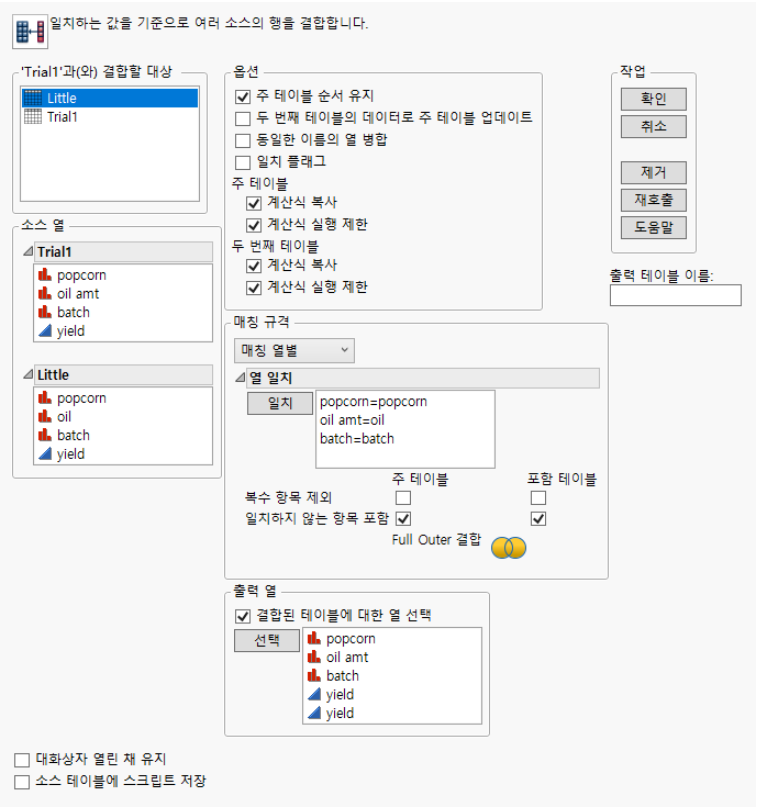
예 : 데이터 테이블 결합

"결합" 옵션을 사용하여 여러 데이터 테이블의 정보를 단일 데이터 테이블에 결합할 수 있습니다. 예를 들어 팝콘 산출량에 대한 실험 결과가 포함된 데이터 테이블이 있다고 가정해 보겠습니다. 다른 데이터 테이블에는 팝콘 산출량에 대한 두 번째 실험 결과가 있습니다. 두 실험을 비교하거나, 두 결과 모두를 사용하여 실험을 분석하려면 동일한 테이블에 데이터가 있어야 합니다. 또한 실험 데이터는 동일한 순서로 데이터 테이블에 입력되지 않았습니다. 한 열은 이름이 다르며 두 번째 실험은 불완전합니다. 따라서 한 테이블에서 다른 테이블로 복사하여 붙여넣을 수 없습니다.

예 : 두 개의 데이터 테이블 결합

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Trial1.jmp 및 Little.jmp 를 엽니다 .
2. Trial1.jmp 데이터 테이블을 클릭하여 활성화합니다 .
3. **테이블 > 결합**을 선택합니다 .
4. **'Trial1' 과 (와) 결합할 대상** 상자에서 Little 을 선택합니다 .
5. **매칭 규격** 메뉴에서 **매칭 열별**이 선택되어 있지 않으면 이 항목을 선택합니다 .
6. **소스 열** 상자 안의 두 상자 모두에서 popcorn 을 선택하고 **매칭**을 클릭합니다 .
7. 같은 방법으로 두 상자의 batch 와 batch 를 매칭하고 oil amt 와 oil 을 매칭합니다 .
   매칭 열의 이름이 같을 필요는 없습니다 .
8. 두 테이블 모두에서 **매칭되지 않는 항목 포함**을 선택합니다 .
   한 실험은 부분적이기 때문에 결측 데이터가 있는 행을 비롯해서 모든 행을 포함하고자 합니다.
9. 열 중복을 피하기 위해 **결합된 테이블의 열 선택** 옵션을 선택합니다 .
10. Trial1 에서 4 개의 열을 모두 선택하고 **선택**을 클릭합니다 .
11. Little 에서 yield 만 선택하고 **선택**을 클릭합니다 .

그림 3.27 완료된 결합 창



12. **확인**을 클릭합니다 .

그림 3.28 결합된 테이블

제목 없음 15
노트 append to TRIAL2 da
노트 2 Join with Trial1.jmp
소스
열(5/0)
popcorn ✱
oil amt ✱
batch ✱
Trial1의 yield
Little의 yield

	popcorn	oil amt	batch	Trial1의 yield	Little의 yield
1	plain	little	large	8.2	8.8
2	gourmet	little	large	8.6	8.2
3	plain	lots	large	10.4	•
4	gourmet	lots	large	9.2	•
5	plain	little	small	9.9	10.1
6	gourmet	little	small	12.1	15.9
7	plain	lots	small	10.6	•
8	gourmet	lots	small	18.0	•

데이터 정렬

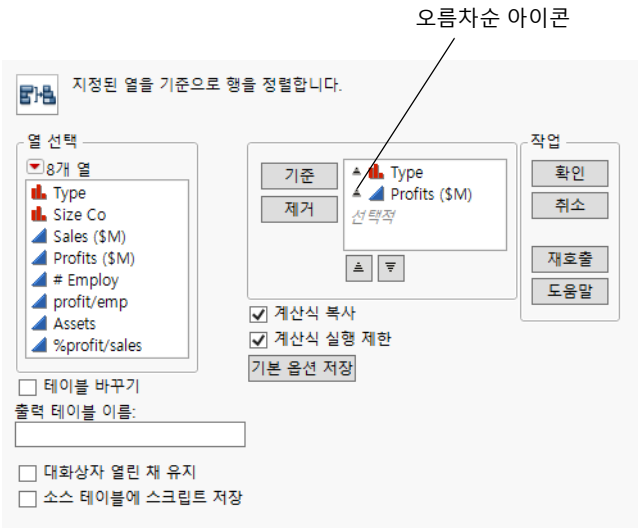
" 정렬 " 명령을 사용하여 데이터 테이블에 있는 하나 이상의 열을 기준으로 데이터 테이블을 정렬할 수 있습니다 . 예를 들어 컴퓨터 및 제약 회사의 재무 데이터를 살펴보십시오 . 처음에는 Type,

그 다음에는 Profits ($M) 를 기준으로 하여 데이터 테이블을 정렬한다고 가정해 보겠습니다 . 또한 각 Type 내에서 Profits ($M) 를 내림차순으로 표시하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **테이블 > 정렬**을 선택합니다 .
3. Type 을 선택하고 **기준**을 클릭하여 Type 을 정렬 변수로 할당합니다 .
4. Profits ($M) 를 선택하고 **기준**을 클릭합니다 .

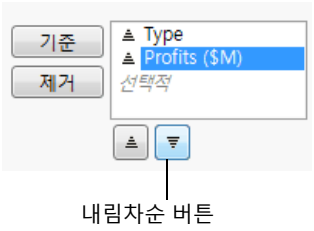
현재 두 변수는 모두 오름차순으로 정렬되도록 설정되어 있습니다. 그림 3.29 에서 변수 옆에 있는 오름차순 아이콘을 확인하십시오.

그림 3.29 오름차순 정렬 아이콘



5. Profits ($M) 를 내림차순으로 정렬하도록 변경하기 위해 Profits ($M) 를 선택하고 내림차순 버튼을 클릭합니다 .

그림 3.30 수익을 내림차순으로 변경



"Profits ($M)" 옆의 아이콘이 내림차순으로 변경됩니다 .

6. **테이블 바꾸기** 체크박스를 선택합니다 .

테이블 바꾸기를 선택하면 정렬된 값으로 새 테이블이 생성되지 않고 원래 데이터 테이블이 정렬됩니다. 원래 데이터 테이블에서 생성된 보고서 창이 열려 있으면 이 옵션을 사용할 수 없습니다. 관련 보고서 창이 열려 있는 상태에서 데이터 테이블을 정렬하면 데이터 중 일부가 보고서 창, 특히 그래프에 표시되는 방식이 변경될 수 있습니다.

7. **확인**을 클릭합니다.

이제 데이터 테이블이 유형 기준 사전순으로 정렬되고, 유형 내에서는 수익 합계 기준 내림차순으로 정렬됩니다.

4 장

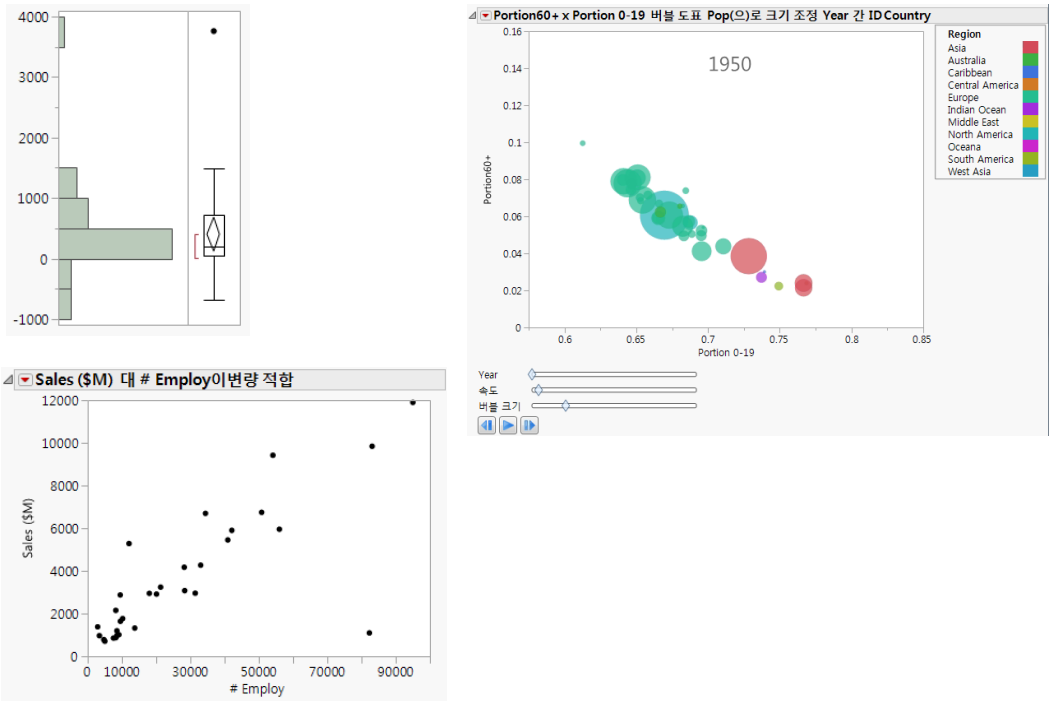
데이터 시각화

일반적인 그래프

데이터 시각화는 중요한 첫 번째 단계입니다. 이 장에서 설명하는 그래프는 데이터에 대한 중요한 상세 정보를 발견하는 데 도움이 됩니다. 예를 들어 히스토그램은 데이터의 모양과 범위를 보여 주고 비정상적인 데이터 점을 찾는 데 도움을 줍니다.

이 장에서는 JMP에서 데이터를 시각화하고 탐색하는 데 사용되는 가장 일반적인 그래프 및 그림 중 몇 가지를 설명합니다. 이 장에서는 JMP의 그래픽 도구 및 플랫폼을 소개합니다. JMP를 사용하여 단일 변수의 분포나 여러 변수 간의 관계를 시각화할 수 있습니다.

그림 4.1 JMP 를 사용하여 데이터 시각화



목차

단변량 그래프의 단일 변수 분석 93
연속형 변수에 히스토그램 사용 93
범주형 변수에 막대 차트 사용 96
다중 변수 비교 98
산점도를 사용하여 여러 변수 비교 99
산점도 행렬을 사용하여 여러 변수 비교 102
병렬 상자 그림을 사용하여 여러 변수 비교 105
그래프 빌더를 사용하여 여러 변수 비교 107
버블 그림을 사용하여 여러 변수 비교 113
중첩 그림을 사용하여 여러 변수 비교 118
계량형 차트를 사용하여 여러 변수 비교 122

단변량 그래프의 단일 변수 분석

단일 변수 그래프, 즉 **단변량** 그래프를 사용하여 한 번에 하나의 변수를 자세히 살펴볼 수 있습니다. 데이터를 살펴보기 시작할 때는 각 변수 간의 교호작용을 확인하기 전에 각 변수에 대해 알아두는 것이 중요합니다. 단변량 그래프를 사용하면 각 변수를 개별적으로 시각화할 수 있습니다.

이 섹션에서는 단일 변수의 분포를 보여 주는 두 가지 그래프를 다룹니다 .

- " 연속형 변수에 히스토그램 사용 "(93 페이지)
- " 범주형 변수에 막대 차트 사용 "(96 페이지)

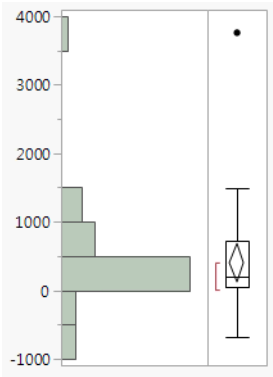
분포 플랫폼을 사용하여 두 그래프를 모두 생성할 수 있습니다. 분포 플랫폼은 각 변수에 대한 그래픽 설명과 기술 통계량을 생성합니다.

연속형 변수에 히스토그램 사용

히스토그램은 연속형 변수의 분포를 파악하는 데 가장 유용한 그래픽 도구 중 하나입니다 . 데이터에서 다음 정보를 확인하려면 히스토그램을 사용합니다 .

- 평균 값 및 변동
- 극단값

그림 4.2 히스토그램의 예



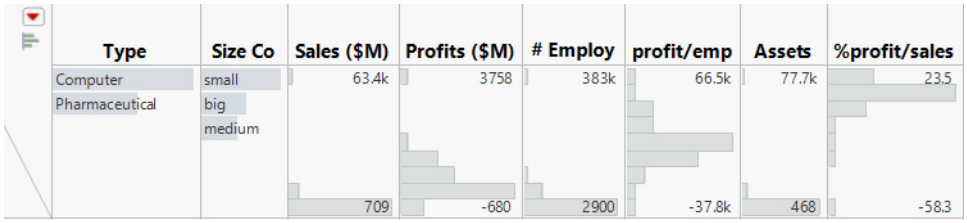
즉석 히스토그램

열 머리글에 있는 히스토그램 아이콘을 클릭하면 곧바로 히스토그램을 볼 수 있습니다. 이 히스토그램은 열 머리글 아래에 표시됩니다.

그림 4.3 즉석 히스토그램

	Type	Size Co	Sales ($M)	Profits ($M)	# Employ	profit/emp	Assets	%profit/sales
1	Computer	small	855.1	31.0	7523	4120.70	615.2	3.63
2	Pharmaceutical	big	5453.5	859.8	40929	21007.11	4851.6	15.77
3	Computer	small	2153.7	153.0	8200	18658.54	2233.7	7.10

히스토그램 아이콘을 클릭합니다 .



시나리오

이 예에서는 특정 회사 그룹의 수익 데이터가 포함된 Companies.jmp 데이터 테이블을 사용합니다. 재무 분석가가 다음과 같은 질문에 대한 답을 구하려고 합니다 .

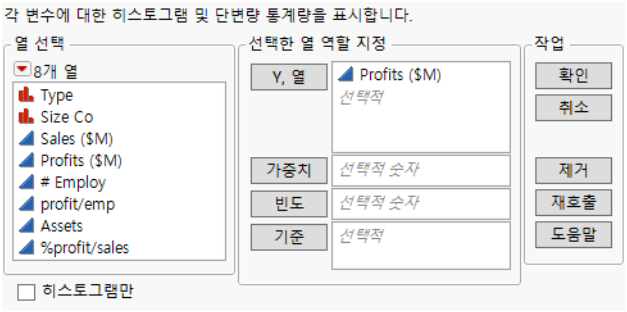
- 일반적으로 각 회사는 얼마나 많은 수익을 올립니까 ?
- 평균 수익은 얼마입니까 ?
- 다른 회사에 비해 극단적으로 높거나 낮은 수익을 올리는 회사가 있습니까 ?

이러한 질문에 답하려면 Profits ($M) 의 히스토그램을 사용합니다.

히스토그램 생성

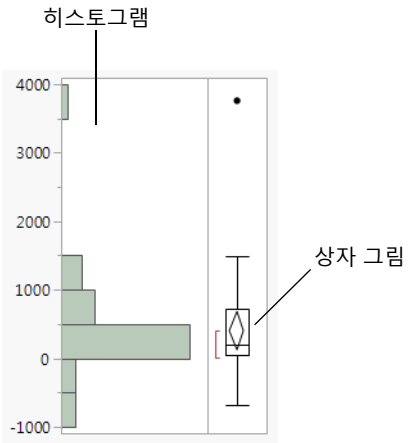
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Profits ($M) 를 선택하고 **Y, 열**을 클릭합니다 .

그림 4.4 Profits ($M) 의 분포 창



4. **확인**을 클릭합니다 .

그림 4.5 Profits ($M) 의 히스토그램



히스토그램 해석

이 히스토그램에서는 다음을 알 수 있습니다 .

- 대부분의 회사는 수익이 -1,000 달러 ~ 1,500 달러 사이입니다 .

  하나를 제외한 모든 막대가 이 범위에 있습니다 . 또한 수익이 0 달러 ~ 500 달러 범위에 있는 회사가 가장 많습니다 . 해당 범위를 나타내는 막대가 다른 모든 막대보다 훨씬 깁니다 .
- 평균 수익은 500 달러보다 약간 적습니다 .

  상자 그림에 있는 마름모의 중심은 평균 값을 나타냅니다 . 여기서는 평균이 500 달러보다 약간 낮습니다 .
- 한 회사는 다른 회사들보다 훨씬 더 높은 수익을 올리고 있어 **이상치**로 간주될 수 있습니다 . 이상치는 다른 데이터 점의 일반적인 패턴과 동떨어져 있는 데이터 점입니다 .

이 이상치는 히스토그램의 상단에 하나의 매우 짧은 막대로 표시됩니다 . 이 작은 막대는 소규모 그룹 (여기서는 단일 회사) 을 나타내며 나머지 히스토그램 막대와 상당히 동떨어져 있습니다 .

이 보고서에는 히스토그램 외에도 다음이 포함됩니다 .

- 데이터의 또 다른 그래픽 요약인 상자 그림 . 상자 그림에 대한 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .
- **분위수** 및 **요약 통계량** 보고서 이러한 보고서에 대해서는 "데이터 분석" 장의 "분포 분석"(136 페이지) 에서 설명합니다 .

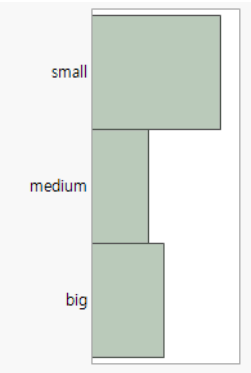
히스토그램에서의 상호 작용

JMP에서 데이터 테이블과 보고서는 모두 연결되어 있습니다. 히스토그램 막대를 클릭하면 데이터 테이블에서 해당 행이 선택됩니다.

범주형 변수에 막대 차트 사용

막대 차트를 사용하여 범주형 변수의 분포를 시각화할 수 있습니다. 막대 차트와 히스토그램은 둘 다 변수의 각 수준에 해당하는 막대가 있기 때문에 모양이 비슷합니다. 막대 차트는 변수의 모든 수준에 대한 막대를 표시하지만 히스토그램은 변수의 값 범위를 표시합니다.

그림 4.6 막대 차트의 예



시나리오

이 예에서는 특정 회사 그룹의 규모 및 유형과 관련된 데이터가 포함된 Companies.jmp 데이터 테이블을 사용합니다 .

재무 분석가가 다음과 같은 질문에 대한 답을 구하려고 합니다.

- 가장 일반적인 회사 유형은 무엇입니까 ?
- 가장 일반적인 회사 규모는 어떠합니까 ?

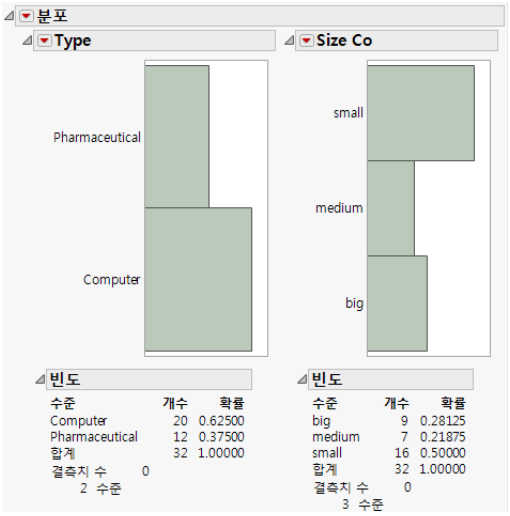
이러한 질문에 답하기 위해 Type 및 Size Co의 막대 차트를 사용합니다.

막대 차트 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. Type 및 Size Co 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.

그림 4.7 Type 및 Size Co 의 막대 차트

막대 차트



요약 정보

빈도

수준	개수	확률
Computer	20	0.62500
Pharmaceutical	12	0.37500
합계	32	1.00000

결측치 수 0
2 수준

빈도

수준	개수	확률
big	9	0.28125
medium	7	0.21875
small	16	0.50000
합계	32	1.00000

결측치 수 0
3 수준

막대 차트 해석

이 막대 차트에서는 다음을 알 수 있습니다.

- 제약 회사보다 컴퓨터 회사가 많습니다.

  컴퓨터 회사를 나타내는 막대가 제약 회사를 나타내는 막대보다 큽니다.
- 가장 일반적인 회사 규모는 소규모입니다.

  소규모 회사를 나타내는 막대가 중간 규모 및 대규모 회사를 나타내는 막대보다 큽니다.

추가적인 요약 출력을 통해 상세한 빈도가 제공됩니다. 이 보고서에 대해서는 "데이터 분석" 장의 "범주형 변수의 분포"(139 페이지)에서 설명합니다.

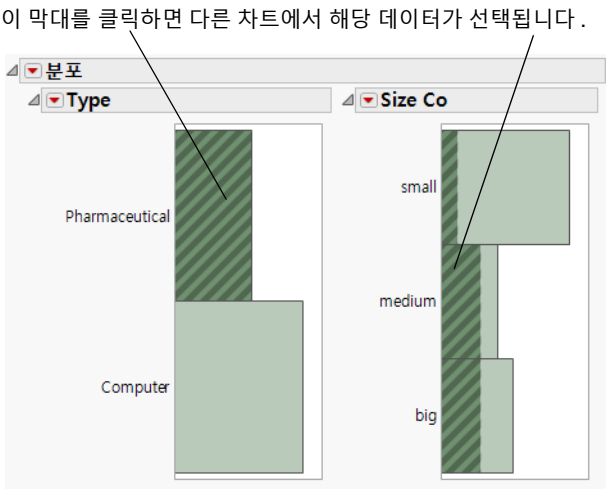
막대 차트에서의 상호 작용

히스토그램과 마찬가지로 개별 막대를 클릭하면 데이터 테이블의 해당 행이 강조 표시됩니다. 그래프가 두 개 이상 생성된 경우 한 막대 차트의 막대를 클릭하면 다른 막대 차트에서 그에 해당하는 막대가 강조 표시됩니다.

예를 들어 제약 회사의 회사 규모 분포를 보려고 한다고 가정해 보겠습니다. Type 막대 차트에서 제약 회사 막대를 클릭하면 Size Co 막대 차트에서 제약 회사가 강조 표시됩니다. 그림 4.8에서는 이 데이터 테이블에 있는 대부분의 회사가 소규모이지만 대부분의 제약 회사는 중간 규모 또는 대규모임을 보여 줍니다.

또한 데이터 테이블에서도 해당 행이 선택됩니다.

그림 4.8 막대 클릭



다중 변수 비교

다중 변수 그래프를 사용하여 둘 이상의 변수 사이에 존재하는 관계 및 패턴을 시각화할 수 있습니다. 이 섹션에서는 다음 그래프를 다룹니다.

표 4.1 다중 변수 그래프

"산점도를 사용하여 여러 변수 비교"(99 페이지)	산점도를 사용하면 두 개의 연속형 변수를 비교할 수 있습니다.
"산점도 행렬을 사용하여 여러 변수 비교"(102 페이지)	산점도 행렬을 사용하면 여러 쌍의 연속형 변수를 비교할 수 있습니다.
"병렬 상자 그림을 사용하여 여러 변수 비교"(105 페이지)	병렬 상자 그림을 사용하면 하나의 연속형 변수와 하나의 범주형 변수를 비교할 수 있습니다.
"계량형 차트를 사용하여 여러 변수 비교"(122 페이지)	계량형 차트를 사용하면 하나의 연속형 Y 변수를 하나 이상의 범주형 X 변수와 비교할 수 있습니다. 계량형 차트는 몇 가지 범주형 X 변수 간의 평균 및 변동 차이를 보여 줍니다.

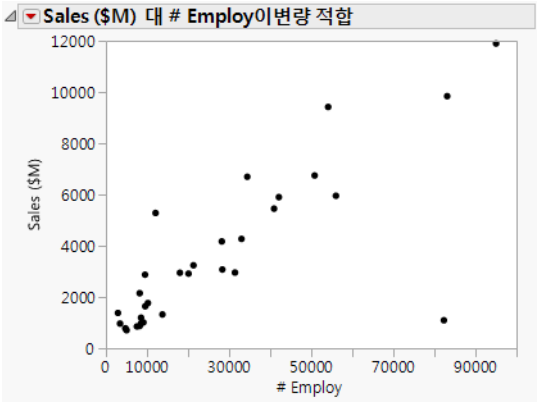
표 4.1 다중 변수 그래프 (계속)

"그래프 빌더를 사용하여 여러 변수 비교"(107 페이지)	그래프 빌더를 사용하면 대화식으로 그래프를 생성 및 변경할 수 있습니다.
"중첩 그림을 사용하여 여러 변수 비교"(118 페이지)	중첩 그림을 사용하면 Y 축에 있는 하나 이상의 변수를 X 축의 다른 변수와 비교할 수 있습니다. 중첩 그림에서는 두 개 이상의 변수가 시간에 따라 어떻게 변하는지 비교할 수 있기 때문에 X 변수가 시간 변수일 때 특히 유용합니다.
"버블 그림을 사용하여 여러 변수 비교"(113 페이지)	버블 그림은 색상 및 버블 크기를 사용하여 한 번에 최대 5개의 변수를 표시하는 특수한 산점도입니다. 변수 중 하나가 시간 변수이면 그림에 애니메이션을 적용하여 다른 변수가 시간에 따라 변화하는 것을 볼 수 있습니다.

산점도를 사용하여 여러 변수 비교

산점도는 모든 다중 변수 그래프 중에서 가장 단순합니다. 두 연속형 변수 간의 관계를 확인하고 두 연속형 변수 사이에 **상관관계**가 있는지 확인하려면 산점도를 사용합니다. 상관관계는 두 변수가 얼마나 밀접하게 관련되어 있는지를 나타냅니다. 상관관계가 높은 두 변수가 있으면 한 변수가 다른 변수에 영향을 미칠 수 있습니다. 또는 두 변수가 모두 비슷한 방식으로 다른 변수의 영향을 받을 수 있습니다.

그림 4.9 산점도의 예



시나리오

이 예에서는 특정 회사 그룹의 매출액과 직원 수가 포함된 Companies.jmp 데이터 테이블을 사용합니다 .

재무 분석가가 다음과 같은 질문에 대한 답을 구하려고 합니다.

- 매출과 직원 수 사이에 어떤 관계가 있습니까 ?

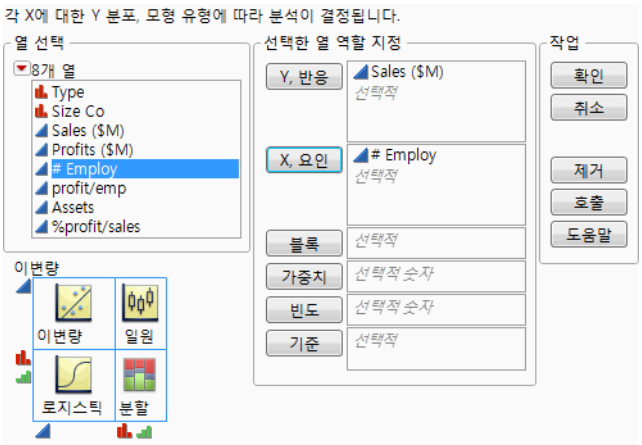
- 직원 수에 따라 매출액이 증가합니까 ?
- 직원 수를 토대로 평균 매출을 예측할 수 있습니까 ?

이러한 질문에 답하려면 Sales ($M) 대 # Employ 산점도를 사용하십시오 .

산점도 생성

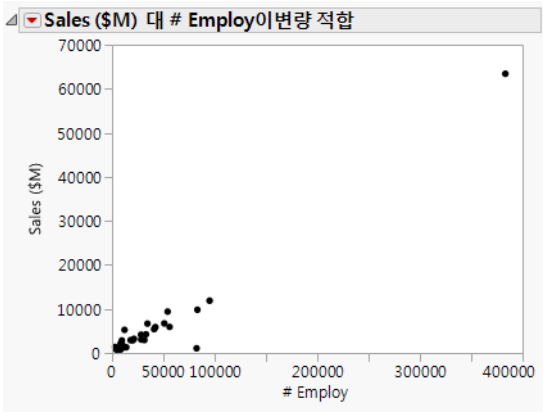
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Sales ($M) 를 선택하고 **Y, 반응**을 클릭합니다 .
4. # Employ 를 선택하고 **X, 요인**을 클릭합니다 .

그림 4.10 X 로 Y 적합 창



5. **확인**을 클릭합니다 .

그림 4.11 Sales ($M) 대 # Employ 산점도



산점도 해석

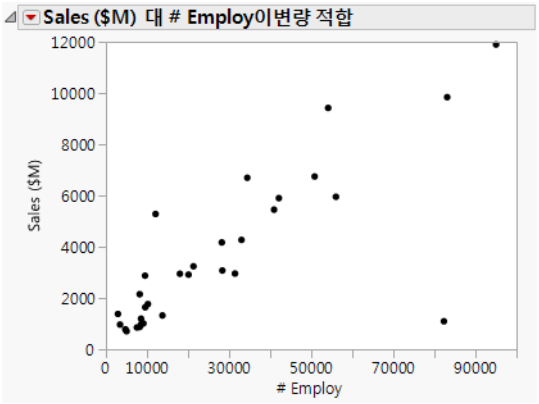
한 회사는 직원 수가 매우 많고 매출이 상당히 높습니다 . 이 회사는 그림 오른쪽 상단에 단일 점으로 표시됩니다 . 이 데이터 점과 나머지 모든 점 사이에는 상당한 거리가 있어 나머지 회사 간의 관계를 시각화하기가 어렵습니다 . 다음 단계에 따라 그림에서 해당 점을 제거하고 그림을 다시 생성하십시오 .

1. 해당 점을 클릭하여 선택합니다 .
2. **행 > 숨기고 제외하기**를 선택합니다 . 해당 데이터 점이 숨겨지고 더 이상 계산에 포함되지 않습니다 .

   참고 : 숨기기와 제외 사이에는 중요한 차이가 있습니다 . 점을 숨기면 해당 점이 모든 그래프에서 제거되지만 통계 계산에는 계속 사용됩니다 . 점을 제외하면 해당 점이 통계 계산에서 제거되지만 그래프에서는 제거되지 않습니다 . 점을 숨기고 제외하면 모든 계산과 모든 그래프에서 해당 점이 제거됩니다 .

3. 이상치를 제외하고 그림을 다시 생성하려면 " 이변량 " 의 빨간색 삼각형을 클릭하고 **다시 실행 > 분석 다시 실행**을 선택합니다 . 원래 보고서 창은 닫아도 됩니다 .

그림 4.12 이상치가 제거된 산점도



업데이트된 산점도에서는 다음을 알 수 있습니다 .

- 매출과 직원 수 사이에는 관계가 있습니다 .

  데이터 점에 눈에 띄는 패턴이 있습니다 . 데이터 점은 그래프 전체에 무작위로 흩어져 있지 않습니다 . 대부분의 데이터 점에 근접하는 주대각선을 그릴 수 있습니다 .

- 직원 수에 따라 매출이 늘어나며 이 관계는 선형입니다 .

  주대각선을 그린다면 왼쪽 하단에서 오른쪽 상단으로 향하는 기울기를 보일 것입니다 . 이 기울기는 직원 수가 증가함에 따라 (하단 축의 왼쪽에서 오른쪽으로) 매출 또한 증가함 (왼쪽 축의 아래쪽에서 위쪽으로) 을 보여 줍니다 . 대부분의 데이터 점이 직선 근처에 위치하므로 선형 관계를 나타냅니다 . 데이터 점에 근접하는 선을 곡선으로만 그릴 수 있는 경우라도 점의 패턴 때문에 관계가 유지됩니다 . 그러나 이 관계는 선형적이지 않습니다 .

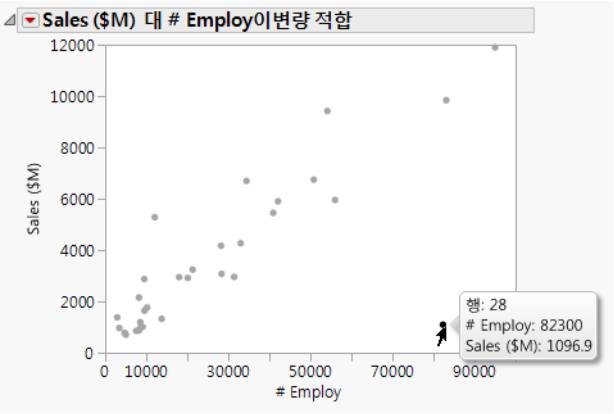
- 직원 수를 토대로 평균 매출을 예측할 수 있습니다 .

  이 산점도는 일반적으로 직원 수가 증가할수록 매출도 증가한다는 것을 보여 줍니다. 따라서 회사의 직원 수만 알고 있으면 회사의 매출을 예측할 수 있습니다 . 이러한 예측점은 가상의 선 위에 있을 것입니다 . 예측은 정확하지는 않겠지만 실제 매출과 비슷할 것입니다 .

산점도에서의 상호 작용

산점도는 다른 JMP 그래픽과 마찬가지로 대화식입니다. 오른쪽 하단의 점을 마우스 커서로 가리키면 행 번호와 x 및 y 값이 표시됩니다.

그림 4.13 커서로 점 가리키기



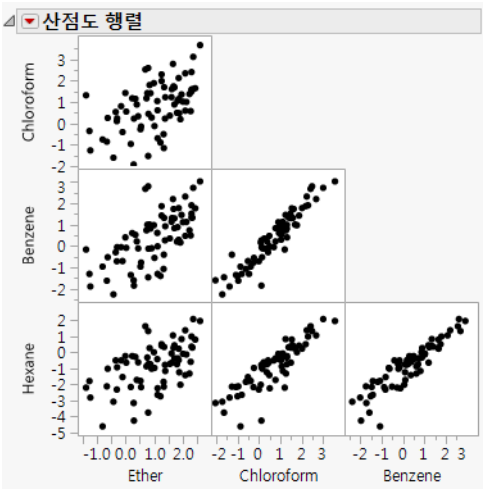
점을 클릭하면 데이터 테이블에서 해당 행이 강조 표시됩니다 . 여러 점을 선택하려면 다음 중 하나를 수행하십시오 .

- 점 주변에 커서를 놓고 클릭하여 드래그합니다 . 그러면 사각형 영역의 점이 선택됩니다 .
- 올가미 도구를 선택한 후 여러 점 주변을 클릭하여 드래그합니다 . 그러면 모양이 불규칙한 영역이 선택됩니다 .

산점도 행렬을 사용하여 여러 변수 비교

산점도 행렬은 격자(또는 행렬)로 구성된 산점도 모음입니다. 각 산점도는 한 쌍의 변수 사이에 있는 관계를 보여 줍니다.

그림 4.14 산점도 행렬의 예



시나리오

이 예에서는 72가지 용질의 용해도 측정 데이터가 들어 있는 Solubility.jmp 데이터 테이블을 사용합니다 .

실험실 기술자가 다음과 같은 질문에 대한 답을 구하려고 합니다.

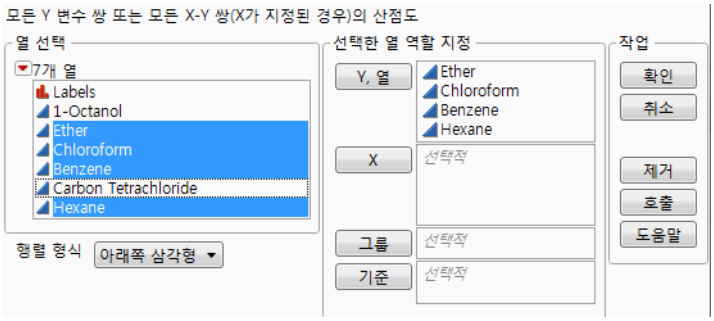
- 관련성이 있는 화학 물질 쌍이 있습니까 ? 가능한 6 개의 쌍이 있습니다 .
- 어떤 쌍의 관련성이 가장 큽니까 ?

이 질문에 답하려면 네 가지 용매의 산점도 행렬을 사용하십시오.

산점도 행렬 생성

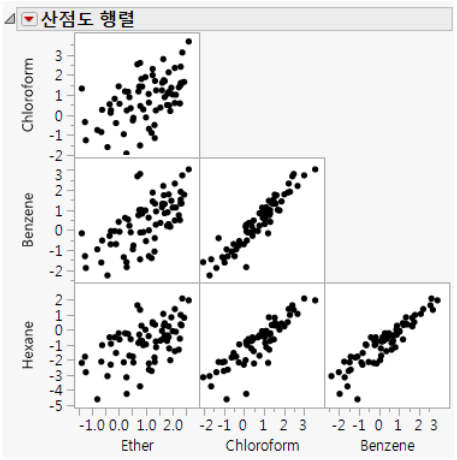
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Solubility.jmp 를 엽니다 .
2. **그래프 > 산점도 행렬**을 선택합니다 .
3. Ether, Chloroform, Benzene 및 Hexane 을 선택하고 **Y, 열**을 클릭합니다 .

그림 4.15 산점도 행렬 창



4. **확인**을 클릭합니다.

그림 4.16 산점도 행렬



산점도 행렬 해석

이 산점도 행렬에서는 다음을 알 수 있습니다.

- 6 개의 변수 쌍은 모두 양의 상관관계에 있습니다.

  한 변수가 증가하면 다른 변수도 증가합니다.
- Benzene 과 Chloroform 사이에 가장 강한 관계가 있는 것으로 나타납니다.

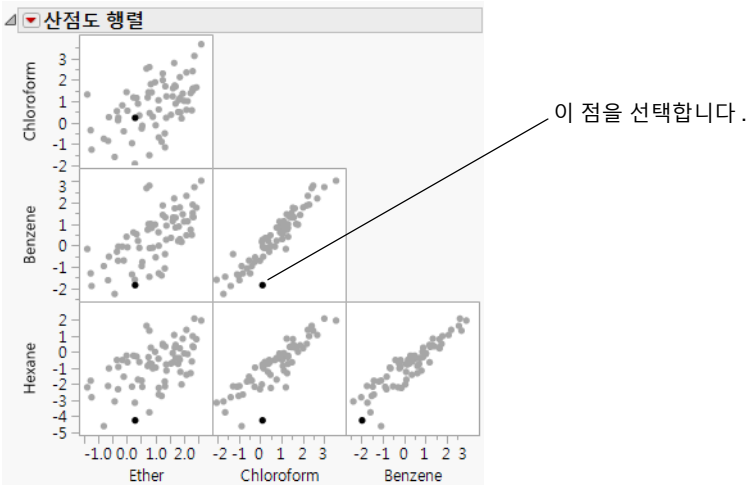
  Benzene 과 Chloroform 의 산점도에 있는 데이터 점은 가상의 선을 따라 가장 조밀하게 밀집되어 있습니다.

산점도 행렬에서의 상호 작용

한 산점도에서 점을 선택하면 다른 모든 산점도에서도 해당 점이 선택됩니다.

예를 들어 Benzene 대 Chloroform 산점도에서 한 점을 선택하면 다른 다섯 개의 산점도에서 동일한 점이 선택됩니다.

그림 4.17 선택된 점

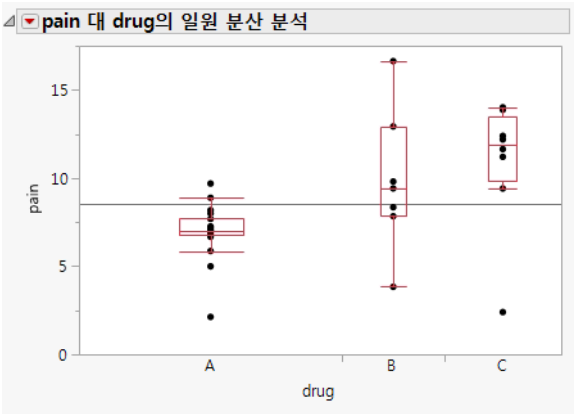


병렬 상자 그림을 사용하여 여러 변수 비교

병렬 상자 그림은 다음을 보여 줍니다.

- 연속형 변수 하나와 범주형 변수 하나 사이의 관계
- 범주형 변수의 전체 수준에서 나타나는 연속형 변수의 차이

그림 4.18 병렬 상자 그림의 예



시나리오

이 예에서는 세 가지 약물의 투약 시 환자 통증 측정 데이터가 포함된 Analgesics.jmp 데이터 테이블을 사용합니다.

연구원이 다음과 같은 질문에 대한 답을 구하려고 합니다 .

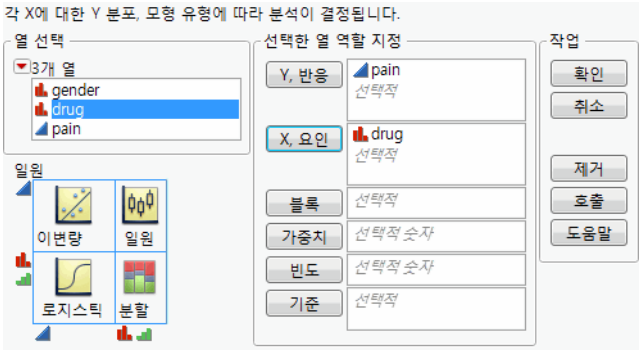
- 약물의 평균적인 통증 조절 수준 간에 차이가 있습니까 ?
- 약물마다 통증 조절 수준의 **변동성**이 다릅니까 ? 변동성이 높은 약물은 변동성이 낮은 약물보다 신뢰도가 떨어집니다 .

이러한 질문에 답하려면 통증 수준과 약물 범주에 대한 병렬 상자 그림을 사용하십시오.

병렬 상자 그림 생성

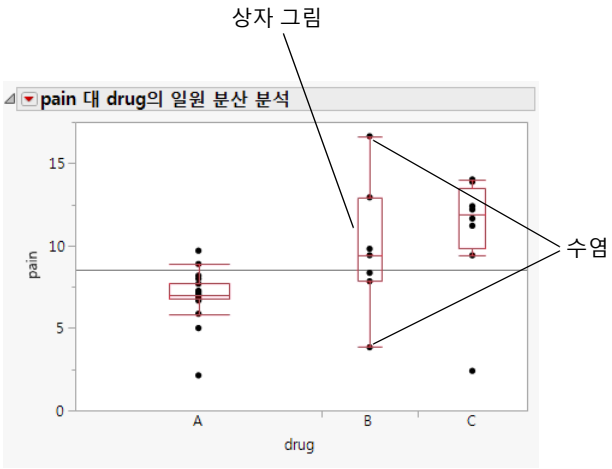
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Analgesics.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. pain 을 선택하고 **Y, 반응**을 클릭합니다 .
4. drug 를 선택하고 **X, 요인**을 클릭합니다 .

그림 4.19 X 로 Y 적합 창



5. **확인**을 클릭합니다 .
6. "pain 대 drug 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림**을 선택합니다 .

그림 4.20 병렬 상자 그림



병렬 상자 그림 해석

상자 그림은 다음과 같은 원리에 따라 만들어졌습니다 .

- 상자 안의 선은 중앙값을 나타냅니다 .
- 데이터의 가운데 절반은 상자 안에 있습니다 .
- 대부분의 데이터는 양쪽 수염 끝 사이에 있습니다 .
- 수염을 벗어나는 데이터 점은 이상치일 수 있습니다 .

그림 4.20 의 상자 그림에서는 다음을 알 수 있습니다 .

- 약물 A 의 상자 그림이 다른 약물보다 통증 수준이 낮기 때문에 약물 A 환자의 통증이 덜하다고 믿을 만한 근거가 됩니다 .
- 약물 B 는 약물 A 및 C 보다 상자 그림이 더 길기 때문에 변동성이 높은 것으로 보입니다 .

약물 C의 한 점은 약물 C의 다른 점보다 훨씬 낮습니다. 이 점을 커서로 가리키면 데이터 테이블의 26 행에 해당하는 것을 알 수 있습니다. 이 점은 약물 그룹 A 또는 B의 데이터와 더 유사해 보이므로 26 행의 정보는 조사할 만합니다. 데이터를 기록할 때 잘못 적었을 수도 있습니다.

그래프 빌더를 사용하여 여러 변수 비교

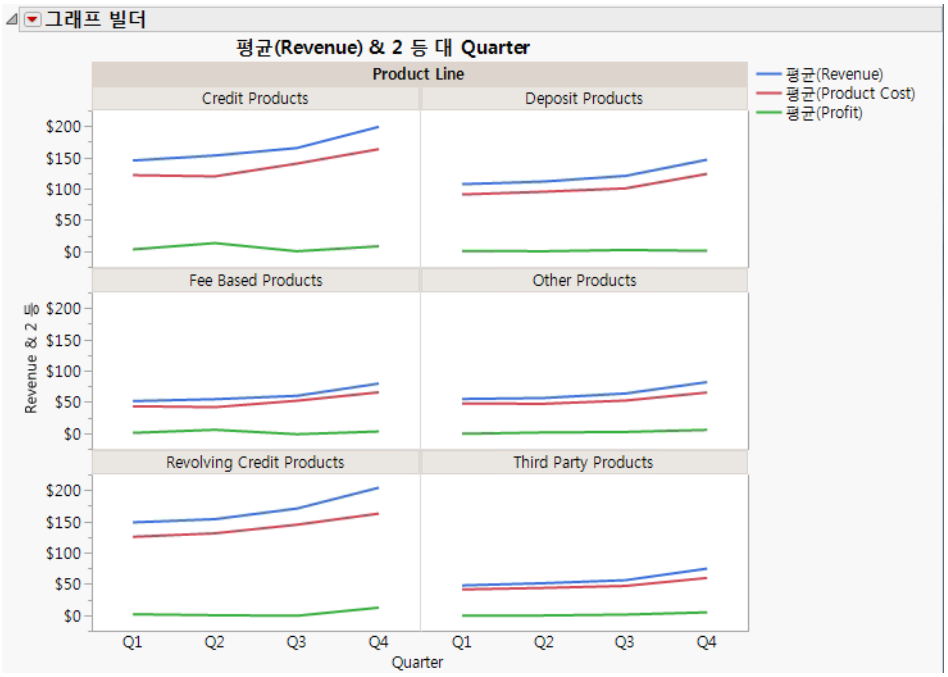
그래프 빌더를 사용하여 대화식으로 그래프를 생성하고 수정할 수 있습니다. JMP의 그래프는 대부분 플랫폼을 시작하고 변수를 지정하는 방법으로 생성됩니다. 다른 유형의 그래프를 생성하려면 "그래프" 메뉴에서 특정 플랫폼을 시작합니다. 하지만 그래프 빌더를 사용하면 언제든지 변수를 변경하고 그래프 유형을 변경할 수 있습니다.

그래프 빌더를 사용하여 다음 작업을 수행할 수 있습니다 .

- 변수를 그래프 안팎으로 드래그하여 놓는 방법으로 변경합니다 .

- 몇 번의 마우스 클릭으로 다른 유형의 그래프를 생성합니다 .
- 그래프를 가로 또는 세로로 분할합니다 .

그림 4.21 그래프 빌더로 생성된 그래프의 예



참고 : 여기에서는 그래프 빌더의 일부 기능만 다룹니다 . 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .

시나리오

이 예에서는 여러 제품 라인의 수익 데이터가 포함된 Profit by Product.jmp 데이터 테이블을 사용합니다.

비즈니스 분석가가 다음과 같은 질문에 대한 답을 구하려고 합니다 .

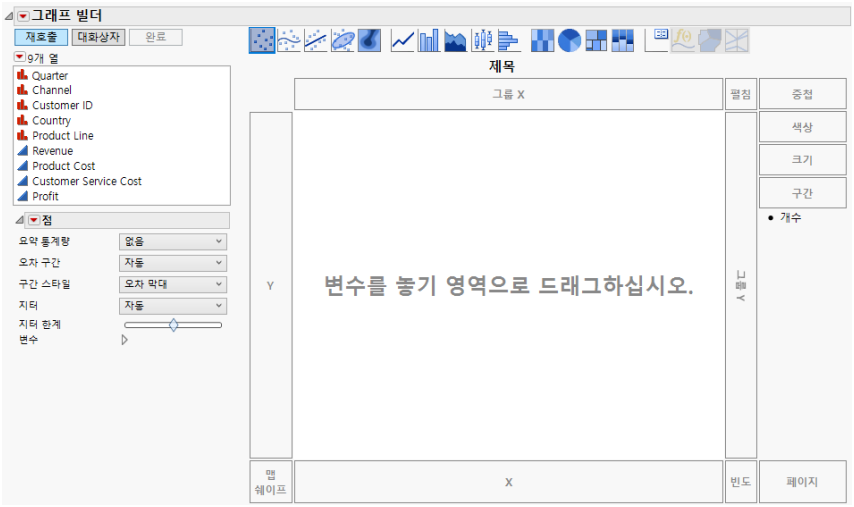
- 제품 라인 간의 수익성이 어떻게 다릅니까 ?

이 질문에 답하려면 다양한 제품 라인의 매출, 제품 비용 및 수익 데이터를 표시하는 선 그림을 사용하십시오.

그래프 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Profit by Product.jmp 를 엽니다 .
2. **그래프 > 그래프 빌더**를 선택합니다 .

그림 4.22 그래프 빌더 작업 공간

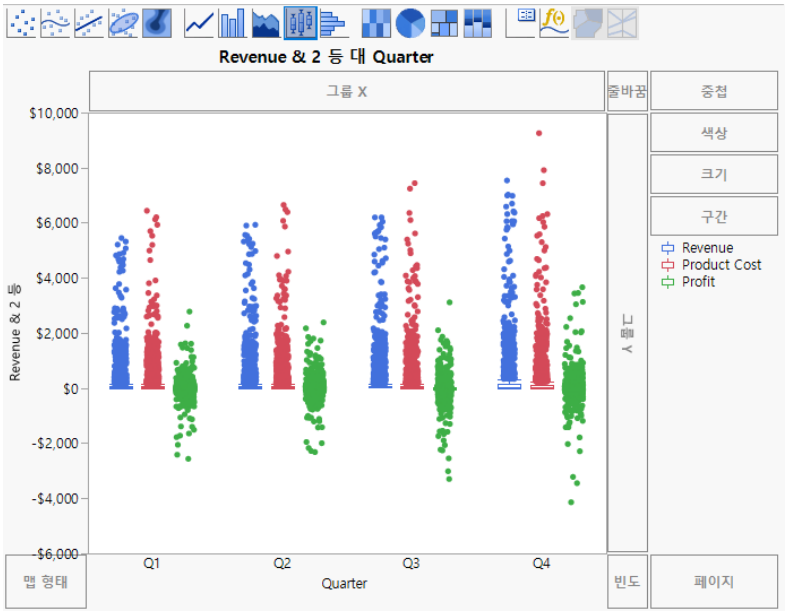


3. Quarter 를 클릭한 후 X 영역으로 드래그하여 놓음으로써 Quarter 를 X 변수로 할당합니다 .
4. Revenue, Product Cost 및 Profit 을 클릭한 후 Y 영역으로 드래그하여 놓음으로써 이 세 변수를 모두 Y 변수로 할당합니다 .

   이제 X 및 Y 영역이 축이 되었습니다 .

참고 : 변수를 클릭한 후 영역을 클릭하여 할당할 수도 있습니다 . 그러나 영역이 축이 된 후에는 변수와 축을 클릭하는 대신 추가 변수를 축으로 드래그하여 놓으십시오 .

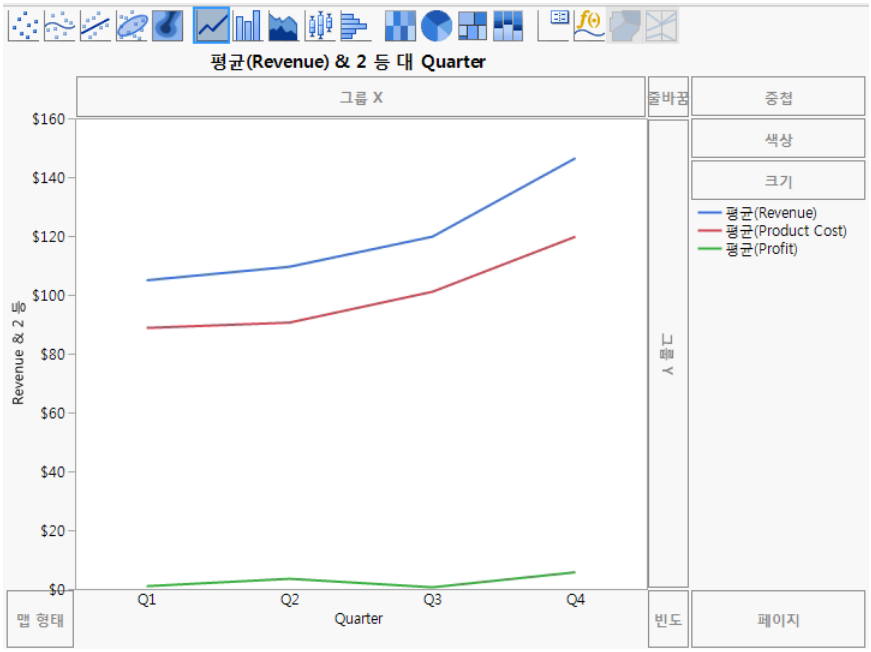
그림 4.23 Y 및 X 변수 추가 후



그래프 빌더는 사용하고 있는 변수에 따라 병렬 상자 그림을 표시합니다.

5. 상자 그림을 선 그림으로 변경하려면 선 아이콘 을 클릭합니다.

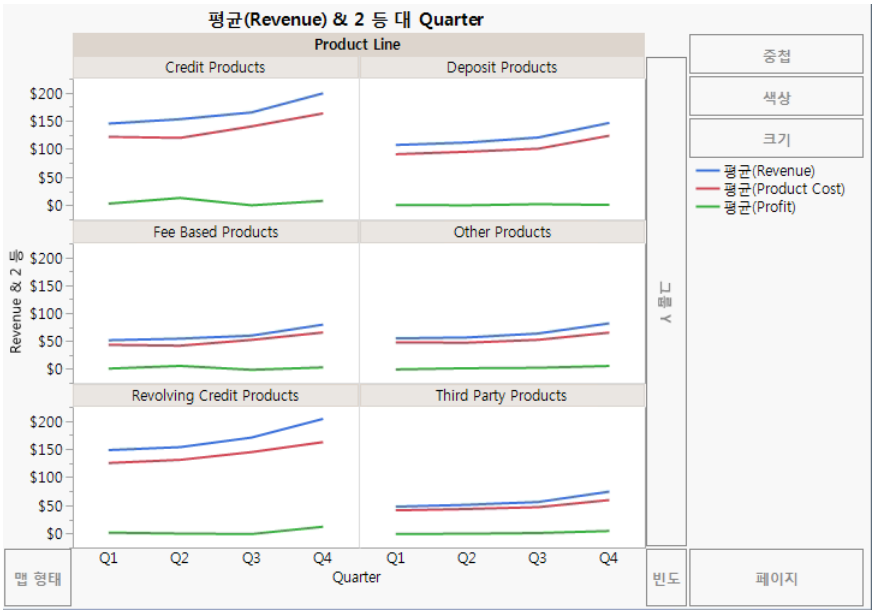
그림 4.24 선 그림



6. 각 제품에 대해 개별 차트를 생성하려면 Product Line 을 클릭한 후 **펼침** 영역으로 드래그하여 놓습니다 .

   각 제품에 대해 개별 선 그림이 생성됩니다 .

그림 4.25 최종 선 그림



그래프 해석

그림 4.25 에서는 제품 라인별로 분류된 매출 , 비용 및 수익을 보여 줍니다 . 비즈니스 분석가는 제품 라인 간의 수익성 차이를 확인하는 데 관심이 있었습니다 . 그림 4.25 의 선 그림을 통해 다음과 같은 몇 가지 사항을 확인할 수 있습니다 .

- 신용 상품 , 예금 상품 및 회전 신용 상품이 수수료 기반 상품 , 타사 상품 및 기타 상품보다 많은 매출을 올립니다 .
- 그러나 모든 제품 라인의 수익은 비슷합니다 .

이 데이터 테이블에는 판매 채널에 대한 데이터도 포함되어 있습니다 . 비즈니스 분석가는 매출 , 제품 비용 및 수익이 서로 다른 판매 채널 간에 어떻게 다른지 확인하고자 합니다 .

1. 그래프에서 Product Line 을 제거하려면 그래프의 제목 (Product Line) 을 클릭하고 그래프 빌더 안의 빈 영역으로 드래그하여 놓습니다 .
2. Channel을 펼침 변수로 추가하려면 Channel을 클릭하고 **펼침** 영역으로 드래그하여 놓습니다.

그림 4.26 판매 채널을 보여주는 선 그림

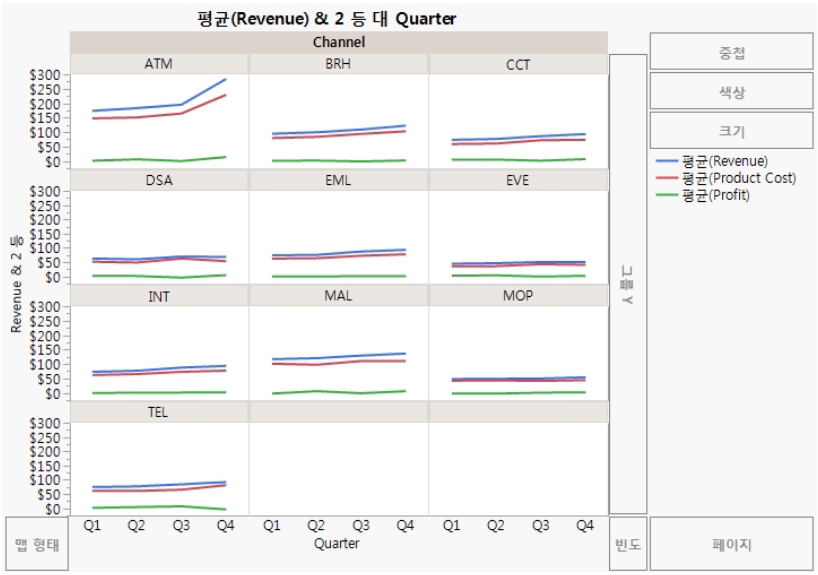
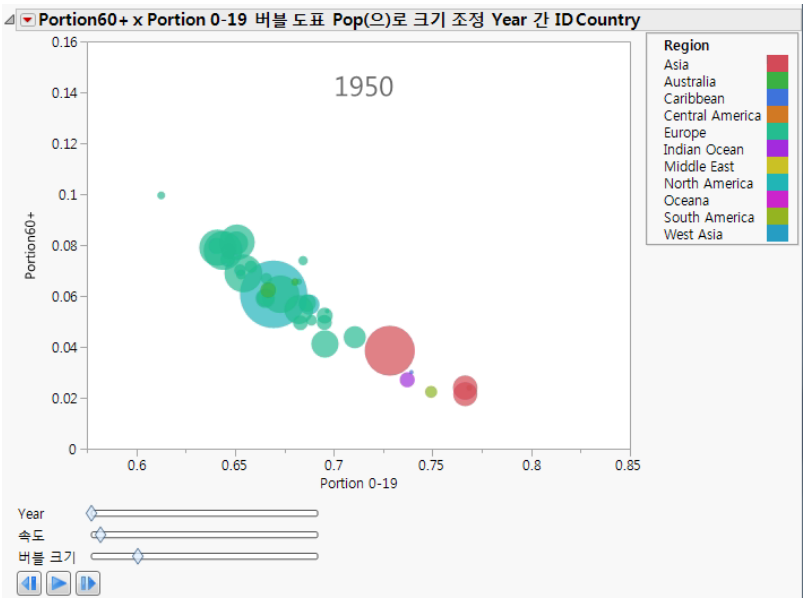


그림 4.26에서는 ATM의 매출 및 제품 비용이 가장 높으며 가장 빠르게 증가하고 있음을 알 수 있습니다.

버블 그림을 사용하여 여러 변수 비교

버블 그림은 점을 버블로 나타내는 산점도입니다. 버블의 크기와 색상을 변경하고 시간 경과에 따라 애니메이션도 적용할 수 있습니다. 버블 그림은 최대 5개의 차원(x 위치, y 위치, 크기, 색상 및 시간)을 나타낼 수 있기 때문에 인상적인 시각적 표현을 생성하며 이를 통해 데이터를 손쉽게 탐색할 수 있습니다.

그림 4.27 버블 그림의 예



시나리오

이 예에서는 1950년부터 2004년까지 116개 국가 또는 지역의 인구 통계가 포함된 PopAgeGroup.jmp 데이터 테이블을 사용합니다. 총 인구 수는 연령대별로 분류되며, 모든 국가의 데이터가 매년 있지는 않습니다.

사회학자가 다음과 같은 질문에 대한 답을 구하려고 합니다.

- 세계 인구의 연령대가 바뀌고 있습니까?

이 질문에 답하려면 인구 중 가장 높은 연령대(60세 이상)와 가장 낮은 연령대(20세 미만) 간의 관계를 살펴보아야 합니다. 버블 그림을 사용하여 이 관계가 시간에 따라 어떻게 변하는지 확인하십시오.

버블 그림 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 PopAgeGroup.jmp 를 엽니다.
2. **그래프 > 버블 그림**을 선택합니다.
3. Portion60+ 를 선택하고 **Y** 를 클릭합니다.
   이 열은 버블 그림의 *Y* 변수에 해당합니다.
4. Portion 0-19 를 선택하고 **X** 를 클릭합니다.
   이 열은 버블 그림의 *X* 변수에 해당합니다.
5. Country 를 선택하고 **ID** 를 클릭합니다.

"ID" 변수의 고유 수준이 각각 그림에 버블로 표시됩니다 .

6. Year 를 선택하고 **시간**을 클릭합니다 .

   버블 그림에 애니메이션이 적용될 때 이 열에 따라 시간 인덱싱이 제어됩니다 .

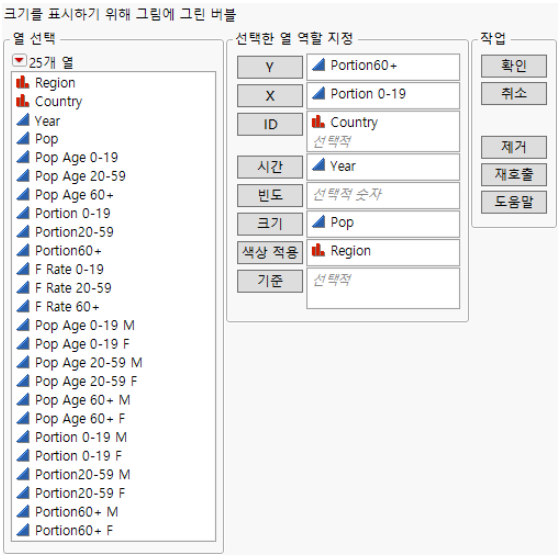
7. Pop 을 선택하고 **크기**를 클릭합니다 .

   이 열에 따라 버블의 크기가 제어됩니다 .

8. Region 을 선택하고 **색상 적용**을 클릭합니다 .

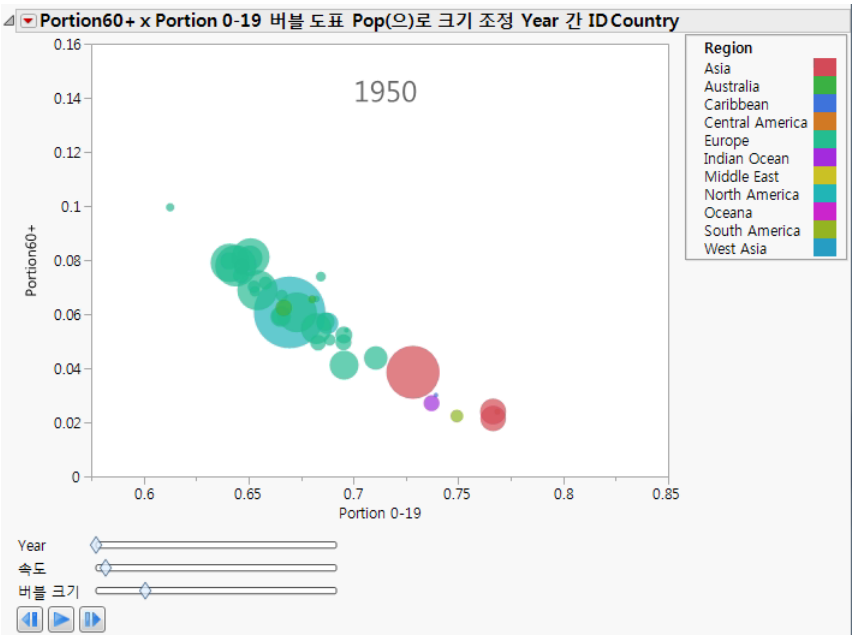
   " 색상 적용 " 변수의 각 수준에 고유한 색상이 할당됩니다 . 따라서 이 예에서 같은 지역에 위치한 국가의 모든 버블은 색상이 동일합니다 . 그림 4.29 에 나타나는 버블 색상은 JMP 기본 색상입니다 .

그림 4.28 버블 그림 시작 창



9. **확인**을 클릭합니다 .

그림 4.29 초기 버블 그림



버블 그림 해석

시간 변수 (이 예의 경우 연도) 가 1950 년부터 시작되기 때문에 초기 버블 그림은 1950 년의 데이터를 보여 줍니다 . 재생 / 일시 중지 버튼을 클릭하여 모든 연도를 순환하도록 버블 그림에 애니메이션을 적용하십시오 . 각각의 연이은 버블 그림은 해당 연도의 데이터를 보여 줍니다 . 각 연도의 데이터에 따라 다음이 결정됩니다 .

- *X* 및 *Y* 좌표
- 버블 크기
- 버블 색상
- 버블 집계

참고 : 버블 그림에서 여러 행의 정보가 집계되는 방식에 대한 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .

1950 년의 버블 그림에서는 국가의 20 세 미만 인구 비율이 높으면 60 세 이상 인구 비율이 낮다는 것을 보여 줍니다 .

버블 그림의 전체 연도 범위에 애니메이션을 적용하려면 재생 / 일시 중지 버튼을 클릭하십시오 . 시간이 지나면서 Portion 0-19 의 비율이 감소하고 Portion60+ 의 비율은 증가합니다 .

애니메이션을 재생하며 , 클릭 후에는 일시 중지 버튼으로 바뀝니다 .

애니메이션을 일시 중지합니다 .

애니메이션을 시간 단위 하나만큼 뒤로 수동 제어합니다 .

애니메이션을 시간 단위 하나만큼 앞으로 수동 제어합니다 .

Year 시간 인덱스를 수동으로 변경합니다 .

속도 애니메이션의 속도를 제어합니다 .

버블 크기 상대 크기를 유지하면서 버블의 절대 크기를 제어합니다 .

사회학자는 세계 인구의 연령이 어떻게 변화하고 있는지 알고 싶어했습니다. 이 버블 그림은 세계 인구가 점점 고령화되고 있음을 보여 줍니다.

버블 그림에서의 상호 작용

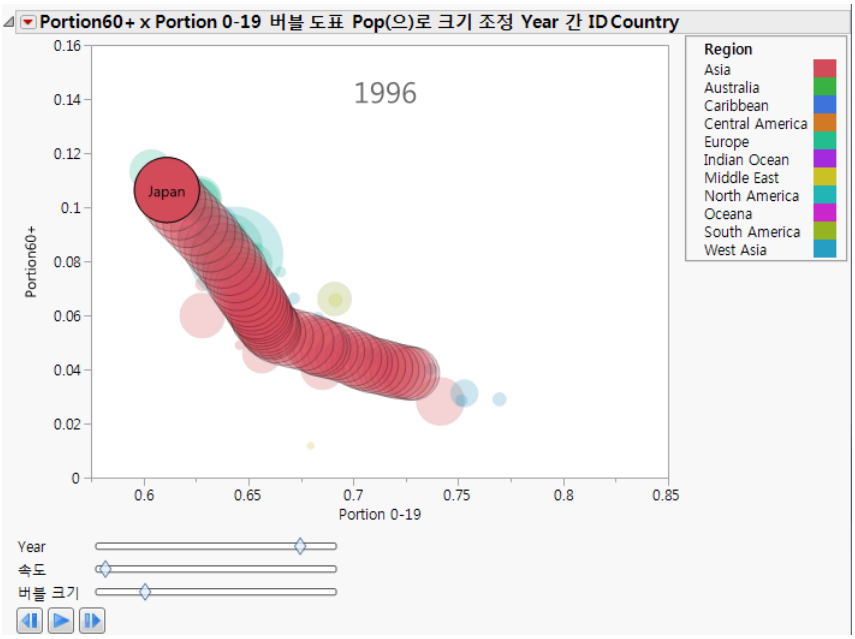
버블을 클릭하여 선택하면 시간 경과에 따른 버블의 추세를 확인할 수 있습니다. 예를 들어 1950 년 그림에서 중간에 있는 큰 버블은 일본의 데이터입니다.

수년에 걸친 일본의 인구 변화 패턴을 보려면 다음을 수행하십시오 .

1. 일본 버블의 가운데를 클릭하여 선택합니다 .
2. " 버블 그림 " 의 빨간색 삼각형을 클릭하고 **이동 경로 버블 > 선택됨**을 선택합니다 .
3. 재생 버튼을 클릭합니다 .

애니메이션이 시간이 지남에 따라 진행되면서 일본 버블은 그 기록을 보여 주는 자취를 남깁니다.

그림 4.30 일본의 인구 변동 기록



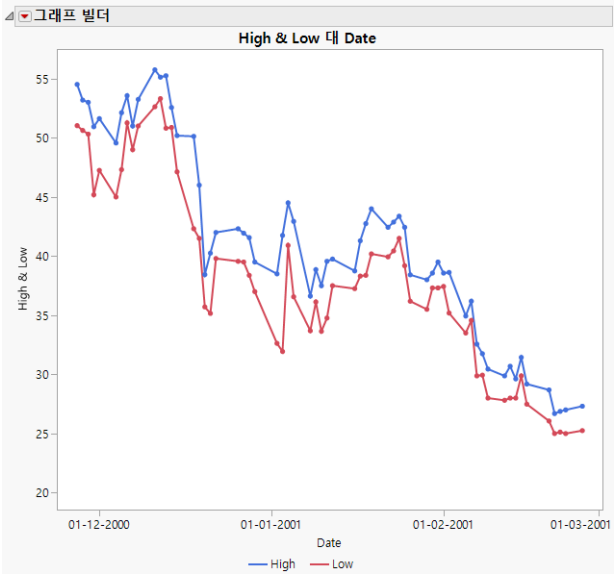
일본 버블에 초점을 맞추면 시간 경과에 따라 다음 정보를 확인할 수 있습니다 .

- 19 세 이하 인구 비율이 감소했습니다 .
- 60 세 이상 인구 비율이 증가했습니다 .

중첩 그림을 사용하여 여러 변수 비교

중첩 그림은 산점도와 마찬가지로 두 개 이상의 변수 사이에 존재하는 관계를 보여 줍니다. 그러나 변수 중 하나가 시간 변수인 경우에는 산점도보다 중첩 그림이 시간에 따른 추세를 더 잘 나타냅니다.

그림 4.31 중첩 그림의 예



참고 : 시간에 따른 데이터 그림을 생성하려는 경우 버블 그림 , 관리도 및 계량형 차트를 사용할 수도 있습니다 . 그래프 빌더와 버블 그림에 대한 자세한 내용은 Essential Graphing 의 에서 확인하십시오 . 관리도와 계량형 차트에 대한 자세한 내용은 Quality and Process Methods 의 및 에서 확인하십시오 .

시나리오

이 예에서는 3 개월 동안의 주가에 대한 데이터가 포함된 Stock Prices.jmp 데이터 테이블을 사용합니다.

잠재 투자자가 다음과 같은 질문에 대한 답을 구하려고 합니다 .

- 지난 3 개월 동안 주식 종가에 변화가 있습니까 ?

  이 질문에 답하려면 시간 경과에 따른 주식 종가의 중첩 그림을 사용해야 합니다 .
- 주식의 고가와 저가는 서로 어떤 관련이 있습니까 ?

  이 질문에 답하려면 시간 경과에 따른 주식의 고가 및 저가를 나타내는 또 다른 중첩 그림을 사용해야 합니다 .

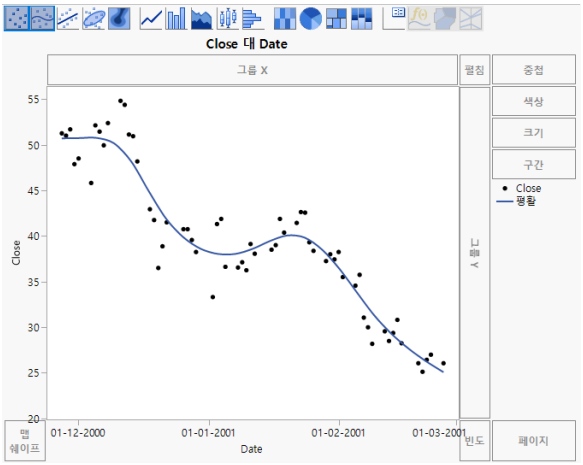
첫 번째 질문에 답하기 위해 첫 번째 중첩 그림을 생성한 후 두 번째 질문에 답하기 위해 두 번째 중첩 그림을 생성합니다.

시간 경과에 따른 주가의 중첩 그림 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Stock Prices.jmp 를 엽니다 .

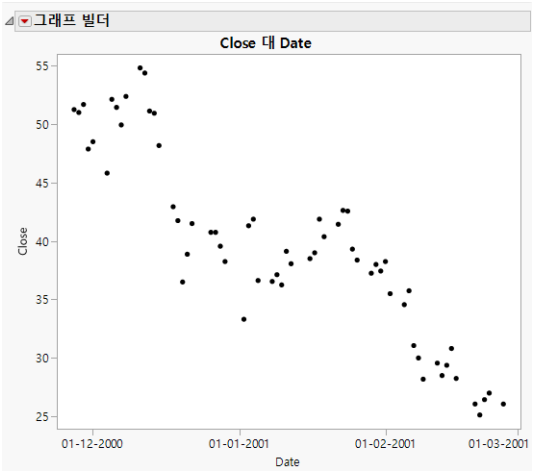
2. **그래프 > 그래프 빌더**를 선택합니다.
3. Close 를 선택하고 **Y** 를 클릭합니다.
4. Date 를 선택하고 **X** 를 클릭합니다.

그림 4.32 평활기를 사용한 중첩 그림



5. Ctrl 키를 누르고 그래프 위의 평활기 아이콘 을 클릭하여 평활기 선을 제거합니다.

그림 4.33 시간에 따른 종가의 중첩 그림

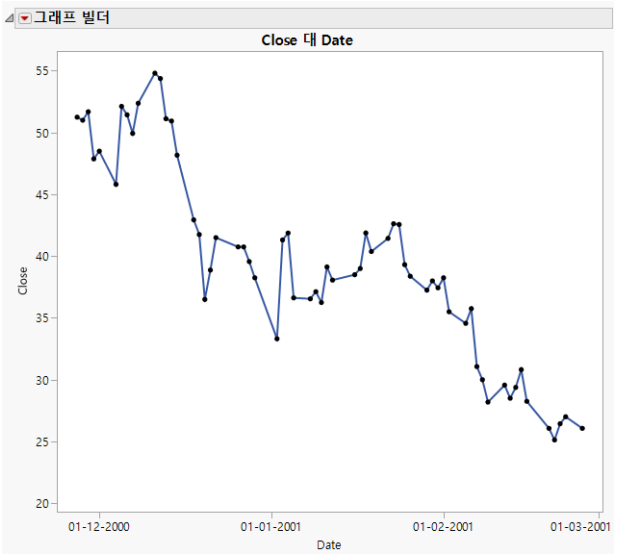


중첩 그림 해석 및 상호 작용

이 중첩 그림에서는 주식 종가가 지난 몇 개월 동안 하락하고 있음을 보여 줍니다 . 추세를 더 명확하게 보려면 점을 연결하십시오 .

1. Shift 키를 누르고 그래프 위의 선 아이콘 을 클릭합니다 .

그림 4.34 연결된 점



이로써 잠재 투자자는 지난 3 개월 동안 주가가 오르내렸지만 전반적으로는 하락 추세였다는 것을 알 수 있습니다 .

주식의 고가 및 저가에 대한 중첩 그림 생성

중첩 그림을 사용하여 둘 이상의 Y 변수에 대한 그림을 생성할 수 있습니다 . 예를 들어 고가와 저가를 모두 동일한 그림에서 보고 싶다고 가정해 보겠습니다 .

1. " 시간 경과에 따른 주가의 중첩 그림 생성 "(119 페이지) 의 단계를 따르되 , 이번에는 **Y** 역할에 High 와 Low 를 할당합니다 .
2. " 중첩 그림 해석 및 상호 작용 "(121 페이지) 에 표시된 대로 점을 연결하고 격자선을 추가합니다 .

그림 4.35 두 개의 Y 변수

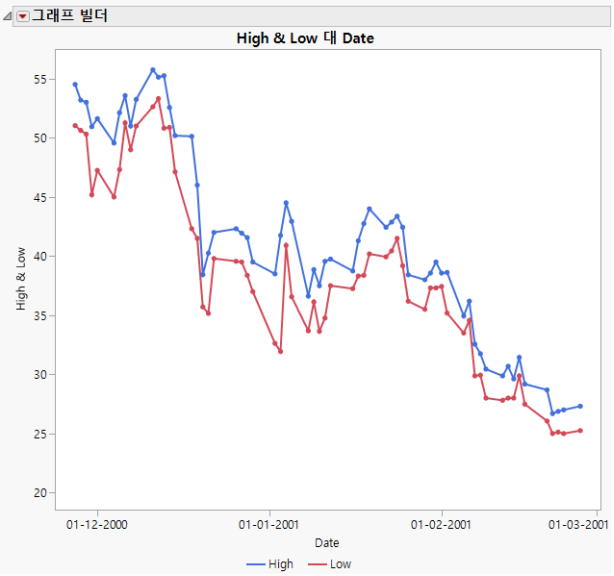


그림 하단의 범례는 그래프에서 High 및 Low 변수에 사용된 색상과 표식을 보여 줍니다. 이 중첩 그림은 High 가격과 Low 가격이 서로 매우 밀접하게 얽혀 있음을 보여 줍니다.

질문에 대한 답

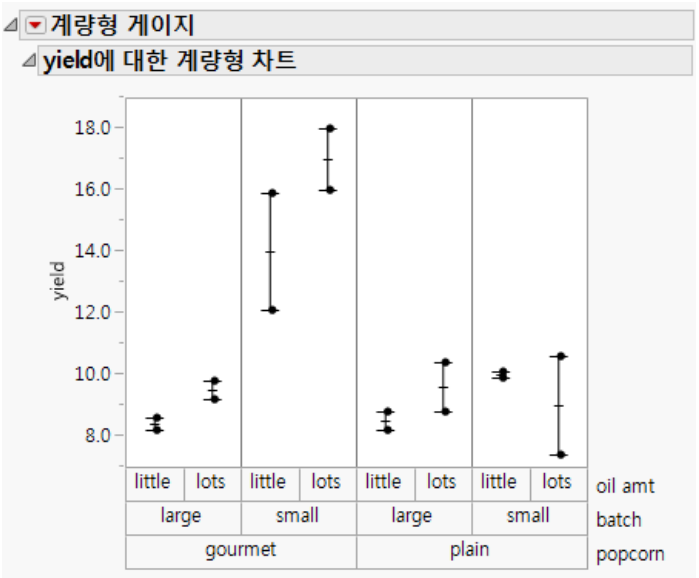
두 중첩 그림은 이 예의 시작 부분에서 제기했던 두 가지 질문에 대한 답을 제공합니다.

- 첫 번째 그림은 이 주식의 가격이 같은 수준으로 유지되지 않고 하락해 왔음을 보여 줍니다.
- 두 번째 그림은 이 주식의 고가와 저가 사이에 그다지 차이가 없다는 것을 보여 줍니다. 즉, 일간 주가 변동이 그리 크지 않습니다.

계량형 차트를 사용하여 여러 변수 비교

지금까지 설명한 그래프에서는 단일 X 변수만 지정했습니다. 계량형 차트를 사용하면 여러 개의 X 변수를 지정하고 모든 변수의 평균 및 변동성 차이를 한 번에 볼 수 있습니다.

그림 4.36 계량형 차트의 예



시나리오

이 예에서는 팝콘 제조업체의 데이터가 포함된 Popcorn.jmp 데이터 테이블을 사용합니다. 이 테이블에는 팝콘 스타일, 배치 크기 및 기름 사용량의 각 조합에 따른 산출량(주어진 낟알 측정값에 대한 팝콘 부피)을 측정한 결과가 있습니다.

팝콘 제조업체에서 다음과 같은 질문에 대한 답을 구하려고 합니다 .

- 어떤 요인의 조합에서 팝콘 산출량이 가장 높습니까 ?

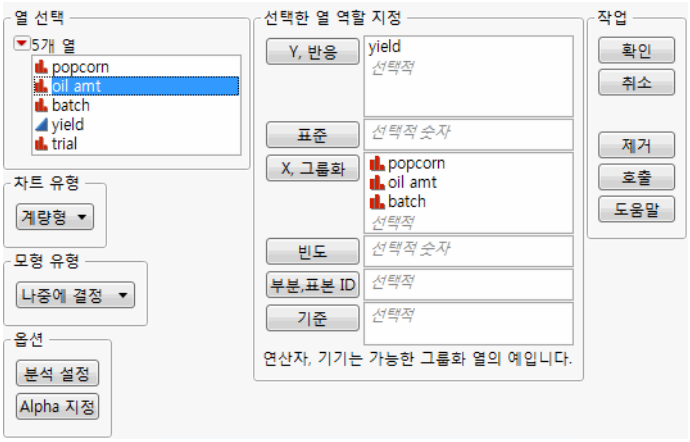
이 질문에 답하려면 산출량과 스타일, 배치 크기 및 기름 양의 관계를 보여 주는 계량형 차트를 사용해야 합니다.

계량형 차트 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Popcorn.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트**를 선택합니다 .
3. yield 를 선택하고 **Y, 반응**을 클릭합니다 .
4. popcorn 을 선택하고 **X, 그룹화**를 클릭합니다 .
5. batch 를 선택하고 **X, 그룹화**를 클릭합니다 .
6. oil amt 을 선택하고 **X, 그룹화**를 클릭합니다 .

참고 : 이 창에서의 순서에 따라 계량형 차트의 내포 순서가 결정되므로 변수를 **X, 그룹화** 역할에 할당하는 순서가 중요합니다 .

그림 4.37 계량형 차트 창

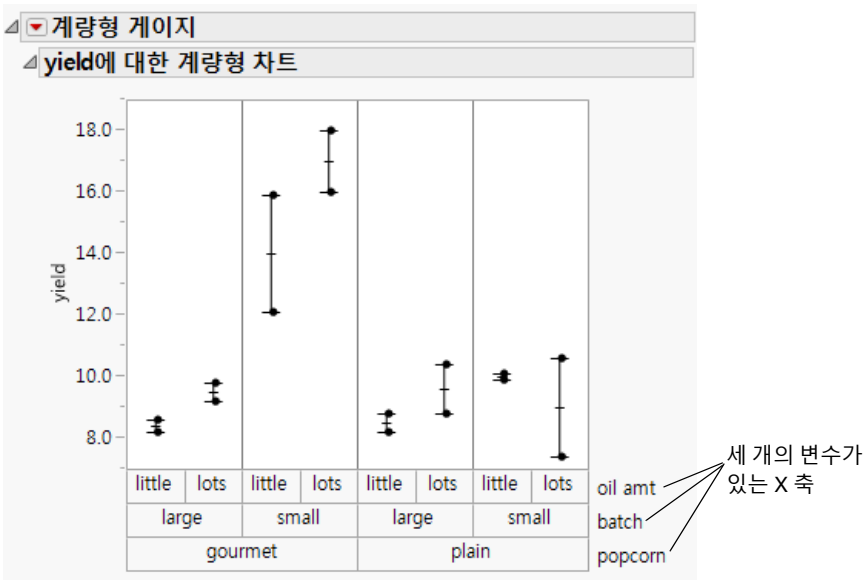


7. **확인**을 클릭합니다.

   위쪽 차트는 계량형 차트로, 세 변수의 조합에 따라 분류된 산출량을 보여 줍니다. 아래쪽 차트는 세 변수의 각 조합에 대한 표준편차를 보여 줍니다. 아래쪽 차트는 산출량을 표시하지 않으므로 숨겨 둡니다.

8. " 계량형 게이지 " 의 빨간색 삼각형을 클릭하고 **S 관리도**를 선택 취소합니다.

그림 4.38 결과 창



계량형 차트 해석

산출량의 계량형 차트는 작은 고메이 배치의 산출량이 가장 높음을 나타냅니다 .

팝콘 제조업체는 보다 구체적으로 다음과 같은 추가 질문을 제기할 수 있습니다. 산출량이 높은 이유가 이 배치가 작기 때문입니까, 이 배치가 고메이이기 때문입니까?

계량형 차트는 다음을 보여 줍니다 .

- 작은 플레인 배치의 산출량이 낮습니다 .
- 큰 고메이 배치의 산출량이 낮습니다 .

이 정보를 감안할 때 팝콘 제조업체는 작은 크기와 고메이 스타일의 조합에서 배치의 산출량이 높다고 결론 내릴 수 있습니다. 단일 변수만 허용하는 차트로는 이 결론에 도달하는 것이 불가능했을 것입니다.

5 장

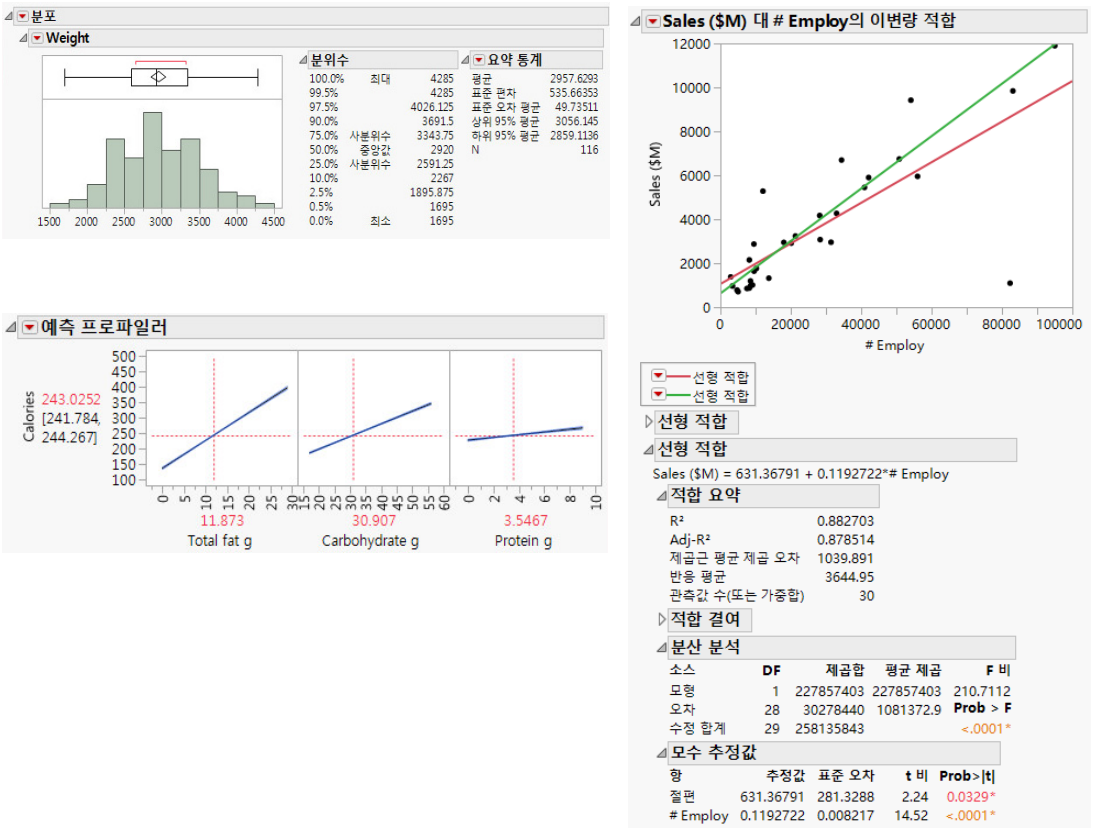
데이터 분석

분포 , 관계 및 모형

JMP에서 데이터를 분석하면 정보를 기반으로 의사 결정을 내릴 수 있습니다. 데이터 분석에는 대개 다음과 같은 작업이 포함됩니다.

- 분포 조사
- 관계 발견
- 가설 검정
- 모형 구축

그림 5.1 분석 예



목차

내용 소개 129
데이터 그래프 생성의 중요성 129
모델링 유형 이해 132
예: 모델링 유형 결과 보기 133
모델링 유형 변경 134
분포 분석 136
연속형 변수의 분포 137
범주형 변수의 분포 139
관계 분석 142
하나의 예측 변수가 있는 회귀 사용 143
한 변수에 대한 평균 비교 148
비율 비교 151
여러 변수의 평균 비교 153
다중 예측 변수가 있는 회귀 사용 157

내용 소개

데이터를 분석하기 전에 기본 개념에 대한 다음 정보를 검토하십시오 .

- " 데이터 그래프 생성의 중요성 "(129 페이지)
- " 모델링 유형 이해 "(132 페이지)

이 장의 나머지 부분에서는 JMP 에서 몇 가지 기본 분석 방법을 사용하는 방법을 소개합니다 .

- " 분포 분석 "(136 페이지)
- " 관계 분석 "(142 페이지)

고급 예측 모형 및 분석 방법에 대한 설명은 다음 JMP 설명서에서 확인하십시오 .

- Fitting Linear Models
- Multivariate Methods
- Predictive and Specialized Modeling
- Consumer Research
- Reliability and Survival Methods
- Quality and Process Methods

데이터 그래프 생성의 중요성

데이터를 그래프로 생성하거나 시각화하는 것은 모든 데이터 분석에 중요하며 항상 통계 검정 또는 모형 구축 전에 수행되어야 합니다 . 데이터 시각화가 데이터 분석 프로세스의 초기 단계에 이루어져야 하는 이유를 분명하게 파악하려면 다음 예를 참조하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Anscombe.jmp(F. J. Anscombe(1973), American Statistician, 27, 17-21) 를 엽니다 .

   이 데이터는 4 쌍의 *X* 및 *Y* 변수로 구성되어 있습니다 .
2. 테이블 패널에서 **The Quartet** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   이 스크립트는 **X 로 Y 적합**을 사용하여 각 변수 쌍에 대한 단순 선형 회귀를 생성합니다 . **점 표시** 옵션이 해제되어 있으므로 산점도에서 어떤 데이터도 볼 수 없습니다 . 그림 5.2 에서는 각 회귀에 대한 모형 적합 및 기타 요약 정보를 보여 줍니다 .

그림 5.2 네 가지 모형

쌍 1 에 대한 모형

선형 적합

Y1 = 3.0000909 + 0.5000909*X1

적합 요약

R^2	0.666542
Adj-R^2	0.629492
제곱근 평균 제곱 오차	1.236603
반응 평균	7.500909
관측값 수(또는 가중합)	11

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	1	27.510001	27.5100	17.9899
오차	9	13.762690	1.5292	Prob > F
수정 합계	10	41.272691		0.0022*

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	3.0000909	1.124747	2.67	0.0257*
X1	0.5000909	0.117906	4.24	0.0022*

쌍 3 에 대한 모형

선형 적합

Y3 = 3.0024545 + 0.4997273*X3

적합 요약

R^2	0.666324
Adj-R^2	0.629249
제곱근 평균 제곱 오차	1.236311
반응 평균	7.5
관측값 수(또는 가중합)	11

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	1	27.470008	27.4700	17.9723
오차	9	13.756192	1.5285	Prob > F
수정 합계	10	41.226200		0.0022*

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	3.0024545	1.124481	2.67	0.0256*
X3	0.4997273	0.117878	4.24	0.0022*

쌍 2 에 대한 모형

선형 적합

Y2 = 3.0009091 + 0.5*X2

적합 요약

R^2	0.666242
Adj-R^2	0.629158
제곱근 평균 제곱 오차	1.237214
반응 평균	7.500909
관측값 수(또는 가중합)	11

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	1	27.500000	27.5000	17.9656
오차	9	13.776291	1.5307	Prob > F
수정 합계	10	41.276291		0.0022*

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	3.0009091	1.125302	2.67	0.0258*
X2	0.5	0.117964	4.24	0.0022*

쌍 4 에 대한 모형

선형 적합

Y4 = 3.0017273 + 0.4999091*X4

적합 요약

R^2	0.666707
Adj-R^2	0.629675
제곱근 평균 제곱 오차	1.235695
반응 평균	7.500909
관측값 수(또는 가중합)	11

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	1	27.490001	27.4900	18.0033
오차	9	13.742490	1.5269	Prob > F
수정 합계	10	41.232491		0.0022*

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	3.0017273	1.123921	2.67	0.0256*
X4	0.4999091	0.117819	4.24	0.0022*

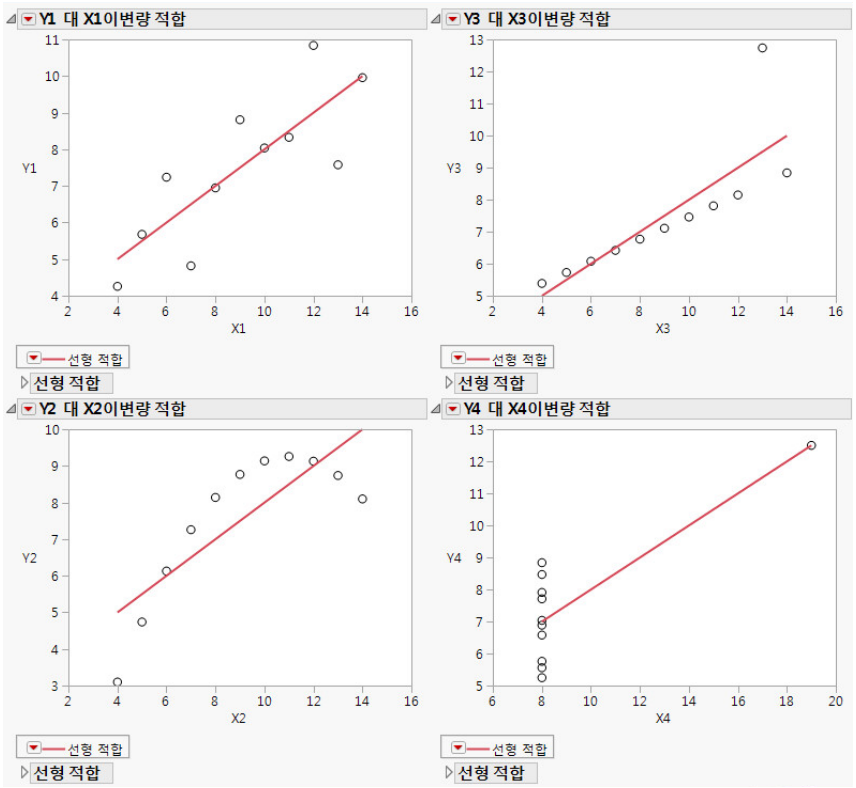
네 가지 모형과 R^2 값이 거의 동일하다는 것에 유의하십시오. 각 사례에서 적합 모형은 기본적으로 $Y = 3 + 0.5X$ 이며 R^2 값은 기본적으로 0.66 입니다. 데이터 분석에서 위의 요약 정보만 고려하면 각 사례의 X 와 Y 간 관계가 동일하다고 결론 내릴 수 있습니다. 그러나 이 시점에는 데이터를 시각화하지 않았습니다. 결론이 잘못되었을 수도 있습니다.

데이터 시각화를 위해 네 가지 산점도 모두에 점 추가

1. Ctrl 키를 누릅니다 .
2. " 이변량 적합 " 중 하나의 옆에 있는 빨간색 삼각형을 클릭하고 **점 표시**를 선택합니다 .

그림 5.3 점이 추가된 산점도

쌍 1 에 대한 산점도



쌍 3 에 대한 산점도

쌍 2 에 대한 산점도

쌍 4 에 대한 산점도

산점도는 관계를 나타내는 선이 같더라도 *X* 와 *Y* 간의 관계가 네 쌍에서 동일하지 않음을 보여 줍니다 .

- 산점도 1 은 선형 관계를 나타냅니다 .
- 산점도 2 는 비선형 관계를 나타냅니다 .
- 산점도 3 은 하나의 이상치를 제외하고 선형 관계를 나타냅니다 .
- 산점도 4 에서는 한 점을 제외하고 x = 8 에 모든 데이터가 있습니다 .

이 예는 통계만을 기반으로 한 결론이 부적절할 수 있음을 보여 줍니다. 모든 데이터 분석의 초기 단계에는 데이터를 시각적으로 탐색하는 과정이 있어야 합니다.

모델링 유형 이해

JMP 에서는 데이터가 여러 가지 유형일 수 있습니다 . JMP 에서는 이것을 데이터의 모델링 유형이라고 합니다 . 표 5.1 에서는 JMP 의 세 가지 모델링 유형을 보여 줍니다 .

표 5.1 모델링 유형

모델링 유형 및 설명	예	구체적인 예
연속형 숫자 데이터에만 해당됩니다. 합계 및 평균 같은 연산에서 사용됩니다.	높이 온도 시간	시험을 완료하는 데 소요되는 시간은 2 시간 또는 2.13 시간일 수 있습니다.
순서형 숫자 또는 문자 데이터에 해당됩니다. 값은 정렬된 범주에 속합니다.	월 (1,2,...,12) 문자 등급 (A, B,...F) 크기 (소, 중, 대)	월은 2(2 월) 또는 3(3 월) 일 수 있지만 2.13 일 수 없습니다. 2 월은 3 월 전에 옵니다.
명목형 숫자 또는 문자 데이터에 해당됩니다. 값은 범주에 속하지만 순서는 중요하지 않습니다.	성별 (남 또는 여) 색상 시험 결과 (통과 또는 실패)	성별은 남 또는 여가 될 수 있으며 순서는 없습니다. 성별 범주를 숫자 (남 = 1 및 여 = 2) 로 나타낼 수도 있습니다.
다중 반응 문자 데이터에만 해당됩니다. 단일 셀에서 쉼표로 구분된 고유 항목입니다.	이를 닦는 시간 대학 학위 즐겨 하는 스포츠	하루에 여러 번 이를 닦을 수 있습니다. 아침에 맨 먼저 , 아침 식사 후 , 식사 후 , 잠자기 전 또는 이들의 조합입니다.
비정형 텍스트 문자 데이터에만 해당됩니다. 일반적으로 텍스트 탐색기를 사용하여 분석해야 하는 모든 고유 값입니다.	제품 후기 노래 가사 설문 조사의 자유 의견 작성 필드	제품 후기는 대부분 고유하며 텍스트 탐색기를 사용하여 기본적인 유사성을 결정합니다.
벡터 표현식 데이터에만 해당됩니다. 셀의 값은 열 벡터 또는 행 벡터입니다.	예측 계산식	

표 5.1 모델링 유형 (계속)

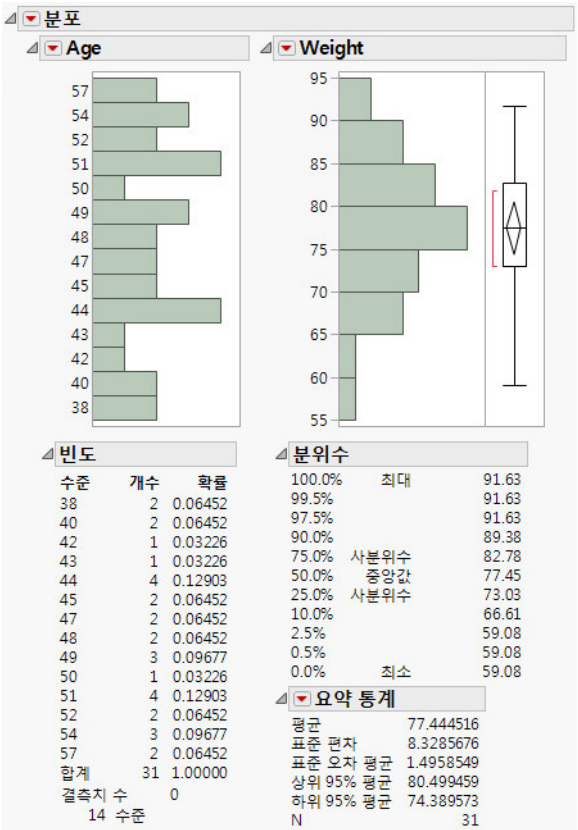
모델링 유형 및 설명	예	구체적인 예
없음 모든 데이터 유형에 해당됩니다. 열이 다른 모델링 유형으로 잘 표현되지 않는 시나리오에서 사용됩니다.	그림 ID 값	데이터 테이블의 그림 열은 모델링에 사용되지 않지만 그래프에서 표식 등으로 사용할 수 있습니다.

예 : 모델링 유형 결과 보기

모델링 유형이 다르면 JMP 에서 생성되는 결과도 달라집니다 . 차이점을 보려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Linnerud.jmp 를 엽니다 .
2. **분석** > **분포**를 선택합니다 .
3. Age 및 Weight 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

그림 5.4 Age 및 Weight 에 대한 분포 결과



Age 및 Weight 는 모두 숫자 변수이지만 동일하게 처리되지 않습니다. 표 5.2 에서는 Weight 및 Age 에 대한 결과 간의 차이점을 비교합니다.

표 5.2 Weight 및 Age 에 대한 결과

변수	모델링 유형	결과
Weight	연속형	히스토그램, 사분위수 및 요약 통계량
Age	순서형	막대 차트 및 빈도

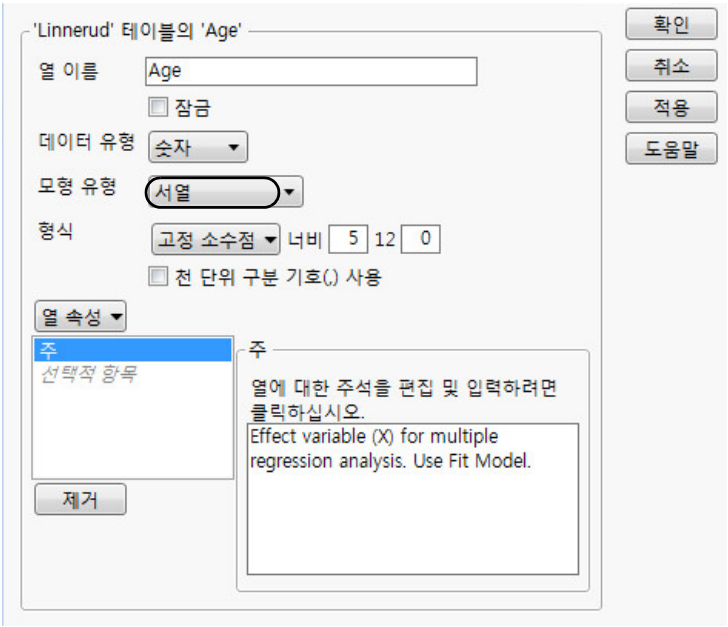
모델링 유형 변경

변수를 다르게 처리하려면 모델링 유형을 변경하십시오 . 예를 들어 그림 5.4 에서 Age 의 모델링 유형은 순서형입니다 . 순서형 변수에서 JMP 는 빈도 수를 계산합니다 . 빈도 수 대신 평균 연령

을 확인하려고 한다고 가정해 보겠습니다 . 모델링 유형을 평균 연령을 보여 주는 연속형으로 변경합니다 .

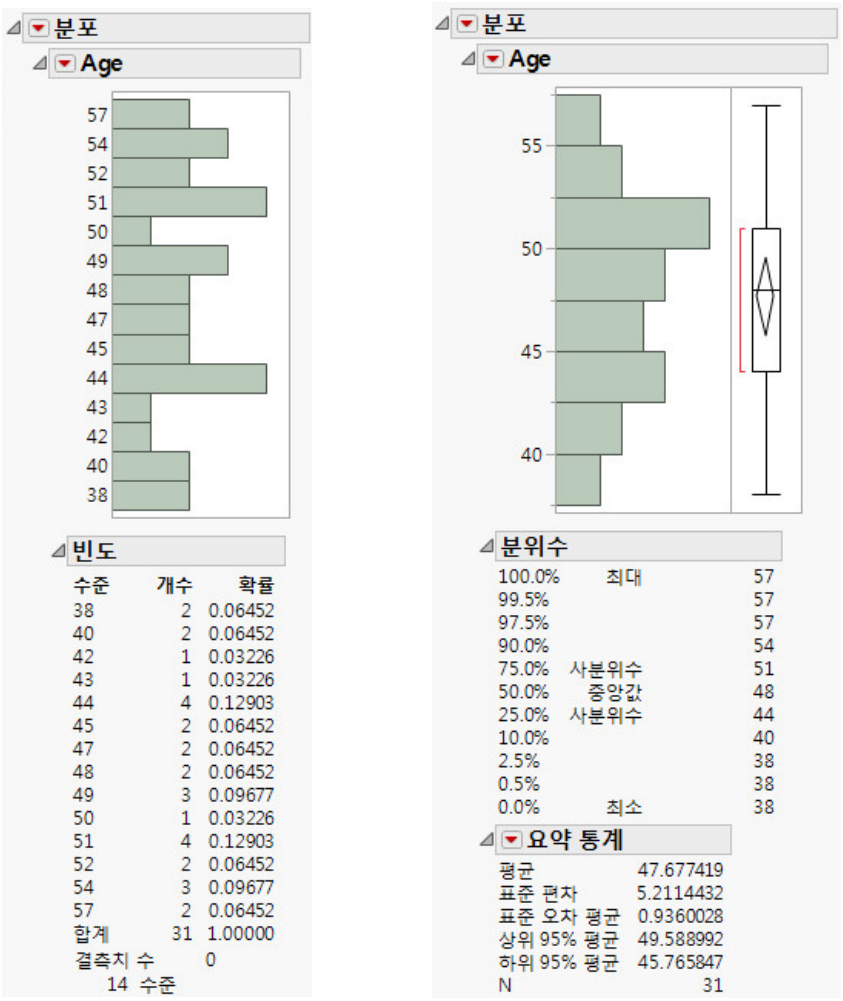
1. Age 열 머리글을 두 번 클릭합니다 . " 열 정보 " 창이 나타납니다 .
2. " 모델링 유형 " 을 **연속형**으로 변경합니다 .

그림 5.5 열 정보 창



3. **확인**을 클릭합니다 .
4. 이 예의 단계를 반복하여 (" 예 : 모델링 유형 결과 보기 "(133 페이지) 참조) 분포를 생성합니다 . 그림 5.6 에서는 Age 가 순서형일 때와 연속형일 때의 분포 결과를 보여 줍니다 .

그림 5.6 Age 의 서로 다른 모델링 유형



순서형 연속형

Age가 순서형이면 각 연령의 빈도 수를 볼 수 있습니다. 예를 들어 연령 48은 두 번 나타납니다. 연령이 연속형이면 평균 연령(이 예의 경우 약 48세, 즉 47.677세)을 확인할 수 있습니다.

분포 분석

단일 변수를 분석하려는 경우 분포 플랫폼을 사용하여 변수의 분포를 조사할 수 있습니다. 각 변수의 보고서 내용은 변수가 범주형(명목형 또는 순서형)인지 연속형인지에 따라 달라집니다.

참고 : 분포 플랫폼에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오 .

연속형 변수의 분포

연속형 변수를 분석할 때는 다음과 같은 질문을 던질 수 있습니다 .

- 데이터의 모양이 알려진 분포와 매칭됩니까 ?
- 데이터에 이상치가 있습니까 ?
- 데이터의 평균은 얼마입니까 ?
- 평균이 목표값 또는 과거 값과 통계적으로 다릅니까 ?
- 데이터가 얼마나 퍼져 있습니까 ? 즉 , 표준편차는 얼마입니까 ?
- 최소값과 최대값은 얼마입니까 ?

그래프, 요약 통계량 및 간단한 통계 검정을 통해 이러한 질문과 그 밖의 질문에 대한 답을 얻을 수 있습니다.

시나리오

이 예에서는 116가지 자동차 모델에 대한 정보가 들어 있는 Car Physical Data.jmp 데이터 테이블을 사용합니다.

계획 전문가가 철도 회사로부터 기차로 자동차를 운송할 때 발생할 수 있는 문제를 파악해 달라는 요청을 받았습니다 . 계획 전문가는 데이터를 사용하여 다음과 같은 질문에 대한 답을 구하려고 합니다 .

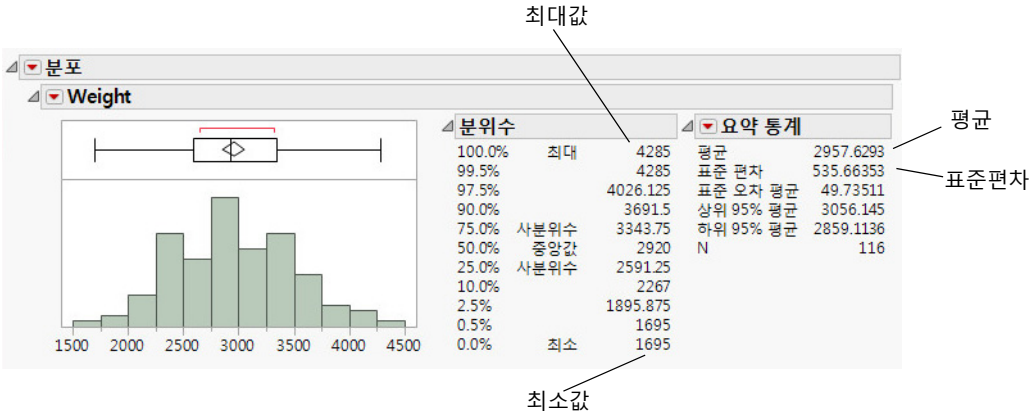
- 평균 자동차 중량은 얼마입니까 ?
- 자동차의 중량은 얼마나 퍼져 있습니까 (표준편차)?
- 자동차의 최소 중량과 최대 중량은 얼마입니까 ?
- 데이터에 이상치가 있습니까 ?

이러한 질문에 답하기 위해 중량에 대한 히스토그램을 사용합니다.

히스토그램 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Physical Data.jmp 를 엽니다 .
2. **분석** > **분포**를 선택합니다 .
3. Weight 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. 보고서 창을 회전하려면 "Weight" 의 빨간색 삼각형을 클릭하고 **표시 옵션** > **가로 레이아웃**을 선택합니다 .

그림 5.7 Weight 분포



보고서 창에는 다음 세 가지 섹션이 있습니다 .

- 히스토그램과 상자 그림은 데이터를 시각화합니다 .
- 분위수 보고서는 분포의 백분위수를 보여 줍니다 .
- 요약 통계량 보고서는 평균 , 표준편차 및 기타 통계량을 보여 줍니다 .

분포 결과 해석

계획 전문가는 그림 5.7 에 제시된 결과를 기반으로 질문에 답할 수 있습니다 .

평균 자동차 중량은 얼마입니까 ? 히스토그램은 중량이 약 3,000 파운드임을 보여 줍니다 .

자동차의 중량은 얼마나 퍼져 있습니까 (표준편차)? 요약 통계량에서 중량은 약 2,958 파운드이고 표준편차는 약 536 파운드입니다 .

최소 중량과 최대 중량은 얼마입니까 ? 히스토그램에서 최소 중량은 약 1,500 파운드이고 최대 중량은 약 4,500 파운드입니다 . 분위수에서 최소 중량은 약 1,695 파운드이고 최대 중량은 약 4,285 파운드입니다 .

이상치가 있습니까 ? 없습니다 .

그림 5.7의 기본 보고서 창에서는 최소한의 그래프와 통계량이 제공됩니다. 추가적인 그래프 및 통계량은 빨간색 삼각형 메뉴를 클릭하여 확인할 수 있습니다.

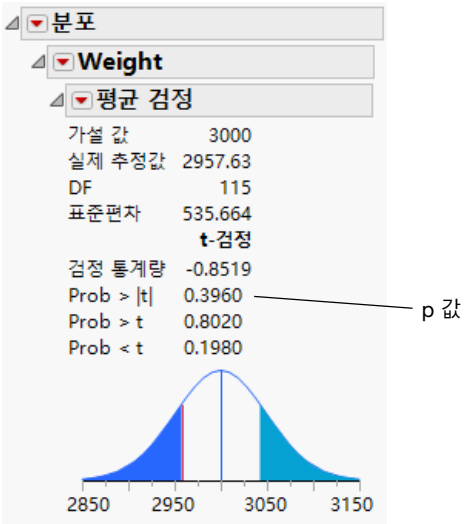
결론

다른 조사 결과에 따라, 철도 회사는 평균 3,000파운드를 운송하는 것이 가장 효율적이라고 판단했습니다. 이제 계획 전문가는 운송할 수 있는 일반적인 자동차 모집단의 평균 자동차 중량이 3,000파운드인지 알아내야 합니다. *t*-검정을 통해 이 모집단 표본을 바탕으로 보다 광범위한 모집단에 대한 추론을 이끌어 내야 합니다.

결론 검정

1. "Weight" 의 빨간색 삼각형을 클릭하고 **평균 검정**을 선택합니다 .
2. 창이 나타나면 " 가설 평균 지정 " 상자에 3000 을 입력합니다 .
3. **확인**을 클릭합니다 .

그림 5.8 평균 검정 결과



t- 검정 해석

t-검정의 기본 결과는 *p* 값입니다. 이 예에서 *p* 값은 0.396이고 분석가는 0.05의 유의 수준을 사용합니다. 0.396은 0.05보다 크기 때문에 더 광범위한 모집단에서 자동차 모델의 평균 중량이 3,000파운드와 유의하게 다르다는 결론을 내릴 수는 없습니다. *p* 값이 유의 수준보다 낮았다면 계획 전문가는 더 광범위한 모집단의 평균 자동차 중량이 3,000파운드와 유의하게 다르다는 결론을 내렸을 것입니다.

범주형 변수의 분포

범주형 (순서형 또는 명목형) 변수를 분석할 때는 다음과 같은 질문을 던질 수 있습니다 .

- 변수에는 몇 개의 수준이 있습니까 ?
- 각 수준에는 몇 개의 데이터 점이 있습니까 ?
- 데이터가 균등하게 분포되어 있습니까 ?
- 각 수준이 차지하는 비율은 어느 정도입니까 ?

시나리오

자세한 내용은 "연속형 변수의 분포"(137 페이지)의 시나리오에서 확인하십시오.

철도 회사에서는 자동차의 평균 중량이 목표 중량과 유의하게 다르지 않다고 판단했으므로 해결해야 할 질문이 더 있습니다.

계획 전문가는 철도 회사의 다음 질문에 대답하고자 합니다.

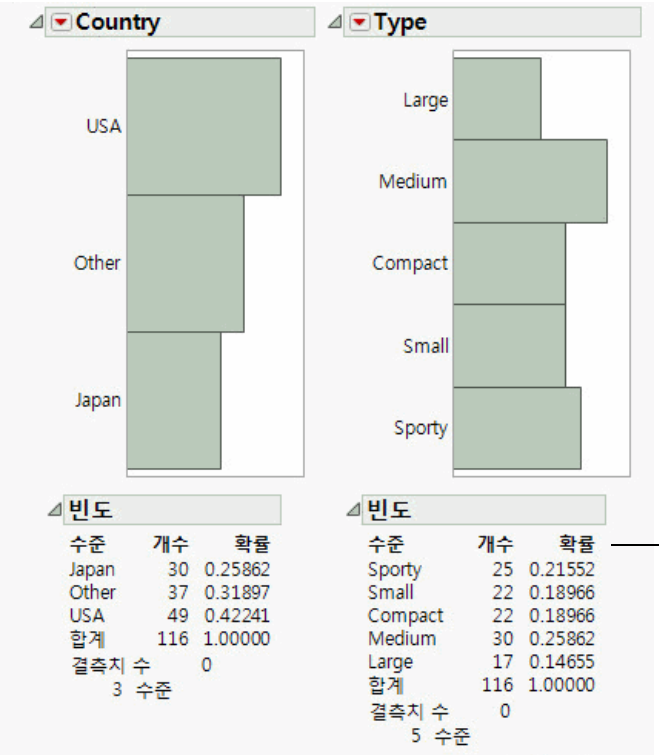
- 자동차의 종류는 무엇입니까?
- 원산지 국가는 어디입니까?

이러한 질문에 답하려면 Type 및 Country 분포를 보십시오.

분포 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Physical Data.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. Country 및 Type 을 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.

그림 5.9 Country 및 Type 분포



분포 결과 해석

보고서 창에는 Country 및 Type 에 대한 막대 차트와 " 빈도 " 보고서가 있습니다 . 막대 차트는 빈도 보고서에서 제공된 빈도 정보를 그래픽으로 나타낸 것입니다 . " 빈도 " 보고서에는 다음 항목이 포함됩니다 .

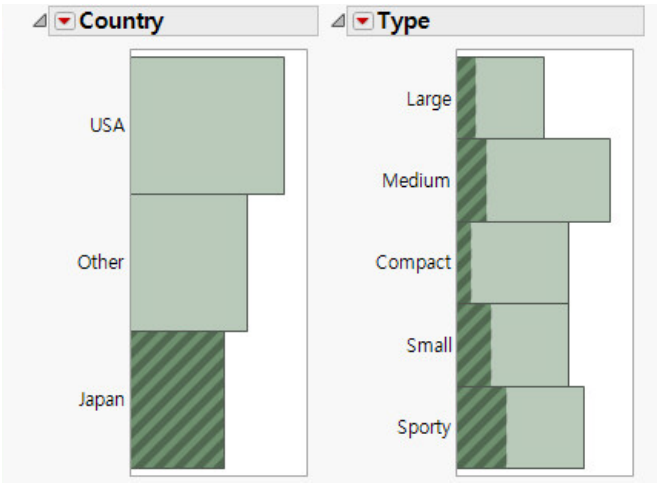
- 데이터의 범주. 예를 들어 "Japan" 은 "Country" 의 범주이고 "Sporty" 는 "Type" 의 범주입니다.
- 각 범주의 총 수
- 각 범주가 차지하는 전체 대비 비율

예를 들어 경차 ("Compact") 는 22 대가 있으며 116 개의 관측값 중 약 19% 를 차지합니다.

분포 결과에서의 상호 작용

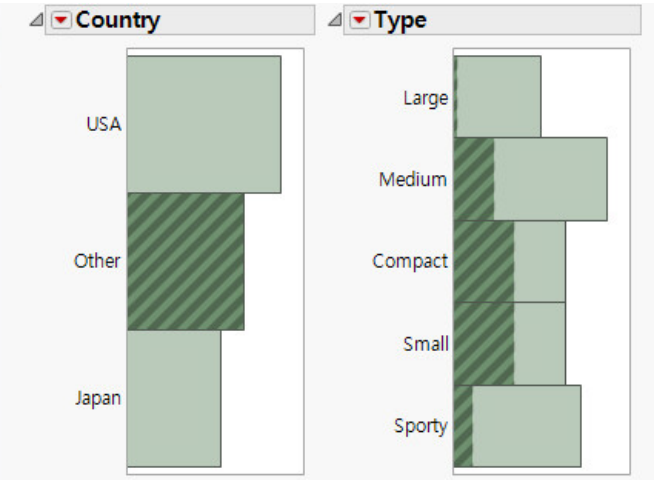
한 차트에서 막대를 선택하면 다른 차트에서 해당하는 데이터가 선택됩니다. 예를 들어 "Country" 막대 차트에서 "Japan" 막대를 선택하면 많은 수의 일본 자동차가 스포츠카인 것을 볼 수 있습니다.

그림 5.10 일본 자동차



"Other" 범주를 선택하면 이러한 자동차의 대다수가 소형 ("Small") 또는 경차 ("Compact") 이고 대형 ("Large") 은 거의 없음을 알 수 있습니다.

그림 5.11 기타 국가의 자동차



관계 분석

산점도 및 기타 그래프는 변수 간의 관계를 시각화하는 데 도움이 됩니다. 관계를 시각화한 후에는 다음 단계로 그 관계를 분석하여 수치로 설명할 수 있습니다. 변수 간의 관계를 수치로 정의한 것을 **모형**이라고 합니다. 더욱 중요한 것은 모형에서 한 변수(*Y*)의 평균 값을 다른 변수(*X*)의 값을 바탕으로 예측할 수도 있다는 것입니다. *X* 변수를 예측 변수라고도 합니다. 일반적으로 이러한 모형을 **회귀** 모형이라고 합니다.

JMP의 **X로 Y 적합** 플랫폼과 **모형 적합** 플랫폼에서는 회귀 모형을 생성합니다.

참고 : 여기서는 기본 플랫폼과 옵션만 다룹니다 . 모든 플랫폼 옵션에 대한 설명은 **기본 분석** , Essential Graphing, 그리고 " 내용 소개 "(129 페이지) 에 나열된 설명서에서 확인하십시오 .

표 5.3에서는 네 가지 기본 유형의 관계를 보여 줍니다.

표 5.3 관계 유형

X	Y	섹션
연속형	연속형	• " 하나의 예측 변수가 있는 회귀 사용 "(143 페이지) • "다중 예측 변수가 있는 회귀 사용"(157페이지)
범주형	연속형	• " 한 변수에 대한 평균 비교 "(148 페이지) • " 여러 변수의 평균 비교 "(153 페이지)

표 5.3 관계 유형 (계속)

X	Y	섹션
범주형	범주형	"비율 비교"(151 페이지)
연속형	범주형	로지스틱 회귀에 대해서는 자세한 설명이 필요합니다. 자세한 내용은 **기본 분석**의 에서 확인하십시오.

하나의 예측 변수가 있는 회귀 사용

시나리오

이 예에서는 제약 및 컴퓨터 업계의 32 개 회사에 대한 재무 데이터가 포함된 Companies.jmp 데이터 테이블을 사용합니다.

직원 수가 많은 회사가 직원 수가 적은 회사보다 많은 매출 수익을 창출한다는 것은 직관적으로 알 수 있습니다. 데이터 분석가는 직원 수에 따라 각 회사의 전체 매출 수익을 예측하려고 합니다.

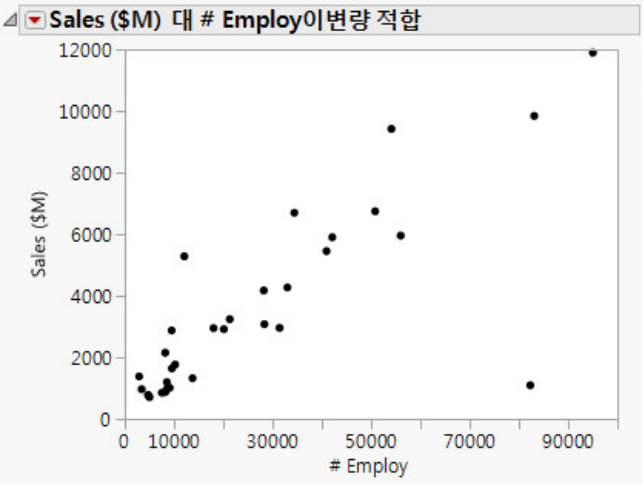
이 작업을 완수하려면 다음을 수행하십시오 .

- " 관계 발견 "(143 페이지)
- " 회귀 모형 적합 "(144 페이지)
- " 평균 매출 예측 "(146 페이지)

관계 발견

먼저, 직원 수와 매출 수익 간의 관계를 확인하기 위해 산점도를 생성합니다. 이 산점도는 " 데이터 시각화" 장의 "산점도 생성"(100 페이지)에서 생성되었습니다. 그림 5.12 의 산점도는 여기에서 이상치 하나(직원 수와 매출이 현저하게 높은 회사)를 숨기고 제외한 후의 결과를 보여 줍니다.

그림 5.12 Sales ($M) 대 # Employ 산점도

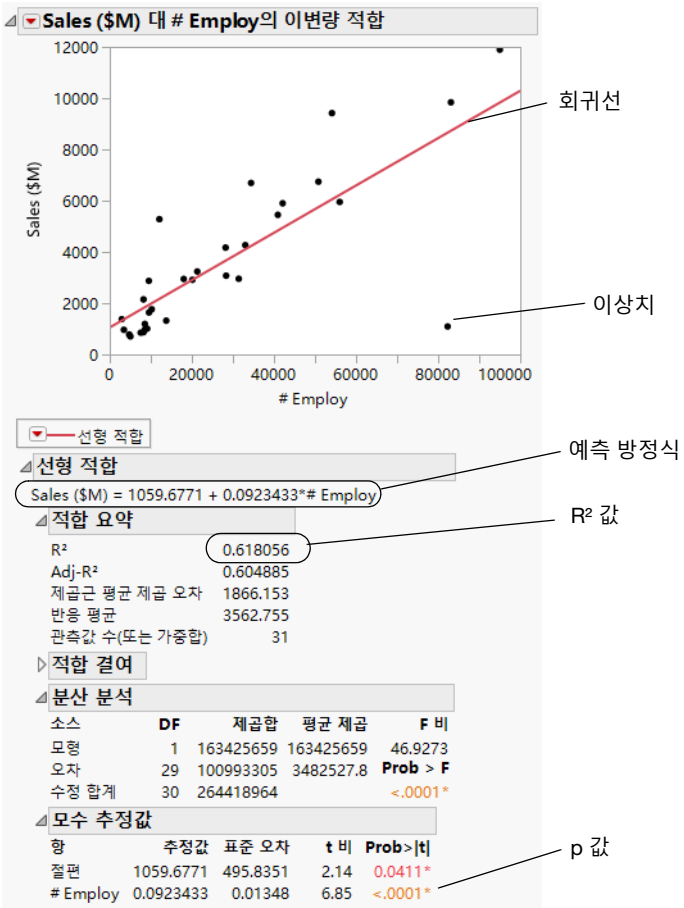


이 산점도는 매출과 직원 수 간의 관계를 명확하게 보여 줍니다. 예상대로, 회사의 직원이 많을수록 매출액이 높아질 수 있습니다. 이 산점도로 데이터 분석가의 추측을 시각적으로 확인할 수 있지만 특정 수의 직원에 대한 매출을 예측할 수는 없습니다.

회귀 모형 적합

직원 수로 매출 수익을 예측하려면 회귀 모형을 적합시켜야 합니다. "이변량 적합"의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다. 산점도에 회귀선이 추가되고 보고서 창에 보고서가 추가됩니다.

그림 5.13 회귀선



보고서에서 다음 결과를 확인하십시오 .

- *p* 값이 .0001 보다 작음
- R^2 값이 0.618 임

이러한 결과를 바탕으로 데이터 분석가는 다음과 같은 결론을 내릴 수 있습니다 .

- "# Employ" 모형 항의 *p* 값이 작습니다 . 이는 유의 수준 0.05 에서 "# Employ" 에 대한 계수가 0 이 아니라는 것을 뒷받침합니다 . 따라서 예측 모형에 직원 수를 포함하면 직원 수가 포함되지 않은 모형에 비해 평균 매출을 예측하는 능력이 크게 향상됩니다 .
- R^2 값 0.618 은 이 모형이 매출 변동의 약 62% 를 설명하고 있음을 나타냅니다 . R^2 값은 결정계수로서 , 모형으로 설명되는 종속 (반응) 변수의 분산 비율을 나타냅니다 . R^2 는 0 에서 1 사이입니다 . R^2 가 0 인 모형은 설명력이 없습니다 . R^2 가 1 인 모형은 반응을 완벽하게 예측합니다 .

평균 매출 예측

회귀 모형을 사용하면 특정 수의 직원이 있을 때 회사에서 기대할 수 있는 평균 매출을 예측할 수 있습니다 . 이 모형에 대한 예측 방정식이 보고서에 포함되어 있습니다 .

평균 매출 = 1059.68 + 0.092* 직원 수

예를 들어 직원 수가 70,000 명인 회사의 매출은 약 7,500 달러로 예측됩니다 .

7,499.68 달러 = 1059.68 + 0.092*70,000

현재 산점도의 오른쪽 아래에는 다른 회사의 일반적인 패턴을 따르지 않는 이상치가 있습니다. 데이터 분석가는 이 이상치가 제외될 때 예측 모형이 바뀌는지에 대해 알고 싶습니다.

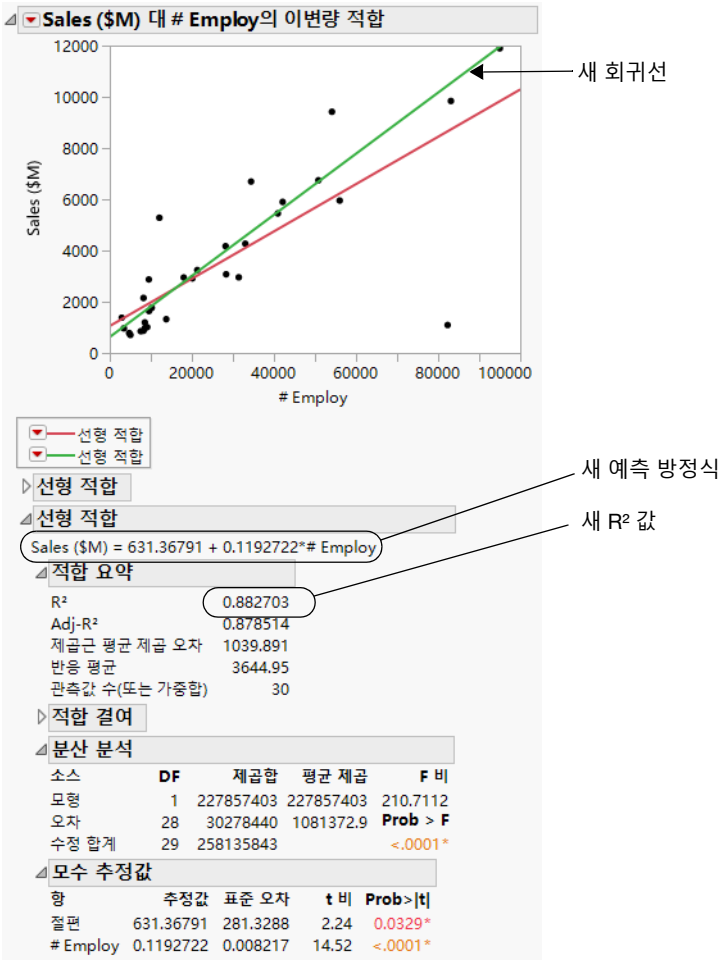
이상치 제외

1. 이상치를 클릭합니다 .
2. **행** > **제외 / 제외 해제**를 선택합니다 .
3. 이 모형을 적합시키려면 "Sales (SM) 대 # Employ 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 .

다음 항목이 보고서 창에 추가됩니다 (그림 5.14).

- 새 회귀선
- 다음을 포함하는 새 선형 적합 보고서
  - 새 예측 방정식
  - 새 R^2 값

그림 5.14 모형 비교



결과 해석

그림 5.14 의 결과를 바탕으로 데이터 분석가는 다음과 같은 결론을 내릴 수 있습니다 .

- 이상치는 큰 회사의 회귀선을 끌어내리고 작은 회사의 회귀선을 끌어올립니다 .
- 이상치가 없는 데이터에 대한 새 모형은 첫 번째 모형보다 강력한 모형입니다 . 새 R^2 값 0.88 은 초기 분석보다 높고 1 에 더 가깝습니다 .

결론

새 예측 방정식을 사용하면 7 만 명의 직원이 있는 회사의 예측 평균 매출을 다음과 같이 계산할 수 있습니다 .

8,961.37 달러 = 631.37 + 0.119*70,000

첫 번째 모형의 예측값은 약 7,500달러입니다. 두 번째 모형에서는 총 매출을 첫 번째 모형에 비해 1,460달러 늘어난 약 8,960달러로 예측합니다.

이상치를 제거한 두 번째 모형은 직원 수를 기준으로 한 매출 총액을 첫 번째 모형보다 더욱 정확하게 설명하고 예측합니다. 데이터 분석가는 이제 사용하기 적절한 모형을 확보했습니다.

한 변수에 대한 평균 비교

연속형 Y 변수와 범주형 X 변수가 있으면 X 변수의 수준에서 평균을 비교할 수 있습니다.

시나리오

이 예에서는 제약 및 컴퓨터 업계의 32개 회사에 대한 재무 데이터가 포함된 Companies.jmp 데이터 테이블을 사용합니다.

재무 분석가가 다음과 같은 질문에 대한 답을 구하려고 합니다.

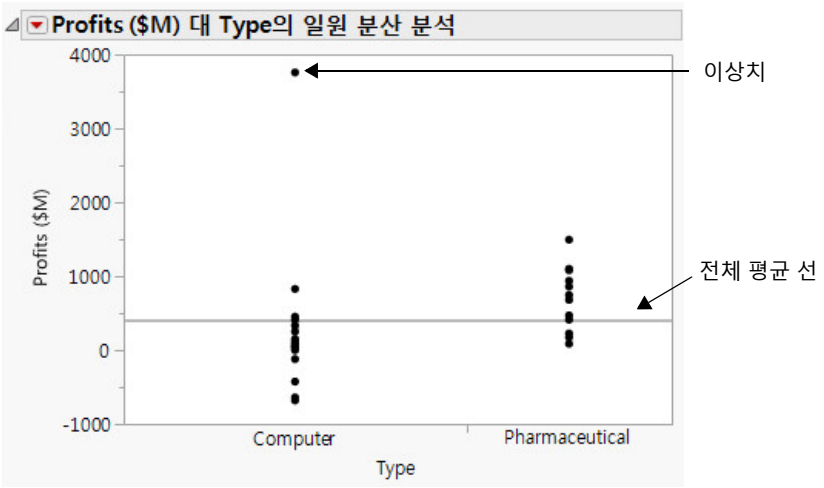
- 컴퓨터 회사의 수익은 제약 회사의 수익과 어떻게 비교됩니까?

이 질문에 답하려면 Type을 기준으로 Profits ($M)를 적합시킵니다.

관계 발견

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다.
2. Companies.jmp 샘플 데이터 테이블을 열어 둔 상태라면 제외되었거나 숨겨진 행이 있을 수 있습니다. 행을 기본 상태로 되돌려 숨겨진 행 없이 모든 행을 포함하려면 **행 > 행 상태 지우기**를 선택합니다.
3. **분석 > X 로 Y 적합**을 선택합니다.
4. Profits ($M) 를 선택하고 **Y, 반응**을 클릭합니다.
5. Type 을 선택하고 **X, 요인**을 클릭합니다.
6. **확인**을 클릭합니다.

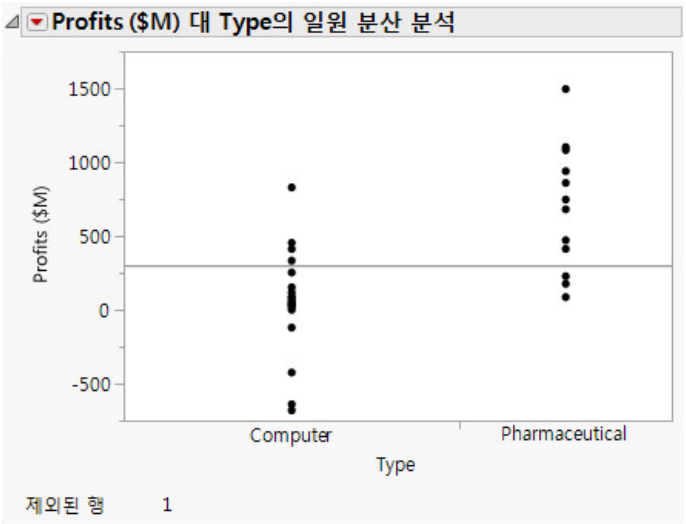
그림 5.15 회사 유형별 수익



컴퓨터 유형에 이상치가 있습니다 . 이 이상치는 산점도의 범위를 늘리고 수익을 비교하기 어렵게 만듭니다 . 이상치를 제외하고 숨깁니다 .

1. 이상치를 클릭합니다 .
2. **행** > **제외 / 제외 해제**를 선택합니다 . 해당 데이터 점이 더 이상 계산에 포함되지 않습니다 .
3. **행** > **숨기기 / 숨기기 해제**를 선택합니다 . 해당 데이터 점이 모든 그래프에서 숨겨집니다 .
4. 이상치를 제외하고 그림을 다시 생성하려면 "Profits ($M) 대 Type 의 일원 분석 " 을 클릭하고 **다시 실행** > **분석 다시 실행**을 선택합니다 . 원래의 산점도 창은 닫아도 됩니다 .

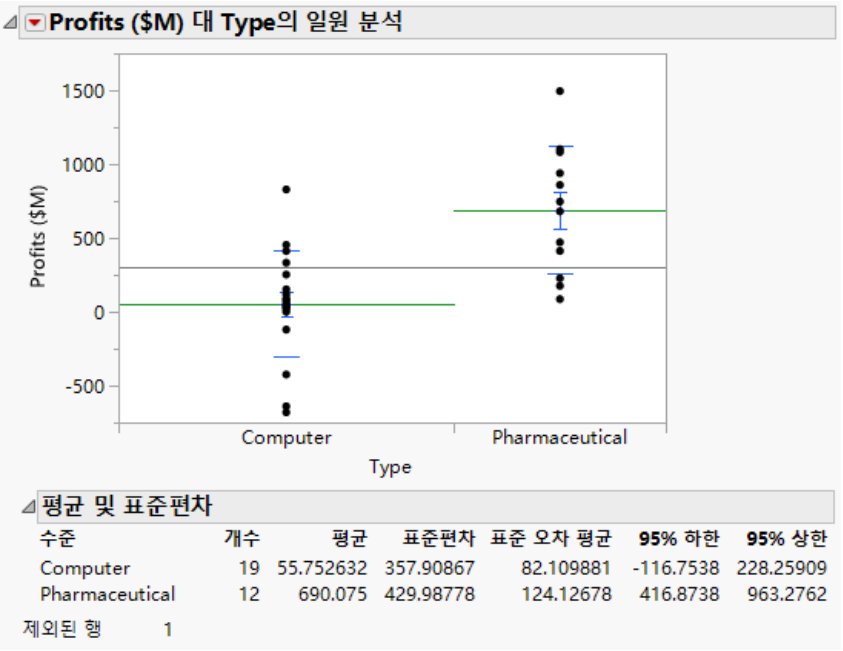
그림 5.16 업데이트된 그림



이상치를 제거하면 재무 분석가가 데이터를 보다 명확하게 파악할 수 있습니다 .

5. 관계 분석을 계속하려면 "Profits ($M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형에서 다음 옵션을 선택합니다 .
   - **표시 옵션 > 평균 선** . 이 옵션을 선택하면 산점도에 평균 선이 추가됩니다 .
   - **평균 및 표준편차** . 이 옵션을 선택하면 평균 및 표준편차를 제공하는 보고서가 표시됩니다 .

그림 5.17 평균 선 및 보고서



결과 해석

재무 분석가는 컴퓨터 회사의 수익이 제약 회사의 수익과 어떻게 비교되는지 알고자 했습니다. 업데이트된 산점도는 제약 회사의 평균 수익이 컴퓨터 회사보다 높다는 것을 보여 줍니다. 이 보고서의 경우 한 쪽 평균 값에서 다른 쪽의 평균 값을 뺀 수익 차이는 약 6 억 3,500 만 달러입니다. 이 산점도는 또한 컴퓨터 회사 중 일부의 수익이 마이너스이고 모든 제약 회사의 수익은 플러스임을 보여 줍니다.

t- 검정 수행

재무 분석가는 회사 표본 (데이터 테이블에 있는 회사) 만을 살펴보았습니다 . 재무 분석가는 이제 다음 질문을 검토하고자 합니다 .

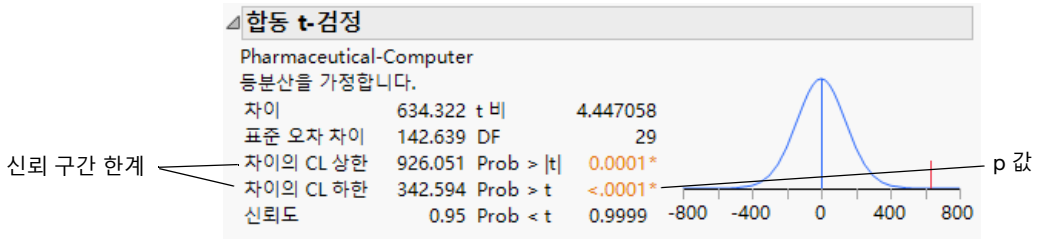
- 더 광범위한 모집단에서 차이가 있습니까 ? 아니면 6 억 3,500 만 달러의 차이는 우연입니까 ?

- 차이가 있다면 얼마입니까 ?

이 질문에 답하려면 2 표본 *t*-검정을 수행합니다. *t*-검정을 통해 표본의 데이터를 사용하여 더 많은 모집단에 대한 추론을 할 수 있습니다.

t-검정을 수행하려면 "일원 분석"의 빨간색 삼각형을 클릭하고 **평균/ANOVA/합동 t**를 선택합니다.

그림 5.18 t- 검정 결과



p 값 0.0001 은 유의 수준 0.05 보다 작으므로 통계적으로 유의함을 나타냅니다. 따라서 재무 분석가는 표본 데이터에 대해 관측된 평균 수익의 차이가 통계적으로 유의하다고 결론 내릴 수 있습니다. 즉, 더 큰 모집단에서 제약 회사의 평균 수익은 컴퓨터 회사의 평균 수익과 다릅니다.

결론

신뢰 구간 한계를 사용하여 두 회사 유형의 수익에 얼마나 많은 차이가 있는지 파악합니다. 그림 5.18 에서 **차이 CL 상한** 및 **차이 CL 하한** 값을 살펴봅니다. 재무 분석가는 제약 회사의 평균 수익이 컴퓨터 회사의 평균 수익보다 높은 3 억 4 천 3 백만 달러에서 9 억 2 천 6 백만 달러 사이라고 결론을 내립니다.

비율 비교

범주형 *X* 변수와 *Y* 변수를 사용할 때는 Y 변수 내 수준의 비율을 X 변수 내의 수준과 비교할 수 있습니다.

시나리오

이 예에서는 계속해서 Companies.jmp 데이터 테이블을 사용합니다. "한 변수에 대한 평균 비교"(148 페이지)에서 재무 분석가는 제약 회사가 컴퓨터 회사보다 평균적으로 더 높은 수익을 얻는 것으로 판단했습니다.

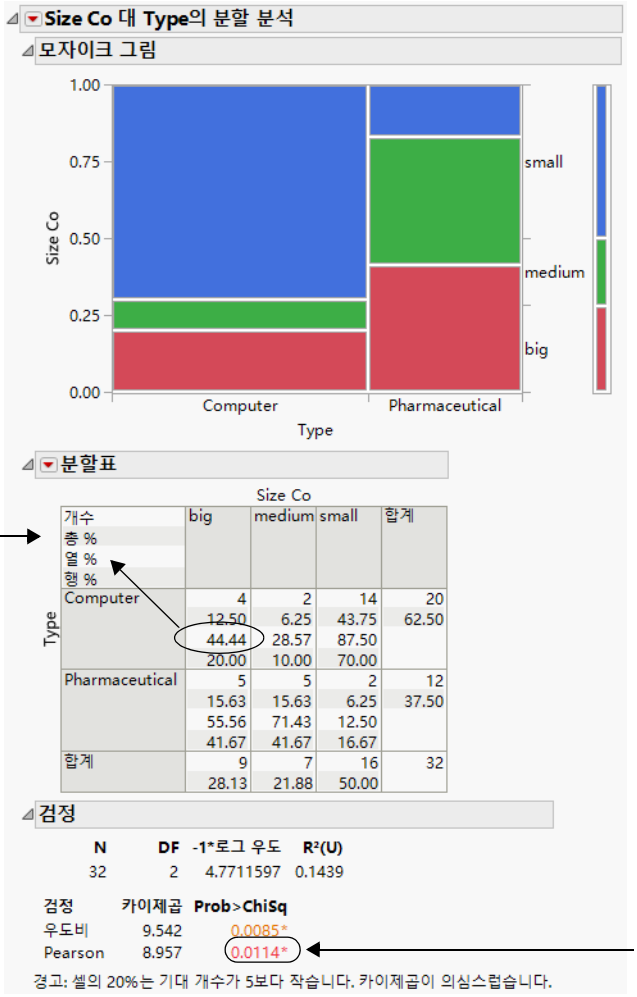
재무 분석가는 회사의 규모가 수익에 미치는 영향이 회사 유형별로 달라지는지 알아보려고 합니다. 그러나 이 질문을 검토하기 전에 재무 분석가는 컴퓨터 및 제약 회사의 모집단에서 소규모, 중간 규모 및 대규모 기업이 동일한 비율로 구성되었는지 알아야 합니다.

관계 발견

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .

2. 이전 예의 Companies.jmp 데이터 파일을 계속 열어 둔 상태라면 제외되거나 숨겨진 행이 있을 수 있습니다. 행을 기본 상태로 되돌려 숨겨진 행 없이 모든 행을 포함하려면 **행** > **행 상태 지우기**를 선택합니다.
3. **분석** > **X 로 Y 적합**을 선택합니다.
4. Size Co 를 선택하고 **Y, 반응**을 클릭합니다.
5. Type 을 선택하고 **X, 요인**을 클릭합니다.
6. **확인**을 클릭합니다.

그림 5.19 회사 유형별 회사 규모



테이블 범례

Pearson 검정

분할표에는 이 예에 적용할 수 없는 정보가 포함되어 있습니다. "분할표"의 빨간색 삼각형을 클릭하고 **총 %** 및 **열 %**를 선택 취소하여 해당 정보를 제거합니다. 그림 5.20에서는 업데이트된 테이블을 보여 줍니다.

그림 5.20 업데이트된 분할표

분할표

		Size Co			
	개수 행 %	big	medium	small	합계
Type	Computer	4 20.00	2 10.00	14 70.00	20
	Pharmaceutical	5 41.67	5 41.67	2 16.67	12
	합계	9	7	16	32

결과 해석

분할표의 통계는 모자이크 그림에 그래픽으로 표시됩니다. 모자이크 그림과 분할표는 두 업계의 소규모, 중간 규모 및 대규모 기업 비율을 비교합니다. 예를 들어 모자이크 그림은 컴퓨터 업계가 제약 업계에 비해 소규모 기업의 비율이 높음을 보여 줍니다. 분할표에서는 정확한 통계량을 보여 줍니다. 즉, 컴퓨터 회사의 70%가 소규모이고 제약 회사의 약 17%가 소규모입니다.

검정 해석

재무 분석가는 회사 표본(데이터 테이블에 있는 회사)만을 살펴보았습니다. 재무 분석가는 모든 컴퓨터 및 제약 회사의 더 광범위한 모집단에서 비율이 다른지 알아야 합니다.

이 질문에 답하려면 **검정** 보고서에서 Pearson 검정의 *p* 값을 사용해야 합니다(그림 5.19(152 페이지)). *p* 값 0.011 이 유의 수준 0.05 보다 작기 때문에 재무 분석가는 다음과 같이 결론을 내립니다.

- 표본 데이터의 차이는 통계적으로 유의합니다.
- 더 광범위한 모집단에서는 비율이 달라집니다.

이제 재무 분석가는 소규모, 중간 규모 및 대규모 기업의 비율이 다르다는 것을 알고 있으며 다음 질문에 답할 수 있습니다.

여러 변수의 평균 비교

"한 변수에 대한 평균 비교"(148 페이지) 섹션에서는 범주형 변수 하나의 여러 수준에서 평균을 비교했습니다. 한 번에 두 개 이상의 변수에 대해 여러 수준에서 평균을 비교하려면 **분산 분석** 기법(또는 ANOVA)을 사용해야 합니다.

시나리오

재무 분석가는 비율 비교 섹션에서 검토하기 시작한 질문에 대답할 수 있습니다. 유형(제약 또는 컴퓨터)에 따라 회사의 규모가 회사의 수익에 미치는 영향이 달라집니까?

이 질문에 대답하려면 다음 두 변수를 기준으로 회사 수익을 비교하십시오 .

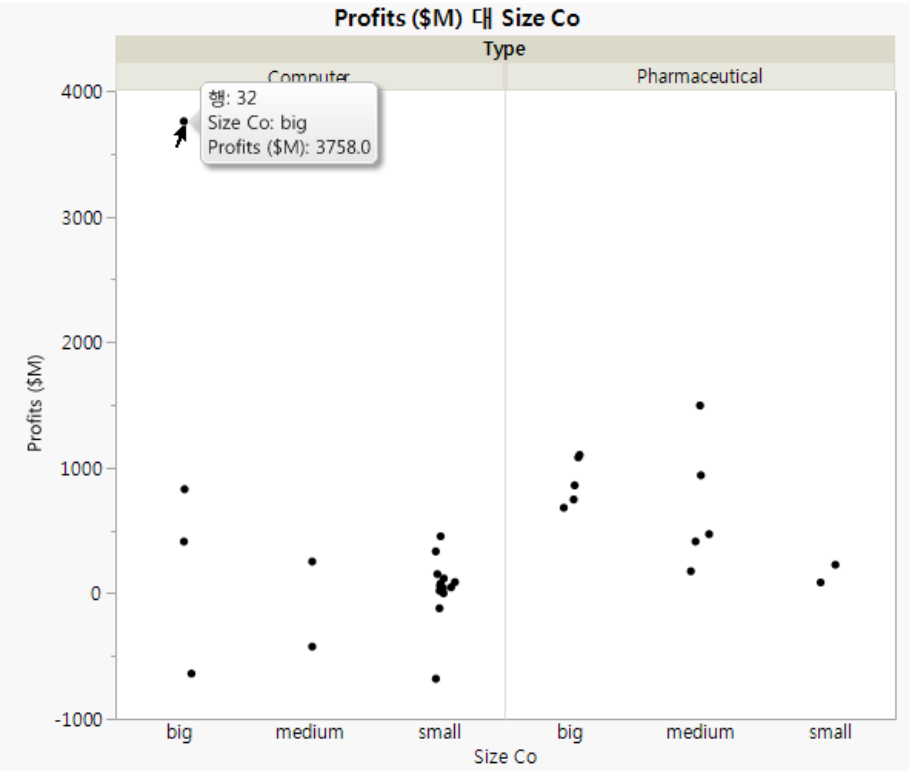
- Type(Pharmaceutical 또는 Computer)
- Size(small, medium, big)

관계 발견

유형과 크기의 모든 조합에 대해 수익 간의 차이를 시각화하려면 그래프를 사용하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **그래프** > **그래프 빌더**를 선택합니다 . " 그래프 빌더 " 창이 나타납니다 .
3. Profits ($M) 를 클릭하고 **Y** 영역으로 드래그하여 놓습니다 .
4. Size Co 를 클릭하고 **X** 영역으로 드래그하여 놓습니다 .
5. Type 을 클릭하고 **그룹 X** 영역으로 드래그하여 놓습니다 .

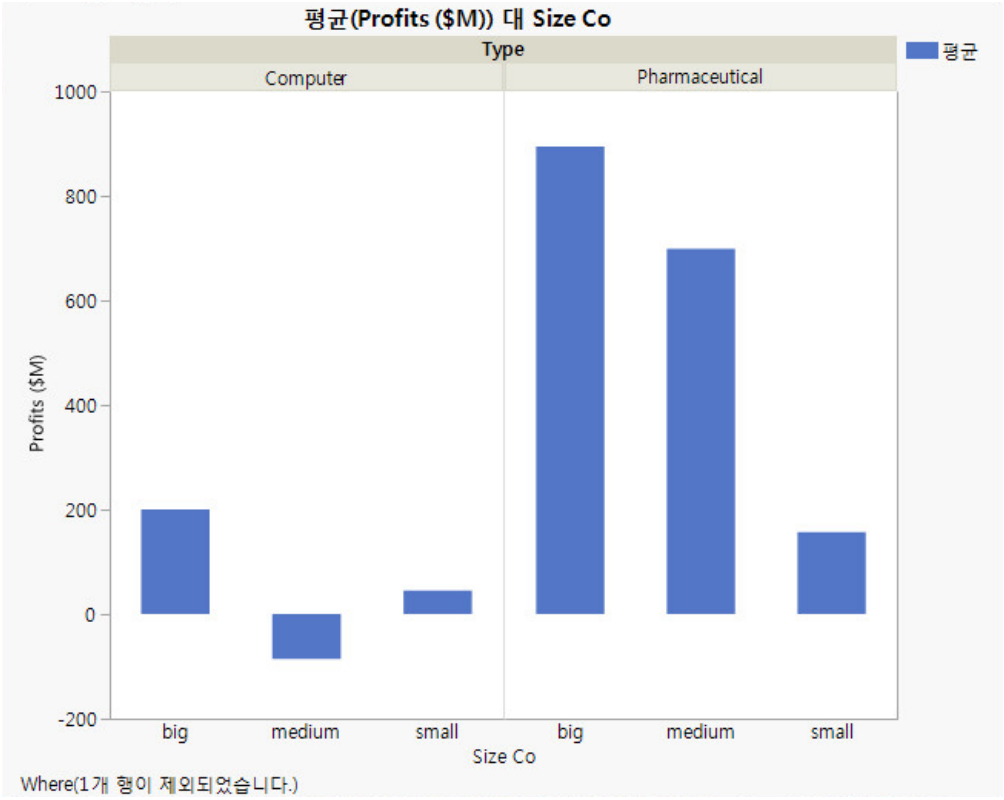
그림 5.21 회사 수익 그래프



그래프에서는 한 대규모 회사의 수익이 매우 크다는 것을 보여 줍니다 . 이 이상치는 그래프의 범위를 늘려 다른 데이터 점을 비교하기 어렵게 만듭니다 .

6. 이 이상치를 선택한 후 마우스 오른쪽 버튼을 클릭하고 **행 > 행 제외**를 선택합니다 . 해당 점이 제거되고 그래프 범위가 자동으로 업데이트됩니다 .
7. 막대 아이콘을 클릭합니다. 막대 차트를 사용하면 점을 사용할 때보다 쉽게 평균 수익을 비교할 수 있습니다 .

그림 5.22 이상치가 제거된 그래프



업데이트된 그래프는 제약 회사의 평균 수익이 더 높다는 것을 보여 줍니다. 이 그래프는 또한 제약 회사만 회사 규모에 따라 수익이 달라진다는 것을 보여 줍니다. 한 변수(회사 규모)의 효과가 다른 변수(회사 유형)의 여러 수준에 대해 변경될 때 이를 **교호작용**이라고 합니다.

관계 수량화

이 데이터는 표본일 뿐이므로 재무 분석가는 다음을 판단해야 합니다 .

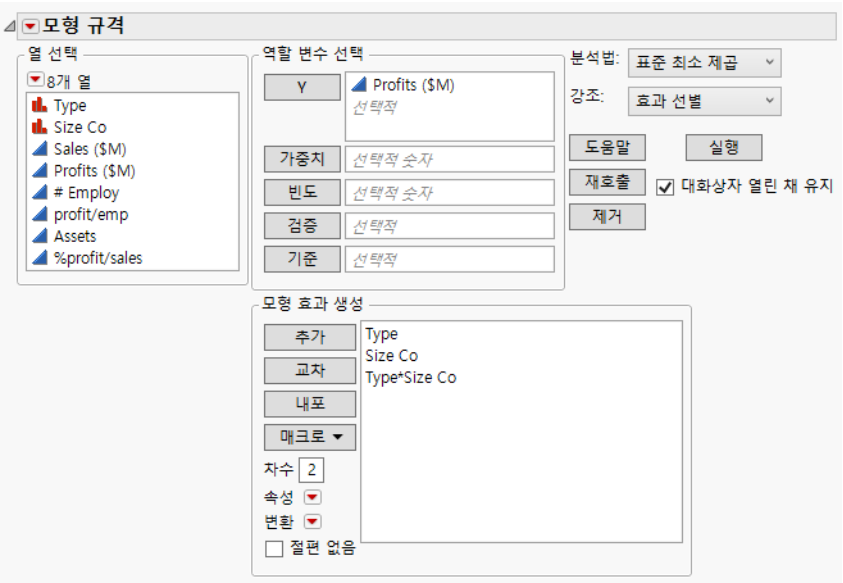
- 차이가 이 표본에 국한되고 우연에 기인하는지 여부

  또는

- 더 광범위한 모집단에 동일한 패턴이 존재하는지 여부

1. 이상치 데이터 점이 제거된 Companies.jmp 샘플 데이터로 돌아갑니다 . 자세한 내용은 " 관계 발견 "(154 페이지) 에서 확인하십시오 .
2. **분석** > **모형 적합**을 선택합니다 .
3. Profits ($M) 를 선택하고 **Y** 를 클릭합니다 .
4. Type 과 Size Co 를 모두 선택합니다 .
5. **매크로** 버튼을 클릭하고 **완전 요인**을 선택합니다 .
6. " 강조 " 메뉴에서 **효과 선별**을 선택합니다 .
7. **대화상자 열린 채 유지** 옵션을 선택합니다 .

그림 5.23 완료된 모형 적합 창



8. **실행**을 클릭합니다 . 보고서 창에 모형 결과가 나타납니다 .

수익의 차이가 실제로 발생했는지 아니면 우연히 발생했는지를 판단하려면 **효과 검정** 보고서를 검토해야 합니다.

참고 : 모든 모형 적합 결과에 대한 자세한 내용은 Fitting Linear Models 의 에서 확인하십시오 .

효과 검정 보기

"효과 검정" 보고서(그림 5.24)는 통계 검정 결과를 보여 줍니다. "모형 적합" 창에는 모형에 포함된 "Type", "Size Co" 및 "Type*Size Co" 효과에 대한 검정이 있습니다.

그림 5.24 효과 검정 보고서

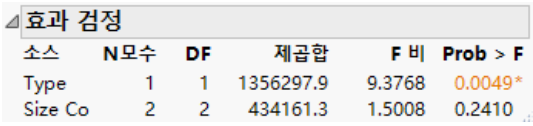
효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F
Type	1	1	1401847.4	10.1368	0.0039*
Size Co	2	2	724616.2	2.6198	0.0927
Type*Size Co	2	2	448061.5	1.6200	0.2180

먼저, 모형에서의 교호작용에 대한 검정인 "Type*Size Co" 효과를 살펴봅니다. 그림 5.22 에서는 제약 회사의 규모에 따라 수익이 다르게 나타난다는 것을 보여 주었습니다. 그러나 이 효과 검정은 수익과 관련하여 유형과 규모 간에 교호작용이 없음을 나타냅니다. *p* 값 0.218 은 매우 큰 값으로, 유의 수준 0.05 보다도 큽니다. 따라서 해당 효과를 모형에서 제거하고 모형을 다시 실행합니다.

1. " 모형 적합 " 창으로 돌아갑니다.
2. " 모형 효과 생성 " 상자에서 **Type*Size Co** 효과를 선택하고 **제거**를 클릭합니다.
3. **실행**을 클릭합니다.

그림 5.25 업데이트된 효과 검정 보고서



효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F
Type	1	1	1356297.9	9.3768	0.0049*
Size Co	2	2	434161.3	1.5008	0.2410

"Size Co" 효과의 *p* 값은 크며, 이는 더 광범위한 모집단에서 크기에 따른 차이가 없음을 나타냅니다. "Type" 효과의 *p* 값은 작아서, 컴퓨터와 제약 회사 간의 데이터에서 확인된 차이가 우연히 발생한 것이 아님을 나타냅니다.

결론

재무 분석가는 유형 (제약 또는 컴퓨터) 에 따라 회사의 규모가 회사의 수익에 미치는 영향에 차이가 있는지 알아보려고 했습니다. 이제 재무 분석가는 질문에 다음과 같이 답할 수 있습니다.

- 더 광범위한 모집단에서 컴퓨터와 제약 회사 간의 수익에는 실제로 차이가 있습니다.
- 회사의 규모와 유형 및 수익 간에는 상관관계가 없습니다.

다중 예측 변수가 있는 회귀 사용

"하나의 예측 변수가 있는 회귀 사용"(143 페이지) 섹션에서는 하나의 예측 변수와 하나의 반응 변수로 구성된 단순 회귀 모형을 생성하는 방법을 소개했습니다. **다중 회귀**는 둘 이상의 예측 변수를 사용하여 평균 반응 변수를 예측합니다.

시나리오

이 예에서는 초코바의 영양 정보가 포함된 Candy Bars.jmp 데이터 테이블을 사용합니다.

영양사는 다음 정보를 사용하여 칼로리를 예측하려고 합니다 .

- 총 지방
- 탄수화물
- 단백질

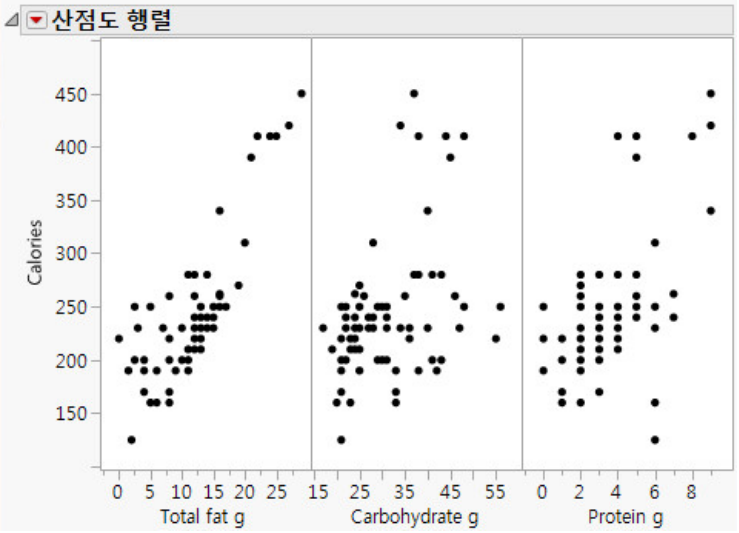
다중 회귀를 사용하여 이 세 가지 예측 변수를 통해 평균 반응 변수를 예측합니다.

관계 발견

칼로리와 총 지방 , 탄수화물 및 단백질 간의 관계를 시각화하려면 산점도 행렬을 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Candy Bars.jmp 를 엽니다 .
2. **그래프** > **산점도 행렬**을 선택합니다 .
3. Calories 를 선택하고 **Y, 열**을 클릭합니다 .
4. Total fat g, Carbohydrate g 및 Protein g 를 선택하고 **X** 를 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 5.26 산점도 행렬 결과



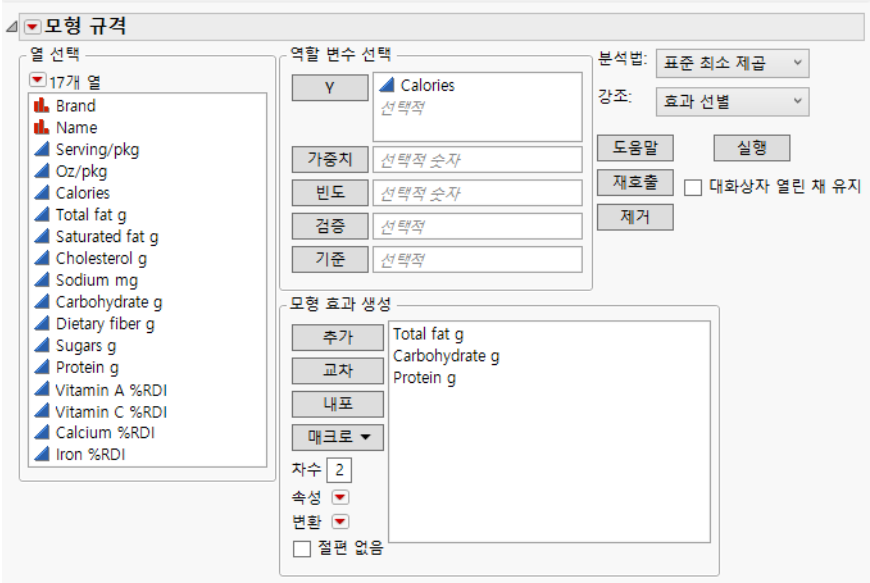
산점도 행렬은 칼로리와 세 변수 간에 양의 상관관계가 있음을 보여 줍니다. 칼로리와 총 지방 간의 상관관계가 가장 강합니다. 이제 영양사는 관계가 있다는 것을 알고 있으므로 평균 칼로리를 예측하기 위해 다중 회귀 모형을 생성할 수 있습니다.

다중 회귀 모형 생성

Candy Bars.jmp 샘플 데이터 테이블을 계속 사용합니다 .

1. **분석** > **모형 적합**을 선택합니다 .
2. Calories 를 선택하고 **Y** 를 클릭합니다 .
3. Total fat g, Carbohydrate g 및 Protein g 를 선택하고 **추가**를 클릭합니다 .
4. " 강조 " 옆에 있는 **효과 선별**을 선택합니다 .

그림 5.27 모형 적합 창



5. **실행**을 클릭합니다 .

보고서 창에 모형 결과가 나타납니다 . 모형 결과를 해석하려면 다음 영역에 중점을 둡니다 .

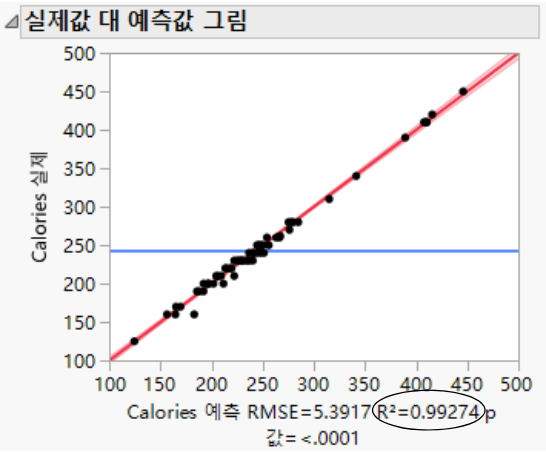
- " 실제값 대 예측값 그림 보기 "(159 페이지)
- " 모수 추정값 해석 "(160 페이지)
- " 예측 프로파일러 사용 "(161 페이지)

참고 : 모든 모형 결과에 대한 자세한 내용은 Fitting Linear Models 의 에서 확인하십시오 .

실제값 대 예측값 그림 보기

실제값 대 예측값 그림은 실제 칼로리와 예측된 칼로리를 보여 줍니다. 예측값이 실제값에 가까워질수록 산점도의 점이 빨간색 선 주위로 모입니다(그림 5.28). 점이 모두 선에 매우 가깝기 때문에 모형이 선택한 요인을 기반으로 칼로리를 정확하게 예측한다는 것을 알 수 있습니다.

그림 5.28 실제값 대 예측값 그림



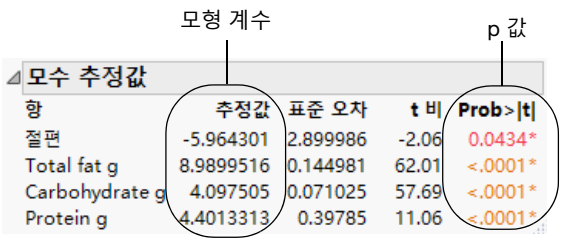
모형 정확도의 또 다른 측도는 R^2 값입니다. 이 값은 그림 5.28의 그림 아래에 나타납니다. R^2 값은 모형에 의해 설명된 대로 칼로리의 변동률을 측정합니다. 1에 가까운 값은 모형이 정확하게 예측한다는 것을 의미합니다. 이 예에서 R^2 값은 0.99입니다.

모수 추정값 해석

" 모수 추정값 " 보고서는 다음과 같은 정보를 보여 줍니다 .

- 모형 계수
- 각 모수의 *p* 값

그림 5.29 모수 추정값 보고서



이 예에서 *p* 값은 모두 매우 작습니다(<.0001). 이것은 칼로리를 예측할 때 세 가지 효과(지방, 탄수화물 및 단백질)가 모두 유의하게 기여함을 나타냅니다.

모형 계수를 사용하여 지방 , 탄수화물 및 단백질의 특정 값에 대한 칼로리 값을 예측할 수 있습니다 . 예를 들어 다음 특성을 가진 초코바의 평균 칼로리를 예측한다고 가정해 보겠습니다 .

- 지방 = 11g
- 탄수화물 = 43g

- 단백질 = 2g

이 값을 사용하여 예측 평균 칼로리를 다음과 같이 계산할 수 있습니다.

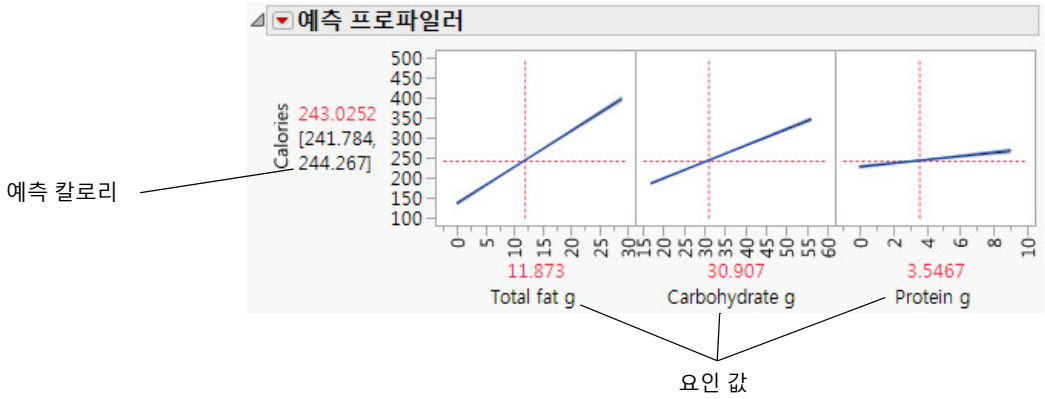
277.92 = -5.9643 + 8.99*11 + 4.0975*43 + 4.4013*2

이 예의 특성은 "Milky Way" 초코바(데이터 테이블의 59 행)와 동일합니다. "Milky Way"의 실제 칼로리는 280이며 모형이 정확하게 예측한다는 것을 보여 줍니다.

예측 프로파일러 사용

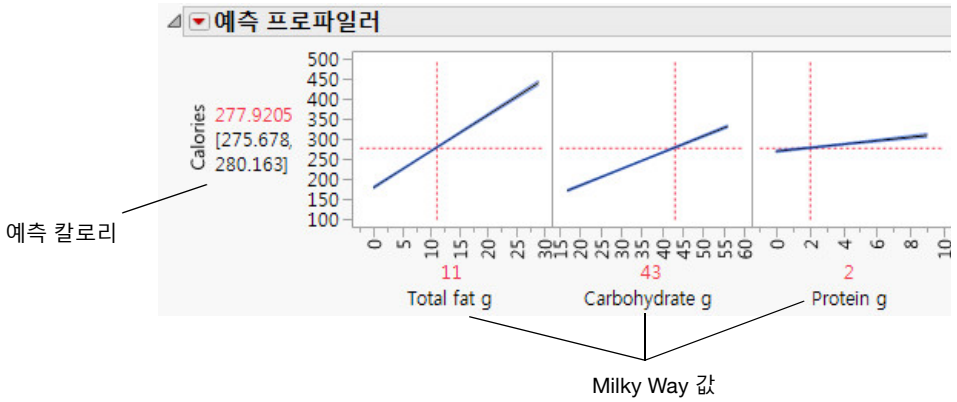
예측 프로파일러를 사용하여 요인의 변화가 예측값에 어떻게 영향을 미치는지 확인할 수 있습니다. 프로파일 선은 요인이 바뀜에 따라 변화되는 칼로리 크기를 보여 줍니다. Total fat g의 선이 가장 가파르며, 이는 총 지방의 변화가 칼로리에 가장 큰 영향을 미친다는 것을 의미합니다.

그림 5.30 예측 프로파일러



각 요인의 수직선을 클릭하고 드래그하여 예측값이 어떻게 변하는지 확인하십시오. 현재 요인 값을 클릭하고 변경할 수도 있습니다. 예를 들어 요인 값을 클릭하고 "Milky Way" 초코바(59 행)의 값을 입력합니다.

그림 5.31 Milky Way 의 요인 값



참고 : 예측 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

결론

영양사는 이제 총 지방, 탄수화물 및 단백질을 기준으로 초코바의 칼로리를 예측할 수 있는 적절한 모형을 확보했습니다.

6 장

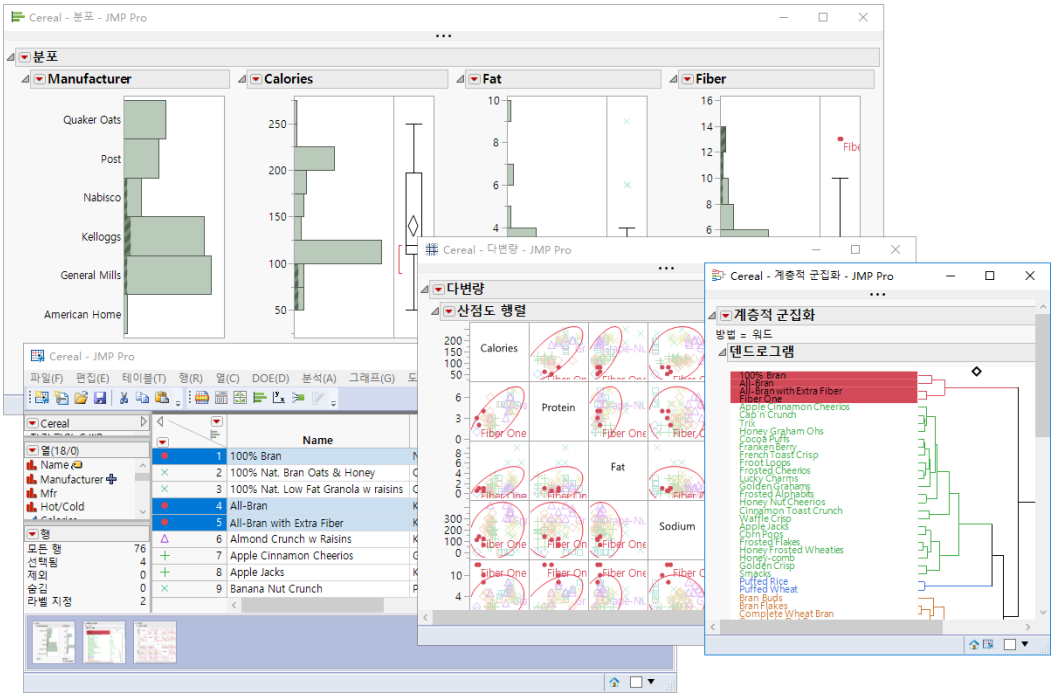
전체를 보는 시각

여러 플랫폼에서 데이터 탐색

JMP는 데이터의 다양한 측면을 탐색하는 데 도움이 되는 통계적 발견 플랫폼의 호스트를 제공합니다. 먼저 히스토그램에서 개별 변수를 간단하게 살펴본 다음 다변량 및 군집 분석으로 진행하여 더욱 자세히 파악할 수 있습니다. 각 단계를 거치면서 데이터에 대해 더 많이 알게 됩니다.

이 장에서는 JMP와 함께 설치되는 샘플 데이터 테이블 Cereal.jmp의 분석 과정을 단계별로 설명합니다. 이를 통해 분포, 다변량 및 계층적 군집화 플랫폼에서 데이터를 탐색하는 방법을 살펴봅니다.

그림 6.1 JMP의 연결된 분석



목차

재미있는 사실 : 연결된 분석 165
여러 플랫폼에서 데이터 탐색 165
분포 플랫폼에서 분포 분석 165
다변량 플랫폼에서 패턴 및 관계 분석 169
군집화 플랫폼에서 유사 값 분석 172

재미있는 사실 : 연결된 분석

JMP의 강력한 기능 중 하나는 연결된 분석입니다. 생성된 그래프와 보고서는 데이터 테이블을 통해 서로 연결됩니다. 그림 6.1에 표시된 것처럼 데이터 테이블에서 선택된 데이터는 세 개의 보고서 창에서도 선택됩니다. 연결된 분석 기능 덕분에 한 창에서 데이터를 선택하면 다른 창에서 해당 데이터의 위치를 확인할 수 있습니다. 이 장의 예를 진행하면서 JMP 창을 열어 두면 이러한 상호 작용을 직접 볼 수 있습니다.

여러 플랫폼에서 데이터 탐색

어떤 시리얼이 건강한 식단에 도움이 될까요? Cereal.jmp 샘플 데이터(인기 시리얼의 상자에서 수집한 실제 데이터)는 섬유질 함량, 칼로리 및 기타 영양 정보에 대한 통계량을 제공합니다. 가장 건강한 시리얼을 파악하려면 히스토그램 및 기술 통계량, 상관관계 및 이상치 탐지, 산점도 및 군집 분석을 단계별로 해석합니다.

분포 플랫폼에서 분포 분석

분포 플랫폼은 히스토그램, 추가 그래프 및 보고서를 사용하여 단일 변수 (**단변량** 분석) 의 분포를 보여 줍니다 . **단변량**이란 관련 변수가 두 개 (이변량) 또는 여러 개 (다변량) 가 아니라 하나뿐임을 의미합니다 . 그러나 단일 보고서에서 여러 개별 변수의 분포를 검토할 수도 있습니다 . 각 변수의 보고서 내용은 변수가 범주형 (명목형 또는 순서형) 인지 연속형인지에 따라 달라집니다 .

- 범주형 변수의 경우 초기 그래프는 히스토그램입니다 . 히스토그램에는 순서형 또는 명목형 변수의 각 수준에 대한 막대가 표시됩니다 . 보고서에는 개수와 비율이 표시됩니다 .
- 연속형 변수의 경우 초기 그래프에는 히스토그램과 이상치 상자 그림이 표시됩니다 . 히스토그램에는 연속형 변수의 그룹화된 값에 대한 막대가 표시됩니다 . 보고서에는 선택한 사분위수와 요약 통계량이 표시됩니다 .

데이터가 어떻게 분포되어 있는지 알게 되면 적절한 유형의 분석을 계획할 수 있습니다.

참고 : 분포 플랫폼에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오 .

시나리오

더 건강한 음식을 먹을 수 있도록 시리얼의 영양가를 확인하려고 합니다 . 시리얼 데이터의 분포를 분석하면 다음 질문에 대한 답을 얻을 수 있습니다 .

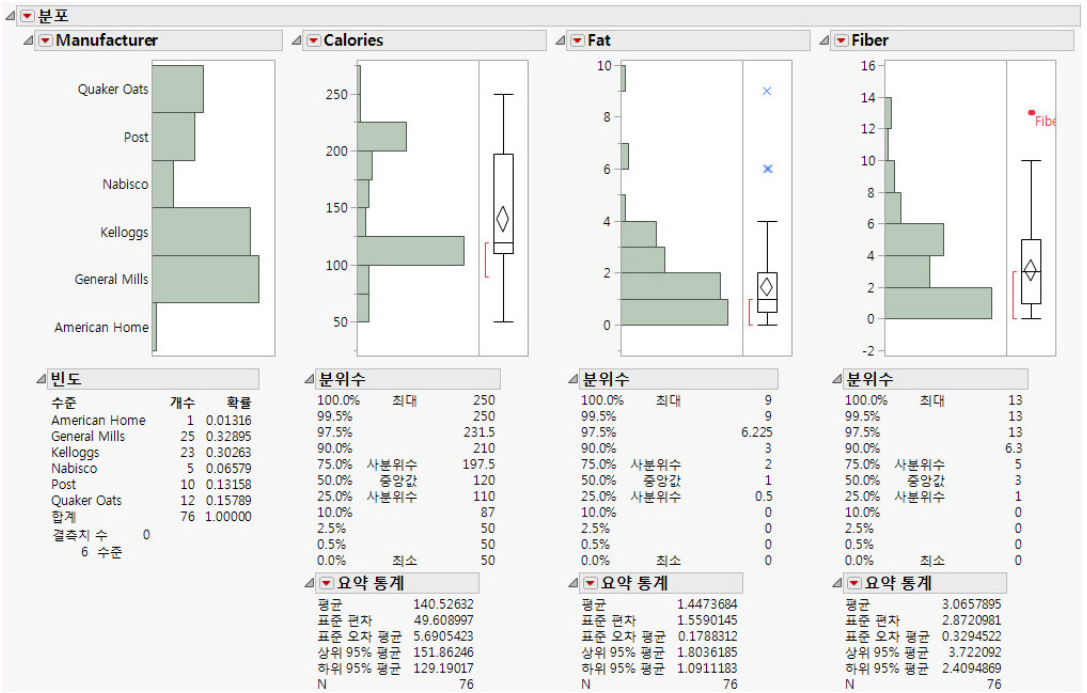
- 어떤 시리얼이 섬유질이 가장 많습니까 ?
- 평균 , 최소 및 최대 칼로리는 얼마입니까 ?
- 지방 함량의 중앙값은 얼마입니까 ?
- 어떤 시리얼이 가장 많은 지방을 함유하고 있습니까 ?

- 데이터에 이상치가 있습니까 ?

분포 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cereal.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Ctrl 키를 누른 채로 Manufacturer, Calories, Fat 및 Fiber 를 클릭합니다 .
4. **Y, 열**을 클릭한 후 **확인**을 클릭합니다 .

그림 6.2 Manufacturer, Calories, Fat 및 Fiber 분포



"Fiber" 분포에서는 다음을 확인할 수 있습니다 .

– "Fiber" 상자 그림에 표시된 것처럼 "Fiber One" 과 "All-Bran with Extra Fiber" 가 가장 많은 섬유질을 함유하고 있습니다 . 이들 시리얼은 섬유질 함량에 있어서 이상치입니다 .

Cereal.jmp 에서 "Fiber One" 이 포함된 행에는 라벨이 있습니다 . 이 라벨은 그래프의 데이터 점 옆에 시리얼 이름을 표시합니다 . 전체 라벨을 표시하려면 맨 오른쪽 세로 경계선을 오른쪽으로 드래그하십시오 . 라벨이 없는 데이터 점을 커서로 가리키면 "All Bran with Extra Fiber" 가 표시됩니다 .

"Fat" 분포에서는 다음을 확인할 수 있습니다 .

– "Fat" 상자 그림에서 맨 위 데이터 점 (x 표식) 을 커서로 가리키면 "100% Nat. Bran Oats & Honey" 가 지방 함량이 가장 높음을 알 수 있습니다 .

– "Fat" 의 " 분위수 " 보고서에서 지방 함량의 중앙값은 1g 입니다 .

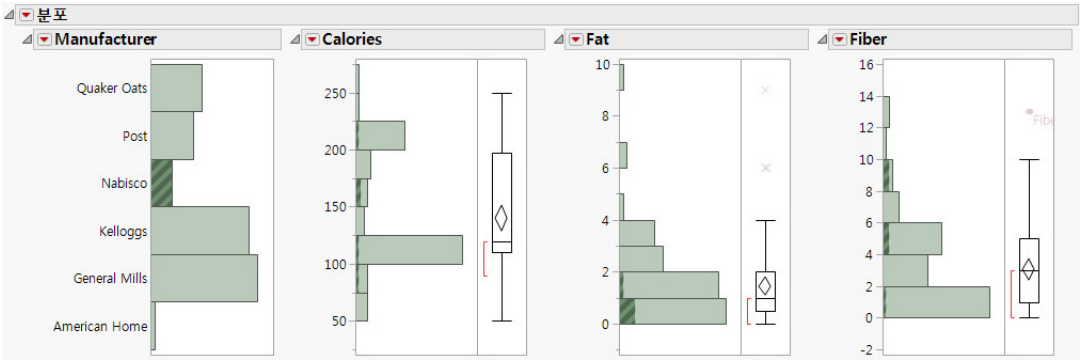
"Calories" 의 " 분위수 " 보고서에서는 다음을 확인할 수 있습니다 .

– 최대 칼로리는 250 입니다 .

– 최소 칼로리는 50 입니다 .

5. "Manufacturer" 히스토그램에서 "Nabisco" 막대를 클릭합니다 .

그림 6.3 Nabisco 시리얼의 분포

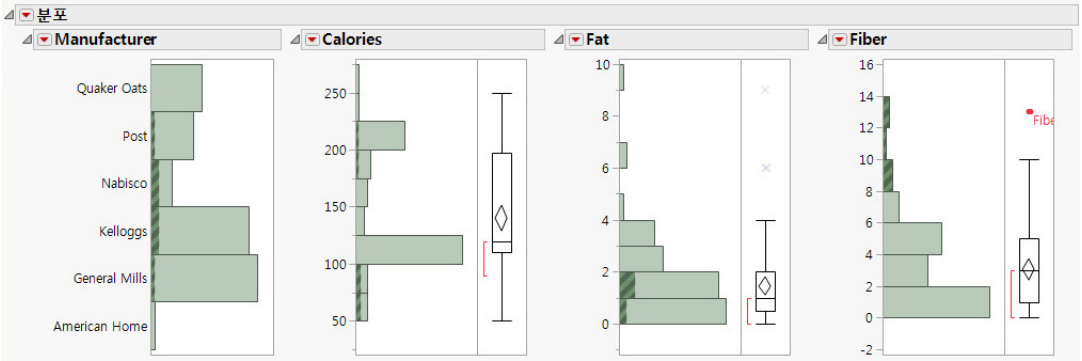


"Nabisco" 시리얼의 칼로리 ("Calories"), 지방 ("Fat") 및 섬유질 ("Fiber") 분포가 다른 히스토그램에서 강조 표시됩니다 . "Nabisco" 시리얼의 칼로리 , 지방 및 섬유질 분포를 전체 데이터의 칼로리 , 지방 및 섬유질 분포와 비교해서 볼 수 있습니다 . 예를 들어 "Nabisco" 시리얼의 지방 분포는 전체 데이터의 지방 분포보다 낮습니다 .

6. 마지막 "Fiber" 막대 아래를 클릭하여 모든 막대를 선택 취소합니다 .

7. Shift 키를 누른 상태로 "Fiber" 히스토그램에서 값이 8 보다 큰 모든 히스토그램 막대를 클릭합니다 .

그림 6.4 섬유질이 많은 시리얼

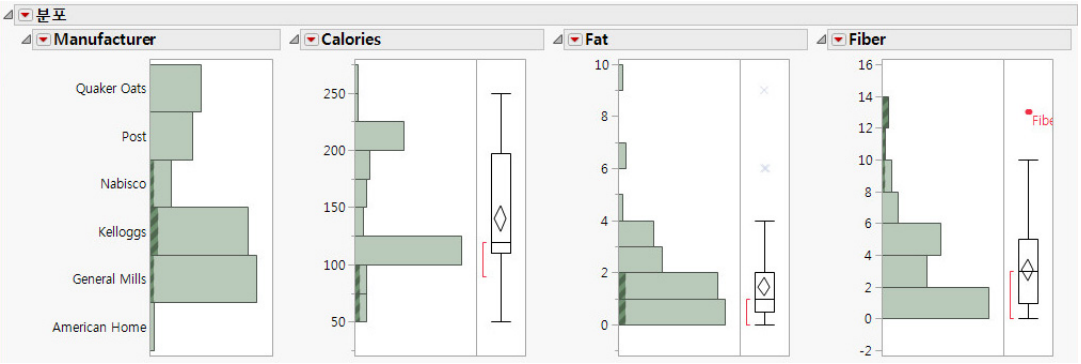


섬유질이 가장 많은 시리얼이 "Calories" 및 "Fat" 히스토그램에서 강조 표시됩니다 . 히스토그램이 연결되어 있기 때문에 섬유질이 많은 시리얼 중 일부는 지방도 적다는 것을 알 수 있습니다 .

8. Ctrl 키와 Shift 키를 누른 상태로 "Calories" 히스토그램에서 값이 200 또는 그 근처에 있는 두 개의 막대를 선택 취소합니다 .

   칼로리가 높은 시리얼이 히스토그램에서 제거됩니다 .

그림 6.5 고섬유질 저지방 시리얼



팁 : " 분포 " 보고서를 열어 두십시오 . 나중에 군집 분석에서 사용할 것입니다 . 자세한 내용은 " 군집화 플랫폼에서 유사 값 분석 "(172 페이지) 에서 확인하십시오 .

결과 해석

결과를 살펴보면 다음 질문에 대한 답을 얻을 수 있습니다 .

어떤 시리얼이 섬유질이 가장 많습니까 ? "Fiber" 상자 그림에서 "All-Bran with Extra Fiber" 와 "Fiber One" 이 가장 많은 섬유질을 함유하고 있음을 알 수 있습니다 . 이 두 시리얼은 이상치입니다 .

평균 , 최소 및 최대 칼로리는 얼마입니까 ? "Calories" 히스토그램에서 칼로리가 50 에서 275 사이임을 알 수 있습니다 . "Calories" 의 " 분위수 " 에서는 칼로리가 50 에서 250 사이이며 , 칼로리의 중앙값이 120 임을 알 수 있습니다 . 이 분포는 균등하지 않습니다 .

지방 함량의 중앙값은 얼마입니까 ? "Fat" 의 " 분위수 " 보고서에서 지방 함량의 중앙값이 1g 임을 알 수 있습니다 .

어떤 시리얼이 가장 많은 지방을 함유하고 있습니까 ? "Fat" 상자 그림에서 "100% Nat. Bran Oats & Honey" 가 지방 함량이 가장 높음을 알 수 있습니다 . 이 시리얼은 이상치입니다 .

결론

섬유질 섭취량을 늘리기 위해 "All-Bran with Extra Fiber" 와 "Fiber One" 을 먹어 보기로 결정했습니다 . 이들 시리얼은 칼로리와 지방 함량이 낮습니다 . 대부분의 시리얼이 지방 섭취량을 크게

늘리지 않지만, 지방 함량이 높은 "100% Nat. Bran Oats & Honey"는 피하기로 했습니다. 또한 대부분의 시리얼이 지방 함량이 비교적 적지만 칼로리가 반드시 낮은 것은 아닙니다.

다변량 플랫폼에서 패턴 및 관계 분석

이제는 먹어야 할 시리얼과 피해야 할 시리얼을 확인했으므로 시리얼 변수 간에 어떤 관련이 있는지 파악해 보려고 합니다. 다변량 플랫폼을 사용하면 변수 간의 패턴과 관계를 관찰할 수 있습니다. " 다변량 " 보고서를 통해 다음이 가능합니다.

- " 상관 " 테이블을 사용하여 각 반응 변수 쌍 간의 선형 관계 강도를 요약할 수 있습니다.
- " 산점도 행렬 " 을 사용하여 종속성, 이상치 및 군집을 식별할 수 있습니다.
- 다른 기법을 사용하여 부분 상관, 역상관, 쌍별 상관, 공분산 행렬 및 주성분 같은 여러 변수를 검토할 수 있습니다.

참고 : 다변량 플랫폼에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오.

시나리오

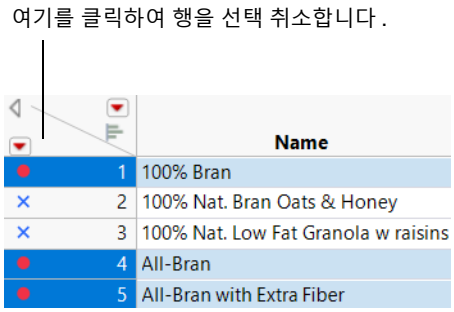
지방과 칼로리 같은 변수 사이의 관계를 확인하려고 합니다. 다변량 플랫폼에서 시리얼 데이터를 분석하면 다음 질문에 대한 답을 얻을 수 있습니다.

- 어떤 변수 쌍이 상관관계가 높습니까 ?
- 상관관계가 없는 변수 쌍은 무엇입니까 ?

다변량 보고서 생성

1. Cereal.jmp 데이터 테이블에서 " 열 " 패널 상단에 있는 아래쪽 삼각형을 클릭하여 행을 선택 취소합니다.

그림 6.6 행 선택 취소



2. **분석 > 다변량 방법 > 다변량**을 선택합니다.
3. Calories 부터 Potassium 까지 선택하고 **Y, 열**을 클릭한 후 **확인**을 클릭합니다.

" 다변량 " 보고서가 나타납니다 . 이 보고서에는 기본적으로 " 상관 " 보고서와 " 산점도 행렬 " 이 포함됩니다 . " 상관 " 보고서는 각 반응 변수 (Y) 쌍 간 선형 관계의 강도가 요약된 상관계수의 행렬입니다 . 숫자 색상이 어두울수록 상관관계가 낮음을 나타냅니다 .

그림 6.7 상관 보고서

상관관계

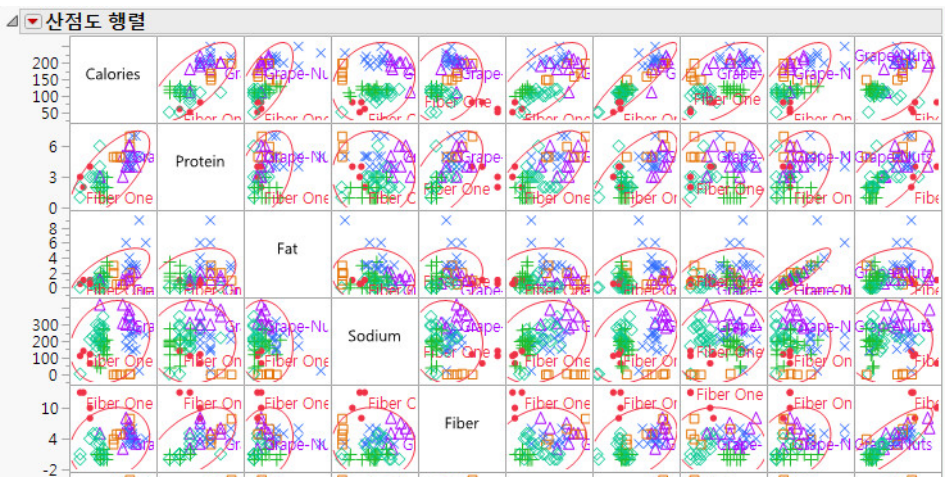
	Calories	Protein	Fat	Sodium	Fiber	Complex Carbos	Tot Carbo	Sugars	Calories fr Fat	Potassium
Calories	1.0000	0.7041	0.6460	0.1996	0.1953	0.6688	0.9076	0.5060	0.6709	0.4451
Protein	0.7041	1.0000	0.4080	0.0050	0.5470	0.6486	0.6937	-0.0010	0.4288	0.6782
Fat	0.6460	0.4080	1.0000	-0.0768	0.1824	0.1037	0.3860	0.4148	0.9013	0.3420
Sodium	0.1996	0.0050	-0.0768	1.0000	-0.0448	0.2619	0.3066	0.1767	0.0572	0.0459
Fiber	0.1953	0.5470	0.1824	-0.0448	1.0000	0.1769	0.3668	-0.1264	0.2553	0.8326
Complex Carbos	0.6688	0.6486	0.1037	0.2619	0.1769	1.0000	0.7773	-0.1601	0.1558	0.2693
Tot Carbo	0.9076	0.6937	0.3860	0.3066	0.3668	0.7773	1.0000	0.4263	0.4636	0.5375
Sugars	0.5060	-0.0010	0.4148	0.1767	-0.1264	-0.1601	0.4263	1.0000	0.4369	0.1166
Calories fr Fat	0.6709	0.4288	0.9013	0.0572	0.2553	0.1558	0.4636	0.4369	1.0000	0.3694
Potassium	0.4451	0.6782	0.3420	0.0459	0.8326	0.2693	0.5375	0.1166	0.3694	1.0000

다음 사항에 유의하십시오 .

- "Calories" 열에서 칼로리는 나트륨 및 섬유질을 제외한 모든 변수와 높은 상관관계가 있습니다 .
- "Fiber" 열에서 섬유질과 칼륨은 높은 상관관계가 있는 것으로 나타납니다 .
- "Sodium" 열에서 나트륨은 다른 변수와 높은 상관관계가 없습니다 .

" 산점도 행렬 " 의 밀도 타원은 변수 간의 관계를 더 자세히 나타냅니다 .

그림 6.8 산점도 행렬의 일부



기본적으로 각 산점도에는 95% 의 이변량 정규 밀도 타원이 있습니다 . 각 변수 쌍이 이변량 정규분포를 따른다고 가정하면 이 타원은 전체 점의 약 95% 를 둘러쌉니다 . 타원이 상당히 둥근 형태이고 대각선 방향이 아니면 해당 변수 간에 상관관계가 없는 것입니다 . 타원이 좁고 대각선 방향이면 해당 변수 간에 상관관계가 높은 것입니다 .

다음 사항을 알 수 있습니다 .

– "Sodium" 행에서는 타원이 상당히 둥근 형태입니다 . 이 모양은 나트륨이 다른 변수와 상관관계가 없음을 나타냅니다 .

– "Fat" 행에서는 "Nat. Bran Oats & Honey", "Cracklin' Oat Bran" 및 "Banana Nut Crunch" 를 나타내는 파란색 x 표식이 타원 바깥쪽에 나타납니다 . 이러한 배치는 해당 데이터가 시리얼의 지방 함량 때문에 이상치임을 나타냅니다 .

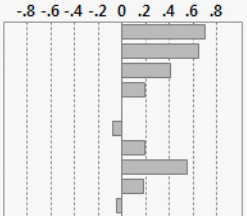
나중에 산점도 행렬을 자세히 살펴볼 것입니다 .

4. " 다변량 " 의 빨간색 삼각형을 클릭하고 **쌍별 상관**을 선택하여 " 쌍별 상관 " 보고서를 표시합니다 .

그림 6.9 쌍별 상관 보고서의 일부

쌍별 상관관계

변수	기준 변수	상관관계	개수	하위 95%	상위 95%	유의성 확률
Protein	Calories	0.7041	76	0.5690	0.8022	<.0001*
Fat	Calories	0.6460	76	0.4922	0.7607	<.0001*
Fat	Protein	0.4080	76	0.2011	0.5801	0.0003*
Sodium	Calories	0.1996	76	-0.0270	0.4068	0.0838
Sodium	Protein	0.0050	76	-0.2207	0.2302	0.9655
Sodium	Fat	-0.0768	76	-0.2971	0.1513	0.5099
Fiber	Calories	0.1953	76	-0.0315	0.4030	0.0908
Fiber	Protein	0.5470	76	0.3668	0.6877	<.0001*
Fiber	Fat	0.1824	76	-0.0449	0.3918	0.1147
Fiber	Sodium	-0.0448	76	-0.2675	0.1825	0.7010



" 쌍별 상관 " 보고서에는 각 Y 변수 쌍에 대한 Pearson 곱적률 상관 계수가 나열됩니다 . 또한 이 보고서는 유의 확률을 보여 주고 막대 차트로 상관관계를 비교합니다 .

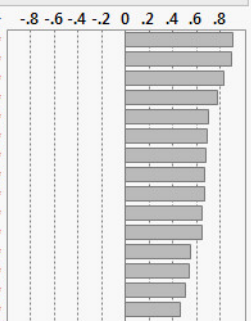
5. 상관관계가 높은 쌍을 빠르게 보려면 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열별 정렬** , **유의 확률** , **오름차순** 체크박스를 선택한 후 **확인**을 클릭합니다 .

가장 관련성이 높은 쌍이 보고서의 맨 위에 나타납니다 . 쌍의 *p* 값이 작은 것은 상관관계의 증거를 나타냅니다 . 가장 유의한 상관관계는 "Tot Carbo"(총 탄수화물) 와 "Calories" 사이의 상관관계입니다 .

그림 6.10 *p* 값이 작은 변수 쌍

쌍별 상관관계

변수	기준 변수	상관관계	개수	하위 95%	상위 95%	유의성 확률
Tot Carbo	Calories	0.9076	76	0.8576	0.9406	<.0001*
Calories fr Fat	Fat	0.9013	76	0.8483	0.9365	<.0001*
Potassium	Fiber	0.8326	76	0.7474	0.8908	<.0001*
Tot Carbo	Complex Carbos	0.7773	76	0.6691	0.8532	<.0001*
Protein	Calories	0.7041	76	0.5690	0.8022	<.0001*
Tot Carbo	Protein	0.6937	76	0.5551	0.7949	<.0001*
Potassium	Protein	0.6782	76	0.5344	0.7838	<.0001*
Calories fr Fat	Calories	0.6709	76	0.5248	0.7786	<.0001*
Complex Carbos	Calories	0.6688	76	0.5221	0.7771	<.0001*
Complex Carbos	Protein	0.6486	76	0.4956	0.7625	<.0001*
Fat	Calories	0.6460	76	0.4922	0.7607	<.0001*
Fiber	Protein	0.5470	76	0.3668	0.6877	<.0001*
Potassium	Tot Carbo	0.5375	76	0.3551	0.6805	<.0001*
Sugars	Calories	0.5060	76	0.3167	0.6566	<.0001*
Calories fr Fat	Tot Carbo	0.4636	76	0.2659	0.6238	<.0001*



결과 해석

결과를 살펴보면 다음 질문에 대한 답을 얻을 수 있습니다 .

어떤 변수 쌍이 상관관계가 높습니까 ? " 상관 " 보고서와 " 산점도 행렬 " 은 "Calories" 가 "Sodium" 및 "Fiber" 를 제외한 모든 변수와 높은 상관관계에 있음을 보여 줍니다 . " 쌍별 상관 " 보고서는 "Tot Carbo"(총 탄수화물) 와 "Calories" 가 가장 상관관계가 높은 변수 쌍이라는 것을 보여 줍니다 .

상관관계가 없는 변수 쌍은 무엇입니까 ? " 상관 " 보고서와 " 산점도 행렬 " 은 "Sodium" 이 다른 변수와 상관관계가 없음을 보여 줍니다 .

결론

지방 함량이 높은 "100% Nat. Bran Oats & Honey" 를 피해야 한다는 이전 결정에 더 확신을 갖게 되었습니다 . "All-Bran with Extra Fiber" 와 "Fiber One" 을 먹어 본다는 것도 현명한 결정이었습니다 . 이 두 가지 고섬유질 시리얼에는 칼로리 , 지방 및 당분 섭취량을 줄이고 칼륨 섭취량을 높여 주는 추가적인 이점이 있습니다 . 또한 칼로리가 높은 고탄수화물 시리얼도 피하기로 결정했습니다 .

군집화 플랫폼에서 유사 값 분석

군집화는 여러 변수 중에서 유사한 값을 공유하는 관측값들을 함께 그룹화하는 다변량 기법입니다 . 계층적 군집화는 행을 계층적 순서로 결합하여 트리로 표현합니다 . 고섬유질과 같은 특정 특성을 지닌 시리얼을 군집으로 그룹화하여 시리얼 간의 유사점을 확인할 수 있습니다 .

참고 : 계층적 군집화에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

시나리오

어떤 시리얼들이 서로 유사하고 어떤 시리얼들이 서로 유사하지 않은지 알아보려고 합니다 . 시리얼 데이터의 군집을 분석하면 다음 질문에 대한 답을 얻을 수 있습니다 .

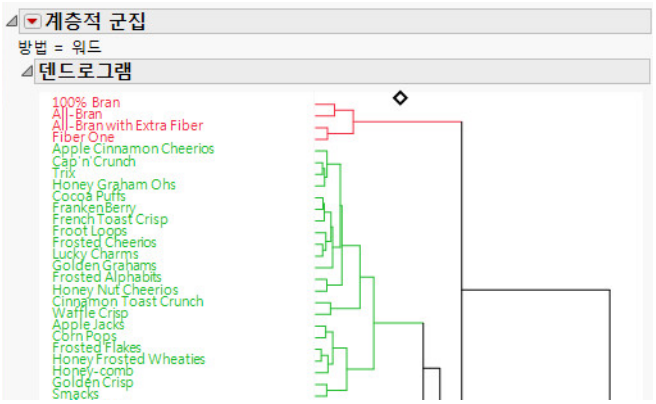
- 어떤 시리얼 군집의 영양가가 적습니까 ?
- 어떤 시리얼 군집이 비타민과 무기질이 풍부하고 당분과 지방은 적습니까 ?
- 어떤 시리얼 군집이 섬유질이 풍부하고 칼로리가 낮습니까 ?

계층적 군집화 그래프 생성

1. Cereal.jmp 가 표시된 상태에서 **분석 > 군집화 > 계층적 군집화**를 선택합니다 .
2. Calories 부터 Enriched 까지 선택하고 **Y, 열**을 클릭한 후 **확인**을 클릭합니다 .

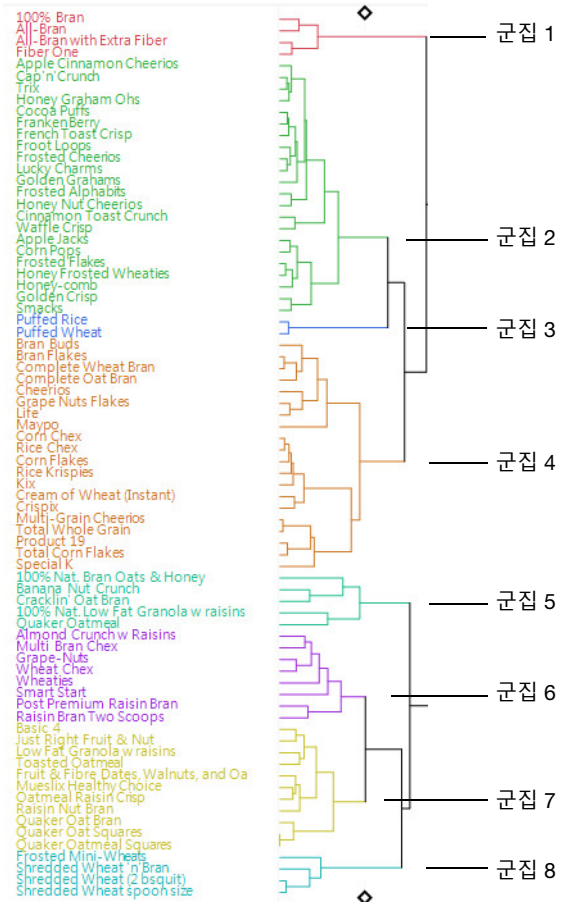
   " 계층적 군집화 " 보고서가 나타납니다 . 데이터 테이블 행 상태에 따라 군집에 색상이 적용됩니다 .

그림 6.11 계층적 군집화 보고서의 일부



3. " 계층적 군집화 " 의 빨간색 삼각형을 클릭하고 **군집 색 표시**를 선택합니다 . 덴드로그램의 관계에 따라 군집에 색상이 적용됩니다 .

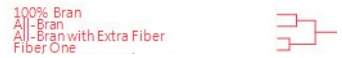
그림 6.12 색상이 적용된 군집



각 군집 내의 시리얼은 유사한 특성을 갖습니다 . 예를 들어 군집 1 에 있는 시리얼은 시리얼 이름으로 판단할 때 섬유질이 많다고 추측할 수 있습니다 .

"All-Bran with Extra Fiber" 와 "Fiber One" 이 군집 1 로 그룹화된 방식에 주목하십시오 . 이들 시리얼은 해당 군집 내 다른 두 시리얼보다 더 유사합니다 .

그림 6.13 군집 1 의 유사한 시리얼



4. 군집 1 을 선택하려면 오른쪽의 빨간색 수평선을 클릭합니다 .

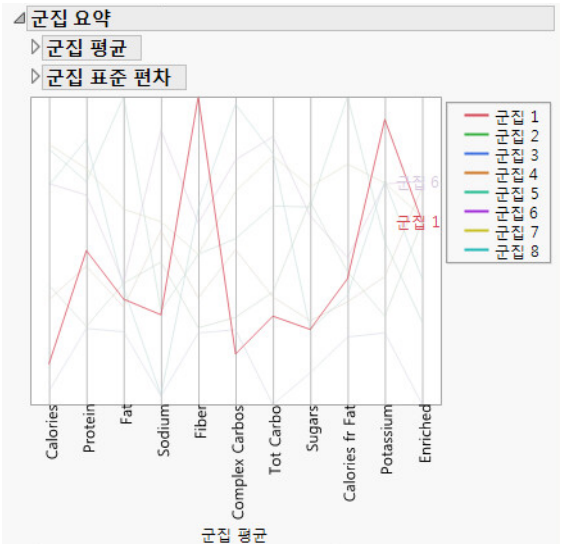
   4 개의 시리얼이 빨간색으로 강조 표시됩니다 .

그림 6.14 군집 선택

5. 군집 내 유사 특성을 확인하려면 " 계층적 군집화 " 의 빨간색 삼각형을 클릭하고 **군집 요약**을 선택합니다 .

   보고서 하단의 " 군집 요약 " 그래프에 각 군집의 변수별 평균 값이 표시됩니다 . 예를 들어 이 군집의 시리얼은 다른 군집의 시리얼보다 섬유질 및 칼륨 함량이 많습니다 .

그림 6.15 군집 요약

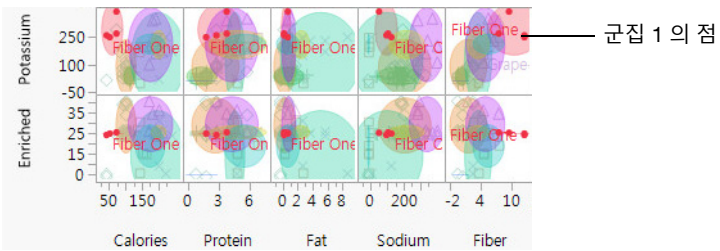


6. " 계층적 군집화 " 의 빨간색 삼각형을 클릭하고 **산점도 행렬**을 선택합니다 .

   다변량 플랫폼에서 산점도 행렬을 만드는 대신 이 옵션을 사용할 수 있습니다 .

   "Potassium" 행의 "Fiber" 그림에 주목하십시오 . 선택한 시리얼은 그림 오른쪽의 8g 과 13g 사이에 있습니다 . 이 위치는 군집 1 의 시리얼에 섬유질과 칼륨이 많다는 것을 나타냅니다 .

그림 6.16 군집 1 특성



참고 : 해당 점은 이전에 생성한 산점도 행렬이 아직 열려 있으면 여기서도 선택됩니다 .

결과 해석

군집을 각각 클릭하고 " 군집 요약 " 보고서를 보면 다음과 같은 특성을 알 수 있습니다 .

- "Fiber One" 과 "All-Bran" 같은 군집 1 의 시리얼은 섬유질 및 칼륨 함량은 높고 칼로리는 낮습니다 .
- 어린이들이 좋아하는 시리얼이 많이 포함된 군집 2 의 시리얼은 당분 함량이 높고 섬유질 , 복합 탄수화물 및 단백질 함량은 낮습니다 .
- 군집 3 의 시리얼 ("Puffed Rice" 와 "Puffed Wheat") 은 칼로리는 낮지만 영양가가 거의 없습니다 .
- "Total Corn Flakes" 와 "Multi-Grain Cheerios" 같은 군집 4 의 시리얼은 하루에 필요한 비타민과 무기질을 100% 공급합니다 . 이들 시리얼은 지방 , 섬유질 및 당분 함량이 낮습니다 .
- 군집 5 의 시리얼은 단백질 및 지방 함량이 높고 나트륨 함량은 적습니다 . 이 군집은 "Banana Nut Crunch" 와 "Quaker Oatmeal" 같은 시리얼로 구성되어 있습니다 .
- 군집 6 의 시리얼은 지방 함량은 낮고 나트륨 및 탄수화물 함량은 높습니다 . "Wheaties" 와 "Grape-Nuts" 같은 전통적인 시리얼이 이 군집에 속합니다 .
- 군집 7 의 시리얼은 칼로리는 높고 섬유질 함량은 낮습니다 . 말린 과일이 들어 있는 많은 시리얼이 이 군집에 속합니다 ("Mueslix Healthy Choice", "Low Fat Granola w Raisins", "Oatmeal Raisin Crisp", "Raisin Nut Bran" 및 "Just Right Fruit & Nut").
- 군집 8 의 시리얼은 나트륨 및 당분 함량은 낮고 복합 탄수화물 , 단백질 및 칼륨 함량은 풍부합니다 . "Shredded Wheat" 및 "Mini-Wheat" 시리얼이 이 군집에 속합니다 .

덴드로그램에서 결합 상태를 보면 각 군집에서 어떤 시리얼들이 가장 유사한지 알 수 있습니다.

- 군집 1 에서는 "Fiber One" 이 영양가 면에서 "All-Bran with Extra Fiber" 와 유사합니다 . "100% Bran" 과 "All-Bran" 도 유사합니다 . 유사한 각 쌍의 시리얼들은 서로 다른 회사에서 제조되므로 서로 경쟁 관계에 있습니다 .
- 군집 2 에서 "Frosted Flakes" 와 "Honey Frosted Wheaties" 는 하나는 옥수수 플레이크이고 다른 하나는 밀 플레이크임에도 불구하고 서로 유사합니다 . "Lucky Charms" 와 "Frosted Cheerios" 도 유사하고 , "Cap'n'Crunch" 와 "Trix" 도 유사합니다 .

결론

섬유질은 더 많이, 칼로리는 더 적게 섭취하려는 바람에 따라 군집 1의 시리얼을 먹어 보기로 결정합니다. 튀긴 밀과 튀긴 쌀로 구성되고 영양가가 거의 없는 군집 3의 시리얼은 피하려고 합니다. 또한 영양가가 높은 군집 4의 시리얼도 먹어 볼 것입니다.

7 장

작업 저장 및 공유

결과 저장 및 다시 생성

JMP에서는 데이터에서 생성된 결과를 여러 가지 방법으로 다른 사람들과 공유할 수 있습니다. 다음은 작업을 공유할 수 있는 몇 가지 방법입니다.

- 플랫폼 결과를 저널 , 프로젝트 또는 웹 보고서로 저장
- 결과 , 데이터 테이블 및 기타 파일을 프로젝트에 저장
- 스크립트를 저장하여 데이터 테이블에서 결과 재현
- 결과를 대화식 HTML(.htm, html) 로 저장
- 결과를 PowerPoint 프레젠테이션 (.pptx) 으로 저장
- 결과를 대시보드에서 공유

그림 7.1 웹 보고서의 예

목차

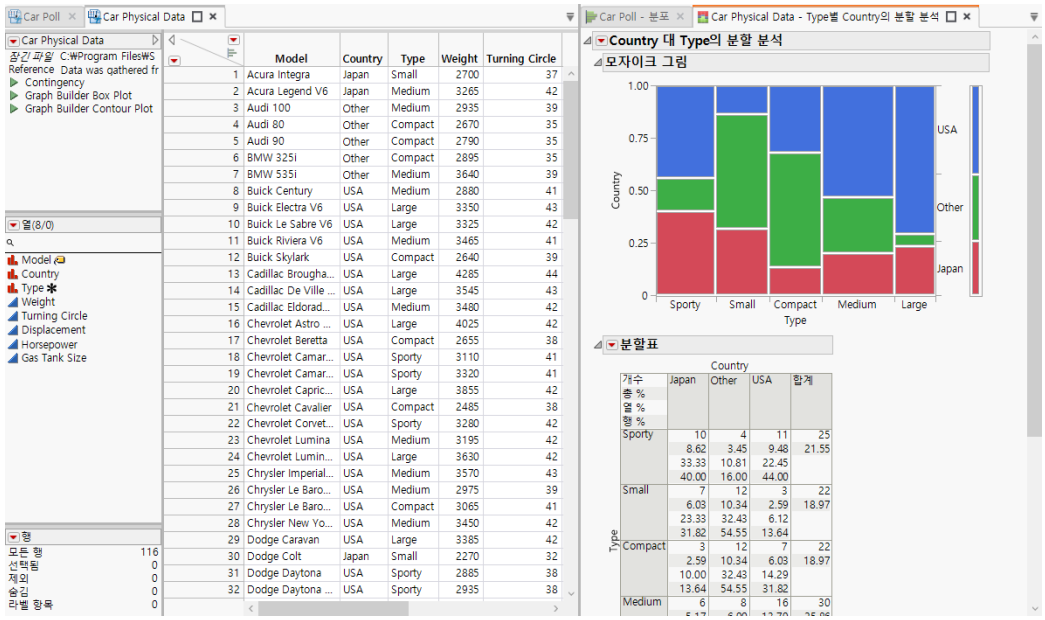
프로젝트 작업 . . . 179
새 프로젝트 생성 . . . 180
프로젝트에서 파일 열기 . . . 180
프로젝트에서 파일 재배열 . . . 180
프로젝트 저장 . . . 182
프로젝트 작업 영역 . . . 182
프로젝트 책갈피 . . . 183
프로젝트 내용 . . . 184
프로젝트의 최근 사용한 파일 . . . 184
프로젝트 로그 . . . 185
파일을 프로젝트로 이동 . . . 185
프로젝트 내의 열기 스크립트 작성 . . . 186
새 프로젝트 생성의 예 . . . 186
플랫폼 결과를 저널로 저장 . . . 189
예: 저널 생성 . . . 189
저널에 분석 추가 . . . 190
스크립트 저장 및 실행 . . . 191
예: 스크립트 저장 및 실행 . . . 191
스크립트 및 JSL . . . 192
보고서를 대화식 HTML 로 저장 . . . 192
대화식 HTML의 데이터 포함 . . . 193
예: 대화식 HTML 생성 . . . 194
보고서를 대화식 HTML로 공유 . . . 195
보고서를 PowerPoint 프레젠테이션으로 저장 . . . 199
대시보드 생성 . . . 199
예: 창 결합 . . . 200
예: 두 개의 보고서가 포함된 대시보드 생성 . . . 201

프로젝트 작업

JMP 프로젝트를 사용하여 다음을 수행할 수 있습니다.

- 탭 형식의 단일 JMP 창에서 데이터를 더 효율적으로 탐색합니다.
- 관련 파일 및 보고서 집합을 빠르게 저장하고 다시 엽니다.
- 자체 포함 프로젝트 파일에 테이블과 스크립트를 포함하여 작업을 쉽게 공유합니다.

그림 7.2 데이터 테이블 및 보고서가 포함된 프로젝트 파일



이 섹션에는 다음과 같은 정보가 포함되어 있습니다.

- "새 프로젝트 생성"
- "프로젝트에서 파일 열기"
- "프로젝트에서 파일 재배열"
- "프로젝트 저장"
- "프로젝트 작업 영역"
- "프로젝트 책갈피"
- "프로젝트 내용"
- "프로젝트의 최근 사용한 파일"
- "프로젝트 로그"
- "파일을 프로젝트로 이동"

- "프로젝트 내의 열기 스크립트 작성"
- "새 프로젝트 생성의 예"

새 프로젝트 생성

비어 있는 새 JMP 프로젝트를 생성하려면 **파일** > **새로 만들기** > **프로젝트**(Windows) 또는 **파일** > **새로 만들기** > **새 프로젝트**(macOS)를 선택합니다.

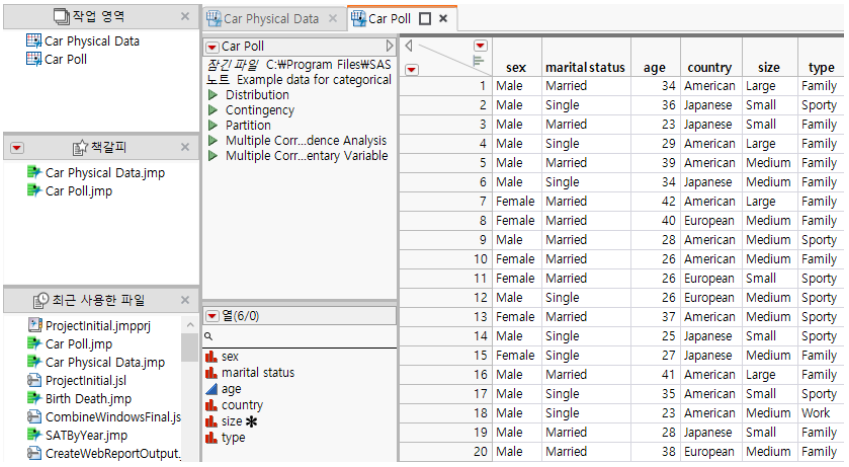
프로젝트에서 파일 열기

JMP 프로젝트에서 데이터 테이블을 열고 데이터에 대한 분석을 실행할 수 있습니다. 각 데이터 테이블과 분석 보고서는 새 탭에서 열립니다.

1. 프로젝트 창에서 **파일** > **열기**를 선택하고 열려는 데이터 테이블로 이동합니다.
2. 데이터 테이블에서 분석을 실행합니다.

컴퓨터에서 데이터 테이블이 업데이트된 경우 프로젝트를 다시 열면 프로젝트의 모든 관련 보고서가 업데이트됩니다.

그림 7.3 초기 프로젝트



팁 : 프로젝트에서 Ctrl+S 또는 Ctrl+W 같은 JMP 바로 가기 키를 사용하여 파일을 저장하거나 활성 창 페인을 닫을 수 있습니다. 전체 목록을 보려면 **도움말** > **빠른 참조 카드**를 선택하십시오.

프로젝트에서 파일 재배열

JMP 프로젝트에서 데이터 테이블과 보고서는 개별 탭에 나타납니다. 탭을 도킹 영역으로 드래그하여 재배열할 수 있습니다.

팁 : 도킹된 항목을 실행 취소하거나 다시 실행하려면 **프로젝트** > **레이아웃 실행 취소** 또는 **프로젝트** > **레이아웃 다시 실행**을 선택합니다 . 모든 테이블과 보고서를 개별 탭으로 되돌리려면 **프로젝트** > **레이아웃 재설정**을 선택합니다 .

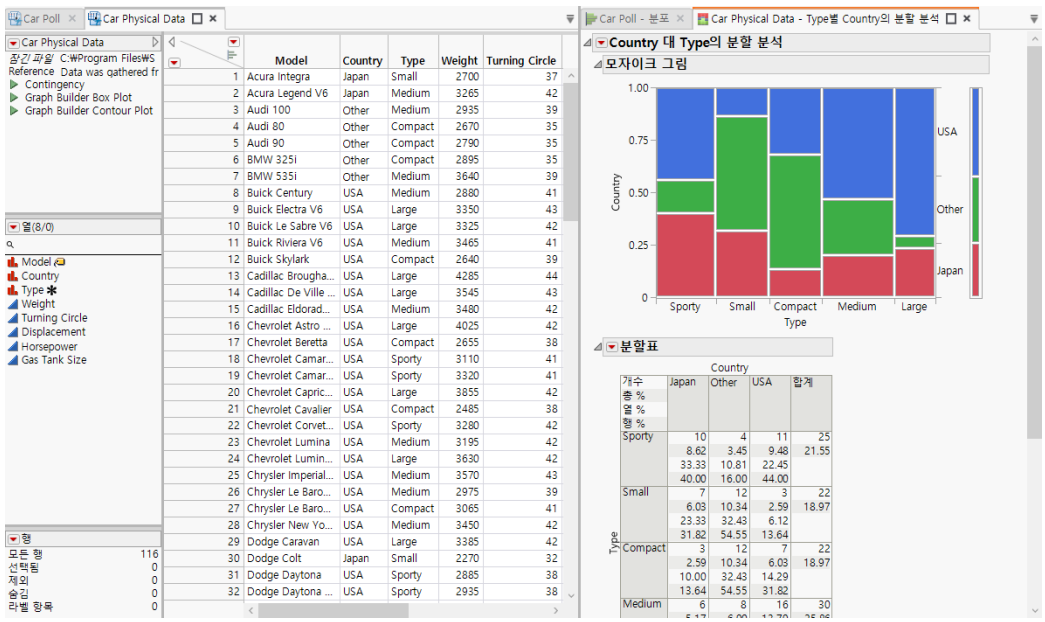
프로젝트 파일 재배열의 예

1. **파일** > **새로 만들기** > **프로젝트** (Windows) 또는 **파일** > **새로 만들기** > **새 프로젝트** (macOS) 를 선택합니다 .
2. **도움말** > **샘플 데이터 라이브러리**를 선택하고 Car Physical Data.jmp 및 Car Poll.jmp를 엽니다.
3. Car Poll.jmp 데이터 테이블에서 Distribution 스크립트를 실행합니다 .
4. Car Physical Data.jmp 데이터 테이블에서 Contingency 스크립트를 실행합니다 .
5. **Car Poll - 분포** 보고서 탭을 오른쪽으로 드래그하여 오른쪽 도킹 영역에 놓습니다 .

   프로젝트 창 오른쪽에 " 분포 " 보고서 페인이 나타납니다 . 보고서가 도킹되어 탭 간에 전환할 때 계속 표시됩니다 .
6. **Car Physical Data - 분할 분석** 탭을 **Car Poll - Distribution** 보고서의 가운데로 드래그하여 탭 도킹 영역에 놓습니다 .

   팁 : 데이터 테이블 창의 크기를 조정하여 두 보고서를 모두 표시할 수 있습니다 .

그림 7.4 탭으로 그룹화된 보고서 및 데이터 테이블



두 데이터 테이블과 마찬가지로 "분포" 보고서와 "분할 분석" 보고서가 탭 형식으로 함께 표시됩니다.

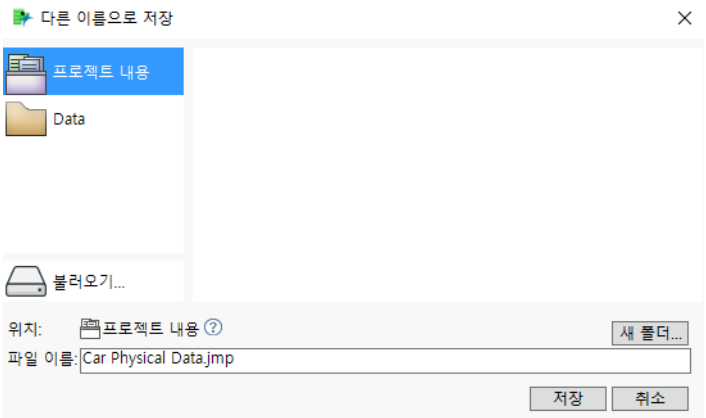
프로젝트 저장

프로젝트를 공유, 배포 또는 보관하려면 모든 데이터 테이블과 스크립트를 프로젝트 내용에 저장하여 자체 포함 프로젝트를 생성합니다. 이렇게 하면 모든 프로젝트 파일과 폴더를 단일 프로젝트 파일에 포함하여 다른 JMP 사용자와 공유할 수 있습니다.

1. (선택 사항) 프로젝트를 자체 포함 프로젝트로 저장하려면 **파일** > **다른 이름으로 저장**을 선택한 후 **프로젝트 내용**을 선택하여 각 데이터 테이블과 스크립트를 프로젝트에 저장합니다. **새 폴더**를 클릭하여 프로젝트 내용에 폴더를 생성할 수 있습니다.

   참고: 보고서는 프로젝트에 자동 저장되므로 저장할 필요가 없습니다.

그림 7.5 프로젝트 내용에 데이터 테이블 저장



2. **파일** > **프로젝트 저장**을 선택하여 프로젝트 파일을 저장합니다.

그림 7.6 프로젝트 파일 저장

프로젝트 작업 영역

JMP 프로젝트에서 현재 열려 있는 모든 파일을 "작업 영역" 도구 페인에서 볼 수 있습니다. 보고서는 해당 데이터 테이블 아래에 나타납니다. 활성 데이터 테이블은 굵게 표시됩니다.

팁 : " 작업 영역 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **작업 영역 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

"작업 영역" 도구 페인에서 다음을 수행할 수 있습니다.

- 파일을 두 번 클릭하여 활성화합니다 .
- 파일을 마우스 오른쪽 버튼으로 클릭하여 파일 닫기 , 숨기기 (탭을 제거하고 파일 이름을 흐리게 표시), 책갈피 지정 등을 수행합니다 .

프로젝트 책갈피

JMP 프로젝트에서 파일 또는 폴더의 바로 가기를 " 책갈피 " 도구 페인에 생성할 수 있습니다. 책갈피 그룹을 생성하여 책갈피 지정된 파일과 폴더를 구성할 수도 있습니다.

팁 : " 책갈피 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **책갈피 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

프로젝트에서 열린 파일 책갈피 지정

- 탭을 마우스 오른쪽 버튼으로 클릭하고 **책갈피**를 선택합니다 .
- 열려 있는 모든 파일을 책갈피 지정하려면 탭을 마우스 오른쪽 버튼으로 클릭하고 **모두 책갈피**를 선택합니다 .

컴퓨터의 파일 또는 폴더 책갈피 지정

컴퓨터의 파일 또는 폴더를 " 책갈피 " 도구 페인으로 드래그하거나 , " 책갈피 " 의 빨간색 삼각형 메뉴를 클릭하고 **파일 추가** 또는 **폴더 추가**를 선택합니다.

참고 : 컴퓨터의 책갈피 지정된 폴더에서 파일을 추가하거나 제거하면 폴더 내용이 프로젝트에서 자동으로 업데이트됩니다 .

책갈피 지정된 파일 열기

" 책갈피 " 도구 페인에서 파일을 두 번 클릭합니다.

책갈피 제거

" 책갈피 " 도구 페인에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **책갈피 제거**를 선택합니다.

책갈피 그룹 생성

1. " 책갈피 " 의 빨간색 삼각형 메뉴를 클릭하고 **새 그룹**을 선택합니다 .

2. 그룹 이름을 지정하고 **확인**을 클릭합니다 .
3. 새 책갈피 또는 기존 책갈피를 그룹으로 드래그하거나 , " 책갈피 " 도구 페인에서 그룹을 마우스 오른쪽 버튼으로 클릭하고 **파일 추가** 또는 **폴더 추가**를 선택합니다 .

프로젝트 내용

자체 포함 JMP 프로젝트에서 프로젝트 내용에 저장된 모든 파일이 "내용" 도구 페인에 나타납니다.

팁 : " 내용 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **내용 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

프로젝트 내용에 있는 파일 열기

"내용" 도구 페인에서 파일을 두 번 클릭합니다.

프로젝트 내용에 폴더 생성

1. " 내용 " 의 빨간색 삼각형 메뉴를 클릭하고 **새 폴더**를 선택합니다 .
2. 폴더 이름을 지정하고 **확인**을 클릭합니다 .

파일을 폴더로 이동

"내용" 도구 페인에서 파일을 폴더로 드래그합니다.

컴퓨터의 파일 또는 폴더를 프로젝트 내용에 복사

1. " 내용 " 의 빨간색 삼각형 메뉴를 클릭하고 **프로젝트에 파일 복사** 또는 **프로젝트에 폴더 복사**를 선택합니다 .
2. 파일을 탐색한 후 **열기**를 클릭하거나 폴더를 탐색한 후 **선택**을 클릭합니다 .

프로젝트 내용에 있는 파일 이름 바꾸기

"내용" 도구 페인에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **이름 바꾸기**를 선택합니다.

프로젝트 내용에 있는 파일 삭제

"내용" 도구 페인에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **삭제**를 선택합니다.

프로젝트의 최근 사용한 파일

JMP 프로젝트의 "최근 사용한 파일" 도구 페인에서 파일을 열 수 있습니다.

팁 : " 최근 사용한 파일 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **최근 사용한 파일 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

파일 열기

"최근 사용한 파일" 도구 페인에서 파일을 두 번 클릭하거나 프로젝트로 드래그합니다.

참고: 프로젝트 내용에 저장된 열린 파일은 "최근 사용한 파일" 도구 페인에 표시되지 않습니다.

프로젝트 로그

JMP 프로젝트의 "로그" 도구 페인에서 로그 메시지를 볼 수 있습니다.

팁 :

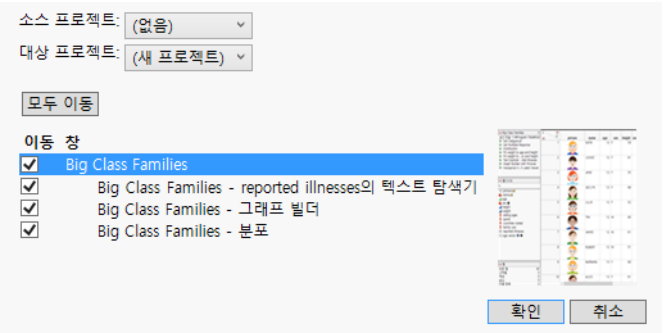
- " 로그 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **로그 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .
- 기본적으로 프로젝트에서 발생하는 로그 메시지는 기본 JMP 로그가 아니라 프로젝트 로그에만 나타납니다 . 이 설정을 변경하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 **프로젝트 로그 페인 사용**을 **열려 있는 경우** 또는 **안 함**으로 업데이트합니다 .

파일을 프로젝트로 이동

JMP에서는 파일을 프로젝트로 이동하거나 프로젝트 외부 또는 프로젝트 간에 이동할 수 있습니다.

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 Big Class Families.jmp 를 엽니다 .
2. "Distribution", "Text Explorer" 및 "Graph Builder" 스크립트를 실행합니다 .
3. 어느 창에서든 **창** > **프로젝트 간 이동**을 선택합니다 .
4. 프로젝트에 열려 있지 않은 파일을 보려면 " 소스 프로젝트 " 를 **(없음)** 으로 그대로 둡니다 .
5. 선택한 파일을 새로 생성한 프로젝트로 이동하려면 " 대상 프로젝트 " 를 **(새 프로젝트)** 로 그대로 둡니다 .
6. Big Class Families.jmp 옆의 체크박스를 선택합니다 . 데이터 테이블과 연결된 그래프도 선택됩니다 .

그림 7.7 프로젝트 간 창 이동



7. 확인을 클릭합니다. 선택한 테이블과 보고서로 새 프로젝트가 생성됩니다.

프로젝트 내의 열기 스크립트 작성

"프로젝트 내의 열기 스크립트" 상자에는 이 특정 프로젝트를 열 때마다 실행할 JSL 스크립트를 추가할 수 있습니다. 예를 들어 보고서를 생성 또는 수정하고, 사용자에게 정보를 요청하거나, 사용 정보를 로그 창에 쓰는 스크립트를 생성할 수 있습니다.

1. **프로젝트** > **프로젝트 설정**을 선택합니다.
2. JSL 스크립트를 붙여 넣고 **확인**을 클릭합니다.

팁 :

- 자세한 내용은 Scripting Guide 의 확인하십시오. 여기에서는 특정 프로젝트 대신 모든 프로젝트를 열 때 시작 스크립트를 생성하는 방법을 설명합니다.
- 기존 프로젝트를 새 프로젝트의 템플릿으로 사용하려면 **파일** > **환경 설정** > **프로젝트**에서 기존 프로젝트를 새 프로젝트 템플릿으로 지정합니다.

새 프로젝트 생성의 예

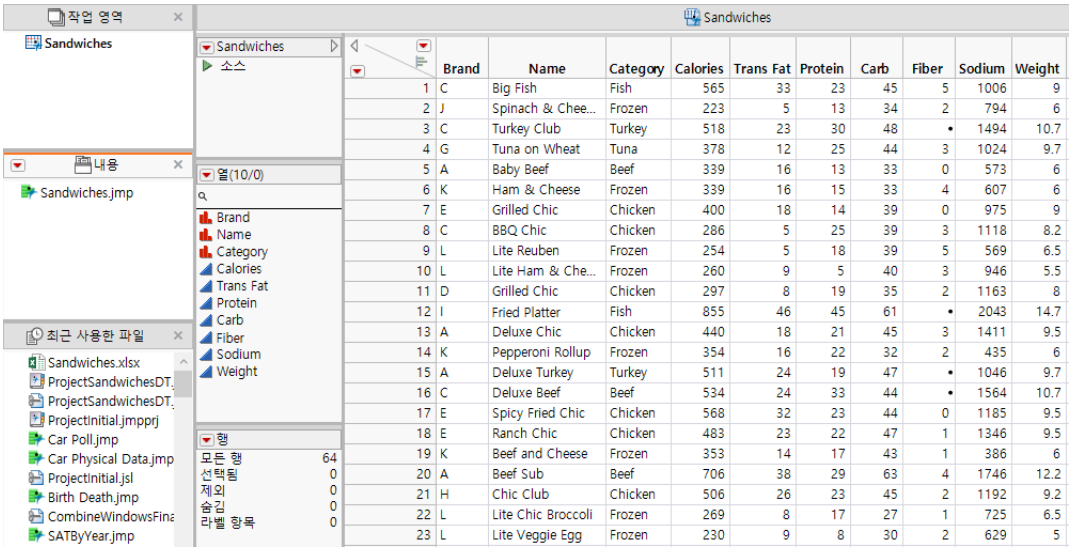
이 예에서는 프로젝트 생성, 데이터 가져오기, 프로젝트 창에서 분석 생성 및 보고서 도킹, 부분 집합 테이블 생성을 차례로 수행한 후 프로젝트를 저장하고, 닫았다가 다시 엽니다.

1. **파일** > **새로 만들기** > **프로젝트** (Windows) 또는 **파일** > **새로 만들기** > **새 프로젝트** (macOS) 를 선택합니다.
2. 프로젝트 창에서 **파일** > **열기**를 선택합니다.
3. sandwiches.xlsx 파일을 엽니다. 이 파일의 기본 위치는 다음과 같습니다.
   C:/Program Files/SAS/JMP/16/Samples/Import Data

   팁 : 하단에서 **모든 JMP 파일**을 **Excel 파일**로 변경해야 할 수도 있습니다.
4. 가져오기를 클릭합니다.

5. **파일** > **저장**을 선택합니다 .
6. **프로젝트 내용**이 선택되어 있는지 확인합니다 .
7. 파일 이름을 Sandwiches.jmp 로 변경하고 **저장**을 클릭합니다 .

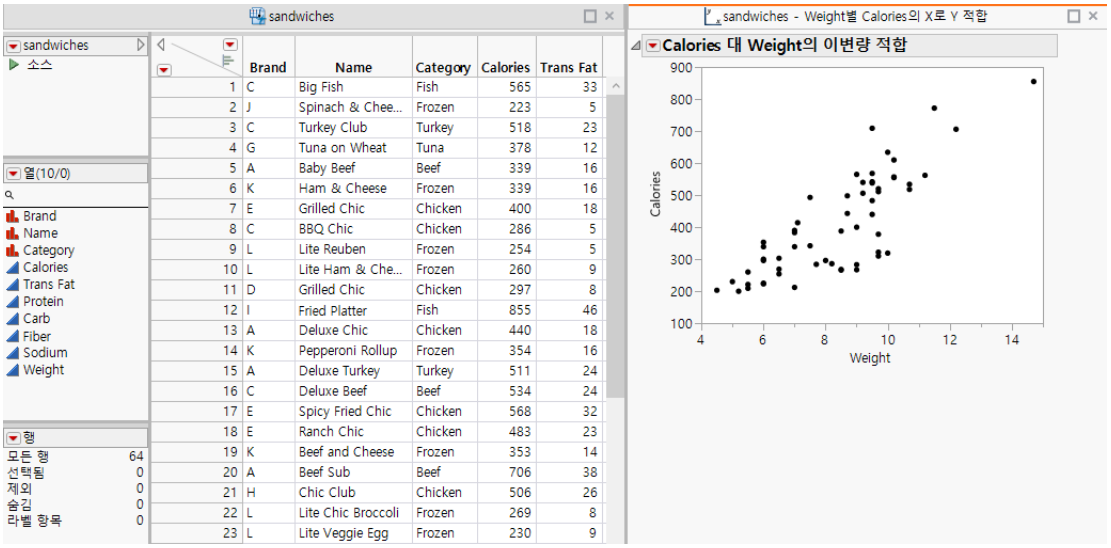
그림 7.8 Sandwiches 데이터를 가져온 프로젝트



8. 분석 > X 로 Y 적합을 선택합니다 .
9. Calories 를 선택하고 **Y, 반응**을 클릭합니다 .
10. Weight 를 선택하고 **X, 요인**을 클릭합니다 .
11. 확인을 클릭합니다 .
12. **Sandwiches - X 로 Y 적합** 탭을 오른쪽으로 드래그하여 오른쪽 도킹 영역에 놓습니다 .

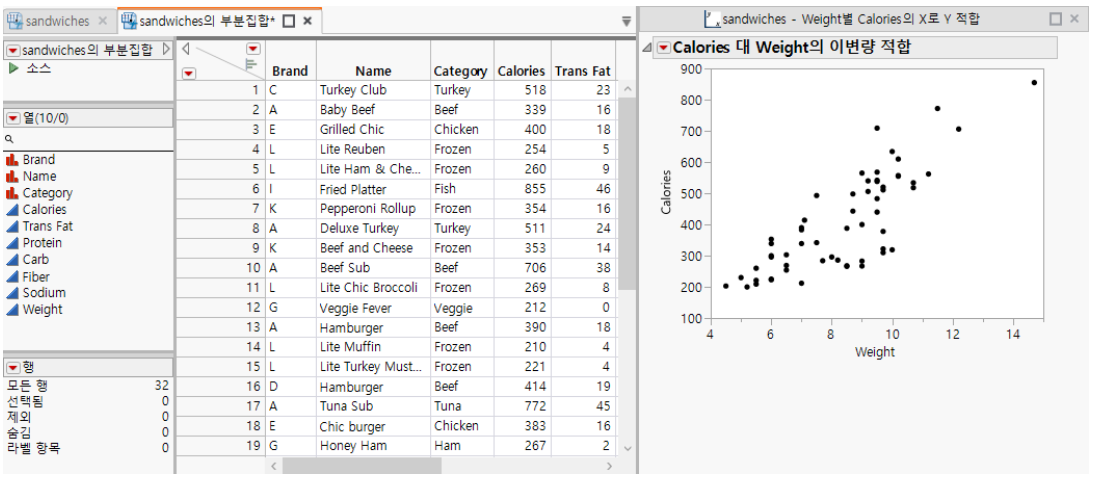
팁 : 전체 보고서를 표시하려면 데이터 테이블과 보고서 사이의 선을 왼쪽으로 드래그합니다 .

그림 7.9 오른쪽에 도킹된 X 로 Y 적합 (이변량) 보고서



13. **테이블** > **부분집합**을 선택합니다 .
14. " 행 " 아래에서 **랜덤 - 표집 비율 : 0.5** 를 선택합니다 .
15. 확인을 클릭합니다 .

그림 7.10 저장되지 않은 부분집합 테이블이 있는 프로젝트



16. **파일** > **프로젝트 저장**을 선택합니다 .
17. 프로젝트를 저장할 폴더로 이동하여 프로젝트 파일 이름을 지정하고 **저장**을 클릭합니다 .
18. 프로젝트를 닫습니다 .

19. **파일** > **열기**를 선택하고 프로젝트 파일을 엽니다 .

저장하지 않은 부분집합 테이블이 프로젝트 파일에 자동으로 저장되어 있습니다.

플랫폼 결과를 저널로 저장

보고서 창의 저널을 생성하여 플랫폼 보고서를 나중에 볼 수 있도록 저장합니다. 저널은 보고서 창의 사본입니다. 기존 저널을 편집하거나 기존 저널에 보고서를 추가할 수 있습니다. 저널은 데이터 테이블에 연결되지 않습니다. 저널은 여러 보고서 창의 결과를 단일 보고서 창에 저장하여 다른 사람들과 간편하게 공유할 수 있는 기능입니다.

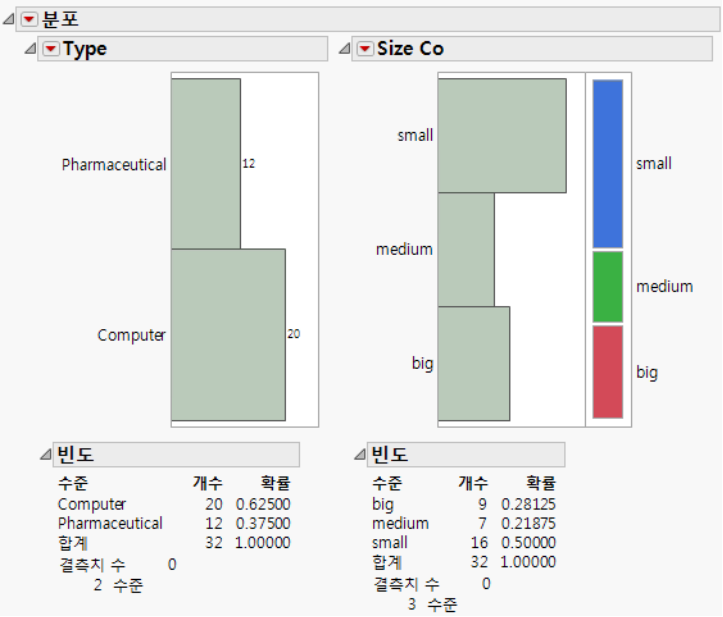
이 섹션에는 다음과 같은 정보가 포함되어 있습니다 .

- " 예 : 저널 생성 "
- " 저널에 분석 추가 "

예 : 저널 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석** > **분포**를 선택합니다 .
3. Type 및 Size Co 를 모두 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "Type" 의 빨간색 삼각형을 클릭하고 **히스토그램 옵션** > **개수 표시**를 선택합니다 .
6. "Size Co" 의 빨간색 삼각형을 클릭하고 **모자이크 그림**을 선택합니다 .
7. **편집** > **저널**을 선택하여 이 결과에 대한 저널을 생성합니다 . 결과가 저널 창에 복제됩니다 .

그림 7.11 분포 결과의 저널



저널의 결과는 데이터 테이블에 연결되지 않습니다. "Type" 막대 차트에서 "Computer" 막대를 클릭해도 데이터 테이블에서 행이 선택되지 않습니다.

저널은 결과의 사본이기 때문에 빨간색 삼각형 메뉴가 대부분 존재하지 않습니다. 그러나 저널에 새로 추가한 보고서에는 빨간색 삼각형 메뉴가 있습니다. 이 메뉴에는 두 가지 옵션이 있습니다.

새 창에서 다시 실행 원래 보고서를 생성하는 데 사용된 원래 데이터 테이블이 있을 때 이 옵션은 분석을 다시 실행합니다. 그 결과로 새 보고서 창이 표시됩니다.

스크립트 편집 이 옵션은 분석을 다시 생성하는 JSL 스크립트가 포함된 스크립트 창을 엽니다. JSL 에 대해서는 좀 더 자세한 설명이 필요하므로 Scripting Guide 및 JSL Syntax Reference 에서 별도로 다룹니다.

저널에 분석 추가

다른 분석을 수행하는 경우 분석 결과를 기존 저널에 추가할 수 있습니다.

1. 저널이 열린 상태에서 **분석** > **분포**를 선택합니다.
2. profit/emp 를 선택하고 **Y, 열**을 클릭합니다.
3. **확인**을 클릭합니다.
4. **편집** > **저널**을 선택합니다. 결과가 저널 하단에 추가됩니다.

스크립트 저장 및 실행

JMP 의 플랫폼 옵션은 대부분 스크립트 가능합니다 . 즉 , 수행하는 대부분의 작업을 JSL(JMP 스크립트 언어) 스크립트로 저장할 수 있습니다 . 스크립트를 사용하여 언제든지 작업 또는 결과를 재현할 수 있습니다 .

이 섹션에는 다음과 같은 정보가 포함되어 있습니다 .

- " 예 : 스크립트 저장 및 실행 "
- " 스크립트 및 JSL"

예 : 스크립트 저장 및 실행

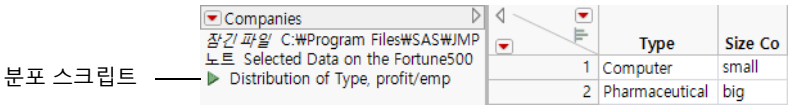
보고서 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Type 및 profit/emp 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "Type" 의 빨간색 삼각형을 클릭하고 다음 옵션을 선택합니다 .
   - **히스토그램 옵션 > 개수 표시**
   - **신뢰 구간 > 0.95**
6. "profit/emp" 의 빨간색 삼각형을 클릭하고 다음 옵션을 선택합니다 .
   - **이상치 상자 그림** (이상치 상자 그림 제거)
   - **CDF 그림**
7. " 분포 " 의 빨간색 삼각형을 클릭하고 **쌓기**를 선택합니다 .

스크립트를 데이터 테이블에 저장하고 실행

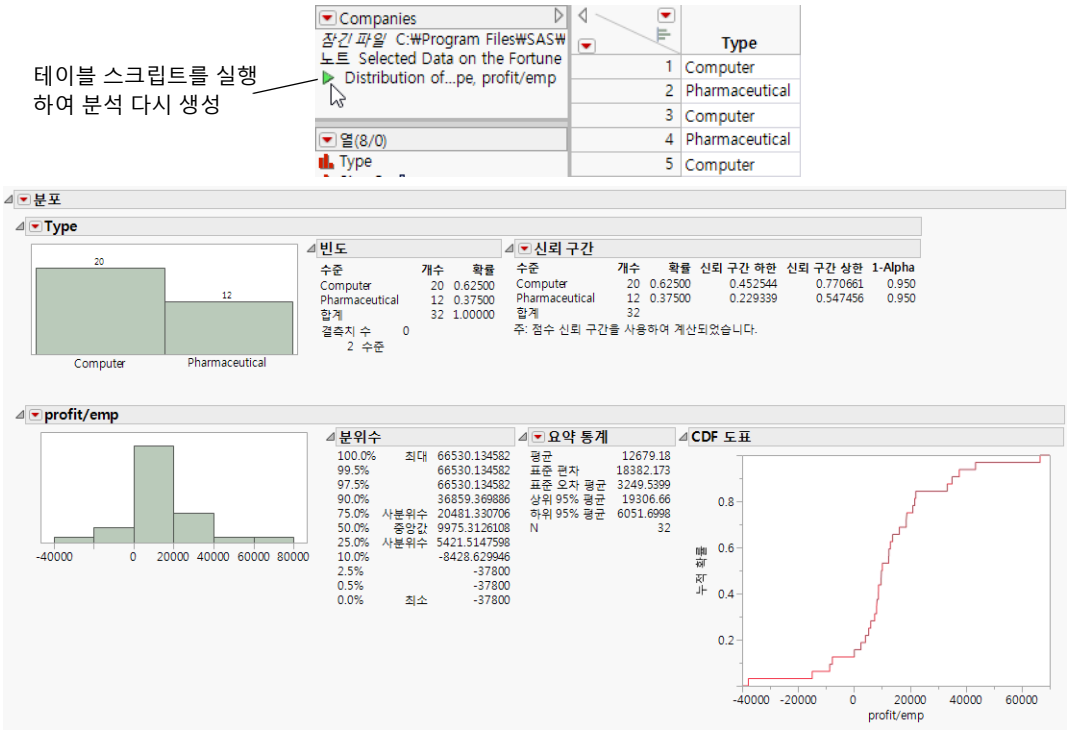
1. 이 분석을 저장하려면 " 분포 " 의 빨간색 삼각형을 클릭하고 **스크립트 저장 > 데이터 테이블에**를 선택합니다 . 테이블 패널에 새 스크립트가 나타납니다 .

그림 7.12 분포 스크립트



2. " 분포 " 보고서 창을 닫습니다 .
3. 분석을 다시 생성하려면 " 분포 " 스크립트 옆의 녹색 삼각형을 클릭합니다 .

그림 7.13 분포 스크립트 실행



팁 : 테이블 스크립트를 마우스 오른쪽 버튼으로 클릭하면 추가 옵션이 표시됩니다 .

스크립트 및 JSL

이 섹션에서 저장한 스크립트에는 JSL(JMP 스크립트 언어) 명령이 포함되어 있습니다. JSL에 대해서는 좀 더 자세한 설명이 필요하므로 Scripting Guide 및 JSL Syntax Reference에서 별도로 다룹니다.

보고서를 대화식 HTML 로 저장

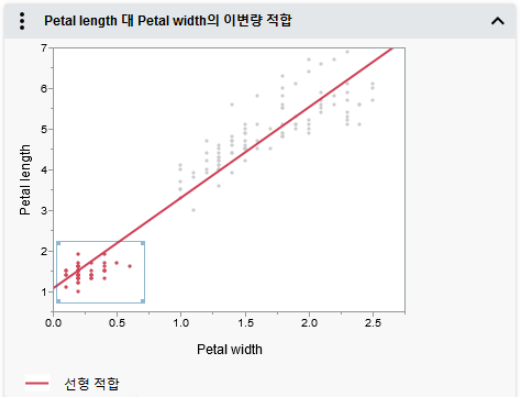
JMP 사용자가 대화식 HTML을 통해 동적 그래프가 포함된 보고서를 공유하면 JMP를 사용하지 않는 사람도 데이터를 탐색할 수 있습니다. JMP 보고서는 사용자에게 전자 우편으로 보내거나 웹 사이트에 게시할 수 있는 HTML 5 형식의 웹 페이지로 저장됩니다. 그러면 사용자들이 JMP에서처럼 데이터를 탐색할 수 있습니다.

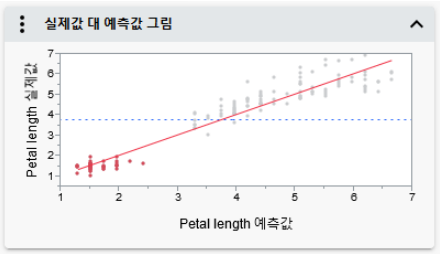
대화식 HTML 은 다음과 같은 JMP 의 일부 기능을 제공합니다 .

- 히스토그램 막대를 선택하고 데이터 값을 보는 등 대화식 그래프 기능을 탐색합니다 .

- 브러시 방식으로 데이터를 확인합니다.
- 보고서 섹션을 표시하거나 숨깁니다.
- 보고서를 커서로 가리켜 툴팁을 봅니다.
- 표식 크기를 늘립니다.

그림 7.14 대화식 HTML 에서 브러시 방식으로 데이터 확인





정렬된 변수, 가로 히스토그램, 배경 색상 및 색상이 적용된 데이터 점 같은 그래프의 많은 변경 사항이 웹 페이지에 저장됩니다. 내용을 저장할 때 닫힌 그래프와 테이블은 사용자가 열 때까지 웹 페이지에서 닫힌 채로 유지됩니다.

대화식 HTML 의 데이터 포함

JMP 에서 보고서를 대화식 HTML 로 저장하면 데이터가 HTML 에 포함됩니다. 웹 브라우저는 암호화된 데이터를 읽을 수 없기 때문에 내용이 암호화되지 않습니다. 중요한 데이터를 공유하지 않으려면 결과를 대화식이 아닌 웹 페이지로 저장하십시오. **파일 > 내보내기 > 데이터가 포함된 대화식 HTML** 을 선택하십시오.

예 : 대화식 HTML 생성

보고서 생성

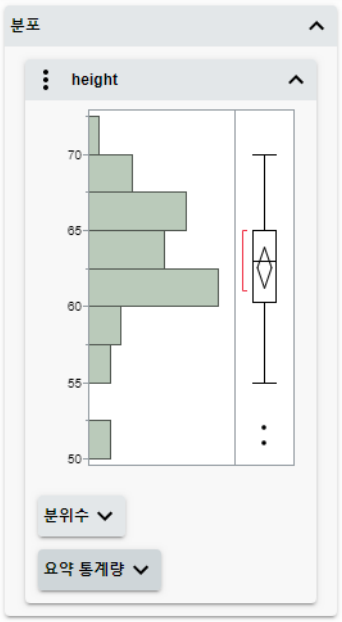
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. height 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

대화식 HTML 로 저장

1. (Windows) **파일 > 내보내기**를 선택하고 **데이터가 포함된 대화식 HTML** 을 선택한 후 **확인**을 클릭합니다 .
2. (macOS) **파일 > 내보내기**를 선택하고 **데이터가 포함된 대화식 HTML** 을 선택한 후 **다음**을 클릭합니다 .
3. " 내보내기 " 창에서 **저장 후 파일 열기**가 선택되어 있지 않으면 이 항목을 선택합니다 .
4. 파일 이름을 지정하고 파일을 저장합니다 .

   출력이 기본 브라우저에 나타납니다 .

그림 7.15 대화식 HTML 출력



대화식 HTML 출력을 탐색하는 방법에 대한 자세한 내용은 https://www.jmp.com/interactive 에서 확인하십시오.

보고서를 대화식 HTML 로 공유

대화식 HTML 을 통해 동적 그래프가 포함된 JMP 보고서를 공유하면 JMP 를 사용하지 않는 사람도 데이터를 탐색할 수 있습니다. JMP 보고서는 공유 네트워크 드라이브나 웹 사이트에서 또는 이메일을 통해 다른 사람과 공유할 수 있는 대화식 웹 페이지로 저장됩니다. 그러면 사용자들이 JMP 에서처럼 데이터를 탐색할 수 있습니다.

대화식 HTML 의 데이터 포함

보고서를 대화식 HTML 로 내보내거나 게시하면 데이터가 HTML 에 포함됩니다. 웹 브라우저는 암호화된 데이터를 읽을 수 없기 때문에 내용이 암호화되지 않습니다. 중요한 데이터를 공유하지 않으려면 **파일 > 내보내기 > HTML** 을 선택하여 결과를 대화식이 아닌 웹 페이지로 내보내십시오.

대화식으로 지원되는 기능

대화식 HTML 은 다음과 같은 JMP 의 일부 기능을 제공합니다.

- 보고서의 기능이 완전히 지원되면 경고 없이 웹 페이지가 생성됩니다 .
- 보고서에 지원되지 않는 기능이 포함되어 있으면 내보내기 또는 게시 창에 대화식 HTML 이 부분적으로 구현되었다는 메시지가 표시됩니다 . 자세한 내용은 JMP 로그 (**보기** > **로그**) 에서 확인하십시오 .
- 부분적으로 지원되거나 지원되지 않는 기능은 웹 페이지에 정적 상태로 나타납니다. 웹 페이지에서 지원되지 않는 기능을 커서로 가리키면 해당 기능이 아직 대화식이 아니라는 툴팁이 표시됩니다 .

대화식 HTML 출력을 탐색하는 방법에 대한 자세한 내용은 https://www.jmp.com/interactive 에서 확인하십시오.

단일 보고서를 대화식 HTML 로 내보내기

단일 JMP 보고서를 대화식 HTML 로 내보내려면 " 데이터가 포함된 대화식 HTML 로 내보내기 " 옵션을 사용하여 단일 웹 페이지를 생성합니다.

1. JMP 에서 보고서를 생성하고 해당 창을 활성화합니다 .

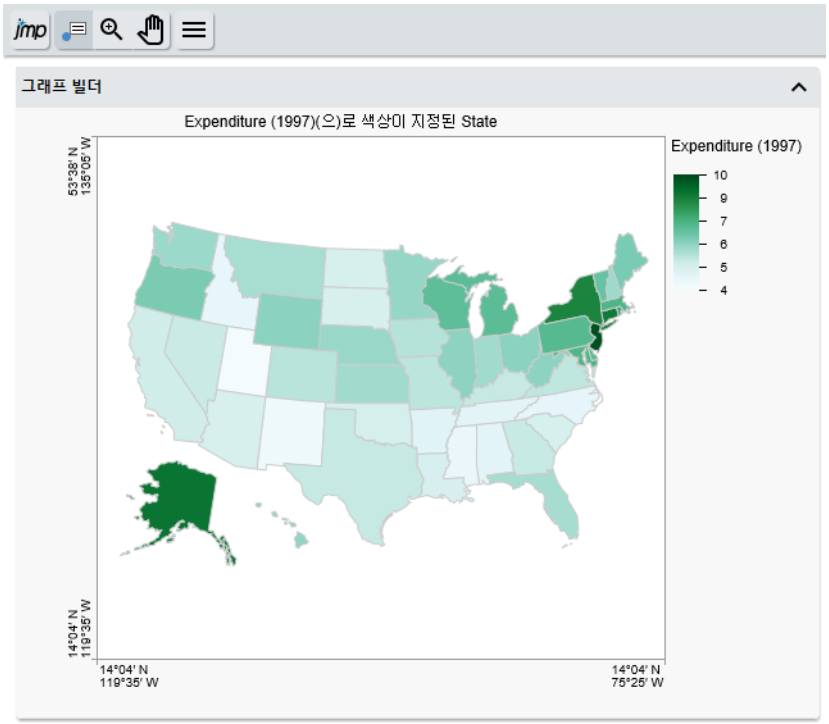
   참고 : 보고서에 로컬 데이터 필터가 포함되어 있으면 웹 보고서에서 정적 상태이며 사용자가 선택을 변경할 수 없습니다 . 로컬 데이터 필터를 대화식으로 만들려면 **포함** 모드를 선택 취소하십시오 . 그래프 빌더에서 **포함** 모드를 표시하려면 " 로컬 데이터 필터 " 의 빨간색 삼각형 메뉴에서 **모드 표시** 를 선택합니다 .

2. 파일 > 내보내기 를 선택하고 데이터가 포함된 대화식 HTML 을 선택한 후 다음 을 클릭합니다 .
3. 파일 이름을 지정합니다 .
4. (선택 사항) HTML 파일을 내보낸 후 기본 브라우저에서 열려면 **저장 후 파일 열기** 를 선택합니다 .

5. 저장을 클릭합니다 .

   출력이 선택한 폴더에 저장됩니다 .

그림 7.16 단일 대화식 HTML 보고서에 대한 웹 페이지



여러 보고서를 대화식 HTML 로 게시

여러 JMP 보고서를 대화식 HTML 로 게시하려면 " 파일에 게시 " 옵션을 사용하여 보고서가 포함된 인덱스 페이지를 생성합니다 .

팁 : 여러 JMP 보고서를 게시하는 경우 인덱스 페이지와 보고서 파일을 저장할 위치를 지정해야 합니다 . 공유 네트워크 폴더를 선택하고 다른 사람에게 위치를 알려주거나 , 공유하기 전에 사용자 컴퓨터의 폴더를 선택하고 파일을 압축할 수 있습니다 . 이 방법은 JMP 를 사용하지 않는 사람과 공유할 때 특히 유용합니다 .

1. JMP 에서 보고서를 생성합니다 .

   참고 : 보고서에 로컬 데이터 필터가 포함되어 있으면 웹 보고서에서 정적 상태이며 사용자가 선택을 변경할 수 없습니다 . 로컬 데이터 필터를 대화식으로 만들려면 **포함** 모드를 선택 취소하십시오 . 그래프 빌더에서 **포함** 모드를 표시하려면 " 로컬 데이터 필터 " 의 빨간색 삼각형 메뉴에서 **모드 표시**를 선택합니다 .

2. 보고서 창에서 **파일** > **게시** > **파일에 게시**를 선택합니다 .
3. 게시할 보고서를 선택합니다 .
4. (선택 사항) 상위 폴더 위치를 변경하거나 보고서를 포함할 하위 폴더의 이름을 변경합니다 .
5. 다음을 클릭합니다 .
6. 인덱스 페이지의 제목을 입력합니다 . 보고서 제목을 업데이트할 수도 있습니다 .
7. (선택 사항) 추가 옵션을 변경합니다 . 자세한 내용은 " 대화식 HTML 보고서 옵션 " 에서 확인하십시오 .
8. 게시를 클릭합니다 .

그림 7.17 여러 대화식 HTML 보고서에 대한 인덱스 페이지

9. 썸네일을 클릭하여 보고서를 엽니다 .

대화식 보고서 작업에 대한 자세한 내용을 보려면 HTML 보고서에서 ☰ > **도움말**을 클릭합니다 . 그러면 도움말 페이지 (https://www.jmp.com/interactive) 가 열립니다 .

대화식 HTML 보고서 옵션

제목 인덱스 페이지 (보고서가 여러 개인 경우만 해당) 또는 보고서의 제목을 추가합니다 .

설명 (선택 사항) 인덱스 페이지 또는 보고서에 설명을 추가합니다 . 이 설명은 제목 아래에 표시됩니다 .

사용자 정의 (보고서가 여러 개인 경우에만 표시) 웹 페이지의 모양을 변경합니다 . 스타일 , 테마 , 글꼴 , 로고 및 날짜 / 시간 표시 여부를 변경할 수 있습니다 . 자세한 내용은 " 인덱스 페이지 사용자 정의 옵션 " 에서 확인하십시오 .

데이터 게시 대화식 보고서의 경우 이 옵션을 선택합니다. 이 옵션을 선택 취소하면 정적 보고서가 됩니다 .

참고 : 중요한 데이터를 공유하지 않으려면 결과를 대화식이 아닌 웹 페이지로 저장하십시오 . **파일 > 내보내기 > HTML** 을 선택합니다 .

이미지 추가 웹 페이지 아래쪽에 이미지를 추가합니다 .

게시된 웹 보고서 열기 " 게시 " 를 클릭하면 브라우저에서 웹 페이지를 엽니다 .

실행 후 보고서 닫기 웹 보고서를 게시하면 JMP 에서 보고서를 닫습니다 .

삭제 아이콘 (보고서가 여러 개인 경우에만 표시) 보고서를 삭제합니다 .

화살표 아이콘 (보고서가 여러 개인 경우에만 표시) 보고서 순서를 변경합니다 .

인덱스 페이지 사용자 정의 옵션

스타일 형식 보고서의 레이아웃을 결정합니다 .

큰 목록 보고서를 큰 썸네일 그래픽과 함께 열 하나에 표시합니다 .

작은 목록 보고서를 작은 썸네일 그래픽과 함께 열 하나에 표시합니다 .

격자 보고서를 여러 행에 표시합니다 .

사용자 CSS 웹 페이지의 형식을 지정하기 위한 CSS 파일을 지정할 수 있습니다 . CSS 파일은 _css 라는 하위 폴더에 복사됩니다 . 보고서와 지원 파일을 다른 사용자에게 전송하면서 CSS 형식 지정을 유지할 수 있도록 CSS 파일에 대한 링크는 상대적입니다 .

색상 테마 웹 페이지 , 머리글 및 테두리의 색상을 지정합니다 . 기본 웹 페이지의 배경은 흰색 , 머리글은 주황색 , 테두리는 파란색입니다 .

글꼴 변경 보고서에 적용된 글꼴을 변경합니다 .

로고 변경 보고서와 함께 표시할 이미지를 지정합니다 . 이미지를 보고서의 위 또는 아래로 이동하려면 이미지 옆의 위쪽 또는 아래쪽 화살표를 클릭합니다 .

날짜 / 시간 표시 웹 페이지가 생성된 날짜 및 시간을 표시하거나 숨깁니다 .

보고서를 PowerPoint 프레젠테이션으로 저장

JMP 결과를 PowerPoint 프레젠테이션 (.pptx) 으로 저장하여 프레젠테이션을 생성할 수 있습니다 . .pptx 파일로 저장한 후에는 PowerPoint 에서 JMP 내용을 재배열하고 텍스트를 편집합니다 . JMP 보고서를 PowerPoint 로 내보낼 경우 다음과 같이 섹션이 달라집니다 .

- 보고서 머리글은 편집 가능한 텍스트 상자로 내보내집니다 .
- 그래프는 이미지로 내보내집니다 . 범례와 같은 특정 그래픽 요소는 별도의 이미지로 내보내집니다 . PowerPoint 의 슬라이드에 맞게 이미지의 크기가 조정됩니다 .

프레젠테이션에 저장할 섹션을 선택하려면 선택 도구를 사용합니다 . PowerPoint 에서 파일을 연 후에는 원치 않는 콘텐츠를 삭제하십시오 .

참고 : Windows 에서 JMP 를 통해 생성한 .pptx 파일을 열려면 PowerPoint 2007 이상이 필요합니다 . macOS 에서는 PowerPoint 2011 이상이 필요합니다 .

1. JMP 에서 보고서를 생성합니다 .
2. **파일 > 내보내기**를 선택하고 **Microsoft PowerPoint** 를 선택한 후 **다음**을 클릭합니다 .
3. 목록에서 그래픽 파일 형식을 선택합니다 .

   Windows 에서는 EMF 가 기본 형식입니다 . macOS 에서는 PDF 가 기본 형식입니다 .
4. 파일 이름을 지정하고 파일을 저장합니다 . macOS 에서는 파일 이름을 지정하고 **내보내기**를 클릭합니다 .

   저장 후 파일 열기가 기본적으로 선택되어 있으므로 파일이 Microsoft PowerPoint 에서 열립니다 .

참고 : Windows 에서 생성된 기본 EMF 그래픽은 macOS 에서 지원되지 않습니다 . macOS 에서 생성된 기본 PDF 그래픽은 Windows 에서 지원되지 않습니다 . 플랫폼 간 호환성을 위해서는 **파일 > 환경 설정 > 일반**을 선택하여 기본 그래픽 파일 형식을 변경하십시오 . 그런 다음 **PowerPoint 용 이미지 형식**을 PNG 또는 JPEG 로 변경하십시오 .

대시보드 생성

대시보드는 정기적으로 보고서를 실행하고 표시할 수 있는 시각적 도구입니다. 보고서, 데이터 필터, 선택 필터, 데이터 테이블 및 그래픽을 대시보드에 표시할 수 있습니다. 대시보드를 열면 대시보드에 표시된 내용이 업데이트됩니다.

이 섹션에는 다음과 같은 정보가 포함되어 있습니다.

- " 예 : 창 결합 "
- " 예 : 두 개의 보고서가 포함된 대시보드 생성 "

예 : 창 결합

JMP 에서 열려 있는 여러 창을 병합하여 신속하게 대시보드를 생성할 수 있습니다 . 창을 결합하면 통계량 요약을 볼 수 있는 옵션이 제공되고 선택 필터가 포함됩니다 .

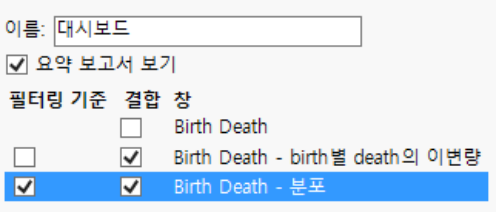
1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Birth Death.jmp 를 엽니다 .
2. "Distribution" 및 "Bivariate" 테이블 스크립트를 실행합니다 .
3. 보고서 창 중 하나에서 **창 > 창 결합**을 선택합니다 .

   " 창 결합 " 창이 나타납니다 .

   팁 : Windows 에서는 JMP 창의 오른쪽 하단에 있는 " 배열 메뉴 " 옵션에서 " 창 결합 " 을 선택할 수도 있습니다 .

4. **요약 보고서 보기**를 선택하여 그래프를 표시하고 통계 보고서를 생략합니다 .
5. " 결합 " 열에서 **Birth Death - birth 별 death 의 이변량** 및 **Birth Death - 분포**를 선택합니다 .
6. " 필터링 기준 " 열에서 **Birth Death - 분포**를 선택합니다 .

그림 7.18 창 결합 옵션



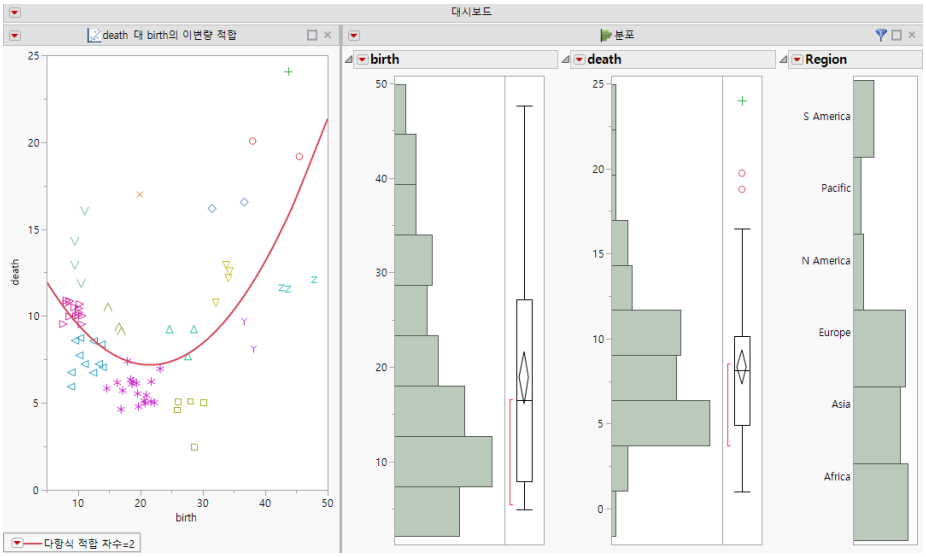
7. 확인을 클릭합니다 .

   두 개의 보고서가 하나의 창으로 결합됩니다 . " 분포 " 보고서 상단의 필터 아이콘 을 확인합니다 . 히스토그램 중 하나에서 막대를 선택하면 이변량 그래프의 해당 데이터가 선택됩니다 .

참고 :

- Windows 에서 보고서를 결합하려면 JMP 창의 오른쪽 하단에 있는 " 배열 메뉴 " 옵션에서 창 결합을 선택해도 됩니다 .
- 보고서에 그래프만 표시하고 통계량은 생략하려면 " 창 결합 " 창에서 **요약 보고서 보기**를 선택하십시오 .

그림 7.19 결합된 창



예 : 두 개의 보고서가 포함된 대시보드 생성

두 개의 보고서를 생성했고 다음날 업데이트된 데이터 집합에 대해 보고서를 다시 실행하려고 한다고 가정해 보겠습니다 . 이 예에서는 대시보드 빌더의 보고서에서 대시보드를 생성하는 방법을 보여 줍니다 .

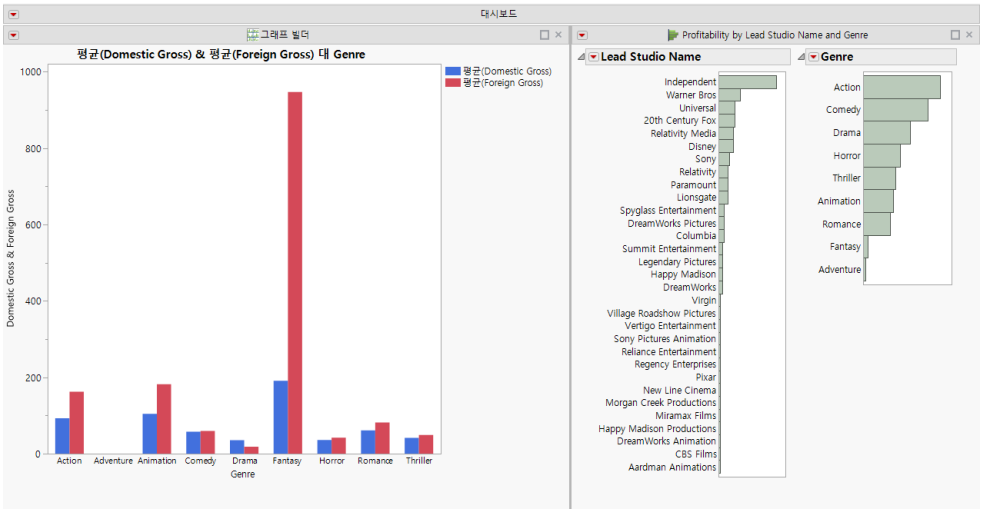
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hollywood Movies.jmp 를 엽니다 .
2. "Distribution: Profitability by Lead Studio and Genre" 및 "Graph Builder: World and Domestic Gross by Genre" 라는 테이블 스크립트를 실행합니다 .
3. 어느 창에서든 **파일 > 새로 만들기 > 대시보드**를 선택합니다 .

   일반적인 레이아웃을 위한 템플릿이 나타납니다 .
4. **2x1 대시보드** 템플릿을 선택합니다 .

   두 개의 보고서를 위한 공간이 있는 상자가 작업 공간에 나타납니다 .
5. " 보고서 " 목록에서 보고서 썸네일을 두 번 클릭하여 대시보드에 배치합니다 .
6. " 대시보드 빌더 " 의 빨간색 삼각형을 클릭하고 **미리보기 모드**를 선택합니다 .

   대시보드의 미리보기가 나타납니다 . 그래프가 서로 연결되고 데이터 테이블과 연결되어 있습니다 . 또한 " 분포 " 및 " 그래프 빌더 " 플랫폼과 동일한 빨간색 삼각형 옵션이 있습니다 .
7. **미리보기 닫기**를 클릭합니다 .

그림 7.20 두 개의 보고서가 있는 대시보드



대시보드 생성에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

8 장

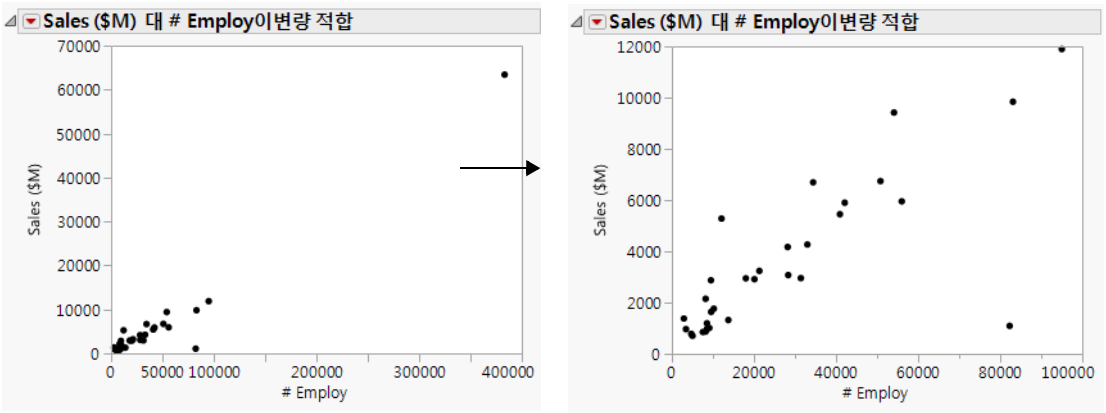
특수 기능

자동 분석 업데이트 및 SAS Integration

JMP의 일부 특수 기능을 사용하여 다음을 수행할 수 있습니다.

- 자동으로 분석 또는 그래프 업데이트
- 플랫폼 결과 사용자 정의
- SAS 와 통합하여 고급 분석 기능 사용

그림 8.1 특수 기능의 예




```
DATA Candy_Bars; INPUT  Calories Total_fat_g Carbohydrate_g Protein_g; Lines;
310  20  28  6
230  12  27  4
220  12  24  3
170  8  21  3
200  2.5  43  1
260  16  25  5
190  1.5  42  2
190  11  21  2
230  12  28  3
;
RUN;

PROC GLM  DATA=Candy_Bars ALPHA=0.05;
MODEL  Calories =  Total_fat_g Carbohydrate_g Protein_g;
RUN;
```

목차

분석 및 그래프 자동 업데이트 205
예: 자동 재계산 사용 205
환경 설정 변경 209
예: 환경 설정 변경 209
JMP 와 SAS 통합 212
예: SAS 코드 생성 212
예: SAS 코드 전송 213

분석 및 그래프 자동 업데이트

데이터 테이블을 변경할 때 자동 재계산 기능을 사용하여 데이터 테이블과 연결된 분석 및 그래프를 자동으로 업데이트할 수 있습니다. 예를 들어 데이터 테이블에서 값을 제외, 포함 또는 삭제하면 해당 변경 사항이 연결된 분석 또는 그래프에 즉시 반영됩니다. 다음 사항에 유의하십시오.

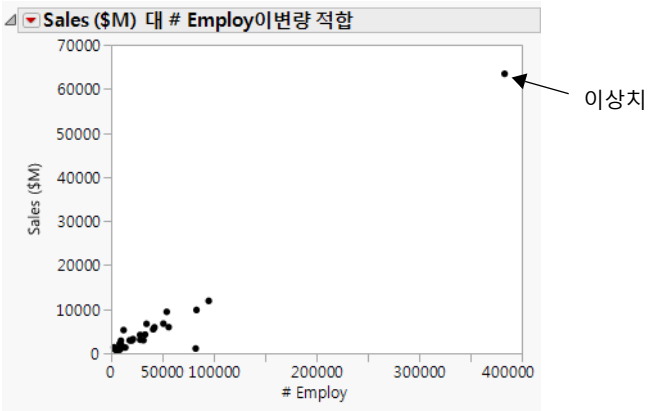
- 일부 플랫폼은 자동 재계산 기능을 지원하지 않습니다. 자세한 내용은 **JMP 사용**의 에서 확인하십시오.
- **분석** 메뉴의 지원되는 플랫폼에서는 자동 재계산 기능이 기본적으로 해제되어 있습니다. 그러나 **품질 및 공정** 메뉴에서 지원되는 플랫폼의 경우에는 계량형 / 계수형 게이지 차트, 공정 능력 및 관리도를 제외하고는 자동 재계산 기능이 기본적으로 설정되어 있습니다.
- **그래프** 메뉴에서 지원되는 플랫폼의 경우에는 자동 재계산 기능이 기본적으로 설정되어 있습니다.

예 : 자동 재계산 사용

이 예에서는 제약 및 컴퓨터 업계의 32 개 회사에 대한 재무 데이터가 포함된 Companies.jmp 샘플 데이터 테이블을 사용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다.
2. **분석** > **X 로 Y 적합**을 선택합니다.
3. Sales ($M) 를 선택하고 **Y, 반응**을 클릭합니다.
4. # Employ 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.

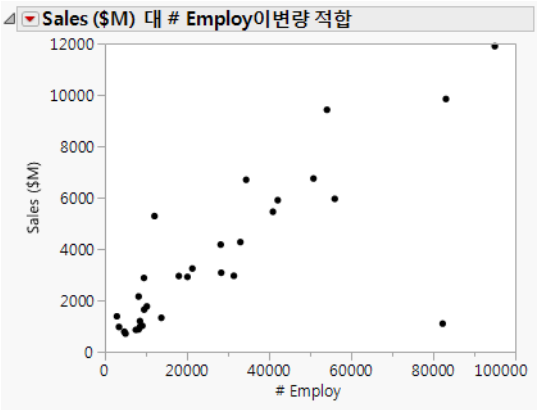
그림 8.2 초기 산점도



초기 산점도에서는 한 회사의 직원 수와 매출이 다른 모든 회사보다 훨씬 많음을 보여 줍니다. 이 회사가 이상치라고 판단하고 해당 데이터 점을 제외하고 싶습니다. 변경 시 산점도가 자동으로 업데이트되도록 하려면 해당 점을 제외하기 전에 자동 재계산 기능을 설정해야 합니다.

6. 자동 재계산 기능을 설정하려면 "Sales ($M) 대 # Employ 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **다시 실행** > **자동 재계산**을 선택합니다 .
7. 이상치를 클릭하여 선택합니다 .
8. **행** > **제외 / 제외 해제**를 선택합니다 . 해당 점이 분석에서 제외되고 산점도가 자동으로 업데이트됩니다 .

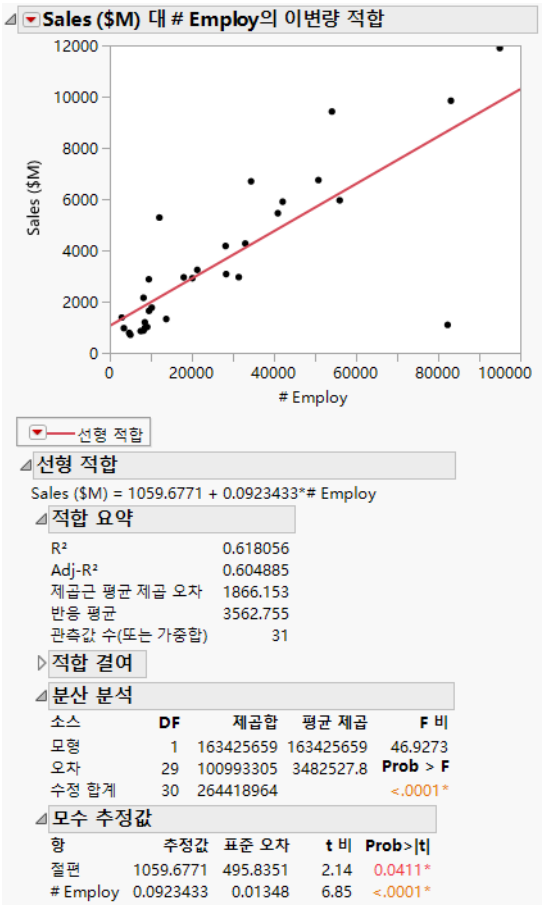
그림 8.3 업데이트된 산점도



데이터에 회귀선을 적합시킬 때 오른쪽 하단의 점은 이상치이며 선의 기울기에 영향을 줍니다 . 자동 재계산 기능이 설정된 상태에서 이상치를 제외하면 선의 기울기가 변경되는 것을 볼 수 있습니다 .

9. 회귀선을 적합시키려면 "Sales ($M) 대 # Employ 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 . 그림 8.4 에서는 보고서 창에 추가된 회귀선과 분석 결과를 보여 줍니다 .

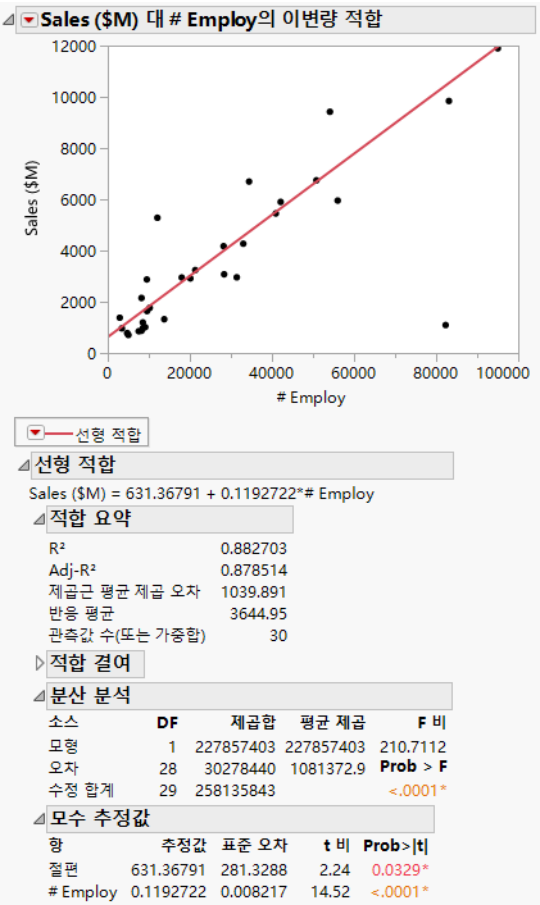
그림 8.4 회귀선 및 분석 결과



10. 이상치를 클릭하여 선택합니다 .
11. **행 > 제외 / 제외 해제**를 선택합니다 . 해당 점이 제외되었음을 반영하여 회귀선과 분석 결과가 자동으로 업데이트됩니다 .

팁 : 점을 제외하면 분석은 해당 데이터 점 없이 다시 계산되지만 해당 데이터 점이 산점도에서 숨겨지지는 않습니다 . 산점도에서도 해당 점을 숨기려면 점을 선택한 후 **행 > 숨기고 제외하기**를 선택하십시오 .

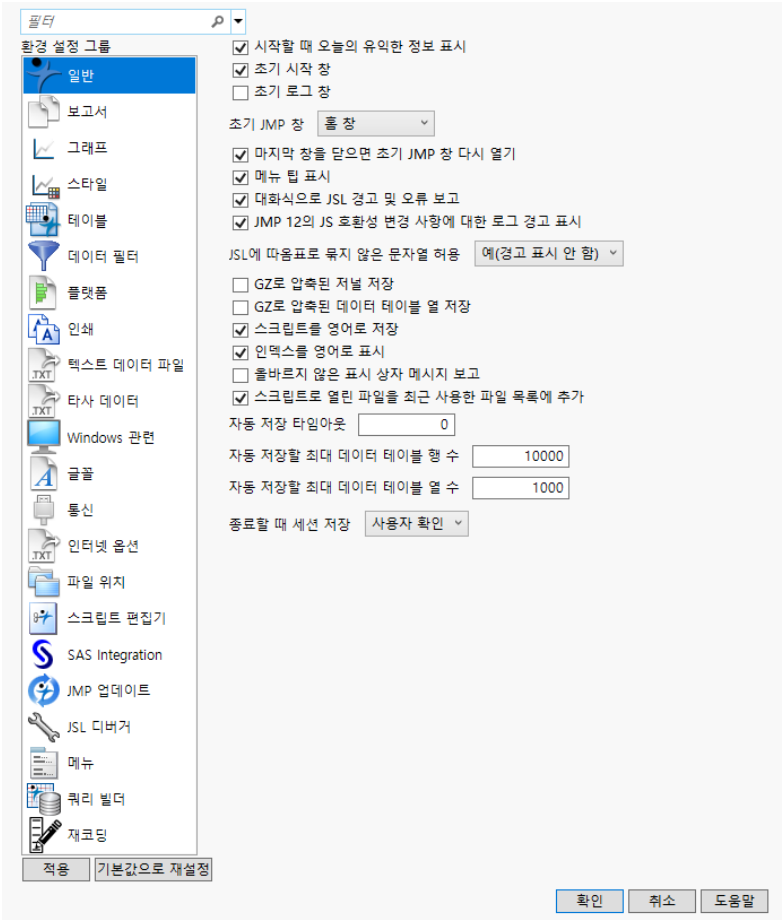
그림 8.5 업데이트된 회귀선 및 분석 결과



환경 설정 변경

" 환경 설정 " 창을 사용하여 JMP 의 환경 설정을 변경할 수 있습니다 . " 환경 설정 " 창을 열려면 **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 .

그림 8.6 환경 설정 창



"환경 설정" 창의 왼쪽에는 "환경 설정 그룹" 목록이 있습니다. 창의 오른쪽에는 선택한 범주에 대해 변경할 수 있는 모든 환경 설정이 있습니다.

예 : 환경 설정 변경

모든 플랫폼 보고서 창에는 설정하거나 해제할 수 있는 옵션이 있습니다. 그러나 이러한 옵션에 대한 변경 사항은 저장되지 않으므로 다음에 플랫폼을 사용할 때는 적용되지 않습니다. JMP 에

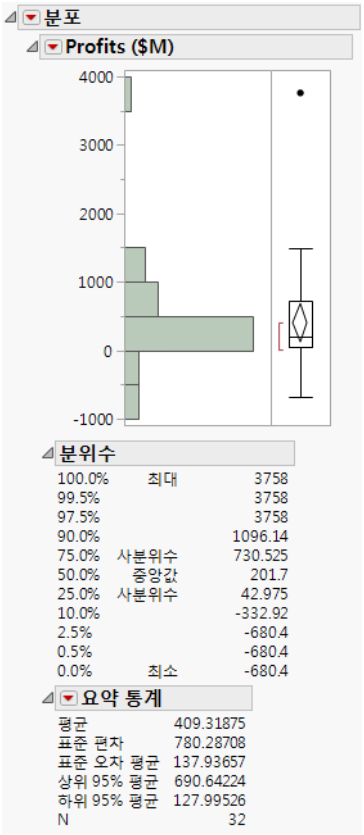
서 변경 사항을 저장하여 플랫폼을 사용할 때마다 적용하도록 하려면 "환경 설정" 창에서 해당 옵션을 변경하십시오.

이 예에서는 분포 플랫폼에서 이상치 상자 그림이 초기 보고서에 추가되지 않도록 설정하는 방법을 보여 줍니다.

기본 환경 설정을 사용하여 분포 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Profits ($M) 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

그림 8.7 분포 보고서 창

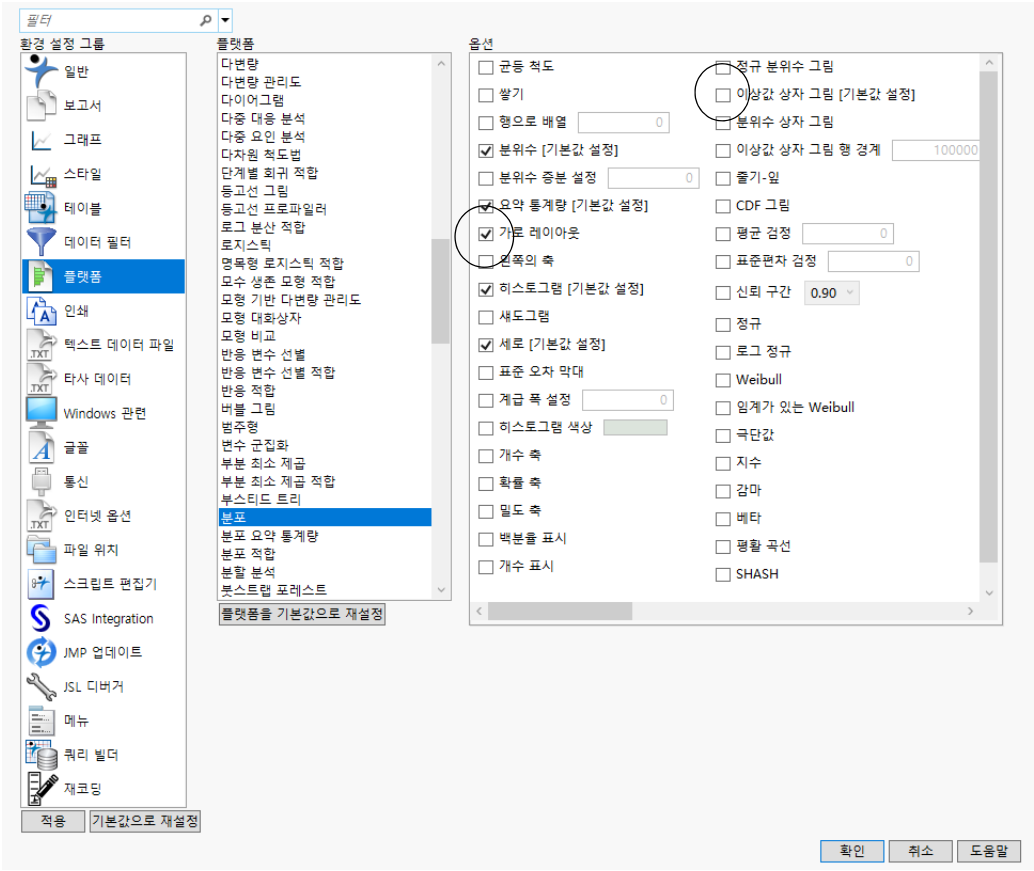


세로 히스토그램이 표시되며 그래프에는 이상치 상자 그림이 포함됩니다. 히스토그램을 가로로 변경하고 이상치 상자 그림을 제거하려면 "Profits ($M)"의 빨간색 삼각형 메뉴에서 적절한 옵션을 선택합니다. 그러나 플랫폼을 사용할 때마다 해당 환경 설정이 적용되도록 하려면 "환경 설정" 창에서 해당 설정을 변경하십시오.

이상치 상자 그림에 대한 환경 설정 변경 및 분포 재실행

1. **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 .
2. " 환경 설정 그룹 " 에서 **플랫폼**을 선택합니다 .
3. " 플랫폼 " 목록에서 **분포**를 선택합니다 .
4. **가로 레이아웃** 옵션을 선택하여 해당 기능을 설정합니다 .
5. **이상치 상자 그림** 옵션을 선택 취소하여 해당 기능을 해제합니다 .

그림 8.8 분포 환경 설정



6. **확인**을 클릭합니다 .
7. 분포 분석을 반복합니다 . 자세한 내용은 " 기본 환경 설정을 사용하여 분포 생성 "(210 페이지) 에서 확인하십시오 .

이제 가로 히스토그램이 표시되며 이상치 상자 그림은 표시되지 않습니다. 이러한 환경 설정은 변경하기 전까지 동일하게 유지됩니다.

모든 환경 설정에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

JMP 와 SAS 통합

참고 : JMP 를 통해 SAS 를 사용하려면 로컬 컴퓨터나 서버에서 SAS 에 액세스해야 합니다 .

JMP 를 사용하면 다음과 같은 방식으로 SAS 와 상호 작용할 수 있습니다 .

- JMP 에서 SAS 코드 작성 또는 생성
- JMP 에서 SAS 코드 전송 및 결과 보기
- 원격 컴퓨터에서 SAS Metadata 서버 또는 SAS 서버에 연결
- 로컬 컴퓨터에서 SAS 에 연결
- SAS 데이터 집합 열기 및 불러오기
- SAS 에서 생성된 데이터 집합 가져오기 및 보기

JMP 와 SAS 의 통합에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

예 : SAS 코드 생성

이 예에서는 초코바의 영양 데이터가 포함된 Candy Bars.jmp 샘플 데이터 테이블을 사용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Candy Bars.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Calories 를 선택하고 **Y** 를 클릭합니다 .
4. Total fat g, Carbohydrate g 및 Protein g 를 선택하고 **추가**를 클릭합니다 .
5. " 모형 규격 " 의 빨간색 삼각형을 클릭하고 **SAS 작업 생성**을 선택합니다 .

그림 8.9 에서는 SAS 코드를 보여 줍니다 . 모든 데이터가 표시되지는 않았습니다.

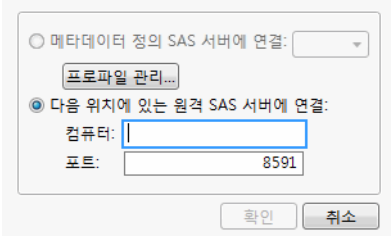
그림 8.9 SAS 코드

```
DATA Candy_Bars; INPUT  Calories Total_fat_g Carbohydrate_g Protein_g; Lines;
310  20  28  6
290  12  27  4
220  12  24  3
170  8  21  3
200  2.5  43  1
260  16  26  5
190  11.5  42  2
190  11  21  2
230  12  28  3
;
RUN;

PROC GLM  DATA=Candy_Bars ALPHA=0.05;
MODEL  Calories =  Total_fat_g Carbohydrate_g Protein_g;
RUN;
```

예 : SAS 코드 전송

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Candy Bars.jmp 를 엽니다 .
2. **분석** > **모형 적합**을 선택합니다 .
3. Calories 를 선택하고 **Y** 를 클릭합니다 .
4. Total fat g, Carbohydrate g 및 Protein g 를 선택하고 **추가**를 클릭합니다 .
5. " 모형 규격 " 의 빨간색 삼각형을 클릭하고 **SAS 로 전송**을 선택합니다 .
6. **SAS 서버에 연결** 창 (그림 8.10) 에서 SAS 에 연결할 방법을 선택합니다 (아직 연결되지 않은 경우). 이 예에서는 **이 컴퓨터의 SAS 에 연결**을 선택합니다 .

그림 8.10 SAS 서버에 연결



7. **확인**을 클릭합니다 .

JMP 가 SAS 에 연결됩니다 . SAS 가 모형을 실행하고 결과를 다시 JMP 로 보냅니다 . 결과는 SAS 출력 , HTML, RTF, PDF 또는 JMP 보고서 형식 (JMP 환경 설정에서 형식을 선택할 수 있음) 으로 나타날 수 있습니다 . 그림 8.11 에서는 JMP 보고서로 형식이 지정된 결과를 보여 줍니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

그림 8.11 JMP 보고서로 형식이 지정된 SAS 결과

SAS 시스템
The GLM Procedure

GLM 프로시저

Data

Number of Observations

Number of Observations Read	75
Number of Observations Used	75

Dependent Variable: Calories

Analysis of Variance

Calories

Overall ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	282358	94119.3	3237.58	<.0001*
Error	71	2064.03	29.0709	.	.
Corrected Total	74	284422	.	.	.

Fit Statistics

R-Square	Coeff Var	Root MSE	Calories Mean
0.99274	2.21858	5.39174	243.027

Type I Model ANOVA

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Total_fat_g	1	185260	185260	6372.68	<.0001*
Carbohydrate_g	1	93540.4	93540.4	3217.67	<.0001*
Protein_g	1	3557.86	3557.86	122.386	<.0001*

Type III Model ANOVA

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Total_fat_g	1	111777	111777	3844.97	<.0001*
Carbohydrate_g	1	96756.1	96756.1	3328.28	<.0001*
Protein_g	1	3557.86	3557.86	122.386	<.0001*

Solution

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-5.9643	2.89999	-2.0567	0.0434*
Total_fat_g	8.98995	0.14498	62.0078	<.0001*
Carbohydrate_g	4.0975	0.07102	57.6913	<.0001*
Protein_g	4.40133	0.39785	11.0628	<.0001*

부록 A

기술 공지사항

- Scintilla - Copyright © 1998-2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library - Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages - Copyright © 2009-2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software - Copyright © 1995-2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 996-2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999-2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998 - 2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball - Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako - Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 2006 –2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright (C) 1999-2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS@AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright(C) January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

버전 16

JMP 사용

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는 것이 아니라 새로운 눈을 가지는 데 있다."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.0

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP 사용® 16. Cary, NC: SAS Institute Inc.

JMP 사용® 16

Copyright © 2020-2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP를 처음 사용하든 오랫동안 사용해왔든, JMP와 관련해서 배워야 하는 것은 언제나 있습니다.

JMP.com을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다.

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티, 추가 기능 및 스크립트 예를 포함한 사용자용 리소스, 포럼, 블로그, 컨퍼런스 정보 등

https://www.jmp.com/getstarted

목차
JMP 사용

1 JMP 알아보기 19
설명서 및 추가 리소스
JMP 설명서에 사용되는 서식 규칙 21
JMP 도움말 22
JMP 설명서 라이브러리 22
JMP 학습을 위한 추가 리소스 28
JMP 자습서 28
샘플 데이터 테이블 28
통계 및 JSL 용어에 대해 알아보기 29
JMP 팁 및 힌트 알아보기 29
JMP 툴팁 29
JMP 사용자 커뮤니티 29
무료 온라인 Statistical Thinking 교육 과정 30
JMP 새로운 사용자를 위한 입문 키트 30
Statistics Knowledge 포털 30
JMP 교육 과정 30
사용자가 작성한 JMP 설명서 30
JMP 시작하기 창 30
JMP 기술 지원 31

2 JMP 시작하기 33
기본 기능 소개
JMP 세션의 구조 35
데이터 테이블 기능 36
데이터 테이블 패널 37
데이터 테이블의 데이터 격자 45
데이터 파일을 여는 옵션 50
플랫폼 및 보고서 51
플랫폼의 시작 창 52
보고서의 분석 결과 54
파일 및 열린 창 관리 55
JMP 홈 창 (Windows) 55

JMP 홈 창 (macOS) 59
창에서 최근에 열었던 파일 검색 60
여러 파일 닫기 61
열려 있는 창 표시 및 배열 61
JMP 파일 미리보기 62
창 나타내기 기능 사용 63
JMP 시작하기 창 64

3 데이터 가져오기 65

데이터 테이블 생성

데이터 가져오기 정보 68
Microsoft Excel 파일 가져오기 69
Microsoft Excel 데이터 미리 보기 및 가져오기 69
Microsoft Excel 파일 직접 가져오기 80
텍스트 파일 가져오기 81
가져온 데이터에서 결측 문자 찾기 83
JMP 텍스트 가져오기 마법사의 옵션 84
JMP 로 가져오기 전에 텍스트 편집 88
스크립트 편집기에서 텍스트 가져오기 90
여러 파일 가져오기 91
PDF 파일 가져오기 96
여러 테이블을 하나의 데이터 테이블로 연결 96
열 머리글을 기준으로 테이블 연결 98
JMP 가 자동 감지하지 않는 테이블 가져오기 99
PDF 미리보기의 빨간색 삼각형 옵션 102
데이터 테이블 미리보기 옵션 102
PDF 미리보기의 마우스 오른쪽 버튼 클릭 옵션 103
XML 및 JSON 파일 가져오기 103
XML 및 JSON 데이터 미리보기 및 가져오기 103
XML 파일 바로 가져오기 109
JSON 파일 바로 가져오기 109
쿼리 빌더에서 SQL 쿼리 작성 110
SQL 데이터베이스에 연결 111
SQL 데이터베이스에서 테이블 선택 112
SQL 쿼리 생성 116
JMP 쿼리 저장 및 실행 129
선택한 데이터 열기 130
쿼리 빌더의 빨간색 삼각형 메뉴 옵션 130
쿼리 빌더에서 SQL 문 작성 131
SAS 에서 데이터 가져오기 132

JRE(Java Runtime Environment) 요구 사항 ... 132
SAS 메뉴 옵션 ... 132
SAS 데이터 집합 가져오기 ... 132
JMP 에서 SAS 의 SAS 전송 파일 생성 ... 135
JMP 에서 SAS 에 연결 ... 135
SAS 쿼리 빌더를 사용하여 SAS 데이터 집합 열기 ... 141
SAS 서버를 통해 SAS 데이터 집합 열기 ... 142
SAS 확장 속성을 사용하여 메타데이터 가져오기 ... 149
스토어드 프로세스 실행 ... 150
SAS 코드 전송 ... 151
ODS 결과 생성 ... 153
생성된 SAS 데이터 집합 가져오기 ... 154
데이터베이스에서 데이터 가져오기 ... 156
데이터베이스에 연결 ... 157
데이터베이스에서 데이터 열기 ... 157
SQL 문을 작성하여 데이터베이스 쿼리 ... 160
SQL(Structured Query Language): 참조 ... 161
Google 스프레드시트에서 데이터 가져오기 ... 167
Google 스프레드시트 가져오기 ... 167
Google 스프레드시트 설정 ... 168
SPSS 파일 가져오기 ... 168
Triple-S 설문 조사 데이터 가져오기 ... 170
HDF5 데이터 가져오기 ... 170
MATLAB 에서 데이터 가져오기 ... 171
R 에서 데이터 가져오기 ... 171
Excel 추가기능을 사용하여 데이터 가져오기 ... 171
Excel 데이터를 데이터 테이블로 전송 ... 172
Excel 용 JMP 추가기능 ... 173
JMP Excel 추가기능 제거 ... 174
인터넷 또는 원격 컴퓨터에서 데이터 가져오기 ... 174
새 데이터 테이블 생성 ... 175

4 데이터 입력 및 편집 ... 177

기본적인 데이터 테이블 작업 수행

데이터 테이블에 데이터 입력 ... 180
데이터 테이블에 데이터 복사 및 붙여넣기 ... 180
데이터 테이블에 행 추가 ... 181
데이터 테이블에 열 추가 ... 181
순차적 데이터로 데이터 테이블 열 채우기 ... 183
결측값을 데이터 테이블 열의 이전 값으로 바꾸기 ... 184

데이터 테이블 열에 셀 계산식 입력 185
데이터 테이블에서 행 선택 186
행에서 매칭 셀 찾기 188
특정 값이 포함된 행 선택 188
다음 선택 행 찾기 190
데이터 테이블에서 열 선택 191
열을 선택하는 방법 191
JMP 열 보기에서 열 선택 192
데이터 테이블에서 다음 및 이전 선택 열 찾기 197
행 및 열 크기 조정 197
데이터 테이블에서 데이터 구성 198
데이터 테이블에서 행 삭제 198
데이터 테이블에서 열 삭제 199
데이터 테이블에서 열 재배열 199
데이터 테이블에서 열 그룹화 199
데이터 테이블에서 값 이동 200
데이터 테이블에서 셀에 색상 적용 201
데이터 테이블에서 셀 편집 또는 삭제 201
데이터 테이블에서 열 이름 편집 201
데이터 테이블에서 행 숨기고 제외하기 202
데이터 테이블에서 행 제외 203
데이터 테이블에서 행 숨기기 203
데이터 테이블에서 열 제외 204
데이터 테이블에서 열 숨기기 205
데이터 테이블에서 결측 데이터의 패턴 보기 206
데이터 테이블에서 셀 값 찾기 및 바꾸기 207
데이터 테이블에서 셀 편집 209
행 및 열에 대한 마우스 오른쪽 버튼 클릭 메뉴 211
데이터 테이블 비교 211
행 및 열에 특성 할당 216
행 및 열에 대한 옵션을 찾는 방법 216
행 및 열 라벨 지정 217
행에 색상 또는 표식 할당 218
그래프의 색상 테마 생성 220
그래프에서 사용자 색상 테마 삭제 224
행 상태가 매칭되는 셀 선택 225
행 특성 삭제 225
열 잠그기 225
데이터 재구성 226
데이터 테이블 열을 여러 열로 분할 226

텍스트 매칭으로 새 열 생성 227
표시자 열 생성 228
데이터 테이블에서 열 결합 229
데이터 테이블에서 선택 열 압축 229
데이터 테이블 열에서 계급화 계산식 만들기 231
라벨과 코드 간 변환 235
새 계산식 열 생성 236
JMP 플랫폼의 열 변환 236
열의 데이터 재코딩 245
데이터 테이블 편집 259
데이터 테이블 이름 바꾸기 259
Lock Data Table 259
데이터 테이블 압축 260
테이블 스크립트 복사 260
데이터 테이블 변수 사용 261
데이터 테이블에서 스크립트 생성 및 저장 262

5 JMP 열 특성 설정 267
열 정보 창
열 정보 창 269
모델링 유형 271
짧은 정수 형식 272
숫자 형식 273
행 상태 열에 정보 저장 277
열의 데이터 초기화 279
JMP 의 열 특성 280
기본 열 특성 281
열 값을 검증하는 특성 282
열 값에 정보를 추가하는 특성 283
열 표시 방식을 제어하는 특성 288
모델링 및 DOE 에 사용되는 특성 289
관리도 및 공정 능력과 관련된 특성 290
플랫폼에서 열이 사용되는 방식을 제어하는 특성 291
SAS 특성 299
가상 결합 특성을 사용하여 데이터 테이블 연결 299
추가 열 특성 301
JMP 에서 할당 및 제어되는 특성 303
열 간에 속성 및 특성 표준화 305
열 값 재코딩 305
열 속성 표준화 306

열 특성 표준화 306
열 특성 삭제 306
예 : 계산식 표준화 306
열에 사전 선택된 분석 역할 할당 308

6 데이터 재구성 309
부분집합 생성 , 데이터 정렬 등

부분집합 데이터 테이블 생성 311
층화 부분집합 313
보고서에서 부분집합 데이터 테이블 생성 313
데이터 테이블 정렬 314
열 쌓기 317
예 : 열을 단일 열로 쌓기 318
예 : 열을 둘 이상의 열로 쌓기 319
열 분할 321
예 : 열 분할 322
행 및 열 전치 324
예 : 행 및 열 전치 326
데이터 테이블 연결 328
예 : 데이터 테이블 연결 329
예 : 데이터 테이블 및 테이블 변수 연결 330
데이터 테이블 결합 331
예 : 데이터 테이블 결합 333
JMP 쿼리 빌더를 사용하여 데이터 테이블 쿼리 및 결합 341
쿼리의 데이터 테이블 링크 정보 343
데이터 테이블 가상 결합 344
예 : 열 가상 결합 345
예 : 열 가상 결합 및 행 상태 표시 348
예 : 가상으로 결합된 데이터 테이블에서 참조된 데이터 병합 349
단일 열의 연결 ID 및 연결 참조 예 351
데이터 테이블 업데이트 354
예 : 데이터 테이블 업데이트 356
데이터 익명화 357

7 JMP 의 계산식 생성 359
계산식 편집기

계산식 개요 361
계산식 작성 361
계산식에서 데이터 테이블 값 참조 363
계산식에서 지역 변수 사용 365

계산식에서 모수 포함 367
계산식에서 상수 삽입 368
JMP 계산식에 연산자 추가 369
계산식 편집기의 키패드 참조 369
계산식에서 함수 사용 371
계산식 편집기에서 자주 사용하는 함수 찾기 374
계산식 복사 및 반복 375
계산식의 표현식 정렬 377
우선 순위대로 계산식 작성 378
JMP 계산식 편집기 옵션 379
JMP 함수의 도함수 계산 379
복잡한 계산식 단순화 380
계산식 실행 381
계산식에서 오류 무시 381
JSL 에서 계산식 보기 382
계산식 편집 382
계산식에서 오류 수정 382
계산식에서 표현식 선택 383
함수 계산식 삭제 383
계산식 잘라내기 , 복사 및 붙여넣기 383
계산식 클릭 및 드래그 384
JMP 계산식 편집기에서 계산식 사용자 정의 384
계산식에서 개요 숨기기 및 표시 385
계산식에서 글꼴 크기 변경 385
계산식에서 행렬 크기 지정 385
JMP 계산식의 방향 변경 385
계산식에서 인수 열기 및 닫기 386
JMP 계산식 편집기의 계산식 작성 예 386
JMP 계산식 편집기에서 기본 기능 사용 386
계산식에서 지역 변수 사용 388
계산식에서 Munger 함수 사용 388
계산식에서 Match 조건 함수 사용 390
계산식에서 표현식 삭제 버튼 사용 390
JMP 계산식 편집기의 바로 가기 키 391
JMP 계산식 편집기의 용어 392

8 데이터 요약 395

테이블 요약 명령

열 요약 397
요약 시작 창 397

요약 테이블 생성 397
기존 요약 테이블에 통계량 열 추가 401
요약 통계량에 대한 설명 402
예 : 그룹을 사용하여 요약 테이블 생성 403

9 JMP 보고서 405
보고서 창 탐색 및 사용자 정의

JMP 보고서 탐색 408
손 도구 사용 409
JMP 보고서의 표시 옵션 409
JMP 보고서 요소 표시 및 숨기기 411
여러 보고서 결합 411
보고서의 제목 바꾸기 412
보고서의 글꼴 크기 늘리기 412
보고서 창 크기 조정 412
JMP 보고서의 빨간색 삼각형 옵션 412
보고서의 로컬 데이터 필터 413
보고서의 다시 실행 메뉴 416
보고서의 스크립트 저장 메뉴 416
데이터 필터 419
필터 열의 유형 420
필터링 모드 421
데이터 필터 제어판 422
데이터 필터의 빨간색 삼각형 메뉴 옵션 424
데이터 필터의 애니메이션 기능 427
데이터 필터의 변수에 대한 빨간색 삼각형 메뉴 옵션 429
보고서의 테이블 형식 조정 432
JMP 보고서의 표시 상자 특성 수정 434
JMP 보고서의 테이블을 행렬로 변환 435
보고서에서 조건부 형식 사용 435
그래프에서 점 선택 441
그래프에서 행 선택 441
JMP 그래프에서 직사각형 영역으로 점 선택 442
JMP 그래프에서 불규칙한 형태의 영역 내에 있는 점 선택 443
그래프에서 표식 사용 444
표식 형태 변경 444
표식 색상 변경 445
표식 크기 변경 445
많은 수의 표식 처리 445
표식 주위에 외곽선 추가 446

표식 선택 모드 변경 446
표식 투명도 지정 447
표식 제외 및 숨기기 448
이미지를 표식으로 사용 448
값을 기반으로 표식 형태 또는 색상 변경 449
그래프 모양 변경 451
그래프 크기 조정 452
그래프의 확대 453
그래프의 선 너비 변경 453
그래프의 배경 색상 변경 453
히스토그램 막대의 색상 변경 454
그래프의 좌표 및 임시 참조선 표시 454
그래프의 축 스크롤 및 척도 조정 454
그래프의 축 및 축 라벨 사용자 정의 455
보고서의 값 순서 변경 465
그래프에서 선택한 개체의 패턴 및 형식 변경 466
데이터 테이블에 그래프 추가 466
그래프의 표식, 선, 텍스트 등 사용자 정의 467
JMP 그래프 또는 보고서에 이미지 추가 473
JMP 그래프에 배경 이미지 붙여넣기 473
JMP 보고서 끝에 이미지 붙여넣기 473
이미지를 JMP 그래프로 드래그하여 놓기 474
JMP 그래프에 지리 이미지 및 경계 추가 476
이미지에서 데이터 추출 476
JMP 보고서 및 그래프에 주석 및 형태 추가 477
JMP 보고서 및 그래프에 주석 추가 477
JMP 보고서에 형태 추가 479
그래프의 가리키기 라벨 사용자 정의 482
가리키기 라벨에 그래프 또는 이미지 추가 483
그래프의 가리키기 라벨에 서식 있는 텍스트 추가 498
가리키기 라벨의 이름 및 값 업데이트 501
그래프의 가리키기 라벨 편집 511
데이터 테이블에서 가리키기 라벨에 이미지 추가 516
그래프의 가리키기 라벨 고정 517
가리키기 라벨의 마우스 오른쪽 버튼 클릭 옵션 518

10 데이터 저장 및 공유 521
JMP 데이터 활용
데이터 테이블 저장 및 공유 524
데이터 테이블을 Microsoft Excel 파일로 저장 525

JMP 데이터를 CSV 파일로 저장 526
JMP 데이터를 텍스트 파일로 저장 527
JMP 데이터를 JSON 파일로 저장 528
JMP 데이터를 SAS 전송 파일로 저장 528
JMP 데이터를 Google 스프레드시트로 저장 529
JMP 데이터를 SAS 데이터 집합으로 저장 531
데이터 테이블을 데이터베이스에 저장 531
데이터베이스를 데이터 테이블로 바꾸기 533
보고서 저장 및 공유 533
보고서를 대화식 HTML 로 공유 535
JMP Live 또는 JMP Public 에 보고서 게시 539
보고서를 PDF 파일로 저장 548
보고서를 PowerPoint 프레젠테이션으로 저장 549
내보낸 그래픽에 대한 그래픽 DPI 설정 550
JMP 분석을 스크립트로 저장 551
보고서 요소를 그래픽 형식으로 저장 551
보고서 인쇄 552
보고서 복사 및 붙여넣기 553
프로젝트 작업 554
새 프로젝트 생성 555
프로젝트에서 파일 열기 555
프로젝트에서 파일 재배열 556
프로젝트 저장 557
프로젝트 작업 영역 558
프로젝트 책갈피 558
프로젝트 내용 559
프로젝트의 최근 사용한 파일 560
프로젝트 로그 560
파일을 프로젝트로 이동 561
프로젝트 내의 열기 스크립트 작성 561
새 프로젝트 생성의 예 562
저널 생성 565
예 : 프레젠테이션용 저널 생성 566
새 저널 생성 568
공유 저널의 링크 관리 569
저널에 개요 추가 및 편집 570
저널의 마우스 오른쪽 버튼 클릭 옵션 571
저널에서 개요 수준 표시 제어 571
저널에 그래프 또는 그래픽 추가 572
저널의 사용자 정의 항목 572

저널 수정 방지 573
저널에 보고서 추가 573
다른 형식으로 저널 저장 573
저널에서 항목 삭제 574
JMP 세션 저장 574
JMP 종료 시 세션 저장 574
수동으로 세션 저장 575
로그 창 저장 576

11 대시보드를 사용하여 보고서 결합 및 표시 577
JMP 확장
대시보드를 생성하여 JMP 보고서 결합 579
예 : 두 개의 JMP 보고서가 포함된 대시보드 생성 579
대시보드 빌더 사용 580
대시보드 빌더의 빨간색 삼각형 옵션 581
샘플 대시보드를 사용하여 시작 582
대시보드 편집 582
실행 중인 대시보드 보기 586
예 : 창을 결합하여 대시보드 생성 590
예 : 대시보드에 선택 필터 추가 591
예 : JMP 쿼리 대시보드 및 추가기능 생성 593
예 : 두 개의 데이터 테이블에서 대시보드 생성 595

12 메뉴 및 도구 모음 개인 설정 597
JMP 사용자 정의
Windows 에서 JMP 도구 모음 및 메뉴 개인 설정 599
도구 모음 개인 설정 600
메인 메뉴 막대 개인 설정 603
메뉴 항목 및 도구 모음 버튼 생성 604
사용자 정의된 도구 모음 및 버튼에 대한 사용자 권한 설정 607
도구 모음 재배열 608
메뉴 , 메뉴 항목 , 도구 모음 및 버튼 복사하여 붙여넣기 609
사용자 메뉴 , 메뉴 항목 및 버튼 재배열 610
사용자 항목 삭제 612
사용자 정의 항목 표시 또는 숨기기 613
사용자 정의 가져오기 614
JMP 에서 사용자 정의 제거 614
macOS 에서 JMP 도구 모음 개인 설정 615
macOS 에서 JMP 메뉴 개인 설정 617
추가기능에 대한 JMP 메뉴 및 도구 모음 사용자 정의 618

13 JMP 환경 설정 619

환경 설정 창

JMP 환경 설정 개요 621
일반 JMP 환경 설정 621
JMP 프로젝트에 대한 환경 설정 625
JMP 보고서의 환경 설정 626
JMP 그래프의 환경 설정 628
JMP 그래프 및 보고서의 스타일 환경 설정 631
데이터 테이블의 환경 설정 634
JMP 데이터 필터의 환경 설정 638
JMP 플랫폼의 환경 설정 639
인쇄 환경 설정 641
텍스트 파일 가져오기 및 내보내기의 환경 설정 642
타사 데이터의 환경 설정 647
Windows 환경 설정 649
Mac OS 환경 설정 653
글꼴 환경 설정 655
통신 환경 설정 657
인터넷 옵션의 환경 설정 657
파일 위치의 환경 설정 659
JSL 스크립트 편집기의 환경 설정 660
SAS Integration 의 환경 설정 663
JMP 업데이트의 환경 설정 666
JSL 디버거의 환경 설정 667
메뉴 환경 설정 668
쿼리 빌더의 환경 설정 669
재코딩 열의 환경 설정 670
열 전환기의 환경 설정 672
로그의 환경 설정 673

A 계산식에서 JMP 함수에 대한 참조 675

계산식 편집기의 함수에 대한 설명

행 함수 677
숫자 함수 679
초월함수 680
삼각함수 683
문자 함수 684
비교 함수 689
조건 함수 690
확률 함수 695

이산 확률 함수 704
통계 함수 706
난수 함수 710
날짜 / 시간 함수 715
행 상태 함수 718
할당 함수 723
모수 모형 함수 724
금융 함수 724

B JMP 기술 상세 정보 727

멀티스레딩을 지원하는 JMP 기능 729
JMP 속성을 SAS 확장 속성에 매핑하기 위한 규칙 730

C 기술 라이선스 표시 737

1 장

JMP 알아보기
설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 21
JMP 도움말 22
JMP 설명서 라이브러리 22
JMP 학습을 위한 추가 리소스 28
JMP 자습서 28
샘플 데이터 테이블 28
통계 및 JSL 용어에 대해 알아보기 29
JMP 팁 및 힌트 알아보기 29
JMP 툴팁 29
JMP 사용자 커뮤니티 29
무료 온라인 Statistical Thinking 교육 과정 30
JMP 새로운 사용자를 위한 입문 키트 30
Statistics Knowledge 포털 30
JMP 교육 과정 30
사용자가 작성한 JMP 설명서 30
JMP 시작하기 창 30
JMP 기술 지원 31

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 Lucida Sans Typewriter 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 "JMP 설명서 라이브러리 "(22 페이지) 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
Quality and Process Methods	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기" 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기

" 창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

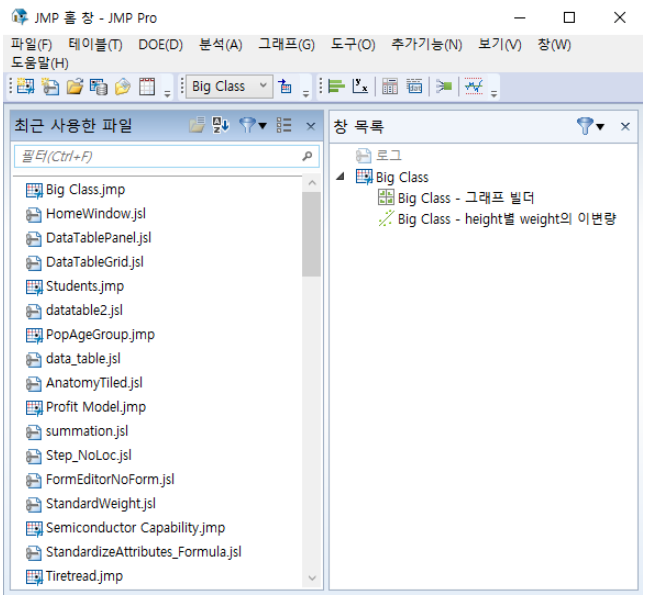
2 장

JMP 시작하기
기본 기능 소개

JMP를 처음 사용하는 사용자를 위해 이 장에서는 다음과 같은 주제를 다룹니다.

- JMP 시작 시 나타나는 초기 창 알아보기
- 데이터 테이블 이해
- 데이터 파일 열기
- 열려 있는 창 관리
- 일반적인 JMP 사용자 세션의 구조 알아보기

그림 2.1 JMP 홈 창 (Windows)



목차

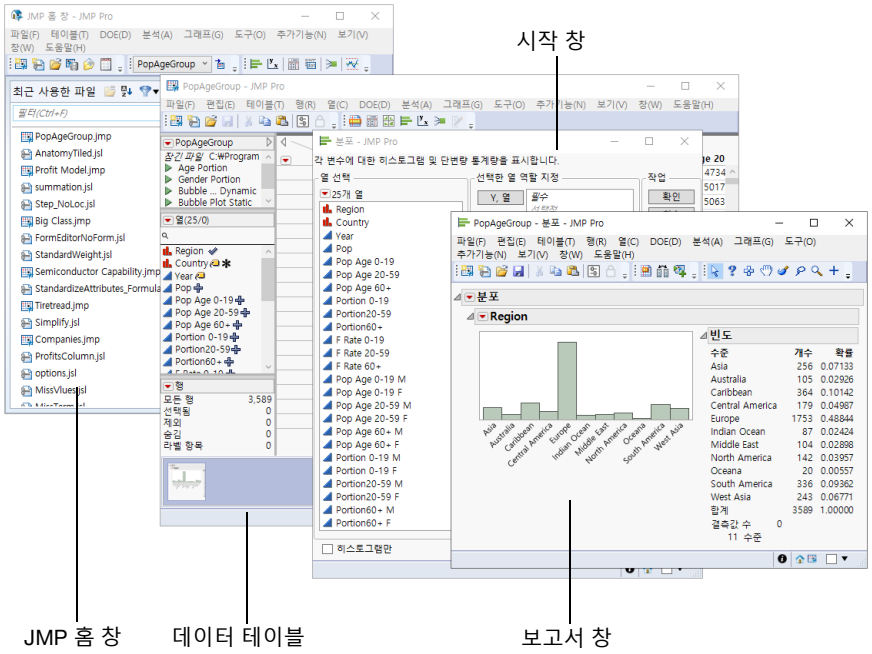
JMP 세션의 구조 35
데이터 테이블 기능 36
데이터 테이블 패널 37
데이터 테이블의 데이터 격자 45
데이터 파일을 여는 옵션 50
플랫폼 및 보고서 51
플랫폼의 시작 창 52
보고서의 분석 결과 54
파일 및 열린 창 관리 55
JMP 홈 창(Windows) 55
JMP 홈 창(macOS) 59
창에서 최근에 열었던 파일 검색 60
여러 파일 닫기 61
열려 있는 창 표시 및 배열 61
JMP 파일 미리보기 62
창 나타내기 기능 사용 63
JMP 시작하기 창 64

JMP 세션의 구조

JMP는 통계 분석 유형별로 체계화된 여러 플랫폼으로 구성되어 있습니다. 예를 들어 분포 플랫폼에서는 히스토그램, 추가 그래프 및 보고서를 사용한 단변량 분석을 생성합니다. 분포 플랫폼에서 데이터를 분석한 후 다른 플랫폼을 선택하여 데이터를 보다 철저하게 검토할 수 있습니다. 그 결과, 데이터 테이블, 플랫폼 시작 창 및 분석 결과를 기본으로 여러 개의 창이 동시에 열려 있게 됩니다.

그림 2.2에서는 일반적인 JMP 세션을 보여 줍니다. 창 사이를 빠르게 전환할 수 있도록 창은 최대화되어 있지 않습니다.

그림 2.2 일반적인 JMP 세션의 구조 (Windows)



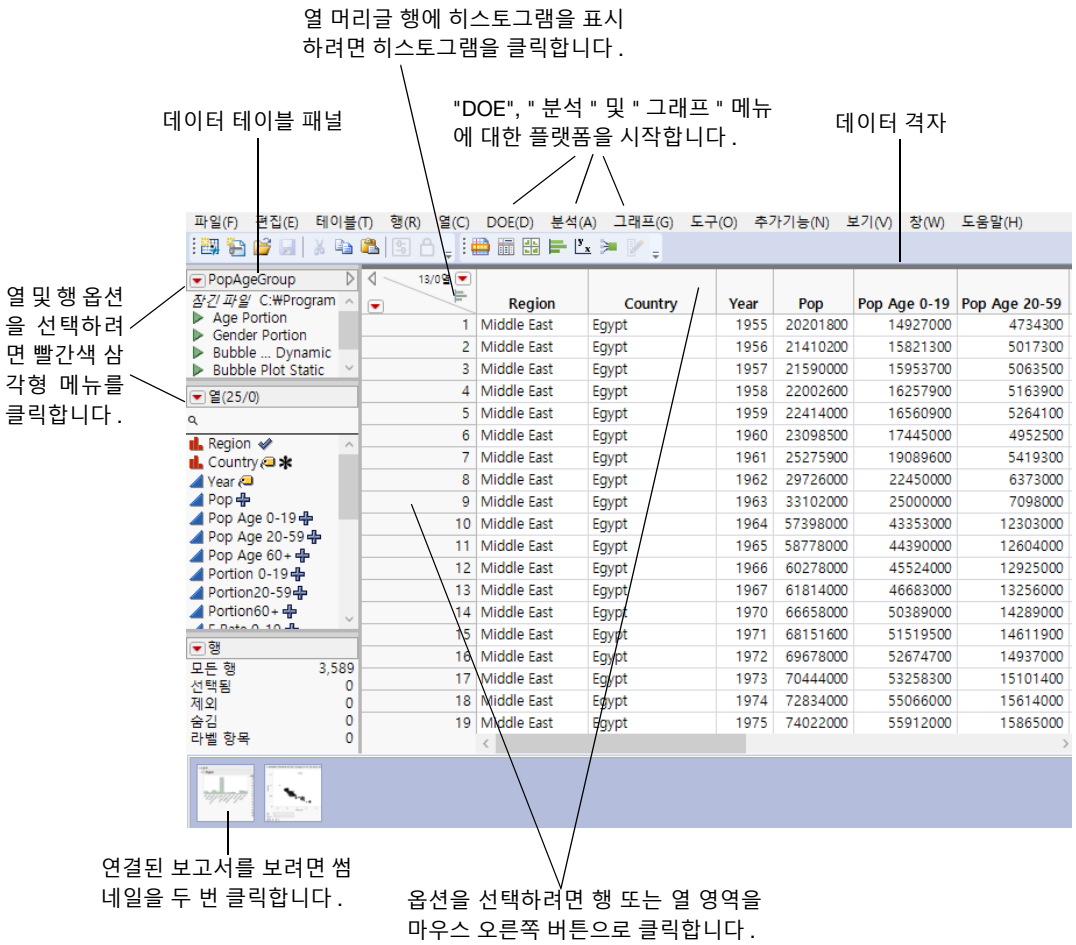
다음 섹션에서는 일반적으로 사용하는 창에 대해 설명합니다.

데이터 테이블 기능

JMP에서 데이터 점은 **데이터 테이블**이라고 하는 행 및 열 구조로 구성됩니다. 데이터 테이블은 왼쪽의 **데이터 테이블 패널**과 오른쪽의 **데이터 격자**로 이루어져 있습니다.

데이터 테이블을 사용하여 데이터를 입력, 확인, 편집 및 조작할 수 있습니다. 데이터 테이블에서 각 변수는 열이 되고 각 관측값은 행이 됩니다.

그림 2.3 데이터 테이블 기능



데이터 테이블에는 다음과 같은 특성이 있습니다.

- 열 이름에는 공백을 포함한 모든 키보드 문자가 포함될 수 있으며, 길이는 최대 255 자일 수 있습니다.
- 데이터 테이블 이름의 최대 길이는 사용하는 컴퓨터의 운영 체제에 따라 다릅니다.

- " 일반 " 환경 설정에서 " 자동 저장 타임아웃 " 값을 설정하여 데이터 테이블을 지정된 시간 (분) 마다 자동으로 저장하는 것이 좋습니다 . 이 자동 저장 값은 저널 , 스크립트 , 프로젝트 및 보고서에도 적용됩니다 .
- **파일 > 환경 설정 > 글꼴**을 선택하여 이름과 값의 기본 크기 및 글꼴을 변경할 수 있습니다 . macOS 의 경우에는 **JMP > 환경 설정 > 글꼴**을 선택합니다 .
- 열 이름 영역에서는 지정된 열 너비에 맞게 열 이름에 자동 줄바꿈이 적용됩니다 .
- 긴 값을 보려면 열 경계를 이동하고 열을 확장합니다 . 선택한 모든 열의 너비를 한 번에 조정하려면 Alt 키를 누른 채로 선택한 열 중 하나의 경계에 있는 양방향 화살표 커서를 드래그합니다 .
- 데이터 테이블의 행 및 열 수는 컴퓨터 메모리에 따라서만 제한됩니다 .
- 행 크기를 조정하려면 행 테두리 중 하나를 드래그합니다 . 그러면 모든 행이 동일한 높이로 조정됩니다 . 각 셀 안에 표시되는 그래픽은 행 높이에 따라 축소됩니다 .

  열을 동일한 너비로 조정하려면 열을 선택한 후 Alt(Windows) 또는 Option(macOS) 을 누른 채로 열 테두리 중 하나를 드래그합니다 . 그러면 모든 열이 동일한 너비로 조정됩니다 .

팁 : Windows 에서 데이터 테이블의 보고서 썸네일을 닫으려면 마우스 가운데 버튼을 클릭하십시오 .

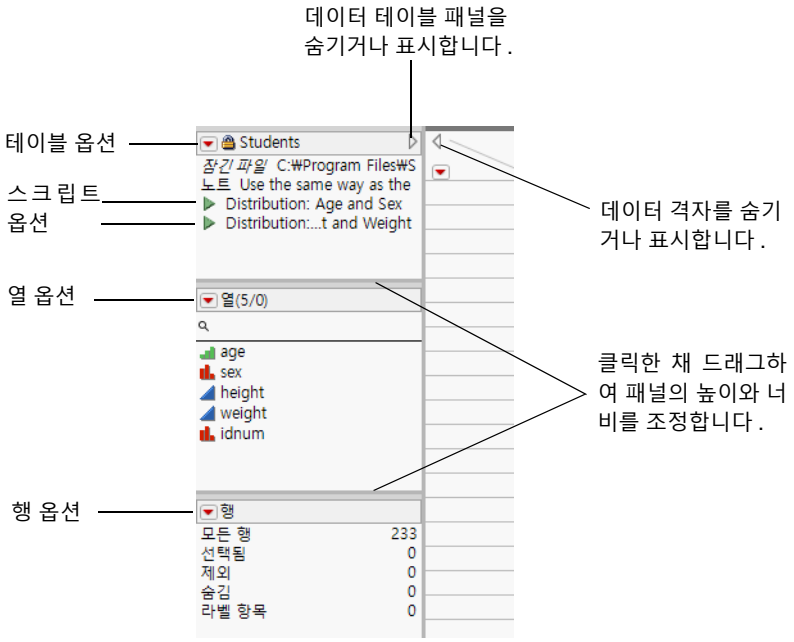
데이터 테이블 패널

데이터 테이블에는 다음과 같은 세 개의 패널이 포함되어 있습니다.

- 테이블 패널
- 열 패널
- 행 패널

이러한 패널은 데이터 격자의 왼쪽에 있습니다. 여기에는 테이블과 해당 내용에 대한 정보가 포함됩니다. 각 패널에는 대화식 영역이 있습니다.

그림 2.4 데이터 테이블 패널에서의 상호 작용

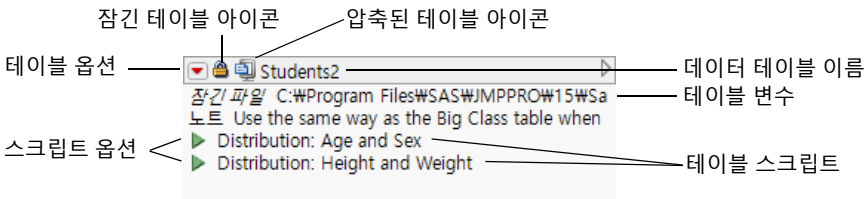


데이터 테이블의 테이블 패널

테이블 패널에는 다음과 같은 요소가 포함되어 있습니다.

- 데이터 테이블의 이름
- 테이블 상태를 나타내는 아이콘
- 테이블 옵션이 포함된 빨간색 삼각형 메뉴
- (선택 사항) 테이블 변수
- (선택 사항) 테이블 스크립트에 대한 녹색 삼각형

그림 2.5 테이블 패널의 예



테이블 옵션

데이터 테이블 이름 옆의 빨간색 삼각형 메뉴를 클릭하면 다음과 같은 옵션이 표시됩니다.

테이블 테이블 메뉴와 동일한 옵션이 포함되어 있습니다 . 자세한 내용은 " 데이터 재구성 " 장 (309 페이지) 에서 확인하십시오 .

새 테이블 변수 데이터 테이블에서 항상 사용할 수 있도록 하려는 새 테이블 변수 (텍스트 또는 다른 모든 상수 문자 값 가능) 를 생성합니다 . 테이블 변수는 일반적으로 테이블을 문서화하는 데 사용됩니다 . 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블 변수 사용 "(261 페이지) 에서 확인하십시오 .

참고 : 테이블 변수의 이름을 바꾸려면 테이블 변수를 두 번 클릭하고 이름 필드에 새 이름을 입력하십시오 .

새 스크립트 데이터 테이블과 함께 저장할 JSL 스크립트를 생성합니다 . 이 명령을 선택한 후 스크립트 이름을 지정하고 값 (JSL 명령) 을 입력합니다 . **확인**을 클릭하면 테이블 패널에 새 스크립트가 나열되며 녹색 삼각형을 클릭하여 해당 스크립트를 실행 , 편집 또는 삭제할 수 있습니다 . 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블에서 스크립트 생성 및 저장 "(262 페이지) 에서 확인하십시오 .

스크립트 그룹화 선택한 스크립트를 그룹으로 이동합니다 . 그룹 이름을 클릭하여 이름을 바꿀 수 있습니다 . 스크립트를 정리하거나 긴 스크립트 목록을 최소화하려는 경우에 유용합니다 .

스크립트 그룹 해제 선택한 그룹에 있는 스크립트를 그룹 해제하여 상위 수준으로 되돌립니다 .

계산식 실행 제한 계산식을 자동으로 실행하는 기능을 해제합니다 . 실행 기능을 해제하고 계산식 섹션을 작성할 수 있습니다 . 이후에 계산식을 테스트하려면 " 계산식 실행 제한 " 을 선택 취소합니다 .

데이터 테이블 잠그기 데이터 및 열 특성을 편집하거나 추가할 수 없도록 데이터 테이블을 잠급니다 . 잠근 후에도 분석 실행과 특성 할당은 가능합니다 . 자세한 내용은 " 데이터 입력 및 편집 " 장의 "Lock Data Table"(259 페이지) 에서 확인하십시오 .

저장 시 파일 압축 데이터 테이블을 저장할 때 압축합니다 . 데이터 테이블이 저장된 후에는 테이블 패널의 데이터 테이블 이름 옆에 압축된 파일 아이콘 이 나타납니다 . 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블 압축 "(260 페이지) 에서 확인하십시오 .

저장 시 파일 압축 옵션은 파일 크기만 줄여 줍니다 . 이 명령은 데이터를 분석하는 데 필요한 메모리에는 영향을 주지 않습니다 . 파일 크기와 분석에 필요한 메모리를 모두 줄이려면 **열 > 유틸리티 > 선택 열 압축**을 사용합니다 . 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블에서 선택 열 압축 "(229 페이지) 에서 확인하십시오 .

팁 : 환경 설정 > 일반 > GZ 로 압축된 데이터 테이블 열 저장을 선택하여 JMP 에서 테이블 저장 시 항상 GZ 압축이 사용되도록 구성할 수도 있습니다 .

실행 취소 비활성화 실행 취소 기록에서 모든 작업을 제거하고 더 이상 작업을 기록하지 않습니다 . 해당 데이터 테이블이 열려 있는 동안에만 실행 취소가 비활성화되며 , 해당 설정이 데이터 테이블과 함께 저장되지는 않습니다 . 특히 데이터 테이블에서 많은 행을 삭제하거나 , 데이터를 기록하기 위해 많은 양의 메모리를 필요로 하는 다른 작업을 수행할 때 이 옵션을 사용하면 메모리가 절약됩니다 .

테이블 스크립트 복사 테이블을 다시 생성하는 스크립트를 복사합니다. 테이블을 다시 생성하려면 복사한 스크립트를 새 스크립트에 붙여넣은 후 실행합니다. 가상으로 결합된 테이블에서 참조된 열은 스크립트에 포함되지 않습니다.

테이블 스크립트 복사 (데이터 제외) 테이블을 다시 생성하되 데이터는 제외하는 스크립트를 복사합니다.

계산식 다시 실행 데이터 테이블 내에서 계산식이 포함된 모든 열을 다시 실행합니다.

스크립트 옵션

데이터 테이블 패널에서 스크립트를 실행하려면 스크립트 이름 옆의 녹색 삼각형 ▶을 클릭합니다. 스크립트 이름이나 녹색 삼각형을 마우스 오른쪽 버튼으로 클릭하면 다음과 같은 옵션이 표시됩니다.

스크립트 실행 스크립트를 실행합니다.

팁 : 여러 테이블 스크립트를 한 번에 실행하려면 Ctrl 키를 누른 채 실행하려는 테이블 스크립트를 선택합니다. 그런 다음 테이블 스크립트 목록 아래의 빈 영역을 마우스 오른쪽 버튼으로 클릭하고 **스크립트 실행**을 선택합니다.

편집 대부분의 스크립트를 편집할 수 있도록 스크립트 편집기에서 엽니다. JMP 응용 프로그램 스크립트는 응용 프로그램 빌더에서 엽니다.

재코딩을 사용하여 편집 ... 스크립트 편집기에서 열을 재코딩하고 스크립트를 저장하여 생성한 재코딩 스크립트를 엽니다.

삭제 스크립트를 삭제합니다.

스크립트 그룹화 선택한 스크립트를 그룹으로 이동합니다 . 그룹 이름을 클릭하여 이름을 바꿀 수 있습니다. 스크립트를 정리하거나 긴 스크립트 목록을 최소화하려는 경우에 유용합니다.

스크립트 그룹 해제 선택한 그룹에 있는 스크립트를 그룹 해제하여 상위 수준으로 되돌립니다.

복사 스크립트를 복사합니다 . 그런 다음 이를 다른 데이터 테이블의 테이블 패널에 붙여넣을 수 있습니다.

스크립트 디버그 스크립트를 JSL 디버거에서 엽니다. 자세한 내용은 Scripting Guide 의 에서 확인하십시오.

붙여넣기 다른 데이터 테이블의 스크립트를 붙여넣습니다.

추가 옵션

테이블 패널에서는 다음 작업도 수행할 수 있습니다.

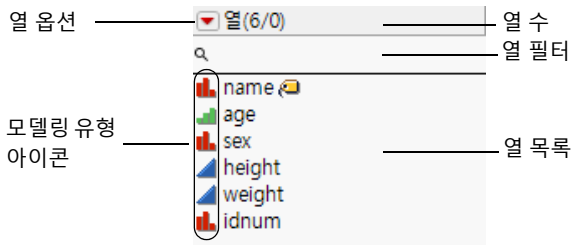
- 테이블 변수 또는 스크립트 이름을 두 번 클릭하여 이름 및 내용을 편집합니다.
- 테이블 변수 또는 스크립트를 드래그하여 재배열합니다.

데이터 테이블의 열 패널

"열" 패널에서는 열을 보고 이동할 수 있습니다. 이 패널에는 다음과 같은 정보가 포함됩니다.

- 열 옵션 (**열** 메뉴와 동일한 옵션 (일부) 포함)
- 데이터 테이블의 총 열 수 및 선택된 열 수
- 열을 빠르게 찾는 데 도움이 되는 필터 상자 (돋보기 아이콘을 클릭하여 검색 필터 옵션 및 모수를 설정함)
- 데이터 테이블의 열 목록
- 각 열의 모델링 유형 ("JMP 열 특성 설정 " 장의 " 모델링 유형 "(271 페이지) 참조) 을 나타내는 아이콘
- 열에 할당된 특성을 나타내는 아이콘 (아래 그림에는 표시되지 않음 . 그림 2.7 참조)

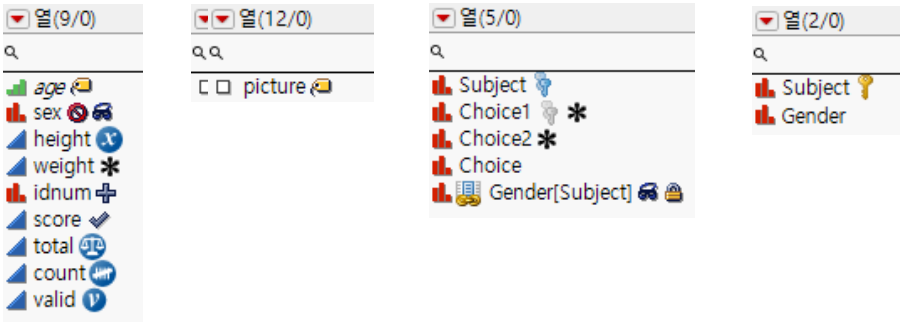
그림 2.6 열 패널의 예



열 특성을 나타내는 아이콘

각 열 이름의 오른쪽에 있는 아이콘은 열에 포함된 특성을 나타냅니다.

그림 2.7 열 특성을 나타내는 아이콘



참고 : 기울임꼴로 표시된 열은 해당 위치에 고정되어 있음을 나타냅니다 . 즉 , 가로로 스크롤해도 해당 열은 계속 표시됩니다 .

다음은 열 패널에 나타날 수 있는 아이콘입니다.

그림에서 해당 열에 대응하는 점의 라벨에 행 번호 대신 값이 사용됨을 나타냅니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "행 및 열 라벨 지정"(217페이지)에서 확인하십시오.

해당 열이 계산에서 제외됨을 나타냅니다. 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블에서 행 제외 "(203 페이지) 에서 확인하십시오.

해당 열이 그래프에 포함되지 않음을 나타냅니다. 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블에서 행 숨기기 "(203 페이지) 에서 확인하십시오.

X 또는 Y 로 표시되며 , 해당 열에 사전 선택된 x 또는 y 역할이 할당되었음을 나타냅니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열에 사전 선택된 분석 역할 할당 "(308 페이지) 에서 확인하십시오.

해당 열에 사전 선택된 **검증** 역할이 할당되었음을 나타냅니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열에 사전 선택된 분석 역할 할당 "(308 페이지) 에서 확인하십시오.

해당 열에 하나 이상의 특성이 포함되어 있음을 나타냅니다 . 이 아이콘을 클릭하면 열에 포함된 특성의 목록이 표시됩니다.

열의 값이 계산식으로 얻은 결과임을 나타냅니다. 계산식 실행이 제한되어 있으면 이 아이콘이 회색으로 나타납니다. 계산식을 보고 편집하려면 이 아이콘을 두 번 클릭합니다. 자세한 내용은 "JMP 의 계산식 생성 " 장의 "JMP 계산식 편집기 옵션 "(379 페이지) 에서 확인하십시오.

범위 확인 또는 목록 확인 옵션이 설정되어 있음을 나타냅니다. 범위 또는 목록을 보고 편집하려면 이 아이콘을 클릭합니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 범위 확인 "(282 페이지) 및 "JMP 열 특성 설정 " 장의 " 목록 확인 "(282 페이지) 에서 확인하십시오.

해당 열에 사전 선택된 **가중치** 역할이 할당되었음을 나타냅니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열에 사전 선택된 분석 역할 할당 "(308 페이지) 에서 확인하십시오.

해당 열에 사전 선택된 **가중치** 역할이 할당되었음을 나타냅니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열에 사전 선택된 분석 역할 할당 "(308 페이지) 에서 확인하십시오.

해당 열의 값을 편집할 수 없음을 나타냅니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열 정보 창 "(269 페이지) 에서 확인하십시오.

해당 표현식 열에 표식으로 사용할 수 있는 이미지가 포함되어 있음을 나타냅니다. 자세한 내용은 "JMP 보고서 " 장의 " 이미지를 표식으로 사용 "(448 페이지) 에서 확인하십시오.

가상 결합 아이콘

해당 열에 " 연결 참조 " 열 특성이 있음을 나타납니다 . 이 특성은 현재 데이터 테이블의 한 열을 참조된 데이터 테이블의 ID 열에 연결합니다 . 파란색은 참조된 데이터 테이블이 연결되어 있음을 나타냅니다 .

해당 열에 " 연결 참조 " 열 특성이 있음을 나타납니다 . 회색은 참조된 데이터 테이블이 열려 있지 않거나 올바르게 연결되지 않았음을 나타냅니다 .

해당 열에 연결 ID 열 특성이 있음을 나타냅니다 . 연결 ID 특성은 데이터 테이블의 특정 열을 ID 열로 표시합니다 . 즉 , 데이터 테이블의 행이 ID 열의 값으로 고유하게 식별됩니다 .

해당 열이 참조된 데이터 테이블에서 연결된 열임을 나타냅니다 .

가상 결합에 대한 자세한 내용은 " 데이터 재구성 " 장의 " 데이터 테이블 가상 결합 "(344 페이지) 에서 확인하십시오 .

검색 필터 옵션

열 목록에 검색 필터 옵션을 표시하려면 열 목록의 빨간색 삼각형 메뉴에서 "필터 표시"를 선택합니다. 그러면 열 검색을 사용자 정의할 수 있으며 이 기능은 특히 열 목록이 길 때 유용합니다.

검색 상자 옆의 아래쪽 화살표 버튼을 클릭하여 검색을 구체화할 수 있습니다 .

용어 포함 검색 기준의 일부가 포함된 항목을 반환합니다 . "ease oom" 을 검색하면 "Release Zoom" 과 같은 메시지가 반환됩니다 .

구 포함 검색 기준이 정확히 포함된 항목을 반환합니다 . "text box" 를 검색하면 "text" 와 "box" 가 바로 연이어 포함된 항목 (예 : "Context Box" 및 "Text Box") 이 반환됩니다 .

구로 시작 검색 기준으로 시작하는 항목을 반환합니다 .

구로 끝남 검색 기준으로 끝나는 항목을 반환합니다 .

전체 구 전체 문자열로 구성된 항목을 반환됩니다 . "text box" 를 검색하면 "text box" 만 포함된 항목이 반환됩니다 .

정규 표현식 검색 상자에서 와일드카드 (*) 와 마침표 (.) 를 사용할 수 있습니다 . "get.*name" 을 검색하면 "get" 다음에 하나 이상의 단어가 포함된 항목을 찾을 수 있습니다 . 즉 , "Get Color Theme Names", "Get Name Info" 및 "Get Effect Names" 등이 반환됩니다 .

결과 반전 검색 기준과 매칭되지 않는 항목을 반환합니다 .

모든 용어 일치 문자열이 모두 포함된 항목을 반환합니다 . "t test" 를 검색하면 검색 문자열 중 하나 또는 둘 모두가 포함된 요소 (예 : "Pat Test", "Shortest Edit Script" 및 "Paired t test") 가 반환됩니다 .

대 / 소문자 무시 검색 기준의 대 / 소문자를 무시합니다 .

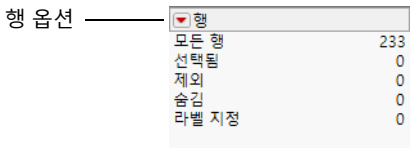
전체 단어 일치 " 모든 용어 일치 " 설정에 따라 문자열의 각 단어가 포함된 항목을 반환합니다 . " 모든 용어 일치 " 옵션이 선택되어 있는 경우 "data filter" 를 검색하면 "data" 와 "filter" 가 반환됩니다 .

데이터 테이블의 행 패널

"행" 패널에는 다음과 같은 정보가 포함됩니다.

- 행 옵션 (**행** 메뉴와 동일한 옵션)
- 총 행 수
- 선택 (강조 표시) 된 행 , 제외된 행 , 숨겨진 행 및 라벨이 지정된 행의 개수

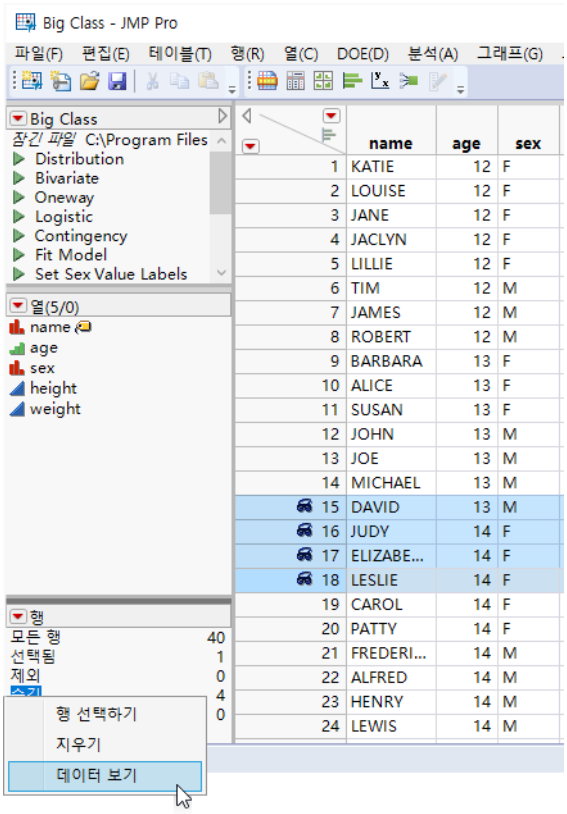
그림 2.8 행 패널의 예



행을 선택 또는 선택 해제하거나 데이터 보기를 생성하려면 "행" 패널에서 해당 범주를 마우스 오른쪽 버튼으로 클릭합니다.

데이터 보기를 생성할 경우 주 데이터 테이블의 연결된 부분집합이 생성됩니다. 예를 들어 몇 개의 행이 숨겨진 행으로 표시된 경우 숨겨진 행만 표시하는 창을 열 수 있습니다. 행 패널에서 **숨김**을 마우스 오른쪽 버튼으로 클릭하고 **데이터 보기**를 선택합니다.

그림 2.9 행 패널에서 데이터 보기 생성

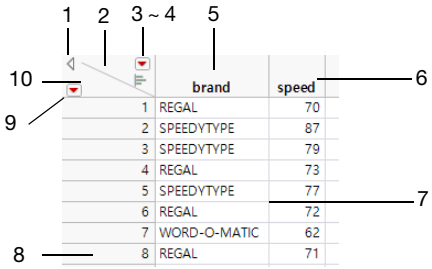


데이터 보기를 사용하는 경우 대부분의 편집 작업은 계속해서 주 데이터 테이블에서 수행하십시오. 주 데이터 테이블이나 데이터 보기 중 어느 쪽에서든 변경을 수행하면 둘 모두에 변경 사항이 반영됩니다. 일부 데이터를 변경하거나 열을 추가하는 등의 사소한 변경 작업은 데이터 보기에서 수행할 수 있습니다. 하지만 계산식을 추가하는 등의 중요한 변경 작업은 주 데이터 테이블에서 수행해야 합니다.

데이터 테이블의 데이터 격자

데이터 격자는 데이터 테이블에서 데이터가 들어 있는 주요 부분입니다. 그림 2.10에서는 데이터 격자에서의 상호 작용 방식을 보여 줍니다. "데이터 입력 및 편집" 장의 "데이터 테이블에서 행 선택"(186 페이지)도 참조하십시오.

그림 2.10 데이터 격자에서의 상호 작용



없습니다.	설명
1	데이터 테이블 패널을 숨기거나 표시합니다.
2	선택된 모든 열을 선택 취소하려면 클릭합니다. 모든 열을 선택하려면 Shift 키를 누른 채 클릭합니다. 여러 열을 추가하려면 두 번 클릭합니다.
3	빨간색 삼각형을 클릭하면 열 메뉴가 표시됩니다.
4	히스토그램을 클릭하면 열 머리글 행에 히스토그램이 표시됩니다.
5	열을 선택하려면 클릭합니다. 열 정보 창을 표시하려면 두 번 클릭합니다. 자세한 내용은 "열 정보 창"(269 페이지)에서 확인하십시오. 열 옵션을 표시하려면 마우스 오른쪽 버튼을 클릭합니다. 자세한 내용은 "열의 마우스 오른쪽 버튼 클릭 옵션"(47 페이지)에서 확인하십시오.
6	열 이름을 두 번 클릭하여 열 정보 창에서 열 이름을 편집하거나, 열 이름을 선택하고 Enter 키를 누릅니다.
7	열 너비를 변경하려면 클릭한 채로 드래그합니다. 여러 열의 너비를 변경하려면 해당 열을 선택한 후 Alt 키를 누른 채로 클릭하여 드래그합니다.
8	행을 선택하려면 클릭합니다. 행 편집기를 열려면 두 번 클릭합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 셀 편집"(209 페이지)에서 확인하십시오. 행 옵션을 표시하려면 마우스 오른쪽 버튼을 클릭합니다. 자세한 내용은 "행의 마우스 오른쪽 버튼 클릭 옵션"(48 페이지)에서 확인하십시오.
9	행 메뉴를 표시합니다.
10	선택된 모든 행을 선택 취소하려면 클릭합니다. 모든 행을 선택하려면 Shift 키를 누른 채 클릭합니다. 행을 추가하려면 두 번 클릭합니다.

열의 마우스 오른쪽 버튼 클릭 옵션

열 머리글을 마우스 오른쪽 버튼으로 클릭하면 다음과 같은 옵션이 표시됩니다.

열 정보 열 정보 창을 엽니다. 자세한 내용은 "열 정보 창"(269 페이지) 에서 확인하십시오.

속성 표준화 한 열의 속성 (데이터 유형, 모델링 유형, 숫자 형식 등) 또는 특성 (계산식, 노트, 목록 및 범위 확인 등) 을 다른 열에 복사합니다. 또한 선택한 속성 또는 특성을 열에서 삭제할 수도 있습니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "열 간에 속성 및 특성 표준화"(305 페이지) 에서 확인하십시오.

열 특성 열 특성 목록이 포함되어 있습니다. 특성을 선택하면 열 정보 창을 열고 열에 해당 특성을 적용할 수 있습니다. 이 목록은 열 정보 창에서도 사용할 수 있습니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "JMP 의 열 특성"(280 페이지) 에서 확인하십시오.

계산식 계산식 편집기를 엽니다. 자세한 내용은 "JMP 의 계산식 생성" 장 (359 페이지) 에서 확인하십시오.

재코딩 열의 모든 값을 한 번에 변경할 수 있습니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "열의 데이터 재코딩"(245 페이지) 에서 확인하십시오.

새 계산식 열 계산식 열을 생성합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "새 계산식 열 생성"(236 페이지) 에서 확인하십시오.

열 삽입 선택된 한 열 또는 여러 열 앞에 하나 이상의 열을 삽입합니다.

열 삭제 선택된 모든 열을 삭제합니다.

라벨 / 라벨 해제 모든 그림에서 선택된 열의 라벨을 표시하거나 표시하지 않습니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "행 및 열 라벨 지정"(217 페이지) 에서 확인하십시오.

연결 ID 가상으로 결합된 테이블에서 보조 데이터 테이블의 한 열을 ID 열로 표시하는 데 사용됩니다. 즉, 데이터 테이블의 행이 ID 열의 값으로 고유하게 식별됩니다. 데이터 테이블의 ID 열에 해당하는 열에는 "연결 ID" 열 특성 체크박스가 선택되어 있습니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "가상 결합 특성을 사용하여 데이터 테이블 연결"(299 페이지) 에서 확인하십시오.

연결 참조 가상으로 결합된 데이터 테이블에서 주 데이터 테이블의 한 열을 보조 데이터 테이블의 ID 열에 매핑하는 데 사용됩니다. "연결 참조" 열 특성은 보조 데이터 테이블의 경로 이름을 지정합니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "가상 결합 특성을 사용하여 데이터 테이블 연결"(299 페이지) 에서 확인하십시오.

정렬 선택된 열의 값을 기준으로 테이블의 모든 행을 정렬합니다. 오름차순 또는 내림차순으로 행을 정렬하도록 선택할 수 있습니다. 자세한 내용은 "데이터 재구성" 장의 "데이터 테이블 정렬"(314 페이지) 에서 확인하십시오.

열 특성 복사 열 특성을 복사합니다.

열 특성 붙여넣기 열 특성을 붙여넣습니다.

열 복사 열을 다시 생성하기 위한 스크립트를 복사합니다. 모든 속성과 특성을 복사하지만 데이터는 복사하지 않습니다.

열 붙여넣기 "열 복사"로 생성된 스크립트에서 새 열을 생성합니다. 이 열은 선택된 열 뒤에 배치됩니다. 선택된 열이 없으면 데이터 테이블의 맨 끝에 새 열이 배치됩니다. 스크립트에 포함된 데이터 행 수만큼의 새 열이 대상 테이블에 채워집니다.

행의 마우스 오른쪽 버튼 클릭 옵션

행 머리글을 마우스 오른쪽 버튼으로 클릭하면 다음과 같은 옵션이 표시됩니다.

숨기고 제외하기 선택된 행을 모든 그림과 그래프에서 숨기고 분석에서 제외합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 행 숨기고 제외하기"(202 페이지)에서 확인하십시오.

제외 / 제외 해제 선택된 행을 분석에서 제외하거나 분석에 포함합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 행 제외"(203 페이지)에서 확인하십시오.

숨기기 / 숨기기 해제 선택된 행을 모든 그림 및 그래프에서 숨기거나 표시합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 행 숨기기"(203 페이지)에서 확인하십시오.

라벨 / 라벨 해제 모든 그림에서 선택된 행의 라벨을 표시하거나 표시하지 않습니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "행 및 열 라벨 지정"(217 페이지)에서 확인하십시오.

색상 색상 팔레트를 제공합니다. 여기서 색상을 선택하여 선택된 행에 적용합니다. 이 색상은 그림과 그래프에서 사용됩니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "행에 색상 할당"(218 페이지)에서 확인하십시오.

표식 표식 또는 기호 팔레트를 제공합니다. 여기서 표식을 선택하여 선택된 행에 적용합니다. 표식은 그림과 그래프에서 점 대신 사용됩니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "행에 표식 추가"(218 페이지)에서 확인하십시오.

행 상태를 기준으로 행에 색상 적용 행에 현재 행 상태 색상과 동일한 색상을 적용합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "열 값에 따라 행에 색상 또는 표식 할당"(218 페이지)에서 확인하십시오.

매칭 셀 선택 활성 데이터 테이블에서 선택된 행과 매칭되는 값이 포함된 행을 선택합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "행 및 열 크기 조정"(197 페이지)에서 확인하십시오.

선택 반전 이전에 선택 취소된 모든 행을 선택하고, 현재 선택된 모든 행을 선택 취소합니다.

행 상태 지우기 데이터 테이블의 모든 활성 행 상태를 지웁니다. 모든 행이 포함되고, 표시되고, 라벨 해제되고, 그림에 검은 점으로 표시됩니다. 행 상태 열에 저장된 행 상태에는 영향이 없습니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "행 특성 삭제"(225 페이지)에서 확인하십시오.

행 추가 데이터 테이블에 지정된 수만큼 행을 추가합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에 행 추가"(181 페이지)에서 확인하십시오.

행 삭제 선택된 모든 행을 데이터 테이블에서 제거합니다. 실수로 삭제한 경우 작업을 실행 취소하려면 **편집** 메뉴에서 **실행 취소** 명령을 사용합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 행 삭제"(198 페이지)에서 확인하십시오.

커서 형태

커서는 데이터 격자의 어느 위치에 있는지에 따라 형태가 달라집니다.

화살표 커서 다음 위치에서는 표준 화살표 커서가 나타납니다.

- 데이터 테이블 왼쪽의 패널 영역
- 데이터 격자의 왼쪽 상단에 있는 삼각형 행 및 열 영역

화살표 커서로 수행할 수 있는 작업은 다음과 같습니다.

- 화살표 커서를 사용하여 열을 선택하려면 " 열 " 패널에서 열 이름을 클릭합니다.
- " 열 " 패널에서 열 이름을 두 번 클릭하여 편집합니다.
- 데이터 격자의 왼쪽 상단에 있는 삼각형 영역을 클릭하여 행 및 열을 선택 취소합니다.

선택 (큰 더하기 기호) 커서 커서가 열 머리글 위나 행 번호 영역에 있을 때는 행 또는 열을 선택할 수 있음을 나타내는 큰 더하기 기호가 됩니다. 마우스를 클릭하면 해당 행 또는 열이 선택되고 강조 표시됩니다. 여러 행 또는 열을 선택하려면 마우스를 클릭한 채로 드래그하고, 인접하지 않은 여러 행 또는 열을 선택하려면 Ctrl 키를 누른 채로 해당 행 또는 열을 클릭합니다.

- 열에 대한 열 정보 창을 표시하려면 해당 열 머리글 영역을 두 번 클릭합니다.
- 열 이름을 변경하려면 열을 선택하여 열을 강조 표시합니다. 그런 다음 입력을 시작합니다 (해당 열 또는 데이터 테이블이 잠기지 않은 경우).
- 행 편집기를 사용하여 행을 편집하려면 행 번호 영역을 두 번 클릭합니다.

I- 빔 커서 편집 가능한 텍스트를 선택하면 커서가 표준 I- 빔 형태가 됩니다. 텍스트를 편집하려면 I-빔 커서를 강조 표시된 텍스트에 놓습니다. 그런 다음 삽입 지점을 클릭하여 표시하거나, 마우스를 드래그하여 바꿀 텍스트를 선택합니다.

양방향 화살표 커서 열 또는 패널 경계에서는 커서가 양방향 화살표로 바뀝니다. 열 또는 패널의 너비를 변경하려면 이 커서를 왼쪽 또는 오른쪽으로 드래그합니다. 열의 너비를 변경해도 열 정보 창 (열 이름을 두 번 클릭하여 액세스) 에서 지정한 열 필드 너비에는 영향이 없습니다.

팁 : Alt 키를 누른 채로 선택한 열 중 하나의 경계에 있는 양방향 화살표 커서를 드래그하면 선택한 모든 열의 너비를 한 번에 조정할 수 있습니다.

목록 확인 커서 데이터 검증이 적용된 열의 값 위에 마우스를 놓으면 커서 형태가 바뀝니다. 즉, 목록 확인 기능이 설정된 열에서는 커서가 아래쪽을 가리키는 작은 화살표 모양이 됩니다. 마우스를 클릭하면 해당 값이 강조 표시되고 커서는 I- 빔 형태가 됩니다. 평소처럼 유효한 텍스트 또는 유효한 숫자로 정의된 값으로 데이터를 입력하거나 편집합니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 목록 확인 "(282 페이지) 에서 확인하십시오.

포인터 커서 다음 개체 위에 마우스를 놓으면 커서가 포인터로 바뀝니다.

- 빨간색 삼각형 메뉴 - 옵션 표시
- 삼각형 표시 아이콘 - 패널 열기 또는 닫기

– 모델링 유형 아이콘 - 모델링 유형 변경

데이터 파일을 여는 옵션

참고 : 파일 열기에 대한 자세한 내용은 " 데이터 가져오기 " 장 (65 페이지) 에서 확인하십시오 .

데이터 집합을 열려면 **파일 > 열기**를 선택하고 파일 유형을 선택합니다. 일부 파일 유형의 경우 "데이터 파일 열기" 창에서 추가 기능 및 옵션이 나타납니다(표 2.1).

팁 :

- Windows 에만 해당 : 매번 동일한 파일 유형을 열려면 **다음에 이 창이 호출될 때 이 필터 선택** 체크박스를 선택합니다 .
- 파일을 JMP 홈 창으로 드래그하는 방법으로 열 수도 있습니다 .
- **파일 > 열기** 명령 사용 시 기본 검색 디렉터리를 변경하는 방법은 "JMP 환경 설정 " 장의 " 파일 위치의 환경 설정 "(659 페이지) 에서 확인하십시오 .

표 2.1 파일 유형별 파일 > 열기 옵션

파일 유형	추가 기능 및 옵션
JMP 데이터 테이블	테이블 노트, 열 수 및 행 수가 나타납니다. 데이터 테이블로 가져올 열을 선택하려면 **열 선택** 옵션을 사용합니다. Windows의 경우, **열기** 옆의 화살표를 클릭한 후 "열 선택" 옵션을 사용하여 데이터 테이블로 가져올 열을 지정합니다. macOS의 경우, **열기**를 클릭한 후 가져올 열을 선택할 수 있습니다.
Excel 파일	• 레이아웃을 사용자 정의하고 파일을 가져오기 전에 데이터를 미리 볼 수 있도록 기본적으로 Excel 가져오기 마법사에서 파일을 가져옵니다 . • **열기** 옆의 화살표를 클릭한 후 옵션 중 하나를 선택합니다. 첫 번째 스프레드시트 행을 열 머리글로 변환합니다 .
텍스트 파일	• 데이터 배열을 자동으로 결정하려면 다음 옵션 중 하나를 선택합니다 . – **데이터 (환경 설정 사용)** – **데이터 (최적 추측)** – **스크립트 편집기의 텍스트** • 데이터 배열을 수동으로 지정하려면 **데이터 (미리보기 사용)** 옵션을 선택합니다 . 자세한 내용은 " 데이터 가져오기 " 장의 "JMP 텍스트 가져오기 마법사의 옵션 "(84 페이지) 에서 확인하십시오 .

표 2.1 파일 유형별 파일 > 열기 옵션 (계속)

파일 유형	추가 기능 및 옵션
SAS 데이터 집합	• 열 이름에 SAS 변수를 사용합니다 . • (Windows 에만 해당) 암호로 보호된 데이터 집합을 열 경우 암호를 입력합니다 . • (SAS 전송 파일에만 해당) 열기 전에 열을 선택합니다 .
SPSS 데이터 파일	열 이름에 SPSS 변수 또는 라벨 이름을 사용합니다.
XML 데이터 파일	옵션에 대한 자세한 내용은 텍스트 파일에서 확인하십시오.
JSON 데이터 파일	옵션에 대한 자세한 내용은 텍스트 파일에서 확인하십시오.
모든 파일	• 데이터 배열을 자동으로 결정하려면 다음 옵션 중 하나를 선택합니다 . – 열기 형식 : 데이터 (텍스트 가져오기 환경 설정 사용) – 열기 형식 : 데이터 (최적 추측 사용) – 열기 형식 : 일반 텍스트 (스크립트 창에) • 데이터 배열을 수동으로 지정하려면 **데이터 (미리보기 사용)** 옵션을 선택합니다 . • (Windows 에만 해당) 컴퓨터의 기본 텍스트 편집기를 사용하여 텍스트 파일을 열려면 **기본 프로그램을 사용하여 엽니다.** 를 선택합니다 .

플랫폼 및 보고서

JMP 는 통계 분석 유형별로 체계화된 여러 플랫폼으로 구성되어 있습니다. 예를 들어 분포 플랫폼에서는 히스토그램, 추가 그래프 및 보고서를 사용한 단변량 분석 (단일 변수 분석) 을 생성합니다. 분포 플랫폼에서 데이터를 분석한 후 다른 플랫폼을 선택하여 데이터를 보다 철저하게 검토할 수 있습니다.

플랫폼과 보고서는 다음과 같은 방식으로 함께 작동합니다.

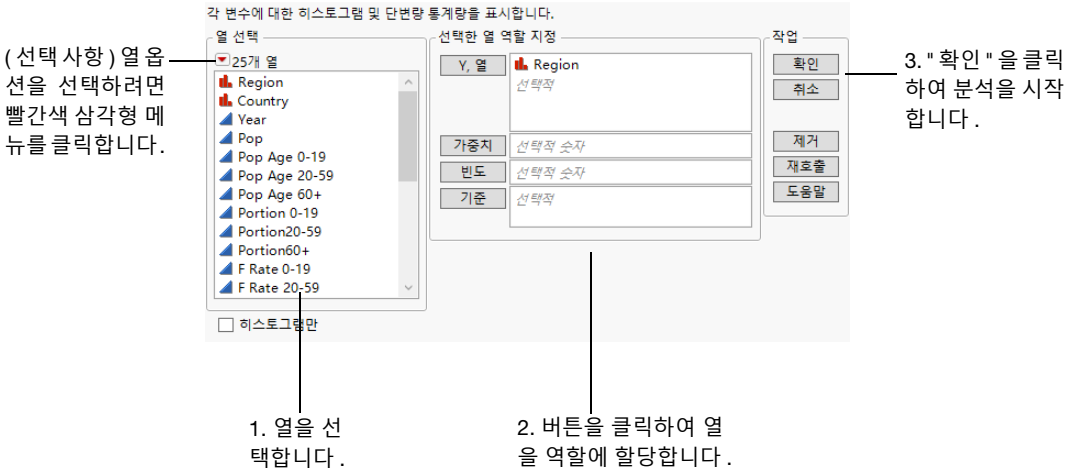
1. 데이터 테이블은 플랫폼의 입력 소스입니다 . 자세한 내용은 " 데이터 테이블 기능 "(36 페이지) 에서 확인하십시오 .
2. 플랫폼 분석은 시작 창에서 시작됩니다 . 자세한 내용은 " 플랫폼의 시작 창 "(52 페이지) 에서 확인하십시오 .
3. 플랫폼 결과는 보고서 창에 나타납니다 . 자세한 내용은 " 보고서의 분석 결과 "(54 페이지) 에서 확인하십시오 .

그림 2.2 "일반적인 JMP 세션의 구조(Windows)"(35 페이지)에서는 플랫폼이 데이터 테이블과 상호 작용한 다음 결과를 보고서 창에 표시하는 방식을 보여 줍니다. 플랫폼 사용에 대한 정보는 JMP 설명서 라이브러리에서 제공됩니다.

플랫폼의 시작 창

시작 창은 플랫폼 진입점으로서, 여기에서 분석할 열을 지정할 수 있습니다. 그림 2.11에서는 분포 시작 창을 보여 줍니다.

그림 2.11 시작 창 기능



모든 시작 창에서 공통적으로 제공되는 옵션은 다음과 같습니다.

열 선택 현재 데이터 테이블의 모든 변수를 나열합니다. 다음 사항에 유의하십시오.

- 모델링 유형을 변경하려면 열 이름 옆의 모델링 유형 아이콘을 마우스 오른쪽 버튼으로 클릭합니다.
- 변환 열을 생성하려면 열 이름을 마우스 오른쪽 버튼으로 클릭합니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "JMP 플랫폼의 열 변환"(236 페이지)에서 확인하십시오.
- 빨간색 삼각형 메뉴의 옵션을 사용하여 열을 필터링하고 정렬할 수 있습니다. 자세한 내용은 "열 필터 메뉴"(53 페이지)에서 확인하십시오.

선택한 열 역할 지정 선택한 열을 역할 (**Y**, **X** 등)로 전환합니다. 열에 변수 역할을 지정할 수 있습니다. 예를 들어 관련 요소에 역할을 지정합니다. 자세한 내용은 "선택한 열 역할 지정 버튼"(53 페이지)에서 확인하십시오.

그래프 빌더 창에서 이 패널을 표시하려면 왼쪽의 **대화상자** 버튼을 클릭하거나 **실행 취소 > 분석 다시 시작**을 선택합니다.

작업 다음 옵션을 사용할 수 있습니다.

확인 분석을 수행합니다.

취소 분석을 중지하고 시작 창을 종료합니다 .

제거 선택한 모든 변수를 역할에서 삭제합니다 .

재호출 최근에 수행한 분석으로 시작 창을 채웁니다 .

도움말 시작 창에 대한 도움말을 표시합니다 .

팁 : 마지막으로 열었던 플랫폼 시작 창을 다시 열려면 **보기 > 마지막 플랫폼 재호출** (Windows) 또는 **창 > 마지막 플랫폼 재호출** (macOS) 을 선택합니다 .

선택한 열 역할 지정 버튼

다음은 전체 시작 창에서 자주 나타나는 버튼입니다. 특정 플랫폼과 관련된 버튼은 특정 플랫폼에 대한 장에서 설명합니다.

Y 열을 반응 변수 또는 종속 변수로 식별합니다. 이러한 변수는 분포를 살펴보려는 대상입니다.

X 열을 독립 변수 , 분류 변수 또는 설명 변수로 식별합니다 . 이러한 변수는 Y 변수의 분포를 예측합니다 .

가중치 데이터에 가중치 (중요도 또는 영향력) 를 할당하는 데 사용할 변수가 있는 데이터 테이블 열을 식별합니다 .

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 데이터 테이블 열을 식별합니다 . 이 옵션은 요약된 데이터에서 각 행에 빈도를 할당하는 경우에 유용합니다 . 0 또는 양의 정수 값은 여러 항목이 기록된 경우 각 행의 관측값 빈도 또는 개수를 나타냅니다 .

JMP PRO **검증** 모형 적합 시 교차 검증용 훈련 , 검증 및 테스트 데이터 집합에 행을 할당하는 데 사용되는 값이 있는 데이터 테이블 열을 식별합니다 .

참고 :

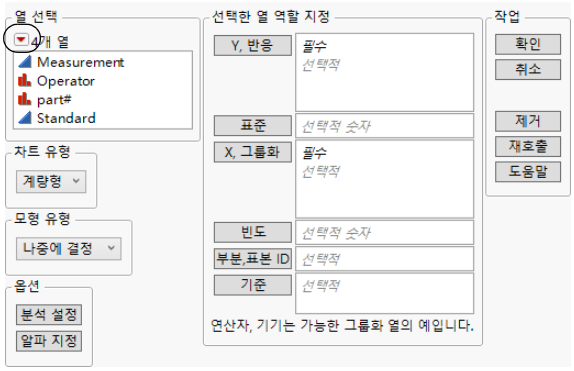
- 일부 플랫폼에서는 검증 열에 4 개 이상의 수준을 지정할 경우 K 폴드 검증을 사용할 수 있습니다 .
- " 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 검증 열을 생성하기 위한 옵션이 표시됩니다 ." 검증 열 생성 " 유틸리티에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다 .

열 필터 메뉴

열 필터 메뉴는 대부분의 시작 창에서 나타납니다. "열 선택" 패널에서 빨간색 삼각형을 클릭하면 열 필터 메뉴를 볼 수 있습니다. 열 정렬, 열 표시 또는 숨김, 열 검색 등을 수행하려면 이 메뉴의 옵션을 사용합니다.

그림 2.12 열 필터 메뉴



재설정 열을 원래 목록으로 재설정합니다.

필터 표시 찾으려는 열 이름을 입력할 수 있는 상자를 엽니다. 상자 옆의 화살표를 클릭하고 필터를 사용자 정의하기 위한 옵션을 선택할 수 있습니다.

이름순 정렬 이름의 사전순서를 활용해 열을 오름차순, 내림차순 또는 원래 순서로 정렬합니다.

모델링 유형 특정 모델링 유형의 열을 표시하거나 숨기기 위한 옵션을 제공합니다.

숫자 데이터 유형이 숫자인 열을 표시하거나 숨깁니다.

문자 데이터 유형이 문자인 열을 표시하거나 숨깁니다.

표현식 데이터 유형이 표현식인 열을 표시하거나 숨깁니다.

형식 제외 특정 형식의 열을 열 선택 목록에서 제외합니다.

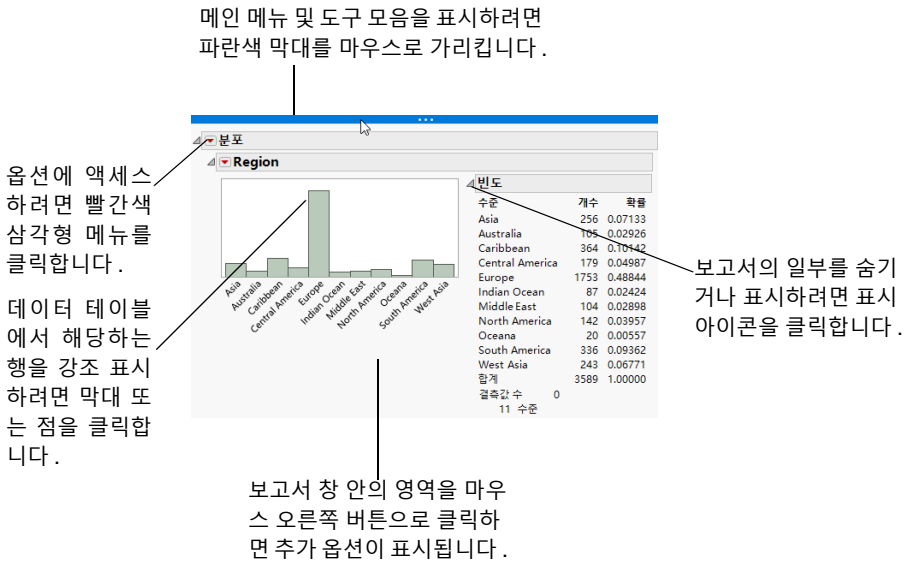
열 그룹 열 그룹을 표시하거나 숨깁니다. 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블에서 열 그룹화 "(199 페이지) 에서 확인하십시오.

그룹 해제된 열 그룹화되지 않은 열을 표시하거나 숨깁니다.

보고서의 분석 결과

분석을 시작하면 JMP 보고서 창이 나타납니다. 보고서 창에서는 대화식 그래프 및 텍스트 보고서를 사용하여 분석 결과가 표시됩니다. 자세한 내용은 "JMP 보고서" 장(405 페이지)에서 확인하십시오.

그림 2.13 보고서 창 기능



파일 및 열린 창 관리

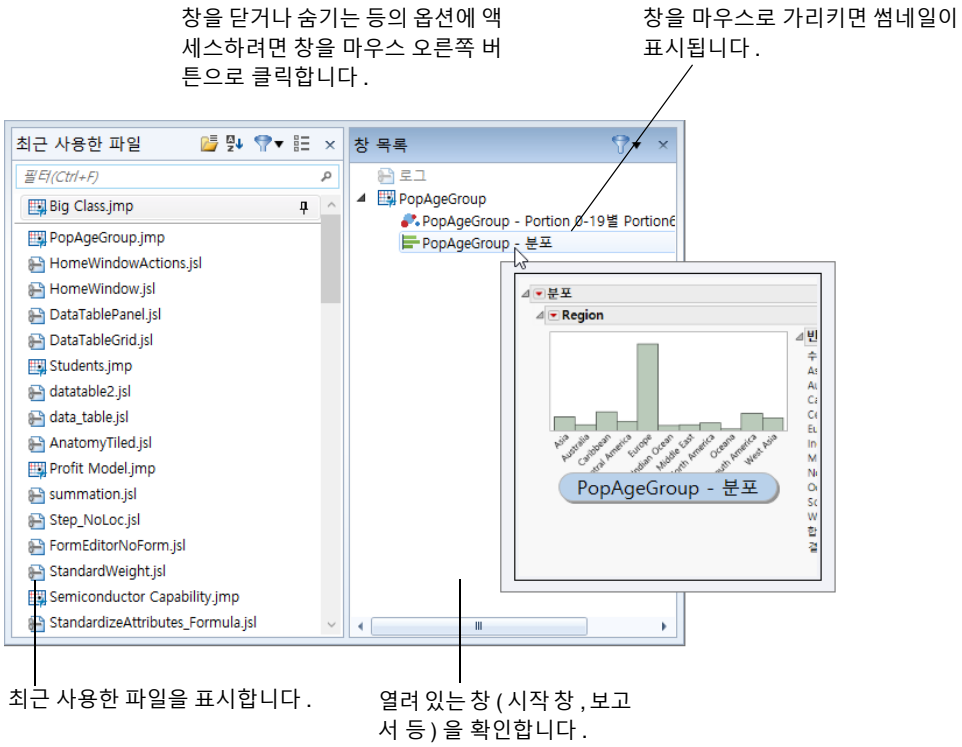
JMP 홈 창에서는 열려 있는 창 및 파일에 바로 액세스할 수 있습니다. 최근에 열었던 파일 열기, 창 닫기, 자주 사용하는 파일 고정 등의 옵션을 사용할 수 있습니다.

JMP 홈 창 (Windows)

Windows에서 JMP를 열면 JMP 홈 창이 나타납니다.

- "최근 사용한 파일" 목록에서 최근에 사용한 파일을 열거나 자주 사용하는 파일을 고정합니다.
- " 최근 사용한 파일 " 목록의 파일을 마우스 오른쪽 버튼으로 클릭하여 일반적인 작업을 수행합니다.
- " 창 목록 " 에서 활성 JMP 창을 열거나 닫습니다.
- 상단 데이터 테이블 목록에서 열려 있는 데이터 테이블을 선택하여 현재 데이터 테이블로 설정합니다.

그림 2.14 JMP 홈 창 작업



JMP 홈 창 버튼

Windows에서는 JMP 홈 창 패널에 다음 버튼이 포함될 수 있습니다.

" 선택 열기 " 버튼은 " 최근 사용한 파일 " 목록에서 선택된 파일을 엽니다.

" 이름순 정렬 " 버튼은 " 최근 사용한 파일 " 을 사전순으로 정렬합니다.

" 필터 " 버튼은 " 최근 사용한 파일 " 및 " 창 목록 " 패널에 나타나는 파일 유형을 필터링합니다.

" 최근 사용한 파일 " 목록에 표시되는 항목의 글꼴 및 아이콘 크기를 늘립니다.

" 닫기 " 버튼은 패널을 닫습니다.

최근 사용한 파일

Windows에서는 "최근 사용한 파일" 목록을 통해 최근에 열었던 파일에 빠르게 액세스할 수 있습니다.

자주 사용하는 파일은 빠르게 액세스할 수 있도록 목록의 맨 위에 유지하십시오. 파일 이름을 마우스로 가리키고 고정 아이콘 을 클릭합니다. 파일을 고정 해제하려면 고정 아이콘 을 클릭합니다.

"최근 사용한 파일" 목록에서 파일을 다음과 같은 위치로 드래그할 수 있습니다.

- JMP " 창 목록 " 으로 드래그하여 파일을 엽니다 .
- Windows 폴더 또는 바탕 화면으로 드래그하여 파일의 복사본을 생성합니다 .
- Microsoft 메모장 또는 Microsoft Word 같은 응용 프로그램으로 드래그하여 파일을 편집합니다 .

파일을 마우스 오른쪽 버튼으로 클릭하면 다음 옵션에 액세스할 수 있습니다.

- 파일 열기
- 파일의 위치 경로 복사
- 파일이 포함된 폴더 내에서 파일 열기
- 이름 기준 사전순으로 또는 최근에 사용한 순서대로 파일 정렬
- 목록에서 선택한 파일 제거
- 더 이상 동일한 위치에 있지 않은 파일 제거 (또는 파일을 선택한 후 Delete 키를 눌러도 됨)
- (JSL 스크립트) 스크립트 편집 , 실행 또는 디버깅
- (텍스트 파일) 가져오기 방법 변경 또는 일반 텍스트로 열기
- (SQL 쿼리) 쿼리 빌더에 저장된 SQL 쿼리 실행 . 기존 쿼리를 템플릿으로 사용하여 새 쿼리를 생성할 수도 있습니다 .
- JMP 응용 프로그램 실행 또는 편집

  " 최근 사용한 파일 " 목록에서 JMP 파일이 아닌 파일을 열면 JMP 가 가져오기 환경 설정을 적용하여 데이터를 배열합니다 . 텍스트 파일을 마우스 오른쪽 버튼으로 클릭하면 가져오기 방법을 변경하거나 파일을 텍스트 편집 창에서 열 수 있습니다 . 현재 가져오기 환경 설정은 팝업 메뉴에서 굵게 표시됩니다 .

팁 : 스크립트를 통해 열었던 파일이 " 최근 사용한 파일 " 창에 표시되지 않도록 하려면 **파일** > **환경 설정** > **일반**을 선택하고 **스크립트로 열린 파일을 최근 사용한 파일 목록에 추가**를 선택 취소하십시오 .

최근 사용한 파일 목록에서 JMP 파일 검색

파일 목록이 길 때 최근에 열었던 파일을 찾으려면 "최근 사용한 파일" 목록의 필터 상자에 파일 이름 중 일부를 입력합니다.

다음 옵션을 사용할 수 있습니다.

- 파일 목록 위쪽의 " 필터 " 상자에 파일 이름의 일부를 입력합니다 .
- " 필터 " 상자를 마우스로 가리키면 검색 팁이 표시됩니다 .
- 파일이 필터링된 상태에서 위쪽 및 아래쪽 화살표 키로 선택된 파일을 변경합니다 . Enter 키를 눌러 선택된 파일을 열거나 , 커서로 파일을 선택합니다 .
- "최근 사용한 파일" 목록에 선택된 파일이 있는 경우 Ctrl+F를 눌러 필터 상자를 활성화합니다.

- 검색어를 입력한 후 필터링 모드를 종료하려면 Esc 키를 누르거나 X 버튼을 클릭합니다 .

창 목록

Windows의 경우, "창 목록"에는 데이터 테이블, 보고서 및 스크립트 같은 열려 있는 JMP 창이 표시됩니다. JMP 창을 열거나, 닫거나, 재배열하거나, 숨길 수 있습니다.

- " 창 목록 " 의 파일을 마우스로 가리키면 썸네일이 나타납니다 .
- " 창 목록 " 을 별도의 창에서 열려면 **보기 > 창 목록**을 선택합니다 .
- 여러 창을 나란히 표시하려면 선택한 창을 마우스 오른쪽 버튼으로 클릭하고 **배열**을 선택합니다 .
- " 창 목록 " 을 항상 최대화된 창 안에 표시하려면 **파일 > 환경 설정 > Windows 관련**을 선택하고 **최대화된 창에 창 목록 도킹**을 선택합니다 .

파일을 마우스 오른쪽 버튼으로 클릭하면 다음 옵션에 액세스할 수 있습니다.

- 창 보기
- 창 닫기 (또는 파일을 선택하고 Delete 키를 눌러도 됨)
- 현재 선택된 창을 제외한 모든 창 닫기 (선택된 창이 보고서 창이면 종속된 데이터 테이블도 열린 상태로 유지)
- 창을 뒤로 이동
- 창을 Windows 작업 표시줄에서 숨기거나 숨기기 해제
- 모든 창 선택 또는 모두 선택 해제

참고 : 활성 파일은 굵은 텍스트 형식으로 표시됩니다 .

홈 창의 패널 재배열

Windows에서 홈 창의 패널을 재배열하려면 패널의 제목 표시줄을 클릭한 채로 드래그합니다. 제목 표시줄을 위쪽, 아래쪽, 왼쪽 또는 오른쪽 화살표 위에 놓으면 패널이 해당 위치에 배치됩니다. 파란색 상자는 패널이 배치될 위치를 나타냅니다. 패널을 탭으로 전환하려면 패널을 홈 창의 가운데로 드래그하여 놓습니다.

팁 :

- 홈 창을 다시 원래 순서로 되돌리려면 **보기 > 홈 창 페인 > 레이아웃 초기화**를 선택합니다 .
- 패널을 닫은 후 다시 열려면 **보기 > 홈 창 페인**을 선택하고 열려는 패널을 선택합니다 .
- JMP 홈 창을 열려면 **보기 > 홈 창**을 선택하거나 , 대부분의 JMP 창에서 오른쪽 하단에 있는 **JMP 홈 창** 버튼을 클릭합니다 . **JMP 홈 창** 버튼이 표시되지 않으면 **보기 > 상태 표시줄**을 선택합니다 .

- 시작 시 기본적으로 "JMP 시작하기 " 또는 " 창 목록 " 을 표시하려면 환경 설정에서 해당 설정을 지정하면 됩니다 (Windows 에만 해당). **파일 > 환경 설정 > 일반**을 선택하고 " 초기 JMP 창 " 목록에서 옵션을 선택하십시오 .

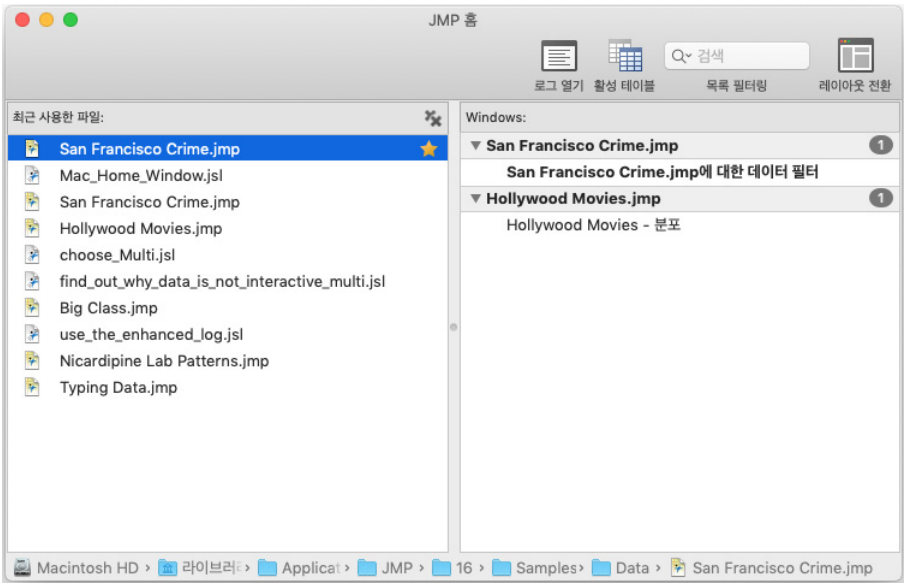
JMP 홈 창 (macOS)

macOS에서 JMP 홈 창을 사용하여 JMP 파일과 열려 있는 창에 빠르게 액세스할 수 있습니다.

- " 최근 사용한 파일 " 목록에서 최근에 사용한 파일을 엽니다 .
- " 최근 사용한 파일 " 목록에서 Control 키를 누르고 파일을 선택하여 즐겨찾기 추가 또는 제거 같은 일반적인 작업을 수행합니다 .
- " 창 목록 " 에서 활성 JMP 창을 열거나 닫습니다 .
- " 활성 테이블 " 목록에서 열려 있는 데이터 테이블을 선택하여 현재 데이터 테이블로 설정합니다 . 데이터 테이블을 맨 앞에 표시하지 않고 활성화하려면 목록에서 Option 키를 누른 채 데이터 테이블을 선택합니다 .
- 로그를 엽니다 .
- 검색 상자에 검색어를 입력하여 " 최근 사용한 파일 " 목록에서 파일을 찾거나 " 창 목록 " 에서 열려 있는 창을 찾습니다 .
- **레이아웃 전환**을 클릭하여 " 최근 사용한 파일 " 목록과 " 창 목록 " 을 세로 대신 가로로 표시합니다 .

macOS 에서 JMP 홈 창을 열려면 **창 > JMP 홈** (그림 2.15) 을 선택합니다 . JMP 를 열 때마다 이 창을 표시하려면 **JMP > 환경 설정 > 일반 > 홈 창**을 선택합니다 .

그림 2.15 JMP 홈 창의 예 (macOS)



팁 :

- 상태 표시줄 영역에는 현재 선택된 파일의 경로가 표시됩니다. 폴더를 클릭하면 해당 폴더가 열리므로 폴더 내용을 볼 수 있습니다 .
- JMP 창에서 JMP 홈 창을 빠르게 보려면 Ctrl 키를 누른 채 창을 클릭하고 , **도구 모음 사용자 정의**를 선택한 후 , JMP 홈 아이콘을 도구 모음으로 드래그합니다 .
- " 최근 사용한 파일 " 목록의 파일을 다른 문서 또는 Finder 에 복사하려면 파일을 선택한 후 **편집 > 복사**를 선택합니다 .

최근 사용한 파일

macOS 에서는 "최근 사용한 파일" 목록을 통해 최근에 열었던 파일에 빠르게 액세스할 수 있습니다.

파일 이름을 마우스로 가리키면 표시되는 별 모양을 클릭하여 해당 파일을 즐겨찾기에 추가할 수도 있습니다.

다른 옵션에 액세스하려면 Ctrl 키를 누른 채 파일을 선택합니다. 예를 들어 Ctrl 키를 누른 채 "스크립트 실행" 또는 "스크립트 디버그"를 선택하면 JSL 스크립트를 실행하거나 디버깅할 수 있습니다.

팁 : 스크립트를 통해 열었던 파일이 " 최근 사용한 파일 " 창에 표시되지 않도록 하려면 **JMP** > **환경 설정** > **일반**을 선택하고 **스크립트로 열린 파일을 최근 사용한 파일 목록에 추가**를 선택 취소하십시오 .

창 목록

macOS의 경우, "창 목록"에는 데이터 테이블, 보고서 및 스크립트 같은 열려 있는 JMP 창이 표시됩니다.

창을 닫으려면 "창 목록"의 창을 마우스로 가리키면 나타나는 **x** 버튼을 클릭합니다.

닫으려는 창이 하나 이상의 하위 창이 있는 상위 창이면 **데이터 테이블 숨기기**, **취소** 또는 **모두 닫기** 중 하나를 선택하라는 메시지가 표시됩니다. 상위 창과 해당하는 모든 하위 창을 닫으려면 **모두 닫기**를 선택합니다.

창에서 최근에 열었던 파일 검색

Windows에서 최근에 연 파일은 홈 창에 나타납니다. 또한 모든 JMP 창에서 최근에 열었던 파일을 검색할 수 있습니다.

파일 > 빠르게 열기를 선택하거나, Alt+Shift+O를 눌러 "필터" 상자를 열고 검색어를 입력합니다. 이 옵션을 사용할 경우 파일 경로도 볼 수 있습니다. 검색어를 입력한 후 창을 닫으려면 Esc 키를 누르거나 X 버튼을 클릭합니다.

여러 파일 닫기

Windows의 경우, JMP 홈 창에서 여러 파일을 한 번에 닫을 수 있습니다. "창 목록"에서 파일을 선택하고 마우스 오른쪽 버튼을 클릭한 후 **닫기**를 선택합니다. 한 파일만 열어 두고 나머지는 닫으려면 단일 파일을 선택한 후 **이 항목을 제외하고 모두 닫기**를 선택합니다.

macOS의 경우, **파일 > 여러 항목 닫기**를 선택한 후 닫으려는 파일을 선택합니다.

열려 있는 창 표시 및 배열

예를 들어 데이터 테이블, 보고서 및 JMP 홈 창과 같은 여러 개의 창이 JMP에 동시에 열려 있는 경우가 많습니다. JMP에서는 몇 가지 방법으로 이러한 열려 있는 창을 배열하고 표시할 수 있습니다.

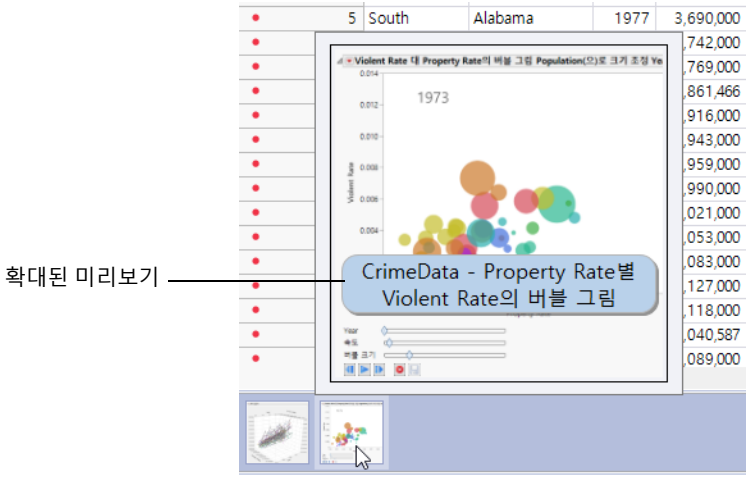
Windows의 경우:

- 창 사이를 전환하려면 Ctrl+Tab 을 누릅니다 .
- "JMP 창 나타내기 " 기능 (F9 또는 창 > JMP 창 나타내기) 을 사용합니다 . 자세한 내용은 " 창 나타내기 기능 사용 "(63 페이지) 에서 확인하십시오 .
- "창" 메뉴의 옵션을 사용합니다. 대부분의 창에서 오른쪽 하단에 나타나는 "배열 메뉴" 옵션을 통해 배열 옵션을 사용할 수도 있습니다 . " 배열 메뉴 ", "JMP 홈 창 " 및 " 연결된 데이터 보기 " 버튼이 표시되지 않으면 **보기 > 상태 표시줄**을 선택합니다 .

  팁 : Windows 의 경우 , " 배열 메뉴 " 에서 " 창 결합 " 을 선택하여 열려 있는 창을 병합할 수 있습니다 . 자세한 내용은 " 대시보드를 사용하여 보고서 결합 및 표시 " 장의 " 예 : 창을 결합하여 대시보드 생성 "(590 페이지) 에서 확인하십시오 .

- JMP 홈 창을 표시하려면 대부분의 창에서 오른쪽 하단에 나타나는 **JMP 홈 창** 버튼을 클릭하거나 , Ctrl+1 을 누릅니다 .
- 보고서의 데이터 테이블을 표시하려면 보고서의 오른쪽 하단에 있는 **연결된 데이터 보기** 버튼을 클릭합니다 .
- 데이터 테이블을 볼 때 연결된 보고서를 열려면 하단 페인에서 보고서의 썸네일 미리보기를 두 번 클릭합니다 . 미리보기를 확대하려면 썸네일을 마우스로 가리킵니다 (그림 2.16).
- 연결된 보고서를 닫으려면 썸네일 미리보기를 마우스 가운데 버튼으로 클릭합니다 .

그림 2.16 열려 있는 보고서의 썸네일 미리보기



열려 있는 보고서의 미리보기

macOS의 경우:

- "창" 메뉴의 옵션을 사용합니다.
- "JMP 창 나타내기" 기능 (F9 또는 F3) 을 사용합니다.

JMP 파일 미리보기

일부 JMP 파일 유형의 경우 Windows 탐색기나 macOS Finder를 사용하여 선택한 파일의 일부분을 볼 수 있습니다. Windows에서는 JMP 데이터 테이블, 저널 및 스크립트의 미리보기가 지원됩니다. macOS에서는 데이터 테이블 및 스크립트의 미리보기가 지원됩니다.

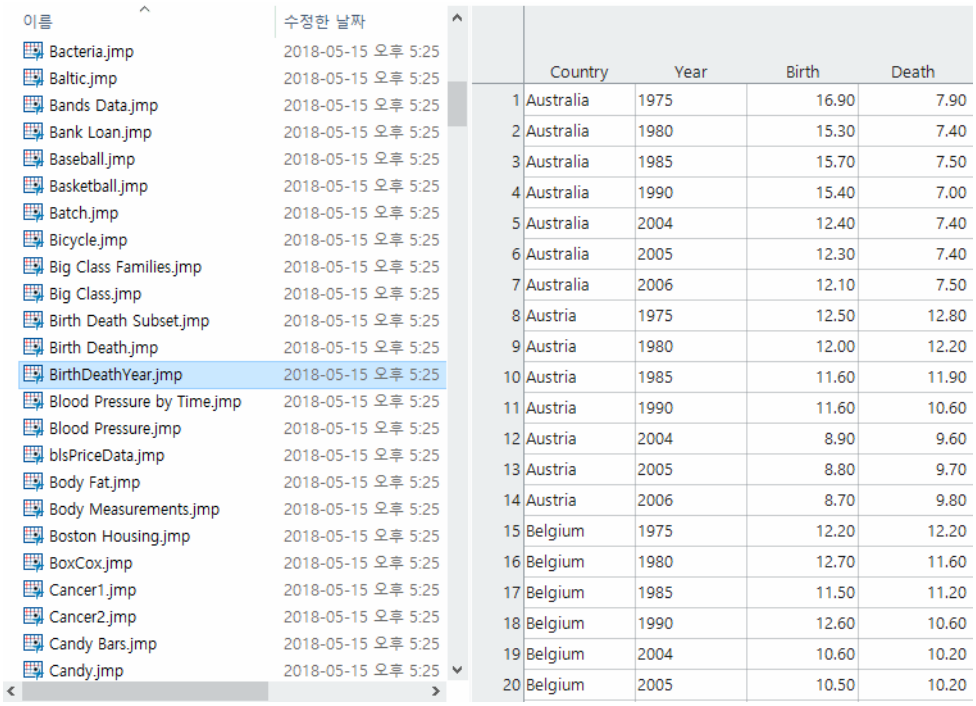
JMP 파일을 미리 보려면

1. Windows 탐색기 또는 macOS Finder 에서 미리 보려는 파일을 선택합니다.
2. Windows 탐색기의 경우, 오른쪽 상단에서 **미리 보기 창 표시** 버튼을 선택합니다. 전자우편 프로그램의 지원은 다양합니다.

   macOS Finder 의 경우, 파일을 선택하고 스페이스바를 누르거나, 팝업 메뉴 또는 "File" 메뉴에서 **Quick Look** 옵션을 선택합니다. Mail 에서는 팝업 메뉴 또는 "File" 메뉴의 **Quick Look Attachment** 옵션을 선택합니다.

   미리보기 페인에 선택한 파일의 일부분이 표시됩니다.

그림 2.17 데이터 테이블 미리보기 페인 (Windows)



이름	수정한 날짜
Bacteria.jmp	2018-05-15 오후 5:25
Baltic.jmp	2018-05-15 오후 5:25
Bands Data.jmp	2018-05-15 오후 5:25
Bank Loan.jmp	2018-05-15 오후 5:25
Baseball.jmp	2018-05-15 오후 5:25
Basketball.jmp	2018-05-15 오후 5:25
Batch.jmp	2018-05-15 오후 5:25
Bicycle.jmp	2018-05-15 오후 5:25
Big Class Families.jmp	2018-05-15 오후 5:25
Big Class.jmp	2018-05-15 오후 5:25
Birth Death Subset.jmp	2018-05-15 오후 5:25
Birth Death.jmp	2018-05-15 오후 5:25
BirthDeathYear.jmp	2018-05-15 오후 5:25
Blood Pressure by Time.jmp	2018-05-15 오후 5:25
Blood Pressure.jmp	2018-05-15 오후 5:25
blsPriceData.jmp	2018-05-15 오후 5:25
Body Fat.jmp	2018-05-15 오후 5:25
Body Measurements.jmp	2018-05-15 오후 5:25
Boston Housing.jmp	2018-05-15 오후 5:25
BoxCox.jmp	2018-05-15 오후 5:25
Cancer1.jmp	2018-05-15 오후 5:25
Cancer2.jmp	2018-05-15 오후 5:25
Candy Bars.jmp	2018-05-15 오후 5:25
Candy.jmp	2018-05-15 오후 5:25

	Country	Year	Birth	Death
1	Australia	1975	16.90	7.90
2	Australia	1980	15.30	7.40
3	Australia	1985	15.70	7.50
4	Australia	1990	15.40	7.00
5	Australia	2004	12.40	7.40
6	Australia	2005	12.30	7.40
7	Australia	2006	12.10	7.50
8	Austria	1975	12.50	12.80
9	Austria	1980	12.00	12.20
10	Austria	1985	11.60	11.90
11	Austria	1990	11.60	10.60
12	Austria	2004	8.90	9.60
13	Austria	2005	8.80	9.70
14	Austria	2006	8.70	9.80
15	Belgium	1975	12.20	12.20
16	Belgium	1980	12.70	11.60
17	Belgium	1985	11.50	11.20
18	Belgium	1990	12.60	10.60
19	Belgium	2004	10.60	10.20
20	Belgium	2005	10.50	10.20

창 나타내기 기능 사용

"JMP 창 나타내기" 기능은 최상위 창과 열려 있는 모든 JMP 창의 썸네일을 표시합니다. Windows의 경우, F9 키를 눌러 "JMP 창 나타내기" 기능을 사용할 수 있습니다.

- 창을 활성화하려면 해당 창을 클릭합니다.
- 미리보기 모드를 설정하려면 스페이스바를 누릅니다. 그러면 커서 아래에 있는 썸네일의 전체 크기 보기가 화면 가운데에 표시됩니다.
- 기본적으로 창은 화면의 원래 순서와 매칭되도록 위에서 아래, 왼쪽에서 오른쪽 순서로 정렬됩니다. 썸네일을 이름순으로 정렬하려면 S 키를 누릅니다.

  "JMP 창 나타내기" 기능을 종료하려면 F9 또는 Esc 키를 누르거나, Alt+Tab 을 누른 후 다른 응용 프로그램을 선택합니다.

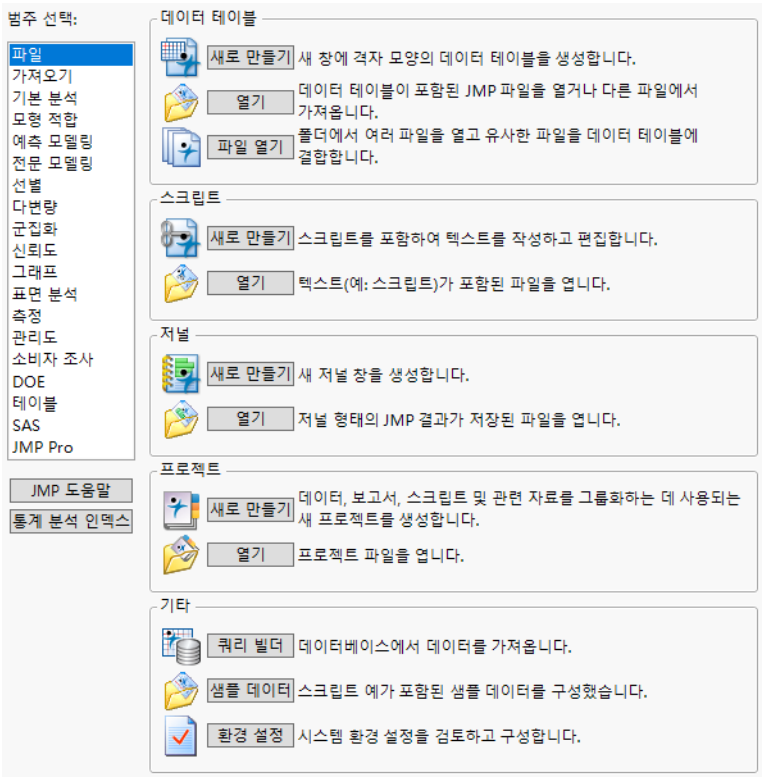
JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기" 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기" 창에는 "분석", "그래프", "테이블" 및 "파일" 메뉴에 있는 다양한 옵션이 포함됩니다.

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반** 을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 의 경우 , **JMP > 환경 설정 > 일반** 페이지에서 이 옵션을 선택합니다 .

그림 2.18에서는 Windows의 "JMP 시작하기" 창을 보여 줍니다. macOS의 "JMP 시작하기" 창도 이와 동일합니다.

그림 2.18 JMP 시작하기 창 (Windows)



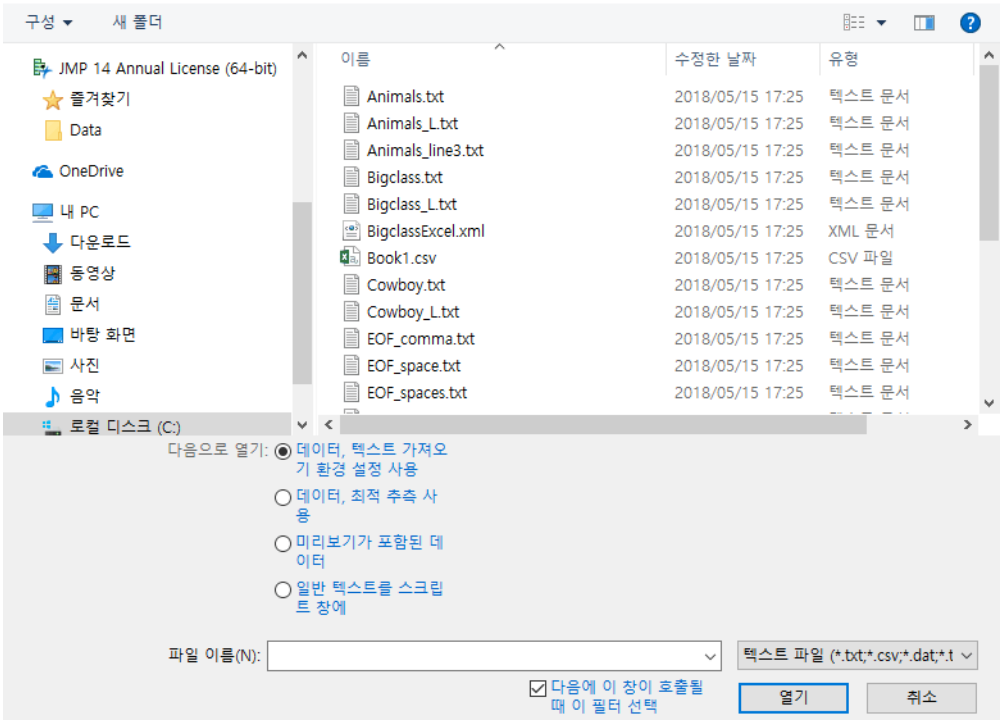
3 장

데이터 가져오기
데이터 테이블 생성

이 장에서는 다음과 같은 주제를 다룹니다.

- 텍스트 파일 , PDF 파일 , SPSS 파일 및 SAS 파일과 같은 데이터를 가져오는 방법
- Excel 데이터를 전송하는 방법
- 실시간 데이터를 읽는 방법
- 새 데이터 테이블을 생성하는 방법

그림 3.1 텍스트 파일 가져오기



목차

데이터 가져오기 정보 68
Microsoft Excel 파일 가져오기 69
Microsoft Excel 데이터 미리 보기 및 가져오기 69
Microsoft Excel 파일 직접 가져오기 80
텍스트 파일 가져오기 81
가져온 데이터에서 결측 문자 찾기 83
JMP 텍스트 가져오기 마법사의 옵션 84
JMP로 가져오기 전에 텍스트 편집 88
스크립트 편집기에서 텍스트 가져오기 90
여러 파일 가져오기 91
PDF 파일 가져오기 96
여러 테이블을 하나의 데이터 테이블로 연결 96
열 머리글을 기준으로 테이블 연결 98
JMP가 자동 감지하지 않는 테이블 가져오기 99
PDF 미리보기의 빨간색 삼각형 옵션 102
데이터 테이블 미리보기 옵션 102
PDF 미리보기의 마우스 오른쪽 버튼 클릭 옵션 103
XML 및 JSON 파일 가져오기 103
XML 및 JSON 데이터 미리보기 및 가져오기 103
XML 파일 바로 가져오기 109
JSON 파일 바로 가져오기 109
쿼리 빌더에서 SQL 쿼리 작성 110
SQL 데이터베이스에 연결 111
SQL 데이터베이스에서 테이블 선택 112
SQL 쿼리 생성 116
JMP 쿼리 저장 및 실행 129
선택한 데이터 열기 130
쿼리 빌더의 빨간색 삼각형 메뉴 옵션 130
쿼리 빌더에서 SQL 문 작성 131
SAS 에서 데이터 가져오기 132
JRE(Java Runtime Environment) 요구 사항 132
SAS 메뉴 옵션 132
SAS 데이터 집합 가져오기 132
JMP에서 SAS의 SAS 전송 파일 생성 135
JMP에서 SAS에 연결 135
SAS 쿼리 빌더를 사용하여 SAS 데이터 집합 열기 141
SAS 서버를 통해 SAS 데이터 집합 열기 142
SAS 확장 속성을 사용하여 메타데이터 가져오기 149

스토어드 프로세스 실행 . . . 150
SAS 코드 전송 . . . 151
ODS 결과 생성 . . . 153
생성된 SAS 데이터 집합 가져오기 . . . 154
데이터베이스에서 데이터 가져오기 . . . 156
데이터베이스에 연결 . . . 157
데이터베이스에서 데이터 열기 . . . 157
SQL 문을 작성하여 데이터베이스 쿼리 . . . 160
SQL(Structured Query Language): 참조 . . . 161
Google 스프레드시트에서 데이터 가져오기 . . . 167
Google 스프레드시트 가져오기 . . . 167
Google 스프레드시트 설정 . . . 168
SPSS 파일 가져오기 . . . 168
Triple-S 설문 조사 데이터 가져오기 . . . 170
HDF5 데이터 가져오기 . . . 170
MATLAB 에서 데이터 가져오기 . . . 171
R 에서 데이터 가져오기 . . . 171
Excel 추가기능을 사용하여 데이터 가져오기 . . . 171
Excel 데이터를 데이터 테이블로 전송 . . . 172
Excel용 JMP 추가기능 . . . 173
JMP Excel 추가기능 제거 . . . 174
인터넷 또는 원격 컴퓨터에서 데이터 가져오기 . . . 174
새 데이터 테이블 생성 . . . 175

데이터 가져오기 정보

다양한 파일 형식을 JMP로 가져와 데이터 테이블로 저장할 수 있습니다. JMP에서는 기본적으로 다양한 파일이 열립니다. JMP에서 기본적으로 지원되지 않는 파일 형식의 경우 ODBC(Open Database Connectivity) 드라이버가 필요합니다.

기본적으로 지원되는 파일 형식은 다음과 같습니다 .

- 쉼표로 구분된 값 (.csv)
- 텍스트로 구성된 .dat 파일
- ESRI 쉐이프 파일 (.shp)
- Flow Cytometry 버전 2.0 및 3.0(.fcs)
- Hierarchical Data Format, 버전 5(.h5)
- HTML(.htm, .html)
- JSON(.json)
- MATLAB(.m, .M)
- macOS 용 Microsoft Excel 1997 ~ 2019(.xls, .xlsx)
- Windows 용 Microsoft Excel 2010 ~ 2019(*.xlsx, *.xlsm)
- Minitab Portable Worksheet(.mtp)
- 일반 텍스트 (.txt)
- R(.r)
- SAS 전송 형식 (.xpt, .stx)
- macOS 용 SAS 버전 7 ~ 9(.sas7bdat)
- Windows 용 SAS 버전 7 ~ 9(.sas7bdat, .sas7bxat)
- SPSS(.sav)
- SQLite 3.0 이상 (.sqlite, .db, .sqlite3, .db3)
- 탭으로 구분된 값 (.tsv)
- Teradata 데이터베이스 (.trd)
- Triple-S(.sss, .xml)
- xBase 데이터 파일 (.dbf)
- XML 데이터 파일 (.xml)

SAS 지원에 관한 노트 :

Windows와 macOS 모두에서 **파일 > 열기** 명령을 통해 SAS 데이터 집합을 직접 열 수 있습니다. 자세한 내용은 "SAS 데이터 집합 가져오기"(132 페이지)에서 확인하십시오.

또는 **파일 > SAS > 데이터 불러오기**를 선택하여 SAS 서버에 연결할 수도 있습니다. 자세한 내용은 "SAS 서버를 통해 SAS 데이터 집합 열기"(142 페이지)에서 확인하십시오.

ODBC 드라이버가 필요한 파일 :

- 데이터베이스 (dBASE)(.ndx, .mdx) 는 V3+ 규격 ODBC 드라이버를 통해 지원됩니다 . .dbf 파일에는 ODBC 드라이버가 필요하지 않습니다 .
- Microsoft Access 데이터베이스 (.mdb) 는 V3+ 규격 ODBC 드라이버를 통해 지원됩니다 .

데이터베이스 작업에 대한 자세한 내용은 " 데이터베이스에서 데이터 가져오기"(156 페이지)에서 확인하십시오.

데이터 가져오기는 사용 가능한 컴퓨터 메모리의 영향을 받습니다. 파일 크기가 매우 크면 로드 속도가 느리거나 전혀 로드되지 않을 수 있습니다. 따라서 큰 파일은 가져오기 전에 분할하는 것이 좋습니다. 그런 다음 테이블을 결합하거나 연결할 수 있습니다. 자세한 내용은 " 데이터 재구성 " 장의 " 데이터 테이블 연결"(328 페이지) 및 " 데이터 재구성 " 장의 " 데이터 테이블 결합"(331 페이지)에서 확인하십시오.

참고 : JMP 에서 R 코드 (.R) 및 SAS 프로그램 파일 (.sas) 을 열 수 있지만 해당 텍스트는 데이터 테이블이 아니라 " 스크립트 " 창에서 열립니다 .

Microsoft Excel 파일 가져오기

Microsoft Excel 파일은 기본적으로 Excel 가져오기 마법사에서 열립니다. 마법사에는 데이터 미리보기가 표시됩니다. 마법사를 사용하면 데이터를 가져오기 전에 설정을 수정할 수 있습니다. 예를 들어 데이터가 시작되는 행과 워크시트에 열 머리글이나 숨겨진 행 또는 열이 포함되어 있는지 여부를 지정할 수 있습니다. Microsoft Excel .xls, .xlsm 및 .xlsx 파일 형식이 지원됩니다.

마법사를 사용하지 않고 Microsoft Excel 파일을 여는 방법에 대한 자세한 내용은 "Microsoft Excel 파일 직접 가져오기"(80 페이지)에서 확인하십시오.

참고 :

- 가져온 데이터의 소수점 이하 자릿수 및 날짜 형식은 Windows 와 macOS 에서 서로 다를 수 있습니다 . 예를 들어 "10/25/2012" 가 macOS 에서는 "25Oct2012" 형식이 될 수 있습니다 . macOS 에서는 열을 문자 열로 가져올 수 있지만 Windows 에서는 그럴 수 없습니다 .
- " 일반 " 환경 설정에서 " 자동 저장 타임아웃 " 값을 설정하여 열려 있는 데이터 테이블을 자동으로 저장하는 것이 좋습니다 . 이 자동 저장 값은 다른 JMP 문서 유형에도 적용됩니다 .

Microsoft Excel 데이터 미리 보기 및 가져오기

워크시트를 가져오기 전에 Excel 에서 스프레드시트를 열고 최종 데이터 테이블에서 데이터를 구성할 방식을 결정하십시오. 예를 들어 워크시트에 숨겨진 셀 또는 병합된 셀이 포함되어 있는지 여부를 알고 있어야 합니다. 그러면 마법사에서 숨겨진 열 또는 행을 제외할 수 있습니다.

워크시트가 여러 개 포함된 Microsoft Excel 파일을 가져오려면 다음 단계를 수행하십시오 .

1. Microsoft Excel 에서 워크시트를 엽니다 .

   이 예의 그림에서는 JMP Samples/Import Data 폴더에 있는 Team Results.xlsx 파일을 보여 줍니다 . 이 파일에는 다음과 같은 특성이 있습니다 .

   – 데이터는 4 행 , 2 열에서 시작되고 9 행 , 5 열에서 끝납니다 .
   – 워크시트가 두 개입니다 .
   – 두 번째 워크시트에는 병합된 셀 집합이 두 개 있습니다 .
   – 숨겨진 행 또는 열은 없습니다 .

그림 3.2 Team Results.xlsx 워크시트

	A	B	C	D	E	F
1		Here are the team results for this year's game.				
2						
3		**Team**	**Member Name**	**Age**	**Winnings**	
4		1	Joe	21	$50.22	
5		1	Mary	22	$24.34	
6		1	Cindy	23	$25.54	
7		2	Mark	22	$52.11	
8		2	Bill	23	$43.32	
9		2	Jennifer	24	$11.23	

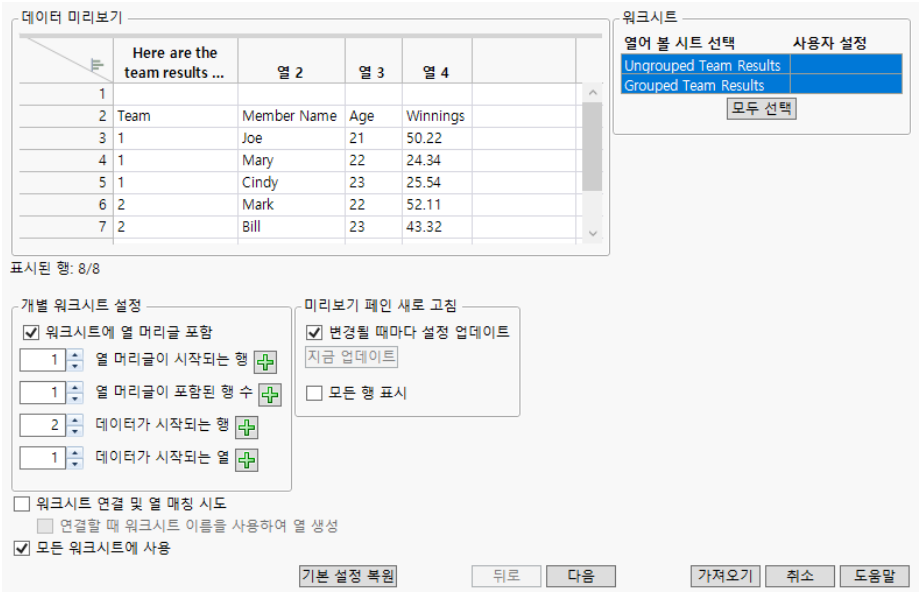
2. Excel 파일을 열려면 **파일 > 열기**를 선택합니다 .

   " 데이터 파일 열기 " 창이 나타납니다 .

3. " 파일 이름 " 옆에서 **Excel 파일**을 선택합니다 .
4. Excel 파일을 선택하고 **열기**를 클릭합니다 .

   워크시트가 Excel 가져오기 마법사에서 열리며 가져오기 옵션과 함께 데이터 미리보기가 나타납니다 .

그림 3.3 초기 데이터 미리보기의 예



이 데이터 미리보기에는 다음과 같은 특성이 있습니다 .

- 오른쪽 상단에 두 개의 워크시트가 모두 선택되어 있습니다 .
- 첫 번째 열은 자동으로 제거되어 있습니다 .
- 워크시트의 첫 번째 행에 있던 텍스트는 열 머리글로 나타납니다 . 하지만 워크시트의 3 행 텍스트를 열 머리글로 사용하려고 합니다 .
- 첫 번째 데이터 행은 비어 있습니다 .

참고 : JMP 는 사용자가 워크시트를 가져올 때마다 저장된 이전 변경 사항을 적용하며 이는 JMP 를 닫았다가 다시 연 경우에도 해당됩니다 . 이 기능은 동일한 워크시트를 여러 번 다시 가져와서 여러 옵션으로 시험해 보는 경우에 유용합니다 . 다른 워크시트를 가져올 때 변경 사항을 지우려면 **기본 설정 복원**을 클릭하십시오 .

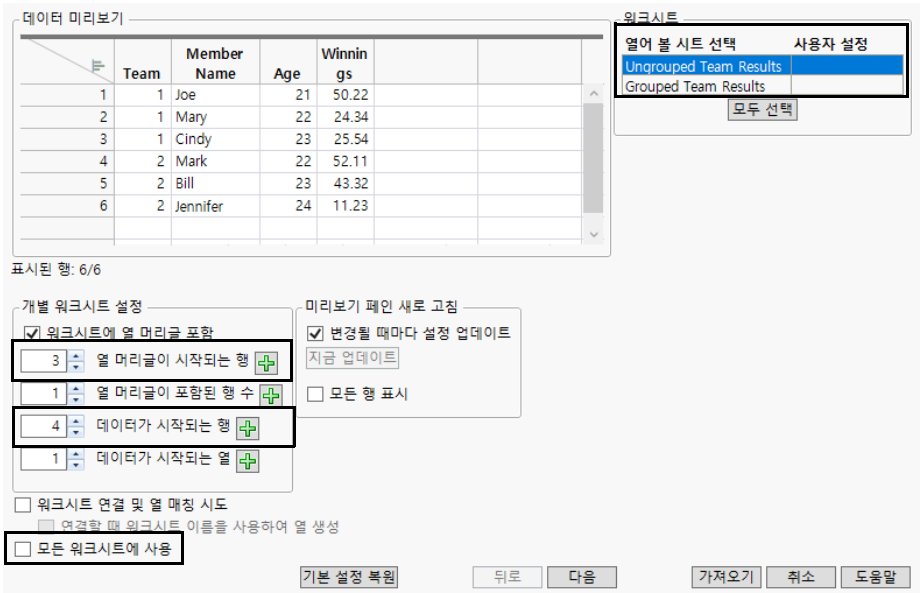
---

5. **열 머리글이 시작되는 행**에 3 을 입력합니다 .
6. **데이터가 시작되는 행**이 4 로 변경됩니다 .
7. " 워크시트 " 페인에서 **Ungrouped Team Results** 를 선택합니다 .

   이 워크시트만 가져오게 됩니다 .
8. **모든 워크시트에 사용**을 선택 취소합니다 .

   "Ungrouped Team Results" 에만 이 설정이 적용됩니다 .

그림 3.4 열 머리글 행 선택



팁 : 데이터 미리보기 페인에서 숫자 열 머리글을 마우스 오른쪽 버튼으로 클릭하여 형식을 변경할 수 있습니다 .

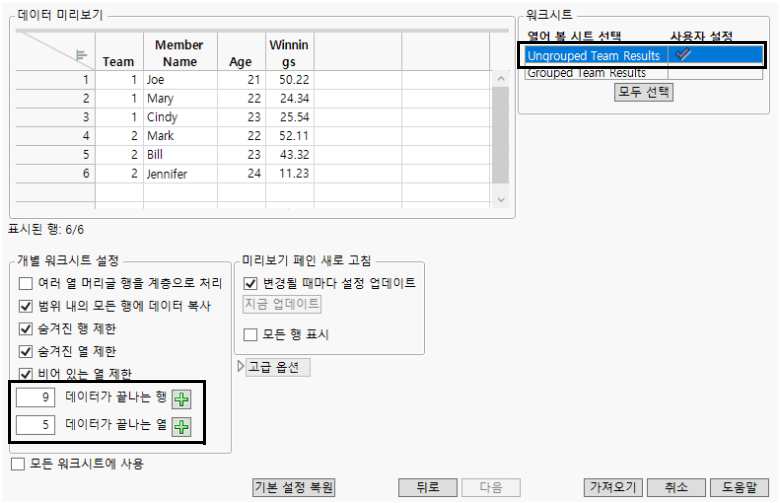
모든 옵션에 대한 자세한 내용은 " 개별 워크시트 설정 "(73 페이지) 에서 확인하십시오 .

9. **다음**을 클릭하여 다른 가져오기 설정을 구성합니다 .

   창에 추가 가져오기 설정이 표시됩니다 .

10. **데이터가 끝나는 행**에 9 를 입력합니다 .
11. **데이터가 끝나는 열**에 5 를 입력합니다 .

그림 3.5 마지막 열 지정

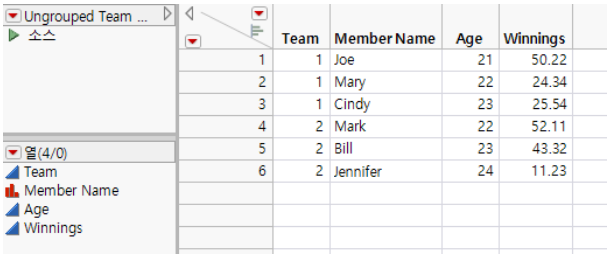


가져오기 옵션을 수정하면 사용자 설정이 자동으로 선택됩니다.

모든 옵션에 대한 자세한 내용은 "추가 개별 워크시트 설정"(74 페이지)에서 확인하십시오.

12. **가져오기**를 클릭하여 워크시트를 지정한 대로 변환합니다.

그림 3.6 최종 데이터 테이블



다음 섹션에서는 Excel 가져오기 마법사의 옵션을 설명합니다.

개별 워크시트 설정

워크시트에 열 머리글 포함 워크시트에 열 머리글이 있는 행이 포함되어 있으면 이 옵션을 선택합니다.

열 머리글이 시작되는 행 워크시트에서 열 머리글이 시작되는 행을 나타냅니다. 올바른 행에서 머리글이 시작될 때까지 위쪽 화살표 를 클릭하거나, 해당 행 번호를 입력하고 Enter 키를 누릅니다.

열 머리글이 포함된 행 수 워크시트에 열 머리글로 사용되는 행이 여러 개 있는지 여부를 나타냅니다. 머리글 행이 올바르게 나타날 때까지 위쪽 화살표 를 클릭하거나, 해당 행 수를 입력하고 Enter 키를 누릅니다.

데이터가 시작되는 행 워크시트에서 데이터가 시작되는 행을 나타냅니다.

데이터가 시작되는 열 워크시트에서 데이터가 시작되는 열을 나타냅니다 .

워크시트 연결 및 열 매칭 시도 모든 워크시트를 하나의 데이터 테이블에 병합합니다 . JMP 는 머리글이 동일한 열을 매칭합니다 .

연결할 때 워크시트 이름을 사용하여 열 생성 가져온 각 테이블의 워크시트 이름을 나열하는 새 소스 테이블 열을 추가합니다 . 이 옵션은 앞의 연결 옵션을 선택한 후에 사용할 수 있습니다 .

모든 워크시트에 사용 현재 가져오기 설정을 오른쪽 상단에서 선택된 모든 워크시트에 적용합니다 .

추가 개별 워크시트 설정

여러 열 머리글 행을 계층으로 처리 워크시트에 열 머리글로 사용되는 행이 여러 개 포함되어 있고 이러한 머리글을 계층으로 처리하려고 함을 나타냅니다. 이 옵션은 쌓기 가능한 데이터에만 적용됩니다 .

범위 내의 모든 행에 데이터 복사 워크시트에 여러 행에서 병합된 셀이 있음을 나타냅니다 . JMP 는 해당 셀을 분리하고 셀 내용을 모든 결과 셀에 복사합니다 . 이 옵션은 기본적으로 선택됩니다 .

범위 내의 모든 행에 데이터 복사를 선택 취소하면 JMP 는 해당 셀을 분리하고 셀 내용을 최상위 셀에 복사합니다 . 분리된 나머지 셀은 빈 상태로 남습니다 .

숨겨진 행 제한 숨겨진 행이 데이터 테이블에 나타나지 않도록 합니다 .

숨겨진 열 제한 숨겨진 열이 데이터 테이블에 나타나지 않도록 합니다 .

비어 있는 열 제한 열 머리글이 있는 빈 열을 가져올지 여부를 나타냅니다 . 해당 열을 가져오려면 이 옵션을 선택 취소합니다 .

데이터가 끝나는 행 워크시트에서 데이터가 포함된 마지막 행을 나타냅니다 .

데이터가 끝나는 열 워크시트에서 데이터가 포함된 마지막 열을 나타냅니다 .

고급 옵션

열 이름 구분 기호 문자열 머리글이 원래 서로 다른 행에 있었던 경우 열 머리글의 각 단어 간 구분 기호를 나타냅니다 . 열 머리글이 포함된 행의 수는 Excel 가져오기 마법사의 첫 번째 창에서 지정합니다 . 그런 다음 " 열 이름 구분 기호 문자열 " 상자에 문자 또는 공백을 입력합니다 . 기본 문자열인 하이픈을 사용하면 열 머리글이 예를 들어 "First-Second-Third" 와 같은 형식이 됩니다 .

다중 계열 쌓기 계층적 머리글이 있는 워크시트에서 하위 범주를 개별 열로 나눕니다. **여러 열 머리글 행을 계층으로 처리**도 선택해야 합니다 . 주 범주는 라벨 열로 가져오게 됩니다 .

범위 내의 모든 행에 머리글 복사 워크시트에 병합된 행이 있는 경우 각 셀에 머리글 텍스트를 반복합니다 . 예를 들어 그림 3.7 에 표시된 "State" 열 머리글은 워크시트에서 병합된 셀에 있었습니다 . 왼쪽에서는 머리글이 복사되어 있습니다 . 오른쪽에 표시된 것처럼 머리글을 반복하지 않으려면 이 옵션을 선택 취소합니다 .

그림 3.7 Microsoft Excel 파일의 복사된 머리글

복사된 머리글

	State-State
1	TX
2	TX
3	TX
4	TX

복사되지 않은 머리글

	State
1	TX
2	TX
3	TX
4	TX

셀 색상 가져오기 워크시트의 셀 색상을 데이터 테이블에 적용합니다 . macOS 에서는 기본 색상과 보조 색상만 확실하게 가져올 수 있습니다 .

열 유형 감지 제한 최대 100 개의 행을 스캔하여 열 유형을 확인합니다 . 큰 워크시트를 빠르게 가져오려면 이 옵션을 선택합니다 .

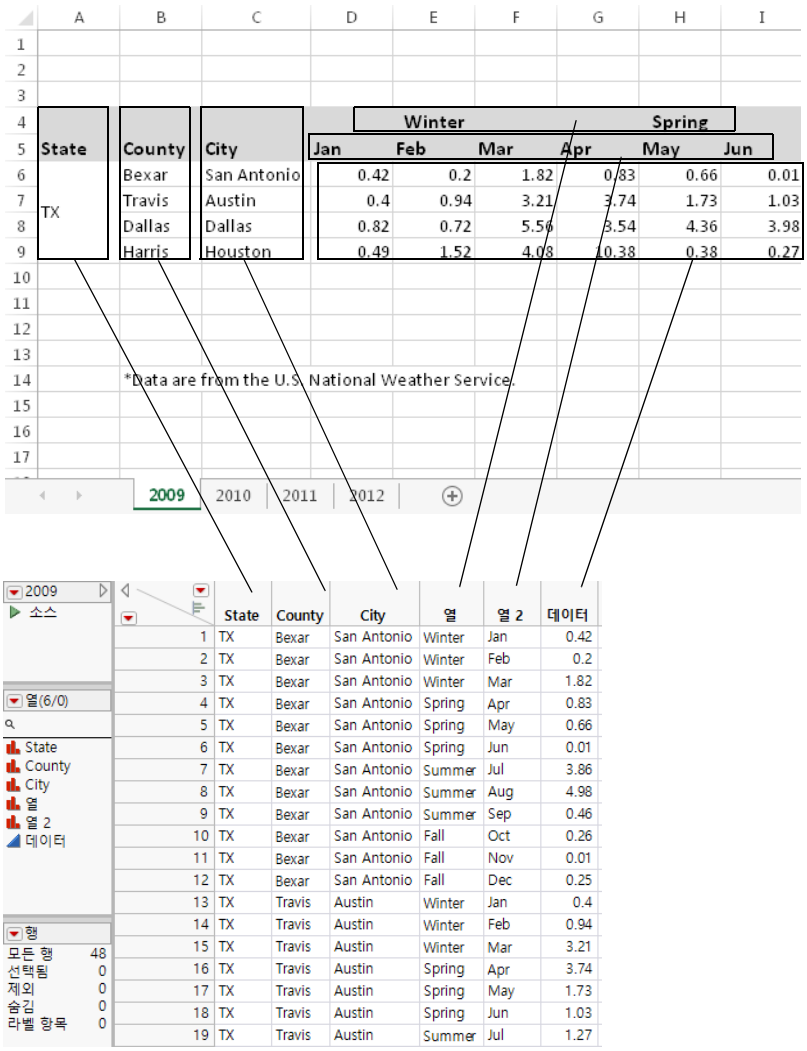
팁 :

- JMP 는 사용자가 워크시트를 가져올 때마다 저장된 이전 변경 사항을 적용하며 이는 JMP 를 닫았다가 다시 연 경우에도 해당됩니다 . 이 기능은 동일한 워크시트를 여러 번 다시 가져와서 여러 옵션으로 시험해 보는 경우에 매우 유용합니다 . 다른 워크시트를 가져올 때 변경 사항을 지우려면 **기본 설정 복원**을 클릭하십시오 .
- 가져오기 설정은 소스라는 데이터 테이블 스크립트에 저장됩니다 . 동일한 설정을 사용하여 워크시트를 다시 가져오려면 이 스크립트를 실행합니다 . 이 스크립트에는 워크시트 경로가 포함되어 있으므로 다른 사용자가 해당 위치에 액세스할 수 있는지 확인하십시오 .
- 큰 워크시트에서 데이터 미리보기 속도를 높이려면 마법사의 첫 번째 창에서 **변경될 때마다 설정 업데이트**를 선택 취소합니다 . 설정을 수정한 후 **지금 업데이트**를 클릭하면 데이터 미리보기가 새로 고쳐집니다 .
- 데이터 미리보기 페인에서 모든 행을 보려면 **모든 행 표시**를 선택합니다 . 스프레드시트의 크기에 따라 미리보기가 다소 지연될 수도 있습니다 .
- 동일한 통합 문서에 있는 두 개의 워크시트를 하나의 데이터 테이블로 결합할 수 있습니다 . 가져올 때 열 이름이 매칭되므로 열 순서는 중요하지 않습니다 .

계층적 머리글이 있는 Microsoft Excel 파일 가져오기

Excel 워크시트에서 여러 머리글 행에 계층이 적용되어 있을 수 있습니다. 즉, 두 번째 머리글 행에 첫 번째 머리글 행의 범주에 해당하는 데이터가 포함되어 있을 수 있습니다. 그림 3.8에서는 그러한 예를 보여 줍니다. 워크시트의 맨 위를 보면 계절 "Winter" 및 "Spring"은 해당 계절의 월 위에 있는 여러 셀에 병합되어 있습니다. JMP 데이터 테이블에서는 계절을 하나의 열로 만들고 해당하는 월을 또 다른 열로 만들려고 합니다.

그림 3.8 2012 년에 대한 Excel 의 원래 데이터와 JMP 의 최종 데이터



	State	County	City	Winter			Spring		
				Jan	Feb	Mar	Apr	May	Jun
6	TX	Bexar	San Antonio	0.42	0.2	1.82	0.83	0.66	0.01
7		Travis	Austin	0.4	0.94	3.21	3.74	1.73	1.03
8		Dallas	Dallas	0.82	0.72	5.56	3.54	4.36	3.98
9		Harris	Houston	0.49	1.52	4.08	10.38	0.38	0.27

*Data are from the U.S. National Weather Service.

2009 | 2010 | 2011 | 2012

2009
소스

열(6/0)
State
County
City
열
열 2
데이터

행
모든 행 48
선택됨 0
제외 0
숨김 0
라벨 항목 0

	State	County	City	열	열 2	데이터
1	TX	Bexar	San Antonio	Winter	Jan	0.42
2	TX	Bexar	San Antonio	Winter	Feb	0.2
3	TX	Bexar	San Antonio	Winter	Mar	1.82
4	TX	Bexar	San Antonio	Spring	Apr	0.83
5	TX	Bexar	San Antonio	Spring	May	0.66
6	TX	Bexar	San Antonio	Spring	Jun	0.01
7	TX	Bexar	San Antonio	Summer	Jul	3.86
8	TX	Bexar	San Antonio	Summer	Aug	4.98
9	TX	Bexar	San Antonio	Summer	Sep	0.46
10	TX	Bexar	San Antonio	Fall	Oct	0.26
11	TX	Bexar	San Antonio	Fall	Nov	0.01
12	TX	Bexar	San Antonio	Fall	Dec	0.25
13	TX	Travis	Austin	Winter	Jan	0.4
14	TX	Travis	Austin	Winter	Feb	0.94
15	TX	Travis	Austin	Winter	Mar	3.21
16	TX	Travis	Austin	Spring	Apr	3.74
17	TX	Travis	Austin	Spring	May	1.73
18	TX	Travis	Austin	Spring	Jun	1.03
19	TX	Travis	Austin	Summer	Jul	1.27

워크시트를 가져오고 여러 열 계층을 유지하려면 다음 단계를 수행하십시오 .

1. JMP 에서 **파일 > 열기**를 선택합니다 .
2. " 데이터 파일 열기 " 창에서 JMP Samples/Import Data 폴더에 있는 Texas Precipitation.xlsx 를 선택한 후 **열기**를 클릭합니다 .

   워크시트가 Excel 가져오기 마법사에서 열리며 가져오기 옵션과 함께 데이터 미리보기가 나타납니다 .

그림 3.9 Excel 가져오기 마법사 미리보기

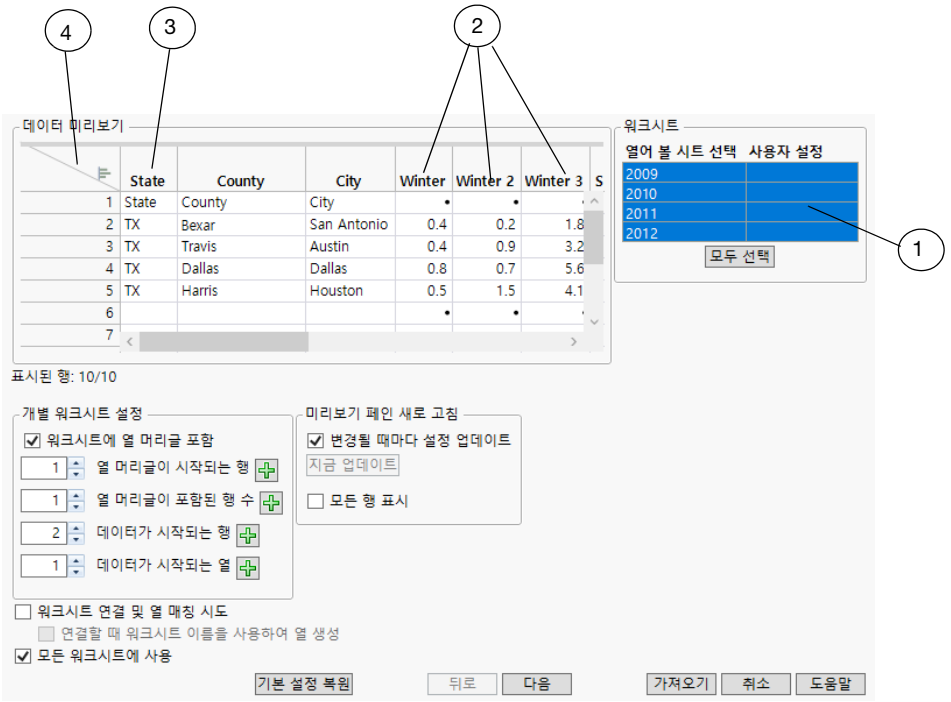


그림 3.9에서는 이 워크시트의 기본 설정을 보여 줍니다.

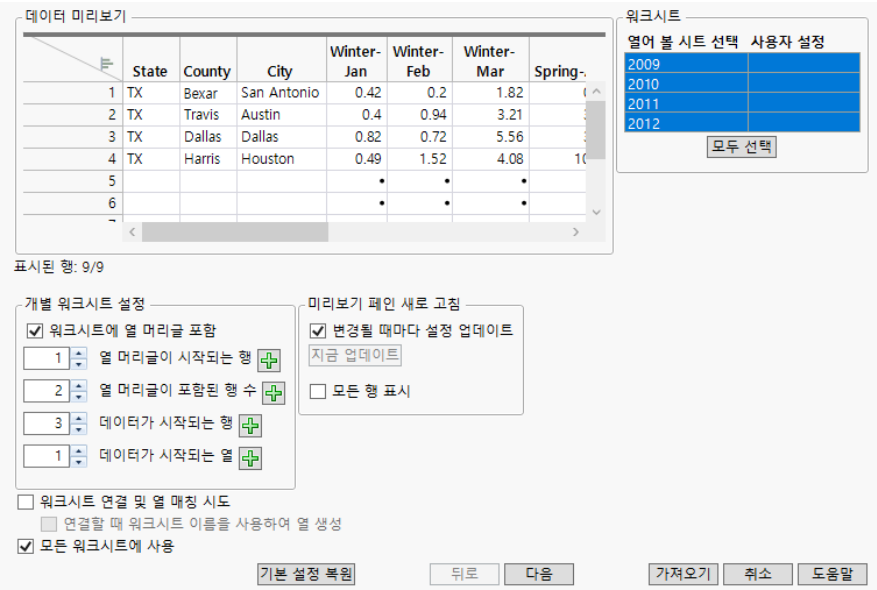
1	모든 워크시트가 가져올 대상으로 선택되어 있습니다. Ctrl 키를 누른 채 워크시트를 클릭하여 가져오기에서 제외할 수 있습니다.
2	각 계절이 여러 열로 분할되었습니다.
3	워크시트의 두 번째 머리글 행 데이터가 첫 번째 행에 나타납니다.
4	맨 위의 빈 행은 제거되었습니다.

3. " 미리보기 페인 새로 고침 " 영역에서 **변경될 때마다 설정 업데이트**가 선택되어 있는지 확인합니다 .

   그러면 변경 사항이 있을 때마다 데이터 미리보기가 자동으로 새로 고쳐집니다 .
4. **열 머리글이 포함된 행 수** 옆의 위쪽 화살표를 한 번 클릭합니다 .

   데이터가 시작되는 행이 자동으로 3 으로 업데이트됩니다 .

그림 3.10 1 페이지의 업데이트된 설정

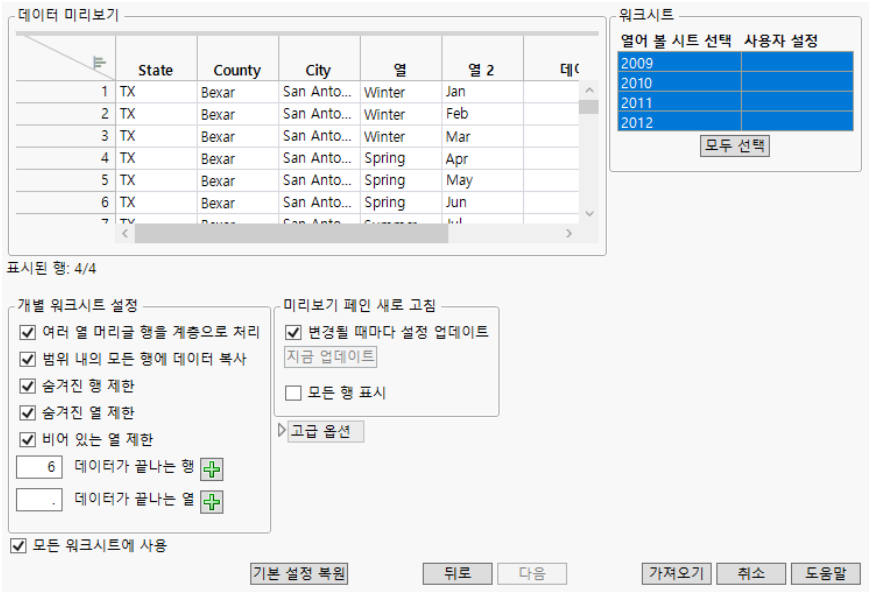


5. **다음**을 클릭합니다 .
6. **데이터가 끝나는 행** 옆에 6 을 입력하고 Enter 키를 누릅니다 .

   워크시트에서 처음 세 개의 빈 행이 제거된 후 데이터가 6 행에서 끝납니다 .

팁 : 끝 행 번호를 입력하는 대신 JMP 에서 행을 자동으로 계산하도록 할 수 있습니다 . 데이터 미리보기 페인에서 4 행을 선택합니다 . 그런 다음 **데이터가 끝나는 행** 옆의 더하기 기호 를 클릭합니다 .

그림 3.11 2 페이지의 업데이트된 설정



범위 내의 모든 행에 데이터 복사는 기본적으로 선택되어 있습니다 . JMP 는 워크시트에서 병합되어 있던 데이터를 병합 해제하고 해당 셀 내용을 개별 범주로 복사합니다 . "TX" 는 원래 병합된 셀에 있었습니다 . JMP 에서는 해당 내용이 개별 셀에 복사됩니다 .

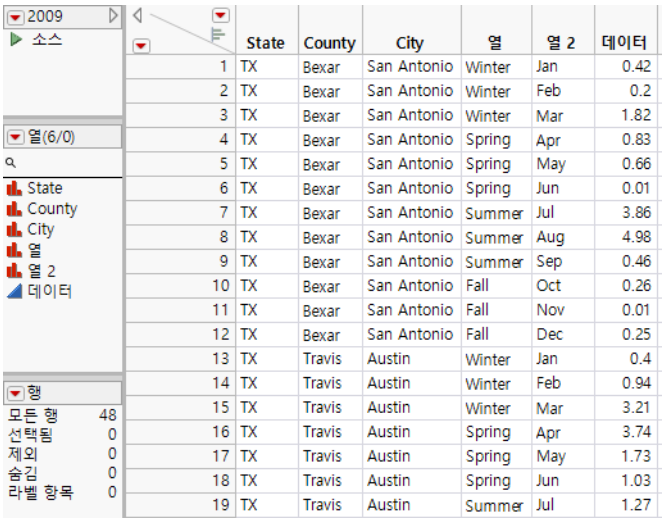
7. **여러 열 머리글 행을 계층으로 처리**를 선택합니다 .

   워크시트의 머리글 행에 있던 각 계절 및 월이 데이터 테이블에서 범주가 됩니다 .

8. **가져오기**를 클릭합니다 .

   네 개의 워크시트가 개별 데이터 테이블로 열립니다 . 데이터 테이블에는 " 소스 " 스크립트가 포함되어 있으며 , 이 스크립트를 실행하면 동일한 가져오기 설정을 사용하여 데이터를 새 데이터 테이블로 가져올 수 있습니다 (그림 3.12). 가져오기 설정은 다음에 JMP 세션에서 Excel 마법사로 워크시트를 열 때도 저장됩니다 .

그림 3.12 최종 데이터



2009
소스

열(6/0)

State
County
City
열
열 2
데이터

행
모든 행 48
선택됨 0
제외 0
숨김 0
라벨 항목 0

	State	County	City	열	열 2	데이터
1	TX	Bexar	San Antonio	Winter	Jan	0.42
2	TX	Bexar	San Antonio	Winter	Feb	0.2
3	TX	Bexar	San Antonio	Winter	Mar	1.82
4	TX	Bexar	San Antonio	Spring	Apr	0.83
5	TX	Bexar	San Antonio	Spring	May	0.66
6	TX	Bexar	San Antonio	Spring	Jun	0.01
7	TX	Bexar	San Antonio	Summer	Jul	3.86
8	TX	Bexar	San Antonio	Summer	Aug	4.98
9	TX	Bexar	San Antonio	Summer	Sep	0.46
10	TX	Bexar	San Antonio	Fall	Oct	0.26
11	TX	Bexar	San Antonio	Fall	Nov	0.01
12	TX	Bexar	San Antonio	Fall	Dec	0.25
13	TX	Travis	Austin	Winter	Jan	0.4
14	TX	Travis	Austin	Winter	Feb	0.94
15	TX	Travis	Austin	Winter	Mar	3.21
16	TX	Travis	Austin	Spring	Apr	3.74
17	TX	Travis	Austin	Spring	May	1.73
18	TX	Travis	Austin	Spring	Jun	1.03
19	TX	Travis	Austin	Summer	Jul	1.27

Microsoft Excel 파일 직접 가져오기

Microsoft Excel 파일은 기본적으로 Excel 가져오기 마법사에서 열립니다. 이 옵션은 워크시트의 데이터 구조가 불규칙한 경우에 유용합니다. 예를 들어 숨겨진 열을 제외하거나 세 번째 행의 텍스트를 열 머리글로 변환할 수 있습니다.

스프레드시트를 Excel 가져오기 마법사에서 여는 대신 **파일 > 열기**를 선택하여 파일을 열 수도 있습니다. 기본적으로 JMP는 첫 번째 행에 라벨이 포함되어 있는지 여부를 감지하고 해당 행을 열 머리글로 변환합니다. 이 설정은 "일반" 환경 설정에서 변경할 수 있습니다. "Excel 열기 방법" 환경 설정은 "파일" > "환경 설정" > "일반"(Windows) 및 "JMP" > "환경 설정" > "일반"(macOS)에 있습니다. "Excel 라벨을 머리글로 사용" 목록에서 **항상** 또는 **안 함**을 선택하십시오.

"Excel 열기 방법 " 환경 설정을 지정하려면

Microsoft Excel 파일을 열 때 항상 마법사를 사용하지 않으려면 "Excel 열기 방법" 환경 설정을 변경하십시오. 모든 워크시트를 한 번에 열거나, 목록에서 워크시트를 선택할 수 있도록 설정하십시오.

Windows 에서 Microsoft Excel 파일을 열려면

1. 위의 설명대로 Excel 열기 방법을 설정한 후에 **파일 > 열기**를 선택합니다 .
2. 첫 번째 행의 텍스트를 열 머리글로 변환하려면 **항상 Excel 1 행을 라벨로 사용** 옆에서 **항상**을 선택합니다 .
3. **Excel 파일** 파일 유형을 선택하거나 , 파일을 선택하거나 , URL 을 입력합니다 .
4. 모든 워크시트를 가져오려면 **열기**를 클릭합니다 .

   또는

열려는 워크시트를 선택하려면 **열기** 버튼의 화살표를 클릭한 후 **선택한 워크시트 열기**를 선택합니다. 워크시트를 하나 이상 선택하고 **확인**을 클릭합니다.

macOS 에서 Microsoft Excel 파일을 열려면

1. **파일 > 열기**를 선택하고 파일을 선택합니다.
2. (.xls 에만 해당) 첫 번째 행의 텍스트를 열 머리글로 변환하려면 **Excel 라벨을 머리글로 사용**을 선택합니다.
3. (.xls 에만 해당) 특정 워크시트를 열려면 **개별 Excel 워크시트 선택**을 선택합니다.
4. **열기**를 클릭합니다.

   특정 워크시트를 열도록 선택한 경우 목록에서 해당 워크시트를 선택한 후 **확인**을 클릭합니다. 마음이 바뀌어서 모든 워크시트를 가져오려는 경우 **모두 선택**을 클릭할 수도 있습니다.

.xlsx 파일을 선택한 경우 Excel 마법사에 데이터 미리보기가 나타납니다. 자세한 내용은 "Microsoft Excel 데이터 미리 보기 및 가져오기"(69 페이지)에서 확인하십시오.

텍스트 파일 가져오기

확장자가 .txt, .csv 및 .tsv인 텍스트 파일을 열 수 있으며, 이때 텍스트는 데이터 테이블로 변환됩니다. 텍스트로 구성되고 확장자가 .dat인 파일도 지원됩니다. 텍스트 파일은 거의 모든 문자를 사용하여 구분될 수 있으며, 고정 너비 파일일 수도 있습니다.

텍스트 파일을 JMP로 가져올 때 후행 공백은 제거됩니다.

JMP가 텍스트 파일에서 9,007,199,254,740,991 보다 큰 정수를 감지하면 해당 열은 문자 데이터로 간주됩니다. 텍스트 가져오기 마법사 창이나 가져오기 스크립트를 사용하여 이 열을 숫자로 설정할 수 있습니다.

팁 : " 여러 파일 가져오기 " 옵션을 사용하면 한 번에 여러 개의 텍스트 파일을 가져올 수 있습니다. 자세한 내용은 " 여러 파일 가져오기 "(91 페이지) 에서 확인하십시오.

가져오기 설정을 조정하려면 (다음 옵션 중 하나 선택)

- **파일 > 환경 설정 > 텍스트 데이터 파일**을 선택하여 JMP 가 데이터 테이블의 구성 및 형식 지정을 위한 최적의 방법을 결정하도록 가져오기 설정을 변경합니다.
- 파일을 열 때 가져오기 설정을 수동으로 선택합니다 (이 섹션에 설명됨).
- 스크립트 편집기에서 파일을 열고, 내용을 편집한 후, 내용을 가져옵니다. 이 옵션은 텍스트 구분자를 추가하거나 텍스트를 수정해야 하는 경우에 유용합니다.

텍스트 파일을 가져오려면

1. **파일 > 열기**를 선택합니다.
2. Windows 의 경우, 파일 유형을 **텍스트 파일**로 설정할 수 있습니다.

3. 열려는 텍스트 파일을 선택합니다 .

   옵션에 대한 자세한 내용은 표 3.1 및 표 3.2 에서 확인하십시오 .

표 3.1 Windows 에서 텍스트 파일 열기

데이터 배열 자동 결정	데이터 배열 수동 지정
1. **파일 > 열기**를 선택합니다 . 2. **파일 이름** 옆의 목록에서 **텍스트 파일**을 선택합니다 . 3. 환경 설정의 가져오기 규칙을 사용하려면 **데이터 (환경 설정 사용)** 를 선택합니다 . 자세한 내용은 "JMP 환경 설정 " 장의 " 텍스트 파일 가져오기 및 내보내기의 환경 설정 "(642 페이지) 에서 확인하십시오 . 텍스트를 가져올 때 JMP 에서 최적의 추측을 통해 데이터를 배열하도록 하려면 **데이터 (최적 추측)** 옵션을 선택합니다 . (선택 사항) 선택한 필터를 기본적으로 적용하려면 **다음에 이 창이 호출될 때 이 필터 선택** 옵션을 선택합니다 . 4. 열려는 파일을 선택합니다 . 5. **열기**를 클릭합니다 . **팁 :** 최근에 해당 파일을 연 경우 위의 단계 대신 JMP 홈 창에서 파일을 빠르게 열 수 있습니다 . " 최근 사용한 파일 " 목록에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **가져오기 (환경 설정)** 또는 **가져오기 (최적 추측)** 를 선택하십시오 . 현재 가져오기 환경 설정은 마우스 오른쪽 버튼 클릭 시 나타나는 메뉴에서 굵게 표시됩니다 .	1. **파일 > 열기**를 선택합니다 . 2. **파일 이름** 옆의 목록에서 **텍스트 파일**을 선택합니다 . 3. 창 하단에서 **열기 형식** 옆의 **데이터 (미리보기 사용)** 를 선택합니다 . 4. 열려는 파일을 선택합니다 . 5. **열기**를 클릭합니다 . 6. 텍스트 가져오기 창을 완료합니다 . 자세한 내용은 "JMP 텍스트 가져오기 마법사의 옵션 "(84 페이지) 에서 확인하십시오 . 7. **가져오기**를 클릭합니다 . **팁 :** 최근에 해당 파일을 연 경우 위의 단계 대신 JMP 홈 창에서 파일을 빠르게 열 수 있습니다 . " 최근 사용한 파일 " 목록에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **가져오기 (미리보기)** 를 선택하십시오 .

표 3.2 macOS 에서 텍스트 파일 열기

데이터 배열 자동 결정	데이터 배열 수동 지정
1. **파일 > 열기**를 선택합니다 . 2. 열려는 파일을 선택합니다 . 3. **열기 형식** 필드에서 **데이터 (최적 추측)** 또는 **데이터 (환경 설정 사용)** 를 선택합니다 . 4. **열기**를 클릭합니다 .	1. **파일 > 열기**를 선택합니다 . 2. 열려는 파일을 선택합니다 . **열기 형식** 필드에서 **데이터 (미리보기 사용)** 를 선택합니다 . 3. **열기**를 클릭합니다 . 4. 텍스트 가져오기 마법사 창을 완료합니다 . 자세한 내용은 "JMP 텍스트 가져오기 마법사의 옵션 "(84 페이지) 에서 확인하십시오 . 5. **가져오기**를 클릭합니다 .

참고 :

- Windows 의 경우 , JMP 에서 컴퓨터의 기본 텍스트 편집기로 텍스트 파일을 열 수 있습니다 . **파일 > 열기**를 선택한 후 **파일 이름** 목록에서 **모든 파일 (*.*)** 을 선택합니다 . 텍스트 파일을 선택한 후 **기본 프로그램을 사용하여 엽니다 . 텍스트로 열려면 선택을 해제하십시오 . 를 선택합니다 .**
- JMP 에서 데이터가 스캔되는 도중에 가져오기를 취소하면 JMP 에서 감지된 사항에 따라 데이터를 가져오게 됩니다 . 따라서 데이터가 예상대로 나타나지 않을 수 있습니다 . 예를 들어 숫자로 잘못 감지된 열에 결측값이 나타날 수 있습니다 . 데이터가 스캔된 **후에** 가져오기를 취소하면 해당 시점까지 읽은 데이터만 가져오게 됩니다 .

스크립트 창에서 텍스트를 가져오는 방법에 대한 자세한 내용은 "스크립트 편집기에서 텍스트 가져오기"(90 페이지) 에서 확인하십시오 .

가져온 데이터에서 결측 문자 찾기

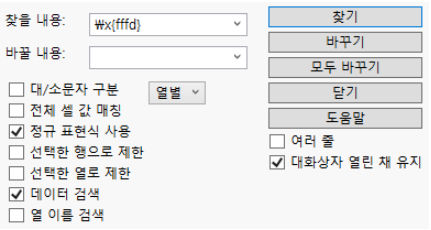
JMP는 문서의 문자 인코딩을 감지하려고 시도합니다. 문서에서 JMP가 인식하지 못하는 인코딩을 사용하거나 JMP가 잘못된 인코딩을 선택하는 경우가 있을 수 있습니다.

JMP가 인식하지 못하는 텍스트는 유니코드 문자 로 대체됩니다. `\x{fffd}`를 검색하여 이 텍스트를 찾습니다.

1. 데이터 테이블에서 **편집 > 검색 > 찾기**를 선택합니다 .
2. **찾을 내용** 옆에 `\x{fffd}` 를 입력합니다 .
3. **정규 표현식 사용**을 선택했는지 확인합니다 .
4. **데이터 검색**을 선택했는지 확인합니다 .
5. **찾기**를 클릭하여 첫 번째 인스턴스를 찾습니다 .

6. 모든 인스턴스를 다른 문자열로 바꿉니다 .

그림 3.13 유니코드 문자 검색



JMP 텍스트 가져오기 마법사의 옵션

JMP에서 지원되는 텍스트 파일을 열면 파일을 데이터 테이블로 열기 전에 테스트 미리보기가 표시될 수 있습니다. 이 옵션을 사용하여 데이터의 배열과 형식을 수동으로 지정할 수 있습니다. 예를 들어 행 끝 문자를 지정하거나 인용 부호를 제거할 수 있습니다.

JMP는 파일의 구조를 감지하고, 구분자 또는 고정 너비 필드를 사용하여 텍스트를 가져오기 위한 옵션을 표시합니다. JMP에서 선택된 파일 구조가 올바르지 않은 경우에는 **구분된 필드** 또는 **고정 너비 필드** 라디오 버튼을 클릭하여 데이터를 올바른 형식으로 가져오십시오. 예를 들어 파일이 실제로는 구분자로 구분되어 있는데 고정 너비 창이 나타날 수 있습니다.

그림 3.14 및 그림 3.15에서는 텍스트 가져오기 마법사 옵션을 보여 줍니다.

그림 3.14 고정 너비 파일을 위한 텍스트 가져오기 마법사

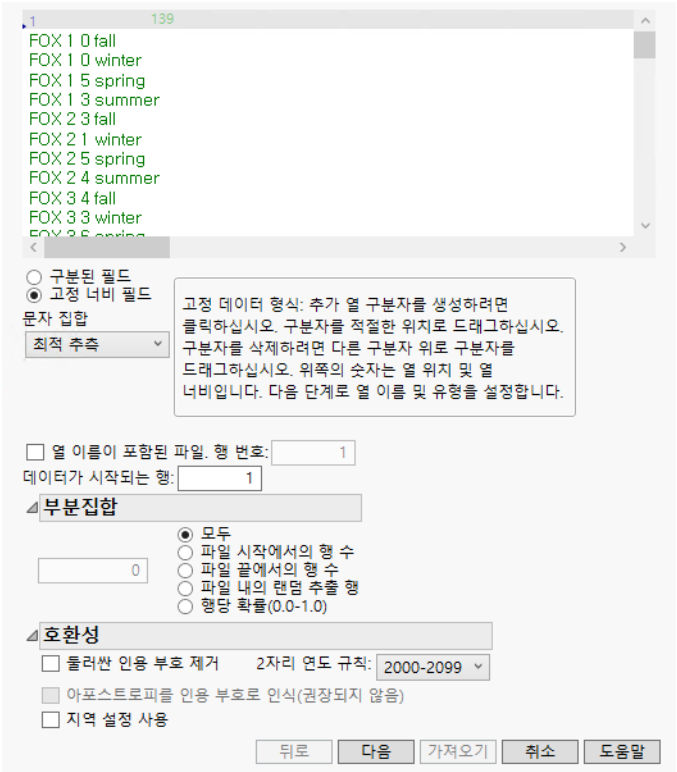


그림 3.15 구분자로 구분된 파일을 위한 텍스트 가져오기 마법사

문자 집합 가져오는 파일에 사용된 문자 집합을 선택하거나, JMP 에서 문자 집합이 감지되도록 합니다. 가져온 파일에 올바르지 않은 문자가 표시되면 파일을 다시 열고 다른 문자 집합을 선택하십시오.

필드 끝 (구분된 파일의 가져오기 창에서만 사용 가능) 필드 끝을 표시하는 문자 옆의 체크박스를 선택합니다. 또는 적절한 문자가 나열되지 않는 경우 **기타** 옆의 체크박스를 선택합니다.

행 끝 (구분된 파일의 가져오기 창에서만 사용 가능) 줄 (행) 끝을 표시하는 문자 옆의 체크박스를 선택합니다. 또는 적절한 문자가 나열되지 않는 경우 **기타** 옆의 체크박스를 선택합니다. JMP 에서 큰따옴표가 발견된 경우 닫는 큰따옴표를 찾기 위해 구분자 규칙이 변경됩니다. 따옴표 내에 포함된 공백을 비롯하여 다른 텍스트 구분자는 무시되고 텍스트 문자열의 일부로 처리됩니다.

열 이름이 포함된 파일. 행 번호 JMP 에 열 이름으로 사용할 데이터의 위치를 알려 줍니다. 예를 들어 텍스트 파일의 3 행에 열 이름이 있으면 이 옵션을 선택하고 체크박스에 3 을 입력합니다. 그러지 않으면 JMP 는 가져온 파일의 첫 번째 행에 있는 데이터를 JMP 데이터 테이블의 열 이름으로 사용하거나, 첫 번째 행을 데이터로 처리합니다.

데이터가 시작되는 행 데이터가 포함된 첫 번째 행의 번호를 지정합니다.

행 수 가져올 행 수를 지정합니다.

둘러싼 인용 부호 제거 (고정 너비 파일을 가져오는 경우에만 사용 가능) JMP 가 텍스트 파일의 데이터를 둘러싼 인용 부호를 제거하도록 하려면 이 체크박스를 선택합니다 .

2 자리 연도 규칙 2 자리 연도가 포함된 날짜의 표시 방법을 지정합니다 . 해당 날짜가 속한 100 년 단위의 범위를 선택합니다 . 예를 들어 가장 빠른 날짜가 1979 년 2 월 2 일이고 , 해당 연도가 1979 년이면 **1970-2069** 를 선택합니다 . 예를 들어 가장 빠른 날짜가 1979 년 2 월 2 일이고 , 해당 연도가 1979 년이면 **1970-2069** 를 선택합니다 . 날짜가 몇 세기에 걸쳐 있는 경우에는 데이터를 가져오기 전에 4 자리 연도를 사용하여 날짜를 재코딩해야 합니다 .

아포스트로피를 인용 부호로 인식 (권장되지 않음) (구분된 파일의 가져오기 창에서만 사용 가능) 이 옵션은 데이터 필드를 인용 부호로 묶지 않고 아포스트로피로 묶는 비표준 소스에서 데이터를 가져오는 경우에만 사용합니다 .

지역 설정 사용 텍스트 파일을 가져올 때 운영 체제의 지역 설정을 사용할지 여부를 지정합니다 . 이 옵션이 선택 취소되어 있으면 (기본값) 소수점으로 마침표를 사용하고 값 구분 기호로 쉼표를 사용하는 파일을 올바르게 가져올 수 있습니다 . " 텍스트 데이터 파일 " 가져오기 환경 설정에서 값 구분 기호를 지정해야 합니다 .

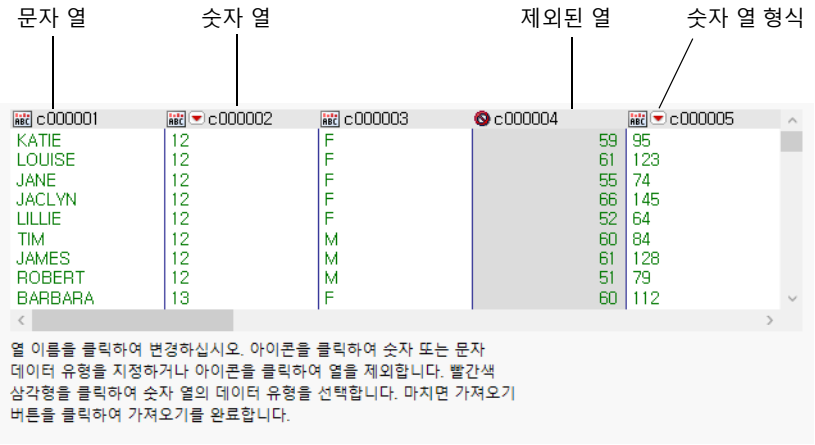
파일에서 소수점으로 쉼표가 사용된 경우에는 추가 옵션도 설정해야 합니다 . 즉 , OS 지역 설정에서 소수점으로 쉼표를 사용하거나 (Windows 및 macOS), "Windows 관련 " > " 표시 언어 " 환경 설정에서 소수점으로 쉼표를 사용하는 언어를 설정해야 합니다 .

설정 선택을 마쳤으면 **다음**을 클릭합니다 . 다음 창에서는 각 열의 모델링 유형을 보여 줍니다 . 기본 모델링 유형을 변경하려면 다음 중 하나를 수행하십시오 .

- 데이터 유형 아이콘을 클릭하여 데이터 유형을 숫자 ( ) 에서 문자 ( ) 로 변경합니다 . 아이콘을 클릭할 때마다 모델링 유형과 제외 ( ) 사이에서 순환됩니다 . 제외는 열을 가져오지 않음을 의미합니다 .
- 숫자 열의 데이터 형식을 변경하려면 빨간색 삼각형 메뉴에서 형식을 선택합니다 .
- 열 머리글을 클릭하여 텍스트를 수정합니다 .

텍스트 가져오기 창의 상단에는 JMP 데이터 파일로 가져왔을 때의 텍스트 파일 모양을 보여 주는 미리보기가 표시됩니다 . 데이터를 가져오려면 **가져오기** 버튼을 클릭합니다 .

그림 3.16 열 옵션이 있는 텍스트 가져오기 마법사 창



참고 : 날짜 형식을 지정하려면 열 이름 옆의 빨간색 삼각형을 선택하고 **날짜**를 선택한 다음 날짜 형식을 선택합니다 . 데이터를 가져온 후 날짜가 표시되는 방식을 지정하려면 데이터 테이블에서 날짜 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 다음 " 형식 " 옆의 날짜를 변경합니다 .

7. 날짜 설정을 마쳤으면 **가져오기**를 클릭하여 텍스트 가져오기를 완료합니다 .

참고 : 텍스트 가져오기 마법사와 여러 파일 가져오기 기능은 모두 기본적으로 숫자 열을 숫자로 가져옵니다 . 유지해야 하는 선행 0 을 포함하는 부품 번호나 15 자리를 초과하는 부품 번호 등의 경우에는 열을 강제로 문자로 가져오도록 설정할 수 있습니다 . 텍스트 가져오기 마법사에서는 두 번째 마법사 페인의 열 상단에서 열별로 이 작업을 수행할 수 있습니다 . 여러 파일 가져오기 기능에서는 **모든** 열을 문자로 가져오도록 하는 체크박스를 선택할 수 있습니다 . 이렇게 하지 않으면 선행 0 이 유지되지 않고 15 자리를 초과하는 숫자가 손실됩니다 . 여러 파일 가져오기에 대한 자세한 내용은 " 여러 파일 가져오기 "(91 페이지) 에서 확인하십시오 .

JMP 로 가져오기 전에 텍스트 편집

텍스트 파일을 스크립트 창에서 열고 이 창에서 텍스트를 편집할 수 있습니다. 그런 다음 텍스트를 데이터 테이블로 가져올 수 있습니다. 이 기능은 텍스트를 데이터 테이블로 가져오기 전에 텍스트 형식을 조정하려는 경우에 유용합니다. 예를 들어 올바른 구분자를 삽입하거나 텍스트를 수정해야 할 수 있습니다.

또 다른 옵션은 JMP 추가기능 정의 (.def) 파일을 텍스트로 연 후 스크립트 창에서 이를 편집하는 것입니다.

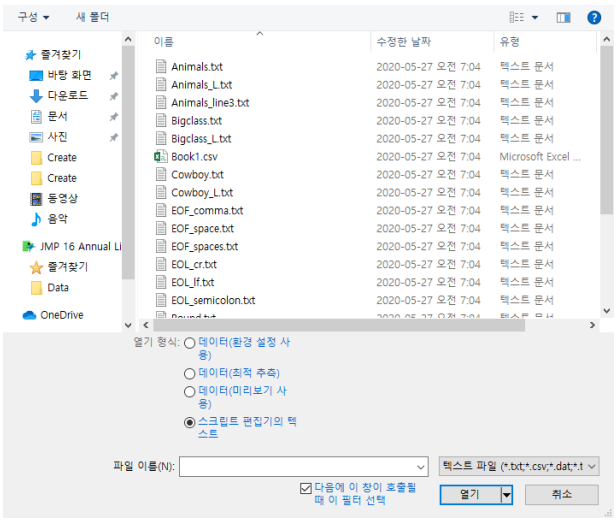
텍스트 파일을 텍스트 편집 창에서 열려면 (Windows)

JMP 홈 창에는 최근에 열었던 파일이 나열됩니다. 대개 텍스트 파일을 마우스 오른쪽 버튼으로 클릭하고 **일반 텍스트로 열기**를 선택하면 파일이 텍스트 편집 창에서 열립니다. JMP 추가기능 정의 파일은 JMP 홈 창에서 일반 텍스트로 열 수 없습니다.

파일을 처음 여는 경우 다음 단계를 수행하십시오.

1. **파일 > 열기**를 선택합니다.
2. **파일 이름** 옆의 목록에서 **텍스트 파일**을 선택합니다.

그림 3.17 텍스트 파일 선택



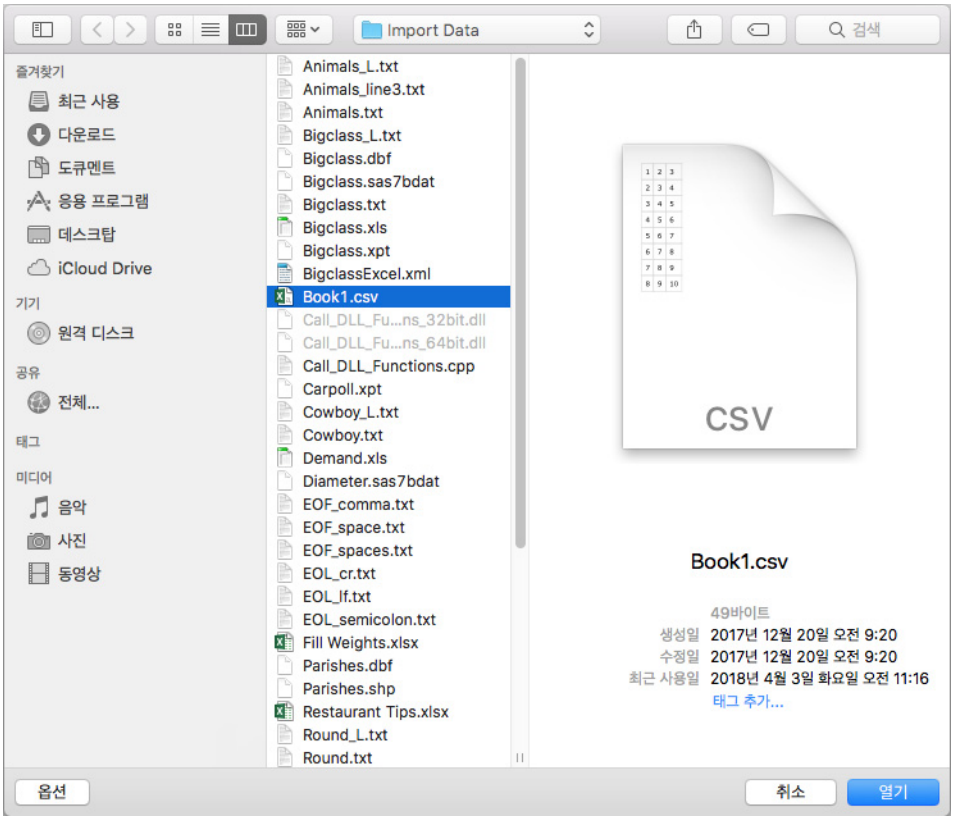
3. (선택 사항) 기본 옵션 파일 유형을 **텍스트 파일**로 설정하려면 **다음에 이 창이 호출될 때 이 필터 선택** 옆의 체크박스를 선택합니다.
4. 파일을 선택합니다.
5. **열기 형식** 옆의 **일반 텍스트 (스크립트 창에)**를 선택합니다.
6. **열기**를 클릭합니다.

   스크립트 창에 텍스트가 나타납니다.

텍스트 파일을 텍스트 편집 창에서 열려면 (macOS)

1. **파일 > 열기**를 선택합니다.

그림 3.18 macOS 에서 텍스트 문서 열기



2. 파일을 선택합니다 .
3. **옵션**을 클릭합니다 .
4. **열기 형식** 목록에서 **스크립트 편집기의 텍스트**를 선택합니다 .
5. **열기**를 클릭합니다 .

   스크립트 창에 텍스트가 나타납니다 .

텍스트를 데이터 테이블로 변환하는 자세한 방법은 "스크립트 편집기에서 텍스트 가져오기"(90 페이지)에서 3단계를 참조하십시오.

스크립트 편집기에서 텍스트 가져오기

스크립트 편집기에서 텍스트를 데이터 테이블로 가져올 수 있습니다. 텍스트는 테이블 형식(예: Microsoft Word 문서 또는 웹 페이지의 테이블)이거나 일반 텍스트 형식일 수 있습니다. 이 기능은 텍스트를 데이터 테이블로 가져오기 전에 텍스트 형식을 조정하려는 경우에 유용합니다. 예를 들어 올바른 구분자를 삽입하거나 텍스트를 수정해야 할 수 있습니다.

JMP는 환경 설정의 가져오기 설정을 사용하여 텍스트의 구성 및 형식 지정 방법을 결정합니다. 텍스트를 둘러싼 인용 부호를 제거하는 옵션과 열 머리글 및 데이터가 포함된 행을 지정하는 옵션을 비롯하여 몇 가지 옵션이 있습니다. 자세한 내용은 "JMP 환경 설정" 장의 "텍스트 파일 가져오기 및 내보내기의 환경 설정"(642 페이지)에서 확인하십시오.

참고: 전체 웹 페이지를 데이터 테이블로 가져올 수도 있습니다. 자세한 내용은 "인터넷 또는 원격 컴퓨터에서 데이터 가져오기"(174 페이지)에서 확인하십시오.

이 섹션에서는 스크립트 편집기에 붙여넣은 텍스트를 가져오는 방법에 대해 설명합니다. 텍스트 파일을 스크립트 편집기에서 여는 방법에 대한 자세한 내용은 "JMP로 가져오기 전에 텍스트 편집"(88 페이지)에서 확인하십시오.

스크립트 편집기에서 텍스트를 가져오려면

1. **파일 > 새로 만들기 > 스크립트** (Windows) 또는 **파일 > 새로 만들기 > 새 스크립트** (macOS) 를 선택하여 새 스크립트 창을 엽니다.
2. 텍스트를 복사하여 스크립트 편집기에 붙여넣습니다.
3. 다음 중 하나를 수행합니다.
   - 스트립트 편집기에서 모든 텍스트를 가져오려면 **편집 > 데이터로 가져오기**를 선택합니다.
   - 특정 텍스트를 가져오려면 텍스트를 선택한 후 **편집 > 데이터로 가져오기**를 선택합니다.

   텍스트가 JMP 데이터 테이블로 가져와집니다.

참고: 스크립트 편집기에서 가져오는 텍스트를 미리 보려면 Shift 키를 누른 채로 "파일" > "데이터로 가져오기"를 선택하십시오.

여러 파일 가져오기

MFI(여러 파일 가져오기)는 한 디렉터리의 CSV 데이터 및 비정형 텍스트 파일을 가져와서 하나의 데이터 테이블에 쌓거나 개별 데이터 테이블에 가져오는 기능입니다. 가져오기 설정을 스크립트로 저장한 후 나중에 데이터가 업데이트될 때 해당 스크립트를 실행할 수 있습니다.

여러 파일 가져오기가 지원되는 파일 유형은 다음과 같습니다.

- XLSX
- CSV
- TSV
- DAT
- TXT
- SAS7BDAT
- XML

- JMP
- JSON
- PNG
- JPG
- GIF

여러 파일 가져오기 기능으로 한 파일만 가져올 수도 있습니다. 표 3.3는 표준 텍스트 가져오기, 여러 파일 가져오기 또는 JSL 중에서 사용할 방법을 결정하는 데 유용합니다.

표 3.3 텍스트 가져오기 마법사와 MFI 비교

가져오기 특징	한 파일	여러 파일
CSV, TSV, DAT	텍스트를 가져올 때 미리보기 마법사를 사용하여 열 이름을 지정하고, 날짜 형식을 선택하고, 데이터 부분집합을 추출할 수 있습니다.	MFI 기능으로 유사한 여러 파일을 자동으로 쌓을 수 있습니다. 그런 다음 나중에 형식을 변경할 수 있습니다.
비정형 텍스트	MFI. `Load Text File()`을 사용하여 JSL로 가져오기 가능.	MFI
100MB를 초과하는 큰 파일	이전 텍스트 가져오기의 소스 스크립트를 사용하여 다음 가져오기의 속도를 높일 수 있습니다.	MFI는 병렬 가져오기를 수행하며, 텍스트 가져오기보다 훨씬 더 빠를 수 있습니다.
형식을 확실히 알 수 없는 경우	텍스트 가져오기를 통해 형식을 추측할 수 있습니다. 텍스트를 가져오기 전에 미리보기 마법사를 사용하여 결과를 확인할 수 있습니다.	MFI는 형식을 추측하지 않습니다. 사용자가 탭 구분자를 지정해야 합니다. 미리보기는 없지만 **대화상자 열린 채 유지**를 선택하면 다른 가져오기 설정을 쉽게 시도할 수 있습니다.
그림	JSL `Open()` 함수를 사용해 봅니다.	MFI는 PNG, JPG, 및 GIF 파일을 인식합니다. 이 경우 그림 열이 있는 테이블로 가져오게 됩니다.

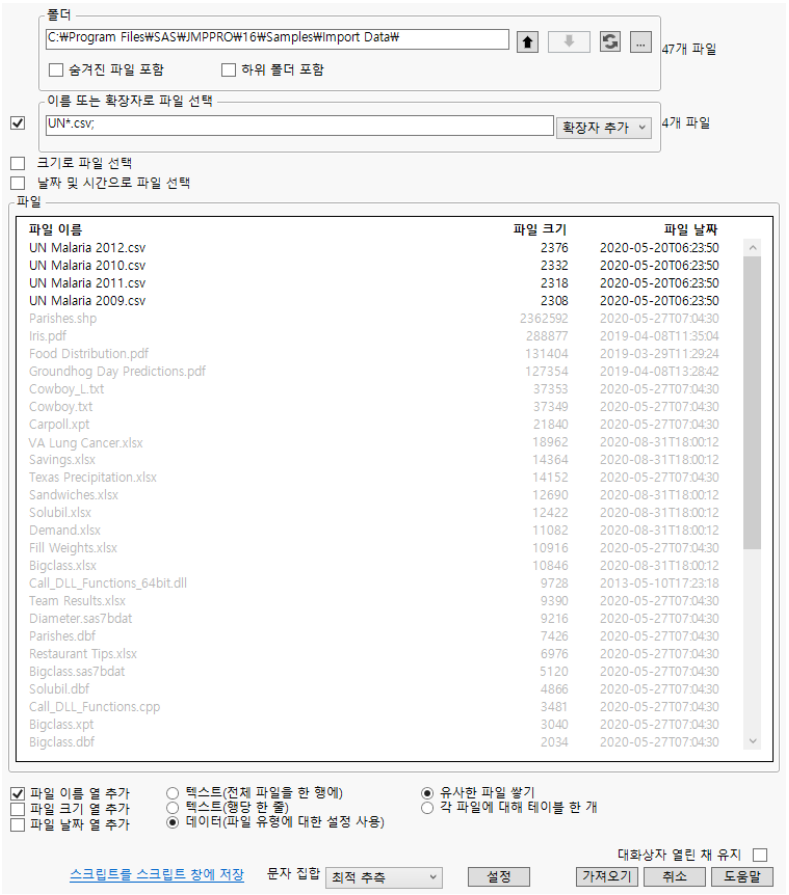
표 3.3 텍스트 가져오기 마법사와 MFI 비교 (계속)

가져오기 특징	한 파일	여러 파일
JMP, XLSX 및 JSON	JSL `Open()` 함수를 사용해 봅니다.	MFI는 JSL `Open()`을 사용하여 JMP, XLSX 및 JSON 파일을 열며, 요청된 경우 이를 쌓습니다. MFI는 JMP 테이블을 열 때 테이블 및 열 스크립트를 유지하지 않으므로 이러한 스크립트가 모두 필요한 경우 `Files In Directory()` 및 `Open()`을 사용해야 합니다.

여러 파일을 가져오려면 다음 단계를 수행하십시오 .

1. 가져오려는 파일을 동일한 디렉터리에 배치합니다 .
   이 예에서는 JMP Samples/Import Data 폴더에 있는 가져오기 데이터를 사용합니다 .
2. **파일 > 여러 파일 가져오기** (Windows) 또는 **파일 > 여러 항목 열기** (macOS) 를 선택합니다 .
3. JMP Samples/Import Data 폴더를 찾아 선택합니다 .
4. **이름 또는 확장자로 파일 선택** 상자에 UN*.csv 를 입력하고 Enter 키를 누릅니다 .
   이 예에서 사용할 파일은 UN 접두사가 있고 CSV 파일입니다 .
   파일 목록이 업데이트되어 선택된 파일이 표시됩니다 .
   여러 파일 이름 확장자를 세미콜론으로 구분하여 지정하거나 , **확장자 추가**를 클릭하고 확장자를 선택합니다 .

그림 3.19 선택된 CSV 파일



47 개의 파일이 이 필터와 매칭됩니다 .

4 개의 파일이 이 필터와 매칭됩니다 .

마지막 새로 고침 이후 해당 디렉터리의 파일이 변경되었으면 새로 고침 버튼 이 깜박입니다 .

5. 데이터 테이블에 추가할 열을 지정합니다 .

   기본적으로는 다른 열이 추가되지 않습니다. 파일 이름 열에 가져온 파일의 이름을 표시하도록 할 수 있습니다 . **파일 이름 열 추가**를 선택합니다 .

   참고 : CSV 설정은 " 텍스트 데이터 파일 " 환경 설정과 유사하기는 하지만 CSV 설정에서 선택되는 옵션이 " 텍스트 데이터 파일 " 환경 설정에 따라 결정되는 것은 아닙니다 .

6. 해당 CSV 파일에는 동일한 열이 포함되어 있으므로 **유사한 파일 쌓기**를 선택된 상태로 둡니다.
7. **가져오기**를 클릭합니다 .

   JMP 는 파일 이름 , 파일 크기 및 파일 날짜 / 시간 필터를 적용하여 파일을 쌓습니다 . 열 머리글이 동일한 열과 데이터를 동일한 데이터 테이블로 가져옵니다 . 열이 매칭되지 않을 경우 매칭되지 않는 열은 별도의 데이터 테이블로 가져옵니다 .

그림 3.20 가져온 CSV 데이터

	Country or Area	Year	Number of Incidents	파일 이름
1	Afghanistan	2009	64880	UN Malaria 2009.csv
2	Algeria	2009	4	UN Malaria 2009.csv
3	Angola	2009	1573422	UN Malaria 2009.csv
4	Argentina	2009	86	UN Malaria 2009.csv
5	Azerbaijan	2009	78	UN Malaria 2009.csv
6	Bangladesh	2009	63873	UN Malaria 2009.csv
7	Belize	2009	256	UN Malaria 2009.csv
8	Benin	2009	889597	UN Malaria 2009.csv
9	Bhutan	2009	972	UN Malaria 2009.csv
10	Bolivia (Plurinati...	2009	9743	UN Malaria 2009.csv
11	Botswana	2009	1024	UN Malaria 2009.csv
12	Brazil	2009	309316	UN Malaria 2009.csv
13	Burkina Faso	2009	182527	UN Malaria 2009.csv

참고 :

- 숨겨진 파일을 가져오려면 **숨겨진 파일 포함**을 선택합니다 .
- 하위 디렉터리에 있는 파일을 가져오려면 **하위 폴더 포함**을 선택합니다 .
- 특정 크기 범위의 파일을 가져오려면 **크기로 파일 선택** 상자에 파일 크기의 하한 및 상한을 입력합니다 .

  특정 한계를 넘는 큰 파일을 가져오지 않으려면 한계를 지정하는 것이 좋습니다 .

  파일 크기는 KB(킬로바이트) 단위로 지정하십시오 . 1KB 는 1,000 바이트에 해당합니다 .
- 특정 날짜 및 시간 범위 동안 수정된 파일을 가져오려면 **날짜 및 시간으로 파일 선택** 상자에서 달력 버튼 을 클릭하여 범위를 지정합니다 .
- (비정형 텍스트 파일) 파일의 구성 방식에 따라 **텍스트 (전체 파일을 한 행에)** 또는 **텍스트 (행당 한 줄)** 를 선택합니다 .

  참고 :

  – XLSX, JSON 및 JMP 파일의 경우에는 이 두 옵션이 지원되지 않습니다 .

  – " 텍스트 (전체 파일을 한 행에)" 를 선택하면 전체 파일이 가져오기에 포함되며 첫 번째 데이터 행 앞의 행은 가져오기에서 제외됩니다 .
- JMP는 가져오지 못한 파일(예: SAS 데이터 집합)을 나열하는 데이터 테이블을 생성합니다.

다른 인용 부호를 사용하여 정형 데이터 가져오기

일부 경우에는 내보내기 필터가 필드를 인용 부호가 아닌 다른 것으로 묶습니다. CSV 설정에서 구분자를 지정해야 합니다.

CSV 파일의 텍스트에는 인용 부호를 두 번 반복하는 대신 백슬래시 같은 이스케이프 시퀀스가 포함될 수 있습니다. 이 경우에도 CSV 설정에서 구분자를 지정해야 합니다.

팁 :

- **대화상자 열린 채 유지**는 기본적으로 해제되어 있지만 테스트 중에는 이를 설정할 수 있습니다 . 그러면 파일을 가져와서 데이터의 모양을 확인하고 , 필요한 경우 소스 파일을 편집한 후 , 동일한 설정으로 파일을 다시 가져올 수 있습니다 .

- 지원되지 않는 파일을 가져오면 "@ 사용하지 않은 파일 " 이라는 데이터 테이블이 생성됩니다 . 파일 이름은 " 파일 " 테이블 스크립트에 나열됩니다 . 스크립트를 마우스 오른쪽 버튼으로 클릭하고 " 편집 " 을 클릭하십시오 . 이 데이터 테이블에 지원되지 않는 파일의 파일 이름을 출력하려면 " 여러 파일 가져오기 " 창에서 **파일 이름 열 추가**를 선택하십시오 .
- 파일 목록의 정렬 방식을 변경하려면 " 파일 " 목록을 마우스 오른쪽 버튼으로 클릭하고 " 열별 정렬 " 을 선택하십시오 .
- 선택한 필터와 매칭되는 파일이 없으면 " 가져오기 " 버튼을 사용할 수 없습니다 .
- 텍스트 탐색기 스타일의 데이터에 대한 두 가지 옵션 (가운데 하단 라디오 버튼) 중 하나를 선택하면 JMP 가 자동으로 CSV 설정을 변경합니다 . 구분자는 지정되어 있지 않으므로 모든 데이터가 한 열에 로드됩니다 . 또한 JMP 는 행 끝 문자를 사용하지 않도록 CSV 설정을 변경하므로 전체 파일이 한 행에 로드됩니다 . CSV 설정을 변경할 수도 있습니다 . 수동 방법을 사용하려면 " 머리글 있음 " 을 해제하고 " 첫 번째 데이터 행 " 을 1 로 설정해야 합니다 .

PDF 파일 가져오기

PDF 가져오기 마법사를 사용하면 데이터를 가져오기 전에 PDF 파일을 미리 보고, 가져오기 설정을 조정하고, 데이터 테이블 미리보기에서 결과를 확인할 수 있습니다. JMP 가 PDF 파일의 데이터 구조를 자동으로 감지할 수도 있고, 사용자가 구조를 사용자 정의할 수도 있습니다.

참고 : PDF 파일은 데이터가 포함된 테이블로 구성되어 있습니다 . PDF 미리보기에서는 각 페이지의 테이블이 외곽선이 표시된 상태로 선택되어 있습니다 .

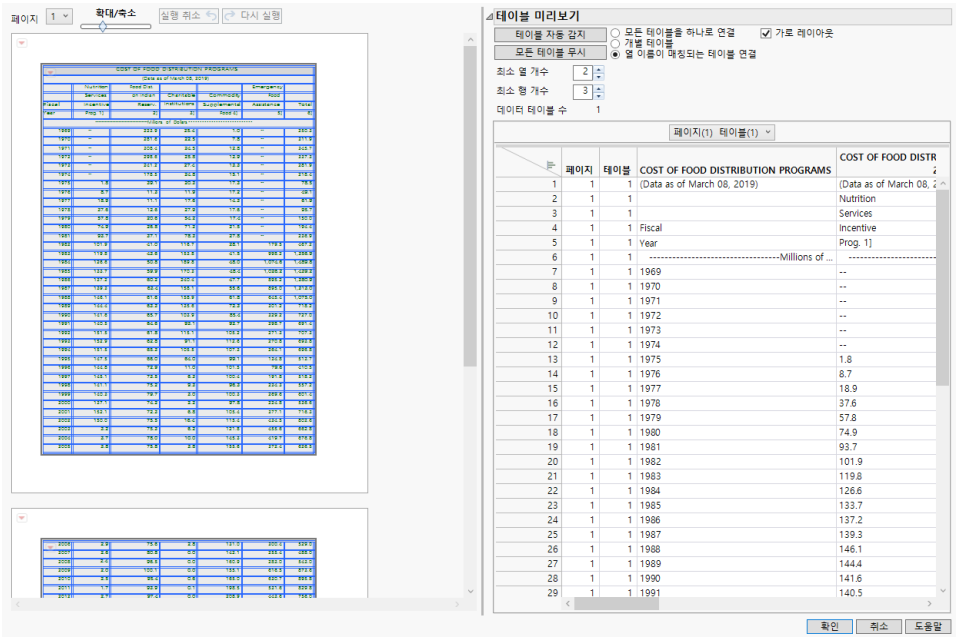
- " 여러 테이블을 하나의 데이터 테이블로 연결 "
- " 열 머리글을 기준으로 테이블 연결 "
- "JMP 가 자동 감지하지 않는 테이블 가져오기 "
- "PDF 미리보기의 빨간색 삼각형 옵션 "
- " 데이터 테이블 미리보기 옵션 "
- "PDF 미리보기의 마우스 오른쪽 버튼 클릭 옵션 "

여러 테이블을 하나의 데이터 테이블로 연결

1. **파일** > **열기**를 선택하고 " 파일 이름 " 옆의 목록에서 **PDF 파일**을 선택합니다 .
2. JMP Samples/Import Data 폴더로 이동하고 Food Distribution.pdf 를 두 번 클릭합니다 .
3. PDF 가져오기 마법사의 왼쪽에 PDF 파일의 미리보기가 표시됩니다 . 오른쪽에는 데이터 테이블에 있는 데이터의 미리보기가 표시됩니다 .

   이 PDF 파일은 페이지당 한 개의 테이블로 구성되어 있습니다 .

그림 3.21 초기 PDF 파일 및 데이터 테이블 미리보기



다음 사항에 유의하십시오 .

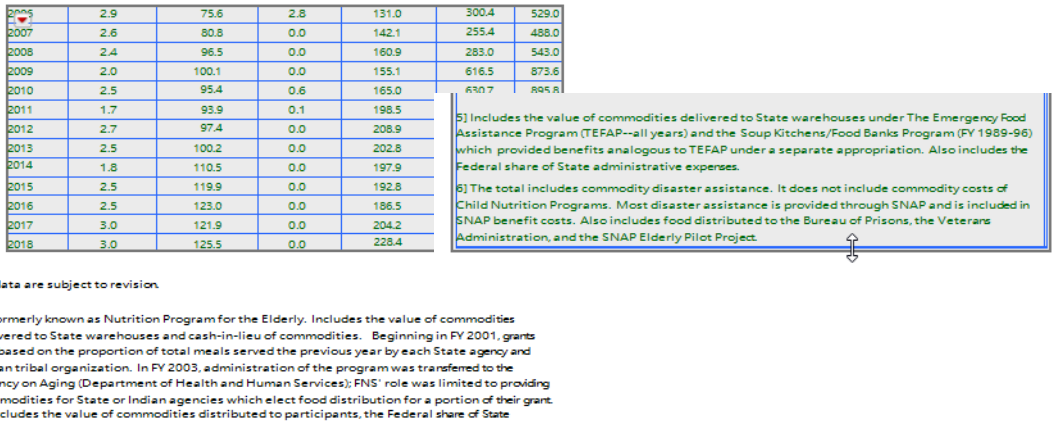
- PDF 의 맨 위에는 가져오지 않으려는 행이 7 개 있습니다 .
- PDF 파일의 마지막 행에 있는 각주도 필요하지 않습니다 .

4. 테이블의 모퉁이에서 빨간색 삼각형을 클릭하고 **머리글로 사용할 행의 개수** > **행 지정**을 선택합니다 .
5. "7" 을 입력하고 **확인**을 클릭합니다 .

   PDF 파일 미리보기에서 처음 7 개의 행이 선택됩니다 .
   데이터 테이블 미리보기에서는 데이터가 1 행부터 시작됩니다 .
6. PDF 파일 미리보기에서 2 페이지로 스크롤합니다 .
7. 각주가 더 이상 선택되지 않을 때까지 PDF 테이블의 하단을 위쪽으로 드래그합니다 (그림 3.22).

참고 : 단순히 최종 데이터 테이블에서 행을 삭제해도 됩니다 .

그림 3.22 테이블을 드래그하여 데이터 제거



2006	2.9	75.6	2.8	131.0	300.4	529.0
2007	2.6	80.8	0.0	142.1	255.4	488.0
2008	2.4	96.5	0.0	160.9	283.0	543.0
2009	2.0	100.1	0.0	155.1	616.5	873.6
2010	2.5	95.4	0.6	165.0	630.7	895.8
2011	1.7	93.9	0.1	198.5		
2012	2.7	97.4	0.0	208.9		
2013	2.5	100.2	0.0	202.8		
2014	1.8	110.5	0.0	197.9		
2015	2.5	119.9	0.0	192.8		
2016	2.5	123.0	0.0	186.5		
2017	3.0	121.9	0.0	204.2		
2018	3.0	125.5	0.0	228.4		

5] Includes the value of commodities delivered to State warehouses under The Emergency Food Assistance Program (TEFAP--all years) and the Soup Kitchens/Food Banks Program (FY 1989-96) which provided benefits analogous to TEFAP under a separate appropriation. Also includes the Federal share of State administrative expenses.

6] The total includes commodity disaster assistance. It does not include commodity costs of Child Nutrition Programs. Most disaster assistance is provided through SNAP and is included in SNAP benefit costs. Also includes food distributed to the Bureau of Prisons, the Veterans Administration, and the SNAP Elderly Pilot Project.

All data are subject to revision.

1] Formerly known as Nutrition Program for the Elderly. Includes the value of commodities delivered to State warehouses and cash-in-lieu of commodities. Beginning in FY 2001, grants are based on the proportion of total meals served the previous year by each State agency and Indian tribal organization. In FY 2003, administration of the program was transferred to the Agency on Aging (Department of Health and Human Services); FNS' role was limited to providing commodities for State or Indian agencies which elect food distribution for a portion of their grant.
2] Includes the value of commodities distributed to participants, the Federal share of State

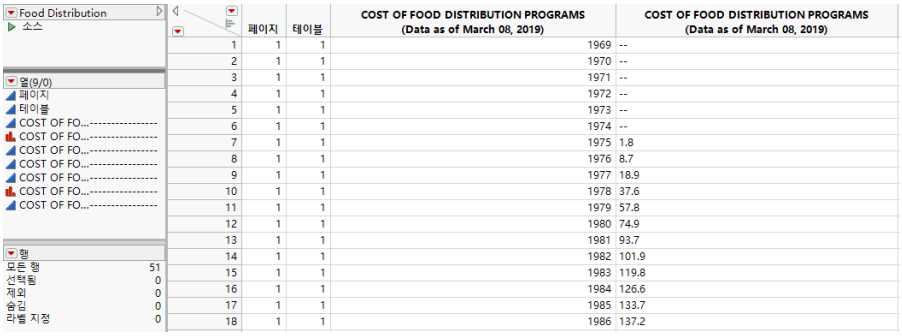
8. 오른쪽의 " 테이블 " 보고서에서 **모든 테이블을 하나로 연결**을 선택합니다 .

   둘 이상의 데이터 테이블을 연결하면 해당 데이터 테이블의 행이 하나의 데이터 테이블에 결합됩니다 .

9. 하단에서 **확인**을 클릭합니다 .

   데이터 테이블로 데이터 가져오기가 수행됩니다 .

그림 3.23 최종 데이터 테이블 (일부 보기)



Food Distribution
▶ 소스

▼ 열(9/0)
페이지
테이블
COST OF FO...
COST OF FO...
COST OF FO...
COST OF FO...
COST OF FO...
COST OF FO...
COST OF FO...

▼ 행
모든 행 51
선택됨 0
제외 0
숨김 0
라벨 지정 0

	페이지	테이블	COST OF FOOD DISTRIBUTION PROGRAMS (Data as of March 08, 2019)	COST OF FOOD DISTRIBUTION PROGRAMS (Data as of March 08, 2019)
1	1	1	1969	--
2	1	1	1970	--
3	1	1	1971	--
4	1	1	1972	--
5	1	1	1973	--
6	1	1	1974	--
7	1	1	1975	1.8
8	1	1	1976	8.7
9	1	1	1977	18.9
10	1	1	1978	37.6
11	1	1	1979	57.8
12	1	1	1980	74.9
13	1	1	1981	93.7
14	1	1	1982	101.9
15	1	1	1983	119.8
16	1	1	1984	126.6
17	1	1	1985	133.7
18	1	1	1986	137.2

열 머리글을 기준으로 테이블 연결

PDF 파일의 각 페이지에 열 머리글이 동일한 테이블 형식 데이터가 포함되어 있으면 열 머리글을 기준으로 테이블을 연결할 수 있습니다 .

1. **파일** > **열기**를 선택하고 " 파일 이름 " 옆의 목록에서 **PDF 파일**을 선택합니다 .
2. JMP Samples/Import Data 폴더로 이동하고 Iris.pdf 를 두 번 클릭합니다 .

   PDF 미리보기를 스크롤하여 각 페이지의 열 머리글이 동일함을 확인합니다 .

3. 오른쪽의 " 테이블 " 보고서에서 **열 이름이 매칭되는 테이블 연결**이 선택되어 있는지 확인합니다 .
4. 하단에서 **확인**을 클릭합니다 .

그림 3.24 최종 데이터 테이블 (일부 보기)

	페이지	테이블	Sepal length	Sepal width	Petal length	Petal width	Species
1	1	1	5.1	3.5	1.4	0.2	setosa
2	1	1	4.9	3	1.4	0.2	setosa
3	1	1	4.7	3.2	1.3	0.2	setosa
4	1	1	4.6	3.1	1.5	0.2	setosa
5	1	1	5	3.6	1.4	0.2	setosa
6	1	1	5.4	3.9	1.7	0.4	setosa
7	1	1	4.6	3.4	1.4	0.3	setosa
8	1	1	5	3.4	1.5	0.2	setosa
9	1	1	4.4	2.9	1.4	0.2	setosa
10	1	1	4.9	3.1	1.5	0.1	setosa
11	1	1	5.4	3.7	1.5	0.2	setosa
12	1	1	4.8	3.4	1.6	0.2	setosa
13	1	1	4.8	3	1.4	0.1	setosa
14	1	1	4.3	3	1.1	0.1	setosa
15	1	1	5.8	4	1.2	0.2	setosa
16	1	1	5.7	4.4	1.5	0.4	setosa
17	1	1	5.4	3.9	1.3	0.4	setosa
18	1	1	5.1	3.5	1.4	0.3	setosa

JMP 가 자동 감지하지 않는 테이블 가져오기

일부 경우에는 PDF 파일의 테이블이 자동으로 감지되지 않습니다. 이러한 경우 열 주위를 상자 모양으로 드래그하여 테이블을 그룹화하고 나눌 수 있습니다.

그림 3.25 일부 PDF 파일 미리보기

1886	Groundhog Day first recognized in Punxsutawney by
	Weathers Wags No Record.
1887	First Official trek to
	Gobbler's Knob Saw Shadow.
1888	Saw Shadow.
1889	No Record.
1890	No Shadow.
1891	No Record.
1892	No Record.
1893	No Record.
1894	No Record.
1895	No Record.
1896	No Record.
1897	No Record.
1898	Saw Shadow.
1899	No Record. Groundhog
	Club founded
1900	Saw Shadow.
1901	Saw Shadow.

첫 번째 열의 행은 세 개의 테이블로 자동 감지되었습니다.

세 번째 테이블의 상단 모서리를 위쪽으로 드래그하여 두 열이 포함되도록 합니다.

그림 3.26 단일 테이블 생성

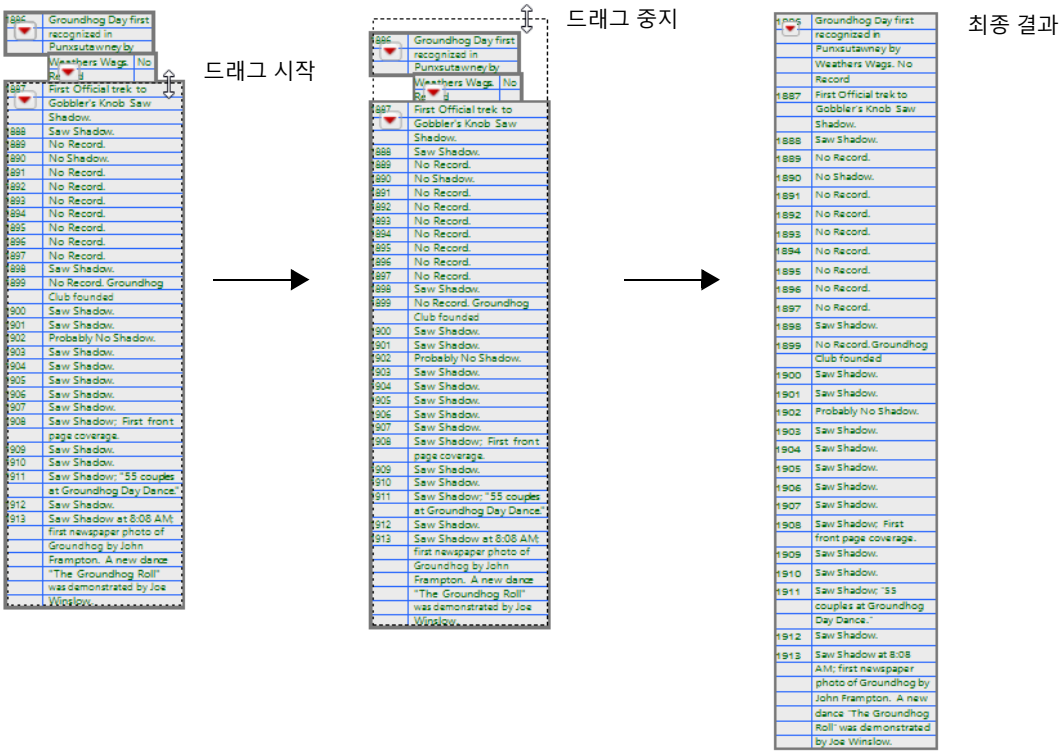
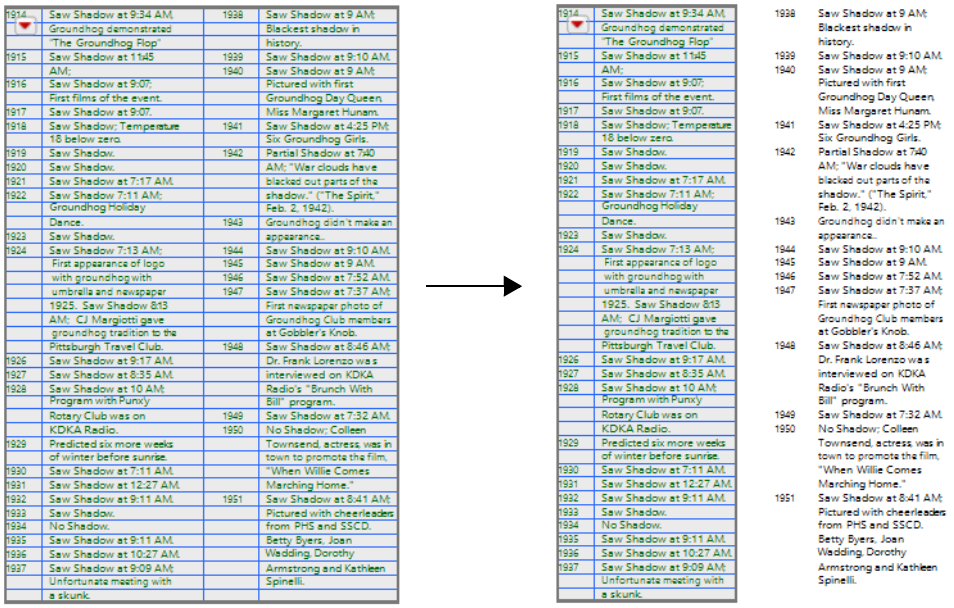


그림 3.27에 표시된 첫 번째 테이블에서는 맨 오른쪽 테두리를 왼쪽으로 드래그하여 1열과 2열만 동일한 테이블에 포함되도록 하고 3열과 4열은 테이블 외부에 있도록 합니다(그림 3.27).

그림 3.27 테이블 분할

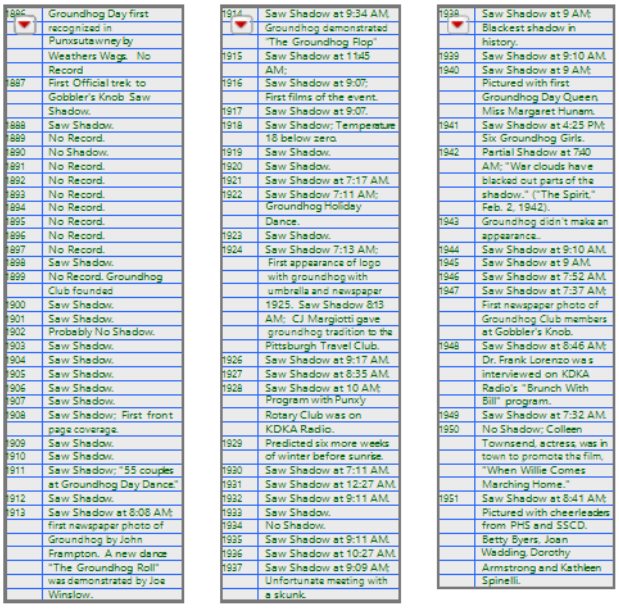


1914	Saw Shadow at 9:34 AM.
	Groundhog demonstrated
	'The Groundhog Flop'
1915	Saw Shadow at 11:45
	AM;
1916	Saw Shadow at 9:07;
	First films of the event.
1917	Saw Shadow at 9:07.
1918	Saw Shadow; Temperature
	18 below zero.
1919	Saw Shadow.
1920	Saw Shadow.
1921	Saw Shadow at 7:17 AM.
1922	Saw Shadow 7:11 AM;
	Groundhog Holiday
	Dance.
1923	Saw Shadow.
1924	Saw Shadow 7:13 AM;
	First appearance of logo
	with groundhog with
	umbrella and newspaper
	1925. Saw Shadow 8:13
	AM; CJ Margiotti gave
	groundhog tradition to the
	Pittsburgh Travel Club.
1926	Saw Shadow at 9:17 AM.
1927	Saw Shadow at 8:35 AM.
1928	Saw Shadow at 10 AM;
	Program with Punxy
	Rotary Club was on
	KDKA Radio.
1929	Predicted six more weeks
	of winter before sunrise.
1930	Saw Shadow at 7:11 AM.
1931	Saw Shadow at 12:27 AM.
1932	Saw Shadow at 9:11 AM.
1933	Saw Shadow.
1934	No Shadow.
1935	Saw Shadow at 9:11 AM.
1936	Saw Shadow at 10:27 AM.
1937	Saw Shadow at 9:09 AM;
	Unfortunate meeting with
	a skunk.

1938	Saw Shadow at 9 AM;
	Blackest shadow in
	history.
1939	Saw Shadow at 9:10 AM.
1940	Saw Shadow at 9 AM;
	Pictured with first
	Groundhog Day Queen,
	Miss Margaret Hunam.
1941	Saw Shadow at 4:25 PM;
	Six Groundhog Girls.
1942	Partial Shadow at 7:40
	AM; "War clouds have
	blacked out parts of the
	shadow." ("The Spirit,"
	Feb. 2, 1942).
1943	Groundhog didn't make an
	appearance.
1944	Saw Shadow at 9:10 AM.
1945	Saw Shadow at 9 AM.
1946	Saw Shadow at 7:52 AM.
1947	Saw Shadow at 7:37 AM;
	First newspaper photo of
	Groundhog Club members
	at Gobbler's Knob.
1948	Saw Shadow at 8:46 AM;
	Dr. Frank Lorenzo was
	interviewed on KDKA
	Radio's "Brunch With
	Bill" program.
1949	Saw Shadow at 7:32 AM.
1950	No Shadow; Colleen
	Townsend, actress, was in
	town to promote the film,
	"When Willie Comes
	Marching Home."
1951	Saw Shadow at 8:41 AM;
	Pictured with cheerleaders
	from PHS and SSCD.
	Betty Byers, Joan
	Wadding, Dorothy
	Armstrong and Kathleen
	Spinelli.

나머지 두 열 주위를 상자 모양으로 드래그하여 하나의 테이블에 넣습니다(그림 3.28).

그림 3.28 1 페이지의 최종 테이블



1886	Groundhog Day first
	recognized in
	Punxsutawney by
	Weathers Wags. No
	Record
1887	First Official trek to
	Gobbler's Knob. Saw
	Shadow.
1888	Saw Shadow.
1889	No Record.
1890	No Shadow.
1891	No Record.
1892	No Record.
1893	No Record.
1894	No Record.
1895	No Record.
1896	No Record.
1897	No Record.
1898	Saw Shadow.
1899	No Record. Groundhog
	Club founded.
1900	Saw Shadow.
1901	Saw Shadow.
1902	Probably No Shadow.
1903	Saw Shadow.
1904	Saw Shadow.
1905	Saw Shadow.
1906	Saw Shadow.
1907	Saw Shadow.
1908	Saw Shadow; First front
	page coverage.
1909	Saw Shadow.
1910	Saw Shadow.
1911	Saw Shadow; "55 couples
	at Groundhog Day Dance."
1912	Saw Shadow.
1913	Saw Shadow at 8:08 AM;
	first newspaper photo of
	Groundhog by John
	Frampton. A new dance
	"The Groundhog Roll"
	was demonstrated by Joe
	Winslow.

1914	Saw Shadow at 9:34 AM.
	Groundhog demonstrated
	'The Groundhog Flop'
1915	Saw Shadow at 11:45
	AM;
1916	Saw Shadow at 9:07;
	First films of the event.
1917	Saw Shadow at 9:07.
1918	Saw Shadow; Temperature
	18 below zero.
1919	Saw Shadow.
1920	Saw Shadow.
1921	Saw Shadow at 7:17 AM.
1922	Saw Shadow 7:11 AM;
	Groundhog Holiday
	Dance.
1923	Saw Shadow.
1924	Saw Shadow 7:13 AM;
	First appearance of logo
	with groundhog with
	umbrella and newspaper
	1925. Saw Shadow 8:13
	AM; CJ Margiotti gave
	groundhog tradition to the
	Pittsburgh Travel Club.
1926	Saw Shadow at 9:17 AM.
1927	Saw Shadow at 8:35 AM.
1928	Saw Shadow at 10 AM;
	Program with Punxy
	Rotary Club was on
	KDKA Radio.
1929	Predicted six more weeks
	of winter before sunrise.
1930	Saw Shadow at 7:11 AM.
1931	Saw Shadow at 12:27 AM.
1932	Saw Shadow at 9:11 AM.
1933	Saw Shadow.
1934	No Shadow.
1935	Saw Shadow at 9:11 AM.
1936	Saw Shadow at 10:27 AM.
1937	Saw Shadow at 9:09 AM;
	Unfortunate meeting with
	a skunk.

1938	Saw Shadow at 9 AM;
	Blackest shadow in
	history.
1939	Saw Shadow at 9:10 AM.
1940	Saw Shadow at 9 AM;
	Pictured with first
	Groundhog Day Queen,
	Miss Margaret Hunam.
1941	Saw Shadow at 4:25 PM;
	Six Groundhog Girls.
1942	Partial Shadow at 7:40
	AM; "War clouds have
	blacked out parts of the
	shadow." ("The Spirit,"
	Feb. 2, 1942).
1943	Groundhog didn't make an
	appearance.
1944	Saw Shadow at 9:10 AM.
1945	Saw Shadow at 9 AM.
1946	Saw Shadow at 7:52 AM.
1947	Saw Shadow at 7:37 AM;
	First newspaper photo of
	Groundhog Club members
	at Gobbler's Knob.
1948	Saw Shadow at 8:46 AM;
	Dr. Frank Lorenzo was
	interviewed on KDKA
	Radio's "Brunch With
	Bill" program.
1949	Saw Shadow at 7:32 AM.
1950	No Shadow; Colleen
	Townsend, actress, was in
	town to promote the film,
	"When Willie Comes
	Marching Home."
1951	Saw Shadow at 8:41 AM;
	Pictured with cheerleaders
	from PHS and SSCD.
	Betty Byers, Joan
	Wadding, Dorothy
	Armstrong and Kathleen
	Spinelli.

참고 : Samples/Import Data 폴더에 있는 Groundhog Day Predictions.pdf 는 PDF 가져오기 마법사에서 자동 감지된 정리된 버전의 데이터를 보여 줍니다 .

PDF 미리보기의 빨간색 삼각형 옵션

이 섹션에서는 PDF 미리보기 페인의 빨간색 삼각형 옵션을 설명합니다 .

전역 설정을 위한 빨간색 삼각형

PDF 미리보기 창의 왼쪽 상단에 있는 빨간색 삼각형에는 다음 두 개의 옵션이 포함되어 있습니다.

이 페이지의 테이블 자동 감지 이전에 테이블을 나누거나 결합한 경우 이 옵션을 사용하면 변경 사항이 되돌려지므로 테이블이 다시 자동 감지됩니다 .

이 페이지의 테이블 무시 현재 페이지의 테이블을 제거합니다. 해당 데이터는 가져오지 않습니다.

각 테이블의 빨간색 삼각형

이 테이블 무시 현재 테이블을 제거합니다 . 해당 데이터는 가져오지 않습니다 .

머리글로 사용할 행의 개수 PDF 파일의 행 중 데이터로 가져오지 않을 머리글 행의 개수를 지정합니다 .

이 페이지의 테이블 자동 감지 이전에 테이블을 나누거나 결합한 경우 이 옵션을 사용하면 변경 사항이 되돌려지므로 테이블이 다시 자동 감지됩니다 .

이 페이지의 테이블 무시 현재 페이지의 테이블을 제거합니다. 해당 데이터는 가져오지 않습니다.

데이터 테이블 미리보기 옵션

이 섹션에서는 데이터 테이블 미리보기 페인의 옵션에 대해 설명합니다 .

테이블 자동 감지 이전에 테이블을 나누거나 결합한 경우 이 옵션을 사용하면 변경 사항이 되돌려지므로 테이블이 다시 자동 감지됩니다 .

모든 테이블 무시 PDF 파일의 테이블을 제거합니다 . 해당 데이터는 가져오지 않습니다 .

최소 열 개수 열 개수가 지정한 것보다 적은 테이블은 가져오지 않도록 합니다 .

최소 행 개수 행 개수가 지정한 것보다 적은 테이블은 가져오지 않도록 합니다 .

모든 테이블을 하나로 연결 둘 이상의 데이터 테이블에 있는 행을 하나의 데이터 테이블로 결합합니다 .

개별 테이블 데이터를 여러 개의 테이블로 가져옵니다 .

열 이름이 매칭되는 테이블 연결 열 이름이 매칭되는 둘 이상의 데이터 테이블에 있는 행을 하나의 데이터 테이블로 결합합니다 .

PDF 미리보기의 마우스 오른쪽 버튼 클릭 옵션

대부분의 PDF 미리보기에서 사용할 수 있는 마우스 오른쪽 버튼 클릭 옵션에 대한 자세한 내용은 "각 테이블의 빨간색 삼각형"(102페이지)에서 확인하십시오. 다음 옵션도 사용할 수 있습니다.

열 구분자 추가 선택한 위치에 열 구분자를 생성합니다. 예를 들어 열이 자동으로 감지되지 않은 경우 등에 데이터 테이블에 새 열을 생성하려면 이 옵션을 선택합니다.

열 구분자 제거 선택한 위치에서 열 구분자를 삭제합니다.

행 구분자 추가 선택한 위치에 행 테두리를 생성합니다. 데이터 테이블에 새 행을 생성하려면 이 옵션을 선택합니다.

행 구분자 제거 예를 들어 행이 자동으로 감지되지 않은 경우 등에 선택한 위치에서 행 테두리를 삭제합니다.

XML 및 JSON 파일 가져오기

XML 및 JSON 데이터 미리보기 및 가져오기

XML 가져오기 마법사는 두 가지 유형의 내포된 텍스트 데이터, 즉 XML 파일과 JSON 파일을 가져옵니다. 두 종류의 파일은 모두 계층 관계를 나타내기 위해 내포될 수 있는 정형화된 텍스트를 포함합니다. 이 마법사를 사용하면 데이터 테이블을 구성할 행을 결정하는 값 및 요소가 포함된 계층 요소를 쉽게 선택할 수 있습니다.

내포된 텍스트 파일의 경우 해당 문서가 짧은 스토리들로 구성된 책이라고 생각하면 이해하기 쉽습니다. 짧은 스토리 중 일부는 장으로 구성될 수 있고 일부는 단락으로만 구성될 수 있습니다. 장은 단락으로 구성됩니다. 단락은 문장으로 구성됩니다. 다음은 이러한 책 구조를 보여 주는 전체 트리의 예입니다.

```
<book>
  <story name="car poll">
    <wheels>4</wheels>
    <para><name>chev</name></para>
    <para><name>ford</name></para>
    <para><name>volk</name></para>
  </story>
  <story name="big class">
    <para><name>ralph</name><height>6</height></para>
    <para><name>billy</name><height>5</height></para>
  </story>
  <story name="cheese" location="NC">
    <chapter name="american">
      <para>1<price>2.00</price><quantity>2</quantity></para>
      <para>2<price>2.10</price><quantity>4</quantity></para>
      <para>3<price>2.20</price><quantity>3</quantity></para>
```

```
    </chapter>
    <chapter name="swiss">
      <para>1<price>3.00</price><quantity>3</quantity></para>
      <para>2<price>3.10</price><quantity>2</quantity></para>
    </chapter>
  </story>
</book>
```

이 XML 문서(book)에는 세 개의 짧은 스토리(story)가 포함되어 있으며 각각의 짧은 스토리는 서로 다른 데이터 테이블입니다. 다음 예에서는 이러한 데이터 테이블을 한 번에 하나씩만 가져오도록 선택합니다. 이 특정 XML 파일은 모든 테이블에서 단락을 사용하여 테이블의 행을 나타냅니다. 단락별 내용 배열은 필요하지 않습니다. 다른 두 테이블보다 약간 더 복잡한 "cheese" 테이블을 가져오려는 경우 각 행에 대한 "american" 또는 "swiss"가 포함된 종류 열이 필요할 수 있습니다. "price" 및 "quantity" 열도 필요합니다. 또한 <para>의 1, 2, 3, 1, 2 값에 대해 명확히 정의된 이름이 없는 열도 필요할 수 있습니다. 이렇게 5개의 행이 필요합니다. 마지막으로, "location" 값 "NC"에 대한 테이블 변수 또는 열이 필요할 수 있습니다.

JSON 파일은 이름 및 값 쌍으로 구성되어 있으며 각 항목을 열 머리글과 데이터로 가져오게 됩니다. 다음 예에서는 각 문자열의 첫 번째 이름이 "Grocery Store Purchases"에 추가된 후 열 머리글로 변환됩니다(예: "Grocery Store Purchases.Item"). "avocado"는 이 테이블의 첫 번째 셀 값이 됩니다. 다음 열의 이름은 "Grocery Store Purchases.Category"로 지정되며, "Produce"는 두 번째 열의 첫 번째 셀 값이 됩니다.

```
{"Grocery Store Purchases":[
    {"Item":"avocado", "Category":"Produce"},
    {"Item":"bread", "Category":"Bakery"},
    {"Item":"chocolate", "Category":"candy"}
]}
```

XML 가져오기 마법사에서 XML 또는 JSON 파일을 가져오려면

1. **파일 > 열기**를 선택합니다.
2. JMP의 Samples/Import Data 폴더로 이동합니다.
3. Windows의 경우, 파일 유형을 **XML 데이터 파일**로 설정합니다.
4. Book.xml을 선택하고 **열기**를 클릭합니다.

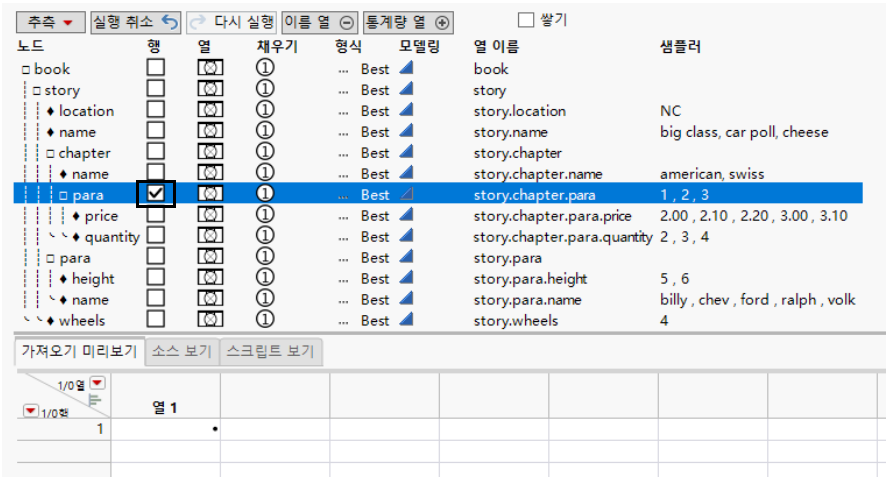
   파일이 XML 가져오기 마법사에서 열립니다.

참고 : " 데이터 파일 열기 " 창에 " 데이터 (미리보기 사용)" 가 선택되어 있으므로 XML 파일이 자동으로 마법사에서 열립니다. XML 및 JSON 파일을 직접 가져올 수도 있습니다. 자세한 내용은 "XML 파일 바로 가져오기 "(109 페이지) 에서 확인하십시오.

데이터 가져오기 방법 결정

1. " 행 " 열에서 데이터 테이블에 새 행을 생성할 요소를 선택합니다. 이 예에서는 </para> 가 처리될 때 행이 생성되고 현재 값으로 해당 행의 열이 작성됩니다.

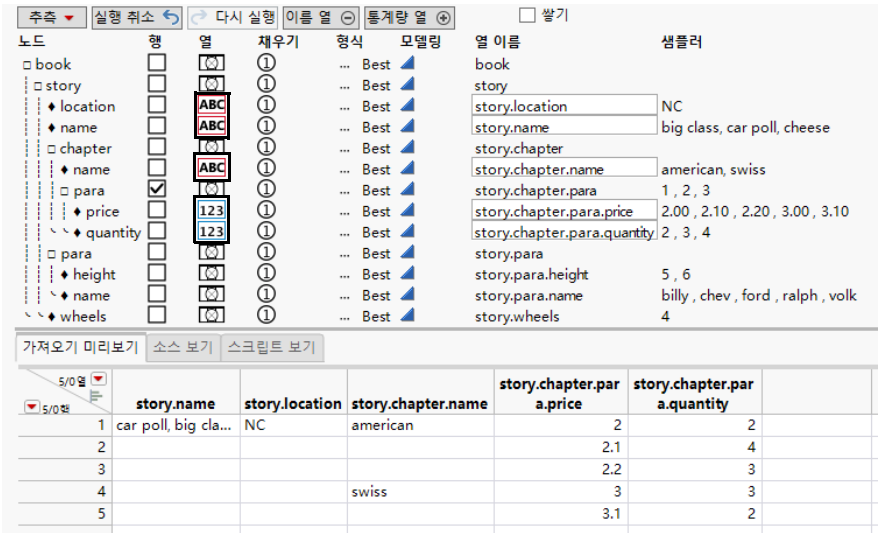
그림 3.29 " 행 " 열에서 선택된 행 요소



2. " 열 " 열의 열 원형 옆에서 해당 열의 데이터 유형을 선택합니다 .
   - location: **문자**
   - name: **문자** (두 인스턴스 모두)
   - price: **숫자**
   - quantity: **숫자**

팁 : " 샘플러 " 열의 값을 참조하면 해당 데이터 유형을 쉽게 결정할 수 있습니다 .

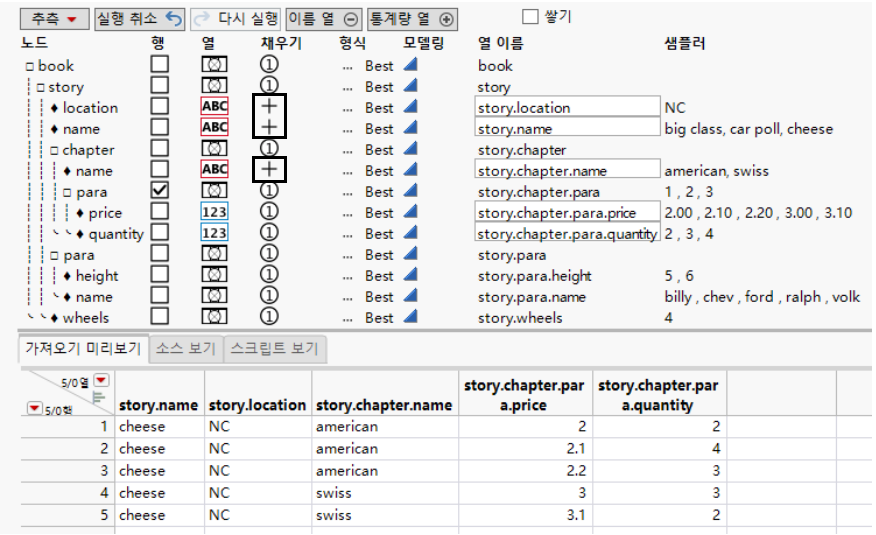
그림 3.30 " 열 " 열에 할당된 데이터 유형



3. " 채우기 " 열에서 셀을 채울 값을 지정합니다 .

   "location" 와 두 개의 "name" 인스턴스 모두에 대해 " 채우기 " 원을 클릭하고 **+book/story** 를 선택합니다. 이렇게 설정하면 <book>/<story> 노드가 다시 시작될 때까지 셀에 값이 채워집니다 .

그림 3.31 " 채우기 " 열에 선택된 채우기 값

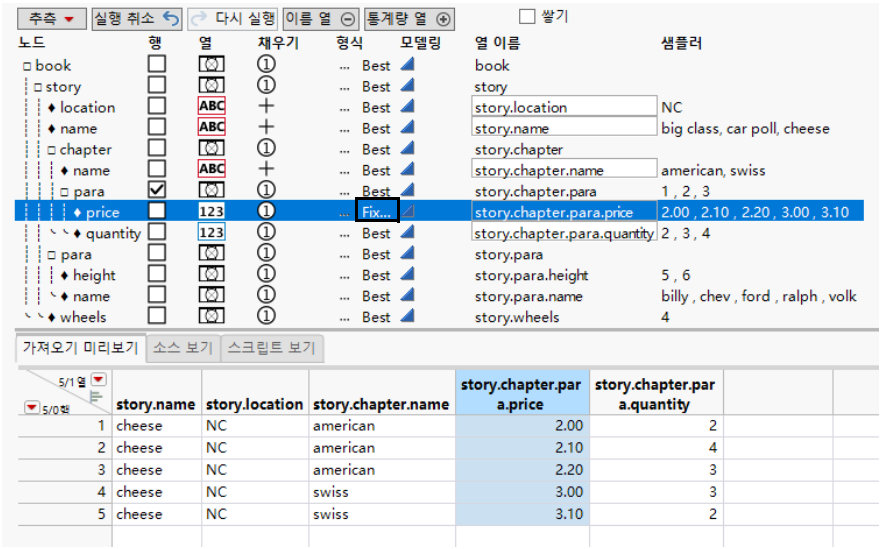


각 행을 데이터 테이블에 추가할 때 JMP 는 열에 새 값이 발견되지 않을 경우 다음 행에 대해 수행할 작업을 결정합니다 . 예를 들어 각 "cheese" 유형에 "chaphter" 이름이 지정되지 않은 경우 "american" 과 "swiss" 다음에는 결측값이 나타납니다 .

- "Use Once" 옵션은 해당 값을 한 번만 사용함을 의미하므로 , 데이터를 다시 제공하지 않으면 두 번째 , 세 번째 및 네 번째 행은 결측값이 됩니다 .
- "Fill Forever" 옵션은 해당 값을 계속 적용하되 새 값이 발견되면 새 값으로 바꿈을 의미합니다 .
- 이 예에서는 "+book/story" 옵션을 선택합니다 . 이 옵션은 여는 <book>/<story> 요소가 발견되면 <book>/<story> 노드가 다시 시작될 때까지 JMP가 해당 셀을 채움을 나타냅니다 . -<element> 옵션은 해당 요소의 닫는 태그 , 즉 노드가 끝나는 위치를 나타냅니다 .

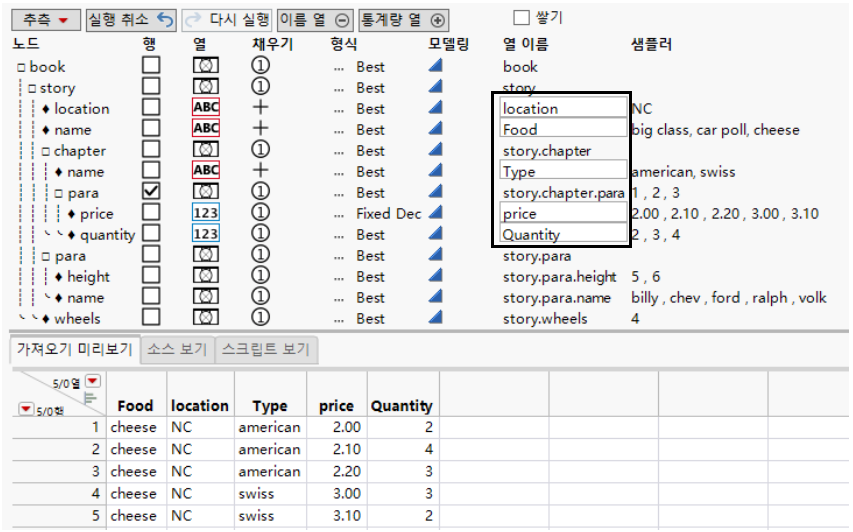
4. " 형식 " 열에서 "price" 옆의 **Best** 를 클릭하고 , **고정 소수점**을 선택하고 , 소수 자릿수로 "2" 를 입력한 다음 **확인**을 클릭합니다 .

그림 3.32 " 형식 " 열에 선택된 형식 유형



5. " 열 이름 " 열에서 그림 3.33 에 표시된 것과 같이 열 이름을 바꿉니다 .

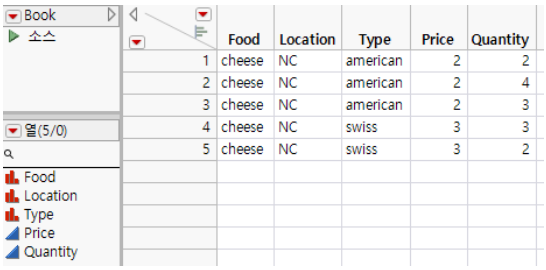
그림 3.33 " 열 이름 " 열에서 이름이 바뀐 열



6. 창 하단에서 **확인**을 클릭합니다 .

   지정한 대로 데이터를 데이터 테이블로 가져오는 작업이 수행됩니다 .

그림 3.34 최종 데이터 테이블



	Food	Location	Type	Price	Quantity
1	cheese	NC	american	2	2
2	cheese	NC	american	2	4
3	cheese	NC	american	2	3
4	cheese	NC	swiss	3	3
5	cheese	NC	swiss	3	2

XML 가져오기 마법사에서 통계량 보기

XML 가져오기 마법사의 상단에 있는 **통계량 열**을 클릭하여 요소에 대한 통계량을 볼 수 있습니다.

발생 횟수 요소의 발생 횟수입니다. 루트 요소는 한 번만 발생하며 대개 행을 만드는 데 사용되는 요소가 아닙니다. 요소에서 만드는 행의 개수는 요소의 발생 횟수와 같습니다.

여러 행을 선택할 경우에는 실제 개수가 총 " 발생 횟수 " 보다 적을 수 있으며, 다른 요소가 끝난 직후에 끝나는 요소는 추가 행을 생성하지 않습니다. 예를 들어 <para> 와 <chapter> 모두를 행 표식으로 선택할 경우 장이 마지막 단락 바로 다음에서 끝날 때는 추가 행이 생성되지 않습니다. 하지만 마지막 단락 다음과 장이 끝나는 곳 앞에 <chapter> 에 대한 <footnote> 가 있으면 추가 행을 생성할 수 있습니다. 이렇게 하려면 "footnote" 열을 선택해야 합니다.

쓰기 횟수 소스에 쓸 수 있는 값이 있는 횟수입니다. 이 횟수가 0 이면 값이 모두 결측값인 것이므로 이 요소에서 열을 만들지 않을 수 있습니다.

고유 값 수 고유 값의 수입니다 (" 쓰기 횟수 " 값보다 적거나 같음).

반복 횟수 상위 요소 내에서 이 요소가 반복된 최대 횟수입니다. 반복 횟수 값이 클수록 행을 만드는 데 적합한 요소일 수 있습니다. 이 요소가 반복되고 열로 사용되는 경우, 문자 열은 목록을 만듭니다. 이 경우 값의 중지 위치와 시작 위치를 나타내는 데 중괄호로 묶은 {lists} 가 사용되므로 표현식 열이 더 적합할 수 있습니다. 숫자 열에는 반복 값 중 하나만 사용됩니다.

샘플러 해당 요소의 값입니다.

XML 가져오기 마법사 옵션

추측 JMP 는 사용자가 선택한 데이터 테이블의 크기에 따라 설정을 선택합니다. **행 중심 추측**은 가능한 한 많은 행을 만드는 요소를 선택합니다 (기본 설정). **열 중심 추측**은 가능한 한 많은 열을 만드는 요소를 선택합니다. **최다 요소 추측**은 모든 데이터를 테이블로 가져오기 위해 필요한 만큼 행을 만드는 요소를 선택하며, 사용자 예상보다 커질 수 있습니다.

실행 취소 창에서 마지막으로 변경한 사항을 되돌립니다.

다시 실행 창에서 마지막으로 변경한 사항을 재호출합니다.

이름 열 " 열 이름 " 열을 표시하거나 숨깁니다.

통계량 열 트리에 요소에 대한 통계량을 표시하거나 숨깁니다.

쌓기 행을 생성하는 상위 노드 내에서 반복되는 노드에 적용됩니다 . 기본적으로 추가 값은 단일 테이블 셀에 쉼표로 구분되어 저장됩니다 . 이 옵션이 선택되어 있으면 추가 행에서 반복 값이 쌓입니다 .

가져오기 미리보기 사용자가 선택한 옵션에 따라 가져오는 데이터의 미리보기를 표시합니다 .

소스 보기 소스 XML 문서를 표시합니다 .

스크립트 보기 사용자가 선택한 설정의 JSL 스크립트를 표시합니다 .

XML 파일 바로 가져오기

JMP 는 XML 또는 JSON 파일의 구조를 감지하여 데이터 테이블에 바로 가져올 수 있습니다. XML 가져오기 마법사에서 데이터 미리보기를 확인하지 않으려면 이 방법을 사용합니다.

1. Windows 의 경우 :
   - **파일 > 열기**를 선택합니다 .
   - 파일 유형을 **데이터 파일**로 설정합니다 .
   - 파일을 선택합니다 .
   - " 열기 형식 " 옆에서 **데이터 (최적 추측)** 를 선택합니다 .
2. macOS 의 경우 :
   - **파일 > 열기**를 선택하고 파일을 선택합니다 .
   - 창의 왼쪽 하단에서 **옵션**을 클릭합니다 .
   - **데이터 (최적 추측)** 를 선택합니다 .
3. **열기**를 클릭합니다 .

파일이 데이터 테이블로 열립니다.

결과가 만족스럽지 않은 경우에는 XML 가져오기 마법사를 사용하여 가져오기를 사용자 정의할 수 있습니다. 자세한 내용은 "XML 및 JSON 데이터 미리보기 및 가져오기"(103 페이지)에서 확인하십시오.

참고 : 스크립트 편집기 창에서 " 편집 " > " 데이터로 가져오기 " 를 선택하여 XML 또는 JSON 텍스트를 가져올 수 있습니다 . 파일에서 아무것도 선택하지 않은 경우 가져오기는 파일 확장자를 기반으로 합니다 . 텍스트를 선택하면 JMP 는 텍스트를 JSON 또는 XML 로 파싱하려고 시도하고 그렇지 않으면 텍스트를 CSV 파일로 가져옵니다 .

JSON 파일 바로 가져오기

JSON 파일은 이름 및 값 쌍으로 구성되어 있으며 각 항목을 열 머리글과 데이터로 가져오게 됩니다. XML 가져오기 마법사를 사용하여 JSON 파일을 가져올 수도 있고("XML 및 JSON 데이터 미리보기 및 가져오기"(103 페이지)) 바로 가져올 수도 있습니다.

JSON 파일을 바로 열려면 다음 단계를 수행하십시오.

1. Windows 의 경우 :
   - **파일 > 열기**를 선택합니다 .
   - 파일 유형을 **JSON 데이터 파일**로 설정합니다 .
   - 파일을 선택합니다 .
   - **열기 형식** 옆에서 **데이터 (최적 추측)** 를 선택합니다 .
2. macOS 의 경우 :
   - **파일 > 열기**를 선택하고 파일을 선택합니다 .
   - 창의 왼쪽 하단에서 **옵션**을 클릭합니다 .
   - **데이터 (최적 추측)** 를 선택합니다 .
3. **열기**를 클릭합니다 .

파일이 데이터 테이블로 열립니다.

참고 :

- 가져온 JSON 파일에는 열 형식을 결정하기 위한 날짜 , 시간 , 통화 , 지리 , 백분율 및 과학 정보가 포함되어 있지 않습니다. 이러한 형식을 사용하려면 숫자 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 다음 " 데이터 유형 " 목록에서 형식을 변경하십시오 .
- 이름 - 값 쌍의 멤버 이름은 대 / 소문자가 구분됩니다 .
- JSON 가져오기 코드에서는 JSON 파일의 숫자 , 부울 또는 문자열 구분을 사용하지 않습니다 . 대신 전체 열이 숫자이거나 숫자가 아닌 것으로 결정합니다 . 추측 결과가 원하는 것과 다르면 명시적으로 열을 문자로 지정할 수 있습니다 .
- 내포된 항목의 경우 각 행에 대해 바깥쪽 항목이 반복됩니다 .
- 이 파일에는 빈 배열 요소가 포함될 수 없습니다 . 이름 - 값 쌍에서 이름은 따옴표로 묶이고 그 다음에 콜론이 와야 합니다 .
- JSON 표준에서는 소수점이 있는 십진수가 허용됩니다 .

쿼리 빌더에서 SQL 쿼리 작성

SQL 문을 작성하지 않고 SAS 데이터베이스에서 데이터를 선택하여 가져오려면 쿼리 빌더를 사용하는 것이 좋습니다. 데이터를 데이터 테이블로 가져오기 전에 미리 볼 수 있습니다. 다른 사용자가 쿼리를 사용자 정의하여 실행할 수 있도록 쿼리를 공유하십시오.

"파일" > "데이터베이스" > "테이블 열기" 기능을 사용하여 쿼리를 직접 작성하는 방법 대신 쿼리 빌더를 사용할 수 있습니다. 하지만 쿼리 빌더에서 쿼리 작성을 시작한 후 SQL 문을 직접 추가할 수도 있습니다.

SAS 쿼리 빌더는 SAS 서버에서 SAS 데이터 집합을 쿼리하는 데도 사용할 수 있습니다. 자세한 내용은 "SAS 쿼리 빌더를 사용하여 SAS 데이터 집합 열기"(141 페이지)에서 확인하십시오.

참고 :

- $# -+/%()&|;? 문자가 포함된 데이터베이스 테이블 이름은 지원되지 않습니다 .
- " 값 순서 " 열 특성은 지원되지 않습니다 . 행을 정렬하려면 " 쿼리 후 스크립트 " 창에서 스크립트를 작성하는 것이 좋습니다 .

SQL 데이터베이스에 연결

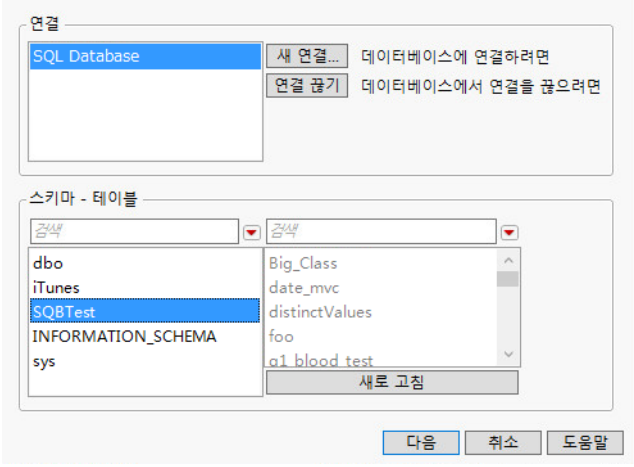
Windows 제어판을 통해 또는 JMP 내에서 ODBC 연결을 설정할 수 있습니다.

1. **파일 > 데이터베이스 > 쿼리 빌더**를 선택하여 " 데이터베이스 연결 선택 " 창을 표시합니다 .

   " 연결 " 상자에는 현재 JMP 세션에서 연결된 데이터 소스가 나열됩니다 .
2. 원하는 데이터 소스가 " 연결 " 상자에 나열되지 **않으면 새 연결**을 클릭하여 데이터 소스를 선택합니다 . 데이터 소스를 선택하는 방법은 사용하는 운영 체제 및 ODBC 드라이버에 따라 다릅니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
3. " 스키마 - 테이블 " 상자에서 테이블 또는 스키마를 선택하고 **다음**을 클릭합니다 .

   쿼리 빌더 예는 영화 대여 데이터가 포함된 SQBTest 라는 테이블을 기반으로 합니다 .

그림 3.35 데이터베이스 스키마 선택



팁 : 긴 목록에서 테이블을 찾으려면 스키마 위의 검색 상자에 이름을 입력하십시오 . 테이블 목록 위에서 테이블을 검색할 수도 있습니다 . 빨간색 삼각형 메뉴를 클릭하면 대 / 소문자를 구분하거나 정규 표현식을 사용하여 검색하기 위한 옵션이 표시됩니다 .

SQL 데이터베이스에서 테이블 선택

SQL 데이터베이스에 연결한 후에는 쿼리할 테이블을 선택합니다. 기본 테이블을 선택하거나, 여러 테이블을 결합하여 모든 테이블을 쿼리할 수 있습니다.

기본적으로 JMP는 테이블에서 할당된 키 관계에 따라 테이블을 결합하려고 시도합니다.

- 기본 키는 예를 들어 고객 ID 번호와 같이 데이터를 고유하게 설명하는 열을 식별합니다. 기본 테이블의 모든 행은 쿼리에 포함됩니다.
- 보조 테이블의 외래 키는 결합된 테이블 중 하나의 기본 키와 매칭됩니다. 보조 테이블의 매칭되는 행만 쿼리에 포함됩니다.

키가 없으면 두 테이블을 결합하는 열 이름을 기준으로 데이터가 매칭됩니다. 기본적으로는 보조 테이블의 매칭되는 행만 쿼리에 포함됩니다.

이 예에서는 여러 테이블을 결합하는 방법을 보여 줍니다. 하지만 단일 테이블을 사용하여 쿼리를 작성할 수도 있습니다. 이 경우에는 결합이 필요하지 않습니다.

참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다.

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고, 데이터베이스에 연결하고, **SQBTest** 스키마를 선택합니다. 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오.
2. " 스키마 - 테이블 " 목록에서 스키마를 선택합니다.

   테이블도 선택할 경우 " 다음 " 버튼을 클릭한 후에는 해당 테이블이 기본 테이블이 됩니다.
3. " 쿼리할 테이블 선택 " 창의 " 사용 가능한 테이블 " 목록에서 **g6_Customers** 를 선택한 후 **기본**을 클릭합니다.

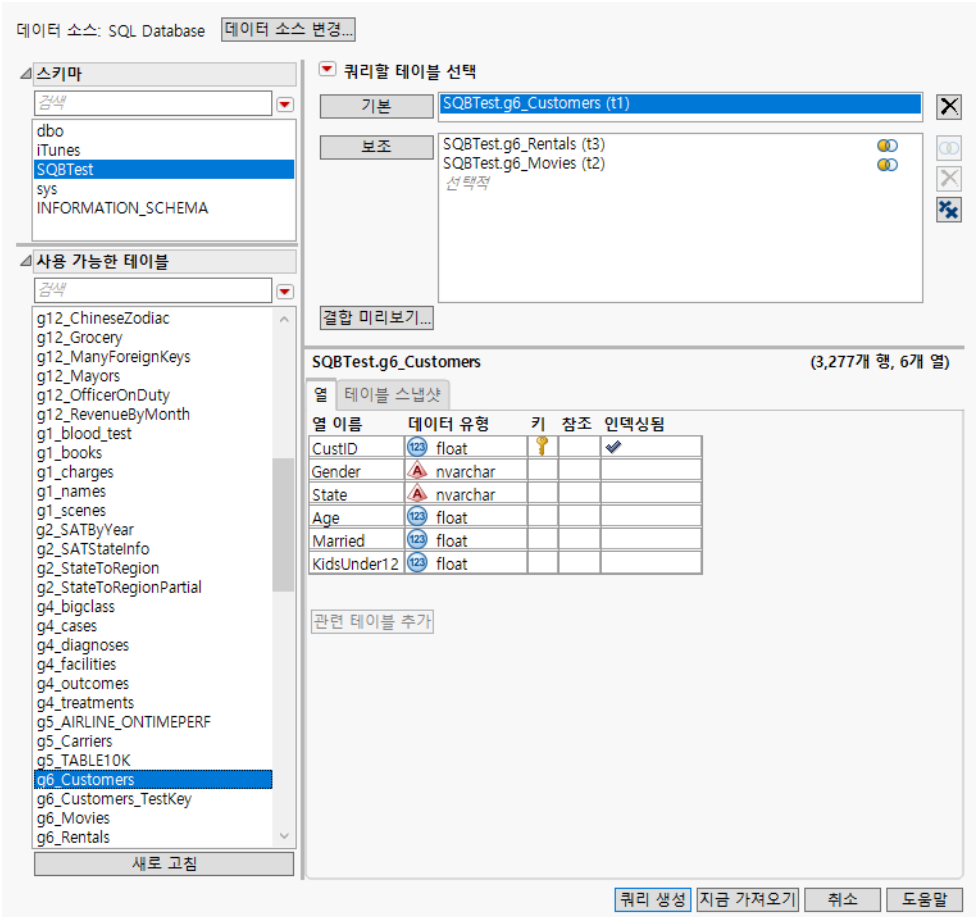
   " 열 " 탭에서는 CustID 가 기본 키임을 보여 줍니다. 또한 데이터가 인덱싱되어 있으므로 쿼리 속도가 빠릅니다.
4. " 사용 가능한 테이블 " 목록에서 **g6_Movies** 및 **g6_Rentals** 를 선택한 후 **보조**를 클릭합니다.

   왼쪽 결합 아이콘 은 테이블이 자동으로 결합되었음을 나타냅니다 (그림 3.36). CustID 는 g6_Customers 의 기본 키이며 다른 테이블 중 하나의 외래 키와 매칭됩니다.

   팁 : 기본 테이블 또는 보조 테이블을 추가한 후에 매칭되는 열이 있는 테이블을 추가하려면 **관련 테이블 추가**를 클릭하십시오. 관련 테이블을 찾을 수 없거나 기본 테이블 또는 보조 테이블이 선택되지 않은 경우에는 이 버튼을 사용할 수 없습니다.

그림 3.36 에서는 완료된 창을 보여 줍니다 .

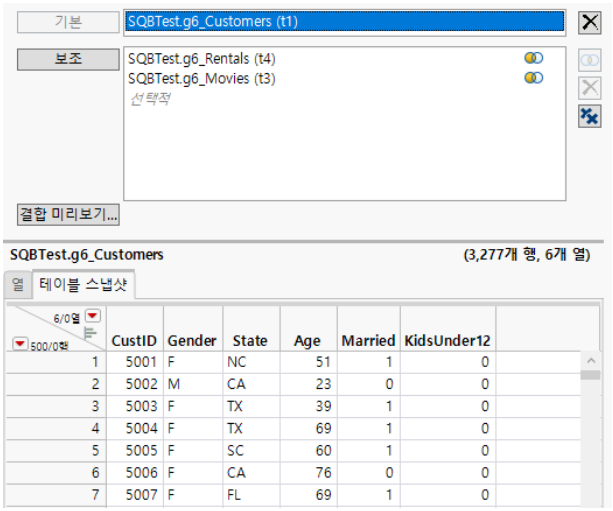
그림 3.36 기본 테이블 및 보조 테이블 선택



팁 : 긴 목록에서 스키마를 찾으려면 스키마 위의 검색 상자에 이름을 입력하십시오 . 테이블 목록 위에서 테이블을 검색할 수도 있습니다 . 빨간색 삼각형 메뉴를 클릭하면 대 / 소문자를 구분하거나 정규 표현식을 사용하여 검색하기 위한 옵션이 표시됩니다 .

5. 각 테이블의 **테이블 스냅샷** 탭을 클릭하여 데이터를 미리 봅니다 .

그림 3.37 테이블 스냅샷 탭



6. 기본 테이블과 보조 테이블 아래에서 **결합 미리보기**를 클릭하여 지정한 결합을 통해 생성된 테이블을 미리 봅니다 .

팁 :

- 보조 테이블 옆의 아이콘은 해당 테이블이 쿼리에서 결합되지 않았음을 나타냅니다 . 결합할 열을 지정하려면 **결합 편집** 버튼 을 클릭합니다 . 결합할 열을 찾을 수 없으면 **제거** 버튼 을 클릭하여 해당 테이블을 제거합니다 . 자세한 내용은 테이블 결합 조건 편집에서 확인하십시오 .
- " 열 " 탭에서 " 키 " 열에는 여러 개의 키가 표시될 수도 있으며 , 일부 열은 기본 키이면서 동시에 외래 키일 수 있습니다 . 기본 키는 모두 고유하므로 기본 키 옆에는 고유 키 아이콘이 나타나지 않습니다 .
- " 열 " 탭에서 다른 스키마의 기본 키와 매칭되는 외래 키에는 " 참조 " 가 지정됩니다 . 참조는 스키마 및 열의 이름입니다 .
- 다른 스키마 또는 데이터베이스를 쿼리하려면 **데이터 소스 변경**을 클릭합니다 .
- 다른 소스 (예 : 데이터베이스 및 Microsoft Excel) 의 데이터를 결합하려면 쿼리 빌더를 사용하여 해당 데이터베이스 데이터를 데이터 테이블로 가져옵니다 . 예를 들어 Excel 데이터를 데이터 테이블로 가져온 후 " 테이블 " 메뉴의 "JMP 쿼리 빌더 " 를 사용하여 테이블을 쿼리하고 결합할 수 있습니다 .
- 기본 키가 포함된 테이블을 가져올 경우 데이터 테이블의 해당 열에 " 연결 ID" 열 특성이 추가됩니다 . 이 열 특성을 사용하면 데이터 테이블을 가상으로 결합할 수 있습니다 . 자세한 내용은 " 데이터 재구성 " 장의 " 데이터 테이블 가상 결합 "(344 페이지) 에서 확인하십시오 .

테이블 결합 조건 편집

"쿼리할 테이블 선택"의 빨간색 삼각형 메뉴에서 "데이터베이스 테이블 자동 결합"은 처음부터 선택되어 있습니다. JMP에서는 키 관계나 매칭되는 열 이름을 기준으로 데이터베이스 테이블이 자동으로 결합됩니다.

키가 없거나 열 이름이 매칭되지 않는 경우에는 **결합 편집** 을 클릭하여 결합할 열을 지정합니다.

참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

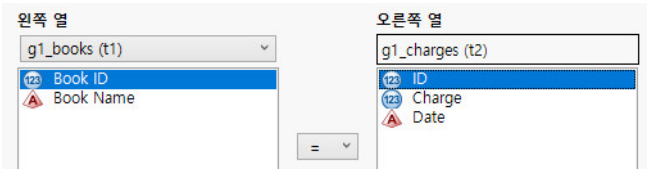
테이블 결합 조건을 편집하려면

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
2. "쿼리할 테이블 선택" 창에서 기본 테이블로 **g1_books**를 선택하고 보조 테이블로 **g1_charges**를 선택합니다 .

   보조 테이블 옆의 아이콘은 해당 테이블이 쿼리에서 결합되지 않았음을 나타냅니다 .
3. " 보조 " 테이블 페인에서 **g1_charges** 를 선택하고 **결합 편집** 을 클릭합니다 .

   " 조건 추가 " 창이 나타납니다 .
4. " 왼쪽 열 " 목록에서 **g1_books** 를 선택합니다 .
5. " 왼쪽 열 " 상자에서 **Book ID** 를 선택합니다 .
6. " 오른쪽 열 " 상자에서 **ID** 를 선택합니다 .
7. 두 상자 사이에 등호가 선택되어 있는지 확인합니다 .

그림 3.38 조건 추가 창



8. **다음**을 클릭합니다 .

   " 결합 편집 " 창에서는 g1_books 의 매칭되지 않는 행이 데이터 테이블에 포함됨을 보여 줍니다 . g1_charges 에만 있는 행은 생략됩니다 .

   전체 결합을 수행하고 모든 행을 가져오려면 **g1_charges 의 매칭되지 않는 행 포함**을 선택합니다 . 매칭되지 않는 행에 대한 옵션 중 하나만 사용할 수 있으면 해당 데이터베이스가 전체 결합을 지원하지 않는 것입니다 .
9. **확인**을 클릭합니다 .

참고 : 모든 보조 테이블이 결합되기 전까지는 **확인** 버튼을 사용할 수 없습니다 .

테이블이 자동으로 결합되지 않도록 하려면

- 기본 테이블 위에 있는 " 쿼리할 테이블 선택 " 의 빨간색 삼각형 메뉴에서 **데이터베이스 테이블 자동 결합**을 선택 취소합니다 .
- 큰 데이터베이스를 자주 쿼리하는 경우 메모리 문제를 방지하려면 **환경 설정 > 쿼리 빌더**에서 **쿼리에 추가된 테이블 자동 결합**을 선택 취소합니다 .

SQL 쿼리 생성

데이터베이스 테이블을 선택한 후에는 데이터를 가져오거나 쿼리를 작성합니다. 쿼리 빌더를 사용하면 SQL 표현식을 작성하지 않고 대화식으로 데이터베이스 쿼리를 생성할 수 있습니다.

데이터베이스 테이블을 선택하고 필요한 경우 이를 결합한 후 **쿼리 생성**을 클릭하여 쿼리 빌더 창을 엽니다. 계속해서 포함할 열을 선택하고 표집 및 필터링 기준을 지정하여 쿼리를 구체화할 수 있습니다. 쿼리를 저장했다가 나중에 다시 편집 및 실행할 수도 있습니다.

"사용 가능한 열" 목록에는 모든 데이터베이스 테이블의 열이 표시됩니다. t1 및 t2와 같은 접두사(**별칭이라고도 함**)는 각 열을 해당하는 데이터베이스 테이블과 연결합니다.

쿼리 빌더 단계를 건너뛰고 모든 데이터를 가져오려면 그 대신 **지금 가져오기**를 클릭합니다.

참고 : " 테이블 " 메뉴의 "JMP 쿼리 빌더 " 에서도 동일한 여러 옵션을 제공하지만 이를 통해서는 JMP 데이터 테이블을 쿼리하고 결합할 수 있습니다. 자세한 내용은 " 데이터 재구성 " 장의 "JMP 쿼리 빌더를 사용하여 데이터 테이블 쿼리 및 결합 "(341 페이지) 에서 확인하십시오 .

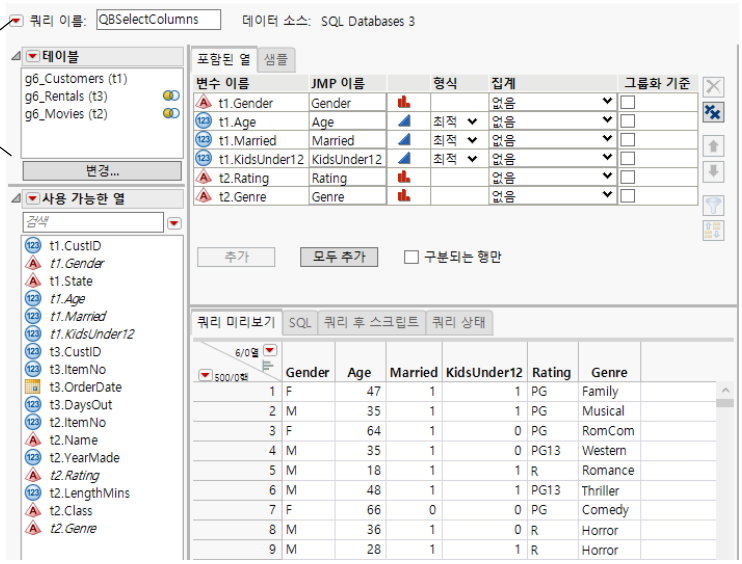
데이터베이스 테이블에서 열 선택

영화 장르, 등급, 그리고 결혼 여부 및 연령 같은 인구 통계학 데이터별로 영화 대여 정보를 확인하려고 한다고 가정하겠습니다.

참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다. 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
2. " 쿼리할 테이블 선택 " 창에서 **g6_Customers** 를 선택하고 **기본**을 클릭합니다 .
3. **g6_Movies** 및 **g6_Rentals** 를 선택하고 **보조**를 클릭합니다 .
4. **쿼리 생성**을 클릭하여 쿼리 빌더 창을 표시합니다 .
5. " 사용 가능한 열 " 상자에서 t1.Gender, t1.Age, t1.Married, t1.KidsUnder12, t2.Rating 및 t2.Genre 를 선택합니다 .
6. " 포함된 열 " 탭에서 **추가**를 클릭합니다 .

그림 3.39 선택한 열



참고 :

- 쿼리에 8 개 이상의 테이블이 포함되어 있으면 왼쪽 상단 부분의 테이블 이름 위에 검색 상자가 나타납니다 . 긴 목록에서 열을 찾으려면 " 사용 가능한 열 " 아래의 검색 상자에 이름을 입력하십시오 . 빨간색 삼각형 메뉴를 클릭하면 대 / 소문자를 구분하거나 정규 표현식을 사용하여 검색하기 위한 옵션이 표시됩니다 . 검색 상자를 숨기려면 " 사용 가능한 열 " 의 빨간색 삼각형 메뉴에서 **검색 상자 표시**를 선택 취소하십시오 .
- 참조 열은 쿼리에서 지원되지 않으므로 " 사용 가능한 열 " 목록에 더 이상 나타나지 않습니다 .
- " 포함된 열 " 목록의 " 형식 " 목록에서 옵션을 선택하여 열 너비를 변경할 수 있습니다 .

7. 열 아래의 **SQL** 탭을 선택하여 쿼리의 SQL 문을 확인합니다 . 이 코드는 쿼리를 실행한 후에 데이터 테이블 특성으로 저장됩니다 .
8. 오른쪽 하단에서 **저장**을 클릭합니다 .

   작업이 g6_Customers.jmpquery 로 저장됩니다 . 나중에 이 파일을 열어 이 지점으로 돌아오거나 쿼리를 실행할 수 있습니다 .
9. **쿼리 실행**을 클릭하여 데이터를 가져옵니다 .

   데이터 테이블에는 다음과 같은 스크립트가 포함되어 있습니다 .
   - 데이터베이스에 다시 연결하고 데이터를 가져오려면 " 소스 " 스크립트를 실행합니다 .
   - 쿼리 빌더에서 쿼리를 열려면 " 쿼리 수정 " 스크립트를 실행합니다 .
   - 데이터를 다시 가져오고 새로 고치려면 " 데이터베이스에서 업데이트 " 스크립트를 실행합니다 .

팁 :

- 열 이름을 바꾸려면 "포함된 열" 탭에서 JMP 이름을 두 번 클릭하고 새 이름을 입력하십시오.
- 별칭 이름을 바꾸려면 " 쿼리할 테이블 선택 " 창에서 테이블을 마우스 오른쪽 버튼으로 클릭하고 **별칭 변경**을 선택하십시오 . 별칭은 대 / 소문자가 구분되지 않습니다 .
- " 쿼리 빌더 " 의 "ODBC" 환경 설정에서 **가능한 경우 백그라운드에서 쿼리 실행**을 선택 취소하지 않은 한 쿼리는 백그라운드에서 실행됩니다 . **보기 > 실행 중인 쿼리**를 선택하여 모든 ODBC 쿼리의 진행률을 확인할 수도 있습니다 .

참고 : SAS 쿼리 빌더의 경우에는 모든 쿼리가 포그라운드에서 실행됩니다 .

- 미리보기 로드 속도가 너무 느리면 **미리보기 자동 업데이트**를 선택 취소하십시오 . 데이터 보기를 업데이트하려면 " 쿼리 미리보기 " 탭 아래의 **업데이트**를 클릭하십시오 . 큰 데이터베이스를 사용하여 작업할 때가 많은 경우 JMP " 쿼리 빌더 " 환경 설정에서 미리보기 옵션을 변경하는 것이 좋습니다 . 미리 볼 수 있는 최대 행 수를 제한하는 것이 좋습니다 . JMP " 쿼리 빌더 " 환경 설정에서 **미리보기 최대 행 수**의 값을 변경하십시오 .
- 데이터베이스에서 중복 행을 생략하려면 "포함된 열" 탭에서 **구분되는 행만**을 선택하십시오.

JMP 12 와의 호환성 유지

쿼리에 JMP 16 기능을 추가하면 해당 쿼리는 JMP 12 에서 더 이상 로드되지 않습니다. JMP 13 을 사용하고 있지만 JMP 12 에서도 실행되는 쿼리를 생성해야 하는 경우 "쿼리 빌더" 환경 설정에서 **이 쿼리를 JMP 12와 호환되는 상태로 유지**를 선택하십시오. 이 옵션을 선택한 후에는 호환성 문제를 발생시키는 기능이 쿼리 빌더에서 숨겨집니다.

JMP 12 쿼리를 최근 버전으로 이동할 준비가 되었으면 이 환경 설정을 선택 취소하십시오.

계산된 열 생성

기존 열에서 새 열을 생성할 수 있습니다. 예를 들어 두 열의 평균을 계산하여 새 열에 저장할 수 있습니다. 날짜/시간 값은 형식이 잘못될 수 있습니다. "사용 가능한 열"의 빨간색 삼각형을 클릭하고 **계산된 열 추가**를 선택한 다음 계산식 편집기에서 새 열을 생성하십시오.

예를 들어 한 영화를 대여 기간 동안 볼 수 있는 최대 횟수를 계산하려고 한다고 가정합니다. 각 영화의 길이와 영화가 체크아웃된 기간(일)이 포함되어 있는 데이터베이스를 쿼리하려고 합니다. 이 예에서는 이러한 데이터에서 계산된 열을 새로 생성하는 방법을 보여 줍니다.

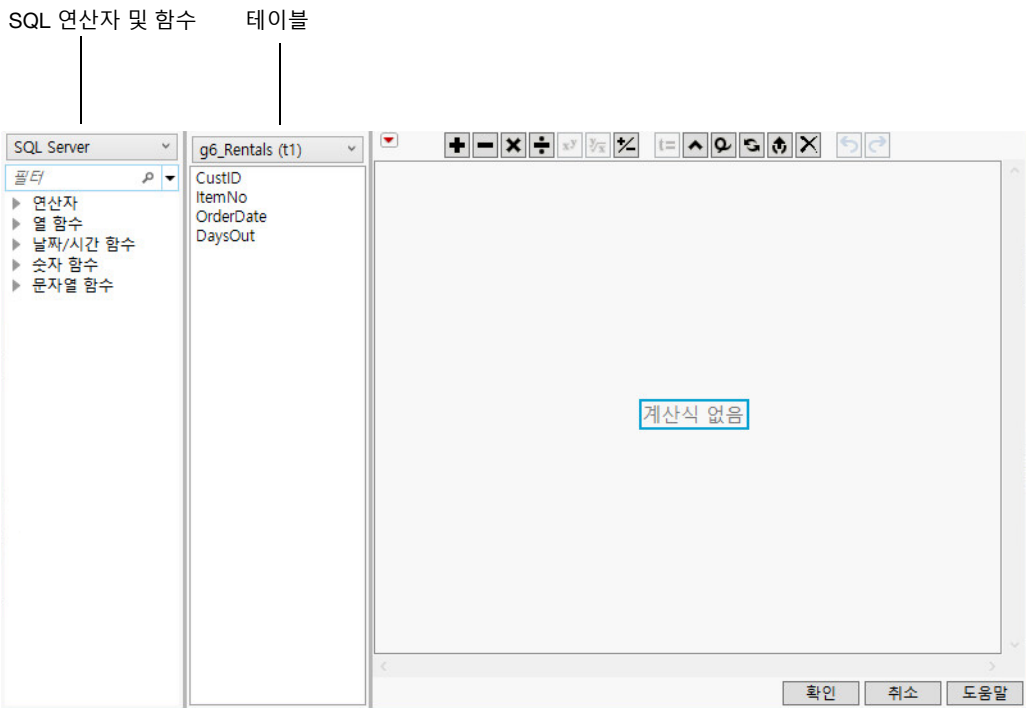
참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
2. " 쿼리할 테이블 선택 " 창에서 기본 테이블로 **g6_Rentals** 를 선택하고 보조 테이블로 **g6_Movies** 를 선택합니다 .
3. **쿼리 생성**을 클릭하여 쿼리 빌더 창을 표시합니다 .

4. " 사용 가능한 열 " 의 빨간색 삼각형을 클릭하고 **계산된 열 추가**를 선택합니다 .

   " 계산된 열 " 창이 나타납니다 . 이 창에는 JMP 계산식 편집기가 포함되어 있습니다 .

그림 3.40 계산식 편집기가 포함된 계산된 열 창



참고 :

- 연산자와 함수는 계산식 편집기의 왼쪽에 있는 목록에서 제공됩니다 (그림 3.40). 일부 경우에는 데이터베이스에 따라 서버 유형을 변경해야 할 수 있습니다 .
- 연산자 목록에서 연결 (||) 연산자는 제공되지 않습니다 . 해당 계산식은 계산식 편집기 상자에서 직접 입력해야 합니다 .

5. 왼쪽의 g6_Rentals 목록에서 **DaysOut** 을 선택하고 곱하기 ✕ 버튼을 클릭합니다 .

그림 3.41 계산된 열

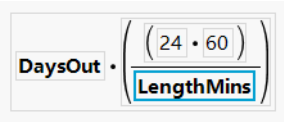
6. 빈 상자를 선택하고 (24 * 60 을 입력한 후 Enter 키를 누릅니다 .

   이 계산식은 하루의 시간 수에 한 시간의 분 수를 곱합니다 . 첫 번째 괄호를 입력하면 두 번째 괄호는 자동으로 삽입됩니다 .

그림 3.42 계산식의 첫 번째 부분

7. 바깥쪽 상자를 클릭하여 전체 방정식을 선택하고 나누기 ÷ 버튼을 클릭합니다.
8. 왼쪽 목록에서 **g6_Movies** 를 선택한 후 **LengthMins** 를 선택합니다.

그림 3.43 계산식의 두 번째 부분



9. **확인**을 클릭합니다.

   Calc1 이라는 새 열이 생성됩니다.
10. 이 열을 마우스 오른쪽 버튼으로 클릭하고 **열 이름 바꾸기**를 선택합니다.
11. MaximumTimesWatched 를 입력하고 **확인**을 클릭합니다.
12. " 사용 가능한 열 " 목록에서 MaximumTimesWatched 를 선택하고 **추가**를 클릭합니다.
13. t2.Name 을 선택하고 **추가**를 클릭합니다.

    " 쿼리 미리보기 " 탭을 확인하면 Nanny McPhee 는 대여한 영화를 대여 기간 동안 160 회 볼 수 있는 것으로 나타납니다.

공통 값 그룹화

데이터를 JMP 로 가져오기 전에 열의 공통 값을 결합(또는 그룹화)할 수 있습니다. 공통 값을 그룹화하려면 집계 함수를 선택하여 공통 값 계산 방식을 결정하십시오.

참고 : 집계 지원은 데이터베이스에 따라 달라집니다. 자세한 내용은 데이터베이스 설명서에서 확인하십시오.

특정 영화가 대여된 횟수에 관심이 있다고 가정해 보겠습니다. 이 예에서는 각 항목 번호의 개수를 계산하고 공통된 영화 값을 단일 행에 그룹화합니다.

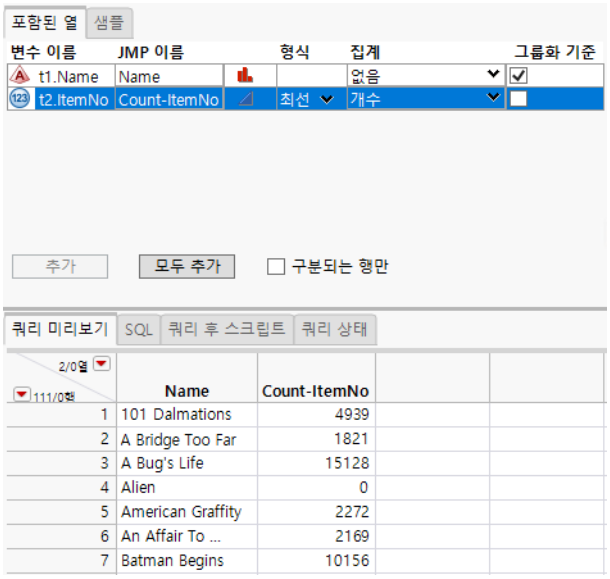
참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다.

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고, 데이터베이스에 연결하고, **SQBTest** 스키마를 선택합니다. 자세한 내용은 " 데이터베이스에 연결 "(157 페이지)에서 확인하십시오.
2. " 쿼리할 테이블 선택 " 창에서 기본 테이블로 **g6_Movies** 를 선택하고 보조 테이블로 **g6_Rentals** 를 선택합니다.
3. **쿼리 생성**을 클릭하여 쿼리 빌더 창을 표시합니다.

4. " 사용 가능한 열 " 상자에서 t1.Name 및 t2.ItemNo 를 선택하고 **추가**를 클릭합니다 .
5. t2.ItemNo 를 선택하고 " 집계 " 목록에서 **개수**를 선택합니다 .

   t1.Name 에는 " 그룹화 기준 " 체크박스가 선택되어 있습니다 . 그러면 특정 영화 이름의 모든 인스턴스가 하나의 행에 그룹화됩니다 .

그림 3.44 그룹화된 열



6. **쿼리 실행**을 클릭하여 데이터를 가져옵니다 .
7. 데이터 테이블에서 Count-ItemNo 열을 마우스 오른쪽 버튼으로 클릭하고 **정렬 > 내림차순**을 선택합니다 .

   Scarface 가 가장 빈번하게 대여되었습니다 .

그림 3.45 정렬된 Count-ItemNo 열 (일부분)

	Name	Count-ItemNo
1	Scarface	15991
2	Rambo	15643
3	Nanny McPhee	15289
4	Marmaduke	15255
5	A Bug's Life	15128
6	Diary of a Wimp...	14989
7	Cars	14905

팁 :

- 그룹화된 행을 지우려면 해당 열의 " 집계 " 목록에서 **없음**을 선택하십시오 .

- DISTINCT 집계 함수는 구분되는 값이 포함된 행만 보여 줍니다 . 중복 값이 있는 행은 생략됩니다 . 이러한 함수는 데이터베이스에 중복 값이 여러 개 포함된 경우에 유용합니다 .

데이터 표본 가져오기

데이터베이스가 클 경우 데이터를 표집하는 것이 좋습니다. 표집 기능은 행의 부분집합을 반환하며 쿼리 시간을 줄여 줍니다. 데이터베이스 쿼리가 실행되어 사용자가 "표본" 탭에서 선택한 옵션에 따라 데이터의 일부만 가져오게 됩니다.

표집 방법은 데이터베이스 벤더에 따라 달라집니다.

- SQL Server 는 기본적으로 블록 표집을 지원합니다 . 블록 표본에는 특정 페이지 전체의 행 (예 : 1 페이지와 5 페이지의 모든 행) 이 포함됩니다 . 1,000 개의 행을 선택하면 약 1,000 개의 행을 가져오게 됩니다 .
- Oracle 및 기타 데이터베이스는 행 표집을 지원합니다 . 5,000 개의 행을 선택하면 일반적으로 Oracle 의 데이터 순환 방식에 따라 표본당 4,800 개에서 5,200 개 사이의 행을 가져오게 됩니다 .

주요 데이터베이스 벤더의 경우 JMP가 기능을 감지하고 가능한 경우 벤더에 맞는 옵션을 제공합니다. 벤더가 지원하지 않는 기능은 "표본" 탭에서 사용할 수 없습니다.

데이터 표본을 가져오려고 한다고 가정해 보겠습니다. 이 예에서는 5,000 개의 랜덤 행을 선택합니다.

참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
2. " 쿼리할 테이블 선택 " 창에서 기본 테이블로 **g6_Rentals** 를 선택하고 보조 테이블로 **g6_Movies** 를 선택합니다 .
3. **쿼리 생성**을 클릭하여 쿼리 빌더 창을 표시합니다 .
4. " 포함된 열 " 탭에서 **모두 추가** 버튼을 클릭합니다 .
5. **표본** 탭을 클릭하고 **이 결과 집합 표집**을 선택합니다 .
6. **임의 N 개 행**을 선택하고 5,000 을 입력합니다 .

   " 표집 기준 " 영역에는 데이터베이스가 지원하는 표집 유형에 따라 " 블록 또는 페이지 " 옵션만 표시됩니다 .
7. **쿼리 실행**을 클릭하여 데이터를 가져옵니다 .

   새 데이터 테이블이 약 5,000 개의 행으로 구성됩니다 . 블록 또는 페이지 표집을 사용하는 경우 , 예를 들어 한 번은 4,900 개의 행 , 다음엔 5,600 개의 행으로 구성된 표본을 얻을 수 있습니다 .

팁 :

- 쿼리를 실행할 때마다 동일한 표본 집합을 다시 생성하려면 시드값을 최대 64,000 인 양의 정수로 설정하십시오. 예를 들어 성별을 기준으로 영화 대여 정보를 쿼리하려고 한다고 가정해 보겠습니다. 시드값으로 1 을 입력하고 쿼리를 실행해 보십시오. 결과에서 남성 고객의 분포가 낮게 나타난다면 시드값으로 2 를 입력하고 쿼리를 다시 실행해 보십시오. 남성과 여성의 분포가 유사하게 되는 시드값을 찾을 때까지 이 프로세스를 반복하십시오.
- " 포함된 열 " 탭에 개별 열을 추가하려면 해당 열을 마우스 오른쪽 버튼으로 클릭하고 **열 포함**을 선택하거나 **추가** 버튼을 클릭하십시오.

필터를 선택하여 데이터 부분집합 가져오기

선택한 필터로 얻은 값의 부분집합을 데이터 테이블로 가져오려면 필터를 추가하십시오. 일부 필터는 쿼리가 JMP 12와 호환되지 않는 경우에는 사용할 수 없습니다. 자세한 내용은 "JMP 12와의 호환성 유지"(118 페이지)에서 확인하십시오.

연속형 및 범주형 열에 사용 가능한 필터

단순 비교 지정된 연산자를 사용하여 값을 매칭합니다.

Age > 14 는 14 세를 초과하는 연령과 매칭됩니다.

범위 지정된 연산자를 사용하여 값 범위를 매칭합니다.

12 ≤ Age ≤ 17 은 12 세 이상 17 세 이하인 연령과 매칭됩니다.

Is NULL 또는 Is Not NULL 결측값을 매칭합니다.

NULL 또는 not NULL 은 결측값과 비결측값을 매칭합니다.

사용자 표현식 SQL 표현식을 직접 작성할 수 있습니다.

```
( ( ( t2.Gender IN ( 'F' ) ) AND ( (t2.Age >= 20) AND (t2.Age <= 50)
  ) ) )
```

이 표현식은 F Gender 와 매칭됩니다. 또한 20 세 이상 50 세 이하인 Age 와도 매칭됩니다.

범주형 열에만 사용 가능한 필터

목록 상자 하나 이상의 값을 선택할 수 있는 목록 상자를 표시합니다. 목록 상자는 " 쿼리 빌더 " 환경 설정에 따라 범주형 열의 기본 필터입니다.

수동 목록 열 값을 직접 입력할 수 있습니다.

체크박스 목록 체크박스 목록을 표시합니다.

참고 : 목록 상자, 수동 목록 및 체크박스 목록에는 선택된 값과 매칭되지 않는 행을 검색할 수 있는 **목록에 없음** 옵션이 포함되어 있습니다.

Contains 지정된 값을 포함하거나 포함하지 않는 문자열을 매칭합니다. **Contains Comedy OR Romance** 는 Comedy 및 Romance 를 매칭합니다.

대부분의 범주형 열에 대한 기본 필터는 목록 상자입니다 . 포함된 수준이 1,000 개를 초과하는 열의 경우에는 " 포함 " 필터가 자동으로 선택됩니다 . " 쿼리 빌더 " 환경 설정에서 이 수준 수를 변경할 수 있습니다 .

같음 또는 같지 않음 지정된 값과 유사하거나 유사하지 않은 문자열을 매칭합니다 . % 와일드카드 (0 개 이상의 문자) 와 _ 와일드카드 (정확히 1 개의 문자) 를 지원합니다 .

Genre Like %com 은 "RomCom" 과 같이 "com" 앞에 임의 개수의 문자가 있는 경우를 매칭합니다 . "Comedy" 도 매칭하려면 **%com%** 또는 **Contains com** 을 사용합니다 .

열 값 매칭 지정된 열 값을 매칭합니다 . 테이블을 선택한 후 열을 선택합니다 . **매칭되지 않는 항목 선택** 옵션을 사용하면 선택한 행을 제외한 모든 행을 필터링할 수 있습니다 . 예는 " 기존 데이터 테이블에서 매칭되는 데이터 가져오기 "(126 페이지) 에서 확인하십시오 .

이 예에서는 연령 30 세 이상인 고객과 RomCom 및 Comedy 장르의 영화에 대한 데이터를 가져오는 방법을 보여 줍니다 .

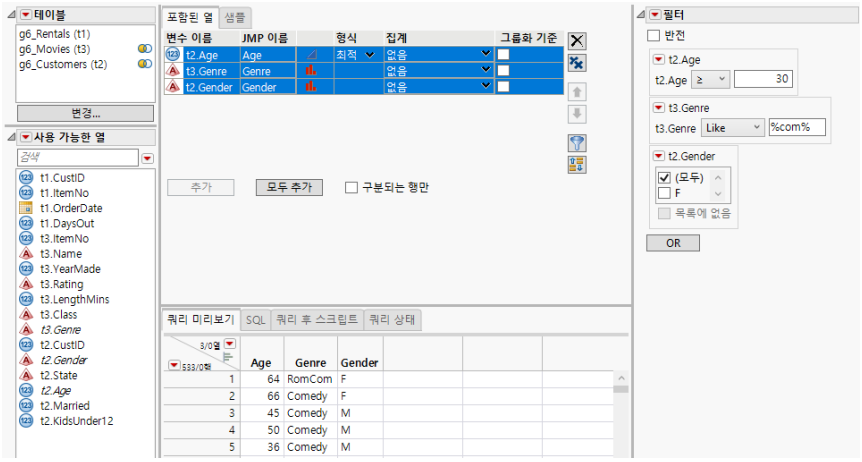
참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. **파일 > 새로 만들기 > 데이터베이스 쿼리** 를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
2. " 쿼리할 테이블 선택 " 창의 " 사용 가능한 테이블 " 목록에서 **g6_Rentals** 를 선택한 후 **기본** 을 클릭합니다 .
3. **g6_Customers** 및 **g6_Movies** 를 선택한 후 **보조** 를 클릭합니다 .
4. **쿼리 생성** 을 클릭하여 쿼리 빌더 창을 표시합니다 .
5. " 사용 가능한 열 " 상자에서 t2.Gender, t2.Age 및 t3.Genre 를 선택한 후 " 포함된 열 " 탭에서 **추가** 를 클릭합니다 .
6. " 포함된 열 " 탭의 모든 열을 선택하고 **선택한 항목을 필터에 추가** 를 클릭합니다 .

   " 필터 " 개요에 해당 열에 대한 필터가 나타납니다 .
7. t2.Age 필터를 " ≥ 30" 으로 설정합니다 .
8. t3.Genre 의 빨간색 삼각형을 클릭하고 **Like** 를 선택하고 %com% 를 입력한 후 **Enter** 키를 누릅니다 .

   % 와일드카드는 "com" 앞뒤에 임의 개수의 문자가 있는 경우를 매칭합니다 . " 쿼리 미리보기 " 탭에 RomCom 및 Comedy 장르의 영화가 모두 표시됩니다 .

그림 3.46 필터 선택



9. " 필터 " 의 빨간색 삼각형을 클릭하고 **실행 시 항상 확인**을 선택합니다 .

   그러면 이 쿼리를 실행하는 사용자가 필터를 사용자 정의할 수 있습니다 .

10. **쿼리 실행**을 클릭합니다 .

11. " 쿼리 확인 " 창에서 **확인**을 클릭하여 사전 선택한 필터를 적용하고 데이터를 가져옵니다 .

참고 :

- 대부분의 범주형 열에 대한 기본 필터는 목록 상자입니다 . 포함된 수준이 1,000 개를 초과하는 열의 경우에는 " 포함 " 필터가 자동으로 선택됩니다 . " 쿼리 빌더 " 환경 설정에서 이 수준 수를 변경할 수 있습니다 .
- 필터의 빨간색 삼각형 메뉴에 있는 " 조건부 " 옵션을 사용하면 계층적 범주 내의 데이터를 필터링할 수 있습니다 . 예를 들어 State(주) 필터와 City(도시) 필터가 있다고 가정해 보겠습니다 . 주를 선택한 후 해당 주에 있는 도시만 표시하려면 "City" 의 빨간색 삼각형을 클릭하고 " 조건부 " 를 선택하십시오 .
- " 필터 " 목록의 상단에 있는 " 반전 " 옵션을 사용하면 모든 필터에 대해 지정한 행을 제외한 모든 행을 선택할 수 있습니다 . 모든 행을 선택하는 필터에는 이 옵션을 사용할 수 없습니다 .
- 필터 목록의 "< 공백 >" 은 데이터베이스에 해당 열에 대한 결측값이 포함되어 있음을 나타냅니다 .
- 범주형 데이터가 있는 큰 열에 대한 필터를 생성할 경우 JMP 는 테이블의 행 수를 확인하려고 합니다 .
  - **크기를 확인할 수 없는 테이블에 대한 범주 수준 가져오기**라는 " 쿼리 빌더 " 환경 설정이 기본적으로 선택되어 있으므로 JMP 는 수준을 자동으로 가져옵니다 . 이 환경 설정을 선택 취소하면 " 쿼리 빌더 " 환경 설정에서 범주형 열에 대한 대체 필터 유형 " 포함 " 이 선택됩니다 .
  - 범주형 열의 행 수가 백만 개를 초과할 경우 JMP 는 필터링된 열에 대해 고유한 범주 수준을 자동으로 가져오지 않습니다 . **범주 수준을 자동으로 가져오는 테이블의 최대 행 수**라는

" 쿼리 빌더 " 환경 설정에서는 최소 -1(제한 없음) 부터 최대 1,000,000,000(10 억 개의 행) 까지 지원됩니다 .

- " 쿼리 빌더 " 환경 설정에서 **이 쿼리를 JMP 12 와 호환되는 상태로 유지**가 선택되어 있지 않은 한 범주형 열의 기본 필터는 목록 상자입니다 .

기존 데이터 테이블에서 매칭되는 데이터 가져오기

열려 있는 데이터 테이블에서 쿼리의 열과 매칭되는 행을 선택할 수도 있습니다. 항공사 데이터의 데이터베이스를 예로 들겠습니다. 이 데이터베이스에는 비행 시간 및 등록 번호 같은 데이터가 포함되어 있습니다. 또한 사용자에게도 등록 번호 데이터가 포함된 데이터 테이블이 있습니다. 이 경우 " 열 값 매칭 " 필터를 사용하여 매칭되는 등록 번호에 대한 데이터만 가져올 수 있습니다.

참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Air Traffic.jmp 를 엽니다 .
2. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
3. " 쿼리할 테이블 선택 " 창의 " 사용 가능한 테이블 " 목록에서 **g5_AIRLINE_ONTIMEPERF** 를 선택한 후 **기본**을 클릭합니다 .
4. **쿼리 생성**을 클릭하여 쿼리 빌더 창을 표시합니다 .
5. " 포함된 열 " 탭에서 **모두 추가**를 클릭합니다 .
6. " 포함된 열 " 탭에서 t1.TailNum 을 선택하고 **선택한 항목을 필터에 추가** 를 클릭합니다 .
7. " 필터 " 열에 있는 t1.TailNum 의 빨간색 삼각형을 클릭하고 **필터 유형**을 선택한 후 **열 값 매칭**을 선택합니다 .
8. **테이블의 매칭 값** 아래에서 **Air Traffic** 을 선택합니다 .
9. Tail Number 열을 선택한 후 목록에서 **모든 행 (38,118)** 을 선택합니다 .

   " 쿼리 미리보기 " 탭의 데이터 보기가 업데이트되어 필터링된 값이 표시됩니다 .
10. **쿼리 실행**을 클릭하여 데이터를 가져옵니다 .

    데이터 테이블에 Tail Number 열에 있는 행의 데이터만 포함됩니다 .

사용자 표현식을 작성하여 데이터 부분집합 가져오기

필터를 선택하여 데이터의 부분집합을 추출할 수 있을 뿐만 아니라 , 제공된 필터를 사용하지 않으려는 경우에는 사용자 SQL 표현식을 작성할 수 있습니다 .

참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. 필터링할 열을 선택합니다("필터를 선택하여 데이터 부분집합 가져오기"(123 페이지) 참조).
2. " 필터 " 의 빨간색 삼각형을 클릭하고 **사용자 표현식 추가**를 선택합니다 .

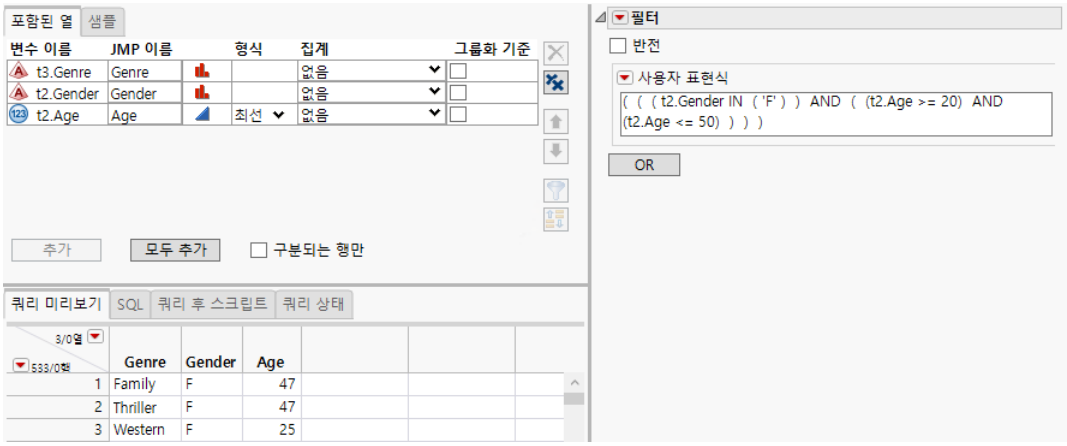
3. " 사용자 표현식 " 상자에 다음 텍스트를 입력합니다 .

```
( ( ( t2.Gender IN ( 'F' ) ) AND ( (t2.Age >= 20) AND (t2.Age <= 50)
) ) )
```

4. " 사용자 표현식 " 상자의 바깥쪽을 클릭하여 " 쿼리 미리보기 " 탭을 업데이트합니다 .

   이 표현식은 F Gender 와 매칭됩니다 . 또한 20 세 이상 50 세 이하인 Age 와도 매칭됩니다 .

그림 3.47 사용자 필터 표현식 작성



선택한 데이터 정렬

특정 열의 행을 값을 기준으로 정렬하여 데이터 테이블에 데이터가 표시되는 방식을 제어할 수 있습니다. 이 예에서는 "Married" 열을 내림차순으로 정렬하고, 연령순 및 키순으로 데이터를 정렬합니다.

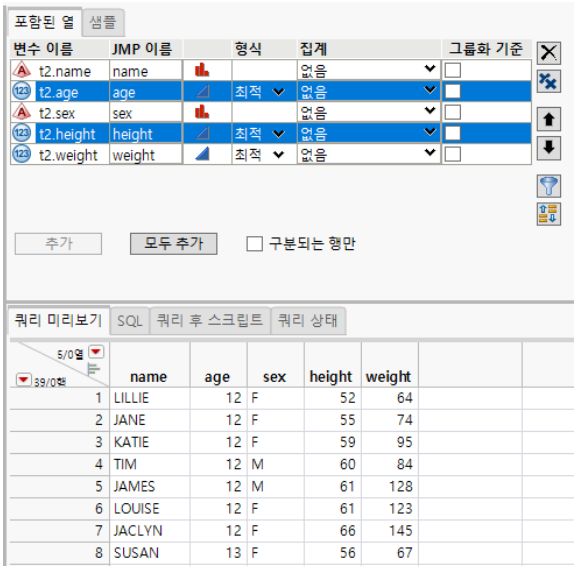
참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
2. " 쿼리할 테이블 선택 " 창에서 **g4_bigclass** 를 기본 테이블로 선택합니다 **.**
3. **쿼리 생성**을 클릭하여 쿼리 빌더 창을 표시합니다 .
4. " 포함된 열 " 탭에서 **모두 추가**를 클릭합니다 .
5. t1.age 및 t1.height 를 선택하고 **선택한 항목을 기준으로 정렬** 을 클릭합니다 .

   오른쪽 열의 " 정렬 기준 " 개요에 해당 열이 나타납니다 .

   열은 먼저 **연령**을 기준으로 (가장 적은 연령부터 가장 많은 연령의 순서) 정렬된 후 **신장**을 기준으로 (가장 작은 키부터 가장 큰 키의 순서) 정렬됩니다 .

그림 3.48 정렬 기준 열 선택



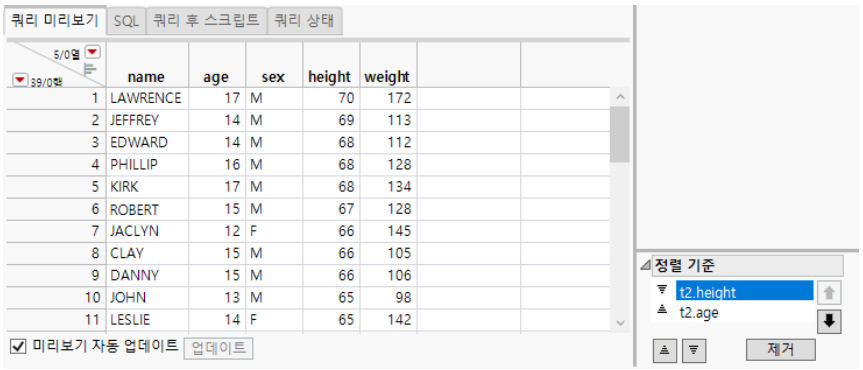
6. " 정렬 기준 " 개요에서 t1.height 를 선택한 후 열 아래에서 **값을 내림차순으로 정렬합니다 .** 를 클릭합니다 .

   height 열이 가장 큰 키부터 가장 작은 키의 순서로 정렬됩니다 .

7. t1.height 를 선택하고 **선택한 항목을 목록에서 위쪽으로 이동** 을 클릭합니다 .

   height 열이 먼저 정렬됩니다 . 그런 다음 height 의 각 수준 내에서 age 열의 값이 정렬됩니다 . height 68 의 경우 age 가 14 부터 17 의 순서로 정렬됩니다 .

그림 3.49 열 재정렬 결과



쿼리 상태 보기

"쿼리 상태" 탭에서는 쿼리가 백그라운드에서 실행되는 동안 쿼리 상태를 볼 수 있습니다. 쿼리 이름, SQL 문 및 처리된 레코드 수가 나타납니다. 언제든지 쿼리를 중지하고 처리된 레코드만

볼 수 있습니다. 다른 JMP 창에서 백그라운드 쿼리를 보려면 **보기 > 실행 중인 쿼리**를 선택합니다. "쿼리 빌더" 환경 설정에서 **가능한 경우 백그라운드에서 쿼리 실행**을 선택 취소한 경우에는 상태 상세 정보를 사용할 수 없습니다.

참고 : 쿼리 빌더에서 SAS 에 연결하는 경우 모든 쿼리 및 쿼리 미리보기는 포그라운드에서 실행됩니다 .

쿼리 후 스크립트 작성

"쿼리 후 스크립트" 탭에서는 쿼리를 실행한 후에 실행되는 JSL 스크립트를 작성할 수 있습니다. 예를 들어 데이터를 가져온 후 분포를 생성하려는 경우가 있습니다.

```
Distribution( Column( :age, :gender ) );
```

이 스크립트는 최종 데이터 테이블에 포함된 소스 스크립트의 일부입니다.

JMP 쿼리 저장 및 실행

쿼리를 .jmpquery 파일로 저장하여 나중에 쿼리를 수정하거나 실행할 수 있습니다. 서버 연결 문자열에 암호가 지정되어 있지 않으면 암호를 입력하라는 메시지가 표시됩니다. JSL 스크립트로 .jmpquery 파일을 열고 실행할 수도 있습니다.

쿼리를 작성한 후 해당 설정을 .jmpquery 파일로 저장하려면 오른쪽 하단의 **저장**을 클릭합니다. **저장**을 다시 클릭하면 파일이 최근 변경 사항으로 덮어써집니다. **다른 이름으로 저장**을 클릭하면 쿼리가 새 .jmpquery 파일에 저장됩니다.

JMP 쿼리 실행

쿼리 파일은 기본적으로 쿼리 빌더에서 열리며, 여기에서 쿼리를 필요한 대로 수정한 후 실행하여 데이터를 가져올 수 있습니다. 쿼리에 만족하면 쿼리를 열 때 쿼리 빌더에서 열리지 않고 바로 실행되도록 구성할 수 있습니다.

쿼리를 작성할 때 왼쪽 상단의 "쿼리 이름" 상자 옆에 있는 빨간색 삼각형 메뉴에서 **열 때 실행**을 선택합니다.

"열 때 실행" 설정을 재정의하려면 파일 시스템에서 파일을 열기 전에 Ctrl 키 (Windows) 또는 Command 키 (macOS) 를 누릅니다. Windows 의 경우, JMP 홈 창에서 쿼리를 마우스 오른쪽 버튼으로 클릭하고 **쿼리 편집**을 선택해도 됩니다.

저장된 쿼리를 템플릿으로 사용

.jmpquery 파일을 새 쿼리의 시작점으로 사용할 수도 있습니다. 이렇게 하면 실수로 **저장**을 클릭하여 원래 쿼리를 덮어쓰는 것을 방지할 수 있습니다. 이 기능은 쿼리를 수정한 후 **다른 이름으로 저장**을 클릭하는 것과 동일합니다.

1. JMP 홈 창의 " 최근 사용한 파일 " 목록에서 .jmpquery 파일을 마우스 오른쪽 버튼으로 클릭합니다 .

2. **템플릿으로 열기**를 선택합니다.
3. 쿼리를 수정하고 **저장**을 클릭하여 새 .jmpquery 파일을 생성합니다.

참고: .jmpquery 파일에는 데이터베이스 로그인 정보가 포함됩니다. 따라서 쿼리를 실행하기 전에 데이터베이스 연결을 설정했어야 합니다. 자세한 내용은 "데이터베이스에 연결"(157 페이지)에서 확인하십시오.

선택한 데이터 열기

SQL 데이터베이스에서 가져올 열과 데이터를 지정한 후에는 **쿼리 실행**을 클릭하여 해당 데이터를 데이터 테이블에서 엽니다. SQL 문은 테이블 변수로 저장됩니다. 다음과 같은 스크립트를 사용할 수 있습니다.

소스 쿼리를 실행합니다.

쿼리 수정 쿼리를 쿼리 빌더에서 엽니다. 쿼리 빌더에서는 가져올 열 및 데이터를 변경하고 쿼리를 보다 세부적으로 사용자 정의할 수 있습니다.

데이터베이스에서 업데이트 데이터베이스에 연결하여 데이터를 새로 고친 후 쿼리를 실행합니다. 데이터 테이블과 데이터베이스 테이블에 동일한 수의 열이 포함된 경우 데이터베이스가 새로 고쳐지면 기존 데이터 테이블의 값이 업데이트됩니다.

또한 새로 고쳐지는 마지막 열 다음에 계산식 열만 있는 경우에도 기존 데이터 테이블이 업데이트됩니다.

그렇지 않으면 업데이트된 데이터가 새 데이터 테이블에 나타납니다.

쿼리 빌더의 빨간색 삼각형 메뉴 옵션

쿼리 빌더의 빨간색 삼각형 메뉴에서는 스크립팅 및 사용자 SQL 옵션을 제공합니다. 수정 및 실행 스크립트는 항상 최종 데이터 테이블에 자동으로 저장됩니다.

수정 스크립트 복사 쿼리를 수정할 수 있도록 스크립트를 컴퓨터의 클립보드에 복사합니다.

실행 스크립트 복사 데이터를 새로 고치고 쿼리를 실행할 수 있도록 스크립트를 컴퓨터의 클립보드에 복사합니다.

수정 스크립트를 스크립트 창에 저장 "쿼리 수정" 스크립트를 스크립트 창에 저장합니다.

실행 스크립트를 스크립트 창에 저장 "데이터베이스에서 업데이트" 스크립트를 스크립트 창에 저장합니다.

열 때 실행 쿼리 파일을 열 때 쿼리를 실행합니다.

사용자 SQL 로 변환 쿼리 문을 새 스크립트 편집기 창에 표시합니다. 이 옵션을 선택하기 전에 필터 확인 기능을 제거해야 합니다.

"사용자 SQL" 창에서 쿼리를 저장하면 사용자 SQL 이 저장됩니다. 쿼리를 사용자 정의하기 전에 있었던 대화식 구성 요소는 저장되지 않습니다. 또한 빨간색 삼각형 메뉴에서 "대화식으로 되돌리기"를 사용할 수 없습니다.

이 쿼리를 JMP 12 와 호환되는 상태로 유지 해당 쿼리에 대한 이후 변경 사항이 JMP 12 와 호환되도록 합니다 . JMP 12 와의 호환성 문제를 일으키는 기능은 쿼리를 편집하는 동안 숨겨집니다 . 이 옵션을 선택 취소하고 JMP 13 이상에서만 지원되는 변경 사항을 적용한 후에는 이 옵션을 더 이상 사용할 수 없습니다 .

JMP " 쿼리 빌더 " 환경 설정에도 동일한 옵션이 포함되어 있으며 이 옵션은 기본적으로 선택 취소되어 있습니다 . JMP 12 와 호환되는 것으로 표시된 쿼리를 JMP 16 에서 열면 환경 설정이 어떻게 설정되어 있든 상관없이 호환성 문제를 발생시키는 기능이 숨겨집니다 .

대화식으로 되돌리기 쿼리 빌더 창에 대화식 쿼리를 표시합니다 . 대화식으로 되돌릴 경우 " 사용자 SQL" 탭에서 적용한 변경 사항은 저장되지 않습니다 .

쿼리 빌더에서 SQL 문 작성

쿼리 빌더를 사용하면 SQL 문을 작성하지 않고 대화식으로 SQL 쿼리를 생성할 수 있습니다. 쿼리 빌더에서 쿼리를 작성한 후 쿼리에 사용자 문을 추가할 수도 있습니다.

참고 : 쿼리 빌더 예는 JMP 와 함께 설치되지 않은 데이터베이스를 기반으로 합니다 .

1. **파일 > 새로 만들기 > 데이터베이스 쿼리**를 선택하고 , 데이터베이스에 연결하고 , **SQBTest** 스키마를 선택합니다 . 자세한 내용은 " 데이터베이스에 연결 "(157 페이지) 에서 확인하십시오 .
2. " 쿼리할 테이블 선택 " 창에서 **g6_Rentals** 를 기본 테이블로 선택합니다 .
3. **g6_Movies** 및 **g6_Customers** 를 보조 테이블로 선택합니다 .
4. **쿼리 생성**을 클릭하여 쿼리 빌더 창을 표시합니다 .
5. " 포함된 열 " 탭에서 **모두 추가** 버튼을 클릭합니다 .
6. " 쿼리 빌더 " 의 빨간색 삼각형을 클릭하고 **사용자 SQL 로 변환**을 선택한 다음 **확인**을 클릭합니다 .

   쿼리 빌더에서 생성된 SQL 이 " 사용자 SQL" 탭에 나타납니다 .
7. 세미콜론 앞을 클릭하고 다음 SQL 문을 입력합니다 .

   ```
   WHERE ( ( ( t2.Gender IN ( 'F' ) ) AND ( (t2.Age >= 20) AND (t2.Age
     <= 50) ) ) )
   ```
8. **쿼리 실행**을 클릭하여 데이터를 JMP 로 가져옵니다 .

   데이터 테이블 스크립트에 사용자 쿼리가 포함됩니다 .

참고 : 빨간색 삼각형 메뉴에서 **대화식으로 되돌리기**를 선택하면 " 사용자 SQL" 탭에서 적용한 변경 사항이 저장됩니다 . 사용자 쿼리를 저장한 후 다시 열면 " 대화식으로 되돌리기 " 를 사용할 수 없습니다 .

SQL 문에 대한 간략한 소개는 "SQL(Structured Query Language): 참조 "(161 페이지) 에서 확인하십시오 .

SAS 에서 데이터 가져오기

SAS 서버에 연결하여 SAS 데이터 집합 직접 처리

JRE(Java Runtime Environment) 요구 사항

- Windows 의 경우 , SAS 에 액세스하려면 컴퓨터에 JRE(Java Runtime Environment) 7 이상이 설치되어 있어야 합니다 . 하지만 JRE 7 을 현재 버전으로 지정할 필요는 없습니다 .
- macOS 의 경우 , SAS Integration 기능을 사용하려면 JRE 7 이상이 설치되어 있어야 합니다 .

참고 : Windows 의 경우 , JRE 비트 수와 JMP 비트 수가 매칭되어야 합니다 . 예를 들어 64 비트 JRE 가 설치된 경우 64 비트 JMP 는 32 비트 SAS 및 64 비트 SAS 모두와 통신합니다 .

SAS 메뉴 옵션

"파일" > "SAS" 메뉴에 다음 옵션이 있습니다.

데이터 불러오기 SAS 서버에 있는 데이터를 찾아서 가져옵니다 .

SAS 쿼리 빌더 SQL 문을 작성하지 않고 SAS 서버에 있는 데이터를 선택하여 가져옵니다 . 자세한 내용은 " 쿼리 빌더에서 SQL 쿼리 작성 "(110 페이지) 에서 확인하십시오 .

SAS 로 데이터 내보내기 JMP 데이터 테이블을 SAS 서버로 내보냅니다 .

SAS 폴더 불러오기 SAS 스토어드 프로세스를 불러와 실행하거나 메타데이터 정의 데이터 테이블을 엽니다 .

SAS 프로그램 새로 만들기 SAS 코드를 작성 및 전송하기 위한 스크립트 창을 엽니다 .

SAS 로 전송 JMP 에서 현재 활성 SAS 서버로 직접 SAS 코드를 전송합니다 .

SAS 로그 창 열기 활성 SAS 서버에 대한 SAS 로그 창을 엽니다 .

SAS 출력 창 열기 활성 SAS 서버에 대한 SAS 출력 창을 엽니다 . 이 창에는 최근 SAS 출력이 표시됩니다 .

서버 연결 SAS 서버에 대한 연결을 관리합니다 .

"JMP 시작하기" 의 "SAS" 페이지에도 SAS 옵션에 대한 바로 가기가 있으며, 이 페이지에는 SAS 도구 모음도 있습니다. "환경 설정" 창 (**파일 > 환경 설정**) 의 "SAS Integration" 페이지에서 SAS Integration 과 관련된 몇 가지 설정을 저장할 수 있습니다. "SAS Integration" 환경 설정 지정에 대한 자세한 내용은 "JMP 환경 설정" 장의 "SAS Integration 의 환경 설정"(663 페이지) 에서 확인하십시오.

SAS 데이터 집합 가져오기

SAS 데이터 집합은 다음과 같은 다양한 SAS 형식 중 하나로 저장됩니다.

- Windows 에서 지원되는 형식은 .sas7bdat 및 .sas7bxat 입니다 .
- macOS 에서는 .sas7bdat 파일의 읽기 및 쓰기가 지원됩니다 .
- Windows 와 macOS 에서 .xpt 파일의 읽기 및 쓰기가 지원됩니다 .

JMP에서 데이터 집합을 열면 파일이 데이터 테이블로 열립니다. JMP에서는 기본적으로 SAS 변수 이름을 열 이름으로 사용합니다. Windows에서 특정 파일의 변수 라벨을 사용하려면 파일을 열 때 해당 옵션을 선택하십시오(아래의 5단계 참조).

.sas7bdat 파일에 지원되는 ISO 날짜 형식은 B8601DA, B8601DN, B8601DT, B8601DZ, B8601LZ, B8601TM, B8601TZ, E8601DA, E8601DN, E8601DT, E8601DZ, E8601LZ, E8601TM, E8601TZ 및 IS8601DA입니다. 지원되지 않는 날짜 형식의 경우에는 데이터를 과학적 표기법 형식으로 가져옵니다.

.xpt 파일에 지원되는 날짜 형식은 BEST, DATE, DATETIME,DAY, MMDDYY, DOWNAME, HHMM,HOUR, TIME,YYMMDD, DDMMYY, MMSS, DTDATE, IS8601DA, E8601DA,PERCENT 및 DOLLAR입니다.

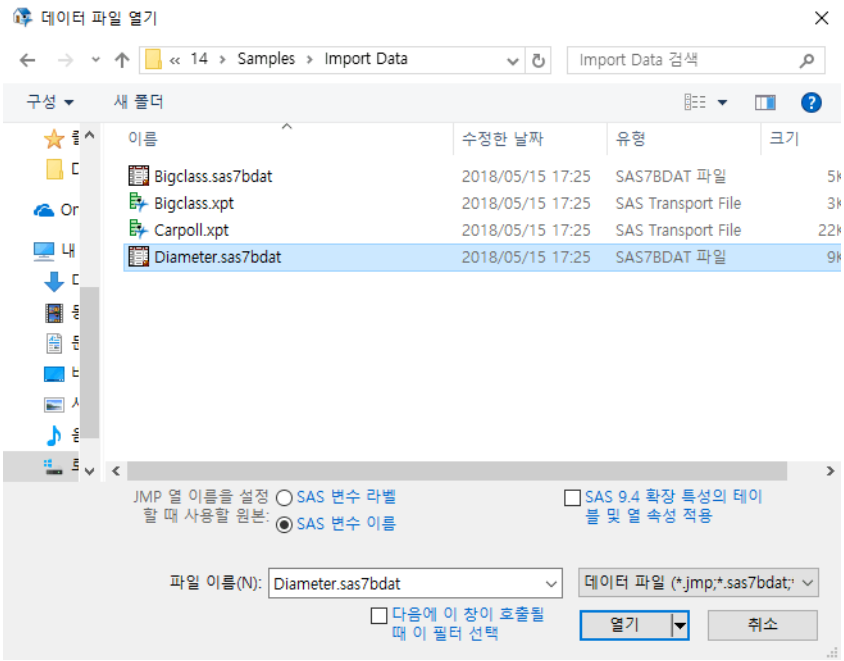
JMP 에서 SAS 데이터 집합을 열려면

1. **파일 > 열기**를 선택합니다 .
2. (Windows 에만 해당) 그림 3.50 에 표시된 것과 같이 **파일 이름** 옆의 목록에서 **SAS 데이터 집합**을 선택합니다 .

   참고 : SAS 변수 이름 및 형식이 유지되며 SAS 데이터 집합을 변경한 후에도 그대로 저장할 수 있습니다 . 자세한 내용은 " 데이터 저장 및 공유 " 장의 "JMP 데이터를 SAS 데이터 집합으로 저장 "(531 페이지) 에서 확인하십시오 .

3. 파일을 선택합니다 .

그림 3.50 SAS 데이터 집합 열기



참고 : 가져올 열을 선택하려면 열기 버튼의 화살표를 클릭하고 **열 선택**을 선택하십시오 .

4. (선택 사항) 다음 옵션 중 하나를 선택합니다 .

   SAS 변수 라벨 JMP 데이터 테이블에서 SAS 변수 이름 대신 SAS 변수 라벨을 열 이름으로 사용합니다 .

   SAS 변수 이름 JMP 데이터 테이블에서 SAS 변수 라벨 대신 SAS 변수 이름을 열 이름으로 사용합니다 .

5. (Windows 의 경우 선택 사항) 다음 옵션 중 하나를 선택합니다 .

   SAS 9.4 확장 속성의 테이블 및 열 특성 적용 SAS 서버가 확장 속성 (SAS 9.4) 을 지원하는 경우 JMP 메타데이터를 저장할 때 확장 속성을 포함합니다 . 이 설정은 "SAS Integration" 페이지의 "SAS 9.4 확장 속성 " 환경 설정보다 우선합니다 .

   다음에 이 창이 호출될 때 이 필터 선택 기본 파일 유형 선택 사항을 **파일 이름** 목록 옆에서 선택하는 옵션으로 설정합니다 . 이 체크박스를 선택할 경우 다음에 이 창을 열 때 기본 파일 유형이 **SAS 데이터 집합**이 됩니다 .

6. (선택 사항) SAS 전송 (.xpt) 파일의 경우 다음 중 하나를 선택합니다 .

   멤버 선택 JMP 에서 열려는 특정 멤버 또는 테이블의 이름을 입력할 수 있습니다 . macOS 의 경우 , **멤버 테이블 > 지정됨**을 선택한 후 이름을 입력합니다 .

모든 멤버 열기 전송 파일의 모든 멤버 또는 테이블을 엽니다 . macOS 의 경우 , **멤버 테이블 > 모두**를 선택합니다 .

모든 멤버 저장 파일을 여는 즉시 JMP 파일로 저장합니다 . SAS 전송 파일을 열었던 디렉터리와 동일한 디렉터리에 파일이 저장됩니다 . macOS 의 경우 , 이 옵션 대신 **모두 저장**이 표시됩니다 .

열 선택 JMP 에서 전송 파일의 특정 열만 열도록 지정합니다 . 표시되는 목록에서 가져올 열을 선택합니다 . macOS 의 경우 , 이 옵션 대신 **열기 전에 열 선택**이 표시됩니다 .

7. **열기**를 클릭합니다 .

참고 : SAS 파일에서 날짜 변수를 가져오는 경우 JMP 는 SAS 날짜 형식을 찾아 JMP 날짜 열로 변환합니다 .

JMP 에서 SAS 의 SAS 전송 파일 생성

JMP에서는 SAS XPORT 엔진을 사용하여 저장된 SAS 전송 파일을 열 수 있습니다. 예를 들어 다음은 test라는 전송 파일을 생성하는 샘플 SAS 코드입니다.

참고 : misc 및 work 는 SAS 라이브러리 참조 이름입니다 .

```
data test;
input name $ age weight;
cards;
Susan 12 72
Melanie 10 68
Jonathan 11 77
Sheila 13 67
;
libname misc xport 'C:/test.xpt';
proc copy in=work out=misc;
run;
```

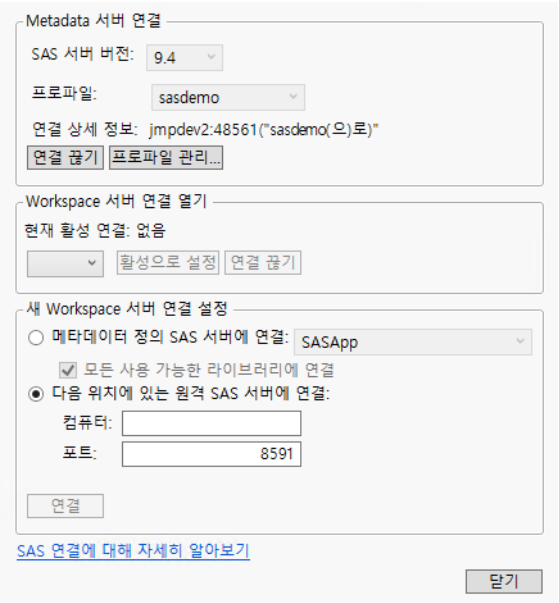
JMP 에서 SAS 에 연결

SAS Metadata 서버에 연결하거나 SAS Workspace 서버에 직접 연결할 수 있습니다. SAS Metadata 서버에 연결된 후에는 SAS 서버, 라이브러리 및 데이터 집합을 탐색할 수 있습니다.

참고 : JMP "SAS Integration" 환경 설정의 기본 설정은 SAS 서버 버전 9.4 이지만 필요한 경우 버전을 9.3 으로 변경할 수 있습니다 .

먼저, **파일 > SAS > 서버 연결**을 선택하십시오. 그러면 그림 3.51에 표시된 것과 같은 "SAS 서버 연결" 창이 나타납니다. 모든 연결은 이 창에서 설정됩니다.

그림 3.51 SAS 서버 연결



다음 섹션에서는 SAS 서버에 연결하는 방법을 설명합니다.

- "JMP 에서 SAS Metadata 서버에 연결 "(136 페이지)
- "JMP 에서 원격 SAS Workspace 서버에 연결 "(139 페이지)
- "JMP 에서 SAS 환경에 연결 (Windows 에만 해당)"(140 페이지)
- "JMP 에서 로컬 컴퓨터의 SAS 에 연결 (Windows 에만 해당)"(140 페이지)

JMP 에서 SAS Metadata 서버에 연결

참고 : 한 번에 하나의 Metadata 서버에만 연결할 수 있습니다 . 두 번째 연결을 설정할 경우 첫 번째 연결은 끊어집니다 .

Metadata 서버에 연결하려면

1. **파일 > SAS > 서버 연결**을 선택합니다 . 그러면 그림 3.51(136 페이지) 에 표시된 것과 같은 "SAS 서버 연결 " 창이 나타납니다 .
2. SAS 서버의 버전을 선택합니다 . 이 정보는 SAS Metadata 서버 관리자에게 문의하십시오 .

   JMP "SAS Integration" 환경 설정에 따라 기본적으로 SAS 서버 버전 9.4가 선택되어 있습니다.
3. 사용하려는 프로파일을 선택합니다 .

   설정된 프로파일이 없는 경우 "SAS Metadata 서버 프로파일을 생성 또는 수정하려면 "(138 페이지) 을 참조하십시오 .

4. **연결**을 클릭합니다 .

   JMP 에서 연결을 설정할 수 없으면 오류 메시지가 나타납니다 . 일반적으로 사용자 이름 또는 암호가 올바르지 않으면 이러한 오류가 발생합니다 . 프로파일의 정보를 업데이트해야 하는 경우에는 "SAS Metadata 서버 프로파일을 생성 또는 수정하려면 "(138 페이지) 을 참조하십시오 .

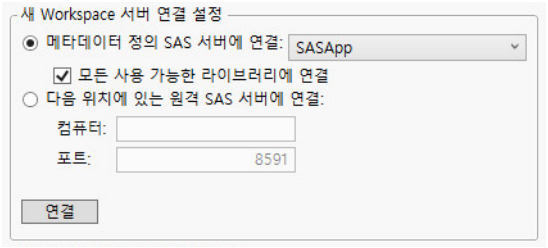
5. **닫기**를 클릭합니다 .

SAS Metadata 서버에 연결된 후 Metadata 서버에서 제공하는 모든 SAS Workspace 서버에 연결할 수 있습니다 .

SAS Workspace 서버에 연결하려면 (Windows 에만 해당)

1. **파일 > SAS > 서버 연결**을 선택합니다 . 그러면 그림 3.51(136 페이지) 에 표시된 것과 같은 "SAS 서버 연결 " 창이 나타납니다 .
2. 연결할 Workspace 서버를 선택합니다 .

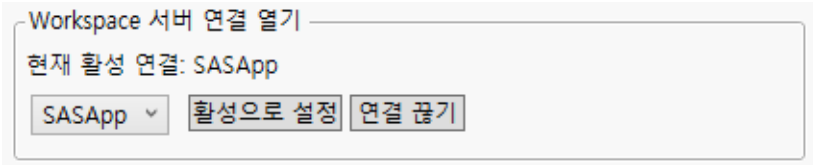
그림 3.52 Workspace 서버에 대한 연결 열기



3. **연결**을 클릭합니다 .

   "Workspace 서버 연결 열기 " 에 Workspace 서버가 현재 활성 연결로 표시되어 있습니다 .

그림 3.53 현재 활성 연결



4. **닫기**를 클릭합니다 .

활성 연결을 변경하려면

참고 : 활성 연결은 SAS 코드를 전송하거나 SAS 스크립트 명령을 처리하는 데 사용되는 연결입니다 .

활성 연결을 변경하려면 먼저 둘 이상의 서버에 연결되어 있어야 합니다. "SAS Workspace 서버에 연결하려면 (Windows 에만 해당)"(137 페이지) 에 설명된 지침을 따라 서버 연결을 둘 이상 추가하십시오.

1. "Workspace 서버 연결 열기 " 섹션에서 드롭다운 메뉴를 클릭하고 원하는 서버를 선택합니다.
2. **활성으로 설정**을 클릭합니다 .
3. **닫기**를 클릭합니다 .

팁 : 언제든지 활성 서버를 변경할 수 있습니다 .

SAS Workspace 서버와의 연결을 끊으려면

1. "SAS 서버 연결 " 창의 "Workspace 서버 연결 열기 " 에서 연결을 끊을 Workspace 서버를 선택합니다 .
2. **연결 끊기**를 클릭합니다 .

SAS Metadata 서버와의 연결을 끊으려면

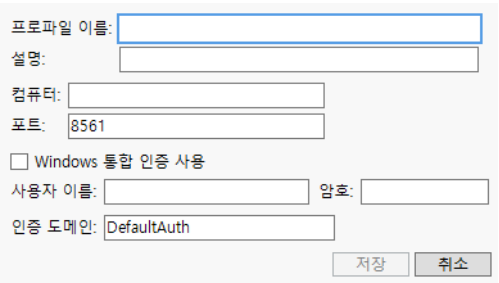
1. "SAS 서버 연결 " 창에서 연결을 끊을 Metadata 서버를 선택합니다 .
2. **연결 끊기**를 클릭합니다 .

SAS Metadata 서버 프로파일을 생성 또는 수정하려면

1. "SAS 서버 연결 " 창에서 **SAS 서버 버전**을 선택합니다 .
2. **프로파일 관리**를 클릭합니다 .
3. 새 프로파일을 추가하려면 **추가**를 클릭하고 , 프로파일의 설정을 변경하려면 **수정**을 클릭합니다 .

   " 프로파일 생성 " 또는 " 프로파일 수정 " 창이 나타납니다 . 새 프로파일을 추가하는 경우 DefaultAuth 가 포함되어 있는 " 인증 도메인 " 필드와 " 포트 " 필드를 제외하고 모든 필드가 비어 있습니다 . 프로파일을 수정하는 경우에는 필드에 현재 정보가 포함됩니다 .

그림 3.54 Metadata 서버 프로파일 생성 또는 수정



4. SAS Metadata 서버에 연결하는 데 필요한 정보를 채웁니다 . 이 정보는 SAS Metadata 서버 관리자에게 문의하십시오 .

   프로파일 이름 이 프로파일의 이름을 선택합니다 . 이 이름은 프로파일 목록에 표시됩니다 .

   설명 (선택 사항) 이 프로파일에 대한 짧은 설명을 입력할 수 있습니다 .

   컴퓨터 Metadata 서버를 호스팅하는 컴퓨터의 이름입니다 . 예 : myserver.mycompany.com

   포트 컴퓨터에 연결하는 데 사용할 포트입니다 . 예 : 8561

   Windows 통합 인증 사용 Windows 로그인 ID 와 암호를 사용하여 서버에 액세스하려면 이 옵션을 선택합니다 . 이 옵션이 활성화되면 " 사용자 이름 " 및 " 암호 " 필드는 비활성화됩니다 . 이 옵션은 기본적으로 비활성화되어 있습니다 .

   사용자 이름 Metadata 서버의 사용자 이름입니다 .

   암호 사용자의 암호입니다 . 항상 별표로 표시됩니다 .

   인증 도메인 사용자가 속해 있는 도메인입니다 .

5. **저장**을 클릭합니다 .

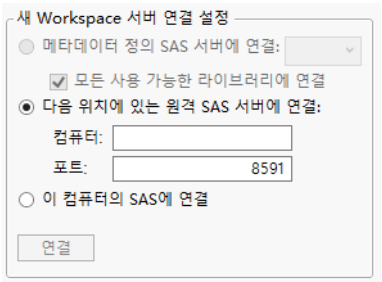
JMP 에서 원격 SAS Workspace 서버에 연결

Metadata 서버를 거치지 않고 SAS Workspace 서버에 직접 연결할 수도 있습니다.

원격 SAS Workspace 서버에 연결하려면

1. **파일 > SAS > 서버 연결**을 선택합니다 . 그러면 그림 3.51(136 페이지) 에 표시된 것과 같은 "SAS 서버 연결 " 창이 나타납니다 .
2. " 새 Workspace 서버 연결 설정 " 에서 **다음 위치에 있는 원격 SAS 서버에 연결**을 선택합니다 .

그림 3.55 원격 SAS 서버에 대한 연결 열기



3. 컴퓨터 이름과 포트 번호를 입력합니다 . 이 정보는 SAS 서버 관리자에게 문의하십시오 .
4. **연결**을 클릭합니다 .
5. 나타나는 창에 사용자 이름 및 암호를 입력합니다 .
6. **확인**을 클릭합니다 .

7. "SAS 서버 연결 " 창에서 **닫기**를 클릭합니다 .

원격 SAS Workspace 서버와의 연결을 끊으려면

1. "SAS 서버 연결 " 창의 "Workspace 서버 연결 열기 " 에서 연결을 끊을 서버를 선택합니다 .
2. **연결 끊기**를 클릭합니다 .

JMP 에서 SAS 환경에 연결 (Windows 에만 해당)

Windows 의 경우, JMP 환경 설정에 SAS 서버 버전 9.3 또는 9.4 가 선택되어 있고 컴퓨터 또는 JMP 가 올바르게 구성되어 있으면 SAS Mid-tier(또는 SAS 환경) 에 연결할 수 있습니다. JMP "SAS Integration" 환경 설정의 기본 설정은 SAS 서버 버전 9.4 입니다.

SAS 설치 관리자가 컴퓨터에서 SAS 환경 정의 파일을 찾을 수 있도록 설정했어야 합니다. 그렇지 않은 경우에는 JMP 환경 설정에서 해당 파일의 경로를 입력할 수 있습니다.

JMP 환경 설정을 구성하려면

1. **파일 > 환경 설정 > SAS Integration** 을 선택합니다 .
2. **SAS 환경에 연결**을 선택한 후 **구성**을 클릭합니다 .
3. JMP 에서 이미 감지된 환경에 연결하려면 **자동 검색**을 클릭한 후 목록에서 URL 을 선택합니다 (필요한 경우).
4. SAS 환경 정의 파일의 경로를 입력하려면 **수동 구성**을 클릭하고 URL 을 입력합니다 .
5. **확인**을 클릭합니다 .

SAS 환경에 연결하려면

1. **파일 > SAS > 서버 연결**을 선택하여 "SAS 서버 연결 " 창을 엽니다 .
2. "Metadata 서버 연결 " 영역에서 **SAS 환경에 연결**을 선택합니다 .

   이 옵션을 사용할 수 없으면 컴퓨터 또는 JMP 가 SAS 환경을 찾을 수 있도록 구성되지 않은 것입니다 . 자세한 내용은 "JMP 환경 설정을 구성하려면 "(140 페이지) 에서 확인하십시오 .
3. 필요한 경우 " 환경 " 목록에서 환경 이름을 선택합니다 .
4. **연결**을 클릭합니다 .
5. 사용자 이름 및 암호를 입력하라는 메시지가 표시되면 해당 사항을 입력합니다 .

JMP 에서 로컬 컴퓨터의 SAS 에 연결 (Windows 에만 해당)

로컬 컴퓨터의 SAS에 직접 연결할 수도 있습니다.

컴퓨터의 SAS 에 연결하려면

1. **파일 > SAS > 서버 연결**을 선택하여 "SAS 서버 연결 " 창을 엽니다 .
2. " 새 연결 설정 " 에서 **이 컴퓨터의 SAS 에 연결**을 선택합니다 .

컴퓨터에 SAS 가 설치되어 있지 않으면 이 옵션을 사용할 수 없습니다 .

3. **연결**을 클릭합니다 .
4. "SAS 서버 연결 " 창에서 **닫기**를 클릭합니다 .

컴퓨터에 있는 SAS 와의 연결을 끊으려면

1. "SAS 서버 연결 " 창에서 " 열려 있는 연결 " 아래의 **로컬**을 선택합니다 .
2. **연결 끊기**를 클릭합니다 .

SAS 쿼리 빌더를 사용하여 SAS 데이터 집합 열기

SAS 서버에서 데이터를 선택하여 가져오려면 SAS 쿼리 빌더를 사용하는 것이 좋습니다. 데이터를 데이터 테이블로 가져오기 전에 미리 볼 수 있습니다. 쿼리를 저장했다가 나중에 수정하여 실행하거나 JSL 스크립트에서 참조할 수도 있습니다.

"파일" > "SAS" > " 데이터 불러오기 " 기능을 사용하여 SAS 데이터 집합을 여는 방법 대신 SAS 쿼리 빌더를 사용할 수 있습니다.

SAS 쿼리 빌더를 사용하여 SAS 데이터 집합을 열려면 다음 단계를 수행하십시오 .

1. **파일 > SAS > SAS 쿼리 빌더**를 선택합니다 .
2. "SAS 서버에 연결 " 창에서 Metadata 서버 또는 원격 서버를 선택합니다 .
   - JMP 에 이미 설정된 Metadata 서버에 연결하려면 **메타데이터 정의 SAS 서버에 연결** 목록에서 서버를 선택합니다 .

     모든 사용 가능한 라이브러리에 연결은 "SAS Integration" 환경 설정에 따라 기본적으로 선택되어 있습니다 . 이 옵션은 모든 JMP 세션에서 자동으로 메타데이터 정의 SAS 라이브러리를 연결합니다 . Workspace 서버에 포함된 메타데이터 정의 SAS 라이브러리가 매우 많은 경우 서버에 빠르게 연결하려면 이 옵션을 선택 취소하는 것이 좋습니다 .
   - Metadata 서버를 추가하거나 구성하려면 **프로파일 관리**를 클릭하고 "SAS Metadata 서버 프로파일을 생성 또는 수정하려면 "(138 페이지) 에 설명된 단계를 따릅니다 .
   - 원격 서버에 연결하려면 해당 컴퓨터 이름과 포트 번호를 입력합니다 . 이 정보는 SAS Metadata 서버 관리자에게 문의하십시오 .
3. **확인**을 클릭합니다 .

   "SAS 쿼리 빌더 " 창이 나타납니다 .

쿼리 빌더를 사용하는 방법에 대한 자세한 내용은 " 쿼리 빌더에서 SQL 쿼리 작성 "(110 페이지) 에서 확인하십시오 .

참고 :

- 모든 쿼리 빌더 쿼리는 포그라운드에서 실행됩니다 .
- 확장 속성은 기본적으로 가져오지 않습니다 . 이를 가져오려면 JMP "SAS Integration" 환경 설정을 수정하십시오 . **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택

합니다 . **SAS Integration** 을 선택한 후 **가져올 때 확장 속성의 테이블 및 열 특성 적용**을 선택합니다 .

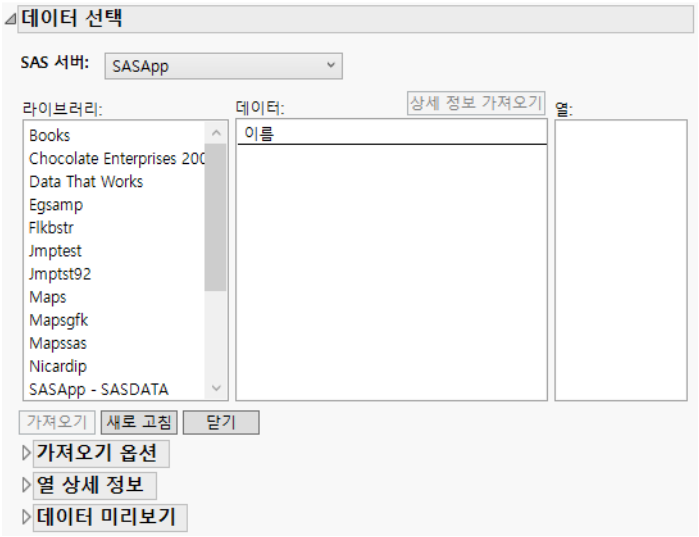
- SAS 쿼리 빌더는 로컬 서버 연결을 지원하지 않습니다 .

SAS 서버를 통해 SAS 데이터 집합 열기

SAS Workspace 서버에 연결한 후에는 해당 서버에서 SAS 라이브러리를 탐색하고 JMP로 데이터를 가져올 수 있습니다.

SAS 서버의 데이터 집합을 불러오려면 **파일 > SAS > 데이터 불러오기**를 선택하십시오. 그러면 "SAS 데이터 불러오기" 창이 나타납니다.

그림 3.56 SAS 데이터 불러오기



이 창은 SAS Metadata 서버(연결된 경우)가 제공하는 서버 목록으로 미리 채워져 있습니다. 모든 물리적 연결 및 로컬 연결도 표시됩니다(그림 3.55(139 페이지) 참조).

- 서버에 포함된 라이브러리의 목록을 보려면 서버를 선택합니다 .
- 라이브러리 내에 있는 데이터 집합의 목록을 보려면 해당 라이브러리를 선택합니다 .
- 데이터 집합 내에 있는 열의 목록을 보려면 해당 데이터 집합을 선택합니다 .

"SAS 데이터 불러오기" 창을 닫았다가 다시 열면 이전에 본 라이브러리와 데이터 집합이 창에 나타납니다. 하지만 언제든지 "SAS 서버" 목록에서 다른 서버를 선택한 후 라이브러리 및 데이터 집합을 선택할 수 있습니다.

팁 : 서버를 사용할 수 없거나 연결에 실패한 경우에는 서버 이름이 연한 색 기울임꼴 텍스트로 표시됩니다 . 연결을 다시 설정해 보려면 해당 서버를 클릭하십시오 .

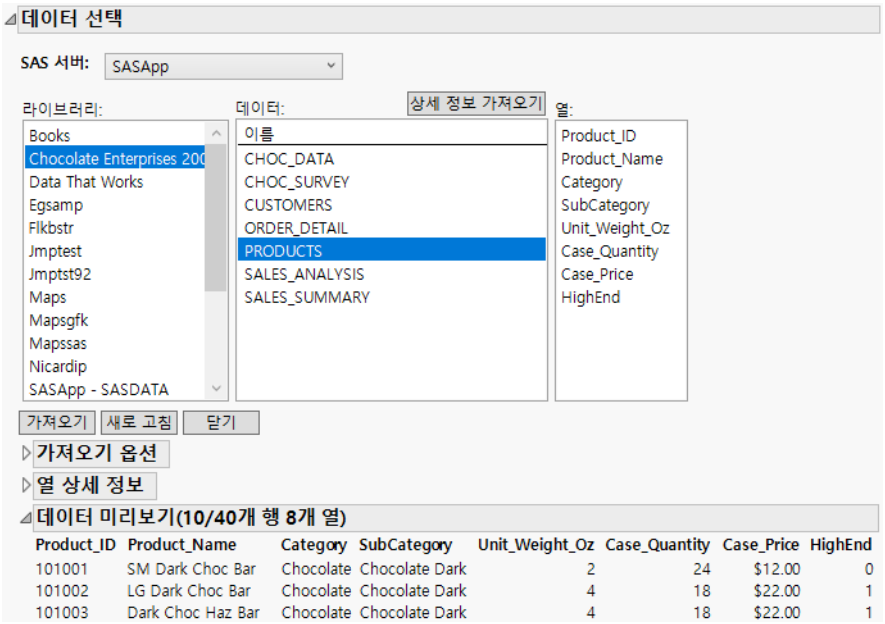
SAS 데이터 정보 불러오기

SAS 데이터 집합을 선택한 후 "상세 정보 가져오기", "열 상세 정보" 및 "데이터 미리보기" 옵션을 사용하여 데이터 집합을 열기 전에 해당 내용에 대한 정보를 확인할 수 있습니다.

데이터 미리보기

데이터 집합을 선택하면 해당 데이터 집합에 있는 처음 10 개의 행 및 열이 "데이터 미리보기" 개요에 표시됩니다.

그림 3.57 데이터 미리보기



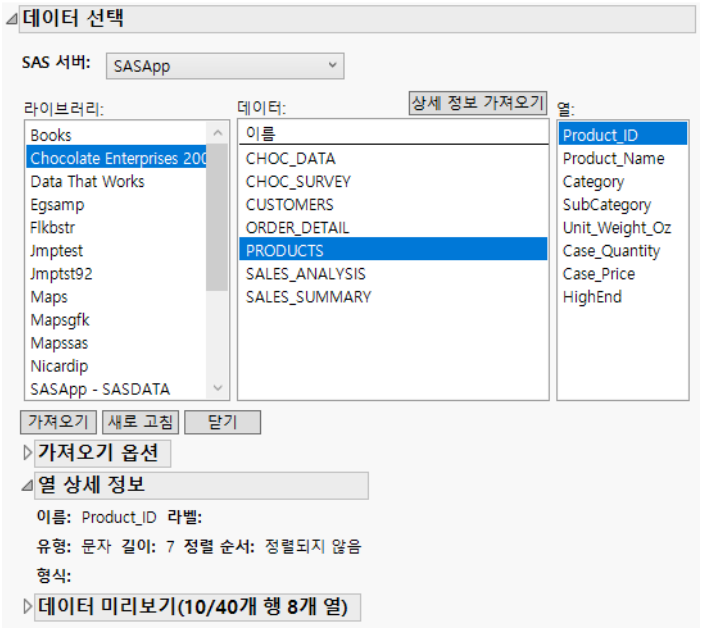
데이터 집합 상세 정보

라이브러리에 있는 각 데이터 집합의 크기 및 마지막 수정 날짜를 확인하려면 "SAS 데이터 불러오기" 창에서 **상세 정보 가져오기**를 클릭하십시오. 이 옵션을 사용하면 컴퓨터에서 전체 데이터 집합을 처리할 수 있는지 여부를 추정하는 데 유용합니다.

열 상세 정보

데이터 집합의 특정 열에 대한 정보를 보려면 해당 열을 선택하십시오. "열 상세 정보" 개요에는 데이터 열에 대한 몇 가지 기본적인 정보가 표시됩니다.

그림 3.58 열 상세 정보



이름 SAS 데이터 집합의 열 이름입니다 .

라벨 열을 설명하는 라벨입니다 . 라벨은 이름보다 길 수 있으며 대개 열 이름의 의미를 확인하는 데 유용합니다 .

유형 열의 데이터 유형이 문자인지 숫자인지를 지정합니다 .

길이 열 데이터의 바이트 길이입니다 .

정렬 순서 SAS 에서 열이 정렬된 방식입니다 .

형식 SAS 열의 형식 (예 : 달러) 입니다 . 이 형식 필드에는 형식이 지정된 값의 너비와 소수 자릿수에 대한 정보도 포함됩니다 .

SAS 데이터 집합 열기

SAS 데이터 집합을 JMP 로 직접 가져올 수 있습니다.

1. "SAS 데이터 불러오기 " 창에서 데이터 집합을 선택합니다 .

   기본적으로 **모든 행**을 가져오도록 지정되어 있습니다 .

2. **가져오기**를 클릭합니다 .

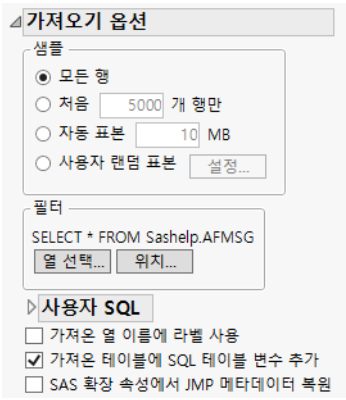
전체 SAS 데이터 집합이 JMP 데이터 테이블로 가져와집니다. SAS 데이터를 가져온 후 JMP 는 SAS 형식과 가장 매칭되는 형식을 찾습니다.

샘플 SAS 데이터 집합 가져오기

샘플 SAS 데이터 집합을 JMP 로 직접 가져올 수 있습니다.

1. "SAS 데이터 불러오기 " 창에서 데이터 집합을 선택합니다 .
2. " 가져오기 옵션 " 개요를 엽니다 .

그림 3.59 가져오기 옵션



3. 데이터 집합의 일부만 가져오려면 다음 중 원하는 작업을 수행합니다 .
   - 처음 *x* 개의 행만 선택합니다 . 자세한 내용은 " 처음 x 개의 행만 가져오려면 "(145 페이지) 에서 확인하십시오 .
   - 지정한 파일 크기로 자동 표집하도록 선택합니다 . 자세한 내용은 " 지정된 크기의 자동 표본 파일을 가져오려면 "(146 페이지) 에서 확인하십시오 .
   - 열의 부분집합을 선택합니다 . 자세한 내용은 " 열 부분집합을 선택하려면 "(146 페이지) 에서 확인하십시오 .
   - WHERE 절을 구성하여 데이터를 필터링합니다 . 자세한 내용은 "WHERE 절을 사용하여 가져오려면 "(146 페이지) 에서 확인하십시오 .
   - 사용자 데이터 표본을 추출합니다 . 자세한 내용은 " 데이터의 랜덤 표본 가져오기 "(147 페이지) 에서 확인하십시오 .

다음 옵션도 사용할 수 있습니다.

가져온 열 이름에 라벨 사용 SAS 라벨을 열 이름으로 변환합니다 .

가져온 테이블에 SQL 테이블 변수 추가 SQL 테이블 변수를 데이터 테이블 변수로 저장합니다 .

SAS 확장 속성에서 JMP 메타데이터 복원 SAS 데이터 집합에 확장 속성으로 저장된 JMP 메타데이터 (테이블 및 열 특성) 를 포함합니다 .

처음 x 개의 행만 가져오려면

1. " 가져오기 옵션 " 섹션에서 **처음 x 개 행만**을 선택하고 가져올 행 수를 지정합니다 .

2. "SAS 데이터 불러오기 " 창에서 **가져오기**를 클릭합니다 .
   JMP 가 지정된 행 수만큼 가져옵니다 .

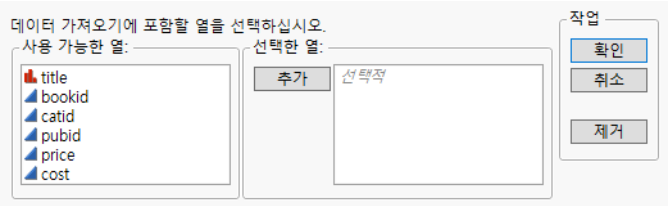
지정된 크기의 자동 표본 파일을 가져오려면

1. " 가져오기 옵션 " 섹션에서 **자동 표본**을 선택하고 가져올 MB 수를 지정합니다 .
2. "SAS 데이터 불러오기 " 창에서 **가져오기**를 클릭합니다 .
   JMP 가 지정된 MB 수만큼 가져옵니다 .

열 부분집합을 선택하려면

1. " 가져오기 옵션 " 섹션에서 **열 선택**을 클릭합니다 .
   " 열 선택 " 창이 나타납니다 .

그림 3.60 열 선택



2. 가져올 열을 선택합니다 .
   한 번에 두 개 이상의 열을 선택하려면 Ctrl 키를 누른 채 각 열을 클릭합니다 .
3. **추가**를 클릭합니다 .
4. 원하는 열을 모두 추가했으면 **확인**을 클릭합니다 .
5. "SAS 데이터 불러오기 " 창에서 **가져오기**를 클릭합니다 .

SAS 데이터 집합에서 선택한 열만 JMP 데이터 테이블로 가져와집니다.

WHERE 절을 사용하여 가져오려면

1. **Where** 를 클릭합니다 .
2. WHERE 절 편집기를 사용하여 WHERE 절을 생성합니다 .
3. **확인**을 클릭하여 "SAS 데이터 불러오기 " 창으로 돌아갑니다 .
4. **가져오기**를 클릭합니다 .

WHERE 절과 매칭되는 데이터만 JMP 데이터 테이블로 가져와집니다.

WHERE 절 생성과 WHERE 절 편집기 사용에 대한 자세한 내용은 "WHERE 절 편집기 사용"(165 페이지)에서 확인하십시오.

참고 : WHERE 절과 표집을 모두 사용하여 데이터를 가져올 경우 WHERE 절이 먼저 적용된 후에 필터링된 데이터의 표본이 추출됩니다 .

사용자가 직접 SQL 문을 작성할 수도 있습니다.

사용자 SQL 문을 사용하여 가져오려면

사용자 SQL 문을 사용하여 SAS 데이터 집합을 열 수도 있습니다.

1. " 가져오기 옵션 " 개요에서 " 사용자 SQL" 개요를 엽니다 .

그림 3.61 사용자 SQL

2. 창에 SQL 문을 입력합니다 .
3. **사용자 SQL 실행**을 클릭합니다 .

참고 : SQL 은 선택한 서버에서 실행되지만 선택한 라이브러리 또는 데이터 집합으로 제한되지는 않습니다 .

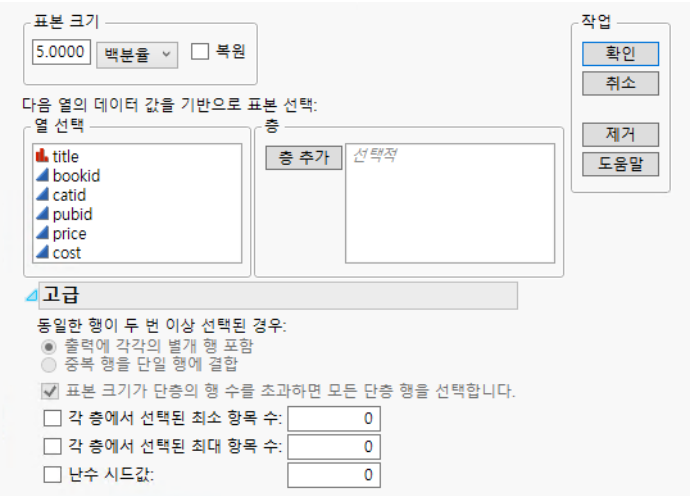
데이터의 랜덤 표본 가져오기

SAS 데이터 집합의 랜덤 행 표본을 가져올 수도 있습니다.

참고 : 표집 기능을 사용하려면 SAS 서버에 SAS/STAT 제품이 라이선스에 따라 설치되어 있어야 합니다 . SAS/STAT 이 없는 경우에는 표집 기능이 비활성화됩니다 .

"가져오기 옵션" 개요의 "가져온 데이터 표집" 영역에서 **사용자 랜덤 표본** 체크박스를 선택합니다. 기본적으로는 5% 의 행을 가져옵니다. 랜덤 표본 가져오기 설정을 변경하려면 **설정** 버튼을 클릭합니다.

그림 3.62 표집 설정



이 창에서는 다음 사항을 지정합니다.

표본 크기 백분율 또는 행 수로 표본 크기를 설정할 수 있습니다. 각 행이 한 번씩만 표집되도록 하려면 **복원** 옵션을 선택 취소합니다. 임의의 행이 두 번 이상 표집되어 가져온 데이터에 나타날 수 있도록 하려면 이 옵션을 선택합니다.

열별 선택 열을 " 층 " 목록으로 이동하여 층을 선택할 수 있습니다.

여러 행 표집 처리 **복원**이 선택된 경우 중복된 각 행을 개별 행으로 추가하거나 중복된 모든 행을 하나의 행에 결합하도록 지정할 수 있습니다. 두 번째 옵션을 선택하면 각 행의 표집 횟수가 포함된 열이 테이블에 추가됩니다.

선택할 최소 및 최대 항목 수 설정 이 옵션을 선택하고 수를 입력합니다.

난수 시드값 설정 이 옵션을 선택하고 시드값을 입력합니다. 시드값을 지정하면 정확히 동일한 표본을 여러 번 재현할 수 있습니다.

참고 : WHERE 절과 표집을 모두 사용하여 데이터를 가져올 경우 WHERE 절이 먼저 적용된 후에 필터링된 데이터의 표본이 추출됩니다.

가져오기 옵션

SAS 데이터를 JMP 로 가져오는 방식을 지정하는 데 사용할 수 있는 추가 옵션이 있습니다.

가져온 열 이름에 라벨 사용 이 옵션을 선택하면 길이에 제한이 있고 의미를 파악하기 어려울 수 있는 열 이름이 열 라벨로 전환됩니다. 이 옵션은 기본적으로 해제되어 있습니다. SAS 데이터 열 이름을 JMP 의 열 이름으로 사용하려면 이 상자를 선택 취소합니다.

가져온 테이블에 SQL 테이블 변수 추가 이 옵션을 선택하면 SQL 쿼리가 데이터 테이블 패널에 변수로 추가됩니다. 이 옵션은 기본적으로 설정되어 있습니다. 이 옵션을 해제하면 데이터 테이블을 가져올 때 두 개의 변수, 즉 SAS 서버 변수와 데이터 집합 변수만 추가됩니다.

SAS 확장 속성에서 JMP 메타데이터 복원

팁 : 데이터가 암호로 보호된 경우 SQL 에 암호가 표시될 수 있으므로 이 옵션을 해제할 수 있습니다 .

테이블 변수

JMP 데이터 테이블을 가져온 후에는 데이터 테이블의 왼쪽 상단 패널에 테이블 변수가 나타납니다. 이러한 변수는 SAS 서버, 데이터 집합, 그리고 SQL 쿼리 및 표집 설정(해당되는 경우)을 보여 줍니다. 언제든지 가져오기를 다시 실행할 수 있도록 소스 스크립트도 추가됩니다.

암호로 보호된 데이터 집합 열기

JMP에서는 암호로 보호된 SAS 버전 7 이상의 데이터 집합을 열 수 있습니다. 암호는 대/소문자가 구분되지 않습니다.

암호로 보호된 데이터 집합을 열려면

1. **파일 > 열기**를 선택합니다 .
2. **파일 유형** 목록에서 **SAS 데이터 집합**을 선택합니다 .
3. 파일을 선택합니다 .
4. **열기**를 클릭합니다 .
5. 암호를 입력한 후 **확인**을 클릭합니다 .

   암호가 올바르지 않으면 올바른 암호를 입력할 때까지 암호를 다시 입력하라는 메시지가 표시됩니다 .

SAS 확장 속성을 사용하여 메타데이터 가져오기

SAS 확장 속성은 테이블 스크립트, 라벨, 길이 및 유형 같은 정보를 가져오기 위해 SAS 코드에서 정의하는 메타데이터입니다. 확장 속성을 데이터 집합 또는 변수와 연결하고 _JMP_TABLESCRIPTNAME_2="OnOpen" 과 같이 이름-값 쌍으로 정의해야 합니다. 자세한 내용은 http://sas.com/ 에서 확장 속성에 대한 SAS 설명서를 참조하십시오.

JMP에서 SAS로 내보낼 때는 열 특성 및 테이블 스크립트 같은 항목이 자동으로 유지됩니다. SAS 코드를 JMP로 가져올 때만 SAS 확장 속성을 JMP 속성에 매핑합니다.

다음은 SAS 코드에서 확장 속성을 정의하는 예입니다. 속성은 테이블 스크립트 이름과 테이블 스크립트 자체를 정의합니다.

```
/* specify two table scripts. */
%LET _DS_ATTRIBUTES=_JMP_TABLESCRIPTCOUNT=2
/* define the table script names */
_JMP_TABLESCRIPTNAME_1="Favorite Movie By State"
_JMP_TABLESCRIPTNAME_2="OnOpen"
```

```
_JMP_TABLESCRIPTVALUE_2="CurrentDataTable() << RunScript(""Favorite Movie By
   State"");"
/* define the OnOpen table script */
_JMP_TABLESCRIPTVALUE_1=
/* a portion of the OnOpen table script */
"
   Current Data Table();
   New Column(""max_name"",
      Character,
      Width(128)
   );
```

다음 예에서는 PROC DATASETS 를 확장 속성과 결합하는 방법을 보여 줍니다.

```
%LET _DS_ATTRIBUTES = attr-name=attr-value...attr-name=attr-value;
%LET _VAR_ATTRIBUTES= var-name (attr-name=attr-value...)...var-name
   (attr-name=attr-value...);

PROC DATASETS NOLIST LIB=WORK;
  MODIFY AttributeReferenceTable;
  XATTR OPTIONS MAXCHUNK=100;
  XATTR SET DS &_DS_ATTRIBUTES;
  XATTR SET VAR &_VAR_ATTRIBUTES;
RUN;
QUIT;
```

SAS 확장 속성과 해당하는 JMP 속성에 대한 자세한 내용은 "JMP 속성을 SAS 확장 속성에 매핑하기 위한 규칙"(730 페이지)에서 확인하십시오.

두 개의 "SAS Integration" 환경 설정에 따라 확장 속성을 가져올지 내보낼지가 결정됩니다. 확장 속성을 내보내려면 **내보낼 때 테이블 및 열 특성을 확장 속성에 저장**을 선택합니다. 또는 **가져올 때 확장 속성의 테이블 및 열 특성 적용**을 선택합니다. 이러한 옵션은 기본적으로 선택 취소되어 있습니다.

확장 속성을 내보내는 예를 보려면 다음 스크립트를 실행하십시오.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
sd = dt << Make SAS Data Step Window( SaveJMPMetadata( 1 ) );
```

스토어드 프로세스 실행

스토어드 프로세스는 연결된 SAS 서버에 저장된 SAS DATA 스텝 코드입니다. JMP 에서 스토어드 프로세서를 실행하고 스크립트 결과를 볼 수 있습니다.

참고 : 오류 메시지는 SAS 에 대해 설정한 환경 설정에 따라 JMP 로그나 별도의 SAS 로그 창으로 전송됩니다 .

스토어드 프로세스를 보고 실행하려면 Metadata 서버에 연결되어 있어야 합니다. Metadata 서버에 연결되지 않은 상태에서 **파일 > SAS > SAS 폴더 불러오기**를 선택하면 연결을 구성하거나 작업을 취소하라는 메시지가 표시됩니다.

스토어드 프로세스를 선택 및 실행하려면

1. **파일 > SAS > SAS 폴더 불러오기**를 선택합니다 .

   "SAS 폴더 불러오기 " 창이 나타납니다 .
2. 스토어드 프로세스를 탐색하여 실행할 스토어드 프로세스를 찾습니다 .
3. 해당 항목을 선택하고 **실행**을 클릭합니다 .

   데이터가 JMP 데이터 테이블로 열립니다 .

Windows의 경우, 스토어드 프로세스를 마우스 오른쪽 버튼으로 클릭하고 **메타데이터 경로 복사**를 선택할 수도 있습니다. 이 옵션은 해당 경로를 클립보드에 복사합니다. 그런 다음 해당 경로를 스크립트 창에 붙여넣어 JSL 연산자 `Meta Get Stored Process()`의 파라미터로 포함할 수 있습니다. 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

참고 : 스트리밍 출력이 선택된 경우에는 SAS 스토어드 프로세스에서 반환된 결과에 정적 그래프가 나타나지 않을 수 있습니다 .

스토어드 프로세스는 기본적으로 보고서를 HTML 로 전송하지만 JMP 환경 설정의 "SAS Integration" 페이지에서 RTF 또는 PDF를 대신 선택할 수 있습니다. JMP 환경 설정을 보려면 **파일 > 환경 설정**(Windows) 또는 **JMP > 환경 설정**(macOS)을 선택합니다.

SAS 코드 전송

JMP에서 현재 활성 SAS 서버로 직접 SAS 코드를 전송할 수 있습니다. 전송된 SAS 코드가 SAS 목록 출력을 생성하는 경우 해당 출력은 자동으로 SAS 서버에서 검색되어 JMP에 표시됩니다. 또한 생성된 SAS 로그가 검색되며, 전송된 코드에 오류가 있는 경우에는 SAS 로그 창에 SAS 로그가 자동으로 표시됩니다.

JMP에서의 모든 분석은 SAS 시스템에 의존하지 않고 JMP 내에서 자체적으로 실행됩니다. JMP에서 생성되는 SAS 코드는 초기 JMP 분석 후 SAS 시스템에서 별도의 유사한 분석을 수행하거나 JMP 내에서 적합된 모형을 사용하여 SAS 시스템에서 새 관측값의 점수를 계산하기 위한 것입니다.

그림 3.63 SAS 코드 전송 예

SAS 코드

```
data a;
x=1;

proc print;
run;
```

SAS 코드를 전송하기 위한 JMP 코드

```
SAS Connect("sasmain");
SAS Submit("data a; x=1; proc print; run;");
```

SAS 코드를 생성하는 JMP 플랫폼은 다음과 같습니다 .

- 표준 최소 제곱 - PROC GLM
- REML - PROC MIXED
- 단계별 - PROC GLM
- 명목형 및 순서형 로지스틱 - PROC LOGISTIC
- GLM - PROC GENMOD
- 시계열 (ARIMA 및 TFM) - PROC ARIMA
- 신경망 - SAS Data 스텝 점수 매기기 코드
- 파티션 (결정 트리 , 붓스트랩 포레스트 , 부스티드 트리) - SAS Data 스텝 점수 매기기 코드

참고 : 예측 계산식을 SAS Data 스텝에서 사용할 수 있는 표현식으로 전환하려면 JSL 함수 `As SAS Expr( formula );` 를 사용하십시오 . 자세한 내용은 JSL Syntax Reference 의 에서 확인하십시오 .

JMP 에서 직접 SAS 코드를 실행하려면

1. **파일 > 열기**를 사용하여 기존 SAS 프로그램을 열거나 , 새 SAS 프로그램을 생성합니다 . 새 SAS 프로그램을 생성하려면 **파일 > SAS > SAS 프로그램 새로 만들기**를 선택하고 SAS 코드를 입력합니다 .
2. **SAS 로 전송** 아이콘 을 클릭합니다 .

   프로그램 편집기 창을 마우스 오른쪽 버튼으로 클릭하고 **SAS 로 전송**을 선택할 수도 있습니다 . 이 메뉴 항목은 SAS 코드를 전송할 대상인 활성 SAS 서버의 이름도 포함합니다 .

   F8 키를 누를 수도 있습니다 (macOS 의 경우 Command+Shift+R 을 누름).

JSL 스크립트를 사용하여 SAS 코드를 실행하려면

JSL 함수 `SAS Submit` 또는 `SAS Submit File`을 사용하는 JSL 스크립트를 작성하여 실행합니다. SAS 코드를 전송하는 JSL 스크립트를 작성하는 방법에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

SAS 목록 출력을 보려면

전송된 SAS 코드가 SAS 목록(텍스트) 출력을 생성하는 경우, 작업이 완료되면 해당 출력이 "SAS 출력" 창에 자동으로 표시됩니다. 나중에 JMP 세션에서 SAS 목록 출력을 다시 봐야 하는 경우에는 **파일 > SAS > SAS 출력 창 열기**를 선택합니다. "SAS 출력" 창에는 이전에 활성 SAS 서버로 전송한 25회의 작업에서 생성된 목록 출력이 유지됩니다.

SAS 로그를 보려면

전송된 SAS 코드에 오류가 있으면 활성 SAS 서버에 대한 SAS 로그 창이 자동으로 열리고 해당 작업에 대한 SAS 로그가 표시됩니다. 하지만 언제든지 **파일 > SAS > SAS 로그 창 열기**를 선택하여 최근에 활성 서버로 전송한 25회의 작업에 대한 SAS 로그를 볼 수 있습니다.

전송이 완료된 후 SAS 로그 정보를 JMP 로그에 추가하려면 다음을 수행하십시오.

1. **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 .
2. "SAS Integration" 범주를 엽니다 .
3. "SAS 로그 표시 " 섹션에서 **별도 창** 대신 **JMP 로그**를 선택합니다 .

또한 SAS 로그 표시 섹션에서 SAS 로그의 표시 방식을 **항상**, **안 함** 또는 **오류 발생 시**(기본값) 중 하나로 설정할 수 있습니다.

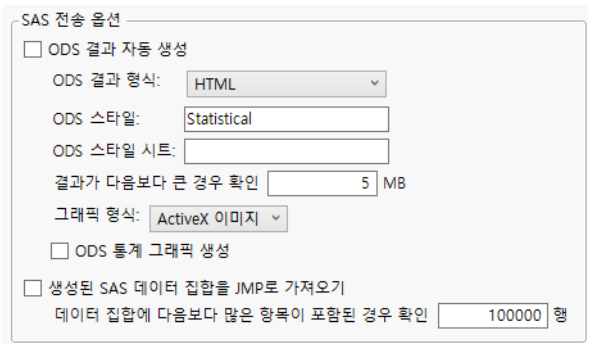
ODS 결과 생성

SAS ODS(Output Delivery System)는 HTML, RTF, PDF 및 기타 형식의 보고서를 생성하기 위한 강력한 메커니즘입니다. ODS 출력은 대개 일반 텍스트 형식인 SAS 목록 출력보다 훨씬 더 유용하며 사용자 정의가 가능합니다. 전송된 SAS 코드로 SAS 목록 출력 대신 ODS 결과가 생성되도록 환경 설정을 지정할 수 있습니다.

전송된 SAS 코드에서 ODS 결과를 생성하려면

1. **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 .
2. "SAS Integration" 범주를 열고 큰 "SAS 전송 옵션 " 그룹을 찾습니다 .

그림 3.64 환경 설정의 SAS 전송 옵션



3. **ODS 결과 자동 생성** 옵션을 선택합니다 .
4. **ODS 결과 형식** 목록에서 ODS 결과를 생성하는 데 사용할 형식 (HTML, PDF, RTF 또는 JMP 보고서) 을 선택합니다 .
5. (선택 사항) 다른 옵션을 사용하여 결과 형식을 지정하는 데 사용할 스타일 또는 스타일 시트를 지정하거나 생성되는 그래픽의 형식을 설정할 수 있습니다 . 자세한 내용은 "JMP 환경 설정 " 장의 "SAS Integration 의 환경 설정 "(663 페이지) 에서 확인하십시오 .

앞의 단계를 수행하면 전송한 SAS 코드를 래핑하는 추가 SAS 코드(ODS 문 포함)가 JMP에서 생성됩니다. 그런 다음에는 전송한 SAS 코드가 지정된 형식의 ODS 결과를 자동으로 생성합니다. 해당 결과는 사용자 컴퓨터로 다운로드되어 JMP 내에서 표시되거나(가능한 경우) 적절한 외부 응용 프로그램에서 표시됩니다.

생성된 SAS 데이터 집합 가져오기

사용자가 전송하는 SAS 코드가 SAS 데이터 집합을 생성할 수도 있습니다. 추가 분석을 위해 이러한 데이터 집합을 자동으로 JMP로 가져오도록 설정할 수 있습니다.

1. **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 .
2. "SAS Integration" 범주를 엽니다 .
3. **생성된 SAS 데이터 집합을 JMP 로 가져오기** 옵션을 선택합니다 .

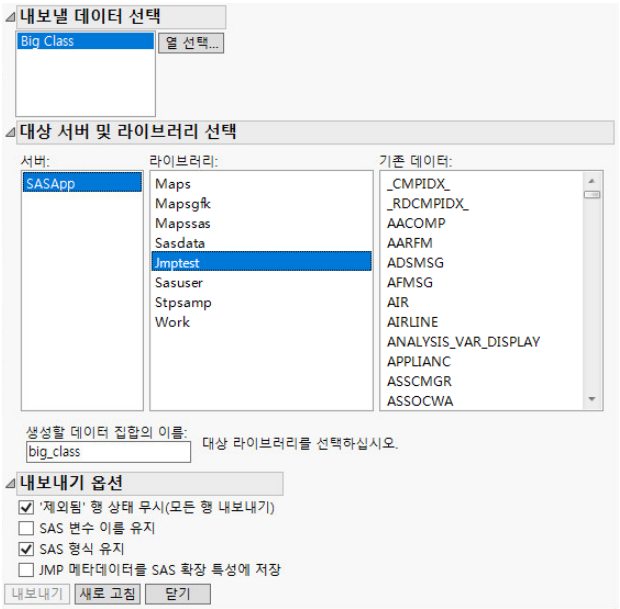
JMP 데이터 테이블을 SAS 로 내보내기

JMP 데이터 테이블을 SAS Workspace 서버로 내보낼 수 있습니다.

1. SAS Workspace 서버에 연결합니다 .
2. 내보내려는 파일을 엽니다 .
3. **파일 > SAS > SAS 로 데이터 내보내기**를 선택합니다 .

   필요한 경우 사용자 프로파일의 사용자 이름 및 암호를 사용하여 사용자가 자동으로 연결됩니다 .
4. " 내보낼 데이터 선택 " 아래의 열려 있는 데이터 테이블 목록에서 SAS 로 내보내려는 데이터 테이블을 선택합니다 .

그림 3.65 SAS 로 데이터 내보내기



5. (선택 사항) 데이터 테이블의 일부 열만 내보내려면 **열 선택**을 클릭합니다 . 자세한 내용은 " 내보낼 열을 선택하려면 "(155 페이지) 에서 확인하십시오 .
6. 대상 서버를 선택합니다 .
7. 라이브러리를 선택합니다 .

팁 : 라이브러리가 나타나지 않는 경우에는 "SAS 로 데이터 내보내기 창에 라이브러리 표시 "(156 페이지) 를 참조하십시오 .

라이브러리에 포함된 데이터 집합의 목록이 나타납니다 .

8. SAS 라이브러리에서 사용할 표시 이름을 입력합니다 .
9. (선택 사항) 사용할 내보내기 옵션을 설정합니다 . 자세한 내용은 " 내보내기 옵션 "(156 페이지) 에서 확인하십시오 .
10. **내보내기**를 클릭합니다 .

내보낼 열을 선택하려면

1. 데이터 테이블의 일부 열만 내보내려면 **열 선택**을 클릭합니다 .
2. 나타나는 창에서 내보낼 열을 선택하고 **추가**를 클릭합니다 .
3. 선택한 열 목록에 열을 모두 추가했으면 **확인**을 클릭합니다 .

내보내기 옵션

' 제외됨 ' 행 상태 무시 (모든 행 내보내기) 데이터 테이블의 모든 행을 내보내려면 이 옵션을 선택합니다 . 제외되지 않은 행만 내보내려면 이 옵션을 선택 취소합니다 . 이 옵션은 기본적으로 설정되어 있습니다 .

SAS 변수 이름 유지 이 옵션은 원래 SAS 에서 가져온 데이터 테이블에 유용합니다 . SAS 데이터 집합을 가져올 때 원래 SAS 변수 이름은 각 열의 열 특성에 저장됩니다 . SAS 로 내보낼 때 각 열에 SAS 변수 이름을 사용하려면 이 옵션을 선택합니다 . JMP 변수 이름을 내보내려면 이 옵션을 선택 취소합니다 . 이 옵션은 기본적으로 해제되어 있습니다 .

SAS 형식 유지 이 옵션은 원래 SAS 에서 가져온 데이터 테이블에 유용합니다 . SAS 데이터 집합을 가져올 때 원래 SAS 형식 및 정보는 각 열의 열 특성에 저장됩니다 . SAS 로 내보낼 때 각 열에 SAS 형식 및 정보를 사용하려면 이 옵션을 선택합니다 . JMP 형식을 대신 내보내려면 이 옵션을 선택 취소합니다 . 이 옵션은 기본적으로 설정되어 있습니다 .

SAS 로 데이터 내보내기 창에 라이브러리 표시

"SAS로 데이터 내보내기" 창에 라이브러리가 표시되지 않는 경우에는 다음 중 한 가지 방법으로 라이브러리를 정의하십시오.

- JSL 을 사용하여 코드를 SAS 서버로 전송합니다 . 이 코드는 `libname` 명령을 사용하여 라이브러리를 정의합니다 .
- SAS 가 호출될 때마다 SAS 코드 snippet 을 실행하는 autoexec.sas 파일을 정의합니다 . 그러면 SAS에 연결할 때마다 동일한 라이브러리 참조가 생성됩니다 . autoexec.sas 파일에 대한 자세한 내용은 SAS 설명서를 참조하십시오 .

SAS Management Console 의 " 데이터 라이브러리 관리자 "에 정의된 라이브러리와 같이 메타데이터에 정의된 라이브러리는 "SAS로 데이터 내보내기" 창에서 액세스할 수 없습니다.

데이터베이스에서 데이터 가져오기

특정 데이터베이스용 ODBC(Open Database Connectivity) 드라이버가 있는 경우 해당 데이터베이스에서 데이터를 가져오고 , 이후에 데이터를 다시 데이터베이스에 저장할 수 있습니다 .

SQL(Structured Query Language) 문을 사용하여 외부 데이터베이스에서 가져올 데이터를 제어할 수 있습니다. JMP에서 데이터베이스 파일을 열 때 실제로는 SQL 문을 데이터베이스로 전송하는 것입니다. 기본적으로 이 문은 데이터베이스 테이블의 모든 열 및 레코드를 가져옵니다. 대화식으로 쿼리를 작성하는 방법에 대한 자세한 내용은 " 쿼리 빌더에서 SQL 쿼리 작성 "(110 페이지)에서 확인하십시오. 직접 쿼리를 작성하는 방법에 대한 자세한 내용은 "SQL 문을 작성하여 데이터베이스 쿼리 "(160 페이지)에서 확인하십시오.

이 섹션에서는 데이터베이스에 연결하고 데이터를 가져오는 방법에 대해 설명합니다.

참고 :

- `$# -+/%()&|;?` 문자가 포함된 데이터베이스 테이블 이름은 지원되지 않습니다 .

- 하나의 데이터 소스에 대한 여러 연결은 허용되지 않습니다 .
- JMP 비트 버전과 매칭되도록 64 비트 버전의 ODBC 드라이버를 사용하십시오 .

데이터베이스에 연결

JMP는 ODBC 관리자를 통해 ODBC 데이터 소스를 사용하여 데이터베이스와 통신할 수 있습니다. ODBC 관리자 및 관련 드라이버는 운영 체제에 따라 다릅니다. 운영 체제 소프트웨어에서 데이터 소스를 생성한 후 JMP에서 다음 단계를 따라 데이터베이스에 연결하십시오.

1. **파일 > 데이터베이스 > 테이블 열기**를 선택합니다 . " 연결 " 상자에는 현재 JMP 세션에서 연결된 데이터 소스가 나열됩니다 .
2. **새 연결**을 클릭합니다 .
3. (Windows) " 데이터 원본 선택 " 창 (그림 3.66) 에서 " 컴퓨터 데이터 원본 " 탭을 클릭하고 , 데이터 소스를 선택하고 , **확인**을 클릭하고 , 사용자 이름 및 암호를 입력한 후 , **확인**을 클릭합니다 .

   (macOS) "Choose DSN" 창에서 데이터 소스를 선택하고 , 사용자 이름 및 암호를 입력한 후 , **Choose DSN** 을 클릭합니다 .

그림 3.66 데이터 소스 선택 (Windows)

" 데이터베이스 테이블 열기 " 창에 새 연결이 표시됩니다 .

데이터베이스에서 데이터 열기

ODBC 데이터베이스에 연결하고 가져올 테이블을 선택하고 나면 해당 데이터가 데이터 테이블에서 열립니다. 데이터 테이블에는 몇 개의 테이블 스크립트가 포함되어 있습니다.

- 데이터베이스에 다시 연결하려면 " 소스 " 스크립트를 실행합니다 .

- 데이터를 다시 가져오고 새로 고치려면 "DB 에서 업데이트 " 스크립트를 실행합니다 . 데이터베이스 테이블에 원래 JMP 테이블과 동일한 수의 열이 포함되어 있으면 해당하는 기존 JMP 테이블의 값이 즉시 업데이트됩니다 . 하지만 열 수가 동일하지 않으면 새 데이터 테이블이 열리고 여기에 업데이트된 데이터가 저장됩니다 .
- 데이터 테이블을 데이터베이스에 저장하려면 "DB 에 저장 " 스크립트를 실행합니다 . 그러면 데이터베이스의 기존 데이터가 바뀝니다 . 이 스크립트에는 사용자 이름 및 암호가 포함될 수 있습니다 . 이러한 중요한 정보가 포함되지 않도록 ODBC 연결 문자열 숨기기라는 환경 설정을 지정할 수 있습니다 . 이 환경을 설정을 찾으려면 Windows 에서 **파일 > 환경 설정 > 테이블**을 선택하거나 macOS 에서 **JMP > 환경 설정 > 테이블**을 선택합니다 .

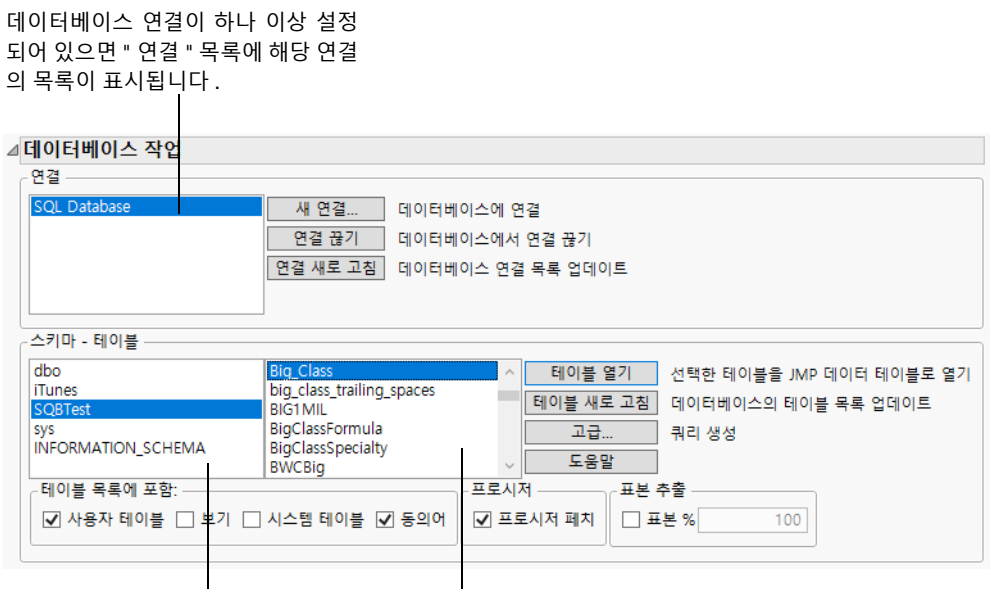
데이터베이스에서 데이터를 가져오려면

1. **파일 > 데이터베이스 > 테이블 열기**를 선택합니다 .

   " 데이터베이스 테이블 열기 " 창이 나타납니다 (그림 3.67).
2. 데이터베이스에 이미 연결되어 있으면 " 연결 " 상자에서 해당 데이터베이스를 선택합니다 . " 데이터베이스에 연결 "(157 페이지) 의 단계를 따르십시오 .

   " 연결 " 상자에는 JMP 가 연결된 데이터 소스가 나열됩니다 . " 스키마 - 테이블 " 상자에는 해당 데이터 소스를 지원하는 데이터베이스에 대한 스키마가 나열됩니다 .

그림 3.67 데이터베이스 테이블 열기 창



참고 : ODBC 드라이버가 프로시저 페치를 지원하지 않는 경우에는 **프로시저 페치** 체크박스를 사용할 수 없습니다 .

3. 원하는 데이터 소스가 " 연결 " 상자에 나열되지 **않으면 새 연결**을 클릭하여 데이터 소스를 선택합니다 . 데이터 소스를 선택하는 방법은 사용하는 운영 체제에 따라 다릅니다 .
4. " 연결 " 상자에서 원하는 데이터 소스를 선택합니다 . 그에 따라 " 테이블 " 상자의 테이블 목록이 업데이트됩니다 . 업데이트에는 몇 초 정도가 걸릴 수 있으며 이 시간은 테이블 수와 데이터베이스 연결 속도에 따라 달라집니다 . 데이터베이스가 스키마를 지원하면 목록의 첫 번째 스키마에 대한 테이블이 로드되며 , 다른 스키마를 클릭하면 해당 스키마에 대한 테이블이 로드됩니다 .
5. **테이블 목록에 포함** 체크박스 그룹에 있는 옵션을 선택하여 나열할 테이블을 제어합니다 .

참고 : 이러한 라벨은 드라이버에 따라 다르게 해석됩니다 .

사용자 테이블 이 옵션을 클릭하면 사용 가능한 모든 사용자 테이블이 **테이블** 목록에 표시됩니다 . 사용자 테이블은 컴퓨터에 로그온한 사용자에 따라 달라집니다 .

보기 이 옵션을 클릭하면 **테이블** 목록에 열 수 있는 다른 모든 파일 유형과 함께 " 보기 " 가 표시됩니다 . " 보기 " 는 가상 테이블로서 , 사용자가 열 때마다 업데이트되는 쿼리 결과 집합입니다 . 보기는 하나 이상의 테이블에서 정보를 추출하여 결합하는 데 사용됩니다 .

시스템 테이블 이 옵션을 클릭하면 사용 가능한 모든 시스템 테이블이 **테이블** 목록에 표시됩니다 . 시스템 테이블은 모든 사용자 또는 시스템 수준의 서비스가 사용할 수 있는 테이블입니다 .

동의어 이 옵션을 클릭하면 사용 가능한 모든 ORACLE 동의어가 **테이블** 목록에 표시됩니다 .

표집 테이블 목록에 표시할 행의 백분율을 입력합니다 . 이 옵션을 선택하면 큰 데이터베이스에서 쿼리 속도가 빨라집니다 . JMP 는 데이터베이스가 지원하는 표집 방법을 사용합니다 . 데이터베이스가 표집을 지원하지 않는 경우에는 이 체크박스를 사용할 수 없습니다 .

6. " 테이블 " 목록에서 원하는 테이블을 선택합니다 .

참고 : dBase 데이터베이스에 연결된 경우에는 연결할 대상 데이터베이스 **폴더**를 선택하십시오 . 개별 파일은 회색으로 표시되며 선택할 수 없습니다 .

7. **테이블 열기**를 클릭하여 선택한 테이블의 모든 데이터를 가져오거나 , **고급**을 클릭하여 가져올 테이블 부분집합을 지정합니다 . 일부 데이터베이스의 경우 사용자 ID 및 암호를 입력해야 데이터에 액세스할 수 있습니다 .

   큰 테이블을 열 때는 약간의 지연이 발생할 수 있습니다 . 모든 활성 ODBC 쿼리의 상태를 보려면 **보기 > 실행 중인 쿼리**를 선택합니다 .

참고 : 이전에 JMP 에서 데이터베이스로 내보냈던 데이터이고 데이터에 표현식 열이 포함되어 있으면 해당 열을 문자 열로 가져오게 됩니다 . **열 > 열 정보**를 선택하고 " 데이터 유형 " 을 " 표현식 " 으로 변경하십시오 .

SQL 문을 작성하여 데이터베이스 쿼리

SQL(Structured Query Language) 문을 사용하여 데이터베이스에서 가져올 데이터를 제어할 수 있습니다. JMP 에서 데이터베이스 파일을 열 때 실제로는 SQL 문을 데이터베이스로 전송하는 것입니다. 기본적으로 이 문은 데이터베이스 테이블의 모든 열 및 레코드를 가져옵니다. 일부 경우에는 가져오는 데이터가 너무 많을 수 있습니다. 테이블의 일부 데이터에만 관심이 있는 경우 SQL 요청을 사용자 정의하여 원하는 데이터만 요청할 수 있습니다. SQL 쿼리를 실행한 후에는 쿼리의 코드가 데이터 테이블의 SQL 테이블 변수에 저장됩니다.

이 섹션에서는 SQL 문을 작성하여 데이터를 가져오는 방법을 설명합니다. SQL 문을 작성하지 않고 대화식으로 데이터를 쿼리하려면 쿼리 빌더를 사용하십시오. 쿼리 빌더에서 쿼리 생성을 시작한 후 SQL 을 직접 추가할 수도 있습니다. 자세한 내용은 "쿼리 빌더에서 SQL 문 작성"(131 페이지)에서 확인하십시오.

1. **파일 > 데이터베이스 > 테이블 열기**를 선택합니다 .

   " 데이터베이스 테이블 열기 " 창이 나타납니다 (그림 3.67).

2. 필요한 경우 데이터베이스에 연결하거나 , 기존 데이터베이스 연결을 선택합니다 . " 데이터베이스에 연결 "(157 페이지) 의 단계를 따르십시오 .

   " 연결 " 상자에는 JMP 가 연결된 데이터 소스가 나열됩니다 . " 스키마 - 테이블 " 상자에는 해당 데이터 소스를 지원하는 데이터베이스에 대한 스키마가 나열됩니다 .

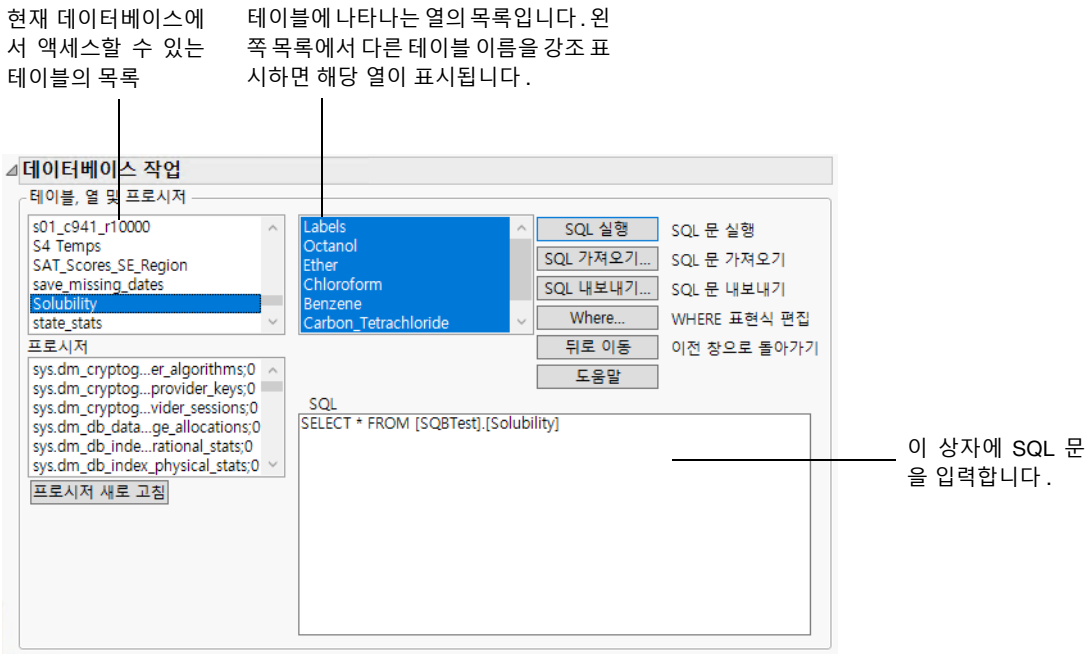
   참고 : 이 창에서 실행하는 SQL 쿼리는 창의 왼쪽 페인에 표시되는 테이블 및 프로시저에 대해서만 작동합니다 . 관련 없는 SQL 을 실행하면 아무 결과도 생성되지 않습니다 .

3. "데이터베이스 테이블 열기" 창에서 **고급** 버튼을 클릭하여 테이블의 특정 부분집합을 엽니다.

4. 올바른 SQL 문을 입력하거나 , 기본 문을 수정합니다 . 그림 3.68 에서는 선택된 파일에 적절한 기본 SQL `Select` 문을 보여 줍니다 . 사용할 수 있는 SQL 문에 대한 자세한 내용은 "SQL(Structured Query Language): 참조 "(161 페이지) 에서 확인하십시오 .

   Where 버튼을 추가하고 WHERE 절 편집기를 사용하여 표현식을 생성하는 방법으로 표현식을 추가할 수도 있습니다 . 자세한 내용은 "WHERE 절 편집기 사용 "(165 페이지) 에서 확인하십시오 .

그림 3.68 Excel 파일에 저장된 Solubility 테이블에서 모든 변수 읽기



5. **SQL 실행**을 클릭합니다. 선택한 열이 포함된 JMP 데이터 테이블이 나타납니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블 변수 사용"(261 페이지)에서 확인하십시오.
6. 실행 중인 모든 쿼리의 상태를 보려면 **보기 > 실행 중인 쿼리**를 선택합니다.

올바른 SQL 문을 입력하고 **SQL 실행**을 클릭하여 명령을 실행할 수 있습니다. 올바른 SQL은 데이터 소스와 ODBC 드라이버에 따라 다릅니다.

SQL(Structured Query Language): 참조

다음 섹션에서는 SQL에 대해 간략히 소개합니다. 이러한 섹션은 쿼리의 기능을 파악하기 위한 것일 뿐, 포괄적인 참조 자료로 활용할 수 있는 것은 아닙니다.

SELECT 문 사용

JMP의 기본적인 SQL 문은 SELECT 문입니다. 이 문은 데이터 소스에서 페치할 행을 데이터베이스에 알려 줍니다. Solubility.jmp 샘플 데이터 테이블을 사용하여 "SQL 문을 작성하여 데이터베이스 쿼리"(160 페이지)의 절차를 완료했다면 실제로는 다음과 같은 SQL 문을 데이터 소스로 보낸 것입니다.

```
SELECT * FROM "Solubility"
```

* 연산자는 "모든 열"을 나타내는 약식 표현입니다. 따라서 이 문은 지정된 데이터 테이블의 모든 열을 반환하라는 요청을 데이터베이스에 보냅니다.

모든 행을 반환하는 대신, *를 데이터 테이블의 특정 열 이름으로 바꿀 수 있습니다. Solubility 데이터 테이블 예의 경우 다음 문을 전송하여 ETHER, OCTANOL 및 CHLOROFORM 열만 선택할 수 있습니다.

```
SELECT ETHER, OCTANOL, CHLOROFORM FROM "Solubility"
```

참고 : JMP 에서는 SQL 문을 세미콜론으로 끝내지 않아도 됩니다 .

JMP에서는 실제 SQL을 입력할 필요 없이 그래픽 인터페이스를 통해 간단한 SELECT 문을 생성할 수 있습니다. 데이터 소스에서 특정 열을 선택하려면 열 목록에서 해당 열을 강조 표시하십시오.

일부 행을 강조 표시하려면

- Shift 키를 누른 채 열 이름 범위를 클릭하여 선택
- Ctrl 키를 누른 채로 클릭 (Windows) 하거나 Command 키를 누른 채로 클릭 (macOS) 하여 개별 열 이름을 선택합니다 .

그러면 선택한 항목에 맞게 SQL 문이 변경됩니다.

때로는 데이터 소스에서 고유한 레코드만 페치하려는 경우가 있습니다. 즉, 중복 레코드를 제거하려고 합니다. 이를 위해서는 DISTINCT 키워드를 사용합니다.

```
SELECT DISTINCT ETHER, OCTANOL, CHLOROFORM FROM "Solubility"
```

결과 정렬

데이터베이스에 있는 하나 이상의 필드를 기준으로 결과를 정렬할 수 있습니다. 정렬할 변수는 ORDER BY 명령을 사용하여 지정합니다.

```
SELECT * FROM "Solubility" ORDER BY LABELS
```

이 문은 모든 필드를 선택하고, 결과 데이터 테이블을 LABELS 변수를 기준으로 정렬합니다. 정렬 기준 변수를 추가로 지정하려면 쉼표로 구분된 목록으로 변수를 추가합니다.

```
SELECT * FROM "Solubility" ORDER BY LABELS, ETHER, OCTANOL
```

WHERE 문 사용

WHERE 문을 사용하면 조건에 따라 데이터 테이블의 일부 행만 페치할 수 있습니다. 예를 들어 ETHER 열의 값이 1보다 큰 모든 행을 선택할 수 있습니다.

```
SELECT * FROM "Solubility" WHERE ETHER > 1
```

WHERE 문은 FROM 문 다음에 배치되며 다음과 같은 논리 연산자를 사용할 수 있습니다.

표 3.4 WHERE 연산자

연산자	의미
=	같음
!= 또는 <>	같지 않음
>	보다 큼
<	보다 작음
>=	보다 크거나 같음
<=	보다 작거나 같음
NOT	논리적 NOT
AND	논리적 AND
OR	논리적 OR

조건을 평가할 때는 전체 문에 대해 NOT 문이 먼저 처리된 후 AND 문과 OR 문이 처리됩니다. 따라서

```
SELECT * FROM "Solubility" WHERE ETHER > -2 OR OCTANOL < 1 AND CHLOROFORM > 0
```

이 문은 다음 문과 동등합니다.

```
SELECT * FROM "Solubility" WHERE ETHER > -2 OR (OCTANOL < 1 AND CHLOROFORM > 0)
```

IN 및 BETWEEN 문 사용

페치할 값 범위를 지정하려면 IN 및 BETWEEN 문을 WHERE와 함께 사용합니다. IN 문은 값 목록을 지정하며 BETWEEN은 값 범위를 지정하는 데 사용할 수 있습니다. 예를 들면 다음과 같습니다.

```
SELECT * FROM "Solubility" WHERE LABELS IN ('Methanol','Ethanol','Propanol')
```

이 문은 LABELS 열의 값이 Methanol, Ethanol 또는 Propanol인 모든 행을 페치합니다.

```
SELECT * FROM "Solubility" WHERE ETHER BETWEEN 0 AND 2
```

이 문은 ETHER 값이 0부터 2 사이인 모든 행을 페치합니다.

LIKE 문 사용

LIKE 문을 사용하여 지정한 문자열과 유사한 값을 선택할 수 있습니다. 임의의 값을 취할 수 있는 여러 문자로 된 문자열을 나타내려면 %를 사용합니다. 예를 들어 Solubility 데이터에서 알코올 화합물, 즉 OL로 끝나는 화합물을 선택할 수 있습니다. 다음 SQL 문은 이 작업을 수행합니다.

```
SELECT * FROM "Solubility" WHERE LABELS LIKE  '%OL'
```

% 연산자는 LIKE 문의 어느 곳에나 배치할 수 있습니다. 다음 예에서는 라벨이 M으로 시작하고 OL로 끝나는 모든 행을 추출합니다.

```
SELECT * FROM "Solubility" WHERE LABELS LIKE  'M%OL'
```

집계 함수 사용

집계 함수는 데이터 자체가 아니라 데이터의 요약을 페치하는 데 사용됩니다. SELECT 문에서는 다음과 같은 집계 함수를 사용합니다.

표 3.5 SELECT 문 함수

함수	의미
SUM( )	열의 합
AVG( )	열의 평균
MAX( )	열의 최대값
MIN( )	열의 최소값
COUNT( )	열의 행 수

다음은 몇 가지 예입니다.

- 다음 문은 ETHER 및 OCTANOL 열의 합을 요청합니다.
  ```
  SELECT SUM(ETHER), SUM(OCTANOL) FROM "Solubility"
  ```
- 다음 문은 ETHER 값이 1 보다 큰 행의 개수를 반환합니다.
  ```
  SELECT COUNT(*) FROM "Solubility" WHERE ETHER > 1
  ```
- 다음 문을 사용하면 알코올에 해당하는 데이터의 OCTANOL 값을 알 수 있습니다.
  ```
  SELECT AVG(OCTANOL) FROM "Solubility" WHERE LABELS LIKE  '%OL'
  ```

참고 : 집계 함수를 사용할 경우 결과 JMP 데이터 테이블에서 열 이름은 Expr1000, Expr1001 등이 됩니다. 페치가 완료된 후 열 이름을 바꿀 수 있습니다.

GROUP BY 및 HAVING 명령

집계 함수와 GROUP BY 및 HAVING 명령을 함께 사용하면 특히 유용합니다. 이러한 명령을 사용하면 데이터 집합의 필드 값에 따라 집계 함수를 여러 번 실행할 수 있습니다.

예를 들어 데이터 테이블에서 각 ETHER 값에 대해 ETHER=0, ETHER=1 등인 레코드의 개수를 계산할 수 있습니다.

- SELECT COUNT(ETHER) FROM "Solubility" GROUP BY (ETHER) 는 각 항목이 ETHER 의 한 수준에 해당하는 단일 데이터 열을 반환합니다.

- SELECT COUNT(ETHER) FROM "Solubility" WHERE OCTANOL > 0 GROUP BY (ETHER) 는 위의 문과 동일한 작업을 수행하되 단 OCTANOL 이 0 보다 큰 행만 대상으로 합니다 .

열 집계 함수와 함께 GROUP BY 를 사용할 때는 해당 열 자체를 SELECT 문에 포함합니다. 예를 들면 다음과 같습니다.

```
SELECT ETHER, COUNT(ETHER) FROM "Solubility" GROUP BY (ETHER)
```

이 문은 개수 외에도 ETHER 의 수준이 포함된 열을 반환합니다.

하위 쿼리 사용

집계 함수는 WHERE 문에서 사용할 값을 계산하는 데도 유용합니다. 예를 들어 ETHER 의 평균 값보다 큰 모든 값을 페치할 수 있습니다. 즉, ETHER 의 평균 값을 구한 후 이 평균보다 큰 값을 가진 레코드만 선택할 수 있습니다. 원하는 평균을 페치하는 데는 SELECT AVG(ETHER) FROM "Solubility"를 사용합니다. 따라서 적절한 SQL 명령은 WHERE 조건에서 이 문을 사용하는 것입니다.

```
SELECT * FROM "Solubility" WHERE ETHER > (SELECT AVG(ETHER) FROM "Solubility")
```

SQL 쿼리 저장 및 로드

쿼리를 생성한 후 나중에 이 쿼리를 반복하려고 할 수 있습니다. 이 경우 쿼리를 사용하려고 할 때마다 매번 쿼리를 입력할 필요가 없습니다. 대신 쿼리를 외부 파일로 내보낼 수 있습니다. 이렇게 하려면 그림 3.68 에 표시된 창에서 **SQL 내보내기** 버튼을 클릭합니다. 그러면 SQL 쿼리를 텍스트 파일로 저장할 수 있는 창이 나타납니다.

저장된 쿼리를 로드하려면 그림 3.68 에 표시된 창에서 **SQL 가져오기** 버튼을 클릭합니다. 그러면 저장된 쿼리로 이동할 수 있는 창이 나타납니다. 쿼리를 열면 해당 쿼리가 창에 로드됩니다.

WHERE 절 편집기 사용

JMP 에서는 ODBC 가져오기 도중 SQL 쿼리의 WHERE 절을 작성하기 위한 도움말이 제공됩니다. 또한 일반적인 SQL 기능을 사용하여 기본적인 표현식을 작성하는 데 유용한 WHERE 절 편집기도 제공되므로 벤더에 맞는 함수를 작성할 수 있습니다. 예를 들어 SQL 에서 비교에 '='을 사용하는지 '=='을 사용하는지, 또는 평균 계산에 avg()를 사용하는지 average()를 사용하는지 몰라도 됩니다.

또한 문자열 리터럴은 큰따옴표("string")가 아니라 작은따옴표('string')로 묶어야 합니다.

WHERE 절 편집기를 열려면

1. " 데이터베이스에 연결 "(157 페이지) 의 단계를 따라 데이터베이스에 연결합니다 .
2. **데이터베이스 테이블 열기** 창에서 **고급** 버튼을 클릭합니다 .
3. **Where** 버튼을 클릭합니다 .

작업 패널에서 표현식, 함수 및 항을 추가하려면 WHERE 절 편집기를 사용합니다. 해당 항목은 강조 표시된 파란색 상자에 적용됩니다.

1. 테이블 이름 브라우저를 클릭하여 테이블을 선택합니다 . 해당 테이블의 열이 목록에 나타납니다 .
2. SQL 벤더 이름 브라우저를 클릭하여 사용할 SQL 의 유형 (GenericSQL, Access, DB2, MySQL, Oracle, SQL Server 또는 이 유형 모두) 을 선택합니다 . 목록에서 함수 또는 연산자를 클릭하고 나타나는 목록에서 연산자를 선택하여 작업을 수행합니다 .

참고 : SQL Server 데이터 유형 Binary, Geography 및 Geometry 는 지원되지 않습니다 .

3. 계산식 편집 영역에서 빈 계산식 요소를 클릭하여 선택합니다. 선택한 요소 주위에는 빨간색 외곽선이 표시됩니다 . 클릭한 위치를 기준으로 가장 작은 내포 상자 내의 모든 항이 선택됩니다 . 이후 작업은 이러한 결합된 요소에 적용됩니다 .
4. 키패드의 버튼을 클릭하여 표현식에 연산자를 추가합니다 .
5. (선택 사항) WHERE 절을 사용자 정의하려면 계산식 편집기에서 빨간색 삼각형 메뉴의 다음 옵션 중 하나를 선택합니다 .

상자 표시 WHERE 절의 항을 둘러싸는 상자를 표시하거나 숨깁니다 .

큰 글꼴 계산식의 글꼴 크기를 늘립니다 .

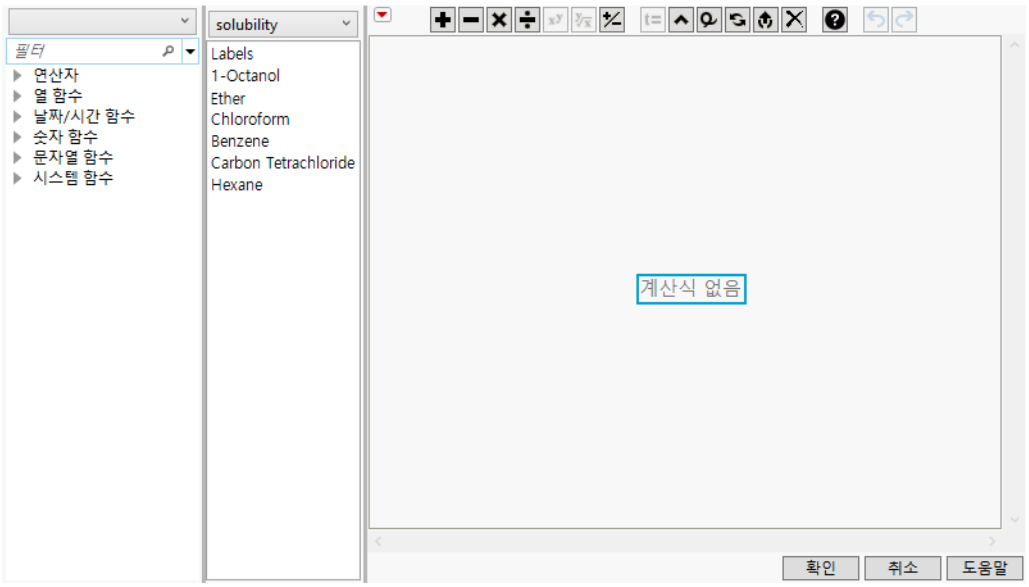
작은 글꼴 계산식의 글꼴 크기를 줄입니다 .

단순화 WHERE 절의 문을 가능한 한 최대로 단순화합니다 .

패널 레이아웃 재설정 그림 3.69 에 표시된 것과 같이 패널을 표시합니다 .

WHERE 절 편집기는 계산식 편집기 ("JMP 의 계산식 생성" 장 (359 페이지) 참조) 와 유사하게 작동합니다 .

그림 3.69 WHERE 절 편집기



Google 스프레드시트에서 데이터 가져오기

Google 스프레드시트는 Google 계정에 저장된 스프레드시트(탭 같은 시트로 구성)입니다. 이 응용 프로그램을 사용하면 스프레드시트를 계속 편집하면서 다른 사람과 공유할 수 있습니다.

Google 계정에 로그인하고 스프레드시트의 URL을 복사한 다음 데이터 구조화 방법을 지정하여 Google 스프레드시트에서 데이터를 가져옵니다. 예를 들어 열 머리글이 있는 경우에는 데이터가 어떤 행에서 시작되는지 알고 있어야 합니다.

Google 스프레드시트에서 데이터를 가져오려면 다음 단계를 수행하십시오 .

1. https://sheets.google.com 에 로그인하고 스프레드시트를 엽니다 .
2. 브라우저의 주소 표시줄에서 URL 을 복사합니다 .
3. JMP 에서 **파일 > 인터넷 열기 > Google 스프레드시트**를 선택합니다 .
4. Google 계정 이름 (@gmail.com 은 필요하지 않음) 을 입력합니다 .
5. 스프레드시트 URL 또는 ID 필드에 URL 을 붙여넣습니다 .
6. **가져오기**를 클릭합니다 .

   JMP 데이터 테이블로 데이터 가져오기가 수행됩니다 .

   지정한 Google 계정을 처음 사용하여 데이터를 가져오는 경우 JMP 에 Google 계정에 대한 액세스 권한을 부여하라는 메시지가 표시됩니다 .

   a. **허용**을 클릭합니다 .

      JMP 데이터 테이블로 데이터 가져오기가 수행됩니다 .

Google 스프레드시트 가져오기

- Google 이 사용자 계정을 인증하는 데 사용하는 정보는 C:/Users/<user name>/AppData/Roaming/SAS/JMP/OAuthTokens.jmp 에 저장됩니다 . Google 암호는 암호화됩니다 . 따라서 인증 정보를 다시 제공하지 않고도 동일한 Google 계정을 사용하여 데이터를 가져오고 내보낼 수 있습니다 . 컴퓨터의 사용자 계정에 대한 액세스 권한이 있는 사용자는 모든 Google 스프레드시트에 액세스할 수 있습니다 . 스프레드시트에 기밀 정보가 포함되어 있는 경우에는 사용자 계정을 공유하는 것을 재고하십시오 .
- 계정 이름에 @gmail.com 은 필요하지 않습니다 . 하지만 이 부분을 사용하면 <user name>@gmail.com 과 <user name> 이 개별 계정으로 처리됩니다 . 따라서 가져오거나 내보낼 때 두 계정 모두의 인증 코드를 입력해야 합니다 .
- 다른 사용자에게 이 컴퓨터의 사용자 계정에 대한 액세스 권한이 있는 경우에는 해당 사용자가 OAuthTokens.jmp 를 자신의 컴퓨터에 복사하여 스프레드시트를 가져갈 수 있습니다 .
- Windows 의 경우 Internet Explorer 에서 **호환성 보기에서 인트라넷 사이트 표시**를 선택하면 Google 스프레드시트를 가져올 수 없습니다 . 이 문제를 방지하려면 다음 단계를 따르십시오 .

  1. Internet Explorer 에서 " 도구 " 버튼 을 선택한 다음 **호환성 보기 설정**을 선택합니다 .

2. **호환성 보기에서 인트라넷 사이트 표시**를 선택 취소합니다 .

참고 : Google 스프레드시트에서는 **파일 > 웹에 게시**을 선택하여 웹 사용자가 볼 수 있도록 시트를 게시할 수 있습니다 . 이 경우 게시되는 시트의 링크에는 pubhtml 접미사가 포함됩니다 . 이러한 시트를 열려면 " 파일 " > " 인터넷 열기 " 를 사용하십시오 . 자세한 내용은 " 인터넷 또는 원격 컴퓨터에서 데이터 가져오기 "(174 페이지) 에서 확인하십시오 .

Google 스프레드시트 설정

필요한 경우 다음 설정을 변경한 다음 **가져오기**를 클릭하여 시트 데이터를 가져오십시오.

시트에 열 머리글이 포함됨 워크시트에 열 머리글이 있는 행이 포함되어 있으면 이 옵션을 선택합니다 .

열 머리글이 시작되는 행 워크시트에서 열 머리글이 시작되는 행을 나타냅니다 . 올바른 행에서 머리글이 시작될 때까지 위쪽 화살표 를 클릭하거나, 해당 행 번호를 입력하고 Enter 키를 누릅니다 .

데이터가 시작되는 행 워크시트에서 데이터가 시작되는 행을 나타냅니다 .

열어 볼 시트 선택 Ctrl 키를 누른 채 가져오지 않으려는 시트를 클릭하여 선택 취소합니다 .

고급 옵션

셀 범위 지정한 범위 (예 : G5:G17) 내의 셀을 가져옵니다 .

셀 색상 가져오기 워크시트의 셀 색상을 데이터 테이블에 적용합니다 .

비어 있는 열 제한 비어 있는 열이 데이터 테이블에 나타나지 않도록 합니다 .

SPSS 파일 가져오기

JMP 는 SPSS 파일을 데이터 테이블로 열고 몇 가지 SPSS 기능을 유지합니다.

- 최소한의 형식이 지정된 일반적인 숫자 및 문자 데이터가 지원됩니다 .
- SPSS 날짜 , 날짜 / 시간 및 시간 형식이 지원됩니다 .
- 기본적으로 라벨은 열 머리글로 변환됩니다 . 데이터에 라벨이 포함되어 있지 않은 경우 이 옵션을 선택하면 열 이름이 **열 1, 열 2** 등으로 지정됩니다 .

  SPSS 파일을 열 때 열 머리글의 변환 방법을 선택할 수도 있습니다 . 이 경우 선택한 방법이 환경 설정보다 우선합니다 .

  기본 변환 방법을 변경하려면 **파일 > 환경 설정** (macOS 의 경우 , **JMP > 환경 설정**) 을 선택합니다 . " 일반 " 페이지에서 **가져오는 동안 열 이름에 SPSS 라벨 사용**을 선택 취소합니다 . 그러면 자동으로 변수 이름을 열 머리글로 가져오게 됩니다 .

- SPSS 파일에서 정의한 값 라벨은 " 값 라벨 " 열 특성으로 저장됩니다 . 그러면 각 데이터 테이블 셀에서 원래 값 대신 이 값 라벨이 표시됩니다 . "값 라벨" 특성에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 값 라벨 "(284 페이지) 에서 확인하십시오 .
- SPSS 에서는 특정 값을 분석 시 결측값으로 처리하도록 지정할 수 있습니다 . 예를 들어 값 64 를 Height 변수의 결측값으로 간주할 수 있습니다 . 그러면 Height 의 분포를 계산할 때 값 64 는 무시됩니다 . SPSS 를 가져올 때 이러한 값은 해당 변수의 결측값 코드 열 특성에 포함됩니다 .

  최대 20 개의 숫자를 포함하는 결측값 범위가 지원됩니다 . 범위가 음수로 시작하면 최대값까지 숫자를 늘려갑니다 (예 : -10 ~ -5). 범위가 양수로 시작하면 최소값부터 숫자를 줄여갑니다 (예 : 1 ~ 12).

SPSS 파일에서 선택된 사용자 통화 형식은 가져올 때 유지되지 않습니다 . 또한 JMP 는 유니코드가 아닌 일본어 문자와 같은 더블바이트 문자가 포함된 SPSS 데이터를 읽지 못합니다 .

SPSS 파일을 열려면 (Windows)

1. **파일 > 열기**를 선택합니다 .
2. **파일 이름** 옆의 목록에서 **SPSS 데이터 파일 (*.sav)** 을 선택합니다 .
3. SPSS 파일을 선택합니다 .
4. (선택 사항) 열 머리글을 지정하려면 **JMP 열 이름을 설정할 때 사용할 원본** 옵션에서 다음 중 하나를 선택합니다 .
   - **SPSS 라벨** : SPSS 라벨을 사용하여 열 머리글을 생성합니다 .
   - **SPSS 변수 이름** : 변수 이름을 사용하여 열 머리글을 생성합니다 .
5. **열기**를 클릭합니다 .

   JMP 에서 파일이 데이터 테이블로 열립니다 .

SPSS 파일을 열려면 (macOS)

1. **파일 > 열기**를 선택합니다 .
2. SPSS 파일을 선택합니다 .
3. (선택 사항) 열 머리글을 지정하려면 다음 중 하나를 수행합니다 .
   - 변수 이름을 열 머리글로 변환하려면 **SPSS 라벨을 머리글로 사용**을 선택 취소합니다 .
   - 라벨을 열 머리글로 변환하려면 **SPSS 라벨을 머리글로 사용**을 선택합니다 .
4. **열기**를 클릭합니다 .

   JMP 에서 파일이 데이터 테이블로 열립니다 .

참고 :

- 다른 파일을 가져올 때와 마찬가지로 큰 SPSS 파일을 열고 저장할 때는 지연이 발생할 수 있습니다 .

- 이제 파이프 (|) 가 .sav 파일의 SPSS 라벨에서 여러 반응 내의 단어를 구분합니다 (예 : "detergent|liquid"). 구분자를 변경하려면 `Preference(SPSSMultiResponseDelimiter (":"))` JSL 스크립트를 실행합니다 . 여기서 , 콜론이 새 구분자입니다 .

Triple-S 설문 조사 데이터 가져오기

Triple-S(SSS) 는 설문 조사 데이터의 규격입니다 . 설문 조사 데이터는 .xml 또는 .sss 파일과 .csv, .dat 또는 .asc 파일로 구성된 한 쌍의 파일에 저장됩니다 . .xml 또는 .sss 파일에서는 변수를 정의하고 설문 조사를 설명합니다 . 함께 제공되는 .csv, .dat 또는 .asc 파일에는 각 응답자로부터 수집한 데이터가 포함됩니다 .

기본적으로 변수 라벨이 열 머리글로 변환됩니다 . 변수 이름을 열 머리글로 변환하려면 JMP " 일반 " 환경 설정에서 **Triple-S 라벨을 머리글로 사용**을 선택 취소합니다 .

Triple-S 파일을 가져오려면 다음 단계를 수행하십시오 .

1. Triple-S 파일 쌍을 동일한 폴더에 넣고 두 파일 모두에 동일한 루트 이름을 사용합니다 .
2. **파일 > 열기**를 선택합니다 .
3. Windows 의 경우 , " 파일 이름 " 목록에서 **Triple-S 설문 조사 파일**을 선택합니다 .
4. SPSS 파일을 선택합니다 .
5. (Windows 의 경우 선택 사항) Triple-S 열 머리글 환경 설정을 재정의하려면 **JMP 열 이름을 설정할 때 사용할 원본** 옆에서 " 변수 라벨 " 또는 " 변수 이름 " 을 선택합니다 .
6. **열기**를 클릭합니다 .

   JMP 에서 파일이 데이터 테이블로 열립니다 .

참고 : 기본 인코딩은 UTF-8 입니다 . Windows-1252 로 인코딩된 Triple-S 파일은 ISO-8859-1 로 가져오게 됩니다 .

HDF5 데이터 가져오기

HDF5(Hierarchical Data Format 버전 5) 는 데이터를 저장하기 위한 호환 가능 파일 형식입니다 . HDF5 파일은 그룹과 데이터 집합으로 구성됩니다 . 파일을 가져오면 JMP 는 그룹을 열어 내부 데이터 집합의 이름을 제공합니다 . 예를 들어 파일에 songs 및 artists 테이블이 포함된 음악 녹음 그룹의 데이터가 포함된 경우 "songs" 를 열지 "artists" 를 열지 선택하라는 메시지가 표시됩니다 . 파일에 그룹이 포함되어 있으면 단일 테이블을 열라는 메시지가 표시됩니다 .

JMP 는 숫자 (integer, float, double) 및 문자열 유형의 테이블과 단순 유형만 포함된 3 차원 이하의 복합 파일만 처리합니다 . 가져오기 창에 데이터 집합이 나타나지 않으면 해당 파일이 지원되지 않을 가능성이 큽니다 .

최대 1,000,000 개의 열을 가져올 수 있으며 행 수에는 제한이 없습니다 .

참고 : HDF5 파일에서는 ASCII 문자만 지원됩니다 . 프랑스어 및 중국어 문자와 같이 ASCII 가 아닌 문자는 지원되지 않습니다 .

HDF5 파일을 가져오려면 다음 단계를 수행하십시오.

1. **파일 > 열기**를 선택합니다 .
2. Windows 의 경우 , " 파일 이름 " 목록에서 **모든 JMP 파일** 또는 **HDF5 데이터 파일**이 선택되어 있는지 확인합니다 .
3. HDF5 파일을 선택합니다 .

   파일에 포함된 그룹 또는 데이터 집합의 목록이 나타납니다 .
4. 가져올 데이터 집합을 선택하고 **가져오기**를 클릭합니다 .

   데이터가 데이터 테이블로 열립니다 .

MATLAB 에서 데이터 가져오기

JMP 에서는 MATLAB 에 대한 스크립팅 인터페이스가 제공됩니다. 자세한 내용은 Scripting Guide 의 및 JSL Syntax Reference 의 에서 확인하십시오.

R 에서 데이터 가져오기

JMP 에서는 R 에 대한 스크립팅 인터페이스가 제공됩니다. 자세한 내용은 Scripting Guide 의 및 JSL Syntax Reference 의 에서 확인하십시오.

Excel 추가기능을 사용하여 데이터 가져오기

Excel 용 추가기능은 Windows 에서 JMP 및 Excel 사용자에게 다음과 같은 새로운 기능을 제공합니다 .

- Excel 의 선택한 셀을 JMP 데이터 테이블로 전송합니다 . 자세한 내용은 "Excel 데이터를 데이터 테이블로 전송 "(172 페이지) 에서 확인하십시오 .
- Excel 통합 문서의 계산 모델과 함께 JMP 프로파일러를 사용합니다 . 이 프로파일러 도구는 Excel 워크시트에 있는 모델에 JMP 프로파일러의 기능을 적용하기 위한 것입니다 . JMP 에서 Excel 모델을 다시 생성하고 , 해당 모델이 올바른지 확인하고 , 모델을 JMP 와 Excel 모두에서 유지할 필요가 없습니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

  참고 : JMP 설치 중에 Excel 추가기능을 선택하십시오 . 그러면 사용 중인 Microsoft Excel 버전에 맞는 추가기능이 설치됩니다 .

- Microsoft Excel 2010, Microsoft Excel 2013, Microsoft Excel 2016, Microsoft Excel 2019 및 Office 365 Desktop Excel 이 지원됩니다 .

Excel 데이터를 데이터 테이블로 전송

Excel용 JMP 추가기능을 사용하여 Excel 워크시트를 다음과 같은 JMP 대상으로 전송할 수 있습니다.

- 데이터 테이블
- 그래프 빌더
- 분포 플랫폼
- X 로 Y 적합 플랫폼
- 모형 적합 플랫폼
- 시계열 플랫폼
- 관리도 플랫폼

JMP에서 프로파일링을 위한 모델을 생성할 수도 있습니다. 자세한 내용은 Profilers의 에서 확인하십시오.

Excel 의 데이터를 JMP 데이터 테이블 및 플랫폼으로 전송하려면

1. Excel 워크시트에서 리본 메뉴에 **JMP** 추가기능을 표시합니다 .
2. **환경 설정** 버튼을 클릭합니다 .
3. 기본 **데이터 테이블 이름** (파일 이름 _ 워크시트 이름) 을 그대로 사용하거나 이름을 입력합니다 .
4. 워크시트의 첫 번째 행에 열 머리글이 있으면 **첫 번째 행을 열 이름으로 사용**을 선택합니다 .
5. 첫 번째 행을 열 머리글로 사용하도록 선택한 경우 사용되는 행 수를 입력합니다 .
6. 워크시트에 있는 숨겨진 행을 JMP 데이터 테이블에 포함하려면 **숨겨진 행 전송**을 선택합니다.
7. 워크시트에 있는 숨겨진 열을 JMP 데이터 테이블에 포함하려면 **숨겨진 열 전송**을 선택합니다.
8. **확인**을 클릭하여 환경 설정을 저장합니다 .
9. 열 이름으로 사용할 셀을 포함하여 JMP 로 전송할 셀을 선택합니다 .

   열 이름으로 사용할 셀은 선택한 범위의 첫 번째 행이어야 합니다 .

   셀을 하나만 선택하거나 아무 셀도 선택하지 않은 경우에는 전체 Excel 워크시트가 JMP 로 전송됩니다 .
10. 도구 모음에서 다음 JMP 대상 중 하나를 선택합니다 .
    - 데이터 테이블
    - 그래프 빌더
    - 분포 플랫폼

– X 로 Y 적합 플랫폼
– 모형 적합 플랫폼
– 시계열 플랫폼
– 관리도 플랫폼

JMP 가 열리고 , 선택한 데이터가 새 JMP 데이터 테이블에 포함되며 , 선택한 시작 창이 나타납니다 .

참고 :

- 빈 셀은 결측 데이터로 가져와지며 , 날짜 , 숫자 및 문자열은 올바르게 인식됩니다 .
- JMP 창을 Excel 창 뒤에 숨길 수 있습니다 (특히 Excel 을 최대화한 경우).

다양한 JMP 플랫폼의 사용 방법에 대한 자세한 내용은 해당 JMP 설명서에서 확인하십시오.

Excel 용 JMP 추가기능

JMP 추가기능 명령은 다음과 같은 두 그룹으로 구성되어 있습니다 .

JMP 로 이동

환경 설정 Excel 의 데이터를 JMP 로 전송하기 위한 환경 설정을 지정합니다 .

데이터 테이블 Excel 파일에서 선택한 데이터를 JMP 데이터 테이블로 전송합니다 .

그래프 빌더 Excel 파일에서 선택한 데이터를 JMP 데이터 테이블로 전송하고 그래프 빌더 플랫폼을 시작합니다 .

분포 Excel 파일에서 선택한 데이터를 JMP 데이터 테이블로 전송하고 분포 플랫폼을 시작합니다 .

X 로 Y 적합 Excel 파일에서 선택한 데이터를 JMP 데이터 테이블로 전송하고 X 로 Y 적합 플랫폼을 시작합니다 .

모형 적합 Excel 파일에서 선택한 데이터를 JMP 데이터 테이블로 전송하고 모형 적합 플랫폼을 시작합니다 .

시계열 Excel 파일에서 선택한 데이터를 JMP 데이터 테이블로 전송하고 시계열 플랫폼을 시작합니다 .

관리도 Excel 파일에서 선택한 데이터를 JMP 데이터 테이블로 전송하고 관리도 플랫폼을 시작합니다 .

JMP 에 프로필

모형 생성 / 편집 Excel 데이터와 함께 JMP 프로파일러를 사용하기 위한 환경 설정을 지정합니다.

모형 실행 JMP 프로파일러를 실행합니다 .

JMP Excel 추가기능 제거

Excel 추가기능을 제거하려면 JMP 설치 관리자를 두 번 클릭하고, **수정**을 클릭한 후, **JMP Excel 추가기능**을 선택 취소하십시오.

인터넷 또는 원격 컴퓨터에서 데이터 가져오기

파일 > 인터넷 열기를 선택하여 인터넷, FTP 사이트 또는 다른 컴퓨터에서 데이터를 가져올 수 있습니다. 파일 경로는 http, ftp 또는 file로 시작하거나, 드라이브 문자로 시작하거나, 네트워크 드라이브의 경로(상대 또는 절대)로 시작할 수 있습니다.

"인터넷 열기"를 사용하면 인터넷 또는 원격 컴퓨터에서 데이터와 이미지를 가져와 데이터 테이블, 웹 페이지 또는 텍스트로 저장할 수 있습니다. 이 옵션을 사용하여 SAS 스토어드 프로세스 보고서를 열 수도 있습니다. "인터넷 열기"를 사용하여 Google 스프레드시트를 가져오는 방법에 대한 자세한 내용은 "Google 스프레드시트에서 데이터 가져오기"(167 페이지)에서 확인하십시오.

웹 페이지에 있는 테이블에서 데이터와 이미지를 가져와 데이터 테이블로 저장하려고 한다고 가정해 보겠습니다. 특정 웹 페이지에 인구별 국가가 나열되어 있습니다.

1. **파일 > 인터넷 열기 > 웹 페이지**를 선택합니다.
2. URL https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population 을 입력합니다.
3. **열기 형식**을 **데이터**로 설정된 상태로 두고 **확인**을 클릭합니다.

   JMP 가 지정된 페이지에서 몇 개의 테이블을 찾습니다.
4. **Rank country (or dependent territory)**... 를 클릭하고 **확인**을 클릭합니다.

   JMP 가 웹 페이지 테이블의 정보가 포함된 데이터 테이블을 엽니다.
5. (선택 사항) 데이터 테이블에 이미지를 추가하려면 **그림 로드** 스크립트를 실행합니다.

참고 :

- 원격 파일을 다음 형식 중 하나로 열 수 있습니다.

  데이터 파일을 데이터 테이블로 엽니다.

  웹 페이지 웹 페이지를 JMP 브라우저(Windows) 또는 기본 브라우저(macOS)에서 엽니다.

  Windows 의 경우, " 파일 " 메뉴에서 파일을 저장할지, 아니면 데이터를 데이터 테이블로 가져올지 선택할 수 있습니다. 웹 페이지 스크립트 및 서버 측 요청에 의해 생성된 데이터를 가져오려면 이 옵션을 선택하십시오.

  텍스트 파일을 스크립트 편집기에서 엽니다. HTML 파일을 가져올 경우 웹 페이지의 HTML 태그가 표시됩니다.

- FTP 서버의 파일을 열면 FTP 로그인 창이 나타납니다 . 익명 계정을 사용하려면 **확인**을 클릭하십시오 . 인증된 계정을 사용하려면 사용자 ID 및 암호를 입력하십시오 . 그러면 파일이 지정한 대로 열립니다 .

  참고 : 일부 익명 FTP 서버의 경우 사용자 ID 가 필요합니다 . 데이터 테이블이 열리지 않으면 " 사용자 ID" 텍스트 상자에 "ftp" 또는 "anonymous" 를 입력해 보십시오 . " 암호 " 텍스트 상자는 빈 상태로 두고 **확인**을 클릭하십시오 .

- JMP 는 오류로 인해 가져오기를 중지하기 전에 60 초 동안 대기합니다 . " 일반 " 환경 설정 페이지에서 이 인터넷 타임아웃 설정을 변경할 수 있습니다 .
- 가져오려는 테이블에 이미지가 포함되어 있으면 먼저 이미지를 텍스트로 가져오게 됩니다 . 데이터 테이블에 이미지를 표시하려면 자동으로 생성된 **그림 로드**라는 테이블 스크립트를 실행하십시오 . 그러면 이미지가 포함된 새 표현식 열이 데이터 테이블에 추가됩니다 .
- 데이터 테이블에서 " 소스 " 스크립트를 실행하면 데이터를 다시 가져오고 새로 고칠 수 있습니다 .

SAS 스토어드 프로세스 보고서를 데이터 테이블로 열려면 (Windows only)

1. **파일** > **인터넷 열기**를 선택하고 파일을 선택합니다 .

   파일이 JMP 브라우저에서 열립니다 .
2. JMP 브라우저에서 **파일 > 테이블을 데이터 테이블로 가져오기**를 선택합니다 .

   웹 페이지에서 찾은 테이블이 나열된 창이 나타납니다 .
3. 가져오려는 테이블을 선택합니다 .
4. **확인**을 클릭합니다 .

   각 테이블이 새 데이터 테이블로 열립니다 .

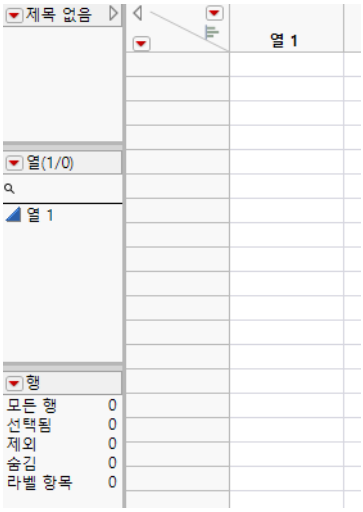
스토어드 프로세스 보고서에 대한 자세한 내용은 " 스토어드 프로세스 실행 "(150 페이지) 에서 확인하십시오 .

새 데이터 테이블 생성

데이터를 수동으로 입력하여 새 데이터 테이블을 생성하려면 다음 단계를 수행하십시오 .

1. **파일 > 새로 만들기 > 데이터 테이블**을 선택합니다 . 그러면 행은 없고 **열 1** 이라는 숫자 열 하나만 있는 빈 데이터 테이블이 표시됩니다 .
2. 커서를 셀에 놓습니다 .
3. 셀을 클릭합니다 . 셀 안에 선 모양의 커서가 나타납니다 .

그림 3.70 새 데이터 테이블



4. 값을 입력합니다 .

다음과 같은 몇 가지 방법으로 테이블에 값을 채울 수 있습니다 .

- 새 행 및 열을 생성하고 데이터 격자에 데이터를 입력하거나 붙여넣습니다 . 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블에 데이터 입력 "(180 페이지) 에서 확인하십시오 .
- 열 값을 계산하기 위한 계산식을 생성합니다 . 자세한 내용은 "JMP 의 계산식 생성 " 장의 " 계산식 작성 "(361 페이지) 에서 확인하십시오 .
- 다른 응용 프로그램에서 데이터를 가져옵니다 . 자세한 내용은 " 데이터 가져오기 정보 "(68 페이지) 에서 확인하십시오 .
- 다른 응용 프로그램에서 값을 복사하여 테이블에 붙여넣습니다 .
- 측정 기기를 사용하여 외부 측정값을 읽습니다 . 데이터 공급에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .
- 한 테이블의 열을 다른 테이블로 드래그합니다 .

데이터 테이블의 형식 지정 , 편집 및 작업 방법에 대한 자세한 내용은 " 데이터 입력 및 편집 " 장 (177 페이지) 에서 확인하십시오 .

4 장

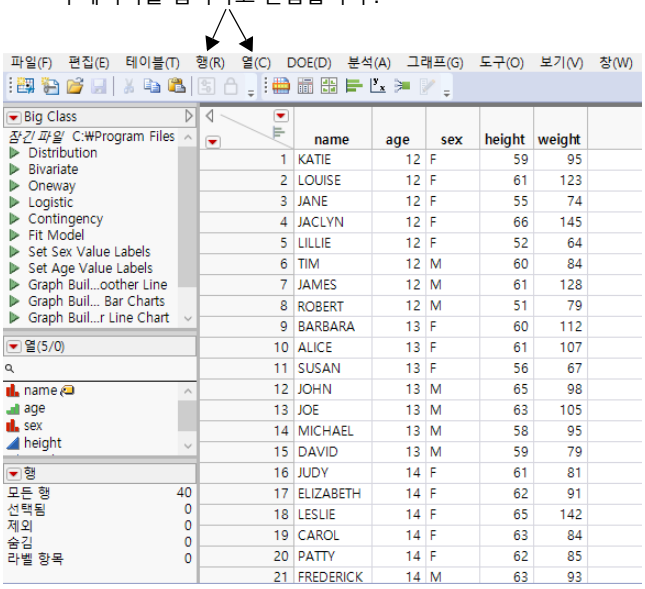
데이터 입력 및 편집
기본적인 데이터 테이블 작업 수행

JMP로 데이터를 가져오거나 새 데이터 테이블을 생성한 후에는 데이터 형식을 지정하여 분석을 위한 준비를 할 수 있습니다.

이 장에는 다음과 같은 정보가 포함되어 있습니다.

- 숫자 값 형식 변경
- 행 및 열 추가, 삭제 및 선택
- 행 편집기를 사용하여 행 탐색 및 행 편집
- 데이터 테이블에 저장되는 스크립트 생성

그림 4.1 행 및 열 메뉴



목차

데이터 테이블에 데이터 입력 . . . 180
데이터 테이블에 데이터 복사 및 붙여넣기 . . . 180
데이터 테이블에 행 추가 . . . 181
데이터 테이블에 열 추가 . . . 181
순차적 데이터로 데이터 테이블 열 채우기 . . . 183
결측값을 데이터 테이블 열의 이전 값으로 바꾸기 . . . 184
데이터 테이블 열에 셀 계산식 입력 . . . 185
데이터 테이블에서 행 선택 . . . 186
행에서 매칭 셀 찾기 . . . 188
특정 값이 포함된 행 선택 . . . 188
다음 선택 행 찾기 . . . 190
데이터 테이블에서 열 선택 . . . 191
열을 선택하는 방법 . . . 191
JMP 열 보기에서 열 선택 . . . 192
데이터 테이블에서 다음 및 이전 선택 열 찾기 . . . 197
행 및 열 크기 조정 . . . 197
데이터 테이블에서 데이터 구성 . . . 198
데이터 테이블에서 행 삭제 . . . 198
데이터 테이블에서 열 삭제 . . . 199
데이터 테이블에서 열 재배열 . . . 199
데이터 테이블에서 열 그룹화 . . . 199
데이터 테이블에서 값 이동 . . . 200
데이터 테이블에서 셀에 색상 적용 . . . 201
데이터 테이블에서 셀 편집 또는 삭제 . . . 201
데이터 테이블에서 열 이름 편집 . . . 201
데이터 테이블에서 행 숨기고 제외하기 . . . 202
데이터 테이블에서 행 제외 . . . 203
데이터 테이블에서 행 숨기기 . . . 203
데이터 테이블에서 열 제외 . . . 204
데이터 테이블에서 열 숨기기 . . . 205
데이터 테이블에서 결측 데이터의 패턴 보기 . . . 206
데이터 테이블에서 셀 값 찾기 및 바꾸기 . . . 207
데이터 테이블에서 셀 편집 . . . 209
행 및 열에 대한 마우스 오른쪽 버튼 클릭 메뉴 . . . 211
데이터 테이블 비교 . . . 211
행 및 열에 특성 할당 . . . 216
행 및 열에 대한 옵션을 찾는 방법 . . . 216
행 및 열 라벨 지정 . . . 217

행에 색상 또는 표식 할당 218
그래프의 색상 테마 생성 220
그래프에서 사용자 색상 테마 삭제 224
행 상태가 매칭되는 셀 선택 225
행 특성 삭제 225
열 잠그기 225
데이터 재구성 226
데이터 테이블 열을 여러 열로 분할 226
텍스트 매칭으로 새 열 생성 227
표시자 열 생성 228
데이터 테이블에서 열 결합 229
데이터 테이블에서 선택 열 압축 229
데이터 테이블 열에서 계급화 계산식 만들기 231
라벨과 코드 간 변환 235
새 계산식 열 생성 236
JMP 플랫폼의 열 변환 236
열의 데이터 재코딩 245
데이터 테이블 편집 259
데이터 테이블 이름 바꾸기 259
Lock Data Table 259
데이터 테이블 압축 260
테이블 스크립트 복사 260
데이터 테이블 변수 사용 261
데이터 테이블에서 스크립트 생성 및 저장 262

데이터 테이블에 데이터 입력

이 섹션에서는 행 및 열을 추가하고, 열에 순차 데이터를 채우고, 셀 계산식을 입력하는 방법을 설명합니다.

- " 데이터 테이블에 데이터 복사 및 붙여넣기 "
- " 데이터 테이블에 행 추가 "
- " 데이터 테이블에 열 추가 "
- " 순차적 데이터로 데이터 테이블 열 채우기 "
- " 결측값을 데이터 테이블 열의 이전 값으로 바꾸기 "
- " 데이터 테이블 열에 셀 계산식 입력 "

팁 : " 일반 " 환경 설정에서 " 자동 저장 타임아웃 " 값을 설정하여 열려 있는 데이터 테이블을 지정된 시간 (분) 마다 자동으로 저장하는 것이 좋습니다 . 이 자동 저장 값은 저널 , 스크립트 , 프로젝트 및 보고서에도 적용됩니다 .

데이터 테이블에 데이터 복사 및 붙여넣기

다른 문서의 데이터를 데이터 테이블에 붙여넣으면 JMP가 필요한 대로 행과 열을 추가합니다. 셀을 클릭하고 데이터를 붙여넣으십시오.

여러 줄로 된 데이터

데이터 테이블 셀 하나를 선택하고 여러 줄로 된 데이터를 붙여넣으면 각 줄이 개별 셀에 붙여넣어집니다.

모든 줄을 한 셀에 붙여넣으려면 다음 중 하나를 수행하십시오.

- 먼저 기존 셀을 두 번 클릭한 후 붙여넣습니다 .
- 기존 셀을 마우스 오른쪽 버튼으로 클릭하고 **셀에 붙여넣기**를 선택합니다 . " 셀에 붙여넣기 " 옵션은 문자 및 표현식 열에서만 사용할 수 있습니다 .

참고 : 존재하지 않는 셀에 붙여넣으려고 하면 텍스트가 여러 개의 개별 셀에 붙여넣어집니다 .

복사한 내용의 데이터 유형은 붙여넣을 셀의 데이터 유형과 매칭되어야 합니다. 새 열을 추가하여 데이터를 붙여넣거나, 올바른 데이터 유형의 셀에 데이터를 붙여넣거나, 필요한 경우 셀의 데이터 유형을 변경할 수 있습니다.

참고 : 숫자 값 전체를 복사하려면 **편집 > 전체 정밀도로 복사**를 선택하십시오 . 이 옵션은 고정 소수점 형식에만 적용됩니다 .

데이터 테이블에 행 추가

데이터 테이블에 임의 개수의 행을 추가하려면

1. **행 > 행 추가**를 선택합니다 .
2. 추가할 행 수를 입력합니다 .
3. 새 행을 추가할 위치 (데이터 테이블의 시작 또는 끝 부분 , 또는 특정 행 뒤) 를 지정합니다 .
4. **확인**을 클릭합니다 .

데이터 테이블의 끝에 단일 행을 추가하려면

- 마지막 행 아래에 있는 셀의 내부를 클릭하고 입력을 시작합니다 .
- 마지막 행 아래에 있는 빈 행의 행 번호 영역을 두 번 클릭합니다 .

데이터 테이블에 열 추가

다음과 같은 몇 가지 방법으로 데이터 테이블에 새 열을 추가할 수 있습니다 .

- 마지막 데이터 테이블 열의 오른쪽에 있는 빈 공간을 두 번 클릭하여 열 하나를 추가합니다 . 열을 추가한 후 속성을 변경하는 방법에 대한 자세한 내용은 " 열 정보 창 "(269 페이지) 에서 확인하십시오 .
- **열 > 새 열**을 선택합니다 . 하나 이상의 열을 추가하고 열 속성을 변경하려면 이 옵션을 사용하십시오 .
- 데이터 격자에서 행 번호 위에 있는 위쪽 삼각형을 두 번 클릭합니다 . " 여러 열 추가 " 창이 나타납니다 .

참고 :

- 열을 처음 생성할 때 열에 초기 데이터 값을 채우도록 선택할 수 있습니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열의 데이터 초기화 "(279 페이지) 에서 확인하십시오 . 하지만 셀을 수정한 후에는 이 옵션이 더 이상 나타나지 않습니다 .
- 데이터 테이블에 선택된 행이 있는 경우 새 열을 생성하기 위해 마우스 버튼을 두 번 클릭하면 해당 행이 선택 취소됩니다 . " 열 " > " 새 열 " 을 사용하여 새 열을 생성할 경우에는 행 선택 상태가 변경되지 않습니다 .
- 기준 그룹 또는 연결된 부분집합의 열에 있는 데이터는 편집할 수 없습니다 .

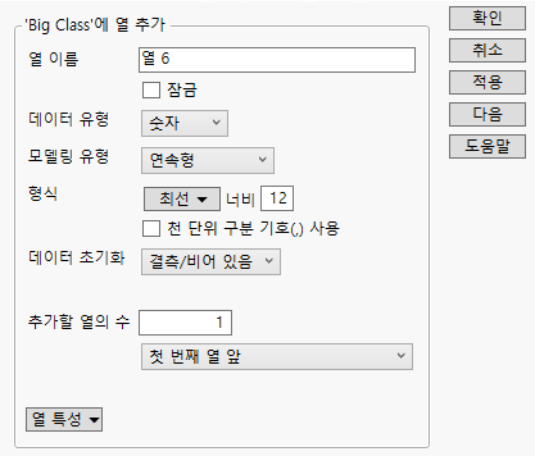
" 새 열 " 을 사용하여 하나 이상의 열을 추가하려면

1. (선택 사항) 하나 이상의 새 열을 추가할 곳의 앞 또는 뒤에 있는 열을 선택합니다 .
2. **열 > 새 열**을 선택합니다 .
3. 열 이름을 변경합니다 .

   기본적으로 새 열의 이름은 열 1, 열 2 등으로 지정됩니다 .

4. 모든 열의 데이터 유형, 모델링 유형 및 형식을 선택합니다. 자세한 내용은 " 열 정보 창 "(269 페이지) 에서 확인하십시오.
5. (선택 사항) 모든 열의 초기 데이터 값을 선택합니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열의 데이터 초기화 "(279 페이지) 에서 확인하십시오.
6. 새 열을 추가할 위치를 지정합니다. 옵션을 선택하지 않으면 해당 열이 마지막 열 뒤에 추가됩니다.
7. (선택 사항) 여러 열을 추가하는 경우 열을 그룹화할지 여부를 지정합니다. 자세한 내용은 " 데이터 테이블에서 열 그룹화 "(199 페이지) 에서 확인하십시오.
8. 열을 더 추가하려면 다음 단계를 수행합니다.
   a. **적용**을 클릭하여 새 열을 추가합니다.
   b. **다음**을 클릭하고, 열 특성을 변경한 후, **적용**을 클릭합니다.
   c. 또 다른 열을 추가하려면 **다음**을 클릭합니다.
9. **확인**을 클릭하여 변경 사항을 저장합니다.

그림 4.2 새 열 창



팁 : 열이 생성된 후에 모델링 유형을 변경하려면 " 열 " 패널에서 모델링 유형을 클릭하고 다른 유형을 선택하십시오.

" 여러 열 추가 " 를 사용하여 하나 이상의 열을 추가하려면

1. 데이터 격자에서 위쪽 삼각형 영역을 두 번 클릭합니다. " 여러 열 추가 " 창이 나타납니다.

그림 4.3 여러 열 추가

2. 각 열의 접두사를 변경합니다 .

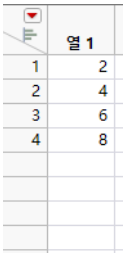
   기본적으로 새 열의 이름은 **열 1, 열 2** 등으로 지정됩니다 .
3. 추가할 열 수를 입력합니다 .
4. 열을 그룹화할지 여부를 지정합니다 . 자세한 내용은 " 데이터 테이블에서 열 그룹화 "(199 페이지) 에서 확인하십시오 .
5. 모든 열의 데이터 유형을 선택합니다 . 자세한 내용은 " 열 정보 창 "(269 페이지) 에서 확인하십시오 .
6. 새 열을 추가할 위치를 지정합니다 .
7. (선택 사항) 모든 열의 초기 데이터 값을 선택합니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 열의 데이터 초기화 "(279 페이지) 에서 확인하십시오 .
8. **확인**을 클릭하여 변경 사항을 저장합니다 .

순차적 데이터로 데이터 테이블 열 채우기

반복적인 데이터 시퀀스 또는 연속되는 값으로 열을 채우려면 다음 단계를 수행하십시오.

1. 열에 데이터 시퀀스를 생성합니다 .

그림 4.4 데이터 시퀀스의 예



	열 1
1	2
2	4
3	6
4	8

2. 연속 데이터가 포함된 셀을 강조 표시합니다. 셀이 서로 다른 열에 있어도 됩니다.
3. 선택한 셀을 마우스 오른쪽 버튼으로 클릭하고 **채우기**에서 옵션을 선택합니다.

채우기 옵션

테이블 끝까지 시퀀스 반복 선택 영역 아래의 셀에 선택한 셀을 반복해서 채우도록 지정합니다.

테이블 끝까지 시퀀스 계속 선택 영역 아래의 셀에 선택한 셀에 있는 패턴을 연속해서 채우도록 지정합니다. 예를 들어 선택한 셀에 숫자 1 과 2 가 포함되어 있으면 나머지 셀에 3, 4, 5, 6 등이 채워집니다. 선택한 셀에 숫자 2 와 4 가 포함되어 있으면 나머지 셀에 6, 8, 10, 12 등이 채워집니다.

시퀀스 반복 선택한 셀에 있는 패턴을 지정한 행 번호까지 반복합니다.

시퀀스 계속 선택한 셀에 있는 패턴을 지정한 행 번호까지 반복합니다.

결측값을 데이터 테이블 열의 이전 값으로 바꾸기

열에 결측값이 포함된 경우 해당 값을 이전 셀의 값으로 바꿀 수 있습니다. 예를 들어 "4, ., ., ., ., 6, ., ., 7, ." 이 선택된 경우 값이 "4, 4, 4, 4, 4, 6, 6, 6, 7, 7" 로 변경됩니다.

1. 결측값과 대체 값이 포함된 셀을 선택합니다.

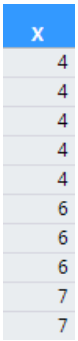
   첫 번째로 선택하는 셀은 결측값 바로 앞에 있어야 합니다. 그 다음에는 인접하지 않은 셀을 선택할 수 있습니다.

그림 4.5 결측값

x
4
•
•
•
•
6
•
•
7
•

2. 마우스 오른쪽 버튼을 클릭하고 **채우기 > 결측값을 이전 값으로 바꾸기**를 선택합니다 .

그림 4.6 바뀐 값



x
4
4
4
4
4
6
6
6
7
7

데이터 테이블 열에 셀 계산식 입력

숫자 데이터 테이블 열에서는 등호(=)를 입력한 후 셀 표현식을 입력할 수 있습니다. JMP는 표현식을 실행한 후 새 숫자를 셀 값으로 저장합니다. 열 계산식과 달리 셀 표현식은 저장되지 않습니다. 셀 표현식에는 연산자, 상수, 전역 변수 및 열 변수가 포함될 수 있습니다.

표현식을 입력하려면

1. 표현식을 입력할 셀을 클릭합니다 .
2. 등호 (=) 를 입력한 후 표현식을 입력합니다 (표 4.1).
3. Enter 키를 누릅니다 .

표 4.1 테이블 셀의 표현식 예

표현식 예	셀 값
=sqrt(2)	1.41
=456+890	1346
=height+weight	이 표현식을 입력한 셀과 동일한 행에 있는 height 및 weight 열의 셀 값을 더합니다.
=height[1]	height 열의 1 행에 있는 값을 표시합니다.

데이터 테이블에서 행 선택

한 행 전체를 선택하려면

- 행 번호가 포함된 빈 공간을 클릭합니다 .

특정 행 번호를 선택하려면

- **행 > 행 선택 > 행으로 이동**을 선택하고 원하는 행 번호를 입력합니다 .

여러 행을 선택하려면

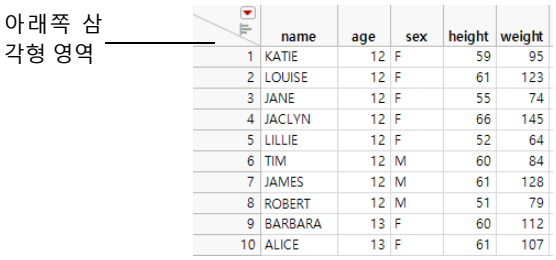
- 연속된 항목을 선택할 경우 :
  - 해당 행 번호 위를 클릭한 채로 드래그합니다 .
  - Shift 키를 누른 채로 원하는 범위의 첫 번째 행과 마지막 행을 클릭합니다 .
  - Shift 키를 누른 채로 위쪽 또는 아래쪽 화살표 키를 누릅니다 .
- 연속되지 않은 항목을 선택할 경우 :
  - Ctrl 키를 누른 채로 각 행을 클릭합니다 .

모든 행을 선택 또는 선택 취소하려면

- 모든 행을 선택하려면 **행 > 행 선택 > 모든 행 선택**을 선택합니다 .

  또는
- Shift 키를 누른 채로 데이터 격자의 왼쪽 상단에 있는 아래쪽 삼각형 영역을 클릭하여 선택합니다 . 동일한 영역을 다시 클릭하면 모든 행이 선택 취소됩니다 (그림 4.7).
- 데이터 테이블에서 강조 표시된 모든 영역을 지우려면 Esc 키를 누릅니다 .

그림 4.7 아래쪽 삼각형 영역



	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107

임의의 행을 선택하려면

1. **행 > 행 선택 > 무작위 선택**을 선택합니다 .
2. 특정 개수의 행 또는 전체 행 수 중 특정 비율을 무작위로 선택할 수 있습니다 .
   - 특정 개수의 행을 선택하려면 해당 정수를 입력합니다 .
   - 특정 비율의 행을 선택하려면 0 에서 1 사이의 값을 입력합니다 .

   예를 들어 10 개의 행을 선택하려면 10 을 입력합니다 . 10% 의 행을 선택하려면 0.1 을 입력합니다 .

행 선택을 반전하려면

- **행 > 행 선택 > 행 선택 반전**을 선택합니다 .

우월 행을 선택하려면

1. **행 > 행 선택 > 우월 행 선택**을 선택합니다 .
2. 우월성을 확인하는 데 사용할 값이 포함된 열을 선택합니다 .
3. 각 열에 대해 우월성을 확인할 기준으로 높은 값을 사용할지 낮은 값을 사용할지 선택합니다.
4. **확인**을 클릭합니다 .

우월 값 및 행에 대해서는 다음을 참고하십시오.

- 사용자가 지정한 사항을 기준으로 다른 값보다 더 높거나 더 낮은 값이 우월 값입니다 .
- " 우월 행 선택 " 옵션은 다른 모든 행보다 우월한 각 행을 선택합니다 . 한 행의 **모든** 값이 다른 행의 값보다 우월한 경우에만 해당 행을 다른 행보다 우월하다고 합니다 .
- 결과 행 집합은 파레토 프론티어라고 합니다 .

현재 선택된 행을 새 열에 저장하려면

1. **행 > 행 선택 > 선택된 행 구분 열 만들기**를 선택합니다 .
2. 열 이름을 입력합니다 .
3. 선택된 행과 선택 취소된 행의 라벨을 지정합니다 .
4. **확인**을 클릭합니다 .

팁 : 새 열을 생성한 후 이 과정을 반복하여 동일한 열 이름을 지정하면 원래 열을 덮어쓰게 됩니다 . 데이터 테이블에서 " 선택됨 " 또는 " 선택 취소됨 " 값에 대해 변경한 내용은 손실됩니다 . 이러한 값을 덮어쓰지 않도록 하려면 " 선택된 행 구분 열 만들기 " 창에서 해당하는 상자를 지우십시오 . 예를 들어 " 선택 취소됨 " 값을 덮어쓰지 않도록 하려면 " 선택 취소됨 " 상자를 지워야 합니다 .

제외된 행 , 숨겨진 행 또는 라벨이 지정된 행을 선택하려면

1. **행 > 행 선택**을 선택합니다 .
2. 다음 옵션 중에서 선택합니다 .
   - 제외 항목 선택
   - 숨김 항목 선택
   - 라벨 항목 선택

참고 : 제외된 행 , 숨겨진 행 또는 라벨이 지정된 행에 대한 자세한 내용은 " 행 및 열에 특성 할당 "(216 페이지) 에서 확인하십시오 .

중복 값이 포함된 행을 선택하려면

1. **행 > 행 선택 > 중복 행 선택**을 선택합니다 .

   첫 번째 행을 제외하고 모든 중복 행이 선택됩니다 . 선택된 열이 없으면 모든 열을 사용하여 중복 행이 선택됩니다 .

   선택된 열이 있으면 해당 열의 행에서 중복 값을 찾습니다 .

   매칭 시 대 / 소문자가 구분됩니다 .

행에서 매칭 셀 찾기

"매칭 셀 선택" 옵션을 사용하여 매칭 값이 있는 셀을 찾을 수 있습니다.

- **행 > 행 선택 > 매칭 셀 선택**을 선택합니다 .

  또는

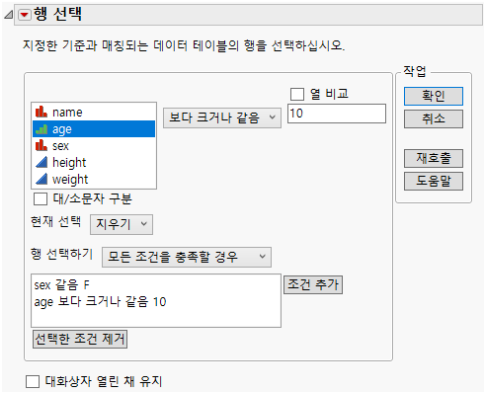
- 강조 표시된 행 번호 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **매칭 셀 선택**을 선택합니다.

특정 값이 포함된 행 선택

JMP에서는 특정 값(또는 텍스트 문자열)을 검색하고 데이터 테이블에서 해당 특정 값이 포함된 모든 셀을 강조 표시할 수 있습니다.

1. **행 > 행 선택 > 선택 조건**을 선택합니다 .

그림 4.8 행 선택 기준 지정



2. 열 목록에서 선택하려는 행이 있는 열의 이름을 강조 표시합니다 .
3. 드롭다운 메뉴를 사용하여 목록의 조건 (같음 , 같지 않음 등) 을 선택합니다 . 그림 4.8 을 참조하십시오 .
4. 검색 값을 입력합니다 .
5. **확인**을 클릭합니다 .

다음과 같은 선택적 기능을 지정할 수도 있습니다 .

- 두 열의 값을 비교하려면 **열 비교** 체크박스를 클릭합니다 . 그런 다음 열 목록에서 비교할 열을 선택합니다 .
- 대 / 소문자를 구분하여 검색하려면 **대 / 소문자 구분** 옆의 상자를 클릭합니다 .
- " 행 선택하기 " 창의 Where 절을 클립보드 , 데이터 테이블 또는 스크립트 창에 저장할 수 있습니다 . **대화상자 열린 채 유지**를 선택하고 , 행과 기타 옵션을 선택한 다음 , " 행 선택 " 의 빨간색 삼각형 메뉴에서 " 스크립트 저장 " 옵션 중 하나를 선택합니다 . 다음과 같이 쿼리를 반환하는 JSL 을 스크립트에 추가합니다 .

```
dt = Open("$SAMPLE_DATA/Big Class.jmp");
Current Data Table() << Row Selection(
   Select where( :age == 15 ),
   Dialog( Edit( Source Column( :age ) == "" ), Keep dialog open )
);
```

현재 데이터 테이블에서 행이 선택된 경우 다음 기능을 지정할 수 있습니다 .

- **현재 선택**의 옵션을 클릭하여 현재 선택 항목을 처리할 방법을 지정합니다 .
  - **지우기** : 현재 선택된 행의 강조 표시를 제거하고 지정된 값이 포함된 모든 행을 선택합니다 .
  - **확장** : 현재 선택된 행을 선택 상태로 유지하고 지정된 값이 있는 행도 선택합니다 .
  - **제한** : 현재 선택된 배열에서 지정된 값이 포함된 행을 선택합니다 .
- **조건 추가**를 클릭하여 목록에 조건을 추가합니다 .

- 검색 조건을 추가하려면 앞의 단계를 반복합니다 . **행 선택하기** 영역에서 적절한 항목을 클릭하여 **모든** 조건을 충족하는 행을 선택할지 , 아니면 **임의의** 조건을 충족하는 행을 선택할지 지정합니다 .

다음 선택 행 찾기

행 > 다음 선택 항목을 선택하면 현재 행 뒤에 있는 다음 선택 행을 찾아 깜박이게 할 수 있습니다.

행 > 다음 선택 항목을 선택할 때마다 다음 선택 행이 검색되어 깜박입니다. 마지막 선택 행에 도달하면 신호음이 울립니다.

큰 데이터 집합에서 행이 간헐적으로 선택되어 있을 때 데이터 테이블의 선택된 행을 찾으려면 이 기능을 사용하면 됩니다.

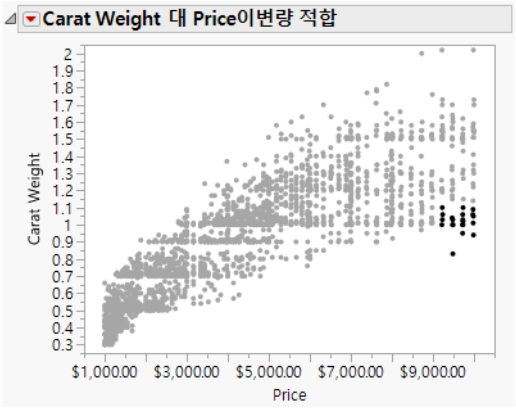
예 : 다음 선택 행 찾기

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diamonds Data.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Carat Weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Price 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. **도구 > 올가미**를 선택합니다 .

   메뉴 표시줄이 보이지 않는 경우 제목 표시줄 아래의 파란색 막대를 마우스로 가리키면 메뉴 표시줄이 표시됩니다 .
7. 그림 하단에서 10,000 달러 주변의 일부 점을 올가미로 묶습니다 (그림 4.9).
8. 데이터 테이블에서 **행 > 다음 선택 항목**을 선택합니다 . 또는 F7 키를 눌러도 됩니다 .

선택된 행 사이를 쉽게 이동하며 각 행의 데이터를 볼 수 있습니다.

그림 4.9 선택된 점



데이터 테이블에서 열 선택

다음과 같은 몇 가지 방법으로 열을 선택할 수 있습니다.

- 데이터 테이블 자체에서 열을 선택합니다. 자세한 내용은 " 열을 선택하는 방법 "(191 페이지) 에서 확인하십시오.
- 많은 열이 있는 데이터 테이블의 경우, " 열 보기 " 에서 속성, 특성 및 통계량을 기준으로 열을 선택합니다. 자세한 내용은 "JMP 열 보기에서 열 선택 "(192 페이지) 에서 확인하십시오.
- 다음 또는 이전 선택 열을 찾습니다. 자세한 내용은 " 데이터 테이블에서 다음 및 이전 선택 열 찾기 "(197 페이지) 에서 확인하십시오.

열을 선택하는 방법

데이터 테이블에서 열을 선택하는 여러 방법 중 하나를 선택합니다.

한 열 전체를 선택하려면

- 데이터 격자에서 열 이름 주변의 빈 공간을 클릭합니다.

  또는

- " 열 " 패널에서 열 이름을 클릭합니다.

특정 열 번호를 선택하려면

1. **열 > 열 선택 > 이동**을 선택합니다.
2. 열 번호 또는 이름을 입력하고 **확인**을 클릭합니다.

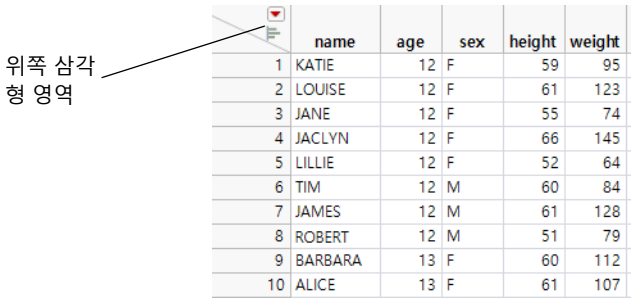
여러 열을 선택하려면

- 연속된 항목을 선택할 경우 :
  - 해당 열 이름 위를 클릭한 채로 드래그합니다.
  - Shift 키를 누른 채로 원하는 범위의 첫 번째 열과 마지막 열을 클릭합니다.
  - Shift 키를 누른 채로 왼쪽 또는 오른쪽 화살표 키를 누릅니다.
- 서로 인접하지 않은 열을 선택할 경우 :
  - Ctrl 키를 누른 채로 각 열을 클릭합니다.

모든 열을 선택 또는 선택 취소하려면

- Shift 키를 누른 채로 데이터 격자의 왼쪽 상단에 있는 위쪽 삼각형 영역을 클릭하여 선택합니다. 동일한 영역을 다시 클릭하면 모든 열이 선택 취소됩니다.

그림 4.10 위쪽 삼각형 영역



	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107

팁 : 데이터 테이블에서 강조 표시된 모든 영역을 지우려면 Esc 키를 누릅니다 .

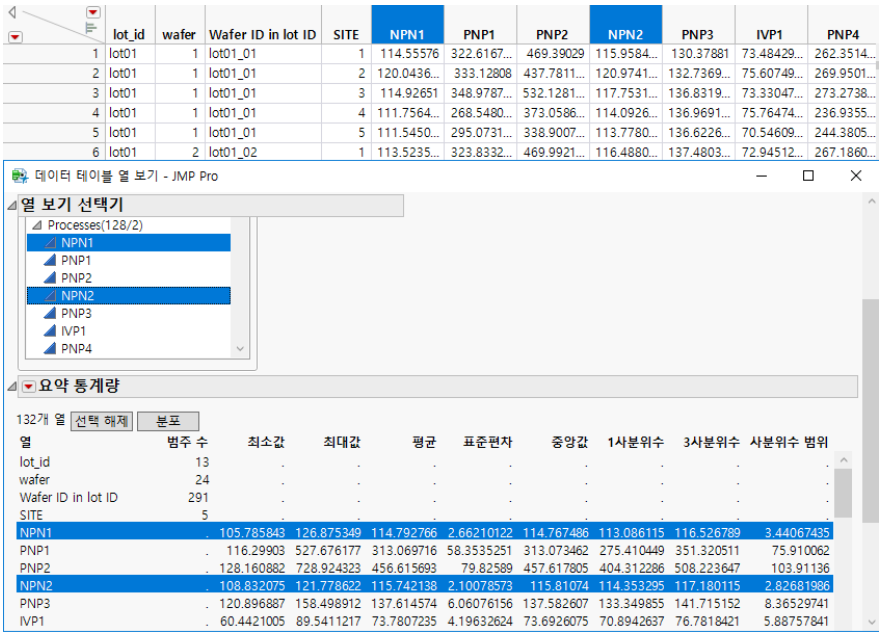
열 선택을 반전하려면

열 > 열 선택 > 열 선택 반전을 선택합니다. 현재 선택 취소되어 있는 모든 열을 선택하고, 현재 선택되어 있는 모든 열을 선택 취소합니다.

JMP 열 보기에서 열 선택

특히 많은 열이 있는 데이터 테이블의 경우, "열" 메뉴의 "열 보기" 옵션을 사용하면 속성, 특성 및 통계량을 기준으로 열을 빠르게 선택할 수 있습니다. 해당 열의 요약 통계량 및 특성을 보고, 요약 통계량의 사분위수를 보고, 데이터 부분집합을 추출하는 등의 작업을 수행할 수 있습니다. 또한 "열 보기" 창의 열은 데이터 테이블 열에 연결되어 있습니다.

그림 4.11 열 보기의 연결된 열



"열 보기"에서는 데이터 테이블 특성을 빠르게 확인할 수 있습니다. 예를 들어 "요약 통계량" 보고서에는 결측값이 포함된 열이 표시됩니다. 보고서에서 해당 열을 선택한 후 데이터 테이블에서 이를 제외할 수 있습니다.

그림 4.12 결측값 확인

"요약 통계량" 보고서에는 다음과 같은 정보가 표시됩니다.

- 총 행 수 (N)
- 결측값이 포함된 행의 수 (결측값 수)
- 범주 수 (범주 수)
- 연속형 데이터의 경우 , 최소값 , 최대값 , 평균 및 표준편차

그 외에도 다음과 같은 옵션이 있습니다.

선택 해제 데이터 테이블과 " 열 보기 " 의 열을 선택 취소합니다 . 이 옵션을 사용하면 열 선택을 시작하기 전에 아무 열도 선택되어 있지 않도록 할 수 있습니다 .

부분집합 선택된 열을 사용하여 연결된 부분집합 데이터 테이블을 생성합니다 .

요약 표시 선택된 열에 대해 연결된 " 요약 통계량 " 보고서를 생성합니다 . 마우스 오른쪽 버튼을 클릭하면 열별 정렬이나 데이터 테이블 생성 등의 옵션을 선택할 수 있습니다 . 1 사분위수 , 3 사분위수 및 사분위수 범위를 포함하려면 **사분위수 표시**를 선택합니다 . 또한 보고서의 모든 열을 사용하여 연결된 데이터 테이블을 생성하려면 " 요약 통계량 " 의 빨간색 삼각형을 클릭하고 " 데이터 테이블 보기 " 를 선택합니다 .

특성을 사용하여 열 찾기 " 특성을 포함하는 열 " 보고서에 열 특성의 목록을 표시합니다 . 찾으려는 특성을 선택한 후 **확인**을 클릭하면 모든 열을 사용하여 연결된 보고서가 생성됩니다 . 또는 먼저 " 열 선택 " 목록에서 열을 선택한 후 해당 열에만 해당되는 특성의 목록을 표시할 수도 있습니다 .

참고 :

- **요약 표시** 또는 **특성을 사용하여 열 찾기** 버튼을 클릭할 때마다 창에 새 보고서가 추가됩니다 . 보고서를 삭제하려면 보고서의 빨간색 삼각형 메뉴에서 **제거**를 선택하십시오 .
- 열 보기를 사용하여 생성된 데이터 테이블 보기에서 행을 삭제할 경우 삭제를 실행 취소할 수 없습니다 .

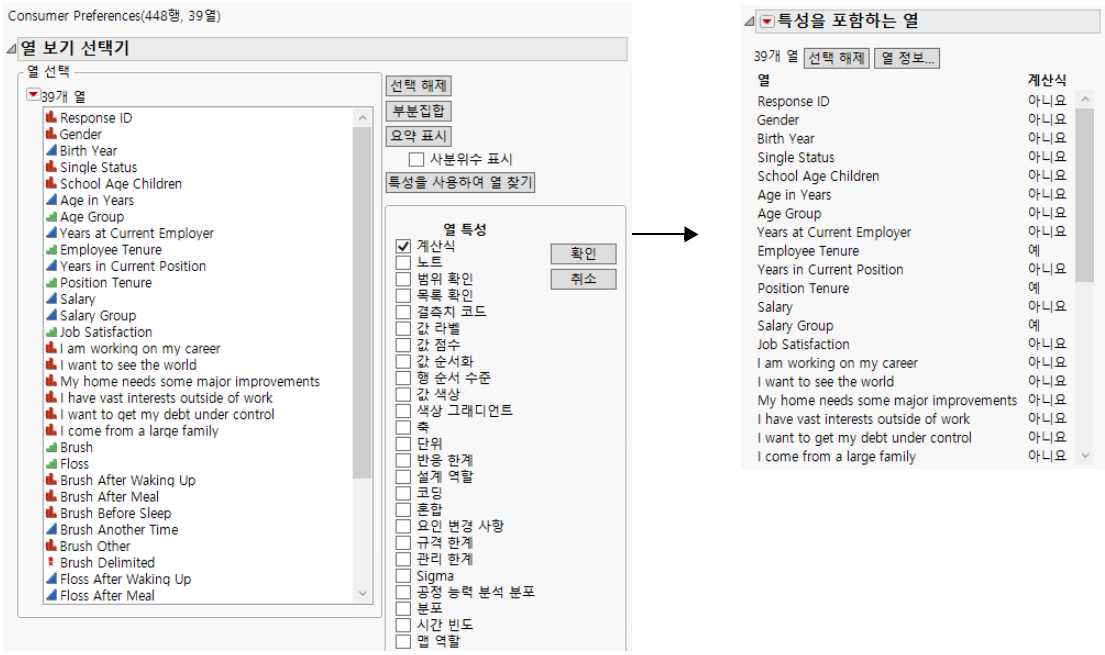
예 : 특정 특성이 있는 열 찾기

이 예에서는 "계산식" 특성이 있는 열을 찾은 후 모든 계산식을 한 번에 보는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Consumer Preferences.jmp 를 엽니다 .
2. **열 > 열 보기**를 선택하여 " 데이터 테이블 열 보기 " 창을 엽니다 .
3. **특성을 사용하여 열 찾기**를 선택하고 **계산식**을 선택한 후 **확인**을 클릭합니다 .

   " 특성을 포함하는 열 " 보고서가 나타납니다 . 일부 열에 " 계산식 " 특성이 포함되어 있습니다 . 이 목록이 너무 길어서 모든 계산식 열을 함께 보려고 합니다 .

그림 4.13 계산식 열 특성 선택



4. 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열별 정렬**, **계산식**을 선택한 후 **확인**을 클릭합니다. " 계산식 " 특성이 있는 열이 보고서의 맨 위에 나타납니다 .

그림 4.14 열별 정렬

5. Employee Tenure, Position Tenure 및 Salary Group 열을 선택하고 **열 정보**를 선택합니다 .

선택한 열의 계산식이 데이터 테이블의 " 열 설정 " 창에 나타납니다 .

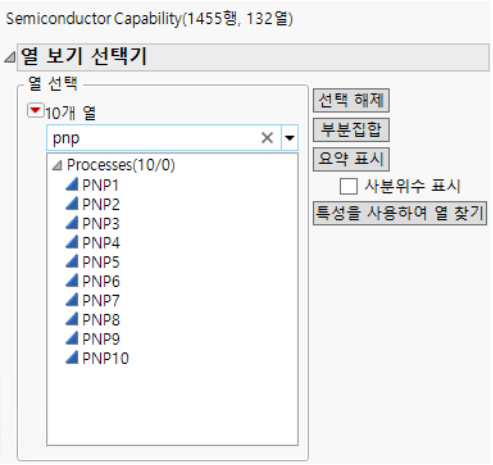
예 : 요약 통계량 표시

이 예에서는 표준편차가 낮은 열을 찾는 방법을 보여 줍니다. 해당 데이터를 삭제하거나 분석에서 제외하려는 경우 이 기능을 사용하면 유용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Semiconductor Capability.jmp 를 엽니다 .
2. **열 > 열 보기**를 선택하여 " 데이터 테이블 열 보기 " 창을 엽니다 .
3. 상자에 PNP 를 입력하고 Enter 키를 누릅니다 .

   그룹화된 processes 가 표시됩니다 .
4. processes 그룹을 펼쳐 PNP 열을 표시합니다 .

그림 4.15 이름별로 열 필터링



5. **요약 표시**를 클릭하여 " 요약 통계량 " 보고서를 추가합니다 .

   각 열의 최소값 , 최대값 , 평균 및 표준편차가 여러 행에 표시됩니다 .

그림 4.16 선택한 열에 대한 요약 통계량

요약 통계량

10개 열 선택 해제 분포

열	최소	최대	평균	표준편차
PNP1	116.29903	527.676177	313.069716	58.3535251
PNP2	128.160882	728.924323	456.615693	79.82589
PNP3	120.896887	158.498912	137.614574	6.06076156
PNP4	149.453273	373.172883	256.375597	32.6073816
PNP5	-55.762191	-40.722372	-47.178172	2.047573
PNP6	0	0	0	0
PNP7	-16.888557	-14.961357	-15.89982	0.28816317
PNP8	-44.913142	-34.771668	-39.21972	1.49081788
PNP9	-79.665058	-66.929598	-73.35004	1.94997812
PNP10	-0.6442798	-0.623852	-0.6335016	0.00299937

6. 보고서 내부를 마우스 오른쪽 버튼으로 클릭하고 **열별 정렬**을 선택합니다 .
7. **표준편차** 및 **오름차순**을 선택한 후 **확인**을 클릭합니다 .

   PNP6 는 최소값 , 최대값 및 평균 값이 0 이므로 표준편차가 없습니다 .

그림 4.17 정렬된 표준편차 열

요약 통계량

10개 열 선택 해제 분포

열	최소	최대	평균	표준편차
PNP6	0	0	0	0
PNP10	-0.6442798	-0.623852	-0.6335016	0.00299937
PNP7	-16.888557	-14.961357	-15.89982	0.28816317
PNP8	-44.913142	-34.771668	-39.21972	1.49081788
PNP9	-79.665058	-66.929598	-73.35004	1.94997812
PNP5	-55.762191	-40.722372	-47.178172	2.047573
PNP3	120.896887	158.498912	137.614574	6.06076156
PNP4	149.453273	373.172883	256.375597	32.6073816
PNP1	116.29903	527.676177	313.069716	58.3535251
PNP2	128.160882	728.924323	456.615693	79.82589

8. " 요약 통계량 " 보고서에서 PNP6 의 행을 선택한 후 데이터 테이블을 표시합니다 .
9. 데이터 테이블을 보고 삭제를 눌러 선택한 열을 제거합니다 .

   해당 열이 데이터 테이블에서 즉시 제거됩니다 .
10. " 열 보기 " 를 닫으려면 창의 오른쪽 상단 (Windows) 또는 왼쪽 상단 (macOS) 에 있는 X 버튼을 클릭합니다 .

데이터 테이블에서 다음 및 이전 선택 열 찾기

열 > 열 선택 > 다음 선택 열을 선택하면 현재 열 뒤에 있는 다음 선택 열을 찾을 수 있습니다. 마찬가지로, **열 > 열 선택 > 이전 선택 열**을 선택하면 현재 열 앞에 있는 이전 선택 열을 찾을 수 있습니다.

이러한 옵션 중 하나를 선택할 때마다 다음 또는 이전 선택 열이 표시되고 깜박입니다. 이러한 옵션은 선택된 열이 있는 경우에만 사용할 수 있습니다.

이 기능을 사용하여 큰 데이터 테이블에서 간헐적으로 선택된 열을 확인할 수 있습니다.

행 및 열 크기 조정

행 크기를 조정하려면

행 크기를 조정하려면 행 테두리 중 하나를 드래그합니다. 그러면 모든 행이 동일한 높이로 조정됩니다. 각 셀 안에 나타나는 그래픽은 행 높이에 따라 축소됩니다.

열 크기를 조정하려면

열 크기를 조정하려면 열의 오른쪽 또는 왼쪽 테두리를 드래그합니다.

여러 열을 동일한 너비로 조정하려면 다음 단계를 수행하십시오.

1. 크기를 조정할 열을 선택합니다.
2. Alt 키 (Windows) 또는 Option 키 (macOS) 를 누른 채로 열 테두리 중 하나를 드래그합니다. 그러면 모든 열이 동일한 너비로 조정됩니다.

데이터 테이블에서 데이터 구성

여러 방법 중 하나를 선택하여 데이터 테이블에서 데이터를 재배열합니다.

- "데이터 테이블에서 행 삭제"
- "데이터 테이블에서 열 삭제"
- "데이터 테이블에서 열 재배열"
- "데이터 테이블에서 열 그룹화"
- "데이터 테이블에서 값 이동"
- "데이터 테이블에서 셀에 색상 적용"
- "데이터 테이블에서 셀 편집 또는 삭제"
- "데이터 테이블에서 열 이름 편집"
- "데이터 테이블에서 행 숨기고 제외하기"
- "데이터 테이블에서 행 제외"
- "데이터 테이블에서 행 숨기기"
- "데이터 테이블에서 열 제외"
- "데이터 테이블에서 열 숨기기"
- "데이터 테이블에서 결측 데이터의 패턴 보기"
- "데이터 테이블에서 셀 값 찾기 및 바꾸기"
- "데이터 테이블에서 셀 편집"
- "행 및 열에 대한 마우스 오른쪽 버튼 클릭 메뉴"
- "데이터 테이블 비교"

데이터 테이블에서 행 삭제

JMP 데이터 테이블에서 행을 삭제하려면 다음 단계를 수행하십시오.

1. 삭제할 행을 강조 표시합니다.
2. Delete 키를 누르거나, 행 번호를 마우스 오른쪽 버튼으로 클릭하고 **행 삭제**를 선택합니다.

주의: 수천 개의 행을 삭제하려는 경우 컴퓨터에 실행 취소를 위해 데이터를 저장하기 위한 메모리가 충분하지 않으면 경고가 나타날 수 있습니다. 삭제할 행을 더 적게 선택하거나, 테이블 패널의 빨간색 삼각형 메뉴에서 **실행 취소 비활성화**를 선택하십시오. 이 옵션은 실행 취소 기록에서 모든 작업을 제거하고 더 이상 작업을 기록하지 않습니다. **실행 취소 비활성화** 옵션을 선택할 경우 해당 옵션은 데이터 테이블이 열려 있는 동안에만 적용되며, 설정이 데이터 테이블과 함께 저장되지는 않습니다.

데이터 테이블에서 열 삭제

JMP 데이터 테이블에서 열을 삭제하려면 다음 단계를 수행하십시오.

1. 삭제할 열을 강조 표시합니다.
2. Delete 키를 누르거나, 마우스 오른쪽 버튼을 클릭하고 **열 삭제**를 선택합니다.

데이터 테이블에서 열 재배열

데이터 테이블 열을 열 이름, 데이터 유형 또는 모델링 유형을 기준으로 재배열 또는 정렬하거나, 현재 순서의 역순으로 바꿀 수 있습니다. 열을 재정렬하려면 **열 > 열 재정렬**을 선택하고 다음 옵션 중 하나를 선택합니다.

선택한 열 이동 선택한 열을 데이터 테이블의 특정 위치로 이동합니다. " 선택한 열 이동 " 창에서 선택한 열을 배치할 위치를 지정하십시오.

처음으로 선택한 열을 데이터 테이블의 맨 왼쪽 위치에 놓이도록 이동합니다.

마지막으로 선택한 열을 데이터 테이블의 맨 오른쪽 위치에 놓이도록 이동합니다.

다음 위치 뒤 선택한 열을 특정 열 뒤에 놓이도록 이동합니다.

원래 순서 열을 데이터 테이블을 마지막으로 저장했을 때의 순서로 되돌립니다.

열 이름에 따라 재정렬 열을 열 이름에 따라 사전순으로 왼쪽에서 오른쪽으로 배열합니다.

데이터 유형에 따라 재정렬 열을 데이터 유형 (표현식, 행 상태, 문자, 숫자) 에 따라 사전순으로 왼쪽에서 오른쪽으로 배열합니다.

모델링 유형에 따라 재정렬 열을 모델링 유형 (없음, 연속형, 순서형, 명목형, 다중 반응, 비정형 텍스트, 벡터) 에 따라 사전순으로 왼쪽에서 오른쪽으로 배열합니다. 행 상태 열은 모델링 유형이 없으므로 마지막에 표시됩니다.

역 순서 데이터 테이블 열의 순서를 역순으로 바꿉니다.

데이터 테이블에서 열 그룹화

많은 수의 열을 관리하고 분석 역할 지정을 쉽게 하기 위해 여러 열을 단일 머리글 내에 그룹화합니다. 그룹화된 열은 "열" 패널 내에 개요 형식으로 나타납니다.

열을 그룹화하거나 그룹 해제하려면

1. 데이터 격자 내에서 그룹화할 열을 선택합니다 .
2. 메인 메뉴에서 **열 > 열 그룹화** 또는 **열 > 열 그룹 해제**를 선택합니다 .

또는

1. " 열 " 패널에서 그룹화할 열을 선택합니다 .
2. 선택된 열을 마우스 오른쪽 버튼으로 클릭하고 **열 그룹화** 또는 **열 그룹 해제**를 선택합니다 .

참고 : 그룹화된 열은 **부분집합** , **정렬** , **요약** , **결합** , **쌓기** 및 **분할** 명령으로 생성된 데이터 테이블에서도 자동으로 유지됩니다 . **쌓기** 명령의 경우 , 쌓기 그룹의 모든 열이 동일한 열 그룹에 속해 있으면 해당 열 이름에 그룹 이름이 사용됩니다 .

데이터 테이블에서 값 이동

데이터 테이블에서 값을 이동하려면 값을 선택한 후 클릭한 채로 잠시 멈췄다가 드래그하여 새 위치에 놓습니다.

팁 : 선택 항목을 클릭한 후 잠시 멈추지 않고 바로 드래그하면 선택 영역이 확장됩니다 .

값을 드래그하여 놓을 때는 다음 사항에 유의하십시오.

- 셀의 모든 특성과 열 특성은 그대로 유지됩니다 .
- 셀을 이동한 후 처음 선택한 셀에는 결측값이 나타납니다 .
- 선택한 셀과 대상 셀의 데이터 유형은 동일해야 합니다 .
- 여러 셀을 테이블의 빈 영역으로 드래그하면 자동으로 새 열이 생성됩니다 .
- 새 열에는 원래 열의 표시 형식과 모델링 유형이 사용됩니다 .

행을 이동할 위치를 지정하려면

1. 이동할 행을 강조 표시합니다 .
2. **행 > 행 이동**을 선택합니다 .
3. " 행 이동 " 창에서 이동하려는 위치를 지정합니다 .
   - 테이블의 처음으로 이동 (**시작**)
   - 테이블의 끝으로 이동 (**끝**)
   - 특정 행 번호 뒤로 이동 (**다음 행 뒤 :**)

내용을 다른 JMP 창으로 이동

Windows에서는 선택한 내용을 최소화된 창 위로 드래그할 수 있습니다. 그러면 최소화된 창이 맨 앞으로 이동하므로 여기에 내용을 붙여넣을 수 있습니다. JMP에서도 동일한 작업을 수행할

수 있습니다. 예를 들어 선택한 내용을 대부분의 창에서 오른쪽 하단에 있는 홈 창 버튼 위로 드래그할 수 있습니다. 그런 다음 "창 목록"에서 해당 내용을 이동하려는 대상 창 위로 드래그합니다. 그러면 해당 창이 맨 앞으로 이동하므로 여기에 내용을 놓을 수 있습니다.

팁 : JMP 홈 창 버튼이 표시되지 않으면 **보기 > 상태 표시줄**을 선택합니다 .

예를 들어 데이터 테이블에서 선택한 열, 행 또는 셀을 다른 데이터 테이블로 드래그하거나, 스크립트 창에서 선택한 텍스트를 다른 스크립트 창으로 드래그하거나, 보고서에서 선택한 내용을 저널로 드래그할 수 있습니다.

데이터 테이블에서 셀에 색상 적용

하나 이상의 데이터 테이블 셀을 선택한 후 색상을 변경하여 특정 값을 강조 표시할 수 있습니다. 하나 이상의 셀을 마우스 오른쪽 버튼으로 클릭하고 **셀에 색상 적용**을 선택한 후 색상을 선택합니다. 셀에서 색상을 제거하려면 셀을 선택하고 마우스 오른쪽 버튼을 클릭한 다음 **열 > 색상 지우기**를 선택합니다.

데이터 테이블에서 셀 편집 또는 삭제

셀의 내용을 편집하거나 삭제하려면 다음 단계를 수행하십시오 .

1. 편집하거나 삭제하려는 값이 포함된 셀을 클릭합니다 .
2. Delete 키를 누릅니다 .
3. 값을 편집하려면 셀을 두 번 클릭한 후 셀 값을 편집합니다 .

데이터 테이블에서 열 이름 편집

단일 열 이름을 편집하려면 열을 선택하고 입력을 시작합니다. 열 정보 창에서 열 이름을 편집하거나, 머리글을 선택하고 Enter 키를 눌러도 됩니다.

모든 열 이름을 한 번에 편집하려면 **열 > 열 이름**을 클릭하고 다음 옵션 중 하나를 선택합니다.

위로 이동 데이터의 첫 번째 행을 위로 이동하여 열 이름을 바꿉니다 .

위로 이동 및 추가 데이터의 첫 번째 행을 위로 이동하고 기존 열 이름에 항목을 추가합니다 .

아래로 이동 데이터 행이 되도록 열 이름을 아래로 이동합니다. 그러면 열 이름은 열 1, 열 2 등으로 지정됩니다 .

열 이름 재코딩 ... 열 이름을 재코딩하기 위해 " 재코딩 " 창을 엽니다 . 자세한 내용은 " 열의 데이터 재코딩 "(245 페이지) 에서 확인하십시오 .

데이터 테이블에서 행 숨기고 제외하기

분석 또는 그림 계산에 포함하지 않고 그림에 표시하지 않으려는 행에는 숨김 및 제외됨 행 상태를 적용하십시오 . 이러한 행은 이후 분석에 포함되지 않습니다 .

참고 :

- " 프로파일러 " 를 제외하고 " 그래프 " 메뉴에 있는 모든 그림은 즉시 다시 계산되어 숨겨지고 제외된 행을 반영합니다 .
- 많은 플랫폼의 경우 , 제외되고 숨겨진 관측값을 반영하기 위해 보고서와 그림이 즉시 업데이트되지 않습니다 . 사용자가 분석을 다시 실행하여 분석 결과와 관련 그림을 다시 계산해야 합니다 . 예외에 대한 자세한 내용은 " 숨기고 제외하기 적용 시 즉시 업데이트되는 플랫폼 "(202 페이지) 에서 확인하십시오 .
- 행을 제외하고 숨기면 해당 행은 현재 열려 있고 자동으로 업데이트되는 모든 보고서 및 그림에서 제외되고 숨겨집니다 .
- 행을 제외하고 숨기면 행 번호 옆에 취소선이 있는 원 모양 아이콘 (, 제외됨을 나타냄) 과 마스크 아이콘 (, 숨겨짐을 나타냄) 이 나타납니다 .
- 해당 행을 강조 표시하고 **숨기고 제외하기**를 다시 선택하기 전까지는 숨김 및 제외 상태가 유지됩니다 .

하나 이상의 행을 숨기고 제외하려면

1. 숨기고 제외할 행을 선택합니다 .
2. 다음 중 하나를 수행합니다 .
   - 행 번호 옆의 강조 표시된 영역을 마우스 오른쪽 버튼으로 클릭하고 **숨기고 제외하기**를 선택합니다 .
   - **행** 메뉴에서 **숨기고 제외하기**를 선택합니다 .

숨기고 제외하기 적용 시 즉시 업데이트되는 플랫폼

행에 "숨기고 제외하기"를 적용할 경우 "분석" 메뉴에 있는 대부분의 플랫폼에서는 일반적으로 해당 행에 직접 대응되는 점 또는 그림 요소가 그림에서 즉시 숨겨집니다. 하지만 포함된 보고서에 표시된 계산은 일반적으로 업데이트되지 **않습니다**. 다음 플랫폼의 경우, "숨기고 제외하기"를 적용하면 그림 **및** 계산이 즉시 업데이트됩니다.

- 테이블 생성
- 관리도 빌더

데이터 테이블에서 행 제외

이후 계산에 포함하지 않으려는 행에는 제외됨 행 상태를 적용하십시오 . 그러면 제외된 행을 포함하지 않고 통계량이 다시 계산됩니다 .

하나 이상의 행을 제외하려면 다음 단계를 수행하십시오 .

1. 제외할 행을 선택합니다 .
2. 다음 중 하나를 수행합니다 .
   - 행 번호 옆의 강조 표시된 영역을 마우스 오른쪽 버튼으로 클릭하고 **제외 / 제외 해제**를 선택합니다 .
   - **행** 메뉴에서 **제외 / 제외 해제**를 선택합니다 .

참고 :

- 대부분의 경우 하나의 점 또는 개별 그림 요소가 하나의 행을 나타내면 해당 행은 제외되더라도 그림에서 숨겨지지 않습니다 .
- 그래프 빌더에서는 하나의 행 그룹이 하나의 그림 요소를 나타내면 제외되지 않은 행만 사용하여 그림이 다시 생성됩니다 . 또한 그림에 제외된 행에 대응하는 그림 요소가 표시되지 않습니다 . 제외된 행은 숨겨진 동시에 제외된 것처럼 처리됩니다 .
- " 프로파일러 ," 평행 그림 " 및 " 산점도 행렬 " 을 제외하고 " 그래프 " 메뉴에 있는 모든 그림은 즉시 다시 계산되어 숨겨지고 제외된 행을 반영합니다 .
- 많은 플랫폼의 경우, 제외된 관측값을 반영하기 위해 보고서와 그림이 즉시 업데이트되지 않습니다 . 사용자가 분석을 다시 실행하여 그림과 분석 결과를 다시 계산해야 합니다 .
- 행을 제외하면 해당 행은 현재 열려 있고 즉시 업데이트되는 모든 그림 및 보고서에서 제외됩니다 .
- 제외된 행의 행 번호 옆에는 취소선이 있는 원 모양 아이콘 ( ) 이 나타납니다 .
- 해당 행을 강조 표시하고 **제외/제외 해제**를 다시 선택하기 전까지는 제외 상태가 유지됩니다.

데이터 테이블에서 행 숨기기

그림에 포함하지 않으려는 행에는 숨김 행 상태를 적용하십시오 .

참고 : 일부 그림에서는 숨겨진 행에 대응하는 요소가 숨겨지지 않습니다 . 그림 요소를 생성하는 데 사용된 계산에서 숨겨진 행의 데이터를 사용하는 경우가 그러한 경우에 해당합니다 . 일반적으로 행과 그림 요소 간에 일대일 대응 관계가 있는 경우에는 숨겨진 행에 대응하는 점 또는 요소가 숨겨집니다 .

하나 이상의 행을 숨기려면 다음 단계를 수행하십시오 .

1. 숨길 행을 선택합니다 .
2. 다음 중 하나를 수행합니다 .

– 행 번호 옆의 선택 영역을 마우스 오른쪽 버튼으로 클릭하고 **숨기기 / 숨기기 해제**를 선택합니다 .
– **행** 메뉴에서 **숨기기 / 숨기기 해제**를 선택합니다 .

참고 :

- 행을 숨기면 해당 행은 현재 열려 있고 즉시 업데이트되는 모든 그림에서 숨겨집니다 .
- 행을 숨겨도 해당 행이 계산 및 분석에서 제외되지는 **않습니다** .
- 숨겨진 행이 각각 하나의 점 또는 개별 그림 요소에 대응하는 경우 행에 숨김 행 상태를 적용하면 그림이 즉시 업데이트됩니다 .
- 플랫폼에 따라서는 그림 요소가 여러 관측값을 사용하여 계산되고, 하나의 행에 직접 대응하지 않을 수도 있습니다 . 이러한 경우에는 일반적으로 행을 숨기는 동시에 제외해야만 그림 요소가 업데이트되어 숨겨진 관측값을 반영합니다 .
- 그림 요소가 행 그룹에 대응하는 일부 그림의 경우 , 행을 숨기면 그림이 업데이트되어 숨겨진 행에 대응하는 그림 요소가 숨겨집니다 . 자세한 내용은 "숨기기 적용 시 그룹화 요소가 업데이트되는 JMP 그림 "(204 페이지) 에서 확인하십시오 .
- 숨겨진 행의 행 번호 옆에는 마스크 아이콘 ( ) 이 나타납니다 .
- 해당 행을 선택하고 **숨기기 / 숨기기 해제**를 다시 선택하기 전까지는 숨김 상태가 유지됩니다 .

숨기기 적용 시 그룹화 요소가 업데이트되는 JMP 그림

행에 숨기기를 적용할 경우 행과 그림 요소 간에 일대일 대응 관계가 있는 그림에서는 일반적으로 점 또는 그림 요소가 즉시 숨겨집니다 . 하지만 다음 플랫폼 및 옵션의 경우에는 숨기기를 적용하면 행 그룹에 대응하는 그림 요소가 즉시 업데이트됩니다 .

- 관리도 빌더 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .
- 그래프 빌더 . 막대 차트 , 히트맵 및 파이 차트는 숨겨진 행을 반영하기 위해 업데이트됩니다 . 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .
- 버블 그림 . 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .

데이터 테이블에서 열 제외

데이터 테이블의 일부 열을 분석 시 고려하지 않으려는 경우 해당 열을 제외하여 시작 창의 선택 목록에 나타나지 않도록 할 수 있습니다 . 다음 사항에 유의하십시오 .

- 제외된 열도 데이터 격자에는 계속 표시됩니다 .
- " 열 " 패널에서 열 이름의 오른쪽에는 취소선이 있는 원 모양 아이콘 ( ) 이 나타납니다 .

시작 창에서 열을 제외하려면

1. 제외할 열을 하나 이상 선택합니다 .
2. 다음 중 하나를 수행합니다 .

- " 열 " 패널에서 강조 표시된 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **제외 / 제외 해제**를 선택합니다 .
- **열** 메뉴에서 **제외 / 제외 해제**를 선택합니다 .

열을 제외 해제하려면

1. " 열 " 패널에서 열을 선택합니다 .
2. 다음 중 하나를 수행합니다 .
   - " 열 " 패널에서 강조 표시된 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **제외 / 제외 해제**를 선택합니다 .
   - **열** 메뉴에서 **제외 / 제외 해제**를 선택합니다 .

데이터 테이블에서 열 숨기기

데이터 테이블의 일부 열을 데이터 격자에 표시하지 않으려는 경우 해당 열을 숨길 수 있습니다. 다음 사항에 유의하십시오.

- 숨겨진 열도 시작 창에는 나타나며 분석에 사용할 수 있습니다 .
- " 열 " 패널에서 열 이름의 오른쪽에는 마스크 아이콘 ( ) 이 나타납니다 .

열을 숨기려면

1. 숨길 열을 하나 이상 선택합니다 .
2. 다음 중 하나를 수행합니다 .
   - " 열 " 패널에서 강조 표시된 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **숨기기 / 숨기기 해제**를 선택합니다 .
   - **열** 메뉴에서 **숨기기 / 숨기기 해제**를 선택합니다 .

열을 숨기기 해제하려면

1. " 열 " 패널에서 열을 선택합니다 .
2. 다음 중 하나를 수행합니다 .
   - " 열 " 패널에서 강조 표시된 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **숨기기 / 숨기기 해제**를 선택합니다 .
   - **열** 메뉴에서 **숨기기 / 숨기기 해제**를 선택합니다 .

데이터 테이블에서 결측 데이터의 패턴 보기

데이터 테이블에 결측 데이터가 포함된 경우 결측 데이터에 일정 패턴이 있는지 여부를 확인할 수 있습니다 . 패턴은 데이터에 대한 정보를 얻는 데 도움이 될 수 있습니다 .

결측 데이터의 패턴을 보려면 다음 단계를 수행하십시오 .

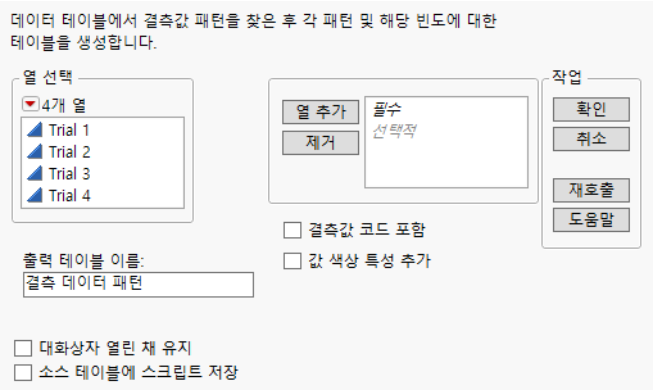
1. 데이터 테이블이 열려 있는 상태에서 **테이블 > 결측 데이터 패턴**을 선택합니다 .
2. 결측 데이터의 패턴을 확인할 열을 선택합니다 .
3. **열 추가**를 클릭합니다 .
4. 결측값 코드를 결측값으로 간주하려면 **결측값 코드 포함**을 선택합니다 .
5. 0 또는 1(결측값 표시자) 이 포함된 열에 일관된 색상을 적용하려면 **값 색상 특성 추가**를 선택합니다 .

   예를 들어 비결측값은 항상 파란색으로 표시하고 결측값은 항상 빨간색으로 표시할 수 있습니다 . 이렇게 하면 여러 열에 유사한 값이 포함될 수 있는 데이터 테이블의 경우에 특히 유용합니다 .
6. **확인**을 클릭합니다 .

예 : 결측 데이터의 패턴 보기

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Missing Data Pattern.jmp 를 엽니다 .
2. **테이블 > 결측 데이터 패턴**을 선택합니다 .

그림 4.18 결측 데이터 패턴 창

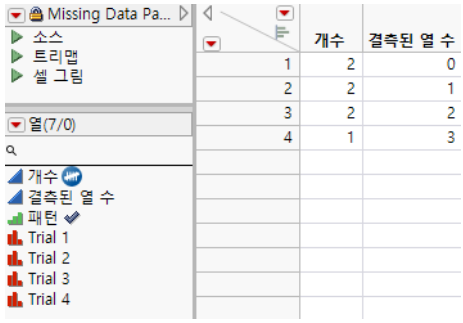


3. 모든 열을 강조 표시합니다 .

   참고 : 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 "JMP 시작하기 " 장의 " 열 필터 메뉴 "(53 페이지) 에서 확인하십시오 .
4. **열 추가**를 클릭합니다 .

5. **확인**을 클릭합니다 .

그림 4.19 결측 데이터 패턴 테이블



팁 : 데이터의 트리맵 또는 셀 그림을 빠르게 생성하려면 테이블 패널에서 " 트리맵 " 또는 " 셀 그림 " 스크립트 옆의 녹색 삼각형을 클릭하십시오 .

그림 4.19 에서는 다음과 같은 패턴을 보여 줍니다 .

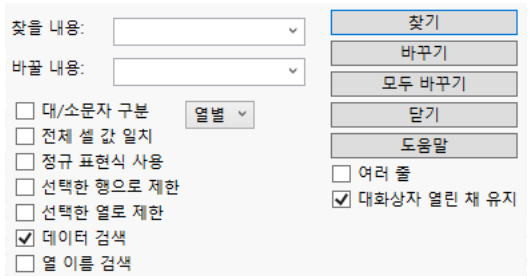
- 1 행은 Trial 1, Trial 2, Trial 3 및 Trial 4 의 모든 행에 결측값이 없는 인스턴스가 두 개 있음을 보여 줍니다 .
- 2 행은 소스 테이블에서 Trial 4 열에 한 개의 결측값이 있는 행이 두 개 있음을 보여 줍니다 .
- 3 행은 소스 테이블에서 Trial 3 및 Trial 4 열에 결측값이 있는 행이 두 개 있음을 보여 줍니다 .
- 4 행은 소스 테이블에서 Trial 2, Trial 3 및 Trial 4 열에 세 개의 결측값이 있는 행이 한 개 있음을 보여 줍니다 .

개수 열에는 빈도 역할이 할당되어 있습니다. 이제 "결측 데이터 패턴" 데이터 테이블을 사용하여 분석을 실행하면 **개수**가 자동으로 빈도로 사용됩니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "열에 사전 선택된 분석 역할 할당"(308 페이지)에서 확인하십시오.

데이터 테이블에서 셀 값 찾기 및 바꾸기

편집 > 검색 > 찾기 옵션을 선택하여 셀 값을 찾고 바꿀 수 있습니다 .

그림 4.20 찾기 창



값 검색 시 다음과 같은 규칙이 적용됩니다 .

- 숨겨진 열의 값을 찾으려면 열을 숨기기 해제하십시오 .
- 잠긴 열에서 찾은 값은 수정할 수 없습니다 .
- **실행 취소** 명령은 **바꾸기**에만 사용할 수 있습니다 . **모두 바꾸기**는 실행 취소할 수 없습니다 .
- 데이터 테이블에 값 라벨이 포함된 경우 **검색** 명령을 사용하면 실제값이 검색되고 라벨은 검색되지 않습니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 값 라벨 "(284 페이지) 에서 확인하십시오 .
- 데이터 테이블에 형식이 지정된 값 (예 : 날짜 , 시간 또는 기간) 이 포함된 경우 " 검색 " 명령을 사용하면 실제값이 아니라 형식이 지정된 값이 검색됩니다 .

찾기 창 옵션

다음과 같은 옵션으로 검색을 구체화하십시오 .

대 / 소문자 구분 대 / 소문자 구분 검색을 수행합니다 . 이 기능은 고유명사나 대문자로 시작하는 그 밖의 단어를 찾는 데 유용할 수 있습니다 .

전체 셀 값 매칭 문자 열에 있는 일련의 단어를 검색하거나 원치 않는 선행 또는 후행 공백이 있는 문자열을 검색할 수 있도록 공백을 감지합니다 .

팁 : 결측 문자 값을 찾으려면 **찾을 내용** 상자를 비워 두고 **전체 셀 값 매칭**을 선택하십시오 . 결측 숫자 값을 찾으려면 **찾기** 상자에 마침표를 넣고 **전체 셀 값 매칭**을 선택하십시오 .

정규 표현식 사용 찾을 문자열이 **찾을 내용** 상자에 입력한 리터럴 문자열이 아니라 정규 표현식인 것으로 가정합니다 . 정규 표현식은 표준 의미 체계를 따릅니다 .

선택한 행으로 제한 검색을 선택한 행으로 제한합니다 .

선택한 열로 제한 검색을 선택한 열로 제한합니다 .

데이터 검색 열 이름은 생략하고 데이터 셀만 검색합니다 .

열 이름 검색 데이터 셀은 생략하고 열 이름만 검색합니다 .

열별 맨 오른쪽 열의 마지막 셀에 도달하거나 사용자가 검색을 중지할 때까지 테이블을 열별로 맨 위에서 맨 아래로 차례로 검색합니다 .

행별 마지막 행의 맨 오른쪽 셀에 도달하거나 사용자가 검색을 중지할 때까지 데이터 테이블을 행별로 왼쪽에서 오른쪽으로 차례로 검색합니다 .

여러 줄 " 찾기 " 및 " 바꾸기 " 상자의 줄 수를 1 줄에서 3 줄로 늘립니다 . Enter 키를 누르면 필드에 줄바꿈 문자가 삽입됩니다 .

팁 : 또는 " 찾기 " 및 " 바꾸기 " 상자를 클릭한 채로 드래그하여 더 크게 늘릴 수 있습니다 . 복사하여 붙여넣을 경우 상자가 1 줄 길이로 조정되지만 모든 텍스트는 상자 내에 유지됩니다 .

대화상자 열린 채 유지 검색하는 동안 " 찾기 " 창을 열린 채로 유지합니다 .

검색 작업

이 섹션에서는 수행 가능한 몇 가지 일반적인 검색 작업에 대해 설명합니다 .

먼저 데이터 테이블에서 값을 검색할 수 있습니다. 검색은 첫 번째 열의 첫 번째 셀에서 시작되며 해당 값을 찾거나 테이블 끝에 도달할 때까지 모든 셀이 검색됩니다.

현재 강조 표시된 셀 값을 바꾸려면

바꿀 내용 상자에 값을 입력하고 **바꾸기**를 클릭합니다. "검색" 창이 닫힌 경우에는 **편집 > 검색 > 바꾸기**를 선택합니다. 바꿀 값이 결측값이면 현재 강조 표시된 셀 내용이 결측값이 됩니다.

지정된 값을 모두 바꾸려면

바꿀 내용 상자에 값을 입력하고 **모두 바꾸기**를 클릭합니다. "검색" 창이 닫힌 경우에는 **편집 > 검색 > 모두 바꾸기**를 선택합니다.

값을 바꾸고 다음 값을 검색하려면

바꿀 내용 상자에 값을 입력하고 **바꾸기**를 클릭합니다. "검색" 창이 닫힌 경우에는 **편집 > 검색 > 바꾸고 다음 찾기**를 선택합니다. 또는 Ctrl+L을 누릅니다.

선택된 값을 찾을 내용 값으로 사용하려면

데이터 테이블에서 값을 선택합니다. **편집 > 검색 > 선택 항목을 찾기에 사용**을 선택합니다. 그런 다음 **편집 > 검색 > 찾기**를 선택합니다. 데이터 테이블에서 선택한 값이 이미 **찾을 내용** 필드에 입력되어 있습니다.

선택된 값을 바꿀 내용 값으로 사용하려면

데이터 테이블에서 값을 선택합니다. 1**편집 > 검색 > 선택 항목을 바꾸기에 사용**을 선택하여 **바꿀 내용** 필드를 채웁니다.

" 검색 " 창이 닫힌 경우 다음 값을 찾으려면

편집 > 검색 > 다음 찾기를 선택합니다. 또는 Ctrl+G나 F3 키 (Windows) 를 누릅니다.

결측값을 찾으려면

- 결측 문자 값을 찾으려면 **찾을 내용** 필드를 비워 두고 **전체 셀 값 매칭**을 선택하십시오 .
- 결측 숫자 값을 찾으려면 **찾을 내용** 텍스트 상자에 마침표를 입력합니다 .

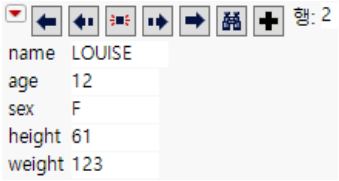
데이터 테이블에서 셀 편집

행 편집기를 사용하여 한 행의 셀을 한 번에 탐색 또는 편집할 수 있습니다 . 다음 중 한 가지 방법으로 행 편집기를 엽니다 .

- **행 > 행 편집기**를 선택합니다 .

- 데이터 테이블에서 행 번호 영역을 두 번 클릭합니다. 사용하는 행이 행 편집기에서 처음 나타나는 행입니다.
- 보고서 창에서 그림 또는 그래프 내부를 마우스 오른쪽 버튼으로 클릭하고 **행 편집기**를 선택합니다.

그림 4.21 행 편집기



다음 사항에 유의하십시오.

- 보고서 창이 열려 있는 경우 편집한 데이터가 보고서에 자동으로 반영되도록 하려면 " 자동 재계산 " 을 설정해야 합니다. 자세한 내용은 "JMP 보고서 " 장의 " 자동 재계산 "(417 페이지) 에서 확인하십시오.
- 데이터 테이블에 값 라벨이 포함되어 있으면 행 편집기에 라벨이 표시되며, 셀이 편집을 위해 강조 표시되어 있으면 행 편집기에 실제값이 표시됩니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 값 라벨 "(284 페이지) 에서 확인하십시오.

행 편집기 버튼

화살표 버튼을 클릭하여 선택한 행 또는 전체 데이터 집합(선택된 행이 없는 경우)을 탐색할 수 있습니다.

- 이전 행을 표시합니다.
- 이전 선택 행을 표시합니다.
- 해당 행이 그래프에서 깜박이도록 합니다.
- 다음 선택 행을 표시합니다.
- 다음 행을 표시합니다.
- 행을 검색합니다. 자세한 내용은 " 행 및 열 크기 조정 "(197 페이지) 에서 확인하십시오.
- 데이터 테이블의 끝에 새 행을 생성합니다.

참고: 행 편집기에서 필드를 변경하면 행 편집기를 사용하여 행에 대해 변경한 사항이 데이터 테이블에 기록됩니다. 데이터 테이블에도 변경 사항을 저장해야 합니다.

행 편집기 옵션

행 편집기의 빨간색 삼각형 메뉴에는 다음과 같은 옵션이 포함되어 있습니다.

다음 선택 항목 현재 행 뒤에 있는 선택된 행에 대한 정보를 표시합니다.

이전 선택 항목 현재 행 앞에 있는 선택된 행에 대한 정보를 표시합니다.

다음 행이 선택되어 있는지 여부에 관계없이 현재 행 뒤에 있는 행에 대한 정보를 표시합니다.

이전 행이 선택되어 있는지 여부에 관계없이 현재 행 앞에 있는 행에 대한 정보를 표시합니다.

저장 데이터 테이블과 행 편집기를 통해 데이터 테이블에 대해 변경한 모든 사항을 저장합니다.

새 행 데이터 테이블에 새 행을 생성합니다.

찾기 **행 > 행 선택 > 선택 조건**을 선택할 때와 동일한 창을 표시합니다. **현재 선택** 메뉴의 옵션 중 하나를 선택한 후 선택하려는 행이 포함된 열을 강조 표시합니다. 그런 다음 검색할 값을 입력합니다. 자세한 내용은 " 행 및 열 크기 조정 "(197 페이지) 에서 확인하십시오.

깜박임 현재 행에서 강조 표시된 항목이 빠른 속도로 깜박이도록 합니다.

참고 : 잠긴 열 또는 잠긴 데이터 테이블의 텍스트는 편집할 수 없습니다. 자세한 내용은 "JMP 열 특성 설정 " 장의 " 잠금 "(269 페이지) 및 "Lock Data Table"(259 페이지) 에서 확인하십시오.

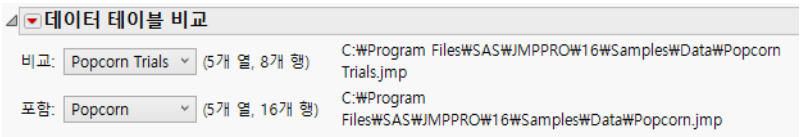
행 및 열에 대한 마우스 오른쪽 버튼 클릭 메뉴

행 번호 영역이나 열 맨 위의 열 이름 영역을 마우스 오른쪽 버튼으로 클릭하면 팝업 메뉴가 나타납니다. 이러한 메뉴에서는 "행" 및 "열" 메뉴의 선택된 명령에 빠르게 액세스할 수 있습니다. 이러한 옵션에 대한 자세한 내용은 "JMP 시작하기" 장의 "열의 마우스 오른쪽 버튼 클릭 옵션"(47 페이지) 및 "JMP 시작하기" 장의 "행의 마우스 오른쪽 버튼 클릭 옵션"(48페이지)에서 확인하십시오.

데이터 테이블 비교

JMP에서는 두 개의 열려 있는 데이터 테이블을 비교하고 데이터, 스크립트, 테이블 변수, 열 이름, 열 특성 및 열 속성 간의 차이점을 보고할 수 있습니다. "데이터 테이블 비교" 창 상단에 각 데이터 테이블의 열 및 행 수가 표시됩니다. 이 예에서는 샘플 데이터 테이블인 Popcorn Trials.jmp와 Popcorn.jmp를 비교합니다.

그림 4.22 데이터 테이블에 대한 기본 분석 정보 정보



매칭 열 지정

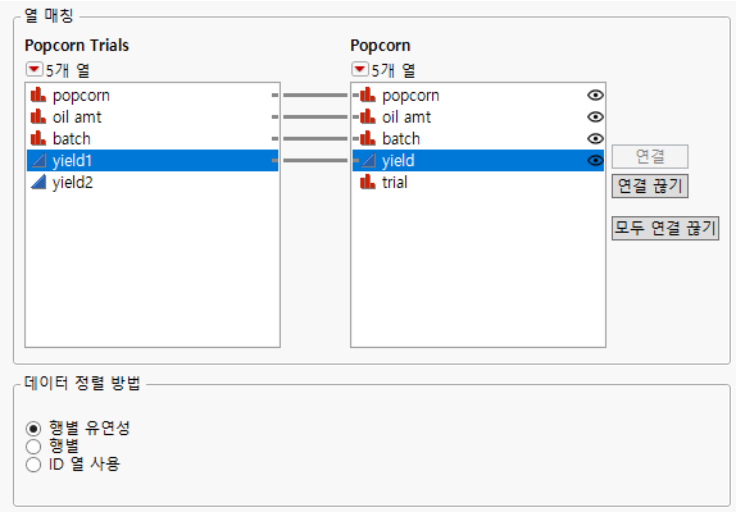
이름이 같은 열은 자동으로 매칭됩니다. 매칭된 열 사이에는 선이 그려집니다. 두 열을 수동으로 연결할 수도 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Popcorn.jmp 및 Popcorn Trials.jmp 를 엽니다.

2. Popcorn Trials.jmp 를 표시하고 **테이블 > 데이터 테이블 비교**를 선택합니다 .
3. " 포함 " 목록에서 Popcorn.jmp 를 선택합니다 .

   " 비교 " 목록에서 Popcorn Trials.jmp 가 자동으로 선택되어야 합니다 .
4. " 열 매칭 " 페인에서 yield 및 yield1 을 선택한 다음 **연결**을 클릭합니다 .

그림 4.23 수동으로 연결된 열



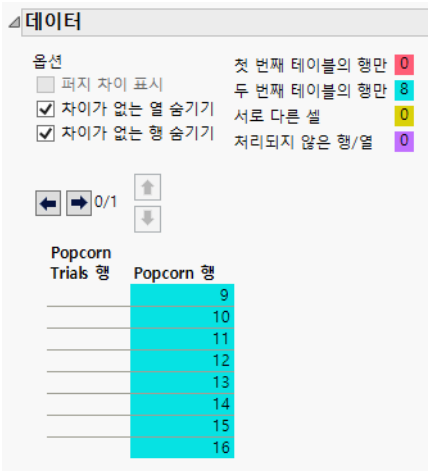
이름이 같은 열은 자동으로 연결됩니다 .

참고 : " 데이터 비교 " 아이콘 을 클릭하면 연결된 열은 비교되지 않습니다 . 해당 열 쌍의 값이 결과에 표시되지만 비교되지는 않습니다 .

5. **비교**를 클릭합니다 .

   결과는 데이터 보고서에 표시됩니다 .

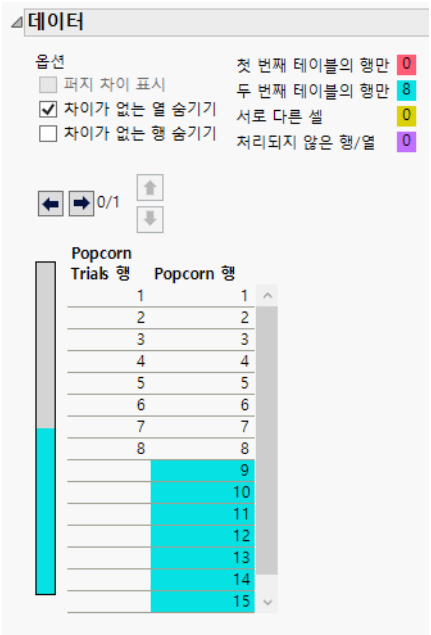
그림 4.24 열 비교



처음 8 개 행은 동일하므로 표시되지 않습니다 . 나머지 행 (파란색이 적용된 행) 은 두 번째 테이블인 Popcorn.jmp 에만 있습니다 .

6. 데이터 보고서에서 **차이가 없는 행 숨기기**를 선택 취소합니다 .

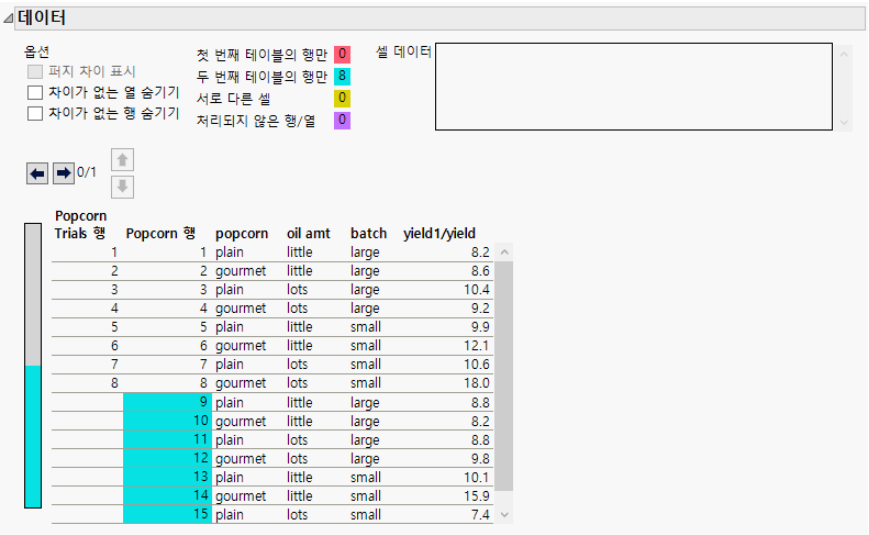
그림 4.25 차이가 없는 행 표시



처음 8 개 행은 두 데이터 테이블에서 동일하기 때문에 표시됩니다 .

7. **차이가 없는 열 숨기기**를 선택 취소합니다 .

그림 4.26 차이가 없는 열 표시



두 데이터 테이블에 있고 매칭되는 열이 표시됩니다 . 이 옵션을 선택하여 매칭된 데이터에 더 많은 컨텍스트를 제공할 수 있습니다 .

참고 : 셀의 데이터가 완전하게 표시되지 않으면 텍스트를 선택한 다음 " 셀 데이터 " 상자에서 데이터를 확인합니다 .

데이터 테이블 비교 옵션

데이터 정렬 옵션

행별 유연성 정렬할 공통 행을 검색합니다 . 소규모 데이터 테이블에서 데이터 테이블이 거의 동일하다고 생각되는 경우 이 옵션을 고려하십시오 .

행별 행을 하나씩 비교합니다 . 행이 행별로 정렬되어야 한다는 것을 이미 알고 있는 경우 이 옵션을 고려하십시오 . 비교가 훨씬 더 빠르게 실행됩니다 . 그러나 기본값인 " 행별 유연성 " 을 선택하여 비교할 때 지나치게 오랜 시간이 걸리는 경우에도 이 옵션을 선택할 수 있습니다 .

ID 열 사용 선택한 ID 열을 사용하여 행을 비교합니다 . 데이터 테이블의 행이 ID 열의 값으로 고유하게 식별됩니다 . 데이터 테이블이 크거나 , 서로 다르게 정렬되거나 , 결측 행이 있는 경우 이 옵션을 고려하십시오 . 둘 이상의 열을 선택할 수 있습니다 .

퍼지 비교 옵션

결측값 무시 결측 데이터를 무시합니다 .

상대 오차 허용 숫자 데이터에 대한 상대 오차 비율을 지정할 수 있습니다 . 사용자가 지정한 상대 오차 비율 내에 있는 숫자 값은 서로 같은 것으로 간주됩니다 . 상대 오차 비율이 작을수록 비교가 정확해집니다 .

대 / 소문자 무시 텍스트를 비교할 때 대 / 소문자를 무시합니다 .

공백 무시 텍스트를 비교할 때 공백을 무시합니다 .

데이터 옵션

퍼지 차이 표시 거의 동일한 숫자 및 문자열 데이터의 차이를 표시합니다 . " 상대 오차 " 필드의 값을 기준으로 유의하지 않은 차이를 제거합니다 .

차이가 없는 열 숨기기 매칭된 모든 열을 표시하거나 숨깁니다 .

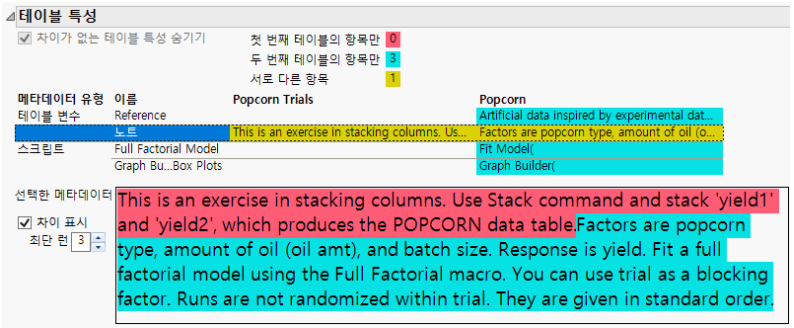
차이가 없는 행 숨기기 매칭된 데이터를 포함하는 행을 표시하거나 숨깁니다 .

테이블 특성 비교

테이블 스크립트 및 변수의 차이를 확인하려면 빨간색 삼각형을 클릭하고 **테이블 특성 비교**를 선택합니다. 그림 4.27에서는 테이블 변수와 스크립트가 다르다는 것을 보여줍니다. 전체 변수 또는 스크립트를 보려면 라인을 선택하고 선택한 메타데이터를 보십시오.

이 예에서는 두 노트가 다르며 참조 변수와 스크립트는 Popcorn.jmp에만 있습니다. 노트 변수는 "선택한 메타데이터" 상자의 두 데이터 테이블에서 내용을 볼 수 있도록 선택됩니다. 빨간색 음영은 해당 텍스트가 Popcorn.jmp 에만 있음을 나타냅니다. 파란색 음영은 해당 텍스트가 Popcorn Trials.jmp에만 있음을 나타냅니다.

그림 4.27 서로 다른 테이블 변수



차이점만 표시하는 대신 각 데이터 테이블의 이름과 각 노트 변수의 전체 내용을 표시하려면 **차이 표시**를 선택 취소합니다.

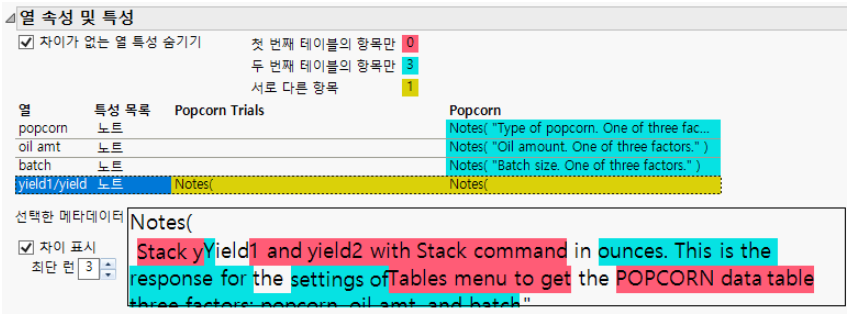
최단 런은 문자를 **공통 하위 세그먼트**로 선언하기 위해 (두 파일 간에서) 동일해야 하는 연속된 문자의 최소 개수를 나타냅니다. 공통 하위 세그먼트는 두 파일에 모두 존재하므로 배경 색상이 없습니다. 최단 런은 3으로 설정되어 있으며, 이는 지나치게 짧은 하위 세그먼트가 공통적으로 표시되어 유용성이 떨어지는 것을 방지하기 위한 것입니다. 예를 들어, 최단 런이 1이면 두 파일에 있는 모든 단일 문자가 매칭됩니다. 이 경우 매우 짧은 공통 텍스트와 차이를 나타내는 많은 세그먼트가 발생하며, 일반적으로 가독성이 떨어집니다.

열 속성 및 특성 비교

열 노트, 값 색상 등의 차이를 확인하려면 빨간색 삼각형을 클릭하고 **열 속성 및 특성 비교**를 선택합니다.

그림 4.28 에서는 Popcorn.jmp 와 Popcorn Trials.jmp 에서 열 노트가 다름을 보여줍니다. yield1/yield 열이 선택되어 있으므로 "선택한 메타데이터" 상자에서 전체 노트와 두 노트 간의 차이점을 볼 수 있습니다.

그림 4.28 열 속성 및 특성 비교



행 및 열에 특성 할당

이 섹션에서는 산점도와 그래프에서의 점 모양을 사용자 정의하기 위해 행 및 열에 제외, 숨기기, 라벨, 색상 또는 표식을 적용하는 방법에 대해 설명합니다.

- " 행 및 열에 대한 옵션을 찾는 방법 "
- " 행 및 열 라벨 지정 "
- " 행에 색상 또는 표식 할당 "
- " 그래프의 색상 테마 생성 "
- " 그래프에서 사용자 색상 테마 삭제 "
- " 행 상태가 매칭되는 셀 선택 "
- " 행 특성 삭제 "
- " 열 잠그기 "

행 및 열에 대한 옵션을 찾는 방법

다음 위치에서 행 작업을 위한 메뉴에 액세스할 수 있습니다.

- 메인 메뉴의 **행** 메뉴
- 행을 마우스 오른쪽 버튼으로 클릭
- " 행 " 패널의 빨간색 삼각형

- 데이터 격자 왼쪽 상단에 있는 왼쪽 빨간색 삼각형

다음 위치에서 열 작업을 위한 메뉴에 액세스할 수 있습니다 .

- 메인 메뉴의 **열** 메뉴
- 열을 마우스 오른쪽 버튼으로 클릭
- " 열 " 패널의 빨간색 삼각형
- 데이터 격자 왼쪽 상단에 있는 오른쪽 빨간색 삼각형

행 및 열 라벨 지정

그림의 점 위에 화살표 커서를 놓으면 해당 점의 라벨이 나타납니다. 기본적으로는 행 번호가 라벨로 사용됩니다. 라벨을 사용자 정의할 수 있습니다.

- 행 번호 대신 열 값을 표시하도록 라벨을 변경할 수 있습니다 .
- 라벨이 점 위에 커서를 놓을 때만 표시되는 것이 아니라 항상 표시되도록 할 수 있습니다 .
- " 열 " 패널의 열 이름 옆에 라벨 또는 노란색 태그 아이콘 ( ) 이 나타나면 그림의 점이 열 값으로 식별되는 것입니다 . 라벨이 지정된 열이 여러 개 있는 경우에는 그림에서 해당 값이 쉼표로 구분되어 표시됩니다 .
- **라벨 / 라벨 해제**를 다시 선택하기 전까지 데이터 라벨은 계속 표시됩니다 .

열 값을 표시하도록 라벨을 변경하려면

1. 그림에 라벨로 표시할 값이 있는 열을 하나 이상 강조 표시합니다 .
2. 메뉴에서 **열 > 라벨 / 라벨 해제**를 선택하거나 , 마우스 오른쪽 버튼을 클릭하고 **라벨 / 라벨 해제**를 선택합니다 .

라벨이 점 위에 커서를 놓을 때만 표시되는 것이 아니라 항상 표시되도록 하려면

1. 그림에 항상 표시할 라벨이 있는 행을 하나 이상 선택합니다 .
2. 메뉴에서 **행 > 라벨 / 라벨 해제**를 선택합니다 .

행 또는 열의 라벨 표시를 해제하려면

1. 라벨이 표시된 행 또는 열 중에서 더 이상 라벨을 표시하지 않을 항목을 강조 표시합니다 .
2. **행** 메뉴 또는 **열** 메뉴에서 **라벨 / 라벨 해제**를 선택합니다 . 열 또는 행을 마우스 오른쪽 버튼으로 클릭하고 **라벨 / 라벨 해제**를 선택해도 됩니다 .

팁 : 데이터를 마우스로 가리킬 때 라벨에 사진이 표시되도록 할 수 있습니다 . 예를 들어 SAS Offices.jmp 샘플 데이터 테이블에는 라벨이 지정된 표현식 열이 포함되어 있습니다 . 이 열에는 각 사무실의 사진이 저장되어 있습니다 . 맵에서 데이터 점을 마우스로 가리키면 해당 행의 사진이 나타납니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 표현식 역할 "(301 페이지) 에서 확인하십시오 .

행에 색상 또는 표식 할당

데이터 테이블의 행을 구분하기 위해 행에 색상이나 표식을 할당합니다.

- 행에 색상을 할당하면 그림에서 해당 행의 값을 나타내는 점이 해당 색상으로 표시됩니다.
- 행에 표식을 할당하면 그림에서 해당 점이 표식으로 바뀝니다.
- 열 값에 따라 색상 또는 표식을 할당할 수도 있습니다.

행에 색상 할당

선택한 행에 색상을 할당하면 그림에서 해당 점이 선택한 색상으로 표시됩니다. 데이터 격자에서는 행에 할당된 활성 색상이 행 번호 옆에 표시됩니다.

행에 색상을 할당하려면

1. 색상을 할당할 행을 하나 이상 강조 표시합니다.
2. 강조 표시한 행을 마우스 오른쪽 버튼으로 클릭하고 **행 > 색상**을 선택합니다.
3. 사용 가능한 색상 중 하나를 선택합니다.

팁 : 선택한 행에서 할당된 색상을 지우려면 검정 색상을 할당하십시오.

행에 표식 추가

그림의 표준 점을 표식으로 바꾸려면 JMP 표식 팔레트를 사용합니다. 해당 표식은 데이터 테이블의 행 번호 옆에도 나타납니다.

1. 표식을 적용할 행을 하나 이상 강조 표시합니다.
2. 선택한 행을 마우스 오른쪽 버튼으로 클릭하고 **표식**을 선택한 후 표식 모양을 선택합니다.

   사용자 표식을 생성하려면 **기타**를 선택합니다. 문자, 숫자 및 기타 키보드 기호를 입력할 수 있습니다.

팁 : 기본 표식으로 되돌리려면 초기 점 표식을 선택하십시오.

열 값에 따라 행에 색상 또는 표식 할당

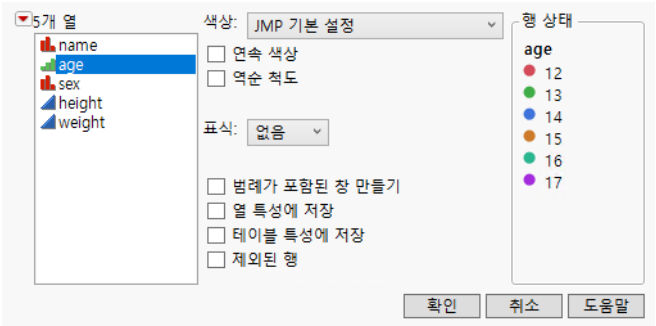
특정 열에 있는 값에 따라 데이터 테이블 행에 색상 또는 표식을 할당할 수 있습니다. 예를 들어 Sex라는 열에서 값이 F인 모든 행에 빨간색 원 표식을 할당할 수 있습니다. 값이 M인 모든 행에는 녹색 더하기 표식을 할당할 수 있습니다. 이러한 색상 및 표식은 그림에서 기본값인 검은색 점을 대체하며, 데이터 테이블에서는 행 번호 옆에 표시됩니다.

열 값에 따라 행에 색상 또는 표식을 할당하려면

1. **행 > 열 값에 따른 색상 또는 표식**을 선택합니다.

2. 색상 및 / 또는 표식을 할당할 열을 선택합니다 .

그림 4.29 열 값에 따른 색상 또는 표식



3. 적용할 **색상** 및 **표식** 구성을 선택합니다 .

   " 행 상태 " 아래에 선택한 구성의 미리보기가 나타납니다 .

4. (선택 사항) 추가 옵션을 선택합니다 . 자세한 내용은 "열 값에 따른 색상 또는 표식 옵션"(219 페이지) 에서 확인하십시오 .

5. **확인**을 클릭합니다 .

6. (선택 사항) 행 상태에 따라 모든 행에 음영을 적용하려면 데이터 격자 내에서 행 번호 영역을 마우스 오른쪽 버튼으로 클릭하고 **행 상태를 기준으로 행에 색상 적용**을 선택합니다 .

   이후부터는 해당 행이 할당된 색상으로 음영 표시됩니다 .

열 값에 따른 색상 또는 표식 옵션

색상 색상 테마를 선택하여 데이터 테이블의 행에 서로 다른 색상을 할당합니다 . 색상 할당은 선택한 열의 값에 따라 달라집니다 .

연속형 척도 강조 표시된 열의 값에 따라 유채색계 시퀀스로 색상을 할당합니다 .

역순 척도 강조 표시된 열의 값에 따라 역순 유채색계 시퀀스로 색상을 할당합니다 .

표식 강조 표시된 열에 있는 값에 따라 데이터 테이블의 각 행에 서로 다른 표식을 할당합니다 .

범례가 포함된 창 만들기 각각의 색상 및 표식이 어떤 행에 해당하는지 쉽게 파악할 수 있도록 새 특성과 함께 범례를 포함합니다 .

열 특성에 저장 색상 및 표식 정보를 열 특성으로 저장합니다 . 데이터 테이블의 선택된 열에 있는 행은 색상 테마에 따라 색상이 적용됩니다 .

테이블 특성에 저장 색상 및 표식 정보를 테이블 특성으로 저장합니다 .

제외된 행 제외된 행에 색상 또는 표식을 할당합니다 .

그래프의 색상 테마 생성

JMP에는 그래프에서 값 범위를 구분할 수 있는 몇 가지 색상 테마가 포함되어 있습니다. 기존 색상 테마를 기초로 고유한 색상 테마를 생성하거나 사용자 테마를 직접 생성할 수도 있습니다.

참고 : 기본 색상 테마를 선택할 경우 열려 있는 보고서에는 해당 색상이 적용되지 않습니다. 기본 색상 테마로 형식을 지정하려면 기존 보고서를 다시 실행해야 합니다.

사용자 색상 테마를 삭제하는 방법에 대한 자세한 내용은 "그래프에서 사용자 색상 테마 삭제"(224 페이지)에서 확인하십시오.

색상 테마를 생성하려면

1. **파일 > 환경 설정 > 그래프**를 선택합니다.
2. 연속 색상 테마 또는 범주형 색상 테마를 새로 생성하려면 적절한 색상 테마를 클릭합니다.
   연속 색상 테마를 새로 생성하려는 경우 " 연속 색상 테마 " 창이 나타납니다.

그림 4.30 연속 색상 테마 창

색상 시퀀스 "Cividis"는 제한된 색상으로 볼 수 있도록 고안된 Viridis의 한 형태입니다.

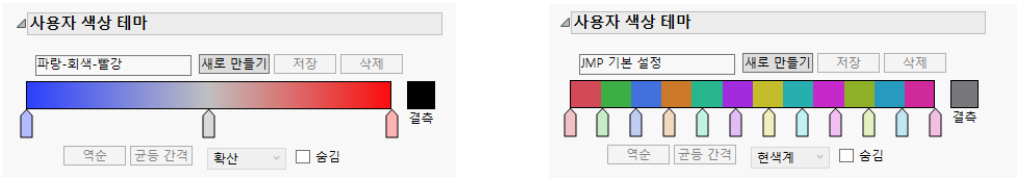
범주형 색상 테마를 새로 생성하려는 경우 " 범주형 색상 테마 " 창이 나타납니다.

그림 4.31 범주형 색상 테마 창

3. (선택 사항) 기존 테마를 기초로 테마를 생성하려면 사용 가능한 색상 테마 중 하나를 선택합니다 .
4. " 사용자 색상 테마 " 패널을 표시하려면 " 사용자 색상 테마 " 표시 아이콘을 클릭합니다 .

   그림 4.32 에서는 연속 테마와 범주형 색상 테마 패널을 각각 보여 줍니다 .

그림 4.32 사용자 색상 테마 패널



5. **새로 만들기**를 클릭하여 새 테마를 생성합니다 .

   선택한 색상 테마를 기초로 새 색상 테마가 생성됩니다. 또한 임시 이름이 테마에 할당됩니다.
6. 임시 라벨 대신 새 이름을 입력합니다 . Windows 의 경우 , Enter 키를 누르지 마십시오 . 그러면 창이 닫힙니다 .
7. 색상 테마를 수정하려면 다음 중 원하는 작업을 수행합니다 .
   - 결측값 색상을 수정하려면 결측값 색상을 클릭하여 색상 팔레트를 엽니다 . 원하는 결측값 색상을 선택하고 **확인**을 클릭합니다 .
   - 연속 색상의 그래디언트를 수정하려면 슬라이더를 왼쪽 또는 오른쪽으로 이동합니다 .
   - 그래디언트에 색상을 추가하려면 색상 표시줄을 클릭하여 색상을 선택합니다. 그러면 색상 표시줄 아래에 새 슬라이더가 표시됩니다 .

- 슬라이더의 색상을 변경하려면 슬라이더를 클릭하여 " 색 " 창을 표시하고 다른 색상을 선택합니다 .
- 그래디언트의 색상 순서를 반대로 하려면 **역순**을 클릭합니다 .
- 그래디언트의 색상이 균등하게 분포되도록 하려면 **균등 간격**을 클릭합니다 .
- 사용자 테마를 " 순차 " 페인에 나열하려면 목록에서 **순차**를 선택합니다 .
- 사용자 테마를 " 확산 " 페인에 나열하려면 목록에서 **확산**을 선택합니다 .
- 사용자 테마를 " 유채색계 " 페인에 나열하려면 목록에서 **유채색계**를 선택합니다 .
- 테마가 색상 테마 목록에 표시되지 않도록 하려면 **숨김**을 선택합니다 .
- 색상 테마에서 특정 색상을 제거하려면 해당 색상의 슬라이더를 클릭한 채로 색상 표시줄의 위 또는 아래로 드래그합니다 .
- 변경을 취소하려면 **취소**를 클릭합니다 .

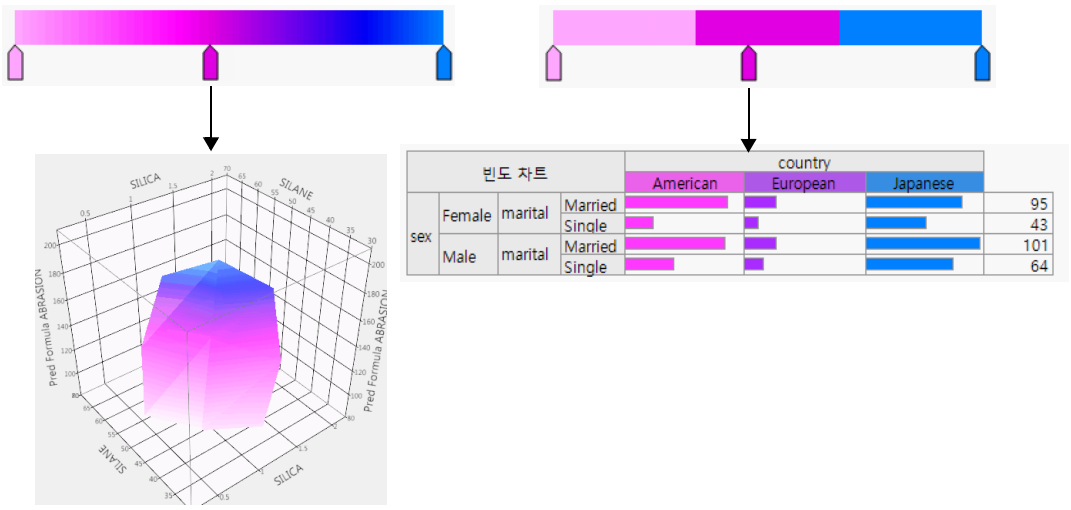
8. **저장**을 클릭하여 사용자 색상 테마를 저장합니다 .

   새 사용자 색상 테마가 선택한 페인의 내용에 추가됩니다 .
9. **확인**을 클릭하여 색상 테마 창을 닫습니다 .

연속 및 범주형 색상 테마

다음 그림에서는 JMP의 두 가지 색상 테마 유형, 즉 연속 색상 테마와 범주형 색상 테마의 예를 보여 줍니다. 연속형 데이터에 사용할 색상 테마를 선택하면 왼쪽과 같이 색상이 그래디언트로 적용됩니다. 범주형 데이터에 동일한 색상 테마를 선택하면 오른쪽과 같이 색상이 여러 개의 개별 색상 블록으로 구성됩니다.

그림 4.33 연속 및 범주형 색상 테마의 예



사용자 색상 테마

사용자 색상 테마는 기본 제공 색상 테마와 동일한 방식으로 적용할 수 있습니다.

- " 그래프 " 환경 설정의 **연속 색상 테마** 및 **범주형 색상 테마** 드롭다운 메뉴에서 사용자 색상 테마를 기본값으로 선택할 수 있습니다 . 연속형 데이터에는 연속 색상 테마만 사용할 수 있습니다 . 범주형 데이터에는 모든 색상 테마를 사용할 수 있습니다 .
- 표식 및 데이터 테이블 행 같은 구성 요소에 사용자 색상 테마를 적용할 수 있습니다 . 자세한 내용은 " 열 값에 따라 행에 색상 또는 표식 할당 "(218 페이지) 에서 확인하십시오 .
- 트리맵 및 표면 그림 같은 일부 보고서에서는 특정 사용자 색상 테마를 선택할 수 있습니다 . 자세한 내용은 Essential Graphing 에서 확인하십시오 .

여러 컴퓨터에서 사용자 색상 테마 사용

Windows 의 경우 , 사용자가 생성하는 색상 테마는 JMP.PFS 라는 JMP 환경 설정 파일에 정의됩니다 . 둘 이상의 컴퓨터 (예 : 가정용 및 직장용) 에서 JMP 를 사용하는 경우 한 컴퓨터에서 JMP 환경 설정의 색상 테마 정의를 다른 컴퓨터에 복사할 수 있습니다 . 그러면 두 컴퓨터 모두에서 사용자 색상을 사용할 수 있습니다 .

환경 설정 파일에서 사용자 색상 테마의 코드는 다음과 같이 나타납니다 .

```
Add Color Theme(
{"Pink to Blue", 2051, {{255, 168, 255}, {255, 0, 255}, {0, 128, 255}}}
),
```

이 예에서 색상 테마의 이름은 "Pink to Blue" 입니다 . 각 색상 슬라이더의 RGB(빨간색 / 녹색 / 파란색) 값은 중괄호 안에 있습니다 . 첫 번째 슬라이더는 RGB 값 255, 168 및 255 를 정의합니다 . 두 번째 및 세 번째 중괄호 그룹은 두 번째 슬라이더와 세 번째 슬라이더의 색상을 정의합니다 .

참고 : 두 번째 값 (2051) 은 색상 테마의 범주 (예 : 범주형 또는 연속형) 입니다 . 범주는 생략해도 되지만 JMP 는 환경 설정 파일 내에 , 또는 `Show Preferences()` 가 실행될 때 항상 이 정보를 포함합니다 .

Microsoft 메모장과 같은 텍스트 편집기에서 다른 컴퓨터의 환경 설정 파일에 이 색상 테마를 추가합니다 . 환경 설정 파일은 Users 폴더 내의 JMP 또는 JMPPro 폴더에 있습니다 .

C:/Users/< 사용자 이름 >/AppData/Roaming/SAS/JMP/16

C:/Users/< 사용자 이름 >/AppData/Roaming/SAS/JMPPro/16

참고: 앞의 폴더를 보려면 Windows 탐색기에서 숨겨진 파일 및 폴더를 표시하도록 구성해야 합니다 . 자세한 내용은 Windows 도움말에서 확인하십시오 .

색상 테마를 다른 Windows 컴퓨터로 전송하려면

1. 사용자 정의한 JMP 환경 설정이 포함된 컴퓨터에서 **파일 > 새로 만들기 > 스크립트**를 선택합니다 .

" 스크립트 " 창이 나타납니다 .

2. 다음 JSL 함수를 입력합니다 .

```
Show Preferences()
```

3. **스크립트 실행** 을 클릭합니다 .

사용자 정의한 환경 설정은 로그에 기록되어 있습니다 .

4. **보기 > 로그**를 선택하거나 , 열려 있는 로그를 표시합니다 .

사용자가 생성한 사용자 색상 테마가 나타납니다 . 예를 들면 다음과 같습니다 .

```
Add Color Theme(
{"Pink to Blue", 2051, {{255, 168, 255}, {255, 0, 255}, {0, 128, 255}}}
),
```

이 정의는 로그에 나타나는 다른 사용자 정의 환경 설정의 가운데에 있을 수도 있습니다 .

5. 로그를 Log.jsl 로 저장하고 환경 설정을 업데이트하려는 컴퓨터에서 이 파일을 엽니다 .
6. 환경 설정을 업데이트하려는 컴퓨터에서 JMP 를 닫습니다 .
7. JMP.PFS 를 백업한 후 원래 JMP.PFS 를 텍스트 편집기에서 엽니다 .
8. Log.jsl 의 사용자 색상 정의를 복사하여 JMP.PFS 에 붙여넣습니다 . 이 정의는 다음 예에 표시된 것과 같이 `Preferences(` 뒤에 나옵니다 .

```
Preferences(
Add Color Theme(
{"Pink to Blue", 2051, {{255, 168, 255}, {255, 0, 255}, {0, 128, 255}}}
),
);
```

참고 : 닫는 괄호와 쉼표를 포함해야 합니다 . 코드를 들여쓸 필요는 없습니다 . 유효한 어떤 위치에든 코드를 배치할 수 있습니다 . `Preferences(` 뒤에 코드를 붙여넣으면 필요한 괄호 또는 쉼표가 삭제되지 않습니다 .

9. 파일을 저장합니다 .

JMP 를 열어 봐서 새 색상 정의가 환경 설정에 표시되지 않으면 업데이트한 환경 설정 파일을 삭제하고 원래 환경 설정 파일에 정의를 추가하십시오 . 정의를 올바른 위치에 복사하여 붙여넣어야 합니다 .

그래프에서 사용자 색상 테마 삭제

1. **파일 > 환경 설정 > 그래프**를 선택합니다 .
2. 색상 테마를 삭제하려면 " 연속 색상 테마 " 또는 " 범주형 색상 테마 " 중 하나를 선택합니다 .

해당 색상 테마 창이 나타납니다 .

3. " 사용자 색상 테마 " 패널을 표시하려면 " 사용자 색상 테마 " 표시 아이콘을 클릭합니다 .
4. 해당 페인에서 삭제할 사용자 색상을 선택합니다 .

참고 : 사용자 색상 테마만 삭제할 수 있습니다 .

5. **삭제**를 클릭합니다 .
6. **확인**을 클릭하여 변경 사항을 저장하고 색상 테마 창을 닫습니다 .

행 상태가 매칭되는 셀 선택

행 상태 열의 셀을 선택한 후 선택한 셀의 속성이 포함된 모든 셀을 선택할 수 있습니다. 이 옵션을 사용하면 제외, 숨기기, 색상 또는 표식과 같은 행 상태 속성의 임의 조합이 적용된 셀을 선택할 수 있습니다. 예를 들어 빨간색 표식이 있고 제외되어 있는 셀이 있는 경우 매칭되는 셀을 선택하면 빨간색 표식이 있고 제외된 행의 모든 셀이 선택됩니다.

1. 선택하려는 행 상태가 적용된 셀을 마우스 오른쪽 버튼으로 클릭합니다 .
2. **매칭 셀 선택**을 선택합니다 .

   매칭되는 행 상태가 적용된 셀이 선택됩니다 .

행 특성 삭제

데이터 테이블에서 모든 행 상태를 지우려면 **행 > 행 상태 지우기**를 선택합니다. 선택한 행의 행 상태만 지우려면 **행 > 선택된 행 상태 지우기**를 선택합니다.

모든 행이 포함되고, 표시되고, 라벨 해제되고, 그림에 검은 점으로 표시됩니다. **행 상태 지우기** 명령은 행 상태 열에 저장된 행 상태에는 영향을 주지 않습니다.

열 잠그기

열을 현재 위치에 잠그면 가로로 스크롤해도 해당 열은 계속 표시됩니다. 열을 강조 표시하고 **열 > 스크롤 잠금/잠금 해제**를 선택합니다. 다음 사항에 유의하십시오.

- 숨겨진 열에는 스크롤 잠금을 적용할 수 없습니다 .
- " 열 " 패널에서 잠긴 열의 이름은 기울임꼴로 나타납니다 .
- 스크롤 잠금이 적용된 열은 데이터 격자의 왼쪽으로 이동됩니다 . 열 잠금을 해제한 후에도 데이터 테이블에서 원래 위치로 돌아가지 않고 왼쪽에 유지됩니다 .
- 스크롤 잠금이 적용된 열은 사용자가 해당 열을 강조 표시하고 **스크롤 잠금 / 잠금 해제**를 다시 선택할 때까지 잠금 상태를 유지합니다 .

데이터 재구성

JMP는 여러 열에서 열 생성, 표시자 열 생성, 데이터 재코딩 등과 같이 데이터 테이블의 데이터를 재구성하는 여러 방법을 제공합니다.

- " 데이터 테이블 열을 여러 열로 분할 "
- " 텍스트 매칭으로 새 열 생성 "
- " 표시자 열 생성 "
- " 데이터 테이블에서 열 결합 "
- " 데이터 테이블에서 선택 열 압축 "
- " 데이터 테이블 열에서 계급화 계산식 만들기 "
- " 라벨과 코드 간 변환 "
- " 새 계산식 열 생성 "
- "JMP 플랫폼의 열 변환 "
- " 열의 데이터 재코딩 "

데이터 테이블 열을 여러 열로 분할

텍스트를 열로 옵션을 사용하여 구분된 필드가 있는 문자 열을 여러 열로 만들 수 있습니다. 데이터 테이블에서 열을 강조 표시하고 **열 > 유틸리티 > 텍스트를 열로**를 선택합니다. 모든 행에 있는 구분된 필드의 최대 개수에 따라 생성되는 새 열의 개수가 결정됩니다.

참고 : 텍스트를 열로 기능은 대 / 소문자를 구분합니다 .

텍스트를 열로 창에는 다음과 같은 옵션이 있습니다.

구분자 소스 열의 데이터가 새 열로 구성되는 방식을 나타내는 텍스트 (예 : 쉼표) 를 지정합니다 . 예를 들어 원래 셀의 내용이 "NY, NJ, PA" 이고 구분자가 쉼표이면 "NY", "NJ" 및 "PA" 가 포함된 세 개의 열이 새로 구성됩니다 .

탭 탭을 구분자로 지정합니다 .

줄바꿈 줄바꿈 문자 (\r, \n 및 \r\n) 를 구분자로 지정합니다 .

표시자 열 생성 소스 열의 구분되는 필드를 따라 이름이 지정되고 셀 값 0 또는 1 을 포함하는 새 열을 생성합니다 .

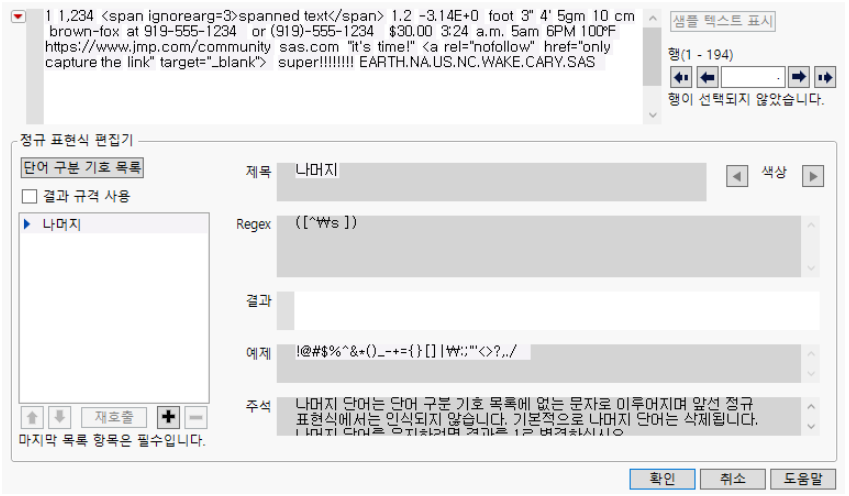
결측값 포함 빈 행이 하나의 범주로 간주되도록 합니다 . 데이터 테이블에 **결측**이라는 열이 추가됩니다 . 값 1 은 빈 행을 나타냅니다 .

텍스트 매칭으로 새 열 생성

"텍스트 매칭으로 새 열 생성" 옵션을 사용하면 정규 표현식(텍스트의 패턴 식별)을 사용하여 열의 비정형 텍스트를 매칭시킨 다음 매칭된 텍스트에서 새 열을 생성할 수 있습니다. 텍스트 탐색기 플랫폼을 사용하여 정규 표현식을 작성한 다음 결과를 열에 저장하는 대신 이 기능을 사용할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Pet Survey.jmp 를 엽니다 .
2. Survey Response 열을 선택합니다 .
3. **열 > 유틸리티 > 텍스트 매칭으로 새 열 생성**을 선택합니다 .

그림 4.34 정규 표현식 편집기



4. " 나머지 " 상자 아래에서 " 추가 " 버튼 ✚ 을 클릭합니다 .
5. 목록에서 **공백**을 선택하고 **확인**을 클릭합니다 .
6. " 제목 " 옆에 Cat 를 입력합니다 .
7. " 정규 표현식 " 옆에 (cat(?:s?))([\s\.]) 를 입력합니다 .

   이 표현식은 "cat" 과 선택적으로 뒤에 공백이나 마침표가 있는 "s" 가 따라오는 "cat" 을 찾습니다 .

   참고 : 정규 표현식에 오류가 있으면 Cat 왼쪽에 X 아이콘 ✘이 나타납니다 .

8. " 추가 " 버튼 ✚ 을 다시 클릭합니다 .
9. 목록에서 **공백**을 선택하고 **확인**을 클릭합니다 .
10. " 제목 " 옆에 Dog 를 입력합니다 .
11. " 정규 표현식 " 옆에 (dog(?:s?))([\s\.]) 를 입력합니다 .

12. (선택 사항) 정규 표현식 편집기를 닫았다가 다시 연 후 다시 채울 필요가 없도록 스크립트를 저장하려면 빨간색 삼각형 메뉴에서 " 스크립트 저장 " 옵션 중 하나를 선택합니다 .
13. **확인**을 클릭합니다 .

그림 4.35 새 열이 있는 데이터 테이블

	Survey Response	Survey Response 2	Survey Response 3
1	The tabby cat do...	cat cat	
2	In the winter I ma...		dog
3	My dogs are not ...		dogs
4	The dog chases c...		dog
5	My dog barks at ...		dog
6	My dogs are so s...		dogs dog
7	I have two dogs t...		dogs
8	My cat slid back...	cat	
9	The dogs bark lik...		dogs

정규 표현식 편집기의 " 결과 규격 사용 " 옵션을 사용하면 새 열에 나타나는 매칭된 텍스트를 제어할 수 있습니다.

- " 결과 규격 사용 "(기본 설정) 을 선택하지 않으면 새 열의 결과는 편집기 오른쪽에 있는 정규 표현식에 의해 생성된 결과입니다 .
- " 결과 규격 사용 "(기본 설정) 을 선택하면 새 열의 결과는 편집기 오른쪽에 있는 " 결과 " 표현식에 의해 생성된 결과입니다 . 정규 표현식이 Air(port) 라면 새 열에는 원래 열에서 찾은 port 라는 단어가 포함됩니다 .

정규 표현식 편집기에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오.

정규 표현식에 대해 자세히 알아보려면 Regular-Expressions.info 사이트를 방문하십시오.

표시자 열 생성

범주형 열을 각각의 개별 범주에 따라 여러 개의 표시자 열로 변환합니다. 열 이름은 해당 열의 개별 수준에 따라 지정됩니다. 예를 들어 Big Class.jmp 의 sex 에서 표시자 열을 생성할 경우 이 표시자 열의 이름은 F 및 M 으로 지정됩니다.

데이터 테이블에서 열을 강조 표시하고 **열 > 유틸리티 > 표시자 열 생성**을 선택합니다. 그런 다음 sex_F 및 sex_M 과 같이 표시자 열 이름 앞에 원래 열 이름을 추가할지 여부와 결측값을 포함할지 여부를 지정할 수 있습니다.

값이 0 또는 1 인 여러 열이 생성됩니다. 값 1 은 원래 열에 해당 특정 범주가 포함되어 있음을 나타냅니다.

지정된 열의 모델링 유형이 " 다중 반응 " 이거나 열 특성이 " 다중 반응 " 인 경우에는 해당 반응 집합에 따라 범주가 결정됩니다.

데이터 테이블에서 열 결합

열 결합 옵션은 텍스트를 열로와 반대의 기능입니다. 여러 열을 생성하는 대신 여러 열을 구분된 필드가 있는 하나의 문자 열로 결합할 수 있습니다.

표시자 열을 결합하려면 다음 단계를 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Consumer Preferences.jmp 를 엽니다 .
2. Floss After Waking Up, Floss After Meal 및 Floss Before Sleep 열을 선택합니다 .
3. **열 > 유틸리티 > 열 결합**을 선택합니다 .
4. 열 이름으로 "Combined Floss" 를 입력하고 기본 구분자인 쉼표를 그대로 둡니다 .
5. **선택한 열을 표시자 열로**를 선택하고 **확인**을 클릭합니다 .

그림 4.36 Combined Floss 열

Combined Floss	Floss After Waking Up	Floss After Meal	Floss Before Sleep
Floss After Waking Up	1	0	0
Floss After Meal	0	1	0
Floss After Meal	0	1	0
Floss After Meal	0	1	0
Floss After Meal	0	1	0
Floss Before Sleep	0	0	1
Floss After Meal	0	1	0
Floss Before Sleep	0	0	1
	0	0	0
Floss After Meal	0	1	0
Floss Before Sleep	0	0	1
Floss After Meal,Floss Before Sleep	0	1	1
Floss After Waking Up,Floss Before Sleep	1	0	1
Floss After Waking Up,Floss After ...	1	1	1
Floss After Waking Up,Floss Before Sleep	1	0	1
Floss After Waking Up,Floss After Meal	1	1	0

선택한 열이 Combined Floss 열에 표현되며 각 필드는 쉼표로 구분됩니다. 각각의 특정 행에 대해 값이 1인 열만 결합된 열에 표현됩니다.

참고 : 값 라벨은 데이터 테이블에서 값 대신 라벨을 표시합니다 . 결합된 열에 있는 각각의 값 인스턴스마다 라벨이 하나씩 나타납니다 . 셀 내의 라벨을 두 번 클릭하면 원래 값을 표시할 수 있습니다 . 값 라벨을 사용하지 않으려면 열을 결합할 때 **값 라벨 없음**을 선택합니다 .

데이터 테이블에서 선택 열 압축

JMP에서는 데이터 테이블의 열을 압축하여 파일 크기를 최소화하고 데이터를 분석하는 데 필요한 메모리 양을 줄일 수 있습니다. 이 기능은 숫자 열에 작은 정수가 많이 포함되어 있거나 열에 포함된 고유 값의 수가 255개 미만일 때 유용합니다. 예를 들어 389개의 열과 85,000개의 행이 있

는 데이터 테이블의 열을 압축하면 데이터 유형에 따라 파일 크기를 250MB에서 33MB까지 줄일 수 있습니다.

열을 압축할 때 JMP는 데이터 유형에 따라 데이터를 보다 압축된 형식으로 저장할 수 있는지 여부를 확인합니다.

- 고유 값이 255 개 미만인 문자 열의 경우, 해당되면 열에 " 목록 확인 " 특성이 추가됩니다 (그림 4.37). 환경 설정에 "16 비트 목록 확인 압축 허용 " 이 선택되어 있으면 255 개 이상의 고유 값이 있는 문자 열에 " 목록 확인 " 특성도 추가됩니다 .

  " 목록 확인 " 특성은 선택된 열의 값을 유효한 값으로 제한합니다 . 선택된 열의 값 수가 너무 많으면 " 목록 확인 " 특성이 적용되지 않습니다 . 예를 들어 값 수와 행 수가 거의 동일하면 데이터 테이블에서 해당 행에 " 목록 확인 " 특성이 추가되지 않습니다 .
- 숫자 열의 경우 , " 최적 ", " 고정 소수점 " 또는 " 데이터 " 형식의 열만 압축됩니다 . 데이터는 가능한 경우 1 바이트 , 2 바이트 또는 4 바이트 정수로 압축됩니다 (그림 4.38). 짧은 정수에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 짧은 정수 형식 "(272 페이지) 에서 확인하십시오 .

  정수가 아닌 값이 포함된 숫자 열도 고유 값의 수가 255 개 미만이면 압축할 수 있습니다 . 이 경우 해당 열에 " 목록 확인 " 특성이 추가됩니다 .

참고 :

- 고유 값의 수가 65,535 개보다 작거나 같은 열을 자동으로 압축하려면 " 일반 " 그룹에서 **16 비트 목록 확인 압축 허용** 환경 설정을 선택하십시오 . 열에 " 목록 확인 " 특성도 추가됩니다 .
- " 목록 확인 " 특성이 있는 열에서는 목록에 있는 값만 입력할 수 있습니다 . 그렇지 않으면 새 값을 입력하려고 할 때 셀에 올바르지 않은 데이터가 포함되어 있다는 경고가 표시됩니다 . 자세한 내용은 " 목록 확인 "(282 페이지) 에서 확인하십시오 .

그림 4.37 압축된 문자 열에 추가된 목록 확인 특성

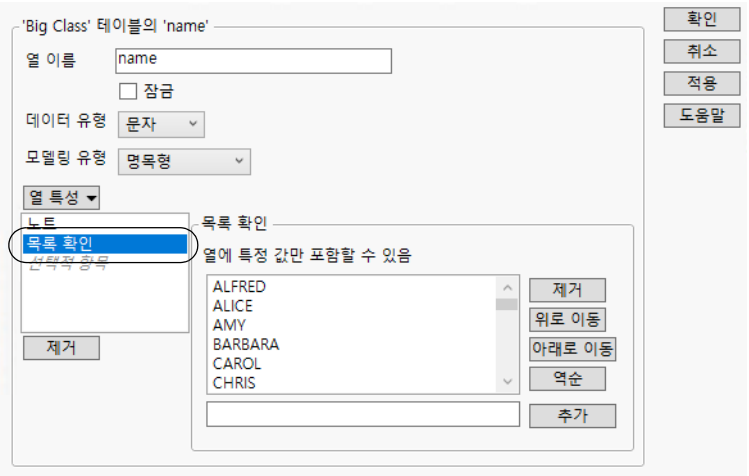
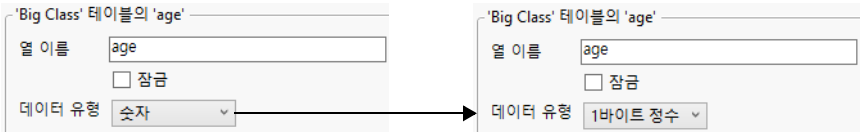


그림 4.38 압축 전후의 숫자 열을 보여 주는 열 정보 창



열을 압축하려면 하나 이상의 열을 선택하고 **열 > 유틸리티 > 선택 열 압축**을 선택합니다. 압축할 수 있는 열이 무엇인지 모르는 경우에는 모든 열을 선택합니다.

가능한 경우 해당 열이 압축됩니다. 로그에는 압축된 열 및 압축 방식이 표시됩니다. 로그를 보려면 **보기 > 로그**를 선택합니다.

참고 : 숫자 열을 수동으로 압축하려면 " 테이블 " 환경 설정에서 짧은 숫자 데이터를 허용하도록 설정한 후 해당 열의 데이터 유형을 **1 바이트 정수** , **2 바이트 정수** 또는 **4 바이트 정수**로 변경하십시오. 이 환경 설정에 대한 자세한 내용은 "JMP 환경 설정" 장의 " 데이터 테이블의 환경 설정"(634 페이지) 에서 확인하십시오 .

데이터 테이블 열에서 계급화 계산식 만들기

"계급화 계산식 만들기" 옵션을 사용하여 데이터를 계급으로 나눌 수 있습니다. 계급으로 나눌 열을 하나 이상 선택하고 **열 > 유틸리티 > 계급화 계산식 만들기**를 선택합니다. 계급의 히스토그램과 값 라벨 범례가 절단점 정의 옵션이 포함된 빨간색 삼각형 메뉴와 함께 표시됩니다. 절단점을 추가하고, 절단점을 제거하고, 계급 크기를 조정하는 방법에는 여러 가지가 있습니다.

- 히스토그램을 마우스 오른쪽 버튼으로 클릭하여 절단점을 추가하거나 제거합니다 .
- 히스토그램에서 수직선을 클릭하고 드래그하여 계급 크기를 조정합니다 .
- 절단점 값 라벨 옆에 있는 상자를 편집하여 절단점 값을 편집하고 계급 크기를 조정합니다 .
- " 값 라벨 " 범례 위에 있는 추가 ✚ 및 제거 ▬ 버튼을 사용하여 행을 추가하거나 제거합니다 . 이 작업은 절단점을 추가하거나 제거하는 것과 같습니다 .
- 실행 취소 ↶ 및 다시 실행 ↷ 버튼을 사용하여 계급 설정에 수행한 작업을 실행 취소하거나 다시 실행합니다 .

절단점에 대한 계산식 열을 데이터 테이블에 추가하는 옵션이 있습니다.

계급화 계산식 만들기의 마우스 오른쪽 버튼 클릭 옵션

"계급화 계산식 만들기" 창에서 히스토그램을 마우스 오른쪽 버튼으로 클릭하여 다음 옵션을 선택합니다.

절단점 추가 히스토그램을 클릭한 위치에 절단점 선을 추가합니다 .

절단점 제거 절단점 선을 제거합니다 . 이 옵션은 기존 절단점을 마우스 오른쪽 버튼으로 클릭한 경우에만 사용할 수 있습니다 .

개수 표시 막대 위에 개수를 표시하거나 숨깁니다 .

백분율 표시 막대 위에 백분율을 표시하거나 숨깁니다 .

부분집합 현재 선택을 기반으로 데이터 테이블 부분집합을 생성합니다 .

채우기 패턴 히스토그램의 막대에 대한 채우기 패턴을 지정합니다 .

절단점에 대한 빨간색 삼각형 메뉴 옵션

절단점 빨간색 삼각형에는 다음과 같은 옵션이 있습니다.

동일 너비 계급을 사용하여 채우기 시작 (오프셋) 및 계급 폭을 지정합니다 . JMP 는 가능한 가장 큰 값을 지날 때까지 테이블을 채웁니다 . 예를 들어 , 오프셋 =50 및 폭 =10 을 입력하고 가장 큰 데이터 값이 95 라면 얻게 되는 계급은 50, 60, 70, 80, 90, 100 입니다 .

계급 수를 사용하여 채우기 시작 , 끝 및 계급 수를 지정합니다 . JMP 는 시작과 끝 사이의 공간을 계급 수로 균등하게 나눕니다 . 계급 수를 변경하면 보고된 계급 폭이 창에서 자동으로 업데이트됩니다 .

백분위수를 사용하여 채우기 1 에서 50 사이의 백분위수를 지정합니다 . JMP 는 데이터 범위를 해당 백분위수로 나눕니다 . 예를 들어 백분위수를 20 으로 지정하면 5 개의 동일한 가중치 계급이 생성됩니다 .

평균 / 표준편차를 사용하여 채우기 데이터의 평균 위치와 창의 선택 항목에 기반한 평균으로부터의 첫 번째 , 두 번째 또는 세 번째 표준 편차 위치에 설정할 계급 절단점을 지정합니다 .

Jenks Natural Breaks 를 사용하여 채우기 계급 수를 지정하고 계급 절단점을 생성하여 각 계급 내의 분산을 최소화하고 계급 간의 분산을 최대화합니다 .

계급 라벨 데이터 값 대신 값 라벨을 표시할지 여부를 지정합니다 .

값 라벨 사용 계급을 정의하는 새 계산식 열에 표시된 값 대신 라벨을 표시합니다 .

범위 라벨 사용 각 범위의 하한 및 상한 값을 라벨에 포함합니다 .

라벨 없음 하한 값을 라벨로 사용합니다 .

문자 문자 데이터 유형을 적용합니다 . 계산식은 라벨을 생성합니다 .

자세한 내용은 "JMP 열 특성 설정 " 장의 " 값 라벨 "(284 페이지) 에서 확인하십시오 .

팁 : 대부분의 플랫폼에서 값 라벨이 권장되며 , 많은 플랫폼은 범위 라벨을 지원하지 않습니다 . 범주형 플랫폼에서는 값 라벨을 사용해야 합니다 . 어떤 축에서는 범위 라벨이 값을 보다 명확하게 식별할 수도 있으며 , 또는 축의 오프셋 및 너비를 조정해 볼 수 있습니다 .

계급 라벨 스타일 값 라벨 또는 문자 출력이 테이블에서 어떻게 보이는지를 정의하는 사전 설정을 지정합니다 .

낮음 - 높음 각 계급의 가장 낮은 점과 가장 높은 점으로 값 라벨을 정의합니다 . 형식은 낮음 – 높음입니다 .

낮음 - 높음 각 계급의 가장 낮은 점과 가장 높은 점으로 값 라벨을 정의합니다 . 형식은 낮음 - 높음입니다 .

낮음 , 높음 각 계급의 가장 낮은 점과 가장 높은 점으로 값 라벨을 정의합니다 . 형식은 낮음 , 높음입니다 .

낮음 - 높음 -1(정수) 각 계급의 가장 낮은 점과 가장 높은 점보다 1 낮은 지점으로 값 라벨을 정의합니다 . 형식은 낮음 – 높음 -1 입니다 .

낮음 - 높음 -1(정수) 각 계급의 가장 낮은 점과 가장 높은 점보다 1 낮은 지점으로 값 라벨을 정의합니다 . 형식은 낮음 - 높음 -1 입니다 .

[낮음 , 높음)(수학) 하한 값은 범위에 포함되고 상한 값은 포함되지 않도록 지정합니다 . 예를 들어 범위가 [50,60) 이면 50 값은 계급에 속하지만 60 값은 속하지 않습니다 .

낮음 각 계급의 가장 낮은 점으로 값 라벨을 정의합니다 .

중간점 각 계급의 중간점으로 값 라벨을 정의합니다 .

정수 시퀀스 ... 연속된 정수 시퀀스로 값 라벨을 정의합니다 . 정수 시퀀스 라벨 창에서 시작 정수를 정의합니다 .

대문자 (사전순) 값 라벨을 대문자 사전순으로 정의합니다 .

소문자 (사전순) 값 라벨을 소문자 사전순으로 정의합니다 .

사용자 정의 사용자 정의 값 라벨을 입력할 수 있습니다 .

열 형식 ... 값 라벨의 형식 지정 방법을 지정합니다 . 기본적으로 열 형식이 적용됩니다 . 그러나 설정을 재정의할 수 있습니다 (예 : 형식을 달러로 변경). 이 설정은 열 형식이 " 최적 " 등이지만 실제로 계급에 소수점이 없도록 하려는 경우 가장 유용합니다 . 모든 값 라벨을 정리하는 대신 형식을 지정합니다 .

히스토그램 막대 위에 개수 또는 백분율을 표시할지 여부를 지정합니다 .

점 표시 막대 아래에 관측된 데이터 점을 표시하거나 숨깁니다 .

축 크기 재조정 축을 원래 척도로 재설정합니다 .

재호출 수행한 마지막 계급화 작업으로 " 계급화 계산식 만들기 " 창을 채웁니다 .

모두 X 처럼 만들기 (여러 열이 선택된 경우에만 표시됨) 첫 번째 열 (X) 에 대해 선택한 사항을 대화상자의 나머지 열에 적용합니다 .

계산식 열 만들기 계산식 열을 만들고 창을 닫습니다 .

팁 : 계산식 열을 생성한 후 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **계급화 계산식 편집**을 선택하여 계산식을 편집할 수 있습니다 . 이렇게 하면 " 계급화 계산식 만들기 " 창이 다시 열립니다 . 계급화 계산식 열을 업데이트하려면 절단점을 편집하고 **열 업데이트**를 클릭합니다 .

예 : 계급화 계산식 만들기 예

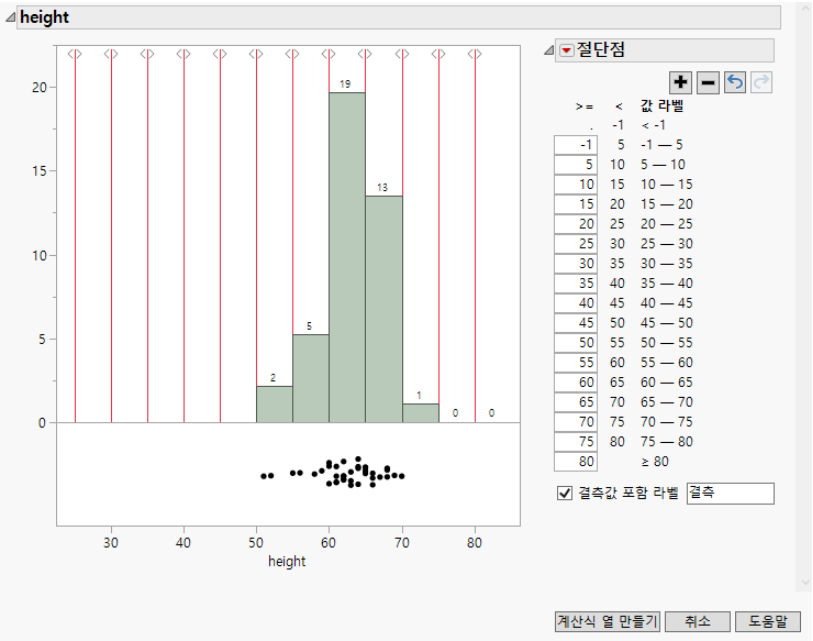
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. height 열을 선택합니다 .

3. **열 > 유틸리티 > 계급화 계산식 만들기**를 선택합니다 .

   값 범위가 X-X 와 같이 나타나도록 하려고 하므로 " 계급 라벨 스타일 " 에서 범위를 **낮음 - 높음**으로 설정된 상태로 둡니다 .
4. 절단점 빨간색 삼각형을 클릭하고 **동일 너비 계급을 사용하여 채우기**를 선택합니다 .
5. 나타나는 창에서 **오프셋**을 -0.5 로 변경합니다 .

   팁 : 정수 데이터의 경우 , 오프셋을 -0.5 로 설정하면 경계 값을 명확하게 할 수 있습니다 . 이 예의 계급 중 하나는 59.5 에서 64.5 사이이므로 59 와 65 는 명확히 이 계급에 포함되지 않습니다 .
6. **너비**를 5 로 설정된 상태로 둡니다 .
7. **확인**을 클릭합니다 .
8. 계급의 값 범위를 볼 수 있도록 **값 라벨 사용**의 " 계급 라벨 " 설정을 유지합니다 .

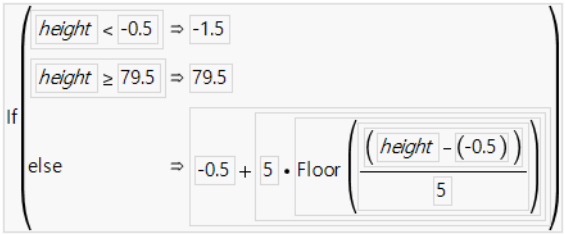
그림 4.39 완료된 계급화 창



9. **계산식 열 만들기**를 클릭합니다 .

   계급화된 height 라는 열이 Big Class.jmp 데이터 테이블에 추가됩니다 .
10. 계산식의 계산 방식을 보려면 계급화된 height 열을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택합니다 .

그림 4.40 계산식



라벨과 코드 간 변환

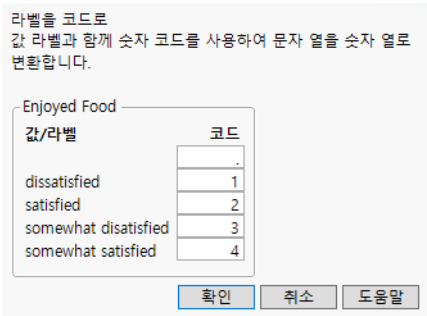
예를 들어 "dissatisfied", "somewhat dissatisfied", "somewhat satisfied" 및 "satisfied" 와 같은 문자열을 포함하는 열이 있다고 가정해 보겠습니다. 이러한 값을 "1", "2", "3" 및 "4" 로 재코딩할 수 있습니다. 원래 문자열은 해당 열의 "값 라벨" 열 특성에 포함됩니다. 열에 "목록 확인" 특성이 있는 경우 해당 특성은 제거됩니다. 열의 데이터 유형은 1 바이트 정수로 변경됩니다. 문자열을 포함하는 열을 짧은 숫자 코드로 변환하면 메모리를 절약하고 처리 속도를 높일 수 있습니다.

코드를 포함하는 숫자 열을 라벨로 변환할 수도 있습니다. 해당 열에는 값 라벨이 있어야 합니다. 열의 데이터 유형은 문자로 변경되고 "값 라벨" 열 특성은 제거됩니다. 사실상 이 기능은 라벨을 코드로 변환하는 기능의 반대 기능입니다.

라벨을 코드로 변환

1. 문자 열을 선택하고 **열 > 유틸리티 > 라벨을 코드로**를 선택합니다 .

그림 4.41 라벨을 코드로 변환



   기본 코드가 값에 할당됩니다 .
2. 원하는 경우 코드를 변경하고 **확인**을 클릭합니다 .
   열이 재코딩되고 값 라벨이 열에 추가됩니다 .

코드를 라벨로 변환

1. 값 라벨이 포함된 숫자 열을 선택하고 **열 > 유틸리티 > 코드를 라벨로**를 선택합니다 .

2. 원하는 경우 라벨을 변경하고 **확인**을 클릭합니다 .

   열이 원래 값 라벨로 재코딩됩니다 .

새 계산식 열 생성

데이터를 보다 세부적으로 분석하려면 기존 데이터 테이블에서 "새 계산식 열" 메뉴 옵션을 사용합니다. 계산식 열은 계산식 또는 계산을 사용하여 열 값을 정의합니다.

데이터 테이블에서 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **새 계산식 열**을 선택합니다. "변환", "문자", "결합", "쌍별", "집계", "분포", "날짜/시간", "난수", "행" 또는 "벡터" 중에서 열 값 계산 옵션을 선택합니다. 그러면 새 계산식 열이 데이터 테이블에 추가됩니다. 이러한 옵션에 대한 자세한 내용은 "JMP 플랫폼의 열 변환"(236 페이지)에서 확인하십시오.

참고 : " 새 계산식 열 " 메뉴와 시작 창에서 마우스 오른쪽 버튼을 클릭하면 나타나는 열 메뉴 모두에 동일한 옵션이 있습니다 . 하지만 시작 창에서 이러한 작업을 수행하면 임시 열이 생성되는 반면 , " 새 계산식 열 " 은 원래 데이터 테이블에 새 열을 추가합니다 .

마우스 오른쪽 버튼을 클릭할 때 나타나는 옵션은 선택한 열의 데이터 유형과 선택한 열 수에 따라 달라집니다. 선택한 열이 문자 열이면 "문자" 및 "행" 옵션이 나타납니다. 자세한 내용은 "문자 메뉴"(243 페이지) 및 " 행 메뉴"(243 페이지)에서 확인하십시오.

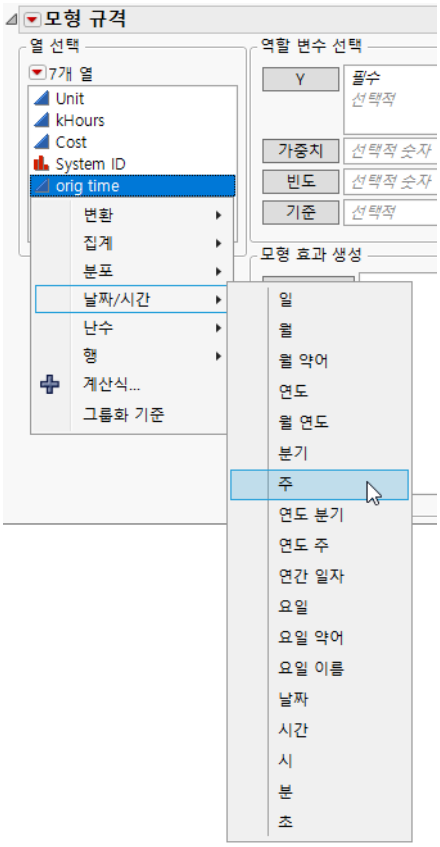
JMP 플랫폼의 열 변환

JMP의 각 시작 창에서는 분석을 수행하는 데 사용할 임시 변환 열을 하나 이상 생성할 수 있습니다. 이러한 변환 열은 소스 데이터 테이블에 포함되지는 않고 현재 시작 창의 컨텍스트 내에서만 사용할 수 있습니다. 변환 열은 계산식 또는 계산을 사용하여 열 값을 정의합니다. 시작 창을 닫으면 모든 변환 열이 삭제됩니다.

시작 창의 "열 선택" 페인에 나열된 각 열에는 열의 모델링 유형 (연속형 , 순서형 또는 명목형) 을 나타내는 아이콘과 열 이름이 포함되어 있습니다. "변환", "문자", "결합", "쌍별", "집계", "분포", "날짜/시간", "난수", "행", "벡터" 또는 "계산식"을 사용하여 열 값을 계산하는 변환 열을 생성하려면 열 이름을 마우스 오른쪽 버튼으로 클릭합니다.

마우스 오른쪽 버튼을 클릭할 때 나타나는 옵션은 선택한 열의 데이터 유형과 선택한 열 수에 따라 달라집니다.

그림 4.42 변환 열 메뉴의 예



그룹화 기준 순서형 및 명목형 데이터의 경우, 데이터를 그룹화하는 데 사용할 열을 지정합니다. 지정된 열의 각 수준에 대해 개별 분석이 실행됩니다.

참고 :

- 변환 열은 현재 시작 창에서만 사용할 수 있습니다. 변환 열을 현재 시작 창 외부에서도 사용할 수 있도록 하려면 변환 열을 마우스 오른쪽 버튼으로 클릭하고 **데이터 테이블에 추가**를 선택하십시오. 그러면 변환 열이 소스 데이터 테이블에 추가됩니다.
- 변환 열을 시작 창의 역할 상자에 붙여넣을 수 있습니다. 예를 들어 스크립트에서 변환 열을 복사할 수 있습니다. 그런 다음 적절한 시작 창에서 " 선택한 열 역할 지정 " 상자를 마우스 오른쪽 버튼으로 클릭하고 " 붙여넣기 " 를 선택하십시오. 이 방법은 " 열 선택 " 목록에서 열을 마우스 오른쪽 버튼으로 클릭하고 변환을 선택한 후 해당 변환 열을 역할에 추가하는 방법 대신 사용할 수 있습니다.
- 변환 함수가 가역적인 경우 다음 플랫폼에서는 예측 프로파일러에 원래 척도를 사용하여 예측된 값과 계산식을 저장합니다.
  - X 로 Y 적합

  - 표준 최소 제곱
  - 파티션
  - 부스티드 트리
  - 붓스트랩 포레스트
- JSL 스크립트를 작성하여 고유한 변환을 정의할 수 있습니다. 자세한 내용은 Scripting Guide의 에서 확인하십시오.

변환 메뉴

선택한 함수를 기반으로 한 계산을 포함하는 변환 열을 생성하려면 "변환" 메뉴에서 함수를 선택하십시오. 자세한 내용은 "도움말" 메뉴의 "스크립트 인덱스" 또는 **선형 모형 적합**의 에서 확인하십시오.

참고 : 여러 열에 단항 함수를 적용하여 여러 변환 열을 생성할 수 있습니다.

표 4.2 변환 메뉴 옵션 설명

Round	날짜 값의 근사값을 계산합니다. 예를 들어 주별 차트에서 "근사값"은 각 날짜 값을 해당 날짜가 속하는 주의 첫 번째 날짜로 변환합니다.
척도 오프셋	일반 선형 변환을 지정할 수 있습니다. 예를 들어 값에 1.8을 곱하고 오프셋 32를 적용하여 섭씨 온도를 화씨 온도로 변환할 수 있습니다.
계급화 사용자 정의 ...	데이터를 계급으로 나눌 수 있는 계급화 창을 엽니다.
제곱근	선택한 열의 값에 대한 제곱근을 취합니다.
제곱	선택한 열 값의 제곱을 계산합니다.
로그	선택한 열에 자연 로그 변환을 적용합니다.
로그 (X+1)	Log(col+1)를 계산합니다.
Exp	선택한 열에 지수 변환을 적용합니다.
상용 로그	선택한 열에 상용 로그 변환을 적용합니다.
10 거듭제곱	선택한 열 값을 지수로 한 10의 거듭제곱을 계산합니다.
세제곱근	선택한 열 값의 세제곱근을 계산합니다.
세제곱	선택한 열 값의 세제곱을 계산합니다.
역수	선택한 열 값의 역수(1/열 값)를 계산합니다.
절대값	선택한 열 값의 절대값을 계산합니다.

표 4.2 변환 메뉴 옵션 설명 (계속)

음수	선택한 열 값의 부정을 계산합니다.
Arrhenius	변수 T(섭씨 온도)에 아레니우스 변환을 적용합니다. $X = \frac{11604.5181215503}{T + 273.15}$ 이 식은 아레니우스 관계식의 구성 요소로, 여기에 활성화 에너지를 곱하게 됩니다.
아레니우스 역	변수 X에 아레니우스 역 변환을 적용합니다. $T = \frac{11604.5181215503}{X} - 273.15$
Logit	선택한 열에 대한 로지스틱 함수의 역함수를 계산합니다(여기서 p는 0 ~ 1 사이). $\text{로짓}(p) = \log\left(\frac{p}{1-p}\right)$
로지스틱	선택한 열에 대한 로지스틱(Squish 및 Logist라고도 함) 함수를 계산합니다(결과 범위 0 ~ 1). $\text{로지스틱}(x) = \frac{1}{(1+e^{-x})}$
Logit Percent	선택한 열에 대한 로짓을 백분율로 계산합니다(여기서 pct는 0 ~ 100 범위의 백분율). $\text{LogitPct}(pct) = \log\left(\frac{\left(\frac{pct}{100}\right)}{1-\left(\frac{pct}{100}\right)}\right)$
로지스틱 백분율	선택한 열에 대한 로지스틱(또는 Logist)을 백분율로 계산합니다(결과 범위 0 ~ 100). $\text{LogisticPct}(x) = \frac{100}{(1+e^{-x})}$

결합 메뉴

"결합" 메뉴에 액세스하려면 여러 열을 선택하십시오. "결합" 메뉴에서는 선택한 함수에 기반한 계산을 포함하는 변환 열을 생성합니다.

이 메뉴에는 다음 함수가 포함되어 있습니다 .

합 첫 번째 열과 두 번째 열의 합 (A + B) 을 계산합니다 .

차 첫 번째 열과 두 번째 열의 차 (A - B) 를 계산합니다 .

차 (역순) 두 번째 열과 첫 번째 열의 차 (B - A) 를 계산합니다 .

곱 첫 번째 열과 두 번째 열의 곱 (A X B) 을 계산합니다 .

비율 두 번째 열에 대한 첫 번째 열의 비율 (A / B) 을 계산합니다 .

비율 (역순) 첫 번째 열에 대한 두 번째 열의 비율 (B / A) 을 계산합니다 .

최소값 선택한 열의 최소값을 반환합니다 .

최대값 선택한 열의 최대값을 반환합니다 .

평균 선택한 열의 평균 값을 반환합니다 .

표준편차 선택한 열에 있는 값의 표준편차를 계산합니다 .

중앙값 선택한 열의 중앙값을 계산합니다 .

분위수 선택한 열에 대해 지정된 백분율의 분위수를 계산합니다 .

기하평균 데이터 곱의 *n* 제곱근을 반환합니다 .

쌍별 메뉴

두 개 이상의 열에 대해 쌍별 결과를 생성하려면 " 쌍별 " 메뉴에서 함수를 선택하십시오 . 이 메뉴는 선택한 열의 개수가 짝수이고 4 보다 크거나 같을 때만 나타납니다 .

합 첫 번째 열과 두 번째 열의 합 (A + B) 을 계산합니다 .

차 첫 번째 열과 두 번째 열의 차 (A - B) 를 계산합니다 .

차 (역순) 두 번째 열과 첫 번째 열의 차 (B - A) 를 계산합니다 .

곱 첫 번째 열과 두 번째 열의 곱 (A X B) 을 계산합니다 .

비율 두 번째 열에 대한 첫 번째 열의 비율 (A / B) 을 계산합니다 .

비율 (역순) 첫 번째 열에 대한 두 번째 열의 비율 (B / A) 을 계산합니다 .

최소값 선택한 열의 최소값을 반환합니다 .

최대값 선택한 열의 최대값을 반환합니다 .

평균 선택한 열의 평균 값을 반환합니다 .

기하평균 데이터 곱의 *n* 제곱근을 반환합니다 .

집계 메뉴

선택한 열 (" 그룹화 기준 " 열을 지정한 경우에는 열의 일부) 에서 계산된 통계량을 포함하는 변환 열을 생성하려면 " 집계 " 메뉴에서 함수를 선택하십시오 .

참고 : 그룹화 기준 옵션은 이러한 함수에 유용합니다 .

이 메뉴에는 다음 함수가 포함되어 있습니다 .

평균 선택한 열의 평균 값을 반환합니다 .

합 선택한 열에 있는 값의 합을 계산합니다 .

개수 선택한 열에 있는 값의 개수를 계산합니다 .

중앙값 선택한 열의 중앙값을 계산합니다 .

분위수 선택한 열에 대해 지정된 백분율의 분위수를 계산합니다 .

최소값 선택한 열의 최소값을 반환합니다 .

최대값 선택한 열의 최대값을 반환합니다 .

표준편차 선택한 열에 있는 값의 표준편차를 계산합니다 .

분포 메뉴

선택한 열에서 계산된 통계량을 포함하는 변환 열을 생성하려면 "분포" 메뉴에서 함수를 선택하십시오. 자세한 내용은 "도움말" 메뉴의 "스크립트 인덱스"에서 확인하십시오.

이 메뉴에는 다음 함수가 포함되어 있습니다 .

중심화 선택한 열의 모든 행에 있는 각 값에서 열 평균을 뺍니다 .

표준화 선택한 열의 모든 행에서 열 값 빼기 평균을 구한 후 표준편차로 나눈 값을 계산합니다 .

범위 0 ~ 1 최소값이 0 보다 크거나 같고 , 최대값이 1 보다 작거나 같도록 데이터를 상향 또는 하향 척도화합니다 .

Box Cox Box-Cox 방정식을 사용하여 데이터를 변환합니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

Johnson 정규화 Johnson 방정식 중 하나를 사용하여 데이터를 변환합니다 . 새 열 이름은 새 데이터를 계산하는 데 사용된 방정식에 따라 "Johnson Su", "Johnson Sb" 또는 " 없음 " 을 나타냅니다 .

결측값 정보화 두 개의 열을 생성합니다 . **정보화** 열에서는 결측값을 열 평균으로 바꿉니다 . **결측됨** 열에서는 결측값의 경우 1 을 나타내고 , 그렇지 않은 경우 0 을 나타냅니다 .

순위 최저 순위 1 부터 시작해서 순위를 반환합니다 (동일한 순위 없음).

순위 (역순) 최고 순위 1 부터 시작해서 순위를 반환합니다 (동일한 순위 없음).

누적 확률 Col Rank(col) / (Col Number(col) + 1) 누적 확률을 계산합니다 .

Normal Quantile 정규 분포의 분위수를 계산합니다 .

SHASH SHASH(SinH-ArcSinH) 분포의 x 에서 평가된 CDF(누적 분포 함수) 를 계산합니다 .

난수 메뉴

숫자 열의 경우, 난수 값을 포함하는 열을 생성하려면 "난수" 메뉴에서 함수를 선택하십시오.

이 메뉴에는 다음 함수가 포함되어 있습니다.

균등 난수 0 에서 1 사이로 균등하게 난수를 생성합니다.

정규 난수 평균이 0 이고 표준편차가 1 인 정규 분포 (인수가 사용되지 않은 경우), 또는 지정된 평균 및 표준편차로 정의되는 정규 분포 (인수가 사용된 경우) 와 근사한 분포를 따르는 난수를 생성합니다.

비복원 표집 값을 평가할 때마다 무작위로 값을 뒤섞습니다. 첫 번째 값의 결과가 두 번째 값의 결과에 영향을 줍니다.

복원 표집 값을 평가할 때마다 무작위로 값을 뒤섞습니다. 첫 번째 값의 결과가 두 번째 값의 결과에 영향을 주지 않습니다.

날짜 / 시간 메뉴

날짜 또는 시간 값을 포함하는 열 값의 경우, 선택한 열에서 계산된 값을 포함하는 변환 열을 생성하려면 **날짜/시간** 메뉴에서 함수를 선택하십시오.

이 메뉴에는 다음 함수가 포함되어 있습니다.

일 선택한 열의 날짜에서 월간 일자를 반환합니다.

월 선택한 열의 날짜에서 월 번호를 반환합니다.

월 약어 (한국어 버전에서는 사용할 수 없음) 선택한 열의 날짜에서 월 약어를 반환합니다.

연도 선택한 열의 날짜에서 연도를 반환합니다.

월 연도 선택한 열의 날짜에서 월 번호와 연도를 반환합니다.

분기 선택한 열의 날짜에서 연간 분기 (1, 2, 3 또는 4) 를 반환합니다.

주 선택한 열의 날짜에서 연간 주 번호를 반환합니다.

연도 분기 선택한 열의 날짜에서 연도와 연간 분기 (1, 2, 3 또는 4) 를 반환합니다.

연도 주 ISO-8601 연간 주 형식을 나타내는 문자열을 반환합니다. 예를 들어 June 12, 2013 의 경우 "2013W24" 를 반환합니다.

연간 일자 선택한 열의 날짜에서 연간 일자를 반환합니다.

요일 선택한 열의 날짜에서 요일을 반환합니다.

요일 약어 선택한 열의 날짜에서 요일 약어를 반환합니다.

요일 이름 선택한 열에 있는 날짜의 전체 요일을 반환합니다.

날짜 선택한 열의 날짜에서 월 , 일 및 연도를 반환합니다.

시간 선택한 열의 날짜에서 시간을 반환합니다.

시 선택한 열의 날짜에서 시 단위 부분을 반환합니다.

분 선택한 열의 날짜에서 분 단위 부분을 반환합니다 .

초 선택한 열의 날짜에서 초 단위 부분을 반환합니다 .

문자 메뉴

선택한 문자 함수로 구성한 문자열을 포함하는 변환 열을 생성하려면 " 문자 " 메뉴에서 함수를 선택하십시오 .

이 메뉴에는 다음 함수가 포함되어 있습니다 .

길이 선택한 열에 있는 각 문자열의 문자 수를 계산합니다 .

연결 선택한 열에 있는 문자열을 하나의 새 문자열로 연결합니다 .

공백으로 연결 선택한 열에 있는 문자열을 하나의 새 문자열로 연결하고 각 하위 문자열을 공백으로 구분합니다 .

쉼표로 연결 선택한 열에 있는 문자열을 하나의 새 문자열로 연결하고 각 하위 문자열을 쉼표 문자로 구분합니다 .

Titlecase 문자열을 첫 글자가 대문자인 형식 (예 : "Sheila Smith") 으로 변환합니다 .

대문자 문자열을 대문자 형식 (예 : "SHEILA SMITH") 으로 변환합니다 .

소문자 문자열을 소문자 형식 (예 : "sheila smith") 으로 변환합니다 .

첫 번째 단어 선택한 열에 있는 문자열에서 첫 번째 단어를 추출합니다 .

마지막 단어 선택한 열에 있는 문자열에서 마지막 단어를 추출합니다 .

순위 최저 순위 1 부터 시작해서 순위를 반환합니다 (동순위는 원래 행 순서대로 처리).

순위 (역순) 최고 순위 1 부터 시작해서 순위를 반환합니다 (동일한 순위 없음).

행 메뉴

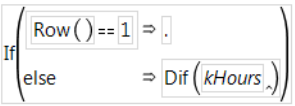
선택한 행 함수로 결정된 계산을 포함하는 변환 열을 생성하려면 " 행 " 메뉴에서 함수를 선택하십시오 .

이 메뉴에는 부록에 설명된 함수 외에 다음 함수가 포함되어 있습니다 .

행 현재 행 번호를 반환합니다 .

선택됨 선택 여부 인덱스를 반환합니다 .

차 다음 계산식을 사용하여 선택한 열에 있는 각 값의 차를 계산합니다 .

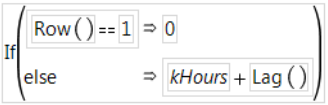


참고 : " 차 " 함수는 " 그룹화 기준 " 옵션도 지원합니다 .

시차 선택한 열의 이전 행에 있는 값을 반환합니다 .

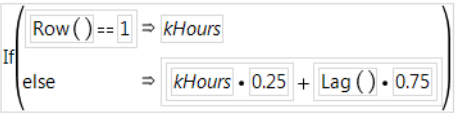
시차 복수 선택한 열의 여러 이전 행에 있는 값을 반환합니다 .

누적합 다음 계산식을 사용하여 선택한 열에 있는 각 값의 누적합을 계산합니다 .



참고 : " 누적합 " 함수는 " 그룹화 기준 " 옵션도 지원합니다 .

이동 평균 0 에서 1.0 사이의 평활 모수를 사용하여 선택한 열에 있는 각 값의 EWMA(지수 가중 이동 평균) 를 계산합니다 . 다음 예에서는 평활 모수 0.25 를 사용합니다 .



가중치 값의 가중 방식을 결정합니다 . 증분 가중은 램프 또는 삼각형입니다 . 지수 이동 평균은 EWMA 또는 EMA 입니다 .

이전 항목 현재 항목 앞에 있는 지정한 개수의 항목을 현재 항목과 함께 평균에 포함하여 범위 (또는 창) 의 크기를 제어합니다 . -1 은 모든 이전 항목을 의미합니다 .

이후 항목 현재 항목 뒤에 있는 지정한 개수의 항목을 현재 항목과 함께 평균에 포함하여 범위 (또는 창) 의 크기를 제어합니다 . -1 은 모든 이후 항목을 의미합니다 .

부분 창에 대한 결측값 보고서 결측값의 처리 방법을 제어합니다 . 기본적으로 결측값은 무시됩니다 .

참고 : JMP 에서는 필요 시 입력된 계산식을 실행하므로 복잡한 계산식은 처리 시간이 많이 필요할 수 있습니다 .

벡터 메뉴

벡터 데이터에서 변환 열을 생성하려면 " 벡터 " 메뉴에서 변환 함수를 선택하십시오 .

여러 열로 벡터를 개별 숫자 열로 확장합니다 .

합 인수의 합 또는 벡터 내에 있는 값의 합을 반환합니다 .

최소값 벡터 내에 있는 값의 최소값을 반환합니다 .

최대값 벡터 내에 있는 값의 최대값을 반환합니다 .

최소 위치 최소값의 첫 번째 위치를 반환합니다 .

최대 위치 최대값의 첫 번째 위치를 반환합니다 .

열 변환 옵션

변환 열을 생성한 후에는 다음 작업을 수행할 수 있습니다 .

이름 바꾸기 변환 열의 이름을 바꿉니다 .

데이터 테이블에 추가 변환 열을 데이터 테이블에 계산식 열로 추가합니다 .

열의 데이터 재코딩

" 재코딩 " 도구를 사용하여 열의 모든 값을 한 번에 변경할 수 있습니다. 예를 들어 컴퓨터 회사와 제약 회사의 매출을 비교하는 데 관심이 있다고 가정해 보겠습니다. 현재 회사 라벨은 Computer 와 Pharmaceutical 입니다. 이를 "Technical" 및 "Drug" 로 변경하려고 합니다. 32 개 행의 모든 데이터를 검토하고 모든 값을 변경하는 것은 지루하고 비효율적이며 오류가 발생하기 쉽습니다. 더 많은 데이터 행이 있을 때 특히 그렇습니다. 재코딩이 더 좋은 방법입니다.

참고 :

- 여러 열에 있는 유사한 값을 재코딩하려면 **열 > 속성 표준화**의 " 재코딩 " 옵션을 사용하십시오. 자세한 내용은 "JMP 열 특성 설정" 장의 "열 속성 표준화"(306 페이지)에서 확인하십시오.
- 열 머리글만 재코딩하려면 **열 > 열 이름 > 열 이름 재코딩**을 선택하십시오 .

데이터를 재코딩하려면 다음 단계를 수행하십시오.

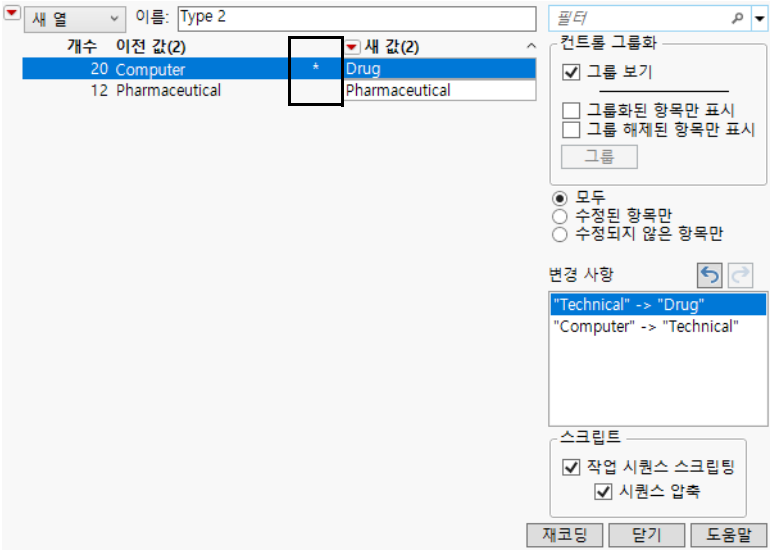
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. Type 열의 머리글을 한 번 클릭하여 해당 열을 선택합니다 .
3. **열 > 재코딩**을 선택합니다 .

   기본적으로 " 새 열 " 이 선택되어 있으므로 재코딩된 값이 새 열에 포함됩니다 . 새 열의 이름이 Type 2 로 지정됩니다 .
4. " 재코딩 " 창의 **새 값** 상자에 원하는 값을 입력합니다. 이 예에서는 "Computer" 행에 "Technical" 을 입력하고 , "Pharmaceutical" 행에 "Drug" 를 입력합니다 .

   " 변경 사항 " 상자에 재코딩된 값이 표시됩니다 .

그림 4.43 재코딩 창

별표는 어떤 값이 변경되었는지를 나타냅니다.



5. **재코딩**을 클릭합니다.

   데이터 테이블의 Type 2 라는 새 열에 새 값이 표시됩니다.

참고 :

- 기본적으로 재코딩된 데이터는 새 열에 표시됩니다. 선택한 열을 항상 그 자리에서 재코딩하려면 **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택하고, **재코딩**을 선택한 다음, " 커밋 스타일 " 목록에서 **현재 위치**를 선택하십시오.
- 기본적으로 " 재코딩 " 창의 데이터는 새 값을 기준으로 오름차순으로 정렬됩니다. 이것은 " 재코딩 " 및 " 열 이름 재코딩 " 모두에 적용됩니다. 데이터가 정렬되는 방식을 변경하려면 **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택하고 **재코딩**을 선택한 다음 " 정렬 스타일 " 환경 설정을 변경하십시오. 데이터 테이블에 나타나는 일반 순서에 따라 열 이름을 정렬하고 지정된 경우 값 순서 열 특성의 순서에 따라 데이터를 정렬하려면 **이전 값**을 선택하십시오. 개수별로 내림차순으로 정렬하려면 **개수**를 선택하십시오.
- 데이터 테이블에 값 라벨이 포함되어 있으면 " 재코딩 " 창에서 " 새 라벨 " 이라는 열에 라벨이 표시됩니다. 이 라벨을 편집하면 " 값 라벨 " 열 특성도 수정됩니다. **값을 값 라벨로 바꾸기**를 선택하여 " 새 값 " 열의 데이터를 수정할 수도 있습니다.
- 열을 재코딩하면 값 라벨, 값 점수, 값 순서, 값 색상, 상위 범주, 코딩 및 결측값 코드 열 특성이 업데이트됩니다.
- 숫자 데이터 유형인 열에 숫자가 아닌 값을 입력하면 데이터 유형을 문자로 변환하라는 메시지가 표시됩니다. 열을 변환하고 새 값을 표시하려면 **예**를 클릭하십시오. 열을 숫자로 유지하고 결측값을 표시하려면 **아니요**를 클릭하십시오.

- "재코딩" 창에서 행을 선택하면 빠른 편집을 위해 데이터 테이블에서도 해당 행이 선택됩니다.

재코딩 옵션

"재코딩"의 빨간색 삼각형 옆에 있는 목록에서는 다음 옵션을 사용할 수 있습니다.

새 열 변경된 데이터를 포함할 새 열을 생성하고 원래 열을 유지합니다.

계산식 열 변경 사항을 값 대신 계산식으로 포함하는 새 열을 생성합니다. 데이터 테이블에서 원래 열의 값을 변경하면 계산식 열에서 해당 값이 자동으로 업데이트됩니다.

현재 위치 모든 변경 사항을 원래 데이터 열에 적용합니다.

"재코딩" 창에서 사용할 수 있는 그 외의 옵션은 다음과 같습니다.

필터 특정 값을 검색합니다. 검색 옵션을 선택하려면 아래쪽 화살표를 클릭합니다. 각 옵션에 대한 자세한 내용은 "재코딩의 검색 필터 옵션"(251 페이지) 에서 확인하십시오. 정규 표현식을 입력하여 필터를 사용자 정의할 수도 있습니다.

그룹 보기 그룹 해제된 값을 보려면 선택 취소합니다. 기본적으로 선택되어 있습니다.

그룹화된 항목만 표시 그룹화된 상태의 재코딩된 값을 표시합니다.

그룹 해제된 항목만 표시 그룹화되지 않은 값을 표시합니다.

그룹 여러 값이 선택되어 있을 때 활성화됩니다. 강조 표시된 값을 동일한 그룹에 포함하려면 **그룹**을 클릭합니다. 그룹화하기 전에 이미 값을 편집한 경우에는 편집된 값이 "새 값" 열의 그룹 대표가 됩니다. 그렇지 않은 경우에는 가장 자주 발생하는 값이 그룹 대표가 됩니다.

모두 모든 데이터 값을 표시합니다.

수정된 항목만 수정된 재코딩 값만 표시합니다.

수정되지 않은 항목만 수정되지 않은 재코딩 값만 표시합니다.

변경 사항 "재코딩" 창에서 적용한 변경 사항의 목록을 표시합니다.

창에서 마지막으로 변경한 사항을 되돌립니다.

창에서 마지막으로 변경한 사항을 재호출합니다.

스크립트

재코딩된 데이터 테이블을 다시 생성하기 위한 스크립트를 저장하려면 빨간색 삼각형 메뉴에서 "스크립트"를 선택한 후 옵션을 선택합니다.

"재코딩" 창에는 다음과 같은 스크립트 옵션이 포함되어 있습니다.

작업 시퀀스 스크립팅 데이터를 제외하고 작업만 스크립트로 저장합니다. 빨간색 삼각형 옵션을 사용하여 모든 문자열을 대문자로 변환할 경우, 생성된 스크립트에는 데이터를 포함하지 않고 입력 문자열을 대문자로 변환하는 코드가 포함됩니다.

시퀀스 압축 불필요한 작업을 제거하여 순차적으로 실행된 작업의 스크립트를 압축합니다. 여러 건의 편집이 포함된 재코딩을 빠르게 완료하려면 이 옵션을 사용하는 것이 좋습니다.

다중 반응

빈 항목 제거 데이터 테이블의 다중 반응 열을 재코딩할 때 열의 빈 항목을 제외합니다.

중복 항목 제거 다중 반응 열에서 중복 항목을 제외합니다. 예를 들어 원래 데이터가 a, b, b, c 이고 이러한 특정 항목을 변경하지 않았다면 데이터 테이블의 재코딩된 열에는 a, b, c 가 포함됩니다.

항목 정렬 다중 반응 열에서 항목을 사전순으로 정렬합니다. 예를 들어 원래 데이터가 b, a, c 이고 이러한 특정 항목을 변경하지 않았다면 데이터 테이블의 재코딩된 열에는 a, b, c 가 포함됩니다.

재코딩을 위한 빨간색 삼각형 메뉴 옵션

"재코딩" 창의 빨간색 삼각형 메뉴에는 옵션이 포함되어 있습니다.

첫 글자 대문자 형식으로 변환 각 단어의 첫 번째 문자를 대문자로 변환하고 나머지 문자는 소문자로 변환합니다.

대문자로 변환 모든 값을 대문자로 변환합니다.

소문자로 변환 모든 값을 소문자로 변환합니다.

공백 자르기 선행 및 후행 공백 문자를 제거합니다. 예를 들어 John 이라는 이름을 앞뒤에 공백이 있는 채로 가져온 경우 이 명령으로 공백을 삭제할 수 있습니다.

공백 제거 후행, 선행 및 중간 공백 문자를 제거합니다.

Collapse Whitespace 선행 및 후행 공백 문자를 제거하고 중간에 있는 중복 공백 문자도 제거합니다. 즉, 공백 문자가 둘 이상 있는 경우 **공백 접기** 명령을 사용하면 두 개의 공백이 한 개의 공백으로 바뀝니다.

구두점 제거 단어의 처음, 중간 및 끝에서 인용 부호 및 앰퍼샌드 같은 특수 문자를 제거합니다.

문자열 바꾸기 지정한 문자를 모두 새 문자열 또는 아무것도 없는 상태로 바꿉니다. 정규 표현식을 입력하여 바꾸기를 사용자 정의할 수도 있습니다.

첫 번째 단어 문자열의 첫 번째 단어를 기준으로 값을 그룹화합니다. 예를 들어 "John Smith" 와 "John Adams" 라는 값이 있는 경우 이 명령을 사용하면 두 값이 "John" 아래에 그룹화됩니다.

마지막 단어 문자열의 마지막 단어를 기준으로 값을 그룹화합니다.

유사한 값 그룹화 데이터 그룹화 방식을 사용자 정의할 수 있습니다. 그룹화 옵션 목록에서 원하는 방식을 선택합니다. 자세한 내용은 " 재코딩을 위한 유사한 값 그룹화 옵션 "(250 페이지) 에서 확인하십시오.

고급

첫 번째 단어를 제외한 모든 단어 첫 번째 단어를 제외한 후 나머지 값을 기준으로 값을 분할합니다.

마지막 단어를 제외한 모든 단어 마지막 단어를 제외한 후 나머지 값을 기준으로 값을 분할합니다.

테이블의 매핑 적용 매핑 데이터 테이블의 열을 사용하여 목표 테이블의 값을 수정할 수 있습니다. 자세한 내용은 " 예 : 매핑 열의 값 적용 "(255 페이지) 에서 확인하십시오.

세그먼트 추출 특정 문자열을 추출할 수 있습니다. 자세한 내용은 " 예 : 쉼표를 기준으로 분할 "(253 페이지) 에서 확인하십시오.

최근접 선택 매핑 데이터 테이블의 열에 있는 값을 재코딩하려는 열에 매핑할 수 있습니다.

숫자로 파싱 데이터를 숫자로 변환합니다.

라벨을 코드로 라벨이 지정된 값을 코드로 변환합니다.

문자로 변환 숫자 열을 문자 열로 변환합니다.

값 라벨 추가 각 값에 값 라벨을 할당할 수 있습니다.

값을 값 라벨로 바꾸기 " 값 라벨 " 열 특성에 지정된 값을 사용자가 지정하는 값 라벨로 바꿉니다. 이 옵션은 열에 " 값 라벨 " 특성이 있는 경우에만 사용할 수 있습니다.

재호출 " 재코딩 " 창에서 이전에 수행한 변경 작업을 재호출합니다.

다시 시작 창을 기본 상태로 되돌립니다.

스크립트 재코딩 스크립트를 가져오거나 병합하거나 저장하기 위한 옵션을 표시합니다.

- **파일에서 가져오기** : JSL 스크립트를 가져와서 이전에 재코딩한 데이터를 재코딩합니다. 다른 데이터에 대해 동일한 스크립트를 실행하여 동일한 방식으로 데이터를 재코딩할 수 있습니다.
- **데이터 테이블에서 가져오기** : 데이터 테이블에 저장된 JSL 스크립트를 가져옵니다.
- **재코딩 결과 열에서 가져오기** : 선택한 열 (일반적으로 이전에 재코딩 결과를 저장한 열) 에서 데이터를 가져옵니다.
- **파일에 저장** : 재코딩 변경 사항을 JSL 스크립트에 저장합니다. " 저장 " 을 선택하면 파일의 이름을 지정한 후 저장하라는 메시지가 표시됩니다.
- **데이터 테이블에 저장** : 재코딩 스크립트를 현재 데이터 테이블에 저장합니다.

  저장된 스크립트에서 재코딩을 수정하려면 데이터 테이블의 스크립트를 마우스 오른쪽 버튼으로 클릭하고 **재코딩을 사용하여 편집**을 선택합니다. 변경 후에는 빨간색 삼각형 메뉴에서 **스크립트 > 데이터 테이블에 저장**을 다시 선택하여 변경 사항을 새 스크립트에 저장합니다.
- **데이터 테이블 스크립트와 병합** : " 재코딩 " 창에서 변경한 사항을 데이터 테이블에 저장된 현재 재코딩 스크립트에 병합합니다. 스크립트가 여러 개이면 재코딩한 데이터를 병합하는 데 사용할 스크립트를 선택하라는 메시지가 표시됩니다.
- **스크립트 창에 저장** : 재코딩 스크립트를 스크립트 창에 추가합니다.

JMP PRO **계산식 저장소에 게시** 다른 스크립트에서 코드를 재사용할 수 있도록 재코딩 스크립트를 계산식 저장소에 게시합니다.

재코딩을 위한 마우스 오른쪽 버튼 클릭 옵션

다음과 같은 처음 6 개의 옵션은 그룹화된 값에 사용할 수 있습니다.

새 값으로 그룹화 (여러 값을 선택한 경우에만 사용 가능) 두 개 이상의 선택된 값에 대한 새 값을 지정할 수 있습니다 .

~(으) 로 그룹화 (여러 값을 선택한 경우에만 사용 가능) 선택한 값을 마우스 오른쪽 버튼으로 클릭하여 다른 그룹화 값 또는 그룹 대표를 선택할 수 있습니다 . 이 목록에는 처음 8 개의 가능한 그룹 대표가 표시됩니다 .

새 값 교환 두 개의 값이 강조 표시된 경우 첫 번째 값의 새 값에 두 번째 값의 새 값을 적용하거나 그 반대로 적용하려면**새 값 교환**을 선택합니다 .

대표 생성 그룹의 단일 값을 마우스 오른쪽 버튼으로 클릭하고 **대표 생성**을 선택하면 선택한 값이 새 값이 됩니다 .

그룹에서 제거 값이 그룹화된 후 하나 이상의 값을 마우스 오른쪽 버튼으로 클릭하여 해당 그룹에서 제거할 수 있습니다 .

유사한 값 그룹화 단일 문자열을 마우스 오른쪽 버튼으로 클릭하여 유사한 값을 찾을 수 있습니다 . 이 경우 " 그룹화 옵션 " 창이 나타납니다 . 자세한 내용은 재코딩을 위한 유사한 값 그룹화 옵션에서 확인하십시오 .

모든 그룹 열기 각 그룹의 모든 값을 볼 수 있도록 그룹을 엽니다 .

모든 그룹 닫기 각 그룹 내의 값을 축소합니다 .

다음 옵션은 모든 데이터에 사용할 수 있습니다.

행 선택하기 데이터 테이블에서 선택된 행을 선택합니다 . " 재코딩 " 창에서 행을 선택하면 데이터 테이블에서 해당 행이 자동으로 선택됩니다 . 그러나 데이터 테이블에서 행을 선택하면 " 재코딩 " 에서 행이 선택되지 않습니다 . " 재코딩 " 창에 표시된 행만 선택하려면 행을 마우스 오른쪽 버튼으로 클릭하고 이 옵션을 선택하십시오 .

데이터 테이블로 만들기 " 재코딩 " 창에 표시된 결과를 포함하는 데이터 테이블을 생성합니다 .

선택한 열 내 행 복사 마우스 오른쪽 버튼으로 클릭한 열의 데이터만 복사합니다. 그런 다음 해당 데이터를 데이터 테이블에 붙여넣을 수 있습니다 .

선택한 테이블 내 행 복사 모든 열의 데이터를 복사합니다 . 그런 다음 해당 데이터를 데이터 테이블에 붙여넣을 수 있습니다 .

재코딩을 위한 유사한 값 그룹화 옵션

다음과 같은 유사한 값 그룹화 명령을 선택하여 그룹화 정확도를 높일 수 있습니다 .

대 / 소문자 무시 항목의 대 / 소문자를 무시합니다 .

인쇄할 수 없는 문자 무시 인쇄할 수 없는 문자를 무시합니다 . 일부 데이터에는 컴퓨터만 인식할 수 있는 파일 구분자 등의 인쇄할 수 없는 문자가 포함되어 있을 수 있습니다 .

공백 무시 공백을 무시합니다 .

구두점 무시 구두점을 무시합니다 .

문자 편집 허용 유사한 값을 그룹화할 때 문자를 새 값으로 바꿀 수 있도록 합니다 .

최대 차이 비율 최대 비율 차이에 따라 값을 그룹화합니다 . 예를 들어 최대 25% 의 차이가 있는 값을 그룹화하려면 "0.25" 를 입력합니다 .

최대 문자 차이 인접하지 않은 문자 차이의 최대 개수에 따라 값을 그룹화합니다 . 예를 들어 5 자 이하의 차이가 있는 값을 그룹화하려면 "5" 를 입력합니다 .

최소 소스 문자열 길이 너무 짧은 문자열은 매칭되지 않도록 합니다 . 이 옵션과 " 최대 차이 비율 " 및 " 최대 문자 차이 " 옵션을 다양하게 사용해 보고 데이터에 적합한 값을 지정하십시오 .

재코딩의 검색 필터 옵션

검색 상자 옆의 아래쪽 화살표 버튼을 클릭하여 검색을 구체화할 수 있습니다 .

용어 포함 검색 기준의 일부가 포함된 항목을 반환합니다 . "ease oom" 을 검색하면 "Release Zoom" 과 같은 메시지가 반환됩니다 .

구 포함 검색 기준이 정확히 포함된 항목을 반환합니다 . "text box" 를 검색하면 "text" 와 "box" 가 바로 연이어 포함된 항목 (예 : "Context Box" 및 "Text Box") 이 반환됩니다 .

구로 시작 검색 기준으로 시작하는 항목을 반환합니다 .

구로 끝남 검색 기준으로 끝나는 항목을 반환합니다 .

전체 구 전체 문자열로 구성된 항목을 반환됩니다 . "text box" 를 검색하면 "text box" 만 포함된 항목이 반환됩니다 .

정규 표현식 검색 상자에서 와일드카드 (*) 와 마침표 (.) 를 사용할 수 있습니다 . "get.*name" 을 검색하면 "get" 다음에 하나 이상의 단어가 포함된 항목을 찾을 수 있습니다 . 즉 , "Get Color Theme Names", "Get Name Info" 및 "Get Effect Names" 등이 반환됩니다 .

결과 반전 검색 기준과 매칭되지 않는 항목을 반환합니다 .

모든 용어 일치 문자열이 모두 포함된 항목을 반환합니다 . "t test" 를 검색하면 검색 문자열 중 하나 또는 둘 모두가 포함된 요소 (예 : "Pat Test", "Shortest Edit Script" 및 "Paired t test") 가 반환됩니다 .

대 / 소문자 무시 검색 기준의 대 / 소문자를 무시합니다 .

전체 단어 일치 " 모든 용어 일치 " 설정에 따라 문자열의 각 단어가 포함된 항목을 반환합니다 . " 모든 용어 일치 " 옵션이 선택되어 있는 경우 "data filter" 를 검색하면 "data" 와 "filter" 가 반환됩니다 .

예 : 유사한 값을 기준으로 한 그룹화

서로 다른 문자의 개수에 따라 유사한 값을 그룹화할 수 있습니다 .

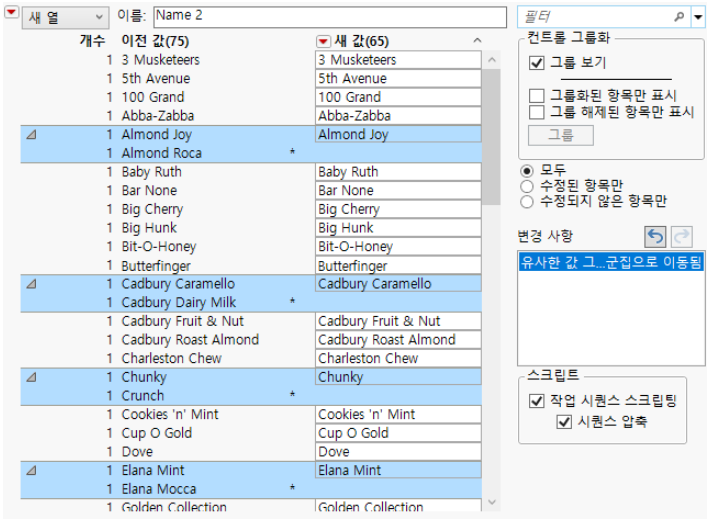
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Candy Bars.jmp 를 엽니다 .
2. 이름 열을 선택합니다 .
3. **열 > 재코딩**을 선택합니다 .
4. 빨간색 삼각형을 클릭하고 **유사한 값 그룹화**를 선택합니다 .

5. " 문자 편집 허용 " 이 선택되어 있는지 확인합니다 .
6. " 최대 차이 비율 " 옆에 "5" 를 입력합니다 .

   그러면 JMP 가 50% 이하의 차이가 있는 값을 그룹화할 수 있습니다 . 즉 , 동일한 문자를 50% 또는 절반 이상 공유하는 값을 그룹화할 수 있습니다 . 차이 비율은 각 값의 최대 문자 수와 지정된 두 값 간의 최대 고유 문자 수를 비교하여 결정됩니다 .
7. " 최대 문자 차이 " 옆에 "6" 을 입력합니다 .

   6 자 이하의 값이 그룹화됩니다 .
8. **확인**을 클릭합니다 .

그림 4.44 유사한 값 그룹화



새 값 열에 표시된 값은 재코딩 대상 데이터 테이블에서 그룹화되는 값을 나타냅니다 .

9. (선택 사항) 그룹의 다른 이전 값을 나타내도록 새 값을 변경하려면 "Almond Roca" 를 마우스 오른쪽 버튼으로 클릭하고 **대표 생성**을 선택합니다 .

그림 4.45 대표 생성

10. (선택 사항) 데이터 테이블의 원래 데이터를 재코딩된 데이터로 바꾸려면 빨간색 삼각형 메뉴에서 **현재 위치**를 선택한 다음 **재코딩**을 클릭합니다 .

예 : 쉼표를 기준으로 분할

Aircraft Incidents.jmp 샘플 데이터 테이블에는 도시와 주 이름이 쉼표로 구분된 열이 포함되어 있습니다. 이 열의 값을 쉼표를 기준으로 분할하고 주 이름만 포함하도록 열을 재코딩할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Aircraft Incidents.jmp 를 엽니다 .
2. Location 열을 선택하고 **열 > 재코딩**을 선택합니다 .

그림 4.46 원래 데이터

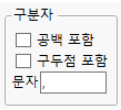
개수	이전 값(1537)	새 값(1537)
1	Abilene, TX	Abilene, TX
1	Adelanto, CA	Adelanto, CA
1	Afton, VA	Afton, VA
1	Aiea, HI	Aiea, HI
1	Aiken, TX	Aiken, TX
1	Ainsworth, NE	Ainsworth, NE
1	Akiachak, AK	Akiachak, AK
1	Akron, CO	Akron, CO
2	Alamogordo, NM	Alamogordo, NM
1	Alamosa, CO	Alamosa, CO
1	Albion, IL	Albion, IL
9	Albuquerque, NM	Albuquerque, NM
1	Alcova, WY	Alcova, WY
1	Alexander City, AL	Alexander City, AL
1	ALEXANDER, AK	ALEXANDER, AK
1	Alice, ND	Alice, ND
1	Allamuchy, NJ	Allamuchy, NJ
1	Almond, WI	Almond, WI
1	Alton, IL	Alton, IL
1	Alturas, CA	Alturas, CA
1	Altus, OK	Altus, OK
2	Alvarado, TX	Alvarado, TX
1	Amargosa Valley, NV	Amargosa Valley, NV
2	Amarillo, TX	Amarillo, TX
1	AMHERST, NH	AMHERST, NH

3. 빨간색 삼각형을 클릭하고 **고급 > 세그먼트 추출**을 선택합니다 .

   미리보기 페인의 데이터에는 도시 이름 다음에 약어 형식의 주 이름이 포함되어 있습니다 . 각 문자열의 마지막 2 자를 추출하려고 합니다 .
4. 창 상단의 **공백 포함**과 **구두점 포함**을 선택 취소합니다 .

   공백과 구두점이 추출 대상에서 제외됩니다 .
5. " 문자 " 상자에 쉼표를 입력합니다 .

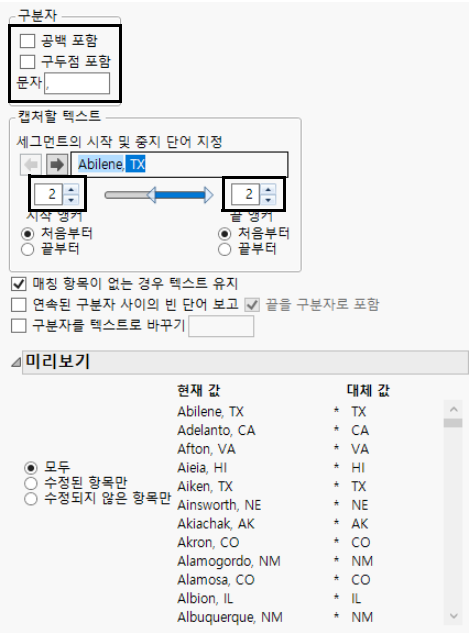
그림 4.47 구분자 지정



   마지막 2 자를 추출하는 데 쉼표가 사용됩니다 .
6. " 세그먼트 추출 " 창에서 " 시작 앵커 " 와 " 끝 앵커 " 위에 "2" 를 입력합니다 .

   세그먼트 추출이 시작되고 두 번째 단어만이 남습니다 .
7. " 시작 앵커 " 와 " 끝 앵커 " 에 " 처음부터 " 가 선택되어 있어야 합니다 .

그림 4.48 세그먼트 추출 창



팁 : " 캡처할 텍스트 " 섹션에서 왼쪽 또는 오른쪽 화살표를 클릭하여 값을 스크롤하면 설정을 미리 볼 수 있습니다 .

8. 그림 4.48 에 표시된 미리보기 페인에서는 대체 값에 주 이름 약어만 포함되어 있습니다 .
9. 미리보기 페인에서 아래로 스크롤합니다 . "Atlantic Ocean" 의 경우 현재 값과 대체 값이 동일합니다 . 이 데이터는 쉼표를 기준으로 분할할 수 없습니다 .

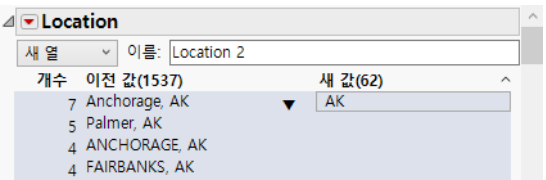
그림 4.49 주 이름이 없는 데이터

현재 값		대체 값
ATHENS, GA	*	GA
Atlanta, GA	*	GA
Atlantic Ocean		Atlantic Ocean

10. **확인**을 클릭합니다 .

    매칭되는 주 이름이 있는 항목이 그룹화됩니다 .

그림 4.50 재코딩된 도시 , 주 이름



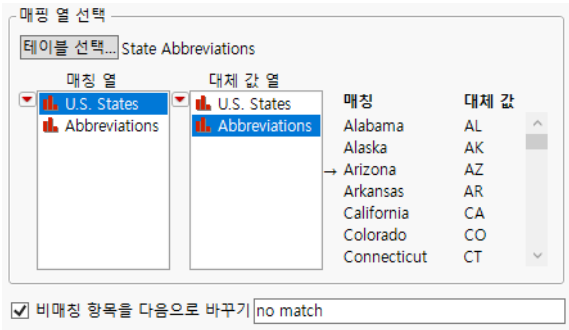
예 : 매핑 열의 값 적용

매핑 데이터 테이블의 열에는 다른 데이터 테이블의 열을 재코딩하는 데 사용할 수 있는 값이 포함되어 있습니다. 이 예에서는 State Capitals.jmp에 포함된 미국 주 이름을 State Abbreviations.jmp (매핑 데이터 테이블)에 정의된 주 약어로 바꿉니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 State Abbreviations.jmp 와 State Capitals.jmp 를 엽니다 .
2. State Capitals.jmp 에서 U.S. States 열을 선택합니다 .
3. **열 > 재코딩**을 선택합니다 .
4. 빨간색 삼각형을 클릭하고 **고급 > 테이블의 매핑 적용**을 선택합니다 .
5. **테이블 선택**을 클릭하고 State Abbreviations.jmp 를 선택한 다음 **확인**을 클릭합니다 .
6. " 매칭 열 " 상자에서 **U.S. States** 를 선택합니다 .
7. " 대체 값 열 " 상자에서 **Abbreviations** 를 선택합니다 .

   매칭 열과 대체 값 열에 매핑 결과가 표시됩니다 .

그림 4.51 열 매칭 및 바꾸기



8. **비매칭 항목을 다음으로 바꾸기** 상자를 선택하고 "no match" 를 입력합니다 .

   미리보기 섹션에 값 매핑 방식이 표시됩니다 . 아래로 스크롤하면 "no match" 로 표시된 주는 철자가 올바르지 않아서 Abbreviations 열의 어떤 주에도 매핑되지 않았음을 알 수 있습니다 .

그림 4.52 매칭되지 않는 열의 미리보기

현재 값		대체 값
Kentucky	*	KY
Louisiana	*	LA
Maine	*	ME
Maryland	*	MD
Massachusets	*	no match
Michigan	*	MI
Minnesota	*	MN
Mississippi	*	no match

9. **확인**을 클릭하고 **재코딩**을 클릭합니다 .

약어 형식 주 이름이 포함된 새 열이 State Capitals.jmp 에 U.S. States 2 라는 이름으로 표시됩니다 .

그림 4.53 약어 형식 주 이름

U.S. States	U.S. States 2
Alabama	AL
Alaska	AK
Arizona	AZ
Arkansas	AR
California	CA
Colorado	CO
Connecticut	CT
Delaware	DE
Florida	FL
Georgia	GA
Hawaii	HI
Idaho	ID
Illinois	IL
Indiana	IN
Iowa	IA
Kansas	KS
Kentucky	KY
Louisiana	LA
Maine	ME
Maryland	MD
Massachusets	no match
Michigan	MI
Minnesota	MN
Mississippi	no match
Missouri	MO

매칭되지 않은 항목을 수정하는 방법은 " 예 : 철자가 올바르지 않은 단어 수정 "(256 페이지) 에서 확인하십시오 .

예 : 철자가 올바르지 않은 단어 수정

데이터 테이블을 사전으로 사용하여 철자가 올바르지 않은 단어를 수정할 수 있습니다. State Abbreviations.jmp 에는 미국의 주 이름 및 약어가 포함된 열이 있습니다. State Capitals.jmp 에는 실제 데이터와 비슷하지만 철자가 올바르지 않은 주 이름이 몇 개 포함되어 있습니다. 이 예에서는 철자가 올바르지 않은 주 이름을 수정하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 State Abbreviations.jmp 와 State Capitals.jmp 를 엽니다 .
2. State Capitals.jmp 에서 U.S. States 열을 선택합니다 .
3. **열 > 재코딩**을 선택합니다 .
4. 빨간색 삼각형을 클릭하고 **고급 > 최근접 선택**을 선택합니다 .
5. **테이블 선택**을 클릭하고 State Abbreviations 를 선택한 다음 **확인**을 클릭합니다 .
6. " 매칭 열 " 상자에서 **U.S. States** 를 선택합니다 .

7. 미리보기 페인에서 아래로 스크롤하면 "Massachusets" 가 "Massachusetts" 로 바뀌고 , "Missisippi" 는 "Mississippi" 로 바뀔 것임을 알 수 있습니다 .

그림 4.54 철자가 올바르지 않은 주 이름 대체

현재 값		대체 값
Kentucky		Kentucky
Louisiana		Louisiana
Maine		Maine
Maryland		Maryland
Massachusets	*	Massachusetts
Michigan		Michigan
Minnesota		Minnesota
Missisippi	*	Mississippi

8. **확인**을 클릭하고 **재코딩**을 클릭합니다 .

   업데이트된 주 이름이 포함된 새 열이 데이터 테이블에 추가됩니다 .

예 : 숫자로 파싱

"숫자로 파싱" 옵션은 문자 데이터를 숫자 데이터로 변환합니다. 이 옵션은 JSL `Num()` 문자 함수와 유사합니다.

날짜 및 시간, 통화, 과학적 표기법과 같은 많은 형식이 지원됩니다. 지원되지 않는 형식의 경우에는 결측값이 표시됩니다.

1. 데이터 테이블을 열고 숫자 값과 문자 값이 모두 포함된 열을 선택합니다 .

그림 4.55 숫자 데이터와 문자 데이터가 포함된 열

	x
1	a
2	4,033.67
3	$8000.56
4	5/4/2019
5	10/21/2017 5:10:33
6	12Jan2018
7	1.07E-6

2. **열 > 재코딩**을 선택합니다 .
3. 빨간색 삼각형 메뉴에서 **숫자로 파싱**을 선택합니다 .

   이전 값과 새 값이 표시됩니다 .

그림 4.56 새 값 미리보기

개수	이전 값(7)		새 값(7)
1	$8000.56	*	8000.56
1	1.07E-6	*	1.07e-6
1	4,033.67	*	4033.67
1	5/4/2019	*	3.64e+9
1	10/21/2017 5:10:33	*	3.591e+9
1	12Jan2018	*	3.599e+9
1	a	*	.

4. **재코딩**을 클릭합니다 .

   재코딩된 값이 포함된 새 열이 데이터 테이블에 추가됩니다. 열의 데이터 유형은 숫자입니다. "a" 는 숫자로 파싱할 수 없으므로 새 열에서 해당 값에는 결측값이 표시됩니다 .

그림 4.57 파싱된 숫자

	x	x 2
1	a	•
2	4,033.67	4033.67
3	$8000.56	8000.56
4	5/4/2019	3.6398e+9
5	10/21/2017 5:10:33	3.5914e+9
6	12Jan2018	3.599e+9
7	1.07E-6	1.07e-6

예 : 재코딩된 값을 다른 열에 적용

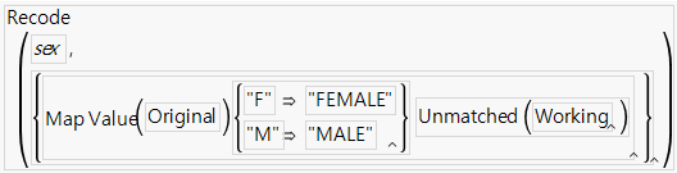
열을 재코딩할 때 재코딩된 값을 스크립트에 저장할 수 있습니다. 이 스크립트를 가져오거나 실행하면 원래 재코딩한 열이 다시 재코딩됩니다.

예를 들어 다른 열을 동일한 값으로 재코딩하려고 한다고 가정해 보겠습니다. 재코딩된 값을 계산식 열에 저장한 후 계산식을 편집할 수 있습니다.

1. " 재코딩 " 창의 " 새 값 " 열에 재코딩된 값을 입력합니다 .
2. 빨간색 삼각형 메뉴 옆의 목록에서 **계산식 열**을 선택합니다 .
3. **재코딩**을 클릭합니다 .
4. 계산식 열을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택합니다 .
5. 계산식에서 재코딩된 열을 선택합니다 .

   이 예에서는 sex 열이 재코딩되었습니다 .

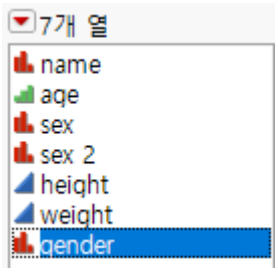
그림 4.58 재코딩된 열



6. 열 목록에서 sex 를 바꿀 열을 선택합니다 .

   이 예에서는 sex 열을 gender 로 바꿀 것이므로 gender 를 선택합니다 .

그림 4.59 gender 열 선택



데이터 테이블 편집

JMP는 데이터 테이블의 기본 특성을 변경하는 다양한 방법을 제공합니다.

- " 데이터 테이블 이름 바꾸기 "
- "Lock Data Table"
- " 데이터 테이블 압축 "
- " 테이블 스크립트 복사 "
- " 데이터 테이블 변수 사용 "
- " 데이터 테이블에서 스크립트 생성 및 저장 "

데이터 테이블 이름 바꾸기

데이터 테이블의 이름은 해당 창, 테이블 패널 및 모든 관련 분석 보고서의 상단에 나타납니다. 다음 중 원하는 방법으로 데이터 테이블의 이름을 변경할 수 있습니다.

- **파일 > 다른 이름으로 저장**을 선택하고 새 이름으로 저장합니다.
- 테이블 패널에서 테이블 이름을 클릭하고 새 이름을 입력한 후 Enter 키를 누릅니다.
- Windows 의 경우 **창 > 제목 설정**을 선택합니다.

Lock Data Table

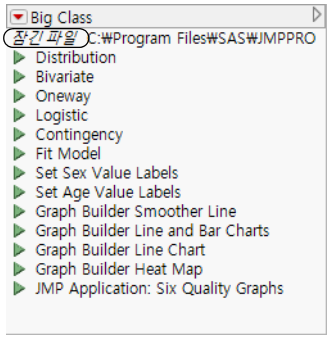
데이터 테이블을 잠그면 데이터 및 열 특성을 추가하거나 편집하지 못하게 됩니다. 행 상태를 할당하거나 분석을 실행하는 등의 작업은 여전히 수행할 수 있습니다. 데이터 테이블을 잠그려면 테이블 패널에서 테이블 패널 옆의 빨간색 삼각형을 클릭하고 **데이터 테이블 잠그기**를 선택합니다.

잠긴 데이터 테이블 이름 옆에는 잠금 아이콘 이 나타납니다. 파일을 잠금 해제하려면 **데이터 테이블 잠그기**를 다시 선택합니다.

Windows에서 데이터 테이블 특성을 변경하는 등의 방법으로 데이터 테이블을 JMP 외부에서 읽기 전용이 되도록 하면 데이터 테이블에 테이블이 잠겨 있음을 알리는 노트가 포함됩니다(그

림 4.60). 이 유형으로 잠글 경우 사용자가 데이터 테이블을 편집할 수 있지만 변경 사항을 저장할 수는 없습니다.

그림 4.60 읽기 전용 파일



데이터 테이블 압축

데이터 테이블을 압축하면 저장된 파일의 크기가 줄어듭니다. 분석을 실행하거나 특성을 할당하는 등의 작업은 여전히 수행할 수 있습니다. 데이터 테이블을 압축하려면 테이블 패널에서 테이블 패널 옆의 빨간색 삼각형을 클릭하고 **저장 시 파일 압축**을 선택한 후 데이터 테이블을 저장합니다.

데이터 테이블을 저장한 후에는 데이터 테이블 이름 옆에 압축 아이콘 이 나타납니다. 파일을 압축 해제하려면 **저장 시 파일 압축**을 다시 선택합니다.

또한 **환경 설정 > 일반 > GZ로 압축된 데이터 테이블 열 저장**을 선택하여 JMP에서 저장 시 항상 GZ 압축이 사용되도록 구성할 수도 있습니다.

참고 : 저장 시 파일 압축 옵션은 파일 크기만 줄여 줍니다 . 이 명령은 데이터를 분석하는 데 필요한 메모리에는 영향을 주지 않습니다 . 파일 크기와 분석에 필요한 메모리를 모두 줄이려면 **열 > 유틸리티 > 선택 열 압축**을 사용합니다 . 자세한 내용은 " 데이터 테이블에서 선택 열 압축 "(229 페이지) 에서 확인하십시오 .

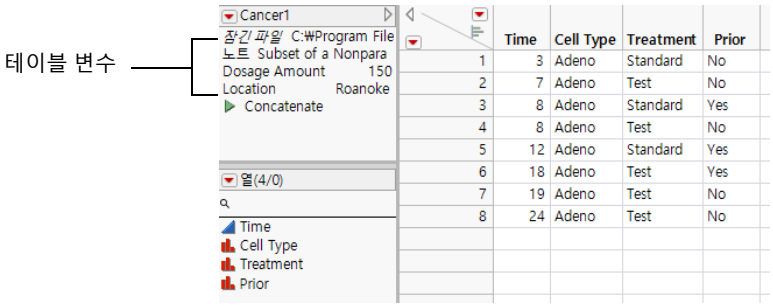
테이블 스크립트 복사

테이블 스크립트를 복사하면 스크립트를 다른 스크립트에 붙여넣어 테이블을 다시 생성할 수 있습니다.

데이터 테이블 변수 사용

데이터 테이블 변수에는 텍스트 정보 (예 : 데이터의 소스 정보) 가 포함되거나 열 계산식 또는 JSL 스크립트에서 사용할 수 있는 값이 포함될 수 있습니다 . 테이블 변수 이름은 데이터 격자 왼쪽의 테이블 패널에 나타납니다 .

그림 4.61 테이블 패널의 테이블 변수



테이블 변수의 용도

테이블 변수는 다음과 같은 경우에 사용합니다.

- 테이블을 문서화하려는 경우
- 계산식에서
- JSL 스크립트에서

테이블 변수를 사용하여 테이블 문서화

테이블 변수는 주로 테이블을 문서화하는 데 사용됩니다. JMP와 함께 설치되는 많은 샘플 데이터 테이블에는 "노트"라는 테이블 변수가 포함되어 있습니다. 이 변수는 데이터에 대한 상세 정보(예: 데이터 소스)를 제공합니다. 그림 4.61의 예에서는 **노트**가 테이블 변수의 하나로 포함되어 있는 데이터 테이블을 보여 줍니다. 또한 JMP에서 "실험 설계" 명령을 사용하여 설계 테이블을 생성하면 테이블 변수가 자동으로 생성됩니다. 설계 테이블에는 설계 유형의 이름이 값으로 사용되는 **설계**라는 테이블 변수가 있습니다.

계산식에서 테이블 변수 참조

계산식 편집기를 사용하여 작성하는 계산식에 테이블 변수를 포함할 수도 있습니다. 이러한 계산식은 테이블 변수를 참조하여 열의 값을 계산합니다. 테이블 변수를 사용하는 계산식의 생성 방법에 대한 자세한 내용은 "JMP의 계산식 생성" 장의 "계산식에서 데이터 테이블 값 참조"(363 페이지)에서 확인하십시오.

JSL 스크립트에서 데이터 테이블 변수 사용

테이블 변수를 JSL 스크립트에 포함할 수도 있습니다. 자세한 내용은 Scripting Guide의 에서 확인하십시오.

데이터 테이블 변수 작업

새 데이터 테이블 변수를 추가하려면

1. 테이블 패널에서 데이터 테이블 이름 왼쪽의 빨간색 삼각형을 클릭합니다.
2. **새 테이블 변수**를 선택합니다.
3. **이름** 및 **값** 상자에서 변수 이름 및 값을 지정합니다.
4. **확인**을 클릭합니다.
   테이블 패널에 해당 테이블 변수가 나타납니다.

테이블 변수를 보거나 편집하려면

1. 기존 테이블 변수의 내용을 두 번 클릭합니다.
2. 내용을 편집합니다.

테이블 변수 이름을 편집하려면

1. 테이블 변수 이름을 두 번 클릭합니다.
2. 이름을 편집합니다.

테이블 변수를 삭제하려면

테이블 변수를 하나 이상 선택하고 Delete 키를 누르거나, 선택한 변수를 마우스 오른쪽 버튼으로 클릭하고 삭제를 선택합니다.

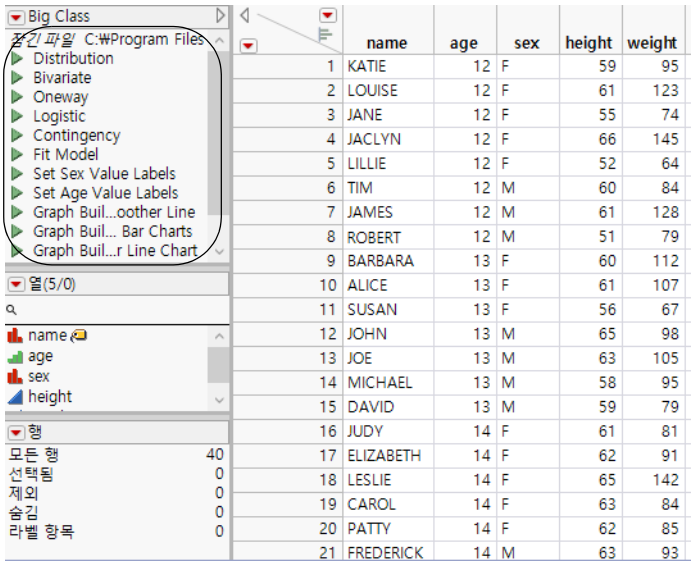
데이터 테이블을 테이블 변수와 연결

자세한 내용은 "데이터 재구성" 장의 "예: 데이터 테이블 및 테이블 변수 연결"(330 페이지)에서 확인하십시오.

데이터 테이블에서 스크립트 생성 및 저장

다양한 분석 및 작업을 자동으로 완료하려면 JSL 스크립트를 생성하여 데이터 테이블에 저장하면 됩니다(그림 4.62). 데이터 테이블 스크립트 작성에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

그림 4.62 데이터 테이블과 함께 저장된 스크립트



	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107
11	SUSAN	13	F	56	67
12	JOHN	13	M	65	98
13	JOE	13	M	63	105
14	MICHAEL	13	M	58	95
15	DAVID	13	M	59	79
16	JUDY	14	F	61	81
17	ELIZABETH	14	F	62	91
18	LESLIE	14	F	65	142
19	CAROL	14	F	63	84
20	PATTY	14	F	62	85
21	FREDERICK	14	M	63	93

데이터 테이블에 보고서 스크립트 저장

분석을 실행한 후 보고서 창이 활성화된 상태에서 데이터 테이블에 스크립트를 추가할 수 있습니다. 이 스크립트는 분석을 재현하는 JSL 을 생성합니다.

데이터 테이블에 스크립트를 저장하려면

보고서 창에서 플랫폼의 빨간색 삼각형을 클릭하고 **스크립트 저장 > 데이터 테이블에**를 선택합니다.

예 : 데이터 테이블에 보고서 스크립트 저장

먼저, 분석을 생성한 후 스크립트를 저장합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. height 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "weight 대 height 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 .
7. "weight 대 height 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **스크립트 저장 > 데이터 테이블에**를 선택합니다 .
8. " 다른 이름으로 스크립트 저장 " 창에서 스크립트의 이름을 입력합니다 .

9. 동일한 이름의 기존 스크립트를 바꾸려면 " 이름 " 목록에서 스크립트를 선택하고 **기존 스크립트 바꾸기**를 선택합니다 .
10. 기존 스크립트와 동일한 이름의 스크립트를 저장하고 동일한 이름을 사용하려면 **고유 접미사 추가**를 선택합니다 .

    스크립트 이름 끝에 숫자가 추가됩니다 .
11. **확인**을 클릭합니다 .

테이블 패널의 맨 아래에 해당 스크립트가 추가됩니다.

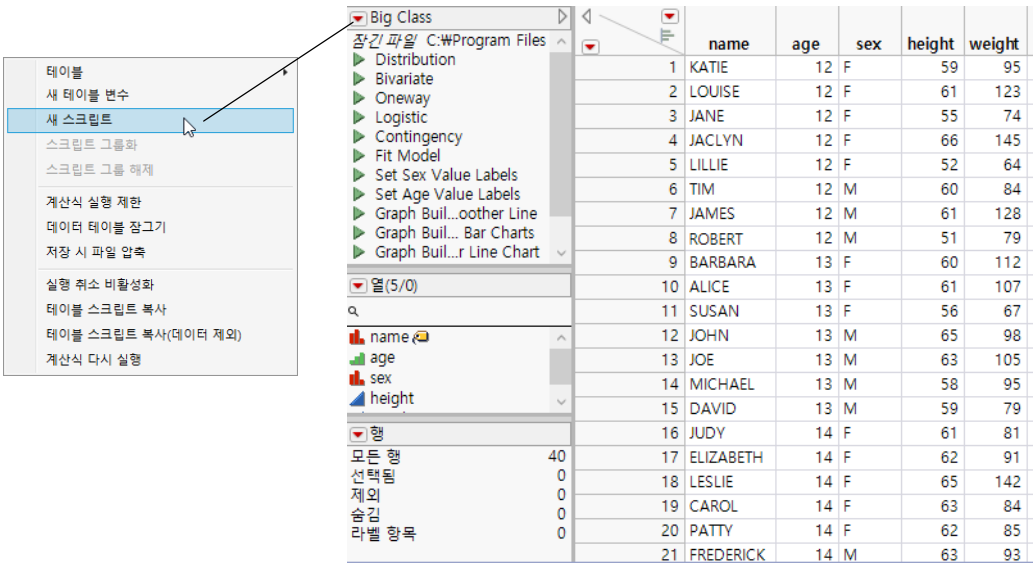
팁 : 해당 데이터 테이블이 열릴 때마다 특정 스크립트가 자동으로 실행되도록 하려면 스크립트 이름을 OnOpen 으로 지정하십시오 . 데이터 테이블에 저장된 스크립트 중 하나만 자동으로 실행되도록 설정할 수 있습니다 . " 모형 적합 " 스크립트에서 스크립트 이름을 **모형** 또는 Model(소문자 가능) 로 지정하면 **분석 > 모형 적합**을 선택할 때 시작 창이 해당 스크립트에 따라 자동으로 채워집니다 .

데이터 테이블에 대한 JSL 스크립트 작성

JSL 을 사용하여 데이터 테이블에 스크립트를 추가하려면

1. 테이블 패널에서 데이터 테이블 이름 왼쪽의 빨간색 삼각형을 클릭합니다 .

그림 4.63 스크립트 생성



2. **새 스크립트**를 선택합니다 .
3. **이름** 옆의 상자에 이름을 입력하여 스크립트 이름을 지정합니다 .

4. **스크립트** 옆의 상자에 JSL 코드를 입력하여 스크립트를 추가합니다 .
5. 다음 작업 중 하나를 수행합니다 .
   - 스크립트에 대해 JSL 디버거를 실행하여 오류가 있는지 확인하려면 **스크립트 디버그**를 클릭합니다 .
   - 스크립트 편집을 마쳤으면 **확인**을 클릭합니다 . 그러면 테이블 패널에 해당 스크립트가 나타나고 창이 닫힙니다 .
   - 스크립트 편집을 마치지 않은 상태로 저장하려면 **저장**을 클릭합니다 . 그러면 테이블 패널에 해당 스크립트가 나타나고 창은 추가 편집을 위해 열린 채로 유지됩니다 .
   - 스크립트를 실행하려면 **실행**을 클릭합니다 .

스크립트 실행 , 편집 , 삭제 또는 복사

데이터 테이블에 저장된 스크립트를 실행 , 편집 , 삭제 또는 복사하려면

1. 테이블 패널에서 스크립트 이름 옆의 녹색 삼각형을 마우스 오른쪽 버튼으로 클릭하거나 스크립트 이름을 마우스 오른쪽 버튼으로 클릭합니다 .
2. 다음 명령 중 하나를 선택합니다 .
   - 스크립트 실행
   - 편집
   - 삭제
   - 복사

스크립트를 복사한 후에는 스크립트 창이나 다른 데이터 테이블의 테이블 패널에 붙여넣을 수 있습니다.

5 장

JMP 열 특성 설정

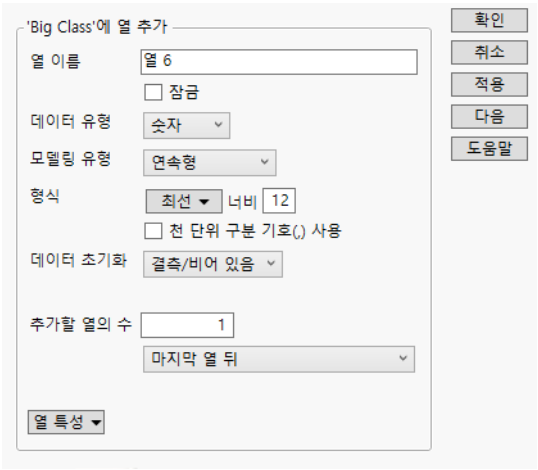
열 정보 창

열 정보 창을 사용하여 데이터 테이블의 열에 특정 특성을 설정할 수 있습니다. 다음은 열에 대해 수행할 수 있는 작업의 몇 가지 예입니다.

- 데이터 및 모델링 유형 변경
- 숫자 형식 변경
- 계산식 추가
- 값 또는 결측값에 대한 제한 지정
- 범주형 값 또는 행 데이터 정렬
- 규격, 관리 또는 반응 한계 저장
- 알려진 시그마 값 입력

여러 열 간에 속성 및 특성을 표준화하고, 열에 사전 선택된 분석 역할을 할당하고, 데이터 테이블의 열을 압축할 수도 있습니다.

그림 5.1 열 정보 창



목차

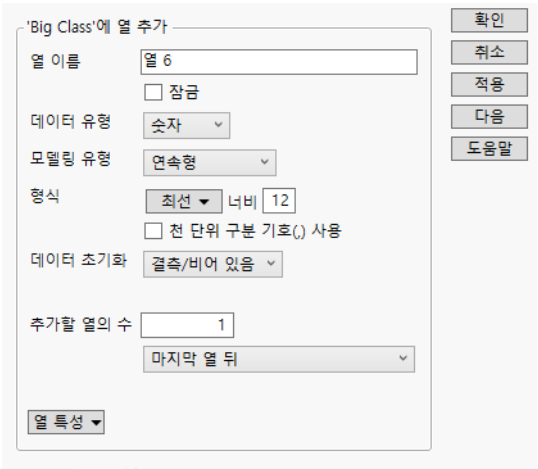
열 정보 창 269
모델링 유형 271
짧은 정수 형식 272
숫자 형식 273
행 상태 열에 정보 저장 277
열의 데이터 초기화 279
JMP 의 열 특성 280
기본 열 특성 281
열 값을 검증하는 특성 282
열 값에 정보를 추가하는 특성 283
열 표시 방식을 제어하는 특성 288
모델링 및 DOE에 사용되는 특성 289
관리도 및 공정 능력과 관련된 특성 290
플랫폼에서 열이 사용되는 방식을 제어하는 특성 291
SAS 특성 299
가상 결합 특성을 사용하여 데이터 테이블 연결 299
추가 열 특성 301
JMP에서 할당 및 제어되는 특성 303
열 간에 속성 및 특성 표준화 305
열 값 재코딩 305
열 속성 표준화 306
열 특성 표준화 306
열 특성 삭제 306
예: 계산식 표준화 306
열에 사전 선택된 분석 역할 할당 308

열 정보 창

열 정보 창을 사용하여 열의 모든 속성 및 특성을 지정할 수 있습니다. 다음 중 한 가지 방법으로 열 정보 창에 액세스할 수 있습니다.

- **열 > 새 열**을 선택합니다.
- 기존 열 머리글을 클릭하고 **열 > 열 정보**를 선택합니다.
- 기존 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다.
- 열 이름 바로 위를 두 번 클릭합니다.
- " 열 " 패널에서 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다.

그림 5.2 열 정보 창



열 정보 창에는 다음과 같은 정보가 포함되어 있습니다.

열 이름 열 이름을 입력하거나 편집합니다.

잠금 모든 값을 편집할 수 없도록 열을 잠급니다. 열을 잠근 후에는 데이터 테이블의 " 열 " 패널에서 해당 열 이름 옆에 잠금 아이콘 ( ) 이 나타납니다. 열에 계산식을 추가하면 열이 자동으로 잠깁니다.

데이터 유형 다음 사항을 결정하는 열 데이터 유형을 선택하거나 변경합니다.

- 데이터 격자에서 열 값의 형식이 지정되는 방식
- 열 값이 내부에 저장되는 방식
- 열 값을 계산에 사용할 수 있는지 여부

자세한 내용은 " 모델링 유형 "(271 페이지) 에서 확인하십시오.

다음 데이터 유형 중에서 선택합니다 .

숫자 이 열에는 숫자만 포함됩니다 (소수점 포함 또는 포함 안 함).

문자 이 열에는 숫자를 포함하여 모든 문자가 포함됩니다 . 문자 열에서 숫자는 문자로 표시되고 연속형 값 대신 이산형 값으로 간주됩니다 . 문자 값의 최대 필드 너비는 32,766 바이트입니다 .

행 상태 이 열에는 행의 제외 , 숨김 , 라벨 지정 , 색상 적용 또는 표식 적용 여부를 나타내는 행 상태 정보가 포함됩니다 . 자세한 내용은 " 행 상태 열에 정보 저장 "(277 페이지) 에서 확인하십시오 .

표현식 이 열에는 단순 입력에서 복잡한 문 시퀀스에 이르는 JSL 표현식이 포함됩니다 . 특히 표현식은 그림 , 행렬 또는 다른 표현식에 대응할 수 있습니다 . 이미지 등의 표현식을 바탕 화면에서 표현식 열의 셀로 드래그하여 놓을 수 있습니다 . 자세한 내용은 " 표현식 역할 "(301 페이지) 에서 확인하십시오 .

참고 :

- 열의 데이터 유형을 문자에서 숫자로 변경하면 모든 문자 값이 결측 데이터 값이 되며 이는 복구할 수 없습니다 .
- 숫자 데이터 유형의 열에 숫자가 아닌 데이터를 입력하면 데이터 유형을 문자로 변경하라는 경고 창이 표시됩니다 . 값을 다시 입력하려면 **다시 시도**를 클릭하여 데이터 테이블로 돌아가십시오 . 데이터 유형을 변경하려면 **변경**을 클릭하십시오 . 편집을 취소하고 셀 값을 되돌리려면 **되돌리기**를 클릭하십시오 .
- 짧은 정수 형식도 사용할 수 있습니다 . 자세한 내용은 " 짧은 정수 형식 "(272 페이지) 에서 확인하십시오 .

모델링 유형 (숫자 또는 문자 데이터 유형만 해당) 열의 모델링 유형을 선택하거나 변경합니다 . 모델링 유형은 분석 중에 열 값을 처리할 방법을 JMP 에 알려 줍니다 . 변수를 다른 방식으로 조사하려는 경우 모델링 유형을 변경할 수 있습니다 . 자세한 내용은 " 모델링 유형 "(271 페이지) 에서 확인하십시오 .

다음 모델링 유형 중에서 선택합니다 .

연속형 숫자 데이터 유형만 해당됩니다 . 연속형 값은 연속형 측정값으로 간주됩니다 . JMP 플랫폼에서는 계산에 숫자 값을 바로 사용합니다 .

순서형 숫자 또는 문자 데이터 유형에 해당됩니다 . JMP 플랫폼에서는 순서형 값을 순서가 있는 이산형 형태의 범주형 값으로 처리합니다 . 그래프 빌더에서는 이산형 값이 50 개를 초과하면 해당 열이 연속형으로 간주됩니다 .

- 열에 숫자 데이터 유형이 있는 경우에는 숫자 크기에 따라 값이 정렬됩니다 .
- 열에 문자 데이터 유형이 있고 값 순서 열 특성이 저장된 경우에는 값 순서 특성에 따라 값이 정렬됩니다 . 그렇지 않은 경우에는 영숫자 데이터 값 순서화에 따라 값이 정렬됩니다 . 하지만 값에 분명한 순서가 있는 특별한 경우에는 값이 그에 따라 자동으로 정렬됩니다 . 자세한 내용은 " 값 순서 "(285 페이지) 에서 확인하십시오 .

명목형 숫자 또는 문자 데이터 유형에 해당됩니다 . JMP 플랫폼에서는 모든 값을 암시적 순서가 없는 이산형 값으로 간주합니다 .

다중 반응 문자 데이터 유형에만 해당됩니다 . 단일 셀의 개별 항목은 쉼표로 구분되어야 합니다 . 다중 반응 열을 지원하는 JMP 플랫폼에서는 쉼표로 구분된 목록의 각 항목을 개별 데이터 값으로 간주합니다 . 항목이 쉼표가 아닌 다른 문자로 구분된 경우에는 " 다중 반응 " 열 특성을 대신 사용하십시오 . 자세한 내용은 " 다중 반응 "(294 페이지) 에서 확인하십시오 .

비정형 텍스트 문자 데이터 유형에만 해당됩니다 . 이러한 값은 일반적으로 고유하며 , 따라서 범주형 분석에는 적합하지 않습니다 . 비정형 텍스트 값을 분석하는 데는 텍스트 탐색기 플랫폼이 이상적입니다 .

벡터 표현식 데이터 유형에만 해당됩니다 . 셀의 항목은 열 또는 행 벡터입니다 . 벡터 모델링 유형을 지원하는 JMP 플랫폼에서는 벡터를 인식하고 계산에서 그에 따라 적절히 처리합니다 .

없음 모든 데이터 유형에 해당됩니다 . 다른 모델링 유형으로 열을 제대로 나타낼 수 없는 경우에는 모델링 유형으로 " 없음 " 을 사용합니다 . 예를 들어 그림 또는 ID 값 열에 " 없음 " 모델링 유형이 할당될 수 있습니다 . JMP 시작 창에서는 모델링 유형이 " 없음 " 인 열에 역할을 할당할 수 없습니다 .

팁 : 열의 모델링 유형을 다중 반응 , 비정형 텍스트 또는 벡터로 변경하려면 열 정보 창을 사용해야 합니다 .

형식 (숫자 데이터 유형만 해당) 숫자 열의 표시 형식을 선택하거나 변경합니다 . 자세한 내용은 " 숫자 형식 "(273 페이지) 에서 확인하십시오 .

데이터 초기화 (새 열을 생성할 때만 표시됨) 열에 표시하려는 초기 데이터 값의 유형을 지정합니다 . 자세한 내용은 " 열의 데이터 초기화 "(279 페이지) 에서 확인하십시오 .

열 특성 (여러 특성의 목록 포함) 열에 특성을 할당합니다 . 자세한 내용은 "JMP 의 열 특성 "(280 페이지) 에서 확인하십시오 .

팁 : 열을 계속 추가하려면 **다음** 버튼을 클릭하십시오 .

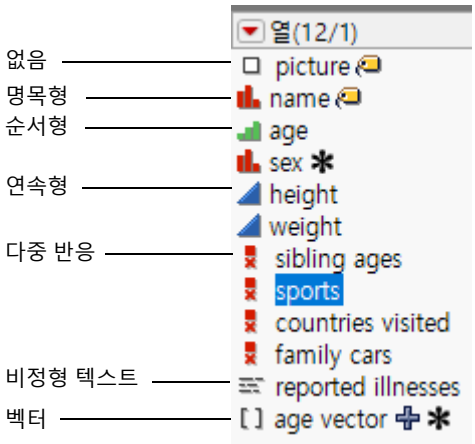
모델링 유형

JMP 데이터 테이블의 열에는 다양한 유형의 정보가 포함될 수 있습니다 . 하지만 단일 열의 모든 정보는 데이터 유형과 모델링 유형이 동일해야 합니다 .

- 데이터를 가져올 때 JMP 는 사용할 데이터 유형과 모델링 유형을 추측합니다 . 따라서 JMP 가 올바르게 추측했는지 확인해야 합니다 .
- 사용자가 수동으로 데이터를 삽입할 경우에는 해당 시점에 데이터 유형과 모델링 유형을 할당해야 합니다 .

그림 5.3에서는 서로 다른 모델링 유형을 식별하는 아이콘을 보여 줍니다 .

그림 5.3 열 패널의 모델링 유형 아이콘



모델링 유형을 **연속형**, **명목형**, **순서형** 또는 **없음**으로 변경하려면 아이콘을 클릭합니다.

자세한 내용은 "모델링 유형"(270 페이지)에서 확인하십시오.

팁 : " 연속형 " 은 데이터 유형이 숫자인 경우에만 선택할 수 있습니다 . 메뉴에서 " 연속형 " 옵션이 흐리게 표시되는 경우 열을 연속형으로 설정하려면 먼저 열 정보 창에서 열의 데이터 유형을 변경해야 합니다 .

짧은 정수 형식

데이터에 올바른 짧은 정수 형식을 사용할 때 숫자가 나타나는 방식에는 눈에 띄는 차이가 없습니다. 하지만 숫자가 차지하는 디스크 공간이 줄어들고 메모리가 덜 사용됩니다. 열 정보 창에 짧은 정수 형식이 나타나도록 하려면 환경 설정에서 짧은 정수 형식을 활성화해야 합니다.

열 정보 창에서 짧은 정수 형식을 사용할 수 있도록 하려면

1. **파일 > 환경 설정**을 선택하고 **테이블**을 클릭합니다 .
2. **짧은 숫자 데이터 형식 허용** 옵션을 선택합니다 .
3. **확인**을 클릭하여 데이터 테이블로 돌아갑니다 .

숫자 데이터를 짧은 정수 형식으로 저장하려면

1. 값을 짧은 정수 형식으로 설정하려는 열의 열 이름 위를 두 번 클릭합니다 .
   " 열 정보 " 창이 나타납니다 .
2. " 데이터 유형 " 메뉴에서 **1 바이트 정수** , **2 바이트 정수** 또는 **4 바이트 정수**를 선택합니다 .

JMP 가 값을 선택된 범위의 정수로 저장합니다 . 숫자의 예는 다음과 같습니다 .

- 1 바이트 정수의 경우 입력할 수 있는 수의 범위는 -126 부터 127 까지입니다 .

- 2 바이트 정수의 경우 입력할 수 있는 수의 범위는 -32,766 부터 32,767 까지입니다 .
- 4 바이트 정수의 경우 입력할 수 있는 수의 범위는 -2,147,483,646 부터 2,147,483,647 까지입니다 .

숫자 형식

숫자 열의 경우 열 정보 창에 " 형식 " 메뉴가 나타납니다. JMP에서 열에 숫자를 표시할 방식을 지정하려면 형식을 지정합니다. 모든 형식 옵션에 대해 해당 열의 셀에 수용 가능한 총 문자 수를 지정할 수 있습니다. 자세한 내용은 " 너비 지정 "(273 페이지) 에서 확인하십시오.

형식 옵션에 대한 자세한 내용은 " 숫자 형식 옵션 "(273 페이지) 에서 확인하십시오.

팁 : 천 이상의 값에 쉼표를 추가하려면 **천 단위 구분 기호 사용** 옵션을 선택하십시오 . " 너비 " 상자에서 쉼표 하나당 공백을 하나씩 추가해야 합니다 . 그러지 않으면 쉼표가 표시되지 않습니다 . 이 옵션은 " 최적 ", " 고정 소수점 ", " 백분율 " 및 " 통화 " 형식에 사용할 수 있습니다 .

너비 지정

"너비" 필드에 숫자를 지정할 때는 가능한 문자의 총 수를 포함해야 합니다. 문자에는 숫자, 소수점, 쉼표 및 통화 기호가 포함됩니다.

숫자 형식 옵션

다음과 같은 숫자 형식 옵션 중에서 선택합니다 .

최적 JMP 가 각 셀 값의 정밀도를 고려하여 가장 적절한 값 표시 방법을 선택할 수 있도록 합니다 . 기본적으로 열의 물리적 너비는 12 자입니다 .

고정 소수점 열의 모든 값을 지정한 소수 자릿수의 근사값으로 표시합니다 .

- 정수만 표시하려면 소수 자릿수를 0 으로 설정합니다 .
- 소수점 다음의 정수 개수가 지정한 소수 자릿수보다 적으면 해당 소수 자릿수에 도달할 때까지 0 이 추가됩니다 . 예를 들어 값이 1.23 인 경우 " 소수 " 상자에 **5** 를 입력하면 소수 자릿수가 5 개인 1.23000 이 표시됩니다 .

백분율 숫자 값에 100 을 곱하고 숫자 다음에 백분율 기호를 표시합니다 .

p 값 확률값을 표시합니다 . 너비 기본값은 12 입니다 . 0.0001 보다 작은 숫자는 <.0001 로 표시됩니다 . 이 형식은 주로 JSL 스크립트에서 사용되며 데이터 테이블 열에는 거의 필요하지 않습니다 .

과학적 숫자를 표준 과학적 표기법으로 표시합니다 . 숫자 123456 을 입력하면 1.23456e+5 로 표시됩니다 . 소수를 표시하려면 " 소수 " 를 선택하고 소수 자릿수를 입력합니다 .

엔지니어링 " 과학적 " 과 유사하지만 지수가 항상 3 의 배수이며 " 소수 " 필드는 유효 숫자를 나타냅니다 .

엔지니어링 SI " 엔지니어링 " 과 동일하지만 지수가 SI 기호로 바뀝니다 .

정밀도 숫자를 지정된 유효 숫자 수를 갖는 근사값으로 계산합니다. " 소수 " 필드에 유효 숫자의 수를 지정합니다.

통화 소수 자릿수 2, 천 단위 구분 기호 및 컴퓨터 로케일 설정에 지정된 통화 기호를 사용하여 숫자 형식을 지정합니다. **통화** 형식의 기본 너비는 15 입니다. 필드 너비가 더 넓어야 하는 숫자가 있는 경우 기본적으로 **최적** 형식이 지정됩니다. 이 형식이 할당되면 해당 열과 해당 열이 포함된 그래프에 통화 기호가 나타납니다.

날짜 열의 모든 값을 날짜로 표시합니다. 자세한 내용은 " 날짜 형식 "(274 페이지) 에서 확인하십시오.

시간 열의 모든 값을 특정 시간 인스턴스 (예 : 12/2/03 2:23 PM) 로 표시합니다. 자세한 내용은 " 시간 형식 "(275 페이지) 에서 확인하십시오.

기간 열의 모든 값을 기간 (예 : 시간 , 분 및 초) 으로 표시합니다.

:day:hr:m, :day:hr:m:s 52:03:01:30(52 일 3 시간 1 분 30 초) 과 같은 기간을 표시합니다.

hr:m, hr:m:s, min:s 17:37(17 시간 37 분) 과 같은 기간을 표시합니다.

지리 지도용 위도 및 경도 숫자 형식을 표시합니다. 위도 및 경도 옵션에는 다음이 포함됩니다.

DDD 도

DMM 도 , 분

DMS 도 , 분 , 초

각 형식에서 마지막 필드에는 소수 부분이 있습니다. 부호 있는 도 (Degree) 필드나 방향 접미사를 사용하여 방향을 지정할 수 있습니다. *59°00′00″* 와 같이 부호 있는 도 (Degree) 필드를 표시하려면 **방향 표시자**를 선택 취소합니다. *59°00′00″ S* 와 같이 방향 접미사를 표시하려면 **방향 표시자**를 선택합니다.

공백을 필드 구분 기호로 사용하려면 **필드 구두점**을 선택 취소합니다. 도 , 분 및 초 기호를 사용하려면 **필드 구두점**을 선택합니다.

사용자 정의 열 너비와 소수 자릿수의 사용자 정의를 비롯한 숫자 열에 대한 사용자 형식을 정의할 수 있습니다. **사용자 정의**를 선택하고 **사용자 형식 설정**을 클릭한 후 계산식 편집기 창에서 형식을 정의합니다. 예를 들어 숫자 옆에 백분율 기호를 표시하지 않으려면 이 옵션을 선택하고 숫자에 100 을 곱합니다. " 백분율 " 형식을 선택하는 대신 이 방법을 사용할 수 있습니다 . 또는 , 미터 단위의 데이터를 단위와 함께 킬로미터 단위로 표시하려면 계산식 `Char(:value / 1000) || "km"` 을 추가합니다. " 날짜 " 또는 " 시간 " 형식 목록에서는 사용할 수 없는 날짜 / 시간 형식 (예 : D/M/Y) 을 적용할 수도 있습니다.

형식 패턴 숫자 값의 날짜 / 시간 형식을 사용자 정의할 수 있습니다. 자세한 내용은 " 형식 패턴으로 날짜 / 시간 형식 사용자 정의 "(276 페이지) 에서 확인하십시오.

날짜 형식

"날짜" 형식을 선택할 경우 "입력 형식"도 지정할 수 있습니다. "날짜" 형식은 데이터 테이블 셀에서 날짜가 나타나는 방식을 나타내고, "입력 형식"은 날짜를 입력하는 방식을 나타냅니다.

이미 데이터가 포함되어 있는 숫자 열에 날짜 형식을 할당하면 숫자 값이 1904 년 1 월 1 일 이후 경과한 초 수로 처리됩니다. 예를 들어 셀 값 1,234,567,890 이 포함된 숫자 열에서 형식을 **날짜 > m/d/y** 로 변경하면 해당 셀 값이 02/13/1943 으로 나타납니다.

표 5.1 의 예에 사용된 날짜는 2004 년 12 월 31 일입니다.

표 5.1 날짜 형식

형식	표시 형식
m/d/y	12/31/2004
mmddyyyy	12312004
m/y	12/2004
yyyyQq	2004Q4
d/m/y	31/12/2004
ddmmyyyy	31122004
ddMonyyyy	31Dec2004
Monddyyyy	Dec312004
y/m/d	2004/12/31
yyyymmdd	20041231
yyyy-mm-dd	2004-12-31
자세한 날짜	2004 년 12 월 31 일 금요일
간단한 날짜	2004-12-31
로케일 날짜	로컬 OS 설정에 따라 달라집니다. 예를 들어 미국의 로컬 OS 설정은 mm/dd/yyyy(12/31/2004) 입니다.

참고: 데이터 테이블의 날짜 표시 방식을 변경하지 않고 그래프의 날짜 표시 방식을 변경하는 방법에 대한 자세한 내용은 "JMP 보고서 " 장의 " 축의 숫자 형식 변경 "(457 페이지) 에서 확인하십시오 .

시간 형식

"시간" 형식을 선택할 경우 "입력 형식"도 지정할 수 있습니다. "시간" 형식은 데이터 테이블 셀에서 시간이 나타나는 방식을 나타내고, "입력 형식"은 시간을 입력하는 방식을 나타냅니다.

- 다음 날짜 형식에 대해 전 날짜의 자정 이후 경과한 시간을 나타내는 시간 , 분 , 초 단위 수를 추가할 수 있습니다 .
  - m/d/y
  - d/m/y
  - y/m/d
  - ddMonyyyy
  - Monddyyyy
  - 로케일 날짜

  예를 들어 2004 년 12 월 31 일의 숫자 값이 3,187,296,600 이면 이 값은 2004 년 12 월 31 일 오전 12:10 을 나타냅니다 .
- **:day:hr:m** 및 **:day:hr:m:s** 는 1904 년 1 월 1 일 이후 경과한 날짜 , 시간 , 분 및 초 단위 수를 표시합니다 . 예를 들어 2004 년 12 월 31 일은 :36890:00:10: 및 :36890:00:10:00 으로 표시됩니다 .
- **h:m:s** 및 **h:m** 은 날짜 필드에 있는 날짜의 시 , 분 및 초 단위 부분을 표시합니다 . 예를 들어 2004 년 12 월 31 일 오전 12:10 은 오전 12:10:00 및 오전 12:10 으로 표시됩니다 .
- **yyyy-mm-ddThh:mm** 및 **yyyy-mm-ddThh:mm:ss** 는 연도 , 월 , 일 및 시간을 표시합니다 (예 : 2004-12-31T12:10:00). T 는 리터럴 값으로 , 그 자체를 나타냅니다 .

참고 : 데이터 테이블의 시간 표시 방식을 변경하지 않고 그래프의 시간 표시 방식을 변경하는 방법에 대한 자세한 내용은 "JMP 보고서 " 장의 " 축의 숫자 형식 변경 "(457 페이지) 에서 확인하십시오 .

형식 패턴으로 날짜 / 시간 형식 사용자 정의

형식 패턴은 날짜/시간 형식을 정의하는 문자열입니다(예 : <YYYY></><MM></><DD> <hh><:><mm><:><ss><ampm>). 패턴에서 꺾쇠 괄호로 묶인 부분을 **필드 설명자**라고 합니다. 필드 설명자는 값 (예 : <YYYY> 4 자리 연도) 또는 기타 날짜/시간 텍스트 (예 : </> 로케일별 날짜 구분 기호) 를 나타냅니다.

형식 패턴을 사용하면 JMP에서 제공하지 않는 형식을 생성할 수 있습니다. 이러한 형식은 데이터 형식 지정 및 입력에 모두 사용할 수 있습니다. 패턴을 직접 입력하거나 기본 제공 형식을 시작 위치로 사용하거나 목록에서 필드 설명자를 선택할 수 있습니다.

기본 제공 날짜/시간 형식을 선택할 수 있는 모든 위치에서 형식 패턴을 사용할 수 있습니다.

형식 패턴 생성

1. 데이터 테이블에 숫자 열을 생성합니다 .
2. 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다 .
3. " 형식 " 목록에서 **형식 패턴**을 선택합니다 .
4. **형식 패턴 설정**을 클릭합니다 .

5. 기본 제공 형식을 시작 위치로 사용하려면 **기본 제공**을 선택하고 날짜 / 시간 형식을 선택한 다음 형식을 편집합니다 .
6. 형식을 편집하려면 " 형식 패턴 " 필드에 필드 설명자를 입력하거나 " 필드 설명자 " 목록에서 형식 패턴을 선택합니다 .
7. **확인**을 클릭합니다 .

입력 형식 패턴 생성

1. 기존 숫자 열에서 마우스 오른쪽 버튼을 클릭하고 **열 정보**를 선택합니다 .
2. " 형식 " 목록에서 **형식 패턴**을 선택한 다음 **출력 형식 사용**을 선택 취소합니다 .
3. **형식 패턴 설정**을 클릭합니다 .
4. 이전 작업에 설명된 대로 패턴을 생성합니다 .

참고 : 일부 형식은 입력에 사용할 수 없습니다 . 예를 들어 가변 길이 숫자 필드 설명자 바로 뒤에 다른 숫자가 오면 일부 날짜가 모호해집니다 . <YYYY><M><D> 형식은 <M> 이 단일 숫자 월이기 때문에 입력을 처리할 수 없습니다 . 예를 들어 2020111 같은 날짜가 모호할 수 있습니다 . 이 날짜는 1 월 11 일이거나 11 월 1 일일 수 있습니다 .

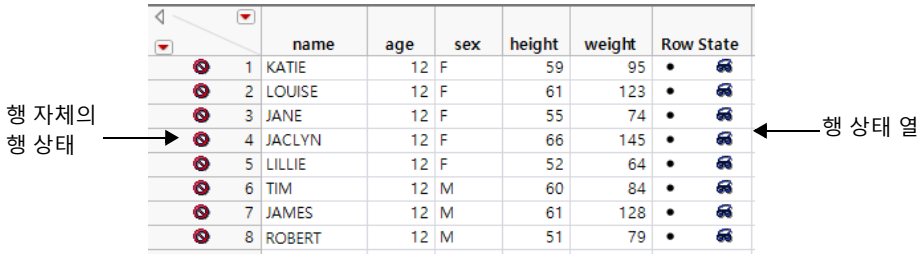
국가별 형식

국가 표준에 따라 달라지는 형식이 포함된 데이터를 가져오거나 입력할 경우 숫자 형식이 올바르게 해석되는지 확인해야 할 수 있습니다. Windows의 경우, 제어판의 "국가 및 언어" 옵션에 액세스하여 숫자 형식을 지정할 때 기준이 되는 국가를 선택합니다. macOS의 경우, Apple 메뉴에서 **Language & Region > Formats**를 선택하고 올바른 국가를 선택합니다. 최근 버전의 macOS에서는 이 옵션이 "System Preferences" > "Language and Region"에 나타날 수 있습니다.

행 상태 열에 정보 저장

행에 행 상태를 할당하는 것과 마찬가지로, 행 상태 정보만 포함된 열을 생성할 수 있습니다. 행 상태 열에는 행의 제외, 숨김, 라벨 지정, 색상 적용, 표식 적용 또는 선택 여부에 대한 정보가 저장됩니다. 열을 행 상태 열로 지정하려면 열 정보 창의 "데이터 유형" 옆에서 **행 상태**를 선택합니다.

그림 5.4 행 자체의 행 상태와 행 상태 열



행 상태 열에는 행 상태가 저장되므로 나중에 행 상태를 다시 적용할 수 있습니다. 행 상태 열을 채우려면 현재 행 상태에서 행 상태를 복사하거나 열 계산식을 사용합니다.

행 상태 열을 생성하려면

1. **열 > 새 열**을 선택합니다.
2. " 데이터 유형 " 옆에서 **행 상태**를 선택합니다.
3. **확인**을 클릭합니다.

새 행 상태 정보로 셀을 채우거나, 행의 기존 행 상태 정보를 복사할 수 있습니다.

새 행 상태 정보로 셀을 채우려면

1. 행 상태 열의 일부 행만 채우려면 해당 행을 강조 표시합니다. 열의 모든 행을 채우려면 행 상태 열을 강조 표시합니다.
2. 마우스 오른쪽 버튼을 클릭하고 **행 상태 셀**을 선택합니다.
3. 적용할 행 상태를 선택합니다.

기존 행 상태 정보를 복사하려면

1. 행 상태 열의 일부 행만 채우려면 해당 행을 강조 표시합니다. 열의 모든 행을 채우려면 행 상태 열을 강조 표시합니다.
2. " 열 " 패널에서 열 이름 옆에 있는 별 모양 아이콘 ( ) 을 클릭합니다.
3. 다음 중 하나를 선택합니다.

   행 상태로부터 복사 해당 열의 행 상태를 행 자체의 행 상태로 바꿉니다.

   행 상태로부터 추가 행 자체의 행 상태를 행 상태 열에 추가합니다.

   행 상태로 복사 행 자체의 행 상태를 해당 열의 행 상태로 바꿉니다.

   행 상태에 추가 행 상태 열의 행 상태를 행 자체의 행 상태에 추가합니다.

셀 영구 선택

다른 행 상태 특성(숨김, 제외, 색상 등)을 저장할 때와 마찬가지로 선택 사항을 행 상태 열에 저장할 수 있습니다. 그러면 셀에 "영구" 강조 표시가 적용됩니다.

셀을 영구적으로 선택하려면

1. 셀을 마우스 오른쪽 버튼으로 클릭하고 **행 상태 셀 > 선택 / 선택 취소**를 선택합니다.
2. 선택할 셀의 수만큼 각 셀에 대해 이 단계를 반복합니다.
3. 강조 표시를 제거하려면 셀을 마우스 오른쪽 버튼으로 클릭하고 **행 상태 셀 > 선택 / 선택 취소**를 선택합니다.

열의 데이터 초기화

데이터 테이블에 새 열을 처음 추가할 때는 열 정보 창에 " 데이터 초기화 " 메뉴가 나타납니다 . 새 열에 표시하려는 초기 데이터 값의 유형을 지정할 수 있습니다 . " 행 상태 " 데이터 유형에는 데이터 초기화 기능을 사용할 수 없습니다 .

다음 옵션 중 하나를 선택합니다.

결측 / 비어 있음 열에 결측값을 넣습니다 . 숫자 데이터의 경우 검은색 점 (•), 문자 데이터의 경우 공백이 결측값을 나타냅니다 .

상수 열의 모든 행에 특정 수 또는 문자를 넣습니다 . 표시되는 상자에 수 또는 문자를 입력합니다 . 임의 개수의 문자를 입력할 수 있습니다 .

오늘 열의 각 행에 오늘에 해당하는 타임스탬프를 넣습니다 . 이 옵션은 **날짜** 또는 **시간** 형식에만 사용할 수 있습니다 .

시퀀스 데이터 사용자가 지정하는 파라미터에 따라 순차적 데이터를 삽입합니다 . 자세한 내용은 " 숫자 또는 문자 시퀀스 데이터 "(280 페이지) 에서 확인하십시오 .

난수 무작위로 생성된 데이터 값을 열에 삽입합니다 . 다음 중에서 난수 값 생성 방법을 선택합니다 .

정수 난수 최소 및 최대 정수 값을 입력합니다 . 이 범위 내에서 정수 값이 생성되며 각 값은 거의 같은 빈도로 발생하게 됩니다 .

균등 난수 연속형 값의 범위를 입력합니다 . 이 범위 내에서 균등 난수 값이 생성됩니다 .

정규 난수 정규 분포의 평균 및 표준편차를 입력합니다 . 이 분포를 따르는 정규 난수 값이 생성됩니다 .

표시자 난수 최대 세 개의 값과 각 값의 원하는 비율을 입력합니다 . 비율의 합은 1 이어야 합니다 . 지정한 비율과 가능한 한 근사한 분포를 갖도록 값이 생성됩니다 .

참고 : 이 방법을 사용할 경우 행 수에 비율 중 하나를 곱한 값이 분수 값이 될 때만 제외하고는 지정한 것과 정확히 일치하는 비율로 값이 생성됩니다 .

예를 들어 n 개의 행이 있고 세 개의 비율이 각각 p_1, p_2 및 p_3 라면 다음은 값이 생성되는 방식입니다 .

- 각 행에 0 에서 1 사이의 균등 난수 값이 할당됩니다 .
- 균등 난수 값에 따라 행이 정렬됩니다 .
- 가장 작은 근사값 (np_1) 에 해당하는 행의 균등 난수 값에 첫 번째 값이 할당되고 , 다음으로 작은 근사값 (np_2) 에 해당하는 행의 균등 난수 값에 두 번째 값이 할당되며 , 나머지 행에는 세 번째 값이 할당됩니다 .

숫자 또는 문자 시퀀스 데이터

숫자 데이터에 대해 순차적 데이터를 삽입하려면

1. **데이터 유형** 옆에 **숫자**가 선택되어 있는지 확인합니다 .
2. **데이터 초기화**에서 **시퀀스 데이터**를 선택합니다 .
3. **시작** 및 **끝** 상자에 시작점과 끝점을 할당합니다 .
4. **단계** 상자에 시퀀스를 할당합니다 .
5. (선택 사항) **각 값을 N 번 반복** 상자에 각 숫자 값을 반복할 횟수를 입력합니다 .
6. **확인**을 클릭합니다 .

예를 들어 열에 2부터 60까지의 짝수를 포함하려면 **시작** 상자에 **2**를 입력하고 **끝** 상자에 **60**을 입력한 후 **단계** 상자에 **2**를 입력합니다.

문자 데이터에 대해 순차적 데이터를 삽입하려면

1. **데이터 유형** 옆에 **문자**가 선택되어 있는지 확인합니다 .
2. **데이터 초기화**에서 **시퀀스 데이터**를 선택합니다 .
3. **추가** 옆의 상자에 문자 데이터를 입력하고 **추가**를 클릭합니다 .
4. (선택 사항) **각 값을 N 번 반복** 상자에 각 문자 값을 반복할 횟수를 입력합니다 .
5. **확인**을 클릭합니다 .

JMP 의 열 특성

열 특성은 열에 연결된 정보입니다. 이 정보는 입력 시 데이터 테이블의 일부로 저장됩니다. 열 특성을 선택하여 열에 적용할 수 있습니다. 사용자 편의를 위해 JMP에서 자동으로 생성되는 열 특성이 몇 가지 있습니다. 종종 이러한 열 특성을 수정할 수 있습니다.

열 특성은 열 정보 창의 "열 특성" 메뉴에 나열됩니다. "열 특성" 메뉴 아래의 영역에는 열에 저장된 열 특성이 표시됩니다. 특성을 보거나 편집하려면 이 목록에서 특성을 선택합니다.

데이터 테이블의 " 열 " 패널에서 " 노트 " 특성 외의 특성이 포함된 각 열의 이름 옆에는 아이콘이 나타납니다 . " 노트 " 특성은 아이콘으로 표시되지 않습니다 . 아이콘은 다음과 같습니다 .

- 범위 또는 목록 확인 특성이 적용되었음을 나타냅니다 .
- 열에 계산식이 포함되어 있음을 나타냅니다 .
- 열에 " 노트 ", " 범위 확인 " 또는 " 목록 확인 " 열 특성이 포함되어 있음을 나타냅니다 .

더 많은 열 특성 아이콘에 대한 자세한 내용은 "JMP 시작하기" 장의 "열 특성을 나타내는 아이콘"(41 페이지) 에서 확인하십시오.

하나 이상의 열에 열 특성을 할당하려면

1. 특성을 할당할 열을 하나 이상 선택합니다 .
2. 다음 중 하나를 수행합니다 .
   - 머리글 영역을 마우스 오른쪽 버튼으로 클릭하고 **열 특성**을 선택한 후 특성을 선택합니다 .
   - 머리글 영역을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 후 " 열 특성 " 메뉴에서 특성을 선택합니다 .
   - **열 > 열 정보**를 선택한 후 " 열 특성 " 메뉴에서 특성을 선택합니다 .
3. 표시된 열 특성 패널에서 값을 지정하고 적절한 옵션을 선택합니다 .

   적용을 클릭하여 열 특성을 추가하거나 , **확인**을 클릭하여 열 특성을 추가하고 열 특성 창을 닫습니다 .

열에 이미 포함된 특성을 다른 열에 적용하려는 경우가 있습니다. 해당 특성을 다른 열에 적용하려면 "속성 표준화" 명령을 사용합니다. 자세한 내용은 "열 간에 속성 및 특성 표준화"(305 페이지)에서 확인하십시오.

다음 섹션에서는 열에 추가할 수 있는 특성에 대해 설명합니다.

기본 열 특성

일반적으로 적용되는 기본 열 특성으로는 " 계산식 " 과 " 노트 " 가 있습니다 .

계산식

열에 해당 열의 값을 계산하는 계산식을 삽입합니다 . 계산식이 추가된 후에는 데이터 값을 수동으로 편집할 수 없도록 열이 잠깁니다 (계산식 무효화 방지).

- 계산식을 생성하려면 **계산식 편집**을 클릭합니다. 계산식 생성에 대한 자세한 내용은 "JMP 의 계산식 생성 " 장 (359 페이지) 에서 확인하십시오 .
- JMP 가 계산식을 실행하지 않도록 하려면 **실행 제한**을 클릭합니다 .
- JMP 가 계산식의 오류에 대해 알리지 않도록 하려면 **오류 무시**를 클릭합니다 .
- 계산식 생성 후 :
  - 열 정보 창의 오른쪽에 계산식이 표시됩니다 . 하지만 계산식이 긴 경우에는 계산식의 일부만 표시될 수 있습니다 . 계산식 상자의 테두리를 클릭한 채로 드래그하여 크기를 조정할 수 있습니다 .
  - 데이터 테이블에서 " 열 " 패널의 열 이름 옆에 있는 아이콘 ( ) 을 클릭하면 계산식을 편집할 수 있습니다 .

팁 : 계산식을 생성할 때 열 정보 창을 사용하지 않으려면 열을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하십시오 .

노트

선택한 열에 노트를 추가합니다.

열 값을 검증하는 특성

"범위 확인", "목록 확인" 및 "결측값 코드" 특성은 열의 값을 검증하는 데 유용합니다.

범위 확인

범위 확인은 열의 데이터를 검증하는 기능입니다. 열에서 지정한 범위에 속하는 숫자만 허용되도록 설정할 수 있습니다.

범위를 설정하는 데 사용할 계산식을 선택해야 합니다 . x 는 열에 입력되는 값이고 , a 는 범위 시작 값 , b 는 범위 끝 값입니다 .

- a = 열에 허용되는 가장 낮은 값
- b = 열에 허용되는 가장 높은 값
- 단일 방향 범위 확인의 경우에는 a 또는 b 를 비워 둡니다 .
- 데이터 테이블에서 " 열 " 패널의 열 이름 옆에 있는 아이콘 ( ) 을 클릭하면 범위 확인 특성을 수정할 수 있습니다 .

목록 확인

목록 확인은 열의 데이터를 검증하는 기능입니다 . 열에서 지정한 개별값만 허용되도록 설정할 수 있습니다 . 목록 확인은 그래프 또는 그림에서 데이터를 정렬할 방법을 지정하려는 경우에 유용합니다 .

- 최대 127 개의 값을 사용할 수 있습니다 .
- 버튼을 사용하여 새 값을 추가하고, 값 순서를 변경하거나 되돌리고, 값을 제거할 수 있습니다.
- 열에 목록 확인 특성이 설정되면 커서가 해당 셀 위에 놓일 때의 모양이 으로 바뀝니다 . 검증 목록에 포함되지 않은 값을 입력하려고 하면 경고 메시지가 나타납니다 .
- 허용 가능한 값의 메뉴를 보려면 셀을 마우스 오른쪽 버튼으로 클릭하고 **목록 확인 값**을 선택합니다 . 셀에 값을 직접 입력하지 않고 메뉴에서 셀 값을 선택할 수 있습니다 .
- 데이터 테이블에서 " 열 " 패널의 열 이름 옆에 있는 아이콘 ( ) 을 클릭하면 목록 확인 특성을 수정할 수 있습니다 .

결측값 코드

결측값 코드는 결측값으로 처리해야 할 열 값을 지정하는 데 사용합니다. 예를 들어 값 99가 결측값을 나타내는 자리 표시자로 사용되는 경우도 있고, 몇 개의 값이 서로 다른 유형의 결측값을 나타내는 데 사용되는 경우도 있습니다.

참고 : " 결측값 코드 " 열 특성은 확장 속성입니다 . SAS 9.4 서버로 내보내거나 " 다른 이름으로 저장 " 을 통해 SAS 9.4 파일 형식으로 저장하려면 "SAS Integration" 환경 설정 페이지에서 확장 속성을 활성화해야 합니다. 자세한 내용은 "JMP 환경 설정" 장의 "SAS Integration의 환경 설정"(663 페이지) 에서 확인하십시오 . 또는 " 다른 이름으로 저장 " 창에서 **테이블 및 열 특성을 SAS 9.4 확장 속성에 저장**을 선택해야 합니다 . 확장 속성이 선택되어 있지 않으면 JMP 는 결측값 코드를 결측값으로 내보내거나 저장합니다 . 확장 속성은 SAS 9.4 에서만 지원됩니다 . 이전 버전의 SAS 에서는 모든 확장 속성이 무시됩니다 .

단일 열의 결측값 코드를 제거하려면

1. 데이터 테이블의 " 열 " 패널에서 열 옆의 ✱ > **결측값 코드**를 클릭합니다 .
2. 목록에서 결측값 코드가 강조 표시됩니다 . **제거**를 클릭합니다 .
3. **확인**을 클릭합니다 .

여러 열에서 결측값 코드를 제거하려면

1. 삭제하려는 결측값 코드가 포함된 열을 선택합니다 .
2. **열** > **속성 표준화**를 선택합니다 .
3. " 특성 삭제 " 에서 **열 특성** > **결측값 코드**를 클릭합니다 .
4. **확인**을 클릭합니다 .

열 값에 정보를 추가하는 특성

다음 열 특성은 열 값에 보고서와 그림에서 사용되는 정보를 추가합니다.

- 값 라벨
- 값 점수
- 값 순서
- 행 순서 수준
- 값 색상
- 색상 그래디언트
- 짧은 이름
- SPSS 이름
- SPSS 라벨
- SAS 이름
- SAS 라벨
- 긴 이름
- Question

값 라벨

데이터 테이블에서 값 대신 라벨을 표시하려면 값 라벨을 사용합니다. 각각의 값 인스턴스마다 라벨이 하나씩 나타납니다. 셀 내의 라벨을 두 번 클릭하면 원래 값을 표시할 수 있습니다.

- " 값 " 상자에 라벨을 할당할 값을 입력합니다 .
- " 라벨 " 상자에 표시할 라벨을 입력합니다 .
- 범위를 사용하려면 **허용 범위**를 클릭한 후 하한 및 상한 값을 지정합니다 . " 허용 범위 " 가 선택된 경우 값 라벨은 정수가 아닐 수도 있습니다 . 그래프 빌더에서는 X 축과 Y 축에 값 라벨을 사용합니다 .

팁 : 결측값에 라벨을 할당하려면 하한으로 마침표 (**.**) 를 입력하고 상한은 비워 두십시오 . 다른 모든 값에 라벨을 할당하려면 하한으로 세 개의 마침표 (**...**) 를 입력하고 상한은 비워 두십시오 .

- 라벨을 추가 , 변경 또는 제거할 수 있습니다 .

팁 :

- 설정한 값 라벨을 삭제하지 않고 데이터 테이블에서 값 라벨을 해제하려면 " 열 특성 " 창에서 **값 라벨 사용**을 선택 취소하십시오 . 데이터 테이블에서 값 라벨을 빠르게 표시하거나 숨기려면 열을 마우스 오른쪽 버튼으로 클릭하고 **값 라벨 사용**을 선택 또는 선택 해제하십시오 .
- 데이터 테이블에 값 라벨이 포함된 경우 **검색** 명령을 사용하면 실제값이 검색되고 라벨은 검색되지 않습니다 .
- 데이터 테이블에 값 라벨이 포함되어 있으면 **행 편집기**에 라벨이 표시되며 , 셀이 편집을 위해 강조 표시되어 있으면 행 편집기에 실제값이 표시됩니다 .
- 값 라벨이 있는 셀을 복사하여 붙여넣으면 실제값이 붙여넣어집니다 .
- 계산식에서 값 라벨을 사용하여 열을 참조하는 경우 값 라벨을 마우스로 가리키면 실제 데이터 값이 표시됩니다 .

값 점수

범주형 데이터 열의 값-점수 쌍을 나타내려면 값 점수를 사용합니다. 값은 데이터 값이고 점수는 숫자입니다. 이 특성은 데이터 값을 점수와 연결합니다. 예를 들어 열의 데이터 값이 "not satisfied", "satisfied", "very satisfied" 인 경우, "not satisfied" 에 0 점, "satisfied" 에 50 점, "very satisfied" 에 100 점의 점수를 할당할 수 있습니다. 그런 다음 평균을 계산하는 등의 계산 용도로 이러한 점수를 사용할 수 있습니다. 자세한 내용은 Consumer Research 의 에서 확인하십시오.

값 점수를 사용하려면

1. 데이터 테이블에서 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다 .
2. **열 특성** 목록에서 **값 점수**를 선택합니다 .
3. " 값 " 상자에 라벨을 할당할 값 (예 : "satisfied") 을 입력합니다 .

4. " 점수 " 상자에 표시할 라벨 (예 : "50") 을 입력합니다 .
5. **추가** , **변경** 또는 **제거**를 클릭합니다 .
6. **확인**을 클릭합니다 .

값 순서

" 값 순서 " 는 열의 값에 보고서 및 그림에서 사용되는 정렬 방식을 할당합니다 . 보고서 및 그림의 기본 정렬 방식은 사전순 데이터 값 순서화입니다 . 단 , 값이 자동으로 적절하게 정렬되는 다음과 같은 경우는 예외입니다 .

- 1 월 , 2 월 , 3 월 , 4 월 , 5 월 , 6 월 , 7 월 , 8 월 , 9 월 , 10 월 , 11 월 , 12 월
- 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
- 일요일 , 월요일 , 화요일 , 수요일 , 목요일 , 금요일 , 토요일
- 매우 낮음 , 낮음 , 중간 낮음 , 중간 , 중간 높음 , 높음 , 매우 높음
- 매우 비동의 , 비동의 , 보통 , 관심 없음 , 동의 , 매우 동의
- 실패 , 허용 불가 , 매우 나쁨 , 나쁨 , 조금 나쁨 , 허용 가능 , 평균 , 조금 좋음 , 좋음 , 매우 좋음 , 탁월 , 최상

팁 : 값 순서를 익명화하여 값 순서 특성의 번호가 다시 매겨지도록 할 수 있습니다 . 이렇게 하려면 " 값 순서 " 열 특성이 있는 열을 강조 표시하고 **테이블 > 익명화**를 선택하십시오 .

DOE 플랫폼을 사용하여 설계를 구성할 경우 범주형 요인에 "값 순서" 열 특성이 자동으로 할당됩니다. "값 순서" 열 특성을 할당 및 편집하는 예는 실험 설계 가이드의 에서 확인하십시오.

참고 : 문자 데이터를 비교하는 경우에는 순서를 결정할 때 로케일 정보를 고려하도록 **문자에 로케일 비교 사용**을 선택하십시오 .

사용자 순서

기본적으로 "사용자 순서"가 선택되어 있습니다. 이 옵션을 사용하면 데이터 값을 사용자가 정의하는 순서로 순서화할 수 있습니다. 기본적으로 문자 데이터는 사전순으로 순서화되고 숫자 데이터는 숫자순으로 순서화됩니다. 화살표를 사용해서 값과 더하기 기호를 이동하여 값을 추가할 수 있습니다.

정렬된 순서

값을 "정렬된 순서" 상자로 이동하려면 **사용자 순서**의 선택을 취소합니다. 그런 다음 정렬된 항목에 대한 규칙을 지정합니다.

사용자 순서를 따르지 않는 항목은 정렬된 순서의 영향을 받습니다. 항목을 사용자 순서에서 제거하고 해당 항목에 정렬 규칙이 적용되도록 하려면 항목을 아래의 "정렬된 순서" 영역으로 드래그합니다. 모든 항목에 정렬 규칙을 적용하려면 **사용자 순서**를 선택 취소합니다.

정렬 순서 규칙은 해당 체크박스를 선택하거나 선택 취소할 때 적용됩니다.

행 순서 수준 열 값을 값을 기준으로 정렬하지 않고 데이터 테이블에 나타나는 대로 정렬합니다. 이 옵션은 데이터 테이블 열을 정렬하는 데는 적용되지 않습니다.

참고 : " 행 순서 수준 " 과 " 일반 순서 " 가 모두 선택되어 있으면 행 순서 수준이 적용됩니다.

일반 순서 모든 값에 사전 정의된 정렬 순서 (예 : 요일) 가 있으면 나머지 항목이 해당 기준에 따라 정렬됩니다. 일반 순서는 일반적인 ASCII 순서입니다. 이 옵션은 문자 데이터에만 사용할 수 있습니다.

수치 순서 데이터를 문자열과 숫자 세그먼트로 분할하여 문자 열 값을 정렬합니다.

다음은 수치 순서 규칙입니다.

- 문자열은 숫자 (+/- 를 포함하지 않지만 중간에 마침표가 포함됨) 와 숫자가 아닌 문자의 세그먼트로 구분됩니다.
- 세그먼트는 숫자가 선행 0 으로 시작하지 않는 한 크기별로 숫자를 비교하는 방식으로 두 문자열간에서 비교됩니다.

  선행 0 이 있는 숫자는 선행 0 이 없는 숫자보다 먼저 정렬됩니다.

  두 숫자 모두 선행 0 이 있으면 숫자를 자릿수마다 비교합니다 . 예를 들어 "0CH2" 가 "0 CH2" 보다 먼저 정렬됩니다.
- 두 문자열의 세그먼트 유형이 다르면 숫자 세그먼트가 문자열 세그먼트보다 먼저 정렬됩니다.
- 로케일 비교는 기호에 영향을 줍니다. 하이픈 같은 문자와 소문자 또는 대문자는 로케일 정렬에서 무시됩니다. ASCII 순서화에서 소문자는 대문자보다 먼저 정렬됩니다. 이러한 규칙은 언어마다 다릅니다. 목표는 운영 체제가 파일 이름을 정렬하는 것과 동일한 방식으로 정렬하는 것입니다.

  ASCII 정렬을 사용하려면 " 값 순서 " 열 특성에서 **수치 순서**를 선택 취소합니다.
- 문자열 세그먼트는 기본적으로 로케일 비교를 사용하여 비교됩니다 . 이 비교는 언어를 인식합니다. 비교는 운영 체제에서 제공하므로 Windows 와 macOS 의 동작간에 미묘한 차이가 있습니다.

값 색상

명목형 또는 순서형 열의 값에 특정 색상이나 색상 테마 범위를 할당하려면 "값 색상"을 사용합니다. 그러면 모자이크 그림과 색상으로 구분된 범례를 사용하는 그림 등 해당되는 모든 그래프에서 열 값이 할당된 색상으로 나타납니다. 데이터 테이블 열의 값에 색상을 적용할 수도 있습니다.

- 특정 값의 색상을 변경하려면 색상 원을 마우스 오른쪽 버튼으로 클릭하고 색상을 선택합니다.
- 색상 테마를 사용하려면 값 색상 목록의 하단에 있는 **색상 테마**를 클릭한 후 색상을 선택합니다.

- 사용자 색상 테마를 생성하는 방법은 " 사용자 색상 테마 생성 "(287 페이지) 에서 확인하십시오 .
- 데이터 테이블의 셀에도 색상을 적용하려면 **값을 기준으로 셀에 색상 적용**을 선택합니다 .
- " 매크로 " 메뉴에서 옵션을 선택합니다 . 자세한 내용은 " 매크로 옵션 "(287 페이지) 에서 확인하십시오 .

사용자 색상 테마 생성

사용자 색상 테마를 생성하려면

1. 열 특성의 값 색상 목록 아래에서 **색상 테마**를 클릭합니다 .
2. " 사용자 색상 테마 " 개요를 엽니다 .
3. **새로 만들기**를 클릭합니다 .
4. 슬라이더를 사용하여 색상을 생성합니다 .
5. 색상에 이름을 지정하고 **저장**을 클릭합니다 .
6. **확인**을 클릭합니다 .

다음에 "값 색상" 특성을 사용할 때 "색상 테마" 메뉴에서 새로 생성한 이 사용자 색상에 액세스할 수 있습니다. 해당 색상은 환경 설정에도 저장됩니다. 색상 테마 생성에 대한 자세한 내용은 "데이터 입력 및 편집" 장의 "그래프의 색상 테마 생성"(220 페이지) 에서 확인하십시오.

매크로 옵션

"매크로" 메뉴에는 다음 옵션이 포함되어 있습니다.

끝 사이의 그래디언트 맨 위 값과 맨 아래 값의 색상을 설정합니다 . 그러면 전체 범위의 값에 색상 그래디언트가 적용됩니다 . 다른 수준에 지정한 색상 사이의 모든 색상을 적용하려면 이 명령을 사용합니다 .

선택한 점 사이의 그래디언트 " 값 색상 " 목록에서 강조 표시한 값 범위에 색상 그래디언트를 적용할 수 있도록 맨 위 값과 맨 아래 값의 색상을 설정합니다 .

색상 반전 맨 위에서 맨 아래로 또는 맨 아래에서 맨 위로 값 색상을 반전합니다 .

이전 색상으로 되돌리기 색상을 다시 원래 색상 값으로 설정합니다 .

색상 그래디언트

그림에서 연속형 열에 색상을 적용하려면 색상 그래디언트를 선택합니다 . 색상 그래디언트는 그래프 빌더 , 버블 그림 , 트리맵 및 셀 그림 플랫폼에서 지원됩니다 .

- 데이터 테이블의 셀에도 색상을 적용하려면 값 색상 목록의 위에 있는 **값을 기준으로 셀에 색상 적용**을 선택합니다 .
- 메뉴에서 색상 그래디언트를 선택합니다 .
- " 최소값 ", " 최대값 " 및 " 가운데 " 값을 입력합니다 .

- " 최소값 " 은 그래디언트의 왼쪽 색상을 반영합니다 .
- " 최대값 " 은 그래디언트의 오른쪽 색상을 반영합니다 .
- " 가운데 " 값은 그래디언트의 가운데 색상을 반영합니다 .

참고 : 그래프 빌더에 색상 그래디언트를 표시하려면 " 색상 " 영역에 열을 할당해야 합니다 . 버블 그림과 트리맵에 색상 그래디언트를 표시하려면 " 색상 적용 " 역할에 열을 할당해야 합니다 .

짧은 이름 , SPSS 이름 및 SAS 이름

이러한 열 특성은 빨간색 삼각형 메뉴의 "보고서에 사용된 열 표시" 옵션이 선택된 경우 범주형 정형 보고서에 지정된 짧은 이름을 삽입합니다.

또한 "SAS 이름"은 SAS 데이터를 가져올 때 추가되는 열 특성입니다. 이 열 특성에는 SAS 데이터 집합의 변수 이름이 포함됩니다. 이 열 특성을 수정하면 데이터를 다시 SAS로 내보낼 때 SAS 데이터 집합에 영향을 주게 됩니다.

SPSS 라벨 , 긴 이름 , SAS 라벨 및 질문

이러한 열 특성은 빨간색 삼각형 메뉴의 "보고서에 사용된 열 표시" 옵션이 선택된 경우 범주형 정형 보고서에 지정된 긴 이름을 삽입합니다.

또한 SAS 라벨은 가져온 SAS 데이터에 대한 열 특성입니다. 이 열 특성에는 SAS 데이터 집합에 있는 열에 대한 설명이 포함됩니다. 이 열 특성을 수정하면 데이터를 다시 SAS로 내보낼 때 SAS 데이터 집합에 영향을 주게 됩니다.

열 표시 방식을 제어하는 특성

"축" 및 "단위" 열 특성은 열 값이 표시되는 방식을 제어합니다.

축

열의 기본 축 설정을 변경하려면 "축" 특성을 사용합니다. JMP에서는 분석에 열이 표시될 때 자동으로 사용자 설정이 사용됩니다. 특성에 대한 자세한 내용은 "JMP 보고서" 장의 "그래프의 축 및 축 라벨 사용자 정의"(455 페이지)에서 확인하십시오.

"축" 열 특성을 추가할 때 모든 열 특성을 저장하거나 변경한 특성만 저장할 수 있습니다. 예를 들어 축이 0 에서 시작되도록 하고 축 상한은 그대로 두려는 경우, 최소 축 값을 설정하고 "열에 저장" 목록에서 **변경된 특성만**을 선택합니다.

그래프 내의 열에 대한 기본 축 특성을 설정하려면

1. 그래프를 생성합니다 .
2. 축을 변경하여 원하는 대로 지정합니다 .
3. 축을 마우스 오른쪽 버튼으로 클릭하고 **열 특성에 저장**을 선택합니다 .

단위

열의 데이터를 수집하는 데 사용된 측정 단위를 지정하려면 "단위" 특성을 사용합니다. 단위는 데이터 테이블과 그림(해당 열이 그림에 나타나는 경우)에서 열 이름 뒤에 괄호로 표시됩니다. 예를 들어 열의 연령 값이 개월 단위로 측정되었거나 금액 값이 천 달러 단위임을 나타낼 수 있습니다.

모델링 및 DOE 에 사용되는 특성

DOE 플랫폼을 사용하여 설계를 생성할 경우 결과 설계 테이블에 열 특성이 저장됩니다. 하지만 일부 열 특성은 일반적인 모델링 상황에 유용합니다. 특성을 보다 일반적으로 사용하려는 경우 특성을 직접 지정할 수 있습니다.

참고 : 실험 설계 가이드의 에서는 DOE 와 관련된 특성을 자세히 설명하고 예를 제공합니다 . 여기에서는 이러한 특성에 대한 간략한 설명만 제공합니다 .

모델링 및 DOE 에서 사용되는 열 특성은 다음과 같습니다 .

- Response Limits
- 감지 한계
- Design Role
- 코딩
- 혼합
- 요인 변경

반응 한계

"반응 한계" 열 특성은 반응의 만족도 함수를 정의합니다. 예측 및 등고선 프로파일러에서는 만족도 함수를 사용하여 최적 설정을 찾습니다. 자세한 내용은 실험 설계 가이드의 에서 확인하십시오.

JMP PRO 감지 한계

"감지 한계" 열 특성은 경계를 정의하며, 이 경계를 넘으면 반응을 측정할 수 없습니다. 이 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다. 자세한 내용은선형 모형 적합의 에서 확인하십시오.

설계 역할

"설계 역할" 열 특성은 설계된 실험과 데이터를 적합시키는 데 사용된 모형 모두에서 열이 사용된 방식을 나타냅니다. 예를 들어 열은 연속형 요인, 범주형 요인, 블록 요인 등을 나타낼 수 있습니다. 자세한 내용은 실험 설계 가이드의 에서 확인하십시오.

코딩

"코딩" 열 특성은 숫자 열의 데이터에 선형 변환을 적용합니다. 데이터는 지정된 경계에 따라 –1 과 +1 로 변환됩니다. 그런 다음 JMP 의 모형 적합 플랫폼에서 해당 열이 모형 효과로 입력될 때마다 이 변환된 데이터 값이 사용됩니다. 자세한 내용은 **실험 설계 가이드** 의 에서 확인하십시오.

혼합

"혼합" 열 특성은 데이터 테이블의 열이 혼합 성분을 나타내는 경우에 유용합니다. 혼합 성분의 합은 특정 상수가 되어야 합니다. "혼합" 열 특성은 열을 혼합 성분으로 식별하고 해당 열의 코딩을 정의합니다. 자세한 내용은 **실험 설계 가이드** 의 에서 확인하십시오.

요인 변경

"요인 변경" 열 특성은 설계된 실험에서 요인 설정을 변경하기가 얼마나 어려운지를 나타냅니다. 이 열 특성은 분할구, 분할-분할구 및 이원 분할구 설계를 생성하고 분석하는 데 사용됩니다. 자세한 내용은 **실험 설계 가이드** 의 에서 확인하십시오.

관리도 및 공정 능력과 관련된 특성

"규격 한계", "관리 한계", "Sigma" 및 "공정 능력 분포" 특성은 관리도 및 공정 능력과 관련되어 있습니다.

규격 한계

규격 한계는 분포 및 공정 능력 플랫폼을 사용하여 공정 능력 분석을 수행할 때 사용됩니다 . 이전 데이터 테이블에서는 이 열 특성을 " 공정 능력 분석 " 이라고 할 수도 있습니다 .

"규격 하한", "규격 상한" 또는 "목표값" 을 원하는 대로 조합해서 지정할 수 있습니다. **그래프 참조선으로 표시** 옵션은 지정된 규격 한계와 목표값을 적절한 그림에 참조선으로 표시합니다.

팁 : 몇 개의 열에 규격 한계를 한 번에 추가하는 방법은 Quality and Process Methods 의 에서 확인하십시오 .

관리 한계

JMP 가 열의 값에서 관리 한계를 계산하도록 하는 대신 "관리 한계" 열 특성을 사용하여 열의 관리 한계를 직접 지정할 수 있습니다. 이 열 특성은 공정 데이터와 과거 관리 한계를 비교할 때 유용합니다.

"관리 한계" 열 특성은 숫자 열에만 할당할 수 있습니다. "관리 한계" 열 특성을 할당하려면 관리도 유형을 선택한 후 요청된 값을 입력합니다. 이러한 항목 중 어느 것이라도 누락하면 관리도에서 자동으로 계산된 값으로 대체됩니다.

Sigma

알려진 시그마 값을 입력하려면 "Sigma" 특성을 사용합니다. 이 값은 관리도 같은 적용 분야나 계산을 완료하는 데 시그마 값이 필요한 모든 적용 분야에서 사용됩니다. 시그마 값이 제공되지 않으면 표본에서 시그마가 계산됩니다.

공정 능력 분포

"공정 능력 분포" 열 특성은 열의 공정 분포를 지정합니다. "공정 능력 분포" 열 특성은 "공정 능력" 플랫폼에서만 사용됩니다. "분포 및 공정 능력 분포"(292 페이지)도 확인하십시오.

QC 경보 스크립트

스크립트를 작성한 후 QC 경보 스크립트라는 열 특성 또는 데이터 테이블 특성으로 저장하여 자동 경보를 생성할 수 있습니다.

다음 QC 경보 스크립트는 검정에 실패할 때마다 자동으로 다음과 같은 메시지를 로그에 기록합니다.

```
Write(
        "Out of Control for test ",
        qc_test,
        " in column ",
        qc_col,
        " in sample ",
        qc_sample
    )
```

플랫폼에서 열이 사용되는 방식을 제어하는 특성

다음 특성은 플랫폼에서 열이 사용되는 방식을 제어합니다.

- 분포
- 시간 주기
- 맵 역할
- 상위 범주
- 다중 반응
- 목표 수준
- 수익 행렬
- 결측값 정보화

분포

연속형 숫자 데이터가 포함된 열의 경우, 열에 적합시킬 분포 유형을 선택하려면 "분포" 특성을 사용합니다. 이 분포는 분포 플랫폼에서 사용되며, 일부 경우에는 공정 능력 플랫폼에서도 사용됩니다. 자세한 내용은 "분포 및 공정 능력 분포"(292 페이지)에서 확인하십시오.

분석 > 분포를 선택하여 열에 대한 "분포" 보고서를 생성할 때 JMP는 지정된 분포를 사용하여 자동으로 적합을 추정합니다. 히스토그램에 적합된 분포를 나타내는 곡선이 추가됩니다.

"분포" 특성과 "규격 한계" 특성을 모두 설정하면 분포 플랫폼에서는 "분포" 열 특성에 지정된 분포를 기준으로 한 "공정 능력 분석" 보고서를 생성합니다.

참고 : 이전 데이터 테이블에서는 규격 한계를 지정하는 열에 " 공정 능력 분석 " 열 특성이 할당되었을 수 있습니다 . 이 특성 대신 " 규격 한계 " 특성을 사용하는 것이 좋습니다 .

분포 및 공정 능력 분포

" 공정 능력 분포 " 특성이 포함되지 않은 열을 공정 능력 플랫폼을 사용하여 분석할 경우 " 분포 " 열 특성에 지정된 분포로 인해 공정 능력 플랫폼에서 비정규 적합 결과가 생성됩니다 . 이 분포가 공정 능력 플랫폼에서 지원되는 경우 해당 분포가 사용됩니다 . 그러나 이 분포가 지원되지 않으면 플랫폼에서는 Johnson 적합을 사용합니다 .

지정된 열에 "분포" 및 "공정 능력 분포" 열 특성이 모두 저장되어 있으면 "분포" 열 특성에 지정된 분포는 분포 플랫폼에서 사용되며 "공정 능력 분포" 열 특성에 지정된 분포는 공정 능력 플랫폼에 사용됩니다.

시간 주기

시계열 플랫폼을 사용할 때는 데이터에 "시간 주기" 특성을 할당할 수 있습니다. "시간 주기" 특성은 데이터가 보고되는 주기(매년, 매분기, 매월 등)를 지정합니다. 시간 주기를 지정하면 JMP가 윤년 및 윤일 같은 사항을 고려할 수 있습니다. 주기를 지정하지 않으면 데이터가 일정한 간격의 수치형 데이터로 간주됩니다.

맵 역할

경계 데이터(국가, 주, 도, 자치주 등)가 포함된 데이터 테이블을 생성한 경우 그래프 빌더에서 해당 맵을 보려면 "맵 역할" 특성을 사용합니다.

다음 사항에 유의하십시오 .

- 사용자 경계 파일이 기본 사용자 맵 디렉터리에 있는 경우에는 -Name 파일에서만 " 맵 역할 " 특성을 지정하면 됩니다 .
- 사용자 경계 파일이 다른 위치에 있는 경우에는 -Name 파일과 분석하려는 데이터 테이블에서 " 맵 역할 " 특성을 지정해야 합니다 .
- " 맵 역할 " 특성이 포함된 열에는 동일한 경계 이름이 포함되어야 하지만 열 이름은 달라도 됩니다 .

"맵 역할" 특성의 사용 예는 Essential Graphing 의 에서 확인하십시오.

-Name 데이터 테이블에 " 맵 역할 " 특성을 추가하려면

1. 경계가 포함된 열을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 맵 역할**을 선택합니다.
2. **쉐이프 이름 정의**를 선택합니다.
3. **확인**을 클릭합니다.
4. 데이터 테이블을 저장합니다.

분석하려는 데이터 테이블에 " 맵 역할 " 특성을 추가하려면

참고 : 사용자 경계 파일이 기본 사용자 맵 디렉터리에 있지 않은 경우에만 이 단계를 수행하십시오.

1. 경계가 포함된 열을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 맵 역할**을 선택합니다.
2. **쉐이프 이름 사용**을 선택합니다.
3. **맵 이름 데이터 테이블** 옆에서 을 클릭하여 -Name 맵 데이터 테이블을 불러옵니다. 상대 경로 또는 절대 경로를 입력할 수 있습니다.

   맵 데이터 테이블이 동일한 폴더에 있으면 파일 이름만 입력합니다. 경로에 공백이 포함되어 있어도 인용 부호가 필요하지 않습니다.
4. **형태 정의 열** 옆에 선택한 열의 값과 매칭할 값이 있는 맵 데이터 테이블의 열 이름을 입력합니다.
5. **확인**을 클릭합니다.
6. 데이터 테이블을 저장합니다.

그래프 빌더에서 그래프를 생성하고 수정된 열을 "형태" 영역에 할당하면 그래프에 경계가 나타납니다.

상위 범주

데이터 집합에 등급(예: 5 점 척도)이 포함된 경우 각 등급 부분집합에 속하는 반응의 백분율을 확인할 수 있습니다. 여러 개의 특정 범주를 하나의 범주로 그룹화하려면 "상위 범주" 열 특성을 추가합니다.

상위 범주는 범주형 플랫폼에서만 지원됩니다.

데이터 열에 " 상위 범주 " 특성을 추가하려면

1. 그룹화하려는 범주가 포함된 열을 마우스 오른쪽 버튼으로 클릭합니다.
2. **열 특성 > 상위 범주**를 선택합니다.

   열 특성 창에 " 상위 범주 " 옵션이 표시됩니다 (그림 5.5).
3. " 열의 범주 " 목록에서 그룹화할 범주를 선택합니다.

4. " 상위 범주 이름 " 옆에 이름을 입력합니다 .

   이름을 비워 두면 선택한 범주에 따라 자동으로 상위 범주 이름이 지정됩니다 .

5. **추가**를 클릭하여 상위 범주를 생성합니다 .
6. **상위 범주**의 빨간색 삼각형을 클릭하고 다음 옵션 중 하나를 선택합니다 .

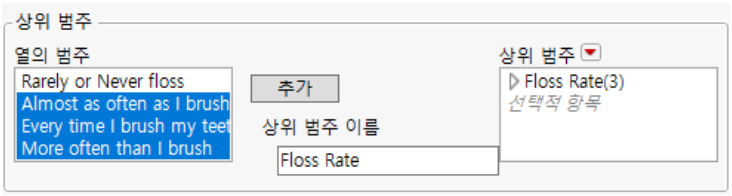
   옵션 > 숨기기 선택한 상위 범주의 데이터를 보고서와 그래프에서 숨깁니다 .

   모두 추가 열의 모든 범주에서 상위 범주를 생성합니다 .

   평균 추가 및 표준편차 추가 값 점수에 대한 통계량을 계산합니다 . 자세한 내용은 Consumer Research 에서 확인하십시오 .

7. **확인**을 클릭하여 열에 특성을 추가합니다 .

그림 5.5 상위 범주 구성의 예



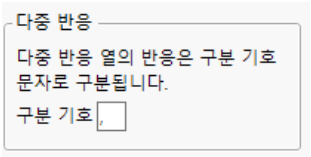
다중 반응

다중 반응이란 용어는 열의 셀에 둘 이상의 반응 값이 포함된 경우를 나타냅니다. 예를 들어 Consumer Preferences.jmp 샘플 데이터 테이블의 Brush Delimited 열에 있는 많은 셀에는 여러 값이 포함되어 있습니다. 예를 들어 6 행에는 "Wake, After Meal, Before Sleep" 이 포함되어 있습니다.

쉼표 외의 구분자를 지정하려면 "다중 반응" 열 특성을 추가합니다. 그렇지 않으면 열 정보 창에서 열의 모델링 유형을 "다중 반응" 으로 변경합니다. 다중 반응 모델링 유형에 대한 자세한 내용은 "모델링 유형"(271 페이지)에서 확인하십시오.

JMP 는 JMP 12 이하에서 저장된 데이터 테이블에 자동으로 다중 반응 모델링 유형을 할당합니다. 열에 "다중 반응" 열 특성이 포함되고 구분자가 쉼표여야 자동 할당이 이루어집니다. 사용자가 "다중 반응" 열 특성을 제거하도록 선택하더라도 JMP 는 "다중 반응" 열 특성을 제거하지 않습니다.

그림 5.6 다중 반응 구성 창



참고 : 범주형 플랫폼에서 " 다중 반응 " 특성을 사용할 수 있습니다 . 자세한 내용은 Consumer Research 의 에서 확인하십시오 . 데이터 필터에서 이 특성을 사용할 수도 있습니다 . 자세한 내용은 "JMP 보고서 " 장의 " 데이터 필터 "(419 페이지) 에서 확인하십시오 . 구분자가 쉼표인 경우에는 이 특성 대신 다중 반응 모델링 유형을 사용하는 것이 좋습니다 .

목표 수준

로지스틱 회귀 플랫폼에 대한 관심 반응 수준을 지정하려면 "목표 수준" 열 특성을 사용합니다. 이 수준은 로지스틱 회귀 모델에서 양성 반응 수준으로 사용되며 통계, 확률 곡선 및 기타 적합 측도에 영향을 줍니다. 이 열 특성은 이항 반응용으로 만들어졌지만 모든 범주형 열에 사용할 수 있습니다.

참고 : 이항 반응 열이 지정된 경우 " 목표 수준 " 열 특성은 "X 로 Y 적합 (로지스틱)" 및 모형 적합의 " 명목형 로지스틱 " 분석법과 같은 플랫폼에서 사용됩니다 .

수익 행렬

예측 모형의 명목형 또는 순서형 반응 변수 수준에 가중치를 할당하려면 "수익 행렬" 열 특성을 사용합니다. 명목형 반응의 경우 확률 임계를 사용하여 수익 행렬 항목을 지정할 수 있습니다.

참고 : " 수익 행렬 " 열 특성은 모형 비교 , 파티션 , PLS 및 모형 적합 플랫폼에서 사용됩니다 . " 수익 행렬 " 열 특성을 사용하는 모형 적합 분석법은 일반화 회귀 , 명목형 로지스틱 및 순서형 로지스틱입니다 .

수익 행렬

열 특성 > 수익 행렬을 선택하면 선택한 열에 있는 각 값에 대한 행과 열이 포함된 행렬 템플릿이 나타납니다. "실제" 수준은 행으로 표시되고 예측 수준은 열로 표시됩니다. 올바른 결정 항목은 예측 수준과 실제 수준이 일치하는 주대각 항목입니다.

- 주대각 항목에 대해 올바른 결정의 수익 또는 가중치를 반영하는 값을 입력합니다 .
- 비대각 항목에 대해 올바르지 않은 결정의 수익 (즉 , 손실) 또는 가중치를 반영하는 값을 입력합니다 .
- 예측이 수행되지 않은 경우에는 " 결정되지 않음 " 열을 사용하여 관련 수익 또는 손실을 나타냅니다 .

수익 행렬의 확률 임계 지정

이항 반응인 경우 수익 행렬 템플릿 아래에 추가 옵션이 나타납니다. 이러한 옵션을 사용하면 수익 행렬에 가중치를 직접 입력하는 대신 확률 임계를 지정할 수 있습니다.

목표 수준 및 확률 임계를 지정하십시오. 그런 다음 "설정"을 클릭하여 수익 행렬을 업데이트하십시오.

목표 수준 확률을 모델링할 수준입니다 .

확률 임계 목표 수준의 확률에 대한 임계입니다 . 관측값이 목표 수준 이내가 될 확률이 확률 임계를 초과할 경우 해당 관측값은 해당 수준으로 분류됩니다 .

설정 목표값 및 확률 임계에 지정한 사항을 반영하는 값을 수익 행렬 템플릿에 입력합니다 . 자세한 내용은 " 확률 임계 계산 "(296 페이지) 에서 확인하십시오 .

확률 임계 계산

임계 확률은 *t* 로 나타냅니다 . "설정 " 을 클릭하면 수익 행렬의 항목이 다음과 같이 할당됩니다 .

- 0: 각 주대각 항목에 할당 . 올바른 결정으로 손실이 없음을 반영
- -1: 실제값이 비목표 수준인 경우 목표 수준의 예측에 할당
- *-t*/(*1 - t*): 실제값이 목표 수준인 경우 비목표 수준의 예측에 할당

이렇게 하면 두 개의 수준으로 분류하기 위한 수익이 다음과 같이 지정됩니다 .

$$\text{목표 수준의 수익} = \left(\frac{-t}{1-t}\right)\text{Prob[Non-Target Level]}$$

$$\text{비목표 수준의 수익} = -\text{Prob[Target Level]}$$

수익성이 가장 높은 예측은 수익이 이 두 값보다 큰 수준입니다 . 위의 두 수익 방정식에 따르면 Prob[Target Level] 이 *t* 이상일 때마다 목표 수준에 관측값이 할당됩니다 .

수익 행렬 및 예측 모형

" 수익 행렬 " 열 특성이 있는 명목형 또는 순서형 열의 경우 , 대부분의 모델링 플랫폼에서 수익 행렬 항목을 반영하는 계산식 열을 저장할 수 있습니다 . 모형을 적합시킨 후 " 예측 계산식 저장 " 또는 " 확률 계산식 저장 " 옵션을 선택합니다 . JMP 에서는 일반적인 예측 계산식을 데이터 테이블에 저장할 뿐 아니라 다음과 유사한 일반적인 계산식 열도 저장합니다 .

- <level> 에 대한 수익 : 열의 각 반응 수준에 대해 각 관측값을 해당 수준으로 분류하기 위한 기대 수익을 지정합니다 .
- <column name> 에 대한 수익성이 가장 높은 예측 : 각 관측값에 대해 기대 수익이 가장 높은 반응의 수준을 지정합니다 .
- <column name> 에 대한 기대 수익 : 각 관측값에 대해 수익성이 가장 높은 예측 열에 의해 정의된 분류의 기대 수익을 지정합니다 .
- <column name> 에 대한 실제 수익 : 각 관측값에 대해 해당 관측값을 수익성이 가장 높은 예측 열에 의해 지정된 수준으로 분류하기 위한 실제 수익을 지정합니다 .

자세한 내용은 " 세 개 이상의 수준에 대한 수익 행렬의 예 "(297 페이지) 에서 확인하십시오 . 모델링에 수익 행렬을 사용하는 예는 Predictive and Specialized Modeling 의 에서 확인하십시오 .

세 개 이상의 수준에 대한 수익 행렬의 예

아래 예에서는 Travel Costs.jmp 샘플 데이터 테이블의 Airline 열에 대한 수익 행렬을 보여 줍니다.

그림 5.7 수익 행렬 창의 예

수익 행렬

행의 반응이 실제 반응일 때 열의 반응을 예측할 경우 각 행렬 항목이 수익입니다.

주대각에 올바른 결정에 대한 수익을 반영하는 값을 입력하십시오(양수).
비대각에 올바르지 않은 결정에 대한 손실을 반영하는 값을 입력하십시오(음수).
대체 결정에 대한 수익성을 반영하려면 결정되지 않음 열을 사용하십시오.

예측 계산식을 저장할 때 이러한 값을 사용하여 최선의 결정 열이 생성됩니다.
최선의 결정은 기대 이익이 가장 큰 의사결정입니다.

결정 또는 예측

실제	Carrier 1	Carrier 2	Carrier 3	Carrier 4	Undecided
Carrier 1	25	5	-5	-25	.
Carrier 2	5	20	-10	-30	.
Carrier 3	55	50	70	20	.
Carrier 4	45	40	30	60	.

이 수익 행렬의 값이 할당된 방식을 보려면 Carrier 1 부터 Carrier 4 까지의 네 항공사에서 고객 서비스에 이용하는 여행사를 고려해야 합니다. 여행사는 티켓 판매 시마다 수익을 실현하며 이 수익은 고객이 선택한 항공사에 따라 달라집니다. 여행사에서는 특정 항공사를 권장하거나 **예측**할 때 소정의 수수료를 내고 티켓을 예약합니다. 고객이 예측된 항공사를 사용하기로 결정하면 여행사는 일정 금액에서 예약 수수료를 뺀 만큼의 수익을 얻습니다. 하지만 고객이 다른 항공사를 이용하기로 결정하면 항공사는 예약 수수료만큼의 손실을 보며 예약 수수료를 한 번 더 지불해야 합니다. 즉, 예측이 올바르지 않으면 여행사의 수익이 낮아집니다.

Carrier 1 ~ Carrier 4 의 예약 수수료가 각각 $15, $20, $30, $50 이고 티켓 판매 수익은 각각 $40, $40, $100, $110 라고 가정해 보겠습니다.

여행사에서 고객에게 Carrier 1 을 권장하고 고객이 해당 티켓을 구매하기로 결정하면 항공사는 $15 의 수수료를 내고 티켓을 예약한 다음 $40 를 얻으므로 순수익은 $25 가 됩니다. 여행사에서 고객이 Carrier 4 를 선택할 것이라고 예측했는데 고객이 Carrier 1 을 선택하면 여행사는 Carrier 4 의 예약 수수료 $50 만큼 손실을 보며 Carrier 1 에 예약 수수료 $15 도 지불해야 합니다. 결과적으로 여행사는 $40 - $50 - $15 = -$25 의 순손실을 보게 됩니다.

예 : 확률 임계를 사용하여 수익 행렬 정의

샘플 데이터 테이블 Liver Cancer.jmp 에는 환자 136 명의 질병에 대한 Severity 등급이 포함되어 있습니다. BMI 부터 Jaundice 까지의 열에 지정된 예측 변수를 사용하여 Severity 를 모델링하려고 합니다. 모형에 대한 일반적인 예측 계산식은 환자를 확률이 가장 높은 Severity 수준으로 분류합니다. 하지만 실제로는 심각도가 높은 환자를 심각도가 낮은 것으로 분류하는 것은 실제로는 심각도가 낮은 환자를 심각도가 높은 것으로 분류하는 것보다 더 중대한 오류입니다. 따라서

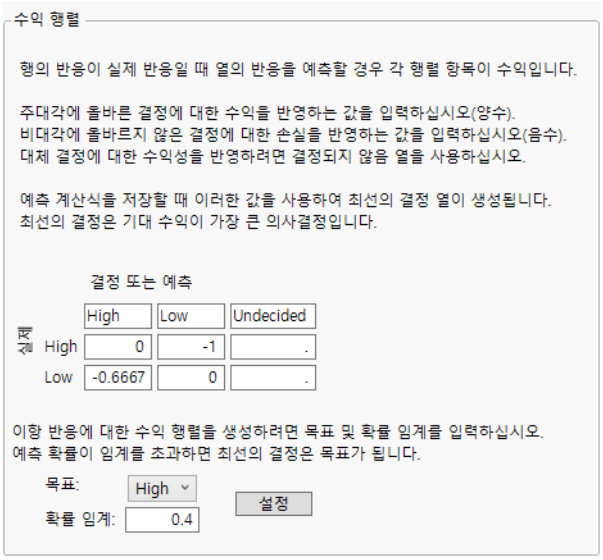
실제로는 심각도가 높은 환자를 심각도가 낮은 것으로 잘못 분류할 경우에 대해 더 높은 비용을 할당하려고 합니다.

확률 임계를 설정하면 더 높은 비용을 할당할 수 있습니다. 전문가의 의견을 토대로 결정한 좋은 전략은 심각도 수준이 높을 것으로 예측되는 확률이 0.4를 초과하는 환자를 Severity 수준 "High" 로 분류하는 것입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Liver Cancer.jmp 를 엽니다 .
2. "Severity" 열을 선택하고 **열 > 열 정보**를 선택합니다 .
3. " 수익 행렬 " 열 특성을 선택합니다 .
4. " 목표값 " 을 **High** 로 변경합니다 .
5. **확률 임계**를 0.4 로 입력합니다 .
6. **설정**을 클릭합니다 .

   수익 행렬이 업데이트되어 해당 가중치를 표시합니다 .

그림 5.8 확률 임계에 해당하는 가중치를 보여 주는 수익 행렬



이 수익 행렬은 심각도가 높은 환자를 심각도가 낮은 것으로 잘못 분류할 경우의 손실이 -1인 반면 , 심각도가 낮은 환자를 심각도가 높은 것으로 잘못 분류할 경우의 손실은 그보다 작은 -0.6667 임을 보여 줍니다 .

결측값 정보화

"결측값 정보화" 열 특성은 대부분의 적합 플랫폼에서 결측값이 포함된 열에 대해 코딩 시스템을 사용하도록 지정합니다. 연속형 열의 경우 코딩 시스템은 두 개의 열로 구성됩니다. 첫 번째 열은 원래 값을 포함하되 결측값은 비결측값의 평균으로 바꾼 열이고, 두 번째 열은 결측값이 있

는 행을 나타내는 표시자 열입니다. 범주형 열의 경우 결측값은 해당 열의 개별 수준으로 간주됩니다.

SAS 특성

SAS 데이터 집합을 가져올 경우 열에 "SAS 라벨", "SAS 이름", "SAS 출력형식" 및 "SAS 입력형식"이 할당됩니다. 데이터를 다시 SAS로 내보낼 경우 열 특성에서 값을 변경하여 SAS 데이터 집합에서 생성되는 열의 라벨, 이름, 출력형식 또는 입력형식을 제어할 수 있습니다.

가상 결합 특성을 사용하여 데이터 테이블 연결

가상 결합은 주 데이터 테이블을 하나 이상의 보조 데이터 테이블에 연결하는 기능입니다. 이 기능을 사용하면 테이블을 실제로 결합하지 않고도 주 데이터 테이블에서 보조 데이터 테이블의 데이터에 액세스할 수 있습니다. 자세한 내용은 "데이터 재구성" 장의 "데이터 테이블 가상 결합"(344 페이지)에서 확인하십시오.

" 연결 ID" 및 " 연결 참조 " 열 특성은 이러한 연결을 가능하게 합니다 .

참고 :

- " 연결 ID" 및 " 연결 참조 " 열의 데이터 유형은 서로 매칭되어야 합니다 .
- 행 상태 및 표현식 데이터 유형의 경우 연결 ID 및 연결 참조가 작동하지 않습니다 .
- 주 테이블의 참조하는 열에서 연결 대상 열을 지정할 수 있도록 " 연결 참조 " 열 특성을 추가하기 전에 " 연결 ID" 열 특성을 추가하는 것이 좋습니다 .

연결 ID

"연결 ID" 열 특성은 보조 데이터 테이블의 열을 ID 열로 표시합니다. 즉, 데이터 테이블의 행이 ID 열의 값으로 고유하게 식별됩니다. "연결 ID" 열 특성이 있는 데이터 테이블을 **참조되는** 데이터 테이블이라고 합니다.

"연결 ID" 열 특성을 추가하려면 열을 선택하고 **열 > 열 정보**를 선택한 후 "열 특성" 목록에서 **연결 ID**를 선택합니다. 그런 다음 "연결 ID" 체크박스가 선택되어 있는지 확인하고 **확인**을 클릭합니다. 선택한 체크박스는 해당 열이 데이터 테이블의 ID 열임을 나타냅니다.

연결 참조

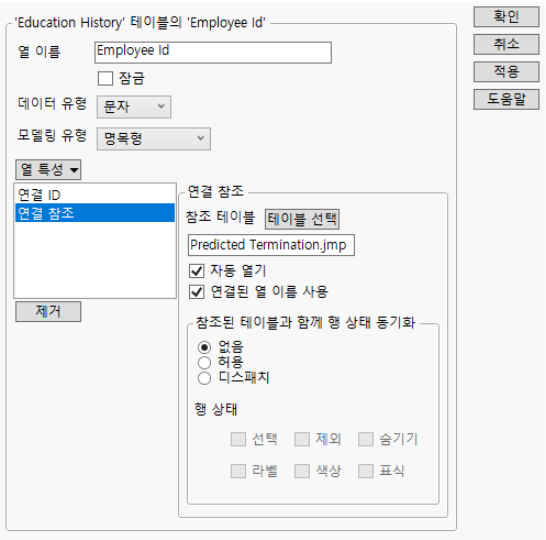
"연결 참조" 열 특성은 주 데이터 테이블의 열을 참조되는 데이터 테이블의 ID 열에 매핑합니다. 연결 참조가 있는 열을 **참조하는** 열이라고 합니다. 참조하는 열은 ID 열을 통해 보조 데이터 테이블의 데이터를 조회할 수 있습니다.

"연결 참조" 열 특성을 추가하려면 열을 선택하고 **열 > 열 정보**를 선택한 후 "열 특성" 목록에서 **연결 참조**를 선택합니다. 그런 다음 주 테이블의 경로를 입력하고 **확인**을 클릭합니다. 경로를 올바르게 입력한 경우 "확인"을 클릭하면 주 테이블의 "열" 패널에 참조되는 열이 나타납니다.

참고 :

- 경로는 대 / 소문자가 구분됩니다 .
- 데이터 테이블을 모두 동일한 폴더에 넣는 것이 좋습니다 . 그러면 " 연결 참조 " 특성에서 디렉터리 이름을 생략하고 데이터 테이블 이름만 포함해도 됩니다 . 그렇지 않으면 데이터 테이블의 전체 경로 (예 : c:/users/marie/My Data.jmp) 를 포함해야 합니다 .
- 연결 참조 경로에 JMP 경로 변수를 사용하면 호환성을 높일 수 있습니다 . 그림 5.9 에서는 예를 보여 줍니다 . 데이터 테이블을 Windows 사용자 및 macOS 사용자 모두와 공유하려면 경로에 정방향 슬래시를 입력하십시오 . 경로 변수에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .
- 열을 마우스 오른쪽 버튼으로 클릭하고 " 연결 참조 " 특성을 추가하면 경로를 입력하지 않아도 됩니다 . 자세한 내용은 " 데이터 재구성 " 장의 " 예 : 열 가상 결합 "(345 페이지) 에서 확인하십시오 .
- 가상으로 결합된 열이 포함된 데이터 테이블의 행 상태를 동기화하려면 행 상태 동기화 옵션을 선택하십시오 . 자세한 내용은 " 예 : 열 가상 결합 및 행 상태 표시 "(348 페이지) 에서 확인하십시오 .
- 주 데이터 테이블이 열릴 때 연결된 테이블을 자동으로 열려면 " 자동 열기 " 를 선택하십시오 . 자세한 내용은 " 데이터 재구성 " 장의 " 데이터 테이블 업데이트 "(354 페이지) 에서 확인하십시오 .
- 가상으로 결합된 열에서 연결된 열 이름이 사용되도록 지정할 수 있습니다 . 이 옵션을 사용하면 더 짧은 열 이름을 지정할 수 있습니다 (예를 들어 Cheese[Choice] 대신 Cheese 지정). " 연결 참조 " 열 특성에서 **연결된 열 이름 사용**을 선택하십시오 .

그림 5.9 연결 참조 열 특성 정의



추가 열 특성

"표현식 역할","이벤트 처리기" 및 "기타"라는 세 가지 추가 특성이 있습니다.

표현식 역할

이미지 데이터가 포함된 데이터 열을 가져오면 데이터를 그림으로 표시하기 위해 데이터 유형이 표현식으로 설정됩니다. 또는 사용자가 직접 "열 정보" 창에서 표현식 데이터 유형을 할당할 수 있습니다.

데이터 테이블 셀에서 이미지가 나타나는 모양을 사용자 정의하거나 행렬 및 표현식을 이미지로 표시하려면 "표현식 역할" 열 특성을 추가합니다. 기본값은 "그림"으로 설정되어 있습니다. "표현식 역할"이 "그림"으로 설정되어 있고 표현식이 그림이면 표현식 실행 결과는 그림이 되며 이 그림이 데이터 테이블에 나타납니다. 그렇지 않으면 표현식이 JSL 표현식으로 표시됩니다. 표현식이 사용되는 방식을 변경하려면 이 설정을 바꿉니다.

참고 : 크기가 32KB 를 초과하는 이미지가 포함된 표현식 열은 Microsoft Excel 에 저장할 경우 잘립니다 . 다시 JMP 로 가져오더라도 해당 표현식 열이 복원되지 않습니다 .

다음 옵션 중 하나를 선택합니다 .

그림 표현식에 그림 데이터가 포함되어 있으면 열이 실행된 후 그림이 표시됩니다 .

행렬 표현식에 행렬이 포함되어 있으면 열이 실행된 후 행렬이 표시됩니다 . 그렇지 않으면 데이터가 표현식으로 표시됩니다 .

표현식 표현식이 단순히 표현식인 경우 열에 표현식을 표시하려면 이 옵션을 사용합니다 .

그림 데이터가 포함되어 있을 수 있는 표현식을 사용하되 JMP 가 이미지를 표시하지는 않도록 하려면 " 표현식 " 을 사용합니다 .

그림 표현식 데이터 테이블 셀에서 이미지가 표시되는 방식을 사용자 정의하려면 목록에서 " 그림 " 을 선택한 후 다음 옵션 중 하나를 선택합니다 .

최대 너비 및 높이 데이터 테이블 셀에 표시되는 이미지의 최대 너비 및 높이를 지정합니다 .

늘이기 작은 이미지를 데이터 테이블 셀에 맞게 크기를 늘립니다 .

화면비율 유지 데이터 테이블 셀에서 이미지의 너비 / 높이 비율을 유지합니다 . 기본적으로 선택되어 있습니다 .

데이터 테이블의 작은 이미지에 장식 프레임 추가 데이터 테이블 셀에서 작은 이미지 주위에 입체적 가장자리를 표시합니다 .

이미지를 표현식 열로 드래그하려면

1. 데이터 테이블에 새 열을 생성합니다 .
2. 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 후 **표현식** 데이터 유형을 선택합니다 .
3. 열을 마우스 오른쪽 버튼으로 클릭하고 **열 특성**을 선택한 후 **표현식 역할**을 선택합니다 .

4. 앞에서 설명한 대로 옵션을 선택합니다 .
5. 데이터 테이블에 표시할 이미지가 포함된 웹 사이트로 이동합니다 . 컴퓨터에 있는 이미지를 사용할 수도 있습니다 .
6. 이미지를 선택하고 표현식 열의 빈 셀로 드래그합니다 .

   셀 크기를 조정하여 이미지를 더 크게 만듭니다 .

   SAS Offices.jmp 및 Big Class Families.jmp 샘플 데이터 테이블에서는 표현식 열의 예를 보여줍니다 .

팁 :

- 이미지 열이 포함된 데이터를 가져와 표현식 열을 생성할 수도 있습니다 .
- 표현식 열의 이미지를 사용하여 수행할 수 있는 작업에 대한 자세한 내용은 " 이미지를 표식으로 사용 "(448 페이지) 및 "JMP 보고서 " 장의 " 데이터 테이블에서 가리키기 라벨에 이미지 추가 "(516 페이지) 에서 확인하십시오 .

요약을 사용하여 표현식 열을 추가하려면

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 CrimeData.jmp 를 엽니다 .
2. **테이블 > 요약**을 선택합니다 .
3. State 를 선택하고 **그룹**을 클릭합니다 .
4. Total Rate 를 선택하고 **통계량 > 히스토그램**을 클릭합니다 .
5. **확인**을 클릭합니다 .

   히스토그램 이미지가 있는 새 표현식 열이 포함된 새 데이터 테이블이 나타납니다. 이미지를 더 크게 만들려면 행 크기를 조정합니다 .

   요약에 대한 자세한 내용은 8 장 , " 데이터 요약 " 에서 확인하십시오 .
6. 히스토그램 (Total Rate) 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다 .

   열의 데이터 유형이 표현식으로 설정되어 있습니다 .
7. " 열 특성 " 메뉴에서 **표현식 역할**을 선택합니다 .

   " 표현식 역할 " 은 자동으로 " 그림 " 으로 설정되어 있습니다 . 드물긴 하지만 그림을 표현식으로 표시해야 하는 경우에는 설정을 " 표현식 " 으로 변경합니다 .
8. **확인**을 클릭합니다 .

   표현식 열에 표현식 역할이 할당되었습니다 .

이벤트 처리기

텍스트 열에 하이퍼링크를 정의하려면 "이벤트 처리기" 열 특성을 사용합니다. 하이퍼링크는 웹 페이지 및 기타 문서를 여는 데 사용할 수 있습니다.

브라우저에서 URL 을 열려면 UTL 열에 이벤트 처리기를 추가합니다. 열 특성에서 자리 표시자 스크립트를 수정하여 다른 종류의 리소스를 열 수도 있습니다. 예를 들어 지리 데이터가 포함된 데이터 테이블에서 클릭 스크립트를 편집하여 사용자가 데이터 테이블의 하이퍼링크를 클릭하

면 Google 지도가 표시되도록 할 수 있습니다. SAS Offices.jmp 샘플 데이터 테이블에서는 예를 제공합니다.

스크립트 편집에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

기타

고유한 열 특성을 생성하여 선택한 이름을 할당하려면 "기타" 열 특성을 사용합니다. 그런 다음 JSL 프로그래밍에 이 특성을 사용할 수 있습니다.

1. 열을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 기타**를 선택합니다.
2. 새 특성의 이름을 입력합니다.
3. 특성 값을 입력합니다.

JMP 에서 할당 및 제어되는 특성

플랫폼 간의 정보 전달 방식을 제어하기 위해 JMP는 사용자가 제어할 수 없는 몇 가지 특성을 할당합니다. 이러한 특성은 "열 특성" 메뉴에 나타나지 않습니다.

반응 확률

프로파일러 보고서에서 "반응 확률" 특성은 범주형 반응 변수의 모든 수준이 단일 행에 나타나도록 합니다. 데이터 테이블에 특정 확률 계산식이 저장되면 JMP는 자동으로 "반응 확률" 특성을 할당합니다.

이 특성을 저장하려면 다음 단계를 수행합니다.

1. 모형 적합 플랫폼을 사용하여 로지스틱 회귀 모형을 적합시킵니다.
2. 보고서의 빨간색 삼각형을 클릭하고 **확률 계산식 저장**을 선택합니다.

   JMP 가 자동으로 새 확률 열에 " 반응 확률 " 특성을 할당합니다.

예측 프로파일러에 대한 자세한 내용은 Profilers에서 확인하십시오.

계산식 확장

프로파일러 및 비선형 플랫폼에서 모형의 성분 열에 계산식이 포함된 경우 해당 열이 다른 열을 참조하면 JMP는 내부 계산식을 대체합니다. 성분 열이 확장되지 않도록 하려면 "계산식 확장" 특성을 추가합니다. 그런 다음 값 0을 할당합니다.

자세한 내용은 Profilers의 에서 확인하십시오.

예측

연속형 반응에 모형을 적합시키고 예측 계산식 또는 예측값을 저장하면 JMP는 자동으로 "예측" 열 특성을 할당합니다. "예측" 열 특성은 예측 계산식을 생성하는 데 사용된 플랫폼을 식별합니다. 이 플랫폼은 모형 비교 플랫폼에서 생성자로 표시됩니다.

결측값 정보화 항

일부 JMP 플랫폼에서는 기본적으로, 또는 사용자가 결측값 정보화 적합을 요청한 경우, 범주형 변수의 결측 수준이 결측값 정보화로 처리됩니다. 이러한 경우 JMP는 범주형 변수가 포함된 예측 계산식 열에 "결측값 정보화 항" 열 특성을 할당합니다. 이 열 특성은 예측 계산식을 포함하는 그림 및 분석에서 결측값 범주가 범주형 변수의 개별 수준으로 처리되도록 합니다. 특히 프로파일러 그림에서는 결측값이 수준으로 표시됩니다.

JMP PRO 열 ID

초기 붓스트랩 결과 데이터 테이블을 분할하는 데 사용된 열에는 JMP가 자동으로 "열 ID" 열 특성을 할당합니다. 이 열 특성은 분할 작업을 쉽게 해주고, 분할된 붓스트랩 결과 테이블의 열이 의미 있는 이름을 갖도록 합니다.

제약 조건 상태

"제약 조건 저장"을 선택하면 각 선형 제약 조건의 계수가 데이터 테이블의 열에 나타납니다. 각 제약 조건 열에는 "제약 조건 상태" 열 특성이 할당됩니다. 이 특성은 제약 조건을 정의하는 부등식의 방향을 지정합니다. 설계 플랫폼에서 "제약 조건 불러오기"를 선택하면 "제약 조건 상태" 열 특성이 JMP에 부등식의 방향을 알려 줍니다.

블록당 런 수

"블록당 런 수" 특성은 각 블록에서 허용되는 최대 런 수를 나타냅니다. 이 특성은 설계 평가 및 설계 확대 플랫폼에서 요인의 블로킹 구조를 나타내는 데 사용됩니다.

DF

"DF" 열 특성은 많은 모델링 플랫폼에서 예측값의 표준 오차를 계산하는 계산식 열에 추가됩니다. 예측 프로파일러에서는 예측 계산식 열과 표준 오차 계산식 열이 모두 제공될 경우 이 열 특성의 값을 사용하여 신뢰 구간을 도출합니다.

주의 : 이 특성의 값은 제거하거나 변경하지 않는 것이 좋습니다 . 그렇게 하면 예측 프로파일러에서 신뢰 구간에 부정적인 영향을 줄 수 있습니다 .

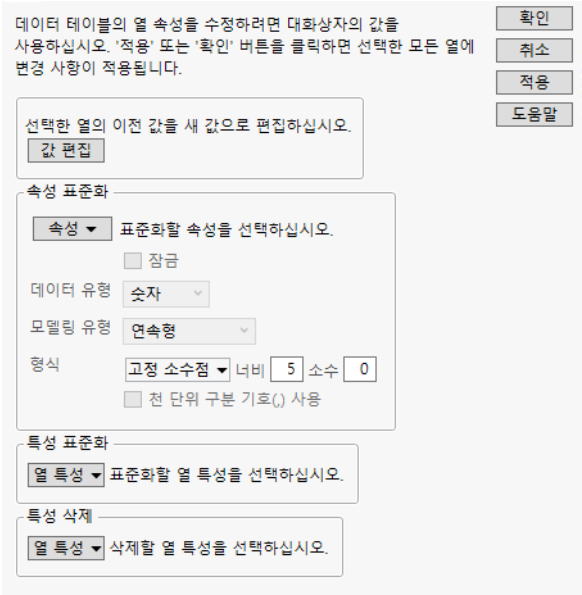
열 간에 속성 및 특성 표준화

열에 포함된 속성(데이터 유형, 모델링 유형, 숫자 형식 등) 또는 특성(계산식, 노트, 목록 및 범위 확인 등)을 다른 열에 복사하고 싶은 경우가 있습니다. 기존 열을 사용하여 열 간에 속성 및 특성을 **표준화**할 수 있습니다. 여기에는 속성 및 특성을 추가하는 기능과 삭제하는 기능이 모두 포함됩니다.

기존 열의 속성 및 특성을 여러 열에 적용하려면

1. 원하는 속성 또는 특성이 포함된 열을 하나 이상 선택합니다.
2. **열 > 속성 표준화**를 선택합니다.

그림 5.10 열 간에 속성 표준화



열 값 재코딩

"재코딩" 옵션은 여러 열 내의 유사한 값을 동일한 방식으로 재코딩해야 하는 경우에만 사용합니다. **재코딩**을 클릭한 후에 표시되는 값은 선택한 모든 열을 결합한 결과를 토대로 생성됩니다. 예를 들어 사용자 반응이 포함된 두 개의 열이 있다고 가정해 보겠습니다. 한 열에는 Agree 및 Disagree라는 값이 포함되어 있고, 다른 열에는 Agree, Disagree, Unsure 및 Strongly Disagree라는 값이 포함되어 있습니다. 두 열 모두에서 모든 값을 A, D, U 및 SD로 단순화하려고 합니다.

참고 : 단일 열에서만 값을 재코딩하려면 " 열 " > " 재코딩 " 옵션을 사용해도 됩니다. 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 열의 데이터 재코딩 "(245 페이지) 에서 확인하십시오.

열 속성 표준화

기본적으로 "속성 표준화" 패널 내의 항목은 흐리게 표시됩니다. 항목에 액세스하려면 **속성** 버튼을 클릭하고 열 간에 중복 사용할 항목을 선택합니다.

참고 : " 입력 형식 " 항목은 날짜 , 시간 및 기간 형식에만 적용할 수 있습니다 .

속성의 값을 변경하려면 "속성 표준화" 패널의 메뉴를 사용합니다.

열 특성 표준화

열 간에 특성을 표준화하려면

1. 표준화하려는 열을 선택합니다 . 해당 열 중 하나에는 다른 열에 추가할 특성이 포함되어 있어야 합니다 .
2. **열 > 속성 표준화**를 선택합니다 .
3. " 특성 표준화 " 영역에서 **열 특성**을 선택합니다 .
4. 표준화하려는 특성을 선택합니다 . 필요한 경우 특성 설정을 수정합니다 .
5. **적용**을 클릭하여 선택한 열 간에 특성을 표준화합니다 . 데이터 테이블 , 열려 있는 보고서 및 열려 있는 그래프에 변경 사항이 표시됩니다 .

   또는 **확인**을 클릭하여 선택한 열 간에 특성을 표준화하고 열 특성 창을 닫습니다 .

열 특성 삭제

여러 열에서 동일한 특성을 삭제하려면

1. 삭제하려는 속성 또는 특성이 포함된 열을 선택합니다 .
2. **열 > 속성 표준화**를 선택합니다 .
3. " 특성 삭제 " 에서 **열 특성**을 클릭하고 삭제하려는 특성을 선택합니다 .
4. **확인**을 클릭합니다 .

참고 : " 제거 " 가 아니라 " 확인 " 을 클릭해야 합니다 . " 제거 " 를 클릭하면 특성이 단순히 목록에서만 제거되고 삭제되지는 않습니다 .

예 : 계산식 표준화

열에 계산식을 적용한 경우 동일한 계산식을 데이터 테이블의 추가 열에 적용하려면 **열 참조 대체** 옵션을 사용합니다.

참고 : 이 옵션은 원래 계산식에서 참조되는 열의 위치에 따라 달라집니다 . 예를 들어 원래 계산식이 이전 열을 기준으로 하는 경우 추가 열에 적용된 다른 모든 계산식도 이전 열을 기준으로 합니다 .

예를 들어 Blood Pressure.jmp 샘플 데이터 테이블에는 5명의 피실험자를 대상으로 하루에 세 번씩 사흘 동안 혈압을 측정한 결과가 포함되어 있습니다. 각 혈압(BP) 열의 로그를 찾으려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Blood Pressure.jmp 를 엽니다 .

기존 BP 열당 하나씩, 총 9 개의 새 열을 생성합니다.

2. **열 > 새 열**을 선택합니다 .
3. 9 개의 열을 추가합니다 .
4. **확인**을 클릭합니다 .

원래 계산식을 적용합니다.

5. 열 12 1 을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택합니다 .
6. BP 8M 을 선택합니다 .
7. **초월함수 > Log** 를 선택합니다 .
8. **확인**을 클릭합니다 .

   이제 열 12 1 에 BP 8M 열의 로그가 포함되어 있습니다 . 나머지 빈 열에도 나머지 BP 열의 로그가 포함되도록 하려고 합니다 .
9. 데이터 테이블에서 원래 계산식이 있는 열을 포함하여 새로 생성한 모든 열 (열 12 1 ~ 12 9) 을 선택합니다 .
10. **열 > 속성 표준화**를 선택합니다 .
11. " 특성 표준화 " 패널에서 **열 특성**을 클릭하고 **계산식**을 선택합니다 .
12. **열 참조 대체** 옆의 체크박스를 선택합니다 .
13. BP 8M 을 선택된 상태로 둡니다 .

    이후 열 계산식에서 해당 열 참조가 대체되도록 하려고 합니다 .
14. **확인**을 클릭합니다 .

    이제 모든 새 열에 BP 열의 로그가 열 순서대로 채워집니다 . 열 12 1 에는 BP 8M 의 로그가 포함되고 , 열 12 2 에는 BP 12M 의 로그가 포함되는 식입니다 .

자세한 내용은 "JMP의 계산식 생성" 장의 "계산식 복사 및 반복"(375페이지)에서 확인하십시오.

열에 사전 선택된 분석 역할 할당

x, *y*, 가중치 또는 빈도 같은 분석 역할을 선택한 열에 할당하고 해당 역할을 데이터 테이블과 함께 저장할 수 있습니다. 이렇게 한 다음 분석을 실행하면 JMP는 사전 선택된 역할을 사용하여 창의 역할 상자를 자동으로 채웁니다. 따라서 분석을 실행할 때마다 역할을 지정할 필요가 없습니다. 예를 들어 height라는 열이 해당 데이터 테이블에 대한 모든 분석에서 *x* 역할을 하도록 하려고 합니다. *x* 역할을 적용하려면 사전 선택된 *x* 역할을 열에 할당합니다.

"빈도"를 선택하면 JMP는 해당 열의 값을 관측값 빈도로 사용합니다. *n*이 지정된 행에 대한 "빈도" 변수의 값인 경우 해당 행은 계산에서 *n*번 사용됩니다. 이 값이 1보다 작거나 없으면 JMP는 분석을 계산하는 데 이 행을 사용하지 않습니다.

"가중치"를 선택하면 해당 열의 값이 데이터 테이블의 각 관측값에 대한 가중치를 제공합니다. 이 변수는 정수가 아니어도 되지만, 값이 0보다 클 때만 분석에 포함됩니다.

JMP PRO "검증"을 선택하면 해당 열이 검증 열 역할을 지원하는 플랫폼에서 검증 열로 지정됩니다.

열에 사전 선택된 역할을 할당하려면

1. 열을 강조 표시합니다.
2. **열 > 역할 사전 선택**을 선택합니다.
3. 역할 (**역할 없음**, **X**, **Y**, **가중치**, **빈도** 또는 **검증**)을 선택합니다.

적절한 역할을 선택한 후에는 "열" 패널의 아이콘이 할당된 역할을 나타냅니다. 역할 목록에 액세스하여 다른 역할을 선택하려면 이 아이콘을 클릭합니다. 자세한 내용은 "JMP 시작하기" 장의 "열 특성을 나타내는 아이콘"(41 페이지)에서 확인하십시오.

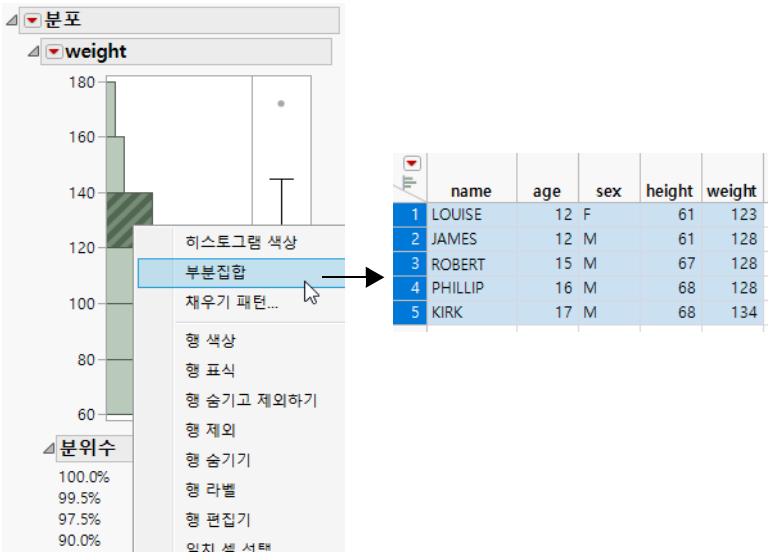
6 장

데이터 재구성
부분집합 생성 , 데이터 정렬 등

이 장에서는 JMP 데이터에 대해 수행할 수 있는 다음 작업에 대해 설명합니다.

- 행 및 열의 부분집합에서 새 데이터 테이블 생성
- 임의 개수의 열을 기준으로 정렬
- 여러 열을 단일 열에 쌓기
- 하나의 열을 두 개 이상의 열로 분할
- 행 및 열 전치
- 여러 테이블을 처음부터 끝까지 연결
- 두 개의 테이블을 나란히 결합
- 테이블을 실제로 결합하지 않고 가상으로 결합
- 테이블의 열을 다른 테이블의 값으로 업데이트
- 데이터 익명화

그림 6.1 보고서에서 부분집합 데이터 테이블 생성



목차

부분집합 데이터 테이블 생성 . . . 311
층화 부분집합 . . . 313
보고서에서 부분집합 데이터 테이블 생성 . . . 313
데이터 테이블 정렬 . . . 314
열 쌓기 . . . 317
예: 열을 단일 열로 쌓기 . . . 318
예: 열을 둘 이상의 열로 쌓기 . . . 319
열 분할 . . . 321
예: 열 분할 . . . 322
행 및 열 전치 . . . 324
예: 행 및 열 전치 . . . 326
데이터 테이블 연결 . . . 328
예: 데이터 테이블 연결 . . . 329
예: 데이터 테이블 및 테이블 변수 연결 . . . 330
데이터 테이블 결합 . . . 331
예: 데이터 테이블 결합 . . . 333
JMP 쿼리 빌더를 사용하여 데이터 테이블 쿼리 및 결합 . . . 341
쿼리의 데이터 테이블 링크 정보 . . . 343
데이터 테이블 가상 결합 . . . 344
예: 열 가상 결합 . . . 345
예: 열 가상 결합 및 행 상태 표시 . . . 348
예: 가상으로 결합된 데이터 테이블에서 참조된 데이터 병합 . . . 349
단일 열의 연결 ID 및 연결 참조 예 . . . 351
데이터 테이블 업데이트 . . . 354
예: 데이터 테이블 업데이트 . . . 356
데이터 익명화 . . . 357

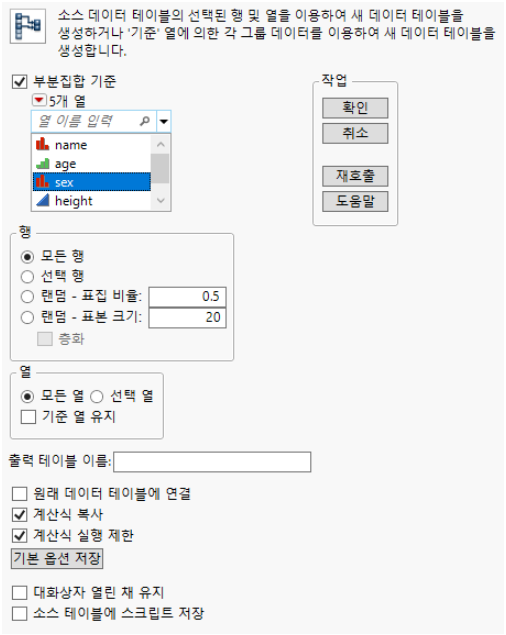
부분집합 데이터 테이블 생성

활성 데이터 테이블의 모든 행 및 열, 강조 표시된 행 및 열, 또는 무작위로 선택된 행을 포함하는 부분집합으로 새 데이터 테이블을 생성할 수 있습니다.

부분집합을 생성하려면

1. **테이블 > 부분집합**을 선택합니다.

그림 6.2 부분집합 창



2. 부분집합으로 만들 내용을 지정합니다. 다음 옵션을 조합해서 선택합니다.
   - 부분집합 기준 (선택한 열 내의 수준)
   - 행 (모든 행 , 선택한 행 또는 랜덤)
   - 열 (모든 열 또는 선택한 열)

   그림 6.2 에서는 sex 열을 부분집합으로 만듭니다.
3. 추가 옵션을 사용하여 부분집합 테이블을 세부적으로 사용자 정의합니다.
4. **확인**을 클릭하여 부분집합 테이블을 생성합니다.

부분집합 옵션

부분집합 기준 특정 열의 수준을 기준으로 부분집합을 만듭니다 . **부분집합 기준**을 선택한 후 부분집합을 만들기 위해 범주화할 열을 선택합니다 .

새 데이터 테이블을 여러 개 생성할 수 있습니다 . " 부분집합 " 창에서 지정한 열의 각 수준마다 새 데이터 테이블이 나타납니다 .

모든 행 활성 테이블의 모든 행을 포함하는 부분집합 테이블을 생성합니다 .

선택 행 활성 테이블의 선택된 행만 포함하는 부분집합 테이블을 생성합니다 . 기본적으로 선택되어 있습니다 .

랜덤 - 표집 비율 활성 데이터 테이블에서 일정 비율로 랜덤 추출한 데이터를 포함하는 부분집합 테이블을 생성합니다 . 원하는 표본 비율을 텍스트 상자에 입력합니다 . 예를 들어 50% 의 랜덤 데이터를 새 테이블에 포함하려면 텍스트 상자에 0.5 를 입력합니다 .

랜덤 - 표본 크기 활성 데이터 테이블의 랜덤 표본 데이터를 포함하는 부분집합 테이블을 생성합니다 . 원하는 표본 크기를 텍스트 상자에 입력합니다 . 예를 들어 16 개의 랜덤 행을 새 테이블에 포함하려며 텍스트 상자에 16 을 입력합니다 .

소스 테이블 전체에 해당하는 랜덤 표본을 선택하면 데이터 테이블의 행이 무작위로 뒤섞인 테이블이 생성됩니다 . 층화 열을 지정하면 각 그룹의 각 행이 무작위로 뒤섞인 테이블이 생성됩니다 . 자세한 내용은 " 층화 부분집합 "(313 페이지) 에서 확인하십시오 .

모든 열 활성 테이블의 모든 열을 포함하는 부분집합 테이블을 생성합니다 . 기본적으로 선택되어 있습니다 .

선택 열 활성 테이블의 선택된 열만 포함하는 부분집합 테이블을 생성합니다 .

기준 열 유지 부분집합을 만드는 데 기준이 된 열을 출력 데이터 테이블에 유지합니다 .

출력 테이블 이름 부분집합 테이블의 이름을 지정합니다 .

원래 데이터 테이블에 연결 부분집합 테이블을 원래 테이블에 연결합니다 . 한 테이블에서 값을 변경하면 다른 테이블이 업데이트됩니다 .

계산식 복사 원래 테이블의 계산식을 출력 열에 포함합니다 . 계산식의 계산에 필요한 모든 열을 포함해야 합니다 . 기본적으로 선택되어 있습니다 .

계산식 실행 제한 새 테이블이 생성될 때 JMP 가 열의 계산식을 실행하지 않도록 합니다 . 기본적으로 선택되어 있습니다 .

기본 옵션 저장 현재 설정을 저장합니다 .

참고 : " 기본 옵션 저장 " 은 " 선택 행 ", " 선택 열 ", " 원래 데이터 테이블에 연결 ", " 계산식 복사 " 및 " 계산식 실행 제한 " 에 대한 설정만 저장합니다 .

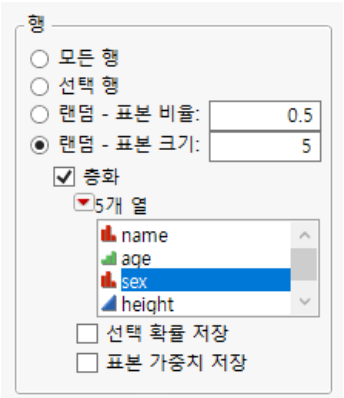
대화상자 열린 채 유지 " 확인 " 을 클릭한 후 " 부분집합 " 창을 열린 채로 유지합니다 .

소스 테이블에 스크립트 저장 동일한 설정을 사용하여 데이터의 부분집합을 다시 생성하기 위한 스크립트를 원래 데이터 테이블에 저장합니다 .

층화 부분집합

표본 크기를 지정하고 층화 열을 추가할 경우 지정한 표본 크기는 전체 부분집합의 크기가 아니라 층당 크기를 나타냅니다.

그림 6.3 층화 부분집합



지정한 표본 크기의 층별 랜덤 표본을 선택할 경우 "선택 확률"과 "표집 가중치"라는 두 개의 열을 저장할 수 있습니다. 이러한 열을 저장하려면 해당 체크박스를 선택합니다.

보고서에서 부분집합 데이터 테이블 생성

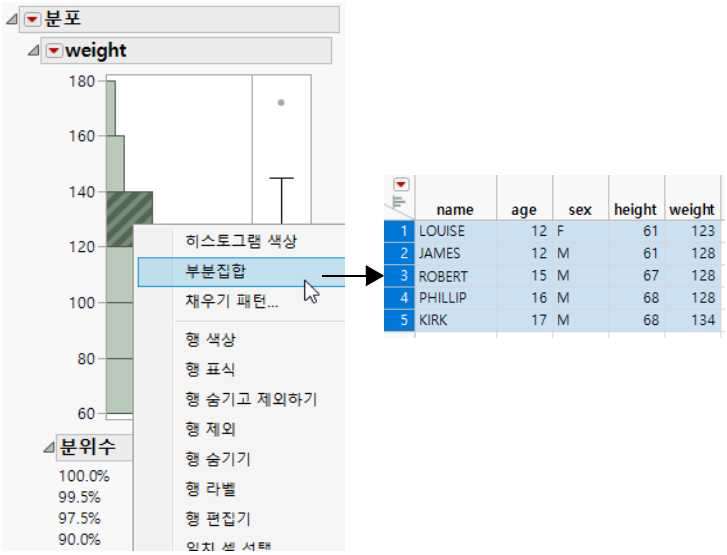
히스토그램 및 파레토도는 데이터 테이블의 연결된 부분집합을 생성합니다.

히스토그램 사용

분석 > 분포를 선택하여 히스토그램이 포함된 출력을 생성한 후 히스토그램을 사용하여 새 데이터 테이블을 생성할 수 있습니다. 새 데이터 테이블에는 히스토그램에서 강조 표시된 막대의 데이터가 포함됩니다.

부분집합을 생성하려면 강조 표시된 막대를 두 번 클릭합니다. 또는 히스토그램의 아무 곳이나 마우스 오른쪽 버튼으로 클릭하고 메뉴에서 **부분집합**을 선택합니다. 부분집합 테이블이 나타납니다.

그림 6.4 히스토그램에서 생성된 부분집합



참고 : 부분집합 데이터 테이블에는 실제 데이터가 아니라 데이터 보기가 포함됩니다 . 복사본을 만들려면 테이블을 명시적으로 저장해야 합니다 . 저장된 데이터 테이블 복사본은 더 이상 소스 테이블에 연결되어 있지 않습니다 .

파레토도 사용

분석 > 품질 및 공정 > 파레토도를 선택하여 파레토도가 포함된 출력을 생성한 후 파레토도를 사용하여 새 데이터 테이블을 생성할 수 있습니다. 새 데이터 테이블에는 파레토도에서 강조 표시된 막대의 데이터가 포함됩니다. 부분집합을 생성하려면 강조 표시된 막대를 두 번 클릭합니다.

데이터 테이블 정렬

JMP 데이터 테이블을 열 기준으로 오름차순 또는 내림차순 정렬할 수 있습니다. 기본적으로 열은 오름차순으로 정렬됩니다. 정렬된 값을 포함하는 새 테이블을 생성하거나, 원래 테이블을 정렬된 테이블로 바꿀 수 있습니다.

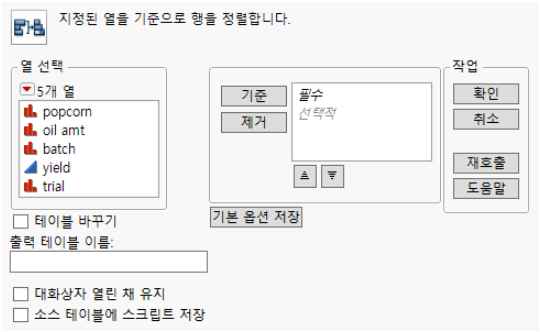
열에 값 라벨이 포함된 경우 정렬은 값 라벨이 아니라 실제 데이터 값을 기준으로 합니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "값 라벨"(284 페이지)에서 확인하십시오. 하지만 정렬된 데이터 테이블에서는 값 라벨이 표시됩니다.

정렬된 열에서 "값 순서" 특성이나 "목록 확인" 특성을 사용하는 경우에는 해당 순서에 따라 열이 정렬됩니다.

예 : 데이터 테이블 정렬

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Popcorn.jmp 를 엽니다 .
2. **테이블 > 정렬**을 선택합니다 .

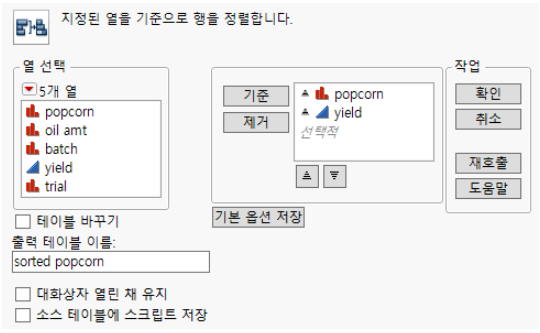
그림 6.5 정렬 창



3. 정렬 기준으로 사용할 열의 이름을 강조 표시합니다 . 이 예에서는 popcorn 과 yield 를 선택합니다 .
4. **기준**을 클릭하여 정렬 목록에 열을 추가합니다 .

   목록에 열을 추가하는 순서에 따라 정렬 우선 순위가 결정됩니다 . 목록의 첫 번째 열이 주 정렬 필드입니다 . 그 다음의 각 변수는 정렬 목록의 이전 변수 내에서 정렬됩니다 . " 기준 " 목록 내에서 열을 드래그하여 놓는 방법으로 정렬 순서를 변경할 수 있습니다 .
5. 추가 옵션을 사용하여 정렬을 세부적으로 사용자 정의합니다 . 이 예에서는 yield 를 강조 표시하고 내림차순 버튼 을 클릭합니다 .
6. **출력 테이블 이름** 옆의 상자에 정렬된 새 테이블의 이름을 입력합니다 . 이 예에서는 sorted popcorn 을 입력합니다 .

그림 6.6 완료된 정렬 창



7. **확인**을 클릭합니다 .

그림 6.7 오름차순 및 내림차순 정렬

	popcorn	oil amt	batch	yield	trial
1	gourmet	lots	small	18.0	1
2	gourmet	lots	small	16.0	2
3	gourmet	little	small	15.9	2
4	gourmet	little	small	12.1	1
5	gourmet	lots	large	9.8	2
6	gourmet	lots	large	9.2	1
7	gourmet	little	large	8.6	1
8	gourmet	little	large	8.2	2
9	plain	lots	small	10.6	1
10	plain	lots	large	10.4	1
11	plain	little	small	10.1	2
12	plain	little	small	9.9	1
13	plain	little	large	8.8	2
14	plain	lots	large	8.8	2
15	plain	little	large	8.2	1
16	plain	lots	small	7.4	2

정렬 옵션

열 선택 필터 메뉴 열을 검색 및 필터링하기 위한 옵션이 포함되어 있습니다 . 자세한 내용은 "JMP 시작하기 " 장의 " 열 필터 메뉴 "(53 페이지) 에서 확인하십시오 .

테이블 바꾸기 정렬된 값을 포함하는 새 테이블을 생성하는 대신 원래 데이터 테이블을 정렬된 테이블로 바꿉니다 . 원래 테이블에서 생성된 보고서 창이 열려 있으면 이 옵션을 사용할 수 없습니다 .

출력 테이블 이름 (선택 사항) 정렬된 테이블의 이름을 지정합니다 .

대화상자 열린 채 유지 **확인**을 클릭한 후 " 정렬 " 창을 열린 채로 유지합니다 .

소스 테이블에 스크립트 저장 동일한 설정을 사용하여 데이터를 다시 정렬하기 위한 스크립트를 원래 데이터 테이블에 저장합니다 .

기준 정렬 기준으로 사용할 열을 추가합니다 . 목록에 열을 추가하는 순서에 따라 정렬 우선 순위가 결정됩니다 . 목록의 첫 번째 열이 주 정렬 필드입니다 . 그 다음의 각 변수는 정렬 목록의 이전 변수 내에서 정렬됩니다 .

제거 강조 표시된 열을 제거합니다 .

오름차순 및 내림차순 버튼 ( ) 그룹화 변수의 값 목록 순서를 오름차순 또는 내림차순으로 변경할 수 있습니다 . " 기준 " 변수 목록에서 변수를 선택하고 필요에 따라 오름차순 또는 내림차순 버튼을 클릭합니다 . 그러면 변수 옆의 아이콘이 정렬 순서를 나타내는 모양으로 바뀝니다 .

기본 옵션 저장 현재 설정을 새 기본 옵션으로 저장합니다 .

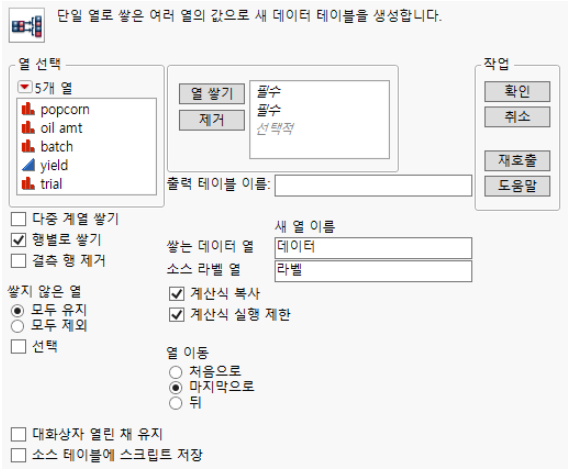
열 쌓기

두 개 이상의 열을 새로운 단일 열에 쌓고 다른 열의 값은 그대로 유지하는 방법으로 데이터 테이블을 재배열할 수 있습니다. 또는 일련의 열을 여러 그룹으로 쌓을 수 있습니다. 열을 쌓는 다양한 방법은 "쌓기 옵션"(317 페이지)에서 확인하십시오.

열을 쌓으려면 다음 단계를 수행하십시오.

1. **테이블 > 쌓기**를 선택합니다.

그림 6.8 쌓기 창



2. 쌓으려는 열의 이름을 강조 표시하고 **열 쌓기**를 클릭합니다.
3. 추가 옵션을 사용하여 쌓을 방법을 세부적으로 사용자 정의합니다.
4. **확인**을 클릭합니다.

쌓기 옵션

열 선택 필터 메뉴 열을 검색 및 필터링하기 위한 옵션이 포함되어 있습니다. 자세한 내용은 "JMP 시작하기" 장의 "열 필터 메뉴"(53 페이지)에서 확인하십시오.

다중 계열 쌓기 선택한 열을 두 개 이상의 열로 쌓습니다. "계열 수" 상자에 숫자를 입력하여 선택한 열을 쌓을 대상 열의 수를 지정합니다. 이 상자는 **다중 계열 쌓기** 옆의 상자를 선택하면 나타납니다.

계열을 인접한 여러 열로 구성할 경우에는 **인접**을 선택합니다.

자세한 내용은 "예 : 열을 둘 이상의 열로 쌓기"(319 페이지)에서 확인하십시오.

참고 : 오른쪽 상자에 열을 추가하는 순서에 따라 열이 속하는 그룹이 결정됩니다.

행별로 쌓기 열을 행별로 쌓습니다. 한 열 아래에 다른 열을 쌓으려면 이 옵션을 선택 취소합니다. 기본적으로 선택되어 있습니다.

결측 행 제거 새 테이블에서 결측 데이터를 제거합니다. " 행별로 쌓기 " 도 선택한 경우 **모든** 데이터가 결측된 행만 제거됩니다.

쌓지 않은 열 쌓지 않은 열을 새 데이터 테이블에 포함하거나 제외합니다. 다음 옵션 중 하나를 선택합니다.

모두 유지 원래 테이블의 열 중 쌓지 않은 모든 열을 새 테이블에 포함합니다. 기본적으로 선택되어 있습니다.

모두 제외 쌓지 않은 열을 새 테이블에서 제외합니다.

선택 쌓지 않은 열 중에서 새 테이블에 포함하거나 제외할 열을 선택합니다.

대화상자 열린 채 유지 **확인**을 클릭한 후 " 쌓기 " 창을 열린 채로 유지합니다.

소스 테이블에 스크립트 저장 동일한 설정을 사용하여 데이터를 다시 쌓기 위한 스크립트를 원래 데이터 테이블에 저장합니다.

열 쌓기 쌓을 열을 추가합니다.

제거 강조 표시된 열을 제거합니다.

출력 테이블 이름 (선택 사항) 새 테이블의 이름을 지정합니다.

쌓는 데이터 열 쌓는 열의 데이터를 포함할 열에 이름을 할당합니다. 기본 이름은 **데이터**입니다. 새 테이블에 이 열이 표시되지 않도록 하려면 이 상자를 비워 둡니다.

소스 라벨 열 원래 테이블의 열 이름을 포함할 열에 이름을 할당합니다. 기본 이름은 **라벨**입니다. 새 테이블에 이 열이 표시되지 않도록 하려면 이 상자를 비워 둡니다.

계산식 복사 원래 테이블의 계산식을 출력 열에 포함합니다. 기본적으로 선택되어 있습니다.

계산식 실행 제한 새 테이블이 생성될 때 JMP 가 열의 계산식을 실행하지 않도록 합니다. 기본적으로 선택되어 있습니다.

열 이동 누적 열을 이동할 위치를 지정합니다. " 마지막으로 " 가 기본값입니다.

예 : 열을 단일 열로 쌓기

연구원이 사용하는 데이터 테이블에 수율을 나타내는 두 개의 열이 있으며 이 두 열을 단일 열에 쌓으려고 합니다. 새로 생성되는 이 단일 열은 기본적으로 **데이터**라고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Popcorn Trials.jmp 를 엽니다.
2. **테이블 > 쌓기**를 선택합니다.
3. yield1 과 yield2 를 선택하고 **열 쌓기**를 클릭합니다.
4. **확인**을 클릭합니다.

그림 6.9 쌓인 데이터 테이블

	popcorn	oil amt	batch	Label	Data
1	plain	little	large	yield1	8.2
2	plain	little	large	yield2	8.8
3	gourmet	little	large	yield1	8.6
4	gourmet	little	large	yield2	8.2
5	plain	lots	large	yield1	10.4
6	plain	lots	large	yield2	8.8
7	gourmet	lots	large	yield1	9.2
8	gourmet	lots	large	yield2	9.8
9	plain	little	small	yield1	9.9
10	plain	little	small	yield2	10.1
11	gourmet	little	small	yield1	12.1
12	gourmet	little	small	yield2	15.9
13	plain	lots	small	yield1	10.6
14	plain	lots	small	yield2	7.4
15	gourmet	lots	small	yield1	18.0
16	gourmet	lots	small	yield2	16.0

라벨 열은 데이터의 소스를 식별하는 소스 라벨 열을 나타냅니다. 해당 값은 쌓은 값의 소스에 해당하는 원래 테이블의 열 이름입니다.

예 : 열을 둘 이상의 열로 쌓기

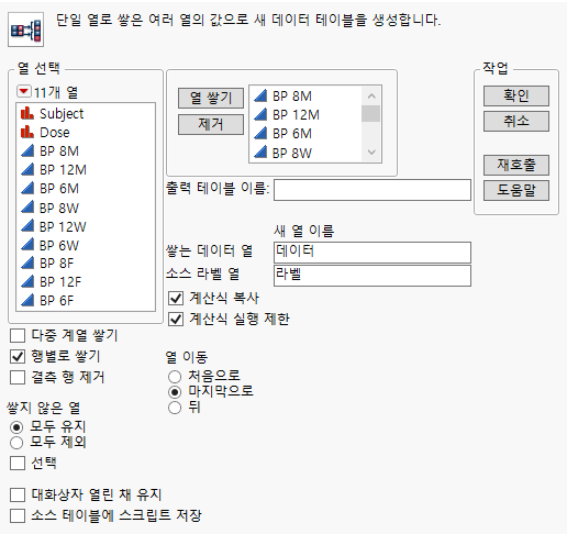
한 연구원이 혈압 측정값에 대한 데이터를 갖고 있다고 가정해 보겠습니다. 이 측정값은 월요일, 수요일, 금요일 3 일 동안 측정한 것입니다. 매일 오전 8 시, 오후 12 시 및 오후 6 시에 측정하여 세 개의 측정값을 얻었습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Blood Pressure.jmp 를 엽니다 .

   각 BP(혈압) 열은 날짜 및 시간에 따라 구분되어 있습니다 . BP 8M 열은 월요일 오전 8 시의 측정값에 해당합니다 . BP 12W 열은 수요일 , 오후 12 시의 측정값에 해당합니다 . 연구원은 모든 혈압 열을 일별로 각각 **월요일** , **수요일** 및 **금요일**에 해당하는 세 개의 열로 쌓으려고 합니다 .
2. **테이블 > 쌓기**를 선택합니다 .
3. 모든 BP 측정값을 선택하고 **열 쌓기**를 클릭합니다 .

   열 순서는 계열의 열이 그룹화되는 방식을 반영합니다 .
4. **다중 계열 쌓기**를 선택합니다 .
5. " 계열 수 " 옆에 3 을 입력합니다 .
6. 열을 세로로 쌓으려고 하므로 **인접**을 선택합니다 .
7. " 쌓는 데이터 열 " 의 이름을 데이터에서 BP(혈압) 로 바꿉니다 .
8. " 소스 라벨 열 " 의 이름을 Label 에서 Day 로 바꿉니다 .

그림 6.10 완료된 쌓기 창



9. **확인**을 클릭합니다 .

그림 6.11 쌓인 데이터 테이블

	Subject	Dose	Day	BP	Day 2	BP 2	Day 3	BP 3
1	1	A	BP 8M	183	BP 8W	174	BP 8F	171
2	1	A	BP 12M	174	BP 12W	178	BP 12F	178
3	1	A	BP 6M	180	BP 6W	181	BP 6F	171
4	2	A	BP 8M	173	BP 8W	170	BP 8F	175
5	2	A	BP 12M	181	BP 12W	179	BP 12F	185
6	2	A	BP 6M	181	BP 6W	176	BP 6F	188
7	3	A	BP 8M	181	BP 8W	188	BP 8F	183
8	3	A	BP 12M	189	BP 12W	175	BP 12F	183
9	3	A	BP 6M	177	BP 6W	182	BP 6F	180
10	4	A	BP 8M	181	BP 8W	176	BP 8F	183
11	4	A	BP 12M	177	BP 12W	173	BP 12F	187
12	4	A	BP 6M	182	BP 6W	184	BP 6F	183
13	5	A	BP 8M	184	BP 8W	172	BP 8F	170

쌓인 데이터 테이블 결과는 다음과 같습니다 .

- 첫 번째 Day 열은 월요일을 나타냅니다 .
- Day 2 열은 수요일을 나타냅니다 .
- Day 3 열은 금요일을 나타냅니다 .

열 분할

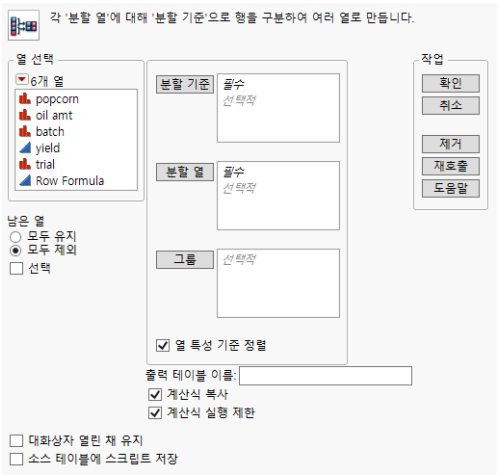
한 개의 열을 몇 개의 새 열로 분할하는 방법으로 활성 테이블에서 새 데이터 테이블을 생성할 수 있습니다. 이 열은 **분할 기준** 열이라고 하는 다른 열에 있는 값에 따라 분할됩니다. 그룹화 변수 하나 이상의 값에 따라 열을 분할할 수도 있습니다.

참고: 결측값이 포함된 범주형 열에서 분할할 경우 결측값 범주의 열 이름은 숫자 열에 해당하는 "." 가 됩니다.

열을 분할하려면 다음 단계를 수행하십시오.

1. **테이블 > 분할**을 선택합니다.

그림 6.12 분할 창



2. 분할할 열의 이름을 하나 강조 표시하고 **분할 열**을 클릭합니다.
3. 열을 분할하는 데 기준으로 사용할 값이 있는 열을 강조 표시합니다.
4. **확인**을 클릭합니다.
5. 추가 옵션을 사용하여 분할 방법을 세부적으로 사용자 정의합니다.
6. **확인**을 클릭합니다.

분할 옵션

열 선택 필터 메뉴 열을 검색 및 필터링하기 위한 옵션이 포함되어 있습니다. 자세한 내용은 "JMP 시작하기" 장의 " 열 필터 메뉴 "(53 페이지) 에서 확인하십시오.

모두 유지 모든 열을 새 테이블에 포함합니다.

모두 제외 분할에 사용된 열만 새 테이블에 포함합니다. 기본적으로 선택되어 있습니다.

선택 새 테이블에 유지할 열을 선택합니다 .

대화상자 열린 채 유지 **확인**을 클릭한 후 " 분할 " 창을 열린 채로 유지합니다 .

소스 테이블에 스크립트 저장 동일한 설정을 사용하여 데이터를 다시 분할하기 위한 스크립트를 원래 데이터 테이블에 저장합니다 .

분할 기준 새 열 이름과 열 분할 기준으로 사용할 값이 있는 열을 추가합니다 .

분할 열 분할할 열을 추가합니다 .

그룹 선택한 변수의 각 그룹 내에서 데이터를 분할하려면 " 그룹 " 변수를 지정합니다 . 각 그룹은 출력 테이블에서 하나의 행이 됩니다 . 필수 변수인 " 분할 기준 " 및 " 분할 열 " 도 지정해야 합니다 .

참고 : 그룹화 기준으로 사용할 변수가 동일하지 않은 그룹을 포함하거나 임의의 순서로 정렬된 경우 해당 변수를 그룹화 변수로 지정하면 데이터가 올바르게 재구성되고 결측값은 적절한 위치에 할당됩니다 .

열 특성 기준 정렬 " 값 순서 " 열 특성을 기준으로 출력 열의 순서를 정렬합니다 . 열에 " 값 순서 " 특성이 없고 데이터에 요일 또는 월 같은 묵시적 순서가 있으면 묵시적 순서가 적용됩니다 .

출력 테이블 이름 (선택 사항) 새 테이블의 이름을 지정합니다 .

계산식 복사 (데이터 테이블에 계산식이 있는 경우에 표시됨) 원래 테이블의 계산식을 출력 열에 포함합니다 .

계산식 실행 제한 (데이터 테이블에 계산식이 있는 경우에 표시됨) 새 테이블이 생성될 때 JMP 가 열의 계산식을 실행하지 않도록 합니다 .

예 : 열 분할

이 섹션에는 " 분할 " 명령을 사용하는 두 가지 예가 포함되어 있습니다 .

- 첫 번째 예에서는 하나의 열이 두 번째 열을 기준으로 분할됩니다 . 자세한 내용은 " 열 분할 : 기본 예 "(322 페이지) 에서 확인하십시오 .
- 두 번째 예에서는 그룹화 변수를 사용합니다 . 자세한 내용은 " 열 분할 : 행 그룹화 예 "(323 페이지) 에서 확인하십시오 .

열 분할 : 기본 예

Popcorn.jmp 데이터 테이블에서 trial 열의 데이터는 두 번의 시험 (1 및 2) 이 있었음을 보여 줍니다 . 이 예에서는 yield 열을 trial 1 과 trial 2 에 대해 각각 하나씩 , 모두 두 개의 새 열로 분할합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Popcorn.jmp 를 엽니다 .
2. **테이블 > 분할**을 선택합니다 .
3. yield 열을 선택하고 **분할 열**을 클릭합니다 .
4. trial 열을 선택하고 **분할 기준**을 클릭합니다 .

5. **남은 열** 아래에서 **모두 유지**를 선택합니다 .

   기본값은 " 모두 제외 " 로 , 이 경우 " 분할 기준 ", " 분할 열 " 또는 " 그룹 " 필드에 포함되지 않은 모든 열이 제외됩니다 . **모두 유지**를 선택하면 이러한 열이 새 테이블에 포함됩니다 .
6. (선택 사항) **출력 테이블 이름** 필드에 Yield column split by Trial column 을 입력합니다 .
7. **확인**을 클릭합니다 .

   새 데이터 테이블이 생성됩니다 (그림 6.13). 다음을 확인할 수 있습니다 .
   - yield 및 trial 열이 사라졌습니다 .
   - 데이터 테이블에 원래 열 trial 의 고유 값 (1 및 2) 을 따라 이름이 지정된 두 개의 새 열이 있습니다 .
   - 원래 열 yield 의 값은 이제 1 및 2 라는 새 열로 분할되었습니다 .
   - trial 및 yield 이외의 열은 원래 테이블과 정확히 동일합니다 .
8. (선택 사항) 새 열의 이름을 의미 있는 이름으로 바꿉니다 . 예를 들어 이름 "1" 은 yield(trial 1) 로 바꾸고 , 이름 "2" 는 yield(trial 2) 로 바꿉니다 .

그림 6.13 trial 열을 기준으로 yield 열을 분할하여 생성된 새 테이블

	popcorn	oil amt	batch	1	2
1	plain	little	large	8.2	8.8
2	gourmet	little	large	8.6	8.2
3	plain	lots	large	10.4	8.8
4	gourmet	lots	large	9.2	9.8
5	plain	little	small	9.9	10.1
6	gourmet	little	small	12.1	15.9
7	plain	lots	small	10.6	7.4
8	gourmet	lots	small	18.0	16.0

열 분할 : 행 그룹화 예

Drug Measurements.jmp 샘플 데이터 테이블에는 서로 다른 12 명의 피실험자에게 세 가지 약품 (a, b, c) 을 투여한 후의 측정값이 포함되어 있습니다 . 이 측정값을 약품 유형별로 서로 다른 열로 분할하려고 합니다 . 또한 측정값을 피실험자별로 그룹화하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug Measurements.jmp 를 엽니다 .
2. **테이블 > 분할**을 선택합니다 .
3. Drug Type 을 선택하고 **분할 기준**을 클릭합니다 .
4. Measurement 를 선택하고 **분할 열**을 클릭합니다 .

   Subject 변수에 포함된 그룹은 균일하지 않습니다 . 대부분의 피실험자에게 세 가지 약품이 모두 투여되었지만 피실험자 2 에게는 한 가지 약품만 투여되었고 피실험자 7 및 12 에게는 두 가지 약품만 투여되었기 때문입니다 . 이 경우 측정값과 피실험자가 올바르게 연결되도록 하려면 Subject 를 " 그룹 " 변수로 지정해야 합니다 .
5. Subject 를 선택하고 **그룹**을 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 6.14 그룹화 변수를 기준으로 분할된 Drug Measurements.jmp

	Subject	a	b	c
1	1	91	99	91
2	2	•	96	•
3	3	95	99	93
4	4	97	93	96
5	5	95	100	94
6	6	92	96	95
7	7	93	96	•
8	8	100	94	92
9	9	98	99	97
10	10	91	96	99
11	11	98	100	93
12	12	•	95	95

이제 피실험자 2, 7 및 12에 해당하는 결측값이 표시된 것을 확인할 수 있습니다.

행 및 열 전치

활성 데이터 테이블의 행과 열을 전치한 버전의 새 JMP 테이블을 생성할 수 있습니다. 활성 테이블의 열은 새 테이블의 행이 되고 활성 테이블의 행은 새 테이블의 열이 됩니다.

열을 전치할 때는 다음을 수행합니다.

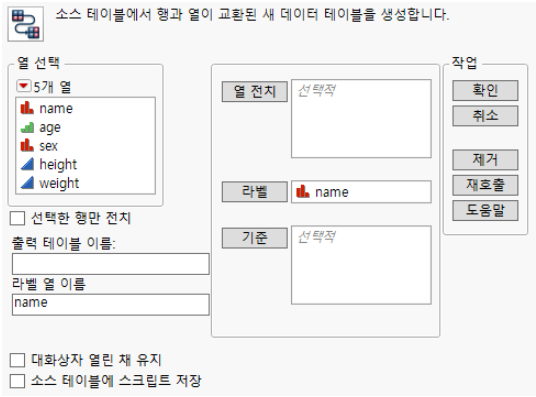
- 전치할 열을 선택합니다 .
- 새 열의 이름으로 사용할 값이 있는 " 라벨 " 열을 지정합니다 (선택 사항).
- 그룹 내에서 데이터를 전치하도록 " 기준 " 열을 지정합니다 (선택 사항).

참고 : 전치하려는 열의 데이터 유형은 동일해야 합니다 . 또한 열에 값 라벨이 포함된 경우에는 값 라벨이 아니라 실제 데이터 값이 전치에 사용됩니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 값 라벨 "(284 페이지) 에서 확인하십시오 .

행 및 열을 전치하려면 다음 단계를 수행하십시오 .

1. 전치하려는 행 및 열이 포함된 데이터 테이블을 엽니다 .
2. **테이블 > 전치**를 선택합니다 .

그림 6.15 전치 창



3. 왼쪽의 **열 선택** 상자에서 전치할 열 이름을 강조 표시합니다 .
4. **열 전치**를 클릭합니다 .
5. (선택 사항) 추가 옵션을 사용하여 전치되는 테이블을 세부적으로 사용자 정의합니다 .
6. **확인**을 클릭합니다 .

전치 옵션

열 선택 필터 메뉴 열을 검색 및 필터링하기 위한 옵션이 포함되어 있습니다 . 자세한 내용은 "JMP 시작하기 " 장의 " 열 필터 메뉴 "(53 페이지) 에서 확인하십시오 .

선택한 행만 전치 활성 테이블에서 현재 강조 표시된 행만 전치합니다 .

출력 테이블 이름 (선택 사항) 부분집합 테이블의 이름을 지정합니다 .

라벨 열 이름 (라벨 열을 지정한 경우에만 해당) 라벨 열의 대체 이름을 지정합니다 . 지정하지 않을 경우 기본 열 이름은 **라벨**입니다 .

대화상자 열린 채 유지 **확인**을 클릭한 후 " 전치 " 창을 열린 채로 유지합니다 .

소스 테이블에 스크립트 저장 동일한 설정을 사용하여 데이터를 다시 전치하기 위한 스크립트를 원래 데이터 테이블에 저장합니다 .

열 전치 전치할 열을 추가합니다 .

라벨 원래 테이블에 있는 열의 데이터를 새 테이블의 열 이름으로 사용합니다 . 다음 단계를 수행하십시오 .

1. 왼쪽의 **열 선택** 상자에서 열을 강조 표시합니다 .
2. **라벨**을 클릭합니다 . **라벨** 상자에 해당 열 이름이 나타납니다 .

기본 열 이름은 **라벨**입니다 . **라벨 열 이름** 옵션을 사용하여 열에 다른 이름을 지정할 수 있습니다 . 라벨 열에 있는 각 개별 값마다 열이 하나씩 생성됩니다 . 따라서 라벨 열에 중복 값이 있으면 마지막 중복 행의 값을 사용하여 중복 값의 열이 하나만 생성됩니다 .

기준 전치하는 열을 " 기준 " 상자에 추가한 열에 따라 그룹으로 구성합니다 . 다음 단계를 수행하십시오 .

1. " 열 선택 " 상자에서 그룹으로 표시할 값이 있는 열의 이름을 강조 표시합니다 .
2. **기준**을 클릭합니다 .

표 6.1에서는 전치에 적용되는 규칙을 설명합니다.

표 6.1 전치 규칙

If	결과
원래 테이블에 열은 있지만 행은 없는 경우	새 테이블에 해당 열 이름만 나열된 한 개의 열이 포함됩니다.
원래 테이블에 한 개의 열이 있고 이 열이 라벨에 할당된 경우	전치된 테이블에서 해당 값이 열 이름이 됩니다.
원래 테이블에 여러 개의 열이 있고 라벨 열이 포함된 경우	창이 나타날 때 "라벨" 상자에 라벨 열이 자동으로 삽입됩니다. 이 열을 표시하지 않으려는 경우 제거할 수 있습니다.
원래 테이블에 라벨 열이 없는 경우	전치된 테이블의 열 이름이 Row 1, Row 2, …, Row n이 됩니다. 여기서 *n*은 원래 테이블의 행 수입니다.

예 : 행 및 열 전치

이 섹션에는 단순한 전치 , " 라벨 " 옵션을 사용한 전치 , 그리고 " 기준 " 그룹을 사용한 전치를 보여 주는 세 가지 예가 포함되어 있습니다 .

예 : 단순 전치

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Materials1.jmp 를 엽니다 .
2. **테이블 > 전치**를 선택합니다 .
3. plastic, tin 및 gold 를 선택하고 **열 전치**를 클릭합니다 .
4. **확인**을 클릭합니다 .

그림 6.16 단순 전치된 테이블

	라벨	1행	2행
1	plastic	1	4
2	tin	2	5
3	gold	3	6

그림 6.16의 원래 테이블에는 두 개의 행과 plastic, tin 및 gold라는 세 개의 연속형 열이 있습니다. 전치된 테이블에는 원래 테이블에 있는 세 개의 열 각각에 해당하는 행이 하나씩 있고 원래 테이블의 행에 해당하는 Row 1 및 Row 2라는 열이 있습니다. Label이라는 추가 열에는 원래 테이블의 열 이름(plastic, tin 및 gold)이 값으로 포함되어 있습니다.

예 : 라벨 옵션 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Materials2.jmp 를 엽니다 .
2. **테이블 > 전치**를 선택합니다 .
3. plastic, tin 및 gold 를 선택하고 **열 전치**를 클릭합니다 .
4. item 을 선택하고 **라벨**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 6.17 라벨을 사용한 전치

	라벨	nails	hooks
1	plastic	1	4
2	tin	2	5
3	gold	3	6

원래 테이블의 item 열에 있던 값이 전치된 테이블에서 열 라벨로 사용되었습니다.

예 : 기준 그룹 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Animals Subset.jmp 를 엽니다 .
2. **테이블 > 전치**를 선택합니다 .
3. subject 및 miles 를 선택하고 **열 전치**를 클릭합니다 .
4. season 을 선택하고 **라벨**을 클릭합니다 .
5. species 를 선택하고 **기준**을 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 6.18 기준 그룹을 사용한 전치

	species	라벨	fall	winter	spring	summer
1	COYOTE	subject	1	1	1	1
2	COYOTE	miles	4	2	7	8
3	FOX	subject	3	3	3	3
4	FOX	miles	4	3	6	2

전치된 테이블에서는 전치된 값이 그룹에 포함되어 있습니다.

데이터 테이블 연결

JMP에서 둘 이상의 데이터 테이블을 연결하면 해당 데이터 테이블의 행이 결합됩니다. 새 데이터 테이블을 생성할 수도 있고, **또는** 첫 번째 데이터 테이블에 행을 추가할 수도 있습니다. 연결하려는 데이터 테이블의 열 이름이 동일한 경우가 있습니다. 이 경우 새 데이터 테이블의 열에는 모든 데이터 테이블의 값이 연결 순서대로 나열됩니다. 두 개의 원래 데이터 테이블에 서로 다른 이름의 열이 있는 경우에는 해당 열이 새 데이터 테이블에 포함되고 결측값이 표시됩니다.

열 이름이 동일한 두 개의 데이터 테이블을 연결하려면 다음 단계를 수행하십시오.

1. **테이블 > 연결**을 선택합니다.

그림 6.19 연결 창

2. 결합할 데이터 테이블의 이름을 강조 표시하고 **추가**를 클릭합니다.

   선택한 데이터 테이블을 개수에 상관없이 연결할 수 있으며 , 동일한 데이터 테이블을 여러 번 추가할 수도 있습니다. 새 데이터 테이블의 행 수는 모든 데이터 테이블에 있는 행 수의 합입니다.

3. (선택 사항) 모든 계산식이 포함되도록 하려면 **계산식 저장 및 실행** 옵션을 클릭합니다.

   이 옵션을 선택하지 않으면 새 데이터 테이블에 계산식이 포함되지 않습니다.

   참고 : 동일한 이름의 열에 서로 다른 계산식이 포함되어 있을 수도 있습니다. 이 경우 연결된 데이터 테이블에는 해당 열에 대한 계산식이 포함된 첫 번째 데이터 테이블의 계산식이 저장됩니다. 세 개 이상의 테이블을 연결하려고 하는데 두 번째 또는 세 번째 테이블에 동일한 이름의 열에 대한 계산식이 포함되어 있으면 이러한 경우가 발생합니다.

4. (선택 사항) 새 데이터 테이블에 **소스 테이블**이라는 열을 추가하려면 **소스 열 생성** 옵션을 클릭합니다.

   이 열은 해당 행에서 소스 데이터 테이블의 이름을 식별합니다.

5. (선택 사항) **연결할 데이터 테이블** 필드에 첫 번째로 나열된 데이터 테이블에 행을 추가하려면 **첫 번째 테이블에 추가** 옵션을 선택합니다. 새 데이터 테이블을 생성하는 대신 이 옵션을 사용할 수 있습니다.

6. (선택 사항) **출력 테이블 이름** 필드에 새 데이터 테이블의 이름을 입력합니다.

이름을 입력하지 않으면 데이터 테이블에 **제목 없음 #**(예 : **제목 없음 1**) 같은 이름이 자동으로 지정됩니다 . **첫 번째 테이블에 추가** 옵션을 선택한 경우에는 **출력 테이블 이름** 필드를 사용할 수 없습니다 .

7. **확인**을 클릭합니다 .

예 : 데이터 테이블 연결

두 개의 데이터 테이블(Trial1 및 Trial2)을 새 데이터 테이블에 연결하려고 한다고 가정해 보겠습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Trial1.jmp 및 Trial2.jmp 를 엽니다 .
2. Trial1.jmp 테이블에서 **테이블 > 연결**을 선택합니다 .
3. **열린 데이터 테이블** 목록에서 Trial2 를 선택하고 **추가**를 클릭합니다 .
4. **확인**을 클릭합니다 .

데이터 테이블이 연결된 새 테이블로 결합되어 첫 번째 데이터 테이블의 모든 행 다음에 두 번째 데이터 테이블의 모든 행이 나열됩니다.

그림 6.20 두 개의 데이터 테이블을 연결한 결과

	popcorn	oil amt	batch	yield
1	plain	little	large	8.2
2	gourmet	little	large	8.6
3	plain	lots	large	10.4
4	gourmet	lots	large	9.2
5	plain	little	small	9.9
6	gourmet	little	small	12.1
7	plain	lots	small	10.6
8	gourmet	lots	small	18.0
9	plain	little	large	8.8
10	gourmet	little	large	8.2
11	plain	lots	large	8.8
12	gourmet	lots	large	9.8
13	plain	little	small	10.1
14	gourmet	little	small	15.9
15	plain	lots	small	7.4
16	gourmet	lots	small	16.0

연결된 데이터 테이블에는 항상 원래 데이터 테이블에 있는 각 열 이름마다 하나씩의 열이 포함됩니다. 하지만 열 이름이 정확히 매칭되지 않는 경우에는 열이 병합되지 않습니다. 예를 들어 yield 열 대신 yield1 및 yield2 라는 열 이름이 사용된 경우에는 연결된 데이터 테이블에서 각 열마다 별도의 개별 열이 생성됩니다.

테이블 변수가 포함된 두 개 이상의 데이터 테이블을 연결할 경우에는 각 테이블 변수마다 별도의 열이 생성됩니다. 따라서 데이터 테이블을 연결할 때 중요한 구분이 손실되지 않습니다.

참고 : 이름이 노트로 시작하는 테이블 변수에 대한 열은 생성되지 않습니다 .

예 : 데이터 테이블 및 테이블 변수 연결

예를 들어 서로 다른 두 병원에서 암에 관한 시험을 두 차례 실시했다고 가정해 보겠습니다. 한 번의 시험에서 얻은 데이터는 Cancer1.jmp 데이터 테이블에 있고, 다른 한 번의 시험에서 얻은 데이터는 Cancer2.jmp 데이터 테이블에 있습니다.

데이터와 변수를 한 테이블에 통합하려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cancer1.jmp 및 Cancer2.jmp 를 엽니다 .

   구분되는 두 개의 테이블 변수 Dosage Amount 와 Location 이 있습니다 . 연결된 테이블에서는 이 두 테이블 변수에 대한 열이 생성됩니다 .
2. Cancer1.jmp 데이터 테이블에서 **테이블 > 연결**을 선택합니다 .
3. Cancer2 를 선택하고 **추가**를 클릭합니다 .
4. **확인**을 클릭합니다 .

그림 6.21 연결된 데이터 및 변수

	Time	Cell Type	Treatment	Prior	Dosage Amount	Location
1	3	Adeno	Standard	No	150	Roanoke
2	7	Adeno	Test	No	150	Roanoke
3	8	Adeno	Standard	Yes	150	Roanoke
4	8	Adeno	Test	No	150	Roanoke
5	12	Adeno	Standard	Yes	150	Roanoke
6	18	Adeno	Test	Yes	150	Roanoke
7	19	Adeno	Test	No	150	Roanoke
8	24	Adeno	Test	No	150	Roanoke
9	3	Adeno	Standard	No	250	Asheville
10	7	Adeno	Test	No	250	Asheville
11	8	Adeno	Standard	Yes	250	Asheville
12	8	Adeno	Test	No	250	Asheville
13	12	Adeno	Standard	Yes	250	Asheville
14	18	Adeno	Test	Yes	250	Asheville
15	19	Adeno	Test	No	250	Asheville
16	24	Adeno	Test	No	250	Asheville

데이터 및 변수가 연결됩니다. 변수는 연결된 테이블에서 열로 나타납니다. 각 데이터 테이블의 노트는 새 데이터 테이블에 테이블 변수로 추가됩니다.

데이터 테이블 결합

테이블 > 결합을 선택하여 두 개의 데이터 테이블을 하나의 새 테이블로 결합할 수 있습니다. 두 개의 데이터 테이블을 결합하는 방법에 대한 자세한 설명은 "두 개의 데이터 테이블을 새 데이터 테이블에 결합하려면 다음 단계를 수행하십시오."(331 페이지)에서 확인하십시오. 다음과 같은 세 가지 방법으로 테이블을 결합할 수 있습니다.

- 행 번호에 따라 결합. 자세한 내용은 " 예 : 행 번호별 결합 "(333 페이지) 에서 확인하십시오 .
- 원래의 두 테이블에 있는 행의 가능한 모든 조합으로 새 테이블을 구성하는 Cartesian 방식으로 결합. 자세한 내용은 " 예 : Cartesian 결합 "(335 페이지) 에서 확인하십시오 .
- 두 데이터 테이블 모두에 있거나 단일 데이터 테이블에 있는 하나 이상의 열에 포함된 값을 매칭. 자세한 내용은 " 예 : 매칭 열을 기준으로 결합 "(337 페이지) 에서 확인하십시오 .

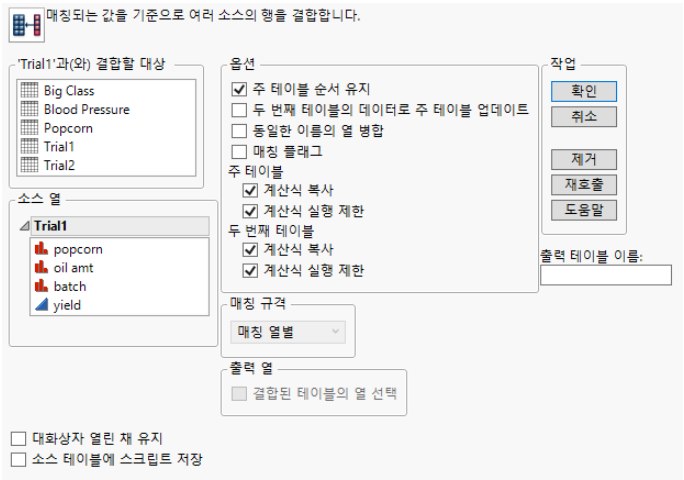
참고 : " 테이블 " 메뉴의 "JMP 쿼리 빌더 " 옵션을 사용하면 단순 결합을 수행하기 전에 데이터를 쿼리할 수 있습니다 . 자세한 내용은 " 데이터 재구성 " 장의 "JMP 쿼리 빌더를 사용하여 데이터 테이블 쿼리 및 결합 "(341 페이지) 에서 확인하십시오 .

두 개의 데이터 테이블을 새 데이터 테이블에 결합하려면 다음 단계를 수행하십시오 .

1. 결합하려는 두 개의 데이터 테이블을 엽니다 .
2. **테이블 > 결합**을 선택합니다 .

   창이 나타나고 **... 과 (와) 결합할 대상** 아래에 열려 있는 모든 테이블의 이름이 나타납니다 .

그림 6.22 결합 창



3. **... 과 (와) 결합할 대상** 상자에서 활성 테이블과 결합할 테이블을 선택합니다 .
4. " 매칭 규격 " 영역에서 테이블을 결합할 방법을 지정하는 옵션을 선택합니다 .

5. **출력 테이블 이름** 텍스트 상자에 새 테이블의 이름을 입력합니다 .
6. (선택 사항) 추가 옵션을 사용하여 결합 절차를 세부적으로 사용자 정의합니다 .
7. **확인**을 클릭하여 결합된 데이터 테이블을 생성합니다 .

결합 옵션

대화상자 열린 채 유지 **확인**을 클릭한 후 " 결합 " 창을 열린 채로 유지합니다 .

소스 테이블에 스크립트 저장 동일한 설정을 사용하여 데이터를 다시 결합하기 위한 스크립트를 원래 데이터 테이블에 저장합니다 .

주 테이블 순서 유지 결합된 테이블에서 매칭되는 열을 기준으로 정렬하지 않고 원래 데이터 테이블의 순서를 유지합니다 . 기본적으로 선택되어 있습니다 .

두 번째 테이블의 데이터로 주 테이블 업데이트 원래 테이블의 데이터가 동일한 이름을 가진 열의 데이터로 업데이트됩니다 . 결과는 새 데이터 테이블에 표시됩니다 .

다음 사항에 유의하십시오 .

- JMP 는 데이터를 결측값으로 바꾸지 않습니다 .
- 출력 테이블에는 원래 테이블과 동일한 열이 사용됩니다 . 따라서 **두 번째 테이블의 데이터로 주 테이블 업데이트**를 사용할 경우에는 **결합된 테이블의 열 선택**을 적용할 수 없습니다 .
- **두 번째 테이블의 데이터로 주 테이블 업데이트** 옵션은 행 번호 또는 매칭되는 열을 기준으로 결합할 때만 사용할 수 있습니다 .

동일한 이름의 열 병합 두 번째 테이블의 데이터로 원래 테이블에 동일한 이름으로 있는 열의 데이터를 바꿉니다 . 첫 번째 테이블의 결측값은 두 번째 테이블의 비결측값으로 바뀝니다 .

열을 기준으로 매칭할 경우 **동일한 이름의 열 병합**을 선택하면 " 매칭 플래그 " 가 선택됩니다 . 새로 결합된 테이블에는 **매칭 플래그**라는 명목형 열이 포함됩니다 .

- 이 열에 " 주 " 가 나타나면 해당 데이터는 첫 번째 (활성) 테이블에 있던 것입니다 .
- 이 열에 " 포함 " 이 나타나면 해당 데이터는 두 번째 테이블에 있던 것입니다 .
- 이 열에 " 둘 다 " 가 나타나면 해당 데이터는 첫 번째 테이블과 두 번째 테이블 모두에 있던 것입니다 .

매칭 플래그 열을 기준으로 매칭할 때 **매칭 플래그** 열을 생성할지 여부를 결정합니다 . **동일한 이름의 열 병합**을 선택하면 자동으로 선택됩니다 .

계산식 복사 (주 테이블 및 두 번째 테이블) 주 테이블 및 / 또는 두 번째 테이블의 계산식을 출력 열에 포함합니다 . 기본적으로 선택되어 있습니다 .

계산식 실행 제한 (주 테이블 및 두 번째 테이블) 새 테이블을 생성하는 동안 JMP 가 열의 계산식을 실행하지 않도록 합니다 . 기본적으로 선택되어 있습니다 .

매칭 열별 행을 결합하려면 두 테이블 모두에서 값과 데이터 유형이 매칭되는 열을 선택합니다 . 다음 단계를 수행하십시오 .

1. **소스 열**의 각 목록에서 열 이름을 강조 표시합니다 . 위쪽 목록의 첫 번째 강조 표시 열은 아래쪽 목록의 첫 번째 강조 표시 열과 쌍을 이루고 , 그 다음 강조 표시 열도 그런 식으로 쌍을 이룹니다 . 모든 열 쌍의 값과 데이터 유형이 매칭되는 경우에만 행이 결합됩니다 .
2. **매칭**을 클릭합니다 . 선택한 열 쌍이 **열 매칭** 상자에 나타납니다 . 매칭 열은 이름이 동일하지 않아도 되며 , 두 테이블 모두에서 상대적 열 위치가 동일하지 않아도 됩니다 .
3. (선택 사항) 첫 번째로 찾은 매칭 항목만 포함하려면 두 테이블 모두의 **복수 항목 제외** 상자를 선택합니다 . 그러면 첫 번째로 찾은 매칭 항목만 새 테이블에 기록됩니다 . 한 테이블에 대해서만 이 옵션을 지정하면 첫 번째 매칭 값이 다른 테이블의 모든 매칭 값과 결합됩니다 . 어느 테이블에서도 **복수 항목 제외** 상자를 선택하지 않으면 매칭되는 열 값의 각 그룹 내에서 Cartesian 결합이 수행됩니다 .
4. (선택 사항) 데이터 테이블의 모든 행을 포함하려면 (매칭되는 값이 없는 경우도 포함) **매칭되지 않는 항목 포함** 상자를 선택합니다 . 결합할 데이터 테이블 중 하나 또는 둘 모두에 대해 이 옵션을 지정할 수 있습니다 .

행 번호별 두 개의 테이블을 나란히 결합합니다 .

Cartesian 결합 원래의 두 테이블에 있는 행의 가능한 모든 조합으로 새 테이블을 구성하는 Cartesian 방식을 사용하여 두 테이블을 결합합니다 . JMP 는 첫 번째 테이블의 데이터를 두 번째 테이블의 데이터와 교차 조합하여 각 집합의 모든 값 조합을 표시합니다 .

결합된 테이블의 열 선택 어느 한 테이블에서 출력 테이블에 포함할 일부 열을 선택합니다 . 다음 단계를 수행하십시오 .

1. " 소스 열 " 영역에서 새 테이블에 포함할 각 테이블의 열을 강조 표시합니다 .
2. " 출력 열 " 영역에서 **선택**을 클릭합니다 .

출력 테이블 이름 결합된 테이블의 이름을 지정합니다 .

예 : 데이터 테이블 결합

다음 섹션에서는 "결합" 명령을 사용하는 예를 제공합니다.

예 : 행 번호별 결합

행 번호별로 테이블을 결합하면 두 테이블이 나란히 결합됩니다. 일부 열만 포함하도록 지정하지 않은 한 새 테이블에는 두 테이블의 모든 열이 포함됩니다.

행 수가 서로 동일하지 않은 테이블을 결합하려면

결합하려는 두 테이블의 행 수가 동일하지 않은 경우 새 테이블에는 두 테이블 모두에 있는 행의 값이 포함됩니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Species1.jmp 및 Species2.jmp 를 엽니다 .

   Species1.jmp 테이블에는 두 개의 행이 있고, Species2.jmp 테이블에는 네 개의 행이 있습니다.
2. Species1.jmp 테이블에서 **테이블 > 결합**을 선택합니다 .

3. **... 과 (와) 결합할 대상** 상자에서 Species2 를 선택합니다 .
4. " 매칭 규격 " 영역에서 **행 번호별**을 선택합니다 .
5. **확인**을 클릭합니다 .

그림 6.23 행 번호별로 결합된 테이블

	species	season
1	FOX	spring
2	COYOTE	summer
3		fall
4		winter

두 개의 행이 있는 테이블과 네 개의 행이 있는 테이블을 결합하면 새 테이블에는 네 개의 행이 포함됩니다.

동일한 이름의 열을 결합하려면

두 테이블에 동일한 이름의 열이 있는 경우 새 테이블에서 해당 열의 이름은 "table name 의 column name" 형식으로 나타납니다. 예를 들어 그림 6.24 에 표시된 것과 같이 Trial1.jmp 및 Trial2.jmp 데이터 테이블의 8 개 행을 단일 테이블에 결합하려고 한다고 가정해 보겠습니다. 두 테이블을 결합하여 새 테이블에 두 테이블의 모든 행이 포함되도록 하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Trial1.jmp 및 Trial2.jmp 를 엽니다 .
2. Trial1.jmp 데이터 테이블에서 **테이블 > 결합**을 선택합니다 .
3. **... 과 (와) 결합할 대상** 상자에서 Trial2 를 선택합니다 .
4. " 매칭 규격 " 메뉴에서 **행 번호별**을 선택합니다 .
5. **확인**을 클릭합니다 .

그림 6.24 원래 테이블과 결합된 테이블

	Trial1의 popcorn	Trial1의 oil amt	Trial1의 batch	Trial1의 yield	Trial2의 popcorn	Trial2의 oil amt	Trial2의 batch	Trial2의 yield
1	plain	little	large	8.2	plain	little	large	8.8
2	gourmet	little	large	8.6	gourmet	little	large	8.2
3	plain	lots	large	10.4	plain	lots	large	8.8
4	gourmet	lots	large	9.2	gourmet	lots	large	9.8
5	plain	little	small	9.9	plain	little	small	10.1
6	gourmet	little	small	12.1	gourmet	little	small	15.9
7	plain	lots	small	10.6	plain	lots	small	7.4
8	gourmet	lots	small	18.0	gourmet	lots	small	16.0

두 개의 원래 테이블에서 열 이름이 동일할 수 있습니다. 이 경우 출력 열 이름은 소스 테이블 이름을 사용하여 한정됩니다. 예를 들어 새 테이블의 열 이름은 "table name의 <variable name>" 형식으로 나타납니다.

지정한 열만 결합하려면

원래 데이터 테이블의 모든 열을 결합된 테이블에 포함하지는 않으려고 한다고 가정해 보겠습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Trial1.jmp 및 Trial2.jmp 를 엽니다 .
2. Trial1.jmp 데이터 테이블에서 **테이블 > 결합**을 선택합니다 .
3. **... 과 (와) 결합할 대상** 상자에서 Trial2 를 선택합니다 .
4. " 매칭 규격 " 메뉴에서 **행 번호별**을 선택합니다 .
5. **결합된 테이블의 열 선택**을 클릭하여 포함하려는 일부 열을 지정합니다 .
6. " 소스 열 " 목록에서 Trial1 목록의 popcorn 및 yield 를 선택하고 , Trial2 목록의 yield 를 선택합니다 .

   두 테이블의 popcorn 열에는 동일한 데이터가 있으므로 한 열만 선택하면 됩니다 .
7. **선택**을 클릭합니다 .
8. **확인**을 클릭합니다 .

그림 6.25 지정한 열만 결합

	popcorn	Trial1의 yield	Trial2의 yield
1	plain	8.2	8.8
2	gourmet	8.6	8.2
3	plain	10.4	8.8
4	gourmet	9.2	9.8
5	plain	9.9	10.1
6	gourmet	12.1	15.9
7	plain	10.6	7.4
8	gourmet	18.0	16.0

예 : Cartesian 결합

Cartesian 결합을 수행하면 JMP는 두 테이블을 Cartesian 방식으로 결합하여 원래의 두 테이블에 있는 행의 가능한 모든 조합으로 새 테이블을 구성합니다. 따라서 출력 테이블에는 각 열 값 조합마다 하나씩, 여러 가지 경우가 포함됩니다.

예 : 단순 Cartesian 결합

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Species1.jmp 및 Species2.jmp 를 엽니다 .
2. Species1.jmp 테이블에서 **테이블 > 결합**을 선택합니다 .
3. **... 과 (와) 결합할 대상** 상자에서 Species2 를 선택합니다 .
4. " 매칭 규격 " 메뉴에서 **Cartesian 결합**을 선택합니다 .
5. **확인**을 클릭합니다 .

그림 6.26 Cartesian 결합을 사용한 테이블 결합

	species	season
1	FOX	spring
2	FOX	summer
3	FOX	fall
4	FOX	winter
5	COYOTE	spring
6	COYOTE	summer
7	COYOTE	fall
8	COYOTE	winter

Species1.jmp 의 데이터가 Species2.jmp 의 데이터와 교차 조합되어 각 집합의 모든 값 조합을 보여 주는 결합 테이블이 생성됩니다.

예 : 복잡한 Cartesian 결합

이 예에서는 **테이블 > 결합** 명령을 두 번 사용합니다.

- 첫 번째 결합 시에는 **Cartesian** 옵션을 사용하여 Oil Amount.jmp 테이블을 Batch.jmp 테이블과 결합합니다 .
- 두 번째 결합 시에는 결과 테이블 (Cartesian oil amount + batch) 을 Popcorn Type.jmp 테이블과 결합하여 모든 테이블이 결합된 최종 테이블을 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Oil Amount.jmp, Batch.jmp 및 Popcorn Type.jmp 를 엽니다 .
2. Oil Amount.jmp 테이블에서 **테이블 > 결합**을 선택합니다 .
3. **... 과 (와) 결합할 대상** 상자에서 Batch 를 선택합니다 .
4. " 매칭 규격 " 메뉴에서 **Cartesian 결합**을 선택합니다 .
5. **출력 테이블 이름**에 Oil Amount and Batch 를 입력합니다 .
6. **확인**을 클릭합니다 .

그림 6.27 결합된 Oil Amount and Batch 테이블

	oil amount	batch
1	little	small
2	little	large
3	lots	small
4	lots	large

결합된 테이블에는 Oil Amount.jmp 및 Batch.jmp 테이블의 모든 열이 포함되어 있습니다 . Popcorn Type.jmp 열을 추가합니다 .

7. 방금 생성한 Oil Amount and Batch 테이블에서 **테이블 > 결합**을 선택합니다 .
8. **... 과 (와) 결합할 대상** 상자에서 Batch 를 선택합니다 .
9. " 매칭 규격 " 메뉴에서 **Cartesian 결합**을 선택합니다 .

10. **확인**을 클릭합니다 .

그림 6.28 Popcorn Type 과 결합된 Oil Amount and Batch 테이블

	oil amount	batch	popcorn
1	little	small	gourmet
2	little	small	plain
3	little	large	gourmet
4	little	large	plain
5	lots	small	gourmet
6	lots	small	plain
7	lots	large	gourmet
8	lots	large	plain

최종 테이블에는 세 개의 원래 테이블 모두에 있는 모든 열이 포함됩니다. Cartesian 결합으로 생성되는 행의 수는 원래 테이블의 행 수를 곱한 것입니다.

예 : 매칭 열을 기준으로 결합

매칭 열을 기준으로 데이터 테이블을 결합하면 JMP가 매칭 열 값을 기준으로 두 테이블의 행을 정렬합니다.

주의 : 매칭된 열이 고유한 행을 생성하지 않으면 신중하게 진행하십시오 .

참고 :

- 매칭에 사용되는 열은 동일한 데이터 유형이어야 합니다 (수치 , 문자 또는 행 상태).
- 데이터 테이블에서 중복된 값 또는 행을 제거하기 위해 데이터 테이블을 해당 데이터 테이블과 자체 결합할 수 있습니다 .
- 새 데이터 테이블을 생성하는 다른 방법으로 가상 결합을 사용할 수도 있습니다 . 자세한 내용은 " 데이터 테이블 가상 결합 "(344 페이지) 에서 확인하십시오 .

고유 식별자로 테이블 결합

핫도그에 대한 정보 파일이 두 개 있습니다. 두 번째 파일에는 Size 열이 포함되어 있지만, 첫 번째 파일에는 이 열이 없습니다. 매칭 열로 결합을 사용하여 데이터 테이블을 결합합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hot Dogs.jmp 및 Hot Dogs2.jmp 를 엽니다 .

   Product Name 열은 데이터 테이블의 행에 대한 고유 식별자입니다 . 각 테이블에는 54 개의 고유한 제품 이름이 있습니다 .
2. Hot Dogs 데이터 테이블에서 **테이블 > 결합**을 선택합니다 .
3. 'Hot Dogs' 과 결합할 대상 목록에서 Hot Dogs2 를 선택합니다 .
4. " 매칭 규격 " 영역에서 **매칭 열별**을 선택합니다 .
5. Hot Dogs 및 Hot Dogs2 목록에서 Product Name 을 선택합니다 .
6. **매칭**을 클릭합니다 .

7. " 옵션 " 에서 **동일한 이름의 열 병합**을 선택합니다 . 열 병합을 사용하면 " 매칭 플래그 " 가 자동으로 선택됩니다 .
8. 결합된 데이터 테이블이 예상과 다를 경우 쉽게 결합 창으로 돌아갈 수 있도록 창의 왼쪽 하단에서 **대화상자 열린 채 유지**를 선택합니다 .
9. 적용을 클릭합니다 .

그림 6.29 매칭 열을 기준으로 결합된 테이블

	매칭 플래그	Product Name	Type	Taste	$/oz	$/lb Protein	Calories	Sodium	Protein/Fat	Size
1	둘 다	Happy Hill Supers	Beef	Bland	0.11	14.23	186	495	1	Jumbo
2	둘 다	Georgies Skinles...	Beef	Bland	0.17	21.70	181	477	2	Jumbo
3	둘 다	Special Market's ...	Beef	Bland	0.11	14.49	176	425	1	Jumbo
4	둘 다	Spike's Beef	Beef	Medium	0.15	20.49	149	322	1	Regular
5	둘 다	Hungry Hugh's J...	Beef	Medium	0.10	14.47	184	482	1	Regular
6	둘 다	Great Dinner Beef	Beef	Medium	0.11	15.45	190	587	1	Regular
7	둘 다	RJB Kosher Beef	Beef	Medium	0.21	25.25	158	370	2	Regular
8	둘 다	Wonder Kosher ...	Beef	Medium	0.20	24.02	139	322	2	Regular
9	둘 다	Happy Fats Jum...	Beef	Medium	0.14	18.86	175	479	1	Regular
10	둘 다	Midwest Beef	Beef	Medium	0.14	18.86	148	375	1	Regular
11	둘 다	General Kosher ...	Beef	Medium	0.23	30.65	152	330	1	Regular

이제 결과 테이블에 Hot Dogs2 테이블에 있었던 Size 열이 포함됩니다. 매칭 플래그 열은 모든 행이 두 테이블에 원래 있었던 데이터임을 나타냅니다.

행 수 및 열 이름이 서로 다른 테이블을 결합하려면

Sarah 와 Joe 가 팝콘에 대한 실험을 수행하고 있다고 가정해 보겠습니다. 두 사람은 서로 다른 유형의 팝콘 (gourmet 및 plain) 을 기름 양을 달리 하여 튀겨 본 후 만들어진 팝콘의 양 (yield) 을 기록하고 있습니다. Sarah 는 첫 번째 실험 데이터를 Trial1.jmp 라는 파일로 제공했습니다. Joe 는 두 번째 실험 데이터를 Little.jmp 라는 파일로 제공했습니다. 이 두 테이블을 하나의 테이블로 결합하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Trial1.jmp 및 Little.jmp 를 엽니다 .
2. Trial1.jmp 테이블에서 **테이블 > 결합**을 선택합니다 .
3. **... 과 (와) 결합할 대상** 상자에서 Little 을 선택합니다 .

   두 테이블에서 세 개의 열 (popcorn, oil amt/oil 및 batch) 에는 동일한 값이 포함되어 있습니다 . 이러한 열을 매칭 열로 지정합니다 . 또한 Sarah 와 Joe 가 oil 및 oil amt 라는 서로 다른 이름의 열을 제공했으므로 oil amt 와 oil 이 매칭됨을 JMP 에 알릴 수 있습니다 .
4. **주 테이블 순서 유지**를 선택 취소합니다 .

   결합된 테이블은 Trial1.jmp 의 데이터 순서가 아니라 매칭 열을 기준으로 정렬됩니다 .
5. " 매칭 규격 " 영역에서 **매칭 열별**을 선택합니다 .
6. Trial1 목록에서 popcorn, oil amt 및 batch 를 선택합니다 .
7. Little 목록에서 popcorn, oil 및 batch 를 선택합니다 .
8. **매칭**을 클릭합니다 .

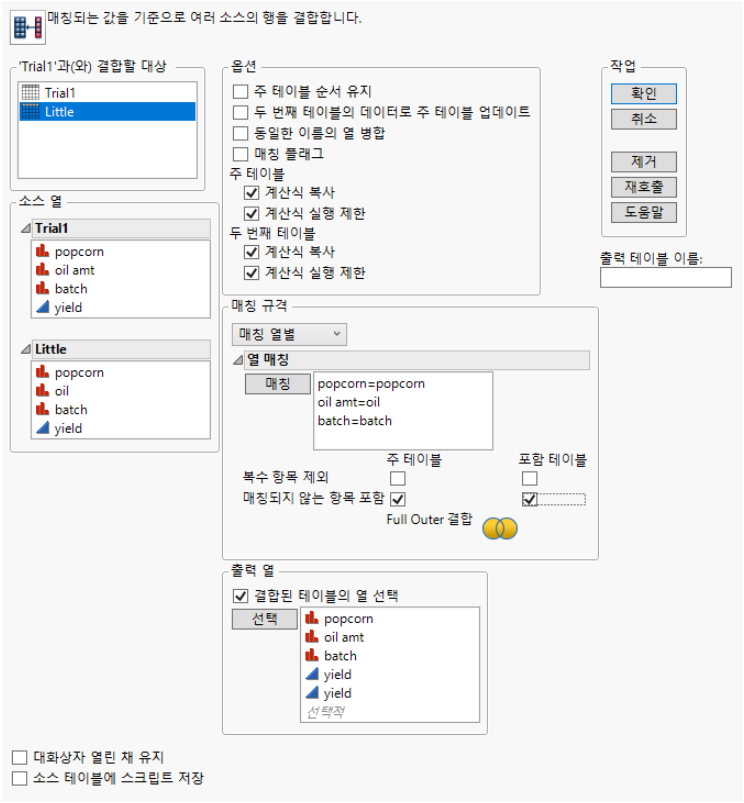
두 데이터 테이블을 살펴보면 행 수가 서로 다름을 알 수 있습니다. Trial1.jmp 에는 8 개의 실험 조건에 대한 값이 있고, Little.jmp 에는 4 개의 실험 조건에 대한 값만 있습니다. Sarah 는 실험을 완료했지만 Joe 는 실험을 부분적으로만 완료했기 때문입니다. Little.jmp 테이블의 행에 결측값이 포함되더라도 결합된 테이블에는 Trial1.jmp 의 모든 행을 포함하려고 합니다.

9. 두 테이블 모두에 대해 **매칭되지 않는 항목 포함** 상자를 선택합니다.

   결합된 테이블에는 popcorn, oil 및 batch 에 대한 열을 각각 하나씩만 포함하려고 합니다. 하지만 yield 에 대한 열은 두 개, 즉 Trial1.jmp 의 yield 를 나타내는 열 하나와 Little.jmp 의 yield 를 나타내는 열 하나를 포함하려고 합니다.

10. **결합된 테이블의 열 선택** 옆의 상자를 선택합니다.
11. Trial1 목록에서 모든 열을 선택합니다.
12. **선택**을 클릭합니다.
13. Little 목록에서 yield 를 선택합니다.
14. **선택**을 클릭합니다.

그림 6.30 완료된 결합 창



15. **확인**을 클릭합니다.

그림 6.31 결합된 Trial1.jmp 및 Little.jmp

	popcorn	oil amt	batch	Trial1의 yield	Little의 yield
1	gourmet	little	large	8.6	8.2
2	gourmet	little	small	12.1	15.9
3	gourmet	lots	large	9.2	•
4	gourmet	lots	small	18.0	•
5	plain	little	large	8.2	8.8
6	plain	little	small	9.9	10.1
7	plain	lots	large	10.4	•
8	plain	lots	small	10.6	•

결합된 테이블은 매칭 열을 기준으로 정렬되어 있습니다. Little.jmp 테이블에서 가져온 yield 열 (Little의 yield)에는 Trial1.jmp 테이블과 매칭되는 값이 없음을 나타내는 결측값이 있습니다.

중복 항목을 제거하기 위해 테이블을 해당 테이블과 자체 결합하려면

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Coffee Shop Purchases.jmp 를 엽니다 .

   일부 고객이 같은 날짜에 같은 음료를 구매했음을 알 수 있습니다 . **결합**을 사용하여 이러한 중복 행을 통합하려고 합니다 .
2. **테이블 > 결합**을 선택합니다 .
3. **... 과 (와) 결합할 대상** 상자에서 Coffee Shop Purchases 를 선택합니다 .
4. " 매칭 규격 " 영역에서 **매칭 열별**을 선택합니다 .
5. 두 개의 "Coffee Shop Purchases" 목록에서 세 개의 열 Date, Customer 및 Beverage 를 모두 선택합니다 .
6. **매칭**을 클릭합니다 .
7. 두 테이블 (주 테이블 및 포함 테이블) 모두에 대해 **복수 항목 제외** 상자를 선택합니다 .
8. **출력 테이블 이름**에 Coffee Shop Purchases Final 을 입력합니다 .
9. **확인**을 클릭합니다 .

그림 6.32 Coffee Shop Purchases 의 원래 데이터 테이블과 결합된 데이터 테이블

	Date	Customer	Beverage
1	01-05-2013	Jeff	Espresso
2	01-05-2013	Mike	Chai tea
3	01-05-2013	Sergio	Latte
4	01-05-2013	Mike	Espresso
5	01-05-2013	Mike	Espresso
6	01-05-2013	Lisa	Espresso
7	01-05-2013	Lisa	Espresso
8	02-13-2013	Lisa	Chai tea
9	02-13-2013	Lisa	Chai tea
10	02-13-2013	Mike	Chai tea
11	02-13-2013	Sergio	Espresso
12	02-13-2013	Sergio	Espresso

	Date	Customer	Beverage
1	01-05-2013	Jeff	Espresso
2	01-05-2013	Mike	Chai tea
3	01-05-2013	Sergio	Latte
4	01-05-2013	Mike	Espresso
5	01-05-2013	Lisa	Espresso
6	02-13-2013	Lisa	Chai tea
7	02-13-2013	Mike	Chai tea
8	02-13-2013	Sergio	Espresso

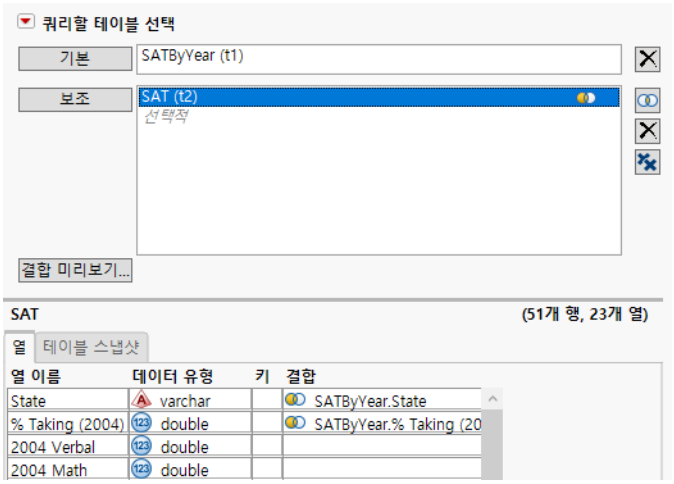
JMP 쿼리 빌더를 사용하여 데이터 테이블 쿼리 및 결합

" 테이블 " 메뉴의 "JMP 쿼리 빌더 " 옵션을 사용하면 데이터 테이블을 쿼리하고 선택한 데이터를 새 데이터 테이블에 저장할 수 있습니다 . 이 기능은 " 결합 " 명령을 사용하는 것과 유사하지만 데이터를 저장하기 전에 쿼리를 수행할 수 있다는 차이점이 있습니다 . 예를 들어 SAT 데이터를 쿼리하고 2004 년 시험 점수만 데이터 테이블에 저장할 수 있습니다 . 사용자가 쿼리를 실행하고 2004 년 시험 점수 중 일부를 선택할 수 있도록 사용자 확인 기능을 포함할 수도 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 SAT.jmp 및 SATByYear.jmp 를 엽니다 .
2. SATByYear.jmp 를 표시합니다 .
3. **테이블 > JMP 쿼리 빌더**를 선택합니다 .

   현재 데이터 테이블인 SATByYear.jmp 가 기본 테이블로 선택되어 있습니다 .
4. " 사용 가능한 테이블 " 목록에서 SAT.jmp 를 선택하고 **보조**를 클릭합니다 .
5. " 보조 " 버튼 옆의 SAT.jmp 를 선택하고 " 열 " 탭을 봅니다 . " 결합 " 열에서는 두 개의 열이 동일한 이름이고 결합되었음을 보여 줍니다 .

그림 6.33 결합된 열



6. **쿼리 생성**을 클릭합니다 .

열 추가

1. " 사용 가능한 열 " 목록에서 t1.State, t1.%Taking (2004), t2.2004 Verbal 및 t2.2004 Math 를 선택합니다 .
2. **추가**를 클릭합니다 .

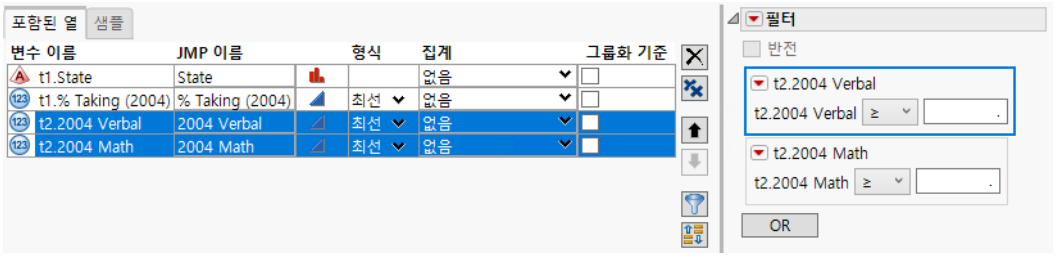
   해당 열이 " 포함된 열 " 탭에 추가됩니다 .
3. 중복 행이 저장되지 않도록 **구분되는 행만**을 선택합니다 .

4. 선택한 열을 볼 수 있도록 " 쿼리 미리보기 " 탭에서 **미리보기 자동 업데이트**가 선택되어 있는지 확인합니다 .

필터 추가

1. " 포함된 열 " 목록에서 t2.2004 Verbal 및 t2.2004 Math 를 선택하고 **선택한 항목을 필터에 추가** 를 클릭합니다 .
   해당 열이 " 필터 " 목록에 추가됩니다 .

그림 6.34 선택된 필터



2. 각 필터 옆의 빨간색 삼각형을 클릭하고 **실행 시 확인**을 선택합니다 . 기본 사용자 확인 메시지를 그대로 적용합니다 .

쿼리 실행

1. **쿼리 실행**을 클릭합니다 .
   데이터 테이블에 저장할 수학 및 언어 점수를 지정하라는 메시지가 표시됩니다 . 모든 데이터를 저장하려면 사용자 확인 창의 값을 변경하지 마십시오 .
2. **확인**을 클릭하여 데이터 테이블을 생성합니다 .

그림 6.35 쿼리 결과 데이터

	State	% Taking (2004)	2004 Verbal	2004 Math
1	Alabama	0.1	560	553
2	Alaska	0.53	518	514
3	Arizona	0.32	523	524
4	Arkansas	0.06	569	555
5	California	0.49	501	519
6	Colorado	0.27	554	553
7	Connecticut	0.85	515	515
8	District of Colum...	0.77	489	476
9	Delaware	0.73	500	499
10	Florida	0.67	499	499
11	Georgia	0.73	494	493

데이터베이스용 쿼리 빌더에도 나타나는 기능에 대한 자세한 내용은 " 데이터 가져오기 " 장의 " 쿼리 빌더에서 SQL 쿼리 작성 "(110 페이지) 에서 확인하십시오 .

팁 :

- 다른 소스 (예 : 데이터베이스 및 Microsoft Excel) 의 데이터를 결합하려면 쿼리 빌더를 사용하여 해당 데이터베이스 데이터를 데이터 테이블로 가져옵니다 . 예를 들어 Excel 데이터를 데이터 테이블로 가져온 후 " 테이블 " 메뉴의 "JMP 쿼리 빌더 " 를 사용하여 테이블을 쿼리하고 결합할 수 있습니다 .
- 쿼리를 열면 쿼리에 포함된 데이터 테이블이 숨겨진 파일로 열리며 JMP 홈 창에서 해당 파일을 열 수 있습니다 . 쿼리를 생성할 때 " 테이블 " 패널에서 숨겨진 파일을 열 수도 있습니다 .
- 열 수가 255 개를 초과하는 Microsoft Excel 파일을 쿼리하려고 하면 255 개의 열만 가져오게 됩니다 . 대신 **파일 > 열기**를 선택하여 Excel 파일을 Excel 가져오기 마법사에서 연 후 파일을 데이터 테이블로 저장하십시오 .

쿼리의 데이터 테이블 링크 정보

데이터 테이블을 쿼리할 경우 결과 데이터 테이블에는 쿼리 재실행, 데이터 업데이트 및 쿼리 편집을 위한 스크립트가 포함됩니다. 이러한 스크립트에서 데이터 테이블에는 **절대 경로**가 사용됩니다. 예를 들어 이러한 스크립트의 다음 부분에서는 두 개의 JMP 샘플 데이터 테이블이 저장되는 위치를 정의합니다.

```
JMP Tables(
     ["SAT" => "C:\Program Files\SAS\JMPPRO\16\Samples\Data\SAT.jmp",
     "SATByYear" => "C:\Program
  Files\SAS\JMPPRO\16\Samples\Data\SATByYear.jmp"]
  ),
```

하지만 쿼리를 .jmpquery 파일로 저장하면 파일에 원래 데이터 테이블의 **상대 경로**(생성 가능한 경우)가 포함됩니다. 다음 예에서는 .jmpquery 파일을 기준으로 한 상대 경로인 FuelData 하위 폴더에 저장된 데이터 테이블을 보여 줍니다.

```
JMP Tables( ["Cars" => "FuelData\Cars.jmp",
  "Trains" => "FuelData\Trains.jmp"] )
```

.jmpquery 파일에서 상대 경로를 생성할 수 없는 경우 가능하면 경로 변수 대체를 통해 절대 경로가 사용됩니다. 쿼리 실행 시 테이블을 찾을 수 없으면 데이터 테이블을 선택하라는 메시지가 표시됩니다. 그런 다음 JMP는 쿼리 내의 다른 누락 데이터 테이블이 선택된 폴더에 있는지 여부를 확인합니다.

경로 변수를 사용하여 데이터 테이블을 찾을 수도 있습니다. 예를 들어 $DOCUMENTS 폴더에 있는 데이터 테이블을 선택하는 스크립트를 작성할 수 있습니다. 또는 JSL 스크립트에서 경로 변수를 정의한 후 이 스크립트에서 쿼리를 실행할 수 있습니다. 경로 변수에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

데이터 테이블 가상 결합

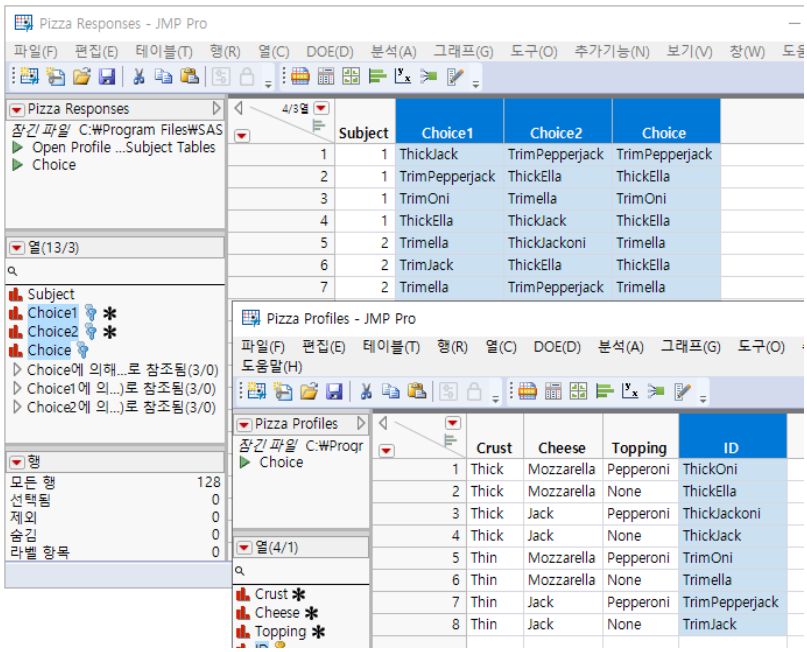
가상 결합은 주 데이터 테이블을 하나 이상의 보조 데이터 테이블에 연결하는 기능입니다. 이 기능을 사용하면 테이블을 실제로 결합하지 않고도 주 데이터 테이블에서 보조 데이터 테이블의 데이터에 액세스할 수 있습니다. 테이블을 가상으로 결합하면 동일한 데이터가 해당 데이터를 참조하는 모든 테이블마다 복사되지 않으므로 메모리가 절약됩니다. 또한 연결된 데이터를 업데이트하기가 더 간편합니다. 연결된 데이터는 참조하는 테이블에서 업데이트하지 않고 소스 테이블에서 독립적으로 업데이트할 수 있기 때문입니다.

" 연결 ID" 및 " 연결 참조 " 열 특성은 이러한 연결을 가능하게 합니다 .

- " 연결 ID" 열 특성은 **보조** 데이터 테이블의 열을 ID 열로 표시합니다 . 즉 , 데이터 테이블의 행이 ID 열의 값으로 고유하게 식별됩니다 . " 연결 ID" 열 특성이 있는 데이터 테이블을 **참조되는** 데이터 테이블이라고 합니다 .
- " 연결 참조 " 열 특성은 **주** 데이터 테이블의 열을 참조되는 데이터 테이블의 연결 ID 열에 연결합니다 . 이 열 특성은 참조되는 데이터 테이블의 경로 이름을 지정합니다 . 연결 참조가 있는 열을 **참조하는** 열이라고 합니다 . 참조하는 열은 연결 ID 열을 통해 보조 데이터 테이블의 데이터를 조회할 수 있습니다 .

그림 6.36에서는 가상으로 결합된 데이터 테이블의 예를 보여 줍니다. Pizza Profiles.jmp는 ID 열에 "연결 ID" 열 특성이 포함되어 있으므로 참조되는 데이터 테이블입니다. ID 열에는 참조하는 데이터 테이블인 Pizza Responses.jmp의 데이터에 해당하는 고유 값이 포함되어 있습니다.

그림 6.36 가상으로 결합된 데이터 테이블



참고 : 열에 " 연결 참조 " 열 특성이 있더라도 참조되는 데이터 테이블의 열에 반드시 연결해야 하는 것은 아닙니다 . " 열 " 목록의 열 옆에 있는 가상 결합 아이콘이 회색 인 경우에는 참조되는 데이터 테이블이 열려 있지 않거나 참조하는 열이 제외된 것일 수 있습니다 . 기타 가상 결합 아이콘에 대한 자세한 내용은 "JMP 시작하기 " 장의 " 가상 결합 아이콘 "(43 페이지) 에서 확인하십시오 .

연결 열 특성이 설정되어 있으면 참조되는 데이터 테이블의 모든 열이 주 데이터 테이블의 일부가 됩니다. 따라서 주 데이터 테이블에 참조하는 열이 있으면 참조되는 테이블의 해당 열이 시작 창의 "열 선택" 목록에 자동으로 나타납니다.

예 : 열 가상 결합

피자 실험에 참여한 응답자들이 각자 좋아하는 크러스트 , 토핑 및 치즈를 선택했다고 가정해 보겠습니다 . 응답자들이 좋아하는 치즈를 확인하려면 다음 단계를 수행하십시오 .

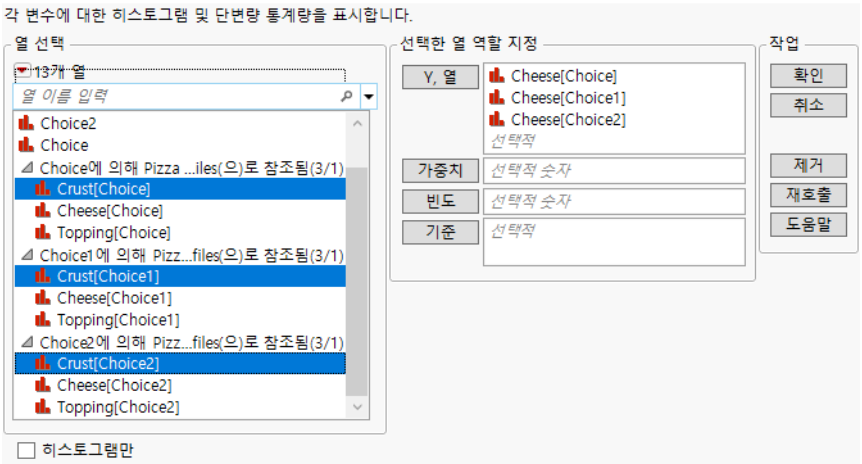
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Pizza Profiles.jmp 및 Pizza Responses.jmp 를 엽니다 .
2. 먼저 , Pizza Profiles.jmp 의 ID 열을 마우스 오른쪽 버튼으로 클릭하고 **연결 ID** 를 선택합니다 .

   이 열에는 Pizza Responses.jmp 의 값에 대응하는 고유 값이 포함되어 있습니다 . 예를 들어 Pizza Profiles.jmp 의 1 행은 두꺼운 크러스트 , 모짜렐라 치즈 및 페페로니 토핑을 사용한 피자의 별칭이 ThickOni 임을 나타냅니다 . Pizza Responses.jmp 에서 모든 응답이 "ThickOni" 이면 해당 피실험자가 그러한 속성의 피자를 선택한 것입니다 .
3. Pizza Responses.jmp 에서 "Choice" 로 시작하는 세 개의 열을 선택합니다 .
4. 마우스 오른쪽 버튼을 클릭하고 **연결 참조** > Pizza Profiles.jmp 를 선택합니다 .

   선택한 열이 Pizza Profiles.jmp 의 ID 열에 연결되고 주 데이터 테이블의 " 열 " 목록에 나타납니다 (그림 6.36(344 페이지)).

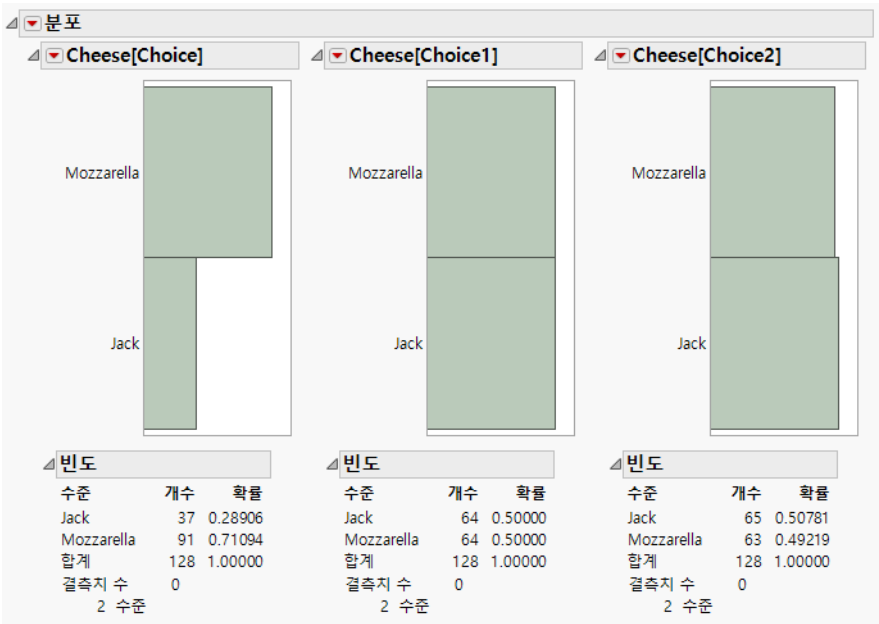
   참고 : 열의 " 연결 참조 " 특성 중 하나를 다시 살펴보면 Pizza Profiles.jmp 데이터 테이블이 선택되어 있음을 알 수 있습니다 . 링크 변경에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 가상 결합 특성을 사용하여 데이터 테이블 연결 "(299 페이지) 에서 확인하십시오 .
5. Pizza Responses.jmp 에서 **분석 > 분포**를 선택합니다 .
6. 아래로 스크롤하여 참조되는 열 그룹 중 세 개의 "Cheese" 열을 선택하고 **Y, 열**을 클릭한 후 **확인**을 클릭합니다 .

그림 6.37 분포 시작 창



Choice 열에서는 Mozzarella를 선택한 응답자가 더 많고, Choice1 및 Choice2에서는 Mozzarella와 Jack 의 비율이 거의 같습니다 .

그림 6.38 피자 응답 분포



참고 :

- 가상으로 결합된 열의 또 다른 예는 Movie Rentals.jmp 샘플 데이터 테이블에서 확인하십시오 . 이 데이터는 이미 Movie Customers.jmp 및 Movie Inventory.jmp 와 결합되어 있으므로 연결 참조를 지정할 필요가 없습니다 .

- 데이터 테이블 dp " 연결 ID" 열 특성은 하나만 포함될 수 있습니다 . " 연결 ID" 열 특성에는 중복 값이 포함될 수 없습니다 .
- 데이터 테이블에는 참조하는 열이 여러 개 포함될 수 있습니다 . 열에서 다른 데이터 테이블 또는 동일한 데이터 테이블을 참조할 수 있습니다 .
- 참조되는 데이터 테이블에 참조하는 열 (다른 데이터 테이블을 참조하는 열) 이 포함될 수도 있습니다 .
- 열에서 " 연결 참조 " 열 특성이 제거되거나 참조되는 데이터 테이블이 닫히면 해당하는 참조되는 열이 주 데이터 테이블에서 제거됩니다 .
- 연결된 열의 데이터 유형은 매칭되어야 합니다 .
- 표현식 및 행 상태 데이터 유형은 " 연결 참조 " 또는 " 연결 ID" 열 특성을 지원하지 않습니다 .
- 주 데이터 테이블을 저장하면 " 연결 참조 " 열 특성이 저장되지만 참조되는 열은 저장되지 않습니다 . 데이터 테이블을 열면 데이터 테이블이 다시 연결됩니다 .
- 참조되는 데이터 테이블의 이름을 바꾸면 해당하는 " 연결 참조 " 열 특성이 자동으로 업데이트됩니다 .
- " 테이블 " 패널의 빨간색 삼각형 메뉴에서 " 테이블 스크립트 복사 " 를 선택해도 참조되는 열은 포함되지 않습니다 .
- 가상으로 결합된 열이 포함된 데이터 테이블의 행 상태를 동기화하려면 행 상태 동기화 옵션을 선택하십시오 . 자세한 내용은 " 예 : 열 가상 결합 및 행 상태 표시 "(348 페이지) 에서 확인하십시오 .
- 열에는 " 연결 ID" 열 특성과 " 연결 참조 " 열 특성이 모두 있을 수 있습니다 .

팁 :

- 열을 마우스 오른쪽 버튼으로 클릭하여 " 연결 ID" 및 " 연결 참조 " 열 특성을 추가할 수 있을 뿐 아니라 열 정보 창을 통해서도 특성을 추가할 수 있습니다 . 자세한 내용은 "JMP 열 특성 설정" 장의 "가상 결합 특성을 사용하여 데이터 테이블 연결"(299페이지)에서 확인하십시오.
- " 열 " 목록에 표시된 가상으로 결합된 열에서 연결된 열 이름이 사용되도록 지정할 수 있습니다 . 이 옵션을 사용하면 더 짧은 열 이름을 지정할 수 있습니다 (예를 들어 Cheese[Choice] 대신 Cheese 지정). " 연결 참조 " 열 특성에서 **연결된 열 이름 사용**을 선택하십시오 . 일반적인 방식대로 열 이름을 직접 바꿀 수도 있습니다 . 이 테이블 환경 설정을 " 가상 결합 - 연결된 열 이름 사용 " 이라고 합니다 . 기본적으로 선택 취소되어 있습니다 .
- 데이터를 다른 사용자와 공유하려는 경우 데이터 테이블이 영구적으로 결합되도록 데이터를 병합할 수 있습니다 . 이 옵션을 사용할 경우 주 데이터 테이블과 보조 데이터 테이블을 제공하는 것이 아니라 데이터 테이블을 하나만 제공하면 되므로 유용합니다. 결합된 열을 연결 참조가 있는 테이블에 병합하려면 데이터 테이블의 왼쪽 페인에 있는 빨간색 삼각형 메뉴에서 **참조된 데이터 병합**을 선택하십시오 . 그러면 보조 데이터 테이블의 데이터가 주 데이터 테이블에 복사됩니다 . 주 데이터 테이블을 저장하면 실제 데이터도 저장됩니다 . 주 데이터 테이블의 참조하는 열에서는 " 연결 참조 " 열 특성이 제거됩니다 . 자세한 내용은 " 예 : 가상으로 결합된 데이터 테이블에서 참조된 데이터 병합 "(349 페이지) 에서 확인하십시오 .

예 : 열 가상 결합 및 행 상태 표시

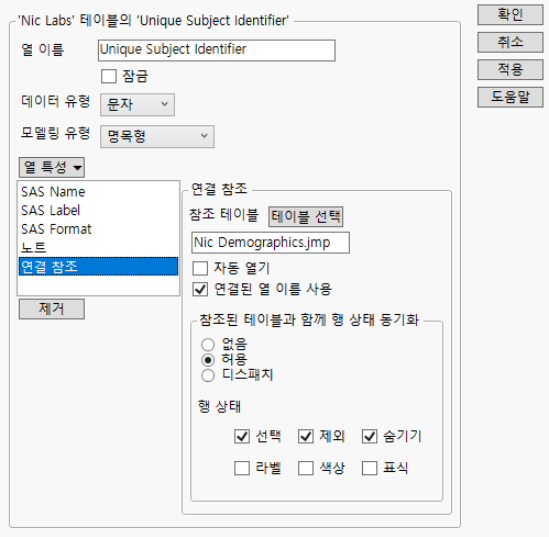
가상으로 결합된 열이 있는 데이터 테이블 하나에서 행 상태가 변경되면 연결된 다른 데이터 테이블의 가상으로 결합된 열에서 해당 행 상태가 자동으로 표시될 수 있습니다. 예를 들어 주 데이터 테이블과 보조 데이터 테이블에 피실험자 ID 열이 있다고 가정해 보겠습니다. 피실험자 ID 열에 "연결 ID" 특성이 포함된 데이터 테이블에서 행 상태가 변경되었습니다. 그러면 가상으로 결합된 "연결 참조" 열이 포함된 데이터 테이블에서 행 상태가 업데이트됩니다. "연결 참조" 열이 다른 데이터 테이블로 행 상태를 전송(**디스패치**)할지 아니면 행 상태를 수신(**허용**)할지 지정할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nic Demographics.jmp, Nic Labs.jmp 및 Nic Adverse Events.jmp 를 엽니다 .
2. Nic Labs.jmp 를 표시합니다 .

   " 열 " 목록에서 Unique Subject Identifier 옆에는 파란색 가상 결합 아이콘 이 나타납니다 . 파란색은 해당 열에 " 연결 참조 " 열 특성이 있고 참조되는 데이터 테이블 (Nic Demographics.jmp) 이 열려 있음을 나타냅니다 . 이 " 연결 참조 " 열 특성은 현재 데이터 테이블의 Unique Subject Identifier 열을 참조되는 데이터 테이블의 Unique Subject Identifier 열에 연결합니다 . 회색 아이콘 이 표시되면 참조되는 데이터 테이블이 열려 있지 않거나 열을 올바르게 연결할 수 없는 것입니다 .

   Nic Adverse Events.jmp 도 동일하게 설정되어 있습니다 .

그림 6.39 Unique Subject Identifier 의 연결 참조 열 특성



3. Nic Labs.jmp 에서 Unique Subject Identifier 열을 선택하고 **열 > 열 정보**를 선택합니다 .
4. **연결 참조** 열 특성을 선택합니다 .

   " 허용 " 이 선택되어 있습니다 . 이는 참조되는 데이터 테이블 (Nic Demographics.jmp) 의 행에 행 상태가 포함되어 있으면 현재 데이터 테이블에 행 상태가 자동으로 나타남을 의미합니다 .

또한 " 선택 ", " 제외 " 및 " 숨기기 " 행 상태도 선택되어 있습니다 . Nic Demographics.jmp 에서 이러한 행 상태를 선택하면 Nic Lib.jmp 및 Nic Adverse Events.jmp 에 해당 행 상태가 나타납니다 .

5. Nic Demographics.jmp 에서 첫 번째 행을 마우스 오른쪽 버튼으로 클릭하고 **숨기고 제외하기** 를 선택합니다 .

   첫 번째 행에는 Unique Subject Identifier 101001 에 대한 데이터가 포함되어 있습니다 .

6. Nic Labs.jmp 를 표시하여 Unique Subject Identifier 101001 에 대해 숨겨지고 제외된 행을 확인합니다 .

주의 :

- 여러 열이 있는 여러 테이블에서는 행 상태를 디스패치하기가 복잡할 수 있으므로 주의를 기울여야 합니다 .
- 행 상태 디스패치 기능은 동일한 ID 값에 여러 행 상태가 할당되지 않은 경우에 사용하는 것이 좋습니다 . " 연결 참조 " 열이 색상 행 상태를 디스패치하는 경우 ID 값이 동일하지만 색상 행 상태는 다른 행이 여러 개 있으면 모호성이 빠르게 증가합니다 . 예를 들어 데이터 테이블에 ID 값이 001 인 행이 두 개 있는 경우 , 한 행에는 빨간색을 지정하고 다른 행에는 녹색을 지정하면 JMP 는 연결된 데이터 테이블에서 행 상태를 설정할 방법을 알 수 없습니다 .

참고 :

- 테이블 A 의 여러 열이 테이블 B 에 연결되고 , 테이블 B 에는 테이블 C 에 연결되는 여러 열이 있을 수 있습니다 . 연결된 열이 포함된 데이터 테이블에는 모든 행의 행 상태가 나타납니다 . 예를 들어 테이블 A 에서 1 행이 제외되어 있고 테이블 B 에서 1 행이 숨겨져 있다면 주 데이터 테이블에서는 해당 행이 제외되고 숨겨진 것으로 나타납니다 .
- 데이터 테이블에서 특정 행 상태를 허용하고 있는 경우에는 데이터 테이블에서 해당 행 상태를 수동으로 적용하지 않는 것이 좋습니다 . 디스패치 열에서 해당 행 상태를 제어하도록 하십시오 .

예 : 가상으로 결합된 데이터 테이블에서 참조된 데이터 병합

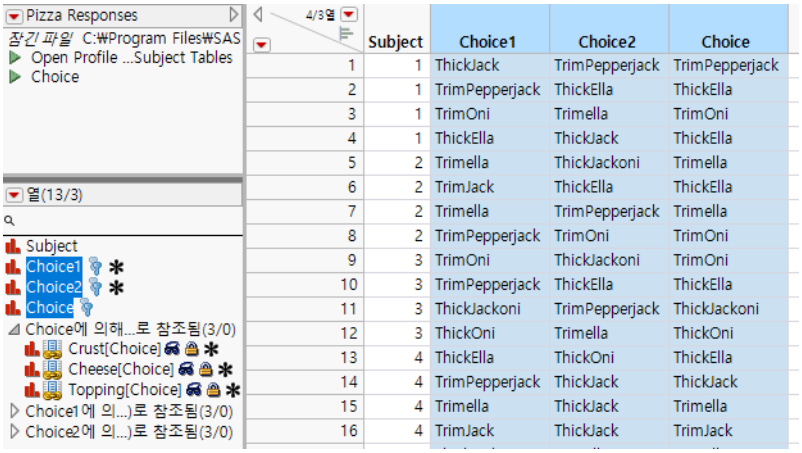
가상으로 결합된 데이터 테이블의 데이터를 다른 사용자와 공유하려는 경우 데이터 테이블이 영구적으로 결합되도록 데이터를 병합할 수 있습니다 . 이 옵션을 사용할 경우 주 데이터 테이블과 보조 데이터 테이블을 제공하는 것이 아니라 데이터 테이블을 하나만 제공하면 되므로 유용합니다 .

참조되는 데이터를 병합하면 보조 데이터 테이블의 데이터가 주 데이터 테이블에 복사됩니다 . 주 데이터 테이블을 저장하면 실제 데이터도 저장됩니다 .

1. **도움말 > 샘플 데이터 라이브러리** 를 선택하고 Pizza Profiles.jmp 및 Pizza Responses.jmp 를 엽니다 .
2. Pizza Profiles.jmp 의 ID 열을 마우스 오른쪽 버튼으로 클릭하고 **연결 ID** 를 선택합니다 .
3. Pizza Responses.jmp 에서 Choice1, Choice2 및 Choice 열을 선택한 다음 마우스 오른쪽 버튼을 클릭하고 **연결 참조** > Pizza Profiles.jmp 를 선택합니다 .

선택한 열이 Pizza Profiles.jmp 의 ID 열에 연결됩니다 .

그림 6.40 가상으로 결합된 데이터 테이블의 선택된 열

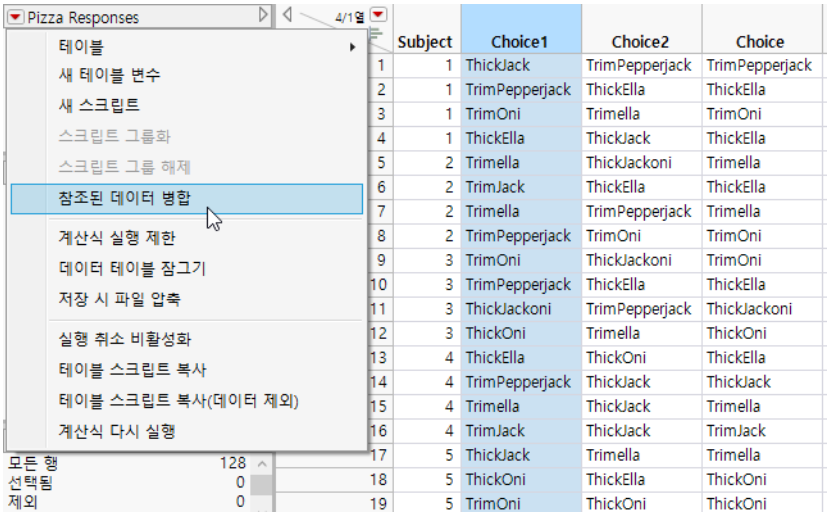


4. Pizza Responses.jmp 에서 Choice1 열을 선택합니다 (선택되어 있지 않은 경우).

참고 : 열을 선택하지 않으면 모든 열이 병합된 데이터에 포함됩니다 .

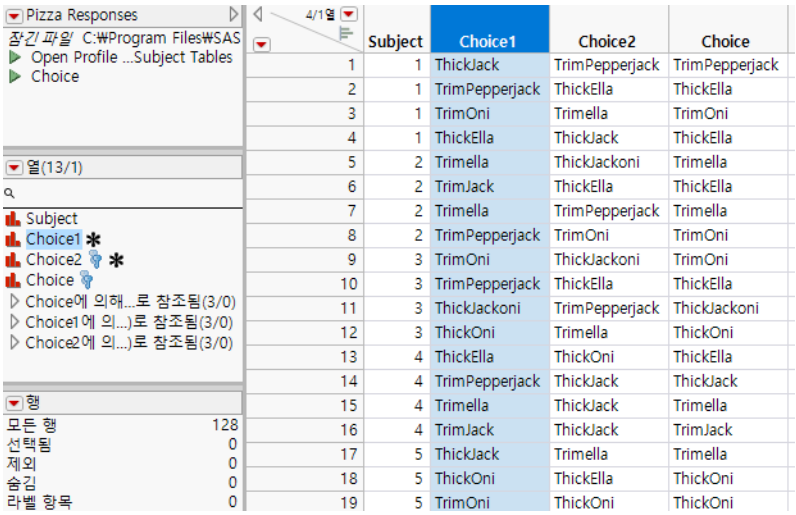
5. 데이터 테이블 이름 옆의 빨간색 삼각형을 클릭하고 **참조된 데이터 병합**을 선택합니다 .

그림 6.41 참조된 데이터 병합



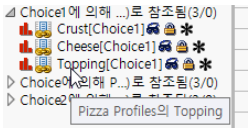
선택된 열이 병합되었다는 메시지가 표시됩니다 . 열 목록에서 Choice1 에 표시되던 가상 결합 아이콘은 더 이상 표시되지 않습니다 . 연결 참조 열 특성이 제거되었습니다 .

그림 6.42 병합된 데이터



"열" 패널에서 연결된 열 중 하나를 마우스로 가리키면 해당 열이 연결된 대상 데이터 테이블을 나타내는 툴팁이 표시됩니다.

그림 6.43 연결된 열이 포함된 데이터 테이블



단일 열의 연결 ID 및 연결 참조 예

직원 데이터를 여러 데이터 테이블에서 사용할 수 있으며 "연결 ID" 및 "연결 참조" 특성이 모두 있는 단일 열을 통해 서로 연결할 수 있다고 가정해 보겠습니다. 데이터 테이블 설정 방법의 예를 보려면 다음 단계를 수행하십시오.

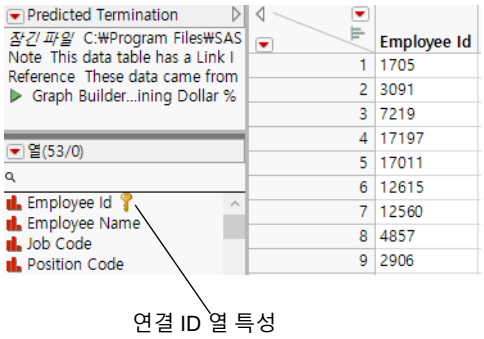
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Employee Master.jmp 를 엽니다 .

   참고 : "Employee Id" 열의 " 연결 참조 " 열 특성에서 " 자동 열기 " 가 선택되어 있으면 Education History.jmp 와 Predicted Termination.jmp 라는 다른 두 테이블이 자동으로 열립니다 . 이 두 테이블은 홈 창의 " 창 목록 " 에 표시됩니다 .

   – Employee Master.jmp 에는 각 직원에 대한 정보가 있습니다 .
   – Education History.jmp 에는 각 직원의 학력에 대한 정보가 있습니다 .
   – Predicted Termination.jmp 에는 각 직원에 대한 직무 성과 정보가 있습니다 .
2. 홈 창의 " 창 목록 " 에서 Predicted Termination.jmp 를 엽니다 .

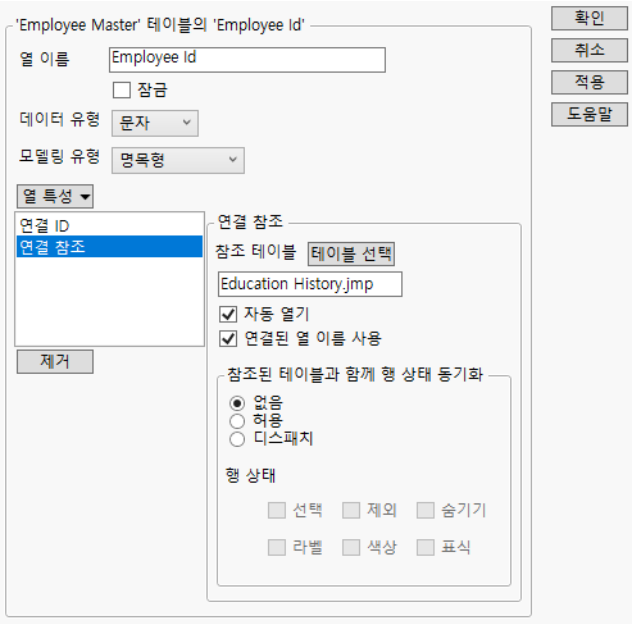
"Employee Id" 열에는 " 연결 ID" 열 특성이 있습니다 .

그림 6.44 연결 ID 열 특성



3. Employee Master.jmp 에서 Employee Id 열을 선택하고 **열 > 열 정보**를 선택합니다 .

그림 6.45 Employee Master.jmp 의 연결 ID 및 연결 참조 열 특성



이 열에는 " 연결 ID" 열 특성과 Education History.jmp 에 대한 " 연결 참조 " 가 있습니다 . 데이터 테이블의 열 목록에서 Employee Id 열 옆에는 연결 참조 아이콘 과 연결 ID 아이콘 이 표시됩니다 .

4. 홈 창의 " 창 목록 " 에서 Education History.jmp 를 엽니다 .
5. Employee Id 열을 선택하고 **열 > 열 정보**를 선택합니다 .

그림 6.46 Education History.jmp 의 연결 ID 및 연결 참조 열 특성

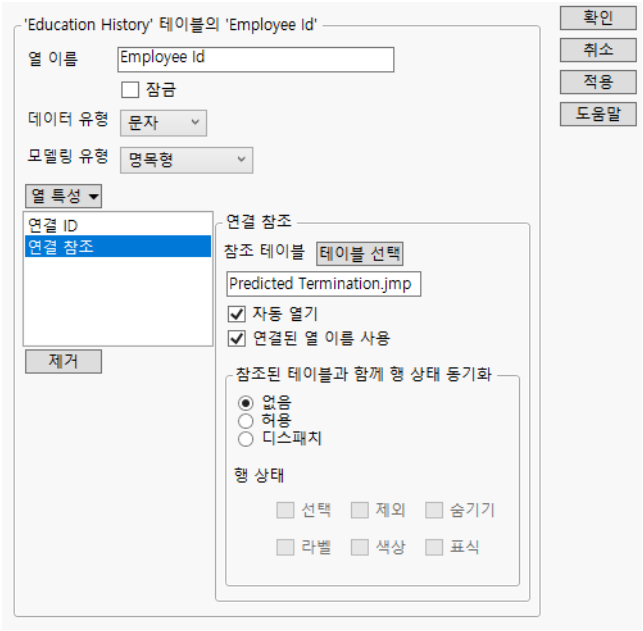
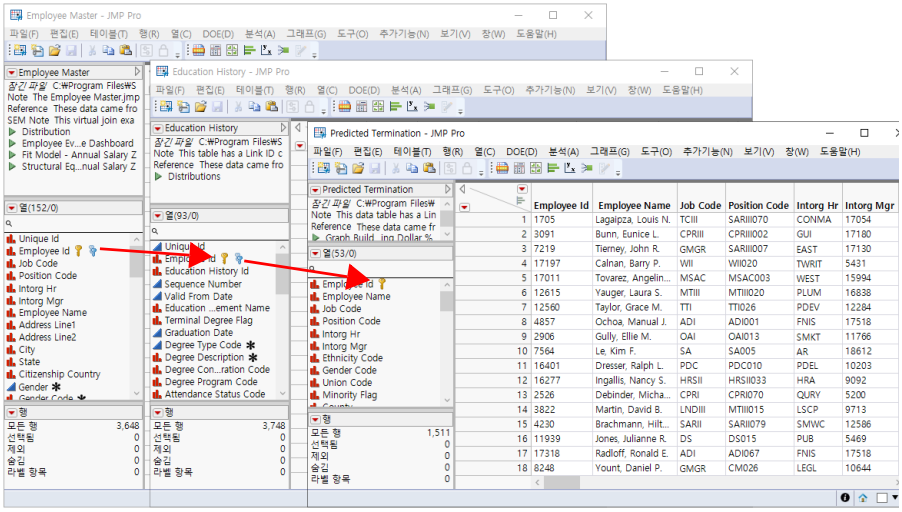


그림 6.46에 표시된 것과 같이 Employee Id에는 "연결 ID" 열 특성과 Predicted Termination.jmp에 대한 " 연결 참조 " 가 있습니다 .

그림 6.47 연결된 데이터 테이블



Employee Master.jmp 의 Employee Id 열은 Education History.jmp 를 참조하고 이 데이터 테이블은 Predicted Termination.jmp 를 참조합니다 . 모든 데이터는 주 데이터 테이블인 Employee Master.jmp 의 가상 결합을 통해 사용할 수 있습니다 .

참고 :

- 열의 " 연결 참조 " 특성에 있는 " 자동 열기 " 는 주 데이터 테이블이 열릴 때 연결된 테이블을 자동으로 열지 여부를 지정합니다 .
- 데이터 테이블이 저장될 때 참조된 열이 숨겨져 있지 않으면 해당 테이블이 열릴 때 참조되는 해당 열이 자동으로 연결되고 숨김 해제됩니다 .
- 이 테이블과 다른 모든 테이블 사이에는 연결이 있으므로 이 데이터 테이블을 닫을 때는 주 테이블 (Employee Master.jmp) 을 먼저 닫는 것이 좋습니다 . 테이블을 다른 순서로 닫으면 닫으려는 테이블에 다른 창이 열려 있다는 메시지가 표시될 수 있습니다 . 이 경우 닫기를 취소하고 테이블을 다른 순서로 저장 후 닫을 수 있습니다 .

데이터 테이블 업데이트

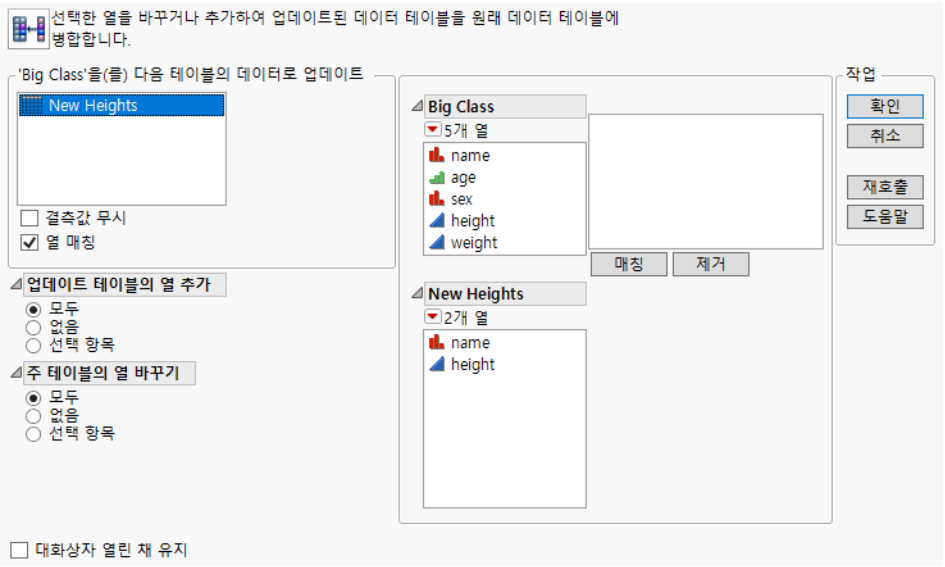
두 개의 데이터 테이블이 있는 경우 원래 테이블을 새 테이블의 데이터로 업데이트하려면 **테이블 > 업데이트**를 선택합니다 . "업데이트 "는 " 결합 "을 특정 상황에 맞게 조정한 명령입니다 . 이 명령은 " 업데이트 " 옵션이 선택된 " 결합 "에 해당하며 새 테이블을 생성하지는 않습니다 .

테이블을 업데이트하기 전에 바꾸려는 원래 값이 포함된 열의 이름이 바꾸려는 새 데이터가 포함된 열의 이름과 동일한지 확인해야 합니다 .

활성 테이블의 값을 열려 있는 다른 테이블의 값으로 바꾸려면

1. 업데이트하려는 원래 테이블 (바꾸려는 값이 포함된 테이블) 을 클릭하여 활성화합니다 .
2. **테이블 > 업데이트**를 선택합니다 .

그림 6.48 테이블 업데이트

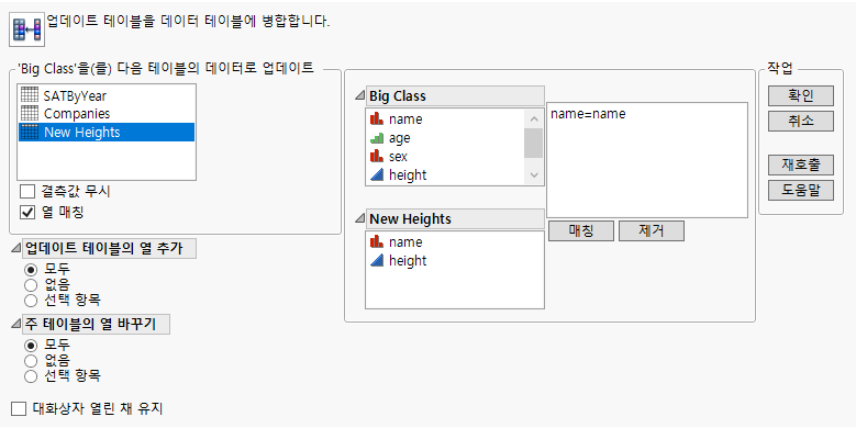


3. 원래 테이블로 전송할 데이터가 포함된 새 테이블을 강조 표시합니다 .
4. (선택 사항) 원래 테이블의 값이 새 테이블에 있는 결측값으로 바뀌지 않도록 하려면 **결측값 무시** 옆의 상자를 선택합니다 . 그러면 원래 테이블에서 새 테이블의 결측값에 대응하는 원래 값이 그대로 유지됩니다 .
5. 두 테이블에 각 행을 고유하게 식별하는 값이 포함된 열이 하나 이상 있으면 JMP 는 해당 열을 **매칭 열 값**으로 사용합니다 . 즉 , JMP 는 매칭 열 값이 매칭되는 행을 업데이트합니다 . JMP 는 이러한 열을 사용하여 데이터의 정렬 순서를 유지합니다 . 테이블에 매칭되는 열 값이 없는 경우에는 다음 작업을 계속하여 행 순서에 따라 값을 통합할 수 있습니다 . 매칭되는 열 값이 포함된 테이블의 경우 업데이트 방법은 " 매칭 열을 사용하여 테이블을 업데이트하려면 "(355 페이지) 에서 확인하십시오 .
6. " 업데이트 테이블의 열 추가 " 영역에서 옵션을 선택합니다 . 다음과 같은 옵션을 사용하여 존재하지 않는 열을 새 테이블에서 원래 테이블에 추가할 수 있습니다 .
   - 새 테이블의 모든 열을 원래 테이블에 추가하려면 **모두**를 선택합니다 .
   - 새 테이블에서 선택한 열만 원래 테이블에 추가하려면 **선택 항목**을 선택합니다 .
   - 존재하지 않는 열을 새 테이블에서 원래 테이블로 추가하지 않으려면 **없음**을 선택합니다 .
7. (선택 사항) " 주 테이블 열 바꾸기 " 영역에서 옵션을 선택합니다 . 이 옵션을 사용하여 같은 이름을 가진 열의 값을 바꿀 수 있습니다 .
   - 새 테이블에서 같은 이름을 가진 모든 열을 원래 테이블에 추가하려면 **모두**를 선택합니다 .
   - 새 테이블에서 선택한 같은 이름의 열만 원래 테이블에 추가하려면 **선택 항목**을 선택합니다 .
   - 같은 이름의 존재하지 않는 열을 새 테이블에서 원래 테이블로 추가하지 않으려면 **없음**을 선택합니다 .
8. **확인**을 클릭합니다 .

매칭 열을 사용하여 테이블을 업데이트하려면

1. 앞의 섹션 (" 활성 테이블의 값을 열려 있는 다른 테이블의 값으로 바꾸려면 "(354 페이지)) 에 설명된 처음 세 단계를 수행합니다 .
2. **열 매칭**을 선택합니다 .

그림 6.49 열 매칭 클릭 시 나타나는 창



3. 각 테이블에서 매칭하려는 두 개의 열 이름을 강조 표시합니다 .
4. **매칭**을 클릭합니다 .
5. (선택 사항) 추가 열을 매칭하려면 이전 단계를 반복합니다 .
6. **확인**을 클릭합니다 .

참고 : 업데이트하려는 테이블 (또한 매칭 항목이 할당되지 않은 테이블) 의 열과 이름이 다른 열은 별도의 열로 나타납니다 . 이 문제를 방지하려면 " 업데이트 테이블의 열 추가 " 영역에서 " 없음 " 옵션을 선택합니다 .

예 : 데이터 테이블 업데이트

한 연구원에게 학생의 키 측정값이 포함된 데이터 테이블이 있다고 가정해 보겠습니다. 이 연구원이 보다 최근의 학생 키 측정값이 포함된 업데이트된 테이블을 받았습니다. 연구원은 데이터 테이블을 스크롤해가며 키가 변경된 학생을 찾고 새 값을 복사해서 붙여넣는 작업을 피하고 싶어 합니다. 이 경우 **업데이트** 명령을 사용하면 원래 데이터 테이블을 새 키 값으로 빠르게 업데이트할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 및 New Heights.jmp 를 엽니다 .

   Big Class.jmp 테이블에는 원래 데이터가 포함되어 있고 , New Heights.jmp 테이블에는 업데이트된 데이터가 포함되어 있습니다 .
2. Big Class.jmp 테이블에서 **테이블 > 업데이트**를 선택합니다 .
3. **다음 테이블의 데이터로 업데이트** 상자에서 New Heights 를 선택합니다 .
4. **열 매칭**을 선택합니다 .
5. Big Class 및 New Heights 목록에서 name 을 선택합니다 .
6. **매칭**을 클릭합니다 .

name 열의 값은 각 행을 고유하게 설명하므로 이렇게 하면 JMP 는 해당 열을 매칭 열 값으로 사용합니다 .

7. **확인**을 클릭합니다 .

그림 6.50 업데이트된 Big Class.jmp 테이블

	name	age	sex	height	weight
1	KATIE	12	F	62	95
2	LOUISE	12	F	61	123
3	JANE	12	F	58	74
4	JACLYN	12	F	68	145
5	LILLIE	12	F	52	64
6	TIM	12	M	64	84
7	JAMES	12	M	63	128
8	ROBERT	12	M	70	79
9	BARBARA	13	F	62	112
10	ALICE	13	F	61	107
11	SUSAN	13	F	58	67
12	JOHN	13	M	68	98
13	JOE	13	M	64	105
14	MICHAEL	13	M	60	95

이제 업데이트된 테이블의 키 값은 더 이상 원래 테이블의 값과 매칭되지 않습니다. New Heights.jmp 테이블의 최신 값을 사용하도록 값이 업데이트되었습니다.

데이터 익명화

데이터 익명화 기능을 사용하면 일부 고유 식별자가 제거된 새 데이터 테이블을 생성할 수 있습니다.

- 열 머리글 , 문자 데이터 및 값 라벨이 수정됩니다 .
- 명목형 열의 데이터가 수정됩니다 .
- 순서형 열의 수정된 데이터가 원래 데이터 테이블에 동일한 순서로 나타납니다 .
- 연속형 열의 데이터는 수정되지 않습니다 .
- 열 특성의 열 이름이 수정됩니다 .

참고 : 테이블 스크립트와 일부 테이블 변수는 익명화할 수 없습니다. 데이터 테이블에 익명화할 수 없는 내용이 포함되어 있으면 로그에 경고 메시지가 나타납니다 .

"노트" 열 특성은 제거됩니다. 다른 열 특성은 데이터 테이블의 변경 사항에 따라 수정됩니다. 예를 들어 "값 순서" 특성의 이름이 바뀝니다. 예측 계산식을 제외한 계산식은 업데이트됩니다. 일부 열 특성은 열에 의도한 영향을 더 이상 미칠 수 없게 됩니다.

데이터의 이름을 바꾸려면 다음 중 하나를 수행하십시오 .

- 특정 열에 있는 데이터의 이름을 바꾸려면 열을 선택한 후 **테이블 > 익명화**를 선택합니다 .

- 전체 데이터 테이블에 있는 데이터의 이름을 바꾸려면 **테이블 > 익명화**를 선택합니다 .

새 데이터 테이블에 데이터가 나타납니다.

고지 사항: 데이터 익명화 유틸리티는 위에 설명된 것과 같이 단순히 데이터 테이블에서 일부 고유 식별자를 제거하려고 하는 사용자를 지원하기 위해 제공됩니다. 데이터 익명화 유틸리티로 민감하거나 개인 식별이 가능한 정보를 데이터 테이블에서 완전히 제거하지 못할 수도 있습니다. 법률 또는 정책에 따라 데이터에 대한 완전한 익명화, 유사 익명화 또는 개인 정보 비식별화가 요구되는 경우(데이터가 타인에게 공개될 수 있는 경우 등) 이러한 정보를 제거할 때는 전적으로 데이터 익명화 유틸리티에만 의존해서는 안 됩니다. SAS는 데이터 익명화 유틸리티를 사용하는 것만으로 사용자가 민감하거나 개인 식별이 가능한 정보의 개인 정보 보호, 개인 정보 비식별화 또는 익명화 등과 관련된 국가, 주, 지역 또는 국제적 법률, 규정 또는 정책을 준수할 수 있다고 단언하지 않으며 이를 명확히 부인합니다.

---

7 장

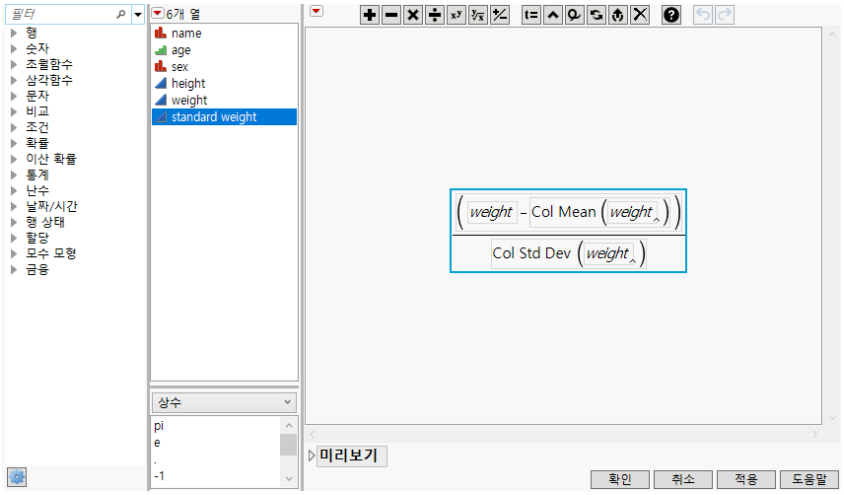
JMP 의 계산식 생성

계산식 편집기

JMP 계산식 편집기를 사용하면 계산식에 따라 값이 계산되는 열을 생성하고 계산식을 열 정보의 일부로 저장할 수 있습니다. 계산식은 단순히 숫자, 문자 또는 행 상태 상수를 대입하는 것일 수도 있고, 조건 절에 기반한 복잡한 계산을 포함할 수도 있습니다. 계산식 편집기 창은 버튼, 표시 화면 및 함수 목록이 있는 계산기처럼 작동합니다.

계산식 편집기에서 사용할 수 있는 함수에 대한 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 장(675 페이지)에서 확인하십시오.

그림 7.1 계산식 편집기



목차

계산식 개요 . . . 361
계산식 작성 . . . 361
계산식에서 데이터 테이블 값 참조 . . . 363
계산식에서 지역 변수 사용 . . . 365
    계산식에서 모수 포함 . . . 367
계산식에서 상수 삽입 . . . 368
JMP 계산식에 연산자 추가 . . . 369
    계산식 편집기의 키패드 참조 . . . 369
계산식에서 함수 사용 . . . 371
    계산식 편집기에서 자주 사용하는 함수 찾기 . . . 374
계산식 복사 및 반복 . . . 375
계산식의 표현식 정렬 . . . 377
    우선 순위대로 계산식 작성 . . . 378
JMP 계산식 편집기 옵션 . . . 379
    JMP 함수의 도함수 계산 . . . 379
    복잡한 계산식 단순화 . . . 380
    계산식 실행 . . . 381
    계산식에서 오류 무시 . . . 381
    JSL에서 계산식 보기 . . . 382
계산식 편집 . . . 382
    계산식에서 오류 수정 . . . 382
    계산식에서 표현식 선택 . . . 383
    함수 계산식 삭제 . . . 383
    계산식 잘라내기, 복사 및 붙여넣기 . . . 383
    계산식 클릭 및 드래그 . . . 384
JMP 계산식 편집기에서 계산식 사용자 정의 . . . 384
    계산식에서 개요 숨기기 및 표시 . . . 385
    계산식에서 글꼴 크기 변경 . . . 385
    계산식에서 행렬 크기 지정 . . . 385
    JMP 계산식의 방향 변경 . . . 385
    계산식에서 인수 열기 및 닫기 . . . 386
JMP 계산식 편집기의 계산식 작성 예 . . . 386
    JMP 계산식 편집기에서 기본 기능 사용 . . . 386
    계산식에서 지역 변수 사용 . . . 388
    계산식에서 Munger 함수 사용 . . . 388
    계산식에서 Match 조건 함수 사용 . . . 390
    계산식에서 표현식 삭제 버튼 사용 . . . 390
JMP 계산식 편집기의 바로 가기 키 . . . 391
JMP 계산식 편집기의 용어 . . . 392

계산식 개요

계산식은 다음과 같은 이유로 데이터 테이블의 핵심 요소입니다.

- 계산식은 데이터 테이블을 저장할 때 열 정보의 일부로 저장됩니다 .
- 언제든지 계산식 편집기를 열어 계산식을 검토하거나 변경할 수 있습니다 .
- 계산식 값은 다른 열의 값에 연결되거나 종속될 수 있습니다 . 계산식이 연결된 열의 값을 편집할 때마다 계산식 값이 자동으로 다시 계산됩니다 .
- 데이터 테이블에서는 수동으로 편집할 수 없도록 계산식 값이 잠겨 있습니다 .

이 장에서는 계산식 편집기에 대해 설명하고 계산식 작성 방법을 보여 줍니다. 각 함수에 대한 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 장 (675 페이지) 에서 확인하십시오.

계산식 작성

계산식은 열에 저장되는 표현식으로, 해당 열에 값을 삽입하기 위한 연산을 수행합니다. 계산식은 덧셈 및 곱셈과 같은 수학적 연산을 수행할 수도 있고, 동일한 데이터 테이블의 다른 영역을 참조하여 열 값을 비교하거나 값을 결합할 수도 있습니다. 계산식은 JSL(JMP 스크립트 언어) 명령으로 구성될 수 있습니다. 열에 계산식을 삽입한 후에는 열이 잠기므로 계산식을 변경하거나 제거하는 방법으로만 계산식 값을 편집할 수 있습니다.

다음은 계산식을 작성하는 기본적인 세 단계입니다.

1. 계산식을 적용할 열의 이름을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식 편집기를 엽니다 .

   또는

   계산식을 적용할 열의 이름을 두 번 클릭하고 , **열 특성** 메뉴에서 **계산식**을 선택한 후 **계산식 편집**을 클릭합니다 .

   계산식 편집기에는 빈 계산식이 선택되어 있습니다 (파란색으로 강조 표시됨). 입력을 시작하거나 열 또는 함수를 선택할 수 있습니다 .

   – 직접 입력할 경우 계산식을 편집할 수 있는 작은 텍스트 편집기 창이 나타납니다 . 더 큰 공간이 필요한 경우 **편집기 최대화** 버튼 을 클릭하면 됩니다 . 긴 계산식은 기본적으로 최대화된 편집기에서 열립니다 .

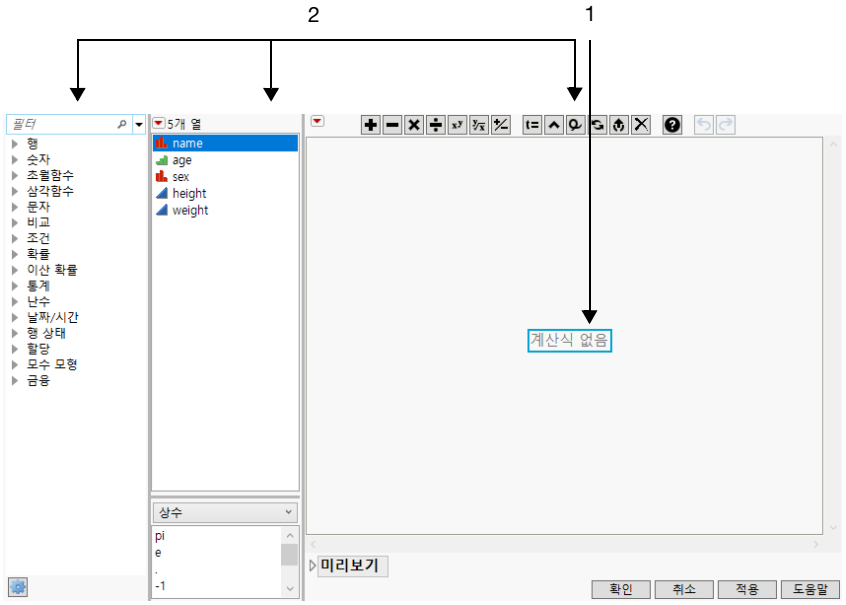
   – 열 또는 함수를 선택하면 해당 항목이 파란색 선택 항목 상자에 추가됩니다 .

   – " 열 " 목록의 열을 선택 항목 상자로 드래그할 수도 있습니다 .

   참고 : 선택된 요소 주위에는 파란색 외곽선이 표시됩니다 . 클릭한 위치를 기준으로 가장 작은 내포 상자 내의 모든 항이 선택됩니다 . 이후 작업은 이러한 결합된 요소에 적용됩니다 .

2. 표현식 , 함수 및 항을 추가합니다 . 그러면 추가한 항목이 강조 표시된 파란색 상자에 나타납니다 . 이 장의 이후 섹션에서는 상수 , 요소 , 연산자 및 함수를 추가하는 방법에 대해 자세히 설명합니다 .

그림 7.2 계산식 작성



계산식 편집기의 사용 방법을 보여 주는 예는 "JMP 계산식 편집기에서 기본 기능 사용"(386 페이지)에서 확인하십시오.

참고 :

- 계산식에서 값 라벨을 사용하여 열을 참조하는 경우 값 라벨을 마우스로 가리키면 실제 데이터 값이 표시됩니다 .
- 계산식 편집기는 제외됨 상태를 인식하며 해당 행을 계산에서 제외합니다 . 다음은 Col Maximum() 함수를 사용한 예입니다 .

  ```
  Col Maximum( :height, :sex, Excluded( Row State() ) )
  ```

- JMP 는 자동으로 계산식에서 후행 공백을 제거합니다 .
- 계산식 편집기를 사용하여 사용자 형식을 정의할 수도 있습니다. 예를 들어 계산 Char(:height * 2.54) || " cm" 을 사용하여 숫자를 인치 단위에서 센티미터 단위로 변환할 수 있습니다 . 하지만 데이터 테이블 셀을 두 번 클릭할 때 표시되는 기본 데이터의 값은 여전히 :height 입니다 .
- 커서를 계산식의 한 필드에서 다른 필드로 이동하려면 필드를 선택하고 화살표 키를 누르십시오 .
- Col Mean 과 같은 집계 함수의 경우 , 계산식을 두 개의 열로 분리하면 더 빠르게 실행할 수 있습니다 . 한 열에는 집계 함수의 대상을 나타내는 표현식이 포함되어야 합니다 . 두 번째 열

에는 새 대상 열을 참조하는 집계 함수 자체가 포함되어야 합니다 . 다음은 원래 계산식의 예입니다 .

```
:Y - Col Mean( If( :Status == "nonsmoker", :Y, . ), :Cycle )
```

새 Y2 열에는 다음 계산식을 생성하십시오 .

```
If( :Status == "nonsmoker",
   :Y,
   .
)
```

또 다른 열에는 다음 계산식을 생성하십시오 .

```
:Y - Col Mean( :Y2, :Cycle )
```

계산식에서 데이터 테이블 값 참조

데이터 테이블의 다른 부분(예: 다른 열 및 테이블 변수)에 있는 값을 참조하는 계산식을 생성할 수 있습니다.

계산식에서 다른 열의 값을 사용하는 경우 해당 계산식이 있는 열의 값은 다른 열의 값에 따라 달라집니다. 계산식에서 참조하는 열이 변경될 때마다 종속된 열도 변경됩니다. 참조되는 열을 삭제하면 해당 계산식이 포함된 열에 빈 항이 나타납니다.

테이블 변수에 있는 값을 참조하는 계산식을 생성할 경우 해당 테이블 변수는 데이터 테이블에 이미 있어야 합니다. 테이블 변수는 테이블 전체에서 사용할 수 있는 문자열입니다. 테이블 변수의 이름은 데이터 테이블 왼쪽의 테이블 패널에 표시됩니다. 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블 변수 사용 "(261 페이지)에서 확인하십시오.

팁 : 계산식 편집기에서 특정 값을 사용하여 테이블 변수를 생성하려면 해당 값을 마우스 오른쪽 버튼으로 클릭하고 **값에서 테이블 변수 생성**을 선택하십시오 .

다음 예에서는 Companies.jmp 샘플 데이터 테이블의 계산식 열을 다시 생성하는 방법을 보여 줍니다.

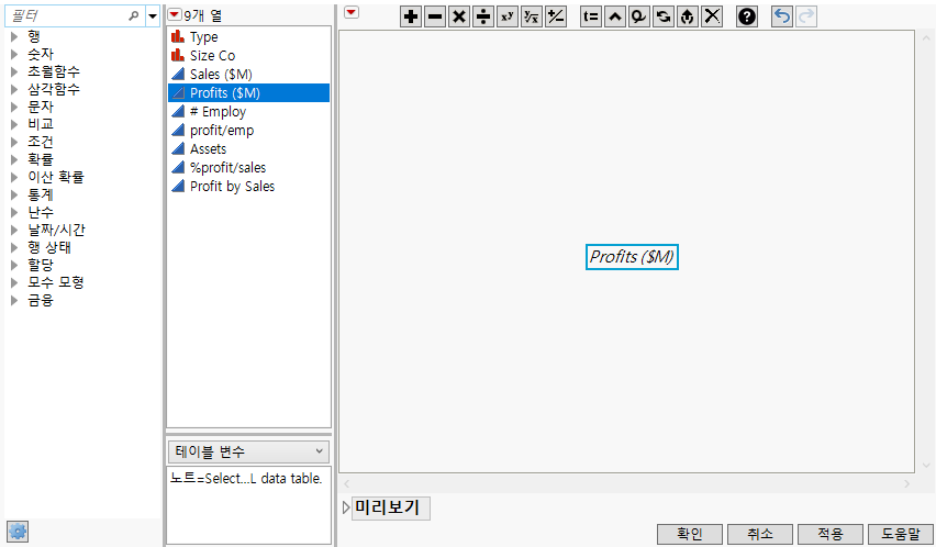
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. 마지막의 빈 열을 마우스 오른쪽 버튼으로 클릭하고 **새 열**을 선택합니다 .

   " 열 정보 " 창이 나타납니다 .
3. " 열 이름 " 옆에 "Profit by Sales" 를 입력합니다 .

   기존 %profit/sales 열에 있는 계산식을 다시 생성하려고 합니다 .
4. " 열 특성 " 목록에서 " 계산식 " 을 선택합니다 .

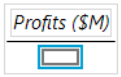
   계산식 편집기가 열립니다 . 편집 창의 상자 주위에 표시되는 파란색 외곽선은 해당 상자가 선택되어 있음을 나타냅니다 .
5. 상자가 선택된 상태로 열 목록에서 **Profits ($M)** 를 선택합니다 .

그림 7.3 Profits ($M) 열 추가



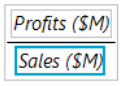
6. 상자가 선택된 상태로 나누기 버튼 ÷을 클릭합니다 .

그림 7.4 나누기 기호 추가



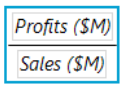
7. 표시된 것과 같이 상자가 선택된 상태로 열 목록에서 **Sales ($M)** 를 선택합니다 .

그림 7.5 Sales ($M) 열 추가



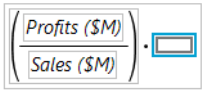
8. 계산식 편집 영역에서 바깥쪽 상자를 클릭하여 선택합니다 .

그림 7.6 계산식 선택



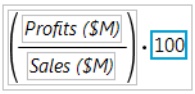
9. 곱하기 버튼 ×을 클릭합니다 .

그림 7.7 곱하기 기호 추가



10. 새 상자가 선택된 상태로 "100" 을 입력하고 Enter 키를 누릅니다 .

그림 7.8 계산식에 100 곱하기



11. **확인**을 클릭합니다 .

    열에 계산된 수가 채워집니다 .

12. 마지막 열을 마우스 오른쪽 버튼으로 클릭하고 " 열 정보 " 를 선택합니다 .
13. 열 정보 창의 " 형식 " 목록에서 **고정 소수점**을 선택하고 " 너비 " 옆에 "10" 을 입력한 후 " 소수 " 옆에 "2" 를 입력합니다 .

    열 너비와 소수 자릿수가 지정되었습니다 .

14. **확인**을 클릭합니다 .

    새 계산식 열은 %profit/sales 계산식 열과 매칭됩니다 .

팁 :

- Alt(macOS 의 경우 Option) 키를 누른 채로 열을 클릭하면 선택한 필드가 비어 있든 비어 있지 않든 상관없이 해당 열로 바뀝니다 .
- 열 목록에서 마우스 오른쪽 버튼을 클릭하면 나타나는 메뉴에는 두 가지 바꾸기 옵션이 있습니다 . **선택한 하위 표현식의 모든 항목 바꾸기**는 선택한 하위 표현식을 모두 현재 선택한 열로 바꿉니다. **선택한 하위 표현식을 열로 바꾸기**는 예를 들어 합 또는 더하기와 같이 여러 열을 지원하는 함수일 경우 선택한 표현식을 여러 열로 바꿉니다 .
- JSL 스크립트를 작성하여 고유한 함수를 정의하고 열을 변환할 수 있습니다 . 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

계산식에서 지역 변수 사용

표현식에서 임시 숫자 변수를 생성하여 사용할 수 있습니다 . 순서형 지역 변수를 사용하거나 특수한 유형의 지역 변수인 모수를 사용할 수 있습니다 . 지역 변수는 해당 지역 변수가 정의된 계산식을 실행하기 위한 용도로만 존재합니다 . 지역 변수는 계산식에서 굵은 기울임꼴 항으로 표시됩니다 .

지역 변수는 주로 할당 함수와 함께 사용되며, 할당 함수는 복잡한 방정식에서 사용되는 지역 변수에 표현식을 할당할 수 있습니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 장의 "할

당 함수"(723 페이지) 에서 확인하십시오 . 일부 경우에 이 방법으로 방정식의 작성을 단순화하고 실행 효율성을 높일 수 있습니다 .

지역 변수에 있는 값을 참조하는 계산식을 작성하려면 다음 단계를 수행하십시오 .

1 단계 : 지역 변수 생성

1. 데이터 테이블에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식 편집기를 엽니다 .
2. 계산식 편집 영역에서 상자를 클릭하여 선택합니다 . 선택 항목 주위에는 파란색 외곽선이 표시됩니다 .
3. 계산식 편집기의 가운데 페인에 있는 목록에서 **지역 변수**를 선택합니다 .
4. **새 지역 변수**를 클릭합니다 .
5. 지역 변수의 이름을 입력합니다 . 기본적으로 지역 변수는 *t0*, *t1* 등의 이름으로 지정되며 결측값을 갖습니다 .
6. 시작 값을 할당하고 **확인**을 클릭합니다 .
7. (선택 사항) 지역 변수를 복사 , 편집 또는 삭제하려면 지역 변수 이름을 마우스 오른쪽 버튼으로 클릭 (macOS 의 경우 Ctrl 키를 누른 채 클릭) 하고 **복사** , **편집** 또는 **삭제**를 선택합니다 .

2 단계 : 계산식에 지역 변수 삽입

1. 계산식 편집 영역에서 항을 클릭하여 선택합니다 . 선택 항목 주위에는 파란색 외곽선이 표시됩니다 .
2. " 지역 변수 " 목록에서 지역 변수 이름을 클릭합니다 . 계산식에 지역 변수가 굵은 기울임꼴 항으로 나타납니다 .

참고 :

- 지역 변수를 두 번 클릭하거나 Alt 키를 누른 채 클릭하면 선택된 계산식 부분이 대체됩니다 .
- 지역 변수를 생성하는 또 다른 방법은 계산식 편집기 키패드에 있는 " 임시 변수 만들기 " t= 버튼을 사용하는 것입니다 . 이 버튼을 사용하면 자동으로 지역 변수가 생성되어 표시되며 그 뒤에는 세미콜론이 추가됩니다 . 자세한 내용은 "JMP 계산식에 연산자 추가 "(369 페이지) 에서 확인하십시오 .
- 표현식에서 지역 변수를 생성할 수도 있습니다 . 계산식에서 표현식을 마우스 오른쪽 버튼으로 클릭하고 **표현식에서 지역 변수 생성**을 선택합니다 .
- 변수를 Alt 키와 Shift 키를 누른 채 클릭하면 선택한 변수를 새 창이 아니라 현재 위치에서 편집할 수 있습니다 .

계산식에서 지역 변수를 참조하는 예는 " 계산식에서 지역 변수 사용 "(388 페이지) 에서 확인하십시오 .

계산식에서 모수 포함

모수는 비선형 적합과 같은 일부 플랫폼에서 모형 모수로 인식되는 특수한 유형의 지역 변수입니다. 계산식에서 순서형 지역 변수를 사용할 때와 같은 방식으로 모수를 사용할 수 있습니다.

모수의 예를 보려면

1. JMP 와 함께 설치된 샘플 데이터 폴더의 Nonlinear Examples 폴더에서 US Population.jmp 를 엽니다 .
2. 열 이름 x-formula 를 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택합니다 .
3. 계산식 편집기의 가운데 페인에 있는 목록에서 **모수**를 선택합니다 .

참고 :

- 비선형 적합을 완료한 후 또는 비선형 제어판에서 " 재설정 " 버튼을 클릭한 후의 모수 값은 비선형 플랫폼에서 계산된 최근 값입니다 .
- 적합 알고리즘이 한 단계를 거칠 때마다 비선형 보고서에 업데이트된 모수 값이 표시됩니다.
- 모수가 있는 계산식을 열에 붙여넣을 경우 해당 열에 동일한 이름의 기존 모수가 있지 않은 한 해당 열에 대한 모수가 자동으로 생성됩니다 .
- 모수를 두 번 클릭하거나 Alt 키를 누른 채 클릭하면 선택된 계산식 부분이 대체됩니다 .
- 모수를 Alt 키와 Shift 키를 누른 채 클릭하면 선택한 모수를 새 창이 아니라 현재 위치에서 편집할 수 있습니다 .

모수에 있는 값을 참조하는 계산식을 작성하려면

1 단계 : 모수 생성

1. 데이터 테이블에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식 편집기를 엽니다 .
2. 계산식 편집 영역에 계산식이 선택되어 있는지 확인합니다. 선택 항목 주위에는 파란색 외곽선이 표시됩니다 .
3. 계산식 편집기의 가운데 페인에 있는 목록에서 **모수**를 선택합니다 .
4. **새 모수**를 클릭합니다 .
5. 모수의 이름을 입력합니다 . 기본적으로 모수는 *b0, b1* 등의 이름으로 지정되며 결측값을 갖습니다 .
6. 시작 값을 할당합니다 . 비선형 플랫폼의 모형에서 모수를 사용할 때는 이 값을 입력하는 것이 중요합니다 . 비선형 적합을 완료한 후의 모수 값은 비선형 플랫폼에서 계산된 최근 값입니다 .
7. (선택 사항) 여러 개의 모수 (예를 들어 범주형 변수의 각 수준마다 하나씩) 를 한 번에 할당하려면 **범주로 확장 , 열 선택**을 선택합니다 . 그런 다음 모수를 확장할 열을 선택합니다 .
8. **확인**을 클릭합니다 .

9. (선택 사항) 모수를 복사 , 편집 또는 삭제하려면 모수 이름을 마우스 오른쪽 버튼으로 클릭 (macOS 의 경우 Ctrl 키를 누른 채 클릭) 하고 **복사** , **편집** 또는 **삭제**를 선택합니다 .

2 단계 : 계산식에 모수 삽입

1. 계산식 편집 영역에서 항을 클릭하여 선택합니다. 선택 항목 주위에는 파란색 외곽선이 표시됩니다 .
2. "모수" 상자에서 모수 이름을 클릭합니다. 계산식에 모수가 굵은 글꼴 유형으로 나타납니다.

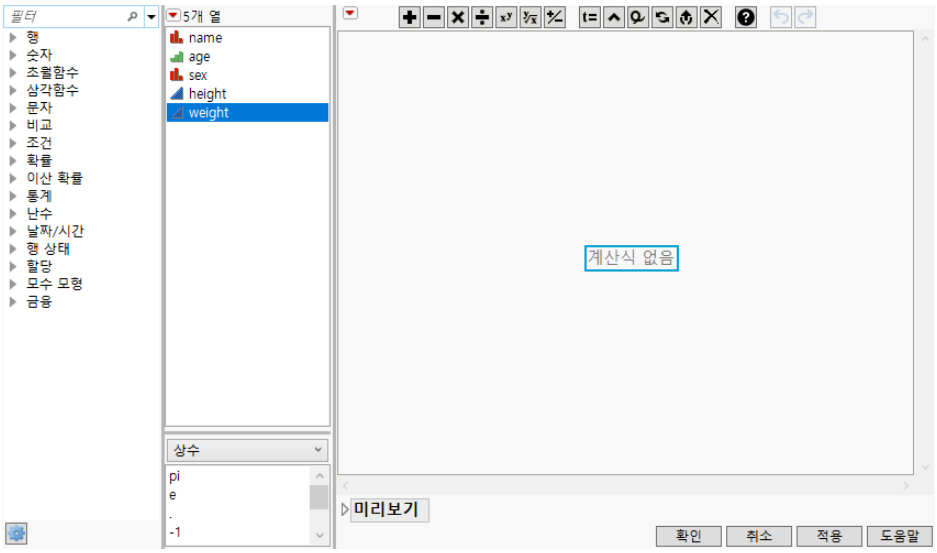
팁 : 모수를 Alt 키와 Shift 키를 누른 채 클릭하면 선택한 모수를 편집할 수 있습니다 .

계산식에서 상수 삽입

계산식은 단순히 숫자, 문자 또는 행 상태 상수를 대입하는 것일 수도 있고, 조건 절에 기반한 복잡한 계산을 포함할 수도 있습니다. 상수에는 e, pi, -1, 0, 1, 2와 같이 일반적으로 사용되는 숫자 항이 포함됩니다. 다음 두 가지 방법으로 계산식에 상수를 추가할 수 있습니다.

- 키보드를 사용하여 직접 입력
- 가운데 페인의 목록에서 선택 (그림 7.9 참조)

그림 7.9 상수가 선택되어 있는 계산식 요소 브라우저



계산식에 상수를 추가하려면 다음 단계를 수행하십시오 .

1. 데이터 테이블에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식 편집기를 엽니다 .

2. 계산식 편집 영역에서 상자를 클릭하여 선택합니다. 선택 항목 주위에는 파란색 외곽선이 표시됩니다 .
3. 숫자를 직접 입력하거나 , 가운데 페인에서 **상수**를 선택합니다 (그림 7.9 참조). 그런 다음 나타나는 목록에서 값(0, 1, 2, -1, pi, e)을 클릭합니다. 외곽선이 표시된 상자에 값이 나타납니다.
4. 키패드와 함수를 사용하여 계산식의 나머지를 완성합니다 . 자세한 내용은 " 계산식에서 데이터 테이블 값 참조 "(363 페이지), "JMP 계산식에 연산자 추가 "(369 페이지) 및 " 계산식에서 함수 사용 "(371 페이지) 에서 확인하십시오 .

참고 : 상수를 두 번 클릭하거나 Alt 키를 누른 채 클릭하면 선택된 계산식 부분이 대체됩니다 .

JMP 계산식에 연산자 추가

계산식 작성에 도움이 되는 버튼이 포함되어 있는 키패드를 사용하여 계산식에 연산자를 추가할 수 있습니다 . 키패드에는 일반적인 연산자 (함수라고도 함) 가 포함되어 있습니다 .

창 상단의 키패드 연산자를 사용하여 계산식을 작성하려면 다음 단계를 수행하십시오 .

1. 데이터 테이블에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식 편집기를 엽니다 .
2. 계산식 편집 영역에서 상자를 클릭하여 선택합니다. 선택 항목 주위에는 파란색 외곽선이 표시됩니다 . 연산자는 강조 표시된 영역에서 관련 작업을 수행합니다 .
3. 계산식에 사용할 열 또는 변수를 선택합니다 .
4. 키패드 버튼을 클릭합니다 .

계산식 편집기의 키패드 참조

표 7.1에서는 키패드 버튼에 대해 설명합니다.

표 7.1 계산식 편집기의 키패드 버튼

이름	설명
산술 버튼	일반적인 휴대용 계산기와 같이 작동하여 더하기, 곱하기, 빼기 및 나누기 연산자를 제공합니다.
지수	주어진 값을 지정된 지수만큼 거듭제곱합니다. 기본 지수는 2입니다.

표 7.1 계산식 편집기의 키패드 버튼 (계속)

이름	설명
제곱근	주어진 값에 대해 지정된 지수만큼의 제곱근을 계산합니다. 기본적으로 암시적 지수 2가 적용되며(제곱근) 이 지수는 표시되지 않습니다.
단항 부호 함수	인수의 부호를 바꿉니다. 변수 표현식에 단항 부호 함수를 적용하거나, 단항 부호 함수를 사용하여 음수 상수를 입력할 수 있습니다.
지역 변수(임시 변수 만들기)	지역 변수를 생성하여 표시하고 여기에 선택된 표현식의 값을 할당합니다. 표현식에서 지역 변수는 기본 이름 *t0* 으로 표시되며 그 뒤에 세미콜론이 추가됩니다 지역 변수 생성 및 삽입에 대한 자세한 내용은 "계산식에서 지역 변수 사용"(365 페이지)에서 확인하십시오. 예는 "계산식에서 지역 변수 사용"(388 페이지)에서 확인하십시오.
삽입	새 절 또는 함수 인수를 삽입합니다. 먼저 새 요소를 추가할 위치의 바로 앞에 있는 기존 절 또는 인수를 선택한 후 이 버튼을 클릭합니다. 그러면 새 절이 선택된 상태로 나타납니다. 쉼표를 입력하여 새 절 또는 인수를 삽입할 수도 있습니다. Alt 키를 누른 채 "삽입" 아이콘 을 클릭하면 선택된 필드 앞에 새 필드가 삽입됩니다.
삭제	요소의 값을 삭제하거나 절을 삭제합니다. 삭제 버튼은 키보드의 Delete 키와 동일하게 작동합니다.
항 전환	선택된 표현식의 가운데에 있는 연산자를 확인하여 해당 연산자의 양쪽에 있는 표현식을 서로 바꿉니다.
표현식 삭제(벗기기)	첫 번째 인수를 포함하는 가장 바깥쪽 표현식을 제거합니다. 이 과정을 반복하여 계산식 항을 하나씩 삭제할 수 있습니다. 예는 "계산식에서 표현식 삭제 버튼 사용"(390 페이지)에서 확인하십시오.
지우기	전체 계산식 지우기
도움말	계산식의 선택한 부분에서 JSL 함수의 항목에 대한 스크립트 인덱스를 엽니다. 이 버튼은 수식의 시각적 영역에서 선택한 상자가 알려진 함수인 경우에만 적용할 수 있습니다.
실행 취소	마지막으로 변경한 사항을 되돌립니다.

표 7.1 계산식 편집기의 키패드 버튼 (계속)

이름	설명
다시 실행	마지막으로 되돌린 변경 사항을 다시 실행합니다.

계산식에서 함수 사용

팁 : 함수를 마우스로 가리키면 함수에 대한 자세한 정보를 볼 수 있습니다 .

계산식에 여러 유형의 함수를 추가할 수 있습니다. 모든 함수는 함수 목록에 정리되어 있습니다. 브라우저에서는 함수 집합을 주제별로 구성된 목록에 그룹화합니다. 목록에는 계산식에 일반적으로 사용되는 함수가 포함되어 있습니다. 함수 목록을 사용하여 계산식의 요소에 대해 수행할 계산 유형을 지정할 수 있습니다.

함수가 포함된 계산식을 생성하려면 다음 단계를 수행하십시오 .

1. 데이터 테이블에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식 편집기를 엽니다 .
2. 계산식 편집 영역에서 표현식을 클릭하여 선택합니다 . 선택 항목 주위에는 파란색 외곽선이 표시됩니다 . 함수는 강조 표시된 영역에서 관련 작업을 수행합니다 .
3. 함수 목록에서 표시할 함수 그룹을 선택합니다 . 자세한 내용은 표 7.2 에서 확인하십시오 .

   그러면 해당 그룹에 속한 함수가 메뉴 아래의 목록에 표시됩니다 . 함수 그룹에 대해서는 뒤에 나오는 목록에서 간략히 설명합니다 .
4. 함수 목록에서 함수를 클릭하여 선택한 항목에 적용합니다. 일부 항목의 경우 항목을 클릭하면 추가 선택을 위한 하위 메뉴가 표시됩니다 .
5. 계속해서 항을 강조 표시하고 계산식 요소 브라우저 , 키패드 또는 함수 목록에서 항목을 클릭하여 계산식을 작성합니다 .

팁 :

- Alt 키를 누른 채 함수를 선택하면 계산식의 선택된 함수가 함수 목록에서 선택한 함수 또는 키패드에서 선택한 버튼에 해당하는 함수로 바뀝니다 .
- Alt 키 (macOS 의 경우 Option 키) 를 누른 채 열 또는 다른 요소를 클릭하면 해당 항목이 비어 있지 않더라도 항목이 바뀝니다 .
- 대부분의 함수는 계산식 편집 영역의 상자에 삽입된 회색 텍스트를 통해 적절한 인수에 대한 힌트를 제공합니다 . 함수에 다른 인수를 추가할 수 있는 경우 인수 영역에 작은 캐럿이 표시되기도 합니다 .

- 함수가 수행하는 작업을 확인하려면 함수 목록 또는 계산식에서 함수를 마우스 오른쪽 버튼으로 클릭하고 **스크립트 인덱스에서 보기**를 선택하십시오 .
- JSL 스크립트를 작성하여 고유한 함수를 정의할 수 있습니다 . 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

표 7.2 함수 그룹

그룹	포함된 함수
Row	Lag, Dif, Subscript, Row, NRow 등의 기타 함수가 포함된 함수 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "행 함수"(677 페이지)에서 확인하십시오.
숫자	계산식 항으로 자주 사용되는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "숫자 함수"(679 페이지)에서 확인하십시오.
초월함수	자연 로그, 상용 로그, 지수, 제곱근, 계승, 조합, 베타, 감마 등이 포함된 함수 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "초월함수"(680 페이지)에서 확인하십시오.
삼각함수	사인, 코사인, 탄젠트, 역함수, 쌍곡선 함수 같은 표준 삼각함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "삼각함수"(683 페이지)에서 확인하십시오.
문자	자르기, 문자열 길이 구하기, 숫자와 문자 간의 변환 같이 문자 인수에 적용되는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "문자 함수"(684 페이지)에서 확인하십시오.
비교	보다 작음, 보다 작거나 같음, 같지 않음 등의 표준 논리적 비교를 수행하는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "비교 함수"(689 페이지)에서 확인하십시오.
조건부	If, Match, Choose 같은 프로그래밍 방식 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "조건 함수"(690 페이지)에서 확인하십시오.
확률	정규 분포, 스튜던트 *t* 분포, 카이제곱 분포, *F* 분포 같이 표준 통계 분포의 확률 및 분위수를 계산하는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "확률 함수"(695 페이지)에서 확인하십시오.
이산 확률	Poisson, 감마 Poisson, 초기하 같이 이산 확률을 계산하는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "이산 확률 함수"(704 페이지)에서 확인하십시오.

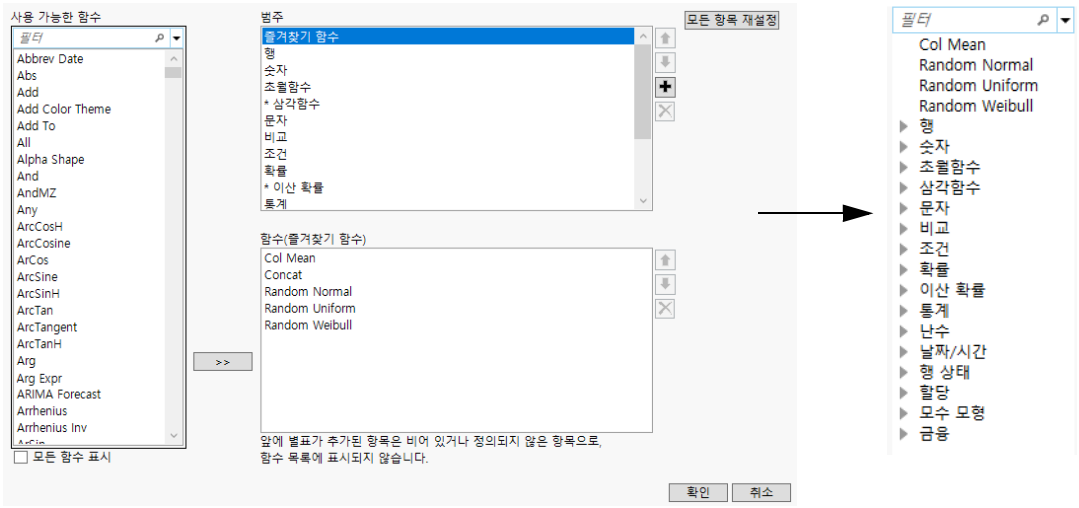
표 7.2 함수 그룹 (계속)

그룹	포함된 함수
통계	평균, 표준편차 같은 표준 통계량을 계산하는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "통계 함수"(706 페이지)에서 확인하십시오.
난수	균등 분포, 정규 분포, Cauch 분포 등의 사전 정의된 분포를 기반으로 난수를 생성하는 함수의 목록을 표시합니다. 테이블 행 순서를 랜덤화하는 함수도 있습니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "난수 함수"(710 페이지)에서 확인하십시오.
날짜 / 시간	1904 년 1 월 1 일 이후의 경과한 초 수로 해석되는 **날짜** 데이터 유형의 인수를 필요로 하는 함수의 목록을 표시합니다. 날짜/시간 함수는 날짜, 주, 월 등의 값을 반환합니다. 날짜를 계산하고 데이터 간격을 구할 수도 있습니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "날짜/시간 함수"(715 페이지)에서 확인하십시오.
행 상태	행 상태의 색상, 표식, 라벨, 숨김, 제외 또는 선택 상태를 할당하거나 감지하는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "행 상태 함수"(718 페이지)에서 확인하십시오.
할당	할당 연산자의 오른쪽 값을 연산자의 왼쪽 변수에 대입하는 함수의 목록을 표시합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "할당 함수"(723 페이지)에서 확인하십시오.
모수 모형	선택한 열에 대해 선형, 교호작용 또는 완전 2 차 모형을 구성할 수 있습니다. 모형 유형을 선택하면 모형을 구성하는 데 사용할 변수를 선택할 수 있는 창이 나타납니다. Ctrl 키를 누른 채 두 개 이상을 클릭하여 선택합니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "모수 모형 함수"(724 페이지)에서 확인하십시오.
금융	지급 원금, 이자율, 수익률 등을 계산하기 위한 계산식을 생성할 수 있습니다. 자세한 내용은 "계산식에서 JMP 함수에 대한 참조" 부록의 "금융 함수"(724 페이지)에서 확인하십시오.

계산식 편집기에서 자주 사용하는 함수 찾기

주로 사용하는 함수를 쉽게 찾을 수 있도록 함수 목록을 사용자 정의할 수 있습니다. 지정한 함수는 목록의 맨 위에 나타납니다. 자주 사용하는 함수를 목록의 맨 위에 표시하고, 고유한 함수 범주를 생성하고, 함수 또는 범주를 삭제하고, 함수를 목록의 위 또는 아래로 이동할 수 있습니다.

그림 7.10 함수 목록 사용자 정의



즐겨찾기 함수를 빠르게 책갈피에 추가하려면

1. 함수 목록에서 함수를 마우스 오른쪽 버튼으로 클릭합니다 .
2. **즐겨찾기로 추가**를 선택합니다 .

즐겨찾기를 책갈피에 추가하고 위치를 지정하려면

1. 계산식 편집기 영역의 위에 있는 빨간색 삼각형 메뉴에서 **함수 목록 사용자 정의**를 선택합니다.
2. " 사용 가능한 함수 " 목록에서 함수를 선택합니다 .
3. 오른쪽 화살표 버튼을 클릭하여 해당 함수를 " 함수 " 목록에 추가합니다 .
4. 위로 또는 아래로 버튼을 클릭하여 즐겨찾기를 재배열합니다 .

즐겨찾기를 재배열 또는 삭제하려면

1. 계산식 편집기 영역의 위에 있는 빨간색 삼각형 메뉴에서 **함수 목록 사용자 정의**를 선택합니다.
2. " 범주 " 목록에서 **즐겨찾기 함수**를 선택합니다 .
3. 재배열하려는 즐겨찾기 함수를 선택하고 위로 또는 아래로 버튼을 클릭합니다 .
4. 삭제하려는 즐겨찾기 함수를 선택하고 X 버튼을 클릭합니다 .

범주에 함수를 추가하려면

함수가 계산식에 포함되어 있지만 함수 목록에는 포함되지 않은 경우 함수를 범주에 추가하는 방법으로 목록에 추가할 수 있습니다.

1. 함수 목록에서 범주에 추가하려는 함수를 선택합니다 .
2. 함수를 마우스 오른쪽 버튼으로 클릭하고 **범주에 추가**를 선택합니다 .
3. 범주를 선택합니다 .
4. **확인**을 클릭합니다 .

범주를 재배열 , 추가 또는 삭제하려면

1. 계산식 편집기 영역의 위에 있는 빨간색 삼각형 메뉴에서 **함수 목록 사용자 정의**를 선택합니다.
2. " 범주 " 목록에서 범주를 선택하고 범주를 재배열 , 추가 또는 삭제합니다 .

함수를 재배열 또는 삭제하려면

1. 계산식 편집기 영역의 위에 있는 빨간색 삼각형 메뉴에서 **함수 목록 사용자 정의**를 선택합니다.
2. " 범주 " 목록에서 재배열 또는 삭제할 함수가 포함된 범주를 선택합니다 .
3. 위로 , 아래로 또는 X 버튼을 클릭하여 선택한 함수를 재배열하거나 삭제합니다 .

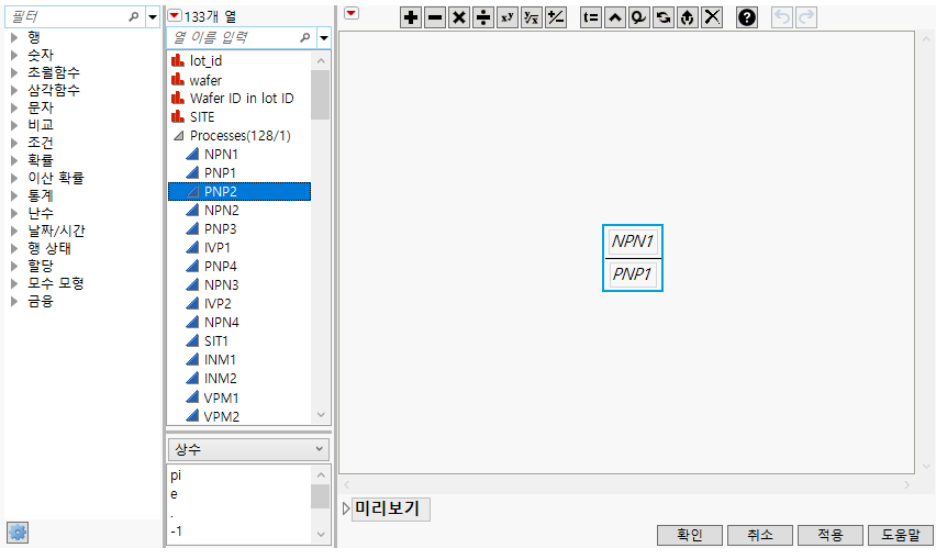
참고 : 함수를 삭제할 경우 해당 함수는 " 함수 " 목록에서만 삭제됩니다 . 계산식에서 함수를 직접 입력하는 것은 여전히 가능합니다 .

계산식 복사 및 반복

"열" 메뉴의 "속성 표준화" 옵션을 사용하면 계산식에서 각 열 참조를 증분하면서 열 계산식을 복사할 수 있습니다. 이후의 각 열에서 열이 1 씩 증분하는 새 열 계산식을 수동으로 생성하는 방법 대신 이 옵션을 사용할 수 있습니다. 예를 들어 열 B 에 계산식 Ax2 가 포함된 경우 표준화된 열 C 에는 계산식 Bx2 가 포함되고 표준화된 열 D 에는 계산식 Cx2 가 포함됩니다.

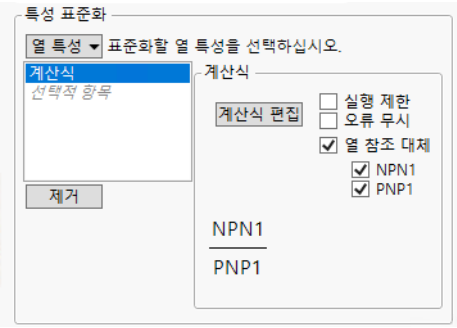
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Semiconductor Capability.jmp 를 엽니다 .
2. **열 > 새 열**을 선택합니다 .
3. **추가할 열의 수** 옆에 3 을 입력하고 **확인**을 클릭합니다 .
4. 첫 번째 새 열을 선택하고 **열**을 선택한 후 **계산식**을 선택합니다 .
5. 계산식 NPN1/PNP1 을 생성하고 **확인**을 클릭합니다 .

그림 7.11 계산식 NPN1/PNP1 생성



6. 세 개의 새 열을 선택한 후 **열 > 속성 표준화**를 선택합니다 .
7. " 특성 표준화 " 섹션의 목록에서 **계산식**을 선택합니다 .
8. **열 참조 대체**를 선택하고 NPN1 및 PNP1 을 선택된 상태로 둡니다 .

그림 7.12 특성 표준화 옵션



참고 : 이 예에서는 NPN1 과 PNP1 을 모두 변경합니다 . 이후 계산식에서 변경하지 않을 열은 선택 취소해도 됩니다 .

9. **확인**을 클릭합니다 .

   첫 번째 새 열의 계산식을 기준으로 조정된 계산식이 두 번째 열과 세 번째 열에 복사됩니다 . 두 번째 및 세 번째 선택 열의 계산식은 한 열씩 증분합니다 .

   – 첫 번째 열의 원래 계산식 : NPN1/PNP1

– 두 번째 열의 조정된 계산식 : PNP1/PNP2
데이터 테이블에서 PNP2 는 PNP1 뒤에 있습니다 .
– 세 번째 열의 조정된 계산식 : PNP2/NPN2
데이터 테이블에서 NPN2 는 PNP2 뒤에 있습니다 .

그림 7.13 결과 데이터 테이블

NPN1	PNP1	PNP2	NPN2	X	Y	Z
114.55576	322.616752	469.39029	115.958473	0.3550831117	0.687310238	4.0479171331
120.043693	333.12808	437.781113	120.974137	0.3603529704	0.7609466694	3.618799212
114.92651	348.978757	532.128147	117.753124	0.3293223673	0.6558171354	4.5190151087
111.756446	268.548094	373.058629	114.092608	0.4161505811	0.7198549341	3.2697879024
111.545085	295.073166	338.900737	113.778098	0.3780251749	0.8706772615	2.9786113754
113.523596	323.833299	469.992171	116.488049	0.3505618367	0.6890184949	4.0346814413

참고 : 그림 7.13 에서 Semiconductor Capability.jmp 의 일부 열은 새 열 X, Y 및 Z 를 표시하기 위해 삭제되었습니다 .

또 다른 예는 "JMP 열 특성 설정" 장의 "예: 계산식 표준화"(306 페이지)에서 확인하십시오.

계산식의 표현식 정렬

계산식을 작성할 때는 모든 함수가 다음 표(수준 1 이 가장 높은 우선 순위)에 표시된 것과 같은 우선 순위를 갖고 있다는 점을 염두에 두어야 합니다. 우선 순위가 높은 표현식은 낮은 수준의 표현식보다 먼저 계산됩니다. 표현식에 우선 순위가 같은 연산자가 여러 개 있으면 왼쪽 연산자부터 오른쪽 연산자의 순서로 계산됩니다. 괄호 안의 표현식은 항상 가장 먼저 계산되므로 필요한 경우 괄호를 사용하여 다른 우선 순위 규칙보다 우선하도록 할 수 있습니다. 항은 더 이상 계산할 수 없으므로 우선 순위가 없습니다. 표 7.3에서는 처음 6 개의 우선 순위 수준을 보여 줍니다.

표 7.3 계산식의 연산자 우선 순위

수준 1	괄호
수준 2	함수 목록의 함수, And, Or, Not
수준 3	*, ÷, 모듈로
수준 4	+, -
수준 5	비교: <, ≤, =, ≠, ≥, >, ≤ x <, < x ≤, < x <
수준 6	논리 연산자 &, |, !

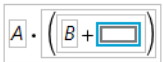
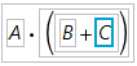
참고 : 함수 인수로 표현식이 포함된 경우 해당 인수는 함수 밖에서 괄호로 묶여 있는 경우보다 우선 순위가 더 높습니다 .

우선 순위대로 계산식 작성

계산식을 작성할 때는 인수 역할을 하는 표현식부터 시작하는 것이 가장 좋습니다. 함수는 우선 순위가 높고 항상 해당하는 인수와 함께 그룹화되기 때문입니다. 또한 가능하면 우선 순위가 높은 것부터 낮은 것 순서대로 계산되도록 표현식을 생성하는 것이 좋습니다. 괄호가 필요한 경우에는 괄호에 넣을 표현식을 입력하기 전에 여는 괄호를 입력해야 합니다.

예를 들어 데이터 테이블에 열 A, B 및 C가 있는 경우 표현식 $A(B + C)$를 작성하려면 다음 단계를 따라야 합니다. 이 표현식은 $(A \times B) + C$로 계산되는 $A \times B + C$와 동일하지 않습니다.

이 표현식을 입력하려면 다음 단계를 수행하십시오 .

1. 열 목록에서 열 A 를 클릭합니다 .
2. 계산식 편집기의 키패드에서 곱하기 버튼 을 클릭합니다 .
3. 여는 괄호를 입력합니다 .
4. 열 목록에서 열 B 를 클릭합니다 .
5. 계산식 편집기의 키패드에서 더하기 버튼 을 클릭합니다 . 
6. 열 목록에서 열 C 를 클릭합니다 . 

우선 순위에 따라 각 함수의 영향을 받는 인수가 결정되므로 표현식 그룹화도 우선 순위의 영향을 받습니다. 우선 순위 규칙이 적용된 방식을 확인하려면 계산식에서 함수를 선택하십시오.

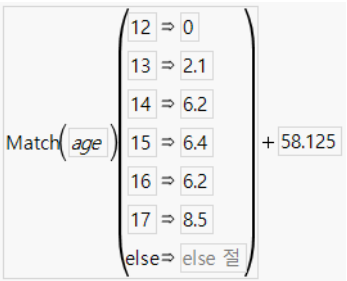
효율적인 계산을 위한 계산식 정형화

일반적으로는 효율적인 계산을 위해 계산식을 정형화할 필요가 없습니다. 대부분의 계산식은 구조에 상관없이 거의 즉각적으로 계산됩니다. 통계 함수 및 상수 표현식은 열 값이 계산될 때 한 번만 계산되기 때문입니다.

하지만 조건부 표현식을 생성할 경우에는 Match가 그와 동등한 조건 함수인 If보다 계산이 더 빠르고 메모리 사용량이 적다는 점을 염두에 두어야 합니다. Match는 후행 공백을 무시하지만 If는 그렇지 않기 때문입니다.

예를 들어 Big Class.jmp를 사용하는 경우 그림 7.14에 표시된 것과 같이 아동의 연령으로부터 키를 예측할 수 있습니다. 여기에서는 age 변수의 값에 따라 달라지는 수에 기본 키 58.125 인치를 더합니다.

그림 7.14 Match 조건 함수가 IF 함수보다 더 빨리 계산됨



Match 조건 함수가 If 함수보다 빨리 계산됩니다. 이는 데이터 테이블의 각 행에 대해 age 변수가 한 번만 계산되기 때문입니다. If 조건은 If 절의 결과가 true가 될 때까지 각 행에 대한 각각의 If 절에서 age 변수를 계산해야 합니다.

JMP 계산식 편집기 옵션

계산식 편집기는 계산식 생성 및 관리를 위한 강력한 옵션을 제공합니다.

- "JMP 함수의 도함수 계산 "
- " 복잡한 계산식 단순화 "
- " 계산식 실행 "
- " 계산식에서 오류 무시 "
- "JSL 에서 계산식 보기 "

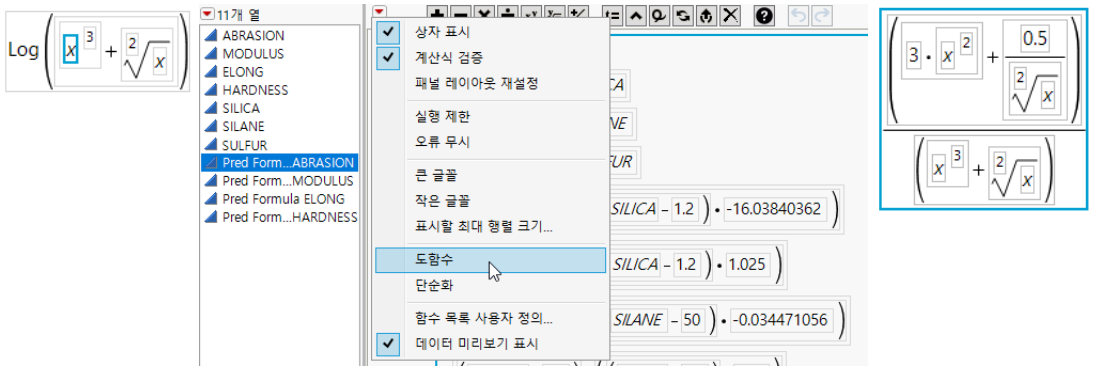
JMP 함수의 도함수 계산

JMP 계산식 편집기는 함수의 도함수를 구해 표시할 수 있습니다. 사용자가 강조 표시한 함수 인수(단일 변수 이름)에 관해 도함수를 구할 수 있습니다. 따라서 x에 관해 미분하려면 x가 표현식의 인수 중 하나여야 합니다. 키패드 옆의 빨간색 삼각형 메뉴에 "도함수" 명령이 포함되어 있습니다.

도함수를 계산하려면 다음 단계를 수행하십시오 .

1. 함수를 입력합니다 .
2. 변수를 강조 표시합니다 .
3. 메뉴에서 **도함수**를 선택합니다 .

그림 7.15 도함수 옵션 사용



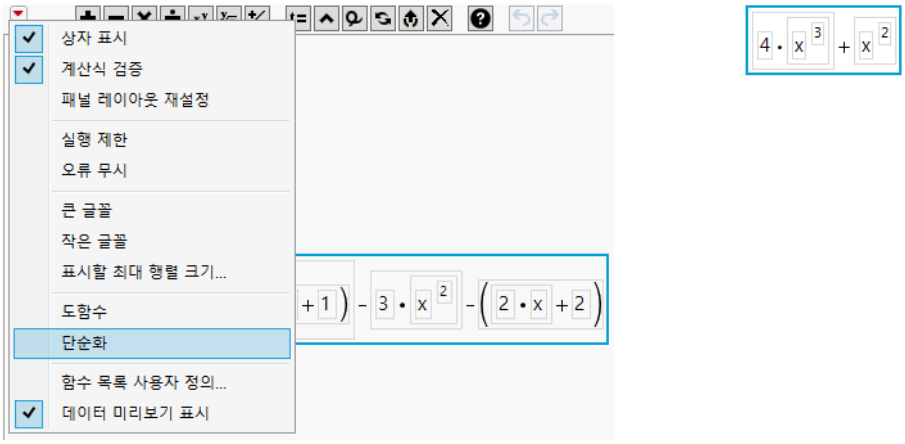
복잡한 계산식 단순화

계산식 편집기에 복잡한 계산식이 포함된 경우 JMP에서는 다양한 대수 규칙을 사용하여 계산식을 단순화할 수 있습니다. 상수 표현식을 찾고, 분배 법칙을 적용하고, 항을 결합하는 등의 작업을 수행할 수 있습니다.

복잡한 계산식을 단순화하려면 다음 단계를 수행하십시오 .

1. 계산식 일부 또는 모두를 선택합니다 .
2. 키패드 옆의 빨간색 삼각형을 클릭합니다 .
3. 그림 7.16 에 표시된 것과 같이 목록에서 **단순화**를 선택합니다 .

그림 7.16 단순화 선택 (왼쪽) 후 단순화된 계산식 (오른쪽)



계산식 실행

기본적으로 JMP 는 사용자가 생성한 각 계산식을 실행합니다. 이 계산식 실행을 해제하거나, 계산식 생성을 마치기 전에 계산식을 실행해 볼 수 있습니다.

실행 제한

복잡한 계산식을 작성하는 경우 개발 단계에서 계산식 실행을 해제하면 유용합니다. 실행 기능을 해제하고 계산식 섹션을 작성한 후 테스트 목적으로만 실행할 수 있습니다. 특히, 계산식 편집기를 닫았다가 나중에 다시 열더라도 계산식이 실행되지 않으므로 계산식 작성을 계속할 수 있습니다.

계산식 실행을 제한하려면 키패드 옆의 빨간색 삼각형을 클릭하고 **실행 제한**을 선택합니다. 실행이 제한되면 계산식 아이콘 이 흐리게 표시됩니다.

다음 사항에 유의하십시오.

- 계산식 편집기에서 빨간색 삼각형의 오른쪽에 아이콘이 나타나면 해당 계산식에 대해 계산식 실행이 제한되어 있는 것입니다.
- 열 패널에서 열 이름 옆에 아이콘이 나타나면 해당 열의 값은 계산식에서 얻은 결과입니다. 계산식 실행이 제한되면 이 계산식 아이콘 이 흐리게 표시됩니다. 자세한 내용은 "JMP 시작하기 " 장의 " 열 특성을 나타내는 아이콘 "(41 페이지) 에서 확인하십시오.
- 모든 계산식에 대해 계산식 실행을 해제하려면 데이터 테이블 이름 옆의 빨간색 삼각형을 클릭하고 **계산식 실행 제한**을 선택하십시오. 계산식을 다시 실행하려면 데이터 테이블 이름 옆의 빨간색 삼각형을 클릭하고 **계산식 다시 실행**을 선택하십시오. 보류 중인 계산식이 포함된 각 열에 대해 계산식 편집기에서 **적용** 버튼을 클릭할 수도 있습니다.
- 계산식 실행을 해제한 후 계산식을 편집하고 **적용**을 클릭하면 계산식을 실행할지 묻는 메시지가 표시됩니다. 계산식 실행을 한 번 허용하려면 **이번에 계산식 실행**을 클릭합니다.

계산식에서 오류 무시

계산식을 생성하고 **확인**을 클릭하면 JMP 가 자동으로 계산식에 오류가 있는지 확인하고 오류가 있으면 사용자에게 알립니다. 오류가 있을 때마다 오류 메시지가 나타나 추가 오류를 무시할지 여부를 사용자에게 묻습니다.

일부 경우 계산식을 개발 중일 때는 오류 메시지를 표시하지 않도록 할 수 있습니다. 예를 들어 오류를 일으키는 각 행에 대한 오류 메시지를 보지 않고 일부 행에 대해 계산된 값만 얻을 수 있습니다.

오류를 무시하려면 다음 단계를 수행하십시오.

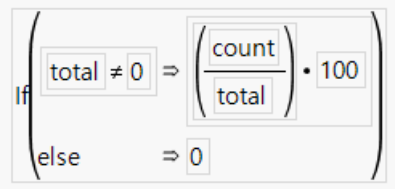
1. 계산식을 생성합니다.
2. 키패드 옆의 빨간색 삼각형을 클릭하고 **오류 무시**를 선택합니다.

JSL 에서 계산식 보기

계산식의 일부를 **텍스트 모드**에서 입력하거나 변경할 수 있습니다. 텍스트 모드에서는 계산식이 JSL(JMP 스크립트 언어)로 표시됩니다. 계산식을 두 번 클릭하면 전체 계산식(또는 일부 항)이 스크립트 편집기에 나타납니다. 긴 계산식은 최대화된 스크립트 편집기 창에서 열립니다. 또는, **편집기 최대화** 버튼 을 클릭하여 창을 최대화할 수도 있습니다.

계산식의 모든 요소를 스크립트 편집기에 표시한 후 편집할 수 있습니다. 계산식 스크립팅 명령을 편집한 후에는 계산식 외부를 클릭(또는 최대화된 창 아래의 **확인**을 클릭)하여 형식이 지정된 형태를 확인하고 변경 사항을 저장합니다. If 문을 텍스트 형식으로 입력하려면 각 If/Then 절의 인수 쌍을 문에 추가하고, 필요한 경우 else 절에 대한 마지막 인수 하나를 추가합니다. 그림 7.17 의 If 문은 텍스트 형식으로는 `If(:total!=0, (:count/:total)*100, 0)` 으로 나타납니다.

그림 7.17 계산식 모드의 If 문



계산식 편집

키패드를 사용하여 계산식을 편집합니다. 화살표를 사용하여 표현식을 선택하고, 잘라내고, 복사하고, 붙여넣고, 함수를 삭제하고, 표현식을 드래그할 수 있습니다.

- " 계산식에서 오류 수정 "
- " 계산식에서 표현식 선택 "
- " 함수 계산식 삭제 "
- " 계산식 잘라내기 , 복사 및 붙여넣기 "
- " 계산식 클릭 및 드래그 "

계산식에서 오류 수정

JMP 계산식 편집기에서 계산식을 입력하는 동안 실수를 한 경우 Ctrl+Z를 누릅니다. 그러면 실행 취소 가능한 마지막 명령의 효과를 되돌립니다.

계산식을 수정하는 데 유용한 다른 명령은 다음과 같습니다 .

- 선택한 표현식을 제거하려면 계산식 편집기 키패드의 삭제 버튼 ( ) 을 클릭합니다 .
- 잘라내기 , 복사 또는 붙여넣기 바로 가기 키 명령을 사용하거나 , 계산식의 강조 표시된 부분을 마우스 오른쪽 버튼으로 클릭하고 메뉴에서 **잘라내기** , **복사** 또는 **붙여넣기**를 선택합니다 .

- 계산식의 일부분을 드래그하여 재배열합니다 .

계산식에서 표현식 선택

키보드의 화살표 키를 사용하여 JMP 계산식 편집기에서 편집할 표현식을 선택할 수 있습니다. 괄호 또는 둘러싸기 옵션이 없는 경우 화살표 키를 사용하여 계산식의 우선 순위를 확인할 수도 없습니다. 자세한 내용은 "계산식에서 개요 숨기기 및 표시"(385 페이지)에서 확인하십시오.

표현식의 연산자 (+, –, *, ÷)를 클릭하면 해당 연산자 및 피연산자가 선택됩니다. 해당 항목 주위에는 파란색 상자가 표시됩니다. 연산자를 선택한 후에는 다음과 같은 작업이 가능합니다.

- 왼쪽 및 오른쪽 화살표 키는 표현식 내에서 같은 우선 순위를 갖는 다른 연관 연산자 사이에서 선택을 전환합니다 .
- 위쪽 화살표는 선택 항목에 더 높은 우선 순위의 피연산자 및 연산자를 추가하여 현재 선택 범위를 확장합니다 .
- 아래쪽 화살표는 선택 항목에서 피연산자 및 연산자를 제거하여 현재 선택 범위를 축소합니다.

함수 계산식 삭제

함수를 삭제하면 함수 인수도 삭제됩니다. 일부 경우에는 함수에서 필수 인수나 결측 항을 삭제하면 함수도 삭제됩니다. JMP 계산식 편집기에서 함수를 벗겨 인수에서 삭제할 수 있습니다.

단일 인수에서 함수를 벗기려면 다음 단계를 수행하십시오 .

1. 함수를 선택합니다 .
2. 계산식 편집기의 키패드에서 벗기기 버튼 을 클릭합니다 . 또는 손 도구를 사용하여 인수를 함수의 위로 드래그합니다 .

그림 7.18 인수 벗기기

3. 계산식 변경을 완료합니다 .
4. **적용**을 클릭하면 해당 데이터 테이블 열에 새 값이 자동으로 채워집니다 .
5. 계산식을 생성한 후에는 계산식에서 참조하는 열의 값을 변경할 수 있습니다. 그러면 계산식의 열에서 영향을 받는 모든 값을 JMP 가 자동으로 다시 계산합니다 .

계산식 잘라내기 , 복사 및 붙여넣기

표현식 또는 전체 계산식을 잘라내거나 복사하여 JMP 계산식 편집기의 다른 계산식에 붙여넣을 수 있습니다. 잘라내기, 복사 또는 붙여넣기 바로 가기 명령을 사용하거나, 계산식의 강조 표시

된 부분을 마우스 오른쪽 버튼으로 클릭하고 메뉴에서 **잘라내기**, **복사** 또는 **붙여넣기**를 선택합니다. 계산식을 잘라내거나 복사하거나 붙여넣을 때는 다음 사항이 적용됩니다.

- 계산식을 다른 계산식 창에 붙여넣을 때는 계산식이 형식이 지정된 형태로 나타납니다 .
- 계산식은 클립보드에 JSL 문으로 저장됩니다 . 따라서 계산식을 다른 응용 프로그램에 복사할 경우에는 계산식이 JSL(JMP 스크립트 언어) 문으로 나타납니다 .

참고: Shift 키를 누른 채 빨간색 삼각형을 클릭하면 **SAS 계산식으로 복사**라는 명령이 나타납니다.

계산식 클릭 및 드래그

JMP 계산식 편집기에서 선택 가능한 모든 계산식 요소를 선택 가능한 다른 위치로 드래그할 수 있습니다.

계산식 요소를 클릭하여 드래그하려면 다음 단계를 수행하십시오 .

1. 표현식 안에 화살표 커서를 놓습니다 .
2. 표현식을 클릭합니다 .
3. 계산식에서 마우스를 드래그합니다 . 대상 표현식에 점선으로 된 외곽선이 표시됩니다 .

그림 7.19 표현식 드래그

4. 선택한 표현식을 원하는 새 위치로 드래그합니다. 선택한 표현식이 새 위치로 이동되어 기존 표현식을 대체합니다 .

JMP 계산식 편집기에서 계산식 사용자 정의

JMP 계산식 편집기에서 몇 가지 방법으로 계산식을 사용자 정의할 수 있습니다.

- " 계산식에서 개요 숨기기 및 표시 "
- " 계산식에서 글꼴 크기 변경 "
- " 계산식에서 행렬 크기 지정 "
- "JMP 계산식의 방향 변경 "
- " 계산식에서 인수 열기 및 닫기 "

계산식에서 개요 숨기기 및 표시

기본적으로 JMP는 계산식 내의 특정 항을 외곽선으로 둘러쌉니다. 이 기능을 둘러싸기라고 합니다. 둘러싸기는 계산식의 특정 부분을 선택하여 수정하려는 경우나 계산 순서를 확인해야 하는 경우에 유용합니다.

둘러싸기를 설정하거나 해제하려면 다음 단계를 수행하십시오 .

1. 계산식을 작성합니다 .
2. 키패드 옆의 빨간색 삼각형을 클릭하고 **상자 표시**를 선택합니다 . 메뉴에서 **상자 표시** 옆에 선택 표시가 있으면 계산식에 외곽선이 표시됩니다. 그렇지 않으면 외곽선이 표시되지 않습니다.

계산식에서 글꼴 크기 변경

형식이 지정된 계산식을 표시하는 데 사용되는 글꼴을 한 수준씩 늘리거나 줄이려면 다음 단계를 수행하십시오 .

1. 키패드 옆의 빨간색 삼각형을 클릭합니다 .
2. **큰 글꼴** 또는 **작은 글꼴**을 선택합니다 .
3. 글꼴 크기를 더 늘리거나 줄이려면 이 과정을 반복합니다 .

계산식에서 행렬 크기 지정

빨간색 삼각형 메뉴에서 **표시할 최대 행렬 크기**를 선택하여 행렬의 최대 너비 및 높이를 설정할 수 있습니다. 너비 또는 높이가 지정한 크기보다 크면 데이터 대신 행렬 차원이 표시됩니다. 이 경우 텍스트 편집기 창에서 계산식을 편집하면 데이터를 볼 수 있습니다.

JMP 계산식의 방향 변경

기본적으로 JMP는 계산식의 크기를 확인한 후 계산식 편집기에서 가장 적절한 방향(수평 또는 수직)으로 계산식을 표시합니다. 하지만 긴 계산식을 생성하는 경우에는 계산식을 다른 레이아웃으로 표시할 수도 있습니다.

계산식의 방향을 변경하려면 다음 단계를 수행하십시오 .

1. 계산식을 작성합니다 .
2. 인수 또는 계산식을 강조 표시합니다 .
3. 강조 표시한 항목을 마우스 오른쪽 버튼으로 클릭합니다 .
4. **방향**을 선택합니다 .
5. 방향 옵션 (최선 , 수평 또는 수직) 을 선택합니다 .

계산식에서 인수 열기 및 닫기

계산식이 너무 커서 화면에 모두 표시되지 않는 경우 계산식과 인수를 닫을 수 있습니다.

인수를 닫으려면 다음 단계를 수행하십시오 .

1. 계산식을 작성합니다 .
2. 인수 또는 계산식을 강조 표시합니다 .
3. 강조 표시한 항목을 마우스 오른쪽 버튼으로 클릭합니다 .
4. 나타나는 메뉴에서 **닫기** 또는 **인수 닫기**를 선택합니다 .

JMP 계산식 편집기의 계산식 작성 예

계산식 작성 방법을 보다 잘 숙지하려면 다음의 예 및 자습서를 검토하십시오 .

- "JMP 계산식 편집기에서 기본 기능 사용 "
- " 계산식에서 지역 변수 사용 "
- " 계산식에서 Munger 함수 사용 "
- " 계산식에서 Match 조건 함수 사용 "
- " 계산식에서 표현식 삭제 버튼 사용 "

JMP 계산식 편집기에서 기본 기능 사용

다음 예에서는 Big Class.jmp 샘플 데이터 테이블을 사용하여 계산식 편집기의 기본 기능 사용 방법을 안내합니다. Big Class.jmp 는 JMP 와 함께 설치되는 샘플 데이터 폴더에 있습니다.

Big Class.jmp 에는 weight 라는 열이 있습니다. 표준화된 체중 값을 계산하는 새 열이 필요하다고 가정해 보겠습니다.

계산식으로 이 열을 생성하여 값을 구하려면 다음 단계를 수행하십시오 .

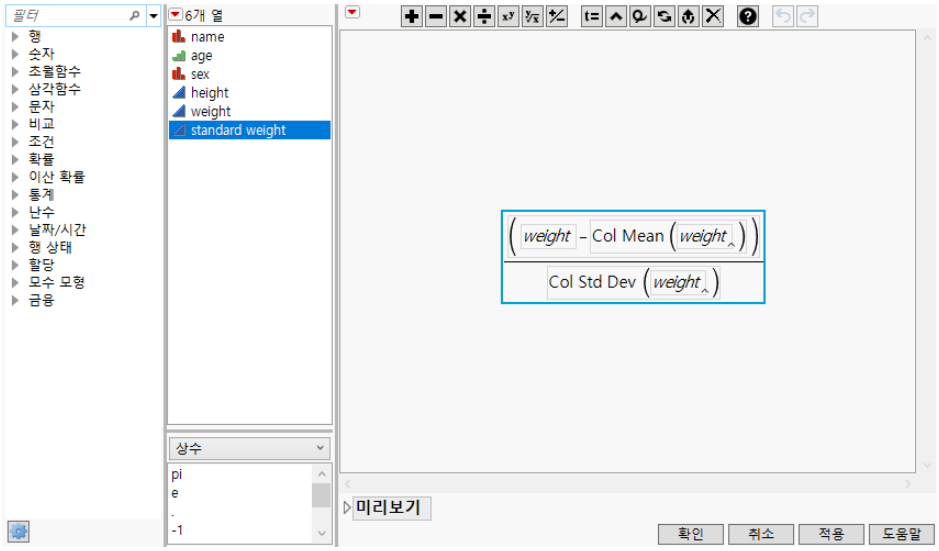
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **열 > 새 열**을 선택합니다 .
3. " 열 이름 " 옆의 상자에 새 이름 standard weight 를 입력합니다 .
4. " 열 특성 ' 메뉴에서 **계산식**을 선택합니다 .
5. 계산식 편집 영역에서 빈 계산식 요소를 클릭하여 선택합니다 .

   " 상자 표시 " 가 선택된 상태로 계산식을 생성하면 계산식에서 선택한 부분에 가는 파란색 외곽선이 표시됩니다 . 자세한 내용은 " 계산식에서 개요 숨기기 및 표시 "(385 페이지) 에서 확인하십시오 . 클릭한 위치를 기준으로 가장 작은 내포 상자 내의 모든 항이 선택되며 , 이후 작업은 이 결합된 요소에 적용됩니다 .

다음으로 , weight 값을 표준화하는 계산식을 입력합니다.

1. 초기의 빈 항이 선택되어 있는 상태에서 열 목록의 weight 를 클릭합니다 .
2. 계산식 편집기의 키패드에서 빼기 버튼 ➖을 클릭합니다 .
   그림 7.20 에 표시된 것과 같이 빼기 기호 다음에 새로운 빈 항이 나타납니다 .
3. weight 를 다시 클릭합니다 .
4. 함수 목록을 클릭하고 **통계**를 선택합니다 .
5. " 통계 " 목록에서 **Col Mean** 을 선택합니다 .
6. 전체 표현식을 선택합니다 .
   이제 파란색 상자가 전체 계산식을 둘러쌉니다 .
7. 키패드에서 나누기 버튼 ➗을 클릭합니다 .
   그러면 전체 표현식에 대한 빈 분모가 선택된 상태로 표시됩니다 .
8. 열 선택 목록에서 weight 를 다시 클릭합니다 .
   해당 항목이 분모에 선택된 상태로 표시됩니다 .
9. **통계** 목록에서 **Col Std Dev** 를 선택합니다 .

그림 7.20 계산식 작성



10. **확인**을 클릭하여 계산식 편집기를 닫습니다 .
11. **확인**을 클릭하여 열 정보 창을 닫습니다 .

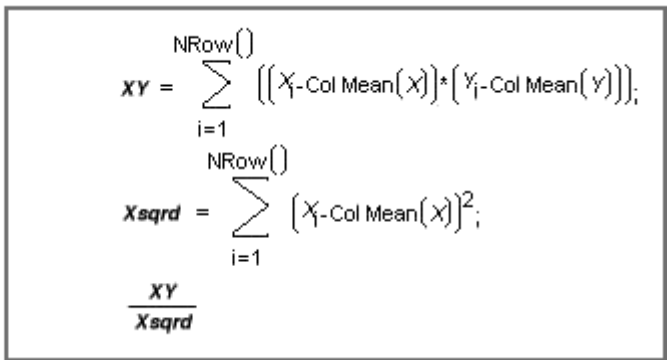
데이터 테이블에서 새 standard weight 열에 값이 채워집니다. weight 값이 변경되면 계산된 standard weight 값이 자동으로 다시 계산됩니다.

계산식에서 지역 변수 사용

그림 7.21 에 표시된 표준 계산식을 사용하여 x 에 대한 y 의 단순 선형 회귀에서 기울기를 계산하려고 합니다. 이를 수행하는 한 가지 방법은 두 개의 지역 변수를 생성하여 XY 및 Xsqrd 라는 이름을 지정하는 것입니다. 자세한 내용은 "계산식에서 지역 변수 사용"(365 페이지)에서 확인하십시오. 그런 다음 기울기 계산식의 분자 및 분모 계산에 이 두 지역 변수를 할당합니다. 각 할당은 그림 7.21 에 표시된 것과 같이 세미콜론으로 구분합니다. 계산식 편집기의 문은 실제로는 JSL 프로그래밍 문입니다. 계산식에 문이 여러 개 포함될 경우 각 문을 세미콜론으로 구분해야 합니다. 결과 기울기 계산은 XY 나누기 Xsqrd 로 단순화됩니다.

그림 7.21 표준 기울기 계산식의 지역 변수

$$\frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$



참고 : 화면의 키패드에 있는 t= 버튼을 사용하여 지역 변수를 생성할 수도 있습니다 . 이 버튼을 클릭하면 표현식에 기본 이름의 지역 변수가 생성되고 그 뒤에 세미콜론이 추가됩니다 . 기본 이름은 *t1* 이고 추가 지역 변수의 이름은 *t2*, *t3* 등으로 지정됩니다 . 이름을 두 번 클릭한 후 편집하여 이러한 기본 이름을 변경할 수 있습니다 . 하지만 변수의 각 인스턴스 이름을 바꿀 때는 오류가 없도록 주의해야 합니다 .

계산식에서 Munger 함수 사용

다음의 몇 가지 예에서는 Munger 함수의 사용 방법을 보여 줍니다. 이러한 예에서는 문자 유형의 이름 열이 있고 이 열의 값 중 하나로 "Veronica Layman" 이 포함되어 있다고 가정합니다. 예를 단순하게 하기 위해 열 이름 대신 리터럴 이름인 "Veronica Layman" 을 검색 문자열로 사용합니다.

Munger 같은 문자 함수를 계산식에 포함하는 방법은 "계산식에서 JMP 함수에 대한 참조" 부록의 "문자 함수"(684 페이지)에서 확인하십시오.

문자 삽입

이 Munger 예에서는 이름과 성 사이의 공백을 찾아 중간 이니셜인 "J."를 삽입합니다. 계산식 **Munger("Veronica Layman", 1, " ", " J. ")**는 중간 이니셜 "J."를 삽입하여 결과가 "Veronica J. Layman"이 되도록 합니다.

Munger 함수에서는 공백으로 구성된 문자열이나 선행 또는 후행 공백이 포함된 문자열을 비롯한 리터럴 문자열에 큰따옴표를 사용해야 합니다.

문자 삭제

문자열에서 하나 이상의 문자를 삭제하려면 다음 단계를 수행하십시오 .

1. Munger 함수에서 삭제할 문자를 Find 문자열로 지정합니다 .
2. Replace 문자열은 빈 문자열 , 즉 가운데에 아무것도 없는 두 개의 인용 부호로 입력합니다 .

예를 들어 함수 **Munger("Veronica Layman", 1, "onic", "")**는 "Veronica"에서 "onic"을 제거하여 "Vera Layman"이 되도록 합니다.

참고 : null(값 없음) 문자열을 인용 부호로 묶은 Replace 필드는 값이 없는 Replace 필드와는 다릅니다 . Replace 문자열을 모두 삭제하면 Munger 는 계산식 편집기 창에 인수 이름 ("Replace") 을 표시하고 해당 선택적 인수가 없는 것처럼 작동합니다 . 또한 **Find/Length** 인수의 값에 따라 결과 데이터 유형이 문자에서 숫자로 변경될 수도 있습니다 .

값 위치 (인덱스) 찾기

Find/Length 필드에 문자가 포함되어 있으면 Munger는 인덱스 함수처럼 작동하여 검색 문자열의 첫 번째 인스턴스(있는 경우)가 있는 숫자 위치를 반환합니다. 예를 들어 **Munger("Veronica Layman", 1, " ")**는 단일 공백을 검색한 결과 위치 9에서 공백을 찾습니다. 검색 문자열을 찾을 수 없으면 Munger는 0을 반환합니다. 이와 같이 Munger를 사용하면 "계산식에서 JMP 함수에 대한 참조" 부록의 "문자 함수"(684 페이지)에 설명된 Contains 함수를 사용하는 것과 동일한 결과가 생성됩니다.

하위 문자열 찾기

Munger로 하위 문자열을 추출할 수 있습니다. 예를 들어 이름만 추출하려는 경우 **Munger ("Veronica Layman", 1, 8,)**는 위치 1 에서부터 위치 8 까지의 문자를 읽습니다. Replace 인수가 정의되지 않았으므로 나머지 문자는 무시됩니다. 결과는 "Veronica"가 되어 "계산식에서 JMP 함수에 대한 참조" 부록의 "문자 함수"(684 페이지)에 설명된 Substring과 동일한 결과를 생성하게 됩니다.

하위 문자열을 찾는 다른 방법은 start 값, 음수 find 값 및 **no replace** 인수를 사용하는 것입니다. **Munger("Veronica Layman", 9, –1,)**는 "Layman"을 반환합니다.

계산식에서 Match 조건 함수 사용

이 예에서는 Match 조건 함수의 사용 방법을 안내합니다.

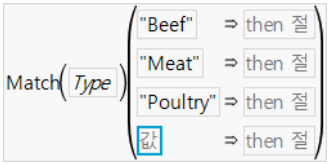
Hot Dogs.jmp 샘플 데이터 테이블에서 명목형 변수 Type에 대해 Match 조건 함수를 실행하려고 합니다.

1. 계산식 편집기의 열 목록에서 Type 을 선택합니다 .
2. 함수 목록에서 **조건**을 선택합니다 .
3. **Match** 를 선택합니다 .
4. **데이터에서 Match 인수 추가**를 선택합니다 .

   값이 자동으로 채워집니다 .

참고 : 2 단계 ~ 4 단계를 완료하는 대신 , Shift 키를 누른 채 함수 목록의 **조건**을 선택하고 **Match** 를 선택해도 됩니다 .

그림 7.22 Match 조건 문 자동 채우기



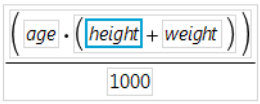
값을 자동으로 채우지 않으려면 Match 목록에서 **데이터에서 Match 인수 추가** 대신 **추가 안 함**을 선택합니다.

계산식에서 표현식 삭제 버튼 사용

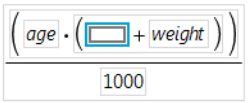
다음 단계에서는 표현식 삭제 (또는 벗기기) 버튼 을 반복해서 클릭할 때의 결과를 보여 줍니다 .

1. 계산식에서 시작합니다 .

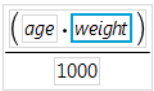
2. 계산식 요소를 선택합니다 .



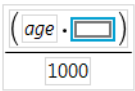
3. 표현식 삭제 버튼 을 클릭합니다 .



4. 표현식 삭제 버튼을 클릭합니다 .



5. 표현식 삭제 버튼을 클릭합니다 .



6. 표현식 삭제 버튼을 클릭합니다 .

7. 표현식 삭제 버튼을 클릭합니다 .

8. 표현식 삭제 버튼을 클릭합니다 .

9. 표현식 삭제 버튼을 클릭합니다 .

JMP 계산식 편집기의 바로 가기 키

표 7.4에서는 JMP 계산식 편집기에서 사용할 수 있는 바로 가기 키를 설명합니다.

표 7.4 계산식 편집기의 바로 가기 키

삽입할 항목	사용할 바로 가기 키
결측 요소(•)	삭제
서브스크립트	[
지수	^(Shift 키를 누른 채 6을 입력합니다.)
괄호 쌍 ()	(

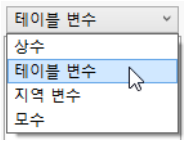
표 7.4 계산식 편집기의 바로 가기 키 (계속)

삽입할 항목	사용할 바로 가기 키
인용 부호 쌍 " "	"
선택한 필드 앞의 새 인수	Alt + ^
*(곱하기 기호)	키패드 또는 키보드의 *
+	키패드 또는 키보드의 +
–	키패드 또는 키보드의 -
÷	키패드 또는 키보드의 /
+/–	Shift 키를 누른 채 +-(?기) 누르기
And	&
Not	!
Or	|
Power	^
새 인수	,
x<y	<
x==y	=
x>y	>

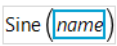
JMP 계산식 편집기의 용어

계산식 편집기에서는 다음과 같은 용어가 사용됩니다 .

요소 요소 브라우저 목록에 표시되는 상수 , 테이블 변수 , 지역 변수 또는 모수의 이름입니다 .

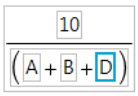


인수 함수의 연산 대상인 요소 또는 전체 표현식 (수학적 피연산자 포함) 입니다 . 인수는 항상 함수와 함께 그룹화됩니다 . 함수의 인수 역할을 하는 표현식을 확인하려면 계산식에서 함수를 선택합니다 .

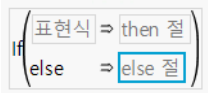


항 상수 및 변수와 같이 더 이상 나눌 수 없는 표현식 요소입니다 .

표현식 계산식에서 단일 단위로 선택할 수 있는 모든 부분으로 , 항 , 빈 항 , 인수와 함께 그룹화된 함수뿐만 아니라 전체 계산식이 이에 해당합니다 .



절 조건 함수의 완전한 세그먼트입니다 .



함수 하나 이상의 인수에 대해 특정 작업을 수행하는 수학적 또는 논리적 연산입니다 . 함수 목록에 있는 대부분의 항목과 모든 키패드 연산자가 함수에 해당합니다 . 함수는 항상 선택된 표현식에 대해 작동하며 인수는 항상 함수와 함께 그룹화됩니다 . 함수의 인수 역할을 하는 표현식을 확인하려면 계산식에서 함수를 선택합니다 .

또한 상자로 둘러싸인 그룹에는 우선 순위 규칙이 적용되는 방식과 함수를 삭제할 경우에 삭제되는 인수도 표시됩니다 . 자세한 내용은 " 계산식의 표현식 정렬 "(377 페이지) 에서 확인하십시오 .

빈 항 표현식을 추가하기 위한 빈 자리 표시자로 , 빈 상자로 표시됩니다 .

결측값 숫자 데이터의 경우 결측값 표시 (•), 문자 데이터의 경우 null 문자열로 구성된 제외된 데이터 또는 null 데이터입니다 .

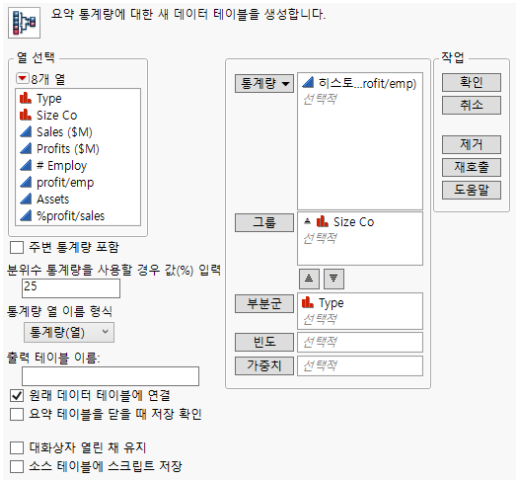
	name	age	sex	height	weight
1	KATIE	•	F	59	95
2		12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84

8 장

데이터 요약
테이블 요약 명령

이 장에서는 중앙값, 표준편차, 최소값 및 최대값과 같은 요약 통계량이 포함된 요약 데이터 테이블을 생성하는 방법에 대해 설명합니다.

그림 8.1 Companies.jmp 의 요약 테이블



	Size Co	행 수	평균(profit/emp, Computer)	평균(profit/emp, Pharmaceutical)
1	big	9	4530.478	17140.699
2	medium	7	-3462.506	24035.115
3	small	16	7998.815	38337.191

목차

열 요약 397
요약 시작 창 397
요약 테이블 생성 397
기존 요약 테이블에 통계량 열 추가 401
요약 통계량에 대한 설명 402
예 : 그룹을 사용하여 요약 테이블 생성 403

열 요약

"테이블" > "요약" 명령은 평균, 중앙값, 표준편차, 최소값 및 최대값 등의 다양한 요약 통계량을 계산합니다.

요약 테이블에는 다음과 같은 특성이 있습니다.

- 지정한 그룹화 변수의 각 수준마다 한 개씩의 행이 있습니다. 그룹화 변수가 지정되지 않은 경우에는 전체 데이터 테이블에 대한 한 개의 행이 있습니다.
- 그룹화 변수가 여러 개인 경우 테이블에는 모든 그룹화 변수 수준의 각 조합에 대한 행이 포함됩니다.
- 이 테이블에는 각 그룹화 변수마다 한 개씩의 열이 포함될 뿐 아니라, **행 수**라는 열에 각 그룹화 수준의 개수를 보여 주는 빈도 수가 포함됩니다.
- 요약 테이블은 소스 테이블에 연결될 수 있습니다. 요약 테이블에서 행을 선택하면 소스 테이블에서 해당하는 행이 강조 표시됩니다.
- 소스 테이블의 열에 값 라벨이 포함되어 있으면 새 테이블에 값 라벨이 표시됩니다.
- 요약 테이블은 테이블을 닫을 때 저장되지 않으며 **파일 > 다른 이름으로 저장**을 선택하여 파일 이름 및 위치를 지정해야만 저장됩니다.

요약 시작 창

통계량 소스 테이블의 숫자 열에 대한 요약 통계량 (평균, 표준편차, 중앙값 등) 을 정의합니다. 자세한 내용은 " 요약 통계량 추가 "(398 페이지) 에서 확인하십시오.

그룹 그룹별로 통계량을 요약합니다. 자세한 내용은 " 하나 이상의 그룹화 열 사용 "(398 페이지) 에서 확인하십시오.

부분군 부분군별로 통계량을 요약합니다. 자세한 내용은 " 부분군 변수를 추가하여 이원 요약 통계량 테이블 생성 "(401 페이지) 에서 확인하십시오.

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 열을 식별합니다.

가중치 각 행에 가중치를 할당하는 데 사용할 값이 있는 열을 식별합니다.

요약 테이블 생성

요약 테이블을 생성하려면 다음 단계를 수행하십시오.

1. 데이터 테이블을 엽니다.
2. **테이블 > 요약**을 선택합니다.
3. 요약할 열을 강조 표시합니다.

참고 : 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 "JMP 시작하기 " 장의 " 열 필터 메뉴 "(53 페이지) 에서 확인하십시오.

4. 요약 통계량, 그룹, 부분군, 빈도 변수 및 가중치 변수를 추가하고 필요한 옵션을 선택합니다.
   - " 요약 통계량 추가 "(398 페이지)
   - " 하나 이상의 그룹화 열 사용 "(398 페이지)
   - " 분위수 통계량 사용 "(399 페이지)
   - " 통계량 열 이름의 형식 변경 "(400 페이지)
   - " 원래 데이터 테이블에 연결 "(400 페이지)
   - " 연결된 요약 테이블을 닫을 때 저장 확인 "(400 페이지)
   - " 요약 창 열린 채 유지 "(400 페이지)
   - " 부분군 변수를 추가하여 이원 요약 통계량 테이블 생성 "(401 페이지)
5. **출력 테이블 이름** 옆의 상자에 이름을 입력하여 요약 테이블의 이름을 지정합니다 .
6. **확인**을 클릭합니다 .

참고 :

- 요약 테이블이 원래 데이터 테이블에 연결되어 있으면 요약 테이블의 데이터를 편집할 수 없습니다 . 그러면 원래 데이터가 수정되어 손상되기 때문입니다 . 데이터를 편집하려면 요약 테이블의 부분집합을 생성하거나 테이블을 저장한 후 테이블을 닫았다가 다시 여십시오 .
- " 단위 " 열 특성이 있는 열을 선택하면 해당 값이 열 머리글에 포함됩니다 .

요약 통계량 추가

소스 테이블의 숫자 열에 대한 요약 통계량(평균, 표준편차, 중앙값 등)을 표시하는 열을 추가할 수 있습니다.

1. " 요약 " 창에서 통계량을 계산하는 데 사용할 열을 강조 표시합니다 .
2. **통계량** 버튼을 클릭합니다 .
3. " 통계량 " 목록에서 표준 단변량 기술 통계량 중 하나를 선택합니다 . 통계량에 대한 자세한 내용은 " 요약 통계량에 대한 설명 "(402 페이지) 에서 확인하십시오 .

하나 이상의 그룹화 열 사용

그룹별로 통계량을 요약하려면 그룹화 변수로 사용할 열을 강조 표시하고 **그룹**을 클릭하여 해당 변수를 그룹화 변수 목록으로 이동합니다. 예는 " 예 : 그룹을 사용하여 요약 테이블 생성 "(403 페이지)에서 확인하십시오. 그룹화 변수만 추가하면 요약 테이블에 각 그룹의 수가 표시됩니다.

그룹화 변수의 순서를 변경하려면

그룹화 변수의 순서(오름차순 또는 내림차순)를 변경하려면 그룹화 변수 목록에서 변수를 선택하고 오름차순 또는 내림차순 버튼( )을 클릭합니다. 그러면 변수 옆의 아이콘이 정렬 순서를 나타내는 모양으로 바뀝니다.

"값 순서" 열 특성을 사용하여 그룹화 변수의 순서를 변경할 수도 있습니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "값 순서"(285 페이지)에서 확인하십시오.

주변 통계량을 포함하려면

그룹화 변수에 대한 주변 통계량을 출력 열에 추가하려면 **주변 통계량 포함** 옆의 상자를 클릭합니다. 그러면 각 그룹화 변수에 대한 주변 통계량이 추가될 뿐만 아니라, 첫 번째 그룹화 변수의 각 수준을 요약하는 행이 테이블 끝에 추가됩니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **테이블 > 요약**을 선택합니다 .
3. Profits ($M) 를 선택하고 **통계량**을 클릭합니다 .
4. **평균**을 선택합니다 .
5. Type 및 Size Co 를 선택하고 **그룹**을 클릭합니다 .
6. **주변 통계량 포함**을 선택합니다 .
7. **확인** (또는 **생성**) 을 클릭합니다 . 그림 8.2 의 왼쪽을 참조하십시오 .

그림 8.2 주변 통계량이 포함된 요약 테이블과 포함되지 않은 요약 테이블

	Type	Size Co	행 수	평균(Profits ($M))
1	Computer	big	4	1089.93
2	Computer	medium	2	-85.75
3	Computer	small	14	44.94
4	Computer		20	240.87
5	Pharmaceutical	big	5	894.42
6	Pharmaceutical	medium	5	698.98
7	Pharmaceutical	small	2	156.95
8	Pharmaceutical		12	690.08
9			32	409.32

	Type	Size Co	행 수	평균(Profits ($M))
1	Computer	big	4	1089.93
2	Computer	medium	2	-85.75
3	Computer	small	14	44.94
4	Pharmaceutical	big	5	894.42
5	Pharmaceutical	medium	5	698.98
6	Pharmaceutical	small	2	156.95

주변 통계량이 포함된 요약 테이블(왼쪽)과 주변 통계량이 포함되지 않은 테이블(오른쪽)을 비교해 보십시오. 주변 통계량이 추가되면 총 32 개의 컴퓨터 회사와 제약 회사가 있음을 보여 주는 행이 추가됩니다.

분위수 통계량 사용

특정 분위수 통계량을 추가하려면 다음 단계를 수행하십시오 .

1. **분위수 통계량을 사용할 경우 값 (%) 입력** 아래의 상자에 첫 번째 분위수에 대해 원하는 분위수 값 (예 : 25) 을 입력합니다 .
2. 적절한 열을 선택하고 **통계량**을 클릭합니다 .
3. **분위수**를 선택합니다 .
4. (선택 사항) 분위수를 더 추가하려면 이 과정을 반복합니다 .

통계량 열 이름의 형식 변경

요약 테이블에 포함되는 통계량 열 이름의 형식을 변경하려면 **통계량 열 이름 형식** 메뉴의 형식 중 하나를 선택합니다. 표 8.1에서는 사용 가능한 옵션을 보여 줍니다. 여기에서는 회사 평균 수익의 요약 테이블을 생성한다고 가정합니다. 원래 열 이름은 Profits ($M)입니다.

표 8.1 통계량 열 이름 형식 옵션 및 예

옵션	예제
통계량(열)	평균(Profits ($M))
열	Profits ($M)
열의 통계량	Profits ($M)의 평균
열 통계량	Profits ($M) 평균
통계량	Mean

원래 데이터 테이블에 연결

요약 테이블을 원래 데이터 테이블에 연결할지 여부를 선택할 수 있습니다. 기본적으로는 **원래 데이터 테이블에 연결** 옵션이 선택되어 있습니다. 요약 테이블에서 데이터를 편집하려면 **원래 데이터 테이블에 연결** 옵션을 선택 취소하십시오. 요약 테이블이 원래 데이터 테이블에 연결되어 있으면 요약 테이블의 데이터를 편집할 수 없습니다. 그러면 원래 데이터가 수정되어 손상되기 때문입니다. 데이터를 편집하려면 요약 테이블의 부분집합을 생성하거나 테이블을 저장한 후 테이블을 닫았다가 다시 여십시오.

연결된 테이블 내에서 요약 테이블의 열을 원래 데이터 테이블에 있는 새 열의 열 머리글로 드래그하면 값이 마치 그룹화 열을 기준으로 매칭된 것처럼 확장됩니다.

연결된 요약 테이블을 닫을 때 저장 확인

연결된 요약 테이블을 저장하지 않고 닫을 때 파일을 저장할지 묻는 메시지를 표시하려면 **요약 테이블을 닫을 때 저장 확인**을 선택합니다. "요약" 창에서 이 옵션을 선택하면 "테이블" 환경 설정에도 해당 옵션이 선택됩니다. 이후에 연결된 요약 테이블을 저장할지 묻는 메시지를 표시하지 않으려면 해당 환경 설정을 선택 취소하십시오.

요약 창 열린 채 유지

대화상자 열린 채 유지 옵션을 선택하면 **적용**을 클릭한 후에도 "요약" 창이 열린 채로 유지됩니다. 이 옵션을 선택하면 "확인" 버튼이 "생성" 버튼으로 바뀝니다.

부분군 변수를 추가하여 이원 요약 통계량 테이블 생성

1. 내포 변수로 설정할 열을 강조 표시합니다 . 해당 열은 " 부분군 변수 " 가 됩니다 .
2. **부분군**을 클릭하여 해당 변수를 부분군 목록으로 이동합니다 .
3. 부분군별로 통계량을 요약할 열을 강조 표시합니다 .
4. " 통계량 " 목록에서 원하는 특정 통계량을 선택합니다 .
5. **확인**을 클릭합니다 .

통계량 유형에 대한 자세한 내용은 " 요약 통계량에 대한 설명 "(402 페이지) 에서 확인하십시오.

기존 요약 테이블에 통계량 열 추가

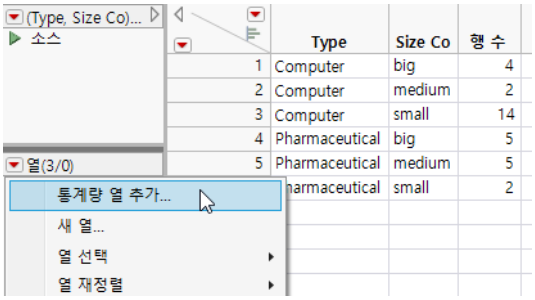
요약 테이블을 생성한 후에 소스 테이블의 숫자 열에 대한 기술 요약 통계량 열을 추가할 수 있습니다. 이렇게 하려면 기존 요약 테이블에서 데이터 격자의 위쪽 빨간색 삼각형을 클릭하고 **통계량 열 추가**를 선택합니다.

예 : 기존 테이블에 통계량 열 추가

요약 테이블을 이미 생성했는데 기존 요약 테이블에 통계량을 더 추가하려고 한다고 가정해 보겠습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **테이블 > 요약**을 선택합니다 .
3. Type 및 Size Co 를 선택하고 **그룹**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. 열 패널에서 빨간색 삼각형을 클릭하고 **통계량 열 추가**를 선택합니다 .

그림 8.3 데이터 테이블 내에서 요약 통계량 열 생성



수정된 버전의 "요약" 창이 나타납니다.

6. 원하는 열을 선택하고 **통계량**을 클릭한 후 원하는 특정 통계량을 선택합니다 . 이 예에서는 profit/emp 를 선택하고 **통계량**을 클릭한 후 **평균**을 선택합니다 .

7. **확인**을 클릭합니다 .

그림 8.4 요약 통계량 열이 있는 요약 테이블의 예

	Type	Size Co	행 수	평균(profit/emp)
1	Computer	big	4	4530.478
2	Computer	medium	2	-3462.506
3	Computer	small	14	7998.815
4	Pharmaceutical	big	5	17140.699
5	Pharmaceutical	medium	5	24035.115
6	Pharmaceutical	small	2	38337.191

요약 통계량 열

기존 요약 테이블에 "Mean(profit/emp)" 열이 추가되었습니다.

요약 통계량에 대한 설명

통계량 버튼을 클릭하고 메뉴에서 옵션을 선택하여 소스 테이블의 숫자 열에 대한 기술 요약 통계량 열을 추가할 수 있습니다.

"통계량" 메뉴에서는 숫자 열에 대해 다음과 같은 요약 통계량을 제공합니다.

N 비결측값의 개수입니다 .

평균 열 값의 산술평균입니다 . 비결측값 (정의된 경우 가중치 변수를 곱한 값) 의 합을 가중합으로 나눈 결과입니다 .

표준편차 비결측값에 대해 지정된 가중치 변수를 사용하여 계산된 표본 표준편차입니다 . 표본 분산의 제곱근입니다 . 계산식은 다음과 같습니다 .

$$s = \sqrt{s^2}$$

$$s^2 = \sum_{i=1}^{N} \frac{w_i(y_i - \bar{y}_w)^2}{N-1}$$

$\bar{y}_w$ = 가중 평균

최소값 열에 있는 가장 작은 비결측값입니다 .

최대값 열에 있는 가장 큰 비결측값입니다 .

범위 최대값과 최소값의 차이입니다 .

% 총계 각 그룹의 총 수가 차지하는 백분율입니다 . 또는 각 그룹의 총 수에 대한 비결측 열 값의 백분율입니다 (지정된 경우).

결측값 수 결측값의 개수입니다 .

범주 수 개별 범주의 개수입니다 .

합 열에 있는 모든 값의 합입니다 .

가중합 열에 있는 모든 **가중치** 값의 합입니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 "JMP 의 열 특성 "(280 페이지) 에서 확인하십시오 . 가중치 역할에 할당된 열이 없는 경우 , **가중합**은 비결측값의 총 수입니다 .

분산 비결측값에 대해 계산된 표본 분산입니다 . 평균으로부터의 편차에 대한 제곱합을 비결측값의 수에서 1 을 뺀 값으로 나눈 것입니다 .

표준 오차 평균의 표준 오차입니다 . 표준편차를 **N** 의 제곱근으로 나눈 것입니다 . 열에 가중치 역할이 할당된 경우에는 분모가 가중치 합의 제곱근입니다 .

CV(변동 계수) 표준편차를 평균으로 나누고 100 을 곱한 산포 측도입니다 .

중앙값 데이터의 위쪽 절반과 아래쪽 절반을 나누는 기준이 되는 50 번째 백분위수 , 또는 50 번째 분위수 (중앙값) 입니다 .

중앙 절대 편차 데이터 중앙값과의 절대 편차에 대한 중앙값입니다 .

기하평균 데이터 곱의 *n* 제곱근입니다 . 예를 들어 기하평균은 이자율을 계산하는 데 종종 사용됩니다 . 이 통계량은 데이터에 많은 값이 비대칭적 분포로 포함된 경우에도 유용합니다 .

참고 : 음수 값이 있으면 결측값이 발생하고 , 음수가 아니고 0 인 값이 있으면 결과가 0 이 됩니다 .

사분위수 범위 3 사분위수와 1 사분위수 간의 차이입니다 .

분위수 특정 비율 지점을 나타내는 값으로 , 해당 비율만큼의 인수가 해당 값보다 작거나 같습니다 . 예를 들어 75% 의 데이터는 75 분위수보다 작습니다 . 요약 창에는 원하는 분위수 백분율을 입력할 수 있는 편집 상자가 있습니다 .

히스토그램 서로 다른 여러 그룹에 대한 히스토그램을 생성합니다 . 히스토그램의 이미지는 데이터 테이블 표현식 열에 저장됩니다 .

예 : 그룹을 사용하여 요약 테이블 생성

한 연구원이 회사를 Type 및 Size 별로 그룹화하는 Companies.jmp 를 사용하고 있다고 가정해 보겠습니다 . JMP 와 함께 설치된 샘플 데이터 폴더에서 Companies.jmp 를 열어 이 예의 단계를 차례로 수행해 보십시오 .

이 연구원은 다음과 같은 작업을 수행하려고 합니다 .

- 소규모 , 중간 규모 및 대규모인 컴퓨터 회사와 제약 회사의 직원당 평균 수익을 보여 주는 테이블을 생성합니다 . 즉 , 각 회사 규모에 해당하는 행 하나와 각 회사 유형의 직원당 평균 수익에 해당하는 열 하나가 포함된 테이블을 생성합니다 .
- 이 테이블을 생성할 때 행과 열의 교차를 통해 정의된 부분군의 평균이 셀에 포함되도록 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **테이블 > 요약**을 선택합니다 .

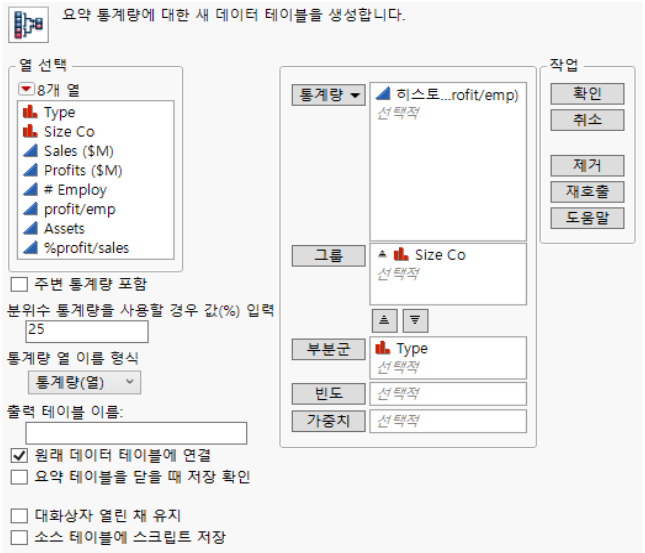
3. Size Co 를 선택하고 **그룹**을 클릭합니다 .

   연구원은 Size Co 열의 값이 새 테이블의 행이 되기를 원하므로 이 열을 그룹화 변수로 선택합니다 .

4. profit/emp 를 선택하고 **통계량**을 클릭합니다 .
5. **평균**을 선택합니다 .
6. Type 을 선택하고 **부분군**을 클릭합니다 .

   이렇게 하면 부분군 변수 (type) 별 각 수준 (computer, pharmaceutical) 의 직원당 평균 수익 (Mean(profit/emp)) 을 보여 주는 열이 생성됩니다 .

그림 8.5 부분군별 요약 통계량



	Size Co	행 수	평균(profit/emp, Computer)	평균(profit/emp, Pharmaceutical)
1	big	9	4530.478	17140.699
2	medium	7	-3462.506	24035.115
3	small	16	7998.815	38337.191

참고 : 요약 테이블의 이름은 그룹 열 이름을 따라 지정됩니다 . 열 이름에 파일 이름에 사용할 수 없는 문자가 포함되어 있는 경우 해당 특수 문자는 하이픈으로 대체됩니다 .

9 장

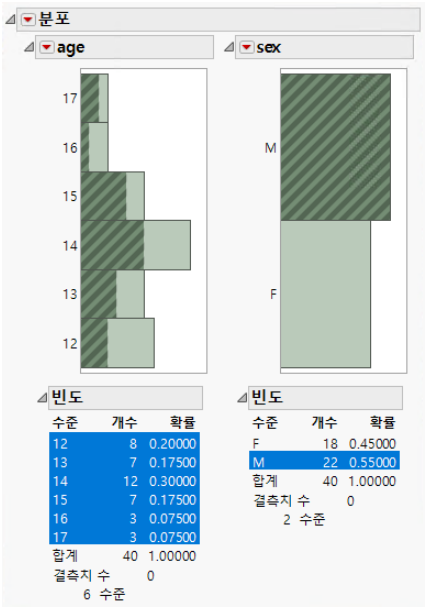
JMP 보고서
보고서 창 탐색 및 사용자 정의

대부분의 JMP 플랫폼은 분석 결과를 보고서 창에 표시합니다. 보고서 창은 그래프와 통계 보고서로 구성되며, 이 둘은 대개 데이터 테이블에 연결되어 있습니다. 그래프 또는 보고서에서 데이터를 선택하면 데이터 테이블에서 해당 데이터가 강조 표시됩니다. 그림 9.1에서는 보고서 창과 데이터 테이블에서 선택된 데이터의 예를 보여 줍니다.

이 장에서는 보고서 창의 색상, 선, 데이터 점 및 기타 그래픽 요소를 사용자 정의하는 방법에 대해 설명합니다.

그림 9.1 JMP 보고서 창 및 데이터 테이블의 예

	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107
11	SUSAN	13	F	56	67
12	JOHN	13	M	65	98
13	JOE	13	M	63	105
14	MICHAEL	13	M	58	95
15	DAVID	13	M	59	79
16	JUDY	14	F	61	81
17	ELIZABETH	14	F	62	91
18	LESLIE	14	F	65	142
19	CAROL	14	F	63	84
20	PATTY	14	F	62	85
21	FREDERICK	14	M	63	93
22	ALFRED	14	M	64	99



목차

JMP 보고서 탐색 408
손 도구 사용 409
JMP 보고서의 표시 옵션 409
JMP 보고서 요소 표시 및 숨기기 411
여러 보고서 결합 411
보고서의 제목 바꾸기 412
보고서의 글꼴 크기 늘리기 412
보고서 창 크기 조정 412
JMP 보고서의 빨간색 삼각형 옵션 412
보고서의 로컬 데이터 필터 413
보고서의 다시 실행 메뉴 416
보고서의 스크립트 저장 메뉴 416
데이터 필터 419
필터 열의 유형 420
필터링 모드 421
데이터 필터 제어판 422
데이터 필터의 빨간색 삼각형 메뉴 옵션 424
데이터 필터의 애니메이션 기능 427
데이터 필터의 변수에 대한 빨간색 삼각형 메뉴 옵션 429
보고서의 테이블 형식 조정 432
JMP 보고서의 표시 상자 특성 수정 434
JMP 보고서의 테이블을 행렬로 변환 435
보고서에서 조건부 형식 사용 435
그래프에서 점 선택 441
그래프에서 행 선택 441
JMP 그래프에서 직사각형 영역으로 점 선택 442
JMP 그래프에서 불규칙한 형태의 영역 내에 있는 점 선택 443
그래프에서 표식 사용 444
표식 형태 변경 444
표식 색상 변경 445
표식 크기 변경 445
많은 수의 표식 처리 445
표식 주위에 외곽선 추가 446
표식 선택 모드 변경 446
표식 투명도 지정 447
표식 제외 및 숨기기 448
이미지를 표식으로 사용 448
값을 기반으로 표식 형태 또는 색상 변경 449

그래프 모양 변경 451
그래프 크기 조정 452
그래프의 확대 453
그래프의 선 너비 변경 453
그래프의 배경 색상 변경 453
히스토그램 막대의 색상 변경 454
그래프의 좌표 및 임시 참조선 표시 454
그래프의 축 스크롤 및 척도 조정 454
그래프의 축 및 축 라벨 사용자 정의 455
보고서의 값 순서 변경 465
그래프에서 선택한 개체의 패턴 및 형식 변경 466
데이터 테이블에 그래프 추가 466
그래프의 표식, 선, 텍스트 등 사용자 정의 467
JMP 그래프 또는 보고서에 이미지 추가 473
JMP 그래프에 배경 이미지 붙여넣기 473
JMP 보고서 끝에 이미지 붙여넣기 473
이미지를 JMP 그래프로 드래그하여 놓기 474
JMP 그래프에 지리 이미지 및 경계 추가 476
이미지에서 데이터 추출 476
JMP 보고서 및 그래프에 주석 및 형태 추가 477
JMP 보고서 및 그래프에 주석 추가 477
JMP 보고서에 형태 추가 479
그래프의 가리키기 라벨 사용자 정의 482
가리키기 라벨에 그래프 또는 이미지 추가 483
그래프의 가리키기 라벨에 서식 있는 텍스트 추가 498
가리키기 라벨의 이름 및 값 업데이트 501
그래프의 가리키기 라벨 편집 511
데이터 테이블에서 가리키기 라벨에 이미지 추가 516
그래프의 가리키기 라벨 고정 517
가리키기 라벨의 마우스 오른쪽 버튼 클릭 옵션 518

JMP 보고서 탐색

JMP 보고서는 스크롤 막대와 크기 조정 옵션이 있는 표준 창에 표시됩니다. 이 창에는 그림 9.2에 표시된 것과 같은 그 밖의 특수 버튼 및 메뉴도 있습니다. 이러한 버튼 및 메뉴에 대해서는 이후 섹션에서 설명합니다.

그림 9.2 보고서 창의 기본 요소

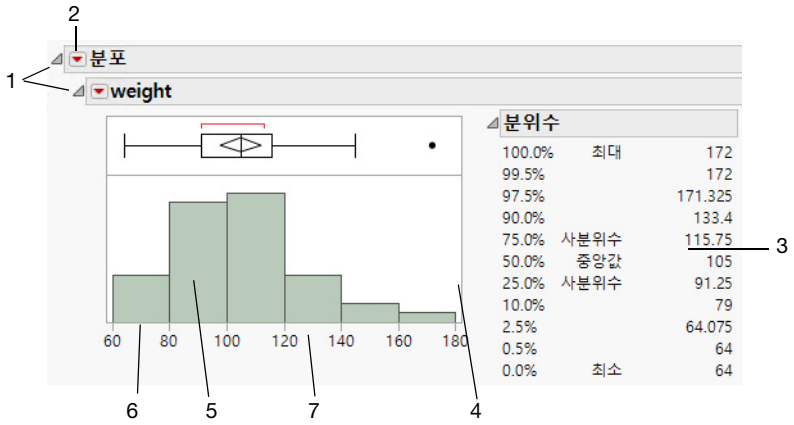


표 9.1 보고서 창 작업

개수	작업
1	보고서의 특정 섹션을 숨기거나 표시하려면 이 표시 아이콘을 클릭합니다.
2	보고서 옵션에 액세스하려면 빨간색 삼각형을 클릭합니다.
3	테이블에서 형식 지정 옵션에 액세스하려면 마우스 오른쪽 버튼을 클릭합니다.
4	그래프 크기를 조정하려면 테두리를 클릭한 채로 드래그합니다.
5	그래프에서 형식 지정 옵션에 액세스하려면 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다.
6	축에서 형식 지정 옵션에 액세스하려면 마우스 오른쪽 버튼을 클릭합니다.
7	축 위에 포인터를 놓으면 커서가 손 모양으로 바뀝니다. 축을 스크롤하거나 축의 척도를 조정하려면 이 커서를 클릭한 채로 드래그합니다. 자세한 내용은 "그래프의 축 스크롤 및 척도 조정"(454 페이지)에서 확인하십시오.

팁 : " 일반 " 환경 설정에서 " 자동 저장 타임아웃 " 값을 설정하여 열려 있는 보고서를 지정된 시간 (분) 마다 자동으로 저장하는 것이 좋습니다 . 이 자동 저장 값은 데이터 테이블 , 저널 , 스크립트 , 프로젝트 및 보고서에도 적용됩니다 .

손 도구 사용

"도구" > "손 도구" 옵션을 사용하여 손 도구를 선택할 수 있습니다. 보고서에는 손 도구로 사용할 수 있는 기능이 여러 가지 있습니다. 다음은 그래프 및 그림에서 손 도구가 작동하는 방식을 알 수 있는 몇 가지 예입니다.

- 히스토그램에 연속형 변수가 있는 경우 손 도구를 사용하여 막대의 수를 변경하거나 막대의 경계를 이동할 수 있습니다 .
- 모든 보고서 테이블에서 손 도구로 열을 클릭한 채 드래그하여 재배열할 수 있습니다 .
- 손 도구를 사용하여 축 값의 표시 범위를 변경할 수 있습니다 . 자세한 내용은 " 그래프의 축 스크롤 및 척도 조정 "(454 페이지) 에서 확인하십시오 .

JMP 보고서의 표시 옵션

표시 아이콘 을 마우스 오른쪽 버튼으로 클릭하면 보고서를 재배열하거나 보고서 개요 수준을 관리할 수 있는 메뉴가 표시됩니다. 표시되는 메뉴에는 다음과 같은 보고서 형식 지정 옵션이 포함되어 있습니다.

닫기 보고서의 해당 섹션을 닫습니다 . 즉 , 숨깁니다 . 표시 아이콘을 클릭하여 이 작업을 수행할 수도 있습니다 .

수평 가능한 경우 보고서의 개요 방향을 수직과 수평 사이에서 전환합니다 .

아래 항목 모두 열기 이 명령이 선택된 수준과 그 아래의 모든 개요 수준을 엽니다 .

아래 항목 모두 닫기 이 명령이 선택된 수준과 그 아래의 모든 개요 수준을 닫습니다 .

유사 항목 모두 열기 분석 창에서 현재 항목과 동일한 유형의 보고서를 모두 엽니다 . 한 번에 여러 개의 변수를 분석할 경우 동일한 유형의 여러 보고서 테이블을 모두 한꺼번에 열 수 있습니다 . 단일 분석에서 여러 옵션을 선택할 경우에도 동일한 유형의 보고서 테이블을 모두 한꺼번에 열 수 있습니다 .

유사 항목 모두 닫기 분석 창에서 현재 항목과 동일한 유형의 보고서를 모두 닫습니다 .

개요 없는 항목 모두 닫기 보고서에서 하위 수준이 없는 모든 부분을 닫습니다 . 이 명령은 일반적으로 보고서 개요의 최상위 수준에서 사용됩니다. 이 명령을 사용하면 보고서의 내포 구조를 빠르게 단순화하여 볼 수 있습니다 .

개요 닫기 방향 표시 아이콘을 클릭하여 닫을 때의 개요 상자 방향을 지정합니다 . 사용 가능한 옵션은 **자동** , **수평** 또는 **수직**입니다 . 기본 방향은 **자동**입니다 .

항목 추가 보고서에 구조 항목을 추가할 수 있는 방법이 나열된 하위 메뉴를 표시합니다 . 해당 항목으로는 텍스트 , 개요 제목 표시줄 , 다른 JMP 파일 및 창에 대한 참조 , 열려 있는 모든 JMP 파일의 목록 , URL, 그리고 스크립트가 포함됩니다 .

텍스트 항목 추가 최대 6 개의 텍스트 항목을 추가할 수 있는 창을 엽니다 . 텍스트 항목을 글머리 기호와 함께 표시할지 여부와 처음에는 숨김 상태로 표시할지 여부를 선택할 수 있습니다 . **참고 :** 숨겨진 텍스트 상자를 정확히 클릭해야 텍스트 항목이 표시됩니다 .

개요 항목 추가 제목 있는 개요 상자를 추가할 수 있는 창을 엽니다 . **참고 :** 추가된 개요 상자에는 추가 메뉴 항목을 나열하는 빨간색 삼각형 메뉴가 포함됩니다 .

창 참조 추가 열려 있는 JMP 창 중에서 선택한 개요 상자에 링크로 추가할 창을 선택할 수 있는 창을 엽니다 .

파일 참조 추가 선택한 개요 상자에 링크로 추가할 파일을 선택할 수 있는 " 데이터 파일 열기 " 창을 표시합니다 .

파일 디렉터리 추가 선택한 개요 상자에 링크로 추가할 디렉터리를 선택할 수 있는 **디렉터리 선택** 창을 표시합니다 . **참고 :** 선택한 디렉터리의 각 파일에 대한 링크가 개요 상자에 추가됩니다 .

모든 열려 있는 파일 추가 JMP 에 열려 있는 각 파일에 대한 새 개요 상자가 하나씩 추가됩니다 . 여기에는 해당 파일에 대한 링크도 포함됩니다 .

URL 참조 추가 **링크 이름** 및 **URL** 을 추가할 수 있는 "URL 에 대한 링크 생성 " 창을 엽니다 . 해당 URL 참조 링크는 선택한 개요 상자에 추가됩니다 .

스크립트 버튼 추가 선택한 개요 상자에 명명된 링크를 추가할 수 있는 " 스크립트 버튼 추가 " 창을 엽니다 . 추가된 링크를 클릭하면 지정된 스크립트가 실행됩니다 .

화면 크기에 맞게 내부의 그래프 또는 보고서 너비에 맞게 개요 크기를 조정합니다 .

특성 표시 보고서에 있는 표시 상자의 형식을 수정할 수 있습니다 . 특성에는 눈금 선과 같은 그래프 축 요소가 포함됩니다 .

팁 : 선택 도구로 하나 이상의 표시 상자를 선택한 후 표시 상자를 마우스로 가리켜 팝업 특성 편집기를 여는 방식으로 , 보고서에서 일부 특성을 직접 수정할 수 있습니다 .

편집 사용되는 개요 수준의 모든 보고서에 영향을 주는 다음과 같은 옵션이 포함된 하위 메뉴를 표시합니다 .

선택 해당 개요 수준의 모든 보고서를 강조 표시합니다 .

선택 취소 해당 개요 수준의 선택된 모든 보고서를 선택 취소합니다 .

저널 해당 보고서를 편집하거나 여기에 다른 보고서를 추가할 수 있도록 " 저널 " 이라는 별도의 창에 보고서를 복사합니다 . 자세한 내용은 " 데이터 저장 및 공유 " 장의 " 저널 생성 "(565 페이지) 에서 확인하십시오 .

그림 복사 해당 보고서를 클립보드에 복사합니다 . 그런 다음 다른 응용 프로그램을 열어 복사 내용을 붙여넣을 수 있습니다 .

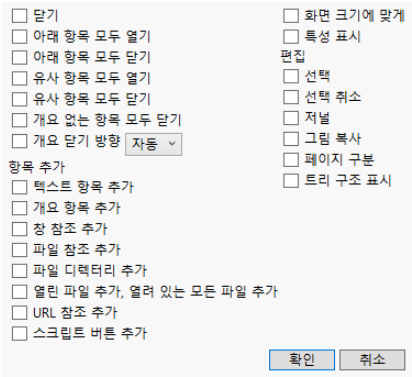
페이지 구분 인쇄용 페이지 구분을 삽입합니다.

트리 구조 표시 보고서를 구성하는 표시 상자를 보여 주는 창을 엽니다. 이 창은 주로 보고서 요소를 조작하거나 읽는 JSL 프로그래머가 사용합니다. 자세한 내용은 Scripting Guide의 에서 확인하십시오.

창 크기 복원 (창 크기가 원래 크기에서 수정된 경우에만 나타납니다.) 선택한 창을 원래 크기로 되돌립니다.

다른 방법으로 이러한 옵션에 액세스하려면 Alt 키를 누른 채 표시 아이콘 ◿ 을 마우스 오른쪽 버튼으로 클릭합니다. 그러면 그림 9.3에 표시된 것과 같이 여러 작업을 동시에 선택할 수 있도록 명령 및 옵션에 대한 체크박스가 포함된 창이 나타납니다. 빨간색 삼각형 메뉴 아래의 메뉴에도 동일한 기능을 사용할 수 있습니다.

그림 9.3 개요 상자 메뉴 항목 창



JMP 보고서 요소 표시 및 숨기기

JMP 보고서는 계층적 개요 구조로 구성되어 있습니다. 개요의 각 수준에는 삼각형 형태의 표시 아이콘 ◿ 이 있습니다. 표시 아이콘을 클릭하면 보고서의 해당 섹션이 열리거나 닫힙니다.

보고서에 표시되는 열을 변경할 수도 있습니다. 보고서를 마우스 오른쪽 버튼으로 클릭하고 "열" 목록에서 열을 선택하면 됩니다. 한 번에 여러 열을 선택하려면 Alt 키를 누른 채로 마우스 오른쪽 버튼을 클릭합니다.

여러 보고서 결합

여러 가지 분석을 수행하면서 모든 결과 및 데이터 테이블을 하나의 창에 표시하려는 경우가 있습니다. 여러 가지 방법으로 보고서 및 데이터 테이블을 선택하고 결합할 수 있습니다. 자세한 내용은 "대시보드를 사용하여 보고서 결합 및 표시" 장의 "예: 창을 결합하여 대시보드 생성"(590 페이지)에서 확인하십시오.

보고서의 제목 바꾸기

보고서에 표시되는 다음과 같은 제목을 바꾸려면 해당 제목을 두 번 클릭합니다 .

- 빨간색 삼각형 메뉴 옆의 제목
- 표시 아이콘 옆의 제목
- 열 제목

보고서의 글꼴 크기 늘리기

Windows 의 경우 , JMP 에서 보고서 및 데이터 테이블에 사용되는 글꼴 크기를 변경하려면 **창 > 글꼴 크기**를 선택합니다 . 그런 다음 하위 메뉴 항목 중 하나를 선택합니다 .

글꼴 크기 늘리기 글꼴 크기를 늘립니다 . 글꼴 크기를 한 번 더 늘리려면 이 항목을 다시 선택합니다 .

글꼴 크기 줄이기 글꼴 크기를 줄입니다 . 글꼴 크기를 한 번 더 줄이려면 이 항목을 다시 선택합니다 .

macOS 의 경우에는 **보기 > 텍스트를 크게** 또는 **보기 > 텍스트를 작게**를 선택합니다.

보고서 창 크기 조정

보고서 창은 창 내용에 맞춰 자동으로 크기가 조정됩니다. Windows 의 경우, 창 크기를 수동으로 조정한 후에 자동으로 조정할 수 있습니다. Ctrl 키를 누른 채 창의 오른쪽 하단을 클릭합니다.

JMP 보고서의 빨간색 삼각형 옵션

보고서에서 빨간색 삼각형 메뉴를 클릭하면 해당 특정 보고서에 적용되는 옵션 목록에 액세스할 수 있습니다 . 빨간색 삼각형 메뉴를 클릭하는 것 외에 다음 작업도 가능합니다 .

- 여러 작업을 동시에 선택할 수 있습니다 . Alt 키를 누른 채 빨간색 삼각형 메뉴를 클릭합니다 . 그러면 모든 명령 및 옵션이 체크박스와 함께 포함된 패널이 나타납니다 .
- 보고서 창의 모든 유사 보고서에 같은 명령을 적용할 수 있습니다 . Ctrl 키를 누른 채 빨간색 삼각형 메뉴를 클릭합니다 . 예를 들어 일원 분석에서 Ctrl 키를 누른 채 이 아이콘을 클릭하고 **평균 /ANOVA/ 합동 t** 검정을 선택하면 활성 보고서 창의 모든 일원 분석에 대해 분산 분석이 수행됩니다 .

" 분석 " 및 " 그래프 " 메뉴의 각 플랫폼에 적용되는 빨간색 삼각형 메뉴 옵션에 대해서는 다음 JMP 설명서에서 설명합니다 .

- 기본 분석
- Consumer Research
- Essential Graphing

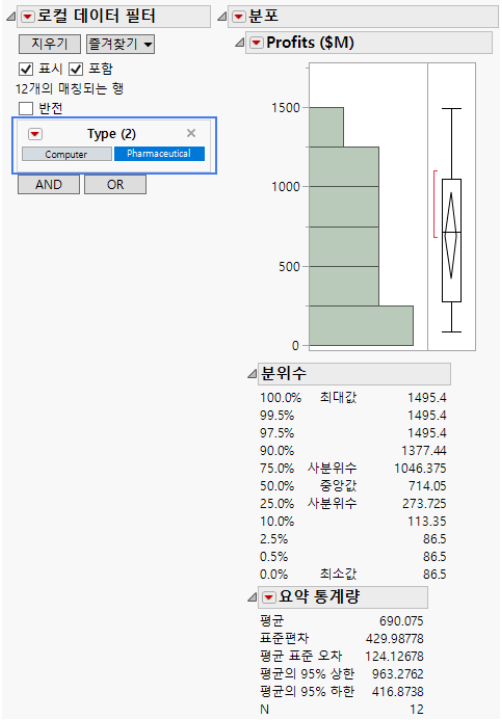
- 선형 모형 적합
- Multivariate Methods
- Profilers
- Quality and Process Methods
- Reliability and Survival Methods
- Predictive and Specialized Modeling

보고서의 로컬 데이터 필터

로컬 데이터 필터는 필터 범위를 현재 플랫폼의 출력으로 제한합니다. 예를 들어 컴퓨터 및 제약 업종 30 개 회사의 수익에 대한 데이터가 있는 경우 회사 유형별 수익을 검토하려면 분포를 실행한 후 로컬 데이터 필터를 사용하여 분포 플랫폼의 결과를 특정 회사 유형으로 좁히면 됩니다.

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석** > **분포**를 선택합니다 .
3. Profits ($M) 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 분포 " 의 빨간색 삼각형을 클릭하고 **로컬 데이터 필터**를 선택합니다 .
6. Type 을 클릭하고 **추가** ✚ 를 클릭합니다 .
7. "Type" 상자에서 **Pharmaceutical** 을 클릭하여 제약 회사의 수익만 살펴본 후 **Computer** 를 클릭하여 컴퓨터 회사의 수익만 살펴봅니다 .

그림 9.4 제약 회사의 수익을 보여 주는 로컬 데이터 필터



참고 : 변수 옆의 빨간색 삼각형 메뉴에는 선택을 반전하거나 체크박스 목록에 값을 표시하는 등의 옵션이 포함되어 있습니다. 자세한 내용은 " 데이터 필터의 변수에 대한 빨간색 삼각형 메뉴 옵션 "(429 페이지) 에서 확인하십시오 .

로컬 필터링

로컬 데이터 필터를 사용할 경우 해당 필터는 데이터 필터와 달리 플랫폼에서만 적용되고 데이터 테이블에는 적용되지 않습니다. 예를 들어 로컬 데이터 필터에서 행을 제외하더라도 데이터 테이블에서는 해당 행이 제외되지 않습니다. 또한 데이터 테이블에 행 상태가 포함된 경우 행 상태는 로컬 데이터 필터의 영향을 받지 않습니다.

필터 저장

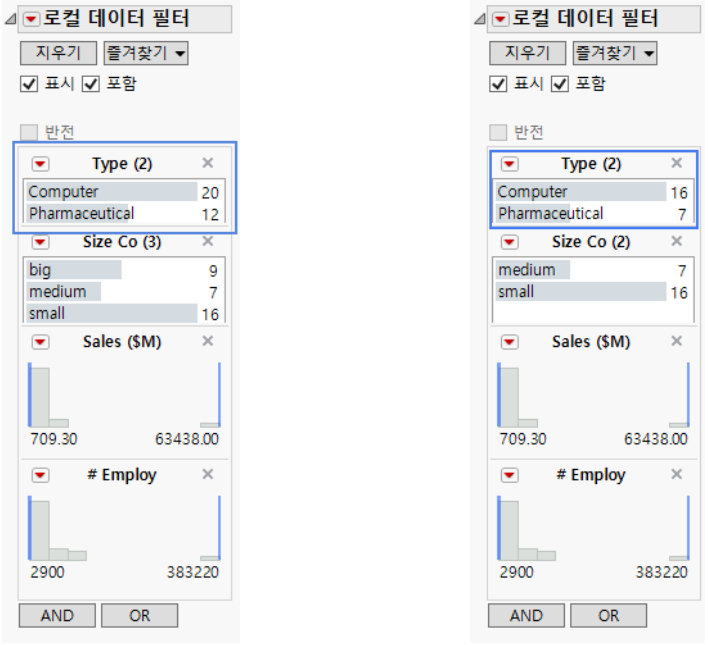
로컬 데이터 필터는 스크립팅이 가능하고 따라서 저장도 가능하므로 사용한 필터 설정에 빠르게 다시 액세스할 수 있습니다. 필터를 저장하려면 "로컬 데이터 필터" 옆의 빨간색 삼각형을 클릭하고 "스크립트 저장" 옵션 중 하나를 선택합니다.

제외된 행의 개수 포함

로컬 데이터 필터에는 기본적으로 각 변수 수준의 행 수(제외된 행 포함)가 표시됩니다. 제외된 행을 행 수에 포함하지 않으려면 "로컬 데이터 필터"의 빨간색 삼각형을 클릭하고 **제외된 행 수 포함**을 선택 취소합니다.

그림 9.5에서는 제외된 행을 행 수에 포함한 경우와 그렇지 않은 경우의 차이를 보여 줍니다. 제외된 행을 포함할 경우 로컬 데이터 필터는 값 "Computer"가 포함된 20개의 행을 표시합니다. 또한 9개의 "big" 값이 있습니다. 제외된 행을 포함하지 않을 경우에는 16개의 행에만 "Computer"가 포함된 것으로 나타납니다.

그림 9.5 제외된 행을 포함한 경우 (왼쪽) 와 포함하지 않은 경우 (오른쪽)



참고 : " 표시 옵션 " 이 " 블록 표시 " 로 설정되어 있으면 개수가 표시되지 않습니다 (그림 9.4(414 페이지)).

로컬 데이터 필터를 사용한 고급 스크립팅에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

기준 그룹 및 부분집합 보고서 사용

분석에서 기준 그룹을 사용하거나 동일한 테이블의 연결된 부분집합을 사용하는 여러 보고서를 결합하는 경우, 상위 테이블에 대한 로컬 데이터 필터를 다른 모든 기준 그룹 및 부분집합 보고서 간에 공유할 수 있습니다.

대시보드 및 응용 프로그램에서 로컬 데이터 필터 사용

대시보드 또는 응용 프로그램에 로컬 데이터 필터를 포함할 수 있습니다.

- 대시보드의 사용 예는 " 대시보드를 사용하여 보고서 결합 및 표시 " 장의 " 예 : JMP 쿼리 대시보드 및 추가기능 생성 "(593 페이지) 에서 확인하십시오 .
- 응용 프로그램의 사용 예는 Scripting Guide 의 에서 확인하십시오 .

시계열 데이터 확대

시계열 데이터를 필터링한 후 로컬 데이터 필터 히스토그램을 확대해서 더 작은 시간 증분을 확인할 수 있습니다. 예를 들어 히스토그램에 1년간의 데이터가 선택되어 있는 경우, 해당 기간 내의 더 작은 시간 증분을 보려면 변수 옆의 빨간색 삼각형 메뉴에서 **선택 영역 확대**를 선택합니다. 히스토그램을 재설정하려면 "로컬 데이터 필터"의 빨간색 삼각형 메뉴에서 **확대/축소 재설정**을 선택합니다.

보고서의 다시 실행 메뉴

분석을 수행한 후 다른 옵션 또는 변수를 사용하여 분석을 다시 실행할 수 있습니다 . 데이터 테이블의 데이터가 변경된 경우에도 분석을 다시 수행할 수 있습니다 . " 다시 실행 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

열 전환기 그래프의 열을 대화식으로 서로 교환할 수 있습니다 . 자세한 내용은 " 열 전환기 "(417 페이지) 에서 확인하십시오 .

분석 다시 실행 이전 시작 창 설정을 기반으로 중복된 분석을 수행합니다 . 분석 결과는 새 보고서 창에 나타납니다 . 이 옵션은 데이터가 변경된 경우에 유용합니다 .

분석 다시 시작 시작 창을 열고 보고서를 생성하는 데 사용되었던 설정을 재호출합니다 .

자동 재계산 데이터 테이블 값이 변경되면 분석과 그래픽을 자동으로 업데이트합니다 . 자세한 내용은 " 자동 재계산 "(417 페이지) 에서 확인하십시오 .

보고서의 스크립트 저장 메뉴

분석을 완료하고 보고서를 받은 후 해당 보고서를 생성하는 데 사용한 프로세스를 JSL 스크립트로 저장할 수 있습니다. 이렇게 하려면 스크립트의 빨간색 삼각형 메뉴에서 옵션을 선택합니다. 대부분의 "스크립트 저장" 옵션은 JMP 전체에서 동일합니다. 일부 플랫폼에서는 추가 옵션이 제공되며 이러한 옵션에 대해서는 플랫폼 관련 장에서 설명합니다. "스크립트 저장" 메뉴의 다음 옵션은 모든 플랫폼에서 공통적으로 사용됩니다.

클립보드에 스크립트를 컴퓨터의 클립보드에 저장합니다 . 그런 다음 스크립트를 다른 스크립트 또는 문서에 복사할 수 있습니다 .

데이터 테이블에 보고서를 생성하는 데 사용된 데이터 테이블에 스크립트를 저장합니다 . 그러면 데이터 테이블에서 스크립트를 다시 실행하여 결과를 다시 생성할 수 있습니다 . 데이터 테이블에 스크립트를 저장하려면 이 옵션을 선택합니다 .

저널에 분석을 위해 스크립트를 실행하는 저널 창에 링크를 저장합니다 . 저널 창이 이미 열려 있으면 해당 저널 창에 스크립트가 추가됩니다 . 스크립트에는 데이터 테이블의 경로가 포함됩니다 . 데이터 테이블을 찾을 수 없으면 스크립트가 실행되지 않습니다 .

스크립트 창에 스크립트 편집기 창을 열고 이 창에 스크립트를 추가합니다 . 스크립트 창이 이미 열려 있으면 이 열려 있는 창의 맨 아래에 스크립트가 추가됩니다 .

보고서에 보고서 창의 맨 위에 스크립트를 추가합니다 .

모든 기준 그룹 스크립팅

시작 창에서 "기준" 변수를 지정한 경우 "그룹별 스크립트 저장" 메뉴를 사용하여 기준 변수의 모든 수준에 대한 스크립트를 데이터 테이블 또는 스크립트 창에 저장할 수 있습니다.

"스크립트 저장" 메뉴에서 **데이터 테이블에(모든 개체)** 및 **스크립트 창에(모든 개체)**는 기준 변수의 모든 수준에 대한 스크립트를 저장합니다.

자동 재계산

자동 재계산은 데이터 테이블에서 변경한 사항을 해당하는 보고서 창에 즉시 반영하는 기능입니다. 데이터 테이블의 변경 사항에는 다음이 포함될 수 있습니다.

- 데이터 테이블 행 제외 또는 제외 해제
- 데이터 테이블 행 삭제 또는 추가

이 강력한 기능은 보고서 창에 있는 해당하는 분석, 통계 및 그래프에 변경 사항을 즉시 반영합니다.

보고서 창에 대해 자동 재계산 기능을 설정하려면 플랫폼의 빨간색 삼각형을 클릭하고 **다시 실행 > 자동 재계산**을 선택합니다. 이 기능을 해제하려면 동일한 옵션을 선택 취소합니다. JSL을 사용하여 자동 재계산을 설정할 수도 있습니다.

자동 재계산을 지원하는 플랫폼의 경우 "플랫폼" 환경 설정에 해당 환경 설정이 포함되어 있습니다.

참고 : 일부 플랫폼의 경우 자동 재계산 기능이 적합하지 않으므로 지원되지 않습니다 . 이러한 플랫폼으로는 DOE, 프로파일러 , 선택 , 파티션 , 비선형 , 신경망 , 신경망 분석 , 부분 최소 제곱 , 모형 적합 (REML, GLM, 로그 분산), 가우스 과정 , 항목 분석 , Cox 비례 위험 , 반응 변수 선별 , 관리도 (런 차트 제외) 가 포함됩니다 .

열 전환기

보고서 내에서 열 전환기를 사용하면 분석 결과를 다시 생성하지 않고도 다른 변수를 빠르게 분석할 수 있습니다. 열 전환기를 활성화하려면 보고서 창에서 빨간색 삼각형을 클릭하고 **다시 실행 > 열 전환기**를 선택합니다.

열이 여러 개이면 버튼을 사용하여 열 전환 시 애니메이션을 적용하거나 각 열을 수동으로 하나씩 전환하십시오. 애니메이션 속도를 변경하려면 슬라이더 컨트롤을 이동하십시오.

참고 : 보고서 내에서 열 전환기를 복사하거나 이동할 수는 없습니다 . 또한 열 전환기는 JMP 저널에 저장할 수 없습니다 .

예 : 열 전환기 사용

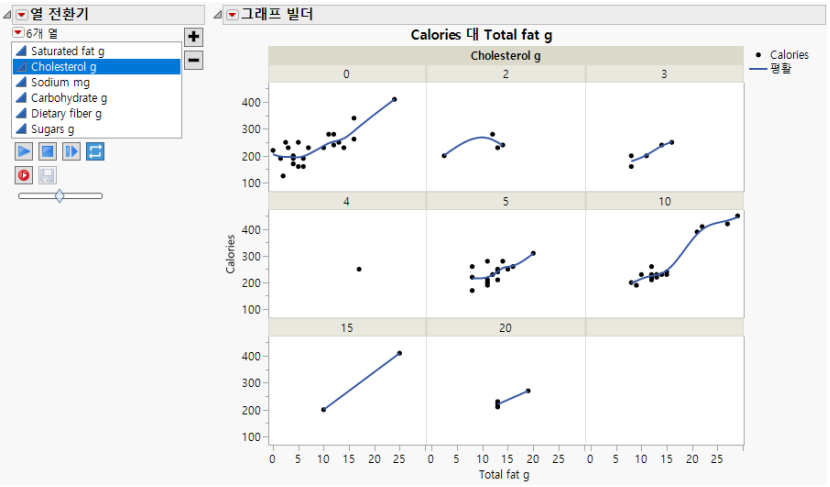
초코바의 영양 정보에 대한 데이터가 있다고 가정해 보겠습니다. 다양한 요인을 검토하여 칼로리 수준을 가장 잘 예측하는 요인을 확인하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Candy Bars.jmp 를 엽니다 .
2. **그래프 > 그래프 빌더**를 선택합니다 .
3. Total fat g 를 클릭하여 **X** 영역으로 드래그합니다 .
4. Calories 를 클릭하여 **Y** 영역으로 드래그합니다 .
5. Cholesterol g 를 클릭하여 **펼침** 영역으로 드래그합니다 .
6. " 그래프 빌더 " 의 빨간색 삼각형을 클릭하고 **제어판 표시**를 선택합니다 .
7. " 그래프 빌더 " 의 빨간색 삼각형을 클릭하고 **다시 실행 > 열 전환기**를 선택합니다 .
8. Cholesterol g 를 선택합니다 .

   그러면 전환할 초기 열이 선택됩니다 .
9. Saturated fat g, Cholesterol g, Sodium mg, Carbohydrate g, Dietary fiber g 및 Sugars g 를 선택합니다 .

   그러면 대체 열이 선택됩니다 .
10. **확인**을 클릭합니다 .

그림 9.6 그래프 빌더 창의 열 전환기



11. **재생** 버튼을 클릭하여 여러 요인을 차례로 순환하며 검토합니다 . 애니메이션의 속도를 제어하려면 슬라이더를 사용합니다. 또는 각 요인을 개별적으로 전환하며 검토할 수도 있습니다.

    열 목록의 마지막 열에 도달할 때 애니메이션의 동작을 지정하려면 "루프" 버튼을 클릭합니다.

각각의 탄수화물 수준에서 칼로리와 지방 함량의 연관성이 비교적 강하다는 것을 알 수 있습니다 . 따라서 Carbohydrate g 가 칼로리 수준에 대한 가장 적절한 예측 변수라고 할 수 있습니다 .

참고 :

- 열을 전환할 때 축 설정을 일정하게 유지하려면 열 전환기 빨간색 삼각형 메뉴에서 **축 설정 유지**를 선택합니다 . 예를 들어 참조선을 추가하여 축을 사용자 정의한 경우 유용합니다 .
- 보고서에 열 전환기가 여러 개 있는 경우 열 전환기 빨간색 삼각형 메뉴의 " 레이아웃 " 옵션을 사용하여 열 전환기의 레이아웃을 조정할 수 있습니다 .
- 새 열을 추가하려면 다음 옵션 중 하나를 수행하십시오 .
  - 열 전환기에서 더하기 버튼을 클릭하고 열을 선택한 다음 **확인**을 클릭합니다 .
  - 그래프 빌더에서 열 목록의 열을 열 전환기 목록으로 드래그합니다 .
  - 데이터 테이블에서 열 패널의 열을 열 전환기 목록으로 드래그합니다 .
- 열을 제거하려면 빼기 버튼을 클릭하고 제거할 열을 선택한 다음 **확인**을 클릭합니다 . 또는 열을 선택하고 Delete 키를 눌러도 됩니다 .
- 변경 사항을 되돌리려면 열 빨간색 삼각형 메뉴에서 **재설정**을 선택합니다 .
- 열 전환기의 기록 및 저장 버튼은 애니메이션을 기록한 후 애니메이션 GIF 로 저장합니다. 이 기능은 Windows 에서만 사용할 수 있습니다 .

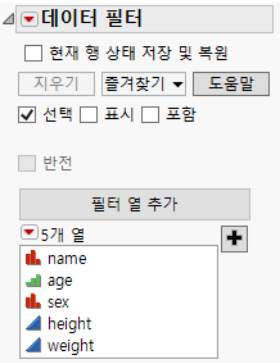
데이터 필터

JMP 의 데이터 필터는 데이터 부분집합을 확인하기 위한 다양한 방법을 제공합니다. 복잡한 데이터 부분집합을 대화식으로 선택하거나, 이러한 부분집합을 그림에서 숨기거나, 분석에서 제외할 수 있습니다.

1. **행** > **데이터 필터**를 선택합니다 .

팁 : 기본 데이터 필터 외에 플랫폼 보고서 내의 로컬 데이터 필터를 실행할 수도 있습니다 . 보고서 도구 모음에서 로컬 데이터 필터 아이콘 을 클릭하거나, 보고서의 빨간색 삼각형 메뉴에서 **로컬 데이터 필터**를 선택하십시오 . 자세한 내용은 " 보고서의 로컬 데이터 필터 "(413 페이지) 에서 확인하십시오 .

그림 9.7 초기 데이터 필터 창



2. 필터로 사용할 열을 선택한 후 **추가** ✚ 를 클릭합니다 .

참고 :

- 데이터 필터 창을 닫을 때 현재 행 상태를 복원하려면 **현재 행 상태 저장 및 복원** 옵션을 선택하십시오 .
- 열 목록이 긴 경우 목록의 열을 정렬하거나 , 표시하거나 , 숨기거나 , 검색할 수 있습니다 . " 필터 열 추가 " 의 빨간색 삼각형 메뉴에 포함된 옵션을 사용하십시오 .
- 기본적으로 데이터 필터 창은 데이터 테이블에 연결되어 있습니다 . 연결을 일시적으로 또는 영구적으로 해제할 수 있습니다 .
  - " 데이터 필터 " 의 빨간색 삼각형 메뉴에서 **부동 창 사용** 옵션을 선택 취소하여 연결을 일시적으로 해제합니다 .
  - **파일 > 환경 설정 > 데이터 필터**를 선택한 다음 **부동 창 사용**을 선택 취소하여 연결을 영구적으로 해제합니다 .

필터 열의 유형

네 가지 유형의 필터 열이 있습니다 .

연속형 열 모델링 유형이 연속형으로 설정된 숫자 열을 필터링합니다 . 연속형 필터 열의 경우 데이터 범위를 따라 히스토그램이 표시됩니다 . 히스토그램에서 필터링된 최소값 및 최대값을 나타내는 수직선을 조정할 수 있습니다 . 옵션에 대한 자세한 내용은 " 모든 유형의 변수에 대한 옵션 "(430 페이지) 및 " 연속형 변수에 대한 옵션 "(430 페이지) 에서 확인하십시오 .

범주형 열 명목형 및 순서형 열을 필터링합니다 . 데이터 필터는 각각의 범주형 열에 대해 일련의 개별 범주를 생성합니다 . 이러한 범주는 서로 다른 형식으로 표시될 수 있습니다 . 옵션에 대한 자세한 내용은 " 모든 유형의 변수에 대한 옵션 "(430 페이지) 및 " 명목형 또는 순서형 변수에 대한 옵션 "(430 페이지) 에서 확인하십시오 .

참고 : 값 라벨이 있는 범주형 열의 경우 데이터에 없는 반응을 포함하려면 **파일 > 환경 설정 > 플랫폼 > 범주형**에서 **데이터에 없는 반응 포함** 옵션을 선택하십시오 .

다중 반응 열 " 다중 반응 " 열 특성이나 " 다중 반응 " 모델링 유형이 할당된 문자 열입니다 . 이 열의 각 데이터 셀은 대개 쉼표 같은 일반적인 구분 기호로 구분된 여러 범주로 구성됩니다 . 다중 반응 열에서는 각 데이터 셀에 둘 이상의 범주가 포함될 수 있으므로 필터링 옵션이 보다 다양합니다 . 옵션에 대한 자세한 내용은 " 모든 유형의 변수에 대한 옵션 "(430 페이지) 및 " 다중 반응 모델링 유형 또는 열 특성이 설정된 변수에 대한 옵션 "(431 페이지) 에서 확인하십시오 .

비정형 텍스트 모델링 유형이 " 비정형 텍스트 " 로 설정된 텍스트 열을 필터링합니다 . 필터 텍스트 편집 상자에 단어나 구를 입력하여 필터 텍스트 목록에 추가합니다 . 그런 다음 이러한 단어와 구를 기준으로 행을 필터링할 수 있습니다 . 옵션에 대한 자세한 내용은 "모든 유형의 변수에 대한 옵션"(430페이지) 및 "비정형 텍스트에 대한 옵션"(431페이지)에서 확인하십시오.

필터링 모드

필터링 모드로는 **선택**, **표시** 및 **포함**이라는 세 가지 모드가 있습니다. 데이터 필터에서 해당하는 체크박스를 사용하여 이러한 모드를 설정하거나 해제할 수 있습니다.

선택 (로컬 데이터 필터에 사용할 수 없음) 데이터 테이블에서 선택한 행을 강조 표시된 상태로 표시합니다 .

데이터 필터 선택 모드 자동 선택 환경 설정을 사용하여 이 옵션의 자동 선택을 해제할 수 있습니다 . 자세한 내용은 " 데이터 필터 선택 후 데이터 테이블에서 행 상태 변경 "(422 페이지) 에서 확인하십시오 .

표시 선택되지 않은 행을 숨김 아이콘 으로 표시합니다 .

데이터 필터 표시 모드 자동 선택 환경 설정을 사용하여 이 옵션의 자동 선택을 설정할 수 있습니다 . 행 상태에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 행 상태 열에 정보 저장 "(277 페이지) 에서 확인하십시오 .

포함 선택되지 않은 행을 제외 아이콘 으로 표시합니다 .

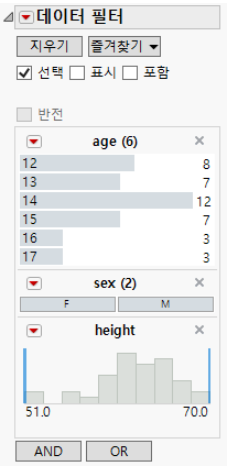
데이터 필터 포함 모드 자동 선택 환경 설정을 사용하여 이 옵션의 자동 선택을 설정할 수 있습니다 . 행 상태에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 행 상태 열에 정보 저장 "(277 페이지) 에서 확인하십시오 .

필터링 시 사용할 수 있는 두 가지 추가 옵션으로 **자동 지우기**와 **조건부**도 있습니다. 이러한 옵션은 "데이터 필터"의 빨간색 삼각형 메뉴에서 사용할 수 있습니다. 자세한 내용은 "데이터 필터의 빨간색 삼각형 메뉴 옵션"(424페이지)에서 확인하십시오.

데이터 필터 제어판

초기 창에서 열을 추가한 후에는 " 데이터 필터 " 제어판이 나타납니다 .

그림 9.8 데이터 필터 제어판



데이터 필터의 기본 컨트롤은 다음과 같습니다 .

지우기 " 데이터 필터 " 창에서 모든 변수 선택 상태를 지웁니다 .

즐겨찾기 현재 데이터 필터 기준을 즐겨찾기로 저장합니다 . 즐겨찾기를 생성한 후 이 즐겨찾기를 선택하면 현재 조건이 즐겨찾기의 기준으로 재설정됩니다 . 즐겨찾기를 제거할 수도 있습니다 . 현재 세션을 종료한 후에도 즐겨찾기를 유지하려면 " 데이터 필터 " 의 빨간색 삼각형 메뉴에서 " 스크립트 저장 " 옵션 중 하나를 선택하여 데이터 필터 스크립트를 저장합니다 .

선택 , 표시 및 포함 자세한 내용은 " 필터링 모드 "(421 페이지) 에서 확인하십시오 .

반전 현재 행 선택 상태를 데이터 테이블에서 반전합니다 .

> **참고 :** " 데이터 필터 " 에서의 선택 상태가 아니라 데이터 테이블의 행만 반전됩니다 . " 데이터 필터"에서 선택 상태를 반전하려면 열의 빨간색 삼각형 메뉴에서 **선택 반전**을 선택합니다.

AND "AND" 버튼은 **선택한 그룹에 추가** 목록을 엽니다 . and 연산자는 선택을 제한합니다 . 언제든지 필터 프로세스에 변수를 추가할 수 있습니다 .

OR "OR" 버튼은 **새 OR 그룹 추가** 목록을 엽니다 . or 연산자는 선택을 확장합니다 . 언제든지 필터 프로세스에 변수를 추가할 수 있습니다 .

데이터 필터 선택 후 데이터 테이블에서 행 상태 변경

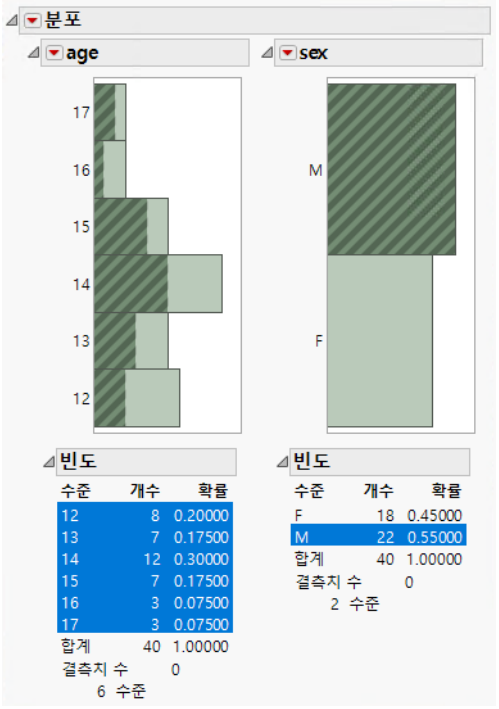
"데이터 필터"에서 필터를 선택하고 이후에 데이터 테이블에서 행 상태를 변경하거나 그래프 또는 그림에서 점을 선택하면 "데이터 필터"에서의 선택 상태와 데이터 테이블에서의 선택 상태가 매칭되지 않게 될 수 있습니다 . "데이터 필터"에는 "현재 선택 상태가 다른 창에서 변경되었습니

다."라는 경고 메시지가 표시됩니다. 또한 **선택 재설정** 버튼이 나타납니다. **선택 재설정** 버튼을 클릭하면 데이터 테이블의 선택 상태가 "데이터 필터"에서의 선택 상태를 반영하도록 다시 변경됩니다.

예 : 선택 상태 수정

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. age 및 sex 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. **행 > 데이터 필터**를 선택합니다 .
6. sex 를 선택하고 **추가**를 클릭합니다 .
7. " 데이터 필터 " 제어판에서 "M" 블록을 클릭하여 모든 남자를 선택합니다 .

그림 9.9 데이터 테이블 및 히스토그램에서 강조 표시된 남자 포함 행



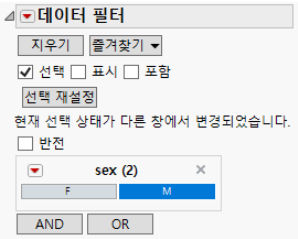
	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107
11	SUSAN	13	F	56	67
12	JOHN	13	M	65	98
13	JOE	13	M	63	105
14	MICHAEL	13	M	58	95
15	DAVID	13	M	59	79
16	JUDY	14	F	61	81
17	ELIZABETH	14	F	62	91
18	LESLIE	14	F	65	142
19	CAROL	14	F	63	84
20	PATTY	14	F	62	85
21	FREDERICK	14	M	63	93
22	ALFRED	14	M	64	99

데이터 테이블과 히스토그램에서 남자가 포함된 모든 행이 강조 표시됩니다 . 이제 12 세인 학생만 확인하기로 합니다 .

8. "age" 히스토그램에서 12 세를 나타내는 막대를 선택합니다 .

이제 선택 상태가 " 데이터 필터 " 에서의 선택 상태와 매칭되지 않습니다 . " 데이터 필터 " 창에 경고 메시지와 **선택 재설정** 버튼이 나타납니다 .

그림 9.10 데이터 필터 경고 메시지 및 재설정 버튼



데이터 필터의 빨간색 삼각형 메뉴 옵션

" 데이터 필터 " 옆의 빨간색 삼각형 메뉴에는 다음과 같은 옵션이 포함되어 있습니다 .

자동 지우기 " 데이터 필터 " 에 명목형 또는 순서형 열이 여러 개 선택된 경우 이 옵션을 사용하면 열을 새로 선택하기 전에 다른 선택 상태를 지울 수 있습니다 . 예를 들어 Big Class.jmp 를 사용하고 있고 데이터 필터에 sex(명목형) 및 age(순서형) 열이 포함되어 있다고 가정합니다 . "sex" 에서 "M"(남자) 을 선택하고 age 의 한 그룹 (예 : 12 세) 을 클릭하면 남자 행의 선택 상태가 자동으로 지워집니다 . 즉 , 12 세를 선택한 것이 남자를 선택하는 데 조건부로 적용되지 않습니다 . 반대로 , " 자동 지우기 " 를 해제하면 남자이면서 동시에 12 세인 행을 선택할 수 있습니다 . " 자동 지우기 " 는 기본적으로 해제되어 있습니다 .

조건부 선택된 필터 열에 대해 표시되는 범주를 제한합니다 . 자세한 내용은 " 조건부 데이터 필터 "(425 페이지) 에서 확인하십시오 .

AND 로 그룹화 열 필터의 여러 그룹에 대한 AND 및 OR 동작을 제어할 수 있습니다 . 자세한 내용은 " 데이터 필터의 AND 로 그룹화 옵션 "(426 페이지) 에서 확인하십시오 .

히스토그램 및 막대 표시 데이터 필터에 히스토그램과 막대를 표시하거나 숨깁니다 .

개수 표시 범주형 열 데이터 필터에서 개수를 표시하거나 숨깁니다 .

부동 창 사용 " 데이터 필터 " 창을 관련 데이터 테이블의 위에 유지합니다 . " 데이터 필터 " 창을 맨 위에 유지하지 않으려면 이 옵션을 선택 취소합니다 .

다시 시작 현재 " 데이터 필터 " 세션을 닫고 원래 " 데이터 필터 " 창을 표시합니다 .

제어판 표시 " 데이터 필터 " 창에서 " 지우기 " 와 " 반전 " 을 제외한 모든 버튼과 체크박스를 표시하거나 숨깁니다 .

모드 표시 " 데이터 필터 " 창에서 필터링 모드 체크박스를 표시하거나 숨깁니다 .

애니메이션 데이터 테이블에서 단일 변수의 값을 순차적으로 강조 표시합니다 . 자세한 내용은 " 데이터 필터의 애니메이션 기능 "(427 페이지) 에서 확인하십시오 .

부분집합 다음 내용만 포함하는 새 데이터 테이블을 생성합니다 .

– 데이터 필터로 지정된 행

– 활성 데이터 테이블에서 선택된 열 . 선택된 열이 없으면 모든 열이 포함됩니다 .

이 옵션은 " 테이블 ">" 부분집합 " 명령과 유사하며 부분집합 추출 옵션이 없다는 점만 다릅니다 .

Where 절 저장 선택한 값에 따라 Where 절을 작성합니다 .

팁 : 계산식 열로를 선택하면 표시자 열이 생성됩니다 . 많은 플랫폼의 표시자 열을 그래프 빌더에서 기준 역할 , 그룹 또는 ID 등으로 사용할 수 있습니다 .

스크립트 저장 스크립트를 저장하기 위한 옵션을 제공합니다 . 자세한 내용은 " 보고서의 스크립트 저장 메뉴 "(416 페이지) 에서 확인하십시오 .

로컬 데이터 필터 복사 (로컬 데이터 필터에만 사용할 수 있음) 로컬 데이터 필터의 스크립트를 클립보드에 복사합니다 . 그런 다음 보고서의 빨간색 삼각형 메뉴에서 " 로컬 데이터 필터 붙여넣기 " 옵션을 사용하여 이 로컬 데이터 필터를 새 보고서에 적용할 수 있습니다 .

로컬 데이터 필터 제거 (로컬 데이터 필터에만 사용할 수 있음) 보고서에서 로컬 데이터 필터를 제거합니다 .

조건부 데이터 필터

계층이 있는 필터 열의 경우 " 조건부 " 옵션을 사용하여 열 목록에 표시되는 열을 필터링할 수 있습니다 . 예를 들어 선택한 지역의 주만 목록에 표시되도록 지역을 기준으로 필터링할 수 있습니다 .

참고 : OR 그룹이 여러 개인 경우에는 조건부 필터를 사용할 수 없습니다 .

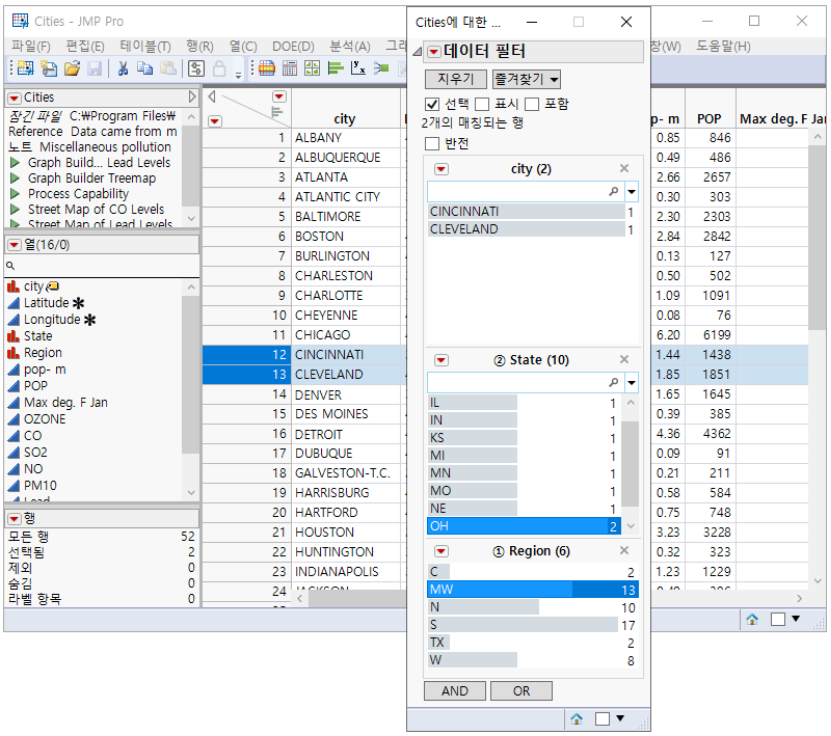
다음 예에서는 **조건부** 옵션을 사용하여 해당되지 않는 추가 범주를 제외하고 하위 범주를 간결하게 표시하는 방법을 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cities.jmp 를 엽니다 .
2. **행 > 데이터 필터**를 선택합니다 .
3. " 데이터 필터 " 창에서 city, State 및 Region 을 선택하고 **추가**를 클릭합니다 .

   " 데이터 필터 " 창이 나타나고 각 변수의 목록이 표시됩니다 .
4. " 데이터 필터 " 의 빨간색 삼각형을 클릭하고 **조건부**를 선택합니다 .
5. "Region" 목록에서 MW 를 선택합니다 .
6. "State" 목록에서 OH 를 선택합니다 .

   데이터 테이블에서 오하이오 (OH) 및 중서부 (MW) 지역의 도시가 선택됩니다 . " 데이터 필터 " 창에서는 "State" 목록에 중서부의 주만 표시되고 , "city" 목록에 오하이오 주의 도시만 표시됩니다 .

그림 9.11 조건부 옵션 사용



열 이름 앞에 있는 괄호 안의 숫자는 열 값이 선택된 순서를 나타냅니다. 그림 9.11에서는 Region이 첫 번째로 선택되었으므로 해당 열 이름 앞에 "1"이 표시됩니다. "State"는 두 번째로 선택되었으므로 해당 열 이름 앞에 "2"가 표시됩니다.

오름차순으로 표시된 계층적 필터를 재배열하면 계층에서의 오름차순 위치와 매칭되도록 필터 번호가 변경됩니다.

선택 상태를 지우고 계층 순서를 재설정하려면 **지우기**를 클릭합니다.

데이터 필터의 AND 로 그룹화 옵션

필터 열을 추가하고 "AND"를 클릭한 후 또 다른 필터 열을 추가하면 필터 그룹이 생성됩니다. 해당 그룹 내에서 모든 기준을 충족하는 열이 데이터 테이블에서 선택됩니다.

그런 다음 "OR"를 클릭하고 하나 이상의 필터를 추가하면 두 번째 필터 그룹이 생성됩니다. "AND로 그룹화"가 선택 취소되어 있는 경우 이것이 기본 동작입니다.

"AND로 그룹화"를 선택하면 이 동작이 반대가 됩니다.

다음 예에서는 "AND로 그룹화"가 선택되어 있을 때의 동작을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .

2. **행 > 데이터 필터**를 선택합니다 .
3. " 데이터 필터 " 의 빨간색 삼각형을 클릭하고 **AND 로 그룹화**를 선택합니다 .
4. " 데이터 필터 " 창에서 age 열을 선택하고 **추가**를 클릭합니다 .
5. **OR** 를 클릭하고 sex 열을 추가합니다 .

   sex 가 age 와 동일한 그룹에 추가됩니다 .
6. age 필터에서 **14** 를 선택하고 , sex 필터에서 **M** 을 선택합니다 .

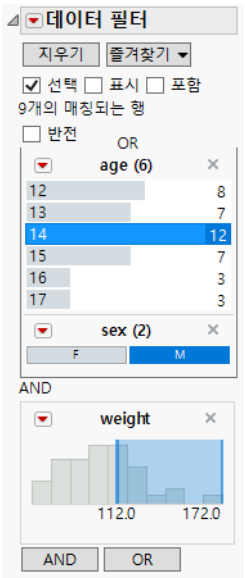
   age 와 sex 는 동일한 그룹에 포함되어 있습니다 . 데이터 테이블에서 age 또는 sex 조건을 충족하는 행이 선택됩니다. 즉, 데이터 테이블에서 모든 14세 학생과 모든 남학생이 선택됩니다.
7. **AND** 를 클릭하고 weight 열을 추가합니다 .

   새 그룹이 추가됩니다 .
8. 왼쪽 weight 슬라이더를 약 112 로 이동합니다 .

   또 다른 필터 그룹이 추가되었습니다 . 두 번째 필터 그룹은 체중이 112 를 초과하는 행으로 선택 범위를 제한합니다 .

   두 필터 그룹이 함께 적용되어 첫 번째 그룹의 조건과 두 번째 그룹의 조건을 모두 충족하는 행 , 즉 14 세이거나 남학생이되 체중이 112 를 초과하는 학생이 선택됩니다 .

그림 9.12 AND 로 그룹화 조건을 사용한 데이터 필터

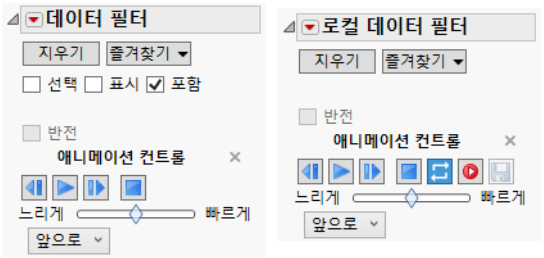


데이터 필터의 애니메이션 기능

애니메이션 기능은 단일 변수의 값을 순차적으로 강조 표시합니다. 즉, 데이터 테이블에서 해당 변수 값이 강조 표시됩니다. 하지만 먼저 그림을 생성한 다음 데이터 필터를 사용하여 변수에 애니메이션을 적용하면 그림에서 변수의 작용을 보다 간편하게 파악할 수 있습니다.

애니메이션 기능을 사용하려면 "데이터 필터"의 빨간색 삼각형을 클릭하고 **애니메이션**을 선택합니다. 그런 다음 애니메이션을 적용할 변수를 선택합니다. 변수 주위의 강조 표시된 프레임은 애니메이션을 적용하도록 선택된 변수를 나타냅니다.

그림 9.13 데이터 필터 및 로컬 데이터 필터의 애니메이션 컨트롤 패널



" 애니메이션 컨트롤 " 패널에는 다음과 같은 옵션이 포함되어 있습니다 .

- 가운데 버튼 ( ): 애니메이션을 시작하거나 일시 중지합니다 . 애니메이션 사이클을 시작한 후에는 버튼이 일시 중지 버튼 ( ) 으로 바뀝니다 . 기본적으로 애니메이션은 맨 위 변수의 첫 번째 값부터 시작합니다 .
- 뒤로 화살표 ( ): 애니메이션을 한 사이클 뒤로 이동합니다 . 여러 사이클 뒤로 이동하려면 이 버튼을 여러 번 클릭합니다 .
- 앞으로 화살표 ( ): 애니메이션을 한 사이클 앞으로 이동합니다 . 여러 사이클 앞으로 이동하려면 이 버튼을 여러 번 클릭합니다 .
- 사각형 버튼 ( ): " 데이터 필터 " 창에서 " 애니메이션 컨트롤 " 섹션을 숨깁니다 . " 애니메이션 컨트롤 " 을 표시하려면 " 데이터 필터 " 제목 표시줄의 메뉴에서 **애니메이션**을 선택합니다 .
- 루프 버튼 ( ) 은 루프를 실행하며 기본적으로 설정되어 있습니다 . 루프 버튼을 해제하면 애니메이션이 값 시퀀스를 한 번 순환한 다음 중지합니다 .
- 기록 버튼 ( ) 은 애니메이션 기록을 시작합니다 . 기록을 시작하면 버튼이 기록 중지 버튼 ( ) 으로 변경됩니다 .
- 저장 버튼 ( ) 은 기록된 애니메이션을 저장합니다 .

애니메이션 속도를 조정하려면 슬라이더를 사용합니다(느린 속도에서 빠른 속도로 전환).

애니메이션 드롭다운 메뉴에는 다음과 같은 옵션이 포함되어 있습니다 .

앞으로 첫 번째 값부터 마지막 값의 순서로 강조 표시합니다 .

뒤로 마지막 값부터 첫 번째 값의 순서로 강조 표시합니다 .

바운스 앞에서부터 강조 표시한 후 뒤에서부터 강조 표시하기를 반복합니다 .

데이터 필터에서 WHERE 절 저장

데이터 필터에서 변수 값을 필터링한 후 해당 정보를 JMP WHERE 절로 표현할 수 있습니다. WHERE 절은 JSL(JMP 스크립트 언어) 프로그램에서 처리 또는 분석을 위해 특정 데이터 행을 식별하는 데 사용됩니다. 데이터 필터는 사용자가 선택한 값에 따라 WHERE 절을 작성합니다.

"Where 절 저장" 메뉴에는 다음과 같은 옵션이 포함되어 있습니다.

클립보드에 필터 기준을 토대로 WHERE 절을 생성하여 클립보드에 추가합니다.

행 상태 열에 데이터 테이블에 필터 기준과 동등한 계산식이 있는 행 상태 열을 생성합니다.

계산식 열로 데이터 테이블에 필터 기준과 동등한 계산식이 있는 계산식 열을 생성합니다.

데이터 테이블에 필터 기준을 토대로 WHERE 절을 생성하여 **필터**라는 테이블 특성에 JSL 명령으로 포함한 후 현재 데이터 테이블과 함께 저장합니다.

스크립트 창에 필터 기준을 토대로 WHERE 절을 생성하여 현재 스크립트 텍스트 창에 추가합니다. 스크립트가 아직 없는 경우에는 새 스크립트를 생성합니다.

저널에 필터 기준을 토대로 WHERE 절을 생성하여 현재 저널에 추가합니다. 저널이 아직 없는 경우에는 새 저널을 생성합니다.

예 : WHERE 절 저장

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **행 > 데이터 필터**를 선택합니다.
3. age, sex 및 height 를 선택하고 **추가**를 클릭합니다.

12 세 또는 14 세이고 키가 56 인치에서 60 인치 사이인 모든 여학생을 선택합니다.

4. "12" 및 "14" 블록을 클릭하고 "F" 블록을 클릭합니다.
5. "51" 을 클릭하고 "56" 을 입력합니다.
6. "70" 을 클릭하고 "60" 을 입력합니다.
7. "데이터 필터"의 빨간색 삼각형을 클릭하고 **Where 절 저장 > 스크립트 창에**에를 선택합니다.

이 예에서 생성된 WHERE 절이 스크립트 창에 나타납니다.

```
Select Where(
(:age == 12 | :age == 14) & :sex == "F" & (:height >= 56 & :height <= 60)
);
```

데이터 필터의 변수에 대한 빨간색 삼각형 메뉴 옵션

데이터 필터에 열을 추가한 후에는 빨간색 삼각형 메뉴가 나타납니다. 변수에 대한 빨간색 삼각형 메뉴의 일부 옵션은 변수 유형에 따라 달라집니다.

모든 유형의 변수에 대한 옵션

모든 유형의 변수 옆에 있는 빨간색 삼각형 메뉴에는 다음과 같은 일반적인 옵션이 포함되어 있습니다.

선택 해제 해당 변수에 대해서만 현재 선택 상태를 해제합니다.

선택 반전 해당 변수에 대해서만 선택된 값은 선택 취소하고, 이전에 선택되지 않았던 모든 값은 선택합니다.

모델링 유형 열 필터에 대한 모델링 유형의 하위 메뉴를 표시합니다. 기본적으로 열 필터 유형은 열의 모델링 유형으로 설정됩니다. 하위 메뉴에서 다른 모델링 유형을 선택하여 필터 유형을 변경할 수 있습니다.

연속형 변수에 대한 옵션

연속형 변수의 경우 값은 최소값과 최대값에 수직선이 있는 히스토그램에 나타납니다. 수직선을 클릭하고 드래그하여 선택된 범위를 조정할 수 있습니다. 업데이트된 범위는 파란색으로 강조 표시됩니다.

다음 옵션은 연속형 변수에만 사용할 수 있습니다.

결측값 선택 결측값이 포함된 행을 선택합니다.

최소값 포함 선택한 범위에서 최소값을 포함하거나 제외합니다. 최소값을 제외하면 히스토그램의 수직 실선이 수직 파선이 됩니다.

최대값 포함 선택한 범위에서 최대값을 포함하거나 제외합니다. 최대값을 제외하면 히스토그램의 수직 실선이 수직 파선이 됩니다.

계급 수 설정 히스토그램의 계급 수를 설정합니다.

선택 영역 확대 히스토그램에 선택된 범위만 표시합니다. 이 메뉴 항목은 선택되면 "확대/축소 재설정"으로 변경됩니다. "확대/축소 재설정"은 표시를 재설정합니다.

명목형 또는 순서형 변수에 대한 옵션

명목형 및 순서형 변수의 경우 값이 블록, 목록 또는 메뉴에 나타납니다. 변수에 적은 수의 범주만 포함되어 있으면 값이 블록으로 나타납니다. 변수에 많은 수의 범주가 포함되어 있으면 값이 목록 또는 메뉴로 나타납니다. 하지만 이러한 기본 설정을 변경할 수 있습니다.

다음 옵션은 명목형 또는 순서형 변수에만 사용할 수 있습니다.

표시 옵션 표시 모양을 변경합니다. 다음과 같은 옵션이 포함될 수 있습니다.

블록 표시 각 수준을 블록으로 표시합니다.

목록 표시 각 수준을 목록 멤버로 표시하고 그 다음에 빈도를 표시합니다.

단일 범주 표시 각 수준을 메뉴에 표시하고 그 다음에 빈도를 표시합니다.

체크박스 표시 각 값 옆에 체크박스를 추가합니다 . 이 옵션을 기본 설정으로 만들려면 **파일 > 환경 설정 > 데이터 필터**에서 **체크박스 표시** 옵션을 선택합니다 .

라디오 상자 표시 각 값 옆에 라디오 상자를 추가합니다 .

개수별 정렬 값을 개수 기준으로 내림차순 정렬합니다 .

찾기 선택한 열에 대한 검색 문자열을 입력할 수 있는 텍스트 상자를 제공합니다 . 검색을 수행하려면 Enter 키를 누릅니다 .

다중 반응 모델링 유형 또는 열 특성이 설정된 변수에 대한 옵션

"다중 반응" 모델링 유형 또는 열 특성이 설정된 변수에는 다음 옵션을 사용할 수 있습니다.

표시 옵션 표시 모양을 변경합니다 . 다음과 같은 옵션이 포함될 수 있습니다 .

목록 표시 각 수준을 목록 멤버로 표시하고 그 다음에 빈도를 표시합니다 .

단일 범주 표시 각 수준을 메뉴에 표시하고 그 다음에 빈도를 표시합니다 .

체크박스 표시 각 값 옆에 체크박스를 추가합니다 . 이 옵션을 기본 설정으로 만들려면 파일 > 환경 설정 > 데이터 필터에서 **체크박스 표시** 옵션을 선택합니다 .

라디오 상자 표시 각 값 옆에 라디오 상자를 추가합니다 .

어떤 것과도 매칭되지 않음 선택 항목 중 어떤 것과도 매칭되지 않는 값이 포함된 행만 선택합니다.

임의 매칭 선택한 값 중 임의의 값과 매칭되는 값이 포함된 모든 행을 선택합니다 . 기본적으로 이 옵션은 선택되어 있습니다 .

모두 매칭 선택한 모든 값이 포함된 값이 있는 행만 선택합니다 .

정확하게 매칭 선택한 값과 정확히 매칭되는 값이 포함된 행만 선택합니다 .

매칭만 선택한 값과 정확히 매칭되는 내용이 있는 행만 선택합니다 .

매칭 최소 개수 선택한 값 중 *n* 개 이상을 선택합니다 .

매칭 최대 개수 선택한 값 중 최대 *n* 개를 선택합니다 .

매칭 개수 범위 선택한 값의 *n* 과 *m* 사이를 선택합니다 .

찾기 선택한 열에 대한 검색 문자열을 입력할 수 있는 텍스트 상자를 제공합니다 . 검색을 수행하려면 Enter 키를 누릅니다 .

참고 : " 다중 반응 " 특성에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 다중 반응 "(294 페이지) 에서 확인하십시오 .

비정형 텍스트에 대한 옵션

다음 옵션은 "비정형 텍스트" 모델링 유형의 변수에 사용할 수 있습니다.

대 / 소문자 구분 이 체크박스를 선택하면 필터 텍스트 편집 상자에 입력한 단어 또는 구에 기반하는 필터가 대 / 소문자를 구분합니다 .

정규 표현식 필터 텍스트 편집 상자에 입력된 문자열이 리터럴 문자열이 아닌 정규 표현식이라고 가정합니다.

표시 옵션 표시 모양을 변경합니다. 다음과 같은 옵션이 포함될 수 있습니다.

목록 표시 각 단어 또는 구를 목록 멤버로 표시하고 그 다음에 빈도를 표시합니다.

단일 범주 표시 메뉴에 각 단어 또는 구를 표시하고 그 다음에 빈도를 표시합니다.

체크박스 표시 각 단어 또는 구 옆에 체크박스를 추가합니다. 이 옵션을 기본 설정으로 만들려면 파일 > 환경 설정 > 데이터 필터에서 **체크박스 표시** 옵션을 선택합니다.

라디오 상자 표시 각 단어 또는 구 옆에 라디오 상자를 추가합니다.

결측 추가 텍스트가 포함되지 않은 항목에 대한 필터를 추가합니다.

필터 텍스트 목록 지우기 필터 텍스트 목록에 추가된 모든 단어 또는 구를 제거합니다.

필터 텍스트 편집 상자 표시 필터 텍스트 편집 상자를 표시하거나 숨깁니다.

개수별 정렬 개수를 기준으로 내림차순으로 단어 또는 구를 정렬합니다.

어떤 것과도 매칭되지 않음 선택된 어떤 단어 또는 구와도 매칭되지 않는 텍스트가 포함된 행만 선택합니다.

임의 매칭 선택된 단어 또는 구와 매칭되는 텍스트가 포함된 모든 행을 선택합니다. 기본적으로 이 옵션은 선택되어 있습니다.

모두 매칭 선택된 모든 단어 또는 구를 포함하는 텍스트가 있는 행만 선택합니다.

보고서의 테이블 형식 조정

보고서 테이블의 형식을 변경하는 방법에는 여러 가지가 있습니다. 보고서 테이블을 마우스 오른쪽 버튼으로 클릭하면 다음과 같은 형식 옵션에 액세스할 수 있습니다.

테이블 스타일 테이블에 테두리, 음영 및 구분선을 추가합니다. 다음 옵션 중에서 선택합니다.

머리글에 밑줄 적용 테이블 머리글에 밑줄을 적용하기 위한 환경 설정을 표시합니다.

머리글에 음영 적용 테이블 머리글에 음영을 적용하기 위한 환경 설정을 표시합니다.

머리글 열 테두리 머리글의 열 사이에 구분선을 표시합니다.

열 테두리 열 바깥쪽 테두리와 열 사이의 구분선을 포함합니다.

행 테두리 행 바깥쪽 테두리와 행 사이의 구분선을 포함합니다.

행에 교대로 음영 적용 행에 교대로 음영을 적용합니다.

셀에 음영 적용 테이블 본문에 음영을 적용합니다. " 행에 교대로 음영 적용 " 과 함께 사용할 경우 한 행씩 번갈아서 더 어두운 음영이 사용됩니다.

참고: 모든 보고서 테이블의 형식을 변경하려면 JMP 환경 설정에서 위의 옵션을 선택하십시오. Windows 의 경우, 보고서 테이블 옵션은 " 파일 " > " 환경 설정 " > " 스타일 " 에 있습니다. macOS 의 경우, **JMP > 환경 설정 > 스타일**에서 보고서 테이블 옵션을 선택하십시오.

열 테이블에 열을 표시하거나 숨깁니다.

참고 : ~ 편향과 같이 이름이 물결표 (~) 로 시작하는 열은 이름 옆의 체크박스를 선택하더라도 사용자가 실행하는 분석에 적용되지 않으며 테이블에도 나타나지 않습니다.

열별 정렬 선택한 열을 기준으로 열을 내림차순 또는 오름차순 정렬합니다. " 수치 순서 " 를 선택하면 정렬된 문자열 열을 수치로 정렬할 수도 있습니다. 이렇게 하면 열을 마우스 오른쪽 버튼으로 클릭할 때 " 수치 순서 " 옵션을 사용할 수 있습니다. " 수치 순서 " 옵션에서 사용하는 정렬 규칙에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 수치 순서 "(286 페이지) 에서 확인하십시오.

데이터 테이블로 만들기 보고서 테이블을 사용하여 JMP 데이터 테이블을 생성합니다. 데이터 테이블 스크립트는 보고서와 " 데이터 테이블로 만들기 " 를 통해 생성된 데이터 테이블을 다시 생성합니다.

결합 데이터 테이블 생성 보고서에서 선택한 테이블과 유사한 다른 테이블을 검색하여 단일 데이터 테이블에 결합합니다. 데이터 테이블에는 원래 데이터 테이블을 식별하는 열이 있습니다. 데이터 테이블 스크립트는 보고서와 " 결합 데이터 테이블 생성 " 을 통해 생성된 데이터 테이블을 다시 생성합니다.

행렬로 만들기 보고서 테이블을 사용하여 JMP 행렬을 생성합니다. 자세한 내용은 "JMP 보고서의 테이블을 행렬로 변환 "(435 페이지) 에서 확인하십시오.

선택 조건 (범주형 열에만 사용 가능) 선택된 열 또는 복합 표현식을 사용하여 Where 절을 생성합니다.

특성 표시 보고서에 있는 표시 상자의 형식을 수정할 수 있습니다.

열 형식 지정 테이블에 있는 하나 이상의 열에 사용되는 숫자 형식을 변경합니다.

열 복사 마우스 오른쪽 버튼으로 클릭한 열의 내용을 다른 창 또는 응용 프로그램에 붙여넣기 위해 클립보드에 복사합니다.

테이블 복사 마우스 오른쪽 버튼으로 클릭한 테이블을 다른 창 또는 응용 프로그램에 붙여넣기 위해 클립보드에 복사합니다.

시뮬레이션 보고서에서 선택된 통계량 열에 대해 시뮬레이션된 결과를 제공합니다.

JMP PRO **붓스트랩** 통계량의 표본 분포에 근사한 값을 산출합니다. 자세한 내용은 **기본 분석**의 에서 확인하십시오.

JMP 보고서의 표시 상자 특성 수정

보고서에 있는 각 표시 상자에 대해 글꼴, 텍스트 색상, 여백 및 맞춤과 같은 특성을 변경할 수 있습니다. 보고서를 스크립트, 그래픽 또는 저널로 저장할 때 변경 사항도 포함됩니다.

표시 상자 특성을 수정하려면 다음 단계를 수행하십시오 .

1. 보고서 옆의 표시 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **특성 표시**를 선택합니다 .
   " 특성 " 페인이 나타납니다 .
2. 보고서 특성을 수정합니다 .
3. 보고서를 클릭하여 변경 사항을 확인합니다 .

팁 :

- " 특성 " 페인에서 표시 트리의 표시 상자 사이를 이동하려면 화살표 버튼을 사용하십시오 .
- " 특성 " 페인을 연 후 특정 보고서를 선택하고 해당 특성을 변경하십시오 .

숫자 열의 형식 수정

"특성 표시" 페인에서 선택한 숫자 열의 형식을 수정하는 등 보고서의 모양을 수정할 수 있습니다. 예를 들어 모든 숫자 열의 소수 자릿수를 2자리로 지정할 수 있습니다. 보고서를 스크립트, 그래픽 또는 저널로 저장할 때 변경 사항도 포함됩니다.

숫자 형식을 수정하려면 다음 단계를 수행하십시오 .

1. 보고서 옆의 표시 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **특성 표시**를 선택합니다 .
   " 특성 " 페인이 나타납니다 .
2. Shift 키를 누른 채로 형식을 변경할 열을 선택합니다 .
3. " 특성 " 페인의 " 형식 " 목록에서 " 형식 " 특성을 선택합니다 .

팁 :

- 특정 유형의 모든 표시 상자 (예 : 모든 NumberColBox) 를 선택하려면 보고서에서 해당 상자 중 하나를 선택하십시오 . 그런 다음 " 특성 " 페인의 " 선택 " 목록에서 **'NumberColBox' 유형의 상자**를 선택하십시오 .
- 특정 형식이 있는 모든 열을 선택하려면 보고서에서 해당 상자 중 하나를 선택하십시오 . 그런 다음 " 특성 " 페인의 " 선택 " 목록에서 **매칭되는 형식이 있는 열**을 선택하십시오 .

JMP 보고서의 테이블을 행렬로 변환

보고서 테이블을 사용하여 JMP 행렬을 생성할 수 있습니다. 예를 들어 JSL(JMP 스크립트 언어)로 작업하는 경우 JSL 변수에 저장된 보고서 테이블에 액세스할 수 있습니다. 또는 보고서 테이블의 값을 테이블 특성에 테이블 특성으로 저장하거나 JSL 할당 값으로 저장할 수 있습니다. 그러면 보고서 테이블의 값이 데이터 테이블 내에 저장되어 스크립트 또는 계산식 편집기를 통해 액세스할 수 있게 됩니다.

테이블을 행렬 형식의 전역 변수에 저장하거나, 테이블 특성에 저장하거나, 테이블 특성에 할당 값으로 저장하려면 다음 단계를 수행하십시오.

1. 보고서 테이블의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다.
2. **행렬로 만들기**를 선택합니다.
3. 나타나는 창에서 테이블을 저장할 방법을 지정합니다.
4. (선택 사항) "이름" 옆의 상자에 새 이름을 입력하여 변수 또는 특성의 이름을 지정합니다.

보고서에서 조건부 형식 사용

참고 : JMP 보고서에 조건부 형식이 나타나도록 하려면 "보고서" 환경 설정에서 **조건부 형식 표시**을 활성화해야 합니다.

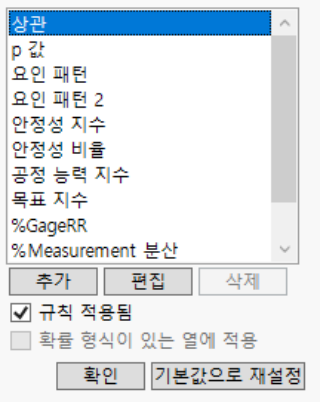
보고서에 조건부 형식이 사용되도록 구성하려면 먼저 "보고서" 환경 설정에서 조건부 형식 표시 기능을 활성화해야 합니다. 자세한 내용은 "JMP 환경 설정" 장의 "JMP 보고서의 환경 설정"(626 페이지)에서 확인하십시오.

조건부 형식을 구성하려면 다음 단계를 수행하십시오.

1. JMP "환경 설정" 창을 엽니다.
2. "보고서" 환경 설정 그룹을 선택합니다.
3. **규칙 관리**를 클릭합니다.

   "조건부 형식 규칙" 창이 나타납니다.

그림 9.14 조건부 형식 규칙 창



모든 규칙은 기본적으로 활성화되어 있습니다. p 값 규칙은 확률값이 포함된 열에 적용됩니다. 이 창에서 규칙을 추가, 편집 또는 삭제할 수 있습니다.

규칙 비활성화 또는 활성화

규칙을 비활성화 또는 활성화하려면

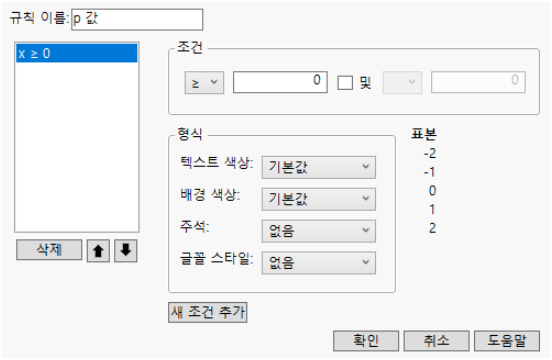
1. JMP " 환경 설정 " 창을 엽니다 .
2. " 보고서 " 환경 설정 그룹을 선택합니다 .
3. **규칙 관리**를 클릭합니다 .
   " 조건부 형식 규칙 " 창이 나타납니다 .
4. 목록에서 규칙을 선택합니다 .
5. **규칙 적용됨**이 선택되어 있는 경우 이를 비활성화하려면 클릭합니다 .
6. **규칙 적용됨**이 선택되어 있지 않은 경우 이를 활성화하려면 클릭합니다 .

규칙 추가

규칙을 추가하려면

1. JMP " 환경 설정 " 창을 엽니다 .
2. " 보고서 " 환경 설정 그룹을 선택합니다 .
3. **규칙 관리**를 클릭합니다 .
   " 조건부 형식 규칙 " 창이 나타납니다 .
4. **추가**를 클릭합니다 .
   " 조건부 형식 규칙 " 창이 나타납니다 . 왼쪽에 자리 표시자 규칙이 나타납니다 .

그림 9.15 조건부 형식 규칙 추가



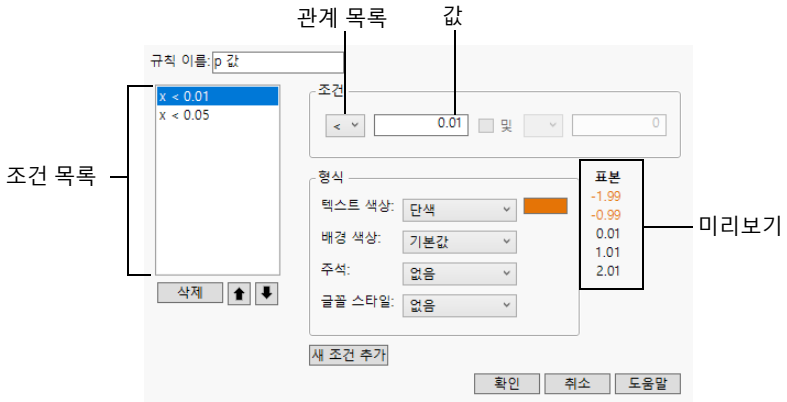
5. " 조건 추가 "(438 페이지) 에 설명된 대로 조건을 추가하고 형식을 지정합니다 .
6. **확인**을 클릭하여 새 규칙을 저장하고 이전의 " 조건부 형식 규칙 " 창으로 돌아갑니다 .
7. **규칙 적용됨**을 선택하여 규칙을 활성화합니다 .

규칙 편집

규칙의 조건부 형식을 변경하려면

1. " 조건부 형식 규칙 " 창의 목록에서 규칙을 선택합니다 .
2. **편집**을 클릭합니다 .
   선택한 규칙의 현재 조건을 보여 주는 " 조건부 형식 규칙 " 창이 나타납니다 .

그림 9.16 조건부 형식 규칙 창



p 값의 기본값에는 두 가지 조건이 포함되어 있습니다 . 목록의 순서는 규칙이 적용되는 순서를 나타냅니다 . 예를 들어 *p* 값이 0.04 이면 첫 번째 규칙 (x < 0.01) 은 호출되지 않지만 두 번째 규칙 (x < 0.05) 은 호출됩니다 .

참고 : 규칙 내의 기존 조건은 편집할 수 없습니다 . 기존 조건을 편집하려면 조건을 삭제한 후 다시 추가해야 합니다 . 자세한 내용은 " 기존 조건 편집 "(440 페이지) 에서 확인하십시오 .

조건 추가

규칙에 조건을 추가하려면

1. 규칙을 열어 " 조건부 형식 규칙 " 창을 표시합니다 .
2. " 조건 " 영역의 목록에서 관계를 선택하고 텍스트 상자에 값을 입력합니다 .

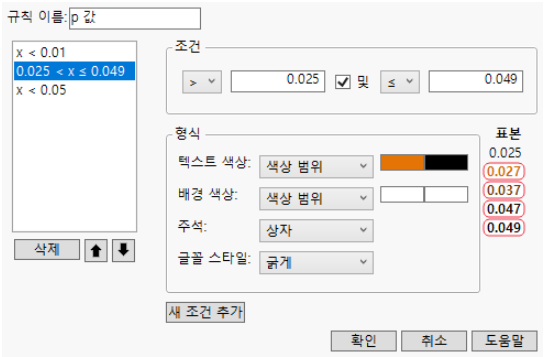
   또는

   조건에 값 범위를 포함하려면 다음을 수행합니다 .
   - 관계 드롭다운 목록에서 > 또는 ≥ 을 선택하고 텍스트 상자에 값을 입력합니다 .
   - 체크박스를 선택하고 < 또는 ≤ 관계를 선택한 후 값을 입력합니다 .
3. " 조건 형식 지정 "(439 페이지) 에 설명된 절차를 따라 조건의 형식을 지정합니다 .

팁 : **새 조건 추가**를 클릭하면 조건과 해당 형식 설정이 조건 목록의 맨 아래에 즉시 추가되고 " 조건 " 및 " 형식 " 영역이 기본 설정으로 돌아갑니다 . **새 조건 추가**를 클릭하기 전에 설정을 확인하십시오 .

4. 설정을 확인한 후 **새 조건 추가**를 클릭하여 해당 조건을 규칙에 추가합니다 .
5. 및 버튼을 사용하여 목록 내에서 해당 조건의 위치를 지정합니다 .

그림 9.17 예 : 형식 규칙과 함께 조건 추가



6. **확인**을 클릭하여 변경 사항을 저장하고 이전의 " 조건부 형식 규칙 " 창으로 돌아갑니다 .

조건 삭제

규칙에서 조건을 삭제하려면

1. " 조건부 형식 규칙 " 창에서 조건을 선택하고 **삭제**를 클릭합니다 .

2. **확인**을 클릭하여 변경 사항을 저장하고 이전의 " 조건부 형식 규칙 " 창으로 돌아갑니다 .
3. **확인**을 클릭하여 " 환경 설정 " 창으로 돌아갑니다 .

조건 형식 지정

조건을 충족하는 값의 형식을 지정하려면

1. 조건을 입력한 후 드롭다운 목록에서 " 텍스트 색상 " 을 선택합니다 .

   기본값 값이 색상 없이 표시됩니다 .

   단색 값이 지정된 색상으로 표시됩니다 .

   흐리게 값이 지정된 투명도만큼 흐리게 표시됩니다 .

   색상 범위 범위가 지정된 조건의 경우 , 값이 지정된 색상 범위 내에서 색상 그래디언트로 표시됩니다 .

   흐린 범위 범위가 지정된 조건의 경우 , 값이 지정된 투명도 범위만큼의 흐린 그래디언트로 표시됩니다 .
2. 텍스트 색상을 지정하려면 색상 견본을 클릭한 후 색상을 선택합니다 .
3. 투명도를 지정하려면 백분율 값 (예 : '60%') 을 입력합니다 .

   " 샘플 " 영역에 텍스트 열의 모양이 미리보기로 표시됩니다 .
4. 드롭다운 목록에서 배경 색상을 선택합니다 .

   기본값 값이 색상 배경 없이 표시됩니다 .

   단색 값이 지정된 색상 배경과 함께 표시됩니다 .

   색상 범위 범위가 지정된 조건의 경우, 값이 배경과 함께 지정된 색상 범위 내에서 색상 그래디언트로 표시됩니다 .
5. 배경 색상을 지정하려면 색상 견본을 클릭한 후 색상을 선택합니다 .

   " 샘플 " 영역에 텍스트 열의 모양이 미리보기로 표시됩니다 .
6. 조건 값에 노트를 추가하려면 드롭다운 목록에서 " 주석 " 을 선택합니다 .

   없음 값에 주석이 사용되지 않습니다 .

   상자 값이 테이블에서 상자로 둘러싸여 표시됩니다 .

   원 값이 테이블에서 타원으로 둘러싸여 표시됩니다 .

   ***(별표)** 테이블에서 값의 오른쪽에 별표 (*) 가 표시됩니다 .

   " 샘플 " 영역에 텍스트 열의 모양이 미리보기로 표시됩니다 .
7. 값 텍스트의 스타일을 지정하려면 드롭다운 목록에서 **글꼴 스타일**을 선택합니다 .

   없음 값 형식을 지정하는 데 스타일이 사용되지 않습니다 .

   일반 값 형식을 지정하는 데 스타일이 사용되지 않습니다 .

굵게 값이 테이블에서 굵게 표시됩니다 .

기울임꼴 값이 테이블에서 기울임꼴로 표시됩니다 .

굵은 기울임꼴 값이 테이블에서 굵은 기울임꼴로 표시됩니다 .

조건 형식의 예는 그림 9.17을 참조하십시오 .

참고 : 형식을 구성한 후에는 **새 조건 추가**를 클릭하여 규칙에 새 조건을 추가해야 합니다 .

기존 조건 편집

기존 조건의 형식을 변경하려면

1. " 조건부 형식 규칙 " 창에서 조건을 선택하고 **삭제**를 클릭합니다 .
2. " 조건 추가 "(438 페이지) 에 설명된 절차를 따라 조건을 다시 생성합니다 .
3. " 조건 형식 지정 "(439 페이지) 에 설명된 절차를 따라 조건의 형식을 지정합니다 .

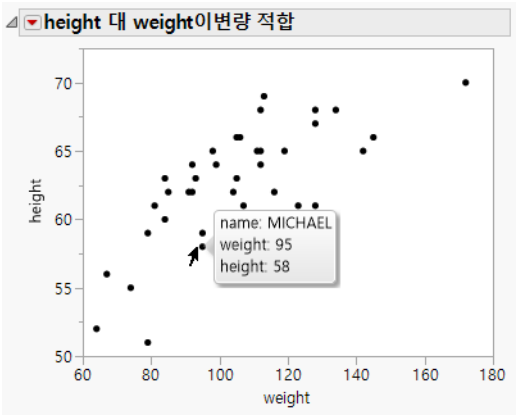
팁 : 새 조건 추가를 클릭하면 조건과 해당 형식 설정이 조건 목록의 맨 아래에 즉시 추가되고 " 조건 " 및 " 형식 " 영역이 기본 설정으로 돌아갑니다 . **새 조건 추가**를 클릭하기 전에 설정을 확인하십시오 .

4. 설정을 확인한 후 **새 조건 추가**를 클릭하여 해당 조건을 규칙에 추가합니다 .
5. 위로 및 아래로 버튼을 사용하여 목록 내에서 해당 조건의 위치를 지정합니다 .
6. **확인**을 클릭하여 변경 사항을 저장하고 이전의 " 조건부 형식 규칙 " 창으로 돌아갑니다 .

그래프에서 점 선택

그래프에서 점을 선택하려면 점을 클릭합니다. 그러면 점이 선택될 뿐만 아니라 현재 데이터 테이블에서 해당 행도 선택됩니다. 여러 개의 점을 선택하려면 Shift 키를 누른 채로 점을 선택합니다. 점 위에 커서를 놓으면 클릭 여부에 상관없이 점의 라벨이 표시됩니다.

그림 9.18 데이터 점의 라벨



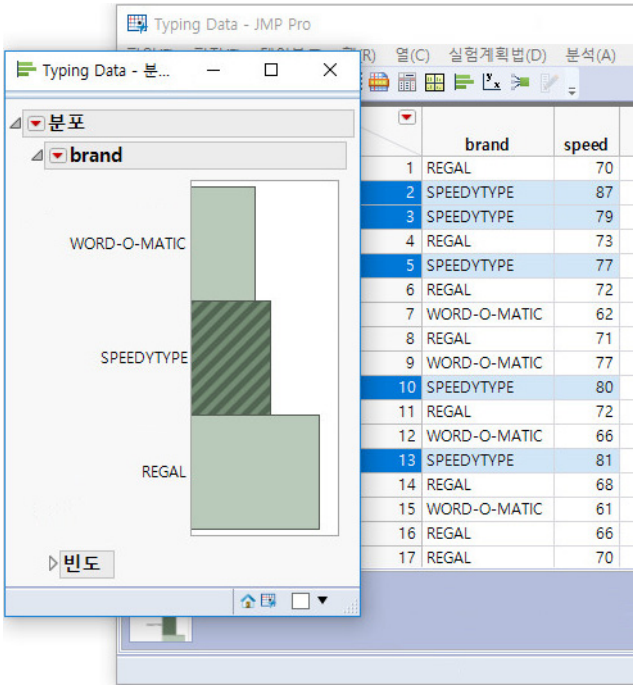
팁 :

- 라벨 내용을 다른 문서에 붙여넣을 수 있습니다 . 라벨을 마우스 오른쪽 버튼으로 클릭하고 **복사**를 선택합니다 . 그런 다음 텍스트를 다른 문서에 붙여넣습니다 .
- 표현식 역할 열에 사진 이미지를 저장한 경우 라벨에 사진을 표시할 수 있습니다 . 자세한 내용은 "JMP 열 특성 설정 " 장의 " 표현식 역할 "(301 페이지) 에서 확인하십시오 .

그래프에서 행 선택

동일한 데이터 테이블을 나타내는 모든 그래프 및 그림은 서로 연결되어 있으며 해당 데이터 테이블에도 연결되어 있습니다. 그림의 점이나 그래프의 막대를 클릭하면 데이터 테이블에서 해당 행이 강조 표시됩니다. 그림 9.19 의 예에서는 히스토그램에서 SPEEDYTYPE 막대가 강조 표시되고 테이블에서 해당 행이 강조 표시된 상태를 보여 줍니다. 히스토그램에서 Shift 키를 누른 채 막대를 선택하여 더 많은 막대를 선택할 수도 있습니다.

그림 9.19 히스토그램에서 행 강조 표시



JMP 그래프에서 직사각형 영역으로 점 선택

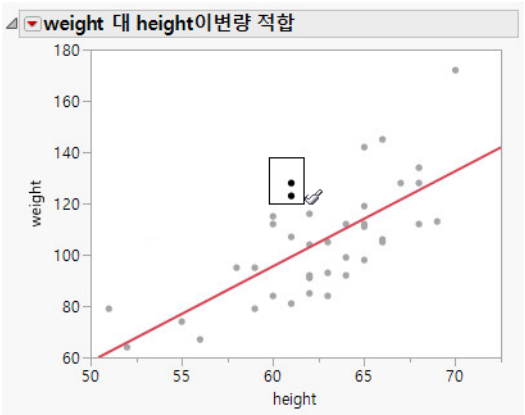
커서를 사용하여 직사각형 영역 안에 들어가는 모든 점을 선택할 수 있습니다. 화살표를 클릭한 채로 드래그하여 점을 강조 표시합니다. 또는 브러시 도구를 사용할 수도 있습니다. 그래프 위에서 브러시를 움직이면 직사각형 안에 들어가는 점이 선택됩니다. 데이터 테이블에서 숨김 상태로 표시된 점은 선택되지 않습니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 행 숨기기"(203 페이지)에서 확인하십시오.

브러시 도구를 사용하여 점을 선택하려면 다음 단계를 수행하십시오.

1. 도구 모음에서 브러시 도구 를 클릭합니다.
2. 그림에서 브러시 도구를 클릭한 채로 있습니다. 직사각형이 나타납니다.
3. 점 위에서 직사각형을 이동합니다. 직사각형이 점 위로 지나가면 해당 점이 그림과 활성 데이터 테이블 모두에서 강조 표시됩니다.

   브러시 도구를 점 위에서 이동할 때 모든 점을 선택된 상태로 유지하려면 Shift 키를 누른 채로 그림의 한 지점을 클릭합니다. 선택된 점은 데이터 테이블에서도 선택됩니다.

그림 9.20 브러시 도구 사용



4. 마우스를 놓습니다. 직사각형과 데이터 테이블 내의 해당 점이 선택된 상태로 유지됩니다.
   - 선택 직사각형의 크기를 변경하려면 Alt 키를 누른 채 그림 내부를 클릭합니다. 그런 다음 커서를 드래그하여 선택 상자의 크기를 조정합니다. 이 형태는 어느 한 축을 따라 이동하면서 여러 점을 포함하는 조각을 강조 표시할 수 있는 조각 도구처럼 작동합니다. **참고 :** 선택 상자의 크기는 저장되었다가 다음에 브러시 도구를 사용할 때도 적용됩니다.
   - 선택한 점에서 Ctrl 키를 누른 채 브러시 도구를 클릭하면 선택 직사각형 내의 점이 선택 취소됩니다. 선택 직사각형 외부의 점은 선택 상태로 유지됩니다.

JMP 그래프에서 불규칙한 형태의 영역 내에 있는 점 선택

올가미 도구를 사용하여 불규칙한 형태의 영역에 들어가는 점을 선택할 수 있습니다. 데이터 테이블에서 숨김 상태로 표시된 점은 선택되지 않습니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 행 숨기기"(203 페이지)에서 확인하십시오.

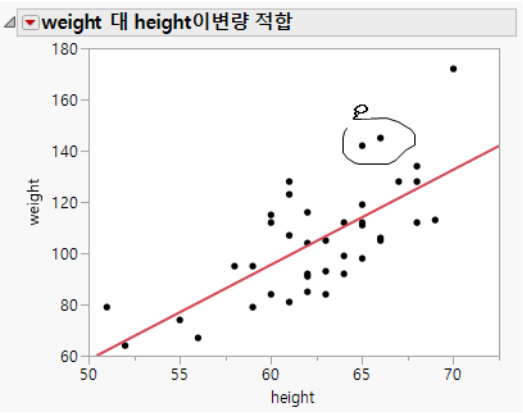
불규칙한 형태의 영역 내에 있는 점을 선택하려면 다음 단계를 수행하십시오.

1. 도구 모음에서 올가미 도구 를 클릭합니다.
2. 그림에서 올가미 도구를 클릭한 채로 있습니다.

   참고 : 여러 점의 주위를 올가미 도구로 드래그할 때 모든 점을 선택된 상태로 유지하려면 Shift 키를 누른 채로 그림 내부를 클릭합니다.

3. 여러 점의 주위를 올가미 도구로 드래그합니다.

그림 9.21 올가미 도구 사용



4. 마우스를 놓습니다. 올가미가 자동으로 닫히고 올가미로 둘러싼 영역 내의 점이 선택됩니다.

그래프에서 표식 사용

표식은 그래프에서 데이터를 나타내는 점입니다. 기본 설정에서 다른 설정으로 변경한 경우 데이터 테이블의 행 옆에도 해당 표식이 나타납니다.

- "표식 형태 변경"
- "표식 색상 변경"
- "표식 크기 변경"
- "많은 수의 표식 처리"
- "표식 주위에 외곽선 추가"
- "표식 선택 모드 변경"
- "표식 투명도 지정"
- "표식 제외 및 숨기기"
- "이미지를 표식으로 사용"
- "값을 기반으로 표식 형태 또는 색상 변경"

표식 형태 변경

JMP 표식 팔레트에서 문자를 할당하여 산점도의 표준 점을 바꿀 수 있습니다. 해당 표식은 데이터 테이블의 행 번호 옆에도 나타납니다.

1. 형태를 변경할 표식을 하나 이상 강조 표시합니다.
2. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다. 히스토그램에서는 오른쪽의 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭합니다.

3. **행 표식**을 선택합니다 .
4. 나타나는 옵션에서 표식 형태를 선택하거나 , **기타**를 클릭하여 표식으로 사용할 문자를 입력합니다 .

표식 색상 변경

강조 표시된 행에 색상을 할당할 수 있습니다. 이렇게 하면 산점도의 점이 색상 팔레트에서 선택한 색상으로 표시됩니다. 데이터 격자에서는 행에 할당된 활성 색상이 행 번호 옆에 표시됩니다.

그래프에서 표식 (또는 점) 의 색상을 변경하려면

1. 색상을 변경할 표식을 하나 이상 강조 표시합니다 .
2. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다 . 히스토그램에서는 오른쪽의 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭합니다 .
3. **행 색상**을 선택합니다 .
4. 색상 중 하나를 선택하거나 **기타**를 클릭하여 사용자 색상을 적용합니다 .

표식 크기 변경

그래프의 표식 크기를 늘리거나 줄이려면

1. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다. 명령을 동일한 창에 있는 동일한 유형의 모든 그림에 일괄적으로 적용하려면 Ctrl 키를 누른 채로 마우스 오른쪽 버튼을 클릭합니다. 히스토그램에서는 오른쪽의 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭합니다.
2. **표식 크기**를 선택합니다 .
3. 나열된 표식 크기 중 하나를 선택합니다 .

   기본값인 " 기본 크기 " 는 JMP 환경 설정의 " 그래프 " 페이지에서 선택한 크기입니다 .

참고 : 그리기 속도가 " 빠름 " 으로 설정되어 있으면 가장 작은 크기의 표식 (점) 이 적용됩니다 . 자세한 내용은 " 많은 수의 표식 처리 "(445 페이지) 에서 확인하십시오 .

많은 수의 표식 처리

그래프의 표식이 너무 많으면 그래프에서 표식이 조밀하게 나타날 수 있습니다. 이 경우 그래프를 보기 쉽게 투명도를 변경해야 할 수 있습니다. 투명도를 변경하면 JMP에서 보고서 창을 새로 고칠 때 사용되는 모드인 표식 그리기 모드에도 영향이 있을 수 있습니다. JMP에서는 그림에 표식을 그릴 때 보통 속도와 빠른 속도 중 하나가 사용됩니다.

표식 그리기 속도를 변경하려면

1. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다 . 히스토그램에서는 오른쪽의 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭합니다 .

2. **표식 그리기 모드**를 선택한 후 **보통** 또는 **빠름**을 선택합니다 .

정규 보통 그리기 모드에 있는 경우 그래프의 표식 수가 지정된 임계 개수보다 많으면 자동으로 빠름 모드로 전환됩니다 . 표식 임계를 설정하는 방법에 대한 자세한 내용은 "JMP 환경 설정 " 장의 "JMP 보고서의 환경 설정 "(626 페이지) 에서 확인하십시오 .

빠름 표식 그리기 속도를 " 빠름 " 으로 설정하면 표식 수가 많은 그래프가 보다 빠르게 표시됩니다 . 그리기 속도가 " 빠름 " 으로 설정되어 있으면 가장 작은 크기의 표식 (점) 이 적용되고 표식 투명도 설정은 기본값인 불투명으로 돌아갑니다 .

외곽선 표시 자세한 내용은 " 표식 주위에 외곽선 추가 "(446 페이지) 에서 확인하십시오 .

빈 표식을 채우려면

JMP 환경 설정에서 **그래프**를 선택한 후 **빈 표식 채우기**를 선택합니다. 그러면 모든 그래프의 빈 표식이 불투명한 배경으로 채워집니다.

표식 주위에 외곽선 추가

그림의 표식에 검은색 외곽선(또는 프레임)을 추가할 수 있습니다. 외곽선이 있는 표식은 표식 크기가 "중간", "큼", "XL", "XXL" 및 "XXXL"일 때 사용할 수 있습니다. 자세한 내용은 "표식 크기 변경"(445페이지)에서 확인하십시오.

외곽선을 추가하려면

1. 그림 또는 그래프를 마우스 오른쪽 버튼으로 클릭합니다 .
2. **표식 그리기 모드**를 선택합니다 .
3. **외곽선 표시**를 선택합니다 .

외곽선을 효과적으로 사용하려면 표식 색상이 검은색이 아닌 다른 색인 것이 좋습니다.

표식 색상을 변경하려면

1. 색상을 변경할 표식을 강조 표시합니다 .
2. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다 .
3. **행 색상**을 선택합니다 .
4. 나타나는 옵션에서 표식 색상을 선택합니다 .

표식 선택 모드 변경

그래프에서 표식을 선택하면 선택한 표식만 강조 표시됩니다. 현재 그래프에서 표식이 강조 표시되는 방식을 변경할 수 있습니다. 다음 그림의 맨 위에 있는 두 개의 삼각형에 이 옵션이 적용되어 있습니다.

그림 9.22 강조 표시된 삼각형 표식의 예

현재 그래프의 표식 강조 표시를 변경하려면 다음 단계를 수행하십시오 .

1. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭하고 **표식 선택 모드**를 선택합니다 .
2. 다음 옵션 중 하나를 선택합니다 .

기본 모드 JMP 환경 설정의 " 그래프 " 페이지에서 " 표식 선택 모드 " 로 선택한 모드입니다 . 기본값은 " 선택 취소 시 흐리게 " 입니다 .

선택 취소 시 흐리게 선택된 표식만 강조 표시합니다 . 그 외의 모든 표식은 " 환경 설정 " > " 그래프 " 에서 **선택 취소된 표식을 흐리게 나타내는 정도**로 지정한 백분율만큼 흐리게 표시합니다 .

선택 시 크게 선택된 표식을 선택되지 않은 표식보다 크게 표시합니다 .

선택 시 후광 표시 선택된 표식의 가장자리를 파란색 외곽선으로 표시합니다 .

선택 시 외곽선 표시 선택된 표식을 검은색 외곽선으로 표시합니다 .

선택 시 지정한 색상 적용 선택된 표식을 " 그래프 " 환경 설정에서 선택한 " 표식 선택 색상 " 으로 음영 처리합니다 .

표식 투명도 지정

그래프에 표시된 표식 (점) 의 투명도를 변경할 수 있습니다 . 예를 들어 이 기능으로 겹치는 점을 보기 쉽게 할 수 있습니다 .

참고 : 그리기 속도가 " 빠름 " 으로 설정되어 있는 경우에는 기본 표식 투명도가 적용됩니다 . 자세한 내용은 " 많은 수의 표식 처리 "(445 페이지) 에서 확인하십시오 .

1. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭하고 **투명도**를 선택합니다 . 히스토그램에서는 오른쪽의 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭하고 **투명도**를 선택합니다 .
2. 그래프의 표식 (점) 에 적용할 투명도 수준을 입력하고 **확인**을 클릭합니다 .

값 1 은 완전 불투명을 나타내고 0 은 표시되지 않음을 나타냅니다 . 1 과 0 사이의 값은 반투명을 나타냅니다 .

표식 제외 및 숨기기

"행">"숨기고 제외하기" 명령을 사용하여 강조 표시된 행이 표시되지 않고 통계 분석에서 제외되도록 할 수 있습니다. "행">"숨기고 제외하기"를 다시 선택할 때까지 해당 데이터는 숨기고 제외된 상태로 유지됩니다.

"제외/제외 해제" 명령을 사용하면 강조 표시된 행을 통계 분석에서 제외할 수 있습니다. 해당 행을 강조 표시하고 "행">"제외/제외 해제"를 다시 선택할 때까지 해당 데이터는 제외된 상태로 유지됩니다.

참고 : 제외된 데이터는 텍스트 보고서와 그래프의 계산에서는 제외되지만 그림에서 자동으로 숨겨지지는 않습니다 .

산점도에서 "숨기기/숨기기 해제" 명령을 사용하면 강조 표시된 점이 표시되지 않도록 숨길 수 있습니다. 예를 들어 점을 분석에서 제외한 후 동일한 점을 산점도에서 숨길 수 있습니다. 숨겨진 행을 강조 표시하고 "행">"숨기기/숨기기 해제"를 선택할 때까지 해당 데이터는 숨김 상태로 유지됩니다.

참고 : 숨겨진 점은 그림에 표시되지는 않지만 텍스트 보고서와 그래프에 영향을 주는 통계 계산에서 자동으로 제외되지는 않습니다 . 숨겨진 관측값을 분석에서 제외하려면 해당 관측값을 강조 표시하고 " 행 ">" 제외 / 제외 해제 " 특성을 선택해야 합니다 .

표식 (점) 을 분석에서 제외하거나 숨기려면 다음 단계를 수행하십시오 .

1. 제외하거나 숨길 표식을 강조 표시합니다 .
2. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다 .
3. **행 제외** 또는 **행 숨기기**를 선택합니다 .

이미지를 표식으로 사용

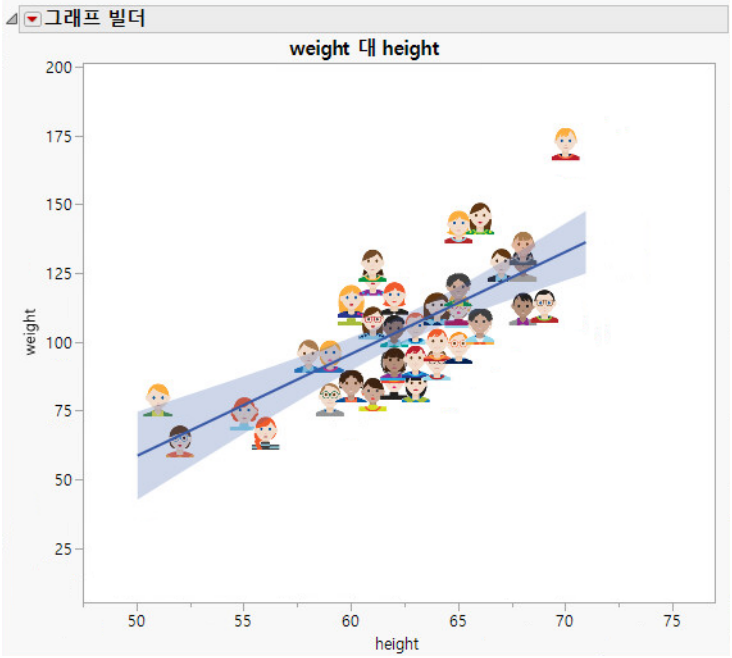
표현식 열에 이미지를 포함하고 이 이미지를 그래프에서 표식으로 사용할 수 있습니다. 그런 다음 표식에 사용할 열을 지정합니다. 이 기능을 사용하면 표식 색상을 보고 범례를 참조하여 표식의 의미를 확인할 필요가 없으므로 많은 수준이 포함된 데이터에서 특히 유용합니다.

문자 또는 숫자 열을 표식으로 사용하는 경우에는 표식에 데이터 값이 사용됩니다 .

1. 데이터 테이블에 새 열을 생성합니다 .
2. 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 후 **표현식** 데이터 유형을 선택합니다 .
3. **확인**을 클릭합니다 .

4. 컴퓨터에 있는 그래픽을 열의 각 셀로 드래그합니다 . 이미지 열이 포함된 데이터를 가져올 수도 있습니다 .
5. 열을 선택하고 **열** 메뉴를 선택한 후 **표식에 사용**을 선택합니다 .
6. 표식을 사용하는 그래프를 생성합니다 . 그림 9.23 는 그래프 빌더에서 생성되었습니다 .
7. (선택 사항) 표식 크기를 조정하려면 그래프를 마우스 오른쪽 버튼으로 클릭하고 **그래프 > 표식 크기**를 선택합니다 . 일부 플랫폼에서는 마우스 오른쪽 버튼을 클릭하고 **표식 크기**를 선택합니다 . 이 예에서는 " 중간 " 크기의 표식이 선택되어 있습니다 .
8. " 그래프 빌더 " 의 빨간색 삼각형을 클릭하고 **범례 표시**를 선택 취소합니다 .

그림 9.23 표식으로 사용된 그림



참고 :

- Big Class Families.jmp 샘플 데이터 테이블에서는 그래프의 표식에 picture 열이 사용되었습니다 .
- 버블 그림에서는 이 기능이 지원되지 않습니다 .

값을 기반으로 표식 형태 또는 색상 변경

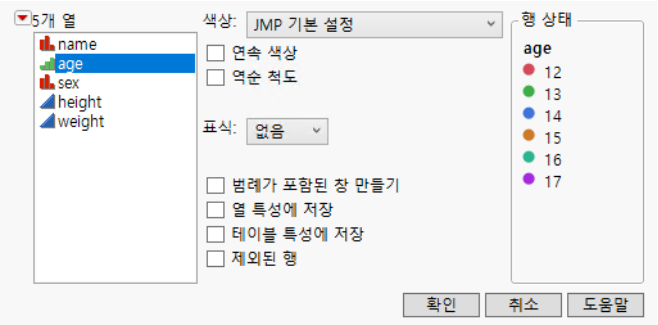
일부 그림에서는 **행 범례**를 추가하여 점의 값에 따라 표식 형태 또는 색상을 변경할 수 있습니다. 이를 행 범례라고 하는 것은 JMP가 행 색상 또는 행 표식 설정을 사용하여 범례를 자동으로 삽입

하기 때문입니다. 이 방법으로 표식 또는 색상을 할당하면 선택된 점에 관계없이 그래프의 모든 점에 특성이 할당됩니다. 이전의 모든 표식 및 색상 설정은 무시됩니다.

열 값에 따라 형태 또는 색상을 추가하려면 다음 단계를 수행하십시오 .

1. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다 . 히스토그램에서는 오른쪽의 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭합니다 .
2. **행 범례**를 선택합니다 .
3. 나타나는 창에서 값에 색상 또는 표식을 적용하려는 열을 강조 표시합니다 . 오른쪽에 범례 미리보기가 표시됩니다 .

그림 9.24 행 범례 추가



4. 다음 옵션을 사용하여 행 범례를 세부적으로 정의합니다 .

색상 사전 정의된 여러 색상 구성 중에서 선택할 수 있습니다 .

팁 : 데이터 테이블에 적용된 색상 구성이 이미 있는 경우 기존 색상 구성을 유지하려면 " 없음 " 을 선택하십시오 . 색상은 일관되게 적용되고 고유하게 그룹화되어야 합니다 .

연속형 척도 오름차순 또는 내림차순 값에 해당하는 스펙트럼으로 색상을 할당합니다. 강조 표시된 열에 연속형 값이 포함되어 있으면 이 옵션을 사용합니다 .

역순 척도 색상 척도를 반대로 적용합니다 .

표식 여러 표식 구성 중에서 선택할 수 있습니다 .

팁 : 데이터 테이블에 적용된 표식 구성이 이미 있는 경우 기존 표식 구성을 유지하려면 " 없음 " 을 선택하십시오 . 표식은 일관되게 적용되고 고유하게 그룹화되어야 합니다 .

범례가 포함된 창 만들기 각 값에 해당하는 색상 및 형태를 알려 주는 별도의 범례 창을 생성합니다 .

열 특성에 저장 선택한 색상 테마를 저장하는 열 특성을 추가합니다 .

테이블 특성에 저장 선택한 색상 및 표식 구성을 보존하는 테이블 특성을 추가합니다 .

제외된 행 제외된 행에 색상 또는 표식을 할당합니다 .

대부분의 범례는 하나의 열에 포함됩니다. 하지만 다음 플랫폼의 경우에는 수준 수가 20 개를 초과할 경우 여러 열로 된 범례가 사용됩니다.

- 재발 분석
- 일원 분석 (CDF 그림 및 세 가지 밀도 모두의 빨간색 삼각형 메뉴 명령)
- 모형 적합 (표준 최소 제곱의 회귀 그림)
- 생존

행 범례 제거

행 범례를 마우스 오른쪽 버튼으로 클릭하고 **제거**를 선택하여 행 범례를 삭제할 수 있습니다.

그래프 모양 변경

여러 가지 방법으로 보고서 형식을 필요에 맞게 지정할 수 있습니다. 이후 섹션에서는 출력 보고서의 그래픽 부분을 변경하는 방법에 대해 자세히 설명합니다.

팁 : 터치스크린을 사용하는 경우 JMP 에서 대부분의 그래프와 축을 이동하고 확대 / 축소할 수 있습니다. 그래프와 축의 척도를 변경할 수 있으며, 손가락을 모아 축소하거나 손가락을 벌려 확대할 수 있습니다 .

- " 그래프 크기 조정 "
- " 그래프의 확대 "
- " 그래프의 선 너비 변경 "
- " 그래프의 배경 색상 변경 "
- " 히스토그램 막대의 색상 변경 "
- " 그래프의 좌표 및 임시 참조선 표시 "
- " 그래프의 축 스크롤 및 척도 조정 "
- " 그래프의 축 및 축 라벨 사용자 정의 "
- " 보고서의 값 순서 변경 "
- " 그래프에서 선택한 개체의 패턴 및 형식 변경 "
- " 데이터 테이블에 그래프 추가 "
- " 그래프의 표식 , 선 , 텍스트 등 사용자 정의 "

그래프 크기 조정

JMP 에서 그래프 크기를 조정하는 두 가지 기본적인 방법은 클릭 후 드래그하는 방법과 픽셀 크기에 따라 크기를 조정하는 방법입니다 .

참고 : " 파일 " > " 환경 설정 " > " 그래프 " 에서 " 그래프 높이 " 옵션을 사용하여 그래프의 기본 크기를 변경할 수도 있습니다 .

클릭 후 드래그 방법 사용

클릭 후 드래그 방법을 사용하여 그림 또는 그래프의 크기를 조정하려면

1. 그림 프레임의 오른쪽 모서리 , 아래쪽 모서리 또는 오른쪽 하단 모퉁이 위에 커서를 놓습니다 . 커서가 작은 양방향 화살표로 바뀝니다 .
2. 클릭 후 드래그하여 그림의 크기를 변경합니다 . 크기를 조정할 때 해당 프레임에 있는 모든 그림의 높이 및 너비는 동일한 보고서 창의 다른 프레임과는 독립적으로 조정됩니다 . 표 9.2 에서는 그림을 조정하는 방법에 대해 설명합니다 .

표 9.2 크기 조정 작업

작업	방법
그림 프레임을 조정하되 비율(화면비율)은 유지	Shift 키를 누른 채로 프레임 모퉁이를 클릭하여 드래그합니다.
8픽셀씩 그림 크기 조정	Alt 키를 누른 채로 프레임 모퉁이를 클릭하여 드래그합니다.
동일한 유형의 모든 그림을 동시에 조정	Ctrl 키를 누른 채로 그림 중 하나의 모퉁이를 클릭하여 드래그합니다. 하나의 산점도에 대해 이 작업을 수행하면 해당 창의 모든 산점도에 같은 작업이 적용되어 함께 크기 조정됩니다. 다른 모든 유형의 그림은 변경되지 않은 채로 유지됩니다.

픽셀 단위로 그래프 크기 지정

그림 또는 그래프를 특정 픽셀 크기로 조정하려면

1. 그림 또는 그래프를 마우스 오른쪽 버튼으로 클릭합니다 .
2. **크기 / 척도 > 프레임 크기**를 선택합니다 .
3. 프레임의 높이 및 너비에 해당하는 픽셀 수를 입력합니다 .

참고 : "크기/척도" 메뉴의 다른 옵션에 대한 자세한 내용은 "그래프의 축 스크롤 및 척도 조정"(454 페이지) 에서 확인하십시오 .

그래프의 확대

돋보기 를 사용하여 그림의 영역을 자동으로 확대할 수 있습니다 . 돋보기를 클릭하면 클릭한 위치의 점 또는 영역이 새 데이터 보기의 가운데가 됩니다. 새 보기의 척도가 늘어나므로 관심 있는 점 또는 패턴을 보다 자세히 볼 수 있습니다 . 다음과 같은 작업을 수행할 수 있습니다 .

- 돋보기를 클릭 후 드래그하여 그림의 특정 영역을 중점적으로 확인합니다 .
- 삼원 그림에서 돋보기를 드래그하여 삼각 축을 확대 / 축소합니다 .
- 반복적으로 확대 / 축소하여 데이터를 보다 자세히 봅니다 .
- Ctrl 키를 누른 채 돋보기를 클릭하여 마지막 확대 / 축소 전의 상태로 돌아갑니다 .
- 돋보기를 두 번 클릭하거나 Alt 키를 누른 채 클릭하여 원래 그림을 복원합니다 .

Windows 에서는 Ctrl 키를 누른 채로 마우스 휠을 움직여 확대/축소할 수 있습니다.

macOS 의 경우, Option 키를 누른 채 해당 영역을 클릭하여 확대/축소합니다.

참고 : 축에서 그래프 안쪽으로 드래그하면 중앙점을 계속 표시하면서 영역을 확대할 수 있습니다. 세로 축의 경우에는 위치를 클릭한 후 그래프를 향해 오른쪽으로 드래그하십시오. 그러면 척도가 확대되면서 척도의 양 끝점이 모두 조정되지만 클릭한 값은 대략적으로 동일한 위치에 유지됩니다 . 왼쪽으로 드래그하면 영역이 축소됩니다 .

그래프의 선 너비 변경

그래프에 선을 적합시키거나 선이 이미 있는 그래프를 생성한 선의 너비를 조정할 수 있습니다.

1. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다 .
2. **사용자 정의**를 선택합니다 .
3. **평활기** (예를 들어 그래프에 평활기가 포함된 경우) 를 선택합니다 .
4. 선 너비를 변경합니다 . 값은 픽셀 단위입니다 .
5. **확인**을 클릭합니다 .

그래프의 배경 색상 변경

그래프의 배경 색상을 변경하려면 다음 단계를 수행하십시오.

1. 그래프의 아무 곳에서나 마우스 오른쪽 버튼을 클릭합니다 . 상자 그림의 색상만 변경하려면 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭합니다 .
2. **배경 색상**을 선택합니다 .
3. 사전 정의된 색상 중 하나를 선택하거나 , 고유한 색상을 생성합니다 .
4. **확인**을 클릭합니다 .

히스토그램 막대의 색상 변경

히스토그램 막대의 색상을 변경하려면 다음 단계를 수행하십시오.

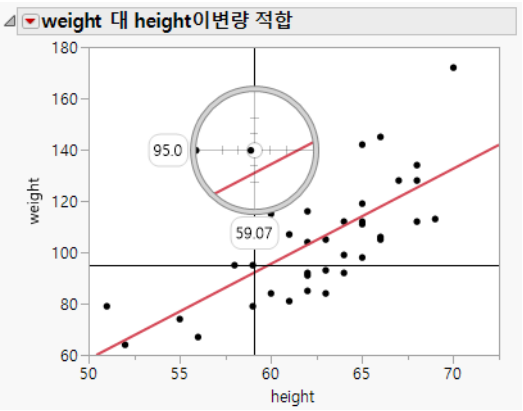
1. 히스토그램의 아무 곳에서나 마우스 오른쪽 버튼을 클릭하고 **히스토그램 색상**을 선택합니다.
2. 색상을 클릭하거나, **기타**를 클릭하고 고유한 색상을 생성합니다.

고유한 색상을 생성하는 방법에 대한 자세한 내용은 운영 체제 설명서에서 확인하십시오.

그래프의 좌표 및 임시 참조선 표시

그래프에서 점과 거리를 측정하거나, 그림 및 그래프에서 점의 정확한 값(또는 좌표)과 거리를 측정할 수 있습니다. 이렇게 하려면 십자기호 도구를 클릭하고 그래프의 아무 곳이나 클릭한 채로 있습니다. 그림 내에서 십자기호 도구를 드래그하면 십자기호 도구가 세로 축 및 가로 축과 만나는 지점의 좌표 값이 나타납니다.

그림 9.25 십자기호 도구 사용



적합된 선 또는 곡선에서 십자기호 도구는 예측값에 대한 반응 값을 식별합니다. 삼원 그림에서 이 도구는 삼각 십자기호 선을 표시합니다.

그래프의 축 스크롤 및 척도 조정

손 도구()를 사용하여 그림의 축 및 보기를 변경할 수 있습니다.

- *Y* 축에서 을 드래그하면 *Y* 축이 스크롤되고, 또는 을 드래그하면 *Y* 축의 척도가 조정됩니다.
- *X* 축에서 을 드래그하면 *X* 축이 스크롤되고, 또는 을 드래그하면 *X* 축의 척도가 조정됩니다.

팁 : 축을 드래그하여 척도를 변경할 경우 JMP 는 새 축 너비에 따라 자동으로 주 눈금 및 보조 눈금 증분을 업데이트합니다 . 이를 방지하여 원래 증분을 유지하려면 Shift 키를 누른 채로 드래그하십시오 .

그림 또는 그래프를 마우스 오른쪽 버튼으로 클릭하고 **크기/척도**(또는 **그래프 > 크기/척도**)를 선택해도 됩니다. 다음 옵션 중 하나를 선택합니다.

- X 축의 척도를 조정하려면 **X 축**을 선택합니다 . Y 축의 척도를 조정하려면 **Y 축** 또는 **오른쪽 Y 축**을 선택합니다 . 이 창에 대한 자세한 내용은 " 그래프의 축 및 축 라벨 사용자 정의 "(455 페이지) 에서 확인하십시오 .
- " 프레임 크기 " 옵션에 대한 자세한 내용은 " 픽셀 단위로 그래프 크기 지정 "(452 페이지) 에서 확인하십시오 .
- *X* 및 *Y* 축이 동일한 단위로 측정되는 경우 그래프에서 방향에 관계없이 거리가 정확하게 표시되도록 하려면 **등간격 가로 - 세로 축**을 선택합니다 .

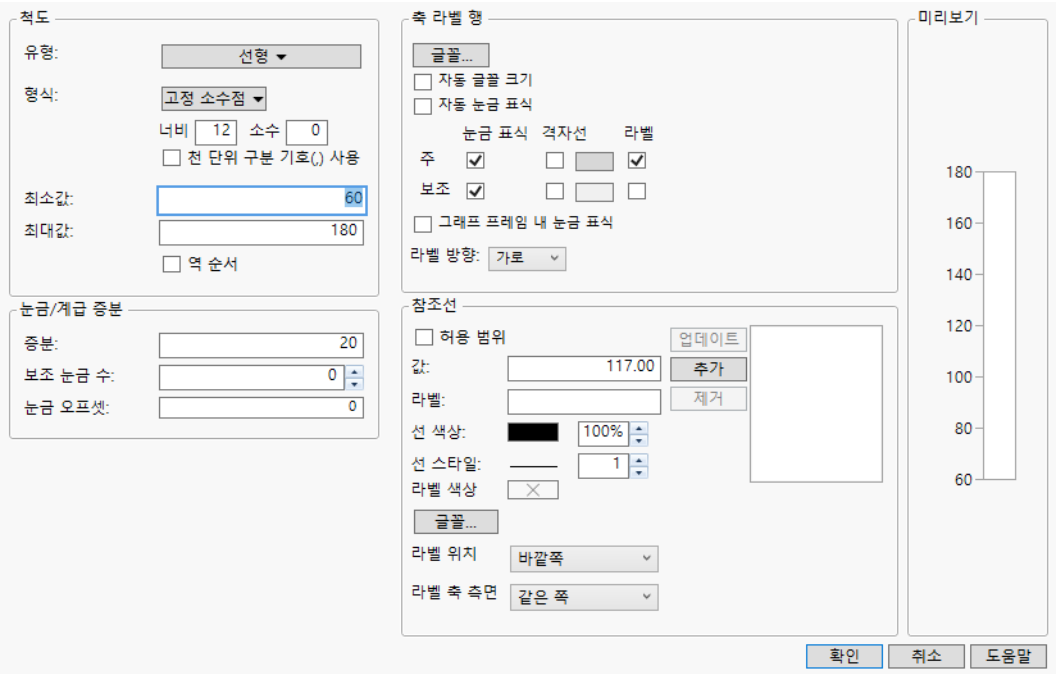
그래프의 축 및 축 라벨 사용자 정의

"축 설정" 창을 사용하여 축을 사용자 정의하려면 숫자 축을 두 번 클릭합니다. 또는 축 영역을 마우스 오른쪽 버튼으로 클릭하고 **축 설정**을 선택하여 이 창에 액세스합니다.

참고 : 그래프 빌더 및 분포 플랫폼에서는 " 축 설정 " 창을 사용하여 명목형 축도 사용자 정의할 수 있습니다 .

이 창의 사용자 정의 기능은 축의 데이터 유형과 JMP 가 그림 또는 차트를 생성하는 데 사용하는 특정 플랫폼에 따라 달라집니다. 그림 9.26 에서는 숫자 (연속형) 축에 대한 일반적인 "축 설정" 창을 보여 줍니다.

그림 9.26 숫자 (연속형) 축에 대한 축 설정 창



"축 설정" 창에는 다음과 같은 패널이 포함되어 있습니다.

- " 척도 "(456 페이지)
- " 눈금 / 계급 증분 "(458 페이지)
- " 축 라벨 행 "(460 페이지)
- " 참조선 "(463 페이지)
- 미리보기 패널에서는 현재 선택한 설정이 축에 어떻게 표시될지를 보여 줍니다 .

척도

" 척도 " 패널에서는 다음을 수행할 수 있습니다 .

- " 축 척도 유형 변경 "(456 페이지)
- " 축의 숫자 형식 변경 "(457 페이지)
- " 최소 및 최대 축 값 설정 "(458 페이지)

축 척도 유형 변경

숫자 축이 있는 그래프를 볼 때 축 척도를 다음 유형 중 하나로 변경할 수 있습니다 .

- 선형

- 로그
- 검정력
- 측지
- 측지 (미국)
- 확률 척도 (정규 , Weibull, Fréchet, 로지스틱 , 지수 , 감마 , 베타 , 혼합)

참고 :

- 특정 플랫폼에서는 고정되어 변경할 수 없는 다른 척도 유형을 사용할 수도 있습니다 . 예를 들어 파괴 열화 , 열화 , 그리고 X 로 Y 적합 산점도의 X 축에서는 사용자 척도를 사용합니다 . 이러한 그래프에서는 " 축 설정 " 창을 통해 척도를 변경하지 마십시오 .
- 척도 유형을 " 로그 " 로 선택한 경우 사용할 로그 밑을 입력하십시오 .
- 척도 유형을 " 거듭제곱 " 으로 선택한 경우 사용할 거듭제곱을 입력하십시오 .
- 거듭제곱은 로그 척도와 유사하게 작동하지만 척도 변환이 log(x) 가 아니라 x^p 입니다 . 여기서 , p 는 사용자 정의 가능한 축 옵션입니다 . 예를 들어 Power = 2 는 x^2 축을 생성합니다 .

분석을 실행할 때마다 변경할 필요가 없도록 변수의 기본 척도 유형을 설정하는 방법은 "JMP 열 특성 설정 " 장의 " 축 "(288 페이지) 에서 확인하십시오 .

축의 숫자 형식 변경

숫자 축 영역이 포함된 그림 및 차트의 경우 축의 형식을 변경할 수 있습니다 . 숫자 형식에 대한 자세한 내용은 "JMP 열 특성 설정 " 장의 " 숫자 형식 "(273 페이지) 에서 확인하십시오 .

참고 :

- " 날짜 ", " 시간 " 또는 " 기간 " 을 선택한 경우 증분 형식을 지정해야 합니다 . 자세한 내용은 " 축 라벨 추가 및 제거 "(464 페이지) 에서 확인하십시오 . 라벨 행 내포도 지정할 수 있습니다 . 자세한 내용은 " 라벨 행 내포 "(460 페이지) 에서 확인하십시오 .
- " 고정 소수점 " 을 선택한 경우 " 소수 " 상자에 표시할 소수 자릿수를 입력하십시오 .
- " 정밀도 " 를 선택한 경우 후행 0 과 모든 전체 자릿수를 유지할지 여부를 선택하십시오 .
- 천 이상의 값에 쉼표를 추가하려면 **천 단위 구분 기호 사용** 옵션을 선택하십시오 . " 너비 " 상자에서 쉼표 하나당 공백을 하나씩 추가해야 합니다 . 그러지 않으면 쉼표가 표시되지 않습니다 . 이 옵션은 " 최적 ", " 고정 소수점 ", " 백분율 " 및 " 통화 " 형식에 사용할 수 있습니다 .
- 숫자 축의 경우 " 너비 " 상자를 사용하여 눈금 표식 라벨의 너비를 조정할 수 있습니다 .

참고 : 축의 숫자 형식을 변경할 때 해당 데이터 테이블에서 값이 표시되는 방식과 관련된 숫자 형식은 변경하면 안 됩니다 . 데이터 테이블에서 날짜 또는 시간이 표시되는 방식을 변경하는 방법은 "JMP 열 특성 설정 " 장의 " 숫자 형식 "(273 페이지) 에서 확인하십시오 .

날짜 증분 드롭다운 메뉴에서 날짜 간격을 선택하면 JMP 날짜 (초 수) 가 적절한 단위로 나뉩니다 . 이런 방식으로 데이터에 원하는 그림 척도를 지정할 수 있습니다 . 날짜 축은 JMP 날짜 값이

포함된 열이어야 하고 "축 설정" 창에서 "열 정보" 창에 설정된 날짜 형식으로 표시되어야 합니다. 하지만 "축 설정" 창을 사용하여 그림의 날짜 표시 형식을 지정할 수 있습니다.

최소 및 최대 축 값 설정

숫자 축 영역이 포함된 그림 및 차트의 경우 그래프에 표시할 최소값 및 최대값을 변경할 수 있습니다.

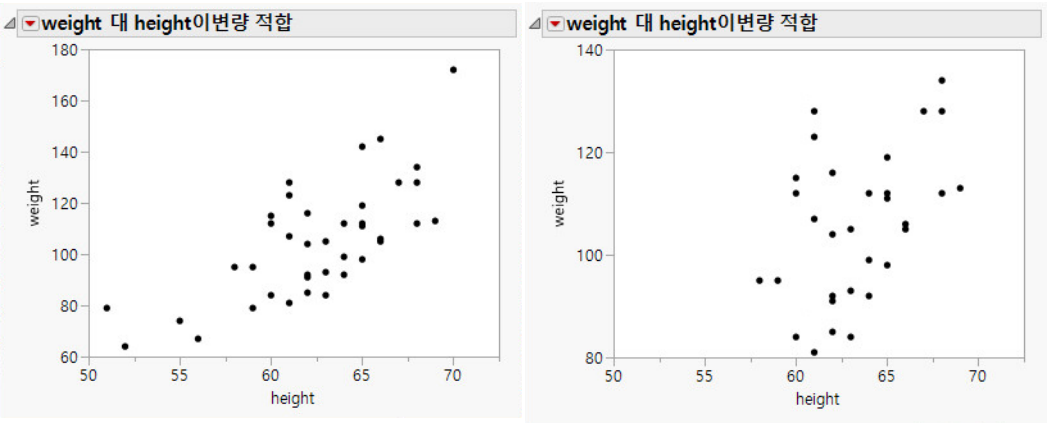
다음 사항에 유의하십시오.

- 숫자 축의 기본 최소 및 최대 축 설정을 복원하려면 숫자 축을 마우스 오른쪽 버튼으로 클릭하고 **축 되돌리기**를 선택합니다.
- 최소값과 최대값을 바꿔서 축을 역 순서로 표시하려면 **역 순서**를 선택합니다. 삼원 그림의 경우에는 이 옵션이 지원되지 않습니다.

팁 : 분석을 실행할 때마다 변경할 필요가 없도록 변수의 기본 최소 및 최대 축 값을 설정하는 방법은 "JMP 열 특성 설정" 장의 "축"(288 페이지) 에서 확인하십시오.

그림 9.27의 오른쪽 예는 왼쪽 그림의 80에서 140 사이를 표시하도록 점 군집을 확대한 것입니다. 이 확대된 그림은 최대 및 최소 축 값을 다시 할당하여 생성한 것입니다.

그림 9.27 축 척도를 재조정하여 그림의 일부 확대



눈금 / 계급 증분

"눈금 / 계급 증분" 패널에서는 다음을 수행할 수 있습니다.

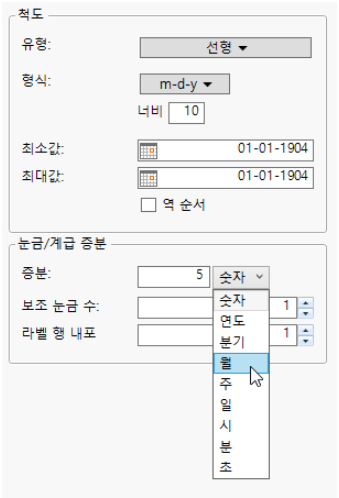
- "축 증분 변경"(459 페이지)
- "보조 눈금 표식 추가"(459 페이지)
- "눈금 오프셋 변경"(459 페이지)
- "라벨 행 내포"(460 페이지)

축 증분 변경

그래프를 볼 때 축 증분을 변경할 수 있습니다.

축의 "형식"이 날짜, 시간 또는 기간으로 설정되어 있으면 증분 옆에 형식 메뉴가 표시됩니다(그림 9.28). 증분의 형식을 선택합니다.

그림 9.28 날짜 및 시간 증분의 형식 선택



팁 : 분석을 실행할 때마다 변경할 필요가 없도록 변수의 기본 축 증분을 설정하는 방법은 "JMP 열 특성 설정 " 장의 " 축 "(288 페이지) 에서 확인하십시오 .

보조 눈금 표식 추가

숫자 축에 눈금 표식을 추가하거나, 숫자 축에 표시할 보조 눈금 표식의 수를 변경할 수 있습니다. "축 라벨 행" 패널에서 **보조** 눈금 표식 상자를 선택해야 축에 해당 눈금 표식이 표시됩니다.

팁: 분석을 실행할 때마다 변경할 필요가 없도록 변수의 기본 보조 눈금 표식을 설정하는 방법은 "JMP 열 특성 설정 " 장의 " 축 "(288 페이지) 에서 확인하십시오 .

눈금 오프셋 변경

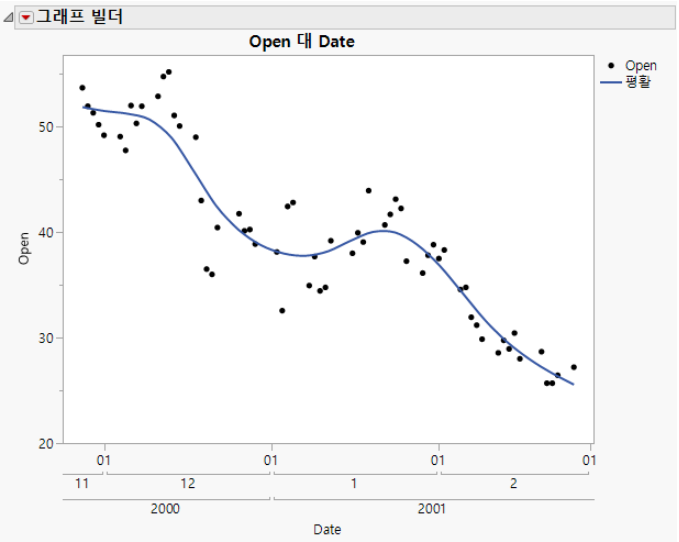
눈금 표식의 시작점을 변경하려면 눈금 오프셋 상자에 숫자를 입력합니다. 예를 들어 "눈금 오프셋"이 현재 0으로 설정되어 있는 경우 이를 1로 설정하면 축의 모든 값이 1씩 위로 이동합니다.

라벨 행 내포

이 옵션은 축의 "형식"으로 "날짜", "시간" 또는 "기간"을 선택한 경우에만 표시됩니다. 라벨 행 내포 기능을 사용하면 해당 형식에 따라 날짜 축을 여러 행으로 분할할 수 있습니다. 예를 들어 가장 바깥쪽 행에 연도를 넣고 그 다음 월 및 날짜를 차례로 넣을 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Stock Prices.jmp 를 엽니다 .
2. **그래프 > 그래프 빌더**를 선택합니다 .
3. Date 를 **X** 영역으로 드래그합니다 .
4. Open 을 **Y** 영역으로 드래그합니다 .
5. X 축을 두 번 클릭합니다 .
6. **형식**에서 **날짜 > ddMonyyyy** 를 선택합니다 .
7. **라벨 행 내포**의 값을 3 으로 늘립니다 .
8. **확인**을 클릭합니다 .
9. **완료**를 클릭하여 완료된 그래프를 확인합니다 .

그림 9.29 라벨 행 내포의 예



축 라벨 행

"축 라벨 행" 패널에서는 다음을 수행할 수 있습니다.

- " 축 라벨 글꼴 변경 "(461 페이지)
- " 눈금 표식 , 격자선 또는 라벨 추가 "(461 페이지)
- " 눈금 라벨의 방향 변경 "(462 페이지)

참고 : 그래프 빌더에서 축의 라벨을 내포한 경우 " 축 라벨 행 " 패널에 탭이 나타납니다 . 각 라벨 행을 해당 탭에서 개별적으로 수정할 수 있습니다 .

축 라벨 글꼴 변경

모든 축 유형에서 축 라벨 글꼴을 수정할 수 있습니다. 변경 사항은 활성 그래프에만 적용됩니다. 기본 축 라벨 글꼴을 설정하는 방법은 "JMP 환경 설정" 장의 "글꼴 환경 설정"(655 페이지)에서 확인하십시오.

현재 글꼴 유형 및 크기를 변경하려면 축을 두 번 클릭하고 **축 설정**을 선택한 다음 **글꼴**을 클릭합니다 .

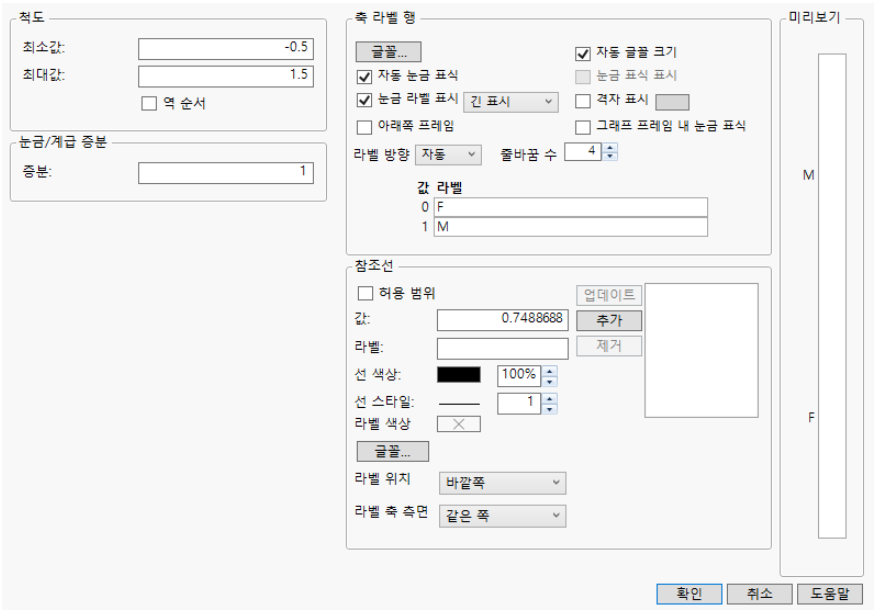
참고 :

- 기본 크기로 모든 라벨을 제대로 표시할 수 없는 경우 최소값 한도 내에서 글꼴 크기를 줄이려면 **자동 글꼴 크기**를 선택하십시오 .
- 축 라벨 글꼴 색상을 변경하려면 축을 마우스 오른쪽 버튼으로 클릭하고 **특성 표시**를 선택합니다 . 오른쪽의 " 편집 " 옵션에서 텍스트 색상을 변경합니다 .

눈금 표식 , 격자선 또는 라벨 추가

이러한 옵션은 축이 연속형인지 범주형인지에 따라 달라질 수 있습니다. 그림 9.30에서는 범주형 축에 대한 축 설정을 보여 줍니다. 그림 9.26(456 페이지)에서는 연속형 축에 대한 축 설정을 보여 줍니다.

그림 9.30 범주형 축의 축 라벨 행



연속형 축에 대한 옵션 :

- 공간이 충분하지 않아 하나 이상의 라벨이 숨겨진 경우에 한해 눈금 표식을 설정하려면 **자동 눈금 표식**을 선택합니다 .
- 주 눈금 또는 보조 눈금 표식, 격자선, 또는 라벨을 추가하려면 해당 체크박스를 선택합니다.
- 격자선의 색상을 변경하려면 색상 상자를 클릭합니다 .
- 그래프 내에서 눈금 표식을 이동하려면 **그래프 프레임 내 눈금 표식**을 선택합니다 .

범주형 축에 대한 옵션 :

- 기본값은 모든 라벨이 표시된 경우 눈금 표식을 표시하지 않는 것입니다 . 눈금 표식을 추가하려면 **자동 눈금 표식**을 선택합니다 . 눈금 표식 스타일을 변경하려면 해당 옵션 중 하나를 선택합니다 .
- 공간이 충분하지 않아 하나 이상의 라벨이 숨겨진 경우에 한해 눈금 표식을 설정하려면 **자동 눈금 표식**을 선택합니다 .
- 격자선을 추가하려면 **격자 표시**를 선택합니다 .
- 격자선의 색상을 변경하려면 색상 상자를 클릭합니다 .
- 그래프 내에서 눈금 표식을 이동하려면 **그래프 프레임 내 눈금 표식**을 선택합니다 .
- 라벨 주위에 프레임을 표시하려면 **아래쪽 프레임**을 선택합니다 .
- 텍스트가 줄바꿈되는 행 수를 변경합니다 .
- 라벨을 수정하려면 " 라벨 " 상자에 다른 값을 입력합니다 .

팁 :

- 분석을 실행할 때마다 변경할 필요가 없도록 변수의 기본 눈금 표식을 설정하는 방법은 "JMP 열 특성 설정 " 장의 " 열 표시 방식을 제어하는 특성 "(288 페이지) 에서 확인하십시오 .
- 추가적인 축 사용자 정의를 위해서는 축 마우스 오른쪽 버튼 클릭 메뉴에서 **특성 표시**를 선택합니다 .

눈금 라벨의 방향 변경

눈금 라벨의 방향을 변경하려면 "라벨 방향" 목록에서 옵션을 선택합니다.

자동 눈금 증분과 라벨 길이에 따라 눈금 라벨을 읽기 쉽도록 라벨 방향이 자동으로 결정됩니다 .

가로 아래쪽 기준선 사용

세로 오른쪽 기준선을 사용하여 세로로

직교 세로 축에는 가로로 , 가로 축에는 세로로

> **참고 :** " 평행 " 과 " 직교 " 는 다변량 그림에서처럼 축이 쌍으로 있는 경우에 유용합니다 .

평행 세로 축에는 세로로 , 가로 축에는 가로로

기울임 약 45 도 기울임

참조선

그래프에 참조선을 추가합니다.

1. 숫자 축을 두 번 클릭합니다. 또는, 숫자 축을 마우스 오른쪽 버튼으로 클릭하고 **축 설정**을 선택합니다.
2. " 값 " 텍스트 상자에 원하는 참조선에 해당하는 값을 입력합니다. 그래프에서 이 위치에 선이 추가됩니다.

   참고 : 그래프를 두 번 클릭하여 " 축 설정 " 창을 연 경우 기본값은 그래프에서 클릭한 지점에 해당합니다.

3. 선의 라벨을 입력합니다.
4. 다음 옵션 중 하나를 선택하여 참조선을 보다 세부적으로 사용자 정의할 수 있습니다.

   허용 범위 참조선의 시작 및 끝을 정의하는 최소값과 최대값을 입력할 수 있습니다. 그러면 해당 데이터 범위에 대해 참조선이 표시됩니다.

   선 색상 선 색상을 변경합니다. 불투명도를 지정할 수도 있습니다.

   선 스타일 선 스타일을 지정합니다 (" 허용 범위 " 가 선택되지 않은 경우). 선 너비를 지정할 수도 있습니다.

   글꼴 참조선 라벨의 글꼴을 조정할 수 있습니다.

   라벨 위치 참조선 라벨의 위치를 선택할 수 있습니다. 옵션은 그래프 프레임 바깥쪽 또는 안쪽입니다.

   라벨 축 측면 참조선 라벨의 그래프 측면을 선택할 수 있습니다.

5. **추가**를 클릭합니다. 해당 값이 **추가** 버튼의 오른쪽에 있는 상자에 표시되어 그래프에 추가될 참조선을 보여 줍니다. 변경 사항은 " 미리보기 " 창에 표시됩니다.
6. 선을 더 추가하려면 위의 단계를 반복합니다.
7. 기존 참조선을 수정하려면 참조선을 변경한 후 **업데이트**를 클릭합니다.

팁 : 분석을 실행할 때마다 변경할 필요가 없도록 변수의 기본 참조선을 설정하는 방법은 "JMP 열 특성 설정 " 장의 " 열 표시 방식을 제어하는 특성 "(288 페이지) 에서 확인하십시오.

범주형 축의 구분선 및 프레임 확장

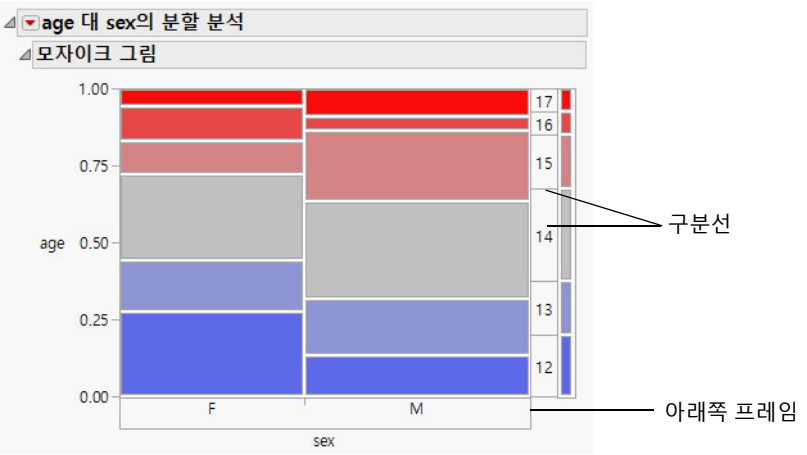
명목형 또는 순서형 (범주형) 변수의 수준이 많은 경우 눈금 라벨 사이의 세로 구분선을 확장하면 유용합니다.

구분선을 *X* 축 라벨까지 확장하려면 다음 단계를 수행하십시오.

1. 명목형 또는 순서형 축을 두 번 클릭하여 축 설정을 표시합니다.

2. **눈금 표식 표시 > 눈금 표식 표시의 짧은 구분자 > 긴 구분자**를 선택하여 구분선을 추가하거나 **아래쪽 프레임**을 선택하여 축 영역을 둘러싸는 프레임을 추가합니다 .

그림 9.31 구분선



축 라벨 추가 및 제거

숫자 축의 라벨을 추가하거나 제거할 수 있습니다. 축 라벨을 추가하려면

1. 숫자 축을 마우스 오른쪽 버튼으로 클릭하고 **축 라벨 추가**를 선택합니다 .
2. 축 라벨의 이름을 입력합니다 . 축 영역이 확장되어 입력한 만큼 라벨 줄이 표시됩니다 .

이 명령을 여러 번 사용하여 라벨을 여러 개 추가할 수 있습니다. 축에 추가한 라벨을 편집하려는 경우 라벨을 클릭하면 라벨이 편집 상자로 바뀝니다.

축 라벨을 제거하려면 숫자 축을 마우스 오른쪽 버튼으로 클릭하고 **축 라벨 제거**를 선택합니다. 그러면 마지막으로 추가된 라벨이 제거됩니다.

축 라벨 및 눈금 표식 회전

모든 축 유형에서 축 라벨을 수정할 수 있습니다.

축 라벨을 회전하려면 다음 단계를 수행하십시오 .

1. 축을 두 번 클릭하여 축 설정을 표시합니다 .
2. " 라벨 방향 " 목록에서 텍스트를 회전할 방향을 선택합니다 .

팁 : 분석을 실행할 때마다 변경할 필요가 없도록 변수의 기본 축 라벨 위치를 설정하는 방법은 "JMP 열 특성 설정 " 장의 " 열 표시 방식을 제어하는 특성 "(288 페이지) 에서 확인하십시오 .

눈금 라벨을 마우스 오른쪽 버튼으로 클릭하고 **회전된 눈금 라벨**을 선택하여 명목형 축에서 눈금 표식을 회전할 수 있습니다.

그래프 내용 복사 및 붙여넣기

적합선 같은 요소를 추가하여 그래프를 사용자 정의한 후 그래프 내용을 복사하여 호환되는 다른 그래프에 붙여넣을 수 있습니다.

1. 사용자 정의한 그래프를 마우스 오른쪽 버튼으로 클릭합니다.
2. **편집 > 프레임 내용 복사**를 선택합니다.
3. 설정을 복사할 대상 그래프를 마우스 오른쪽 버튼으로 클릭합니다.
4. **편집 > 프레임 내용 붙여넣기**를 선택합니다.

팁 : 히스토그램 막대를 붙여넣는 데 어려움이 있는 경우 Scripting Guide 의 에서 JSL 해결 방법을 확인하십시오.

축 설정 복사 및 붙여넣기

"그래프의 축 및 축 라벨 사용자 정의"(455 페이지)에 설명된 대로 축을 사용자 정의한 후 새 설정을 복사하여 다른 축에 붙여넣을 수 있습니다.

1. 사용자 정의한 축을 마우스 오른쪽 버튼으로 클릭합니다.
2. **편집 > 축 설정 복사**를 선택합니다.
3. 설정을 복사할 대상 축을 마우스 오른쪽 버튼으로 클릭합니다.
4. **편집 > 축 설정 붙여넣기**를 선택합니다.

팁 : 축 설정을 여러 그래프에 일괄 적용하려면 Ctrl 키를 누른 채로 축을 마우스 오른쪽 버튼으로 클릭하고 **편집 > 축 설정 붙여넣기**를 선택합니다.

보고서의 값 순서 변경

JMP 보고서의 데이터가 원하는 순서대로 표시되지 않을 수 있습니다. 보고서의 데이터가 원하는 순서대로 표시되도록 데이터에 특정 순서를 지정하려면 "JMP 열 특성 설정" 장의 "값 순서"(285 페이지)에 설명된 대로 분석을 실행하기 전에 열에 "값 순서" 특성을 할당합니다.

값에 다음이 포함된 경우 이러한 값은 보고서에서 자동으로 적절한 순서로 표시됩니다.

- 1 월, 2 월, 3 월, 4 월, 5 월, 6 월, 7 월, 8 월, 9 월, 10 월, 11 월, 12 월
- 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
- 일요일, 월요일, 화요일, 수요일, 목요일, 금요일, 토요일
- 매우 낮음, 낮음, 중간 낮음, 중간, 중간 높음, 높음, 매우 높음
- 매우 비동의, 비동의, 보통, 관심 없음, 동의, 매우 동의
- 실패, 허용 불가, 매우 나쁨, 나쁨, 조금 나쁨, 허용 가능, 평균, 조금 좋음, 좋음, 매우 좋음, 탁월, 최상

그래프에서 선택한 개체의 패턴 및 형식 변경

막대 및 맵 같은 일부 개체의 경우 개체를 선택하면 패턴이 나타납니다. 범례에서 특정 개체의 패턴을 변경할 수 있습니다.

범례의 항목을 마우스 오른쪽 버튼으로 클릭하고 **채우기 패턴**을 선택하여 새 패턴을 선택합니다.

패턴을 제거하려면 항목을 다시 마우스 오른쪽 버튼으로 클릭하고 **되돌리기**를 선택합니다.

모든 그래프에서 선택한 개체의 형식 수정

파일 > 환경 설정 > 그래프(Windows) 또는 **JMP > 환경 설정 > 그래프**(macOS)를 선택하여 " 채우기 영역 " 환경 설정을 변경합니다.

- 선택한 개체의 강조 표시를 변경하려면 " 채우기 선택 모드 " 환경 설정에서 다음 중 하나를 선택합니다 .

  선택 시 패턴화 줄무늬 패턴 (기본 설정)

  선택 시 어둡게 어두운 단색

  선택 시 외곽선 표시 외곽선

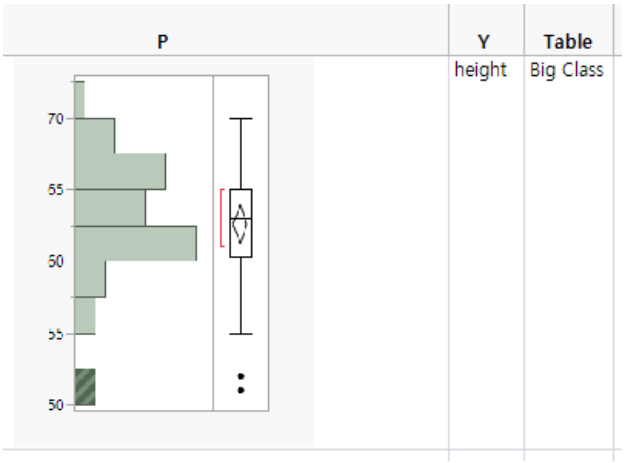
  선택 시 지정한 색상 적용 환경 설정의 " 채우기 선택 색상 " 에 따라 선택한 개체의 형식을 지정합니다 .

- " 채우기 선택 모드 " 가 " 선택 시 지정한 색상 적용 " 으로 설정된 경우 그래픽 개체에서 선택된 행의 색상을 변경하려면 " 채우기 선택 색상 " 막대를 클릭합니다 . 기본 색상은 빨간색입니다 .

데이터 테이블에 그래프 추가

새 데이터 테이블에 그래프를 추가할 수 있습니다. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **편집 > 이와 같은 그래프를 포함하는 테이블 생성**을 선택합니다. 새 데이터 테이블에 해당 그래프 및 변수가 표시됩니다. 그런 다음 행 및 열 크기를 조정하여 그래프를 확대할 수 있습니다.

그림 9.32 데이터 테이블에 추가된 히스토그램 및 Y 변수



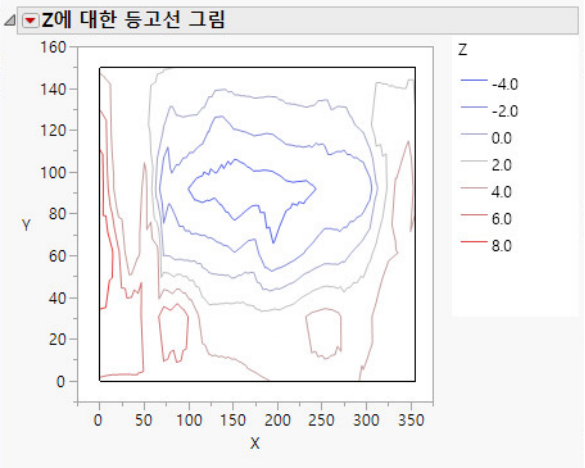
그래프의 표식 , 선 , 텍스트 등 사용자 정의

그래프는 표식, 선, 텍스트 및 기타 그래픽 요소로 구성되며 이러한 요소를 사용자 정의할 수 있습니다. 예를 들어 산점도 행렬 그래프에서 이변량 그래프 중 하나의 데이터 점을 분홍색 단색 표식으로 강조 표시할 수 있습니다. 등고선 그림 그래프에서는 등고선의 너비를 늘리거나 투명도를 높일 수 있습니다.

예 : 등고선 그림 사용자 정의

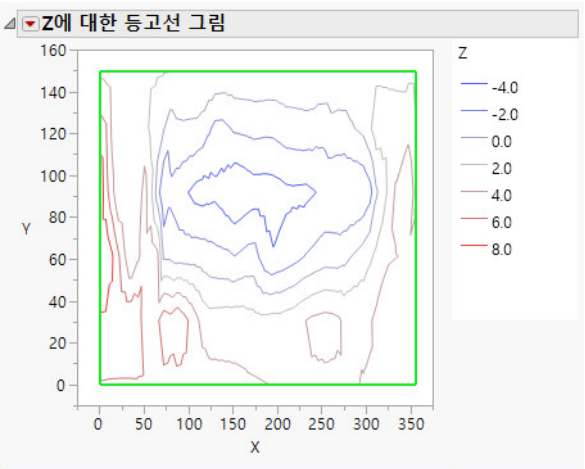
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Little Pond.jmp 를 엽니다 .
2. 테이블 패널에서 "Contour Plot" 스크립트 옆의 녹색 삼각형을 클릭합니다 .
3. "Z 에 대한 등고선 그림 " 옆의 빨간색 삼각형을 클릭하고 **영역 채우기**를 선택 취소합니다 .

그림 9.33 사용자 정의 전 등고선 그림



4. 그림을 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다.
   " 그래프 사용자 정의 " 창이 나타납니다.
5. **형태**를 클릭합니다.
6. **선 너비**를 2.0 으로 변경합니다.
7. **선 색상** 옆의 색상을 클릭하고 녹색 음영을 선택합니다.
8. **확인**을 클릭합니다.

그림 9.34 사용자 정의된 등고선 그림



사용자 정의할 수 있는 그래픽 요소는 각 그래프에 따라 다릅니다. 예를 들어 관리도 그래프에는 세 가지 선 요소(하한, 상한 및 연결선)가 있습니다. 다른 그래프에서는 선 요소를 선 또는 사용자 선이라고 할 수 있습니다.

현재 그래프의 그래픽 요소를 사용자 정의하려면 다음 단계를 수행하십시오 .

1. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다 .
2. 변경하려는 요소를 선택한 후 특성을 수정합니다 .

   변경 사항은 그래프에 즉시 표시됩니다 .
3. **확인**을 클릭하여 변경 사항을 현재 그래프에 저장합니다 .

참고 : 사용자 정의 설정은 범례에 표시된 설정을 재정의하지 않습니다 . 예를 들어 막대 차트의 채우기 영역을 빨간색으로 설정할 경우 범례가 파란색이면 범례는 파란색으로 유지됩니다 .

각 요소에 선 색상, 표식 스타일 및 크기, 텍스트 스타일 등에 대한 특성이 포함될 수 있습니다. 일반적인 특성은 다음과 같습니다.

선 색상 선 색상을 변경합니다 . 창에서 색상을 클릭하여 선택합니다 . 더 많은 색상 중에서 선택하려면 마우스 오른쪽 버튼을 클릭합니다 .

선 스타일 선 스타일을 변경합니다 . 다섯 가지 스타일 중 하나를 클릭하여 선택합니다 .

선 너비 선 너비를 변경합니다 . 상자를 클릭하고 원하는 선 너비를 포인트 단위로 입력합니다 .

표식 표식 형태를 변경합니다 . 표식에 대한 자세한 내용은 " 그래프에서 표식 사용 "(444 페이지) 에서 확인하십시오 .

표식 크기 표식 크기를 변경합니다 .

화살표 선의 시작점 , 끝점 또는 양쪽에 화살촉을 추가하거나 어느 쪽에도 추가하지 않습니다 .

적합선 선의 색상 , 너비 또는 스타일을 변경합니다 .

텍스트 색상 텍스트 색상을 변경합니다 .

글꼴 텍스트의 글꼴 , 스타일 및 크기 (포인트) 를 변경합니다 .

텍스트 스타일 텍스트 맞춤을 가운데 맞춤 , 왼쪽 맞춤 또는 오른쪽 맞춤으로 변경합니다 . **채우기** 옵션은 선택한 색상을 배경에 적용합니다 .

채우기 색상 상자 그림 및 적합 신뢰 영역 같은 개체의 색상을 변경합니다 .

채우기 패턴 채우기 개체의 패턴을 변경합니다 .

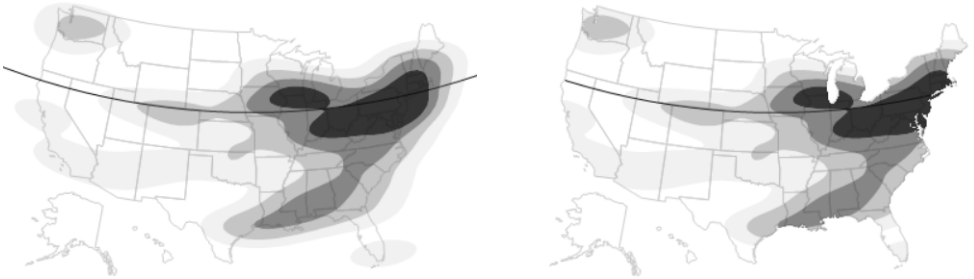
투명도 표식 또는 라벨의 투명도를 변경합니다 . 그래프에 표식 (점) 을 그리려면 투명도 수준을 입력합니다 . 불투명도는 0(완전 투명) 부터 1(불투명) 까지입니다 . 표식 투명도 변경에 대한 자세한 내용은 " 표식 투명도 지정 "(447 페이지) 에서 확인하십시오 .

자르기 쉐이프 선택한 개체 (등고선 등) 를 선택한 자르기 쉐이프를 기준으로 경계를 따라 자릅니다 . 그러면 개체를 구성하는 표시 세그먼트가 맵의 경계와 일치하게 됩니다 . 경계의 일부분 위에 있는 개체만 표시하려면 하위 경계를 선택합니다 . 예를 들어 뉴욕에 있는 개체만 표시할 수 있습니다 .

자르기 쉐이프 옵션

그래픽 요소(참조선, 격자선 및 등고선 같은 표시 세그먼트)를 지도의 경계에 따라 자를 수 있습니다. "그래프 사용자 정의" 창에서는 쉐이프를 자르기 위한 기준 경계를 선택할 수 있는 "자르기 쉐이프" 옵션을 제공합니다. 그림 9.35에서는 미국 지도에 따라 자른 참조선을 보여 줍니다.

그림 9.35 자르지 않은 참조선 (왼쪽) 과 자른 참조선 (오른쪽)



팁 :

- 그림 9.35에 표시된 것과 같은 맵을 생성하는 예는 Essential Graphing의 에서 확인하십시오.
- " 사용자 정의 " 창의 " 자르기 쉐이프 " 옵션은 지도 경계를 따라 등고선을 자르는 용도로도 설정됩니다 . 자세한 내용은 " 자르기 쉐이프 "(469 페이지) 에서 확인하십시오 .
- 스크립팅 시 행렬 또는 문자열로 자르기 경로를 지정하는 등의 추가 옵션을 사용할 수 있습니다 . 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

그래픽 요소에 대한 스크립트 생성

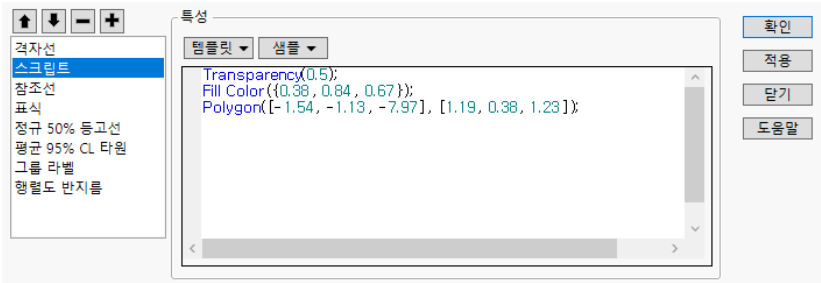
그래픽 요소를 사용자 정의할 수 있을 뿐만 아니라 요소를 추가하는 JSL 스크립트를 작성할 수도 있습니다. 이러한 스크립트는 그래프를 표시할 때 실행됩니다. 스크립트를 처음부터 작성할 수도 있고, **새 스크립트 추가 버튼** ✚ 을 클릭하여 액세스할 수 있는 다음과 같은 명령 및 스크립트 목록에서 선택할 수도 있습니다.

- 단일 JSL 명령을 삽입하려면 " 템플릿 " 목록을 사용합니다 . 예를 들어 "Polygon" 옵션은 Polygon 명령을 삽입합니다. 밑줄로 둘러싸인 텍스트는 점 값의 자리 표시자이며, 사용자가 이를 원하는 값으로 바꿔야 합니다 .

  ```
  Polygon([_x0_, _x1_, ...], [_y0_, _y1_, ...]);
  ```

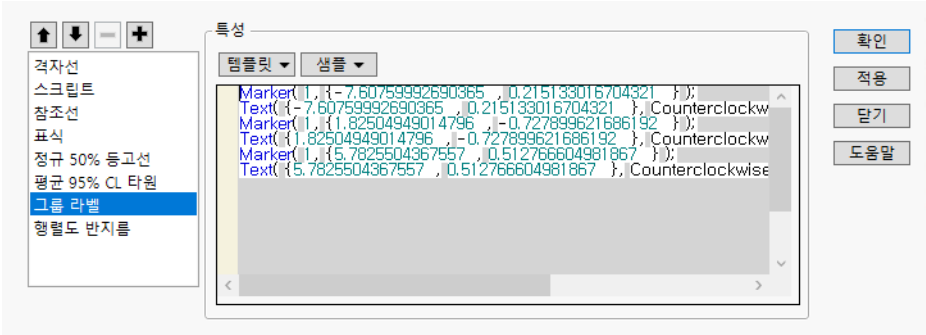
- 버블 그림 및 사인파 같은 요소를 생성하는 스크립트를 삽입하려면 " 샘플 " 목록을 사용합니다 . 이 목록에서 " 다각형 " 샘플 스크립트는 Transparency, Fill Color 및 Polygon 값의 예를 보여 주며 , 사용자가 이를 원하는 값으로 바꿔야 합니다 .

그림 9.36 다각형 샘플 스크립트



일부 그래프의 경우 그래픽 요소를 생성하는 JSL을 볼 수 있습니다. 그림 9.37에서는 판별 분석 그래프의 "그룹 라벨" 스크립트를 보여 줍니다. 이 스크립트는 그룹 라벨의 표식 및 텍스트를 정의합니다. 이 기본 제공 스크립트에 포함된 명령은 수정하거나 삭제할 수 없습니다. "템플릿" 또는 "샘플" 목록에서 명령을 삽입할 수 있지만 창 내부를 클릭하여 새 명령을 입력할 수는 없습니다.

그림 9.37 그룹 라벨을 생성하는 JSL 의 예



JSL에 대한 자세한 내용은 Scripting Guide 및 JSL Syntax Reference에서 확인하십시오.

그래픽 스크립트를 생성하려면 다음 단계를 수행하십시오.

1. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다.
2. **새 스크립트 추가** 버튼 (+) 을 클릭하여 새 스크립트를 생성합니다.

   기본 이름 " 스크립트 " 가 강조 표시됩니다.
3. 기본 이름이 강조 표시된 상태에서 보다 구분하기 쉬운 이름을 입력한 후 Enter 키를 누릅니다. 이미 커서를 이동해서 해당 이름이 더 이상 강조 표시되어 있지 않은 경우에는 " 스크립트 " 를 두 번 클릭한 후 새 이름을 입력합니다.
4. 다음 중 하나 이상을 수행합니다.
   - " 특성 " 창에 JSL 을 입력합니다.
   - " 템플릿 " 목록에서 JSL 명령을 하나 이상 선택하고 자리 표시자 텍스트를 수정합니다. 예를 들어 "Pen Color" 옵션을 "blue" 에서 "red" 로 변경합니다.

– " 샘플 " 목록에서 샘플 스크립트를 하나 이상 선택하고 , 필요한 경우 수정합니다 .

5. (선택 사항) 창을 닫지 않고 변경 사항으로 그래프를 업데이트하려면 **적용**을 클릭합니다 . 다른 특성을 변경할 때와 달리 스크립트 변경 사항은 **적용** 또는 **확인**을 클릭해야만 적용됩니다 .
6. **확인**을 클릭하여 변경 사항을 저장합니다 .

   다음 중 하나가 발생합니다 .

   – 생성한 요소가 그래프에 표시됩니다 .
   – 스크립트에 오류가 있는 경우 오류 메시지가 나타납니다 . 이 경우 **보기 > 로그**를 선택하여 오류를 자세히 확인하고 스크립트를 수정합니다 .

생성한 스크립트를 삭제하려면 스크립트를 선택한 후 **선택한 개체 삭제** 버튼 ( ) 을 선택합니다 .

그래픽 요소의 그리기 순서 변경

그래픽 요소는 목록에 나열된 순서대로 그려집니다 . 목록의 첫 번째 요소는 가장 먼저 그려지므로 다른 모든 그래픽 요소보다 **뒤쪽에** 표시됩니다 . 한 요소가 다른 요소를 가리는 경우 요소 순서를 조정하여 재배열할 수 있습니다 .

그래픽 요소를 재배열하려면

1. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다 .
2. 이동할 요소를 선택합니다 .
3. 요소가 원하는 그리기 순서로 설정될 때까지 **그리기 순서에서 위로 이동** 버튼 ( ) 또는 **그리기 순서에서 아래로 이동** 버튼 ( ) 을 한 번 이상 클릭합니다 .
4. **확인**을 클릭합니다 .

사용자 정의 설정 저장

그래프의 사용자 정의 설정은 현재 그래프에 적용되며 분석을 다시 실행할 때도 사용됩니다 . 나중에 사용자 정의 설정으로 그래프를 다시 생성하려면 빨간색 삼각형 메뉴에서 **스크립트 저장**을 선택한 후 저장 옵션 중 하나를 선택합니다 . 예를 들어 스크립트를 데이터 테이블에 저장하여 해당 스크립트를 실행할 때마다 사용자 정의한 특성을 적용할 수 있습니다 . 자세한 내용은 " 데이터 저장 및 공유 " 장의 "JMP 분석을 스크립트로 저장"(551 페이지)에서 확인하십시오 .

사용자 정의 복사

그래프의 사용자 정의를 다른 그래프에 복사할 수 있습니다 . 그러면 색상이 적용된 텍스트 또는 선과 같이 , 생성되거나 수정된 모든 개체를 다른 그래프에 붙여넣을 수 있습니다 .

그래프의 선을 다른 그래프에 복사하는 방법은 "그래프 내용 복사 및 붙여넣기"(465 페이지)에서 확인하십시오 .

다른 개체를 복사하려면

1. 사용자 요소가 있는 그래프를 마우스 오른쪽 버튼으로 클릭하고 **편집 > 사용자 정의 복사**를 선택합니다 .
2. 대상 그래프를 마우스 오른쪽 버튼으로 클릭하고 **편집 > 사용자 정의 붙여넣기**를 선택합니다 .

   해당 개체가 현재 그래프에 나타나고 사용자 정의 요소 목록에 추가됩니다 .

참고 :

- Ctrl 키를 누른 채로 클릭하여 붙여넣으면 유사한 모든 그래프에 사용자 정의를 붙여넣을 수 있습니다 .
- 사용자 정의 복사 기능은 사용자가 추가하거나 수정한 요소만 복사합니다. 그래프의 다른 내용은 복사하지 않습니다 .

JMP 그래프 또는 보고서에 이미지 추가

로고나 다른 유형의 그래픽으로 분석 보고서가 보다 돋보이게 하기 위해 그래프 또는 보고서에 그래픽을 붙여넣거나 드래그하여 놓을 수 있습니다. .bmp, .emf, .jpeg, .png, 및 .rtf 그래픽 형식이 지원됩니다.

- "JMP 그래프에 배경 이미지 붙여넣기 "
- "JMP 보고서 끝에 이미지 붙여넣기 "
- " 이미지를 JMP 그래프로 드래그하여 놓기 "
- "JMP 그래프에 지리 이미지 및 경계 추가 "
- " 이미지에서 데이터 추출 "

JMP 그래프에 배경 이미지 붙여넣기

1. 그래픽 프로그램에서 그래픽 파일을 열고 그래픽을 선택합니다 . 그래픽을 컴퓨터의 클립보드에 복사합니다 .
2. 그래프 내부를 마우스 오른쪽 버튼으로 클릭합니다 .
3. **편집 > 배경 이미지 붙여넣기**를 선택합니다 . 그래프에서 마우스 오른쪽 버튼을 클릭한 지점에 그래픽이 삽입됩니다 .

JMP 보고서 끝에 이미지 붙여넣기

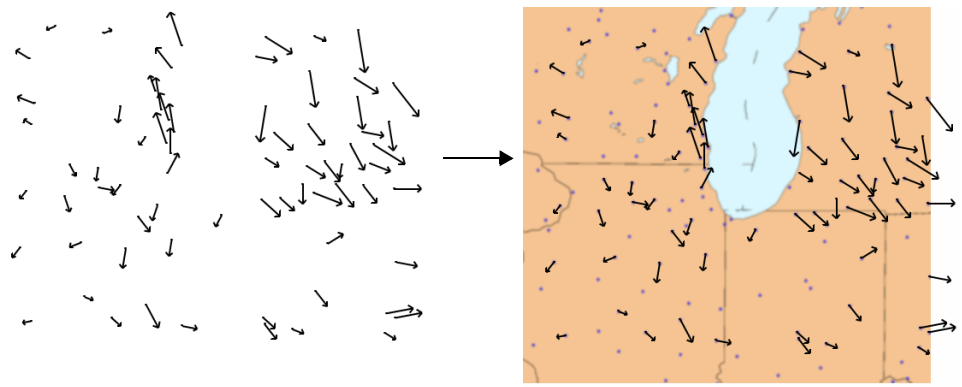
보고서 창의 끝에 그래픽을 추가할 수도 있습니다. 그래픽 프로그램에서 그래픽을 컴퓨터의 클립보드에 복사하고, 보고서 창을 표시한 후, **편집 > 붙여넣기**를 선택합니다.

이미지를 JMP 그래프로 드래그하여 놓기

사용자가 생성한 이미지를 추가하려면 보고서를 생성한 후 컴퓨터의 파일 시스템에 있는 해당 이미지를 보고서의 그래프로 드래그하여 놓습니다. 이미지를 추가한 후에는 JMP에서 이미지 크기 조정, 형식 지정, 회전 등을 위한 몇 가지 옵션이 제공됩니다. 이미지 옵션을 표시하려면 이미지를 마우스 오른쪽 버튼으로 클릭하고 **이미지**를 선택합니다.

다음 예에서는 시카고 지역의 풍속에 대한 이변량 그림을 보여 줍니다. 왼쪽의 그림에는 바람의 방향과 속도를 보여 주는 화살표가 포함되어 있습니다. 이 그림에 지도 이미지를 드래그하여 놓고 표식이 바람에 대한 데이터를 제공한 기상관측소(각 점이 기상관측소를 나타냄)와 일치하도록 크기를 조정했습니다.

그림 9.38 사용자 지도 이미지의 예



이미지 처리

JMP에는 다양한 이미지 처리 옵션이 포함되어 있습니다. 이미지 처리에는 이미지를 선명하게 하거나, 흐리게 하거나, 크기 조정하거나, 품질을 향상시킬 수 있는 필터가 포함됩니다. 또한 JMP에서는 이미지를 반전하거나, 회전하거나, 잠글 수도 있습니다. 이 섹션에서는 이러한 처리 옵션에 대해 설명합니다.

참고 :

- 이미지를 반복적으로 변경할 경우 이미지 품질이 저하될 수 있습니다 . 이미지의 크기 조정 , 척도 조정 및 필터 적용 결과가 만족스럽지 않은 경우에는 해당 이미지를 제거하고 다시 시작하십시오 .
- 모든 JMP 이미지 필터는 운영 체제 수준에서 지원되므로 Windows 와 macOS 에서 이미지가 서로 다르게 처리될 수 있습니다 .

잠금

이미지를 이동할 수 없도록 현재 위치에 잠급니다.

크기 / 척도

- **그래프 채우기** : 이미지 크기를 그래프 크기에 맞게 조정합니다 .
- **크기 지정** : 입력한 값에 따라 이미지 크기를 조정합니다 . 값 단위는 그래프 축과 동일합니다 .
- **자르기** : 입력한 값에 따라 이미지를 자릅니다 . 값 단위는 그래프 축과 동일합니다 . 예를 들어 왼쪽 모서리가 50 위치에 있을 수 있습니다 . " 왼쪽 " 옆에 60 을 입력하면 50 에서 60 사이의 이미지 부분이 이미지에서 제거됩니다 .

반전

- **상하 반전** : 이미지를 상하로 반전합니다 .
- **좌우 반전** : 이미지를 좌우로 반전합니다 .
- **상하좌우 반전** : 이미지를 상하좌우로 반전합니다 .

회전

이미지를 지정된 각도만큼 회전합니다. 이미지를 반시계 방향으로 회전하려면 음수 값을 입력합니다.

투명도

이미지의 투명도 수준을 변경합니다. 0에서 1 사이의 값을 입력합니다(1은 완전 불투명을 나타냄).

필터

많은 그래픽 편집 프로그램에서 이미지의 모양을 변경하는 데 사용되는 필터를 제공합니다. 특정 필터를 반복해서 선택하면 이미지에서 해당 효과가 증가합니다.

- **대비** : 밝고 어두운 색상을 최적화합니다 . 값이 클수록 이미지가 밝아집니다 .
- **얼룩 제거** : 주변 픽셀과 이질적인 픽셀을 제거합니다 . 예를 들어 사방이 흰색 픽셀로 둘러싸인 검은색 픽셀은 흰색 픽셀로 변환됩니다 .
- **모서리** : 개체의 외곽선을 제외한 모든 요소를 어둡게 합니다 .
- **향상** : 잡음이 심한 이미지의 픽셀 간 대비를 줄입니다 .
- **감마** : 이미지의 밝기와 RGB(빨간색 , 녹색 , 파란색) 비율 간에 균형을 맞춥니다 . 값이 클수록 이미지가 밝아집니다 .
- **가우시안 흐림** : 지정한 반지름만큼의 픽셀을 흐리게 표시합니다 . 반지름이 클수록 부드러운 이미지가 생성됩니다 . JSL 에서는 시그마 값을 지정할 수도 있습니다 . 시그마 값이 클수록 부드러운 이미지가 생성됩니다 .
- **중앙값** : 각 픽셀의 색상 값을 주변 픽셀의 중앙값으로 바꿉니다 .
- **반전** : 각 픽셀을 보색으로 변환합니다 (예 : 분홍색은 녹색으로 , 흰색은 검은색으로).
- **정규화** : 최상위 및 최하위 색상 값의 백분율을 제거합니다 . 그러면 색상 값이 늘어나서 나머지 이미지를 채웁니다 . 이 프로세스는 색상의 강도를 높입니다 .

- **잡음 감소** : 최소 및 최대 색상 값을 찾아 주변 픽셀과 보다 일관된 값으로 바꿉니다 . 값이 클수록 부드러운 이미지가 생성됩니다 .
- **선명하게** : 픽셀 모서리를 보다 분명하게 표시합니다 .

제거

보고서에서 이미지를 제거합니다.

JMP 그래프에 지리 이미지 및 경계 추가

그래프에 지도 이미지 및 행정 경계를 추가하면 지리적 데이터의 맥락을 시각적으로 파악할 수 있습니다. 예를 들어 그래프에 미국 이미지를 표시하는 지도를 추가할 수 있습니다. 데이터에 미국의 위도 및 경도가 포함된 경우에는 각 주의 경계를 표시하는 방법도 있습니다.

데이터에 위도 및 경도 좌표가 포함되어 있어야 합니다. 그렇지 않으면 지도가 있어도 데이터와 관련된 의미를 파악할 수 없습니다. 또한 X 및 Y 축에는 지도 유형에 따른 범위 요구 사항이 있습니다. 이후 섹션에서는 이러한 요구 사항에 대해 설명합니다.

지도를 표시할 때는 다음 두 가지 도구가 특히 유용합니다.

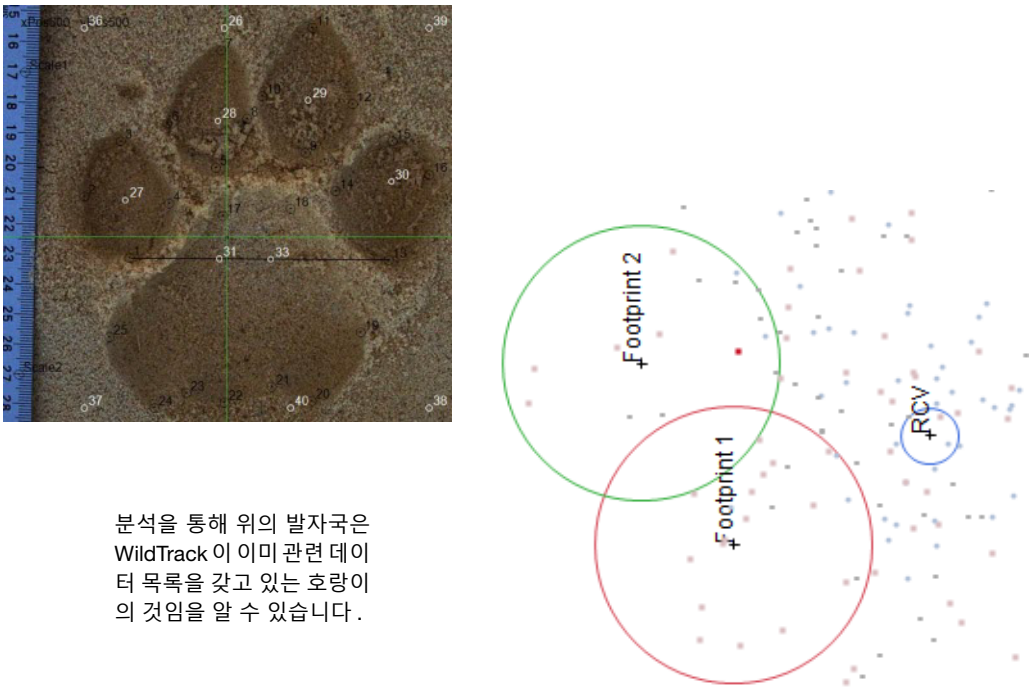
- 손 도구 ( ) 는 지도에서 가로 및 세로로 스크롤하는 데 사용할 수 있습니다 .
- 돋보기 도구 ( ) 는 확대 / 축소에 사용할 수 있습니다 .

맵을 추가하는 방법에 대한 자세한 내용은 Essential Graphing 의 에서 확인하십시오.

이미지에서 데이터 추출

JMP에서는 이미지에서 정보를 추출하여 데이터 테이블에 추가한 후 해당 정보를 분석하는 기능을 제공합니다. WildTrack.org 의 연구원들은 디지털 발자국 사진을 JMP 에서 분석하여 멸종 위기종 동물을 추적합니다. 연구원들은 발자국 이미지를 JMP 보고서로 드래그하여 놓은 다음 발자국의 크기와 형태를 나타내는 데이터 점을 그립니다. JSL(JMP 스크립트 언어) 스크립트는 이 측정값을 데이터 테이블에 추출합니다. 그러면 데이터를 분석하여 발자국이 새로운 동물의 것인지 여부를 확인할 수 있습니다. 이 방법은 전 세계의 특정 지역에서 멸종 위기종 동물의 모집단을 추적하는 데 유용합니다.

그림 9.39 데이터 추출 및 분석의 예



이 기능을 구현하는 자세한 방법은 사용자 요구 사항에 따라 매우 다양하므로 JMP 사용에서 다루기에는 적합하지 않습니다. JSL 스크립트 작성에 대한 자세한 내용은 Scripting Guide 및 JSL Syntax Reference에서 확인하십시오.

JMP 보고서 및 그래프에 주석 및 형태 추가

도구 모음에 있는 그래픽 도구를 사용하여 보고서 또는 그래프에 텍스트 노트 , 선 , 다각형 , 타원 및 직사각형을 추가할 수 있습니다 .

JMP 보고서 및 그래프에 주석 추가

주석 달기 도구 를 사용하여 JMP 보고서 또는 그래프에 편집 가능한 텍스트 노트를 추가할 수 있습니다. 주석을 추가하려면 다음 단계를 수행하십시오.

1. " 도구 " 메뉴 또는 도구 모음에서 주석 달기 도구 를 선택합니다 .
2. 보고서 또는 그래프에서 주석을 추가할 위치를 클릭합니다 . 또는 클릭 후 드래그하여 주석 노트의 크기를 지정합니다 . 편집 가능한 흰색 텍스트 상자가 나타납니다 .
3. 텍스트를 입력합니다 .
4. 주석 외부를 클릭합니다 .

5. (선택 사항) 태그 선을 추가하려면 주석을 마우스 오른쪽 버튼으로 클릭하고 **태그 선**을 선택합니다 .

   또는 다음 옵션을 편집할 수 있습니다 .

   배경 색상 주석의 배경 색상을 선택합니다 .

   텍스트 색상 텍스트 색상을 선택합니다 . 텍스트 색상은 주석 외곽선의 색상도 정의합니다 .

   글꼴 글꼴 유형 , 스타일 및 크기를 변경합니다 . 기본 글꼴을 변경하는 방법은 "JMP 환경 설정 " 장의 " 글꼴 환경 설정 "(655 페이지) 에서 확인하십시오 .

   태그 선 주석과 주석이 고정된 점 또는 텍스트 사이에 선을 그립니다 .

   채우기 주석이 투명하게 보이도록 주석에서 배경 색상을 제거합니다 . 투명한 노트는 그래프에 제목 및 각주를 추가하는 데 유용합니다 .

   편집 가능 주석을 두 번 클릭하여 편집할 수 있도록 설정합니다 .

   고정 크기 상자에 텍스트를 추가할 때 주석 상자의 크기가 조정되지 않습니다 .

   재고정 (주석이 이동되는 경우에만 사용 가능) 주석은 보고서 내의 위치에 고정되지만 고정 위치와 독립적으로 이동할 수 있습니다 . 주석의 동작은 고정 위치에 따라 다릅니다 . 예를 들어 그래프 내부에 고정된 주석은 축과 함께 이동하고 외곽선 내부에 고정된 주석은 외곽선이 열리거나 닫힐 때 표시되거나 숨겨집니다 . 주석을 재고정하려면 태그 선을 해제하고 주석을 새 위치로 이동한 다음 마우스 오른쪽 버튼으로 클릭하여 재고정합니다 .

   삭제 전체 주석을 삭제합니다 .

   참고 : 주석을 여러 개 추가하려면 주석 달기 도구를 처음 선택할 때 Shift 키를 누릅니다 . 그러면 이후에는 각 주석을 추가할 때마다 도구 모음에서 주석 달기 도구를 선택할 필요 없이 마우스 클릭만으로 주석을 추가할 수 있습니다 .

주석을 추가한 후에는 다음 작업을 수행할 수 있습니다.

표 9.3 주석 사용

작업	방법
주석 내용 추가 또는 편집	텍스트 상자 내부를 클릭합니다.
주석 이동	주석 상자 내부를 클릭한 채로 드래그합니다. 주석을 이동하면 주석이 선택 상태(경계를 따라 핸들이 있는 파란색 이중선으로 표시)가 됩니다.
주석 크기 조정	선택된 노트의 핸들(모서리의 가운데와 모퉁이에 표시)에 커서를 놓습니다. 커서가 양방향 화살표로 표시되면 드래그하여 주석 크기를 조정합니다.

표 9.3 주석 사용 (계속)

작업	방법
주석 삭제	노트의 핸들을 클릭하여 주석을 강조 표시한 다음 Delete 또는 Backspace 키를 누릅니다.
주석의 태그 선 이동	1. 주석을 마우스 오른쪽 버튼으로 클릭하고 **태그 선**을 선택 취소합니다 . 2. 태그 선을 시작할 위치 (첨부 점) 로 주석을 드래그합니다 . 3. 주석을 마우스 오른쪽 버튼으로 클릭하고 **태그 선**을 선택합니다 . 이제 태그 선은 주석을 이동한 점에서 시작됩니다 . 4. 태그 선을 표시할 위치로 주석을 이동합니다 .
주석으로 태그 선 이동	Ctrl 키를 누른 채 주석을 드래그하여 태그 선 끝점을 주석과 함께 이동합니다.

JMP 보고서에 형태 추가

그리기 도구 를 사용하여 편집 가능한 선, 다각형 및 단순 형태(타원 또는 직사각형)를 JMP 보고서에 추가할 수 있습니다. 다음 섹션에서는 이러한 각 도구를 사용하는 방법에 대해 설명합니다.

참고 : 각 그래픽 도구는 최근에 선택된 옵션을 기억합니다. 따라서 동일한 특성을 갖는 주석 또는 기타 그래픽이 많이 필요한 경우에 유용합니다. 예를 들어 한쪽 끝에 화살표가 있는 두꺼운 녹색 선이 많이 필요하다고 가정해 보겠습니다. 원하는 방법으로 선택 생성하고 옵션을 설정하면 이후 선은 해당 옵션이 적용된 상태로 나타납니다 . 옵션은 사용자가 변경하기 전까지 유지됩니다 .

선 추가

보고서 창에 선 주석을 추가합니다.

1. 도구 팔레트에서 선 도구 를 클릭합니다 .
2. 선을 삽입할 위치를 클릭하고 드래그합니다 . 선이 양쪽 끝에 핸들이 표시된 선택 상태로 나타납니다 .
3. 선을 이동하려면 선을 클릭하여 드래그합니다 .
4. 선을 회전하려면 핸들을 클릭하여 드래그합니다 .
5. 선 모양을 사용자 정의하기 위한 옵션이 포함된 메뉴를 표시하려면 선을 마우스 오른쪽 버튼으로 클릭합니다 .

끝 점 및 시작 점 선의 한쪽 끝 (또는 양쪽 끝) 에 화살표를 추가합니다 .

굵게 선의 너비를 두껍게 변경합니다. 기본적으로 얇은 선이 사용됩니다.

파선 선을 파선으로 변경합니다. 기본적으로 실선이 사용됩니다.

색상 형태의 색상을 변경하기 위해 JMP 색상 팔레트를 표시합니다.

재고정 선 주석은 보고서 내의 위치에 고정되지만 고정 위치와 독립적으로 이동할 수 있습니다. 동작은 고정 위치에 따라 다릅니다. 예를 들어 그래프 내부에 고정된 선 주석은 축과 함께 이동하고 외곽선 내부에 고정된 선 주석은 외곽선이 열리거나 닫힐 때 표시되거나 숨겨집니다. 선 주석을 재고정하려면 선을 새 위치로 이동하고 마우스 오른쪽 버튼을 클릭하여 재고정합니다. 주석이 그래프 바깥쪽에 위치한 경우에만 사용할 수 있습니다.

삭제 형태를 보고서 화면에서 제거합니다. 형태를 선택한 다음 Delete 또는 Backspace 키를 눌러 제거할 수도 있습니다.

다각형 또는 베지어 곡선 추가

보고서 창에 다각형 (또는 베지어 곡선) 주석을 추가하려면

1. 도구 팔레트에서 다각형 도구 를 클릭합니다.
2. 다각형의 첫 번째 모서리를 그릴 시작점을 클릭하여 생성합니다.
3. 첫 번째 모서리를 완성할 위치에서 다시 클릭하고 인접한 모서리를 시작합니다. 다각형 영역 주위에는 핸들이 있는 사각형 선택 상자가 나타납니다.
4. 세 번째로 마우스를 클릭하여 두 번째 모서리를 완성합니다.
5. 원하는 대로 다각형이 생성될 때까지 이 과정을 계속합니다. 모서리가 완성될 때마다 다각형 모서리를 둘러싸기 위해 선택 상자가 조정됩니다.
6. 두 번 클릭하여 다각형 도구를 해제합니다.

다각형을 추가한 후에는 다음 작업을 수행할 수 있습니다.

표 9.4 다각형 작업

작업	방법
다각형 선택 또는 선택 취소	완성된 다각형의 모서리를 클릭합니다.
다각형 크기 조정	다각형을 선택하고 선택 상자의 핸들 중 하나를 드래그합니다.
다각형 이동	상자의 핸들 사이를 클릭하고 선택 상자를 드래그합니다.
다각형의 모서리 수 변경	모서리를 클릭한 채로 드래그하여 새 형태를 만듭니다.

다각형의 모양을 사용자 정의하는 옵션을 보려면 다각형을 마우스 오른쪽 버튼으로 클릭합니다.

채우기 형태 영역을 채우거나 비웁니다.

양각 형태 주위에 음영이 적용된 두꺼운 선을 표시합니다 . 형태를 색상으로 채우기도 한 경우 도형의 아래쪽 모서리가 양각으로 표시되어 3 차원적인 모양이 만들어집니다 .

매끄러운 다각형의 꼭지점을 부드럽게 해서 베지어 곡선을 생성합니다 . 부드럽게 다듬어진 도형은 다각형과 동일한 방식으로 형태 및 크기를 조정할 수 있으며 색상을 채우거나 양각 처리할 수도 있습니다 .

닫기 다각형의 마지막 선분을 열거나 닫습니다 .

색상 **채우기** 옵션이 적용된 경우 형태의 모서리 색상과 채우기 색상을 변경하기 위한 JMP 색상 팔레트를 표시합니다 .

재고정 다각형 주석은 보고서 내의 위치에 고정되지만 고정 위치와 독립적으로 이동할 수 있습니다 . 동작은 고정 위치에 따라 다릅니다 . 예를 들어 그래프 내부에 고정된 다각형 주석은 축과 함께 이동하고 외곽선 내부에 고정된 다각형 주석은 외곽선이 열리거나 닫힐 때 표시되거나 숨겨집니다 . 다각형 주석을 재고정하려면 다각형을 새 위치로 이동하고 마우스 오른쪽 버튼을 클릭하여 재고정합니다 .

삭제 형태를 보고서에서 제거합니다 . 형태를 선택한 다음 Delete 또는 Backspace 키를 눌러 제거할 수도 있습니다 .

타원 또는 직사각형 추가

보고서 창에 타원 또는 직사각형 주석을 추가하려면

1. 도구 팔레트에서 원모양 도구 를 클릭합니다 .
2. 형태를 삽입할 위치를 클릭하고 드래그합니다 . 선택 상자로 둘러싸인 타원이 나타납니다 .
3. (선택 사항) 타원을 직사각형으로 전환하려면 타원을 마우스 오른쪽 버튼으로 클릭하고 **형태 > 직사각형**을 선택합니다 .
4. 두 번 클릭하여 원 모양 도구를 해제합니다 .

형태를 추가한 후에는 다음을 수행할 수 있습니다.

- 형태를 선택하고 선택 상자 핸들 중 하나를 드래그하여 형태 또는 크기를 조정합니다 .
- 선택 상자의 모서리 (핸들 사이) 를 클릭한 후 드래그하여 이동합니다 .
- 형태를 선택한 후 마우스 오른쪽 버튼을 클릭하여 모양을 사용자 정의하기 위한 옵션을 표시합니다 .

  채우기 형태 영역을 채우거나 비웁니다 .

  양각 형태 주위에 음영이 적용된 두꺼운 선을 표시합니다 . 형태를 색상으로 채우기도 한 경우 도형의 아래쪽 모서리가 양각으로 표시되어 3 차원적인 모양이 만들어집니다 .

  형태 선택 시 형태를 타원 또는 직사각형으로 변환하는 옵션이 포함된 하위 메뉴를 표시합니다 .

  색상 채우기 옵션이 적용된 경우 형태의 모서리 색상과 채우기 색상을 변경하기 위한 JMP 색상 팔레트를 표시합니다 .

재고정 형태 주석은 보고서 내의 위치에 고정되지만 고정 위치와 독립적으로 이동할 수 있습니다. 동작은 고정 위치에 따라 다릅니다. 예를 들어 그래프 내부에 고정된 형태 주석은 축과 함께 이동하고 외곽선 내부에 고정된 형태 주석은 외곽선이 열리거나 닫힐 때 표시되거나 숨겨집니다. 형태 주석을 재고정하려면 형태를 새 위치로 이동하고 마우스 오른쪽 버튼을 클릭하여 재고정합니다.

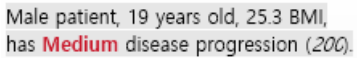
삭제 형태를 보고서에서 제거합니다. 형태를 선택한 다음 Delete 또는 Backspace 키를 눌러 제거할 수도 있습니다.

그래프의 가리키기 라벨 사용자 정의

그래프의 시각적 요소를 마우스로 가리키면 대개 해당 데이터 점에 대한 설명이 가리키기 라벨에 표시됩니다. 다음 방법 중 하나로 가리키기 라벨에 나타나는 내용을 사용자 정의합니다.

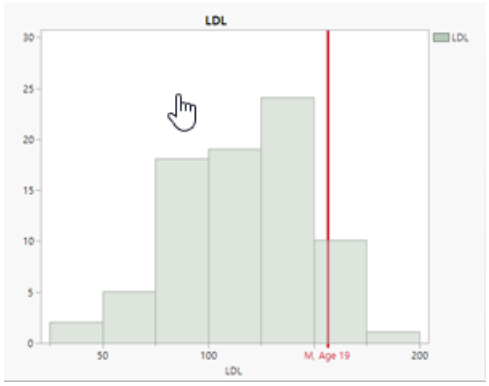
Graphlet을 사용하여 대화식 그래프 또는 이미지를 추가합니다.	자세한 내용은 "가리키기 라벨에 그래프 또는 이미지 추가"(483페이지)에서 확인하십시오.
Textlet을 사용하여 형식이 지정된 텍스트를 추가합니다.	자세한 내용은 "그래프의 가리키기 라벨에 서식 있는 텍스트 추가"(498페이지)에서 확인하십시오.
Gridlet을 사용하여 이름과 값을 추가, 제거 또는 업데이트합니다.	자세한 내용은 "가리키기 라벨의 이름 및 값 업데이트"(501페이지)에서 확인하십시오.
가리키기 라벨 편집기의 사용 방법을 자세히 알아봅니다.	자세한 내용은 "그래프의 가리키기 라벨 편집"(511페이지)에서 확인하십시오.
가리키기 라벨 내의 이름 및 값을 업데이트하거나, 그래프에 라벨을 추가합니다.	자세한 내용은 "데이터 테이블에서 가리키기 라벨에 이미지 추가"(516페이지)에서 확인하십시오.
데이터 테이블에서 이미지를 추가합니다.	자세한 내용은 "데이터 테이블에서 가리키기 라벨에 이미지 추가"(516페이지)에서 확인하십시오.
그래프에 가리키기 라벨을 고정합니다.	자세한 내용은 "그래프의 가리키기 라벨 고정"(517페이지)에서 확인하십시오.
가리키기 라벨의 형식을 지정하고 추가 옵션에 액세스합니다.	자세한 내용은 "가리키기 라벨의 마우스 오른쪽 버튼 클릭 옵션"(518페이지)에서 확인하십시오.

그림 9.40 가리키기 라벨에 표시되는 그래프 , 텍스트 , 이름 및 값의 예



Textlet 을 사용하여 생성된 텍스트

Gridlet 을 사용하여 생성된 이름 및 값



Graphlet 을 사용하여 생성된 가리키기 그래프

가리키기 라벨에 그래프 또는 이미지 추가

Graphlet 을 사용하여 가리키기 라벨에 그래프 또는 이미지의 썸네일을 추가할 수 있습니다.

- 사전 설정 Graphlet: 사전 설정 Graphlet 은 사전 정의되어 있으며 편집 가능합니다 . 그래프를 마우스 오른쪽 버튼으로 클릭하고 **가리키기 라벨** > **그래프 유형**을 선택하기만 하면 손쉽게 사전 설정 Graphlet 를 추가할 수 있습니다 . 또는 가리키기 라벨 편집기에서 사전 설정 Graphlet 을 추가할 수도 있습니다 .

  참고 : 시뮬레이션에 사용되는 데이터와 같이 데이터 테이블에 연결되지 않은 합성 데이터를 기반으로 한 그래프는 사전 설정 그래프를 지원하지 않습니다 .

- Graphlet 붙여넣기 : 그래프 빌더에서 그래프를 생성합니다 . 스크립트를 클립보드에 저장하는 방법으로 복사한 다음 기준 그래프에 Graphlet 으로 붙여넣습니다 .

가리키기 그래프를 추가한 후에는 기준 그래프의 요소 위에 커서를 놓습니다. 그러면 해당 요소에 대한 가리키기 그래프의 썸네일 미리보기가 표시됩니다. 예를 들어 집계 데이터의 경우 기본 데이터 분포의 가리키기 그래프를 볼 수 있습니다.

일부 JMP 플랫폼에는 주성분, 공정 변수 선별, 모형 기반 다변량 관리도, 함수 데이터 탐색기와 같은 기본 제공 가리키기 라벨 그래프가 있습니다. 별도의 작업 없이도 이러한 그래프를 시각화 요소에 추가할 수 있습니다. 시각화 요소를 마우스로 가리키면 가리키기 라벨에 자동으로 그래프가 나타납니다.

JSL 을 사용하여 가리키기 라벨 그래프를 사용자 정의할 수도 있습니다. 자세한 내용은 Scripting Guide 의 에서 확인하십시오.

가리키기 그래프 작업

- 가리키기 그래프를 클릭하여 새 창에서 해당 그래프를 시작하거나 , 이미지를 클릭하여 관련 링크를 열 수 있습니다 .
- Ctrl 키를 누른 채 가리키기 라벨 그래프를 클릭하면 기준 그래프를 가리키기 라벨 그래프로 바꿀 수 있습니다 .
- Alt 키를 누른 채 가리키기 라벨 그래프를 클릭하면 새 창을 여는 대신 마지막으로 시작한 창의 내용을 바꿀 수 있습니다 .

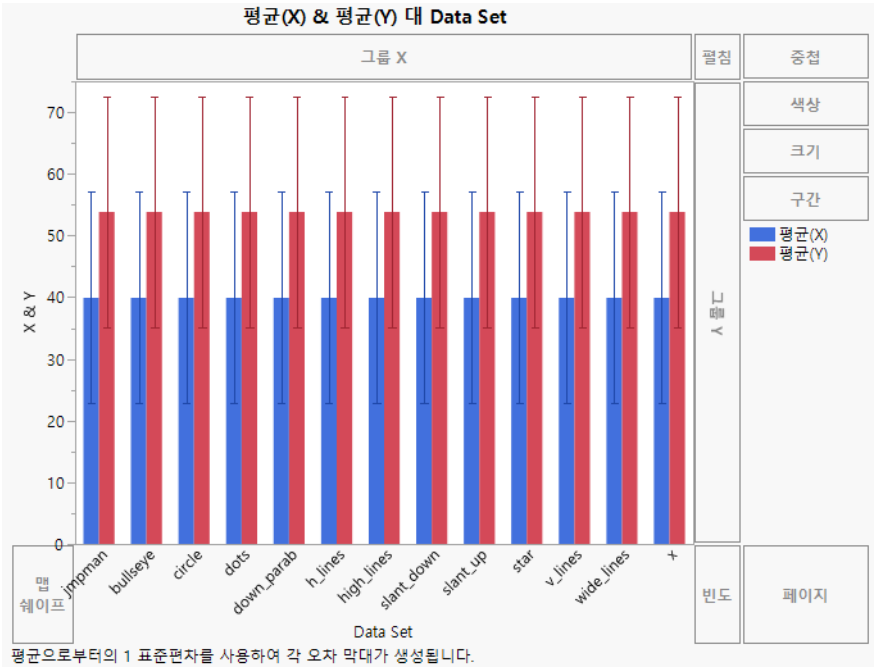
예 : 가리키기 라벨에 사전 설정 그래프 추가

이 예에서는 그래프를 생성하고, 가리키기 라벨에 사전 설정 그래프를 추가한 다음, 가리키기 라벨 그래프를 사용자 정의합니다.

원래 그래프 생성

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 JMP Man Dozen.jmp 를 엽니다 .
2. **그래프** > **그래프 빌더**를 선택합니다 .
3. X 및 Y 를 선택하고 **Y** 영역으로 드래그합니다 .
4. Data Set 를 선택하고 **X** 영역으로 드래그합니다 .
5. 막대 요소 를 클릭합니다 .
6. " 오차 구간 " 에서 **표준편차**를 선택합니다 .

그림 9.41 JMP Man Dozen.jmp 의 원래 그래프



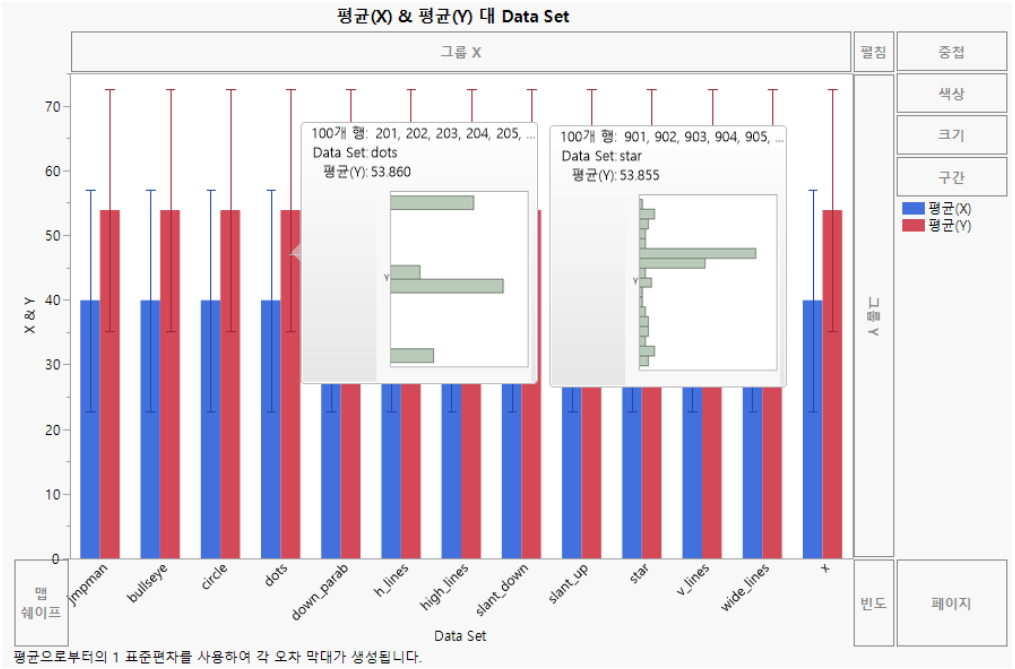
이 관측값들은 평균 및 표준편차가 동일하므로 이 그래프로는 많은 것을 알 수 없습니다. 사전 설정 그래프를 사용하면 데이터에 대한 더 많은 정보를 얻을 수 있습니다.

사전 설정 그래프 추가

1. 그래프에서 마우스 오른쪽 버튼을 클릭하고 **가리키기 라벨** > **히스토그램**을 선택합니다 .
   그러면 가리키기 라벨에 사전 설정 히스토그램이 추가됩니다 .
2. 일부 막대를 가리킵니다 . 이 예에서는 dots 및 star 의 빨간색 막대를 가리킵니다 .

팁 : 가리키기 라벨을 다른 가리키기 라벨과 비교할 수 있도록 화면에 유지하려면 오른쪽 상단의 고정 아이콘 📌을 클릭하십시오 . 자세한 내용은 " 그래프의 가리키기 라벨 고정 "(517 페이지) 에서 확인하십시오 .

그림 9.42 사전 설정 히스토그램의 예



측정값의 분포가 완전히 다르다는 것을 알 수 있습니다. 하지만 다른 유형의 그래프, 즉 산점도도 확인하려고 합니다.

3. 그래프에서 마우스 오른쪽 버튼을 클릭하고 **가리키기 라벨** > **점**을 선택합니다.

   그러면 사전 설정 그래프가 히스토그램에서 산점도로 변경됩니다.

4. down_parab 의 파란색 막대와 down_parab 의 빨간색 막대를 마우스로 가리킵니다. 축의 Y 변수와 X 변수가 서로 바뀌어 있음에 유의하십시오.

   모든 그래프에 동일한 X 및 Y 역할 할당을 사용하려고 합니다. 즉, X 축에 X 변수를 할당하고, Y 축에 Y 변수를 할당하려고 합니다. 또한 평활기를 적합선으로 변경하고, 산점도에 요약 통계량이 포함된 캡션 상자를 추가하려고 합니다.

   그래프 빌더에서 그래프를 사용자 정의한 다음 이 그래프를 템플릿으로 복사하여 이 그래프의 모든 가리키기 라벨에 적용할 수 있습니다.

그래프 사용자 정의

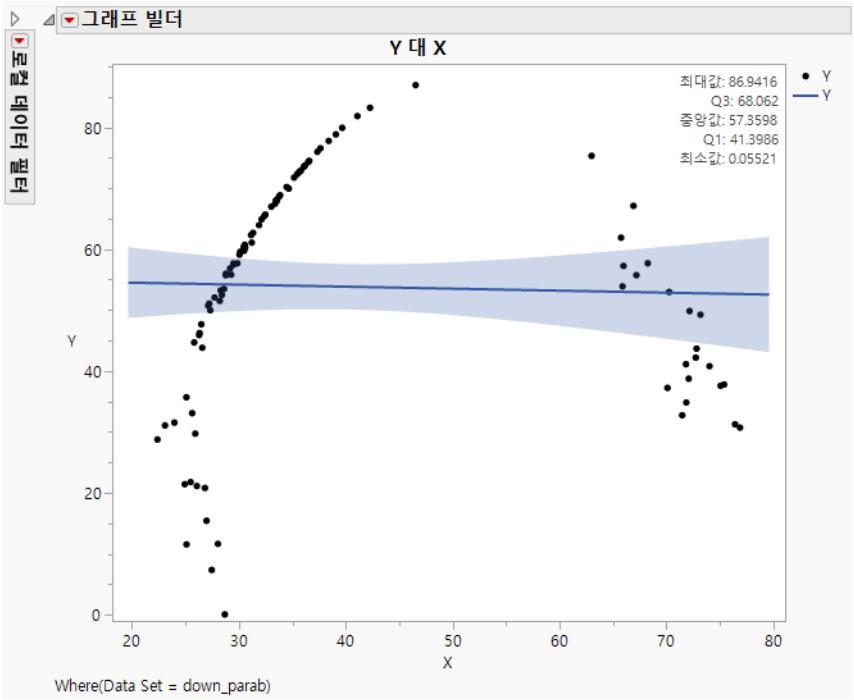
1. down_parab 의 파란색 막대를 마우스로 가리키고 해당 그래프를 클릭하여 새 창에서 엽니다.

   해당 그래프가 그래프 빌더에서 열립니다.

참고 : 사전 설정 그래프를 시작하면 왼쪽에 " 로컬 데이터 필터 " 가 표시됩니다. 로컬 데이터 필터를 사용하여 그래프를 확장하거나 수정할 수 있습니다. 자세한 내용은 " 보고서의 로컬 데이터 필터 "(413 페이지) 에서 확인하십시오.

2. " 그래프 빌더 " 의 빨간색 삼각형을 클릭하고 **제어판 표시**를 선택합니다 .
3. 적합선 요소 를 클릭합니다 .
4. 캡션 상자 아이콘 을 클릭합니다 .
5. " 캡션 상자 " 에서 " 요약 통계량 " 을 **5 가지 숫자 요약**으로 변경합니다 .
6. **완료**를 클릭합니다 .

그림 9.43 down_parab 에 대한 사용자 정의 그래프



7. " 그래프 빌더 " 의 빨간색 삼각형 메뉴를 클릭하고 **스크립트 저장** > **클립보드에**를 선택합니다 .
8. 이 그래프 빌더 창을 닫습니다 .
9. 원래 그래프에서 마우스 오른쪽 버튼을 클릭하고 **가리키기 라벨** > **Graphlet 붙여넣기**를 선택합니다 .

   그러면 이 사용자 정의 그래프가 모든 가리키기 라벨 그래프에 템플릿으로 적용됩니다. 일부 막대를 마우스로 가리켜서 변경 사항을 확인합니다 .

팁 : 원래 그래프와 가리키기 라벨에서 시작된 그래프를 더 잘 비교하려면 프로젝트를 사용하십시오 . 자세한 내용은 10 장 , " 데이터 저장 및 공유 " 에서 확인하십시오 .

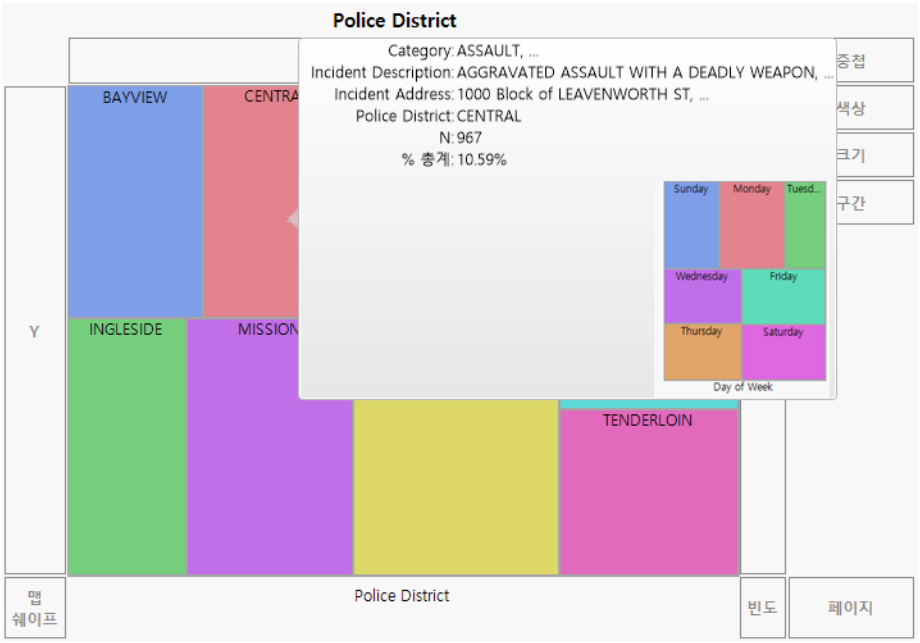
예 : 가리키기 라벨의 드릴다운 열 계층 지정

범주형 열을 사용하는 사전 설정 그래프에서는 필터링 기준으로 사용할 열을 지정할 수 있습니다.

원래 그래프 생성

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 San Francisco Crime.jmp 를 엽니다 .
   이 데이터는 샌프란시스코의 경찰 관할구별 범죄 통계량을 보여 줍니다 .
2. **그래프** > **그래프 빌더**를 선택합니다 .
3. Police District 를 선택하여 **X** 영역으로 드래그합니다 .
4. 트리맵 요소 를 클릭합니다 .
   샌프란시스코의 경찰 관할구별 트리맵이 표시됩니다 .
5. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **가리키기 라벨** > **트리맵**을 선택합니다 .
6. "Central" 위에 커서를 놓아 가리키기 라벨 그래프를 표시합니다 .

그림 9.44 Central 에 대한 초기 가리키기 라벨 그래프



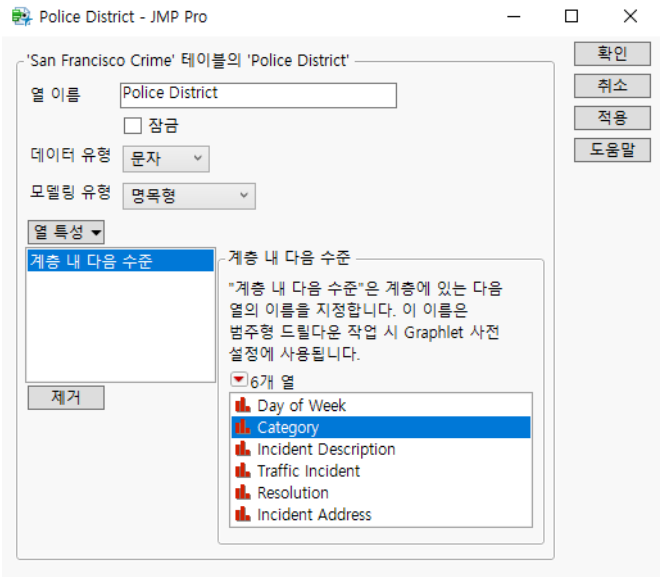
이 가리키기 라벨 그래프는 각 관할구의 요일별 발생 사건 수를 보여 줍니다 . 이 예에서는 그 대신 각 관할구의 범주별 발생 사건 수를 보려고 합니다 .

드릴다운 열 계층 지정

1. San Francisco Crime.jmp 데이터 테이블에서 Police District 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **열 특성** > **계층 내 다음 수준**을 선택합니다 .

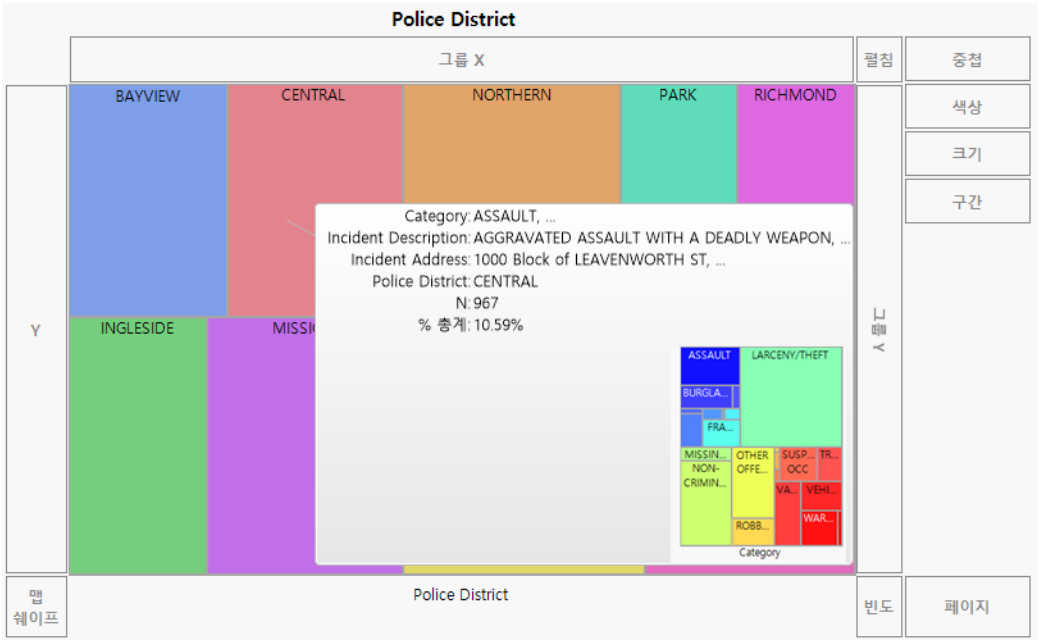
2. " 계층 내 다음 수준 " 에서 Category 를 클릭합니다 .

그림 9.45 계층 내 다음 수준으로 선택된 Category



3. **확인**을 클릭합니다 .
4. 첫 번째 가리키기 라벨 그래프가 실행될 경우에 표시되는 계층의 두 번째 수준을 추가합니다.
   a. San Francisco Crime.jmp 데이터 테이블에서 Category 열을 마우스 오른쪽 버튼으로 클릭합니다 .
   b. **열 특성** > **계층 내 다음 수준**을 선택합니다 .
   c. " 계층 내 다음 수준 " 에서 Incident Description 을 클릭합니다 .
   d. **확인**을 클릭합니다 .
5. 그래프 빌더에서 다시 "Central" 위에 커서를 놓아 업데이트된 가리키기 라벨 그래프를 표시합니다 .

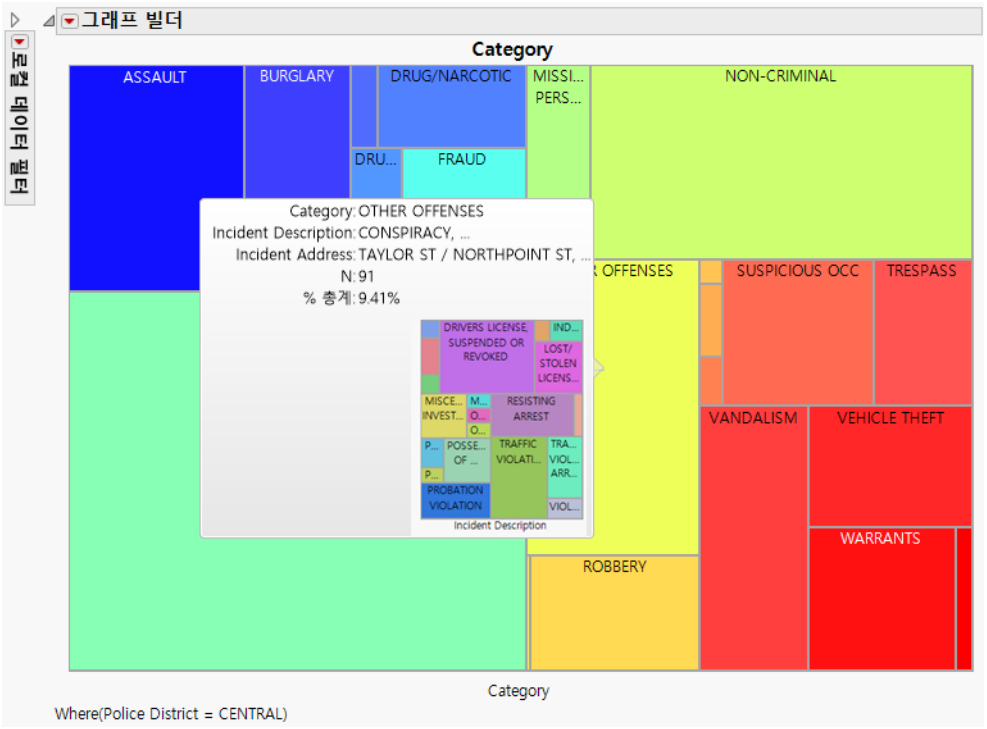
그림 9.46 Central 에 대한 업데이트된 가리키기 라벨 그래프



이 가리키기 라벨 그래프는 이제 각 관할구의 범주별 발생 사건 수를 보여 줍니다 .

6. 가리키기 라벨 그래프를 클릭하여 시작합니다 .
7. "Other Offenses" 위에 커서를 놓습니다 .

그림 9.47 Other Offenses 범주에 대한 두 번째 수준 가리키기 라벨 그래프



두 번째 수준의 가리키기 라벨 그래프에서는 "Other Offenses" 범주의 사건 설명별 사건 수를 보여 줍니다.

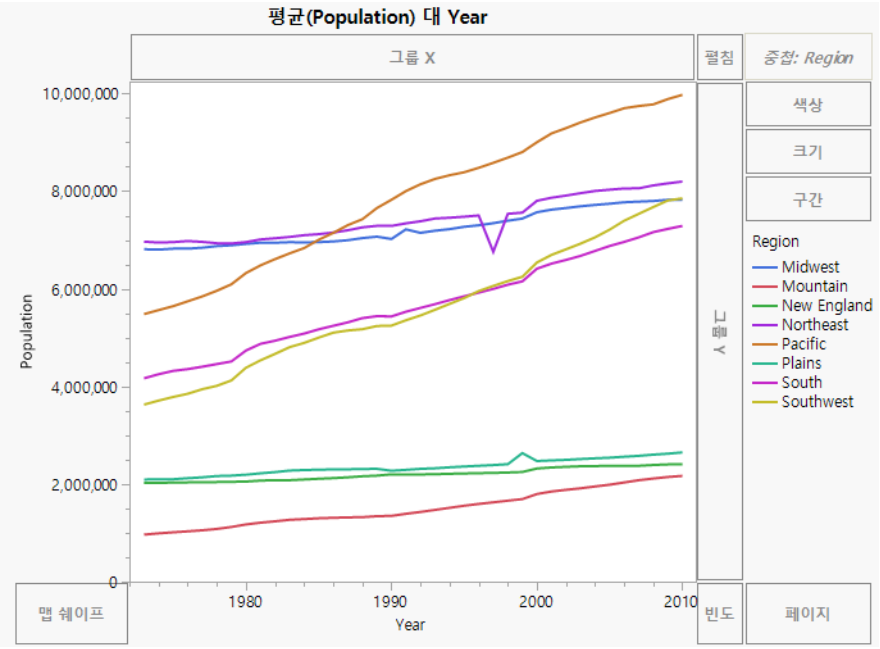
예 : 가리키기 라벨 그래프에서 필터 열 제거

일부 경우에는 제거하려는 열을 기준으로 가리키기 라벨 그래프를 필터링할 수 있습니다. 예를 들어 필터 열이 너무 세부적이어서 유용하지 않은 경우가 있을 수 있습니다. 이 경우 가리키기 라벨 편집기를 사용하여 해당 열을 제거할 수 있습니다.

원래 그래프 생성

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 CrimeData.jmp 를 엽니다 .
   이 데이터는 미국의 지역 및 주별 범죄율을 보여 줍니다 .
2. **그래프** > **그래프 빌더**를 선택합니다 .
3. Population 을 선택하여 **Y** 영역으로 드래그합니다 .
4. Year 를 선택하여 **X** 영역으로 드래그합니다 .
5. Region 을 선택하여 **중첩** 영역으로 드래그합니다 .
6. 선 요소 를 클릭합니다 .

그림 9.48 미국의 연도별 / 지역별 평균 인구 및 범죄율



이 그래프는 미국 각 지역의 평균 인구를 보여 주며 , 이는 연도별로 범죄율에 따라 영향을 받았습니다 . 가리키기 라벨 그래프에서는 지역 대신 미국 주에 따른 영향을 보다 세부적으로 확인하려고 합니다 .

7. 이 창을 열어 둡니다 .

그래프 복사 및 붙여넣기

1. " 그래프 빌더 " 의 빨간색 삼각형을 클릭하고 **다시 실행 > 분석 다시 실행**을 선택합니다 .

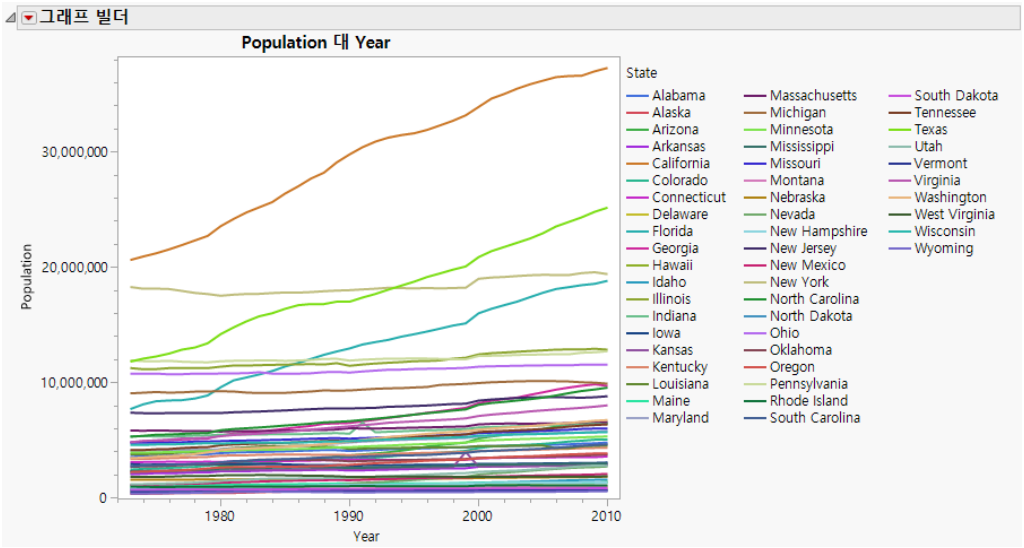
   원래 그래프의 복사본이 나타납니다 .

2. 그래프 복사본에서 State 를 선택하여 **중첩** 영역으로 드래그합니다 .

   그러면 Region 이 State 로 바뀝니다 .

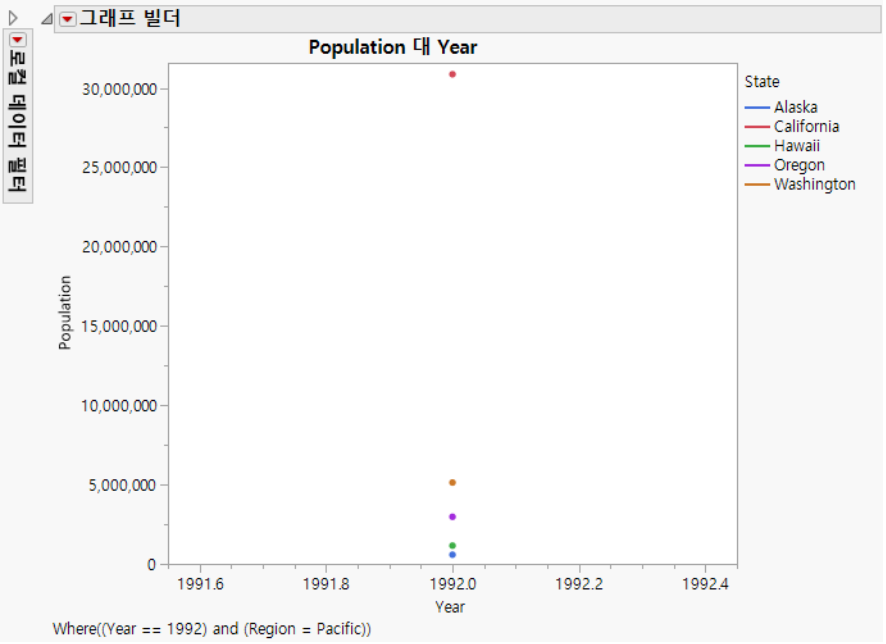
3. **완료**를 클릭합니다 .

그림 9.49 미국의 연도별 / 주별 평균 인구 및 범죄율



4. " 그래프 빌더 " 의 빨간색 삼각형을 클릭하고 **스크립트 저장 > 클립보드에**를 선택합니다 .
5. 이 그래프를 닫습니다 .
6. 원래 그래프에서 마우스 오른쪽 버튼을 클릭하고 **가리키기 라벨** > **Graphlet 붙여넣기**를 선택합니다 .
7. "Pacific" 선을 마우스로 가리키고 클릭하여 가리키기 라벨 그래프를 시작합니다 .

그림 9.50 Pacific 지역의 주별 / 연도별 범죄율을 보여 주는 가리키기 라벨 그래프



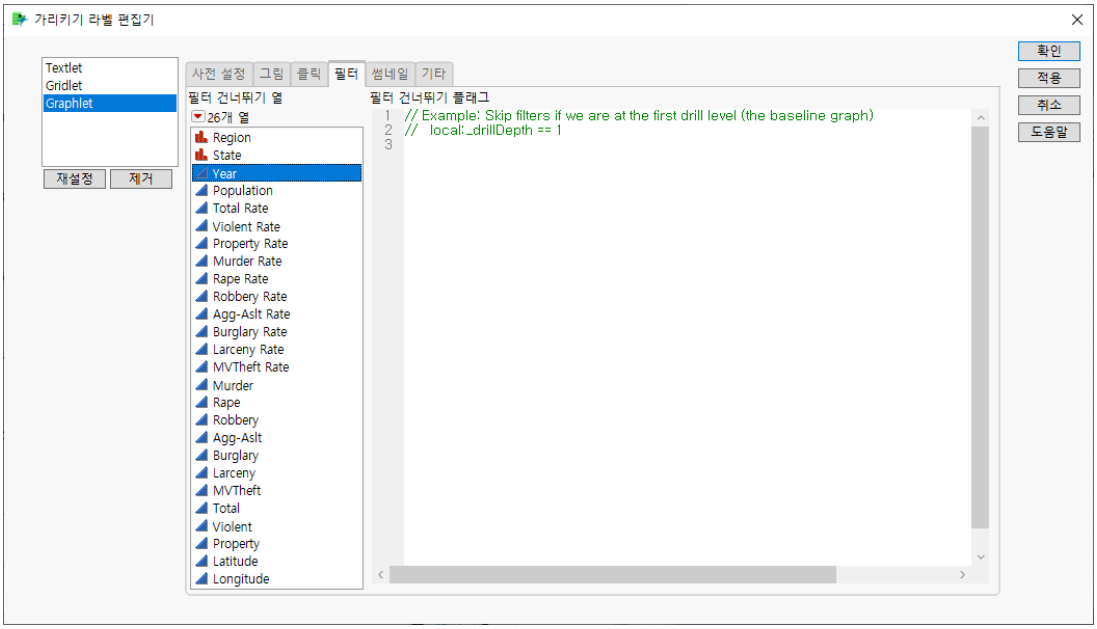
아직 연도를 기준으로 필터링하고 있기 때문에 이 보기는 너무 세부적이어서 유용하지 않습니다 . 연도 필터를 제거하여 지역만을 기준으로 그래프를 필터링하려고 합니다 .

8. "Pacific" 지역에 대한 가리키기 라벨 그래프를 닫습니다 .

필터 열 제거

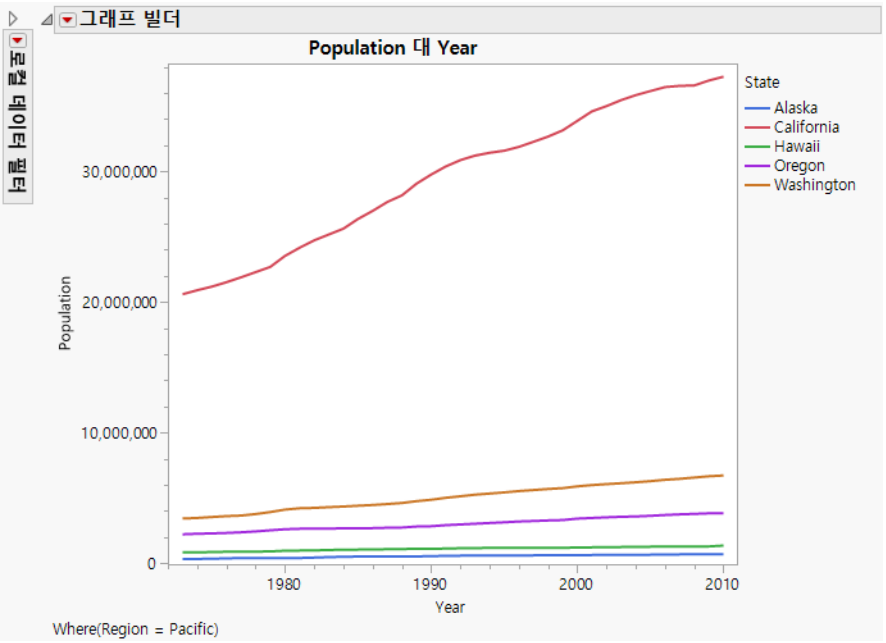
1. 원래 그래프를 마우스 오른쪽 버튼으로 클릭하고 **가리키기 라벨** > **가리키기 라벨 편집기**를 선택합니다 .
2. "Graphlet" 창에서 **필터** 탭을 클릭합니다 .
3. 더 이상 필터링 기준으로 사용하지 않을 열인 **Year** 를 선택합니다 .

그림 9.51 가리키기 라벨 편집기에서 필터 건너뛰기



4. **확인**을 클릭합니다 .
5. "Pacific" 선을 마우스로 가리키고 클릭하여 가리키기 라벨 그래프를 시작합니다 .

그림 9.52 Pacific 지역의 주별 범죄율을 보여 주는 가리키기 라벨 그래프



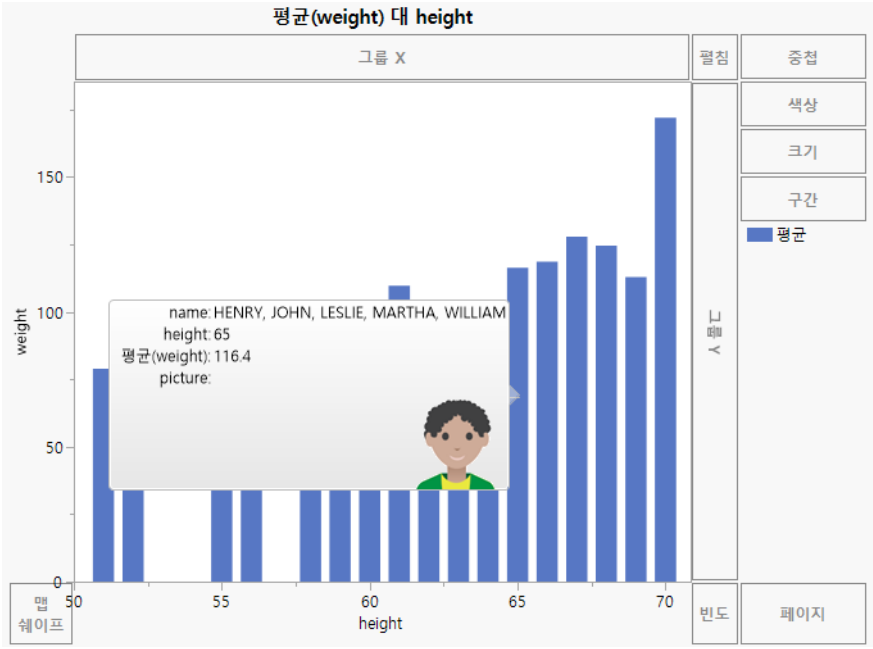
이 가리키기 라벨 그래프는 더 이상 연도를 기준으로 필터링하지 않고 지역만을 기준으로 필터링하여 보다 유용한 시각화 자료를 생성합니다.

예 : 가리키기 라벨에 여러 이미지 표시

라벨 열의 경우 가리키기 라벨은 첫 번째 라벨 이미지 열의 이미지와 라벨 텍스트 열의 값 최대 4 개를 포함할 수 있습니다. 더 많은 라벨 이미지나 텍스트 관측값을 표시하려면 라벨 보기를 사용하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class Families.jmp 를 엽니다 .
2. **그래프 > 그래프 빌더**를 선택합니다 .
3. weight 를 선택하여 **Y** 영역으로 드래그합니다 .
4. height 를 선택하여 **X** 영역으로 드래그합니다 .
5. 막대 요소 를 클릭합니다 .
6. 높이가 65 인치인 막대를 마우스로 가리키고 고정 을 클릭합니다 .

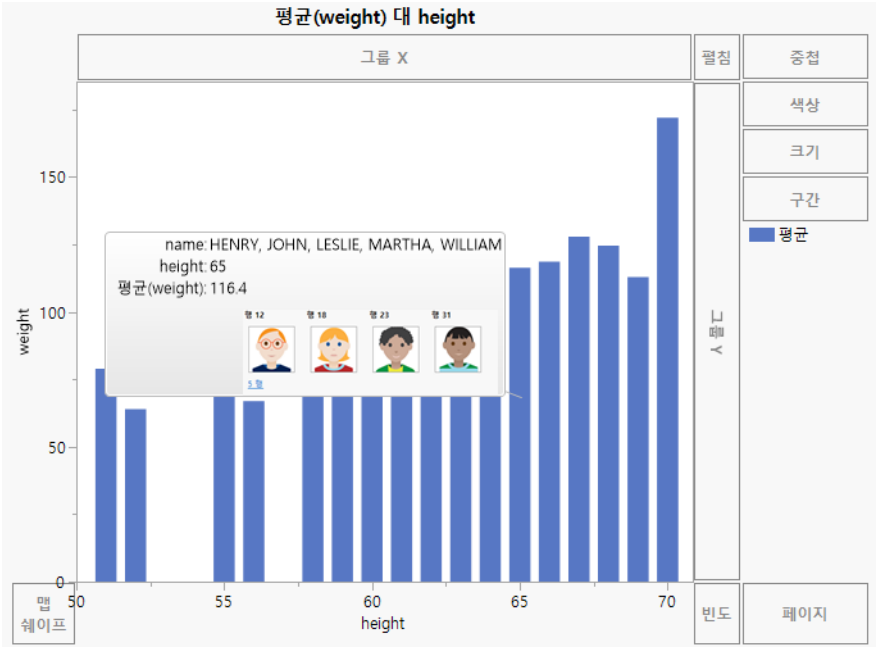
그림 9.53 초기 이미지만 있는 가리키기 라벨



이 막대에는 5 개의 이름이 포함되어 있지만 첫 번째 이미지 (Henry) 만 가리키기 라벨에 나타납니다 . 가리키기 라벨에 더 많은 이미지를 표시하고 싶습니다 .

7. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **가리키기 라벨** > **라벨 보기**를 선택합니다 .

그림 9.54 다중 이미지가 있는 가리키기 라벨



이제 최대값인 4 개의 이미지가 가리키기 라벨에 나타납니다 .

8. 추가 이미지를 보려면 가리키기 라벨에서 썸네일을 클릭합니다 .
9. 라벨 보기에서 앞으로 화살표 를 클릭하여 5 번째 이미지를 표시합니다 .

그래프의 가리키기 라벨에 서식 있는 텍스트 추가

Textlet 은 가리키기 라벨에 형식이 지정된 텍스트를 추가하는 가리키기 라벨 확장입니다. Textlet 을 사용하면 설명 단락에서 데이터를 설명할 수도 있고, 정보를 굵게 표시하거나 다른 색상으로 표시하는 등의 방법으로 중요한 정보를 눈에 띄게 할 수도 있습니다.

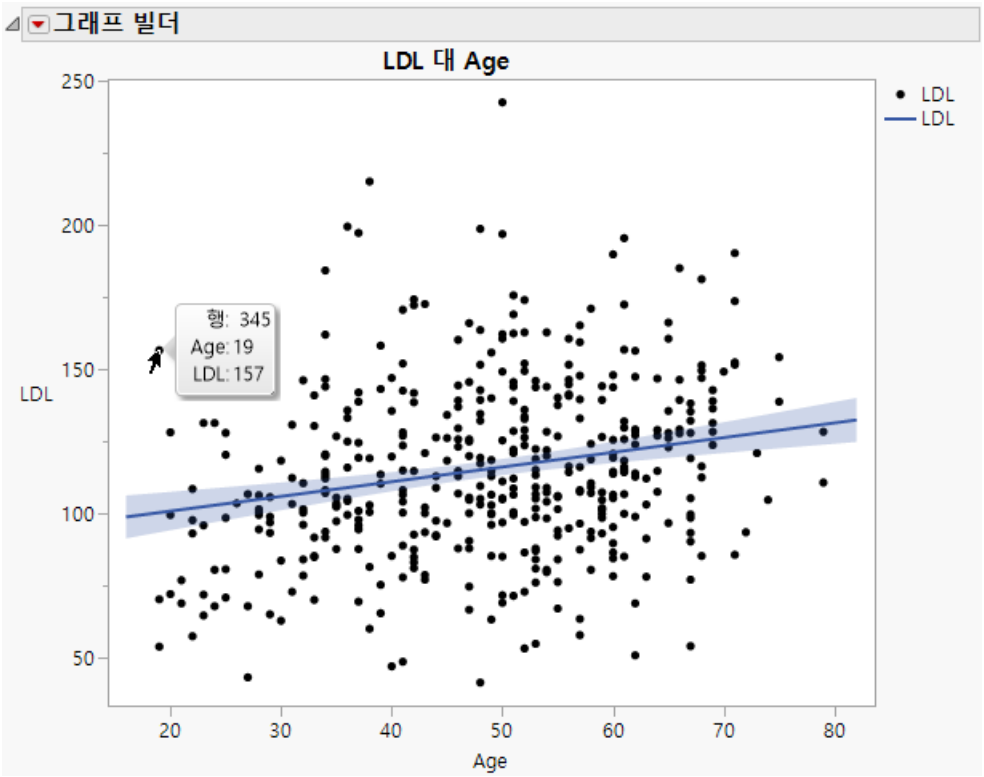
예 : 가리키기 라벨에 텍스트 추가

원래 그래프 생성

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
   이 데이터는 당뇨병 환자의 질병 경과를 추적한 측정값을 보여 줍니다 .
2. **그래프** > **그래프 빌더**를 선택합니다 .
3. LDL 을 선택하여 **Y** 영역으로 드래그합니다 .
4. Age 를 선택하여 **X** 영역으로 드래그합니다 .
5. 적합선 요소 를 클릭합니다 .

6. **완료**를 클릭합니다 .
7. 345 행의 점을 마우스로 가리킵니다 .

그림 9.55 345 행의 내용에 대한 초기 가리키기 라벨



345 행에 있는 데이터 점의 행 번호, 환자 연령 및 LDL 측정값을 알 수 있습니다. 가리키기 라벨에 나타나는 내용을 변경합니다.

- 환자의 질병 경과 수준을 한 문장으로 추가합니다 .
- 심각도에 따라 질병 경과 수준에 서로 다른 색상을 할당합니다 .

텍스트 추가

1. 345 행에 대한 가리키기 라벨 위에 커서를 놓고 고정 📌을 클릭합니다 .
   가리키기 라벨을 고정하면 변경 사항이 적용되는 것을 볼 수 있습니다 .
2. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **가리키기 라벨** > **가리키기 라벨 편집기**를 클릭합니다 .
3. 왼쪽에서 **Textlet** 을 클릭합니다 .

4. "JSL 변수 " 상자에서 JSL 을 사용하여 변수를 정의합니다 . 다음 JSL 스크립트를 복사하여 붙여넣습니다 .

```
// gets the value for the Y Ordinal column for the underlying visual element
local:YOrdinal = As Column( "Y Ordinal" )[local:_firstRow];
// maps the value to HTML color names that are used to style the text
local:color = Match( local:YOrdinal,
"Low", "Medium Dark Green",
"Medium", "Medium Dark Blue",
"High", "Medium Dark Red");
```

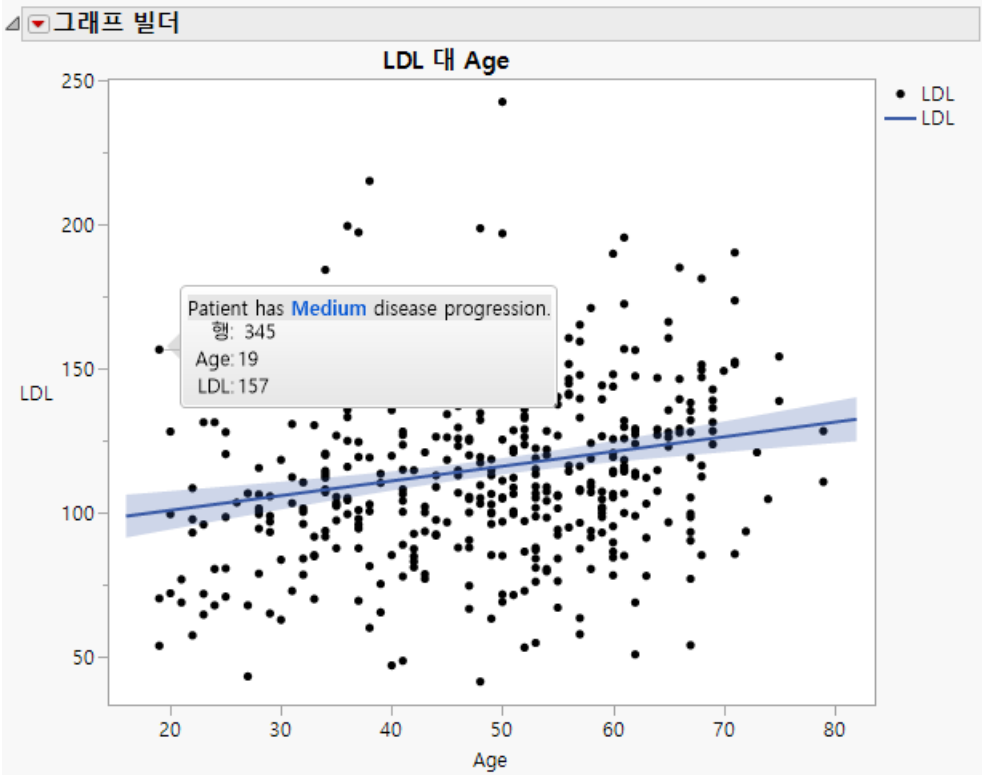
5. "HTML 마크업 " 에서 HTML 과 방금 정의한 JSL 변수를 사용하여 텍스트를 작성합니다 . 다음 코드를 복사하여 붙여넣습니다 .

```
Patient has <font color='{local:color}'><b>{local:YOrdinal}</b></font>
   disease progression.
```

그림 9.56 Textlet 의 JSL 변수 및 HTML 마크업

6. **확인**을 클릭합니다 .

그림 9.57 서식 있는 텍스트를 포함하도록 업데이트된 가리키기 라벨



가리키기 라벨에 질병 경과 수준을 나타내는 새 문장이 표시되는 것을 확인할 수 있습니다. 이 환자의 질병 경과는 중간 수준이므로 파란색으로 표시되었습니다.

가리키기 라벨의 이름 및 값 업데이트

그래프의 요소(예: 점, 막대 등)를 마우스로 가리키면 가리키기 라벨에 해당 요소에 대한 내용이 표시됩니다. 기본적으로 가리키기 라벨에는 격자에 있는 이름 및 값 쌍이 포함됩니다. 이 값은 데이터 테이블에서 가져온 것일 수도 있고, 임의의 계산 또는 JSL 표현식일 수도 있습니다. Gridlet을 사용하여 이러한 이름 또는 값을 사용자 정의할 수 있습니다. 예를 들어 다음과 같은 작업을 수행할 수 있습니다.

- 이름과 관련 값을 제거합니다.
- 이름과 관련 값을 추가합니다.
- 이름을 변경합니다.
- 이름 스타일(예: 굵은 글꼴) 또는 값 스타일(예: 소수 자릿수)을 변경합니다.

팁 : 가리키기 라벨 격자 항목을 사용자 정의하는 또 다른 방법은 라벨을 사용하는 것입니다 . 자세한 내용은 "데이터 테이블에서 가리키기 라벨에 이미지 추가"(516 페이지)에서 확인하십시오.

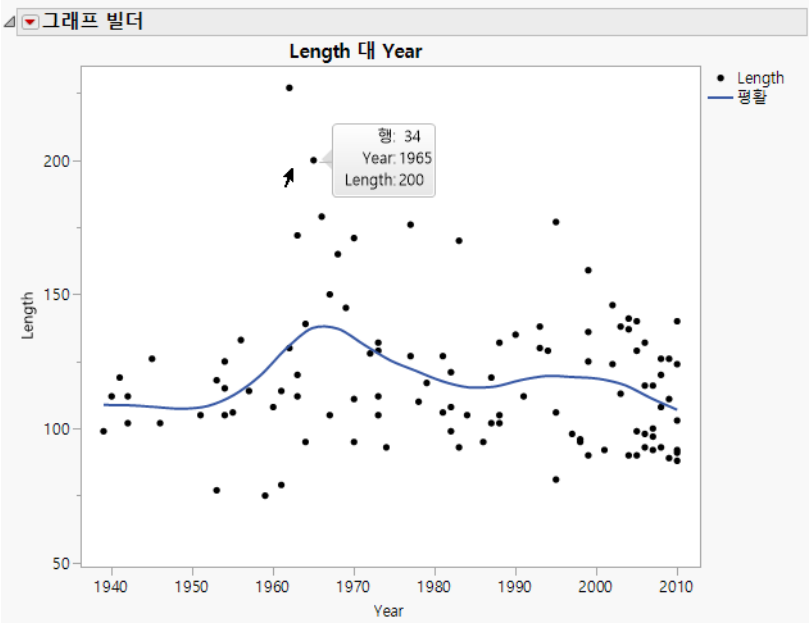
예 : 가리키기 라벨의 이름 및 값 업데이트

이 예에서는 그래프를 생성하고 Gridlet 을 사용하여 가리키기 라벨의 내용을 업데이트합니다.

원래 그래프 생성

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 Movie Inventory.jmp 를 엽니다 .
   이 데이터는 영화 대여 데이터를 보여 줍니다 .
2. **그래프** > **그래프 빌더**를 선택합니다 .
3. Length 를 선택하여 **Y** 영역으로 드래그합니다 .
4. Year 를 선택하여 **X** 영역으로 드래그합니다 .
5. **완료**를 클릭합니다 .
6. 34 행의 점을 마우스로 가리킵니다 (그림 9.58).

그림 9.58 34 행에 대한 초기 가리키기 라벨 내용



34 행에 있는 데이터 점의 행 번호, 영화 제작 연도 및 영화 길이 (분) 가 표시됩니다. 가리키기 라벨에 나타나는 내용을 변경합니다.

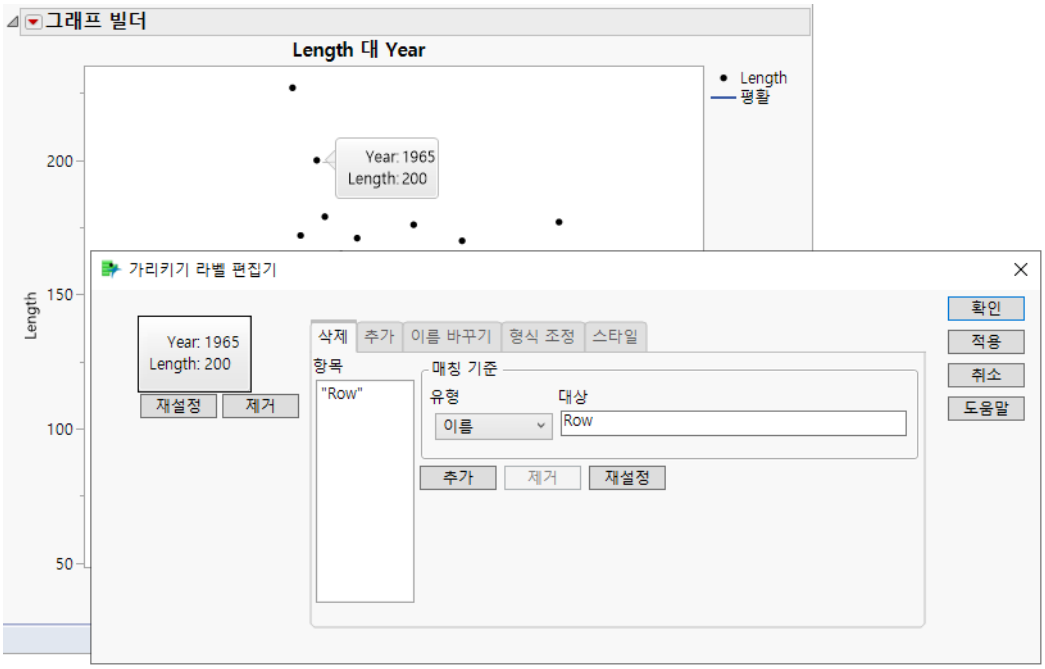
- 행 번호 및 길이 항목을 제거합니다 .

- 영화 이름 및 장르를 추가합니다 .
- 영화 길이를 시간 및 분 단위로 표시하는 새 항목을 생성합니다 .
- 장르 이름을 굵은 글꼴로 표시합니다 .

이름 및 값 업데이트

1. 34 행에 대한 가리키기 라벨을 마우스로 가리키고 고정 을 클릭합니다 .
   가리키기 라벨을 고정하면 변경 사항이 적용되는 것을 볼 수 있습니다 .
2. 가리키기 라벨을 마우스 오른쪽 버튼으로 클릭하고 **편집**을 클릭합니다 .
3. 썸네일의 왼쪽에서 **행**을 클릭합니다 .
4. " 삭제 " 탭에서 **추가**를 클릭합니다 .
   삭제할 항목 아래에 " 행 " 이 나타납니다 .
5. **적용**을 클릭합니다 .
   그래프 창에서 가리키기 라벨의 **행**이 제거된 것을 확인할 수 있습니다 .

그림 9.59 " 행 " 이 제거된 가리키기 라벨



6. 썸네일 왼쪽에서 **길이** (제거하려는 다른 항목) 를 클릭합니다 .
7. " 삭제 " 탭에서 **추가**를 클릭합니다 .
   삭제할 항목 아래에 " 길이 " 가 나타납니다 .

8. **적용**을 클릭합니다 .

   그래프 창에서 가리키기 라벨의 길이가 제거되어 연도만 남은 것을 확인할 수 있습니다 .

9. 영화 이름 및 장르를 추가합니다 .

   a. **추가** 탭을 클릭합니다 . " 유형 " 이 **이름**으로 설정되어 있어야 합니다 .

   b. " 유형 " 에 Name(추가하려는 이름) 을 입력합니다 .

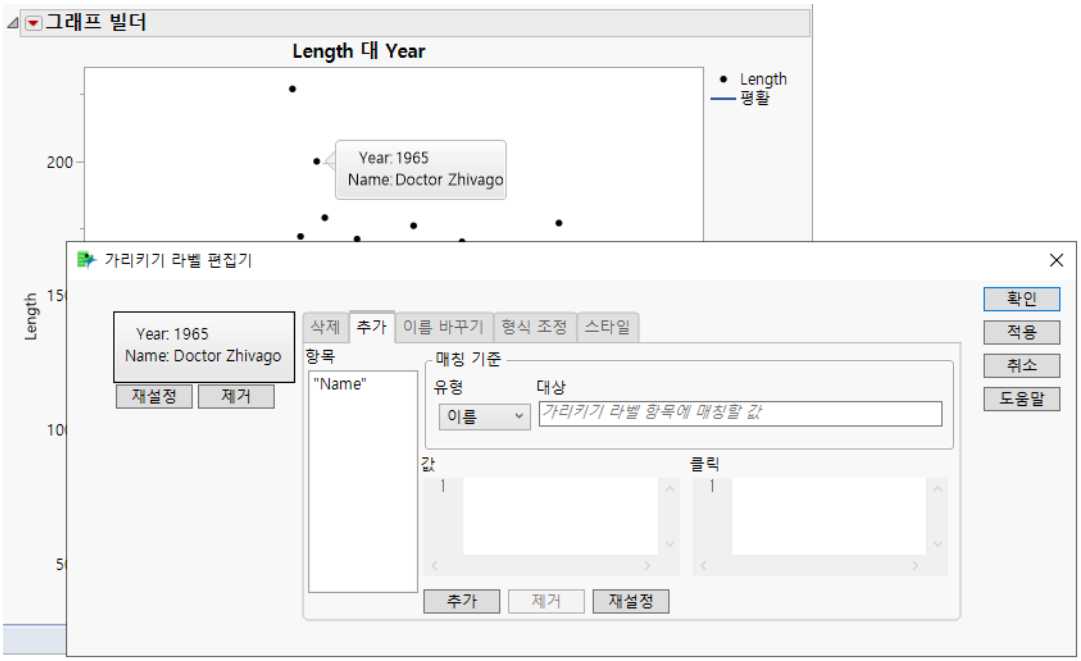
   c. " 값 " 에 JSL 표현식 "`:Name[local:_firstRow]`" 를 복사하여 붙여넣습니다 .

      "Name" 은 열 참조입니다 . 대괄호는 지정된 인덱스에 대해 해당하는 열 값을 가져오도록 지정합니다 . `local:_firstRow` 는 HLEC(가리키기 라벨 실행 컨텍스트) 변수인 색인입니다 . 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

   d. **추가**를 클릭하고 **적용**을 클릭합니다 .

      썸네일과 그래프 가리키기 라벨에 영화 이름이 추가됩니다 .

그림 9.60 "Name" 이 추가된 가리키기 라벨



   e. " 대상 " 에 Genre(추가하려는 두 번째 이름) 를 입력합니다 .

   f. " 값 " 에 JSL 표현식 "`:Genre[local:_firstRow]`" 를 복사하여 붙여넣습니다 .

   g. **추가**를 클릭하고 **적용**을 클릭합니다 .

      썸네일과 그래프 가리키기 라벨에 영화 장르가 추가됩니다 .

10. "Length" 데이터를 분 단위에서 시간 및 분 단위로 변환합니다 .

   a. **추가** 탭의 " 대상 " 상자에 Duration(추가하려는 새 가리키기 라벨의 이름) 을 입력합니다 .

b. " 값 " 에 JSL 표현식 "`local:_Length * 60`" 을 복사하여 붙여넣습니다 .

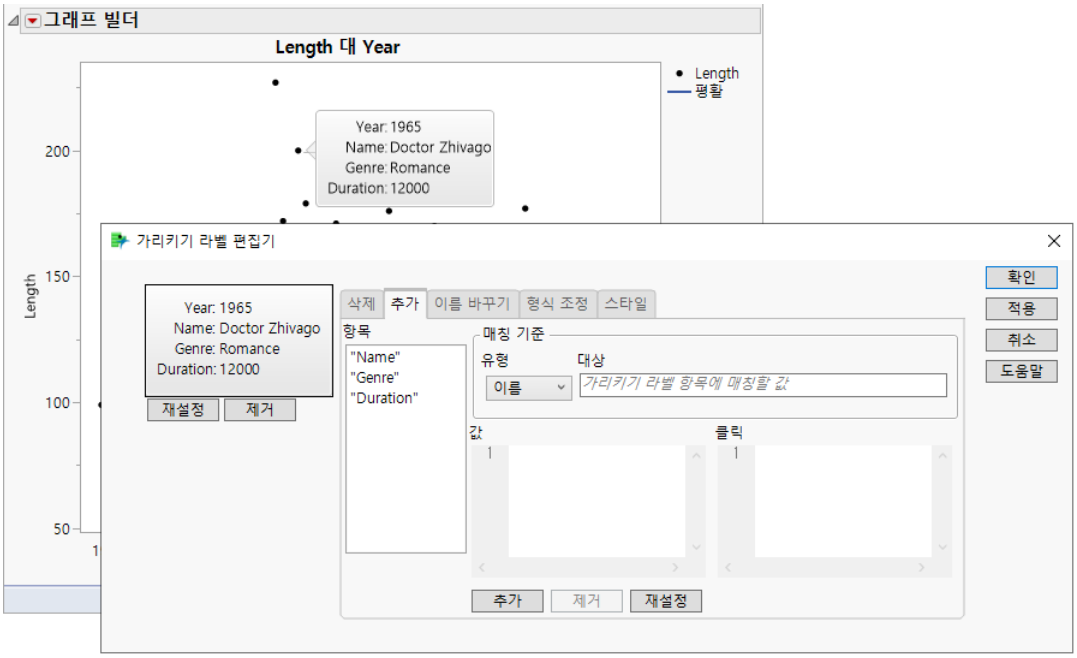
"local:_Length" 는 격자 항목 "Lengh" 의 값이 있는 HLEC 변수입니다 . "Length" 는 그래프에 있으므로 가리키기 라벨 컨텍스트에서 이 값에 액세스할 수 있습니다 . 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

"* 60" 은 "Duration" 을 분 단위에서 초 단위로 변환합니다 . 시간 : 분 형식을 적용하려면 "Duration" 이 초 단위여야 합니다 .

c. **추가**를 클릭하고 **적용**을 클릭합니다 .

썸네일과 그래프 가리키기 라벨에 "Duration" 라벨이 추가됩니다. 하지만 이 값의 형식을 올바르게 지정해야 합니다 .

그림 9.61 "Duration" 이 추가된 가리키기 라벨



d. **형식 조정** 탭을 클릭합니다 .

e. " 대상 " 에 Duration(형식을 조정하려는 가리키기 라벨 항목의 이름) 을 입력합니다 .

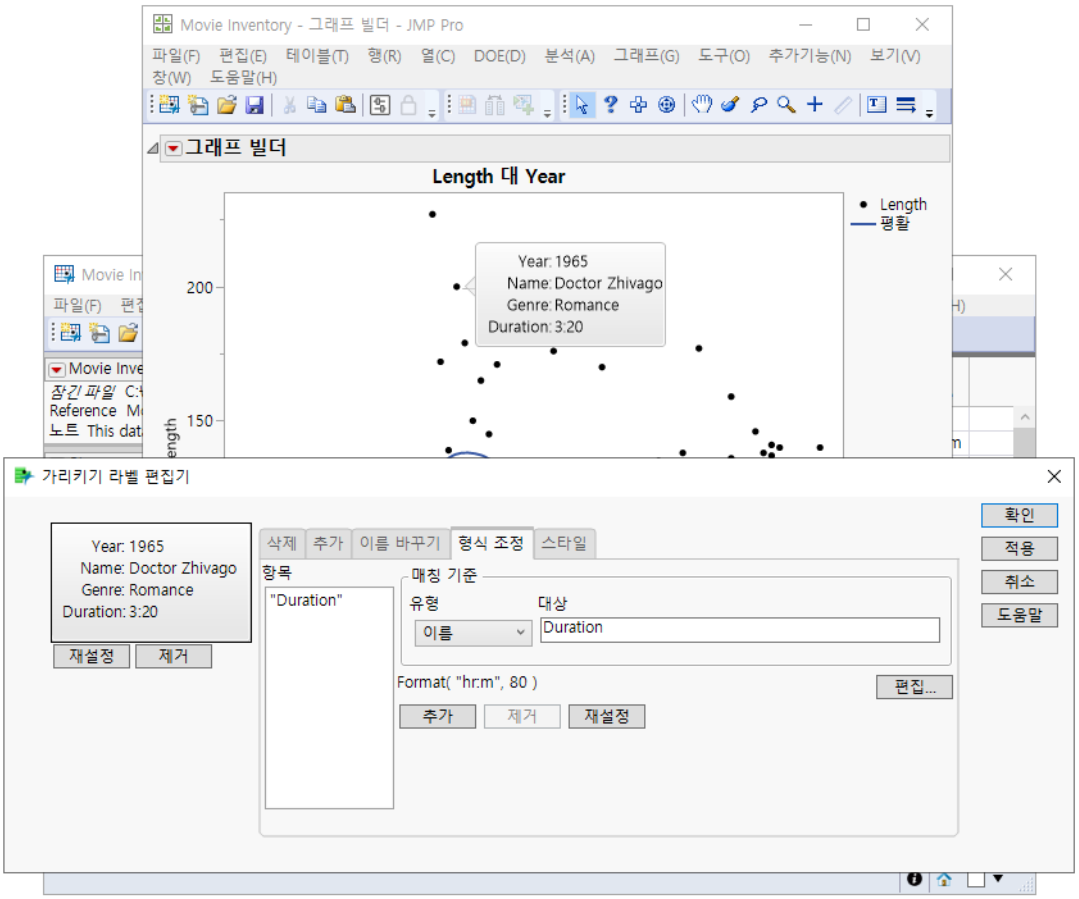
f. **편집**을 클릭합니다 .

g. " 형식 " 옆에서 **기간** > **hr:m** 을 선택하고 **확인**을 클릭합니다 .

h. **추가**를 클릭하고 **적용**을 클릭합니다 .

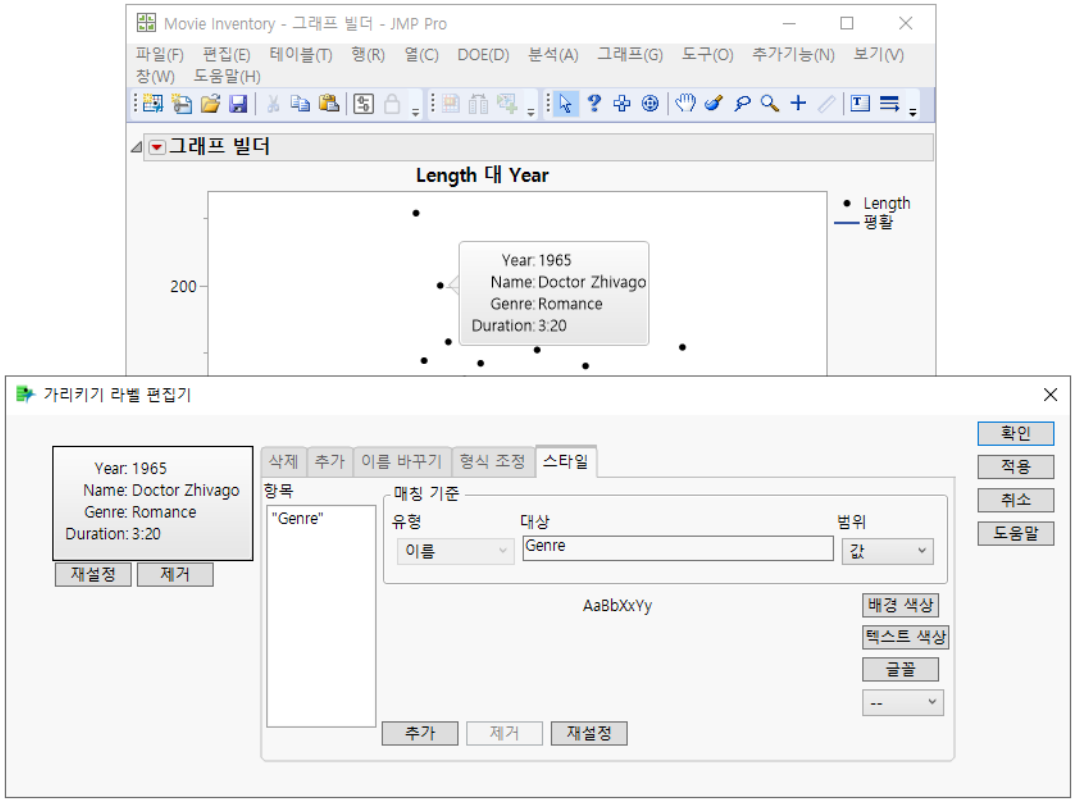
가리키기 라벨에서 "Duration" 이 시간 및 분 단위로 표시됩니다 .

그림 9.62 시간 및 분 단위의 "Duration" 이 포함된 가리키기 라벨



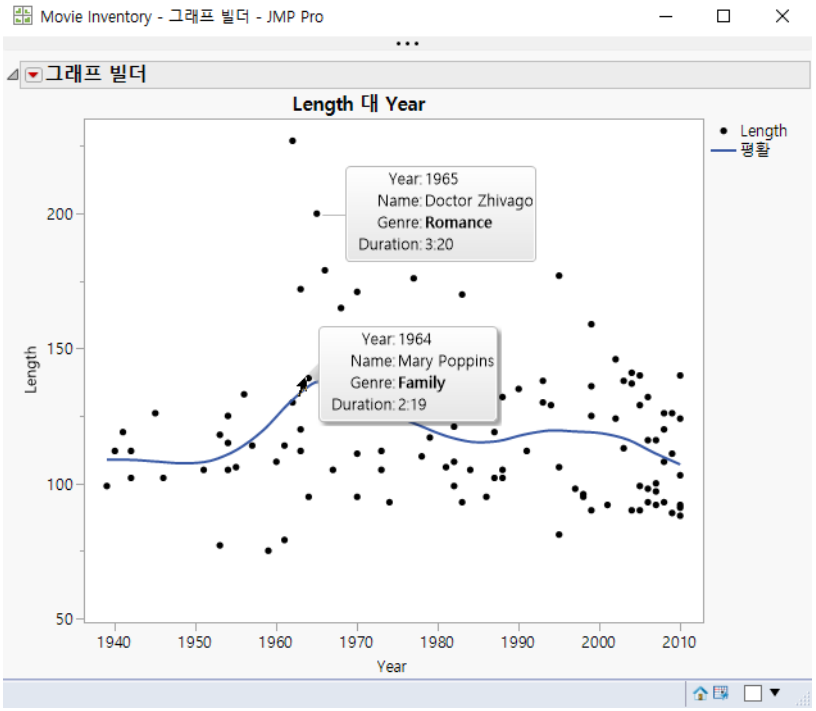
11. "Genre" 값을 굵게 표시합니다 .
    a. **스타일** 탭을 클릭합니다 .
    b. " 대상 " 에 Genre(굵게 표시하려는 가리키기 라벨 항목의 이름) 를 입력합니다 .
       " 범위 " 는 **값**으로 설정되어 있습니다 . 그러면 "Genre" 라는 이름 대신 "Genre" 값이 굵게 표시됩니다 .
    c. **글꼴**을 클릭합니다 .
    d. " 글꼴 스타일 " 에서 **굵게**를 클릭하고 **확인**을 클릭합니다 .
    e. **추가**를 클릭하고 **적용**을 클릭합니다 .
       가리키기 라벨에서 "Genre" 값 "Romance" 가 굵게 표시됩니다 .

그림 9.63 "Genre" 값이 굵게 표시된 가리키기 라벨



다른 라벨을 마우스로 가리키고 모두 업데이트되었는지 확인합니다.

그림 9.64 업데이트된 가리키기 라벨



데이터 테이블에서 라벨 사용

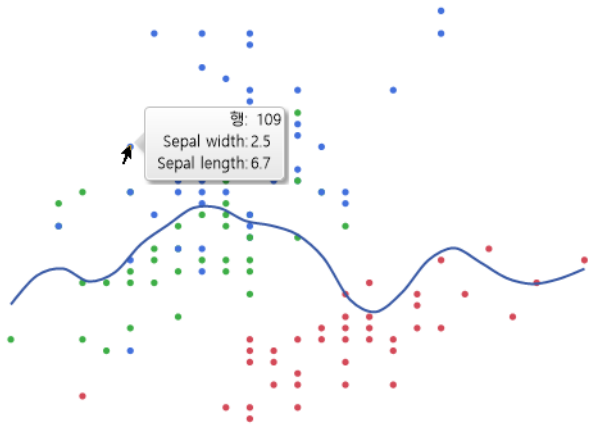
Gridlet 외에 데이터 테이블에서 열에 라벨을 추가하는 방법으로도 가리키기 라벨에 대한 몇 가지 기본 사용자 정의를 수행할 수 있습니다.

- 가리키기 라벨에 " 행 " 항목이 포함되어 있는 경우 라벨을 추가하는 방법으로 이를 열의 값으로 바꿀 수 있습니다 .
- 라벨을 추가한 후 이 라벨을 그래프에 유지할 수 있습니다 . 그래프에서 라벨의 위치를 이동할 수도 있습니다 . 또한 태그 선을 추가하여 라벨을 해당 점에 연결할 수 있습니다 .

예 : 데이터 테이블의 라벨 사용

1. **도움말** > **샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다 .
2. **그래프** > **그래프 빌더**를 선택합니다 .
3. Sepal length 를 선택하여 **Y** 영역으로 드래그합니다 .
4. Sepal width 를 선택하여 **X** 영역으로 드래그합니다 .
5. **완료**를 클릭합니다 .
6. 109 행의 점을 마우스로 가리킵니다 .

그림 9.65 109 행에 대한 가리키기 라벨



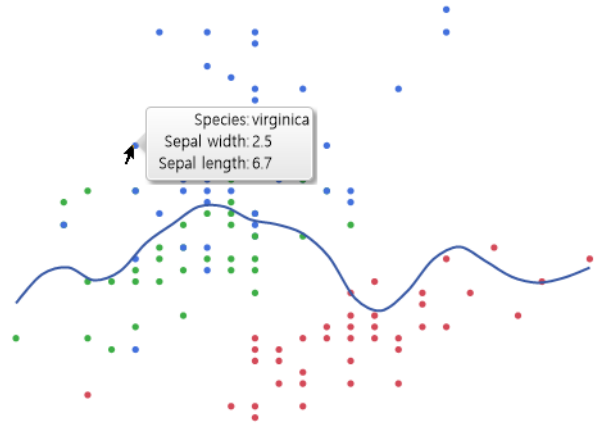
가리키기 라벨에 " 행 " 항목이 표시되는 것을 확인할 수 있습니다 . 이 항목을 데이터 테이블의 "Species" 라는 열과 해당 값으로 바꾸려고 합니다 .

7. 데이터 테이블에서 "Species" 열을 마우스 오른쪽 버튼으로 클릭하고 **라벨 / 라벨 해제**를 선택합니다 .

   " 열 " 패널의 해당 열 이름 옆에 라벨 아이콘 이 표시됩니다 .

8. 그래프 빌더에서 동일한 데이터 점을 마우스로 가리킵니다 .

그림 9.66 "Species" 열을 보여 주는 가리키기 라벨



" 행 " 항목이 "Species" 열로 바뀌어 있음을 알 수 있습니다 .

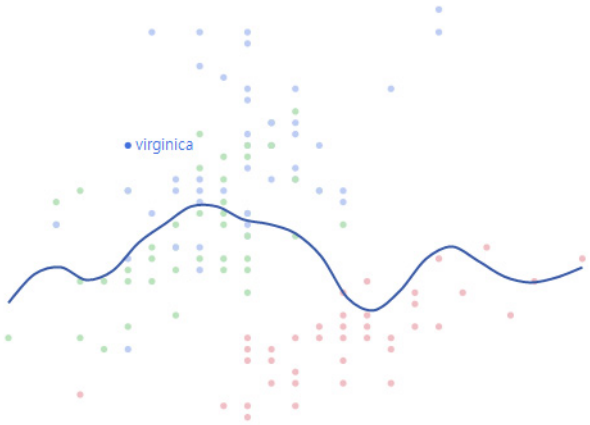
참고 : 라벨 열을 여러 개 지정하면 가리키기 라벨에 해당 값들이 사전순으로 표시됩니다 .

이제 일부 라벨을 그래프에 유지하려고 합니다 .

9. 동일한 데이터 점을 클릭하여 강조 표시합니다 .
10. 마우스 오른쪽 버튼으로 아무 곳이나 클릭하고 **행** > **행 라벨**을 선택합니다 .

참고 : 히스토그램에서는 오른쪽의 상자 그림 영역을 마우스 오른쪽 버튼으로 클릭합니다 .

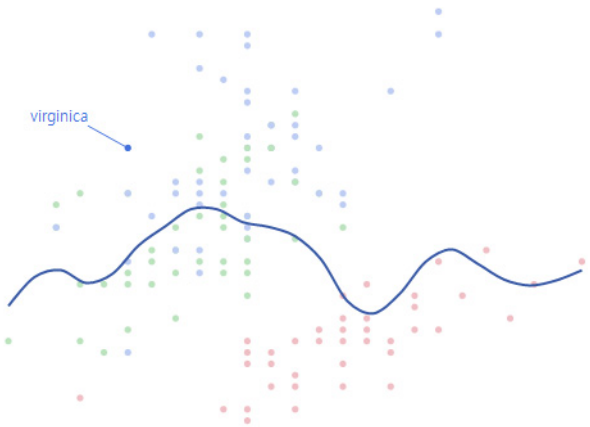
그림 9.67 그래프에 표시된 "Species" 라벨



이 데이터 점에 대한 "Species" 값이 포함된 라벨이 그래프에 표시됩니다 .

11. 라벨을 클릭한 채 드래그하여 위치를 이동합니다 .
    라벨과 점을 연결하는 태그 선이 추가됩니다 .

그림 9.68 "Species" 라벨 위치 조정



데이터 테이블에서 해당 행 번호 옆에 라벨 아이콘 이 표시됩니다 .

해당 열 또는 행을 강조 표시하고 **라벨/라벨 해제**를 다시 선택할 때까지 해당 데이터는 라벨 표시 상태로 유지됩니다.

그래프의 가리키기 라벨 편집

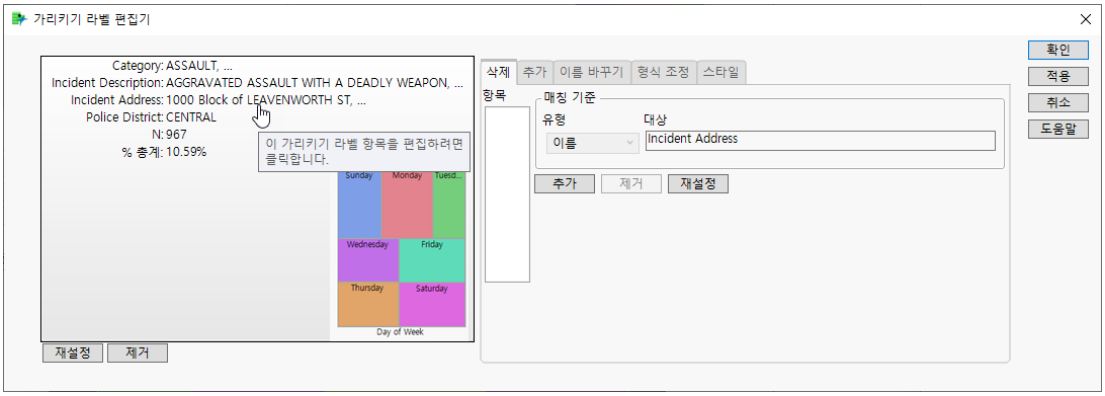
가리키기 라벨 편집기를 사용하여 가리키기 라벨에 표시되는 그래프, 이미지, 형식이 지정된 텍스트, 이름 및 값을 사용자 정의하고 추가할 수 있습니다.

가리키기 라벨을 대화식으로 편집하려면

1. 그래프에서 가리키기 라벨을 표시할 데이터를 마우스로 가리킵니다.
2. 가리키기 라벨에서 마우스 오른쪽 버튼을 클릭하고 **편집**을 선택합니다.

가리키기 라벨 편집기에서 가리키기 라벨의 썸네일이 왼쪽에 나타납니다. 썸네일 안쪽 영역을 클릭하여 오른쪽에서 편집합니다. 변경한 사항은 그래프의 모든 가리키기 라벨에 적용됩니다.

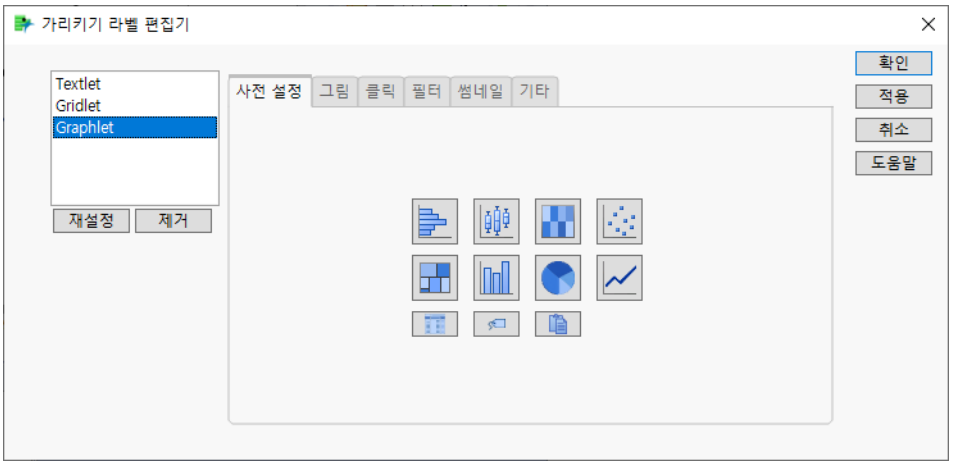
그림 9.69 가리키기 라벨 대화식 편집



이 예에서는 "Incident Address" 데이터를 클릭하면 "삭제" 탭이 표시되므로 그래프의 모든 가리키기 라벨에서 "Incident Address" 항목을 삭제할 수 있습니다. 이름을 바꾸거나 형식을 조정하거나 글꼴 스타일을 변경할 수도 있습니다.

일반적으로 가리키기 라벨을 추가하거나 편집하려면 "가리키기 라벨 편집기"를 시작합니다. 그래프를 마우스 오른쪽 버튼으로 클릭하고 **가리키기 라벨** > **가리키기 라벨 편집기**를 선택합니다.

그림 9.70 가리키기 라벨 편집기



왼쪽에서는 가리키기 라벨에서 수행하려는 작업에 따라 **Textlet**, **Gridlet** 또는 **Graphlet**을 선택할 수 있습니다.

Textlet JSL 을 사용하여 가리키기 라벨에 형식이 지정된 텍스트를 추가합니다 . 예는 " 예 : 가리키기 라벨에 텍스트 추가 "(498 페이지) 에서 확인하십시오 . 이 창에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

Gridlet 가리키기 라벨에 표시되는 이름 및 값을 삭제하거나 , 추가하거나 , 이름을 바꾸거나 , 형식을 조정하거나, 스타일을 선택합니다. 예는 " 예: 가리키기 라벨의 이름 및 값 업데이트"(502 페이지) 에서 확인하십시오 . 이 창에 대한 자세한 내용은 "Gridlet 탭 "(514 페이지) 에서 확인하십시오 .

Graphlet 가리키기 라벨에 대화식 그래프 또는 이미지를 추가합니다 . 예는 " 예 : 가리키기 라벨에 사전 설정 그래프 추가 "(484 페이지) 에서 확인하십시오 . 이 창에 대한 자세한 내용은 "Graphlet 탭 "(516 페이지) 에서 확인하십시오 .

Textlet 탭

마크업 가리키기 라벨에 서식 있는 텍스트를 생성하는 데 사용되는 HTML 마크업을 정의합니다.

사전 설정 사전 구성된 Textlet 의 예가 표시됩니다 .

- "Showcase Textlet" 은 기본 데이터에 따라 Textlet 내용에 색상 및 스타일을 추가하는 방법을 보여 주는 예입니다 . 예는 " 그래프의 가리키기 라벨에 서식 있는 텍스트 추가 "(498 페이지) 에서 확인하십시오 .
- "Tabulate Textlet" 은 Textlet 내용에 기본 데이터에 대한 요약 통계량을 추가하는 방법을 보여 주는 예입니다 .
- "Context Textlet" 은 디버깅 목적으로 Textlet 에 실행 컨텍스트 변수를 표시하는 방법을 보여 주는 예입니다 . 실행 컨텍스트 변수는 Graphlet 및 Textlet 스크립트가 실행되는 환경을 정의합니다 . 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

JSL 변수 (선택 사항) 로컬 JSL 변수에 표현식 결과를 할당할 수 있습니다 . 이 변수는 "HTML 마크업" 필드에서 동적이고 데이터에 기반한 서식 있는 텍스트를 생성하는 데 사용됩니다 .

HTML 마크업 일부 HTML 태그에 따라 서식 있는 텍스트 단락 템플릿을 지정합니다 . 마크업에는 구분자로 묶인 JSL 변수 참조가 포함될 수 있습니다 . 기본 구분자는 중괄호 ({}) 입니다 . 표 9.6 에서는 지원되는 HTML 태그를 설명합니다 .

표 9.5 Textlet 에서 지원되는 HTML 태그

HTML 태그	설명
	굵은 텍스트
	기울임꼴 텍스트
<u> </u>	밑줄이 적용된 텍스트
 	다음 속성으로 스타일이 지정된 글꼴 • color=*color* : 따옴표로 묶은 전경 색상 값입니다 . 색상 이름이거나 16 진수 RGB 색상 값일 수 있습니다 . • face=*family* : 따옴표로 묶은 설치된 글꼴군 이름입니다 . • size=*size* : 따옴표로 묶은 글꼴 크기입니다 .
<background> </background>	배경 색상 값 • color=*color*: 따옴표로 묶은 색상 이름 또는 16 진수 RGB 색상 값입니다 .

기타 가리키기 라벨에 표시되는 텍스트의 구분자와 너비를 지정합니다 .

구분자 " 마크업 " 영역에서 대체 변수를 나타내는 구분자를 선택합니다 . 예를 들어 중괄호도 사용하는 C++ 코드나 기타 코드가 스크립트에 포함된 경우 JSL 변수에는 대괄호 구분자를 사용할 수 있습니다 .

너비 가리키기 라벨의 기본 너비를 픽셀 단위로 지정합니다 .

참고 : 가리키기 라벨을 고정한 경우 Textlet 구성을 변경한 후 " 적용 " 을 클릭하면 변경 사항을 바로 확인할 수 있습니다 . 업데이트된 가리키기 라벨에 만족하면 **확인**을 클릭하여 변경 사항을 저장하십시오 .

Gridlet 탭

"Gridlet"에는 "삭제", "추가", "이름 바꾸기", "형식 조정" 및 "스타일" 탭이 포함되어 있습니다. 이 모든 탭에서 수정할 가리키기 라벨 행목을 나타내는 "매칭 기준"을 지정합니다.

유형 및 대상

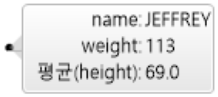
매칭 기준 수정하려는 가리키기 라벨 항목을 지정합니다 .

선택할 항목의 유형과 대상 항목을 선택합니다.

참고 : 모든 옵션이 모든 탭에서 표시되는 것은 아닙니다 .

이름 수정하려는 항목의 정확한 이름을 입력합니다. 예를 들어 해당 항목 이름이 "Mean(height)" 이면 " 대상 " 값으로 "Mean(height)" 을 입력합니다 .

그림 9.71 가리키기 라벨 예



인덱스 수정하려는 격자 내 항목의 인덱스를 입력합니다 . 예를 들어 가리키기 라벨의 세 번째 항목이 "Mean(height)" 인 경우 이 항목을 지정하려면 " 대상 " 값으로 "3" 을 입력합니다 .

열 수정하려는 항목의 열 이름을 입력합니다 . 예를 들어 해당 항목이 "Mean(height)" 이면 " 대상 " 값으로 "height" 를 입력합니다 .

정규 표현식 수정하려는 이름과 매칭되는 정규 표현식을 입력합니다 . 예를 들어 해당 항목이 "Mean(height)" 이면 " 대상 " 값으로 "^.*\(height\)$" 를 입력합니다 . "Pat Regex", 괄호 및 따옴표는 자동으로 추가됩니다 .

모두 모든 가리키기 라벨 항목의 스타일을 변경합니다 .

Gridlet 버튼

추가 지정한 " 대상 " 값을 " 항목 " 에 추가합니다 . 가리키기 라벨에 변경 사항을 적용하려면 **적용**을 클릭합니다 .

제거 지정한 " 대상 " 값을 " 항목 " 에서 제거합니다 . 가리키기 라벨에 변경 사항을 적용하려면 **적용**을 클릭합니다 .

바꾸기 " 항목 " 에서 선택한 " 대상 " 값을 바꿉니다 . 항목을 클릭하고 새 대상 값을 입력한 다음 **바꾸기**를 클릭합니다 . 가리키기 라벨에 변경 사항을 적용하려면 **적용**을 클릭합니다 .

Gridlet 탭

삭제 가리키기 라벨에서 삭제할 항목을 지정합니다 .

참고 : JSL 을 사용하는 경우 " 삭제 " 는 Expunge 명령입니다 .

추가 가리키기 라벨에 추가할 항목을 지정합니다 .

값 가리키기 라벨에 값을 추가하기 위한 JSL을 지정합니다. 예는 "이름 및 값 업데이트"(503 페이지) 에서 확인하십시오 .

클릭 항목에 링크로 표시할 작업을 추가합니다 . 링크를 클릭하면 실행되는 스크립트를 지정합니다 . 예를 들어 스크립트로 시각화 요소를 생성하거나 , 웹 페이지를 열거나 , 플랫폼을 시작할 수 있습니다 . 다음은 스크립트를 복사하여 붙여넣을 수 있는 몇 가지 방법입니다 .

연결할 시각화 요소 생성 : 그래프 빌더에서 원하는 시각화 요소를 생성합니다 . " 그래프 빌더 " 의 빨간색 삼각형 메뉴를 클릭하고 **스크립트 저장 > 클립보드에**를 선택합니다 . 스크립트를 " 클릭 " 아래에 붙여넣습니다 .

사전 설정 그래프에 연결 : 그래프 빌더에서 사전 설정 가리키기 라벨 그래프가 있는 기본 그래프를 마우스 오른쪽 버튼으로 클릭하고 **가리키기 라벨 > 가리키기 라벨 편집기**를 선택합니다 . "Graphlet" 의 " 그림 " 탭에서 스크립트를 복사합니다 . 스크립트를 " 클릭 " 아래에 붙여넣습니다 .

가리키기 라벨에 변경 사항을 적용하려면 **적용**을 클릭합니다 .

참고 :JSL 을 사용하는 경우 " 추가 " 는 Annex 명령입니다 .

이름 바꾸기 가리키기 라벨에서 이름을 바꾸려는 항목을 지정합니다 . 현재 " 대상 " 값을 바꿀 새 이름을 입력합니다 .

형식 조정 가리키기 라벨의 항목 값에 사용할 다른 형식을 지정합니다 . 예를 들어 날짜 또는 기간 형식을 변경하거나 소수 자릿수를 변경할 수 있습니다 . 다른 형식을 선택하려면 **편집**을 클릭합니다 . " 편집 " 옵션에 대한 자세한 내용은 " 숫자 형식 옵션 "(273 페이지) 에서 확인하십시오 .

스타일 가리키기 라벨에 있는 항목의 배경 색상 , 텍스트 색상 또는 글꼴 상세 정보 (굵게 등) 를 지정합니다 .

범위 스타일 업데이트를 항목 이름에 적용할지 , 값에 적용할지 , 둘 모두에 적용할지를 지정합니다 . 예를 들어 모든 매칭 기준에 스타일 변경 사항을 적용하려는 경우 손쉽게 왼쪽에 모든 이름을 할당하거나 오른쪽에 모든 값을 할당할 수 있습니다 .

배경 색상 가리키기 라벨에서 대상 값, 이름 또는 둘 모두에 사용할 배경 색상을 지정합니다.

텍스트 색상 가리키기 라벨에서 대상 값 , 이름 또는 둘 모두에 사용할 텍스트 색상을 지정합니다 .

글꼴 가리키기 라벨에서 대상 값 , 이름 또는 둘 모두에 사용할 글꼴 , 글꼴 스타일 , 크기 등의 글꼴 상세 정보를 지정합니다 .

맞춤 가리키기 라벨에서 대상 값, 이름 또는 둘 모두에 사용할 텍스트 맞춤 방식을 지정합니다.

Graphlet 탭

사전 설정 사전 설정 그래프를 클릭하여 가리키기 라벨에 추가합니다. 사전 설정 그래프에 대한 자세한 내용은 "가리키기 라벨에 그래프 또는 이미지 추가"(483 페이지) 에서 확인하십시오.

그림 JSL 스크립트를 추가하여 가리키기 라벨에 추가할 그래프 또는 이미지의 유형을 정의합니다.

클릭 사용자가 " 그림 " 스크립트로 정의된 썸네일 (또는 가리키기 그래프) 을 클릭하면 실행되는 JSL 스크립트를 추가합니다. 여기에서 아무것도 정의하지 않으면 " 그림 " 스크립트가 다시 사용되므로 " 그림 " 에서 정의된 이미지 또는 그래프가 새 창에서 시작됩니다.

필터 가리키기 그래프에서 필터 열을 제거합니다. 제거하려는 열을 클릭하거나, " 필터 건너뛰기 플래그 " 에 "1" 을 추가합니다. 자세한 내용은 " 예 : 가리키기 라벨 그래프에서 필터 열 제거 "(491 페이지) 에서 확인하십시오.

썸네일 그래프 썸네일의 크기와 축 값 포함 여부를 지정합니다.

기타 JSL 을 사용하여 다음 정보를 지정합니다.

제목 가리키기 그래프가 시작될 때 생성되는 새 창의 제목을 설정합니다.

재적용 현재 가리키기 라벨 그래프의 사용자 정의 사항을 가리키기 그래프를 클릭하면 시작되는 새 창에 적용할지 여부를 정의합니다.

데이터 테이블에서 가리키기 라벨에 이미지 추가

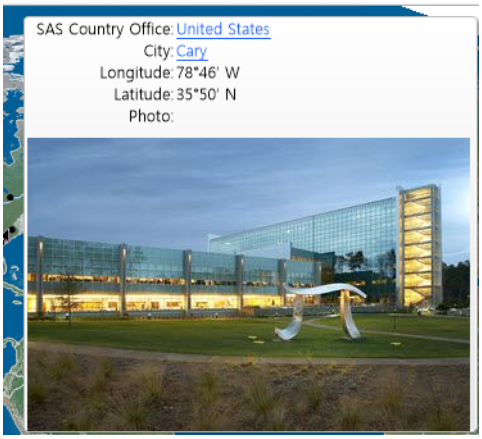
Graphlet 을 사용하여 가리키기 라벨에 이미지를 추가할 수 있지만 데이터 테이블에서도 이미지를 추가할 수 있습니다.

팁 : Graphlet 을 사용하여 가리키기 라벨에 JMP 그래프 또는 이미지를 추가하는 방법에 대한 자세한 내용은 " 그래프의 가리키기 라벨 사용자 정의 "(482 페이지) 에서 확인하십시오.

1. 데이터 테이블에 새 열을 생성합니다.
2. 열을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 후 **표현식** 데이터 유형을 선택합니다.
3. **확인**을 클릭합니다.
4. 컴퓨터에 있는 그래픽을 열의 각 셀로 드래그합니다 . 이미지 열이 포함된 데이터를 가져올 수도 있습니다.
5. 데이터 테이블의 열을 마우스 오른쪽 버튼으로 클릭하고 **라벨 / 라벨 해제**를 선택하여 라벨을 적용합니다.
6. 예를 들어 " 그래프 빌더 " 나 " 분포 " 에서 데이터 테이블을 사용하여 그래프를 생성합니다.
7. 데이터 점 또는 히스토그램 막대를 마우스로 가리킵니다.

   라벨에 이미지가 나타납니다. 예를 보려면 SAS Offices.jmp 샘플 데이터 테이블에서 "Graph Builder" 스크립트를 실행하십시오.

그림 9.72 가리키기 라벨의 이미지



팁 :

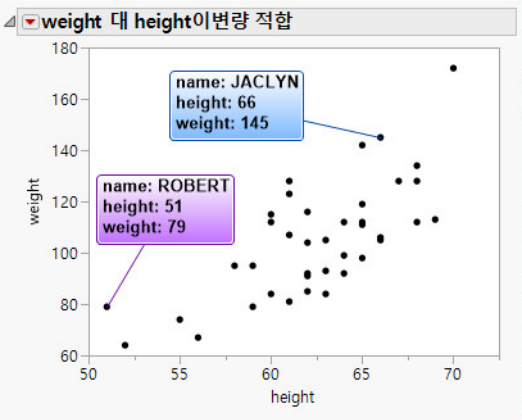
- 표현식 열의 이미지를 표식으로 사용할 수도 있습니다 . 자세한 내용은 " 이미지를 표식으로 사용 "(448 페이지) 에서 확인하십시오 .
- 그래프에 라벨이 항상 표시되도록 가리키기 라벨을 고정할 수 있습니다 . 자세한 내용은 " 그래프의 가리키기 라벨 고정 "(517 페이지) 에서 확인하십시오 .

그래프의 가리키기 라벨 고정

그래프의 데이터를 마우스로 가리키면 데이터 값이 포함된 가리키기 라벨이 표시됩니다. 이 가리키기 라벨이 단순 라벨로 그래프에 항상 표시되도록 가리키기 라벨을 고정할 수 있습니다. 이렇게 하면 여러 개의 라벨을 비교하거나, 라벨 정보를 표시하려는 경우에 유용합니다. 그래프를 저장하면 라벨이 저장됩니다. 그래프를 인쇄하면 라벨도 인쇄됩니다.

참고 : 가리키기 라벨은 분석을 다시 실행한 후에도 고정된 상태로 유지됩니다 . 하지만 데이터가 변경된 경우에는 라벨 위치를 조정해야 할 수 있습니다 .

그림 9.73 고정된 가리키기 라벨



가리키기 라벨을 고정하려면 데이터 점을 마우스로 가리켜 라벨이 나타날 때까지 기다립니다. 라벨의 오른쪽 상단에 있는 고정 아이콘을 클릭합니다. 그런 다음 라벨을 드래그하여 위치를 조정할 수 있습니다.

가리키기 라벨의 마우스 오른쪽 버튼 클릭 옵션

배경 색상 라벨 배경에 선택한 색상을 적용합니다.

프레임 색상 라벨 테두리 및 태그 선에 선택한 색상을 적용합니다.

텍스트 색상 텍스트에 선택한 색상을 적용합니다.

채우기 라벨에 배경 색상 및 프레임 색상을 적용합니다. 기본적으로 선택되어 있습니다.

그래디언트 사용 배경 색상을 세로 방향으로 밝은 색상부터 어두운 색상까지 음영 처리합니다. 기본적으로 선택되어 있습니다.

글꼴 텍스트에 선택한 글꼴군을 적용합니다. 기본 글꼴은 JMP "글꼴" 환경 설정에 따라 결정됩니다.

고정된 라벨이 있는 스크립트에서 컴퓨터에 없는 글꼴이 사용되고 있으면 스크립트가 실행되지 않습니다.

텍스트 추가 라벨의 맨 아래에 텍스트를 추가합니다.

텍스트 바꾸기 툴팁의 텍스트를 변경합니다. 분석을 다시 실행해도 툴팁 텍스트가 업데이트된 상태를 유지합니다.

태그 선 라벨과 데이터 점 사이에 선을 그립니다. 이 선에는 프레임 색상이 적용됩니다.

복사 가리키기 라벨의 콘텐츠를 클립보드에 복사합니다.

부분집합 가리키기 라벨과 연결된 부분집합 데이터가 있는 데이터 테이블을 엽니다.

편집 가리키기 라벨 편집기를 엽니다. 이 편집기에서 가리키기 라벨의 내용을 추가로 편집할 수 있습니다. 자세한 내용은 "그래프의 가리키기 라벨 편집"(511 페이지)에서 확인하십시오.

닫기 라벨을 제거합니다. 라벨의 오른쪽 상단에 있는 X를 클릭하여 라벨을 제거할 수도 있습니다.

참고 : 기본 라벨 형식으로 되돌리려면 라벨을 제거하고 다시 추가하십시오 .

10 장

데이터 저장 및 공유
JMP 데이터 활용

이 장에서는 다음과 같은 주제를 다룹니다.

- 데이터 테이블을 텍스트 파일 , SAS 데이터 집합 , Excel 파일 및 기타 형식으로 저장
- 보고서를 대화식 HTML 및 PowerPoint 로 저장
- 보고서 및 데이터 테이블을 전자 우편으로 전송
- 분석을 스크립트로 저장
- 보고서의 일부를 그래픽으로 저장
- 보고서 인쇄
- 보고서 복사 및 붙여넣기
- JMP 저널 또는 프로젝트 생성
- JMP 세션 저장
- 로그 창 저장

이후에 분석을 활용할 구체적인 방식에 따라 다음 중 적절한 방법을 선택할 수 있습니다.

- 데이터를 대화식으로 제공하려면 JMP 저널을 생성합니다 .
- 데이터 테이블 , 연결된 보고서 및 기타 항목을 다른 사용자에게 전송하려면 JMP 프로젝트를 생성합니다 .
- 나중에 보고서를 계속 탐색하려면 JMP 보고서를 생성합니다 .
- 보고서를 JMP 사용자가 아닌 사람에게 전송하려면 보고서를 대화식 HTML 또는 웹 보고서로 변환합니다 .

목차

데이터 테이블 저장 및 공유 524
데이터 테이블을 Microsoft Excel 파일로 저장 525
JMP 데이터를 CSV 파일로 저장 526
JMP 데이터를 텍스트 파일로 저장 527
JMP 데이터를 JSON 파일로 저장 528
JMP 데이터를 SAS 전송 파일로 저장 528
JMP 데이터를 Google 스프레드시트로 저장 529
JMP 데이터를 SAS 데이터 집합으로 저장 531
데이터 테이블을 데이터베이스에 저장 531
데이터베이스를 데이터 테이블로 바꾸기 533
보고서 저장 및 공유 533
보고서를 대화식 HTML로 공유 535
JMP Live 또는 JMP Public에 보고서 게시 539
보고서를 PDF 파일로 저장 548
보고서를 PowerPoint 프레젠테이션으로 저장 549
내보낸 그래픽에 대한 그래픽 DPI 설정 550
JMP 분석을 스크립트로 저장 551
보고서 요소를 그래픽 형식으로 저장 551
보고서 인쇄 552
보고서 복사 및 붙여넣기 553
프로젝트 작업 554
새 프로젝트 생성 555
프로젝트에서 파일 열기 555
프로젝트에서 파일 재배열 556
프로젝트 저장 557
프로젝트 작업 영역 558
프로젝트 책갈피 558
프로젝트 내용 559
프로젝트의 최근 사용한 파일 560
프로젝트 로그 560
파일을 프로젝트로 이동 561
프로젝트 내의 열기 스크립트 작성 561
새 프로젝트 생성의 예 562
저널 생성 565
예: 프레젠테이션용 저널 생성 566
새 저널 생성 568
공유 저널의 링크 관리 569
저널에 개요 추가 및 편집 570

저널의 마우스 오른쪽 버튼 클릭 옵션 571
저널에서 개요 수준 표시 제어 571
저널에 그래프 또는 그래픽 추가 572
저널의 사용자 정의 항목 572
저널 수정 방지 573
저널에 보고서 추가 573
다른 형식으로 저널 저장 573
저널에서 항목 삭제 574
JMP 세션 저장 574
JMP 종료 시 세션 저장 574
수동으로 세션 저장 575
로그 창 저장 576

데이터 테이블 저장 및 공유

파일 > 내보내기 또는 **파일 > 다른 이름으로 저장**을 선택하여 Microsoft Excel 또는 CSV와 같은 다른 형식으로 데이터 테이블을 저장합니다.

참고 : 데이터 테이블을 저널로 저장하려면 **편집 > 저널**을 선택한 후 **파일 > 다른 이름으로 저장**을 선택하십시오 .

JMP에서는 데이터 테이블을 표 10.1에 나열된 형식으로 저장합니다.

표 10.1 데이터 테이블을 저장하는 데 지원되는 파일 유형

파일 유형	사용 가능 OS
JMP 데이터 테이블(.jmp)	모두
Excel 통합 문서(.xls, .xlsx)	모두
텍스트 내보내기 파일(.txt)	모두
쉼표로 구분된 파일(.csv)	Windows
탭으로 구분된 파일(.tsv)	Windows
SAS 데이터 집합(.sas7bdat)	모두
확장 속성이 있는 SAS 데이터 집합(.sas7bxat) **참고 :** JMP 에서는 열 특성과 테이블 특성을 SAS 9.4 .sas7bxat 파일에 확장 속성으로 저장할 수 있습니다 . **파일 > 내보내기**를 선택하고 **SAS** 파일 유형을 선택한 후 확장 속성을 저장하기 위한 체크박스를 선택합니다 . JMP 에서는 이 파일의 이름을 .sas7bxat 확장자를 사용하여 지정합니다 . "SAS Integration" 환경 설정에도 이 옵션이 기본적으로 해제된 상태로 포함되어 있습니다 .	Windows
SAS 전송 파일(.xpt, .stx)	모두
JSON 데이터 파일(.json)	모두

해당하는 ODBC 드라이버가 있는 유형의 파일을 저장할 수도 있습니다. 자세한 내용은 "데이터 테이블을 데이터베이스에 저장"(531 페이지)에서 확인하십시오.

참고 :

- 데이터 테이블 이름의 최대 길이는 사용하는 컴퓨터의 운영 체제에 따라 다릅니다 . 자세한 내용은 운영 체제 설명서에서 확인하십시오 .
- JMP 는 충돌이 감지되면 수정된 파일을 저장하려고 합니다 .

데이터 테이블을 Microsoft Excel 파일로 저장

데이터 테이블을 Microsoft Excel 파일로 저장하기 위한 옵션은 다음 두 가지가 있습니다.

- 한 개의 데이터 테이블을 Microsoft Excel 형식 (.xls 또는 .xlsx) 으로 저장합니다 . 단일 테이블의 데이터가 해당 데이터 테이블을 따라 이름이 지정된 통합 문서에 저장됩니다 .
- 여러 테이블을 사용하여 .xlsx 형식의 통합 문서를 생성합니다 . 각 테이블의 데이터가 통합 문서의 서로 다른 워크시트에 나타납니다 .

.xlsx 파일로 저장하면 데이터 테이블 셀 색상, 숨겨진 데이터 및 제외된 데이터가 유지됩니다. 계산식과 대부분의 형식은 포함되지 않습니다.

Excel 파일 이름의 최대 길이는 컴퓨터의 운영 체제에 따라 결정됩니다.

한 개의 데이터 테이블을 Microsoft Excel 통합 문서로 저장

데이터 테이블을 Microsoft Excel 통합 문서로 저장하려면 다음 단계를 수행하십시오.

Windows

1. **파일 > 내보내기**를 선택합니다 .
2. **Microsoft Excel** 을 선택하고 **다음**을 클릭합니다 .
3. **파일 이름** 상자에 파일 이름을 입력합니다 .
4. (선택 사항) 파일을 .xls 파일로 저장하려면 **Excel 97-2003 형식 (.xls) 사용**을 선택합니다 .
5. (선택 사항) 저장 후 기본 스프레드시트 프로그램에서 파일이 열리지 않도록 하려면 " 내보내기 " 창에서 **저장 후 파일 열기**를 선택 취소합니다 .
6. **저장**을 클릭합니다 .

macOS

1. **파일 > 내보내기**를 선택합니다 .
2. **Microsoft Excel** 을 선택하고 형식을 선택합니다 .
3. **다음**을 클릭합니다 .
4. **다른 이름으로 저장** 상자에 파일 이름을 입력합니다 .
5. (선택 사항) 저장 후 기본 스프레드시트 프로그램에서 파일이 열리지 않도록 하려면 **저장 후 파일 열기**를 선택 취소합니다 .
6. **내보내기**를 클릭합니다 .

참고 : Windows 의 경우 , 열 또는 행 수가 Excel 에서 지원되는 개수를 초과하면 데이터가 여러 워크시트로 분할됩니다 .

여러 개의 데이터 테이블을 Microsoft Excel 통합 문서로 저장

여러 개의 데이터 테이블을 새 Microsoft Excel 통합 문서 또는 기존 Microsoft Excel 통합 문서에 저장하려면 다음 단계를 수행하십시오.

1. 데이터 테이블을 엽니다 .
2. (Windows) **파일 > 내보내기 > Microsoft Excel**을 선택하고 **여러 데이터 테이블 내보내기**를 선택한 후 **다음**을 클릭합니다 .
3. (macOS) **파일 > 내보내기**를 선택하고 Excel 목록에서 **여러 데이터 테이블 내보내기**를 선택한 후 **다음**을 클릭합니다 .
4. "JMP 데이터 테이블 " 열에서 저장할 데이터 테이블만 선택합니다 .
5. 필요한 경우 " 워크시트 이름 " 열에서 워크시트 탭의 이름을 지정합니다 .
6. " 통합 문서 이름 " 상자에 통합 문서의 이름을 입력합니다 .

   기본적으로 통합 문서의 이름은 현재 데이터 테이블의 이름을 따라 지정됩니다 .
7. **선택**클 클릭하고 파일을 저장한 디렉터리를 찾아 선택합니다 .

   기존 통합 문서를 덮어쓸 수 있습니다 . 이 경우 통합 문서에 이미 있던 워크시트가 대체됩니다 . 기존 통합 문서에 없던 워크시트는 다른 시트 끝에 나타납니다 .
8. **저장** (macOS 의 경우 **내보내기**) 을 클릭합니다 .

   파일이 .xlsx 파일로 저장되었습니다 .

JMP 데이터를 CSV 파일로 저장

Windows 의 JMP 에서는 JMP 데이터 테이블의 데이터를 CSV(쉼표로 구분된 값) 파일로 변환할 수 있습니다. 데이터 테이블을 CSV 파일로 저장하려면 다음을 수행하십시오.

1. 특정 데이터 테이블이 JMP 에 열려 있는 상태에서 **파일 > 다른 이름으로 저장**을 선택합니다 .
2. **파일 형식** 목록에서 "CSV(쉼표로 구분) (*.csv)" 를 선택합니다 .
3. 파일 이름을 입력합니다 .
4. CSV 파일의 위치를 선택합니다 .
5. **저장**을 클릭합니다 .

참고 : 데이터 테이블을 CSV 파일로 내보내면 쉼표와 CRLF(캐리지 리턴 줄바꿈) 문자가 파일 구분자로 사용됩니다 . 다른 구분자를 사용해야 하는 경우에는 다른 확장자를 사용하고 나중에 파일 이름을 CSV 확장자를 사용한 이름으로 바꾸십시오 .

JMP 데이터를 텍스트 파일로 저장

JMP 에서는 JMP 데이터 테이블의 데이터를 행 및 열이 있는 표준 텍스트 형식으로 변환할 수 있습니다 .

텍스트 형식 지정 옵션

열 이름을 텍스트 파일로 내보내기 열 머리글을 텍스트 첫 줄에 저장하고 라벨 또는 머리글 정보를 데이터와 함께 저장하려면 **테이블 머리글 내보내기**를 선택합니다 .

모든 열 이름에 인용 부호 추가 , 모든 문자 값에 인용 부호 추가 , 모든 숫자 값에 인용 부호 추가

인용 부호가 필요하지는 않지만 인용 부호를 추가하려면 옵션을 선택합니다 . CSV 파일을 가져오는 일부 프로그램에서는 인용 부호가 있어야 텍스트를 올바르게 표시할 수 있는 경우도 있습니다 . 인용 부호는 항상 필드 구분자 (일반적으로 쉼표 또는 탭), 줄 구분 기호 (일반적으로 캐리지 리턴) 또는 인용 부호 등의 특수 문자가 포함된 필드에 추가됩니다 .

필드 끝 필드 (또는 셀) 끝을 표시하는 문자 옆의 라디오 버튼을 선택합니다 . 적절한 문자가 나열되지 않는 경우에는 **기타**를 선택하고 문자를 입력합니다 .

행 끝 줄 (또는 행) 끝을 표시하는 문자 옆의 라디오 버튼을 선택합니다 . 적절한 문자가 나열되지 않는 경우에는 **기타**를 선택하고 문자를 입력합니다 .

참고 :

- 텍스트 데이터를 가져올 때 큰따옴표가 발견되면 구분자 규칙이 변경되며 JMP 는 닫는 큰따옴표를 찾습니다 . 따옴표 내에 포함된 공백을 비롯하여 다른 텍스트 구분자는 무시되고 텍스트 문자열의 일부로 처리됩니다 .
- 텍스트 데이터를 가져올 때 이미지 데이터가 포함된 열에는 표현식 데이터 유형이 할당됩니다 .

Windows

1. **파일 > 내보내기**를 선택합니다 .
2. **텍스트**를 선택하고 **다음**을 클릭합니다 .
3. " 파일 이름 " 상자에 파일 이름을 입력합니다 .
4. **옵션** 버튼을 클릭한 후 행 끝 및 필드 끝 문자를 지정하고 열 머리글을 텍스트로 내보낼지 여부를 선택합니다 .
5. **확인**을 클릭합니다 .
6. (선택 사항) 저장 후 기본 텍스트 편집기에서 파일이 열리지 않도록 하려면 **저장 후 파일 열기**를 선택 취소합니다 .
7. **저장**을 클릭합니다 .
8. **예**를 클릭합니다 .

macOS

1. **파일 > 내보내기**를 선택합니다 .
2. **텍스트**를 선택하고 **다음**을 클릭합니다 .
3. 다른 이름으로 저장 상자에 파일 이름을 입력합니다 .
4. 행 끝 및 필드 끝 문자를 지정하고 열 머리글을 텍스트로 내보낼지 여부를 선택합니다 .
5. (선택 사항) 저장 후 기본 텍스트 편집기에서 파일이 열리지 않도록 하려면 **저장 후 파일 열기**를 선택 취소합니다 .
6. **내보내기**를 클릭합니다 .

JMP 데이터를 JSON 파일로 저장

데이터 테이블을 이름 및 값 쌍으로 구성되는 JSON 파일로 저장할 수 있습니다. JSON 파일은 대개 응용 프로그램 간에 데이터를 교환하는 데 사용됩니다.

데이터 테이블을 JSON 파일로 저장하려면 다음 단계를 수행하십시오.

Windows

1. **파일 > 내보내기**를 선택합니다 .
2. **JSON 데이터**를 선택하고 **다음**을 클릭합니다 .
3. " 파일 이름 " 상자에 파일 이름을 입력합니다 .
4. **저장**을 클릭합니다 .

macOS

1. **파일 > 내보내기**를 선택합니다 .
2. **JSON 데이터**를 선택하고 **다음**을 클릭합니다 .
3. 다른 이름으로 저장 상자에 파일 이름을 입력합니다 .
4. **내보내기**를 클릭합니다 .

JMP 데이터를 SAS 전송 파일로 저장

JMP 데이터 테이블을 SAS 전송 파일 형식으로 저장할 수 있습니다. JMP는 파일 이름 및 열 이름의 공백을 밑줄로 바꾸고, 열 머리글을 대문자 SAS 변수로 변환한 후, 기타 사항을 전송 파일 규격에 맞게 변경합니다.

데이터 파일을 전송 파일로 저장할 때 기존 전송 파일에 추가할 수도 있습니다. 전송 파일을 JMP에서 다시 열면 두 개의 데이터 테이블의 개별 창에서 열립니다.

참고 :

- SAS 전송 파일의 최대 파일 이름 길이는 8 자입니다 . 이 한계를 초과하는 문자는 생략된다는 경고가 표시됩니다 .
- 표현식 열의 이미지는 200 자 이내여야 합니다 . 그렇지 않은 이미지는 지원되지 않습니다 .

Windows

1. **파일 > 내보내기**를 선택합니다 .
2. **SAS 전송 파일**을 선택합니다 .
3. **다음**을 클릭합니다 .
4. " 파일 이름 " 상자에 파일 이름을 입력합니다 .
5. **저장**을 클릭합니다 .
6. **확인**을 클릭합니다 .

   다른 SAS 형식의 수정에 대한 알림이 나타납니다 .
7. 각 알림에서 **확인**을 클릭합니다 .
8. (선택 사항) 데이터 테이블을 기존 SAS 전송 파일에 추가합니다 . **파일 > 내보내기**를 선택하고 **추가** 체크박스를 선택한 후 **다음**을 클릭합니다 . 데이터 테이블을 추가할 대상 파일을 선택한 후 **저장**을 클릭합니다 .

macOS

1. **파일 > 내보내기**를 선택합니다 .
2. **SAS 전송**을 선택합니다 .
3. **다음**을 클릭합니다 .
4. **다른 이름으로 저장** 상자에 파일 이름을 입력한 후 **내보내기**를 클릭합니다 .

   SAS 형식의 수정에 대한 알림이 나타납니다 .
5. 각 알림에서 **확인**을 클릭합니다 .
6. (선택 사항) 데이터 테이블을 기존 SAS 전송 파일에 추가합니다 . **파일 > 내보내기**를 선택하고 **추가** 체크박스를 선택한 후 **다음**을 클릭합니다 . 데이터 테이블을 추가할 대상 파일을 선택한 후 **추가**를 클릭합니다 .

JMP 데이터를 Google 스프레드시트로 저장

Google 스프레드시트는 Google 계정에 저장된 스프레드시트(탭 같은 시트로 구성)입니다. JMP 데이터 테이블을 Google 스프레드시트로 내보내면 다른 사람과 데이터를 공유하고 JMP 외부에서 데이터를 편집할 수 있습니다.

JMP 데이터 테이블을 기존 Google 스프레드시트로 내보내거나 새 Google 스프레드시트를 생성할 수 있습니다.

JMP 데이터 테이블을 Google 스프레드시트로 내보내려면 다음 단계를 수행하십시오 .

1. 기존 스프레드시트로 내보내려면 먼저 다음 단계를 수행합니다 . 그렇지 않으면 2 단계로 이동합니다 .
   a. 브라우저에서 https://sheets.google.com 에 로그인합니다 .
   b. 기존 스프레드시트를 엽니다 .
   c. 브라우저의 주소 표시줄에서 URL 을 복사합니다 .
2. JMP 에서 내보내려는 데이터가 있는 데이터 테이블을 엽니다 .
3. **파일 > 내보내기 > Google 스프레드시트**를 선택합니다 .
4. " 스프레드시트 유형 " 에서 **기존 스프레드시트** 또는 **새 스프레드시트**를 선택합니다 .
5. Google 계정 이름 (@gmail.com 은 필요하지 않음) 을 입력합니다 .
6. (기존 스프레드시트의 경우만 해당) 1 단계에서 복사한 Google 스프레드시트의 URL 을 " 시트 이름 " 옆에 붙여넣습니다 .
7. (새 스프레드시트의 경우만 해당) 새 스프레드시트의 이름을 입력합니다 .
8. 시트 이름 (탭 이름) 을 입력합니다 .
9. **저장**을 클릭합니다 .
10. JMP 에서 데이터 테이블을 이미 내보낸 경우 해당 데이터가 Google 스프레드시트에 표시됩니다 . Google 스프레드시트로 이동하여 지정한 시트의 데이터를 확인합니다 .

    또는

    지정한 Google 계정을 처음 사용하여 데이터를 내보내는 경우 JMP 에 Google 계정에 대한 액세스 권한을 부여하라는 메시지가 표시됩니다 .
    a. **허용**을 클릭합니다 .
    b. Google 스프레드시트로 이동하여 지정한 시트의 데이터를 확인합니다 .

    데이터 테이블의 크기에 따라 저장 작업에 몇 시간이 걸릴 수도 있습니다 .

Google 스프레드시트 내보내기

- 계산식 및 " 목록 확인 " 열 특성과 같은 JMP 기능은 Google 스프레드시트에서 지원되지 않습니다 .
- 스프레드시트가 비어 있으면 JMP 로그에서 오류 메시지를 확인하십시오 . Windows 의 경우 , **보기 > 로그**를 선택합니다 . macOS 의 경우 , **창 > 로그**를 선택합니다 .
- Google 스프레드시트에 시트 하나만 있는 경우 시트를 새 시트로 바꿀 수 없습니다 . 고유한 이름의 두 번째 시트를 추가한 다음 원래 시트를 삭제합니다 . "Google 스프레드시트로 내보내기 " 창에서 **기존 스프레드시트**를 선택하고 데이터 테이블을 다시 내보냅니다 . 지정한 시트 이름이 시트에 나타납니다 .
- 보안 , 국가별 제한 등에 대한 자세한 내용은 " 데이터 가져오기 " 장의 "Google 스프레드시트 가져오기 "(167 페이지) 에서 확인하십시오 .

- Windows 의 경우 Internet Explorer 에서 **호환성 보기에서 인트라넷 사이트 표시**를 선택하면 Google 스프레드시트를 내보낼 수 없습니다. 이 문제를 방지하려면 다음 단계를 따르십시오.
  1. Internet Explorer 에서 " 도구 " 버튼 을 선택한 다음 **호환성 보기 설정**을 선택합니다 .
  2. **호환성 보기에서 인트라넷 사이트 표시**를 선택 취소합니다 .

JMP 데이터를 SAS 데이터 집합으로 저장

데이터 테이블을 SAS 버전 7 이상의 데이터 집합으로 저장할 수 있습니다. JMP는 열을 SAS 변수로 쓰고, 행을 SAS 관측값으로 쓴 후, 기타 사항을 SAS 데이터 집합 규격에 맞게 변경합니다.

데이터를 SAS 데이터 집합으로 내보낼 경우 JMP 날짜 열은 적절한 SAS 형식의 SAS 날짜 값이 됩니다.

Windows의 경우, JMP에서는 열 특성과 테이블 특성을 SAS 9.4 .sas7bxat 파일에 확장 속성으로 저장할 수 있습니다. 이 옵션은 "다른 이름으로 JMP 파일 저장" 창과 "SAS Integration" 환경 설정에서 사용할 수 있습니다.

Windows

1. **파일 > 내보내기**를 선택합니다 .
2. **SAS 데이터 집합**을 선택하고 **다음**을 클릭합니다 .
3. " 파일 이름 " 옆에 파일 이름을 입력합니다 .
4. (선택 사항) SAS 변수 이름 또는 SAS 형식을 저장하려면 **SAS 변수 이름 유지** 또는 **SAS 형식 유지** 체크박스를 선택합니다 .
5. (선택 사항) JMP 메타데이터를 SAS 9.4 용 확장 속성으로 저장하려면 **테이블 및 열 특성을 SAS 9.4 확장 속성에 저장**을 선택합니다 . JMP 는 자동으로 .sas7bxat 파일 확장자를 할당합니다 .
6. **저장**을 클릭합니다 .
7. **예**를 클릭합니다 .
8. 다른 알림이 나타나면 각 알림에서 **확인**을 클릭합니다 .

macOS

1. **파일 > 내보내기**를 선택합니다 .
2. **SAS** 를 선택하고 **다음**을 클릭합니다 .
3. 파일 이름을 입력하고 내보내기 옵션을 선택합니다 .
   – 행 상태가 " 제외됨 " 인 행을 내보내지 않으려면 **제외된 행 제외 유지**를 선택합니다 . 다른 옵션은 이전 섹션에서 설명한 것과 같습니다 .
4. **내보내기**를 클릭합니다 .

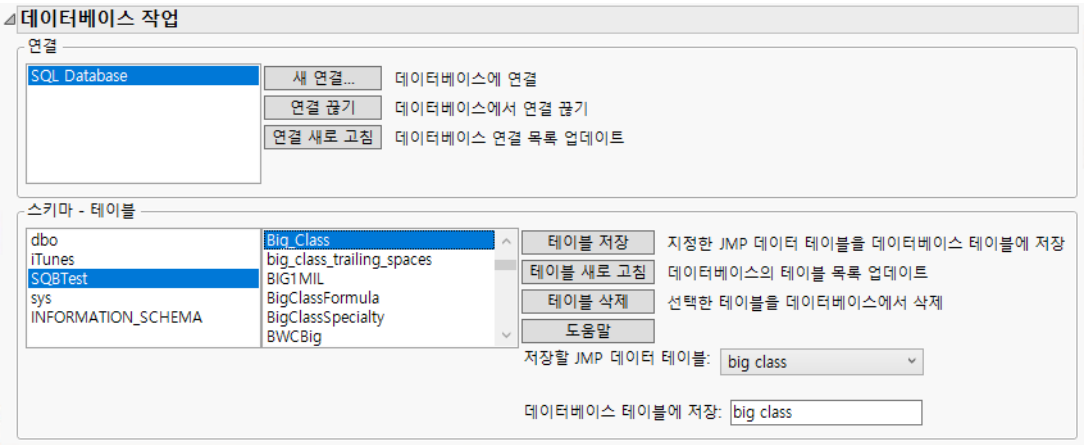
데이터 테이블을 데이터베이스에 저장

데이터 테이블을 해당하는 규격의 ODBC(Open Database Connectivity) 드라이버가 있는 시스템의 데이터베이스에 저장할 수 있습니다.

데이터 테이블을 ODBC를 사용하는 데이터베이스로 내보낼 경우 JMP는 데이터를 문자(CHAR 또는 NCHAR) 또는 숫자(FLOAT 또는 DOUBLE PRECISION) 데이터 유형으로 내보냅니다. JMP 날짜 또는 시간으로 정의된 열이 있으면 해당 열은 문자 데이터로 내보내집니다. 수신하는 데이터베이스에서 데이터 유형을 올바른 날짜 형식으로 변경하는 작업을 처리합니다.

1. 먼저, 데이터베이스에 연결합니다. 자세한 내용은 "새 데이터베이스 연결을 추가하려면"(532 페이지) 에서 확인하십시오 .
2. **파일 > 데이터베이스 > 테이블 저장**을 선택합니다 .

그림 10.1 데이터베이스 테이블 저장 창



3. " 연결 " 상자에서 파일을 저장할 대상 데이터베이스의 이름을 강조 표시합니다 (그림 10.1). " 연결 " 상자에는 시스템이 연결된 데이터베이스의 목록이 포함되어 있습니다 .
4. **저장할 JMP 데이터 테이블** 메뉴에서 데이터베이스에 저장할 열려 있는 JMP 데이터 테이블을 선택합니다 .
5. **데이터베이스 테이블에 저장** 상자에 해당 데이터 테이블을 데이터베이스에 저장할 때 사용할 테이블 이름을 입력합니다 .
6. **테이블 저장**을 클릭합니다 .

새 데이터베이스 연결을 추가하려면

1. **연결**을 클릭합니다 .
2. 원하는 데이터 소스를 선택하고 **확인**을 클릭합니다 . 새 소스를 생성하려면 **새로 만들기** 버튼 (Windows) 또는 **DSN 선택** 버튼 (macOS) 을 클릭합니다 .

선택하는 데이터 소스와 컴퓨터에 설치된 데이터베이스 드라이버에 따라 다양한 창이 나타날 수 있습니다 . 나타나는 창을 사용하여 데이터베이스 소스를 생성합니다 .

3. 파일을 저장할 대상 데이터베이스를 선택합니다 .

데이터베이스를 데이터 테이블로 바꾸기

데이터베이스를 데이터 테이블로 바꾸려면 다음 단계를 수행하십시오 .

1. 먼저 , 데이터베이스에 연결합니다 . 자세한 내용은 "새 데이터베이스 연결을 추가하려면"(532 페이지) 에서 확인하십시오 .
2. **파일 > 데이터베이스 > 테이블 저장**을 선택합니다 .
3. " 연결 " 상자에서 파일을 바꿀 대상 데이터베이스의 이름을 선택합니다 . " 연결 " 상자에는 시스템이 연결된 데이터베이스의 목록이 포함되어 있습니다 .
4. " 스키마 - 테이블 " 상자에서 바꾸려는 데이터베이스 테이블을 선택합니다 .

참고 : 이 데이터베이스 테이블의 이름을 기억했다가 대체 데이터베이스 테이블에 이와 동일한 이름을 사용하십시오 .

5. **테이블 삭제**를 클릭합니다 .

   해당 데이터베이스 테이블이 데이터베이스에서 삭제됩니다 .
6. **저장할 JMP 데이터 테이블** 목록에서 데이터베이스에 저장할 JMP 데이터 테이블을 선택합니다 .
7. **데이터베이스 테이블에 저장** 상자에 삭제한 데이터베이스 테이블의 이름을 입력합니다 .
8. **테이블 저장**을 클릭합니다 .

   해당 데이터베이스 테이블이 데이터베이스에 저장됩니다 .
9. **연결 끊기**를 클릭합니다 .
10. " 데이터베이스 테이블 저장 " 창을 닫습니다 .

보고서 저장 및 공유

보고서를 저장하려면 **파일 > 내보내기**를 선택합니다 .

- Windows 의 경우 , 보고서를 표 10.2 에 나열된 파일 형식으로 저장할 수 있습니다 .
- macOS 의 경우 , 보고서는 JMP 보고서 (.jrp) 로 저장됩니다 . 다음 사항에 유의하십시오 .
  - **편집 > 저널**을 선택한 후 **파일 > 저장**을 선택하여 보고서를 저널로 저장할 수 있습니다 .
  - 보고서를 텍스트 , 이미지 , HTML, PowerPoint 또는 RTF 파일로 저장하려면 **파일 > 내보내기**를 선택합니다 . 지원되는 이미지 파일에는 .png, .emf, .jpg, .gif, .tiff, .svg 및 .eps 가 포함됩니다 .

표 10.2 Windows 에서 지원되는 보고서 파일 유형

파일 유형	설명
Microsoft Word(.doc)	그림, 텍스트 및 테이블이 혼합된 워드 프로세서 형식입니다.
확장 메타파일(.emf)	벡터 구성 요소와 비트맵 구성 요소를 모두 포함할 수 있습니다.
EPS(Encapsulated PostScript) 파일(.eps)	이미지의 비트맵 미리보기를 표시할 수 있는 선 그림(또는 **벡터** 이미지)입니다.
CompuServe 그래픽 교환 형식(.gif)	압축된 비트맵 그림입니다.
HTML(Hypertext Markup Language) 형식(.htm, .html)	마크업 텍스트와 별도의 그림 파일에 대한 참조를 포함하는 브라우저 형식입니다. 그림은 PNG, JPEG, SVG 또는 GIF 중 한 가지 형식을 선택하여 HTML 파일 내에 저장합니다.
HTML 또는 데이터가 포함된 대화식 HTML 형식(.htm, .html)	JMP를 사용하지 않는 사람이 데이터를 볼 수 있도록 데이터, 보고서 및 그래프를 대화식 웹 페이지에 저장합니다. 자세한 내용은 "보고서를 대화식 HTML로 공유"(535페이지)에서 확인하십시오.
JPEG(Joint Photographics Expert Group) 형식(.jpg, .jpeg)	압축된 비트맵 그림으로, 사진의 표준 형식입니다. 자세한 내용은 "내보낸 그래픽에 대한 그래픽 DPI 설정"(550페이지)에서 확인하십시오.
JMP 저널(.jrn)	저널이라는 별도의 창에 복사된 분석 보고서입니다. 이를 편집하거나 여기에 다른 보고서를 추가할 수 있습니다. 자세한 내용은 "저널 생성"(565페이지)에서 확인하십시오. macOS의 경우, **편집 > 저널**을 선택한 후 **파일 > 저장**을 선택합니다. Windows의 경우, **파일 > 저장**을 선택한 후 저널 형식을 선택합니다.
JMP 보고서(.jrp)	JMP에서 생성된 원본 분석 보고서입니다. 분석을 계속하기 위해 이 보고서를 다시 열 수 있습니다. JSL 데이터와 포함된 데이터, 또는 실시간 보고서를 다시 생성하기 위한 데이터 참조를 포함합니다. 예를 들어 새 데이터가 있을 때 보고서 작업을 다시 수행하려면 이 형식을 선택합니다.
PDF(Portable Document Format) 형식(.pdf)	문서가 생성된 운영 체제 또는 응용 프로그램에 관계없이 문서를 공유하기 위한 형식입니다.

표 10.2 Windows 에서 지원되는 보고서 파일 유형 (계속)

파일 유형	설명
PNG(Portable Network Graphics) 형식 (.png)	GIF 에 이어 등장한 압축된 비트맵 그림입니다. 자세한 내용은 "내보낸 그래픽에 대한 그래픽 DPI 설정"(550 페이지) 에서 확인하십시오.
PowerPoint 프레젠테이션 (.pptx)	그림, 텍스트 및 테이블이 혼합된 Microsoft PowerPoint 형식입니다. 자세한 내용은 "보고서를 PowerPoint 프레젠테이션으로 저장"(549 페이지) 에서 확인하십시오.
RTF(서식 있는 텍스트 형식) 형식 (.rtf)	그림, 텍스트 및 테이블이 혼합된 워드 프로세서 형식입니다. 그림은 PNG, JPEG 또는 EMF 중 한 가지 형식을 선택하여 RTF 파일 내에 저장합니다.
SVG(Scalable Vector Graphic) 형식 (.svg)	텍스트로 저장되는 그림으로, 2 차원 그래픽에 사용하기에 적합합니다.
TIFF(Tagged Image File Format) 형식 (.tif, .tiff)	래스터 파일 형식입니다. 자세한 내용은 "내보낸 그래픽에 대한 그래픽 DPI 설정"(550 페이지) 에서 확인하십시오.
텍스트 형식 (.txt)	일반 텍스트 형식으로, 그림을 포함하지 않습니다.

보고서를 대화식 HTML 로 공유

대화식 HTML 을 통해 동적 그래프가 포함된 JMP 보고서를 공유하면 JMP 를 사용하지 않는 사람도 데이터를 탐색할 수 있습니다. JMP 보고서는 공유 네트워크 드라이브나 웹 사이트에서 또는 이메일을 통해 다른 사람과 공유할 수 있는 대화식 웹 페이지로 저장됩니다. 그러면 사용자들이 JMP 에서처럼 데이터를 탐색할 수 있습니다.

대화식 HTML 의 데이터 포함

보고서를 대화식 HTML 로 내보내거나 게시하면 데이터가 HTML 에 포함됩니다. 웹 브라우저는 암호화된 데이터를 읽을 수 없기 때문에 내용이 암호화되지 않습니다. 중요한 데이터를 공유하지 않으려면 **파일 > 내보내기 > HTML** 을 선택하여 결과를 대화식이 아닌 웹 페이지로 내보내십시오.

대화식으로 지원되는 기능

대화식 HTML 은 다음과 같은 JMP 의 일부 기능을 제공합니다.

- 보고서의 기능이 완전히 지원되면 경고 없이 웹 페이지가 생성됩니다 .

- 보고서에 지원되지 않는 기능이 포함되어 있으면 내보내기 또는 게시 창에 대화식 HTML 이 부분적으로 구현되었다는 메시지가 표시됩니다 . 자세한 내용은 JMP 로그 (**보기** > **로그**) 에서 확인하십시오 .
- 부분적으로 지원되거나 지원되지 않는 기능은 웹 페이지에 정적 상태로 나타납니다. 웹 페이지에서 지원되지 않는 기능을 커서로 가리키면 해당 기능이 아직 대화식이 아니라는 툴팁이 표시됩니다 .

대화식 HTML 출력을 탐색하는 방법에 대한 자세한 내용은 https://www.jmp.com/interactive 에서 확인하십시오.

단일 보고서를 대화식 HTML 로 내보내기

단일 JMP 보고서를 대화식 HTML로 내보내려면 " 데이터가 포함된 대화식 HTML로 내보내기 " 옵션을 사용하여 단일 웹 페이지를 생성합니다.

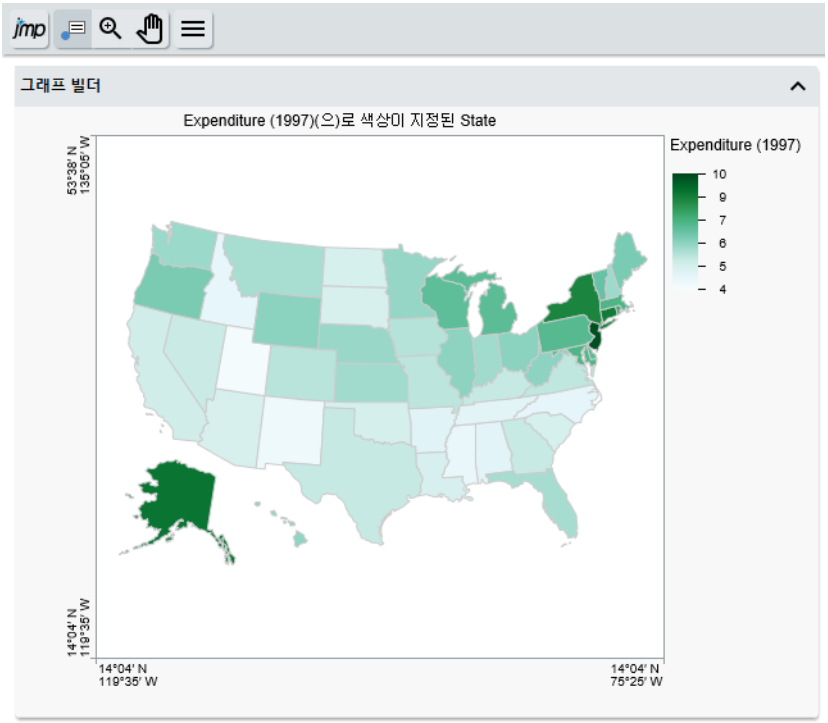
1. JMP 에서 보고서를 생성하고 해당 창을 활성화합니다 .

   참고 : 보고서에 로컬 데이터 필터가 포함되어 있으면 웹 보고서에서 정적 상태이며 사용자가 선택을 변경할 수 없습니다 . 로컬 데이터 필터를 대화식으로 만들려면 **포함** 모드를 선택 취소하십시오 . 그래프 빌더에서 **포함** 모드를 표시하려면 " 로컬 데이터 필터 " 의 빨간색 삼각형 메뉴에서 **모드 표시**를 선택합니다 .

2. **파일 > 내보내기**를 선택하고 **데이터가 포함된 대화식 HTML** 을 선택한 후 **다음**을 클릭합니다 .
3. 파일 이름을 지정합니다 .
4. (선택 사항) HTML 파일을 내보낸 후 기본 브라우저에서 열려면 **저장 후 파일 열기**를 선택합니다 .
5. **저장**을 클릭합니다 .

   출력이 선택한 폴더에 저장됩니다 .

그림 10.2 단일 대화식 HTML 보고서에 대한 웹 페이지



여러 보고서를 대화식 HTML 로 게시

여러 JMP 보고서를 대화식 HTML 로 게시하려면 "파일에 게시" 옵션을 사용하여 보고서가 포함된 인덱스 페이지를 생성합니다.

팁 : 여러 JMP 보고서를 게시하는 경우 인덱스 페이지와 보고서 파일을 저장할 위치를 지정해야 합니다. 공유 네트워크 폴더를 선택하고 다른 사람에게 위치를 알려주거나, 공유하기 전에 사용자 컴퓨터의 폴더를 선택하고 파일을 압축할 수 있습니다. 이 방법은 JMP 를 사용하지 않는 사람과 공유할 때 특히 유용합니다.

1. JMP 에서 보고서를 생성합니다.

   참고 : 보고서에 로컬 데이터 필터가 포함되어 있으면 웹 보고서에서 정적 상태이며 사용자가 선택을 변경할 수 없습니다. 로컬 데이터 필터를 대화식으로 만들려면 **포함** 모드를 선택 취소하십시오. 그래프 빌더에서 **포함** 모드를 표시하려면 " 로컬 데이터 필터 " 의 빨간색 삼각형 메뉴에서 **모드 표시**를 선택합니다.

2. 보고서 창에서 **파일** > **게시** > **파일에 게시**를 선택합니다.
3. 게시할 보고서를 선택합니다.
4. (선택 사항) 상위 폴더 위치를 변경하거나 보고서를 포함할 하위 폴더의 이름을 변경합니다.

5. **다음**을 클릭합니다 .
6. 인덱스 페이지의 제목을 입력합니다 . 보고서 제목을 업데이트할 수도 있습니다 .
7. (선택 사항) 추가 옵션을 변경합니다 . 자세한 내용은 대화식 HTML 보고서 옵션에서 확인하십시오 .
8. **게시**를 클릭합니다 .

그림 10.3 여러 대화식 HTML 보고서에 대한 인덱스 페이지

9. 썸네일을 클릭하여 보고서를 엽니다 .

대화식 보고서 작업에 대한 자세한 내용을 보려면 HTML 보고서에서 ☰ > **도움말**을 클릭합니다. 그러면 도움말 페이지 (https://www.jmp.com/interactive) 가 열립니다.

대화식 HTML 보고서 옵션

제목 인덱스 페이지 (보고서가 여러 개인 경우만 해당) 또는 보고서의 제목을 추가합니다 .

설명 (선택 사항) 인덱스 페이지 또는 보고서에 설명을 추가합니다 . 이 설명은 제목 아래에 표시됩니다 .

사용자 정의 (보고서가 여러 개인 경우에만 표시) 웹 페이지의 모양을 변경합니다 . 스타일 , 테마 , 글꼴 , 로고 및 날짜 / 시간 표시 여부를 변경할 수 있습니다 . 자세한 내용은 인덱스 페이지 사용자 정의 옵션에서 확인하십시오 .

데이터 게시 대화식 보고서의 경우 이 옵션을 선택합니다. 이 옵션을 선택 취소하면 정적 보고서가 됩니다 .

참고 : 중요한 데이터를 공유하지 않으려면 결과를 대화식이 아닌 웹 페이지로 저장하십시오 . **파일 > 내보내기 > HTML** 을 선택합니다 .

이미지 추가 웹 페이지 아래쪽에 이미지를 추가합니다 .

게시된 웹 보고서 열기 " 게시 " 를 클릭하면 브라우저에서 웹 페이지를 엽니다 .

실행 후 보고서 닫기 웹 보고서를 게시하면 JMP 에서 보고서를 닫습니다 .

삭제 아이콘 (보고서가 여러 개인 경우에만 표시) 보고서를 삭제합니다 .

화살표 아이콘 (보고서가 여러 개인 경우에만 표시) 보고서 순서를 변경합니다 .

인덱스 페이지 사용자 정의 옵션

스타일 형식 보고서의 레이아웃을 결정합니다 .

큰 목록 보고서를 큰 썸네일 그래픽과 함께 열 하나에 표시합니다 .

작은 목록 보고서를 작은 썸네일 그래픽과 함께 열 하나에 표시합니다 .

격자 보고서를 여러 행에 표시합니다 .

사용자 CSS 웹 페이지의 형식을 지정하기 위한 CSS 파일을 지정할 수 있습니다 . CSS 파일은 _css 라는 하위 폴더에 복사됩니다 . 보고서와 지원 파일을 다른 사용자에게 전송하면서 CSS 형식 지정을 유지할 수 있도록 CSS 파일에 대한 링크는 상대적입니다 .

색상 테마 웹 페이지 , 머리글 및 테두리의 색상을 지정합니다 . 기본 웹 페이지의 배경은 흰색 , 머리글은 주황색 , 테두리는 파란색입니다 .

글꼴 변경 보고서에 적용된 글꼴을 변경합니다 .

로고 변경 보고서와 함께 표시할 이미지를 지정합니다 . 이미지를 보고서의 위 또는 아래로 이동하려면 이미지 옆의 위쪽 또는 아래쪽 화살표를 클릭합니다 .

날짜 / 시간 표시 웹 페이지가 생성된 날짜 및 시간을 표시하거나 숨깁니다 .

JMP Live 또는 JMP Public 에 보고서 게시

JMP Live는 사용자의 회사에서 JMP 콘텐츠를 공유할 용도로 구매한 전용 웹 사이트입니다. JMP Live는 SAS 또는 사용자 회사에서 호스팅할 수 있습니다. JMP Public은 JMP Live와 동일한 기술을 사용하는 웹 사이트이지만 누구나 JMP 콘텐츠를 모든 사용자와 공유할 수 있다는 차이점이 있습니다.

JMP Live와 JMP Public은 JMP 데이터, 시각화 요소 및 대시보드를 공유하기 위한 플랫폼입니다. JMP에서 보고서를 준비한 후 게시물로 게시하여 동료(JMP가 없을 수도 있음)와 공동 작업을 수행하십시오.

JMP Live와 JMP Public에서는 데이터 탐색을 위해 JMP의 시각적 대화식 작업을 제공하며 이를 통해 다음을 수행할 수 있습니다.

- 모든 보고서를 한 곳에 저장합니다 .

- 개별 보고서에 대해 해당 보고서를 봐야 하는 사람과의 공유 상태를 제어합니다 .
- 팀이 최신 보고서를 볼 수 있도록 보고서를 액세스하기 쉬운 위치에 보관합니다 .
- 보고서의 기능을 다른 위치까지 확장하여 웹 페이지 , 블로그 , 소셜 미디어 게시물 등에 보고서를 포함하고 대화식 작업을 수행합니다 .

SAS 프로파일 등록 및 로그인 (JMP Public 에만 해당)

JMP Public에 게시물을 게시하려면 먼저 SAS 프로파일을 사용하여 등록하고 로그인해야 합니다.

참고 : JMP Public 에 로그인하지 않은 경우 일부 게시물만 볼 수 있습니다 .

1. 브라우저에서 https://public.jmp.com 으로 이동합니다 .
   지원되는 브라우저는 Chrome, Firefox, Safari, Edge(Chromium 기반) 입니다 .
2. 오른쪽 상단에서 **Sign in** 을 클릭합니다 .
3. 이미 SAS 프로파일이 있으면 해당 자격 증명을 사용하여 로그온한 다음 **Submit** 를 클릭합니다.
4. SAS 프로파일이 없으면 다음을 수행합니다 .
   a. **Register** 를 클릭합니다 . 양식을 완성하고 **Create profile** 을 클릭합니다 .
   b. 전자 우편의 지침을 따라 암호를 설정하고 프로파일을 활성화합니다 .

JMP Live 서버 추가

JMP Live에 게시하려면 먼저 JMP Live 서버를 추가해야 합니다.

1. JMP 에서 **파일** > **게시** > **연결 관리**를 선택합니다 .
2. **새로 추가**를 클릭합니다 .
3. 연결을 식별하는 이름 (예 : JMP Live) 을 입력합니다 .
4. https://jmpliveserver.com:3501 과 같은 형식으로 지정된 서버 URL 을 입력합니다 (3501 은 기본 웹 서버 포트).
5. (선택 사항) JSL 을 사용하여 JMP Live 보고서를 생성하려면 API 키를 입력하십시오 .
6. **다음**을 클릭합니다 .
7. 사용자 이름과 암호를 사용하여 JMP Live 서버에 로그인합니다 .
8. 연결이 성공적으로 생성되면 **저장**을 클릭합니다 .
   " 연결 선택 " 아래에 서버 이름이 나타납니다 .
9. **종료**를 클릭합니다 .

JMP 보고서 생성 및 게시

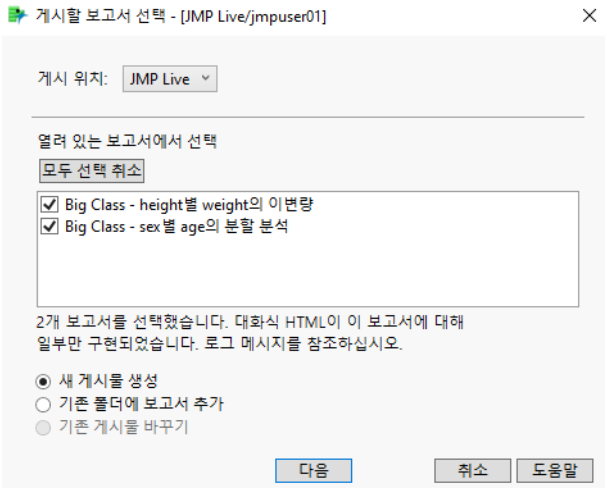
1. JMP 에서 보고서를 생성합니다 .

JMP Live 또는 JMP Public 에 자체 게시물로 표시되는 단일 보고서를 생성하거나 폴더에 표시되는 여러 보고서를 생성할 수 있습니다 .

2. (보고서에 로컬 데이터 필터가 포함되지 않은 경우 건너뜀) 보고서에 로컬 데이터 필터가 포함된 경우 게시하기 전에 열을 추가하십시오 . 그렇지 않으면 로컬 데이터 필터가 활성 필터가 아닌 정적 필터가 됩니다 . 열을 추가하려면 로컬 데이터 필터에서 열을 선택하고 " 추가 " 아이콘을 클릭하십시오 .
3. **파일 > 게시 > JMP Live 에 게시** 또는 **JMP Public 에 게시**를 선택합니다 .
   - (JMP Live 에만 해당) JMP Live 서버를 추가하지 않은 경우 지금 추가하십시오 . 자세한 내용은 "JMP Live 서버 추가 "(540 페이지) 에서 확인하십시오 .
   - JMP Live 또는 JMP Public 에 로그인해야 할 수 있습니다 .

   참고 :SAS 로그인을 등록하려면 https://www.jmp.com 을 방문하고 페이지 상단에 있는 검색 상자 옆의 **로그인**을 클릭하십시오 .

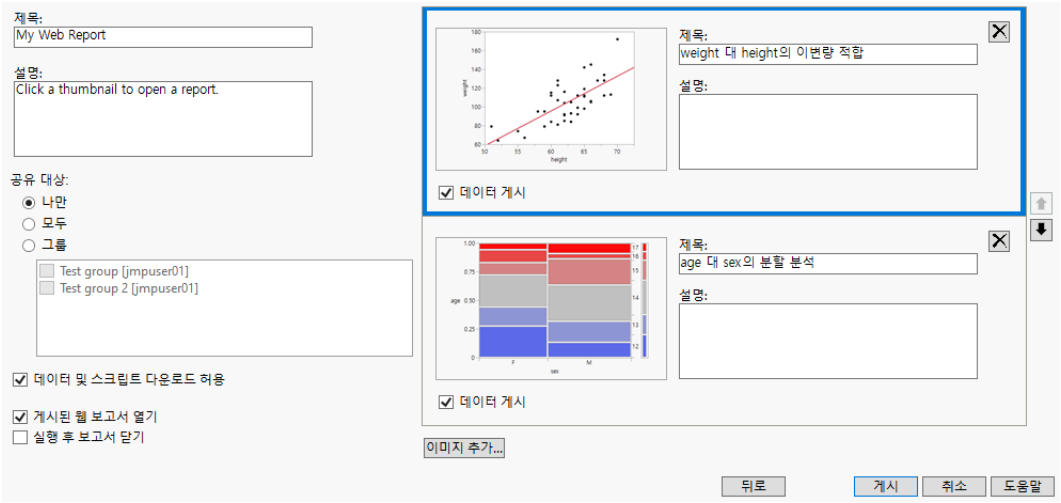
그림 10.4 보고서 선택 창



4. (JMP Live 에만 해당) **게시 위치** 옆에서 JMP Live 서버를 선택합니다 .
5. 게시할 보고서를 선택하여 엽니다 .
6. 작업을 선택한 후 **다음**을 클릭합니다 .
   - " 새 게시물 생성 "(542 페이지)
   - " 기존 폴더에 보고서 추가 "(543 페이지)
   - " 게시물의 기존 보고서 바꾸기 "(545 페이지)(**정확한 이름의 게시물이 있는 경우**)

새 게시물 생성

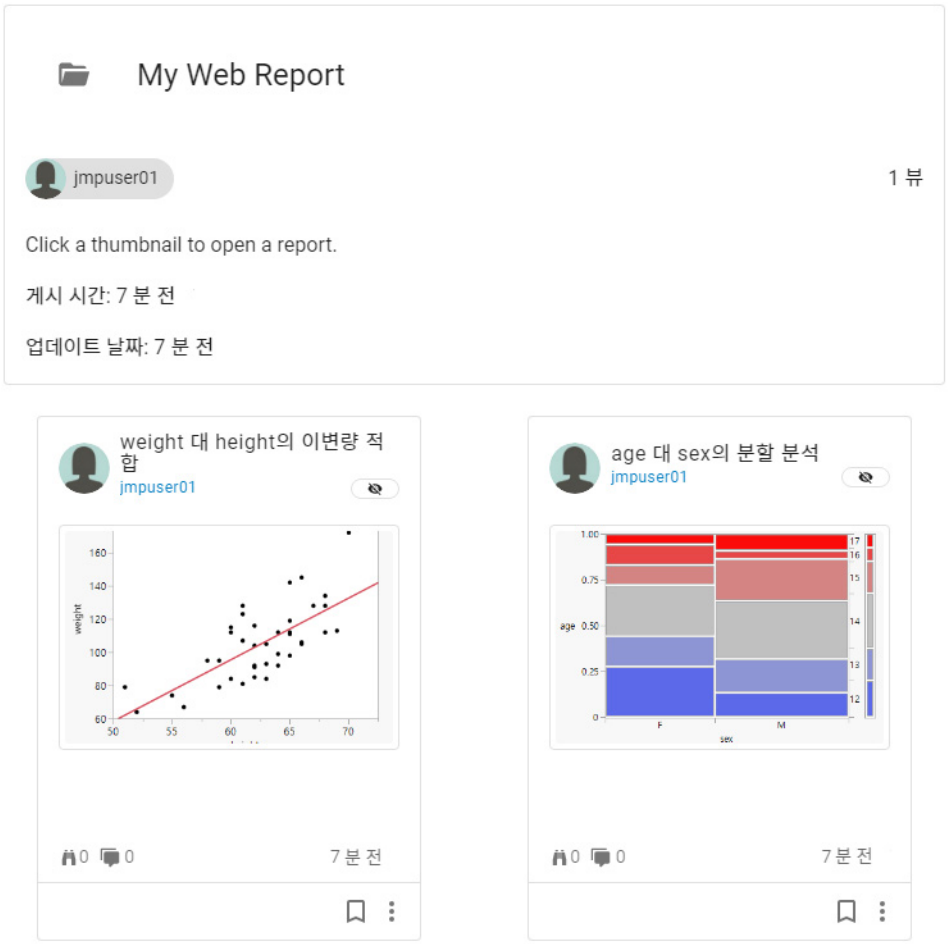
그림 10.5 여러 보고서가 있는 새 게시물 생성



참고 : 단일 보고서를 게시하는 경우 그림 10.5 의 창과 옵션이 약간 다릅니다 .

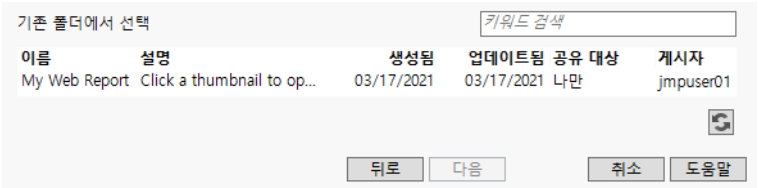
1. (다중 게시물에만 해당) 폴더 제목과 선택적 설명을 입력합니다 .
2. (선택 사항) 각 보고서의 제목을 업데이트하고 설명을 추가합니다 .
3. 게시물 공유 방법을 선택합니다 .
   - **나만**은 게시물을 비공개로 유지하고 나한테만 표시합니다 .
   - **모두**는 게시물을 JMP Public 또는 JMP Live 의 모든 사용자와 공유합니다 .
   - **그룹**은 게시물을 선택한 그룹과 공유합니다 . 게시할 수 있는 그룹이 이 옵션 아래에 표시됩니다 .
4. 추가 옵션을 변경합니다 . 자세한 내용은 " 게시물 옵션 "(547 페이지) 에서 확인하십시오 .
5. **게시**를 클릭합니다 .
6. **확인**을 클릭하여 정보 창을 종료합니다 .

그림 10.6 보고서가 포함된 JMP Live 폴더



기존 폴더에 보고서 추가

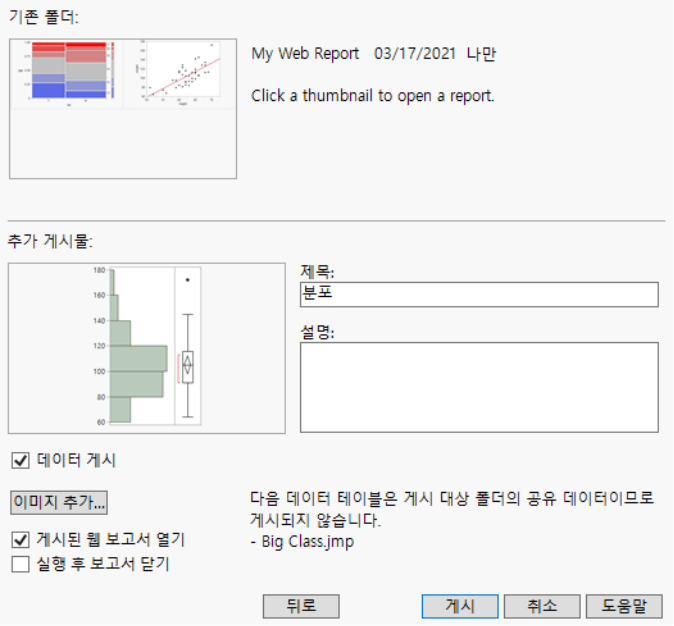
그림 10.7 기존 폴더 선택



참고 : 업데이트 권한이 있는 폴더만 표시됩니다 . 사용자가 소유한 폴더 또는 멤버인 그룹의 폴더입니다 .

1. 폴더를 선택합니다 . 다음 중 하나를 수행합니다 .
   - 목록에서 폴더를 클릭하고 **다음**을 클릭합니다 .
   - 폴더 검색 - " 키워드 검색 " 상자에 폴더 이름 또는 설명에 나타나는 단어를 입력한 다음 폴더를 클릭합니다 . **다음**을 클릭합니다 .

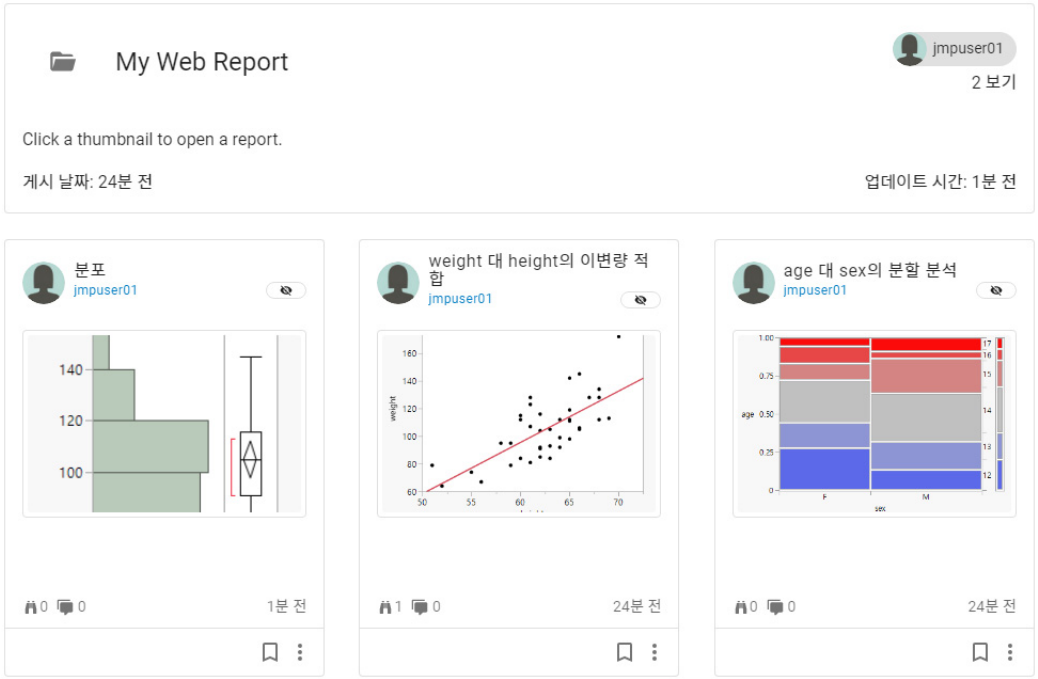
그림 10.8 기존 폴더에 보고서 추가



참고 : 다중 보고서를 게시하는 경우 그림 10.8 의 창과 옵션이 약간 다릅니다 .

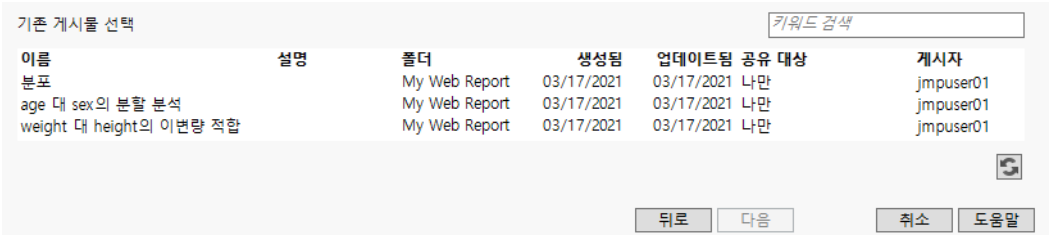
2. (선택 사항) 각 보고서의 제목을 업데이트하고 설명을 추가합니다 .
3. 추가 옵션을 변경합니다 . 자세한 내용은 " 게시물 옵션 "(547 페이지) 에서 확인하십시오 .
4. **게시**를 클릭합니다 .
5. **확인**을 클릭하여 정보 창을 종료합니다 .

그림 10.9 기존 JMP Live 폴더에 추가된 보고서



게시물의 기존 보고서 바꾸기

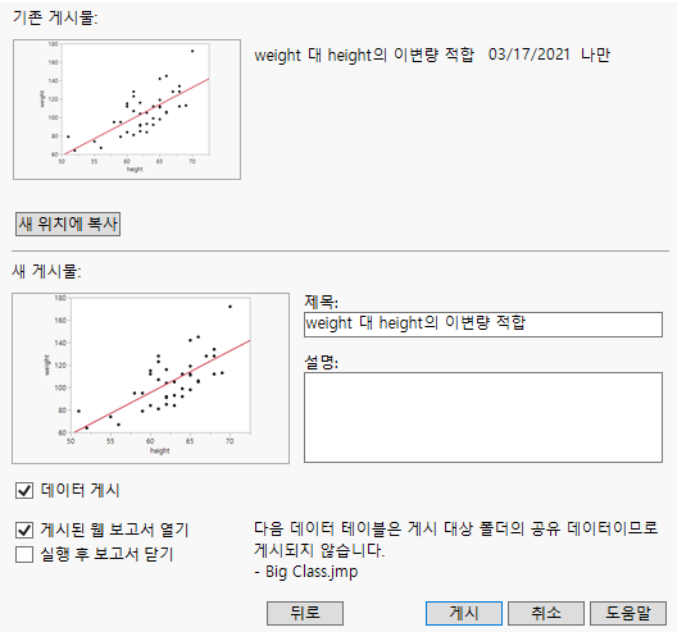
그림 10.10 기존 게시물 선택



참고 : 업데이트 권한이 있는 게시물만 표시됩니다 .

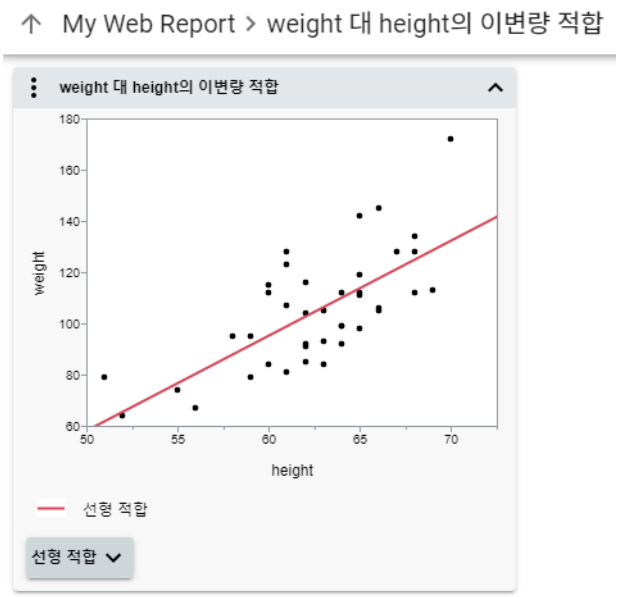
1. 게시물을 선택합니다 . 다음 중 하나를 수행합니다 .
   - 목록에서 게시물을 클릭하고 **다음**을 클릭합니다 .
   - 게시물 검색 - " 키워드 검색 " 상자에 게시물 이름 또는 설명에 나타나는 단어를 입력한 다음 게시물을 클릭합니다 . **다음**을 클릭합니다 .

그림 10.11 기존 게시물 바꾸기



2. 다음 옵션 중 하나를 선택합니다 .
   - 기존 게시물과 동일한 제목과 설명 (있는 경우) 을 사용하려면 **새 위치에 복사**를 클릭합니다 .
   - 업데이트된 제목과 설명을 사용하려면 새 제목과 설명을 입력하십시오 .
3. 추가 옵션을 변경합니다 . 자세한 내용은 " 게시물 옵션 "(547 페이지) 에서 확인하십시오 .
4. **게시**를 클릭합니다 .
5. **확인**을 클릭하여 정보 창을 종료합니다 .

그림 10.12 JMP Live 에서 바뀐 보고서



게시물 옵션

제목 웹 페이지의 제목을 추가합니다 . 보고서 제목을 변경할 수도 있습니다 .

설명 (선택 사항) 웹 페이지에 제목 아래에 나타나는 설명을 추가합니다 . 보고서 제목 아래에 나타나는 보고서 설명을 추가할 수도 있습니다 .

데이터 게시 대화식 보고서의 경우 이 옵션을 선택합니다 . 이 옵션을 선택 취소하면 정적 보고서가 됩니다 . 이 옵션을 선택하면 보고서의 배경 데이터와 보고서의 JSL 스크립트가 게시됩니다 .

> **참고 :** 중요한 데이터를 공유하지 않으려면 결과를 대화식이 아닌 웹 페이지로 저장하십시오 . (단일 보고서인 경우 **파일 > 내보내기 > HTML** 을 선택합니다 . 다중 보고서인 경우 **파일 > 게시 > 파일에 게시**를 선택하고 **데이터 게시**를 선택 취소합니다 .)

경고 사용 (관리도를 지원하는 플랫폼에만 적용 가능) 게시물에 관리도 경고가 있을 때 관련 당사자에게 알립니다 . 관련 당사자에는 게시물 게시자와 경고 알림이 활성화되어 있고 게시물이 공유되는 모든 그룹이 포함됩니다 .

이미지 추가 이미지를 이미지 전용 보고서로 추가합니다 .

데이터 및 스크립트 다운로드 허용 보고서와 관련된 데이터 테이블 및 스크립트를 다운로드하도록 허용합니다 . 이 옵션을 선택하지 않으면 JMP Live 에서 데이터 테이블 이름 옆에 다운로드 옵션이 나타나지 않고 보고서를 재생성하는 JSL 스크립트가 없습니다 . 자세한 내용은 " 데이터 접근성에 대한 참고 사항 "(548 페이지) 에서 확인하십시오 .

게시된 웹 보고서 열기 " 게시 " 를 클릭하면 브라우저에서 웹 페이지를 엽니다 .

실행 후 보고서 닫기 웹 보고서를 게시하면 JMP 에서 보고서를 닫습니다 .

삭제 아이콘 보고서를 삭제합니다.

화살표 아이콘 보고서 순서를 변경합니다.

JMP Live 및 JMP Public 도움말

JMP Live 또는 JMP Public 사용에 관한 도움말을 보려면 > **도움말**을 클릭합니다. 그러면 다음 도움말 페이지 중 하나가 열립니다.

- JMP Live: https://www.jmp.com/jmplive
- JMP Live(관리자): https://www.jmp.com/jmpliveadmin
- JMP Public: https://www.jmp.com/jmppublic

데이터 접근성에 대한 참고 사항

게시된 보고서에 표시되는 데이터는 HTML5 파일과 JMP 데이터 테이블 형식으로 온라인에 존재합니다.

"데이터 다운로드 허용" 옵션을 비활성화하면 JMP Live 페이지에 대한 액세스 권한이 있는 모든 사용자가 JMP 데이터 테이블 형식의 데이터를 다운로드할 수 없게 되지만 XML 형식의 데이터가 포함된 HTML 파일은 가져올 수 있습니다. "데이터 다운로드 허용"을 활성화하면 JMP Live 페이지에 액세스할 수 있는 사용자가 두 가지 방법 중 하나로 데이터를 가져올 수 있습니다. JMP 형식은 JMP Live 페이지에 액세스할 수 있으며 JMP 소프트웨어 또는 JMP 파일을 읽는 방법을 알고 있는 타사 소프트웨어에 대한 별도의 라이선스가 있는 사용자에게만 유용합니다. 하지만 JMP Live 페이지 사용자는 HTML 파일의 데이터를 볼 수 있는 웹 브라우저만 있으면 됩니다.

보고서를 PDF 파일로 저장

"파일" > "다른 이름으로 저장" > "PDF 파일 (*.pdf)"(Windows) 또는 "파일" > "내보내기" > "이미지" > "PDF"(macOS)를 선택하면 보고서를 PDF 파일로 저장할 수 있습니다.

Windows

- 파일을 내보낼 때 머리글, 바닥글 및 여백과 같은 페이지 설정 기능을 제어할 수 있습니다.
  - **파일 > 내보내기**를 선택하고 **PDF** 를 선택한 후 **다음**을 클릭합니다.
  - 파일 이름을 지정하고 **저장**을 클릭한 다음 " 페이지 설정 " 창에서 설정을 변경합니다.
- **파일 > 페이지 설정**을 선택하여 페이지 설정을 제어할 수도 있습니다. **기본값으로 설정**을 클릭하여 이후 페이지 설정에 현재 설정을 적용할 수 있습니다.
- **파일 > 인쇄 미리보기**를 선택하여 페이지 설정을 제어할 수도 있습니다.

macOS

- **파일 > 내보내기 > 이미지 > PDF** 를 선택합니다. 결과 PDF 파일에는 머리글과 바닥글이 없는 단일 PDF 이미지가 포함됩니다.

- **파일 > 인쇄**를 선택한 후 PDF 창에서 **PDF 로 저장**을 선택할 수도 있습니다 . 결과 PDF 파일은 필요한 수만큼의 여러 페이지로 생성되며 , 머리글 및 바닥글이 포함되고 " 인쇄 " 및 " 페이지 설정 " 창에서 선택한 페이지 크기 , 방향 , 배율 및 여백이 사용됩니다 .

Windows 및 macOS

JSL 을 통해 Windows 와 macOS 모두에서의 페이지 설정을 제어할 수 있습니다. 페이지 설정 지정, 인쇄 머리글 설정, 인쇄 바닥글 설정 및 PDF 저장에 대한 자세한 내용은 JSL Syntax Reference 에서 확인하십시오 .

보고서를 PowerPoint 프레젠테이션으로 저장

JMP 결과를 Microsoft PowerPoint 프레젠테이션 (.pptx) 으로 저장하여 프레젠테이션을 생성할 수 있습니다 . .pptx 파일로 저장한 후에는 PowerPoint 에서 JMP 내용을 재배열하고 텍스트를 편집합니다 . JMP 보고서를 PowerPoint 로 가져올 경우 다음과 같이 섹션이 달라집니다 . 보고서 머리글은 편집 가능한 텍스트 상자로 가져오고 , 그래프는 이미지로 가져오게 됩니다 . 범례 같은 일부 그래픽 요소는 별도의 이미지로 가져올 수 있습니다 . 이미지를 PowerPoint 에서 열면 슬라이드 크기에 맞게 이미지 크기가 조정됩니다 . 일반적으로 JMP 보고서의 각 개요는 개별 PowerPoint 슬라이드에 표시됩니다 . .pptx 파일로 저장할 때 보고서에서 닫혀 있던 개요는 PowerPoint 에서 제외됩니다 . 내용이 많으면 여러 개의 PowerPoint 슬라이드로 나뉘어 표시될 수 있습니다 . 필요하지 않은 내용은 PowerPoint 파일에서 삭제할 수 있습니다 .

참고 : Windows 에서 JMP 를 통해 생성한 .pptx 파일을 열려면 PowerPoint 2007 이상이 필요합니다 . macOS 에서는 PowerPoint 2011 이상이 필요합니다 .

1. JMP 에서 보고서를 생성합니다 .
2. **파일 > 내보내기 > Microsoft PowerPoint** 를 선택하고 **다음**을 클릭합니다 .
3. 그래픽 파일 형식을 선택합니다 .

   Windows 에서는 EMF 가 기본 형식입니다 . macOS 에서는 PDF 가 기본 형식입니다 .
4. (선택 사항) 저장 후 기본 프레젠테이션 프로그램에서 파일이 열리지 않도록 하려면 **저장 후 파일 열기**를 선택 취소합니다 .
5. 파일 이름을 지정하고 파일을 저장합니다 (macOS 에서는 내보냄).

팁 :

- Windows 에서 생성된 기본 EMF 그래픽은 macOS 에서 지원되지 않습니다 . macOS 에서 생성된 기본 PDF 그래픽은 Windows 에서 지원되지 않습니다 . 플랫폼 간 호환성을 위해서 **파일 > 환경 설정 > 타사 데이터** (Windows) 또는 **JMP > 환경 설정 > 타사 데이터** (macOS). 그런 다음 **PowerPoint 용 이미지 형식**을 PNG 또는 JPEG 로 변경하십시오 .
- 일부 경우에는 PowerPoint 를 사용하여 열 너비를 자동으로 조정해야 할 수 있습니다 . PowerPoint 에서 개체 크기가 JMP 와 다르게 조정되어 테이블 열의 크기가 비정상적으로 설정되는 경우도 있습니다 .

- 레이아웃을 개선하려면 PowerPoint 에서 열 구분선을 드래그하여 열 너비를 조정하십시오 . 열 테두리를 두 번 클릭하여 열의 공백을 줄을 수도 있습니다 .
- JMP 디렉터리의 pptx 폴더에는 기본 PowerPoint 템플릿이 설치되어 있습니다 . 내보낼 때 항상 대체 템플릿을 사용하려는 경우 이 PowerPoint 템플릿의 복사본을 저장하고 기본 템플릿을 원하는 템플릿으로 바꿀 수 있습니다 . 예를 들어 독일어 기본 글꼴이 포함된 독일어 PowerPoint 템플릿을 생성할 수 있습니다 . 자세한 내용은 사용자 PowerPoint 템플릿에서 확인하십시오 .

사용자 PowerPoint 템플릿

PowerPoint 프레젠테이션에서는 기본 JMP 템플릿을 사용합니다 . 기본 템플릿 파일을 덮어써서 사용자 템플릿을 사용할 수 있습니다 . 이렇게 하려면 다음 위치 중 하나에 있는 JMPExportTemplate.pptx 를 엽니다 .

C:/Program Files/SAS/JMP/16/pptx/

C:/Program Files/SAS/JMPPro/16/pptx/

그런 다음 사용할 PowerPoint 템플릿을 선택하고 **파일 > 저장**을 선택합니다 . 다음에 JMP 내용을 PowerPoint 로 저장할 때는 파일이 이 사용자 템플릿을 사용하여 열립니다 . 또는 , 먼저 PowerPoint 에서 사용자 템플릿을 생성하고 이를 JMPExportTemplate.pptx 로 저장해도 됩니다 .

참고 :

- macOS 의 사용자 정의 PowerPoint 템플릿은 편집할 수 없습니다 .
- 나중에 JMP 기본 템플릿을 사용하려면 파일을 덮어쓰기 전에 컴퓨터에 복사본을 만들어 두십시오 .

JSL 스크립트와 함께 사용자 템플릿을 사용할 수도 있습니다 . 자세한 내용은 JSL Syntax Reference 의 에서 확인하십시오 .

내보낸 그래픽에 대한 그래픽 DPI 설정

Windows 에서는 몇 가지 방법으로 내보낸 그래픽의 해상도 (또는 DPI) 를 설정할 수 있습니다 .

- 모든 그래픽의 기본 DPI 를 변경하려면 "Windows 관련 " 환경 설정에서 DPI 를 설정합니다 . 기본 옵션은 운영 체제의 기본 설정을 사용합니다 .
- 다음 스크립트를 사용하면 기본 DPI 를 다른 값으로 설정할 수 있습니다 . 다른 DPI 를 원하는 경우 숫자를 변경한 후 스크립트를 실행합니다 .

  ```
  Pref( Save Image DPI( 600 ) );
  ```

  위의 스크립트를 실행한 후 기본 DPI 를 확인하려면 다음 스크립트를 실행합니다 .

  ```
  Get Preferences( Save Image DPI );
  ```

  이 스크립트를 실행한 후에는 "Windows 관련 " 환경 설정과 " 다른 이름으로 보고서 저장 " 창에서 이 DPI 가 기본값으로 표시됩니다 . 이 설정은 다시 변경하기 전까지 적용됩니다 . 선

택한 설정을 여러 JMP 세션에 저장하려면 " 다른 이름으로 보고서 저장 " 창에서 **항상 이 설정 사용**을 선택합니다 .

- 특정 그래픽의 경우 , " 다른 이름으로 보고서 저장 " 창에서 DPI 를 설정합니다 .

macOS에서는 기본 DPI인 72를 변경할 수 없습니다.

JMP 분석을 스크립트로 저장

분석을 완료하고 보고서를 받은 후 해당 보고서를 생성하기까지 사용한 프로세스를 JSL 스크립트로 저장할 수 있습니다. 데이터 테이블 내에, 보고서 내에, 또는 별도의 스크립트 창에 스크립트를 저장할 수 있습니다. 이렇게 하려면 보고서 제목에서 빨간색 삼각형을 클릭하고 "스크립트 저장"에서 원하는 옵션을 선택합니다. 자세한 내용은 "보고서의 스크립트 저장 메뉴"(416 페이지)에서 확인하십시오.

영어 이외 버전의 JMP에서는 스크립트를 영어 또는 로컬 언어로 저장할 수 있습니다. 자세한 내용은 "JMP 환경 설정" 장의 "일반 JMP 환경 설정"(621 페이지)에서 확인하십시오.

보고서 요소를 그래픽 형식으로 저장

분석 보고서 창의 요소를 그래픽으로 저장할 수 있습니다. 선택 항목을 .png, .jpg, .gif, .eps, .emf, .pdf, .tiff 및 .svg 형식으로 저장할 수 있습니다. macOS에서는 선택한 요소를 .png, .jpg, .gif, .svg, .eps, .pdf 및 .pptx 형식으로 저장할 수 있습니다.

선택한 요소를 저장하려면 다음 단계를 수행하십시오 .

1. 선택 도구 ( ) 를 클릭합니다 .
2. 저장할 영역을 강조 표시합니다 .
3. Windows 의 경우 , **편집 > 다른 이름으로 선택 항목 저장**을 선택합니다 . MacOS 에서는 Ctrl 키를 누르고 **파일 > 다른 이름으로 선택 항목 저장**을 누릅니다 .

   " 다른 이름으로 선택 항목 저장 " 창 (macOS 의 경우 " 저장 " 창) 이 나타납니다 .
4. 그래픽의 파일 이름을 입력합니다 .

   macOS 의 경우 , 파일을 기본 형식인 .png 이외의 다른 형식으로 저장하려면 파일 확장자를 변경합니다 .

   참고 : " 다른 이름으로 선택 항목 저장 " 을 사용하여 선택 항목에서 PDF 파일을 생성하는 경우에는 PDF 파일에 머리글과 바닥글이 포함됩니다 . JSL 을 사용하여 Windows 와 macOS 모두에서의 페이지 설정을 제어할 수 있습니다 . 자세한 내용은 JSL Syntax Reference 의 에서 확인하십시오 .
5. Windows 에서 옵션을 변경합니다 .

- 선택 항목을 저장할 그래픽 파일 형식 유형을 선택합니다 . 자세한 내용은 " 보고서 저장 및 공유 "(533 페이지) 에서 확인하십시오 .
- " 이미지 DPI 설정 " 을 선택합니다 . 늘리거나 인쇄해야 하는 그래픽에는 300 이상이 적합합니다 .
- 현재 세션에서 선택한 DPI 설정을 유지하려면 **항상 이 설정 사용**을 선택합니다 .
- 저장 후 기본 그래픽 프로그램에서 파일이 열리지 않도록 하려면 **저장 후 파일 열기**를 선택 취소합니다 .
- 현재 세션에서 선택한 파일 유형을 저장하려면 **다음에 이 창이 호출될 때 이 필터 선택**을 선택합니다 .

참고 : " 다음에 이 창이 호출될 때 이 필터 선택 " 옵션이 표시되지 않으면 그래픽 파일을 저장할 폴더 위치를 선택하십시오 .

6. **저장**을 클릭합니다 .

팁 :

- "Windows 관련 " 환경 설정에서 기본적으로 선택되는 DPI 설정을 96DPI 에서 300DPI 로 변경할 수 있습니다 . 다음 스크립트를 사용하면 기본 DPI 를 다른 값으로 설정할 수 있습니다 . 다른 DPI 를 원하는 경우 숫자를 변경한 후 스크립트를 실행합니다 .

```
Pref( Save Image DPI( 600 ) );
```

- macOS 에서는 기본 DPI 인 72 를 변경할 수 없습니다 .
- 비트맵 이미지가 포함된 그래프 또는 보고서를 .svg 파일로 내보내면 이미지가 .svg 파일에 포함됩니다 . 그러면 내보낸 내용을 다른 문서에서 쉽게 재사용할 수 있습니다 .

보고서 인쇄

활성 창의 보고서를 인쇄하려면 **파일 > 인쇄**를 선택합니다. 이 명령은 표준 인쇄 창을 표시합니다. 이 창은 운영 체제와 프린터 드라이버에 따라 다르게 나타납니다.

인쇄용 페이지 구분을 삽입하려면 다음 단계를 수행하십시오 .

1. 보고서 창에서 표시 아이콘 을 마우스 오른쪽 버튼으로 클릭합니다 .
2. **편집 > 페이지 구분**을 선택합니다 .

보고서 복사 및 붙여넣기

JMP 보고서 또는 데이터 테이블을 다른 프로그램에서 사용해야 하는 경우 보고서 또는 테이블의 일부분을 다른 프로그램에 붙여넣거나 드래그할 수 있습니다. 예를 들어 동료에게 JMP 보고서의 복사본을 Microsoft Word 파일로 보낼 수 있습니다.

1. 선택 도구 를 클릭합니다.
2. 보고서 창 또는 데이터 테이블에서 원하는 항목을 클릭 후 드래그하거나 Shift 키를 누른 채 클릭하여 선택합니다. 보고서 창의 모서리 부근을 클릭하면 보고서 전체가 선택됩니다.
3. 선택한 항목을 클릭하고 JMP 에서 다른 프로그램으로 드래그합니다. 또는 JMP 에서 선택한 항목을 복사한 후 다른 프로그램에 붙여넣습니다. 요소를 다른 응용 프로그램에 붙여넣을 때 사용되는 형식은 붙여넣는 대상 응용 프로그램에 따라 달라집니다. 응용 프로그램의 선택하여 붙여넣기 명령을 사용하는 경우에는 텍스트 (.rtf), 형식 없는 텍스트 (.txt) 또는 비트맵 (.bmp) 중에서 사용할 형식을 선택할 수 있습니다.

참고 :

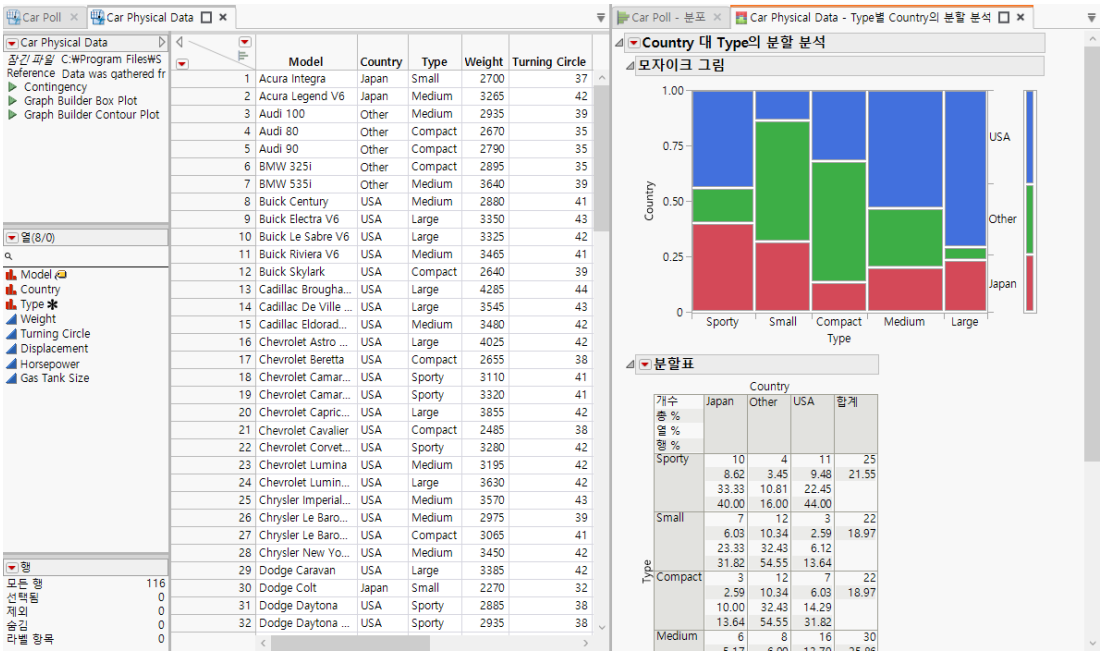
- 활성 보고서 창의 모든 텍스트 (그래프 제외) 를 형식 없는 텍스트로 복사하려면 **편집 > 텍스트로 복사**를 선택하십시오. macOS 의 경우 , Microsoft Word 에서 **붙여넣기 > 선택하여 붙여넣기**를 선택하거나 다른 macOS 응용 프로그램에서 **Paste and Match Style** 을 선택합니다.
- 라벨과 축이 포함된 그래프를 복사하려면 그래프를 마우스 오른쪽 버튼으로 클릭하고 **편집 > 그래프 복사**를 선택하십시오.
- 그래프 외의 요소를 복사하려면 마우스 오른쪽 버튼을 클릭하고 **편집 > 그림 복사**를 선택하십시오. 예를 들어 보고서를 복사하려면 보고서 제목이 포함된 막대를 마우스 오른쪽 버튼으로 클릭하고 옵션을 선택하십시오.
- 텍스트 상자의 텍스트를 클립보드에 복사하려면 텍스트를 마우스 오른쪽 버튼으로 클릭하고 **텍스트 복사**를 선택하십시오.
- 한 열의 텍스트를 클립보드에 복사하려면 열을 마우스 오른쪽 버튼으로 클릭하고 **열 복사**를 선택하십시오.
- 전체 테이블을 클립보드에 복사하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **테이블 복사**를 선택하십시오.
- " 보고서 " 환경 설정의 **보고서 PNG 이미지에 투명 배경 사용**은 PNG 이미지로 저장된 보고서에 투명 배경을 사용할지 여부를 결정합니다. 이 환경 설정은 기본적으로 선택 취소되어 있습니다.

프로젝트 작업

JMP 프로젝트를 사용하여 다음을 수행할 수 있습니다.

- 탭 형식의 단일 JMP 창에서 데이터를 더 효율적으로 탐색합니다.
- 관련 파일 및 보고서 집합을 빠르게 저장하고 다시 엽니다.
- 자체 포함 프로젝트 파일에 테이블과 스크립트를 포함하여 작업을 쉽게 공유합니다.

그림 10.13 데이터 테이블 및 보고서가 포함된 프로젝트 파일



이 섹션에는 다음과 같은 정보가 포함되어 있습니다.

- "새 프로젝트 생성"
- "프로젝트에서 파일 열기"
- "프로젝트에서 파일 재배열"
- "프로젝트 저장"
- "프로젝트 작업 영역"
- "프로젝트 책갈피"
- "프로젝트 내용"
- "프로젝트의 최근 사용한 파일"
- "프로젝트 로그"

- "파일을 프로젝트로 이동"
- "프로젝트 내의 열기 스크립트 작성"
- "새 프로젝트 생성의 예"

새 프로젝트 생성

비어 있는 새 JMP 프로젝트를 생성하려면 **파일** > **새로 만들기** > **프로젝트**(Windows) 또는 **파일** > **새로 만들기** > **새 프로젝트**(macOS)를 선택합니다.

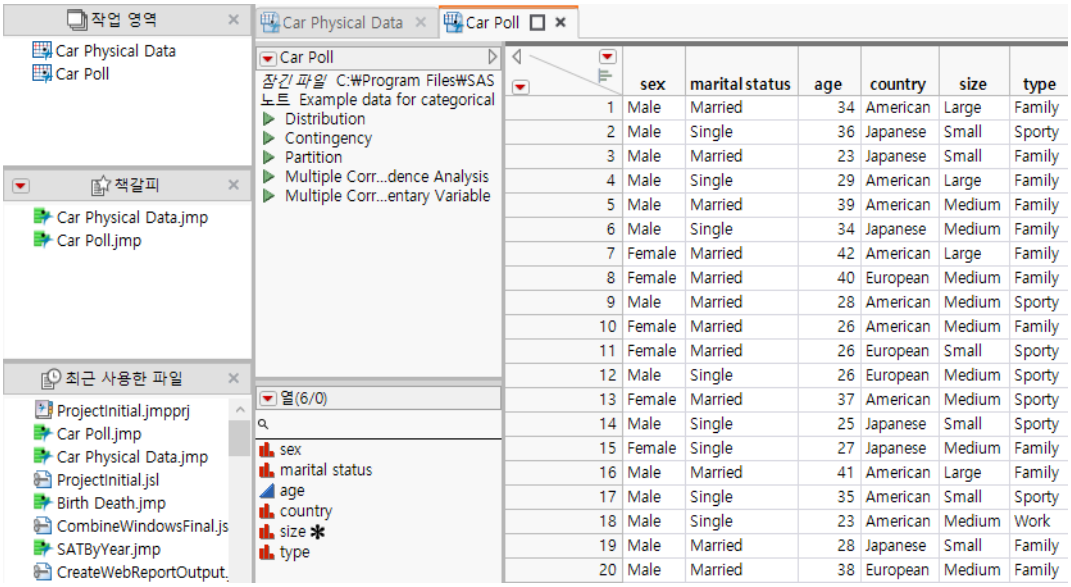
프로젝트에서 파일 열기

JMP 프로젝트에서 데이터 테이블을 열고 데이터에 대한 분석을 실행할 수 있습니다. 각 데이터 테이블과 분석 보고서는 새 탭에서 열립니다.

1. 프로젝트 창에서 **파일** > **열기**를 선택하고 열려는 데이터 테이블로 이동합니다.
2. 데이터 테이블에서 분석을 실행합니다.

컴퓨터에서 데이터 테이블이 업데이트된 경우 프로젝트를 다시 열면 프로젝트의 모든 관련 보고서가 업데이트됩니다.

그림 10.14 초기 프로젝트



팁 : 프로젝트에서 Ctrl+S 또는 Ctrl+W 같은 JMP 바로 가기 키를 사용하여 파일을 저장하거나 활성 창 페인을 닫을 수 있습니다. 전체 목록을 보려면 **도움말** > **빠른 참조 카드**를 선택하십시오.

프로젝트에서 파일 재배열

JMP 프로젝트에서 데이터 테이블과 보고서는 개별 탭에 나타납니다. 탭을 도킹 영역으로 드래그하여 재배열할 수 있습니다.

팁 : 도킹된 항목을 실행 취소하거나 다시 실행하려면 **프로젝트** > **레이아웃 실행 취소** 또는 **프로젝트** > **레이아웃 다시 실행**을 선택합니다 . 모든 테이블과 보고서를 개별 탭으로 되돌리려면 **프로젝트** > **레이아웃 재설정**을 선택합니다 .

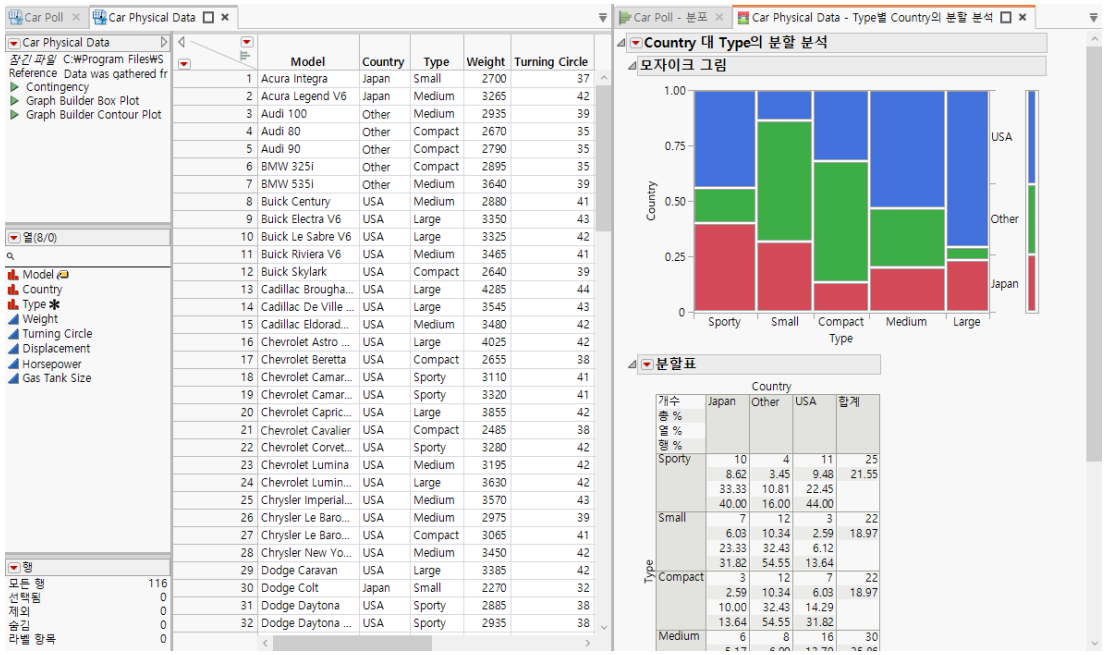
프로젝트 파일 재배열의 예

1. **파일** > **새로 만들기** > **프로젝트** (Windows) 또는 **파일** > **새로 만들기** > **새 프로젝트** (macOS) 를 선택합니다 .
2. **도움말** > **샘플 데이터 라이브러리**를 선택하고 Car Physical Data.jmp 및 Car Poll.jmp를 엽니다.
3. Car Poll.jmp 데이터 테이블에서 Distribution 스크립트를 실행합니다 .
4. Car Physical Data.jmp 데이터 테이블에서 Contingency 스크립트를 실행합니다 .
5. **Car Poll - 분포** 보고서 탭을 오른쪽으로 드래그하여 **오른쪽 도킹** 영역에 놓습니다 .

   프로젝트 창 오른쪽에 " 분포 " 보고서 페인이 나타납니다 . 보고서가 도킹되어 탭 간에 전환할 때 계속 표시됩니다 .
6. **Car Physical Data - 분할 분석** 탭을 **Car Poll - Distribution** 보고서의 가운데로 드래그하여 **탭 도킹** 영역에 놓습니다 .

   팁 : 데이터 테이블 창의 크기를 조정하여 두 보고서를 모두 표시할 수 있습니다 .

그림 10.15 탭으로 그룹화된 보고서 및 데이터 테이블



두 데이터 테이블과 마찬가지로 "분포" 보고서와 "분할 분석" 보고서가 탭 형식으로 함께 표시됩니다.

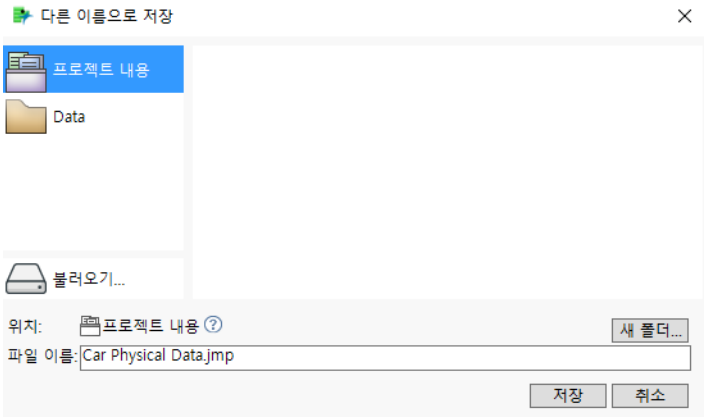
프로젝트 저장

프로젝트를 공유, 배포 또는 보관하려면 모든 데이터 테이블과 스크립트를 프로젝트 내용에 저장하여 자체 포함 프로젝트를 생성합니다. 이렇게 하면 모든 프로젝트 파일과 폴더를 단일 프로젝트 파일에 포함하여 다른 JMP 사용자와 공유할 수 있습니다.

1. (선택 사항) 프로젝트를 자체 포함 프로젝트로 저장하려면 **파일** > **다른 이름으로 저장**을 선택한 후 **프로젝트 내용**을 선택하여 각 데이터 테이블과 스크립트를 프로젝트에 저장합니다. **새 폴더**를 클릭하여 프로젝트 내용에 폴더를 생성할 수 있습니다.

참고 : 보고서는 프로젝트에 자동 저장되므로 저장할 필요가 없습니다.

그림 10.16 프로젝트 내용에 데이터 테이블 저장



2. **파일** > **프로젝트 저장**을 선택하여 프로젝트 파일을 저장합니다 .

그림 10.17 프로젝트 파일 저장

프로젝트 작업 영역

JMP 프로젝트에서 현재 열려 있는 모든 파일을 " 작업 영역 " 도구 페인에서 볼 수 있습니다. 보고서는 해당 데이터 테이블 아래에 나타납니다. 활성 데이터 테이블은 굵게 표시됩니다.

팁 : " 작업 영역 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **작업 영역 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

" 작업 영역 " 도구 페인에서 다음을 수행할 수 있습니다.

- 파일을 두 번 클릭하여 활성화합니다 .
- 파일을 마우스 오른쪽 버튼으로 클릭하여 파일 닫기 , 숨기기 (탭을 제거하고 파일 이름을 흐리게 표시), 책갈피 지정 등을 수행합니다 .

프로젝트 책갈피

JMP 프로젝트에서 파일 또는 폴더의 바로 가기를 " 책갈피 " 도구 페인에 생성할 수 있습니다. 책갈피 그룹을 생성하여 책갈피 지정된 파일과 폴더를 구성할 수도 있습니다.

팁 : " 책갈피 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **책갈피 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

프로젝트에서 열린 파일 책갈피 지정

- 탭을 마우스 오른쪽 버튼으로 클릭하고 **책갈피**를 선택합니다 .
- 열려 있는 모든 파일을 책갈피 지정하려면 탭을 마우스 오른쪽 버튼으로 클릭하고 **모두 책갈피**를 선택합니다 .

컴퓨터의 파일 또는 폴더 책갈피 지정

컴퓨터의 파일 또는 폴더를 "책갈피" 도구 페인으로 드래그하거나, "책갈피"의 빨간색 삼각형 메뉴를 클릭하고 **파일 추가** 또는 **폴더 추가**를 선택합니다.

참고 : 컴퓨터의 책갈피 지정된 폴더에서 파일을 추가하거나 제거하면 폴더 내용이 프로젝트에서 자동으로 업데이트됩니다 .

책갈피 지정된 파일 열기

"책갈피" 도구 페인에서 파일을 두 번 클릭합니다.

책갈피 제거

"책갈피" 도구 페인에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **책갈피 제거**를 선택합니다.

책갈피 그룹 생성

1. " 책갈피 " 의 빨간색 삼각형 메뉴를 클릭하고 **새 그룹**을 선택합니다 .
2. 폴더 이름을 지정하고 **확인**을 클릭합니다 .
3. 새 책갈피 또는 기존 책갈피를 그룹으로 드래그하거나 , " 책갈피 " 도구 페인에서 그룹을 마우스 오른쪽 버튼으로 클릭하고 **파일 추가** 또는 **폴더 추가**를 선택합니다 .

프로젝트 내용

자체 포함 JMP 프로젝트에서 프로젝트 내용에 저장된 모든 파일이 "내용" 도구 페인에 나타납니다.

팁 : " 내용 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **내용 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

프로젝트 내용에 있는 파일 열기

"내용" 도구 페인에서 파일을 두 번 클릭합니다.

프로젝트 내용에 폴더 생성

1. " 내용 " 의 빨간색 삼각형 메뉴를 클릭하고 **새 폴더**를 선택합니다 .
2. 폴더 이름을 지정하고 **확인**을 클릭합니다 .

파일을 폴더로 이동

"내용" 도구 페인에서 파일을 폴더로 드래그합니다.

컴퓨터의 파일 또는 폴더를 프로젝트 내용에 복사

1. " 내용 " 의 빨간색 삼각형 메뉴를 클릭하고 **프로젝트에 파일 복사** 또는 **프로젝트에 폴더 복사**를 선택합니다 .
2. 파일을 탐색한 후 **열기**를 클릭하거나 폴더를 탐색한 후 **선택**을 클릭합니다 .

프로젝트 내용에 있는 파일 이름 바꾸기

"내용" 도구 페인에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **이름 바꾸기**를 선택합니다.

프로젝트 내용에 있는 파일 삭제

"내용" 도구 페인에서 파일을 마우스 오른쪽 버튼으로 클릭하고 **삭제**를 선택합니다.

프로젝트의 최근 사용한 파일

JMP 프로젝트의 "최근 사용한 파일" 도구 페인에서 파일을 열 수 있습니다.

팁 : " 최근 사용한 파일 " 도구 페인을 표시하거나 숨기려면 **프로젝트** > **최근 사용한 파일 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일** > **환경 설정** > **프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .

파일 열기

"최근 사용한 파일" 도구 페인에서 파일을 두 번 클릭하거나 프로젝트로 드래그합니다.

참고: 프로젝트 내용에 저장된 열린 파일은 "최근 사용한 파일" 도구 페인에 표시되지 않습니다.

프로젝트 로그

JMP 프로젝트의 "로그" 도구 페인에서 로그 메시지를 볼 수 있습니다.

팁 :

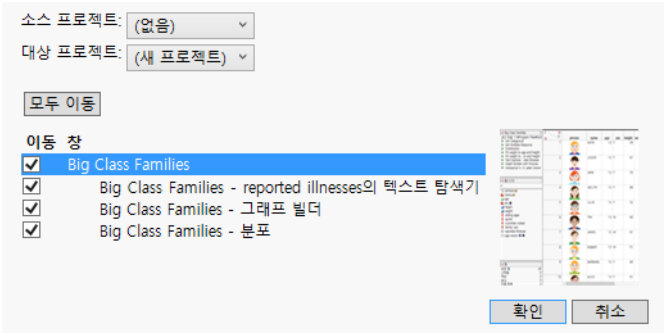
- " 로그 " 도구 페인을 표시하거나 숨기려면 **프로젝트 > 로그 표시**를 선택합니다 . 프로젝트를 생성할 때 기본적으로 표시되는 도구 페인을 지정하려면 **파일 > 환경 설정 > 프로젝트**를 선택하고 초기 도구 페인을 선택합니다 .
- 기본적으로 프로젝트에서 발생하는 로그 메시지는 기본 JMP 로그가 아니라 프로젝트 로그에만 나타납니다 . 이 설정을 변경하려면 **파일 > 환경 설정 > 프로젝트**를 선택하고 **프로젝트 로그 페인 사용**을 **열려 있는 경우** 또는 **안 함**으로 업데이트합니다 .

파일을 프로젝트로 이동

JMP 에서는 파일을 프로젝트로 이동하거나 프로젝트 외부 또는 프로젝트 간에 이동할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class Families.jmp 를 엽니다 .
2. "Distribution", "Text Explorer" 및 "Graph Builder" 스크립트를 실행합니다 .
3. 어느 창에서든 **창 > 프로젝트 간 이동**을 선택합니다 .
4. 프로젝트에 열려 있지 않은 파일을 보려면 " 소스 프로젝트 " 를 **(없음)** 으로 그대로 둡니다 .
5. 선택한 파일을 새로 생성한 프로젝트로 이동하려면 " 대상 프로젝트 " 를 **(새 프로젝트)** 로 그대로 둡니다 .
6. Big Class Families.jmp 옆의 체크박스를 선택합니다 . 데이터 테이블과 연결된 그래프도 선택됩니다 .

그림 10.18 프로젝트 간 창 이동



7. **확인**을 클릭합니다 . 선택한 테이블과 보고서로 새 프로젝트가 생성됩니다 .

프로젝트 내의 열기 스크립트 작성

" 프로젝트 내의 열기 스크립트 " 상자에는 이 특정 프로젝트를 열 때마다 실행할 JSL 스크립트를 추가할 수 있습니다 . 예를 들어 보고서를 생성 또는 수정하고 , 사용자에게 정보를 요청하거나 , 사용 정보를 로그 창에 쓰는 스크립트를 생성할 수 있습니다 .

1. **프로젝트** > **프로젝트 설정**을 선택합니다 .
2. JSL 스크립트를 붙여 넣고 **확인**을 클릭합니다 .

팁 :

- 자세한 내용은 Scripting Guide 의 확인하십시오 . 여기에서는 특정 프로젝트 대신 모든 프로젝트를 열 때 시작 스크립트를 생성하는 방법을 설명합니다 .
- 기존 프로젝트를 새 프로젝트의 템플릿으로 사용하려면 **파일** > **환경 설정** > **프로젝트**에서 기존 프로젝트를 새 프로젝트 템플릿으로 지정합니다 .

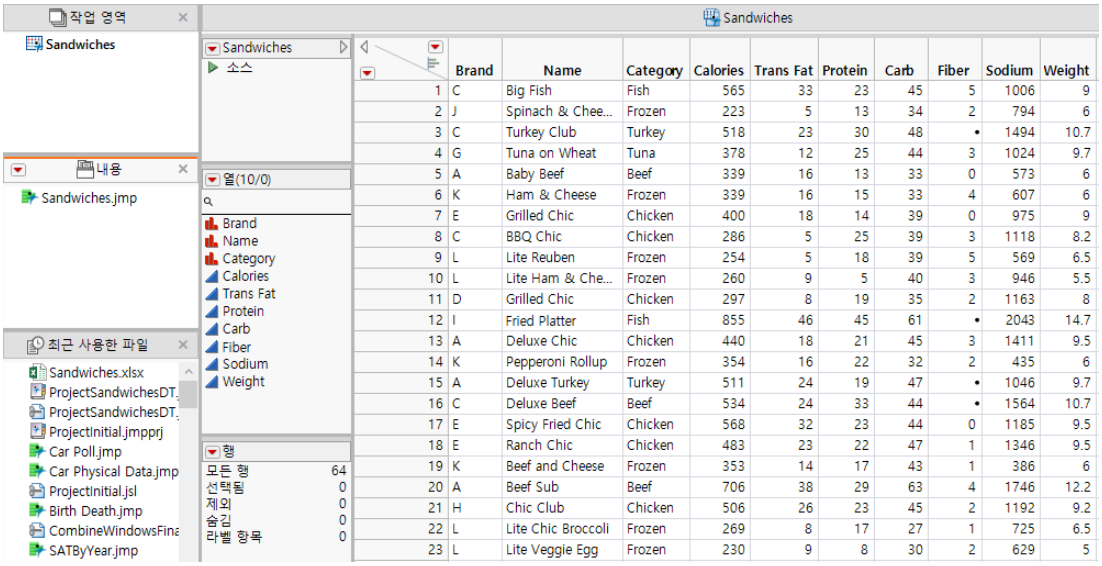
새 프로젝트 생성의 예

이 예에서는 프로젝트 생성, 데이터 가져오기, 프로젝트 창에서 분석 생성 및 보고서 도킹, 부분 집합 테이블 생성을 차례로 수행한 후 프로젝트를 저장하고, 닫았다가 다시 엽니다.

1. **파일** > **새로 만들기** > **프로젝트** (Windows) 또는 **파일** > **새로 만들기** > **새 프로젝트** (macOS) 를 선택합니다 .
2. 프로젝트 창에서 **파일** > **열기**를 선택합니다 .
3. sandwiches.xlsx 파일을 엽니다 . 이 파일의 기본 위치는 다음과 같습니다 .
   C:/Program Files/SAS/JMP/16/Samples/Import Data

   팁 : 하단에서 **모든 JMP 파일**을 **Excel 파일**로 변경해야 할 수도 있습니다 .
4. **가져오기**를 클릭합니다 .
5. **파일** > **저장**을 선택합니다 .
6. **프로젝트 내용**이 선택되어 있는지 확인합니다 .
7. 파일 이름을 Sandwiches.jmp 로 변경하고 **저장**을 클릭합니다 .

그림 10.19 Sandwiches 데이터를 가져온 프로젝트

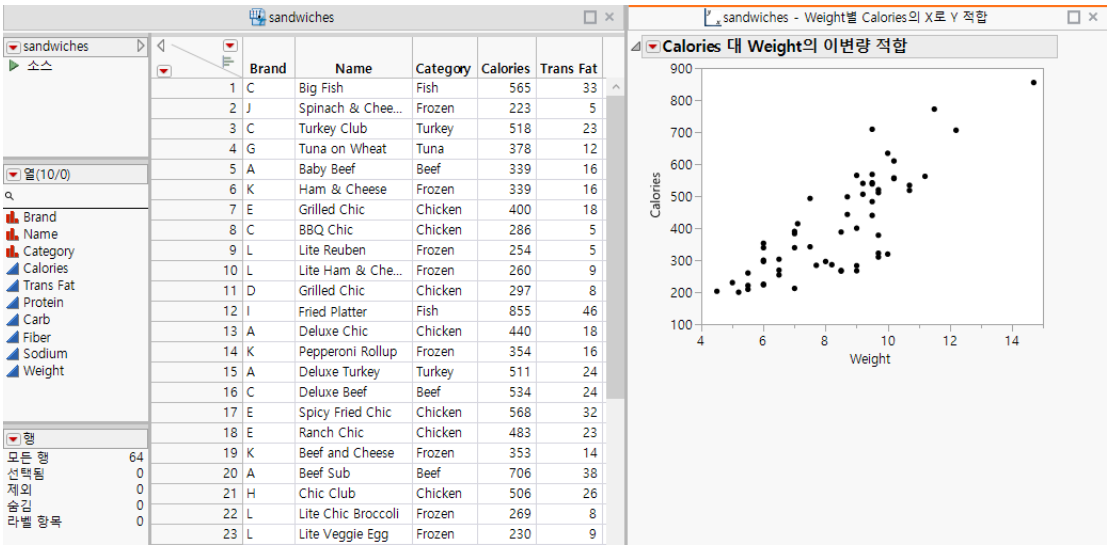


	Brand	Name	Category	Calories	Trans Fat	Protein	Carb	Fiber	Sodium	Weight
1	C	Big Fish	Fish	565	33	23	45	5	1006	9
2	J	Spinach & Chee...	Frozen	223	5	13	34	2	794	6
3	C	Turkey Club	Turkey	518	23	30	48	•	1494	10.7
4	G	Tuna on Wheat	Tuna	378	12	25	44	3	1024	9.7
5	A	Baby Beef	Beef	339	16	13	33	0	573	6
6	K	Ham & Cheese	Frozen	339	16	15	33	4	607	6
7	E	Grilled Chic	Chicken	400	18	14	39	0	975	9
8	C	BBQ Chic	Chicken	286	5	25	39	3	1118	8.2
9	L	Lite Reuben	Frozen	254	5	18	39	5	569	6.5
10	L	Lite Ham & Che...	Frozen	260	9	5	40	3	946	5.5
11	D	Grilled Chic	Chicken	297	8	19	35	2	1163	8
12	I	Fried Platter	Fish	855	46	45	61	•	2043	14.7
13	A	Deluxe Chic	Chicken	440	18	21	45	3	1411	9.5
14	K	Pepperoni Rollup	Frozen	354	16	22	32	2	435	6
15	A	Deluxe Turkey	Turkey	511	24	19	47	•	1046	9.7
16	C	Deluxe Beef	Beef	534	24	33	44	•	1564	10.7
17	E	Spicy Fried Chic	Chicken	568	32	23	44	0	1185	9.5
18	E	Ranch Chic	Chicken	483	23	22	47	1	1346	9.5
19	K	Beef and Cheese	Frozen	353	14	17	43	1	386	6
20	A	Beef Sub	Beef	706	38	29	63	4	1746	12.2
21	H	Chic Club	Chicken	506	26	23	45	2	1192	9.2
22	L	Lite Chic Broccoli	Frozen	269	8	17	27	1	725	6.5
23	L	Lite Veggie Egg	Frozen	230	9	8	30	2	629	5

8. **분석** > **X 로 Y 적합**을 선택합니다 .
9. Calories 를 선택하고 **Y, 반응**을 클릭합니다 .
10. Weight 를 선택하고 **X, 요인**을 클릭합니다 .
11. **확인**을 클릭합니다 .
12. **Sandwiches - X 로 Y 적합** 탭을 오른쪽으로 드래그하여 오른쪽 도킹 영역에 놓습니다 .

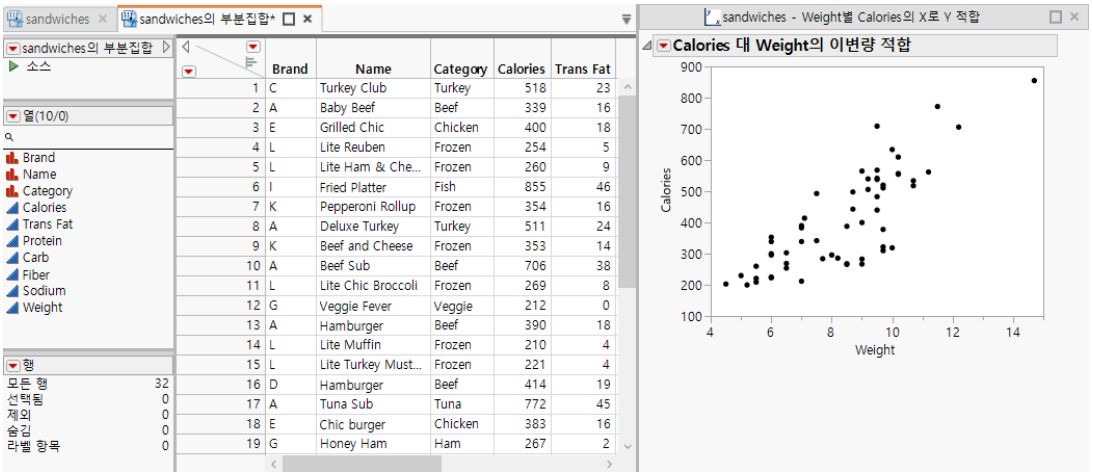
팁 : 전체 보고서를 표시하려면 데이터 테이블과 보고서 사이의 선을 왼쪽으로 드래그합니다 .

그림 10.20 오른쪽에 도킹된 X 로 Y 적합 (이변량) 보고서



13. **테이블** > **부분집합**을 선택합니다 .
14. " 행 " 아래에서 **랜덤 - 표집 비율 : 0.5** 를 선택합니다 .
15. **확인**을 클릭합니다 .

그림 10.21 저장되지 않은 부분집합 테이블이 있는 프로젝트



16. **파일** > **프로젝트 저장**을 선택합니다 .
17. 프로젝트를 저장할 폴더로 이동하여 프로젝트 파일 이름을 지정하고 **저장**을 클릭합니다 .
18. 프로젝트를 닫습니다 .
19. **파일** > **열기**를 선택하고 프로젝트 파일을 엽니다 .

저장하지 않은 부분집합 테이블이 프로젝트 파일에 자동으로 저장되어 있습니다.

저널 생성

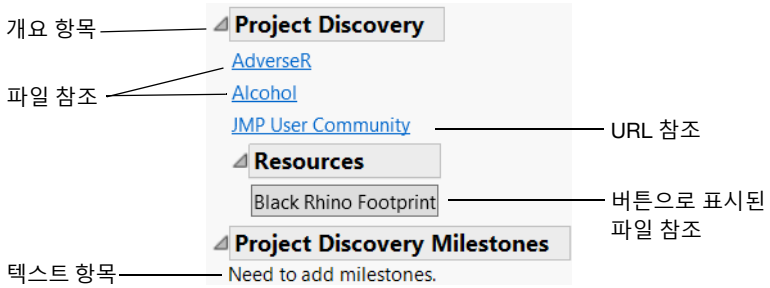
결과를 저장한 후 다른 사용자에게 제공하려는 경우 JMP 저널을 생성할 수 있습니다. 저널 생성의 이점 중 한 가지는 외부 파일에 의존하지 않고 분석을 캡처할 수 있다는 점입니다. 또 다른 이점은 특정 보고서 또는 그래프를 저널에 복사하여 붙여넣을 수 있다는 점입니다. 이 기능을 사용하면 사용자 보고서를 생성하여 다른 사용자에게 제공할 수 있습니다.

저널을 사용하여 생성할 수 있는 프레젠테이션에는 다음과 같은 두 가지 유형이 있습니다.

- 정적 프레젠테이션은 고정된 시점을 기준으로 하나 이상의 JMP 보고서에 출력된 내용을 포함합니다. 이 출력은 출력을 PDF 파일로 저장할 때 생성되는 것과 유사합니다. 여기에는 데이터 테이블이 포함되지 않으며, 점을 선택하거나 점에 색상을 적용하거나 점을 숨기거나 제외하는 등의 작업도 불가능합니다. 글꼴, 축, 크기 등 표시 모양과 관련된 변경만 가능합니다. 하지만 외부 파일에 의존할 필요가 없다는 이점이 있습니다.
- 동적 프레젠테이션은 일련의 데이터 테이블 및 보고서를 구성하거나 숨기거나 표시할 수 있도록 하는 텍스트 및 버튼 (또는 링크) 이 포함된 개요 구조를 사용하여 생성됩니다. 이 유형의 저널은 사용자가 보고서 및 테이블을 열어 대화식으로 작업하거나, 점을 선택하거나, 데이터 부분집합을 추출하거나, 더 많은 그래프를 만들거나, 분석을 다시 실행할 수 있으므로 동적입니다.

그림 10.22에서는 저널에 추가할 수 있는 항목 중 일부를 보여 줍니다.

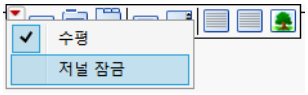
그림 10.22 저널의 예



참고 :

- 저널 생성을 마쳤으면 **편집 > 저널 잠금**을 선택합니다. 편집을 위해 저널을 잠금 해제하려면 **편집**을 선택하고 **저널 잠금**을 선택 취소합니다.

그림 10.23 저널 잠금



- 저널을 Microsoft Word, HTML, PDF 등의 다른 형식으로 저장할 수 있습니다 . 자세한 내용은 " 다른 형식으로 저널 저장 "(573 페이지) 에서 확인하십시오 .
- 저널은 파일 크기를 줄이기 위한 GZ 압축을 지원합니다 . 자세한 내용은 "JMP 환경 설정 " 장의 " 일반 JMP 환경 설정 "(621 페이지) 에서 확인하십시오 . 이 기능은 특히 큰 데이터 테이블을 저널에 포함하는 경우에 유용합니다 .
- " 일반 " 환경 설정에서 " 자동 저장 타임아웃 " 값을 설정하여 열려 있는 저널을 지정된 시간 (분) 마다 자동으로 저장하는 것이 좋습니다 . 이 자동 저장 값은 데이터 테이블 , 스크립트 , 프로젝트 및 보고서에도 적용됩니다 .

예 : 프레젠테이션용 저널 생성

대부분의 사람들은 프레젠테이션 시 PowerPoint 같은 프레젠테이션 응용 프로그램을 사용합니다. JMP 저널을 사용하면 프레젠테이션 응용 프로그램을 사용하지 않아도 됩니다. 모든 글머리 기호를 실시간 링크 및 버튼과 결합하여 표시하려는 분석을 쉽게 자동화할 수 있습니다. 보고서에서의 대화식 작업도 유지됩니다.

다음 예를 따라 저널을 사용한 프레젠테이션을 생성해 보십시오 .

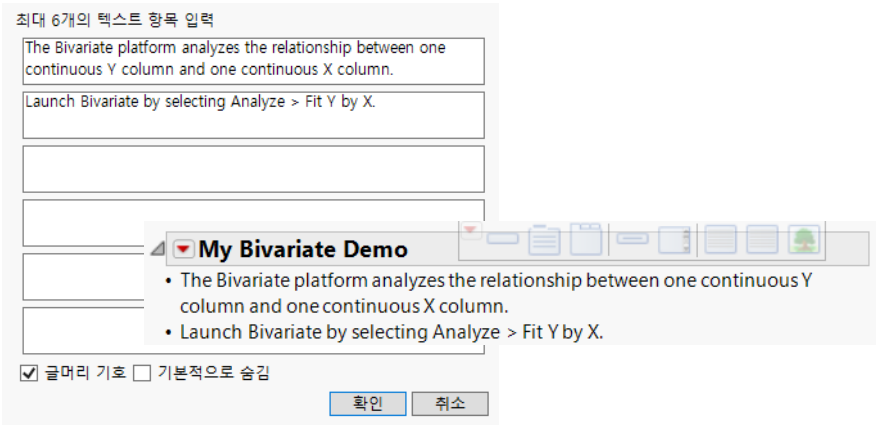
1. **파일 > 새로 만들기 > 저널**을 선택합니다 .
2. 저널을 마우스 오른쪽 버튼으로 클릭합니다 . **개요 항목 추가**를 선택하여 개요를 시작합니다 .
3. 프레젠테이션 제목 "My Bivariate Demo" 를 입력합니다 .

그림 10.24 개요 항목 추가

프레젠테이션에는 개요 노드가 있어야 합니다 . 노드는 내포 구조로 표시되어 프레젠테이션 시 차례로 열렸다 닫힙니다 .

4. 개요에 글머리 기호를 추가하려면 "My Bivariate Demo" 제목 표시줄의 빨간색 삼각형을 클릭하고 **텍스트 항목 추가**를 선택합니다 .
5. 그림 10.25 에 표시된 텍스트를 입력합니다 .

그림 10.25 텍스트 항목 추가



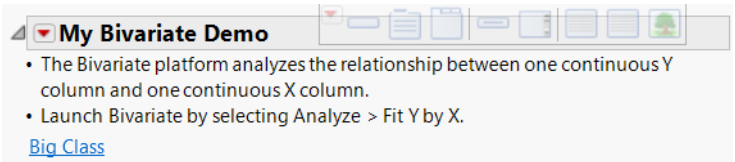
6. **글머리 기호** 체크박스를 선택하고 **확인**을 클릭합니다 .

팁 : 저널 상단에 깜빡이는 수평선이 나타납니다 . 화살표 키를 사용하여 개체를 삽입할 위치로 선을 이동합니다 .

이제 창 참조를 링크로 추가합니다. 이러한 링크는 프레젠테이션 시 클릭 한 번으로 파일을 열 수 있도록 합니다.

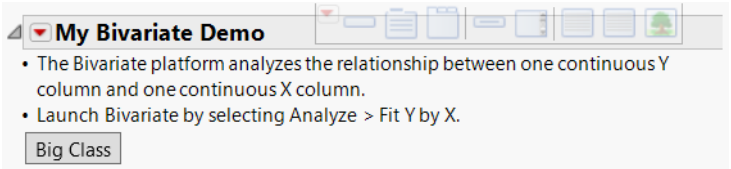
7. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
8. "My Bivariate Demo" 제목 표시줄의 빨간색 삼각형을 클릭하고 **창 참조 추가**를 선택합니다 .
9. "Big Class" 를 선택하고 **확인**을 클릭합니다 .

그림 10.26 창 참조 추가



10. 창 참조를 클릭하여 링크를 확인합니다 .
11. 링크를 마우스 오른쪽 버튼으로 클릭하고 **스타일 > 보통**을 선택하여 창 참조를 표시합니다 .

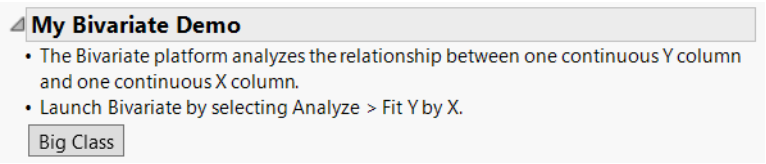
그림 10.27 링크를 버튼으로 변경



12. **편집 > 저널 잠금**을 선택하여 최종 저널을 확인합니다 .

    저널을 잠그면 저널 도구 모음과 삽입 커서가 더 이상 나타나지 않습니다 . 플랫폼에서 " 스크립트 저장 " > " 저널에 " 를 선택하면 저널이 열려 있는 잠긴 저널에 추가되지 않습니다 .

그림 10.28 최종 저널



팁 : 저널 도구 모음에서 개체를 드래그하거나 , 개체가 필요한 위치에 커서를 놓고 저널 도구 모음에서 개체를 두 번 클릭하면 저널에 개체를 추가할 수도 있습니다. 도구 모음은 커서를 도구 모음 위에 놓으면 사용할 수 있습니다 .

새 저널 생성

1. 열려 있는 모든 저널을 닫고 다음 중 하나를 수행합니다 .
   - 빈 저널을 생성하려면 **파일 > 새로 만들기 > 저널**을 선택합니다 . 또는 "JMP 시작하기 " 창의 " 저널 " 섹션에서 **새로 만들기**를 선택합니다 .
   - 전체 보고서 또는 데이터 테이블에서 저널을 생성하려면 **편집 > 저널**을 선택합니다 .
   - 특정 보고서 또는 보고서 내의 텍스트에서 저널을 생성하려면 선택 도구 ( ) 를 클릭한 채로 드래그하여 보고서 또는 데이터 테이블 내의 인접한 항목을 선택합니다 . 연속되지 않은 항목을 선택하려면 Shift 키를 누른 채로 선택 도구 ( ) 를 사용하여 항목을 클릭합니다 . 그런 다음 **편집 > 저널**을 선택합니다 .
   - 데이터 테이블 내의 특정 텍스트에서 저널을 생성하려면 텍스트를 선택한 후 **편집 > 저널**을 선택합니다 .
   - 대부분의 그래프에서 저널을 생성하려면 그래픽을 마우스 오른쪽 버튼으로 클릭한 후 **편집 > 저널**을 선택합니다 . 표면 그림과 3D 산점도의 경우에는 이 기능을 사용할 수 없습니다 .

팁 : 바로 가기 키를 선호하는 경우 Ctrl+Shift+J(Windows) 또는 Command+J(macOS) 를 눌러 새 저널을 생성합니다 .

2. **파일 > 저장**을 선택하고 저널을 .jrn 형식으로 저장합니다 .

공유 저널의 링크 관리

컴퓨터에 있는 데이터 테이블에 대한 링크를 삽입하면 해당 데이터 테이블의 전체 경로가 저널에 포함됩니다. 저널을 공유하거나 다른 컴퓨터로 이동하려면 연결된 파일을 저널과 동일한 디렉터리 또는 공유 네트워크 드라이브에 저장해야 합니다. 이 경우 파일이 동일한 디렉터리에 있기 때문에 전체 경로 이름을 사용하지 않아도 됩니다. 그러면 저널을 어디에서 열든 상관없이 링크가 그대로 유지됩니다.

팁 :

- 저널 및 지원 파일을 압축하여 다른 사용자가 저널 및 지원 파일을 한 디렉터리에 추출할 수 있도록 하십시오. 파일이 해당 디렉터리에 유지되어 있으면 링크가 정상적으로 작동합니다.
- 샘플 데이터 테이블에 대한 링크를 생성하는 경우에는 경로에 $SAMPLE_DATA 변수를 사용하는 것이 좋습니다 . 마찬가지로 , 파일이 바탕 화면에 있으면 $DESKTOP 변수를 사용하고 , 사용자 개인의 문서 폴더에 있으면 $DOCUMENTS 를 사용할 수 있습니다 .
- 개요 내에 있는 몇 개의 링크를 한 번에 편집할 수 있습니다. 링크의 문자열을 바꿀 수도 있습니다 . 개요 옆의 표시 아이콘을 마우스 오른쪽 버튼으로 클릭한 후 **항목 추가 > 링크 편집**을 선택하면 됩니다 .

저널에 이동형 링크를 생성하려면

1. 참조하려는 데이터 테이블 및 기타 파일을 동일한 폴더에 넣습니다 .
2. 저널을 해당 폴더에 저장합니다 .
3. 저널 하단을 마우스 오른쪽 버튼으로 클릭하고 **파일 참조 추가**를 선택한 후, 파일을 선택하고 **열기**를 클릭합니다 .
4. 저널의 링크를 마우스 오른쪽 버튼으로 클릭하고 **편집 > 스크립트 설정**을 선택합니다 .

   전체 파일 경로는 다음과 같습니다 .

   ```
   Open("C:\Users\shlori\journals\San Francisco Traffic.jmp")
   ```
5. 파일 이름을 제외하고 이 문자열의 모든 부분을 삭제합니다 .

   ```
   Open("San Francisco Traffic.jmp")
   ```

저널에 데이터 테이블을 포함하려면

외부 데이터 테이블에 연결할 필요가 없도록 데이터 테이블을 스크립트에 복사하여 저널에 포함할 수 있습니다. 이렇게 하면 링크를 편집하지 않아도 다른 사용자가 항상 해당 데이터에 액세스할 수 있다는 장점이 있습니다. 반면, 데이터를 편집하려는 경우에는 이 스크립트를 편집해야 한다는 단점도 있습니다.

1. 데이터 테이블에서 데이터 테이블 이름 옆의 빨간색 삼각형을 클릭하고 **테이블 스크립트 복사**를 선택합니다 .

그림 10.29 테이블 스크립트 복사

2. 저널에서 마우스 오른쪽 버튼을 클릭하고 **스크립트 버튼 추가**를 선택합니다.
3. " 이름 " 상자에 스크립트 이름을 입력합니다.
4. " 스크립트 " 상자에 테이블 스크립트를 붙여넣고 **확인**을 클릭합니다.
5. 링크를 마우스 오른쪽 버튼으로 클릭하고 **밑줄 스타일**을 선택 취소하여 링크를 버튼으로 표시합니다.

저널에 개요 추가 및 편집

텍스트 및 링크를 개요로 그룹화한 후 회색 표시 아이콘을 클릭하여 그룹을 펼치거나 접을 수 있습니다.

- 개요를 추가하려면 저널을 마우스 오른쪽 버튼으로 클릭한 후 **항목 추가 > 개요 항목 추가**를 선택합니다. 생성한 개요 수준에 내포되는 개요 수준을 추가할 수도 있습니다. 저널을 생성한 그래프의 빨간색 삼각형 메뉴를 클릭하면 **개요 항목 추가** 옵션을 사용할 수 없습니다. 개요의 빨간색 삼각형을 클릭하고 **개요 항목 추가**를 선택합니다.
- 개요 머리글을 편집하려면 머리글을 두 번 클릭하고 새 이름을 입력한 후 Enter 키를 누릅니다.
- 개요 내에 텍스트 또는 링크를 추가하려면 개요의 빨간색 삼각형을 클릭하고 " 저널의 마우스 오른쪽 버튼 클릭 옵션 "(571 페이지) 에 설명된 옵션을 선택합니다. 스크립트를 생성한 항목의 빨간색 삼각형을 클릭할 때는 이러한 옵션을 사용할 수 없습니다.
- 개요 외부에 텍스트 또는 링크를 추가하려면 저널을 마우스 오른쪽 버튼으로 클릭한 후 " 저널의 마우스 오른쪽 버튼 클릭 옵션 "(571 페이지) 에 설명된 옵션 중 하나를 선택합니다.
- 링크를 편집하려면 링크를 마우스 오른쪽 버튼으로 클릭하고 **편집 > 스크립트 설정**을 선택하여 경로를 편집한 후 **확인**을 클릭합니다.
- 저널에 표시된 링크의 이름을 편집하려면 링크를 마우스 오른쪽 버튼으로 클릭하고 **버튼 이름 설정**을 선택하여 새 이름을 입력한 후 **확인**을 클릭합니다.

저널의 마우스 오른쪽 버튼 클릭 옵션

개요 항목 추가 텍스트 및 링크를 접기 및 펼치기 가능한 목록으로 그룹화합니다 .

텍스트 항목 추가 최대 6 개의 단락을 입력하며 , 글머리 기호를 추가하거나 텍스트를 숨기는 옵션을 사용할 수 있습니다. 숨기기를 해제하려면 숨겨진 텍스트를 마우스 오른쪽 버튼으로 클릭하고 **숨기기**를 선택 취소합니다 .

참고 : 개요 수준 외부에 텍스트를 추가하려는 경우 저널의 끝부분을 두 번 클릭한 후 나타나는 상자에 텍스트를 입력할 수도 있습니다 . 편집 모드를 종료하려면 상자 외부를 클릭하십시오 .

창 참조 추가 JMP 에 이미 열려 있는 창에 대한 링크를 생성합니다 .

파일 참조 추가 컴퓨터에 있는 파일에 대한 링크를 생성합니다 .

파일 디렉터리 추가 선택한 디렉터리에 있는 모든 파일 (JMP 파일 외의 파일 포함) 에 대한 링크를 생성합니다 .

열려 있는 파일 추가 JMP 에 열려 있는 모든 파일 (로그 제외) 에 대한 개요 수준 및 링크를 생성합니다 . 예를 들어 열려 있는 Big Class.jmp 데이터 테이블 및 차트에 대한 링크는 Big Class 라는 접기 가능한 머리글 아래에 표시됩니다 . 저널 , 스크립트 및 자습서 같은 파일은 **기타 파일**이라는 접기 가능한 머리글 아래에 그룹화됩니다 .

URL 참조 추가 HTTP, FTP 또는 FILE:// 등의 인터넷 프로토콜을 통해 제공되는 파일에 대한 링크를 생성합니다 .

스크립트 버튼 추가 JSL 스크립트에 대한 링크를 생성합니다. 링크를 마우스 오른쪽 버튼으로 클릭하고 **밑줄 스타일**을 선택 취소하여 링크를 버튼으로 표시합니다 .

스크립트 상자 추가 스크립트를 입력하는 상자를 추가합니다 .

잠금 저널 도구 모음과 삽입 커서를 숨깁니다 . 플랫폼에서 " 스크립트 저장 " > " 저널에 " 를 선택하면 결과가 잠긴 저널에 추가되지 않습니다 .

링크 편집 예를 들어 하드 코딩된 경로를 제거하기 위해 파일 링크의 이름을 바꿀 수 있습니다 . 링크를 테스트하는 옵션도 제공됩니다 . 경로가 올바르면 " 링크 테스트 " 클릭 시 아무것도 발생하지 않습니다 . 웹 주소가 올바르지 않으면 로그에 오류 메시지가 나타납니다 . 올바르지 않은 링크를 업데이트하려면 **바꾸기**를 클릭합니다 .

화면 크기에 맞게 버튼 이름이 완전히 표시될 수 있도록 창 크기를 조정합니다 .

저널에서 개요 수준 표시 제어

예를 들어 선택한 수준 아래의 모든 하위 수준을 닫는 등, 표시되는 개요 수준을 즉석에서 제어할 수 있습니다. 해당 수준의 빨간색 삼각형을 클릭하고 **개요 상자**를 선택한 후 " 열기 " 또는 " 닫기 " 명령 중 하나를 선택합니다. " 열기 " 및 " 닫기 " 명령에 대한 자세한 내용은 "JMP 보고서" 장의 "JMP 보고서의 표시 옵션"(409 페이지)에서 확인하십시오 .

팁 : Windows 의 경우 , Alt 키를 누른 채 개요 수준의 회색 표시 아이콘을 마우스 오른쪽 버튼으로 클릭하면 개요 상자 옵션이 표시됩니다 .

저널에 그래프 또는 그래픽 추가

저널에서는 PNG, GIF, JPG, BMP 및 TIF 파일이 지원됩니다. 저널에 그래픽을 추가하려면 다음 중 하나를 수행하십시오.

- JMP 내의 그래프 또는 다른 응용 프로그램의 그래픽을 복사한 후 **편집 > 붙여넣기**를 선택합니다 . 그러면 해당 그래픽이 저널 끝에 추가됩니다 .
- 다른 창의 그래프 또는 그래픽을 저널로 드래그하여 놓습니다 . 자세한 내용은 "JMP 보고서 " 장의 "JMP 그래프에 배경 이미지 붙여넣기 "(473 페이지) 에서 확인하십시오 . 파란색 선은 그래픽을 놓을 수 있는 위치를 나타냅니다 .

팁 : 그래픽 크기는 외부 프로그램에서 조정하는 것이 좋습니다 . 그러면 드래그 방법으로 그래픽 크기를 조정하지 않고 100% 해상도의 그래픽을 저널에 삽입할 수 있습니다 .

저널의 사용자 정의 항목

- 저널 항목을 클릭하여 다른 위치로 드래그하려면 선택 도구 ( ) 를 선택하고 항목을 선택한 후 해당 항목을 새 위치로 드래그합니다 . 선은 항목을 드래그할 수 있는 위치를 나타냅니다 .
- 텍스트 줄바꿈을 조정하려면 텍스트를 마우스 오른쪽 버튼으로 클릭하고 **줄바꿈 설정**을 선택합니다 . 원하는 줄당 픽셀 수를 입력하고 **확인**을 클릭합니다 . 모든 텍스트 항목에 이 줄바꿈을 적용하려면 **줄바꿈 설정**을 선택하고 **이 설정을 다른 텍스트 상자에도 적용**을 선택한 후 **확인**을 클릭합니다 .
- 그림 축을 수정하려면 축을 두 번 클릭하거나 드래그합니다 . 자세한 내용은 "JMP 보고서 " 장의 " 그래프의 축 및 축 라벨 사용자 정의 "(455 페이지) 에서 확인하십시오 .
- 보고서의 임의 위치에 텍스트 또는 형태를 추가하려면 그리기 도구 (주석 달기 , 선 , 다각형 및 원 모양 도구) 를 사용합니다 . 그리기 도구에 대한 자세한 내용은 "JMP 보고서 " 장의 "JMP 보고서 및 그래프에 주석 및 형태 추가 "(477 페이지) 에서 확인하십시오 .
- 그림 및 그래프의 크기를 조정하려면 다음 중 하나를 수행합니다 .
  - 모서리를 클릭한 채로 드래그합니다 .
  - 그림 또는 그래프를 마우스 오른쪽 버튼으로 클릭한 후 **크기 / 척도 > 프레임 크기**를 선택합니다 . 자세한 내용은 "JMP 보고서 " 장의 " 그래프 크기 조정 "(452 페이지) 에서 확인하십시오 .
- 숫자 열의 형식을 수정하려면 표시 상자의 " 특성 " 페인을 편집합니다 . 자세한 내용은 "JMP 보고서 " 장의 " 숫자 열의 형식 수정 "(434 페이지) 에서 확인하십시오 .

저널 수정 방지

저널이 열려 있는 경우 다른 보고서 또는 데이터 테이블의 저널을 생성하면 두 번째 보고서 또는 데이터 테이블이 열려 있는 저널의 끝에 추가됩니다.

저널 수정을 방지하려면 저널 하단의 빈 영역을 마우스 오른쪽 버튼으로 클릭하고 **잠금**을 선택합니다. 그러면 추가 보고서가 다음 섹션에 설명된 것처럼 이 저널에 추가되지 않습니다. **편집 > 저널** 명령을 다시 선택하면 JMP는 해당 저널 대신 열려 있고 잠기지 않은 다른 저널을 찾거나 새 저널을 생성합니다.

저널을 잠금 해제하려면 저널 하단의 빈 영역을 마우스 오른쪽 버튼으로 클릭하고 **잠금**을 선택 취소합니다.

저널에 보고서 추가

이미 저널을 생성한 보고서에 다른 보고서를 추가하려면 해당 저널을 열고 **편집 > 저널**을 선택합니다. 분석 창에 선택된 영역이 있는 경우 **편집 > 저널**을 선택하면 전체 창 대신 선택된 영역만 저장됩니다.

참고 :

- 저널 창에는 보고서 창과 동일한 기능이 있습니다 . 즉 , 아이콘을 클릭하고 , 클릭 후 드래그하고 , 마우스 오른쪽 버튼을 클릭하여 메뉴에 액세스할 수 있습니다 .
- 보고서의 저널이 생성된 경우 저널 복사본은 더 이상 데이터 테이블에 연결되어 있지 않습니다 . 연결을 유지하려면 보고서를 JMP 보고서로 저장하고 이 보고서에 대한 링크를 저널에 삽입해야 합니다 .

다른 형식으로 저널 저장

저널을 HTML 또는 PDF 등의 다른 형식으로 저장할 수 있습니다. 저널을 HTML로 저장하면 웹 브라우저에서 파일을 볼 때 버튼 및 링크가 활성화됩니다.

참고 : PDF 파일에는 프린터 설정에 정의된 머리글 및 바닥글이 나타납니다 . 하지만 JSL 스크립트에서 왼쪽 , 가운데 및 오른쪽에 표시되는 머리글 및 바닥글을 설정할 수 있습니다 . JMP " 스크립트 인덱스 " 에서 `Set Print Footers` 메시지를 검색하십시오 . JMP 에서 **도움말 > 스크립트 인덱스**를 선택하면 " 스크립트 인덱스 " 에 액세스할 수 있습니다 .

Windows

1. **파일 > 내보내기**를 선택합니다 .
2. 형식을 선택합니다 . 자세한 내용은 " 보고서 저장 및 공유 "(533 페이지) 에서 확인하십시오 .
3. **다음**을 클릭합니다 .
4. 파일 이름을 지정하고 파일을 저장합니다 .

macOS

1. **파일 > 내보내기**를 선택합니다 .
2. 형식을 선택합니다 . 자세한 내용은 " 보고서 저장 및 공유 "(533 페이지) 에서 확인하십시오 .
3. **다음**을 클릭합니다 .
4. 파일 이름을 지정한 후 내보냅니다 .

저널에서 항목 삭제

1. 선택 도구 ( ) 를 클릭합니다 .
2. 삭제할 항목을 선택하고 Delete 키를 누릅니다 .

JMP 세션 저장

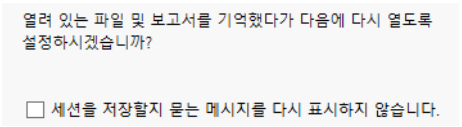
JMP를 사용하는 매 시기를 **세션**이라고 합니다. **저장된 세션**은 문서를 다시 열고 분석을 다시 실행하여 해당 세션 스크립트가 저장된 시점의 JMP 상태를 복원하는 JSL 스크립트입니다. 저장된 세션은 수동으로 파일을 다시 열고 분석을 다시 실행할 필요 없이 이전 상태로 다시 돌아가는 데 유용합니다.

고급 사용자라면 저장된 세션에 보존되는 세션 정보가 무엇인지 이해하는 것이 중요합니다. 저장된 모든 문서(데이터 테이블, 스크립트, 저널 등)가 다시 열립니다. 또한 스크립트 저장을 지원하는 JMP 창이 다시 실행됩니다. 이는 보고서의 빨간색 삼각형을 클릭하고 **다시 실행 > 분석 다시 실행**을 선택하는 것과 동등합니다. 스크립트 실행을 통해 부가적으로 얻어지는 결과(예: 전역 변수 값 또는 사용자 창)는 저장되지 않습니다. 대화식 작업이 많은 분석의 상태도 저장되지 않습니다.

JMP 종료 시 세션 저장

저장된 세션은 주로 JMP 종료 시 상태를 저장했다가 JMP 재시작 시 상태를 복원하려는 경우에 사용합니다. 기본적으로 프로그램을 종료할 때마다 세션 상태를 저장할지 묻는 메시지가 표시됩니다(그림 10.30). 이 기능을 사용하면 JMP를 종료한 후 나중에 이전 작업에 사용한 파일을 열지 않고도 이전 세션으로 돌아갈 수 있습니다.

그림 10.30 종료 시 세션 저장



종료 시 항상 세션을 저장하려면 **세션을 저장할지 묻는 메시지를 다시 표시하지 않습니다.**를 선택하고 **예** 또는 **아니요**를 클릭합니다. 이 옵션으로 JMP가 종료 시 세션을 저장하지 않도록 할 수도 있습니다.

나중에 환경 설정에서 이 옵션을 변경할 수 있습니다.

1. **파일 > 환경 설정**을 선택합니다.
2. **일반** 아이콘을 선택합니다.
3. **종료할 때 세션 저장** 옵션 옆에서 JMP가 언제 세션을 저장하도록 할지 선택합니다. 항상 저장하거나, 저장하지 않거나, 저장 여부를 사용자에게 확인하도록 선택할 수 있습니다.

수동으로 세션 저장

선택한 위치에 세션을 저장하고 작업을 계속했다가 필요할 때 언제든지 저장된 상태를 복원할 수도 있습니다. 세션 저장 및 복원을 보다 세부적으로 제어하려는 경우에 세션을 수동으로 저장하면 유용합니다. 이 옵션은 특히 여러 개의 독립된 세션 상태를 보관하려는 경우에 유용합니다. 각 상태마다 서로 다른 파일 및 분석이 관련됩니다.

JMP 세션의 스크립트를 생성하려면

1. **파일 > 세션 스크립트 저장**을 선택합니다.
2. 창에 스크립트의 이름을 입력하고 **저장**을 클릭합니다.

세션을 수동으로 저장하는 또 다른 방법은 각 세션의 저널을 생성하는 것입니다. 저널은 노트북 스타일 또는 프로젝트 스타일의 파일입니다. 저널을 사용하면 프로젝트 내의 파일에 대한 참조를 취합하고, 프레젠테이션 시작 패드를 만들고, 프로젝트를 문서화하고, 여러 스크립트를 한 곳에 저장할 수 있습니다. 자세한 내용은 "저널 생성"(565 페이지)에서 확인하십시오.

JMP 세션의 저널을 생성하려면

1. 저널에 포함할 파일을 엽니다.
2. **파일 > 새로 만들기 > 저널**을 선택합니다. 열려 있는 파일을 기존 저널에 추가하려면 해당 저널을 엽니다.
3. 빈 저널을 마우스 오른쪽 버튼으로 클릭하고 **모든 열려 있는 파일 추가**를 선택합니다.

로그 창 저장

보기 > 로그를 선택하면 JMP 작업 상태(JSL 문의 실행 시 상태와 스크립트 오류 등)를 모니터링하는 창이 표시됩니다. 계산식, 행렬 또는 다른 연산을 실행하는 중에 중요한 정보가 로그 창에 기록되는 경우도 있을 수 있습니다.

로그를 저장하려면

1. 로그 창을 엽니다.
2. **파일 > 다른 이름으로 저장**을 선택합니다.
3. (Windows에만 해당) 로그에 텍스트가 추가될 때 로그를 자동으로 열려면 **파일 > 환경 설정 > Windows 관련**을 선택합니다. 그런 다음 **JMP 로그 창 열기** 목록에서 **텍스트가 추가될 때마다**를 선택합니다.

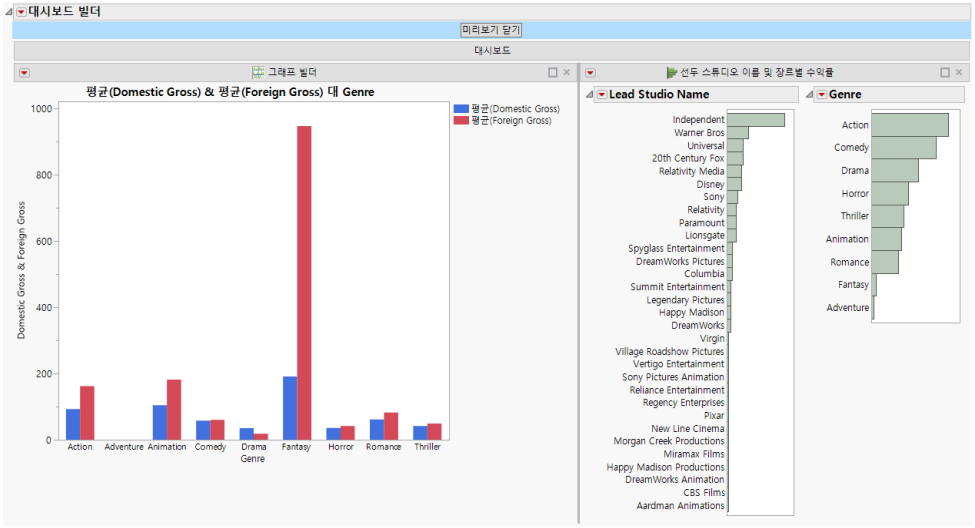
11 장

대시보드를 사용하여 보고서 결합 및 표시

JMP 확장

JMP는 확장 가능한 프로그램이므로 사용자가 JMP의 기능을 확장하여 고유한 JMP 구성 요소를 생성할 수 있습니다. 대시보드 빌더를 사용하면 정기적으로 보고서를 실행하고 표시할 수 있는 시각적 도구인 대시보드에 여러 보고서를 결합할 수 있습니다. JSL에 대한 지식이 없어도 됩니다. 템플릿에서 시작하여 보고서 구성 요소를 클릭하고 그래프와 그림을 삽입하기만 하면 대시보드가 완성됩니다. 대시보드 빌더 외부에서 창을 결합하여 대시보드를 생성할 수도 있습니다.

그림 11.1 두 개의 보고서가 포함된 대시보드



목차

대시보드를 생성하여 JMP 보고서 결합 579
예: 두 개의 JMP 보고서가 포함된 대시보드 생성 579
대시보드 빌더 사용 580
대시보드 빌더의 빨간색 삼각형 옵션 581
샘플 대시보드를 사용하여 시작 582
대시보드 편집 582
실행 중인 대시보드 보기 586
예: 창을 결합하여 대시보드 생성 590
예: 대시보드에 선택 필터 추가 591
예: JMP 쿼리 대시보드 및 추가기능 생성 593
예: 두 개의 데이터 테이블에서 대시보드 생성 595

대시보드를 생성하여 JMP 보고서 결합

대시보드 빌더를 사용하면 JSL 스크립트를 작성하지 않고도 여러 보고서를 결합하여 대시보드를 생성할 수 있습니다. 응용 프로그램 빌더와 마찬가지로 대시보드 빌더에도 드래그앤드롭 인터페이스가 있지만 사전 구성된 표시 상자가 사용된다는 차이점이 있습니다. 대시보드 빌더에서는 응용 프로그램 빌더와 달리 표시 상자를 사용자 정의할 필요가 없습니다. 탭에 보고서와 그래프를 추가하고, 데이터베이스 쿼리를 포함하고, 대시보드를 추가기능으로 저장하는 등의 작업을 수행하기만 하면 됩니다.

"창" > "창 결합" 옵션으로 여러 개의 열려 있는 창을 결합하여 대시보드를 빠르게 생성할 수도 있습니다. 예를 들어 두 개의 보고서와 한 개의 데이터 테이블을 결합하여 하나의 대시보드에 배열할 수 있습니다. 예는 예: 창을 결합하여 대시보드 생성에서 확인하십시오.

예 : 두 개의 JMP 보고서가 포함된 대시보드 생성

두 개의 보고서를 생성했고 다음날 업데이트된 데이터 집합에 대해 보고서를 다시 실행하려고 한다고 가정해 보겠습니다. 이 예에서는 보고서에서 대시보드를 생성하는 방법을 보여 줍니다.

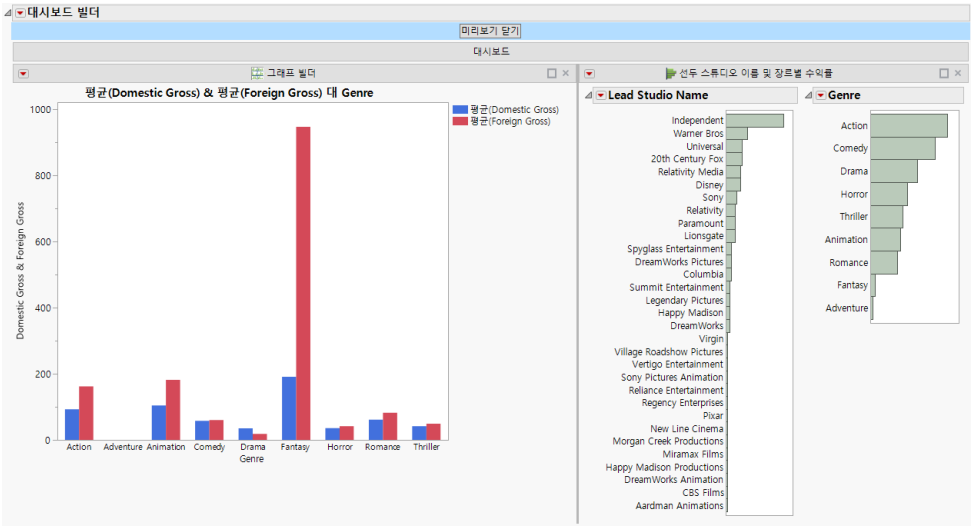
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hollywood Movies.jmp 를 엽니다 .
2. 보고서를 생성하기 위해 "Distribution: Profitability by Lead Studio and Genre" 및 "Graph Builder: World and Domestic Gross by Genre" 라는 테이블 스크립트를 실행합니다 .
3. 어느 창에서든 **파일 > 새로 만들기 > 대시보드**를 선택합니다 .

   일반적인 레이아웃을 위한 템플릿이 나타납니다 .
4. **2x1 대시보드** 템플릿을 선택합니다 .

   두 개의 보고서를 위한 공간이 있는 상자가 작업 공간에 나타납니다 .
5. " 보고서 " 목록에서 보고서 썸네일을 두 번 클릭하여 대시보드에 배치합니다 (그림 11.3).
6. " 대시보드 빌더 " 의 빨간색 삼각형을 클릭하고 **미리보기 모드**를 선택합니다 .

   대시보드의 미리보기가 나타납니다 . 그래프가 서로 연결되고 데이터 테이블과 연결되어 있습니다 . 또한 " 분포 " 및 " 그래프 빌더 " 플랫폼과 동일한 빨간색 삼각형 옵션이 있습니다 .

그림 11.2 두 개의 JMP 보고서가 포함된 대시보드



7. **미리보기 닫기**를 클릭합니다 .
8. 대시보드 빌더에서 **파일 > 저장**을 선택한 후 파일을 .jmpapp 파일로 저장합니다 .

   JMP 사용자는 데이터 테이블을 열고 이 파일을 두 번 클릭하여 대시보드를 표시할 수 있습니다 . 단 , .jmpapp 파일은 편집할 수 없습니다 .
9. 대시보드 빌더 창으로 돌아가서 **파일 > 다른 이름으로 저장**을 선택한 후 파일을 .jmpappsource 파일로 저장합니다 .

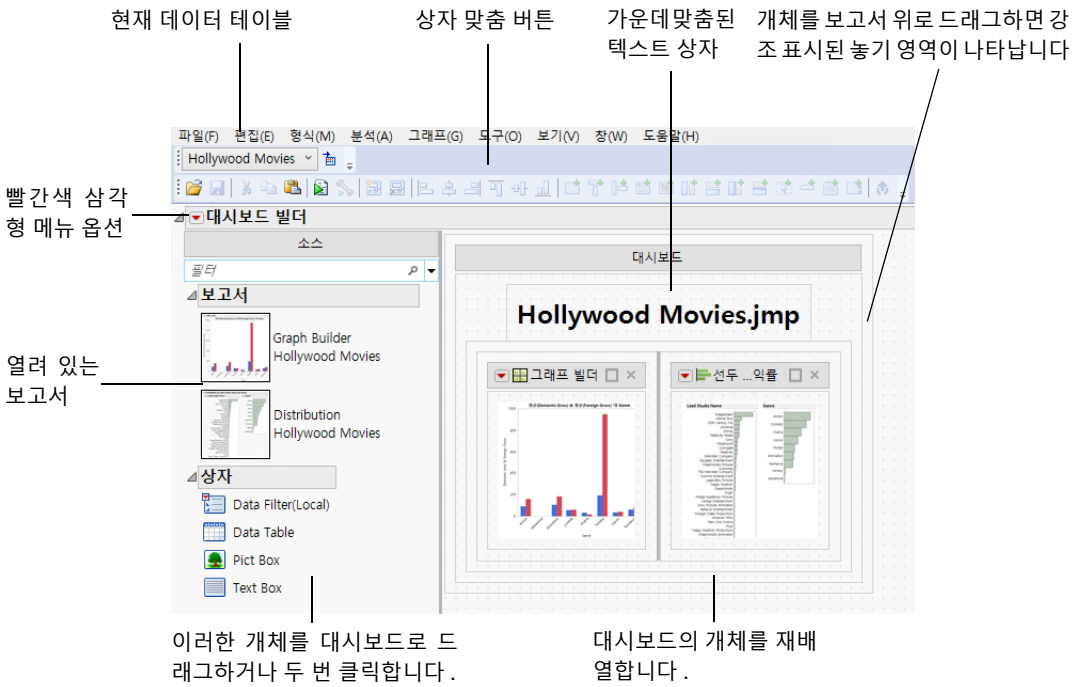
   대시보드가 이후 편집할 수 있는 파일로 저장됩니다 .

참고 : 대시보드를 추가기능으로 저장할 수도 있습니다 . 그러면 사용자가 .jmpapp 파일을 여는 대신 메인 " 추가기능 " 메뉴에서 메뉴 항목을 선택하여 대시보드를 볼 수 있습니다 . 자세한 내용은 " 예 : JMP 쿼리 대시보드 및 추가기능 생성 "(593 페이지) 에서 확인하십시오 .

대시보드 빌더 사용

그림 11.3 에서는 개발 중인 대시보드를 보여 줍니다. 왼쪽 페인에 있는 두 개의 보고서는 Hollywood Movies.jmp 샘플 데이터 테이블에서 생성되었습니다. 이 두 보고서가 오른쪽 페인에 추가되었습니다. 대시보드에 추가되고 도구 모음 버튼을 사용하여 가운데 맞춤된 텍스트 상자에는 "Hollywood Movies.jmp"가 표시되어 있습니다.

그림 11.3 대시보드 빌더 창



대시보드 빌더의 빨간색 삼각형 옵션

"대시보드 빌더"의 빨간색 삼각형 메뉴에서는 대시보드를 실행하고, 대시보드를 미리 보고, 대시보드를 추가기능으로 저장하는 등의 작업을 위한 옵션이 제공됩니다.

대시보드 실행 대시보드를 시작합니다.

미리보기 모드 대시보드의 미리보기를 표시합니다. 대시보드 모드로 돌아가려면 대시보드에서 **미리보기 닫기**를 클릭합니다.

다시 시작 작업 공간에서 개체를 제거하고 새 템플릿을 선택합니다.

격자 표시 작업 공간에 점선으로 된 격자선을 표시합니다. 기본적으로 선택되어 있습니다.

특성 표시 표시 상자를 편집하기 위한 " 개체 " 및 " 특성 " 패널을 표시하거나 숨깁니다. 특성 편집에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오.

자동 스크롤 개체를 작업 공간의 모서리 가까이로 드래그할 때 자동으로 가로 또는 세로로 스크롤합니다. 기본적으로 선택되어 있습니다.

대시보드 모드 응용 프로그램의 모든 요소를 사용자 정의하고 스크립트를 작성할 수 있는 응용 프로그램 빌더에 대시보드를 표시하려면 이 옵션을 선택 취소합니다 . 자세한 내용은 Scripting Guide의 에서 확인하십시오. 대시보드로 돌아가려면 **대시보드 모드**를 선택합니다.

소스 패널 보고서 구성 요소가 표시되는 왼쪽 열 , 즉 " 소스 " 패널을 표시하거나 숨기고 , 아이콘을 범주화합니다 . 기본적으로 " 소스 표시 " 가 선택되어 있습니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 일부 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

데이터 테이블에 보고서를 생성하는 데 사용된 데이터 테이블에 스크립트를 저장합니다 . 그러면 데이터 테이블에서 스크립트를 다시 실행하여 결과를 다시 생성할 수 있습니다 .

저널에 분석을 위해 스크립트를 실행하는 저널 창에 링크를 저장합니다 . 저널 창이 이미 열려 있으면 해당 저널 창에 스크립트가 추가됩니다 . 스크립트에는 데이터 테이블의 경로가 포함됩니다 . 데이터 테이블을 찾을 수 없으면 스크립트가 실행되지 않습니다 .

스크립트 창에 스크립트 편집기 창을 열고 이 창에 스크립트를 추가합니다 . 스크립트 창이 이미 열려 있으면 이 열려 있는 창의 맨 아래에 스크립트가 추가됩니다 .

추가기능에 JMP 사용자가 응용 프로그램을 설치하고 JMP 메뉴에서 이를 시작할 수 있도록 합니다 . 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

템플릿으로 응용 프로그램을 새 대시보드를 생성하는 데 사용할 수 있는 템플릿으로 저장합니다 . 템플릿은 사용하는 운영 체제 및 JMP 버전에 따라 다음 위치에 저장됩니다 .

- Windows(JMP):
  "/C:/Users/<username>/AppData/Roaming/SAS/JMP/16/DashboardTemplates"
- Windows(JMP Pro):
  "/C:/Users/<username>/AppData/Roaming/SAS/JMPPro/16/DashboardTemplates"
- macOS:
  "/Users/<username>/Library/Application Support/JMP/16/DashboardTemplates"

샘플 대시보드를 사용하여 시작

대시보드 빌더를 열면 몇 가지 대시보드 예가 표시됩니다. 빈 대시보드로 시작하는 대신 이러한 샘플을 수정할 수 있습니다. 이 대시보드는 JMP Samples/Dashboards 폴더에도 설치되어 있습니다.

사용 가능한 샘플 대시보드는 다음과 같습니다 .

Instant Dashboard.jmpappsource 두 개의 다변량 보고서 (" 주성분 / 요인 분석 " 및 " 모든 주성분이 포함된 T 제곱 ") 를 결합합니다 .

Six Quality Graphs Dashboard.jmpappsource 세 개의 관리도 , 한 개의 분포 보고서 및 한 개의 공정 능력 분석 보고서를 생성합니다 .

대시보드 편집

" 예 : 두 개의 JMP 보고서가 포함된 대시보드 생성 "(579 페이지) 에서는 두 개의 보고서로 대시보드를 생성하는 방법을 보여 줍니다. 이 섹션에서는 해당 보고서를 배열하고 최종 대시보드를 미리 보는 방법을 보여 줍니다.

대시보드 편집 모드에서 보고서 재배열

대시보드를 편집하는 동안 보고서를 쉽게 재배열할 수 있도록 보고서 주위의 외곽선에는 놓기 영역이 포함되어 있습니다. 보고서를 상자의 다른 영역으로 드래그하면 상자의 일부분이 강조 표시됩니다. 강조 표시는 해당 상자를 놓을 수 있는 위치를 보여 줍니다.

그림 11.4에서는 탭 형식 보고서를 생성하는 방법을 보여 줍니다. "분포" 보고서를 탭 안에 넣으려면 해당 막대 차트 보고서를 보고서의 오른쪽 상단이 강조 표시될 때까지 다른 보고서 위로 드래그합니다. 그러면 "분포" 보고서가 다른 보고서의 위에 탭으로 나타납니다.

탭 형식 보고서를 제거하려면 보고서의 가운데를 클릭하고 탭 바깥쪽으로 드래그합니다.

그림 11.4 대시보드 편집 모드에서 탭 형식 보고서 작업

탭 형식 표시를 위한 놓기 영역

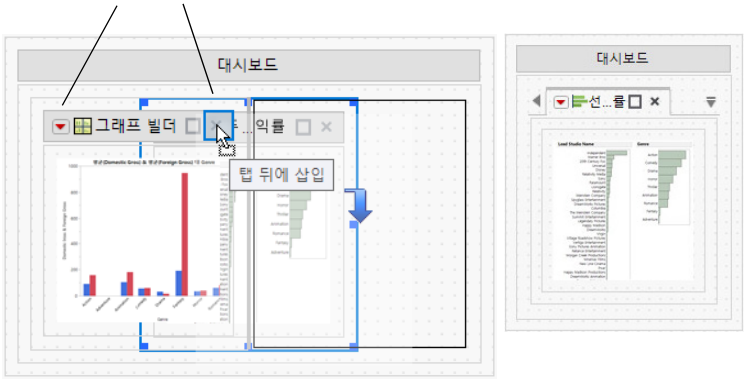


그림 11.5에서는 보고서의 하단 테두리 안에 데이터 테이블을 넣는 예를 보여 줍니다.

데이터 테이블을 보고서 내부가 아니라 보고서 **아래에** 표시하려면 "Data Table" 상자를 대시보드의 보고서 아래로 드래그합니다. 그러면 상자가 보고서의 하단 테두리 안쪽에 맞춰집니다.

그림 11.5 보고서 내부에 데이터 테이블 표시

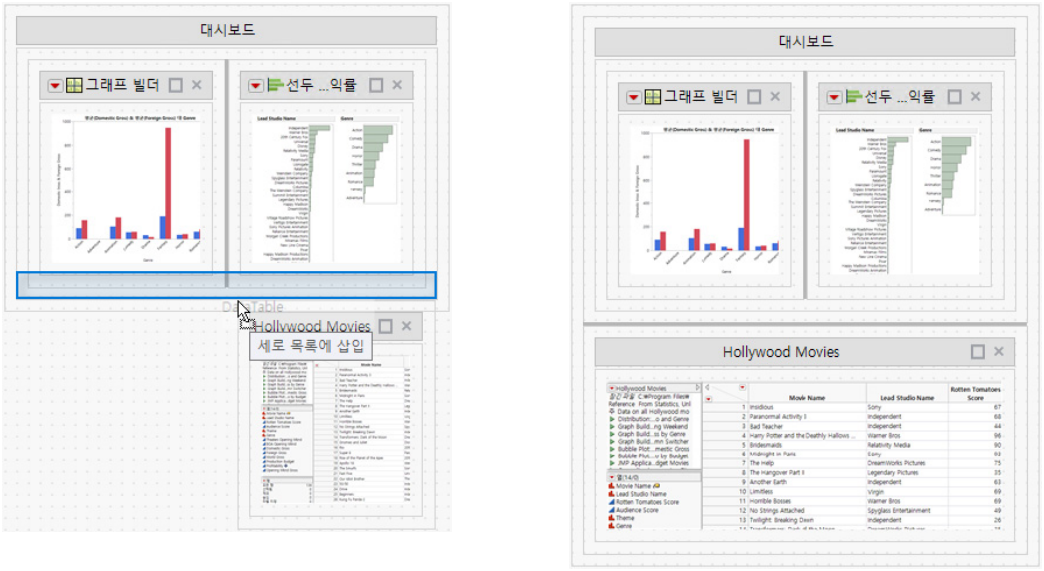
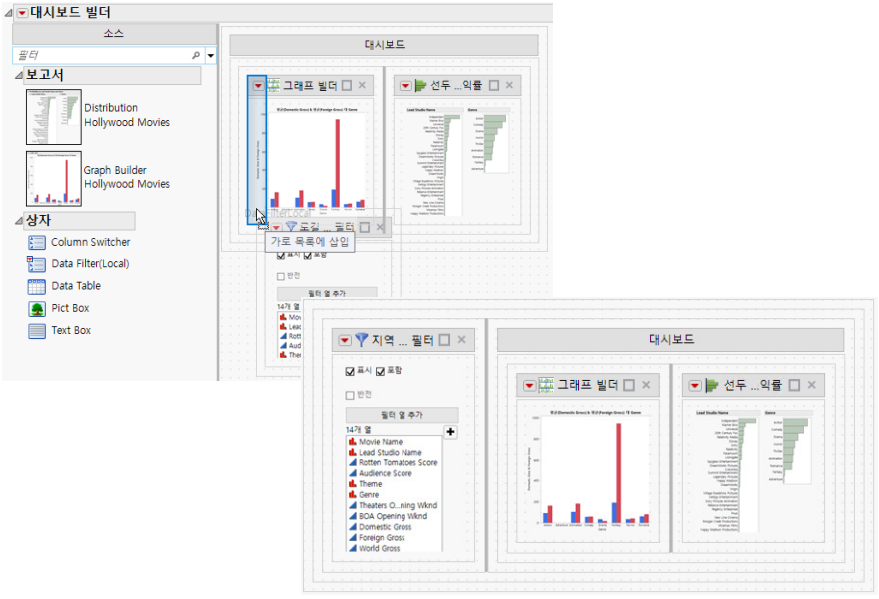


그림 11.6에서는 보고서의 왼쪽에 표시되는 데이터 필터를 보여 줍니다.

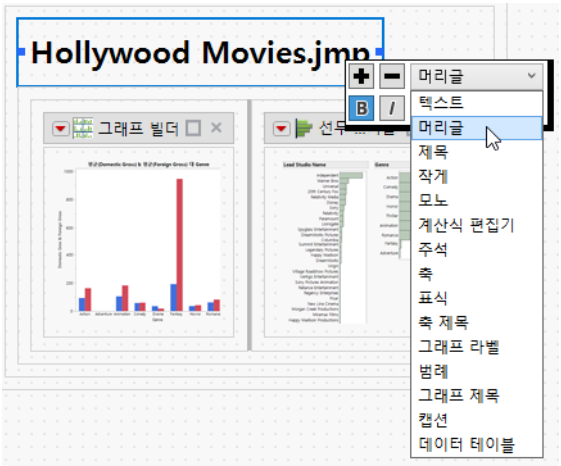
그림 11.6 보고서의 왼쪽에 데이터 필터 추가



대시보드의 텍스트 형식

대시보드를 편집할 때 텍스트 상자에 있는 텍스트의 형식을 지정할 수 있습니다. 텍스트 상자를 대시보드로 드래그하고 텍스트 상자의 가운데를 클릭합니다. 그러면 형식 지정 옵션이 표시됩니다.

그림 11.7 텍스트 상자의 형식 지정 옵션



대시보드 미리보기

대시보드를 편집할 때 "대시보드 빌더"의 빨간색 삼각형 메뉴에서 **미리보기 모드**를 선택하여 최종 대시보드의 미리보기를 볼 수 있습니다. 대시보드를 미리 보면 열 전환기, 데이터 필터 및 탭 형식 보고서와 같은 대화식 요소를 테스트할 때 특히 유용합니다.

참고 : 미리보기 모드에서 페인 크기를 조정할 경우에는 변경 사항이 저장되지 않습니다 .

팁 :

- 열 전환기는 처음에 테이블의 모든 열을 표시하지만 대시보드 또는 응용 프로그램의 보고서를 적극적으로 변경하지는 않습니다 .
- 미리보기 모드로 전환하여 다음을 수행할 수 있습니다 .
  - 모든 보고서에서 전환할 초기 열을 선택합니다 .
  - 보고서에 사용되는 다른 모든 열을 포함하여 열 전환기에 표시하지 않을 열을 제거합니다.
- 대시보드 또는 응용 프로그램을 실행하면 열 전환기가 동일한 데이터 테이블을 사용하는 대시보드의 모든 플랫폼에 자동으로 연결됩니다 .

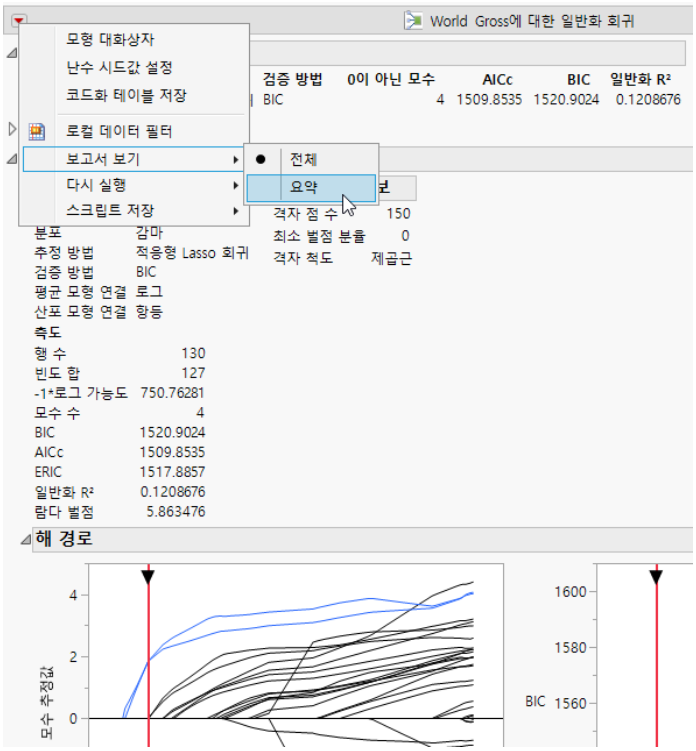
실행 중인 대시보드 보기

실행 중인 대시보드를 볼 때는 보고서에서 JMP 플랫폼에 표시되는 것과 동일한 빨간색 삼각형 메뉴 옵션을 사용할 수 있습니다. 대시보드 빌더에서는 다른 옵션도 몇 가지 제공합니다.

보고서 요약 보기

그래프와 통계량이 포함된 보고서는 크기가 매우 커질 수 있습니다. 보고서를 실행하는 동안 그래프만 보려면 보고서의 위쪽 빨간색 삼각형을 선택한 후 **보고서 보기 > 요약**을 선택합니다. 그림 11.8에 표시된 보고서는 원래 모수 추정값을 보여 주었습니다. 요약 보기에서는 그래프만 보여 줍니다.

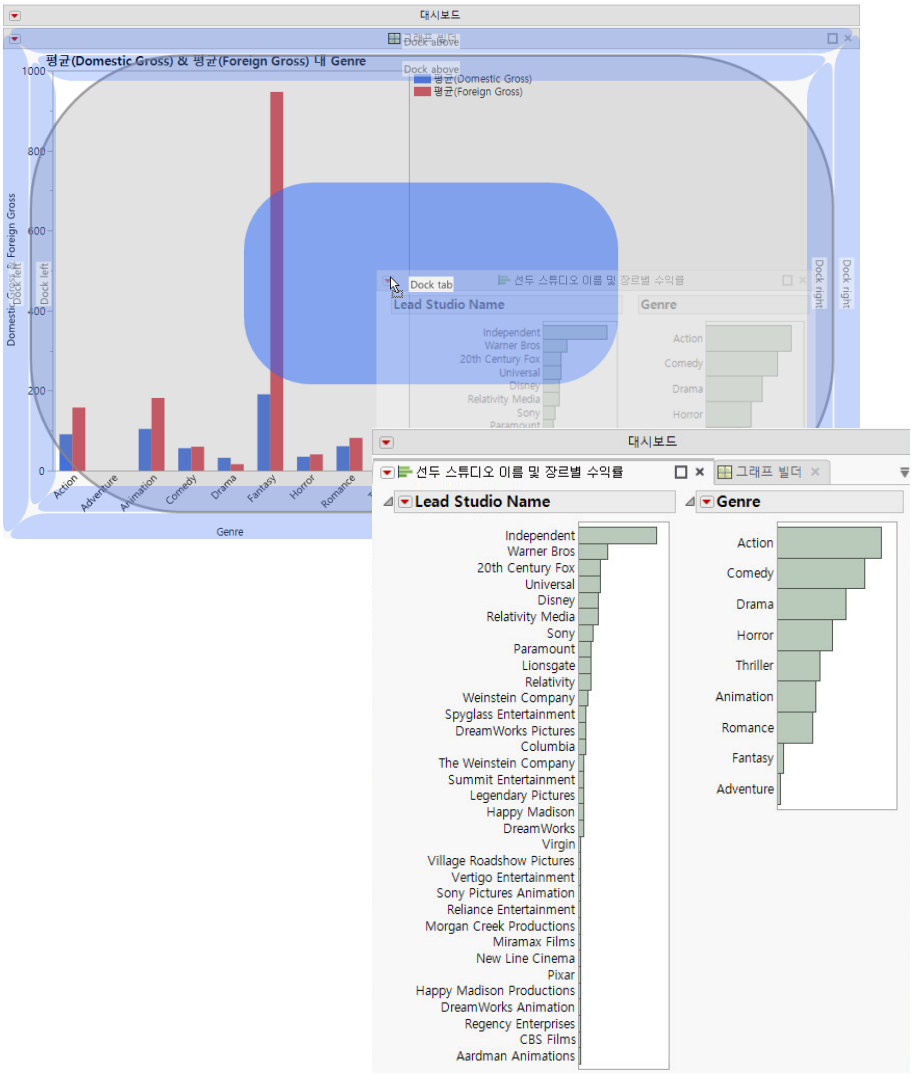
그림 11.8 보고서 요약 보기



실행 중인 대시보드에서 보고서 재배열

실행 중인 대시보드에서 보고서를 강조 표시된 놓기 영역으로 드래그하여 보고서를 재배열할 수 있습니다. 놓기 영역은 보고서를 놓을 수 있는 위치를 나타냅니다. 그림 11.9에서는 보고서를 탭 안으로 이동하는 예를 보여 줍니다. 분포 보고서를 "탭 도킹" 놓기 영역 안에 놓으면 첫 번째 탭에 해당 보고서가 배치됩니다.

그림 11.9 실행 중인 대시보드에서 탭 형식 보고서 생성

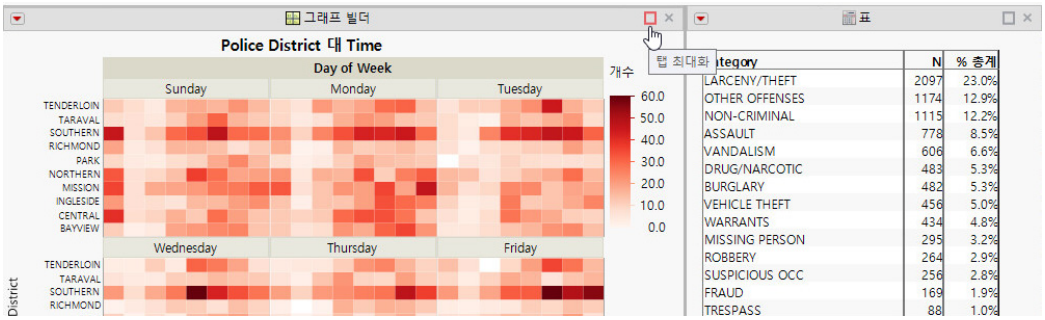


탭 형식 보고서를 제거하려면 보고서의 가운데를 클릭하고 탭 바깥쪽으로 드래그합니다.

실행 중인 대시보드에서 보고서 최대화

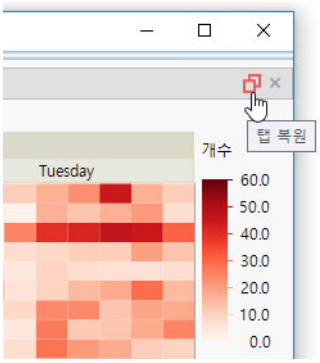
대시보드를 실행하는 동안 특정 보고서만 보려면 보고서의 **탭 최대화** 상자를 선택합니다.

그림 11.10 탭 최대화



보고서를 원래 크기로 되돌리려면 **탭 복원** 상자를 선택합니다.

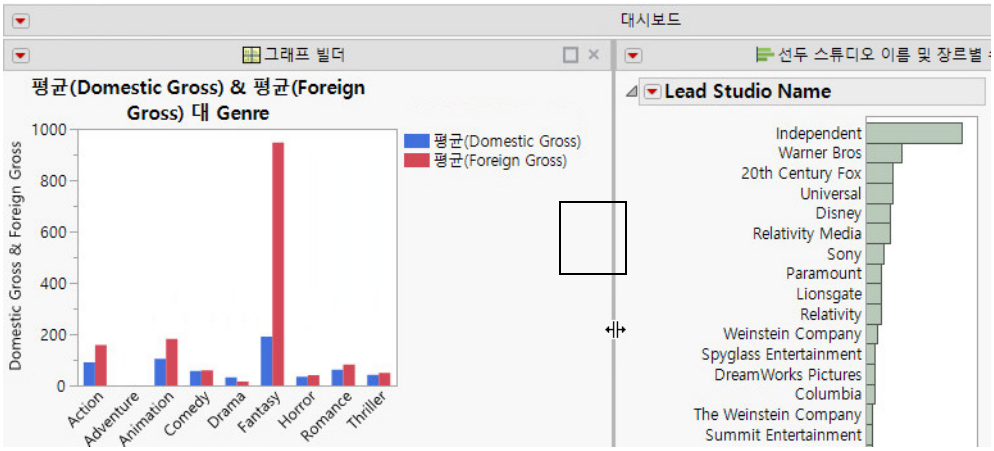
그림 11.11 최대화된 보고서 복원



실행 중인 대시보드에서 보고서 크기 조정

대시보드에서 보고서 사이에 표시되는 두꺼운 회색 선을 **분할선**이라고 합니다. 페인 크기를 조정하려면 이 분할선을 드래그합니다. 페인이 너무 넓어서 포함된 보고서가 표시되지 않으면 페인이 닫힙니다. 분할선을 대시보드의 바깥쪽 모서리에서 안쪽으로 드래그하면 다시 모든 보고서가 표시됩니다. 그림 11.12에서는 축소된 보고서와 확대된 보고서를 보여 줍니다.

그림 11.12 실행 중인 대시보드에서 페인 크기 조정



그래프의 모서리를 드래그할 수도 있습니다. 창 너비와 다른 보고서를 고려하여 허용되는 한도 내에서 그래프의 너비를 늘릴 수 있습니다.

JMP 외부에서 작업하는 사용자와 보고서 공유

대시보드를 대화식 HTML로 저장하면 JMP 외부에서 작업하는 사용자와 보고서를 공유할 수 있습니다. 먼저 "대시보드 빌더"의 빨간색 삼각형을 클릭하고 **대시보드 실행**을 선택하여 대시보드를 실행합니다. 그런 다음 실행 중인 대시보드에서 **파일 > 다른 이름으로 저장**을 선택한 후 목록에서 **데이터가 포함된 대화식 HTML**을 선택합니다. 그러면 보고서가 데이터도 포함된 웹 페이지로 저장됩니다. 이 웹 페이지를 사용자에게 전자 우편으로 보내거나 웹 사이트에 게시할 수 있습니다. 그러면 사용자들이 JMP에서처럼 데이터를 탐색할 수 있습니다. 자세한 내용은 보고서를 대화식 HTML로 공유에서 확인하십시오.

JMP 플랫폼 스크립트 편집

대시보드는 JSL 스크립트로 구성되며, 이 스크립트는 대시보드 빌더에서 편집할 수 있습니다. 예를 들어 다변량 플랫폼에서 산점도 행렬을 표시하거나 다른 열을 사용하도록 스크립트를 편집할 수 있습니다.

대시보드에서 플랫폼 개체를 마우스 오른쪽 버튼으로 클릭하고 **플랫폼 스크립트 편집**을 선택합니다. 그러면 해당 스크립트가 스크립트 편집기 창에서 열립니다.

스크립트를 편집하고 "확인"을 클릭하면 작업 공간의 개체가 업데이트됩니다.

예 : 창을 결합하여 대시보드 생성

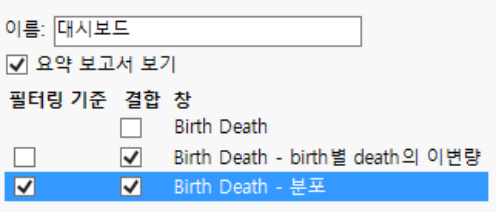
JMP 에서 열려 있는 여러 창을 병합하여 신속하게 대시보드를 생성할 수 있습니다 . 창을 결합하면 통계량 요약을 볼 수 있는 옵션이 제공되고 선택 필터가 포함됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Birth Death.jmp 를 엽니다 .
2. "Distribution" 및 "Bivariate" 테이블 스크립트를 실행합니다 .
3. 보고서 창 중 하나에서 **창 > 창 결합**을 선택합니다 .

   " 창 결합 " 창이 나타납니다 .

   팁 : Windows 에서는 JMP 창의 오른쪽 하단에 있는 " 배열 메뉴 " 옵션에서 " 창 결합 " 을 선택할 수도 있습니다 .
4. **요약 보고서 보기**를 선택하여 그래프를 표시하고 통계 보고서를 생략합니다 .
5. " 결합 " 열에서 **Birth Death - birth 별 death 의 이변량** 및 **Birth Death - 분포**를 선택합니다 .
6. " 필터링 기준 " 열에서 **Birth Death - 분포**를 선택합니다 .

그림 11.13 창 결합 옵션



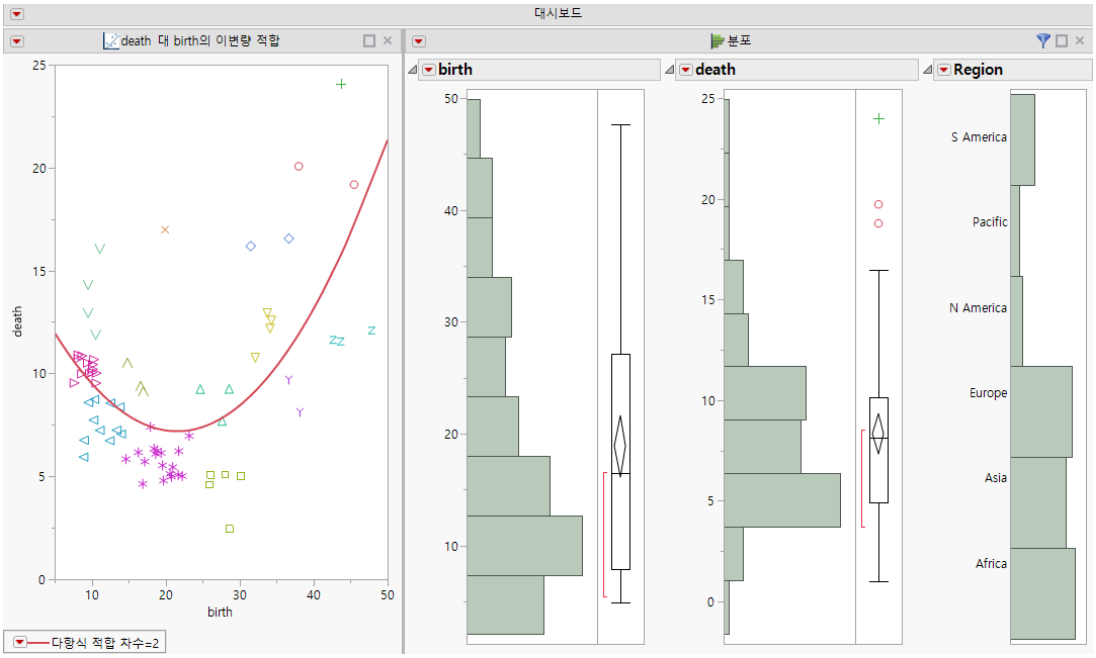
7. **확인**을 클릭합니다 .

   두 개의 보고서가 하나의 창으로 결합됩니다 . " 분포 " 보고서 상단의 필터 아이콘 을 확인합니다 . 히스토그램 중 하나에서 막대를 선택하면 이변량 그래프의 해당 데이터가 선택됩니다 .

참고 :

- Windows 에서 보고서를 결합하려면 JMP 창의 오른쪽 하단에 있는 " 배열 메뉴 " 옵션에서 창 결합을 선택해도 됩니다 .
- 보고서에 그래프만 표시하고 통계량은 생략하려면 " 창 결합 " 창에서 **요약 보고서 보기**를 선택하십시오 .

그림 11.14 결합된 창



예 : 대시보드에 선택 필터 추가

여러 개의 보고서가 포함된 대시보드에서는 한 보고서에서 데이터를 선택한 다음 동일한 창에 포함된 다른 보고서에서는 해당 데이터만 표시되도록 할 수 있습니다. 이 예에서는 대시보드를 편집할 때 기본 보고서에 선택 필터를 추가하는 방법을 보여 줍니다.

보고서 열기 및 대시보드 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hollywood Movies.jmp 를 엽니다 .
2. 보고서를 생성하기 위해 "Distribution: Profitability by Lead Studio and Genre", "Graph Builder: World and Domestic Gross by Genre" 및 "Bubble Plot: Theaters Open Week by Domestic Gross" 라는 테이블 스크립트를 실행합니다 .
3. 어느 창에서든 **파일 > 새로 만들기 > 대시보드**를 선택합니다 .

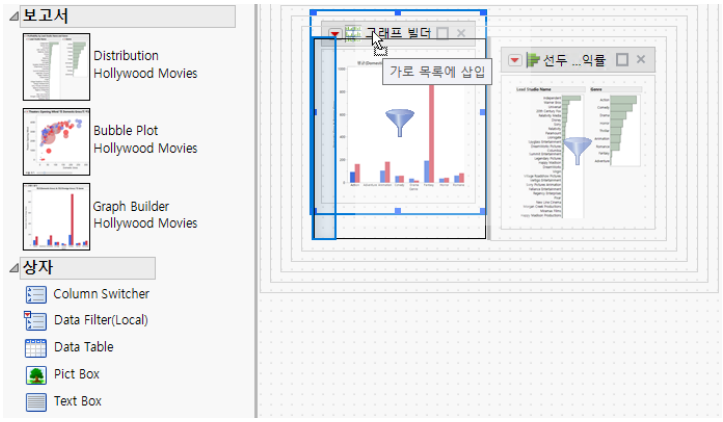
   일반적인 레이아웃을 위한 템플릿이 나타납니다 .
4. **2x1 대시보드** 템플릿을 선택합니다 .

   두 개의 보고서를 위한 공간이 있는 상자가 작업 공간에 나타납니다 .
5. " 보고서 " 목록에서 "Graph Builder" 및 "Distribution" 보고서 썸네일을 두 번 클릭하여 대시보드에 배치합니다 .

선택 필터 추가 및 사용

1. 왼쪽 보고서 (" 그래프 빌더 ") 의 가운데를 마우스 오른쪽 버튼으로 클릭하고 **선택 필터로 사용**을 선택합니다 .
2. 또 다른 보고서를 두 번째 선택 필터로 추가하기 위해 "Bubble Plot" 썸네일을 클릭하고 " 그래프 빌더 " 그래프 주위의 파란색 선으로 드래그합니다 .

그림 11.15 버블 그림을 그래프 빌더 그래프로 드래그



두 썸네일 모두에 필터 아이콘이 나타납니다 .

3. " 대시보드 빌더 " 의 빨간색 삼각형을 클릭하고 **미리보기 모드**를 선택합니다 .
4. 왼쪽 보고서에서 "Fantasy" 장르의 막대 중 하나를 선택합니다 .

그림 11.16 데이터 필터링

오른쪽 보고서를 보면 "Warner Bros" 및 "Independent" 스튜디오에서 판타지 영화를 가장 많이 제작했음을 알 수 있습니다. 아래쪽 보고서에서는 Harry Potter 가 가장 많은 수익을 낸 판타지 영화였음을 알 수 있습니다.

예 : JMP 쿼리 대시보드 및 추가기능 생성

JMP 쿼리 빌더를 사용하여 두 데이터 테이블의 일부분을 JMP 로 가져왔다고 가정해 보겠습니다. 대시보드에는 데이터 보고서가 표시됩니다. 데이터를 쿼리하고 업데이트된 보고서를 대시보드에 표시하는 추가기능을 생성하려고 합니다.

쿼리 작성

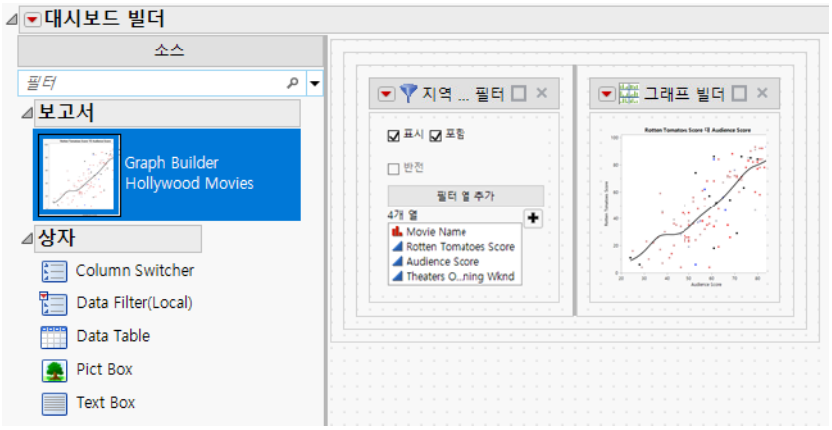
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hollywood Movies.jmp 를 엽니다.
2. **테이블 > JMP 쿼리 빌더**를 선택합니다.
3. **쿼리 생성**을 클릭합니다.
4. " 사용 가능한 열 " 목록에서 t1.Movie Name, t1.Rotten Tomatoes Score, t1.Audience Score 및 t1.Theaters Opening Wknd 를 두 번 클릭하여 " 포함된 열 " 탭에 추가합니다.
5. **쿼리 실행**을 클릭합니다.

   새 데이터 테이블에 쿼리한 데이터가 나타납니다. " 소스 " 테이블 스크립트를 확인합니다. 이 대시보드를 열면 JMP 는 이 " 소스 " 스크립트를 실행하여 Hollywood Movies.jmp 의 데이터를 쿼리합니다.
6. 새 데이터 테이블에서 **Graph Builder: Scores by Opening Weekend** 테이블 스크립트를 실행하여 보고서를 생성합니다.

대시보드 생성

1. 보고서 창에서 **파일 > 새로 만들기 > 대시보드**를 선택합니다 .
2. **빈 대시보드** 템플릿을 선택합니다 .
3. 왼쪽 페인에서 **Data Filter(Local)** 상자를 두 번 클릭하여 대시보드에 추가합니다 .
4. 왼쪽 페인에서 "Graph Builder" 보고서를 두 번 클릭하여 로컬 데이터 필터 옆에 배치합니다 .

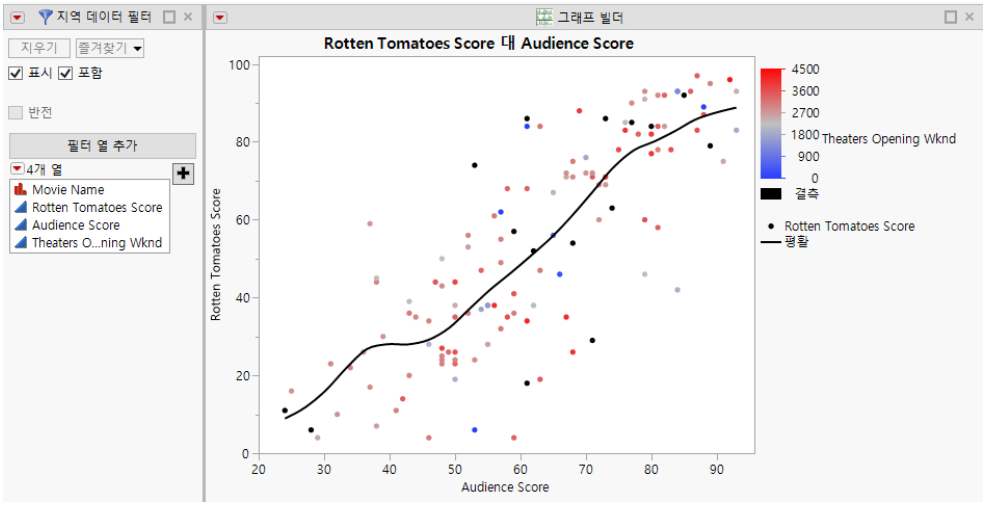
그림 11.17 Hollywood Movies 에 대한 대시보드



추가기능 생성 및 실행

1. " 대시보드 빌더 " 의 빨간색 삼각형을 클릭하고 **스크립트 저장 > 추가기능에**를 선택합니다 .
   추가기능 빌더가 열립니다 .
2. "추가기능 이름" 옆에 "Hollywood Movies Dashboard"(추가기능 파일 이름)를 입력합니다.
3. **메뉴 항목** 탭을 클릭하고 **메뉴 항목 이름** 옆에 "Hollywood Movies Dashboard"(추가기능 메뉴 항목의 이름) 를 입력합니다 .
4. **저장**을 클릭하여 추가기능을 데스크톱에 저장합니다 .
   추가기능이 저장되고 " 추가기능 " 메뉴에 설치됩니다 .
5. " 추가기능 빌더 " 창에서 **닫기**를 클릭합니다 .
6. JMP 메인 메뉴에서 **추가기능**을 선택하고 **Hollywood Movies Dashboard** 를 선택합니다 .
   Hollywood Movies.jmp 데이터 테이블이 쿼리되고, 로컬 데이터 필터와 업데이트된 보고서를 보여 주는 대시보드가 열립니다 .

그림 11.18 Hollywood Movies 추가기능



예 : 두 개의 데이터 테이블에서 대시보드 생성

여러 개의 데이터 테이블에서 대시보드의 그래프 또는 보고서를 생성할 수 있습니다. 대시보드의 첫 번째 그래프에 사용된 데이터 테이블은 기본적으로 현재 데이터 테이블로 간주됩니다. 따라서 사용자는 현재 데이터 테이블을 사용하여 그래프 또는 보고서를 유연하게 생성할 수 있습니다. 대시보드 작성자는 경로를 하드 코딩할 수도 있습니다. 이름을 지정하면 대시보드 작성자가 사용되는 데이터 테이블을 제어할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 연 후 "Bivariate" 스크립트를 실행합니다 .
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hollywood Movies.jmp 를 연 후 "Distribution" 스크립트를 실행합니다 .
3. 어느 JMP 창에서든 **파일 > 새로 만들기 > 대시보드**를 선택합니다 .
4. 대시보드 빌더에서 **2x1 대시보드** 템플릿을 선택합니다 .
5. 먼저 , "Distribution" 그래프를 두 번 클릭하여 대시보드에 추가합니다 .

   기본적으로 이 그래프는 현재 데이터 테이블에 해당하는 데이터 테이블을 사용합니다. 이 예에서는 Samples/Data 폴더에 저장된 Hollywood Movies.jmp 가 현재 데이터 테이블입니다 .
6. "Bivariate" 그래프를 두 번 클릭하여 대시보드에 추가합니다 .

   대시보드에서 첫 번째 그래프에 현재 데이터 테이블을 사용할지 여부 또는 데이터 테이블의 경로를 하드 코딩 (테이블을 이름으로 참조) 할지 여부를 지정하라는 메시지가 표시됩니다 .
7. **이름으로 테이블 참조**를 선택하여 그래프를 생성할 때 항상 지정한 데이터 테이블이 사용되도록 합니다 .
8. **확인**을 클릭합니다 .

이 예에서는 열려 있는 현재 데이터 테이블에 상관없이 항상 Samples/Data 폴더의 Hollywood Movies.jmp 가 첫 번째 그래프의 현재 데이터 테이블로 사용됩니다 .

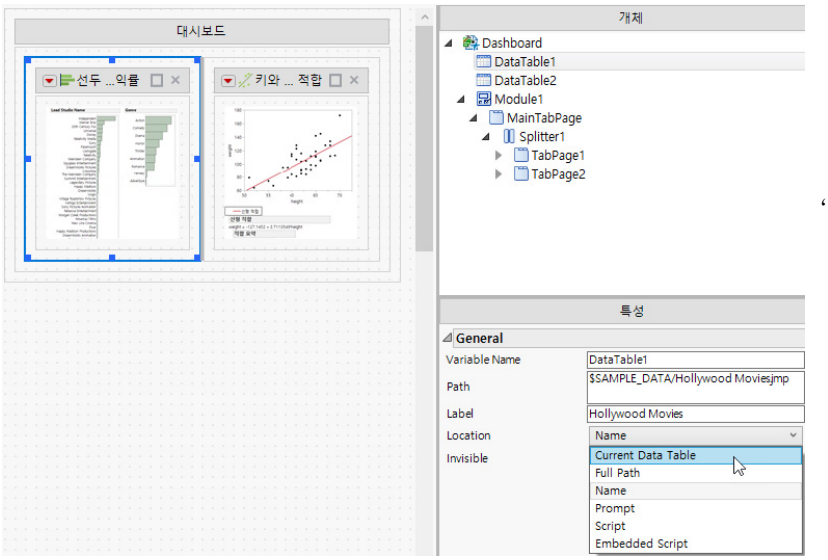
팁 : 대시보드가 실행될 때 명명된 데이터 테이블을 찾을 수 없는 경우도 있습니다 . 이는 특히 대시보드가 다른 컴퓨터로 이동된 경우에 주로 발생합니다 . 대시보드를 실행하기 전에 데이터 테이블을 열거나 , 대시보드를 실행한 후 해당 메시지가 표시될 때 데이터 테이블을 선택할 수 있습니다 .

데이터 테이블의 위치를 변경하려면

데이터 테이블(현재 데이터 테이블 또는 이름)의 위치를 할당한 후 대시보드의 "개체" 및 "특성" 페인에서 설정을 변경할 수 있습니다.

1. " 대시보드 빌더 " 의 빨간색 삼각형을 클릭하고 **특성 표시**를 선택합니다 .

그림 11.19 개체 및 특성 페인



2. " 개체 " 페인에서 **DataTable1** 을 선택합니다 .
3. " 특성 " 페인에서 위치를 그림 11.19 에 표시된 것과 같이 변경합니다 .

   예를 들어 "Location" 에서 "Name" 을 "Current Data Table" 로 변경하면 대시보드의 선택된 그래프가 현재 데이터 테이블을 기준으로 하게 됩니다 .

참고 : "Location" 값 "Current Data Table" 은 대시보드를 실행할 때만 적용됩니다 . 대시보드를 편집할 때는 대시보드 빌더에서 데이터 테이블의 위치로 전체 경로가 사용됩니다 .

12 장

메뉴 및 도구 모음 개인 설정
JMP 사용자 정의

이 장에서는 JMP 메뉴 및 도구 모음을 사용자 정의하여 필요한 명령만 표시하는 방법을 설명합니다. 예를 들어 SAS를 사용하지 않는 경우 "파일" 메뉴에서 SAS 옵션을 제거할 수 있습니다. 또는 "스크립트 실행" 명령에 바로 가기 키를 할당할 수 있습니다.

JMP를 사용자 정의하면 JMP를 사용자 그룹별로 특정 관심사에 맞게 설정할 수 있습니다. 예를 들어 어떤 그룹에서 실험 설계를 사용하지 않는 경우에는 JMP 메뉴에서 "DOE"를 제거할 수 있습니다.

참고 : JMP 16 구독으로 업그레이드할 때 환경 설정 변경 사항이 유지됩니다 .

목차

Windows 에서 JMP 도구 모음 및 메뉴 개인 설정 599
도구 모음 개인 설정 600
메인 메뉴 막대 개인 설정 603
메뉴 항목 및 도구 모음 버튼 생성 604
사용자 정의된 도구 모음 및 버튼에 대한 사용자 권한 설정 607
도구 모음 재배열 608
메뉴, 메뉴 항목, 도구 모음 및 버튼 복사하여 붙여넣기 609
사용자 메뉴, 메뉴 항목 및 버튼 재배열 610
사용자 항목 삭제 612
사용자 정의 항목 표시 또는 숨기기 613
사용자 정의 가져오기 614
JMP에서 사용자 정의 제거 614
macOS 에서 JMP 도구 모음 개인 설정 615
macOS 에서 JMP 메뉴 개인 설정 617
추가기능에 대한 JMP 메뉴 및 도구 모음 사용자 정의 618

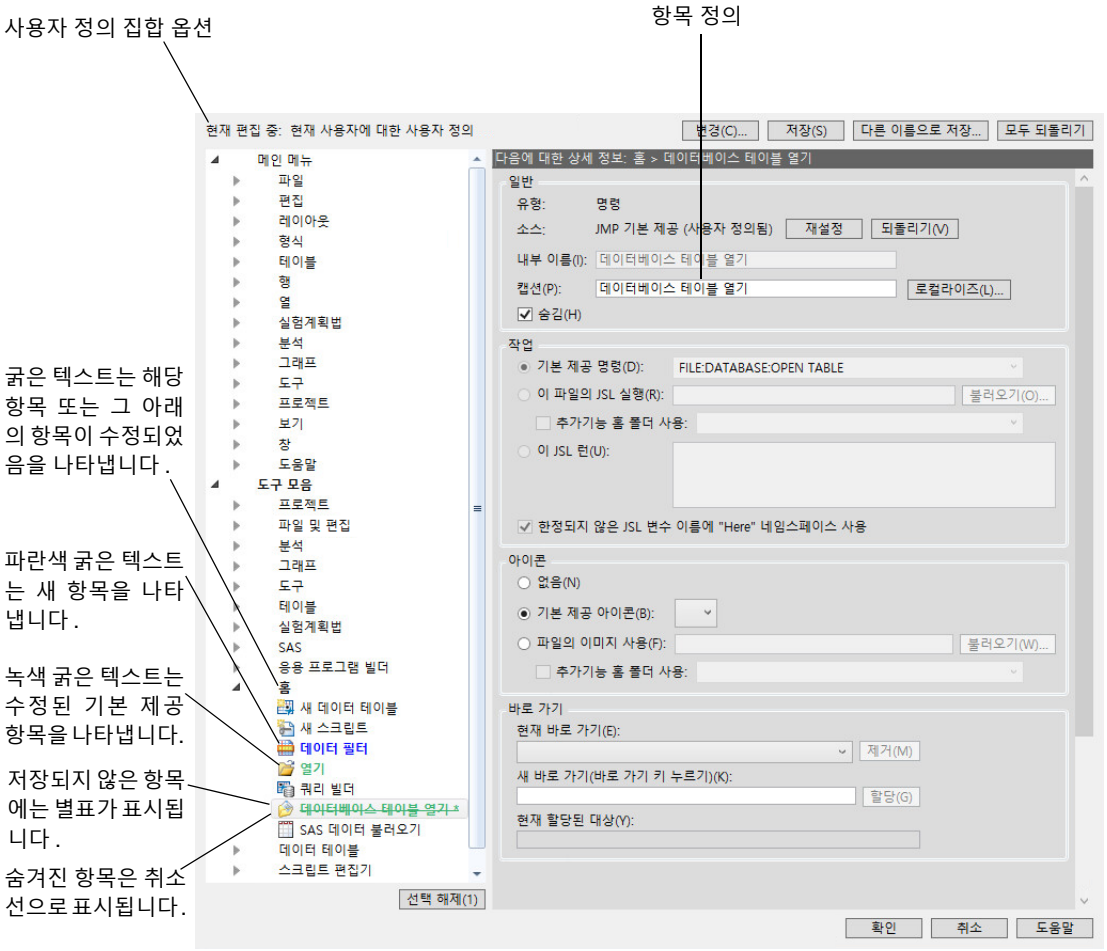
Windows 에서 JMP 도구 모음 및 메뉴 개인 설정

도구 모음은 명령을 실행하는 버튼과 버튼을 정리하는 데 도움이 되는 구분선으로 구성됩니다. Windows에서는 여러 가지 방법으로 도구 모음을 개인 설정할 수 있습니다. 예를 들어 도구 모음 또는 버튼을 생성하고, 이름을 바꾸고, 숨기고, 수정할 수 있습니다. 생성한 도구 모음 및 버튼을 재배열하거나 삭제할 수도 있습니다.

메뉴는 명령, 하위 메뉴 및 구분선으로 구성됩니다. 도구 모음 및 버튼과 마찬가지로 메뉴를 생성하고, 이름을 바꾸고, 숨기고, 수정할 수 있습니다. 생성한 메뉴 및 메뉴 항목을 재배열하거나 삭제할 수도 있습니다.

메뉴 편집기에서 도구 모음과 메뉴를 개인 설정할 수 있습니다. 수정된 항목의 이름은 적절한 형식으로 표시되어, 생성 또는 수정되거나 숨겨진 항목과 저장되지 않은 항목을 나타냅니다.

그림 12.1 메뉴 편집기 옵션



사용자 정의 집합 옵션

수정할 사용자 정의 집합

현재 사용자 현재 사용자만 변경 사항을 볼 수 있습니다 . 해당 사용자 정의 파일의 이름은 usercust.jmpcust 로 지정됩니다 . 이 파일을 직접 수정하면 예기치 않은 결과가 발생할 수 있습니다 .

모든 사용자 컴퓨터에서 JMP 를 여는 모든 사용자가 변경 사항을 볼 수 있습니다 . 해당 사용자 정의 파일의 이름은 admincust.jmpcust 로 지정됩니다 .

JMP 추가기능 선택한 JMP 추가기능을 사용하는 사용자와 " 시작 집합 " 으로 **JMP 추가기능 사용자 정의**를 선택하는 사용자만 변경 사항을 볼 수 있습니다. 비활성화된 추가기능에 대한 사용자 정의 집합을 수정할 경우 해당 추가기능을 활성화하기 전까지는 변경 사항이 표시되지 않습니다 .

기타 파일 선택한 .jmpcust 파일이 사용자 정의 집합으로 선택될 때 변경 사항이 표시됩니다 . 이 파일은 텍스트 편집기에서 생성합니다 .

새로 만들기 이 새 사용자 정의가 선택될 때 변경 사항이 표시됩니다 . 변경 후에는 **저장**을 클릭하여 변경 사항을 새 파일 이름으로 선택한 폴더에 저장합니다 .

시작 집합

JMP 기본 제공 항목 기본 JMP 도구 모음 및 메뉴를 선택한 사용자 정의 집합의 기초로 사용합니다 . 이 옵션은 모든 사용자 정의 집합에 적용되며 선택 취소할 수 없습니다 .

모든 사용자의 사용자 정의 " 모든 사용자 " 사용자 정의 집합을 선택한 사용자 정의 집합의 기초로 사용합니다 . 이 옵션은 " 기타 파일 " 또는 " 새로 만들기 " 사용자 정의 집합에 사용할 수 있습니다 . " 현재 사용자 " 사용자 정의 집합에서는 이 옵션을 선택 취소할 수 없습니다 .

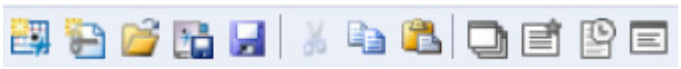
JMP 추가기능 사용자 정의 "JMP 추가기능" 사용자 정의를 선택한 사용자 정의 집합의 기초로 사용합니다 . 추가기능 사용자 정의는 일반적으로 추가기능과 함께 설치됩니다 . 이 옵션은 " 모든 사용자 " 를 제외한 모든 사용자 정의 집합에 사용할 수 있습니다 .

현재 사용자의 사용자 정의 " 현재 사용자 " 사용자 정의 집합을 선택한 사용자 정의 집합의 기초로 사용합니다 . 이 옵션은 " 기타 파일 " 및 " 새로 만들기 " 사용자 정의 집합에만 사용할 수 있습니다 .

도구 모음 개인 설정

JMP에는 새 버튼을 추가할 수 있는 도구 모음이 포함되어 있으며, 사용자가 직접 도구 모음을 생성할 수도 있습니다. 기본적으로 JMP 창 유형에 따라 표시되는 도구 모음이 결정됩니다. 예를 들어 모든 창에는 "파일 및 편집" 도구 모음이 있습니다. 데이터 테이블에서는 "데이터 테이블" 도구 모음도 표시됩니다.

프로젝트 프로젝트에 대한 아이콘을 표시합니다 .



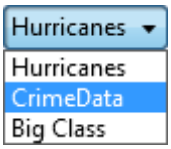
파일 및 편집 " 파일 " 및 " 편집 " 메뉴에 있는 명령에 대한 아이콘을 표시합니다 .

분석 " 분석 " 메뉴에 있는 일반적인 명령에 대한 아이콘을 표시합니다 .

그래프 " 그래프 " 메뉴에 있는 명령에 대한 아이콘을 표시합니다 .

도구 클릭하여 커서처럼 사용할 수 있는 도구의 아이콘을 표시합니다 . 일부 창에서는 이 도구 모음이 숨겨집니다 . 대신 파란색 선이 표시됩니다 . 파란색 선을 마우스로 가리키면 이 도구 모음이 표시됩니다 .

데이터 테이블 목록 열려 있는 데이터 테이블의 목록을 표시합니다 . 이 목록에서 데이터 테이블을 선택하면 해당 데이터 테이블이 현재 테이블이 됩니다 . 현재 테이블 창이 반드시 맨 앞의 창인 것은 아닙니다 . 테이블 (또는 임의의 창) 을 맨 앞으로 가져오려면 " 창 " 메뉴의 목록에서 해당 이름을 선택합니다 .



테이블 " 테이블 " 메뉴의 명령에 대한 아이콘을 표시합니다 .

DOE "DOE" 메뉴의 명령에 대한 아이콘을 표시합니다 .

SAS SAS 데이터 및 폴더 액세스 / 불러오기 , SAS 쿼리 빌더 열기 , 새 SAS 프로그램 생성 , SAS 로 데이터 전송, SAS 로그 창 및 출력 창 열기, 서버 연결 보기 등을 위한 아이콘을 표시합니다.

응용 프로그램 빌더 응용 프로그램 빌더의 명령에 대한 아이콘을 표시합니다 .

홈 "파일" 메뉴의 자주 사용하는 명령에 대한 아이콘을 표시합니다.

데이터 테이블 데이터 테이블에서 " 데이터 필터 ", " 모형 적합 " 및 " 재코딩 " 창을 여는 등의 일반적인 작업을 수행하기 위한 아이콘을 표시합니다 .

스크립트 편집기 스크립트 실행 , 디버깅 및 형식 조정을 위한 아이콘을 표시합니다 .

보고서 로컬 데이터 필터 및 열 전환기를 열거나 스크립트를 데이터 테이블에 저장하기 위한 아이콘을 표시합니다 .

기본 도구 모음에는 포함되어 있지 않지만 자주 사용하는 명령을 표시하기 위한 고유한 도구 모음을 생성할 수 있습니다. 버튼에 명령을 할당하고 바로 가기 키, 아이콘, JSL(JMP 스크립트 언어) 스크립트 등의 다른 특성을 추가할 수 있습니다.

기존 도구 모음을 기초로 도구 모음을 생성하려면 기존 도구 모음의 복사본을 생성한 후 설정을 필요한 대로 변경합니다. 자세한 내용은 "메뉴, 메뉴 항목, 도구 모음 및 버튼 복사하여 붙여넣기"(609 페이지)에서 확인하십시오.

1 단계 : 도구 모음 생성

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   왼쪽에 도구 모음 목록이 표시됩니다 .
2. (선택 사항) 이 사용자 정의를 볼 수 있는 사용자를 제어하려면 사용자 정의 집합을 변경합니다 . 자세한 내용은 " 사용자 정의 집합 옵션 "(600 페이지) 에서 확인하십시오 .
3. 메뉴 편집기가 아직 열려 있는 상태에서 도구 모음 중 하나 (예 : **파일 및 편집**) 를 마우스 오른쪽 버튼으로 클릭합니다 .Not

   모든 새 도구 모음은 " 도구 모음 " 목록의 끝에 나타나므로 임의의 도구 모음을 선택합니다 .
4. **새 도구 모음**을 선택합니다 .

   목록 끝에 도구 모음이 삽입됩니다 . 모든 도구 모음에는 버튼이 하나 이상 있어야 하므로 이 도구 모음에는 제목 없는 버튼이 하나 포함되어 있습니다 .

2 단계 : 캡션 및 내부 이름 지정

도구 모음에는 몇 가지 기본적인 특성이 있습니다 .

- 캡션 : **보기 > 도구 모음** 목록에 표시되며 이를 통해 도구 모음을 표시하거나 숨길 수 있습니다 . 각 도구 모음에 고유한 이름을 지정해야 합니다 . 메뉴 편집기를 닫았다가 다시 열면 JMP 는 동일한 이름의 도구 모음을 병합합니다 .
- 내부 이름 (대 / 소문자 구분 안 함): JMP 는 항목의 위치를 캡션이 아니라 내부 이름으로 식별합니다 .

이러한 도구 모음 특성을 지정하려면 다음 단계를 수행하십시오 .

1. 메뉴 편집기가 아직 열려 있는 상태에서 도구 모음을 선택합니다 .

   오른쪽에 " 일반 " 특성이 나타납니다 .
2. **내부 이름** 옆에 내부 이름을 입력합니다 .
3. **캡션** 상자에 도구 모음 또는 버튼의 고유 이름을 입력합니다 .
4. (선택 사항) 캡션에 대한 번역을 지정하려면 다음을 수행합니다 .
   - 사용자 정의하려는 항목에 대한 **로컬라이즈** 버튼을 클릭합니다 .
   - 언어를 선택합니다 .
   - **텍스트** 옆에 번역을 입력합니다 .
   - **확인**을 클릭합니다 .
5. (선택 사항) 생성 후 도구 모음을 숨김 상태로 유지하려면 **숨김**을 클릭합니다 .
6. " 메뉴 항목 및 도구 모음 버튼 생성 "(604 페이지) 에 설명된 대로 제목 없는 버튼을 사용자 정의합니다 .
7. **저장**을 클릭하여 변경 사항을 저장합니다 .

메인 메뉴 막대 개인 설정

JMP의 메인 메뉴 표시줄에 메뉴 항목을 추가할 수 있습니다. 이러한 메뉴는 **파일**, **편집** 및 **테이블** 같은 메뉴의 앞 또는 뒤에 표시됩니다. 모든 메뉴에는 명령이 하나 이상 있어야 하므로 추가한 메뉴에 제목 없는 명령이 자동으로 삽입됩니다.

기존 메뉴를 기초로 메인 메뉴를 생성하려면 기존 메뉴의 복사본을 생성한 후 설정을 필요한 대로 변경합니다. 자세한 내용은 " 메뉴, 메뉴 항목, 도구 모음 및 버튼 복사하여 붙여넣기 "(609 페이지) 에서 확인하십시오.

메인 메뉴를 생성하려면

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   왼쪽에 " 메인 메뉴 " 목록이 표시됩니다 .
2. (선택 사항) 이 사용자 정의를 볼 수 있는 사용자를 제어하려면 사용자 정의 집합을 변경합니다 . 자세한 내용은 " 사용자 정의 집합 옵션 "(600 페이지) 에서 확인하십시오 .

3. 새 메뉴를 추가할 위치 옆의 메뉴를 마우스 오른쪽 버튼으로 클릭합니다 .
4. **앞에 삽입** 또는 **뒤에 삽입**을 선택합니다 .

   제목 없는 메뉴와 메뉴 항목이 목록에 추가됩니다 .
5. "2 단계 : 캡션 및 내부 이름 지정 "(603 페이지) 을 완료하여 메인 메뉴의 이름을 바꿉니다 .
6. " 메뉴 항목 및 도구 모음 버튼 생성 "(604 페이지) 에 설명된 대로 메뉴 항목을 사용자 정의합니다 .
7. **저장**을 클릭하여 변경 사항을 저장합니다 .

메뉴 항목 및 도구 모음 버튼 생성

메뉴는 다음과 같은 메뉴 항목으로 구성됩니다 .

- 명령 : " 데이터 테이블 " 및 " 스크립트 " 같은 명령을 실행하기 위해 클릭하는 항목입니다 .
- 하위 메뉴 : 추가 메뉴 옵션을 표시하기 위해 클릭하는 메뉴 항목입니다 . 예를 들어 최상위 " 파일 " 메뉴에 있는 " 새로 만들기 " 가 하위 메뉴에 해당합니다 . " 새로 만들기 " 는 " 데이터 테이블 " 및 " 스크립트 " 같은 명령을 포함하기 때문에 하위 메뉴입니다 .
- 구분선 (): 명령 및 하위 메뉴를 나누거나 그룹화하는 선입니다 .

도구 모음의 버튼도 일종의 명령으로 간주되므로 새 도구 모음 버튼을 생성하려면 다음 단계를 완료해야 합니다.

기존 항목을 기초로 메뉴 항목 또는 버튼을 생성하려면 기존 항목을 복사한 후 필요한 대로 설정을 변경합니다. 자세한 내용은 "메뉴, 메뉴 항목, 도구 모음 및 버튼 복사하여 붙여넣기"(609 페이지)에서 확인하십시오.

1 단계 : 새 메뉴 항목 또는 버튼 생성

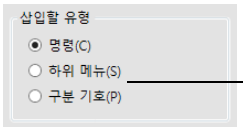
1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   메뉴 편집기가 나타납니다 .
2. (선택 사항) 이 사용자 정의를 볼 수 있는 사용자를 제어하려면 사용자 정의 집합을 변경합니다 . 자세한 내용은 " 사용자 정의 집합 옵션 "(600 페이지) 에서 확인하십시오 .
3. 메뉴 항목 또는 버튼을 추가할 위치를 마우스 오른쪽 버튼으로 클릭합니다 .

   가능한 위치의 목록이 표시됩니다 .
4. **앞에 삽입** 또는 **뒤에 삽입**을 선택합니다 .

   " 유형 지정 " 창이 나타납니다 .

그림 12.2 유형 지정 창



" 하위 메뉴 " 는 버튼의 경우 에는 표시되지 않습니다 .

5. 다음 중 하나를 수행합니다 .
   - 버튼을 생성하려면 **명령**을 선택합니다 .
   - 메뉴 내의 메뉴를 생성하려면 **하위 메뉴**를 선택합니다 . 버튼의 경우에는 사용할 수 없습니다 .
   - 메뉴 항목 또는 도구 모음 버튼 사이에 구분선을 생성하려면 **구분 기호**를 선택합니다 .
6. **확인**을 클릭합니다 .

   제목 없는 버튼 , 하위 메뉴 또는 구분선이 새로 추가됩니다 .

2 단계 : 캡션 , 툴팁 및 내부 이름 지정

버튼에는 몇 가지 기본적인 특성이 있습니다 .

- 캡션 : " 보기 " > " 도구 모음 " 목록에 표시되며 이를 통해 도구 모음을 표시하거나 숨길 수 있습니다 . 각 도구 모음에 고유한 이름을 지정해야 합니다 . 메뉴 편집기를 닫았다가 다시 열면 JMP 는 동일한 이름의 도구 모음을 병합합니다 .
- 툴팁 : 메뉴 항목 또는 버튼 위에 커서를 놓으면 표시됩니다 .
- 내부 이름 (대 / 소문자 구분 안 함): JMP 는 항목의 위치를 캡션이 아니라 내부 이름으로 식별합니다 .

하위 메뉴 항목에는 툴팁이 없다는 점만 제외하고 버튼과 동일한 " 일반 " 특성이 포함됩니다.

이러한 특성을 지정하려면 다음 단계를 수행하십시오 .

1. 메뉴 편집기가 아직 열려 있는 상태에서 메뉴 항목 또는 버튼을 선택합니다 .

   오른쪽에 " 일반 " 특성이 나타납니다 .
2. **내부 이름** 옆에 내부 이름을 입력합니다 .
3. " 캡션 " 상자에 메뉴 항목 또는 버튼의 고유 이름을 입력합니다 .
4. " 팁 " 상자에 버튼에 대한 설명을 입력합니다 .
5. (선택 사항) 캡션 또는 팁에 대한 번역을 지정하려면 다음을 수행합니다 .
   - 사용자 정의하려는 항목에 대한 **로컬라이즈** 버튼을 클릭합니다 .
   - 언어를 선택합니다 .
   - 텍스트 옆에 번역을 입력합니다 .
   - **확인**을 클릭합니다 .
6. 다음 절차에 설명된 대로 제목 없는 버튼 또는 메뉴 항목을 사용자 정의합니다 .

3 단계 : 기능 할당

메뉴 항목 또는 버튼을 클릭하면 사전 정의된 명령이 실행되거나 JSL 스크립트가 실행됩니다. 스크립트는 별도의 파일에 저장되어 있을 수도 있고, 메뉴 편집기의 " 이 JSL 실행 " 영역에 사용자가 직접 JSL 을 입력할 수도 있습니다. 사용자가 네트워크 등을 통해 중앙 위치에 액세스할 수 있는 경우에는 일반적으로 해당 위치에서 스크립트를 실행하도록 합니다. 이렇게 하면 사용자

가 업데이트된 스크립트에 간편하게 액세스할 수 있습니다. 그렇지 않으면 메뉴 항목 또는 버튼 정의에 JSL 을 저장합니다.

메뉴 항목 또는 버튼을 추가기능 JSL 스크립트와 연결하려면 다음 두 가지 방법을 사용합니다.

- **이 파일의 JSL 실행**을 선택한 후 외부 파일을 찾아 불러옵니다 .
- JSL 스크립트의 상대 경로를 입력하고 , **추가기능 홈 폴더 사용** 목록에서 추가기능을 선택합니다. 예를 들어 다음 명령은 선택한 추가기능의 mds_application.jsl 스크립트를 실행합니다.

  ```
  $ADDIN_HOME(com.jmp.mdswithr)/mds_application.jsl
  ```

또한 내부 또는 외부 JSL 스크립트에 네임스페이스가 포함된 경우에는 **한정되지 않은 JSL 변수 이름에 "Here" 네임스페이스 사용**을 선택할 수 있습니다. 네임스페이스에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오.

이 기능을 할당하려면 다음 단계를 수행하십시오 .

1. 메뉴 편집기가 아직 열려 있는 상태에서 수정하려는 메뉴 항목 또는 버튼을 선택합니다 .

   오른쪽에 " 작업 " 특성이 표시됩니다 .
2. 실행하려는 작업을 선택합니다 .
3. 내부 JSL 스크립트를 실행하려면 **이 JSL 실행** 영역의 자리 표시자 텍스트 print(" 구현되지 않음 "); 를 삭제하고 JSL 을 입력합니다 .
4. (선택 사항) 내부 또는 외부 JSL 스크립트에 네임스페이스가 포함되어 있으면 **한정되지 않은 JSL 변수 이름에 "Here" 네임스페이스 사용**을 선택합니다 .
5. **저장**을 클릭하여 변경 사항을 저장합니다 .

4 단계 : 메뉴 항목 또는 버튼에 아이콘 표시 (선택 사항)

새 메뉴 항목의 항목 옆에는 아이콘이 없습니다. 새 버튼은 도구 모음에 파란색 사각형으로 표시됩니다. 메뉴 항목 또는 버튼에 설명적인 아이콘을 할당할 수 있습니다. 컴퓨터 또는 추가기능 폴더에 있는 그래픽을 사용할 수 있으며, .ico, .png, .jpg 및 .bmp 그래픽이 지원됩니다.

버튼에 아이콘을 할당하는 것은 특히 중요합니다. 그렇지 않으면 사용자가 버튼의 툴팁을 통해서만 버튼의 기능을 알 수 있습니다.

메뉴 항목 또는 버튼에 아이콘을 표시하려면 다음 단계를 수행하십시오 .

1. 메뉴 편집기가 아직 열려 있는 상태에서 메뉴 항목 또는 버튼을 선택합니다 .

   오른쪽에 " 아이콘 " 특성이 표시됩니다 .
2. 다음 옵션 중 하나를 선택합니다 .

   없음 아이콘 대신 파란색 사각형을 표시하려면 이 옵션을 선택합니다 .

   기본 제공 아이콘 : JMP 에서 제공되는 아이콘을 표시하려면 이 옵션을 선택한 후 목록에서 아이콘을 선택합니다 .

   파일의 이미지 사용 생성한 이미지를 표시하려면 이 옵션을 선택합니다 . **불러오기**를 클릭하고 그래픽을 선택한 후 **확인**을 클릭합니다 .

추가기능 홈 폴더 사용 메뉴 항목 왼쪽에 추가기능의 아이콘을 표시하려면 이 옵션을 선택합니다 . 체크박스를 선택한 후 추가기능을 선택하고 **불러오기**를 클릭하여 아이콘을 선택합니다 . 그러면 해당 이미지 경로가 **파일의 이미지 사용** 필드에 표시됩니다 . 이미지는 ICO, PNG, JPG 또는 BMP 그래픽일 수 있습니다 .

3. **저장**을 클릭하여 변경 사항을 저장합니다 .

5 단계 : 바로 가기 키 할당 (선택 사항)

바로 가기 키는 메뉴 항목을 선택하거나 도구 모음의 버튼을 클릭하지 않고도 작업을 실행할 수 있도록 합니다. 이 옵션은 버튼의 도구 모음이 표시되어 있지 않은 상태에서 해당 작업을 실행하려는 경우에도 유용합니다.

바로 가기는 Ctrl, Ctrl+Shift 또는 Ctrl+Alt로 시작하고 키보드에 나타나는 숫자, 문자 또는 기호로 끝납니다. F1 같은 기능 키도 지원됩니다. Ctrl + 문자 조합과 기능 키 바로 가기는 대부분 JMP에서 이미 할당되어 있지만 원하는 경우 사용자 정의 버튼에 이러한 바로 가기를 재할당할 수 있습니다.

바로 가기 키는 메뉴에서 각 메뉴 항목 옆에 표시됩니다.

바로 가기 키를 할당하려면 다음 단계를 수행하십시오 .

1. 메뉴 편집기가 아직 열려 있는 상태에서 메뉴 항목 또는 버튼을 선택합니다 .

   오른쪽에 " 바로 가기 " 특성이 표시됩니다 .
2. 다음 중 하나를 수행합니다 .
   - 새 바로 가기를 할당하려면 **새 바로 가기** 영역을 클릭하고 바로 가기 키를 누릅니다 . 해당 키 입력이 **현재 할당된 대상** 상자에 표시되면 다른 바로 가기 키 조합을 누릅니다 . 현재 키 입력을 삭제할 필요 없이 바로 다른 바로 가기 키를 누르면 됩니다 .
   - 이전에 지정한 바로 가기를 변경하려면 **제거**를 클릭하고 **새 바로 가기** 영역을 클릭한 다음 바로 가기 키를 누릅니다 .
3. **할당**을 클릭합니다 .

   현재 바로 가기 목록에 해당 바로 가기가 표시됩니다 .
4. **저장**을 클릭하여 변경 사항을 저장합니다 .

바로 가기를 제거하려면 바로 가기를 선택하고 **제거**를 클릭합니다.

사용자 정의된 도구 모음 및 버튼에 대한 사용자 권한 설정

JMP 메뉴 편집기를 사용하여 사용자 정의된 도구 모음 및 버튼을 볼 수 있는 사용자를 지정합니다. 변경 사항은 .jmpcust 확장자를 사용하는 사용자 정의 집합이라는 텍스트 파일로 저장됩니다. 기본적으로 사용자 정의는 현재 사용자에게만 표시되므로, 해당 컴퓨터에 로그온하여 JMP를 여는 다른 사용자는 현재 사용자의 개인 도구 모음 및 메뉴를 볼 수 없습니다.

모든 사용자 정의 집합에는 기본 제공 JMP 버튼 및 메뉴가 포함됩니다. 다른 사용자 정의 집합을 기초로 새 사용자 정의 집합을 만들 수도 있습니다. 예를 들어 사용자가 수정한 사항과 함께 JMP 추가기능 메뉴 및 도구 모음을 표시하는 새 사용자 정의 집합을 생성할 수 있습니다.

현재 사용자 사용자 정의 집합을 수정하면 JMP 에서는 해당 집합의 백업 파일을 생성합니다. 이 파일은 Windows Users 폴더의 JMP 또는 JMPPro 폴더에 배치됩니다.

C:/Users/<user name>/AppData/Roaming/SAS/JMP/16/

모든 사용자 사용자 정의 집합을 수정하면 JMP 에서는 해당 집합의 백업 파일을 생성합니다. 이 파일은 Windows All Users 폴더의 JMP 또는 JMPPro 폴더에 배치됩니다.

C:/Users/All Users/SAS/JMP/16/

폴더를 찾아볼 때 All Users 폴더는 표시되지 않으므로 Windows 탐색기 주소 필드에 경로를 직접 입력해야 합니다.

참고 : Windows 탐색기에서 사용자 정의 집합 파일을 찾으려면 Windows 탐색기의 " 폴더 옵션 " 에서 숨겨진 파일을 모두 표시하십시오 . 자세한 내용은 운영 체제 설명서에서 확인하십시오 .

사용자 정의 집합을 변경하려면 다음 단계를 수행하십시오 .

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   메뉴 편집기가 나타납니다 .
2. **변경**을 클릭합니다 .

   사용자 정의 집합을 변경하려고 할 때 선택한 집합에 대한 변경 사항이 저장되어 있지 않으면 변경 사항을 저장하라는 메시지가 표시됩니다 . 이 경우 **변경 사항 저장**을 클릭합니다 .
3. " 수정할 사용자 정의 집합 " 영역에서 사용자 정의 집합을 선택합니다 . 기본적으로 변경 사항은 **현재 사용자** 사용자 정의 집합에만 적용됩니다 . " 사용자 정의 집합 옵션 "(600 페이지) 에서는 해당 옵션을 설명합니다 .
4. (선택 사항) 선택한 사용자 정의 집합의 기초로 사용된 집합을 변경하려면 " 시작 집합 " 영역에서 집합을 선택합니다 . **모든 사용자** 사용자 정의 집합에는 이 기능을 사용할 수 없습니다 . " 사용자 정의 집합 옵션 "(600 페이지) 에서는 해당 옵션을 설명합니다 .
5. **확인**을 클릭합니다 .

도구 모음 재배열

Windows 에서는 다음과 같은 두 가지 방법으로 도구 모음을 재배열할 수 있습니다.

- 현재 사용자 환경에서 도구 모음을 빠르게 재배열하려면 도구 모음을 인접한 도구 모음의 위 , 아래 , 왼쪽 또는 오른쪽으로 드래그합니다 .
- 도구 모음을 드래그하지 않고 새 위치를 지정하려면 도구 모음을 마우스 오른쪽 버튼으로 클릭하고 " 위치 " 메뉴에서 위치를 선택합니다 . 이 방법을 사용하면 도구 모음을 창의 상단 또는 하단으로 이동할 수 있습니다 . 도구 모음을 창의 왼쪽 또는 오른쪽에 세로로 배치할 수도 있습니다 .

일부 창에서는 이 도구 모음이 기본적으로 숨겨집니다 . 새 위치를 지정하면 해당 유형의 모든 창에서 같은 도구 모음이 표시됩니다 . Windows 의 경우 , 이 도구 모음을 항상 표시하도록 도구 모음 환경 설정을 변경할 수 있습니다 . 자세한 내용은 "JMP 환경 설정 " 장의 "Windows 환경 설정 "(649 페이지) 에서 확인하십시오 .

도구 모음을 드래그하려면

1. 이동 커서가 나타날 때까지 도구 모음의 왼쪽 모퉁이를 가리킵니다 .
2. 도구 모음을 새 위치로 드래그합니다 . 도구 모음이 너무 넓어서 완전히 표시되지 않으면 그림 12.3 에 표시된 것과 같이 화살표를 클릭하여 모든 버튼이 표시되도록 합니다 .

그림 12.3 도구 모음 확장

도구 모음의 위치를 조정하려면

1. 이동하려는 도구 모음을 마우스 오른쪽 버튼으로 클릭합니다 .
2. **위치**를 선택한 후 새 위치를 선택합니다 .

참고 : 도구 모음을 왼쪽 , 오른쪽 또는 아래쪽 위치로 이동한 후에 동일한 위치에 다른 도구 모음을 추가할 수 있습니다 . 예를 들어 **파일 및 편집** 도구 모음을 JMP 창의 오른쪽에 재배치할 수 있습니다 . 오른쪽에 다른 도구 모음을 추가하려면 도구 모음을 마우스 오른쪽 버튼으로 클릭하고 추가할 도구 모음을 선택합니다 .

메뉴 , 메뉴 항목 , 도구 모음 및 버튼 복사하여 붙여넣기

기존 항목을 복사하여 메뉴 또는 도구 모음에 새 항목을 간편하게 생성할 수 있습니다. 이 옵션을 사용하면 새 항목에서 가능한 일부 설정만 변경할 수 있습니다.

메뉴 및 메뉴 항목을 복사하여 붙여넣으려면

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   메뉴 편집기가 나타납니다 .
2. (선택 사항) 이 사용자 정의를 볼 수 있는 사용자를 제어하려면 사용자 정의 집합을 변경합니다 . 자세한 내용은 " 사용자 정의 집합 옵션 "(600 페이지) 에서 확인하십시오 .
3. 복사할 메뉴 또는 메뉴 항목을 마우스 오른쪽 버튼으로 클릭하고 **복사**를 선택합니다 .
4. 해당 항목을 표시할 상위 메뉴 또는 메뉴 항목이나 그 앞뒤의 메뉴 또는 메뉴 항목을 마우스 오른쪽 버튼으로 클릭하고 **붙여넣기**를 선택합니다 .

   다음 중 하나가 발생합니다 .

   – 가능한 위치의 목록이 표시됩니다 .

– 선택한 항목 아래에 해당 메뉴 또는 메뉴 항목이 표시됩니다. 이 경우 7단계로 건너뜁니다.

5. 다음 중 하나를 수행하여 메뉴의 위치를 선택합니다.
   – 선택한 메뉴 **앞에** 항목을 붙여넣으려면 **~ 앞에 붙여넣기**를 선택합니다.
   – 선택한 메뉴 **뒤에** 메뉴를 붙여넣으려면 **~ 뒤에 붙여넣기**를 선택합니다.
   – 선택한 메뉴 **안에** 메뉴를 붙여넣으려면 **~ 에 붙여넣기**를 선택합니다. 이 옵션은 메뉴를 새 위치로 선택한 경우에만 사용할 수 있습니다.
   – 작업을 취소하려면 **취소**를 선택합니다.

   지정한 대로 항목이 표시됩니다.
6. 항목을 수정합니다.
7. **저장**을 클릭하여 변경 사항을 저장합니다.

항목을 수정하는 방법에 대한 자세한 내용은 "메인 메뉴 막대 개인 설정"(603 페이지) 및 "메뉴 항목 및 도구 모음 버튼 생성"(604 페이지)에서 확인하십시오.

도구 모음 및 버튼을 복사하여 붙여넣으려면

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다.

   메뉴 편집기가 나타납니다.
2. (선택 사항) 이 사용자 정의를 볼 수 있는 사용자를 제어하려면 사용자 정의 집합을 변경합니다. 자세한 내용은 " 사용자 정의 집합 옵션 "(600 페이지)에서 확인하십시오.
3. 복사하려는 도구 모음 또는 버튼을 마우스 오른쪽 버튼으로 클릭하고 **복사**를 선택합니다.
4. 다음 중 하나를 수행합니다.
   – 도구 모음을 붙여넣으려면 **도구 모음**을 마우스 오른쪽 버튼으로 클릭하고 **붙여넣기**를 선택합니다. 그러면 도구 모음 목록의 끝에 해당 도구 모음이 표시됩니다.
   – 버튼을 붙여넣으려면 새 버튼을 표시할 위치의 앞 또는 뒤에 있는 버튼을 마우스 오른쪽 버튼으로 클릭하고 **붙여넣기**를 선택합니다. 도구 모음의 첫 번째 버튼을 선택한 경우에는 버튼 위치를 선택합니다. 그러지 않으면 선택한 버튼 아래에 버튼이 붙여넣어집니다.
5. 항목을 수정합니다.
6. **저장**을 클릭하여 변경 사항을 저장합니다.

항목을 수정하는 방법에 대한 자세한 내용은 "도구 모음 개인 설정"(600 페이지) 및 "메뉴 항목 및 도구 모음 버튼 생성"(604 페이지)에서 확인하십시오.

사용자 메뉴 , 메뉴 항목 및 버튼 재배열

Windows 에서는 사용자가 생성한 메뉴, 메뉴 항목 및 버튼의 순서를 재배열할 수 있습니다. 예를 들어 **파일 > 새로 만들기** 메뉴 아래에서 사용자 메뉴를 **데이터 테이블** 명령 앞으로 이동할 수 있습니다.

항목을 이동할 수 있는지 여부는 항목의 "일반" 특성에 있는 "소스"에 따라 결정됩니다. "소스"가 **사용자 항목**이면 항목을 이동할 수 있습니다. 기본 제공 항목은 이동할 수 없습니다. 다른 사용자 정의 집합에 정의된 항목은 해당 사용자 정의 집합 내에서만 이동할 수 있습니다. 자세한 내용은 "사용자 정의 집합 옵션"(600 페이지)에서 확인하십시오.

참고: 마우스 오른쪽 버튼 클릭 시 나타나는 메뉴에는 항목을 잘라서 붙여넣기 위한 옵션이 있습니다 . 항목을 잘라내려면 Ctrl+X 를 누르거나 Shift+Delete 를 누를 수도 있습니다 . 붙여넣으려면 Ctrl+V 를 누르거나 Shift+Insert 를 누릅니다 .

사용자 메뉴 및 메뉴 항목을 재배열하려면

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   왼쪽에 " 메인 메뉴 " 목록이 표시됩니다 .
2. 다음을 수행합니다 .
   - 이동할 항목을 선택합니다 .
   - " 소스 " 가 **사용자 항목**인지 확인합니다 .
   - "소스"가 **모든 사용자** 또는 **JMP 추가기능**이면 지정된 사용자 정의 집합을 선택하고 " 소스 " 가 **사용자 항목**인지 확인합니다 .
3. 이동하려는 사용자 메뉴 또는 메뉴 항목을 마우스 오른쪽 버튼으로 클릭하고 **잘라내기**를 선택합니다 .
4. 잘라낸 항목의 새 위치를 마우스 오른쪽 버튼으로 클릭하고 **붙여넣기**를 선택합니다 .

   다음 중 하나가 발생합니다 .
   - 잘라낸 항목의 가능한 위치가 포함된 목록이 표시됩니다 . 메뉴 (예 : **파일** 또는 **새로 만들기**) 또는 메뉴 내의 첫 번째 항목 (예 : **파일 > 새로 만들기** 메뉴의 **데이터 테이블**) 을 선택하면 이 결과가 발생합니다 .
   - 잘라낸 메뉴 항목이 선택한 메뉴 항목 아래에 표시됩니다 . 이 경우 6 단계로 건너뜁니다 .
5. 다음 중 하나를 수행하여 잘라낸 항목의 위치를 선택합니다 .
   - 잘라낸 항목을 선택한 항목 **앞에** 붙여넣으려면 **~ 앞에 붙여넣기**를 선택합니다 .
   - 잘라낸 항목을 선택한 항목 **뒤에** 붙여넣으려면 **~ 뒤에 붙여넣기**를 선택합니다 .
   - 잘라낸 항목을 선택한 메뉴 **안에** 붙여넣으려면 **~ 에 붙여넣기**를 선택합니다 . 이 옵션은 메뉴를 새 위치로 선택한 경우에만 사용할 수 있습니다 .
   - 이동을 취소하려면 **취소**를 선택합니다 .

   지정한 대로 항목이 표시됩니다 .
6. **저장**을 클릭하여 변경 사항을 저장합니다 .

사용자 버튼을 재배열하려면

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

왼쪽에 도구 모음 목록이 표시됩니다 .

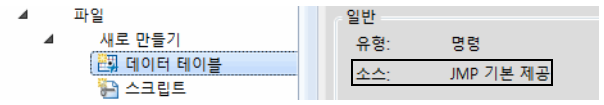
2. 다음을 수행합니다 .
   - 이동하려는 사용자 버튼을 선택합니다 .
   - " 소스 " 가 **사용자 항목**인지 확인합니다 .
   - " 소스 " 가 **모든 사용자** 또는 **JMP 추가기능**이면 지정된 사용자 정의 집합을 선택하고 " 소스 " 가 **사용자 항목**인지 확인합니다 .
3. 이동하려는 사용자 버튼을 마우스 오른쪽 버튼으로 클릭하고 **잘라내기**를 선택합니다 .
4. 잘라낸 버튼의 새 위치를 마우스 오른쪽 버튼으로 클릭합니다 .

   다음 중 하나가 발생합니다 .
   - 잘라낸 버튼의 가능한 위치가 포함된 목록이 표시됩니다 . 도구 모음의 첫 번째 버튼 (예 : **파일 및 편집** 도구 모음의 **새 데이터 테이블**) 을 선택하면 이 결과가 발생합니다 .
   - 도구 모음의 끝에 버튼이 붙여넣어집니다 . 도구 모음의 이름을 선택하면 이 결과가 발생합니다 . 예를 들어 **파일 및 편집**을 선택하여 붙여넣으면 마지막 버튼인 **스크립트 실행** 뒤에 버튼이 표시됩니다 . 이 경우 6 단계로 건너뜁니다 .
   - 잘라낸 버튼이 선택한 버튼 뒤에 표시됩니다 . 이 경우 6 단계로 건너뜁니다 .
5. 다음 중 하나를 수행하여 잘라낸 버튼의 위치를 선택합니다 .
   - 잘라낸 버튼을 선택한 버튼 **앞**에 붙여넣으려면 **~ 앞에 붙여넣기**를 선택합니다 .
   - 잘라낸 버튼을 선택한 버튼 **뒤**에 붙여넣으려면 **~ 뒤에 붙여넣기**를 선택합니다 .
6. **저장**을 클릭하여 변경 사항을 저장합니다 .

사용자 항목 삭제

사용자가 생성한 도구 모음, 버튼, 메뉴 및 메뉴 항목을 일시적으로 표시하거나 숨기는 방법 대신 삭제하는 방법으로 JMP에서 영구적으로 제거할 수 있습니다. 항목을 삭제할 수 있는지 여부와 항목을 포함하는 사용자 정의 집합은 "소스"에 따라 결정됩니다. 그림 12.4에서는 사용자 정의 소스 유형의 예를 보여 줍니다.

그림 12.4 사용자 정의 소스의 예



이 항목은 삭제할 수 없습니다 .

현재 사용자 정의 집합에서 항목을 삭제합니다 .

모든 사용자 사용자 정의 집합에서 항목을 삭제합니다 .

빈 메뉴는 지원되지 않습니다. 메뉴의 유일한 항목을 삭제하면 선택한 항목뿐만 아니라 전체 메뉴가 삭제됩니다.

참고: 사용자 항목을 정말로 삭제할지 확인하십시오. 사용자 항목은 확인 없이 즉시 제거됩니다.

사용자 항목을 삭제하려면 다음 단계를 수행하십시오 .

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .
   메뉴 편집기가 나타납니다 .
2. 삭제하려는 항목을 선택합니다 .
3. 해당 항목이 삭제 가능한지 확인하고 , 필요한 경우 사용자 정의 집합을 변경합니다 .
4. 선택한 항목을 마우스 오른쪽 버튼으로 클릭하고 **삭제**를 선택합니다 .
   항목이 즉시 삭제됩니다 .
5. **저장**을 클릭합니다 .
   변경 사항이 저장됩니다 .

사용자 정의 항목 표시 또는 숨기기

메뉴 편집기에서 특정 사용자 정의 집합의 도구 모음, 버튼, 메뉴 및 메뉴 항목을 표시하거나 숨길 수 있습니다. 항목을 표시하거나 숨기는 위치는 항목의 "소스"에 따라 결정됩니다. 예를 들면 다음과 같습니다.

- "소스"가 **모든 사용자**인 경우 모든 사용자에게서 해당 항목을 숨기려면 사용자 정의 집합을 **모든 사용자**로 변경합니다 .
- "소스"가 JMP **추가기능**이면 사용자 정의 집합을 **JMP 추가기능**으로 변경하고 추가기능 이름을 선택합니다 . 그러면 선택한 JMP 추가기능이 포함된 모든 집합에서 해당 항목이 표시되거나 숨겨집니다 .

사용자 정의 집합을 변경하지 않으면 현재 선택된 집합에서만 해당 항목이 표시되거나 숨겨집니다. 사용자 정의 집합에 대한 자세한 내용은 "사용자 정의 집합 옵션"(600 페이지)에서 확인하십시오 .

참고 : 현재 사용자 환경에서 도구 모음을 빠르게 숨기거나 표시하려면 **보기 > 도구 모음** 목록에서 도구 모음을 선택 또는 선택 취소하십시오 .

사용자 정의 집합에서 항목을 표시하거나 숨기려면 다음 단계를 수행하십시오 .

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .
   메뉴 편집기가 나타납니다 .
2. 수정하려는 사용자 정의 집합을 선택합니다 .
3. 표시하거나 숨기려는 항목을 선택합니다 .

4. " 일반 " 특성에서 **숨김**을 선택하거나 선택 취소합니다 . 일반 특성에 대한 자세한 내용은 "2 단계 : 캡션 및 내부 이름 지정 "(603 페이지) 에서 확인하십시오 .
5. **저장**을 클릭하여 변경 사항을 저장합니다 .

사용자 정의 가져오기

도구 모음 및 메뉴에 대한 변경 사항은 사용자 정의 집합, 즉 일반 텍스트 파일에 저장됩니다. 수정한 도구 모음 및 메뉴를 가져오려면 JMP에서 사용자 정의 집합을 엽니다. 그러면 해당 사용자 정의가 메뉴 편집기에 표시됩니다. 메뉴 사용자 정의 파일에 대한 자세한 내용은 "사용자 정의 집합 옵션"(600 페이지) 에서 확인하십시오.

사용자 정의를 가져오려면 다음 단계를 수행하십시오 .

1. **파일 > 열기**를 선택합니다 .
2. " 파일 이름 " 목록에서 **모든 JMP 파일** (또는 **JMP 메뉴 파일**) 을 선택한 후 사용자 정의 집합을 선택합니다 .
3. **열기**를 클릭합니다 .

   사용자 정의를 가져왔다는 확인 메시지가 표시됩니다 . 또한 메뉴 편집기가 열립니다 .
4. 메뉴 편집기를 표시하고 **확인**을 클릭하여 변경 사항을 저장하거나 , **취소**를 클릭하여 변경 사항을 삭제합니다 .

JMP 에서 사용자 정의 제거

항목을 수정할 때 메뉴 편집기의 **취소** 버튼을 클릭하고 **예**를 클릭하여 취소를 확인하면 저장하지 않은 변경 사항을 삭제할 수 있습니다. 사용자 정의를 저장한 후에 모든 사용자 정의를 제거하고 원래 메뉴 및 도구 모음으로 되돌릴 수도 있습니다.

현재 사용자의 모든 사용자 정의를 제거하려면

1. **보기 > 사용자 정의 > 초기 기본 설정으로 되돌리기**를 선택합니다 .
2. **예**를 클릭하여 현재 사용자의 사용자 정의를 제거합니다 .

   JMP 가 이전 버전의 JMP 에서 사용자 정의를 발견하면 확인 메시지가 나타납니다 . 다음 중 하나를 수행합니다 .
   - 해당 사용자 정의를 현재 설치된 JMP 버전에 추가하지 않으려면 **아니요**를 선택합니다 .
   - 해당 사용자 정의를 현재 설치된 JMP 버전에 추가하려면 **예**를 선택합니다 .

다른 사용자 정의 집합에서 사용자 정의를 제거하려면

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   메뉴 편집기가 나타납니다 .
2. **변경**을 클릭합니다 .

3. 사용자 정의 집합을 선택하고 **확인**을 클릭합니다 .
4. **모두 되돌리기**를 클릭합니다 .

   확인 창이 나타납니다 .
5. **예**를 클릭하여 사용자 정의를 제거합니다 .

선택한 항목에서 사용자 정의를 제거하려면

1. **보기 > 사용자 정의 > 메뉴 / 도구 모음**을 선택합니다 .

   메뉴 편집기가 나타납니다 .
2. (선택 사항) 해당 항목이 포함된 사용자 정의를 선택합니다 .
3. 수정하려는 항목을 선택합니다 .
4. 다음 중 하나를 수행합니다 .
   - 저장하지 않은 항목의 이전 특성을 복원하려면 **재설정**을 클릭합니다 .
   - 기본 제공 항목의 원래 특성을 복원하려면 **모두 되돌리기**를 클릭한 후 **예**를 클릭하여 확인합니다 .
5. **저장**을 클릭하여 변경 사항을 저장합니다 .

macOS 에서 JMP 도구 모음 개인 설정

macOS 에서는 도구 모음에 필요한 아이콘만 표시되도록 설정할 수 있습니다 . 사용할 수 있는 아이콘은 창 유형에 따라 달라집니다 . 예를 들어 스크립트 창에서는 스크립트 형식 조정 , 암호화 및 실행 같은 옵션에 대한 아이콘을 사용할 수 있습니다 . 테이블 아이콘은 데이터 테이블 도구 모음에만 사용할 수 있습니다 .

도구 모음에 아이콘만 표시할지 , 아이콘과 아이콘 이름을 표시할지 , 이름만 표시할지 지정할 수 있습니다 . 필요한 경우 작은 아이콘을 표시할 수도 있습니다 .

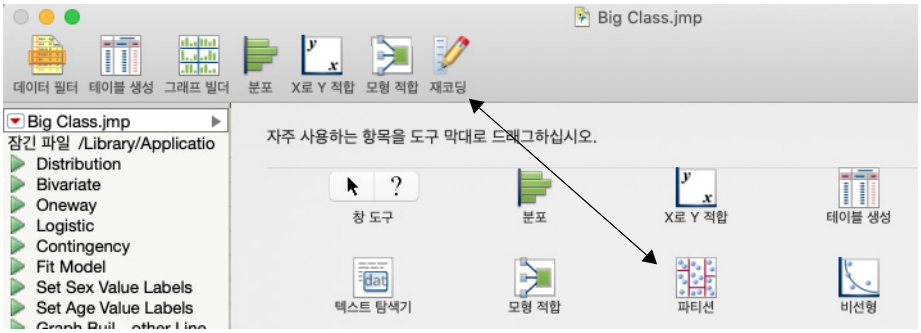
도구 모음 아이콘을 추가 , 제거 또는 재배열하려면

1. 도구 모음을 사용자 정의하려는 유형의 창을 엽니다 .
2. **보기 > 도구 모음 사용자 정의**를 선택합니다 .

   현재 창 유형과 관련된 아이콘만 보여 주는 창이 나타납니다 .
3. 다음 중 원하는 작업을 수행합니다 .
   - 도구 모음에 아이콘을 추가하려면 창의 아이콘을 도구 모음으로 드래그합니다 .
   - 도구 모음의 아이콘을 이동하려면 아이콘을 도구 모음의 새 위치로 드래그합니다 .
   - 도구 모음에서 아이콘을 제거하려면 도구 모음의 아이콘을 창으로 드래그합니다 .

   변경 사항이 현재 창 및 동일한 유형의 열려 있는 모든 창에 적용됩니다 .

그림 12.5 macOS 에서 도구 모음 버튼 추가 , 재배열 및 제거



4. **완료**를 클릭합니다 .

   도구 모음이 너무 넓어서 완전히 표시되지 않으면 도구 모음의 오른쪽 화살표를 클릭하여 다른 아이콘을 표시한 후 선택합니다 . 또는 전체 도구 모음이 표시될 때까지 창을 클릭하여 드래그해도 됩니다 .

참고 : 원래 도구 모음 아이콘 집합을 표시하려면 도구 모음 사용자 정의 창의 아래쪽에 있는 기본 집합을 도구 모음으로 드래그하십시오 .

도구 모음 아이콘의 모양을 변경하려면

1. 도구 모음을 사용자 정의하려는 유형의 창을 엽니다 .
2. **보기 > 도구 모음 사용자 정의**를 선택합니다 .

   도구 모음 아이콘이 포함된 창이 나타납니다 .
3. 창 하단의 " 표시 " 목록에서 다음 옵션 중 하나를 선택합니다 .
   - 아이콘만 표시하려면 **아이콘만**을 선택합니다 .
   - DTP TEAM please delete this segment.
   - 아이콘 이름만 표시하려면 **텍스트만**을 선택합니다 .
4. 작은 아이콘을 표시하거나 표준 아이콘을 표시하려면 **작은 크기 사용**을 선택하거나 선택 취소합니다 .

   변경 사항이 현재 창 및 동일한 유형의 열려 있는 모든 창에 적용됩니다 .
5. **완료**를 클릭합니다 .

macOS 에서 JMP 메뉴 개인 설정

빠르게 파일에 액세스하거나 기존 메뉴 항목을 수정할 수 있도록 JMP 메뉴를 사용자 정의할 수 있습니다. 메뉴 항목을 선택하면 컴퓨터에서 JSL 스크립트가 실행되거나 사용자가 지정한 JSL 명령이 실행됩니다. 예를 들어 자주 사용하는 스크립트를 여는 메뉴 항목을 생성할 수 있습니다. 이 옵션은 스크립트를 홈 창에 즐겨찾기로 추가하는 방법 대신 사용할 수 있습니다.

메뉴 항목 추가 및 이름 바꾸기, 메뉴 항목 숨기기, 메뉴 항목 구분선 삽입, 메뉴 항목과 JSL 연결 등의 기능을 사용할 수 있습니다.

JMP 사용자 정의는 <username>/Library/Application Support/JMP/16/usercust.jmpcust에 저장됩니다.

참고 : 기본 JMP 메뉴를 삭제하거나 다른 메뉴 항목을 표시하기 위해 열리는 하위 메뉴를 삭제할 수는 없습니다 . 기본 및 사용자 하위 메뉴의 이름을 바꿀 수도 없습니다 .

예 : 메뉴 항목 추가

이 예에서는 스크립트를 실행하는 메뉴 항목을 생성합니다.

1. **보기 > 메뉴 사용자 정의**를 선택합니다 .
2. JMP 메뉴 편집기에서 **파일**을 선택한 후 **열기**를 선택합니다 .
3. Control 키를 누른 채 **뒤에 새 메뉴 항목 추가**를 선택합니다 .
4. " 메뉴 라벨 " 상자에 What is the Current Time? 을 입력합니다 .
5. **이 파일의 JSL 실행**을 선택하고 Library/Application Support/JMP/16/Samples/Scripts/timeAnnouncer.jsl 을 선택합니다 .
6. **추가**를 클릭하고 메뉴 편집기 창을 닫습니다 .
7. **파일 > What is the Current Time?** 을 선택합니다 .

   스케줄러 창이 표시되고 현재 시간이 안내됩니다 .

예 : 하위 메뉴 항목 추가

하위 메뉴는 하위 메뉴 항목을 표시하기 위해 열립니다. 이 예에서는 데이터 테이블을 여는 하위 메뉴 항목을 생성합니다.

1. **보기 > 메뉴 사용자 정의**를 선택합니다 .
2. JMP 메뉴 편집기에서 **파일**을 선택한 후 **열기**를 선택합니다 .
3. Control 키를 누른 채 **뒤에 새 하위 메뉴 추가**를 선택합니다 .
4. "Open My Data Tables" 를 입력하고 **JSL 명령**을 선택합니다 .

   참고 : JSL 스크립트를 선택하면 컴퓨터에서 JSL 스크립트를 선택할 수 있습니다 .

5. **추가**를 클릭합니다 .

6. **Open My Data Tables** 를 선택하고 " 메뉴 라벨 " 상자에서 **하위 항목 자리 표시자**를 삭제한 후 Open Napoleons March.jmp 를 입력합니다 .
7. " 이 JSL 실행 " 상자에 `Open("$SAMPLE_DATA/Napoleons March.jmp");` 을 입력합니다 .
8. 이 경우 **변경 사항 저장**을 클릭합니다 .
9. **파일** 메뉴를 선택하여 새 메뉴 항목을 확인합니다 .

팁 : 새 메뉴 항목이 메뉴에 표시되지만 메뉴 편집기에는 표시되지 않으면 메뉴 편집기를 닫았다가 다시 여십시오 .

사용자 메뉴 또는 하위 메뉴 항목을 삭제하려면

- 사용자 메뉴 항목을 삭제하려면 항목을 선택하고 옵션 목록에서 **삭제**를 선택합니다 .
- 모든 사용자 메뉴 항목을 삭제하려면 옵션 목록에서 **모두 되돌리기**를 선택합니다 .
- 하위 메뉴 내의 항목을 삭제하려면 전체 하위 메뉴를 삭제해야 합니다 . 빈 메뉴는 허용되지 않습니다 . 하위 메뉴를 선택하고 옵션 목록에서 **삭제**를 선택합니다 .

추가기능에 대한 JMP 메뉴 및 도구 모음 사용자 정의

JMP에 새 기능을 추가하려면 스크립트를 작성하고 파일을 컴파일하여 추가기능을 생성하면 됩니다. 스크립트를 컴파일하는 한 가지 방법은 Scripting Guide의 에 설명된 대로 추가기능 빌더를 사용하는 것입니다.

추가기능 빌더에서는 "추가기능" 메뉴에 메뉴 항목을 생성하는 데 사용할 수 있는 메뉴 편집기를 제공합니다. JMP 메뉴 편집기를 사용하여 메뉴 항목을 생성하고 추가기능용 도구 모음 버튼을 추가할 수도 있습니다.

13 장

JMP 환경 설정

환경 설정 창

JMP 환경 설정을 사용하면 일반 설정 및 특수 설정을 지정하고 설정을 저장할 수 있습니다.

환경 설정을 변경하려면

1. **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 . 그림 13.1 의 창이 나타납니다 . 이전에 선택한 환경 설정 페이지가 표시됩니다 .
2. 범주를 클릭하고 설정을 선택합니다 . **적용**을 클릭하여 결과를 확인한 후 **확인**을 클릭합니다 .

참고 : JMP 16 구독으로 업그레이드할 때 환경 설정 변경 사항이 유지됩니다 . 다른 JMP 버전에서는 업그레이드할 때마다 환경 설정을 사용자 정의해야 합니다 .

환경 설정을 검색하려면

1. **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 .
2. 상단의 " 필터 " 상자에 텍스트를 입력합니다 .
   매칭되는 텍스트가 강조 표시됩니다 .
3. " 환경 설정 그룹 " 에서 결과를 보려는 범주를 클릭합니다 .

다음과 같은 위치에서 검색어가 나타날 수 있습니다 .

- 환경 설정 자체
- 환경 설정에 대한 툴팁
- 환경 설정 옵션 목록

기본 설정으로 환경 설정 재설정

1. **파일 > 환경 설정** (Windows) 또는 **JMP > 환경 설정** (macOS) 을 선택합니다 .
2. **모두 재설정**을 클릭한 다음 **확인**을 클릭합니다 .

목차

JMP 환경 설정 개요 621
일반 JMP 환경 설정 621
JMP 프로젝트에 대한 환경 설정 625
JMP 보고서의 환경 설정 626
JMP 그래프의 환경 설정 628
JMP 그래프 및 보고서의 스타일 환경 설정 631
데이터 테이블의 환경 설정 634
JMP 데이터 필터의 환경 설정 638
JMP 플랫폼의 환경 설정 639
인쇄 환경 설정 641
텍스트 파일 가져오기 및 내보내기의 환경 설정 642
타사 데이터의 환경 설정 647
Windows 환경 설정 649
Mac OS 환경 설정 653
글꼴 환경 설정 655
통신 환경 설정 657
인터넷 옵션의 환경 설정 657
파일 위치의 환경 설정 659
JSL 스크립트 편집기의 환경 설정 660
SAS Integration 의 환경 설정 663
JMP 업데이트의 환경 설정 666
JSL 디버거의 환경 설정 667
메뉴 환경 설정 668
쿼리 빌더의 환경 설정 669
재코딩 열의 환경 설정 670
열 전환기의 환경 설정 672
로그의 환경 설정 673

JMP 환경 설정 개요

"파일" 메뉴(macOS의 경우 "JMP" 메뉴)의 "환경 설정" 명령은 "환경 설정" 창을 표시합니다. 자세한 내용은 그림 13.1에서 확인하십시오. 이 장에서는 각 범주에 대해 설명합니다.

Windows의 경우, 환경 설정에서 변경한 사항은 다음 폴더 중 하나에 있는 버전별 JMP.PFS 파일에 저장됩니다.

C:/Users/<user name>/AppData/Roaming/SAS/JMP/16/

C:/Users/<user name>/AppData/Roaming/SAS/JMPPro/16/

macOS의 경우, ~/Library/Preferences/에 있는 com.sas.jmp.plist에 환경 설정이 저장됩니다.

JMP 프로그램을 열면 JMP는 이 파일을 찾은 후 이 파일의 변경 사항을 초기 기본 설정으로 간주합니다. 이후의 모든 변경 사항도 이 파일에 저장됩니다.

일반 JMP 환경 설정

"일반" 페이지에 있는 대부분의 환경 설정은 시작 시 JMP 세션을 사용자 정의합니다. 기타 환경 설정은 파일 처리와 JMP 작업 공간의 일반적인 모양에 대한 환경 설정을 지정합니다.

그림 13.1 일반 환경 설정

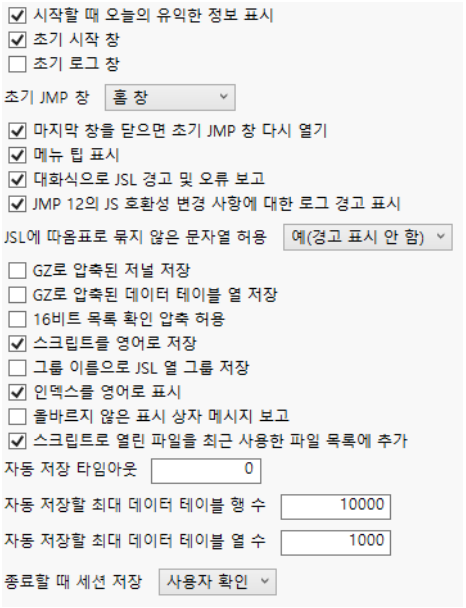


표 13.1 일반 페이지의 환경 설정

환경 설정	설명
시작할 때 오늘의 유익한 정보 표시	시작할 때 "오늘의 유익한 정보" 창을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
초기 시작 창	시작할 때 초기 시작 창을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
초기 로그 창	(Windows에만 해당) JMP가 열릴 때 로그를 열려면 이 옵션을 선택합니다.
초기 JMP 시작하기 창	(macOS에만 해당) JMP를 시작할 때마다 "JMP 시작하기" 창을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
초기 JMP 창	(Windows에만 해당) 나열된 창 중 하나를 기본 창으로 선택합니다. JMP가 시작되면 이 창이 표시됩니다. 홈 창이 기본 창입니다.
홈 창	(macOS) macOS에서 JMP를 시작할 때 JMP 홈 창을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
마지막 창을 닫으면 초기 JMP 창 다시 열기	(Windows에만 해당) 마지막 JMP 창을 닫을 경우 사용자가 초기 JMP 창으로 설정한 창을 열려면 이 옵션을 선택합니다. 이 옵션이 해제되어 있을 때 마지막 JMP 창을 닫으면 JMP도 종료됩니다. 이 옵션은 기본적으로 선택되어 있습니다.
메뉴 팁 표시	(Windows에만 해당) 메인 메뉴 옵션과 빨간색 삼각형 메뉴의 옵션에 툴팁을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
대화식으로 JSL 경고 및 오류 보고	스크립트 오류를 로그에만 표시하지 않고 새 창에도 표시합니다. 이 옵션은 기본적으로 선택되어 있습니다.
JMP 12 의 JS 호환성 변경 사항에 대한 로그 경고 표시	JSL 호환성에 대한 경고를 로그에 출력하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.

표 13.1 일반 페이지의 환경 설정 (계속)

환경 설정	설명
JSL 에 따옴표로 묶지 않은 문자열 허용	플랫폼 메시지의 문자열 값아 따옴표로 묶이지 않은 경우 로그에 경고를 출력하려면 이 옵션을 선택합니다. 리터럴 이름보다 따옴표로 묶은 문자열 값을 사용하는 것이 좋습니다. **예 (경고 표시 안 함)** 따옴표로 묶지 않은 문자열 값을 허용하고 로그에 경고를 출력하지 않습니다 . 이 옵션은 기본적으로 선택되어 있습니다 . **예 (경고 표시)** 따옴표로 묶지 않은 문자열 값을 허용하고 로그에 경고를 출력합니다 . **아니요** 따옴표로 묶지 않은 문자열 값을 무시하고 로그에 경고를 출력하지 않습니다 .
GZ 로 압축된 저널 저장	디스크 공간을 절약하기 위해 JMP 저널을 압축 형식으로 저장하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
GZ 로 압축된 데이터 테이블 열 저장	JMP가 데이터 테이블을 GZ 압축을 사용하여 저장할 수 있도록 하려면 이 옵션을 선택합니다. 이전 버전과의 호환성을 위해 JMP 10은 압축된 파일을 읽을 수 있지만 저장할 수는 없습니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
16비트 목록 확인 압축 허용	고유 값이 65,535 개 이하인 열을 압축합니다. 열에 "목록 확인" 특성을 추가합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
스크립트를 영어로 저장	영어 이외 버전의 JMP에서는 스크립트를 영어 또는 로컬 언어로 저장할 수 있습니다. 이러한 스크립트에서 열 이름 및 텍스트 문자열과 같이 사용자가 지정하는 값은 데이터 테이블에서와 같은 방식으로 표시됩니다. 하지만 스크립트가 영어 버전 JMP에서 실행될 수 있도록 `Distribution` 및 `Set Value`와 같은 명령어는 로컬 언어가 아니라 영어로 표시됩니다. 영어 버전 JMP에서도 글꼴이 필요한 문자를 지원하기만 하면 일본어 변수 이름과 같이 로마 문자가 아닌 문자가 올바르게 표시됩니다. 이 옵션은 기본적으로 선택됩니다. 스크립트를 로컬 언어로 저장하려면 이 옵션을 선택 취소합니다. 이러한 스크립트는 JMP가 동일한 언어로 실행되는 경우에 한해 올바르게 실행됩니다.
그룹 이름으로 JSL 열 그룹 저장	모든 열 이름을 나열하는 대신 플랫폼에서 저장된 JSL 스크립트에 열 그룹 이름을 지정합니다.

표 13.1 일반 페이지의 환경 설정 (계속)

환경 설정	설명
인덱스를 영어로 표시	JMP 인덱스 섹션을 영어로 표시하거나 현재 로케일의 언어로 표시합니다. 이 옵션은 기본적으로 선택됩니다. 이 옵션이 선택 취소된 경우 다음 항목은 현재 로케일의 언어로 표시됩니다. • 개체에 대한 메시지 목록 • 표시 상자에 대한 메시지 목록 • JSL 함수에 대한 범주 목록 **참고 :** 각 인덱스 항목에 대한 설명은 항상 현재 로케일의 언어로 표시됩니다 . 예는 항상 영어로 표시됩니다 .
올바르지 않은 표시 상자 메시지 보고	올바르지 않은 표시 상자 메시지에 대한 정보를 로그로 전송합니다. 이 옵션은 기본적으로 선택되어 있습니다. 이 옵션은 스크립트 개발 중에는 유용할 수 있지만 기존 스크립트에 대해 불필요한 로그가 발생할 수 있습니다.
스크립트로 열린 파일을 최근 사용한 파일 목록에 추가	(Windows에만 해당) JSL `Open()` 명령으로 열린 파일을 홈 창의 "최근 사용한 파일" 페인과 파일 > 최근 사용한 파일 목록에 포함하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
자동 저장 타임아웃	제목 없는 파일을 제외하고 열려 있는 모든 데이터 테이블, 저널, 보고서, 프로젝트 및 스크립트의 내용을 지정된 시간 간격마다 자동으로 저장합니다. 기본 설정은 0분으로, 자동으로 저장하지 않음을 나타냅니다.
자동 저장할 최대 데이터 테이블 행 수	행 수가 지정된 값을 초과하지 않는 경우 데이터 테이블을 자동 저장합니다. 기본 설정은 10,000입니다.
자동 저장할 최대 데이터 테이블 열 수	열 수가 지정된 값을 초과하지 않는 경우 데이터 테이블을 자동 저장합니다. 기본 설정은 1,000입니다.
종료할 때 세션 저장	이 옵션을 사용하면 JMP를 종료할 때 JMP 창의 상태를 저장할 수 있습니다. JMP를 열면 열려 있던 파일 및 창을 비롯하여 저장된 상태가 복원됩니다. 기본 설정은 "사용자 확인"입니다.

JMP 프로젝트에 대한 환경 설정

프로젝트 환경 설정을 사용하면 비어 있는 새 프로젝트를 생성하는 방법, 표시되는 도구 창 등을 제어할 수 있습니다.

그림 13.2 프로젝트 환경 설정

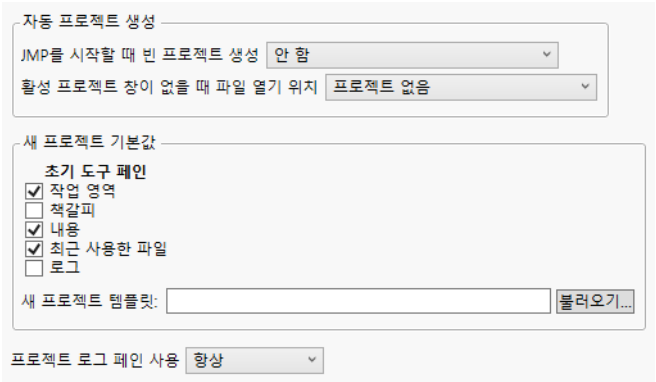


표 13.2 프로젝트 페이지의 환경 설정

환경 설정	설명
JMP를 시작할 때 빈 프로젝트 생성	JMP를 시작할 때 비어 있는 프로젝트를 생성합니다. "안 함", "항상" 또는 "다른 프로젝트가 열려 있지 않은 경우"를 선택할 수 있습니다. 기본 설정은 "안 함"입니다.
활성 프로젝트 창이 없을 때 파일 열기 위치	새 프로젝트가 자동으로 생성되는 방법을 지정합니다. **프로젝트 없음** 프로젝트에 파일을 배치하지 않습니다. 이것이 기본 설정입니다. **프로젝트 열기 또는 프로젝트 없음** 열려 있는 프로젝트를 찾은 경우 해당 프로젝트에, 아니면 프로젝트 없음에 파일을 배치합니다. **프로젝트 열기 또는 새 프로젝트** 열려 있는 프로젝트를 찾은 경우 해당 프로젝트에, 아니면 새 프로젝트에 파일을 배치합니다. **새 프로젝트** 파일을 새 프로젝트에 배치합니다.
초기 도구 페인	프로젝트에 처음 나타나는 도구 페인을 지정합니다. "작업 영역", "내용" 및 "최근 사용한 파일" 도구가 기본적으로 나타납니다.

표 13.2 프로젝트 페이지의 환경 설정 (계속)

환경 설정	설명
새 프로젝트 템플릿	새 프로젝트의 템플릿으로 사용할 프로젝트를 지정합니다. 새 프로젝트에는 템플릿과 동일한 창 레이아웃, 연결된 문서, 그래프, 도구 창 및 기타 프로젝트 요소를 갖습니다.
프로젝트 로그 페인 사용	JMP 로그를 사용하는 대신 프로젝트 로그를 사용하려는 경우 지정하십시오. "항상"이 기본 설정입니다.

JMP 보고서의 환경 설정

"보고서" 환경 설정에서는 보고서의 모양을 사용자 정의합니다.

그림 13.3 보고서 환경 설정

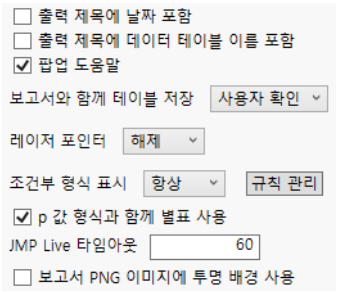


표 13.3 보고서 페이지의 환경 설정

환경 설정	설명
출력 제목에 날짜 포함	분석이 수행된 날짜 및 시간을 보고서 창에 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
출력 제목에 데이터 테이블 이름 포함	보고서 맨 위에 데이터 테이블의 이름과 노트를 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
팝업 도움말	일부 숫자 출력에서 JMP는 마우스로 결과 위에 원을 그릴 때 툴팁 스타일의 도움말을 제공합니다. 사용 가능한 툴팁이 있는 경우 툴팁을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다. **참고 :** 이 옵션을 선택 취소하면 가리키기 라벨 그래프가 더 이상 표시되지 않습니다 .

표 13.3 보고서 페이지의 환경 설정 (계속)

환경 설정	설명
보고서와 함께 테이블 저장	메뉴의 옵션 중 하나를 선택하여 보고서에 테이블이 저장되는 방식을 결정합니다. **포함** 보고서를 저장할 때 테이블이 보고서에 포함됩니다 . 보고서를 다른 사람과 공유하려면 이 옵션을 선택합니다 . 보고서에 저장된 데이터 테이블은 보고서를 열 때마다 변경되지 않은 상태로 다시 열립니다 . 테이블을 변경하더라도 보고서를 다시 저장하지 않는 한 변경 사항이 보고서에는 저장되지 않습니다 . **별도** 보고서를 저장할 때 테이블이 보고서에서 참조됩니다 . **사용자 확인** 보고서를 저장할 때 보고서 저장 방법을 지정하라는 메시지가 표시됩니다 . 보고서 내에 테이블을 포함하거나 , 보고서에 테이블에 대한 참조를 추가할 수 있습니다 . 이것이 기본 설정입니다 .
레이저 포인터	JMP에는 보고서의 일부분을 시각적으로 강조할 수 있는 기본 제공 레이저 포인터가 있습니다. 이 옵션은 기본적으로 선택됩니다. 이 기능을 설정하려면 목록에서 레이저 포인터의 색상을 선택합니다.
조건부 형식 표시	값을 나타내는 텍스트의 색상 형식을 조건부로 지정합니다. 상관관계 값, *p* 값 및 요인 패턴 값에 사용할 수 있습니다. 메뉴에서 다음 옵션 중 하나를 선택합니다. **항상** 상관관계 값, *p* 값 및 요인 패턴 값에 조건부 형식 지정이 항상 적용됩니다 . 이것이 기본 설정입니다 . **화면만** 상관관계 값, *p* 값 및 요인 패턴 값에 조건부 형식 지정이 항상 적용됩니다 . 하지만 보고서를 인쇄할 때는 조건부 형식이 표시되지 않습니다 . **안 함** 상관관계 값 , *p* 값 및 요인 패턴 값에 조건부 형식 지정이 적용되지 않습니다 .
규칙 관리	숫자 열 또는 행렬에 적용할 수 있는 사용자 조건부 형식 지정 규칙을 생성합니다. 자세한 내용은 "JMP 보고서" 장의 "보고서에서 조건부 형식 사용"(435 페이지)에서 확인하십시오. **참고 :** *p* 값 조건에 별표 (*) 가 사용되거나 사용되지 않도록 하려면 **p 값 형식과 함께 별표 사용** 환경 설정을 확인하십시오 .

표 13.3 보고서 페이지의 환경 설정 (계속)

환경 설정	설명
p 값 형식과 함께 별표 사용	유의한 *p* 값 옆에 별표(*)를 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다. **참고 :** 정의된 조건에 별표 (*) 가 사용되지 않도록 하려면 p 값에 대한 조건부 형식 지정 규칙을 확인하십시오 .
JMP Live 타임아웃	JMP Live에 게시하기 위한 타임아웃 값을 지정합니다. 기본 시간은 60초입니다.
보고서 PNG 이미지에 투명 배경 사용	보고서를 PNG 파일로 저장할 때 투명한 배경을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.

JMP 그래프의 환경 설정

"그래프" 환경 설정에서는 그래프의 모양을 사용자 정의합니다. 그래프 모양을 구성하는 방법에 대한 자세한 내용은 "JMP 그래프 및 보고서의 스타일 환경 설정"(631페이지)에서 확인하십시오.

그림 13.4 그래프 환경 설정

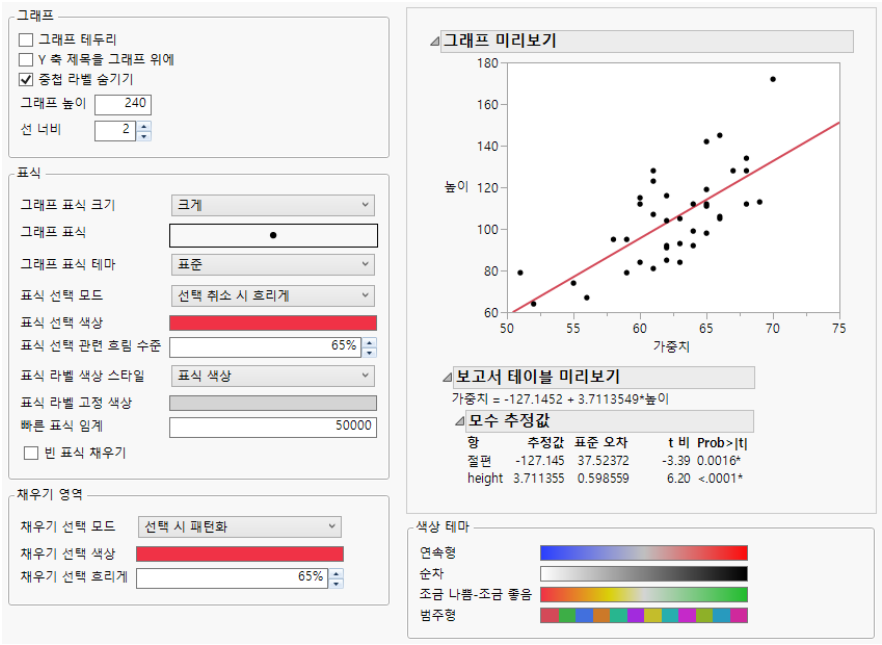


표 13.4 그래프 페이지의 환경 설정

환경 설정	설명
그래프 테두리	그래프 영역 주위에 테두리 선을 그립니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
Y 축 제목을 그래프 위에	Y 축의 제목을 선 옆이 아니라 선 위쪽에 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
중첩 라벨 숨기기	중첩 차트 라벨을 숨기려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
그래프 높이	그래프의 기본 높이(픽셀)를 설정합니다. 기본 설정은 240 픽셀입니다. 일부 플랫폼에서는 이 값에 일정 값을 곱합니다. **참고 :** 그래프 빌더 창을 사용할 경우 " 그래프 높이 " 를 지정하면 창이 비례적으로 커지지만 정확한 값이 적용되지 않습니다 .
선 너비	그래프 내용과 관련된 선의 기본 너비(픽셀)를 설정합니다. 기본 설정은 2 픽셀입니다. **참고 :** 이 환경 설정은 격자선에 적용되지 않습니다 .
그래프 표식 크기	그래프 표식의 기본 크기(점, 작게, 중간, 크게, XL, XXL 또는 XXXL)를 선택합니다. 이 환경 설정은 그래프에서 표식 크기가 기본 크기로 설정된 경우에 적용됩니다. 기본 설정은 "크게"입니다. 그래프의 표식 수가 "빠른 표식 임계"를 초과하면 그래프의 "표식 그리기 모드"가 "빠름"으로 설정되어 도트 크기가 조정된 표식이 그려집니다.
그래프 표식	그래프 표식의 기본 표식 형태를 선택합니다. 기본 설정은 작은 점입니다.
그래프 표식 테마	행 또는 열별로 표시하는 경우 그래프 표식의 기본 테마를 선택합니다. 기본 설정은 "표준"입니다.
표식 선택 모드	선택된 표식의 기본 형식을 선택합니다. 기본 설정은 "선택 취소 시 흐리게"입니다. 자세한 내용은 "JMP 보고서" 장의 "표식 선택 모드 변경"(446 페이지)에서 확인하십시오.
표식 선택 색상	"표식 선택 모드"가 "선택 시 지정한 색상 적용"인 경우 이 설정은 선택된 표식에 지정된 색상을 적용합니다. 기본 색상은 빨간색입니다.

표 13.4 그래프 페이지의 환경 설정 (계속)

환경 설정	설명
표식 선택 관련 흐림 수준	"표식 선택 모드"가 "선택 취소 시 흐리게"인 경우 이 설정은 선택되지 않은 표식을 지정된 수준(백분율)만큼 흐리게 표시합니다. 기본 설정은 65% 입니다.
표식 라벨 색상 스타일	데이터 테이블의 열에 라벨을 지정할 경우 그래프에 표시되는 표식 라벨의 색상을 선택합니다. **표식 색상** 표식 라벨에 표식과 동일한 색상을 적용합니다. 이 옵션은 기본적으로 선택되어 있습니다. **표식 색상 흐리게** 표식 라벨을 " 표식 선택 관련 흐림 수준 " 환경 설정에 따라 흐리게 표시합니다. **고정 색상** 표식 라벨에 지정된 색상을 적용합니다. " 표식 라벨 고정 색상 " 환경 설정에 따라 색상이 결정됩니다.
표식 라벨 고정 색상	"표식 라벨 색상 스타일" 환경 설정에서 "고정 색상"을 선택한 경우 고정 색상 표식 라벨의 색상을 선택합니다. 기본 색상은 회색입니다.
빠른 표식 임계	JMP는 보고서 창을 새로 고칠 때 그림의 표식 또는 점을 두 가지 속도, 즉 보통 속도와 빠른 속도로 그릴 수 있습니다. 보통 그리기 모드에 있는 경우 그래프의 표식 수가 지정된 임계 개수보다 많으면 자동으로 빠름 모드로 전환됩니다. 보통 모드와 빠름 모드의 전환 기준이 되는 표식 수를 입력합니다. 그래프의 표식 수가 "빠른 표식 임계"를 초과하면 그래프의 "표식 그리기 모드"가 "빠름"으로 설정되어 도트 크기가 조정된 표식이 그려집니다. 기본 설정은 50,000 입니다.
빈 표식 채우기	표식에 불투명한 배경을 적용하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
채우기 선택 모드	그래픽 개체에서 선택된 행에 대한 기본 강조 표시를 선택합니다. **선택 시 패턴화** 줄무늬 패턴 (기본값) **선택 시 어둡게** 어두운 단색 **선택 시 외곽선 표시** 외곽선 **선택 시 지정한 색상 적용** 환경 설정의 " 채우기 선택 색상 " 에 따라 선택한 개체의 형식을 지정합니다. **선택 취소 시 흐리게** 선택 취소된 행은 흐리게 표시하고 선택된 행은 원래 색상으로 표시합니다.

표 13.4 그래프 페이지의 환경 설정 (계속)

환경 설정	설명
채우기 선택 색상	"채우기 선택 모드"가 "선택 시 지정한 색상 적용"으로 설정된 경우 그래픽 개체에서 선택된 행의 기본 색상을 선택합니다. 기본 색상은 빨간색입니다.
연속 색상 테마	연속형 데이터의 기본 색상 테마를 선택합니다. 기본 테마는 "파란->회색->빨간색"입니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "그래프의 색상 테마 생성"(220페이지)에서 확인하십시오.
순차적 색상 테마	의미 있는 중간점이 없는 순차적 데이터에 대한 기본 색상 테마를 선택합니다. 기본 테마는 "흰색-검정"입니다.
조금 나쁨 - 조금 좋음 색상 테마	양의 데이터에서 음의 데이터까지 포함하는 확산 데이터에 대한 기본 색상 테마를 선택합니다. 기본 테마는 "신호등 조금 나쁨-조금 좋음"입니다.
범주형 색상 테마	범주형 데이터의 기본 색상 테마를 선택합니다. 기본 테마는 "JMP 기본 설정"입니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "그래프의 색상 테마 생성"(220페이지)에서 확인하십시오.

JMP 그래프 및 보고서의 스타일 환경 설정

"스타일" 페이지에서는 색상, 눈금 표식, 격자선 및 보고서 테이블을 사용자 정의할 수 있습니다.

그래프 모양을 구성하는 방법에 대한 자세한 내용은 "JMP 그래프의 환경 설정"(628페이지)에서 확인하십시오.

그림 13.5 스타일 환경 설정

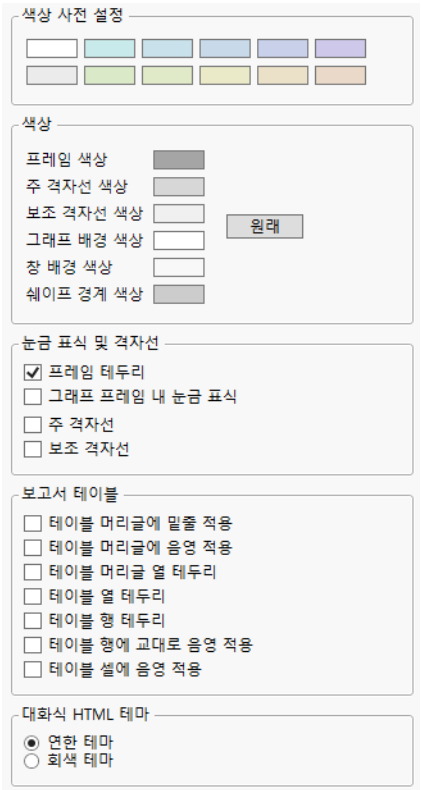


표 13.5 스타일 페이지의 환경 설정

환경 설정	설명
색상 사전 설정	"색상 사전 설정"에서는 보고서, 그래프, 차트 등의 모양을 사용자 정의하는 데 사용할 사전 정의된 색상 집합을 제공합니다. 다양한 요소에 선택한 색상의 여러 색조가 적용됩니다. **참고 :** 창에도 선택한 색상이 적용됩니다 .
프레임 색상	그래프 프레임 및 눈금 표식에 적용되는 색상을 정의합니다. 기본 색상을 클릭하여 변경할 수 있습니다.
주 격자선 색상	주 격자선 색상을 정의합니다. 기본 색상을 클릭하여 변경할 수 있습니다.
보조 격자선 색상	보조 격자선 색상을 정의합니다. 기본 색상을 클릭하여 변경할 수 있습니다.

표 13.5 스타일 페이지의 환경 설정 (계속)

환경 설정	설명
그래프 배경 색상	색상 상자를 클릭하여 모든 그래프의 배경 색상을 선택합니다. 기본 색상을 클릭하여 변경할 수 있습니다.
창 배경 색상	색상 상자를 클릭하여 모든 보고서 및 데이터 테이블의 배경 색상을 선택합니다. 기본 색상을 클릭하여 변경할 수 있습니다.
쉐이프 경계 색상	쉐이프 파일 경계의 색상을 선택하려면 색상 상자를 클릭합니다.
원래	기본 색상 설정을 복원하려면 클릭합니다.
프레임 테두리	축에만 테두리를 표시합니다. 이 옵션은 기본적으로 선택되어 있습니다.
그래프 프레임 내 눈금 표식	그래프 프레임 내에 축 눈금 표식을 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
주 격자선	그래프에 주 격자선을 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
보조 격자선	그래프에 보조 격자선을 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
테이블 머리글에 밑줄 적용	테이블 머리글 아래에 테두리를 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
테이블 머리글에 음영 적용	테이블 머리글에 음영을 표시합니다. 이 옵션은 기본적으로 선택되어 있습니다.
테이블 머리글 열 테두리	테이블 머리글의 열 사이에 구분선을 표시합니다. 이 옵션은 기본적으로 선택되어 있습니다.
테이블 열 테두리	테이블 열 사이에 테두리를 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
테이블 행 테두리	테이블 행 사이에 테두리를 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
테이블 행에 교대로 음영 적용	테이블 행에 교대로 음영을 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
테이블 셀에 음영 적용	테이블 셀에 음영을 표시합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
대화식 HTML 테마	대화식 HTML 페이지의 색상 테마를 지정합니다. 기본 테마는 "연한 테마"입니다.

데이터 테이블의 환경 설정

"테이블" 환경 설정에서는 JMP 데이터 테이블과 관련하여 계산식 처리, 모양, SAS 데이터 집합과의 호환성 등을 사용자 정의합니다.

그림 13.6 테이블 환경 설정

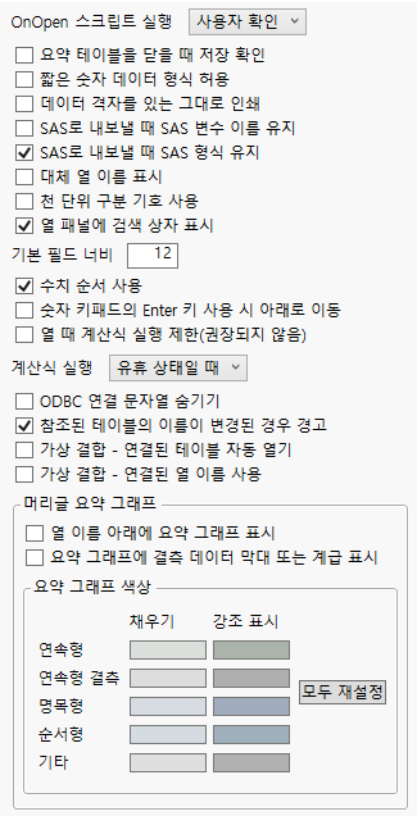


표 13.6 테이블 페이지의 환경 설정

환경 설정	설명
OnOpen 스크립트 실행	스크립트를 데이터 테이블에 저장하고 이름을 OnOpen 또는 On Open으로 지정하면 해당 데이터 테이블을 열 때마다 JMP가 이 스크립트를 자동으로 실행할 수 있습니다. 다른 프로그램을 실행하는 On Open 스크립트는 실행되지 않습니다. 기본적으로 사용자 확인이 선택되어 있습니다. 현재 JMP 세션에서 데이터 테이블을 열 때마다 선택한 설정에 대한 확인이 요청됩니다. **항상** 사용자 확인 없이 On Open 스크립트가 실행되도록 허용합니다. **안 함** 데이터 테이블을 열 때 On Open 스크립트가 자동으로 실행되지 않도록 합니다.
요약 테이블을 닫을 때 저장 확인	연결된 요약 테이블을 저장하지 않고 닫을 때 사용자에게 저장 여부를 묻습니다. "요약" 창에서 이 옵션을 선택하면 "테이블" 환경 설정에도 해당 옵션이 선택됩니다. 이후에 연결된 모든 요약 테이블을 저장할지 묻는 메시지를 표시하지 않으려면 해당 환경 설정을 선택 취소하십시오. 이 옵션은 기본적으로 선택 취소되어 있습니다.
짧은 숫자 데이터 형식 허용	JMP에는 숫자 데이터를 8비트(1바이트)로 짧게 저장하는 기능이 있습니다. 이 옵션을 선택하면 "열" > "열 정보"를 선택하고 열에 데이터 유형을 할당할 때 짧은 정수 형식을 사용할 수 있습니다. 데이터에 올바른 짧은 정수 형식을 사용하면 숫자가 다르게 표시되지는 않지만 데이터 테이블에 사용되는 디스크 공간이 줄어듭니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "짧은 정수 형식"(272 페이지)에서 확인하십시오.
데이터 격자를 있는 그대로 인쇄	JMP 데이터 테이블을 화면에 표시된 그대로 인쇄하려면 이 옵션을 선택합니다. 내용 너비에 맞게 열 너비를 조정하려면 이 옵션을 선택 해제합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
SAS 로 내보낼 때 SAS 변수 이름 유지	JMP 데이터 테이블을 SAS 데이터 집합으로 내보낼 때 SAS 규격의 변수 이름을 사용하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
SAS 로 내보낼 때 SAS 형식 유지	JMP 데이터 테이블을 SAS 데이터 집합으로 내보낼 때 SAS 규격의 형식을 사용하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.

표 13.6 테이블 페이지의 환경 설정 (계속)

환경 설정	설명
대체 열 이름 표시	열의 이름-라벨 쌍을 표시하려면 이 옵션을 선택합니다. 열의 이름 특성과 라벨 특성이 정의되어 있고 열 이름이 두 특성 중 하나와 일치하면 해당 열은 이름-라벨 쌍을 갖게 됩니다. JMP는 다음과 같은 이름-라벨 쌍을 인식합니다. • SAS 이름-라벨 쌍: 특성이 "SAS 이름" 및 "SAS 라벨" 인 경우 • SPSS 이름 - 라벨 쌍 : 특성이 "SPSS 이름 " 및 "SPSS 라벨 " 인 경우 • 짧은 이름과 긴 이름으로 구성된 이름 - 라벨 쌍 : 특성이 " 짧은 이름 " 및 " 긴 이름 " 인 경우 이 옵션이 선택되어 있으면 대화상자에 이름과 라벨이 모두 표시됩니다.
천 단위 구분 기호 사용	숫자를 현재 로케일에 적합한 천 단위 구분 기호를 사용하여 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
열 패널에 검색 상자 표시	열 패널에서 열 이름 위에 검색 상자를 표시합니다. 이 옵션은 기본적으로 선택되어 있습니다.
기본 필드 너비	숫자 열에 표시되는 숫자의 자릿수를 변경합니다. 기본 설정은 12입니다.
수치 순서 사용	숫자가 포함된 텍스트를 수치 순서로 정렬하도록 지정합니다. 이 옵션은 기본적으로 선택되어 있습니다.
숫자 키패드의 Enter 키 사용 시 아래로 이동	(Windows에만 해당) 데이터 테이블 셀이 선택되어 있을 때 키보드의 Enter 키를 누르면 아래쪽 다음 셀이 선택됩니다. Tab 키나 키보드 오른쪽에 있는 숫자 키패드의 Enter 키를 누르면 오른쪽 다음 셀이 선택됩니다. 숫자 키패드에 있는 Enter 키의 동작을 변경하여 오른쪽 대신 아래쪽의 다음 테이블 셀이 선택되도록 하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
열 때 계산식 실행 제한	데이터 테이블을 열 때 열 계산식이 실행되지 않도록 하려면 이 옵션을 선택합니다. 데이터 테이블을 열 때 계산식이 실행되도록 하려면 이 옵션을 선택 취소합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다. 일반적으로 계산식의 결과를 알아야 하기 때문에 이 옵션을 선택하는 것은 권장되지 않습니다.

표 13.6 테이블 페이지의 환경 설정 (계속)

환경 설정	설명
계산식 실행	계산식이 실행되는 시기를 지정합니다. **유휴 상태일 때** JMP 가 작업을 수행하지 않을 때 계산식을 실행합니다. 진행률 표시줄이 나타나지 않습니다. 이 옵션은 기본적으로 선택되어 있습니다 . **즉시** 계산식을 즉시 실행하고 진행률 표시줄을 표시합니다 .
ODBC 연결 문자열 숨기기	Open Database 명령에서 ODBC 연결 문자열, 즉 사용자 ID 및 암호를 숨기려면 이 옵션을 선택합니다. 자세한 내용은 Scripting Guide의 에서 확인하십시오. 이 옵션은 기본적으로 선택 취소되어 있습니다.
참조된 테이블의 이름이 변경된 경우 경고	가상으로 연결된(참조된) 테이블의 이름이 변경된 경우 경고 메시지를 표시합니다. 이 옵션은 기본적으로 선택되어 있습니다.
가상 결합 - 연결된 테이블 자동 열기	가상으로 결합된 데이터 테이블을 자동으로 엽니다. 데이터 테이블에 대한 링크는 JMP 홈 창에 있습니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
가상 결합 - 연결된 열 이름 사용	"열" 목록에 표시된 가상으로 결합된 열에서 연결된 열 이름이 사용되도록 지정합니다. 이 옵션을 사용하면 더 짧은 열 이름을 지정할 수 있습니다(예를 들어 Cheese[Choice] 대신 Cheese 지정). 이 옵션은 기본적으로 선택 취소되어 있습니다.
머리글 요약 그래프	**열 이름 아래에 요약 그래프 표시** 열 머리글에 요약 그래프 (분포) 를 표시합니다 . 행 수가 임계값 (예 : 3 백만 개) 미만인 경우 적용됩니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 . **요약 그래프에 결측 데이터 막대 또는 계급 표시** 요약 그래프에 결측 데이터 막대 또는 계급을 기본적으로 표시할지 여부를 지정합니다 . 개별 그래프에 결측값 막대를 표시하려면 그래프를 마우스 오른쪽 버튼으로 클릭하고 **결측값 막대**를 선택합니다 . 결측값 막대는 기본적으로 표시되지 않습니다 . **요약 그래프 색상** 데이터 유형을 포함하여 채우기 및 강조 표시의 색상 (막대를 선택할 때 나타나는 색상) 을 지정합니다 .

JMP 데이터 필터의 환경 설정

"데이터 필터" 환경 설정에서는 데이터 필터에 대한 설정을 사용자 정의합니다.

그림 13.7 데이터 필터링 환경 설정

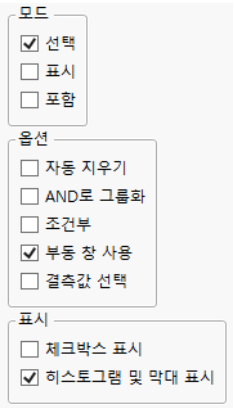


표 13.7 데이터 필터 페이지의 환경 설정

환경 설정	설명
모드	필터링 모드를 지정합니다. **선택** 데이터 테이블에서 선택한 행을 강조 표시된 상태로 표시합니다 . 이 옵션은 기본적으로 선택되어 있습니다 . **표시** 선택되지 않은 행을 숨김 아이콘으로 표시합니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 . **포함** 선택되지 않은 행을 제외 아이콘으로 표시합니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 .

표 13.7 데이터 필터 페이지의 환경 설정 (계속)

환경 설정	설명
옵션	데이터 필터의 옵션을 지정합니다.
	자동 지우기 " 데이터 필터 " 에 명목형 또는 순서형 열이 여러 개 선택된 경우 이 옵션을 사용하면 열을 새로 선택하기 전에 다른 선택 상태를 지울 수 있습니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 .
	AND 로 그룹화 열 필터의 여러 그룹에 대한 AND 및 OR 동작을 제어할 수 있습니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 .
	조건부 선택된 필터 열에 대해 표시되는 범주를 제한합니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 .
	부동 창 사용 " 데이터 필터 " 창을 관련 데이터 테이블의 위에 유지합니다 . 이 옵션은 기본적으로 선택되어 있습니다 .
	결측값 선택 데이터 테이블에서 결측값을 강조 표시합니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 .
표시	범주형 열의 표시 옵션을 지정합니다.
	체크박스 표시 각 값 옆에 체크박스를 표시합니다 . 이 옵션은 기본적으로 선택 취소되어 있습니다 .
	히스토그램 및 막대 표시 데이터 필터에 히스토그램과 막대를 표시합니다 . 이 옵션은 기본적으로 선택되어 있습니다 .

JMP 플랫폼의 환경 설정

각 분석 보고서에는 기본적으로 표시되는 다양한 그림 및 테이블 옵션이 있습니다. 하지만 특정 분석을 실행할 때마다 표시하려는 추가 옵션이 있을 수도 있습니다. 예를 들어 이변량 분석은 기본적으로 산점도를 표시하지만 분석 시마다 항상 선형 적합을 표시하도록 할 수도 있습니다. "환경 설정" 창의 왼쪽 패널에서 "플랫폼" 옵션을 선택하면 기본 분석 옵션을 설정할 수 있습니다. 분석은 이변량 플랫폼과 같은 플랫폼을 사용하여 실행됩니다. 따라서 이 범주의 이름은 "플랫폼" 입니다.

"플랫폼" 목록에서 분석 이름을 강조 표시합니다. 그러면 사용 가능한 옵션이 "옵션" 상자에 표시되며 기본값은 이미 선택되어 있습니다.

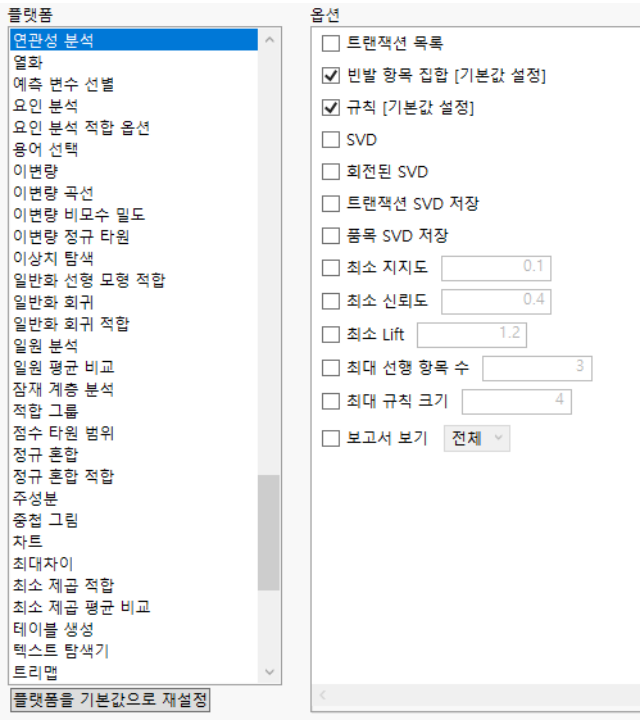
일부 플랫폼 환경 설정에는 보고서에 옵션을 표시할지 여부를 제어하기 위한 "설정" 체크박스가 포함되어 있습니다.

- 옵션이 보고서에 표시되지 않도록 하려면 옵션 옆의 **설정**을 선택하고 해당 옵션은 선택하지 않습니다 .
- 옵션이 보고서에 표시되도록 하려면 **설정**을 선택하고 해당 옵션을 선택합니다 .

선택한 플랫폼 옵션을 기본 설정으로 되돌리려면 **플랫폼을 기본값으로 재설정**을 클릭합니다.

모든 플랫폼을 기본 설정으로 되돌리려면 **기본값으로 재설정**을 클릭합니다.

그림 13.8 플랫폼 환경 설정



인쇄 환경 설정

" 인쇄 " 페이지에서는 여백 , 머리글 및 바닥글과 같은 기본 인쇄 설정을 구성할 수 있습니다 .

그림 13.9 인쇄 환경 설정

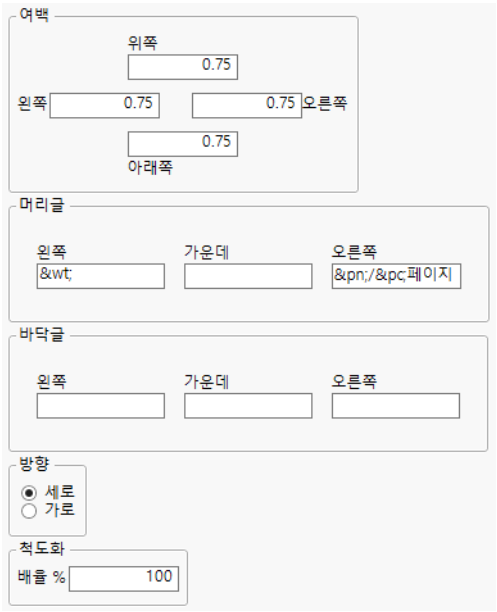


표 13.8 인쇄 페이지의 환경 설정

환경 설정	설명
위쪽 (여백)	인쇄 페이지의 위쪽 여백을 설정합니다. 기본 설정은 0.75 인치입니다.
왼쪽 (여백)	인쇄 페이지의 왼쪽 여백을 설정합니다. 기본 설정은 0.75 인치입니다.
오른쪽 (여백)	인쇄 페이지의 오른쪽 여백을 설정합니다. 기본 설정은 0.75 인치입니다.
아래쪽 (여백)	인쇄 페이지의 아래쪽 여백을 설정합니다. 기본 설정은 0.75 인치입니다.
왼쪽 (머리글)	페이지의 왼쪽 머리글 정보를 설정합니다. 기본 설정은 **&wt;**(즉, 제목)입니다.
가운데 (머리글)	페이지의 가운데 머리글 정보를 설정합니다. 기본 설정은 공백입니다.
오른쪽 (머리글)	페이지의 오른쪽 머리글 정보를 설정합니다. 기본 설정은 **Page &pn; of &pc;**(즉, 페이지 번호/페이지 수)입니다.
왼쪽 (바닥글)	페이지의 왼쪽 바닥글 정보를 설정합니다. 기본 설정은 공백입니다.
가운데 (바닥글)	페이지의 가운데 바닥글 정보를 설정합니다. 기본 설정은 공백입니다.

표 13.8 인쇄 페이지의 환경 설정 (계속)

환경 설정	설명
오른쪽 (바닥글)	페이지의 오른쪽 바닥글 정보를 설정합니다. 기본 설정은 공백입니다.
방향	페이지의 인쇄 방향을 "세로" 또는 "가로"로 설정합니다. 기본 설정은 "세로"입니다.
배율	페이지의 인쇄 배율 인수(백분율)를 설정합니다. 기본 설정은 100%입니다.

참고 : 머리글 또는 바닥글에 현재 날짜를 인쇄하기 위한 코드는 **&d** 입니다 .

텍스트 파일 가져오기 및 내보내기의 환경 설정

"텍스트 데이터 파일" 환경 설정에서는 텍스트 파일의 가져오기 및 내보내기를 사용자 정의합니다.

그림 13.10 텍스트 데이터 파일 환경 설정

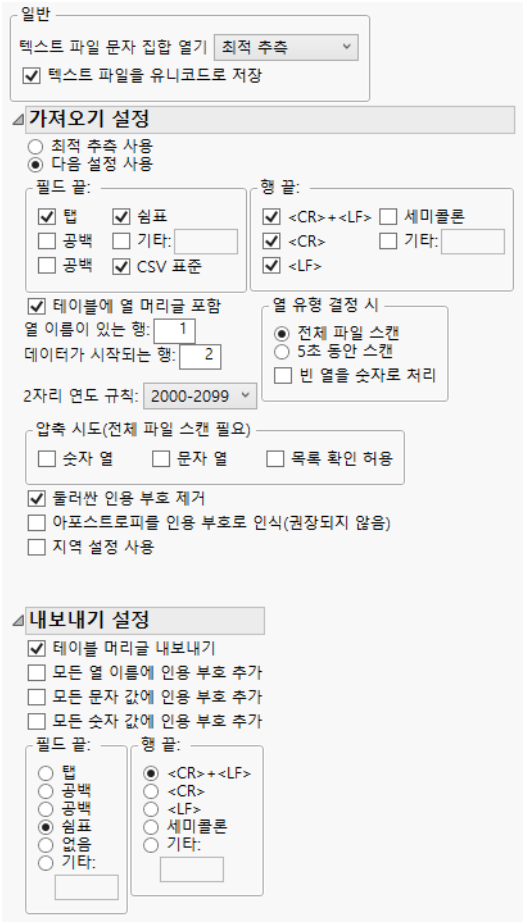


표 13.9 텍스트 파일의 가져오기 설정에 대한 환경 설정

환경 설정	설명
텍스트 파일 문자 집합 열기	메뉴 옵션 중 하나를 선택하여 JMP에서 파일을 여는 데 사용되는 문자 인코딩을 결정합니다. 기본 설정은 "최적 추측"입니다. Windows-1252는 일부 시스템에서 ANSII로 간주되며, UTF-8-BOM은 지원되지 않습니다.
텍스트 파일을 유니코드로 저장	JMP는 é 및 ½과 같은 특수 문자를 지원하는 유니코드 문자 집합을 사용합니다. 유니코드 특수 문자가 없는 파일은 자동으로 일반 텍스트로 저장됩니다. 이 옵션은 기본적으로 선택되어 있습니다. 모든 파일을 일반 텍스트로 저장하려면 이 체크박스를 선택 취소합니다.

표 13.9 텍스트 파일의 가져오기 설정에 대한 환경 설정 (계속)

환경 설정	설명
가져오기 설정	JMP에서 텍스트 파일을 여는 데 사용되는 전략을 선택합니다. 기본적으로 **다음 설정 사용**이 선택되어 있습니다. 이 경우 해당 설정이 사용하는 텍스트 파일에 맞는지 확인해야 합니다. **최적 추측 사용**을 선택하면 JMP는 텍스트 파일에서 탭, 쉼표, 공백 및 그 밖의 일부 문자에 대한 통계를 수집하고 규칙 기반 시스템을 사용하여 적절한 파일 형식을 결정합니다. 이 규칙에 따라 적절한 필드 너비와 적절한 줄당 필드 수가 결정됩니다. 데이터 형식이 규칙에 따라 추측된 것과 너무 다른 경우에는JMP가 잘못 추측한 것입니다. 이 경우에는 마법사를 사용하거나 이러한 환경 설정에서 데이터를 명시적으로 설명합니다.
필드 끝	텍스트 데이터를 가져올 때 필드 끝을 나타내는 구분자로 사용할 문자를 하나 이상 선택합니다. 탭, 쉼표, 및 CSV 표준은 기본적으로 선택되어 있습니다. 목록에 없는 구분자를 지정하려면 **기타** 옵션을 선택하고 문자를 입력합니다.
행 끝	줄(행) 끝을 나타내는 구분자로 사용할 문자를 하나 이상 선택합니다. <CR>+<LF>, <CR> 및 <LF>가 기본적으로 선택되어 있습니다. 목록에 없는 구분자를 지정하려면 **기타** 옵션을 선택하고 문자를 입력합니다. 텍스트 데이터를 가져올 때 큰따옴표가 발견되면 닫는 큰따옴표를 찾기 위해 구분자 규칙이 변경됩니다. 따옴표 내에 포함된 공백을 비롯하여 다른 텍스트 구분자는 무시되고 텍스트 문자열의 일부로 처리됩니다.
테이블에 열 머리글 포함	텍스트 파일에 열 이름이 포함되어 있으면 이 옵션을 선택합니다. 이 옵션을 선택할 경우 **열 이름이 있는 행** 옆의 필드에 열 이름이 있는 행의 번호를 입력합니다. 이 옵션은 기본적으로 선택되어 있습니다.
열 이름이 있는 행	**테이블에 열 머리글 포함** 옵션을 선택한 경우 열 이름이 있는 행의 번호를 이 필드에 입력합니다. 행 1이 기본 설정입니다.
데이터가 시작되는 행	텍스트 파일에서 데이터가 시작되는 행의 번호를 입력합니다. 행 2가 기본 설정입니다.

표 13.9 텍스트 파일의 가져오기 설정에 대한 환경 설정 (계속)

환경 설정	설명
열 유형 결정 시	열의 데이터 유형을 확인하기 위해 JMP에서 텍스트 파일을 얼마나 오래 스캔할지 설정합니다. **전체 파일 스캔**이 기본적으로 선택되어 있습니다. **전체 파일 스캔** 옵션을 선택할 경우 파일이 크면 텍스트 파일을 가져오는 속도가 느려질 수 있습니다. 이 경우 **5초 동안 스캔**을 대신 선택하는 것이 좋습니다. 텍스트 파일에 결측 데이터의 열이 포함되어 있으면 **빈 열을 숫자로 처리**를 선택하여 열을 문자가 아닌 숫자로 가져옵니다. 마침표, 유니코드 점, NaN 또는 빈 문자열이 있으면 결측값이 있는 것일 수 있습니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
2 자리 연도 규칙	4자리 대신 2자리 연도가 있는 날짜를 가져오는 데 사용할 규칙을 선택합니다. 2000-2099가 기본 설정입니다. 이러한 규칙에 대한 자세한 내용은 "데이터 가져오기" 장의 "2 자리 연도 규칙"(87 페이지)에서 확인하십시오.
압축 시도	텍스트 파일을 압축하는 데 사용되는 옵션을 선택합니다. 다음 옵션을 사용할 수 있으며 기본적으로 모두 선택 취소되어 있습니다. • 숫자 열 • 문자 열 • 목록 확인 허용 **참고 :** 이 기능을 사용하려면 전체 파일을 스캔해야 합니다 .
둘러싼 인용 부호 제거	텍스트 파일의 데이터를 둘러싼 인용 부호를 제거하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
아포스트로피를 인용 부호로 인식	아포스트로피를 인용 부호로 처리하고 생략하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다. **참고 :** 데이터 필드를 인용 부호로 묶지 않고 아포스트로피로 묶는 비표준 소스에서 데이터를 가져오는 경우가 아니면 이 옵션을 사용하지 않는 것이 좋습니다 .

표 13.9 텍스트 파일의 가져오기 설정에 대한 환경 설정 (계속)

환경 설정	설명
지역 설정 사용	텍스트 파일을 가져올 때 운영 체제의 지역 설정을 사용하려면 이 옵션을 선택합니다. • 이 옵션이 선택 취소되어 있으면 (기본값) 소수점으로 마침표를 사용하고 값 구분 기호로 쉼표를 사용하는 파일을 올바르게 가져올 수 있습니다 . • 소수점으로 쉼표를 사용하고 다른 값 구분 기호를 사용하는 파일의 경우 이 옵션을 선택하면 텍스트를 올바르게 가져올 수 있습니다 . " 텍스트 데이터 파일 " 가져오기 환경 설정에서 값 구분 기호를 지정해야 합니다 .

표 13.10 텍스트 파일 내보내기 설정에 대한 환경 설정

환경 설정	설명
테이블 머리글 내보내기	데이터 테이블을 텍스트 파일로 저장할 때 열 이름을 포함하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
모든 열 이름에 인용 부호 추가	열 이름 양쪽에 인용 부호를 삽입하려면 이 옵션을 선택합니다. CSV 보다 요구 사항이 까다로운 프로그램으로 데이터를 내보내는 데 사용됩니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
모든 문자 값에 인용 부호 추가	문자 값 양쪽에 인용 부호를 삽입하려면 이 옵션을 선택합니다. CSV 보다 요구 사항이 까다로운 프로그램으로 데이터를 내보내는 데 사용됩니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
모든 숫자 값에 인용 부호 추가	숫자 값 양쪽에 인용 부호를 삽입하려면 이 옵션을 선택합니다. CSV 보다 요구 사항이 까다로운 프로그램으로 데이터를 내보내는 데 사용됩니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
필드 끝	텍스트 데이터를 내보낼 때 필드 끝을 나타내는 구분자로 사용할 문자를 하나 이상 선택합니다. 쉼표가 기본 설정입니다. 목록에 없는 구분자를 지정하려면 **기타** 옵션을 선택하고 문자를 입력합니다.

표 13.10 텍스트 파일 내보내기 설정에 대한 환경 설정 (계속)

환경 설정	설명
행 끝	줄 (행) 끝을 나타내는 구분자로 사용할 문자를 하나 이상 선택합니다. <CR>+<LF>가 기본 설정입니다. 목록에 없는 구분자를 지정하려면 **기타** 옵션을 선택하고 문자를 입력합니다.

타사 데이터의 환경 설정

이 환경 설정에서는 Microsoft Excel 및 SPSS 파일과 같은 타사 데이터에 대한 설정을 지정합니다.

그림 13.11 타사 데이터 환경 설정

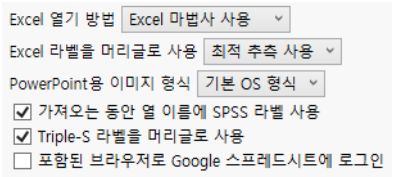


표 13.11 타사 데이터 페이지의 환경 설정

환경 설정	설명
Excel 열기 방법	JMP가 Microsoft Excel 워크시트를 가져올 방법을 결정하는 옵션 중 하나를 선택합니다. **Excel 마법사 사용** 데이터를 가져오기 전에 Microsoft Excel 워크시트를 미리 보고 설정을 수정할 수 있습니다. Windows 의 경우 기본 설정입니다. macOS 의 경우에는 모든 .xlsx 파일이 Excel 마법사에서 열립니다. **모든 시트 열기** 통합 문서의 모든 워크시트를 엽니다. macOS 의 경우, .xls 파일에만 이 환경 설정이 적용됩니다. **개별 워크시트 선택** 통합 문서에서 열려는 워크시트를 선택할 수 있습니다. macOS 의 경우, .xls 파일에만 이 환경 설정이 적용됩니다. macOS 에서는 Excel(.xls) 파일을 열 때도 이 옵션이 제공됩니다. Windows에서는 **파일 > 열기**를 선택하고, 통합 문서를 선택하고, **열기** 버튼의 화살표를 클릭한 후, **선택한 워크시트 열기** 또는 **Excel 마법사 사용**을 선택할 수도 있습니다.

표 13.11 타사 데이터 페이지의 환경 설정 (계속)

환경 설정	설명
Excel 라벨을 머리글로 사용	Excel 파일을 가져올 때 JMP에서 Excel 라벨 이름을 열 머리글로 사용할 수 있도록 하려면 이 옵션을 선택합니다. 사용 가능한 설정은 다음과 같습니다. • 최적 추측 사용 (기본값) • 항상 • 안 함
PowerPoint 용 이미지 형식	Microsoft PowerPoint로 내보낸 이미지의 기본 파일 형식을 선택합니다. Windows는 macOS에서 생성된 기본 PDF 그래픽을 지원하지 않습니다. macOS는 Windows에서 생성된 기본 EMF 그래픽을 지원하지 않습니다. 플랫폼 간 호환성을 위해서는 PNG 또는 JPEG 그래픽 형식을 지정합니다. 추가 파일 형식은 스크립팅을 통해 사용할 수 있습니다. 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.
가져오는 동안 열 이름에 SPSS 라벨 사용	SPSS 파일을 가져올 때 JMP에서 SPSS 라벨 이름을 열 머리글로 사용할 수 있도록 하려면 이 옵션을 선택합니다.
Triple-S 라벨을 머리글로 사용	JMP에서 Triple-S 라벨 이름을 열 머리글로 사용할 수 있도록 하려면 이 옵션을 선택합니다.
포함된 브라우저로 Google 스프레드시트에 로그인	Google 스프레드시트에 로그인하는 데 기본 시스템 브라우저 대신 내장 JMP 브라우저를 사용하려면 이 옵션을 선택합니다.

Windows 환경 설정

Windows 관련 환경 설정에서는 메뉴 및 도구 모음 자동 숨기기와 기본 그래픽 형식 선택을 비롯하여 Windows 컴퓨터에 대한 설정을 사용자 정의합니다.

그림 13.12 Windows 관련 환경 설정

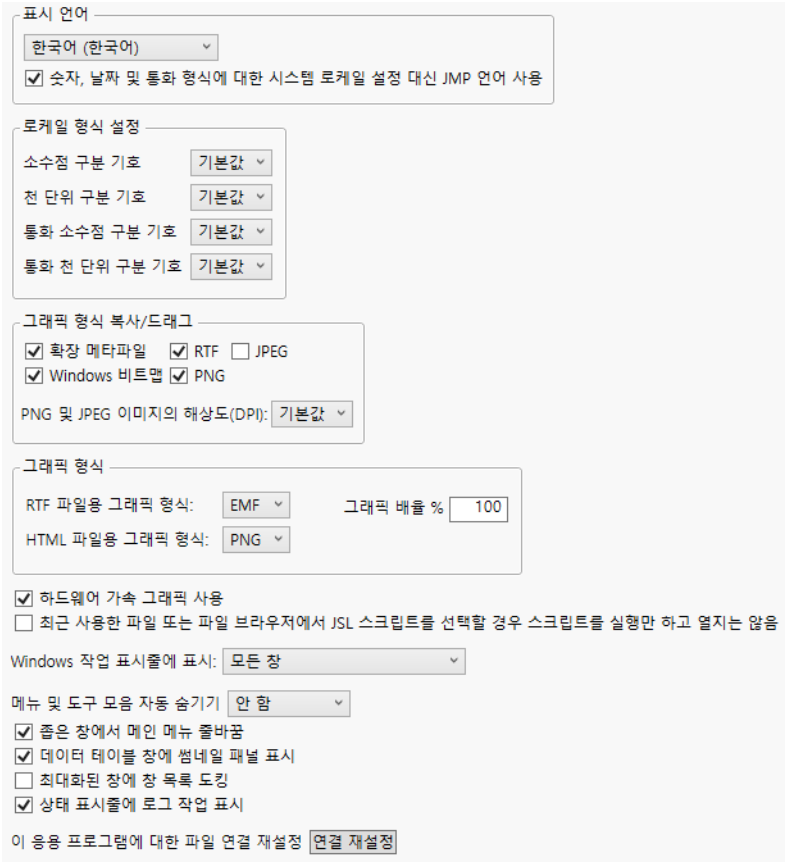


표 13.12 Windows 관련 페이지의 환경 설정

환경 설정	설명
표시 언어	JMP를 실행하는 데 사용할 언어를 선택합니다. 일반적으로 숫자, 날짜 및 통화 형식에 대한 설정은 운영 체제의 로케일 설정에 따라 결정됩니다. 이러한 형식을 결정하는 데 선택한 언어를 사용하려면 언어 메뉴 아래의 옵션을 선택합니다. **파일 > 열기**와 같이 운영 체제에서 제공되는 창에 표시되는 텍스트는 언어 설정의 변경 사항을 반영하지 않습니다. **참고 :** • 표시 언어를 변경한 후에는 JMP 를 닫았다가 다시 시작해야 언어 설정이 완전히 적용됩니다 . • 사용 가능한 언어는 JMP 설치 프로그램에서 선택한 언어에 따라 결정됩니다. 그 외의 언어를 사용하려면 JMP 설치 프로그램을 다시 실행하고 " 수정 " 을 선택한 다음 언어를 선택해야 합니다 .
숫자, 날짜 및 통화 형식에 대한 시스템 로케일 설정 대신 JMP 언어 사용	JMP 표시 언어를 사용하여 숫자, 날짜 및 통화 형식을 지정합니다.
로케일 형식 설정	숫자 및 통화 값에 사용되는 소수점 및 천 단위 구분 기호를 지정합니다. 기본 설정에서는 시스템 로케일과 "표시 언어" 환경 설정의 설정을 사용합니다. 기본이 아닌 설정을 지정하면 해당 설정이 "표시 언어" 환경 설정과 시스템 로케일을 재정의합니다.
그래픽 형식 복사 / 드래그	JMP에서 다른 응용 프로그램으로 그래픽을 복사하여 붙여넣거나 드래그하여 놓을 때 사용할 그래픽 형식을 하나 이상 선택합니다. 확장 메타파일, Windows 비트맵, RTF 및 PNG가 기본적으로 선택되어 있습니다. PNG 및 JPEG 이미지의 해상도는 기본값입니다.

표 13.12 Windows 관련 페이지의 환경 설정

환경 설정	설명

PNG 및 JPEG 이미지의 해상도 (DPI)	그래픽 형식을 복사할 때 사용할 DPI를 지정합니다. "기본값" (96) 또는 "300" 중에서 선택할 수 있습니다. 늘리거나 간행물에 포함하거나 인쇄해야 하는 이미지에는 "300"이 더 적합합니다. 하지만 이 설정을 선택하면 메모리 사용량이 늘어나고 이미지의 경우 생성 속도가 느려집니다.
	
	다음 스크립트를 사용하면 기본 DPI를 설정할 수 있습니다. 원하는 경우 숫자를 다른 DPI로 변경한 후 스크립트를 실행합니다.
	
	`Pref( Save Image DPI( 600 ) );`
	
	위의 스크립트를 실행한 후 기본 DPI를 확인하려면 다음 스크립트를 실행합니다.
	
	`Get Preferences( Save Image DPI );`
	
	더 높은 DPI를 지정하는 방법에 대한 자세한 내용은 "데이터 저장 및 공유" 장의 "내보낸 그래픽에 대한 그래픽 DPI 설정"(550 페이지)에서 확인하십시오.

그래픽 형식	JMP 보고서를 RTF 파일(EMF, PNG 또는 JPEG) 또는 HTML 파(PNG, JPEG, GIF 및 SVG)일로 저장할 때 그래픽에 사용할 형식을 선택합니다. "그래픽 배율 %" 상자에는 다른 응용 프로그램에서 그래픽을 표시할 크기를 백분율로 입력합니다. 선택한 응용 프로그램의 일부 버전에서는 이 기능이 작동하지 않을 수도 있습니다. 100%가 기본 설정입니다.
하드웨어 가속 그래픽 사용	비디오 성능을 높이기 위해 컴퓨터의 GPU(그래픽 처리 장치)를 활용합니다. JMP 창을 그릴 때 컴퓨터 속도가 느린 경우 이 옵션을 선택하는 것이 좋습니다. 이 기능은 아직 시험 단계인 기능입니다.
최근 사용한 파일 또는 파일 브라우저에서 JSL 스크립트를 선택할 경우 스크립트를 실행만 하고 열지는 않음	모든 스크립트가 스크립트를 열 때 실행되도록 하려면 이 옵션을 선택합니다. 이 옵션이 선택되어 있으면 스크립트에 대한 스크립트 창은 열리지 않습니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
Windows 작업 표시줄에 표시	Windows 작업 표시줄에 표시할 JMP 창을 선택합니다. 기본적으로 "모든 창"이 선택되어 있습니다. 기본 JMP 창과 데이터 테이블 창을 표시하도록 선택할 수도 있습니다.

표 13.12 Windows 관련 페이지의 환경 설정

환경 설정	설명
메뉴 및 도구 모음 자동 숨기기	메뉴 및 도구 모음을 숨길 시기를 결정하는 규칙을 선택합니다. 메뉴 및 도구 모음은 보고서 및 레이아웃 창에서 항상 숨겨집니다. 다른 창에서는 메뉴 및 도구 모음이 창 높이를 기준으로 표시됩니다. 기본 설정은 "안 함"입니다.
좁은 창에서 메인 메뉴 줄바꿈	창 너비가 메뉴 너비보다 좁을 경우 메뉴를 줄바꿈해서 여러 줄로 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
데이터 테이블 창에 썸네일 패널 표시	데이터 테이블 하단에 보고서의 썸네일 패널을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
최대화된 창에 창 목록 도킹	JMP 창을 최대화할 경우 "창 목록"을 자동으로 도킹하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
상태 표시줄에 로그 작업 표시	로그 창에서 활동이 감지되었을 때 나타나는 상태 표시줄에 아이콘을 표시하려면 이 옵션을 선택합니다. 그런 다음 아이콘을 클릭하여 로그 창을 맨 앞으로 가져올 수 있습니다. 이 옵션은 기본적으로 선택되어 있습니다.
이 응용 프로그램에 대한 파일 연결 재설정	모든 JMP 파일 유형을 실행 중인 JMP 버전과 연결하려면 이 버튼을 클릭합니다.

Mac OS 환경 설정

"Mac OS 설정" 환경 설정에서는 그래픽 형식과 파일 연결을 비롯하여 macOS 컴퓨터에 대한 설정을 사용자 정의합니다.

그림 13.13 Mac OS 설정의 환경 설정

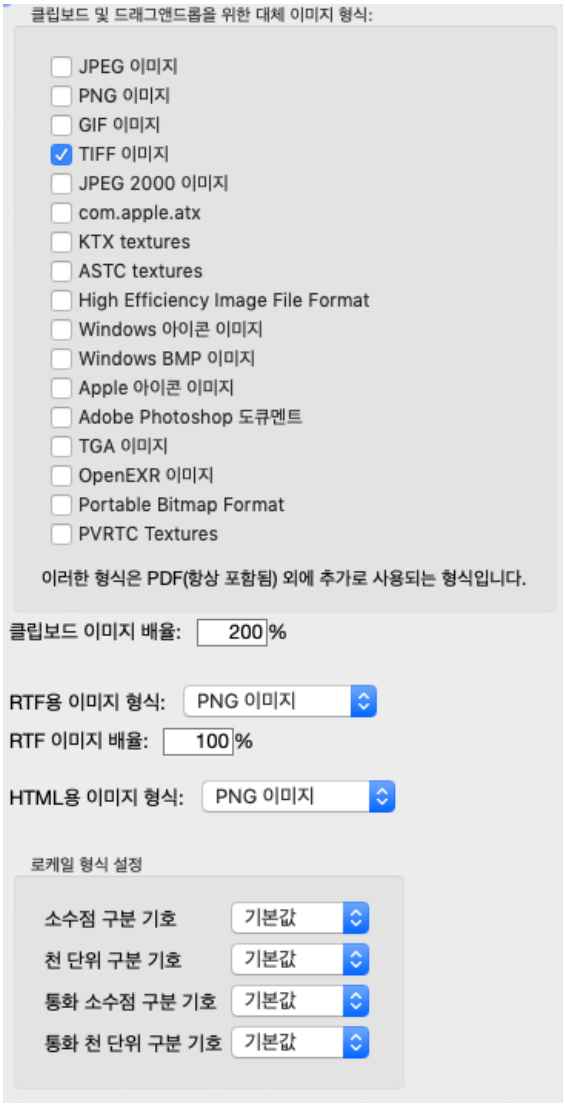


표 13.13 Mac OS 설정 페이지의 환경 설정

환경 설정	설명
클립보드 및 드래그앤드롭을 위한 대체 이미지 형식	JMP에서 다른 응용 프로그램으로 그래픽을 복사하여 붙여넣거나 드래그하여 놓을 때 사용할 그래픽 형식을 하나 이상 선택합니다. 기본 설정은 TIFF입니다.
클립보드 이미지 배율	다른 응용 프로그램에서 그래픽을 표시할 크기를 백분율로 입력합니다. 기본 설정은 200%입니다.
RTF 용 이미지 형식	JMP 보고서를 RTF 파일로 저장할 때 그래픽에 사용할 형식을 선택합니다. 기본 설정은 PNG입니다.
RTF 이미지 배율	RTF 문서에서 그래픽을 표시할 크기를 백분율로 입력합니다. 기본 설정은 100%입니다.
HTML 용 이미지 형식	JMP 보고서를 HTML 파일로 저장할 때 그래픽에 사용할 형식을 선택합니다. 기본 설정은 PNG입니다.
로케일 형식 설정	숫자 및 통화 값에 사용되는 소수점 및 천 단위 구분 기호를 지정합니다. 기본 설정에서는 시스템 로케일을 사용합니다. 기본이 아닌 설정을 지정하면 해당 설정이 시스템 로케일을 재정의합니다.

글꼴 환경 설정

"글꼴" 환경 설정에서는 보고서, 데이터 테이블 및 스크립트의 글꼴, 텍스트 크기 및 글꼴 스타일 같은 모양을 사용자 정의합니다.

그림 13.14 글꼴 환경 설정

표 13.14 글꼴 페이지의 환경 설정

환경 설정	설명
텍스트	JMP 분석 보고서의 텍스트 부분에 사용할 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
머리글	분석 보고서 및 데이터 테이블에서 열의 머리글에 사용할 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트, 굵게입니다.
제목	모든 제목 표시줄에 표시되는 제목의 글꼴을 설정합니다. 기본 설정은 Segoe UI, 11포인트, 굵게입니다.

표 13.14 글꼴 페이지의 환경 설정 (계속)

환경 설정	설명
작게	데이터 격자의 왼쪽 상단에서 열 및 행 번호를 표시하는 데 사용되는 작은 텍스트의 글꼴을 설정합니다. 기본 설정은 Segoe UI, 7포인트입니다.
모노	JSL(JMP 스크립트 언어) 편집기와 스크립트 명령용 편집기 상자에 사용되는 글꼴을 설정합니다. 기본 설정은 Consolas, 10포인트입니다.
계산식 편집기	계산식 편집기에 입력되는 표현식의 글꼴을 설정합니다. 기본 설정은 Segoe UI, 11포인트입니다.
주석	주석에 사용되는 기본 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
축	축 눈금 라벨에 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
표식	영숫자 표식에 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
축 제목	축 제목에 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
그래프 라벨	그래프에서 표식 라벨 또는 막대 라벨과 같은 데이터 라벨에 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
범례	범례 제목, 값 및 기타 범례 텍스트에 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
그래프 제목	부제목 또는 각주를 제외하고 주 그래프 제목에만 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 11포인트, 굵게입니다.
캡션	캡션 요소 및 캡션과 유사한 텍스트(예: 적합선 통계량 및 상자 그림의 5가지 숫자 요약)에 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 8포인트입니다.
데이터 테이블	데이터 테이블에 데이터를 표시하는 데 사용되는 글꼴을 설정합니다. 기본 설정은 Segoe UI, 9포인트입니다.
글꼴군	가변폭 글꼴을 선택합니다. 선택하는 글꼴은 스크립트에 사용되는 "모노"를 제외하고 모든 설정에 즉시 적용됩니다.
그리스 문자 사용	그리스 문자를 풀어 쓰지 않고 그대로 사용하려면 이 옵션을 선택합니다(예: pi 대신 π 사용). 이 옵션은 기본적으로 선택되어 있습니다.
수학 기호 사용	수학 기호의 단순한 텍스트 표현 대신 수학 기호를 사용하려면 이 옵션을 선택합니다(예: +/- 대신 ± 사용). 이 옵션은 기본적으로 선택됩니다.

통신 환경 설정

(Windows 에만 해당) " 통신 " 환경 설정에서는 외부 소스의 데이터를 읽기 위한 설정을 사용자 정의합니다 . 이러한 설정은 통신 기기를 사용하는 경우에만 지정해야 합니다 .

그림 13.15 통신 환경 설정

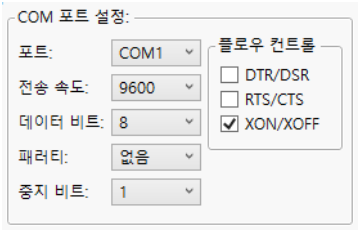


표 13.15 통신 페이지의 환경 설정

환경 설정	설명
포트	데이터 소스가 사용하는 포트를 선택합니다.
전송 속도	데이터 소스의 전송 속도를 설정합니다.
데이터 비트	데이터 비트 수(7 또는 8)를 설정합니다.
패러티	패러티 비트를 설정합니다. "없음"이 기본 설정입니다.
중지 비트	중지 비트(1 또는 2)를 설정합니다.
플로우 컨트롤	플로우 컨트롤을 설정합니다. XON/XOFF 가 기본 설정입니다.

인터넷 옵션의 환경 설정

"인터넷 옵션"에서는 인터넷에서 파일을 열기 위한 설정을 지정합니다.

그림 13.16 인터넷 옵션

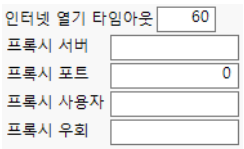


표 13.16 인터넷 옵션 페이지의 환경 설정

환경 설정	설명
인터넷 열기 타임아웃	"인터넷 열기"를 사용하여 웹 페이지를 여는 경우 지정된 시간(초) 내에 파일이 열리지 않으면 JMP는 이를 오류로 처리하고 가져오기를 중지합니다. 기본 설정은 60초입니다.
프록시 서버	프록시 호스트를 지정합니다. `http`(또는 `https`)를 함께 지정해야 합니다. 프로토콜을 사용하지 않으면 `http`로 간주됩니다.
프록시 포트	프록시 호스트에 연결하는 데 사용되는 포트를 지정합니다. 아무것도 지정하지 않으면 포트 443(`http`의 경우) 및 포트 1080(`https` 등 그 외 모든 경우)이 사용됩니다. 기본 설정은 0입니다.
프록시 사용자	프록시의 선택적 사용자를 지정합니다. 모든 프록시 설정에 사용자가 필요하지는 않습니다.
프록시 우회	우회할(즉, 프록시를 거치지 않을) 호스트가 포함된 선택적 쉼표 구분 목록을 지정합니다. 회사 방화벽 뒤에 있어 프록시가 필요하지 않은 호스트를 제외할 수 있습니다.

파일 위치의 환경 설정

(Windows에만 해당) "파일 위치" 환경 설정에서는 JMP 시스템 파일의 기본 위치를 설정합니다.

그림 13.17 파일 위치 환경 설정

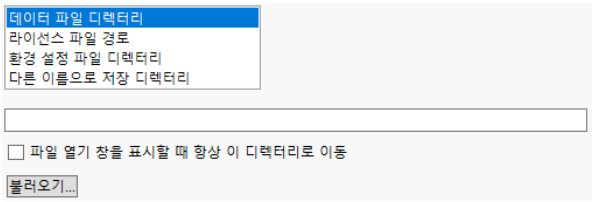


표 13.17 파일 위치 페이지의 환경 설정

환경 설정	설명
데이터 파일 디렉터리	JMP가 데이터 테이블 등의 데이터 파일을 검색하는 폴더를 변경합니다.
라이선스 파일 경로	JMP가 JMP 라이선스 파일을 검색하는 폴더를 변경합니다.
환경 설정 파일 디렉터리	JMP가 환경 설정 정보를 검색하는 폴더를 변경합니다.
다른 이름으로 저장 디렉터리	"파일">"다른 이름으로 저장"을 선택할 때 JMP가 데이터 파일을 저장하는 폴더를 변경합니다.
파일 열기 창을 표시할 때 항상 이 디렉터리로 이동	"파일">"열기"를 선택할 때 JMP가 항상 지정된 경로를 사용하도록 하려면 선택합니다.

JSL 스크립트 편집기의 환경 설정

" 스크립트 편집기 " 환경 설정에서는 탭 너비 , 구문 색상 , 툴팁을 비롯하여 JSL 스크립트 편집기 및 스크립트 편집기 상자의 모양을 사용자 정의합니다 .

그림 13.18 스크립트 편집기 환경 설정

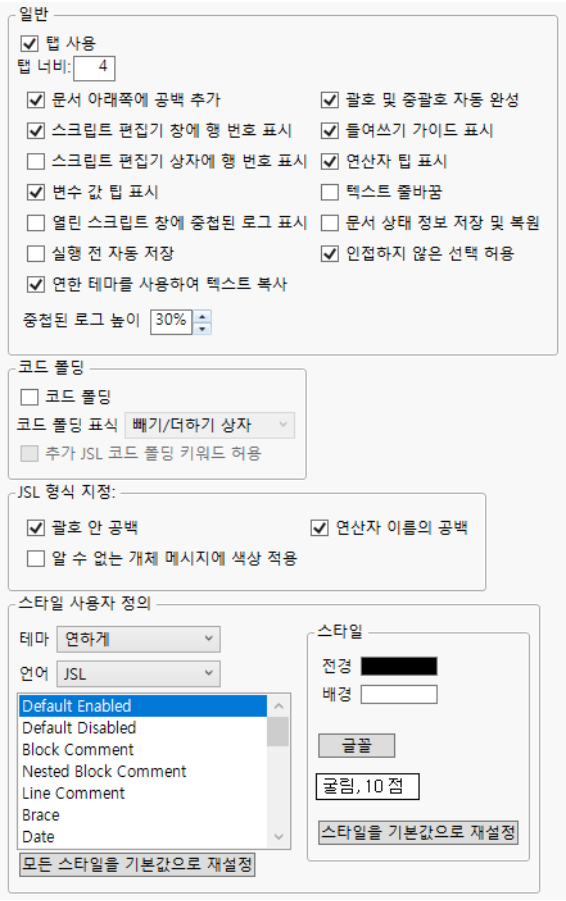


표 13.18 스크립트 편집기 페이지의 환경 설정

환경 설정	설명
탭 사용	스크립트에 탭을 사용할 수 있도록 하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다. 탭 입력 시 공백으로 대체하려면 이 옵션을 선택 해제합니다.

표 13.18 스크립트 편집기 페이지의 환경 설정 (계속)

환경 설정	설명
탭 너비	탭 사용 시 들여쓸 공백 수를 입력합니다. 탭을 비활성화한 경우 탭을 입력할 때마다 지정한 수만큼의 공백으로 대체됩니다. 기본 설정은 4입니다.
문서 아래쪽에 공백 추가	스크립트의 마지막 빈 줄에서 위로 스크롤할 수 있도록 하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 Windows에서는 선택되어 있고 macOS에서는 선택 취소되어 있습니다.
괄호 및 중괄호 자동 완성	스크립트 편집기에서 여는 소괄호, 중괄호 및 대괄호를 입력할 경우 그에 해당하는 닫는 괄호가 자동으로 추가되도록 하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
스크립트 편집기 창에 행 번호 표시	스크립트 편집기의 왼쪽에 행 번호를 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
들여쓰기 가이드 표시	들여쓰기를 표시하는 희미한 세로선을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
스크립트 편집기 상자에 행 번호 표시	"스크립트 인덱스" 등에서 스크립트 편집기 상자의 왼쪽에 행 번호를 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 해제되어 있습니다.
연산자 팁 표시	JSL 연산자에 대한 툴팁을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
변수 값 팁 표시	변수 값에 대한 툴팁을 표시하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
텍스트 줄바꿈	스크립트 편집기에서 항상 텍스트를 줄바꿈하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 해제되어 있습니다.
열린 스크립트 창에 중첩된 로그 표시	스크립트를 편집하거나 실행할 때 스크립트 창에 포함된 로그 창이 표시되도록 하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 해제되어 있습니다.
문서 상태 정보 저장 및 복원	접히거나 펼쳐진 코드의 상태를 저장하고, 스크립트를 다시 열 때 해당 상태를 복원합니다.
실행 전 자동 저장	스크립트를 실행하기 전에 자동으로 저장합니다. 이 옵션은 기본적으로 해제되어 있습니다.
인접하지 않은 선택 허용	인접하지 않은 텍스트를 선택할 수 있습니다. Ctrl 키 (Windows) 또는 Command 키 (macOS) 를 누른 채 커서로 텍스트를 드래그합니다. 다른 텍스트를 더 추가하려면 이 작업을 계속합니다. 이 옵션은 기본적으로 선택되어 있습니다.

표 13.18 스크립트 편집기 페이지의 환경 설정 (계속)

환경 설정	설명
연한 테마를 사용하여 텍스트 복사	코드의 배경 색상을 다른 문서에 붙여넣을지 여부를 결정합니다. 이 옵션은 기본적으로 선택되어 있습니다. 즉, 스크립트 편집기 환경 설정에서 연한 테마를 선택하면 코드가 연한 배경과 함께 다른 문서에 붙여넣어집니다.
중첩된 로그 높이	스크립트 창에 포함된 로그의 높이를 지정합니다. 기본 설정은 30% 입니다.
코드 폴딩	스크립트 편집기에서 사용자가 정의하는 함수 및 Expr() 블록의 시작과 끝을 표시하는 코드 폴딩 표식을 사용하려면 이 옵션을 선택합니다. 이 표식이 사용된 코드 블록은 펼치거나 접을 수 있습니다. 이 옵션은 기본적으로 해제되어 있습니다. JSL 코드 폴딩 표식 메뉴를 사용하여 표식의 모양을 선택할 수도 있습니다. 자세한 내용은 Scripting Guide의 에서 확인하십시오.
추가 JSL 코드 폴딩 키워드 허용	스크립트 편집기에서 폴딩 표식에 추가 키워드를 사용할 수 있도록 하려면 이 옵션을 선택합니다. 자세한 내용은 Scripting Guide의 에서 확인하십시오.
괄호 안 공백	자동으로 형식이 지정되는 스크립트에 대해 스크립트 편집기에서 소괄호, 중괄호 및 대괄호와 괄호 내용 사이에 공백이 추가되도록 하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 설정되어 있습니다.
연산자 이름의 공백	스크립트 편집기에서 연산자 이름 내의 단어 사이에 공백이 추가되도록 하려면 이 옵션을 선택합니다. 예를 들어 이 옵션을 설정하면 NewWindow 대신 New Window가 사용됩니다. 이 옵션은 기본적으로 선택되어 있습니다.
알 수 없는 개체 메시지에 색상 적용	알 수 없는 개체 메시지에 항상 색상을 적용하려면 이 옵션을 선택합니다. 알 수 없는 개체 메시지는 "알 수 없는 메시지에 색상 적용"에 지정된 색상으로 표시됩니다. 이 옵션은 기본적으로 해제되어 있습니다.
스타일 사용자 정의	JSL 스크립트, SAS 스크립트, 텍스트 탐색기 사용자 정의 창, 텍스트 파일 등에 표시되는 코드 요소의 색상과 글꼴을 선택합니다. "모든 스타일을 기본값으로 재설정"은 목록에 있는 모든 항목의 원래 색상을 복원합니다. "스타일을 기본값으로 재설정"은 오른쪽에 있는 선택된 항목의 원래 색상 및 글꼴을 복원합니다. "글꼴" 환경 설정을 재설정해도 이 페이지에서 선택한 글꼴에는 영향이 없습니다.

SAS Integration 의 환경 설정

"SAS Integration" 환경 설정에서는 SAS 서버와의 연동을 위한 기본 설정을 사용자 정의합니다. SAS Integration 기능을 사용하는 방법에 대한 자세한 내용은 "데이터 가져오기" 장의 "SAS에서 데이터 가져오기"(132 페이지)에서 확인하십시오.

참고: "SAS 환경" 옵션은 "SAS 서버 버전" 목록에서 SAS 9.3 또는 9.4를 선택한 후에 표시됩니다.

그림 13.19 SAS Integration 환경 설정

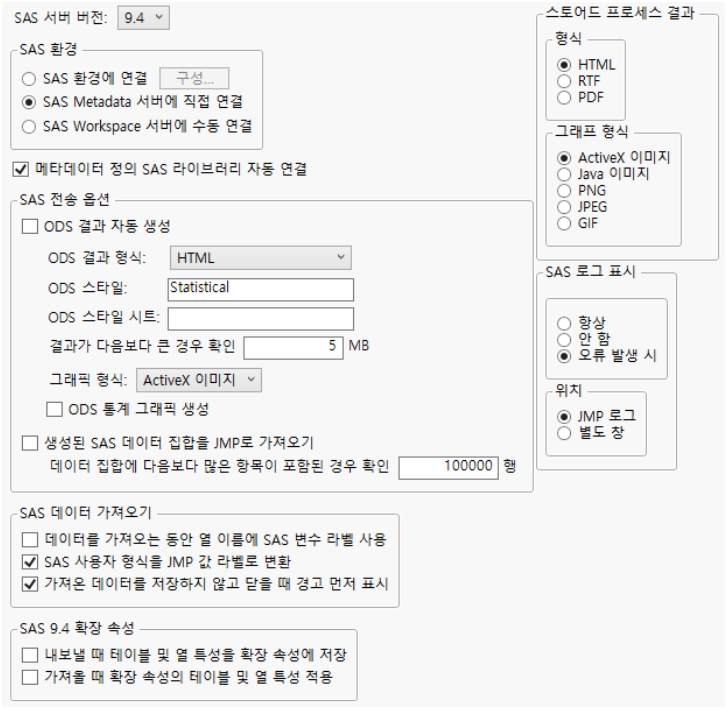


표 13.19 SAS Integration 페이지의 환경 설정

환경 설정	설명
SAS 서버 버전	SAS 서버의 기본 버전을 선택합니다. "서버 연결" 창에서도 이 설정을 변경할 수 있습니다. 자세한 내용은 "데이터 가져오기" 장의 "JMP에서 SAS Metadata 서버에 연결"(136 페이지)에서 확인하십시오. 이 옵션은 Windows에서 로컬 SAS에 연결할 때는 적용되지 않습니다. 기본 설정은 9.4입니다.

표 13.19 SAS Integration 페이지의 환경 설정 (계속)

환경 설정	설명
SAS 환경에 연결	항상 SAS 환경에 연결하려면 이 옵션을 선택하고 **구성**을 클릭하여 해당 위치의 URL을 구성합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
SAS Metadata 서버에 연결	항상 SAS Metadata 서버에 연결하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다. 사용 가능한 SAS Metadata 서버가 없고 대신 SAS Workspace 서버에 직접 연결하려면 이 옵션을 선택 해제합니다.
SAS Workspace 서버에 수동 연결	SAS Workspace 서버에 수동으로 연결하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
메타데이터 정의 SAS 라이브러리 자동 연결	모든 JMP 세션에서 SAS Workspace 서버에 연결할 때 자동으로 메타데이터 정의 SAS 라이브러리에 연결하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다. Workspace 서버에 포함된 메타데이터 정의 SAS 라이브러리가 매우 많은 경우 서버에 빠르게 연결하려면 이 옵션을 선택 취소합니다.
ODS 결과 자동 생성	ODS(Output Delivery System) 결과를 생성하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
ODS 결과 형식	ODS 보고서의 형식을 선택합니다. 기본 설정은 HTML입니다.
ODS 스타일	ODS 보고서의 스타일 이름을 입력합니다. 기본 설정은 "통계" 입니다.
ODS 스타일 시트	ODS 보고서의 스타일 시트 이름을 입력합니다.
결과가 다음보다 큰 경우 확인 __ MB	사용자에게 작업을 계속할지 취소할지 묻는 확인 메시지를 트리거할 기준 MB 수를 입력합니다. 5가 기본 설정입니다.
그래픽 형식	ODS 보고서의 그래픽 형식을 선택합니다. "ActiveX 이미지"가 기본 설정입니다.
ODS 통계 그래픽 생성	ODS 보고서에 통계 그래픽을 포함하려면 이 옵션을 선택합니다.
생성된 SAS 데이터 집합을 JMP 로 가져오기	생성된 데이터 집합을 자동으로 JMP로 가져오려면 이 옵션을 선택합니다.
데이터 집합에 다음보다 많은 항목이 포함된 경우 확인 __ 행	사용자에게 작업을 계속할지 취소할지 묻는 확인 메시지를 트리거할 기준 행 수를 입력합니다. 기본 설정은 100,000입니다.

표 13.19 SAS Integration 페이지의 환경 설정 (계속)

환경 설정	설명
데이터를 가져오는 동안 열 이름에 SAS 변수 라벨 사용	SAS 데이터 집합을 JMP 데이터 테이블로 가져올 때 SAS 데이터 집합의 열 라벨을 JMP 데이터 테이블의 열 이름으로 사용하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
SAS 사용자 형식을 JMP 값 라벨로 변환	SAS 데이터 집합을 JMP 데이터 테이블로 가져올 때 SAS 열 형식의 정보를 사용하여 JMP 값 라벨을 설정하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
가져온 데이터를 저장하지 않고 닫을 때 경고 먼저 표시	SAS 데이터 집합을 JMP로 가져온 후 저장하지 않은 상태에서 닫으려고 할 때 SAS 데이터 집합을 저장하라는 메시지를 트리거하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택되어 있습니다.
내보낼 때 테이블 및 열 특성을 확장 속성에 저장	SAS 9.4 데이터 내보내기 시 확장 속성 지원을 허용하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
가져올 때 확장 속성의 테이블 및 열 특성 적용	SAS 9.4 데이터 가져오기 시 확장 속성 지원을 허용하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
스토어드 프로세스 결과: 형식	HTML(기본값), RTF 또는 PDF 중에서 보고서 형식을 선택합니다.
스토어드 프로세스 결과: 그래프 형식	ActiveX 이미지(기본값, Windows에서 실행되는 SAS 서버에서만 사용 가능), Java 이미지, PNG, JPEG 또는 GIF 중에서 그래프 형식을 선택합니다.
SAS 로그 표시	"항상", "안 함" 또는 "오류 발생 시"(기본값)를 선택하여 SAS 로그가 표시되는 시기를 설정합니다.
위치	SAS 로그 정보의 위치(JMP 로그 창 또는 별도의 SAS 로그 창)를 선택합니다. 기본 설정은 JMP 로그 창입니다.

JMP 업데이트의 환경 설정

"JMP 업데이트" 환경 설정에는 현재 사용 중인 JMP 버전, 마지막 업데이트 시간, JMP 업데이트 자동 확인 여부 및 간격이 표시됩니다.

그림 13.20 JMP 업데이트 환경 설정

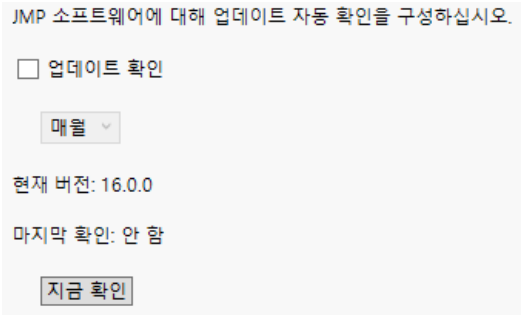


표 13.20 JMP 업데이트 페이지의 환경 설정

환경 설정	설명
업데이트 확인	JMP에서 자동으로 업데이트를 확인하려면 이 옵션을 선택합니다. 이 옵션을 선택할 경우 확인 간격(매일, 매주 또는 매월)을 지정합니다.
현재 버전	현재 실행 중인 JMP 버전입니다.
마지막 확인	마지막으로 업데이트를 확인한 시간입니다.
지금 확인	즉시 업데이트를 확인하려면 이 버튼을 클릭합니다.

JSL 디버거의 환경 설정

"JSL 디버거 " 환경 설정에서는 행 번호를 표시하거나 숨기고 , 스크립트에 중단점을 생성하고 , 경고를 표시할 수 있습니다 .

그림 13.21 JSL 디버거 환경 설정

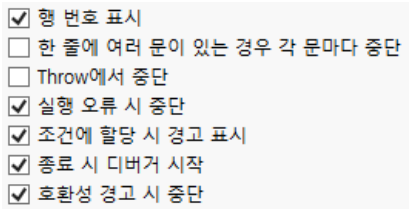


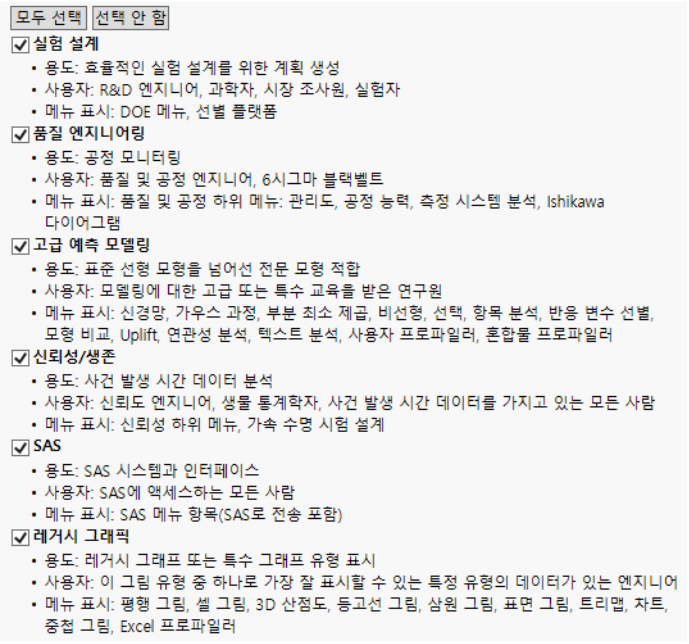
표 13.21 JSL 디버거 페이지의 환경 설정

환경 설정	설명
행 번호 표시	디버거에서 스크립트의 행 번호를 표시하거나 숨깁니다. 이 옵션은 기본적으로 선택되어 있습니다.
한 줄에 여러 문이 있는 경우 각 문마다 중단	한 줄에 표현식이 여러 개 있을 때 각 표현식 사이에서 스크립트 실행을 중지합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
Throw 에서 중단	`Try()` 함수에서 예외 오류가 발생할 때 중단합니다. 예를 들어 `Try()` 함수의 예외 오류는 `Throw()` 함수에 의해 발생합니다. 이 경우 나머지 코드를 계속 진행하지 않고 `Throw()` 에서 디버거가 중단됩니다. 따라서 스크립트의 어느 부분에서 문제가 발생했는지 파악한 후 디버깅으로 돌아갈 수 있습니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
실행 오류 시 중단	오류가 발생하면 디버거를 닫지 않고 스크립트 실행을 중지합니다. 이 옵션은 기본적으로 선택됩니다.
조건에 할당 시 경고 표시	중단점 조건에 대한 할당 표현식을 입력할 때 경고를 표시하려면 이 옵션을 선택합니다. 예를 들어 중단점 조건에서 `x == 1` 대신 `x = 1`을 지정하면 지정한 내용을 확인하라는 메시지가 표시됩니다. 이 옵션은 기본적으로 선택됩니다.
종료 시 디버거 시작	JMP 종료 시 디버거를 시작하려면 이 옵션을 선택합니다. 이 옵션은 기본적으로 선택됩니다.
호환성 경고 시 중단	JSL 호환성 문제가 발생할 경우 스크립트 실행을 중지합니다. 이 옵션은 기본적으로 선택됩니다.

메뉴 환경 설정

"메뉴" 환경 설정에서는 JMP 사용 방식에 따라 메뉴를 숨기거나 표시합니다. 이를 통해 메뉴 항목을 줄이면 JMP 인터페이스를 보다 간편하게 탐색하면서 효율적으로 사용할 수 있습니다. 예를 들어 실험을 설계하지 않는 경우에는 **실험 설계**를 선택 취소합니다. 다른 메뉴는 품질 공학, 고급 예측 모델링, 신뢰성 및 생존, SAS 옵션과 같은 관심 분야별로 그룹화되어 있습니다.

그림 13.22 메뉴 환경 설정



쿼리 빌더의 환경 설정

" 쿼리 빌더 " 환경 설정에서는 쿼리 빌더에서 수행하는 SQL 쿼리를 사용자 정의합니다 .

그림 13.23 쿼리 빌더 환경 설정

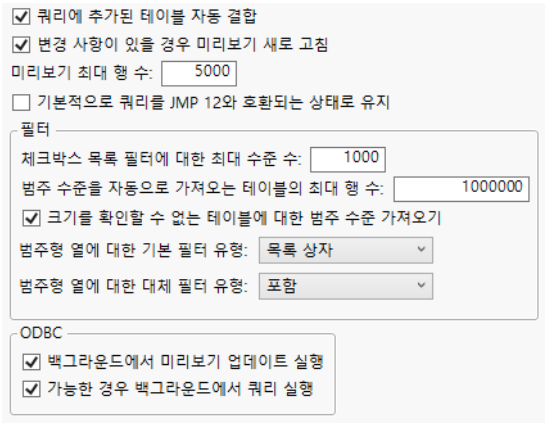


표 13.22 쿼리 빌더 페이지의 환경 설정

환경 설정	설명
쿼리에 추가된 테이블 자동 결합	쿼리 빌더에서 키와 공통 열 이름을 기준으로 데이터베이스 테이블을 자동으로 결합합니다. 이 옵션은 기본적으로 선택되어 있습니다. 큰 데이터베이스로 인한 메모리 문제를 방지하려면 이 옵션을 선택 취소합니다.
변경 사항이 있을 경우 미리보기 새로 고침	변경 사항이 발생하면 "쿼리 미리보기" 탭의 데이터 보기를 업데이트합니다. 이 옵션은 기본적으로 선택되어 있습니다.
미리보기 최대 행 수	"쿼리 미리보기" 탭에 표시되는 행의 개수를 제한합니다. 기본 개수는 5,000 개입니다.
기본적으로 쿼리를 JMP 12 와 호환되는 상태로 유지	쿼리를 JMP 12 와 호환되는 상태로 유지합니다. JMP 12 와 호환되는 것으로 표시된 쿼리를 JMP 13 이상에서 열면 환경 설정이 어떻게 설정되어 있든 상관없이 호환성 문제를 발생시키는 기능이 숨겨집니다.
체크박스 목록 필터에 대한 최대 수준 수	범주형 변수의 필터를 "체크박스 목록" 필터로 추가할지 여부를 결정합니다. 기본적으로 수준 수가 1,000 개를 초과하는 범주형 변수를 추가하면 해당 필터가 포함 필터로 추가됩니다.

표 13.22 쿼리 빌더 페이지의 환경 설정 (계속)

환경 설정	설명
범주 수준을 자동으로 가져오는 테이블의 최대 행 수	필터링된 열의 고유 범주 수준을 가져올 최대 행 수를 지정합니다. 최소값은 -1(제한 없음)입니다. 최대값은 10억입니다. 기본적으로 범주형 열의 행 수가 백만 개를 초과할 경우 JMP는 고유한 범주 수준을 자동으로 가져오지 않습니다.
크기를 확인할 수 없는 테이블에 대한 범주 수준 가져오기	행 수를 확인할 수 없는 경우 필터링된 열에 대해 범주 수준을 가져올지 여부를 결정합니다. 기본적으로 선택되어 있습니다. 이 옵션이 선택 취소되어 있으면 해당 필터가 포함 필터로 추가됩니다.
범주형 열에 대한 기본 필터 유형	"목록 상자"를 범주형 열의 기본 필터 유형으로 지정합니다.
범주형 열에 대한 대체 필터 유형	행 수를 확인할 수 없거나 쿼리가 취소될 때 포함 필터를 범주형 열의 기본 필터 유형으로 지정합니다.
백그라운드에서 미리보기 업데이트 실행	쿼리 실행 중에 작업을 계속할 수 있도록 상태를 표시하지 않고 데이터 보기의 "쿼리 미리보기" 탭을 업데이트합니다. 이 옵션은 기본적으로 선택되어 있습니다.
가능한 경우 백그라운드에서 쿼리 실행	쿼리 실행 중에 쿼리 빌더 작업을 계속할 수 있도록 상태를 표시하지 않고 데이터베이스 쿼리를 실행합니다. 이 옵션은 기본적으로 선택되어 있습니다.

재코딩 열의 환경 설정

"재코딩" 환경 설정에서는 열의 재코딩을 사용자 정의합니다.

그림 13.24 재코딩 환경 설정

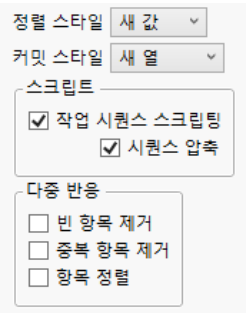


표 13.23 재코딩 페이지의 환경 설정

환경 설정	설명
정렬 스타일	기본적으로 값이 정렬되는 방식을 지정합니다. **새 값** 새 값을 기준으로 오름차순으로 정렬합니다. **이전 값** 이전 값을 기준으로 오름차순으로 정렬합니다. **개수** 가장 자주 나타나는 값이 맨 위에 오도록 내림차순으로 정렬합니다.
커밋 스타일	재코딩 열의 기본 동작을 정의합니다. **새 열** 새 열에 값을 넣는 것이 기본 설정입니다. **계산식 열** 값을 계산식 열에 넣습니다 . 값이 어떻게 바뀌는지 확인하려면 이 옵션을 선택합니다. **현재 위치** 재코딩된 열을 " 현재 위치 " 에서 바꿉니다.
작업 시퀀스 스크립팅	모든 작업을 스크립트에 저장합니다. 이 옵션은 기본적으로 선택되어 있습니다. 이 옵션이 선택 취소되어 있으면 원래 값과 대체 값만 스크립트에 포함됩니다.
시퀀스 압축	표현식을 결합하여 저장되는 스크립트를 단순화합니다. **작업 시퀀스 스크립팅**을 선택하면 기본적으로 선택됩니다.
빈 항목 제거	데이터 테이블의 다중 반응 열을 재코딩할 때 열의 빈 항목을 제외합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
중복 항목 제거	다중 반응 열에서 중복 항목을 제외합니다. 예를 들어 원래 데이터가 a, b, b, c이고 이러한 특정 항목을 변경하지 않았다면 데이터 테이블의 재코딩된 열에는 a, b, c가 포함됩니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
항목 정렬	다중 반응 열에서 항목을 사전순으로 정렬합니다. 예를 들어 원래 데이터가 b, a, c이고 이러한 특정 항목을 변경하지 않았다면 데이터 테이블의 재코딩된 열에는 a, b, c가 포함됩니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.

열 전환기의 환경 설정

" 열 전환기 " 환경 설정은 열 전환기 유틸리티의 설정을 사용자 정의합니다 .

그림 13.25 열 전환기 환경 설정

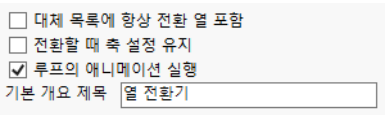


표 13.24 열 전환기 페이지의 환경 설정

환경 설정	설명
대체 목록에 항상 전환 열 포함	이 옵션을 선택하면 현재 열이 자동으로 열 전환기의 열 목록에 포함됩니다. 이 옵션을 선택 취소하면(기본 설정) 지정된 열 전환기에 현재 열이 포함되지 않은 경우 경고가 나타납니다.
전환할 때 축 설정 유지	열 전환기에서 열을 전환할 때 축 설정을 일정하게 유지하도록 지정합니다. 이 옵션은 기본적으로 선택 취소되어 있습니다.
루프의 애니메이션 실행	애니메이션 루프가 끝날 때의 기본 동작을 지정합니다. 이 옵션이 선택되어 있으면(기본 설정) 애니메이션이 루프 끝에 도달한 경우 처음부터 다시 시작됩니다. 이 옵션이 선택 취소되어 있으면 애니메이션이 루프 끝에서 중지됩니다.
기본 개요 제목	열 전환기의 제목을 지정합니다. 기본 설정은 "열 전환기" 입니다.

로그의 환경 설정

" 로그 " 환경 설정은 로그에 기록되는 대화식 JMP 작업 , 로그 모드 , 확장 로그의 모양 및 동작에 대한 사용자 정의를 지정합니다 . 로그 미리보기가 환경 설정 옵션 오른쪽에 나타납니다 .

그림 13.26 로그 환경 설정

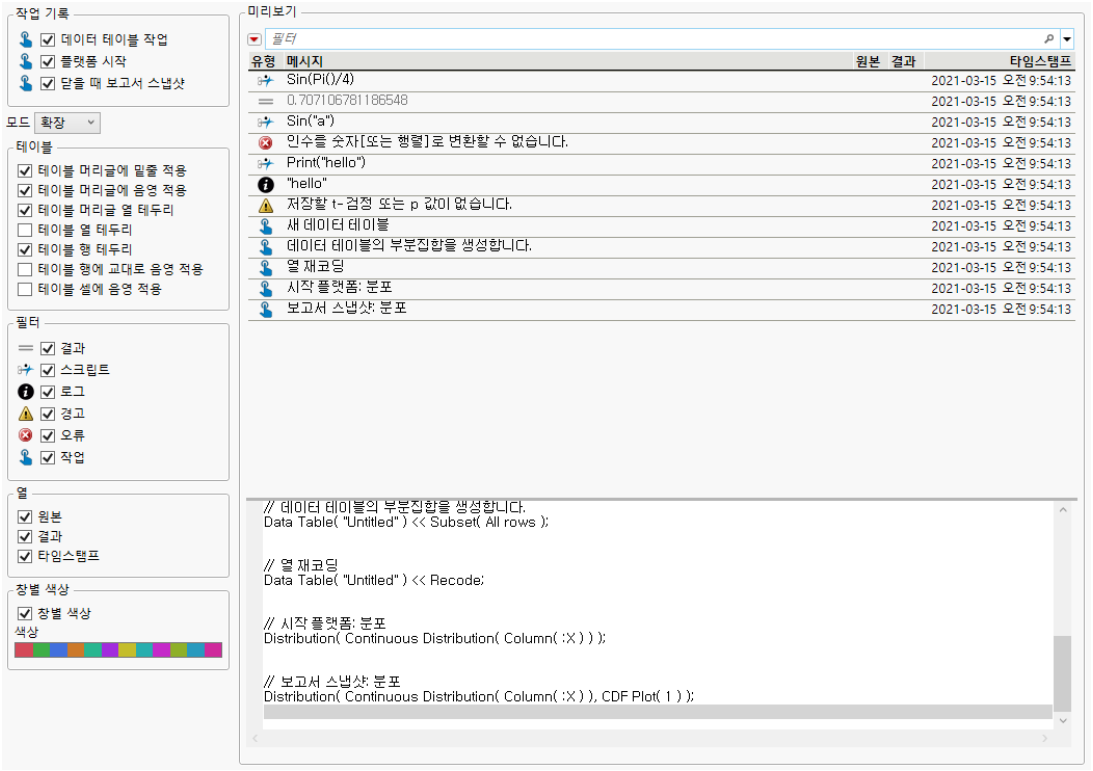


표 13.25 로그 페이지의 환경 설정

환경 설정 그룹	설명
작업 기록	로그에 기록되는 작업 유형을 제어하는 옵션이 포함되어 있습니다. 보고서를 닫을 때 데이터 테이블, 플랫폼 시작 및 각 보고서의 스크립트에 대한 작업을 기록하도록 선택할 수 있습니다. "작업 기록" 환경 설정은 두 로그 모드에 모두 적용할 수 있습니다. "모두" 옵션은 기본적으로 선택되어 있습니다.
모드	확장 로그와 텍스트 로그(기본 설정) 사이를 전환할 수 있습니다.

표 13.25 로그 페이지의 환경 설정 (계속)

환경 설정 그룹	설명
테이블	확장 로그 내용의 모양을 제어하는 옵션이 포함되어 있습니다. "테이블 머리글에 밑줄 적용", "테이블 머리글에 음영 적용", "테이블 머리글 열 테두리" 및 "테이블 행 테두리"는 기본적으로 선택되어 있습니다.
필터	확장 로그에 나타나는 출력 유형을 제어하는 옵션이 포함되어 있습니다. "모두" 옵션은 기본적으로 선택되어 있습니다.
열	확장 로그에 나타나는 정보 열을 제어하는 옵션이 포함되어 있습니다. 메시지의 원본 유형, 메시지의 결과 유형 및 메시지의 타임스탬프와 같은 정보 열을 표시하거나 숨길 수 있습니다. "모두" 옵션은 기본적으로 선택되어 있습니다.
창별 색상	확장 로그의 행에 로그 출력의 행을 생성한 JMP 창에 따라 색상을 지정할지 여부를 지정합니다. 이 옵션은 기본적으로 선택되어 있습니다. "JMP 기본 설정"이 기본 색상 테마입니다.

부록 A

계산식에서 JMP 함수에 대한 참조

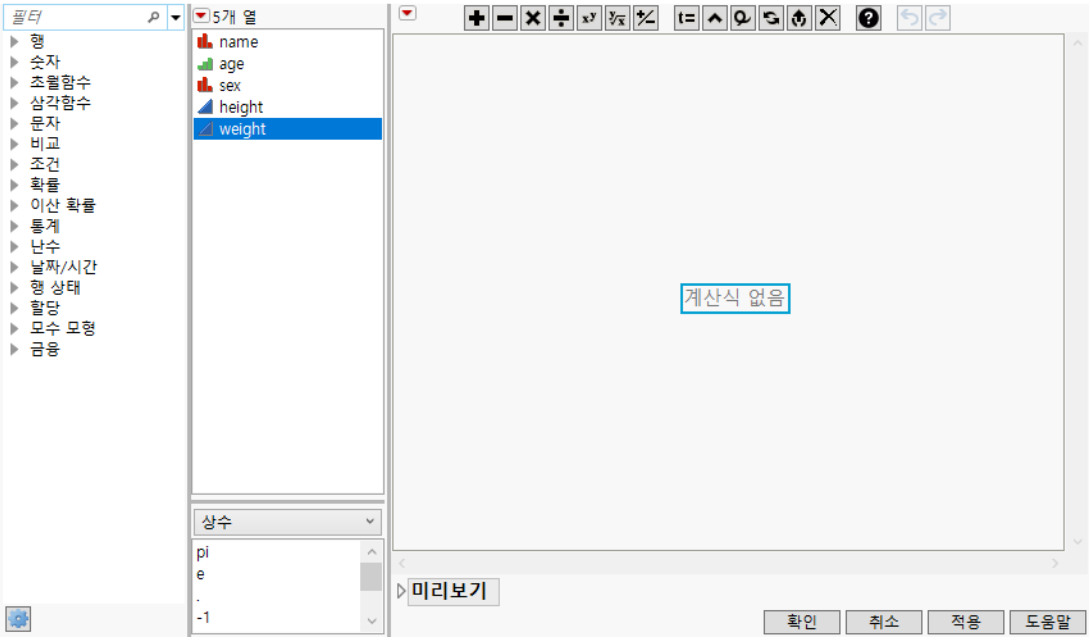
계산식 편집기의 함수에 대한 설명

계산식에 함수를 추가할 수 있습니다. 이러한 모든 함수는 계산식 편집기의 "함수" 목록에서 제공됩니다. 이 장에서는 계산식 편집기의 함수에 대해 설명합니다.

다음 리소스에서도 함수에 대한 자세한 내용을 확인할 수 있습니다.

- 스크립트 인덱스에서는 모든 함수와 인수에 대해 설명하고, 함수의 작동 방식을 보여 주고, 온라인 도움말에 연결합니다. JMP 에서 이 대화식 리소스를 보려면 **도움말 > 스크립트 인덱스**를 선택하십시오.
- JSL Syntax Reference 에서는 계산식 편집기에서 사용할 수 있는 JMP 함수뿐만 아니라 JMP 함수의 인수도 제공합니다. JMP 에서 **도움말 > JMP 도움말**을 선택하고 왼쪽 열에서 JSL Syntax Reference 를 찾으십시오.

그림 A.1 계산식 편집기의 함수



함수가 포함된 계산식을 생성하는 방법에 대한 자세한 내용은 "JMP의 계산식 생성" 장의 "계산식 작성"(361 페이지)에서 확인하십시오.

목차

행 함수 677
숫자 함수 679
초월함수 680
삼각함수 683
문자 함수 684
비교 함수 689
조건 함수 690
확률 함수 695
이산 확률 함수 704
통계 함수 706
난수 함수 710
날짜 / 시간 함수 715
행 상태 함수 718
할당 함수 723
모수 모형 함수 724
금융 함수 724

행 함수

계산식에 행 함수를 추가하여 특정 행 또는 특정 행 내의 셀을 참조할 수 있습니다. 등차수열을 기반으로 값을 삽입할 수도 있습니다. 구문에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

Col Stored Value

열 특성이 적용되지 않은 열 값을 반환합니다. 행을 지정하지 않으면 현재 행이 사용됩니다.

예를 들어 "999"를 결측값으로 처리하기 위해 x1 열에 "결측값 코드" 열 특성이 할당되었다고 가정합니다. 또 다른 열에는 평균을 계산하는 계산식이 포함되어 있습니다. 결측값 대신 값 "999"를 사용하여 평균을 계산하려면 계산식에서 다음과 같이 "Col Stored Value"를 사용하십시오.

```
Mean( Col Stored Value( :x1 ), :x2, :x3 )
```

개수

from 값으로 시작하고 **to** 값으로 끝나는 값 목록을 생성합니다. **step** 수는 목록에서 **from** 값과 **to** 값 사이에 있는 값(경계값 포함)의 개수를 지정합니다. count 함수의 처음 세 인수에 의해 결정되는 각 값은 **times** 인수를 사용하여 지정한 횟수만큼 연속적으로 발생합니다. **to** 값에 도달하면 **from** 값에서 **Count**가 다시 시작됩니다.

또한 키보드의 삽입 버튼 ^을 사용하여 **times** 인수를 추가할 수 있습니다. 이 인수는 기본적으로 1이지만 그림 A.2의 데이터 테이블에서 Count4 열에 표시된 것과 같이 지정된 횟수만큼 count 프로세스를 반복합니다. **Count** 함수에 인수를 추가하려면 인수를 입력할 위치 앞의 인수를 강조 표시하십시오. 쉼표를 입력하거나 계산식 편집기 키패드의 삽입 버튼 ^을 사용합니다.

아래 데이터 테이블의 열은 다음 계산식에서 반환된 결과입니다.

- **Count (1, 9, 2)** 에서 Count 1 이 반환됨
- **Count (1, 9, 3)** 에서 Count 2 가 반환됨
- **Count (1, 9, 9)** 에서 Count 3 이 반환됨
- **Count (1, 9, 3, 3)** 에서 Count 4 가 반환됨

그림 A.2 Count 함수의 예

	count1	count2	count3	count4
1	1	1	1	1
2	9	5	2	1
3	1	9	3	1
4	9	1	4	5
5	1	5	5	5

Count 함수는 격자 값의 열을 생성하는 데 유용합니다. 예를 들어 다음 계산식은 **NRow()** 만큼씩 증분되는 정사각형 격자를 생성합니다. **NRow()**는 데이터 테이블의 총 행 수를 반환하는 **Row** 함수입니다. 또한 **NRow()**는 범위가 –5부터 5까지인 축을 생성합니다.

```
Count(-5, 5, Root( NRow() ) );
Count(-5, 5, Root( NRow() ), Root( NRow() ) );
```

Dif

현재 행에 있는 첫 번째 인수의 값에서, 현재 행 번호에서 두 번째 인수를 뺀 수로 정의되는 행의 해당 인수 값을 뺀 차이를 반환합니다. 기본 **Dif**는 1이며, 이를 임의의 숫자로 변경할 수 있습니다. Dif(X, n)는 $X_{Row()}-X_{Row()-n}$ 또는 $X_{Row()}$-Lag(X, n)와 동일한 결과를 반환합니다.

Dim

현재 데이터 테이블, 지정된 데이터 테이블 또는 행렬의 차원이 포함된 행 벡터를 반환합니다. 차원은 행과 열의 개수이며 해당 순서로 나열됩니다.

Lag

현재 행 번호에서 두 번째 인수를 뺀 수로 정의되는 행에서 첫 번째 인수의 값을 반환합니다. 기본 **Lag**는 1이며, 이를 임의의 숫자로 변경할 수 있습니다. 1보다 작은 행 번호를 나타내는 Lag에 대해 반환되는 값은 결측값입니다. Lag(X, n)는 서브스크립트 표기인 $X_{Row( )-n}$과 동일한 결과를 반환합니다.

NRow

활성 데이터 테이블의 총 행 수를 반환합니다.

Row

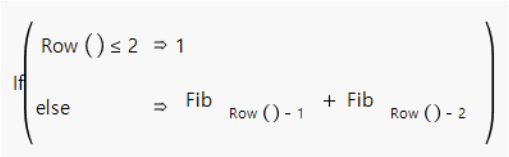
표현식이 실행되는 현재 행 번호를 반환합니다. 열 이름 서브스크립트를 포함한 모든 표현식에서 **Row()**를 사용할 수 있습니다. 열 이름의 기본 서브스크립트는 달리 지정되지 않은 한 **Row()**입니다.

Subscript

현재 행이 아닌 다른 행의 열 값을 사용할 수 있도록 합니다. 목록에서 **Subscript**를 선택한 후 subscript 인수에 숫자 표현식을 입력합니다. 존재하지 않는 행 번호로 평가되는 Subscript는 결측값을 생성합니다. 서브스크립트가 없는 열 이름은 현재 행을 참조합니다. 서브스크립트를 제거하려면 서브스크립트를 선택한 후 삭제합니다. 그런 다음 빈 상자를 삭제합니다.

계산식 $Count_{Row()} - Count_{Row()-1}$(여기서 **Row()**는 아래에 설명된 행 번호)은 서브스크립트를 사용하여 Count라는 열에서 각 값 쌍 사이의 차이를 계산합니다. 이 결과는 **Dif()** 함수에서 반환되는 결과와 동일합니다. **Row()**가 1인 경우 계산 결과는 결측값이 됩니다.

아래의 계산식은 **피보나치** 급수(각 값이 계산 대상 열에 있는 이전 2개 값의 합인 급수)의 항이 포함된 Fib라는 열을 산출합니다.



이 계산식은 서브스크립트를 사용하여 재귀적 계산을 수행하는 방법을 보여 줍니다. 재귀적 계산식에는 이전에 계산된 행(행 1 ~ (*i*–1))만 참조하도록 계산 대상 열의 이름이 서브스크립트되어 포함됩니다. 표시된 피보나치 급수의 계산에는 조건부 표현식과 비교가 포함되어 있습니다. 자세한 내용은 "조건 함수"(690 페이지) 및 "비교 함수"(689 페이지)에서 확인하십시오.

Sequence

데이터 테이블의 행에 대해 등차수열을 생성합니다. 시작 값, 종료 한계 및 증분이 인수로 지정됩니다.

숫자 함수

계산식에 일반적으로 사용되는 산술 연산자가 포함된 계산식을 생성할 수 있습니다. 구문에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

Abs

인수 값과 동일한 크기의 양수를 반환합니다. 예를 들어 |5| 와 |–5| 는 모두 5가 됩니다.

Ceiling

인수보다 크거나 같은 정수 중 가장 작은 정수를 반환합니다. 예를 들어 **Ceiling(2.3)**은 3을 반환하는 반면, **Ceiling(–2.3)**은 –2를 반환합니다.

Floor

인수보다 작거나 같은 정수 중 가장 큰 정수를 반환합니다. 예를 들어 **Floor(2.7)**는 2를 반환하지만, **Floor(–0.5)**는 –1을 반환합니다.

Integrate

Gander와 Gautschi(2000)의 적응형 구적법을 사용하여 관련된 스칼라 값에 대해 표현식을 적분합니다. 예를 들어 **Integrate(Exp(-x), x, 0, .)**는 0.99999999986846을 반환합니다.

Modulo

두 번째 인수를 첫 번째 인수로 나눈 나머지를 반환합니다. 예를 들어 **Modulo(6, 5)**는 1을 반환합니다.

Round

첫 번째 인수를 두 번째 인수로 지정된 소수 자릿수로 반올림합니다. 예를 들어 **Round(3.554, 2)**는 3.55로 반올림되고 **Round(3.555, 2)**는 3.56으로 반올림됩니다.

초월함수

밑에 대한 로그 함수, 조합 계산을 위한 함수, **베타** 함수 및 일부 감마 함수와 같은 초월함수를 지원하는 계산식을 생성할 수 있습니다. 구문에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

Exp

*e*를 지정한 횟수만큼 거듭제곱합니다. 따라서 **Exp(1) = e**입니다.

ExpM1

x가 매우 작을 경우 **Exp(x) - 1**에 대한 더 정확한 계산을 반환합니다.

Ln

*x*의 자연 로그를 계산합니다.

Log 및 Log10

*e*를 밑으로 하는 자연 로그를 계산합니다. 기본 밑을 변경하려면 인수를 강조 표시하고 쉼표를 입력하거나 계산식 편집기 키패드의 삽입 키를 클릭하십시오. 그러면 밑이 표시되며 이를 편집할 수 있습니다. **Log** 인수는 임의의 숫자 표현식일 수 있습니다. 표현식 **Log(e)**의 결과는 1이 되고, **Log(32,2)**의 결과는 5가 됩니다. **Log10** 함수는 상용 로그만 계산합니다.

Log1P

*x*가 매우 작을 경우 **Log(1+x)**에 대한 더 정확한 계산을 반환합니다.

Squash

함수 **1 / (1 + ex)**를 계산합니다. 여기서 *x*는 숫자 열, 변수 또는 표현식입니다.

Logist

Squish 또는 Logistic이라고도 하며, 함수 **1 / (1+e-x)**를 효율적으로 계산하는 방법입니다. 여기서 *x*는 숫자 열, 변수 또는 표현식입니다.

Root(Square Root)

인덱스로 지정된 인수의 제곱근을 계산합니다. **Root**는 기본적으로 인덱스 2와 함께 표시됩니다. 인덱스를 변경하려면 인덱스 인수를 강조 표시하고 원하는 값을 입력하십시오.

Factorial

1부터 지정한 인수까지 모든 수를 곱한 결과를 반환합니다. 예를 들어 **Factorial(5)**의 결과는 120이 됩니다.

NChooseK

*n*개의 항목 중 한 번에 *k*개를 선택하는 경우의 수(*n* select *k*)를 반환하며, 계승을 사용하는 표준 방법, 즉 $n! / (k!(n - k)!)$로 계산됩니다. 예를 들어 **NChooseK(5,2)**의 결과는 10이 됩니다.

베타

2모수 **Beta** 함수를 추가하며, Gamma 함수로는 다음과 같이 작성됩니다.

$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$$

감마

$\Gamma(i)$로 나타내는 **Gamma** 함수를 추가하며, 다음과 같이 정의됩니다.

$$\Gamma(i) = \int_0^\infty (x^{i-1})(e^{-x})dx$$

단일 인수를 사용하는 Gamma는 **Gamma(x, infinity)**와 동일합니다. 선택적인 두 번째 인수는 상한 정수를 무한대에서 사용자가 입력한 값으로 변경합니다. 다른 흥미로운 감마 함수 관계식은 다음과 같습니다.

- 1 보다 큰 임의의 α 에 대해 , $\Gamma(\alpha) = (\alpha-1) \bullet \Gamma(\alpha-1)$
- 임의의 양의 정수에 대해 , *n*, $\Gamma(n) = (n-1)!$
- $\Gamma(0.5) = \pi$ 의 제곱근

LGamma

감마 함수 계산 결과의 자연 로그입니다. **Log**(자연 로그) 함수와 **Gamma** 함수를 함께 사용할 때와 동일한 결과가 반환됩니다. 하지만 **LGamma** 함수를 사용한 계산이 **Log**(자연 로그) 및 **Gamma** 함수를 함께 사용한 계산보다 더 효율적입니다. **NChooseK**는 **LGamma** 함수를 사용하여 구현됩니다. 결과가 항상 정확한 정수가 되지는 않습니다. 결과가 정수에 가까우면 **Floor** 함수를 사용하여 올림됩니다.

Digamma

Gamma 함수의 로그 도함수입니다.

Trigamma

Digamma 함수의 도함수, 또는 **Gamma** 함수의 로그 2차 도함수입니다.

Arrhenius

아레니우스 관계식의 비특정 성분을 계산한 다음 아레니우스 방정식의 활성화 에너지로 곱합니다.

$$\frac{11604.5181215503}{T + 273.15}$$

Arrhenius Inv

Arrhenius 함수의 역함수를 반환합니다.

$$\left(\frac{11604.5181215503}{y}\right) - 273.15$$

Logit

다음을 사용하여 인수에 로짓 변환을 적용합니다.

$$\text{logit}(p) = \log\left(\frac{p}{(1-p)}\right)$$

Logit Percent

인수에 대한 로짓 백분율을 계산합니다.

$$\text{LogitPct}(pct) = \log\left(\frac{\left(\frac{pct}{100}\right)}{1-\left(\frac{pct}{100}\right)}\right)$$

Logist Percent

인수에 대한 로지스틱 백분율을 계산합니다.

$$\text{LogisticPct}(x) = \frac{100}{(1 + e^{-x})}$$

Scheffe Cubic

특정 모형의 적합에 사용됩니다. **Scheffe Cubic (X1, X2)**는 **X1*X2*(X1-X2)**와 동등합니다.

SbInv

Johnson Sb 역 변환입니다. 인수가 정규 분포를 따를 경우 결과는 Johnson Sb입니다.

SbTrans

이중 경계 변수에서 표준 정규(0, 1) 분포로의 Johnson Sb 변환입니다.

SHASHInv

표준 정규 분포를 따르는 변수를 SHASH(SinH-ArcSinH) 분포를 따르는 변수로 변환한 결과를 반환합니다.

SHASHTrans

SHASH(SinH-ArcSinH) 분포를 따르는 변수를 표준 정규 분포를 따르는 변수로 변환한 결과를 반환합니다.

SlInv

Johnson Sl 역 변환입니다. 인수가 정규 분포를 따를 경우 결과는 Johnson Sl입니다.

SlTrans

이중 경계 변수에서 표준 정규(0, 1) 분포로의 Johnson Sl 변환입니다.

SuInv

Johnson Su 역 변환입니다. 인수가 정규 분포를 따를 경우 결과는 Johnson Su입니다.

SuTrans

이중 경계 변수에서 표준 정규(0, 1) 분포로의 Johnson Su 변환입니다.

삼각함수

Sine 및 Cosine과 같은 삼각함수를 지원하는 계산식을 생성할 수 있습니다. JMP의 삼각함수에서는 모든 각도 인수가 라디안 단위여야 합니다. 구문에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

Sine, Cosine, Tangent

Sine 및 **Cosine** 함수는 라디안으로 지정된 해당 인수의 사인 및 코사인을 계산합니다. 예를 들어 표현식 **Sine(0)**의 결과는 0이 되고, **Cosine(0)**의 결과는 1이 됩니다. Tangent 함수는 라디안으로 지정된 인수의 탄젠트를 계산합니다. 표현식 **Tan(Pi()/4)**의 결과는 1이 됩니다.

ArcSine, ArcCosine, ArcTangent

ArcSine 및 **ArcCosine** 함수는 해당 인수의 역사인 및 역코사인을 반환합니다. 반환되는 값은 라디안 단위입니다. 예를 들어 표현식 **ArcSine(1)**과 **ArcCosine(0)**의 결과는 모두 1.57080이 됩니다. **ArcTangent** 함수는 인수의 역탄젠트를 반환합니다. 반환되는 값은 라디안 단위입니다. 표현식 **ArcTangent(1)**의 결과는 0.78540(=3.14159/4)이 됩니다.

SinH, CosH, TanH

SinH 및 **CosH** 함수는 해당 인수의 쌍곡선 사인 및 쌍곡선 코사인을 반환합니다. 표현식 **SinH(1)**의 결과는 1.175201이 되고, **CosH(0)**의 결과는 1.0이 됩니다. **TanH** 함수는 인수의 쌍곡선 탄젠트를 반환합니다. 표현식 **TanH(1)**의 결과는 0.761594가 됩니다.

ArcSinH, ArcCosH, ArcTanH

ArcSinH 및 **ArcCosH** 함수는 해당 인수의 역쌍곡선 사인 및 역쌍곡선 코사인을 반환합니다. 표현식 **ArcSinH(1)**의 결과는 0.881374가 되고, **ArcCosH(1)**의 결과는 0이 됩니다. **ArcTanH** 함수는 인수의 역쌍곡선 탄젠트를 반환합니다. 표현식 **ArcTanH(0.5)**는 0.549306이 됩니다.

문자 함수

문자 인수를 받거나 문자열을 반환하고, 값의 데이터 유형을 숫자에서 문자로 변환하거나 문자에서 숫자로 변환하는 계산식을 생성할 수 있습니다. 이러한 계산식을 생성할 경우 다음 사항에 유의하십시오.

- 문자 함수의 결과는 문자 또는 숫자 데이터일 수 있습니다. 지정된 것과 다른 데이터 유형을 계산할 경우 계산된 열의 데이터 유형이 결과와 매칭되도록 자동으로 변경됩니다.
- 리터럴 문자열 인수는 인용 부호로 묶어야 합니다.

구문에 대한 자세한 내용은 Scripting Guide의 에서 확인하십시오.

Char

숫자 인수의 소수 자릿수와 일치하는 문자열을 생성합니다. 예를 들어 **Char(1.123)**의 결과는 1.123이 됩니다. 자세한 내용은 Scripting Guide의 에서 확인하십시오.

Collapse Whitespace

선행 및 후행 공백을 제거하고 중간에 있는 공백을 단일 공백으로 바꿉니다. 즉, 공백 문자가 둘 이상 있는 경우 **Collapse Whitespace** 명령을 사용하면 두 개의 공백이 한 개의 공백으로 바뀝니다.

Concat ||

문자열을 연결하여 함수의 첫 번째 문자 인수에 두 번째 문자 인수가 추가된 새 문자열을 생성합니다. 예를 들어 **"Dr." || " " || name**은 직함 **Dr.** 다음에 공백과 name 문자열의 내용을 포함하는 새 문자열을 생성합니다.

Contains

첫 번째 인수 내에 두 번째 인수가 있는 경우 이 인수의 첫 번째 인스턴스가 나타나는 숫자 위치를 반환합니다. 두 번째 인수에는 하나 이상의 문자가 포함될 수 있습니다. 두 번째 인수가 나타나지 않는 경우 **Contains**는 0을 반환합니다. 예를 들어 **Contains("Veronica Layman", "ron")**의 결과는 3이 됩니다. **Contains("Lillie Layman", "L")**의 결과는 1이 됩니다. **Contains("Lillie**

Layman", "Veronica")의 결과는 **0이 됩니다.** 세 번째 인수인 offset은 선택 사항입니다. 오프셋은 검색 시작 위치를 지정하는 숫자 값입니다. **Contains("Lillie Layman", "L", 5)**의 결과는 검색이 5번째 위치에서 시작하기 때문에 8이 되고, 첫 번째 "L"은 건너뜁니다. offset이 음수이면 **Contains**는 문자열 끝에서 offset만큼 이동한 위치부터 역방향으로 검색합니다.

Munger

문자를 삽입하거나 삭제하는 방법으로 기존 문자열에서 새 문자열을 생성합니다. 인수를 지정하는 방식에 따라 하위 문자열을 생성하거나, 인덱스를 계산하거나, 그 밖의 작업을 수행할 수도 있습니다. **Munger** 함수는 대문자와 소문자를 서로 다른 문자로 간주합니다.

Text는 문자 표현식입니다. **Munger**는 다른 세 개의 인수를 이 문자열에 적용하여 결과를 계산합니다.

Offset은 문자열 내의 검색 시작 위치를 나타내는 숫자 표현식입니다. **Offset**이 Find 인수의 첫 번째 인스턴스가 나타나는 위치보다 크면 첫 번째 인스턴스는 무시됩니다.

Find/Length는 문자 또는 숫자 표현식입니다. 문자열을 검색 기준으로 사용할 수도 있고, 양의 정수를 사용하여 **Offset** 위치에서부터 해당 수만큼의 연속된 문자를 반환할 수도 있습니다. **Length** 값으로 음의 정수를 지정하면 **Munger**는 **Offset**부터 문자열 끝까지의 모든 문자를 반환합니다.

Replace(선택적 인수)는 문자열로 지정할 수도 있고 지정하지 않을 수도 있습니다. 이 인수가 문자열이고 **Find/Offset** 값이 숫자이면 **Munger**는 검색 기준을 **Replace** 문자열로 바꿔 결과를 생성합니다. Find/Offset 값이 숫자이고 문자열이 지정되지 않은 경우 **Munger**는 하위 문자열을 산출합니다. **Find/Length** 값이 문자열이면 **Munger**는 **Replace** 값이 있더라도 이를 무시하고 항상 숫자 오프셋을 반환합니다. **Replace** 인수를 삽입하려면 **Munger** 함수에서 아무 인수나 클릭한 후 삽입 버튼을 클릭하십시오. **Replace** 인수를 제거하려면 Delete 키를 누르거나 계산식 편집기 키패드의 삭제 버튼( )을 클릭합니다.

Lowercase, Uppercase

Lowercase 함수는 인수에 있는 대문자를 동등한 소문자로 변환합니다. 예를 들어 **Lowercase ("VERONICA LAYMAN")**의 결과는 "veronica layman"이 됩니다. **Uppercase** 함수는 인수에 있는 소문자를 동등한 대문자로 변환합니다. 예를 들어 **Uppercase("Veronica Layman")**의 결과는 "VERONICA LAYMAN"이 됩니다.

Length

인수의 길이를 계산합니다. 예를 들어 **Length("Veronica")**의 결과는 8이 됩니다. 인수에 따라 결과는 다음과 같이 달라집니다.

- 인수가 문자열이면 문자의 개수가 반환됩니다 .
- 인수가 목록이면 목록에 있는 항목의 개수가 반환됩니다 .
- 인수가 BLOB(이진 개체) 이면 바이트 수가 반환됩니다 .

Num

문자열 인수가 숫자로만 구성된 경우 해당 문자열에 대응하는 숫자 값을 생성합니다. 문자열에 숫자가 아닌 값이 포함되어 있는 경우 결과는 결측값입니다. 예를 들어 **Num("1.123")**의 결과는 1.123이 됩니다.

Substr

첫 번째 인수의 일부 문자를 추출합니다. 두 번째 인수로 지정된 위치에서 시작되며, 끝 위치는 세 번째 인수로 지정된 문자 수에 따라 결정됩니다. 첫 번째 인수는 문자 열이거나 리터럴 값일 수 있습니다. 시작 인수와 길이 인수는 숫자이거나 결과가 숫자가 되는 표현식일 수 있습니다. 예를 들어 **Substr("Veronica Layman", 10, 6)**은 전체 이름 중 이름만 표시하기 위해 위치 10에서부터 위치 15까지 읽어 **Layman**을 산출합니다.

start가 음수인 경우 **Substr**은 문자열의 끝에서 start만큼 이동한 위치부터 역방향으로 검색합니다. length가 음수이거나 지정되지 않은 경우 **Substr**은 start 위치에서 시작하여 텍스트 문자열의 끝까지 포함하는 문자열을 반환합니다.

Substr은 목록과 함께 사용할 수도 있습니다.

Titlecase

문자열을 첫 글자가 대문자인 형식, 즉 첫 번째 문자는 대문자이고 이후 문자는 소문자인 형식으로 변환합니다. 예를 들어 **Titlecase("Veronica layman")**은 **Veronica Layman**을 반환합니다.

Trim

인수에서 선택 및 후행 공백을 제거하여 새 문자열을 생성합니다. 두 번째 인수는 문자열의 **왼쪽**, **오른쪽** 또는 **양쪽** 중 어디에서 공백을 제거할지 결정합니다. 두 번째 인수가 사용되지 않은 경우에는 양쪽에서 공백이 제거됩니다. 예를 들어 **Trim("john ")**의 결과는 **john**이 됩니다. 또한 **Trim(" john ", both)**의 결과도 **john**이 됩니다.

Word

문자열에서 *n*번째 단어를 추출합니다. 선택적인 **delimiters** 인수가 지정되지 않은 경우 하나 이상의 공백으로 각 단어의 시작 및 끝 위치가 정의됩니다. 예를 들어 **Word(2, "Veronica Layman")**은 단어 **Layman**을 반환합니다.

delimiters 인수를 삽입하려면 **Word** 함수의 아무 인수나 클릭한 후 계산식 편집기 키패드의 삽입 버튼 을 클릭하십시오. **delimiters 인수를** 제거하려면 Delete 키를 누르거나 계산식 편집기 키패드의 삭제 버튼( )을 클릭합니다. 구분자를 지정하지 않으면 공백이 구분자로 사용됩니다. 구분자를 빈 문자열로 정의하면 각 문자가 개별 단어로 간주됩니다.

대부분의 특수 문자는 단일 구분자 역할을 합니다. 임의의 문자 또는 문자 집합을 입력하여 단어 구분자 역할을 하도록 할 수 있습니다. 예를 들어 다음 예에서는 전체 이름 중 성을 추출하기 위해 쉼표와 공백을 구분 문자로 사용하고 첫 번째 단어를 요청합니다. **Word(1, "Layman, Veronica", ", ")**는 단어 **Layman**을 반환합니다.

Words

선택적인 두 번째 인수에 나열된 구분자에 따라 text의 단어들을 추출합니다. 기본 구분자는 공백입니다. 예를 들어 **Words("the quick brown fox")**는 **{"the","quick","brown","fox"}**를 반환합니다.

두 번째 인수를 포함하면 해당 인수의 모든 문자가 구분자로 간주됩니다. 예를 들어 **Words ("Doe, Jane P.",", .")**는 **{"Doe","Jane","P"}**를 반환합니다.

delimiters 인수를 삽입하려면 **Words** 함수의 아무 인수나 클릭한 후 계산식 편집기 키패드의 삽입 버튼 을 클릭하십시오. **delimiters 인수를** 제거하려면 Delete 키를 누르거나 계산식 편집기 키패드의 삭제 버튼( )을 클릭합니다. 구분자를 지정하지 않으면 공백이 구분자로 사용됩니다. 구분자를 빈 문자열로 정의하면 각 문자가 개별 단어로 간주됩니다.

Left, Right

text로 지정된 문자열 중 맨 왼쪽 또는 맨 오른쪽 *n*개의 문자로 구성된 하위 문자열을 반환합니다. 두 함수 모두 목록과 함께 사용할 수 있습니다.

Starts With, Ends With

각각 whole이 part로 시작하거나 끝나면 1을 반환합니다. 그렇지 않으면 0을 반환합니다. 두 함수 모두 목록과 함께 사용할 수 있습니다.

Item

Word 함수와 유사하지만 단어 구분자의 처리 방식이 다릅니다. 한 구분자가 여러 번 발견되거나 여러 문자로 된 구분자가 입력된 경우 **Word** 함수는 이를 단일 문자로 처리합니다. **Item** 함수는 각 구분자를 사용하여 새 단어 위치를 정의합니다. 비교를 위해 "성, 이름" 형식으로 된 이름을 예로 들자면, 구분자는 쉼표와 그 다음의 공백이므로 함수를 다음과 같이 지정할 수 있습니다.

Item(2, "Layman, Veronica", ", ")

Word(2, "Layman, Veronica", ", ")

Item 함수의 경우 쉼표와 공백을 별개로 처리하는데 그 사이에 아무것도 없으므로 결측값이 반환됩니다. **Word** 함수는 쉼표와 공백을 단일 구분자로 처리하므로 두 번째 단어로 **Veronica**를 찾습니다.

구분자를 지정하지 않으면 공백이 구분자로 사용됩니다. 구분자를 빈 문자열로 정의하면 각 문자가 개별 항목으로 간주됩니다.

Hex 및 Hex to Number

16진수와 숫자 사이에서 변환합니다.

Hex는 인수의 16진수 표현을 반환합니다. 인수가 문자(따옴표로 묶임)이면 결과는 해당 문자값의 16진수 코드를 포함하는 두 배 길이의 문자열입니다. 예를 들어 **Hex("A")**는 문자열 "41"을 반환합니다.

인수가 숫자이고 "integer"가 지정된 경우 Hex 함수는 반환된 정수의 8자리 16진수 문자 표현을 반환합니다. 예를 들어 **Hex(12, "integer")**는 문자열 **0000000C**를 반환합니다.

Hex to Number는 16진수를 숫자로 변환합니다.

자세한 내용은 Scripting Guide의 에서 확인하십시오.

Repeat

첫 번째 인수가 두 번째 인수로 지정된 횟수만큼 반복된 문자열을 생성합니다. 첫 번째 인수는 문자 리터럴, 문자 변수 또는 문자 표현식일 수 있습니다. 예를 들어 **Repeat("Katie", 3)**는 **KatieKatieKatie**를 생성합니다.

세 번째 인수는 **Repeat**가 JSL 스크립트에서 행렬을 반복하는 데 사용될 때 적용됩니다. 첫 번째 인수가 행렬인 경우 두 번째 인수는 행별 반복이고 세 번째 인수는 열별 반복입니다.

Insert

Insert는 list 또는 expression의 지정된 위치에 새 항목을 삽입합니다. 위치가 지정되지 않은 경우에는 맨 끝에 새 항목이 삽입됩니다.

Reverse

Reverse는 문자열의 문자를 역순으로 정렬합니다.

Substitute

첫 번째 인수는 문자열이고, 두 번째 인수는 패턴이며, 세 번째 인수는 대체 문자열입니다. **Substitute**는 지정된 문자열의 패턴과 매칭되는 모든 부분을 찾아 대체 문자열로 바꿉니다.

Regex

첫 번째 인수는 Regex가 패턴과 매칭되는 부분을 검색할 소스 문자열입니다. 두 번째 인수는 정규 표현식 형식의 패턴입니다. 계산식 편집기에서는 이 두 개의 필수 인수를 입력하라는 메시지가 표시됩니다.

팁: 정규 표현식 사용 방법에 대한 자세한 내용은 인터넷에서 **정규 표현식 자습서**를 검색하여 확인하십시오 .

기본적으로 Regex는 대/소문자가 구분되는 검색을 수행하여 소스 문자열에서 지정된 패턴과 매칭되는 부분을 반환합니다. 매칭되는 부분이 없으면 결측값을 반환합니다. 두 개의 선택적 인수를 추가할 수 있습니다. 세 번째 인수로는 반환할 문자열을 지정하는 format을 입력할 수 있습니다. 선택한 경우 정규 표현식을 사용하여 반환되는 문자열 내의 대체 텍스트를 지정할 수 있습니다. 세 번째 인수를 지정할 경우, Regex가 소스 문자열에서 매칭되는 부분을 검색할 때 대/소문자를 무시하도록 **IGNORECASE**도 지정할 수 있습니다.

표 A.1 Regex 예제

Regex 함수 샘플	반환되는 문자열
`Regex( "@ q3 #", "([a-z])([0-9])" )`	q3 이 함수는 대/소문자를 구분하므로 q3는 매칭되지만 Q3는 매칭되지 않습니다.
`Regex( "@ Q3 #", "([a-z])([0-9])", "\0",IGNORECASE)`	Q3 \0은 기본 인수이지만 이 예에서는 IGNORECASE를 지정하기 위해 필요합니다.
`Regex( "@ Q3 #", "([a-z])([0-9])", "\2\1",IGNORECASE)`	3Q

자세한 내용과 실행 가능한 예제를 보려면 **도움말 > 스크립트 인덱스**를 선택하고 **Regex**를 검색하십시오.

비교 함수

비교 함수를 사용하여 두 인수의 값을 비교하는 계산식을 생성할 수 있습니다. 각 비교 관계식의 결과는 숫자 크기 또는 문자 순위에 따라 true 또는 false가 됩니다. true 관계식의 결과는 1이 되고, false 관계식의 결과는 0이 됩니다.

비교 함수는 조건부 표현식에 포함할 경우에 유용하지만, 비교의 어느 항도 누락되지 않았다면 독립적인 숫자 표현식으로 사용할 수도 있습니다. 관계 기호의 인수는 두 개의 표현식일 수 있습니다. 하지만 비교 함수의 두 인수가 모두 동일한 데이터 유형이어야 합니다. 또한 다음 사항에 유의하십시오.

- "==" 가 필요한 조건에 단일 "=" 를 사용하면 오류가 표시됩니다.
- 계산식 편집기는 문자열 비교 시 International Utilities 패키지를 사용합니다 . 이 패키지는 각 국제 문자 집합에 대한 서로 다른 순위를 포함하고 있으며 분음 부호도 고려합니다.
- 결측값과 비교하려는 경우에는 비교 연산자를 사용하면 안 됩니다. 대신 **Is Missing** 함수를 사용하여 결측값을 확인해야 합니다.

구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

< 보다 작음

> 보다 큼

<= 보다 작거나 같음

>= 보다 크거나 같음

== 같음

!= 같지 않음

a<b<=c *b* 가 *a* 보다 크고 *c* 보다 작거나 같음

a<=b<c *b* 가 *a* 보다 크거나 같고 *c* 보다 작음

Is Missing 현재 행의 인수 값이 결측값이면 1 을 반환하고 , 결측값이 아니면 0 을 반환합니다 .

조건 함수

계산식에 조건부 표현식 (줄여서 **조건**이라고 함) 을 포함할 수 있습니다 . 이러한 표현식을 사용하여 절과 결과 표현식으로 구성된 일련의 쌍을 생성할 수 있습니다 . 일련의 절을 생성함으로써 계산된 열의 셀에 조건부로 값을 할당할 수 있습니다 . 조건은 다음과 같은 규칙을 따릅니다 .

- true 가 되는 절이 없으면 계산식 편집기는 else 절에서 제공되는 결과 표현식을 실행합니다 .
- 조건부 표현식의 모든 결과 표현식은 동일한 데이터 유형이 되어야 합니다 .
- 결측 항은 모든 데이터 유형과 매칭됩니다 .
- 정의에 따라 , 결과가 0 이 되는 표현식은 false 입니다 .
- 표현식 결과가 결측값이 되면 어떤 절도 실행되지 않고 결측값이 반환됩니다 . 다른 모든 숫자 표현식은 true 입니다 .

에서 확인하십시오 .

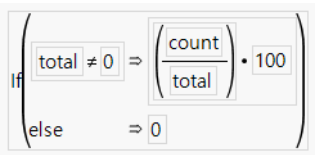
표현식을 확장하려면 계산식 편집기 패널의 삽입 및 삭제 버튼을 사용하십시오 . 효율성을 최대화하려면 가장 자주 실행되는 절 및 결과 쌍을 조건 시퀀스의 맨 앞에 나열하십시오 .

참고 : Interpolate, **Step**, **For** 및 **While** 는 주로 다른 명령과 함께 JSL 스크립트를 생성하는 데 사용됩니다 . 계산식 편집기를 사용하여 열 내에서 스크립트를 생성하고 실행할 수 있지만 종속성 및 모호성 문제가 발생할 수 있으므로 이 방법은 권장되지 않습니다 . 대개 스크립트는 .jsl 파일로 저장되며 , 데이터 테이블과 함께 테이블 특성으로 저장될 수 있습니다 . 테이블 특성에 대한 자세한 내용은 "JMP 시작하기 " 장의 " 데이터 테이블의 테이블 패널 "(38 페이지) 에서 확인하십시오 . 스크립트 명령에 대한 설명은 JSL Syntax Reference 에서 확인하십시오 .

If

빈 표현식 하나와 빈 **then 절** 하나가 포함된 단일 **If** 조건을 표시합니다 . **표현식** 또는 **then 절**을 강조 표시하고 값을 입력하십시오 . 예를 들어 그림 A.3 에서는 **count** 및 **total** 열을 사용하여 총 수가 0 이 아닐 때 개수를 총 수 대비 백분율로 계산합니다 .

그림 A.3 조건부 표현식



If 조건에 **새 조건을 추가하려면 then 절**을 강조 표시하고 계산식 편집기 키패드의 삽입 버튼(⌃)을 클릭하십시오. 이렇게 하면 처음에는 기존 **else** 조건이 **표현식** 절로 변경됩니다. 삽입 버튼을 다시 클릭하면 **else 절**이 추가됩니다. **then** 또는 **else**를 강조 표시하고 삽입 버튼을 반복적으로 클릭하면 **else**가 **표현식**으로 변경되거나 새 **표현식** 절이 추가됩니다.

절을 삭제하려면 해당 절 위의 **then 절**을 선택하고 Delete 키를 누르거나 계산식 편집기 키패드의 삭제 버튼(⌫)을 클릭합니다.

정의에 따라, 결과가 0이 되는 표현식은 false입니다. 표현식 결과가 결측값이 되면 어떤 절도 실행되지 않고 결측값이 반환됩니다. 다른 모든 숫자 표현식은 true입니다.

Match

표현식을 절 목록과 비교한 후 첫 번째로 매칭되는 절의 결과 표현식 값을 반환합니다. 매칭 표현식을 한 번만 지정한 후 각 절의 매칭 값을 지정해야 합니다.

계산식 편집기에서 **Match**를 선택하면 다음과 같은 두 개의 옵션이 있는 목록이 나타납니다.

- **데이터에서 Match 인수 추가**를 선택하면 데이터의 모든 수준에 해당하는 절이 자동으로 추가됩니다. 또는 Shift 키를 누른 채 **조건부**를 선택하고 **매칭**를 선택합니다. 그림 A.4 의 왼쪽 예에서는 자동으로 추가된 절을 보여 줍니다.
- 각 절을 개별적으로 추가할 수 있도록 **추가 안 함**을 선택합니다. 그림 A.4 의 오른쪽 예에서는 사용자가 직접 빈 표현식을 채울 수 있는 빈 절을 보여 줍니다.

그림 A.4 Match 함수의 사용 예

자동으로 추가된 Match 인수

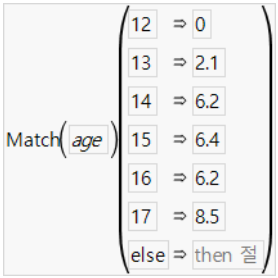
빈 Match 인수

자동으로 채워진 인수에서는 **then 절**을 강조 표시한 후 표현식을 입력해야 합니다. 빈 인수에서는 **표현식**, **값** 또는 **then 절**을 강조 표시한 후 표현식을 입력합니다. 또는 표현식을 강조 표시하고 **Match**를 클릭해도 됩니다. 그러면 계산식 편집기가 원래의 강조 표시된 표현식이 **표현식**으로 포함된 새 **Match** 조건을 추가하고 **값** 및 **else** 절에는 아무 작업도 수행하지 않습니다. 또한 다음 사항에 유의하십시오.

- **Match** 는 변수가 데이터베이스의 각 행에 대해 한 번씩만 계산되므로 동등한 기능의 **If** 에 비해 실행 속도가 빠르고 메모리 사용량은 적습니다. **If** 조건은 If 절의 결과가 true 가 될 때까지 각 행에 대한 각각의 **If** 절에서 변수를 계산해야 합니다. **Match** 조건과 **If** 조건의 비교는 " 비교 함수 "(689 페이지) 에서 확인하십시오.
- **If** 및 **Match** 를 사용할 경우 계산식 편집기는 시퀀스의 맨 위부터 아래로 검색하여 첫 번째 true 절을 찾고 해당하는 결과 표현식의 값을 구합니다. 이후의 true 절은 무시됩니다.

다음 예에서는 *age* 변수의 값에 따라 각각의 값이 할당됩니다.

그림 A.5 Match 함수의 사용 예



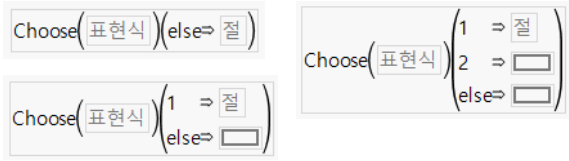
참고 : Match 는 후행 공백을 무시하지만 **If** 는 그렇지 않습니다 .

Match는 결측값에 대해 결측값을 반환하지만 명시적으로 결측값을 매칭할 수도 있습니다.

Choose

Choose는 조건의 인수가 1부터 시작하는 정수 수열인 특수한 형태의 **Match**입니다. **절**의 값은 매칭 조건을 대체합니다. 그림 A.6에서는 **Choose** 조건의 예를 보여 줍니다. **Choose**를 사용할 경우 계산식 편집기는 올바른 선택 절로 바로 이동하여 결과 표현식을 계산합니다.

그림 A.6 Choose 조건의 예



표현식을 강조 표시하고 **Choose**를 클릭하면 계산식 편집기가 한 개의 **절**을 포함하는 새 조건부 표현식을 생성합니다. 새 절을 추가하거나 필요하지 않은 절을 제거하려면 **If** 조건에서 이미 설명한 대로 삽입( ) 또는 삭제( ) 버튼을 사용하십시오.

Choose는 **Choose** 표현식을 계산하고 즉시 해당하는 결과 표현식으로 이동하여 반환 값을 생성합니다. **Choose**를 사용할 경우에는 1부터 시작하여 순차적 정수를 한 번씩만 생성하는 선택 표현식을 지정한 후 수열의 각 정수를 지정해야 합니다.

IfMax

각 인수 쌍의 첫 번째 인수를 계산하고 표현식의 최대값과 연결된 결과 표현식(각 쌍의 두 번째 인수)에 대한 계산 결과를 반환합니다. 최대값을 갖는 표현식이 둘 이상이면 첫 번째 최대값이 반환됩니다. 모든 표현식이 결측되고 최종 결과가 지정되지 않은 경우에는 결측값이 반환됩니다. 모든 표현식이 결측되고 최종 결과가 지정된 경우에는 해당 최종 결과가 반환됩니다. 테스트 표현식의 결과는 숫자 값이어야 하지만 결과 표현식은 무엇이든 될 수 있습니다.

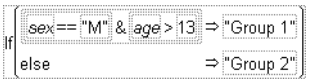
IfMin

각 인수 쌍의 첫 번째 인수를 계산하고 표현식의 최소값과 연결된 결과 표현식(각 쌍의 두 번째 인수)에 대한 계산 결과를 반환합니다. 최소값을 갖는 표현식이 둘 이상이면 첫 번째 최소값이 반환됩니다. 모든 표현식이 결측되고 최종 결과가 지정되지 않은 경우에는 결측값이 반환됩니다. 모든 표현식이 결측되고 최종 결과가 지정된 경우에는 해당 최종 결과가 반환됩니다. 테스트 표현식의 결과는 숫자 값이어야 하지만 결과 표현식은 무엇이든 될 수 있습니다.

And &

두 인수가 모두 true일 때 결과가 1이 됩니다. 그렇지 않으면 결과가 0이 됩니다(그림 A.9). 그림 A.7의 계산식은 두 비교가 모두 true인 경우에만 Group 1 운전자 라벨을 적용합니다.

그림 A.7 And 함수 생성



Or |

인수 중 하나가 true일 때 결과가 1이 됩니다. 두 인수가 모두 false인 경우 Or 표현식의 결과는 0이 됩니다(그림 A.9). 그림 A.8의 계산식은 남성과 13세 이상의 모든 참가자를 Group 1에 할당합니다.

그림 A.8 Or 함수 생성

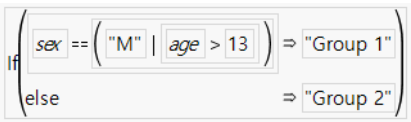
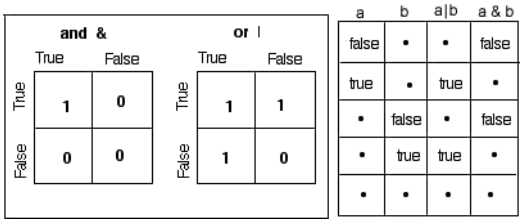


그림 A.9의 왼쪽에 있는 true/false 테이블에서는 두 인수가 모두 true 또는 false가 되는 비결측값일 때 **And**(&) 및 **Or**(|) 함수의 결과를 보여 줍니다. 오른쪽의 테이블에서는 왼쪽 또는 오른쪽 표현식(**a** 및 **b**) 중 하나가 결측값이거나 둘 모두 결측값일 때의 결과를 보여 줍니다.

그림 A.9 And 및 Or 표현식의 결과



a	b	a|b	a & b
false	•	•	false
true	•	true	•
•	false	•	false
•	true	true	•
•	•	•	•

Not !

인수가 false일 때 결과가 1이 됩니다. 그렇지 않은 경우 **Not**의 결과는 **0**이 됩니다. **Not** 함수를 적용할 경우 모호성을 방지하기 위해 필요하면 괄호를 사용합니다. `Not`은 0 또는 1만 반환할 수 있으므로 예를 들어 `!(weight==64)`는 true 또는 false(1 또는 0)일 수 있지만 `(!weight)==64`는 항상 false(0)입니다. `!(weight==64)`와 같은 표현식을 `weight != 64`로 입력할 수도 있습니다.

Interpolate

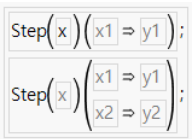
두 점, 즉 지정된 인수에 해당하는 *x1, y1*과 *x2, y2* 사이에서 *y* 값을 선형으로 보간합니다. 삽입 키를 사용하여 *x, y* 인수 쌍을 추가로 삽입할 수 있습니다. **Interpolate**는 *x* 값에 해당하는 *x, y* 점의 쌍을 찾고 보간을 완료합니다.

그림 A.10 Interpolate 의 예

Step

Step은 **Interpolate**와 유사하지만, *x* 및 *y* 인수보다 작거나 같으면서 가장 큰 *x* 값에 해당하는 *y* 값을 반환한다는 점이 다릅니다. 즉, 이 함수는 점 사이의 선형 적합이 아니라 step 함수에서 지정된 *x*에 대해 해당하는 *y*를 구합니다. **Interpolate**와 마찬가지로, *x* 및 *y* 인수 쌍을 원하는 만큼 추가할 수 있습니다.

그림 A.11 Step 의 예



For

while 조건이 true일 때 body 인수의 문을 반복합니다. **init**와 **next**는 반복을 제어합니다.

While

expr 조건을 반복적으로 테스트하고 **expr**이 더 이상 true가 아닐 때까지 **body**를 실행합니다.

Break, Continue

Break는 루프의 실행을 완전히 중지하고 계속해서 루프 다음의 문을 실행합니다. **Continue**는 현재 루프 반복을 끝내고 다음 반복의 루프를 시작합니다.

둘 모두 **For, While** 및 **For Each Row** 루프에서 사용됩니다.

Stop

실행 중인 스크립트를 즉시 중지합니다.

Return

사용자 정의 함수에서 표현식 값을 반환합니다.

확률 함수

베타, 카이제곱, F, 감마, 정규, 스튜던트 t 및 Weibull 분포를 비롯한 통계 분포에 대해 확률 및 분위수를 계산하는 계산식을 작성할 수 있습니다. 분포 파라미터화 및 함수 인수에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Beta Density

Beta Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Beta Distribution

Beta Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Beta Quantile

Beta Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Cauchy Density

Cauchy Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Cauchy Distribution

Cauchy Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Cauchy Quantile

Cauchy Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

ChiSquare Density

ChiSquare Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

ChiSquare Distribution

ChiSquare Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

ChiSquare Log CDistribution

ChiSquare Log 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

ChiSquare Log Density

ChiSquare Log Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

ChiSquare Log Distribution

ChiSquare Log Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

ChiSquare Noncentrality

ChiSquare Noncentrality 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

ChiSquare Quantile

ChiSquare Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Dunnett P Value

Dunnett P Value 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Dunnett Quantile

Dunnett Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Exp Density

Exp Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Exp Distribution

Exp Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Exp Quantile

Exp Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Density

F Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Distribution

F Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Log CDistribution

F Log CDistribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Log Density

F Log Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Log Distribution

F Log Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Noncentrality

F Noncentrality 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Power

F Power 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Quantile

F Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

F Sample Size

F Sample Size 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Frechet Density

Frechet Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Frechet Distribution

Frechet Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Frechet Quantile

Frechet Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Density

Gamma Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Distribution

Gamma Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Log CDistribution

Gamma Log CDistribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Log Density

Gamma Log Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Log Distribution

Gamma Log Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Quantile

Gamma Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

GenGamma Density

GenGamma Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

GenGamma Distribution

GenGamma Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

GenGamma Quantile

GenGamma Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

GLog Density

GLog Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

GLog Distribution

GLog Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

GLog Quantile

GLog Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Sb Density

Johnson Sb Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Sb Distribution

Johnson Sb Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Sb Quantile

Johnson Sb Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Sl Density

Johnson Sl Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Sl Distribution

Johnson Sl Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Sl Quantile

Johnson Sl Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Su Density

Johnson Su Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Su Distribution

Johnson Su Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Johnson Su Quantile

Johnson Su Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

LEV Density

LEV(Largest Extreme Value) Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

LEV Distribution

LEV(Largest Extreme Value) Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

LEV Quantile

LEV(Largest Extreme Value) Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

LogGenGamma Density

LogGenGamma Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

LogGenGamma Distribution

LogGenGamma Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

LogGenGammaQuantile

LogGenGamma Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Logistic Density

Logistic Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Logistic Distribution

Logistic Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Logistic Quantile

Logistic Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Loglogistic Density

Loglogistic Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Loglogistic Distribution

Loglogistic Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Loglogistic Quantile

Loglogistic Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Lognormal Density

Lognormal Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Lognormal Distribution

Lognormal Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Lognormal Quantile

Lognormal Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Biv Distribution

Normal Biv Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Density

Normal Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Distribution

Normal Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Log CDistribution

Normal Log CDistribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Log Density

Normal Log Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Log Distribution

Normal Log Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Mixture Density

Normal Mixture Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Mixture Distribution

Normal Mixture Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Mixture Quantile

Normal Mixture Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Normal Quantile

Normal Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

SEV Density

SEV(Smallest Extreme Value) Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

SEV Distribution

SEV(Smallest Extreme Value) Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

SEV Quantile

SEV(Smallest Extreme Value) Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

SHASH Density

SHASH Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

SHASH Distribution

SHASH Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

SHASH Quantile

SHASH Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

t Density

t Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

t Distribution

t Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

t Log CDistribution

t Log CDistribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

t Log Density

t Log Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

t Log Distribution

t Log Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

t Noncentrality

t Noncentrality 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

t Quantile

t Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Tukey HSD P Value

Tukey HSD P Value 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Tukey HSD Quantile

Tukey HSD Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Weibull Density

Weibull Density 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Weibull Distribution

Weibull Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Weibull Quantile

Weibull Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

이산 확률 함수

Beta Binomial Distribution

Beta Binomial Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Beta Binomial Probability

Beta Binomial Probability 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Beta Binomial Quantile

Beta Binomial Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Binomial Distribution

Binomial Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Binomial Probability

Binomial Probability 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Binomial Quantile

Binomial Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Poisson Distribution

Gamma Poisson Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Poisson Probability

Gamma Poisson Probability 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Gamma Poisson Quantile

Gamma Poisson Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Hypergeometric Distribution

Hypergeometric Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Hypergeometric Probability

Hypergeometric Probability 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Neg Binomial Distribution

Neg Binomial Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Neg Binomial Probability

Neg Binomial Probability 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Poisson Distribution

Poisson Distribution 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Poisson Probability

Poisson Probability 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Poisson Quantile

Poisson Quantile 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

통계 함수

계산식에 사용할 수 있는 **통계** 함수에는 다음 두 가지 유형이 있습니다.

- 이름에 접두사 **Col** 이 있는 함수: 이러한 함수는 숫자 열 또는 열과 관련된 표현식에 대한 통계량을 계산합니다.
- **Mean**, **Std Dev**, **Number**, **Sum**, **Quantile**, **Maximum**, **Minimum** 및 **N Missing** 함수: 이러한 함수는 여러 열 또는 인수를 대상으로 계산됩니다. 여러 인수를 사용하여 각 행에 대한 통계량이 계산됩니다. 여러 인수를 사용할 수 있는 함수의 경우 계산식 편집기 키패드에 있는 삽입 키 ( ) 를 클릭하거나 쉼표를 입력하여 인수를 추가할 수 있습니다. 여러 개의 인접한 인수가 있는 경우에는 함수를 선택하고 첫 번째 인수를 선택한 후 Shift 키를 누른 채 그룹의 마지막 인수를 클릭하십시오. 그러면 해당 함수가 자동으로 전체 목록과 함께 표시됩니다.

참고 : 통계 함수에서는 결측값이 무시됩니다.

구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Col Cumulative Sum

현재 행에 대한 누적 합을 반환합니다. Col Cumulative Sum은 정렬할 필요가 없는 기준 열을 지원합니다.

Col Mean

인수로 지정된 숫자 값의 평균(산술평균)을 계산합니다. 계산식 **Col Mean(age)**는 **age** 열에 있는 모든 비결측값의 평균을 계산합니다.

Col Moving Average

현재 행을 기준으로 지정된 간격의 이동 평균을 반환합니다. 결측값은 무시됩니다.

Col Std Dev

인수로 지정된 분포의 평균을 중심으로 한 퍼짐 정도를 측정합니다. 정규 분포에서는 분포의 약 68%가 평균으로부터 1 표준편차 범위 내에 있습니다. 또한 분포의 95%는 평균으로부터 2 표준편차 범위 내에 있고, 99%는 평균으로부터 3 표준편차 범위 내에 있습니다.

Col Number

지정한 열에 있는 비결측값의 수를 계산합니다. 숫자 결측값은 셀에 할당된 값이 없거나 셀에 올바르지 않은 연산(예: 0으로 나누기)의 결과가 있는 경우에 발생합니다. 스프레드시트에서 결측값은 결측값 표식(•)으로 표시됩니다. 문자 결측값은 null 문자열입니다. 행 상태 열에 대한 계산식에서 제외된 행 상태 특성은 결측값으로 간주됩니다.

Col N Missing

지정한 열에 있는 결측값의 수를 계산합니다. 숫자 결측값은 셀에 할당된 값이 없거나 셀에 올바르지 않은 연산(예: 0으로 나누기)의 결과가 있는 경우에 발생합니다. 데이터 격자에서 결측값은 결측값 문자(•)와 함께 표시됩니다. 문자 결측값은 null 문자열입니다.

Col Sum

숫자 인수에 있는 값의 합을 계산합니다. 결측값은 무시됩니다.

Col Minimum 및 Col Maximum

숫자 인수의 최소값을 구합니다. **Col Minimum**은 결측값을 무시합니다. **Col Maximum**은 숫자열 인수의 최대값을 구하며 결측값을 무시합니다.

Col Quantile

지정된 비율만큼의 값이 해당 값보다 작거나 같은 지점에 있는 값을 계산합니다. 예를 들어 50% 분위수로 계산된 값(중앙값이라고도 함)은 데이터의 50%보다 크거나 같습니다. 데이터 값의 절반은 50번째 분위수보다 작습니다.

Col Quantile 함수의 quantile 인수는 100으로 나눈 분위수 백분율을 나타냅니다. 25% 분위수(1사분위수라고도 함)는 $p = 0.25$에 해당하고, 75% 분위수(3사분위수라고도 함)는 $p = 0.75$에 해당합니다.

계산식 편집기에서는 n개의 비결측값이 있는 열에 대해 값을 오름차순으로 배열하여 분위수를 계산합니다. 정렬된 열 값의 서브스크립트 $y_1, y_2, \ldots, y_n$은 오름차순 순위를 나타냅니다.

p번째 분위수 값은 계산식 $p(n + 1)$을 사용하여 계산됩니다. 여기서 p는 백분율 값이고 n은 비결측값의 총 수입니다. $p(n + 1)$이 정수이면 분위수 값은 $y_{p(n+1)}$이 됩니다. $p(n + 1)$이 정수가 아니면 i에 결과의 정수부를 할당하고 f에 소수부를 할당한 후 계산식 $(1 - f)y_i + (f)y_{i+1}$을 적용하여 값이 보간됩니다.

예를 들어 열에 값 2, 4, 6, 8, 10, 12, 14, 16, 18 및 20이 있는 경우, 50% 분위수는 `0.5(10 + 1) = 5.5`로 계산됩니다.

결과가 소수이므로 50% 분위수 값은 다음과 같이 보간됩니다.

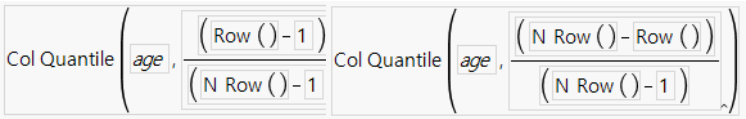
$$(1 - 0.5) \times 10 + (0.5) \times 12 = (0.5)10 + (0.5)12 = 6 + 5 = 11$$

다음은 **ColQuantile** 계산식의 예입니다.

- **ColQuantile(age, 1)** Age의 최대값을 계산합니다.
- **ColQuantile(age, 0.75)** Age의 3사분위수를 계산합니다.
- **ColQuantile(age, 0.5)** Age의 중앙값을 계산합니다.
- **ColQuantile(age, 0.25)** Age의 1사분위수를 계산합니다.
- **ColQuantile(age, 0)** Age의 최소값을 계산합니다.

ColQuantile 인수는 0과 1 사이의 값(경계값 포함)으로 계산되는 표현식일 수 있습니다. 예를 들어 그림 A.12의 첫 번째 계산식은 각 행에 대해 오름차순 분위수 값을 계산합니다. 그런 다음 계산된 열에는 보간된 오름차순 **age** 값이 포함됩니다. 두 번째 계산식은 보간된 age 값을 내림차순으로 나열합니다.

그림 A.12 Quantile 함수의 예



Col Rank

각 행의 값을 순위화합니다. 가장 낮은 값에는 순위 1이 할당되고 가장 높은 값에는 비결측 열의 수에 해당하는 순위가 할당됩니다. 동순위가 발생할 경우 가능한 순위의 평균을 보고하거나, 원래 나타나던 순서대로 순위를 할당하거나, 두 값 모두에 가능한 가장 낮은 순위를 부여하거나, 임의로 순위를 매길 수 있습니다. JMP 12에서는 임의 방식으로 동순위를 처리합니다. 분포의 정규 분위수 출력에서는 평균을 사용하여 동순위를 처리합니다.

Col Standardize

숫자 표현식에 대해 일반적인 표준화를 수행합니다. 각 행 i에 대해 **Col Standardize(height)**는 **(Height$_{Row()}$ – Col Mean(Height)) / Col Std Dev(Height)**입니다.

평균

지정한 인수 목록에 있는 비결측값의 산술평균을 계산합니다. 인수는 상수, 숫자 또는 표현식일 수 있습니다. **Mean** 함수는 기본적으로 단일 인수와 함께 표시됩니다. 계산식 편집기의 키패드에 있는 삽입 버튼(▲)을 사용하거나 쉼표를 입력하여 인수를 추가할 수 있습니다.

Std Dev

지정한 인수 목록에 있는 비결측값의 표준편차를 계산합니다. 인수는 상수, 숫자 또는 표현식일 수 있습니다. **Std Dev** 함수는 기본적으로 단일 인수와 함께 표시됩니다. 계산식 편집기의 키패드에 있는 삽입 버튼(▲)을 클릭하거나 쉼표를 입력하여 인수를 추가할 수 있습니다.

Number

지정한 인수 목록에 있는 비결측값의 수를 계산합니다.

Sum

지정한 인수 목록에 있는 비결측값의 합을 반환합니다.

Quantile

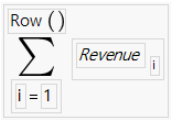
지정한 인수 목록의 모든 비결측값에 대해 첫 번째 인수로 지정된 값이 나머지 목록에 대해 갖는 분위수를 계산합니다.

Summation(Σ)

합 인덱스로 지정된 열의 명시적 값 범위에 대해 합을 계산합니다. 이 동작은 **항상 모든 행**에 대해 계산하는 다른 모든 통계 함수(**Product** 제외)와는 다릅니다. **Summation** 함수는 그림 A.13에 표시된 합 표기를 사용합니다. 합을 계산하려면 빈 body 항을 인덱스 변수 *i* 또는 사용자가 할당한 인덱스 변수를 포함하는 표현식으로 바꾸십시오. **Summation**은 $i = 1, i = 2, ..., i =$ **NRow()**에 대해 표현식을 반복적으로 계산한 후 비결측 결과를 모두 더해 최종 결과를 결정합니다.

활성 스프레드시트의 행 수를 나타내는 **NRow()**와 인덱스 상수 *i*는 계산식에 적절한 표현식으로 바꿀 수 있습니다. 예를 들어 그림 A.13의 Summation 계산식에서는 1행부터 현재 행 번호까지의 모든 revenue 값에 대한 행별 총합을 계산하고 계산 열에 revenue 열의 누적 총합을 채웁니다.

그림 A.13 Summation 함수의 예



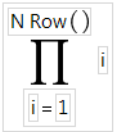
Product(Π)

항상 모든 행에 대해 계산되는 다른 모든 통계 함수(**Summation** 제외)와 달리 합 인덱스로 지정된 열의 명시적 값 범위에 대해 곱을 계산합니다. **Product**는 그림 A.14의 계산식에 표시된 표기를 사용합니다. 곱을 계산하려면 빈 body 항을 인덱스 변수 *j*를 포함하는 표현식으로 바꾸십시오. **Product**는 $i = 1, i = 2, ..., i = n$에 대해 표현식을 반복적으로 계산한 후 비결측 결과를 모두 곱해 최종 결과를 결정합니다.

활성 스프레드시트의 행 수를 나타내는 **NRow()**와 인덱스 상수 *i*는 계산식에 적절한 표현식으로 바꿀 수 있습니다.

예를 들어 그림 A.14의 Product 예에서는 *i*!(각 행 번호의 계승)을 계산합니다.

그림 A.14 Product 함수의 예



Minimum 및 Maximum

지정한 비결측 인수 목록의 최소값 및 최대값을 각각 반환합니다.

N Missing

지정한 인수 목록에 있는 결측값의 수를 계산합니다.

SSQ

모든 요소의 제곱합을 반환합니다.

난수 함수

지정한 분포의 제약 조건 내에서 효과적으로 "주사위를 굴려" 난수를 추출하는 계산식을 생성할 수 있습니다. 이 함수는 사용자가 계산식 편집기 창에서 **적용**을 클릭할 때마다 새로운 난수 집합을 생성합니다.

참고 : 난수는 Mersenne-Twister 기술을 사용하여 생성됩니다 . 이 기술의 주기 길이는 2^{19937}-1 입니다 . 이 난수 생성기에 대한 자세한 내용은 "Matsumoto and Nishimura"(1998) 를 참조하십시오 . 이 새로운 난수 생성기들은 "Marshalled"(1996) 에 설명된 모든 DIEHARD 검증을 통과하는 것으로 확인되었습니다 .

함수 인수에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오. 분포 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random ChiSquare

지정한 자유도를 갖는 카이제곱 분포에서 난수 열을 생성합니다. 선택적인 비중심성 모수는 0보다 크거나 같아야 합니다. 기본적으로 비중심성 모수는 0입니다. 카이제곱 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random F

지정한 분자 자유도 및 분모 자유도를 갖는 F 분포에서 난수 열을 생성합니다. 선택적인 비중심성 모수는 0보다 크거나 같아야 합니다. 기본적으로 비중심성 모수는 0입니다. F 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Index

1에서 *n* 사이의 랜덤 정수로 구성된 *k* x 1 행렬을 반환합니다(중복 값 없음).

Random Uniform

균등 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 균등 분포의 최소값은 0이고 최대값은 1입니다. 즉, 0과 1 사이의 모든 수가 다른 수와 동일한 확률로 추출됩니다. 결과는 거의 균일한 분포를 보입니다. 선택적 인수를 사용하여 균등 분포의 최소 숫자와 최대 숫자에 다른 값을 지정할 수 있습니다.

Random Normal

지정한 평균 및 표준편차를 갖는 정규 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 정규 분포의 평균은 0, 표준편차는 1로 간주됩니다. 정규 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Normal Mixture

지정된 인수가 있는 정규 혼합 분포에서 난수를 반환합니다. Random Normal Mixture 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Exp

람다=1인 지수 분포에서 난수 열을 생성합니다. 다른 람다를 사용하도록 지수 함수의 척도를 조정할 수 있습니다. 예를 들어 `Random Exp() * 0.1`은 람다=0.1인 지수 분포를 생성합니다. 지수 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Gamma

지정한 형태 및 척도 모수를 갖는 감마 분포에서 난수 열을 생성합니다. 척도 모수의 기본값은 1입니다. 척도 모수가 1인 경우 감마 분포는 지수 분포와 동등합니다. 감마 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random GenGamma

지정한 위치, 척도 및 형태 모수를 갖는 확장된 일반화 감마 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 일반화 감마 분포의 위치는 0, 척도는 1, 형태는 0입니다. 일반화 감마 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random GenGamma

지정한 위치, 척도 및 형태 모수를 갖는 로그 일반화 감마 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 확장된 일반화 감마 분포의 위치는 0, 척도는 1, 형태는 0입니다. 로그 일반화 감마 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Beta

두 개의 지정한 형태 모수와 선택적인 임계 및 척도 모수를 갖는 베타 분포에서 난수 열을 생성합니다. 기본 임계는 0이고 기본 척도는 1입니다. 베타 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Cauchy

위치 모수가 0이고 척도 모수가 1인 Cauchy 분포에서 난수 열을 생성합니다. 계산식 `alpha + beta * Random Cauchy()`를 사용하면 위치 모수가 alpha이고 척도 모수가 beta인 Cauchy 변량을 생성할 수 있습니다. Cauchy 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Category

확률 및 결과 표현식 집합을 기준으로 랜덤 범주 할당 열을 생성합니다(예: `Random Category(.2, "A", .3, "B", .4, "C", "D");`). 균등 난수가 생성되고 반환할 결과 표현식을 결정하기 위해 확률 인수와 비교됩니다.

Random Johnson Su

지정한 형태 모수 두 개, 위치 모수 한 개, 척도 모수 한 개를 갖는 Johnson Su 분포에서 난수 열을 생성합니다. Johnson Su 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Johnson Sb

지정한 형태 모수 두 개, 위치 모수 한 개, 척도 모수 한 개를 갖는 Johnson Sb 분포에서 난수 열을 생성합니다. Johnson Sb 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Johnson Sl

지정한 형태 모수 두 개, 위치 모수 한 개, 선택적 왜도 모수 한 개를 갖는 Johnson Sl 분포에서 난수 열을 생성합니다. 왜도 모수는 양수 왜도의 경우 +1, 음수 왜도의 경우 -1일 수 있습니다. 기본값은 +1입니다. Johnson Sl 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Seed State

BLOB 개체에서 난수 시드값 상태를 가져오거나 복원합니다.

Random Triangular

지정한 최소값, 최대값 및 최빈값 모수를 갖는 삼각 분포에서 난수 열을 생성합니다. 인수를 하나만 지정할 경우 해당 인수는 최빈값을 나타내며 삼각 분포의 범위는 0에서 1 사이가 됩니다. 인수를 두 개 지정할 경우 해당 인수는 각각 최빈값과 최대값을 나타내며 삼각 분포의 범위는 0에서 최대값 사이가 됩니다. 인수를 세 개 지정할 경우 해당 인수는 각각 최소값, 최빈값 및 최대값을 나타내며 삼각 분포의 범위는 최소값에서 최대값 사이가 됩니다.

Random Integer

정수의 균등 분포에서 난수 열을 생성합니다. 인수를 하나만 지정할 경우 정수의 범위는 1에서 해당 인수 사이가 됩니다. 인수를 두 개 지정할 경우 정수의 범위는 두 인수 값의 사이(경계값 포함)가 됩니다.

Random Binomial

지정한 모수를 갖는 이항 분포에서 난수 열을 생성합니다. 첫 번째 인수는 이항 실험의 시도 횟수를 나타내는 n입니다. 두 번째 인수는 관심 사건의 발생 확률을 나타내는 p입니다. n이 1일 때 이

이항 함수는 베르누이 시행의 분포를 생성합니다. 이항 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Negative Binomial

지정한 모수를 갖는 음이항 분포에서 난수 열을 생성합니다. 첫 번째 인수는 관심 사건의 성공 횟수를 나타내는 r입니다. 두 번째 인수는 성공 확률을 나타내는 p입니다. 관심 있는 확률 변수는 r번째 성공 이전에 발생한 실패 횟수입니다. 시도 횟수가 고정적이고 성공 횟수가 가변적인 이항 변량과 달리 음이항 변량은 성공 횟수가 고정적이고 시도 횟수가 무작위한 경우에 사용됩니다. 음이항 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Beta Binomial

지정한 모수를 갖는 베타 이항 분포에서 난수 열을 생성합니다. 첫 번째 인수는 시도 횟수를 나타내는 n입니다. 두 번째 인수는 각 시도의 성공 확률을 나타내는 p입니다. 세 번째 인수는 선택적 인수로, 과대산포 모수를 나타냅니다. 기본 과대산포 모수는 0입니다. 베타 이항 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Frechet

지정한 위치 및 척도 모수를 갖는 Fréchet 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 Fréchet 분포의 위치 모수는 0, 시그마는 1이 됩니다. Fréchet 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Geometric

지정한 확률 모수를 갖는 기하 분포에서 난수 열을 생성합니다. 인수는 한 번의 시도에서 특정 사건이 발생할 확률입니다. 관심 있는 확률 변수는 특정 사건이 처음 발생할 때까지의 시도 횟수입니다.

Random Poisson

지정한 형태 모수를 갖는 Poisson 분포에서 난수 열을 생성합니다. Poisson 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Gamma Poisson

지정한 형태 모수와 과대산포 모수를 갖는 감마 Poisson 분포에서 난수 열을 생성합니다. 두 번째 인수를 지정하지 않을 경우 감마 Poisson 분포의 과대산포 모수는 1이 됩니다. 감마 Poisson 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Weibull

지정한 형태 및 척도 모수를 갖는 Weibull 분포에서 난수 열을 생성합니다. 두 번째 인수를 지정하지 않을 경우 Weibull 분포의 척도 모수는 1이 됩니다. Weibull 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Logistic

지정한 위치 및 척도 모수를 갖는 로지스틱 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 로지스틱 분포의 위치 모수는 0, 척도 모수는 1이 됩니다. 로지스틱 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Loglogistic

지정한 위치 및 척도 모수를 갖는 로그로지스틱 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 로그로지스틱 분포의 위치 모수는 0, 척도 모수는 1이 됩니다. 로그로지스틱 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Lognormal

지정한 위치 및 척도 모수를 갖는 로그 정규 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 로그 정규 분포의 위치 모수는 0, 척도 모수는 1이 됩니다. 로그 정규 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random GLog

지정한 위치, 척도 및 형태 모수를 갖는 일반화 로그 분포에서 난수 열을 생성합니다. 형태 모수가 0일 경우 일반화 로그 분포는 로그 분포와 동등합니다. 일반화 로그 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Reset

지정한 시드값을 사용하여 난수 시퀀스를 다시 시작합니다.

Random LEV

지정한 위치 및 척도 모수를 갖는 LEV(최대 극단값) 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 LEV 분포의 위치 모수는 0, 척도 모수는 1이 됩니다. LEV 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random SEV

지정한 위치 및 척도 모수를 갖는 SEV(최소 극단값) 분포에서 난수 열을 생성합니다. 인수를 지정하지 않을 경우 SEV 분포의 위치 모수는 0, 척도 모수는 1이 됩니다. SEV 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random SHASH

지정한 형태 모수 두 개, 위치 모수 한 개, 척도 모수 한 개를 갖는 SHASH(SinH-ArcSinH) 분포에서 난수 열을 생성합니다. SHASH 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Random Shuffle

행렬 인수의 숫자에서 임의 순서의 숫자 열을 생성합니다.

Random t

지정한 자유도를 갖는 t 분포에서 난수 열을 생성합니다. 선택적인 비중심성 인수는 음수 또는 양수일 수 있습니다. 기본적으로 비중심성 모수는 0입니다. t 분포의 파라미터화에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Col Shuffle

현재 데이터 테이블에서 임의의 행 번호를 선택합니다. 각 행 번호는 한 번씩만 선택됩니다. **Col Shuffle**을 서브스크립트로 사용할 경우 이 함수는 인수 역할을 하는 열에서 무작위로 선택된 값을 반환합니다. 원래 열의 각 값은 한 번씩만 **Col Shuffle**의 결과로 할당됩니다. Col Shuffle 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

Resample Freq

랜덤 복원 추출의 빈도 수를 생성하며 붓스트랩에 유용합니다. 예를 들어 이 함수는 두 번째 인수인 **Freq Column**을 지원하며, 따라서 두 번째 인수에 지정된 기존 빈도 열과 관련하여 붓스트랩 표집을 수행할 수 있습니다. `Resample Freq()`는 100% 재표집 표본을 생성합니다. `ResampleFreq(rate)`는 `rate`로 지정된 빈도의 표본을 생성합니다. `Resample(rate, column)`은 rate에 지정된 열의 합을 곱한 비율로 계산된 표본을 생성합니다. Resample Freq 함수의 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

날짜 / 시간 함수

JMP는 날짜 및 시간을 숫자 열에 저장할 때 macOS 표준인 1904년 1월 1일 이후 경과한 초 수를 사용합니다. 열에 날짜 값이 있는 경우 열 이름을 두 번 클릭하고 **형식** 메뉴에서 **날짜** 또는 **시간**을 선택하여 해당 열에 날짜 형식을 할당할 수 있습니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "숫자 형식 옵션"(273페이지)에서 확인하십시오.

구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

In Minutes, In Hours, In Days, In Weeks, In Years

함수 이름의 단위를 인수와 동등한 초 수로 변환합니다. 인수는 숫자이거나 숫자 표현식이어야 합니다. 예를 들어 **In Minutes(2)**는 **120**으로 산출되고, **In Years(1)**은 **31,557,600**(60초 * 60분 * 24시간 * 365.25일)으로 산출됩니다.

Date DMY, Date MDY

일, 월 및 연도에 대한 숫자 표현식을 받아 연관된 JMP 날짜를 반환합니다. 예를 들어 **Date DMY (20, 3, 1991)**과 **Date MDY(3, 20, 1991)**의 결과는 **2,752,272,000**이 됩니다.

Today

1904년 1월 1일부터 현재 날짜까지의 초 수를 반환합니다. 예를 들어 1991년 3월 20일(수요일) 자정에 **Today** 함수는 **2752272000**(2,752,272,000초)을 반환하고 계산을 계속합니다. 같은 날 더 나중에 **Today** 함수를 실행하면 추가된 초 수가 반영됩니다.

Day, Month, Year

각각 월간 일자, 월(1에서 12 사이의 숫자), 4자리 연도를 반환합니다. 이러한 함수의 인수는 JMP 날짜로 해석됩니다. 예를 들어 1991년 3월 20은 다음과 같이 계산됩니다.

- **Day(2752272000)** 은 숫자 **20** 을 반환합니다 .
- **Month(2752272000)** 은 숫자 **3** 을 반환합니다 .
- **Year(2752272000)** 은 숫자 **1991** 을 반환합니다 .

Quarter

datetime 값의 연간 분기를 정수 1 ~ 4로 반환합니다.

Hour, Minute, Second

각각 datetime 값의 시, 분 및 초 부분을 반환합니다. 이러한 함수의 인수는 JMP 날짜로 해석됩니다. 예를 들어 1991년 3월 20은 다음과 같이 계산됩니다.

- **Hour(2752572649)** 는 숫자 **11** 을 반환합니다 .
- **Minute(2752572649)** 는 숫자 **30** 을 반환합니다 .
- **Second(2752572649)** 는 숫자 **49** 를 반환합니다 .

Day of Week, Day of Year, Week of Year, Time of Day

이러한 함수의 인수는 JMP 날짜입니다. **Day Of Week**는 1에서 7 사이의 숫자를 반환하며, 여기서 1은 일요일을 나타냅니다. **Day Of Year**는 연초를 기준으로 한 날짜를 반환합니다. **Week Of Year**는 지정된 규칙에 따라 1에서 52 사이의 숫자를 반환합니다. 규칙 1(기본값)에서는 주가 일요일에 시작되며, 첫 번째 일요일은 2주에 속하고 1주는 불완전한 주이거나 빈 주입니다. 규칙 2에서는 첫 번째 일요일부터 1주가 시작되며, 그 이전 날짜는 0주에 속합니다. 규칙 3에서는 주가 월요일에 시작되며, 해당 연도의 날짜가 4일 포함된 첫 번째 주가 1주가 되는 ISO 주 번호를 반환합니다. ISO 주를 사용할 경우 해당 연도의 처음 3일 또는 마지막 3일은 직전 또는 직후 연도의 주 번호가 될 수 있습니다. **Time Of Day**는 0에서 86399 사이의 숫자(초 단위 시간)를 반환합니다. 예를 들어 1991년 3월 20 수요일은 다음과 같이 계산됩니다.

- **Day Of Week(2752272000)** 은 숫자 **4** 를 반환합니다 .
- **Day Of Year(2752272000)** 은 숫자 **79** 를 반환합니다 .
- **Week Of Year(2752272000)** 은 숫자 **12** 를 반환합니다 .
- **Time Of Day(2752272000)** 은 숫자 **0** 을 반환합니다 .

Informat

Informat 함수의 인수는 날짜 문자열입니다. 예를 들어 **Informat("03/20/1991")**은 해당하는 JMP 날짜 값인 **2752272000**을 반환합니다. JMP는 **날짜 약어** 및 **긴 날짜**를 제외한 모든 날짜 형식을 인식할 수 있습니다.

Abbrev Date, Long Date, Short Date

이러한 날짜 함수의 인수는 JMP 날짜입니다. 이러한 함수는 인수의 형식 지정 표현인 날짜 문자열을 반환합니다. 예를 들면 다음과 같습니다.

- **Abbrev Date(2752272000)** 은 **1991-03-20** 을 반환합니다 .
- **Long Date(2752272000)** 은 **1991 년 3 월 20 일 수요일**을 반환합니다 .
- **Short Date(2752272000)** 은 **03-20-1991** 을 반환합니다 .

형식

Format 함수의 첫 번째 인수는 JMP 날짜입니다. 이 함수는 두 번째 인수에서 지정한 날짜 형식에 따라 날짜의 문자열 표현을 인용 부호로 묶어 반환합니다. 숫자 열에 이 함수를 적용하면 해당 열의 데이터 유형이 자동으로 문자로 변경됩니다.

첫 번째 인수에 열을 지정하고 나머지는 비워 둘 수도 있습니다. 결과는 열 참조에 대한 형식이 지정된 값입니다. 이 함수는 값 라벨이 해제된 경우 열의 값 라벨을 추출하는 데 사용할 수 있습니다.

MDYHMS

MDYHMS의 인수는 JMP 날짜입니다. 이 함수는 모든 날짜 및 시간 필드를 표시하며, 시간 정보가 없는 경우에는 일련의 0이 시간 필드로 추가됩니다. 이 함수는 일부 날짜 정보가 표시되지 않도록 날짜 열을 구성하려는 경우에 유용합니다. **MDYHMS** 함수는 사용 가능한 모든 날짜 및 시간 정보를 표시하는 데 사용할 수 있습니다.

Date Increment

시작 날짜/시간 값에 1개 이상의 간격을 더합니다. 예를 들어 **Date Increment(Today(), "Day", 3)**은 현재 날짜에 3일을 더합니다. **Date Increment(Today(), "Year", 3)**은 현재 날짜에 3년을 더합니다.

Date Difference

두 날짜/시간 값의 차이를 반환합니다. 간격 인수는 Second, Minute, Hour, Day, Week, Month, Quarter, Year일 수 있습니다. 다음은 맞춤 인수에 대한 설명입니다.

Start 간격 시작 횟수를 계산하는 데 사용됩니다 .

Actual 정수 형식의 간격을 계산하는 데 사용됩니다 .

Fractional 소수 형식의 간격을 계산하는 데 사용됩니다 .

예를 들어 다음 계산식은 두 날짜 사이의 일수를 나타내는 **207.890243055556**을 반환합니다.

```
Date Difference(
   01Jan2010:00:00:00,
   27Jul2010:21:21:57,
   "Day",
   "fractional"
);
```

다음 계산식은 두 날짜 사이에서 지나간 일수를 나타내는 **207**을 반환합니다.

```
Date Difference(
   01Jan2010:00:00:00,
   27Jul2010:21:21:57,
   "Day",
   "actual"
);
```

다음 계산식은 두 시간 사이에서 지나간 시간 수를 나타내는 9를 반환합니다.

```
Date Difference(
   01Jan2010:00:00:00,
   01Jan2010:09:22:57,
   "Hour",
   "actual"
);
```

다음 계산식은 두 시간 사이에서 새로 시간이 시작된 횟수를 나타내는 1을 반환합니다.

```
Date Difference(
   31Dec2010:23:59:59,
   01Jan2011:00:59:59,
   "Hour",
   "start"
);
```

행 상태 함수

데이터 테이블의 행에는 여섯 가지 특성, 즉 선택, 숨김, 제외, 라벨 지정, 색상 적용 및 표식 적용 상태가 할당될 수 있습니다. 행에 이러한 특성을 한 개 이상 할당한 후 행 상태 데이터 테이블 열을 생성한 경우 행 상태 조건을 계산하고 저장하는 계산식을 생성할 수 있습니다. 자세한 내용은 "JMP 열 특성 설정" 장의 "JMP의 열 특성"(280페이지) 및 "JMP 열 특성 설정" 장의 "행 상태 열에 정보 저장"(277페이지)에서 확인하십시오. 이 계산식은 행 상태 데이터를 문자 및 숫자 데이터와 같은 방식으로 처리합니다.

구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

참고 : 하나의 행에 임의의 행 상태 조합을 할당할 수 있으며, 하나의 행 상태 열에 여러 행 상태가 값으로 포함될 수 있습니다.

표 A.2에서는 각 행 상태 함수에 필요한 인수의 유형과 각 함수의 반환 항목에 대해 설명합니다.

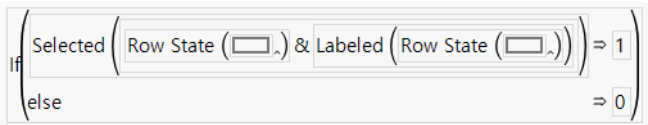
표 A.2 행 상태 함수

함수 이름	필요한 인수 유형	함수 반환 항목 (열 데이터 유형과 동일 유형)
행 상태	없음	현재 행의 행 상태
As Row State	숫자	현재 행의 모든 행 상태
Combine States	여러 행 상태 인수	여러 행 상태 할당
Excluded State	양의 정수 또는 0	행 상태 - 제외됨 또는 제외되지 않음
Hidden State	양의 정수 또는 0	행 상태 - 숨김 또는 숨기지 않음
Labeled State	양의 정수 또는 0	행 상태 - 라벨 지정 또는 라벨 지정 안 함
Color State	정수, 색상 이름 또는 {red, green, blue}	행 상태 색상
Marker State	정수 또는 문자	행 상태 표식
Selected State	양의 정수 또는 0	행 상태 - 선택됨 또는 선택되지 않음
Hue State	정수	행 상태 색조
Shade State	정수 1~5	행 상태 음영 정도
Excluded	**Row State()** 또는 행 상태 열	정수 0(제외되지 않음) 또는 1(제외됨)
Hidden	**Row State()** 또는 행 상태 열	정수 0(숨김) 또는 1(숨기지 않음)
Labeled	**Row State()** 또는 행 상태 열	정수 0(라벨 지정 안 함) 또는 1(라벨 지정)
Color Of	**Row State()** 또는 행 상태 열	색상 맵 정수
Marker Of	**Row State()** 또는 행 상태 열	표식 맵 정수
선택됨	**Row State()** 또는 행 상태 열	정수 0(선택되지 않음) 또는 1(선택됨)

Row State

현재 행의 활성 행 상태 조건을 true 또는 false로 반환합니다. 이 함수를 사용하여 현재 행의 상태에 따라 달라지는 조건부 절을 간편하게 작성할 수 있습니다. 예를 들어 그림 A.15에서는 현재 선택되고 라벨이 지정된 행에 1을 할당하고, 그렇지 않은 행에 0을 할당합니다.

그림 A.15 Row State



As Row State

숫자 인수를 행 상태 또는 행 상태 집합 조건으로 변환합니다. 행 상태는 JMP 내부에 16비트 숫자로 저장되며, 각 비트는 그림 A.3에 표시된 것과 같이 가능한 행 상태 중 하나를 나타내는 데 할당됩니다. 예를 들어 **1327**의 이진 표현은 **0000010100101111**입니다. 따라서 **As Row State(1327)**은 해당 행 상태를 선택됨, 제외됨, 숨김, 라벨 지정됨 상태로 설정하고 표식 2와 색상 10을 적용합니다.

표 A.3 16 비트 숫자로 저장된 행 상태 (각 비트가 하나의 행 상태를 나타냄)

비트	행 상태
0	선택되지 않음(0) 또는 선택됨(1)
1	제외되지 않음(0) 또는 제외됨(1)
2	숨기지 않음(0) 또는 숨김(1)
3	라벨 지정되지 않음(0) 또는 라벨 지정됨(1)
4~7	표식
8~14	색상

Combine States

두 개 이상의 인수를 사용하여 행 상태 조합을 생성합니다. **Combine States** 함수에 인수를 추가하려면 계산식 편집기 키패드의 삽입 버튼(^)을 사용하거나 쉼표를 입력하십시오. **Combine States**를 선택할 때 현재 선택된 표현식이 첫 번째 인수가 됩니다. 각 인수를 결과가 행 상태가 되는 표현식으로 **바꾸십시오**. 다음 계산식을 예로 들겠습니다.

```
Combine States(
   Selected State( Modulo( Row(), 2 ) ),
   Labeled State( Modulo( Row() + 1, 2 ) ) );
```

이 계산식은 계산 대상 행 상태 열의 각 행을 교대로 라벨을 지정하거나 선택합니다. **Selected State** 및 **Labeled State** 함수는 이 섹션의 뒷부분에서 설명합니다. 인수를 추가하거나 필요하지 않은 인수를 제거하려면 계산식 편집기 키패드의 삽입(^) 또는 삭제(⌫) 버튼을 사용하십시오.

하나의 조합에 서로 충돌하는 행 상태를 포함하면 예측할 수 없는 결과가 발생합니다.

Excluded State

숫자 인수를 true 또는 false로 해석합니다. 인수가 true로 계산되면 **Excluded State** 함수는 해당 행의 열 값으로 제외됨 조건을 할당합니다. 예를 들어 **Excluded State(Modulo(Row(),2))**는 각 홀수 번호 행의 행 상태 열 값으로 제외됨 행 상태를 할당합니다.

Hidden State

인수가 0보다 클 경우 숨김 행 상태 조건을 할당합니다. 인수가 0이면 해당 행의 열 값은 숨겨지지 않습니다.

Labeled State

인수가 0보다 클 경우 라벨 지정됨 행 상태 조건을 할당합니다. 인수가 0이면 해당 행의 열에 있는 행 값에 라벨이 지정되지 않습니다.

Color State

JMP 색상 맵에서 정수 인수에 해당하는 색상을 반환합니다. JMP 색상은 0부터 84까지의 번호로 되어 있습니다. 0은 검정에 매핑됩니다.

Marker State

JMP 표식 맵에서 정수 인수에 해당하는 표식을 반환합니다. JMP 표식은 0부터 16까지의 번호로 되어 있습니다. 계산식 **Marker State(Row())**는 반복 시퀀스에서 현재 행 번호로 결정된 모든 행 상태 표식을 계산된 행 상태 열에 할당합니다. 행 상태 열에는 여러 개의 행 상태가 값으로 포함될 수 있습니다.

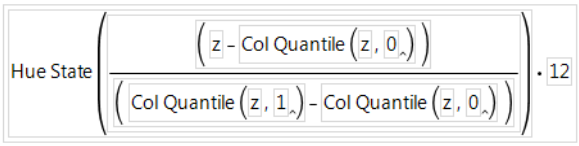
Selected State

인수가 0보다 클 경우 선택됨 행 상태 조건을 할당합니다. 인수가 0이면 해당 행의 열 값은 선택되지 않습니다.

Hue State

JMP 색조 맵에서 정수 인수에 해당하는 색상을 반환합니다. JMP 색조는 0부터 11까지의 번호로 되어 있지만 더 큰 정수는 modulo 12로 처리됩니다. **Hue State** 함수는 검정, 회색 또는 흰색에는 매핑되지 않습니다. 색조 0은 빨강에 매핑되고 색조 11은 자홍색에 매핑됩니다. 그림 A.16의 계산식은 *z*의 값에 따라 퍼져 나가는 유채색계로 행 상태 색상을 할당합니다. **Hue State** 함수는 행 상태 데이터 유형 열과 함께 사용됩니다.

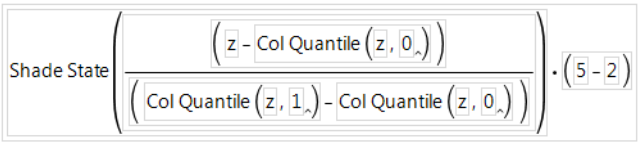
그림 A.16 Hue State 함수의 예



Shade State

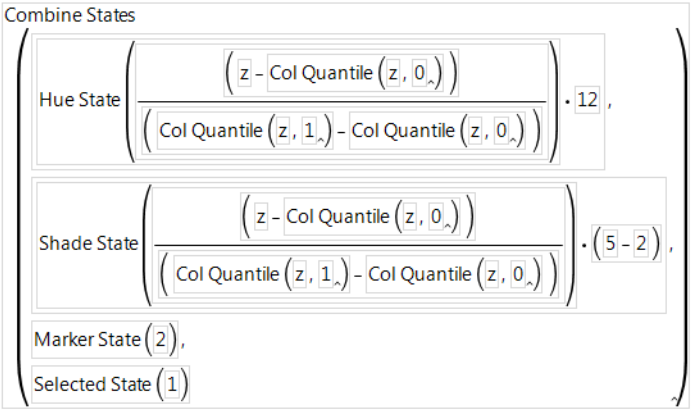
색상 또는 색조에 5 가지 음영 수준을 할당합니다 . –2 수준의 음영이 가장 어둡고 +2 수준의 음영이 가장 밝습니다 . 0 수준의 음영은 순수 색상입니다 . 그림 A.17 의 계산식은 *z* 의 값에 따라 음영 값을 할당합니다 .

그림 A.17 Shade State 함수의 예



색상 팔레트에 있는 모든 색상의 모든 음영을 할당하려면 **Hue State** 및 **Shade State** 할당을 함께 사용해야 합니다. 그림 A.18의 계산식은 이 섹션의 처음 부분에서 설명한 **Combine States** 함수를 사용합니다. **Combine States** 함수의 첫 번째 인수는 앞에 표시된 **Hue State** 계산식이고, 두 번째 인수는 **Shade State** 계산식입니다. 또한 인수가 2인 **Marker State** 함수는 각 행에 **X** 표식을 할당하고, 인수가 1인 **Selected State** 함수는 각 행을 선택합니다.

그림 A.18 Hue State 와 Row State 를 모두 사용하기 위한 Combine States 의 예



Excluded, Hidden, Labeled, Selected

결과가 1 또는 0(true 또는 false)이 되는 Row State 표현식 인수(행 상태 열 또는 행 상태 상수)가 사용됩니다. 이러한 특성은 기본적으로 비활성 상태입니다. 때로는 **Row()** 함수가 인수로 사용되며, 이 경우 가가 행의 활성 행 상태 조건이 감지됩니다. 예를 들어 그림 A.19의 계산식은 행이 활성화 상태로 선택되어 있을 때마다 99를 할당하고, 그렇지 않으면 0을 할당합니다. 이 계산식은 숫자 데이터 유형의 열에서 사용됩니다.

그림 A.19 Selected 함수를 사용한 계산식의 예

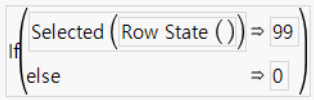
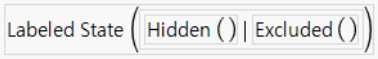
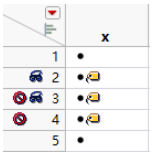


그림 A.20의 예에서는 행 상태 열에 행 상태 조건을 할당합니다. 행 상태 열(x라는 열)에 대한 계산식은 활성 행 상태가 **숨김** 또는 **제외됨**인지 확인하고, 그러한 경우에는 **라벨 지정됨** 행 상태를 할당합니다.

그림 A.20 행 상태 열에서 행 상태 정보 계산





Color Of

행 상태 표현식, 행 상태 열 또는 **Row State()** 함수가 인수로 사용됩니다. 이 함수는 JMP 색상 맵에서 활성 행 상태에 해당하는 번호를 반환하거나, 할당된 색이 없는 경우 0을 반환합니다.

Marker Of

행 상태 표현식, 행 상태 열 또는 **Row State()** 함수가 인수로 사용됩니다. 이 함수는 JMP 표식 맵에서 활성 표식에 해당하는 번호를 반환하거나, 할당된 표식이 없는 경우 0을 반환합니다.

할당 함수

할당 함수는 현재 위치에서 작동합니다. 즉, 연산에 의해 연산자의 오른쪽에 반환된 결과가 연산자의 왼쪽에 있는 인수에 저장되어 현재 값을 바꿉니다.

할당 문은 주로 다른 명령과 함께 JSL 스크립트를 생성하는 데 사용됩니다. 계산식 편집기를 사용하여 열 내에서 스크립트를 생성하고 실행할 수 있지만 종속성 및 모호성 문제가 발생할 수 있으므로 이 방법은 권장되지 않습니다. 대개 스크립트는 .jsl 파일로 저장되며, 데이터 테이블과 함께 저장될 수 있습니다. 자세한 내용은 "데이터 입력 및 편집" 장의 "데이터 테이블에서 스크립트 생성 및 저장"(262 페이지)에서 확인하십시오. 구문에 대한 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

참고 : 할당 함수의 첫 번째 인수는 할당 가능한 항목이어야 합니다 . 즉 , 3+=4 와 같은 할당은 3 이 재할당할 수 없는 상수 값이기 때문에 불가능합니다 . 먼저 , 값이 3 인 변수 (테이블 변수 또는 지역 변수) 를 생성해야 합니다 . 테이블 변수에 대한 자세한 내용은 " 데이터 입력 및 편집 " 장의 " 데이터 테이블 변수 사용 "(261 페이지) 에서 확인하십시오 . 로컬 변수에 대한 자세한 내용은 "JMP 의 계산식 생성 " 장의 " 계산식에서 데이터 테이블 값 참조 "(363 페이지) 에서 확인하십시오 . 그런 다음 이 변수를 할당 함수의 왼쪽 인수로 사용하십시오 .

=(assign) *b* 의 값을 *a* 에 넣습니다 . 예 : *a=b*

+=(add to) *b* 의 값을 *a* 에 더하고 결과를 다시 *a* 에 넣습니다 . 예 : *a+=b*

-=(subtract to) *b* 의 값을 빼고 결과를 다시 *a* 에 넣습니다 . 예 : *a-=b*

***=(multiply to)** *b* 를 *a* 로 곱하고 결과를 다시 *a* 에 넣습니다 . 예 : *a*=b*

/=(divide to) *b* 를 *a* 로 나누고 결과를 다시 *a* 에 넣습니다 . 예 : *a/=b*

++(post increment) *a* 에 1 을 바로 더해 *a++* 로 만듭니다 . 예를 들어 *a* 의 초기값이 4 이면 표현식 *a++* 는 a 를 5 로 변경합니다 .

--(post decrement) *a*에서 1을 바로 빼서 *a—*로 만듭니다. 예를 들어 *a*의 초기값이 4이면 표현식 *a--* 는 a 를 3 으로 변경합니다 .

모수 모형 함수

이 범주는 창 선택 열 집합에 대한 선형 함수인 세 가지 모수 모형을 생성하기 위한 바로 가기입니다.

Linear Model, Interactions Model, Full Quadratic Model

이러한 각 함수를 선택하면 모형에 포함할 열을 하나 이상 선택할 수 있는 열 선택 상자가 열립니다. 열을 선택하면 함수가 선택한 모형을 생성하고 채웁니다.

금융 함수

지급 원금, 이자율, 수익률 등을 계산하기 위한 계산식을 생성할 수 있습니다.

Double Declining Balance

지정된 기간 동안의 자산 감가상각을 반환합니다. 이 함수는 이중정률감가상각 방법 또는 기타 감가상각 요인을 사용합니다.

Future Value

기간, 고정 지급 및 고정 이자율을 기반으로 투자의 미래가치를 반환합니다.

Interest Payment

기간, 고정 지급 및 고정 이자율을 기반으로 지정된 기간 동안 투자의 이자 지급액을 반환합니다.

Interest Rate

대출 또는 투자의 기간별 이자율을 반환합니다.

Internal Rate of Return

values 인수에 지정된 일련의 현금 흐름에 대한 IRR(내부 수익률)을 반환합니다.

Modified Internal Rate of Return

일련의 주기적 현금 흐름에 대한 MIRR(수정된 내부 수익률)을 반환합니다. 투자 비용 및 현금 재투자 시 받은 이자가 포함됩니다.

Net Present Value

할인율, 일련의 미래 지급(음수 값) 및 수입(양수 값)을 사용하여 투자의 순현재가치를 반환합니다.

Number of Periods

기간, 고정 지급 및 고정 이자율을 기반으로 투자에 대한 기간 수를 반환합니다.

Payment

고정 지급 및 고정 이자율을 기반으로 대출에 대한 상환액을 반환합니다.

Present Value

투자의 순현재가치를 반환합니다.

Principal Payment

기간, 고정 지급 및 고정 이자율을 기반으로 지정된 기간의 투자에 대한 원금 지급액을 반환합니다.

Straight Line Depreciation

한 기간 동안의 자산의 정액감가상각을 반환합니다.

Sum Of Years Digits Depreciation

지정된 기간 동안의 자산의 연수합 감가상각을 반환합니다.

부록 B

JMP 기술 상세 정보

이 장에는 JMP의 여러 플랫폼과 SAS Integration에 적용되는 기술 상세 정보가 포함되어 있습니다.

목차

멀티스레딩을 지원하는 JMP 기능 729
JMP 속성을 SAS 확장 속성에 매핑하기 위한 규칙 730

멀티스레딩을 지원하는 JMP 기능

JMP의 일부 기능은 컴퓨터에 있는 여러 CPU(중앙 처리 장치)를 활용하도록 코딩되어 있으므로 매우 빠르게 실행될 수 있습니다. 이 프로세스를 멀티스레딩이라고 합니다.

멀티스레딩을 지원하는 기능은 다음과 같습니다.

- JMP PRO 부스티드 트리
- JMP PRO 붓스트랩 포레스트
- 선택
- JMP PRO 덮기 배열
- 분포에서의 공정 능력 분석
- 이상치 탐색
- 패턴 탐색
- 요인 분석
- 수명 분포 적합
- 모형 적합 : 모수 생존, JMP PRO 혼합 모형, JMP PRO 일반화 회귀, JMP PRO 반응 변수 선별
- 함수 데이터 탐색기
- 가우스 (JMP PRO 빠른 GASP 와 범주형 변수 사용은 JMP Pro 에서만 가능)
- JMP PRO K 최근접 이웃
- 잠재 계층 분석
- 수명 분포
- 다변량
- 신경망 (JMP PRO 일부 기능은 JMP Pro 에만 포함됨)
- 명목형 로지스틱
- 비선형 및 비선형 곡선
- 군집 분석의 정규 혼합
- 부분 최소 제곱 (JMP PRO 일부 기능은 JMP Pro 에만 포함됨)
- 파티션
- 예측 변수 선별
- 주성분
- 공정 변수 선별
- 프로파일러 (다른 플랫폼 내에서 시작된 프로파일러에는 적용되지 않음)
- 신뢰도 예측
- JMP PRO 수리 가능 시스템 시뮬레이션

- 반응 변수 선별
- JMP PRO 희소 주성분
- JMP PRO 구조 방정식 모형
- JMP PRO 서포트 벡터 머신
- 결측값 선별의 SVD 대치법
- 텍스트 탐색기 (JMP PRO SVD 및 잠재 계층 분석은 JMP Pro 에만 포함됨)
- 시계열 예측
- JMP PRO Uplift
- 변수 군집화

JMP 속성을 SAS 확장 속성에 매핑하기 위한 규칙

SAS 확장 속성은 테이블 스크립트, 라벨, 길이 및 유형 같은 정보를 가져오기 위해 SAS 코드에서 정의하는 메타데이터입니다. 확장 속성을 데이터 집합 또는 변수와 연결하고 `_JMP_TABLESCRIPTNAME_2="OnOpen"` 과 같이 이름-값 쌍으로 정의해야 합니다.

이 섹션에서는 SAS 확장 속성과 그에 해당하는 JMP 속성에 대한 정보를 제공합니다.

표 B.1 테이블 속성

JMP 테이블 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Table Variable Count	_JMP_TABLEVARCOUNT	숫자, 0	테이블 변수의 개수
Table Variable Name 1	_JMP_TABLEVARNAME_1	문자열	첫 번째 테이블 변수의 이름
Table Variable Value 1	_JMP_TABLEVARVALUE_1	문자열	첫 번째 테이블 변수의 값
Table Variable Name 2	_JMP_TABLEVARNAME_2	문자열	두 번째 테이블 변수의 이름
Table Variable Value 2	_JMP_TABLEVARVALUE_2	문자열	두 번째 테이블 변수의 값
Table Variable Name n	_JMP_TABLEVARNAME_*n*	문자열	*n*번째 테이블 변수의 이름
Table Variable Value n	_JMP_TABLEVARVALUE_*n*	문자열	*n*번째 테이블 변수의 값
Table Script Count	_JMP_TABLESCRIPTCOUNT	숫자	테이블 스크립트의 개수
Table Script Name 1	_JMP_TABLESCRIPTNAME_1	문자열	첫 번째 테이블 스크립트의 이름

표 B.1 테이블 속성 (계속)

JMP 테이블 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Table Script Value 1	_JMP_TABLESCRIPTVALUE_1	문자열	첫 번째 테이블 스크립트의 값
Table Script Name 2	_JMP_TABLESCRIPTNAME_2	문자열	두 번째 테이블 스크립트의 이름
Table Script Value 2	_JMP_TABLESCRIPTVALUE_2	문자열	두 번째 테이블 스크립트의 값
Table Script Name n	_JMP_TABLESCRIPTNAME_*n*	문자열	테이블 스크립트 *n* 의 이름
Table Script Value n	_JMP_TABLESCRIPTVALUE_*n*	문자열	테이블 스크립트 *n* 의 값
Lock Data Table	_JMP_ISLOCKED	부울, 0	잠긴 테이블
Suppress Formula Eval	_JMP_SUPPRESSEVAL	부울, 0	제한된 계산식 실행
Column Group Count	_JMP_COLGRPCOUNT	숫자, 0	열 그룹의 수
Column Group Name 1	_JMP_COLGRPNAME_1	문자열	열 그룹 1 의 이름
Column Group Start Col 1	_JMP_COLGRPSTARTCOL_1	문자열	그룹 1 에 포함된 첫 번째 열의 이름
Column Group NCols 1	_JMP_COLGRPNCOLS_1	숫자	그룹 1 의 열 수
Column Group Name 2	_JMP_COLGRPNAME_2	문자열	열 그룹 2 의 이름
Column Group Start Col 2	_JMP_COLGRPSTARTCOL_2	문자열	그룹 2 에 포함된 첫 번째 열의 이름
Column Group NCols 2	_JMP_COLGRPNCOLS_2	숫자	그룹 2 의 열 수
Column Group Name n	_JMP_COLGRPNAME_*n*	문자열	열 그룹 *n* 의 이름
Column Group Start Col n	_JMP_COLGRPSTARTCOL_*n*	문자열	그룹 *n* 에 포함된 첫 번째 열의 이름
Column Group NCols n	_JMP_COLGRPNCOLS_*n*	숫자	그룹 *n* 의 열 수

표 B.2 열 속성

JMP 열 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Column Name	_JMP_COLNAME	문자열	JMP의 원래 열 이름
Data Length	_JMP_DATALENGTH	숫자, 0	데이터의 길이. 숫자 유형의 경우 값은 1, 2, 4 또는 8일 수 있습니다. 문자 유형의 경우 0(또는 음수)은 가변 길이를 의미하고 양수 값은 고정 길이를 의미합니다.
Row State column	_JMP_ISROWSTATE	숫자, 0	0 = 행 상태 아님, 1 = 테이블의 기본 행 상태, 2 = 행 상태 열
Label	_JMP_ISLABEL	부울, 0	라벨 열. 1 = 예. 0 = 아니요.
Scroll Lock	_JMP_ISSCROLLLOCK	부울, 0	스크롤 잠금 열. 1 = 예. 0 = 아니요.
Hidden	_JMP_ISHIDDEN	부울, 0	숨겨진 열. 1 = 예. 0 = 아니요
Excluded	_JMP_ISEXCLUDED	부울, 0	제외된 열. 1 = 예. 0 = 아니요
Notes	_SAS_NOTES	문자열	열에 대한 노트
Modeling Type	_SAS_LEVEL	문자열	유효한 값은 Binary, Interval, Ordinal, Nominal, Unary입니다. JMP는 Binary와 Unary를 Nominal에 매핑하고 Interval을 Continuous에 매핑합니다.

표 B.2 열 속성 (계속)

JMP 열 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Preselect Role	_SAS_MININGROLE	문자열	유효한 값은 Assessment, Censor, Classification, Cost, Cross ID, Decision, Frequency, ID, Input, Key, Label, Prediction, Referrer, Rejected, Residual, Segment, Sequence, Target, Text, Text Location, Time ID, Treatment, Web Address, Weight입니다. JMP는 Y를 Target에, X를 Input에, Freq를 Frequency에, Weight를 Weight에 매핑합니다.
Distribution	_SAS_DISTRIBUTION	문자열	이 열에 대해 적합시킬 분포 유형
Format Name	_JMP_FORMATNAME	문자열	열의 형식(예: 고정 소수점 또는 백분율)
Format Width	_JMP_FORMATWIDTH	숫자	열의 너비
Format Decimals	_JMP_FORMATDECIMALS	숫자	소수 자릿수
Format Code	_JMP_FORMATCODE	문자열	일반적으로 통화 형식의 통화 코드
Use Thousands Separator	_JMP_USETHOUSANDSSEP	부울, 0	천 단위 구분 기호 포함. 1 = 예. 0 = 아니요.
Input Format	_JMP_INFORMAT	문자열	입력 형식(날짜, 시간 또는 기간 형식 열의 경우)
Formula	_JMP_FORMULA	문자열	열 계산식
Suppress Formula Eval	_JMP_SUPPRESSEVAL	부울, 0	제한된 계산식 실행. 1 = 예. 0 = 아니요.
Lock	_JMP_ISLOCKED	부울, 0	잠긴 열. 1 = 예. 0 = 아니요.

표 B.2 열 속성 (계속)

JMP 열 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Range Check Code	_JMP_RANGECHECK_CODE	문자열	범위 확인 유형을 나타내는 코드. 예: LE, LT, LTLT, !LT, !LELE
Range Check Low	_JMP_RANGECHECK_LOW	숫자	범위 확인의 첫 번째 인수
Range Check High	_JMP_RANGECHECK_HIGH	숫자	범위 확인의 두 번째 선택적 인수
List Check	_JMP_LISTCHECK	문자열	목록 확인
Missing Value Codes	_JMP_MISSINGCODES	문자열	결측값으로 처리할 값
Value Label Count	_JMP_VALUELABEL_COUNT	숫자	값 라벨의 개수
Value Label Code	_JMP_VALUELABELCODE_1	문자열	값 라벨의 코드(예: EQ, LT, LE, LTLT, LELT). 문자 변수의 경우 항상 EQ입니다.
Value Label Lower Bound	_JMP_VALUELABELLOW_1	가변	숫자 범위의 하한. 범위의 한쪽이 제한되지 않은 경우에는 한계가 하나만 있습니다 문자 열의 경우 문자열이고, 숫자 열의 경우 숫자입니다.
Value Label Upper Bound	_JMP_VALUELABELHIGH_1	숫자	숫자 범위의 상한. 문자 값 라벨에는 사용되지 않습니다.
Value Label Display	_JMP_VALUELABELDISPLAY_1	문자열	이 값 라벨의 표시 문자열
Use Value Labels	_JMP_USEVALUELABELS	부울, 0	이 열을 표시할 때 값 라벨을 사용합니다. 1 = 예. 0 = 아니요.
값 점수	_JMP_VALUESCORES	문자열	데이터 값을 숫자 점수와 연결합니다.
값 순서	_JMP_VALUEORDER	문자열	열의 값 순서
Value Colors	_JMP_VALUECOLORS	문자열	값을 색상에 매핑합니다.
Color Gradient	_JMP_COLORGRADIENT	문자열	값을 색상 그래디언트에 매핑합니다.

표 B.2 열 속성 (계속)

JMP 열 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Color Cells	_JMP_COLORCELLS	부울, 0	지정된 값 색상 또는 색상 그래디언트를 사용하여 이 열의 셀에 색상을 적용합니다. 1 = 예. 0 = 아니요.
Axis	_JMP_AXIS	문자열	축 정의
Coding	_JMP_CODING	문자열	모델링을 위해 값 범위를 -1, 1로 코딩합니다.
Mixture	_JMP_MIXTURE	문자열	열 혼합 관여도로, 합은 특정 값이 됩니다. 혼합 열 특성은 숫자 값(한계)과 문자열 값(L 유사성분 및 U 유사성분)을 갖습니다.
Row Order Levels	_JMP_ROWORDER	부울, 0	값이 아니라 발생 순서를 기준으로 열을 정렬합니다. 1 = 예. 0 = 아니요.
Spec Limits	_JMP_SPECLIMITS	문자열	목표값 및 규격 상한/하한
Control Limits	_JMP_CONTROLLIMITS	문자열	관리도 유형의 관리 한계
Response Limits	_JMP_RESPONSELIMITS	문자열	반응의 허용 가능한 범위에 대한 한계(프로파일러에서 사용됨)
Design Role	_JMP_DESIGNROLE	문자열	설계된 실험에서의 역할
Factor Changes	_JMP_FACTORCHANGES	문자열	요인의 변경 난이도(DOE의 경우). 유효한 값: Easy, Hard, Very Hard
Sigma	_JMP_SIGMA	문자열	관리도의 시그마
Units	_JMP_UNITS	문자열	이 변수의 단위(문자열 가능)
Time Frequency	_JMP_TIMEFREQ	문자열	Numeric, Annual, Quarterly, Monthly, Weekly, Daily, Hourly, By Minutes, By Seconds
Map Role	_JMP_MAPROLE	문자열	맵 데이터 집합에서의 역할

표 B.2 열 속성 (계속)

JMP 열 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Multiple Response Separator	_JMP_MRSEP	문자열, ","	다중 반응 열의 구분 기호
Label Column	_JMP_LABELCOLUMN	부울, 0	이 열이 라벨 열인지 여부. 1 = 예. 0 = 아니요.
Expression Column	_JMP_ISEXPRESSIONCOLUMN	부울, 0	이 열이 표현식 열인지 여부 1 = 예. 0 = 아니요.

표 B.3 사용자 열 특성

JMP 열 속성	SAS 속성 이름	SAS 속성 유형 및 기본값	설명
Custom Property Count	_JMP_CUSTCOLPROPCOUNT	숫자, 0	사용자 열 특성의 개수
Custom Property Name 1	_JMP_CUSTCOLPROPNAME_1	문자열	사용자 열 특성 1의 이름
Custom Property Value 1	_JMP_CUSTCOLPROPVALUE_1	문자열	사용자 열 특성 1의 값
Custom Property Name 2	_JMP_CUSTCOLPROPNAME_2	문자열	사용자 열 특성 2의 이름
Custom Property Value 2	_JMP_CUSTCOLPROPVALUE_2	문자열	사용자 열 특성 2의 값
Custom Property Name n	_JMP_CUSTCOLPROPNAME_*n*	문자열	사용자 열 특성 *n*의 이름
Custom Property Value n	_JMP_CUSTCOLPROPNAME_*n*	문자열	사용자 열 특성 *n*의 값

부록 C

기술 라이선스 표시

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

버전 16

기본 분석

" 진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는
것이 아니라 새로운 눈을 가지는 데 있다 ."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.0

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. 기본 분석® 16. Cary, NC: SAS Institute Inc.

기본 분석® 16

Copyright © 2020-2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP를 처음 사용하든 오랫동안 사용해왔든, JMP와 관련해서 배워야 하는 것은 언제나 있습니다.

JMP.com을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다.

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티, 추가 기능 및 스크립트 예를 포함한 사용자용 리소스, 포럼, 블로그, 컨퍼런스 정보 등

https://www.jmp.com/getstarted

목차
기본 분석

1 JMP 알아보기 15
설명서 및 추가 리소스
JMP 설명서에 사용되는 서식 규칙 17
JMP 도움말 18
JMP 설명서 라이브러리 18
JMP 학습을 위한 추가 리소스 24
JMP 자습서 24
샘플 데이터 테이블 24
통계 및 JSL 용어에 대해 알아보기 25
JMP 팁 및 힌트 알아보기 25
JMP 툴팁 25
JMP 사용자 커뮤니티 25
무료 온라인 Statistical Thinking 교육 과정 26
JMP 새로운 사용자를 위한 입문 키트 26
Statistics Knowledge 포털 26
JMP 교육 과정 26
사용자가 작성한 JMP 설명서 26
JMP 시작하기 창 26
JMP 기술 지원 27

2 기본 분석 소개 29
기본 분석 방법 개요

3 분포 31
분포 플랫폼 사용
분포 플랫폼의 개요 34
분포 플랫폼의 예 34
분포 플랫폼 시작 36
분포 보고서 37
히스토그램 39
빈도 보고서 41
분위수 보고서 42

요약 통계량 보고서 42
분포 플랫폼 옵션 44
범주형 변수에 대한 옵션 45
범주형 변수에 대한 표시 옵션 45
범주형 변수에 대한 히스토그램 옵션 46
범주형 변수에 대한 저장 옵션 46
연속형 변수에 대한 옵션 46
연속형 변수에 대한 표시 옵션 47
연속형 변수에 대한 히스토그램 옵션 48
정규 분위수 그림 49
이상치 상자 그림 50
분위수 상자 그림 51
줄기 - 잎 53
CDF 그림 53
평균 검정 54
표준편차 검정 55
동등성 검정 55
신뢰 구간 56
예측 구간 57
공차 구간 57
공정 능력 57
분포 적합 59
연속형 변수에 대한 저장 옵션 64
분포 플랫폼의 추가 예 65
예 : 여러 히스토그램에서 데이터 선택 65
예 : 기준 변수 사용 66
검정 확률 옵션의 예 67
예측 구간의 예 69
공차 구간의 예 71
공정 능력의 예 72
분포 플랫폼에 대한 통계 상세 정보 73
표준 오차 막대 74
분위수에 대한 통계 상세 정보 74
요약 통계량 75
정규 분위수 그림 76
Wilcoxon 부호 순위 검정 76
표준편차 검정 78
정규 분위수 78
표준화된 데이터 저장 79
예측 구간 79

공차 구간 ... 80
연속형 적합 분포 ... 82
이산형 적합 분포 ... 88
레거시 분포 적합기에 대한 상세 정보 ... 91
분포 적합 옵션 (레거시) ... 91
연속형 적합 분포에 대한 통계 상세 정보 (레거시) ... 95
이산형 적합 분포에 대한 통계 상세 정보 (레거시) ... 100
적합 분위수에 대한 통계 상세 정보 (레거시) ... 101
분포 적합 옵션에 대한 통계 상세 정보 (레거시) ... 102

4 X 로 Y 적합 소개 ... 105
두 변수 간의 관계 분석
X 로 Y 적합 플랫폼의 개요 ... 107
X 로 Y 적합 플랫폼 시작 ... 107
JMP 시작하기 창에서 특정 분석 시작 ... 108

5 이변량 분석 ... 109
두 연속형 변수 간의 관계 분석
이변량 분석의 예 ... 112
이변량 플랫폼 시작 ... 112
이변량 그림 ... 113
적합 옵션 ... 114
적합 옵션 ... 115
적합 옵션 범주 ... 117
동일한 적합 옵션을 여러 번 적용 ... 117
히스토그램 테두리 ... 118
평균 적합 ... 119
평균 적합 보고서 ... 119
선형 적합 및 다항식 적합 ... 120
선형 적합 및 다항식 적합 보고서 ... 120
특수 적합 ... 125
특수 적합 보고서 및 메뉴 ... 126
유연 ... 127
스플라인 적합 ... 127
커널 평활기 ... 128
각 값 적합 ... 129
직교 적합 ... 129
직교 적합 비율 보고서 ... 130
로버스트 ... 130
로버스트 적합 ... 131

Cauchy 적합 131
밀도 타원 131
상관 보고서 132
비모수 밀도 133
분위수 밀도 등고선 보고서 133
그룹화 기준 133
적합 메뉴 134
적합 메뉴 옵션 135
진단 그림 137
이변량 플랫폼의 추가 예 138
특수 적합 옵션의 예 138
직교 적합 옵션의 예 140
로버스트 적합 명령의 예 142
밀도 타원을 사용한 그룹화의 예 144
회귀선을 사용한 그룹화의 예 145
기준 변수를 사용한 그룹화의 예 146
이변량 플랫폼에 대한 통계 상세 정보 147
선형 적합 148
스플라인 적합 148
직교 적합 148
적합 요약 보고서 149
적합 결여 보고서 149
모수 추정값 보고서 150
평활 적합 보고서 150
상관 보고서 150

6 일원 분석 153
연속형 Y 와 범주형 X 변수 간의 관계 분석

일원 분석 개요 156
일원 분석의 예 156
일원 분석 플랫폼 시작 158
데이터 형식 159
일원 분석 그림 159
일원 분석 플랫폼 옵션 160
표시 옵션 163
분위수 164
이상치 상자 그림 165
평균 /ANOVA 및 평균 /ANOVA/ 합동 t 165
적합 요약 보고서 166
t- 검정 보고서 166

분산 분석 보고서 167
일원 ANOVA 에 대한 평균 보고서 168
블록 평균 보고서 168
평균 다이아몬드 및 X 축 비례 168
평균 선 , 오차 막대 및 표준편차 선 169
평균 분석 방법 170
위치에 대한 평균 분석 170
척도에 대한 평균 분석 171
평균 분석 차트 172
평균 분석 옵션 173
평균 비교 173
비교 원 사용 175
개별 쌍 비교 (스튜던트 t) 177
전체 쌍 비교 (Tukey HSD) 177
최량 비교 (Hsu MCB) 177
대조군 비교 (Dunnett) 178
단계별 개별 쌍 비교 (Newman-Keuls) 178
평균 비교 옵션 179
비모수 검정 180
Wilcoxon, 중앙값 , Van der Waerden 및 Friedman 순위 검정 보고서 181
Kolmogorov-Smirnov 2 표본 검정 보고서 182
비모수 다중 비교 184
이분산 186
분산 동일 여부 검정 보고서 187
동등성 검정 188
동등성 검정 보고서 188
로버스트 190
로버스트 적합 190
Cauchy 적합 190
검정력 191
검정력 상세 정보 창 및 보고서 191
정규 분위수 그림 192
CDF 그림 192
밀도 193
매칭 열 193
일원 분석 플랫폼의 추가 예 194
평균 분석 차트의 예 194
분산에 대한 평균 분석 차트의 예 195
개별 쌍 비교 (스튜던트 t) 검정의 예 196
전체 쌍 비교 (Tukey HSD) 검정의 예 198

최량 비교 (Hsu MCB) 검정의 예 ... 200
대조군 비교 (Dunnett) 의 예 ... 201
단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예 ... 202
예 : 4 가지 평균 비교 검정 대조 ... 203
비모수 Wilcoxon 검정의 예 ... 204
이분산 옵션의 예 ... 207
동등성 검정의 예 ... 208
로버스트 적합 옵션의 예 ... 209
검정력 옵션의 예 ... 211
정규 분위수 그림의 예 ... 213
CDF 그림의 예 ... 214
밀도 옵션의 예 ... 215
매칭 열 옵션의 예 ... 216
예 : 일원 분석을 위한 데이터 쌓기 ... 217
일원 분석 플랫폼에 대한 통계 상세 정보 ... 224
비교 원 ... 224
검정력 ... 225
적합 요약 보고서 ... 226
분산 동일 여부 검정 ... 227
비모수 검정 통계량 ... 228

7 분할 분석 ... 233

두 범주형 변수 간의 관계 분석

분할 분석의 예 ... 235
분할 분석 플랫폼 시작 ... 236
데이터 형식 ... 236
분할 분석 보고서 ... 237
분할 분석 플랫폼 옵션 ... 238
모자이크 그림 ... 239
팝업 메뉴 ... 240
분할표 ... 241
분할표 설명 ... 243
검정 ... 243
검정 보고서 설명 ... 244
Fisher 정확 검정 ... 244
비율에 대한 평균 분석 ... 244
대응 분석 ... 245
대응 분석 그림 이해 ... 245
대응 분석 옵션 ... 245
상세 정보 보고서 ... 246

Cochran-Mantel-Haenszel 검정 246
합치도 통계량 247
상대 위험도 247
비율에 대한 2 표본 검정 248
연관성 측도 248
Cochran Armitage 추세 검정 249
정확 검정 250
분할 분석 플랫폼의 추가 예 250
비율에 대한 평균 분석의 예 251
대응 분석의 예 252
Cochran Mantel Haenszel 검정의 예 254
합치도 통계량 옵션의 예 256
상대 위험도 옵션의 예 257
비율에 대한 2 표본 검정의 예 258
연관성 측도 옵션의 예 259
Cochran Armitage 추세 검정의 예 260
분할 분석 플랫폼에 대한 통계 상세 정보 261
합치도 통계량 261
승산비 옵션 262
검정 보고서 262
대응 분석의 상세 정보 보고서 263

8 로지스틱 분석 265
범주형 Y 변수와 연속형 X 변수 간의 관계 분석

로지스틱 회귀의 개요 267
명목형 로지스틱 회귀 267
순서형 로지스틱 회귀 267
명목형 로지스틱 회귀의 예 268
로지스틱 플랫폼 시작 269
데이터 구조 270
로지스틱 보고서 271
로지스틱 그림 272
반복 272
전체 모형 검정 272
적합 상세 정보 273
모수 추정값 274
로지스틱 플랫폼 옵션 275
ROC 곡선 276
확률 계산식 저장 277
역추정 예측 277

로지스틱 회귀의 추가 예 277
순서형 로지스틱 회귀의 예 277
로지스틱 그림의 추가 예 279
ROC 곡선의 예 281
십자기호 도구를 사용한 역추정 예측의 예 282
역추정 예측 옵션을 사용한 역추정 예측의 예 283
로지스틱 플랫폼에 대한 통계 상세 정보 284

9 테이블 생성 287
대화식으로 요약 테이블 생성

테이블 생성 플랫폼의 예 289
테이블 생성 플랫폼 시작 293
대화상자 사용 294
통계량 추가 295
테이블 생성 출력 298
분석 열 299
그룹화 열 300
열 및 행 테이블 300
테이블 편집 301
테이블 생성 플랫폼 옵션 302
테스트 빌드 패널 표시 303
마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴 303
테이블 생성 플랫폼의 추가 예 304
예 : 여러 테이블 생성 및 내용 재배열 304
예 : 여러 열을 단일 테이블에 결합 308
예 : 페이지 열 사용 310

10 시뮬레이션 313
모수 재표집을 통해 난제 해결 방안 모색

시뮬레이션 플랫폼 개요 315
시뮬레이션 사용 예 315
분산 성분에 대한 준모수 신뢰 구간 구성 316
순열 검정 수행 321
일반화 회귀의 요인 유지 탐색 323
비선형 모형에 대한 전향적 검정력 분석 328
시뮬레이션 창 시작 337
시뮬레이션 창 337
시뮬레이션 결과 테이블 338
시뮬레이션 결과 보고서 338
시뮬레이션 검정력 보고서 339

11 붓스트랩 341
재표집을 통한 통계량 분포 근사

붓스트랩 개요 343
- 붓스트랩 지원 플랫폼 344

붓스트랩의 예 345
붓스트랩 창 옵션 347
누적 붓스트랩 결과 테이블 348
비누적 붓스트랩 결과 테이블 350
붓스트랩 결과 분석 350
붓스트랩의 추가 예 352
붓스트랩에 대한 통계 상세 정보 357
- 부분 가중치 계산 357
- 편향 수정 백분위수 구간 357

12 텍스트 탐색기 359
데이터의 비정형 텍스트 탐구

텍스트 탐색기 플랫폼 개요 361
- 텍스트 처리 단계 362

텍스트 탐색기 플랫폼의 예 363
텍스트 탐색기 플랫폼 시작 366
- 정규 표현식 편집기에서 정규 표현식 사용자 정의 368

텍스트 탐색기 보고서 374
- 요약 개수 보고서 374
- 용어 및 구 목록 375

텍스트 탐색기 플랫폼 옵션 378
- 텍스트 준비 옵션 378
- 텍스트 분석 옵션 384
- 저장 옵션 386
- 보고서 옵션 387

잠재 계층 분석 387
잠재 의미 분석 (SVD) 389
- SVD 보고서 390
- SVD 보고서 옵션 391

주제 분석 392
- 주제 분석 보고서 393
- 주제 분석 보고서 옵션 393

판별 분석 394
- 판별 분석 보고서 394
- 판별 분석 보고서 옵션 395

용어 선택 396

용어 선택 설정 396
용어 선택 보고서 397
용어 선택 보고서 옵션 399
감정 분석 399
감정 분석 보고서 401
감정 분석 보고서 옵션 403
텍스트 탐색기 플랫폼의 추가 예 404

A 참조 자료 409

B 기술 라이선스 표시 413

1 장

JMP 알아보기
설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 17
JMP 도움말 18
JMP 설명서 라이브러리 18
JMP 학습을 위한 추가 리소스 24
JMP 자습서 24
샘플 데이터 테이블 24
통계 및 JSL 용어에 대해 알아보기 25
JMP 팁 및 힌트 알아보기 25
JMP 툴팁 25
JMP 사용자 커뮤니티 25
무료 온라인 Statistical Thinking 교육 과정 26
JMP 새로운 사용자를 위한 입문 키트 26
Statistics Knowledge 포털 26
JMP 교육 과정 26
사용자가 작성한 JMP 설명서 26
JMP 시작하기 창 26
JMP 기술 지원 27

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 *Lucida Sans Typewriter 기울임꼴* (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 "JMP 설명서 라이브러리 "(18 페이지) 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
Quality and Process Methods	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기" 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기

" 창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

2 장

기본 분석 소개

기본 분석 방법 개요

기본 분석에서는 JMP에서 종종 수행하는 다음과 같은 초기 분석 유형에 대해 설명합니다.

- 분포 플랫폼에서는 히스토그램, 추가 그래프 및 보고서를 사용하여 단일 변수의 분포를 보여줍니다. 데이터가 어떻게 분포되어 있는지 알게 되면 적절한 유형의 분석을 계획할 수 있습니다. 자세한 내용은 3 장, " 분포 " 에서 확인하십시오.
- X 로 Y 적합 플랫폼에서는 지정된 X 및 Y 변수 쌍을 모델링 유형에 따라 컨텍스트별로 분석합니다. 자세한 내용은 4 장, "X 로 Y 적합 소개 " 에서 확인하십시오. 네 가지 분석 유형으로는 다음이 포함됩니다.
  - 이변량 플랫폼 : 두 연속형 X 변수 간의 관계를 분석합니다. 자세한 내용은 5 장, " 이변량 분석 " 에서 확인하십시오.
  - 일원 분석 플랫폼 : 범주형 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지를 분석합니다. 자세한 내용은 6 장, " 일원 분석 " 에서 확인하십시오.
  - 분할 분석 플랫폼 : 범주형 X 요인의 값에 따라 좌우되는 범주형 반응 변수의 분포를 분석합니다. 자세한 내용은 7 장, " 분할 분석 " 에서 확인하십시오.
  - 로지스틱 플랫폼 : 반응 범주 (Y) 에 대한 확률을 연속형 X 예측 변수에 적합시킵니다. 자세한 내용은 8 장, " 로지스틱 분석 " 에서 확인하십시오.
- 테이블 생성 플랫폼 : 기술 통계량 테이블을 대화식으로 생성합니다. 자세한 내용은 9 장, " 테이블 생성 " 에서 확인하십시오.
- 시뮬레이션 기능으로는 모수 및 비모수 시뮬레이션을 수행할 수 있습니다. 자세한 내용은 10 장, " 시뮬레이션 " 에서 확인하십시오.
- 붓스트랩 분석은 통계량의 표본 분포에 근사한 값을 산출합니다. 복원 표집 방식으로 데이터가 재표집되어 통계량이 계산됩니다. 통계량의 값 분포가 생성될 때까지 이 과정이 반복됩니다. 자세한 내용은 11 장, " 붓스트랩 " 에서 확인하십시오.
- 텍스트 탐색기 플랫폼에서는 형식이 지정되지 않은 텍스트 데이터를 범주화하고 분석할 수 있습니다. 분석을 진행하기 전에 정규 표현식을 사용하여 데이터를 정리할 수 있습니다. 자세한 내용은 12 장, " 텍스트 탐색기 " 에서 확인하십시오.

3 장

분포

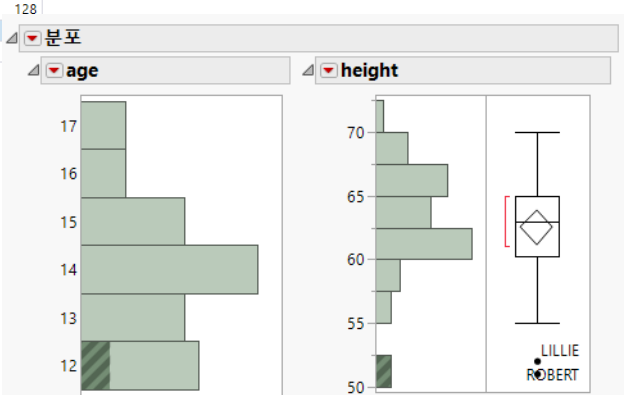
분포 플랫폼 사용

히스토그램, 상자 그림 및 요약 통계량을 사용하여 단일 변수의 분포를 탐색하려면 분포 플랫폼을 사용합니다. *t*-검정, *z*-검정, 카이제곱 검정, 동등성 검정 등을 비롯한 여러 가지 유형의 가설 검정을 수행할 수 있습니다. 신뢰도, 공차 및 예측 구간을 생성하고 공정 능력을 추정할 수도 있습니다. **단변량**이란 관련 변수가 두 개(이변량) 또는 여러 개(다변량)가 아니라 하나뿐임을 의미합니다. 그러나 하나의 보고서에서 여러 개별 변수의 분포를 검토할 수도 있습니다. 각 변수의 보고서 내용은 변수가 범주형(명목형 또는 순서형)인지 연속형인지에 따라 달라집니다.

분포 보고서 창은 대화식입니다. 히스토그램 막대를 클릭하면 다른 모든 히스토그램과 데이터 테이블에서 해당 데이터가 강조 표시됩니다.

그림 3.1 분포 플랫폼의 예

	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	
9	BARBARA	13	F	60	



목차

분포 플랫폼의 개요 34
분포 플랫폼의 예 34
분포 플랫폼 시작 36
분포 보고서 37
히스토그램 39
빈도 보고서 41
분위수 보고서 42
요약 통계량 보고서 42
분포 플랫폼 옵션 44
범주형 변수에 대한 옵션 45
범주형 변수에 대한 표시 옵션 45
범주형 변수에 대한 히스토그램 옵션 46
범주형 변수에 대한 저장 옵션 46
연속형 변수에 대한 옵션 46
연속형 변수에 대한 표시 옵션 47
연속형 변수에 대한 히스토그램 옵션 48
정규 분위수 그림 49
이상치 상자 그림 50
분위수 상자 그림 51
줄기-잎 53
CDF 그림 53
평균 검정 54
표준편차 검정 55
동등성 검정 55
신뢰 구간 56
예측 구간 57
공차 구간 57
공정 능력 57
분포 적합 59
연속형 변수에 대한 저장 옵션 64
분포 플랫폼의 추가 예 65
예: 여러 히스토그램에서 데이터 선택 65
예: 기준 변수 사용 66
검정 확률 옵션의 예 67
예측 구간의 예 69
공차 구간의 예 71
공정 능력의 예 72
분포 플랫폼에 대한 통계 상세 정보 73

표준 오차 막대 74
분위수에 대한 통계 상세 정보 74
요약 통계량 75
정규 분위수 그림 76
Wilcoxon 부호 순위 검정 76
표준편차 검정 78
정규 분위수 78
표준화된 데이터 저장 79
예측 구간 79
공차 구간 80
연속형 적합 분포 82
이산형 적합 분포 88
레거시 분포 적합기에 대한 상세 정보 91
분포 적합 옵션(레거시) 91
연속형 적합 분포에 대한 통계 상세 정보(레거시) 95
이산형 적합 분포에 대한 통계 상세 정보(레거시) 100
적합 분위수에 대한 통계 상세 정보(레거시) 101
분포 적합 옵션에 대한 통계 상세 정보(레거시) 102

분포 플랫폼의 개요

분포 플랫폼에서는 변수의 모델링 유형, 즉 변수가 범주형(명목형 또는 순서형)인지 연속형인지에 따라 변수 처리가 달라집니다.

범주형 변수

범주형 변수의 경우 초기에 표시되는 그래프는 히스토그램입니다. 히스토그램에는 순서형 또는 명목형 변수의 각 수준에 대한 막대가 표시됩니다. 나뉜(모자이크) 막대 차트를 추가할 수도 있습니다.

빈도 보고서에는 개수와 비율이 표시됩니다. 빨간색 삼각형 메뉴의 옵션을 사용하여 신뢰 구간을 추가하고 확률을 검정할 수 있습니다.

연속형 변수

연속형 숫자 변수의 경우 초기 그래프에는 히스토그램과 이상치 상자 그림이 표시됩니다 . 히스토그램에는 연속형 변수의 그룹화된 값에 대한 막대가 표시됩니다 . 다음 옵션도 사용할 수 있습니다 .

- 정규 분위수 그림
- 분위수 상자 그림
- 줄기 - 잎 그림
- CDF 그림

보고서에는 선택한 사분위수와 요약 통계량이 표시됩니다 . 빨간색 삼각형 메뉴에서는 다음 작업을 위한 추가 보고서 옵션을 사용할 수 있습니다 .

- 순위 , 확률 점수 , 정규 분위수 값 등을 데이터 테이블에 새 열로 저장 (" 저장 " 옵션)
- 지정한 상수를 기준으로 열의 평균 및 표준편차 검정 (" 평균 검정 " 및 " 표준편차 검정 " 옵션)
- 다양한 분포 및 비모수 평활 곡선 적합 (" 연속형 적합 " 및 " 이산형 적합 " 옵션)
- 품질 관리 응용 프로그램에 대한 공정 능력 분석 수행
- 신뢰 구간 , 예측 구간 및 공차 구간

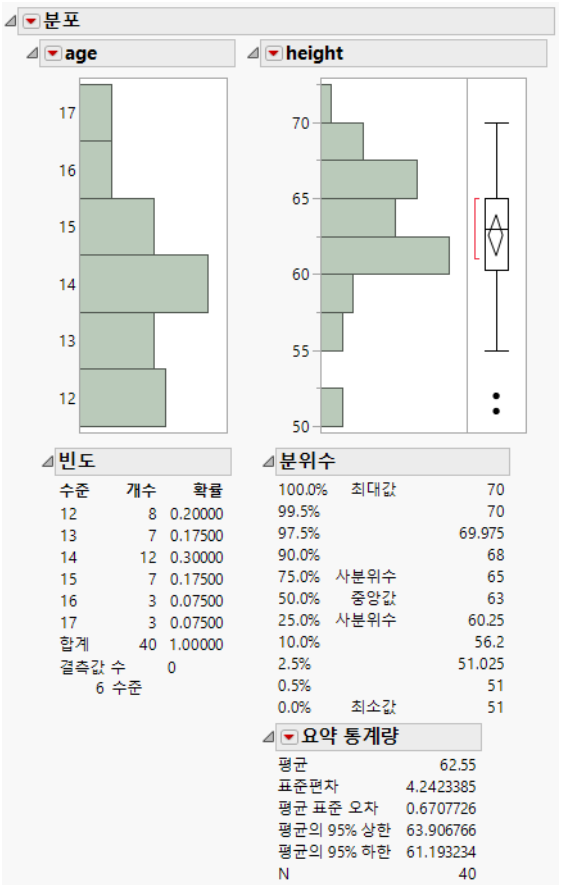
분포 플랫폼의 예

JMP 에서 분포를 생성하는 방법을 알아봅니다 . 40 명의 학생에 대한 데이터가 있고 이 학생들의 연령 및 키 분포를 보려고 한다고 가정해 보겠습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. age 및 height 를 선택하고 **Y, 열**을 클릭합니다 .

4. **확인**을 클릭합니다 .

그림 3.2 분포 플랫폼의 예



히스토그램에서 다음을 확인할 수 있습니다 .

- 연령은 균일하게 분포되어 있지 않습니다 .
- 키의 경우 극단값 (이상치일 수 있음) 을 갖는 두 개의 점이 있습니다 .

"height" 히스토그램에서 50 에 해당하는 막대를 클릭하여 잠재적 이상치를 보다 자세히 확인합니다 .

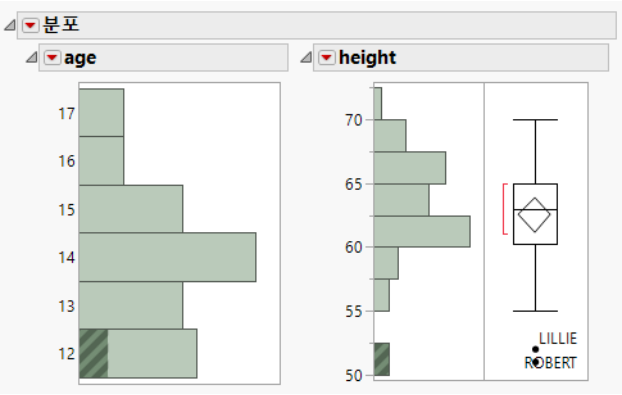
- "age" 히스토그램에서 해당 연령이 강조 표시됩니다 . 잠재적 이상치의 연령은 12 세입니다 .
- 데이터 테이블에서 해당 행이 강조 표시됩니다 . 잠재적 이상치의 이름은 Lillie 와 Robert 입니다 .

"height" 히스토그램에서 잠재적 이상치에 라벨을 추가합니다 .

1. 두 개의 이상치를 모두 선택합니다 .

2. 이상치 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **행 라벨**을 선택합니다 .
   데이터 테이블에서 해당 행에 라벨 아이콘이 추가됩니다 .
3. 전체 라벨이 표시되도록 상자 그림을 더 넓게 조정합니다 .

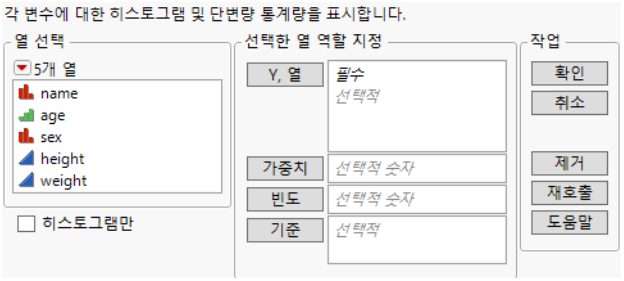
그림 3.3 라벨이 지정된 잠재적 이상치



분포 플랫폼 시작

분석 > 분포를 선택하여 분포 플랫폼을 시작할 수 있습니다 .

그림 3.4 분포 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 열 분석하려는 변수를 할당합니다 . 각 변수에 대해 히스토그램과 관련 보고서가 표시됩니다 .

가중치 연속형 Y 의 관측값에 대한 가중치를 지정하는 변수를 할당합니다 . 범주형 Y 의 경우에는 " 가중치 " 열이 무시됩니다 . 가중치 합계를 기준으로 하는 모든 통계량은 가중치의 영향을 받습니다 .

빈도 이 역할에 빈도 변수를 할당합니다 . 이 기능은 데이터를 요약할 경우에 유용합니다 . 이 경우 Y 값에 대한 열 하나와 Y 값의 발생 빈도에 대한 또 다른 열이 있게 됩니다 . 이 변수의 합

계는 " 요약 통계량 " 보고서에 표시되는 전체 개수에 포함됩니다 (N 으로 표시). 다른 모든 적률 통계량 (평균 , 표준편차 등) 도 **빈도** 변수의 영향을 받습니다 .

기준 **기준** 변수의 각 수준에 대해 개별 보고서를 생성합니다 . **기준** 변수가 둘 이상 할당되면 **기준** 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

공정 능력 생성 (열에 " 규격 한계 " 열 특성이 포함된 경우에만 나타남) " 규격 한계 " 열 특성이 포함된 분석 열에 대한 공정 능력 보고서를 추가합니다 .

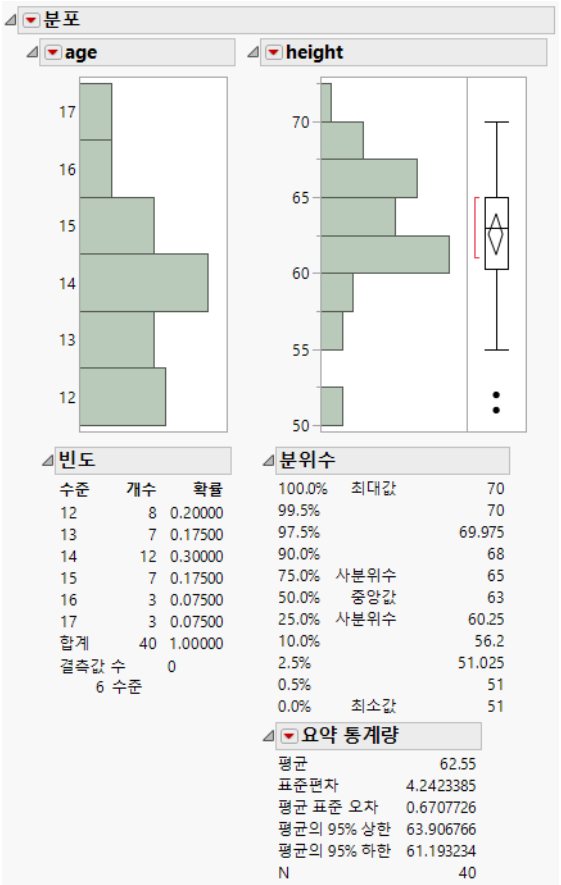
히스토그램만 보고서 창에서 히스토그램을 제외한 모든 내용을 제거합니다 .

시작 창에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

분포 보고서

초기 " 분포 " 보고서에는 각 변수에 대한 히스토그램과 보고서가 포함됩니다 .

그림 3.5 초기 분포 보고서 창



참고 : 그림 3.5 에 표시된 것과 같은 보고서를 생성하려면 " 분포 플랫폼의 예 "(34 페이지) 에 설명된 방법을 따르십시오 .

빨간색 삼각형 메뉴

"분포" 옆의 빨간색 삼각형 메뉴에는 모든 변수에 영향을 주는 옵션이 포함되어 있습니다. 자세한 내용은 "분포 플랫폼 옵션"(44 페이지) 에서 확인하십시오.

각 변수 옆의 빨간색 삼각형 메뉴에는 해당 변수에만 영향을 주는 옵션이 포함되어 있습니다. 자세한 내용은 "범주형 변수에 대한 옵션"(45 페이지) 또는 "연속형 변수에 대한 옵션"(46 페이지) 에서 확인하십시오.

팁 : Ctrl 키를 누른 채 변수 옵션을 선택하면 보고서에서 모델링 유형이 동일한 모든 변수에 해당 옵션이 적용됩니다 .

히스토그램 및 보고서

히스토그램에는 데이터가 시각적으로 표시됩니다. 자세한 내용은 "히스토그램"(39 페이지) 에서 확인하십시오.

범주형 변수에 대한 초기 보고서에는 "빈도" 보고서가 포함됩니다. 자세한 내용은 "빈도 보고서"(41 페이지) 에서 확인하십시오.

연속형 변수에 대한 초기 보고서에는 "분위수" 및 "요약 통계량" 보고서가 포함됩니다. 자세한 내용은 "분위수 보고서"(42 페이지) 및 "요약 통계량 보고서"(42 페이지) 에서 확인하십시오.

히스토그램의 변수 바꾸기 , 추가 또는 제거

히스토그램의 변수를 바꾸려면 관련 데이터 테이블의 "열" 패널에서 원하는 변수를 히스토그램의 축으로 드래그하여 놓습니다.

새 변수를 사용하여 보고서에 새 히스토그램을 생성하려면 기존 히스토그램 외부로 변수를 드래그하여 놓습니다. 새 변수와 히스토그램은 기존 히스토그램 앞, 사이 또는 뒤에 추가할 수 있습니다.

히스토그램에서 변수를 제거하려면 빨간색 삼각형 메뉴에서 **제거**를 선택하십시오.

숨겨진 행 및 제외된 행

데이터 테이블의 행에 "숨김" 행 상태를 적용하면 데이터 점을 표시하는 그림에 해당 점이 표시되지 않습니다. 하지만 히스토그램을 생성하는 데는 여전히 숨겨진 행이 사용됩니다.

생성되는 히스토그램과 분석 결과에서 행을 제외하려면 "제외" 행 상태를 적용하십시오. 그런 다음 "분포" 옆의 빨간색 삼각형 메뉴에서 **다시 실행 > 분석 다시 실행**을 선택하십시오. 데이터 테이블에서 제외된 모든 행은 점을 표시하는 그림에서도 숨겨집니다.

히스토그램

히스토그램에는 데이터가 시각적으로 표시됩니다 . 범주형 (명목형 또는 순서형) 변수의 경우에는 히스토그램에 순서형 또는 명목형 변수의 각 수준에 대한 막대가 표시됩니다 . 연속형 변수의 경우에는 히스토그램에 연속형 변수의 그룹화된 값에 대한 막대가 표시됩니다 .

데이터 강조 표시 히스토그램 막대 또는 그래프의 이상점을 클릭합니다 . 그러면 데이터 테이블에서 해당 행이 강조 표시되고 , 다른 히스토그램의 해당 부분도 강조 표시됩니다 (해당되는 경우). 자세한 내용은 " 막대 강조 표시 및 행 선택 "(40 페이지) 에서 확인하십시오 .

부분집합 생성 히스토그램 막대를 두 번 클릭하거나 , 히스토그램 막대를 마우스 오른쪽 버튼으로 클릭한 후 **부분집합**을 선택합니다 . 그러면 선택한 데이터만 포함된 새 데이터 테이블이 생성됩니다 .

전체 히스토그램 크기 조정 히스토그램 테두리 위로 마우스를 이동하여 양방향 화살표가 표시되도록 합니다 . 그런 다음 테두리를 클릭한 채로 드래그합니다 .

축 배율 조정 축 배율을 조정하려면 축을 클릭한 채로 드래그합니다 .

또는 축 위로 마우스를 이동하여 손 모양 포인터가 표시되도록 합니다 . 그런 다음 축을 두 번 클릭하고 " 축 설정 " 창에서 파라미터를 설정합니다 .

히스토그램 막대 크기 조정 (연속형 변수에만 사용 가능) 여러 가지 옵션으로 히스토그램 막대의 크기를 조정할 수 있습니다 . 자세한 내용은 " 연속형 변수의 히스토그램 막대 크기 조정 "(39 페이지) 에서 확인하십시오 .

선택 항목 지정 여러 히스토그램에서 선택한 데이터를 지정할 수 있습니다 . 자세한 내용은 " 여러 히스토그램에서 선택 영역 지정 "(41 페이지) 에서 확인하십시오 .

히스토그램 또는 관련 데이터 테이블에 사용 가능한 추가 옵션을 보려면 다음을 수행하십시오 .

- 히스토그램을 마우스 오른쪽 버튼으로 클릭합니다 . 자세한 내용은 JMP 사용에서 확인하십시오 .
- 축을 마우스 오른쪽 버튼으로 클릭합니다 . 라벨을 추가하거나 축을 수정할 수 있습니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .
- 변수 옆의 빨간색 삼각형을 클릭하고 **히스토그램 옵션**을 선택합니다 . 옵션은 변수의 모델링 유형에 따라 조금씩 다릅니다 . 자세한 내용은 " 범주형 변수에 대한 옵션 "(45 페이지) 또는 " 연속형 변수에 대한 옵션 "(46 페이지) 에서 확인하십시오 .

연속형 변수의 히스토그램 막대 크기 조정

다음을 사용하여 연속형 변수의 히스토그램 막대 크기를 조정할 수 있습니다 .

- 손 도구
- **계급 폭 설정** 옵션
- **증분** 옵션

손 도구 사용

손 도구를 사용하면 데이터를 빠르게 탐색할 수 있습니다 .

1. **도구 > 손 도구**를 선택합니다 .

   참고 : (Windows 에만 해당) 메뉴 표시줄을 표시하려면 창 제목 아래의 막대를 마우스로 가리켜야 합니다 . **파일 > 환경 설정 > Windows 관련**에서 이 설정을 변경할 수도 있습니다 .

2. 손 도구를 히스토그램의 아무 곳에나 놓습니다 .
3. 히스토그램 막대를 클릭한 채로 드래그합니다 .

각 막대는 여러 관측값을 포함하는 계급으로 생각할 수 있습니다 . 세로 히스토그램의 경우는 다음과 같습니다 .

- 손 도구를 왼쪽으로 이동하면 계급 폭이 늘어나고 구간이 결합됩니다 . 막대 크기가 커지면 막대 수는 줄어듭니다 .
- 손 도구를 오른쪽으로 이동하면 계급 폭이 줄어들어 더 많은 막대가 생성됩니다 .
- 손 도구를 위 또는 아래로 이동하면 축에서의 계급 위치가 이동되어 각 계급의 내용 및 크기가 변경됩니다 .

참고 : 히스토그램 방향을 가로로 변경한 경우에는 위의 방향을 반대로 적용하십시오 . 즉 , 손 도구를 아래로 이동하면 계급 폭이 늘어나고 , 위로 이동하면 계급 폭이 줄어들며 , 왼쪽 또는 오른쪽으로 이동하면 축에서의 계급 위치가 이동됩니다 .

계급 폭 설정 옵션 사용

계급 폭 설정 옵션을 사용하면 히스토그램에 있는 모든 막대의 너비를 보다 정확하게 설정할 수 있습니다 . "계급 폭 설정" 옵션을 사용하려면 변수의 빨간색 삼각형 메뉴에서 **히스토그램 옵션 > 계급 폭 설정**을 선택합니다 . 그런 다음 계급 폭 값을 변경합니다 .

증분 옵션 사용

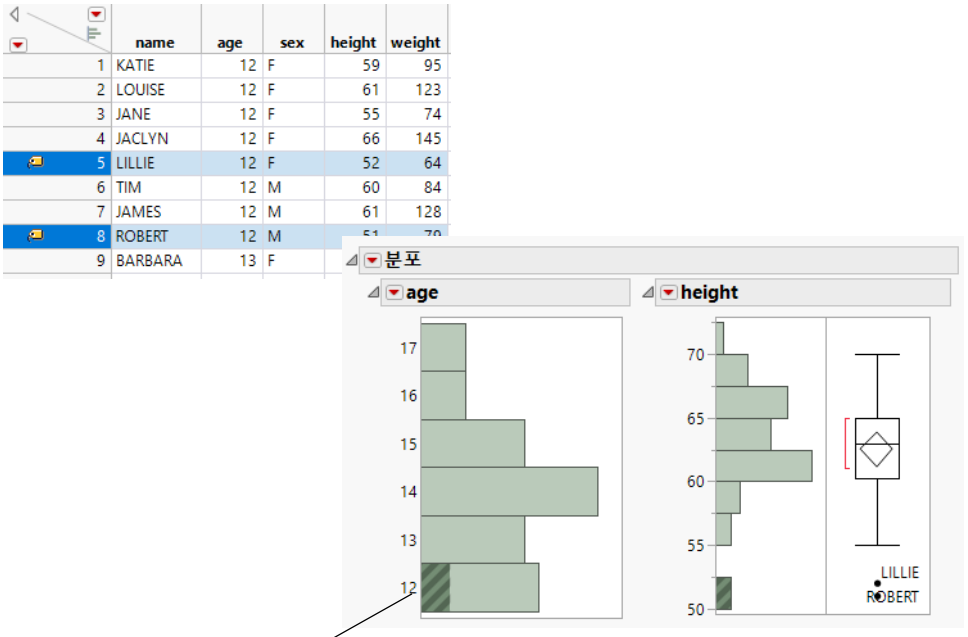
증분 옵션으로도 막대 너비를 정확하게 설정할 수 있습니다 . **증분** 옵션을 사용하려면 축을 두 번 클릭하고 "증분" 값을 변경합니다 .

막대 강조 표시 및 행 선택

히스토그램 막대를 클릭하면 해당 막대가 강조 표시되고 데이터 테이블에서 해당하는 행이 선택됩니다 . 다른 모든 그래픽 표시의 해당 부분에서도 선택 항목이 강조 표시됩니다 . 그림 3.6 에서는 height 히스토그램의 막대를 강조 표시한 결과를 보여 줍니다 . 데이터 테이블에서는 해당 행이 선택됩니다 .

팁 : 특정 히스토그램 막대를 선택 취소하려면 Ctrl 키를 누른 채로 강조 표시된 막대를 클릭하십시오 .

그림 3.6 막대 및 행 강조 표시



여러 히스토그램에서 선택 영역 지정

여러 히스토그램에서 선택 영역을 넓히거나 좁힐 수 있습니다.

- 선택 영역을 넓히려면 Shift 키를 누른 채로 다른 막대를 선택합니다. 이 방법은 or 연산자를 사용하는 것과 동등합니다.
- 선택 영역을 좁히려면 Ctrl 키와 Alt 키 (Windows), 또는 Command 키와 Option 키 (macOS)를 누른 채로 다른 막대를 선택합니다. 이 방법은 and 연산자를 사용하는 것과 동등합니다.

예는 "예: 여러 히스토그램에서 데이터 선택"(65 페이지)에서 확인하십시오.

빈도 보고서

명목형 및 순서형 변수의 경우 "빈도" 보고서에 변수의 수준과 함께 관련 발생 빈도 및 확률이 나열됩니다.

"빈도" 보고서에는 범주형(명목형 또는 순서형) 변수의 각 수준별로 다음 목록에 설명된 정보가 포함됩니다. 결측값은 분석에서 생략됩니다.

팁 : " 빈도 " 보고서에서 값을 클릭하면 히스토그램과 데이터 테이블에서 해당하는 데이터가 선택됩니다 .

수준 반응 변수에 대해 발견된 각 값을 나열합니다 .

개수 반응 변수의 각 수준에 대해 발견된 행 수를 나열합니다 . " 빈도 " 변수를 사용할 경우 이 개수는 반응 변수의 각 수준에 대한 빈도 변수의 합계입니다 .

확률 반응 변수의 각 수준에 대한 발생 확률 (또는 비율) 을 나열합니다 . 이 확률은 각 수준의 개수를 테이블 맨 아래에 표시된 변수의 총 빈도로 나눠서 계산됩니다 .

표준 오차 확률 확률의 표준 오차를 나열합니다 . 이 열이 숨겨져 있을 수도 있습니다 . 이 열을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열 > 표준 오차 확률**을 선택합니다 .

누적 확률 확률 열의 누적 합계를 포함합니다 . 이 열이 숨겨져 있을 수도 있습니다 . 이 열을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열 > 누적 확률**을 선택합니다 .

분위수 보고서

연속형 변수의 경우 "분위수 보고서"에 선택된 분위수(**백분위수**라고도 함)의 값이 나열됩니다. 통계 상세 정보는 "분위수에 대한 통계 상세 정보"(74 페이지)에서 확인하십시오.

요약 통계량 보고서

연속형 변수의 경우 "요약 통계량" 보고서에 평균, 표준편차 및 기타 요약 통계량이 표시됩니다. "요약 통계량" 옆의 빨간색 삼각형 메뉴에서 **요약 통계량 사용자 정의**를 선택하면 이 보고서에 표시되는 통계량을 제어할 수 있습니다.

팁 : 연속형 변수에 대한 분포 분석을 실행할 때마다 보고서에 표시할 요약 통계량을 지정할 수 있습니다 . **파일 > 환경 설정 > 플랫폼 > 분포 요약 통계량**을 선택하고 표시하려는 요약 통계량을 선택하십시오 .

- 요약 통계량 보고서 설명에서는 기본적으로 표시되는 통계량에 대해 설명합니다 .
- 추가 요약 통계량에서는 **요약 통계량 사용자 정의** 창을 사용하여 보고서에 추가할 수 있는 기타 통계량에 대해 설명합니다 .

요약 통계량 보고서 설명

평균 반응 변수에 대해 주어진 분포하에서의 기대값을 추정합니다 . 이 값은 열 값의 산술평균입니다 . 즉 , 비결측값의 합을 비결측값의 개수로 나눈 값입니다 .

표준편차 정규 분포는 주로 평균과 표준편차로 정의됩니다 . 이러한 모수를 사용하면 표본 크기가 커질 경우 데이터를 손쉽게 요약할 수 있습니다 .

- 값의 68% 는 평균의 1 표준편차 범위 내에 있습니다 .
- 값의 95% 는 평균의 2 표준편차 범위 내에 있습니다 .

– 값의 99.7% 는 평균의 3 표준편차 범위 내에 있습니다 .

평균 표준 오차 평균의 표준 오차로 , 평균 분포의 표준편차를 추정합니다 .

평균의 신뢰 상한 / 하한 평균에 대한 95% 신뢰 한계로 , 실제 모집단 평균을 포함할 가능성이 매우 높은 구간을 정의합니다 .

N 비결측값의 총 개수입니다 .

추가 요약 통계량

가중합 시작 창에서 " 가중치 " 역할에 할당된 열의 합계입니다 . 가중합은 평균 계산 시 *N* 을 대신해서 분모로 사용됩니다 .

합 반응 값의 합입니다 .

분산 표본 분산으로 , 표본 표준편차의 제곱입니다 .

왜도 편중성 또는 대칭성을 측정합니다 .

첨도 꼬리의 뾰족함 또는 두꺼움을 측정합니다 . 계산식에 대한 자세한 내용은 " 첨도 "(76 페이지) 에서 확인하십시오 .

CV 변동 계수 백분율로 , 표준편차를 평균으로 나눈 후 100 을 곱해서 계산됩니다 . 변동 계수는 상대적 변동을 평가하는 데 사용할 수 있습니다 . 예를 들어 서로 다른 단위 또는 서로 다른 크기로 측정된 데이터의 변동을 비교할 때 이 통계량을 사용할 수 있습니다 .

결측값 수 결측 관측값의 개수입니다 .

0 개수 0 값의 개수입니다 .

고유 값 수 고유 값의 개수입니다 .

비수정 SS 비수정 제곱합 , 즉 제곱 값의 합입니다 .

수정 SS 수정 제곱합 , 즉 평균으로부터의 편차를 제곱한 값의 합입니다 .

자기상관 (" 빈도 " 변수를 지정하지 않은 경우에만 표시됨) 잔차가 행 간에 상관되어 있는지 여부를 검정하는 1 차 자기상관입니다 . 이 검정은 데이터의 비랜덤성을 감지하는 데 유용합니다 .

최소값 데이터의 0 번째 백분위수를 나타냅니다 .

최대값 데이터의 100 번째 백분위수를 나타냅니다 .

중앙값 데이터의 50 번째 백분위수를 나타냅니다 .

최빈값 데이터에서 발생 빈도가 가장 큰 값입니다 . 최빈값이 여러 개이면 가장 작은 최빈값이 표시됩니다 .

절사평균 데이터의 하위 p% 와 상위 p% 를 제거한 후 계산된 평균입니다 . p 값은 창의 맨 아래에 있는 **절사평균 백분율 입력** 텍스트 상자에 입력합니다 . " 가중치 " 변수를 지정한 경우에는 " 절사평균 " 옵션을 사용할 수 없습니다 .

기하평균 데이터 곱의 *n* 제곱근입니다 . 예를 들어 기하평균은 이자율을 계산하는 데 종종 사용됩니다 . 이 통계량은 데이터에 많은 값이 비대칭적 분포로 포함된 경우에도 유용합니다 .

참고 : 음수 값이 있으면 결측값이 발생하고 , 음수가 아니고 0 인 값이 있으면 결과가 0 이 됩니다 .

범위 데이터 최대값과 최소값 사이의 차이입니다 .

사분위수 범위 3 사분위수와 1 사분위수 사이의 차이입니다 .

중앙 절대 편차 (" 가중치 " 변수를 지정한 경우에는 표시되지 않음) 중앙값과의 절대 편차에 대한 중앙값입니다 .

0 인 비율 0 인 비결측값의 비율입니다 .

0 이 아닌 비율 0 이 아닌 비결측값의 비율입니다 .

로버스트 평균 로버스트 평균은 Huber M 추정을 사용하여 이상치의 영향을 받지 않는 방식으로 계산됩니다 . 자세한 내용은 Huber and Ronchetti 연구 자료 (2009) 에서 확인하십시오 .

로버스트 표준편차 로버스트 표준편차는 Huber M 추정을 사용하여 이상치의 영향을 받지 않는 방식으로 계산됩니다 . 자세한 내용은 Huber and Ronchetti 연구 자료 (2009) 에서 확인하십시오 .

평균 신뢰 구간의 (1-α) 입력 평균 신뢰 구간의 유의 수준을 지정합니다 .

절사평균 백분율 입력 절사평균 백분율을 지정합니다 . 데이터의 양쪽에서 이 백분율만큼 절사됩니다 .

요약 통계량 옵션

" 요약 통계량 " 옆의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

요약 통계량 사용자 정의 표시하려는 통계량을 목록에서 선택합니다 . 모든 요약 통계량을 선택하거나 선택 취소할 수 있습니다 .

모든 최빈값 표시 최빈값이 여러 개 있는 경우 최빈값을 모두 표시합니다 .

통계 상세 정보는 " 요약 통계량 "(75 페이지) 에서 확인하십시오 .

분포 플랫폼 옵션

" 분포 " 옆의 빨간색 삼각형 메뉴에는 분포 플랫폼의 모든 보고서 및 그래프 수에 영향을 주는 옵션이 포함되어 있습니다 .

균등 척도 분포를 쉽게 비교할 수 있도록 모든 축 척도의 최소값 , 최대값 및 간격을 동일하게 설정합니다 .

쌓기 히스토그램과 보고서의 방향을 가로로 변경하고 개별 분포 보고서를 세로로 쌓습니다 . 보고서 창을 원래 레이아웃으로 되돌리려면 이 옵션을 선택 취소합니다 .

행으로 배열 한 행에 표시할 그림의 개수를 입력합니다 . 이 옵션을 사용하면 여러 그림을 하나의 넓은 행으로 표시하지 않고 세로로 표시할 수 있습니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

범주형 변수에 대한 옵션

보고서 창에서 각 변수 옆에 있는 빨간색 삼각형 메뉴에는 해당 변수에 적용되는 추가 옵션이 포함되어 있습니다. 이 섹션에서는 범주형 (명목형 또는 순서형) 변수에 사용할 수 있는 옵션을 설명합니다.

표시 옵션 자세한 내용은 " 범주형 변수에 대한 표시 옵션 "(45 페이지) 에서 확인하십시오.

히스토그램 옵션 자세한 내용은 " 범주형 변수에 대한 히스토그램 옵션 "(46 페이지) 에서 확인하십시오.

모자이크 그림 각각의 명목형 또는 순서형 반응 변수에 대한 모자이크 막대 차트를 표시합니다. 모자이크 그림은 각 세그먼트가 해당 그룹의 빈도 수에 비례하는 누적 막대 차트입니다.

정렬 기준 히스토그램, 모자이크 그림 및 빈도 보고서를 오름차순 / 내림차순 또는 개수별로 재정렬합니다. 새 순서를 열 특성으로 저장하려면 **저장 > 값 순서화** 옵션을 사용합니다.

검정 확률 가설 확률을 검정하는 보고서를 표시합니다. 자세한 내용은 "검정 확률 옵션의 예"(67 페이지) 에서 확인하십시오.

신뢰 구간 이 메뉴에는 신뢰 구간이 포함되어 있습니다. 나열된 값 중에서 선택하거나, **기타**를 클릭한 후 값을 직접 입력합니다. 그러면 JMP 가 점수 신뢰 구간을 계산합니다.

저장 자세한 내용은 " 범주형 변수에 대한 저장 옵션 "(46 페이지) 에서 확인하십시오.

제거 해당 변수와 모든 관련 보고서를 분포 보고서에서 영구적으로 제거합니다.

범주형 변수에 대한 표시 옵션

빈도 " 빈도 " 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 빈도 보고서 "(41 페이지) 에서 확인하십시오.

가로 레이아웃 히스토그램과 보고서의 방향을 세로 또는 가로로 변경합니다.

왼쪽의 축 **개수**, **확률** 및 **밀도** 축을 오른쪽 대신 왼쪽으로 이동합니다.

이 옵션은 **가로 레이아웃**이 선택된 경우에만 적용할 수 있습니다 .

범주형 변수에 대한 히스토그램 옵션

히스토그램 히스토그램을 표시하거나 숨깁니다 . 자세한 내용은 " 히스토그램 "(39 페이지) 에서 확인하십시오 .

세로 히스토그램의 방향을 세로에서 가로 방향으로 변경합니다 .

표준 오차 막대 히스토그램의 각 수준에 표준 오차 막대를 그립니다 .

막대 분리 히스토그램 막대를 분리합니다 .

히스토그램 색상 히스토그램 막대의 색상을 변경합니다 .

개수 축 히스토그램 막대가 나타내는 열 값의 빈도를 보여 주는 축을 추가합니다 .

확률 축 히스토그램 막대가 나타내는 열 값의 비율을 보여 주는 축을 추가합니다 .

밀도 축 히스토그램에 막대 길이를 보여 주는 축을 추가합니다 .

개수 및 확률 축은 다음과 같은 계산을 기준으로 합니다 .

확률 = (막대 너비) * 밀도

개수 = (막대 너비) * 밀도 * (총 개수)

백분율 표시 각 히스토그램 막대가 나타내는 열 값의 백분율을 라벨로 표시합니다 .

참고 : 소수 자릿수를 지정하려면 히스토그램을 마우스 오른쪽 버튼으로 클릭한 후 **사용자 정의 > 히스토그램**을 선택하십시오 .

개수 표시 각 히스토그램 막대가 나타내는 열 값의 빈도를 라벨로 표시합니다 .

범주형 변수에 대한 저장 옵션

수준 수 데이터 테이블에 **수준 < 열 이름 >** 이라는 새 열을 생성합니다 . 각 관측값의 수준 수는 해당 관측값이 포함된 히스토그램 막대에 해당합니다 .

값 순서화 (**정렬 기준** 옵션과 함께 사용) 새 순서를 반영하여 데이터 테이블에 새로운 값 순서화 열 특성을 생성합니다 .

로그에 스크립트로 현재 보고서를 생성하기 위한 스크립트 명령을 로그 창에 표시합니다 . 로그 창을 보려면 **보기 > 로그**를 선택합니다 .

연속형 변수에 대한 옵션

보고서 창에서 각 변수 옆에 있는 빨간색 삼각형 메뉴에는 해당 변수에 적용되는 추가 옵션이 포함되어 있습니다 . 이 섹션에서는 연속형 변수에 사용할 수 있는 옵션을 설명합니다 .

표시 옵션 자세한 내용은 " 연속형 변수에 대한 표시 옵션 "(47 페이지) 에서 확인하십시오 .

히스토그램 옵션 자세한 내용은 " 연속형 변수에 대한 히스토그램 옵션 "(48 페이지) 에서 확인하십시오 .

정규 분위수 그림 변수의 정규 분포 정도를 쉽게 시각화할 수 있습니다 . 자세한 내용은 " 정규 분위수 그림 "(49 페이지) 에서 확인하십시오 .

이상치 상자 그림 분포를 표시하며 가능한 이상치를 쉽게 확인할 수 있습니다 . 자세한 내용은 " 이상치 상자 그림 "(50 페이지) 에서 확인하십시오 .

분위수 상자 그림 " 분위수 " 보고서의 특정 분위수를 표시합니다 . 자세한 내용은 " 분위수 상자 그림 "(51 페이지) 에서 확인하십시오 .

줄기 - 잎 자세한 내용은 " 줄기 - 잎 "(53 페이지) 에서 확인하십시오 .

CDF 그림 경험적 누적 분포 함수 그림을 생성합니다 . 자세한 내용은 "CDF 그림 "(53 페이지) 에서 확인하십시오 .

평균 검정 평균에 대한 1 표본 검정을 수행합니다 . 자세한 내용은 " 평균 검정 "(54 페이지) 에서 확인하십시오 .

표준편차 검정 표준편차에 대한 1 표본 검정을 수행합니다 . 자세한 내용은 " 표준편차 검정 "(55 페이지) 에서 확인하십시오 .

동등성 검정 모집단 평균이 가설 값과 동등한지 평가합니다 . 자세한 내용은 " 동등성 검정 "(55 페이지) 에서 확인하십시오 .

신뢰 구간 평균 및 표준편차의 신뢰 구간을 선택합니다 . 자세한 내용은 " 신뢰 구간 "(56 페이지) 에서 확인하십시오 .

예측 구간 단일 관측값이나 다음 무작위 선택 표본의 평균 및 표준편차에 대한 예측 구간을 선택합니다 . 자세한 내용은 " 예측 구간 "(57 페이지) 에서 확인하십시오 .

공차 구간 모집단의 지정된 비율 이상을 포함하기 위한 구간을 계산합니다 . 자세한 내용은 " 공차 구간 "(57 페이지) 에서 확인하십시오 .

공정 능력 지정된 규격 한계에 대한 공정 적합성을 측정합니다 . 자세한 내용은 " 공정 능력 "(57 페이지) 에서 확인하십시오 .

연속형 적합 연속형 변수에 분포를 적합시킵니다 . 자세한 내용은 " 분포 적합 "(59 페이지) 에서 확인하십시오 .

이산형 적합 (모든 데이터 값이 정수인 경우에만 사용 가능) 이산형 변수에 분포를 적합시킵니다 . 자세한 내용은 " 분포 적합 "(59 페이지) 에서 확인하십시오 .

저장 연속형 또는 범주형 변수에 대한 정보를 저장합니다 . 자세한 내용은 " 예측 구간 "(57 페이지) 에서 확인하십시오 .

제거 해당 변수와 모든 관련 보고서를 분포 보고서에서 영구적으로 제거합니다 .

연속형 변수에 대한 표시 옵션

분위수 " 분위수 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 분위수 보고서 "(42 페이지) 에서 확인하십시오 .

분위수 증분 설정 분위수 증분을 변경하거나 기본 분위수 증분으로 되돌립니다 .

사용자 분위수 값 또는 증분을 기준으로 사용자 분위수를 설정합니다 . 신뢰 수준을 지정하고 , 평활된 경험적 가능도 분위수를 계산할지 여부를 선택할 수 있습니다 . 데이터 집합이 큰 경우에는 다소 시간이 걸릴 수 있습니다 .

- 가중 평균 분위수의 추정 방법에 대한 자세한 내용은 " 분위수에 대한 통계 상세 정보 "(74 페이지) 에서 확인하십시오 .
- 가중 평균 분위수의 분포 무관 신뢰 한계에 대한 자세한 내용은 Meeker et al. 연구 자료의 섹션 5.2 (2017) 에서 확인하십시오 .
- 평활된 경험적 가능도 분위수는 커널 밀도 추정값을 기반으로 합니다 . 이러한 분위수와 해당 신뢰 한계에 대한 자세한 내용은 Chen and Hall 연구 자료 (1993) 에서 확인하십시오 .
- 소수 빈도를 사용하는 경우에는 신뢰 구간과 평활된 경험적 가능도 분위수를 사용할 수 없습니다 .

요약 통계량 " 요약 통계량 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 요약 통계량 보고서 "(42 페이지) 에서 확인하십시오 .

요약 통계량 사용자 정의 " 요약 통계량 " 보고서의 통계량을 추가하거나 제거합니다 . 자세한 내용은 " 요약 통계량 보고서 "(42 페이지) 에서 확인하십시오 .

가로 레이아웃 히스토그램과 보고서의 방향을 세로 또는 가로로 변경합니다 .

왼쪽의 축 개수 , 확률 , 밀도 및 **정규 분위수 그림** 축을 오른쪽 대신 왼쪽으로 이동합니다 .

이 옵션은 **가로 레이아웃**이 선택된 경우에만 적용할 수 있습니다 .

연속형 변수에 대한 히스토그램 옵션

히스토그램 히스토그램을 표시하거나 숨깁니다 . 자세한 내용은 " 히스토그램 "(39 페이지) 에서 확인하십시오 .

새도그램 히스토그램을 새도그램으로 바꿉니다 . 히스토그램은 계급 폭이 변경될 경우 히스토그램 모양이 변경되지만 새도그램은 계급 폭을 달리 한 여러 히스토그램을 중첩 표시한 것입니다 . 새도그램에서는 분포의 주요 특징이 덜 뚜렷하게 나타납니다 .

새도그램에서는 다음 옵션을 사용할 수 없습니다 .

- 표준 오차 막대
- 개수 표시
- 백분율 표시

세로 히스토그램의 방향을 세로에서 가로 방향으로 변경합니다 .

표준 오차 막대 표준 오차를 사용하여 히스토그램의 각 수준에 표준 오차 막대를 그립니다 . 손 도구로 막대 수를 조정하면 표준 오차 막대가 자동으로 조정됩니다 . 자세한 내용은 " 연속형 변수의 히스토그램 막대 크기 조정 "(39 페이지) 및 " 표준 오차 막대 "(74 페이지) 에서 확인하십시오 .

계급 폭 설정 히스토그램 막대의 계급 폭을 변경합니다 . 자세한 내용은 " 연속형 변수의 히스토그램 막대 크기 조정 "(39 페이지) 에서 확인하십시오 .

히스토그램 색상 히스토그램 막대의 색상을 변경합니다 .

개수 축 히스토그램 막대가 나타내는 열 값의 빈도를 보여 주는 축을 추가합니다 .

참고 : 히스토그램 막대의 크기를 조정하면 개수 축의 크기도 조정됩니다 .

확률 축 히스토그램 막대가 나타내는 열 값의 비율을 보여 주는 축을 추가합니다 .

참고 : 히스토그램 막대의 크기를 조정하면 확률 축의 크기도 조정됩니다 .

밀도 축 밀도는 히스토그램의 막대 길이입니다. 개수와 확률은 모두 다음과 같은 계산을 기준으로 합니다 .

확률 = (막대 너비) * 밀도

개수 = (막대 너비) * 밀도 * (총 개수)

적합 분포에 의해 추가된 밀도 곡선을 살펴볼 때 밀도 축에는 곡선의 점 추정값이 표시됩니다.

참고 : 히스토그램 막대의 크기를 조정하면 밀도 축의 크기도 조정됩니다 .

백분율 표시 각 히스토그램 막대가 나타내는 열 값의 비율을 라벨로 표시합니다 .

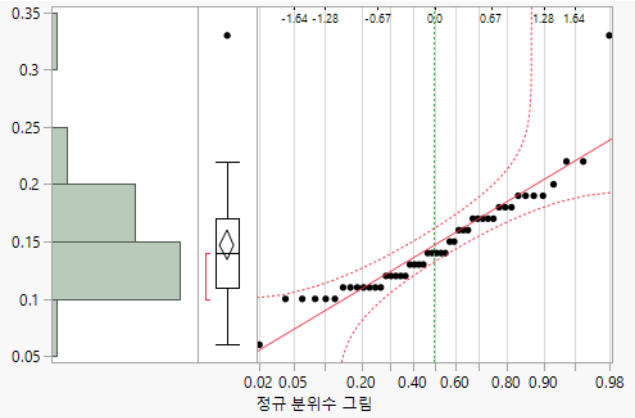
개수 표시 각 히스토그램 막대가 나타내는 열 값의 빈도를 라벨로 표시합니다 .

정규 분위수 그림

변수의 정규 분포 정도를 시각화하려면 **정규 분위수 그림** 옵션을 사용합니다. 변수가 정규 분포된 경우 정규 분위수 그림은 대각 직선에 근사합니다. 이 유형의 그림을 분위수-분위수 그림 또는 Q-Q 그림이라고도 합니다.

정규 분위수 그림에서는 Lilliefors 신뢰경계 (Conover 1980) 와 확률 및 정규 분위수 척도도 보여 줍니다.

그림 3.7 정규 분위수 그림



다음 사항에 유의하십시오 .

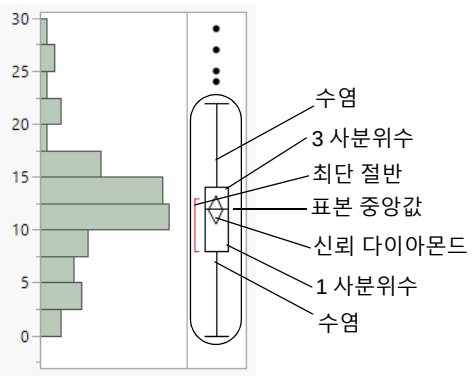
- 세로 축은 열 값을 보여 줍니다 .
- 위쪽 가로 축은 정규 분위수 척도를 보여 줍니다 .
- 아래쪽 가로 축은 각 값의 경험적 누적 확률을 보여 줍니다 .
- 빨간색 파선은 Lilliefors 신뢰경계를 보여 줍니다 .

통계 상세 정보는 "정규 분위수 그림"(76 페이지)에서 확인하십시오 .

이상치 상자 그림

분포를 보고 가능한 이상치를 확인하려면 이상치 상자 그림 (Tukey 이상치 상자 그림이라고도 함) 을 사용합니다 . 일반적으로 상자 그림에는 연속형 분포 중 선택된 분위수가 표시됩니다 .

그림 3.8 이상치 상자 그림



이상치 상자 그림의 경우 다음 사항에 유의하십시오 .

- 상자 내의 가로 선은 표본 중앙값을 나타냅니다 .
- 신뢰 다이아몬드에는 평균과 평균의 95% 상한 및 하한이 포함됩니다 . 다이아몬드의 중앙을 지나는 선을 그리면 평균을 알 수 있습니다 . 다이아몬드의 위쪽 및 아래쪽 점은 평균의 95% 상한 및 하한을 나타냅니다 .
- 상자의 끝은 25 번째 및 75 번째 분위수를 나타내며 각각 **1 사분위수 및 3 사분위수**라고도 합니다 .
- 1 사분위수와 3 사분위수 사이의 차이를 **사분위수 범위** . 라고 합니다 .
- 상자에는 양 끝에서 시작되는 선이 있으며 이 선은 **수염**이라고도 합니다 . 수염은 상자 끝에서 시작해서 다음과 같이 계산된 거리 내에 있는 점들 중 가장 바깥쪽 데이터 점까지 이어집니다 .

  1 사분위수 - 1.5 * (사분위수 범위)

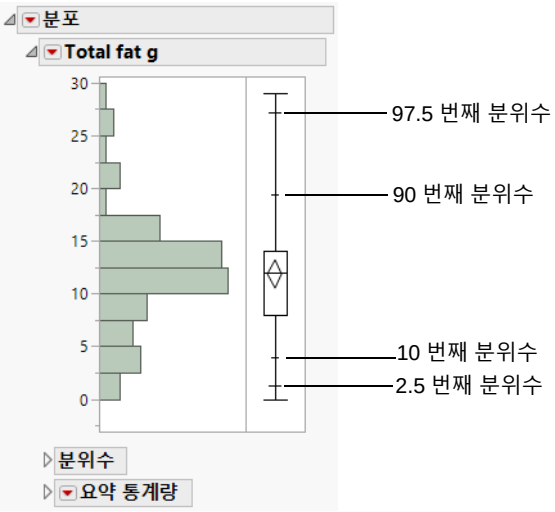
  3 사분위수 + 1.5 * (사분위수 범위)

  데이터 점이 계산된 범위에 포함되지 않으면 상한 및 하한 데이터 점 값 (이상치 제외) 에 따라 수염이 결정됩니다 .
- 상자 외부의 괄호는 관측값 중 밀도가 가장 높은 50% 에 해당하는 **최단 절반**을 나타냅니다 (Rousseeuw and Leroy 1987).
- 이상치 상자 그림에서 개체를 제거하는 방법은 " 이상치 또는 분위수 상자 그림에서 개체 제거 "(52 페이지) 에서 확인하십시오 .

분위수 상자 그림

분위수 상자 그림에는 "분위수" 보고서의 특정 분위수가 표시됩니다. 분포가 대칭적인 경우 상자 그림의 분위수는 서로 거의 같은 거리로 나타납니다. 따라서 분포가 대칭적인지 여부를 한눈에 알 수 있습니다. 예를 들어 분위수 표식이 한 쪽 끝에 밀접해서 모여 있지만 다른 쪽 끝에 더 큰 간격이 있는 경우 분포는 더 큰 간격을 갖는 끝 쪽으로 치우칩니다.

그림 3.9 분위수 상자 그림



분위수는 값 중 하나로, *p* 번째 분위수는 *p*% 의 값보다 큰 값입니다. 예를 들어 데이터의 10% 는 10 번째 분위수 아래에 있으며, 데이터의 90% 는 90 번째 분위수 아래에 있습니다.

이상치 또는 분위수 상자 그림에서 개체 제거

이상치 또는 분위수 상자 그림에서 신뢰 다이아몬드와 최단 절반을 제거할 수 있습니다. 단일 그래프에서 제거할 수도 있고, 이후의 모든 그래프에서 제거할 수도 있습니다.

개별 그래프에서 개체를 제거하려면

1. 이상치 상자 그림을 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다 .
2. **상자 그림**을 클릭합니다 .
3. **신뢰 다이아몬드** 또는 **최단 절반** 옆의 체크박스를 선택 취소합니다 .

"그래프 사용자 정의" 창에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

이후의 모든 그래프에서 개체를 제거하려면

1. **파일** > **환경 설정** > **플랫폼** > **분포**를 선택합니다 .
2. 다음 옵션을 선택 취소합니다 .
   - **상자 그림 신뢰 다이아몬드 표시**
   - **이상치 상자 그림 최단 절반 표시**
3. **확인**을 클릭합니다 .

   이제 분포 플랫폼에서 추가하는 모든 상자 그림에는 신뢰 다이아몬드나 최단 절반이 포함되지 않습니다 .

줄기 - 잎

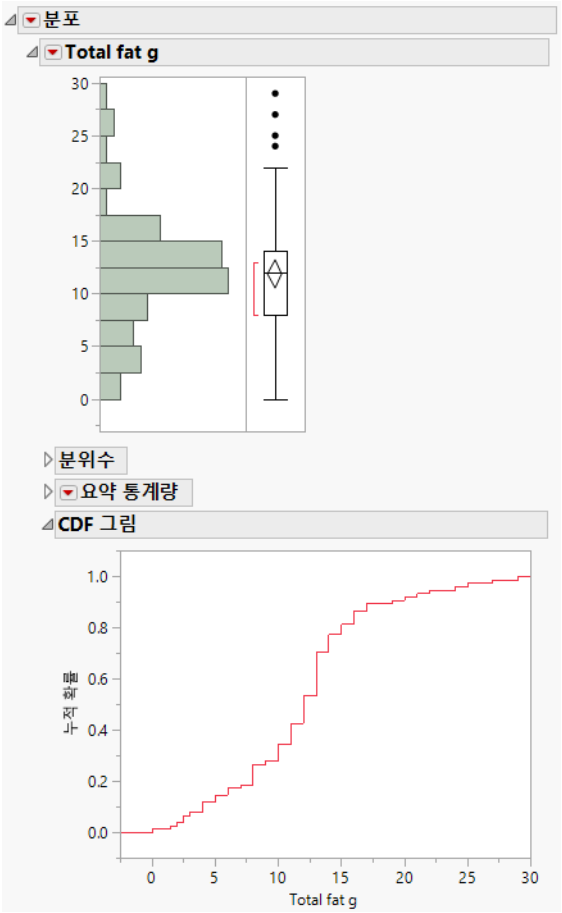
이 그림의 각 행에는 열 값 범위의 선행 숫자를 나타내는 **줄기** 값이 있습니다. **잎** 값은 값의 뒷자리 숫자로 이루어집니다. 줄기와 잎을 결합하면 데이터 점을 확인할 수 있습니다. 일부 경우에 줄기-잎 그림의 숫자는 테이블의 실제 데이터에 대한 근사값입니다. 줄기-잎 그림은 마우스 버튼을 클릭하거나 브러시 도구를 사용할 때 그에 따라 변경됩니다.

참고 : 줄기 - 잎 그림에서는 소수 빈도가 지정된 빈도 이상의 가장 작은 정수로 변환됩니다 .

CDF 그림

분포의 CDF 그림은 경험적 누적 분포 함수 그림을 생성합니다. 가로 축의 특정 값보다 작거나 같은 데이터의 백분율을 확인하려면 CDF 그림을 사용합니다. 데이터에 분포가 적합되면 적합된 분포에 대한 누적 분포 함수가 CDF 그림에 중첩 표시됩니다.

그림 3.10 CDF 그림



예를 들어 이 CDF 그림에서 총 지방 값이 10g인 데이터는 약 34% 입니다.

평균 검정

평균에 대한 1표본 검정 옵션을 지정하고 검정을 수행하려면 **평균 검정** 창을 사용합니다. 표준편차 값을 지정하면 *z*-검정이 수행됩니다. 그렇지 않으면 표본 표준편차를 사용하여 *t*-검정이 수행됩니다. 비모수 Wilcoxon 부호 순위 검정을 요청할 수도 있습니다.

서로 다른 여러 값을 검정하려면 **평균 검정** 옵션을 반복적으로 사용합니다. 평균을 검정할 때마다 새 "평균 검정" 보고서가 표시됩니다.

평균 검정 보고서 설명

평균 검정을 위해 계산되는 통계량

t- 검정 (또는 z- 검정) 양측 및 단측 대립 가설에 대한 검정 통계량 값과 *p* 값을 나열합니다 .

부호 순위 ("Wilcoxon 부호 순위 " 가 선택된 경우에만 표시됨) 양측 및 단측 대립 가설에 대한 Wilcoxon 부호 순위 통계량과 *p* 값을 차례로 나열합니다 . 이 검정에서는 Pratt 방법을 사용하여 0 값을 처리합니다 . 이 검정은 비모수 검정으로서 중앙값이 가정된 값과 같다는 것을 귀무가설로 하며 , 분포가 대칭적이라고 가정합니다 . 자세한 내용은 "Wilcoxon 부호 순위 검정 "(76 페이지) 에서 확인하십시오 .

확률 값

Prob > |t| 모집단 평균이 가설 값과 같을 때 절대 *t* 값이 관측된 *t* 값보다 클 확률입니다 . 이 값은 양쪽 꼬리 *t*- 검정의 관측된 유의성에 대한 *p* 값입니다 .

Prob > t 모집단 평균이 가설 값과 다르지 않을 때 *t* 값이 계산된 표본 *t* 비보다 클 확률입니다 . 이 값은 위쪽 꼬리 검정의 *p* 값입니다 .

Prob < t 모집단 평균이 가설 값과 다르지 않을 때 *t* 값이 계산된 표본 *t* 비보다 작을 확률입니다 . 이 값은 아래쪽 꼬리 검정의 *p* 값입니다 .

평균 검정 옵션 설명

p 값 애니메이션 *p* 값에 대한 대화식 시각적 표현을 시작합니다 . 가설 평균 값을 변경하면서 변경 사항이 *p* 값에 어떤 영향을 주는지 살펴볼 수 있습니다 .

검정력 애니메이션 검정력 및 베타에 대한 대화식 시각적 표현을 시작합니다 . 가설 평균과 표본 평균을 변경하면서 변경 사항이 검정력과 베타에 어떤 영향을 주는지 살펴볼 수 있습니다 .

검정 제거 평균 검정을 제거합니다 .

표준편차 검정

표준편차에 대한 1 표본 검정을 수행하려면 **표준편차 검정** 옵션을 사용합니다. 서로 다른 여러 값을 검정하려면 **표준편차 검정** 옵션을 반복적으로 사용합니다. 표준편차를 검정할 때마다 새로운 " 표준편차 검정 " 보고서가 표시됩니다 .

검정 통계량 카이제곱 검정 통계량의 값을 제공합니다 . 자세한 내용은 " 표준편차 검정 "(78 페이지) 에서 확인하십시오 .

최소 p 값 모집단 표준편차가 가설 값과 다르지 않을 때 보다 극단적인 카이제곱 값을 얻을 확률입니다 . 자세한 내용은 " 표준편차 검정 "(78 페이지) 에서 확인하십시오 .

Prob>ChiSq 모집단 표준편차가 가설 값과 다르지 않을 때 카이제곱 값이 계산된 표본 카이제곱보다 클 확률입니다 . 이 값은 한쪽 꼬리 *t*- 검정의 관측된 유의성에 대한 *p* 값입니다 .

Prob<ChiSq 모집단 표준편차가 가설 값과 다르지 않을 때 카이제곱 값이 계산된 표본 카이제곱보다 작을 확률입니다 . 이 값은 한쪽 꼬리 *t*- 검정의 관측된 유의성에 대한 *p* 값입니다 .

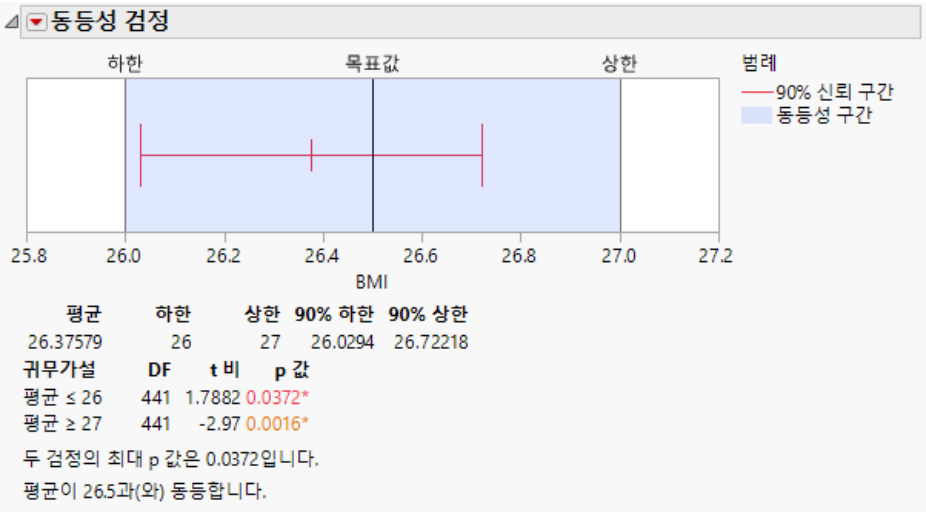
동등성 검정

동등성 검정에서는 모집단 평균이 가설 값과 동등한지 평가합니다. 차이가 없는 것과 동등하게 간주되는 임계 차이를 정의해야 합니다. "동등성 검정" 옵션은 TOST(Two One-Sided Tests: 두 개의 단측 검정) 방법을 사용합니다. 두 개의 단측 *t* 검정은 실제 평균과 가설 값 사이의 차이가 임계를 초과하는 귀무가설에 대해 구성됩니다. 두 귀무가설이 모두 기각되면 실제 차이가 임계를 초과하지 않는 것입니다. 이 경우 평균이 실제적으로 가설 값과 동등한 것으로 간주할 수 있다는 결론을 내릴 수 있습니다.

"동등성 검정" 옵션을 선택할 때는 가설 평균, 임계 차이(실제적으로 0 으로 간주되는 차이) 및 신뢰 수준을 지정합니다. 신뢰 수준은 1-α 이며, 여기서 α 는 각 단측 검정의 유의 수준입니다.

그림 3.11 에서는 Diabetes.jmp 샘플 데이터 테이블의 BMI 변수에 대한 "동등성 검정" 보고서를 보여 줍니다. 가설 평균은 26.5 이고, 실제적으로 0 으로 간주되는 차이는 0.5 로 지정되어 있습니다.

그림 3.11 동등성 검정 보고서



이 보고서에는 다음이 표시됩니다 .

- " 목표값 " 과 경계 (" 하한 " 및 " 상한 " 이라는 라벨이 표시된 수직선으로 정의됨) 를 보여 주는 정의된 동등성 구간의 그림
- 계산된 평균의 신뢰 구간 . 이 신뢰 구간은 1 - 2 * 유의 수준 구간입니다 .
- 계산된 평균 , 지정된 하한 및 상한 경계 , 평균의 (1 - 2 * 알파) 수준 신뢰 구간을 보여 주는 테이블
- 두 개의 단측 검정 결과를 보여 주는 테이블
- 결과를 요약하고 , 평균을 목표값과 동등한 것으로 간주할 수 있는지 여부를 언급하는 노트

신뢰 구간

연속형 변수의 **신뢰 구간** 옵션은 평균과 표준편차의 신뢰 구간을 표시합니다. **0.90**, **0.95** 및 **0.99** 옵션은 평균과 표준편차의 양측 신뢰 구간을 계산합니다. **신뢰 구간 > 기타** 옵션을 사용하여 신뢰 수준을 선택하고 단측 또는 양측 신뢰 구간을 선택할 수 있습니다. 알려진 시그마를 입력할 수도 있습니다. 알려진 시그마를 사용할 경우 평균의 신뢰 구간은 *t* 값이 아니라 *z* 값을 기준으로 합니다.

"신뢰 구간" 보고서에서는 평균 및 표준편차 모수 추정값과 함께 1 - α에 대한 신뢰 상한 및 하한을 보여 줍니다.

예측 구간

예측 구간은 단일 관측값이나 다음 무작위 선택 표본의 평균 및 표준편차와 관련이 있습니다. 계산 시에는 정규 분포에서 특정 표본이 무작위로 선택되었다고 가정합니다. 단측 또는 양측 예측 구간을 선택합니다.

변수의 **예측 구간** 옵션을 선택하면 "예측 구간" 창이 나타납니다. 이 창을 사용하여 신뢰 수준, 미래 표본 수, 그리고 단측 또는 양측 한계를 지정합니다.

관련 정보

- 통계 상세 정보는 " 예측 구간 "(79 페이지) 에서 확인하십시오 .
- 예는 " 예측 구간의 예 "(69 페이지) 에서 확인하십시오 .

공차 구간

공차 구간에는 모집단의 지정된 비율 이상이 포함됩니다. 이 구간은 평균이나 표준편차가 아니라 모집단의 지정된 비율에 대한 신뢰 구간입니다. 공차 구간에 대한 자세한 내용은 Meeker et al. 연구 자료(2017) 및 Tamhane and Dunlop 연구 자료(2000)에서 확인할 수 있습니다.

변수의 **공차 구간** 옵션을 선택하면 "공차 구간" 창이 나타납니다. 이 창을 사용하여 신뢰 수준, 포함할 비율, 단측 또는 양측 한계, 방법을 지정합니다. 사용할 수 있는 두 가지 방법은 "정규 분포 가정"과 "비모수"입니다. "정규 분포 가정" 옵션은 표본이 정규 분포에서 무작위로 선택되었다는 가정에 기반한 공차 구간을 계산합니다. "비모수" 옵션은 분포 무관 공차 구간을 계산합니다.

관련 정보

- 통계 상세 정보는 " 공차 구간 "(80 페이지) 에서 확인하십시오 .
- 예는 " 공차 구간의 예 "(71 페이지) 에서 확인하십시오 .

공정 능력

공정 능력 분석에서는 지정된 규격 한계와 비교하여 공정이 얼마나 잘 수행되고 있는지를 측정합니다. 양호한 공정은 안정적이고 규격 한계 이내인 제품을 일관되게 생산하는 공정입니다. 공정 능력 지수는 공정 중심화 및 변동성별로 요약된 공정 성능을 규격 한계와 관련짓는 측정값입니다.

규격 한계

열에 "규격 한계" 열 특성이 포함되어 있고 시작 창에서 "공정 능력 생성" 옵션이 선택되었으면 "공정 능력" 보고서가 자동으로 생성됩니다. 열에 "분포" 열 특성도 포함되어 있지 않은 한 이 보고서는 정규 분포를 기반으로 합니다. 열에 "분포" 열 특성이 포함되어 있는 경우 "공정 능력" 보고서는 해당 열 특성에 지정된 분포를 기반으로 합니다.

팁 : 여러 열에 대하여 규격 한계를 한 번에 추가하는 방법은 Quality and Process Methods 의 에서 확인하십시오 .

열에 규격 한계가 포함되어 있지 않으면 분석 변수 이름 옆의 빨간색 삼각형에서 **공정 능력**을 선택하고 "공정 능력 분석" 창에서 규격 한계를 설정합니다.

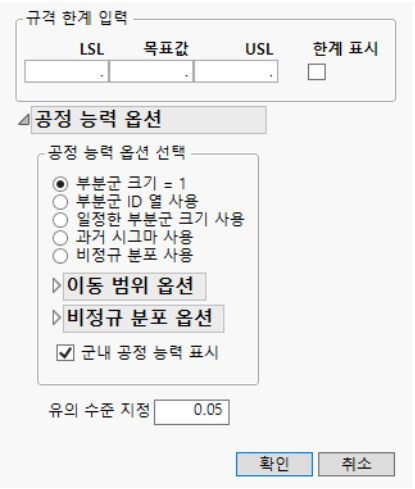
보고서의 규격 한계를 데이터 테이블에 열 특성으로 저장하려면 "공정 능력"의 빨간색 삼각형에서 **규격 한계를 열 특성으로 저장**을 선택합니다. 공정 능력 분석을 반복할 때 저장된 규격 한계를 자동으로 가져오게 됩니다.

공정 능력 분석 창

분석 변수 이름 옆의 빨간색 삼각형에서 "공정 능력" 옵션을 선택하면 "공정 능력 분석" 창이 나타납니다.

"공정 능력 분석" 창을 사용하여 규격 한계, 분석을 위한 기본 분포, 시그마 추정 방법 등의 공정 능력 분석 옵션을 지정합니다. 공정 능력을 분석하려면 시그마 추정 방법과 그룹 내(단기) 변동을 선택해야 합니다. 선택한 공정 능력 옵션에 따라 다른 하위 옵션이 나타납니다.

그림 3.12 공정 능력 분석 창



규격 한계 입력 공정 능력 분석을 위한 규격 하한 , 목표값 및 규격 상한을 지정합니다 . 이 중 적어도 하나는 비결측값이어야 합니다 . " 한계 표시 " 옵션을 선택하면 분포 플랫폼 보고서의 히스토그램에 규격 한계가 나타납니다 .

공정 능력 옵션 선택한 옵션에 따라 다른 추가 옵션이 나타납니다 . 다음 옵션 중 하나를 선택합니다 .

부분군 크기 = 1 부분군 크기를 1 로 설정하고 추가 이동 범위 옵션을 제공합니다 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

부분군 ID 열 사용 부분군 ID 열을 선택하고 추가 부분군 지정 및 이동 범위 옵션을 제공합니다 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

일정한 부분군 크기 사용 일정한 부분군 크기를 설정하고 추가 부분군 지정 및 이동 범위 옵션을 제공합니다 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

과거 시그마 사용 과거에 승인된 시그마 값을 할당합니다 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

비정규 분포 사용 비정규 분포를 선택하고 추가 비정규 분포 옵션을 제공합니다 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

군내 공정 능력 표시 (" 부분군 크기 " 가 1 이거나 " 공정 능력 " 옵션에서 " 비정규 분포 사용 " 을 선택한 경우에만 사용 가능합니다 .) 군내 표준편차의 추정값을 보고서에 표시할지 여부를 지정합니다 .

유의 수준 지정 신뢰 한계의 유의 수준을 지정합니다 .

공정 능력 분석 보고서

" 공정 능력 분석 " 창에서 " 확인 " 을 클릭하면 선택한 변수에 대한 공정 능력 보고서가 포함된 공정 능력 보고서가 나타납니다. 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오.

- 통계 상세 정보는 Quality and Process Methods 의 에서 확인하십시오 .
- 예는 " 공정 능력의 예 "(72 페이지) 에서 확인하십시오 .
- " 공정 능력 " 플랫폼에 대한 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

참고 : 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 공정 능력 보고서의 다양한 옵션에 대한 환경 설정을 지정할 수 있습니다 .

분포 적합

" 연속형 적합 " 또는 " 이산형 적합 " 하위 메뉴의 옵션을 사용하여 연속형 변수에 분포를 적합시킬 수 있습니다 . 연속형 변수에 분포를 적합시키면 히스토그램에 곡선이 중첩 표시되고 보고서 창에 분포 비교 보고서와 적합 분포 보고서가 추가됩니다 . 적합 분포 보고서의 빨간색 삼각형 메뉴에는 추가 옵션이 포함되어 있습니다 . 자세한 내용은 " 분포 적합 옵션 "(62 페이지) 에서 확인하십시오 . 열에 " 분포 " 열 특성이 포함된 경우 분포 보고서에서 해당 열 특성의 분포는 기본값으로 적합됩니다 .

참고 : 수명 분포 플랫폼에도 분포 적합을 위한 옵션이 포함되어 있으며 여기에는 다른 파라미터화 방법이 사용되고 중도절단된 관측값이 허용될 수 있습니다 . 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

연속형 적합

" 연속형 적합 " 하위 메뉴에는 연속형 분포를 적합시키기 위한 옵션이 포함되어 있습니다 . 이러한 분포의 파라미터화 방법에 대한 자세한 내용은 " 연속형 적합 분포 "(82 페이지) 에서 확인하십시오 .

정규 적합 데이터에 정규 분포를 적합시킵니다 . 정규 분포는 대개 대부분의 값이 곡선의 가운데에 있는 대칭적 데이터를 모델링하는 데 사용됩니다 .

Cauchy 적합 데이터에 Cauchy 분포를 적합시킵니다 . Cauchy 분포는 평균과 표준편차가 정의되어 있지 않습니다 . 대부분의 데이터가 본질적으로 Cauchy 분포를 따르기는 하지만 데이터에 포함된 이상치의 비율이 높은 경우 (최대 50%) 데이터의 로버스트 위치 및 척도를 추정하는 데 이 분포가 특히 유용할 수 있습니다 .

스튜던트 t 적합 스튜던트 *t* 분포를 데이터에 적합시킵니다 . 스튜던트 *t* 분포는 정규 분포와 Cauchy 분포 사이의 공간을 차지하는 로버스트 옵션입니다 . 스튜던트 *t* 분포의 자유도가 무한대에 접근하면 해당 분포는 정규 분포와 동등해집니다 . 스튜던트 *t* 분포의 자유도가 1 과 같아지면 해당 분포는 Cauchy 분포와 동등해집니다 . 분포 플랫폼은 자유도 값을 추정합니다 .

SHASH 적합 데이터에 SHASH(SinH-ArcSinH) 분포를 적합시킵니다 . SHASH 분포는 정규 분포로 변환된다는 점에서 Johnson 분포와 유사하지만 SHASH 분포에는 정규 분포가 특수한 경우로 포함되어 있습니다 . 이 분포는 대칭적일 수도 있고 비대칭적일 수도 있습니다 .

지수 적합 (모든 관측값이 음수가 아닌 경우에만 사용 가능) 데이터에 지수 분포를 적합시킵니다 . 지수 분포는 오른쪽으로 편중된 분포로 , 대개 수명이나 연속된 이벤트 사이의 시간을 모델링하는 데 사용됩니다 .

감마 적합 (모든 관측값이 양수인 경우에만 사용 가능) 데이터에 감마 분포를 적합시킵니다 . 감마 분포는 양수 값을 모델링하기 위한 유연한 분포입니다 .

로그 정규 적합 (모든 관측값이 양수인 경우에만 사용 가능) 데이터에 로그 정규 분포를 적합시킵니다 . 로그 정규 분포는 오른쪽으로 편중된 분포로 , 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다 .

Weibull 적합 (모든 관측값이 양수인 경우에만 사용 가능) 데이터에 Weibull 분포를 적합시킵니다 . Weibull 분포는 유연한 분포로 , 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다 .

정규 2 혼합 적합 두 가지 정규 분포의 혼합 분포를 적합시킵니다 . 이 유연한 분포는 이봉 분포 데이터를 적합시킬 수 있습니다 .

정규 3 혼합 적합 세 가지 정규 분포의 혼합 분포를 적합시킵니다 . 이 유연한 분포는 다봉 분포 데이터를 적합시킵니다 .

평활 곡선 적합 비모수 밀도 추정 (커널 밀도 추정) 을 사용하여 평활 곡선을 적합시킵니다 . " 비모수 밀도 " 보고서에 나타나는 슬라이더로 대역폭을 변경하여 평활 정도를 조절할 수 있습니다 .

Johnson 적합 데이터에 Johnson 분포를 적합시킵니다 . 세 가지 유형의 Johnson 분포 (Su, Sb 및 Sl) 중 가장 적절한 분포가 적합되어 보고됩니다 . Johnson 분포 계열은 왜도와 첨도의 가

능한 모든 조합을 지원하므로 데이터 적합 기능에 유용합니다 . Johnson 분포의 선택 프로시저 및 모수 추정에 대한 자세한 내용은 Slifker and Shapiro 연구 자료 (1980) 에서 확인할 수 있습니다 .

베타 적합 (모든 관측값이 0 에서 1 사이인 경우에만 사용 가능) 데이터에 베타 분포를 적합시킵니다 . 베타 분포는 0 에서 1 사이의 데이터를 모델링하는 데 유용하며 , 대개 비율을 모델링하는 데 사용됩니다 .

전체 적합 변수에 사용 가능한 모든 연속형 분포를 적합시킵니다 . " 분포 비교 " 보고서에는 각 적합 분포에 대한 통계량이 포함됩니다 . 기본적으로는 최량 적합 분포가 선택되어 히스토그램에 표시됩니다 . 체크박스를 사용하여 선택한 분포에 대한 적합 보고서를 표시하거나 숨기고 분포 곡선을 중첩 표시할 수 있습니다 . 기본적으로 " 분포 비교 " 목록은 AICc 를 기준으로 오름차순 정렬됩니다 .

팁 : " 분포 " 열에서 분포 이름을 두 번 클릭하면 " 분포 비교 " 목록의 분포를 빠르게 제거할 수 있습니다 . 이렇게 하면 해당하는 적합 분포 보고서도 제거됩니다 .

레거시 적합기 사용 " 레거시 적합기 " 하위 메뉴를 표시하거나 숨깁니다 . JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 옵션을 활성화하면 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 사용할 수 있습니다 . 자세한 내용은 " 레거시 분포 적합기에 대한 상세 정보 "(91 페이지) 에서 확인하십시오 .

이산형 적합

" 이산형 적합 " 하위 메뉴는 모든 데이터 값이 정수일 때 사용할 수 있습니다 . " 이산형 적합 " 하위 메뉴에는 이산형 분포를 적합시키기 위한 옵션이 포함되어 있습니다 . 이러한 분포의 파라미터화 방법에 대한 자세한 내용은 " 이산형 적합 분포 "(88 페이지) 에서 확인하십시오 .

Poisson 적합 데이터에 Poisson 분포를 적합시킵니다 . Poisson 분포는 특정 구간 내의 사건 수를 모델링하는 데 유용하며 대개 개수 데이터로 표현됩니다 .

음이항 적합 데이터에 음이항 분포를 적합시킵니다 . 음이항 분포는 지정된 실패 횟수에 도달하기 전의 성공 횟수를 모델링하는 데 유용합니다 . 또한 음이항 분포는 감마 Poisson 분포와 동등합니다 .

ZI Poisson 적합 (데이터에 0 값이 있는 경우에만 사용 가능) 데이터에 영과잉 Poisson 분포를 적합시킵니다 . 영과잉 Poisson 분포에서는 표준 Poisson 분포에 비해 더 많은 비율의 데이터가 0 값이라고 가정합니다 .

ZI 음이항 적합 (데이터에 0 값이 있는 경우에만 사용 가능) 데이터에 영과잉 음이항 분포를 적합시킵니다 . 영과잉 음이항 분포에서는 표준 음이항 분포에 비해 더 많은 비율의 데이터가 0 값이라고 가정합니다 .

이항 적합 데이터에 이항 분포를 적합시킵니다 . 이항 분포는 n 번의 독립 시도 (단 , 모든 시도의 성공 확률 p 는 고정됨) 시 총 성공 횟수를 모델링하는 데 사용됩니다 . 표본 크기는 모든 관측값에 대해 고정된 표본 크기로 지정할 수도 있고 , 각 행의 표본 크기가 포함된 데이터 테이블 내의 다른 열로 지정할 수도 있습니다 .

참고 : 일정하지 않은 표본 크기를 지정할 경우에는 밀도 곡선, 진단 그림 및 프로파일러를 사용할 수 없습니다 .

베타 이항 적합 데이터에 베타 이항 분포를 적합시킵니다 . 베타 이항 분포는 이항 분포의 과대산포 버전입니다 . 이 분포를 적합시키려면 각 관측값에 대해 표본 크기가 1 보다 커야 합니다 . 표본 크기는 모든 관측값에 대해 고정된 표본 크기로 지정할 수도 있고 , 각 행의 표본 크기가 포함된 데이터 테이블 내의 다른 열로 지정할 수도 있습니다 .

참고 : 일정하지 않은 표본 크기를 지정할 경우에는 밀도 곡선, 진단 그림 및 프로파일러를 사용할 수 없습니다 .

분포 적합 옵션

각 적합 분포 보고서에는 추가 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다 .

밀도 곡선 분포의 추정 모수를 사용하여 히스토그램에 밀도 곡선을 중첩 표시합니다 .

진단 그림 다음 옵션이 포함되어 있습니다 .

QQ 그림 QQ(분위수 - 분위수) 그림을 표시하거나 숨깁니다 . 이 그림은 추정 모수를 사용하여 구한 관측값과 분위수 간의 관계를 보여 줍니다 .

PP 그림 PP(백분위수 - 백분위수) 그림을 표시하거나 숨깁니다 . 이 그림은 추정 모수를 사용하여 구한 경험적 CDF(누적 분포 함수) 와 적합 CDF 간의 관계를 보여 줍니다 .

프로파일러 다음 옵션이 포함되어 있습니다 .

분포 프로파일러 CDF(누적 분포 함수) 의 예측 프로파일러를 표시하거나 숨깁니다 .

분위수 프로파일러 분위수 함수의 예측 프로파일러를 표시하거나 숨깁니다 .

열 저장 다음 옵션이 포함되어 있습니다 .

밀도 계산식 저장 추정 모수 값을 사용하여 계산된 밀도 계산식이 포함된 열을 데이터 테이블에 저장합니다 .

분포 계산식 저장 추정 모수 값을 사용하여 계산된 CDF(누적 분포 함수) 가 포함된 열을 데이터 테이블에 저장합니다 .

시뮬레이션 계산식 저장 추정 모수를 사용하여 시뮬레이션된 값을 생성하는 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 이 열은 " 시뮬레이션 " 유틸리티에서 " 스위치 인 할 열 " 로 사용할 수 있습니다 . 자세한 내용은 " 시뮬레이션 " 장 (313 페이지) 에서 확인하십시오 .

변환 저장 (Johnson 및 SHASH 분포 적합의 경우에만 사용 가능) 변환 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 이 계산식은 적합된 분포를 사용하여 분석 열을 정규 분포로 변환하는 데 사용할 수 있습니다 .

적합도 (Johnson, 평활 곡선 , 정규 혼합 , 이항 또는 베타 음이항 분포에는 사용 불가능) 적합된 분포에 대한 적합도 검정이 포함된 적합도 검정 보고서를 표시하거나 숨깁니다 . 연속형 적합

의 경우에는 Anderson-Darling 검정이 포함됩니다. 검정의 p 값은 Stephens 연구 자료(1974)의 섹션 4.1 에 설명된 절차와 유사하게 모수 붓스트랩을 사용하여 시뮬레이션됩니다 . 이산형 적합의 경우에는 Pearson 카이제곱 검정이 포함됩니다 . 정규 분포의 경우 표본 크기가 2,000 보다 작거나 같고 고정 모수가 없으면 Shapiro-Wilk 정규성 검정도 보고됩니다 .

모수 고정 (Johnson 분포 또는 평활 곡선 적합에는 사용할 수 없습니다 .) 모수를 고정하고 고정되지 않은 모수를 다시 추정할 수 있습니다 . 새 모수를 검정하여 해당 모수가 데이터를 적합시키는지 여부를 확인하는 " 적합성 LR 검정 "(LR: 가능도비) 보고서도 표시됩니다 .

공정 능력 (Cauchy, 스튜던트 t 또는 이산형 분포 적합에는 사용할 수 없습니다 .) 적합된 분포를 사용하여 공정 능력 분석을 생성할 수 있습니다 . 공정 능력 분석에서는 규격 한계를 기준으로 공정이 얼마나 잘 수행되고 있는지를 측정합니다 . 적합 분포의 빨간색 삼각형 메뉴에서 " 공정 능력 " 옵션을 선택하면 다음 옵션이 포함된 창이 나타납니다 .

규격 한계 입력 규격 한계를 수동으로 입력할 수 있습니다 . 적합된 분포를 사용하여 규격 한계를 계산하려면 이 섹션을 비워 두고 " 분위수 규격 한계 계산 옵션 " 아래의 옵션을 사용하십시오 .

분위수 규격 한계 계산 옵션 적합된 분포를 기준으로 규격 한계를 계산할 수 있습니다 . 두 가지 방법을 사용할 수 있습니다 .

첫 번째 방법에서는 적합된 분포의 분위수와 연관된 확률을 입력하여 규격 한계를 계산합니다 .

두 번째 방법에서는 규격 한계를 계산하는 데 사용되는 K 시그마 승수 값을 입력합니다 . 이 방법을 사용할 경우 양측 또는 단측 한계를 생성할 수 있습니다 .

확률 또는 시그마 승수 값을 입력한 후 **규격 한계 계산**을 클릭하여 규격 한계를 계산합니다 . 이러한 한계는 " 규격 한계 입력 " 패널에 입력됩니다 . **확인**을 클릭하여 입력된 한계를 적용하고 공정 능력 보고서를 생성합니다 .

공정 능력 옵션 다음 옵션이 포함되어 있습니다 .

" 이동 범위 옵션 " 개요에는 이동 범위 통계량의 유형을 선택할 수 있는 옵션이 포함되어 있습니다 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

" 비정규 분포 옵션 " 개요에는 비정규 공정 능력 계산에 사용되는 방법을 선택할 수 있는 옵션이 포함되어 있습니다 . 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

" 공정 능력 " 옵션 및 보고서에 대한 자세한 내용은 Quality and Process Methods 의 에서 확인하십시오 .

참고 : 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 공정 능력 보고서의 다양한 옵션에 대한 환경 설정을 지정할 수 있습니다 .

적합 제거 보고서 창에서 분포 적합을 제거합니다 .

연속형 변수에 대한 저장 옵션

"저장" 메뉴의 옵션을 사용하여 연속형 변수에 대한 정보를 저장할 수 있습니다. 각 저장 옵션은 현재 데이터 테이블에 새 열을 생성합니다. 새 열의 이름은 저장 명령 이름에 변수 이름(다음 정의에서 <열 이름>으로 표시)을 추가한 형식으로 지정됩니다(표 3.1).

히스토그램 막대를 결합하기 전과 후처럼 서로 다른 상황에서 동일한 정보를 여러 번 저장하려면 저장 옵션을 반복해서 선택합니다. 저장 옵션을 여러 번 사용하면 열 이름을 고유하게 유지하기 위해 열 이름에 번호가 매겨집니다(name1, name2 등).

표 3.1 저장 옵션 설명

옵션	데이터 테이블에 추가되는 열	설명
수준 수	수준 <열 이름>	각 관측값의 수준 수는 해당 관측값이 포함된 히스토그램 막대에 해당합니다. 히스토그램 막대는 낮은 것부터 높은 것 순서로 1로 시작하는 번호가 매겨집니다. **참고 :** 소스 정보를 유지하기 위해 새 열에 값 라벨을 추가할 수 있지만 값 라벨은 기본적으로 해제되어 있습니다.
수준 중간점	중간점 <열 이름>	각 관측값의 중간점 값은 수준 하한에 수준 너비의 1/2을 더해서 계산됩니다. **참고 :** 소스 정보를 유지하기 위해 새 열에 값 라벨을 추가할 수 있지만 값 라벨은 기본적으로 해제되어 있습니다.
순위	순위화 <열 이름>	해당 열의 각 값에 대한 순위(1부터 시작)를 제공합니다. 중복 반응 값의 경우 데이터 테이블에서의 발생 순서에 따라 연속 순위가 할당됩니다.
순위 평균	평균 순위 <열 이름>	값이 고유한 경우에는 평균 순위와 순위가 동일합니다. 값이 k번 나타나는 경우 평균 순위는 값의 순위 합계를 k로 나눠서 계산됩니다.
확률 점수	확률 <열 이름>	비결측 점수가 N개인 경우 값의 확률 점수는 해당 값의 평균 순위를 $N+1$로 나눠서 계산됩니다. 이 열은 경험적 누적 분포 함수와 유사합니다.
정규 분위수	N-분위수 <열 이름>	정규 분위수를 저장합니다. 자세한 내용은 "정규 분위수 그림"(76 페이지)에서 확인하십시오.

표 3.1 저장 옵션 설명 (계속)

옵션	데이터 테이블에 추가되는 열	설명
표준화	표준화 <열 이름>	표준화된 값을 저장합니다. 자세한 내용은 "표준화된 데이터 저장"(79 페이지) 에서 확인하십시오.
중심화됨	중심화 <열 이름>	0 을 기준으로 중심화한 값을 저장합니다.
로버스트 표준화	로버스트 표준화 <열 이름>	로버스트 평균을 중심으로 중심화되고 로버스트 표준편차를 사용하여 표준화된 반응 값을 포함하는 열을 저장합니다.
로버스트 중심화	로버스트 중심화 <열 이름>	로버스트 평균을 중심으로 중심화된 반응 값을 포함하는 열을 저장합니다.
로그에 스크립트로	(없음)	스크립트를 로그 창에 출력합니다. 분석을 다시 생성하려면 스크립트를 실행합니다.

분포 플랫폼의 추가 예

- " 예 : 여러 히스토그램에서 데이터 선택 "
- " 예 : 기준 변수 사용 "
- " 검정 확률 옵션의 예 "
- " 예측 구간의 예 "
- " 공차 구간의 예 "
- " 공정 능력의 예 "

예 : 여러 히스토그램에서 데이터 선택

이 예에서는 한 번에 여러 히스토그램에서 데이터를 선택하고 강조 표시하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Type 및 Size Co 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

   소규모 회사의 유형 분포를 확인하려고 합니다 .
5. "small" 옆의 막대를 클릭합니다 .

제약 회사보다 컴퓨터 회사에 소규모 회사가 더 많음을 확인할 수 있습니다 . 선택 범위를 넓히기 위해 중간 규모 회사를 추가합니다 .

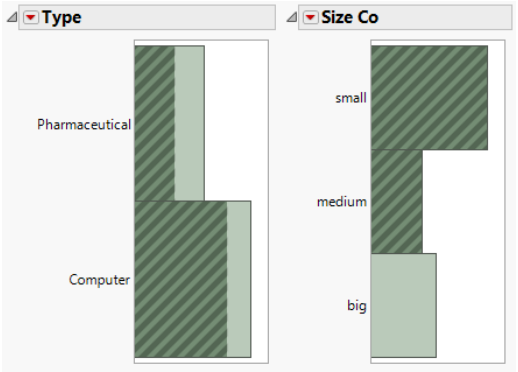
6. Shift 키를 누른 채로 있습니다 . "Size Co" 히스토그램에서 "medium" 옆의 막대를 클릭합니다 .

   소규모 및 중간 규모 회사의 유형 분포를 확인할 수 있습니다 . 그림 3.13 의 왼쪽을 참조하십시오 . 선택 범위를 좁히기 위해 소규모 및 중간 규모의 제약 회사만 확인하려고 합니다 .

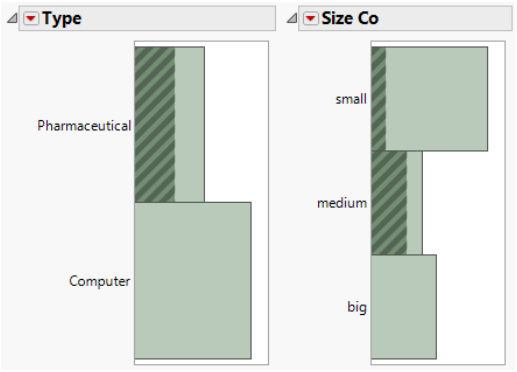
7. Ctrl 및 Shift 키 (Windows 의 경우) 또는 Command 및 Shift 키 (macOS 의 경우) 를 누른 채로 있습니다 . "Type" 히스토그램에서 "Computer" 막대를 클릭하여 선택 취소합니다 .

   소규모 및 중간 규모 회사 중 제약 회사의 수를 확인할 수 있습니다 . 그림 3.13 의 오른쪽을 참조하십시오 .

그림 3.13 여러 히스토그램에서 데이터 선택



Shift 키를 사용하여 선택 범위를 넓힙니다 .



Ctrl+Shift(Windows) 또는 Command+Shift(macOS) 를 사용하여 선택 범위를 좁힙니다 .

예 : 기준 변수 사용

이 예에서는 기준 변수를 사용하여 범주형 변수의 각 수준에 대해 별도의 분석을 생성하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Lipid Data.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Cholesterol 을 선택하고 **Y, 열**을 클릭합니다 .
4. Gender 를 선택하고 **기준**을 클릭합니다 .

   이렇게 하면 Gender 의 각 수준 (female 및 male) 에 대해 개별 분석이 수행됩니다 .

5. **확인**을 클릭합니다 .

히스토그램과 보고서의 방향을 변경합니다.

6. " 분포 " 의 빨간색 삼각형을 클릭하고 **쌓기**를 선택합니다 .

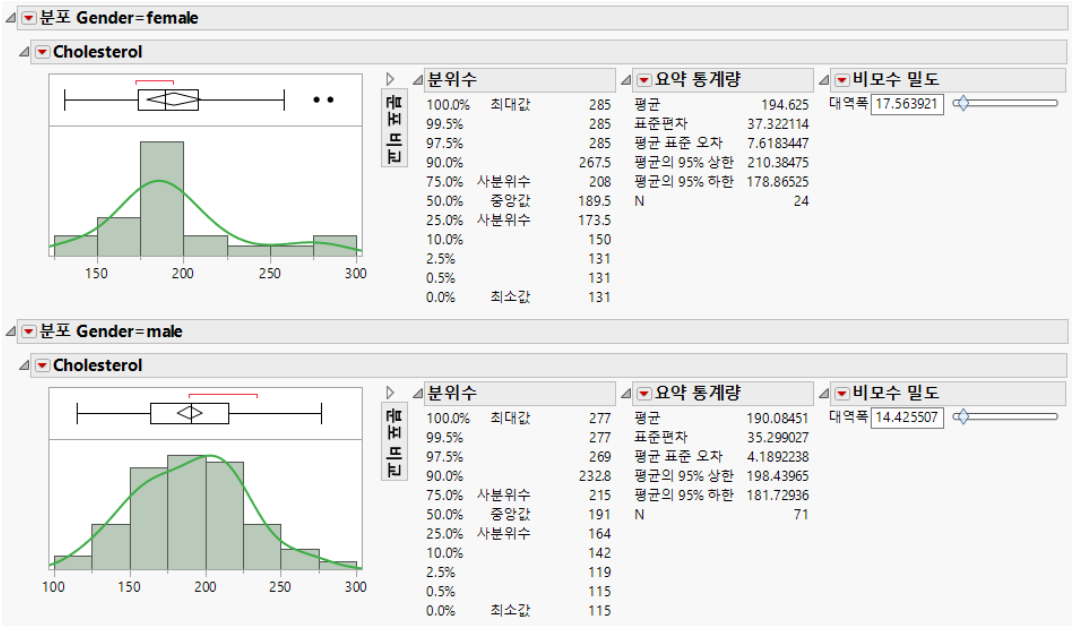
두 히스토그램 모두에 평활 곡선을 추가합니다.

7. Ctrl 키를 누른 채로 있습니다 . "Cholesterol" 의 빨간색 삼각형을 클릭하고 **연속형 적합 > 평활 곡선 적합**을 선택합니다 .

" 분포 비교 " 보고서를 숨깁니다 .

8. Ctrl 키를 누른 채로 있습니다 . **분포 비교** 옆의 회색 표시 아이콘을 클릭합니다 .

그림 3.14 성별별 개별 분포



검정 확률 옵션의 예

이러한 예에서는 수준이 두 개 이상인 변수에 대한 검정 확률 보고서를 생성하는 방법을 보여 줍니다 .

수준이 세 개 이상인 변수에 대한 검정 확률 보고서를 시작합니다.

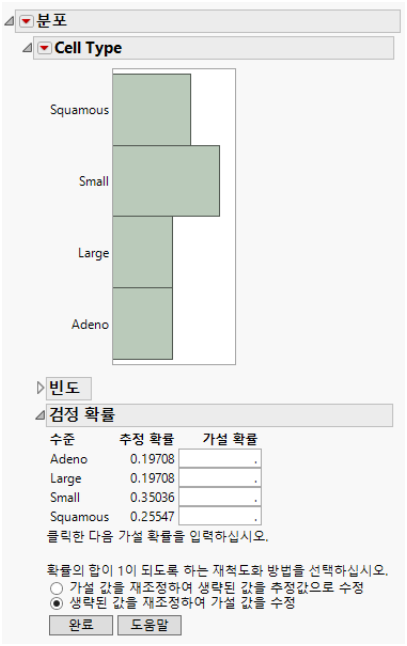
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 VA Lung Cancer.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Cell Type 을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "Cell Type" 의 빨간색 삼각형을 클릭하고 **검정 확률**을 선택합니다 .

   그림 3.15 의 왼쪽을 참조하십시오 .

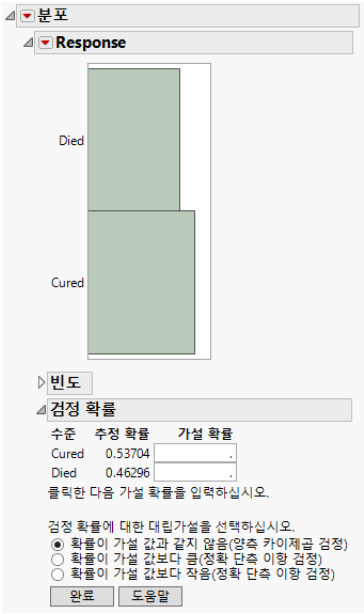
수준이 정확히 두 개인 변수에 대한 검정 확률 보고서를 시작합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penicillin.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Response 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "Response" 의 빨간색 삼각형을 클릭하고 **검정 확률**을 선택합니다 .
   그림 3.15 의 오른쪽을 참조하십시오 .

그림 3.15 검정 확률 옵션의 예



수준이 세 개 이상인 변수에 대한 보고서 옵션



수준이 정확히 두 개인 변수에 대한 보고서 옵션

예 : 검정 확률 보고서 생성

수준이 세 개 이상인 변수에 대한 검정 확률 보고서를 생성하려면 다음을 수행하십시오 .

1. 그림 3.15 의 왼쪽을 참조하십시오 . 네 개의 " 가설 확률 " 필드 모두에 0.25 를 입력합니다 .
2. **생략된 값을 재조정하여 가설 값을 수정** 버튼을 클릭합니다 .
3. **완료**를 클릭합니다 .

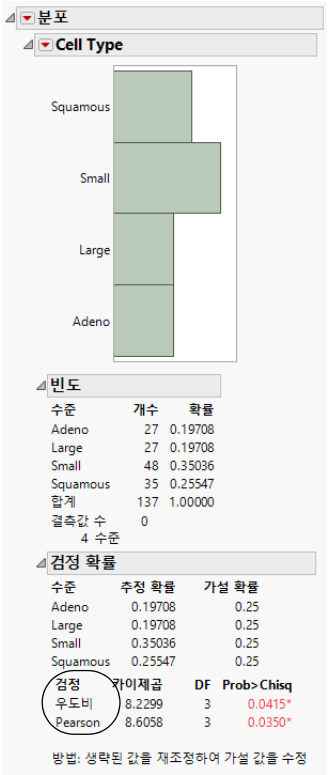
가능도비 및 Pearson 카이제곱 검정이 계산됩니다. 그림 3.16의 왼쪽을 참조하십시오.

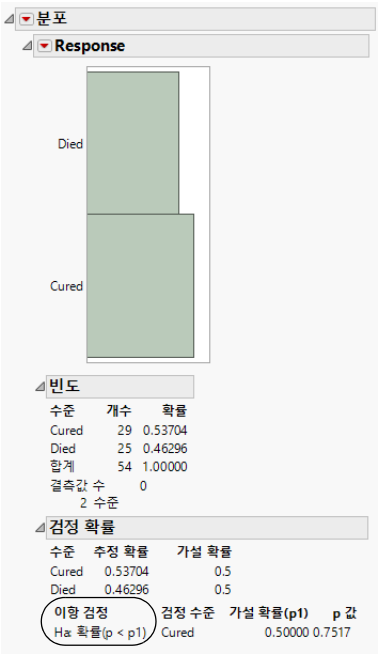
수준이 정확히 두 개인 변수에 대한 검정 확률 보고서를 생성하려면 다음을 수행하십시오.

1. 그림 3.15 의 오른쪽을 참조하십시오 . 두 " 가설 확률 " 필드 모두에 0.5 를 입력합니다 .
2. **확률이 가설 값보다 작음** 버튼을 클릭합니다 .
3. **완료**를 클릭합니다 .

이항 검정의 정확한 확률이 계산됩니다. 그림 3.16의 오른쪽을 참조하십시오.

그림 3.16 검정 확률 보고서의 예



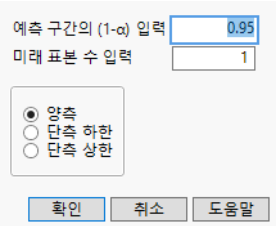


예측 구간의 예

이 예에서는 분포 보고서에 예측 구간을 추가하는 방법을 보여 줍니다 . 다음 10 개의 오존 수준 관측값에 대한 예측 구간을 계산하는 데 관심이 있다고 가정해 보겠습니다 .

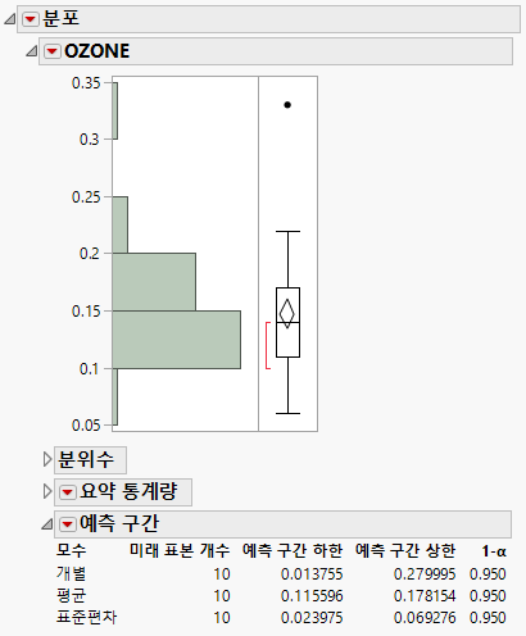
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cities.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. OZONE 을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "OZONE" 의 빨간색 삼각형을 클릭하고 **예측 구간**을 선택합니다 .

그림 3.17 예측 구간 창



6. " 예측 구간 " 창에서 **미래 표본 수 입력** 옆에 10 을 입력합니다 .
7. **확인**을 클릭합니다 .

그림 3.18 예측 구간 보고서의 예



모수	미래 표본 개수	예측 구간 하한	예측 구간 상한	1-α
개별	10	0.013755	0.279995	0.950
평균	10	0.115596	0.178154	0.950
표준편차	10	0.023975	0.069276	0.950

이 예에서는 다음 사항을 95% 신뢰할 수 있습니다 .

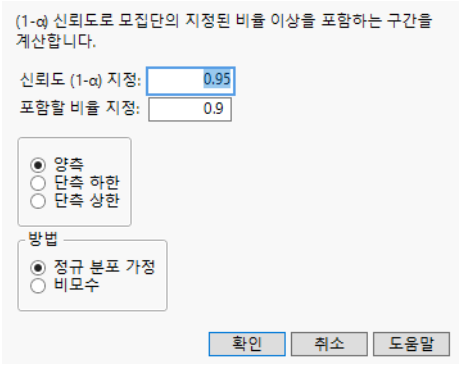
- 다음 10 개의 관측값은 각각 0.013755 에서 0.279995 사이에 있습니다 .
- 다음 10 개 관측값의 평균은 0.115596 에서 0.178154 사이에 있습니다 .
- 다음 10 개 관측값의 표준편차는 0.023975 에서 0.069276 사이에 있습니다 .

공차 구간의 예

이 예에서는 분포 보고서에 공차 구간을 추가하는 방법을 보여 줍니다 . 오존 수준 측정값의 90% 를 포함하는 구간을 추정하려고 한다고 가정해 보겠습니다 .

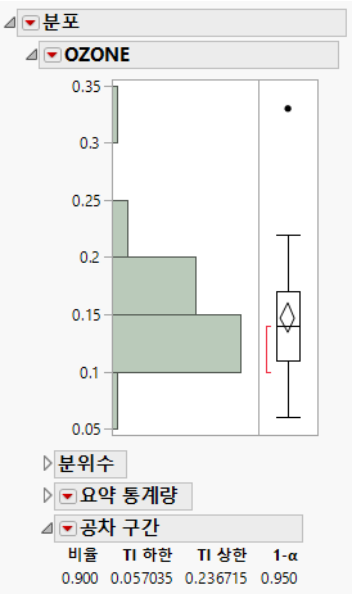
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cities.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. OZONE 을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "OZONE" 의 빨간색 삼각형을 클릭하고 **공차 구간**을 선택합니다 .

그림 3.19 공차 구간 창



6. 기본 선택 사항을 유지하고 **확인**을 클릭합니다 .

그림 3.20 공차 구간 보고서의 예



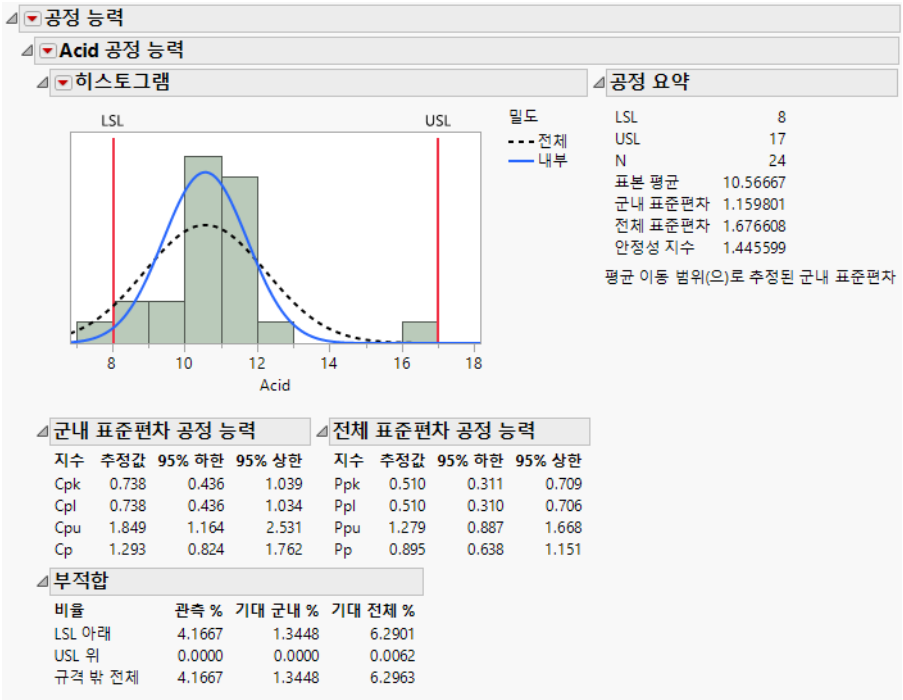
이 예에서는 TI(공차 구간) 하한 및 TI 상한 값에 따라 모집단의 90% 이상이 0.057035 에서 0.236715 사이에 있음을 95% 신뢰할 수 있습니다.

공정 능력의 예

이 예에서는 분포 보고서에 공정 능력 결과를 추가하는 방법을 보여 줍니다. 피클의 산도를 특성화하려고 한다고 가정해 보겠습니다. 규격 하한과 규격 상한은 각각 8 과 17 입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Pickles.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. Acid 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "Acid" 의 빨간색 삼각형을 클릭하고 **공정 능력**을 선택합니다.
6. **LSL**(규격 하한) 로 8 을 입력합니다.
7. **USL**(규격 상한) 로 17 을 입력합니다.
8. **확인**을 클릭합니다.

그림 3.21 공정 능력 보고서의 예



보고서에는 공정 능력 결과가 추가되었습니다. 공정 능력 보고서의 히스토그램에는 규격 한계가 표시되므로 시각적으로 데이터를 이 한계와 비교할 수 있습니다. 그림에서 알 수 있듯이 일부 산도 수준은 규격 하한보다 낮고 일부는 규격 상한에 매우 가깝습니다. Ppk 값은 0.510 이며, 이는 지정된 규격 한계를 기준으로 할 때 공정 능력이 없음을 나타냅니다.

분포 플랫폼에 대한 통계 상세 정보

- " 표준 오차 막대 "
- " 분위수에 대한 통계 상세 정보 "
- " 요약 통계량 "
- " 정규 분위수 그림 "
- "Wilcoxon 부호 순위 검정 "
- " 표준편차 검정 "
- " 정규 분위수 "
- " 표준화된 데이터 저장 "
- " 예측 구간 "
- " 공차 구간 "

- "연속형 적합 분포"
- "이산형 적합 분포"

표준 오차 막대

표준 오차 막대는 표준 오차($\sqrt{np_i(1-p_i)}$)를 사용하여 계산됩니다($p_i=n_i/n$).

분위수에 대한 통계 상세 정보

이 섹션에서는 분위수 계산 방법에 대해 설명합니다.

한 열에 있는 n 개의 비결측값 중 p 번째 분위수를 계산하려면 n 개의 값을 오름차순으로 배열하며 이 열 값을 $y_1, y_2, \ldots, y_n$ 이라고 합니다. p 번째 분위수의 순위 번호는 $p/100(n+1)$ 로 계산합니다.

- 결과가 정수인 경우 p 번째 분위수는 이 순위에 해당하는 값이 됩니다.
- 결과가 정수가 아닌 경우에는 보간법을 사용하여 p 번째 분위수를 구합니다. p 번째 분위수, 즉 q_p 는 다음과 같이 정의됩니다.

  $$q_p = (1-f)y_i + (f)y_{i+1}$$

  다음은 각 요소에 대한 설명입니다.
  - n 은 변수의 비결측값 수입니다.
  - $y_1, y_2, \ldots, y_n$ 은 정렬된 변수 값을 나타냅니다.
  - y_{n+1} 은 y_n 으로 간주됩니다.
  - i 는 $(n+1)p$ 의 정수부이고 f 는 소수부입니다.
  - $(n+1)p = i + f$

예를 들어 15 개의 행이 있는 데이터 테이블에서 연속형 열의 75 번째 및 90 번째 분위수 값을 구하려고 한다고 가정해 보겠습니다. 열을 오름차순으로 배열한 후 이러한 분위수가 포함된 순위를 다음과 같이 계산합니다.

$$\frac{75}{100}(15+1) = 12 \text{ 및 } \frac{90}{100}(15+1) = 14.4$$

값 y_{12} 가 75 번째 분위수입니다. 90 번째 분위수는 14 번째 순위 값과 15 번째 순위 값의 가중 평균을 계산하여 보간됩니다($y_{90} = 0.6y_{14} + 0.4y_{15}$).

요약 통계량

이 섹션에는 " 분포 요약 통계량 " 보고서의 특정 통계량에 대한 통계 상세 정보가 포함되어 있습니다 .

평균

평균은 비결측값의 합을 비결측값의 개수로 나눈 값입니다 . **가중치** 또는 **빈도** 변수를 할당한 경우 평균은 다음과 같이 계산됩니다 .

1. 각 열 값에 해당 가중치 또는 빈도를 곱합니다 .
2. 결과 값을 모두 더한 후 가중치 또는 빈도의 합으로 나눕니다 .

표준편차

표준편차는 평균을 중심으로 한 분포의 퍼짐 정도를 측정한 것입니다 . 종종 s 로 표시하며 , 표본 분산 (s^2 으로 표시) 의 제곱근과 같습니다 .

$$s = \sqrt{s^2}$$

다음은 각 요소에 대한 설명입니다 .

$$s^2 = \sum_{i=1}^{N} \frac{w_i(y_i - \bar{y}_w)^2}{N-1}$$

$\bar{y}_w$ = 가중 평균

평균 표준 오차

평균 표준 오차는 표본 표준편차 s를 N의 제곱근으로 나눠서 계산됩니다. 시작 창에서 가중치 또는 빈도 열을 지정한 경우에는 가중치 또는 빈도 합의 제곱근이 분모가 됩니다.

왜도

왜도는 평균에 대한 3 차 적률을 기준으로 하며 다음과 같이 계산됩니다 .

$$\sum w_i^{\frac{3}{2}} z_i^3 \frac{N}{(N-1)(N-2)} \quad \text{여기서 , } z_i = \frac{x_i - \bar{x}}{s}$$

또한 w_i는 가중치 항 (균일 가중 항목의 경우 1) 입니다.

첨도

첨도는 평균에 대한 3 차 적률을 기준으로 하며 다음과 같이 계산됩니다 .

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n} w_i^2\left(\frac{x_i-\bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

여기서 w_i는 가중치 항(균일 가중 항목의 경우 1)입니다. 이 계산식을 사용할 경우 정규 분포의 첨도는 0입니다. 이 계산식을 종종 초과 첨도라고 합니다.

정규 분위수 그림

각 값의 경험적 누적 확률은 다음과 같이 계산됩니다 .

$$\frac{r_i}{N+1}$$

여기서 r_i는 i번째 관측값의 순위이고, N은 비결측이면서 제외되지 않은 관측값의 개수입니다.

정규 분위수 값은 다음과 같이 계산됩니다 .

$$\Phi^{-1}\left(\frac{r_i}{N+1}\right)$$

여기서 Φ는 정규 분포의 누적 확률 분포 함수입니다.

이 정규 분위수 값은 정규 분포에 기대되는 순서 통계량에 대한 Van Der Waerden 근사값입니다.

Wilcoxon 부호 순위 검정

Wilcoxon 부호 순위 검정을 사용하여 단일 모집단의 중앙값을 검정하거나 일반적인 중앙값의 매칭 쌍 데이터를 검정할 수 있습니다. 매칭 쌍의 경우에는 검정 범위를 줄여 중앙값 0에 대한 쌍체 차이로 구성된 단일 모집단을 검정합니다. 검정 시 기본 모집단은 대칭적인 것으로 가정됩니다.

Wilcoxon 검정에서는 동일 값이 허용됩니다. 차이가 0인 경우에는 Pratt이 제안한 방법을 사용하여 검정 통계량이 조정됩니다. 자세한 내용은 Lehmann and D'Abrera 연구 자료(2006), Pratt 연구 자료(1959) 및 Cureton 연구 자료(1967)에서 확인하십시오.

단일 모집단의 중앙값 검정

- 다음과 같이 N 개의 관측값이 있습니다 .
  $X_1, X_2, \ldots, X_N$
- 귀무가설은 다음과 같습니다 .
  H_0: X 의 분포가 m 을 기준으로 대칭적입니다 .

- 관측값과 가설 값 m 사이의 차이는 다음과 같이 계산됩니다 .

  $D_j = X_j - m$

매칭 쌍 데이터를 사용하여 두 모집단 중앙값의 동일 여부 검정

매칭 쌍 데이터에는 특수한 형태의 Wilcoxon 부호 순위 검정이 적용됩니다 .

- 다음과 같이 두 개의 모집단에서 얻은 N 쌍의 관측값이 있습니다 .

  $X_1, X_2, ..., X_N$ 및 $Y_1, Y_2, ..., Y_N$
- 귀무가설은 다음과 같습니다 .

  H_0: $X - Y$ 의 분포가 0 을 기준으로 대칭적입니다 .
- 관측값 쌍 사이의 차이는 다음과 같이 계산됩니다 .

  $D_j = X_j - Y_j$

Wilcoxon 부호 순위 검정 통계량

이 검정 통계량은 부호 순위의 합을 기준으로 합니다 . 부호 순위는 다음과 같이 정의됩니다 .

- 차이 절대값 $|D_j|$ 는 작은 값부터 큰 값의 순서로 순위화됩니다 .
- 차이가 0 인 경우가 있더라도 순위는 값 1 부터 시작합니다 .
- 동일한 절대 차이가 있는 경우에는 해당 관측값 순위의 평균 또는 **중간 순위**가 할당됩니다 .

차이 D_j 의 순위 또는 중간 순위는 R_j 로 나타냅니다 . D_j 의 부호 순위는 다음과 같이 정의합니다 .

- 차이 D_j 가 양수이면 부호 순위는 R_j 입니다 .
- 차이 D_j 가 0 이면 부호 순위는 0 입니다 .
- 차이 D_j 가 음수이면 부호 순위는 $-R_j$ 입니다 .

부호 순위 통계량은 다음과 같이 계산됩니다 .

$$S = \frac{1}{2} \sum_{j=1}^{N} \text{부호 순위}$$

다음을 정의합니다 .

d_0 - 0 과 동일한 부호 순위의 개수입니다 .
R^+ - 양수 부호 순위의 합입니다 .

그런 후 다음을 구합니다 .

$$S = R^+ - \frac{1}{4}[N(N+1) - d_0(d_0+1)]$$

Wilcoxon 부호 순위 검정 p 값

$N \leq 20$인 경우 정확한 p 값이 계산됩니다.

$N > 20$인 경우 아래에 정의된 것과 같이 이 통계량에 대한 스튜던트 t 근사가 사용됩니다. 이때 동일 값에 대해서는 수정이 적용됩니다. 자세한 내용은 Iman 연구 자료(1974) 및 Lehmann and D'Abrera 연구 자료(2006)에서 확인하십시오.

귀무가설하에서 S 의 평균은 0 입니다 . S 의 분산은 다음과 같이 구합니다 .

$$Var(S) = \frac{1}{24}\left[N(N+1)(2N+1) - d_0(d_0+1)(2d_0+1) - \frac{1}{2}\sum_{i>0} d_i(d_i+1)(d_i-1)\right]$$

Var(S) 표현식의 마지막 합은 동일 값을 수정한 것입니다 . i > 0 인 경우에 대한 표기인 d_i 는 0 이 아닌 부호 순위의 i 번째 그룹에 있는 값의 개수를 나타냅니다 . 특정 부호 순위에 대해 동일 값이 없으면 $d_i = 1$ 이고 합은 0 입니다 .

다음 계산식으로 구한 통계량 t 는 자유도가 N - 1 인 근사 t 분포를 갖습니다 .

$$t = \frac{S}{\sqrt{\frac{N \cdot Var(S) - S^2}{N-1}}}$$

표준편차 검정

다음은 이 검정 통계량을 계산하기 위한 계산식입니다 .

$$\frac{(n-1)s^2}{\sigma^2}$$

이 검정 통계량은 모집단이 정규 데이터일 때 자유도가 n - 1 인 카이제곱 변수로 분포됩니다.

최소 p 값은 양쪽 꼬리 검정의 p 값으로 , 다음과 같이 계산됩니다 .

2 * min(p1,p2)

여기서 p1 은 아래쪽 꼬리의 p 값이고 p2 는 위쪽 꼬리의 p 값입니다.

정규 분위수

정규 분위수 값은 다음과 같이 계산됩니다 .

$$\Phi^{-1}\left(\frac{r_i}{N+1}\right)$$

다음은 각 요소에 대한 설명입니다 .

Φ - 정규 분포의 누적 확률 분포 함수입니다 .

r_i - i 번째 관측값의 순위입니다 .
N - 비결측 관측값의 개수입니다 .

표준화된 데이터 저장

표준화된 값은 다음 계산식을 사용하여 계산됩니다 .

$$\frac{X-\overline{X}}{S_X}$$

다음은 각 요소에 대한 설명입니다 .

X - 원래 열입니다 .
$\overline{X}$ - X 열의 평균입니다 .
S_X - X 열의 표준편차입니다 .

예측 구간

예측 구간은 다음과 같이 정의됩니다 .

- 미래 관측값이 m 개인 경우 :

$$[\underset{\sim}{y}_m, \tilde{y_m}] = \overline{X} \pm t_{(1-\alpha/2m;n-1)} \times \sqrt{1+\frac{1}{n}} \times s \quad (\text{단 , } m \geq 1)$$

- 미래 관측값 m 개의 평균을 사용하는 경우 :

$$[Y_l, Y_u] = \overline{X} \pm t_{(1-\alpha/2,\, n-1)} \times \sqrt{\frac{1}{m}+\frac{1}{n}} \times s \quad (\text{단 , } m \geq 1)$$

- 미래 관측값 m 개의 표준편차를 사용하는 경우 :

$$[s_l, s_u] = \left[s \times \sqrt{\frac{1}{F_{(1-\alpha/2;(n-1,\, m-1))}}},\; s \times \sqrt{F_{(1-\alpha/2;(m-1,\, n-1))}} \right] \quad (\text{단 , } m \geq 2)$$

여기서 m 은 미래 관측값의 개수이고 , n 은 현재 분석 표본에 포함된 점의 개수입니다 .

- 단측 구간은 분위수 함수에서 1-α 를 사용하여 구합니다 .

자세한 내용은 Meeker et al. 연구 자료 (2017, ch. 4) 에서 확인하십시오 .

공차 구간

이 섹션에는 단측 및 양측 공차 구간에 대한 통계 상세 정보가 포함되어 있습니다 .

정규 분포 기반의 구간

단측 구간

단측 구간은 다음과 같이 계산됩니다 .

하한 = $\bar{x} - g's$

상한 = $\bar{x} + g's$

다음은 각 요소에 대한 설명입니다 .

$$g' = t(1-\alpha, n-1, \Phi^{-1}(p)\cdot\sqrt{n})/\sqrt{n}$$

s - 표준편차입니다 .

t - 비중심 t 분포의 분위수입니다 .

Φ^{-1} - 표준 정규 분위수입니다 .

양측 구간

양측 구간은 다음과 같이 계산됩니다 .

$$[T_{p_L}, T_{p_U}] = [\bar{x} - g_{(1-\alpha/2;p,n)}s, \bar{x} + g_{(1-\alpha/2;p,n)}s]$$

여기서 s 는 표준편차이고 , $g_{(1-\alpha/2;p,n)}$ 은 상수입니다 .

g 를 확인하려면 공차 구간으로 구한 모집단의 비율을 고려합니다 . Tamhane and Dunlop 연구 자료 (2000) 에서는 이 비율을 다음과 같이 정의합니다 .

$$\Phi\left(\frac{\bar{x}+gs-\mu}{\sigma}\right) - \Phi\left(\frac{\bar{x}-gs-\mu}{\sigma}\right)$$

여기서 Φ 는 표준 정규 CDF(누적 분포 함수) 입니다 .

따라서 g 는 다음 방정식을 해결합니다 .

$$P\left\{\Phi\left(\frac{\bar{X}+gs-\mu}{\sigma}\right) - \Phi\left(\frac{\bar{X}-gs-\mu}{\sigma}\right) \geq 1-\gamma\right\} = 1-\alpha$$

여기서 1 - γ 는 공차 구간에 포함된 모든 미래 관측값의 비율입니다 .

정규 분포 기반의 공차 구간에 대한 자세한 내용은 Meeker et al. 연구 자료 (2017) 의 표 J.1a, J.1b, J.6a 및 J.6b 에서 확인하십시오 .

비모수 구간

단측 하한

크기가 n 인 표본에서 비율 β 이상의 표본 분포를 포함하기 위한 100(1 - α)% 단측 공차 하한은 순서 통계량 $x_{(l)}$ 입니다 . 지수 l 은 다음과 같이 계산됩니다 .

$$l = n - \Phi_{bin}^{-1}(1 - \alpha, n, \beta)$$

여기서 $\Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포의 (1 - α) 번째 분위수입니다 .

실제 신뢰 수준은 $\Phi_{bin}(n\text{-}l, n, \beta)$로 계산되며, 여기서 $\Phi_{bin}(x, n, \beta)$는 시행 횟수가 n이고 성공 확률이 β인 이항 분포 확률 변수가 x보다 작거나 같을 확률입니다.

단측 분포 무관 공차 하한을 계산하려면 표본 크기 n이 $(\log\alpha)/(\log\beta)$ 이상이어야 합니다.

단측 상한

크기가 n 인 표본에서 비율 β 이상의 표본 분포를 포함하기 위한 100(1 - α)% 단측 공차 상한은 순서 통계량 $x_{(u)}$ 입니다 . 지수 u 는 다음과 같이 계산됩니다 .

$$u = 1 + \Phi_{bin}^{-1}(1 - \alpha, n, \beta)$$

여기서 $\Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포의 (1 - α) 번째 분위수입니다 .

실제 신뢰 수준은 $\Phi_{bin}(u\text{-}1, n, \beta)$로 계산되며, 여기서 $\Phi_{bin}(x, n, \beta)$는 시행 횟수가 n이고 성공 확률이 β 인 이항 분포 확률 변수가 x보다 작거나 같을 확률입니다.

단측 분포 무관 공차 상한을 계산하려면 표본 크기 n이 $(\log\alpha)/(\log\beta)$ 이상이어야 합니다.

양측 구간

크기가 n 인 표본에서 비율 β 이상의 표본 분포를 포함하기 위한 100(1 - α)% 양측 공차 구간은 다음과 같이 계산됩니다 .

$$[\tilde{T}_{p_L}, \tilde{T}_{p_U}] = [x_{(l)}, x_{(u)}]$$

여기서 $x_{(i)}$ 는 i 번째 순서 통계량이며 , l 및 u 는 다음과 같이 계산됩니다 .

$\nu = n - \Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$라고 가정합니다. 여기서 $\Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$는 시행 횟수가 n이고 성공 확률이 β 인 이항 분포의 (1 - α) 번째 분위수입니다 . ν 가 2 보다 작으면 양측 분포 무관 공차 구간을 계산할 수 없습니다 . ν 가 2 보다 크거나 같으면 $l = \text{floor}(\nu/2)$ 이고 $u = \text{floor}(n + 1 - \nu/2)$ 입니다 .

실제 신뢰 수준은 $\Phi_{bin}(u\text{-}l\text{-}1, n, \beta)$로 계산되며, 여기서 $\Phi_{bin}(x, n, \beta)$는 시행 횟수가 n이고 성공 확률이 β 인 이항 분포 확률 변수가 x보다 작거나 같을 확률입니다.

양측 분포 무관 공차 구간을 계산하려면 표본 크기 n 이 다음 방정식의 이상이어야 합니다.

$$1 - \alpha = 1 - n\beta^{n-1} + (n-1)\beta^{n}$$

분포 무관 공차 구간에 대한 자세한 내용은 Meeker et al. 연구 자료(2017, sec. 5.3)에서 확인하십시오.

연속형 적합 분포

이 섹션에는 " 연속형 적합 " 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다.

정규 적합

"정규 적합" 옵션은 정규 분포의 두 모수를 추정합니다.

- μ(평균) - x 축에서의 분포 위치를 정의합니다.
- σ(표준편차) - 분포의 산포 또는 퍼짐 정도를 정의합니다.

표준 정규 분포는 $\mu = 0$ 이고 $\sigma = 1$ 일 때 나타납니다.

$$\text{pdf:} \quad \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (\text{단}, \ -\infty < x < \infty, \ -\infty < \mu < \infty, \ 0 < \sigma)$$

$E(x) = \mu$

$Var(x) = \sigma^2$

Cauchy 적합

"Cauchy 적합 " 옵션은 위치가 μ 이고 척도가 σ 인 Cauchy 분포를 적합시킵니다.

$$\text{pdf:} \quad \left\{\pi\sigma\left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]\right\}^{-1} \quad (\text{단}, \ -\infty < x < \infty, \ -\infty < \mu < \infty, \ 0 < \sigma)$$

$E(x)$ = 정의되지 않음

$Var(x)$ = 정의되지 않음

스튜던트 t 적합

" 스튜던트 *t* 적합 " 옵션은 위치가 μ 이고 , 척도가 σ 이고 , 자유도가 ν 인 스튜던트 *t* 분포를 적합시킵니다 .

$$\text{pdf:}\quad \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\nu\pi\sigma^2}}\left[1+\frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-\left(\frac{\nu+1}{2}\right)} \qquad (\text{단 }, -\infty < x < \infty,\ -\infty < \mu < \infty,\ 0 < \sigma,\ 1 \le \nu)$$

SHASH 적합

"SHASH 적합 " 옵션은 SHASH(SinH-ArcSinH) 분포를 적합시킵니다 . SHASH 분포는 정규 분포의 변환을 기반으로 하며 정규 분포를 특수한 경우로 포함합니다 . 이 분포는 대칭적일 수도 있고 비대칭적일 수도 있습니다 . 형태는 두 개의 형상 모수 γ 및 δ 에 의해 결정됩니다 . SHASH 분포에 대한 자세한 내용은 Jones and Pewsey 연구 자료 (2009) 에서 확인하십시오 .

$$\text{pdf:}\quad f(x) = \frac{\delta\cosh(w)}{\sqrt{\sigma^2+(x-\theta)^2}}\phi[\sinh(w)] \qquad (\text{단 },\ -\infty < \gamma, x, \theta < \infty,\ 0 < \delta, \sigma)$$

다음은 각 요소에 대한 설명입니다 .

$\phi(\cdot)$ - 표준 정규 pdf 입니다 .

$$w = \gamma + \delta\sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)$$

- $\gamma = 0$ 이고 $\delta = 1$ 인 경우 SHASH 분포는 위치가 θ 이고 척도가 σ 인 정규 분포와 동등합니다 .
- 변환 $\sinh(w)$ 는 $\mu = 0$ 이고 $\sigma = 1$ 인 정규 분포를 따릅니다 .

지수 적합

지수 분포는 생존 데이터와 같이 시간 경과에 따라 무작위로 발생하는 사건을 설명하는 데 특히 유용합니다. 지수 분포는 중복되지 않는 사건의 발생 사이 경과 시간을 모델링하는 데도 유용할 수 있습니다. 중복되지 않는 사건의 예로는 사용자의 컴퓨터 쿼리 후 서버에서 응답할 때까지의 시간, 서비스 데스크에서의 고객 내방, 전화 교환대에서의 전화 수신 등이 포함됩니다.

지수 분포는 $\beta = 1$ 이고 $\alpha = \sigma$ 일 경우 특수한 형태의 2 모수 Weibull 분포가 되며 , $\alpha = 1$ 일 경우에는 특수한 형태의 감마 분포가 되기도 합니다 .

$$\text{pdf:}\quad \frac{1}{\sigma}\exp\left(-\frac{x}{\sigma}\right) \qquad (\text{단 },\ 0 < \sigma,\ 0 \le x)$$

$E(x) = \sigma$

$\text{Var}(x) = \sigma^2$

Devore 연구 자료(1995)에 따르면 지수 분포는 **무기억성**을 가집니다. 무기억성은 t 시간 후 여전히 유효한 어떤 성분을 확인할 경우 추가 수명의 분포(t 시간 후까지도 해당 성분이 유지된 경우 추가 수명의 조건부 확률)가 원래 분포와 동일함을 의미합니다.

감마 적합

" 감마 적합 " 옵션은 감마 분포 모수 $\alpha > 0$ 및 $\sigma > 0$ 을 추정합니다 . 적합 감마 보고서에서 모수 α(알파)는 형태 또는 곡률을 설명합니다. 모수 σ(시그마)는 분포의 척도 모수입니다. 데이터는 0 보다 커야 합니다 .

$$\text{pdf:} \quad \frac{1}{\Gamma(\alpha)\sigma^{\alpha}} x^{\alpha-1}\exp(-x/\sigma) \qquad (\text{단}, \ 0 < x, \ 0 < \alpha,\sigma)$$

$$E(x) = \alpha\sigma$$

$$\text{Var}(x) = \alpha\sigma^2$$

- **표준** 감마 분포의 경우 $\sigma = 1$ 입니다 . 시그마는 값이 1 이 아닐 경우 가로 축을 따라 분포를 늘리거나 줄이므로 척도 모수라고 합니다 .
- 카이제곱 $\chi^2_{(\nu)}$ 분포는 $\sigma = 2$ 이고 $\alpha = \nu/2$ 일 때 발생합니다 .
- 지수 분포는 $\alpha = 1$ 일 때 발생합니다 .

표준 감마 밀도 함수는 $\alpha \leq 1$일 때는 감소하기만 합니다. $\alpha > 1$일 때 밀도 함수는 0에서 시작하여 최대값까지 증가하다가 감소합니다.

로그 정규 적합

" 로그 정규 적합 " 옵션은 2 모수 로그 정규 분포의 μ(척도) 및 σ(형상) 모수를 추정합니다 . $X = \ln(Y)$가 정규 분포인 경우에만 Y 변수가 로그 정규 분포로 나타납니다. 데이터는 0보다 커야 합니다 .

$$\text{pdf:} \quad \frac{1}{\sigma\sqrt{2\pi}} \frac{\exp\left[\frac{-(\log(x)-\mu)^2}{2\sigma^2}\right]}{x} \qquad (\text{단}, \ 0 \leq x, \ -\infty < \mu < \infty, \ 0 < \sigma)$$

$$E(x) = \exp(\mu + \sigma^2/2)$$

$$\text{Var}(x) = \exp(2(\mu + \sigma^2)) - \exp(2\mu + \sigma^2)$$

Weibull 적합

Weibull 분포는 α(척도) 및 β(형태) 값에 따라 다른 형태로 나타납니다. 이 함수는 특히 기계 장치 및 생물학 분야에서 수명 길이를 추정하는 데 적절한 모형을 제공하는 경우가 많습니다.

Weibull 분포의 pdf 는 다음과 같이 정의됩니다 .

$$\text{pdf:} \quad \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] \quad (\text{단}, \ \alpha,\beta>0, \ \theta<x)$$

$$E(x) = \alpha\Gamma\left(1+\frac{1}{\beta}\right)$$

$$\text{Var}(x) = \alpha^2\left\{\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right\}$$

여기서 Γ(·) 는 Gamma 함수입니다 .

정규 2 혼합 적합 및 정규 3 혼합 적합

" 정규 2 혼합 적합 " 및 " 정규 3 혼합 적합 " 옵션은 두 개 또는 세 개의 정규 분포를 적합시킵니다 . 이러한 유연한 분포는 이봉 또는 다봉 데이터를 적합시킬 수 있습니다 . 각 그룹에 대해 평균 , 표준편차 및 전체 대비 비율이 개별적으로 추정됩니다 . 다음 방정식에서 *k* 는 혼합되는 정규 분포의 수와 같습니다 .

$$\text{pdf:} \quad \sum_{i=1}^{k}\frac{\pi_i}{\sigma_i}\phi\left(\frac{x-\mu_i}{\sigma_i}\right)$$

$$E(x) = \sum_{i=1}^{k}\pi_i\mu_i$$

$$\text{Var}(x) = \sum_{i=1}^{k}\pi_i(\mu_i^2+\sigma_i^2)-\left(\sum_{i=1}^{k}\pi_i\mu_i\right)^2$$

여기서 μ_i, σ_i 및 π_i 는 각각 *i* 번째 그룹의 평균 , 표준편차 및 비율이며 , ϕ(·) 는 표준 정규 pdf 입니다 .

Johnson 적합

"Johnson 적합 " 옵션은 Johnson 분포 체계에서 최적의 적합 분포를 선택하여 적합시키며 , Johnson 분포 체계에 포함된 세 개의 분포는 모두 하나의 변환된 정규 분포를 기반으로 합니다 . 이 세 가지 분포는 다음과 같습니다 .

- Johnson Su - 경계가 없습니다 .
- Johnson Sb - 양쪽 꼬리에 경계가 있습니다 . 이 경계는 추정 가능한 모수에 의해 정의됩니다 .
- Johnson Sl - 한쪽 꼬리에 경계가 있습니다 . 이 경계는 추정 가능한 모수에 의해 정의됩니다 . Johnson Sl 계열에는 로그 정규 분포 계열이 포함됩니다 .

선택된 분포에 대한 적합만 보고됩니다. Johnson 분포의 선택 프로시저 및 모수 추정에 대한 자세한 내용은 Slifker and Shapiro 연구 자료(1980)에서 확인할 수 있습니다. 모수 추정에는 최대 가능도가 사용되지 않습니다.

Johnson 분포는 유연하기 때문에 널리 사용됩니다. 특히 Johnson 분포 체계는 왜도와 첨도의 가능한 모든 조합을 지원하므로 데이터 적합 기능 면에서 주목됩니다. 하지만 SHASH 분포 또한 매우 유연하므로 Johnson 분포보다 권장됩니다.

Z 가 표준 정규 변량인 경우 이 분포 체계는 다음과 같이 정의됩니다 .

$$Z = \gamma + \delta f(Y)$$

다음은 Johnson Su 의 경우에 해당하는 각 요소에 대한 설명입니다 .

$$f(Y) = \ln\left(Y + \sqrt{1 + Y^2}\right) = \sinh^{-1} Y$$

$$Y = \frac{X - \theta}{\sigma} \qquad -\infty < X < \infty$$

다음은 Johnson Sb 의 경우에 해당하는 각 요소에 대한 설명입니다 .

$$f(Y) = \ln\left(\frac{Y}{1 - Y}\right)$$

$$Y = \frac{X - \theta}{\sigma} \qquad \theta < X < \theta + \sigma$$

다음은 Johnson Sl 의 경우 (단 , $\sigma = \pm 1$) 에 해당하는 각 요소에 대한 설명입니다 .

$$f(Y) = \ln(Y)$$

$$Y = \frac{X - \theta}{\sigma} \qquad \begin{array}{ll} \theta < X < \infty & \text{단, } \sigma = 1 \\ -\infty < X < \theta & \text{단, } \sigma = -1 \end{array}$$

Johnson Su

$$\text{pdf:} \quad \frac{\delta}{\sigma}\left[1 + \left(\frac{x - \theta}{\sigma}\right)^2\right]^{-1/2} \phi\left[\gamma + \delta \sinh^{-1}\left(\frac{x - \theta}{\sigma}\right)\right] \quad (\text{단 , } -\infty < x, \theta, \gamma < \infty, \ 0 < \theta, \delta)$$

Johnson Sb

$$\text{pdf:} \quad \phi\left[\gamma + \delta \ln\left(\frac{x - \theta}{\sigma - (x - \theta)}\right)\right]\left(\frac{\delta\sigma}{(x - \theta)(\sigma - (x - \theta))}\right) \quad (\text{단 , } \theta < x < \theta + \sigma, \ 0 < \sigma)$$

Johnson Sl

$$\text{pdf:} \quad \frac{\delta}{|x - \theta|}\phi\left[\gamma + \delta \ln\left(\frac{x - \theta}{\sigma}\right)\right] \quad (\text{단 , } \sigma = 1 \text{ 일 경우 } \theta < x, \ \sigma = -1 \text{ 일 경우 } \theta > x)$$

여기서 $\phi(\cdot)$ 는 표준 정규 pdf 입니다 .

베타 적합

베타 분포는 간격 0,1 사이에 들어가도록 제한된 확률 변수의 동작을 모델링하는 데 유용합니다. 예를 들어 비율은 항상 0에서 1 사이입니다. "베타 적합" 옵션은 두 개의 형상 모수, 즉 $\alpha > 0$ 및 $\beta > 0$을 추정합니다. 베타 분포에서는 값이 0,1 구간에만 있습니다.

$$\text{pdf:} \quad \frac{1}{B(\alpha, \beta)\sigma^{\alpha+\beta-1}} x^{\alpha-1} x^{\beta-1} \quad (\text{단}, \ 0 < x < 1, \ 0 < \sigma, \alpha, \beta)$$

$$E(x) = \sigma \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(x) = \frac{\sigma^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

여기서 B(·) 는 베타 함수입니다 .

전체 적합

" 분포 비교 " 보고서에서는 분포 목록이 AICc 를 기준으로 오름차순 정렬됩니다 . 체크박스를 사용하여 선택한 분포에 대한 적합 보고서를 표시하거나 숨기고 분포 곡선을 중첩 표시할 수 있습니다 .

AICc 와 BIC 의 계산식은 다음과 같이 정의됩니다 .

$$\text{AICc} = -2\text{logL} + 2k + \frac{2k(k+1)}{n-(k+1)}$$

$$\text{BIC} = -2\text{logL} + k\ln(n)$$

다음은 각 요소에 대한 설명입니다 .

- logL 은 로그 가능도입니다 .
- n 은 표본 크기입니다 .
- k 는 모수의 개수입니다 .

"AICc 가중치" 열은 합계가 1이 되는 정규화된 AICc 값을 보여 줍니다. AICc 가중치는 적합된 분포 중 하나가 true일 때 특정 분포가 true 분포일 확률로 해석됩니다. 따라서 AICc 가중치가 1과 가장 가까운 분포가 더 적절한 적합입니다. AICc 가중치는 비결측 AICc 값만 사용해서 계산됩니다.

AICcWeight = exp[-0.5(AICc-min(AICc))] / sum(exp[-0.5(AICc-min(AICc))])

여기서 min(AICc) 는 적합된 분포 중에서 가장 작은 AICc 값입니다 .

"분포 비교" 보고서의 측도에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

이산형 적합 분포

이 섹션에는 " 이산형 적합 " 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다 .

Poisson 적합

Poisson 분포는 단일 척도 모수 $\lambda > 0$ 을 갖습니다 .

pmf: $\frac{e^{-\lambda}\lambda^{x}}{x!}$ (단 , $0 \le \lambda < \infty$, $x = 0,1,2,...$)

$E(x) = \lambda$

$Var(x) = \lambda$

Poisson 분포는 이산형 분포이므로 해당 중첩 곡선은 각 정수에서 도약이 나타나는 단계 함수입니다.

음이항 적합

음이항 분포는 지정된 실패 횟수에 도달하기 전의 성공 횟수를 모델링하는 데 유용합니다. 다음 파라미터화 계산식에는 평균 모수 λ와 산포 모수 σ가 포함되어 있습니다.

pmf: $\frac{\Gamma[x + (1/\sigma)]}{\Gamma[x+1]\Gamma[1/\sigma]}\left[\frac{(\lambda\sigma)^{x}}{(1+\lambda\sigma)^{x+(1/\sigma)}}\right]$, $x = 0, 1, 2, \ldots$

$E(x) = \lambda$

$Var(x) = \lambda + \sigma\lambda^{2}$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

음이항 분포와 감마 Poisson 분포 간의 관계

음이항 분포는 감마 Poisson 분포와 동등합니다. 감마 Poisson 분포는 데이터가 몇 가지 Poisson(μ) 분포의 조합이고 각 Poisson(μ) 분포의 μ이 다른 경우에 유용합니다.

감마 Poisson 분포는 $x|\mu$가 Poisson 분포를 따르고 μ가 Gamma(α,τ) 분포를 따른다고 가정한 결과입니다. 감마 Poisson에는 $\lambda = \alpha\tau$ 및 $\sigma = \tau+1$ 모수가 있습니다. σ 모수는 산포 모수입니다. $\sigma > 1$이면 과대산포가 나타납니다. 즉, x에서 Poisson만으로 설명할 경우보다 더 많은 변동이 나타납니다. $\sigma = 1$이면 x가 Poisson(λ)으로 축소됩니다.

pmf: $\frac{\Gamma\left(x + \frac{\lambda}{\sigma-1}\right)}{\Gamma(x+1)\Gamma\left(\frac{\lambda}{\sigma-1}\right)}\left(\frac{\sigma-1}{\sigma}\right)^{x}\sigma^{-\frac{\lambda}{\sigma-1}}$ (단 , $0 < \lambda$, $1 \le \sigma$, $x = 0,1,2,...$)

$E(x) = \lambda$

$\text{Var}(x) = \lambda\sigma$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

감마 Poisson 분포는 $\sigma_{negbin} = (\sigma_{gp} - 1) / \lambda_{gp}$ 인 음이항 분포와 동등합니다 .

모수 λ 및 σ 를 갖는 감마 Poisson 분포를 모수 λ 를 갖는 Poisson 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoGammaPoisson.jsl 을 실행하십시오 .

ZI Poisson 적합

ZI(영과잉) Poisson 분포는 척도 모수 $\lambda > 0$ 과 영과잉 모수 π 를 갖습니다 .

$$\text{pmf:} \quad \begin{cases} \pi + (1-\pi)\exp[-\lambda], \text{ 단, } x = 0 \\ (1-\pi)\dfrac{\lambda^x}{x!}\exp[-\lambda], \text{ for } x = 1, 2, \ldots \end{cases}$$

$E(x) = (1 - \pi)\lambda$

$\text{Var}(x) = \lambda(1 - \pi)(1 + \lambda\pi)$

ZI 음이항 적합

ZI(영과잉) 음이항 분포는 척도 모수 $\lambda > 0$, 산포 모수 $\sigma > 0$ 및 영과잉 모수 π 를 갖습니다 .

$$\text{pmf:} \quad \begin{cases} \pi + (1-\pi)(1+\lambda\sigma)^{-(1/\sigma)}, \text{ for } x = 0 \\ (1-\pi)\dfrac{\Gamma[x + (1/\sigma)]}{\Gamma[x+1]\Gamma[1/\sigma]}\left[\dfrac{(\lambda\sigma)^x}{(1+\lambda\sigma)^{x+(1/\sigma)}}\right], \text{ for } x = 1, 2, \ldots \end{cases}$$

$E(x) = (1 - \pi)\lambda$

$\text{Var}(x) = \lambda(1 - \pi)[1 + \lambda(\sigma + \pi)]$

이항 적합

" 이항 적합 " 옵션에서는 두 가지 형식의 데이터 , 즉 표본 크기 상수나 표본 크기를 포함하는 열이 허용됩니다 .

$$\text{pmf:} \quad \binom{n}{x} p^x (1-p)^{n-x} \quad (\text{ 단 , } 0 \le p \le 1,\ x = 0,1,2,\ldots,n)$$

$E(x) = np$

$\text{Var}(x) = np(1-p)$

여기서 n 은 독립 시행 횟수입니다 .

참고 : 이항 분포 모수의 신뢰 구간은 점수 구간입니다. 자세한 내용은 Agresti and Coull 연구 자료 (1998) 에서 확인하십시오 .

베타 이항 적합

베타 이항 분포는 데이터가 몇 가지 Binomial(p) 분포의 조합이고 각 Binomial(p) 분포의 p 가 다른 경우에 유용합니다. 한 예로, 제조 라인 간에 평균 결함 수 (p) 가 다른 경우 여러 라인에서 결합된 전체 결함 수를 들 수 있습니다.

베타 이항 분포는 x|π 가 Binomial(n,π) 분포를 따르고 π 가 Beta(α,β) 를 따른다고 가정한 결과입니다 . 베타 이항 분포는 모수 p = α/(α+β) 및 δ = 1/(α+β+1) 를 갖습니다 . 모수 δ 는 산포 모수입니다 . δ > 1 이면 과대산포가 나타납니다 . 즉 , x 에서 이항 분포만으로 설명할 경우보다 더 많은 변동이 나타납니다 . δ < 0 이면 과소산포가 나타납니다 . δ = 0 이면 x 가 Binomial(n,p) 분포를 따릅니다 . 베타 이항 분포는 $n \geq 2$ 일 때만 존재합니다 .

$$\text{pmf:} \quad \binom{n}{x} \frac{\Gamma\left(\frac{1}{\delta}-1\right)\Gamma\left[x+p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-x+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right]\Gamma\left(n+\frac{1}{\delta}-1\right)}$$

(단 , $0 \leq p \leq 1$, $\max\left(-\frac{p}{n-p-1}, -\frac{1-p}{n-2+p}\right) \leq \delta \leq 1$, $x = 0,1,2,\ldots,n$)

$E(x) = np$

$\text{Var}(x) = np(1-p)[1+(n-1)\delta]$

여기서 Γ(·) 는 Gamma 함수입니다 .

$x|\pi \sim$ Binomial(n,π) 인 반면 π ~ Beta(α,β) 임을 기억하십시오 . 모수 p = α/(α+β) 및 δ = 1/(α+β+1) 는 플랫폼에 의해 추정됩니다 . α 및 β 의 추정값을 구하려면 다음 계산식을 사용합니다 .

$$\hat{\alpha} = \hat{p}\left(\frac{1-\hat{\delta}}{\hat{\delta}}\right)$$

$$\hat{\beta} = (1-\hat{p})\left(\frac{1-\hat{\delta}}{\hat{\delta}}\right)$$

δ 의 추정값이 0 인 경우에는 이 계산식이 적용되지 않습니다. 이 경우에는 베타 이항 분포가 Binomial(n,p) 로 축소된 것이며 $\hat{p}$ 는 p 의 추정값입니다.

베타 이항 분포 모수의 신뢰 구간은 프로파일 가능도 구간입니다.

산포 모수 δ 를 갖는 베타 이항 분포를 모수 p 및 n = 20 을 갖는 이항 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoBetaBinomial.jsl 을 실행하십시오.

레거시 분포 적합기에 대한 상세 정보

JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

- " 분포 적합 옵션 (레거시)"
- " 연속형 적합 분포에 대한 통계 상세 정보 (레거시)"
- " 이산형 적합 분포에 대한 통계 상세 정보 (레거시)"
- " 적합 분위수에 대한 통계 상세 정보 (레거시)"
- " 분포 적합 옵션에 대한 통계 상세 정보 (레거시)"

분포 적합 옵션 (레거시)

연속형 변수에 분포를 적합시키려면 연속형 적합 또는 이산형 적합 옵션을 사용합니다 .

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

그러면 히스토그램에 곡선이 중첩 표시되고 보고서 창에 "모수 추정값" 보고서가 추가됩니다. 빨간색 삼각형 메뉴에는 추가 옵션이 포함되어 있습니다. 자세한 내용은 "분포 적합 옵션 (레거시)"(93 페이지) 에서 확인하십시오.

참고 : 수명 분포 플랫폼에도 분포 적합을 위한 옵션이 포함되어 있으며 여기에는 다른 파라미터화 방법이 사용되고 중도절단이 허용될 수 있습니다 . 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

연속형 적합 (레거시)

이 섹션에서는 " 레거시 적합기 " 하위 메뉴의 분포 중 업데이트된 " 연속형 적합 " 옵션의 해당 분포와는 다른 분포에 대해 설명합니다 .

- Weibull 분포 , 임계가 있는 Weibull 분포 및 극단값 분포는 특히 기계 장치 및 생물학 분야에서 수명 길이를 추정하는 데 적절한 모형을 제공하는 경우가 많습니다 .
- 감마 분포는 0 으로 제한되며 형태가 유연합니다 .
- 베타 분포는 간격 0,1 사이에 들어가도록 제한된 확률 변수의 동작을 모델링하는 데 유용합니다 . 예를 들어 비율은 항상 0 에서 1 사이입니다 .
- 평활 곡선 분포는 비모수 밀도 추정 (커널 밀도 추정) 을 사용하여 평활 곡선을 적합시킵니다 . 평활 곡선은 히스토그램에 중첩 표시되며 그림 아래에는 슬라이더가 나타납니다 . 슬라

이더로 커널 표준편차를 변경하여 평활 정도를 조절할 수 있습니다. 초기 **커널 표준편차** 추정값은 데이터의 표준편차를 사용하여 계산됩니다.

- Johnson Su, Johnson Sb 및 Johnson Sl 분포는 왜도와 첨도의 가능한 모든 조합을 지원하므로 데이터 적합 기능에 유용합니다.
- GLog(일반화 로그) 분포는 생물학적 분석 데이터처럼 거의 정규 분포를 따르지 않으며 대개 분산이 일정하지 않은 데이터를 적합시키는 데 유용합니다.

모든 분포 비교

모두 옵션은 변수에 적용 가능한 모든 연속형 분포를 적합시킵니다. "분포 비교" 보고서에는 각 적합 분포에 대한 통계량이 포함됩니다. 체크박스를 사용하여 선택한 분포에 대한 적합 보고서를 표시하거나 숨기고 분포 곡선을 중첩 표시할 수 있습니다. 기본적으로는 최량 적합 분포가 선택됩니다.

분포 표시 목록은 AICc를 기준으로 오름차순 정렬됩니다.

변수에 음수 값이 포함되었으면 양수 값을 갖는 데이터를 필요로 하는 분포는 분포 표시 목록에 포함되지 않습니다. 이 명령으로는 연속형 분포만 적합됩니다. 베타 분포와 Johnson Sb 분포처럼 임계 모수가 있는 분포는 가능한 분포 목록에 포함되지 않습니다.

관련 정보

통계 상세 정보는 다음 섹션에서 확인하십시오.

- " 연속형 적합 분포에 대한 통계 상세 정보 (레거시)"(95 페이지)
- " 적합 분위수에 대한 통계 상세 정보 (레거시)"(101 페이지)
- " 분포 적합 옵션 (레거시)"(93 페이지)

이산형 적합 (레거시)

" 이산형 적합 " 옵션은 모든 데이터 값이 정수일 때 사용할 수 있습니다. 이산형 변수에 분포 (예 : Poisson 또는 이항 분포) 를 적합시키려면 " 이산형 적합 " 옵션을 사용합니다. 다음 분포를 사용할 수 있습니다.

- Poisson
- 감마 Poisson
- 이항
- 베타 이항

관련 정보

통계 상세 정보는 다음 섹션에서 확인하십시오.

- " 이산형 적합 분포에 대한 통계 상세 정보 (레거시)"(100 페이지)
- " 적합 분위수에 대한 통계 상세 정보 (레거시)"(101 페이지)

- " 분포 적합 옵션 (레거시)"(93 페이지)

분포 적합 옵션 (레거시)

각 적합 분포 보고서에는 추가 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다 .

진단 그림 분위수 그림 또는 확률도를 생성합니다 . 자세한 내용은 " 진단 그림 "(93 페이지) 에서 확인하십시오 .

밀도 곡선 분포의 추정 모수를 사용하여 히스토그램에 밀도 곡선을 중첩 표시합니다 .

적합도 적합 분포의 적합도 검정을 계산합니다 . 자세한 내용은 " 적합도 "(95 페이지) 에서 확인하십시오 .

모수 고정 모수를 고정하고 고정되지 않은 모수를 다시 추정할 수 있습니다 . 새 모수를 검정하여 해당 모수가 데이터를 적합시키는지 여부를 확인하는 " 적합성 LR 검정 "(LR: 가능도비) 보고서도 표시됩니다 .

분위수 지정한 특정 확률 하한 값의 비척도화 및 비중심화 분위수를 반환합니다 .

K 시그마 규격 한계 설정 공정의 규격 한계를 모르는 경우 분포를 기준으로 규격 한계를 설정하려면 이 옵션을 사용합니다 .

일반적으로 규격 한계는 엔지니어링 고려 사항을 사용하여 구합니다 . 엔지니어링 고려 사항이 없고 데이터가 잘 작동하는 공정에서 얻은 것일 경우 적합 분포의 분위수를 사용하면 규격 한계를 설정하는 데 도움이 될 수 있습니다. 자세한 내용은 "K 시그마 규격 한계 설정"(103 페이지) 에서 확인하십시오 .

규격 한계 지정한 규격 한계와 목표값을 기준으로 표준 공정 능력 지수의 일반화 값을 계산합니다 . 자세한 내용은 " 규격 한계 "(95 페이지) 에서 확인하십시오 .

적합 분위수 저장 적합 분위수 값을 데이터 테이블에 새 열로 저장합니다 . 자세한 내용은 " 적합 분위수에 대한 통계 상세 정보 (레거시)"(101 페이지) 에서 확인하십시오 .

밀도 계산식 저장 현재 데이터 테이블에 밀도 계산식으로 계산된 적합 값을 포함하는 새 열을 생성합니다 . 밀도 계산식에서는 추정된 모수 값을 사용합니다 .

규격 한계 저장 규격 한계를 열 특성으로 저장합니다 .

변환 저장 새 열을 생성하고 계산식을 저장합니다 . 이 계산식은 적합 분포를 사용하여 열을 정규 분포로 변환할 수 있습니다 . 이 옵션은 Johnson 분포 , GLog 분포 또는 SHASH 분포 중 하나를 적합시키는 경우에만 사용할 수 있습니다 .

적합 제거 보고서 창에서 분포 적합을 제거합니다 .

진단 그림

진단 그림 옵션은 분위수 그림 또는 확률도를 생성합니다 . 적합 분포에 따라 다음 네 가지 형식 중 하나의 그림이 표시됩니다 .

적합 분위수 대 데이터

- 임계가 있는 Weibull

- 감마
- 베타
- Poisson
- 감마 Poisson
- 이항
- 베타 이항

적합 확률 대 데이터

- 정규
- 정규 혼합
- 지수

적합 확률 대 로그 척도 데이터

- Weibull
- 로그 정규
- 극단값

적합 확률 대 표준 정규 분위수

- SHASH
- Johnson Sl
- Johnson Sb
- Johnson Su
- GLog

진단 그림의 빨간색 삼각형 메뉴에서는 다음과 같은 옵션을 사용할 수 있습니다 .

회전 x 축과 y 축을 서로 바꿉니다 .

신뢰 한계 정규 분위수 그림의 경우 Lilliefors 95% 신뢰 한계를 그리고 , 다른 모든 분위수 그림의 경우 $a = 0.001$ 이고 $b = 0.99$ 인 95% 균등 정밀 띠를 그립니다 (Meeker and Escobar 1998).

적합선 직선으로 된 대각 참조선을 그립니다 . 변수가 선택된 분포에 적합하면 값이 이 참조선 주위에서 나타납니다 .

중앙값 참조선 반응 중앙값에 가로선을 그립니다 .

적합도

적합도 옵션은 적합 분포의 적합도 검정을 계산합니다 . 적합도 검정은 카이제곱 검정이 아니라 EDF(경험적 분포 함수) 검정입니다 . EDF 검정은 카이제곱 검정에 비해 향상된 검정력과 히스토그램 중간점의 불변성을 포함한 장점을 제공합니다 .

- 정규 분포의 경우 표본 크기가 2,000 보다 작거나 같으면 Shapiro-Wilk 정규성 검정이 보고됩니다 . 2,000 보다 큰 표본의 경우에는 KSL 검정이 계산됩니다 .
- 표본 크기가 30 보다 작거나 같은 이산형 분포의 경우 거의 정확한 검정을 구성하기 위해 결합된 두 개의 단측 정확 Kolmogorov 검정을 사용하여 적합도 검정이 구성됩니다 . 자세한 내용은 Conover 연구 자료(1972)에서 확인하십시오 . 표본 크기가 30 보다 큰 경우에는 Pearson 카이제곱 적합도 검정이 수행됩니다 .

관련 정보

- 통계 상세 정보는 " 분포 적합 옵션 (레거시)"(93 페이지) 에서 확인하십시오 .

규격 한계

규격 한계 옵션을 선택하면 규격 한계와 목표값을 입력할 수 있는 창이 열립니다 . 규격 한계와 목표값을 입력하고 나면 일반화된 표준 공정 능력 지수가 계산됩니다 . 정규 분포의 경우 3σ는 하위 0.135 백분위수부터 중앙값 (또는 평균) 까지의 거리이자 중앙값 (또는 평균) 부터 상위 99.865 백분위수까지의 거리입니다 . 이러한 백분위수는 적합 분포에서 추정되며 , 표준 계산식에서 3σ는 적절한 백분위수부터 중앙값까지의 거리로 대체됩니다 .

관련 정보

- 통계 상세 정보는 " 분포 적합 옵션 (레거시)"(93 페이지) 에서 확인하십시오 .

연속형 적합 분포에 대한 통계 상세 정보 (레거시)

이 섹션에는 " 연속형 적합 " 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다 .

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

정규

정규 분포 적합에 대한 자세한 내용은 "정규 적합"(82 페이지) 에서 확인하십시오 .

로그 정규

로그 정규 분포 적합에 대한 자세한 내용은 "로그 정규 적합"(84 페이지) 에서 확인하십시오 .

Weibull, 임계가 있는 Weibull 및 극단값

Weibull 분포는 α(척도) 및 β(형태) 값에 따라 다른 형태로 나타납니다. 이 함수는 특히 기계 장치 및 생물학 분야에서 수명 길이를 추정하는 데 적절한 모형을 제공하는 경우가 많습니다.

Weibull 분포와 임계가 있는 Weibull 분포의 pdf 는 다음과 같이 정의됩니다 .

$$\text{pdf:}\quad \frac{\beta}{\alpha}\left(\frac{x-\theta}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{x-\theta}{\alpha}\right)^{\beta}\right] \quad (\text{단},\ \alpha,\beta>0,\ \theta<x)$$

$$E(x)=\theta+\alpha\Gamma\left(1+\frac{1}{\beta}\right)$$

$$\text{Var}(x)=\alpha^2\left\{\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right\}$$

여기서 Γ(·) 는 Gamma 함수입니다 .

Weibull 옵션은 임계 모수(θ)를 0 으로 설정합니다. **임계가 있는 Weibull** 옵션은 최소 관측값을 사용하여 임계 모수(θ)를 추정하고, 나머지 관측값을 사용하여 α 및 β를 추정합니다. 필요한 임계를 알고 있으면 **모수 고정** 옵션을 사용하여 임계를 설정합니다. 자세한 내용은 "분포 적합 옵션"(62 페이지)에서 확인하십시오.

참고 : 분포 플랫폼에서 임계가 있는 Weibull 분포의 임계 모수에 사용되는 추정 기법은 수명 분포 플랫폼에서 사용되는 추정 기법과는 다릅니다 . 이 분포를 적합시키는 데는 수명 분포에서의 추정 방법을 사용하는 것이 좋습니다 . 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

극단값 분포는 δ = 1 / β 및 λ = ln(α) 로 재파라미터화된 2 모수 Weibull(α, β) 분포와 동등합니다.

지수

지수 분포 적합에 대한 자세한 내용은 "지수 적합"(83 페이지)에서 확인하십시오.

감마

감마 적합 옵션은 감마 분포 모수 α > 0 및 σ > 0 을 추정합니다 . 적합 감마 보고서에서 모수 α(알파) 는 형태 또는 곡률을 설명합니다 . 모수 σ(시그마) 는 분포의 척도 모수입니다 . 세 번째 모수인 θ(임계) 는 하한점 모수입니다 . 음수 값이 없는 한 이 모수는 기본적으로 0 으로 설정되어 있습니다 . **모수 고정** 옵션을 사용하여 이 값을 직접 설정할 수도 있습니다 . 자세한 내용은 " 분포 적합 옵션 "(62 페이지) 에서 확인하십시오 .

$$\text{pdf:}\quad \frac{1}{\Gamma(\alpha)\sigma^{\alpha}}(x-\theta)^{\alpha-1}\exp(-(x-\theta)/\sigma) \quad (\text{단},\ 0\le x,\ 0<\alpha,\sigma)$$

$$E(x)=\alpha\sigma+\theta$$

$\mathrm{Var}(x) = \alpha\sigma^2$

- **표준** 감마 분포의 경우 $\sigma = 1$ 입니다 . 시그마는 값이 1 이 아닐 경우 가로 축을 따라 분포를 늘리거나 줄이므로 척도 모수라고 합니다 .
- 카이제곱 $\chi^2_{(\nu)}$ 분포는 $\sigma = 2$ 이고 $\alpha = \nu/2$ 이며 $\theta = 0$ 일 때 발생합니다 .
- 지수 분포는 $\alpha = 1$ 이고 $\theta = 0$ 일 때 발생하는 감마 곡선에 속합니다 .

표준 감마 밀도 함수는 $\alpha \leq 1$ 일 때는 감소하기만 합니다 . $\alpha > 1$ 일 때 밀도 함수는 0 에서 시작하여 최대값까지 증가하다가 감소합니다 .

베타

표준 베타 분포는 간격 0,1 사이에 들어가도록 제한된 확률 변수의 동작을 모델링하는 데 유용합니다 . 예를 들어 비율은 항상 0 에서 1 사이입니다 . **베타** 적합 옵션은 두 개의 형상 모수 $\alpha > 0$ 및 $\beta > 0$ 과 두 개의 임계 모수 θ 및 σ 를 추정합니다 . 임계 하한은 θ 로 나타내고 임계 상한은 $\theta + \sigma$ 로 나타냅니다 . 베타 분포는 $\theta \leq x \leq (\theta + \sigma)$ 구간에서만 값을 갖습니다 . θ 는 최소값을 사용하여 추정되고 σ 는 범위를 사용하여 추정됩니다 . 표준 베타 분포는 $\theta = 0$ 이고 $\sigma = 1$ 일 때 발생합니다 .

모수 고정 옵션을 사용하면 모수를 고정 값으로 설정할 수 있습니다 . 임계 상한은 최대 데이터 값보다 크거나 같아야 하며 , 임계 하한은 최소 데이터 값보다 작거나 같아야 합니다 . " 모수 고정 " 옵션에 대한 자세한 내용은 " 분포 적합 옵션 "(62 페이지) 에서 확인하십시오 .

$$\text{pdf:} \quad \frac{1}{B(\alpha, \beta)\sigma^{\alpha+\beta-1}}(x-\theta)^{\alpha-1}(\theta+\sigma-x)^{\beta-1} \qquad (\text{단}, \ \theta \leq x \leq \theta+\sigma, \ 0<\sigma,\alpha,\beta)$$

$$E(x) = \theta + \sigma\frac{\alpha}{\alpha+\beta}$$

$$\mathrm{Var}(x) = \frac{\sigma^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

여기서 $B(\cdot)$ 는 베타 함수입니다 .

정규 혼합

정규 혼합 분포 적합에 대한 자세한 내용은 " 정규 2 혼합 적합 및 정규 3 혼합 적합 "(85 페이지) 에서 확인하십시오 .

평활 곡선

평활 곡선 옵션은 비모수 밀도 추정 (커널 밀도 추정) 을 사용하여 평활 곡선을 적합시킵니다 . 평활 곡선은 히스토그램에 중첩 표시되며 그림 아래에는 슬라이더가 나타납니다 . 슬라이더로 커널 표준편차를 변경하여 평활 정도를 조절할 수 있습니다 . 초기 **커널 표준편차** 추정값은 데이터의 표준편차를 사용하여 계산됩니다 .

SHASH

SHASH 분포 적합에 대한 자세한 내용은 "SHASH 적합"(83 페이지)에서 확인하십시오.

Johnson Su, Johnson Sb, Johnson Sl

Johnson 분포 체계에는 세 가지 분포가 포함되어 있으며 이러한 분포는 모두 변환된 정규 분포를 기반으로 합니다. 이 세 가지 분포는 다음과 같습니다.

- Johnson Su - 경계가 없습니다.
- Johnson Sb - 양쪽 꼬리에 경계가 있습니다. 이 경계는 추정 가능한 모수에 의해 정의됩니다.
- Johnson Sl - 한쪽 꼬리에 경계가 있습니다. 이 경계는 추정 가능한 모수에 의해 정의됩니다. Johnson Sl 계열에는 로그 정규 분포 계열이 포함됩니다.

S는 체계, 즉 범위의 서브스크립트를 나타냅니다. 레거시 적합기에는 다른 방법이 구현되기는 했지만, Slifker and Shapiro 연구 자료(1980)에서 우리가 사용한 특정 Johnson 체계의 선택 기준을 확인할 수 있습니다.

Johnson 분포는 유연하기 때문에 널리 사용됩니다. 특히 Johnson 분포 체계는 왜도와 첨도의 가능한 모든 조합을 지원하므로 데이터 적합 기능 면에서 주목됩니다.

Z 가 표준 정규 변량인 경우 이 분포 체계는 다음과 같이 정의됩니다.

$$Z = \gamma + \delta f(Y)$$

다음은 Johnson Su 의 경우에 해당하는 각 요소에 대한 설명입니다.

$$f(Y) = \ln\left(Y + \sqrt{1 + Y^2}\right) = \sinh^{-1} Y$$

$$Y = \frac{X - \theta}{\sigma} \qquad -\infty < X < \infty$$

다음은 Johnson Sb 의 경우에 해당하는 각 요소에 대한 설명입니다.

$$f(Y) = \ln\left(\frac{Y}{1 - Y}\right)$$

$$Y = \frac{X - \theta}{\sigma} \qquad \theta < X < \theta + \sigma$$

다음은 Johnson Sl 의 경우 (단 , $\sigma = \pm 1$) 에 해당하는 각 요소에 대한 설명입니다.

$$f(Y) = \ln(Y)$$

$$Y = \frac{X - \theta}{\sigma} \qquad \begin{array}{ll} \theta < X < \infty & \text{단, } \sigma = 1 \\ -\infty < X < \theta & \text{단, } \sigma = -1 \end{array}$$

Johnson Su

pdf: $\frac{\delta}{\sigma}\left[1+\left(\frac{x-\theta}{\sigma}\right)^2\right]^{-1/2}\phi\left[\gamma+\delta\sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)\right]$ (단 , $-\infty < x, \theta, \gamma < \infty$, $0 < \theta, \delta$)

Johnson Sb

pdf: $\phi\left[\gamma+\delta\ln\left(\frac{x-\theta}{\sigma-(x-\theta)}\right)\right]\left(\frac{\delta\sigma}{(x-\theta)(\sigma-(x-\theta))}\right)$ (단 , $\theta < x < \theta+\sigma$, $0 < \sigma$)

Johnson Sl

pdf: $\frac{\delta}{|x-\theta|}\phi\left[\gamma+\delta\ln\left(\frac{x-\theta}{\sigma}\right)\right]$ (단 , $\sigma = 1$ 일 경우 $\theta < x$, $\sigma = -1$ 일 경우 $\theta > x$)

여기서 $\phi(\cdot)$ 는 표준 정규 pdf 입니다 .

다음 사항에 유의하십시오 .

- 작업 순서와 컴퓨터 정밀도로 인해 컴퓨터마다 모수 추정값이 다를 수 있습니다 .
- 기본 보고서에는 모수 신뢰 구간이 숨겨져 있습니다 . Johnson 분포는 정규 분포로 변환되므로 이 경우에는 모수 신뢰 구간이 그다지 의미가 없습니다 . 모수 신뢰 구간을 표시하려면 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 하한** 및 **95% 상한**을 선택합니다 .

GLog(일반화 로그)

이 분포는 생물학적 분석 데이터처럼 거의 정규 분포를 따르지 않으며 대개 분산이 일정하지 않은 데이터를 적합시키는 데 유용합니다 . GLog 분포는 μ(위치), σ(척도) 및 λ(형상) 모수로 설명됩니다 .

pdf: $\phi\left\{\frac{1}{\sigma}\left[\log\left(\frac{x+\sqrt{x^2+\lambda^2}}{2}\right)-\mu\right]\right\}\frac{x+\sqrt{x^2+\lambda^2}}{\sigma(x^2+\lambda^2+x\sqrt{x^2+\lambda^2})}$

(단 , $0 \le \lambda$, $0 < \sigma$, $-\infty < \mu < \infty$)

GLog 분포는 정규 분포로 변환되며 , 다음과 같은 관계에 있습니다 .

$z = \frac{1}{\sigma}\left[\log\left(\frac{x+\sqrt{x^2+\lambda^2}}{2}\right)-\mu\right] \sim N(0,1)$ 이면 $x \sim Glog(\mu,\sigma,\lambda)$ 입니다 .

$\lambda = 0$ 이면 GLog 가 LogNormal(μ,σ) 로 축소됩니다 .

참고 : 기본 보고서에는 모수 신뢰 구간이 숨겨져 있습니다 . GLog 분포는 정규 분포로 변환되므로 이 경우에는 모수 신뢰 구간이 그다지 의미가 없습니다 . 모수 신뢰 구간을 표시하려면 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 하한** 및 **95% 상한**을 선택합니다 .

모두

"분포 비교" 보고서에서는 분포 목록이 AICc를 기준으로 오름차순 정렬됩니다.

AICc 는 다음과 같이 정의됩니다 .

$$AICc = -2\log L + 2\nu + \frac{2\nu(\nu+1)}{n-(\nu+1)}$$

다음은 각 요소에 대한 설명입니다 .

- logL 은 로그 가능도입니다 .
- n 은 표본 크기입니다 .
- ν 은 모수의 개수입니다 .

열에 음수 값이 포함되었으면 양수 값을 갖는 데이터를 필요로 하는 분포는 분포 목록에 포함되지 않습니다. 연속형 분포만 나열됩니다. 베타 분포와 Johnson Sb 분포와 같이 임계 모수가 있는 분포는 가능한 분포 목록에 포함되지 않습니다.

이산형 적합 분포에 대한 통계 상세 정보 (레거시)

이 섹션에는 " 이산형 적합 " 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다 .

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

Poisson

Poisson 분포 적합에 대한 자세한 내용은 "Poisson 적합"(88 페이지)에서 확인하십시오.

감마 Poisson

이 분포는 데이터가 몇 가지 Poisson(μ) 분포의 조합이고 각 Poisson(μ) 분포의 μ 이 다른 경우에 유용합니다. 한 예로, 교차로 간에 평균 사고 발생 횟수(μ)가 다른 경우 여러 교차로에서 결합된 전체 사고 발생 횟수를 들 수 있습니다.

감마 Poisson 분포는 x|μ 가 Poisson 분포를 따르고 μ 가 Gamma(α,τ) 분포를 따른다고 가정한 결과입니다 . 감마 Poisson 에는 $\lambda = \alpha\tau$ 및 $\sigma = \tau+1$ 모수가 있습니다 . σ 모수는 산포 모수입

니다 . $\sigma > 1$ 이면 과대산포가 나타납니다 . 즉 , x 에서 Poisson 만으로 설명할 경우보다 더 많은 변동이 나타납니다 . $\sigma = 1$ 이면 x 가 Poisson(λ) 으로 축소됩니다 .

$$\text{pmf: } \frac{\Gamma\left(x+\frac{\lambda}{\sigma-1}\right)}{\Gamma(x+1)\Gamma\left(\frac{\lambda}{\sigma-1}\right)}\left(\frac{\sigma-1}{\sigma}\right)^{x}\sigma^{-\frac{\lambda}{\sigma-1}} \quad (\text{단 , } 0<\lambda,\ 1\le\sigma,\ x=0,1,2,\ldots)$$

$E(x) = \lambda$

$\text{Var}(x) = \lambda\sigma$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

$x \mid \mu \sim \text{Poisson}(\mu)$ 인 반면 $\mu \sim \text{Gamma}(\alpha,\tau)$ 임을 기억하십시오 . 이 플랫폼은 $\lambda = \alpha\tau$ 이고 $\sigma = \tau+1$ 이라고 추정합니다 . α 및 τ 의 추정값을 구하려면 다음 계산식을 사용합니다 .

$$\hat{\tau} = \hat{\sigma} - 1$$

$$\hat{\alpha} = \frac{\hat{\lambda}}{\hat{\tau}}$$

σ 의 추정값이 1 인 경우에는 이 계산식이 적용되지 않습니다 . 이 경우에는 감마 Poisson 이 Poisson(λ) 으로 축소된 것이며 $\hat{\lambda}$ 는 λ 의 추정값입니다 .

α 의 추정값이 정수인 경우 감마 Poisson 은 다음 pmf 를 갖는 음이항과 동등합니다 .

$$p(y) = \binom{y+r-1}{y}p^{r}(1-p)^{y} \quad (\text{단 , } 0 \le y)$$

이때 $r = \alpha$ 이고 $(1\text{-}p)/p = \tau$ 입니다 .

모수 λ 및 σ 를 갖는 감마 Poisson 분포를 모수 λ 를 갖는 Poisson 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoGammaPoisson.jsl 을 실행하십시오 .

이항

이항 분포 적합에 대한 자세한 내용은 "이항 적합"(89 페이지)에서 확인하십시오 .

베타 이항

베타 이항 분포 적합에 대한 자세한 내용은 "베타 이항 적합"(90 페이지)에서 확인하십시오 .

적합 분위수에 대한 통계 상세 정보 (레거시)

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

진단 그림의 적합 분위수와 **적합 분위수 저장** 명령으로 저장된 적합 분위수는 다음 방법을 사용하여 구합니다 .

1. 데이터를 정렬하고 순위화합니다 . 동일 값에는 서로 다른 순위가 할당됩니다 .
2. $p_{[i]} = rank_{[i]}/(n+1)$ 를 계산합니다 .
3. $quantile_{[i]} = Quantile_d(p_{[i]})$ 를 계산합니다 . 여기서 $Quantile_d$ 는 특정 적합 분포에 대한 분위수 함수이고 $i = 1,2,...,n$ 입니다 .

분포 적합 옵션에 대한 통계 상세 정보 (레거시)

이 섹션에서는 분포 적합을 위한 적합도 검정과 적합된 분포와 관련된 규격 한계의 통계 상세 정보를 설명합니다.

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

적합도

표 3.2 JMP 적합도 검정 설명

분포	모수	적합도 검정
정규[a]	μ 및 σ 를 알 수 없음	Shapiro-Wilk($n \leq 2000$ 인 경우), Kolmogorov-Smirnov-Lillefors ($n > 2000$ 인 경우)
	μ 및 σ 를 모두 알고 있음	Kolmogorov-Smirnov-Lillefors
	μ 또는 σ 중 하나를 알고 있음	(없음)
로그 정규	μ 및 σ 를 알고 있거나 알 수 없음	Kolmogorov *D*
Weibull	α 및 β 를 알고 있거나 알 수 없음	Cramér-von Mises W^2
임계가 있는 Weibull	α, β 및 θ 를 알고 있거나 알 수 없음	Cramér-von Mises W^2
극단값	α 및 β 를 알고 있거나 알 수 없음	Cramér-von Mises W^2
지수	σ 를 알고 있거나 알 수 없음	Kolmogorov *D*
감마	α 및 σ 를 알고 있음	Cramér-von Mises W^2
	α 또는 σ 중 하나를 알 수 없음	(없음)

표 3.2 JMP 적합도 검정 설명 (계속)

분포	모수	적합도 검정
베타	α 및 β를 알고 있음	Kolmogorov *D*
	α 또는 β 중 하나를 알 수 없음	(없음)
이항	ρ를 알고 있거나 알 수 없으며, *n*을 알고 있음	Kolmogorov *D*(n ≤ 30인 경우), Pearson χ^2(n > 30인 경우)
베타 이항	ρ 및 δ를 알고 있거나 알 수 없음	Kolmogorov *D*(n ≤ 30인 경우), Pearson χ^2(n > 30인 경우)
Poisson	λ를 알고 있거나 알 수 없음	Kolmogorov *D*(n ≤ 30인 경우), Pearson χ^2(n > 30인 경우)
감마 Poisson	λ 또는 σ를 알고 있거나 알 수 없음	Kolmogorov *D*(n ≤ 30인 경우), Pearson χ^2(n > 30인 경우)

a. 세 가지 Johnson 분포와 GLog 분포의 경우에는 데이터가 정규 데이터로 변환된 후 적절한 정규성 검정이 수행됩니다 .

K 시그마 규격 한계 설정

공정 능력 분석을 위해 K 값을 입력하고 단측 또는 양측을 선택할 수 있습니다. K 표준편차에 해당하는 꼬리 확률은 정규 분포에서 계산됩니다. 이 확률은 적합시킨 특정 분포에 맞는 분위수로 변환됩니다. 결과 분위수는 공정 능력 분석에서 규격 한계에 사용됩니다. 이 옵션은 **분위수** 옵션과 유사하지만 확률 대신 K를 제공해야 합니다. K는 규격 한계와 평균 사이의 거리를 나타내는 표준편차 수에 해당합니다.

예를 들어 정규 분포에서 K = 3인 경우 평균의 하위 및 상위 3 표준편차는 각각 0.00135번째 분위수와 0.99865번째 분위수에 해당합니다. 규격 하한은 적합 분포의 0.00135번째 분위수에서 설정되고, 규격 상한은 0.99865번째 분위수에서 설정됩니다. 공정 능력 분석은 이러한 규격 한계를 기준으로 반환됩니다.

4 장

X 로 Y 적합 소개

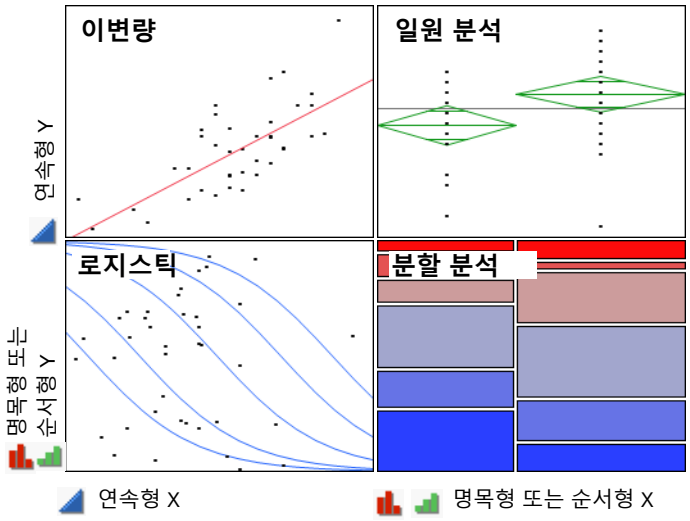
두 변수 간의 관계 분석

변수 쌍을 분석하려면 "X로 Y 적합" 플랫폼을 사용합니다. 산점도, 선형 회귀, ANOVA, 다중 비교, 로지스틱 회귀, 분할표 등을 사용하여 이 작업을 수행할 수 있습니다. 특정 분석은 변수의 모델링 유형에 따라 달라집니다.

"X로 Y 적합"은 네 가지 플랫폼 중 하나를 시작합니다.

- 연속형 Y 및 연속형 X 를 분석하려면 이변량을 사용합니다 .
- 연속형 Y 및 범주형 X 를 분석하려면 ANOVA 를 포함한 일원 분석을 사용합니다 .
- 범주형 Y 및 연속형 X 를 분석하려면 로지스틱 회귀를 사용합니다 .
- 범주형 Y 및 범주형 X 를 분석하려면 분할 분석을 사용합니다 .

그림 4.1 네 가지 분석 유형의 예



목차

X 로 Y 적합 플랫폼의 개요 107
X 로 Y 적합 플랫폼 시작 107
JMP 시작하기 창에서 특정 분석 시작 108

X 로 Y 적합 플랫폼의 개요

X 로 Y 적합 플랫폼은 네 가지 상세 플랫폼 (또는 분석 유형) 을 모은 것입니다 .

상세 플랫폼	모델링 유형	설명
이변량	연속형 X로 연속형 Y 적합	두 연속형 변수 간의 관계를 분석합니다. 자세한 내용은 이변량 분석에서 확인하십시오.
일원 분석	명목형 또는 순서형 X로 연속형 Y 적합	범주형 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지를 분석합니다. 자세한 내용은 일원 분석에서 확인하십시오.
로지스틱	연속형 X로 명목형 또는 순서형 Y 적합	반응 범주에 대한 확률을 연속형 X 예측 변수에 적합시킵니다. 자세한 내용은 로지스틱 분석에서 확인하십시오.
분할 분석	명목형 또는 순서형 X로 명목형 또는 순서형 Y 적합	범주형 X 요인의 값에 따라 좌우되는 범주형 반응 변수의 분포를 분석합니다. 자세한 내용은 분할 분석에서 확인하십시오.

X 로 Y 적합 플랫폼 시작

분석 > X 로 Y 적합을 선택하여 X 로 Y 적합 플랫폼을 시작할 수 있습니다 .

그림 4.2 X 로 Y 적합 시작 창

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

이변량 , 일원 분석 , 로지스틱 , 분할 분석 이 격자에는 서로 다른 데이터 유형 조합의 분석 결과가 표시됩니다 . 열을 할당하면 적용 가능한 플랫폼이 격자 위에 라벨로 표시됩니다 .

블록 (선택 사항 . 일원 분석 및 분할 분석에만 사용 가능):

- 일원 분석 플랫폼의 경우 블록 변수를 지정하면 교호작용이 없는 이원 분석을 구성하는 두 번째 요인이 식별됩니다 . 일원 분석 그림에서는 Y 변수의 값이 블록 변수에 의해 중심화됩니다 . 그룹 셀별로 각 블록에 동일한 개수가 있는 경우 사용 가능한 일원 분석 플랫폼 옵션이 이원 분석 설정에 맞게 조정됩니다 . 그룹 셀별 블록에 동일한 개수가 있지 않은 경우 모형 적합 보고서가 일원 분석 그림 아래에 표시되고 일원 분석 플랫폼 옵션을 사용할 수 없습니다 .
- 분할 분석 플랫폼의 경우 블록 변수를 지정하면 두 번째 요인이 식별되고 Cochran-Mantel-Haenszel 검정이 수행됩니다 .

시작 창에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

JMP 시작하기 창에서 특정 분석 시작

"JMP 시작하기" 창에서 특정 분석(**이변량**, **일원 분석**, **로지스틱** 또는 **분할 분석**)을 시작할 수 있습니다. 이 옵션을 선택할 경우 분석에 적절한 모델링 유형(Y 및 X 변수)을 지정합니다(표 4.1).

"JMP 시작하기 " 창에서 특정 분석을 시작하려면 **기본 분석** 범주를 클릭하고 특정 분석을 선택합니다 .

플랫폼 시작 옵션은 대부분 동일합니다. 하지만 Y 및 X 플랫폼 버튼 중 일부는 수행하는 특정 분석에 맞게 사용자 정의됩니다.

표 4.1 플랫폼 및 버튼

플랫폼 또는 분석	Y 버튼	X 버튼
X로 Y 적합	Y, 반응	X, 요인
이변량	Y, 반응	X, 회귀변수
일원 분석	Y, 반응	X, 그룹화
로지스틱	Y, 범주형 반응	X, 연속형 회귀변수
분할 분석	Y, 반응 범주	X, 그룹화 범주

5 장

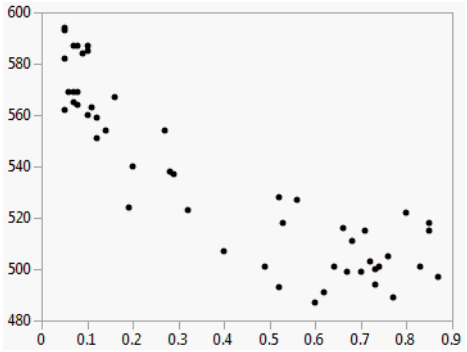
이변량 분석

두 연속형 변수 간의 관계 분석

이변량 플랫폼을 사용하여 두 연속형 변수 간의 관계를 조사합니다. 단순 선형 회귀선, 다항 회귀 곡선 또는 평활기를 산점도에 대화식으로 추가합니다. 산점도를 사용하면 두 변수 사이의 관계를 한눈에 확인할 수 있습니다.

이변량 플랫폼은 X로 Y 적합 플랫폼의 **연속형 대 연속형** 분석법입니다. 이변량이란 단어는 단순히 관련 변수가 한 개(단변량) 또는 여러 개(다변량)가 아니라 두 개임을 의미합니다.

그림 5.1 이변량 분석의 예



목차

이변량 분석의 예 . . . 112
이변량 플랫폼 시작 . . . 112
이변량 그림 . . . 113
적합 옵션 . . . 114
적합 옵션 . . . 115
적합 옵션 범주 . . . 117
동일한 적합 옵션을 여러 번 적용 . . . 117
히스토그램 테두리 . . . 118
평균 적합 . . . 119
평균 적합 보고서 . . . 119
선형 적합 및 다항식 적합 . . . 120
선형 적합 및 다항식 적합 보고서 . . . 120
특수 적합 . . . 125
특수 적합 보고서 및 메뉴 . . . 126
유연 . . . 127
스플라인 적합 . . . 127
커널 평활기 . . . 128
각 값 적합 . . . 129
직교 적합 . . . 129
직교 적합 비율 보고서 . . . 130
로버스트 . . . 130
로버스트 적합 . . . 131
Cauchy 적합 . . . 131
밀도 타원 . . . 131
상관 보고서 . . . 132
비모수 밀도 . . . 133
분위수 밀도 등고선 보고서 . . . 133
그룹화 기준 . . . 133
적합 메뉴 . . . 134
적합 메뉴 옵션 . . . 135
진단 그림 . . . 137
이변량 플랫폼의 추가 예 . . . 138
특수 적합 옵션의 예 . . . 138
직교 적합 옵션의 예 . . . 140
로버스트 적합 명령의 예 . . . 142
밀도 타원을 사용한 그룹화의 예 . . . 144
회귀선을 사용한 그룹화의 예 . . . 145
기준 변수를 사용한 그룹화의 예 . . . 146

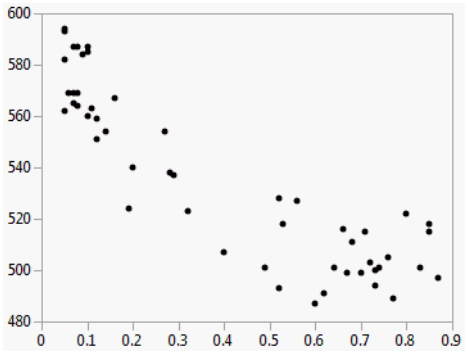
이변량 플랫폼에 대한 통계 상세 정보 147
선형 적합 148
스플라인 적합 148
직교 적합 148
적합 요약 보고서 149
적합 결여 보고서 149
모수 추정값 보고서 150
평활 적합 보고서 150
상관 보고서 150

이변량 분석의 예

두 연속형 변수에 대한 산점도를 생성하는 방법을 알아봅니다 . 이 예에서는 SAT.jmp 샘플 데이터 테이블을 사용합니다 . 여기에는 미국 50 개 주와 컬럼비아 특별구에 거주하는 학생의 SAT 시험 점수가 언어 영역과 수학 영역으로 나뉘어 있습니다 . 2004 년도 SAT 시험의 학생 응시율과 언어 영역 시험 점수 간에 어떤 관계가 있는지 알아보려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 SAT.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. 2004 Verbal 을 선택하고 **Y, 반응**을 클릭합니다 .
4. % Taking (2004) 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 5.2 응시율별 SAT 점수의 예



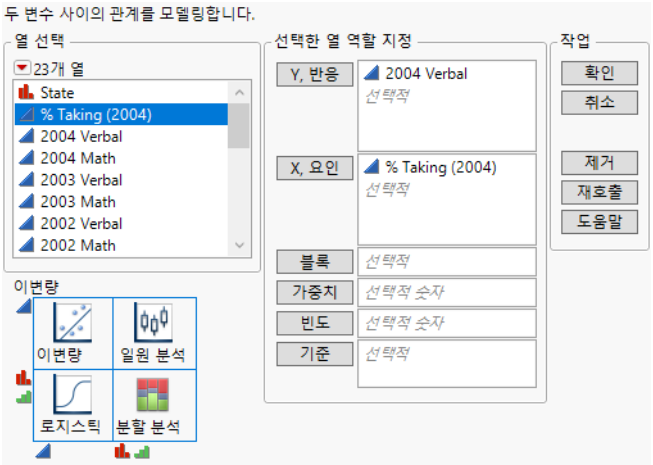
시험에 응시한 모집단의 비율이 낮을수록 언어 영역 점수가 상승했음을 알 수 있습니다.

이변량 플랫폼 시작

이변량 분석에서 두 연속형 변수 간의 관계를 조사하려면 다음을 수행하십시오 .

1. **분석 > X 로 Y 적합**을 선택합니다 .
2. **Y, 반응**에 연속형 열을 입력합니다 .
3. **X, 요인**에 연속형 열을 입력합니다 .

그림 5.3 이변량 시작 창



다이어그램 위에 "이변량"이라고 표시되면 이변량 분석을 수행할 것임을 나타냅니다.

참고 : "JMP 시작하기 " 창에서 이변량 분석을 시작할 수도 있습니다 . **보기** > **JMP 시작하기** > **기본 분석** > **이변량**을 선택합니다 .

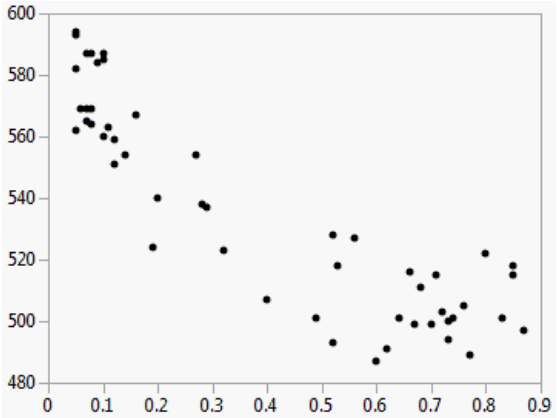
이 시작 창에 대한 자세한 내용은 "X 로 Y 적합 소개" 장(105 페이지)에서 확인하십시오 . "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

이변량 그림

이변량 보고서는 각 X 및 Y 변수 쌍에 대한 산점도로 시작됩니다 .

참고 : 그림 5.4 에 표시된 것과 같은 그림을 생성하려면 " 이변량 분석의 예 "(112 페이지) 에 설명된 방법을 따르십시오 .

그림 5.4 이변량 그림



두 가지 방법 중 하나로 변수를 드래그하여 놓아 그림의 변수를 바꿀 수 있습니다. 한 가지 방법은 한 축의 변수를 다른 축으로 드래그하여 놓아 기존 변수를 바꾸는 것이고, 다른 방법은 관련 데이터 테이블의 "열" 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다.

다른 JMP 그림과 마찬가지로 이 그림에서도 그림 크기를 조정하거나, 화살표 또는 브러시 도구로 점을 강조 표시하거나, 점에 라벨을 지정하는 등의 상호 작용이 가능합니다. 이러한 기능에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

빨간색 삼각형 메뉴 내에 있는 적합 옵션을 사용하여 그림에서 곡선을 적합시키고 통계 보고서 및 추가 메뉴를 볼 수 있습니다. 자세한 내용은 " 적합 옵션 "(114 페이지) 에서 확인하십시오.

적합 옵션

" 이변량 적합 " 의 빨간색 삼각형 메뉴에는 표시 옵션 , 적합 옵션 및 제어 옵션이 포함되어 있습니다 .

참고 : " 적합 그룹 " 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에 표시됩니다 . 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 를 기준으로 정렬할 수 있습니다 . 자세한 내용은 선형 모형 적합 의 에서 확인하십시오 .

점 표시 산점도에서 점을 숨기거나 표시합니다. 체크 표시는 점이 표시되어 있음을 나타냅니다.

히스토그램 테두리 산점도의 *x* 및 *y* 축에 히스토그램을 추가합니다 . 체크 표시는 히스토그램 테두리가 설정되어 있음을 나타냅니다 . 자세한 내용은 " 히스토그램 테두리 "(118 페이지) 에서 확인하십시오 .

참고 : 데이터 테이블의 행에 " 숨김 " 행 상태만 적용하면 산점도에 해당 점이 표시되지 않습니다. 하지만 히스토그램을 생성하는 데는 숨겨진 행이 사용됩니다. 생성되는 히스토그램과 분석 결과에서 행을 제외하려면 " 제외 " 행 상태를 적용하고 " 이변량 적합 " 의 빨간색 삼각형 메뉴에서 **다시 실행 > 분석 다시 실행**을 선택하십시오 .

요약 통계량 그림과 관련된 상관, 신뢰 구간, 평균, 표준편차 등의 요약 통계량을 표시합니다.

그룹화 기준 분류 (또는 그룹화) 변수를 선택할 수 있습니다 . 그러면 그룹화 변수의 각 수준에 대해 개별 분석이 계산되고 산점도에 회귀 곡선 또는 타원이 중첩 표시됩니다 . 자세한 내용은 " 그룹화 기준 "(133 페이지) 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

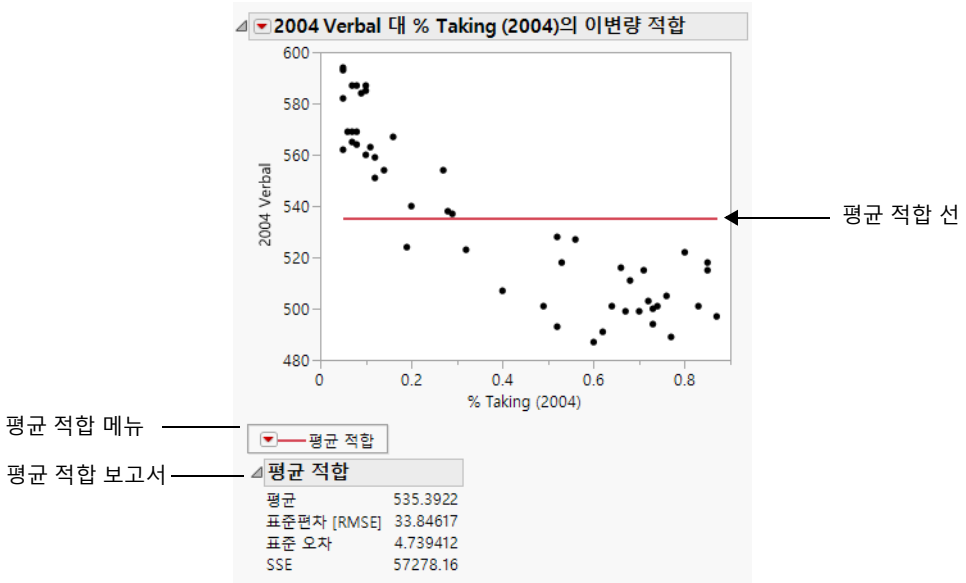
그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

적합 옵션

각 적합 옵션으로 다음을 추가할 수 있습니다 .

- 선 , 곡선 또는 분포 (산점도)
- 빨간색 삼각형 메뉴 (보고서 창)
- 특정 보고서 (보고서 창)

그림 5.5 평균 적합 옵션의 예



다음과 같은 적합 옵션을 사용할 수 있습니다 .

평균 적합 Y 반응 변수의 평균을 나타내는 수평선을 산점도에 추가합니다 . 자세한 내용은 " 평균 적합 "(119 페이지) 에서 확인하십시오 .

선형 적합 최소 제곱 회귀를 사용하여 산점도에 적합 직선을 추가합니다 . 자세한 내용은 " 선형 적합 및 다항식 적합 "(120 페이지) 에서 확인하십시오 .

다항식 적합 최소 제곱 회귀를 사용하여 특정 차수의 다항식 곡선을 적합시킵니다 . 자세한 내용은 " 선형 적합 및 다항식 적합 "(120 페이지) 에서 확인하십시오 .

특수 적합 *Y* 및 *X* 를 변환합니다 . 변환에는 로그, 제곱, 역수 및 지수가 포함됩니다 . 중심화 다항식을 해제하거나 , 절편 및 기울기를 제약하거나 , 다항식 모형을 적합시킬 수도 있습니다 . 자세한 내용은 " 특수 적합 "(125 페이지) 에서 확인하십시오 .

유연 추정된 회귀 곡선의 평활도를 제어할 수 있는 옵션을 제공합니다 . 자세한 내용은 " 유연 "(127 페이지) 에서 확인하십시오 .

직교 적합 X 가 변동하는 것으로 간주되는 경우에 유용한 직교 회귀 적합을 위한 옵션을 제공합니다 . 이 옵션의 경우 X 및 Y 의 분산에 대한 다양한 가정을 반영하는 하위 옵션이 제공됩니다 . 자세한 내용은 " 직교 적합 "(129 페이지) 에서 확인하십시오 .

로버스트 적합 모형에서 데이터 집합에 포함된 이상치의 영향을 줄이는 옵션을 제공합니다 . 자세한 내용은 " 로버스트 "(130 페이지) 에서 확인하십시오 .

밀도 타원 *X* 및 *Y* 변수에 대한 이변량 정규 분포 적합을 위해 밀도 타원을 그립니다 . 자세한 내용은 " 밀도 타원 "(131 페이지) 에서 확인하십시오 .

비모수 밀도 평활된 표면을 기반으로 밀도 등고선을 그립니다 . 이 등고선은 데이터 점의 밀도를 설명합니다 . 자세한 내용은 " 비모수 밀도 "(133 페이지) 에서 확인하십시오 .

참고 : 적합 제거 옵션을 사용하여 적합을 제거할 수 있습니다 . 자세한 내용은 " 적합 메뉴 옵션 "(135 페이지) 에서 확인하십시오 .

적합 옵션 범주

적합 옵션 범주에는 회귀 적합과 밀도 추정이 포함됩니다.

범주	설명	적합 옵션
회귀 적합	회귀 방법은 관측된 데이터 점에 곡선을 적합시킵니다. 이 적합 방법에는 최소 제곱 적합과 함께 스플라인 적합, 커널 평활, 직교 적합 및 로버스트 적합이 포함됩니다.	평균 적합 선형 적합 다항식 적합 특수 적합 유연 직교 적합 로버스트
밀도 추정	밀도 추정은 점에 이변량 분포를 적합시킵니다. 타원형 등고선이 특징인 이변량 정규 밀도나 일반 비모수 밀도 중 하나를 선택할 수 있습니다.	밀도 타원 비모수 밀도

동일한 적합 옵션을 여러 번 적용

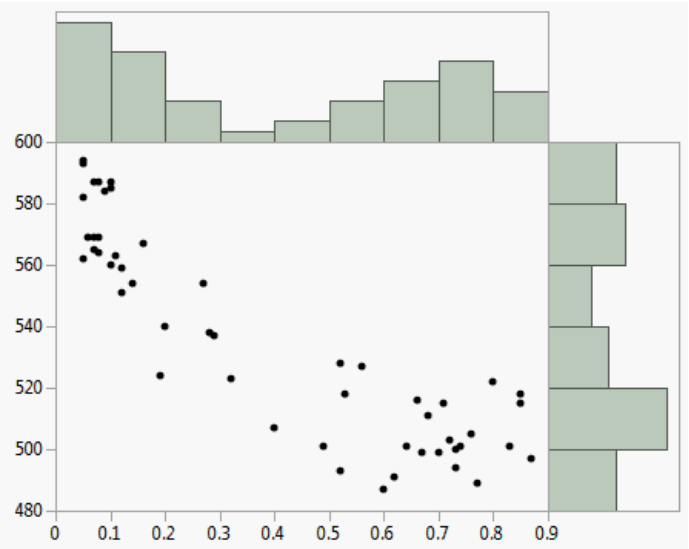
동일한 적합 옵션을 여러 번 선택하여 각각의 새 적합이 산점도에 중첩 표시되도록 할 수 있습니다. 적합을 시도한 후 특정 점을 제외하고 다시 적합시킬 수 있으며, 각 적합을 동일한 산점도에서 비교할 수 있습니다.

보고서 창에서 한 가지 적합 옵션을 여러 분석에 적용하려면 Ctrl 키를 누른 채로 적합 옵션을 선택합니다.

히스토그램 테두리

히스토그램 테두리 옵션은 산점도의 x 및 y 축에 히스토그램을 추가합니다. 히스토그램을 사용하여 X 및 Y 변수의 주변 분포를 시각화할 수 있습니다 .

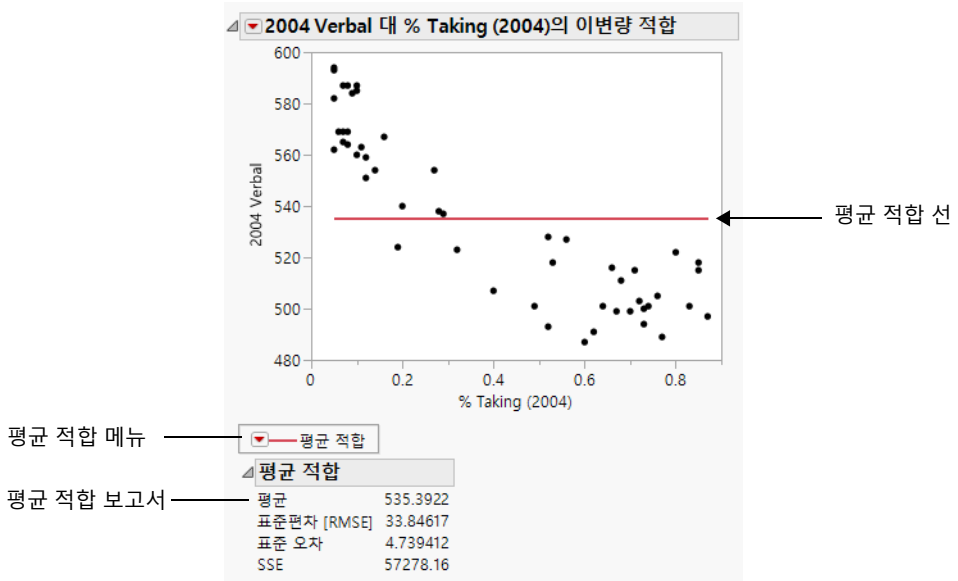
그림 5.6 히스토그램 테두리의 예



평균 적합

평균 적합 옵션을 사용하여 *Y* 반응 변수의 평균을 나타내는 수평선을 산점도에 추가할 수 있습니다 . 먼저 평균을 적합시킨 후 평균 선을 다른 적합 (직선 , 신뢰 곡선 , 다항식 곡선 등) 에 참조로 사용할 수 있습니다 .

그림 5.7 평균 적합의 예



평균 적합 보고서

" 평균 적합 " 보고서에는 평균 적합에 대한 요약 통계량이 표시됩니다 .

평균 반응 변수의 평균입니다 . 모형에 지정된 효과가 없는 경우에는 예측된 반응입니다 .

표준편차 [RMSE] 반응 변수의 표준편차입니다 . 평균 제곱 오차의 제곱근으로 , RMSE(제곱근 평균 제곱 오차) 라고도 합니다 .

표준 오차 반응 평균의 표준편차입니다 . RMSE 를 값 개수의 제곱근으로 나눠서 계산됩니다 .

SSE 단순 평균 모형의 오차제곱합입니다 . 각 모형 적합에 대한 분산 분석 테이블에서 **오차**의 제곱합으로 표시됩니다 .

"평균 적합" 메뉴의 옵션에 대한 자세한 내용은 "적합 메뉴"(134 페이지)에서 확인하십시오.

선형 적합 및 다항식 적합

선형 적합 옵션을 선택하면 최소 제곱 회귀를 사용하여 산점도에 직선 적합을 추가할 수 있습니다. **다항식 적합** 옵션을 선택하면 최소 제곱 회귀를 사용하여 특정 차수의 다항식 곡선을 적합시킬 수 있습니다.

그림 5.8 선형 적합 및 다항식 적합의 예

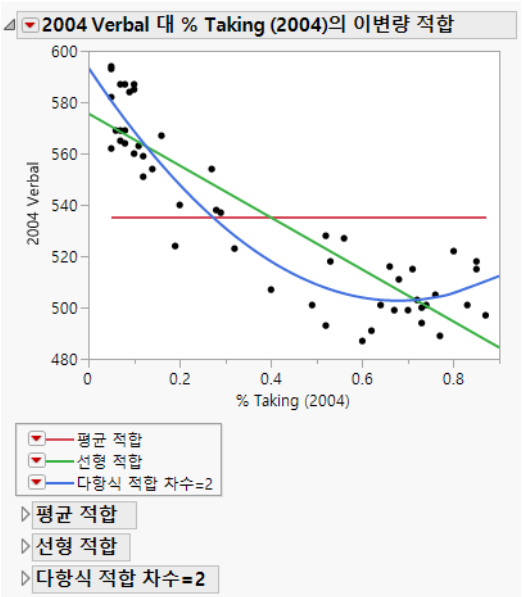


그림 5.8에서는 평균 선에 대한 선형 적합과 2차 다항식 적합을 비교하는 예를 보여 줍니다.

다음 사항에 유의하십시오.

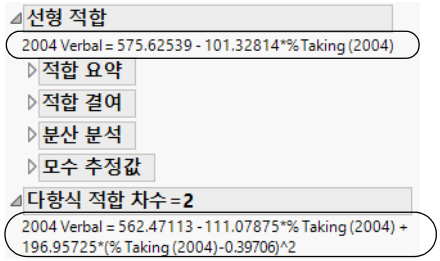
- **선형 적합** 출력은 1 차 다항식 적합과 동등합니다.
- **평균 적합** 출력은 0 차 다항식 적합과 동등합니다.

"선형 적합"과 "다항식 적합 차수" 메뉴의 옵션에 대한 자세한 내용은 "적합 메뉴"(134 페이지)에서 확인하십시오. 이 적합에 대한 통계 상세 정보는 "선형 적합"(148 페이지)에서 확인하십시오.

선형 적합 및 다항식 적합 보고서

"선형 적합" 및 "다항식 적합" 보고서는 적합 방정식으로 시작됩니다.

그림 5.9 적합 방정식의 예



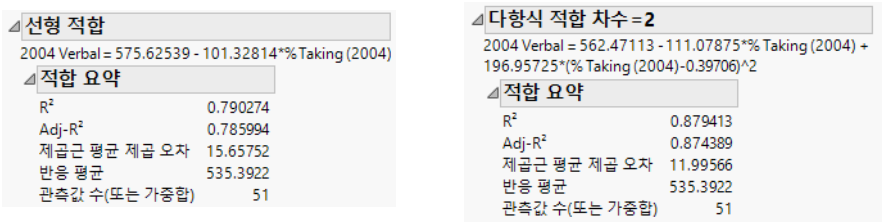
팁 : 방정식을 클릭하여 편집할 수 있습니다 .

각 "선형 적합" 및 "다항식 적합 차수" 보고서에는 세 개 이상의 보고서가 포함됩니다. 네 번째인 "적합 결여" 보고서는 데이터에 *X* 반복 실험이 있는 경우에만 표시됩니다.

적합 요약 보고서

이변량 플랫폼의 "적합 요약" 보고서에는 동일한 데이터에 대한 선형 적합 및 2 차 다항식 적합의 반응과 관련한 수치 요약이 표시됩니다. 여러 "적합 요약" 보고서를 비교하여 더 나은 모형을 확인할 수 있습니다. 더 나은 모형일수록 R^2 값이 크고 **제곱근 평균 제곱 오차**는 작습니다.

그림 5.10 선형 적합 및 다항식 적합의 적합 요약 보고서



" 적합 요약 " 보고서에는 다음 열이 포함됩니다 .

R^2 모형에 의해 설명된 변동의 비율을 측정합니다 . 나머지 변동은 해당 모형으로 설명되지 않으며 이는 랜덤 오차에 기인합니다 . 모형이 완벽하게 적합되는 경우에는 R^2 가 1 입니다 .

참고 : R^2 값이 낮으면 설명되지 않은 변동을 설명하는 변수가 모형에 없는 것일 수 있습니다 . 하지만 데이터의 내재 변동 범위가 큰 경우에는 유용한 회귀 모형이라도 R^2 값이 낮을 수 있습니다 . 일반적인 R^2 값에 대해 알아보려면 연구 영역의 문헌을 읽어 보십시오 .

그림 5.10 의 R^2 값은 2 차 다항식 적합이 선형 적합보다 좀더 낫다는 것을 나타냅니다 . 자세한 내용은 " 적합 요약 보고서 "(149 페이지) 에서 확인하십시오 .

Adj-R^2 모수의 개수가 서로 다른 모형들을 비교하기 좋도록 계산에 자유도를 사용하여 R^2 값을 조정합니다 . 자세한 내용은 " 적합 요약 보고서 "(149 페이지) 에서 확인하십시오 .

제곱근 평균 제곱 오차 랜덤 오차의 표준편차를 추정합니다 . " 분산 분석 " 보고서에 표시된 **오차** 에 대한 평균 제곱의 제곱근입니다 (그림 5.12).

반응 평균 반응 변수의 표본 평균 (산술평균) 을 제공합니다 . 모형 효과가 지정되지 않은 경우에는 예측된 반응입니다 .

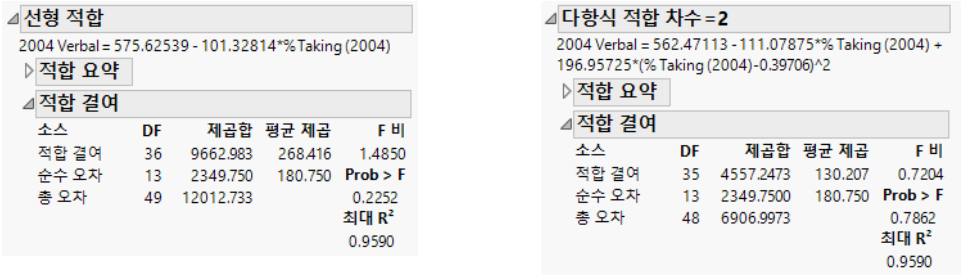
관측값 적합을 추정하는 데 사용된 관측값의 개수를 제공합니다 . 가중치 변수가 있는 경우에는 가중치의 합계입니다 .

적합 결여 보고서

참고 : " 적합 결여 " 보고서는 동일한 x 값을 갖는 행이 여러 개 있는 경우에만 표시됩니다 .

"적합 결여" 보고서를 사용하면 모형이 적절한지 여부에 관계없이 오차를 추정할 수 있습니다. 이 보고서는 동일한 x 값에서 여러 관측값이 나타나는 경우에 표시됩니다. 이러한 정확 반복 실험에서 측정되는 오차를 **순수 오차**라고 합니다. 이 값은 어떤 형태의 모형을 사용하더라도 설명하거나 예측할 수 없는 표본 오차의 비율입니다. 하지만 반복 실험되는 x 값이 매우 적어서 자유도가 매우 낮은 경우에는 적합 결여 검정이 그다지 유용하지 않을 수 있습니다.

그림 5.11 선형 적합 및 다항식 적합에 대한 적합 요약 보고서의 예



모형의 잔차와 순수 오차 사이의 차이를 **적합 결여 오차**라고 합니다. 회귀변수의 함수 형태가 잘못된 경우에는 적합 결여 오차가 순수 오차보다 매우 클 수 있습니다. 이 경우 다른 유형의 모형 적합을 시도해야 합니다. "적합 결여" 보고서에서는 적합 결여 오차가 0인지 여부를 검정합니다.

" 적합 결여 " 보고서에는 다음 열이 포함됩니다 .

소스 변동의 소스로는 **적합 결여** , **순수 오차** 및 **총 오차**의 세 가지가 있습니다 .

DF 각 오차 소스의 DF(**자유도**) 입니다 .

- **총 오차**의 DF 는 분산 분석 테이블의 **오차** 행 (" 분산 분석 보고서 "(123 페이지) 참조) 에 나오는 자유도입니다 . 이 값은 이 테이블에 나오는 **합계** DF 와 **모형** DF 사이의 차이입니다 . **오차** DF 는 적합 결여의 자유도와 순수 오차의 자유도로 분할됩니다 .
- **순수 오차** DF 는 각 효과에 대해 동일한 값을 갖는 여러 행이 있는 경우 각 그룹에서 합동화됩니다 . 자세한 내용은 " 적합 결여 보고서 "(149 페이지) 에서 확인하십시오 .
- **적합 결여** DF 는 **총 오차** DF 와 **순수 오차** DF 사이의 차이입니다 .

제곱합 각 오차 소스의 제곱합 (줄여서 SS 라고 함) 입니다 .

- **총 오차** SS 는 해당 분산 분석 테이블의 **오차** 행 (" 분산 분석 보고서 "(123 페이지) 참조) 에 나오는 제곱합입니다 .
- **순수 오차** SS 는 x 변수에 대해 동일한 값을 갖는 여러 행이 있는 경우 각 그룹에서 합동화됩니다 . 이 값은 모형 x 효과로 설명되지 않는 실제 랜덤 오차의 비율을 추정합니다 . 자세한 내용은 " 적합 결여 보고서 "(149 페이지) 에서 확인하십시오 .
- **적합 결여** SS 는 **총 오차** 제곱합과 **순수 오차** 제곱합 사이의 차이입니다 . 적합 결여 SS 가 크다면 해당 모형이 데이터에 적절하지 않은 것일 수 있습니다 . 아래에 설명된 F 비는 적합 결여로 인한 변동이 순수 오차의 무시 가능한 부분으로서 허용될 만큼 작은지 여부를 검정합니다 .

평균 제곱 제곱합을 관련 자유도로 나눈 것입니다 . 이 계산에 따라 제곱합이 평균 (평균 제곱) 으로 변환됩니다 . 통계 검정의 F 비는 평균 제곱의 비율입니다 .

F 비 적합 결여의 평균 제곱 대비 순수 오차의 평균 제곱에 대한 비율입니다 . 이 값은 적합 결여 오차가 0 이라는 가설을 검정합니다 .

Prob > F 적합 결여 분산과 순수 오차 분산으로 인한 변동이 동일한 경우 F 값이 더 클 확률입니다 . p 값이 높으면 적합 결여 정도가 크지 않은 것입니다 .

최대 R^2 모형에서 모형의 변수만으로 얻을 수 있는 최대 R^2 입니다 . 자세한 내용은 " 적합 결여 보고서 "(149 페이지) 에서 확인하십시오 .

분산 분석 보고서

회귀에 대한 ANOVA(분산 분석)에서는 표본의 총 변동을 여러 성분으로 분할합니다. 이러한 성분은 모형의 유효성을 평가하는 F 비를 계산하는 데 사용됩니다. F 비와 관련된 확률이 작으면 해당 모형은 반응 평균만 사용할 때보다 통계적 적합도가 더 높은 것으로 간주됩니다.

그림 5.12 의 "분산 분석" 보고서에서는 선형 적합(**선형 적합**)과 2 차 다항식 적합(**다항식 적합**)을 비교합니다. 두 적합 모두 평균 가로 선보다 통계적으로 더 낫습니다.

그림 5.12 선형 적합 및 다항식 적합에 대한 분산 분석 보고서의 예

선형 적합

2004 Verbal = 575.62539 - 101.32814*% Taking (2004)

▷ 적합 요약

▷ 적합 결여

◢ 분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	1	45265.424	45265.4	184.6379
오차	49	12012.733	245.2	Prob > F
수정 합계	50	57278.157		<.0001*

다항식 적합 차수 =2

2004 Verbal = 562.47113 - 111.07875*% Taking (2004) + 196.95725*(% Taking (2004)-0.39706)^2

▷ 적합 요약

▷ 적합 결여

◢ 분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	2	50371.160	25185.6	175.0265
오차	48	6906.997	143.9	Prob > F
수정 합계	50	57278.157		<.0001*

" 분산 분석 " 보고서에는 다음과 같은 열이 포함됩니다 .

소스 변동 소스로 , **모형** , **오차** 및 **수정 합계** 중 하나입니다 .

DF 각 변동 소스의 DF(자유도) 입니다 .

- 계산에 사용된 각 모수 추정값에 대해 비결측값의 총 개수 (*N*) 에서 자유도를 뺍니다 . 총 표본 변동의 계산에는 평균 추정값이 사용됩니다 . 따라서 합계에서 자유도 1 을 빼서 50 을 유지합니다 . 수정된 자유도의 합계는 **모형** 및 **오차** 항으로 분할됩니다 .
- 선형 적합의 경우 합계 중 자유도 1(**모형** 행에 표시) 이 단일 회귀 모수 (기울기) 를 추정하는 데 사용됩니다 . 2 차 다항식 적합의 경우에는 자유도 2 가 모수 (β_1 및 β_2) 를 추정하는 데 사용됩니다 .
- **오차** 자유도는 **수정 합계** DF 와 **모형** DF 사이의 차이입니다 .

제곱합 각 변동 소스의 제곱합 (줄여서 SS 라고 함) 입니다 .

- 이 예에서 각 반응과 표본 평균 사이의 거리에 대한 총 제곱합 (**수정 합계**) 은 그림 5.12 에 표시된 것과 같이 57,278.157 입니다 . 이 값은 다른 모든 모형과의 비교에 사용되는 기본 모형 (또는 단순 평균 모형) 의 제곱합입니다 .
- 선형 회귀의 경우 각 점과 적합선 사이의 거리에 대한 제곱합이 12,012.733 에서부터 줄어듭니다 . 이 값은 모형 적합 후의 잔차 또는 설명되지 않은 **(오차)** SS 입니다 . 2 차 다항식 적합의 잔차 SS 는 6,906.997 로 , 선형 적합보다 좀더 많은 변동을 설명합니다 . 즉 , 모형 SS 는 선형 적합보다 2 차 다항식 적합의 경우에 더 높으므로 이 모형은 더 많은 변동을 설명합니다 . **수정 합계** SS 에서 **오차** SS 를 뺀 값은 모형에 기인한 제곱합입니다 .

평균 제곱 제곱합을 관련 자유도로 나눈 것입니다 . 통계 검정의 *F* 비는 다음과 같은 평균 제곱의 비율입니다 .

- 선형 적합의 **모형** 평균 제곱은 45,265.4 입니다 . 이 값은 오차 분산을 추정하지만 , 단 모형 모수가 0 이라는 가설에 기반합니다 .
- **오차** 평균 제곱은 245.2 입니다 . 이 값은 오차 분산을 추정합니다 .

F 비 모형 평균 제곱을 오차 평균 제곱으로 나눈 것입니다 . 적합의 기본 가설은 절편을 제외한 모든 회귀 모수가 0 이라는 것입니다 . 이 가설이 참이면 오차의 평균 제곱과 모형의 평균 제곱이 모두 오차 분산을 추정하며 , 해당 비율은 *F* 분포를 따릅니다 .

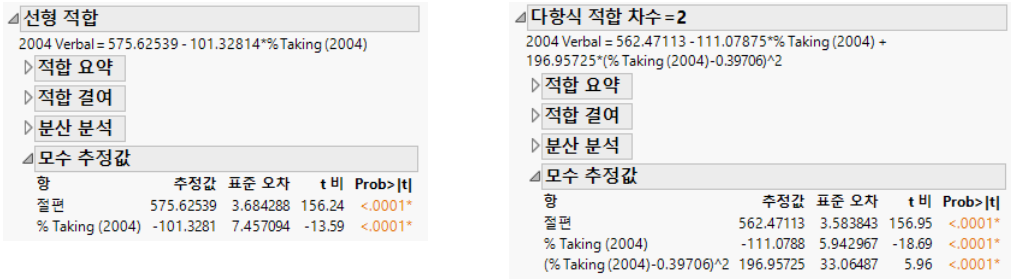
Prob > F 지정된 모형이 전체 반응 평균보다 더 잘 적합시키지 않는 경우 *F* 값이 더 클 것으로 관측된 유의 확률 (*p* 값) 입니다 . 관측된 유의 확률이 0.05 이하이면 회귀 효과의 증거로 간주됩니다 .

모수 추정값 보고서

선형 적합에 대한 "모수 추정값" 보고서의 절편 및 단일 *x* 변수입니다.

k 차 다항식 적합의 경우 모형 절편에 대한 추정값과 *X* 변수의 *k* 제곱 각각에 대한 모수 추정값이 있습니다.

그림 5.13 선형 적합 및 다항식 적합에 대한 모수 추정값 보고서의 예



" 모수 추정값 " 보고서에는 다음 열이 포함됩니다 .

항 요청된 모형의 각 모수 이름을 나열합니다 . 절편은 모든 모형에서 상수 항입니다 .

추정값 선형 모형의 모수 추정값을 나열합니다 . 예측 계산식은 이러한 추정값과 해당 변수 값의 선형 결합입니다 .

표준 오차 모수 추정값의 표준 오차에 대한 추정값을 나열합니다 . 이러한 값은 검정 및 신뢰 구간을 구성하는 데 사용됩니다 .

t 비 각 모수가 0 이라는 가설에 대한 검정 통계량을 나열합니다 . 이 값은 표준 오차에 대한 모수 추정값의 비율입니다 . 가설이 참이면 이 통계량은 스튜던트 *t* 분포를 따릅니다 .

Prob>|t| 각 *t* 비에서 계산된 관측된 유의 확률을 나열합니다 . 이 값은 귀무가설이 참일 때 *t* 비 (절대값 기준) 가 계산된 값보다 클 확률입니다 . 종종 0.05(일부 경우에는 0.01) 보다 작은 값은 모수가 0 과 유의하게 다르다는 증거로 해석됩니다 .

추가 통계량을 표시하려면 보고서에서 마우스 오른쪽 버튼을 클릭하고 **열** 메뉴를 선택합니다. 다음 통계량은 기본적으로 표시되지 않습니다.

95% 하한 모수 추정값에 대한 95% 신뢰 구간의 하한점입니다 .

95% 상한 모수 추정값에 대한 95% 신뢰 구간의 상한점입니다 .

표준화 베타 표준화된 모수 추정값입니다 . 이 값은 서로 다른 척도로 측정된 *X* 변수들의 효과를 비교하는 데 유용합니다 . 자세한 내용은 " 모수 추정값 보고서 "(150 페이지) 에서 확인하십시오 .

VIF 분산 팽창 계수입니다 .

설계 표준 오차 모수 추정값의 설계 표준 오차입니다 . 자세한 내용은 " 모수 추정값 보고서 "(150 페이지) 에서 확인하십시오 .

특수 적합

특수 적합 옵션을 사용하여 Y 및 X를 변환할 수 있습니다. 다양한 이유로 데이터를 변환할 수 있지만, 대개는 데이터가 좀더 충분히 정규 분포를 따르도록 하기 위해 변환을 수행합니다. 이렇게 하면 적절한 검정을 수행할 수 있습니다. 기울기와 절편을 제약하거나, 특정 차수의 다항식을 적합시키거나, 다항식을 중심화할 수도 있습니다.

참고 : 이 옵션의 예는 " 특수 적합 옵션의 예 "(138 페이지) 에서 확인하십시오 .

" 변환 또는 제약 조건 지정 " 창에는 다음과 같은 옵션이 포함되어 있습니다 .

Y 변환 또는 X 변환 Y 또는 X 변수를 변환하려면 다음 옵션 중 하나를 사용합니다 .

- 자연 로그
- 제곱근
- 제곱
- 역수
- 지수

차수 지정된 차수의 다항식을 적합시키려면 이 옵션을 사용합니다 .

중심화 다항식 다항식 중심화를 해제하려면 **중심화 다항식** 체크박스를 선택 취소합니다 (그림 5.19). *X* 변수의 변환에 대해서는 다항식 중심화가 수행되지 않습니다 . 다항식을 중심화하면 회귀 계수가 안정화되고 다중공선성이 감소합니다 .

절편을 다음 값으로 제약 모형 절편을 지정한 값으로 제약하려면 이 체크박스를 선택합니다 .

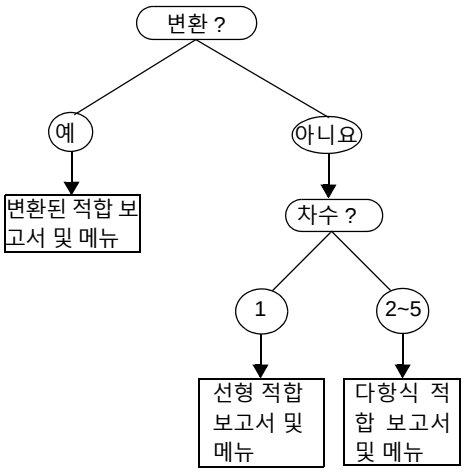
기울기를 다음 값으로 제약 모형 기울기를 지정한 값으로 제약하려면 이 체크박스를 선택합니다.

이 적합의 빨간색 삼각형 옵션에 대한 자세한 내용은 " 적합 메뉴 "(134 페이지) 에서 확인하십시오 .

특수 적합 보고서 및 메뉴

"특수 적합" 창에서 선택한 항목에 따라 보고서 및 메뉴가 다르게 표시됩니다. 그림 5.14 의 순서도에서는 선택 항목에 따라 표시되는 보고서 및 메뉴를 보여 줍니다.

그림 5.14 특수 적합 순서도의 예



변환된 적합 보고서

"변환된 적합" 보고서에는 "선형 적합 및 다항식 적합 보고서"(120 페이지)에 설명된 보고서가 포함됩니다. 하지만 *Y*를 변환한 경우에는 "원래 척도에서 적합" 보고서도 표시됩니다. 이 보고서에는 원래 *Y* 변수를 기준으로 한 적합 측도와 원래 척도로 다시 변환된 적합 모형이 표시됩니다.

유연

추정된 회귀 곡선의 평활도를 제어하려면 " 유연 " 메뉴의 옵션을 사용합니다 .

- " 스플라인 적합 " 에서는 벌점 최소 제곱 방법을 사용합니다 . 람다 모수를 사용하여 평활도를 조정할 수 있습니다 . 자세한 내용은 " 스플라인 적합 "(127 페이지) 에서 확인하십시오 .
- " 커널 평활기 " 는 로컬 가중 적합을 기반으로 합니다 . 알파 모수를 사용하여 로컬 동작의 영향을 제어할 수 있습니다 . 자세한 내용은 " 커널 평활기 "(128 페이지) 에서 확인하십시오 .
- " 각 값 적합 " 은 각 X 값에서의 평균 반응을 계산합니다 . 자세한 내용은 " 각 값 적합 "(129 페이지) 에서 확인하십시오 .

스플라인 적합

스플라인 적합 옵션을 사용하여 람다(λ) 값에 따라 평활도(또는 유연도)가 달라지는 평활 스플라인을 적합시킬 수 있습니다. 람다 값은 스플라인 계산식의 조정 모수입니다. λ 값이 감소하면 스플라인 모형의 오차 항 가중치가 커지며 적합은 더 유연해지고 곡선 형태가 됩니다. λ 값이 증가하면 적합이 덜 유연해져 곡선보다는 직선에 가깝게 됩니다.

다음 사항에 유의하십시오 .

- 적합 스플라인은 *X* 에 따른 *Y* 분포의 기대값을 확인하는 데 유용합니다 .
- 적합 곡선의 각 부분에 가장 근접한 점이 해당 부분에 가장 많은 영향을 미칩니다 . λ 값이 낮을수록 이 영향이 커져 매우 유연한 곡선이 생성됩니다 .
- 메뉴에 나열되지 않은 람다 값을 사용하려면 **스플라인 적합 > 기타**를 선택합니다 . *X* 변수의 척도가 변경되면 적합 모형도 변경됩니다. 이를 방지하려면 **표준화 X** 옵션을 선택합니다. 그러면 원래 *X* 변수나 척도화된 *X* 변수의 적합 모형이 동일하게 유지됩니다 .
- λ 값을 여러 개 시험해 보는 것이 유용할 수 있습니다 . " 평활 스플라인 적합 " 보고서 아래의 **람다** 슬라이더를 사용하여 다른 λ 값을 시험해 볼 수 있습니다 . 하지만 λ 는 데이터 척도에 따라 변동됩니다 . 예를 들어 인치 단위로 측정된 *X* 의 λ 값은 센티미터 단위로 측정된 *X* 의 λ 값과 같지 않습니다 .

"평활 스플라인 적합" 메뉴의 옵션에 대한 자세한 내용은 "적합 메뉴"(134 페이지)에서 확인하십시오. 이 적합에 대한 통계 상세 정보는 "스플라인 적합"(148 페이지)에서 확인하십시오.

평활 스플라인 적합 보고서

" 평활 스플라인 적합 " 보고서에는 스플라인 적합의 R² 와 **오차 제곱합**이 포함됩니다 . 이러한 값을 사용하여 스플라인 적합을 다른 적합과 비교하거나 여러 스플라인 적합을 서로 비교할 수 있습니다 .

R² 평활 스플라인 모형으로 설명되는 변동의 비율을 측정합니다 . 자세한 내용은 " 평활 적합 보고서 "(150 페이지) 에서 확인하십시오 .

오차 제곱합 각 점에서 적합 스플라인까지의 거리에 대한 제곱합입니다 . 이 값은 스플라인 모형을 적합시킨 후 설명되지 않은 오차 (**잔차**) 입니다 .

람다 변경 숫자를 직접 입력하거나 슬라이더를 이동하는 방법으로 λ 값을 변경할 수 있습니다 .

커널 평활기

커널 평활기 옵션은 영역 내의 표집된 점에서 단순 곡선 (선형 또는 2 차) 의 로컬 가중 적합을 반복적으로 찾아 구성된 곡선을 생성합니다 . 다수의 로컬 적합 (총 128 개) 을 결합하여 전체 영역에 대한 평활 곡선을 생성합니다 . 이 방법을 Loess 또는 Lowess 라고도 하며 , 이는 Locally Weighted Scatterplot Smoother, 즉 로컬 가중 산점도 평활기의 머리글자어입니다 . "커널 평활기 " 옵션은 완벽에 가깝게 적합되는 경우에 대한 작은 조정이 있는 Cleveland 방법 (1979) 을 구현합니다 . Cleveland 의 Biweight 함수에서 6 * q50 인수는 max(6 * q50, 2 * q90) 으로 대체됩니다 . 여기서 , q50 및 q90 은 각각 50 번째 및 90 번째 백분위수입니다 .

이 방법을 사용하면 변수 간의 관계를 빠르게 확인할 수 있으므로 수행할 분석 또는 적합 유형을 결정하는 데 도움이 됩니다 .

" 로컬 평활기 " 메뉴의 옵션에 대한 자세한 내용은 " 적합 메뉴 "(134 페이지) 에서 확인하십시오 .

로컬 평활기 보고서

" 로컬 평활기 " 보고서에는 평활기 적합의 R² 와 **오차 제곱합**이 포함됩니다 . 이러한 값을 사용하여 평활기 적합을 다른 적합과 비교하거나 여러 평활기 적합을 서로 비교할 수 있습니다 .

R² 평활기 모형으로 설명되는 변동의 비율을 측정합니다 . 자세한 내용은 " 평활 적합 보고서 "(150 페이지) 에서 확인하십시오 .

오차 제곱합 각 점에서 적합 평활기까지의 거리에 대한 제곱합입니다 . 이 값은 평활기 모형을 적합시킨 후 설명되지 않은 오차 (**잔차**) 입니다 .

로컬 적합 (람다) 각 로컬 적합의 다항식 차수를 선택합니다 . 2 차 다항식은 로컬 비평탄 정도를 보다 간편하게 추적할 수 있습니다 . 람다는 이 방법으로 적합된 다항식의 차수입니다 . 람다는 0, 1 또는 2 일 수 있습니다 .

가중치 함수 각 로컬 적합의 근방에 있는 데이터에 가중치를 적용할 방법을 지정합니다 . Loess 방법에서는 트라이큐브를 사용합니다 . 가중치 함수는 각 x_i 및 y_i 가 선형 적합에 미치는 영향을 결정합니다 . 이 영향은 x_i 와 x 의 거리가 멀어질수록 감소하다가 최종적으로 0 이 됩니다 .

평활 (알파) 각 로컬 적합의 대상이 되는 점의 수를 제어합니다 . 슬라이더를 사용하거나 값을 직접 입력합니다 . 알파는 평활 모수로 , 임의의 양수일 수 있지만 , 일반적인 값은 1/4 ~ 1 입니다 . 알파가 증가할수록 곡선이 평활해집니다 .

표집 델타 적합 프로세스에 사용되는 표집 양을 제어합니다 . 기본 표집 델타는 아무 점도 건너뛰지 않음을 의미하는 0 입니다 . 표집 델타를 늘리면 마지막 표본 점을 기준으로 지정된 델타 범위 내의 점들을 적합 프로세스에서 건너뛰게 됩니다 . 이 옵션을 사용하여 데이터가 조밀할 때 사용되는 점의 수를 줄일 수 있습니다 .

로버스트 적합 곡선에서 멀리 떨어진 점의 영향을 줄이기 위해 점의 가중치를 다시 적용합니다 . 이 프로세스를 반복할 횟수 (패스 수) 를 지정합니다 . 이 옵션의 목적은 이상치에 작은 가중치를 할당하여 곡선을 수렴시키고 이상치를 자동으로 제외하는 것입니다 .

각 값 적합

각 값 적합 옵션은 각각의 고유 *X* 값에 하나의 값을 적합시킵니다 . 적합된 값은 각각의 고유 *X* 값에 대한 반응 평균입니다 .

"각 값 적합" 메뉴의 옵션에 대한 자세한 내용은 " 적합 메뉴 "(134 페이지) 에서 확인하십시오 .

각 값 적합 보고서

" 각 값 적합 " 보고서에는 모형 적합에 대한 요약 통계량이 표시됩니다 .

관측값 수 관측값의 총 개수입니다 .

고유 값 수 고유 X 값의 개수입니다 .

자유도 순수 오차 자유도입니다 .

제곱합 순수 오차 제곱합입니다 .

평균 제곱 순수 오차 평균 제곱입니다 .

직교 적합

직교 적합 옵션은 *X* 와 *Y* 모두의 변동성을 설명하는 선형 모형을 적합시킵니다 . Deming 회귀라고도 합니다 .

참고 : 이 옵션의 예는 " 직교 적합 옵션의 예 "(140 페이지) 에서 확인하십시오 .

직교 적합 옵션

다음 옵션 중 하나를 선택하여 분산 비율을 지정할 수 있습니다 .

단변량 분산 , 주성분 *X* 및 *Y* 의 표본에서 계산된 단변량 분산 추정값을 사용합니다 . 이 값은 표준화된 첫 번째 주성분이 됩니다 . 측정 시스템 응용 프로그램에서는 오차 분산이 모집단 분산에 비례하지 않을 수 있으므로 이 옵션을 사용하지 않는 것이 좋습니다 .

등분산 분산 비율로 1 을 사용합니다 . 이 값은 오차 분산이 동일하다고 가정합니다 . 등분산을 사용하는 것은 표준화되지 않은 첫 번째 주성분 선을 사용하는 것과 동등합니다 . 예를 들어 산점도가 *X* 및 *Y* 방향에서 동일하게 척도화된 경우 , 정규 밀도 타원을 표시하면 이 선이 타원의 최장축이 됩니다 .

X 를 Y 에 적합 분산 비율 0 을 사용합니다 . 이 값은 *Y* 에 사실상 분산이 없음을 나타냅니다 .

지정된 분산 비율 비율을 원하는 대로 입력할 수 있으므로 *X* 의 측정 오차와 *Y* 의 반응 오차에 대해 알려진 정보를 사용할 수 있습니다 .

"직교 적합 비율" 메뉴의 옵션에 대한 자세한 내용은 "적합 메뉴"(134 페이지)에서 확인하십시오 . 이 적합에 대한 통계 상세 정보는 "직교 적합"(148 페이지)에서 확인하십시오 .

직교 적합 비율 보고서

" 직교 적합 비율 " 보고서에는 직교 회귀 모형에 대한 요약 통계량이 표시됩니다 .

변수 선을 적합시키는 데 사용되는 변수의 이름입니다 .

평균 각 변수의 평균입니다 .

표준편차 각 변수의 표준편차입니다 .

분산 비율 선을 적합시키는 데 사용되는 분산 비율입니다 .

상관 두 변수 간의 상관계수입니다 .

절편 적합선의 절편입니다 .

기울기 적합선의 기울기입니다 .

CL 하한 기울기의 신뢰 하한입니다 .

CL 상한 기울기의 신뢰 상한입니다 .

알파 신뢰 구간을 계산하는 데 사용되는 유의 수준을 입력합니다 .

로버스트

로버스트 옵션은 데이터 집합에서 이상치의 영향을 줄이기 위한 두 가지 방법을 제공합니다 . 이상치는 부정확한 추정 및 결정을 초래할 수 있습니다 .

참고 : 이 옵션의 예는 " 로버스트 적합 명령의 예 "(142 페이지) 에서 확인하십시오 .

"로버스트 적합" 및 "Cauchy 적합" 메뉴의 옵션에 대한 자세한 내용은 "적합 메뉴"(134 페이지)에서 확인하십시오 .

로버스트 적합

"로버스트 적합" 옵션은 반응 변수에서 이상치의 영향을 줄입니다. Huber M-추정 방법이 사용됩니다. Huber M-추정은 다음과 같이 Huber 손실 함수를 최소화하는 모수 추정값을 찾는 방법입니다.

$$l(\underline{e}) = \sum_i \rho(e_i)$$

다음은 각 요소에 대한 설명입니다.

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & \text{단, } |e| < k \\ k|e| - \frac{1}{2}k^2 & \text{단, } |e| \geq k \end{cases}$$

e_i는 잔차를 나타냅니다.

Huber 손실 함수는 이상치에 벌점을 부과하며, 작은 오차의 경우에는 2차 형태로 증가하고 큰 오차의 경우에는 선형적으로 증가합니다. JMP 구현에서는 $k = 2$입니다. 로버스트 적합에 대한 자세한 내용은 Huber 연구 자료(1973) 및 Huber and Ronchetti 연구 자료(2009)에서 확인하십시오.

Cauchy 적합

Cauchy 옵션은 최대 가능도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다. 이 방법에서는 오차가 Cauchy 분포를 따른다고 가정합니다. Cauchy 분포는 정규 분포보다 꼬리가 더 두꺼우므로 이상치의 영향이 줄어듭니다. 이 옵션은 데이터에 포함된 이상치의 비율이 높은 경우에 유용할 수 있습니다. 하지만 데이터가 정규 분포에 가깝고 이상치가 일부만 있는 경우에 이 옵션을 사용하면 추론이 잘못될 수 있습니다.

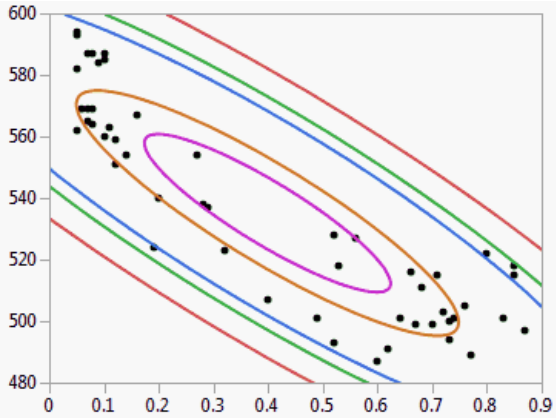
밀도 타원

밀도 타원 옵션을 사용하면 지정된 점 집합을 포함하는 타원을 그릴 수 있습니다. 점의 개수는 **밀도 타원** 메뉴에서 선택하는 확률에 따라 결정됩니다.

참고 : 이 옵션의 예는 "밀도 타원을 사용한 그룹화의 예"(144 페이지)에서 확인하십시오.

"이변량 정규 타원" 메뉴의 옵션에 대한 자세한 내용은 "적합 메뉴"(134 페이지)에서 확인하십시오.

그림 5.15 밀도 타원의 예



밀도 타원은 *X* 및 *Y* 변수에 대한 이변량 정규 분포 적합을 통해 계산됩니다. 이변량 정규 밀도는 *X* 및 *Y* 변수의 평균 및 표준편차와 변수 사이의 상관계수를 사용하는 함수입니다. **기타**를 선택하면 0보다 크고 1보다 작거나 같은 범위에서 원하는 확률을 지정할 수 있습니다.

이러한 타원은 밀도 등고선이자 신뢰 곡선입니다. 신뢰 곡선으로 볼 때 밀도 타원은 이변량 정규 분포를 따른다는 가정하에 지정된 백분율의 데이터가 놓일 것으로 기대되는 영역을 보여 줍니다.

밀도 타원은 두 변수 사이의 상관관계를 나타내는 데 유용한 그래픽입니다. 두 변수 간의 상관계수가 1 또는 –1에 가까워질수록 타원이 대각선 방향으로 축소합니다. 두 변수 간에 상관관계가 낮을수록 타원이 원에 가까워지고 대각선 방향의 기울어짐이 덜하게 됩니다.

상관 보고서

각 **밀도 타원** 적합에 대한 상관 보고서에는 *X* 및 *Y* 변수의 상관계수가 표시됩니다.

참고 : 여러 변수 쌍에 대한 밀도 타원 및 상관계수를 행렬 형식으로 보려면 **분석 > 다변량 방법** 메뉴에서 **다변량** 플랫폼을 사용합니다 .

변수 타원을 생성하는 데 사용된 변수의 이름입니다 .

평균 *X* 및 *Y* 변수 모두의 평균입니다 .

표준편차 *X* 및 *Y* 변수 모두의 표준편차입니다 .

평균과 표준편차에 대한 자세한 내용은 " 분포 " 장의 " 요약 통계량 보고서 "(42 페이지) 섹션에서 확인하십시오 .

상관 Pearson 상관계수입니다 . 두 변수가 정확히 선형 관계에 있는 경우 상관계수는 해당 변수 간 관계가 양의 관계인지 음의 관계인지에 따라 1 또는 –1 입니다 . 관계가 없으면 상관계수가 0 에 가까워집니다 . 자세한 내용은 " 상관 보고서 "(150 페이지) 에서 확인하십시오 .

유의 확률 *X* 변수와 *Y* 변수 간에 선형 관계가 있지 않은 경우 상관계수의 절대값이 계산된 값보다 더 클 확률입니다.

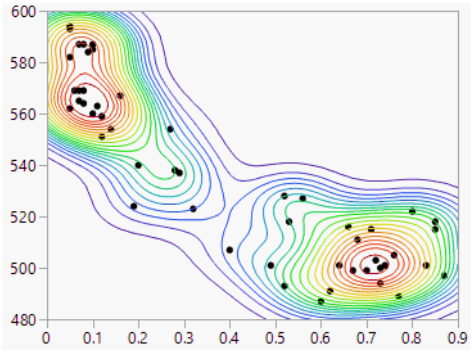
개수 계산에 사용된 관측값의 개수입니다.

비모수 밀도

그림에 표시되는 점이 수천 개이면 점 집합이 너무 진하게 표시되어 밀도 패턴을 알아보기 어려울 수 있습니다. "비모수 밀도" 옵션을 사용하면 패턴을 확인하기가 쉬워집니다.

"비모수 밀도"는 데이터 점의 밀도를 설명하는 평활 비모수 이변량 표면을 추정합니다. 그림에는 밀도를 보여 주는 일련의 등고선이 추가됩니다(그림 5.16). 이 등고선은 5% 간격의 분위수 등고선입니다. 즉, 추정된 비모수 분포에서 생성된 점 중 약 5%가 최저 등고선 아래에 있고 다음 등고선 아래에 10%가 있는 식입니다. 최고 등고선 아래에는 약 95%의 점이 있습니다.

그림 5.16 비모수 밀도의 예



비모수 밀도 등고선 격자의 크기를 변경하려면 "이변량"의 빨간색 삼각형 메뉴에서 Shift 키를 누른 채로 **비모수 밀도**를 선택합니다. 기본값인 102보다 큰 값을 입력합니다.

"분위수 밀도 등고선" 메뉴의 옵션에 대한 자세한 내용은 "적합 메뉴"(134페이지)에서 확인하십시오.

분위수 밀도 등고선 보고서

"분위수 밀도 등고선" 보고서에는 비모수 밀도를 생성하는 데 사용된 표준편차가 표시됩니다.

그룹화 기준

그룹화 기준 옵션을 사용하면 분류(그룹화) 변수를 선택할 수 있습니다. 그룹화 변수가 적용되면 이변량 플랫폼에서는 그룹화 변수의 각 수준에 대해 개별 분석을 수행합니다. 산점도에는 각 그

룹화 변수에 대한 적합 모형(선, 곡선, 등고선 또는 타원)이 표시됩니다. 그룹화 변수의 각 수준에 대한 적합은 산점도 아래에서 개별 팝업 메뉴를 사용하여 확인할 수 있습니다.

그룹화 변수가 적용된 경우에는 "이변량 적합"의 빨간색 삼각형 메뉴에서 **그룹화 기준** 옵션이 선택된 상태로 표시됩니다. 그룹화 변수를 변경하려면 먼저 **그룹화 기준** 옵션을 선택하여 기존 변수를 제거(선택 해제)합니다. 그런 다음 **그룹화 기준** 옵션을 다시 선택하고 표시되는 창에서 이전과 같이 처리합니다.

그룹화 기준 옵션을 사용하면 그룹화 변수별로 별도의 분석을 생성하되 적합 그림은 동일한 산점도에 표시할 수 있습니다. 이렇게 하면 그룹별 적합을 시각적으로 비교할 수 있습니다.

참고 : 이 옵션의 예는 " 밀도 타원을 사용한 그룹화의 예 "(144 페이지) 및 " 회귀선을 사용한 그룹화의 예 "(145 페이지) 에서 확인하십시오 .

적합 메뉴

적합 옵션을 선택할 때마다 해당 보고서와 함께 적합 메뉴가 보고서 창에 추가됩니다. 다음 표에서는 각 적합 옵션에 해당하는 적합 메뉴를 보여 줍니다.

적합 옵션	적합 메뉴
평균 적합	평균 적합
선형 적합	선형 적합
다항식 적합	다항식 적합 차수=X*
특수 적합	선형 적합 다항식 적합 차수=X* 변환된 적합 X* 제약된 적합
스플라인 적합	평활 스플라인 적합, 람다=X*
커널 평활기	로컬 평활기
각 값 적합	각 값 적합
직교 적합	직교 적합 비율=X*
로버스트 적합	로버스트 적합
Cauchy 적합	Cauchy 적합
밀도 타원	이변량 정규 타원 P=X*

적합 옵션	적합 메뉴
비모수 밀도	분위수 밀도 등고선

*X= 변수 (문자 또는 숫자)

적합 메뉴 옵션

적합 메뉴 옵션은 선택한 적합에 따라 다릅니다 .

- " 대부분의 적합에 적용되는 옵션 "(135 페이지)
- " 다중 적합에 적용되는 옵션 "(135 페이지)
- " 이변량 정규 타원에 적용되는 옵션 "(136 페이지)
- " 분위수 밀도 등고선에 적용되는 옵션 "(137 페이지)

대부분의 적합에 적용되는 옵션

적합선 모형 적합을 설명하는 선 또는 곡선을 표시하거나 숨깁니다 . " 이변량 정규 타원 " 보고서의 경우 이 옵션은 등고선 테두리를 나타나는 타원을 표시하거나 숨깁니다 . " 분위수 밀도 등고선 " 의 경우에는 사용할 수 없습니다 .

선 색상 색상 팔레트에서 각 적합에 할당할 색상을 선택할 수 있습니다 . " 분위수 밀도 등고선 " 의 경우에는 사용할 수 없습니다 .

선 스타일 선 스타일 팔레트에서 각 적합의 선 스타일을 선택할 수 있습니다 . " 분위수 밀도 등고선 " 의 경우에는 사용할 수 없습니다 .

선 너비 적합선의 선 너비를 변경할 수 있습니다 . 기본 선 너비는 가장 가는 선입니다 . " 분위수 밀도 등고선 " 의 경우에는 사용할 수 없습니다 .

보고서 적합 보고서 표시를 설정하거나 해제합니다 . 이변량 그림은 수정하지 않습니다 .

적합 제거 그래프에서 적합을 제거하고 해당 보고서도 제거합니다 .

다중 적합에 적용되는 옵션

적합 신뢰 곡선 기대값 (평균) 에 대한 신뢰 한계를 표시하거나 숨깁니다 . 이 옵션은 " 스플라인 적합 ", " 밀도 타원 ", " 각 값 적합 " 및 " 직교 적합 " 의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다 .

개별값 신뢰 곡선 개별 예측값의 신뢰 한계를 표시하거나 숨깁니다 . 신뢰 한계는 오차의 변동과 모수 추정값의 변동을 반영합니다 . 이 옵션은 " 평균 적합 ", " 스플라인 적합 ", " 밀도 타원 ", " 각 값 적합 " 및 " 직교 적합 " 의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다 .

예측값 저장 현재 데이터 테이블에 **예측 < 열 이름 >** 이라는 새 열을 생성합니다 . 여기서 **열 이름** 은 *Y* 변수의 이름입니다 . 이 열에는 예측 계산식과 계산된 표본 예측값이 포함됩니다 . 예측

계산식은 사용자가 테이블에 추가하는 행에 대해 값을 자동으로 계산합니다 . 이 옵션은 " 각 값 적합 " 및 " 밀도 타원 " 의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다 .

각 적합에 **예측값 저장** 및 **잔차 저장** 옵션을 사용할 수 있습니다 . 이러한 옵션을 여러 번 사용하거나 그룹화 변수와 함께 사용하는 경우에는 데이터 테이블에 생성되는 결과 열의 이름을 각 적합에 맞게 변경하는 것이 좋습니다 .

잔차 저장 현재 데이터 테이블에 **잔차 < 열 이름 >** 이라는 새 열을 생성합니다 . 여기서 **열 이름**은 *Y* 변수의 이름입니다 . 각 값은 실제값 (관측값) 과 예측값 사이의 차이입니다 . **예측값 저장** 옵션과 달리 이 옵션은 새 열에 계산식을 생성하지 않습니다 . 이 옵션은 " 각 값 적합 " 및 " 밀도 타원 " 의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다 .

각 적합에 **예측값 저장** 및 **잔차 저장** 옵션을 사용할 수 있습니다 . 이러한 옵션을 여러 번 사용하거나 그룹화 변수와 함께 사용하는 경우에는 데이터 테이블에 생성되는 결과 열의 이름을 각 적합에 맞게 변경하는 것이 좋습니다 .

스튜던트화 잔차 저장 잔차를 잔차의 표준 오차로 나눈 결과를 포함하는 새 열을 데이터 테이블에 생성합니다 .

평균 신뢰 한계 계산식 평균 신뢰 구간의 계산식을 포함하는 새 열을 데이터 테이블에 생성합니다 .

개별값 신뢰 한계 계산식 개별 신뢰 구간의 계산식을 포함하는 새 열을 데이터 테이블에 생성합니다 .

잔차 그림 (선형 , 다항식 및 특수 적합만 해당) 5 가지 진단 그림 , 즉 잔차 대 예측값 그림 , 실제값 대 예측값 그림 , 잔차 대 행 번호 그림 , 잔차 대 X 그림 , 잔차 정규 분위수 그림을 생성합니다 . 자세한 내용은 " 진단 그림 "(137 페이지) 에서 확인하십시오 .

α **수준 설정** 다양한 적합의 신뢰 대역을 계산하는 데 사용되는 유의 수준을 설정할 수 있습니다 .

적합 신뢰 구간 음영 **적합 신뢰 곡선** 옵션과 동일한 곡선을 그리되 곡선 사이의 영역에 음영을 적용합니다 .

개별값 신뢰 구간 음영 **개별값 신뢰 곡선** 옵션과 동일한 곡선을 그리되 곡선 사이의 영역에 음영을 적용합니다 .

계수 저장 스플라인 계수를 X, A, B, C 및 D 라는 열을 포함하는 새 데이터 테이블로 저장합니다 . X 열에는 매듭 점이 포함됩니다 . A, B, C 및 D 는 절편과 3 차 다항식의 1 차 , 2 차 및 3 차 계수입니다 . 이러한 계수는 X 열의 해당 값에서 다음으로 높은 값까지 확장됩니다 .

이변량 정규 타원에 적용되는 옵션

음영 등고선 밀도 타원의 내부 영역에 음영을 적용합니다 .

내부 점 선택 타원 내부의 점을 선택합니다 .

외부 점 선택 타원 외부의 점을 선택합니다 .

분위수 밀도 등고선에 적용되는 옵션

커널 제어 각 변수에 대한 슬라이더를 표시합니다. 슬라이더로 등고선 밀도를 결정하기 위한 *X* 및 *Y* 값 범위를 정의하는 표준편차를 변경할 수 있습니다.

5% 등고선 5% 등고선을 표시하거나 숨깁니다.

등고선 10% 등고선을 표시하거나 숨깁니다.

등고선 채우기 등고선 사이의 영역을 채웁니다.

색상 테마 등고선의 색상 테마를 변경합니다.

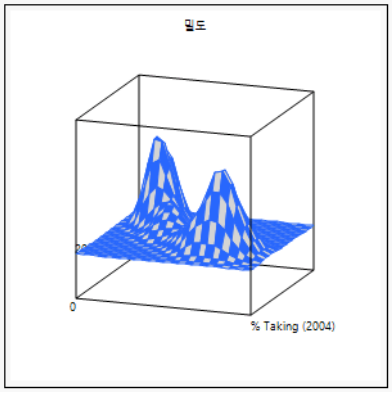
밀도별 점 선택 사용자가 지정한 분위수 범위에 있는 점을 선택합니다.

밀도 분위수별 색상 밀도에 따라 다른 점 색상을 적용합니다.

밀도 분위수 저장 각 점의 밀도 분위수를 포함하는 새 열을 생성합니다.

그물 그림 두 분석 변수의 격자 위에 표시되는 3 차원 밀도 그림입니다.

그림 5.17 그물 그림의 예



최빈 군집화 현재 데이터 테이블에 새 열을 생성하고 해당 열에 군집 값을 채웁니다.

> **참고 :** 최빈 군집화 값을 먼저 저장한 다음 밀도 격자를 저장하면 격자 테이블에 군집 값도 포함됩니다. 군집 값은 그림의 점에 색상 및 표식을 적용하는 데 유용합니다.

밀도 격자 저장 밀도 추정값과 관련 분위수를 새 데이터 테이블에 저장합니다. 이 격자 데이터를 사용하여 밀도를 3D 산점도나 등고선 그림 플랫폼에서처럼 다른 방식으로 시각화할 수 있습니다.

진단 그림

잔차 그림 옵션은 잔차 그림과 모형 적합을 진단하기 위한 다른 그림을 생성합니다. 다음과 같은 그림을 사용할 수 있습니다.

잔차 대 예측값 그림 잔차 대 예측값의 그림입니다. 잔차 히스토그램도 생성됩니다.

실제값 대 예측값 그림 실제값 대 예측값의 그림입니다 .

잔차 대 행 번호 그림 잔차 값 대 행 번호의 그림입니다 .

잔차 대 X 그림 잔차 값 대 X 변수의 그림입니다 .

잔차 정규 분위수 그림 잔차의 정규 분위수 그림입니다 .

이변량 플랫폼의 추가 예

다음 섹션에서는 다양한 이변량 적합의 예를 제공합니다.

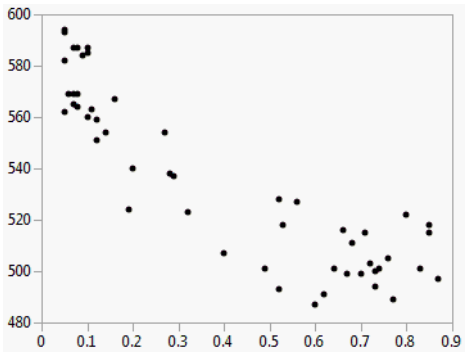
- " 특수 적합 옵션의 예 "
- " 직교 적합 옵션의 예 "
- " 로버스트 적합 명령의 예 "
- " 밀도 타원을 사용한 그룹화의 예 "
- " 회귀선을 사용한 그룹화의 예 "
- " 기준 변수를 사용한 그룹화의 예 "

특수 적합 옵션의 예

이 예에서는 데이터에 변환을 추가하는 방법을 보여 줍니다 . Y 를 로그로 변환하고 X 를 제곱근으로 변환하는 것부터 시작합니다 .

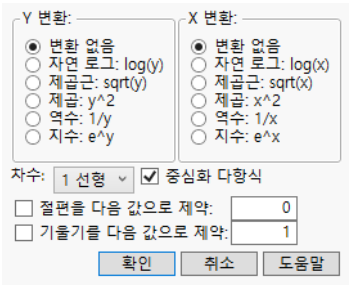
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 SAT.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. 2004 Verbal 을 선택하고 **Y, 반응**을 클릭합니다 .
4. % Taking (2004) 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 5.18 응시율별 SAT 점수의 예



6. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **특수 적합**을 선택합니다 . " 변환 또는 제약 조건 지정 " 창이 나타납니다 . 이 창에 대한 자세한 내용은 " 특수 적합 "(125 페이지) 에서 확인하십시오 .

그림 5.19 변환 또는 제약 조건 지정 창



7. "Y 변환 " 에서 **자연 로그 : log(y)** 를 선택합니다 .
8. "X 변환 " 에서 **제곱근 : sqrt(x)** 를 선택합니다 .
9. **확인**을 클릭합니다 .

그림 5.20 특수 적합 보고서의 예

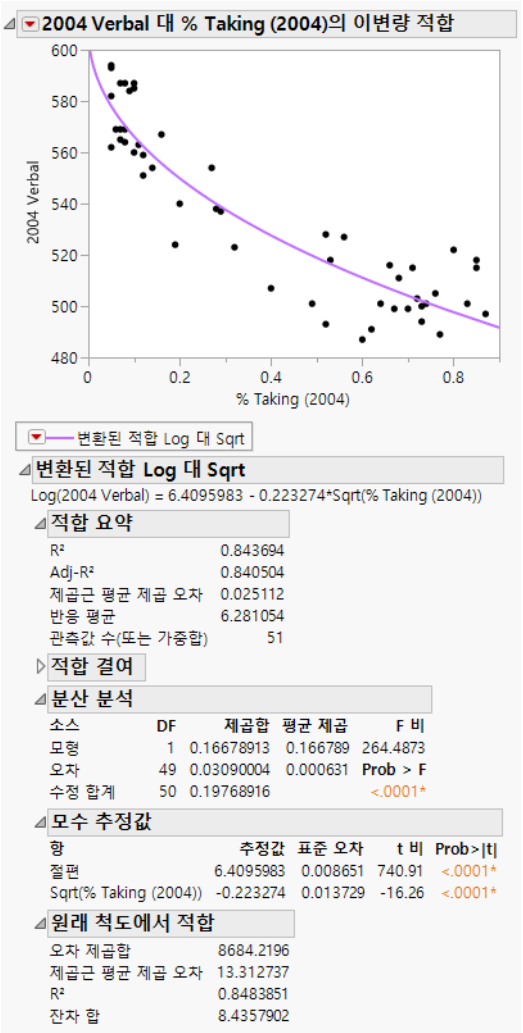


그림 5.20에서는 원래 척도로 표시된 적합선을 보여 줍니다. 표시된 선은 점들이 몰려 있는 곳을 통과하므로 이 모형은 데이터를 잘 적합시키는 것으로 보입니다.

직교 적합 옵션의 예

이 예에서는 분포 플랫폼을 사용하여 변수를 표준화한 다음 표준화된 변수를 사용하여 직교 모형을 적합시키는 방법을 보여 줍니다.

변수 표준화

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.

2. **분석 > 분포**를 선택합니다 .
3. height 및 weight 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. Ctrl 키를 누른 채로 있습니다 . height 의 빨간색 삼각형을 클릭하고 **저장 > 표준화**를 선택합니다 .

   Ctrl 키를 누른 채로 작업을 수행하면 해당 작업이 보고서 창의 모든 변수에 일괄 적용됩니다 . Big Class.jmp 샘플 데이터 테이블에는 두 개의 새 열이 추가되었습니다 .
6. " 분포 " 보고서 창을 닫습니다 .

표준화된 변수를 사용하여 직교 모형 적합

1. Big Class.jmp 샘플 데이터 테이블에서 **분석 > X 로 Y 적합**을 선택합니다 .
2. 표준화 weight 를 선택하고 **Y, 반응**을 클릭합니다 .
3. 표준화 height 를 선택하고 **X, 요인**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 표준화 weight 대 표준화 height 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 .
6. " 표준화 weight 대 표준화 height 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **직교 적합**을 선택합니다 . 그런 후 다음을 각각 선택합니다 .
   - **등분산**
   - **X 를 Y 에 적합**
   - **지정된 분산 비율** (0.2 입력)
   - **지정된 분산 비율** (5 입력)

그림 5.21 직교 적합 옵션의 예

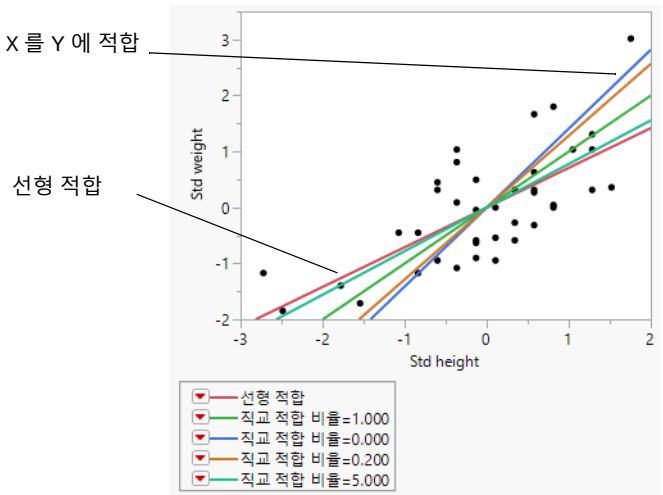


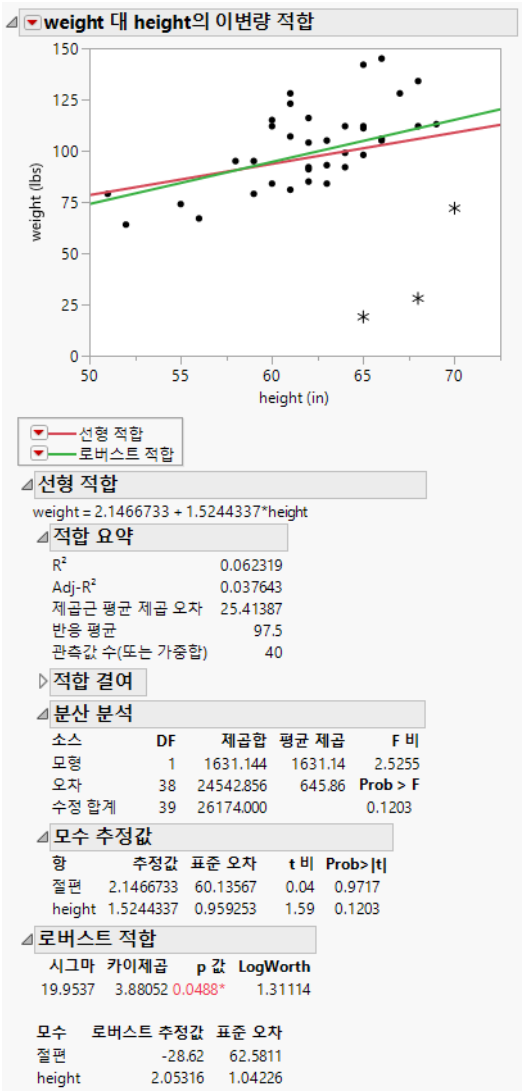
그림 5.21 의 산점도에서는 표준화된 height 및 weight 값과 함께 선택한 직교 분산 비율의 동작을 나타내는 다양한 선형 적합을 보여 줍니다. 표준 선형 회귀(**선형 적합**)는 *X* 변수의 분산이 매우 작은 것으로 간주될 때 발생합니다. **X를 Y에 적합**은 반대로 *Y* 변수의 분산이 무시될 때의 극단 적합입니다. 다른 모든 선은 이 두 극단 사이에 존재하며 분산 비율이 변경되면 선이 이동합니다. 분산 비율이 높을수록 *Y* 반응의 변동이 우월하게 되고 적합선의 기울기가 *X*로 *Y* 적합에 가까워집니다. 마찬가지로, 비율을 낮추면 선의 기울기가 *Y*로 *X* 적합에 가까워집니다.

로버스트 적합 명령의 예

이 예에서는 Huber M- 추정 방법을 사용하여 로버스트 모형을 적합시키는 방법을 보여 줍니다. Weight Measurements.jmp 샘플 데이터 테이블의 데이터는 40 명의 학생을 대상으로 한 키 및 체중 측정값을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Weight Measurements.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. weight 를 선택하고 **Y, 반응**을 클릭합니다.
4. height 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "weight 대 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.
7. "weight 대 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **로버스트 > 로버스트 적합**을 선택합니다.

그림 5.22 로버스트 적합의 예



표준 분산 분석 보고서를 보면 p 값이 0.1203이므로 키와 체중이 선형 관계에 있지 않다고 잘못된 결론을 내릴 수 있습니다. 하지만 "로버스트 적합" 보고서를 보면 여기서는 p 값이 0.0489이므로 키와 체중이 선형 관계에 있다는 결론을 내릴 수 있습니다. 일부 측정값은 비정상적으로 낮은데 이는 사용자 입력이 잘못된 데 기인한 것일 수 있습니다. 이러한 측정값은 분석에 큰 영향을 주지 않았습니다.

밀도 타원을 사용한 그룹화의 예

이 예에서는 그룹화 (기준) 변수를 사용하고 데이터에 밀도 타원을 추가하는 방법을 보여 줍니다 . Hot Dogs.jmp 샘플 데이터 테이블에서 Type 열은 핫도그의 세 가지 종류 (소고기 , 돼지고기 또는 가금육) 를 나타냅니다 . 이 세 종류의 핫도그를 비용 변수에 따라 그룹화하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hot Dogs.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. $/oz 를 선택하고 **Y, 반응**을 클릭합니다 .
4. $/lb Protein 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "$/oz 대 $/lb Protein 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **그룹화 기준**을 선택합니다 .
7. 목록에서 Type 을 선택합니다 .
8. **확인**을 클릭합니다 .

   그룹화 기준 옵션을 다시 보면 그 옆에 체크 표시가 표시되어 있습니다 .
9. "$/oz 대 $/lb Protein 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **밀도 타원 > 0.90** 을 선택합니다 .

Type 에 따라 다른 점 색상을 적용합니다 .

10. 산점도를 마우스 오른쪽 버튼으로 클릭하고 **행 범례**를 선택합니다 .
11. 열 목록에서 Type 을 선택하고 **확인**을 클릭합니다 .

그림 5.23 그룹화의 예

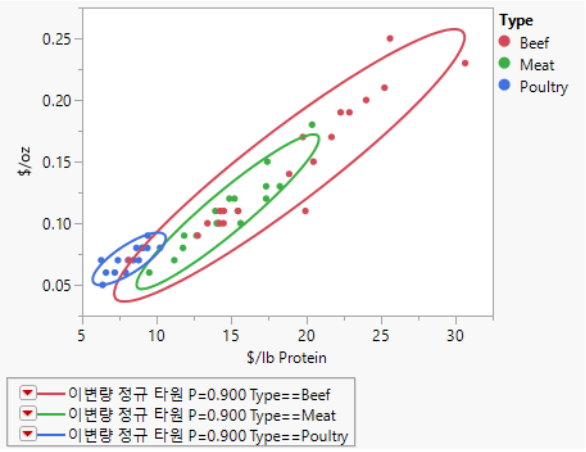


그림 5.23 의 타원은 서로 다른 종류의 핫도그들이 비용 변수에 따라 어떻게 군집화되는지를 명확히 보여 줍니다 .

회귀선을 사용한 그룹화의 예

이 예에서는 그룹화 변수를 통해 회귀선을 중첩 표시하여 서로 다른 그룹의 기울기를 비교하는 방법을 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. height 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 5.24 의 왼쪽 예와 같은 그림을 생성하려면 다음을 수행하십시오 .

6. "weight 대 height 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 .

그림 5.24 의 오른쪽 예와 같은 그림을 생성하려면 다음을 수행하십시오 .

7. " 선형 적합 " 메뉴에서 **적합 제거**를 선택합니다 .
8. "weight 대 height 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **그룹화 기준**을 선택합니다 .
9. 목록에서 sex 를 선택합니다 .
10. **확인**을 클릭합니다 .
11. "weight 대 height 의 이변량 적합 " 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 .

그림 5.24 전체 표본 및 그룹화된 표본에 대한 회귀 분석의 예

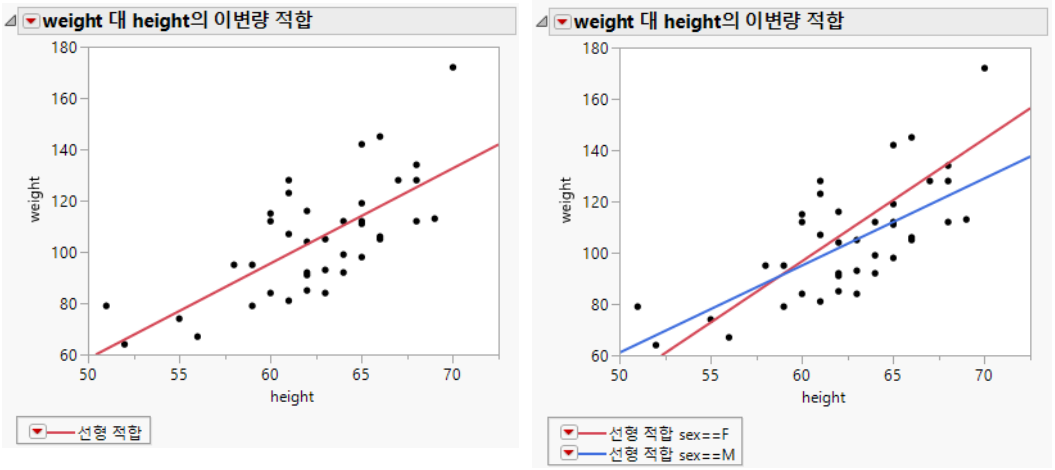


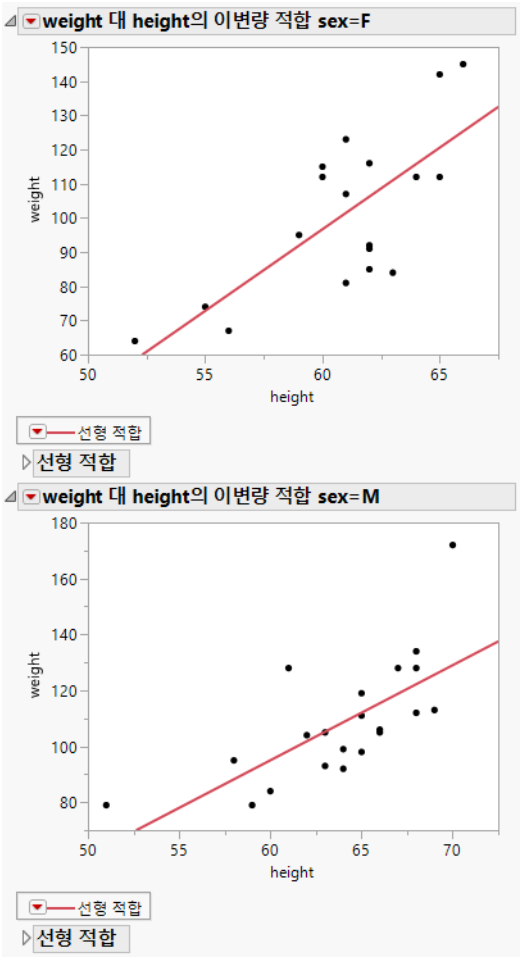
그림 5.24의 왼쪽 산점도에는 체중과 키의 상관관계를 나타내는 단일 회귀선이 있습니다. 오른쪽 산점도에서는 남학생과 여학생의 회귀선을 개별적으로 보여 줍니다.

기준 변수를 사용한 그룹화의 예

이 예에서는 시작 창에서 기준 변수를 지정하여 그룹화하는 또 다른 방법을 보여 줍니다 . 그러면 기준 변수 (또는 여러 기준 변수의 조합) 의 각 수준별로 별도의 보고서와 그래프가 표시됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. height 를 선택하고 **X, 요인**을 클릭합니다 .
5. sex 를 선택하고 **기준**을 클릭합니다 .
6. **확인**을 클릭합니다 .
7. Ctrl 키를 누른 채로 "weight 대 height 의 이변량 적합 sex =F" 옆의 빨간색 삼각형 메뉴를 클릭하고 빨간색 삼각형 메뉴에서 **선형 적합**을 선택합니다 .

그림 5.25 기준 변수 그림의 예



기준 변수(sex)의 각 수준에 대해 개별 분석이 표시됩니다. 따라서 여학생에 대한 산점도와 남학생에 대한 산점도가 표시됩니다.

이변량 플랫폼에 대한 통계 상세 정보

- "선형 적합"
- "스플라인 적합"
- "직교 적합"
- "적합 요약 보고서"
- "적합 결여 보고서"

- "모수 추정값 보고서"
- "평활 적합 보고서"
- "상관 보고서"

선형 적합

선형 적합 옵션은 잔차 제곱합을 최소화하도록 점을 적합시키는 직선의 모수 β_0 및 β_1을 찾습니다. i번째 행에 대한 모형은 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$로 기록됩니다.

2 차 다항식은 포물선으로 나타나고 3 차 다항식은 3 차 곡선으로 나타납니다. k 차의 경우 i 번째 관측값에 대한 모형은 다음과 같습니다.

$$y_i = \sum_{j=0}^{k} \beta_j x_i^j + \varepsilon_i$$

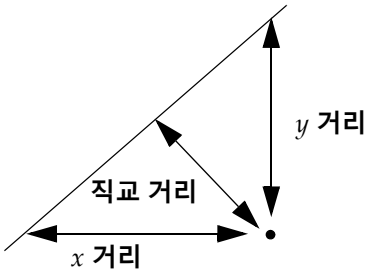
스플라인 적합

3차 스플라인 방법에서는 결과 곡선이 연속적이고 부드럽게 되도록 매듭지어진 일련의 3차 다항식이 사용됩니다. 추정은 오차 제곱합과 곡선 범위에 통합된 곡률에 대한 벌점의 조합인 목적 함수를 최소화하는 방법으로 수행됩니다. 이 방법에 대한 자세한 내용은 Reinsch 연구 자료 (1967) 또는 Eubank 학술 자료 (1999)에서 확인하십시오.

직교 적합

표준 최소 제곱 적합 방법에서는 X 변수가 고정되어 있고 Y 변수는 X + 오차에 따라 달라진다고 가정합니다. X 측정값에 랜덤 변동이 있으면 수직 거리 차의 제곱합을 최소화하는 선을 적합시켜야 합니다(그림 5.26). 하지만 수직 거리는 X 및 Y의 척도화 방법에 따라 달라지며, 수직 거리의 척도화는 그래픽 문제가 아니라 통계적 문제로 남아 있습니다.

그림 5.26 적합선에 수직인 선



적합을 수행하려면 X의 오차 분산에 대한 Y의 오차 분산의 비율을 지정해야 합니다. 이는 표본점의 분산이 아니라 오차의 분산이므로 신중하게 선택해야 합니다. 표준 최소 제곱에서는 σ_x^2이 0이기 때문에 $(\sigma_y^2)/(\sigma_x^2)$ 비율이 무한대가 됩니다. 오차 비율이 큰 경우에 직교 적합을 수행하면

적합선이 표준 최소 제곱 적합선에 가까워집니다. 비율을 0으로 지정하면 X에 대한 Y가 아니라 Y에 대한 X의 회귀와 동등한 적합이 수행됩니다.

이 방법의 가장 일반적인 용도는 동일한 값을 측정하는 데 오차가 있는 두 측정 시스템을 비교하는 것입니다. 따라서 Y 반응 오차와 X 측정 오차가 모두 동일한 유형의 측정 오차입니다. 측정 오차 분산은 어디에서 구할까요? 이변량 데이터에서는 어느 측정 시스템에서 어떤 비율의 오차가 나타나는지 알 수 없으므로 이를 구할 수 없습니다. 따라서 1과 같은 비율을 임의로 가정하거나, 두 측정 시스템에 의한 동일한 단위의 개별 반복 측정에 의존해야 합니다.

이 방법을 사용하면 Y와 X 모두의 예측값을 계산할 수 있다는 장점이 있습니다. 예측값은 데이터 점과 가장 가까운 적합선 상의 점이며, 가깝다는 것은 분산 비율을 기준으로 합니다.

신뢰 한계는 Tan and Iglewicz 연구 자료(1999)에 설명된 것과 같이 계산됩니다.

적합 요약 보고서

R^2

해당 분산 분석 테이블에서 구한 통계량을 사용할 경우 모든 연속형 반응 적합의 R^2 는 다음과 같이 계산됩니다 .

$$\frac{\text{모형의 제곱합}}{\text{Sum of Squares for C. Total}}$$

Adj-R^2

Adj-R^2 는 제곱합 대신 사용되는 평균 제곱의 비율로 , 다음과 같이 계산됩니다 .

$$1 - \frac{\text{오차의 평균 제곱}}{\text{Mean Square for C. Total}}$$

오차의 평균 제곱은 "분산 분석" 보고서에 표시됩니다(그림 5.12). 수정 합계의 제곱합은 수정 합계의 제곱합을 해당 자유도로 나눠서 계산할 수 있습니다.

적합 결여 보고서

순수 오차 DF

순수 오차 DF 의 경우 키 값이 동일한 관측값이 둘 이상 있는 경우를 고려합니다 . 일반적으로 각 효과의 값이 동일한 행이 여러 개 있는 g 개의 그룹이 있는 경우 합동 DF(DF_p) 는 다음과 같이 정의됩니다 .

$$DF_p = \sum_{i=1}^{g} (n_i - 1)$$

여기서 n_i는 i번째 그룹에 있는 관측값의 개수입니다.

순수 오차 SS

순수 오차 SS 의 경우 , 일반적으로 x 값이 동일한 행이 여러 개 있는 g 개의 그룹이 있다면 합동 SS(SS_p) 는 다음과 같이 정의됩니다 .

$$SS_p = \sum_{i=1}^{g} SS_i$$

여기서 SS_i는 평균이 수정된 i번째 그룹의 제곱합입니다.

최대 R²

순수 오차는 모형 형태에 따라 변동되지 않으며 가능한 최소 분산이므로 최대 R² 는 다음과 같이 계산됩니다 .

$$1 - \frac{\text{SS(순수 오차)}}{\text{SS(Total for whole model)}}$$

모수 추정값 보고서

표준화 베타

표준화 베타는 다음과 같이 계산됩니다 .

$$\hat{\beta}(s_x / s_y)$$

여기서 $\hat{\beta}$ 는 추정된 모수이고, s_x 및 s_y는 X 및 Y 변수의 표준편차입니다.

설계 표준 오차

설계 표준 오차는 모수 추정값의 표준 오차를 RMSE로 나눠서 계산됩니다.

평활 적합 보고서

R²는 1 - (SSE / 수정 합계 SS)와 같습니다. 여기서 수정 합계 SS는 선형 적합 ANOVA 보고서에서 확인할 수 있습니다.

상관 보고서

Pearson 상관계수는 r 로 표기되며 다음과 같이 계산됩니다 .

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} \text{ 여기서 , } s_{xy} = \frac{\sum w_i (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{df}$$

여기서 w_i는 i번째 관측값의 체중(체중 열이 지정된 경우) 또는 1(체중 열이 할당되지 않은 경우)입니다.

6 장

일원 분석

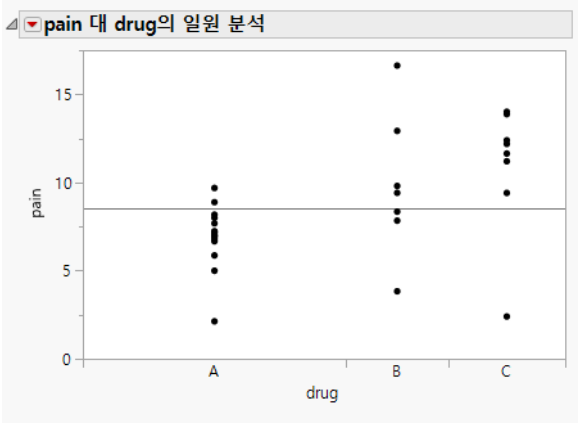
연속형 Y 와 범주형 X 변수 간의 관계 분석

범주형 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지를 탐색하려면 일원 분석 플랫폼을 사용합니다. ANOVA, 비모수 검정 또는 다중 비교를 사용하여 그룹 간 평균의 차이를 평가할 수 있습니다. 밀도 및 CDF 그림을 사용하면 X 범주별 반응의 분포를 시각화할 수 있습니다. 예를 들어 동일한 약품 유형(X)의 서로 다른 범주가 수치 척도로 측정된 환자의 통증 수준(Y)에 어떤 영향을 미치는지 확인할 수 있습니다.

일원 분석 플랫폼은 X로 Y 적합 플랫폼의 **연속형 대 명목형 또는 순서형** 분석법입니다. 분석 결과는 그림에 표시되며, 다음과 같은 분석을 대화식으로 추가할 수 있습니다.

- 평균을 적합시키고 평균이 동일한지 검정하기 위한 일원 분산 분석
- 비모수 검정
- 분산 동질성 검정
- 평균에 대한 다중 비교 검정 (평균 비교 원 사용)
- 각 그룹에 중첩 표시되는 이상치 상자 그림
- 일원 배치의 검정력 상세 정보

그림 6.1 일원 분석



목차

일원 분석 개요 156
일원 분석의 예 156
일원 분석 플랫폼 시작 158
데이터 형식 159
일원 분석 그림 159
일원 분석 플랫폼 옵션 160
표시 옵션 163
분위수 164
이상치 상자 그림 165
평균 /ANOVA 및 평균 /ANOVA/ 합동 t 165
적합 요약 보고서 166
t-검정 보고서 166
분산 분석 보고서 167
일원 ANOVA에 대한 평균 보고서 168
블록 평균 보고서 168
평균 다이아몬드 및 X 축 비례 168
평균 선, 오차 막대 및 표준편차 선 169
평균 분석 방법 170
위치에 대한 평균 분석 170
척도에 대한 평균 분석 171
평균 분석 차트 172
평균 분석 옵션 173
평균 비교 173
비교 원 사용 175
개별 쌍 비교(스튜던트 t) 177
전체 쌍 비교(Tukey HSD) 177
최량 비교(Hsu MCB) 177
대조군 비교(Dunnett) 178
단계별 개별 쌍 비교(Newman-Keuls) 178
평균 비교 옵션 179
비모수 검정 180
Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서 181
Kolmogorov-Smirnov 2표본 검정 보고서 182
비모수 다중 비교 184
이분산 186
분산 동일 여부 검정 보고서 187
동등성 검정 188
동등성 검정 보고서 188

로버스트 190
로버스트 적합 190
Cauchy 적합 190
검정력 191
검정력 상세 정보 창 및 보고서 191
정규 분위수 그림 192
CDF 그림 192
밀도 193
매칭 열 193
일원 분석 플랫폼의 추가 예 194
평균 분석 차트의 예 194
분산에 대한 평균 분석 차트의 예 195
개별 쌍 비교(스튜던트 t) 검정의 예 196
전체 쌍 비교(Tukey HSD) 검정의 예 198
최량 비교(Hsu MCB) 검정의 예 200
대조군 비교(Dunnett)의 예 201
단계별 개별 쌍 비교(Newman-Keuls) 검정의 예 202
예: 4가지 평균 비교 검정 대조 203
비모수 Wilcoxon 검정의 예 204
이분산 옵션의 예 207
동등성 검정의 예 208
로버스트 적합 옵션의 예 209
검정력 옵션의 예 211
정규 분위수 그림의 예 213
CDF 그림의 예 214
밀도 옵션의 예 215
매칭 열 옵션의 예 216
예: 일원 분석을 위한 데이터 쌓기 217
일원 분석 플랫폼에 대한 통계 상세 정보 224
비교 원 224
검정력 225
적합 요약 보고서 226
분산 동일 여부 검정 227
비모수 검정 통계량 228

일원 분석 개요

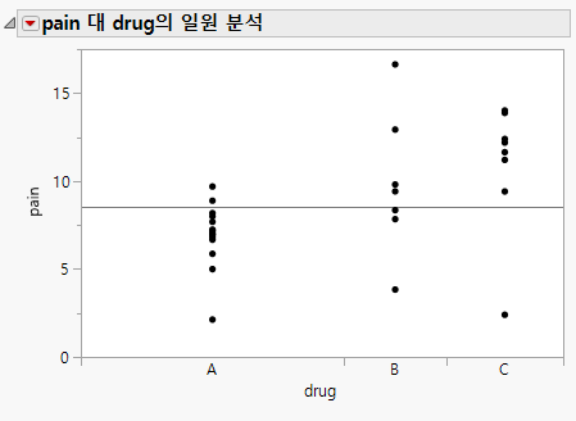
일원 분산 분석에서는 그룹 평균 간에 차이가 있는지 검정합니다. 반응의 전체 변동성은 그룹 내 변동성과 그룹 간 변동성의 두 부분으로 분할됩니다. 그룹 간 변동성이 그룹 내 변동성에 비해 크면 그룹 평균 간의 차이가 유의한 것으로 간주됩니다.

일원 분석의 예

그룹 평균의 차이가 유의한지 검정하기 위해 JMP 에서 일원 분석을 수행하는 방법을 알아봅니다 . 이 데이터에서는 33 명의 참가자에게 세 가지 유형의 진통제 (A, B 및 C) 를 투여한 후 통증 수준을 슬라이딩 척도로 평가하도록 했습니다 . A, B 및 C 에 대한 평균이 유의하게 다른지 여부를 확인하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Analgesics.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. pain 을 선택하고 **Y, 반응**을 클릭합니다 .
4. drug 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 6.2 일원 분석의 예



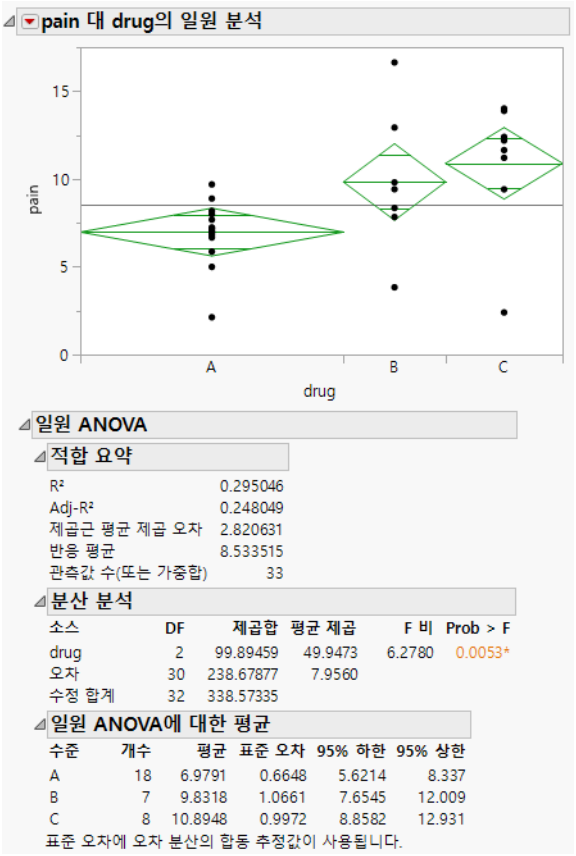
한 가지 약물(A)이 다른 약물보다 일관되게 낮은 점수를 받았음을 알 수 있습니다. 또한 *x* 축의 눈금 간격이 일정하지 않습니다. 눈금 사이의 길이는 각 약물에 대한 점수(관측값)의 개수와 비례합니다.

데이터에 대해 분산 분석을 수행합니다 .

6. "pain 대 drug 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .

참고 : X 요인의 수준이 두 개뿐인 경우 **평균 /ANOVA** 옵션은 **평균 /ANOVA/ 합동 t** 로 표시되며 이 옵션을 선택하면 보고서 창에 " 합동 *t*- 검정 " 보고서가 추가됩니다 .

그림 6.3 평균 /ANOVA 옵션의 예



다음 관측값에 유의하십시오 .

- 신뢰 구간을 나타내는 평균 다이아몬드가 표시됩니다 .
  - 각 다이아몬드의 중심 부근에 있는 선은 그룹 평균을 나타냅니다 . 각 약물의 평균에 유의한 차이가 있음을 알 수 있습니다 .
  - 각 다이아몬드의 세로 범위는 각 그룹의 평균에 대한 95% 신뢰 구간을 나타냅니다 .

  자세한 내용은 " 평균 다이아몬드 및 X 축 비례 "(168 페이지) 에서 확인하십시오 .
- " 적합 요약 " 테이블에서는 분석에 대한 전체적인 요약 정보를 제공합니다 .
- " 분산 분석 " 보고서에서는 표준 ANOVA 정보를 보여 줍니다 . Prob > F(*p* 값) 는 0.0053 으로 나타나는데 이는 약물 간에 유의한 차이가 있어 보인다는 결론을 뒷받침합니다 .

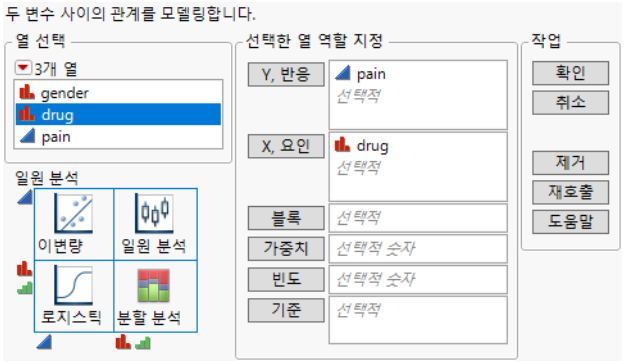
- " 일원 ANOVA 에 대한 평균 " 보고서에서는 범주형 요인의 각 수준에 대한 평균 , 표본 크기 및 표준 오차를 보여 줍니다 .

일원 분석 플랫폼 시작

일원 분산 분석을 수행하려면 다음을 수행하십시오 .

1. **분석 > X 로 Y 적합**을 선택합니다 .
2. **Y, 반응**에 연속형 열을 입력합니다 .
3. **X, 요인**에 명목형 또는 순서형 열을 입력합니다 .

그림 6.4 X 로 Y 적합 시작 창



다이어그램 위에 "일원 분석"이라고 표시되면 일원 분석을 수행할 것임을 나타냅니다.

참고 : "JMP 시작하기 " 창에서 일원 분석을 시작할 수도 있습니다 . **보기 > JMP 시작하기 > 기본 분석 > 일원 분석**을 선택합니다 .

이 시작 창에 대한 자세한 내용은 "X 로 Y 적합 소개" 장 (105 페이지)에서 확인하십시오 . "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

블록 변수를 지정하면 교호작용이 없는 이원 분석을 구성하는 두 번째 요인이 식별됩니다. 일원 분석 그림에서는 Y 변수의 값이 블록 변수에 의해 중심화됩니다. 그룹 셀별로 각 블록에 동일한 개수가 있는 경우 사용 가능한 일원 분석 플랫폼 옵션이 이원 분석 설정에 맞게 조정됩니다. 그룹 셀별 블록에 동일한 개수가 있지 않은 경우 모형 적합 보고서가 일원 분석 그림 아래에 표시되고 일원 분석 플랫폼 옵션을 사용할 수 없습니다.

데이터 형식

일원 분석 플랫폼을 실행하려면 각 행에 X 변수 수준이 동일한 하나 이상의 관측값에 대한 정보가 포함되어 있어야 합니다. 한 행이 둘 이상의 관측값을 나타내는 경우에는 가중치 또는 빈도 변수를 사용하여 행이 나타내는 관측값의 수를 지정해야 합니다.

일원 데이터가 JMP 데이터 테이블이 아닌 다른 형식으로 존재하는 경우 데이터 배열이 한 행에 여러 관측값에 대한 정보가 포함되는 방식일 수도 있습니다. JMP에서 데이터를 분석하려면 데이터를 가져온 후 JMP 데이터 테이블의 각 행에 단일 관측값에 대한 정보가 포함되도록 데이터를 재구성해야 합니다. 자세한 내용은 "예: 일원 분석을 위한 데이터 쌓기"(217페이지)에서 확인하십시오.

일원 분석 그림

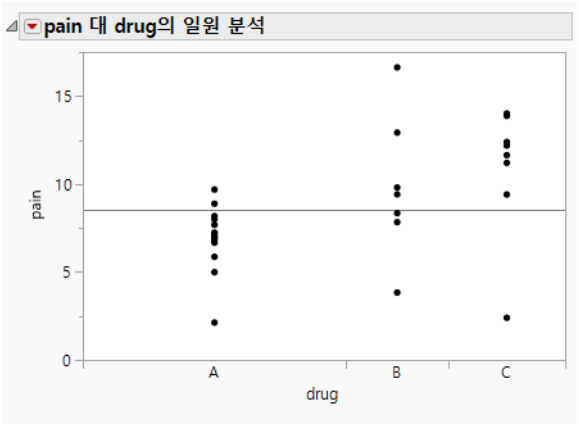
일원 분석 그림에는 각 X 요인 값의 반응 점이 표시됩니다. X 요인의 수준 간에 반응 분포를 비교할 수 있습니다. X의 구분되는 값을 수준이라고도 합니다. 시작 창에 블록 변수가 지정된 경우 일원 분석 그림의 Y 변수 값은 블록 변수에 의해 중심화됩니다.

두 가지 방법 중 하나로 그림의 변수를 바꿀 수 있습니다. 한 가지 방법은 한 축의 변수를 다른 축으로 드래그하여 놓아 기존 변수를 바꾸는 것이고, 다른 방법은 관련 데이터 테이블의 "열" 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다.

"일원 분석"의 빨간색 삼각형 메뉴에 있는 옵션을 사용하여 보고서 창에 보고서, 추가 그림 및 검정을 추가할 수 있습니다. 자세한 내용은 "일원 분석 플랫폼 옵션"(160페이지)에서 확인하십시오.

그림 6.5에 표시된 것과 같은 그림을 생성하려면 "일원 분석의 예"(156페이지)에 설명된 방법을 따르십시오.

그림 6.5 일원 분석 그림



참고 : 데이터 테이블에서 제외된 모든 행은 일원 분석 그림에서도 숨겨집니다 .

일원 분석 플랫폼 옵션

일원 분석 플랫폼의 빨간색 삼각형 메뉴에는 그래프 및 분석 옵션이 포함되어 있습니다. 여기에는 ANOVA, 비모수 검정 및 그룹 간 평균의 차이를 평가하기 위한 다중 비교 방법이 포함됩니다. 밀도 및 CDF 그림을 사용하면 X 범주별 반응의 분포를 시각화할 수 있습니다. 시작 창에 블록 변수가 지정된 경우 이러한 옵션 중 많은 수를 사용할 수 없습니다.

참고 : " 적합 그룹 " 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에 표시됩니다 . 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 를 기준으로 정렬할 수 있습니다 . 자세한 내용은 **선형 모형 적합** 의 에서 확인하십시오 .

표 6.1 옵션 및 요소의 예

플랫폼 옵션	그림에 추가되는 개체	보고서 창에 추가되는 보고서
분위수	상자 그림	분위수 보고서
평균 /ANOVA	평균 다이아몬드	일원 ANOVA 보고서
평균 및 표준편차	평균 선, 오차 막대 및 표준편차 선	평균 및 표준편차 보고서
평균 비교	비교 원 (" 단계별 개별 쌍 비교 (Newman-Keuls)" 옵션 제외)	평균 비교 보고서

" 일원 분석 " 의 빨간색 삼각형 메뉴에는 다음과 같은 옵션이 있습니다 . 일부 옵션은 특정 조건이 충족되는 경우에만 표시될 수 있습니다 .

분위수 각 그룹에 대해 다음 분위수를 나열합니다 .

- 0%(최소값)
- 10%
- 25%
- 50%(중앙값)
- 75%
- 90%
- 100%(최대값)

표시 옵션 메뉴에서 **상자 그림**을 활성화합니다 . 자세한 내용은 " 분위수 "(164 페이지) 에서 확인하십시오 .

평균 /ANOVA 각 그룹의 평균을 적합시키고 일원 분산 분석을 수행하여 평균 간에 차이가 있는지 검정합니다 . 자세한 내용은 " 평균 /ANOVA 및 평균 /ANOVA/ 합동 t"(165 페이지) 에서 확인하십시오 .

참고 : X 요인의 수준이 두 개인 경우에는 이 메뉴 옵션이 **평균 /ANOVA/ 합동 t** 로 변경되며 이 옵션을 선택하면 분산이 동일하다는 가정하에 " 합동 *t*- 검정 " 보고서가 생성됩니다 .

평균 및 표준편차 각 그룹에 대한 요약 통계량을 제공합니다 . 평균에 대한 표준 오차에는 표준편차의 합동 추정값 대신 개별 그룹의 표준편차가 사용됩니다 .

그림에는 평균 선 , 오차 막대 및 표준편차 선이 포함됩니다 . 이러한 요소에 대한 간략한 설명은 " 표시 옵션 "(163 페이지) 에서 확인하십시오 . 이러한 요소에 대한 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 "(169 페이지) 에서 확인하십시오 .

t- 검정 (X 요인의 수준이 두 개인 경우에만 사용 가능) 분산이 동일하지 않다는 가정하에 *t*- 검정 보고서를 생성합니다 . 자세한 내용은 "t- 검정 보고서 "(166 페이지) 에서 확인하십시오 .

평균 분석 방법 ANOM(평균 분석) 절차를 수행하기 위한 5 가지 명령을 제공합니다 . 평균 , 분산 및 범위를 비교하기 위한 명령이 있습니다 . 자세한 내용은 " 평균 분석 방법 "(170 페이지) 에서 확인하십시오 .

평균 비교 여러 그룹 평균을 비교하기 위한 다중 비교 방법을 제공합니다 . 자세한 내용은 " 평균 비교 "(173 페이지) 에서 확인하십시오 .

비모수 그룹 위치에 대한 비모수 비교 기능을 제공합니다 . 자세한 내용은 " 비모수 검정 "(180 페이지) 에서 확인하십시오 .

이분산 그룹 분산의 동일 여부에 대한 4 가지 검정을 수행합니다 . 또한 그룹 내의 분산이 동일하지 않을 때 평균을 비교하기 위한 ANOVA 검정인 Welch 검정을 제공합니다 . 자세한 내용은 " 이분산 "(186 페이지) 에서 확인하십시오 .

동등성 검정 차이가 지정된 임계값보다 작은지 검정합니다 . 검정의 분산 가정을 합동 분산 또는 이분산으로 지정할 수 있습니다 . 자세한 내용은 " 동등성 검정 "(188 페이지) 에서 확인하십시오 .

로버스트 이상치가 데이터에 미치는 영향을 줄이기 위한 두 가지 방법을 제공합니다 . 자세한 내용은 " 로버스트 "(190 페이지) 에서 확인하십시오 .

검정력 통계 검정력에 대한 계산과 지정된 가설 검정에 대한 기타 상세 정보를 제공합니다 . 자세한 내용은 " 검정력 "(191 페이지) 에서 확인하십시오 .

" 모형 적합 " 플랫폼 내에 " 검정력 상세 정보 " 창 및 보고서도 표시됩니다 . 검정력 계산에 대한 자세한 내용과 예는 **선형 모형 적합**의 에서 확인하십시오 .

α **수준 설정** 가장 일반적인 유의 수준 옵션 중에서 하나를 선택하거나 , **기타**를 선택하고 원하는 수준을 직접 지정할 수 있습니다 . 유의 수준을 변경하면 다음 작업이 수행됩니다 .

- 신뢰 한계가 다시 계산됩니다 .
- 그림의 평균 다이아몬드가 조정됩니다 (표시된 경우).

- 보고서의 신뢰 상한 및 신뢰 하한 값이 수정됩니다 .
- 모든 " 평균 비교 " 보고서의 임계값 및 비교 원이 변경됩니다 .
- 모든 " 비모수 다중 비교 " 보고서의 임계값이 변경됩니다 .

정규 분위수 그림 각 그룹의 데이터 분위수를 표시하기 위한 다음 옵션을 제공합니다 .

분위수별 실제값 그림 일원 분석 그림에 분위수 그림을 추가합니다. 분위수 그림은 반응 변수에 대한 일원 분석 세로 축을 공유합니다 . 각 그룹의 누적 확률은 가로 축에 있습니다 . 이 그림에는 범주형 X 요인의 각 수준 내에서 계산된 분위수가 표시됩니다 .

실제값별 분위수 그림 가로 축에 반응 변수가 표시되고 세로 축에 누적 확률이 표시되는 분위수 그림을 생성합니다 . 이 그림에는 범주형 X 요인의 각 수준 내에서 계산된 분위수가 표시됩니다 .

적합선 X 변수의 각 수준에 대한 데이터에 적합된 참조선을 표시하거나 숨깁니다 .

정규 분위수 라벨 정규 분위수 그림의 오른쪽 세로 축 내부 또는 분위수별 실제값 그림의 위쪽 축 내부에 라벨을 표시하거나 숨깁니다 . 라벨은 정규 분위수 척도를 나타냅니다 .

CDF 그림 일원 분석 보고서의 모든 그룹에 대한 누적 분포 함수 그림을 표시합니다 . 자세한 내용은 "CDF 그림 "(192 페이지) 에서 확인하십시오 .

밀도 그룹 간에 밀도를 비교합니다 . 자세한 내용은 " 밀도 "(193 페이지) 에서 확인하십시오 .

매칭 열 매칭 모형 분석을 수행하기 위한 매칭 변수를 지정합니다 . 서로 다른 그룹의 관측값을 동일한 참가자로부터 얻은 경우처럼 일원 분석의 데이터를 매칭 (쌍체) 데이터로부터 얻은 경우에 이 옵션을 사용합니다 .

그러면 매칭되는 점들을 연결하는 매칭 선이 그림에 포함됩니다. 자세한 내용은 "매칭 열"(193 페이지) 에서 확인하십시오 .

저장 다음과 같은 통계량을 현재 데이터 테이블에 새 열로 저장합니다 .

잔차 저장 반응 변수에서 요인 변수의 각 수준 내 반응 변수 평균을 빼서 계산된 값을 저장합니다 .

표준화한 데이터 저장 요인 변수의 각 수준 내에서 계산된 표준화된 반응 변수 값을 저장합니다 . 이 값은 중심화된 반응을 각 수준 내의 표준편차로 나눈 값입니다 .

정규 분위수 저장 범주형 요인 변수의 각 수준 내에서 계산된 정규 분위수 값을 저장합니다 .

예측값 저장 요인 변수의 각 수준에 대한 반응 변수의 예측 평균을 저장합니다 .

표시 옵션 그림의 요소를 추가하거나 제거합니다 . 자세한 내용은 " 표시 옵션 "(163 페이지) 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

표시 옵션

" 표시 옵션 " 을 사용하여 일원 분석 데이터 그림의 요소를 추가하거나 제거할 수 있습니다 . 일부 옵션은 관련이 없으면 표시되지 않을 수 있습니다 .

모든 그래프 모든 그래프를 표시하거나 숨깁니다 .

점 그림에 데이터 점을 표시하거나 숨깁니다 .

상자 그림 각 그룹에 대한 이상치 상자 그림을 표시하거나 숨깁니다 . 예는 " 일원 분석 수행 "(222 페이지) 에서 확인하십시오 .

평균 다이아몬드 가로 축에 비례하여 각 그룹의 평균을 통과하는 수평선을 그립니다 . 평균 다이아몬드의 맨 위 점과 맨 아래 점은 각 그룹에 대한 95% 신뢰 구간의 상한 및 하한 점을 보여줍니다 . 자세한 내용은 " 평균 다이아몬드 및 X 축 비례 "(168 페이지) 에서 확인하십시오 .

평균 선 각 그룹의 평균 위치에 선을 그립니다 . 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 "(169 페이지) 에서 확인하십시오 .

평균 CI 선 각 그룹의 95% 신뢰 상한 및 하한 위치에 선을 그립니다 .

평균 오차 막대 각 그룹의 평균을 식별하고 평균의 1 표준 오차 위와 아래에 오차 막대를 표시합니다 . 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 "(169 페이지) 에서 확인하십시오 .

총 평균 그림에 *Y* 변수의 전체 평균을 그립니다 .

표준편차 선 각 그룹 평균의 1 표준편차 위와 아래에 선을 표시합니다 . 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 "(169 페이지) 에서 확인하십시오 .

비교 원 비교 원을 표시하거나 숨깁니다 . 이 옵션은 **평균 비교** 옵션 중 하나를 선택한 경우에만 사용할 수 있습니다 . 자세한 내용은 " 비교 원 "(224 페이지) 에서 확인하십시오 . 예는 " 일원 분석 수행 "(222 페이지) 에서 확인하십시오 .

여러 평균 연결 그룹 평균을 직선으로 연결합니다 .

평균의 평균 그룹 평균의 평균 위치에 선을 그립니다 .

X 축 비례 각 수준의 표본 크기에 비례하여 *x* 축에 간격을 지정합니다 . 자세한 내용은 " 평균 다이아몬드 및 X 축 비례 "(168 페이지) 에서 확인하십시오 .

퍼진 점 구간 너비 전체에 점을 분산시킵니다 .

지터링된 점 동일한 *y* 값에 중첩 표시되는 여러 점 사이에 약간의 간격을 추가합니다 . 가로로 조정되는 점 간격은 0.375 ~ 0.625 이며 4 * (Uniform(0,1)-0.5)^5 분포를 따릅니다 .

매칭 선 (**매칭 열** 옵션을 선택한 경우에만 표시됨) 매칭되는 점들을 연결합니다 .

매칭 점선 (**매칭 열** 옵션을 선택한 경우에만 표시됨) 테이블에 있는 결측 셀의 셀 평균을 연결하는 점선을 그립니다. 선의 끝점으로 사용되는 값은 이원 ANOVA 모형을 사용하여 구합니다.

히스토그램 원래 그림의 오른쪽에 가로 배열 히스토그램을 그립니다 .

로버스트 평균 선 (" 로버스트 " 옵션을 선택한 경우에만 표시됨) 각 그룹의 로버스트 평균 위치에 선을 그립니다 .

범례 " 정규 분위수 그림 ", "CDF 그림 " 및 " 밀도 " 옵션에 대해 범례를 표시합니다 .

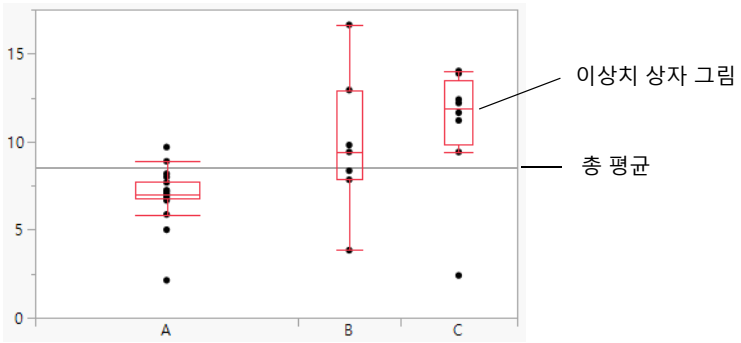
분위수

" 분위수 " 보고서에는 X 요인 변수의 각 수준에 대해 선택된 백분위수가 나열됩니다 . 중앙값은 50 번째 백분위수이고 , 25 번째 및 75 번째 백분위수는 **사분위수**라고 합니다 .

분위수 옵션은 그림에 다음과 같은 요소를 추가합니다 .

- Y 변수의 전체 평균을 나타내는 총 평균
- 각 요인 수준의 점 분포를 요약한 이상치 상자 그림

그림 6.6 이상치 상자 그림 및 총 평균



참고 : 이러한 요소를 숨기려면 " 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림** 또는 **총 평균**을 선택하십시오 .

이상치 상자 그림

이상치 상자 그림은 데이터의 분포를 그래픽으로 요약한 것입니다 . 이상치 상자 그림의 경우 다음 사항에 유의하십시오 (그림 6.7).

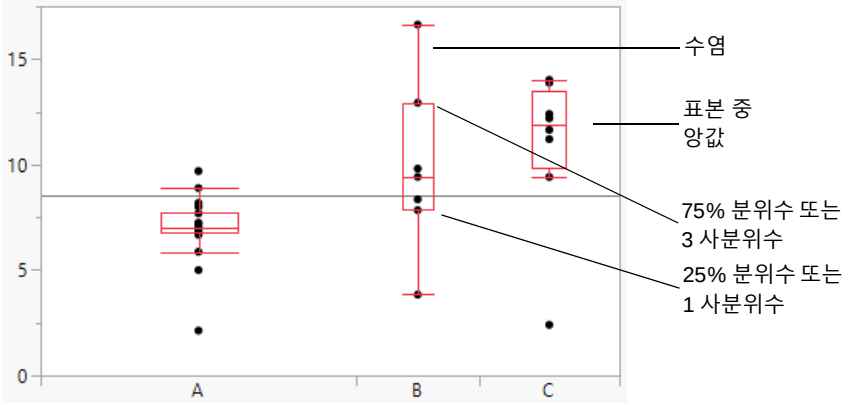
- 상자 내의 가로 선은 표본 중앙값을 나타냅니다 .
- 상자 끝은 75번째 및 25번째 분위수를 나타내며 각각 3사분위수 및 1사분위수라고도 합니다.
- 1 사분위수와 3 사분위수 사이의 차이를 **사분위수 범위** . 라고 합니다 .
- 각 상자에는 양 끝에서 시작되는 선이 있으며 이 선은 **수염**이라고도 합니다 . 수염은 상자 끝에서 시작해서 다음과 같이 계산된 거리 내에 있는 점들 중 가장 바깥쪽 데이터 점까지 이어집니다 .

  3 사분위수 + 1.5 * (사분위수 범위)

  1 사분위수 - 1.5 * (사분위수 범위)

  데이터 점이 계산된 범위에 포함되지 않으면 상한 및 하한 데이터 점 값 (이상치 제외) 에 따라 수염이 결정됩니다 .

그림 6.7 이상치 상자 그림의 예



평균 /ANOVA 및 평균 /ANOVA/ 합동 t

평균 /ANOVA 옵션은 분산 분석을 수행합니다 . X 요인에 정확히 두 개의 수준만 포함되어 있는 경우 이 옵션은 **평균 /ANOVA/ 합동 t** 로 표시됩니다 . 이 경우 다른 보고서 외에도 합동 분산 (또는 등분산) 을 가정하는 합동 *t*- 검정 보고서가 표시됩니다 .

일원 분석 그림에 평균 다이아몬드가 추가됨 자세한 내용은 " 표시 옵션 "(163 페이지) 및 " 평균 다이아몬드 및 X 축 비례 "(168 페이지) 에서 확인하십시오 .

보고서 자세한 내용은 " 적합 요약 보고서 "(166 페이지), " 분산 분석 보고서 "(167 페이지), " 일원 ANOVA 에 대한 평균 보고서 "(168 페이지), "t- 검정 보고서 "(166 페이지) 및 " 블록 평균 보고서 "(168 페이지) 에서 확인하십시오 .

- "*t*- 검정 " 보고서는 **평균 /ANOVA/ 합동 t** 옵션을 선택한 경우에만 표시됩니다 .
- " 블록 평균 " 보고서는 시작 창에서 블록 변수를 지정했고 블록 및 X 요인 수준의 각 조합에 동일한 개수가 있는 경우에만 나타납니다 .

적합 요약 보고서

일원 분석 플랫폼의 " 적합 요약 " 보고서에는 일원 분산 분석에 대한 요약이 표시됩니다 .

R² 각 요인 수준에 평균을 적합시킴으로써 설명되는 변동의 비율을 측정합니다. 나머지 변동은 랜덤 오차에 기인합니다 . 그룹 평균을 적합시킨 결과 오차 없이 모든 변동이 설명되면 R^2 값이 1 이 됩니다 . R^2 가 0 이면 해당 적합이 전체 반응 평균보다 더 적절한 예측 모형이 아님을 나타냅니다 . 자세한 내용은 " 적합 요약 보고서 "(226 페이지) 에 대한 통계 상세 정보에서 확인하십시오 .

R^2 를 **결정 계수**라고도 합니다 .

참고 : R² 값이 낮으면 설명되지 않은 변동을 설명하는 변수가 모형에 없는 것일 수 있습니다 . 하지만 데이터의 내재 변동 범위가 큰 경우에는 유용한 ANOVA 모형이라도 R² 값이 낮을 수 있습니다 . 일반적인 R² 값에 대해 알아보려면 연구 영역의 문헌을 읽어 보십시오 .

Adj-R² 모수의 개수가 서로 다른 모형들을 비교하기 좋도록 계산에 자유도를 사용하여 R^2 값을 조정합니다 . 자세한 내용은 " 적합 요약 보고서 "(226 페이지) 에 대한 통계 상세 정보에서 확인하십시오 .

제곱근 평균 제곱 오차 랜덤 오차의 표준편차를 추정합니다 . " 분산 분석 " 보고서에 표시된 **오차** 에 대한 평균 제곱의 제곱근입니다 .

반응 평균 반응 변수의 전체 평균 (산술평균) 입니다 .

관측값 수 (또는 가중합) 적합 추정에 사용된 관측값의 개수입니다 . 가중치가 사용된 경우에는 가중치의 합입니다 . 자세한 내용은 " 적합 요약 보고서 "(226 페이지) 에 대한 통계 상세 정보에서 확인하십시오 .

t- 검정 보고서

t- 검정에는 등분산 및 이분산의 두 가지 유형이 있습니다 .

- **평균 /ANOVA/ 합동 t** 옵션을 선택하면 " 합동 t- 검정 " 보고서가 표시됩니다 . 이 *t*- 검정에서는 등분산을 가정합니다 .
- 빨간색 삼각형 메뉴에서 **t- 검정** 옵션을 선택하면 "*t*- 검정 " 보고서가 표시됩니다 . 이 *t*- 검정에서는 등분산을 가정합니다 .

"t- 검정 " 보고서에는 다음 열이 포함됩니다 .

t- 검정 그림 귀무가설이 참이라는 가정하에 평균의 차이에 대한 표본 분포를 보여 줍니다 . 빨간색 세로 선은 평균의 실제 차이입니다 . 음영 영역은 *p* 값에 해당합니다 .

차이 두 X 수준 간의 추정 차이를 보여 줍니다. 이 그림에서 " 차이 " 값은 두 수준을 비교하는 빨간색 선으로 표시됩니다.

차이 표준 오차 차이의 표준 오차를 보여 줍니다.

차이 CL 상한 차이의 신뢰 상한을 보여 줍니다.

차이 CL 하한 차이의 신뢰 하한을 보여 줍니다.

신뢰도 신뢰 수준 (1-α) 을 보여 줍니다. 신뢰 수준을 변경하려면 플랫폼의 빨간색 삼각형 메뉴에 있는 α **수준 설정** 명령에서 새 유의 수준을 선택합니다.

t 비 *t* 통계량의 값입니다.

DF *t*- 검정에 사용된 자유도입니다.

Prob > |t| 양쪽 꼬리 검정과 관련된 *p* 값입니다.

Prob > t 위쪽 꼬리 검정과 관련된 *p* 값입니다.

Prob < t 아래쪽 꼬리 검정과 관련된 *p* 값입니다.

분산 분석 보고서

" 분산 분석 " 보고서에서는 표본의 총 변동을 두 성분으로 분할합니다. 두 평균 제곱의 비율은 *F* 비가 됩니다. *F* 비와 관련된 확률이 낮으면 해당 모형은 전체 반응 평균보다 통계적으로 더 적절한 적합인 것입니다.

참고 : 블록 열을 지정한 경우에는 " 분산 분석 " 보고서에 **블록** 변수가 포함됩니다.

소스 세 가지 변동 소스를 나열합니다. 모형 소스 , **오차** 및 **수정 합계**가 이러한 소스에 해당합니다.

DF 각 변동 소스의 관련 DF(자유도) 를 기록합니다.

- **수정 합계**의 자유도는 N - 1 이며 , 여기서 N 은 분석에 사용된 관측값의 총 개수입니다.
- X 요인의 수준이 k 개이이면 모형의 자유도는 k - 1 입니다.

오차 자유도는 **수정 합계** 자유도와 **모형** 자유도의 차이 (즉 , N - k) 입니다.

제곱합 각 변동 소스의 제곱합 (줄여서 SS 라고 함) 을 기록합니다.

- 전체 반응 평균과 각 반응의 차이에 대한 총 제곱합 (**수정 합계**). **수정 합계** 제곱합은 다른 모든 모형과의 비교에 사용되는 기본 모형입니다.
- 각 점으로부터 해당 그룹 평균까지의 거리에 대한 제곱합. 이 값은 분산 분석 모형을 적합시킨 후 설명되지 않은 나머지 **오차** (잔차) 의 SS 입니다.

총 SS 에서 오차 SS 를 뺀 값은 모형에 기인한 제곱합입니다. 이를 통해 총 변동 중 모형으로 설명되는 변동의 비율을 알 수 있습니다.

평균 제곱 제곱합을 관련 자유도로 나눈 것입니다.

- **모형** 평균 제곱은 그룹 평균이 동일하다는 가설을 전제로 오차의 분산을 추정합니다.

– **오차** 평균 제곱은 모형 평균 제곱과는 독립적으로 오차 항의 분산을 추정하며 모형 가설이 전제되지 않습니다 .

F 비 모형 평균 제곱을 오차 평균 제곱으로 나눈 것입니다 . 그룹 평균이 동일하다는 가설 , 즉 그룹 평균 간에 실질적인 차이가 없다는 가설이 참인 경우에는 오차의 평균 제곱과 모형의 평균 제곱이 모두 오차 분산을 추정합니다 . 이 둘의 비율은 *F* 분포를 따릅니다 . 분산 분석 모형의 결과가 합계보다 유의하게 낮은 변동을 보이면 *F* 비가 기대값보다 높게 나타납니다 .

Prob>F 모집단 그룹 평균에 차이가 없는 경우 *F* 값이 계산된 값보다 클 확률입니다 . 관측된 유의 확률이 0.05 이하인 것은 종종 그룹 평균에 차이가 있는 증거로 간주됩니다 .

일원 ANOVA 에 대한 평균 보고서

" 일원 ANOVA 에 대한 평균 " 보고서에는 명목형 또는 순서형 요인의 각 수준에 대한 반응 정보가 요약됩니다 .

수준 X 변수의 수준을 나열합니다 .

개수 각 그룹의 관측값 수를 나열합니다 .

평균 각 그룹의 평균을 나열합니다 .

표준 오차 그룹 평균에 대한 표준편차 추정값을 나열합니다 . 이 표준 오차는 각 수준에서 반응의 분산이 동일하다는 가정하에 추정됩니다 . 이 값은 " 적합 요약 " 보고서에 나오는 제곱근 평균 제곱 오차를 그룹 평균 계산에 사용된 값 개수의 제곱근으로 나눈 값입니다 .

95% 하한 및 95% 상한 그룹 평균에 대한 95% 신뢰 하한 및 상한을 나열합니다 .

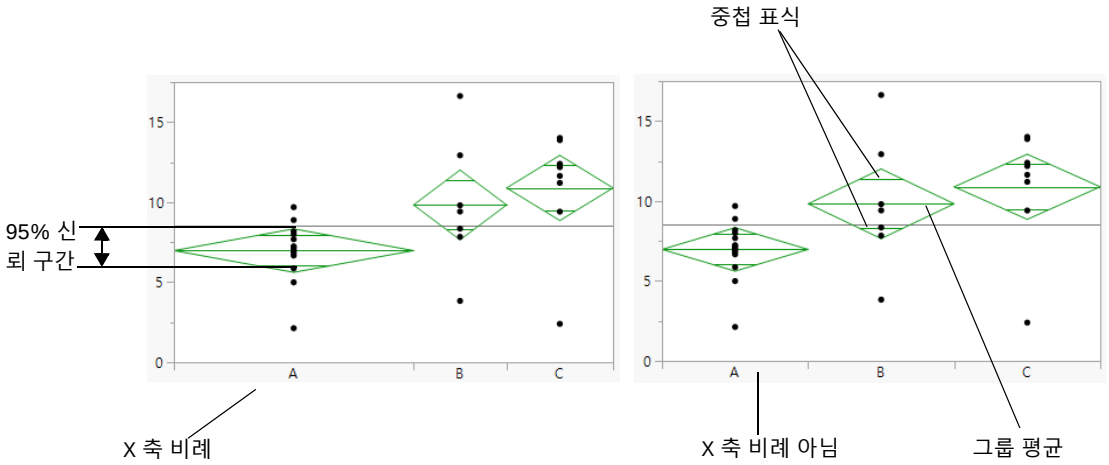
블록 평균 보고서

시작 창에서 블록 변수를 지정한 경우 **평균 /ANOVA** 및 **평균 /ANOVA/ 합동 t** 명령을 선택하면 " 블록 평균 " 보고서가 생성됩니다 . 이 보고서에는 각 블록의 평균과 각 블록의 관측값 수가 표시됩니다 .

평균 다이아몬드 및 X 축 비례

평균 다이아몬드는 표본 평균과 신뢰 구간을 나타낸 것입니다 .

그림 6.8 평균 다이아몬드 및 X 축 비례 옵션의 예



다음 관측값에 유의하십시오 .

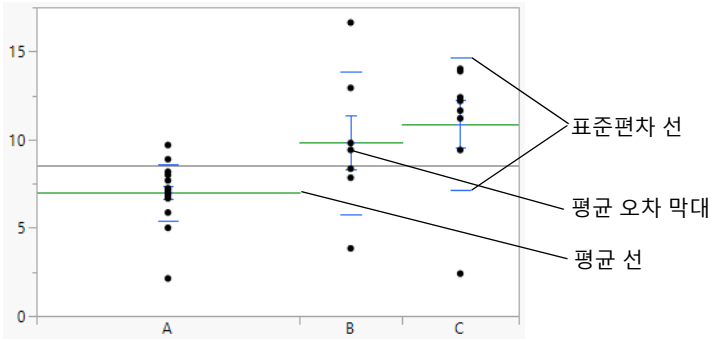
- 각 다이아몬드의 맨 위와 맨 아래는 각 그룹에 대한 $(1 - \alpha) \times 100$ 신뢰 구간을 나타냅니다 . 신뢰 구간은 관측값 간에 분산이 동일하다는 가정하에 계산됩니다 . 따라서 다이아몬드의 높이는 그룹 내의 관측값 수에 대한 제곱근의 역수에 비례합니다 .
- **X 축 비례** 옵션이 선택된 경우 가로 축에서 각 그룹이 차지하는 가로 범위 (다이아몬드의 가로 크기) 는 X 변수의 각 수준에 대한 표본 크기에 비례합니다 . 따라서 일반적으로 다이아몬드의 가로가 좁을수록 세로가 길어지는데, 이는 데이터 점이 적을수록 신뢰 구간이 넓어지기 때문입니다 .
- 각 다이아몬드의 가운데를 가로지르는 평균 선은 그룹 평균을 나타냅니다 .
- 중첩 표식은 그룹 평균의 위쪽과 아래쪽에 선으로 표시됩니다 . 표본 크기가 동일한 그룹의 경우 중첩 표식은 해당 신뢰 수준에서 두 그룹 평균이 유의하게 다르지 않음을 나타냅니다 . 중첩 표식은 그룹 평균 $\pm (\sqrt{2})/2 \times \mathrm{CI}/2$ 로 계산됩니다 . 한 다이아몬드의 중첩 표식이 해당 다이아몬드의 중첩 표식보다 다른 다이아몬드의 평균에 더 가까우면 해당 신뢰 수준에서 관련된 두 그룹이 다르지 않음을 나타냅니다 .
- 플랫폼 메뉴에서 **평균 /ANOVA/ 합동 t** 또는 **평균 /ANOVA** 옵션을 선택하면 자동으로 평균 다이아몬드가 표시됩니다 . 하지만 언제든지 빨간색 삼각형 메뉴에서 **표시 옵션 > 평균 다이아몬드**를 선택하여 이를 표시하거나 숨길 수도 있습니다 .

평균 선 , 오차 막대 및 표준편차 선

표시 옵션 > 평균 선을 선택하여 평균 선을 표시할 수 있습니다 . 평균 선은 X 변수의 각 수준에 대한 반응 평균을 나타냅니다 .

빨간색 삼각형 메뉴에서 **평균 및 표준편차** 옵션을 선택하면 평균 오차 막대와 표준편차 선이 표시됩니다 . 각 옵션을 개별적으로 설정하거나 해제하려면 **표시 옵션 > 평균 오차 막대** 또는 **표준편차 선**을 선택합니다 .

그림 6.9 평균 선 , 평균 오차 막대 및 표준편차 선



평균 분석 방법

평균 분석 (ANOM) 은 평균 , 비율 또는 분산에 대한 다중 비교 방법입니다 . 다음과 같은 경우에 이러한 ANOM 을 사용할 수 있습니다 .

- 전체 (표본) 평균과 통계적으로 다른 그룹 평균이 있는지 여부를 검정하려는 경우
- RMSE(제곱근 평균 제곱 오차) 와 통계적으로 다른 그룹 표준편차가 있는지 여부를 검정하려는 경우
- 전체 범위 평균과 통계적으로 다른 그룹 범위가 있는지 여부를 검정하려는 경우

참고: 분할 분석 플랫폼 내에서는 반응에 두 가지 범주가 있는 경우에 **비율에 대한 평균 분석**을 사용할 수 있습니다 . 자세한 내용은 " 분할 분석 " 장 (233 페이지) 에서 확인하십시오 .

ANOM 방법과 JMP 에서의 ANOM 구현 방식에 대한 자세한 내용은 Nelson et al. 의 연구 자료 (2005) 에서 확인하십시오 .

참고 : 시작 창에서 블록 변수를 지정하고 블록 및 X 요인 수준의 각 조합에 동일한 개수가 있는 경우 ANOM 차트가 사용 가능한 유일한 " 평균 분석 " 차트입니다 . 시작 창에서 블록 변수를 지정하고 개수가 동일하지 않은 경우 사용 가능한 " 평균 분석 " 옵션이 없습니다 .

위치에 대한 평균 분석

다음 옵션을 사용하여 그룹의 평균 또는 중심 값이 공통적인지 여부를 검정할 수 있습니다 .

- ANOM
- 변환 순위를 사용한 ANOM

ANOM

그룹 평균을 전체 평균과 비교하려면 ANOM을 사용합니다. 이 방법은 데이터가 대략적으로 정규 분포되었다고 가정합니다. 자세한 내용은 "평균 분석 차트의 예"(194 페이지)에서 확인하십시오.

변환 순위를 사용한 ANOM

이 방법은 **ANOM** 분석의 비모수 버전입니다. 데이터가 명확하게 비정규 데이터이고 정규 데이터로 변환할 수 없는 경우 이 방법을 사용합니다. "변환 순위를 사용한 ANOM"은 각 그룹 평균의 변환 순위를 전체 평균의 변환 순위와 비교합니다. ANOM 검정을 위해서는 변환된 관측값에 일반적인 ANOM 절차와 임계값을 적용해야 합니다.

변환 순위

n 개의 관측값이 있다고 가정해 보겠습니다. 변환된 관측값은 다음과 같이 계산됩니다.

- 각각의 동일 값을 포함한 모든 관측값에 최소값에서 최대값의 순서로 순위를 매깁니다. 동일한 관측값의 경우 해당 관측값들이 공유하는 순위 블록의 평균을 각 관측값에 할당합니다.
- 순위는 $R_1, R_2, ..., R_n$ 으로 나타냅니다.
- i 번째 관측값에 해당하는 변환 순위는 다음과 같이 결정됩니다.

$$\text{변환된 } R_i = Normal\ Quantile\left[\left(\frac{R_i}{2n+1}\right) + 0.5\right]$$

ANOM 절차는 변환된 R_i 값에 적용됩니다. 순위는 균등 분포를 따르므로 변환 순위는 접힌 정규 분포를 따릅니다. 자세한 내용은 Nelson et al. 연구 자료(2005)에서 확인하십시오.

척도에 대한 평균 분석

다음 옵션을 사용하여 그룹 내의 변동 동질성을 검정할 수 있습니다.

- 분산에 대한 ANOM
- Levene(ADM) 을 사용한 분산에 대한 ANOM
- 범위에 대한 ANOM

분산에 대한 ANOM

그룹 표준편차(또는 분산)를 제곱근 평균 제곱 오차(또는 평균 제곱 오차)와 비교하려면 이 방법을 사용합니다. 이 방법은 데이터가 대략적으로 정규 분포되었다고 가정합니다. 이 방법을 사용하려면 각 그룹마다 4 개 이상의 관측값이 있어야 합니다. 분산에 대한 ANOM 검정과 관련된 자세한 내용은 Wludyka and Nelson 연구 자료(1997) 및 Nelson et al. 연구 자료(2005)에서 확인하십시오. 예는 "분산에 대한 평균 분석 차트의 예"(195 페이지)에서 확인하십시오.

Levene(ADM) 을 사용한 분산에 대한 ANOM

이 방법은 ADM(중앙값과의 절대 편차)의 그룹 평균을 전체 평균 ADM과 비교하는 로버스트 검정을 제공합니다. 데이터가 비정규 데이터이고 정규 데이터로 변환할 수 없는 것으로 의심되는 경우에 "Levene(ADM)을 사용한 분산에 대한 ANOM"을 사용합니다. "Levene(ADM)을 사용한 분산에 대한 ANOM"은 "분산에 대한 ANOM" 분석과 유사한 비모수 분석 방법입니다. "Levene(ADM)을 사용한 분산에 대한 ANOM" 검정에 대한 자세한 내용은 Levene 연구 자료 (1960) 또는 Brown and Forsythe 연구 자료(1974)에서 확인하십시오.

범위에 대한 ANOM

그룹 범위를 그룹 범위의 평균과 비교하려면 이 검정을 사용합니다. 이 검정은 퍼짐에 대한 측도로서 범위를 기준으로 한 척도 차이에 대한 검정입니다. 자세한 내용은 Wheeler 연구 자료(2003)에서 확인하십시오.

참고 : "범위에 대한 ANOM"은 균형 설계와 특정 그룹 크기에만 사용할 수 있습니다. 자세한 내용은 "범위에 대한 ANOM 검정의 제한 사항"(172 페이지)에서 확인하십시오.

범위에 대한 ANOM 검정의 제한 사항

다른 ANOM 결정 한계와 달리 "범위에 대한 ANOM" 차트의 결정 한계에는 제공된 임계값만 사용됩니다. 따라서 "범위에 대한 ANOM"은 다음의 경우에만 사용할 수 있습니다.

- 크기가 동일한 그룹
- 특정 크기 (2~10, 12, 15 및 20) 의 그룹
- 2~30 개의 그룹
- 유의 수준 0.10, 0.05 및 0.01

평균 분석 차트

"평균 분석 방법"의 각 옵션은 보고서 창에 다음을 표시하는 의사결정 차트를 추가합니다.

- UDL(결정 상한)
- LDL(결정 하한)
- 결정 한계 사이에 있는 가로 중심선입니다. 중심선 위치는 차트 유형에 의해 결정됩니다.
  - ANOM: 전체 평균
  - 변환 순위를 사용한 ANOM: 변환 순위의 전체 평균
  - 분산에 대한 ANOM: 제곱근 평균 제곱 오차 (분산 척도의 경우에는 평균 제곱 오차)
  - Levene(ADM) 을 사용한 분산에 대한 ANOM: 평균과의 전체 절대 편차
  - 범위에 대한 ANOM: 그룹 범위의 평균

그림에 표시된 그룹의 통계량이 결정 한계를 벗어나는 경우 이 검정은 해당 그룹의 통계량과 모든 그룹에 대한 전체 평균 통계량 사이에 통계적인 차이가 있음을 나타냅니다.

평균 분석 옵션

"평균 분석 방법"의 각 옵션은 보고서 창의 "평균 분석" 옆에 빨간색 삼각형 메뉴를 추가합니다.

유의 수준 설정 가장 일반적인 유의 수준 옵션 중에서 하나를 선택하거나, **기타**를 선택하고 원하는 수준을 직접 지정합니다. 유의 수준을 변경하면 결정 상한 및 결정 하한이 수정됩니다.

참고 : " 범위에 대한 ANOM" 의 경우에는 0.10, 0.05 및 0.01 만 선택할 수 있습니다.

요약 보고서 표시 보고서는 다음과 같이 평균 분석 방법에 따라 달라집니다.

- ANOM: 그룹 평균 및 결정 한계를 보여 주는 보고서가 생성됩니다.
- 변환 순위를 사용한 ANOM: 그룹 평균 변환 순위 및 결정 한계를 보여 주는 보고서가 생성됩니다.
- 분산에 대한 ANOM: 그룹 표준편차 (또는 분산) 및 결정 한계를 보여 주는 보고서가 생성됩니다.
- Levene(ADM) 을 사용한 분산에 대한 ANOM: 그룹 평균 ADM 및 결정 한계를 보여 주는 보고서가 생성됩니다.
- 범위에 대한 ANOM: 그룹 범위 및 결정 한계를 보여 주는 보고서가 생성됩니다.

분산 척도의 그래프 (**분산에 대한 ANOM** 에만 사용 가능) 세로 축의 척도를 표준편차에서 분산으로 변경합니다.

표시 옵션 표시 모양을 사용자 정의하기 위한 다음 옵션이 포함되어 있습니다.

결정 한계 표시 결정 한계선을 표시하거나 숨깁니다.

결정 한계 음영 표시 결정 한계 음영을 표시하거나 숨깁니다.

중심선 표시 중심선 통계량을 표시하거나 숨깁니다.

점 옵션 : 바늘 표시 바늘을 표시합니다. 이 옵션은 기본 옵션입니다. **연결된 점 표시**는 각 그룹의 평균을 연결하는 선을 표시합니다. **점만 표시**는 각 그룹의 평균을 나타내는 점만 표시합니다.

평균 비교

그룹 평균에 대한 다중 비교를 수행하려면 "평균 비교" 옵션을 사용합니다. 방법에는 Tukey-Kramer 전체 쌍 비교 검정과 대조군과 비교하기 위한 Dunnett 다중 비교 검정이 포함됩니다. 모든 평균 비교 방법에는 평균에 대한 합동 분산 추정값이 사용됩니다. "단계별 개별 쌍 비교(Newman-Keuls)" 방법을 제외한 각 "평균 비교" 옵션은 그림 옆에 비교 원을 추가하고 보고

서 창에 특정 보고서를 추가합니다. 비교 원에 대한 자세한 내용은 "비교 원 사용"(175 페이지)에서 확인하십시오.

참고 : 평균을 비교하는 또 다른 방법은 ANOM 입니다 . 자세한 내용은 " 평균 분석 방법 "(170 페이지) 에서 확인하십시오 .

옵션	설명	비모수 메뉴 옵션
개별 쌍 비교 (스튜던트 t)	스튜던트 *t* 검정을 사용하여 개별 쌍별 비교를 수행합니다. 다중 쌍별 검정을 수행하는 경우 추론 간에 보호가 적용되지 않습니다. 따라서 가설 검정의 유의 수준 (제 1 종 오류 비율) 이 개별 검정의 유의 수준보다 높습니다. 자세한 내용은 " 개별 쌍 비교 (스튜던트 t)"(177 페이지) 에서 확인하십시오.	비모수 > 비모수 다중 비교 > Wilcoxon 개별 쌍 비교
전체 쌍 비교 (Tukey HSD)	평균 간의 모든 차이에 대해 크기 조정된 검정을 표시합니다. 이는 Tukey 또는 Tukey-Kramer HSD(Honestly Significant Difference) 검정 (Tukey 1953, Kramer 1956) 입니다. 이 검정은 표본 크기가 동일한 경우에는 정확한 유의 수준 검정이고, 표본 크기가 서로 다른 경우에는 보수적인 검정입니다(Hayter 1984). 자세한 내용은 "전체 쌍 비교 (Tukey HSD)"(177 페이지) 에서 확인하십시오.	비모수 > 비모수 다중 비교 > Steel-Dwass 전체 쌍 비교
최량 비교 (Hsu MCB)	평균이 알 수 없는 최대값보다 작거나 알 수 없는 최소값보다 큰지 여부를 검정합니다. 이는 Hsu MCB 검정 (Hsu 1996, Hsu 1981) 입니다. 자세한 내용은 " 최량 비교 (Hsu MCB)"(177 페이지) 에서 확인하십시오.	없음
대조군 비교 (Dunnett)	평균이 대조군의 평균과 다른지 여부를 검정합니다. 이는 Dunnett 검정 (Dunnett 1955) 입니다. 자세한 내용은 " 대조군 비교 (Dunnett)"(178 페이지) 에서 확인하십시오.	비모수 > 비모수 다중 비교 > Steel 대조군 비교

옵션	설명	비모수 메뉴 옵션
단계별 개별 쌍 비교 (Newman-Keuls)	단계별 절차에서 스튜던트화 범위 검정을 사용하여 평균 간에 차이가 있는지 여부를 검정합니다. 이는 Newman-Keuls 또는 Student-Newman-Keuls 방법 (Keuls, 1952) 입니다. 이 검정은 Tukey HSD 검정보다 덜 보수적입니다. 자세한 내용은 "단계별 개별 쌍 비교 (Newman-Keuls)"(178 페이지)에서 확인하십시오.	없음

참고 : 블록 열을 지정한 경우에는 블록 평균에 맞게 조정된 데이터에 대해 다중 비교 방법이 수행됩니다 .

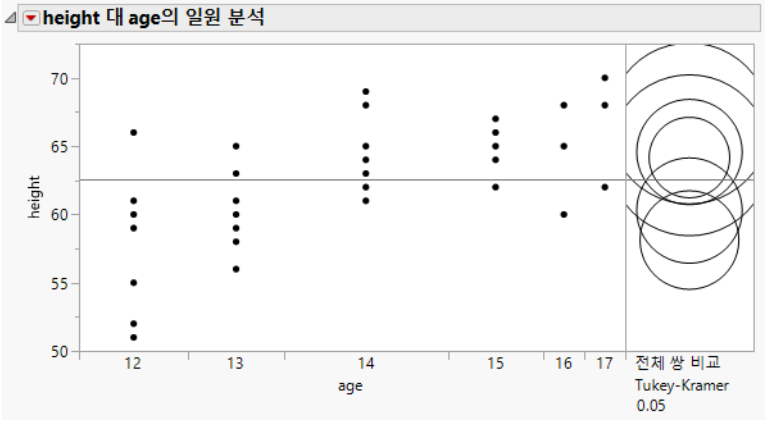
이러한 모든 검정을 보여 주는 예는 "예 : 4 가지 평균 비교 검정 대조"(203 페이지)에서 확인하십시오.

비교 원 사용

"단계별 개별 쌍 비교(Newman-Keuls)" 방법을 제외한 각 다중 비교 검정은 그룹 평균 비교를 시각적으로 나타내는 **비교 원** 그림으로 시작됩니다. 그림 6.10 에서는 "전체 쌍 비교(Tukey HSD)" 방법에 사용되는 비교 원을 보여 줍니다. 다른 비교 검정의 경우에는 원의 반지름이 이보다 길거나 짧습니다.

참고 : 비교 원 그림을 영구적으로 숨기려면 **파일 > 환경 설정 > 플랫폼 > 일원 분석**을 선택하고 **비교 원** 옵션을 선택 취소하십시오 .

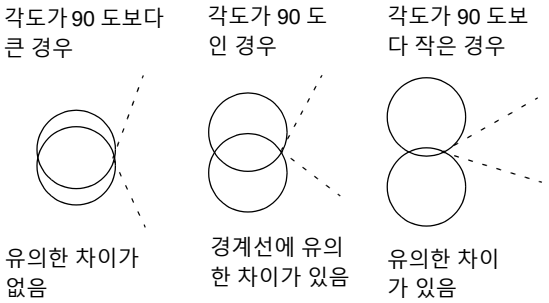
그림 6.10 그룹 평균의 시각적 비교



비교 원의 교차 영역을 검토하여 각 그룹 평균 쌍을 비교할 수 있습니다 . 교차점의 외각은 해당 그룹 평균 간에 유의한 차이가 있는지 여부를 알려 줍니다 (그림 6.11).

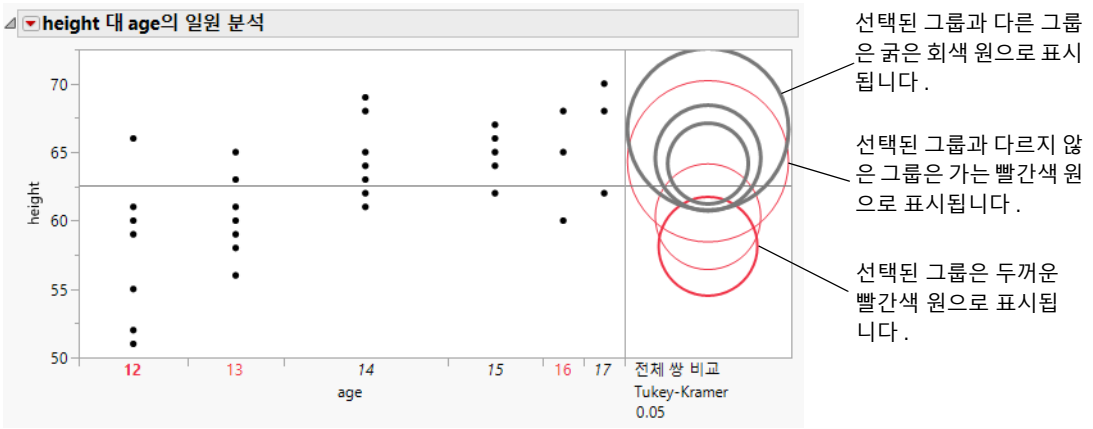
- 유의한 차이가 있는 평균의 경우 비교 원이 교차하지 않거나 약간 교차하므로 교차점의 외각이 90 도 미만입니다 .
- 원의 교차 각도가 90도를 초과하거나 원이 내포되는 경우에는 평균 간에 유의한 차이가 없는 것입니다 .

그림 6.11 교차 각도 및 유의성



교차 각도가 90도에 가까운 경우 비교 원을 클릭하여 선택하면 평균 간에 유의한 차이가 있는지 여부를 확인할 수 있습니다(그림 6.12). 원을 선택 취소하려면 원 외부의 공백을 클릭합니다.

그림 6.12 비교 원 강조 표시



관련 정보

- " 비교 원 "(224 페이지)

개별 쌍 비교 (스튜던트 t)

개별 쌍 비교(스튜던트 t) 검정에서는 각 그룹 수준 쌍에 대한 스튜던트 *t*-검정을 보여 주고 개별 비교만 검정합니다. 이 검정의 예는 "개별 쌍 비교(스튜던트 t) 검정의 예"(196 페이지)에서 확인하십시오.

전체 쌍 비교 (Tukey HSD)

전체 쌍 비교(Tukey HSD) 검정(Tukey-Kramer 라고도 함)에서는 모든 쌍 조합의 유의성 검정을 보호하므로 HSD 구간이 스튜던트 *t* 쌍별 LSD 보다 커집니다. 그래픽에서는 비교 원이 더 크고 차이가 덜 유의하게 나타납니다.

q 통계량은 q* = (1/sqrt(2)) * q 로 계산됩니다. 여기서 q 는 스튜던트화 범위 분포에 필요한 백분위수입니다. 이 검정의 예는 "전체 쌍 비교(Tukey HSD) 검정의 예"(198 페이지)에서 확인하십시오.

최량 비교 (Hsu MCB)

최량 비교(Hsu MCB) 검정에서는 특정 수준의 평균이 나머지 수준의 최대 평균보다 크거나 나머지 수준의 최소 평균보다 작은지 여부를 검정합니다. 자세한 내용은 Hsu 연구 자료(1996)에서 확인하십시오. 이 검정의 예는 "최량 비교(Hsu MCB) 검정의 예"(200 페이지)에서 확인하십시오.

Hsu MCB 검정의 분위수는 범주형 변수의 수준에 따라 달라집니다. 각 수준의 표본 크기가 동일하지 않으면 비교 원 기법이 정확하지 않습니다. 비교 원의 반지름은 수준의 표준 오차에 최대 분위수 값을 곱해서 구합니다. 유의한 차이가 있는지 정확히 평가하려면 *p* 값을 사용합니다. 자세한 내용은 "최대값 및 최소값을 사용한 비교"(177 페이지)에서 확인하십시오.

참고 : MCB 에 의해 최대값 또는 최소값으로 간주되지 않는 평균은 잠재적 최대 또는 최소 평균에 대한 Gupta(1965) 의 선택된 부분집합에도 포함되지 않습니다 .

신뢰 분위수

이 보고서에서는 범주형 변수의 각 수준에 대한 분위수를 제공합니다. 이러한 분위수는 지정된 알파 값에 대응합니다.

최대값 및 최소값을 사용한 비교

이 보고서에는 단측 Dunnett 검정의 *p* 값이 표시됩니다. 최량 수준이 아닌 각 수준에 대해 주어진 *p* 값은 표본 최량 수준의 평균을 대조 수준(잠재적 최량 수준)으로 처리되는 나머지 각 수준의 평균과 비교하는 검정을 위한 것입니다. 표본 최량 수준의 *p* 값은 두 번째 표본 최량 수준의 평균을 대조 수준으로 처리되는 표본 최량 수준의 평균과 비교해서 구합니다.

이 보고서에는 세 개의 열이 표시됩니다 .

수준 범주형 변수의 수준입니다.

최대 p 값 사용 이 열에서는 범주형 변수의 각 수준에 대해 해당 수준의 평균이 나머지 수준의 최대 평균을 초과하는지 검정하기 위한 *p* 값을 제공합니다. 평균이 최대 실제 평균 (알 수 없음) 보다 작거나 같은 수준을 선별해서 제외하려면 이 열의 검정을 사용합니다.

최소 p 값 사용 이 열에서는 범주형 변수의 각 수준에 대해 해당 수준의 평균이 나머지 수준의 최소 평균보다 작은지 검정하기 위한 *p* 값을 제공합니다. 평균이 최소 실제 평균 (알 수 없음) 보다 크거나 같은 수준을 선별해서 제외하려면 이 열의 검정을 사용합니다.

LSD 임계 행렬

첫 번째로 표시되는 보고서는 최대값에 대한 것이고 두 번째는 최소값에 대한 것입니다.

최대값 보고서의 경우 행 평균에서 열 평균을 빼고 LSD를 뺀 값을 포함하는 열이 표시됩니다. 값이 양수이면 해당 행 평균이 열 평균보다 유의하게 높으며 열 평균이 최대값이 아님을 나타냅니다.

최소값 보고서의 경우 행 평균에서 열 평균을 빼고 LSD를 더한 값을 포함하는 열이 표시됩니다. 값이 음수이면 해당 행 평균이 열 평균보다 유의하게 낮으며 열 평균이 최소값이 아님을 나타냅니다.

대조군 비교 (Dunnett)

대조군 비교(Dunnett)에서는 일련의 평균을 대조군 평균과 비교합니다. 산출되는 LSD는 중간 비교 개수를 제한할 수 있도록 크기가 조정되므로 스튜던트 *t*와 Tukey-Kramer LSD 사이에 있습니다. 이 검정의 예는 "대조군 비교(Dunnett)의 예"(201 페이지)에서 확인하십시오.

Dunnett 검정 보고서에는 |*d*| 분위수가 표시되며 이를 스튜던트 *t* 통계량과 유사한 방식으로 사용할 수 있습니다. LSD 임계 행렬에는 차이 - LSD의 절대값이 표시됩니다. 값이 양수이면 해당 평균은 대조군 평균에서 멀리 떨어져 LSD보다 크므로 유의한 차이가 있는 것입니다.

단계별 개별 쌍 비교 (Newman-Keuls)

단계별 개별 쌍 비교(Newman-Keuls) 검정에서는 반복적인 단계별 절차를 사용하여 표본 평균을 비교합니다. 각 반복에서는 두 그룹 평균 사이의 차이를 검정하기 위해 Tukey HSD 검정이 사용됩니다. 이 검정의 예는 " 단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예 "(202 페이지) 에서 확인하십시오.

주의 : Newman-Keuls 검정은 모임별 오차율을 제어하지 않습니다. 이 절차의 결과를 해석할 때 주의하십시오.

J 그룹 평균 검정에 다음 절차가 사용됩니다.

다음을 정의합니다.

J = 그룹 수 (그룹 평균을 기준으로 오름차순 정렬됨)
N = 관측값 수

d = 자유도 . *N* - *J* 로 계산

i = 비교 대상 중 가장 작은 그룹 평균의 인덱스입니다 .

j = 비교 대상 중 가장 큰 그룹 평균의 인덱스입니다 .

k = 절차 중의 모든 비교에서 *j* 의 최소값

절차를 시작할 때 *i* = 1, *j* = *J* 및 *k* = 2 로 설정합니다 .

1. 그룹 *i* 및 *j* 에 대해 Tukey HSD 검정을 수행합니다 . 여기서 적절한 분위수를 찾기 위한 그룹 수는 *j* - i + 1 입니다 .
   - 검정이 유의하면 그룹 *i* 및 *j* 에 유의한 차이가 있는 것으로 결정됩니다 . *j* 를 1 만큼 줄입니다 .

     그 결과 *j* 가 *k* 보다 작아지면 *i* 를 1 만큼 늘리고 *k* = max(*i*, *j*) + 1, *j* = *J* 로 설정한 다음 2 단계로 진행합니다 .

     그 결과 *j* 가 *k* 보다 크거나 같아지면 2 단계로 진행합니다 .
   - 검정이 유의하지 않으면 그룹 *i* 및 *j* 에 유의한 차이가 있는 것으로 결정되지 않습니다 . *i* 를 1 만큼 늘리고 *k* = *j* + 1, *j* = *J* 로 설정한 다음 2 단계로 진행합니다 .
2. *k* 값에 따라 절차를 계속할지 중지할지 결정합니다 .
   - *k* 가 *i* 보다 크고 *k* 가 *J* 보다 작거나 같은 경우 1 단계를 반복합니다 .
   - *k* 가 *i* 보다 작거나 같거나 *k* 가 *J* 보다 크면 절차를 중지합니다 . 검정되지 않은 나머지 모든 범위는 유의한 차이가 없는 것으로 간주됩니다 .

Tukey HSD 에 사용되는 분위수는 검정마다 다르며 , 검정 대상 평균 사이에 있는 그룹 평균의 개수에 따라 달라집니다 . "ewman-Keuls" 보고서에서 " 고려되는 최소 분위수 "(" 최소 q*" 라는 라벨이 지정됨) 는 위의 절차에 사용된 최소 스튜던트화 범위 분위수를 2 의 제곱근으로 나눈 것입니다 .

검정 결과는 " 연결 문자 보고서 " 에 보고됩니다 .

Newman-Keuls 검정에 대한 자세한 내용은 Howell 연구 자료 (2013) 에서 확인하십시오 .

참고 : 단계별 개별 쌍 비교 (Newman-Keuls) 검정을 사용할 때는 " 비교 원 " 그래프에 평균 원이 추가되지 않습니다 . 이는 검정되는 두 평균 사이에 있는 평균의 개수에 따라 각 비교 방법마다 경계가 다르기 때문입니다 . 이로 인해 각 원의 크기가 다르게 됩니다 .

평균 비교 옵션

" 평균 비교 " 보고서에는 다음과 같은 옵션을 제공하는 빨간색 삼각형 메뉴가 포함됩니다 .

차이 행렬 모든 평균 차이가 포함된 테이블을 표시하거나 숨깁니다 .

신뢰 분위수 평균 비교 절차에 사용되는 임계값 및 유의 수준 (α) 을 표시하거나 숨깁니다 .

LSD 임계 행렬 (Newman-Keuls 검정에는 사용할 수 없음) 평균의 쌍별 차이에서 이러한 평균에 대한 최소 유의차를 뺀 행렬을 표시하거나 숨깁니다 . 양수 값은 유의차가 있는 평균의 쌍을 나타냅니다 .

연결 문자 보고서 (스튜던트 *t*, Tukey HSD 및 Newman-Keuls 검정에만 사용 가능) 기존의 문자로 코딩된 보고서를 표시하거나 숨깁니다 . 이 보고서에서 문자를 공유하지 않는 평균에는 유의차가 있습니다 .

정렬된 차이 보고서 (스튜던트 *t* 및 Tukey HSD 검정에만 사용 가능) 모든 쌍의 양수 쪽 차이 , 차이의 표준 오차 , 신뢰 구간 및 *p* 값과 함께 신뢰 구간이 중첩 표시된 차이 크기 그림을 표시하거나 숨깁니다 . 해당 차이 막대가 완전히 포함되지 않은 신뢰 구간은 서로 유의한 차이가 없는 평균 쌍을 나타냅니다 .

상세 비교 보고서 (스튜던트 *t* 검정에만 사용 가능) 각 비교에 대한 상세 보고서를 표시하거나 숨깁니다 . 각 섹션에서는 수준 , 표준 오차 및 신뢰 구간 간의 차이와 *t* 비 , *p* 값 및 자유도를 보여 줍니다 . 각 보고서의 오른쪽에는 비교를 보여 주는 그림이 표시됩니다 .

비모수 검정

비모수 검정은 일반적인 분산 분석의 정규성 가정이 실용적이지 않을 때 유용합니다 . "비모수" 옵션은 그룹 간에 평균 또는 중앙값이 동일하다는 가설을 검정하기 위한 몇 가지 방법을 제공합니다 . "비모수 다중 비교" 절차는 쌍별 비교의 전체 오차율을 제어하는 데도 사용할 수 있습니다 . 비모수 검정에는 순위 점수라고 하는 반응 순위의 함수가 사용됩니다 . 자세한 내용은 Hajek 연구 자료 (1969) 및 SAS Institute Inc. 연구 자료 (2020b) 에서 확인하십시오 .

참고 : 시작 창에서 블록 변수를 지정하고 블록 및 X 요인 수준의 각 조합에 동일한 개수가 있는 경우 Friedman 순위 검정이 사용 가능한 유일한 " 비모수 " 옵션입니다 . 시작 창에서 블록 변수를 지정하고 개수가 동일하지 않은 경우 사용 가능한 " 비모수 " 옵션이 없습니다 .

다음 사항에 유의하십시오 .

- Wilcoxon, 중앙값 , Van der Waerden 및 Friedman 순위 검정의 경우 X 요인의 수준이 세 개 이상이면 일원 검정에 대한 카이제곱 근사가 수행됩니다 .
- 블록 열을 지정하면 블록 평균을 사용하여 중심화된 데이터 값에 대해 비모수 검정 (Friedman 순위 검정 제외) 이 수행됩니다 .

Wilcoxon 검정 Wilcoxon 순위 점수를 기반으로 검정을 수행합니다 . Wilcoxon 순위 점수는 데이터의 단순 순위입니다 . Wilcoxon 검정은 로지스틱 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 요인의 수준이 세 개 이상이면 Kruskal-Wallis 검정이 수행됩니다 . 보고서에 대한 자세한 내용은 "Wilcoxon, 중앙값 , Van der Waerden 및 Friedman 순위 검정 보고서 "(181 페이지) 에서 확인하십시오 . 예는 " 비모수 Wilcoxon 검정의 예 "(204 페이지) 에서 확인하십시오 .

Wilcoxon 검정은 Mann-Whitney 검정이라고도 합니다 .

중앙값 검정 중앙값 순위 점수를 기반으로 검정을 수행합니다 . 중앙값 순위 점수는 순위가 중앙값 순위보다 위인지 아래인지에 따라 0 또는 1 이 됩니다 . 중앙값 검정은 이중 지수 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 보고서에 대한 자세한 내용은 "Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서 "(181 페이지)에서 확인하십시오 .

Van der Waerden 검정 Van der Waerden 순위 점수를 기반으로 검정을 수행합니다 . Van der Waerden 순위 점수는 정규 분포 함수의 역함수를 적용하여 정규 점수로 변환된 관측값의 개수에 1 을 더한 값으로 데이터의 순위를 나눈 값입니다 . Van der Waerden 검정은 정규 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 보고서에 대한 자세한 내용은 "Wilcoxon, 중앙값 , Van der Waerden 및 Friedman 순위 검정 보고서 "(181 페이지) 에서 확인하십시오 .

Kolmogorov Smirnov 검정 (X 요인의 수준이 두 개인 경우에만 사용 가능) 반응의 분포가 그룹 간에 동일한지 여부를 검정하는 경험적 분포 함수를 기반으로 검정을 수행합니다 . 근사 검정과 정확 검정이 모두 제공됩니다 . 보고서에 대한 자세한 내용은 "Kolmogorov-Smirnov 2 표본 검정 보고서 "(182 페이지) 에서 확인하십시오 .

Friedman 순위 검정 (시작 창에서 블록 변수를 지정한 경우에만 사용 가능) Friedman 순위 점수를 기반으로 검정을 수행합니다 . Friedman 순위 점수는 블록 변수의 각 수준 내에 있는 데이터의 순위입니다 . 이 검정의 모수 버전은 반복 측정 ANOVA 입니다 . 보고서에 대한 자세한 내용은 "Wilcoxon, 중앙값 , Van der Waerden 및 Friedman 순위 검정 보고서 "(181 페이지) 에서 확인하십시오 .

참고 : 각 블록 내의 관측값 수는 동일해야 합니다 .

JMP PRO **정확 검정** 정확한 Wilcoxon, 중앙값 , van der Waerden 및 Kolmogorov-Smirnov 검정을 수행하기 위한 옵션을 제공합니다 . 이러한 옵션은 X 요인의 수준이 두 개인 경우에만 사용할 수 있습니다 . 근사 검정과 정확 검정 모두의 결과가 제공됩니다 .

보고서에 대한 자세한 내용은 "2 표본 : 정확 검정 "(182 페이지)에서 확인하십시오 . Wilcoxon 정확 검정의 관련 예는 " 비모수 Wilcoxon 검정의 예 "(204 페이지) 에서 확인하십시오 .

Wilcoxon, 중앙값 , Van der Waerden 및 Friedman 순위 검정 보고서

각 검정의 보고서에는 기술 통계량과 검정 결과가 표시됩니다 . 검정 결과는 " 일원 검정 , 카이제곱 근사 " 보고서에 표시되며 , X 변수의 수준이 정확히 두 개뿐이면 "2 표본 검정 , 정규 근사 " 보고서도 표시됩니다 . 표시되는 기술 통계량은 다음과 같습니다 .

수준 X 의 수준입니다 .

개수 각 수준의 빈도입니다 .

점수 합 각 수준의 순위 점수 합입니다 .

기대 점수 클래스 수준 간에 차이가 없다는 귀무가설하의 기대 점수입니다 .

점수 평균 각 수준의 평균 순위 점수입니다 .

(평균 - 평균 0)/ 표준 0 표준화된 점수입니다 . 평균 0 은 귀무가설하의 기대 평균 점수입니다 . 표준 0 은 귀무가설하의 기대 점수 합에 대한 표준편차입니다 . 귀무가설은 그룹 간에 그룹 평균 또는 중앙값이 동일한 위치에 있다는 것입니다 .

2 표본 검정 , 정규 근사

X 의 수준이 정확히 두 개뿐이면 "2 표본 검정 , 정규 근사 " 보고서가 표시됩니다 . 이 보고서에서는 다음을 제공합니다 .

S 더 적은 수가 관측값이 있는 수준의 순위 점수 합을 제공합니다 .

Z 정규 근사 검정에 대한 검정 통계량을 제공합니다 . 자세한 내용은 "2 표본 정규 근사 "(229 페이지) 에서 확인하십시오 .

Prob>|Z| 표준 정규 분포를 기준으로 정규 근사 검정의 *p* 값을 제공합니다 .

일원 검정 , 카이제곱 근사

이 보고서에서는 위치에 대한 카이제곱 검정의 결과를 제공합니다 . 자세한 내용은 Conover 연구 자료 (1999) 에서 확인하십시오 .

카이제곱 카이제곱 검정 통계량의 값을 제공합니다 . 자세한 내용은 " 일원 카이제곱 근사 "(230 페이지) 에서 확인하십시오 .

DF 검정의 자유도를 제공합니다 .

Prob>ChiSq 검정의 *p* 값을 제공합니다 . *p* 값은 자유도가 X 의 수준 수에서 1 을 뺀 값에 해당하는 카이제곱 분포를 기준으로 합니다 .

JMP PRO 2 표본 : 정확 검정

데이터가 희소 데이터이거나 비대칭적이거나 동일 값이 많은 경우에는 점근적 동작을 기반으로 한 근사보다 정확 검정이 더 적합할 수 있습니다 . X 의 수준이 정확히 두 개뿐인 경우 JMP Pro 에서는 정확 검정에 대한 검정 통계량이 계산됩니다 . **비모수 > 정확 검정**을 선택하고 원하는 검정을 선택하십시오 . 그러면 "2 표본 : 정확 검정 " 보고서가 표시됩니다 . 이 보고서에서는 다음을 제공합니다 .

S 더 작은 그룹의 관측값에 대한 순위 점수 합을 제공합니다 . X 의 두 수준에 동일한 수의 관측값이 있으면 S 값은 값 순서화 시 X 의 마지막 수준에 대응합니다 .

Prob ≤ S 검정의 단측 *p* 값을 제공합니다 .

Prob ≥ |S-Mean| 검정의 양측 *p* 값을 제공합니다 .

Kolmogorov-Smirnov 2 표본 검정 보고서

Kolmogorov-Smirnov 검정은 X 의 수준이 정확히 두 개뿐인 경우에만 사용할 수 있습니다 . 이 보고서에는 기술 통계량과 검정 결과가 표시됩니다 . 표시되는 기술 통계량은 다음과 같습니다 .

수준 X 의 두 수준입니다 .

개수 각 수준의 빈도입니다 .

최대값에서의 EDF X 의 수준별로 두 EDF 간의 차이가 최대인 X 값 위치에서 해당 수준에 대한 EDF(경험적 누적 분포 함수) 의 값을 제공합니다 . " 합계 " 행에서는 두 EDF 간의 차이가 최대인 X 값 위치에서의 합동 EDF(전체 데이터 집합에 대한 EDF) 값을 제공합니다 .

최대값에서 평균과의 편차 각 수준에 대해 다음 단계에서 구한 값을 제공합니다 .

- 해당 수준의 최대값에서의 EDF 와 합동 데이터 집합 (합계) 의 최대값에서의 EDF 사이의 차이를 계산합니다 .
- 이 차이를 해당 수준에 포함된 관측값 수 (" 개수 " 로 제공됨) 의 제곱근으로 곱합니다 .

Kolmogorov-Smirnov 점근적 검정

이 보고서에서는 검정에 대한 상세 정보를 제공합니다 .

KS 다음과 같이 계산된 Kolmogorov-Smirnov 통계량입니다 .

$$\mathrm{KS} = \max_j \sqrt{\frac{1}{n}\sum_i n_i (F_i(x_j) - F(x_j))^2}$$

이 계산식에서는 다음과 같은 표기를 사용합니다 .

- $x_j (j = 1,..., n)$ 은 관측값입니다 .
- n_i 는 X 의 i 번째 수준에 포함된 관측값의 개수입니다 .
- F 는 합동 경험적 누적 분포 함수입니다 .
- F_i 는 X 의 i 번째 수준에 대한 경험적 누적 분포 함수입니다 .

Kolmogorov-Smirnov 통계량의 이 버전은 X 의 수준이 세 개 이상인 경우에도 적용됩니다 . 하지만 JMP 에서는 X 의 수준이 두 개뿐일 때만 Kolmogorov-Smirnov 분석을 수행합니다 .

KSa $\mathrm{KS}\sqrt{n}$ 으로 계산된 점근적 Kolmogorov-Smirnov 통계량입니다 . 여기서 n 은 총 관측값 수입니다 .

D=max|F1-F2| 두 수준의 EDF 간 최대 절대 편차입니다 . 이 값은 두 표본을 비교하는 데 일반적으로 사용되는 버전의 Kolmogorov-Smirnov 통계량입니다 .

Prob > D 검정의 p 값입니다 . 이 값은 수준 간에 차이가 없다는 귀무가설하에 D 가 계산된 값을 초과할 확률입니다 .

D+ = max(F1-F2) 첫 번째 그룹의 수준이 두 번째 그룹의 수준을 초과한다는 대립가설에 대한 단측 검정 통계량입니다 .

Prob > D+ D+ 검정의 p 값입니다 .

D- = max(F2-F1) 두 번째 그룹의 수준이 첫 번째 그룹의 수준을 초과한다는 대립가설에 대한 단측 검정 통계량입니다 .

Prob > D- D- 검정의 p 값입니다 .

JMP PRO Kolmogorov-Smirnov 정확 검정

Kolmogorov-Smirnov 정확 검정에 대한 보고서에서는 점근적 검정과 동일한 통계량을 제공하지만 *p* 값은 정확하게 계산됩니다.

비모수 다중 비교

이 옵션은 비모수 다중 비교를 수행하기 위한 몇 가지 방법을 제공합니다. "Wilcoxon 개별 쌍 비교 검정"을 제외하고 이러한 검정은 모두 순위를 기반으로 하며 전체 실험 오차율을 제어합니다. 이러한 검정에 대한 자세한 내용은 Dunn 연구 자료(1964) 및 Hsu 연구 자료(1996)에서 확인하십시오. 보고서에 대한 자세한 내용은 "비모수 다중 비교 절차"(184 페이지)에서 확인하십시오.

비모수 다중 비교 절차

Wilcoxon 개별 쌍 비교 각 쌍에 대한 Wilcoxon 검정을 수행합니다. 이 절차에서는 전체 유의 수준을 제어하지 않습니다. 이 값은 "평균 비교" 메뉴에 있는 **개별 쌍 비교 (스튜던트 t)** 옵션의 비모수 버전입니다. 자세한 내용은 "Wilcoxon 개별 쌍 비교, Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교"(185 페이지)에서 확인하십시오.

Steel-Dwass 전체 쌍 비교 개별 쌍에 대해 Steel-Dwass 검정을 수행합니다. 이 검정은 "평균 비교" 메뉴에 있는 **전체 쌍 비교 (Tukey HSD)** 옵션의 비모수 버전입니다. 자세한 내용은 "Wilcoxon 개별 쌍 비교, Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교"(185 페이지)에서 확인하십시오.

Steel 대조군 비교 각 수준을 대조 수준과 비교합니다. 이 검정은 "평균 비교" 메뉴에 있는 **대조군 비교 (Dunnett)** 옵션의 비모수 버전입니다. 자세한 내용은 "Wilcoxon 개별 쌍 비교, Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교"(185 페이지)에서 확인하십시오.

Dunnett 전체 쌍 비교 (결합 순위 기준) "Steel-Dwass 전체 쌍 비교" 옵션과 유사하게 각 쌍에 대한 비교를 수행합니다. Dunnett 방법은 비교 대상인 쌍뿐만 아니라 모든 데이터에 대한 순위를 계산합니다. 보고되는 *p* 값은 Bonferroni 조정을 반영합니다. 이 값은 조정되지 않은 *p* 값에 비교 개수를 곱한 것입니다. 조정된 *p* 값이 1 을 초과할 경우 1 로 보고됩니다. 자세한 내용은 "Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준)"(186 페이지)에서 확인하십시오.

Dunnett 대조군 비교 (결합 순위 기준) "Steel 대조군 비교" 옵션과 유사하게 각 수준을 대조 수준과 비교합니다. Dunnett 방법은 비교 대상인 쌍뿐만 아니라 모든 데이터에 대한 순위를 계산합니다. 보고되는 *p* 값은 Bonferroni 조정을 반영합니다. 이 값은 조정되지 않은 *p* 값에 비교 개수를 곱한 것입니다. 조정된 *p* 값이 1 을 초과할 경우 1 로 보고됩니다. 자세한 내용은 "Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준)"(186 페이지)에서 확인하십시오.

Wilcoxon 개별 쌍 비교 , Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교

이러한 다중 비교 절차에 대한 보고서에서는 검정 결과와 신뢰 구간을 제공합니다 . 이러한 검정의 경우 해당 비교에 사용된 두 수준만 결합하여 구한 표본 내에서 관측값의 순위가 매겨집니다 .

q* 신뢰 구간을 계산하는 데 사용된 분위수입니다 .

알파 신뢰 구간을 계산하는 데 사용된 유의 수준입니다 . " 일원 분석 " 메뉴에서 "α 수준 설정 " 옵션을 선택하여 신뢰 수준을 변경할 수 있습니다 .

수준 쌍별 비교에서 사용된 X 변수의 첫 번째 수준입니다 .

- 수준 쌍별 비교에서 사용된 X 변수의 두 번째 수준입니다 .

점수 평균 차이 첫 번째 수준 (" 수준 ") 에 포함된 관측값의 순위 점수 평균에서 두 번째 수준 ("- 수준 ") 에 포함된 관측값의 순위 점수 평균을 뺀 값입니다 . 이때 연속성 수정이 적용됩니다 .

첫 번째 수준의 관측값 수는 n_1 로 나타내고 두 번째 수준의 관측값 수는 n_2 로 나타냅니다 . 이러한 두 수준으로 구성된 표본 내에서 관측값의 순위가 매겨집니다 . 동일 순위에 대해서는 평균을 구합니다 . 첫 번째 수준의 순위 합은 ScoreSum_1 으로 나타내고 두 번째 수준의 순위 합은 ScoreSum_2 로 나타냅니다 .

평균 점수의 차이가 양수인 경우 " 점수 평균 차이 " 는 다음과 같이 정의됩니다 .

$$\text{점수 평균 차이} = (\text{ScoreSum}_1 - 0.5) / n_1 - (\text{ScoreSum}_2 + 0.5) / n_2$$

평균 점수의 차이가 음수인 경우 " 점수 평균 차이 " 는 다음과 같이 정의됩니다 .

$$\text{점수 평균 차이} = (\text{ScoreSum}_1 + 0.5) / n_1 - (\text{ScoreSum}_2 - 0.5) / n_2$$

차이 표준 오차 점수 평균 차이의 표준 오차입니다 .

Z 평균에 차이가 없다는 귀무가설하에서 점근적 표준 정규 분포를 따르는 표준화된 검정 통계량입니다 .

p 값 Z 를 기반으로 한 점근적 검정의 p 값입니다 .

Hodges-Lehmann 위치 이동에 대한 Hodges-Lehmann 추정량입니다 . 첫 번째 수준에 포함된 관측값에서 두 번째 수준에 포함된 관측값을 빼서 구한 모든 쌍체 차이가 생성됩니다 . Hodges-Lehmann 추정량은 이러한 차이의 중앙값입니다 . 차이 그림 막대 차트에는 Hodges-Lehmann 추정값의 크기가 표시됩니다 .

CL 하한 Hodges-Lehmann 통계량의 신뢰 하한입니다 .

참고 : 그룹 표본 크기가 메모리 문제를 일으킬 만큼 큰 경우에는 계산되지 않습니다 .

CL 상한 Hodges-Lehmann 통계량의 신뢰 상한입니다 .

참고 : 그룹 표본 크기가 메모리 문제를 일으킬 만큼 큰 경우에는 계산되지 않습니다 .

Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준)

이러한 비교 절차는 전체 데이터 집합의 관측값 순위를 기반으로 합니다 . "Dunnett 대조군 비교 (결합 순위 기준)" 검정의 경우 대조 수준을 선택해야 합니다 .

수준 쌍별 비교에서 사용된 X 변수의 첫 번째 수준입니다 .

- 수준 쌍별 비교에서 사용된 X 변수의 두 번째 수준입니다 .

점수 평균 차이 첫 번째 수준 (" 수준 ") 에 포함된 관측값의 순위 점수 평균에서 두 번째 수준 ("- 수준 ") 에 포함된 관측값의 순위 점수 평균을 뺀 값입니다 . 이때 연속성 수정이 적용됩니다 . 순위는 전체 표본 내의 관측값에 순위를 매기는 방식으로 결정됩니다 . 동일 순위에 대해서는 평균을 구합니다 . 연속성 수정에 대한 자세한 내용은 " 점수 평균 차이 "(185 페이지) 에서 확인하십시오 .

차이 표준 오차 점수 평균 차이의 표준 오차입니다 .

Z 평균에 차이가 없다는 귀무가설하에서 점근적 표준 정규 분포를 따르는 표준화된 검정 통계량입니다 .

p 값 Z 를 기반으로 한 점근적 검정의 *p* 값입니다 .

이분산

JMP는 그룹 분산의 동일 여부에 대한 4가지 검정을 제공합니다. 그룹 간에 분산이 동일하지 않으면 분산 분석의 일반적인 가정이 충족되지 않습니다. 예를 들어 ANOVA *F* 검정이 유효하지 않으며 동일한 그룹 분산을 가정하지 않는 분석을 사용해야 합니다. 처음 세 가지 등분산 검정의 기본 개념은 각 그룹의 퍼짐 정도를 측정하기 위해 구성된 새 반응 변수에 대해 분산 분석을 수행하는 것입니다. 네 번째 검정은 정규 분포하에서 가능도비 검정과 유사한 Bartlett 검정입니다. 블록 변수가 시작 창에 지정된 경우 이분산 옵션을 사용할 수 없습니다.

참고 : 이분산을 검정하는 또 다른 방법은 ANOMV 입니다 . 자세한 내용은 " 평균 분석 방법 "(170 페이지) 에서 확인하십시오 .

다음과 같은 등분산 검정을 사용할 수 있습니다 .

O'Brien 새 변수의 그룹 평균이 원래 반응의 그룹 표본 분산과 동일하도록 종속 변수를 생성합니다 . O'Brien 변수에 대한 ANOVA 는 실제로는 그룹 표본 분산에 대한 ANOVA 입니다 (O'Brien 1979, Olejnik and Algina 1987).

Brown-Forsythe 반응이 각 관측값과 그룹 중앙값의 차이에 대한 절대값인 경우 ANOVA 에서 구한 *F*- 검정을 표시합니다 (Brown and Forsythe 1974).

Levene 반응이 각 관측값과 그룹 중앙값의 차이에 대한 절대값인 경우 ANOVA 에서 구한 *F*- 검정을 표시합니다 (Levene 1960). 퍼짐 정도는 SAS 기본값인 $z_{ij}^2 = (y_{ij} - \bar{y}_i)^2$ 대신 $z_{ij} = |y_{ij} - \bar{y}_i|$ 으로 측정됩니다 .

Bartlett 표본 분산의 가중 산술평균을 표본 분산의 가중 기하평균과 비교합니다 . 기하평균은 항상 산술평균보다 작거나 같으며 모든 표본 분산이 동일한 경우에만 이 둘이 동일하게 유지

됩니다 . 그룹 분산 간의 변동이 클수록 이 두 평균의 차이가 커집니다 . 이 두 평균으로부터 χ^2 분포 (사실상 특정 공식에서는 F 분포) 에 근사한 값을 산출하는 함수가 생성됩니다 . 값이 크면 산술 또는 기하 비율의 값이 큰 것이므로 그룹 분산의 변동 폭이 커집니다 . Bartlett 카이 제곱 검정 통계량을 자유도로 나누면 테이블에 표시된 F 값이 됩니다 . Bartlett 검정은 정규성 가정 위배에 대해 그다지 로버스트하지 않습니다 (Bartlett and Kendall 1946).

F- 검정 양측 (X 변수의 수준이 두 개인 경우에만 사용 가능) 검정 대상 그룹이 두 개뿐이면 이분산에 대한 표준 F- 검정도 수행됩니다 . F- 검정은 큰 분산 추정값 대 작은 분산 추정값의 비율입니다 . F 분포의 p 값은 양측 검정을 구성하기 위해 두 배가 됩니다 .

참고 : 블록 열을 지정한 경우에는 블록 평균에 맞게 데이터가 조정된 후에 데이터에 대해 분산 검정이 수행됩니다 .

자세한 내용은 " 이분산 옵션의 예 "(207 페이지) 에서 확인하십시오 .

분산 동일 여부 검정 보고서

" 분산 동일 여부 검정 " 보고서에는 그룹 평균과 총 평균 및 중앙값 사이의 차이가 표시되고 검정 절차에 대한 요약이 제공됩니다 .

등분산 검정 결과 그룹 분산에 유의한 차이가 있는 것으로 나타나면 일반적인 ANOVA 검정 대신 Welch 검정을 사용하십시오 . Welch 통계량은 일반적인 ANOVA F- 검정을 기반으로 합니다 . 하지만 그룹 평균 분산의 역수로 평균에 가중치가 적용된다는 차이점이 있습니다 (Welch 1951, Brown and Forsythe 1974, Asiribo and Gurland 1990). 수준이 두 개뿐인 경우 Welch ANOVA 는 이분산 t- 검정과 동등합니다 .

분산 동일 여부 검정 보고서 설명

수준 데이터에서 나타나는 요인 수준을 나열합니다 .

개수 각 수준의 빈도를 기록합니다 .

표준편차 각 요인 수준의 반응 표준편차를 기록합니다 . 이 표준편차는 O'Brien 변수의 평균과 동일합니다 . 데이터에서 한 수준이 한 번만 나타나는 경우에는 표준편차가 계산되지 않습니다 .

평균에 대한 평균 절대 차이 반응 및 그룹 평균의 평균 절대 차이를 기록합니다 . 이 평균 절대 차이는 Levene 변수의 그룹 평균과 동일합니다 .

중앙값에 대한 평균 절대 차이 반응 및 그룹 중앙값의 절대 차이를 기록합니다 . 이 평균 절대 차이는 Brown-Forsythe 변수의 그룹 평균과 동일합니다 .

검정 수행된 검정의 이름을 나열합니다 .

F 비 각 검정에 대해 계산된 F 통계량을 기록합니다 . 자세한 내용은 " 분산 동일 여부 검정 "(227 페이지) 에서 확인하십시오 .

DFNum 각 검정의 분자 자유도를 기록합니다 . 요인에 k 개의 수준이 있는 경우 분자의 자유도는 k - 1 입니다 . 데이터에 한 번만 나타나는 수준은 O'Brien, Brown-Forsythe 또는 Levene 의

검정 통계량을 계산하는 데 사용되지 않습니다 . 이 경우 분자 자유도는 계산에 사용된 수준 수에서 1 을 뺀 값입니다 .

DFDen 각 검정의 분모에 사용된 자유도를 기록합니다 . O'Brien, Brown-Forsythe 및 Levene 의 경우 검정 통계량을 계산하는 데 사용된 각 요인 수준에 대해 자유도를 뺍니다 . 요인에 k 개의 수준이 있는 경우 분모 자유도는 $n - k$ 입니다 .

p 값 모든 수준의 분산이 동일한 경우 *F* 비 값이 계산된 값보다 클 확률입니다 .

참고 : X 변수에 관측값 수가 5 개 미만인 수준이 있으면 경고가 나타납니다 . 표본 크기가 작을 때 위 검정의 성능에 대한 자세한 내용은 Brown and Forsythe 연구 자료 (1974) 및 Miller 연구 자료 (1972) 에서 확인하십시오 .

Welch 검정 보고서 설명

F 비 평균 동일성 검정에 대한 *F*- 검정 통계량을 표시합니다 .

DFNum 검정의 분자 자유도를 기록합니다 . 요인에 *k* 개의 수준이 있는 경우 분자의 자유도는 $k - 1$ 입니다 . 데이터에 한 번만 나타나는 수준은 Welch ANOVA 를 계산하는 데 사용되지 않습니다 . 이 경우 분자 자유도는 계산에 사용된 수준 수에서 1 을 뺀 값입니다 .

DFDen 검정의 분모 자유도를 기록합니다 . 자세한 내용은 " 분산 동일 여부 검정 "(227 페이지) 에서 확인하십시오 .

Prob>F 모든 수준의 평균이 동일한 경우 *F* 값이 계산된 값보다 클 확률입니다 . 관측된 유의 확률이 0.05 이하이면 수준 간에 평균이 동일하지 않다는 증거로 간주됩니다 .

t- 검정 *F* 비와 *t*- 검정 사이의 관계를 보여 줍니다 . *F* 비의 제곱근으로 계산됩니다 . X 요인의 수준이 두 개인 경우에만 표시됩니다 .

동등성 검정

동등성 검정은 **실제적** 관심 영역인 평균의 차이를 검색하려는 경우에 유용합니다. 임계 차이를 지정해야 하며 그보다 작은 차이는 실제적으로 동등한 것으로 간주됩니다. 즉, 두 그룹 평균의 차이가 이 값보다 작거나 같으면 두 그룹이 동등하다고 간주합니다.

자세한 내용은 "동등성 검정의 예"(208 페이지) 에서 확인하십시오.

동등성 검정 보고서

임계값 및 분산 가정을 지정하면 "동등성 검정" 보고서가 나타납니다. 보고서의 전체 제목은 지정된 분산 가정에 따라 "합동 분산을 사용한 동등성 검정" 또는 "이분산을 사용한 동등성 검정" 입니다. 지정된 임계값은 상한과 하한을 의미하며 보고서 위쪽에 표시됩니다. 보고서 위쪽에는 동등성 검정의 α 수준도 포함됩니다. 보고서는 동등성 검정이 포함된 테이블과 이를 표시하는 산점도로 구성됩니다. 동등성 검정 및 신뢰 구간은 스튜던트 *t* 임계값에 기반합니다.

팁 : 동등성 검정의 α 수준을 변경하려면 " 동등성 검정 " 옵션을 선택하기 전에 " 일원 분석 " 의 빨간색 삼각형 메뉴의 α **수준 설정** 옵션을 사용합니다 .

" 합동 분산을 사용한 동등성 검정 " 또는 " 이분산을 사용한 동등성 검정 " 옆에 있는 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

동등성 TOST 검정 동등성 TOST 검정 보고서를 표시하거나 숨깁니다 .

동등성 검정 산점도 동등성 검정 산점도를 표시하거나 숨깁니다 .

동등성 검정 쌍별 비교 모든 쌍별 비교에 대한 실제적 동등성 보고서를 표시하거나 숨깁니다 .

제거 일원 분석 보고서 창에서 동등성 검정 보고서를 제거합니다 .

동등성 TOST 검정

두 개의 단측 검정(TOST) 방법은 평균 사이의 실제적 차이에 대한 검정에 사용됩니다 (Schuirmann 1987). 지정된 분산 가정을 사용하는 두 개의 단측 *t* 검정은 실제 차이가 임계값을 초과한다는 귀무가설에 대해 구성됩니다. 두 검정이 모두 기각되면 평균의 차이가 통계적으로 두 임계값을 초과하지 않습니다. 따라서 그룹은 실제적으로 동등한 것으로 간주됩니다. 검정이 하나만 기각되거나 둘 모두 기각되지 않으면 그룹은 실제적으로 동등하지 않을 수 있습니다.

각 비교에 대해 " 동등성 TOST 검정 " 보고서는 다음과 같은 정보를 제공합니다 .

차이 평균의 추정 차이입니다 .

하한 t 비 , 상한 t 비 단측 합동 분산 유의성 검정의 하한 및 상한 *t* 비입니다 .

하한 p 값 , 상한 p 값 하한 및 상한 *t* 비에 해당하는 *p* 값입니다 .

최대 p 값 하한 및 상한 *p* 값의 최대값입니다 .

하한 및 상한 평균의 차이에 대한 1–2α 신뢰 구간의 한계입니다 .

동등성 검정 산점도

이 산점도는 색상을 사용하여 어떤 평균이 실제로 동등하고 어떤 평균이 실제로 동등하지 않은지를 동등성 검정에 의해 결정된 것을 나타냅니다. 이 그림을 디포그램(diffogram) *또는* 평균-평균 산점도라고도 합니다.

이 그림은 주대각의 단색 참조선과 음영 참조 대역을 표시합니다. 대역의 너비는 실제적 차이의 두 배입니다. 선 세그먼트에서 점의 좌표는 대응 그룹에 대한 평균입니다. 각 선 세그먼트가 쌍별 비교에 대한 1–2α 신뢰 구간에 해당하는 주대각에 암시적인 세 번째 축이 있습니다. 비교 대상인 그룹과 추정 차이를 나타내는 툴팁을 표시하려면 이러한 점 중 하나를 마우스로 가리킵니다. 선 세그먼트가 대각선 대역 내에 완전히 포함되면 평균이 실제적으로 동등합니다.

동등성 검정 산점도에는 다음과 같은 옵션이 있습니다 .

참조선 표시 산점도에 점의 참조선을 표시합니다 . 산점도에 점이 많은 경우 권장되지 않습니다 . 점이 많은 경우 점을 마우스를 가리켜서 라벨을 확인하는 것이 좋습니다 .

로버스트

로버스트 옵션은 데이터 집합에서 이상치 또는 극단 데이터 점의 영향을 줄이기 위한 두 가지 방법으로 " 로버스트 적합 " 과 "Cauchy 적합 " 을 제공합니다 . 블록 변수가 시작 창에 지정된 경우 로버스트 옵션을 사용할 수 없습니다 .

로버스트 적합

" 로버스트 적합 " 옵션은 반응 변수에서 이상치의 영향을 줄입니다 . Huber M- 추정 방법이 사용됩니다 . Huber M- 추정은 다음과 같이 Huber 손실 함수를 최소화하는 모수 추정값을 찾는 방법입니다 .

$$l(\underline{e}) = \sum_i \rho(e_i)$$

다음은 각 요소에 대한 설명입니다 .

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & \text{단, } |e| < k \\ k|e| - \frac{1}{2}k^2 & \text{단, } |e| \geq k \end{cases}$$

e_i 는 잔차를 나타냅니다 .

Huber 손실 함수는 이상치에 벌점을 부과하며, 작은 오차의 경우에는 2 차 형태로 증가하고 큰 오차의 경우에는 선형적으로 증가합니다. 로버스트 적합에 대한 자세한 내용은 Huber 연구 자료 (1973) 및 Huber and Ronchetti 연구 자료 (2009) 에서 확인하십시오. 자세한 내용은 " 로버스트 적합 옵션의 예 "(209 페이지) 에서 확인하십시오.

Cauchy 적합

Cauchy 적합 옵션은 오차가 Cauchy 분포를 따른다고 가정합니다. Cauchy 분포는 정규 분포보다 꼬리가 더 두꺼우므로 이상치의 영향이 줄어듭니다. 이 옵션은 데이터에 포함된 이상치의 비율이 높은 경우에 유용할 수 있습니다. 하지만 데이터가 정규 분포에 가깝고 이상치가 일부만 있는 경우에 이 옵션을 사용하면 추론이 잘못될 수 있습니다. Cauchy 옵션은 최대 가능도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다.

검정력

검정력 옵션은 통계 검정력을 계산하고 지정된 가설 검정에 대한 기타 상세 정보를 제공합니다. 자세한 내용은 " 검정력 옵션의 예 "(211 페이지) 에서 확인하십시오 . 통계 상세 정보는 " 검정력 "(225 페이지) 에서 확인하십시오 .

- LSV(최소 유의값) 는 특정 *p* 값이 알파가 되는 모수 또는 모수 함수의 값입니다 . 즉 , 어떤 *p* 값이 알파가 될 때 효과의 유의성이 얼마나 작게 선언되는지 알 수 있습니다 . LSV 는 확률 척도가 아니라 모수 척도의 유의성에 대한 측정 수단을 제공합니다 . 이를 통해 설계와 데이터의 민감도를 보여 줍니다 .
- LSN(최소 유의 개수) 은 데이터의 형식이 동일한 경우 지정된 *p* 값 알파를 산출할 관측값의 총 개수입니다 . LSN 은 알파 , 시그마 및 델타 값 (유의 수준 , 오차의 표준편차 및 효과 크기) 이 주어진 상태에서 유의한 결과를 얻을 수 있을 정도로 추정값의 분산을 줄이는 데 필요한 관측값의 개수로 정의됩니다 . 유의성을 달성하기 위해 더 많은 데이터가 필요한 경우 LSN 을 통해 필요한 데이터 양을 알 수 있습니다 . LSN 은 검정력이 약 50% 가 되도록 하는 총 관측값 수입니다 .
- **검정력**은 그룹 간에 실제 차이가 존재할 때 유의성을 달성 (*p* 값 < 알파) 할 확률입니다 . 이는 표본 크기 , 효과 크기 , 오차의 표준편차 및 유의 수준에 따라 달라집니다 . 검정력은 특정 유의 수준에서 실험을 통해 차이 (효과 크기) 를 발견할 가능성을 나타냅니다 .

참고 : 일원 배치의 그룹이 두 개뿐인 경우 검정력에 의해 계산된 LSV 는 다중 비교 테이블에 표시된 LSD(최소 유의차) 와 동일합니다 .

검정력 상세 정보 창 및 보고서

검정력 상세 정보 창 및 보고서는 모형 적합 플랫폼을 통해 시작된 일반 적합 플랫폼에서와 동일합니다. 검정력 계산에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

알파 , 시그마 , 델타 및 개수를 나타내는 4 개의 각 열에 단일 값 , 두 개의 값 , 또는 값 시퀀스의 시작 , 끝 및 증분 값을 입력해야 합니다 (그림 6.32). 지정한 값의 가능한 모든 조합에 대해 검정력 계산이 수행됩니다 .

알파 (α) 유의 수준으로 , 0 에서 1 사이입니다 (일반적으로 0.05, 0.01 또는 0.10). 기본적으로 표시되는 값은 0.05 입니다 .

시그마 (σ) 모형의 잔차에 대한 표준 오차입니다 . 기본적으로 제공되는 값은 RMSE(제곱근 평균 제곱 오차) 에서 구한 추정값입니다 .

델타 (δ) 원시 효과 크기입니다 . 효과 크기 계산에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 . 첫 번째 위치는 기본적으로 가설의 제곱근 제곱합을 n 으로 나눈 값 (즉 , $\delta = \sqrt{SS/n}$) 으로 설정되어 있습니다 .

개수 (*n*) 모든 그룹의 총 표본 크기입니다 . 첫 번째 위치에는 기본적으로 실제 표본 크기가 입력되어 있습니다 .

검정력 계산 네 개의 값 , 즉 α, σ, δ 및 *n* 을 모두 사용하여 검정력 (유의한 결과를 얻을 확률) 을 계산합니다 .

최소 유의 개수 계산 주어진 α, σ 및 δ 로 약 50% 의 검정력을 달성하는 데 필요한 관측값 수를 계산합니다 .

최소 유의값 계산 *p* 값이 α 가 되는 모수 또는 선형 적합의 값을 계산합니다 . 계산에는 α, σ, *n*, 그리고 추정값의 표준 오차가 사용됩니다 . 이 기능은 X 요인의 수준이 두 개일 때만 사용할 수 있으며 일반적으로 개별 모수에 사용됩니다 .

조정된 검정력 및 신뢰 구간 후향적으로 검정력을 분석할 경우에는 표준 오차와 검정 모수의 추정값을 사용합니다 .

- 조정된 검정력은 비중심성 모수의 보다 비편향적인 추정값으로부터 계산된 검정력입니다 .
- 조정된 검정력의 신뢰 구간은 비중심성 추정값의 신뢰 구간을 기반으로 합니다 .

랜덤 변동은 원래 **델타**에서 발생하므로 원래 델타에 대해서만 조정된 검정력 및 신뢰 한계가 계산됩니다 .

정규 분위수 그림

다음과 같은 두 가지 유형의 정규 분위수 그림을 생성할 수 있습니다 .

- **분위수별 실제값 그림** : 반응 값 대 정규 분위수 값 그림을 생성합니다 . X 변수의 각 수준에 대해 개별적으로 분위수가 계산되어 그림에 표시됩니다 .
- **실제값별 분위수 그림** : 정규 분위수 값 대 반응 값 그림을 생성합니다 . X 변수의 각 수준에 대해 개별적으로 분위수가 계산되어 그림에 표시됩니다 .

적합선 옵션은 분위수 그림에 적합선을 표시하거나 숨깁니다. **정규 분위수 라벨 옵션은** 정규 분위수 그림의 오른쪽 세로 축 내부 또는 분위수별 실제값 그림의 위쪽 축 내부에 라벨을 표시하거나 숨깁니다. 라벨은 정규 분위수 척도를 나타냅니다.

블록 변수가 시작 창에 지정된 경우 정규 분위수 그림 옵션을 사용할 수 없습니다.

자세한 내용은 "정규 분위수 그림의 예"(213 페이지) 에서 확인하십시오.

CDF 그림

CDF 그림에서는 일원 분석 보고서의 모든 그룹에 대한 누적 분포 함수를 보여 줍니다. CDF 그림은 X 요인의 여러 수준 간에 반응 분포를 비교하려는 경우에 유용합니다. 블록 변수가 시작 창에 지정된 경우 CDF 그림 옵션을 사용할 수 없습니다. 자세한 내용은 "CDF 그림의 예"(214 페이지) 에서 확인하십시오.

밀도

밀도 옵션은 X 요인의 여러 수준 간에 반응의 분포 및 합성을 시각적으로 비교하는 몇 가지 방법을 제공합니다. 다음과 같은 세 가지 밀도 옵션이 있습니다.

- **밀도 비교**: 각 그룹의 밀도 추정값을 나타내는 평활 곡선을 표시합니다. 평활 곡선은 각 그룹의 밀도 추정값입니다.
- **밀도의 합성**: 각 그룹의 반응 개수에 따라 가중치가 적용된 밀도 요약을 표시합니다. "밀도의 합성" 그림에서는 X 값별로 각 그룹이 전체에 얼마나 기여하는지를 보여 줍니다.
- **밀도의 비율**: 각 X 수준에서 그룹이 합계에 기여하는 정도를 전체에 대한 비율로 보여 줍니다.

블록 변수가 시작 창에 지정된 경우 밀도 옵션을 사용할 수 없습니다.

자세한 내용은 "밀도 옵션의 예"(215 페이지)에서 확인하십시오.

매칭 열

매칭 모형 분석을 위해 매칭(ID) 변수를 지정하려면 **매칭 열** 옵션을 사용합니다. **매칭 열** 옵션은 일원 분석에 사용되는 데이터가 매칭(쌍체) 데이터인 경우를 처리합니다. 동일한 참가자로부터 서로 다른 그룹의 관측값을 얻은 경우에 매칭 데이터가 발생할 수 있습니다. 블록 변수가 시작 창에 지정된 경우 매칭 열 옵션을 사용할 수 없습니다. 자세한 내용은 "매칭 열 옵션의 예"(216 페이지)에서 확인하십시오.

참고 : 매칭의 특수한 경우에는 쌍체 *t*- 검정을 수행합니다. **매칭 쌍** 플랫폼에서는 이 유형의 데이터를 처리하지만 데이터가 서로 다른 행이 아니라 서로 다른 열에 쌍으로 구성되어 있어야 합니다.

매칭 열 옵션은 다음과 같은 두 가지 기본 작업을 수행합니다.

- 반복적 비례 적합 알고리즘을 사용하여 그룹화 변수 (X 로 Y 적합 분석의 X 변수) 와 선택한 매칭 변수를 모두 포함하는 가법 모형을 적합시킵니다. 이와 동등한 선형 모형은 매우 느리고 방대한 메모리 리소스를 필요로 하므로 참가자가 수백 명인 경우에는 반복적 비례 적합 알고리즘이 효과를 발휘합니다.
- 그룹 간에 매칭되는 점 사이에 선을 그립니다. 매칭 ID 값이 동일한 관측값이 여러 개이면 관측값 그룹의 평균에서부터 선이 그려집니다.

매칭 열 옵션은 매칭되는 점들을 연결하는 **매칭 선** 옵션을 자동으로 활성화합니다. 이 선을 해제하려면 **표시 옵션 > 매칭 선**을 선택합니다.

"매칭 적합" 보고서에는 *F*-검정을 사용한 효과가 표시됩니다. 이는 교호작용 항이 있는 모형과 없는 모형 두 개를 실행하는 경우 모형 적합 플랫폼에서 얻을 수 있는 검정과 동등합니다. 수준이 두 개뿐인 경우 *F*-검정은 쌍체 *t*-검정과 동등합니다.

참고 : 모형 적합 플랫폼에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

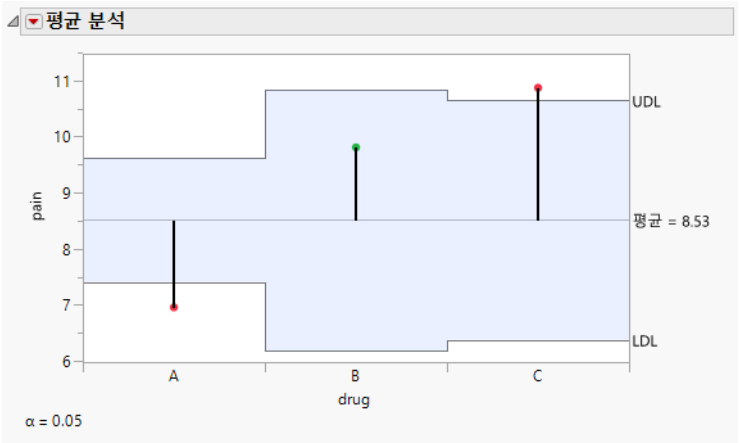
일원 분석 플랫폼의 추가 예

- " 평균 분석 차트의 예 "
- " 분산에 대한 평균 분석 차트의 예 "
- " 개별 쌍 비교 (스튜던트 t) 검정의 예 "
- " 전체 쌍 비교 (Tukey HSD) 검정의 예 "
- " 최량 비교 (Hsu MCB) 검정의 예 "
- " 대조군 비교 (Dunnett) 의 예 "
- " 단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예 "
- " 예 : 4 가지 평균 비교 검정 대조 "
- " 비모수 Wilcoxon 검정의 예 "
- " 이분산 옵션의 예 "
- " 동등성 검정의 예 "
- " 로버스트 적합 옵션의 예 "
- " 검정력 옵션의 예 "
- " 정규 분위수 그림의 예 "
- "CDF 그림의 예 "
- " 밀도 옵션의 예 "
- " 매칭 열 옵션의 예 "
- " 예 : 일원 분석을 위한 데이터 쌓기 "

평균 분석 차트의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Analgesics.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. pain 을 선택하고 **Y, 반응**을 클릭합니다 .
4. drug 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 평균 분석 " 의 빨간색 삼각형을 클릭하고 **평균 분석 방법 > ANOM** 을 선택합니다 .

그림 6.13 평균 분석 차트의 예



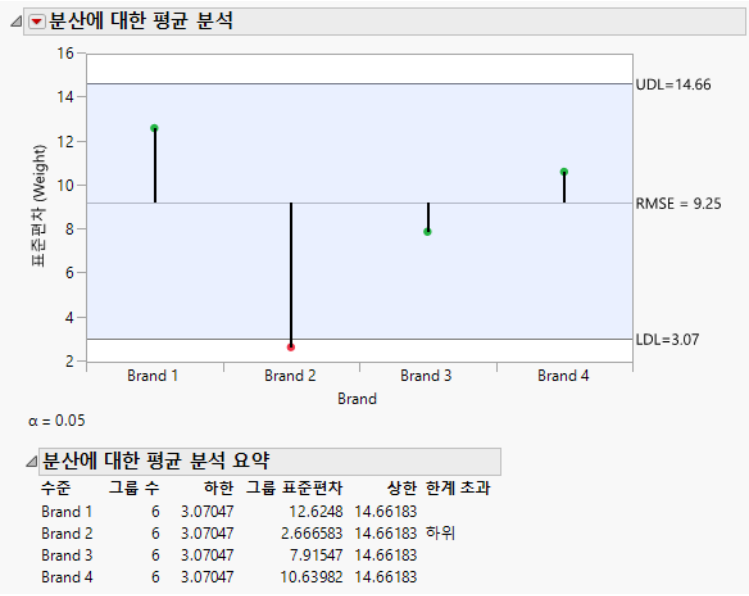
예를 들어 약물 A 와 C 의 평균은 전체 평균과 유의하게 다릅니다. 약물 A 의 평균은 더 낮고 약물 C 의 평균은 더 높습니다. 표본 크기가 다르기 때문에 약물 유형에 대한 결정 한계는 동일하지 않습니다.

분산에 대한 평균 분석 차트의 예

ANOMV 를 사용하여 스프링 브랜드 간에 변동성이 유의한 차이가 있는지 검정합니다 . 스프링 길이를 0.10 인치 늘리기 위해 필요한 중량을 확인하기 위해 4 가지 브랜드의 스프링을 테스트했습니다 . 각 브랜드에서 6 개의 스프링을 테스트했습니다 . 또한 ANOMV 검정은 비정규성에 대해 로버스트하지 않으므로 데이터의 정규성을 확인했습니다 . 이제 각 브랜드를 검토하여 브랜드 간에 변동성이 유의하게 다른지 여부를 확인하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Spring Data.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Brand 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Weight 대 Brand 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 분석 방법 > 분산에 대한 ANOM** 을 선택합니다 .
7. " 분산에 대한 평균 분석 " 옆의 빨간색 삼각형을 클릭하고 **요약 보고서 표시**를 선택합니다 .

그림 6.14 분산에 대한 평균 분석 차트의 예



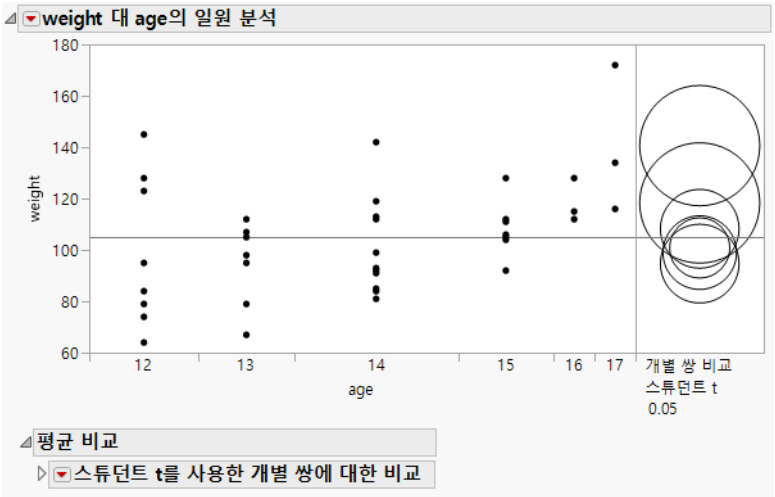
Brand 2의 표준편차는 결정 하한을 넘습니다. 따라서 Brand 2는 다른 브랜드보다 분산이 유의하게 낮습니다.

개별 쌍 비교 (스튜던트 t) 검정의 예

이 예에서는 모든 가능한 *t*- 검정의 사용법을 보여 줍니다 .

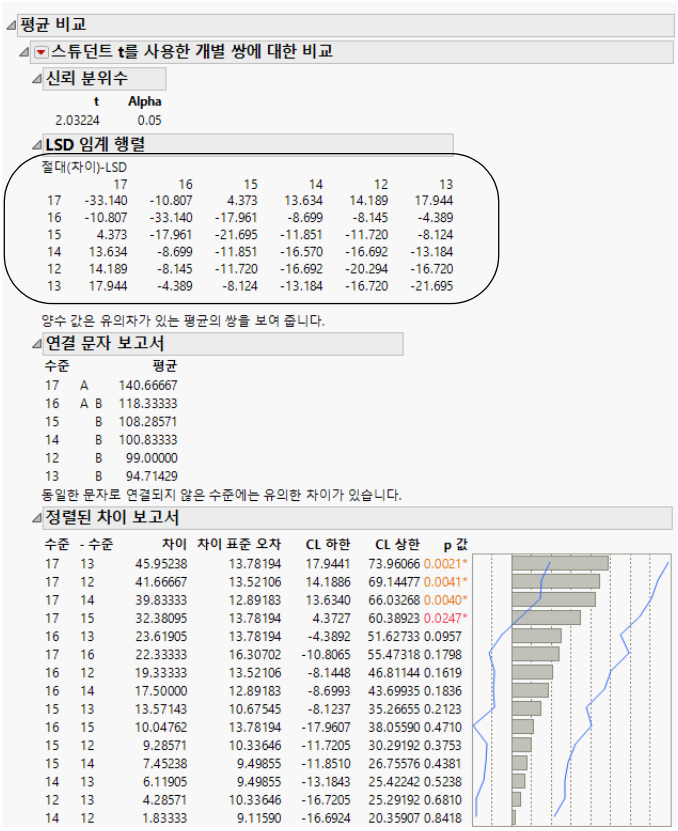
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. age 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "weight 대 age 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 개별 쌍 비교 (스튜던트 t)** 를 선택합니다 .

그림 6.15 개별 쌍 비교 (스튜던트 t) 비교 원의 예



이 평균 비교 방법은 평균의 실제 차이가 유의한 것으로 간주될 차이보다 큰지 확인하는 방법이라고 할 수 있습니다. 이 차이를 LSD(최소 유의차)라고 합니다. LSD라는 용어는 스튜던트 *t* 구간과 다른 검정의 구간 관련 문맥에서 사용됩니다. 비교 원 그래프에서 원 중심 간의 거리는 실제 차이를 나타냅니다. LSD는 원이 직각으로 교차한 경우의 거리입니다.

그림 6.16 개별 쌍 비교 (스튜던트 *t*) 에 대한 평균 비교 보고서의 예



평균 비교

스튜던트 t를 사용한 개별 쌍에 대한 비교

신뢰 분위수

t	Alpha
2.03224	0.05

LSD 임계 행렬

절대(차이)-LSD

	17	16	15	14	12	13
17	-33.140	-10.807	4.373	13.634	14.189	17.944
16	-10.807	-33.140	-17.961	-8.699	-8.145	-4.389
15	4.373	-17.961	-21.695	-11.851	-11.720	-8.124
14	13.634	-8.699	-11.851	-16.570	-16.692	-13.184
12	14.189	-8.145	-11.720	-16.692	-20.294	-16.720
13	17.944	-4.389	-8.124	-13.184	-16.720	-21.695

양수 값은 유의차가 있는 평균의 쌍을 보여 줍니다.

연결 문자 보고서

수준		평균
17	A	140.66667
16	A B	118.33333
15	B	108.28571
14	B	100.83333
12	B	99.00000
13	B	94.71429

동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다.

정렬된 차이 보고서

수준	- 수준	차이	차이 표준 오차	CL 하한	CL 상한	p 값
17	13	45.95238	13.78194	17.9441	73.96066	0.0021*
17	12	41.66667	13.52106	14.1886	69.14477	0.0041*
17	14	39.83333	12.89183	13.6340	66.03268	0.0040*
17	15	32.38095	13.78194	4.3727	60.38923	0.0247*
16	13	23.61905	13.78194	-4.3892	51.62733	0.0957
17	16	22.33333	16.30702	-10.8065	55.47318	0.1798
16	12	19.33333	13.52106	-8.1448	46.81144	0.1619
16	14	17.50000	12.89183	-8.6993	43.69935	0.1836
15	13	13.57143	10.67545	-8.1237	35.26655	0.2123
16	15	10.04762	13.78194	-17.9607	38.05590	0.4710
15	12	9.28571	10.33646	-11.7205	30.29192	0.3753
15	14	7.45238	9.49855	-11.8510	26.75576	0.4381
14	13	6.11905	9.49855	-13.1843	25.42242	0.5238
12	13	4.28571	10.33646	-16.7205	25.29192	0.6810
14	12	1.83333	9.11590	-16.6924	20.35907	0.8418

그림 6.16의 LSD 임계 테이블에서는 평균의 절대 차이와 LSD(최소 유의차) 간의 차이를 보여 줍니다. 값이 양수이면 두 평균의 차이가 LSD보다 크고 두 그룹이 유의하게 다른 것입니다.

전체 쌍 비교 (Tukey HSD) 검정의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. age 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "weight 대 age의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 전체 쌍 비교(Tukey HSD)** 를 선택합니다 .

그림 6.17 전체 쌍 비교 (Tukey HSD) 비교 원의 예

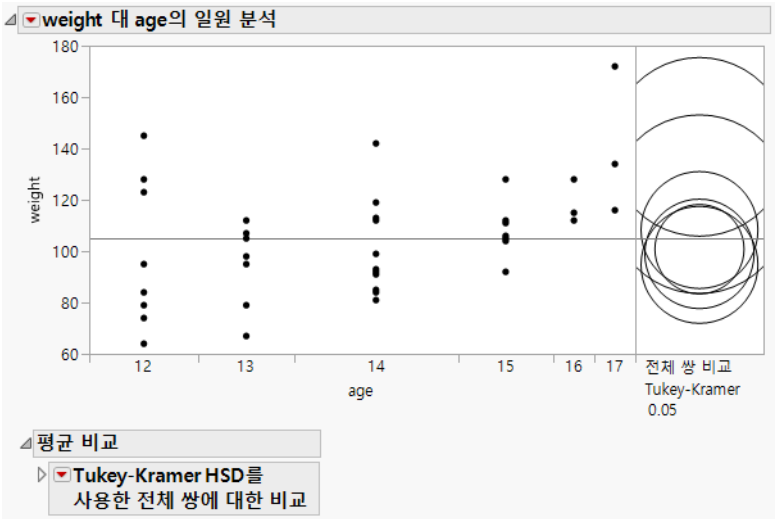


그림 6.18 전체 쌍 비교 (Tukey HSD) 에 대한 평균 비교 보고서의 예

평균 비교

Tukey-Kramer HSD를 사용한 전체 쌍에 대한 비교

신뢰 분위수

q*	Alpha
3.01825	0.05

HSD 임계 행렬

절대(차이)-HSD

	17	16	15	14	12	13
17	-49.219	-26.885	-9.216	0.923	0.857	4.355
16	-26.885	-49.219	-31.550	-21.411	-21.477	-17.978
15	-9.216	-31.550	-32.221	-21.217	-21.912	-18.650
14	0.923	-21.411	-21.217	-24.609	-25.681	-22.550
12	0.857	-21.477	-21.912	-25.681	-30.140	-26.912
13	4.355	-17.978	-18.650	-22.550	-26.912	-32.221

양수 값은 유의차가 있는 평균의 쌍을 보여 줍니다.

연결 문자 보고서

수준		평균
17	A	140.66667
16	A B	118.33333
15	A B	108.28571
14	B	100.83333
12	B	99.00000
13	B	94.71429

동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다.

정렬된 차이 보고서

수준	- 수준	차이	차이 표준 오차	CL 하한	CL 상한	p 값
17	13	45.95238	13.78194	4.3551	87.54968	0.0233*
17	12	41.66667	13.52106	0.8568	82.47657	0.0431*
17	14	39.83333	12.89183	0.9226	78.74404	0.0423*
17	15	32.38095	13.78194	-9.2163	73.97825	0.2029
16	13	23.61905	13.78194	-17.9783	65.21635	0.5325
17	16	22.33333	16.30702	-26.8853	71.55192	0.7443
16	12	19.33333	13.52106	-21.4766	60.14323	0.7091
16	14	17.50000	12.89183	-21.4107	56.41071	0.7512
15	13	13.57143	10.67545	-18.6497	45.79256	0.7982
16	15	10.04762	13.78194	-31.5497	51.64492	0.9769
15	12	9.28571	10.33646	-21.9123	40.48369	0.9442
15	14	7.45238	9.49855	-21.2166	36.12134	0.9683
14	13	6.11905	9.49855	-22.5499	34.78801	0.9867
12	13	4.28571	10.33646	-26.9123	35.48369	0.9983
14	12	1.83333	9.11590	-25.6807	29.34736	1.0000

그림 6.18의 Tukey-Kramer HSD 임계 행렬에서는 평균의 실제 절대 차이에서 HSD를 뺀 값을 보여 줍니다. 이 값은 유의하게 간주될 차이를 나타냅니다. 양수 값을 가진 쌍은 유의하게 다른 것입니다. q*(HSD 임계 행렬 테이블 위에 표시)는 HSD를 척도화하는 데 사용된 분위수입니다. 이 값은 스튜던트 *t*와 유사한 계산 역할을 합니다.

최량 비교 (Hsu MCB) 검정의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. age 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "weight 대 age 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 최량 비교 (Hsu MCB)** 를 선택합니다 .

그림 6.19 최량 비교 (Hsu MCB) 비교 원의 예

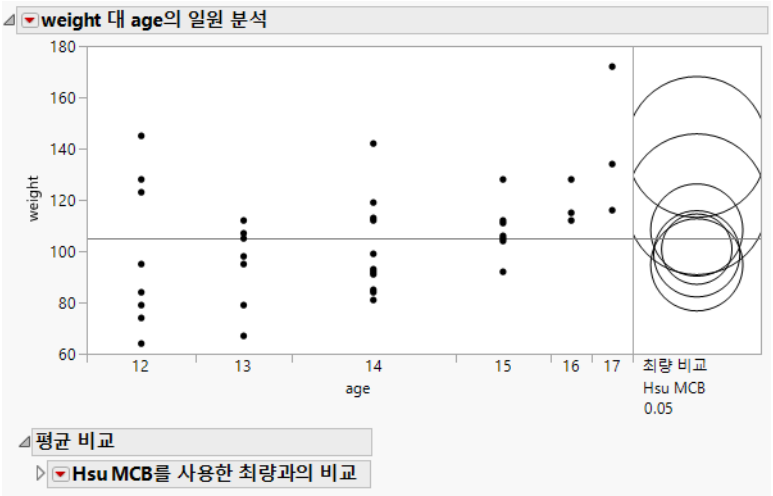


그림 6.20 최량 비교 (Hsu MCB) 에 대한 평균 비교 보고서의 예

Hsu MCB를 사용한 최량과의 비교

신뢰 분위수

d	Alpha
2.23157	0.05
2.23157	
2.34056	
2.38154	
2.35270	
2.34056	

최대값 및 최소값을 사용한 비교

수준	최대 p 값 사용	최소 p 값 사용
17	0.9863	0.0039*
16	0.2219	0.1304
15	0.0491*	0.3172
14	0.0093*	0.6599
12	0.0091*	0.7216
13	0.0046*	0.9408

LSD 임계 행렬

평균[i]-평균[j]-LSD

	17	16	15	14	12	13
17	-36.390	-14.057	0.124	9.131	9.856	13.695
16	-58.724	-36.390	-22.210	-13.202	-12.478	-8.638
15	-63.136	-40.803	-24.987	-15.169	-15.033	-11.415
14	-68.602	-46.269	-29.684	-19.418	-19.614	-16.113
12	-71.840	-49.507	-33.479	-23.543	-23.494	-19.907
13	-76.708	-54.374	-38.558	-28.740	-28.604	-24.987

열에 양수 값이 있는 경우 평균이 최대값보다 상당히 작습니다.

평균[i]-평균[j]+LSD

	17	16	15	14	12	13
17	36.390	58.724	64.638	70.536	73.478	78.210
16	14.057	36.390	42.305	48.202	51.144	55.876
15	-1.626	20.708	24.987	30.074	33.604	38.558
14	-11.064	11.269	14.780	19.418	23.280	28.351
12	-11.493	10.840	14.907	19.877	23.494	28.479
13	-15.197	7.136	11.415	16.502	20.033	24.987

열에 음수 값이 있는 경우 평균이 최소값보다 상당히 큽니다.

"최대값 및 최소값을 사용한 비교" 보고서에서는 각 수준의 평균을 나머지 수준의 최대 및 최소 평균과 비교합니다. 예를 들어 15세의 평균은 나머지 수준의 최대 평균과 유의하게 다릅니다. 17세의 평균은 나머지 수준의 최소 평균과 유의하게 다릅니다. 최대 평균은 16세 또는 17세에서 나타날 수 있습니다. 두 평균 모두 최대 평균과 유의하게 다르지 않기 때문입니다. 동일한 이유로 최소 평균은 17세를 제외한 다른 연령에 해당할 수 있습니다.

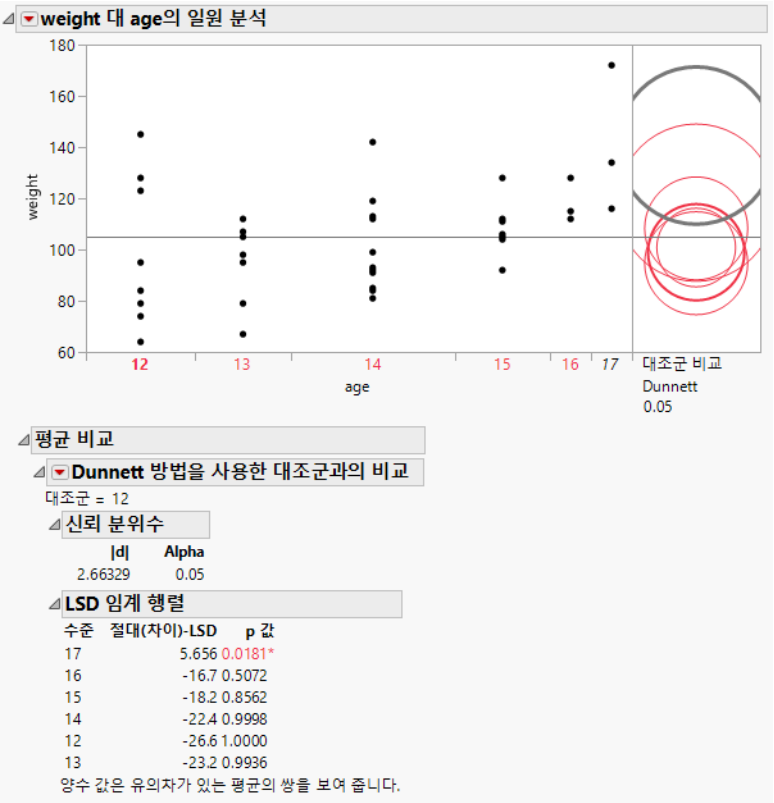
대조군 비교 (Dunnett) 의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. age 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

6. "weight 대 age 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 대조군 비교 (Dunnett)** 를 선택합니다 .
7. 대조군으로 사용할 그룹을 선택합니다 . 이 예에서는 12 세를 선택합니다 .

   또는 산점도에서 행을 클릭하여 강조 표시한 후 **평균 비교 > 대조군 비교 (Dunnett)** 옵션을 선택합니다 . 이 검정에서는 선택된 행이 대조군으로 사용됩니다 .
8. **확인**을 클릭합니다 .

그림 6.21 대조군 비교 (Dunnett) 비교 원의 예



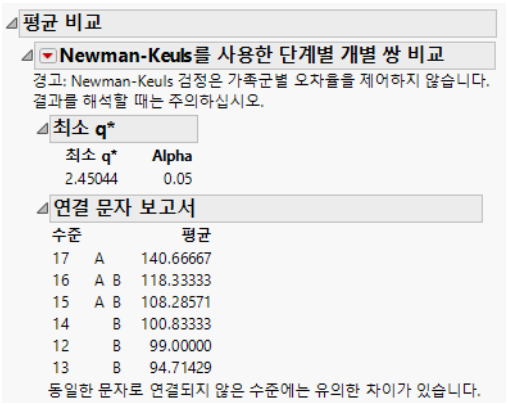
이 비교 원을 사용하면 수준 "17" 이 대조 수준 "12" 와 유의하게 다른 유일한 수준이라고 결론 내릴 수 있습니다.

단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .

4. age 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "weight 대 age 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 단계별 개별 쌍 비교 (Newman-Keuls)** 를 선택합니다 .

그림 6.22 단계별 개별 쌍 비교 (Newman-Keuls) 에 대한 평균 비교 보고서의 예



"연결 문자 보고서"에는 수준 "17"이 "16" 및 "15"를 제외한 다른 모든 수준과 유의하게 다른 것으로 나타납니다.

예 : 4 가지 평균 비교 검정 대조

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. age 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "weight 대 age 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교**의 각 옵션을 하나씩 선택합니다 . " 대조군 비교 (Dunnett)" 옵션의 경우 대조군으로 17 세를 선택합니다 .

4 가지 방법은 모두 그룹 평균 간의 차이를 검정합니다. 각 검정은 특정 가설에 사용되며 서로 다른 사실을 확인할 수 있습니다.

그림 6.23 4 가지 다중 비교 검정의 비교 원

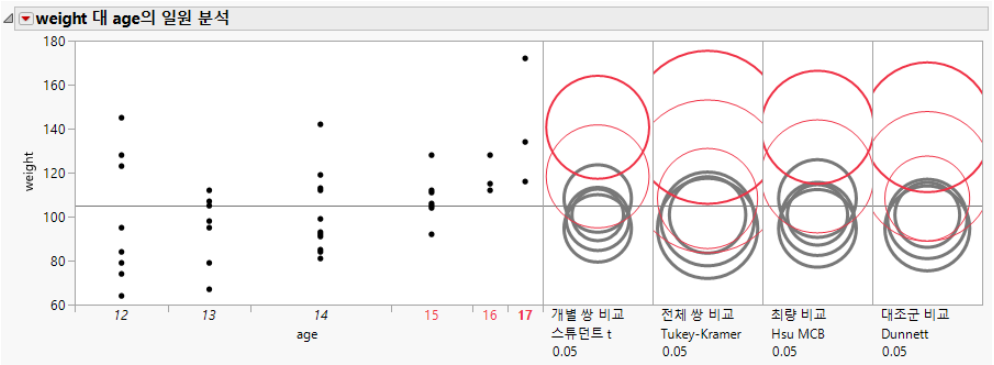


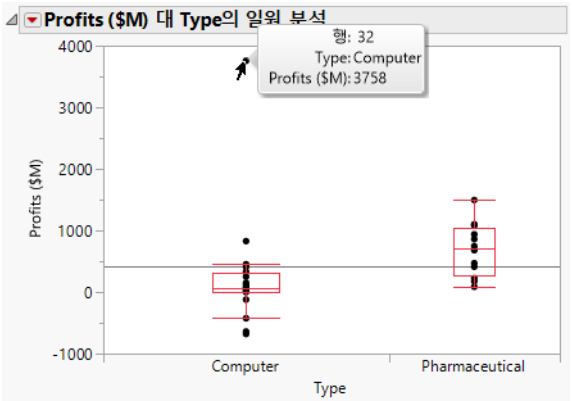
그림 6.23 에서는 17 세 그룹이 강조 표시되어 있습니다. 다른 대조군 원은 17 세 그룹과의 관계에 따라 색상이 적용되어 있습니다. 스튜던트 t 및 Hsu 방법의 경우에는 15 세 그룹 (위에서 세 번째 원) 이 회색입니다. 이는 이 그룹이 17 세 그룹과 유의하게 다름을 나타냅니다. 하지만 Tukey 및 Dunnett 방법의 경우에는 15 세 그룹이 빨간색으로, 17 세 그룹과 유의하게 다르지 않음을 나타냅니다.

비모수 Wilcoxon 검정의 예

Wilcoxon 검정을 사용하여 회사의 평균 수익이 회사 유형에 따라 다른지 여부를 확인합니다 . Companies.jmp 의 데이터는 제약 회사 (12 개) 와 컴퓨터 회사 (20 개) 라는 두 가지 회사 유형에 대한 다양한 측정 기준으로 구성되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Profits ($M) 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Type 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Profits ($M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림**을 선택합니다 .

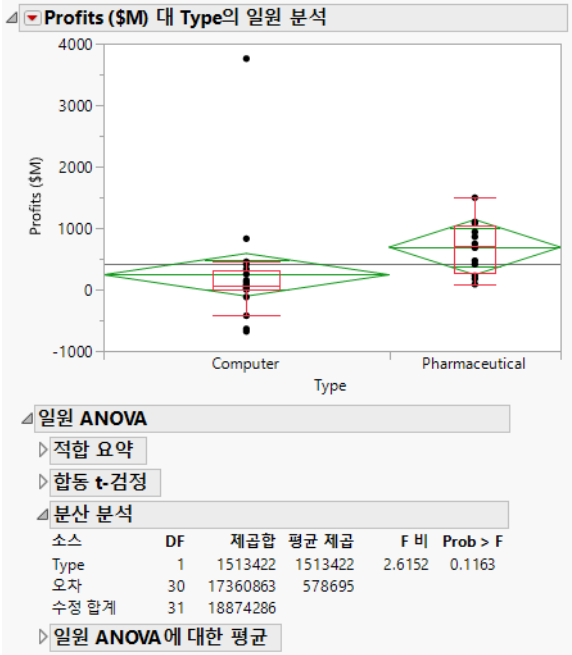
그림 6.24 컴퓨터 회사의 수익 분포



이 상자 그림에서는 데이터가 정규 분포 또는 대칭 분포가 아닌 것으로 나타납니다. 32 행 회사의 경우에는 모수 검정에 영향을 줄 수 있는 매우 큰 값이 있습니다.

7. "Profits ($M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA/ 합동 t** 를 선택합니다 .

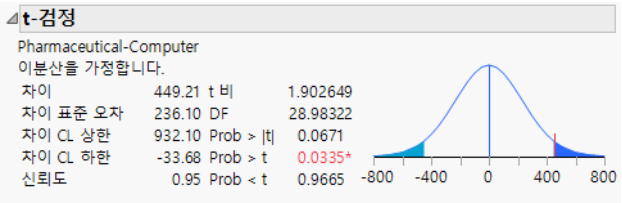
그림 6.25 회사 분산 분석



F- 검정에서는 *p* 값이 크기 때문에 ($p = 0.1163$) 유의성이 없음을 보여 줍니다 . 이는 32 행의 값이 커서 정규성 가정이 위배되었기 때문일 수 있습니다 .

8. "Profits ($M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **t- 검정**을 선택합니다 .

그림 6.26 t- 검정 결과



양측 검정의 "Prob > |t|" 는 0.0671 입니다 . *t*- 검정에서는 등분산을 가정하지 않지만 이분산 *t*- 검정도 모수 검정입니다 .

9. "Profits ($M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **비모수 > Wilcoxon 검정**을 선택합니다 .

그림 6.27 Wilcoxon 검정 결과

Wilcoxon 검정은 비모수 검정입니다 . 이 검정은 순위를 기반으로 하므로 이상치의 영향을 받지 않습니다 . 또한 정규성이 필요하지 않습니다 .

Wilcoxon 검정 통계량에 대한 정규 근사와 카이제곱 근사는 모두 *p* 값이 0.0010 일 때 유의성을 나타냅니다 . 따라서 분포의 이 위치에서 유의한 차이가 있으며 회사 유형에 따라 평균 수익에 차이가 있다는 결론을 내릴 수 있습니다 .

정규 검정과 카이제곱 검정은 검정 통계량의 점근적 분포를 기반으로 합니다 . JMP Pro 를 사용한다면 정확 검정을 수행할 수 있습니다 .

10. JMP PRO "Profits ($M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **비모수 > 정확 검정 > Wilcoxon 정확 검정**을 선택합니다 .

그림 6.28 Wilcoxon 정확 검정 결과

Wilcoxon/Kruskal-Wallis 검정(순위 합)

수준	개수	점수 합	기대 점수	점수 평균	(평균-평균0)/표준0
Computer	20	245.000	330.000	12.2500	-3.289
Pharmaceutical	12	283.000	198.000	23.5833	3.289

2표본 검정, 정규 근사

S	Z	Prob>|Z|
283	3.28916	0.0010*

일원 검정, 카이제곱 근사

카이제곱	DF	Prob>ChiSq
10.9470	1	0.0009*

2표본: 정확 검정

S	Prob≥S	Prob≥|S-Mean|
283	0.0003*	0.0005*

관측된 검정 통계량 값은 S = 283 입니다 . 이는 표본 크기가 더 작은 Type 수준 (Pharmaceutical) 의 순위 합입니다 . 평균 중간 순위와의 절대 차이가 S 의 절대값에서 중간 순위의 평균을 뺀 값을 초과하는 것으로 관측될 확률은 0.0005 입니다 . 이는 위치의 차이에 대한 양측 검정으로 , 회사 유형별로 수익이 다르지 않다는 가설을 기각할 것을 지지합니다 .

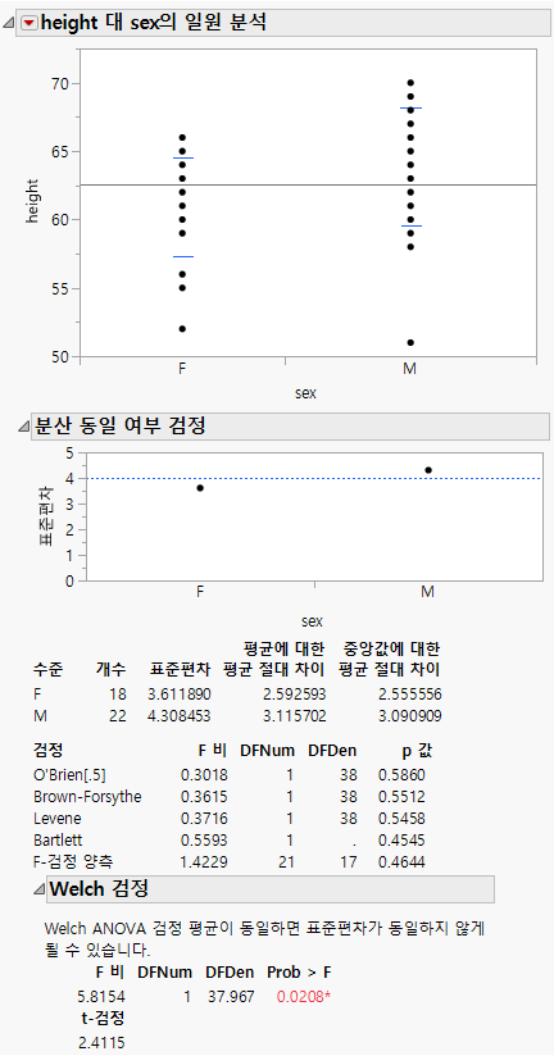
이 예에서는 비모수 검정이 정규성 기반 분산 분석 ANOVA 검정과 이분산 *t*- 검정보다 더 적절합니다 . 비모수 검정은 32 행에 있는 큰 값의 영향을 받지 않으며 정규성 가정이 필요하지 않습니다 .

이분산 옵션의 예

이 예에서는 두 그룹 (남성과 여성) 이 등분산을 갖는지 여부를 검정하려고 한다고 가정해 보겠습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. height 를 선택하고 **Y, 반응**을 클릭합니다 .
4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "height 대 sex 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **이분산**을 선택합니다 .

그림 6.29 이분산 보고서의 예



수준	개수	표준편차	평균에 대한 평균 절대 차이	중앙값에 대한 평균 절대 차이
F	18	3.611890	2.592593	2.555556
M	22	4.308453	3.115702	3.090909

검정	F 비	DFNum	DFDen	p 값
O'Brien[.5]	0.3018	1	38	0.5860
Brown-Forsythe	0.3615	1	38	0.5512
Levene	0.3716	1	38	0.5458
Bartlett	0.5593	1	.	0.4545
F-검정 양측	1.4229	21	17	0.4644

F 비	DFNum	DFDen	Prob > F
5.8154	1	37.967	0.0208*

t-검정
2.4115

양측 *F*-검정의 *p* 값이 크기 때문에 분산이 동일하지 않다는 결론을 내릴 수 없습니다.

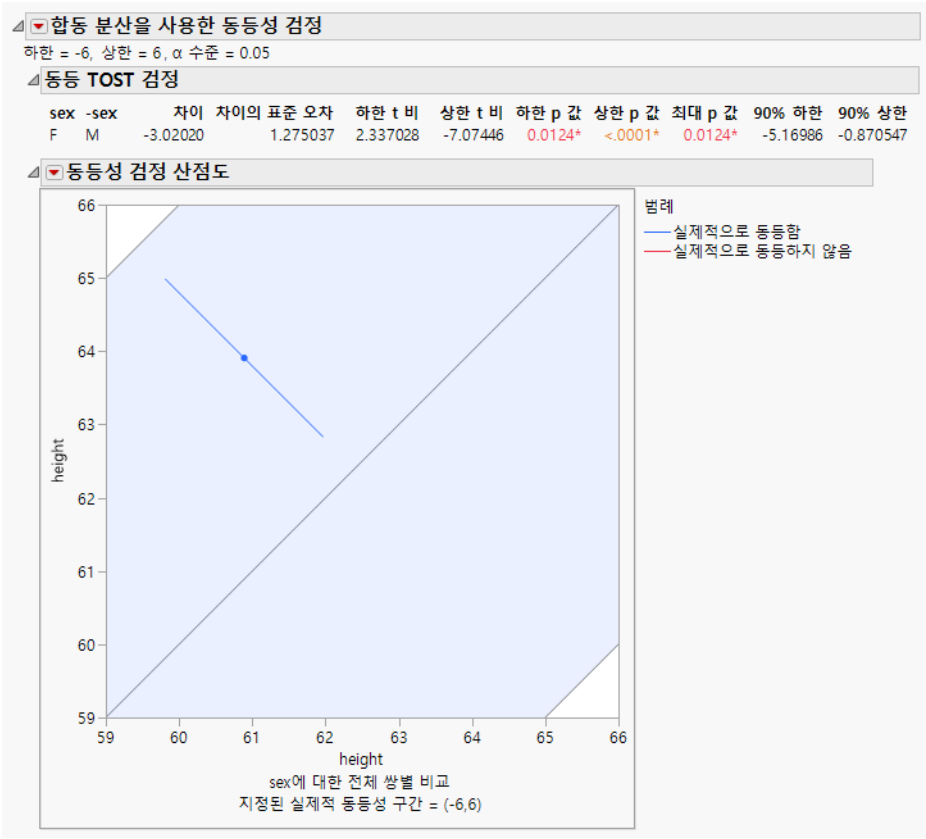
동등성 검정의 예

이 예에서는 남성과 여성 사이의 키 차이가 6 인치 미만인지 여부를 조사하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. height 를 선택하고 **Y, 반응**을 클릭합니다 .

4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "height 대 sex 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **동등성 검정**을 선택합니다 .
7. " 실제적으로 0 으로 간주되는 차이 " 로 6 을 입력합니다 .
8. **확인**을 클릭합니다 .

그림 6.30 동등성 검정의 예



두 개의 단측 검정을 사용하면 둘 모두에서 *p* 값이 작다는 것을 확인할 수 있습니다. 따라서 모집단 평균의 차이가 -6에서 6 사이의 어딘가에 있다고 결론 내릴 수 있습니다. 이 예에서는 평균이 실제적으로 동등하다고 선언할 수 있습니다.

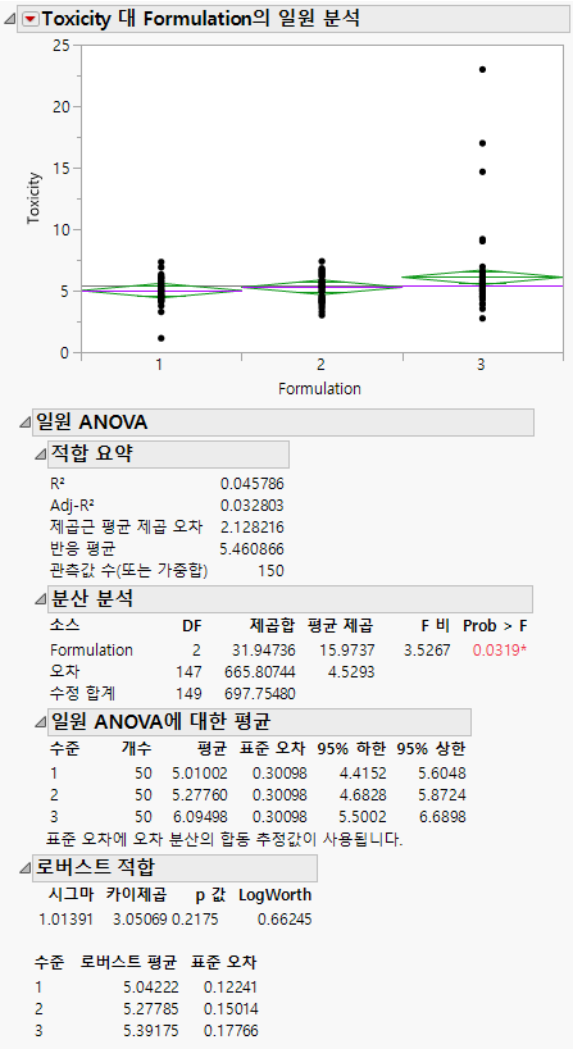
로버스트 적합 옵션의 예

세 그룹 중 하나에 이상치가 포함된 이 예에서는 로버스트 적합을 사용합니다 . Drug Toxicity.jmp 샘플 데이터 테이블의 데이터는 세 가지 약물 제제의 독성 수준을 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug Toxicity.jmp 를 엽니다 .

2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Toxicity 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Formulation 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Toxicity 대 Formulation 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .
7. "Toxicity 대 Formulation 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **로버스트 > 로버스트 적합**을 선택합니다 .

그림 6.31 로버스트 적합의 예

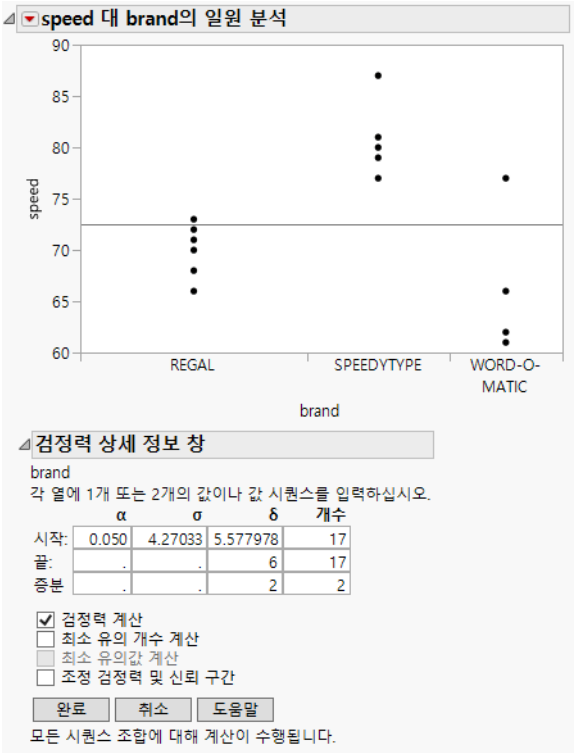


표준 "분산 분석" 보고서를 보면 *p* 값이 0.0319이므로 세 가지 제제 사이에 차이가 있다는 잘못된 결론을 내리게 될 수 있습니다. 하지만 "로버스트 적합" 보고서를 보면 여기서는 *p* 값이 0.21755이므로 세 가지 제제가 유의하게 다르다는 결론을 내리지 않을 것입니다. 일부 관측값의 독성은 비정상적으로 높아서 데이터에 과도한 영향을 주는 것으로 보입니다.

검정력 옵션의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Typing Data.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. speed 를 선택하고 **Y, 반응**을 클릭합니다 .
4. brand 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "speed 대 brand 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **검정력**을 선택합니다 .
7. " 시작 " 행에서 델타 (세 번째 상자) 값으로 2 를 입력하고 " 개수 " 에 11 을 입력합니다 .
8. " 끝 " 행에서 델타 값으로 6 을 입력하고 " 개수 " 상자에 17 을 입력합니다 .
9. " 증분 " 행에서 델타 값과 " 개수 " 모두에 2 를 입력합니다 .
10. **검정력 계산** 체크박스를 선택합니다 .

그림 6.32 검정력 상세 정보 창의 예



11. **완료**를 클릭합니다 .

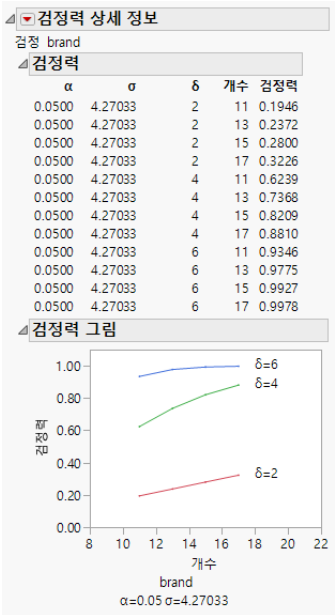
참고 : 필요한 옵션이 모두 적용되기 전까지는 **완료** 버튼이 흐리게 표시됩니다 .

델타 값과 개수의 각 조합에 대해 검정력이 계산되어 "검정력" 보고서에 표시됩니다.

검정력 값을 그림으로 표시하려면 다음을 수행하십시오.

12. " 검정력 상세 정보 " 의 빨간색 삼각형을 클릭하고 **검정력 그림**을 선택합니다 .

그림 6.33 검정력 보고서의 예



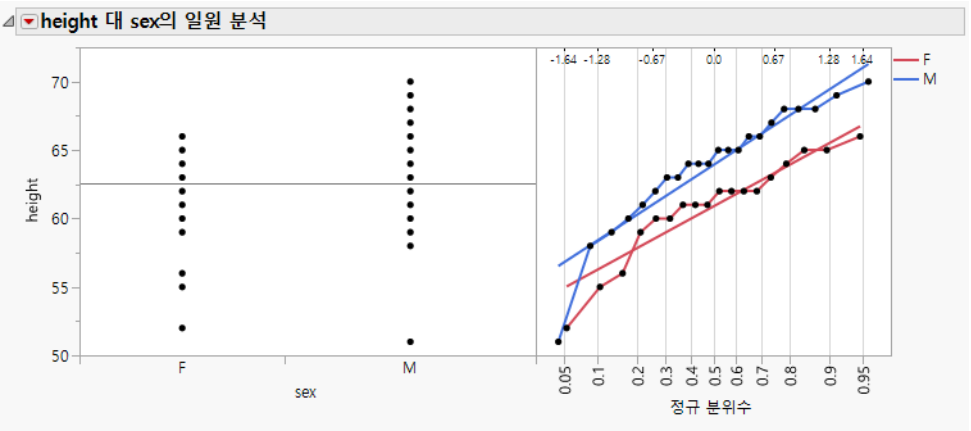
13. 그림의 데이터를 모두 표시하려면 " 검정력 " 축을 클릭한 채로 세로로 드래그해야 할 수 있습니다 .

델타 값과 개수의 각 조합에 대해 검정력이 표시됩니다. 예상한 대로 " 개수 "(표본 크기) 값과 델타 값(평균 차이)이 클수록 검정력이 높아집니다.

정규 분위수 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. height 를 선택하고 **Y, 반응**을 클릭합니다 .
4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "height 대 sex 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **정규 분위수 그림 > 분위수별 실제값 그림**을 선택합니다 .

그림 6.34 정규 분위수 그림의 예



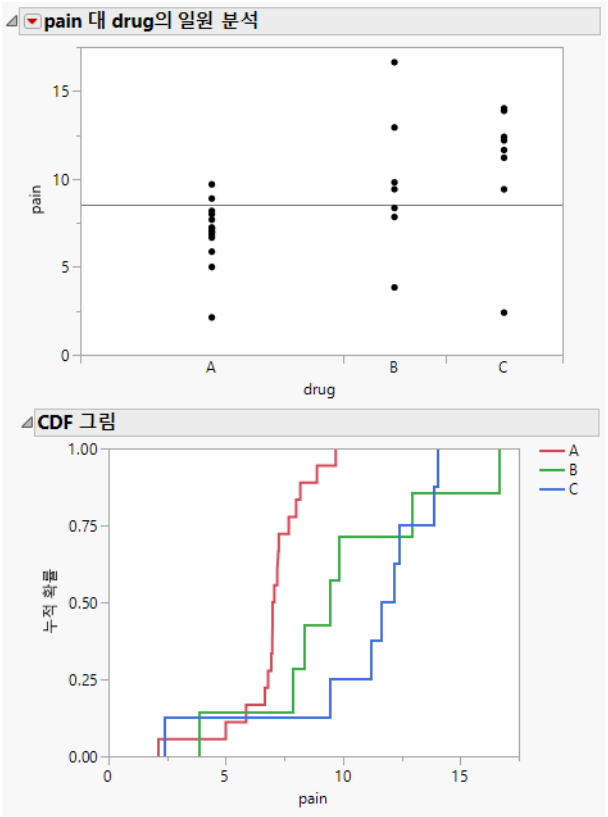
다음 사항에 유의하십시오 .

- 기본적으로 적합선이 표시됩니다 .
- 데이터 점은 적합선에 매우 가깝게 늘어서 있어 정규 분포를 나타냅니다 .

CDF 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Analgesics.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. pain 을 선택하고 **Y, 반응**을 클릭합니다 .
4. drug 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "pain 대 drug 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **CDF 그림**을 선택합니다 .

그림 6.35 CDF 그림의 예

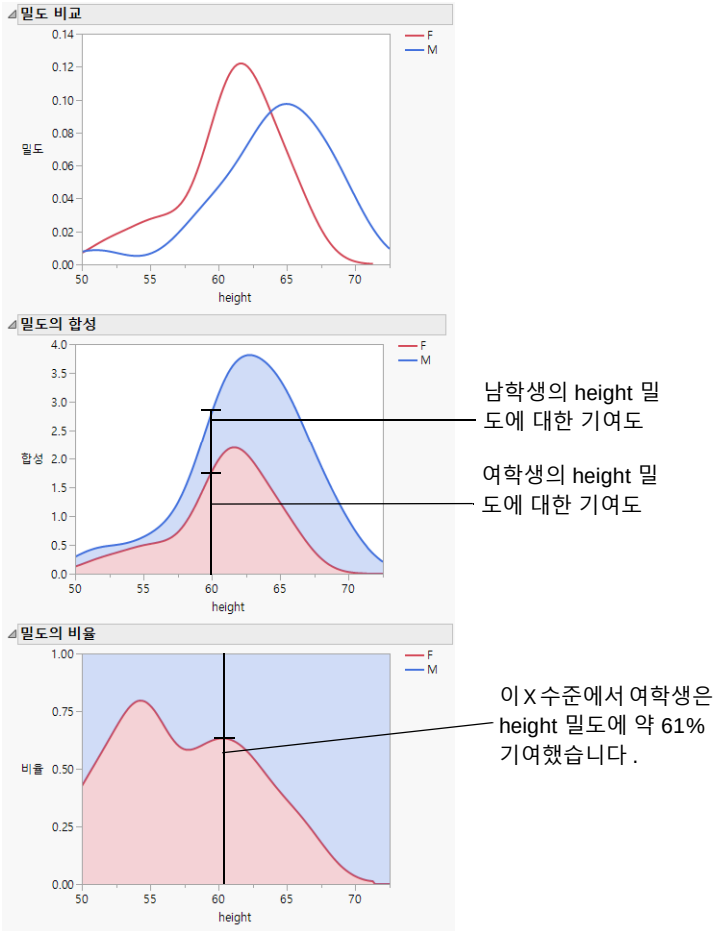


초기 일원 분석의 X 변수 수준은 CDF 그림에 서로 다른 곡선으로 표시됩니다. CDF 그림의 가로 축에는 초기 일원 분석의 y 값이 사용됩니다.

밀도 옵션의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. height 를 선택하고 **Y, 반응**을 클릭합니다 .
4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "height 대 sex 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 세 개의 옵션 , 즉 **밀도 > 밀도 비교** , **밀도 > 밀도의 합성** 및 **밀도 > 밀도의 비율**을 모두 선택합니다 .

그림 6.36 밀도 옵션의 예



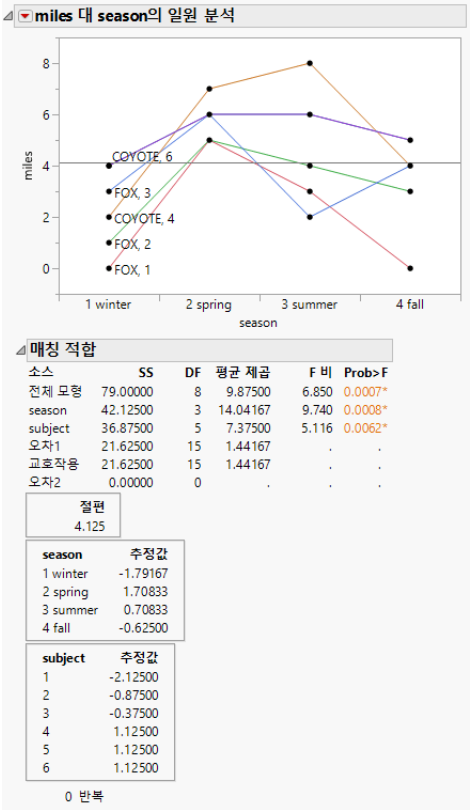
매칭 열 옵션의 예

이 예에서는 6 종의 동물과 이 동물들이 계절별로 이동한 거리 (마일) 가 포함되어 있는 Matching.jmp 샘플 데이터 테이블을 사용합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Matching.jmp 를 엽니다 .
2. **분석** > **X 로 Y 적합**을 선택합니다 .
3. miles 를 선택하고 **Y, 반응**을 클릭합니다 .
4. season 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "miles 대 season 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **매칭 열**을 선택합니다 .
7. 매칭 열로 subject 를 선택합니다 .

8. **확인**을 클릭합니다 .

그림 6.37 매칭 열 보고서의 예



이 그림에서는 개체를 매칭 변수로 하여 계절별 이동 거리(마일)를 그래프로 보여 줍니다. 그래프에서 각 개체의 첫 번째 측정값 옆에 표시되는 라벨은 species 및 subject 변수에 따라 결정되었습니다.

"매칭 적합" 보고서에는 *F*-검정을 사용한 season 및 subject 효과가 표시됩니다. 이는 교호작용 항이 있는 모형과 없는 모형 두 개를 실행하는 경우 모형 적합 플랫폼에서 얻을 수 있는 검정과 동등합니다. 수준이 두 개뿐인 경우 *F*-검정은 쌍체 *t*-검정과 동등합니다.

참고 : 모형 적합 플랫폼에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

예 : 일원 분석을 위한 데이터 쌓기

데이터가 JMP 데이터 테이블이 아닌 다른 형식으로 존재하는 경우 데이터 배열이 한 행에 여러 관측값에 대한 정보가 포함되는 방식일 수도 있습니다. JMP에서 데이터를 분석하려면 데이터를 가져온 후 JMP 데이터 테이블의 각 행에 단일 관측값에 대한 정보가 포함되도록 데이터를 재구

성해야 합니다. 예를 들어 데이터가 스프레드시트에 들어 있다고 가정해 보겠습니다. 세 개의 생산 라인에서 생산된 제품에 대한 데이터가 세 개의 열 집합에 배열되어 있습니다. JMP 데이터 테이블에서는 세 개 생산라인에서 얻은 데이터를 단일 열 집합에 쌓아 각 행이 단일 제품의 데이터를 나타내도록 해야 합니다.

설명 및 목표

이 예에서는 서로 다른 세 개의 생산 라인에서 무작위로 표집된 시리얼 상자의 중량이 포함된 Fill Weights.xlsx 파일을 사용합니다 . 그림 6.38 에서는 이 데이터의 형식을 보여 줍니다 .

- "ID" 열에는 측정된 각 시리얼 상자의 식별자가 포함되어 있습니다 .
- "Line" 열에는 해당 생산 라인에서 표집된 상자의 중량 (온스) 이 포함되어 있습니다 .

그림 6.38 데이터 형식

Weights					
ID	Line A	ID	Line B	ID	Line C
215	12.42	705	13.63	254	11.73
287	12.49	670	12.56	282	11.40
381	12.80	715	12.87	938	12.78
		683	13.09	597	12.19
		514	13.31	179	12.25
		517	12.64		
		946	12.75		

상자의 목표 충진 중량은 12.5 온스입니다. 세 개의 생산 라인이 목표를 충족하고 있는지 여부에 관심이 있지만 우선은 세 개의 라인이 동일한 평균 충진 비율을 달성하고 있는지 확인하려고 합니다. 일원 분석을 사용하여 평균 충진 중량 간에 차이가 있는지 검정할 수 있습니다.

일원 분석 플랫폼을 사용하려면 다음을 수행해야 합니다 .

1. 데이터를 JMP 로 가져옵니다 . 자세한 내용은 " 데이터 가져오기 "(218 페이지) 에서 확인하십시오 .
2. 데이터를 재구성하여 JMP 데이터 테이블에서 각 행이 단일 관측값만 반영하도록 합니다. 데이터를 재구성하려면 시리얼 상자 ID, 라인 식별자 및 중량을 해당 열에 쌓아야 합니다 . 자세한 내용은 " 데이터 쌓기 "(220 페이지) 에서 확인하십시오 .

데이터 가져오기

이 예에서는 Microsoft Excel 의 데이터를 JMP 로 가져오는 두 가지 방법을 보여 줍니다 . 한 가지 방법을 선택하거나 두 방법을 모두 사용해 보십시오 .

- **파일 > 열기** 옵션을 사용하여 Excel 가져오기 마법사로 Microsoft Excel 파일의 데이터를 가져옵니다 . 자세한 내용은 "Excel 가져오기 마법사를 사용하여 데이터 가져오기 "(219 페이지) 에서 확인하십시오 . 이 방법은 모든 Excel 파일에 편리한 방법입니다 .

- Microsoft Excel 의 데이터를 새 JMP 데이터 테이블에 복사하여 붙여넣습니다 . 자세한 내용은 "Excel 에서 데이터 복사 및 붙여넣기 "(219 페이지) 에서 확인하십시오 . 데이터 파일이 작은 경우에 이 방법을 사용할 수 있습니다 .

Microsoft Excel에서 데이터를 가져오는 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Excel 가져오기 마법사를 사용하여 데이터 가져오기

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Samples/Import Data 폴더에 있는 Fill Weights.xlsx 를 엽니다 .

   파일이 Excel 가져오기 마법사에서 열립니다 .
2. **열 머리글이 시작되는 행** 옆에 3 을 입력합니다 .

   Excel 파일의 1 행에는 테이블에 대한 정보가 포함되어 있고 2 행은 비어 있습니다 . 열 머리글 정보는 3 행에서 시작됩니다 .
3. **열 머리글이 포함된 행 수**에 2 를 입력합니다 .

   Excel 파일의 3 행과 4 행에는 열 머리글 정보가 포함되어 있습니다 .
4. **가져오기**를 클릭합니다 .

그림 6.39 Excel 가져오기 마법사를 사용하여 생성된 JMP 테이블

	Weights-ID	Weights-Line A	Weights-ID 2	Weights-Line B	Weights-ID 3	Weights-Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.40
3	381	12.80	715	12.87	938	12.78
4	•	•	683	13.09	597	12.19
5	•	•	514	13.31	179	12.25
6	•	•	517	12.64	•	•
7	•	•	946	12.75	•	•

   7 개의 행에 데이터가 있고 각 행마다 여러 개의 ID 가 나타납니다 . 세 개 라인 각각에 대해 ID 및 중량 열이 있으며 총 열 수는 6 개입니다 .

   ID 열 이름의 "Weights" 부분은 불필요한 데다 혼동을 일으킬 수 있습니다 . 지금 열 이름을 바꿀 수도 있지만 데이터를 쌓은 후 열 이름을 바꾸는 것이 더 효율적입니다 .
5. " 데이터 쌓기 "(220 페이지) 로 진행합니다 .

Excel 에서 데이터 복사 및 붙여넣기

1. Microsoft Excel 에서 Fill Weights.xlsx 를 엽니다 .
2. 테이블 내의 데이터를 선택하되 불필요한 "Weights" 머리글은 제외합니다 .
3. 마우스 오른쪽 버튼을 클릭하고 **복사**를 선택합니다 .
4. JMP 에서 **파일 > 새로 만들기 > 데이터 테이블**을 선택합니다 .

5. **편집 > 열 이름과 함께 붙여넣기**를 선택합니다 .

   편집 > 열 이름과 함께 붙여넣기 옵션은 클립보드의 선택 내용에 열 이름이 포함된 경우에 사용됩니다 .

그림 6.40 " 열 이름과 함께 붙여넣기 " 를 사용하여 생성된 JMP 테이블

	ID	Line A	ID 2	Line B	ID 3	Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.40
3	381	12.80	715	12.87	938	12.78
4	•	•	683	13.09	597	12.19
5	•	•	514	13.31	179	12.25
6	•	•	517	12.64	•	•
7	•	•	946	12.75	•	•

6. " 데이터 쌓기 "(220 페이지) 으로 진행합니다 .

데이터 쌓기

" 쌓기 " 옵션을 사용하여 새 데이터 테이블의 각 행에 하나의 관측값이 포함되도록 할 수 있습니다 . " 쌓기 " 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

1. JMP 데이터 테이블에서 **테이블 > 쌓기**를 선택합니다 .
2. 6 개의 열을 모두 선택하고 **열 쌓기**를 클릭합니다 .
3. **다중 계열 쌓기**를 선택합니다 .

   두 개의 계열 , 즉 ID 와 라인을 쌓을 것이므로 " 계열 수 " 는 변경하지 않고 기본값 2 로 설정된 상태로 둡니다 . 계열이 포함된 열은 인접해 있지 않고 한 열씩 번갈아 표시되어 있습니다 ("ID", "Line A", "ID", "Line B", "ID", "Line C"). 따라서 " 인접 " 은 선택하지 않습니다 .
4. **행별로 쌓기**를 선택 취소합니다 .
5. **결측 행 제거**를 선택합니다 .
6. **출력 테이블 이름** 옆에 Stacked 를 입력합니다 .
7. **확인**을 클릭합니다 .

   새 데이터 테이블에서 데이터 및 데이터 2 는 ID 및 중량 데이터가 포함된 열입니다 .
8. 라벨 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **열 삭제**를 선택합니다 .

   라벨 열 내의 항목은 가져온 데이터 테이블에 있는 상자 ID 열의 머리글입니다 . 이러한 항목은 필요하지 않습니다 .
9. 각 열의 열 머리글을 두 번 클릭하여 이름을 바꿉니다 .
   - 데이터 -> ID
   - 라벨 2 -> Line
   - 데이터 2 -> Weight

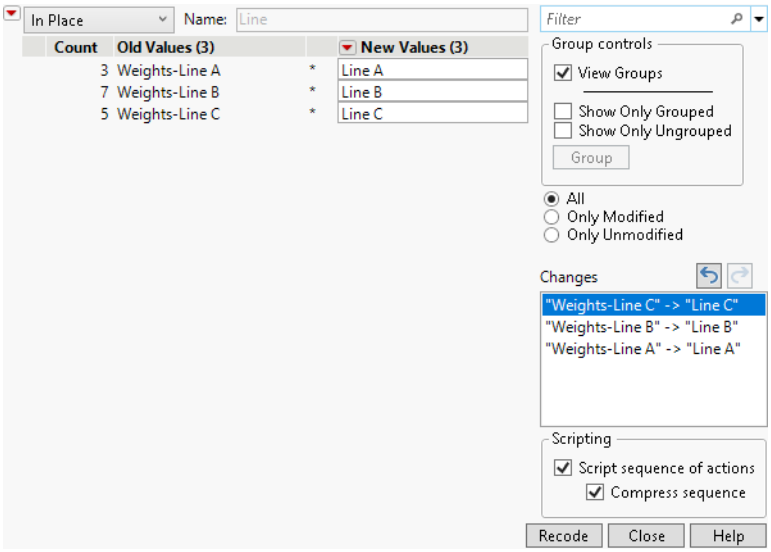
10. " 열 " 패널에서 ID 왼쪽의 아이콘을 클릭하고 **명목형**을 선택합니다 .

    ID 는 숫자로 지정되기는 했지만 식별자이므로 모델링 유형을 명목형으로 처리해야 합니다 . 이 예에서는 이 단계가 중요하지 않지만 항상 열에 적절한 모델링 유형을 할당하는 것이 좋습니다 .

11. (**파일 > 열기**를 사용하여 Excel 에서 데이터를 가져온 경우에만 해당) 다음을 수행합니다 .

    1. Line 열 머리글을 클릭하여 해당 열을 선택하고 **열 > 재코딩**을 선택합니다 .
    2. **새 열을 클릭하고 현재 위치를 선택합니다** .
    3. **새 값** 열의 값을 아래의 그림 6.41 에 표시된 것과 매칭되도록 변경합니다 .

그림 6.41 열 값 재코딩



4. **재코딩**을 클릭합니다 .

이제 JMP 분석에 적절하게 새 데이터 테이블이 구성되었습니다. 각 행에는 단일 시리얼 상자에 대한 데이터가 포함되어 있습니다. 첫 번째 열에는 상자 ID가 있고, 두 번째 열에는 생산 라인이 있으며, 세 번째 열에는 상자의 중량이 있습니다(그림 6.42).

그림 6.42 재코딩된 데이터 테이블

	ID	Line	Weight
1	215	Line A	12.42
2	287	Line A	12.49
3	381	Line A	12.8
4	705	Line B	13.63
5	670	Line B	12.56
6	715	Line B	12.87
7	683	Line B	13.09
8	514	Line B	13.31
9	517	Line B	12.64
10	946	Line B	12.75
11	254	Line C	11.73
12	282	Line C	11.4
13	938	Line C	12.78
14	597	Line C	12.19
15	179	Line C	12.25

일원 분석 수행

이 부분에는 다음과 같은 작업이 포함되어 있습니다 .

- 일원 분산 분석을 수행하여 세 개의 생산 라인 간에 평균 충진 중량의 차이가 있는지 검정합니다 .
- 비교 원을 생성하여 차이가 있을 수 있는 라인을 확인합니다 .
- 가중치를 조정하거나 상자를 보다 상세히 검토하려는 경우 ID 로 점에 라벨을 지정합니다 .

시작하기 전에 Stacked 데이터 테이블을 사용하고 있는지 확인하십시오.

1. **분석 > X 로 Y 적합**을 선택합니다 .
2. Weight 를 선택하고 **Y, 반응**을 클릭합니다 .
3. Line 을 선택하고 **X, 요인**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "Weight 대 Line 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .

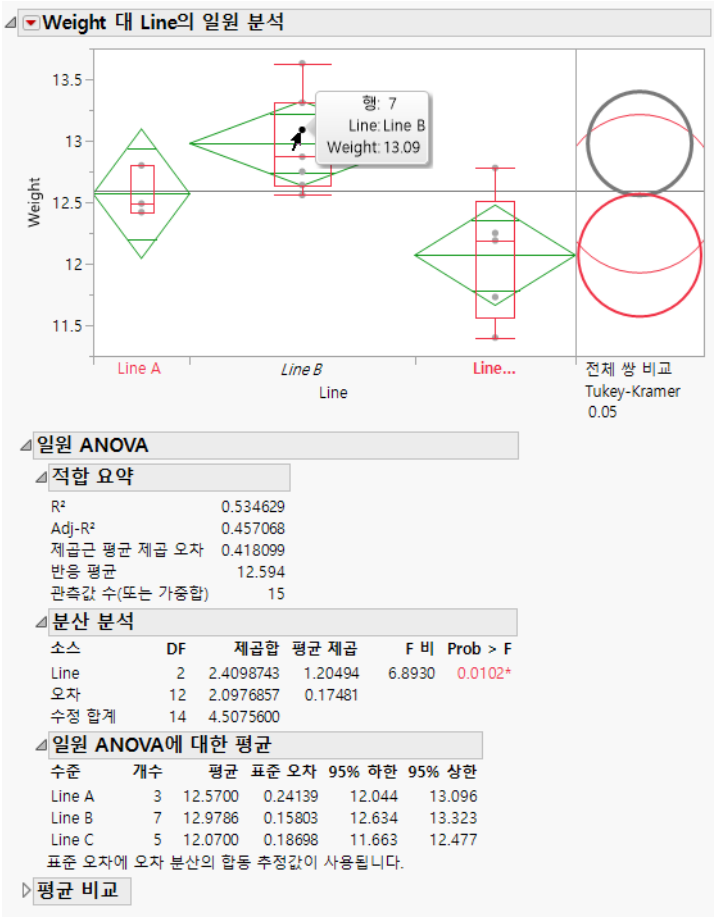
   그림의 평균 다이아몬드는 생산 라인 평균에 대한 95% 신뢰 구간을 보여 줍니다 . 평균 다이아몬드 외부의 점은 이상치인 것처럼 보일 수도 있습니다 . 하지만 그렇지 않습니다 . 이를 확인하기 위해 그림에 상자 그림을 추가합니다 .
6. "Weight 대 Line 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림**을 선택합니다 .

   모든 점이 상자 그림 경계 내에 있습니다 . 따라서 모두 이상치가 아닙니다 .
7. 데이터 테이블의 " 열 " 패널에서 ID 를 마우스 오른쪽 버튼으로 클릭하고 **라벨 / 라벨 해제**를 선택합니다 .

8. 그림에서 점을 마우스로 가리키면 해당 점의 ID 값과 함께 Line 및 Weight 데이터가 표시됩니다 (그림 6.43).
9. "Weight 대 Line 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 전체 쌍 비교 (Tukey HSD)** 를 선택합니다 .

   그림 오른쪽의 패널에 비교 원이 표시됩니다 .
10. 맨 아래 비교 원을 클릭합니다 .

그림 6.43 Weight 대 Line 의 일원 분석



"분산 분석" 보고서에서 p 값이 0.0102인 것은 평균이 모두 동일하지 않다는 증거입니다. 이 그림에는 라인 C에 대한 비교 원이 선택되어서 빨간색으로 표시되어 있습니다. 라인 B에 대한 원은 굵은 회색으로 표시되므로 라인 C의 평균은 유의 수준 0.05에서 라인 B의 평균과 차이가 있습니다. 라인 A 및 B의 평균은 통계적으로 유의한 차이를 나타내지 않습니다.

그림에 표시된 평균 다이아몬드는 평균에 대한 95% 신뢰 구간 범위에 있습니다. 95% 신뢰 구간의 수치 한계는 "일원 ANOVA에 대한 평균" 보고서에서 제공됩니다. 이 두 가지는 모두 라인 B

및 C에 대한 신뢰 구간에 목표 충진 중량인 12.5가 포함되지 않았음을 나타냅니다. 라인 B는 중량 초과로 보이고 라인 C는 중량 미달로 보입니다. 이 두 생산 라인의 경우 목표 충진 중량 미충족을 초래한 근본적인 원인을 해결해야 합니다.

일원 분석 플랫폼에 대한 통계 상세 정보

- "비교 원"
- "검정력"
- "적합 요약 보고서"
- "분산 동일 여부 검정"
- "비모수 검정 통계량"

비교 원

비교 원은 다중 비교 검정에서 LSD(최소 유의차)의 그래픽 표현입니다. 이 최소 유의차는 스튜던트 *t* 통계량에 두 평균의 차이에 대한 표준 오차를 곱한 것으로, 다음과 같이 정의됩니다.

$$\text{LSD} = t_{\alpha/2}\text{std}(\hat{\mu}_1 - \hat{\mu}_2)$$

두 독립 평균의 차이에 대한 표준 오차는 다음 관계식으로 계산됩니다.

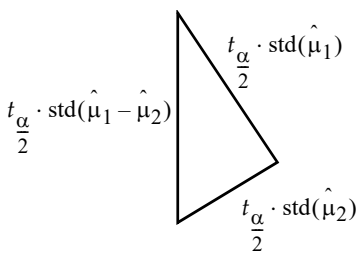
$$[\text{std}(\hat{\mu}_1 - \hat{\mu}_2)]^2 = [\text{std}(\hat{\mu}_1)]^2 + [\text{std}(\hat{\mu}_2)]^2$$

평균 간에 상관관계가 없는 경우 이러한 통계량은 다음과 같은 관계를 갖습니다.

$$\text{LSD}^2 = [t_{\alpha/2}\text{std}((\hat{\mu}_1 - \hat{\mu}_2))]^2 = [t_{\alpha/2}\text{std}(\hat{\mu}_1)]^2 + [t_{\alpha/2}\text{std}\hat{\mu}_2]^2$$

이러한 제곱 값 간에는 그림 6.44에 그래픽으로 표시된 직각 삼각형의 경우처럼 피타고라스 관계식이 성립됩니다.

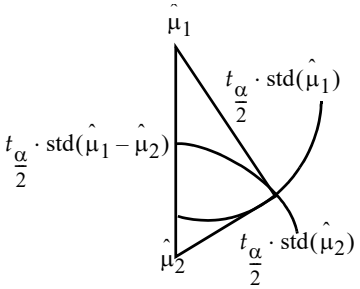
그림 6.44 두 평균 간의 차이 관계



이 삼각형의 빗변은 평균을 비교하기 위한 척도가 됩니다. 실제 차이가 빗변(LSD)보다 큰 경우에만 평균이 유의하게 다른 것으로 간주됩니다.

두 평균이 정확히 경계선에 있고 실제 차이는 최소 유의차와 동일하다고 가정해 보겠습니다. 세로 척도에서 측정된 평균 값을 꼭지점으로 하는 삼각형을 그려 봅니다. 또한 각 평균을 중심으로 하고 해당 평균의 신뢰 구간을 지름으로 하는 원을 그려 봅니다.

그림 6.45 *t*- 검정 통계량의 기하학적 관계



각 원의 반지름은 삼각형의 해당 변 길이, 즉 $t_{\alpha/2}\text{std}(\hat{\mu}_i)$ 입니다.

원은 삼각형의 변과 동일하게 직각으로 교차해야 하며 그 관계는 다음과 같습니다 .

- 평균이 정확히 최소 유의차만큼 다르면 각 평균을 중심으로 한 신뢰 구간 원이 직각으로 교차합니다 . 즉 , 탄젠트 각도가 직각입니다 .

이제 평균 간에 최소 유의차보다 크거나 작은 차이가 있을 때 원이 교차하는 방식을 살펴보겠습니다 .

- 원의 교차 외각이 직각보다 큰 경우에는 평균이 유의하게 **다르지 않은** 것입니다 . 원의 교차 외각이 직각보다 작은 경우에는 평균이 유의하게 **다른** 것입니다 . 즉 , 외각이 90 도 미만이면 평균이 최소 유의차에서 멀리 떨어져 있는 것입니다 .
- 원이 교차하지 않으면 평균이 유의하게 다른 것입니다 . 원이 내포 관계에 있으면 평균이 유의하게 다르지 않은 것입니다 (그림 6.11).

스튜던트 *t* 대신 다른 확률 분위수 값을 사용하여 다양한 다중 비교 검정에 이와 동일한 그래픽 기법을 적용할 수 있습니다.

검정력

검정력을 계산하려면 비중심 *F* 분포를 사용합니다 . 계산식 (O'Brien and Lohr 1984) 은 다음과 같이 정의됩니다 .

검정력 = Prob($F > F_{crit}$, ν_1, ν_2, nc)

다음은 각 요소에 대한 설명입니다 .

F : 비중심 $F(nc, \nu_1, \nu_2)$ 분포를 따르며 , $F_{crit} = F_{(1-\alpha, \nu1, \nu2)}$ 는 자유도가 ν_1 및 ν_2 인 *F* 분포의 1 - α 분위수입니다 .

$\nu_1 = r - 1$: 분자 DF 입니다 .

$\nu_2 = r(n-1)$: 분모 DF 입니다 .

n : 그룹당 개수입니다 .

r : 그룹 수입니다 .

$nc = n(CSS)/\sigma^2$: 비중심성 모수입니다 .

$CSS = \sum_{g=1}^{r} (\mu_g - \mu)^2$: 수정 제곱합입니다 .

μ_g : g 번째 그룹의 평균입니다 .

μ : 전체 평균입니다 .

σ^2 : MSE(평균 제곱 오차) 를 사용하여 추정됩니다 .

적합 요약 보고서

R^2

모형에 대한 분산 분석 보고서의 통계량을 사용하는 경우 연속형 반응 적합에 대한 R^2 는 다음과 같이 계산됩니다 .

$$\frac{\text{Sum of Squares (Model)}}{\text{Sum of Squares (C Total)}}$$

Adj-R^2

Adj-R^2 는 제곱합 대신 사용되는 평균 제곱의 비율로 , 다음과 같이 계산됩니다 .

$$1 - \frac{\text{Mean Square (Error)}}{\text{Mean Square (C Total)}}$$

오차의 평균 제곱은 "분산 분석" 보고서에서 구하며, **수정 합계**의 평균 제곱은 **수정 합계** 제곱합을 해당 자유도로 나눠서 계산할 수 있습니다. 자세한 내용은 "분산 분석 보고서"(167 페이지)에서 확인하십시오.

분산 동일 여부 검정

F 비

O'Brien 검정에서는 새 변수의 그룹 평균이 원래 반응의 그룹 표본 분산과 동일하도록 종속 변수를 생성합니다 . O'Brien 변수는 다음과 같이 계산됩니다 .

$$r_{ijk} = \frac{(n_{ij}-1.5)n_{ij}(y_{ijk}-\overline{y_{ij}})^2 - 0.5s_{ij}^2(n_{ij}-1)}{(n_{ij}-1)(n_{ij}-2)}$$

여기서 n은 y_{ijk} 관측값의 개수를 나타냅니다.

Brown-Forsythe는 $z_{ij} = |y_{ij}-\tilde{y}_i|$ 에 대한 ANOVA에서 구한 모형 F 통계량이며, 여기서 $\tilde{y}_i$는 i번째 수준의 중앙값 반응입니다.

Levene F는 $z_{ij} = |y_{ij}-\overline{y_{i.}}|$ 에 대한 ANOVA에서 구한 모형 F 통계량이며, 여기서 $\overline{y_{i.}}$ 는 i번째 수준의 평균 반응입니다.

Bartlett 검정은 다음과 같이 계산됩니다 .

$$T = \frac{v\log\left(\sum_i \frac{v_i}{v}s_i^2\right) - \sum_i v_i\log(s_i^2)}{1+\left(\frac{\sum_i \frac{1}{v_i} - \frac{1}{v}}{3(k-1)}\right)} \quad \text{단 , } v_i = n_i - 1 \text{ 및 } v = \sum_i v_i$$

여기서 n_i는 i번째 수준의 관측값 수이고, s_i^2은 i번째 수준의 반응 표본 분산입니다. Bartlett 통계량은 χ^2 분포를 따릅니다. 카이제곱 검정 통계량을 자유도로 나누면 보고된 F 값이 됩니다.

Welch 검정 F 비

Welch 검정 F 비는 다음과 같이 계산됩니다 .

$$F = \frac{\left[\frac{\sum_i w_i(\bar{y}_i - \tilde{y}_{..})^2}{k-1}\right]}{\left\{1+\frac{2(k-2)}{k^2-1}\left[\sum_i \frac{\left(1-\frac{w_i}{u}\right)^2}{n_i-1}\right]\right\}} \quad \text{단 , } w_i = \frac{n_i}{s_i^2},\ u = \sum_i w_i,\ \tilde{y}_{..} = \sum_i \frac{w_i\bar{y}_{i.}}{u},$$

여기서 n_i는 i번째 수준의 관측값 수이고, $\bar{y}_{i.}$ 는 i번째1수준의 평균 반응이며, s_i^2은 i번째 수준의 반응 표본 분산입니다.

Welch 검정 DF Den

분모 자유도에 대한 Welch 근사 계산식은 다음과 같이 정의됩니다 .

$$df = \frac{1}{\left(\frac{3}{k^2-1}\right)\left[\sum_i \frac{\left(1-\frac{w_i}{u}\right)^2}{n_i-1}\right]}$$

여기서 w_i, n_i 및 u 는 F 비 계산식에서의 정의와 같습니다.

비모수 검정 통계량

이 섹션에서는 Wilcoxon, 중앙값, van der Waerden 및 Friedman Rank 검정에 사용되는 검정 통계량의 계산식을 제공합니다.

표기

언급된 검정은 점수를 기반으로 하며 다음과 같은 표기를 사용합니다 .

$j = 1,\ldots, n$ 전체 표본의 관측값입니다 .

$i = 1,\ldots, k$ X 의 수준입니다 . 여기서 k 는 총 수준 수입니다 .

$n_1, n_2,\ldots, n_k$ X 의 k 개 수준 각각에 포함된 관측값의 개수입니다 .

R_j j 번째 관측값의 중간 순위입니다 . **중간 순위**는 해당 관측값의 순위 (동일 점수가 없는 경우) 이거나 평균 순위 (동일 점수가 있는 경우) 입니다 .

α 다양한 검정에서 점수를 정의하는 데 사용되는 중간 순위의 함수입니다 .

시작 창에서 블록 변수를 지정한 경우에는 다음 표기가 사용됩니다 .

$b = 1,\ldots, B$ 블록 변수의 수준입니다 . 여기서 B 는 총 블록 수입니다 .

R_{bi} 블록 b 내에 있는 X 의 i 번째 수준에 해당되는 중간 순위입니다 .

함수 α 는 점수를 다음과 같이 정의합니다 .

Wilcoxon 점수

$$\alpha(R_j) = R_j$$

중앙값 점수

$$\alpha(R_j) = \begin{cases} 1 \text{ 단, } R_j > \text{중앙값} \\ 0 \text{ if } R_j < \text{중앙값} \\ t \text{ 단, } R_j = \text{중앙값} \end{cases}$$

n_t 가 중앙값 위치에 있는 동일 관측값의 개수이고 , n_u 이 중앙값보다 큰 관측값의 개수라고 할 때 , t 는 다음과 같이 구합니다 .

$$t = \frac{\text{floor}(n/2) - n_u}{n_t}$$

van der Waerden 점수

$$\alpha(R_j) = \text{표준 정규 분위수}(R_j/(n+1))$$

Friedman 순위 점수

$$\alpha(R_{bi}) = R_{bi}$$

2 표본 정규 근사

정규 근사를 기반으로 하는 검정은 X 의 수준이 정확히 두 개뿐인 경우에만 가능합니다 . 이 섹션에서 사용되는 표기는 " 표기 "(228 페이지) 에 정의되어 있습니다 . 2 표본 정규 근사 보고서에 표시되는 통계량은 아래와 같이 정의됩니다 .

S 통계량 S 는 더 작은 그룹에 있는 관측값에 대한 $\alpha(R_j)$ 값의 합입니다 . X 의 두 수준에 동일한 수의 관측값이 있으면 S 값은 값 순서화 시 X 의 마지막 수준에 대응합니다 .

Z Z 값은 다음과 같이 정의됩니다 .

$$Z = (S - E(S))/\sqrt{Var(S)}$$

참고 : Wilcoxon 검정에서는 연속성 수정이 추가됩니다 . $(S - E(S))$ 가 0 보다 크면 분자에서 0.5 를 뺍니다 . $(S - E(S))$ 가 0 보다 작으면 분자에 0.5 를 더합니다 .

***E*(*S*)** 귀무가설하의 S 기대값입니다 . 더 작은 수준 또는 값 순서화 시 마지막 수준 (두 그룹의 관측값 수가 동일한 경우) 의 관측값 수는 n_l 로 나타냅니다 .

$$E(S) = \frac{n_l}{n} \sum_{j=1}^{n} \alpha(R_j)$$

***Var*(*S*)** 모든 관측값의 평균 점수를 *ave* 라고 정의할 경우 S 의 분산은 다음과 같이 정의됩니다 .

$$Var(S) = \frac{n_1 n_2}{n(n-1)} \sum_{j=1}^{n} (\alpha(R_j) - ave)^2$$

Friedman 순위 검정의 2 표본 정규 근사

Friedman 순위 검정을 사용하는 경우 2 표본 정규 근사에 대한 계산은 위와 동일하되 S 의 분산만 다릅니다 . S 의 분산은 다음과 같이 계산됩니다 .

$$Var(S) = \frac{B}{(n-1)} \sum_{j=1}^{n} (\alpha(R_j) - ave)^2$$

일원 카이제곱 근사

참고 : Wilcoxon 점수를 기반으로 하는 카이제곱 검정은 Kruskal-Wallis 검정이라고 합니다 .

이 섹션에서 사용되는 표기는 " 표기 "(228 페이지) 에 정의되어 있습니다 . 카이제곱 통계량을 계산하는 데는 다음과 같은 통계량이 사용됩니다 .

T_i X 의 i 번째 수준에 대한 총 점수입니다 .

$E(T_i)$ 수준 간에 차이가 없다는 귀무가설하에 수준 i 의 총 점수에 대한 기대값으로 , 다음과 같이 정의됩니다 .

$$E(T_i) = \frac{n_i}{n} \sum_{j=1}^{n} \alpha(R_j)$$

$Var(T)$ 모든 관측값의 평균 점수를 ave 라고 정의할 경우 T 의 분산은 다음과 같이 정의됩니다 .

$$Var(T) = \frac{1}{(n-1)} \sum_{j=1}^{n} (\alpha(R_j) - ave)^2$$

이 검정 통계량의 값은 아래와 같이 구합니다 . 이 통계량은 $k - 1$ 자유도에 대한 점근적 카이제곱입니다 .

$$C = \left(\sum_{i=1}^{k} (T_i - E(T_i))^2 / n_i \right) / Var(T)$$

Friedman 순위 검정에 대한 일원 카이제곱 근사

Friedman 순위 검정에 대한 카이제곱 검정 통계량은 다음과 같이 계산됩니다 .

$$C = \frac{\sum_{i=1}^{k} (T_i - E(T_i))^2 / n_i}{\frac{n}{\frac{1}{(k-1)} \sum_{j=1} (\alpha(R_j) - ave)^2 / n_i}}$$

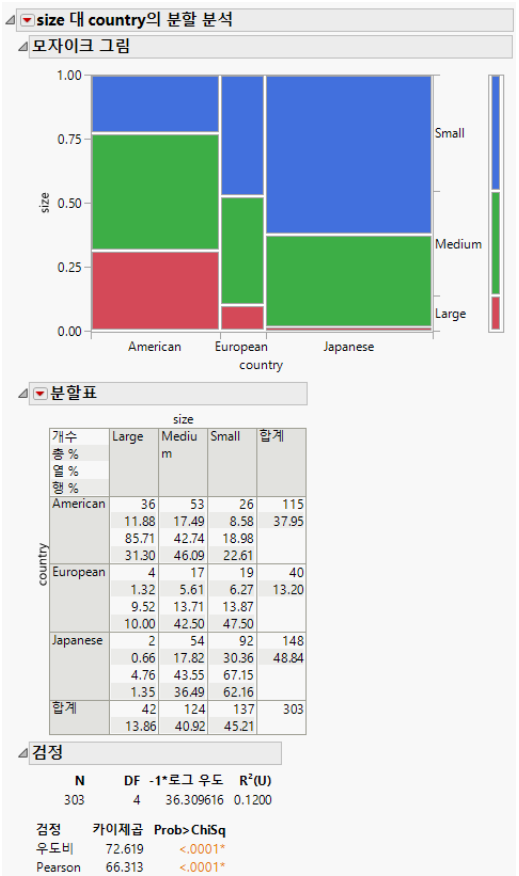
7 장

분할 분석

두 범주형 변수 간의 관계 분석

분할 분석 플랫폼을 사용하여 두 범주형 변수 사이의 관계를 조사합니다. 범주형 변수는 명목형 또는 순서형일 수 있습니다. 분석 결과에 모자이크 그림, 빈도 수 및 비율의 분할표와 유의성 카이제곱 검정이 포함됩니다. 대화식으로 비율에 대한 평균 분석, 대응 분석, 연관성 분석 등의 추가적인 데이터 분석 및 검정을 수행할 수 있습니다. 분할 분석 플랫폼은 X로 Y 적합 플랫폼의 **범주형 대 범주형** 분석법입니다.

그림 7.1 분할 분석의 예



개수 총 % 열 % 행 %	Large	Medium	Small	합계
American	36	53	26	115
	11.88	17.49	8.58	37.95
	85.71	42.74	18.98	
	31.30	46.09	22.61	
European	4	17	19	40
	1.32	5.61	6.27	13.20
	9.52	13.71	13.87	
	10.00	42.50	47.50	
Japanese	2	54	92	148
	0.66	17.82	30.36	48.84
	4.76	43.55	67.15	
	1.35	36.49	62.16	
합계	42	124	137	303
	13.86	40.92	45.21	

목차

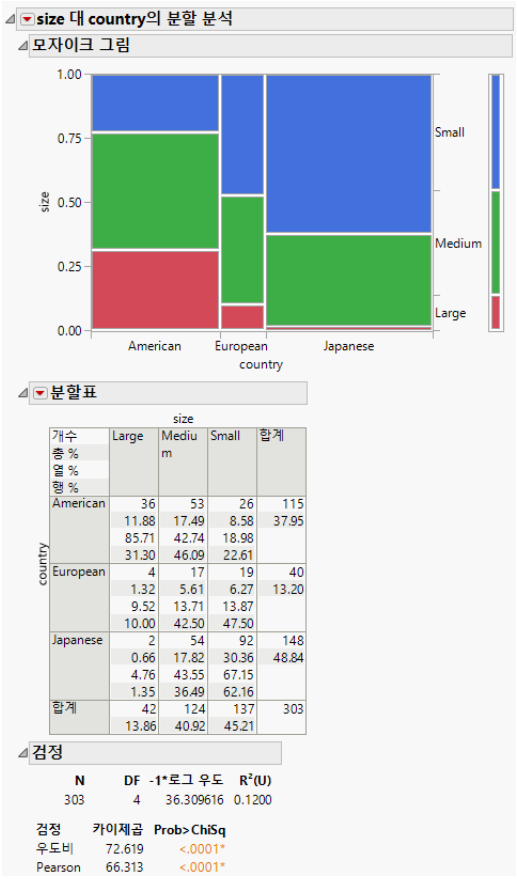
분할 분석의 예 235
분할 분석 플랫폼 시작 236
데이터 형식 236
분할 분석 보고서 237
분할 분석 플랫폼 옵션 238
모자이크 그림 239
팝업 메뉴 240
분할표 241
분할표 설명 243
검정 243
검정 보고서 설명 244
Fisher 정확 검정 244
비율에 대한 평균 분석 244
대응 분석 245
대응 분석 그림 이해 245
대응 분석 옵션 245
상세 정보 보고서 246
Cochran-Mantel-Haenszel 검정 246
합치도 통계량 247
상대 위험도 247
비율에 대한 2 표본 검정 248
연관성 측도 248
Cochran Armitage 추세 검정 249
정확 검정 250
분할 분석 플랫폼의 추가 예 250
비율에 대한 평균 분석의 예 251
대응 분석의 예 252
Cochran Mantel Haenszel 검정의 예 254
합치도 통계량 옵션의 예 256
상대 위험도 옵션의 예 257
비율에 대한 2표본 검정의 예 258
연관성 측도 옵션의 예 259
Cochran Armitage 추세 검정의 예 260
분할 분석 플랫폼에 대한 통계 상세 정보 261
합치도 통계량 261
승산비 옵션 262
검정 보고서 262
대응 분석의 상세 정보 보고서 263

분할 분석의 예

JMP 에서 두 범주형 변수 사이의 관계를 검토하는 방법에 대해 알아봅니다 . 이 예에서는 자동차에 대한 설문조사에서 수집된 데이터를 사용합니다 . 이 데이터에는 응답자 속성 , 즉 성별 , 결혼 여부 및 연령이 포함되어 있습니다 . 또한 응답자 소유 자동차의 속성 , 즉 생산 국가 , 크기 및 유형도 포함되어 있습니다 . 자동차 크기 (소형 , 중형 및 대형) 와 자동차 생산 국가 사이의 관계를 검토하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. size 를 선택하고 **Y, 반응**을 클릭합니다 .
4. country 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 7.2 분할 분석의 예



모자이크 그림 및 범례에서 다음을 확인할 수 있습니다 .

- 대형 크기 범주에 속하는 일본 자동차는 매우 적습니다 .
- 유럽 자동차의 대부분은 소형 및 중형 크기 범주에 속합니다 .
- 미국 자동차의 대부분은 대형 및 중형 크기 범주에 속합니다 .

분할 분석 플랫폼 시작

분할 분석을 수행하려면 다음을 수행하십시오 .

1. **분석 > X 로 Y 적합**을 선택합니다 .
2. **Y, 반응**에 명목형 또는 순서형 열을 입력합니다 .
3. **X, 요인**에 명목형 또는 순서형 열을 입력합니다 .

그림 7.3 X 로 Y 적합 시작 창

다이어그램 위에 "분할 분석"이라고 표시되면 분할 분석을 수행할 것임을 나타냅니다.

참고 : "JMP 시작하기 " 창에서 분할 분석을 시작할 수도 있습니다 . **보기** > **JMP 시작하기** > **기본 분석** > **분할 분석**을 선택합니다 .

이 시작 창에 대한 자세한 내용은 "X 로 Y 적합 소개" 장(105 페이지)에서 확인하십시오 . "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

데이터 형식

범주형 데이터는 종종 Y 및 X 변수의 각 수준 조합마다 하나씩의 행만 데이터 테이블에 포함된 요약 형식으로 제공됩니다 . 이 경우 빈도 또는 가중치 변수를 사용하여 각 행이 나타내는 관측값

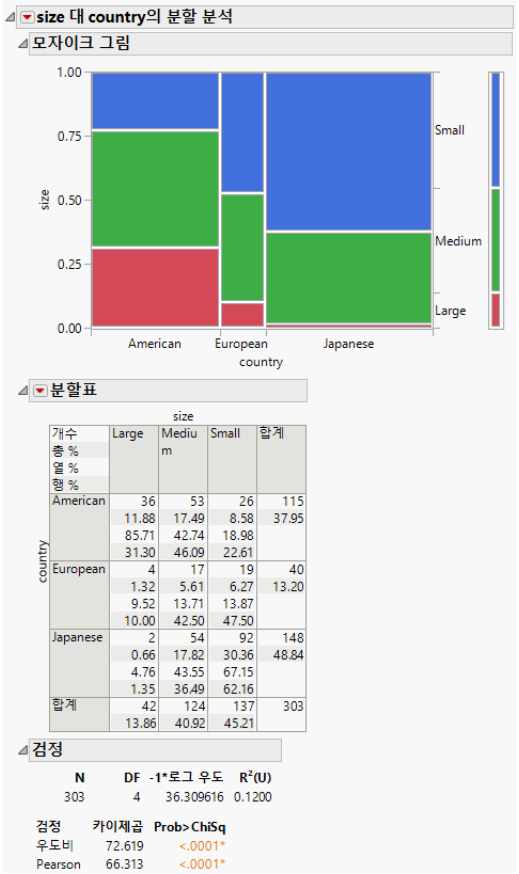
의 개수를 나타낼 수 있습니다. 요약된 범주형 데이터의 예는 "비율에 대한 평균 분석의 예"(251 페이지)에서 확인하십시오.

분할 분석 보고서

분할 분석 보고서에는 기본적으로 모자이크 그림, 분할표 및 검정 보고서가 표시됩니다. 빨간색 삼각형 메뉴 옵션을 사용하여 추가 분석 및 검정을 실행할 수 있습니다. 이러한 모든 보고서 및 옵션에 대한 자세한 내용은 "분할 분석 플랫폼 옵션"(238 페이지)에서 확인하십시오.

그림 7.4 에 표시된 것과 같은 그림을 생성하려면 "분할 분석의 예"(235 페이지)에 설명된 방법을 따르십시오.

그림 7.4 분할 분석 보고서의 예



참고 : 데이터 테이블에서 제외된 모든 행은 모자이크 그림에서도 숨겨집니다.

분할 분석 플랫폼 옵션

분할 분석 빨간색 삼각형 메뉴에 있는 옵션을 사용하여 추가 분석을 수행합니다.

참고 : " 적합 그룹 " 메뉴는 Y 변수를 여러 개 지정한 경우에 표시됩니다 . 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 를 기준으로 정렬할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

모자이크 그림 분할표의 데이터를 그래픽으로 나타낸 것입니다 . 자세한 내용은 " 모자이크 그림 "(239 페이지) 에서 확인하십시오 .

분할표 이원 빈도 테이블입니다 . 각 요인 수준마다 하나씩의 행과 각 반응 수준마다 하나씩의 열이 있습니다 . 자세한 내용은 " 분할표 "(241 페이지) 에서 확인하십시오 .

검정 연속형 데이터에 대한 분산 분석 테이블과 유사합니다 . X 수준 간에 반응 수준 비율이 동일한지 검정한 결과를 보여 줍니다 . 자세한 내용은 " 검정 "(243 페이지) 에서 확인하십시오 .

α 수준 설정 신뢰 구간에 사용된 유의 수준을 변경합니다 . 일반적인 값 (0.10, 0.05, 0.01) 중 하나를 선택하거나 , **기타** 옵션을 사용하여 특정 값을 선택합니다 .

비율에 대한 평균 분석 (반응의 수준이 정확히 두 개뿐인 경우에만 표시됨) X 수준에 대한 반응 비율을 전체 반응 비율과 비교합니다 . 자세한 내용은 " 비율에 대한 평균 분석 "(244 페이지) 에서 확인하십시오 .

대응 분석 빈도 테이블에서 유사한 개수 패턴을 가진 행 또는 열을 표시합니다 . 대응 분석 그림에는 분할표의 각 행과 각 열에 해당하는 점이 하나씩 있습니다 . 자세한 내용은 " 대응 분석 "(245 페이지) 에서 확인하십시오 .

Cochran Mantel Haenszel 세 번째 분류 기준에 따라 블록화한 후 두 범주형 변수 사이에 관계가 있는지 검정합니다 . 자세한 내용은 "Cochran-Mantel-Haenszel 검정 "(246 페이지) 에서 확인하십시오 .

합치도 통계량 (X 및 Y 변수의 수준이 동일한 경우에만 표시됨) 카파 통계량 (Agresti 1990), 해당 표준 오차 , 신뢰 구간 , 가설 검정 및 Bowker 대칭성 검정 (McNemar 검정이라고도 함) 을 표시합니다 . 자세한 내용은 " 합치도 통계량 "(247 페이지) 에서 확인하십시오 .

상대 위험도 (X 및 Y 변수의 수준이 두 개인 경우에만 표시됨) 위험 비율을 계산합니다 . 자세한 내용은 " 상대 위험도 "(247 페이지) 에서 확인하십시오 .

승산비 (X 및 Y 변수의 수준이 두 개인 경우에만 표시됨) 승산비 보고서를 생성합니다 . 자세한 내용은 " 승산비 옵션 "(262 페이지) 에서 확인하십시오 .

보고서에는 이 비율에 대한 신뢰 구간도 표시됩니다 . **α 수준 설정** 옵션을 사용하여 유의 수준을 변경할 수 있습니다 .

비율에 대한 2 표본 검정 (X 및 Y 변수의 수준이 두 개인 경우에만 표시됨) 비율에 대한 2 표본 검정을 수행합니다 . 이 검정에서는 X 변수의 두 수준 간에 Y 변수의 비율을 비교합니다 . 자세한 내용은 " 비율에 대한 2 표본 검정 "(248 페이지) 에서 확인하십시오 .

연관성 측도 분할표의 변수 간 연관성을 설명합니다 . 자세한 내용은 " 연관성 측도 "(248 페이지) 에서 확인하십시오 .

Cochran Armitage 추세 검정 (한 변수의 수준이 두 개이고 다른 변수는 순서형인 경우에만 표시됨) 단일 변수의 수준 간에 이항 비율의 추세를 검정합니다 . 자세한 내용은 "Cochran Armitage 추세 검정 "(249 페이지) 에서 확인하십시오 .

JMP PRO 정확 검정 다음 검정의 정확 검정 버전을 제공합니다 .

- Fisher 검정
- Cochran Armitage 추세 검정
- 합치도 통계량

자세한 내용은 " 정확 검정 "(250 페이지) 에서 확인하십시오 .

표시 옵션 > 가로 모자이크 모자이크 그림을 가로 또는 세로로 회전합니다 .

데이터 테이블로 만들기 보고서 테이블을 사용하여 JMP 데이터 테이블을 생성합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

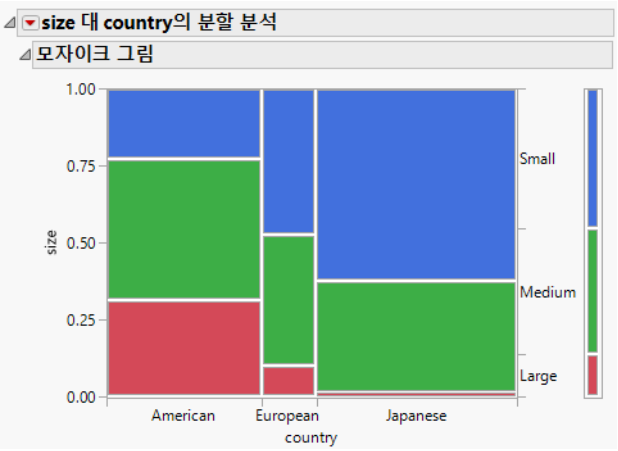
그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

모자이크 그림

모자이크 그림은 이원 빈도 테이블 또는 분할표를 그래픽으로 나타낸 것입니다 . 모자이크 그림은 여러 개의 직사각형으로 나뉘며 , 각 직사각형의 세로 길이는 X 변수의 각 수준에서 Y 변수가 나타나는 비율에 비례합니다 . 모자이크 그림은 Hartigan and Kleiner 연구 자료 (1981) 에서 처음 소개되었으며 Friendly 연구 자료 (1994) 에서 구체화되었습니다 .

그림 7.5 에 표시된 것과 같은 그림을 생성하려면 "분할 분석의 예 "(235 페이지) 에 설명된 방법을 따르십시오 .

그림 7.5 모자이크 그림의 예



이 모자이크 그림의 경우 다음 사항에 유의하십시오 .

- 가로 축에서의 비율은 X 변수 (country) 의 수준별 관측값 수를 나타냅니다 .
- 오른쪽 세로 축에서의 비율은 결합된 수준(American, European 및 Japanese) 전체의 자동차 크기 (Small, Medium 및 Large) 별 비율을 나타냅니다 .
- 왼쪽 세로 축의 척도는 반응 확률을 보여 줍니다 . 전체 축은 전체 표본을 나타내는 확률 1 과 동등합니다 .

모자이크 그림의 직사각형을 클릭하면 선택한 직사각형이 강조 표시되고 관련 데이터 테이블의 해당 데이터도 강조 표시됩니다.

두 가지 방법 중 하나로 변수를 드래그하여 놓아 모자이크 그림의 변수를 바꿀 수 있습니다. 한 가지 방법은 한 축의 변수를 다른 축으로 드래그하여 놓아 기존 변수를 바꾸는 것이고, 다른 방법은 관련 데이터 테이블의 "열" 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다.

팝업 메뉴

셀의 색상을 변경하고 라벨을 지정하려면 모자이크 그림을 마우스 오른쪽 버튼으로 클릭합니다 .

색상 설정 수준에 현재 할당된 색상을 표시합니다 . 자세한 내용은 " 색상 설정 "(241 페이지) 에서 확인하십시오 .

셀 라벨 지정 모자이크 그림에 표시할 라벨을 지정합니다 . 다음 옵션 중 하나를 선택합니다 .

라벨 없음 라벨을 표시하지 않고 , 다른 옵션을 모두 제거합니다 .

개수별 라벨 각 셀의 관측값 수를 표시합니다 .

백분율별 라벨 각 셀의 관측값 백분율을 표시합니다 .

값별 라벨 각 셀에 해당하는 Y 변수의 수준을 표시합니다 .

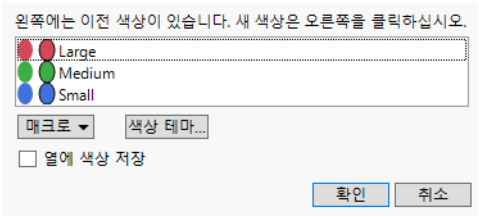
행별 라벨 셀이 나타내는 모든 행의 행 라벨을 표시합니다 .

참고 : 나머지 마우스 오른쪽 버튼 클릭 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

색상 설정

색상 설정 옵션을 선택하면 "값에 대한 색상 선택" 창이 나타납니다.

그림 7.6 값에 대한 색상 선택 창



기본 모자이크 색상은 반응 열이 순서형인지 명목형인지, 그리고 기존의 "값 색상" 열 특성이 있는지에 따라 다릅니다. 수준 색상을 변경하려면 두 번째 색상 열의 타원을 클릭하고 새 색상을 선택합니다.

값에 대한 색상 선택 창 설명

매크로 두 수준 간의 색상 그래디언트를 계산합니다 .

끝 사이의 그래디언트 변수의 모든 수준에 그래디언트를 적용합니다 .

선택한 점 사이의 그래디언트 선택된 수준에 색상 그래디언트를 적용합니다 . 선택하려는 수준 위에서 커서를 드래그하거나 Shift 키를 누른 채 첫 번째 수준과 마지막 수준을 클릭하면 수준 범위를 선택할 수 있습니다 .

색상 반전 색상 순서가 반전됩니다 .

이전 색상으로 되돌리기 변경 사항을 되돌립니다 .

색상 테마 색상 테마를 기반으로 각 값의 색상을 변경합니다 .

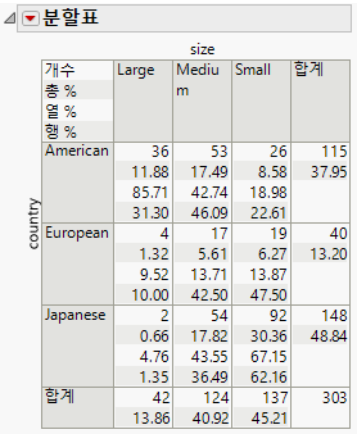
열에 색상 저장 기본 색상 테마를 변경한 다음 이 체크박스를 선택하면 관련 데이터 테이블의 열에 새 열 특성 (**값 색상**) 이 추가됩니다 . 데이터 테이블에서 이 특성을 편집하려면 **열 > 열 정보**를 선택합니다 .

분할표

분할표는 이원 빈도 테이블입니다. 각 요인 수준마다 하나씩의 행과 각 반응 수준마다 하나씩의 열이 있습니다.

그림 7.7에 표시된 것과 같은 그림을 생성하려면 "분할 분석의 예"(235 페이지)에 설명된 방법을 따르십시오.

그림 7.7 분할표의 예



분할표

country	개수 / 총 % / 열 % / 행 %	size: Large	size: Mediu m	size: Small	합계
American		36	53	26	115
		11.88	17.49	8.58	37.95
		85.71	42.74	18.98	
		31.30	46.09	22.61	
European		4	17	19	40
		1.32	5.61	6.27	13.20
		9.52	13.71	13.87	
		10.00	42.50	47.50	
Japanese		2	54	92	148
		0.66	17.82	30.36	48.84
		4.76	43.55	67.15	
		1.35	36.49	62.16	
합계		42	124	137	303
		13.86	40.92	45.21	

분할표의 경우 다음 사항에 유의하십시오.

- "개수", "총 %", "열 %" 및 "행 %"는 행 및 열 머리글이 있는 각 셀(예: "American" 및 "Large" 아래의 셀) 내의 데이터에 해당합니다.
- 마지막 열에는 각 행의 총 개수와 각 행의 총 백분율이 포함됩니다.
- 맨 아래 행에는 각 열의 총 개수와 각 열의 총 백분율이 포함됩니다.

그림 7.7에서 대형이면서 미국에서 생산된 자동차를 집중적으로 살펴보겠습니다. 다음 표에서는 분할표를 사용하여 이러한 자동차에 대해 내릴 수 있는 결론을 설명합니다.

표 7.1 분할표 예를 기반으로 한 결론

개수	설명	표에 사용된 라벨
36	대형이면서 미국에서 생산된 자동차의 수	개수
11.88%	대형이면서 미국에서 생산된 모든 자동차의 백분율[a]	총 %
85.71%	미국에서 생산된 대형 자동차의 백분율(36/42)[b]	열 %
31.30%	대형인 미국 자동차의 백분율(36/115)[c]	행 %
37.95%	미국에서 생산된 모든 자동차의 백분율(115/303)	(없음)
13.86%	모든 대형 자동차의 백분율(42/303)	(없음)

a. 303 은 설문조사 대상 자동차의 총 대수입니다.
b. 42 는 설문조사 대상 대형 자동차의 총 대수입니다.
c. 115 는 설문조사 대상 미국 자동차의 총 대수입니다.

팁 : 분할표의 데이터를 표시하거나 숨기려면 " 분할표 " 옆의 빨간색 삼각형 메뉴에서 표시하거나 숨기려는 옵션을 선택하십시오 .

분할표 설명

개수 셀 빈도 , 빈도 소계 및 총 합계 (총 표본 크기) 입니다 .

총 % 총 합계 대비 셀 개수 및 소계의 백분율입니다 .

행 % 행 합계 대비 각 셀 개수의 백분율입니다 .

열 % 열 합계 대비 각 셀 개수의 백분율입니다 .

기대값 독립성을 가정한 상태에서 각 셀의 기대 빈도 (*E*) 입니다 . 해당 행 합계와 열 합계의 곱을 총 합계로 나눠서 계산됩니다 .

편차 관측된 셀 빈도 (*O*) 에서 기대 셀 빈도 (*E*) 를 뺀 값입니다 .

셀 카이제곱 각 셀에 대해 $(O - E)^2 / E$ 로 계산된 카이제곱 값입니다 .

열 누적 누적 열 합계입니다 .

열 누적 % 누적 열 백분율입니다 .

행 누적 누적 행 합계입니다 .

행 누적 % 누적 행 백분율입니다 .

검정

" 검정 " 보고서에는 X 수준 간에 반응 수준 비율이 동일한지 여부를 확인하기 위한 두 가지 검정의 결과가 표시됩니다 .

그림 7.8에 표시된 것과 같은 보고서를 생성하려면 " 분할 분석의 예 "(235 페이지)에 설명된 방법을 따르십시오 .

그림 7.8 검정 보고서의 예

검정

N	DF	-1*로그 우도	R^2(U)
303	4	36.309616	0.1200

검정	카이제곱	Prob>ChiSq
우도비	72.619	<.0001*
Pearson	66.313	<.0001*

카이제곱 통계량의 경우 다음 사항에 유의하십시오 .

- 두 범주형 변수가 모두 반응 변수 (Y 변수) 인 경우 카이제곱 통계량은 두 변수가 독립적인지를 검증합니다 .

- Y 변수가 하나 있고 고정된 X 변수가 하나 있을 수도 있습니다 . 이 경우 카이제곱 통계량은 각 X 수준 간에 Y 변수의 분포가 동일한지를 검증합니다 .

검정 보고서 설명

N 관측값의 총 개수입니다 .

DF 검정과 관련된 자유도를 기록합니다 . 자유도는 $(c - 1)(r - 1)$ 와 같습니다 . 여기서 c 는 열 수이고 r 은 행 수입니다 .

-1* 로그 우도 적합 및 불확도를 측정하는 음의 로그 가능도로 , 연속형 반응에서의 제곱합과 매우 유사합니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

R^2(U) 모형 적합에 기인한 전체 불확실성의 비율입니다 .

- R^2 가 1 이면 요인이 범주형 반응을 완전히 예측함을 나타냅니다 .
- R^2 가 0 이면 고정된 기준 반응 비율 대신 모형을 사용함으로써 얻는 이득이 없음을 의미합니다 .

자세한 내용은 " 검정 보고서 "(262 페이지) 에서 확인하십시오 .

검정 각 표본 범주에서 반응 비율이 동일하다는 가설에 대한 카이제곱 통계 검정을 나열합니다 . 자세한 내용은 " 검정 보고서 "(262 페이지) 에서 확인하십시오 .

Prob>ChiSq 반응과 요인 사이에 아무런 관계가 없는 경우 카이제곱 값이 계산된 값보다 클 확률을 나열합니다 . 두 변수의 수준이 각각 두 개인 경우에는 한쪽 꼬리 검정과 양쪽 꼬리 검정에 대한 Fisher 정확 확률도 표시됩니다 .

Fisher 정확 검정

이 보고서에서는 2 x 2 테이블에 대한 Fisher 정확 검정 결과를 제공합니다. 결과는 자동으로 2 x 2 테이블에 대해 표시됩니다. Fisher 정확 검정과 r x c 테이블의 검정에 대한 자세한 내용은 " 정확 검정 "(250 페이지) 에서 확인하십시오.

비율에 대한 평균 분석

비율에 대한 평균 분석 (ANOM) 은 개별 그룹의 비율이 전체 비율과 다른지 여부를 검정하기 위한 다중 비교 방법입니다.

참고 : 평균 분석 방법에 대한 자세한 내용은 Nelson et al. 연구 자료 (2005) 에서 확인하십시오 . 자세한 내용은 " 비율에 대한 평균 분석의 예 "(251 페이지) 에서 확인하십시오 .

반응의 수준이 두 개인 경우 이 옵션을 사용하여 X 수준의 반응 비율을 전체 반응 비율과 비교할 수 있습니다. 이 방법에서는 이항 분포에 대한 정규 근사를 사용합니다. 따라서 표본 크기가 너무 작으면 결과에 경고가 표시됩니다.

빨간색 삼각형 메뉴에는 다음과 같은 옵션이 표시됩니다 .

유의 수준 설정 분석에 사용되는 유의 수준을 선택합니다 .

요약 보고서 표시 반응 비율과 X 변수의 각 수준에 대한 결정 한계를 보여 주는 보고서를 생성합니다 . 이 보고서에서는 한계가 초과되었는지 여부를 나타냅니다 .

비율에 대한 반응 수준 전환 분석에 사용된 반응 범주를 변경합니다 .

표시 옵션 결정 한계 , 결정 한계 음영 , 중심선 및 점 옵션을 표시하거나 숨깁니다 .

대응 분석

분할 분석 플랫폼의 대응 분석 옵션은 빈도 테이블에서 유사한 개수 패턴을 가진 행 또는 열을 표시하기 위한 그래픽 기법을 제공합니다. 대응 분석 그림에는 각 행과 각 열에 해당하는 점이 하나씩 있습니다. 수준 수가 많아서 모자이크 그림으로는 유용한 정보를 도출하기 어려운 경우에 대응 분석을 사용합니다.

참고 : 자세한 내용은 " 대응 분석의 예 "(252 페이지) 에서 확인하십시오 .

대응 분석 그림 이해

행 프로파일은 행별 비율의 집합, 즉 한 행의 개수들을 해당 행의 총 개수로 나눈 값들로 정의할 수 있습니다. 두 행의 행 프로파일이 매우 유사한 경우에는 대응 분석 그림에서 해당 점들이 서로 가깝게 나타납니다. 행 점 간의 거리를 제곱한 값은 행 쌍 간의 동질성을 검정하는 카이제곱 거리에 대략적으로 비례합니다.

문제는 대칭적으로 정의되므로 열 및 행 프로파일은 서로 비슷합니다. 행 점과 열 점 간의 거리는 아무 의미도 없습니다. 하지만 원점으로부터의 열 및 행 방향은 의미가 있으며 이 관계는 그림을 해석하는 데 도움이 됩니다.

대응 분석 옵션

" 대응 분석 " 빨간색 삼각형 메뉴에 있는 옵션을 사용하여 분할 분석 보고서에 3D 산점도를 추가하고 열 특성을 데이터 테이블에 추가합니다 .

3D 대응 분석 3D 산점도를 생성합니다 .

값 순서 저장 첫 번째 대응 점수 계수를 기준으로 정렬된 수준 순서를 X 및 Y 열 모두의 열 특성으로 저장합니다 .

상세 정보 보고서

" 상세 정보 " 보고서에는 대응 분석에 대한 통계 정보가 포함되고 그림에 사용된 값이 표시됩니다 .

특이값 분할표에 대한 특이값 분해를 제공합니다 . 계산식은 " 대응 분석의 상세 정보 보고서 "(263 페이지) 에서 확인하십시오 .

관성 정준 차원의 측면을 고려한 상대적인 변동을 반영하는 특이값 제곱을 나열합니다 .

비율 총 관성에 대한 관성 비율입니다 .

누적 관성의 누적 비율을 보여 줍니다 . 처음 두 개의 특이값에서 큰 관성이 발견되는 경우에는 2D 대응 분석 그림으로 테이블에서의 관계를 충분히 나타낼 수 있습니다 .

X 변수 c1, c2, c3 대응 분석 그림에 표시된 값입니다 (그림 7.11).

Y 변수 c1, c2, c3 대응 분석 그림에 표시된 값입니다 (그림 7.11).

Cochran-Mantel-Haenszel 검정

Cochran-Mantel-Haenszel 검정에서는 세 번째 분류 기준에 따라 블록화한 후 두 범주형 변수 사이의 관계를 평가합니다 .

참고 : 자세한 내용은 "Cochran Mantel Haenszel 검정의 예 "(254 페이지) 에서 확인하십시오 .

점수 상관 Y 또는 X 가 모두 순서형이거나 구간을 나타낼 때 사용할 수 있습니다 . 대립가설은 블록화 변수의 최소 한 개 수준에서 Y 와 X 사이에 선형 연관성이 있다는 것입니다 .

열 범주별 행 점수 Y 가 순서형이거나 구간을 나타낼 때 사용할 수 있습니다 . 대립가설은 블록화 변수의 최소 한 개 수준에서 *r* 개 행의 평균 점수가 동일하지 않다는 것입니다 .

행 범주별 열 점수 X 가 순서형이거나 구간을 나타낼 때 사용할 수 있습니다 . 대립가설은 블록화 변수의 최소 한 개 수준에서 *c* 개 열의 평균 점수가 동일하지 않다는 것입니다 .

범주의 일반 연관성 블록화 변수의 최소 한 개 수준에서 X 와 Y 사이에 일종의 연관성이 있는지를 검정합니다 .

합치도 통계량

두 변수의 수준이 동일하면 **합치도 통계량** 옵션을 사용할 수 있습니다 . 이 옵션은 카파 통계량 (Agresti 1990), 해당 표준 오차 , 신뢰 구간 , 가설 검정 및 Bowker 대칭성 검정을 표시합니다 .

이 섹션에서 제공되는 카파 통계량과 관련 p 값은 근사값입니다 . 정확한 합치도 통계량도 사용할 수 있습니다 . 자세한 내용은 " 정확 검정 "(250 페이지) 에서 확인하십시오 .

참고 : 통계 상세 정보는 " 합치도 통계량 "(261 페이지) 에서 확인하십시오 . 자세한 내용은 " 합치도 통계량 옵션의 예 "(256 페이지) 에서 확인하십시오 .

카파 카파 통계량을 보여 줍니다 .

표준 오차 카파 통계량의 표준 오차를 보여 줍니다 .

95% 하한 카파에 대한 신뢰 구간의 하한점을 보여 줍니다 .

95% 상한 카파에 대한 신뢰 구간의 상한점을 보여 줍니다 .

Prob>Z 카파에 대한 단측 검정의 p 값을 보여 줍니다 . 귀무가설은 카파가 0 인지를 검정합니다 .

Prob>|Z| 카파에 대한 양측 검정의 p 값을 보여 줍니다 .

카이제곱 Bowker 검정의 검정 통계량을 보여 줍니다 . Bowker 대칭성 검정에서 귀무가설은 정방 테이블의 확률이 대칭성을 만족한다는 것 , 즉 테이블 셀의 모든 쌍에 대해 $p_{ij}=p_{ji}$ 가 성립한다는 것입니다 . X 와 Y 모두 두 개의 수준이 있는 경우 이 검정은 McNemar 검정과 동일합니다 .

Prob>ChiSq Bowker 검정의 p 값을 보여 줍니다 .

상대 위험도

참고 : 자세한 내용은 " 상대 위험도 옵션의 예 "(257 페이지) 에서 확인하십시오 .

상대 위험도 옵션을 사용하여 2 x 2 분할표의 위험 비율을 계산할 수 있습니다 . 보고서에는 신뢰 구간도 표시됩니다 . 이 방법에 대한 자세한 내용은 Agresti 연구 자료 (1990, sect. 3.4.2) 에서 확인하십시오 .

상대 위험도 옵션을 선택하면 " 상대 위험도 범주 선택 " 창이 나타납니다 . 단일 반응 및 요인 조합을 선택하거나 , 반응 및 요인 수준의 모든 조합에 대한 위험 비율을 계산할 수 있습니다 .

비율에 대한 2 표본 검정

X 및 Y 변수 모두의 수준이 두 개인 경우 두 비율 간의 차이에 대한 신뢰 구간을 생성할 수 있습니다 . 이 경우 신뢰 구간에 해당하는 검정도 계산됩니다 .

참고 : 자세한 내용은 " 비율에 대한 2 표본 검정의 예 "(258 페이지) 에서 확인하십시오 .

설명 수행되는 검정을 보여 줍니다 .

비율 차이 X 변수의 수준 간 비율 차이를 보여 줍니다 .

95% 하한 차이에 대한 신뢰 구간의 하한점을 보여 줍니다 . 조정 Wald 신뢰 구간을 기반으로 합니다 .

95% 상한 차이에 대한 신뢰 구간의 상한점을 보여 줍니다 . 조정 Wald 신뢰 구간을 기반으로 합니다 .

조정 Wald 검정 양쪽 꼬리 및 한쪽 꼬리 검정을 보여 줍니다 .

확률 검정의 *p* 값을 보여 줍니다 .

< 변수 > 반응의 관심 범주 검정에 사용할 반응 수준을 선택합니다 .

연관성 측도

연관성 측도 옵션은 연관성 통계량을 제공합니다 .

참고 : 자세한 내용은 " 연관성 측도 옵션의 예 "(259 페이지) 에서 확인하십시오 .

감마 부합 / 비부합 쌍의 개수를 기반으로 하며 동일 값 쌍은 무시합니다 . -1 ~ 1 범위의 값을 취합니다 .

Kendall Tau-b 감마와 유사하되 , 동일 값에 대한 수정을 사용합니다 . -1 ~ 1 범위의 값을 취합니다 .

Stuart Tau-c 감마와 유사하되 , 테이블 크기에 대한 조정과 동일 값에 대한 수정을 사용합니다 . -1 ~ 1 범위의 값을 취합니다 .

Somers D Tau-b 에 대한 비대칭 수정입니다 .

- C|R 은 행 변수 X 가 독립 변수로 간주되고 열 변수 Y 가 종속 변수로 간주됨을 나타냅니다 .
- 마찬가지로 R|C 는 열 변수 Y 가 독립 변수로 간주되고 행 변수 X 가 종속 변수로 간주됨을 나타냅니다 .

Somers D 는 독립 변수에 동일 값을 가진 쌍이 있을 때만 동일 값에 대한 수정을 사용한다는 점에서 Tau-b 와 다릅니다 . -1 ~ 1 범위의 값을 취합니다 .

람다 비대칭 C|R 의 경우와 R|C 의 경우에 차이가 있습니다 .

- C|R 의 경우 , 행 변수 X 에 대해 알 경우 열 변수 Y 를 예측하는 데 있어서 가능한 향상 정도로 해석됩니다 .
- R|C 의 경우 , 열 변수 Y 에 대해 알 경우 행 변수 X 를 예측하는 데 있어서 가능한 향상 정도로 해석됩니다 .

0 ~ 1 범위의 값을 취합니다 .

람다 대칭 대략적으로 두 람다 비대칭 측도의 평균으로 해석됩니다 . 0 ~ 1 범위의 값을 취합니다 .

불확도 계수

- C|R 의 경우 , 행 변수 X 로 설명되는 열 변수 Y 의 불확실성 비율입니다 .
- R|C 의 경우 , 열 변수 Y 로 설명되는 행 변수 X 의 불확실성 비율로 해석됩니다 .

0 ~ 1 범위의 값을 취합니다 .

불확도 계수 대칭 두 불확도 계수 측도의 대칭 버전입니다 . 0 ~ 1 범위의 값을 취합니다 .

참고 :

- 각 통계량은 표준 오차 및 신뢰 구간과 함께 표시됩니다 .
- 감마 , Kendall Tau-b, Stuart Tau-c 및 Somers D 는 X 가 증가함에 따라 변수 Y 가 증가하는 경향이 있는지 여부를 평가하는 순서형 연관성 측도입니다 . 평가에 따라 관측값 쌍이 부합 / 비부합 쌍으로 분류됩니다 . 관측값의 X 값이 커질수록 Y 값도 커지는 쌍은 부합 쌍입니다 . 관측값의 X 값이 커질수록 Y 값이 작아지는 쌍은 비부합 쌍입니다 . 이러한 측도는 두 변수가 모두 순서형일 경우에만 적합합니다 .
- 람다 및 불확도 측도는 순서형 및 명목형 변수에 적합합니다 .

연관성 측도 통계량에 대한 계산 상세 정보는 SAS Institute Inc. 가이드 (2020b) 의 "FREQ Procedure" 장에서 확인하십시오 . 다음 참조 자료에도 추가 정보가 포함되어 있습니다 .

- Brown and Benedetti(1977)
- Goodman and Kruskal(1979)
- Kendall and Stuart(1979)
- Snedecor and Cochran(1980)
- Somers(1962)

Cochran Armitage 추세 검정

참고 : 자세한 내용은 "Cochran Armitage 추세 검정의 예 "(260 페이지) 에서 확인하십시오 .

이 Cochran Armitage 추세 검정에서는 단일 변수의 수준 간에 이항 비율의 추세를 검정합니다. 이 검정은 한 변수의 수준이 두 개이고 다른 변수는 순서형인 경우에만 적합합니다. 2 수준 변수

는 반응을 나타내고 다른 변수는 순서형 수준이 있는 설명 변수를 나타냅니다. 귀무가설은 추세가 없다는 것으로, 설명 변수의 모든 수준에 대해 이항 비율이 동일함을 의미합니다.

이 검정에서 제공되는 검정 통계량과 *p* 값은 근사값입니다. 정확한 추세 검정도 사용할 수 있습니다. 자세한 내용은 "정확 검정"(250 페이지)에서 확인하십시오.

정확 검정

JMP PRO 분할 분석 플랫폼에서는 다음과 같은 정확 검정을 사용할 수 있습니다.

Fisher 정확 검정 *rxc* 테이블에 대해 Fisher 정확 검정을 수행합니다. 이 검정은 두 변수 간의 연관성에 대한 검정입니다. Fisher 정확 검정에서는 행 및 열 합계가 고정되어 있다고 가정하고 초기하 분포를 사용하여 확률을 계산합니다.

이 검정은 대표본 분포 가정에 의존하지 않습니다. 따라서 표본 크기가 작은 경우나 희소 테이블의 경우처럼 가능도비와 Pearson 검정의 신뢰도가 떨어지는 경우에 적합합니다.

이 보고서에는 다음과 같은 정보가 포함됩니다.

테이블 확률 (P) 관측된 테이블의 확률을 제공합니다. 이 값은 검정의 *p* 값이 아닙니다.

양측 확률 ≤ P 양측 검정의 *p* 값을 제공합니다.

2 x 2 테이블의 경우 한 행 또는 열에 포함된 값이 모두 0 인 경우 (이 경우 검정을 계산할 수 없음) 만 제외하고 자동으로 Fisher 정확 검정이 수행됩니다. 자세한 내용은 " 검정 "(243 페이지) 에서 확인하십시오.

정확 Cochran Armitage 추세 검정 정확한 Cochran Armitage 추세 검정을 수행합니다. 이 검정은 한 변수의 수준이 두 개인 경우에만 사용할 수 있습니다. 추세 검정에 대한 자세한 내용은 "Cochran Armitage 추세 검정 "(249 페이지) 에서 확인하십시오.

정확 합치도 통계량 변수 간의 합치도를 검정하기 위한 정확 검정을 수행합니다. 이는 카파 통계량에 대한 정확 검정입니다. 이 검정은 두 변수의 수준이 동일한 경우에만 사용할 수 있습니다. 합치도 검정에 대한 자세한 내용은 " 합치도 통계량 "(247 페이지) 에서 확인하십시오.

참고 : 전체 표본 크기가 32767 보다 크고 분할표가 2x2 보다 큰 경우 정확 검정 옵션을 사용할 수 없습니다.

분할 분석 플랫폼의 추가 예

- " 비율에 대한 평균 분석의 예 "
- " 대응 분석의 예 "
- "Cochran Mantel Haenszel 검정의 예 "
- " 합치도 통계량 옵션의 예 "
- " 상대 위험도 옵션의 예 "

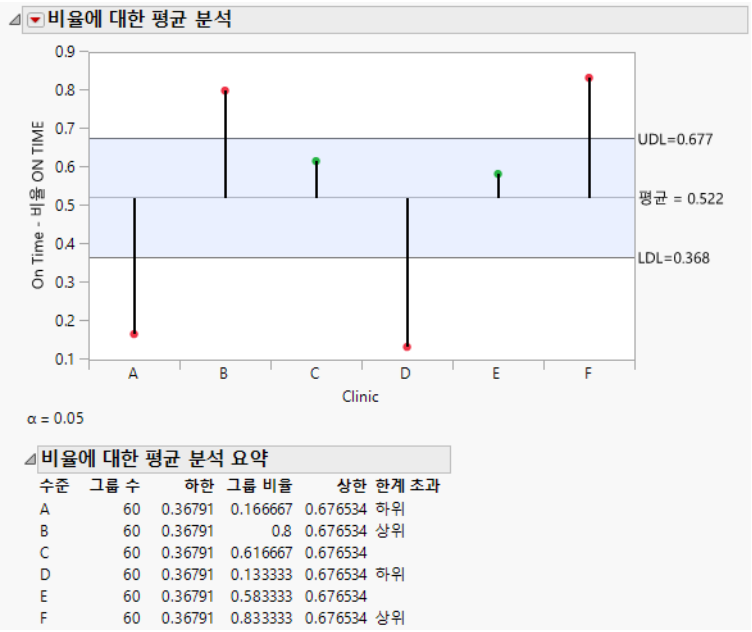
- " 비율에 대한 2 표본 검정의 예 "
- " 연관성 측도 옵션의 예 "
- "Cochran Armitage 추세 검정의 예 "

비율에 대한 평균 분석의 예

이 예에서는 예약 시간 정시에 내원한 환자의 비율을 검토하는 방법을 보여 줍니다 . Office Visits.jmp 샘플 데이터 테이블에는 한 지역에 있는 6 개 병원의 진료 예약 시간 준수 현황이 기록되어 있습니다 . 6 개 병원 각각에서 1 주간의 기록 중 60 건의 예약이 무작위로 선택되었습니다 . 정시 내원으로 간주되기 위해서는 환자가 예약 시간으로부터 5 분 이내에 진료실로 안내되어야 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Office Visits.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. On Time 을 선택하고 **Y, 반응**을 클릭합니다 .
4. Clinic 을 선택하고 **X, 요인**을 클릭합니다 .
5. Frequency 를 선택하고 **빈도**를 클릭합니다 .
6. **확인**을 클릭합니다 .
7. "On Time 대 Clinic 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **비율에 대한 평균 분석**을 선택합니다 .
8. " 비율에 대한 평균 분석 " 옆의 빨간색 삼각형을 클릭하고 **요약 보고서 표시** 및 **비율에 대한 반응 수준 전환**을 선택합니다 .

그림 7.9 비율에 대한 평균 분석의 예



비율에 대한 평균 분석 요약

수준	그룹 수	하한	그룹 비율	상한	한계 초과
A	60	0.36791	0.166667	0.676534	하위
B	60	0.36791	0.8	0.676534	상위
C	60	0.36791	0.616667	0.676534	
D	60	0.36791	0.133333	0.676534	하위
E	60	0.36791	0.583333	0.676534	
F	60	0.36791	0.833333	0.676534	상위

그림 7.9 에서는 각 병원의 예약 시간 정시에 내원한 환자의 비율을 보여 줍니다 . 그림 7.9 에서는 다음을 확인할 수 있습니다 .

- 정시 내원 비율은 F 병원의 경우가 가장 높고 그 다음은 B 병원입니다 .
- D 병원은 정시 내원 비율이 가장 낮고 그 다음은 A 병원입니다 .
- E 병원과 C 병원은 평균에 가까우며 결정 한계를 초과하지 않습니다 .

대응 분석의 예

이 예에서는 Cheese.jmp 샘플 데이터 테이블을 사용합니다 . 이 테이블에는 McCullagh and Nelder 연구 자료 (1989) 에서 보고된 Newell 치즈의 시식 평가 데이터가 포함되어 있습니다 . 이 평가에서는 네 가지 치즈 첨가물에 대해 9 개 이상의 반응 수준 수를 기록했습니다 .

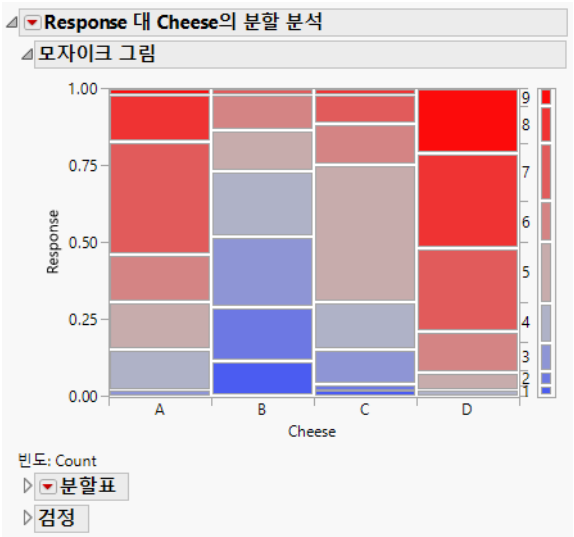
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cheese.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .

   Response 값의 범위는 1 에서 9 까지이며 , 이 중 1 은 호감도가 가장 낮은 것이고 9 는 가장 높은 것입니다 .
4. Cheese 를 선택하고 **X, 요인**을 클릭합니다 .

   A, B, C 및 D 는 4 가지 치즈 첨가물을 나타냅니다 .
5. Count 를 선택하고 **빈도**를 클릭합니다 .

6. **확인**을 클릭합니다 .

그림 7.10 치즈 데이터에 대한 모자이크 그림



이 모자이크 그림에서는 분포가 서로 비슷하지 않음을 알 수 있습니다. 하지만 9 개의 수준에 대한 모자이크 그림을 제대로 파악하기는 어렵습니다. 대응 분석을 수행하면 이러한 경우의 관계를 쉽게 정의할 수 있습니다.

7. 대응 분석 그림을 표시하려면 "Response 대 Cheese 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **대응 분석**을 선택합니다 .

그림 7.11 대응 분석 그림의 예

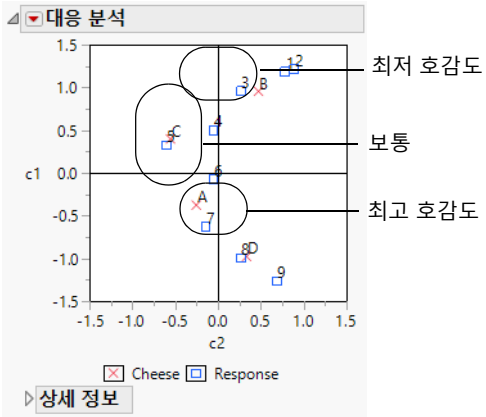
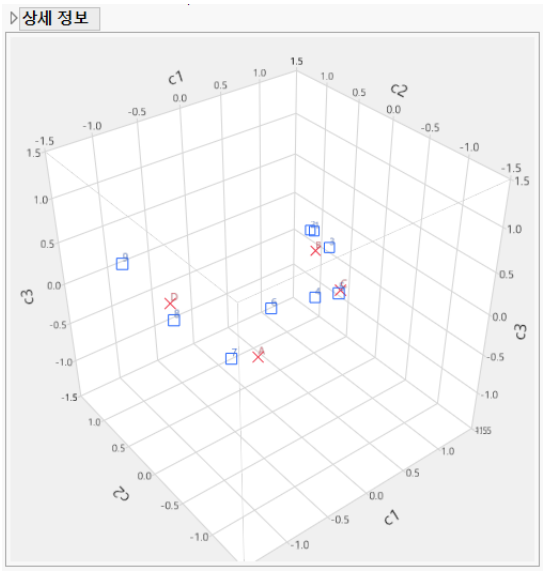


그림 7.11 에서는 대응 분석을 그래프로 보여 줍니다 . 이 그림의 축에는 "c1" 및 "c2" 라벨이 지정되어 있습니다 . 다음을 확인할 수 있습니다 .

- c1 은 일반적인 만족도와 일치하는 것으로 보입니다 . c1 축의 치즈는 맨 위의 최저 호감도에서 맨 아래의 최고 호감도로 변화합니다 .
- 치즈 D 는 반응이 8 및 9 로 , 호감도가 가장 높습니다 .
- 치즈 B 는 반응이 1, 2 및 3 으로 , 호감도가 가장 낮습니다 .
- 치즈 C 와 A 는 반응이 4,5,6 및 7 로 보통 호감도를 보입니다 .

8. " 대응 분석 " 옆의 빨간색 삼각형을 클릭하고 **3D 대응 분석**을 선택합니다 .

그림 7.12 3D 산점도의 예



다음을 확인할 수 있습니다 .

- c1 축을 보면 반응 1 부터 5 까지는 0 의 오른쪽 (양수) 에 나타납니다 . 반응 6 부터 9 까지는 0 의 왼쪽 (음수) 에 나타납니다 .
- c2 축을 보면 A 와 C 가 0 의 오른쪽 (양수) 에 나타납니다 . B 와 D 는 0 의 왼쪽 (음수) 에 나타납니다 .
- c1 은 일반적인 만족도 (최저 호감도부터 최고 호감도까지) 와 일치한다고 결론 내릴 수 있습니다 .

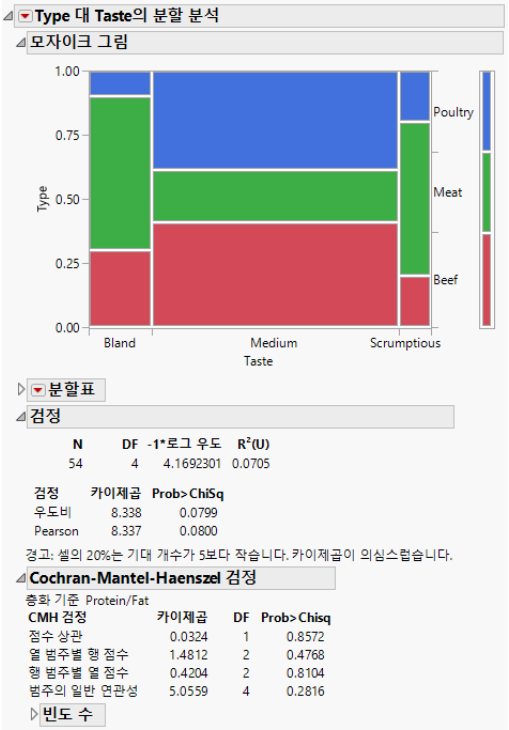
Cochran Mantel Haenszel 검정의 예

이 예에서는 핫도그 유형과 맛 사이의 관계를 검토합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hot Dogs.jmp 를 엽니다 .

2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Type 을 선택하고 **Y, 반응**을 클릭합니다 .
4. Taste 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Type 대 Taste 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **Cochran Mantel Haenszel** 을 선택합니다 .
7. 그룹화 변수로 Protein/Fat 을 선택하고 **확인**을 클릭합니다 .

그림 7.13 Cochran-Mantel-Haenszel 검정의 예



다음을 확인할 수 있습니다 .

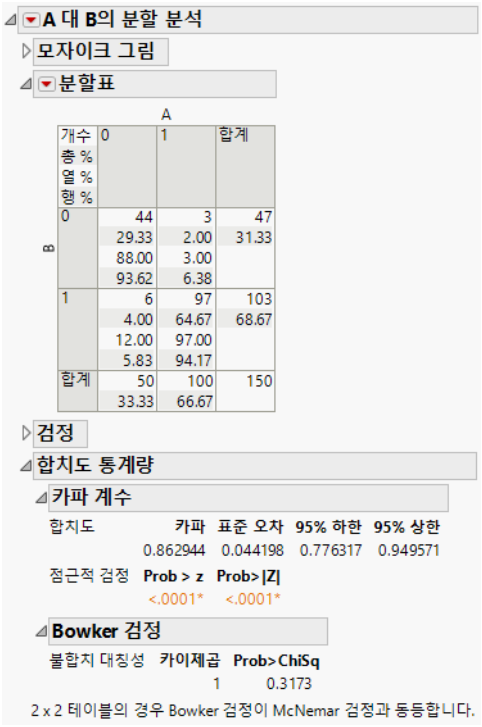
- " 검정 " 보고서에는 약 0.0799 로 다소 유의한 카이제곱 확률이 표시됩니다 . 이는 핫도그의 맛과 유형 간 관계에 어느 정도 유의성이 있음을 나타냅니다 .
- "Cochran-Mantel-Haenszel 검정 " 보고서에는 범주의 일반 연관성에 대한 p 값이 0.2816 으로 , 5% 보다 훨씬 큰 것으로 표시됩니다 .

합치도 통계량 옵션의 예

이 예에서는 두 평가자 사이의 관계를 검토합니다 . 이 데이터는 세 명의 평가자가 50 개의 부품을 각각 세 번씩 평가한 결과를 제공합니다 . 평가자 A 와 B 사이의 관계를 검토하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Attribute Gauge.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. A 를 선택하고 **Y, 반응**을 클릭합니다 .
4. B 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "A 대 B 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **합치도 통계량**을 선택합니다 .

그림 7.14 합치도 통계량 보고서의 예



합치도 통계량이 1 에 가까운 0.86 으로 높게 나타나고 p 값이 <.0001 로 작게 나타납니다. 이는 분할표를 대각선 방향으로 보면 알 수 있는 높은 합치도를 더욱 뒷받침합니다. 평가자 간의 합치도는 두 평가자가 모두 등급 0 을 주거나 등급 1 을 준 경우에 나타납니다.

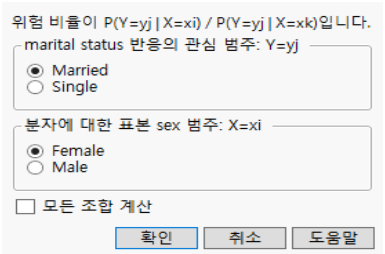
상대 위험도 옵션의 예

이 예에서는 설문조사 참가자의 상대적 기혼 / 비혼 확률을 검토합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. marital status 를 선택하고 **Y, 반응**을 클릭합니다 .
4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "marital status 대 sex 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **상대 위험도**를 선택합니다 .

   " 상대 위험도 범주 선택 " 창이 나타납니다 .

그림 7.15 상대 위험도 범주 선택 창



" 상대 위험도 범주 선택 " 창에서 다음 사항을 알 수 있습니다 .

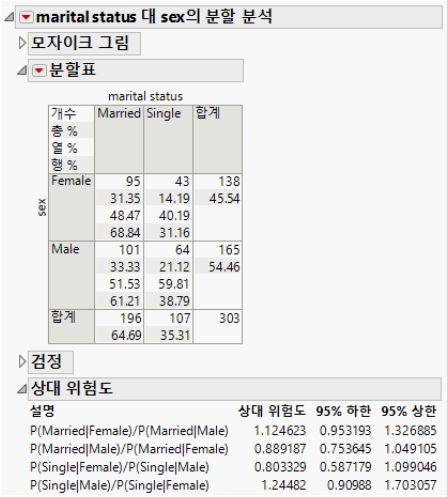
- 단일 반응 및 요인 조합에만 관심이 있는 경우 여기에서 해당 조합을 선택할 수 있습니다 . 예를 들어 그림 7.15 의 창에서 **확인**을 클릭할 경우의 계산은 다음과 같습니다 .

$$\frac{P(Y = \text{Married} | X = \text{Female})}{P(Y = \text{Married} | X = \text{Male})}$$

- 모든 $(2 \times 2 = 4)$ 반응 및 요인 수준 조합에 대해 위험 비율을 계산하려면 **모든 조합 계산** 체크박스를 선택합니다 (그림 7.16).

7. **모든 조합 계산** 체크박스를 선택하여 모든 조합을 요청합니다 . 다른 항목은 모두 기본 선택 상태대로 둡니다 .

그림 7.16 위험 비율 보고서의 예



marital status 대 sex의 분할 분석

모자이크 그림

분할표

marital status

	개수 총 % 열 % 행 %	Married	Single	합계
sex	Female	95 31.35 48.47 68.84	43 14.19 40.19 31.16	138 45.54
	Male	101 33.33 51.53 61.21	64 21.12 59.81 38.79	165 54.46
	합계	196 64.69	107 35.31	303

검정

상대 위험도

설명	상대 위험도	95% 하한	95% 상한
P(Married|Female)/P(Married|Male)	1.124623	0.953193	1.326885
P(Married|Male)/P(Married|Female)	0.889187	0.753645	1.049105
P(Single|Female)/P(Single|Male)	0.803329	0.587179	1.099046
P(Single|Male)/P(Single|Female)	1.24482	0.90988	1.703057

상대 위험도는 다음과 같이 계산됩니다 .

1. " 상대 위험도 " 보고서의 첫 번째 항목 "P(Married | Female)/P(Married | Male)" 을 검토합니다 .
2. 해당 확률은 분할표에서 확인할 수 있습니다 . 이 확률은 테이블의 행별로 달라지는 두 개의 성별 수준을 기반으로 계산되므로 이를 판독하려면 다음과 같은 " 행 %" 를 사용합니다 .

   P(Married | Female) = 0.6884

   P(Married | Male) = 0.6121

   따라서 상대 위험도는 다음과 같이 계산됩니다 .

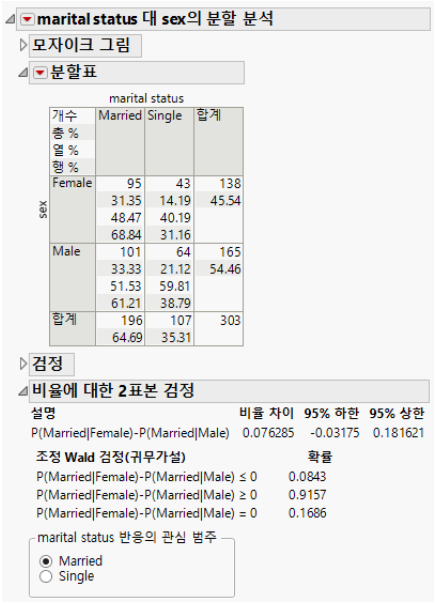
   $$P(\text{Married} \mid \text{Female})/P(\text{Married} \mid \text{Male}) = \frac{0.6884}{0.6121} = 1.1247$$

비율에 대한 2 표본 검정의 예

이 예에서는 남성과 여성의 기혼 확률을 검토합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. marital status 를 선택하고 **Y, 반응**을 클릭합니다 .
4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "marital status 대 sex 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **비율에 대한 2 표본 검정** 을 선택합니다 .

그림 7.17 비율에 대한 2 표본 검정 보고서의 예


marital status 대 sex의 분할 분석

모자이크 그림

분할표

marital status

개수 총 % 열 % 행 %	Married	Single	합계
Female	95 31.35 48.47 68.84	43 14.19 40.19 31.16	138 45.54
Male	101 33.33 51.53 61.21	64 21.12 59.81 38.79	165 54.46
합계	196 64.69	107 35.31	303

sex

검정

비율에 대한 2표본 검정

설명	비율 차이	95% 하한	95% 상한
P(Married|Female)-P(Married|Male)	0.076285	-0.03175	0.181621

조정 Wald 검정(귀무가설)	확률
P(Married|Female)-P(Married|Male) ≤ 0	0.0843
P(Married|Female)-P(Married|Male) ≥ 0	0.9157
P(Married|Female)-P(Married|Male) = 0	0.1686

marital status 반응의 관심 범주

Married

Single

이 예에서는 여성과 남성의 기혼 확률을 비교하려고 합니다 . " 분할표 " 의 " 행 %" 를 참조하여 다음을 구합니다 .

P(Married | Female) = 0.6884

P(Married | Male) = 0.6121

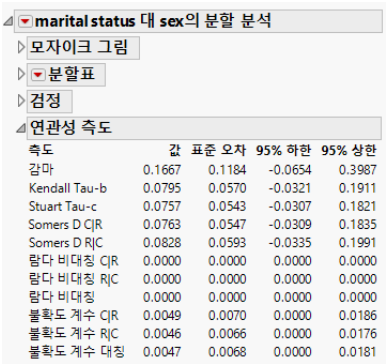
이 두 수치의 차이인 0.0763 이 보고서에 표시된 "비율 차이" 입니다. 양측 신뢰 구간은 [-0.03175, 0.181621] 입니다. 조정 Wald 방법으로 구한 이 신뢰 구간의 p 값은 0.1686 으로, Pearson 카이제곱 검정으로 구한 p 값 (0.1665) 에 가깝습니다. 일반적으로 두 비율의 차이를 검정하는 데는 조정 Wald 검정보다 Pearson 카이제곱 검정이 더 널리 사용됩니다.

연관성 측도 옵션의 예

이 예에서는 성별과 기혼 확률의 연관성을 검토합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. marital status 를 선택하고 **Y, 반응**을 클릭합니다 .
4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "marital status 대 sex 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **연관성 측도**를 선택합니다 .

그림 7.18 연관성 측도 보고서의 예



marital status 대 sex의 분할 분석

모자이크 그림

분할표

검정

연관성 측도

측도	값	표준 오차	95% 하한	95% 상한
감마	0.1667	0.1184	-0.0654	0.3987
Kendall Tau-b	0.0795	0.0570	-0.0321	0.1911
Stuart Tau-c	0.0757	0.0543	-0.0307	0.1821
Somers D C|R	0.0763	0.0547	-0.0309	0.1835
Somers D R|C	0.0828	0.0593	-0.0335	0.1991
람다 비대칭 C|R	0.0000	0.0000	0.0000	0.0000
람다 비대칭 R|C	0.0000	0.0000	0.0000	0.0000
람다 비대칭	0.0000	0.0000	0.0000	0.0000
불확도 계수 C|R	0.0049	0.0070	0.0000	0.0186
불확도 계수 R|C	0.0046	0.0066	0.0000	0.0176
불확도 계수 대칭	0.0047	0.0068	0.0000	0.0181

검토하려는 변수("sex" 및 "marital status")가 명목형이므로 람다 및 불확도 측도를 사용합니다. 이러한 측도가 모두 작으므로 성별과 결혼 여부 간에는 약한 연관성이 있는 것으로 보입니다.

Cochran Armitage 추세 검정의 예

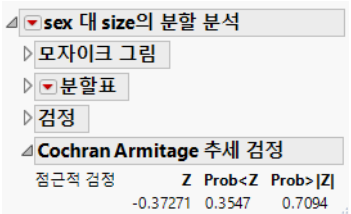
이 예에서는 구매한 차량의 크기와 남성과 여성의 비율 사이에 관계가 있는지 여부를 조사합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .

이 검정을 위해 다음과 같이 "size"를 순서형 변수로 변경하십시오.

2. "열" 패널에서 "size" 옆의 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **순서형**을 선택합니다.
3. **분석 > X 로 Y 적합**을 선택합니다 .
4. sex 를 선택하고 **Y, 반응**을 클릭합니다 .
5. size 를 선택하고 **X, 요인**을 클릭합니다 .
6. **확인**을 클릭합니다 .
7. "sex 대 size 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **Cochran Armitage 추세 검정**을 선택합니다 .

그림 7.19 Cochran Armitage 추세 검정 보고서의 예



sex 대 size의 분할 분석

모자이크 그림

분할표

검정

Cochran Armitage 추세 검정

점근적 검정	Z	Prob<Z	Prob>|Z|
	-0.37271	0.3547	0.7094

양측 p 값(0.7094)이 크게 나타납니다. 따라서 서로 다른 크기의 자동차를 구매하는 남성과 여성의 비율에 관계가 있다고 결론 내릴 수 없습니다.

분할 분석 플랫폼에 대한 통계 상세 정보

- "합치도 통계량"
- "승산비 옵션"
- "검정 보고서"
- "대응 분석의 상세 정보 보고서"

합치도 통계량

두 개의 반응 변수를 n 개의 개체에 대한 두 개의 독립적 등급으로 볼 경우 평가자의 평가가 완전히 합치되면 카파 계수가 +1 이 됩니다. 관측된 합치도가 우연에 의한 합치 확률을 초과하면 카파 계수가 양수가 되며 그 크기는 합치 강도를 반영합니다. 실제로는 흔치 않지만 관측된 합치도가 우연에 의한 합치 확률보다 작을 경우에는 카파가 음수가 됩니다. 카파의 최소값은 주변 비율에 따라 -1 에서 0 사이입니다.

카파 계수는 다음과 같이 계산됩니다.

$$\hat{\kappa} = \frac{P_0 - P_c}{1 - P_c} \text{ 단, } P_0 = \sum_i p_{ii} \text{ 및 } P_c = \sum_i p_{i+} p_{+i}$$

p_{ij} 는 (i, j) 번째 셀에 있는 개체의 비율이므로 $\sum_i \sum_j p_{ij} = 1$ 입니다.

단순 카파 계수의 점근 분산은 다음과 같이 추정됩니다.

$$\text{var} = \frac{A + B - C}{(1 - P_c)^2 n} \text{ 단, } A = \sum_i p_{ii} [1 - (p_{i+} + p_{+i})(1 - \hat{\kappa})]^2, \; B = (1 - \hat{\kappa})^2 \sum_{i \neq j} \sum p_{ij} (p_{+i} + p_{j+})^2 \text{ 및}$$

$$C = [\hat{\kappa} - P_c(1 - \hat{\kappa})]^2$$

자세한 내용은 Cohen 연구 자료(1960) 및 Fleiss et al. 연구 자료(1969)에서 확인하십시오.

Bowker 대칭성 검정에서 귀무가설은 2 x 2 테이블의 확률이 대칭성을 만족($p_{ij}=p_{ji}$)한다는 것입니다.

승산비 옵션

승산비는 다음과 같이 계산됩니다 .

$$\frac{p_{11} \times p_{22}}{p_{12} \times p_{21}}$$

여기서 p_{ij}는 2 x 2 테이블의 i번째 행, j번째 열의 개수입니다.

검정 보고서

R²(U)

R²(U) 는 다음과 같이 계산됩니다 .

$$\frac{-\log \text{likelihood for Model}}{-\log \text{likelihood for Corrected Total}}$$

음의 로드 가능도 합계는 전체 표본에서 고정된 반응 비율을 적합시켜 구합니다.

검정

두 개의 카이제곱 검정은 다음과 같이 계산됩니다.

가능도비 카이제곱 검정 통계량은 " 검정 " 테이블의 모형에 대한 음의 로그 가능도의 2 배로 계산됩니다 . 일부 책에서는 이 통계량에 G^2 표기를 사용합니다 . 두 개의 음의 로그 가능도 (**전체 모집단** 반응 확률을 사용한 로그 가능도와 **각 모집단** 반응 비율을 사용한 로그 가능도) 사이의 차이는 다음과 같이 정의됩니다 .

$$G^2 = 2\left[\sum_{ij}(-n_{ij})\ln(p_j) - \sum_{ij} -n_{ij}\ln(p_{ij})\right] \text{ 단 , } p_{ij} = \frac{n_{ij}}{N} \text{ 및 } p_j = \frac{N_j}{N}$$

이 계산식은 간단히 다음과 같이 쓸 수 있습니다 .

$$G^2 = 2\sum_i\sum_j n_{ij}\ln\left(\frac{n_{ij}}{e_{ij}}\right)$$

Pearson 카이제곱 검정 통계량은 관측된 셀 개수와 기대 셀 개수 사이의 차이를 제곱한 후 합해서 계산합니다 . Pearson 카이제곱 검정은 표본 크기가 매우 클 경우 빈도 수는 정규 분포를 따르는 경향이 있다는 특성을 이용합니다 . 이 카이제곱 통계량의 친숙한 형태는 다음과 같이 정의됩니다 .

$$\chi^2 = \sum\frac{(\mathrm{O}-\mathrm{E})^2}{\mathrm{E}}$$

여기서 O는 관측된 셀 개수이고 E는 기대 셀 개수입니다. 합계에는 모든 셀이 포함됩니다. 여기서는 2 x 2 테이블에서 종종 수행되는 연속성 수정이 수행되지 않습니다.

대응 분석의 상세 정보 보고서

다음과 같은 특이값 분해 (SVD) 의 값을 나열합니다 .

$$\mathbf{D}_r^{-0.5}(\mathbf{P} - rc')\mathbf{D}_c^{-0.5} = \mathbf{U}Diag(\Lambda)\mathbf{V}'$$

다음은 각 요소에 대한 설명입니다 .

- **P 는** 개수를 총 빈도로 나눈 값의 행렬입니다 .
- *r* 및 *c* 는 **P** 의 행 및 열 합계입니다 .
- **D 행렬은** *r* 및 *c* 값의 주대각 행렬입니다 .
- Λ 은 상세 정보 보고서에서 보고된 특이값의 열 벡터입니다 .

특이값 분해에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

상세 정보 보고서의 행 좌표 (*rc*) 와 열 좌표 (*cc*) 는 다음과 같이 계산됩니다 .

$$rc = \mathbf{D}_r^{-0.5}\mathbf{U}Diag(\Lambda)$$

$$cc = \mathbf{D}_c^{-0.5}\mathbf{V}Diag(\Lambda)$$

8 장

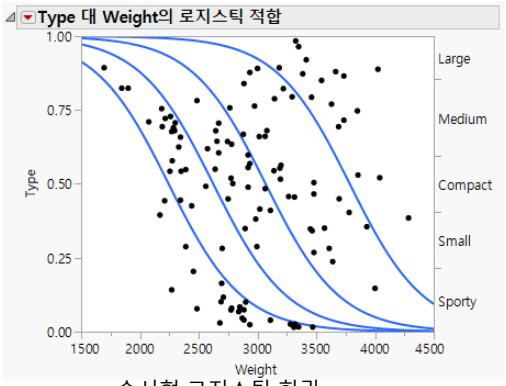
로지스틱 분석

범주형 Y 변수와 연속형 X 변수 간의 관계 분석

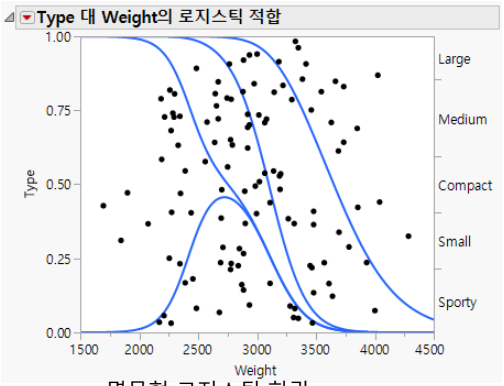
연속형 *x* 예측 변수가 있는 범주형 반응에 로지스틱 회귀 모형을 적합시키려면 로지스틱 플랫폼을 사용합니다. 옵션에는 ROC 곡선, Lift 곡선 및 승산비 추정값이 포함됩니다. 적합된 모형은 각 *x* 값의 확률을 추정합니다. 로지스틱 플랫폼은 X로 Y 적합 플랫폼의 **명목형** 또는 **순서형** 대 **연속형** 분석법입니다. 이 플랫폼에서 명목형 반응과 순서형 반응은 다음과 같이 구분됩니다.

- 명목형 로지스틱 회귀는 반응 간에 확률을 분할하기 위한 일련의 곡선을 추정합니다 .
- 순서형 로지스틱 회귀는 특정 반응보다 작거나 같은 확률을 모델링합니다. 이는 정렬된 범주에 대한 확률을 산출하기 위해 가로로 이동된 단일 로지스틱 곡선을 추정하는 효과가 있습니다 . 이 모형은 덜 복잡하며 순서형 반응에 권장됩니다 .

그림 8.1 로지스틱 회귀의 예



순서형 로지스틱 회귀



명목형 로지스틱 회귀

목차

로지스틱 회귀의 개요 267
명목형 로지스틱 회귀 267
순서형 로지스틱 회귀 267
명목형 로지스틱 회귀의 예 268
로지스틱 플랫폼 시작 269
데이터 구조 270
로지스틱 보고서 271
로지스틱 그림 272
반복 272
전체 모형 검정 272
적합 상세 정보 273
모수 추정값 274
로지스틱 플랫폼 옵션 275
ROC 곡선 276
확률 계산식 저장 277
역추정 예측 277
로지스틱 회귀의 추가 예 277
순서형 로지스틱 회귀의 예 277
로지스틱 그림의 추가 예 279
ROC 곡선의 예 281
십자기호 도구를 사용한 역추정 예측의 예 282
역추정 예측 옵션을 사용한 역추정 예측의 예 283
로지스틱 플랫폼에 대한 통계 상세 정보 284

로지스틱 회귀의 개요

로지스틱 회귀는 오랜 전통을 갖고 있으며 투약 반응 데이터나 구매 선택 데이터를 모델링하는 등 매우 다양하게 적용되고 있습니다. 하지만 많은 통계 입문 과정에서 이 매우 단순한 방법을 다루지 않고 있습니다. 로지스틱 회귀 분야의 교재(Hosmer and Lemeshow 1989)와 범주형 통계 분야의 많은 교재(Agresti 1990)에서 이를 다루고 있습니다. 일부 분석가는 이 방법을 다른 분포 함수, 즉 정규 분포 함수와 함께 사용합니다. 이러한 분석을 **프로빗 분석**이라고 합니다. 일부 분석가는 연속형 변수를 *Y*로 간주하고 범주를 *X*로 간주해서 반대로 작업하는 것을 선호하므로 로지스틱 회귀 대신 판별 분석을 사용합니다. 하지만 판별 분석에서는 연속형 데이터가 고정된 회귀변수가 아니라 정규 분포를 따르는 확률 반응이라고 가정합니다.

단순 로지스틱 회귀는 모형 적합 플랫폼에서 사용되는 범주형 반응에 대한 일반적인 기능을 단순화하고 그래픽 정보를 더 포함한 것입니다. 보다 복잡한 로지스틱 회귀 모형의 예는 **선형 모형 적합**의 에서 확인하십시오.

명목형 로지스틱 회귀

명목형 로지스틱 회귀에서는 반응 수준 중 하나를 선택할 확률을 *x* 요인에 대한 평활 함수로 추정합니다. 적합된 확률은 0에서 1 사이여야 하며, 특정 요인 값에 대한 반응 수준 간의 합계는 1이 되어야 합니다.

로지스틱 확률도에서 세로 축은 확률을 나타냅니다. 반응 수준이 *k*개인 경우 *k* - 1개의 평활 곡선으로 전체 확률(= 1)이 반응 수준 간에 분할됩니다. 로지스틱 회귀의 적합 원칙은 발생하는 반응 사건에 적합된 확률(즉, 최대 가능도)의 음의 자연 로그 합을 최소화하는 것입니다.

순서형 로지스틱 회귀

*Y*가 순서형인 경우 적합에는 수정된 로지스틱 회귀가 사용됩니다. 각 반응 수준과 같은 수준 또는 그 아래 수준의 반응이 나타날 누적 확률은 곡선으로 모델링됩니다. 이 곡선들은 오른쪽 또는 왼쪽으로 이동된다는 점만 제외하고는 각 수준마다 동일합니다.

순서형 로지스틱 모형은 *r* - 1개(여기서 *r*은 반응 수준의 개수)의 누적 로지스틱 비교에 대해 각각 다른 절편과 동일한 기울기를 적합시킵니다. 각 모수 추정값을 개별적으로 검토하고 검정할 수 있지만 이는 거의 필요하지 않습니다.

명목형 모형이라도 추정할 모수가 적어서 순서형 모형이 적합한 경우에는 순서형 모형을 사용하는 것이 좋습니다. 사실상 반응 수준이 수백 개인 경우에는 순서형 반응을 적합시키는 것이 실용적입니다.

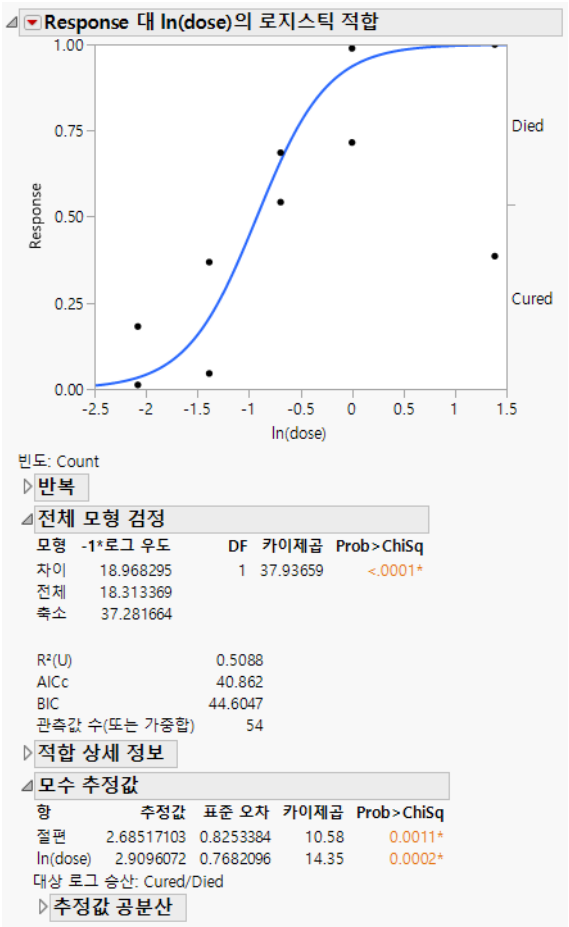
명목형 로지스틱 회귀의 예

명목형 Y 반응과 연속형 X 요인 사이의 관계를 검토하는 방법에 대해 알아봅니다 . 이 예의 데이터는 각각 12 마리의 토끼를 포함하는 5 개 그룹을 대상으로 연쇄상구균 박테리아를 주입한 실험에서 얻은 결과입니다 . 토끼의 신체 조직이 박테리아에 감염된 것으로 확인된 후에는 페니실린을 용량을 달리 해서 투여했습니다 . 투여량의 자연 로그 (ln(dose)) 가 토끼의 치료 여부에 영향을 미치는지 확인하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penicillin.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .
4. ln(dose) 을 선택하고 **X, 요인**을 클릭합니다 .

   빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.
5. " 목표 수준 " 목록에서 Cured 를 선택합니다 .
6. **확인**을 클릭합니다 .

그림 8.2 명목형 로지스틱 보고서의 예



그림에는 적합된 모형이 ln(dose)의 함수로 표시됩니다. 적합된 모형은 예측 치료 확률입니다. *p* 값이 매우 크며, 이는 투여량이 토끼의 치료 여부에 상당한 영향을 미쳤음을 나타냅니다.

팁 : 분석 대상 반응 수준을 변경하려면 시작 창에서 " 목표 수준 " 을 지정하거나 , " 값 순서 " 열 특성을 사용하십시오 .

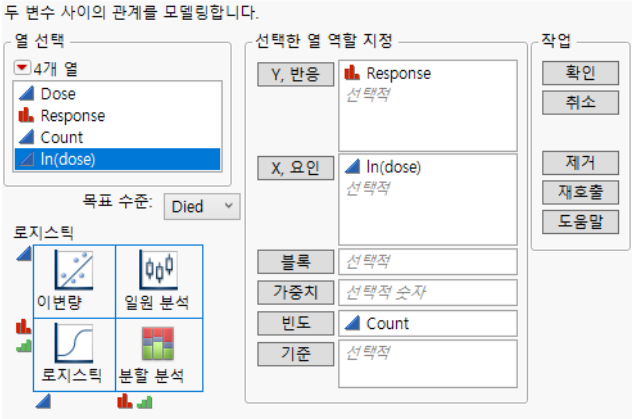
로지스틱 플랫폼 시작

로지스틱 분석을 수행하려면 다음을 수행하십시오 .

1. **분석 > X 로 Y 적합**을 선택합니다 .
2. **Y, 반응**에 명목형 또는 순서형 열을 입력합니다 .

3. **X, 요인**에 연속형 열을 입력합니다 .

그림 8.3 X 로 Y 적합 시작 창



다이어그램 위에 "로지스틱"이라고 표시되면 로지스틱 분석을 수행할 것임을 나타냅니다.

참고 : "JMP 시작하기 " 창에서 로지스틱 분석을 시작할 수도 있습니다 . **보기** > **JMP 시작하기** > **기본 분석** > **로지스틱**을 선택합니다 .

반응이 이항 형식이고 명목형 모델링 유형이면 시작 창에 "목표 수준" 메뉴가 나타납니다. 확률을 모델링할 반응 수준을 지정하려면 이 메뉴를 사용합니다.

팁 : 기본값은 수준 순서를 기준으로 두 수준 중 더 높은 값입니다 . 기본값이 모델링할 수준인지 확인하십시오 .

X로 Y 적합 시작 창에 대한 자세한 내용은 "X로 Y 적합 소개" 장(105 페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

데이터 구조

데이터는 요약되지 않은 데이터나 요약된 데이터로 구성될 수 있습니다 .

요약되지 않은 데이터 각 관측값마다 X 값과 해당하는 Y 값이 들어 있는 행이 하나씩 있습니다 .

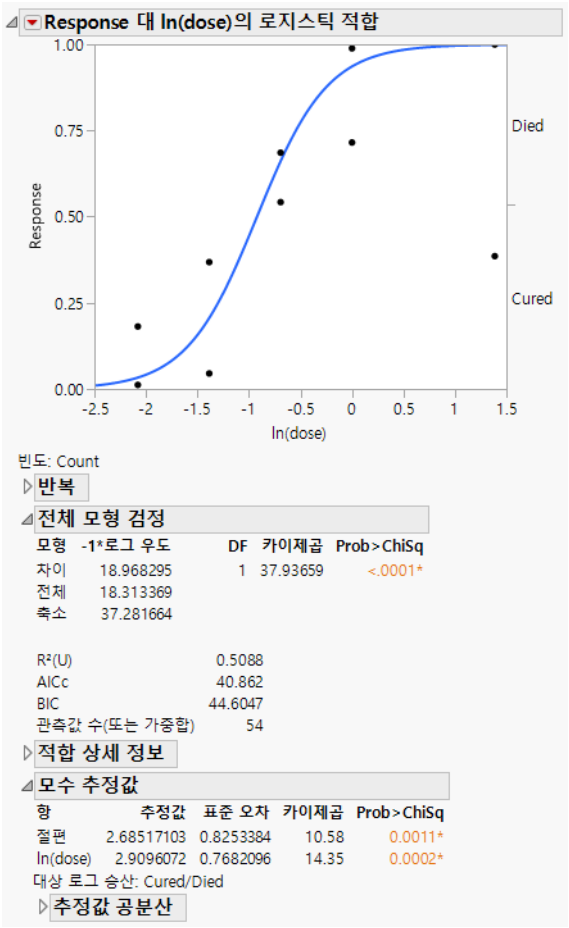
요약된 데이터 각 행이 공통된 X 및 Y 값을 갖는 일련의 관측값을 나타냅니다 . 데이터 테이블에는 각 행의 개수를 제공하는 열이 포함되어야 합니다 . 시작 창에서는 이 열을 " 빈도 " 열로 입력합니다 .

로지스틱 보고서

로지스틱 보고서 창에는 로지스틱 그림, "반복" 보고서, "전체 모형 검정" 보고서, "적합 상세 정보" 보고서 및 "모수 추정값" 보고서가 포함됩니다.

그림 8.4에 표시된 것과 같은 그림을 생성하려면 "명목형 로지스틱 회귀의 예"(268 페이지)에 설명된 방법을 따르십시오.

그림 8.4 로지스틱 보고서의 예



참고 : 빨간색 삼각형 메뉴에서는 초기 보고서 창에 추가할 수 있는 추가 옵션을 제공합니다. 자세한 내용은 " 로지스틱 플랫폼 옵션 "(275 페이지) 에서 확인하십시오 .

로지스틱 그림

로지스틱 확률 그림은 로지스틱 적합을 보여 줍니다. 가로 축의 각 값에서 세로 방향의 확률 척도가 각 반응 범주의 확률로 분할됩니다. 이 확률은 곡선 사이의 세로 거리로 측정되며 모든 반응 범주 확률의 합계는 1이 됩니다.

로지스틱 그림에서 점은 데이터 테이블의 관측값을 나타냅니다. 각 점의 가로 위치는 연속형 요인의 값에 따라 결정됩니다. 각 점의 세로 위치는 반응 범주의 값에 해당하는 곡선 사이에 오도록 무작위로 선택됩니다. 이렇게 점을 지터링하면 점들이 가장 조밀한 곳을 보다 쉽게 확인할 수 있지만 이 경우 세로 위치는 세로 축의 값과 일치하지 않습니다. 고정된 난수 시드값이 사용되므로 세로 위치는 동일한 모형을 여러 번 적합시켜도 달라지지 않습니다.

관련 데이터 테이블의 "열" 패널에서 변수를 클릭하여 축으로 드래그하는 방법으로 그림의 변수를 바꿀 수 있습니다.

참고 : 자세한 내용은 " 로지스틱 그림의 추가 예 "(279 페이지) 에서 확인하십시오 .

반복

"반복" 보고서에서는 각 반복과 로지스틱 모형 수렴 여부를 결정하는 평가 기준을 보여 줍니다. "반복" 보고서는 명목형 로지스틱 회귀의 경우에만 표시됩니다.

전체 모형 검정

"전체 모형 검정" 보고서에서는 모형이 상수 반응 확률보다 더 잘 적합시키는지 여부를 보여 줍니다. 이 보고서는 연속형 반응 모형에 대한 분산 분석 보고서와 유사합니다. 범주형 모형이 데이터를 얼마나 잘 적합시키는지 평가하는 것은 특정 가능도비 카이제곱 검정입니다.

관측된 확률의 음의 자연 로그 합을 음의 로그 가능도(**–1*로그 가능도**)라고 합니다. 범주형 데이터에 대한 음의 로그 가능도는 연속형 데이터에서의 제곱합과 같은 역할을 합니다. 즉, 데이터에 의해 적합된 모형과 동일 확률의 모형에서 얻은 음의 로그 가능도의 차이에 2를 곱하면 카이제곱 통계량이 됩니다. 이 검정 통계량은 x 변수가 반응에 아무 영향도 미치지 않는다는 가설을 검정합니다.

R²(U)(일부 경우에는 R^2로 표기) 값은 0에서 1 사이입니다. R^2 값이 높다는 것은 모형 적합이 적절하다는 것을 나타내며, 범주형 모델에서는 이 값이 높은 경우가 거의 없습니다.

" 전체 모형 검정 " 보고서에는 다음 열이 포함됩니다 .

모형 일부 경우에는 " 소스 " 라고 합니다 .

- **축소** 모형에는 절편만 포함됩니다 .
- **전체** 모형에는 절편뿐만 아니라 모든 효과도 포함됩니다 .
- **차이**는 전체 모형과 축소 모형의 로그 가능도 차이입니다 .

DF 모형과 관련된 자유도를 기록합니다 .

-1* 로그 가능도 표본 내의 변동 (일부 경우에는 **불확도**라고 함) 을 측정합니다 .

전체 (전체 모형) 는 모형 적합 후에 계산된 음의 로그 가능도 (또는 불확도) 입니다 . 적합 프로세스에서는 선형 모형과 로지스틱 반응 함수로 반응 비율을 예측합니다 . 이 값은 적합 프로세스에 의해 최소화됩니다 .

축소 (축소 모형) 는 고정된 기준 비율에 의해 확률이 추정될 경우에 대한 음의 로그 가능도 (또는 불확도) 입니다 . 이는 모형에 효과가 없는 경우의 기준 불확도입니다 .

이 두 개의 음의 로그 가능도의 차이는 모형 적합으로 인한 감소분입니다 . 이 값의 두 배는 가능도비 카이제곱 검정 통계량입니다 .

자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

카이제곱 모형이 전체 표본에서 고정된 반응 비율보다 더 잘 적합된다는 가설에 대한 가능도비 카이제곱 검정입니다 . 이 값은 차이 모형에 대한 **-1* 로그 가능도**의 두 배입니다 . 즉 , 두 개의 음의 로그 가능도 (전체 모집단 반응 확률을 사용한 로그 가능도와 각 모집단 반응 비율을 사용한 로그 가능도) 사이의 차이에 2 를 곱한 것입니다 . 자세한 내용은 " 로지스틱 플랫폼에 대한 통계 상세 정보 "(284 페이지) 에서 확인하십시오 .

Prob>ChiSq 카이제곱 검정의 관측된 유의 확률 (*p* 값이라고도 함) 입니다 . 이 값은 카이제곱 값이 계산된 값보다 클 확률입니다 . 대개 이 확률이 0.05 미만이면 모형이 유의한 것으로 판단됩니다 .

R^2(U) 모형 적합에 기인한 전체 불확실성의 비율로 , **차이**의 음의 로그 가능도 값을 **축소**의 음의 로그 가능도 값으로 나눈 값으로 정의됩니다 . R^2(U) 값이 1 이면 발생하는 사건의 예측 확률이 1 과 같음을 나타냅니다 . 즉 , 예측 확률에 불확실성이 없는 것입니다 . 로짓 모형의 경우 예측 확률은 거의 확실하지 않으므로 R^2(U) 가 작은 경향이 있습니다 . 자세한 내용은 " 로지스틱 플랫폼에 대한 통계 상세 정보 "(284 페이지) 에서 확인하십시오 .

참고 : R^2(U) 를 McFadden 유사 R^2 라고도 합니다 .

AICc 수정된 Akaike 정보 기준입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

BIC 베이지안 정보 기준입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

관측값 일부 경우에는 가중합이라고 합니다 . 계산에 사용된 전체 표본 크기입니다 . **가중치** 변수를 지정한 경우에는 가중치의 합계입니다 .

적합 상세 정보

로지스틱 플랫폼의 " 적합 상세 정보 " 보고서에는 다음 통계량이 포함되어 있습니다 .

측도 다음과 같은 적합 측도가 포함됩니다 .

엔트로피 R^2 적합된 모형과 상수 확률 모형의 로그 가능도를 비교합니다 . R^2(U) 와 동일합니다 . 자세한 내용은 " 로지스틱 플랫폼에 대한 통계 상세 정보 "(284 페이지) 에서 확인하십시오 .

일반화 R² 일반 회귀 모형에 적용할 수 있는 측도입니다. 이 값은 가능도 함수 *L* 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다. 일반화 R² 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R² 로 단순화됩니다. 일반화 R² 를 Nagelkerke 또는 Craig and Uhler R^2 라고도 하는데, 이는 Cox-Snell 유사 R^2 를 정규화한 것입니다. 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오.

평균 -Log p -log(*p*) 의 평균입니다. 여기서 *p* 는 발생한 사건과 연관된 적합 확률입니다.

RASE 평균 제곱근 오차입니다. 여기서 차이는 반응과 *p*(실제로 발생한 사건의 적합 확률) 사이의 차이입니다.

평균 절대 편차 반응과 p(실제로 발생한 사건의 적합 확률) 사이의 차이에 대한 절대값을 평균한 것입니다.

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다.

N 관측값 수입니다.

엔트로피 R² 및 일반화 R² 의 경우 값이 1 에 가까울수록 더 나은 적합을 나타냅니다. 평균 -Log p, RASE, 평균 절대 편차 및 오분류 비율의 경우 값이 작을수록 더 나은 적합을 나타냅니다.

훈련 적합 측도의 값입니다.

정의 적합 측도의 대수 정의입니다.

모수 추정값

명목형 로지스틱 모형은 $k-1$ 개의 로지스틱 비교 각각에 대해 절편 및 기울기 모수를 적합시킵니다. 여기서 *k* 는 반응 수준의 개수입니다. " 모수 추정값 " 보고서에는 이러한 추정값이 나열됩니다. 각 모수 추정값을 개별적으로 검토하고 검정할 수 있지만 이는 거의 필요하지 않습니다.

항 로지스틱 모형의 각 모수를 나열합니다. 반응 변수의 마지막 수준을 제외하고 각 수준별로 요인에 대한 절편 항과 기울기 항이 하나씩 있습니다.

추정값 로지스틱 모형으로 구한 모수 추정값을 나열합니다.

표준 오차 각 모수 추정값의 표준 오차를 나열합니다. 이 값은 각 항을 0 과 비교하는 통계 검정을 계산하는 데 사용됩니다.

카이제곱 각 모수가 0 이라는 가설에 대한 Wald 검정을 나열합니다. Wald 카이제곱은 (추정값 / 표준 오차)^2 으로 계산됩니다.

Prob>ChiSq 카이제곱 검정의 관측된 유의 확률을 나열합니다.

추정값 공분산

모수 추정값의 추정 분산과 모수 추정값 간의 추정 공분산을 보고합니다. 분산 추정값의 제곱근은 **표준 오차** 섹션에 제공된 값과 동일합니다.

로지스틱 플랫폼 옵션

로지스틱 적합의 빨간색 삼각형 메뉴에는 로지스틱 그림에 영향을 미치는 옵션이 포함되어 있습니다 . 다중 Y 반응 변수를 지정한 경우 적합 그룹의 빨간색 삼각형 메뉴가 대신 나타납니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

승산비 " 모수 추정값 " 보고서에 승산비를 추가합니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

이 옵션은 수준이 두 개인 반응에만 사용할 수 있습니다 .

역추정 예측 특정 *y* 값으로부터 *x* 값을 예측합니다 . 자세한 내용은 " 역추정 예측 "(277 페이지) 에서 확인하십시오 .

이 옵션은 2 수준 명목형 반응에만 사용할 수 있습니다 .

로지스틱 그림 로지스틱 그림을 표시하거나 숨깁니다 .

그림 옵션 다음 옵션이 포함되어 있습니다 .

점 표시 로지스틱 그림에 점을 표시하거나 숨깁니다 .

비율 곡선 표시 각 *x* 값마다 점이 여러 개 있는 경우에만 유용합니다 . 이 경우 각 값의 적절한 비율 추정값을 구하고 이 비율을 적합된 로지스틱 곡선과 비교할 수 있습니다 . 퇴화점 (일반적으로 0 또는 1 일 때) 이 너무 많아지지 않도록 JMP 에서는 *x* 값에 점이 세 개 이상 있는 경우에 비율 값만 표시합니다 .

선 색상 그림의 곡선 색상을 선택할 수 있습니다 .

ROC 곡선 ROC(Receiver Operating Characteristic) 곡선을 생성합니다 . ROC 곡선은 각 *x* 값에 대한 민감도 대 (1 - 특이도) 의 그림입니다 . 자세한 내용은 "ROC 곡선 "(276 페이지) 에서 확인하십시오 .

Lift 곡선 모형에 대한 Lift 곡선을 생성합니다 . Lift 곡선은 ROC 곡선과 동일한 정보를 보여 주되 , 시작점에서 정렬의 풍부도를 극화하는 방법을 사용합니다 . 세로 축은 모집단 중 해당 부분이 해당 반응 수준 전체의 비율에 비해 선택된 반응 수준에 얼마나 풍부하게 포함되었는지를 보여 줍니다 . Lift 곡선에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

확률 계산식 저장 계산식을 포함하는 새 데이터 테이블 열을 생성합니다 . 자세한 내용은 " 확률 계산식 저장 "(277 페이지) 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

ROC 곡선

이항 결과에 대해 단순 로지스틱 회귀를 수행할 경우 해당 분석에 대한 ROC 곡선의 플랫폼 옵션을 사용할 수 있습니다. ROC 곡선 옵션을 선택한 후에는 positive(양성) 반응을 지정해야 합니다.

참고 : 자세한 내용은 "ROC 곡선의 예 "(281 페이지) 에서 확인하십시오 .

예를 들어 진단 측도인 *x* 값이 있고 다음을 나타내는 *x* 임계값을 결정하려고 한다고 가정해 보겠습니다 .

- *x* 값이 임계보다 크면 조건이 존재하는 것입니다 .
- *x* 값이 임계보다 작으면 조건이 존재하지 않는 것입니다 .

예를 들어 암 유형을 예측하기 위한 진단 검사로 혈액 성분 수준을 측정할 수 있습니다. 임계가 다양하고 그로 인해 False Positive(가양성) 및 False Negative(가음성) 가 늘어나거나 줄어들 수 있기 때문에 진단 검사를 검토합니다. 그런 다음 해당 비율을 그림으로 표시합니다. 이상적인 상태는 True Negative(진음성) 와 True Positive(진양성) 를 가장 잘 구분하는 *x* 기준 값의 범위가 매우 좁은 것입니다. ROC(Receiver Operating Characteristic) 곡선은 이 전환이 얼마나 빠르게 일어나는지를 보여 줍니다. ROC 곡선의 목적은 곡선 아래 면적을 최대화하는 진단 기준을 얻는 것입니다.

두 가지 표준 정의가 의학에서 사용됩니다 .

- 민감도 - 특정 *x* 값 (검사 또는 측정) 이 존재하는 조건을 올바르게 예측할 확률입니다 . 특정 *x* 가 조건의 존재를 잘못 예측할 확률은 1 – 민감도입니다 .
- 특이도 - 검사를 통해 조건이 존재하지 않음을 올바르게 예측할 확률입니다 .

ROC 곡선은 각 *x* 값에 대한 민감도 대 (1 - 특이도) 의 그림입니다. ROC 곡선 아래의 면적은 곡선에 포함된 정보를 요약하는 데 사용되는 일반적인 지표입니다.

검사를 통한 예측이 완벽했다면 전체 비정상 모집단의 아래쪽, 모든 정상 값의 위쪽에 위치하는 하나의 값을 얻게 됩니다. 민감도가 완벽하게 되면 격자 위의 점 (0,1) 을 지납니다. ROC 곡선이 이 이상적인 점에 가까울수록 판별 능력이 높습니다. 예측 능력이 없는 검사의 경우 격자의 대각선과 일치하는 곡선이 생성됩니다 (DeLong et al. 1988).

ROC 곡선은 False Positive 비율과 True Positive 비율의 관계를 그래픽으로 나타낸 것입니다. 관계를 평가하는 표준 방법은 보고서의 그림 아래에 표시된 곡선 아래 면적을 사용하는 것입니다. 그림에는 ROC 곡선에 45 도로 접하는 노란색 선이 그려집니다. 이는 False Negative 및 False Positive 의 비용이 유사하다는 가정하에 적절한 경계 점을 나타냅니다.

확률 계산식 저장

확률 계산식 저장 옵션은 새 데이터 테이블 열을 생성합니다 . 이 데이터 테이블 열에는 다음 항목이 저장됩니다 .

- *x* 요인의 선형 결합 (일반적으로 로짓이라고 함) 계산식
- 반응 수준 확률에 대한 예측 계산식
- 가장 가능성이 높은 반응을 구하는 예측 계산식

역추정 예측

역추정 예측은 예측을 반대로 하는 것입니다. 즉, 특정 *y* 값으로부터 *x* 값을 예측합니다. 하지만 로지스틱 회귀에서는 *y* 값 대신 *Y* 수준 중 하나에 기인한 확률을 사용합니다. 이 기능은 2수준 명목형 반응에만 사용할 수 있습니다.

모형 적합 플랫폼에도 신뢰 한계를 사용하여 역추정 예측을 수행하는 옵션이 있습니다. 역추정 예측에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

참고 : " 십자기호 도구를 사용한 역추정 예측의 예 "(282 페이지) 및 " 역추정 예측 옵션을 사용한 역추정 예측의 예 "(283 페이지) 도 확인하십시오 .

로지스틱 회귀의 추가 예

이 섹션에는 로지스틱 플랫폼의 로지스틱 회귀를 사용하는 추가 예가 포함되어 있습니다 .

- " 순서형 로지스틱 회귀의 예 "
- " 로지스틱 그림의 추가 예 "
- "ROC 곡선의 예 "
- " 십자기호 도구를 사용한 역추정 예측의 예 "
- " 역추정 예측 옵션을 사용한 역추정 예측의 예 "

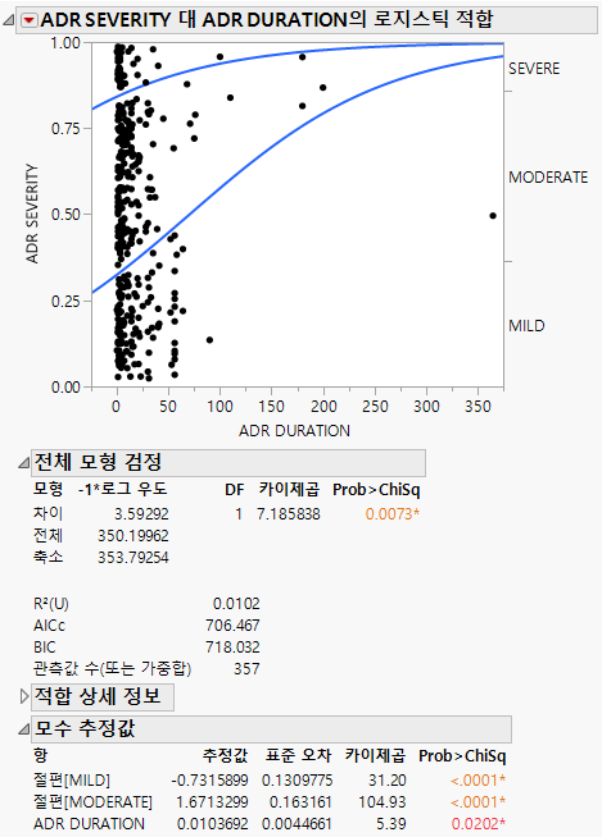
순서형 로지스틱 회귀의 예

이 예에서는 순서형 Y 반응 및 연속형 X 요인 사이의 관계를 검토하는 방법을 보여 줍니다 . 이 예에서는 이상 사례의 심각도를 처치 기간 값의 함수로 모델링하려고 한다고 가정해 보겠습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 AdverseR.jmp 를 엽니다 .
2. ADR SEVERITY 왼쪽의 아이콘을 마우스 오른쪽 버튼으로 클릭하고 모델링 유형을 순서형으로 변경합니다 .
3. **분석 > X 로 Y 적합**을 선택합니다 .

4. ADR SEVERITY 를 선택하고 **Y, 반응**을 클릭합니다 .
5. ADR DURATION 을 선택하고 **X, 요인**을 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 8.5 순서형 로지스틱 보고서의 예



이 보고서의 해석 방법은 명목형 보고서의 경우와 동일합니다. 자세한 내용은 "로지스틱 보고서"(271 페이지)에서 확인하십시오.

그림에서 데이터 표식은 x 좌표에 그려집니다. 여러 개의 데이터 점이 동일한 y 위치에 나타나는 경우에는 점이 지터링됩니다. 즉, 각 점을 더 명확하게 볼 수 있도록 데이터 점 사이에 약간의 간격이 나타납니다.

점이 많은 경우에는 곡선들이 서로 멀어집니다. 점이 거의 없거나 전혀 없으면 곡선이 서로 가까워집니다. 데이터가 곡선을 이러한 방식으로 밀어내는 이유는 최대화되는 기준이 모형으로 적합된 확률의 곱이기 때문입니다. 적합은 확률이 작은 데 기인한 점, 즉 적합 곡선에 의해 밀리는 점을 방지하려고 합니다. 계산 상세 정보에 대한 자세한 내용은 **선형 모형 적합**에서 확인하십시오.

"전체 모형 검정" 보고서와 "모수 추정값" 보고서에 대한 자세한 내용은 "로지스틱 보고서"(271 페이지)에서 확인하십시오. "모수 추정값" 보고서에서 절편 모수는 마지막 반응 수준을 제외한 각 반응 수준에 대해 추정된 값이지만 기울기 모수는 하나만 있습니다. 절편 모수는 반응 수준의 간격을 보여 줍니다. 이들은 항상 단순 증가합니다.

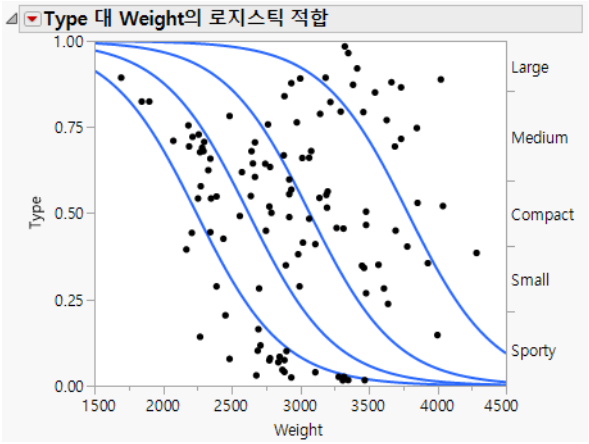
로지스틱 그림의 추가 예

이 예에서는 116 대의 자동차에 대해 중량을 사용하여 자동차 크기 (Type) 를 예측하려고 한다고 가정해 보겠습니다 . 자동차 크기는 작은 것부터 큰 것 순서로 "Sporty", "Small", "Compact", "Medium" 또는 "Large" 중 하나일 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Physical Data.jmp 를 엽니다 .
2. 열 패널에서 Type 왼쪽의 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **순서형**을 선택합니다 .
3. Type 을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다 .
4. " 열 특성 " 메뉴에서 **값 순서**를 선택합니다 .
5. 데이터가 하향식 순서 , 즉 "Sporty", "Small", "Compact", "Medium", "Large" 의 순서로 되어 있는지 확인합니다 .
6. **확인**을 클릭합니다 .
7. **분석 > X 로 Y 적합**을 선택합니다 .
8. Type 을 선택하고 **Y, 반응**을 클릭합니다 .
9. Weight 를 선택하고 **X, 요인**을 클릭합니다 .
10. **확인**을 클릭합니다 .

    보고서 창이 나타납니다 .

그림 8.6 Type 대 Weight 에 대한 로지스틱 그림의 예



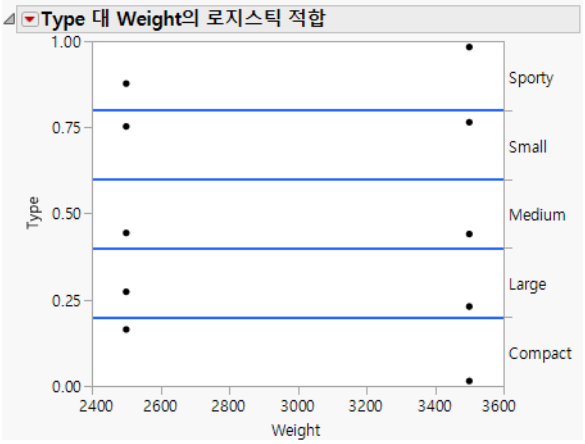
다음 관측값에 유의하십시오 .

- 첫 번째 (맨 아래) 곡선은 주어진 중량의 자동차가 Sporty 일 확률을 나타냅니다 .
- 두 번째 곡선은 자동차가 Small 또는 Sporty 일 확률을 나타냅니다 . 첫 번째 곡선과 두 번째 곡선 사이의 거리만 보면 Small 일 확률에 부합합니다 .
- 예상할 수 있듯이 더 무거운 자동차는 Large 일 가능성이 높습니다 .
- 데이터 표식은 x 좌표에 그려집니다 . y 위치는 해당 행의 반응 범주에 해당하는 범위 내에서 무작위로 지터링됩니다 .

x 변수가 반응에 아무런 영향도 미치지 않는다면 적합선은 수평선이 되고 연속형 요인 범위에서 각 반응의 확률은 상수가 됩니다. 그림 8.7에서는 Weight가 Type을 예측하는 데 유용하지 않은 경우의 로지스틱 그림을 보여 줍니다.

그림 8.7 y 대 x 관계가 없음을 보여 주는 샘플 데이터 테이블 및 로지스틱 그림의 예

	Type	Weight
1	Large	2500
2	Medium	2500
3	Compact	2500
4	Small	2500
5	Sporty	2500
6	Large	3500
7	Medium	3500
8	Compact	3500
9	Small	3500
10	Sporty	3500

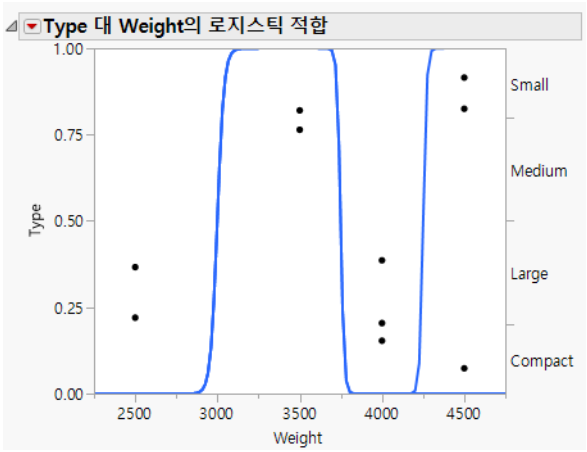


참고 : 그림 8.7 및 그림 8.8 의 그림을 다시 생성하려면 먼저 여기에 표시된 데이터 테이블을 생성한 다음 이 섹션의 처음에 나오는 7~10 단계를 수행해야 합니다 .

요인 값으로 반응이 완전히 예측되는 경우에는 로지스틱 곡선이 거의 수직이 됩니다. 각 요인 수준에서 반응 예측은 거의 확실(확률이 거의 1) 합니다. 그림 8.8에서는 Weight가 Type을 거의 완벽하게 예측하는 경우의 로지스틱 그림을 보여 줍니다.

그림 8.8 거의 완벽한 y 대 x 관계를 보여 주는 샘플 데이터 테이블 및 로지스틱 그림의 예

	Type	Weight
1	Small	2500
2	Small	2500
3	Compact	3500
4	Compact	3500
5	Medium	4000
6	Medium	4000
7	Medium	4000
8	Large	4500
9	Large	4500
10	Large	4500



이 경우에는 모수 추정값이 매우 커지고 회귀 보고서에 **불안정**이라는 라벨이 표시됩니다. 이러한 경우에는 Firth 편향 조정 추정값을 사용하는 일반화 선형 모형 분석법을 사용할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

ROC 곡선의 예

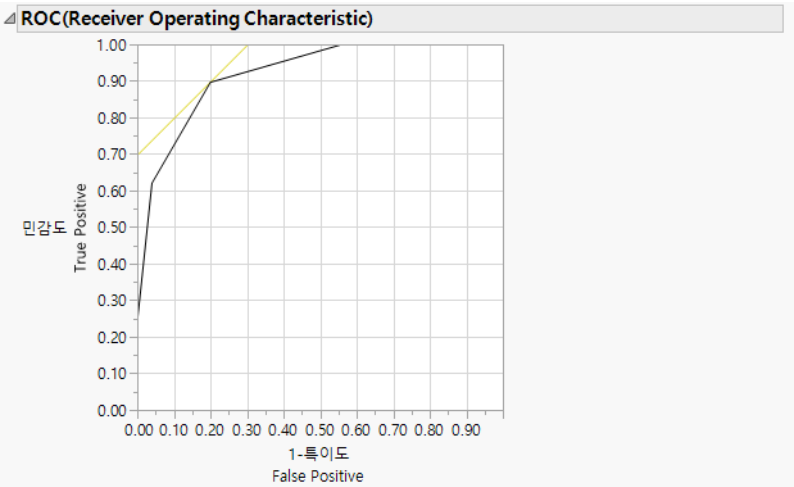
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penicillin.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .
4. ln(dose) 을 선택하고 **X, 요인**을 클릭합니다 .

   빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.
5. " 목표 수준 " 목록에서 Cured 를 선택합니다 .
6. **확인**을 클릭합니다 .
7. "Response 대 ln(dose) 의 로지스틱 적합 " 옆의 빨간색 삼각형을 클릭하고 **ROC 곡선**을 선택합니다 .

참고 : 이 예에서는 명목형 반응의 ROC 곡선을 보여 줍니다. 순서형 ROC 곡선에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

response 대 ln(dose) 예의 결과는 다음과 같습니다. ROC 곡선에는 반응 예측을 위해 위에서 설명한 확률이 표시됩니다. "ROC 테이블"에서 "민감도-(1-특이도)"가 가장 높은 행에는 별표가 표시됩니다.

그림 8.9 ROC 곡선 및 테이블의 예



ROC 테이블

X	확률	1-특이도	민감도	민감도-(1-특이도)		True Pos	True Neg	False Pos	False Neg
.	.	0.0000	0.0000	0.0000		0	25	0	29
1.38629	0.9988	0.0000	0.2414	0.2414		7	25	0	22
0.00000	0.9361	0.0400	0.6207	0.5807		18	24	1	11
-0.69315	0.6611	0.2000	0.8966	0.6966	*	26	20	5	3
-1.38629	0.2061	0.5600	1.0000	0.4400		29	11	14	0
-2.07944	0.0334	1.0000	1.0000	0.0000		29	0	25	0
-2.07944	0.0334	1.0000	1.0000	0.0000		29	0	25	0

이 ROC 곡선은 대각선의 매우 위쪽에 있으므로 모형의 예측 능력이 좋다는 결론을 내릴 수 있습니다.

십자기호 도구를 사용한 역추정 예측의 예

이 예에서는 JMP의 십자기호 도구를 사용하여 역추정 예측의 근사값을 시각적으로 구하는 방법을 보여 줍니다. 페니실린이 투여된 토끼에 대한 연구에서 토끼의 치료 확률이 0.5가 되는 페니실린 투여량을 알아보려고 합니다. 이 경우 0.5에 대한 역추정 예측을 ED50이라고 하며, 이는 생존률 50%에 해당하는 유효 투여량을 나타냅니다.

치료 또는 치사 가능성이 동일한 투여량을 결정합니다 .

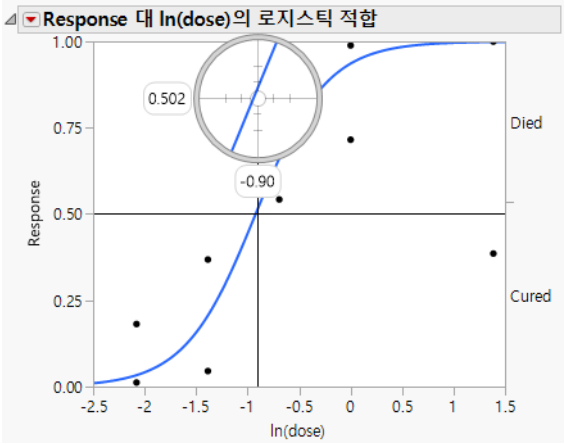
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penicillin.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .
4. ln(dose) 을 선택하고 **X, 요인**을 클릭합니다 .

   빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.

5. " 목표 수준 " 목록에서 Cured 를 선택합니다 .
6. **확인**을 클릭합니다 .
7. 십자기호 도구를 클릭합니다 .
8. 십자기호의 가로 선이 세로 (Response) 확률 축의 약 0.5 지점에 오도록 합니다 .
9. 십자기호의 교차점을 예측선 쪽으로 이동하고 가로 축에 표시되는 ln(dose) 값을 읽습니다 .

이 예에서 ln(dose)이 약 -0.9인 토끼는 치료될 가능성과 죽을 가능성이 동일합니다.

그림 8.10 로지스틱 그림에서 사용되는 십자기호 도구의 예



역추정 예측 옵션을 사용한 역추정 예측의 예

반응 수준이 정확히 두 개뿐인 경우 **역추정 예측** 옵션을 사용하여 정확한 역추정 예측을 요청할 수 있습니다. 하위 반응 범주의 특정 확률에 해당하는 x 값과 해당 x 값의 신뢰 구간을 알 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penicillin.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .
4. ln(dose) 을 선택하고 **X, 요인**을 클릭합니다 .

   빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.
5. " 목표 수준 " 목록에서 Cured 를 선택합니다 .
6. **확인**을 클릭합니다 .
7. "Response 대 ln(dose) 의 로지스틱 적합 " 옆의 빨간색 삼각형을 클릭하고 **역추정 예측**을 선택합니다 (그림 8.11).
8. **신뢰 수준**에 0.95 를 입력합니다 .
9. 신뢰 구간으로 **양측**을 선택합니다 .

10. 관심 있는 반응 확률을 요청합니다. 이 예에서는 0.5 와 0.9 를 입력합니다. 즉, 치료 확률이 0.5 와 0.9 인 ln(dose) 의 값을 요청합니다.
11. **확인**을 클릭합니다.

    역추정 예측 그림이 나타납니다.

그림 8.11 역추정 예측 창

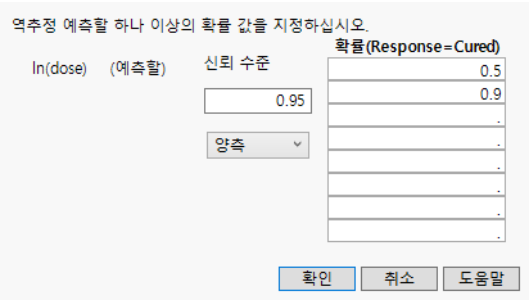
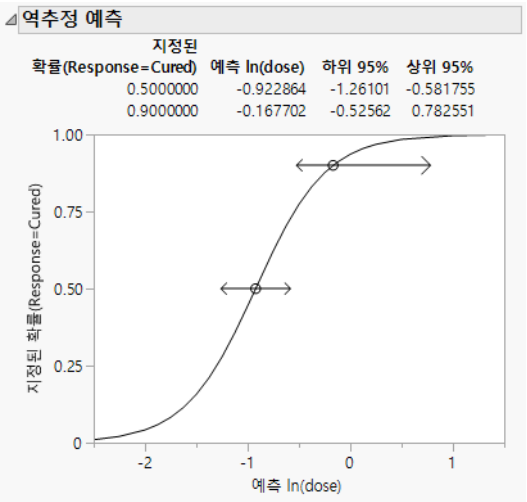


그림 8.12 역추정 예측 그림의 예



x 값의 추정값과 신뢰 구간이 보고서와 확률도에 표시됩니다. 예를 들어 치료 확률이 90%에 이르는 ln(dose)의 값은 -0.526 ~ 0.783으로 추정됩니다.

로지스틱 플랫폼에 대한 통계 상세 정보

이 섹션에는 "전체 모형 검정" 보고서에 대한 통계 상세 정보가 포함되어 있습니다.

카이제곱

카이제곱 통계량은 때때로 G^2 으로 표기되며 다음과 같이 정의됩니다 .

$$G^2 = 2(\sum -\ln p(\text{background}) - \sum -\ln p(\text{모형}))$$

여기서 합계는 모든 셀이 아니라 모든 관측값에 대한 값입니다.

R²(U)

기준 로그 가능도에 대한 이 검정 통계량의 비율을 1 에서 빼서 R^2 를 계산합니다 . R²(U) 를 더 간단하게 계산하는 방법은 다음과 같습니다 .

$$\frac{\text{차이 모형에 대한 음의 로그 가능도}}{\text{축소 모형에 대한 음의 로그 가능도}}$$

이 계산에는 "전체 모형 검정" 보고서의 통계량이 사용됩니다.

참고 : R²(U) 를 McFadden 유사 R² 라고도 합니다 .

9 장

테이블 생성
대화식으로 요약 테이블 생성

테이블 생성 플랫폼을 사용하여 기술 통계량의 요약 테이블이나 피벗 테이블을 대화식으로 생성할 수 있습니다. 테이블 생성 플랫폼은 요약 데이터를 테이블 형식으로 표시하는 쉽고 유연한 방법입니다. 테이블은 데이터 테이블 열을 테이블 생성의 행 또는 열로 할당한 다음 원하는 요약 통계량을 할당하여 작성됩니다.

그림 9.1 테이블 생성 예

	sex											
	F						M					
	age						age					
	12	13	14	15	16	17	12	13	14	15	16	17
	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight
최소값	64	67	81	92	112	116	79	79	92	104	128	134
평균	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
최대값	145	112	142	112	115	116	128	105	119	128	128	172

		sex							
		Female				Male			
		marital status				marital status			
		Married		Single		Married		Single	
		평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
country	size	age	age	age	age	age	age	age	age
American	Large	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	Medium	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	Small	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
European	Large	34.0	7.071	28.0	.	.	.	26.0	.
	Medium	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	Small	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
Japanese	Large	25.0	.	.	.	32.0	.	.	.
	Medium	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	Small	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
country									
American		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
European		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
Japanese		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

		평균		
Type	Size Co	Profits ($M)	Sales ($M)	profit/emp
Computer	big	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
	모두	240.87	5652.02	6159.015
Pharmaceutical	big	894.42	7474.04	17140.70
	medium	698.98	4261.06	24035.11
	small	156.95	1083.75	38337.19
	모두	690.08	5070.25	23546.12
모두	모두	409.32	5433.86	12679.18
Type				
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
모두		409.32	5433.86	12679.18

목차

테이블 생성 플랫폼의 예 289
테이블 생성 플랫폼 시작 293
대화상자 사용 294
통계량 추가 295
테이블 생성 출력 298
분석 열 299
그룹화 열 300
열 및 행 테이블 300
테이블 편집 301
테이블 생성 플랫폼 옵션 302
테스트 빌드 패널 표시 303
마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴 303
테이블 생성 플랫폼의 추가 예 304
예: 여러 테이블 생성 및 내용 재배열 304
예: 여러 열을 단일 테이블에 결합 308
예: 페이지 열 사용 310

테이블 생성 플랫폼의 예

테이블 생성 플랫폼을 사용하여 데이터를 테이블 형식으로 요약하는 방법을 알아봅니다. 이 예에는 남학생과 여학생의 키 측정값이 포함된 데이터가 있습니다. 남학생 및 여학생의 평균 키와 두 성별 모두의 키 집계를 보여 주는 테이블을 생성하려고 합니다.

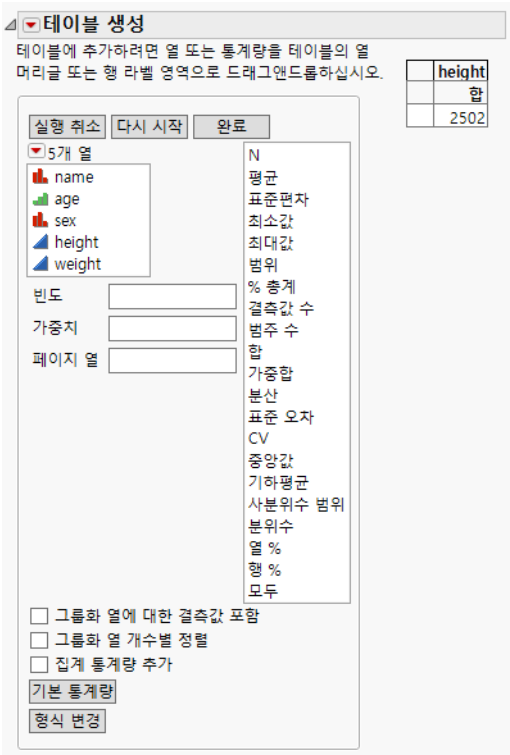
그림 9.2 평균 키를 보여 주는 테이블

	height
sex	평균
F	60.9
M	63.9
모두	62.6

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .

   height 변수를 검토할 것이므로 테이블의 맨 위에 이 변수를 표시하려고 합니다 .
3. height 를 클릭하고 " 열에 대한 놓기 영역 " 으로 드래그합니다 .

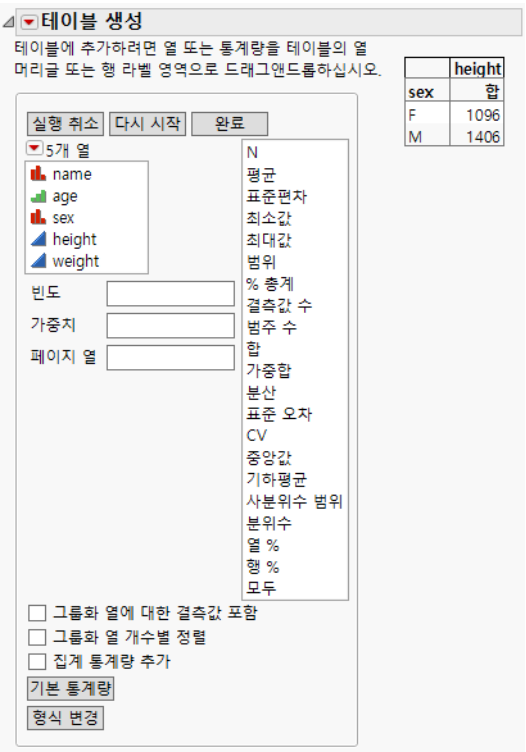
그림 9.3 height 변수 추가



성별별로 통계량을 살펴볼 것이므로 sex 변수를 나란히 표시하려고 합니다.

4. sex 를 클릭하고 숫자 "2502" 옆의 빈 셀로 드래그합니다 .

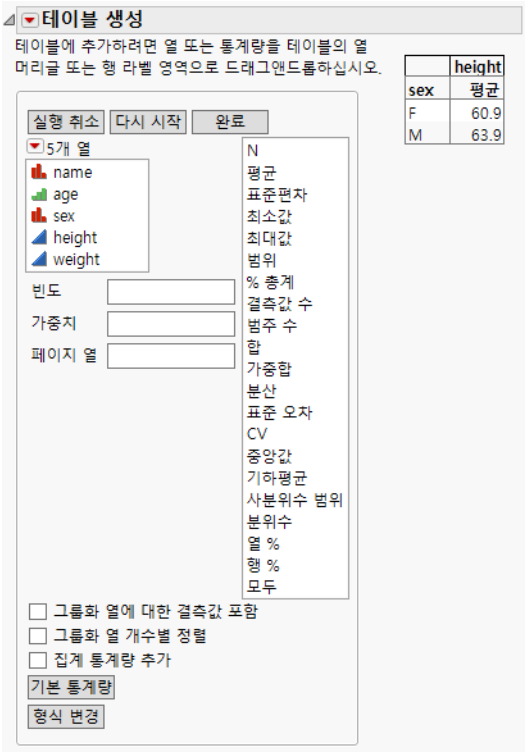
그림 9.4 sex 변수 추가



합 대신 평균을 표시하려고 합니다.

5. **평균**을 클릭하고 **합** 위로 드래그합니다 .

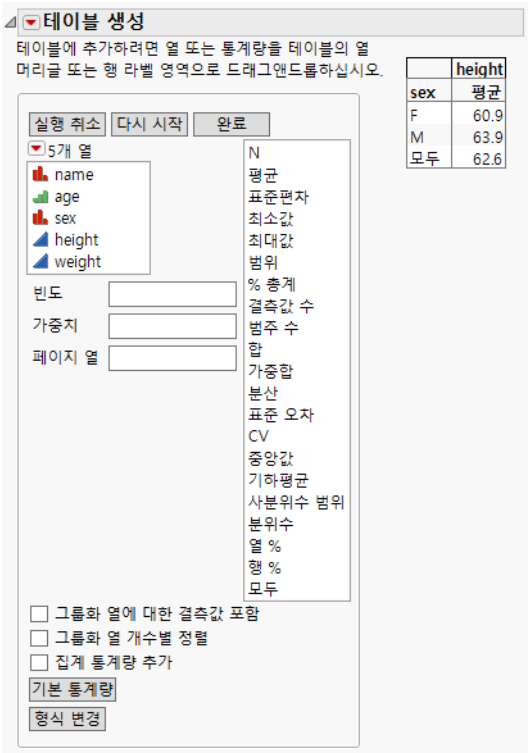
그림 9.5 평균 통계량 추가



또한 남학생과 여학생의 결합 평균을 표시하려고 합니다.

6. **모두**를 클릭하고 sex 위로 드래그합니다 . 또는 단순히 **집계 통계량 추가** 체크박스를 선택해도 됩니다 .

그림 9.6 전체 (" 모두 ") 통계량 추가



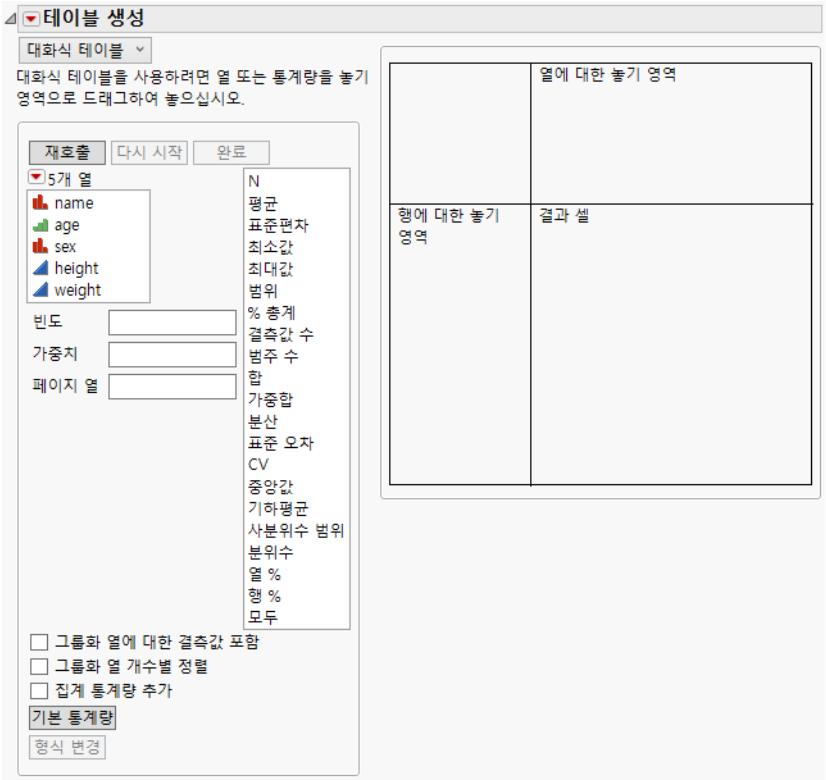
7. (선택 사항) **완료**를 클릭합니다 .

완료된 테이블에 여학생 및 남학생의 평균 키와 둘 모두의 결합 평균 키가 표시됩니다.

테이블 생성 플랫폼 시작

테이블 생성 플랫폼을 시작하려면 **분석 > 테이블 생성**을 선택합니다 .

그림 9.7 대화식 테이블 생성



참고 : 빨간색 삼각형 옵션에 대한 자세한 내용은 " 테이블 생성 플랫폼 옵션 "(302 페이지) 에서 확인하십시오 . " 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

" 테이블 생성 " 창에는 다음과 같은 옵션이 포함되어 있습니다 .

대화식 테이블 / 대화상자 두 모드 사이를 전환합니다. 항목을 드래그하여 놓는 방법으로 사용자 테이블을 생성하려면 대화식 테이블 모드를 사용합니다 . 고정 형식을 사용하여 간단한 테이블을 생성하려면 대화상자 모드를 사용합니다 . 자세한 내용은 " 대화상자 사용 "(294 페이지) 에서 확인하십시오 .

통계량 옵션 표준 통계량을 나열합니다 . 목록의 통계량을 테이블로 드래그하여 테이블에 포함할 수 있습니다 . 자세한 내용은 " 통계량 추가 "(295 페이지) 에서 확인하십시오 .

열에 대한 놓기 영역 열 또는 통계량을 드래그하여 여기에 놓으면 열이 생성됩니다 .

참고 : 데이터 테이블에 "통계량" 옵션의 통계량과 이름이 같은 열이 포함되어 있으면 열 목록의 열 이름을 드래그하여 놓아야 합니다 . 그러지 않으면 JMP 에서는 테이블의 열이 동일한 이름의 통계량으로 대체됩니다 .

행에 대한 놓기 영역 열 또는 통계량을 드래그하여 여기에 놓으면 행이 생성됩니다 .

팁 : 열 목록에서 열을 하나 이상 선택하고 통계량을 하나 이상 선택한 다음 놓기 영역을 Alt 키를 누른 채 클릭 (macOS 의 경우 Option 키를 누른 채 클릭) 하여 테이블의 행 또는 열을 생성할 수도 있습니다 .

결과 셀 드래그하여 놓은 열 또는 통계량을 기준으로 결과 셀을 표시합니다 .

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 데이터 테이블 열을 식별합니다 . 이 옵션은 요약된 데이터에서 각 행에 빈도가 할당된 경우에 유용합니다 .

가중치 데이터에 가중치 (중요도 또는 영향력) 를 할당하는 데 사용할 변수가 있는 데이터 테이블 열을 식별합니다 .

페이지 열 명목형 또는 순서형 열의 각 범주에 대한 별도의 테이블을 생성합니다 . 자세한 내용은 " 예 : 페이지 열 사용 "(310 페이지) 에서 확인하십시오 .

그룹화 열에 대한 결측값 포함 그룹화 열의 결측값에 대한 별도의 그룹을 생성합니다 . 이 옵션을 선택 해제하면 결측값이 테이블에 포함되지 않습니다. 열 특성으로 정의한 모든 결측값 코드가 고려됩니다 .

그룹화 열 개수별 정렬 테이블의 정렬 순서를 그룹화 열의 값에 대한 내림차순으로 변경합니다 .

집계 통계량 추가 모든 행 및 열에 대한 집계 통계량을 추가합니다 .

기본 통계량 분석 열 또는 분석 열이 아닌 열 (예 : 그룹화 열) 을 드래그하여 놓을 때 표시되는 기본 통계량을 변경할 수 있습니다 .

형식 변경 특정 통계량을 표시하기 위한 숫자 형식을 변경할 수 있습니다 . 자세한 내용은 " 숫자 형식 변경 "(297 페이지) 에서 확인하십시오 .

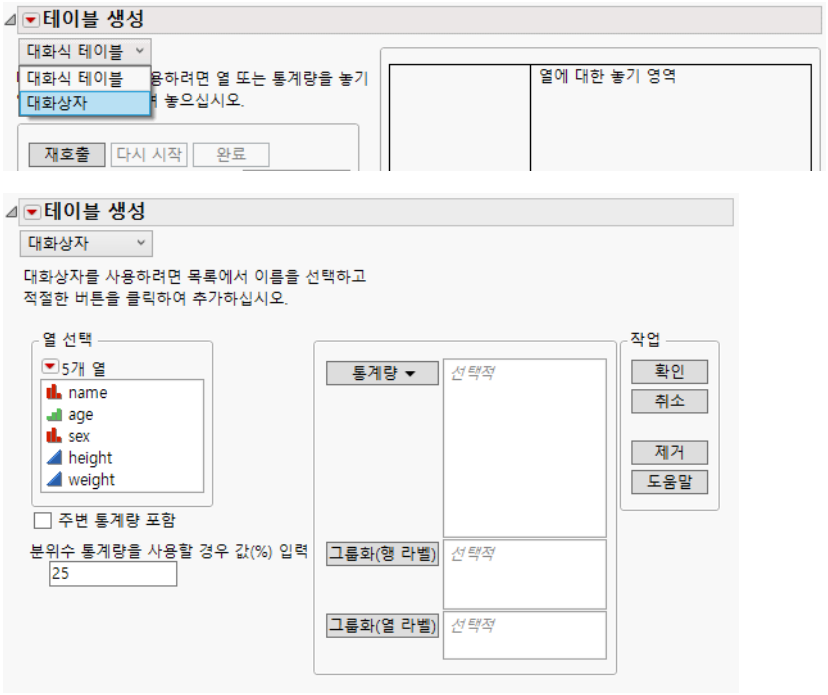
그림 척도 변경 (빨간색 삼각형 메뉴에서 **차트 표시**를 선택한 경우에만 표시됨) 균등 사용자 척도를 지정할 수 있습니다 .

균등 그림 척도 (빨간색 삼각형 메뉴에서 **차트 표시**를 선택한 경우에만 표시됨) 각 막대 열에 대해 표시되는 각 열의 데이터와는 별도로 결정된 척도를 사용하려면 이 상자를 선택 취소합니다 .

대화상자 사용

드래그하여 놓는 방법 대신 대화식 방법으로 테이블을 작성하려는 경우 테이블 생성 플랫폼의 "대화상자" 인터페이스를 사용하여 간단한 테이블을 생성할 수 있습니다. **분석 > 테이블 생성**을 선택한 후 그림 9.8에 표시된 것과 같이 메뉴에서 **대화상자**를 선택합니다. 빨간색 삼각형 메뉴에서 **제어판 표시**를 선택한 다음 새 항목을 테이블로 드래그하여 놓는 방법으로 테이블을 변경할 수 있습니다.

그림 9.8 대화상자 사용



대화상자에는 다음 옵션이 포함되어 있습니다 .

주변 통계량 포함 그룹화 열의 범주에 대한 요약 정보를 집계합니다 .

분위수 통계량을 사용할 경우 값 (%) 입력 특정 비율 지점을 나타내는 값을 입력합니다 . 해당 비율 만큼의 인수가 해당 값보다 작거나 같습니다 . 예를 들어 75% 의 데이터는 75 번째 분위수보다 작습니다 . 이는 모든 그룹화 열에 적용됩니다 .

통계량 열을 선택한 후 해당 열에 적용할 표준 통계량을 선택합니다 . 자세한 내용은 " 통계량 추가 "(295 페이지) 에서 확인하십시오 .

그룹화 (행 라벨) 행 라벨로 사용할 열을 선택합니다 .

그룹화 (열 라벨) 열 라벨로 사용할 열을 선택합니다 .

통계량 추가

테이블 생성 기능은 제어판에 나타나는 표준 통계량 목록을 지원합니다 . 열을 추가할 때와 마찬가지로 이 목록의 키워드를 테이블로 드래그할 수 있습니다 . 다음 사항에 유의하십시오 .

팁 : 열과 통계량을 모두 동시에 선택하고 테이블로 드래그할 수 있습니다 .

- 각 셀과 연관된 통계량은 그룹화 열로 정의된 해당 범주의 모든 관측값에서 분석 열의 값에 대해 계산됩니다 .

- 요청된 통계량은 모두 행 테이블이나 열 테이블에서 동일한 차원으로 있어야 합니다 .
- 연속형 열을 데이터 영역으로 드래그하면 해당 열은 분석 열로 처리됩니다 .

참고 : 분석 열은 통계량을 계산하려는 연속형 숫자 열입니다 . 자세한 내용은 " 분석 열 "(299 페이지) 에서 확인하십시오 .

테이블 생성에는 다음 키워드가 사용됩니다 .

N 열에 있는 비결측값의 수를 제공합니다 . 이는 분석 열이 없을 때의 기본 통계량입니다 .

평균 열 값의 산술평균을 제공합니다 . 비결측값 (정의된 경우 **가중치** 변수를 곱한 값) 의 합을 **가중합**으로 나눈 결과입니다 .

표준편차 비결측값에 대해 계산된 표본 표준편차를 제공합니다 . 이 값은 표본 분산의 제곱근입니다 .

최소값 열에 있는 가장 작은 비결측값을 제공합니다 .

최대값 열에 있는 가장 큰 비결측값을 제공합니다 .

범위 **최대값**과 **최소값**의 차이를 제공합니다 .

% 총계 전체 모집단의 총계에 대한 백분율을 계산합니다 . 계산에 사용되는 분모는 포함된 모든 관측값의 합계이고 , 분자는 해당 범주의 합계입니다 . 분석 열이 없는 경우 "% 총계 " 는 총 개수에 대한 백분율입니다 . 분석 열이 있는 경우 "% 총계 " 는 분석 열의 총 합에 대한 백분율입니다 . 따라서 분모는 포함된 모든 관측값에 대한 분석 열의 합이고 , 문자는 해당 범주에 대한 분석 열의 합입니다 . 키워드를 테이블로 드래그하여 다른 백분율을 요청할 수 있습니다 .

- 테이블의 그룹화 변수 중 하나 이상을 "% 총계 " 머리글에 놓으면 분모 정의가 변경됩니다 . 이 경우 테이블 생성 기능은 해당 그룹화 열의 합을 분모에 사용합니다 .
- 열 합계의 백분율을 구하려면 행 테이블의 모든 그룹화 열을 **% 총계** 머리글 (" 열 %" 와 동일) 로 드래그하여 놓습니다 . 행 합계의 백분율을 구하려면 열 테이블의 모든 그룹화 열을 **% 총계** 머리글 (" 행 %" 와 동일) 로 드래그하여 놓습니다 .

결측값 수 결측값의 개수를 제공합니다 .

범주 수 분석 열에 있는 개별 범주의 개수를 제공합니다 .

합 해당 열에 있는 모든 값의 합을 제공합니다 . 이 통계량은 테이블에 대한 다른 통계량이 없는 경우 분석 열의 기본 통계량입니다 .

가중합 열에 있는 모든 **가중치** 값의 합을 제공합니다 . **가중치** 역할에 할당된 열이 없는 경우 , **가중합**은 비결측값의 총 수입니다 .

분산 비결측값에 대해 계산된 표본 분산을 제공합니다 . 평균으로부터의 편차에 대한 제곱합을 비결측값의 수에서 1 을 뺀 값으로 나눈 것입니다 .

표준 오차 평균의 표준 오차를 제공합니다 . 표준편차를 **N** 의 제곱근으로 나눈 것입니다 . 열에 가중치 역할이 할당된 경우에는 분모가 가중치 합의 제곱근입니다 .

CV (변동 계수) 표준편차를 평균으로 나누고 100 을 곱한 산포 측도입니다 .

중앙값 데이터의 위쪽 절반과 아래쪽 절반을 나누는 기준이 되는 50 번째 백분위수 , 또는 50 번째 분위수 (중앙값) 를 제공합니다 .

기하평균 데이터 곱의 n 제곱근입니다 . 예를 들어 기하평균은 이자율을 계산하는 데 종종 사용됩니다 . 이 통계량은 데이터에 많은 값이 비대칭적 분포로 포함된 경우에도 유용합니다 .

참고 : 음수 값이 있으면 결측값이 발생하고 , 음수가 아니고 0 인 값이 있으면 결과가 0 이 됩니다 .

사분위수 범위 3 사분위수와 1 사분위수 사이의 차이를 제공합니다 .

분위수 특정 비율 지점을 나타내는 값을 제공합니다 . 해당 비율만큼의 인수가 해당 값보다 작거나 같습니다 . 예를 들어 75% 의 데이터는 75 번째 분위수보다 작습니다 . **분위수** 키워드를 클릭하여 테이블로 드래그한 다음 표시되는 상자에 분위수를 입력하는 방법으로 다른 분위수를 요청할 수 있습니다 .

열 % 분석 열이 없는 경우 열 합계에 대한 각 셀 개수의 백분율을 제공합니다 . 분석 열이 있는 경우 " 열 %" 는 분석 열의 열 총 합에 대한 백분율입니다 . 맨 위에 통계량이 있는 테이블의 경우 여러 행 테이블과 함께 테이블에 " 열 %" 를 추가할 수 있습니다 (세로로 누적).

행 % 분석 열이 없는 경우 행 합계에 대한 각 셀 개수의 백분율을 제공합니다 . 분석 열이 있는 경우 " 행 %" 는 분석 열의 행 총 합에 대한 백분율입니다 . 통계량이 나란히 표시되는 테이블의 경우 여러 열 테이블과 함께 테이블에 " 행 %" 를 추가할 수 있습니다 (나란히 표시되는 테이블).

모두 그룹화 열의 범주에 대한 요약 정보를 집계합니다 .

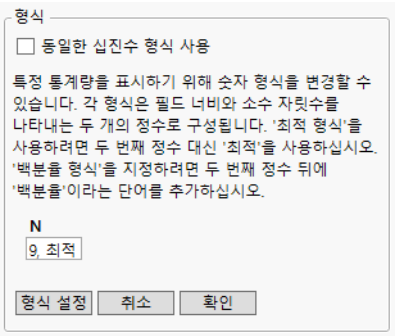
숫자 형식 변경

각 셀의 계산식은 분석 열 및 통계량에 따라 달라집니다 . 개수의 경우 기본 형식에는 소수점 이하 자릿수가 없습니다 . 일부 통계량으로 정의된 각 셀의 경우 , JMP 에서는 요청된 분석 열과 통계량의 형식을 사용하여 적절한 형식을 결정하려고 합니다 . 기본 형식을 재정의하려면 다음을 수행하십시오 .

1. " 테이블 생성 " 창의 맨 아래에 있는 **형식 변경** 버튼을 클릭합니다 .
2. 표시되는 패널에서 필드 너비 , 쉼표 , 테이블에 표시할 소수 자릿수를 차례로 입력합니다 (그림 9.9).
3. 셀 값을 백분율 형식으로 표시하려면 소수 자릿수 다음에 쉼표를 입력하고 **백분율**이라는 단어를 추가합니다 .
4. (선택 사항) 사용할 최적의 형식이 JMP 에서 자동으로 결정되도록 하려면 텍스트 상자에 **최적**이라는 단어를 입력합니다 .

   JMP 가 각 셀 값의 정밀도를 고려하여 가장 적절한 값 표시 방법을 선택합니다 .
5. 변경 사항을 구현하고 " 형식 " 섹션을 닫으려면 **확인**을 클릭하고 , 구현된 변경 사항을 확인하고 " 형식 " 섹션을 닫지는 않으려면 **형식 설정**을 클릭합니다 .

그림 9.9 숫자 형식 변경



테이블 생성 출력

테이블 생성 기능의 출력은 나란히 연결된 하나 이상의 열 테이블과 위에서 아래로 연결된 하나 이상의 행 테이블로 구성됩니다. 출력에 열 테이블 또는 행 테이블이 하나만 포함될 수도 있습니다.

그림 9.10 테이블 생성 출력

	sex											
	F						M					
	age						age					
	12	13	14	15	16	17	12	13	14	15	16	17
	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight
최소값	64	67	81	92	112	116	79	79	92	104	128	134
평균	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
최대값	145	112	142	112	115	116	128	105	119	128	128	172

		sex							
		Female				Male			
		marital status				marital status			
		Married		Single		Married		Single	
		평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
country	size	age	age	age	age	age	age	age	age
American	Large	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	Medium	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	Small	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
European	Large	34.0	7.071	28.0	.	.	.	26.0	.
	Medium	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	Small	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
Japanese	Large	25.0	.	.	.	32.0	.	.	.
	Medium	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	Small	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
country									
American		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
European		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
Japanese		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

		평균		
Type	Size Co	Profits ($M)	Sales ($M)	profit/emp
Computer	big	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
	모두	240.87	5652.02	6159.015
Pharmaceutical	big	894.42	7474.04	17140.70
	medium	698.98	4261.06	24035.11
	small	156.95	1083.75	38337.19
	모두	690.08	5070.25	23546.12
모두	모두	409.32	5433.86	12679.18
Type				
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
모두		409.32	5433.86	12679.18

대화식으로 테이블을 생성하려면 다음 과정을 반복해야 합니다 .

- 해당 목록에서 항목 (열 또는 통계량) 을 클릭하고 행 또는 열에 대한 놓기 영역으로 드래그합니다 . 자세한 내용은 " 테이블 편집 "(301 페이지) 및 " 열 및 행 테이블 "(300 페이지) 에서 확인하십시오 .
- 드래그하여 놓기 과정을 반복하여 테이블에 항목을 추가합니다 . 테이블이 업데이트되어 마지막 추가 항목이 반영됩니다 . 열 머리글 또는 행 라벨이 이미 있는 경우에는 기존 항목을 기준으로 추가 항목의 위치를 결정할 수 있습니다 .

클릭 및 드래그 작업에 대한 다음 사항에 유의하십시오 .

- JMP 에서는 모델링 유형을 사용하여 열의 역할이 결정됩니다 . 연속형 열은 분석 열로 할당됩니다 . 자세한 내용은 " 분석 열 "(299 페이지) 에서 확인하십시오 . 순서형 또는 명목형 열은 그룹화 열로 간주됩니다 . 자세한 내용은 " 그룹화 열 "(300 페이지) 에서 확인하십시오 .
- 초기 테이블에 여러 개의 열을 드래그하여 놓으면 다음 작업이 수행됩니다 .
  - 해당 열이 일련의 공통된 값을 공유하는 경우에는 열이 단일 테이블에 결합됩니다 . 해당 열 이름과 해당 열에서 수집된 범주의 교차표가 생성됩니다 . 각 셀은 하나의 열과 하나의 범주로 정의됩니다 .
  - 해당 열이 공통된 값을 공유하지 않는 경우에는 각 열이 개별 테이블에 포함됩니다 .
  - 언제든지 열을 마우스 오른쪽 버튼으로 클릭하고 **테이블 결합** 또는 **별도의 테이블**을 선택하여 기본 작업을 변경할 수 있습니다 . 자세한 내용은 " 마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴 "(303 페이지) 에서 확인하십시오 .
- 열을 내포하려면 첫 번째 열을 포함하는 테이블을 생성한 다음 추가 열을 첫 번째 열로 드래그합니다 .
- 올바르게 생성된 테이블에서는 모든 그룹화 열 , 모든 분석 열 및 모든 통계량이 각각 함께 표시됩니다 . 따라서 JMP 에서는 분석 열 목록 내에서 통계량 키워드가 변형되지 않습니다 . 또한 그룹화 열 목록 내에 분석 열이 삽입되지도 않습니다 .
- " 테이블 생성 " 제어판을 사용하는 대신 데이터 테이블에서 " 테이블 " 패널의 열을 " 테이블 생성 " 테이블로 드래그할 수 있습니다 .

참고 : 열려 있는 데이터 테이블에서 데이터를 추가하거나 행을 삭제하거나 데이터를 재코딩하면 " 테이블 생성 " 테이블이 업데이트됩니다 .

분석 열

테이블 생성 플랫폼의 분석 열은 통계량을 계산하려는 숫자(연속형) 열입니다. 테이블 생성 기능은 그룹화 열로 구성된 각 범주에 대해 분석 열의 통계량을 계산합니다.

참고 : 모든 분석 열은 행 테이블이나 열 테이블에서 동일한 차원으로 있어야 합니다 .

그룹화 열

테이블 생성 플랫폼의 그룹화 열은 데이터를 여러 정보 범주로 분류하는 데 사용할 열(명목형 또는 순서형)입니다. 이 열에는 문자, 정수 또는 십진수가 포함될 수 있지만 고유 값의 개수는 제한됩니다.

다음 사항에 유의하십시오 .

- 그룹화 열이 내포된 경우 테이블 생성 기능은 열 값의 계층적 내포 구조로부터 개별 범주를 생성합니다 . 예를 들어 F 값 및 M 값을 갖는 Sex 열과 Married 및 Single 값을 갖는 Marital Status 열이 그룹화 열인 경우 테이블 생성 기능은 "F 및 Married", "F 및 Single", "M 및 Married", "M 및 Single" 이라는 네 개의 개별 범주를 생성합니다 .
- 열 테이블과 행 테이블 모두에 그룹화 열을 지정할 수 있습니다. 두 경우 모두 각 테이블 셀을 정의하는 범주가 생성됩니다 .
- 테이블 생성 기능은 하나 이상의 그룹화 열에 대해 결측값이 있는 관측값을 기본적으로 포함하지 않습니다. **그룹화 열에 대한 결측값 포함** 옵션을 선택하면 이러한 관측값을 포함할 수 있습니다 .
- 결측값으로 간주해야 할 코드 또는 값을 지정하려면 " 결측값 코드 " 열 특성을 사용합니다 . **그룹화 열에 대한 결측값 포함** 옵션을 선택하면 이러한 관측값을 포함할 수 있습니다 . 결측값 코드에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

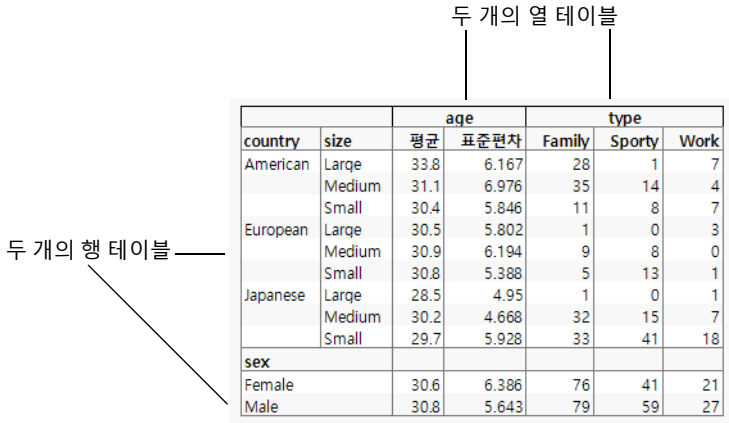
열 및 행 테이블

테이블 생성 플랫폼에서 테이블은 열 머리글과 행 라벨로 정의됩니다. 이러한 하위 테이블을 **행 테이블** 및 **열 테이블**이라고 합니다(그림 9.11).

테이블 생성 플랫폼의 행 및 열 테이블 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .
3. size 를 " 행에 대한 놓기 영역 " 으로 드래그합니다 .
4. country 를 size 머리글의 왼쪽으로 드래그합니다 .
5. 평균을 N 머리글 위로 드래그합니다 .
6. 표준편차를 평균 머리글 아래로 드래그합니다 .
7. age 를 평균 머리글 위로 드래그합니다 .
8. type 을 테이블의 맨 오른쪽으로 드래그합니다 .
9. sex 를 테이블 아래로 드래그합니다 .

그림 9.11 행 및 열 테이블



		age		type		
country	size	평균	표준편차	Family	Sporty	Work
American	Large	33.8	6.167	28	1	7
	Medium	31.1	6.976	35	14	4
	Small	30.4	5.846	11	8	7
European	Large	30.5	5.802	1	0	3
	Medium	30.9	6.194	9	8	0
	Small	30.8	5.388	5	13	1
Japanese	Large	28.5	4.95	1	0	1
	Medium	30.2	4.668	32	15	7
	Small	29.7	5.928	33	41	18
sex						
Female		30.6	6.386	76	41	21
Male		30.8	5.643	79	59	27

열 테이블이 여러 개인 경우 옆쪽의 라벨이 열 테이블 간에 공유됩니다. 이 경우에는 "country" 및 "sex"가 테이블 간에 공유됩니다. 마찬가지로, 행 테이블이 여러 개이면 위쪽의 머리글이 행 테이블 간에 공유됩니다. 이 경우에는 "age" 및 "type"이 테이블 간에 공유됩니다.

테이블 편집

몇 가지 방법으로 테이블에 추가하는 항목을 편집할 수 있습니다. 항목을 삭제하거나, 열 라벨을 제거하거나, 통계 키워드 또는 라벨을 편집할 수 있습니다.

항목 삭제

테이블에 항목을 추가한 후 다음 중 한 가지 방법으로 항목을 제거할 수 있습니다.

- 항목을 테이블 외부로 드래그합니다.
- 마지막 항목을 제거하려면 **실행 취소**를 클릭합니다.
- 항목을 마우스 오른쪽 버튼으로 클릭하고 **삭제**를 선택합니다.

열 라벨 제거

그룹화 열에서는 해당 열과 연관된 범주 위에 열 이름이 표시됩니다. 일부 열의 경우에는 열 이름이 중복될 수 있습니다. 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 라벨 제거**를 선택하면 열 테이블에서 열 이름을 제거할 수 있습니다. 열 라벨을 다시 삽입하려면 연관된 범주 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **열 라벨 복원**을 선택합니다.

통계 키워드 및 라벨 편집

통계 키워드 또는 통계 라벨을 편집할 수 있습니다. 예를 들어 "평균" 대신 "평균값"이라는 단어를 사용할 수 있습니다. 편집하려는 단어를 마우스 오른쪽 버튼으로 클릭하고 **항목 라벨 변경**을

선택합니다. 그런 다음 표시되는 상자에 새 라벨을 입력합니다. 또는 편집 상자에 라벨을 직접 입력해도 됩니다.

한 통계 키워드를 다른 통계 키워드로 변경하면 JMP에서는 단지 라벨이 아니라 통계량을 실제로 변경하려는 것으로 간주됩니다. 즉, 첫 번째 기존 통계량을 테이블에서 삭제한 후 두 번째 통계량을 추가하는 것으로 간주됩니다.

테이블 생성 플랫폼 옵션

" 테이블 생성 " 옆의 빨간색 삼각형 메뉴에서 사용할 수 있는 옵션은 다음과 같습니다 .

제어판 표시 이후 상호 작용을 위해 제어판을 표시합니다 .

테이블 표시 요약된 데이터를 테이블 형식으로 표시합니다 .

차트 표시 요약된 데이터를 요약 통계량 테이블을 반영하는 막대 차트로 표시합니다. 간단한 막대 차트를 사용하여 요약 통계량의 상대적 크기를 시각적으로 비교할 수 있습니다. 기본적으로 모든 막대 열이 동일한 척도를 공유합니다 . **균등 그림 척도** 체크박스를 선택 해제하면 각 막대 열이 각각의 표시된 열에 있는 데이터에서 개별적으로 결정된 척도를 사용하도록 할 수 있습니다 . **그림 척도 변경** 버튼을 사용하면 균등 사용자 척도를 지정할 수 있습니다 . 차트는 0 부터 시작하거나 0 을 중심으로 표시될 수 있습니다 . 데이터가 모두 음수가 아니거나 모두 양수가 아닐 경우 차트 기준선은 0 입니다 . 그렇지 않으면 차트가 0 을 중심으로 표시됩니다 .

음영 표시 행이 여러 개인 경우 테이블에 회색 음영 상자를 표시합니다 .

툴팁 표시 테이블 영역을 마우스로 가리킬 때 나타나는 팁을 표시합니다 .

테스트 빌드 패널 표시 원래 테이블에서 표집한 랜덤 표본을 사용하여 테스트 빌드를 생성할 수 있는 제어 영역을 표시합니다 . 이 기능은 데이터가 많을 때 특히 유용합니다 . 자세한 내용은 " 테스트 빌드 패널 표시 "(303 페이지) 에서 확인하십시오 .

데이터 테이블로 만들기 보고서에서 데이터 테이블을 만듭니다 . 서로 다른 행 테이블의 라벨은 동일한 구조로 매핑되지 않을 수 있으므로 각 행 테이블마다 데이터 테이블이 하나씩 생성됩니다 .

전체 경로 열 이름 생성된 데이터 테이블의 열 이름에 그룹화 열의 정규화된 열 이름을 사용합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

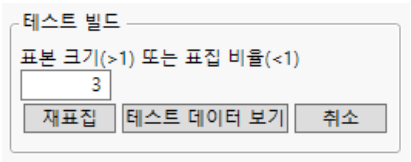
"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

테스트 빌드 패널 표시

데이터 테이블이 매우 큰 경우에는 데이터 테이블의 작은 부분집합을 사용하여 여러 가지 테이블 레이아웃을 시도해 본 후 요약 정보를 가장 잘 보여 주는 테이블을 찾을 수 있습니다. 이 경우 JMP 에서는 지정된 크기의 랜덤 부분집합을 생성한 후, 테이블을 작성할 때 이 부분집합을 사용합니다. 테스트 빌드 기능을 사용하려면 다음을 수행하십시오.

1. " 테이블 생성 " 의 빨간색 삼각형을 클릭하고 **테스트 빌드 패널 표시**를 선택합니다.
2. 그림 9.12 에 표시된 것과 같이 **표본 크기 (>1) 또는 표집 비율 (<1)** 아래의 상자에 원하는 표본 크기를 입력합니다 . 표본 크기는 활성 테이블의 입력한 비율이거나 활성 테이블의 일부 행 수일 수 있습니다.

그림 9.12 테스트 빌드 패널



3. **재표집**을 클릭합니다.
4. JMP 데이터 테이블에서 표집된 데이터를 보려면 **테스트 데이터 보기** 버튼을 클릭합니다. 테스트 빌드 패널을 종료하면 JMP 에서는 전체 데이터 테이블을 사용하여 지정된 대로 테이블을 다시 생성합니다.

마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴

테이블에서 열을 마우스 오른쪽 버튼으로 클릭하면 다음 옵션이 표시됩니다.

삭제 선택한 열을 삭제합니다.

그룹화 열로 사용 분석 열을 그룹화 열로 변경합니다.

분석 열로 사용 그룹화 열을 분석 열로 변경합니다.

항목 라벨 변경 (개별 열 또는 내포된 열의 경우에만 표시됨) 새 라벨을 입력합니다.

테이블 결합 (범주별 열) (개별 열 또는 내포된 열의 경우에만 표시됨) 개별 열 또는 내포된 열을 결합합니다. 자세한 내용은 " 예 : 여러 열을 단일 테이블에 결합 "(308 페이지) 에서 확인하십시오.

그룹화 열 내포 그룹화 열을 세로 또는 가로로 내포합니다.

별도의 테이블 (결합된 테이블의 경우에만 표시됨) 각 열에 대해 별도의 테이블을 생성합니다.

열 라벨 제거 열 테이블에서 열 이름을 제거합니다.

열 라벨 복원 숨겨진 열 이름을 열 테이블에 복원합니다 .

개수별 정렬 그룹화 열의 수준을 각 수준 개수에 따라 높은 것부터 낮은 것 순서로 정렬합니다 . 이 옵션은 특정 그룹화 열에 대한 " 그룹화 열 개수별 정렬 " 옵션의 설정을 재정의하는 데에도 사용할 수 있습니다 .

테이블 생성 플랫폼의 추가 예

- " 예 : 여러 테이블 생성 및 내용 재배열 "
- " 예 : 여러 열을 단일 테이블에 결합 "
- " 예 : 페이지 열 사용 "

예 : 여러 테이블 생성 및 내용 재배열

이 예에는 다음과 같은 단계가 포함되어 있습니다.

1. 개수 테이블 생성
2. 통계량을 보여 주는 테이블 생성
3. 테이블 내용 재배열

개수 테이블 생성

설문 조사 대상자 중 일본, 유럽 및 미국 자동차의 소유자 수가 포함된 테이블을 생성하려고 한다고 가정해 보겠습니다. 또한 이 인원 수를 자동차 크기별로 세분하려고 합니다.

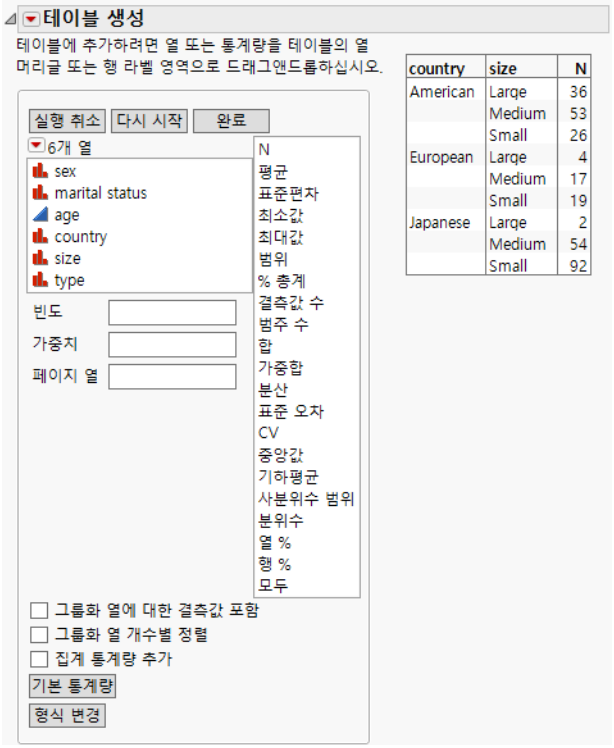
그림 9.13 자동차 소유자 수를 보여 주는 테이블

country	size	N
American	Large	36
	Medium	53
	Small	26
European	Large	4
	Medium	17
	Small	19
Japanese	Large	2
	Medium	54
	Small	92

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .
3. country 를 클릭하고 " 행에 대한 놓기 영역 " 으로 드래그합니다 .

4. size 를 클릭하고 country 머리글의 오른쪽으로 드래그합니다 .

그림 9.14 테이블에 추가된 country 및 size



통계량을 보여 주는 테이블 생성

각 크기의 자동차를 소유한 사람의 나이 평균 및 표준편차를 확인하려고 한다고 가정해 보겠습니다.

그림 9.15 나이의 평균 및 표준편차를 보여 주는 테이블

country	size			
American	Large	age	평균	33.8
			표준편차	6.167
	Medium	age	평균	31.1
			표준편차	6.976
	Small	age	평균	30.4
			표준편차	5.846
European	Large	age	평균	30.5
			표준편차	5.802
	Medium	age	평균	30.9
			표준편차	6.194
	Small	age	평균	30.8
			표준편차	5.388
Japanese	Large	age	평균	28.5
			표준편차	4.95
	Medium	age	평균	30.2
			표준편차	4.668
	Small	age	평균	29.7
			표준편차	5.928

1. 그림 9.14 에서 시작합니다 . age 를 클릭하고 size 머리글의 오른쪽으로 드래그합니다 .
2. **평균**을 클릭하고 " 합 " 위로 드래그합니다 .
3. **표준편차**를 클릭하고 " 평균 " 아래로 드래그합니다 .

   테이블의 " 평균 " 아래에 " 표준편차 " 가 배치됩니다 . " 평균 " 위에 " 표준편차 " 를 놓으면 테이블의 " 평균 " 위에 " 표준편차 " 가 배치됩니다 .

그림 9.16 테이블에 추가된 Age, 평균 및 표준편차

country	size			
American	Large	age	평균	33.8
			표준편차	6.167
	Medium	age	평균	31.1
			표준편차	6.976
	Small	age	평균	30.4
			표준편차	5.846
European	Large	age	평균	30.5
			표준편차	5.802
	Medium	age	평균	30.9
			표준편차	6.194
	Small	age	평균	30.8
			표준편차	5.388
Japanese	Large	age	평균	28.5
			표준편차	4.95
	Medium	age	평균	30.2
			표준편차	4.668
	Small	age	평균	29.7
			표준편차	5.928

테이블 내용 재배열

크기를 맨 위에 배치하여 교차표 레이아웃을 표시하려고 한다고 가정해 보겠습니다.

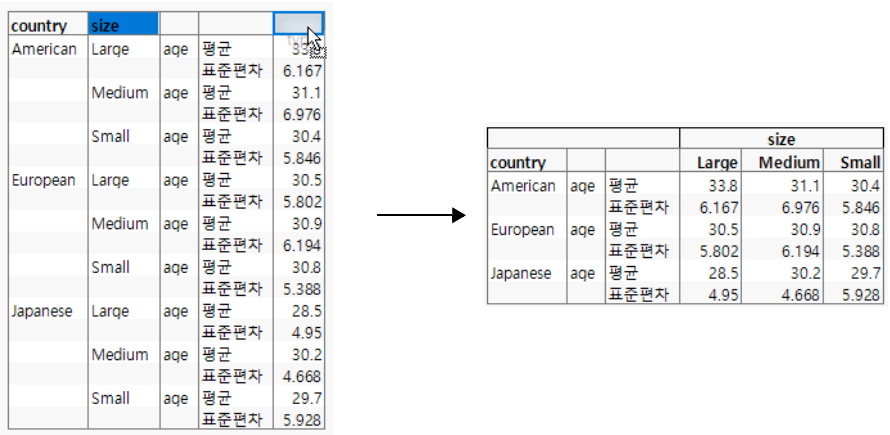
그림 9.17 맨 위에 size 표시

	size					
	Large		Medium		Small	
	평균	표준편차	평균	표준편차	평균	표준편차
country	age	age	age	age	age	age
American	33.8	6.167	31.1	6.976	30.4	5.846
European	30.5	5.802	30.9	6.194	30.8	5.388
Japanese	28.5	4.95	30.2	4.668	29.7	5.928

테이블 내용을 재배열합니다 .

1. 그림 9.16 에서 시작합니다 . size 머리글을 클릭하고 테이블 머리글의 오른쪽으로 드래그합니다 .

그림 9.18 size 이동



country	size			
American	Large	age	평균	33
			표준편차	6.167
	Medium	age	평균	31.1
			표준편차	6.976
	Small	age	평균	30.4
			표준편차	5.846
European	Large	age	평균	30.5
			표준편차	5.802
	Medium	age	평균	30.9
			표준편차	6.194
	Small	age	평균	30.8
			표준편차	5.388
Japanese	Large	age	평균	28.5
			표준편차	4.95
	Medium	age	평균	30.2
			표준편차	4.668
	Small	age	평균	29.7
			표준편차	5.928

			size		
country			Large	Medium	Small
American	age	평균	33.8	31.1	30.4
		표준편차	6.167	6.976	5.846
European	age	평균	30.5	30.9	30.8
		표준편차	5.802	6.194	5.388
Japanese	age	평균	28.5	30.2	29.7
		표준편차	4.95	4.668	5.928

2. "age" 를 클릭하고 Large Medium Small 머리글 아래로 드래그합니다 .
3. " 평균 " 과 " 표준편차 " 를 모두 선택하고 Large 머리글 아래로 드래그합니다 .

이제 테이블의 데이터가 명확하게 표시됩니다. 따라서 자동차 소유자 나이의 평균과 표준편차를 자동차 크기 및 국가별로 세분해서 확인하기가 쉬워졌습니다.

예 : 여러 열을 단일 테이블에 결합

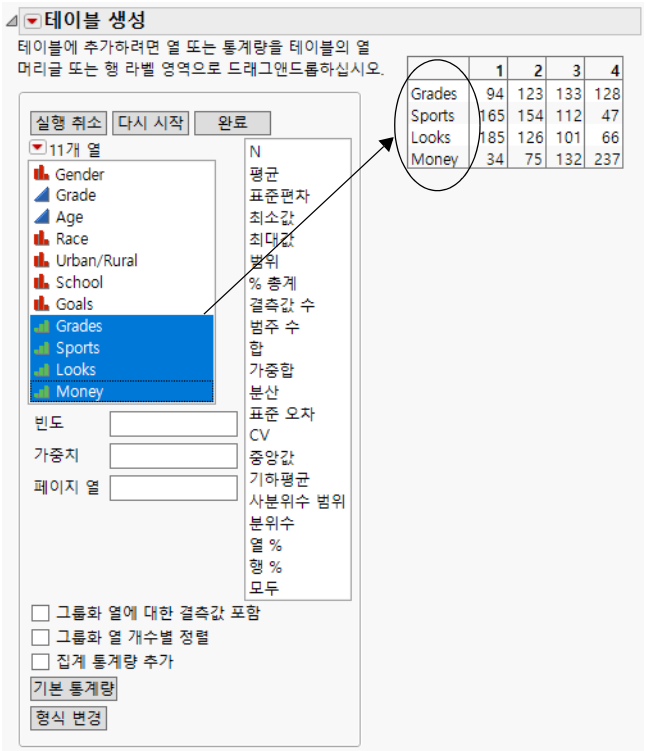
이 예에는 아동의 인기도에 대한 자가 보고 요인(성적, 스포츠, 외모, 경제 여건)의 중요도를 나타내는 학생 데이터가 있습니다. 이러한 모든 요인을 추가 통계량 및 요인을 포함하는 결합된 단일 테이블에서 확인하려고 한다고 가정해 보겠습니다.

그림 9.19 인구 통계학 데이터 추가

		% 총계														
		Urban/Rural														
		Rural					Suburban					Urban				
Gender		1	2	3	4	모두	1	2	3	4	모두	1	2	3	4	모두
boy	Grades	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	Sports	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	Looks	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
	Money	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
girl	Grades	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	Sports	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	Looks	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	Money	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Children's Popularity.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .
3. Grades, Sports, Looks 및 Money 를 선택하고 " 행에 대한 놓기 영역 " 으로 드래그합니다 .

그림 9.20 범주별 열



결합된 단일 테이블이 나타납니다 .

각 범주의 등급 (1 ~ 4) 백분율에 대한 테이블을 생성하십시오 .

4. Gender 를 왼쪽의 빈 머리글로 드래그합니다 .
5. **% 총계**를 번호가 매겨진 머리글 위로 드래그합니다 .
6. **모두**를 번호 4 옆으로 드래그합니다 .

그림 9.21 테이블에 추가된 "Gender", "% 총계 " 및 " 모두 "

		% 총계				
Gender		1	2	3	4	모두
boy	Grades	8.16%	12.76%	14.02%	12.55%	47.49%
	Sports	26.57%	12.55%	6.49%	1.88%	47.49%
	Looks	9.21%	15.48%	12.34%	10.46%	47.49%
	Money	3.56%	6.69%	14.64%	22.59%	47.49%
girl	Grades	11.51%	12.97%	13.81%	14.23%	52.51%
	Sports	7.95%	19.67%	16.95%	7.95%	52.51%
	Looks	29.50%	10.88%	8.79%	3.35%	52.51%
	Money	3.56%	9.00%	12.97%	26.99%	52.51%

인구 통계학 데이터를 추가하여 테이블을 더 세분합니다 .

7. Urban/Rural 을 **% 총계** 머리글 아래로 드래그합니다 .

그림 9.22 테이블에 추가된 "Urban/Rural"

		% 총계														
		Urban/Rural														
		Rural					Suburban					Urban				
Gender		1	2	3	4	모두	1	2	3	4	모두	1	2	3	4	모두
boy	Grades	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	Sports	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	Looks	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
	Money	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
qirl	Grades	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	Sports	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	Looks	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	Money	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

지방, 교외 및 도시 지역에 사는 소년의 경우 스포츠가 인기에 가장 중요한 요인임을 알 수 있습니다. 지방, 교외 및 도시 지역에 사는 소녀의 경우에는 외모가 인기에 가장 중요한 요인입니다.

예 : 페이지 열 사용

이 예에는 남학생과 여학생의 키 측정값이 포함된 데이터가 있습니다. 학생의 나이별 키 평균을 보여 주는 테이블을 생성하려고 합니다. 그런 다음 데이터를 성별별로 서로 다른 테이블에 층화하려고 합니다. 이렇게 하려면 층화 열을 페이지 열로 추가하여 각 그룹에 대한 페이지를 작성합니다.

그림 9.23 학생의 성별별 키 평균

sex = F

age	height 평균
12	58.6
13	59.0
14	62.6
15	63.0
16	62.5
17	62.0

여학생

sex = M

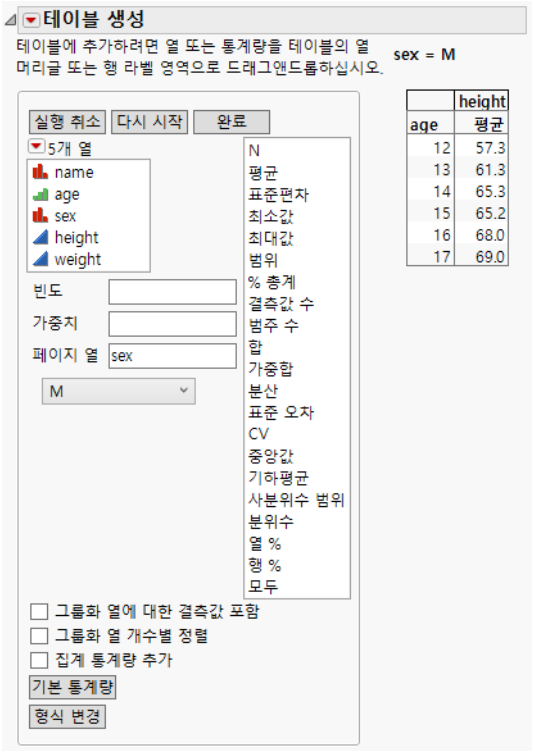
age	height 평균
12	57.3
13	61.3
14	65.3
15	65.2
16	68.0
17	69.0

남학생

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .
   height 변수를 검토할 것이므로 테이블의 맨 위에 이 변수를 표시하려고 합니다 .
3. height 를 클릭하고 " 열에 대한 놓기 영역 " 으로 드래그합니다 .
   나이별로 통계량을 살펴볼 것이므로 age 변수를 나란히 표시하려고 합니다 .
4. age 를 클릭하고 숫자 "2502" 옆의 빈 셀로 드래그합니다 .
5. " 평균 " 을 클릭하고 " 합 " 이라는 셀로 드래그합니다 .
6. sex 를 클릭하고 **페이지 열**로 드래그합니다 .
7. " 페이지 열 " 목록에서 "F" 를 선택하여 여학생만의 평균 키를 표시합니다 .

8. " 페이지 열 " 목록에서 "M" 을 선택하여 남학생만의 평균 키를 표시합니다 . **선택된 항목 없음** 을 선택하여 모든 값을 표시할 수도 있습니다 .

그림 9.24 페이지 열 사용



10 장

JMP PRO 시뮬레이션

모수 재표집을 통해 난제 해결 방안 모색

시뮬레이션 플랫폼은 JMP Pro 에서만 사용할 수 있습니다.

시뮬레이션 기능으로는 강력한 모수 및 비모수 시뮬레이션을 수행할 수 있습니다. 시뮬레이션 기능을 사용하여 수행할 수 있는 작업은 다음과 같습니다.

- 붓스트랩을 확장하여 모수 붓스트랩을 제공합니다.
- 비표준 상황에서 검정력을 계산합니다.
- 비표준 상황에서 예측값 같은 통계량의 분포와 신뢰 구간에 대해 근사값을 산출합니다.
- 순열 검정을 수행합니다.
- 모형의 예측 변수에 대한 가정의 효과를 탐색합니다.
- 모형을 기준으로 다양한 "what if" 시나리오를 탐색합니다.
- 새로운 통계적 방법 또는 기존 통계적 방법을 평가합니다.

시뮬레이션 옵션은 붓스트랩을 지원하는 모든 보고서를 포함하여 다양한 보고서에서 사용할 수 있습니다. 시뮬레이션 옵션에 액세스하려면 보고서를 마우스 오른쪽 버튼으로 클릭합니다.

그림 10.1 시뮬레이션을 통한 검정력 분석

목차

시뮬레이션 플랫폼 개요 315
시뮬레이션 사용 예 315
분산 성분에 대한 준모수 신뢰 구간 구성 316
순열 검정 수행 321
일반화 회귀의 요인 유지 탐색 323
비선형 모형에 대한 전향적 검정력 분석 328
시뮬레이션 창 시작 337
시뮬레이션 창 337
시뮬레이션 결과 테이블 338
시뮬레이션 결과 보고서 338
시뮬레이션 검정력 보고서 339

JMP PRO 시뮬레이션 플랫폼 개요

시뮬레이션 플랫폼에서는 보고서의 통계량 열에 대해 시뮬레이션된 결과를 제공합니다. 보고서의 통계량 열을 마우스 오른쪽 버튼으로 클릭하고 "시뮬레이션"을 선택합니다. 그런 다음 "시뮬레이션" 창에서 시뮬레이션의 기초로 사용할 데이터 테이블 내의 열을 지정합니다. 이 열은 **스위치 아웃**할 열입니다. 이 열은 분석에서 어떤 역할이든 할 수 있습니다. 특히 모형에서는 반응 또는 예측 변수가 될 수 있습니다. 다음으로는 데이터 테이블에서 시뮬레이션에 사용할 계산식이 포함된 열을 지정합니다. 이 열은 **스위치 인**할 열입니다. 이 열은 스위치 아웃한 열의 대리 역할을 합니다.

참고 : 데이터 테이블에는 랜덤 성분이 있는 열이 포함되어 있어야 합니다 .

이 방법의 작동 방식은 다음과 같습니다. 스위치 인한 계산식 열의 계산식을 기반으로 시뮬레이션된 값을 포함하는 열이 생성됩니다. 관심 있는 통계량이 포함된 보고서를 생성했던 전체 분석이 시뮬레이션된 값을 포함하는 이 새 열을 사용하여 다시 실행되어 스위치 아웃한 열을 대체합니다. 이 과정이 N번 반복됩니다. 여기서 N은 사용자가 지정한 총 표본 수입니다.

시뮬레이션 분석에서는 분석 요약을 보여 주는 출력 데이터 테이블을 생성합니다 .

- 데이터 테이블의 각 행은 시뮬레이션된 값을 포함하는 한 열에 대한 분석 결과를 나타냅니다 .
- 시뮬레이션에 관련된 보고서 테이블의 각 행마다 열이 하나씩 있습니다 .
- 분석을 용이하게 하는 스크립트가 있습니다 .

팁 : 시뮬레이션 플랫폼에서는 시뮬레이션 기능이 호출된 플랫폼 보고서에 표시된 전체 분석을 다시 실행합니다 . 따라서 보고서에 포함된 관련 없는 분석으로 인해 선택한 열에 대한 시뮬레이션이 느리게 실행될 수 있습니다 . 시뮬레이션이 오래 걸릴 경우에는 플랫폼 보고서에서 관련 없는 옵션을 제거한 후 시뮬레이션을 실행하십시오 .

시뮬레이션 기능은 연관성 분석, 다이어그램, 다차원 척도법, 다중 요인 분석, 신뢰도 블록 다이어그램, 신뢰도 예측, 수리 가능 시스템 시뮬레이션, 반응 변수 선별, 텍스트 탐색기를 제외한 모든 통계 플랫폼에서 사용할 수 있습니다.

JMP PRO 시뮬레이션 사용 예

다양한 시나리오에서 데이터에 대한 시뮬레이션된 결과를 얻는 방법을 알아봅니다 .

- " 분산 성분에 대한 준모수 신뢰 구간 구성 "
- " 순열 검정 수행 "
- " 일반화 회귀의 요인 유지 탐색 "
- " 비선형 모형에 대한 전향적 검정력 분석 "

- 비정규 변수의 Ppk 및 부적합률에 대한 신뢰 구간 시뮬레이션 방법을 보여 주는 예는 Quality and Process Methods 의 에서 확인하십시오 .

JMP PRO 분산 성분에 대한 준모수 신뢰 구간 구성

이 예에서는 변경하기 힘든 변수의 분산 성분에 대한 준모수 신뢰 구간을 구하는 것이 목표입니다. 이 데이터에 대해 온도, 시간 및 촉매의 양이 반응에 미치는 효과를 알아보려고 합니다. 온도는 매우 변경하기 힘든 변수(주구 요인)이고, 시간은 변경하기 힘든 변수(하위구 요인)이며, 촉매 양은 변경하기 쉬운 변수입니다. 주구 및 하위구 요인에 대한 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

이전 연구에서는 주구 표준편차가 오차 표준편차의 약 2 배이고, 하위구 오차는 오차 표준편차의 약 1.5 배인 것으로 나타났습니다. 분산 성분이 점근적으로 정규 분포를 따른다고 가정하는 REML 보고서에서 제공된 Wald 구간은 포함 범위가 매우 작은 특성이 있습니다. 따라서 시뮬레이션된 분산 성분 분포의 백분위수를 사용하여 신뢰 구간을 구합니다.

JMP PRO 설계 구성

이 섹션에서는 분할 - 분할구 실험을 위한 사용자 설계를 구성합니다 .

팁 : 이 섹션의 단계를 건너뛰려면 **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Catalyst Design.jmp 를 여십시오 . Catalyst Design.jmp 데이터 테이블에서 **DOE Simulate** 스크립트 옆의 녹색 삼각형을 클릭하십시오 . 그런 다음 " 모형 적합 "(318 페이지) 절차로 이동하십시오 .

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 요인 " 개요에서 **N 개 요인 추가** 옆에 3 을 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. 추가된 요인을 두 번 클릭하여 이름을 Temperature, Time 및 Catalyst 로 바꿉니다 .
   이러한 요인의 " 값 " 은 기본값 –1 및 1 을 유지합니다 .
5. Temperature 에서 **쉬움**을 클릭하고 **매우 어려움**을 선택합니다 .
   이렇게 하면 Temperature 가 주구 요인으로 정의됩니다 .
6. Time 에서 **쉬움**을 클릭하고 **어려움**을 선택합니다 .
   이렇게 하면 Time 이 하위구 요인으로 정의됩니다 .
7. **계속**을 클릭합니다 .
8. " 모형 " 개요에서 **교호작용 > 2 차**를 선택합니다 .
   이렇게 하면 모형에 모든 2 원 교호작용이 추가됩니다 .
9. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **반응 시뮬레이션**을 선택합니다 .

이렇게 하면 " 테이블 생성 " 을 선택하여 설계 테이블을 구성한 후에 " 반응 시뮬레이션 " 창이 열립니다 .

참고 : 10 단계에서 " 난수 시드값 " 을 설정하고 11 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 설계가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

10. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 . "12345" 를 입력하고 **확인**을 클릭합니다 .
11. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **시작 수**를 선택합니다 . "1000" 을 입력하고 **확인**을 클릭합니다 .
12. **설계 생성**을 클릭합니다 .
13. **테이블 생성**을 클릭합니다 .

참고 : Y 및 Y 시뮬레이션 열의 항목은 그림 10.2 에 표시된 것과 다를 수 있습니다 .

그림 10.2 설계 테이블

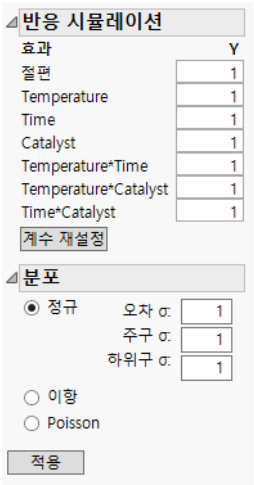
Catalyst Design
잠긴 파일 C:₩Program File
Design Custom Design
Criterion D Optimal
Model
Model for Y Simulated
Evaluate Design
DOE Simulate
DOE Dialog

열(7/0)
Whole Plots ✱
Subplots ✱
Temperature ✱
Time ✱
Catalyst ✱
Y ✱
Y Simulated

행
모든 행 24
선택됨 0
제외 0
숨김 0
라벨 항목 0

	Whole Plots	Subplots	Temperature	Time	Catalyst	Y	Y Simulated
1	1	1	1	1	-1	-0.046768023	2.38362212
2	1	1	1	1	1	6.539690219	6.934627
3	1	1	1	1	1	6.5397108257	6.6911769
4	1	2	1	-1	1	0.494801276	-1.5879342
5	1	2	1	-1	-1	-2.65841644	-3.2406153
6	1	2	1	-1	-1	-0.200740508	-2.5118393
7	2	3	-1	1	1	3.1335699565	2.10432757
8	2	3	-1	1	1	2.2739679213	2.36399771
9	2	3	-1	1	-1	1.1704395547	0.23356541
10	2	4	-1	-1	-1	-0.693941035	1.29953351
11	2	4	-1	-1	-1	2.3525055616	2.18979623
12	2	4	-1	-1	1	-1.545149968	-0.0606938
13	3	5	-1	1	-1	-1.4838427	-1.08124
14	3	5	-1	1	-1	-1.375362439	-2.0077657
15	3	5	-1	1	1	0.7434199911	-0.3799921
16	3	6	-1	-1	1	-1.11989643	-3.4434001
17	3	6	-1	-1	-1	3.4346315881	1.28011609
18	3	6	-1	-1	1	0.2987179478	-2.752904
19	4	7	1	-1	1	-1.245681444	-1.7403239
20	4	7	1	-1	1	0.4951842993	0.74829587
21	4	7	1	-1	-1	-0.6602484	-1.3436848
22	4	8	1	1	-1	0.7779088461	1.23261462
23	4	8	1	1	1	7.9432230978	5.9851515
24	4	8	1	1	-1	1.0668195454	1.35743939

그림 10.3 반응 시뮬레이션 창



설계 테이블과 " 반응 시뮬레이션 " 창이 나타납니다 . 설계 테이블에는 **DOE 시뮬레이션** 스크립트가 포함되어 있습니다 . 언제든지 이 스크립트를 실행하여 다른 모수 값을 지정할 수 있습니다 .

다음 섹션으로 진행하여 주구 및 하위구 오차의 표준편차를 지정하고 첫 번째 시뮬레이션 값 집합에 REML 모형을 적합시킬 수 있습니다.

JMP PRO 모형 적합

이 섹션에서는 REML 방법을 사용하여 모형을 적합시킵니다 . 주구 및 하위구 오차가 정규 분포를 따른다고 가정하십시오 . 표준편차 추정값에 따르면 오차 표준편차가 약 1 단위일 경우 주구 표준편차는 약 2 단위이고 하위구 표준편차는 약 1.5 단위입니다 . 주구 및 하위구 변동에만 관심이 있으므로 " 반응 시뮬레이션 " 개요의 " 효과 " 에 할당된 값은 변경하지 않아도 됩니다 .

1. " 분포 " 패널 (그림 10.3) 에서 **주구** σ 옆에 2 를 입력합니다 .

   기본적으로 " 정규 " 분포가 선택되어 있습니다 . 따라서 계산식에 정규 오차가 추가됩니다 .
2. **하위구** σ 옆에 1.5 를 입력합니다 .
3. **적용**을 클릭합니다 .

   데이터 테이블에서 Y 시뮬레이션의 계산식이 업데이트되어 지정한 내용이 반영됩니다 . 계산식을 보려면 " 열 " 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다 .
4. 데이터 테이블에서 **모형** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
5. **Y** 버튼 옆의 Y 변수를 클릭하고 **제거**를 클릭합니다 .
6. Y 시뮬레이션을 클릭하고 **Y** 버튼을 클릭합니다 .

   이렇게 하면 Y 가 시뮬레이션 계산식이 포함된 열로 대체됩니다 .
7. **실행**을 클릭합니다 .

적합된 모형은 시뮬레이션된 반응의 단일 집합을 기반으로 합니다 .

참고 : Y 시뮬레이션의 값은 무작위로 생성되므로 보고서의 항목은 그림 10.4 에 표시된 것과 다를 수 있습니다 .

그림 10.4 Wald 신뢰 구간을 보여 주는 REML 보고서

REML 분산 성분 추정값

임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
Whole Plots	3.8107856	1.556364	1.8007472	-1.973036	5.0857636	0.3874	68.445
Subplots	0.7568606	0.3091097	0.4661521	-0.604532	1.222751	0.5073	13.594
잔차		0.4084103	0.160228	0.2146172	1.060295		17.961
합계		2.2738839	1.8035082	0.7468087	28.115033		100.000

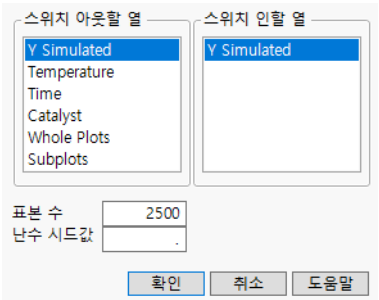
-2*로그 우도 = 63.890021519
참고: 합계는 양의 분산 성분의 합입니다.
음수 추정값을 포함한 합계 = 2.2738839

JMP PRO 신뢰 구간 생성

이 섹션에서는 분산 성분 추정값을 시뮬레이션하고 이를 사용하여 시뮬레이션된 백분위수 신뢰 구간을 구성합니다 .

1. "REML 분산 성분 추정값 " 개요에서 **분산 성분** 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .

그림 10.5 시뮬레이션 창



시뮬레이션에서는 모형을 실행하는 데 사용된 Y 시뮬레이션 열이 각 시뮬레이션에 대해 새로운 시뮬레이션 값 열을 생성하는 Y 시뮬레이션 열의 새 인스턴스로 대체됩니다 . 마우스 오른쪽 버튼으로 클릭하여 선택 상태로 표시된 **분산 성분** 열이 " 모수 추정값 " 테이블에 나열된 각 효과에 대해 시뮬레이션됩니다 .

2. **표본 수** 옆에 **200** 을 입력합니다 .
3. (선택 사항) **난수 시드값** 옆에 456 을 입력합니다 .

   이렇게 하면 그림 10.6 에 표시된 값이 1 행의 값만 제외하고 모두 재현됩니다 .
4. **확인**을 클릭합니다 .

   1 행의 항목은 그림 10.6 에 표시된 것과 다를 수 있습니다 .

그림 10.6 분산 성분에 대한 시뮬레이션 결과 테이블 (일부분)

최소 제곱 적합 시뮬레이...
난수 시드값 456
결합 데이터 테이블 생성
분포

열(6/0)
Y ✱
SimID•
Subplots
Whole Plots
잔차
합계

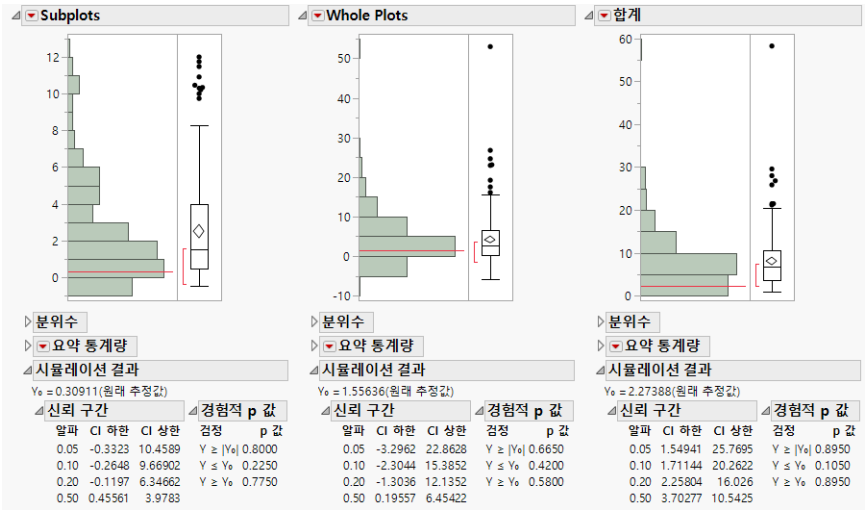
행
모든 행 201
선택됨 0
제외 1
숨김 0
라벨 항목 0

	Y	SimID•	Subplots	Whole Plots	잔차	합계
1	Y Simulated	0	0.3091096508	1.5563639995	0.4084102799	2.2738839302
2	Y Simulated	1	9.9978320675	0.1160090855	1.7118437836	11.825684937
3	Y Simulated	2	4.5617762065	3.3652848741	0.7259348275	8.6529959081
4	Y Simulated	3	3.0383907272	22.95631251	0.8416572537	26.836360491
5	Y Simulated	4	-0.176636748	12.156574946	0.9872034755	13.143778421
6	Y Simulated	5	-0.153702674	-0.060658009	1.3514412755	1.3514412755
7	Y Simulated	6	1.1063669307	11.804365807	0.9837455205	13.894478258
8	Y Simulated	7	0.1630909822	2.6824797817	0.8559296173	3.7015003812
9	Y Simulated	8	0.2154635485	26.765100099	1.0422128078	28.022776455
10	Y Simulated	9	4.0615625437	12.75877132	1.0302645404	17.850598404
11	Y Simulated	10	1.8169128804	-0.433818317	1.2569403028	3.0738531832
12	Y Simulated	11	6.9696566519	2.4634884406	1.6719518802	11.105096973
13	Y Simulated	12	1.9758939136	4.2166345958	1.0326091926	7.2251377019
14	Y Simulated	13	1.1584152266	2.8027318323	0.7050109246	4.6661579836
15	Y Simulated	14	5.2888052789	0.3194564512	0.9281404099	6.53640214
16	Y Simulated	15	1.4827905134	-0.270350632	1.6586718874	3.1414624009
17	Y Simulated	16	0.7636963822	-0.192580268	0.9471174937	1.7108138759
18	Y Simulated	17	1.1919618966	0.756815441	0.9782751744	2.927052512
19	Y Simulated	18	1.0391874862	4.8135329401	1.5524955876	7.4052160139

" 최소 제곱 적합 시뮬레이션 결과 (분산 성분)" 데이터 테이블의 첫 번째 행은 **분산 성분**의 초기값을 포함하므로 제외됩니다 . 나머지 행에는 시뮬레이션된 값이 포함되어 있습니다 .

5. **Distribution** 스크립트를 실행합니다 .

그림 10.7 분산 성분에 대한 분포 그림 (일부분)



" 시뮬레이션 결과 " 보고서에는 각 분산 성분에 대해 다양한 신뢰 수준의 신뢰 구간이 표시됩니다 . 각 테이블의 알파 =0.05 행에 표시된 95% 구간을 REML 보고서에 제공된 구간 (그림 10.4) 과 비교하십시오 .

– 주구 분산 성분에 대해 시뮬레이션된 95% 신뢰 구간은 -3.296 ~ 22.863 입니다 . REML 보고서에 제공된 Wald 구간은 -1.973 ~ 5.086 입니다 .

– 하위구 분산 성분에 대해 시뮬레이션된 95% 신뢰 구간은 -0.332 ~ 10.459 입니다 . REML 보고서에 제공된 Wald 구간은 -0.605 ~ 1.223 입니다 .

시뮬레이션을 사용하여 구한 구간은 단일 값 집합으로부터 계산된 REML 구간보다 상당히 더 넓습니다 . 보다 정확한 구간을 구하려면 더 많은 수의 시뮬레이션을 실행하는 것이 좋습니다 .

JMP PRO 순열 검정 수행

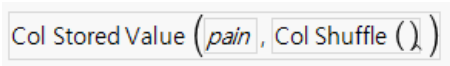
이 예에서는 시뮬레이션 기능을 사용하여 난수화 또는 순열 검정을 수행합니다. 세 가지 약물이 통증에 미치는 효과를 연구하려고 하며 관심 사항은 약물의 효과가 다른지 여부입니다. 표본 크기가 매우 작고 일반적인 ANOVA 가정에 다소 위배되기 때문에 시뮬레이션 기능을 사용하여 순열 검정을 수행할 수 있습니다.

먼저, 세 가지 약물의 투여 시 통증 측정값을 무작위로 뒤섞는 계산식을 구성합니다. 효과가 없다는 귀무가설하에 이러한 배분은 모두 동일한 가능성을 갖습니다. 따라서 이 방식으로 얻은 *F* 비는 귀무가설하의 *F* 비에 근사한 분포를 따릅니다. 마지막으로, 관측된 *F* 비 값을 시뮬레이션으로 구한 귀무 분포와 비교합니다.

JMP PRO 시뮬레이션 계산식 정의

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Analgesics.jmp 를 엽니다 .
2. **열 > 새 열**을 선택합니다 .
3. " 열 이름 " 에 Pain Shuffled 를 입력합니다 .
4. " 열 특성 " 목록에서 **계산식**을 선택합니다 .
5. 함수 목록에서 **행 > Col Stored Value** 를 선택합니다 .
6. 열 목록에서 pain 을 두 번 클릭합니다 .
7. 편집기 패널의 위쪽에 있는 기호 목록에서 삽입 키 (^) 를 클릭합니다 .
8. 함수 목록에서 **난수 > Col Shuffle** 을 선택합니다 .

그림 10.8 완료된 계산식



이 계산식은 pain 열의 항목을 무작위로 뒤섞습니다 .

9. 계산식 편집기 창에서 **확인**을 클릭합니다 .
10. 열 정보 창에서 **확인**을 클릭합니다 .

순열 검정 수행

1. **분석 > X 로 Y 적합**을 선택합니다 .

2. pain 을 선택하고 **Y, 반응**을 클릭합니다 .
3. drug 를 선택하고 **X, 요인**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 일원 분석 " 의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .

그림 10.9 분산 분석 보고서

분산 분석

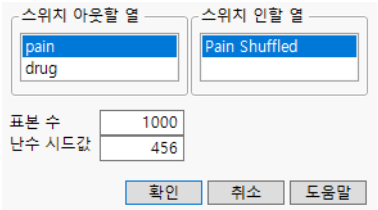
소스	DF	제곱합	평균 제곱	F 비	Prob > F
drug	2	99.89459	49.9473	6.2780	0.0053*
오차	30	238.67877	7.9560		
수정 합계	32	338.57335			

F 비가 6.2780 입니다 .

6. " 분산 분석 " 개요에서 "F 비 " 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .
7. " 스위치 아웃할 열 " 목록에서 pain 을 클릭합니다 .
8. " 스위치 인할 열 " 목록에서 Pain Shuffled 를 클릭합니다 .
9. **표본 수** 옆에 1000 을 입력합니다 .
10. (선택 사항) **난수 시드값** 옆에 456 을 입력합니다 .

    이렇게 하면 이 예의 값이 재현됩니다 .

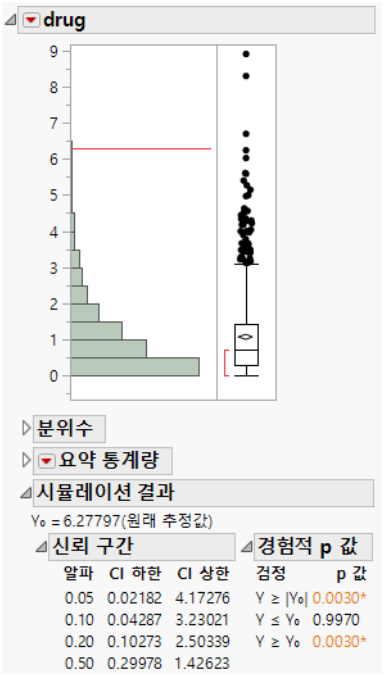
그림 10.10 완료된 시뮬레이션 창



11. **확인**을 클릭합니다 .

    시뮬레이션 결과 테이블에서 " 수정 합계 " 및 " 오차 " 열은 비어 있습니다 . 이는 " 분산 분석 " 테이블의 *F* 비 값이 drug 에만 적용되기 때문입니다 .
12. 시뮬레이션된 값 테이블에서 **분포** 스크립트를 실행합니다 .

그림 10.11 귀무 분포하에서 시뮬레이션된 *F* 비 분포



관측된 *F* 비 값 6.2780 은 히스토그램에 빨간색 선으로 표시됩니다 . 이 값은 시뮬레이션된 *F* 비 귀무 분포의 상위 0.5% 에 속합니다 . 이는 세 가지 약물이 pain 에 미치는 효과가 다르다는 강력한 증거를 제시합니다 .

JMP PRO 일반화 회귀의 요인 유지 탐색

이 예에서는 일반화 회귀 모형을 구성하고 활성 모수를 사용하여 축소 모형을 적합시킵니다. 이 축소 모형에 따라 시뮬레이션을 사용하여 모형에 포함될 요인 중 하나인 가능도를 구합니다.

한 제약 회사가 정제의 체내 용해도와 용해도에 영향을 미칠 수 있는 다양한 요인에 관한 과거 정보를 갖고 있다고 가정합니다. 용해도가 70 미만인 정제는 결함이 있는 것으로 간주됩니다. 따라서 용해도에 영향을 미치는 요인을 파악하려고 합니다.

JMP PRO 모형 적합

이 섹션에서는 일반화 회귀를 사용하여 모형을 적합시킵니다 .

팁 : 이 섹션의 단계를 거치지 않으려면 Tablet Production.jmp 데이터 테이블의 **Generalized Regression** 스크립트 옆에 있는 녹색 삼각형을 클릭하여 모형을 구하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Tablet Production.jmp 를 엽니다 .

2. **분석 > 모형 적합**을 선택합니다 .
3. Dissolution 을 클릭하고 **Y** 를 클릭합니다 .
4. Mill Time 부터 Atomizer Pressure 까지 선택하고 **추가**를 클릭합니다 .
5. " 분석법 " 목록에서 **일반화 회귀**를 선택합니다 .
6. **실행**을 클릭합니다 .
7. " 모형 시작 " 패널에서 **적응형** 상자를 선택합니다 .
8. " 모형 시작 " 패널에서 **시작**을 클릭합니다 .

그림 10.12 적응형 Lasso 회귀 기반의 모형

원래 예측 변수에 대한 모수 추정값

항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	108.6406	24.332975	19.933977	<.0001*	60.948851	156.33236
Mill Time	0.130278	0.028185	21.365139	<.0001*	0.0750364	0.1855195
Screen Size[3-5]	4.1877616	0.541493	59.8106	<.0001*	3.1264548	5.2490685
Screen Size[4-5]	2.3907729	0.567217	17.765534	<.0001*	1.2790481	3.5024977
Mag. Stearate Supplier[Jones Inc-Smith Ind]	0	0	0	1.0000	0	0
Lactose Supplier[Bond Inc-James Ind]	0	0	0	1.0000	0	0
Sugar Supplier[Sour-Sweet]	0	0	0	1.0000	0	0
Talc Supplier[Rough-Smooth]	0	0	0	1.0000	0	0
Blend Time	0.7223423	0.1381848	27.325338	<.0001*	0.4515051	0.9931796
Blend Speed	0.2130888	0.2389205	0.7954524	0.3725	-0.255187	0.6813644
Compressor[Compress1-Compress2]	-0.538112	0.3993673	1.8155148	0.1778	-1.320857	0.2446338
Force	0	0	0	1.0000	0	0
Coating Supplier[Coat-Mac]	0	0	0	1.0000	0	0
Coating Supplier[Down-Mac]	0	0	0	1.0000	0	0
Coating Viscosity	0.1834341	0.0500838	13.414242	0.0002*	0.0852717	0.2815965
Inlet Temp	0	0	0	1.0000	0	0
Exhaust Temp	0	0	0	1.0000	0	0
Spray Rate	-0.204356	0.0420142	23.658267	<.0001*	-0.286702	-0.12201
Atomizer Pressure	0	0	0	1.0000	0	0

정규 분포 모수	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
척도	2.0224259	0.1641142	151.86339	<.0001*	1.700768	2.3440838

"AICc 검증을 사용한 정규 적응형 Lasso 회귀 " 보고서에 표시된 모수 추정값에 관심이 있습니다 . 0 이 아닌 모수 추정값에 따라 이 모형은 Mill Time, Screen Size, Blend Time, Blend Speed, Compressor, Coating Viscosity 및 Spray Rate 가 Dissolution 과 관련이 있음을 나타냅니다 .

JMP PRO 모형 축소

모형을 축소하기 전에 Tablet Production.jmp 데이터 테이블에 선택된 열이 없는지 확인하십시오. 선택된 열은 아래의 첫 번째 단계에서 선택 취소되지 않습니다. 선택된 열이 없도록 하면 0 인 항이 있는 열이 실수로 포함되는 경우를 방지할 수 있습니다.

이 섹션의 단계를 거치지 않으려면 Tablet Production.jmp 데이터 테이블의 **Generalized Regression Reduced Model** 스크립트 옆에 있는 녹색 삼각형을 클릭하여 축소 모형을 구하십시오 .

1. "AICc 검증을 사용한 정규 적응형 Lasso 회귀 " 옆의 빨간색 삼각형을 클릭하고 **활성 집합 다시 시작 > 활성 효과를 사용하여 다시 시작**을 선택합니다 .

   이렇게 하면 " 모수 추정값 " 보고서에서 0 이 아닌 계수 추정값이 있는 항이 " 모형 효과 생성 " 목록에 포함됩니다 . 반응은 Y 로 입력되며 , " 일반화 회귀 " 분석법이 선택됩니다 .
2. **실행**을 클릭합니다 .
3. " 모형 시작 " 패널에서 **적응형** 상자를 선택합니다 .
4. " 모형 시작 " 패널에서 **시작**을 클릭합니다 .

그림 10.13 적응형 Lasso 회귀를 사용한 축소 모형

원래 예측 변수에 대한 모수 추정값

항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	95.142391	23.430178	16.489099	<.0001*	49.220085	141.0647
Mill Time	0.1394103	0.0275609	25.585971	<.0001*	0.0853919	0.1934288
Screen Size[3-5]	4.3323833	0.534237	65.763638	<.0001*	3.285298	5.3794685
Screen Size[4-5]	2.6331283	0.5457852	23.275584	<.0001*	1.563409	3.7028476
Blend Time	0.7583048	0.1385246	29.966364	<.0001*	0.4868017	1.029808
Blend Speed	0.4575802	0.2289168	3.9955744	0.0456*	0.0089116	0.9062488
Compressor[Compress1-Compress2]	-0.877986	0.4127286	4.5252866	0.0334*	-1.686919	-0.069053
Coating Viscosity	0.198673	0.0486798	16.656386	<.0001*	0.1032624	0.2940836
Spray Rate	-0.212753	0.0410929	26.805238	<.0001*	-0.293294	-0.132213

정규 분포 모수	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
척도	1.9982636	0.1574951	160.97992	<.0001*	1.689579	2.3069483

Blend Speed 추정값의 신뢰 구간(95% 하한)은 0을 포함하는 것과 매우 가깝습니다. 다음으로는 시뮬레이션 연구를 수행하여 용해도 분포에서 다른 데이터 값이 관측될 때 Blend Speed가 얼마나 자주 모형에 포함되는지 알아봅니다.

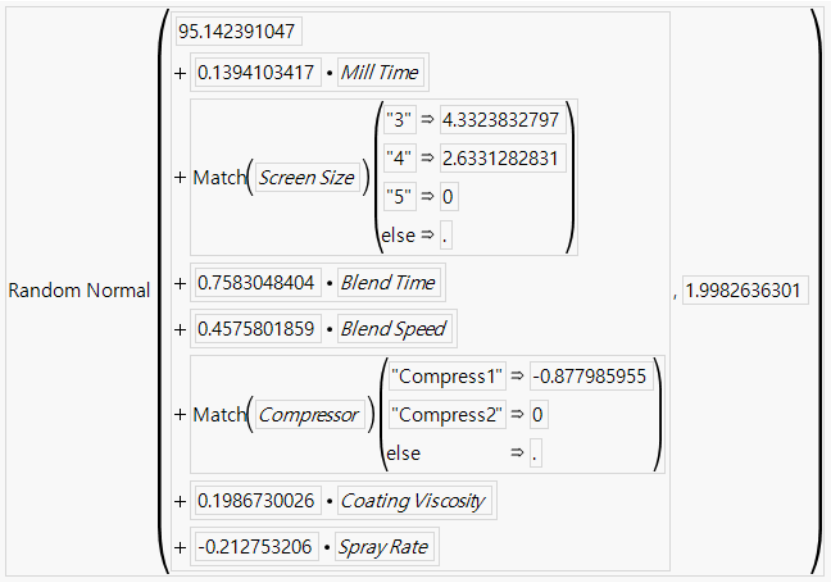
JMP PRO 모형의 Blend Speed 포함 여부 탐색

아래 단계에서는 축소 모형에 대한 보고서 (그림 10.13) 를 사용합니다 .

1. "AICc 검증을 사용한 정규 적응형 Lasso 회귀 " 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 시뮬레이션 계산식 저장**을 선택합니다 .

   이렇게 하면 Dissolution 시뮬레이션 계산식이라는 새 열이 Tablet Production.jmp 데이터 테이블에 추가됩니다 .
2. (선택 사항) 데이터 테이블의 " 열 " 패널에서 Dissolution 시뮬레이션 계산식의 오른쪽에 있는 더하기 기호를 클릭합니다 .

그림 10.14 시뮬레이션 계산식



이 계산식은 각 행에 대해서 주어진 모형과 Dissolution 의 분포 (표준편차가 약 1.998 인 정규 분포로 추정) 로 구할 수 있는 값을 시뮬레이션합니다 .

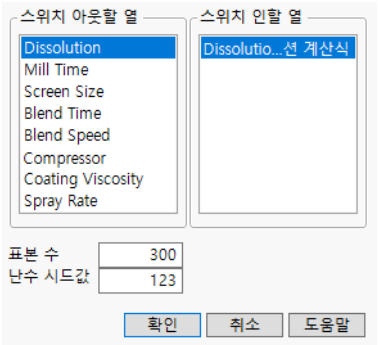
3. **취소**를 클릭합니다 .
4. 축소 모형 보고서 창으로 돌아갑니다 . " 원래 예측 변수에 대한 모수 추정값 " 보고서에서 " 추정값 " 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .

   " 스위치 아웃할 열 " 목록에 Dissolution 이 선택되어 있는지 확인합니다 .
5. **표본 수** 옆에 300 을 입력합니다 .

   이 시뮬레이션에서는 300 개의 각 분석값에서 Dissolution 열을 Dissolution 시뮬레이션 계산식 열을 사용하여 시뮬레이션된 값으로 대체하도록 요청합니다 .
6. (선택 사항) **난수 시드값**을 123 으로 설정합니다 .

   이렇게 하면 이 예의 값이 재현됩니다 .

그림 10.15 완료된 시뮬레이션 창

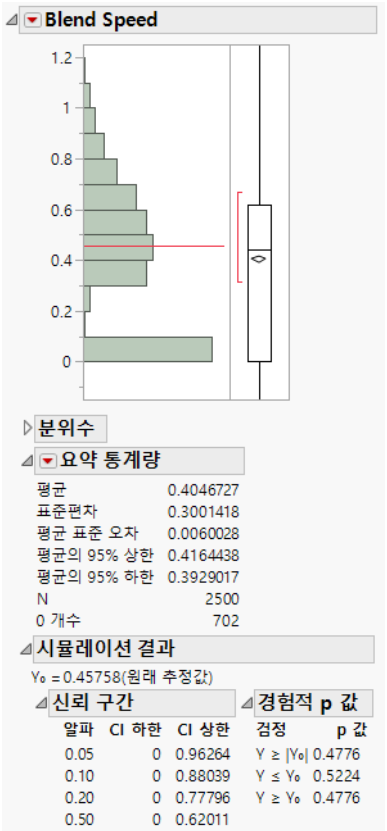


7. **확인**을 클릭합니다 .

   테이블의 첫 번째 행은 추정값의 초기 값을 포함하므로 제외됩니다. 나머지 행에는 시뮬레이션된 값이 포함되어 있습니다 .

8. **Distribution** 스크립트를 실행합니다 .
9. Ctrl 키를 누른 채 Blend Speed 빨간색 삼각형을 클릭하고 **표시 옵션 > 요약 통계량 사용자 정의를 선택합니다** .
10. **0 개수**를 선택합니다 .
11. **확인**을 클릭합니다 .
12. Blend Speed 에 대한 분포 보고서로 스크롤합니다 .

그림 10.16 시뮬레이션된 Blend Speed 계수 추정값의 히스토그램



이 " 요약 통계량 " 보고서에서는 시뮬레이션의 103/300 = 34.3% 에 대해 Blend Speed 추정값이 0 임을 보여 줍니다 .

JMP PRO 비선형 모형에 대한 전향적 검정력 분석

이 예에서는 6가지 연속형 요인이 부품의 검사 합격 또는 불합격에 미치는 주효과를 알아보려고 합니다. 반응은 이항이며 총 60회 런이 가능합니다.

예제용 계획

로짓을 연결 함수로 사용한 일반화 선형 모형을 사용하여 성공 확률을 모델링합니다 . 로짓 연결 함수는 로지스틱 모형을 적합시킵니다 .

$$\pi(\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \ldots + \beta_6 X_6)}}$$

여기서 $\pi(\mathbf{X})$는 주어진 설계 설정 $\mathbf{X} = (X1, X2, ..., X6)$에서 부품이 검사에 합격할 확률을 나타냅니다.

선형 예측 변수는 $L(\mathbf{X})$ 로 나타냅니다.

$$L(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \ldots + \beta_6 X_6$$

다음으로는 선형 예측 변수의 다음 계수 값에 대한 검정력을 구합니다.

표 10.1

계수	값
β_0	0
β_1	1
β_2	0.9
β_3	0.8
β_4	0.7
β_5	0.6
β_6	0.5

선형 예측 변수의 절편이 0이므로 모든 요인이 0으로 설정될 경우 부품의 합격 확률은 50%가 됩니다. 다른 모든 요인이 0으로 설정될 경우 i^t번째 요인의 수준과 연관된 확률은 아래와 같습니다.

표 10.2

요인	$\mathbf{X}_i = 1$ 일 때의 합격률	$\mathbf{X}_i = -1$ 일 때의 합격률	차이
X_1	73.11%	26.89%	46.2%
X_2	71.09%	28.91%	42.2%
X_3	69.00%	31.00%	38.0%
X_4	66.82%	33.18%	33.6%
X_5	64.56%	35.43%	29.1%
X_6	62.25%	37.75%	24.5%

예를 들어 X_1을 제외한 모든 요인이 0으로 설정될 경우에 검출하려는 합격률은 46.2%입니다. 검출하려는 합격률의 최소 차이는 X_6을 제외한 모든 요인이 0으로 설정되고 해당 차이가 24.5%일 때 발생합니다.

JMP PRO 설계 구성

이 섹션에서는 실험을 위한 사용자 설계를 구성합니다.

참고 : 이 섹션의 단계를 건너뛰려면 **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Binomial Experiment.jmp 를 여십시오. **DOE Simulate** 스크립트 옆의 녹색 삼각형을 클릭한 다음 " 시뮬레이션된 반응 정의 "(332 페이지) 절차로 이동하십시오.

1. **DOE > 사용자 설계**를 선택합니다.

   참고 : 사용자 설계는 비선형 상황에는 최적이 아니지만 이 예에서는 간소화를 위해 비선형 설계 플랫폼 대신 사용자 설계 플랫폼을 사용할 수 있습니다. 비선형 설계 플랫폼을 사용하여 구성한 설계가 직교 설계보다 나은 이유를 보여 주는 예는 **실험 설계 가이드**의 에서 확인하십시오.

2. " 요인 " 개요에서 **N 개 요인 추가** 옆에 6 을 입력합니다.
3. **요인 추가 > 연속형**을 클릭합니다.
4. **계속**을 클릭합니다.

   주효과 설계를 구성 중이므로 " 모형 " 개요의 다른 사항은 변경하지 마십시오.

5. " 런 수 " 아래에서 **사용자 지정** 옆에 60 을 입력합니다.
6. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **반응 시뮬레이션**을 선택합니다.

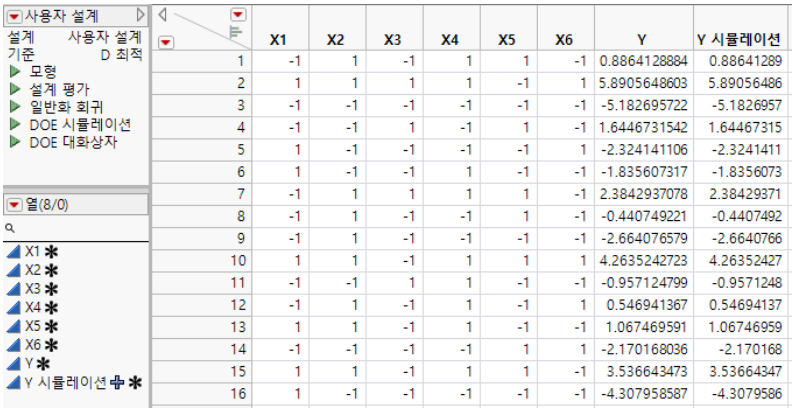
   이렇게 하면 " 테이블 생성 " 을 선택하여 설계 테이블을 구성한 후에 " 반응 시뮬레이션 " 창이 열립니다.

   참고 : 7 단계에서 " 난수 시드값 " 을 설정하고 8 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 설계가 재현됩니다. 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다.

7. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다. "12345" 를 입력하고 **확인**을 클릭합니다.
8. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **시작 수**를 선택합니다. 1 을 입력하고 **확인**을 클릭합니다.
9. **설계 생성**을 클릭합니다.
10. **테이블 생성**을 클릭합니다.

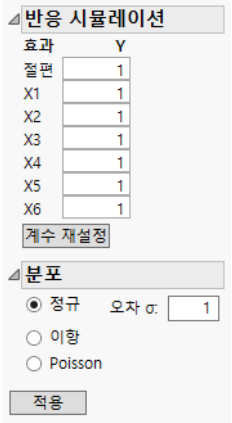
    참고 : Y 및 Y 시뮬레이션 열의 항목은 그림 10.17 에 표시된 것과 다를 수 있습니다.

그림 10.17 설계 테이블 (일부분)



	X1	X2	X3	X4	X5	X6	Y	Y 시뮬레이션
1	-1	1	-1	1	1	-1	0.8864128884	0.88641289
2	1	1	1	1	-1	1	5.8905648603	5.89056486
3	-1	-1	-1	-1	-1	-1	-5.182695722	-5.1826957
4	-1	-1	1	-1	1	-1	1.6446731542	1.64467315
5	1	-1	-1	-1	-1	1	-2.324141106	-2.3241411
6	1	-1	-1	1	-1	-1	-1.835607317	-1.8356073
7	-1	1	1	1	1	-1	2.3842937078	2.38429371
8	-1	1	-1	-1	1	-1	-0.440749221	-0.4407492
9	-1	1	-1	-1	-1	-1	-2.664076579	-2.6640766
10	1	1	-1	1	1	1	4.2635242723	4.26352427
11	-1	-1	1	1	-1	-1	-0.957124799	-0.9571248
12	-1	1	-1	1	-1	1	0.546941367	0.54694137
13	1	1	-1	1	-1	-1	1.067469591	1.06746959
14	-1	-1	-1	-1	1	1	-2.170168036	-2.170168
15	1	1	-1	1	1	-1	3.536643473	3.53664347
16	1	-1	-1	-1	-1	-1	-4.307958587	-4.3079586

그림 10.18 반응 시뮬레이션 창



설계 테이블과 " 반응 시뮬레이션 " 창이 나타납니다 . 설계 테이블에 두 개의 열이 추가되었습니다 .

- Y 에는 " 반응 시뮬레이션 " 창에서 지정한 대로 시뮬레이션된 값 집합이 포함되어 있습니다 .
- Y 시뮬레이션에는 " 반응 시뮬레이션 " 창에서 지정한 모형 계산식을 사용하여 값을 계산하는 계산식이 포함되어 있습니다 . 계산식을 보려면 " 열 " 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다 .

다음 섹션으로 진행하여 이항 반응을 시뮬레이션하고 시뮬레이션된 반응에 일반화 선형 모형을 적합시키십시오.

JMP PRO 시뮬레이션된 반응 정의

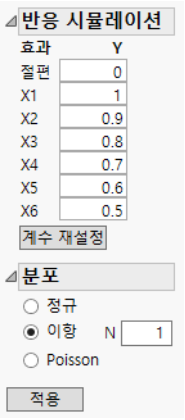
성공 확률이 로지스틱 모형으로 주어지는 이항 반응 데이터를 시뮬레이션할 계획입니다. 반응 시뮬레이션에 대한 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

참고 : 이 섹션의 단계를 건너뛰려면 **Simulate Model Responses** 스크립트 옆의 녹색 삼각형을 클릭하십시오 . 그런 다음 " 일반화 선형 모형 적합 "(333 페이지) 절차로 이동하십시오 .

1. " 반응 시뮬레이션 " 창 (그림 10.18) 에서 Y 아래에 다음 값을 입력합니다 .
   - " 절편 " 옆에 0 을 입력합니다 .
   - X1 옆에는 기본적으로 1 이 입력되어 있습니다 . 이 값을 유지합니다 .
   - X2 옆에 0.9 를 입력합니다 .
   - X3 옆에 0.8 을 입력합니다 .
   - X4 옆에 0.7 을 입력합니다 .
   - X5 옆에 0.6 을 입력합니다 .
   - X6 옆에 0.5 를 입력합니다 .
2. " 분포 " 개요에서 **이항**을 선택합니다 .

   N 값은 1 로 설정된 상태로 둡니다 . 이 값은 시도당 1 단위만 있음을 나타냅니다 .

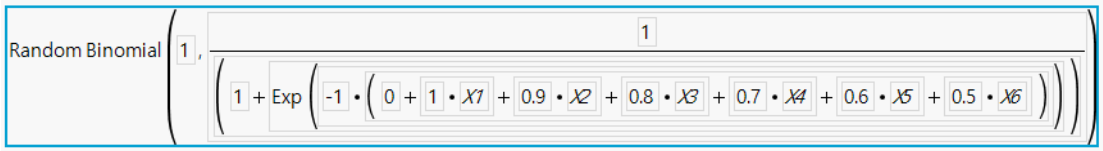
그림 10.19 완료된 반응 시뮬레이션 창



3. **적용**을 클릭합니다 .

   설계 데이터 테이블에서 Y 시뮬레이션 열이 이항 값을 생성하는 계산식 열로 대체됩니다 . Y 시도 수라는 열은 각 런의 시도 횟수를 나타냅니다 .
4. (선택 사항) " 열 " 패널에서 Y 시뮬레이션 오른쪽의 더하기 기호를 클릭합니다 .

그림 10.20 Y 시뮬레이션에 대한 랜덤 이항 계산식



5. **취소**를 클릭합니다 .

JMP PRO 일반화 선형 모형 적합

이 섹션에서는 일반화 선형 모형 분석법을 사용하여 로지스틱 모형을 적합시킵니다.

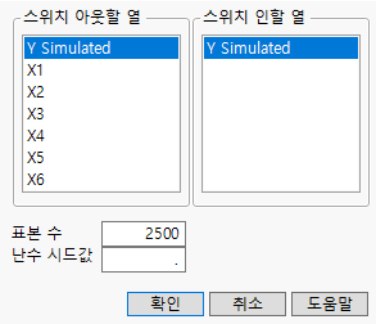
1. 데이터 테이블에서 **모형** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
2. **Y** 버튼 옆의 Y 변수를 클릭하고 **제거**를 클릭합니다 .
3. Y 시뮬레이션을 클릭하고 **Y** 버튼을 클릭합니다 .
   이렇게 하면 Y 가 무작위로 생성된 이항 값을 포함하는 열로 대체됩니다 .
4. " 분석법 " 목록에서 **일반화 선형 모형**을 선택합니다 .
5. " 분포 " 목록에서 **이항**을 선택합니다 .
   " 연결 함수 " 메뉴에 " 로짓 " 함수가 나타납니다 .
6. **실행**을 클릭합니다 .
   적합된 모형은 시뮬레이션된 이항 반응의 단일 집합을 기반으로 합니다 .

JMP PRO 검정력 분석

이 섹션에서는 가능도비 검정 p 값을 시뮬레이션하여 " 예제용 계획 "(328 페이지)에서 구한 계수 값이 활용된 선형 예측 으로부터 결정된 확률 값의 범위에서 차이를 검출하는 검정력을 구합니다.

1. " 효과 검정 " 개요에서 **Prob>ChiSq** 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .

그림 10.21 시뮬레이션 창



" 스위치 아웃할 열 " 목록에서 **Y 시뮬레이션** 열이 선택되어 있는지 확인하십시오 . 이 열에는 모형 적합에 사용된 값이 포함되어 있습니다 . 각 시뮬레이션에 대해 " 스위치 인할 열 " 에서 **Y 시뮬레이션** 열을 선택하면 **Y 시뮬레이션**의 값이 **Y 시뮬레이션** 열의 계산식을 사용하여 시뮬레이션된 값을 포함하는 새 열로 대체됩니다 .

보고서에서 선택한 **Prob>ChiSq** 열은 연관된 주효과가 0 인지 여부에 대한 가능도비 검정의 *p* 값입니다 . " 효과 검정 " 테이블에 나열된 각 효과에 대해 "Prob>ChiSq" 값이 시뮬레이션됩니다 .

2. **표본 수** 옆에 **500** 을 입력합니다 .
3. **확인**을 클릭합니다 .

" 일반화 선형 모형 시뮬레이션 결과 " 데이터 테이블이 나타납니다 .

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션된 *p* 값은 그림 10.22 에 표시된 것과 다를 수 있습니다 .

그림 10.22 시뮬레이션 결과 표 (일부분)

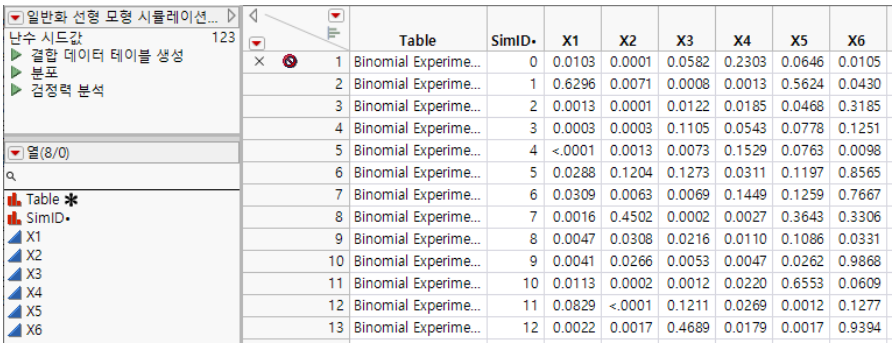


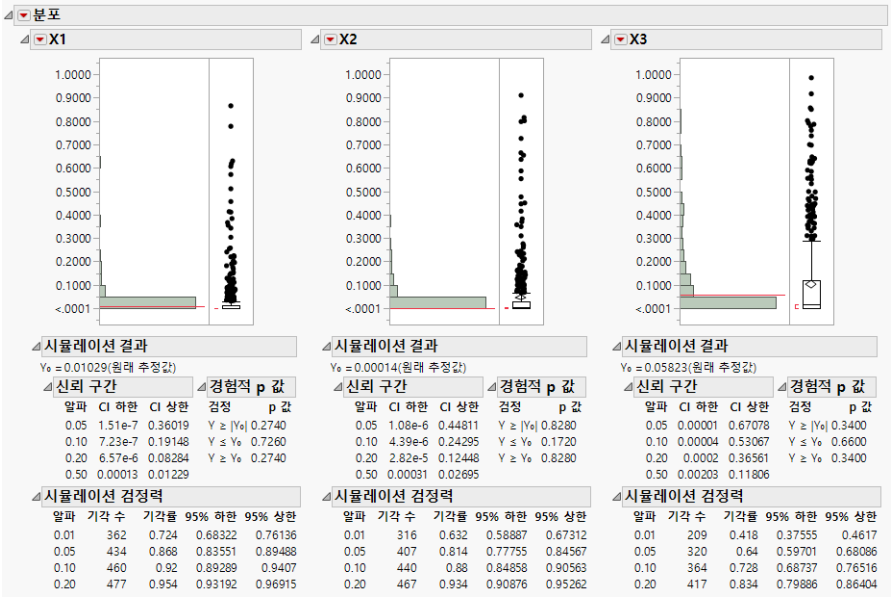
	Table	SimID•	X1	X2	X3	X4	X5	X6
1	Binomial Experime...	0	0.0103	0.0001	0.0582	0.2303	0.0646	0.0105
2	Binomial Experime...	1	0.6296	0.0071	0.0008	0.0013	0.5624	0.0430
3	Binomial Experime...	2	0.0013	0.0001	0.0122	0.0185	0.0468	0.3185
4	Binomial Experime...	3	0.0003	0.0003	0.1105	0.0543	0.0778	0.1251
5	Binomial Experime...	4	<.0001	0.0013	0.0073	0.1529	0.0763	0.0098
6	Binomial Experime...	5	0.0288	0.1204	0.1273	0.0311	0.1197	0.8565
7	Binomial Experime...	6	0.0309	0.0063	0.0069	0.1449	0.1259	0.7667
8	Binomial Experime...	7	0.0016	0.4502	0.0002	0.0027	0.3643	0.3306
9	Binomial Experime...	8	0.0047	0.0308	0.0216	0.0110	0.1086	0.0331
10	Binomial Experime...	9	0.0041	0.0266	0.0053	0.0047	0.0262	0.9868
11	Binomial Experime...	10	0.0113	0.0002	0.0012	0.0220	0.6553	0.0609
12	Binomial Experime...	11	0.0829	<.0001	0.1211	0.0269	0.0012	0.1277
13	Binomial Experime...	12	0.0022	0.0017	0.4689	0.0179	0.0017	0.9394

테이블의 첫 번째 행은 **Prob>ChiSq** 의 초기 값을 포함하므로 제외됩니다 . 나머지 500 개 행에는 시뮬레이션된 값이 포함되어 있습니다 .

4. **검정력 분석** 스크립트를 실행합니다 .

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 그림 10.23 에 표시된 것과 다를 수 있습니다 .

그림 10.23 처음 두 개의 효과에 대한 분포 그림



히스토그램에는 각 주효과에 대해 시뮬레이션된 500 개의 Prob>ChiSq 값이 표시됩니다 . " 시뮬레이션 검정력 " 개요에는 500 회의 시뮬레이션에서 시뮬레이션된 기각률이 표시됩니다 . 좀 더 보기 쉽게 보고서를 세로로 쌓고 그림을 선택 취소할 수 있습니다 .

5. " 분포 " 의 빨간색 삼각형을 클릭하고 **쌓기**를 선택합니다 .
6. Ctrl 키를 누른 채 X1 의 빨간색 삼각형을 클릭하고 **이상치 상자 그림**을 선택 취소합니다 .
7. Ctrl 키를 누른 채 X1 의 빨간색 삼각형을 클릭하고 **히스토그램 옵션**을 선택한 다음 **히스토그램**을 선택 취소합니다 .

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 그림 10.24 에 표시된 것과 다를 수 있습니다 .

그림 10.24 처음 세 개의 효과에 대한 검정력 결과

분포

X1

시뮬레이션 결과

Y_0 = 0.01029(원래 추정값)

신뢰 구간

알파	CI 하한	CI 상한
0.05	1.51e-7	0.36019
0.10	7.23e-7	0.19148
0.20	6.57e-6	0.08284
0.50	0.00013	0.01229

경험적 p 값

검정	p 값
Y ≥ |Y_0|	0.2740
Y ≤ Y_0	0.7260
Y ≥ Y_0	0.2740

시뮬레이션 검정력

알파	기각 수	기각률	95% 하한	95% 상한
0.01	362	0.724	0.68322	0.76136
0.05	434	0.868	0.83551	0.89488
0.10	460	0.92	0.89289	0.9407
0.20	477	0.954	0.93192	0.96915

X2

시뮬레이션 결과

Y_0 = 0.00014(원래 추정값)

신뢰 구간

알파	CI 하한	CI 상한
0.05	1.08e-6	0.44811
0.10	4.39e-6	0.24295
0.20	2.82e-5	0.12448
0.50	0.00031	0.02695

경험적 p 값

검정	p 값
Y ≥ |Y_0|	0.8280
Y ≤ Y_0	0.1720
Y ≥ Y_0	0.8280

시뮬레이션 검정력

알파	기각 수	기각률	95% 하한	95% 상한
0.01	316	0.632	0.58887	0.67312
0.05	407	0.814	0.77755	0.84567
0.10	440	0.88	0.84858	0.90563
0.20	467	0.934	0.90876	0.95262

X3

시뮬레이션 결과

Y_0 = 0.05823(원래 추정값)

신뢰 구간

알파	CI 하한	CI 상한
0.05	0.00001	0.67078
0.10	0.00004	0.53067
0.20	0.0002	0.36561
0.50	0.00203	0.11806

경험적 p 값

검정	p 값
Y ≥ |Y_0|	0.3400
Y ≤ Y_0	0.6600
Y ≥ Y_0	0.3400

시뮬레이션 검정력

알파	기각 수	기각률	95% 하한	95% 상한
0.01	209	0.418	0.37555	0.4617
0.05	320	0.64	0.59701	0.68086
0.10	364	0.728	0.68737	0.76516
0.20	417	0.834	0.79886	0.86404

" 시뮬레이션 검정력 " 개요에서 각 행의 " 기각률 " 은 해당 알파 값보다 작은 p 값의 비율을 나타냅니다 . 예를 들어 계수 값 0.8, 확률 차이 38% 에 해당하는 X3 의 경우 유의 수준 0.05 에 대한 시뮬레이션된 검정력은 379/500 = 0.758 입니다 . 표 10.3 에는 유의 수준 0.05 에서 모든 효과에 대해 추정된 검정력이 요약되어 있습니다 . 이를 통해 " 감지할 차이 " 가 감소할 때 검정력이 얼마나 감소하는지 알 수 있습니다 . 또한 24.5% 의 효과를 감지할 수 있는 검정력 (X6) 은 약 0.37 에 불과함을 알 수 있습니다 .

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 표 10.3 에 표시된 것과 다를 수 있습니다 .

표 10.3 유의 수준 0.05 에서의 시뮬레이션 검정력

요인	$X_i = 1$ 일 때의 합격률	$X_i = -1$ 일 때의 합격률	감지할 차이	알파 =0.05 일 때의 시뮬레이션 검정력 (기각률)
X_1	73.11%	26.89%	46.2%	0.852
X_2	71.09%	28.91%	42.2%	0.828
X_3	69.00%	31.00%	38.0%	0.758
X_4	66.82%	33.18%	33.6%	0.654

표 10.3 유의 수준 0.05 에서의 시뮬레이션 검정력 (계속)

요인	$X_i = 1$ 일 때의 합격률	$X_i = -1$ 일 때의 합격률	감지할 차이	알파 =0.05 일 때의 시뮬레이션 검정력 (기각률)
X_5	64.56%	35.43%	29.1%	0.488
X_6	62.25%	37.75%	24.5%	0.372

JMP PRO 시뮬레이션 창 시작

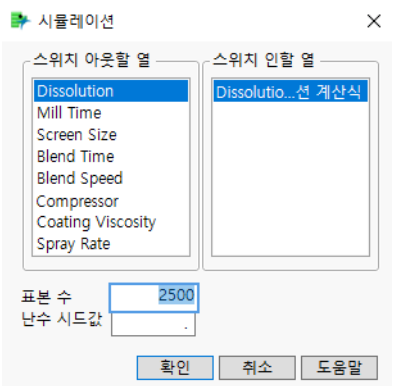
시뮬레이션 플랫폼을 시작하려면 보고서 창에서 계산된 값이 포함된 열을 마우스 오른쪽 버튼으로 클릭하고 "시뮬레이션"을 선택합니다. 시뮬레이션 플랫폼은 붓스트랩을 지원하는 모든 보고서를 포함하여 다양한 보고서에서 사용할 수 있습니다. 시뮬레이션 플랫폼을 사용하려면 데이터를 시뮬레이션하는 랜덤 성분이 있는 계산식이 데이터 테이블에 포함되어 있어야 합니다.

참고 : 기준 변수를 사용하는 보고서에서는 시뮬레이션 플랫폼을 사용할 수 없습니다 .

JMP PRO 시뮬레이션 창

"확인"을 클릭하여 시뮬레이션을 시작하면 진행률 표시줄과 "조기 중지" 버튼이 포함된 창이 나타납니다. 진행률 표시줄 위에는 시뮬레이션 중인 표본 수가 표시됩니다. "조기 중지"를 클릭하면 해당 시점까지 계산된 시뮬레이션 값이 "시뮬레이션 결과" 테이블에 표시됩니다. 이 창에는 특정 시점에 실행 중인 분석도 표시됩니다.

그림 10.25 Tablet Production.jmp 에 대한 시뮬레이션 창



" 시뮬레이션 " 창에는 다음과 같은 패널 및 옵션이 포함되어 있습니다 .

스위치 아웃할 열 ' 스위치 인할 열 " 로 대체되는 열입니다 .

스위치 인할 열 " 스위치 아웃할 열 " 을 대체하는 열입니다 . " 스위치 인할 열 " 의 계산식에 따라 시뮬레이션된 값을 사용하여 분석이 반복됩니다 . " 스위치 인할 열 " 패널에는 계산식이 있는 열만 나열됩니다 .

표본 수 시뮬레이션된 데이터 집합에 대해 보고서를 다시 실행할 횟수입니다 . 기본값은 2500 입니다 .

난수 시드값 시뮬레이션된 결과를 제어하는 값입니다 . 난수 시드값은 결과를 재현 가능하게 합니다 .

JMP PRO 시뮬레이션 결과 테이블

시뮬레이션 결과는 표로 나타납니다 . 다음 사항에 유의하십시오 .

- 테이블의 첫 번째 행에는 보고서에 표시된 테이블 항목에 대한 값이 포함됩니다 . 따라서 첫 번째 행은 항상 제외됩니다 .
- 나머지 행에서는 시뮬레이션 결과를 제공합니다 . 나머지 행의 개수는 시뮬레이션 시작 창에서 지정한 표본 수와 같습니다 .
- 보고서의 행은 선택된 계산 값 열이 포함된 보고서 테이블의 첫 번째 열로 식별됩니다 . 이 첫 번째 열에 있는 각 항목마다 하나씩의 열이 시뮬레이션 결과 테이블에 표시됩니다 .
- 이 테이블에는 " 분포 " 보고서를 생성하는 **분포** 스크립트가 포함되어 있습니다 . 이 보고서에는 시뮬레이션 결과 데이터 테이블의 각 열에 대한 히스토그램 , 분위수 , 요약 통계량 및 시뮬레이션 결과가 포함됩니다 . 표준 " 분포 " 보고서뿐만 아니라 다음 항목도 이 보고서에 포함됩니다 .
  - 히스토그램에 원래 추정값을 나타내는 빨간색 선이 표시됩니다 .
  - 원래 추정값과 함께 시뮬레이션에 대한 신뢰 구간 및 경험적 p 값이 포함된 " 시뮬레이션 결과 " 보고서 . 자세한 내용은 " 시뮬레이션 결과 보고서 "(338 페이지) 에서 확인하십시오 .
  - 시뮬레이션 결과 데이터 테이블의 값이 p 값 형식인 경우에는 " 시뮬레이션 검정력 " 보고서도 제공됩니다 . 자세한 내용은 " 시뮬레이션 검정력 보고서 "(339 페이지) 에서 확인하십시오 .
- p 값 열을 시뮬레이션한 경우에만 **검정력 분석** 스크립트가 테이블에 포함됩니다 . 이 스크립트는 p 값의 히스토그램을 보여 주는 " 분포 " 보고서를 생성하고 " 시뮬레이션 검정력 " 보고서를 제공합니다 . 자세한 내용은 " 시뮬레이션 검정력 보고서 "(339 페이지) 에서 확인하십시오 .

JMP PRO 시뮬레이션 결과 보고서

원래 추정값 원래 추정값의 값으로 , 시뮬레이션 결과 데이터 테이블의 첫 번째 행에도 표시됩니다 . 이 추정값에는 Y_0 이라는 라벨이 지정됩니다 .

신뢰 구간 유의 수준이 0.05, 0.10, 0.20 및 0.50 일 때 분위수 기반 신뢰 구간의 하한 및 상한입니다 .

경험적 p 값 시뮬레이션된 값을 원래 추정값과 비교하는 하나의 양측 검정과 두 개의 단측 검정에 대한 경험적 *p* 값입니다. 이러한 *p* 값은 보고서의 " 검정 " 열에 지정된 범위에 속하는 시뮬레이션된 값의 비율로 계산됩니다.

JMP PRO 시뮬레이션 검정력 보고서

알파 유의 수준으로 , 0.01, 0.05, 0.10 및 0.20 입니다.

기각 수 해당 유의 수준에서 검정이 기각되는 시뮬레이션의 수입니다.

기각률 해당 유의 수준에서 검정이 기각되는 시뮬레이션의 비율입니다.

95% 하한 및 95% 상한 시뮬레이션된 기각률에 대한 95% 신뢰 구간의 하한 및 상한입니다. 신뢰 구간은 Wilson 점수 방법을 사용하여 계산됩니다. 자세한 내용은 Wilson 연구 자료 (1927) 에서 확인하십시오.

팁 : 신뢰 구간을 좁히려면 표본 수를 늘리십시오.

11 장

JMP PRO 붓스트랩
재표집을 통한 통계량 분포 근사

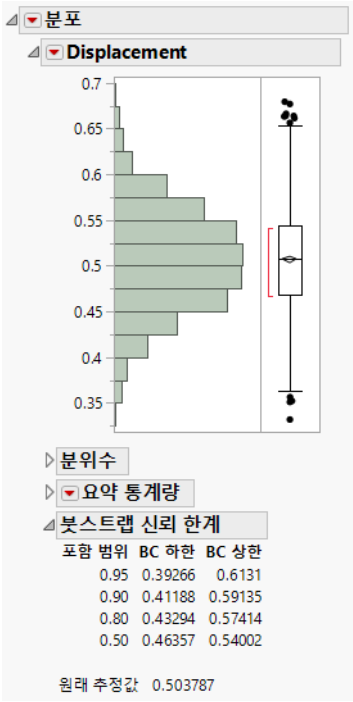
붓스트랩은 JMP Pro 에서만 사용할 수 있습니다.

붓스트랩은 통계량의 표본 분포에 근사한 값을 산출하기 위한 재표집 방법입니다. 붓스트랩을 사용하여 평균, 편향, 표준 오차 및 신뢰 구간 같은 통계량의 분포와 해당 특성을 추정할 수 있습니다. 붓스트랩은 다음과 같은 경우에 특히 유용합니다.

- 통계량의 이론상 분포가 복잡하거나 알 수 없는 경우
- 가정 위배로 인해 모수 방법을 사용한 추론이 가능하지 않은 경우

참고 : 붓스트랩은 보고서에서 마우스 오른쪽 버튼을 클릭하는 방법으로만 사용할 수 있습니다. 플랫폼 명령으로는 제공되지 않습니다.

그림 11.1 기울기 모수에 대한 붓스트랩 결과



목차

붓스트랩 개요 343
붓스트랩 지원 플랫폼 344
붓스트랩의 예 345
붓스트랩 창 옵션 347
누적 붓스트랩 결과 테이블 348
비누적 붓스트랩 결과 테이블 350
붓스트랩 결과 분석 350
붓스트랩의 추가 예 352
붓스트랩에 대한 통계 상세 정보 357
부분 가중치 계산 357
편향 수정 백분위수 구간 357

JMP PRO 붓스트랩 개요

붓스트랩은 보고서에 사용되는 관측값을 반복적으로 재표집하여 통계량의 분포를 추정하는 기법입니다. 이때 관측값은 독립적인 것으로 가정합니다.

단순 붓스트랩에서는 복원 표집 방법으로 n 개의 관측값을 재표집하여 크기가 n 인 붓스트랩 표본을 생성합니다 . 일부 관측값은 붓스트랩 표본에 나타나지 않을 수 있으며 , 다른 관측값은 여러 번 나타날 수 있습니다 . 하나의 관측값이 붓스트랩 표본에 나타나는 횟수를 **붓스트랩 가중치** 라고 합니다 . 붓스트랩 반복 시마다 관심 있는 통계량을 산출한 전체 분석이 다음과 같이 변경되어 다시 실행됩니다 .

- n 개의 관측값을 포함하는 붓스트랩 표본이 데이터 집합이 됩니다 .
- 붓스트랩 가중치는 분석 플랫폼의 빈도 변수입니다 .

관심 있는 통계량의 값 분포가 생성될 때까지 이 과정이 반복됩니다.

하지만 일부 경우에는 단순 붓스트랩이 부적절할 수 있습니다. 예를 들어 데이터 집합이 작거나 로지스틱 회귀를 사용하기 때문에 분리 문제가 발생할 수 있는 경우가 있습니다. 이러한 경우 JMP에서는 부분 가중치를 사용하여 베이지안 붓스트랩을 수행할 수 있습니다. 부분 가중치가 사용되는 경우에는 각 관측값에 부분 가중치가 연관됩니다. 부분 가중치의 합은 n입니다. 관심 있는 통계량을 계산하려면 분석 플랫폼에서 부분 가중치를 빈도 변수로 처리합니다. 부분 가중치에 대한 자세한 내용은 "부분 가중치"(348 페이지) 및 "부분 가중치 계산"(357 페이지)에서 확인하십시오.

보고서에서 붓스트랩 분석을 실행하려면 붓스트랩을 수행할 통계량이 포함된 테이블 열을 마우스 오른쪽 버튼으로 클릭하고 "붓스트랩"을 선택합니다.

참고 : 붓스트랩은 보고서에서 마우스 오른쪽 버튼을 클릭하는 방법으로만 사용할 수 있습니다 . 플랫폼 명령으로는 제공되지 않습니다 .

JMP는 많은 통계 플랫폼에서 붓스트랩 기능을 제공합니다. 전체 목록은 "붓스트랩 지원 플랫폼"(344 페이지)에서 확인하십시오. 표본을 구성하는 관측값은 관심 있는 통계량을 계산하는 데 사용된 모든 관측값입니다. 보고서에 빈도 열이 사용된 경우 해당 열의 관측값은 빈도 변수로 지정된 횟수만큼 반복된 것처럼 처리됩니다. 보고서에 가중치 변수가 사용된 경우 붓스트랩을 사용하면 해당 변수는 보고서의 계산에서 처리되던 방식대로 처리됩니다.

팁 : 붓스트랩 기능은 붓스트랩이 호출된 플랫폼 보고서에 표시된 전체 분석을 다시 실행합니다 . 따라서 보고서에 포함된 관련 없는 분석으로 인해 선택한 열에 대한 붓스트랩이 느리게 실행될 수 있습니다 . 붓스트랩이 느리게 실행되는 경우에는 플랫폼 보고서에서 관련 없는 옵션을 제거한 후 붓스트랩을 실행하십시오 .

붓스트랩 지원 플랫폼

붓스트랩은 다음 통계 플랫폼에서 사용할 수 있습니다 .

- 부스티드 트리
- 붓스트랩 포레스트
- 범주형
- 파괴 열화
- 판별
- 분포
- 곡선 적합
- 수명 분포 적합
- 모수 생존 모형 적합
- 비례 위험 모형 적합
- X 로 Y 적합
- 일반화 선형 모형
- 일반화 회귀
- 수명 분포
- 로지스틱
- 로그 선형 분산
- 다중 대응 분석
- 다변량
- 신경망
- 비선형
- 모수 생존
- 부분 최소 제곱
- 파티션
- 주성분
- 비례 위험
- 표준 최소 제곱
- 생존
- Uplift

JMP PRO 붓스트랩의 예

이 예에서는 Car Physical Data.jmp 샘플 데이터 테이블을 사용합니다. 한 타이어 제조업체에서 엔진 배기량(in^3)으로부터 엔진 출력(마력)을 예측하려고 합니다. 이 회사는 두 변수 간 관계의 기울기를 추정하는 데 가장 관심이 있습니다. 기울기 값은 배기량이 변화할 때 이에 대응하는 출력의 변화를 예측하는 데 유용합니다.

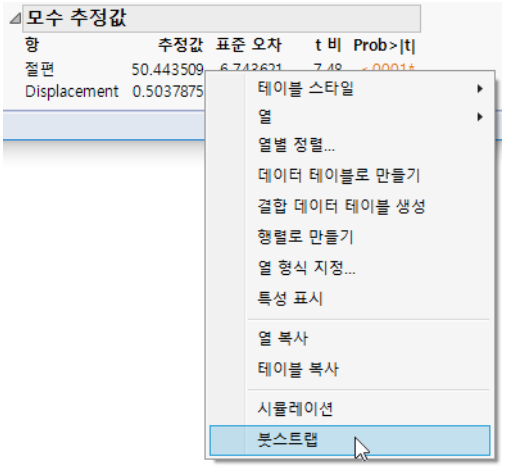
이 예에서는 분산 동질성에 대한 회귀 가정이 위배되므로 기울기에 대한 회귀 분석의 신뢰 한계가 혼동을 일으킬 수 있습니다. 따라서 이 회사에서는 기울기의 신뢰 구간에 대한 붓스트랩 추정값을 사용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Physical Data.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. Horsepower 를 선택하고 **Y, 반응**을 클릭합니다.
4. Displacement 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "Horsepower 대 Displacement 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.

   기울기 추정값은 0.503787, 대략적으로는 0.504 입니다.
7. (선택 사항) **모수 추정값** 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 하한**을 선택합니다.
8. (선택 사항) **모수 추정값** 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 상한**을 선택합니다.

   기울기에 대한 회귀 분석으로 구한 신뢰 한계는 0.4249038 과 0.5826711 입니다.
9. **모수 추정값** 보고서의 **추정값** 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다.

그림 11.2 붓스트랩 옵션



마우스 오른쪽 버튼으로 클릭한 열은 **선택한 열 분할** 옵션을 선택할 때도 관련됩니다. 자세한 내용은 " 붓스트랩 창 옵션 "(347 페이지) 에서 확인하십시오 .

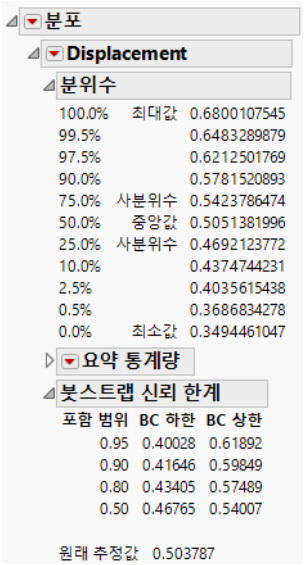
10. **붓스트랩 표본 수**에 1000 을 입력합니다 .
11. (선택 사항) 그림 11.3 의 결과와 매칭되는 결과를 얻으려면 **난수 시드값**에 12345 를 입력합니다 .
12. **확인**을 클릭합니다 .

    붓스트랩 프로세스가 실행된 후 붓스트랩 결과 데이터 테이블이 생성됩니다. 이때 기울기 및 절편에 대한 결과는 쌓이지 않은 상태입니다 .

    다음으로 , 붓스트랩 기울기를 분석합니다 .
13. 붓스트랩 결과 테이블에서 **분포** 스크립트를 실행합니다 .

    " 분포 " 보고서에는 " 붓스트랩 신뢰 한계 " 보고서가 포함됩니다 .

그림 11.3 붓스트랩 보고서



기울기(6단계)의 추정값은 0.504입니다. 95% 포함 범위에 대한 붓스트랩 결과에 따르면 이 회사에서는 기울기가 0.40028에서 0.61892 사이일 것이라고 추정할 수 있습니다. 즉, 배기량이 한 단위 변화할 때 출력이 0.40028 ~ 0.61892 정도로 변화함을 95% 신뢰도로 추정할 수 있습니다. 기울기에 대한 붓스트랩 신뢰 구간(0.400 ~ 0.619)은 7단계 및 8단계의 일반적인 회귀 가정을 사용하여 얻은 신뢰 구간(0.425 ~ 0.583)보다 다소 넓습니다.

참고 : " 붓스트랩 신뢰 한계 " 보고서의 "BC 하한 " 및 "BC 상한 " 열은 **편향 수정 구간**을 나타냅니다 . 자세한 내용은 " 편향 수정 백분위수 구간 "(357 페이지) 에서 확인하십시오 .

JMP PRO 붓스트랩 창 옵션

붓스트랩 분석을 수행하려면 보고서 창의 테이블에서 숫자로 된 표본 통계량 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다. 그러면 선택한 열이 강조 표시되고 "붓스트랩" 창이 나타납니다. "붓스트랩" 창에서 옵션을 선택하고 **확인**을 클릭하면 해당 열의 모든 통계량에 대한 붓스트랩 결과가 기본 결과 테이블에 나타납니다.

참고 : 기준 변수를 사용하는 보고서에서는 붓스트랩 옵션을 사용할 수 없습니다 .

" 붓스트랩 " 창에는 다음 옵션이 포함되어 있습니다 .

붓스트랩 표본 수 데이터를 재표집하여 통계량을 계산할 횟수를 설정합니다 . 값이 클수록 해당 통계량 특성을 더 정확하게 추정할 수 있습니다 . 기본적으로 붓스트랩 표본 수는 2500 으로 설정됩니다 .

난수 시드값 현재 결과를 복제하기 위해 붓스트랩 분석의 이후 런에서 다시 입력할 수 있는 난수 시드값을 설정합니다. 기본적으로 시드값은 설정되어 있지 않습니다.

부분 가중치 베이지안 붓스트랩 분석을 수행합니다. 붓스트랩 반복 시마다 "부분 가중치 계산"(357 페이지)에 설명된 대로 계산된 가중치가 각 관측값에 할당됩니다. 가중치가 적용된 관측값은 관심 있는 통계량을 계산하는 데 사용됩니다. 기본적으로는 "부분 가중치" 옵션이 선택되어 있지 않으며 단순 붓스트랩 분석이 수행됩니다.

팁 : 분석에 사용되는 관측값의 개수가 적거나 로지스틱 회귀 설정 시 분리가 우려되는 경우에 "부분 가중치" 옵션을 사용하십시오.

"부분 가중치" 옵션이 선택된 경우에는 붓스트랩 반복 시마다 보고서에 사용되는 각 관측값에 0 이 아닌 가중치가 할당됩니다. 이러한 가중치의 합은 *n*, 즉 관심 있는 통계량의 계산에 사용된 관측값 수가 됩니다. 가중치의 계산 및 사용 방법에 대한 자세한 내용은 "부분 가중치 계산"(357 페이지)에서 확인하십시오.

선택한 열 분할 붓스트랩을 위해 선택한 열의 각 통계량에 대한 붓스트랩 결과를 붓스트랩 결과 테이블에 **개별** 열로 배치합니다. 붓스트랩 결과 테이블의 각 행(첫 번째 행 제외)은 단일 붓스트랩 표본에 해당합니다.

이 옵션을 선택 취소하면 누적 붓스트랩 결과 테이블이 나타납니다. 붓스트랩 반복 시마다 붓스트랩을 위해 선택한 열이 포함된 전체 보고서 테이블에 대한 결과가 이 테이블에 포함됩니다. 보고서 테이블의 각 행에 대한 결과는 누적 붓스트랩 결과 테이블에 행으로 나타납니다. 보고서 테이블의 각 열은 누적 붓스트랩 결과 테이블의 열 하나를 정의합니다. 예는 "누적 붓스트랩 결과 테이블"(348 페이지)에서 확인하십시오.

분할 적용 시 쌓인 테이블 삭제 (**선택한 열 분할** 옵션이 선택된 경우에만 사용 가능) 붓스트랩을 통해 생성되는 결과 테이블의 수를 결정합니다.

"분할 적용 시 쌓인 테이블 삭제" 옵션이 선택되어 있지 **않으면** 다음과 같은 두 개의 붓스트랩 테이블이 표시됩니다.

- 누적 붓스트랩 결과 테이블 - 붓스트랩을 위해 선택한 열이 포함된 테이블의 각 행에 대한 붓스트랩 결과가 포함됩니다. 이 테이블에서는 보고서의 모든 통계량에 대한 붓스트랩 결과를 제공하며 각 열은 하나의 통계량에 의해 정의됩니다.
- 비누적 붓스트랩 결과 테이블 - 누적 테이블을 분할하여 생성한 테이블입니다. 이 테이블에서는 원래 보고서에서 선택한 열에 대한 결과만 제공합니다.

"분할 적용 시 쌓인 테이블 삭제" 옵션이 **선택되어** 있고 **선택한 열 분할** 작업에 성공한 경우에는 누적 붓스트랩 결과 테이블이 표시되지 않습니다.

JMP PRO 누적 붓스트랩 결과 테이블

붓스트랩 분석의 초기 결과는 누적 결과 테이블로 나타납니다(그림 11.4). **분할 적용 시 쌓인 테이블 삭제** 옵션을 선택한 경우에는 이 테이블이 나타나지 않습니다. 그림 11.4 에서는 Car Physical Data.jmp 에 "X 로 Y 적합"의 이변량 모형을 적합시켜서 생성된 "모수 추정값" 보고

서를 기반으로 한 붓스트랩 테이블을 보여 줍니다. 자세한 내용은 "붓스트랩 개요"(343 페이지)에서 확인하십시오.

그림 11.4 누적 붓스트랩 결과 테이블

	X	Y	항	~편향	추정값	표준 오차	t 비	Prob>|t|	BootID•
1	Displacement	Horsepower	절편		50.443509471	6.7436209999	7.48	<.0001	0
2	Displacement	Horsepower	Displacement		0.5037874592	0.0398202611	12.65	<.0001	0
3	Displacement	Horsepower	절편		45.385604758	5.4724273046	8.29	<.0001	1
4	Displacement	Horsepower	Displacement		0.5451674092	0.0306151772	17.81	<.0001	1
5	Displacement	Horsepower	절편		40.843862813	7.0508187326	5.79	<.0001	2
6	Displacement	Horsepower	Displacement		0.5854988173	0.0457041828	12.81	<.0001	2
7	Displacement	Horsepower	절편		47.765642104	5.4677610087	8.74	<.0001	3
8	Displacement	Horsepower	Displacement		0.4943459579	0.0305765677	16.17	<.0001	3
9	Displacement	Horsepower	절편		59.758060908	7.0436720036	8.48	<.0001	4
10	Displacement	Horsepower	Displacement		0.4385540785	0.042053201	10.43	<.0001	4
11	Displacement	Horsepower	절편		66.341062413	6.4663383596	10.26	<.0001	5
12	Displacement	Horsepower	Displacement		0.3969640071	0.0393100058	10.10	<.0001	5
13	Displacement	Horsepower	절편		41.234989734	5.8543901826	7.04	<.0001	6
14	Displacement	Horsepower	Displacement		0.5723548142	0.034172131	16.75	<.0001	6
15	Displacement	Horsepower	절편		43.876867815	7.8085946052	5.62	<.0001	7
16	Displacement	Horsepower	Displacement		0.5731377467	0.0439823115	13.03	<.0001	7

누적 결과 테이블의 특징은 다음과 같습니다.

- 각 붓스트랩 표본마다 보고서 테이블의 첫 번째 열에 주어진 각 값에 대한 행이 하나씩 있습니다. 이러한 값은 보고서 테이블의 첫 번째 열 이름을 사용하는 열에 표시됩니다. 이 예에서는 각 붓스트랩 표본마다 **항** 열에 나타나는 각 **항**, 즉 " 절편 " 및 "Displacement" 에 대한 결과를 포함하는 행이 하나씩 있습니다.
- 분석에 사용된 데이터 테이블 열은 이 테이블에 나타납니다. 이 예에서는 X 가 Displacement 이고 Y 가 Horsepower 입니다.
- 붓스트랩을 수행하는 보고서 테이블의 각 열마다 하나씩의 열이 있습니다. 이 예에서는 **~ 편향**, **추정값**, **표준 오차**, t 비 및 Prob>|t| 열이 그러한 열에 해당합니다. **~ 편향**은 모수 추정값 중 하나가 편향되지 않은 한 "X 로 Y 적합 " 보고서에서는 숨겨지는 열입니다.
- BootID 열은 붓스트랩 표본을 식별합니다. BootID = 0 인 행은 원래 추정값에 해당합니다. 이러한 행은 X 로 표시되며 ' 제외됨 ' 행 상태를 갖습니다. 이 예에서는 각 붓스트랩 표본이 두 개의 행에 대한 결과, 즉 **절편**에 대한 결과와 Displacement 에 대한 결과를 계산하는 데 사용되었습니다.
- 데이터 테이블 이름에는 " 누적 붓스트랩 결과 " 가 포함됩니다.

선택한 열 분할 옵션을 선택한 경우에는 비누적 결과 테이블도 나타날 수 있습니다. 자세한 내용은 "비누적 붓스트랩 결과 테이블"(350 페이지)에서 확인하십시오.

JMP PRO 비누적 붓스트랩 결과 테이블

선택한 보고서 열에 대한 개별 열이 포함된 붓스트랩 테이블을 생성하려면 **선택한 열 분할**을 선택합니다. 각 열은 보고서 테이블의 항 하나에 해당합니다. 예를 들어 그림 11.5 에서는 그림 11.4 의 **추정값** 열이 **항**의 각 수준에 해당하는 두 개의 열 (Displacement 및 **절편**) 로 분할됩니다.

그림 11.5 비누적 붓스트랩 결과 테이블

Displacement별 Hors...
난수 시드값 12345
결합 데이터 테이블 생성
분포

열(6/0)
X ✱
Y
Table
BootID•
Displacement
절편

	X	Y	Table	BootID•	Displacement	절편
1	Displacement	Horsepower	Car Physical Data	0	0.5037874592	50.443509471
2	Displacement	Horsepower	Car Physical Data	1	0.5451674092	45.385604758
3	Displacement	Horsepower	Car Physical Data	2	0.5854988173	40.843862813
4	Displacement	Horsepower	Car Physical Data	3	0.4943459579	47.765642104
5	Displacement	Horsepower	Car Physical Data	4	0.4385540785	59.758060908
6	Displacement	Horsepower	Car Physical Data	5	0.3969640071	66.341062413
7	Displacement	Horsepower	Car Physical Data	6	0.5723548142	41.234989734
8	Displacement	Horsepower	Car Physical Data	7	0.5731377467	43.876867815
9	Displacement	Horsepower	Car Physical Data	8	0.637861847	33.125519594
10	Displacement	Horsepower	Car Physical Data	9	0.5441967613	45.293111747
11	Displacement	Horsepower	Car Physical Data	10	0.5423927826	45.87058516

비누적 결과 테이블의 특징은 다음과 같습니다.

- 각 붓스트랩 표본마다 하나의 행이 있습니다.
- 분석에 사용된 데이터 테이블 열은 이 테이블에 나타납니다. 이 예에서는 X 가 Displacement 이고 Y 가 Horsepower 입니다.
- 붓스트랩이 수행된 보고서의 각 행마다 하나의 열이 있습니다.
- " 붓스트랩 " 창에서 " 난수 시드값 " 을 지정한 경우에는 해당 값을 제공하는 " 난수 시드값 " 이라는 테이블 변수가 붓스트랩 결과 테이블에 포함됩니다.
- 비누적 붓스트랩 결과 테이블에는 " 소스 " 테이블 스크립트와 " 분포 " 테이블 스크립트가 포함됩니다. " 분포 " 테이블 스크립트를 사용하면 붓스트랩 표본을 기반으로 한 붓스트랩 신뢰구간 등의 통계량을 빠르게 구할 수 있습니다.
- BootID 열은 붓스트랩 표본을 식별합니다. BootID = 0 인 행은 원래 추정값에 해당합니다. 이 행은 X 로 표시되며 ' 제외됨 ' 행 상태를 갖습니다. 비누적 붓스트랩 테이블에서는 각 행이 단일 붓스트랩 표본으로부터 계산됩니다.
- 이 데이터 테이블의 이름은 " 붓스트랩 결과 (< 열 이름 >)" 로 끝납니다. 여기서 < 열 이름 > 은 붓스트랩을 수행한 보고서의 열을 식별합니다.

JMP PRO 붓스트랩 결과 분석

분포 플랫폼을 사용하여 붓스트랩 결과를 분석할 수 있습니다.

- 분석을 통해 비누적 붓스트랩 결과 테이블이 생성된 경우에는 테이블에서 " 분포 " 스크립트를 실행합니다.

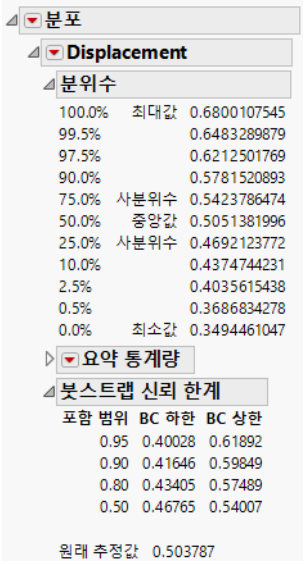
- 분석을 통해 누적 붓스트랩 결과 테이블이 생성된 경우에는 **분석 > 분포**를 선택하고 관심 있는 열을 적절한 역할에 할당합니다. 대부분의 경우 보고서 테이블의 첫 번째 열에 해당하는 열을 기준 역할에 할당하는 것이 좋습니다.

분포 플랫폼에서는 붓스트랩 결과에 대한 요약 통계량을 제공합니다. 또한 BootID 열이 포함된 테이블(그림 11.6)에 대한 "붓스트랩 신뢰 한계" 보고서도 생성됩니다.

"분포" 보고서를 사용하여 다음과 같은 두 가지 유형의 붓스트랩 신뢰 구간을 구할 수 있습니다.

- "분위수" 보고서에서는**백분위수 구간**을 제공합니다. 예를 들어 백분위수 방법을 사용하여 95% 신뢰 구간을 구성하려면 2.5% 및 97.5% 분위수를 구간 경계로 사용합니다.
- "붓스트랩 신뢰 한계" 보고서에서는 **편향 수정 백분위수 구간**을 제공합니다. 이 보고서에는 포함 수준이 95%, 90%, 80% 및 50% 인 구간이 표시됩니다. "BC 하한" 및 "BC 상한" 열에는 각각 하한점과 상한점이 표시됩니다. 편향 수정 백분위수 구간의 계산에 대한 자세한 내용은 "편향 수정 백분위수 구간"(357 페이지)에서 확인하십시오.

그림 11.6 붓스트랩 신뢰 한계 보고서


분포

Displacement

분위수

100.0%	최대값	0.6800107545
99.5%		0.6483289879
97.5%		0.6212501769
90.0%		0.5781520893
75.0%	사분위수	0.5423786474
50.0%	중앙값	0.5051381996
25.0%	사분위수	0.4692123772
10.0%		0.4374744231
2.5%		0.4035615438
0.5%		0.3686834278
0.0%	최소값	0.3494461047

요약 통계량

붓스트랩 신뢰 한계

포함 범위	BC 하한	BC 상한
0.95	0.40028	0.61892
0.90	0.41646	0.59849
0.80	0.43405	0.57489
0.50	0.46765	0.54007

원래 추정값 0.503787

"붓스트랩 신뢰 한계" 보고서의 맨 아래에 있는 **원래 추정값**은 원래 데이터를 사용한 통계량의 추정값입니다.

"붓스트랩 신뢰 한계" 보고서의 해석에 대한 자세한 내용은 "붓스트랩 개요"(343페이지)에서 확인하십시오. Efron 연구 자료(1981)에는 백분위수 구간과 편향 수정 백분위수 구간 모두의 방법이 설명되어 있습니다.

JMP PRO 붓스트랩의 추가 예

이 예에서는 작은 데이터 테이블에 부분 가중치(베이지안 붓스트랩) 옵션을 사용할 경우의 이점을 보여 줍니다. 데이터는 7가지 토양 유형을 각각 포함하는 세 개의 표본에서 측정된 반응(Y)으로 구성되어 있습니다. 한 과학자가 wabash 토양 유형의 평균 반응에 대한 신뢰 구간을 구하려고 합니다.

각 토양 유형마다 관측값이 세 개씩만 있으므로 단순 붓스트랩을 수행하면 wabash의 관측값 세 개가 모두 붓스트랩 표본에서 제외될 가능성이 있습니다. 부분 가중치 옵션을 사용하면 모든 붓스트랩 표본에서 각 토양 유형의 모든 관측값을 나타낼 수 있습니다.

이 과학자는 다음과 같은 두 가지 붓스트랩 방법을 모두 사용하여 wabash 표본 평균의 분포를 확인합니다.

- " 단순 붓스트랩 분석 "(352 페이지)
- " 베이지안 붓스트랩 분석 "(355 페이지)

단순 붓스트랩 분석

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Snapdragon.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Y 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Soil 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Y 대 Soil 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .
7. **일원 ANOVA 에 대한 평균** 보고서에서 **평균** 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다 .
8. **붓스트랩 표본 수**에 1000 을 입력합니다 .
9. (선택 사항) 그림 11.7 의 결과와 매칭되는 결과를 얻으려면 **난수 시드값**에 12345 를 입력합니다 .
10. **확인**을 클릭합니다 .

그림 11.7 단순 붓스트랩에 대한 붓스트랩 결과

	X	Y	BootID	clarion	clinton	compost	knox	o'neill	wabash	webster
1	Soil	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	Soil	Y	1	32.2333	30.9000	28.0000	35.7000	34.3500	31.9000	31.1000
3	Soil	Y	2	32.4333	30.3000	30.2000	33.9667	31.2000	38.0000	•
4	Soil	Y	3	32.5000	30.7500	29.2000	34.0333	32.8000	38.0000	31.1000
5	Soil	Y	4	31.5000	29.4000	29.9000	35.7500	35.0400	34.9500	•
6	Soil	Y	5	32.5000	30.6000	31.8000	35.7000	34.3000	35.0500	31.5667
7	Soil	Y	6	32.4000	30.1000	29.8500	34.4500	36.0000	38.2000	30.4000
8	Soil	Y	7	31.9000	29.3400	•	34.4000	33.6000	35.9667	31.4500
9	Soil	Y	8	32.1400	31.3500	29.6667	33.1000	35.1000	37.9333	30.6333
10	Soil	Y	9	32.7000	30.8000	30.2000	35.2600	31.2000	34.8500	32.5000
11	Soil	Y	10	32.1000	32.1000	28.0000	33.1000	34.1143	34.0000	31.8000
12	Soil	Y	11	31.5000	30.3000	•	33.1000	34.4000	37.2125	30.6333
13	Soil	Y	12	32.7000	30.4200	31.8000	34.0333	35.1000	35.0500	•
14	Soil	Y	13	32.1667	•	30.1000	34.9000	33.9000	38.2000	31.8000
15	Soil	Y	14	32.3000	•	29.2000	33.9667	34.4000	35.6000	31.9400
16	Soil	Y	15	32.2500	29.1000	29.6667	33.1000	35.1000	31.9000	•
17	Soil	Y	16	31.5000	30.4200	30.0000	•	35.2800	37.8000	31.1000
18	Soil	Y	17	32.4333	29.7000	29.2000	33.1000	34.4000	34.0000	31.1000
19	Soil	Y	18	32.3800	32.1000	29.9000	35.8000	34.6800	35.0500	31.8000
20	Soil	Y	19	32.0600	30.1000	30.2800	•	33.1500	31.9000	31.1000
21	Soil	Y	20	31.9000	29.1000	28.4000	35.9000	34.8000	34.8500	30.8200

그림 11.7 의 결측값은 해당 붓스트랩 표본에 주어진 토양 유형의 관측값이 모두 선택되지 않은 붓스트랩 반복을 나타냅니다 .

11. **분석 > 분포**를 선택합니다 .
12. wabash 를 선택하고 **Y, 열**을 클릭합니다 .
13. **확인**을 클릭합니다 .

그림 11.8 단순 붓스트랩으로 구한 wabash 평균의 분포

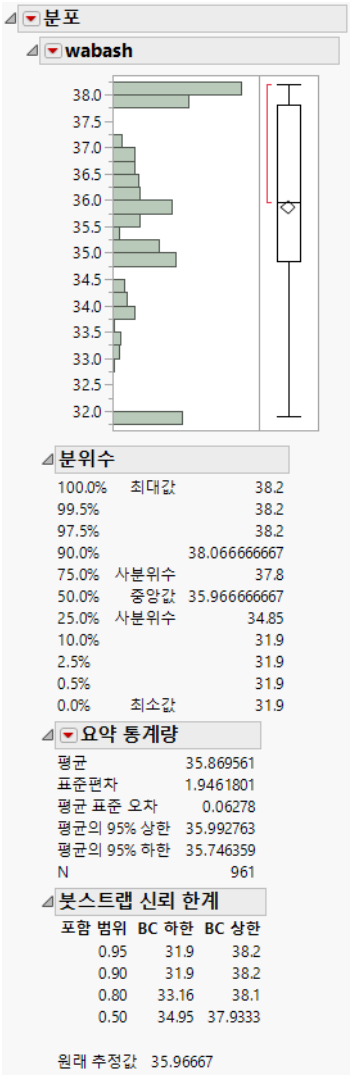


그림 11.8 에서는 단순 붓스트랩 분석으로 구한 wabash 평균의 분포를 보여 줍니다 . 다음을 확인할 수 있습니다 .

- " 요약 통계량 " 보고서는 wabash 에 대한 붓스트랩 평균을 포함하는 행의 개수가 N = 961 임을 나타냅니다 . 1,000 회 반복을 수행했지만 39 개의 붓스트랩 표본에는 wabash 에 대한 세 개의 관측값이 전혀 포함되지 않았습니다 .
- 표본 평균 히스토그램은 평활하지 않고 두 개의 극단값에서 정상점을 보여 줍니다 . wabash 에 대한 세 개의 값은 38.2, 37.8 및 31.9 입니다 . 분포의 낮은 쪽 정상점은 붓스트랩 표본에 값 31.9 만 포함되어서 발생한 경우이고 , 높은 쪽 정상점은 붓스트랩 표본에 값 38.2 와 37.8 이 하나 또는 모두 포함되어서 발생한 경우입니다 .

다음으로는 부분 가중치 (베이지안 붓스트랩) 옵션을 사용하여 붓스트랩 표본에서 결측값이 발생하지 않도록 하고 붓스트랩 평균의 분포를 평활하게 합니다 .

베이지안 붓스트랩 분석

1. " 일원 분석 " 보고서에서 **일원 ANOVA 에 대한 평균** 보고서의 **평균** 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다 .
2. **붓스트랩 표본 수**에 1000 을 입력합니다 .
3. (선택 사항) 그림 11.9 의 결과와 매칭되는 결과를 얻으려면 **난수 시드값**에 12345 를 입력합니다 .
4. **부분 가중치** 옵션을 선택합니다 .
5. **확인**을 클릭합니다 .

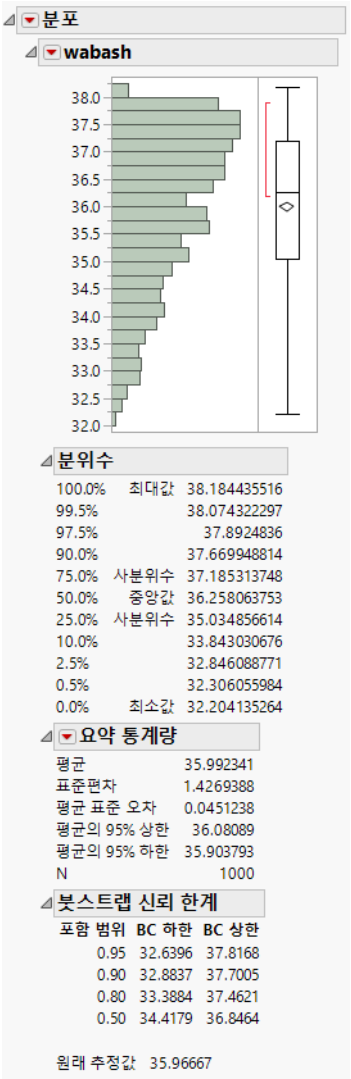
그림 11.9 베이지안 붓스트랩에 대한 붓스트랩 결과

	X	Y	BootID•	clarion	clinton	compost	knox	o'neill	wabash	webster
1	Soil	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	Soil	Y	1	31.9365	31.3493	29.3234	35.4497	34.7105	33.7270	31.4071
3	Soil	Y	2	32.4189	30.3474	29.5143	35.6281	32.7006	34.1027	32.1212
4	Soil	Y	3	32.2339	30.2001	31.4102	34.8758	33.6674	38.0389	31.5428
5	Soil	Y	4	32.4054	30.3242	30.6227	33.8495	32.9495	36.8344	31.9607
6	Soil	Y	5	32.2262	30.8672	29.7999	33.6759	32.2022	35.5792	31.9058
7	Soil	Y	6	32.3222	31.9732	28.8823	35.6307	34.8863	35.3014	30.1325
8	Soil	Y	7	31.9948	30.8828	29.2516	35.3424	33.2094	36.3367	30.8386
9	Soil	Y	8	31.6254	29.7677	28.6390	34.4697	33.0662	36.6183	31.4667
10	Soil	Y	9	32.3499	29.9416	29.5732	35.2564	31.9583	35.8246	29.9302
11	Soil	Y	10	32.5228	30.4506	28.4859	34.9088	35.8126	34.6317	31.6100
12	Soil	Y	11	31.9057	30.0711	29.0693	35.4018	34.4654	33.2086	31.0309
13	Soil	Y	12	31.7275	29.6189	29.1609	34.3984	33.6840	35.0815	31.5446
14	Soil	Y	13	32.5893	30.5210	28.6054	33.4594	34.0958	33.7692	31.7129
15	Soil	Y	14	32.2473	30.5199	31.6080	35.5617	34.1706	35.7146	31.8023
16	Soil	Y	15	32.2329	29.7275	30.1770	35.3286	32.7820	37.4866	30.9706
17	Soil	Y	16	31.5831	29.7508	28.7122	33.5304	34.5348	37.1092	31.2752
18	Soil	Y	17	32.3545	31.3237	29.1542	35.5890	32.2606	37.2005	30.8430
19	Soil	Y	18	32.3811	29.6241	30.6138	35.4308	33.2024	33.0787	31.2926
20	Soil	Y	19	31.7488	29.7763	28.7327	34.7007	33.8910	34.0573	29.9064

베이지안 붓스트랩 결과 테이블에는 결측값이 없습니다 . Snapdragon.jmp 데이터 테이블의 21 개 행이 모두 포함되었으며 각 붓스트랩 표본에서 붓스트랩 가중치만 달라졌습니다 .

6. **분석 > 분포**를 선택합니다 .
7. wabash 를 선택하고 **Y, 열**을 클릭합니다 .
8. **확인**을 클릭합니다 .

그림 11.10 베이지안 붓스트랩으로 구한 wabash 평균의 분포



베이지안 붓스트랩은 wabash 표본 평균에 대해 훨씬 더 평활한 분포를 생성합니다 . 1,000 개의 붓스트랩 표본 모두에 wabash 에 대한 세 개의 관측값이 포함되어 있습니다 . 각 반복 시마다 서로 다른 부분 가중치를 사용하여 wabash 표본 평균이 계산됩니다 .

" 붓스트랩 신뢰 한계 " 보고서에는 평균에 대한 95% 신뢰 구간이 32.6396 ~ 37.8168 로 나타납니다 .

JMP PRO 부스트랩에 대한 통계 상세 정보

JMP PRO 부분 가중치 계산

부분 가중치 옵션은 베이지안 붓스트랩(Rubin 1981)을 기반으로 합니다. 하나의 관측값이 주어진 붓스트랩 표본에 나타나는 횟수를 **붓스트랩 가중치**라고 합니다. 단순 붓스트랩에서는 각 붓스트랩 표본에 대한 붓스트랩 가중치가 단순 랜덤 복원 표집 방식으로 결정됩니다.

베이지안 방법에서는 표집 확률이 알 수 없는 모수로 처리되며 무정보 사전 분포를 사용하여 확률의 사후 분포를 구합니다. 확률 추정값은 이 사후 분포에서 표집을 통해 구합니다. 이러한 추정값은 붓스트랩 가중치를 생성하는 데 사용됩니다.

- 형상 모수가 $(n - 1)/n$ 이고 척도 모수가 1 인 감마 분포의 n 개 값을 포함하는 벡터를 무작위로 생성합니다.

  참고 : Rubin 연구 자료 (1981) 에서는 감마 형상 모수로 1 을 사용합니다. JMP Pro 에서 사용되는 형상 모수는 부분 가중치의 평균 및 분산이 단순 붓스트랩 가중치의 평균 및 분산과 같도록 합니다.

- n 개 값의 합인 S 를 계산합니다.
- n 개 값의 벡터에 N / S 를 곱해서 부분 가중치를 계산합니다. 여기서 N 은 행 수 또는 빈도 합 (빈도 변수가 지정된 경우) 과 같습니다.

참고: 분석에 빈도 변수가 지정된 경우에는 행별로 감마 분포의 형상 모수에 빈도 값을 곱하십시오. 빈도 변수 값의 합은 1 보다 커야 합니다. 그러면 형상 모수가 $f_i(N - 1)/N$ 과 같아집니다. 여기서 f_i 는 i 번째 행의 빈도 값이며 N 은 빈도 값의 합과 같습니다.

이 절차에서는 붓스트랩 표집의 평균 및 분산이 단순 붓스트랩 가중치의 평균 및 분산과 같도록 각 행의 부분 가중치를 척도화합니다. 각 붓스트랩 표본의 부분 붓스트랩 가중치는 양수이고, 합은 n이 되며, 평균은 1이 됩니다.

JMP PRO 편향 수정 백분위수 구간

이 섹션에서는 붓스트랩 결과 테이블에서 "분포" 스크립트를 실행하면 "붓스트랩 신뢰 한계" 보고서에 표시되는 BC(편향 수정) 신뢰 구간의 계산 방법을 설명합니다. 편향 수정 백분위수 구간은 붓스트랩 분포에서 백분위수 구간의 비대칭성 설명 능력을 향상시킵니다. 자세한 내용은 Efron 연구 자료(1981)에서 확인하십시오.

표기

- p^* - 관심 있는 통계량의 추정값이 원래 추정값보다 작거나 같은 붓스트랩 표본의 비율입니다.

- z_0 - 표준 정규 분포의 p^* 분위수입니다 .
- z_α - 표준 정규 분포의 α 분위수입니다 .

편향 수정 신뢰 구간의 끝점

(1 - α) 편향 수정 신뢰 구간의 끝점은 붓스트랩 분포의 분위수로 주어집니다 .

- 하한점은 다음 분위수입니다 .

$$\Phi\left(2z_0 + z_{\frac{\alpha}{2}}\right)$$

- 상한점은 다음 분위수입니다 .

$$\Phi\left(2z_0 + z_{1-\frac{\alpha}{2}}\right)$$

12 장

텍스트 탐색기
데이터의 비정형 텍스트 탐구

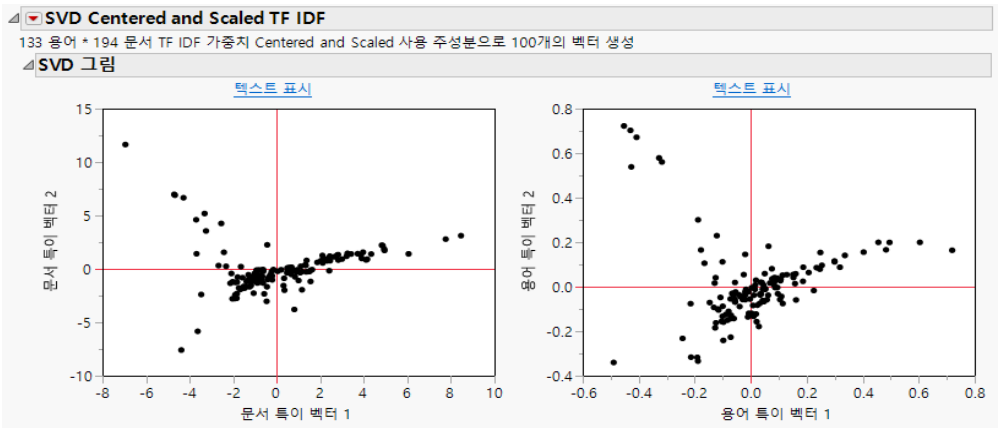
JMP PRO 이 플랫폼의 많은 기능은 JMPPro 에서만 사용할 수 있으며 그러한 경우 이 아이콘이 표시됩니다 .

텍스트 탐색기 플랫폼을 사용하면 설문 조사 또는 사고 보고서의 주석 필드와 같은 비정형 텍스트를 분석할 수 있습니다. 도구를 사용하여 유사한 용어를 결합하고 잘못 지정된 용어를 재코딩하고 텍스트 데이터의 기본 패턴을 이해함으로써 텍스트 데이터와 상호 작용할 수 있습니다.

JMP PRO JMP Pro 버전의 플랫폼에는 SVD(특이값 분해)를 사용하여 유사한 문서를 주제별로 그룹화하는 분석 도구가 포함되어 있습니다. 텍스트 문서를 군집화하거나 문서 모음에 있는 용어를 군집화할 수 있습니다. 잠재 계층 분석을 사용하여 문서를 군집화할 수도 있습니다.

JMP PRO JMP Pro 버전의 플랫폼에는 문서의 중요 용어와 감정을 식별하기 위한 도구도 포함되어 있습니다. 용어 선택을 사용하면 서로 다른 응답을 가장 잘 설명하는 용어를 식별할 수 있습니다. 감정 분석을 사용하면 어휘 분석을 사용하여 문서에서 감정 용어를 식별하고 긍정, 부정 및 전반적 감정에 대해 문서의 점수를 매길 수 있습니다. 감정 분석 기능에는 기본적인 NLP(자연어 처리) 지원이 포함됩니다.

그림 12.1 텍스트 탐색기의 SVD 그림



목차

텍스트 탐색기 플랫폼 개요 361
텍스트 처리 단계 362
텍스트 탐색기 플랫폼의 예 363
텍스트 탐색기 플랫폼 시작 366
정규 표현식 편집기에서 정규 표현식 사용자 정의 368
텍스트 탐색기 보고서 374
요약 개수 보고서 374
용어 및 구 목록 375
텍스트 탐색기 플랫폼 옵션 378
텍스트 준비 옵션 378
텍스트 분석 옵션 384
저장 옵션 386
보고서 옵션 387
잠재 계층 분석 387
잠재 의미 분석 (SVD) 389
SVD 보고서 390
SVD 보고서 옵션 391
주제 분석 392
주제 분석 보고서 393
주제 분석 보고서 옵션 393
판별 분석 394
판별 분석 보고서 394
판별 분석 보고서 옵션 395
용어 선택 396
용어 선택 설정 396
용어 선택 보고서 397
용어 선택 보고서 옵션 399
감정 분석 399
감정 분석 보고서 401
감정 분석 보고서 옵션 403
텍스트 탐색기 플랫폼의 추가 예 404

텍스트 탐색기 플랫폼 개요

텍스트 탐색기 플랫폼을 사용하면 비정형 텍스트를 탐색하여 텍스트의 의미를 더 잘 이해할 수 있습니다. 비정형 텍스트 데이터는 일반적으로 볼 수 있습니다. 예를 들어 설문 조사, 제품 평가 의견 또는 사고 보고서의 자유 응답 필드에서 비정형 텍스트 데이터가 발생할 수 있습니다.

텍스트 분석은 대개 반복적인 프로세스이므로 용어 목록의 큐레이팅과 분석을 번갈아가며 수행할 수 있습니다.

용어 목록 큐레이팅

텍스트 분석에는 몇 가지 고유한 용어가 사용됩니다. **용어** 또는 **토큰**은 가장 작은 텍스트 조각으로, 문장 내의 단어와 유사합니다. 하지만 용어는 정규 표현식을 사용하는 등의 여러 가지 방법으로 정의할 수 있습니다. 텍스트를 용어로 분해하는 프로세스를 **토큰화**라고 합니다.

- **구**는 짧은 용어 모음입니다. 텍스트 탐색기 플랫폼에는 자체적으로 용어로 지정된 구를 관리하는 옵션이 있습니다.
- **문서**는 단어 모음을 나타냅니다. JMP 데이터 테이블에서는 텍스트 열의 각 행에 있는 비정형 텍스트가 문서에 해당합니다.
- **말뭉치**는 문서 모음을 나타냅니다.

때로는 분석에서 몇 가지 일반적인 단어를 제외하는 것이 좋습니다. 이 제외된 단어를 **중지 단어**라고 합니다. 플랫폼에 기본 중지 단어 목록이 있지만 사용자가 특정 단어를 중지 단어로 추가할 수도 있습니다. 중지 단어는 용어가 될 수는 없지만 구에 사용될 수는 있습니다.

용어를 재코딩할 수도 있습니다. 그러면 동의어를 하나의 공통 용어로 결합하는 데 유용합니다.

어간 추출(Stemming)은 서로 다른 어미를 제거함으로써 시작 부분이 동일한 단어(**어간**)를 결합하는 프로세스입니다. 즉, "jump", "jumped" 및 "jumping"이 모두 "jump"라는 용어로 취급됩니다. 어간 추출 절차는 Snowball 문자열 처리 언어에서 사용되는 절차와 유사합니다. 구를 어간 추출할 때는 구의 각 단어가 독립된 용어일 때와 마찬가지 방식으로 어간 추출됩니다.

용어 목록 분석

텍스트 탐색기 플랫폼의 텍스트 분석에는 BoW(bag of words) 방법이 사용됩니다. 따라서 구를 형성할 때 외에는 용어의 순서가 무시됩니다. 이 분석은 용어의 개수를 기준으로 합니다.

정규 표현식, 중지 단어, 재코딩 및 어간 추출을 사용하여 용어 목록을 큐레이팅한 후에는 큐레이팅된 용어 목록에 대한 분석을 수행할 수 있습니다. 플랫폼의 분석 옵션은 DTM(**문서 용어 행렬**)을 기반으로 합니다. DTM의 각 행은 하나의 문서(JMP 데이터 테이블의 텍스트 열에 있는 셀 하나)에 해당합니다. DTM의 각 열은 큐레이팅된 용어 목록의 용어 하나에 해당합니다. 이 방법은 단어 순서를 무시하기 때문에 BoW 방법을 구현합니다. 가장 단순한 형태일 때 DTM의 각 셀에는 행의 문서에 있는 열의 용어 빈도(발생 횟수)가 포함됩니다. DTM을 위한 다른 가중치 체계가 다양하게 존재하며, 이는 "저장 옵션"(386페이지)에 설명되어 있습니다.

JMP PRO 플랫폼에서 사용할 수 있는 분석 옵션은 먼저 문서 용어 행렬에 대해 SVD(특이값 분해)를 수행합니다. 이를 통해 데이터의 용어 정보를 나타내는 데 필요한 열 수가 크게 줄어들 수 있습니다. 특이값 분해에 대한 자세한 내용은 Multivariate Methods의 에서 확인하십시오. 용어 군집화 및 문서 군집화에는 "계층적 군집화" 옵션을 사용할 수 있습니다. 이 옵션을 사용하면 유사한 용어 또는 문서끼리 그룹화할 수 있습니다.

텍스트 탐색기 플랫폼 워크플로우

텍스트 탐색기 플랫폼을 사용하기 위해 필요한 단계가 있습니다.

1. 토큰화 방법 (기본 제공 또는 사용자 정의 정규 표현식) 을 지정합니다.
2. 보고서를 사용하여 추가 중지 단어를 지정하고, 용어 목록에 구를 추가하고, 용어의 재코딩을 수행하고, 어간 추출 규칙에 대한 예외를 지정합니다.
3. 어간 추출에 대한 환경 설정을 지정합니다.
4. 단어 및 문 개수, SVD 및 군집화 방법을 사용하여 중요한 용어 및 구를 식별합니다.

   참고 : JMP PRO SVD 및 군집화 옵션은 JMP Pro 에서만 사용할 수 있습니다.

5. 결과 (용어 테이블, DTM, 특이 벡터 또는 기타 결과) 를 추가 분석에 사용할 수 있도록 저장합니다.

   참고 : JMP PRO 특이 벡터를 저장하는 옵션은 JMP Pro 에서만 사용할 수 있습니다.

6. 구, 재코딩 및 중지 단어 특성을 유사한 텍스트 데이터의 향후 분석에 사용할 수 있도록 저장합니다.

텍스트 처리 단계

JMP 텍스트 탐색기플랫폼에서 텍스트는 토큰화, 구 추출 및 용어 추출의 세 단계로 처리됩니다.

토큰화 단계

토큰화 단계에서는 다음 작업을 수행합니다.

1. 텍스트를 소문자로 변환합니다.
2. 토큰화 방법 (기본 단어 또는 정규 표현식) 을 적용하여 문자를 토큰으로 그룹화합니다.
3. 지정된 재코딩 정의에 따라 토큰을 재코딩합니다. 재코딩은 어간 추출 (Stemming) 전에 수행됩니다.

   참고 : 재코딩 작업은 보고서 창에서 지정된 순서와 관계없이 단일 패스로 내부적으로 처리됩니다.

구 추출 단계

구 추출 단계에서는 말뭉치(문서 모음)에서 나타나는 구를 수집하며, 이를 통해 개별 구를 용어로 처리하도록 지정할 수 있습니다. 구는 중지 단어로 시작하거나 끝날 수 없지만 중지 단어를 포함할 수는 있습니다.

용어 추출 단계

용어 추출 단계에서는 이전 단계에서 추출된 토큰 및 구로부터 용어 목록을 생성합니다.

용어 추출 단계에서는 각 토큰에 대해 다음 작업을 수행합니다.

1. 시작 창에 지정된 최소 및 최대 길이 요구 사항이 충족되는지 확인합니다. 숫자만 포함된 토큰은 이 작업에서 제외됩니다.
2. 토큰이 용어가 될 수 있는 자격이 있는지 확인합니다. 기본 단어 토큰화 방법으로 파싱된 토큰에는 알파벳 문자 또는 유니코드 문자가 하나 이상 포함되어야 합니다. 숫자만 포함된 토큰은 이 작업에서 제외됩니다. 정규 표현식 토큰화 방법에서는 정규 표현식을 사용하여 토큰의 일부인 문자를 확인합니다.
3. 토큰이 중지 단어가 아닌지 확인합니다.
4. 어간 추출 및 어간 예외를 적용합니다.

용어 추출 단계에서는 사용자가 추가하는 각 구에 대해 다음 작업을 수행합니다.

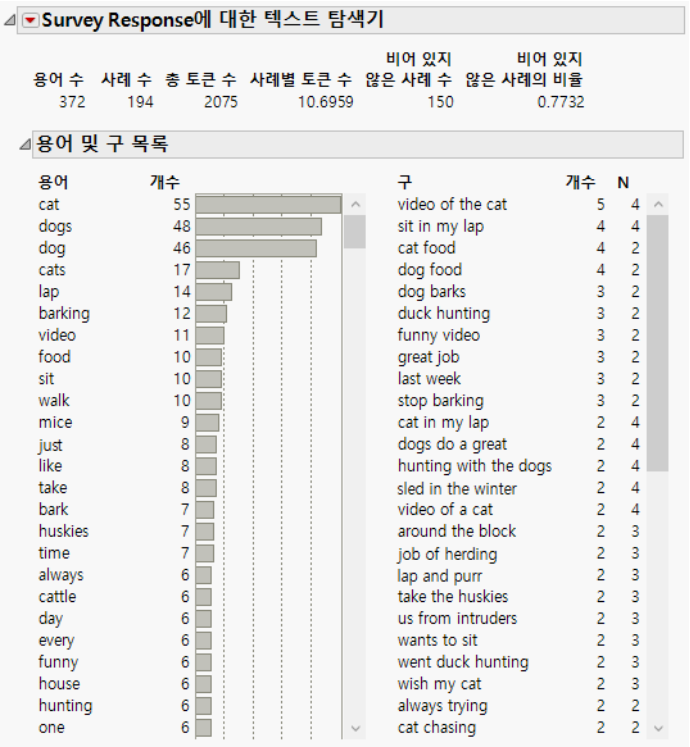
1. 용어 목록에 구를 추가합니다. 구는 용어 목록에 어간 추출된 구의 각 단어에 어간 추출을 적용해야 합니다. 원시 토큰은 다르지만 어간이 동일한 구는 용어 목록에서 결합됩니다.
2. 구에 나타나는 토큰 용어 항목을 제거합니다.

텍스트 탐색기 플랫폼의 예

JMP 에서 텍스트 응답을 탐색하는 방법을 알아봅니다. 이 예에서는 반려 동물에 대한 설문 조사에서 얻은 텍스트 응답을 살펴보려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Pet Survey.jmp 를 엽니다.
2. **분석 > 텍스트 탐색기**를 선택합니다.
3. Survey Response 를 선택하고 **텍스트 열**을 클릭합니다.
4. " 언어 " 목록에서 **영어**를 선택합니다.
5. **확인**을 클릭합니다.

그림 12.2 초기 텍스트 탐색기 보고서의 예



194 개 문서에 372 개의 고유 용어가 있음을 한눈에 알 수 있습니다 . 총 1921 개의 토큰화된 용어가 있습니다 . 가장 자주 나타나는 용어는 55 회 나타나는 "cat" 입니다 .

6. "Survey Response 에 대한 텍스트 탐색기 " 옆의 빨간색 삼각형을 클릭하고 **용어 옵션 > 어간 추출 (Stemming) > 모든 용어의 어간 추출**을 선택합니다 .
7. 구 목록 테이블에서 **cat food** 및 **dog food** 를 선택하고 선택 항목을 마우스 오른쪽 버튼으로 클릭한 다음 **구 추가**를 선택합니다 .

   "cat food" 와 "dog food" 라는 용어가 용어 목록에 포함됩니다 .
8. 용어 목록을 아래로 스크롤하여 "cat food" 및 "dog food" 항목을 찾습니다 .

   각 구가 네 번씩 나타남을 알 수 있습니다 .

그림 12.3 수정 및 스크롤 후 용어 목록

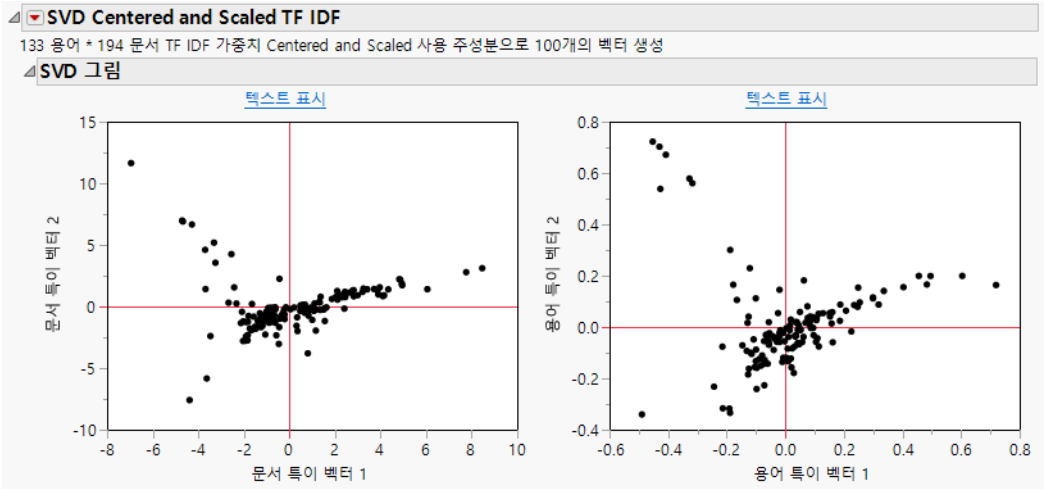
"cat food" 와 "dog food" 는 이제 이 텍스트 탐색기 보고서에 한해 용어로 처리되므로 구 목록에서는 이 두 항목이 회색으로 표시됩니다 .

JMP PRO 이 예의 나머지 단계는 JMP Pro 에서만 수행할 수 있습니다 .

9. JMP PRO "Survey Response 에 대한 텍스트 탐색기 " 옆의 빨간색 삼각형을 클릭하고 **잠재 의미 분석 , SVD** 를 선택합니다 .
10. JMP PRO **확인**을 클릭하여 기본값을 적용합니다 .

    보고서에 두 개의 SVD 그림이 나타납니다 . 왼쪽 그림에서는 문서 공간에 있는 처음 두 개의 특이 벡터를 보여 줍니다 . 오른쪽 그림에서는 용어 공간에 있는 처음 두 개의 특이 벡터를 보여 줍니다 .

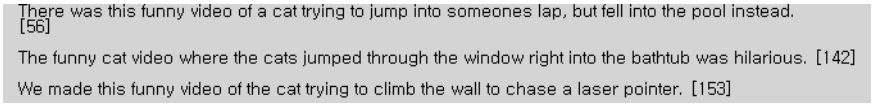
그림 12.4 SVD 그림



11. JMP PRO 왼쪽 SVD 그림에서 맨 오른쪽에 있는 점 세 개를 선택합니다 .

    이 세 개의 점은 나머지 점과는 동떨어져서 군집화된 설문 조사 응답을 나타냅니다 . 이 군집을 보다 자세히 조사하려면 해당 응답의 텍스트를 읽어 보십시오 .
12. JMP PRO 왼쪽 SVD 그림 위에 있는 **텍스트 표시** 버튼을 클릭합니다 .

그림 12.5 선택한 문서의 텍스트



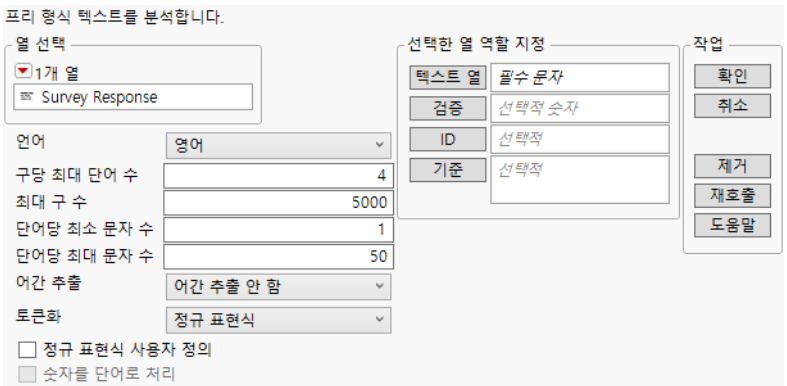
선택한 점이 나타내는 세 문서의 텍스트가 포함된 창이 나타납니다. 이러한 설문 조사 응답은 모두 "funny", "cat" 및 "video" 의 몇 가지 조합을 언급한다는 점에서 유사합니다. 이러한 문서는 첫 번째 특이 벡터에 대해 나머지 문서보다 큰 양의 값을 가집니다. 이렇게 큰 값은 이러한 문서가 해당 차원의 나머지 문서와는 다름을 나타냅니다.

특이 벡터의 차원을 더 조사해 보면 차원이 나타내는 것을 해석할 수 있습니다. 예를 들어 그림의 맨 오른쪽에 있는 많은 문서들은 고양이에 관한 응답입니다. 맨 왼쪽의 많은 응답은 개에 관한 것입니다. 따라서 첫번째 특이 벡터에서는 응답이 고양이에 관한 것인지 개에 관한 것인지에 따라 차이를 찾아냅니다.

텍스트 탐색기 플랫폼 시작

분석 > 텍스트 탐색기를 선택하여 텍스트 탐색기 플랫폼을 시작할 수 있습니다.

그림 12.6 텍스트 탐색기 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 텍스트 탐색기 시작 창에는 다음과 같은 옵션이 포함되어 있습니다.

텍스트 열 텍스트 데이터가 포함된 열을 할당합니다. 여러 열을 지정하면 각 열에 대해 별도의 분석이 생성됩니다.

JMP PRO **검증** JMP Pro 에서는 검증 열을 입력할 수 있습니다. " 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다. " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오.

검증 열을 지정해도 문서 용어 행렬의 계산에는 영향이 없습니다. 하지만 검증 열이 지정된 경우에는 " 잠재 계층 분석 ", " 잠재 의미 분석 ", " 주제 분석 " 및 " 판별 분석 " 옵션에 훈련 데

이터 집합만 사용됩니다 . 검증 열은 " 용어 선택 " 옵션에 대한 일반화 회귀 검증 방법으로 사용됩니다 .

ID " 연관성을 위해 쌓인 형식의 DTM 저장 " 출력 데이터 테이블에서 개별 응답자를 식별하는 데 사용되는 열을 할당합니다 . 이 출력 데이터 테이블은 연관성 분석에 적합합니다 . 이 열은 " 잠재 계층 분석 " 보고서에서 개별 응답자를 식별하는 데도 사용됩니다 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

참고 : 기준 변수를 지정하는 경우 기준 변수의 모든 수준에 " 정규 표현식 사용자 정의 " 옵션 및 설정이 적용됩니다 .

언어 텍스트 처리에 사용되는 언어를 지정합니다 . 이는 어간 추출 (Stemming) 과 중지 단어 , 재코딩 및 구의 기본 제공 목록에 영향을 줍니다 . 이 옵션은 JMP 실행 언어와는 별개입니다 . " 언어 " 플랫폼 환경 설정이 지정되지 않은 경우 이 " 언어 " 옵션은 JMP " 표시 언어 " 환경 설정에 따라 설정됩니다 . 하지만 텍스트 탐색기의 " 언어 " 옵션은 한국어를 지원하지 않습니다 . JMP 표시 언어가 한국어인 경우 이 옵션은 기본적으로 영어로 설정됩니다 .

구당 최대 단어 수 구가 분석에 구로 포함되기 위해 최대로 포함할 수 있는 단어 수를 지정합니다 .

최대 구 수 구 목록에 나타나는 구의 최대 개수를 지정합니다 .

단어당 최소 문자 수 단어가 분석에 용어로 포함되기 위해 반드시 포함해야 하는 문자 수를 지정합니다 .

단어당 최대 문자 수 단어가 분석에 용어로 포함되기 위해 최대로 포함할 수 있는 문자 수 (최대 2000) 를 지정합니다 .

어간 추출 (Stemming) (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 시작 문자는 유사하지만 종료 문자는 다른 용어를 결합하는 방법을 지정합니다 . 다음 옵션을 사용할 수 있습니다 .

어간 추출 (Stemming) 안 함 용어를 결합하지 않습니다 .

결합 가능한 경우 어간 추출 둘 이상의 용어에서 추출된 어간이 동일한 용어일 경우에만 용어의 어간을 추출합니다 .

모든 용어의 어간 추출 모든 용어의 어간을 추출합니다 .

참고 : " 어간 추출 (Stemming)" 옵션을 사용할 경우 용어 목록에 추가된 구에도 영향을 주게 됩니다 . 구 내의 용어가 어간 추출된 후에는 구 식별이 발생합니다 . 예를 들어 "dogs bark" 와 "dog barks" 는 모두 "dog· bark " 라는 특정 구와 매칭됩니다 .

토큰화 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 텍스트를 용어 또는 토큰으로 파싱하기 위한 방법을 지정합니다 . 사용 가능한 토큰화 옵션은 다음과 같습니다 .

정규 표현식 기본 제공 정규 표현식을 사용하여 텍스트를 파싱합니다 . 텍스트를 파싱하는 데 사용되는 정규 표현식을 추가 , 제거 또는 편집하려면 **정규 표현식 사용자 정의** 옵션을 선택합니다 . 자세한 내용은 " 정규 표현식 편집기에서 정규 표현식 사용자 정의 "(368 페이지) 에서 확인하십시오 .

기본 단어 텍스트는 일반적으로 단어를 구분하는 문자를 기준으로 파싱됩니다 . 이러한 문자로는 공백 , 탭 , 줄바꿈 및 대부분의 구두점 표시가 포함됩니다 . 분석을 위해 숫자를 용어로 파싱하려면 **숫자를 단어로 처리** 옵션을 선택합니다 . 이 옵션을 선택하지 않을 경우 구분자 사이에서 숫자만 포함하는 텍스트는 토큰화 단계에서 무시됩니다 .

팁 : 기본 단어 토큰화 방법을 사용하는 텍스트 탐색기 보고서에서 **표시 옵션 > 구분자 표시** 옵션을 사용하여 기본 구분자 집합을 볼 수 있습니다 .

정규 표현식 사용자 정의 (정규 표현식 토큰화 방법을 사용하는 경우에만 사용 가능) 텍스트 탐색기 정규 표현식 편집기 창을 사용하여 정규 표현식 설정을 수정할 수 있습니다 . 일반적이지 않은 단어를 수용하려면 이 옵션을 사용합니다 . 예를 들어 전화 번호나 문자와 숫자의 조합으로 구성된 단어를 들 수 있습니다 . " 정규 표현식 사용자 정의 " 옵션은 기본 정규 표현식 방법으로 필요한 결과를 얻을 수 없는 경우에만 사용하는 것이 좋습니다 . 이러한 경우는 텍스트에 기본 정규 표현식 방법으로 인식되지 않는 구조가 포함된 경우에 발생할 수 있습니다 . 자세한 내용은 " 정규 표현식 편집기에서 정규 표현식 사용자 정의 "(368 페이지) 에서 확인하십시오 .

숫자를 단어로 처리 (기본 단어 토큰화 방법과 함께만 사용 가능) 분석에서 숫자를 용어로 토큰화할 수 있도록 합니다 . 이 옵션이 선택되어 있으면 숫자가 포함된 용어에 대해 " 단어당 최소 문자 수 " 설정이 무시됩니다 .

시작 창에서 **정규 표현식 사용자 정의**를 선택한 경우 시작 창에서 **확인**을 클릭하면 텍스트 탐색기 정규 표현식 편집기 창이 나타납니다. 그렇지 않은 경우에는 텍스트 탐색기 보고서가 나타납니다.

참고 : 텍스트 입력 처리는 대 / 소문자를 구분하지 않습니다 . 모든 텍스트는 토큰화 단계와 모든 분석 단계 이전에 내부적으로 소문자로 변환됩니다 . 이 변환은 텍스트 탐색기 출력에서 정규 표현식 처리와 용어 집계에 영향을 미칩니다 .

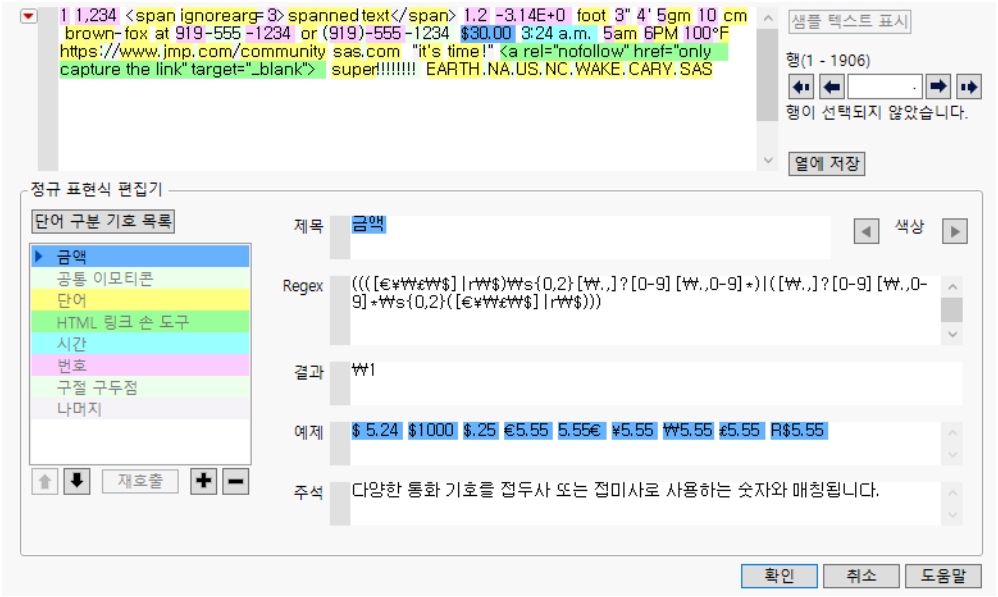
정규 표현식 편집기에서 정규 표현식 사용자 정의

정규 표현식 사용자 정의 옵션을 선택하면 텍스트 탐색기 정규 표현식 편집기가 나타납니다. 이 창에서 전화 번호, 시간 또는 통화 값과 같은 다양한 기본 제공 정규 표현식을 사용하여 텍스트 문서를 파싱할 수 있습니다. 고유한 정규 표현식 정의를 생성할 수도 있습니다.

참고 : " 정규 표현식 사용자 정의 " 옵션은 기본 정규 표현식 방법으로 원하는 결과를 얻을 수 없는 경우에만 사용하는 것이 좋습니다 . 이러한 경우는 텍스트에 기본 정규 표현식 방법으로 인식되지 않는 구조가 포함된 경우에 발생할 수 있습니다 .

팁 : 시작 창에서 " 언어 " 옵션으로 일본어 , 중국어 (간체) 또는 중국어 (번체) 가 지정된 경우에는 정규 표현식 패턴 목록에 지정된 언어의 단일 정규 표현식이 포함됩니다 . 다른 정규 표현식 패턴을 추가하려면 단일 정규 표현식 패턴 뒤에 추가하는 것이 좋습니다 . 단어 패턴은 긴 일련의 아시아 언어 문자를 하나의 단어로 취합할 수 있기 때문에 언어별 정규 표현식 패턴 앞에 단어 패턴을 사용하지 않는 것이 좋습니다 .

그림 12.7 텍스트 탐색기 정규 표현식 편집기



스크립트 편집기 상자를 사용한 파싱

창 맨 위의 스크립트 편집기 상자에서는 샘플 텍스트에 대한 파싱 진행 방식을 보여 줍니다 . 정규 표현식 편집기 목록의 정규 표현식을 파싱한 결과는 정규 표현식 편집기 목록의 색상에 해당하는 색상으로 강조 표시됩니다 .

- 스크립트 편집기 상자에 사용자 데이터의 텍스트를 채우려면 **첫 번째** , **이전** , **다음** 및 **마지막** 행 버튼을 클릭합니다 . 이렇게 하면 지정된 텍스트 데이터 행이 파싱되는 방식을 볼 수 있습니다 . 편집 상자에 행 번호를 입력하여 스크립트 편집기 상자를 데이터 테이블의 특정 행에 있는 텍스트로 채울 수도 있습니다 .
- 정규 표현식 토큰화 결과가 포함된 새 열을 데이터 테이블에 저장하려면 **열에 저장** 버튼을 클릭합니다 . 정규 표현식의 결과를 지정하는 방법에 대한 자세한 내용은 " 정규 표현식 편집 "(370 페이지) 에서 확인하십시오 . **열 > 유틸리티 > 텍스트 매칭으로 새 열 생성에서** 정규 표현식 편집기에 액세스하면 **열에 저장** 버튼이 나타나지 않습니다 .

참고 : 열에 저장 버튼을 클릭할 경우 텍스트 매칭에 정규 표현식만 사용됩니다 . 정규 표현식의 출력을 수정하는 데 중지 단어 , 재코딩 , 어간 추출 (Stemming), 구 , 또는 단어당 최소 / 최대 문자 수 등의 설정은 사용되지 않습니다 .

정규 표현식 추가

토큰화에 사용할 정규 표현식을 추가하려면 목록 아래의 더하기 기호를 클릭합니다. 그러면 "정규 표현식 라이브러리 선택" 창이 나타납니다. 이 창에는 모든 기본 제공 정규 표현식뿐만 아니라 이전에 정규 표현식 편집기에서 생성한 후 최근에 수정한 정규 표현식도 포함되어 있습니다. 기본 제공 정규 표현식에는 라벨이 지정되어 있습니다. 라이브러리에 저장된 사용자 정규 표현식에는 사용자가 지정한 이름으로 라벨이 지정됩니다. 지정된 이름의 가장 최근 표현식만 정규 표현식 라이브러리에 저장됩니다.

재호출 버튼을 클릭하여 정규 표현식 목록을 정규 표현식 편집기의 가장 최근 인스턴스에서 사용한 정규 표현식으로 채웁니다. 재호출된 정규 표현식은 편집기의 이전 인스턴스에서 **열에 저장** 버튼이나 **확인** 버튼을 클릭할 때 존재한 정규 표현식입니다.

선택한 정규 표현식을 토큰화에 사용할 정규 표현식으로 추가하려면 목록에서 정규 표현식을 하나 이상 선택하고 **확인**을 클릭합니다 . 정규 표현식 라이브러리에서 하나 이상의 사용자 정규 표현식을 제거하려면 **선택 항목 삭제** 버튼을 사용합니다 . 각 사용자의 정규 표현식 라이브러리는 TextExplorer 라는 디렉터리에 JSL 파일로 저장됩니다 . 이 디렉터리의 위치는 다음과 같이 컴퓨터 운영 체제에 따라 다릅니다 .

- Windows: "C:/Users/< 사용자 이름 >/AppData/Roaming/SAS/JMP/TextExplorer/"
- macOS: "/Users/< 사용자 이름 >/Library/Application Support/JMP/TextExplorer/"

이러한 파일을 다른 사용자와 공유할 수는 있지만 파일을 직접 편집해서는 안 됩니다. 대신 정규 표현식 편집기를 사용하십시오.

정규 표현식 편집

정규 표현식 편집기 패널에 지정된 순서대로 정규 표현식을 처리하여 용어를 토큰화할 수 있습니다. 정규 표현식의 순서를 변경하려면 목록에서 정규 표현식을 선택하고 목록 아래의 위쪽 또는 아래쪽 화살표 버튼을 클릭합니다. 정규 표현식 목록의 항목을 드래그하여 놓는 방법으로 실행 순서를 변경할 수도 있습니다. 파란색 삼각형은 현재 선택된 정규 표현식을 나타냅니다. 정규 표현식을 제거하고 토큰화에서 제외하려면 목록에서 정규 표현식을 선택하고 목록 아래의 빼기 기호를 클릭합니다. "나머지" 정규 표현식은 제거할 수 없으며 정규 표현식 시퀀스에서 마지막으로 나타나야 합니다.

목록에서 정규 표현식을 선택하면 정규 표현식 편집기 패널의 편집 가능한 필드는 선택한 정규 표현식을 참조합니다. 필드를 편집하려면 해당 필드를 클릭하고 내용을 입력합니다.

각 정규 표현식에는 다음과 같은 속성이 있습니다 .

제목 현재 창과 이후 정규 표현식 라이브러리에서 정규 표현식을 식별하는 데 사용할 이름을 지정합니다 .

정규 표현식 정규 표현식 정의를 지정합니다 . 정규 표현식 캡처를 지정하려면 정규 표현식에 하나 이상의 괄호 쌍이 있어야 합니다 .

결과 정규 표현식과 매칭되는 텍스트를 대체할 항목을 지정합니다 . 이 값은 정적 텍스트 , 공백 또는 정규 표현식 캡처의 값일 수 있습니다 . 정규 표현식 캡처는 정규 표현식 정의의 결과로 정의됩니다 .

- 매칭되는 텍스트를 정적 텍스트로 바꾸려면 " 결과 " 필드에 정적 텍스트를 지정합니다 .
- 매칭되는 텍스트를 무시하려면 " 결과 " 필드를 비워 둡니다 .
- 정규 표현식의 가장 바깥쪽 괄호에서 나온 텍스트를 유지하려면 " 결과 " 필드에 "\1"(따옴표 제외) 을 사용합니다 .
- 정규 표현식의 전체 결과를 유지하려면 " 결과 " 필드에 "\0"(따옴표 제외) 을 사용합니다 .

예제 (선택 사항) 정규 표현식의 동작을 나타내는 색상이 포함된 예제 텍스트 문자열을 지정합니다 .

주석 (선택 사항) 정규 표현식과 해당 동작을 설명하는 주석을 지정합니다 .

색상 스크립트 편집기 상자와 " 예제 " 필드의 텍스트에서 정규 표현식의 매칭 항목을 식별하는 데 사용되는 색상을 지정합니다 . 색상을 변경하려면 화살표 버튼을 사용합니다 .

참고 : " 정규 표현식 " 필드의 정규 표현식 정의가 올바르지 않은 경우에는 정규 표현식 목록에서 해당 정규 표현식 이름 옆에 빨간색 X 가 표시됩니다 .

사용자 정규 표현식 생성

사용자 정규 표현식을 생성하려면 다음 단계를 수행하십시오 .

1. 목록 아래의 더하기 기호를 클릭합니다 .
2. " 정규 표현식 라이브러리 선택 " 창에서 " 공백 " 정규 표현식이 선택되어 있는지 확인합니다 .
3. **확인**을 클릭합니다 .
4. " 정규 표현식 편집기 " 패널에서 정규 표현식 정의를 편집합니다 .
5. " 제목 " 필드에서 사용자 정규 표현식의 고유 이름을 지정합니다 .

팁 : 정규 표현식 정의 필드를 편집할 때 로그 창을 열어 표시 상태로 두면 유용합니다 . 일부 오류 메시지는 로그 창에만 나타나기 때문입니다 . 로그 창을 열려면 **보기 > 로그**를 선택하십시오 . 인터넷에서 정규 표현식 문제를 해결하는 데 사용할 수 있는 다양한 리소스를 참조할 수 있습니다 (예 : https://regexr.com/).

단어 구분 기호 목록

단어 구분 기호 목록 버튼을 사용하면 토큰화 프로세스에서 단어 사이에 나타나는 문자의 목록을 지정할 수 있습니다 . 단어 사이 문자는 단어의 시작 문자가 될 수는 없지만 정규 표현식 중 하나에서 허용될 경우 단어 내에는 나타날 수 있습니다 . 이 버튼을 클릭하면 나타나는 창의 목록에서 문자를 추가하거나 제거할 수 있습니다 . 기본적으로 이 목록에 포함된 문자는 공백 문자뿐입니

다. "구분 기호 문자" 창에서 **재설정** 버튼을 클릭하면 구분 기호 문자 목록의 수정 사항이 모두 실행 취소됩니다. 구분 기호 문자 목록의 수정 사항은 현재 정규 표현식 토큰화에만 적용됩니다.

다음 단계에서는 지정된 정규 표현식과 필요한 " 나머지 " 정규 표현식의 처리에 대해 설명합니다 .

1. 텍스트 스트림의 현재 문자를 구분 기호 문자 목록과 비교합니다 .
   - 구분 기호 문자 목록에 있는 문자일 경우 해당 문자를 무시하고 " 나머지 " 임시 문자열에서 취합된 모든 문자를 처리한 후 다음 문자로 이동해서 1 단계를 반복합니다 .
   - 구분 기호 문자 목록에 없는 문자일 경우 2 단계로 이동합니다 .
2. 현재 문자부터 시작되는 문자열을 " 나머지 " 정규 표현식 바로 앞까지의 각 정규 표현식과 한 번에 하나씩 비교합니다 .
   - 현재 문자부터 시작되는 문자열이 정규 표현식 중 하나와 매칭되면 후속 작업이 수행됩니다 . 즉 , " 나머지 " 임시 문자열에 취합된 모든 문자가 처리되고 , " 결과 " 필드의 값이 용어로 저장됩니다 . 텍스트 스트림의 현재 문자는 매칭되는 문자열 다음의 문자가 됩니다 . 그런 다음 1 단계 처리로 돌아갑니다 .
   - 현재 문자부터 시작되는 문자열이 " 나머지 " 정규 표현식까지의 모든 정규 표현식과 매칭되지 않을 경우에는 3 단계로 이동합니다 .
3. 현재 문자를 추가하고 현재 문자를 텍스트 스트림의 다음 문자로 설정하여 " 나머지 " 임시 문자열에 문자를 수집합니다 . 그런 다음 1 단계로 돌아갑니다 .
   - " 나머지 " 임시 문자열은 다른 정규 표현식 중 하나가 매칭될 때까지 한 번에 한 문자씩 취합됩니다 .
   - "나머지" 정규 표현식의 기본 결과는 취합된 "나머지" 임시 문자열을 삭제하는 것입니다.

팁 :

- " 나머지 " 정규 표현식의 결과를 \1 로 설정하면 구두점 표시와 같은 구분 기호 문자를 더 많이 추가할 수 있습니다 . 이렇게 하면 지정된 구두점 표시가 결과에 포함되지 않습니다 .
- " 나머지 " 정규 표현식의 결과를 \1 로 변경하는 대신 다음 중 하나 이상의 작업을 통해 관심 있는 용어를 캡처할 수 있습니다 .
  - 정규 표현식 라이브러리에서 더 많은 정규 표현식을 추가합니다 .
  - 사용자 정규 표현식을 생성합니다 .

데이터 테이블의 각 행마다 텍스트 문자열의 끝에 도달할 때까지 위의 단계대로 처리가 진행됩니다.

결과를 데이터 테이블에 열로 저장

정규 표현식 토큰화 결과가 포함된 새 열을 데이터 테이블에 저장하려면 **열에 저장** 버튼을 클릭합니다. 새 열은 텍스트 탐색기 시작 창에서 지정된 텍스트 열과 이름이 동일한 문자 열입니다. 단, 열 이름이 고유하도록 이름에 번호가 추가됩니다. "열" > "유틸리티" > "텍스트 매칭으로 새

열 생성"에서 독립 실행형 정규 표현식 유틸리티를 사용할 수도 있습니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

참고 : 사용자 정규 표현식 토큰화의 결과를 데이터 테이블의 열에 저장하면 데이터 테이블의 각 행에 있는 원래 텍스트에 대해 정규 표현식 프로세스가 실행됩니다 . 소문자로 변환된 텍스트 문자열에는 이 프로세스가 실행되지 않습니다 .

텍스트 탐색기 정규 표현식 편집기 닫기

텍스트 탐색기 정규 표현식 편집기 창에서 **확인**을 클릭하면 다음과 같은 작업이 수행됩니다 .

1. 텍스트 탐색기 정규 표현식 편집기 창에서 정의된 사용자 정규 표현식이 정규 표현식 라이브러리에 저장됩니다 .

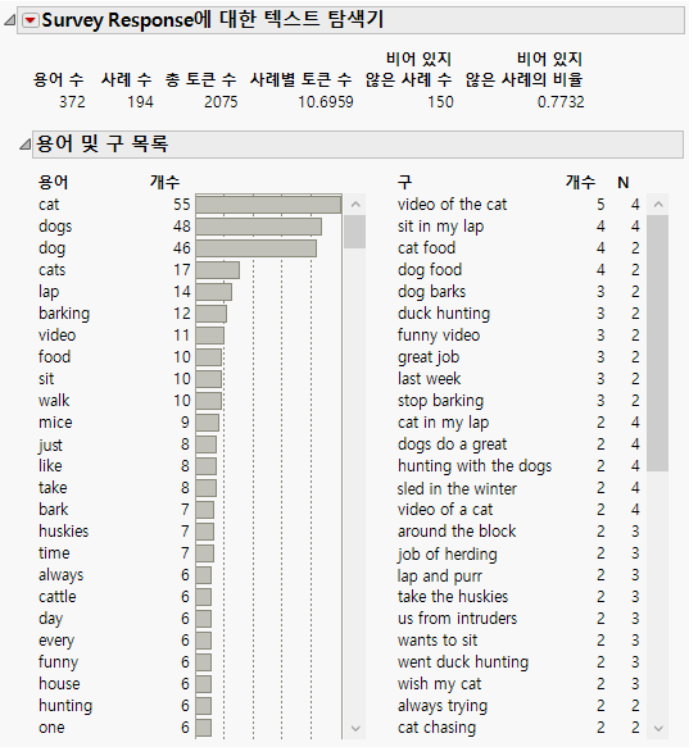
   주의 : 사용자 정규 표현식이 있는 경우 사용자 정규 표현식 라이브러리는 사용자가 **확인**을 클릭할 때만 저장됩니다 . 가장 최근에 저장된 정규 표현식은 다음 번에 사용할 수 있습니다 . 고유한 이름을 사용하여 정규 표현식 라이브러리에 추가 정규 표현식을 보관하십시오 . 나중에 정규 표현식을 사용할 수 있도록 텍스트 탐색기 보고서 창에서 스크립트를 저장할 수 있습니다 .

2. 텍스트 탐색기 보고서가 나타납니다 . 이 보고서에서는 지정된 정규 표현식 설정을 사용하여 텍스트를 토큰화한 결과를 보여 줍니다 .

텍스트 탐색기 보고서

텍스트 탐색기 보고서 창에는 요약 개수 보고서와 " 용어 및 구 목록 " 보고서가 포함됩니다 .

그림 12.8 텍스트 탐색기 보고서의 예



Survey Response에 대한 텍스트 탐색기

용어 수	사례 수	총 토큰 수	사례별 토큰 수	비어 있지 않은 사례 수	비어 있지 않은 사례의 비율
372	194	2075	10.6959	150	0.7732

용어 및 구 목록

용어	개수
cat	55
dogs	48
dog	46
cats	17
lap	14
barking	12
video	11
food	10
sit	10
walk	10
mice	9
just	8
like	8
take	8
bark	7
huskies	7
time	7
always	6
cattle	6
day	6
every	6
funny	6
house	6
hunting	6
one	6

구	개수	N
video of the cat	5	4
sit in my lap	4	4
cat food	4	2
dog food	4	2
dog barks	3	2
duck hunting	3	2
funny video	3	2
great job	3	2
last week	3	2
stop barking	3	2
cat in my lap	2	4
dogs do a great	2	4
hunting with the dogs	2	4
sled in the winter	2	4
video of a cat	2	4
around the block	2	3
job of herding	2	3
lap and purr	2	3
take the huskies	2	3
us from intruders	2	3
wants to sit	2	3
went duck hunting	2	3
wish my cat	2	3
always trying	2	2
cat chasing	2	2

요약 개수 보고서

텍스트 탐색기 보고서의 첫 번째 테이블에는 다음과 같은 요약 통계량이 포함됩니다 .

용어 수 용어 목록의 용어 수입니다 .

사례 수 말뭉치의 문서 수입니다 .

총 토큰 수 말뭉치의 총 용어 수입니다 .

사례별 토큰 수 토큰 수를 사례 수로 나눈 것입니다 .

비어 있지 않은 사례 수 한 개 이상의 용어가 포함된 말뭉치 내 문서의 개수입니다 .

비어 있지 않은 사례의 비율 한 개 이상의 용어가 포함된 말뭉치 내 문서의 비율입니다 .

용어 및 구 목록

"용어 및 구 목록" 보고서에는 토큰화 수행 후 텍스트에 있는 용어 및 구의 테이블이 포함됩니다. "용어 및 구 목록" 보고서의 예는 그림 12.8에서 확인하십시오. 용어 목록의 "개수" 열은 말뭉치에서 각 용어가 나타난 횟수를 나타냅니다. 구 목록의 "개수" 열은 말뭉치에서 각 구가 나타난 횟수를 나타냅니다. "N" 열은 구 내의 단어 수를 나타냅니다.

기본적으로 용어 목록은 개수를 기준으로 내림차순 정렬되며, 개수가 동일한 용어들은 사전순으로 정렬됩니다. 구 목록은 개수를 기준으로 내림차순 정렬되며, 개수가 동일한 구들은 길이("N")를 기준으로 내림차순 정렬됩니다. 구 목록에서 길이도 동일한 구들은 사전순으로 정렬됩니다. 각 목록의 옵션을 사용하여 각 목록의 정렬 순서를 사전순으로 변경할 수 있습니다.

구 목록에 나타나는 구는 시작 창에 있는 **구당 최대 단어 수** 및 **최대 구 수** 옵션의 설정에 따라 결정됩니다. 데이터 테이블에서 한 번만 나타나는 구는 구 목록에 표시되지 않습니다.

구는 다양한 범위에서 용어로 지정될 수 있습니다. 구 목록에서 용어로 지정된 구는 구의 규격 범위를 기준으로 색상이 지정됩니다(표 12.1). 다양한 범위에서 구를 지정하는 방법에 대한 자세한 내용은 "용어 옵션 관리 창"(381 페이지)에서 확인하십시오.

표 12.1 지정된 구의 색상

범위	색상
기본 제공	빨간색
사용자 라이브러리	녹색
프로젝트	파란색
열 특성	주황색
지역	회색

용어 및 구에 대한 작업

각 테이블의 맨 왼쪽 열에서 항목을 선택한 다음 마우스 오른쪽 버튼을 클릭하여 용어 목록 및 구 목록 테이블의 옵션에 액세스할 수 있습니다. 각 테이블의 "개수" 열을 마우스 오른쪽 버튼으로 클릭하고 "데이터 테이블로 만들기"를 선택하여 각 테이블을 데이터 테이블로 저장할 수 있습니다.

용어 목록 팝업 메뉴 옵션

용어 목록 테이블의 " 용어 " 열에서 마우스 오른쪽 버튼을 클릭하면 다음 옵션이 포함된 팝업 메뉴가 나타납니다 .

행 선택하기 데이터 테이블에서 선택된 용어가 포함된 행을 선택합니다 .

텍스트 표시 선택된 용어가 포함된 문서를 표시합니다 .

참고 : 기본적으로 처음 10,000 개의 문서만 표시됩니다 . 선택된 용어가 포함된 문서가 10,000 개를 초과하면 이 한계를 늘릴 수 있는 창이 나타납니다 .

사전순 용어 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 용어가 사전순으로 정렬됩니다 . 이 옵션을 선택하지 않으면 용어가 개수 기준 내림차순으로 정렬됩니다 .

수치 순서 (" 사전순 " 옵션을 선택한 경우에만 사용할 수 있습니다 .) 용어 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 항목이 문자열 및 숫자 세그먼트로 분할되고 숫자 세그먼트가 수치 순서로 정렬됩니다 . " 수치 순서 " 옵션에서 사용되는 정렬 규칙에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

복사 선택된 용어를 클립보드에 추가합니다 .

색상 선택된 용어에 색상을 할당할 수 있습니다 .

라벨 선택된 용어에 대한 용어 SVD 그림의 해당 점에 라벨을 표시합니다 .

포함하는 구 구 목록 테이블에서 선택된 용어가 포함된 구를 선택합니다 .

표시자 저장 용어 목록에서 선택된 각 용어에 대한 표시자 열을 데이터 테이블에 저장합니다 . 각 행의 표시자 열 값은 해당 행의 문서에 선택된 용어가 포함되어 있으면 1 이고 , 그렇지 않으면 0 입니다 .

계산식 저장 용어 목록에서 선택된 각 용어에 대한 열 계산식을 데이터 테이블에 저장합니다 . 각 행의 열 계산식은 해당 행의 문서에 선택된 용어가 포함되어 있으면 1 로 계산되고 , 그렇지 않으면 0 으로 계산됩니다 . 이 옵션은 새 문서에 유용합니다 .

재코딩 하나 이상의 용어에 대한 값을 변경할 수 있습니다 . 이 옵션을 선택하기 전에 목록에서 용어를 선택해야 합니다 . 이 옵션을 선택하면 " 재코딩 " 창이 나타납니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

중지 단어 추가 선택된 용어를 중지 단어 목록에 추가하고 용어 목록에서 해당 용어를 제거합니다 . 이 작업으로 구 목록도 업데이트됩니다 .

참고 : 어간 추출된 단어를 중지 단어로 추가하면 해당 어간과 일치하는 모든 토큰이 중지 단어로 추가됩니다 .

어간 예외 추가 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 선택된 용어를 어간 추출 (Stemming) 에서 제외된 용어의 목록에 추가합니다 .

구 제거 (지정된 구가 용어 목록에서 선택된 경우에만 사용 가능) 지정된 구 집합에서 선택된 구를 제거하고 그에 따라 용어 개수를 업데이트합니다 .

참고 : 구가 감정 구로 추가된 경우 " 구 제거 " 옵션은 현재 감정 분석 보고서의 감정 용어 목록에서도 구를 제거합니다 .

JMP PRO **감정 추가** (현재 보고서 창에 감정 분석 보고서가 열려 있는 경우에만 사용할 수 있습니다 .) 선택한 용어를 현재 감정 분석 보고서의 감정 용어 목록에 추가합니다 .

참고 : 어간 추출된 단어를 감정 용어로 추가하면 해당 어간과 일치하는 모든 토큰이 감정 용어로 추가됩니다 .

필터 표시 용어 목록 위에 검색 필터를 표시하거나 숨깁니다 . 자세한 내용은 " 검색 필터 옵션 "(378 페이지) 에서 확인하십시오 .

데이터 테이블로 만들기 보고서 테이블을 사용하여 JMP 데이터 테이블을 생성합니다 .

결합 데이터 테이블 생성 보고서에서 선택한 테이블과 유사한 다른 테이블을 검색하여 단일 JMP 데이터 테이블에 결합합니다 .

구 목록 팝업 메뉴 옵션

구 목록 테이블의 " 구 " 열에서 마우스 오른쪽 버튼을 클릭하면 다음 옵션이 포함된 팝업 메뉴가 나타납니다 .

행 선택하기 데이터 테이블에서 선택된 구가 포함된 행을 선택합니다 .

텍스트 표시 선택된 구가 포함된 문서를 표시합니다 .

표시자 저장 구 목록에서 선택된 각 구에 대한 표시자 열을 데이터 테이블에 저장합니다 . 각 행의 표시자 열 값은 해당 행의 문서에 선택된 구가 포함되어 있으면 1 이고 , 그렇지 않으면 0 입니다 .

사전순 구 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 용어가 사전순으로 정렬됩니다 . 이 옵션을 선택하지 않으면 용어가 개수 기준 내림차순으로 정렬됩니다 .

수치 순서 (" 사전순 " 옵션을 선택한 경우에만 사용할 수 있습니다 .) 구 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 항목이 문자열 및 숫자 세그먼트로 분할되고 숫자 세그먼트가 수치 순서로 정렬됩니다 . " 수치 순서 " 옵션에서 사용되는 정렬 규칙에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

복사 선택된 구를 클립보드에 추가합니다 .

포함하는 항목 선택 구 목록에서 선택된 구를 포함하는 더 큰 구를 선택합니다 .

포함된 항목 선택 구 목록 또는 용어 목록에서 선택된 구에 포함된 더 작은 구 또는 용어를 선택합니다 .

구 추가 선택된 구를 용어 목록에 추가하고 그에 따라 용어 개수를 업데이트합니다 .

중지 단어 추가 선택된 구를 중지 단어 목록에 추가합니다 . 이 작업으로 용어 목록도 업데이트됩니다 .

JMP PRO **감정 구 추가** (현재 보고서 창에 감정 분석 보고서가 열려 있는 경우에만 사용할 수 있습니다 .) 선택한 구를 용어 목록 및 현재 감정 분석 보고서의 감정 용어 목록에 추가합니다 .

필터 표시 구 목록 위에 검색 필터를 표시하거나 숨깁니다 . 자세한 내용은 " 검색 필터 옵션 "(378 페이지) 에서 확인하십시오 .

데이터 테이블로 만들기 보고서 테이블을 사용하여 JMP 데이터 테이블을 생성합니다 .

결합 데이터 테이블 생성 보고서에서 선택한 테이블과 유사한 다른 테이블을 검색하여 단일 JMP 데이터 테이블에 결합합니다 .

검색 필터 옵션

검색 상자 옆의 아래쪽 화살표 버튼을 클릭하여 검색을 구체화할 수 있습니다 .

용어 포함 검색 기준의 일부가 포함된 항목을 반환합니다 . "ease oom" 을 검색하면 "Release Zoom" 과 같은 메시지가 반환됩니다 .

구 포함 검색 기준이 정확히 포함된 항목을 반환합니다 . "text box" 를 검색하면 "text" 와 "box" 가 바로 연이어 포함된 항목 (예 : "Context Box" 및 "Text Box") 이 반환됩니다 .

구로 시작 검색 기준으로 시작하는 항목을 반환합니다 .

구로 끝남 검색 기준으로 끝나는 항목을 반환합니다 .

전체 구 전체 문자열로 구성된 항목을 반환됩니다 . "text box" 를 검색하면 "text box" 만 포함된 항목이 반환됩니다 .

정규 표현식 검색 상자에서 와일드카드 (*) 와 마침표 (.) 를 사용할 수 있습니다 . "get.*name" 을 검색하면 "get" 다음에 하나 이상의 단어가 포함된 항목을 찾을 수 있습니다 . 즉 , "Get Color Theme Names", "Get Name Info" 및 "Get Effect Names" 등이 반환됩니다 .

결과 반전 검색 기준과 매칭되지 않는 항목을 반환합니다 .

모든 용어 일치 문자열이 모두 포함된 항목을 반환합니다 . "t test" 를 검색하면 검색 문자열 중 하나 또는 둘 모두가 포함된 요소 (예 : "Pat Test", "Shortest Edit Script" 및 "Paired t test") 가 반환됩니다 .

대 / 소문자 무시 검색 기준의 대 / 소문자를 무시합니다 .

전체 단어 일치 " 모든 용어 일치 " 설정에 따라 문자열의 각 단어가 포함된 항목을 반환합니다 . " 모든 용어 일치 " 옵션이 선택되어 있는 경우 "data filter" 를 검색하면 "data" 와 "filter" 가 반환됩니다 .

텍스트 탐색기 플랫폼 옵션

이 섹션에서는 텍스트 탐색기 플랫폼에서 사용할 수 있는 옵션에 대해 설명합니다 .

- " 텍스트 준비 옵션 "
- " 텍스트 분석 옵션 "
- " 저장 옵션 "
- " 보고서 옵션 "

텍스트 준비 옵션

텍스트 탐색기의 빨간색 삼각형 메뉴에는 텍스트 준비를 위한 다음 옵션이 포함되어 있습니다 .

표시 옵션 보고서 표시를 제어하는 옵션의 하위 메뉴를 표시합니다 .

단어 클라우드 표시 " 단어 클라우드 " 보고서를 표시하거나 숨깁니다 . " 단어 클라우드 " 의 빨간색 삼각형 메뉴를 사용하여 단어 클라우드의 레이아웃 및 글꼴을 변경할 수 있습니다 . 자세한 내용은 " 단어 클라우드 옵션 "(381 페이지) 에서 확인하십시오 .

너비를 변경하는 방법으로 단어 클라우드의 크기를 대화식으로 조정할 수 있습니다 . 이 때 높이는 자동으로 결정됩니다 . 용어 목록의 행은 단어 클라우드의 용어에 연결됩니다 .

용어 목록 표시 용어 목록을 표시하거나 숨깁니다 .

구 목록 표시 구 목록을 표시하거나 숨깁니다 .

용어 및 구 옵션 표시 각 목록의 팝업 메뉴에서 사용할 수 있는 옵션에 해당하는 버튼을 "용어 및 구 목록 " 보고서에 표시합니다 . 자세한 내용은 " 용어 및 구 목록 "(375 페이지) 에서 확인하십시오 .

요약 개수 표시 요약 개수 테이블을 표시하거나 숨깁니다 . 자세한 내용은 " 요약 개수 보고서 "(374 페이지) 에서 확인하십시오 .

중지 단어 표시 분석에 사용된 중지 단어의 목록을 표시하거나 숨깁니다 . 처음에는 기본 제공 중지 단어 목록이 사용됩니다 . 중지 단어를 추가하려면 용어 목록을 마우스 오른쪽 버튼으로 클릭한 후 팝업 메뉴에서 **중지 단어 추가**를 선택합니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

재코딩 표시 재코딩된 용어의 목록을 표시하거나 숨깁니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

지정한 구 표시 사용자가 용어로 처리되도록 지정한 구의 목록을 표시하거나 숨깁니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

어간 예외 표시 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출 (stemming) 에서 제외된 용어를 표시하거나 숨깁니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

구분자 표시 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정되고 선택된 토큰화 방법이 " 기본 단어 " 인 경우에만 사용 가능) 기본 단어 토큰화 방법에 사용된 구분자를 표시하거나 숨깁니다 사용된 구분자 집합을 수정하려면 JSL 에서 `Add Delimiters()` 또는 `Set Delimiters()` 메시지를 사용해야 합니다 .

어간 보고서 표시 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정되고 선택된 어간 추출 방법이 " 어간 추출 (Stemming) 안 함 " 인 경우에만 사용 가능) 두 개의 어간 추출 결과 테이블이 포함된 " 어간 추출 (Stemming)" 보고서를 표시하거나 숨깁니다 . 왼쪽의 테이블은 각 어간을 해당 용어에 매핑합니다 . 오른쪽의 테이블은 각 용어를 해당 어간에 매핑합니다 .

선택된 행 표시 현재 선택된 행에 있는 문서의 텍스트가 포함된 창을 엽니다 .

모든 테이블에 대해 필터 표시 보고서의 테이블을 검색하는 데 사용할 수 있는 필터를 표시하거나 숨깁니다 . 이 옵션은 중지 단어 , 지정된 구 , 어간 예외 , 용어 목록 , 구 목록 및 어간 보고서에 적용됩니다 . 필터 도구에 대한 자세한 내용은 " 검색 필터 옵션 "(378 페이지) 에서 확인하십시오 .

용어 옵션 용어 목록에 적용되는 옵션의 하위 메뉴를 표시합니다 .

어간 추출 (Stemming) (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출 옵션에 대한 자세한 내용은 " 텍스트 탐색기 플랫폼 시작 "(366 페이지) 에서 확인하십시오 .

기본 제공 중지 단어 포함 토큰화 프로세스에 사용되는 중지 단어에 기본 제공 중지 단어를 포함할지 여부를 지정합니다 .

기본 제공 구 포함 토큰화 프로세스에 사용되는 구에 기본 제공 구를 포함할지 여부를 지정합니다 .

중지 단어 관리 중지 단어를 추가하거나 제거할 수 있는 창을 표시합니다 . 변경 사항은 사용자 , 열 및 지역 수준에서 적용될 수 있습니다 . 다른 수준에 지정된 중지 단어를 제외하는 지역 예외를 지정할 수도 있습니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

재코딩 관리 재코딩을 추가하거나 제거할 수 있는 창을 표시합니다 . 변경 사항은 사용자 , 열 및 지역 수준에서 적용될 수 있습니다 . 다른 수준에 지정된 재코딩을 제외하는 지역 예외를 지정할 수도 있습니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

구 관리 용어로 처리되는 구를 추가하거나 제거할 수 있는 창을 표시합니다 . 변경 사항은 사용자 , 열 및 지역 수준에서 적용될 수 있습니다 . 다른 수준에 지정된 구를 제외하는 지역 예외를 지정할 수도 있습니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

어간 예외 관리 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출 예외를 추가하거나 제거할 수 있는 창을 표시합니다 . 변경 사항은 사용자 , 열 및 지역 수준에서 적용될 수 있습니다 . 다른 수준에 지정된 어간 예외를 제외하는 지역 예외를 지정할 수도 있습니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

파싱 옵션 파싱 및 토큰화에 적용되는 옵션의 하위 메뉴를 표시합니다 .

토큰화 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 토큰화 옵션에 대한 자세한 내용은 " 텍스트 탐색기 플랫폼 시작 "(366 페이지) 에서 확인하십시오 .

정규 표현식 사용자 정의 (정규 표현식 토큰화 방법을 사용하는 경우에만 사용 가능) " 정규 표현식 사용자 정의 " 창을 표시합니다 . 이 옵션을 사용하여 현재 텍스트 탐색기 보고서에 대한 정규 표현식 설정을 수정할 수 있습니다 .

참고 : 플랫폼 시작 창에서 기준 변수를 지정한 경우에는 " 정규 표현식 사용자 정의 " 옵션이 기준 변수의 모든 수준에 일괄 적용됩니다 .

숫자를 단어로 처리 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정되고 선택된 토큰화 방법이 " 기본 단어 " 인 경우에만 사용 가능) 분석에서 숫자를 용어로 토큰화할 수 있도록 합니다 . 이 옵션은 단어당 최소 문자 수에 대한 설정에 따라 영향을 받습니다 .

단어 클라우드 옵션

" 단어 클라우드 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

레이아웃 단어 클라우드의 용어 배열 방식을 지정합니다 . 기본 레이아웃은 " 정렬됨 " 으로 설정되어 있습니다 .

정렬됨 용어를 빈도가 높은 것부터 낮은 것 순서로 가로 줄로 표시합니다 .

사전순 용어를 오름차순 사전순으로 정렬하여 가로 줄로 표시합니다 .

중심화됨 용어를 빈도에 따른 크기로 클라우드에 표시합니다 .

색상 적용 단어 클라우드의 용어 색상을 지정합니다 . 기본 색상은 " 없음 " 으로 설정되어 있습니다 .

없음 각 용어를 용어 목록의 색상과 동일한 색상으로 표시합니다 .

균등 색상 각 용어를 동일한 색상으로 표시합니다 . 범례에서 이 색상을 변경할 수 있습니다 .

임의 회색 각 용어를 다양한 회색 음영으로 표시합니다 .

임의 색상 각 용어를 다양한 색상으로 표시합니다 . 범례에서 색상을 조정할 수 있습니다 .

열 값별 각 용어를 그래디언트 색상 척도로 표시합니다 . 척도는 " 열을 기준으로 용어 점수 매기기 " 옵션으로 생성된 용어 점수를 기준으로 합니다 . 범례에서 색상과 그래디언트를 조정할 수 있습니다 .

글꼴 단어 클라우드의 용어에 적용할 글꼴 , 스타일 및 크기를 지정합니다 .

범례 표시 단어 클라우드의 범례를 표시하거나 숨깁니다 .

용어 옵션 관리 창

다양한 범위에 대해 중지 단어, 재코딩, 구 및 어간 예외 정보를 지정할 수 있습니다. 이러한 정보는 텍스트 탐색기 사용자 라이브러리(사용자 범위), 현재 프로젝트, 분석 열에 대한 열 특성(열 범위) 또는 플랫폼 스크립트(지역 범위)에 저장할 수 있습니다. 텍스트 탐색기 보고서의 스크립트를 저장하여 특정 텍스트 탐색기 인스턴스에 대한 지역 규격 및 지역 예외를 저장할 수 있습니다.

용어 옵션 관리 창은 중지 단어, 재코딩, 구 및 어간 예외의 모음을 관리할 수 있는 네 가지 유사한 창입니다. 그림 12.9에서는 "중지 단어 관리" 창을 보여 줍니다. "구 관리" 및 "어간 예외 관리" 창은 "중지 단어 관리" 창과 동일합니다. "재코딩 관리" 창은 약간 다릅니다. 자세한 내용은 "재코딩 관리"(383 페이지)에서 확인하십시오.

그림 12.9 중지 단어 관리 창

중지 단어 관리

" 중지 단어 관리 " 창에는 지정된 중지 단어의 다양한 범위 (또는 위치) 를 나타내는 중지 단어 목록이 여러 개 포함되어 있습니다 . 각 목록 아래에는 텍스트 편집 상자와 추가 버튼이 있습니다 . 이러한 컨트롤을 사용하여 각 범위에 사용자 중지 단어를 추가할 수 있습니다 . 중지 단어를 드래그하여 한 범위에서 다른 범위로 이동할 수 있습니다 . 한 목록에서 다른 목록으로 항목을 복사하여 붙여넣을 수도 있습니다 . 창 아래에 있는 두 개의 버튼을 클릭하면 선택된 항목이 한 범위에서 왼쪽 또는 오른쪽의 다음 범위로 이동합니다 . X 버튼을 클릭하면 선택된 항목이 현재 범위에서 제거됩니다 . 또한 항목을 두 번 클릭하고 텍스트를 변경하여 목록의 기존 항목을 편집할 수 있습니다 .

언어 기본 제공 중지 단어 목록과 사용자 라이브러리의 선택 항목을 저장할 언어를 지정합니다 . ' 언어 " 에서 " 항목 적용 " 을 선택하면 변경 사항이 주 사용자 라이브러리에 저장됩니다 . 언어 설정은 기본 제공 범위와 사용자 및 프로젝트 범위에만 적용됩니다 .

기본 제공 (잠김) 지정된 언어의 중지 단어에 대한 기본 제공 목록입니다 . 기본 제공 중지 단어를 " 지역 예외 " 목록에 포함하면 중지 단어 목록에서 제외할 수 있습니다 .

사용자 지정된 언어의 사용자 라이브러리에 있는 중지 단어를 나열합니다 .

프로젝트 (텍스트 탐색기가 프로젝트 내에서 시작된 경우에만 사용할 수 있습니다 .) 지정된 언어의 현재 프로젝트에 있는 중지 단어를 나열합니다 .

열 텍스트 열의 " 중지 단어 " 열 특성에 지정된 중지 단어를 나열합니다 .

지역 지역 범위의 중지 단어를 나열합니다 . JSL 을 통해 텍스트 탐색기를 시작할 때 이러한 단어를 지정할 수 있습니다 . 이러한 중지 단어는 현재 텍스트 탐색기 플랫폼 보고서에서만 사용됩니다 .

지역 예외 현재 텍스트 탐색기 플랫폼에서 중지 단어로 처리되지 않는 단어를 나열합니다 . JSL 을 통해 텍스트 탐색기를 시작할 때 이러한 단어를 지정할 수 있습니다 . " 지역 예외 " 에 나열된 단어는 다른 모든 범위에 나열된 단어를 재정의합니다 .

가져오기 텍스트 파일에서 중지 단어를 가져올 수 있습니다 . 중지 단어는 클립보드에 복사됩니다 . 이를 기본 제공 목록 이외의 목록에 붙여넣을 수 있습니다 .

내보내기 중지 단어를 클립보드 또는 텍스트 파일로 내보낼 수 있습니다 . 중지 단어를 내보낼 범위와 내보낼 위치를 선택할 수 있는 " 내보내기 " 창이 나타납니다 .

사용자 라이브러리 파일은 TextExplorer 디렉터리에 있습니다 . 이 디렉터리의 위치는 다음과 같이 컴퓨터 운영 체제에 따라 다릅니다 .

- Windows: "C:/Users/< 사용자 이름 >/AppData/Roaming/SAS/JMP/TextExplorer/< 언어 >/"
- macOS: "/Users/< 사용자 이름 >/Library/Application Support/JMP/TextExplorer/< 언어 >/"

주 사용자 라이브러리 파일은 TextExplorer 디렉터리 자체에 있습니다. 이러한 파일은 언어별로 다르지 않습니다.

확인을 클릭하면 "사용자", "프로젝트" 및 "열" 목록의 변경 사항이 각각 사용자 라이브러리, 프로젝트 및 열 특성에 저장됩니다. "지역" 및 "지역 예외" 목록에 지정된 모든 항목은 텍스트 탐색기 보고서 스크립트를 저장할 때만 저장됩니다.

중지 단어를 사용자 라이브러리에 저장할 경우 파일 이름은 stopwords.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "중지 단어"라고 합니다.

재코딩 관리

"재코딩 관리" 창은 "중지 단어 관리" 창과는 약간 다릅니다. 각 목록 아래에는 텍스트 편집 상자가 하나가 아니라 두 개 있습니다. 이전 값(위쪽 상자에서 지정)은 새 값(아래쪽 상자에서 지정)으로 재코딩됩니다.

사용자 라이브러리에 재코딩을 저장하는 경우 파일 이름은 recodes.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "재코딩"이라고 합니다.

구 관리

사용자 라이브러리에 구를 저장하는 경우 파일 이름은 phrases.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "구"라고 합니다.

어간 예외 관리

사용자 라이브러리에 어간 예외를 저장하는 경우 파일 이름은 stemExceptions.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "어간 예외"라고 합니다.

참고 : " 어간 예외 관리 " 창의 " 로컬 예외 " 목록에는 어간 예외 목록에서 제외되는 어간 예외가 나열됩니다 . 이 목록에 있는 단어는 어간 추출 작업에 관련되지 않습니다 .

JMP PRO 부정 용어 관리

"부정 용어 관리" 창은 감정 분석 보고서에서 사용할 수 있습니다. 자세한 내용은 "감정 분석"(399 페이지)에서 확인하십시오.

사용자 라이브러리에 부정 용어를 저장하는 경우 파일 이름은 negations.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "부정 용어"라고 합니다.

참고: "지역" 또는 "지역 예외" 목록에 나타나는 용어는 현재 감정 분석 보고서에만 적용됩니다.

JMP PRO 강조사 용어 관리

감정 분석 보고서에서 사용할 수 있는 "강조사 용어 관리" 창은 "중지 단어 관리" 창과는 약간 다릅니다. 각 목록 아래에 텍스트 편집 상자가 있을 뿐만 아니라 승수 컨트롤도 있습니다. 승수 컨트롤을 사용하면 강조사 용어 집합에 추가할 때 용어의 강조 승수를 지정할 수 있습니다. 자세한 내용은 "감정 분석"(399 페이지)에서 확인하십시오.

사용자 라이브러리에 강조사 용어를 저장하는 경우 파일 이름은 intensifiers.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "강조사 용어"라고 합니다.

참고: "지역" 또는 "지역 예외" 목록에 나타나는 용어는 현재 감정 분석 보고서에만 적용됩니다.

JMP PRO 감정 용어 관리

감정 분석 보고서에서 사용할 수 있는 "감정 용어 관리" 창은 "중지 단어 관리" 창과는 약간 다릅니다. 각 목록 아래에 텍스트 편집 상자가 있을 뿐만 아니라 점수 컨트롤도 있습니다. 점수 컨트롤을 사용하면 감정 용어 집합에 추가할 때 용어의 감정 점수를 지정할 수 있습니다. 자세한 내용은 "감정 분석"(399 페이지)에서 확인하십시오.

사용자 라이브러리에 감정 용어를 저장하는 경우 파일 이름은 sentiments.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "감정 용어"라고 합니다.

참고: "지역" 또는 "지역 예외" 목록에 나타나는 용어는 현재 감정 분석 보고서에만 적용됩니다.

JMP PRO 텍스트 분석 옵션

텍스트 탐색기의 빨간색 삼각형 메뉴에는 분석을 위한 다음 옵션이 포함되어 있습니다.

잠재 계층 분석 희소 행렬 루틴을 사용하여 이진 가중 문서 용어 행렬에 대해 잠재 계층 분석을 수행합니다. 자세한 내용은 "잠재 계층 분석"(387 페이지)에서 확인하십시오.

텍스트 탐색기의 빨간색 삼각형 메뉴에서 "잠재 계층 분석"을 선택하면 다음 옵션이 포함된 규격 창이 나타납니다.

최대 용어 수 잠재 계층 분석에 포함되는 용어의 최대 개수입니다.

최소 용어 빈도 한 용어가 잠재 계층 분석에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다.

군집 수 잠재 계층 분석의 군집 수입니다.

잠재 의미 분석 , SVD 문서 용어 행렬의 부분 특이값 분해를 수행합니다. 자세한 내용은 "잠재 의미 분석 (SVD)"(389 페이지) 에서 확인하십시오 .

판별 분석 문서 용어 행렬을 기준으로 그룹 또는 범주에서 각 문서의 소속을 예측합니다 . 자세한 내용은 " 판별 분석 "(394 페이지) 에서 확인하십시오 .

용어 선택 서로 다른 응답을 가장 잘 설명하는 용어를 분석합니다 . 용어 선택은 응답이 평가인 경우 감정 분석에도 유용할 수 있습니다 . 자세한 내용은 " 용어 선택 "(396 페이지) 에서 확인하십시오 .

감정 분석 (" 언어 " 옵션이 영어로 설정된 경우에만 사용할 수 있습니다 .) 어휘 분석을 사용하여 문서에서 감정 용어를 식별하고 긍정 , 부정 및 전반적 감정에 대해 문서의 점수를 매깁니다 . 자세한 내용은 " 감정 분석 "(399 페이지) 에서 확인하십시오 .

JMP PRO 특이값 분해 규격 창

텍스트 탐색기의 분석 옵션은 DTM(문서 용어 행렬)을 기반으로 합니다. DTM은 용어 목록에 있는 각 용어마다 열을 하나씩 생성함으로써 구성됩니다(지정된 최대 용어 수까지). 각 텍스트 문서(데이터 테이블의 행과 동등)는 DTM의 한 행에 해당합니다. DTM의 셀 값은 "규격" 창에서 사용자가 지정한 가중 유형에 따라 달라집니다.

그림 12.10 에서는 특이값 분해를 위한 " 규격 " 창을 보여 줍니다 . 텍스트 탐색기의 빨간색 삼각형 메뉴에서 문서 용어 행렬에 대해 특이값 분해를 수행하는 옵션을 선택하면 다음 옵션이 포함된 " 규격 " 창이 나타납니다 .

최대 용어 수 특이값 분해에 포함되는 용어의 최대 개수입니다 .

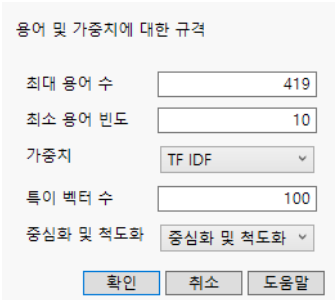
최소 용어 빈도 한 용어가 특이값 분해에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다 .

가중치 문서 용어 행렬의 셀에 들어갈 값을 결정하는 가중치 체계입니다 . 가중치 체계 옵션에 대한 자세한 내용은 " 문서 용어 행렬 규격 창 "(386 페이지) 에서 확인하십시오 .

특이 벡터 수 특이값 분해의 특이 벡터 수입니다 . 기본값은 문서 수 , 용어 수 또는 100 중 최소값입니다 .

중심화 및 척도화 문서 용어 행렬을 중심화하고 척도화하기 위한 옵션입니다 . **중심화 및 척도화** , **중심화** 및 **비중심화** 중에서 선택할 수 있습니다 . 기본적으로 문서 용어 행렬은 중심화되고 척도화됩니다 .

그림 12.10 SVD 규격 창



저장 옵션

텍스트 탐색기의 빨간색 삼각형 메뉴에는 데이터 테이블, 테이블 열 및 열 특성에 정보를 저장하기 위한 다음 옵션이 포함되어 있습니다.

문서 용어 행렬 저장 문서 용어 행렬의 각 열에 해당하는 열을 데이터 테이블에 저장합니다 (지정된 최대 용어 수까지).

JMP PRO **연관성을 위해 쌓인 형식의 DTM 저장** 쌓인 형식의 문서 용어 행렬을 JMP 데이터 테이블에 저장합니다. 쌓인 형식은 연관성 분석 플랫폼의 분석에 적합합니다. 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오. 텍스트 탐색기 시작 창에서 ID 변수를 지정하면 원래 텍스트 데이터 테이블에서 각 용어가 있던 행을 식별하는 데 ID 변수가 사용됩니다. 쌓인 테이블에는 연관성 분석을 시작하기 위한 테이블 스크립트도 포함됩니다.

DTM 계산식 저장 벡터 모델링 유형의 계산식 열을 데이터 테이블에 저장합니다. 벡터의 길이는 최대 용어 수, 최소 용어 빈도 및 가중치에 대해 사용자가 지정한 옵션에 따라 달라집니다. 결과 열에는 `Text Score()` JSL 함수가 사용됩니다. 이 함수에 대한 자세한 내용은 " 도움말 " > " 스크립트 인덱스 " 에서 확인하십시오.

용어 테이블 저장 용어 목록의 각 용어, 발생 횟수 및 각 용어가 포함된 문서 수가 포함된 JMP 데이터 테이블을 생성합니다. " 용어 테이블 저장 " 을 선택한 후 " 열을 기준으로 용어 점수 매기기 " 옵션을 선택하면 용어별 점수가 포함된 열이 " 용어 테이블 저장 " 옵션으로 생성한 데이터 테이블에 추가됩니다.

열을 기준으로 용어 점수 매기기 지정된 열의 값을 기준으로 한 점수를 " 용어 테이블 저장 " 옵션으로 생성된 JMP 데이터 테이블에 저장합니다. 각 용어의 점수는 지정된 열에 대한 평균 값으로, 이 값은 각 행에서 해당 용어가 나타나는 횟수를 기준으로 가중치를 부여한 값입니다. " 용어 테이블 저장 " 옵션을 이미 선택한 경우 " 열을 기준으로 용어 점수 매기기 " 옵션은 " 용어 테이블 저장 " 옵션으로 생성된 데이터 테이블에 점수를 포함하는 열을 추가합니다. 그렇지 않은 경우에는 용어 테이블에 대한 JMP 데이터 테이블이 생성됩니다. 지정된 열이 연속형이 아니면 지정된 열의 각 수준에 대한 점수를 포함하는 열이 생성됩니다.

문서 용어 행렬 규격 창

텍스트 탐색기의 빨간색 삼각형 메뉴에서 " 문서 용어 행렬 저장 " 및 "DTM 계산식 저장 " 옵션을 선택하면 다음 옵션이 포함된 문서 용어 행렬 " 규격 " 창이 나타납니다.

최대 용어 수 문서 용어 행렬에 포함되는 용어의 최대 개수입니다.

최소 용어 빈도 한 용어가 문서 용어 행렬에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다.

가중치 문서 용어 행렬의 셀에 들어갈 값을 결정하는 가중치 체계입니다.

" 가중치 " 에는 다음 옵션을 사용할 수 있습니다.

이진 각 문서에서 용어가 나오면 1 을 할당하고, 그렇지 않으면 0 을 할당합니다. SVD 분석이 이전에 실행되었던 경우를 제외하고는 이 가중 방법이 기본값입니다.

삼진 각 문서에서 용어가 두 번 이상 나오면 2 를 할당하고 , 한 번만 나오면 1 을 할당하며 , 나오지 않으면 0 을 할당합니다 .

빈도 각 문서에서의 용어 발생 횟수를 할당합니다 .

로그 빈도 *log10(1 + x)* 를 할당합니다 . 여기서 *x* 는 각 문서에서의 용어 발생 횟수입니다 .

TF IDF *TF * log10(nDoc / nDocTerm)* 을 할당합니다 . 용어 빈도 - 문서 빈도 역수 (term frequency - inverse document frequency) 의 약어입니다 . 이 옵션이 기본 가중치 옵션입니다 . 계산식의 항은 다음과 같이 정의됩니다 .

TF = 문서 내 용어 빈도

nDoc = 말뭉치 내 문서 수

nDocTerm = 해당 용어를 포함하는 문서의 수

참고 : SVD 분석을 실행한 후 " 문서 용어 행렬 저장 " 또는 "DTM 계산식 저장 " 을 선택하면 " 규격 " 창에 가장 최근의 SVD 분석 규격이 포함됩니다 .

보고서 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO 잠재 계층 분석

텍스트 탐색기 플랫폼에서 잠재 계층 분석을 통해 말뭉치의 문서를 유사 문서의 군집으로 그룹화할 수 있습니다 . 잠재 계층 분석 보고서에는 모형 규격 , 모형에 대한 BIC(베이지안 정보 기준) 값 , 그리고 " 텍스트 표시 " 버튼이 포함됩니다 . " 군집 혼합 확률 " 테이블에서 하나 이상의

군집을 선택하고 " 텍스트 표시 " 버튼을 클릭하면 선택한 군집에 속할 가능성이 가장 높은 문서의 텍스트가 포함된 창이 열립니다 .

" 잠재 계층 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

표시 옵션 " 잠재 계층 분석 " 보고서의 내용을 지정합니다 . 기본적으로는 각 군집의 단어 클라우드를 제외한 모든 보고서 옵션이 표시됩니다 .

군집 혼합 확률 관측값이 각 군집에 속할 확률을 보여 주는 테이블을 표시하거나 숨깁니다 .

팁 : " 군집별 혼합 확률 " 테이블에서 하나 이상의 행을 선택하여 해당 군집에 할당된 관측값을 선택할 수 있습니다 .

군집별 용어 확률 문서가 특정 군집에 속할 경우 군집별로 문서에 용어가 포함될 조건부 확률의 추정값이 들어 있는 용어 테이블을 표시하거나 숨깁니다 . 기본적으로 이 테이블의 용어는 말뭉치에서의 빈도를 기준으로 내림차순 정렬됩니다 .

" 가장 특징적인 군집 " 열에는 각 용어가 가장 높은 비율로 발생하는 군집이 표시됩니다 .

" 가장 확률이 높은 군집 " 열에는 각 용어가 포함된 문서 중에서 무작위로 선택된 문서가 발견될 확률이 가장 높은 군집이 표시됩니다 .

군집별 상위 용어 각 군집에서 점수가 가장 높은 10 개 용어의 테이블을 표시하거나 숨깁니다 . 군집 c 의 용어 t 에 대한 점수 $S_{t,c}$ 는 다음과 같이 계산됩니다 .

$$S_{t,\,c} = 100 \bullet \text{mean}(p_t) \bullet \log 10\left(\frac{p_{t,\,c}}{\text{mean}(p_t)}\right)$$

여기서 $\text{mean}(p_t)$ 는 용어 t 의 군집별 평균 용어 확률이고 , $p_{t,c}$ 는 군집 c 의 용어 t 에 대한 군집별 용어 확률입니다 .

MDS 그림 군집의 근접성을 2 차원으로 표현한 다차원 척도 그림을 표시하거나 숨깁니다 . MDS 그림에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 . " 텍스트 표시 " 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다 .

행별 군집 확률 각 행의 군집 소속 확률을 표시하는 혼합 확률 테이블을 표시하거나 숨깁니다 . " 가장 가능성이 큰 군집 " 은 각 행의 소속 확률이 가장 높은 군집을 나타냅니다 .

군집별 단어 클라우드 단어 클라우드를 군집당 하나씩 포함하는 행렬을 표시하거나 숨깁니다 .

군집 이름 바꾸기 하나 이상의 군집에 대한 이름을 추가할 수 있습니다 .

확률 저장 " 혼합 확률 " 테이블의 값을 데이터 테이블의 해당 행에 저장합니다 .

확률 계산식 저장 각 군집에 대한 계산식 열과 가장 가능성이 높은 군집에 대한 계산식 열을 데이터 테이블에 저장합니다 .

저장되는 점수 계산식에서는 가중치 인수를 "LCA" 로 설정하고 Text Score() JSL 함수를 사용합니다 .

군집별 색상 데이터 테이블의 각 행을 가장 가능성이 큰 군집에 따라 색상을 적용합니다 .

제거 텍스트 탐색기 보고서에서 " 잠재 계층 분석 " 보고서를 제거합니다 .

잠재 계층 분석에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

참고 : 텍스트 탐색기 플랫폼에서 사용되는 LCA 알고리즘은 문서 용어 행렬의 희소성을 활용합니다 . 이러한 이유로 텍스트 탐색기 플랫폼의 LCA 결과는 잠재 계층 분석 플랫폼의 결과와 정확하게 매칭되지 않습니다 .

JMP PRO 잠재 의미 분석 (SVD)

텍스트 탐색기 플랫폼에서 잠재 의미 분석은 DTM(문서 용어 행렬)의 부분 SVD(특이값 분해)를 계산하는 것을 중심으로 이루어집니다. 이 분해 방법에서는 분석을 위해 텍스트 데이터를 관리 가능한 차원 수로 축소합니다. 잠재 의미 분석은 PCA(주성분 분석)를 수행하는 것과 동등합니다.

부분 특이값 분해는 세 개의 행렬 **U**, **S** 및 **V**′ 를 사용하여 DTM 에 근사한 값을 산출합니다 . 이러한 행렬 간의 관계는 다음과 같이 정의됩니다 .

$$DTM \approx \mathbf{U} * \mathbf{S} * \mathbf{V}'$$

*nDoc*는 DTM의 문서(행) 수로 정의되고, *nTerm*은 DTM의 용어(열) 수로 정의되며, *nVec*은 지정된 특이값 수로 정의됩니다. *nVec*은 min(*nDoc*, *nTerm*)보다 작거나 같아야 합니다. 따라서 **U**는 DTM의 왼쪽 특이 벡터를 포함하는 *nDoc* x *nVec* 행렬입니다. **S**는 *nVec*차 주대각 행렬입니다. **S**의 주대각 항목은 DTM의 특이값입니다. **V**′는 *nVec* x *nTerm* 행렬입니다. **V**′의 행(또는 **V**의 열)은 오른쪽 특이 벡터입니다.

오른쪽 특이 벡터는 의미 또는 주제 영역이 유사한 서로 다른 용어 간의 연결성을 포착합니다. 세 개의 용어가 동일한 문서에서 나타나는 경향이 있는 경우 SVD에서는 **V**′에 이 세 개의 용어에 대해 큰 값을 갖는 특이 벡터를 생성할 가능성이 높습니다. **U** 특이 벡터는 이 새로운 용어 공간에 투영된 문서를 나타냅니다.

잠재 의미 분석에서는 간접적 연결성도 포착합니다. 두 개의 단어가 동일한 문서에 함께 나타나지는 않지만 일반적으로 다른 세 번째 단어가 있는 문서에서 나타나는 경우 SVD는 해당 연결성의 일부를 포착할 수 있습니다. 두 개의 문서에 공통된 단어는 없지만 차원 감소 공간에서 연결되는 단어가 포함된 경우 이 두 문서는 SVD 출력에서 유사 벡터에 매핑됩니다.

SVD는 텍스트 데이터를 고정 차원 벡터 공간으로 변환하여 모든 유형의 군집화, 분류 및 회귀 기법에 적용할 수 있도록 합니다. 저장 옵션을 사용하면 이 벡터 공간을 다른 JMP 플랫폼에서 분석할 수 있는 형태로 내보낼 수 있습니다.

기본적으로는 특이값 분해가 수행되기 전에 DTM을 중심화하고 척도화한 후 *nDoc* - 1로 나눕니다. 이 분석은 DTM 상관 행렬의 PCA와 동등합니다.

" 규격 " 창에서 " 중심화 " 또는 " 비중심화 " 를 지정할 수도 있습니다 .

- " 중심화 " 를 지정할 경우 특이값 분해 전에 DTM 이 중심화된 후 *nDoc* - 1 로 나뉩니다 . 이 분석은 DTM 공분산 행렬의 PCA 와 동등합니다 .
- " 비중심화 " 를 지정할 경우에는 특이값 분해 전에 DTM 이 *nDoc* 로 나뉩니다 . 이 분석은 비척도화 DTM 의 PCA 와 동등합니다 .

SVD 구현에서는 DTM이 중심화된 경우에도 DTM의 희소성을 활용합니다.

JMP PRO SVD 보고서

텍스트 탐색기 플랫폼의 "잠재 의미 분석" 옵션은 특이값 분해를 통해 두 개의 SVD 그림과 한 개의 특이값 테이블을 생성합니다 .

JMP PRO SVD 그림

첫 번째 그림에는 각 문서에 대한 점이 포함되어 있습니다. 특정 문서에 대해 표시되는 점은 처음 두 개의 특이 벡터(**U** 행렬의 처음 두 열)에 있는 문서 값을 주대각 특이값 행렬(**S**)로 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 점수 그림과 동등합니다. 이 그림의 각 점은 문서(데이터 테이블의 행)를 나타냅니다. 이 그림에서 점을 선택하여 데이터 테이블의 해당 행을 선택할 수 있습니다.

두 번째 그림에는 각 용어에 대한 점이 포함되어 있습니다. 특정 용어에 대해 표시되는 점은 처음 두 개의 특이 벡터(**V'** 행렬의 처음 두 행)에 있는 용어 값을 주대각 특이값 행렬(**S**)로 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 적재 그림과 동등합니다. 이 그림의 점은 용어 목록 테이블의 행에 해당합니다.

각 SVD 그림 위에 있는 "텍스트 표시" 버튼을 클릭하면 그림에서 선택된 점의 텍스트가 포함된 창을 열 수 있습니다.

JMP PRO 특이값

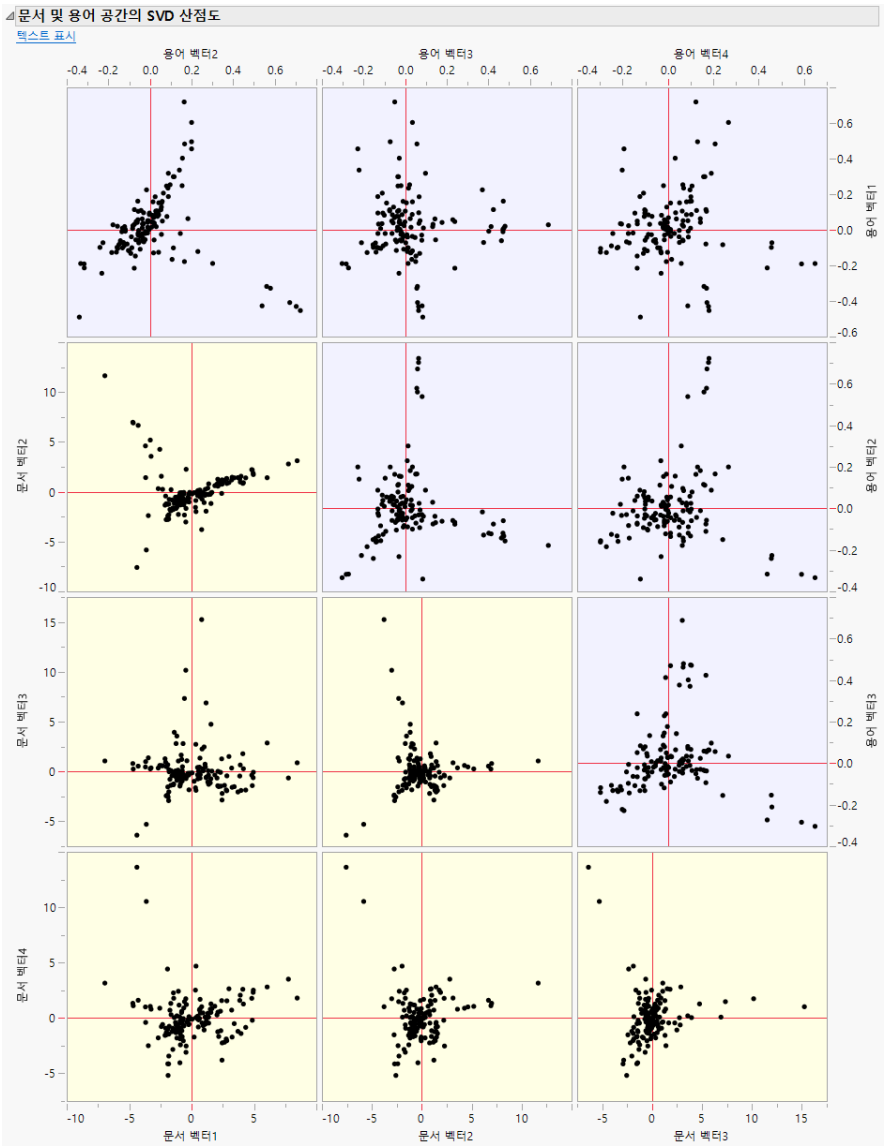
문서 및 용어 SVD 그림 아래에는 특이값 테이블이 표시됩니다. 이 특이값들은 문서 용어 행렬의 특이값 분해에서 **S** 행렬의 주대각 항목입니다. 특이값 테이블에는 동등한 주성분 분석의 해당 고유값이 들어 있는 열이 포함되어 있습니다. 주성분 플랫폼에서와 마찬가지로 각 고유값(또는 특이값)이 설명하는 변동의 백분율 및 누적 백분율을 나타내는 열이 있습니다. "누적 백분율" 열을 사용하여 DTM에서 유지할 분산의 백분율을 결정한 다음 해당하는 수의 특이 벡터를 사용할 수 있습니다.

JMP PRO SVD 보고서 옵션

텍스트 탐색기 플랫폼의 SVD 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

SVD 산점도 행렬 용어 및 문서 특이값 분해 벡터에 대한 산점도 행렬을 표시하거나 숨깁니다 . 이 옵션을 선택할 경우 산점도 행렬의 크기를 선택해야 합니다 . 이 산점도 행렬을 사용하면 특이값 분해의 처음 두 차원보다 더 많은 차원을 시각화할 수 있습니다 . " 텍스트 표시 " 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다 .

그림 12.11 문서 및 용어 공간의 SVD 산점도



주제 분석 , 회전된 SVD 문서 용어 행렬의 Varimax 회전 부분 특이값 분해를 수행하여 용어 그룹 , 즉 주제를 추출합니다 . 이 옵션을 여러 번 선택하여 서로 다른 개수의 주제를 찾을 수 있습니다 . 자세한 내용은 " 주제 분석 "(392 페이지) 에서 확인하십시오 .

군집 용어 데이터에 있는 용어의 계층적 군집화 분석을 표시하거나 숨깁니다 . 덴드로그램 오른쪽에는 군집 수를 설정하고 군집을 데이터 테이블에 저장하기 위한 옵션이 있습니다 . 이 데이터 테이블에는 각 용어의 빈도 , 해당 용어가 포함된 문서의 개수 및 할당된 군집이 포함됩니다 . 계층적 군집화 및 덴드로그램에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

군집 문서 데이터에 있는 문서의 계층적 군집화 분석을 표시하거나 숨깁니다 . 덴드로그램 오른쪽에는 군집 수를 설정하고 , 군집을 데이터 테이블의 열에 저장하고 , 덴드로그램 그림의 선택된 분기에 문서를 표시하기 위한 옵션이 있습니다 .

근접 이웃 선택 문서 SVD 그림에서 선택한 점의 최근접 이웃을 찾아 선택합니다 . 이 알고리즘은 특이값 분해의 **U** 행렬을 사용하여 최근접 이웃을 찾습니다 . 이 옵션을 선택할 때 선택할 최근접 이웃 수를 지정해야 합니다 . 기본적으로 이 옵션은 10 개의 최근접 이웃을 선택합니다 .

문서 특이 벡터 저장 문서 특이값 분해에서 구한 특이 벡터 중 사용자가 지정한 수만큼을 데이터 테이블에 열로 저장합니다 . 처음 두 개의 저장된 열은 문서 SVD 그림에 표시된 점을 나타냅니다 . 자세한 내용은 " 잠재 의미 분석 (SVD)"(389 페이지) 에서 확인하십시오 .

특이 벡터 계산식 저장 문서 특이값 분해를 포함하는 벡터 모델링 유형의 계산식 열을 데이터 테이블에 저장합니다 . 결과 열에는 `Text Score()` JSL 함수가 사용됩니다 . 이 함수에 대한 자세한 내용은 " 도움말 " > " 스크립트 인덱스 " 에서 확인하십시오 .

용어 특이 벡터 저장 용어 특이값 분해에서 구한 특이 벡터 중 사용자가 지정한 수만큼을 새 데이터 테이블에 열로 저장합니다 . 이 데이터 테이블에서 각 행은 하나의 용어에 해당합니다 . 용어 테이블 데이터 테이블이 이미 열려 있는 경우 이 옵션을 사용하면 열이 해당 데이터 테이블에 저장됩니다 . 처음 두 개의 저장된 열은 용어 SVD 그림에 표시된 점을 나타냅니다 . 자세한 내용은 " 잠재 의미 분석 (SVD)"(389 페이지) 에서 확인하십시오 .

제거 텍스트 탐색기 보고서 창에서 SVD 보고서를 제거합니다 .

JMP PRO 주제 분석

" 주제 분석 , 회전된 SVD" 옵션은 DTM(문서 용어 행렬) 의 부분 SVD(특이값 분해) 에 대해 Varimax 회전을 수행합니다 . DTM 에서 유지할 주제의 개수에 해당하는 회전 특이 벡터의 수를 지정해야 합니다 . 주제 수를 지정하면 주제 분석 보고서가 나타납니다 .

주제 분석은 회전 PCA(주성분 분석)와 동등합니다. Varimax 회전은 여러 개의 특이 벡터를 구하고 이를 회전시켜 더 직접적으로 좌표 방향(용어 방향)을 향하게 합니다. 이렇게 회전하면 각 회전 벡터가 용어 집합을 향하므로 텍스트를 설명하는 데 도움이 됩니다. 음수 값은 반발력을 나타냅니다. 음수 값을 갖는 용어는 양수 값을 갖는 용어에 비해 주제에 나타나는 빈도가 낮습니다.

JMP PRO 주제 분석 보고서

주제 분석 보고서에서는 회전 후 각 주제에서 적재량이 가장 큰 용어를 보여 줍니다. 추가 보고서에서는 회전된 특이값 분해의 성분을 보여 줍니다.

"주제별 상위 적재" 보고서에서는 각 주제의 용어 테이블을 보여 줍니다. 각 테이블의 용어는 각 주제에 대한 적재량 절대값이 가장 큰 용어입니다. 각 테이블은 적재량 절대값을 기준으로 내림차순 정렬되어 있습니다. 이러한 테이블을 사용하여 각 주제에 해당하는 개념 테마를 판별할 수 있습니다.

주제 분석 보고서에는 다음 보고서도 포함됩니다.

주제 적재 각 용어의 주제 간 적재량 행렬이 포함됩니다. 이 행렬은 회전 PCA 의 요인 적재 행렬과 동등합니다.

주제별 단어 클라우드 각 주제당 하나씩의 단어 클라우드가 포함됩니다.

주제 점수 각 주제의 문서 점수 행렬이 포함됩니다. 주제 점수가 높은 문서는 해당 주제와 연관성이 있을 가능성이 높습니다.

주제 점수 그림 " 텍스트 표시 " 버튼과 각 문서의 주제 점수 그림이 포함됩니다. " 텍스트 표시 " 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다.

" 주제 점수 그림 " 보고서는 주제 점수 보고서의 행렬을 시각적으로 나타낸 것입니다. 그림의 각 패널은 주제 중 하나, 또는 주제 점수 행렬의 열 중 하나에 해당합니다. 각 패널 내에서 각 점은 말뭉치의 문서 중 하나, 또는 주제 점수 행렬의 행 중 하나에 해당합니다.

각 주제에 의해 설명된 분산 각 주제에 의해 설명되는 분산의 테이블이 포함됩니다. 이 테이블에는 각 주제에 의해 설명되는 변동이 백분율 또는 누적 백분율에 대한 열도 포함됩니다.

회전 행렬 Varimax 회전에 대한 회전 행렬이 포함됩니다.

JMP PRO 주제 분석 보고서 옵션

텍스트 탐색기 플랫폼의 주제 분석 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

주제 산점도 행렬 회전된 특이값 분해 벡터에 대한 산점도 행렬을 표시하거나 숨깁니다. " 텍스트 표시 " 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다.

표시 옵션 주제 분석 보고서에 나타나는 내용을 표시하거나 숨기기 위한 옵션이 포함되어 있습니다. 자세한 내용은 " 주제 분석 보고서 "(393 페이지) 에서 확인하십시오.

주제 이름 바꾸기 하나 이상의 주제에 대한 이름을 추가할 수 있습니다.

문서 주제 벡터 저장 회전 특이값 분해에서 구한 특이 벡터 중 사용자가 지정한 수만큼을 데이터 테이블에 열로 저장합니다.

주제 벡터 계산식 저장 회전 특이값 분해를 포함하는 벡터 모델링 유형의 계산식 열을 데이터 테이블에 저장합니다. 결과 열에는 `Text Score()` JSL 함수가 사용됩니다. 이 함수에 대한 자세한 내용은 " 도움말 " > " 스크립트 인덱스 " 에서 확인하십시오.

용어 주제 벡터 저장 주제 벡터를 " 용어 테이블 저장 " 옵션으로 생성된 데이터 테이블에 열로 저장합니다 .

제거 SVD 보고서에서 주제 분석 보고서를 제거합니다 .

JMP PRO 판별 분석

판별 분석에서는 DTM(문서 용어 행렬) 의 열을 기준으로 그룹 또는 범주에서 각 문서의 소속을 예측합니다. 특히 판별 분석에서는 각 문서가 반응 열 범주로 분류되는지를 예측합니다. 판별 분석 옵션을 선택할 때는 범주 또는 그룹이 포함된 반응 열을 선택해야 합니다 . 소속 그룹은 DTM 의 열을 통해 예측됩니다 . 판별 분석에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

텍스트 탐색기 플랫폼의 판별 분석 방법은 중심화된 DTM 의 특이값 분해를 기반으로 합니다. 반응 열의 각 그룹에는 DTM 을 중심화하는 데 사용되는 고유한 그룹 평균이 있습니다. 텍스트 탐색기 플랫폼의 판별 분석 방법은 DTM 의 희소성을 활용하기 때문에 판별 분석 플랫폼보다 빠릅니다 .

JMP PRO 판별 분석 규격 창

텍스트 탐색기의 판별 분석 옵션은 DTM(문서 용어 행렬)을 기반으로 합니다. DTM 은 용어 목록에 있는 각 용어마다 열을 하나씩 생성함으로써 구성됩니다(지정된 최대 용어 수까지). 각 텍스트 문서 (데이터 테이블의 행과 동등)는 DTM 의 한 행에 해당합니다. DTM 의 셀 값은 " 규격 " 창에서 사용자가 지정한 가중 유형에 따라 달라집니다 .

텍스트 탐색기의 빨간색 삼각형 메뉴에서 " 판별 분석 " 을 선택하면 다음 옵션이 포함된 " 규격 " 창이 나타납니다 .

최대 용어 수 판별 분석에 포함되는 용어의 최대 개수입니다 .

최소 용어 빈도 한 용어가 판별 분석에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다 .

가중치 문서 용어 행렬의 셀에 들어갈 값을 결정하는 가중치 체계입니다 . 가중치 체계 옵션에 대한 자세한 내용은 " 문서 용어 행렬 규격 창 "(386 페이지) 에서 확인하십시오 .

특이 벡터 수 판별 분석의 특이 벡터 수입니다 . 기본값은 문서 수 , 용어 수 또는 100 중 최소값입니다 .

JMP PRO 판별 분석 보고서

텍스트 탐색기 플랫폼의 판별 분석 보고서에는 기본적으로 두 개의 열린 보고서, 즉 분류 요약 보고서와 판별 점수 보고서가 포함됩니다. 다른 보고서는 처음에는 닫혀 있습니다.

판별 분석 보고서에는 다음 보고서가 포함됩니다 .

용어 평균 판별 분석에 사용된 용어의 테이블을 제공합니다 . 이러한 용어는 DTM 의 열에 해당합니다 . 이 테이블에는 각 용어에 대한 각 그룹 내의 평균과 각 용어의 전체 평균 및 가중 표준편차가 포함됩니다 .

각 그룹에 대한 제곱 거리 각 문서에 대해 각 그룹까지의 Mahalanobis 거리를 제곱한 결과를 포함하는 테이블을 제공합니다 . Mahalanobis 거리에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

각 그룹에 대한 확률 문서가 각 그룹에 속할 확률이 포함된 테이블을 제공합니다 .

분류 요약 판별 점수를 요약한 보고서를 제공합니다 . 이 보고서는 판별 분석 플랫폼 보고서의 " 점수 요약 " 보고서에 해당합니다 .

판별 점수 각 문서 및 기타 지원 정보의 예측 분류 테이블을 제공합니다 . 이 테이블은 판별 분석 플랫폼 보고서의 " 판별 점수 " 테이블에 해당합니다 .

JMP PRO 판별 분석 보고서 옵션

" 판별 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

정준 그림 정준 공간의 문서 및 그룹 평균 그림을 숨기거나 표시합니다 . 정준 공간은 그룹을 가장 많이 나누는 공간입니다 . 반응 변수의 수준이 세 개 이상인 경우에는 정준 좌표의 수를 지정해야 합니다 . 정준 좌표를 세 개 이상 지정할 경우 이 옵션은 정준 그림 행렬을 생성합니다 .

확률 저장 각 반응 수준에 대한 확률 열 및 가장 확률이 높은 반응이 포함된 열을 데이터 테이블에 저장합니다 . 가장 확률이 높은 반응 열에는 모형에 따른 확률이 가장 높은 수준이 포함됩니다 .

각 확률 열에서는 관측값이 해당 반응 수준에 속할 사후 확률을 제공합니다 . " 반응 확률 " 열 특성은 각 확률 열에 저장됩니다 . 반응 확률 열 특성에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

확률 계산식 저장 가장 확률이 높은 반응을 예측하기 위한 계산식 열을 데이터 테이블에 저장합니다 . 첫번째로 저장되는 열에는 `Text Score()` 함수를 사용하여 각 반응 수준의 확률을 계산하는 계산식이 포함됩니다 . 각 반응 수준에 대한 확률을 포함하는 열뿐만 아니라 예측 반응을 포함하는 열도 있습니다 .

정준 점수 저장 각 관측값에 대해 정준 공간의 점수가 포함된 열을 데이터 테이블에 저장합니다 정준 공간은 그룹을 가장 많이 나누는 공간입니다 . *k* 번째 정준 점수의 열에는 정준 <k> 라는 이름이 지정됩니다 .

제거 텍스트 탐색기 보고서 창에서 판별 분석 보고서를 제거합니다 .

JMP PRO 용어 선택

용어 선택은 서로 다른 반응을 가장 잘 설명하는 용어를 식별합니다. 이 분석은 일반화 회귀 플랫폼을 사용하여 DTM(문서 용어 행렬)에서 변수를 선택하고 반응에 가장 큰 영향을 미치는 용어를 식별합니다. 용어 선택은 감정 분석 및 다른 유형의 응답과 유사한 이항 반응과 함께 사용할 수 있습니다. 적합 모형은 지정된 반응 열에 대해 적절한 반응 분포를 사용합니다.

팁 : 용어 선택의 예를 보려면 **도움말 > 샘플 데이터 라이브러리를 선택하고** Chips.jmp 를 연 다음 텍스트 탐색기 - 용어 선택 테이블 스크립트를 실행하십시오 .

JMP PRO 용어 선택 설정

설정 보고서를 사용하면 반응 열을 선택하고, 반응의 목표 수준을 지정하고, 모형의 설정을 조정할 수 있습니다. 모형 설정을 지정했으면 "실행" 버튼을 클릭하여 모형을 실행합니다. 그러면 요약 보고서에 적합 모형이 나타납니다. 자세한 내용은 "용어 선택 요약 보고서"(398 페이지)에서 확인하십시오.

목표 수준

반응 열을 선택하면 목표 수준 개요가 나타납니다.

- 명목형 반응의 경우 로지스틱 회귀 모형에서 목표 수준이 될 반응의 한 수준을 선택합니다 . 로지스틱 회귀 모형의 반응은 목표 수준과 다른 모든 수준 간의 결합입니다 .
- 순서형 반응의 경우 처음에는 모든 반응 수준이 모형에 포함됩니다 . 로컬 데이터 필터를 사용하면 모형에서 제외할 반응 수준을 선택할 수 있습니다 . 포함된 수준의 기본 숫자 값은 정규 반응 분포로 모델링됩니다 .

  참고 : 순서형 반응의 경우 용어 선택 모형은 반응 열의 데이터 유형이 숫자인 경우에만 적합시킬 수 있습니다 .

- 연속형 반응의 경우 로컬 데이터 필터 히스토그램을 사용하여 모형에서 제외할 반응 값을 선택합니다 . 포함된 값은 정규 반응 분포로 모델링됩니다 .
- 다중 반응 모델링 유형이 있는 반응 열의 경우 이항 로지스틱 회귀 모형에서 목표 수준이 될 반응 수준을 하나 이상 선택합니다 . 두 개 이상의 수준을 선택하는 경우 문서의 반응 열에 선택한 수준이 하나라도 있으면 해당 문서는 목표 수준에 속합니다 . 선택한 모든 수준이 문서의 반응 열에 있어야 해당 문서가 목표 수준에 포함되게 하려면 **AND 와 결합** 옵션을 선택합니다 .

모형 설정

기본적으로 일반화 회귀 모델은 조기 중지가 있는 Elastic Net 추정 방법과 AICc 검증 방법을 사용합니다. 모형 설정 개요에서 이러한 설정을 변경할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

참고 : 텍스트 탐색기 시작 창에 검증 열이 지정된 경우 용어 선택 보고서의 일반화 회귀 플랫폼은 이 검증 열을 검증 방법으로 사용합니다 .

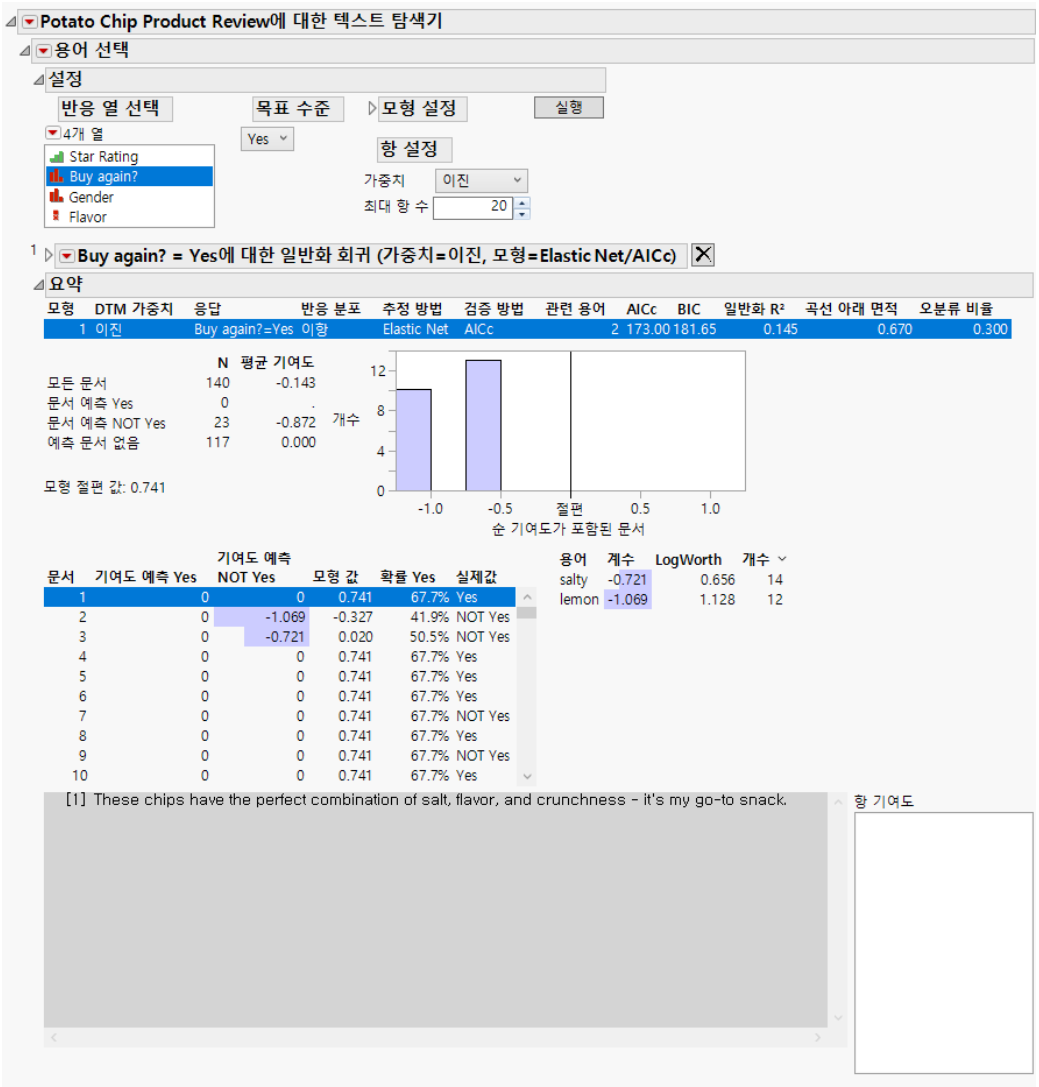
항 설정

항 설정은 회귀 모형에 사용되는 DTM(문서 용어 행렬)을 정의합니다. 가중치 기법과 DTM에 포함된 최대 항 수를 변경할 수 있습니다. 각 항은 DTM의 열에 해당합니다. 말뭉치에 10개 미만의 항목이 있는 항은 모형에서 사용하는 DTM에 포함되지 않습니다. DTM 옵션에 대한 자세한 내용은 "문서 용어 행렬 규격 창"(386 페이지)에서 확인하십시오.

JMP PRO 용어 선택 보고서

분석을 실행한 후 용어 선택 보고서는 세 섹션으로 구성됩니다. 설정 보고서에는 분석을 지정하기 위한 컨트롤이 포함되어 있습니다. 자세한 내용은 "용어 선택 설정"(396 페이지)에서 확인하십시오. 설정 보고서 아래에는 실행한 각 분석에 대해 초기에 닫힌 일반화 회귀 보고서가 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오. 보고서의 마지막 섹션은 요약 보고서입니다.

그림 12.12 용어 선택 보고서



용어 선택 요약 보고서

요약 보고서에는 모형 비교 테이블, 요약 테이블 및 히스토그램, 문서 점수 테이블, 용어 점수 테이블 및 텍스트 상자가 포함됩니다.

모형 비교 테이블에는 각 적합 모형에 대한 행이 포함됩니다. 요약 보고서의 나머지 부분에서는 이 테이블에서 현재 선택된 모형의 결과를 보여 줍니다.

요약 테이블에는 문서의 개수와 평균 점수가 전체 및 모형의 반응 예측값별로 표시됩니다. 평균 기여도는 문서 점수 테이블의 기여도 값 평균입니다. 요약 히스토그램에는 문서의 전체 기여도

값 분포가 표시됩니다. 히스토그램은 대화식이므로 막대를 클릭하면 문서 점수 테이블에서 해당하는 문서가 강조 표시됩니다.

문서 점수 테이블에는 각 문서의 양 및 음의 기여도 값과 각 문서의 예측값 및 실제값이 표시됩니다. 이항 반응 모델의 경우 예측값은 문서가 목표 수준에 속할 확률이고 정규 반응 모델의 경우 예측값은 각 문서에 대한 적합 모형의 예측입니다. 테이블의 행을 선택하면 해당 문서의 텍스트가 테이블 아래의 텍스트 상자에 나타납니다.

용어 점수 테이블에는 적합 모형에서 선택된 각 용어, 용어의 해당 계수, 용어의 LogWorth 및 말뭉치에서 용어가 나타나는 횟수가 나열됩니다. 테이블의 행을 선택하면 해당 문서의 텍스트가 테이블 아래의 텍스트 상자에 나타납니다.

텍스트 상자에는 문서 점수 테이블에서 선택된 문서의 텍스트 또는 용어 점수 테이블에서 선택된 용어의 컨텍스트가 표시됩니다.

JMP PRO 용어 선택 보고서 옵션

용어 선택 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

문서 점수 저장 (요약 테이블에서 분석을 선택한 경우에만 사용할 수 있습니다.) 문서 점수 테이블의 열을 데이터 테이블의 새 열에 저장합니다. 새 열에는 각 문서에 대한 예측값과 함께 양 및 음의 기여도가 포함됩니다.

용어 점수 DTM 저장 (요약 테이블에서 분석을 선택한 경우에만 사용할 수 있습니다.) 현재 선택한 분석의 각 관련 용어에 대한 데이터 테이블에 열을 저장합니다. 열에는 "용어 선택"의 "항 설정"에 지정된 가중치를 사용하여 각 문서에 대한 용어 점수가 포함됩니다.

예측 계산식 저장 (요약 테이블에서 분석을 선택한 경우에만 사용할 수 있습니다.) 현재 선택한 분석에 대한 예측 계산식이 포함된 데이터 테이블에 열을 저장합니다.

용어 클라우드 표시 요약 보고서에서 단어 클라우드를 표시하거나 숨깁니다. 단어 클라우드에는 현재 선택된 분석의 계수 항이 표시됩니다. 단어는 해당 계수의 절대값에 따라 크기가 지정되고 계수의 부호에 따라 색상이 지정됩니다.

제거 텍스트 탐색기 보고서 창에서 용어 선택 보고서를 제거합니다.

JMP PRO 감정 분석

감정 분석에서는 어휘 분석을 사용하여 문서에서 감정 용어를 식별하고 긍정, 부정 및 전반적 감정에 대해 문서의 점수를 매깁니다. 이 분석에서는 각 문서가 단일 주제에 대한 이항 감정이 있는 프리 텍스트라고 가정합니다. 감정 분석에서는 기본적인 NLP(자연어 처리)를 결과에 통합합니다. 자연어 처리에 대한 자세한 내용은 https://opennlp.apache.org/에서 확인하십시오. NLP를 사용하지 않으려면 "문서 파싱" 옵션을 선택 취소하십시오.

팁 : 감정 분석의 예를 보려면 **도움말 > 샘플 데이터 라이브러리를 선택하고** Chips.jmp 를 연 다음 텍스트 탐색기 - 감정 분석 테이블 스크립트를 실행하십시오 .

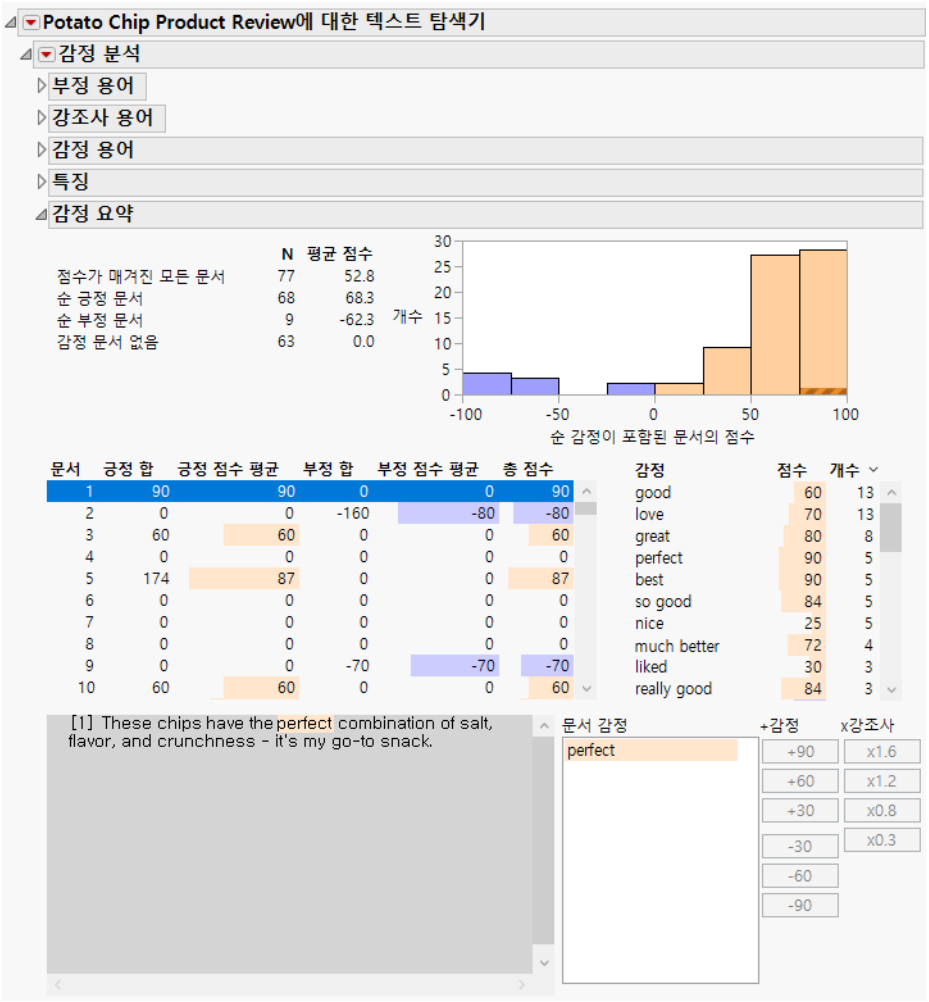
참고 :

- 감정 분석에서 단어는 부정 , 강조사 또는 감정 중 한 용어 클래스에만 적용될 수 있습니다 .
- 감정 분석은 일부 이모티콘 또는 단일 단위로 취급되는 일련의 문자를 인식합니다 . " 감정 용어 " 보고서 또는 " 감정 용어 관리 " 창에서 기본 제공 이모티콘과 해당 기본 감정 점수를 확인할 수 있습니다 .
- 단어를 부정 , 강조사 또는 감정 용어로 지정할 때 해당 단어가 이미 중지 단어로 지정된 경우 감정 분석 보고서가 열려 있는 동안 해당 단어는 일시적으로 중지 단어에서 제거됩니다 . 이 임시 제거는 전체 텍스트 탐색기 보고서에 영향을 미칩니다 . 감정 분석 보고서가 닫히면 중지 단어로 복원됩니다 .

JMP PRO 감정 분석 보고서

기본적으로 텍스트 탐색기 플랫폼의 감정 분석 보고서에는 " 감정 요약 " 이라는 보고서 하나가 열린 상태로 포함되어 있습니다 . 다른 보고서는 처음에는 닫혀 있습니다 .

그림 12.13 감정 분석 보고서



감정 분석 보고서에는 다음 보고서가 포함됩니다 .

부정 용어

현재 감정 분석의 부정 용어 목록을 포함합니다. 추가 옵션 메뉴를 보려면 목록을 마우스 오른쪽 버튼으로 클릭하십시오. 목록에서 용어를 선택하여 제거할 수 있습니다.

강조사 용어

강조사 용어와 해당하는 승수 값의 목록을 포함합니다. 추가 옵션 메뉴를 보려면 목록을 마우스 오른쪽 버튼으로 클릭하십시오. 목록에서 용어를 선택하여 제거할 수 있습니다.

감정 용어

감정 용어와 해당하는 점수 값의 목록을 포함합니다. 이 보고서를 사용하여 새 감정 용어를 추가할 수도 있습니다. 가능한 감정 테이블에는 감정 용어로 추가할 수 있는 용어의 개수가 포함됩니다. 용어를 감정 용어로 추가하려면 "가능한 감정" 테이블에서 해당 용어를 선택하고 "+감정" 아래의 버튼 중 하나를 클릭하십시오. 나열되지 않은 감정 점수 값을 선택하려면 감정 용어 목록에 점수 값을 추가한 후 편집할 수 있습니다.

"가능한 감정" 테이블에서 용어를 선택하면 해당 용어가 포함된 문서가 감정 용어 보고서의 오른쪽에 나타납니다. 이것은 말뭉치에서 용어가 사용되는 방식에 대한 컨텍스트를 제공합니다.

특징

말뭉치의 특징에 점수를 매기는 옵션이 포함되어 있습니다. 특징은 감정 용어로 설명되는 것입니다. 가능한 특징 용어의 목록을 생성하려면 **검색** 버튼을 클릭하십시오. "가능한 특징" 테이블에서 하나 이상의 용어를 선택하면 해당 용어가 포함된 문서의 발췌문이 테이블 오른쪽의 텍스트 상자에 나타납니다. **선택한 특징 점수 매기기** 버튼을 클릭하여 선택한 특징 용어의 점수 매기기 결과를 표시하도록 감정 요약 보고서를 업데이트하십시오.

참고 : " 문서 파싱 " 옵션을 선택하면 단어가 한 감정과 동일한 우산 절 내에서 발생할 때 특징 보고서가 점수를 매깁니다 .

감정 요약

현재 설정을 기반으로 한 감정 분석의 결과를 포함합니다. 이 보고서에는 요약 테이블 및 히스토그램, 문서 점수 테이블, 감정 용어 테이블, 텍스트 상자 및 더 많은 감정 용어와 강조사 용어를 추가할 수 있는 제어판이 포함되어 있습니다.

요약 테이블에는 문서의 점수를 매긴 방법에 따라 세분화된 문서의 개수와 평균 점수가 표시됩니다. 평균 점수는 점수 매기기 옵션의 설정에 따라 결정됩니다. 자세한 내용은 "감정 분석 보고서 옵션"(403 페이지)에서 확인하십시오. 요약 히스토그램에는 문서의 전체 감정 분포가 표시됩니다. 히스토그램은 대화식이므로 막대를 클릭하면 문서 점수 테이블에서 해당하는 문서가 강조 표시됩니다.

문서 점수 테이블에는 각 문서에 대한 전체 감정 점수뿐만 아니라 긍정 및 부정 감정 점수 합계와 평균이 표시됩니다. 테이블의 행을 선택하면 해당 문서의 텍스트가 테이블 아래의 텍스트 상자에 나타납니다. "열 점수 매기기"를 지정하면 테이블에 점수 매기기 열의 값이 포함됩니다.

팁 : 문서 점수 테이블의 셀을 마우스로 가리켜 테이블 결과를 생성하는 데 사용된 점수 매기기 계산을 확인할 수 있습니다 .

감정 용어 테이블에는 각 감정 용어, 해당 점수 값 및 말뭉치에서 해당 용어가 발생하는 횟수가 나열됩니다.

팁 : 여러 단어가 포함된 감정 용어의 경우 점수 열의 셀을 마우스로 가리키면 점수를 생성하는 데 사용된 계산을 확인할 수 있습니다.

텍스트 상자에는 문서 점수 테이블에서 선택된 문서의 텍스트 또는 감정 용어 테이블에서 선택된 용어의 컨텍스트가 표시됩니다. 문서 점수 테이블에서 문서를 선택하면 해당 문서의 감정 목록이 텍스트 상자 오른쪽에 나타납니다.

팁 : 부정, 강조사 또는 감정 용어로 분류된 텍스트 상자의 용어를 마우스로 가리키면 분류를 표시하고 제거 버튼이 포함된 상자가 나타납니다. 부정, 강조사 또는 감정 용어 목록에서 해당 용어를 빠르게 제거하려면 **제거** 버튼을 클릭하십시오.

제어판은 텍스트 상자에서 용어를 선택하면 활성화됩니다. 용어를 감정 용어로 추가하려면 텍스트 상자에서 해당 용어를 선택하고 "+감정" 아래의 버튼 중 하나를 클릭하십시오. 용어를 강조사 용어로 추가하려면 텍스트 상자에서 해당 용어를 선택하고 "×강조사" 아래의 버튼 중 하나를 클릭하십시오.

JMP PRO 감정 분석 보고서 옵션

" 감정 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

점수 매기기 문서의 총 점수를 계산하기 위한 다음 옵션이 포함되어 있습니다.

척도화 긍정적 구와 부정적 구의 점수가 합산됩니다. 그런 다음 합계를 문서의 구 수로 나누어 총 점수를 결정합니다.

최소값 최대값 총 점수는 긍정 점수 최대값과 부정 점수 최소값의 합으로 계산됩니다.

열 점수 매기기 계산된 감정과 비교할 수 있는 알려진 정보가 포함된 데이터 테이블 열을 지정합니다. 점수 열이 문서 점수 테이블에 추가됩니다.

팁 : 총 점수 열과 점수 열을 시각적으로 비교하여 감정 점수 매기기를 평가할 수 있습니다.

문서 파싱 문서 구문 분석에 NLP(자연어 처리) 가 사용되는지 여부를 지정합니다. 자연어 처리에 대한 자세한 내용은 https://opennlp.apache.org/ 에서 확인하십시오.

문서 점수 저장 문서 점수 테이블의 열을 데이터 테이블의 새 열에 저장합니다. 새 열에는 각 문서에 대한 전체 감정 점수뿐만 아니라 양 및 음의 감정 점수 합계와 평균이 포함됩니다.

문서별 감정 점수 계산 저장 각 감정 용어에 대한 열을 데이터 테이블에 저장합니다. 각 열에는 각 문서에서 각 감정 용어가 발생한 횟수가 포함됩니다.

부정 용어 표시 부정 용어 보고서를 표시하거나 숨깁니다.

강조사 용어 표시 강조사 용어 보고서를 표시하거나 숨깁니다.

감정 용어 표시 감정 용어 보고서를 표시하거나 숨깁니다 .

특징 찾기 표시 특징 보고서를 표시하거나 숨깁니다 .

감정 클라우드 표시 감정 요약 보고서에서 감정 용어의 단어 클라우드를 표시하거나 숨깁니다 .

기본 제공 부정 용어 포함 감정 분석에 사용된 부정 용어에 기본 제공 부정 용어가 포함되는지 여부를 지정합니다 .

기본 제공 강조사 용어 포함 감정 분석에 사용된 강조사 용어에 기본 제공 강조사 용어가 포함되는지 여부를 지정합니다 .

기본 제공 감정 용어 포함 감정 분석에 사용된 감정 용어에 기본 제공 감정 용어가 포함되는지 여부를 지정합니다 .

부정 용어 관리 부정 용어를 추가하거나 제거할 수 있는 창을 표시합니다 . 변경 사항은 사용자 , 열 및 지역 수준에서 적용될 수 있습니다 . 다른 수준에 지정된 부정 용어를 제외하는 지역 예외를 지정할 수도 있습니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

강조사 용어 관리 강조사 용어를 추가하거나 제거할 수 있는 창을 표시합니다 . 변경 사항은 사용자 , 열 및 지역 수준에서 적용될 수 있습니다 . 다른 수준에 지정된 강조사 용어를 제외하는 지역 예외를 지정할 수도 있습니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

감정 용어 관리 감정 용어를 추가하거나 제거할 수 있는 창을 표시합니다 . 변경 사항은 사용자 , 열 및 지역 수준에서 적용될 수 있습니다 . 다른 수준에 지정된 감정 용어를 제외하는 지역 예외를 지정할 수도 있습니다 . 자세한 내용은 " 용어 옵션 관리 창 "(381 페이지) 에서 확인하십시오 .

JMP PRO 텍스트 탐색기 플랫폼의 추가 예

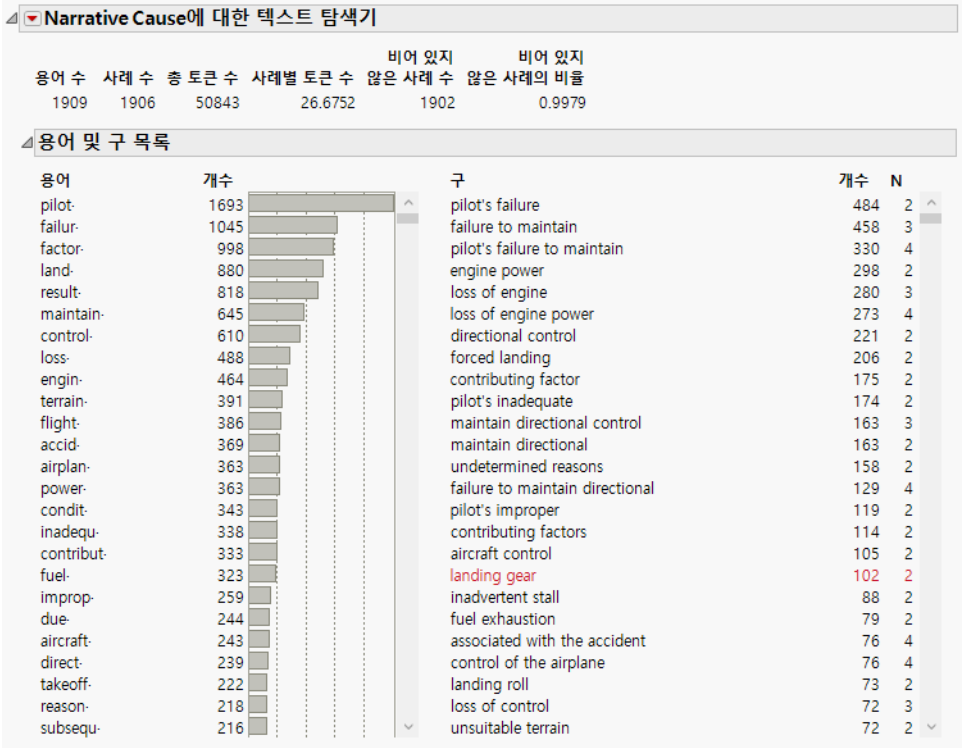
이 예에서는 미국에서 2001 년에 발생한 사고에 대한 미국 연방 교통 안전 위원회의 비행기 사고 보고서를 살펴봅니다 . 조사 결과에 대한 설명이 포함된 텍스트를 탐색해서 각 사고의 원인을 밝히려고 합니다 . 또한 사고 보고서 모음에서 테마를 찾으려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Aircraft Incidents.jmp 를 엽니다 .
2. **행 > 열 값에 따른 색상 또는 표식**을 선택합니다 .
3. 열 목록에서 Fatal 을 선택하고 **확인**을 클릭합니다 .

   심각한 피해를 가져온 사고가 포함되어 있는 행이 빨간색으로 표시됩니다 .
4. **분석 > 텍스트 탐색기**를 선택합니다 .
5. " 열 선택 " 목록에서 Narrative Cause 를 선택하고 **텍스트 열**을 클릭합니다 .
6. " 언어 " 목록에서 **영어**를 선택합니다 .
7. " 어간 추출 (Stemming)" 목록에서 **모든 용어의 어간 추출**을 선택합니다 .
8. " 토큰화 " 목록에서 **기본 단어**를 선택합니다 .

9. **확인**을 클릭합니다 .

그림 12.14 Narrative Cause 에 대한 텍스트 탐색기



보고서에서 약 51,000 개의 토큰과 약 1,900 개의 고유 용어가 있음을 알 수 있습니다 .

10. 용어 목록에서 "pilot" 을 마우스 오른쪽 버튼으로 클릭하고 **행 선택하기**를 선택합니다 .

    데이터 테이블에서 선택된 행의 개수를 통해 1,300 건 이상의 사고 보고서에서 "pilot" 이라는 단어의 몇 가지 형태가 나타남을 알 수 있습니다 .

11. "pilot" 을 마우스 오른쪽 버튼으로 클릭하고 **중지 단어 추가**를 선택합니다 .

    단어 "pilot" 의 일부 형태가 다른 용어들에 비해 빈도 높게 나타나므로 이러한 용어는 문서 간의 차이를 확인하는 데 충분한 정보를 제공하지 못합니다 . 따라서 어간이 "plot" 인 모든 용어를 중지 단어 목록에 추가합니다 .

JMP PRO 이 예의 나머지 단계는 JMP Pro 에서만 수행할 수 있습니다 .

12. JMP PRO "Narrative Cause 에 대한 텍스트 탐색기 " 옆의 빨간색 삼각형을 클릭하고 **잠재 의미 분석 , SVD** 를 선택합니다 .

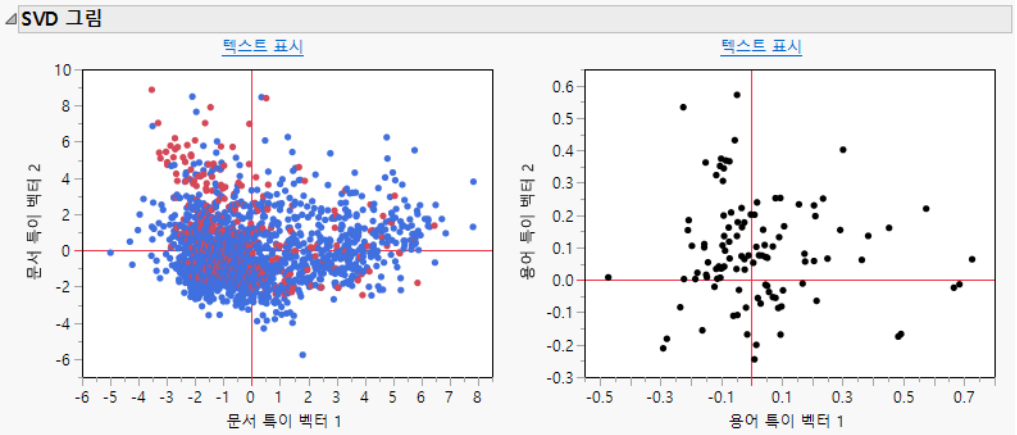
    이 분석은 주제 분석을 위한 첫 번째 분석 단계로서 , SVD 회전을 수행합니다 .

13. JMP PRO " 규격 " 창에서 " 최소 용어 빈도 " 에 50 을 입력합니다 .

    약 51,000 개의 토큰이 있으므로 이 빈도는 모든 용어의 0.1% 이상을 나타내는 용어와 동등합니다 .

14. JMP PRO **확인**을 클릭합니다 .

그림 12.15 Narrative Cause 에 대한 SVD 그림



문서 SVD 그림에서는 fatal한 사건과 non-fatal한 사건 사이에 큰 차이가 나타나지 않습니다.

15. JMP PRO "SVD 중심화 및 척도화 TF IDF" 옆의 빨간색 삼각형을 클릭하고 **주제 분석 , 회전된 SVD** 를 선택합니다 .

주제를 이루는 용어 그룹을 살펴보려고 합니다 .

16. JMP PRO " 주제 수 " 에 5 를 입력합니다 .
17. JMP PRO **확인**을 클릭합니다 .

그림 12.16 Narrative Cause 에 대한 주제별 상위 적재

주제별 상위 적재

주제 1		주제 2		주제 3		주제 4		주제 5	
용어	적재	용어	적재	용어	적재	용어	적재	용어	적재
power·	0.67567	altitud·	0.48052	factor·	0.5093	control·	0.5125	fuel·	0.4864
loss·	0.66539	low·	0.45137	condit·	0.4677	direct·	0.4957	personnel·	0.4630
forc·	0.62046	dark·	0.42289	unsuit·	0.3984	experi·	0.4382	mainten·	0.4546
engin·	0.61866	night·	0.40408	accid·	0.3909	student·	0.4273	result·	0.4153
suitabl·	0.58926	maintain·	0.39283	select·	0.3842	lack·	0.3625	preflight·	0.3930
lack·	0.53292	instrument·	0.39239	associ·	0.3806	maintain·	0.3616	exhaust·	0.3663
reason·	0.47828	clearanc·	0.37881	area·	0.3628	instructor·	0.3559	inspect·	0.3640
undetermin·	0.46599	airspe·	0.33612	compens·	0.3352	supervis·	0.3282	plan·	0.3474
terrain·	0.37013	condit·	0.33473	failur·	-0.3207	power·	-0.3269	reason·	-0.3468
total·	0.31932	stall·	0.33010	wind·	0.3202	failur·	0.3171	undetermin·	-0.3438
land·	0.29402	flight·	0.32838	inadequ·	0.3080	reason·	-0.3142	improp·	0.3389
		continu·	0.31729	result·	-0.2845	undetermin·	-0.3101	inadequ·	0.3353
		maneuv·	0.30931	terrain·	0.2663	crosswind·	0.3085	subsequ·	0.3246
		weather·	0.29484	airspe·	-0.2542	aircraft·	0.3007	maintain·	-0.3117
		adequ·	0.28082			factor·	0.2787	due·	0.2882

적재량이 가장 높은 각 주제에 대한 용어를 사용하면 해당 주제가 사고 보고서의 테마를 포착하는지 여부를 해석할 수 있습니다 .

예를 들어 주제 1 은 "power", "loss" 및 "engine" 에 대한 적재량이 높으므로 사고 원인으로 엔진 동력 손실이라는 테마를 나타냅니다 . 이는 전체 사고 보고서에서 273 회 나타나는 " 엔진 동력 손실 " 이라는 구와 일치합니다 .

주제 2 에서 적재량이 높은 단어를 살펴보면 , 사고가 어둠이나 낮은 고도와 관련이 있다고 설명할 수 있습니다 .

이 텍스트 분석 단계에서는 진행 방법을 다양하게 선택할 수 있습니다. 텍스트 분석은 반복적인 프로세스이므로 주제 정보를 사용하여 중지 단어를 추가하거나 구를 지정하는 방법으로 용어 목록을 세부적으로 큐레이팅할 수 있습니다. 가중치가 부여된 문서 용어 행렬, SVD 또는 회전된 SVD 의 벡터를 데이터 테이블에 숫자 열로 저장하여 다른 JMP 분석 플랫폼에서 사용할 수도 있습니다. 다른 플랫폼에서 이러한 열을 사용하는 경우 추가 분석에 데이터 테이블의 다른 열을 포함할 수도 있습니다.

A 부록

참조 자료

Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.

Agresti, A., and Coull, B. A. (1998). "Approximate Is Better than 'Exact' for Interval Estimation of Binomial Proportions." *American Statistician* 52:119–126.

Asiribo, O., and Gurland, J. (1990). "Coping with Variance Heterogeneity." *Communication in Statistics: Theory and Methods* 19:4029–4048.

Bartlett, M. S., and Kendall, D. G. (1946). "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation." *Supplement to the Journal of the Royal Statistical Society* 8:128–138.

Bissell, A. F. (1990). "How Reliable is Your Capability Index?" *Applied Statistics* 30:331–340.

Brown, M. B., and Benedetti, J. K. (1977). "Sampling Behavior of Tests for Correlation in Two-Way Contingency Tables." *Journal of the American Statistical Association* 72:305–315.

Brown, M. B., and Forsythe, A. B. (1974). "Robust Tests for Equality of Variances." *Journal of the American Statistical Association* 69:364–367.

Chen, S.-X., and Hall, P. (1993). "Empirical Likelihood Confidence Intervals for Quantiles." *The Annals of Statistics* 21:1166–1181.

Chou, Y.-M., Owen, D. B., and Borrego, S. A. (1990). "Lower Confidence Limits on Process Capability Indices." *Journal of Quality Technology* 22:223–229.

Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74:829–836.

Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Education Psychological Measurement* 20:37–46.

Conover, W. J. (1972). "A Kolmogorov Goodness-of-fit Test for Discontinuous Distributions." *Journal of the American Statistical Association* 67:591–596.

Conover, W. J. (1980). *Practical Nonparametric Statistics*. New York: John Wiley & Sons.

Conover, W. J. (1999). *Practical Nonparametric Statistics*. 3rd ed. New York: John Wiley & Sons.

Cureton, E. E. (1967). "The Normal Approximation to the Signed-Rank Sampling Distribution when Zero Differences are Present." *Journal of the American Statistical Association* 62:1068–1069.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics* 44:837–845.

Devore, J. L. (1995). *Probability and Statistics for Engineering and the Sciences*. Pacific Grove, CA: Duxbury Press.

Dunn, O. J. (1964). "Multiple Comparisons Using Rank Sums." *Technometrics* 6:241–252.

Dunnett, C. W. (1955). "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50:1096–1121.

Efron, B. (1981). "Nonparametric Standard Errors and Confidence Intervals." *The Canadian Journal of Statistics* 9:139–158.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. 2nd ed. Boca Raton, Florida: CRC.

Fleiss, J. L., Cohen, J., and Everitt, B. S. (1969). "Large-Sample Standard Errors of Kappa and Weighted Kappa." *Psychological Bulletin* 72:323–327.

Friendly, M. (1994). "Mosaic Displays for Multi-Way Contingency Tables." *Journal of the American Statistical Association* 89:190–200.

Goodman, L. A., and Kruskal, W. H. (1979). *Measures of Association for Cross Classification*. New York: Springer-Verlag.

Gupta, S. S. (1965). "On Some Multiple Decision (Selection and Ranking) Rules. *Technometrics* 7:225–245.

Hajek, J. (1969). *A Course in Nonparametric Statistics*. San Francisco: Holden-Day.

Hartigan, J. A., and Kleiner, B. (1981). "Mosaics for Contingency Tables." In *Computer Science and Statistics: Proceedings of the Thirteenth Symposium on the Interface*, edited by W. F. Eddy, 268–273. New York: Springer-Verlag.

Hayter, A. J. (1984). "A Proof of the Conjecture That the Tukey-Kramer Method Is Conservative." *Annals of Mathematical Statistics* 12: 61–75.

Hosmer, D. W., and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley & Sons.

Howell, D. C. (2013). *Statistical Methods for Psychology.* 8th ed. Belmont, California: Wadsworth.

Hsu, J. (1981). "Simultaneous Confidence Intervals for All Distances from the 'Best'." *Annals of Statistics* 9:1026–1034.

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall.

Huber, P. J. (1973). "Robust Regression: Asymptotics, Conjecture, and Monte Carlo." *Annals of Statistics* 1:799–821.

Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics.* 2nd ed. New York: John Wiley & Sons.

Iman, R. L. (1974). "Use of a t-statistic as an Approximation to the Exact Distribution of Wilcoxon Signed Ranks Test Statistic." *Communications in Statistics—Simulation and Computation* 3:795–806.

Jones, M. C., and Pewsey, A. (2009). "Sinh-Arcsinh Distributions." *Biometrika* 96:761–780.

Kendall, M., and Stuart, A. (1979). *The Advanced Theory of Statistics*. 4th ed. Vol. 2. New York: Macmillan.

Keuls, M. (1952). "The Use of the 'Studentized Range' in Connection with an Analysis of Variance." *Euphytica* 1.2:112–122.

Kramer, C. Y. (1956). "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." *Biometrics* 12:307–310.

Lehmann, E. L., and D'Abrera, H. J. M. (2006). *Nonparametrics: Statistical Methods Based on Ranks*. Rev. ed. San Francisco: Holden-Day.

Levene, H. (1960). "Robust Tests for the Equality of Variance." In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann. Palo Alto, CA: Stanford University Press.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.

Meeker, W. Q., Hahn, G. J., and Escobar, L. A. (2017). *Statistical Intervals: A Guide for Practitioners and Researchers*. 2nd ed. New York: John Wiley & Sons.

Miller, A. J. (1972). "Letter to the Editor." *Technometrics* 38:307.

Nagelkerke, N. J. D. (1991). "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78:691–692.

Nelson, P. R., Wludyka, P. S., and Copeland, K. A. F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Philadelphia: Society for Industrial and Applied Mathematics.

Neter, J., Wasserman, W., and Kutner, M. H. (1996). *Applied Linear Statistical Models*. 4th ed. Boston: Irwin.

O'Brien, R. G. (1979). "A General ANOVA Method for Robust Tests of Additive Models for Variances." *Journal of the American Statistical Association* 74:877–880.

O'Brien, R., and Lohr, V. (1984). "Power Analysis For Linear Models: The Time Has Come." *Proceedings of the Ninth Annual SAS User's Group International Conference*, 840–846. Cary, NC: SAS Institute Inc.

Olejnik, S. F., and Algina, J. (1987). "Type I Error Rates and Power Estimates of Selected Parametric and Nonparametric Tests of Scale." *Journal of Educational Statistics* 12:45–61.

Pratt, J. W. (1959). "Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures." *Journal of the American Statistical Association* 54:655–667.

Reinsch, C. H. (1967). "Smoothing by Spline Functions." *Numerische Mathematik* 10:177–183.

Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.

Rubin, D. (1981). "The Bayesian Bootstrap." *The Annals of Statistics* 9:130–134.

SAS Institute Inc. (2020a). "Introduction to Nonparametric Analysis." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/intronpar.pdf.

SAS Institute Inc. (2020b). "The FREQ Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/freq.pdf.

Schuirmann, D. J. (1987). "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability." *Journal of Pharmacokinetics and Biopharmaceutics* 15:657–680.

Slifker, J. F., and Shapiro, S. S. (1980). "The Johnson System: Selection and Parameter Estimation." *Technometrics* 22:239–246.

Snedecor, G. W., and Cochran, W. G. (1980). *Statistical Methods*. 7th ed. Ames, Iowa: Iowa State University Press.

Somers, R. H. (1962). "A New Asymmetric Measure of Association for Ordinal Variables." *American Sociological Review* 27:799–811.

Stephens, M. A. (1974). "EDF Statistics for Goodness of Fit and Some Comparisons." *Journal of the American Statistical Association* 69:730–737.

Tan, C. Y., and Iglewicz, B. (1999). "Measurement-Methods Comparisons and Linear Statistical Relationship." *Technometrics* 41:192–201.

Tamhane, A. C., and Dunlop, D. D. (2000). *Statistics and Data Analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Tukey, J. W. (1953). "The Problem of Multiple Comparisons." In *Multiple Comparisons, 1948–1983*, edited by H. I. Braun, vol. 8 of *The Collected Works of John W. Tukey* (published 1994), 1–300. London: Chapman & Hall. Unpublished manuscript.

Welch, B. L. (1951). "On the Comparison of Several Mean Values: An Alternative Approach." *Biometrika* 38:330–336.

Wheeler, D. J. (2003). *Range Based Analysis of Means*. Knoxville, TN: SPC Press.

Wilson, E. B. (1927). "Probable Inference, the Law of Succession, and Statistical Inference." *Journal of the American Statistical Association* 22:209–212.

Wludyka, P. S., and Nelson, P. R. (1997). "An Analysis-of-Means-Type Test for Variances From Normal Populations." *Technometrics* 39:274–285.

B 부록

기술 라이선스 표시

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED

WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR

ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT

(INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

Version 16

Essential Graphing

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. *JMP® 16 Essential Graphing*. Cary, NC: SAS Institute Inc.

JMP® 16 Essential Graphing

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

August 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Get the Most from JMP

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

https://www.jmp.com/getstarted

Contents

Essential Graphing

1 Learn about JMP 11
Documentation and Additional Resources
Formatting Conventions in JMP Documentation 13
JMP Help 14
JMP Documentation Library 14
Additional Resources for Learning JMP 20
JMP Tutorials 20
Sample Data Tables 20
Learn about Statistical and JSL Terms 21
Learn JMP Tips and Tricks 21
JMP Tooltips 21
JMP User Community 22
Free Online Statistical Thinking Course 22
JMP New User Welcome Kit 22
Statistics Knowledge Portal 22
JMP Training 22
JMP Books by Users 23
The JMP Starter Window 23
JMP Technical Support 23

2 Introduction to Interactive Graphing 25
Overview of Data Visualization

3 Graph Builder 27
Interactive Data Visualization
How to Use Graph Builder 30
About the Graph Builder Window 30
Example of Features in Graph Builder 32
Move or Remove Variables in Zones 39
Work with Axes 40
Merge Variables on a Common Axis 40
Create Separate Axes in the X and Y Zones 41
Create a Second Y Axis 43

Create Nested Axes for Character Variables ... 45
Order the Levels of a Categorical Variable ... 49
Element Types and Options ... 53
Points ... 55
Smoother ... 57
Line of Fit ... 58
Ellipse ... 60
Contour ... 61
Line ... 64
Bar ... 67
Area ... 78
Box Plot ... 81
Histogram ... 85
Heatmap ... 87
Pie ... 89
Treemap ... 91
Mosaic ... 94
Caption Box ... 95
Formula ... 97
Map Shapes ... 98
Parallel ... 99
Graph Zones ... 103
Categorical and Continuous Variables in Group X or Y Zones ... 107
Move Group X and Group Y Labels ... 108
Red Triangle Options ... 108
Graph Builder Buttons ... 110
Right-Click Menus ... 111
Graph Options ... 111
Grouping Zones Options ... 112
Options for Columns in the Variables Panel ... 112
Options for Axes, Axes Variable Labels, or Graph Titles ... 113
Options for the Dividing Line between Multiple Graphs ... 114
Legend Options ... 115

4 Graph Builder Examples ... 121
Interactive Data Visualization

Example of an Ordered Bar Chart ... 123
Example of an Area and Line Chart ... 126
Example of a Map Chart ... 131
Example of a Butterfly Chart ... 134

Example of a Scatterplot ... 138
Example of a Packed Bar Chart ... 143
Example of an Overlaid Histogram and Ridgeline Chart ... 145
Example of a Bullet Bar Chart ... 147
Example of a Forecast Plot ... 149

5 Bubble Plots ... 151
View Patterns in Multidimensional Data Using Bubble Plots

Example of a Dynamic Bubble Plot ... 153
Launch the Bubble Plot Platform ... 155
Specifying Two ID Variables ... 156
Specifying a Time Variable ... 156
Interact with the Bubble Plot ... 157
Control Animation for Dynamic Bubble Plots ... 157
Select Bubbles ... 158
Use the Brush Tool ... 159
Bubble Plot Platform Options ... 159
Show Roles ... 162
Additional Examples of the Bubble Plot Platform ... 163
Example of Specifying Only a Time Variable ... 163
Example of Specifying Only ID Variables and Splitting a Bubble ... 165
Example of a Static Bubble Plot ... 167
Example of a Bubble Plot with a Categorical Y Variable ... 171

6 Scatterplot Matrix ... 173
View Multiple Bivariate Relationships Simultaneously

Example of a Scatterplot Matrix ... 175
Example using the Scatterplot Matrix Platform ... 175
Example of a Scatterplot Matrix in Graph Builder ... 176
Launch the Scatterplot Matrix Platform ... 177
Change the Matrix Format ... 178
The Scatterplot Matrix Window ... 179
Scatterplot Matrix Platform Options ... 179
Example Using a Grouping Variable ... 181
Create a Grouping Variable ... 182

7 Parallel Plots ... 185
View Patterns in Multidimensional Data by Plotting Parallel Coordinates

Example of a Parallel Plot ... 187
Launch the Parallel Plot Platform ... 189
The Parallel Plot ... 190

Interpreting Parallel Plots ... 190
Parallel Plot Platform Options ... 192
Additional Examples of the Parallel Plot Platform ... 193
Examine Iris Measurements ... 193
Examine Student Measurements ... 194

8 Cell Plots ... 197
View Color-Intensity Plots of Variables

Example of a Cell Plot ... 199
Launch the Cell Plot Platform ... 200
The Cell Plot ... 200
Cell Plot Platform Options ... 201
Right-Click Menu for Cell Plots ... 202
Additional Example of the Cell Plot Platform ... 203

9 Scatterplot 3D ... 205
Create a Rotating Three-Dimensional View of Data

Example of a 3D Scatterplot ... 207
Launch the Scatterplot 3D Platform ... 208
The Scatterplot 3D Report ... 209
Spin the 3D Scatterplot ... 211
Change Variables on the Axes ... 212
Adjust the Axes ... 212
Assign Colors and Markers to Data Points ... 213
Assign Colors and Markers in the Data Table ... 213
Scatterplot 3D Platform Options ... 213
Normal Contour Ellipsoids ... 216
Nonparametric Density Contours ... 216
Pop-Up Menu ... 218
Additional Examples of the Scatterplot 3D Platform ... 221
Example of an Ungrouped Normal Contour Ellipsoid ... 221
Example of Grouped Normal Contour Ellipsoids ... 222
Example of a Grouped Nonparametric Density Contour ... 223

10 Contour Plots ... 227
View Multidimensional Relationships in Two Dimensions

Example of a Contour Plot ... 229
Example Using the Contour Platform ... 229
Example of a Contour Plot in Graph Builder ... 230
Launch the Contour Plot Platform ... 231
The Contour Plot ... 233

Contour Plot Platform Options ... 233
Fill Areas ... 235
Contour Specification ... 235
Contour Plot Save Options ... 239
Use Formulas for Specifying Contours ... 240
Additional Example of Contour Plots ... 241

11 Ternary Plots ... 243
View Plots for Compositional or Mixture Data

Example of a Ternary Plot ... 245
Launch the Ternary Plot Platform ... 247
The Ternary Plot ... 248
Mixtures and Constraints ... 249
Ternary Plot Platform Options ... 249
Additional Examples of the Ternary Plot Platform ... 250
Example Using Mixture Constraints ... 250
Example Using a Contour Function ... 251

12 Maps ... 253
Add Maps or Custom Shapes to Enhance Data Visualization

Overview of Mapping ... 255
Example of Creating a Map in Graph Builder ... 256
Graph Builder ... 257
Map Shape ... 257
Color ... 259
Size ... 260
Custom Map Files ... 263
Background Maps ... 269
Images in Maps ... 271
Boundaries ... 276
Examples of Creating Maps ... 277
Louisiana Parishes Example ... 277
Hurricane Tracking Examples ... 285
Office Temperature Study ... 290
Boundary Map with Clipped Elements ... 295

A Legacy Platforms ... 299
Treemaps ... 299
Charts ... 300
Overlay Plots ... 300

B References 301
C Technology License Notices 303

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation . . . 13
JMP Help . . . 14
JMP Documentation Library . . . 14
Additional Resources for Learning JMP . . . 20
JMP Tutorials . . . 20
Sample Data Tables . . . 20
Learn about Statistical and JSL Terms . . . 21
Learn JMP Tips and Tricks . . . 21
JMP Tooltips . . . 21
JMP User Community . . . 22
Free Online Statistical Thinking Course . . . 22
JMP New User Welcome Kit . . . 22
Statistics Knowledge Portal . . . 22
JMP Training . . . 22
JMP Books by Users . . . 23
The JMP Starter Window . . . 23
JMP Technical Support . . . 23

Formatting Conventions in JMP Documentation

These conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit https://www.jmp.com/software/pro.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from https://community.jmp.com.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Model Screening • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Legacy Control Charts • Pareto Plot • Diagram • Manage Spec Limits • OC Curves

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "JMP Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "JMP Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "JMP New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\16\Samples\Data

On macOS: \Library\Application Support\JMP\16\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit https://jmp.com/tools.

Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > Initial JMP Starter Window**.

JMP Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at https://www.jmp.com/support, including the technical support phone number.

Chapter 2

Introduction to Interactive Graphing

Overview of Data Visualization

Essential Graphing describes all of the different graphs and elements you can use to visualize your data:

- Graph Builder interactively creates many different types of graphs. See "Graph Builder" and "Graph Builder Examples".
- Bubble Plot creates a scatterplot that represents its points as circles, or bubbles. Bubble plots can be dynamic (animated over time) or static (fixed bubbles that do not move). See "Bubble Plots".
- Scatterplot Matrix shows an ordered collection of bivariate graphs. See "Scatterplot Matrix".
- Parallel Plot draws connected line segments that represent each row in a data table. See "Parallel Plots".
- Cell Plot draws a rectangular array of cells where each cell corresponds to a data table entry. See "Cell Plots".
- 3D Scatterplot shows the values of numeric columns in the associated data table in a rotatable, three-dimensional view. See "Scatterplot 3D".
- Contour Plot constructs contours of a response in a rectangular coordinate system. See "Contour Plots".
- Ternary Plot display the distribution and variability of three-part compositional data. See "Ternary Plots".
- Maps can be used in Graph Builder, but also in other platforms, as background maps. See "Maps".
- Treemaps, charts, and overlay plots are deprecated platforms and can be created using Graph Builder. See "Legacy Platforms".

Chapter 3

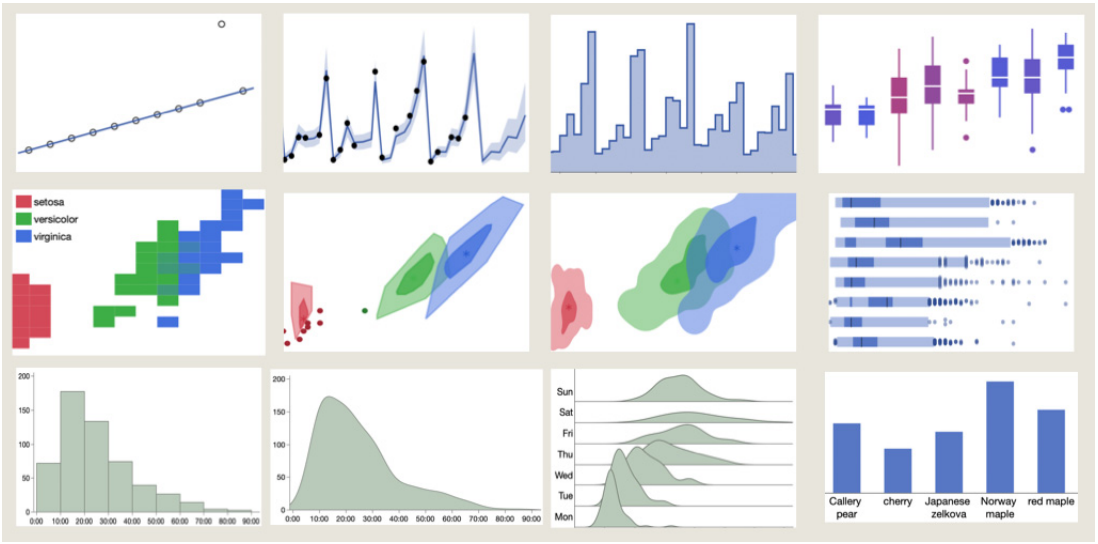
Graph Builder

Interactive Data Visualization

Use Graph Builder to interactively explore your data. You can quickly create and experiment with plots until you find the one you want. Then, you can share your results with others.

This chapter shows you how to use Graph Builder. For detailed examples using Graph Builder, see "Graph Builder Examples".

Figure 3.1 Graphs Created in Graph Builder



Contents

How to Use Graph Builder 30
About the Graph Builder Window 30
Example of Features in Graph Builder 32
Move or Remove Variables in Zones 39
Work with Axes 40
Merge Variables on a Common Axis 40
Create Separate Axes in the X and Y Zones 41
Create a Second Y Axis 43
Create Nested Axes for Character Variables 45
Order the Levels of a Categorical Variable 49
Element Types and Options 53
Points 55
Smoother 57
Line of Fit 58
Ellipse 60
Contour 61
Line 64
Bar 67
Area 78
Box Plot 81
Histogram 85
Heatmap 87
Pie 89
Treemap 91
Mosaic 94
Caption Box 95
Formula 97
Map Shapes 98
Parallel 99
Graph Zones 103
Categorical and Continuous Variables in Group X or Y Zones 107
Move Group X and Group Y Labels 108
Red Triangle Options 108
Graph Builder Buttons 110
Right-Click Menus 111
Graph Options 111
Grouping Zones Options 112
Options for Columns in the Variables Panel 112
Options for Axes, Axes Variable Labels, or Graph Titles 113

Options for the Dividing Line between Multiple Graphs 114
Legend Options 115

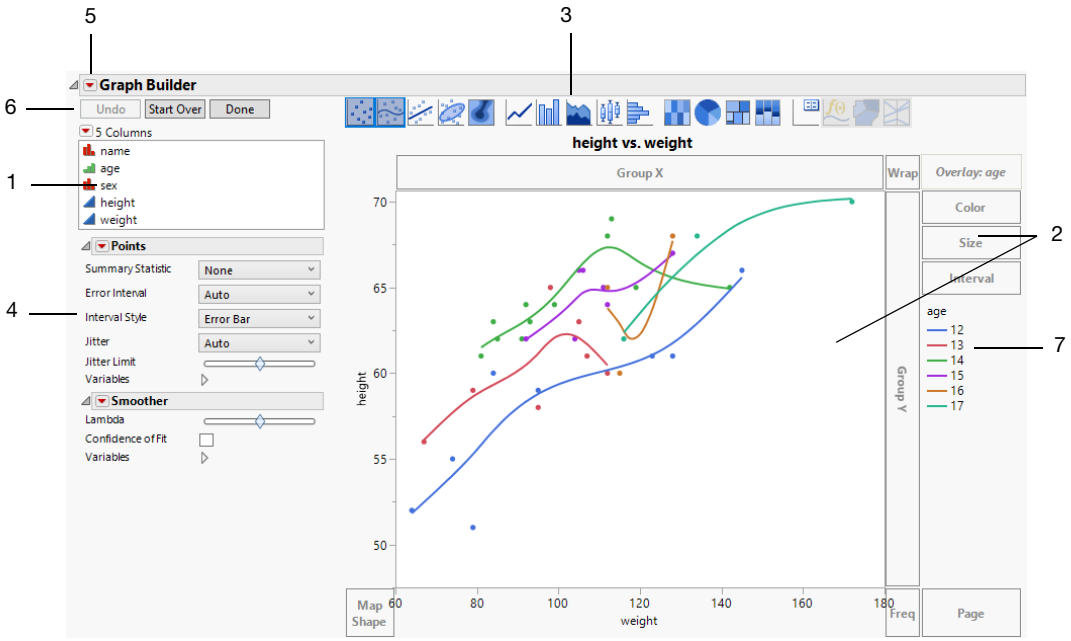
How to Use Graph Builder

1. Open the data table containing the data that you want to graph.
2. Select **Graph** > **Graph Builder**.
3. Drag columns from the Variables list into zones. See "Move or Remove Variables in Zones" and "Graph Zones".
4. Click element type icons to choose different types of graphs or elements. See "Element Types and Options".
5. (Optional) Customize the selected element types. See "Element Types and Options".
6. (Optional) Customize the legend. See "Legend Options".
7. When you are satisfied with the graph, click **Done**.

About the Graph Builder Window

To launch Graph Builder, open your data table and then select **Graph > Graph Builder**. The Variables list contains the columns in your data table, which you can drag into zones.

Figure 3.2 Graph Builder Window for Big Class.jmp



Number	Description
1	The columns in your data table are listed in the Variables panel. Drag columns from the Columns list into zones, or select the column in the Variables list and click the empty zone. You can create a virtual column by right-clicking a column name in the Variables list and selecting from the options provided. • The transformed column appears at the bottom of the list in italics. • You can drag that column into Graph Builder zones. • Right-click the transformed column for further options. **Tip:** For descriptions of the column options in the Variables red triangle menu, see *Using JMP*.
2	The graph area and zones are the main body of Graph Builder. Your graph appears in the middle, and you drag variables into different zones. See "Graph Zones". **Note:** Any excluded rows are hidden in graphs.
3	Click an element type icon to choose the type of graph that you want. You can apply several types of elements to your data simultaneously (using the Shift key). See "Element Types and Options".
4	Customize the graph with specific options for the type of graph you have selected. See "Element Types and Options".
5	Customize the Graph Builder window. See "Red Triangle Options".
6	Undo actions, start over, and more. See "Graph Builder Buttons".
7	If your graph has a legend, it appears here. To customize the legend, see "Legend Options".

To re-create and modify the graph in Figure 3.2, follow these steps:

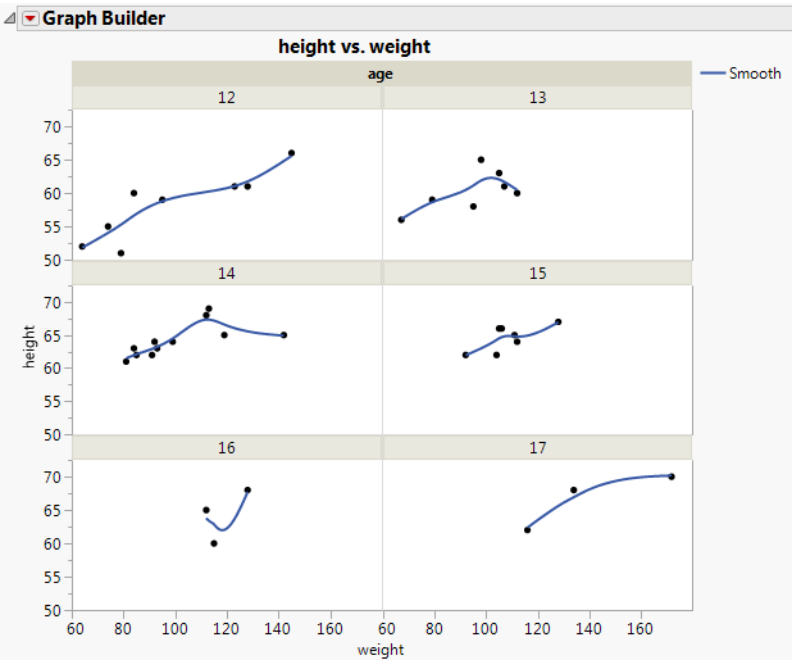
1. Select **Help** > **Sample Data Library** and open Big Class.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select height and drag it to the **Y** zone.
4. Select weight and drag it to the **X** zone.
5. Select age and drag it to the **Overlay** zone.

The graph in Figure 3.2 appears. The relationship between height and weight for each value of age is shown using color in a single graph.

Now you want to see separate graphs for each value of age.

6. Select age in the **Overlay** zone and drag it to the **Wrap** zone.

Figure 3.3 Height and Weight Grouped by Age



For each value of age, the relationship between height and weight is shown in a separate graph.

Example of Features in Graph Builder

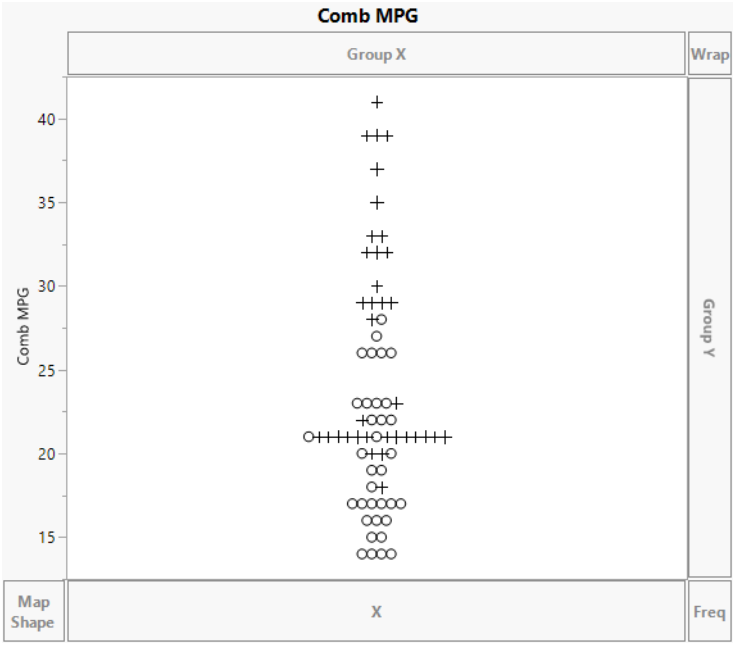
The following example gets you started using some of the features in Graph Builder. This example uses fuel economy data collected for hybrid and non-hybrid cars. You want to get an understanding of miles per gallon (MPG), a measure of fuel economy, and to see which factors might be influencing MPG. This example shows you how to plot several factors of interest:

- "Get an Overall Picture of Combined Miles per Gallon"
- "Plot Mileage Stratified by Engine and Driving Type"
- "Find the Relationship between Hwy and City MPG by Engine Type"
- "Find the Relationship between Hwy and City MPG by Engine Type"

Get an Overall Picture of Combined Miles per Gallon

1. Select **Help > Sample Data Library** and open Hybrid Fuel Economy.jmp.
2. Select **Graph > Graph Builder**.
3. Select Comb MPG (combined miles per gallon) and drag it to the **Y** zone.

Figure 3.4 Initial Graph of Combined MPG



Because markers are assigned to rows in the data table, points are represented by those markers. The value of Comb MPG for each row is plotted at its value on the vertical axis. To avoid over-plotting points that correspond to observations with the same Comb MPG values, JMP automatically jitters the points.

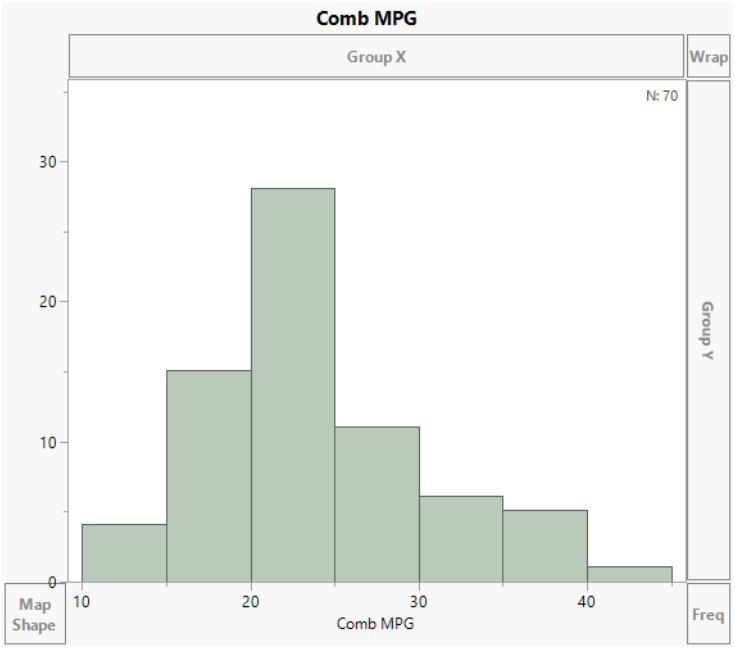
You can get a cleaner picture of the distribution by plotting a histogram.

4. Click the Histogram icon .

   You want the histogram to be vertical, so you need to move Comb MPG to the X zone.

5. Hover over Comb MPG in the Y axis. The cursor turns into a hand. Select and drag Comb MPG into the **X** zone.
6. Click the Caption Box icon to show the mean Comb MPG on the graph.

Figure 3.5 Histogram of Combined MPG



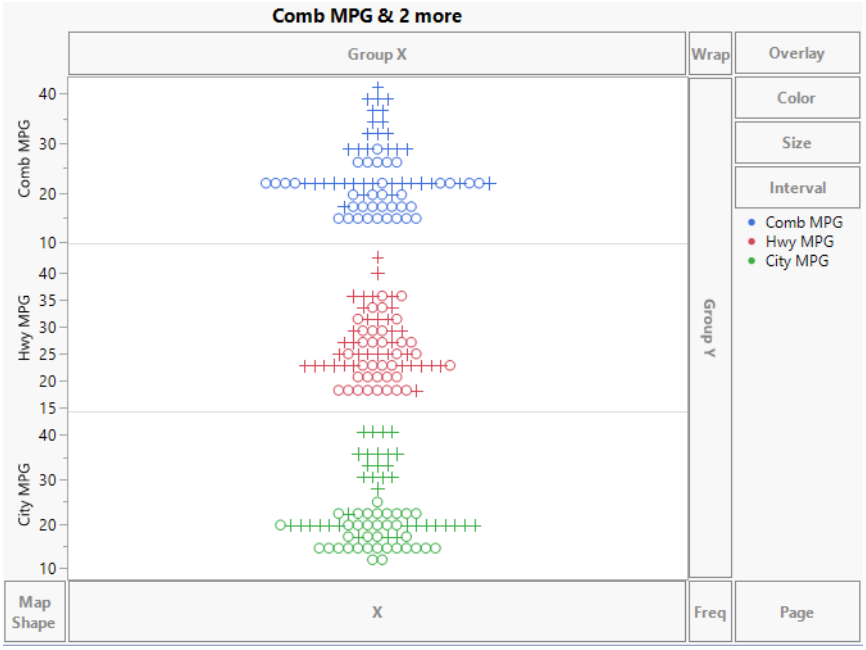
The histogram shows that Comb MPG can range widely, from about 10 to 45. However, a lot of vehicles get mileage between 15 and 30. The mean MPG is 23.3286.

Plot Mileage Stratified by Engine and Driving Type

1. Click **Start Over**.
2. Select City MPG, Hwy MPG, and Comb MPG.
3. Hover over the Y zone, hold down the Shift key and click in the Y zone.

   This creates separate Y axes for City MPG, Hwy MPG, and Comb MPG.

Figure 3.6 MPG Variables with Separate Y Zone Axes

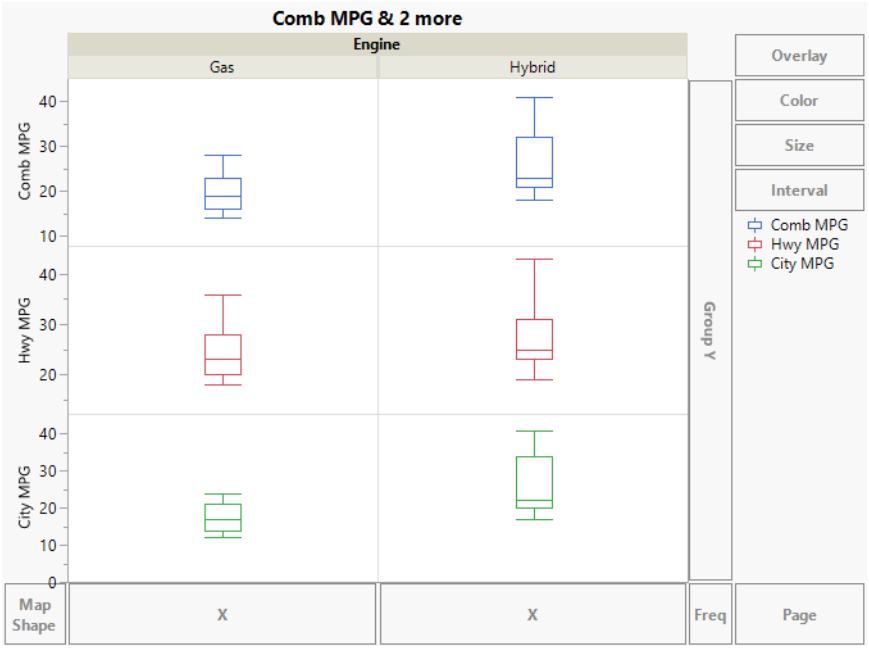


In the legend at right, markers are colored according to the driving type.

Tip: To change the color assigned to a driving type, right-click the marker in the legend.

4. Drag Engine to the Group X zone.
5. Click the Box Plot icon .

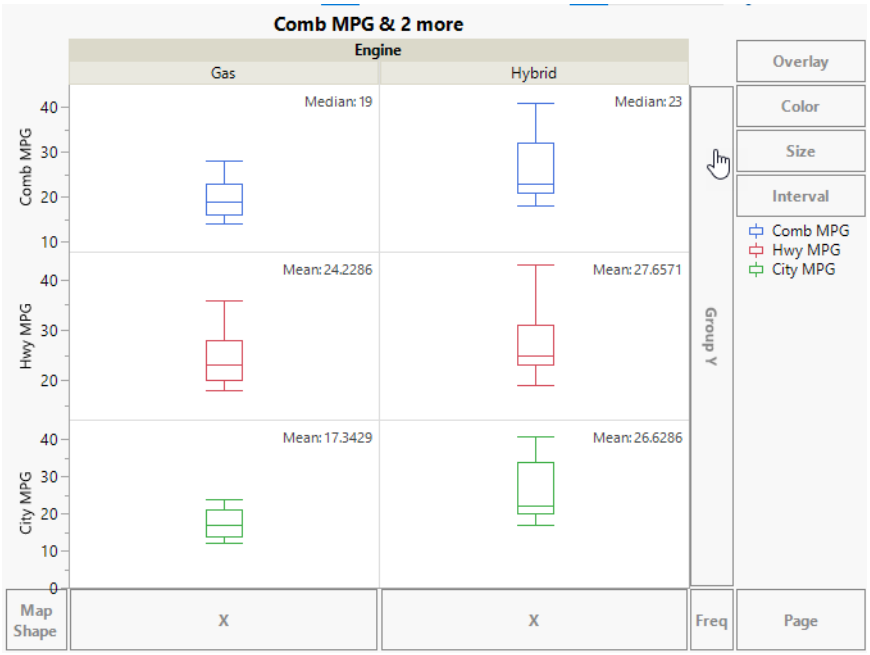
Figure 3.7 Box Plots of MPG Variables by Engine Type



The box plots show the distributions in a very compact form. You can see that all three types of MPG values are substantially lower for the gas vehicles than for the hybrids in the study. From the horizontal lines in the centers of the two City MPG box plots, you can estimate that the median difference is about five miles per gallon. However, you can use the Caption element to be sure.

6. Click the Caption Box icon .
7. In the Caption Box options at the left of the plot, select **Median** from the Summary Statistic list for Comb MPG, Hwy MPG, and City MPG.

Figure 3.8 Box Plots of MPG Variables with Captions

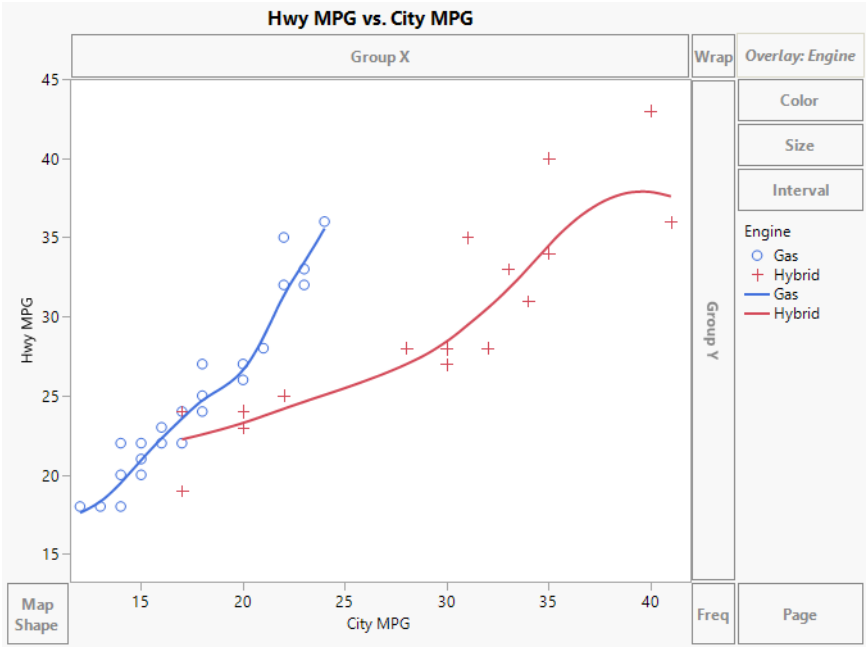


The median city MPG for gas engines is 17 and for hybrid engines is 22.

Find the Relationship between Hwy and City MPG by Engine Type

1. Click **Start Over**.
2. Select City MPG and drag it into the **X** zone.
3. Select Hwy MPG and drag it into the **Y** zone.
4. Select Engine and drag it into the **Overlay** zone.

Figure 3.9 Hwy and City MPG by Engine Type



A Smoother appears for each type of engine, Gas, and Hybrid.

Tip: To change the line properties (color, width, and so on), right-click a line in the legend.

5. Click the Line of Fit icon .

   For each engine type, the smoother is removed and a least squares line, together with confidence bands for the predicted mean, is added.

   Tip: To add the y = x line, right-click in the graph and select **Customize** > **+** > **Templates** > **Y Function**. Replace `_function_of_x_` with `y = x` and click **OK**.

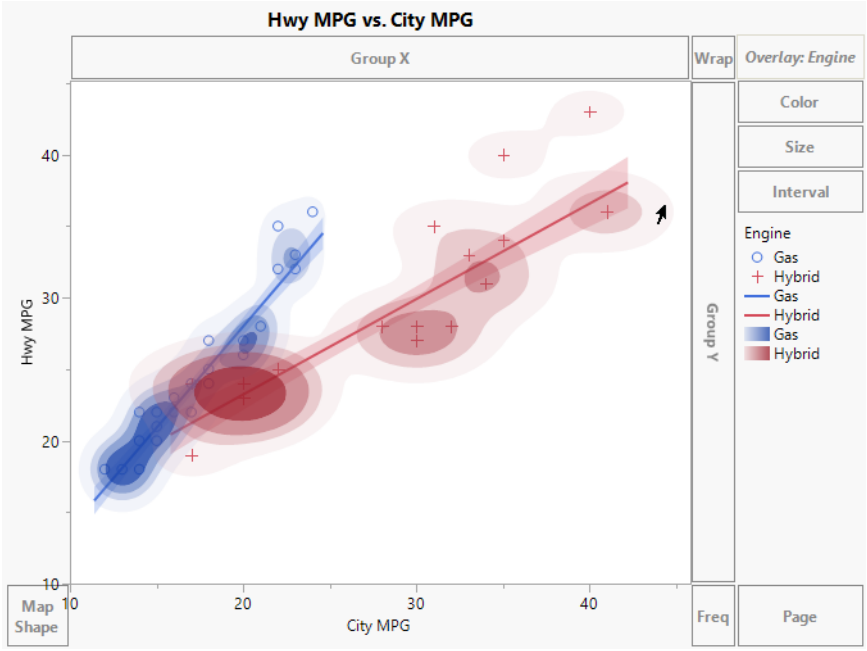
6. Drag a rectangle around the "o" marker with the smallest value of City MPG and check the number of rows selected in the Rows panel of the data table.

   You see that two rows are represented by this marker. You can check that over-plotting happens for other markers as well. When there is over-plotting of observations, namely when several observations are represented by a single point on a plot, density contour plots can help you see the density of points.

7. Drag the Contour icon into the graph.

   Dragging the Contour icon into the graph retains the Line of Fit for each engine type.

Figure 3.10 Density Contour for Hwy and City MPG by Engine Type



The contours indicate that for both engine types, there are more vehicles in the study with lower MPG values than higher MPG values.

Move or Remove Variables in Zones

There are several ways that you can move a variable into a zone:

- To move a variable from one zone to another, click and drag it from the old zone into the new zone.
- To replace a variable in a zone, drag the new variable over the old one.
- To switch a variable from one zone to another, right-click the variable in the zone that you want to move and select **Swap**. Then, select the variable that you want to switch places with.
- To merge variables on the same axis, see "Merge Variables on a Common Axis".

To remove a variable, drag it into empty space, or right-click the name of the variable in the zone and select **Remove**.

Work with Axes

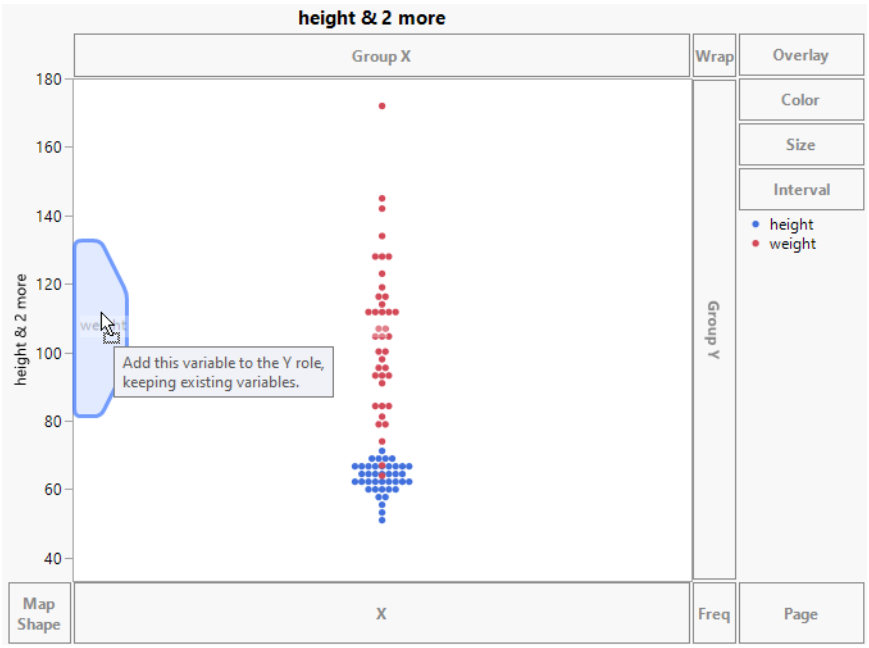
- "Merge Variables on a Common Axis"
- "Create Separate Axes in the X and Y Zones"
- "Create a Second Y Axis"
- "Create Nested Axes for Character Variables"
- "Order the Levels of a Categorical Variable"

Merge Variables on a Common Axis

Merging variables places both variables on the same axis and creates a single graph for both variables. Elements in the plot are colored by the variables, and a legend appears to the right of the plot. When points are plotted, they are randomly jittered for each value or category defined by the combination of the axes.

- If you are starting with an empty zone, highlight multiple variables from the list and drag them into the zone at the same time.
- If you already have variables in the zone and want to add more, drag the new variables to the inner side of the zone, in the graph area, until a symmetric polygonal shape appears.

Figure 3.11 New Variable Merged with Existing Variable



Results Based on Data and Modeling Type

- If only variables with continuous modeling types are merged, their values are plotted against a common axis.
- If variables with categorical (nominal or ordinal) modeling types are merged with a continuous variable, the values of all variables are plotted against a common axis.

  Before you can merge a variable with the character data type with numeric variables, the levels of the character variable have to be assigned numeric values. Integer values are assigned based on the Value Order column property or the default value ordering. These integer values range from 0 (highest level in value ordering) to the number of levels minus one (lowest value in value ordering). These integer values are then plotted against the common axis.
- If only variables with categorical modeling types are merged, then separate, *nested*, axes are constructed for each variable. See "Create Nested Axes for Character Variables".

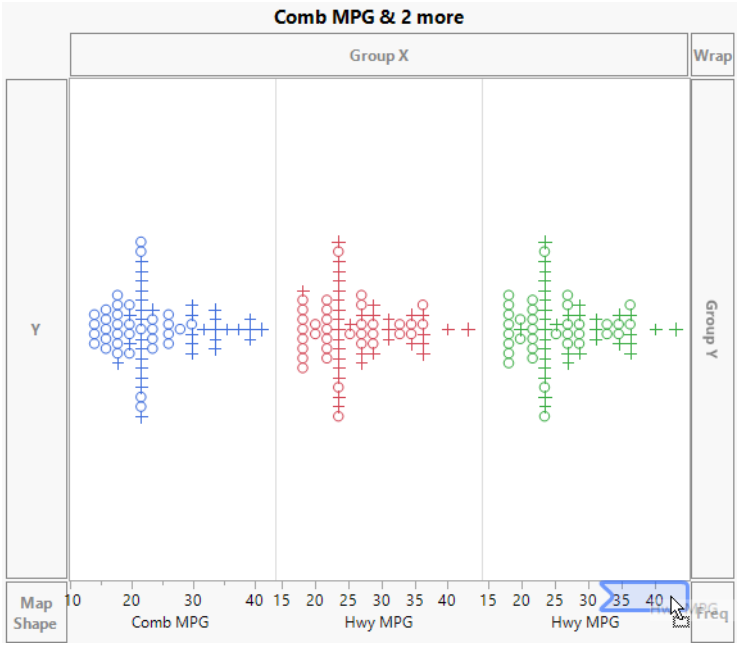
Create Separate Axes in the X and Y Zones

To quickly plot multiple variables against separate axes in the X zone:

1. Select the columns in the Variables list.
2. Click **Shift** and then click the **X** zone.

If you want more control over where the variables appear, drag them individually. Drag additional variables above, below, or between existing variables in the X zone.

Figure 3.12 Dragging a Variable to the Right of the Existing Variable



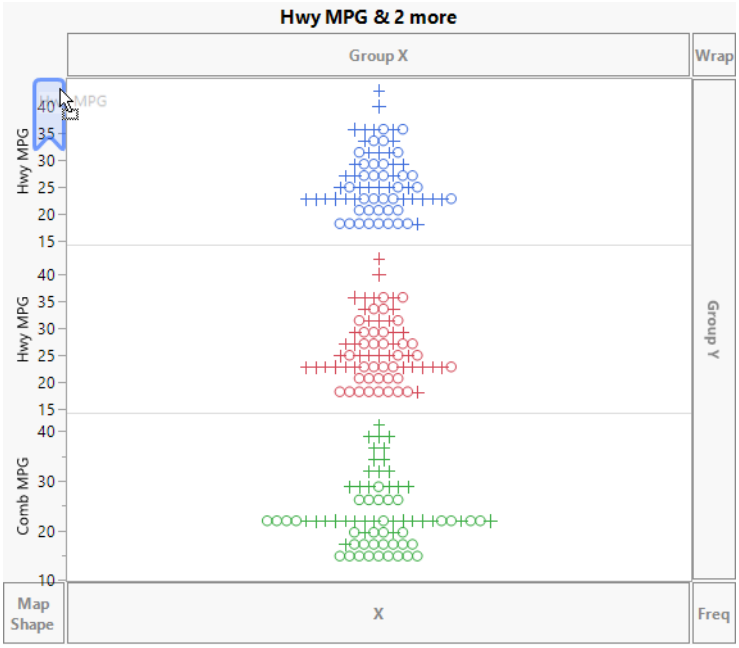
A single polygonal shape indicates when your variable is in the appropriate target zone.

To quickly plot multiple variables against separate axes in the Y zone:

1. Select the columns in the Variables list.
2. Click **Shift** and then click the **Y** zone.

If you want more control over where the variables appear, drag them individually. Drag additional variables above, below, or between existing variables in the Y zone.

Figure 3.13 Dragging a Variable above the Existing Variable



A single polygonal shape indicates when your variable is in the appropriate target zone.

Create a Second Y Axis

If you have two or more Y variables on the same axis, you can reflect the scaling of a second set of variables by creating a second Y axis. This can be useful when two variables measure the same underlying quantities, but have different scales. In general, it is unwise to use a second Y axis in any other situation. See Few (2008).

To create a second Y axis:

1. Right-click one of the Y variable names in the Y zone and select **Move Right**.
2. Select the variable or variables that you want to move to the new axis.

   The new axis is scaled according to the values of the specified variable, and the selected variable is plotted against this axis.
3. Repeat the process to plot additional variables against the new axis.

   The new axis adjusts to accommodate the values of the additional variables.

Change the Graph for a Second Y Axis

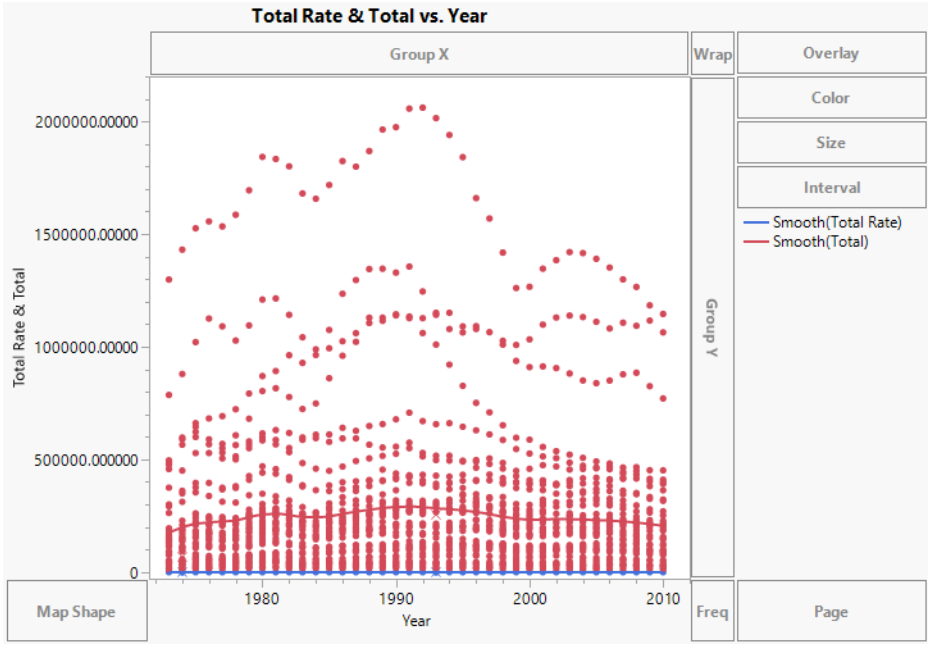
When you click an element type, it applies only to the variables on the left Y axis. To apply an element to the variables on the right Y axis, right-click and change the second element, which corresponds to the right axis.

Example of Creating a Second Y Axis

The CrimeData.jmp sample data table contains data on various types of crime for the 50 US states yearly, from 1973 to 2010. The Total column sums the number of incidents, and the Total Rate column gives a population-adjusted rate. You want to show these two variables on a single chart.

1. Select **Help > Sample Data Library** and open CrimeData.jmp.
2. Select **Graph > Graph Builder**.
3. Select Total Rate and Total and drag them to the Y zone.
4. Select Year and drag it to the X zone.

Figure 3.14 Total Rate and Total Merged

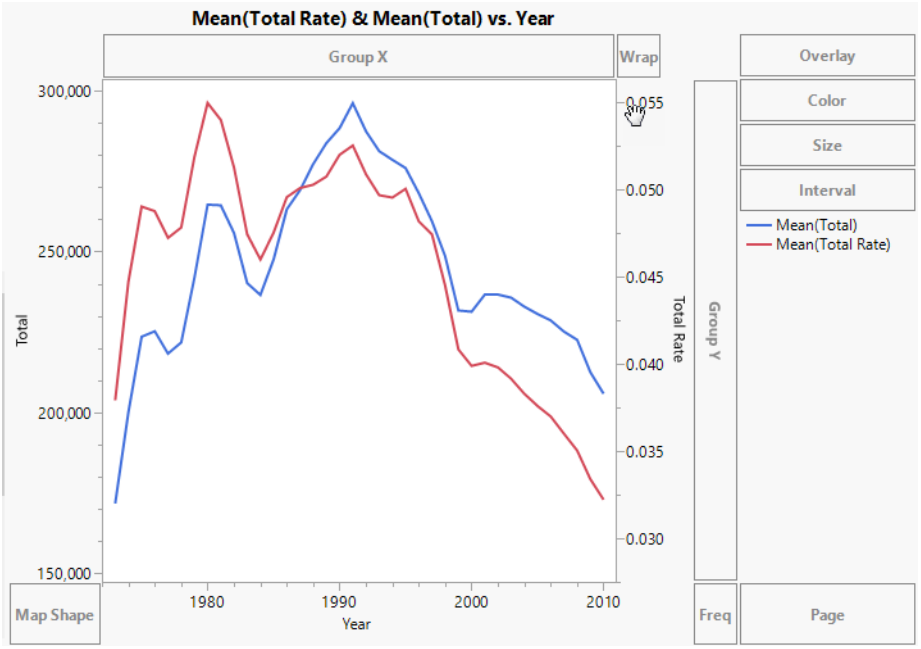


The Total Rate values are all between 0 and 1, and they are barely visible. Calculate yearly means and then use a second Y axis to make these values visible.

5. In the options panel for Points, select **Mean** next to **Summary Statistic**.
6. Click the Line element .

7. Right-click the variable names in the Y zone and select **Move Right** > **Total Rate**.

Figure 3.15 Second Y Axis Added



An axis for Total Rate is added on the right, and the axes rescale to show both sets of values. You can now see relationships between the two measures. For example, both measures began to decrease in 1991.

8. (Optional) Click **Done**.

Create Nested Axes for Character Variables

If you merge variables that all have categorical modeling types, then an individual axis is constructed for each variable. The outermost axis corresponds to the first variable selected, the next to the second, and so on.

To nest the axes in a specific order, drag the variables to the zone individually. First, drag the variable for the innermost axis to the zone. Then drag the variable for the next axis to the outside of the preceding variable and drop the variable once a trapezoid shape appears.

The following example illustrates nested axes.

After changes are made to improve a measurement process, a measurement systems analysis study is conducted to study repeatability and reproducibility with Part, Operator, and Instrument as factors. Each of three operators measures each of eight parts with four instruments. Of particular interest is the consistency of the instruments. The measured quantity is called new Y.

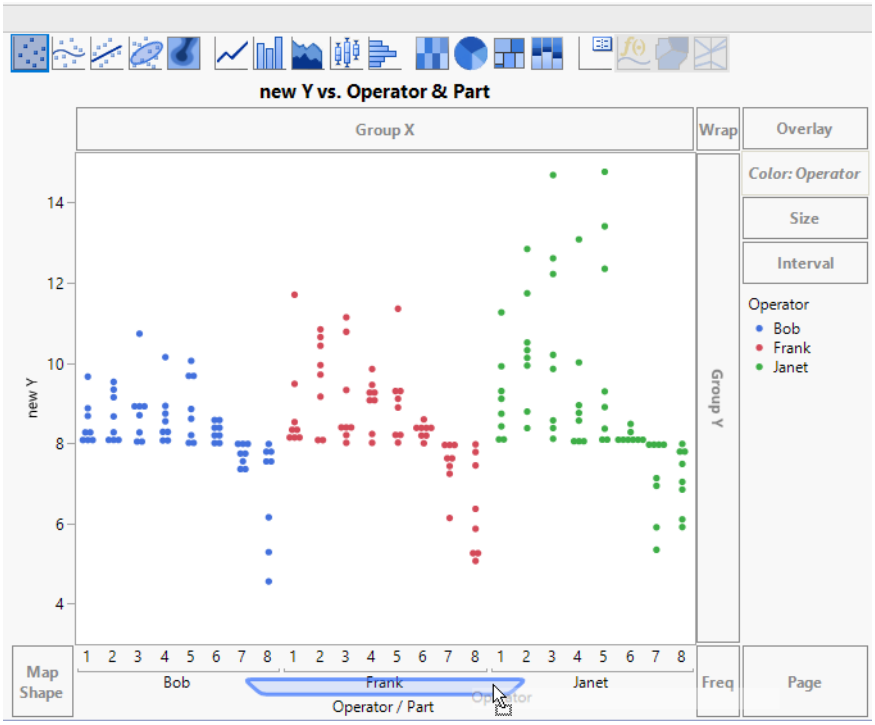
1. Select **Help > Sample Data Library** and open Variability Data/3 Factors Crossed.jmp.
2. Select **Graph > Graph Builder**.
3. Select new Y and drag it to the **Y** zone.
4. Select Part and drag it to the X zone.

   The plot shows variation in the values that are measured for each of the eight parts. There are systematic differences among the parts, which is to be expected. For example, measurements for parts 7 and 8 are lower than those for parts 1 through 6.
5. Drag Operator to the **Color** zone.

   The new Y values are colored by Operator, using the legend at the right of the graph. It appears that Janet might be measuring higher values than the other two operators on most parts. But the Operator effect is not easy to visualize, so you create a separate Operator axis.
6. Select Operator and drag it beneath Part in the X zone.

   The label Operator/Part appears, indicating that Part is associated with the topmost axis, and Operator is associated with the lower axis.

Figure 3.16 Drag Operator to Add a Second Axis

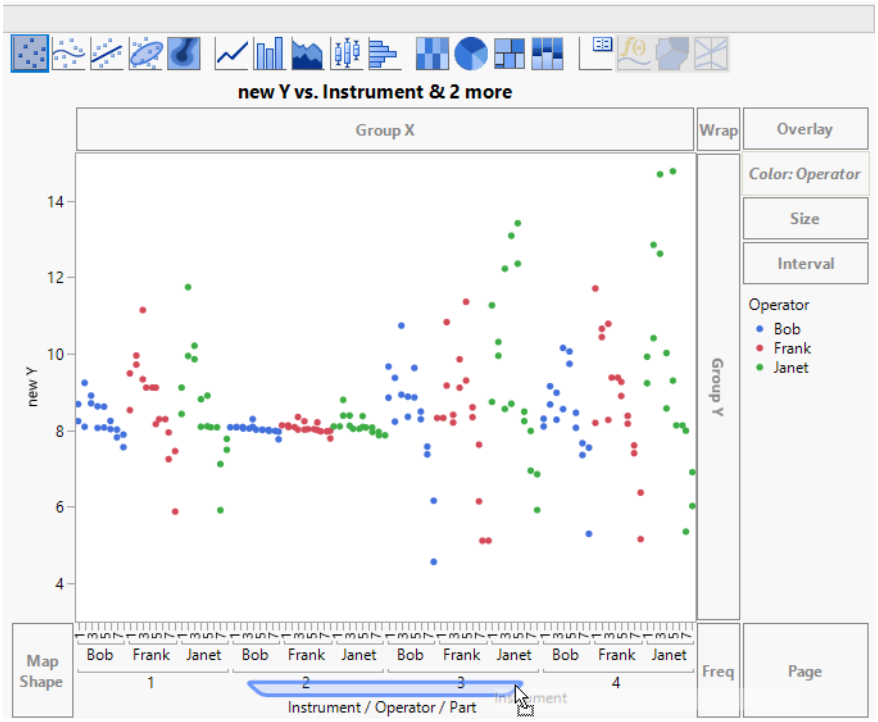


Now it is easier to see that Janet tends to obtain higher measurements for the same parts than do Bob and Frank. But what about the effect of Instrument?

7. Select Instrument and drag it beneath Operator/Part in the X zone.

   The label Instrument/Operator/Part appears, indicating that a third axis for Instrument has been added beneath the Operator axis.

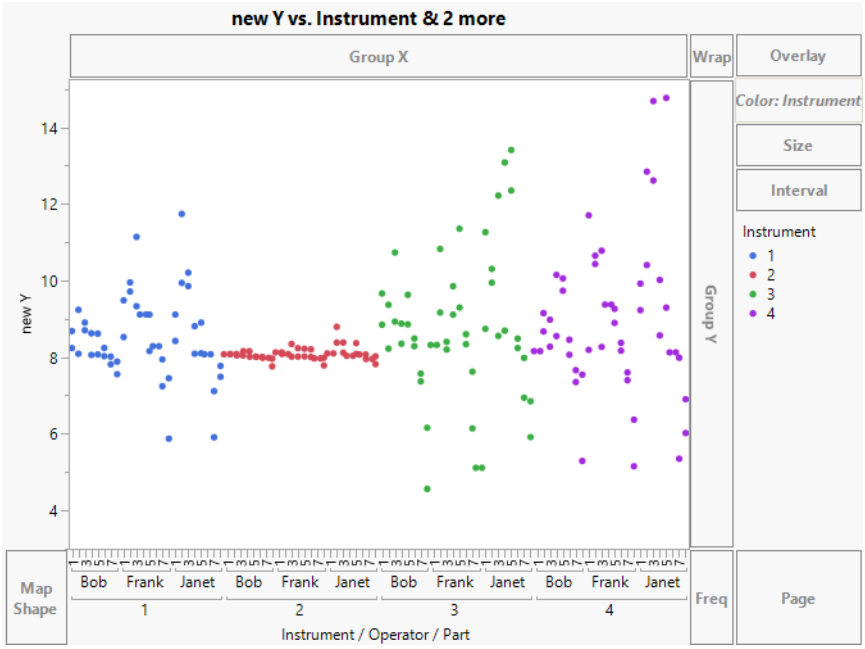
Figure 3.17 Drag Instrument to Add a Third Axis



It is clear that Instrument 2 leads to much more consistent measurements than the other three instruments. For Instrument 2, there is comparatively little variation between or within operators.

8. Select Instrument and drag it to the **Color** zone.

Figure 3.18 Three Nested Axes



Now the new Y values are colored by Instrument, and it is easy to see Instrument differences.

By nesting the axes for the three factors in study, you are able to obtain a visual understanding of the variation attributable to the factors.

Order the Levels of a Categorical Variable

JMP orders the levels of a categorical variable according to the following rules:

- Numeric, nominal data are sorted numerically.
  - White space around numbers is compared: “vt 1” is sorted before “vt1”.
- Character data that are only digits (numbers) are sorted numerically.
- Character data are sorted alphabetically, with the following exceptions:
  - Months and days of the week are in chronological order.
  - Ratings are sorted from low to high:

    Low to high: Very Low, Low, Medium Low, Medium, Medium High, High, Very High

    Agreement: Strongly Disagree, Disagree, Neutral, Indifferent, Agree, Strongly Agree

    Bad or good: Failing, Unacceptable, Very Poor, Poor, Bad, Acceptable, Average, Good, Better, Very Good, Excellent, Best

- Character data that have a character prefix and a numeric suffix are sorted first by prefix and then by suffix. For example, lots 1 through 12 are ordered “lot1”, “lot2”, “lot3”,...”lot10”, “lot11”, and “lot12”.
- If the numeric suffix starts with a “0”, it does not follow numeric ordering by suffix. (“lot1” is ordered *after* “lot02”).

Note: The preceding rules apply only to graphs with categorical axes.

To change the order of the levels of a categorical variable on a graph axis, click and hold a level on the axis to enter edit mode. In the edit mode, denoted by boxes around each label, you can drag one or more labels (using shift-click for multiple selection) to rearrange them. As you rearrange labels, the value order column property is updated. You can also edit labels, which updates the value label column property. Press Esc to exit the edit mode.

You can also order the levels of a categorical (nominal or ordinal) variable on axes using the Value Order column property, which takes precedence over all other ordering rules. To change the ordering, you can do one of the following:

- Use the Value Order column property to ensure that values are ordered as you intend. For more information about the Value Order column property, see *Using JMP*.
- Use the values of a numeric variable already in the graph. In Graph Builder, right-click the categorical axis and select **Order By**.
- Use the values of an arbitrary numeric variable. In Graph Builder, drag the numeric variable next to the categorical variable in the zone until a trapezoid appears and then drop the variable. The axis label changes to <categorical variable> ordered by <numeric variable>. See “Example of Ordering the Levels of a Categorical Variable Using a Numeric Variable”.

Change the order of the levels or the statistic:

1. Right-click the axis label and select **Order By**.
2. Select one of the ascending or descending options or **other** to select a numerical column.

   You can order by a variable’s Order Statistic (these options are prefixed by the variable’s name) or by the number of observations in each level of the categorical variable (Count).
3. The default ordering statistic is the mean. To use another statistic, right-click the axis label and select **Order Statistic**.

Note: If you try to order the values of a numeric variable using another numeric variable, JMP merges the variables. See “Merge Variables on a Common Axis”.

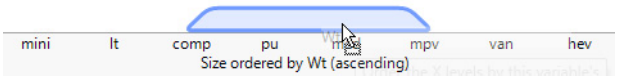
Example of Ordering the Levels of a Categorical Variable Using a Numeric Variable

To order a nominal or ordinal variable by a numeric variable, consider data about vehicle types. You want to see the vehicle types arranged in a meaningful order.

1. Select **Help > Sample Data Library** and open Cars.jmp.
2. Select **Graph > Graph Builder**.
3. Select Size and drag it into the **X** zone.

   This variable represents the type of the vehicle. Eight levels are listed alphabetically on the X axis: compact (comp), heavy (hev), lightweight (lt), medium (med), mini, multi-purpose (mpv), pick-up truck (pu), and van. Since the levels are listed alphabetically, they are not ordered in a meaningful way. For example, heavy comes before mini and lightweight. You want to order the levels by Wt (weight).
4. Select Wt and drag it to the middle of the X axis. Drop it just above the **X** axis. Before you drop the variable, a blue quadrilateral appears.

Figure 3.19 Merging Wt and Size

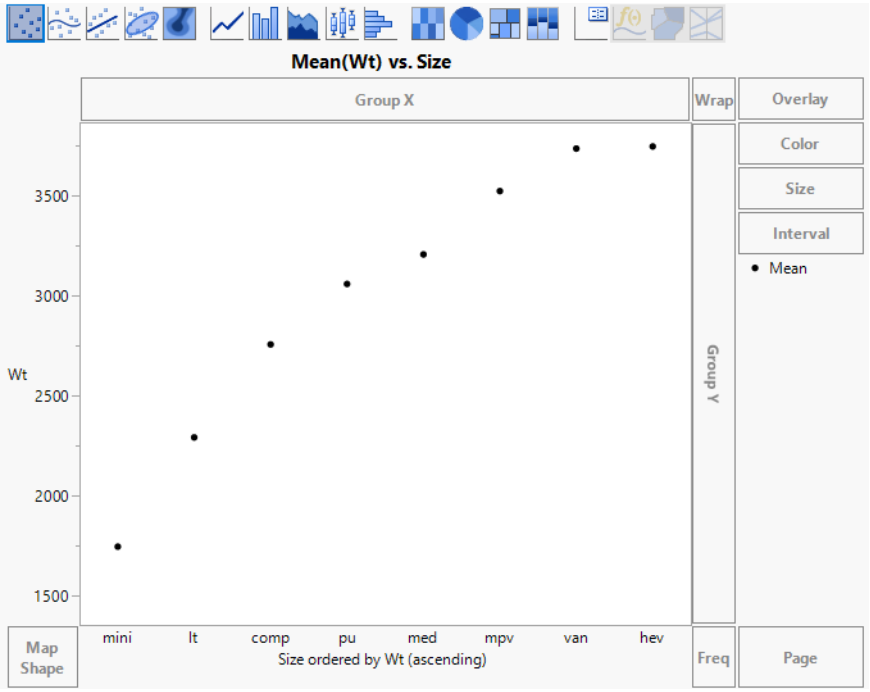


The levels of Size are now arranged in increasing order according to the average Wt of all vehicles in the levels. Notice that mini and lt (lightweight) are now ordered before hev (heavy). The axis label is updated, signifying that an ordering variable is in use.

Next, verify that Size is actually ordered by Wt.

5. Select Wt and drag it to the **Y** zone.
6. Select **Mean** from the Summary Statistic list.

Figure 3.20 Example of Size Ordered by Wt, Ascending



You can see that the average Wt increases from left to right.

Change the order from ascending to descending.

7. Right-click in the **X** zone and select **Order By** > **Wt, descending**.

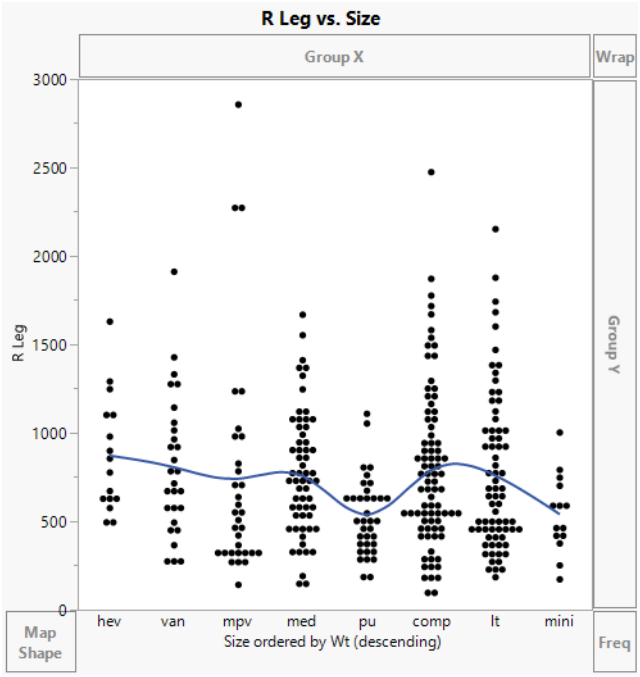
   Now the levels of Size are arranged in decreasing order of mean Wt. You want to see whether the right femur load (R Leg) decreases with vehicle weight.

8. Right-click Wt in the Y zone and select Remove.
9. Select R Leg and drag it to the Y zone.
10. Select **None** from the Summary Statistic list.

    This selection replaces the means with points for all the individual observations.

11. Click the Smoother element .

Figure 3.21 Example of R Leg Ordered by Wt, Descending



R Leg seems unrelated to the weight classes.

The default ordering statistic is the mean. To use another statistic, right-click in the **X** zone, select **Order Statistic** and change it to the statistic that you want.

Element Types and Options

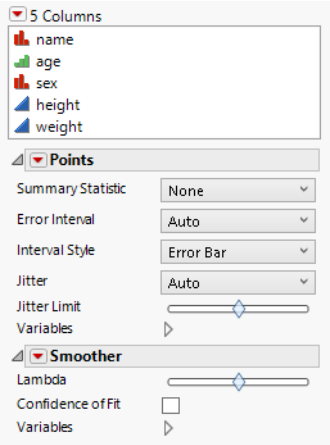
You can change how your data appear by clicking an element type icon, such as a bar chart, line, or histogram.

Figure 3.22 Element Type Icons

Note: Supported element types vary depending on the variable type and zones that are selected. Element types that are not applicable appear dimmed.

For each element that you add to a graph, an outline appears in the Properties area at left. This is where you specify and change properties for each element type.

Figure 3.23 Properties Areas for Points and Smoother

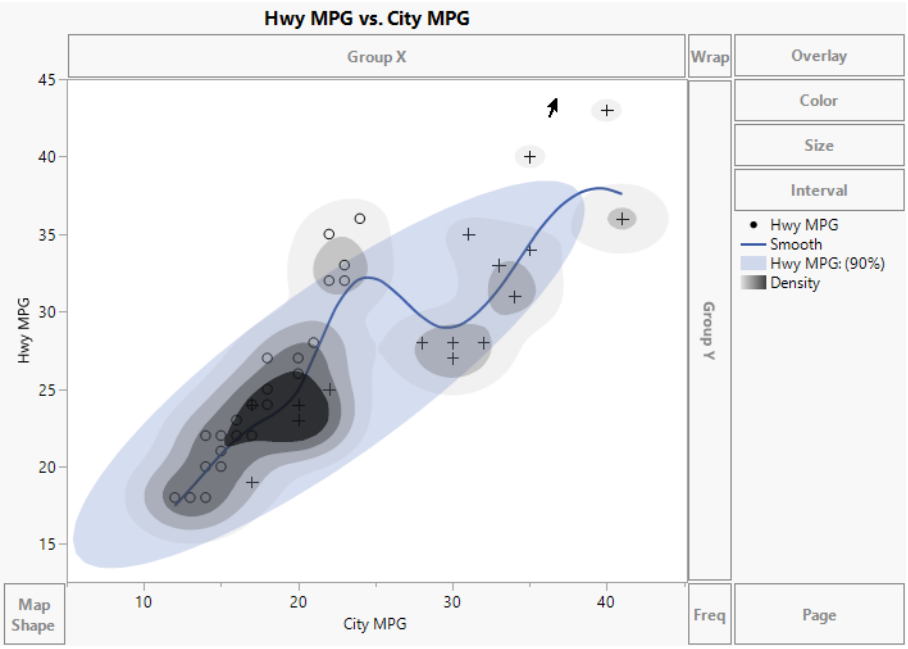


You can also right-click in any graph to change the element, or customize elements in the graph.

Select Multiple Elements at Once

To select and overlay multiple elements, hold down the Shift key and click the elements. Alternatively, you can drag the elements into the graph one at a time.

Figure 3.24 Graph Showing Points, Smoother, Ellipse, and Contour Elements



Labels for Points

In most graphs that display points, when you hover over a point, a label appears showing information about the corresponding row. Hover over the label and right-click for more options. You can pin the label to the graph, copy the contents, or close to remove the label. You can also pin the label by clicking the pin icon in the upper right of the label. See *Using JMP*.

Hidden and Excluded Rows

Hide and exclude are two different row states:

- The goal of a hidden row is to impact visibility but not analytical results.
- The goal of an excluded row is to impact the analytical result.

For graphs, an analytical value is usually part of the appearance. Therefore, excluding rows typically impacts the graph appearance and any analytical result included as part of the graph. Hiding a row typically impacts only the graph appearance.

For example, consider a graph with points and a line of fit:

- If rows are hidden, those rows are included in the calculations to obtain the line of fit but they are not plotted.
- If rows are excluded, those rows are excluded from both the calculations and the display.

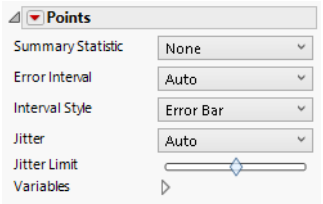
In the two scenarios, the number of points that appear is the same, but the line of fit is different.

When graphing groups, if all rows in a group are hidden, then the visual for that group is also hidden.

Points

The **Points** element shows data values as points.

Figure 3.25 Points Options



Summary Statistic Changes the statistic being plotted. The statistic is calculated using the variable on the Response Axis. None is the default setting, indicating that the data values themselves are plotted.

Error Interval Adds or removes specified error interval in the graph.

Interval Style Draws error bars or error bands for the selected error interval.

Jitter Jitter adds random noise to values to reduce over-plotting. Choose from the following types of jitter:

None No adjustments are made.

Auto Adds various types of jitter when categorical variables are involved. No jitter is added when you have only continuous variables.

Random Uniform Random offset with uniform distribution.

Random Normal Random offset with Gaussian distribution.

Packed Places markers tightly to preserve any non-jittered dimensions.

Centered Grid Similar to Packed, but adjusts the non-jittered dimensions to fall into a grid.

Positive Grid Similar to Center Grid, but in the positive direction.

Density Random For one-dimensional jitter, this option places markers randomly within the bounds of a violin plot. For two-dimensional jitter, this option places markers randomly within an oval.

You can also customize jitter by right-clicking and selecting **Customize** > **Marker**.

Tip: To create a dot plot, select the Positive Grid option. You can move the dot plot to start at the bottom by changing the Y-axis values to 0-1. You can also resize the markers to a larger value, such as 20, to adjust the binning of the dots.

Jitter Limit Controls the spread or amount of overlap.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Points

Response Axis Changes the axis that is associated with the response variable to X (horizontal), Y (vertical), or automatically determines the response direction.

Set Shape Expression Enables the use of a JSL expression to define a marker shape.

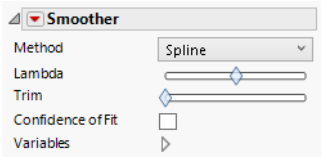
Set Shape Column Sets a column to use as the marker shape.

Smoother

The **Smoother** element shows a smooth curve through the data.

Tip: To label a smoother, right-click on the smoother element in the legend.

Figure 3.26 Smoother Options



Method Specifies the method used for generating the smooth fit. Options include Spline, Local Kernel, Savitzky-Gloay, Moving Average, and Moving Box.

Lambda (Available for Spline.) Enables you to adjust the value of lambda. The default value is 0.05. For more information about lambda, see *Basic Analysis*.

Degree (Available for Savitzky-Golay.) Specifies the degree of the smoother.

Local Width (Not available for Spline.) Enables you to control the smoother fit in terms of a local to a global fit. Use the red triangle Local Region option for further control.

Trim Enables you to trim the sections of the smoother that are based on incomplete or imbalanced data ranges. The slider is scaled from 0 to 1. At 1, all of the incomplete or

imbalanced regions are trimmed with the exception of the spline where 1 corresponds to the penultimate edge knots.

Confidence of Fit Shows or hides the bootstrap confidence region for each fit.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Smoother

Response Axis Changes the axis that is associated with the response variable to X (horizontal), Y (vertical), or Auto (the Y axis for Smoother).

Local Region Specifies options for controlling the region used for the local width option.

Local Weighting Specifies weighting options.

Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

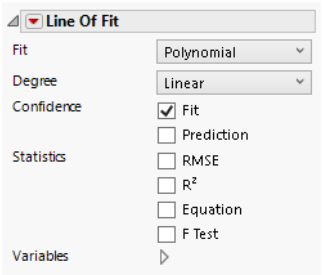
Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Line of Fit

The **Line of Fit** element shows a linear regression line with confidence intervals for the fit.

Tip: To label a line of fit, right-click on the line element in the legend.

Figure 3.27 Line of Fit Options



Fit Specifies the type of fit:

Polynomial A simple linear regression.

Robust Cauchy A linear regression assuming Cauchy distributed residuals, to de-emphasize outliers.

Time Series Smoothing for equally spaced X values with optimal seasonality. Includes options for a forecast model and the number of seasonal and forecast periods. The smoothing model is selected from a subset of state space smoothing models defined by Hyndman et al. (2008). See *Predictive and Specialized Modeling*.

Degree Specifies the polynomial degree of the linear regression fit, which can be linear, quadratic, or cubic.

Confidence Shows or hides confidence intervals for the predicted value (Fit) or for individual values (Prediction). Both types of intervals are fixed at 95% confidence.

Statistics Shows various selections on the graph. You can show the root mean square error (RMSE), R-square, the equation of the regression line, and the F Test value.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.
– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Line of Fit

Response Axis Specifies the axis for the variable that is used as the response in the calculation of the linear regression line. Because standard least squares typically uses the Y (vertical) axis as the response axis, the Auto setting defaults to the Y axis.

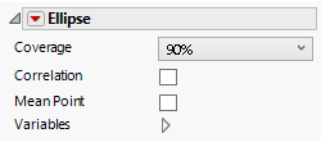
Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Ellipse

The **Ellipse** element shows a bivariate normal density ellipse.

Figure 3.28 Ellipse Options



Coverage Specifies the coverage percentage (99%, 95%, 90%, or 50%) for the density ellipse.

Correlation Shows the Pearson correlation coefficient for the X and Y variables on the graph.

Mean Point Shows the mean on the graph.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option for Ellipse

Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

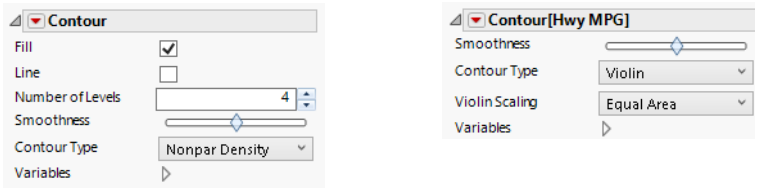
Contour

The **Contour** element shows regions of density (or value contours when used with a Color variable). Density contours are useful when you have a scatterplot with many points where the mass of points makes it difficult to see patterns in density. Multiple contour types are available. The default is a smooth bivariate nonparametric density surface that is fit to reflect the density of the data points. The nonparametric density surface estimates the bivariate probability density function at each point, providing a continuous analog of a bivariate histogram.

- For two continuous variables, four contours are plotted by default. These contours are 100%, 75%, 50%, and 25% density contours. You can specify the number of contour levels to display. Alternatively, you can select a Bagplot or High Density Region (HDR) contours.
- For only one continuous variable, a *violin* plot appears instead of a contour plot. A violin plot illustrates the density of the data by plotting symmetric kernel densities around a common vertical axis. The kernel density estimates the probability density function at each point, providing a continuous analog of the histogram. The violin plot is similar to a box plot with symmetric kernel densities replacing the box and whiskers. Alternatively, you can select High Density Region (HDR) contours.
- If you add a Color variable to a contour plot, the plot shows *value contours* that reflect the levels of the Color variable. The value contours are computed using Delaunay triangulation. You can select an option (Transform) to show a plot where the X and Y ranges have been normalized.

Contour Options

Figure 3.29 Contour Options for a Contour Plot or a Violin Plot



Fill (Not available for Bagplots.) Fills in the contours.

Line (Not available for Bagplots.) Adds lines around the contours.

Number of Levels (Available for Nonpar Density.) For density contours, specifies the number of contours that appear. The number can be between 1 and 1000, the default is 4 contours.

Boundary (Available only when you have a Color variable.) Adds a line around the outside boundary of the contour.

Alpha (Available only when you have a Color variable.) Controls the hull of value contours. Increasing alpha can eliminate some of the long, skinny, or large triangles where interpolation might be undesirable.

Smoothness (Not available for Bagplots.) Smooths the boundaries of the contour plots. The smoothness value is normalized between -1 and 1. This value can be interpreted as a smoothing kernel radius. The original data are interpolated to a grid, and then a Gaussian smoother is applied.

Transform (Available only when you have a Color variable.) Transforms the triangulation to use a normalized scale for X and Y by selecting Range Normalized. This causes both the X and Y values to be scaled to [0,1] before computing the Delaunay triangulation. This option might be desirable in cases where the X and Y units are very different. In these cases, if points are stretched in one dimension and not the other, Delaunay triangulation tries to minimize long, skinny triangles, which can obscure some features.

Contour Type Enables selection of contour types

Nonpar Density (Available only for bivariate plots.) A smooth bivariate nonparametric density surface.

Bagplot Draws a Bagplot, also known as a bivariate box plot. A Bagplot consists of two polygons, a set of outlier points, and a median point. All computations are based on first computing the Tukey depth (bivariate depth) of each point in the data. The median point is the average of all points at maximum depth, which is plotted as an asterisk. The inner polygon is the *bag*, which contains at most 50% of the data points. Not shown in the plot is the *fence*, which is the bag polygon inflated three times relative to the median point. The outer polygon is the convex hull of all points contained within the fence. This is referred to as the *loop*. Points that lie outside the fence are designated outliers, and are shown as points on the plot. For more information about Bagplots, see Rousseeuw (1999).

HDR Draws highest density region rectangles for univariate and contours for bivariate data. The lighter shaded region represents the 99% probability region; the darker shaded region represents the 50% probability region. (Note that the regions can be noncontiguous and may not contain exactly 50% or 99% of the input data.) The density mode within the 50% probability region is represented by a line (univariate data) or an asterisk (bivariate data). If you remove points from the plot, the points that remain represent outliers relative to the 99% probability region. You can adjust the smoothness of the regions using the Smoothness option.

The HDR regions are based on a nonparametric density estimated by applying a Gaussian kernel to the data after the points have been interpolated to a grid. The default bandwidth for determining the smoothness of the HDR regions is computed as follows:

$$\frac{0.9s}{\sqrt[5]{n}}$$

The resulting equally spaced bins are used to compute the density regions. The contours are evenly spaced within the density regions that separate the space into equal-probability regions. For more information about highest density regions, see Hyndman (1996).

Violin (Available only for univariate plots.) Draws a violin plot of the density of the data by plotting symmetric kernel densities around a common vertical axis. The densities are defined using the same procedure as the HDR regions.

Jitter Turns jitter on or off. Jitter adds random noise to data, and can help reduce over-plotting. For descriptions of the specific jitter options, see "Points".

Outliers (Available for Bagplot and HDR contours.) Plots outliers.

Violin Scaling For multiple violin plots, choose a scaling option. Equal area or width means that the areas or maximum width of the violins are the same. Weighted area means that the areas of the violins are proportional to the number of observations in each violin.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Contour

Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

Contour Placement (Available with grouping variables.) Enables you to align or offset the contours.

Line

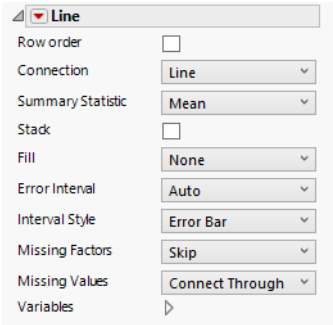
The **Line** element uses line segments to connect summary statistic values (or raw values) of the variable plotted on the Response Axis across the values or categories of another variable.

For an example that combines a line chart with an area plot to emphasize a line, see "Example of an Area and Line Chart".

Notes:

- If any of the rows used in calculating a point on the plot are hidden and not excluded, the point is not shown, but the position that it would occupy remains. The line segment connects the adjoining points.
- To label a line, right-click on the line element in the legend.

Figure 3.30 Line Options



Row order Connects points with line segments in the order of their row numbers.

Connection Changes the connection type to a line, curve, step, centered step, horizontal, or vertical line. If you select a curve, you can also change the smoothness.

Smoothness Changes the smoothness of a Curve connection. To use this option, select Curve for the Connection type.

Summary Statistic Specifies the statistic that is plotted. The statistic is computed for observations at each distinct value of the variable in the X zone (or the Y zone, if X is specified as the Response Axis).

Stack (Available for multiple Y variables.) Accumulates, or stacks, the Y values. Use to build an area chart.

Fill Specify how to fill the area below or between the lines.

Error Interval Adds or removes specified error interval in the graph.

Interval Style Draws error bars or error bands for the selected error interval.

Missing Factors Provides options for connections when all values are missing for a setting of the variable on the X axis:

Skip Skips over the missing factor value, connecting the previous factor to the next factor.

Treat as Missing Skips over the missing factor value, connecting the previous factor to the next factor. Draws the connection in the style specified for Missing Values.

Treat as zero Treats the missing factor values as zero.

Missing Values Provides options for connections when all values are missing for a setting of the variable on the Response axis:

Connect Through Connects the points adjacent to the missing values with the connector type used for the remaining points.

Connect Faded Connects the points adjacent to the missing values with a faded connection.

Connect Dashed Connects the points adjacent to the missing values with a dashed connection.

No Connection Shows no connection between the points adjacent to the missing values.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.
– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Line

Response Axis Specifies the axis for the variable that is used as the response in plotting the line segments. The Auto setting is the Y axis.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Bar

The **Bar** element shows a bar chart of summary statistic values of one or more variables across the values or categories of one or more stratifying variables. The graph can display a single variable or multiple variables of any type arranged in the X and Y zones.

- If you add a single categorical variable, then each bar shows the count of observations in the corresponding level of the variable.
- If you add two variables, with one in each zone, each bar represents the value of the specified summary statistic for the variable on the specified Response Axis. There is a bar for each value or level of the stratifying variable.
- If you add multiple variables in each zone, multiple bars are plotted for each value or combination of levels of the stratifying variables.

Response Axis Variables with Character Data Type

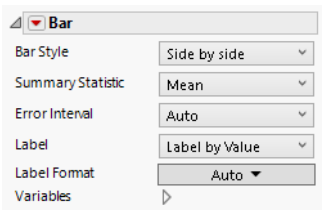
If the variable on the Response Axis has the Character data type, then this variable is assigned numeric values:

- Integer values are assigned based on the Value Order column property or the default value ordering.
- These integer values range from 0 (highest level in value ordering) to the number of levels minus one (lowest value in value ordering).
- These integer values are used in calculating the Summary Statistic.

Note: If any of the rows used in constructing a bar on the plot are hidden and not excluded, the bar is not shown, but the position that it would occupy remains.

For an example of a graph that uses the bar element, see "Example of an Overlaid Histogram and Ridgeline Chart" and "Example of a Packed Bar Chart".

Figure 3.31 Bar Options



Bar Style You can change the appearance of the bars that describe the levels of the stratifying variable. Many bar styles are applicable only when there are merged variables on the response axis, or if an overlay variable is present. For bar style descriptions, see Table 3.1. To change bar widths and space, right-click in the graph and select **Customize** > **Bar**.

Packed Bar Style Options

Packed Primaries Specifies the number of primary bars.

Packed Placement Specifies where the secondary bars are stacked. Options include the following:

First Stack—Start on the top row and fill it before moving to the next row.

Smallest Stack—Place each bar at the stack that is currently the smallest.

Separate Stack—Make a stack called Other where all secondary bars appear.

Packed Ordering Specifies the order in which secondary bars are placed. Can be by size. The biggest bars appear first or by label, which is in natural order and usually alphabetical.

Packed Coloring Specifies how the secondary bars are colored. Options include the following:

Bar color—Applies the same color as the primary bar color.

Faded bar color—Applies a faded version of the primary bar color.

Grays—Applies random shades of light gray.

Packed Labeling Controls the percentage of secondary bars that are labeled.

Packed Primary Labels Specifies where primary labels should be placed, either inside bars or on the axis.

Summary Statistic Specifies the statistic that is used as bar labels and controls the response axis scale. Available statistics depend on the type of bar chart used.

Error Interval Adds or removes error intervals in the graph when applicable, otherwise it is not available. To customize error bars, right-click in the graph and select **Customize** > **Error Bar**.

Label Adds or removes labels in the bar chart. You can label by value, percent of total value, or by row. When **Label by Value** is selected the values are determined by the summary statistic setting. Click and drag to reposition a label.

Label Format (Available for value and percent labels.) Enables you to set the format for the cell labels.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Bar

Response Axis Specifies the axis for the variable that is used as the response in plotting the bars. This option is available only when each axis contains at least one variable with a Continuous modeling type. If only one variable is Continuous, its axis is the response axis. The Auto setting is the Y axis.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Overlap Specifies the amount of overlap (Auto, None, Half, Full) between multiple bars.

Table 3.1 Bar Styles and Descriptions

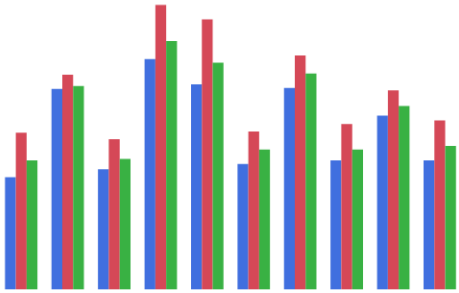
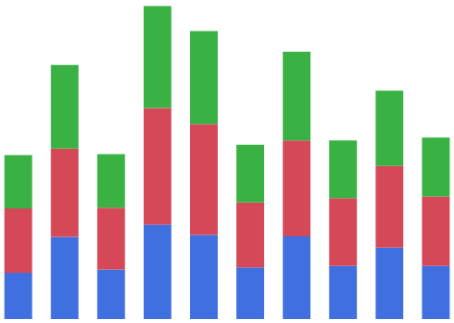
Bar Style	Description
Side by side	When you have two or more merged variables on the response axis, a bar is plotted for each of the merged variables. The bars are arranged next to each other for each level of the stratifying variables. They are colored and a legend appears. 
Stacked	When you have two or more merged variables on the response axis, a single bar is plotted for each level of the stratifying variables. Each bar stacks the bars for each of the merged variables. The part of the bar corresponding to each merged variable is colored and a legend appears. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

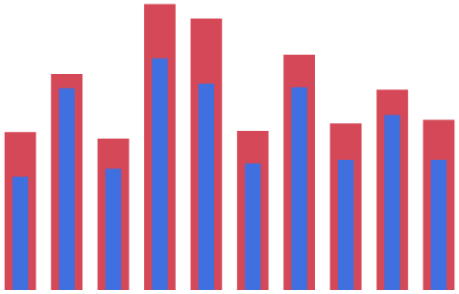
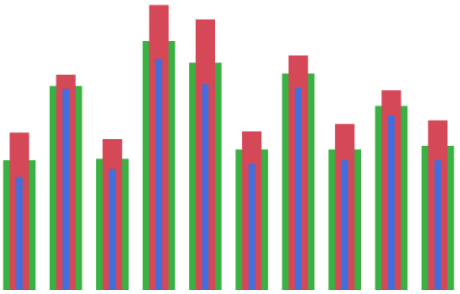
Bar Style	Description
Bullet	When you have two or more merged variables on the response axis, a narrow bar is plotted within a wider and taller bar. The widths of the bars depend on the order of the variables. The part of the bar corresponding to each merged variable is colored and a legend appears. If you have only a single variable, this style plots narrow bars. 
Nested	When you have two or more merged variables on the response axis, the bars for each variable are nested within each other. The narrowest bar corresponds to the first variable listed in the legend, the next narrowest to the second, and so on. **Tip:** You can change the legend order using the arrows in the Variables panel. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

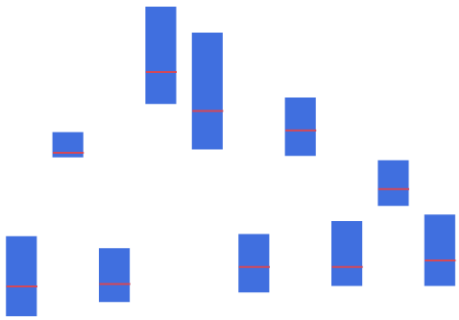
Bar Style	Description
Range	When you have two or more merged variables on the response axis, a rectangle and a line are plotted for each level of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A rectangle is plotted with bounds equal to the summary statistic values for the first two of variables, as indicated in the legend. • Lines are plotted at the values of the summary statistic for the other variables. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

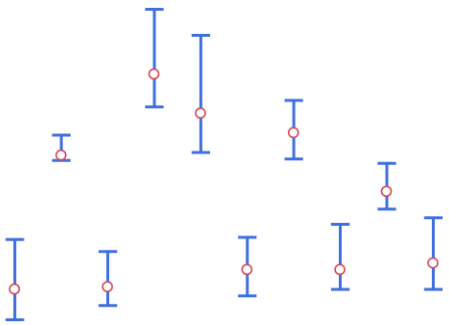
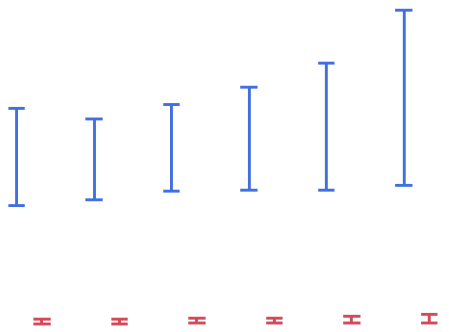
Bar Style	Description
Interval	When you have two or more merged variables on the response axis, a line with boundaries and a circle is plotted for each level of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A line with boundaries is plotted with bounds equal to the summary statistic values for the first two of variables, as indicated in the legend. • A circle is plotted at the overall values of the summary statistic for the other variables. 
Side by side intervals	In this style, each pair of variables defines an interval. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

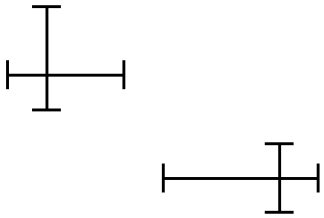
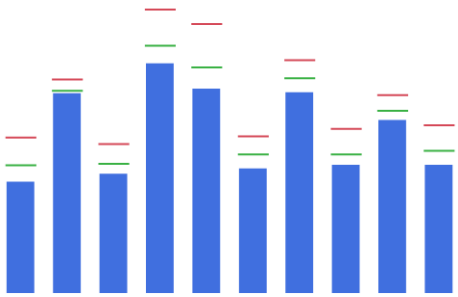
Bar Style	Description
Two-way interval	This style is for two-dimensional bar charts with bars in each direction. 
Single	For each level of the stratifying variables, a bar is plotted. When you have two or more merged variables on the response axis, the bar represents the value of the summary statistic for the first variable in the legend and lines are shown for the values of the summary statistic for the remaining variables. **Tip:** You can change the legend order using the arrows in the Variables panel. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

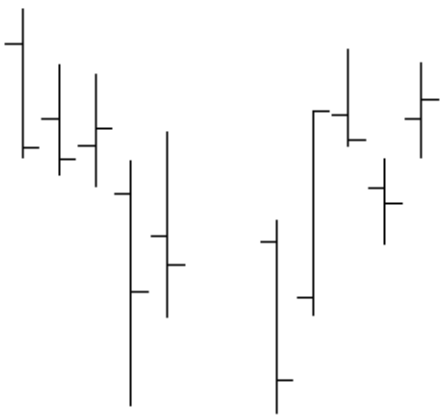
Bar Style	Description
Stock	When you have two or more merged variables on the response axis, line segments are plotted at the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A line segment connecting the summary statistic values for two of the variables, as indicated in the legend. • Bars are plotted at right angles to the line segment at the remaining values of the summary statistic. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

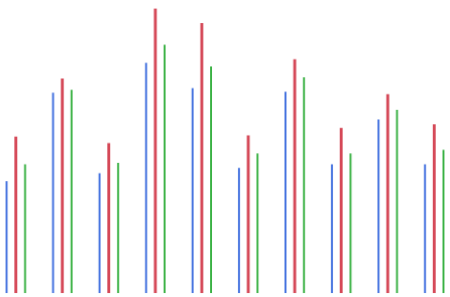
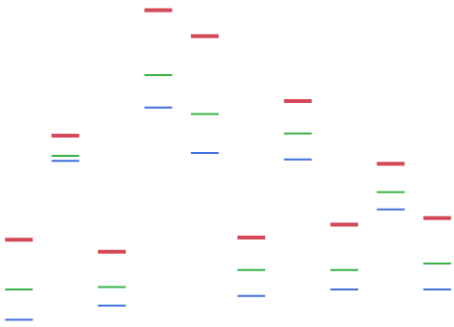
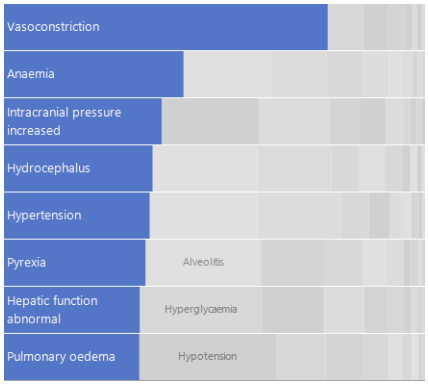
Bar Style	Description
Box Plot	When you have two or more merged variables on the response axis, box plots are shown for the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • An outlier box plot is constructed for these values.
Needle	When you have two or more merged variables on the response axis, needle-shaped bars are plotted side-by-side for each of the merged variables. The bars are colored and a legend appears. This style is useful when you have many levels of the stratifying variables. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

Bar Style	Description
Float	When you have two or more merged variables on the response axis, bars are plotted at the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A bar is plotted at each of these values. This style is useful when you have many merged variables. 
Packed	This bar style shows the top categories as a bar chart and stacks the other categories as bars with an approximately rectangular fill. The top categories are the focus and the other bars provide context. This style is useful when you have many categories and skewed response data. 

Area

Note: You can also create an area plot using the Line element .

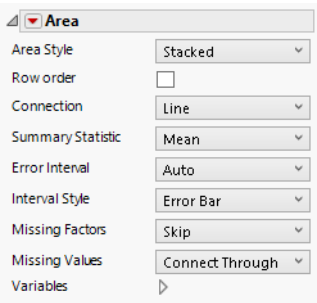
The **Area** element shows a contiguous area that represents the area beneath summary statistic values of a variable across the values or categories of another variable. The Area element can be considered a continuous representation of the Bar element.

- Both variables might be nominal or ordinal. The summary statistic represents the count of observations in a level of the variable on the axis that is not specified as the Response Axis.
- One of the variables might be continuous and the other nominal or ordinal. The summary statistic represents the value of the specified summary statistic for the continuous variable for the corresponding level of the categorical variable.
- If both variables are continuous, the summary statistic is computed for the continuous variable on the axis that is specified as the Response Axis.

Note: The plot is not affected in any way if rows are hidden and not excluded.

For an example of a graph that uses the Area element, see "Example of an Area and Line Chart".

Figure 3.32 Area Options



Area Style Changes how the area is displayed when there are merged variables on the response axis. For area style descriptions, see Table 3.2.

Row order Connects points with line segments in the order of their row numbers.

Connection Changes the connection type to a line, curve, step, centered step, horizontal, or vertical line. If you select a curve, you can also change the smoothness.

Summary Statistic Specifies the statistic that is plotted. The statistic is computed for observations at each distinct value of the variable in the X zone (or the Y zone, if X is specified as the Response Axis).

Error Interval Adds or removes specified error interval in the graph.

Interval Style Draws error bars or error bands for the selected error interval.

Missing Factors Provides options for connections when all values are missing for a setting of the variable on the X axis:

Skip Skips over the missing factor value, connecting the previous factor to the next factor.

Treat as Missing Skips over the missing factor value, connecting the previous factor to the next factor. Draws the connection in the style specified for Missing Values.

Treat as zero Treats the missing factor values as zero.

Missing Values Provides options for connections when all values are missing for a setting of the variable on the Response axis:

Connect Through Connects the points adjacent to the missing values with the connector type used for the remaining points.

Connect Faded Connects the points adjacent to the missing values with a faded connection.

Connect Dashed Connects the points adjacent to the missing values with a dashed connection.

No Connection Shows no connection between the points adjacent to the missing values.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Area

Response Axis Specifies the axis for the variable that is used as the response in plotting the areas. Available only when both variables have a Continuous modeling type. If only one variable is Continuous, its axis is the response axis. The Auto setting is the Y axis.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Table 3.2 Area Styles and Descriptions

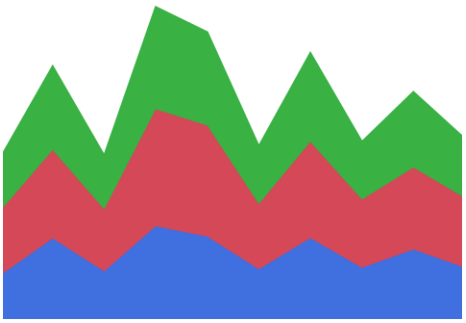
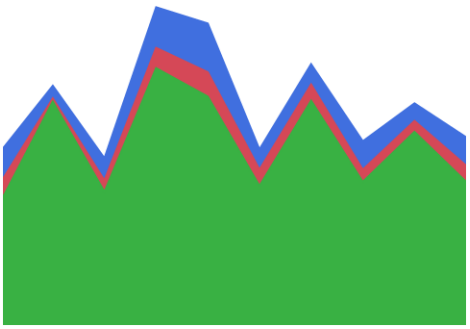
Area Style	Description
Stacked	Shows areas that correspond to summary statistics for each of the merged variables. The plots are colored and a legend appears in the Legend zone. 
Overlaid	Shows the area plots for the merged variables superimposed over each other. The plots are colored and a legend appears in the Legend zone. 

Table 3.2 Area Styles and Descriptions *(Continued)*

Area Style	Description
Range	Shows the area corresponding to the difference in the summary statistic for the merged variables.
Stacked Range	Shows the areas corresponding to the difference in the summary statistic after the variables are stacked.

Box Plot

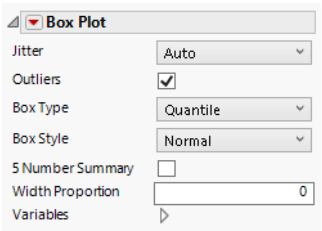
The **Box Plot** element shows outlier or quantile box plots. A box plot provides a compact view of a distribution of values. The box extends from the 25th percentile to the 75th percentile where the distance between the 75th and 25th percentiles is the interquartile range (IQR). The median is marked within the box. For outlier box plots, whiskers extend to the last point that is within 1.5*IQR from the ends of the box. For quantile box plots the whiskers cover the full range of observations with tick marks at specific quantiles. For more information about outlier and quantile box plots, see *Basic Analysis*.

- If one of the variables is continuous and the other is nominal or ordinal, box plots are constructed for values of the continuous variable. There is a box plot for each level of the categorical variable.

- If both variables are continuous, box plots are constructed for values of the continuous variable on the axis specified as the Response Axis. There is a box plot for each level of the other continuous variable.

For an example of a graph using box plots, see "Example of Features in Graph Builder".

Figure 3.33 Box Plot Options



Jitter Turns jitter on or off. Jitter adds random noise to data, and can help reduce over-plotting.

Outliers Shows or hides values that extend beyond the whiskers.

Box Type Specifies whether the box plot is an outlier box plot or a quantile box plot.

Box Style Changes the style the box plot. For box style descriptions, see Table 3.3.

5 Number Summary Adds statistics like the median, maximum, minimum, and quantiles 1 and 3 to the graph.

Width Proportion adjusts the width of the box plots.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Box Plots

Confidence Diamond Shows or hides a confidence diamond about the mean. The diamond covers the 95th confidence interval for the mean.

Notched Displays a box plot that is notched at the median where the notches span:

$$\text{median} \pm (1.58 \times \text{IQR}) / (\sqrt{n})$$

Fences Shows or hides vertical lines, or fences, at the end of the box plot whiskers.

Shortest Half Shows or hides a bracket that indicates the location of the shortest interval that contains 50% of the data.

Shortest Half Color Enables you to select a color for the shortest half bracket.

Response Axis (Available only when the X and Y axes variables both have a continuous modeling type.) Specifies the axis for the variable that is used as the response in constructing the box plots. Available only when both variables have a continuous modeling type. If only one variable is continuous, its axis is the response axis. The Auto setting is the Y axis.

Box Placement (Available with grouping variables.) Enables you to align or offset the box plots.

Tip: When you have a categorical grouping variable for your box plot, right-click on the categorical axis and select **Size By > Count** to draw box plots that are proportional to the number of observations in each category.

Table 3.3 Box Styles and Descriptions

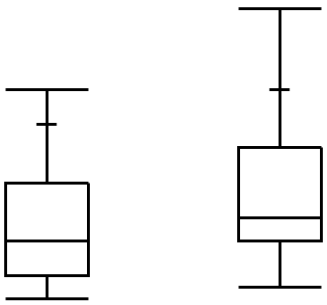
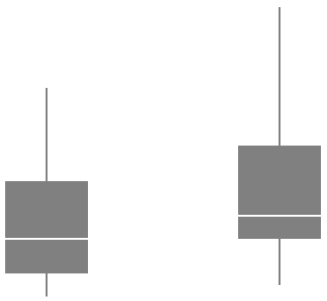
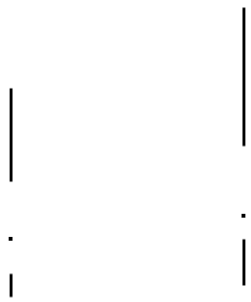
Box Style	Description
Normal	Shows the box defined by the quartiles as hollow with only a line at the median. 
Solid	Shows the box defined by the quartiles with solid fill. The line at the median is shown by white space. 

Table 3.3 Box Styles and Descriptions *(Continued)*

Box Style	Description
Thin	The box defined by the quartiles is not shown. It is discernible due to the whiskers. The median is denoted by a point.



Histogram

The **Histogram** element shows a variable's distribution using binning. For more information about histograms, see *Basic Analysis*.

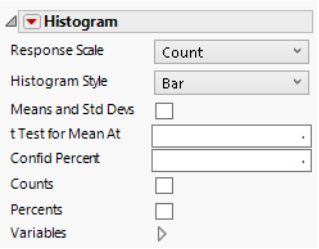
Note: The bars of the histogram align with the tick marks on the axis for the zone containing the variable. If you customize the axis, the histogram adjusts to match the new tick marks.

You can change the primary direction of the graph to X (horizontal), Y (vertical), or Auto using the Response Axis option.

- If both variables are nominal or ordinal, the histogram shows the count of observations in a level of the variable on the axis that is not specified as the Response Axis.
- If one of the variables is continuous and the other is nominal or ordinal, the plot shows a histogram for the continuous variable for each level of the categorical variable.
- If the same variable is specified for both *X* and *Y*, then the *Y* role is ignored and a single histogram appears.
- If both variables are continuous, then the variable on the X axis is treated as discrete. You might need to rescale the X axis to view the graph clearly. For an alternative visual, use a scatterplot with the Contour element.

Tip: You can overlay histograms with transparent color so that you can see the individual histograms. To do this, assign the variable of interest as X or Y. Assign your overlay variable to the **Overlay** zone. Then, click the **Histogram** element icon. See "Example of an Overlaid Histogram and Ridgeline Chart".

Figure 3.34 Histogram Options



Response Scale Specifies the scale for the response axis. The options are count, percent, or fill. Fill has no response axis label, fills the display space, and scales multiple histograms independently.

Overlap (Available with a categorical response.) Adjusts the amount of histogram overlap between vertical categories.

Histogram Style Specifies the histogram style.

Bar Default traditional histogram bars with a height based on the number or percent of observations that fall within each bar or bin.

Polygon Connects the peaks of each histogram bar to construct a polygon representation of the distribution of the data.

Kernel Density Density curve with a smoother control.

Shadowgram Overlays histograms with different bin widths.

Means and Std Devs Shows the means and standard deviations for the levels of the variables in the X or Y zone.

t Test for Mean At Performs a *t* test for the specified mean.

Confid Percent Specifies the confidence interval for the mean.

Counts Displays counts on the histogram bars.

Percents Displays percents on the histogram bars.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option for Histogram

Response Axis Specifies the axis for the variable that is used as the response in constructing the histograms. Available only when both variables have a Continuous modeling type. If only one variable is Continuous, the other axis is the response axis.

Heatmap

The **Heatmap** element shows the counts or average values for groups of observations using a color intensity scale. If you plot a single variable as either Y or X, your plot appears as bars. If you plot a cross tabulation of two variables Y and X, your plot shows rectangles.

- For a categorical variable, the levels define the groups of observations.
- For a continuous variable, Graph Builder constructs non-overlapping intervals of values. These intervals define the groups of observations.

Note: If all of the rows used in constructing a bar or rectangle on the plot are hidden and not excluded, the rectangle is not shown, but the position that it would occupy remains.

When you create a heatmap, an intensity legend appears in the Legend area. See "Discrete or Continuous Legend Items". The intensity legend shows counts or average values:

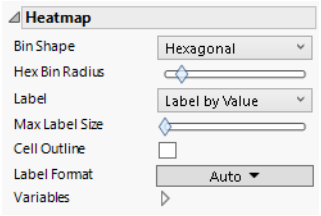
- If no Color variable is assigned, the counts in each bar or rectangle define the scale for the intensity legend.

- If a Color variable is assigned, the average of that variable for the observations in each bar or rectangle define the scale for the intensity legend.

Tip: Hover over a cell to see labels. Click a cell to select the corresponding rows.

To color or size a heatmap by two variables, drag the first variable into the Color or Size zone and then drag the second variable into the top or bottom corner of the zone.

Figure 3.35 Heatmap Options



Bin Shape (Available for continuous variables). Specifies the shape of the bin.

Hex Bin Radius (Available for hexagonal bins.) Enables you to adjust the radius of the bin.

Label Specifies labels for the heatmap. You can label by value, percent of total, or by row.

Max Label Size Enables you to adjust the size of the label.

Cell Outline (Available when a size variable is used.) Enables you to add outlines for the original bin sizes. This highlights the size scaling on the bins.

Label Format (Available for value and percent labels.) Enables you to set the format for the cell labels.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Use a Heatmap to Apply Background Colors

You can also use the heatmap to control the background color when you have plot elements that use the Wrap, Group X, or Group Y zones. To do so:

1. Select the variable for the background color and drag it to the **Color** zone.
2. Select the Heatmap element .
3. Construct the plot elements.

In the Heatmap panel, under Variables, deselect all check boxes except the one for Color.

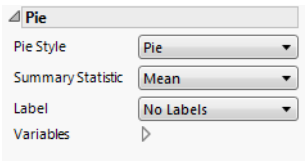
Pie

The **Pie** element plots summary statistics for groups of observations, representing their values as the size of the slices or rings on a pie chart.

Note: If any of the rows used in constructing a slice of the pie chart are hidden and not excluded, the slice is not shown, but the position that it would occupy remains.

- For a single variable, the slice sizes the number of observations in each category.
- For two variables, the variable in the Y zone is used to size the slices according to the selected Summary Statistic. A legend appears in the Legend zone showing the values of the variable in the X zone.

Figure 3.36 Pie Options



When you add an Overlay variable, a ring chart is constructed for each level of the Overlay variable. The outer ring represents the smallest value in the value ordering, and the inner ring represents the largest value.

Pie Style Changes the appearance of the pie chart. For pie style descriptions, see Table 3.4.

Summary Statistic Changes the statistic that is plotted.

Label Adds or removes labels in the pie chart.

Label Format (Available for value and percent labels.) Enables you to set the format for the cell labels.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Table 3.4 Pie Style Descriptions

Pie Style	Description
Pie	Traditional pie chart with each slice sized by the Summary Statistic.

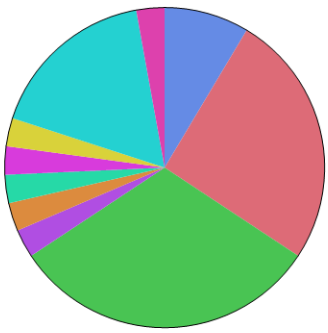
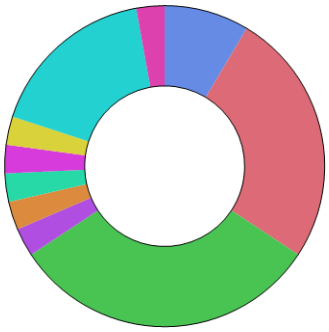
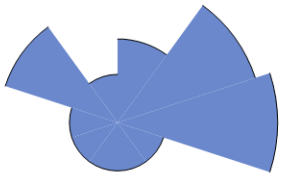


Table 3.4 Pie Style Descriptions *(Continued)*

Pie Style	Description
Ring	Each variable or level of a stratifying variable is represented by a concentric ring. The sections are sized by the Summary Statistic. Ring charts can help you visualize hierarchical data, using concentric rings. 
Coxcomb	The central angles for all slices are equal. The size of each slice is determined by the Summary Statistic. Also known as an *exploded pie chart*, a coxcomb chart can help you see smaller areas. 

Treemap

The **Treemap** element shows the levels or values of a variable as rectangles within a rectangular display. The rectangle sizes represent summary statistic values of the variable in the **Size** zone across the levels or values of one or more X variables.

To construct a treemap that shows summary values across the levels of two or more X variables, merge the variables in the X zone. The rectangles are sized to represent the values of the summary statistic without leaving unused space in the overall rectangular layout.

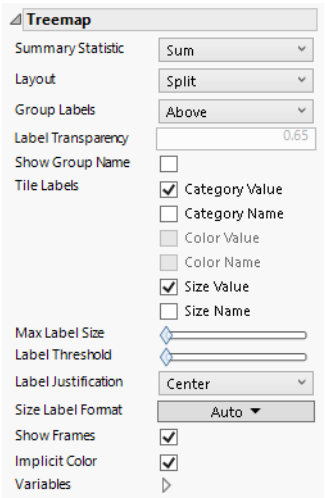
By default, the data table order is used when ordering the tiles in the treemap using the Split layout. Alternatively, you can specify one or two continuous variables as Y variables for ordering:

- If one Y variable is specified, it orders both the X and Y axes.
- If two Y variables are specified, the first orders the X axis and the second orders the Y axis.

If the Squarify layout is specified, then the order of the tiles is determined by the size, and ordering variables are ignored.

Note: The plot is not affected in any way if rows are hidden and not excluded.

Figure 3.37 Treemap Options



Summary Statistic Specifies the summary statistic, corresponding to the variable in the Y zone, that is used to size the rectangles.

Layout Arranges rectangles to the extent possible by the value ordering of the X variable or by the size of the rectangle.

Split Arranges the rectangles according to the value ordering of the levels or values of the X variable. Split is the default setting.

Squarify Arranges the rectangles according to the values of the summary statistic, sorted in descending order. This places the largest rectangles in the top left corner of the plot and the smallest in the bottom right corner.

Mixed Applies only when you have two or more X variables. It applies Split to the outermost variable and Squarify to the other variables. It follows that the large groupings are ordered according to their value ordering, whereas the inscribed rectangles are ordered according to the summary statistic values.

Group Labels (Available when you have more than one X variable.) Shows or hides the additional group labels above each category or floating in the center of each category.

Tip: To filter data in a group, click a group label.

Label Transparency (Available with floating group labels.) Specifies the transparency of the floating group labels.

Show Group Name Adds the variable title to the group labels.

Tile Labels Shows or hides the following labels:

Category Value The values of the X variable.

Category Name The name of the X variable.

Color Value The values of the Color variable.

Color Name The name of the Color variable.

Size Value The values of the Size variable, or the Y variable if no Size variable exists (Size Value).

Size Name The name of the Size variable, or the Y variable if no Size variable exists (Size Name).

Tip: Hover over a rectangle to see details.

Max Label Size Increases or decreases the size of the labels.

Label Threshold Removes labels based on the size of the rectangle. By default, all labels are shown. Move the slider to the right to remove labels for progressively larger rectangles.

Label Justification Places the label in the center, left, or right.

Color Label Format (Available when a color variable is used.) Enables you to set the format for the cell labels based on the color variable.

Size Label Format (Available when Size Value is selected.) Enables you to set the format for the cell labels based on the size value variable.

Show Frames Shows or hides the borders around the rectangles.

Tip: Use the treemap option in the right-click customize menu to adjust frame color, style, and thickness.

Implicit Color Shows or hides the color. If no variable is specified in the Color zone, all rectangles have the same color.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

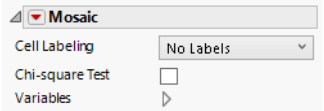
Mosaic

The **Mosaic** element uses rectangle sizes to represent the proportion of observations in categories for one variable across categories of another variable. Although variables can be categorical or continuous, they are treated as categorical. For more information about mosaic plots, see *Basic Analysis*.

The default Response Axis is the Y axis. This means that, for each level of the X variable, vertical stacked bars show the proportion of values in each level of the Y variable. You can change the primary direction of the graph to X (horizontal).

Note: The plot is not affected in any way if rows are hidden and not excluded.

Figure 3.38 Mosaic Options



Cell Labeling Add labels to cells based on counts, percents, values, or row.

Label Format (Available for percent or value labels.) Enables you to set the format for the cell labels.

Chi-square Test Performs a Pearson chi-square test and shows results in a label on the graph. This option appears if each side has 2 or more levels.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option for Mosaic

Response Axis Changes the axis that is associated with the response variable to X (horizontal), Y (vertical), or Auto (the Y axis for Mosaic). The variable on the Response Axis is used to calculate the proportion of observations in each level of the variable on the other axis.

Caption Box

The **Caption Box** element displays summary statistics on the plot. The summary statistic corresponds to the variable that corresponds to the axis that you specify as the Response Axis in the Caption Box options.

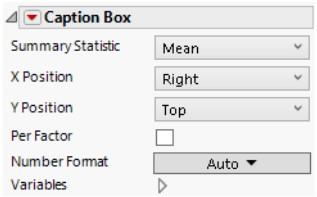
Summary Statistic for Categorical Variables

For a categorical variable, the Summary Statistic that appears is calculated by assigning numerical values to the levels of the categorical variable. The levels are placed in the order defined by their value ordering. They are assigned integer values ranging from 0 to the number of levels minus one.

Note: When using categorical values in zones associated with summarization roles like Y and Color, summary statistics are calculated based on numerical values assigned to the variable levels. (The integer values range from 0 to the number of levels minus one.) The final result is then mapped back as a level to a corresponding categorical value or value range. This approach allows the use of ordinal variables in summarization, but might not lead to clear results for nominal variables.

For an example illustrating the Caption Box element, see "Example of Features in Graph Builder".

Figure 3.39 Caption Box Options



Summary Statistic Specifies the summary statistic to be displayed for the variable defined by the Response Axis setting. The default Summary Statistic is the Mean. For more information about how summary statistics are computed for categorical variables, see "Summary Statistic for Categorical Variables".

X Position Specifies the horizontal position of the caption.

Y Position Specifies the vertical position of the caption.

Per Factor Shows a caption for each X, or each Y if horizontal.

Number Format Enables you to set the format of the summary statistic.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option for Caption Box

Response Axis Specifies the axis for the variable whose summary statistic is displayed. The Y axis is the default. Notice that this axis can differ from the Response Axis specified to define the plot.

Formula

The **Formula** element shows the graphs of a function, its inverse, or a parametric curve. The function or functions are defined by column formulas.

- $y = f(x)$:
  - To plot a function of a single argument across the values of its single argument, drag the column that contains the formula to the Y zone and the column that contains the values of its single argument to the X zone.
  - Make sure that the Response Axis is set to Auto or Y.
- $y = f^{-1}(x)$:
  - To plot the inverse of a function of a single argument for a column of values, drag the column that contains the formula to the X zone. Drag the column that contains the values of interest to the Y zone.
  - Set the Response Axis to X.
- $x = f(t)$ and $y = g(t)$: To plot the parametric curve defined by two functions, drag one to the X zone and the other to the Y zone.

Note: Because the Formula element plots the values of a function, the plot is not affected in any way if rows are hidden and not excluded.

Figure 3.40 Formula Option

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option for Formula

Response Axis Specifies the axis for the formula variable.

Map Shapes

The Map Shapes element shows areas defined by the variable in the Map Shape zone.

When a column in your data table contains the names of geographical regions (such as countries, states, provinces, counties), you can assign the column to the Map Shape zone. When a variable is dropped in the Map Shape zone, Graph Builder searches for internal maps that correspond to the values in the column. If it finds an appropriate map, it draws the map in the graph area.

Alternatively, you can define the Map Role column property for the column of interest. This property tells JMP how to connect the values in the column with map shape data. It is especially useful when you create custom maps. See "Custom Map Files".

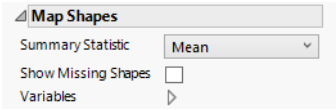
If you have a variable in the Map Shape zone, the X and Y zones disappear. You can add information from other variables using Color and Size.

- To color the map shapes by the values of a summary statistic, drag the column of interest to the **Color** zone. The categorical or continuous color theme selected in your Preferences is applied to each shape.

- To size the map shapes by the values of a summary statistic, drag the column of interest to the **Size** zone. This scales the map shapes according to the summary statistic value of the size variable, minimizing distortion.

See "Maps". For examples, see "Examples of Creating Maps", or run the associated scripts in these sample data tables: PopulationByMSA.jmp or SAT.jmp.

Figure 3.41 Map Shapes Options



Summary Statistic Changes the statistic that is plotted as the Color or Size variable.

Show Missing Shapes Shows or hides missing data from the map.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Parallel

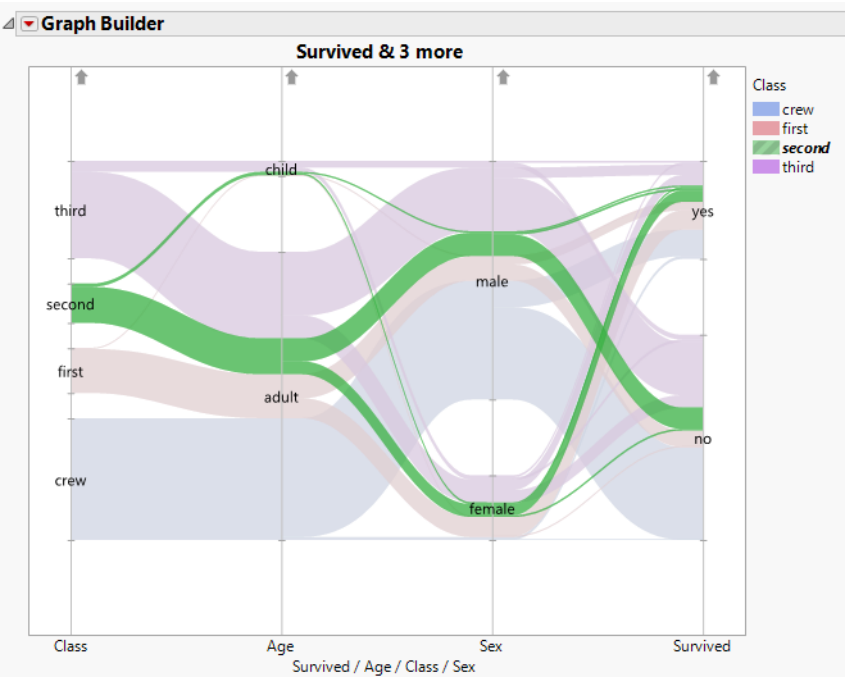
The **Parallel** element connects the values in a row across two or more variables. Drag two or more variables together to either the X or Y zone. The variable names appear as axis labels in the zone to which they were dragged.

- Values for each continuous variable are plotted along lines (axes) parallel to the other axis.

- Levels of categorical variables are represented by intervals on lines (axes) parallel to the other axis. The sizes of the intervals are proportional to the number of observations in each level of the categorical variable.
- The values (for continuous variables) or bands (for categorical variables) are joined with continuous curves.

When all variables are categorical, there is a band for every combination of levels of the categorical variables. The bands split as they move from left to right. In the first interval corresponding to a categorical variable, there is a band for every level of that categorical variable. In the last interval corresponding to categorical variables, there is a band for every possible combination of the categorical variables.

Figure 3.42 Example of Categorical Bands Using Titanic.jmp



In Figure 3.42, the band containing all second-class passengers is selected. The parallel plot shows that most were adults, there were more males than females, and slightly fewer survived than did not survive.

The values or levels of the variables are connected with lines or curves using the Curve Lines option.

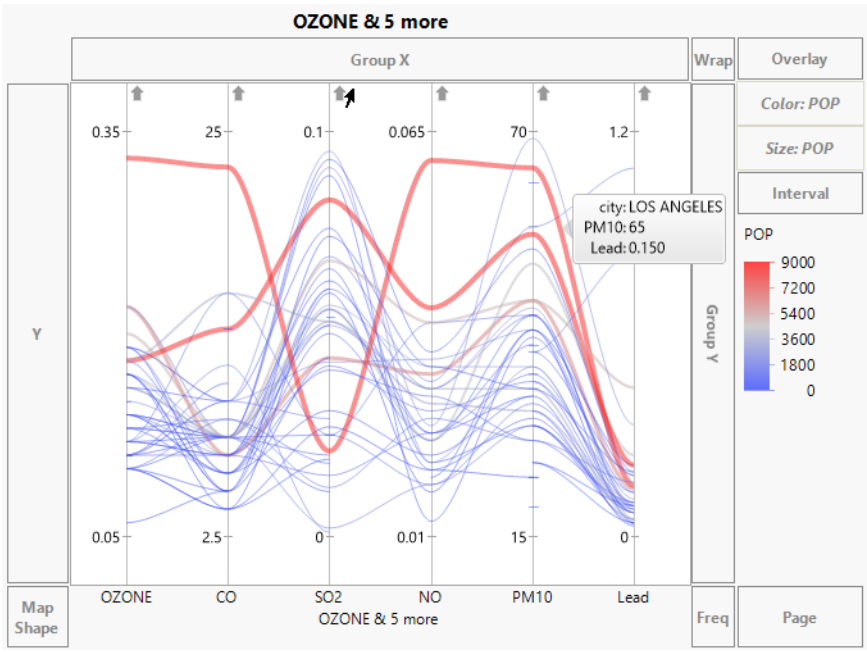
- To color the curves by the values of a variable, drag the column of interest to the **Color** zone. The categorical or continuous color theme selected in your Preferences is shown in the Legend zone.

- To size the curves by the values of a variable, drag the column of interest to the **Size** zone.
- To change the direction of a variable's axis, click the arrow at the top of the vertical line denoting the variable's axis.
- To move a variable and its axis, click and drag the axis.

Tip: Hover over a curve to see a label giving information about the corresponding row.

Figure 3.43 shows a parallel plot for six variables in the Cities.jmp data table. The variable POP is used both as a Color and Size variable. The curve for Los Angeles is labeled.

Figure 3.43 Parallel Plot for Pollution Data in Cities.jmp



Tips:

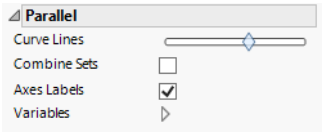
- You can add a line element to a parallel plot to show a summary relationship across the parallel plot.
- You can add reference lines for specification limits. For information, see *Using JMP*.

Scaling

By default, the scales for the values of the variables are adjusted so that the minimum and maximum values are plotted at the same level. For example, in Figure 3.43, the values of each of the variables have an identical vertical spread. Each vertical line is labeled by the minimum and maximum values of the variables.

In Figure 3.43, the scales for CO and PM10 differ greatly from the scales of the other variables. When your variables are measured on very different scales, this scaling enables you to see differences clearly.

Figure 3.44 Parallel Options



Curve Lines Adjusts the amount of curvature of the curves that connect points. Place the slider all the way to the left for lines. As you move the slider to the right, the degree of curvature increases.

Combine Sets (Applicable only if you have three or more categorical factors.) For categorical variables, when Combine Sets is not selected, bands split as they move from left to right. In the last interval that corresponds to a categorical variable, there is a band for every possible combination of the categorical variables. Selecting the Combine Sets option causes the bands not to split. Each interval that follows a categorical variable shows a band for each level of that variable.

Axes Labels Removes the axis labels from the display.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

For an example using Variables, see "Example of an Area and Line Chart".

Graph Zones

The main element in the Graph Builder window is the graph area. The graph area contains zones, and you drag and drop variables from the Variables box into the zones.

Tip: If you drop variables into the center of the graph, JMP guesses which drop zone to put them into, based on whether the variables are continuous, ordinal, or nominal.

There are two types of zones into which you can drag variables:

- *Data zones* include **X**, **Y**, **Map Shape**, **Freq**, **Color**, **Size**, and **Interval**. The X, Y, and Map Shape zones are positional, and influence the types of graph elements that are available. The Freq, Color, Size, and Interval zones modify certain graph elements.
- *Grouping zones* partition the data into subsets that repeat the graph for each possible subset. Grouping zones include **Group X**, **Group Y**, **Wrap**, **Overlay**, and **Page**.

The following zones are available in Graph Builder:

Table 3.5 Graph Builder Zones

Zone	Description
X and Y	Constructs plots that are based on the values, levels, or categories of the variables that you drop into these zones. Use the X zone for a horizontal layout and the Y zone for a vertical layout.

Table 3.5 Graph Builder Zones *(Continued)*

Group X and Group Y	Stratifies your data by the levels of the nominal or ordinal variable and constructs a plot for each level. Group X produces a horizontal display, and Group Y produces a vertical display. The type of variable that you put in Group X or Y also determines aspects of the display: • Adding a continuous variable creates non-overlapping interval subgroups and plots for the values in each interval. • Adding a categorical variable creates plots for each level of the categorical variable. You can add multiple variables to Group X or Group Y: • To add a variable *above* a variable in one of the Group zones, drag it to the left of that variable in the Group X zone or above that variable in the Group Y zone. • To add a variable *below* a variable in one of the zones, drag it to the right of that variable in the Group X zone or below that variable in the Group Y zone. To stratify by the levels of two variables, drag a variable into both Group X and Group Y. To change the number of levels that appear, right-click in the axis border and select Levels in View. For a description of all options, see "Options for Axes, Axes Variable Labels, or Graph Titles".

Table 3.5 Graph Builder Zones *(Continued)*

Map Shape	Adds a map. Your variable must contain geographic place names, such as countries, regions, states, or provinces, or you must provide information in a custom shape file. Note the following: • If you drag and drop columns labeled Latitude and Longitude, JMP automatically assigns them to the Y and X roles (respectively). • When you drop a variable into the Map Shape zone, the X and Y zones disappear. • The Map Shape element can be used with the Points element to place points within the shape. For more information about creating map shapes, see "Red Triangle Option for Formula" and "Graph Builder".
Wrap	Creates rows of plots for the levels of a categorical variable or for the interval subgroups of a continuous variable. Use Wrap when your stratifying variable has many levels. To change the number of levels that appear, right-click in the axis border and select **Levels in View**. For a description of all options, see "Grouping Zones Options". **Note:** Once a variable is placed here, no variable can be placed in Group X.
Overlay	Subgroups and colors the plot elements according to the levels of the Overlay variable. • Adding a nominal or ordinal variable stratifies and colors the plot elements by the levels of the variable. • Adding a continuous variable creates interval subgroups. The plot elements are stratified and colored according to the subgroups. A legend appears to the right of the plot. Right-click on legend items for customization options.

Table 3.5 Graph Builder Zones *(Continued)*

Color	Adds color to points, map shapes, and other objects. • Adding a nominal or ordinal variable colors the plot elements by the variable's levels. • Adding a continuous variable colors the plot elements according to an intensity scale. A legend appears to the right of the plot. Right-click in the legend to change colors or to customize the intensity scale. **Tip:** Once you have added a variable in the Color zone, you can show or hide color using the Variables option in the element properties panel.
Size	Sizes graph elements by a summarizing statistic or other size variable. A legend appears to the right of the plot. Right-click on legend items for customization options.
Interval	Add interval variables to draw custom error bars. • A single interval variable is treated as a delta to compute upper and lower ends of the interval. • Two interval variables are treated as lower and upper values for the interval.
Freq	Add a frequency or weight variable. When all values are integer values, the Freq variable is treated as a frequency, otherwise it is treated as a weight. A footer note indicates how the variable is treated. The Freq variable affects summary statistics.

Table 3.5 Graph Builder Zones *(Continued)*

Page	Drop a By group variable to the Page zone to show each level of the group on a separate graph. By default, the Page zone is limited to 200 pages. The first 200 categories are displayed. Adjust this limit in the Graph Builder preferences. See *Using JMP*. **Tips:** • Once you click Page, you can select the red triangle menu option Link Page Axes. Use this option to link or unlink graph axis scales across levels of the By group variable in the Page zone. • To copy individual graphs using the selection tool, first save the Graph Builder output to a journal (Edit > Journal).

For more information about the legend, see "Legend Options".

Note: When you use categorical values in zones associated with summarization roles like Y and Color, summary statistics are calculated based on numerical values assigned to the variable levels (integer values ranging from 0 to the number of levels minus one). The final result is then mapped back as a level to a corresponding categorical value or value range. This approach allows the use of ordinal variables in summarization, but might not lead to clear results for nominal variables.

Categorical and Continuous Variables in Group X or Y Zones

When a categorical variable is added to the Group X or Group Y zone, a partition is created for each level of the variable.

When a continuous variable is added to a grouping zone, Graph Builder uses quantiles of the data to divide the variable into five groups. To change the number of groups:

1. Right-click the grouping variable label and select **Number of Levels**.
2. Type in the number of levels that you want to display.
3. Click **OK**.

Move Group X and Group Y Labels

Grouping variable labels can be relocated to another position on the graph. The Group X labels can be either on the top or the bottom of the graph. The Group Y labels can be either on the right or the left of the graph.

- To relocate a Group X label, right-click the variable in the Group X zone and select **X Group Edge > Top** or **Bottom**.
- To relocate a Group Y label, right-click the variable in the Group Y zone and select **Y Group Edge > Left** or **Right**.

Red Triangle Options

The red triangle menu for Graph Builder contains these options:

Show Control Panel Shows or hides the platform buttons, the Select Columns box, and the drop zone borders.

Show

Title Shows or hides the graph title. Right-click to change the alignment or the span of the title.

Subtitle Shows or hides the graph subtitle. Right-click to change the alignment or the span of the subtitle.

Legend Shows or hides the legend.

Footer Shows or hides the footer, which contains informative messages such as missing map shapes, error bar notes, frequency notes, and WHERE clauses.

X Axis Shows or hides the X axis.

Y Axis Shows or hides the Y axis.

X Axis Title Shows or hides the X-axis title.

Y Axis Title Shows or hides the Y-axis title.

Legend Position Sets the position of the legend. The legend appears on the right by default. Putting the legend at the bottom places it in the center below the graph. The legend items then appear horizontally instead of vertically. Or, you can put the legend inside the graph, on the left or right.

Legend Settings Opens a window where you can modify legend settings, such as the title, font, and title position.

Color Settings Contains the following options:

Continuous Color Theme Selects the color theme for continuous variables.

Categorical Color Theme Selects the color theme for categorical variables.

Note: For more information about color themes, see *Using JMP*.

Lighten large fills Automatically lightens colors for elements that fill large areas, such as pie, treemap, and mosaic plots.

Use row colors for levels When every level has a unique color, initialize legend levels with row colors.

Lock Scales Prevents axis scales and gradient legend scales from automatically adjusting in response to data or filtering changes.

Error Bar Offset Adjust the offset for overlaid error bars.

Link Page Axes Links or unlinks graph axis scales across levels of the By group variable in the Page zone.

Fit to Window Specifies whether the graph is resized as you resize the JMP window. The default setting is Auto, which is usually set to On except when the window is shared with other content or if there is a Page variable. To always fit the graph inside the window, keep the setting at On. You can also maintain the aspect ratio. To prevent the graph from resizing, change the setting to Off.

Sampling Uses a random sample of the data to speed up graph drawing. If the sample size is zero, or greater than or equal to the number of rows in the data table, then sampling is turned off.

Graph Spacing Sets the amount of space between graph panels.

Include Missing Categories Enables a graph to collect and display missing values for categorical variables.

Set Alpha Level Enables you to set the Alpha level for graphs and statistics that use an alpha value for confidence intervals.

Launch Analysis Launches the Fit Model platform with the variables on the graph placed into roles. It launches the Distribution platform when only one variable is placed.

Make into Data Table Creates a new data table that contains the results from the graph.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Graph Builder Buttons

The Graph Builder window contains these buttons:

Recall Populates the Graph Builder window with the last graph that you created. This button changes into the Undo button once you perform an action.

Dialog Opens the Graph Builder launch window. This button changes into the Start Over button once you perform an action. You can add a By variable in the Graph Builder launch window.

Done Hides the buttons, Columns list, and all drop zone outlines. Equivalent to deselecting the Show Control Panel option.

Use this presentation-friendly format for pasting into other applications. To copy the graph, select **Edit > Copy**. To restore the window to the interactive mode, click the Graph Builder red triangle and select **Show Control Panel**.

Undo Reverses the last change made to the window.

Start Over Returns the window to the default condition, removing all data and graph elements from the window, and all variables from the drop zones.

Right-Click Menus

Numerous options are contained in Graph Builder right-click menus. The specific menu options depend on where in a graph you click, and the types and number of variables in your graph. Descriptions for the right click menu options are grouped in the following categories.

- "Graph Options"
- "Grouping Zones Options"
- "Options for Columns in the Variables Panel"
- "Options for Axes, Axes Variable Labels, or Graph Titles"
- "Options for the Dividing Line between Multiple Graphs"
- "Legend Options"

For multiple graphs (created with grouping zones or the Wrap zone), note the following:

- Generally, any options that you apply to one graph element apply to all graphs, across all grouping and wrap variables.
- To apply an option to only one graph, right-click in that graph area and select the option from the Graph menu.
- To apply an option to all of your multiple graphs simultaneously, first hold down the Ctrl key and right-click and then select the option from the Graph menu.

Graph Options

Right-clicking in a graph shows the following:

1. Element types that appear in the plot (such as Points, Line, and so on). Sub-menus provide options that are specific to each element type, and general options. Alternatively, you can change element-specific options in each element's properties panel, below the Variables panel.
2. The Add option, where you can select other elements to add to the graph. Alternatively, you can add more elements to the graph by dragging the desired element icons.
3. The Rows, Graph, Customize, and Edit menus. For descriptions of these options, see *Using JMP*.

Tip: Select **Rows > Name Selection in Column** to create a new column to name and label selected points in a graph.

Grouping Zones Options

Right-clicking in the X Group, Y Group, Wrap, Color, Size, and Overlay zones can show the following options:

Levels in View Changes the number of levels of the grouping variable that are visible. Use the arrow buttons to scroll forward and backward.

Number of Levels Changes the number of levels. See "Categorical and Continuous Variables in Group X or Y Zones".

Order By Orders the levels of a variable. See "Order the Levels of a Categorical Variable".

Show Title Shows or hides the variable title.

Title Orientation Changes the orientation of the variable text to horizontal or vertical.

Level Orientation Changes the orientation of the level values to horizontal or vertical.

Levels per Row Changes the number of columns included in the graph. Use with a **Wrap** variable.

Color Changes the background color of the grouping zone.

X or Y Group Edge Moves the grouping variable labels. See "Move Group X and Group Y Labels".

Summary Statistic Applies color and or size variables based on a summary statistic. Available only for a summarized Y response.

Swap Swaps the position of two variables. See "Move or Remove Variables in Zones".

Remove Removes a variable.

Options for Columns in the Variables Panel

Right-clicking on a column in the Variables panel shows options that depend on the column's modeling type. For more information about these menus, see *Using JMP*.

Options for Axes, Axes Variable Labels, or Graph Titles

Right-clicking on axes, axes variable labels, or graph titles show options for customizing your graph. Select **Show Properties** from a right-click menu to access the properties panel for additional customizations, such as tick line thickness or color. The following options are included in the right-click menus:

Graph Titles

Alignment Changes the graph title alignment.

Span Changes the width of the graph title span.

Additional options for customizing fonts, font colors, and the visibility of the title are also available in the right-click menu.

Axes

Axes have two levels of right-click menus, one for the axis title and one for the axis labels.

Combine Scales Combines scales when you have multiple variables in the same zone. Auto does the following:

- If there are 4 or more variables on one axis and no variables on the other axis, elements that support parallel mode are combined as parallel scales. If all variables are continuous and overlapping, Parallel Merged is used.
- If all variables are categorical, use Nested.

In addition, consider the following:

  - Merged creates one scale that covers all variables.
  - Nested creates a nested scale with one level per variable.
  - Parallel Merged is for parallel coordinates with a shared scale. You can use this option with the following elements: Points, Contour (Violin), Histogram, Box Plot and Parallel.
  - Parallel Independent is for parallel coordinates with independent scales. You can use this option with the following elements: Points, Contour (Violin), Histogram, Box Plot and Parallel.

Date/Time Bin (Appears only for a date or time axis variable.) Enables you to select a level of granularity to bin the data. For example, you might choose to bin at a high level, by year, or at a low level, by second.

Date/Time Wrap (Appears only for a date or time axis variable.) Enables you to select a cyclic period for your date data.

Swap Switches a variable from one zone to another. Select the variable that you want to switch places with.

Remove Removes the variable from the zone.

Move Right (Appears only if you have more than one variable in the Y zone.) Creates a second Y axis (on the right) for the variable that you select.

Move Left (Appears only after you have moved a variable to the right.) Moves the variable back to the left.

Align Ticks with Left Axis Aligns the major tick marks with those on the left axis.

Order By (Appears only if you have categorical variables in the X or Y zone.) Orders the levels of a categorical variable. See "Order the Levels of a Categorical Variable".

Note: Select **Other** to indicate a continuous variable as the ordering variable.

Order Statistic Enables you to select a statistic to use for the ordering.

Save Value Order (Available once an ordering has been defined.) Saves the defined order to the Value Order column property.

Note: You can edit value ordering from Graph Builder. Right-click on an axis and select **Edit Value Order**.

Size By (Available for a categorical axis.) Enables you to size your box plots by the count of observations within a group or by a numerical column.

Size Statistic (Available when the Size By option indicates a numerical column as the sizing variable.) Enables you to select a statistic to use for the sizing.

For descriptions of the general axis options below the line, see *Using JMP*.

Note: The **Revert Axis** option reverts the axis back to the default settings. The **Revert Scale** option reverts the scale-specific items: min, max, major interval, n minor.

Options for the Dividing Line between Multiple Graphs

This menu appears only if you have two or more graphs in the graph area. Right-click the line that separates graphs. Options appear that change the formatting of the line, such as color, transparency, spacing, and borders.

Tip: If the line between the graphs is too thin, you can resize it. Click the Graph Builder red triangle, select **Graph Spacing**, and enter a larger value.

Legend Options

A legend consists of elements, such as lines, fill colors, and gradients. Right-click on legend elements for customization properties. These properties include color and labeling options when available. Additional legend settings are found in the Legend Settings window.

To change legend settings or revert to initial settings, right-click to the right of the legend or on the title (if applicable). To undo recent changes to the legend, click Undo.

Tip: When you have both an overlay and a color variable, the color variable levels are shown by rounded color icons in the legend.

Legend Settings Change certain aspects of the graphics elements (line, bars, and so on) such as color, fill, transparency, and gradient. You can also add a title to the legend. See "Legend Settings Window".

Revert Legend Returns the legend to its initial settings.

Move the Legend

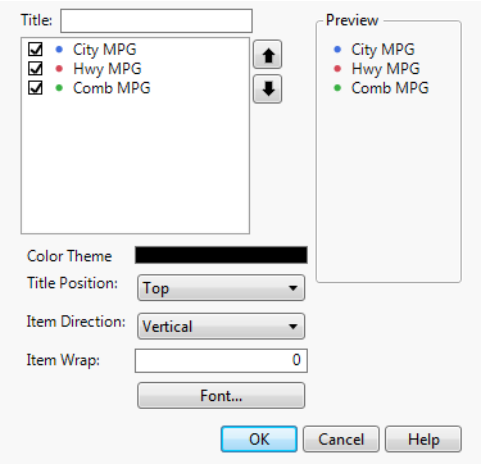
To move the position of the legend, click and drag to the desired location. The location can be a hotspot corresponding to the locations accessible from the Graph Builder red triangle menu or the legend or it can be any other location within the graph.

To use the menu options, click the Graph Builder red triangle and select **Legend Position**. To make the legend vertical or horizontal, change the Item Direction in the Legend Settings window.

Legend Settings Window

Change aspects of the legend through the Legend Settings window.

Figure 3.45 Example of the Legend Settings Window Using Hybrid Fuel Economy.jmp



Title The name of the legend.

Check boxes Shows or hides items in the legend. JMP hides items that appear to be redundant, but you can make them appear by selecting them. If your legend has a large number of entries, some might be hidden in the Graph Builder window to ensure sufficient space for the graph. However, all of the entries appear here, so you can select the ones that you want to show. To select or deselect all entries, right-click and select **Toggle check box**.

Up and down arrows Changes the order of items in the legend.

Color Theme Select a different color theme. For more information about color themes, see *Using JMP*.

Title Position Places the legend title on top or to the left of the items in the legend.

Item Direction Displays the legend horizontally or vertically.

Item Wrap Sets the legend to be *n* items tall (if vertical) or *n* items wide (if horizontal).

Font Changes the font type, style, and size.

Preview Shows your changes to the legend.

OK Commits your changes to the legend.

Cancel Cancels your changes to the legend.

Help Opens the online Help.

Discrete or Continuous Legend Items

Legends can list discrete values or show a scale of continuous values:

- Discrete legends can contain the levels of a categorical variable, or a list of variables.
- Continuous legends describe a color intensity scale for a range of numerical values.

Right-click an item in either legend to see the following options:

Line Color (Discrete legends only.) Changes the color of the element associated with the item.

Marker Changes the type of marker associated with the item.

Marker Size Changes the size of the marker associated with the item.

Pattern or Style Changes the fill or pattern of the element associated with the item.

Line Width Changes the width of the line associated with the item.

Fill Color Assigns a color to the item.

Color by Theme Colors multiple contours by levels in a color theme. This option sets line colors and fill colors for density contours.

Transparency Changes the marker or label transparency. Enter the level of transparency to draw markers (points) on the graph. The degrees of opacity ranges from 0 (clear) to 1 (opaque).

Gradient (Continuous legends only.) Changes the gradient settings. See "Gradient Settings".

Density Gradient (Contour element only.) Changes the gradient colors:

Fade to White The highest density contour uses the fill color, and colors fade to white at lower densities.

Fade to Gray The highest density contour uses the fill color, and colors fade to gray at lower densities.

Full Color Density contours use a predefined or customized color theme with full color control.

Gradient Transparency (Contour element only.) Changes the transparency of the gradient:

None Density contours use one transparency level for all levels.

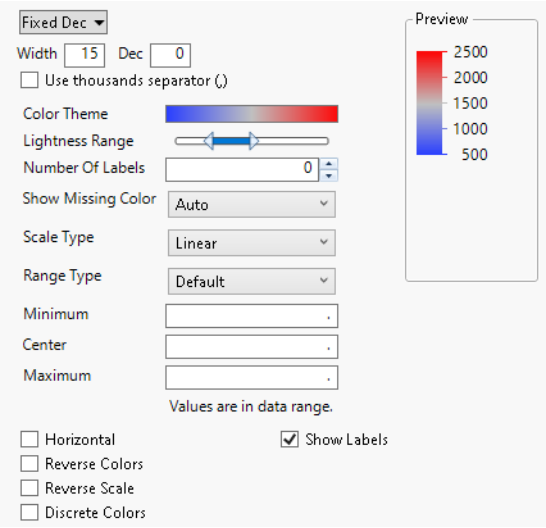
Linear Density contours get gradually more transparent at lower densities.

Revert Restores default settings.

Gradient Settings

You can change various aspects of the gradient legend in the Gradient Settings window.

Figure 3.46 Gradient Settings Window



To format the labels in the legend, use the menu, Width and Dec boxes, and the **Use thousands separator (,)** check box in the top left of the window.

Color Theme Change the color theme or define a custom color theme. For more information about color options, see *Using JMP*.

Lightness Range Set the range of intensities for the gradient coloring.

Number of Labels Specify the number of labels for your legend. The value of zero provides the default number of labels.

Show Missing Color Specifies whether to show missing color in the legend. Auto shows the color only when there are missing values, Off never shows the color, On always shows the color.

Scale Type Sets the scale for the gradient coloring. Look below the Maximum option for a description of the scale type.

Linear The scale is piecewise linear between the Minimum and Center values and between the Center and Maximum values. This is the default scaling.

Quantile The scale is piecewise linear for quantiles of the variable represented by the legend.

Standard Deviation The legend range is divided into offsets from the mean determined by standard deviation increments. The scale is piecewise linear between these offset settings. The number of offsets is determined by the Number of Labels.

Log The scale is linear for the logarithm of the values represented by the legend.

Log Offset The scale is linear for the offset of the base 10 logarithm of the values represented by the legend.

Range Type Sets the range of the legend values.

Default Unless you specify values for the Minimum and Maximum, the minimum and maximum values include the range of the data and are chosen to be appropriately rounded values.

Exact Data Range The minimum and maximum values are the exact minimum and maximum values for the data, or they are the values that you specify as Minimum and Maximum.

Middle 90% The minimum and maximum values are the 5^{th} and 95^{th} quantiles, or they are the values that you specify as Minimum and Maximum. This option is resistant to outliers.

Minimum, Center, Maximum Specify the smallest, middle, and largest values to use for your legend scaling.

Note: These values can override the Range Type values.

Horizontal Changes the orientation of the legend to horizontal.

Reverse Colors Reverses the colors in the color scheme.

Reverse Scale Reverses the scale for the color theme.

Discrete Colors Changes the color scheme from a continuous gradient to a stepped gradient with discrete colors.

Show Labels Shows or hides labels for the legend.

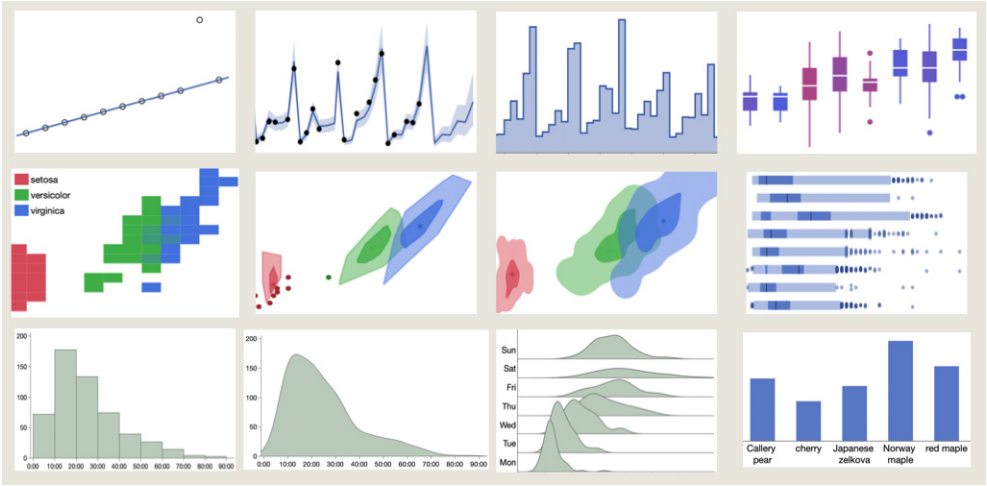
Chapter 4

Graph Builder Examples

Interactive Data Visualization

Graph Builder provides you with a variety of ways to visualize and explore your data. This chapter provides detailed examples of data exploration using bar charts, histograms, area plots, maps, and other types of plots.

Figure 4.1 Graphs Created in Graph Builder



Contents

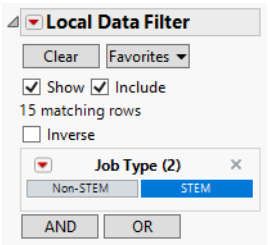
Example of an Ordered Bar Chart 123
Example of an Area and Line Chart 126
Example of a Map Chart 131
Example of a Butterfly Chart 134
Example of a Scatterplot 138
Example of a Packed Bar Chart 143
Example of an Overlaid Histogram and Ridgeline Chart 145
Example of a Bullet Bar Chart 147
Example of a Forecast Plot 149

Example of an Ordered Bar Chart

Data	This example uses data from college graduates in 2012. The table contains counts of degrees for graduates in 15 fields of study. Each degree is subdivided by the graduate's job type. There are categorical columns for degree, degree type, and job type (where type is STEM or non-STEM related). The continuous column Number contains the number of graduates from each unique combination of degree and job type. **Note:** These data come from the U.S. Census Bureau (2012).
Techniques	This example uses color, a local data filter, and sorting options.
Goal	The goal of this example is to visualize the percentage of graduates who are employed in STEM-related jobs, based on their degree type.

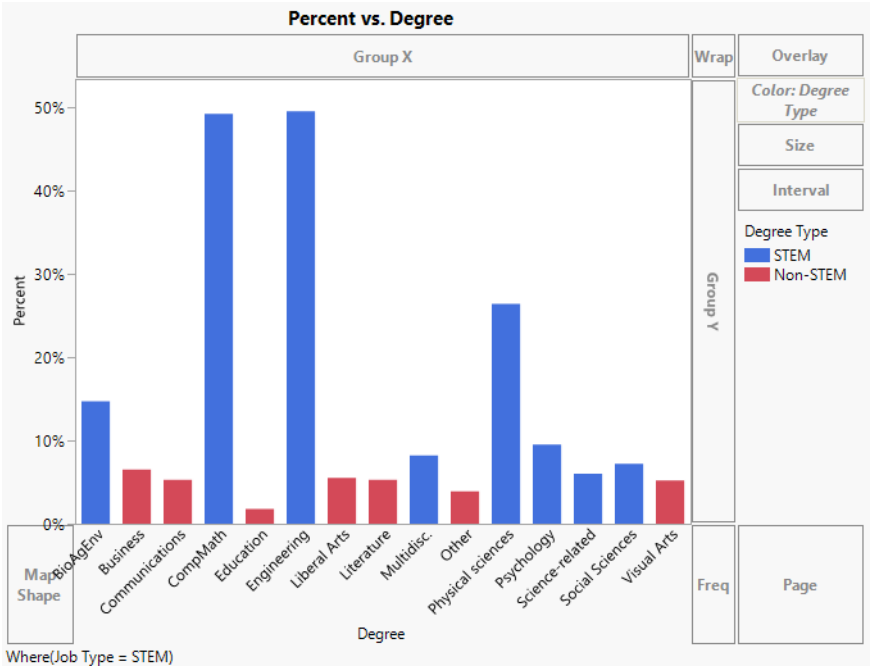
1. Select **Help** > **Sample Data Library** and open STEM Jobs.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select Degree and drag it to the **X** zone.
4. Select Percent and drag it to the **Y** zone.
5. Click the Bar element .
6. Click the Graph Builder red triangle and select **Local Data Filter**. In the Local Data Filter, perform these steps:
   a. Click Job Type and then click .
   b. Click **STEM**.

Figure 4.2 Selections in the Local Data Filter



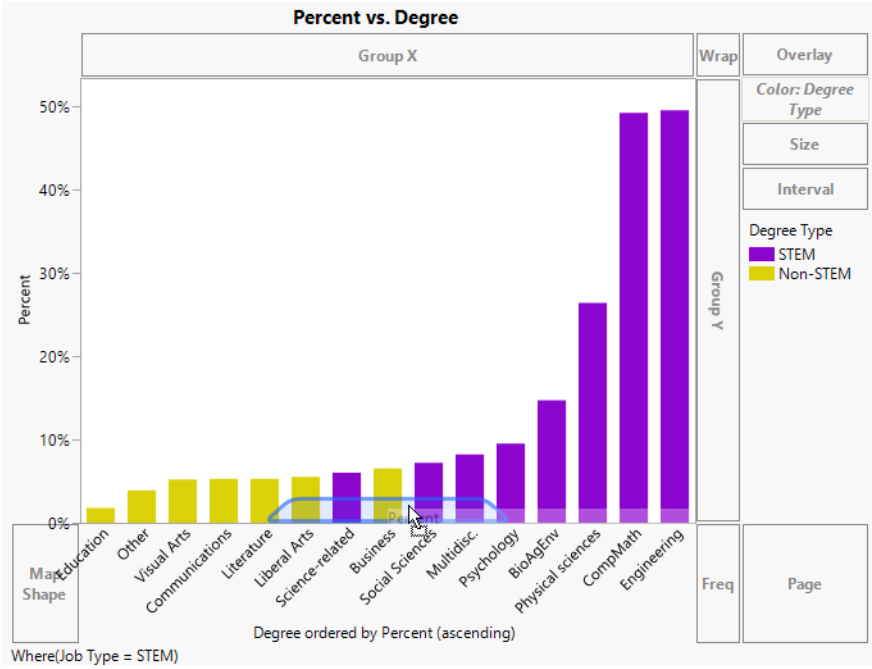
7. In the Graph Builder column list, select Degree Type and drag it to the **Color** zone.

Figure 4.3 Bar Chart Colored by Degree Type



8. In the legend, change the bar colors according to degree type:
   - Right-click **STEM** and change the fill color to purple.
   - Right-click **Non-STEM** and change the fill color to yellow.
9. In the Graph Builder column list, select Percent and drag it above the X axis. See Figure 4.4.

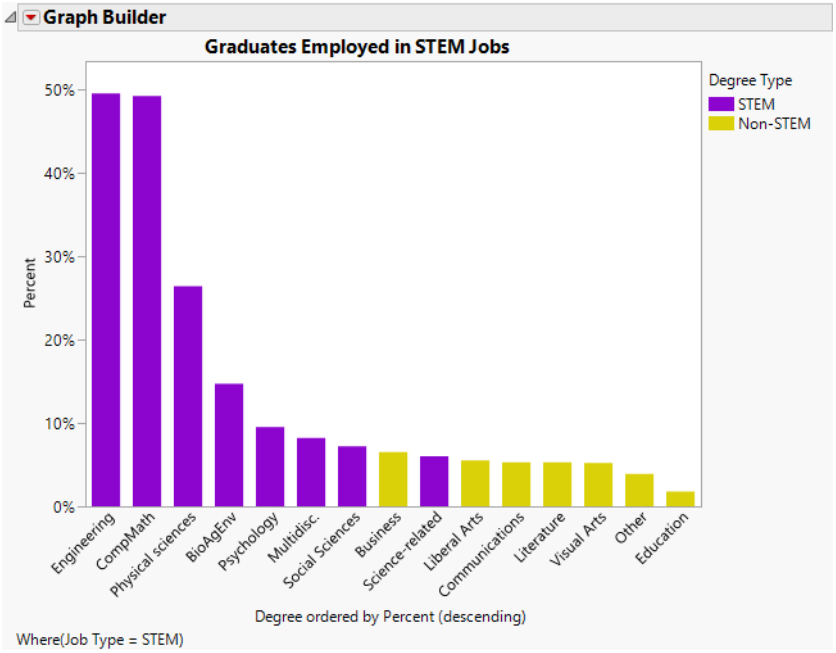
Figure 4.4 Ordering Degree by Percent



This orders the degree by percent in ascending order. However, you want it to be in descending order.

10. Right-click the X axis and select **Order By** > **Percent, descending**.
11. Click the graph title and type Graduates Employed in STEM Jobs and press Enter.
12. Click **Done**.

Figure 4.5 Completed Bar Chart



This graph displays the percent of graduates employed in STEM jobs by their degree type. Nearly 50% of Engineering and CompMath graduates are employed in STEM jobs, as compared to less than 2% of Education graduates.

Example of an Area and Line Chart

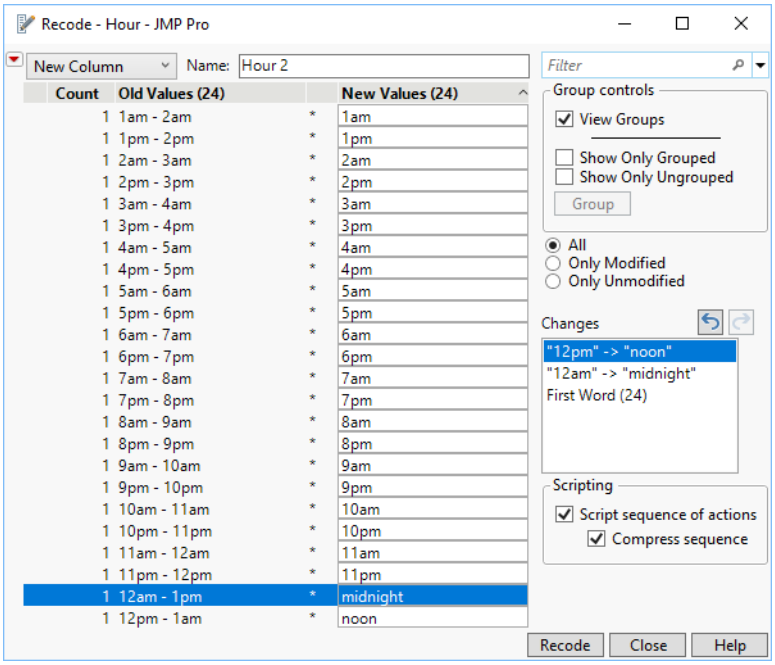
Data	This example uses data based on website traffic. It includes hourly measurements and historical range information.
Techniques	This example uses Recode to improve data labels, multiple graph elements (area and line), and legend customization. It also shows how to remove variables from a graphing element when multiple variables are plotted on a single axis.
Goal	The goal of this example is to display trend lines for website traffic within an overall range.

1. Select **Help** > **Sample Data Library** and open Bands.jmp.

2. Select the Hour column, and then select **Cols** > **Recode**.
3. In the Recode window, make these changes:
   a. Click the red triangle and select **First Word**.

      Notice that the New Values include only the first hour.
   b. Under New Values, click **12am** and enter **midnight**. Click **12pm** and enter **noon**.

Figure 4.6 Completed Recode Window



   c. Click **Recode**.

      A new column called Hour 2 is added to the data table.

Tip: Recode enables you to recode your data to a new column, as a formula in a new column, or to overwrite your current data. For more information about Recode, see *Using JMP*.

4. Select **Graph** > **Graph Builder**.
5. Select Hour 2 and drag it to the **X** zone.

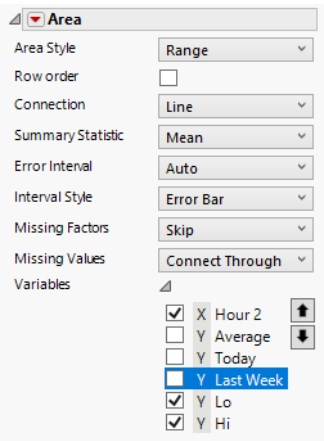
Note: The ordering of the X axis is logical, going from midnight to 11pm. This is because the Hour column has a value ordering column property that was copied to the Hour 2 column when you created it with Recode.

6. Select Average, Today, Last Week, Lo, and Hi and drag them to the **Y** zone.

7. Click the Area element .
8. In the Area properties panel, make these selections:
   a. Change the Area Style to **Range**.
   b. Open Variables and deselect Average, Today, and Last Week.

      Now the area element is applied only to the Lo and Hi variables.

Figure 4.7 Completed Area Properties Panel



9. Hold down the Shift key and click the Line element .

Tip: Holding down the Shift key and clicking adds additional elements to the graph, without replacing any existing elements.

10. In the Line properties panel, open Variables and deselect Lo and Hi.

    Now the line element is applied only to the Avg, Today, and Last Week Y variables.

Figure 4.8 Completed Line Properties Panel

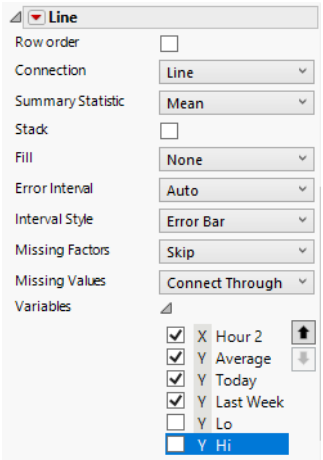
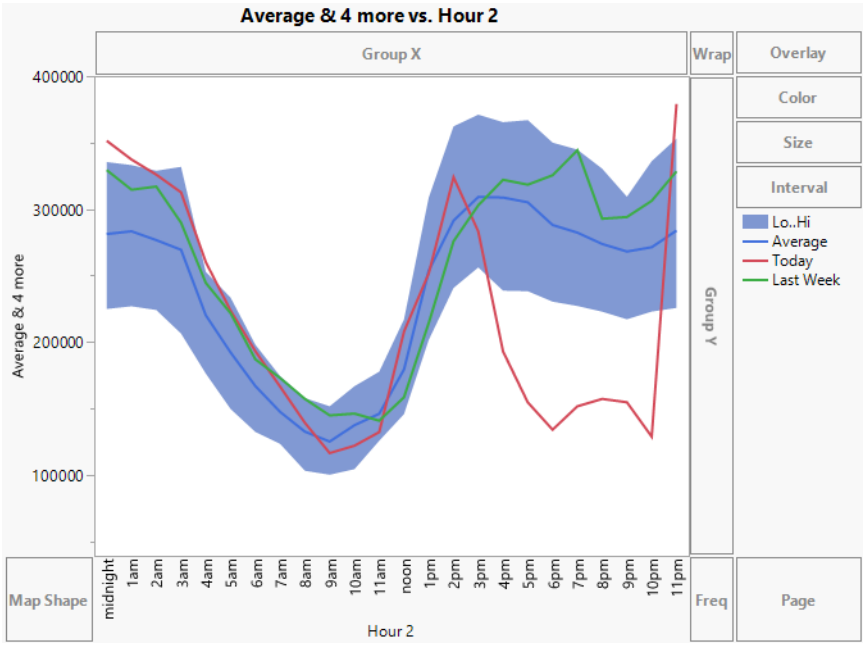


Figure 4.9 Area and Line Graphs

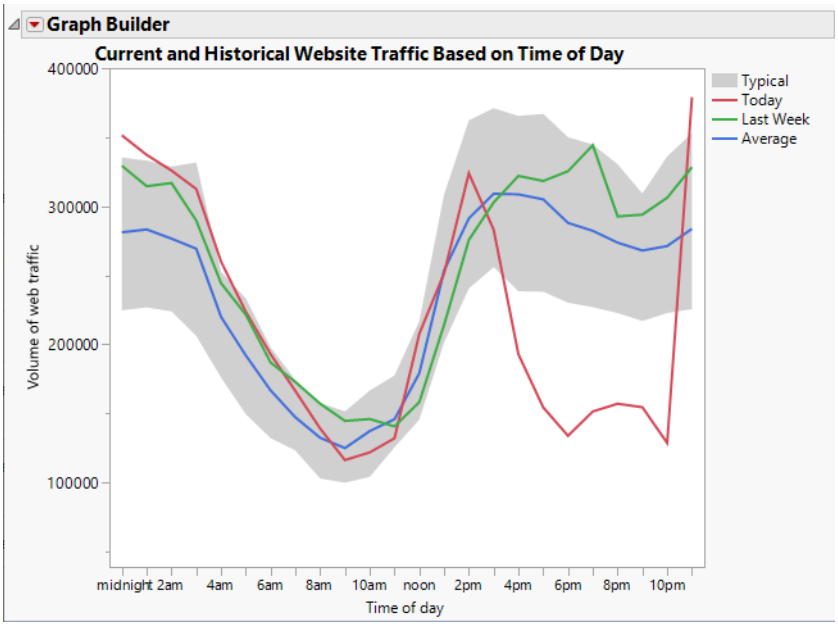


11. Customize the graph elements using the legend:
    a. In the legend at right, double-click **Lo..Hi**.
    b. In the Legend Settings window, double-click **Lo..Hi** and rename it to **Typical.**
    c. Highlight Today and Last Week, and click the up arrow to move them up one, so that the order is: Typical, Today, Last Week, and Average. Click **OK**.

d. Right-click Typical > **Fill Color** and select light gray.

12. Right-click the X axis > **Axis Settings**. For Label Orientation, click **Horizontal**, and then click **OK**.
13. Change the titles of the axes and graph:
    a. Click the Y-axis title and enter Volume of web traffic.
    b. Click the X-axis title and enter Time of day.
    c. Click the graph title and enter Current and Historical Website Traffic Based on Time of Day.
14. Click **Done**.

Figure 4.10 Completed Area and Line Graph



Tip: If an item in the legend is still highlighted, click it to deselect it.

The graph shows that the afternoon website traffic fell to unusually low levels today. The red trend line that represents today's traffic peaks at 1pm and then falls to outside of the gray band that represents the range of expected daily traffic for this website by about 3pm.

Example of a Map Chart

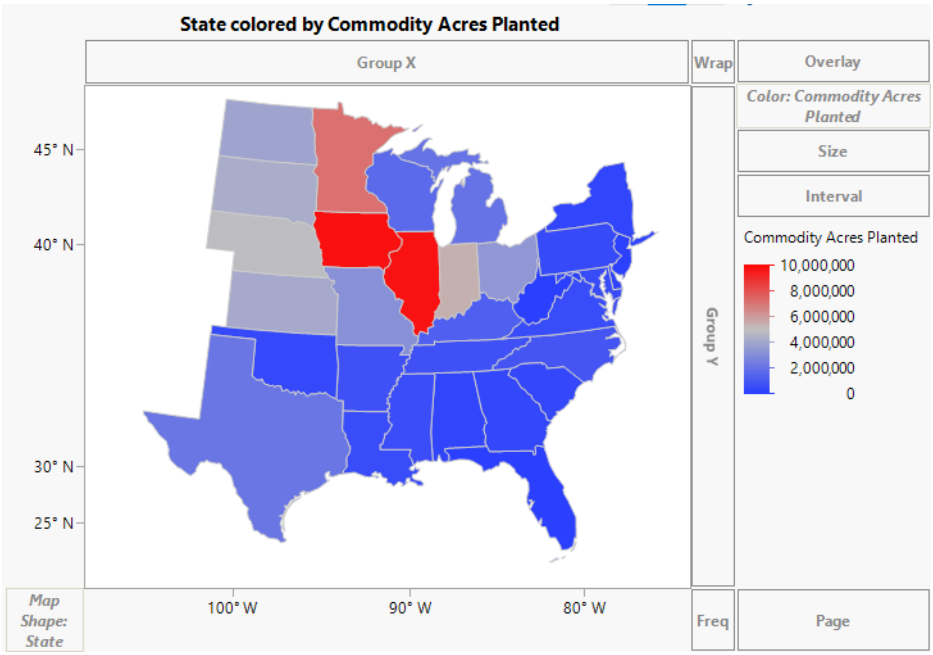
Data	This example uses data from 2017 of commodity acres planted in states within the US that grow corn, wheat, and soybeans. **Note:** These data come from the United States Department of Agriculture (2017).
Techniques	This example uses map shapes and legend customization.
Goal	The goal of this example is to show the median number of commodity acres that have been planted per state.

1. Select **Help** > **Sample Data Library** and open Corn Wheat Soybean Production.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select State and drag it to the **Map Shape** zone.

   Note: JMP recognizes US State names as map shapes. For more information, see "Map Shape".

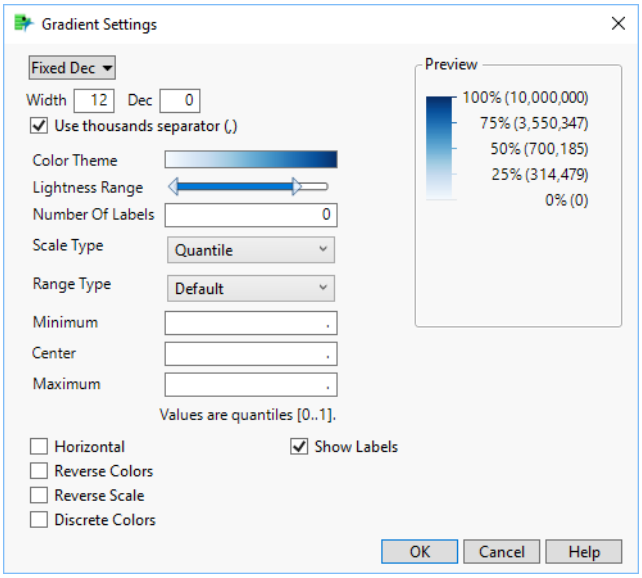
4. Select Commodity Acres Planted and drag it to the **Color** zone.
5. In the Map Shapes properties panel, change the Summary Statistic to **Median**.

Figure 4.11 Map Showing Median Commodity Acres Planted



6. Customize the graph elements in the legend:
   a. In the legend color bar, right-click and select **Gradient**.
   b. Next to Width, enter **12**.
   c. Next to Color Theme, click the color bar. Under Sequential, click the 2nd option (white to blue) and then click **OK**.
   d. From the Scale Type list, choose **Quantile**.

Figure 4.12 Completed Gradient Settings Window



e. Click **OK**.

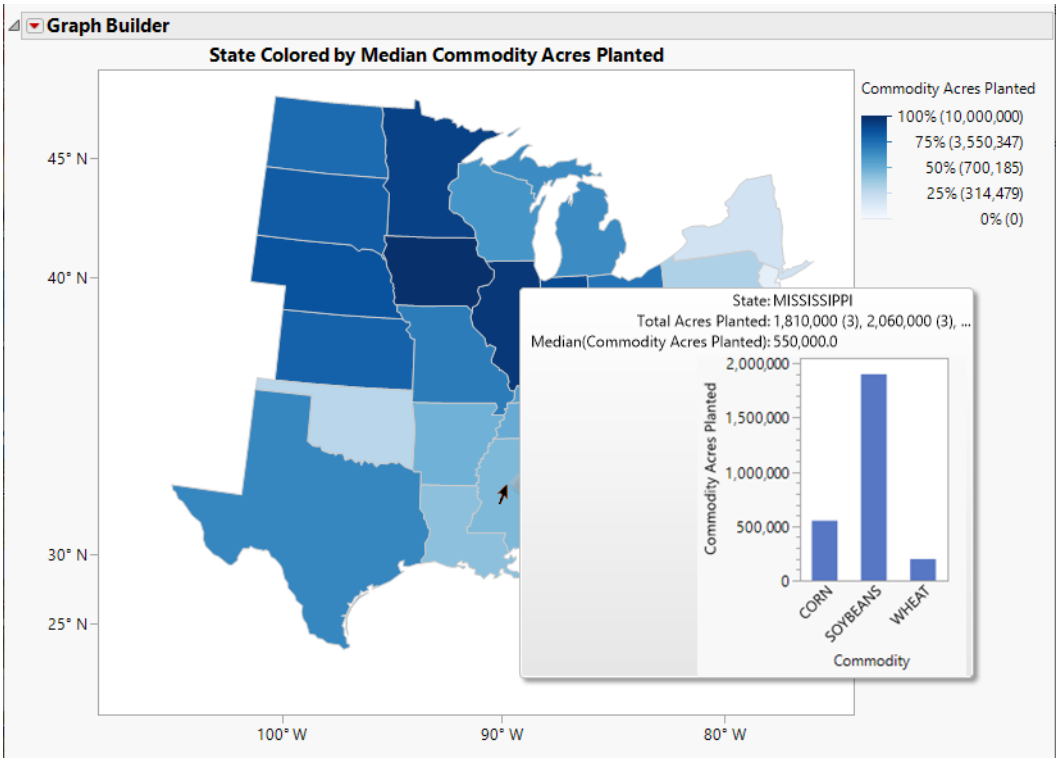
7. Click the title above the graph and type **State Colored by Median Commodity Acres Planted**.
8. Right-click in the graph and select **Hover Label > Bar**.

   This adds a bar chart hover graph. When you hover over a state, you can see the proportions of the commodities grown in that state in a bar chart.

   Note: For more information about Hover Labels, see *Using JMP*.

9. Click **Done**.
10. Hover over a state (in this case, Mississippi).

Figure 4.13 Completed Map Colored by Median Commodity Acres Planted



The coloring on the map highlights the states with the highest number of acres of wheat, corn, and soybeans. The dark blue states are known as America's Breadbasket. The hover graphs enable you to explore the proportions of the three crops grown in a state. This example shows that in Mississippi, soybeans are the prominent crop.

Tip: To see other examples using Graph Builder, run the additional scripts in the Corn Wheat Soybean Production.jmp data table.

Example of a Butterfly Chart

Data	This example uses summarized survey responses to 20 questions all using the same 5-point Likert scale.
Techniques	This example uses a bar chart with variable ordering and color customizations.

Goal	The goal of this example is to create a butterfly chart of the responses to 20 survey questions. This provides a summary visual of the results from a large number of survey questions in one chart.

1. Select **Help** > **Sample Data Library** and open Likert Survey.jmp.

   Notice the following about this data:

   – The data has a row for each question and columns with the counts of responders for each different response.

   – The order of the survey questions in the question column is not sequential. When you use this column in Graph Builder, JMP orders the rows sequentially.

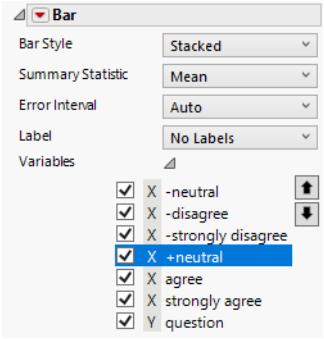
   – Formula columns are included in the table to create negative counts for negative responses. The neutral counts are split so that half contribute to a negative column and half contribute to a positive column.

2. Select **Graph** > **Graph Builder**.
3. Select question and drag it to the **Y** zone.
4. Select agree, strongly agree, -strongly disagree, -disagree, -neutral, and +neutral and drag them to the **X** zone.
5. Click the Bar element .
6. In the Bar properties panel, do the following:

   – Change the Bar Style to **Stacked**.

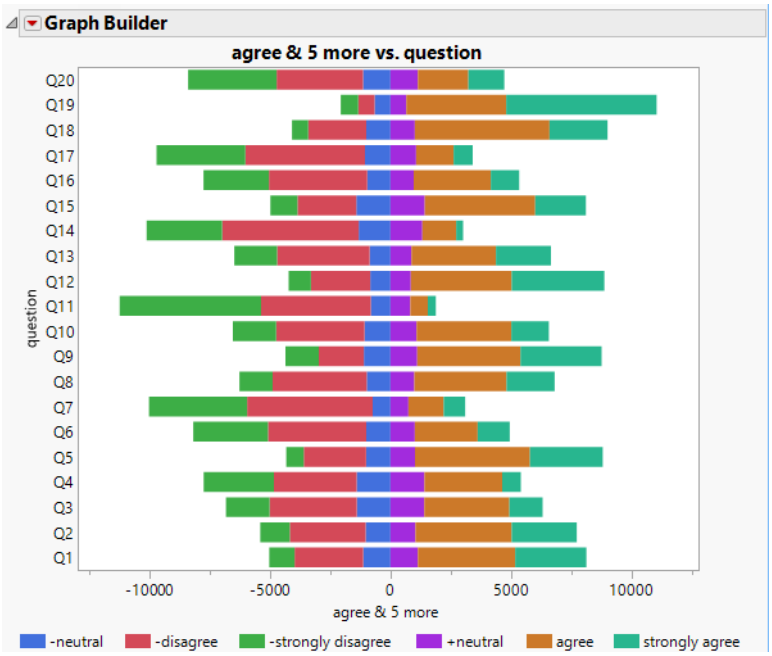
   – Open **Variables**, and then highlight a variable and use the arrow icons to move them into this order, from top to bottom: -neutral, -disagree, -strongly disagree, +neutral, agree, strongly agree.

Figure 4.14 Completed Bar Properties Panel



7. Click the Graph Builder red triangle and select **Legend Position** > **Bottom**.
8. Click **Done**.

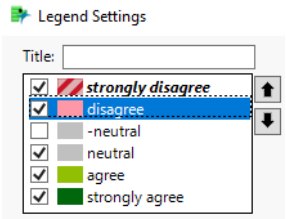
Figure 4.15 Graph Prior to Legend Customizations



Customize the Legend Colors and Ordering

1. Right-click on each color in the legend and change the Fill Color:
   - Light gray for +neutral and -neutral
   - Light pink for -disagree
   - Red for -strongly disagree
   - Light green for agree
   - Dark green for strongly agree
2. Hover over any item in the legend. The cursor should appear as a hand. Double-click to open the Legend Settings window.
3. Highlight a variable and use the arrow icons to change the order from top to bottom: -strongly disagree, -disagree, -neutral, +neutral, agree, strongly agree.
4. Double-click +neutral and rename it to **neutral**.
5. Deselect the box next to -neutral so that it no longer appears in the legend, because it is a duplicate.
6. Double-click -strongly disagree and -disagree and remove the - sign.

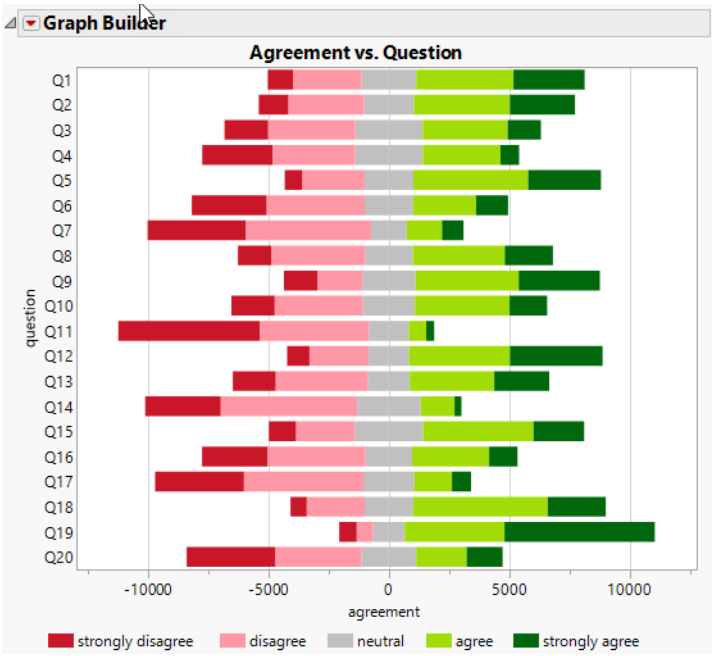
Figure 4.16 Completed Legend Settings



7. Click **OK**.
8. Change the titles of the X axis and graph:
   a. Click the graph title and enter Agreement vs. Question.
   b. Click the X-axis title and enter agreement.
9. Double-click in the Y axis and select the **Reverse Order** box, and then click **OK**.
10. Double-click in the X axis and select the **Major Grid Lines** box, and then click **OK**.

    The grid lines help you visually judge the differences between question results.

Figure 4.17 Final Likert Scale Graph



Tip: If an item in the legend is still highlighted, click it to deselect it.

This graph provides a visual summary of 20 survey questions, ignoring the non-responses. Note that the non-responses are less than 1% of the total responders for any question. Therefore, leaving them out of the visualization does not impact the conclusions. Question 19 has the most positive responses, and Question 11 has the least favorable responses.

Example of a Scatterplot

Data	This example uses data from 442 diabetic patients. The data includes baseline clinical and laboratory data as well as a binary measure of diabetes disease progression obtained one year after each patient's initial visit. This measure classifies disease progression as Low or High.
Techniques	This example uses two scatterplots with a shared X axis, axis customization, and annotation.
Goal	The goal of this example is to compare and understand the predicted probabilities of High disease progression from two classification models.

Run the Classification Models

You want to construct a classification model to predict the disease progression based on clinical and laboratory variables. You will build two different classification models, save prediction formulas, and compare their predicted classifications.

1. Select **Help > Sample Data Library** and open Diabetes.jmp.
2. Click the green triangle next to the **Decision Tree of Y Binary** script to build a decision tree prediction model.
3. Click the Partition for Y Binary red triangle and select **Save Columns > Save Prediction Formula**. You can close this window.

   This saves the probability formulas to the data table.
4. In the data table, right-click the Prob(Y Binary= = High) column and select **Column Info**.
5. Change the column name to Partition Prob High and click **OK**.
6. Click the green triangle next to the **Neural of Y Binary** script to build a neural net prediction model.
7. Click the Model NTanH(3) red triangle and select **Save Profile Formulas**. You can close this window.

   This saves the probability formulas to the data table.
8. In the data table, right-click the Probability(Y Binary=High) column and select **Column Info**.

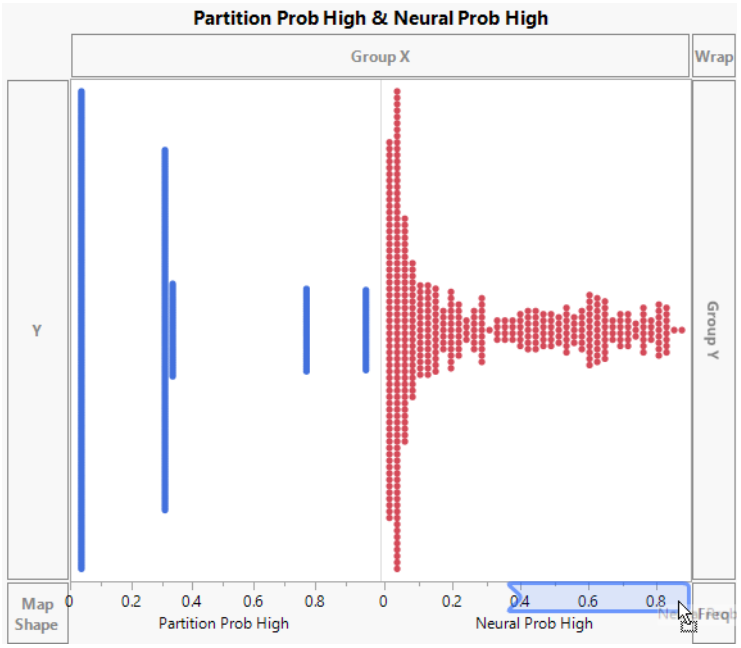
9. Change the column name to Neural Prob High and click **OK**.

Create the Initial Graph

You will compare the model predicted probabilities of a patient having a High disease progression.

1. Select **Graph > Graph Builder**.
2. Select Partition Prob High and drag it to the **X** zone.
3. Select Neural Prob High and drag it to the **X** zone, to the right of Partition Prob High. This creates a second X axis.

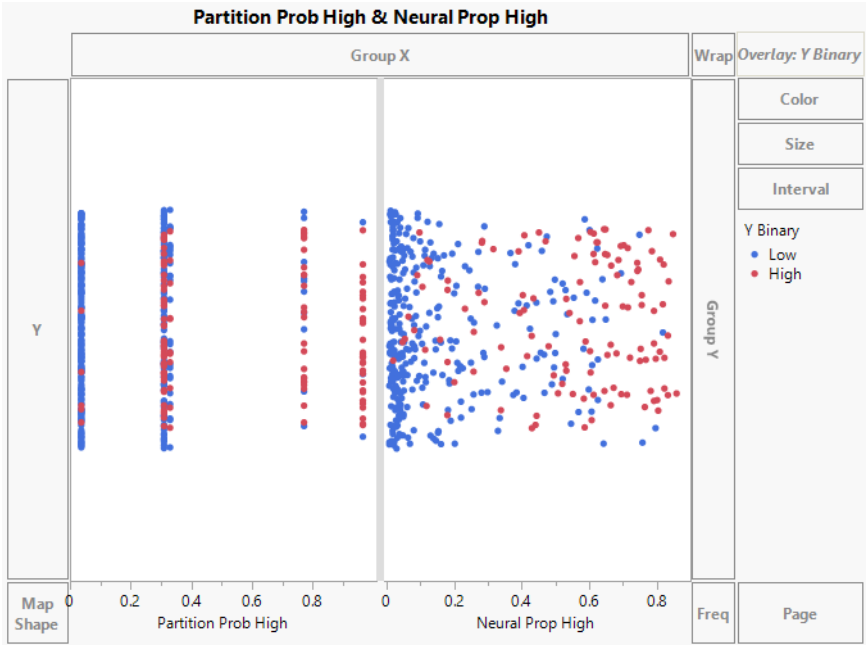
Figure 4.18 Drag Neural Prob High to the Right of Partition Prob High



4. Select Y Binary and drag it to the Overlay zone.
5. Click the Graph Builder red triangle and select **Graph Spacing**.
6. Type 6 next to Graph Spacing and click **OK**.

   This increases the spacing between the two X axes.

Figure 4.19 Initial Graph of Model Probabilities

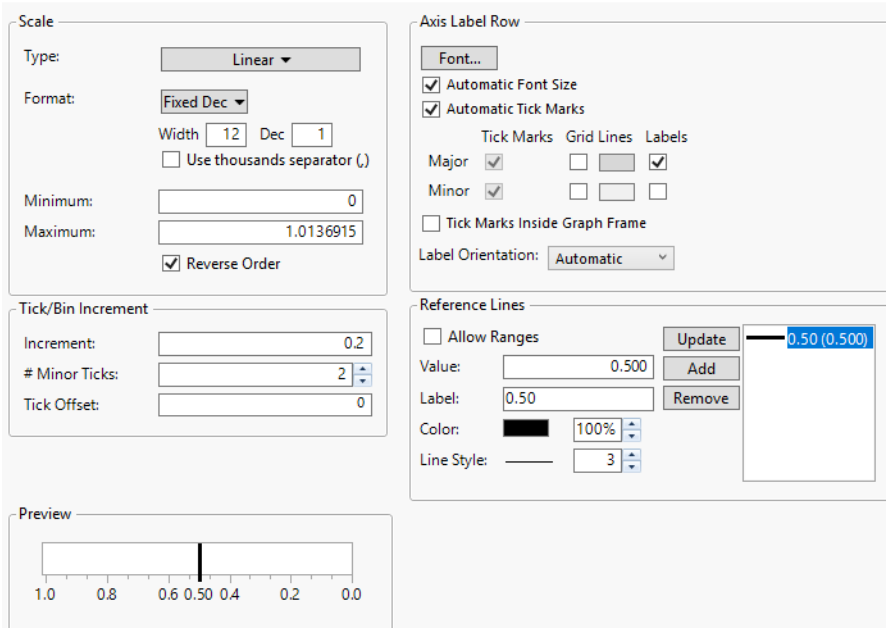


Customize the Graph

In both modeling platforms, the default threshold value is 0.50. This means that if a patient has a predicted probability greater than 0.50 of "High", the model predicts their classification as "High". Use reference lines to show the threshold value in the graph.

1. Right-click the Partition Prob High X axis and select **Axis Settings**.
2. In the Scale section, click the box next to **Reverse Order**.
3. In the Tick/Bin Increment section, set **# Minor Ticks** to 2.
4. In the Reference Lines section, type 0.50 in the boxes next to **Value** and **Label**.
5. In the box next to Line Style, enter 3. The 3 indicates the thickness of the reference line.
6. Click **Add**.

Figure 4.20 X Axis Settings

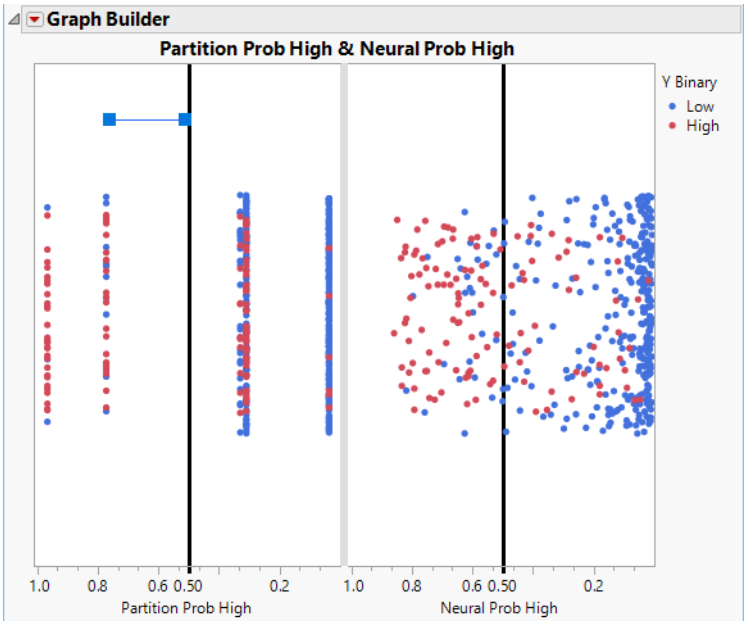


7. Click **OK**.
8. Right-click on the Partition Prob High X axis and select **Edit** > **Copy Axis Settings**. Right-click on the Neural Prob High X axis and select **Edit** > **Paste Axis Settings**.
9. Click **Done**.

Annotate the Graph

1. In the main menu, click **Tools** > **Line**. Draw a horizontal line to the left of the reference line in the Partition Prob High graph.

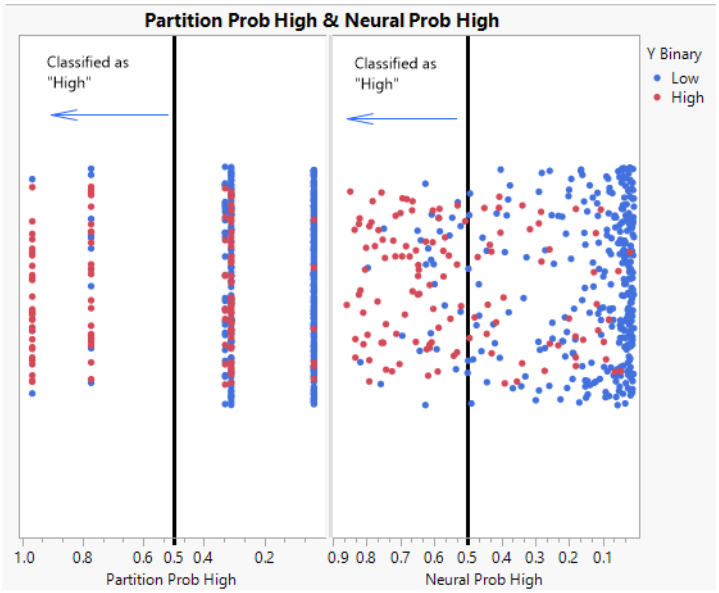
Figure 4.21 Line Drawn to the Left of the Partition Prob High Reference Line



2. Right-click the line and select **Point to**.

   The arrow should be pointing away from the reference line.
3. In the main menu, click **Tools** > **Annotate** and click on the graph above the arrow.
4. Type **Classified as "High"** in the text box. See Figure 4.22.
5. Repeat step 1 through step 4 in the Neural Prob High graph.

Figure 4.22 Distribution of Predicted Probabilities



The graph shows that the distribution of the predicted probabilities differs between the two models. The partition model has five predicted score levels while the neural model scores are dispersed across the score range. For both models, there is a cluster of blue data points on the far right of each plot. These are low subjects that both models classify correctly as they fall to the right of the threshold.

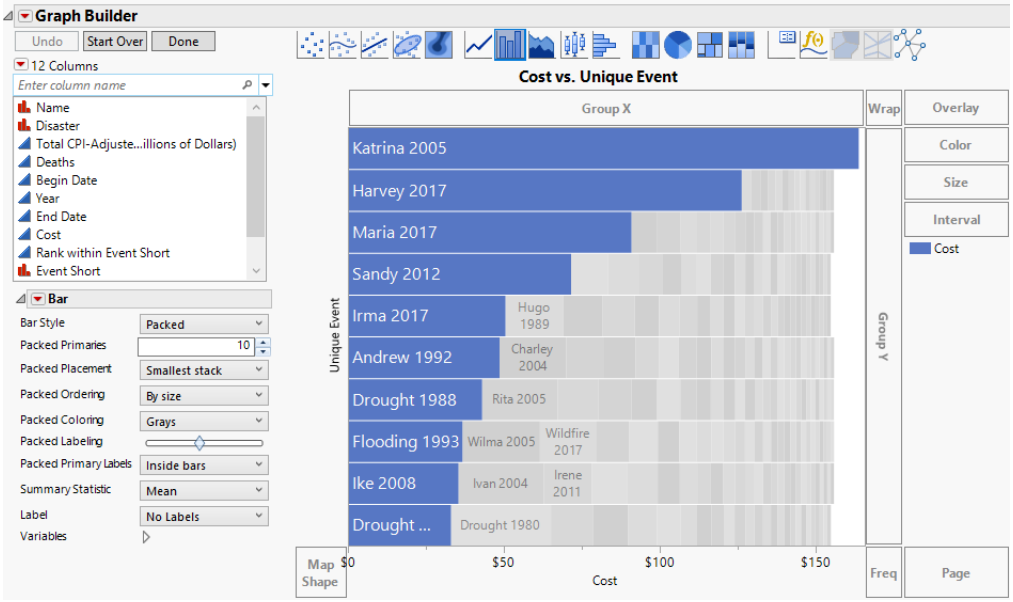
Example of a Packed Bar Chart

Data	This example uses data from the National Oceanic and Atmospheric Administration on the economic impact of weather and climate events in the U.S. during the years 1980 to 2018.
Techniques	This example uses a packed bar chart with 10 primary categories.
Goal	The goal of this example is to create a packed bar chart showing the events that had the greatest economic impact.

1. Select **Help > Sample Data Library** and open Billion Dollar Events.jmp.
2. Select **Graph > Graph Builder**.
3. Select Unique Event and drag it to the **Y** zone.

4. Select Cost and drag it to the **X** zone.
5. Select the Bar element .
6. In the Bar options panel:
   a. For the Bar Style, select **Packed**.
   b. Change Packed Primaries to 10.
   c. Move the Packed Labeling slider to the left until it is about halfway.

Figure 4.23 Weather Events With Large Economic Impacts in a Packed Bar Chart



The top 10 categories appear as a bar chart with blue bars. These are the “Packed Primaries” that you set to 10 in the options. Secondary categories are labeled and in gray. You can hover over any bar for more details.

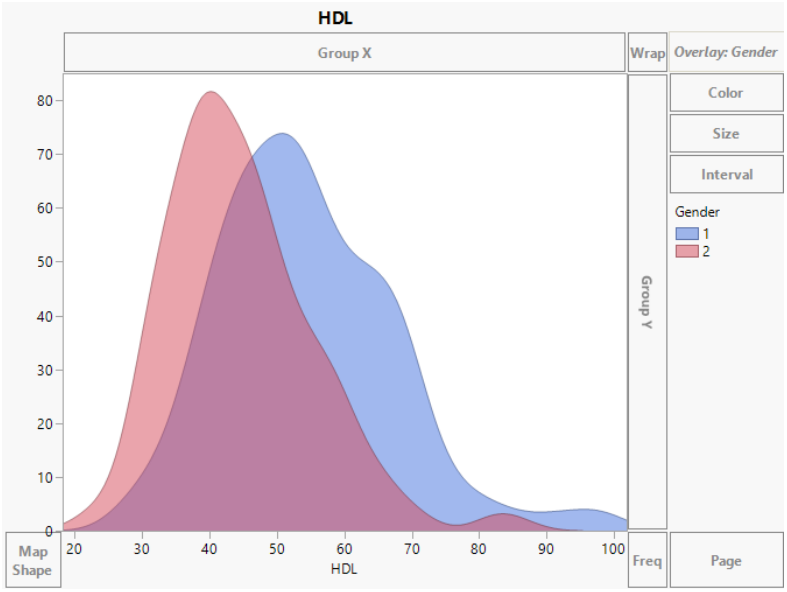
In this graph, you can clearly see that Katrina in 2005 had the largest financial impact of any storm that hit the US between 1980 and 2018. The 2017 hurricanes, Harvey and Maria, also had large financial impacts.

Example of an Overlaid Histogram and Ridgeline Chart

Data	This example uses data from 442 diabetic patients. The data include baseline clinical and laboratory data as well as a binary measure of diabetes disease progression obtained one year after each patient's initial visit. This measure classifies disease progression as Low or High.
Techniques	This example uses overlaid histograms and a ridgeline chart.
Goal	The goal of this example is to explore the distribution of a continuous variable grouped by categories. Specifically, you are interested in exploring how the distribution of HDL varies by gender and then by disease progression.

1. Select **Help > Sample Data Library** and open Diabetes.jmp.
2. Select **Graph > Graph Builder**.
3. Select HDL and drag it to the **X** zone.
4. Select Gender and drag it to the **Overlay** zone.
5. Click the **Histogram** element icon .
6. In the Histogram properties panel, from the Histogram Style list, select **Kernel Density**.
7. Hover over the Y axis near the top of the scale until the cursor turns into a horizontal hand. Click and drag until the maximum value is about 85.

Figure 4.24 Overlaid HDL Histograms by Gender

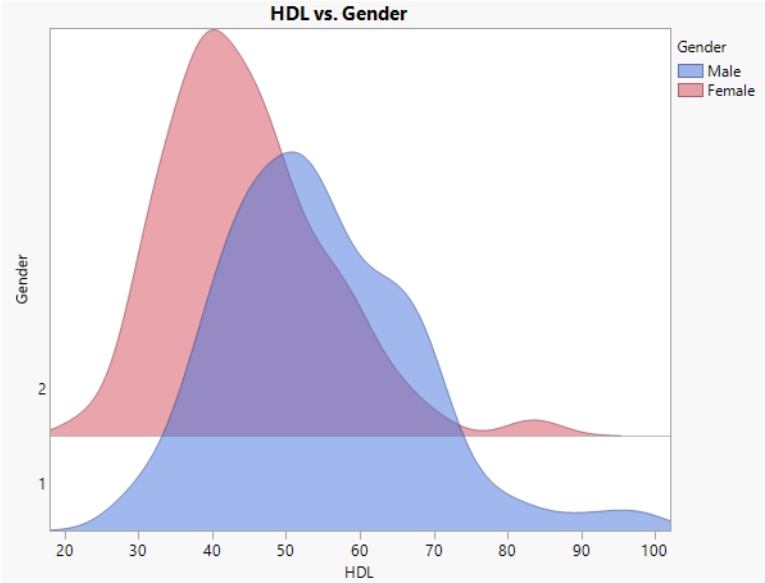


The histograms indicate that HDL levels are higher in males (blue) than females (pink). The blue distribution falls to the right, or higher on the HDL scale, than the pink distribution. Now change the graph to view the differences in gender in a ridgeline plot.

8. Select Gender and drag it to the Y zone.
9. In the Histogram properties panel, move the Overlap slider to overlap the histograms (until the slider is about 3/4 up).
10. Click **Done**.
11. Double-click in the legend. In the Legend Settings window, double-click 1 and type male. Double-click 2 and type female.
12. Click **OK**.

Tip: If an item in the legend is still highlighted, click it to deselect it.

Figure 4.25 Ridgeline Chart of HDL by Gender and Diabetes Level



The ridgeline chart shows the gender distributions offsets.

Example of a Bullet Bar Chart

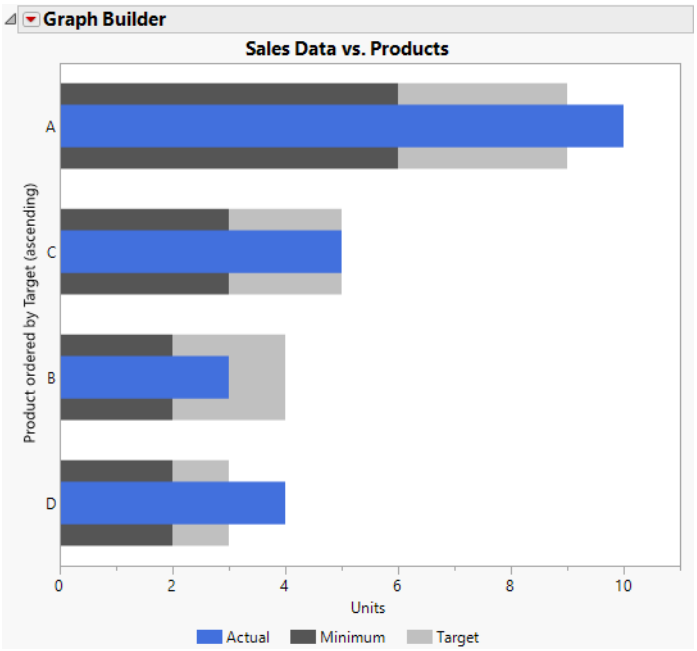
Data	This example uses fabricated sales data for four products.
Techniques	This example uses a bullet chart.
Goal	The goal of this example is to display actual units sold compared to a minimum and target number of units.

1. Select **Help** > **Sample Data Library** and open Bullet Measures.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select Actual, Target, and Minimum and drag them to the **X** zone.
4. Select Product and drag it to the **Y** zone.
5. Click the **Bar** element icon .
6. In the Bar properties panel, perform these actions:
   - From the Bar Style list, select **Bullet**.

- Open **Variables**, and then highlight Minimum and click the Up arrow once.

7. Click the Graph Builder red triangle and select **Legend Position** > **Bottom**.
8. Right-click on a color in the legend and change the Fill Color:
   - No change to Actual
   - Change Minimum to dark gray
   - Change Target to light gray
9. Right-click in the Y axis > **Order By** > **Target, ascending**.
10. Change the titles of the X axis and graph:
    a. Click the graph title and enter Sales Data vs. Products.
    b. Click the X-axis title and enter Units.
11. Click **Done**.

Figure 4.26 Bullet Chart of Actual, Target, and Minimum Units Sold



The products are ordered by the target number of units as shown by the light gray bars. Products A and D have actual sales that are above their targets; the blue bars extend beyond the light gray target bars. Sales for Product C are on target. Sales for Product B are below target but above the minimum number of units.

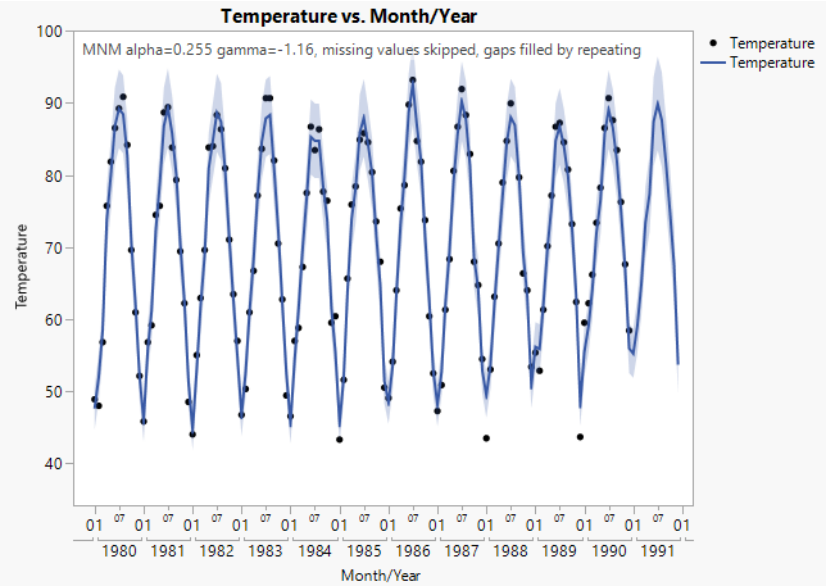
Example of a Forecast Plot

Data	This example uses maximum monthly temperature data from the US city of Raleigh, NC that were recorded from 1980 to 1990.
Techniques	This example uses a time series line of fit.
Goal	The goal of this example is to forecast the maximum monthly temperature for the following year.

1. Select **Help** > **Sample Data Library** and open Time Series/RaleighTemps.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select Month/Year and drag it to the **X** zone.
4. Select Temperature and drag it to the **Y** zone.
5. Click the **Line of Fit** element icon .
6. In the Line of Fit properties panel, do these actions:
   - From the Fit list, select **Time Series**.
   - In the box next to Seasonal Period, type 12.
   - In the box next to Forecast Periods, type 12.

     This forecasts 12 months out. Since the data ends at 1990, it forecasts through 1991.
7. Right-click the X axis and select **Axis Settings**.
   - Under Tick/Bin Increment, change the Increment to 1.
   - Under Tick/Bin Increment, change the Label Row Nesting to 2.
   - Under Axis Label Rows, on the Label Row 1 tab, select **Labels** for the minor tick marks.
   - Click **OK**.
8. (Optional) To save the predicted forecast values and the lower and upper points of the forecast interval as new columns in the data table, click the Line of Fit red triangle and select **Save Formula**.
9. Click **Done**.

Figure 4.27 Forecast Plot of Maximum Monthly Temperature



The graph shows the data points used to estimate the time series function that extends through the forecast period. Note the following about this graph:

- The solid blue line represents the forecast model. The shaded blue area is the 95% confidence interval around the predicted values.
- The forecast period covers 1991, which is 12 months past the last data point. The forecast period is the right most portion of the plot that has no data points.
- The peak forecast high for 1991 is in July with an estimated value of 90°F. This estimate has a confidence interval of about 83°F to 96°F represented by the limits of the blue region above and below the estimate.
- The text at the top indicates the type of model that was fit, the parameter values for that model, and information about missing values. To remove this text from your graph, deselect the Forecast Model option in the Line of Fit properties panel.

Chapter 5

Bubble Plots

View Patterns in Multidimensional Data Using Bubble Plots

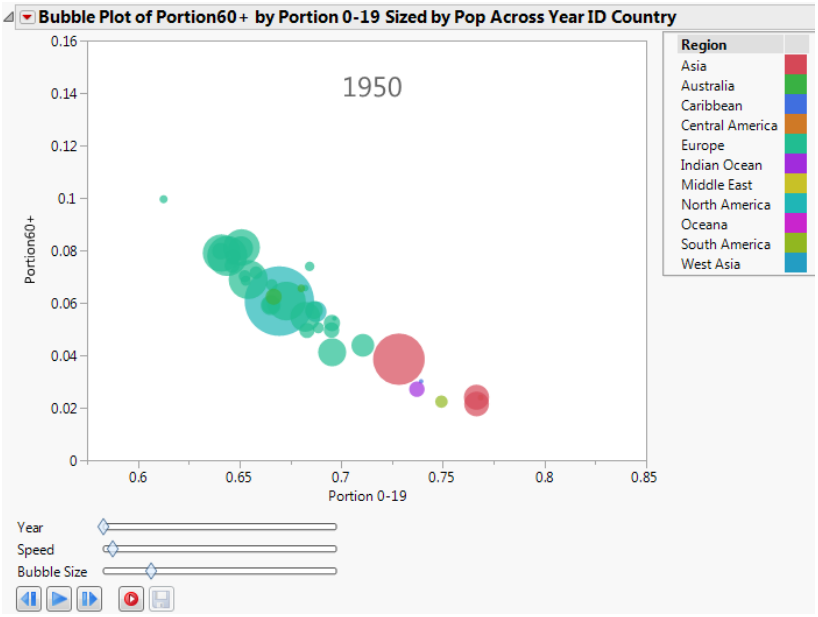
A bubble plot is a scatterplot that represents its points as circles, or bubbles. Bubble plots can be dynamic (animated over time) or static (fixed bubbles that do not move). Use bubble plots to:

- dynamically animate bubbles using a time variable, to see patterns and movement across time
- use size and color to clearly distinguish between different variables
- aggregate data (rows) into a single bubble, simplifying the bubble plot

Because you can see up to five dimensions at once (x position, y position, size, color, and time), bubble plots can produce dramatic visualizations and readily show patterns and trends.

Note: Dynamic bubble plots were pioneered by Hans Rosling, Professor of International Health, Karolinska Institutet, and the people involved in the Gapminder.org project.

Figure 5.1 Example of a Bubble Plot



Contents

Example of a Dynamic Bubble Plot 153
Launch the Bubble Plot Platform 155
Specifying Two ID Variables 156
Specifying a Time Variable 156
Interact with the Bubble Plot 157
Control Animation for Dynamic Bubble Plots 157
Select Bubbles 158
Use the Brush Tool 159
Bubble Plot Platform Options 159
Show Roles 162
Additional Examples of the Bubble Plot Platform 163
Example of Specifying Only a Time Variable 163
Example of Specifying Only ID Variables and Splitting a Bubble 165
Example of a Static Bubble Plot 167
Example of a Bubble Plot with a Categorical Y Variable 171

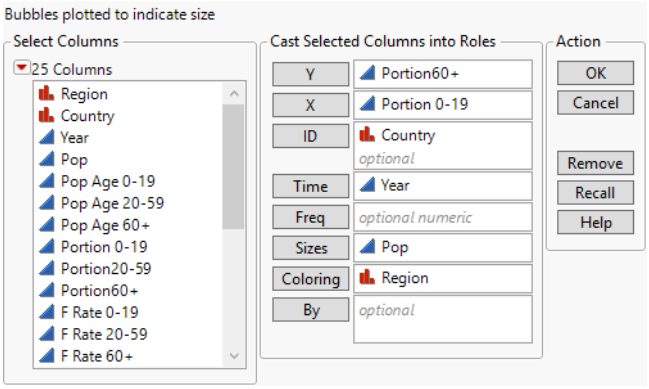
Example of a Dynamic Bubble Plot

This example uses the PopAgeGroup.jmp sample data table, which contains population data for countries and regions around the world. Examine the relationship between the proportion of younger and older people in the sample populations.

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.
2. Select **Graph > Bubble Plot**.

   The launch window appears.

Figure 5.2 The Bubble Plot Launch Window



3. Select Portion60+ and click **Y**.

   The portion of the population that are 60 years or older becomes the *y* coordinate.
4. Select Portion 0-19 and click **X**.

   The portion of the population that are 0-19 years becomes the *x* coordinate.
5. Select Country and click **ID**.

   All the rows for each country are aggregated into a single bubble.
6. Select Year and click **Time**.

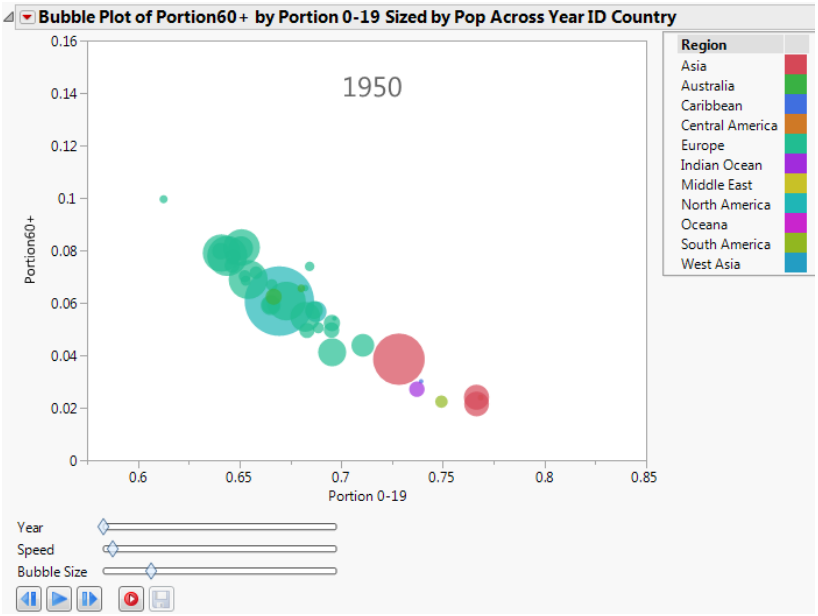
   The bubble plot shows a unique plot for each year's data.
7. Select Pop and click **Sizes**.

   The sizes of the bubbles reflect the overall population values.
8. Select Region and click **Coloring**.

   Bubbles for different regions are assigned different colors. See the Region legend in Figure 5.3. The colors shown in the plot are JMP default colors.
9. Click **OK**.

The report window appears.

Figure 5.3 The Bubble Plot Report Window



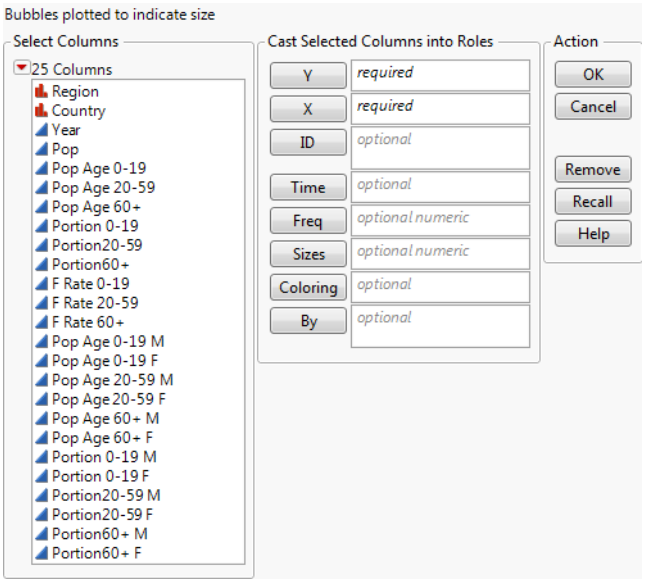
10. Click the play button to see the animated, dynamic report. Alternatively, you can click the previous button to move forward by one year.
11. (Optional) To view a legend that identifies each color with its region, click the Bubble Plot red triangle and select **Legend**.

As time progresses, you can see that the portion of the population that is 0-19 years decreases, and the portion of the population that is 60 years or more increases.

Launch the Bubble Plot Platform

Launch the Bubble Plot platform by selecting **Graph > Bubble Plot**.

Figure 5.4 The Bubble Plot Launch Window



Y, X The Y and X columns become the *y* and *x* coordinates of the bubbles in the plot. These values can be continuous or categorical (nominal or ordinal).

ID ID variables identify rows that should be aggregated and shown as a single bubble. The default coordinates of each bubble are the averaged *x* and *y* values, and the default size of each bubble is the sum of the sizes of all aggregated members. See "Specifying Two ID Variables".

Time Maintains separate coordinates, sizes, and colors for each unique time period. The bubble plot shows these values for a single time period. For example, if the **Time** column contains years, the bubble plot is updated to show data by each year. See "Specifying a Time Variable".

Freq Weights computations when aggregating bubbles using an ID variable.

Note: Negative frequency values are ignored.

Sizes Controls the size of the bubbles. The area of the bubbles is proportional to the **Size** value. There is a minimum bubble size, to keep bubbles visible, even if the size value is

zero. If **Size** is left blank, the default bubble size is proportional to the number of rows in that combination of **Time** and **ID**.

Coloring Colors the bubbles according to the selected variable. If the selected variable is categorical (nominal or ordinal), each category is assigned a unique color. If the selected variable is continuous, a gradient of colors is used. You can set a preference for a color theme by selecting File > Preferences > Graphs and making a selection in the Color Themes panel.

By Place a column here to produce a separate bubble plot for each level of the variable.

After you click **OK**, the Bubble Plot report window appears.

Specifying Two ID Variables

Specifying a second **ID** variable provides a hierarchy of categories, but the bubbles are not split by the second category until they are selected and split interactively. In the launch window, if you specify a second **ID** variable, **Split** and **Combine** buttons appear in the report window.

For example, you might specify a country as the first **ID** variable, resulting in a separate aggregated bubble for each country. A second **ID** variable, perhaps designating regions within each country, would further split each country when the interactive **Split** button under the graph is pressed.

Specifying a Time Variable

Maintains separate coordinates, sizes, and colors for each unique time period. The bubble plot shows these values for a single time period. For example, if the **Time** column contains years, the bubble plot is updated to show data by each year.

To move the time label on the plot, click and drag the label.

If data is missing within a time period, the value is linearly interpolated. If data is missing for the first or last time period, the value is not estimated, but left as missing.

Related Information

- "Control Animation for Dynamic Bubble Plots"
- "Example of Specifying Only a Time Variable"

Interact with the Bubble Plot

Note: If *all* of the rows used in constructing a bubble plot shape are hidden and not excluded, the corresponding bubble plot shape is not shown in the bubble plot. If rows are excluded in the data table, the bubble plot is constructed without the excluded rows.

Use the Bubble Plot platform in one of two modes:

- *Static mode*, where the bubbles are fixed and do not animate over time (no **Time** variable is specified). See "Example of a Static Bubble Plot".
- *Dynamic mode*, where the bubbles are animated over time (a **Time** variable is specified). See "Example of a Dynamic Bubble Plot".

You interact with both static and dynamic bubble plots in different ways.

Control Animation for Dynamic Bubble Plots

Use sliders and buttons to control the animation of dynamic bubble plots.

Figure 5.5 Animation Controls

Table 5.1 Descriptions of the Animation Controls

Slider or Button	Description
<Time variable>	Controls which time values appear in the bubble plot. You manually drag the slider to see a progression of time. Click and drag on the time variable in the bubble plot to move its position. Appears only if you specified a variable for **Time**.
Speed	Adjusts the speed of the animation. Appears only if you specified a variable for **Time**.
Bubble Size	Adjusts the size of the bubbles. The bubbles maintain their relative size, but their absolute size can be adjusted. Appears on all bubble plots.

Table 5.1 Descriptions of the Animation Controls *(Continued)*

Slider or Button	Description
	Adjusts the time value by one unit and shows the previous time value. Appears only if you specified a variable for **Time**.
 	Press Play to animate the bubble plot. Moves through all of the time values in order, and loops back to the beginning when the last time period is reached. Press Pause to stop the animation. Appears only if you specified a variable for **Time**.
	Adjusts the time value by one unit and shows the next time value. Appears only if you specified a variable for **Time**.
	Records the animation (Windows only).
	Saves the animation as an animated GIF file (Windows only).
Split	Separates the bubble represented by the first, larger **ID** variable into its smaller constituent parts, which are defined by the second, smaller **ID** column. Select the bubble and click **Split**. Appears only if you specified two **ID** variables.
Combine	Reverses the action of the **Split** button by recombining the smaller bubbles back into their original bubble. Select any of the smaller bubbles in the group and click **Combine**. Appears only if you specified two **ID** variables.

Select Bubbles

Click a bubble to select it. Note the following:

- Visually, selected bubbles become darker or brighter, and non-selected bubbles are more transparent.
- If the bubble was not filled initially, selection fills it.
- If no bubbles are selected, all of the bubbles are semi-transparent.

When you select a bubble, all of the rows in the data table that correspond to the selected bubble are highlighted. Note the following:

- If the bubble is an aggregate based on an **ID** column, all of the rows for that **ID** are highlighted. Otherwise, the one row represented by that bubble is highlighted.
- If you specify an **ID** and a **Time** variable, selecting a bubble highlights all of the rows for that **ID**, across all of the **Time** levels.

If you select a row from the data table, it is selected in the associated bubble plot. Note the following:

- If you have not specified a **Time** variable, selecting one row from the data table highlights the corresponding bubble in the plot.
- If you have specified a **Time** variable, selecting one row from the data table highlights the corresponding bubble for only that time period in the dynamic bubble plot.

Use the Brush Tool

Use the brush tool to temporarily select bubbles and obtain more information about the selected bubbles. When you select bubbles with the brush tool, the corresponding rows are highlighted in the associated data table.

Note: For a more granular examination of the highlighted rows, use the **Tables > Subset** command or the Row Editor. See *Using JMP*.

Bubble Plot Platform Options

The Bubble Plot red triangle menu provides the following options:

Draw Applies a fill or outline.

Filled Fills all of the bubbles.

Outlined Outlines all of the bubbles.

Filled and Outlined Fills and outlines all of the bubbles.

Set Shape Change the shape of the bubble.

You can create a custom shape using JSL. The **Custom** option opens the custom shape. If no custom shape has been created, the **Custom** option uses the default circle shape. For more information about creating custom shapes, see the *Scripting Guide*.

Orient Shapes Orients the shapes as they move in particular directions over time, following the shape of the data.

This option appears only if you have specified a variable for **Time**.

Trail Bubbles Shows the past history of bubbles as a semi-transparent trail. See "Example of Specifying Only a Time Variable".

Note the following:

- This option appears only if you have specified a variable for **Time**.
- If you do not want to see the bubble labels, select the **Label > None** option.

Trail Lines Shows the past history of bubbles as connected line segments. See "Example of Specifying Only a Time Variable".

Note the following:

- This option appears only if you have specified a variable for **Time**.
- If you do not want to see the bubble labels, select the **Label > None** option.

Label Changes the labels on the plot.

None Labels none of the bubbles in the plot.

All Labels all of the bubbles in the plot.

Selected Labels bubbles only when you select them.

Note: Click and drag on a label to move it.

Color Theme Change the colors representing the high, middle, and low values of the color variable.

This option appears only if you have specified a variable for **Coloring**.

Revert Color Theme Reverts back to the original color theme.

This option appears only if you have applied a color theme.

Legend Shows a legend that describes the colors in the bubble plot.

This option appears only if you have specified a variable for **Coloring**.

Selectable Across Gaps If a bubble is selected, this option keeps the bubble selected during time periods where data is missing. Otherwise, the bubble is not selected during time periods where data is missing.

Show Roles Shows the variables that are used in the bubble plot. You can change and delete the variables. See "Show Roles".

Split All Splits all bubbles into their constituent parts. Unlike the **Split** button, the bubbles do not have to be selected.

This option appears only if you have specified two **ID** variables.

Combine All Combines all constituent bubbles within a group into their larger bubble. Unlike the **Combine** button, the bubbles do not have to be selected.

This option appears only if you have specified two **ID** variables.

Show Time Annotation In bubble plots that contain a time variable, the Show Time Annotation option shows or hides the time that is displayed on the bubble plot.

Lock Scales Prevents axis scales and gradient legend scales from automatically adjusting in response to data or filtering changes.

Fit to Window Determines whether the plot is resized as you resize the JMP window. The default setting is Auto, which bases the scaling on the contents of the plot. For example, a plot with By variables does not stretch to fit the resized window; the plot extends beyond the viewing area. Change the setting to On to always fit the plot inside the window. Change the setting to Off to prevent the plot from resizing.

Aggregation Options Alters how the **X**, **Y**, and **Sizes** roles are computed. By default, the values are calculated using means for **X** and **Y**, and sums for **Sizes**.

X as Sum or Y as Sum Computes the **X** and **Y** values using sums.

Size as Sum Deselecting this option computes **Size** values using means.

Color as Sum Computes the sum of the data values and maps to a color. This option appears only for continuous variables.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

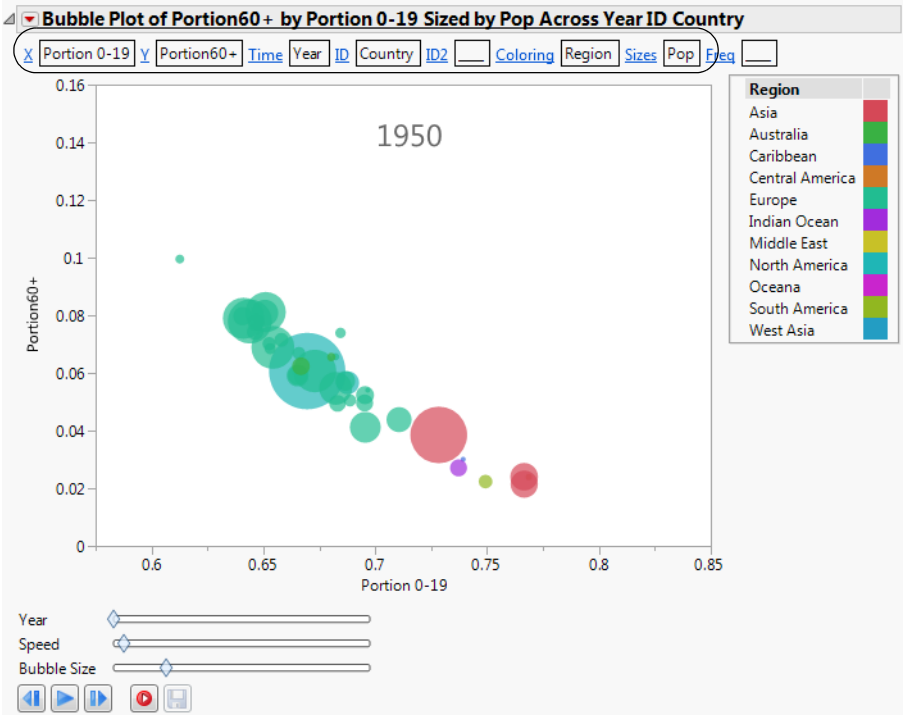
Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Show Roles

Using the **Show Roles** option in the red triangle menu, you can make changes to your existing variables without having to relaunch the platform and start your analysis over.

Follow the instructions in "Example of a Dynamic Bubble Plot" to produce the report window shown in Figure 5.6. (The colors are the JMP default colors.)

Figure 5.6 Example of Bubble Plot with Show Roles Selected



Change the Variable Assigned to a Role

To change the variable assigned to a role, click a blue underlined role name. For example, perform these steps to change the **Coloring** variable in Figure 5.6:

1. Click the **Coloring** link.

   The Select column for Coloring window appears.

2. Click Country.
3. Click **OK**.

   Country now replaces Region as the **Coloring** variable in the bubble plot.

Remove a Variable

To remove an existing variable from the bubble plot, make sure that nothing is selected in the Select column for <*Role*> window, and click **OK**. For example, perform these steps to remove the **Sizes** variable (Pop) in Figure 5.6:

1. Click the Sizes link.

   The Select column for Sizes window appears.
2. Ensure that nothing is selected. If a variable is selected, deselect it by holding down the Ctrl key and clicking on the variable.
3. Click **OK**.

   The Sizes role now appears with an empty box.

Note: The X and Y variables can be changed only and cannot be removed.

Add a Variable

Once you have removed an existing variable from the bubble plot, there are two ways to add a new variable:

- Click the blue underlined role name. See "Change the Variable Assigned to a Role".
- In the data table, click the variable in the column panel, and drag it into the empty role box.

Additional Examples of the Bubble Plot Platform

- "Example of Specifying Only a Time Variable"
- "Example of Specifying Only ID Variables and Splitting a Bubble"
- "Example of a Static Bubble Plot"
- "Example of a Bubble Plot with a Categorical Y Variable"

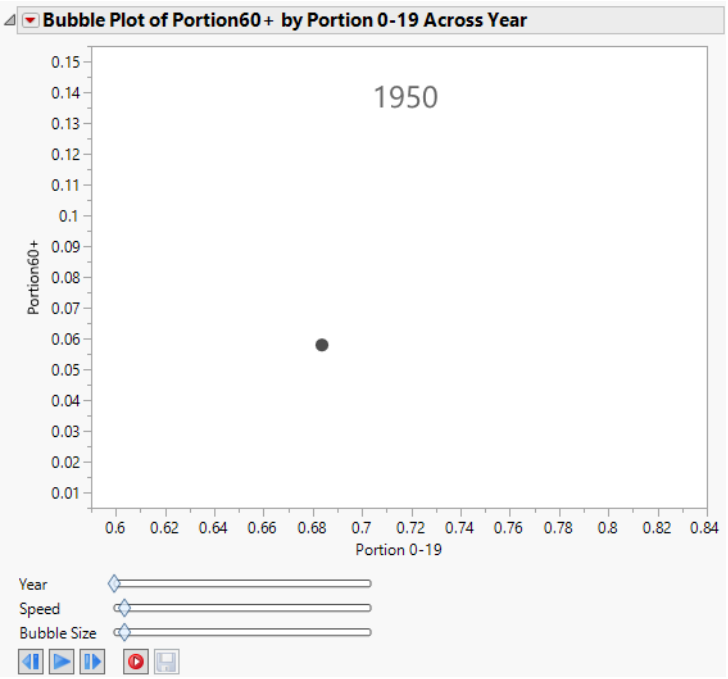
Example of Specifying Only a Time Variable

For dynamic bubble plots, you might specify only a **Time** variable and no **ID** variable. The resulting bubble plot contains a single moving bubble that tracks the series as the **Time** value changes.

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.
2. Select **Graph > Bubble Plot**.

3. Select Portion60+ and click **Y**.
4. Select Portion 0-19 and click **X**.
5. Select Year and click **Time**.
6. Click **OK**.

Figure 5.7 The Initial Report Window with a Time Variable

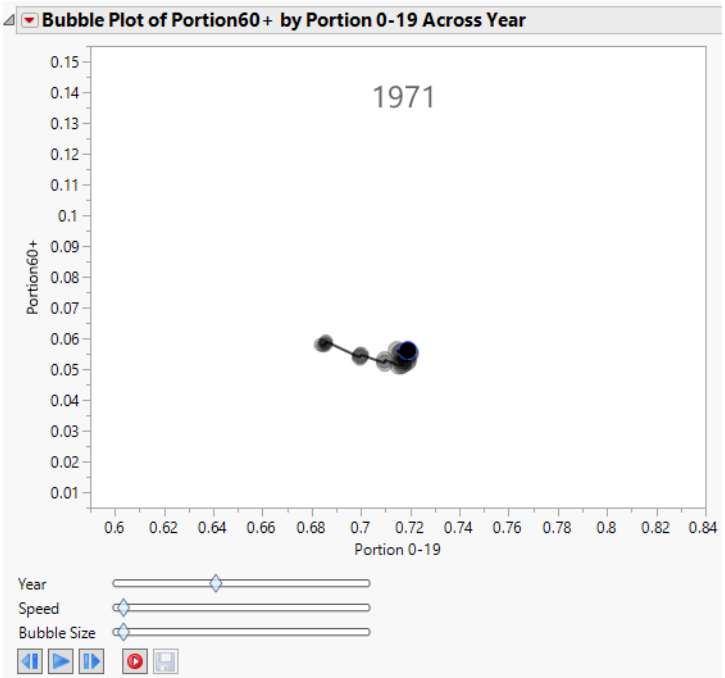


7. Click the bubble to select it.

   All rows in the data table are also highlighted.
8. Click the Bubble Plot red triangle and select **Trail Bubbles > All** and **Trail Lines > All**.
9. Click the play button.

   The bubble plot animates, showing a trail for the single bubble.

Figure 5.8 Animated Bubble Plot over Time



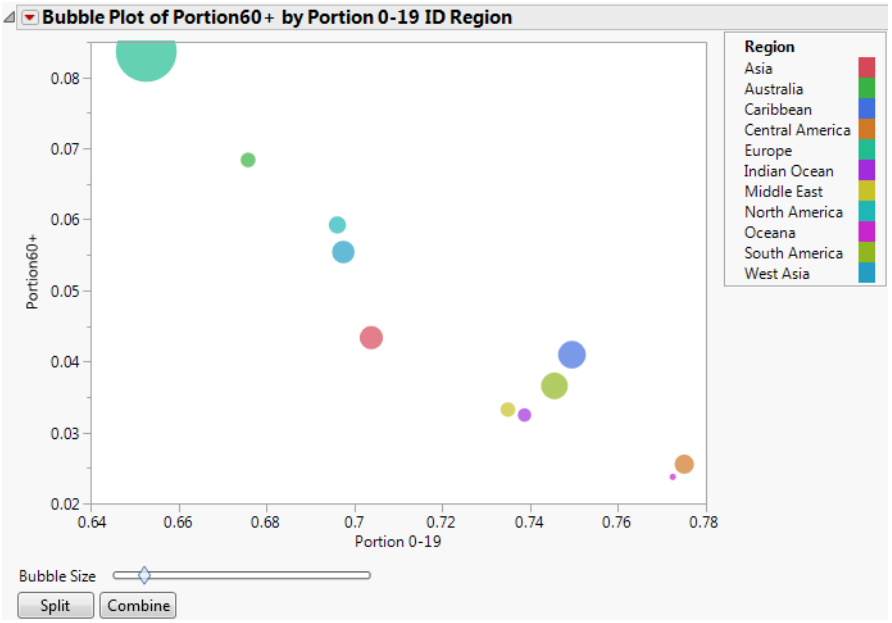
Example of Specifying Only ID Variables and Splitting a Bubble

For static bubble plots, you might specify one or two **ID** variables and no **Time** variable. The resulting bubble plot contains a bubble at each **ID** value. Note that although this bubble plot is static, you can perform splitting on bubbles.

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.
2. Select **Graph > Bubble Plot**.
3. Select Portion60+ and click **Y**.
4. Select Portion 0-19 and click **X**.
5. Select Region and Country and click **ID**.
6. Select Region and click **Coloring**.
7. Click **OK**.

   The initial report window appears. (Figure 5.9 uses the default JMP colors.)

Figure 5.9 Initial Report Window with ID Variables

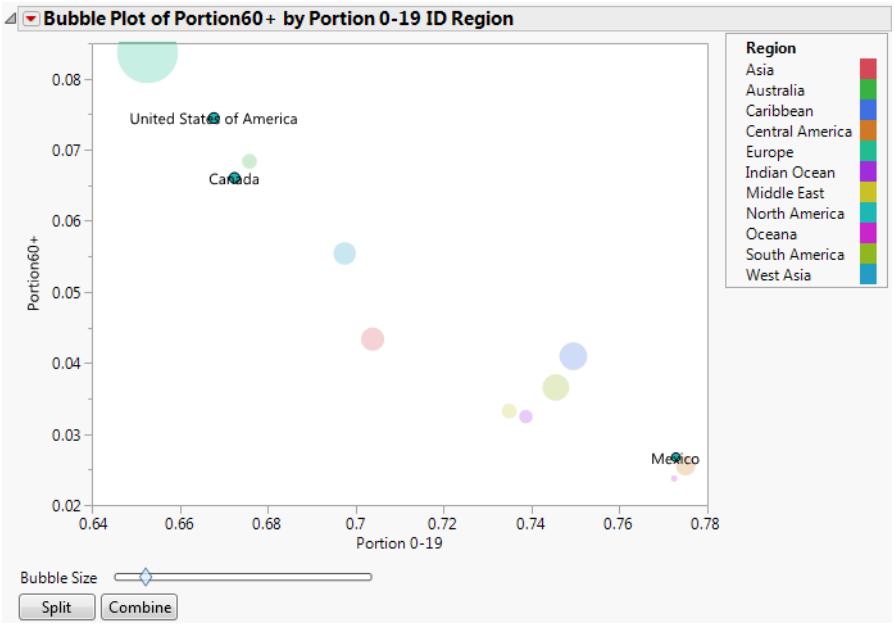


Split the bubble representing the region of North America into countries.

8. Click the bubble representing North America (hover over a bubble to see its label, or use the legend to find the color of North America.)
9. Click **Split**.

   You see that the North America bubble has split into three bubbles, representing the countries within the region of North America (the United States of America, Canada, and Mexico).

Figure 5.10 Splitting the North America Bubble



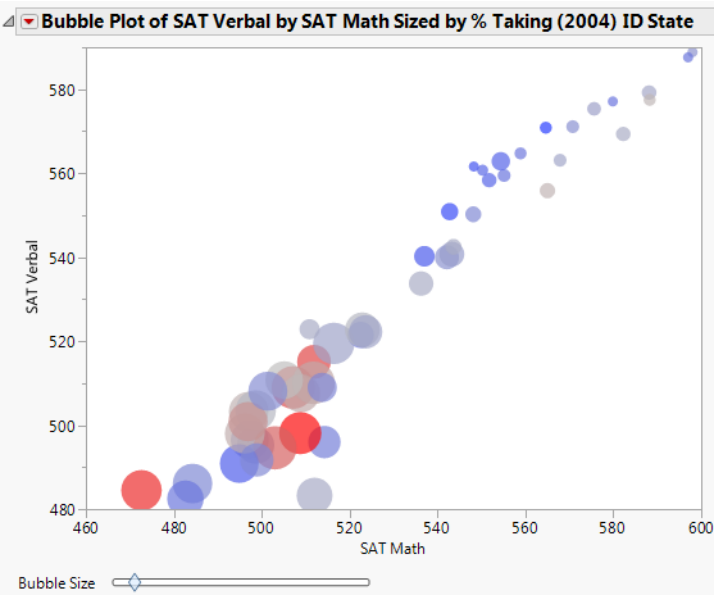
Example of a Static Bubble Plot

This example uses the SATByYear.jmp sample data table, which contains SAT verbal and math test scores for a selection of the US population in 2004.

1. Select **Help > Sample Data Library** and open SATByYear.jmp.
2. Select **Graph > Bubble Plot**.
3. Select SAT Verbal and click **Y**.
4. Select SAT Math and click **X**.
5. Select State and click **ID**.
6. Select % Taking (2004) and click **Sizes**.
7. Click **OK**.

   The report window appears. (Figure 5.11 uses the default JMP colors.)

Figure 5.11 The Static Bubble Plot Report Window



You draw the following conclusions:

- Higher verbal scores appear to be associated with higher math scores, since the two track very closely in the bubble plot. This signifies a correlation between verbal and math scores.
- The larger bubbles represent the US states that have a high percentage of individuals taking the SAT test in 2004. These larger bubbles are all grouped together in the lower left of the graph. This shows that when a state has a high percentage of individuals taking the test, both the math and verbal scores are low.

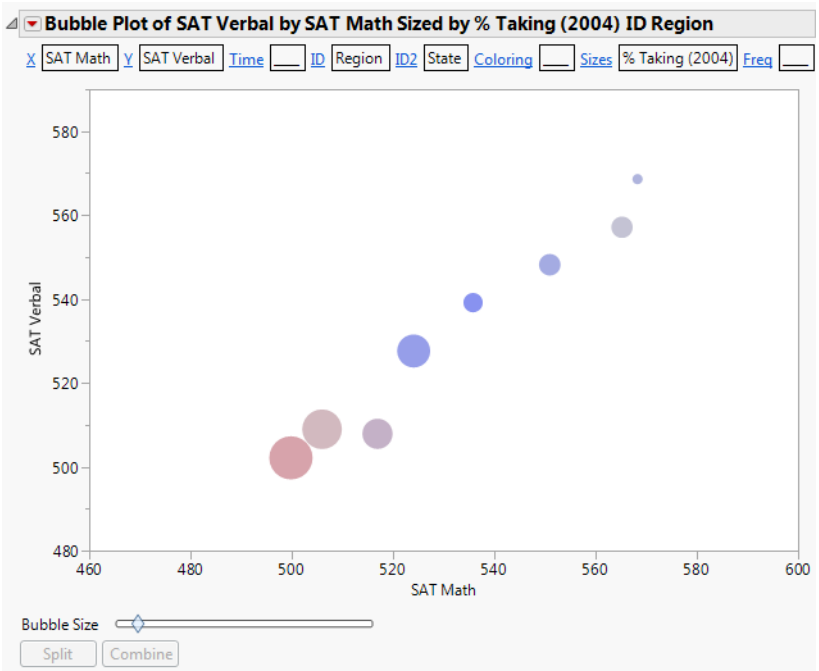
Instead of grouping the bubbles primarily by state, group the bubbles primarily by region:

1. Click the Bubble Plot red triangle and select **Show Roles**.
2. Click the **ID** link.
3. Select Region and click **OK**.

   Region is now the primary **ID** variable.
4. Click the **ID2** link.
5. Select State and click **OK**.

   State is now the secondary **ID** variable.

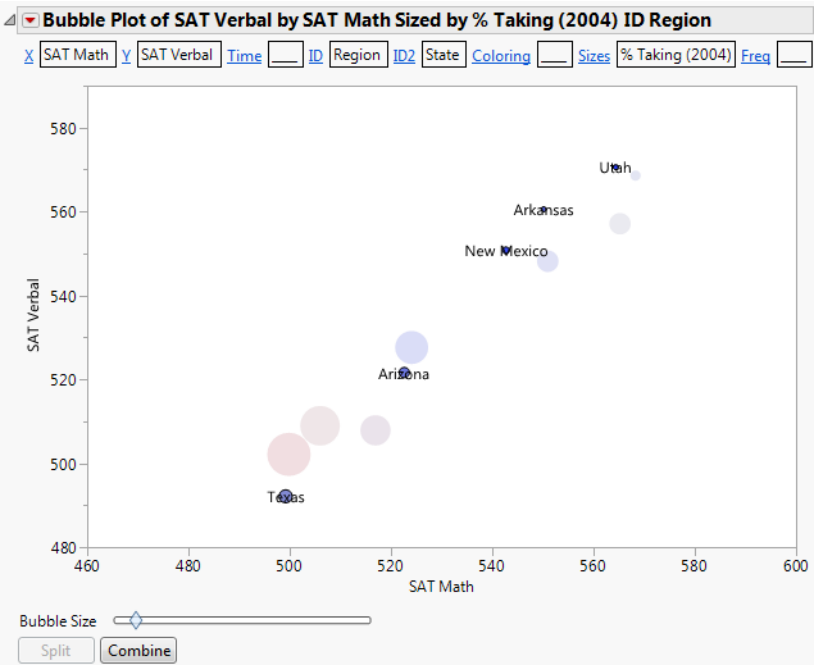
Figure 5.12 Example of Bubble Plot Grouped by Region and State



6. Click the bubble that represents the Southwest region (hover over a bubble or click it to see its label).
7. Click **Split**.

   Now the bubbles are split by the secondary **ID** variable, which is State. You now see each state within the Southwest region.

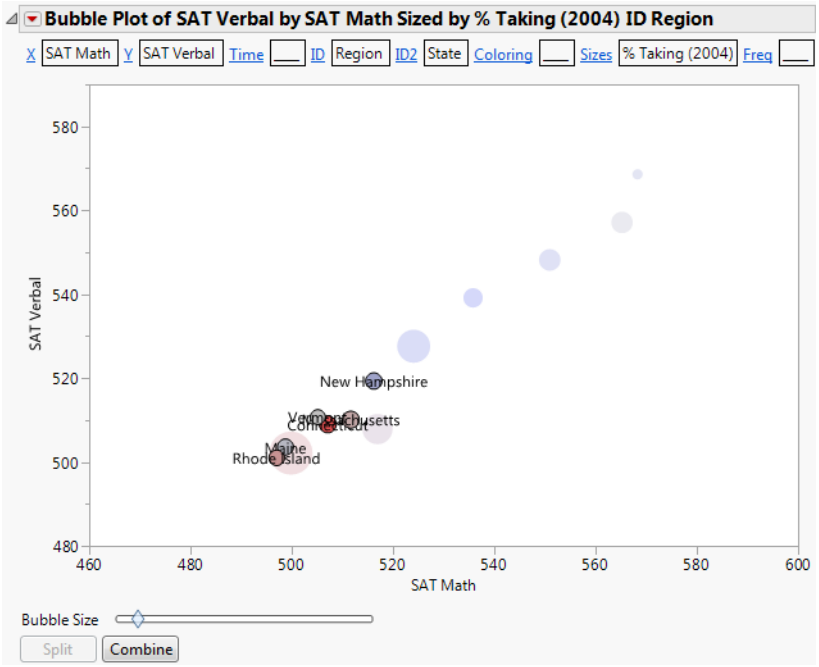
Figure 5.13 Example of Southwest Region Split by State



You see that there is significant variation between the scores from the Southwest states.

8. Click **Combine** to combine the southwest states again.
9. To do a comparison, click the New England bubble (hover over a bubble or click it to see its label).
10. Click **Split**.

Figure 5.14 Example of New England Region Split by State



You see that the New England states do not have as much variation as the Southwest states.

Example of a Bubble Plot with a Categorical Y Variable

All of the examples shown so far use continuous Y variables. If you use a categorical (nominal or ordinal) Y variable, the bubble plot appears differently.

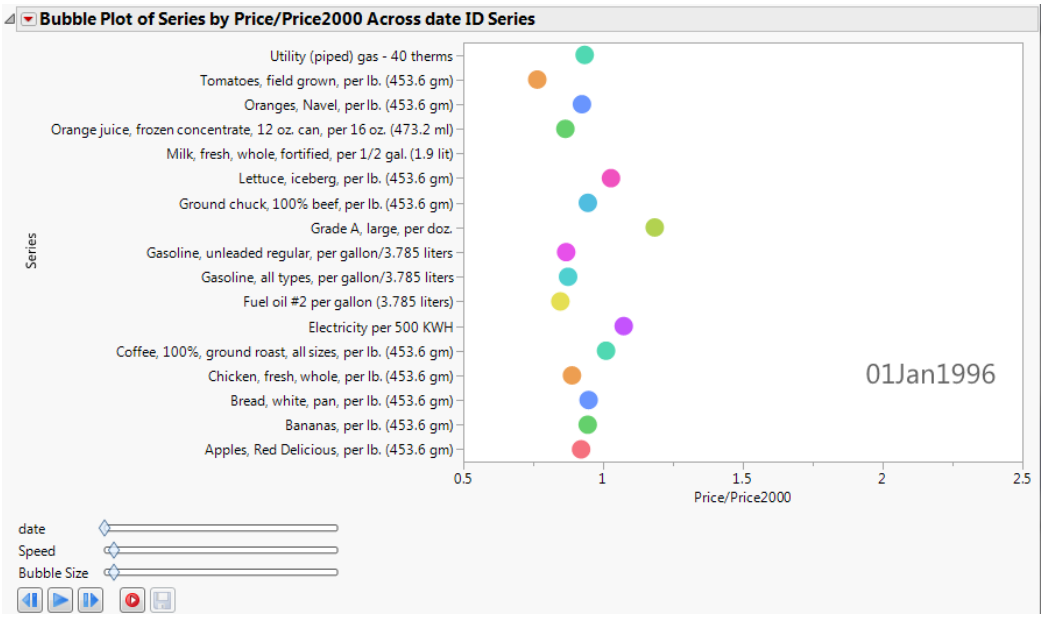
This example uses the blsPriceData.jmp sample data table, which shows the price of commodities over several years. Because the value of the US dollar changes over time, a column named Price/Price2000 shows the ratio of a commodity's price at any given time to the price in the year 2000.

1. Select **Help > Sample Data Library** and open blsPriceData.jmp.
2. Select **Graph > Bubble Plot**.
3. Select Series and click **Y**.
4. Select Price/Price2000 and click **X**.
5. Select date and click **Time**.
6. Click **OK**.

   The report window appears. (Figure 5.15 uses the JMP default colors.)

This produces a bubble plot that, when animated by clicking the play button, shows the price bubbles moving side to side according to their price ratio.

Figure 5.15 Static Example of Animated Bubbles



For easier readability, add grid lines:

7. Double-click the categorical axis.
8. In the Y Axis Settings window, select **Show Grid**.
9. Click **OK**.

To animate the bubble plot, click the play button. The price bubbles move side to side, according to their price ratio.

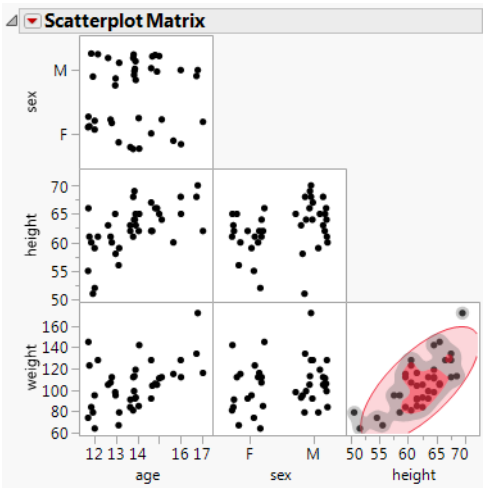
Chapter 6

Scatterplot Matrix

View Multiple Bivariate Relationships Simultaneously

Using the Scatterplot Matrix platform, you can assess the relationships between multiple variables simultaneously. A scatterplot matrix is an ordered collection of bivariate graphs. For further analysis, you can customize the scatterplots with density ellipses for all of your data, or for only groups of your data.

Figure 6.1 Example of a Scatterplot Matrix



Contents

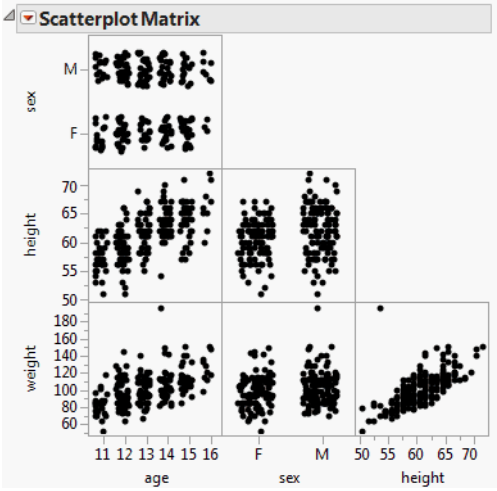
Example of a Scatterplot Matrix 175
Launch the Scatterplot Matrix Platform 176
Change the Matrix Format 178
The Scatterplot Matrix Window 179
Scatterplot Matrix Platform Options 179
Example Using a Grouping Variable 181
Create a Grouping Variable 182

Example of a Scatterplot Matrix

This example shows you how to create a scatterplot matrix.

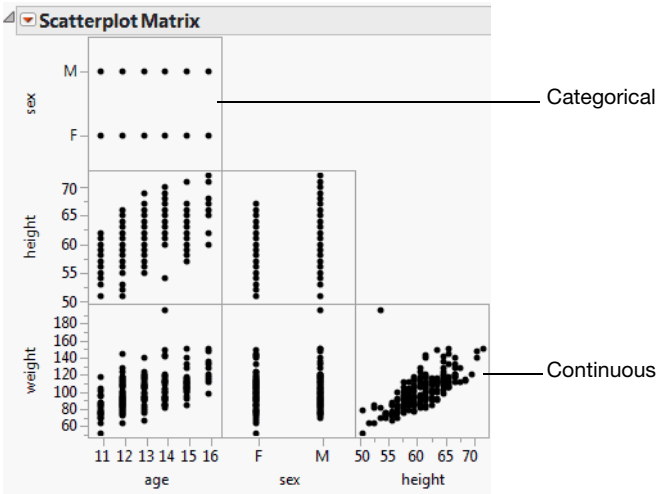
1. Select **Help > Sample Data Library** and open Students.jmp.
2. Select **Graph > Scatterplot Matrix**.
3. Select age, sex, height, and weight and click **Y, Columns**.
4. Click **OK**.

Figure 6.2 Example of a Scatterplot Matrix



In this example, you can see that the graph for weight versus height is different from the graph for sex versus age. If you turn off jitter by clicking on the red triangle menu and selecting **Points Jittered**, the difference becomes even more pronounced.

Figure 6.3 Example of a Scatterplot Matrix with No Jitter

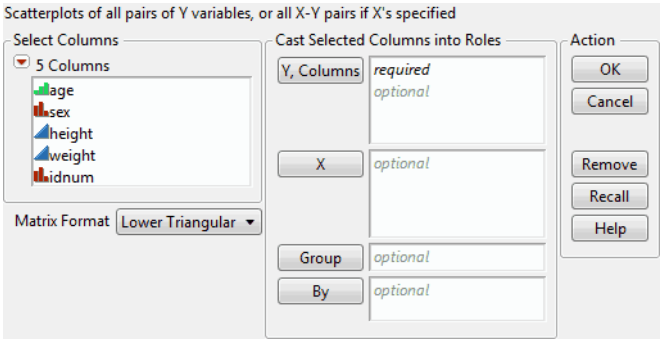


The weight versus height graph shows continuous data, and the sex versus age graph shows categorical data.

Launch the Scatterplot Matrix Platform

Launch the Scatterplot Matrix platform by selecting **Graph > Scatterplot Matrix**.

Figure 6.4 The Scatterplot Matrix Launch Window



Y, Columns, X Specify columns for *Y* and *X*.

- If you assign variables to the **Y, Columns** role only, they appear on both the horizontal and vertical axes.
- If you assign variables to both the **Y, Columns** and **X** role, then the **Y, Columns** variables appear on the vertical axis. The **X** variables appear on the horizontal axis. This

approach enables you to produce rectangular matrices, or matrices that have different, yet overlapping, sets of variables forming the axes of the matrix.

Group If you assign a variable to the **Group** role, you can add shaded density ellipses for each level of the **Group** variable. See "Example Using a Grouping Variable".

By This option produces a separate scatterplot matrix for each level of the **By** variable. If two **By** variables are assigned, a separate graph for each possible combination of the levels of both **By** variables is produced.

Matrix Format The **Matrix Format** can be one of three arrangements: **Upper Triangular**, **Lower Triangular**, or **Square**. See "Change the Matrix Format".

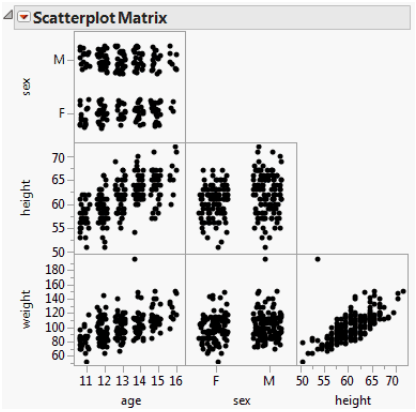
After you click **OK**, the Scatterplot Matrix window appears. See "The Scatterplot Matrix Window".

Change the Matrix Format

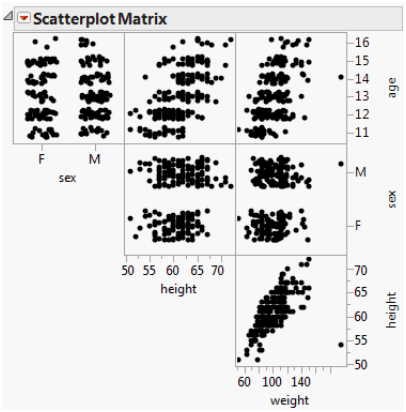
The **Matrix Format** can be one of three arrangements: **Upper Triangular**, **Lower Triangular**, or **Square**.

Figure 6.5 Examples of Matrix Formats

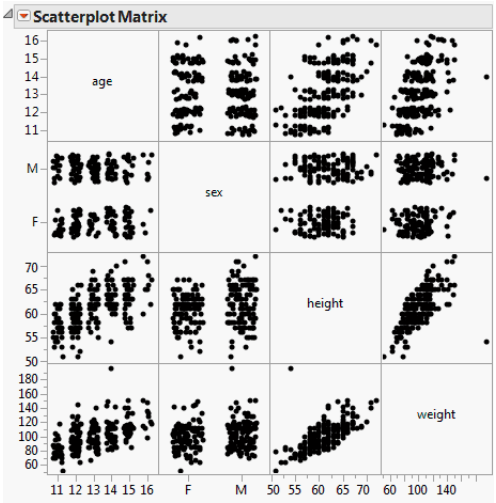
Lower Triangular



Upper Triangular



Square



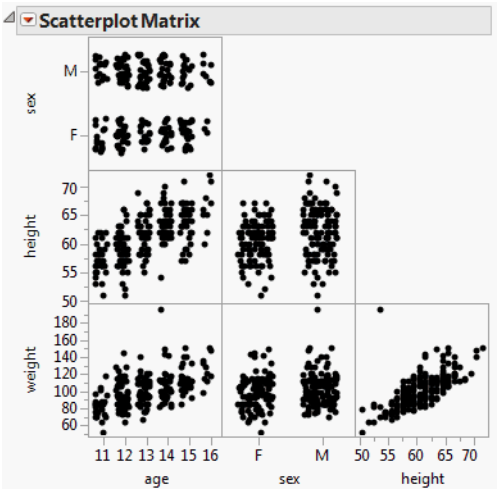
The Scatterplot Matrix Window

The Scatterplot Matrix window shows an ordered grouping of bivariate graphs. In each graph, you can examine the relationships between each pair of variables.

Follow the instructions in "Example of a Scatterplot Matrix" to produce the plot shown in Figure 6.6.

Note: For information about additional options, see "Scatterplot Matrix Platform Options".

Figure 6.6 Example of a Scatterplot Matrix Window



Replace variables in the plot by dragging and dropping a variable, in one of two ways: swap existing variables by dragging and dropping a variable from one axis to the other axis; or, click a variable in the Columns panel of the associated data table and drag it onto an axis. This feature is not available for matrices in the Square format.

Scatterplot Matrix Platform Options

The Scatterplot Matrix red triangle menu provides the following options:

Show Points Shows or hides the points in the scatterplots.

Points Jittered Turns the jittering of the points in the scatterplot on or off. This option is available when at least one variable is either ordinal or nominal.

Fit Line Fits a simple regression line and its mean confidence interval to the scatterplots.

Density Ellipses Shows or hides the outline and area of the density ellipses. See "Example Using a Grouping Variable".

Shaded Ellipses Colors the area within each ellipse. See "Example Using a Grouping Variable".

Ellipses Coverage Enables you to select an α-level for the ellipses to cover.

Ellipses Transparency Enables you to select the transparency of the shaded ellipses, where 0 is completely transparent and 1 is completely opaque.

Ellipses Color Enables you to select a color for the outline and the area within an ellipse.

Nonpar Density Shows or hides the nonparametric density, which represents the areas where the data points are the most dense. The nonparametric density estimation is helpful when you have a lot of points and the density of the points is difficult to see.

There are two quantile density contours. One contour includes 50% of the smoothed density, and the other contour includes 90% of the smoothed density. Since the percentage is based on the smoothed density, the percentage may not agree with the actual proportion of points within a contour.

Group By In the Group By window, you can perform the following actions:

- If you did not select a **Group** variable in the launch window, you can add one now.
- If you did select a **Group** variable in the launch window, you can remove the existing **Group** variable, or you can replace the **Group** variable.

See "Example Using a Grouping Variable".

Lock Scales Prevents axis scales and gradient legend scales from automatically adjusting in response to data or filtering changes.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

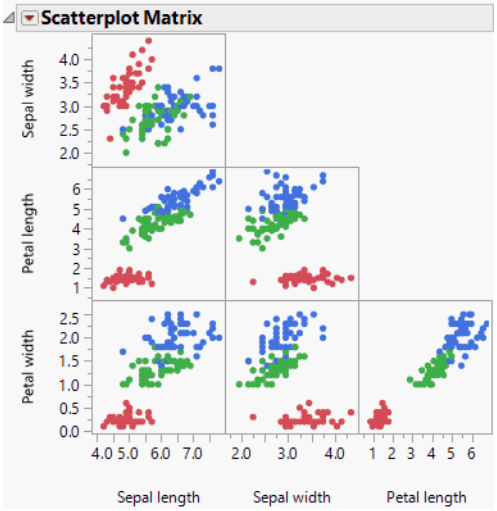
Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Example Using a Grouping Variable

This example shows you how to create a scatterplot matrix using a grouping variable.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot Matrix**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Select Species and click **Group**.
5. Click **OK**.

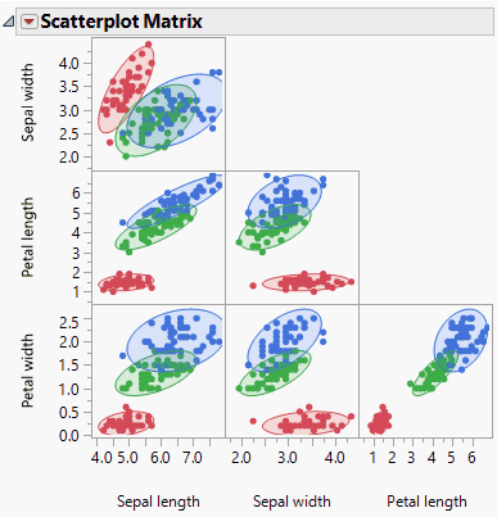
Figure 6.7 Initial Example Using a Grouping Variable



Make the groupings stand out by adding ellipses and shading:

6. Click the Scatterplot Matrix red triangle and select **Density Ellipses > Density Ellipses**.
7. Click the Scatterplot Matrix red triangle and select **Density Ellipses > Shaded Ellipses**.

Figure 6.8 Example of a Scatterplot Matrix with Ellipses



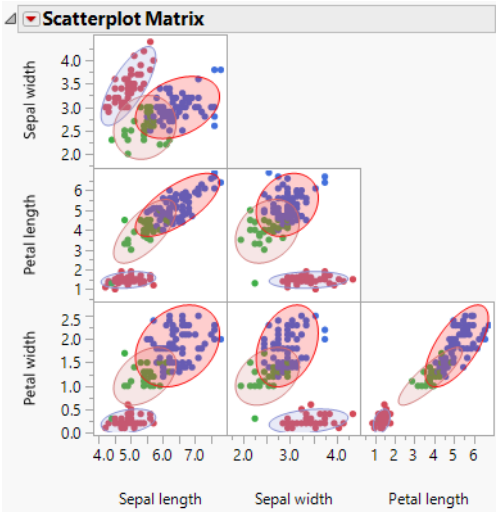
Create a Grouping Variable

If your data does not already have a grouping variable, you can create one using the Cluster platform. Using the Iris.jmp data, assume that the Species column does not exist. You know that the data comes from three species of Iris flowers, so you want to create three clusters within a group.

1. Using the Iris.jmp sample data table, select **Analyze > Clustering > Hierarchical Cluster**.
2. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
3. Click **OK**.
4. Click the Hierarchical Clustering red triangle and select **Number of Clusters**.
5. Type 3 to represent the three different Iris species.
6. Click **OK**.
7. Click the Hierarchical Clustering red triangle and select **Save Clusters**.
8. Close the Hierarchical Cluster report window, and go back to the Iris.jmp data table.

   You can see that a Cluster column has been added to the Iris.jmp data table.
9. Perform the Scatterplot Matrix analysis. Follow the instructions in "Example Using a Grouping Variable", but use Cluster as the grouping variable.

Figure 6.9 Example of a Scatterplot Matrix Using a Cluster Variable



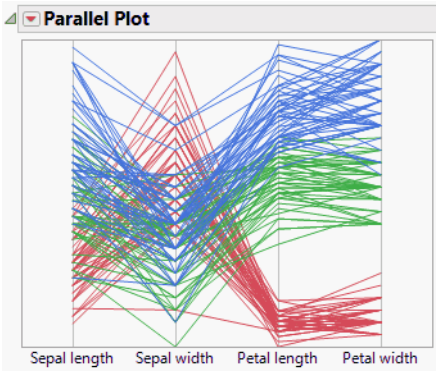
Chapter 7

Parallel Plots

View Patterns in Multidimensional Data by Plotting Parallel Coordinates

Using parallel plots, you can visualize each cell in a data table. Parallel plots draw connected line segments that represent each row in a data table. Parallel plots were initially developed by Inselberg (1985) and later popularized by Wegman (1990).

Figure 7.1 Example of a Parallel Plot



Contents

Example of a Parallel Plot 187
Launch the Parallel Plot Platform 189
The Parallel Plot 190
Interpreting Parallel Plots 190
Parallel Plot Platform Options 192
Additional Examples of the Parallel Plot Platform 193
Examine Iris Measurements 193
Examine Student Measurements 194

Example of a Parallel Plot

This example uses the Dogs.jmp sample data table, which contains histamine level measurements for 16 dogs that were given two different drugs. The histamine levels were taken at zero, one, three, and five minutes. Examine the variation in the histamine levels for each drug.

1. Select **Help > Sample Data Library** and open Dogs.jmp.

   To see the differences by drug, color the parallel plot lines by drug:

2. Select **Rows > Color or Mark by Column**.
3. Select drug.

   If the selection shown in your Colors menu is **JMP Default**, morphine is assigned the color red and trimeth is assigned the color blue.

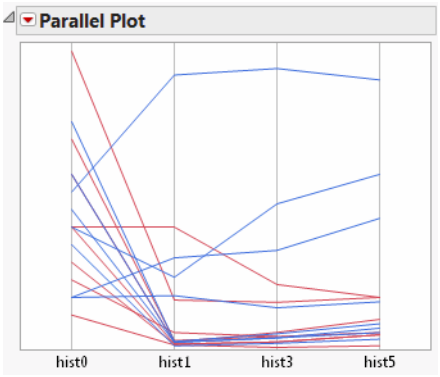
4. Click **OK**.

Create the parallel plot:

5. Select **Graph > Parallel Plot**.
6. Select hist0, hist1, hist3, and hist5 and click **Y, Response**.
7. Click **OK**.

   The report window appears.

Figure 7.2 Parallel Plot of Histamine Variables



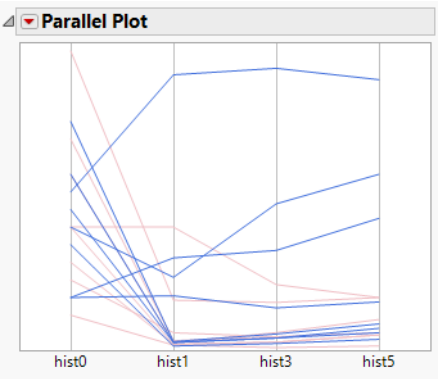
Each connected line segment represents a single observation. Click a line segment to see which observation (or row) it corresponds to in the data table.

For further exploration, isolate the trimeth values:

8. Select **Rows > Data Filter**.
9. Select drug and click Add .
10. Select trimeth.

    Only the trimeth values are highlighted in the parallel plot.

Figure 7.3 Trimeth Values Highlighted



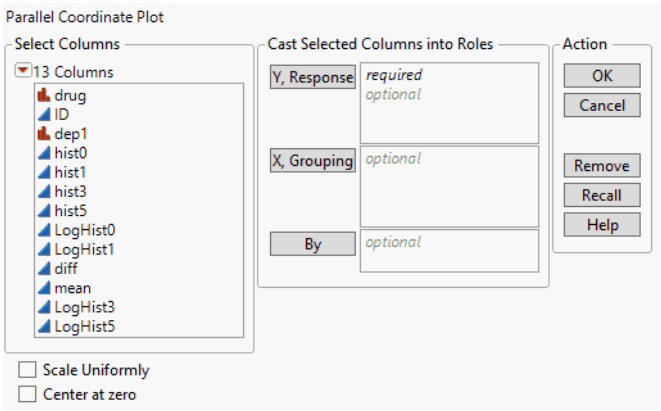
You observe the following about the histamine levels for dogs given trimeth:

- For most of the dogs, the histamine levels had a sharp drop at one minute.
- For four of the dogs, the histamine levels remained high, or rose higher. You might investigate this finding further, to determine why the histamine levels were different for these dogs.

Launch the Parallel Plot Platform

Launch the Parallel Plot platform by selecting **Graph > Parallel Plot**.

Figure 7.4 The Parallel Plot Launch Window



Y, Response Variables appear on the horizontal axis of the parallel plot. These values are plotted and connected in the parallel plot.

X, Grouping Produces a separate parallel plot for each level of the variable.

By Identifies a column that creates a report consisting of separate analyses for each level of the specified variable.

Scale Uniformly Represents all variables on the same scale, adding a *y*-axis to the plot. Without this option, each variable is on a different scale.

To allow for proper comparisons, select this option if your variables are measured on the same scale.

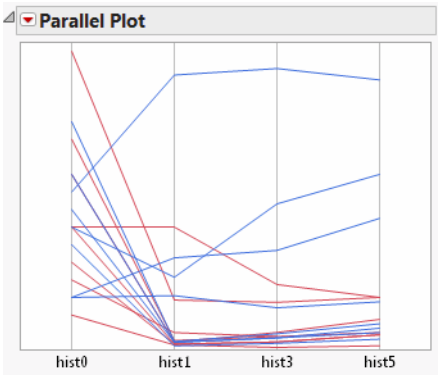
Center at zero Centers the parallel plot (not the variables) at zero.

After you click **OK**, the Parallel plot appears.

The Parallel Plot

To produce the plot shown in Figure 7.5, follow the instructions in "Example of a Parallel Plot".

Figure 7.5 The Parallel Plot Report



A parallel plot is one of the few types of coordinate plots that show any number of variables in one plot. However, the relationships between variables might be evident only in the following circumstances:

- when the variables are side by side
- if you assign a color to a level of a variable to track groups
- if you select lines to track groups

Tip: You can add reference lines for specification limits. For information, see *Using JMP*.

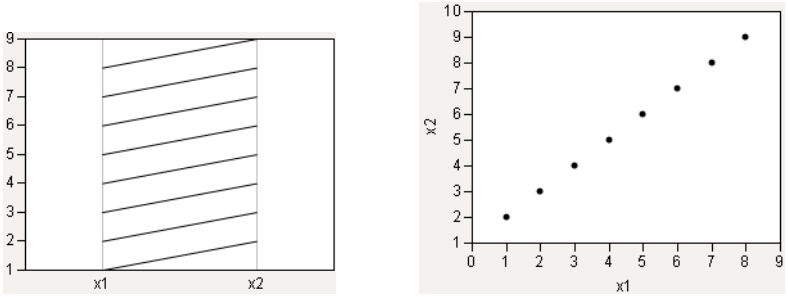
Interpreting Parallel Plots

To help you interpret parallel plots, compare the parallel plot with a scatterplot. In each of the following figures, the parallel plot appears on the left, and the scatterplot appears on the right.

Strong Positive Correlation

The following relationship shows a strong positive correlation. Notice the coherence of the lines in the parallel plot.

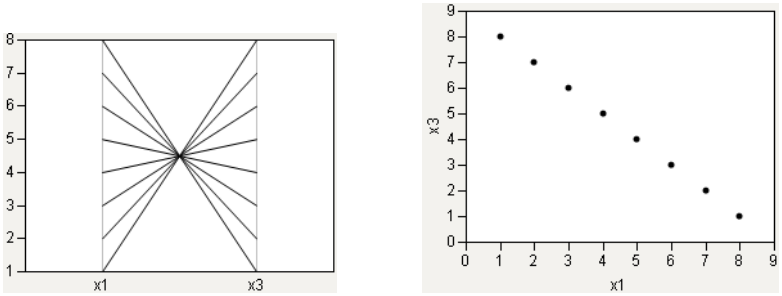
Figure 7.6 Strong Positive Correlation



Strong Negative Correlation

A strong negative correlation, by contrast, shows a narrow neck in the parallel plot.

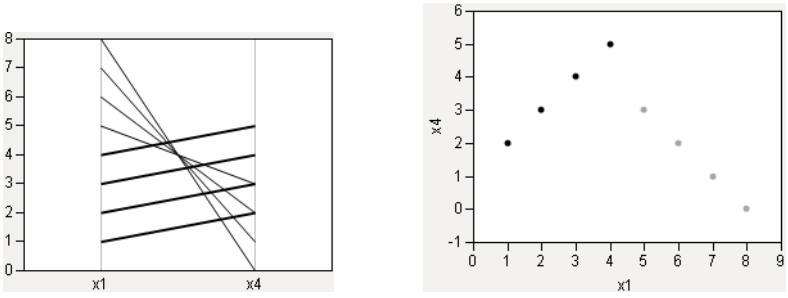
Figure 7.7 Strong Negative Correlation



Collinear Groups

Now, consider a case that encompasses both situations: two groups, both strongly collinear. One has a positive slope, the other has a negative slope. In Figure 7.8, the positively sloped group is highlighted.

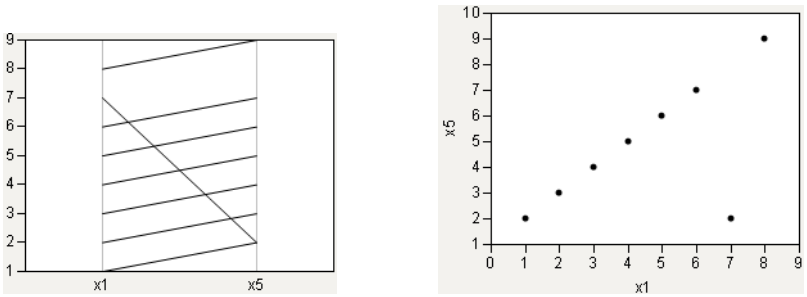
Figure 7.8 Collinear Groups: Parallel Plot and Scatterplot



Single Outlier

Finally, consider the case of a single outlier. The parallel plot shows a general coherence among the lines, with a noticeable exception.

Figure 7.9 Single Outlier: Parallel Plot and Scatterplot



Related Information

- "Additional Examples of the Parallel Plot Platform"

Parallel Plot Platform Options

The Parallel Plot red triangle menu provides the following options:

Show Reversing Checkboxes Reverses the scale for one or more variables.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

For more information about the context menu options that appear when you right-click a parallel plot, see *Using JMP*.

Additional Examples of the Parallel Plot Platform

- "Examine Iris Measurements"
- "Examine Student Measurements"

Examine Iris Measurements

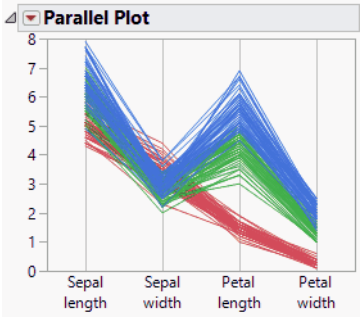
The following example uses the Fisher's Iris data set (Mardia, Kent, and Bibby 1979). The Iris.jmp sample data table contains measurements of the sepal length and width and petal length and width in centimeters for three species of Iris flowers: setosa, versicolor, and virginica. To find characteristics that differentiate the three species, examine these measurements.

Examine Three Species in One Parallel Plot

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Parallel Plot**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Response**.
4. Select the **Scale Uniformly** check box.
5. Click **OK**.

   The report window appears.

Figure 7.10 Three Species in One Parallel Plot



In this parallel plot, the three species are all represented in the same plot. The colors correspond to the three species:

- Blue corresponds to virginica.
- Green corresponds to versicolor.
- Red corresponds to setosa.

From the parallel plot, you observe the following:

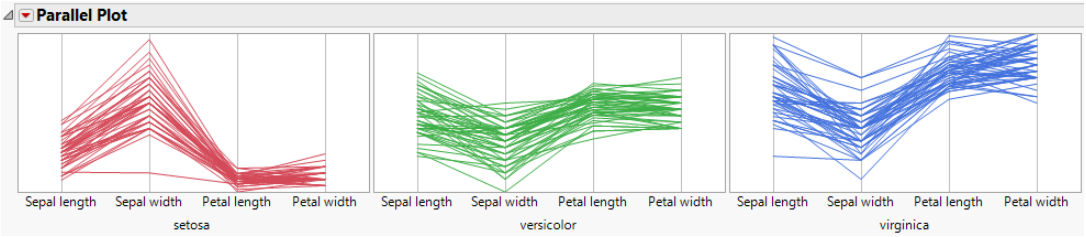
- For sepal width, the setosa values appear to be higher than the virginica and versicolor values.
- For petal width, the setosa values appear to be lower than the virginica and versicolor values.

Examine Three Species in Different Parallel Plots

1. From the Iris.jmp sample data table, select **Graph > Parallel Plot**.
2. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Response**.
3. Select Species and click **X, Grouping**.
4. Click **OK**.

   The report window appears.

Figure 7.11 Three Species in Different Parallel Plots



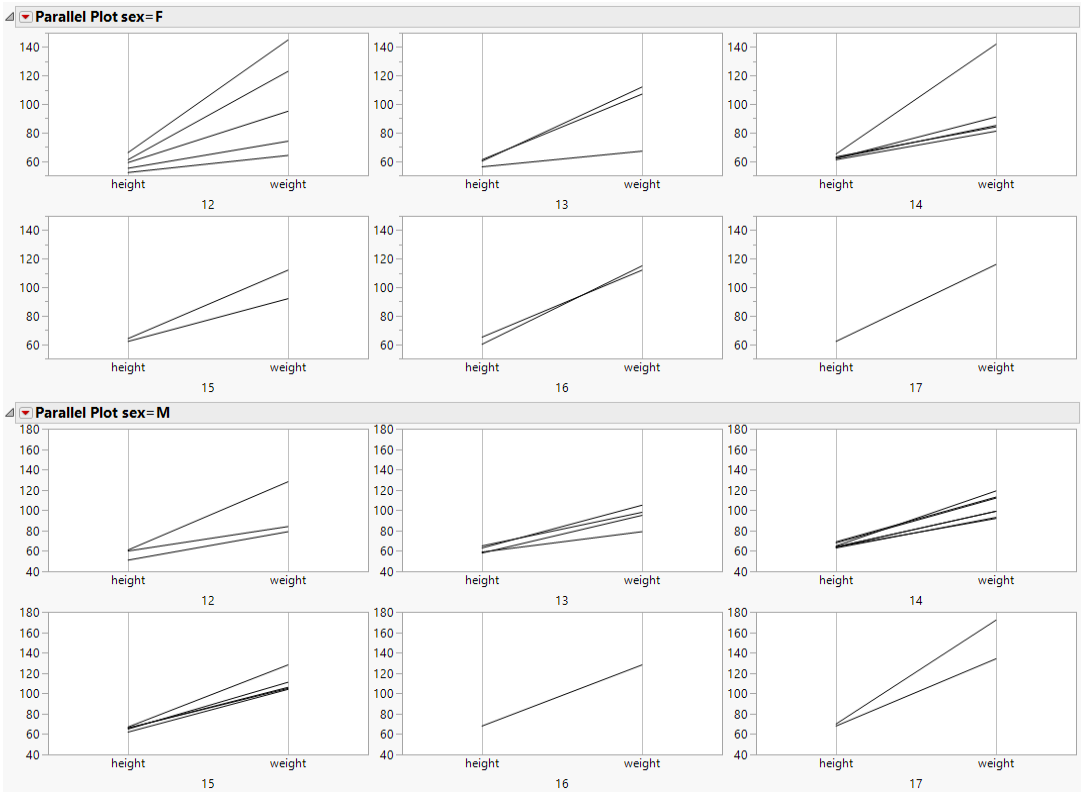
Each species is represented in a separate parallel plot.

Examine Student Measurements

The following example uses the Big Class.jmp sample data table, which contains data on age, sex, height, and weight for 40 students. Examine the relationships between different variables.

1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Select **Graph > Parallel Plot**.
3. Select height and weight and click **Y, Response**.
4. Select age and click **X, Grouping**.
5. Select sex and click **By**.
6. Select the **Scale Uniformly** check box.
7. Click **OK**.

Figure 7.12 Height and Weight by Sex, Grouped by Age



You observe the following:

- Among the 13-year-old females, one female's weight is lower than the other females in her age group. If you click the line representing the lower weight, the respective individual (Susan) is highlighted in the data table.
- Among the 14-year-old females, one female's weight is higher than the other females in her age group. If you click the line representing the higher weight, the respective individual (Leslie) is highlighted in the data table.

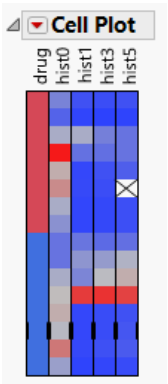
Chapter 8

Cell Plots

View Color-Intensity Plots of Variables

Using cell plots, you can visualize each cell in a data table. Cell plots are direct representations of a data table, since they draw a rectangular array of cells where each cell corresponds to a data table entry. Cell plots were popularized by genomics applications to browse large numbers of values for gene expression levels.

Figure 8.1 Example of a Cell Plot



Contents

Example of a Cell Plot 199
Launch the Cell Plot Platform 200
The Cell Plot 200
Cell Plot Platform Options 201
Right-Click Menu for Cell Plots 202
Additional Example of the Cell Plot Platform 203

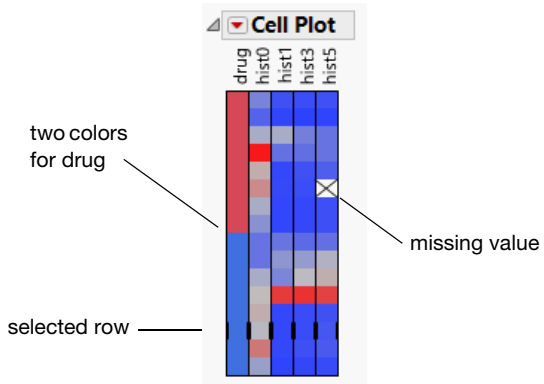
Example of a Cell Plot

Learn how to create and analyze a cell plot in JMP. This example uses the Dogs.jmp sample data table, which contains histamine level measurements for 16 dogs that were given two different drugs. The histamine levels were taken at zero, one, three, and five minutes. Examine the variation in the histamine levels for each drug.

1. Select **Help > Sample Data Library** and open Dogs.jmp.
2. Select the third row from the bottom (row 14).
3. Select **Graph > Cell Plot**.
4. Select drug, hist0, hist1, hist3, and hist5 and click **Y, Response**.
5. Click **OK**.

   The report window appears.

Figure 8.2 Dogs.jmp cell plot



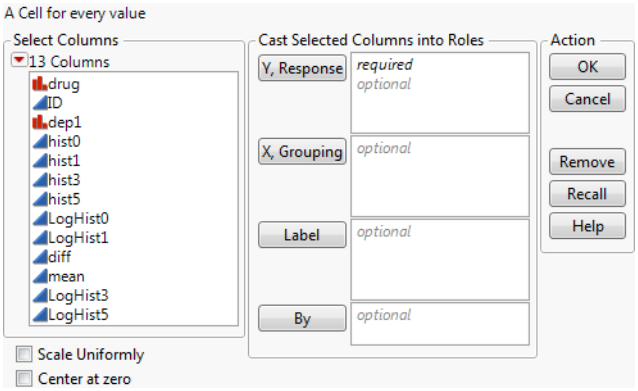
Notice the following:

- There are two types of drugs, represented by two distinct colors.
- Histamine levels are assigned colors from a gradient of blue to red.
- Any missing values are delineated by an X.
- The third row from the bottom is selected, and black lines appear next to the cells.

Launch the Cell Plot Platform

Launch the Cell Plot platform by selecting **Graph > Cell Plot**.

Figure 8.3 The Cell Plot Launch Window



Y, Response Variables appear on the horizontal axis of the cell plot. Each cell represents a value.

X, Grouping Produces a separate cell plot for each level of the variable.

Label Labels each row by the specified variable. See "Additional Example of the Cell Plot Platform".

By Identifies a column that creates a report consisting of separate analyses for each level of the variable.

Scale Uniformly Represents all variables on the same scale. Without this option, each variable is on a different scale.

Center at zero Centers the cell plot at zero.

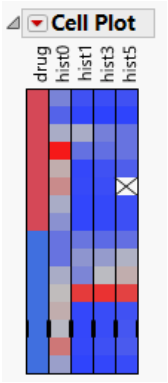
After you click **OK**, the Cell Plot window appears.

The Cell Plot

The cell plot appears with a one-to-one correspondence of a colored cell representing each data table entry. Colors are assigned to each cell based on the range and type of values found in the column.

Note: Any rows that are excluded in the data table are also hidden in the cell plot.

Figure 8.4 The Cell Plot



Note: To produce the plot shown in Figure 8.4, follow the instructions in "Example of a Cell Plot".

Note the following information about cell plots:

- Nominal variables use a distinct color for each level. You can customize nominal and ordinal colors using the **Value Colors** property of data columns, available through the **Column Info** command.
- Continuous variables are assigned a gradient of colors to show the smooth range of values in the variable.
- Ordinal variables are scaled like continuous variables in order.
- When some outliers are present, the scale uses all but the extreme categories for the 90% middle of the distribution, so that the outliers do not overly influence the scale.

Related Information

- "Additional Example of the Cell Plot Platform"

Cell Plot Platform Options

The Cell Plot red triangle menu provides the following options:

Legend Shows or hides a legend.

Arrange Plots Specifies how many plots to put on the same row before starting the next row of plots.

This option is available only if you specify an **X, Grouping** variable.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Right-Click Menu for Cell Plots

Right-click the cell plot to find the following options:

Graph Type Determines the appearance of the graph. See "Change the Graph Type".

Color Theme Shows a list of color themes that affect continuous variables in color maps. The default color theme is **Blue to Gray to Red** (corresponding to small values to middle values to large values). Use **White to Black** to create a gray-scale plot.

Note: To see custom colors, you must first create them. Select **File > Preferences > Graphs**. In the Color Themes area, click the type of color theme that you want to create, click **New**, and then change the colors. See *Using JMP* for more information about creating custom color themes.

Sort Ascending Sorts the rows of the plot from lowest to highest by the values of a column. To sort, right-click in the plot under a column and select **Sort Ascending**. The entire plot is rearranged to accommodate the sorting. See "Additional Example of the Cell Plot Platform".

Sort Descending Sorts the rows of the plot from highest to lowest by the values of a column. To sort, right-click in the plot under a column and select **Sort Descending**. The entire plot is rearranged to accommodate the sorting.

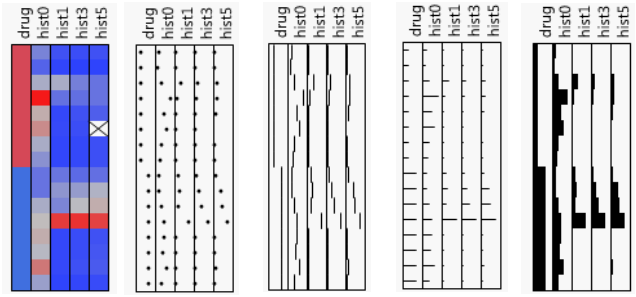
No Separator Lines Draws or removes lines separating the columns.

Note: For more information about the pop-up options that appear when you right-click labels, see *Using JMP*.

Change the Graph Type

Use the **Graph Type** option to change the appearance of the cell plot.

Figure 8.5 Graph Types



Additional Example of the Cell Plot Platform

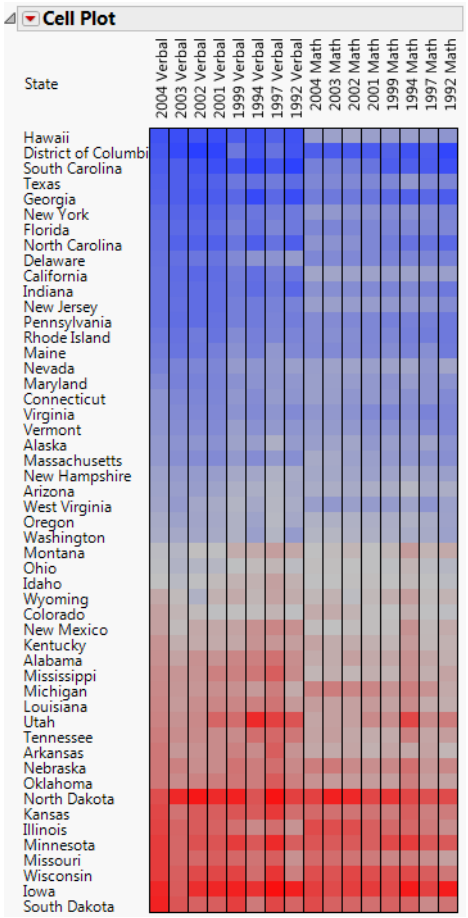
This example uses the SAT.jmp sample data table, which contains SAT test scores (divided into verbal and mathematics portions) for all 50 United States.

1. Select **Help > Sample Data Library** and open SAT.jmp.
2. Select **Graph > Cell Plot**.
3. Select all of the Verbal scores for all of the years, and click **Y, Response**.
4. Select all of the Math scores for all of the years, and click **Y, Response**.
5. Select State and click **Label**.
6. Click **OK**.

   The report window appears.
7. Right-click the plot under 2004 Verbal (the top left cell) and select **Sort Ascending**.

   This sorts the cell plot by the verbal scores for 2004.

Figure 8.6 Cell Plot for SAT Scores



You notice the following:

- Hawaii has the lowest verbal scores for 2004, and South Dakota has the highest verbal scores for 2004.
- There is a contrast between Hawaii's math and verbal scores. Hawaii has average math scores (represented by gray color values) but low verbal scores (represented by blue color values). Hawaii appears to be an outlier, since it has a strikingly different pattern for its math scores and its verbal scores.
- There is very little contrast between North Dakota's math and verbal scores. North Dakota's math and verbal scores are generally high (represented by red color values).

For a description of color themes, see "Right-Click Menu for Cell Plots".

Chapter 9

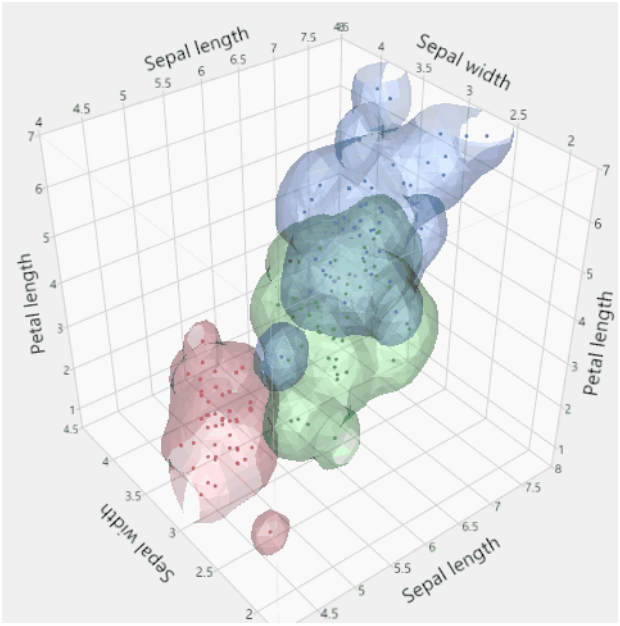
Scatterplot 3D

Create a Rotating Three-Dimensional View of Data

The Scatterplot 3D platform shows the values of numeric columns in the associated data table in a rotatable, three-dimensional view. Up to three columns that you select from the associated data table are displayed at one time.

To help visualize variation in higher dimensions, the 3D scatterplot can show a biplot representation of the points and variables when you request principal components. The most prominent directions of data are displayed on the 3D scatterplot report.

Figure 9.1 Example of a 3D Scatterplot



Contents

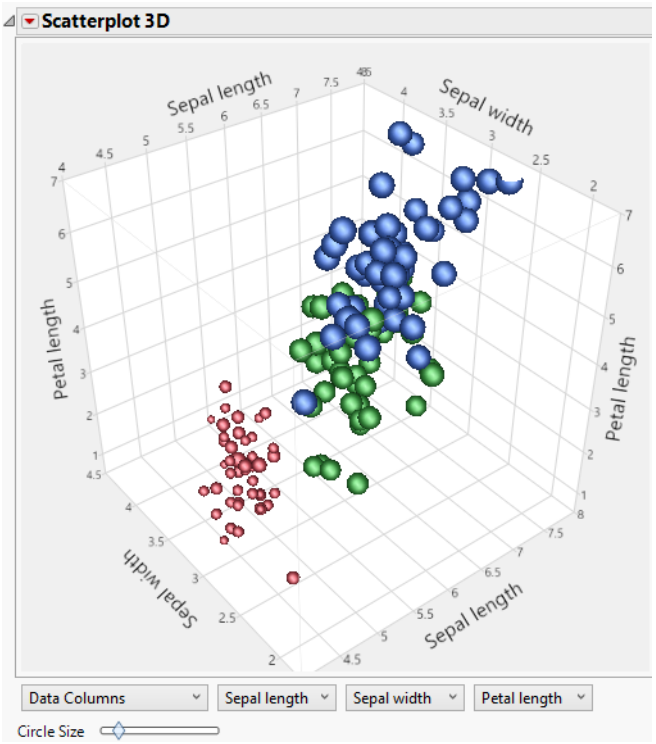
Example of a 3D Scatterplot 207
Launch the Scatterplot 3D Platform 208
The Scatterplot 3D Report 209
Spin the 3D Scatterplot 211
Change Variables on the Axes 212
Adjust the Axes 212
Assign Colors and Markers to Data Points 213
Assign Colors and Markers in the Data Table 213
Scatterplot 3D Platform Options 213
Normal Contour Ellipsoids 216
Nonparametric Density Contours 216
Pop-Up Menu 218
Additional Examples of the Scatterplot 3D Platform 221
Example of an Ungrouped Normal Contour Ellipsoid 221
Example of Grouped Normal Contour Ellipsoids 222
Example of a Grouped Nonparametric Density Contour 223

Example of a 3D Scatterplot

This example uses the Iris.jmp sample data table, which includes measurements of sepal length, sepal width, petal length, and petal width for three species of iris.

1. Open the Iris.jmp sample data table.
2. Select **Graph > Scatterplot 3D**.
3. Select Sepal length, Sepal width, and Petal length, and click **Y, Columns**.
4. Select Petal width and click **Weight**.
5. Click **OK**.

Figure 9.2 Example of an Initial 3D Scatterplot

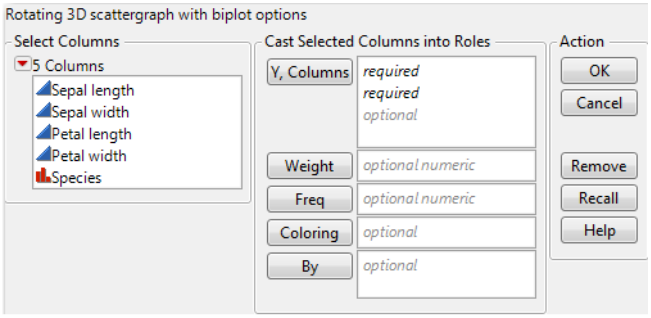


Now you can spin the 3D scatterplot to see the relationships between the variables. In this example, the data points are formatted in blue, red, and green. You might want to spin the scatterplot to see more clearly the relationships between the red and green points.

Launch the Scatterplot 3D Platform

Launch the Scatterplot 3D platform by selecting **Graph > Scatterplot 3D**.

Figure 9.3 The Scatterplot 3D Launch Window



Y, Columns Select the variables to plot on the 3D scatterplot. The order in which you select the variables determines where the data points appear on the axes:

- The first variable appears on the *x* axis.
- The second variable appears on the *y* axis.
- The third variable appears on the *z* axis.

You can assign the remaining variables interactively through the drop-down menus below the scatterplot.

Weight Use the **Weight** variable to:

- Assign a weight (importance or influence) to the data
- Visualize a fourth variable that sizes the points

Note: Red triangle options account for the **Weight** variable. If you do not want this variable accounted for in your analyses, remove it from the launch window.

When you specify a **Weight** variable, JMP draws the points as balls. The balls are scaled so that their volume represents the weight value. You click and drag the **Circle Size** slider below the scatterplot to resize the balls.

Freq Identifies the data table column whose variables assign a frequency to each row. This option is useful when a frequency is assigned to each row in summarized data.

Coloring Colors the points according to the selected variable. If the selected variable is categorical (nominal or ordinal), each category is colored distinctly. If the selected variable is continuous, a gradient of colors is used.

By Produces a separate 3D scatterplot for each By variable value. When two By variables are assigned, a separate graph is produced for each combination of both By variables.

After you click **OK**, the Scatterplot 3D report window appears.

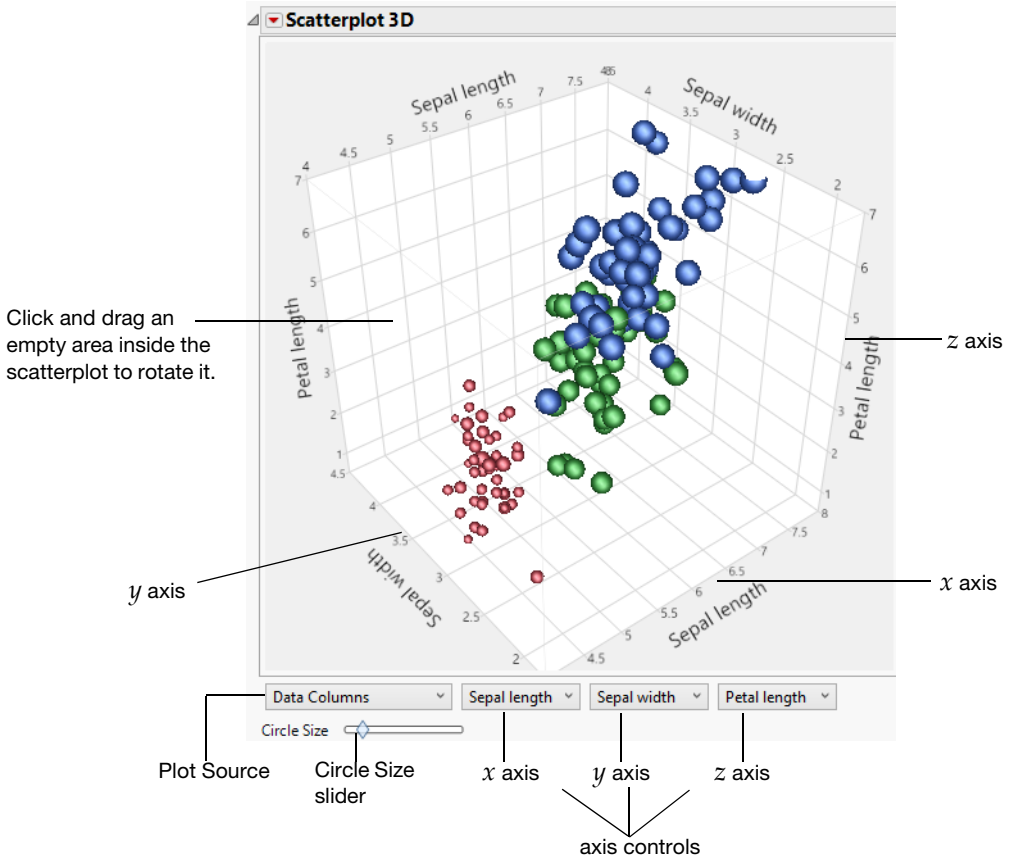
The Scatterplot 3D Report

To produce the 3D scatterplot shown in Figure 9.4, follow the instructions in "Example of a 3D Scatterplot".

The Scatterplot 3D report shows a three-dimensional spinnable view of your data. See Figure 9.4. In the launch window, you select the variables and then create the report. The variables are displayed on the 3D scatterplots' *x*, *y*, and *z* axes. Up to three variables can be displayed at a time.

Note: The Crosshairs tool is not supported in 3D scatterplots.

Figure 9.4 Example of Information Displayed on the Scatterplot 3D Report



Note: Any rows that are excluded in the data table are also hidden in the 3D scatterplot.

Plot source The plot source box indicates the source of the data in the plot.

Circle Size slider Click and drag the **Circle Size** slider to resize the balls while maintaining their relative sizes.

Note: The **Circle Size** slider appears only if you have specified a **Weight** variable.

Axis controls Select which variable appears on each axis. Choose the **Other** option to add a new variable.

Next Axis Set Cycles through the axis controls for any hidden variables. See "Change Variables on the Axes".

Note: The **Next Axis Set** button appears only if your analysis contains more than three variables.

After you create a 3D scatterplot, you can add features such as displaying ellipses around specific data points, showing separate principal components, rotating components, connecting points, and more. See "Scatterplot 3D Platform Options".

You can also assign colors and symbols (or markers) to data points either on the 3D scatterplot itself or in the associated data table. See "Assign Colors and Markers to Data Points" and "Assign Colors and Markers in the Data Table".

Spin the 3D Scatterplot

You spin the 3D scatterplot report in four ways:

- Click and drag an empty area on the 3D scatterplot. The 3D scatterplot spins in the direction you dragged the mouse.

  Note: Click and drag on an empty area on the 3D scatterplot, not on an axis or data point. Dragging the axis rescales the axis. Dragging a data point selects only the point.

- Slide the mouse wheel. The 3D scatterplot spins up and down only.
- Hold down an arrow key. (Before using an arrow on the number keypad, verify that Num Lock is turned off.)
- Hold down the Esc key. The 3D scatterplot spins left and right only.

In each case, the 3D scatterplot spins as long as you hold down the mouse button, arrow key, or Esc key. The spinning also continues as you slide the mouse wheel.

You can also spin the 3D scatterplot continuously in any of these ways:

- Click and drag: Hold down the Shift key, click and drag an empty area on the plot, and release Shift. The faster you drag the mouse, the faster the 3D scatterplot spins.
- Mouse wheel: Hold down the Shift key, slide the wheel, and release the wheel. The 3D scatterplot spins up and down only.
- Arrow keys: Hold down the Shift key and press the arrow key, and then release Shift.
- Esc key: Press Shift+Esc. The 3D scatterplot spins left and right only.

In addition to automatically spinning the plot, you can oscillate the plot. Press Shift+Ctrl and then click and drag the plot. The plot shakes up and down or left to right, depending on the direction in which you dragged the plot.

To stop the spinning or oscillating, click the plot or press the Esc key.

Change Variables on the Axes

The variables on each axis are determined by the order in which you select the variables in the launch window. For example, the first variable that you select is displayed on the *x* axis. The second variable is displayed on the *y* axis, and the third variable is displayed on the *z* axis.

After you create a 3D scatterplot, you can change the variable assigned to an axis, plot a different set of variables, or sequence through all combinations of the variables.

1. To change the variable on a specific axis, select the axis control drop-down menu and select a different variable.
2. To add a different variable, click an axis control drop-down menu, select **Other**, select the variable, and then click **OK**.
3. To sequence through combinations of all variables, click the **Next Axis Set** button until the variables that you want to plot are displayed.

Adjust the Axes

You can manually move or rescale the axis coordinates by clicking and dragging the axis. This option shows a different set of coordinates on the 3D scatterplot. It also lets you change the space displayed between the coordinates (or the coordinate scaling).

You can also specify axis properties by double-clicking the axis and modifying settings in the specifications window.

To Move the Coordinates on the Axis

1. Hover over the middle of the axis.
2. Click and drag the axis.

To Modify Coordinate Scaling

1. Hover over the end of the axis.
2. Click and drag the axis.

To Rescale an Axis Precisely

1. Hover over the middle of the axis (the axis, not the label).
2. Double-click the axis.
3. Change the minimum and maximum coordinate values.

For more information about updating axes, see *Using JMP*.

Assign Colors and Markers to Data Points

Each point in the 3D scatterplot corresponds to a row in the associated data table. To highlight points on the 3D scatterplot, you assign colors and markers to the points. The colors and markers are then displayed on the 3D scatterplot and in the data table.

When you click a point, the following items are selected:

- the point in the 3D scatterplot
- the corresponding row in the associated data table
- the point in any other opened graphs, if applicable

To select one point, click the point.

To select several points, double-click the 3D scatterplot and drag the cursor over the points. A box is displayed to indicate which points are selected.

To deselect points, double-click the 3D scatterplot.

Assign a color or marker to selected data points:

1. To assign a color to the selected point, select **Rows > Colors** and then select the color.
2. To assign a marker to the selected point, select **Rows > Markers** and then select the marker.

Assign Colors and Markers in the Data Table

You can assign colors and markers to rows in the data table. The colors and markers appear next to the row number in the data table and on the 3D scatterplot. This option distinguishes points for each variable, and you can save the settings in the data table. Assigning colors and markers to specific data points (as described in "Assign Colors and Markers to Data Points") only highlights them for the current open graphs.

See *Using JMP* for more information about assigning colors and markers in the data table. For more information about changing the size, quality, or transparency of markers, see "Scatterplot 3D Settings".

Scatterplot 3D Platform Options

The red triangle menu next to Scatterplot 3D contains options to customize the display and to compute, rotate, and save principal or rotated components.

Show Points Shows or hides the data points on the graph.

Show Controls Shows or hides the source and axis controls displayed beneath the 3D scatterplot (Figure 9.4).

Drop Lines Draws lines from each point to the plane created by the *x* and *z* variables that you selected on the launch window.

Connect Points Connects the points with a line. Points can be connected on the data as a whole or in groups. You can also group data by a specific variable.

Jitter Displays small spaces between the data points so that you can see each point more clearly.

Normal Contour Ellipsoids Draws one or more normal contour ellipsoids, that is, three-dimensional ellipses that encompass a specified portion of points. You specify whether you want an ellipsoid for all of the data or for each group. You can also control the size and transparency of the ellipsoids. See "Normal Contour Ellipsoids".

Ellipsoid Coverage Changes the size of normal contour ellipsoids. Type a value *between* 0 and 1, where the greater the value creates a bigger the ellipsoid. The actual values "0" and "1" produce no ellipsoid, so a warning appears if you try to use those values.

This option only appears after you add a normal contour ellipsoid to the 3D scatterplot.

Ellipsoid Transparency Changes the surface of normal contour ellipsoids. The greater the value, the more opaque the ellipsoid. This option only appears after you add a normal contour ellipsoid to the 3D scatterplot.

Nonpar Density Contour Draws nonparametric density contours, which approximately encompass a specified proportion of the points. You specify whether you want a density contour for all of the data or for each group. See "Nonparametric Density Contours".

Drop Line Thickness Changes the width of drop lines. This option only appears after you add drop lines to the 3D scatterplot.

Principal Components Calculates principal components on all variables. This changes the axes of the plot to have principal component scores.

Biplot rays are displayed by default. You can remove them by selecting **Biplot Rays** from the red triangle menu. For more information about principal components, see *Multivariate Methods*.

Std Prin Components Calculates principal components (as with the **Principal Components** option) but scales the principal component scores to have unit variance. If this option is not selected, the scores have variance equal to the corresponding eigenvalue.

With standardized principal components, the correlation between the variables and the principal component scores is equal to the values in the eigenvector. This helps you quickly assess the relative importance of the variables. See *Multivariate Methods*.

Select this option if you want GH' rather than JK' biplots. GH' biplots try to preserve relationships between variables; JK' biplots try to preserve relationships between

observations. The interpoint distance shown by GH' biplots is less meaningful, but the angles of the GH' biplot rays measure correlations better.

Rotated Components Specifies the number of factors that you want to rotate and the rotation method. You rotate components to better align the directions of the factors with the original variables so that the factors might be more interpretable. See *Multivariate Methods*.

Biplot Rays Shows or hides biplot rays that correspond to the principal components. You must have already selected Principal Components, Std Prin Components, or Rotated Components for this option to appear.

Show Ray Labels Shows or hides labels for the biplot rays. You must have already selected Biplot Rays for this option to appear.

Remove Prin Comp Removes principal components, standardized principal components, and rotated components from the scatterplot 3D report. The 3D scatterplot reverts to its original display before principal components were selected. This option, however, does not remove any saved principal components from the data table.

This option only appears after you add principal, standard, or rotated components to the 3D scatterplot.

Save Prin Components Saves the specified number of current principal component scores as new columns in the current data table. These columns also include the formulas used for the principal components. For *n* variables in the components list, *n* principal component columns are created and named Prin1, Prin2, ... Prinn.

This option only appears after you add principal, standard, or rotated components to the 3D scatterplot.

Save Rotated Components Saves all rotated component scores as columns in the current data table. These columns also include the formulas that were used. If you requested *n* rotated components, then *n* rotated component columns are created and named Rot1, Rot2, ... Rotn.

This option only appears after you add rotated components to the 3D scatterplot.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Normal Contour Ellipsoids

A normal contour ellipsoid is a 3-dimensional ellipse that encompasses a specified portion of points. The ellipsoid is computed from a contour of the multivariate normal distribution fit to the points. The ellipsoid is a function of the means, standard deviations, and correlations of variables on the plot. See *Multivariate Methods* for more information about multivariate normal distributions.

When you add an ellipsoid, two formatting options are available:

- *Coverage* changes the portion of data points covered by the ellipsoid. The larger the value, the bigger the ellipsoid.
- *Transparency* changes the surface of the ellipsoid from transparent to opaque. The larger the value, the more opaque the ellipsoid.

The coverage and transparency options also appear in the red triangle menu after you add the ellipsoid.

When you add normal contour ellipsoids to a 3D scatterplot, you specify whether you want an ellipsoid for all of the data or for a specific group of data. The ellipsoid for each set of grouped data is color-coded to differentiate one group from another.

You display and remove normal contour ellipsoids by selecting and deselecting **Normal Contour Ellipsoids** from the red triangle menu.

Related Information

- "Example of an Ungrouped Normal Contour Ellipsoid"
- "Example of Grouped Normal Contour Ellipsoids"

Nonparametric Density Contours

The nonparametric density contour shows contours that approximately encompass a specified proportion of the points. You add nonparametric density contours to see patterns in point density when the scatterplot is darkened by thousands of points.

This feature is particularly valuable when you have many points on a 3D scatterplot; the contours can be so dark that you cannot see the structure. In this situation, you remove the points so that only the contours are displayed. See "Optimizing a Dense Nonparametric Density Contour".

When you add nonparametric density contours to a 3D scatterplot, you specify whether you want a contour for all of the data or for a specific group of data. The contour for each set of grouped data is color-coded to differentiate one group from another.

You display and remove nonparametric density contours by selecting and deselecting **Nonpar Density Contours** from the red triangle menu.

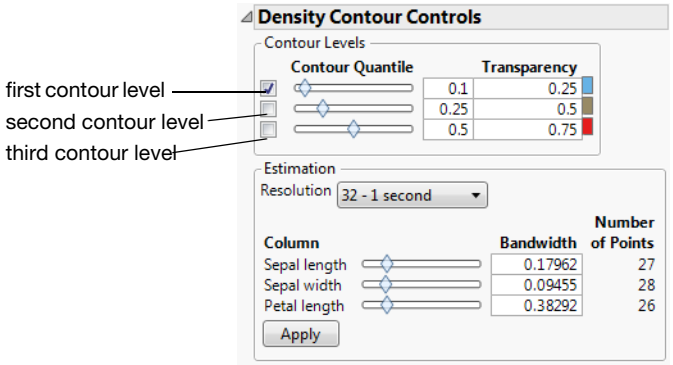
Related Information

- "Example of a Grouped Nonparametric Density Contour"

Density Contour Controls

The Density Contour Controls options are displayed below the 3D scatterplot. These options let you select additional contours and change each contour's formatting.

Figure 9.5 The Density Contour Controls Window



Contour Quantile Controls which contours are shown and lets you customize the contour formatting.

- *Density level* represents the volume and density of the points. As the contours go from smaller to larger values, the contours cover less volume but more dense areas. A 0.9 contour represents the 10% densest part of the total, where the points are closest together. Click and drag the slider below "Contour Quantile," or enter a value next to the slider.
- *Transparency* changes the surface of density contours. The greater the value, the more opaque the contour. Enter a value in the box.
- *Color* changes the color of the contour. Click the colored box and select a different color. (This option only appears for ungrouped density contours.)

Changes to these settings take effect immediately.

Resolution Changes the resolution of the contours. A higher resolution results in a less granular drawing of the contours but takes more time to display.

Column Bandwidth Changes the smoothness of the fitted density. A higher bandwidth results in a smoother fitted density.

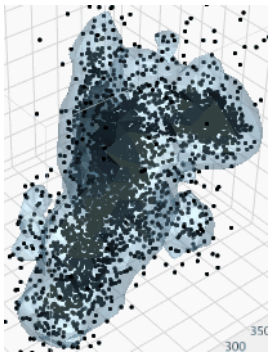
Type a new bandwidth for each variable, or click and drag the sliders. Click **Apply** to display your changes.

Optimizing a Dense Nonparametric Density Contour

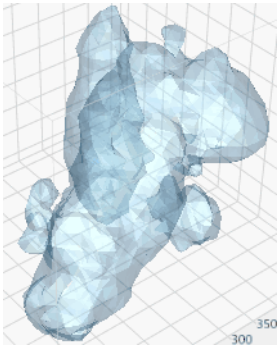
When you have many points on a 3D scatterplot, the contours can be so dark that you cannot see the structure. In this situation, you remove the points so that only the contours are displayed.

To remove points from a 3D scatterplot, select **Show Points** from the red triangle menu. You can further optimize the contours by changing their size, color, and transparency. See "Scatterplot 3D Platform Options".

Figure 9.6 Example of Optimizing a Dense Nonparametric Density Contour



3D scatterplot with density contour and points



3D scatterplot with density contour and no points

Pop-Up Menu

Right-click the 3D scatterplot to find the following options.

Show Legend Shows and hides the color legend for the 3D scatterplot.

Note: This option requires the **Coloring** role.

Reset Returns the orientation of the scatterplot to its original state.

Settings Provides options to change the appearance of the 3D scatterplot. See "Scatterplot 3D Settings".

Hide Lights Border Shows and hides a border that displays the lights. The lights highlight different portions of the 3D scatterplot.

Right-click a light to turn it on or off and to change the color.

Wall Color Changes the color of the 3D scatterplot.

Background Color Changes the color surrounding the 3D scatterplot.

Rows You can color, mark, exclude, hide, and label points that correspond to rows in the associated data table. You must select the points before selecting this option. See *Using JMP*.

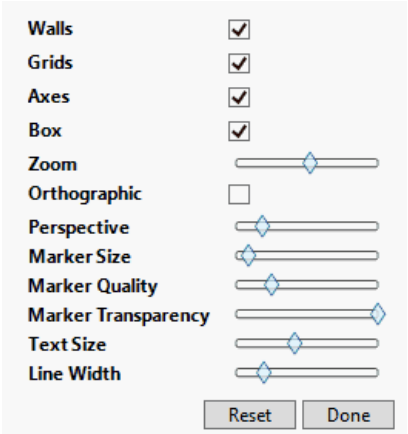
Use Hardware Acceleration Turns hardware acceleration on or off for machines that support acceleration. This option might display the scatterplot faster. If not, try updating your graphics drivers.

Show ArcBall Shows and hides a globe around the 3D scatterplot. This option helps you visualize the rotation of the scatterplot. Select whether you want the ArcBall to appear always, only when you drag the scatterplot, or never.

Scatterplot 3D Settings

To customize properties such as the marker size, text size, and grid lines, right-click the 3D scatterplot and select **Settings**. The Settings window appears. As you modify the settings, a preview appears on the 3D scatterplot.

Figure 9.7 The Scatterplot 3D Settings Window



Note the following:

- Move the sliders left to decrease the selected property or to the right to increase the selected property.
- To move the Settings window around the scatterplot, click and drag the top portion of the window.

The following options are available:

Reset Resets the default settings.

Done Closes the window.

Walls Adds or removes the 3D scatterplot walls. Without walls, the background color of the 3D scatterplot is displayed.

Grids Shows or hides the coordinate lines.

Axes Shows or hides the variable names that appear above each axis.

Box Shows or hides the box. Without the box, the 3D scatterplot is displayed as an open plot.

Zoom Enlarges or shrinks the 3D scatterplot.

Orthographic Changes the view of the scatterplot from 3-dimensional to an orthographic projection. In the orthographic view, the walls of the scatterplot do not converge to a vanishing point. This means that you can compare near and far distances and see the structure between data points.

Note: If you turn off orthographic view and completely decrease the perspective, the walls of the scatterplot do not converge. This is the same effect that you get when you turn on orthographic view.

Perspective Increases or decreases the perspective. Large values create a view that is unnaturally large and visually disorienting. In this case, you need to resize the scatterplot window to show the entire plot.

Marker Size Increases or decreases the size of the data point markers.

Marker Quality Increases and decreases the data marker quality. For example, when you increase the marker quality, some markers have an opaque center. Other symbol markers are formatted in bold. Increase the zoom to see these changes in quality.

Marker Transparency Increases or decreases the transparency of the data markers.

Text Size Increases or decreases the text size.

Line Width Changes the width of the coordinate and axes lines.

Additional Examples of the Scatterplot 3D Platform

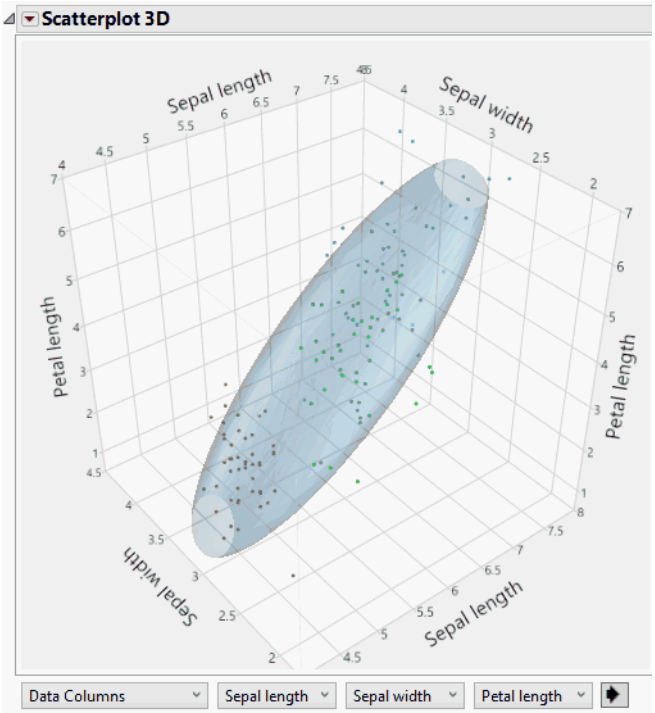
- "Example of an Ungrouped Normal Contour Ellipsoid"
- "Example of Grouped Normal Contour Ellipsoids"
- "Example of a Grouped Nonparametric Density Contour"

Example of an Ungrouped Normal Contour Ellipsoid

This example shows how to add a normal contour ellipsoid to more than 75% of the data points. The ellipsoid is 25% transparent.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot 3D**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Click **OK**.
5. Click the Scatterplot 3D red triangle and select **Normal Contour Ellipsoids**. Notice that **Ungrouped** is already selected.
6. Type 0.75 next to Coverage.
7. Type 0.25 next to Transparency.
8. Click **OK**.

Figure 9.8 Example of an Ungrouped Normal Contour Ellipsoid

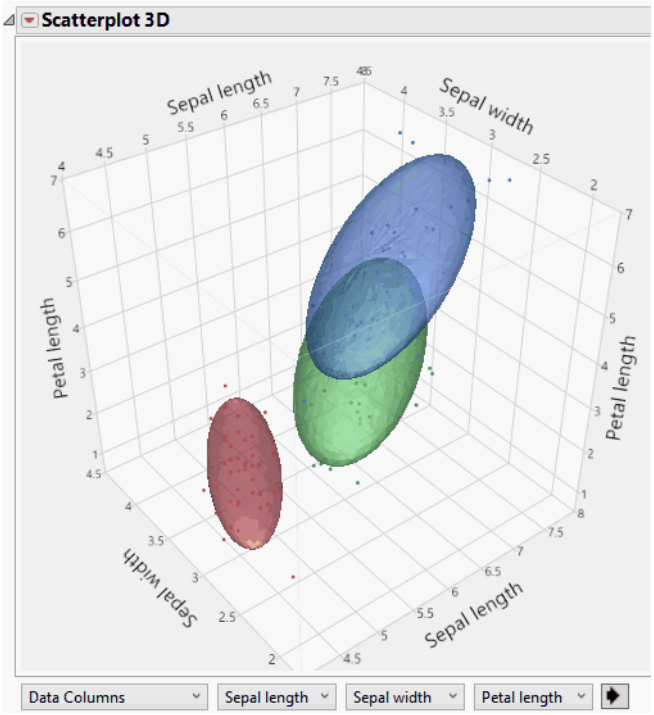


Example of Grouped Normal Contour Ellipsoids

This example shows how to group measurements by species and to format each group with a normal contour ellipsoid. The ellipsoids cover 75% of the data points and are 50% transparent. The contours are color-coded (using JMP default colors) based on species.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot 3D**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Click **OK**.
5. Click the Scatterplot 3D red triangle and select **Normal Contour Ellipsoids**.
6. Select **Grouped by Column**.
7. Select **Species**.
8. Type 0.75 next to Coverage.
9. Type 0.5 next to Transparency.
10. Click **OK**.

Figure 9.9 Example of Grouped Normal Contour Ellipsoids

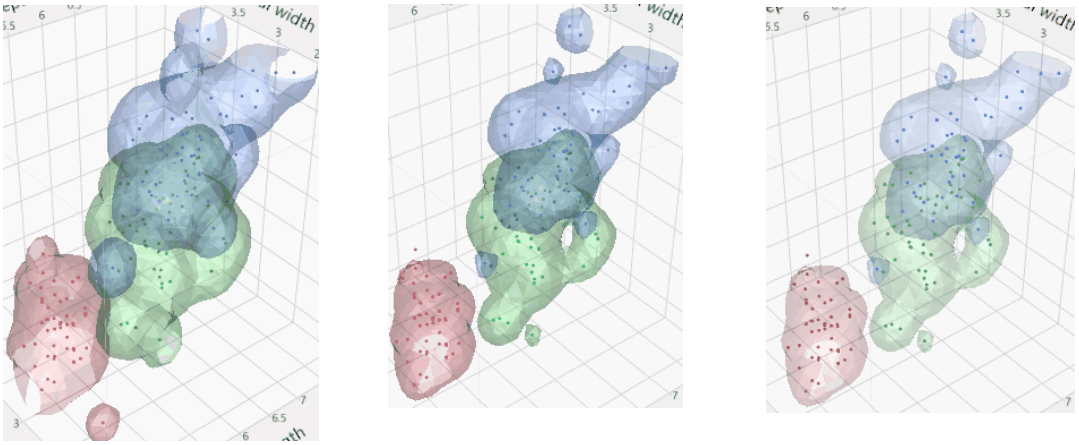


Example of a Grouped Nonparametric Density Contour

This example shows how to group data points and format each nonparametric density contour.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot 3D**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Click **OK**.
5. Click the Scatterplot 3D red triangle and select **Nonpar Density Contour**.
6. Select **Grouped by Column**.
7. Select **Species** and click **OK**. A different colored contour is displayed for each of the three species.
8. Type 0.25 in the first Contour Quantile box. 25% of the data points appear outside the contour surfaces, which results in smaller contours.
9. Type 0.15 in the first Transparency box. The contours are 15% opaque.

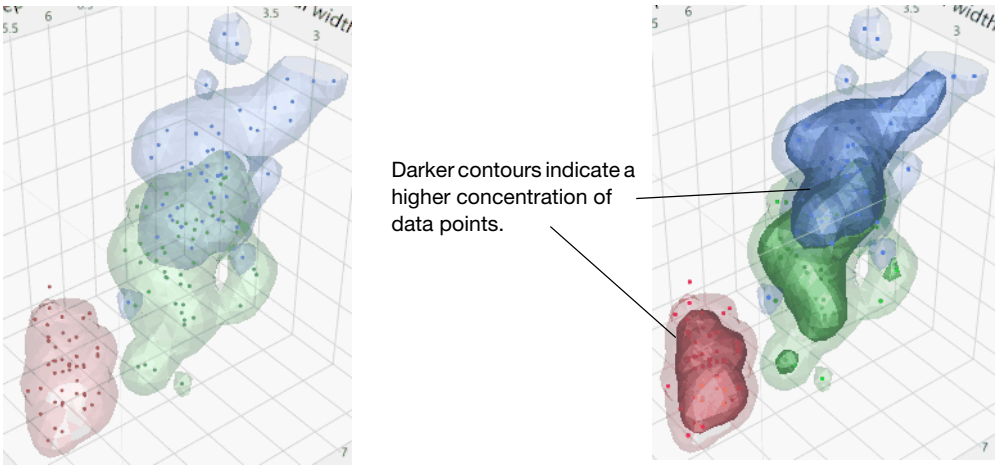
Figure 9.10 Changing the Nonparametric Density Contour Transparency and Density



original contours | decreased density | decreased transparency

10. Select the second check box. The contour quantiles are the same (.25), so the new contours overlap the first ones.
11. Type 0.5 in the second Contour Quantile box. 50% of the data points appear outside the contour surfaces. A second set of contours appears within the first, to further illustrate the density of the data points.

Figure 9.11 Adding a Second Nonparametric Density Contour



one level of contours | two levels of contours

You can now format the second levels of contours and turn on the third level of contours.

The options for formatting the grouped and ungrouped nonparametric density contours are similar. The only difference is that you cannot change the color of each *grouped* nonparametric density contour. See "Scatterplot 3D Platform Options" for options.

Chapter 10

Contour Plots

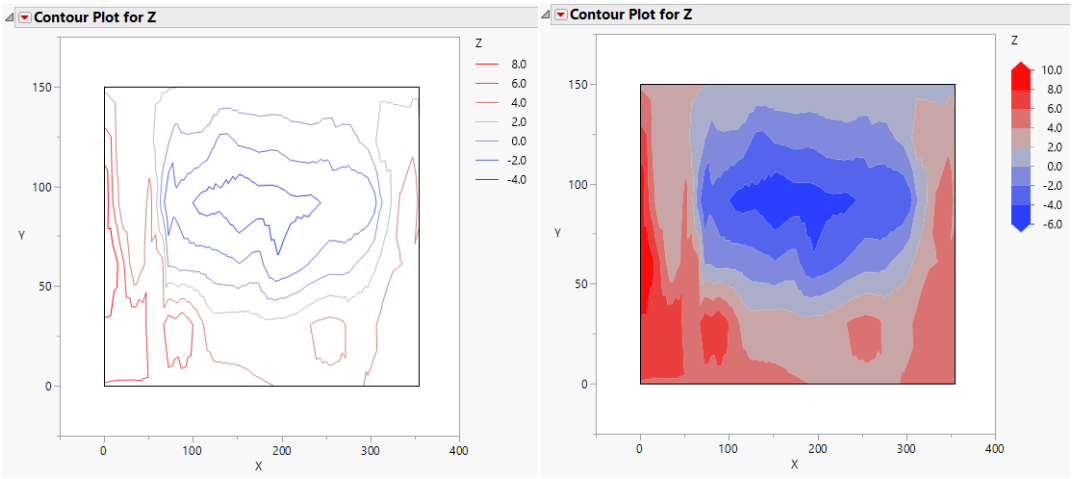
View Multidimensional Relationships in Two Dimensions

The **Contour Plot** command in the **Graph** menu constructs contours of a response in a rectangular coordinate system. A contour plot shows a three-dimensional surface in two dimensions. Contours delineate changes in the third dimension.

Here are some of the options available with the Contour platform:

- specify the number of contour levels
- choose to plot contour lines or filled contours
- show or hide data points
- label contours with response values
- define and use a custom coloring scheme

Figure 10.1 Examples of Contour Plots



Contents

Example of a Contour Plot 229
Example Using the Contour Platform 229
Example of a Contour Plot in Graph Builder 230
Launch the Contour Plot Platform 231
The Contour Plot 233
Contour Plot Platform Options 233
Fill Areas 235
Contour Specification 235
Contour Plot Save Options 239
Use Formulas for Specifying Contours 240
Additional Example of Contour Plots 241

Example of a Contour Plot

To create a contour plot, you can use either the Contour platform or Graph Builder. This section provides examples for both.

- To see the example using the Contour platform, see "Example Using the Contour Platform".
- To see the example using Graph Builder, see "Example of a Contour Plot in Graph Builder".

Example Using the Contour Platform

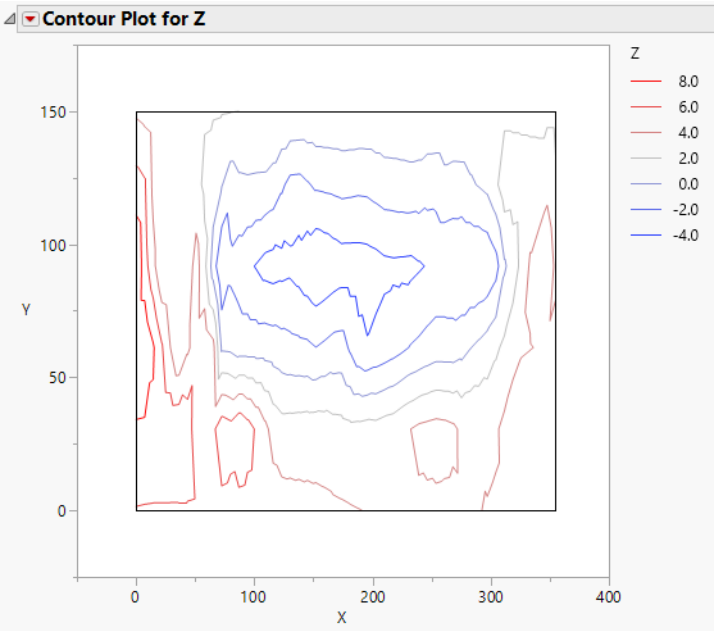
To create a contour plot, you need two variables for the *x*- and *y*-axes and at least one more variable for contours. You can also use several *y*-variables. This example uses the Little Pond.jmp sample data table. X and Y are coordinates of a pond. Z is the depth.

1. Select **Help > Sample Data Library** and open Little Pond.jmp.
2. Select **Graph > Contour Plot**.
3. Select the X and Y coordinates and click **X**.
4. Select the depth, Z, and click **Y**.

   Note: In a contour plot, the X1 and X2 roles are used for the X and Y axes.

5. Click **OK**.

Figure 10.2 Example of a Contour Plot with Legend



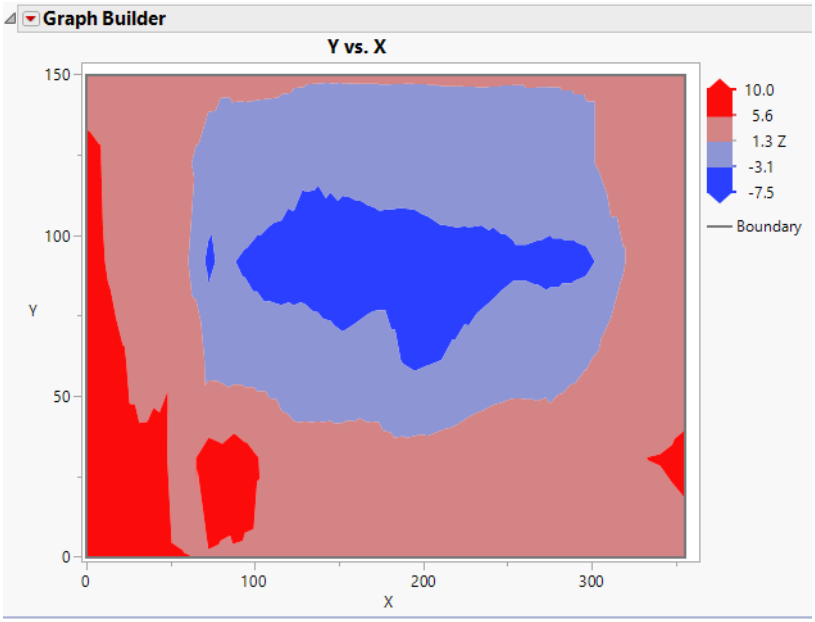
The *x*- and *y*-axes are coordinates and the contour lines are defined by the depth variable. This contour plot is essentially a map of a pond showing depth. To see the contours more clearly, click the red triangle next to Contour Plot for Z and select **Fill Areas**.

Example of a Contour Plot in Graph Builder

You can create the same contour plot shown in Figure 10.2 using Graph Builder.

1. Select **Help > Sample Data Library** and open Little Pond.jmp.
2. Select **Graph** > **Graph Builder**.
3. Click and drag the X coordinate to the **X** zone.
4. Click and drag the Y coordinate to the **Y** zone.
5. Click and drag depth, Z, to the **Color** zone.
6. Click the Contour icon.

Figure 10.3 Contour Plot in Graph Builder



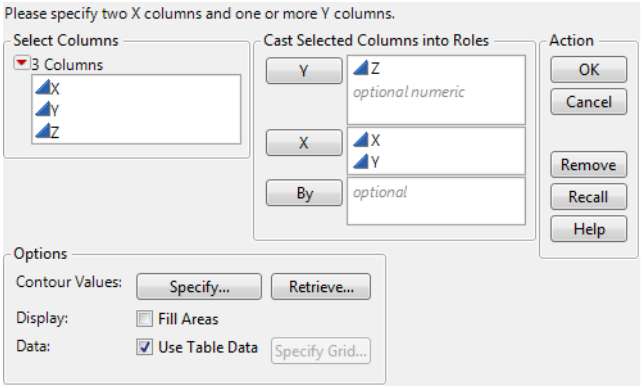
Notice the plot produced from the Contour platform looks slightly different than the plot produced using Graph Builder.

Launch the Contour Plot Platform

Launch the Contour Plot platform by selecting **Graph > Contour Plot**.

By default, the contour levels used in the plot are values computed from the data. You can specify your own number of levels and level increments in the Launch window before you create the plot. You can also do so in the red triangle menu for Contour Plot after you create the plot. You can use a column formula to compute the contour variable values.

Figure 10.4 The Contour Plot Launch Window



Y Columns assigned to the **Y** role are used as variables to determine the contours of the plot. You must specify at least one, and you can specify more than one.

You can also assign a column with a formula to this role. If you do so, the formula should be a function of exactly two variables. Those variables should be the *x* variables entered in the Launch window.

X Columns assigned to the **X** role are used as the variables for the *x*- and *y*-axes. You must specify exactly two columns for **X**.

By This option produces a separate graph for each level of the By variable. If two By variables are assigned, a separate graph for each possible combination of the levels of both By variables is produced.

Options

Contour Values Specify your own number of levels and level increments. See "Contour Specification".

Fill Areas Fill the areas between contour lines using the contour line colors.

Use Table Data and Specify Grid Most often, you construct a contour plot for a table of recorded response values. In that case, **Use Table Data** is selected and the **Specify Grid** button is unavailable.

However, if a column has a formula and you specify that column as the response (**Y**), the **Specify Grid** button becomes available. When you click **Specify Grid**, you can define the contour grid in any way, regardless of the rows in the existing data table. This feature is also available with table templates that have one or more columns defined by formulas but no rows. See "Use Formulas for Specifying Contours".

After you click **OK**, the Contour plot appears.

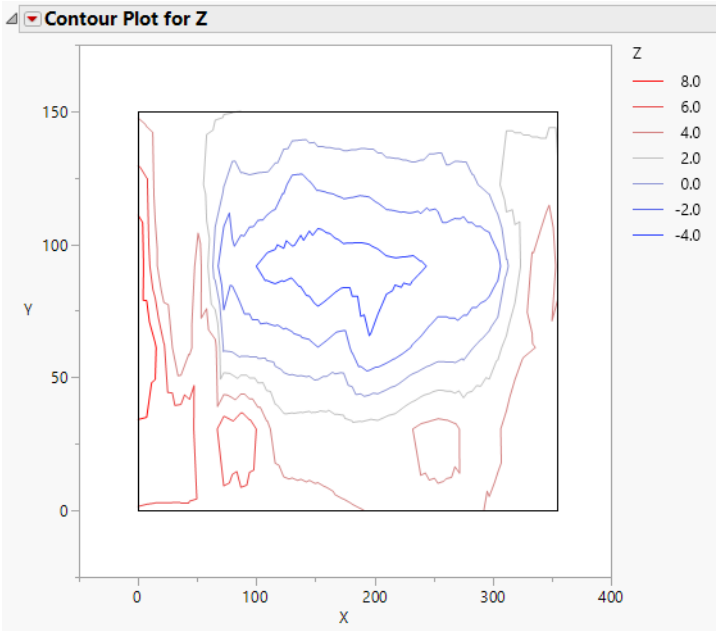
The Contour Plot

Follow the instructions in "Example of a Contour Plot" to produce the plot shown in Figure 10.5.

The legend for the plot shows individual markers and colors for the *Y* variable. Replace variables in the plot by dragging and dropping a variable, in one of two ways: swap existing variables by dragging and dropping a variable from one axis to the other axis; or, click a variable in the Columns panel of the associated data table and drag it onto an axis.

For information about additional options for the report, see "Contour Plot Platform Options".

Figure 10.5 The Contour Plot Report



Contour Plot Platform Options

Using the options in the red triangle menu next to Contour Plot, you can tailor the appearance of your contour plot and save information about its construction.

Show Data Points Shows or hides (*x*, *y*) points. The points are hidden by default.

Show Missing Data Points Shows or hides points with missing *y* values. Available only if **Show Data Points** is selected.

Show Contours Shows or hides the contour lines or fills. The contour lines are shown by default.

Show Boundary Shows or hides the boundary of the total contour area. The boundary is shown by default.

Show Control Panel Shows or hides the **Alpha** slider that allows you to change the **Alpha** shapes filter.

Transform If the contour plot includes a **Color** role, the **Transform** option is enabled. See "Additional Example of Contour Plots".

None The triangulation is computed without any scaling to coordinates using Delaunay triangulation. Delaunay triangles are computed to maximize the minimum angle of the triangles in the triangulation. This value is selected by default.

Range Normalized The X1/X2 values are both scaled to [0,1] prior to computing the triangulation. If the X1/X2 limits are different, then this is a non-uniform scale. This option may be more desirable in cases where the X1/X2 units are very different.

Fill Areas Fills the areas between the contours with a solid color. It is the same option that is available in the Launch window. If you leave it deselected in the Launch window, you can see the line contours before filling the areas. See "Fill Areas".

Label Contours Shows or hides the label (z-value) of the contour lines.

Fit to Window Specifies the stretching behavior of the report.

Change Contours Set your own number of levels and level increments. See "Contour Specification".

Save This menu has options to save information about contours, triangulation, and grid coordinates to data tables. See "Contour Plot Save Options".

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

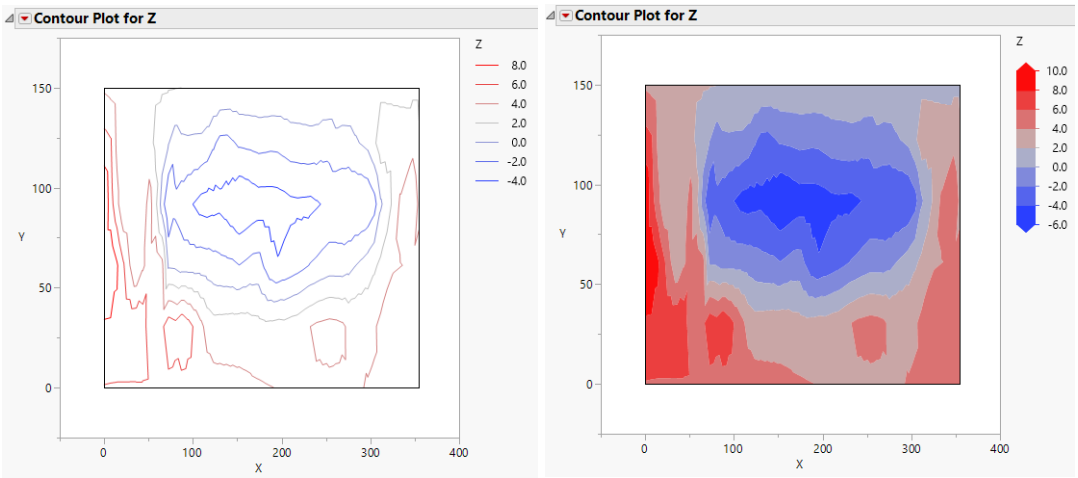
Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Fill Areas

If you select **Fill Areas**, the areas between contour lines are filled with the contour line colors. This option is available in the Launch window and in the red triangle menu for Contour Plot. Figure 10.6 shows a plot with contour lines on the left and a plot with the contour areas filled on the right.

Figure 10.6 Comparison of Contour Lines and Area Fills

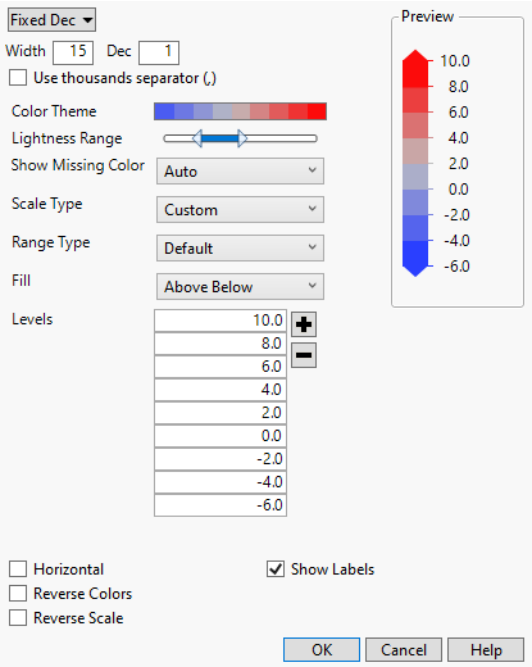


Areas are filled from low to high values on a color gradient. An additional color is added in the filled contour plot for the level above the last, and highest, contour line.

Contour Specification

If you do not select options in the Launch window, the default plot spaces the contour levels equally within the range of the *Y* variable. You can specify a color theme by selecting a Continuous Color Theme in **File > Preferences > Graphs**. You can customize colors for individual contours by right-clicking on the contour color in the Contour Plot legend. You can customize the entire contour color gradient by right-clicking on the Contour Plot legend and selecting Gradient to open the Gradient Settings.

Figure 10.7 Gradient Settings Window



To format the labels in the legend, use the format menu, Width and Dec boxes, and the Use thousands separator (,) check box in the top left of the window.

Color Theme Changes the color theme or defines a custom color theme. For more information about color options, see *Using JMP*.

Lightness Range Sets the range of intensities for the gradient coloring.

Number of Labels (Not available if the Scale Type is Custom.) Specifies the number of labels for your legend. The value of zero provides the default number of labels.

Show Missing Color Specifies whether to show missing color in the legend. Auto shows the color only when there are missing values, Off never shows the color, On always shows the color.

Scale Type Sets the scale for the gradient coloring. A description of the scale type is found below the Maximum value setting.

Linear The scale is piecewise linear between the Minimum and Center values and between the Center and Maximum values. This is the default scaling.

Quantile The scale is piecewise linear for quantiles of the gradient variable.

Standard Deviation The gradient variable range is divided into standard deviation offsets from the mean. The scale is piecewise linear between these offset settings. The number of offsets is determined by the Number of Labels.

Log The scale is linear for the logarithm of the gradient variable.

Log Offset The scale is linear for the offset of the base 10 logarithm of the gradient variable.

Custom The scale labels are specified by the user in the Levels setting.

Range Type Sets the range of the legend values.

Default Unless you specify values for the Minimum and Maximum, the minimum and maximum values include the range of the data and are chosen to be appropriately rounded values.

Exact Data Range The minimum and maximum values are the exact minimum and maximum values for the data, or they are the values that you specify as Minimum and Maximum.

Middle 90% The minimum and maximum values are the 5^{th} and 95^{th} quantiles, or they are the values that you specify as Minimum and Maximum. This option is resistant to outliers.

Fill Determines how the endpoints of the range are specified in the contours.

Between The contours stop at the endpoint and any data beyond the endpoint is not shown.

Above Values above the range are included in the top contour level.

Below Values below the range are included in the bottom contour level.

Above Below Values at either end are included in their respective contour levels.

Levels (Available only when the Custom Scale Type is specified.) Enables you to specify the number of labels and the label values for the legend.

Horizontal Sets the orientation of the legend to horizontal.

Reverse Colors Reverses the colors in the color scheme.

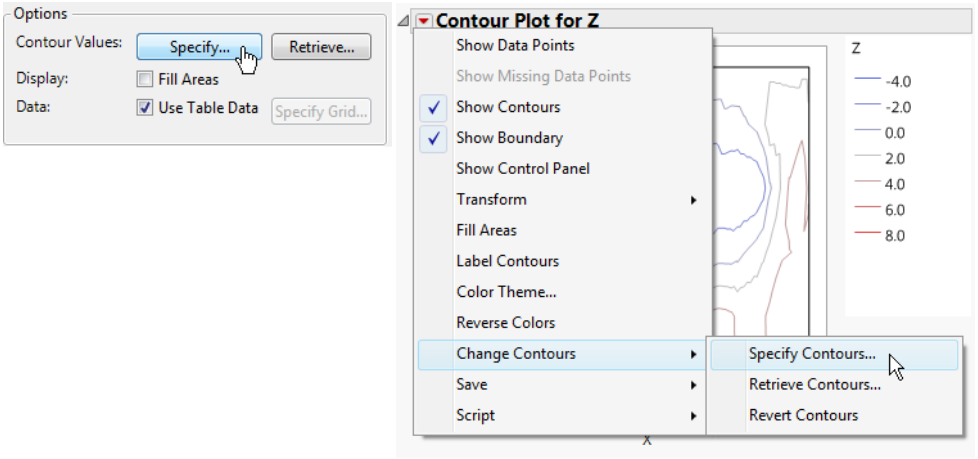
Reverse Scale Reverses the scale for the color theme.

Show Labels Shows or hides labels for the legend.

Specify

You can specify contour levels either in the Launch window (the **Specify** button) or in the report window from the red triangle menu for Contour Plot (the **Specify Contours** option). Selecting this option displays the Contour Specification window.

Figure 10.8 Example of Contour Specification: Launch Window (on the left) and Menu (on the right)

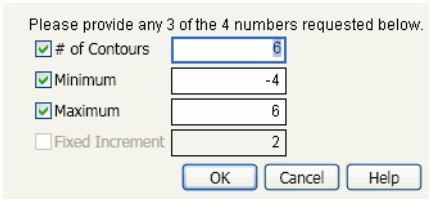


You can use the Contour Specification window to do the following:

- change the number of contours
- specify minimum and maximum values to define the range of the response to be used in the plot
- change the increment between contour values

You supply any three of the four values, and the remaining value is computed for you. Click the check box to deselect one of the numbers and automatically select the remaining check box.

Figure 10.9 The Contour Specification Window



Colors are automatically assigned and are determined by the number of levels in the plot. After the plot appears, you can right-click (press Control and click on macOS) on any contour in the plot legend and choose from the JMP color palette to change that contour color.

Retrieve

This option is both in the Launch window (the **Retrieve** button) and on the red triangle menu for Contour Plot (the **Retrieve Contours** option).

Note: Neither the button nor the menu option are active unless there is an open data table *in addition to* the table that has the contour plotting values. When you click **Retrieve** or select **Retrieve Contours**, a window with a list of open data tables appears.

Using this option, you can retrieve the following from an open JMP data table:

- the number of contours
- an exact value for each level
- a color for each level

From the list of open data tables, select the data table that contains the contour levels.

For level value specification, the Contour Plot platform looks for a numeric column with the same name as the response column that you specified in the Launch window. The number of rows in the data table defines the number of levels.

If there is a row state column with color information, those colors are used for the contour levels. Otherwise, the default platform colors are used.

Revert Contours

This option appears only on the red triangle menu for Contour Plot.

If you have specified your own contours, selecting this option reverts your Contour Plot back to the default contours.

Contour Plot Save Options

This menu has options to save information about contours, triangulation, and grid coordinates to data tables.

Save Contours Creates a new JMP data table with columns for the following:

- the x- and y-coordinate values generated by the Contour platform for each contour
- the response computed for each coordinate set
- the curve number for each coordinate set

The number of observations in this table depends on the number of contours you specified. You can use the coordinates and response values to look at the data with other JMP platforms. For example, you can use the Scatterplot 3D platform to get a three-dimensional view of the pond.

Generate Grid Displays a window that prompts you for the grid size that you want. When you click **OK**, the Contour platform creates a new JMP data table with the following:

- the number of grid coordinates you requested
- the contour values for the grid points computed from a linear interpolation

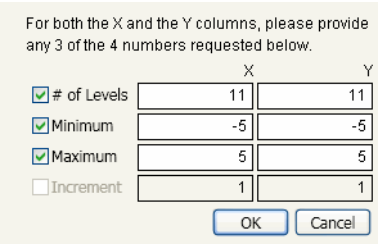
Save Triangulation Creates a new JMP data table that lists coordinates of each triangle used to construct the contours. By default, JMP uses Delaunay triangulation to connect the nearest data points to form triangles. The resulting set of triangles are calculated so that no other data points are inside a triangle's circumscribed circle, that is, the circle that passes through the three vertices of the triangle. To change the triangulation to a normalized scale, select **Transform > Range Normalized**.

Use Formulas for Specifying Contours

Most often you construct a contour plot for a table of recorded response values such as the Little Pond data table. In that case, in the launch window, **Use Table Data** is checked and the **Specify Grid** button is unavailable. However, if a column has a formula and you specify that column as the response (**Y**), the **Specify Grid** button becomes active.

When you click **Specify Grid**, the window shown in Figure 10.10 appears.

Figure 10.10 Example of the Contour Specification for Formula Column



You can complete the Specify Grid window and define the contour grid in any way, regardless of the rows in the existing data table. This feature is also available with table templates that have one or more columns defined by formulas but no rows.

Additional Example of Contour Plots

This example illustrates how to create a triangulation data table, to transform the triangulation to use Delaunay triangles, and to filter Alpha shapes of the triangles.

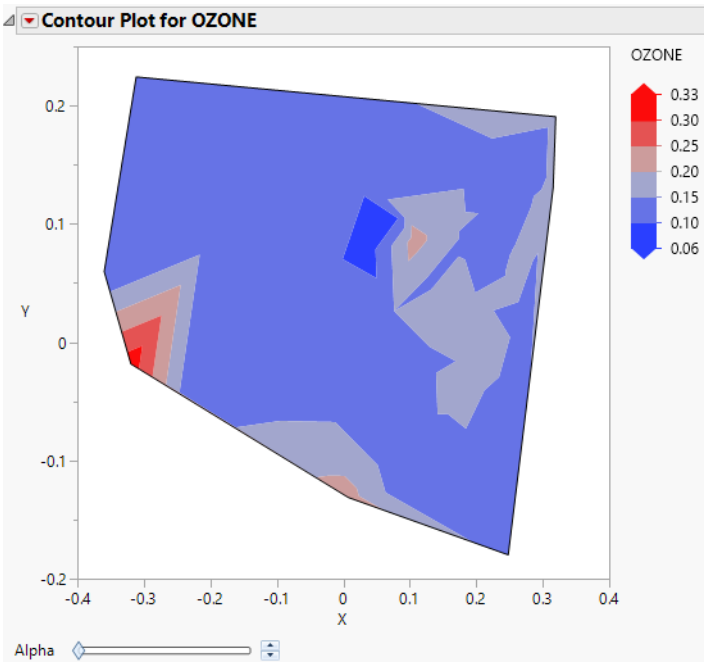
1. Select **Help > Sample Data Library** and open Cities.jmp.
2. Select **Graph > Contour Plot**.
3. Select OZONE and click **Y**.
4. Select X and Y and click **X**.
5. Select to **Fill Areas**.
6. Click **OK**.
7. Click the red triangle next to Contour Plot for OZONE and select the following options:
   a. **Transform** > **Range Normalized**

      Instead of Delaunay triangulation, this changes the method for calculating the triangulations to a normalized scale ([0,1]) in both X and Y.

   b. Show Control Panel

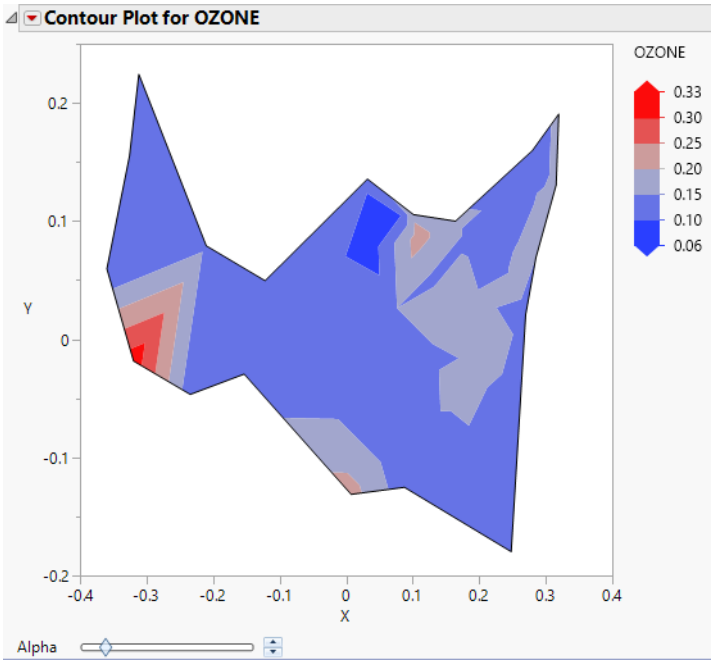
      The **Alpha** slider appears.

Figure 10.11 Contour Plot for OZONE



8. Click and move the **Alpha** slider to the right.

Figure 10.12 Alpha Shapes Filter



Using the Alpha slider filters out the larger Delaunay triangulation areas.

Chapter 11

Ternary Plots

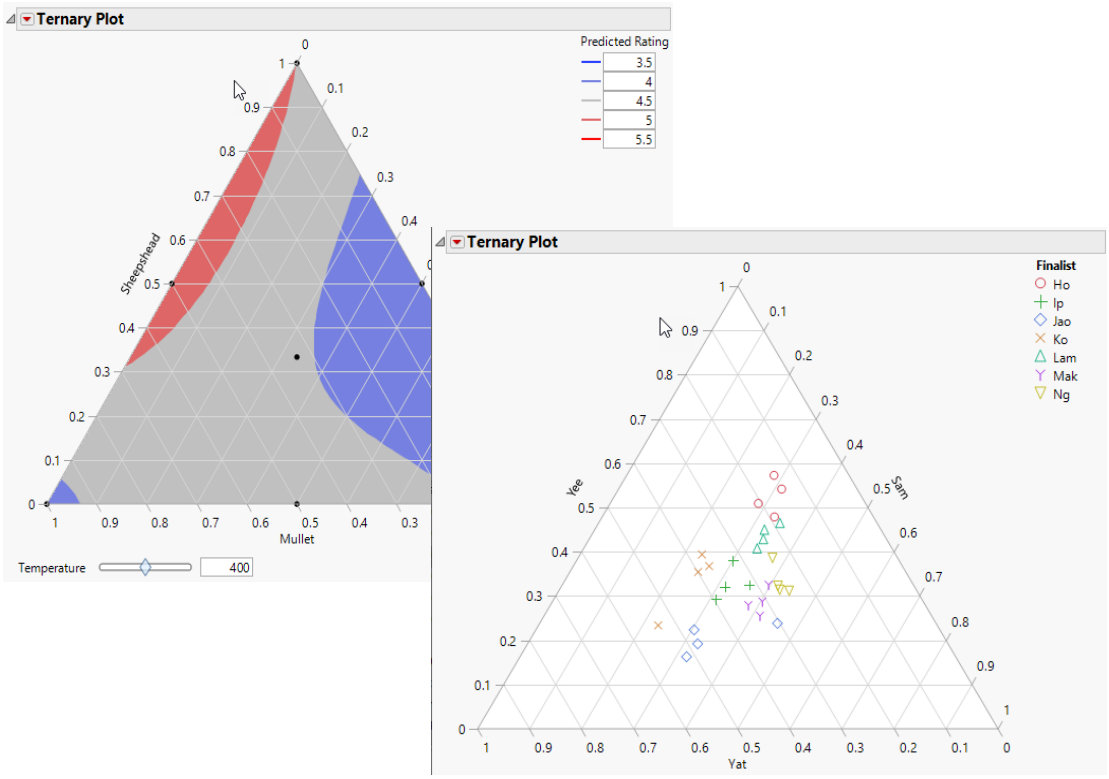
View Plots for Compositional or Mixture Data

The Ternary Plot command in the Graph menu produces a three-axis plot.

Ternary plots are a way of displaying the distribution and variability of three-part compositional data. (For example, the proportion of sand, silt, and clay in soil or the proportion of three chemical agents in a trial drug.) You can use data expressed in proportions or use absolute measures.

The ternary display is a triangle with sides scaled from 0 to 1. Each side represents one of the three components. A point is plotted so that a line drawn perpendicular from the point to each leg of the triangle intersect at the component values of the point.

Figure 11.1 Examples of Ternary Plots



Contents

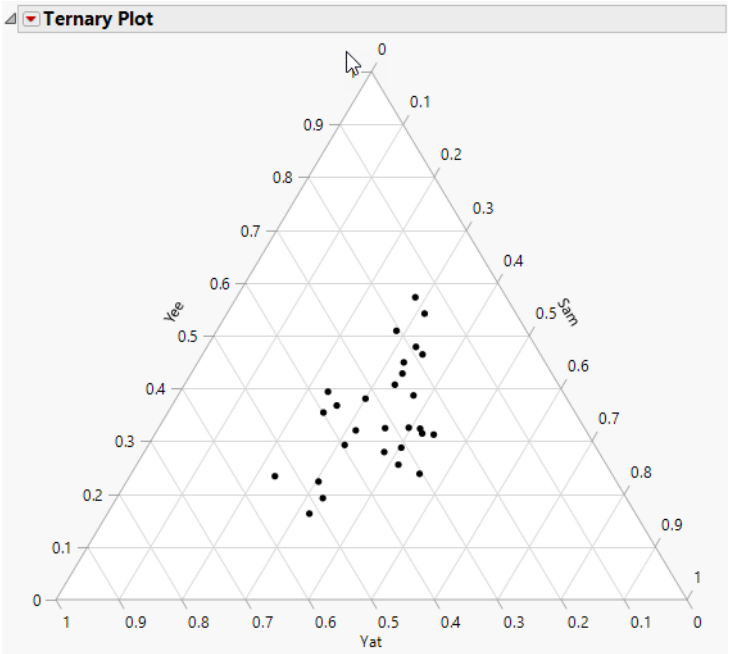
Example of a Ternary Plot 245
Launch the Ternary Plot Platform 247
The Ternary Plot 248
Mixtures and Constraints 249
Ternary Plot Platform Options 249
Additional Examples of the Ternary Plot Platform 250
Example Using Mixture Constraints 250
Example Using a Contour Function 251

Example of a Ternary Plot

This example uses the Pogo Jumps.jmp sample data table. The data, adapted from Aitchison (1986), show measurements for pogo jumps of seven finalists in the 1985 Hong Kong Pogo-Jump Championship. A single pogo jump is the total jump distance in three consecutive bounces, referred to as *yat*, *yee*, and *sam*.

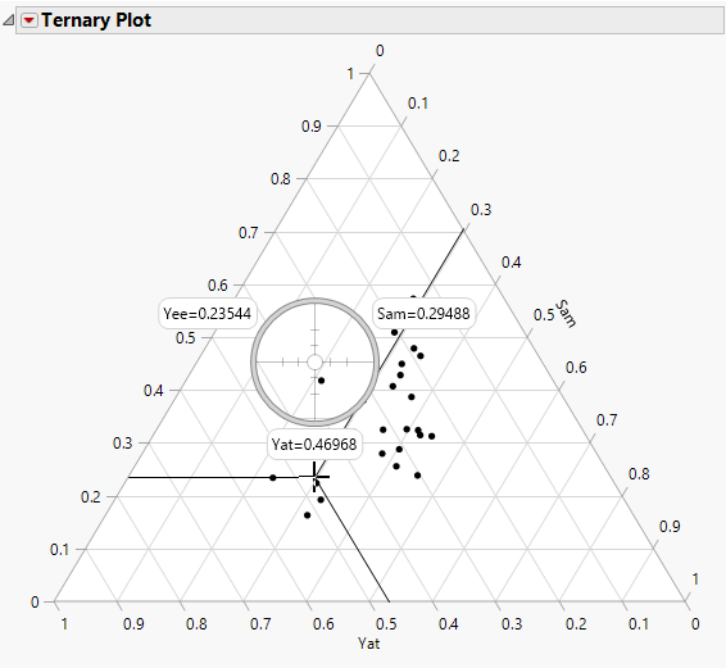
1. Select **Help > Sample Data Library** and open Pogo Jumps.jmp.
2. From the **Graph** menu, select **Ternary Plot**.
3. Select Yat, Yee, and Sam and click **X, Plotting**.
4. Click **OK**.

Figure 11.2 Example of a Ternary Plot



Use the crosshairs tool to determine exact coordinates of points within the plot.

Figure 11.3 Using the Crosshairs Tool

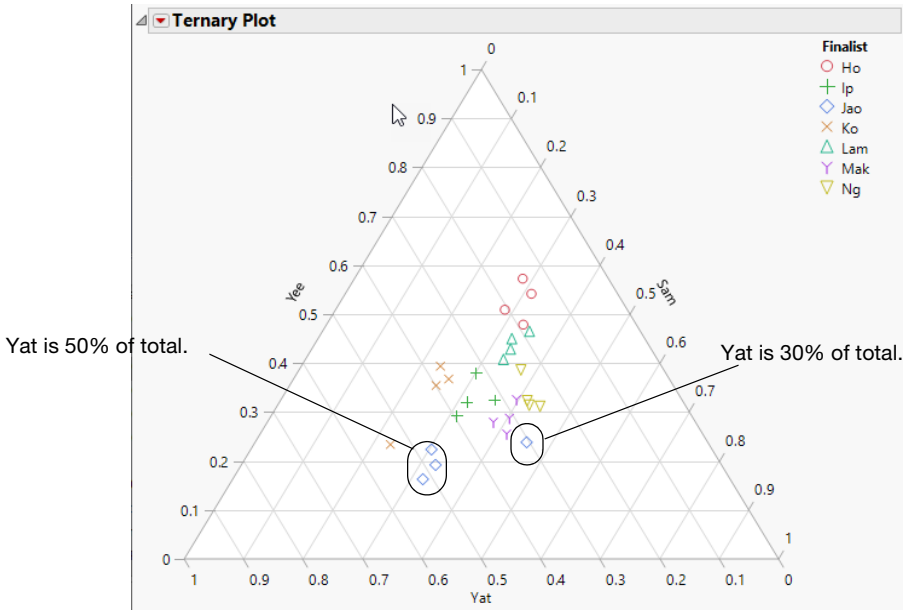


To get a better idea of how the three bounces contribute to total distance, assign each contestant's points a different color and marker.

1. Right-click the plot and select **Row Legend**.
2. Select Finalist in the column list box.

   Colors should be automatically set to **JMP Default**.
3. Select **Standard** from the **Markers** menu.
4. Click **OK**.

Figure 11.4 Pogo Data Colored by Finalist

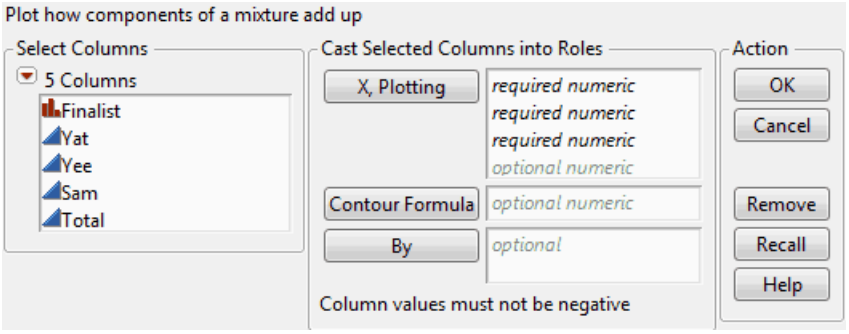


Note that most of the finalists are consistent in the composition of total distance. However, two finalists, Jao and Ko, both have one jump that is not consistent with their other jumps. For example, for three of Jao's jumps, the Yat composed about 50% of the total distance, but for the other jump, the Yat composed only 30% of the total distance. That jump is not consistent with the others. A similar observation can be made about Ko's jumps.

Launch the Ternary Plot Platform

Launch Ternary Plot by selecting **Graph > Ternary Plot**.

Figure 11.5 The Ternary Plot Launch Window



X, Plotting Assign three columns to generate a ternary plot.

If you assign more than 3 variables to the **X, Plotting** role, a matrix of ternary plots is created. A separate variable is assigned to the first two axes of a plot, with the third axis being the sum of the other variables. If necessary, the variables are scaled so they sum to 1.

Contour Formula To plot contours of a response surface, assign a column containing a formula to the **Contour Formula** role. If you have variables in a Contour formula that are not listed as **X, Plotting** variables, JMP appends sliders below the plot so that the values can be interactively adjusted. See "Example Using a Contour Function".

By This option produces a separate graph for each level of the By variable.

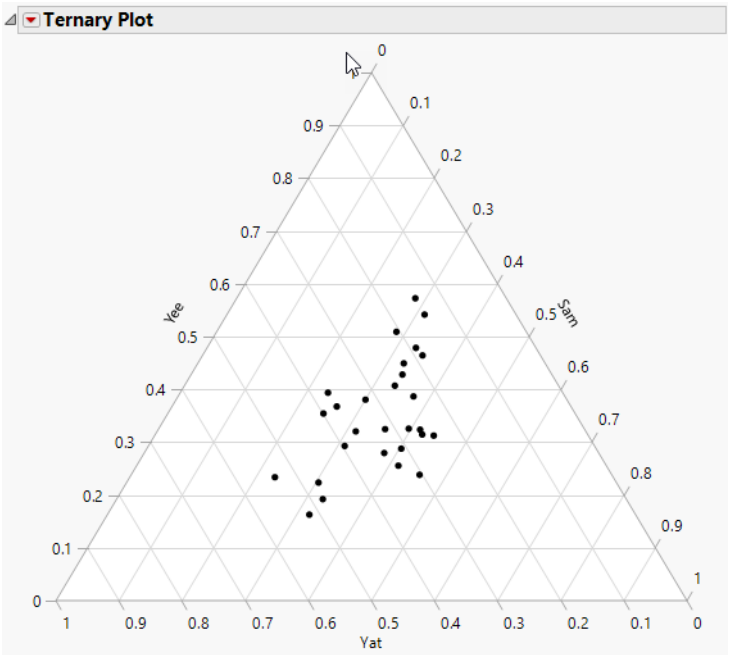
After you click **OK**, the Ternary Plot window appears.

The Ternary Plot

Follow the instructions in "Example of a Ternary Plot" to produce the plot shown in Figure 11.6.

Each of the three sides of a ternary plot represents a proportion of 0%, with the point of the triangle opposite that base representing a proportion of 100%. As a proportion increases in any one sample, the point representing that sample moves from the base to the opposite point of the triangle.

Figure 11.6 The Ternary Plot



Mixtures and Constraints

Ternary Plot uses the Mixture column property to shade the portion of the graph that is out of bounds. The Ternary plot recognizes the mixture sum, mixture bounds, and general linear constraints. For information about setting the Mixture column property in the Column Info window, see *Using JMP*.

Related Information

- "Example Using Mixture Constraints"

Ternary Plot Platform Options

The Ternary Plot red triangle menu contains options to modify the plot.

Note: To view more detailed options, right-click the plot.

Show Constraints Shows or hides the constraints on the plot. The default plot shows the constraints.

Contour Fill Allows filling of contours if a contour formula is specified in the plot. You can select Lines Only, Fill Above, or Fill Below. The default platform shows lines only.

Color Theme Allows you to select a color theme. The default plot shows the Blue to Gray to Red color theme.

Show Points Shows or hides the plotted points. The default plot shows the points.

Fit to Window Specifies whether the graph is resized as you resize the JMP window. The default setting is Auto, which is usually set to On except when the window is shared with other content or if there is a Page variable. To always fit the graph inside the window, keep the setting at On. You can also maintain the aspect ratio. To prevent the graph from resizing, change the setting to Off.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Additional Examples of the Ternary Plot Platform

- "Example Using Mixture Constraints"
- "Example Using a Contour Function"

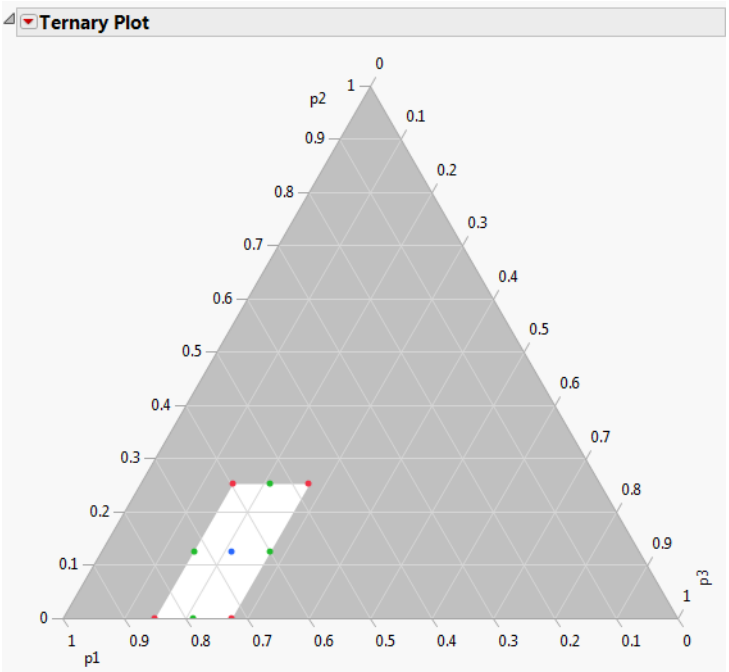
Example Using Mixture Constraints

1. Select **Help > Sample Data Library** and open Plasticizer.jmp.

   The p1, p2, and p3 columns all have Mixture Column Properties defined.
2. From the **Graph** menu, select **Ternary Plot**.

   Notice that p1, p2, and p3 appear next to **X, Plotting**.
3. Click **OK**.

Figure 11.7 Mixture Constraints in a Ternary Plot



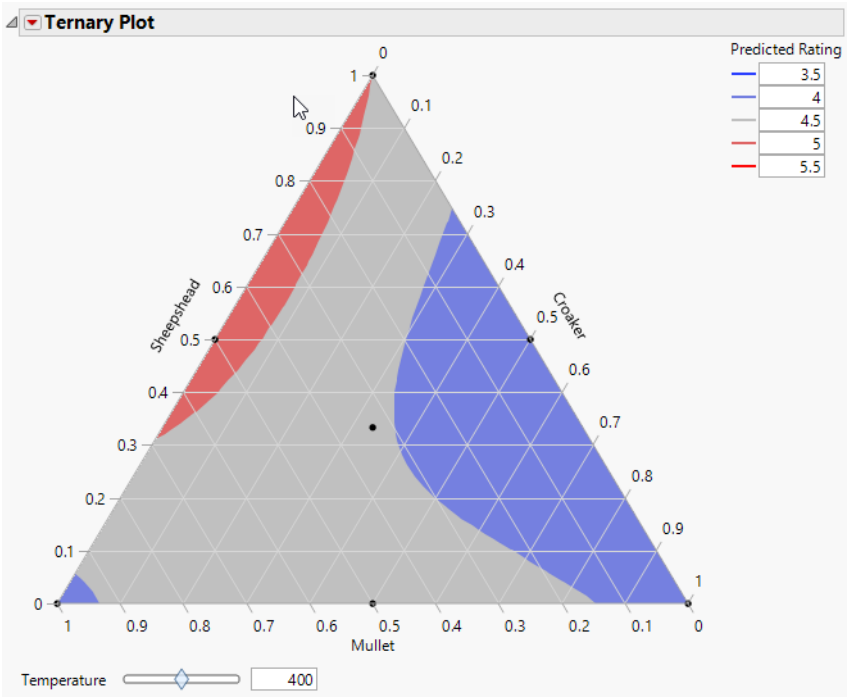
For more information about mixtures, see *Profilers*.

Example Using a Contour Function

The data in Fish Patty.jmp is adapted from Cornell (1990) and comes from an experiment to optimize the texture of fish patties. The columns Mullet, Sheepshead, and Croaker represent what proportion of the patty came from those fish types. The column Temperature represents the oven temperature used to bake the patties. The column Rating is the response and is a measure of texture acceptability, where higher is better. A response surface model was fit to the data and the prediction formula was stored in the column Predicted Rating. See *Profilers*.

1. Select **Help > Sample Data Library** and open Fish Patty.jmp.
2. From the **Graph** menu, select **Ternary Plot**.
3. Select Mullet, Sheepshead, and Croaker and click **X, Plotting**.
4. Select Predicted Rating and click **Contour Formula**.
5. Click **OK**.
6. Click the Ternary Plot red triangle and select **Contour Fill > Fill Above**.

Figure 11.8 Ternary Plot with a Contour Formula



The manufacturer wants the rating to be at least 5. You can drag the slider for Temperature and see the contours for the Predicted Rating change. Each point represents a mixture of the three fish. Any given mixture of fish types receives different ratings according to the temperature at which the patties are baked.

In this example, the red shaded area shows the mixture of fish that results in a rating of 5 to 5.5. Any purple areas show the mixture of fish that results in a rating of 4 and below. At 400 degrees, a mixture of mostly sheepshead and mullet with very little croaker results in a rating of 5 and above.

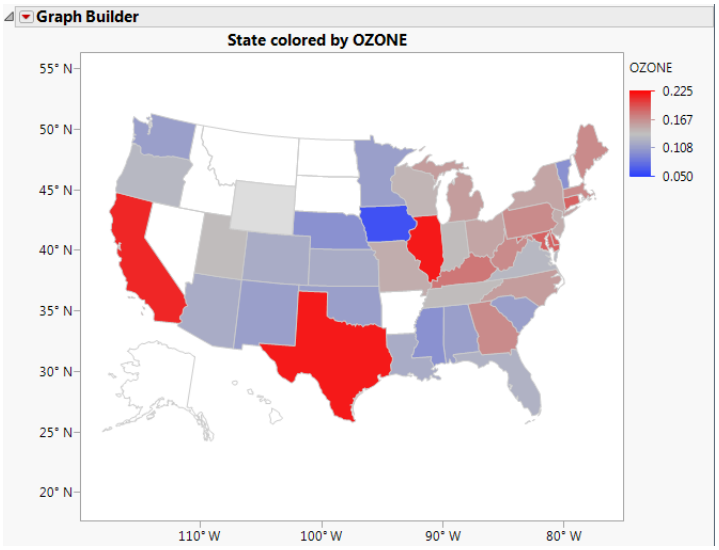
Chapter 12

Maps

Add Maps or Custom Shapes to Enhance Data Visualization

JMP transforms numbers and geographic data into compelling images, and turns simple tables of numbers into captivating pictures that bring the story in your data to life. JMP can help you display your data on geographical maps. Choose from built-in high-quality images. Select Street Map Service or Web Map Service to get custom map images from the Internet. JMP includes shape files for borders or many geographic regions and lets you add your own custom shapes, such as for a manufacturing plant or campus.

Figure 12.1 Example of a Map



Contents

Overview of Mapping 255
Example of Creating a Map in Graph Builder 256
Graph Builder 257
Map Shape 257
Color 259
Size 260
Custom Map Files 263
Background Maps 269
Images in Maps 271
Boundaries 276
Examples of Creating Maps 277
Louisiana Parishes Example 277
Hurricane Tracking Examples 285
Office Temperature Study 290
Boundary Map with Clipped Elements 295

Overview of Mapping

There are two types of map support in JMP: one where a map shows the data (Graph Builder) and one where a map provides context for the data (Background Maps). You can also create your own maps.

Graph Builder

You can interact with Graph Builder to create compelling visualizations of your data. JMP includes graphical support to display analyses using background maps and shape files. You can add color and geographical boundaries to maps through the following zones:

- The **Map Shape** zone assigns geographical boundaries to a map based on variables in the data table. The map shape value determines the x and y axes.

  Boundaries such as U.S. state names, Canadian provinces, and Japanese prefectures are installed with JMP. You can also create your own boundaries (geographical or otherwise) and specify them as a **Map Role** column property in the data table.
- The **Color** zone applies color based on a variable to geographical shapes.
- The **Size** element scales map shapes according to the size variable, minimizing distortion.

Background Maps

You can add background maps to any JMP graph through the Set Background Map window. You can use built-in background maps or connect to a Web Map Service (WMS) to display specialty maps like satellite images, radar images, or roadways. Right-click in a graph and select **Graph > Background Map** to choose from the following images and boundaries:

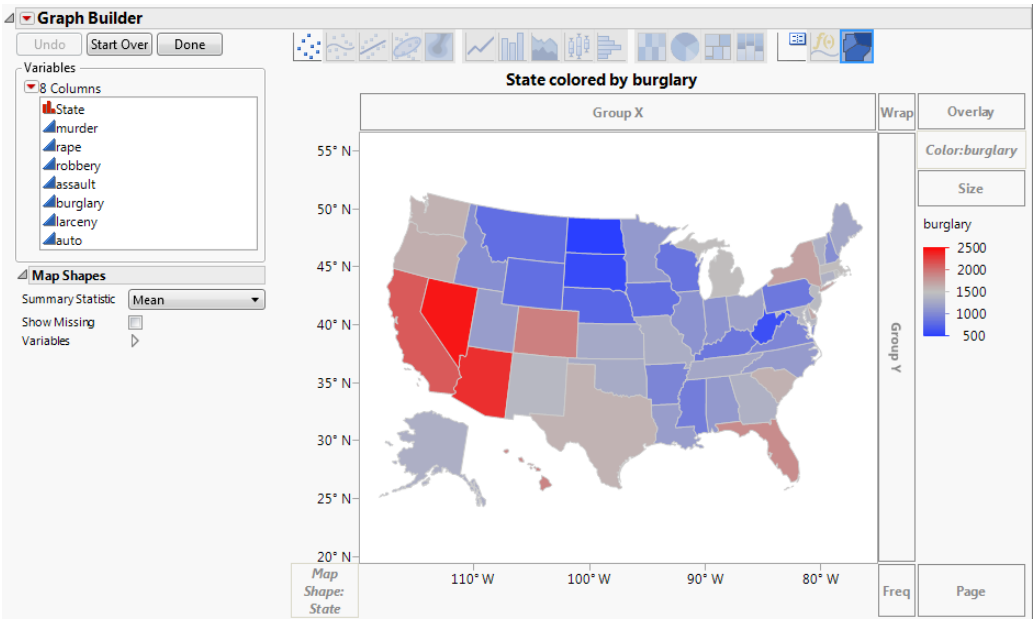
- **Simple Earth** and **Detailed Earth** maps are installed with JMP.
- **NASA server** provides maps using a WMS to show their most up-to-date maps.
- **Street Map Service** provides street maps. The OpenStreetMap and Open Database License links provide further information on the Street Map Service.
- **Web Map Service** lets you enter the URL for a website that provides maps using the WMS protocol. You can also specify the map layer.
- Boundaries for various regions.

Example of Creating a Map in Graph Builder

This example uses the Crime.jmp sample data table, which contains data on crime rates for each US state.

1. Select **Help > Sample Data Library** and open Crime.jmp.
2. Select **Graph > Graph Builder**.
3. Drag and drop State into the **Map Shape** zone.
4. Drag and drop Burglary into the **Color** zone.

Figure 12.2 Example of Burglary by State



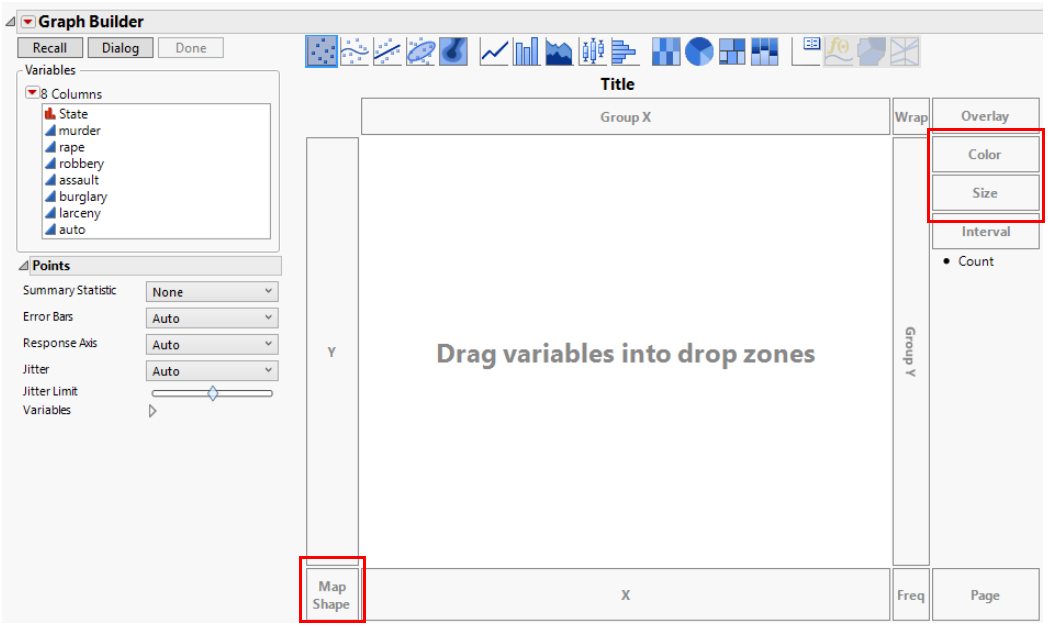
Note the following:

- The latitude and longitude appear on the Y and X axes.
- The legend shows the colors that correspond to the burglary rates. Since Burglary is a continuous variable, the colors are based on a continuous color theme. Figure 12.2 uses the JMP default continuous theme. You can change the theme under File > Preferences > Graphs.
- The map is projected so that relative areas are not distorted (the 49th parallel across the top of the US is not a straight line).

Graph Builder

Open a data table that contains geographic data. Launch Graph Builder by selecting **Graph** > **Graph Builder**. The primary element in the Graph Builder window is the graph area. The graph area contains drop zones (Map Shape, Color and Size), and you can drag and drop variables into the zones. From here you can map shapes for data tables that include place names.

Figure 12.3 The Graph Builder Window



Map Shape

When a column contains the names of geographical regions (such as countries, regions, states, provinces, counties), you can assign the column to the **Map Shape** zone. When a variable is dropped in **Map Shape**, Graph Builder looks for map shapes that correspond to the values of the variable and draws the corresponding map. The variable can have a column property that tells JMP where to find the map data. If not, JMP looks through all known map files. If you have a variable in the Map Shape zone, the X and Y zones disappear. The Map Shape zone is positional and influences the types of graph elements that are available.

Figure 12.4 Example of Cities.jmp After Dragging State to Map Shape

For each map there are two .jmp files; one for the name data (one row per entity) and one for coordinate data (many rows per entity). They are paired via a naming convention; xxx-Name.jmp and xxx-XY.jmp, where "xxx" is some common prefix.

Some examples of sample files that are shipped with JMP are:

- World-Name.jmp
- World-XY.jmp
- US-State-Name.jmp
- US-State-XY.jmp

Map Name Files

Each xxx-Name.jmp can contain any number of shape name columns, which are identified with a column property. Multiple name columns support localizations and alternate names styles (such as abbreviations), but a given graph usage uses only one column of names. The first column of the Name file contains unique Shape ID numbers in ascending order. JMP creates this columns for you. The column values match those in the -XY.jmp map data table Shape ID column.

Note: The Shape ID column can also be named Shape.

Figure 12.5 Example of US-State-Name.jmp

	Shape ID	Name	State FIPS	USPS Code	Abbr	Max Grow ...	Max Grow X Offset	Max Grow Y Offset
1	1	Alabama	01	AL	Ala.	1	0	0
2	2	Alaska	02	AK	Alaska	1	0	0
3	3	Arizona	04	AZ	Ariz.	1	0	0
4	4	Arkansas	05	AR	Ark.	1	0	0
5	5	California	06	CA	Calif.	3	-1.7	-3
6	6	Colorado	08	CO	Colo.	1	0	0
7	7	Connecticut	09	CT	Conn.	4	0.8	-1.1
8	8	Delaware	10	DE	Del.	3	0.5	-0.3
9	9	District of Columbia|Washi...	11	DC	D.C.	70	2	-2.5
10	10	Florida	12	FL	Fla.	4	2	-2
11	11	Georgia	13	GA	Ga.	1	0	0

Map XY Files

Each xxx-XY.jmp file has four columns. Each row is a coordinate in some shape. Each shape is made of one or more parts. Each part is a closed polygon. The first column is the same Shape ID as in the xxx-Name file. The second column is the Part ID. The next two columns are X and Y.

Figure 12.6 Example of US-State-XY.jmp

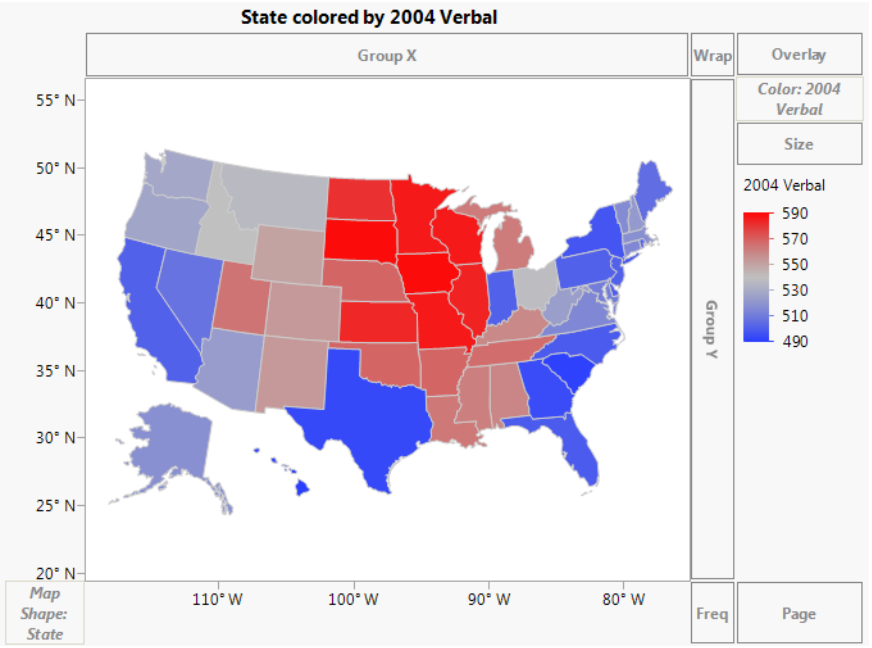
	Shape ID	Part ID	X	Y
1	1	1	85.6052° W	34.9847° N
2	1	1	85.1844° W	32.8613° N
3	1	1	84.9630° W	32.4242° N
4	1	1	85.0071° W	32.3284° N
5	1	1	84.8909° W	32.2615° N
6	1	1	85.0611° W	32.1341° N
7	1	1	85.1407° W	31.8575° N
8	1	1	85.0413° W	31.5410° N
9	1	1	85.1075° W	31.1865° N
10	1	1	85.0024° W	31.0007° N
11	1	1	87.5989° W	30.9975° N

Color

The Graph Builder platform lets you adds color to create choropleth maps. A choropleth map shows statistical differences in a geographic area while maintaining the proportion of the statistical variable.

Drag a column containing geographic place-names, like countries, regions, states, or provinces, into the **Map Shape** zone and create a map. Then drag a column to the **Color** zone to color the map by that column. The categorical or continuous color theme selected in your Preferences is applied to each shape.

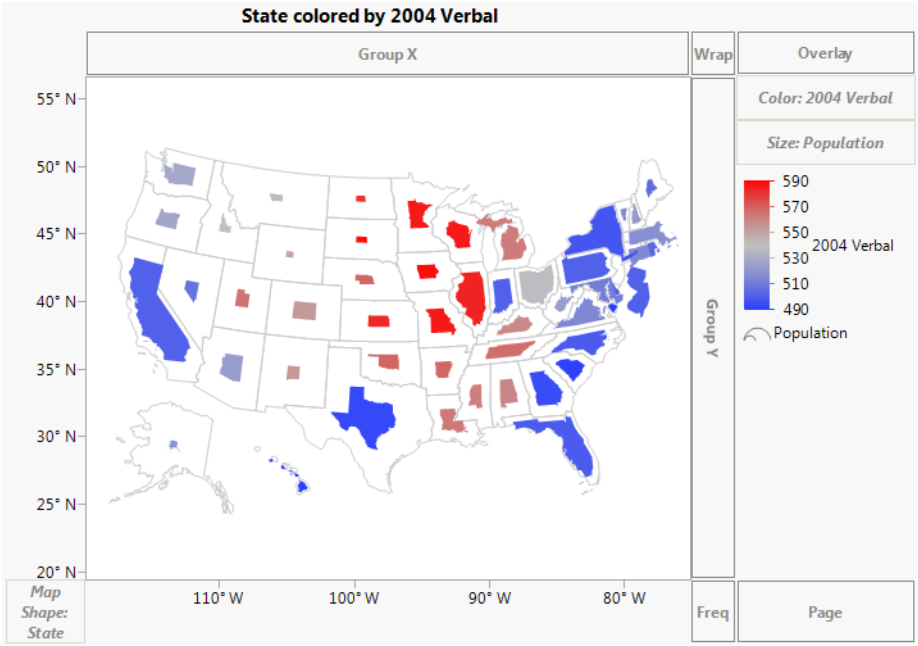
Figure 12.7 Example of SAT.jmp After Dragging 2004 Verbal to Color



Size

Use the **Size** element to scale map shapes according to the size variable, minimizing distortion.

Figure 12.8 Example of SAT.jmp After Dragging Population to Size



Customizing Graphs

To change colors and transparency for a map, right-click the color bar in the legend. The right-click options vary, depending on whether the color variable is continuous or categorical (nominal or ordinal). However, for both types of variables, you can change the transparency.

To change the transparency of a graph:

1. Right-click the color of the variable level on the color bar that you want to change and select **Transparency**.
2. Specify the transparency between 0 (clear) and 1 (opaque).
3. Click **OK**.

You can also change the transparency of images (for example, Simple Earth and Detailed Earth). To set the transparency, right-click over the graph and select **Customize...**. This brings up the Customize Graph window, where you can select the Background Map and assign a value for transparency. A valid value for transparency goes from 0.0 (completely transparent) to 1.0 (completely opaque).

Categorical (nominal or ordinal) variables use a singular coloring system, where each level of the variable is colored differently.

To change the color of one of the variable levels:

1. Right-click the color of the variable level that you want to change and select **Fill Color**.
2. Select the new color.

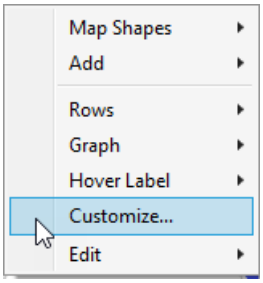
Continuous variables use a color gradient.

To change the color theme:

1. Right-click the color bar and select **Gradient**.
2. In the Gradient Settings window, select a different **Color Theme**.

Graphs consist of markers, lines, text, and other graphical elements that you can customize. If you right-click an image, there are several options for working with the graph. The options differ based on what you clicked. See "Gradient Settings" and *Using JMP*. Below are a few options.

Figure 12.9 Right-click Menu for Graphics



- Map Shapes:
  - Change To - Provides options for changing the map into another graph type.
  - Summary Statistics - Provides options for changing the statistic being plotted.
  - Show Missing Shapes - Shows or hides missing data from a map (turned off by default). Missing Shape means that there are some shape names that exist in the map file but not in the data table for analysis.
  - Remove - Removes the current map shape.
- Customize - You can change the properties of the graph such as contents, grid lines, or reference lines. The graphical elements that you can customize differ for each graph. Select **Background Map** to change the transparency of a background map or **Map Shape** to change the line color, line style and width, fill color, missing shape fill or missing value fill. Click **Help** in the Customize Graph window for a more detailed explanation of the customize options.

To change the shape file:

1. Right-click the Map Shapes zone and select **Set Shape File**.
2. Navigate to the shape file that contains the name data.
3. Click **Open**.
4. (Optional) To revert the changes, right-click the Map Shapes zone and select **Reset Shape File**.

Custom Map Files

You can create your own map files by following the same pattern as the built-in files. To add your own map files, you need two things: a series of XY coordinates for the vertices of the polygons that describe the shape, and a set of names for each polygon. Data and shape attributes are required to map custom shapes so that you can add your own shapes to JMP. There are two common sources for data like this: Esri shapefiles and SAS/GRAPH map data sets.

In order for JMP to automatically find your files, place them in the following directory:

- On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
- On macOS: /Users/<user name>/Library/Application Support/JMP/Maps

Note: On Windows, in JMP Pro, the "JMP" folder is named "JMPPro". Or, you can link the map files to your data files explicitly with the **Map Role** column property.

Note the following when creating map files:

- Each set of map files that you create must contain a -Name file and a -XY file.
- The first column in both files must be the ascending, numeric Shape ID variable. The -Name file can contain any other columns. The shapes are built by rows. The XY coordinates have to go around the shape rather than just define the convex hull of the shape.
- For the **Map Role** column property, columns that are marked with the **Shape Name Definition** are searched for shape identification and must contain unique values.
- If you import an Esri SHP file, it is opened in the correct format. -Name files commonly have a .dbf extension. See "Esri Shapefiles".
- SAS/GRAPH software includes a number of map data sets that can be used with JMP. See "SAS/GRAPH Map Data Sets".

You might want to create choropleth maps of other non-geographic regions (for example, a floor of an office building). Simply, add the two shape files for your non-geographic space. If you do not have XY coordinates, but you do have a graphic image of the space, you can use the Custom Map Creator add-in for JMP. With this add-in, you can trace the outlines of the space and JMP creates the -XY and -Name files for you. You can download this add-in from the JMP File Exchange page.

Map Role

You can specify the attributes and properties of a column in a data table within the Column Info window in Column Properties. The **Map Role** property is set for a column like other column properties in the Column Info window.

If you have created your own data table that contains boundary data (such as countries, regions, states, provinces, or counties) and you want to see a corresponding map in Graph Builder, use the **Map Role** property within Column Properties. Each pair of map files that you create must contain a -Name file and a -XY file.

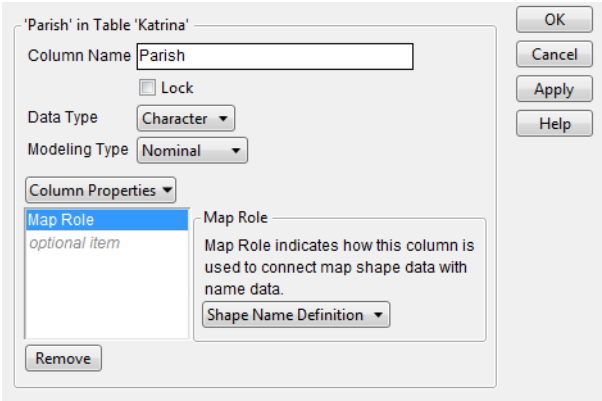
Note the following:

- If the custom boundary files reside in the default custom maps directory, then you need to specify only the **Map Role** property in the -Name file.
- If the custom boundary files reside in an alternate location, specify the **Map Role** property in the -Name file and in the data table that you are analyzing.
- The columns that contain the **Map Role** property must contain the same boundary names, but the column names can be different.

To add the Map Role property into the -Name data table:

1. Right-click the column containing the boundaries and select **Column Properties > Map Role**.
2. Select **Shape Name Definition** below **Map Role**.
3. Click **OK**.
4. Save the data table.

Figure 12.10 Shape Name Definition Example



To add the Map Role property into the data table that you are analyzing:

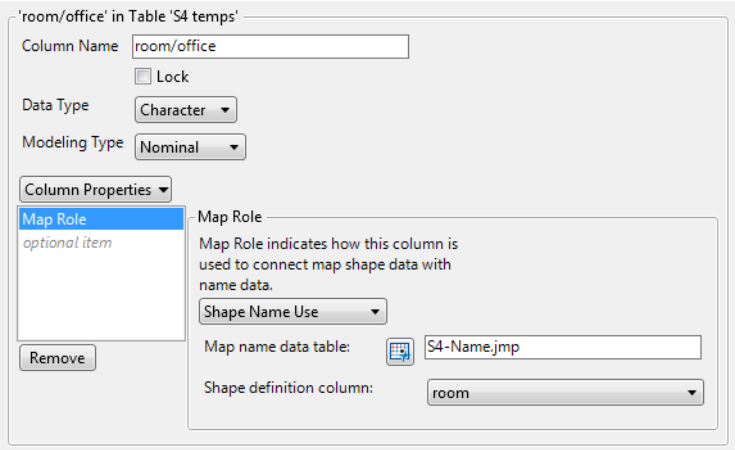
Note: Perform these steps only if your custom boundary files do not reside in the default custom maps directory.

1. Right-click the column containing the boundaries and select **Column Properties > Map Role**.
2. Select **Shape Name Use** below **Map Role**.
3. Next to **Map name data table**, click to browse to a -Name map data table. You can enter the relative or absolute path.

   If the map data table is in the same folder, enter only the filename. Quotation marks are not required when the path contains spaces.
4. From the **Shape definition column** list, select the column in the map data table whose values match those in the selected column.

   Figure 12.11 shows an example of the room/office column in the S4 Temps.jmp sample data table.

Figure 12.11 Shape Definition Column Example



5. Click **OK.**
6. Save the data table.

When you generate a graph in Graph Builder and assign the modified column to the **Map Shape** zone, your boundaries appear on the graph.

For numeric columns, the Format Menu appears in the Column Info window. Specify the format to tell JMP how to display numbers in the column. Latitude and Longitude for geographic maps are located under **Format > Geographic** when customizing axes and axes labels.

Geographic Shows latitude and longitude number formatting for geographic maps. Latitude and longitude options include the following:

- DDD (degrees)
- DMM (degrees and minutes)
- DMS (degrees, minutes, and seconds)

In each format, the last field can have a fraction part. You can specify the direction with either a signed degree field or a direction suffix. To show a signed degree field, such as *-59°00′00″*, deselect **Direction Indicator**. To show the direction suffix, such as *59°00′00″ S*, select **Direction Indicator**.

To use spaces as field separators, deselect **Field Punctuation**. To use degrees, minutes, and seconds symbols, select **Field Punctuation**.

Esri Shapefiles

The Esri shapefile is a vector data format that contains data about geographic features such as terrain and oceans. It is developed and regulated by Esri as a specification for geographic mapping software.

Each shapefile is a set of files with the same name and different extensions.

main file (.shp)

The .shp file contains sequences of points that make up polygons. When opened with JMP, a .shp file is imported as a JMP table.

- The Shape column is added during import to uniquely identify each geographic region. Each coordinate point is in a separate row.
- The Part column to indicate discontiguous regions, and the XY coordinates (in latitude and longitude degrees).

JMP supports two-dimensional .shp files (no elevation information).

dBase Table (.dbf)

You add a Shape ID column to the .dbf table, which maps to the Shape column in the .shp file. Add any number of columns that provide common names or values to refer to specific regions.

To convert an Esri shapefile to a JMP map file:

1. Open the .shp file in JMP.
2. Make sure that the Shape column is the first column in the .shp file. Add formatting and axis settings for the X and Y columns (optional). Graph Builder uses those settings for the X and Y axes.
3. Save the .shp file as a JMP data table to the Maps folder with a name that ends in -XY.jmp.
4. Open the .dbf file.
5. Add a Shape ID column as the first column in the table. This column should be the row numbers from 1 to n, the number of rows in the data table.

   Note: You can use **Cols > New Columns > Initialize Data > Sequence Data**) to fill the column with sequential numbers.

6. Assign the **Map Role** column property to any column that you use for place names in the Shape role of Graph Builder. To do this, right-click at the top of the column and select **Column Properties > Map Role**.
7. Select **Shape Name Definition** from the drop-down box in the property definition.

8. Save the table as a JMP data table with a name that matches the earlier table and that ends in -Name.jmp.

JMP looks for these files in two locations. One location is shared by all users on a machine. This location is:

- Windows: C:/Program Files/SAS/JMP/16/Maps/
- macOS: /Library/Application Support/JMP/16/Maps

The other location is specific for an individual user:

- On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
- On macOS: /Users/<user name>/Library/Application Support/JMP/Maps

Note: On Windows, in JMP Pro, the JMP folder is named JMPPro.

SAS/GRAPH Map Data Sets

SAS/GRAPH® software includes a number of map data sets that can be converted for use with JMP. The data sets are in the Maps library. The traditional map data sets contain the XY coordinate data and the feature table contains the common place names. You need to convert both of these files to JMP data tables for use with JMP.

Most of the traditional map data sets have unprojected latitude and longitude variables in radians. The data sets can be used with JMP once they have been converted to degrees and the longitude variable has been adjusted for projection. The following is a DATA step that shows the conversion process for the Belize data set.

```
data WORK.BELIZE;
keep id segment x y;
rename segment=Part;
set maps.belize;
if x NE .;
if y NE .;
y=lat*(180/constant('pi'));
x=-long*(180/constant('pi'));
run;
```

You can now import the converted file and save it as Belize-XY.jmp.

The next step is to import the matching feature data set (in this case: MAPS.BELIZE2). After importing the feature data set, move the ID column to the first position in the data table. Then assign the Map Role column property to the columns that you use for place names in the Shape role of Graph Builder. To do this, right-click the top of the column and select **Column Properties** > **Map Role**. Then select **Shape Name Definition** from the drop-down box in the property definition. For MAPS.BELIZE2, use the IDNAME column. Save the feature data table as Belize-Name.jmp.

To convert SAS maps, download the SAS to JMP Map Converter add-in from the JMP File Exchange page. For each map, the add-in reads the data from the two SAS map tables, rearranges and formats the data and then places it into the two JMP map tables.

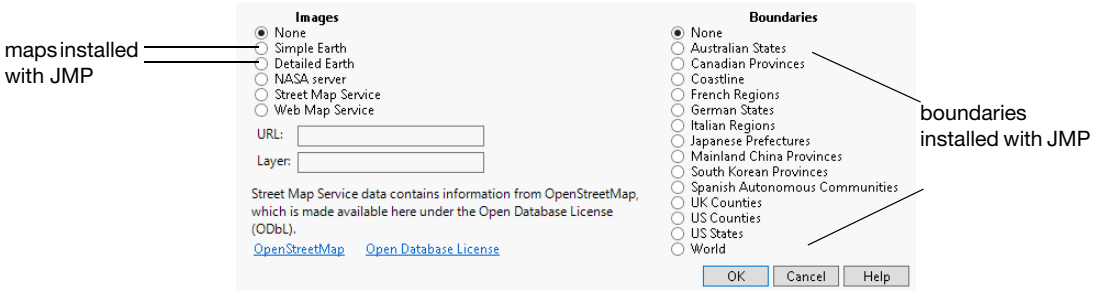
Background Maps

Adding map images and boundaries to graphs provides visual context to geospatial data. Affixing a background map generates an appealing map, providing your data a geographic context and giving you a whole new way to view your data. For example, you can add a map to a graph that displays an image of the U.S. Another option is displaying the boundaries for each state (when data includes the latitudes and longitudes for the U.S.). There are different types of background maps. Some maps are built into JMP and are delivered as part of the JMP install. Other maps are retrieved from an Internet source, and still other maps are user-defined.

The data should have latitudinal and longitudinal coordinates. Otherwise, the map has no meaning in the context of the data. The X and Y axes also have range requirements based on the type of map. These requirements are described in the following sections. Simply plot longitude and latitude on the X and Y axes, and then right-click within the graph and select **Graph** > **Background Map**.

The Background Map window shows two columns of choices: Images and Boundaries. On the left of the window that you can select from two built-in map images, or you can connect to a Web Map Service to retrieve a background image. On the right side of the window, you can select political boundaries for a number of regions.

Figure 12.12 Example of Background Map Options



The following Background Map options are available:

Images

None Removes the background map that you selected in the Images column.

Simple Earth Shows a map of basic terrain. For examples, run the associated scripts in these sample data tables: SAS Offices.jmp and Aircraft Incidents.jmp.

Detailed Earth Shows a high-resolution map with detailed terrain.

NASA Server Shows a map from the NASA server. Requires an Internet connection.

Street Map Service Shows a map with an appropriate amount of detail based on the display's zoom level. This enables you to zoom down to the street level. For example, run the associated script in the sample data table SAS Offices.jmp.

Web Map Service Shows a map from the Uniform Resource Locator (URL) and the layer that you specify. Requires an Internet connection.

Boundaries

None Removes the boundaries that you selected in the Boundaries column.

Boundaries for various regions Shows borders for the map regions, such as Canadian provinces, U.S. counties, U.S. States, and world countries. The list varies based on your location. The maps that you created from Esri shapefiles are also listed here.

Two tools are especially helpful when you are viewing a map:

- The grabber tool ( ) lets you scroll horizontally and vertically through a map.
- The magnifier tool ( ) lets you zoom in and out.

Images in Maps

Every flat map misrepresents the surface of the Earth in some way. Maps cannot match a globe in truly representing the surface of the entire Earth. A map projection is used to portray all or part of the round Earth on a flat surface. This cannot be done without some distortion. Every projection has its own set of advantages and disadvantages. A map can show one or more, but not all, of the following: true direction, distance, area, or shape. JMP uses a couple of projections (Albers Equal Area Conic and Kavrayskiy VII) for its maps. Within Images, you can select from two built-in map images, or you can connect to a Web Map Service to retrieve a background image.

Earth Images Installed with JMP

JMP provides two levels of earth imagery; simple and detailed. Both maps show features such as bodies of water and terrain. However, detailed maps show more precise terrain. And with detailed maps, you can zoom in farther, and the map features remain clear. Image maps are raster images. The maps wrap horizontally, so you continue to see map details as you scroll from left to right. The maps do not wrap vertically. Beyond the -90 and 90 y-axis range, a plain background appears instead of the map.

Figure 12.13 Examples of Simple and Detailed Maps

Simple Earth Image

Detailed Earth Image

As its name suggests, Simple Earth is a relatively unadorned image of the earth's geography. It does not show clouds or arctic ice, and it uses a green and brown color scheme for the land and a constant deep blue for water. Detailed Earth has a softer color scheme than Simple Earth, lighter greens and browns for the land, as well as variation in the blue for the water. Detailed Earth also has a slightly higher resolution than Simple Earth. The higher resolution lets you zoom into a graph further with Detailed Earth than with Simple Earth before the quality of the background image begins to blur.

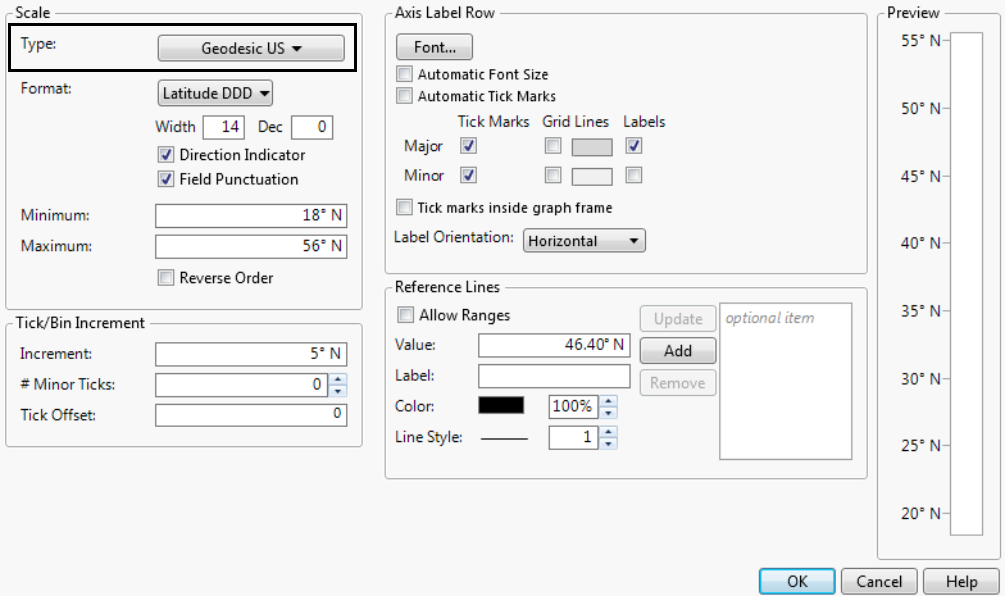
Another feature of Simple Earth and Detailed Earth is the ability to wrap. The Earth is round, and when you cross 180° longitude, the Earth does not end. The longitudinal value continues from -180° and increases. The map wraps continuously in the horizontal direction, much as the Earth does. The background map does not wrap in the vertical direction.

Simple Earth and Detailed Earth both support a geodesic scaling. In Figure 12.13, the Earth appears as a rectangle, where the width is twice as wide as the height. If we were to take this rectangle and roll it up, we would have a cylinder. In reality, we know that the Earth does not form a cylinder, but rather a sphere. You can use a geodesic scaling, which transforms the map to a more realistic representation of the Earth. To use the geodesic scaling, change the type of scale on the axes.

To change the axes scale:

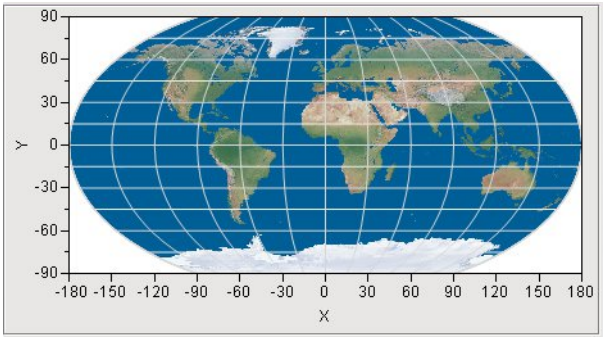
1. Right-click the X or Y axis and then select **Axis Settings.**
2. Change the Scale Type to **Geodesic** or **Geodesic US**.

Figure 12.14 Y Axis Setting Window



Both choices transform the map to a geodesic scaling. Use Geodesic US if you are viewing a map of the continental US and you want Alaska and Hawaii to be included in the map. It is important to note that you must set the scale to geodesic for both axes to get the transformation. You will not see a change in the map after setting only one of the axes. In the following figure, Simple Earth is used as the background map with the axes set to use a geodesic scale. The axes lines are turned on as well. Notice the longitudinal lines are now curved, instead of straight.

Figure 12.15 Simple Earth Background Map - Axes Set to Geodesic Scale

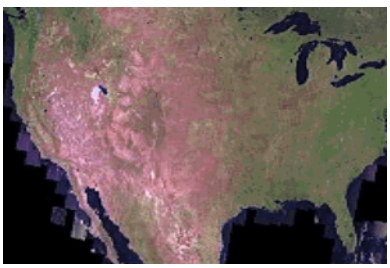


Since Detailed and Simple Earth are built into JMP, these options work anytime, without a network connection. However, these images might not be all that you want, or they might not be detailed at the resolution that you need. If this is the case, and if you have an Internet connection, you can connect to a Web Map Service to retrieve a map image that meets your needs.

Maps from the Internet

The National Aeronautics and Space Administration (NASA) and other organizations provide map image data using a protocol called Web Map Service (WMS). These maps have the advantage of showing the most up-to-date geographical information. However, the display of the maps can be slow depending on the response time of the server, and the sites can change or disappear at any time. An Internet connection is required to access the information.

Figure 12.16 Examples of NASA and WMS Maps

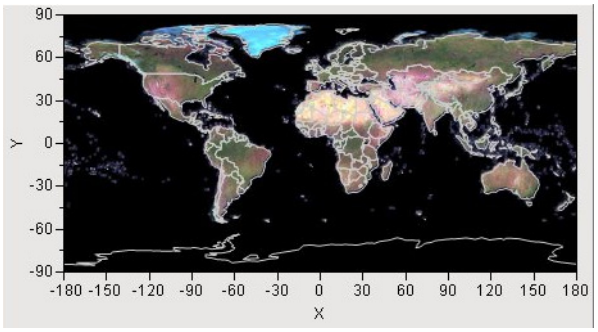


NASA image

WMS image from the MetaCarta™ Server

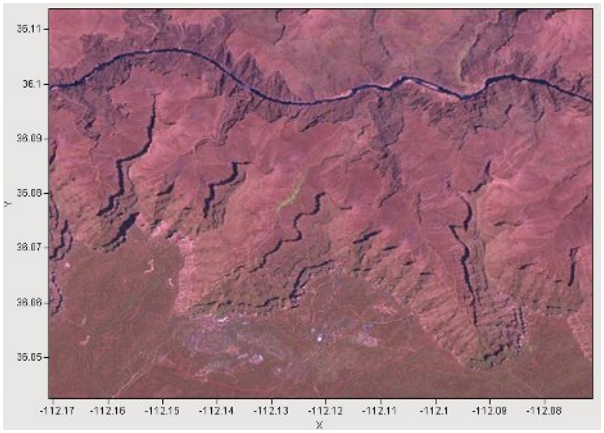
The NASA server provides maps for the entire Earth. The following figure displays the Earth using the NASA server as its source for the background map. The boundary map shows the outlines of the countries.

Figure 12.17 NASA Server Map Example



Not only does this server cover the entire Earth, but you can also zoom in on a much smaller area of the Earth and still get a reasonable map. The following figure displays the Colorado River running through the Grand Canyon in Arizona. The Grand Canyon Village is visible in the bottom of the map.

Figure 12.18 NASA Server Map Example - Zoom In on Colorado



If you look at the axes values, you can see that the area is less than 1/10° by 1/10°. The Simple Earth and Detailed Earth background maps do not display that type of resolution. The NASA server provides a fairly detailed view of any land mass on Earth. Water, however, is simply filled in as black. The NASA server is free to access, but it is also limited in availability. If the server is temporarily unavailable or becomes overloaded with requests, it delivers an error message instead of the requested map.

Another Internet-based option for background maps is a Web Map Service (WMS). The WMS option enables you to specify any server that supports the WMS interface. The NASA server is an example of a WMS server, but we have provided the URL and a layer name for you. With the WMS option, you must know the URL to the WMS server and a layer name supported by the server. Most WMS servers support multiple layers. For example, one layer can show terrain, another layer can show roads, and still another layer can include water, such as rivers and lakes. By specifying the URL for the server and the layer, JMP can make a request to the server and then display the map that is returned.

Unlike with simple and detailed maps, WMS maps do not wrap. You can scroll horizontally and vertically. However, beyond the -180 to 180 (x axis) and -90 to 90 (y axis) ranges, a plain background appears instead of the map. The limits of the axes are used to define the limits of the map that is displayed.

In order to use the WMS option for a background map, you need to decide which WMS server to use. There are many WMS servers freely available from the Internet. Most of them provide maps only for a particular area of the world, and each of them supports their own layers. So you have to search for the appropriate WMS server for your particular situation.

You can search for WMS servers on the Internet using your favorite search engine. Once you find one, you need to discover the layers that it supports. For this, you can use the WMS Explorer add-in. The WMS Explorer add-in generates a list of all the layers available on a server. You can select a layer from the list to see what it looks like. You can download the WMS Explorer add-in from the JMP File Exchange page.

Note: To use the WMS Explorer add-in and the WMS background map capabilities of JMP, your computer must be connected to the Internet.

To locate a server, launch the add-in through the menu items **Add-Ins > Map Images > WMS Explorer**. The add-in presents a text box for entering the url of a known WMS server. Alternatively, you can make a selection from a drop-down list of pre-discovered WMS servers (the list can be out of date). After specifying a WMS server, select **Get Layers**. Using Get Layers is not necessary if selecting from the drop-down list or if clicking **Enter** after entering a URL. This sends a request to the WMS server for a list of layers that the server supports. The returned list appears in the list box on the left, labeled **Layers**. A map of the world appears as an outline in the graph to the right. Selecting a layer makes a request to the WMS server to return a map, using the specified layer, that represents the entire earth. Selecting a different layer generates a different map.

The default maps do not cover the entire earth (for example, some WMS servers might provide mapping data for a particular county, within a state). In that case, it is likely that selecting a layer does not generate any visible map. You might have to zoom in on the appropriate area before any image map is visible. The standard JMP toolbar is available in the add-in window and the zoom tool works just like it does in any JMP window.

The graph is a typical graph in JMP, which means that all the regular JMP controls are available to you. You can adjust the axes or use the zoom tool (found on the menu bar) just as you would in JMP. You can also right-mouse-click to select **Size/Scale > Size to Isometric** to return the graph to a proper aspect ratio. You can also select Background Map, where you can adjust the boundary map.

Once a desirable map is determined, note the URL in the text box at the top and the selected layer in the Layers list. This is the information that you need to enter in the background map window when WMS is selected as the type of image background map.

Because requests are being made to a server across the Internet, there are a number of conditions that can generate an error. WMS servers often have limited availability and sometimes are not available at all. Occasionally a WMS server might return a name of a layer that it no longer supports. In these types of cases (and others), a server usually returns an error message in lieu of a map. If that happens, the error message is displayed below the Layers list in an area labeled **Errors**.

Boundaries

JMP can display boundaries (such as U.S. states or French region boundaries). These boundaries draw an outline around a defined area and can be displayed alone on a graph or combined with image data. Several boundaries are installed with JMP. Alternatively, you can create your own boundaries from Esri shapefiles or from scratch. Because of this, the list of Boundaries that you see in the Set Background Map window can be different.

When you add shape files to the built-in locations in JMP, they are available for the Graph Builder platform and for the Boundaries option in the Background Map window. In this way, you can add more political boundaries for use with background maps. Boundary-style maps are vector-based shapes.

Figure 12.19 Example of U.S. State Boundaries

Add a Background Map and Boundaries

To add a background map and boundaries:

1. Right-click a blank area on the graph and select **Background Map** (or select **Graph > Background Map** in Graph Builder).

   The Set Background Map window appears (Figure 12.12).
2. To display a background, do one of the following:
   - Select **Simple Earth**, **Detailed Earth**, **NASA server**, or **Street Map Service** in the Images column.
   - Select **Web Map Service** and paste a WMS URL next to URL. Enter the layer identifier next to **Layer**.
3. To display geographic borders on the map, select an option in the Boundaries column (If you installed your own boundary shapefiles, they are also listed in this column).
4. Click **OK**.

   If the NASA map, Street Map, or WMS map does not appear after you add it, the map server might not be available. View the error log to verify the problem.

Examples of Creating Maps

- "Louisiana Parishes Example"
- "Hurricane Tracking Examples"
- "Office Temperature Study"
- "Boundary Map with Clipped Elements"

Louisiana Parishes Example

In this example you work with custom map files and then create custom maps in two different ways:

- Set up custom map files initially and save them in the predetermined location. JMP finds and uses them in the future with any appropriate data.
- Point to specific predefined map files directly from your data. This step might be required each time you want to specify custom maps.

Set Up Automatic Custom Maps

Suppose that you have downloaded Esri shapefiles from the Internet and you want to use them as your map files in JMP. The shapefiles are named Parishes.shp and Parishes.dbf. These files contain coordinates and information about the parishes (or counties) of Louisiana.

Note: Pathnames in this section refer to the JMP folder. On Windows, in JMP Pro, the JMP folder is named JMPPro.

Save the .shp File

Save the .shp file with the appropriate name and in the correct directory.

1. In JMP, open the Parishes.shp file from the following default location:
   - On Windows: C:/Program Files/SAS/JMP/16/Samples/Import Data
   - On macOS: /Library/Application Support/JMP/16/Samples/Import Data

   Note: If you cannot see the file, you might need to change the file type to **All Files**.

   JMP opens the file as Parishes. The .shp file contains the x and y coordinates.
2. Save the Parishes file with the following name and extension: Parishes-XY.jmp. Save the file here:
   - On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
   - On macOS: /Users/<user name>/Library/Application Support/JMP/Maps
3. Close the Parishes-XY.jmp file.

Save the .dbf File

Perform the initial setup and save the .dbf file.

1. Open the Parishes.dbf file from the following default location:
   - On Windows: C:/Program Files/SAS/JMP/16/Samples/Import Data
   - On macOS: /Library/Application Support/JMP/16/Samples/Import Data

   Note: If you cannot see the file, you might need to change the file type to **All Files**.

   JMP opens the file as Parishes. The .dbf file contains identifying information.
2. Right-click the PARISH column and select **Column Info**.
3. Select **Column Properties > Map Role**.
4. Select **Shape Name Definition**.
5. Click **OK**.

6. Save the Parishes file with the following name and extension: Parishes-Name.jmp. Save the file here:
   - On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
   - On macOS: /Users/<user name>/Library/Application Support/JMP/Maps
7. Close the Parishes-Name.jmp file.

Create the Map in Graph Builder

Once the map files have been set up, you can use them. The Katrina.jmp data table contains data on Hurricane Katrina's impact by parish. You want to visually see how the population of the parishes changed after Hurricane Katrina.

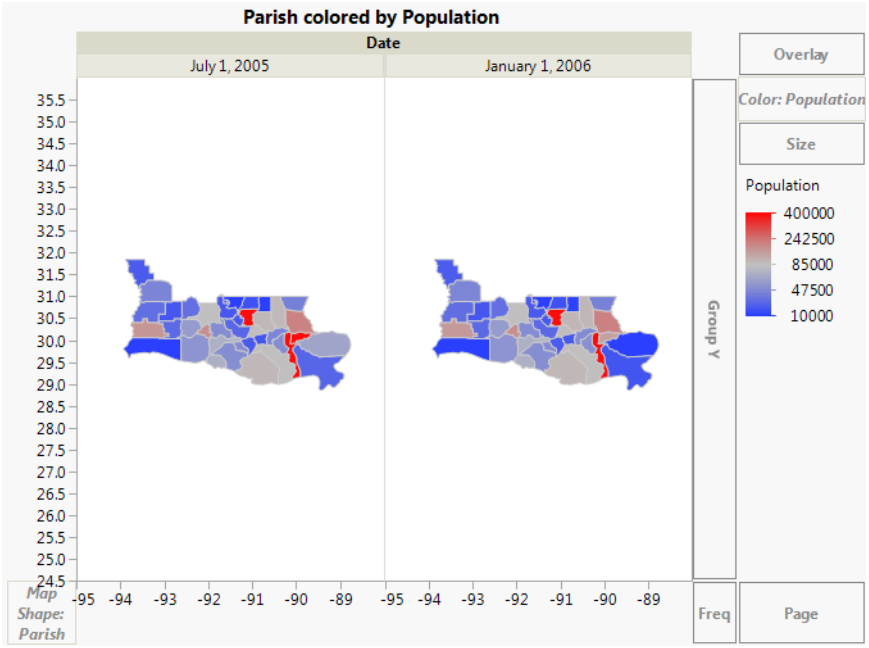
1. Select **Help > Sample Data Library** and open Katrina.jmp.
2. Right-click the Parish column and select **Column Properties > Map Role**.
3. Select **Shape Name Use**.
4. Click the **Map name data table** button and browse to select Parishes-Name.jmp, which you previously created.

   This tells JMP where the data tables containing the map information reside.
5. Select PARISH from the **Shape definition column** list.

   In Parishes-Name.jmp, the PARISH column has the **Shape Name Definition Map Role** property assigned. The column consists of map shape data for each parish.
6. Click **OK**.
7. Select **Graph > Graph Builder**.
8. Drag and drop Parish into the **Map Shape** zone.

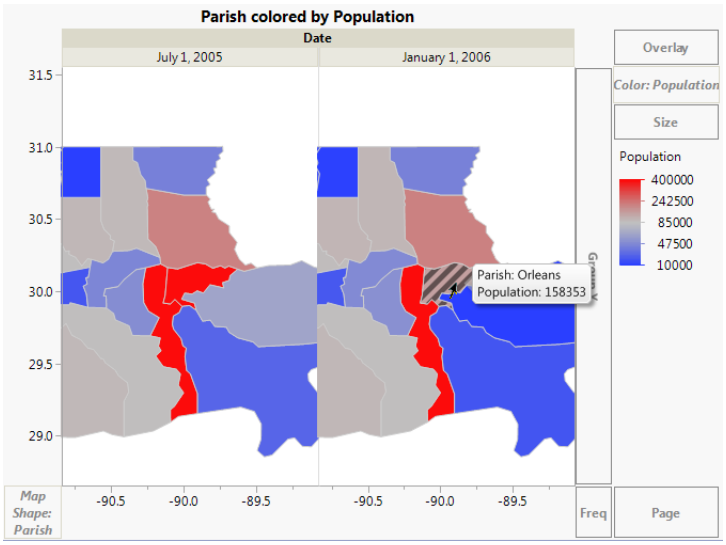
   The map appears automatically, since you defined the Parish column using the custom map files.
9. Drag and drop Population into the **Color** zone.
10. Drag and drop Date into the **Group X** zone.

Figure 12.20 Population of Parishes Before and After Katrina



11. Select the **Magnifier** tool to zoom in on the Orleans parish in both maps (Figure 12.21)

Figure 12.21 Orleans Parish



You can clearly see the drop in population as a result of Hurricane Katrina. The population of the Orleans parish went from 437,186 in July 2005 to 158,353 in January 2006.

Point to Existing Map Files Directly from Your Data

Suppose that you already have your custom map files and they are named appropriately. Your map files are US-MSA-Name.jmp and US-MSA-XY.jmp. They are saved in the sample data folder.

The PopulationByMSA.jmp data table contains population data from the years 2000 and 2010 for the metropolitan statistical areas (MSAs) of the United States. This example shows how the data table has been set up to create a map.

Add the Map Role Column Property

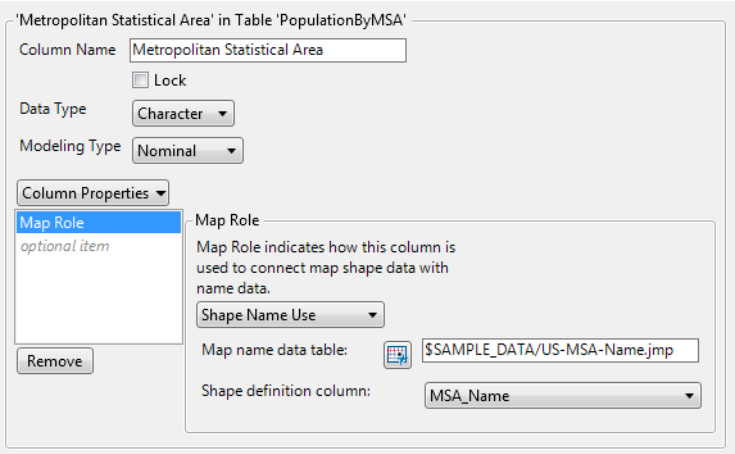
1. Select **Help > Sample Data Library** and open PopulationByMSA.jmp.
2. Right-click the **Metropolitan Statistical Area** column and select **Column Info**.
3. Select **Column Properties > Map Role**.
4. Select **Shape Name Use**.
5. Next to the **Map name data table**, type $SAMPLE_DATA/US-MSA-Name.jmp.

   This tells JMP where the data tables containing the map information reside.
6. Select MSA_Name from the **Shape definition column** list.

   MSA_Name is the specific column within the US-MSA-Name.jmp data table that contains the unique names for each metropolitan statistical area. Notice that the MSA_Name column has the **Shape Name Definition Map Role** property assigned, as part of correctly defining the map files.

Note: Remember, the Shape ID column in the -Name data table maps to the Shape ID column in the -XY data table. This means that indicating where the -Name data table resides links it to the -XY data table, so that JMP has everything that it needs to create the map.

Figure 12.22 Map Role Column Property



7. Click **OK**.

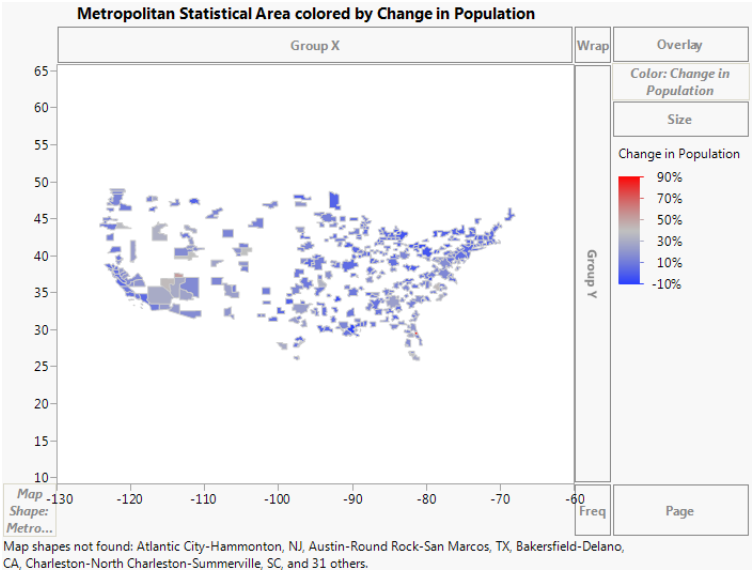
Create the Map in Graph Builder

Once the **Map Role** column property has been set up, you can perform your analysis. You want to visually see how the population has changed in the metropolitan statistical areas of the United States between the years 2000 and 2010.

1. Select **Graph > Graph Builder**.
2. Drag and drop **Metropolitan Statistical Area** into the **Map Shape** zone.

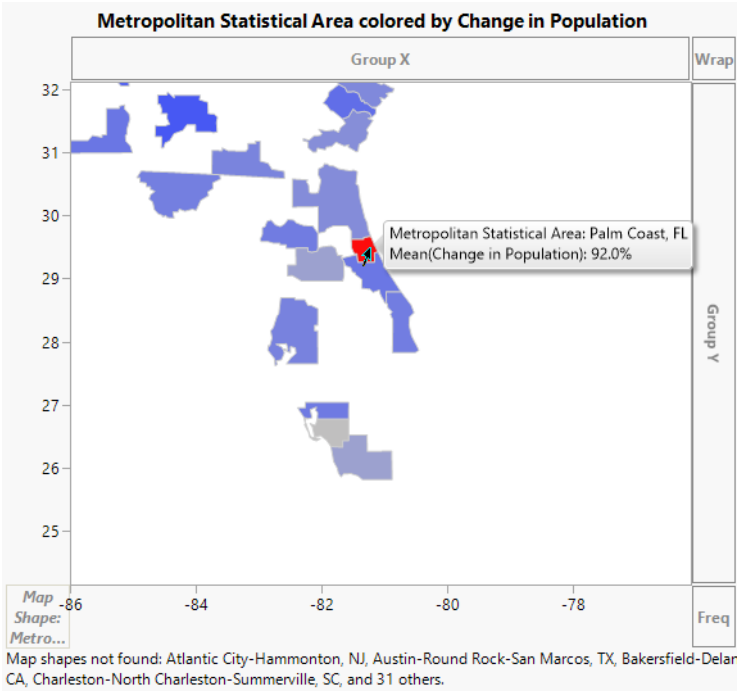
   Since you have defined the **Map Role** column property on this column, the map appears.
3. Drag and drop **Change in Population** to the **Color** zone.

Figure 12.23 Change in Population for Metropolitan Statistical Areas



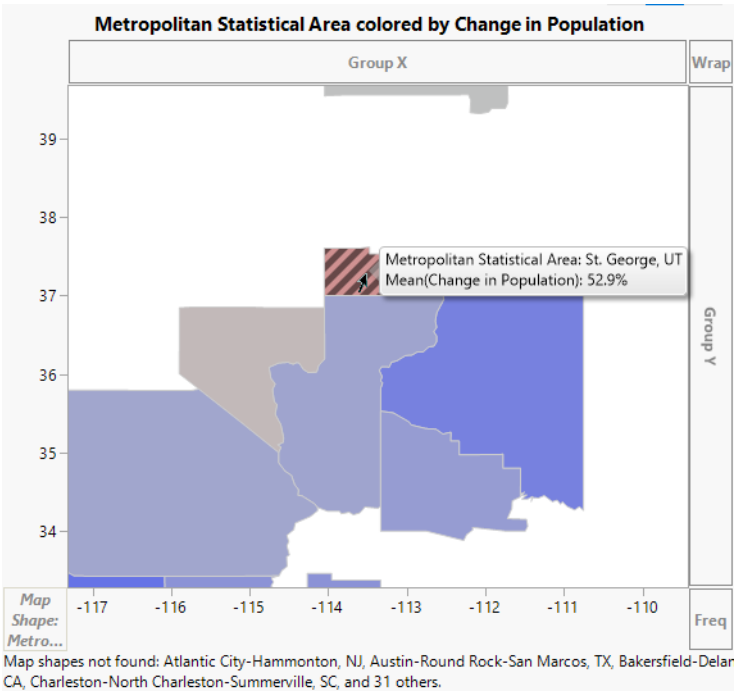
4. Select the **Magnifier** tool to zoom in on the state of Florida.
5. Select the **Arrow** tool and click the red area.

Figure 12.24 Population Change of Palm Coast, Florida



6. Select the **Magnifier** tool and hold down the Alt key while clicking on the map to zoom out.
7. Select the **Magnifier** tool and zoom in on the state of Utah.
8. Select the **Arrow** tool and click the area that is slightly red.

Figure 12.25 Population Change of St. George, Utah



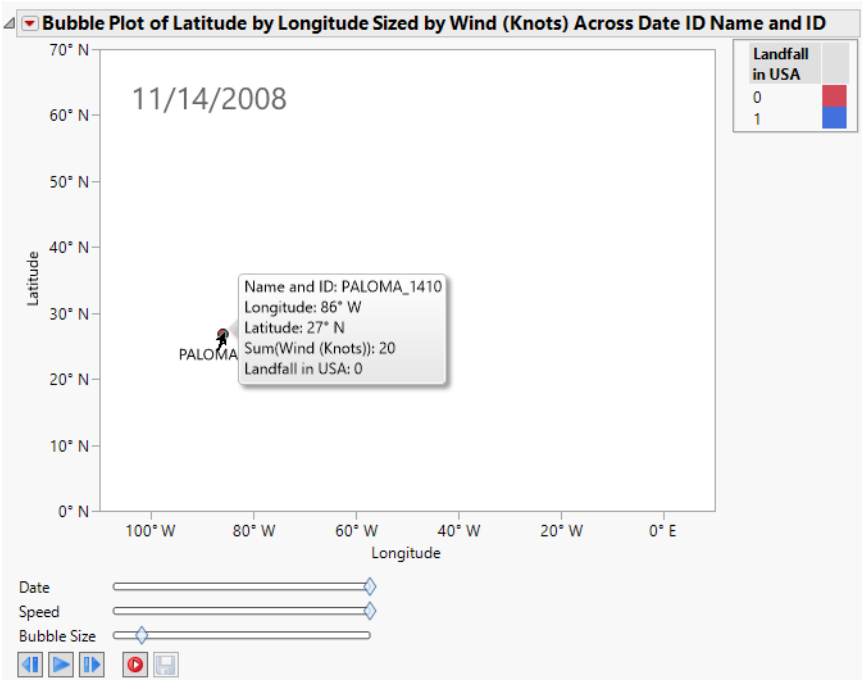
You can see that the areas of Palm Coast, Florida, and St. George, Utah had the most population change between 2000 and 2010. The Palm Coast area saw a population change of 92%, and the St. George area saw a population change of about 53%.

Hurricane Tracking Examples

This example uses the Hurricanes.jmp sample data table, which contains data on hurricanes that have affected the east coast of the United States. Adding a background map helps you see the areas the hurricanes affected. A script has been developed for this example and is part of the data table.

1. Select **Help > Sample Data Library** and open Hurricanes.jmp.
2. In the Table panel, click the green triangle next to the **Bubble Plot** script.
3. Drag the Date slider to the right as shown in Figure 12.26.
4. Click the red dot to display the name of the hurricane. The date appears in the upper left corner of the window. The red dot shows the location of Hurricane Paloma on November 14, 2008.

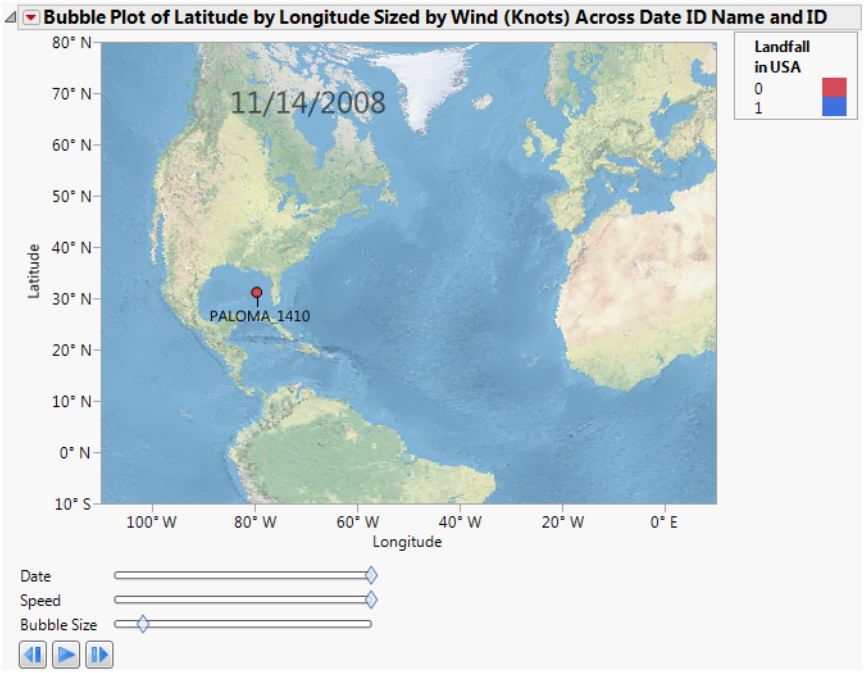
Figure 12.26 Bubble Plot of Hurricanes.jmp



Note that even though the location of the hurricane is plotted, it does not really tell us where it is. The axes information is there (27° North latitude and 86° West longitude), but we need a little more context. It is most likely over the middle of the Atlantic, but is it over a small island? This could make a big difference, especially for the inhabitants of the small island. Obviously, a map in the background of our graph would add a good deal of information.

5. Right-click the graph and select **Background Map**. The Set Background Map window appears (Figure 12.12).
6. Select **Detailed Earth** and click **OK**.

Figure 12.27 Bubble Plot of Hurricanes.jmp with Background Map

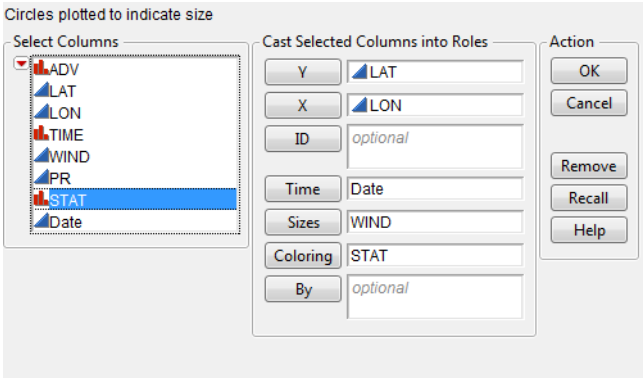


Now the coordinates make geographic sense. Click **Run** to view the animation of the hurricane data moving over the background map. Experiment with different options and view the displays. Adjust the axes or use the zoom tool to change what part of the world you are viewing. The map adjusts as the view does. You can also right-click the graph and select Size/Scale->Size to Isometric to get the aspect ratio of your graph to be proportional.

The next example uses the Katrina Data.jmp sample data table, which contains data on hurricane Katrina such as latitude, longitude, date, wind speed, pressure, and status. Adding a background map helps you see the path the hurricane took and impact on land based on size and strength. A script has been developed for this example and is part of the data table.

1. Select **Help > Sample Data Library** and open Katrina Data.jmp.
2. Select **Graph** > **Bubble Plot**.
3. Select LAT and click **Y**.
4. Select LON and click **X**.
5. Select Date and click **Time**.
6. Select WIND and click **Sizes**.
7. Select Stat and click **Coloring**.

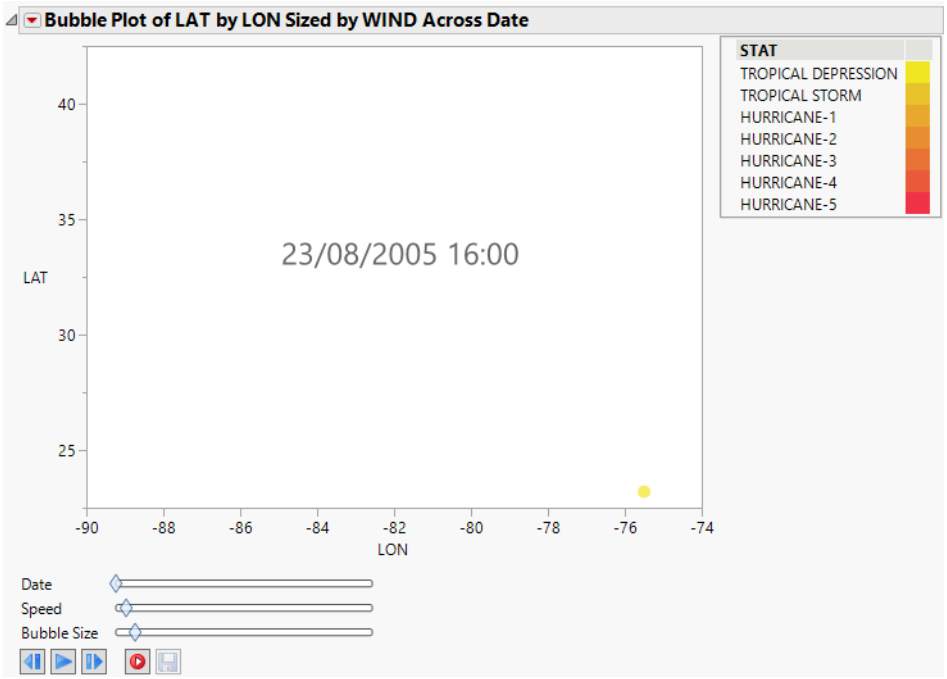
Figure 12.28 Bubble Plot Setup of Katrina Data.jmp



8. Click **OK**.

   The following image appears. The yellow dot shows the location of Tropical Depression Katrina on August 23, 2005.

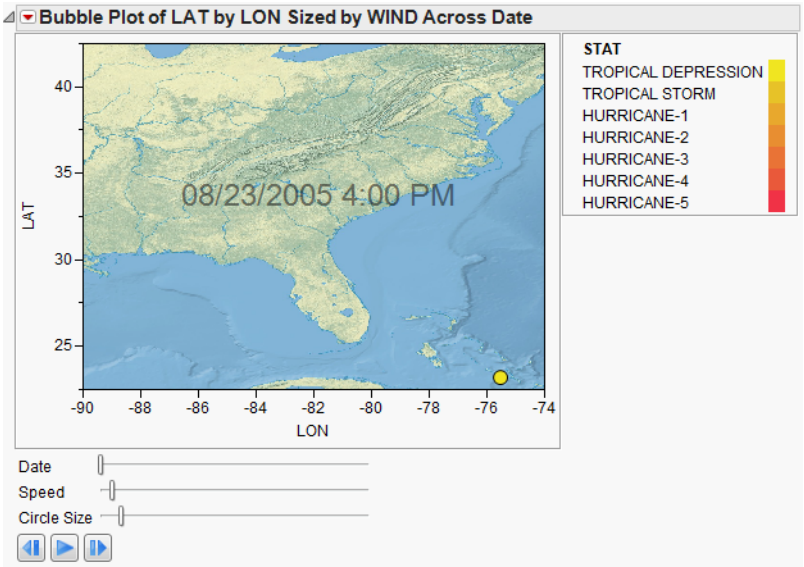
Figure 12.29 Bubble Plot of Katrina Data.jmp



Note that even though the location of the storm is plotted, it does not really tell us where it is. To add more context, add a map in the background.

9. Right-click the graph and select **Background Map**. The Set Background Map window appears.
10. Select **Detailed Earth** and click **OK**.

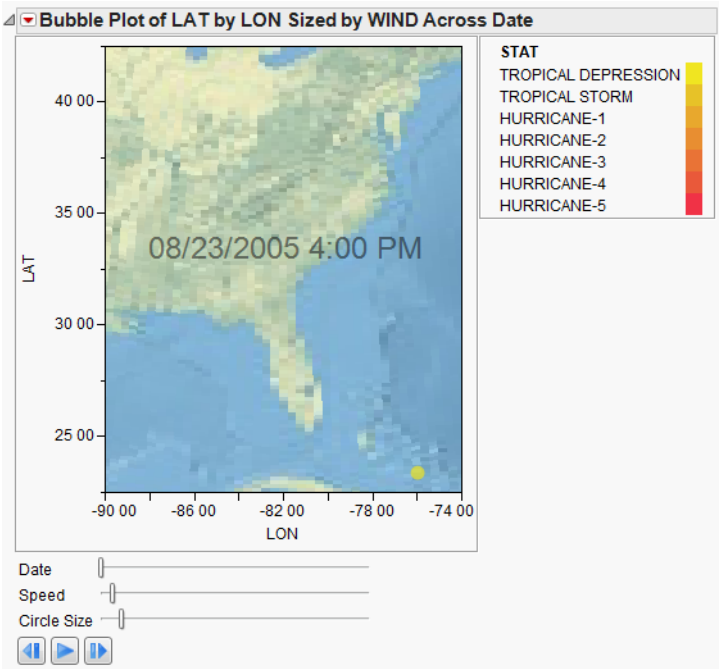
Figure 12.30 Bubble Plot of Katrina Data.jmp with Background Map



Now the coordinates make geographic sense. You can edit the axes and the size/scale to change how the graph appears.

11. Right-click the X axis (LON) and select **Axis Settings**. The X Axis Specification window appears.
12. Select **Scale** > **Geodesic US**.
13. Select **Format** > **Geographic** > **Longitude DMM**.
14. Click **OK**.
15. Repeat the same for the Y axis (LAT) except select **Format** > **Geographic** > **Latitude DMM**.
16. Right-click the map and select **Size/Scale** > **Size to Isometric.**

Figure 12.31 Bubble Plot of Katrina Data.jmp with Background Map



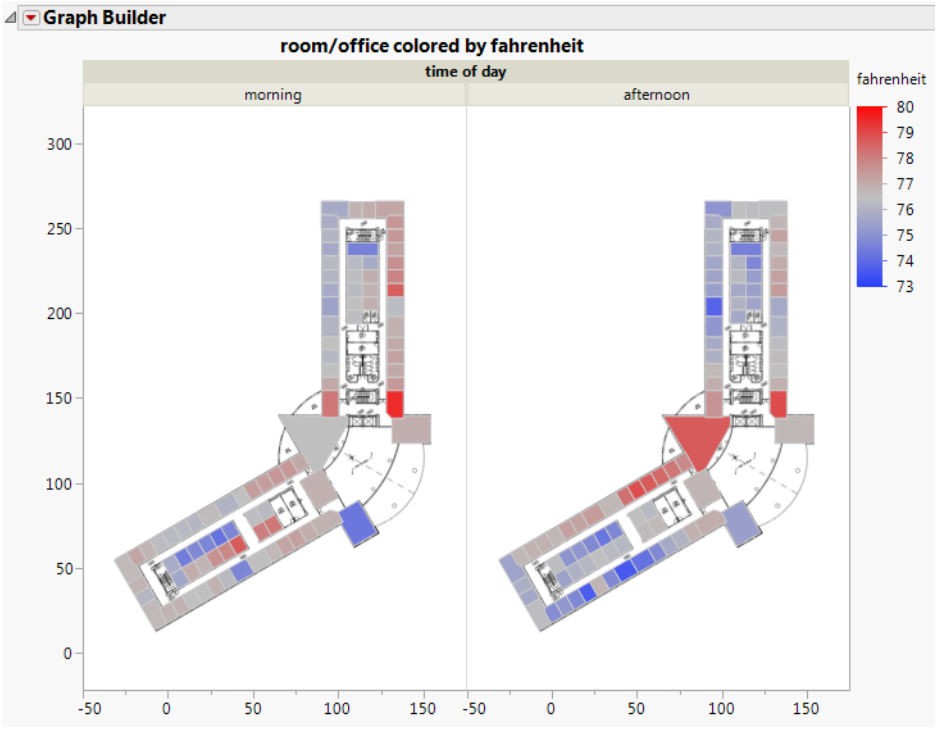
Click **Run** to view the animation of the hurricane data moving over the background map. You can manipulate the speed and the bubble size. Experiment with different options and view the displays. Adjust the axes or use the zoom tool to change what part of the area you are viewing. Add boundaries to the states. The map adjusts as the view does.

Office Temperature Study

This example demonstrates the creation of a custom background map for an office temperature study and how JMP was used to visualize the results. Data was collected concerning office temperatures for a floor within a building. A map was created for the floor using the Custom Map Creator add-in from the JMP File Exchange (https://community.jmp.com/docs/DOC-6218). Using Graph Builder, the office temperature results were then analyzed visually.

The map shown below is the floor, grouped by *time of day*. The color reflects the Fahrenheit value. Exploring data visually in this way can give hints as to what factors are affecting office temperature. Looking at this map, it appears the offices on the east side of the building are warmer in the mornings than they are in the afternoons. On the western side of the building, the opposite appears to be true. From this visualization, we might expect that both of these variables are affecting office temperatures, or perhaps that the interaction between these terms is significant. Such visuals help guide decision-making during the analysis.

Figure 12.32 Room/Office Colored by Fahrenheit and Grouped by Time of Day



First, data was collected and input into a data table (S4 Temps.jmp). Note the Room/Office column. It contains the unique names for each office and was assigned the Map Role to correctly define the map files.

Figure 12.33 Partial View of the S4 Temps.jmp Data Table

	day	time of day	thermometer	sector	room/office
1	1	morning	T4	sector 1	S4068
2	1	morning	T4	sector 1	S4069
3	1	morning	T4	sector 1	S4070
4	1	morning	T4	sector 1	S4072
5	1	morning	T4	sector 1	S4074
6	1	morning	T4	sector 1	S4076
7	1	morning	T4	sector 1	S4078
8	1	morning	T4	sector 1	S4082
9	1	morning	T4	sector 1	S4084
10	1	morning	T4	sector 1	S4086
11	1	morning	T4	sector 1	S4088
12	1	afternoon	T2	sector 1	S4068
13	1	afternoon	T2	sector 1	S4069

Then, a map of the floor was created using the Custom Map Creator add-in, which you can download from the JMP File Exchange at https://community.jmp.com/docs/DOC-6218. The add-in creates two tables to define the shapes; an XY table and a Name table. The instructions below describe how it was built.

Create a Map of the Floor

1. Launch the add-in through the menu items **Add-Ins > Map Shapes > Custom Map Creator**. Two tables open in the background followed by the Custom Map Creator Window.
2. Drag a background image into the graph frame. An image of the floor plan was available.
3. Perform any resizing on the background image and graph the frame.
4. Name the table (for example, S4).
5. Click **Next**.
6. Name the shape that you are about to define. For this example, each office was individually named for the map (for example, S4001).
7. Within the graph frame, use your mouse to click all of the boundaries of the shape that you want to define. A line appears that connects all of the boundary points.
8. As soon as you finish defining the boundaries of the shape, click **Next Shape**. Continue adding shapes until you have completed the floor plan. Note that you do not need to connect the final boundary point; the add-in automatically does that for you when you click **Next Shape**.
9. The line size and color can be changed. In addition, checking **Fill Shapes** fills each shape with a random color.
10. Click **Finish**.

The custom map files were created and named appropriately. The map files are S4-Name.jmp and S4-XY.jmp and have been saved in the JMP Samples/Data folder.

Add the Map Role Column Property

Note: Pathnames in this section refer to the JMP folder. On Windows, in JMP Pro, the JMP folder is named JMPPro.

The S4 Temps.jmp data table contains office data over a three-day period. Set up the **Map Role** column property in the data table:

1. Select **Help > Sample Data Library** and open S4 Temps.jmp.
2. Right-click the **Room/Office** column and select **Column Info**.
3. Select **Column Properties > Map Role**.
4. Select **Shape Name Use**.

5. Click the icon next to **Map name data table** and browse to the S4-Name.jmp file (located in the JMP Samples/Data folder).

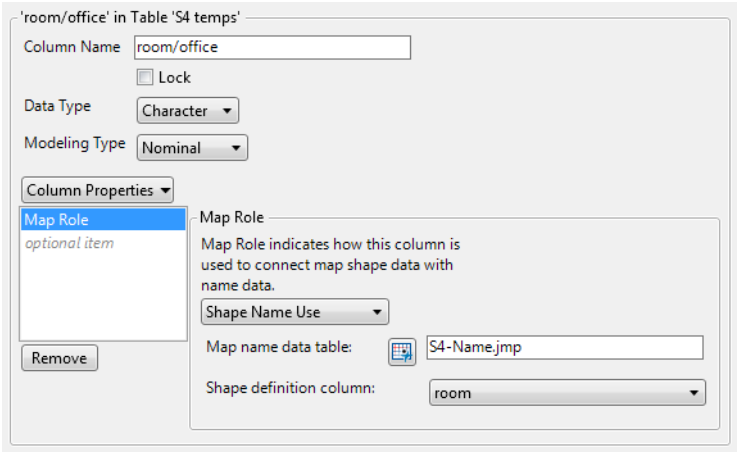
   This tells JMP where the data tables containing the map information reside.

6. Select room from the **Shape definition column** list.

   Room is the specific column within the S4-Name.jmp data table that contains the unique names for each office. Notice that the room column has the **Shape Name Definition Map Role** property assigned, as part of correctly defining the map files.

   Note: Remember, the Shape ID column in the -Name data table maps to the Shape ID column in the -XY data table. This means that indicating where the -Name data table resides links it to the -XY data table, so that JMP has everything that it needs to create the map.

Figure 12.34 Map Role Column Property



7. Click **OK**.

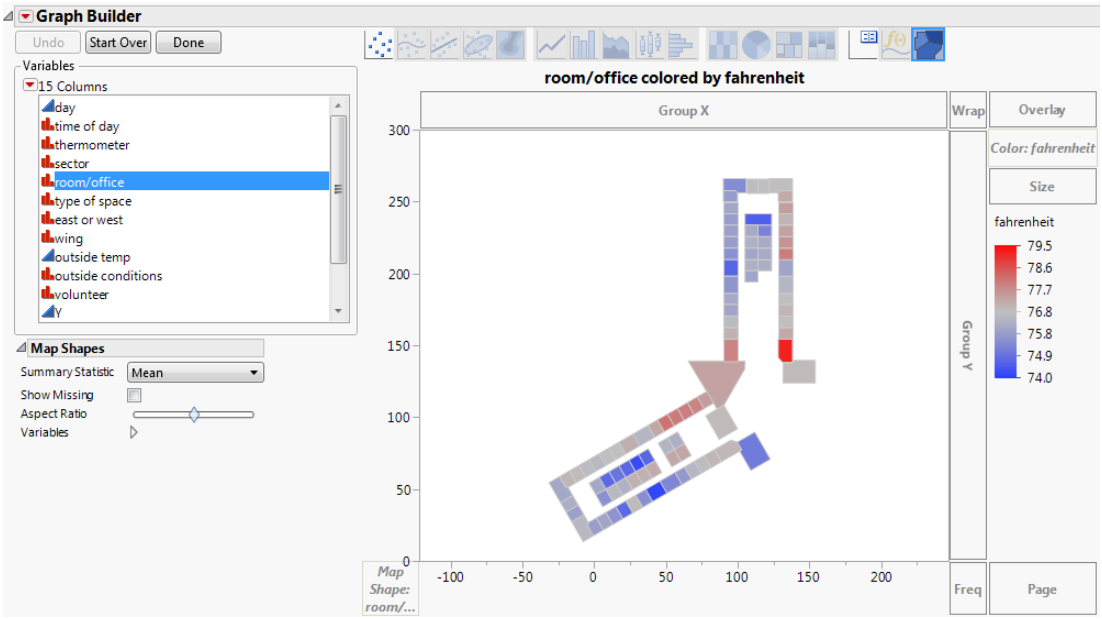
Once the **Map Role** column property has been set up, you can perform your analysis. You want to visually see the differences in office temperatures throughout the floor.

1. Select **Graph > Graph Builder**.
2. Drag and drop room/office into the **Map Shape** zone.

   Since you have defined the **Map Role** column property on this column, the map appears.

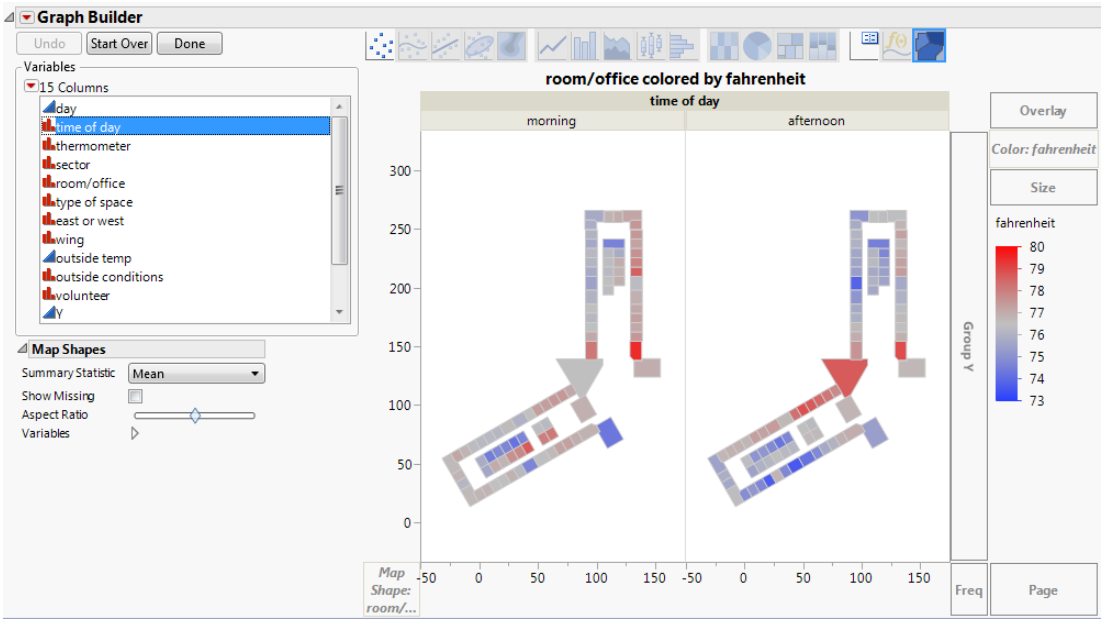
3. Drag and drop Fahrenheit to the **Color** zone.

Figure 12.35 Room/Office Colored by Fahrenheit



4. Drag and drop Time of Day onto Group X zone.

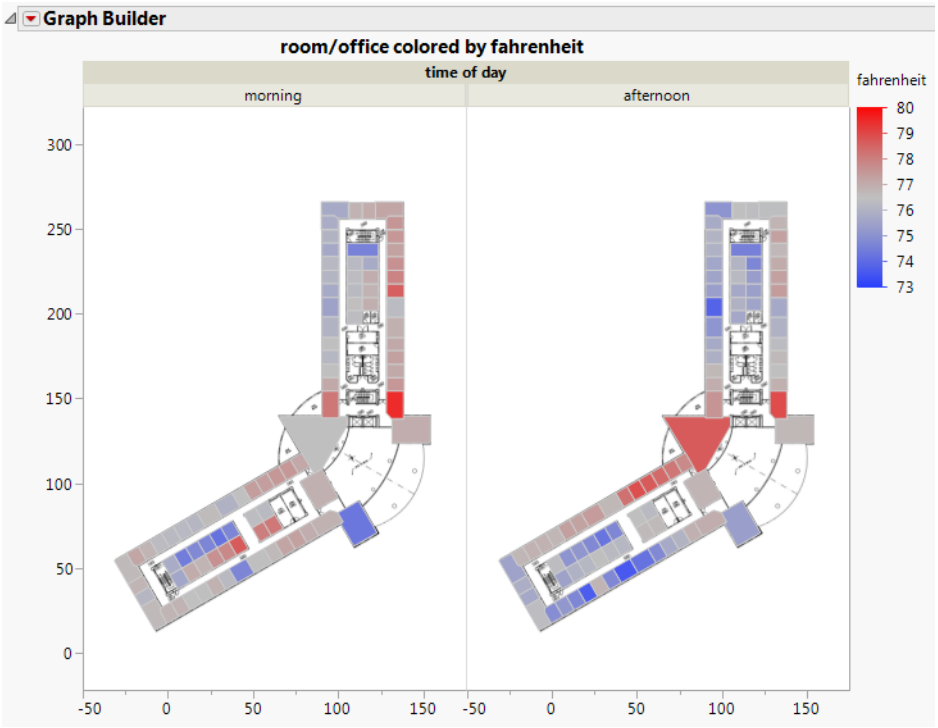
Figure 12.36 Room/Office Colored by Fahrenheit and Grouped by Time of Day



Note that only the offices that were part of the study and were created using the Custom Map Creator add-in are displayed. To add the entire floor plan image, the original floor plan graphic was dragged and dropped onto the Graph Builder window to create Figure 12.37.

To view Figure 12.37, select **Help > Sample Data Library** and open S4 Temps.jmp and run the **by Time of Day** script.

Figure 12.37 Room/Office Map with Original Floor Plan

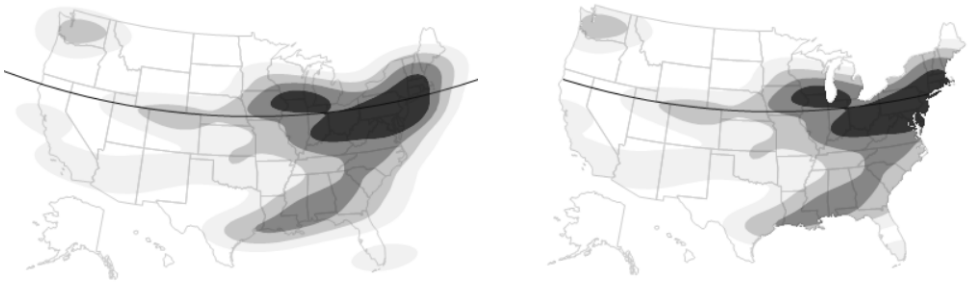


There are several scripts provided with the data table that you can run to view the various analysis and modeling that can be performed and visually displayed.

Boundary Map with Clipped Elements

Graphical elements (display segments such as reference, grid lines, and contours) can be clipped to conform to the boundaries of a geographical map. The Customize Graph window provides Clip Shape options, where you can select a boundary around which to clip the shape.

Figure 12.38 Unclipped (Left) and Clipped (Right) Reference Lines and Contours

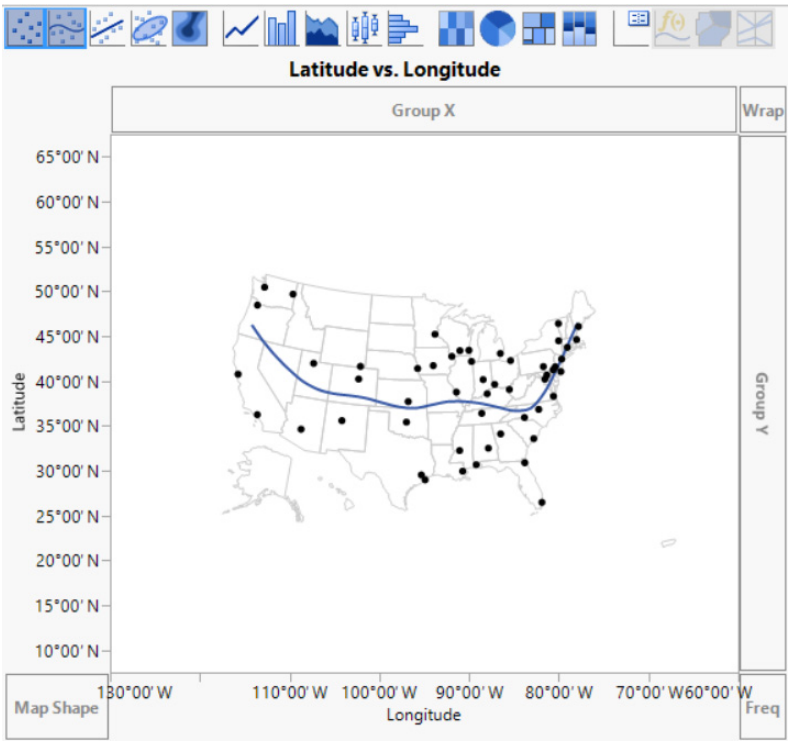


Create the Background Map

1. Select **Help > Sample Data Library** and open Cities.jmp.
2. Select **Graph > Graph Builder**.
3. Drag the Latitude column to the Y axis.
4. Drag the Longitude column to the X axis.
5. Right-click the graph and select **Graph > Background Map**.
6. In the Boundaries column, select **US States** and click **OK**.

   A map of the United States appears on the graph.

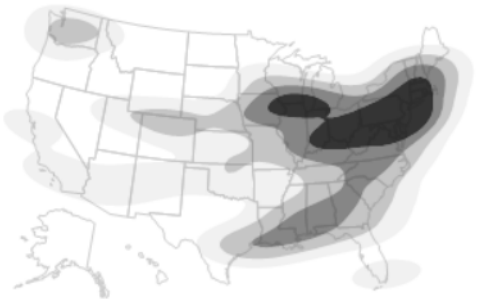
Figure 12.39 Background Map



7. Click the Points icon to remove the points.
8. Click the Smoother icon to remove the smoother.
9. Click the Contour icon .

   A contour plot appears on the map.

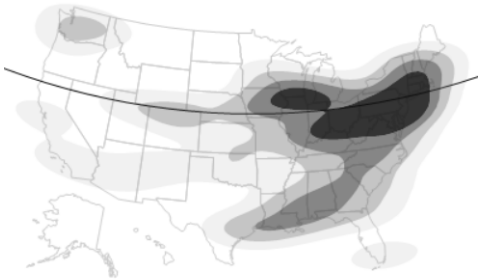
Figure 12.40 Unclipped Contours on Background Map



Add a Reference Line

1. On the Latitude axis, double-click **45°00’ N** to open the Y Axis Settings window.
2. In the Reference Lines pane, change the first number in the Value box to “40”.
   For example, in 44°40.11′ N, change “44” to “40”.
3. Click **Add** and then **OK**.

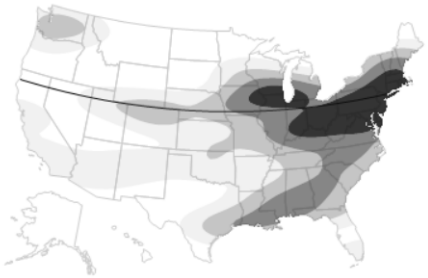
Figure 12.41 Unclipped Reference Line



Assign a Clip Shape

1. Right-click the graph and select **Customize**.
2. Select **Reference Lines** and select **US States** from the Clip Shape list.
3. Select **Contour**, select **US States** from the Clip Shape list, and click **OK**.

Figure 12.42 Clipped Reference Line and Contours



Tips:

- The boundaries that appear in the Background Map window are installed in the JMP Maps installation folder. See “Custom Map Files” for more information about adding your own boundaries.
- Scripting provides additional options, such as specifying the clipping path in a matrix or string. See the *Scripting Guide*.

Appendix A

Legacy Platforms

The following legacy platforms are deprecated. A deprecated platform is an older platform that contains features that can be accessed in a newer platform. The charts and plots in this appendix are all available in the Graph Builder platform. These legacy platforms might not be available in future releases of JMP.

Table A.1 Legacy Platforms

Platform	Graph Builder References
Treemap	• "Treemap"
Chart	• Bar chart examples in "Graph Builder Examples" • "Bar", "Line", or "Use a Heatmap to Apply Background Colors"
Overlay Plot	• "Example of an Overlaid Histogram and Ridgeline Chart" • "Graph Builder"

Treemaps

Treemaps are useful for observing patterns among groups that have many levels. Use treemaps when your data contains many categories, to visualize many groups.

Tip: To create a treemap, consider using Graph Builder. See "Treemap".

Launch the Treemap platform by selecting **Graph > Legacy > Treemap**.

When you click **OK** in the launch window, a treemap in Graph Builder is created by default. To produce a report in the legacy Treemap platform, you must first deselect the Use Graph Builder platform preference under Files > Preferences > Platforms > Treemap. For documentation on the legacy Treemap platform, see the Legacy Platforms section of the JMP 15.2 Help.

Charts

The Chart platform on the **Graph** menu charts continuous variables versus categorical variables. The continuous variables are summarized for each categorical level.

If you want to make a plot of individual data points (rather than summaries of data points), we recommend using an overlay plot instead.

Tip: To create a chart, consider using Graph Builder. See "Example of an Overlaid Histogram and Ridgeline Chart".

Launch the Chart platform by selecting **Graph > Legacy > Chart**.

When you click **OK** in the launch window, a chart in Graph Builder is created by default. To produce a report in the legacy Chart platform, you must first deselect the Use Graph Builder platform preference under Files > Preferences > Platforms > Chart. For documentation on the legacy Chart platform, see the Legacy Platforms section of the JMP 15.2 Help.

Overlay Plots

The Overlay Plot option produces plots of a single X column and one or more numeric Ys and does not accept non-numeric values for the *y*-axis.

Tip: To create an overlay plot, consider using the Overlay zone in Graph Builder. See "Example of an Overlaid Histogram and Ridgeline Chart".

Launch Overlay Plot by selecting **Graph > Legacy > Overlay Plot**.

When you click **OK** in the launch window, an overlay plot in Graph Builder is created by default. To produce a report in the legacy Overlay Plot platform, you must first deselect the Use Graph Builder platform preference under Files > Preferences > Platforms > Overlay Plot. For documentation on the legacy Overlay Plot platform, see the Legacy Platforms section of the JMP 15.2 Help.

Appendix B

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman & Hall.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: John Wiley & Sons.

Cornell, J. A. (1990). *Experiments with Mixtures*. 2nd ed. New York: John Wiley & Sons.

Few, Stephen (2008). "Dual-Scaled Axes in Graphs." *Visual Business Intelligence Newsletter*. http://www.perceptualedge.com/articles/visual_business_intelligence/dual-scaled_axes.pdf.

Hyndman, R. J. (1996). "Computing and Graphing High Density Regions."*The American Statistician* 50: 120 - 126.

Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer-Verlag.

Inselberg, A. (1985). "The Plane with Parallel Coordinates." *Visual Computing* 1: 69–91.

Mardia, K. V., Kent, J. T., and Bibby J. M. (1979). *Multivariate Analysis*. New York: Academic Press.

Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). "The Bagplot: A Bivariate Boxplot." *The American Statistician* 53: 382 - 387.

Wegman, E. J. (1990). "Hyperdimensional Data Analysis using Parallel Coordinates." *Journal of the American Statistical Association* 85: 664–675.

Appendix C

Technology License Notices

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

  - You must give any other recipients of the Work or Derivative Works a copy of this License; and
  - You must cause any modified files to carry prominent notices stating that You changed the files; and
  - You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
  - If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
  - You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

Version 16

Profilers

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 16 *Profilers*. Cary, NC: SAS Institute Inc.

JMP® 16 Profilers

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

August 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Get the Most from JMP

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

https://www.jmp.com/getstarted

Contents

Profilers

1 Learn about JMP 11
Documentation and Additional Resources
Formatting Conventions in JMP Documentation 13
JMP Help 14
JMP Documentation Library 14
Additional Resources for Learning JMP 20
JMP Tutorials 20
Sample Data Tables 20
Learn about Statistical and JSL Terms 21
Learn JMP Tips and Tricks 21
JMP Tooltips 21
JMP User Community 22
Free Online Statistical Thinking Course 22
JMP New User Welcome Kit 22
Statistics Knowledge Portal 22
JMP Training 22
JMP Books by Users 23
The JMP Starter Window 23
JMP Technical Support 23

2 Introduction to Profilers 25
Visualize Response Surfaces and Optimize Processes
Overview of the Profiler Platform 27
Introduction to Profiling 28
Profiler Features in JMP 28
Profiler Launch Windows 30
Fit Group 31
Interpret the Profiles 32
Profiler Platform Options 35
Common Profiler Topics 36
Linear Constraints 36
Noise Factors 38

3 Profiler ... 41
Explore Cross Sections of Responses across Each Factor
Example of the Prediction Profiler ... 43
Launch the Prediction Profiler Platform ... 44
Prediction Profiler Options ... 44
Maximization Options Window ... 52
Desirability Profiling and Optimization ... 53
Construction of Desirability Functions ... 54
How to Use the Desirability Function ... 55
The Desirability Profile ... 57
Customized Desirability Functions ... 57
Assess Variable Importance ... 59
Bagging ... 63
Additional Examples of the Prediction Profiler ... 64
Example of Desirability Profiling for Multiple Responses ... 65
Example of a Noise Factor in the Prediction Profiler ... 66
Example of Variable Importance for One Response ... 72
Example of Variable Importance for Multiple Responses ... 74
Example of Bagging to Improve Prediction ... 76
Example of Bagging to Indicate the Accuracy of Predictions ... 79
Example of Confidence Intervals Using Bagging ... 82
Example of Confidence Intervals Using StdErr Pred Formula Column ... 84
Statistical Details for the Prediction Profiler ... 85
Assess Variable Importance ... 85
Propagation of Error Bars ... 87
Extrapolation Control Metrics ... 88

4 Contour Profiler ... 91
Explore Contours of Responses across Two Factors
Overview of the Contour Profiler ... 93
Example of the Contour Profiler ... 93
Launch the Contour Profiler Platform ... 95
The Contour Profiler Report ... 95
Factor Settings and Controls ... 96
Response Settings and Controls ... 96
Contour Profiler Platform Options ... 97
Constraint Shading Settings ... 99
Additional Example of the Contour Profiler ... 99
Explore Optimal Settings ... 99

5 Surface Plot ... 103
Explore Contours of Responses across Three Factors
Overview of the Surface Plot and Surface Profiler ... 105
Example of the Surface Plot Platform ... 105
Launch the Surface Plot Platform ... 108
The Surface Plot Report ... 110
The Surface Plot ... 110
Appearance Controls ... 111
Independent Variables ... 112
Dependent Variables ... 113
Surface Plot Platform Options ... 115
Pop-Up Menu Options ... 115
Dependent Variables Options ... 116
Sheet or Surface Properties ... 117
Keyboard Shortcuts ... 119
Additional Examples of the Surface Plot Platform ... 119
Construct a Surface Plot for a Single Mathematical Function ... 119
Plotting Isosurfaces ... 120

6 Mixture Profiler ... 123
Explore Factor Effects Using Ternary Plots
Overview of the Mixture Profiler ... 125
Ternary Plot Overview ... 125
Ternary Plots with More Than Three Components ... 126
Example of the Mixture Profiler ... 127
Launch the Mixture Profiler Platform ... 128
The Mixture Profiler Report ... 129
Factor Settings and Controls ... 129
Response Settings and Controls ... 130
Mixture Profiler Platform Options ... 130
Customizations for the Mixture Profiler ... 131
Linear Constraints ... 131
Additional Examples of the Mixture Profiler Platform ... 132
Mixture Variables with a Process Variable ... 132
Multiple Responses with Five Mixture Variables ... 135

7 Custom Profiler ... 141
Explore Response Surfaces Using a Numerical Calculator
Overview of the Custom Profiler ... 143
Example of the Custom Profiler ... 143
Launch the Custom Profiler Platform ... 145

The Custom Profiler Report ... 146
Factor Settings and Controls ... 146
Response Settings and Controls ... 147
Optimization Settings and Controls ... 147
Custom Profiler Platform Options ... 148

8 Simulator ... 149
Explore the Effects of Variation on Responses

Overview of the Simulator ... 151
Example of the Simulator ... 151
Launch the Simulator ... 154
Factors Simulation Settings ... 154
Response Simulation Settings ... 157
The Simulator Report ... 157
The Simulator Report Options ... 158
Defect Profiler ... 159
Introduction to Tolerance Design ... 159
Graph Scale ... 160
Defect Rates ... 160
Simulation Method and Details ... 161
Notes ... 161
Defect Parametric Profiler ... 162
Simulation Experiment ... 163
Specification Limits ... 163
Additional Examples of the Simulator ... 163
Example of the Defect Profiler ... 164
Example of Stochastic Optimization ... 167
Example of Simulating General Formulas ... 176
Statistical Details for the Simulator ... 179
Normal Weighted Distribution ... 179

9 Excel Profiler ... 181
Visualize Models Saved in Microsoft Excel

Overview of the Excel Profiler ... 183
Example of an Excel Model ... 183
Run the JMP Profiler ... 185
Use Linear Constraints ... 186
Resolution of Profile Lines ... 186
Use the Excel Profiler from JMP ... 187

A References 189
B Technology License Notices 191

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation . . . 13
JMP Help . . . 14
JMP Documentation Library . . . 14
Additional Resources for Learning JMP . . . 20
JMP Tutorials . . . 20
Sample Data Tables . . . 20
Learn about Statistical and JSL Terms . . . 21
Learn JMP Tips and Tricks . . . 21
JMP Tooltips . . . 21
JMP User Community . . . 22
Free Online Statistical Thinking Course . . . 22
JMP New User Welcome Kit . . . 22
Statistics Knowledge Portal . . . 22
JMP Training . . . 22
JMP Books by Users . . . 23
The JMP Starter Window . . . 23
JMP Technical Support . . . 23

Formatting Conventions in JMP Documentation

These conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit https://www.jmp.com/software/pro.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from https://community.jmp.com.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Model Screening • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Legacy Control Charts • Pareto Plot • Diagram • Manage Spec Limits • OC Curves

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "JMP Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "JMP Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "JMP New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\16\Samples\Data

On macOS: \Library\Application Support\JMP\16\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit https://jmp.com/tools.

Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > Initial JMP Starter Window**.

JMP Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at https://www.jmp.com/support, including the technical support phone number.

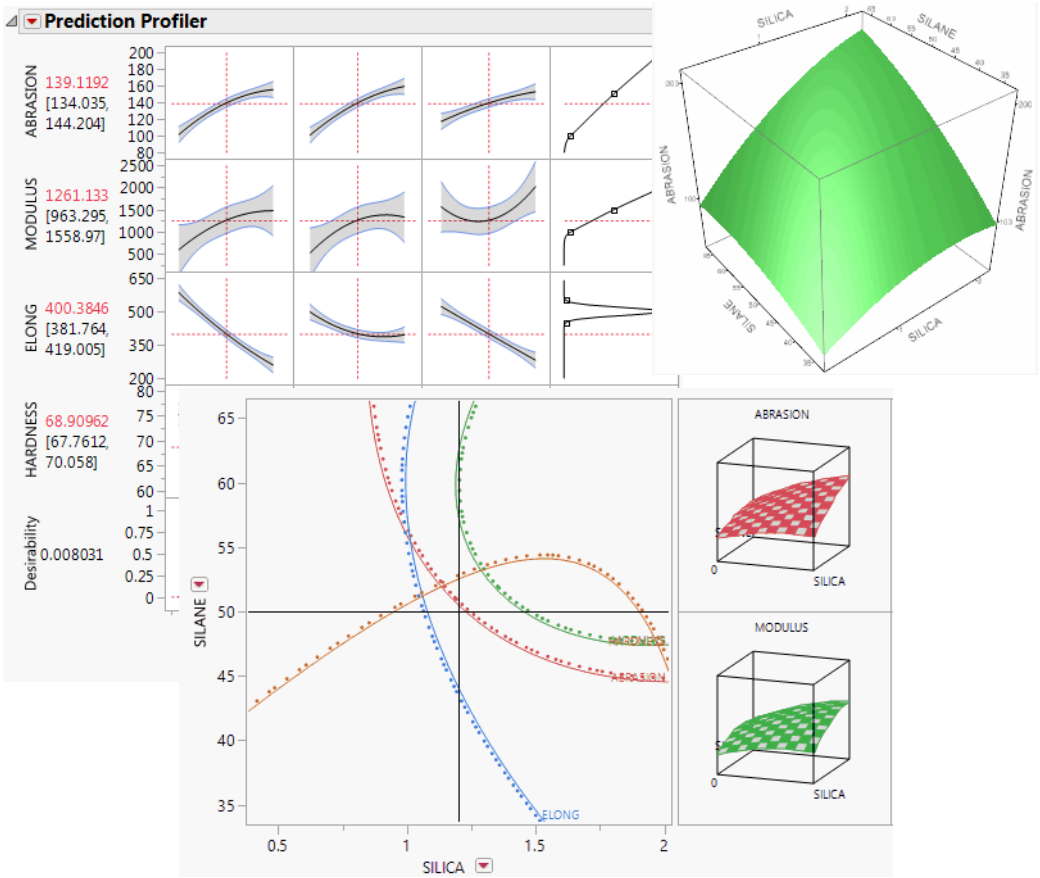
Chapter 2

Introduction to Profilers

Visualize Response Surfaces and Optimize Processes

Profiling is an approach to visualizing response surfaces by seeing what would happen if you change just one or two factors at a time. Essentially, a profile is a cross-section view. The interactive profilers in JMP promote exploring opportunity spaces. In fitting equations to data, the fitting is only the first step. Interpreting the fit, understanding the fitted response surface, and finding factor values to optimize the responses is desirable.

Figure 2.1 Examples of Profilers



Contents

Overview of the Profiler Platform 27
Introduction to Profiling 28
Profiler Features in JMP 28
Profiler Launch Windows 30
Fit Group 32
Interpret the Profiles 32
Profiler Platform Options 36
Common Profiler Topics 37
Linear Constraints 37
Noise Factors 39

Overview of the Profiler Platform

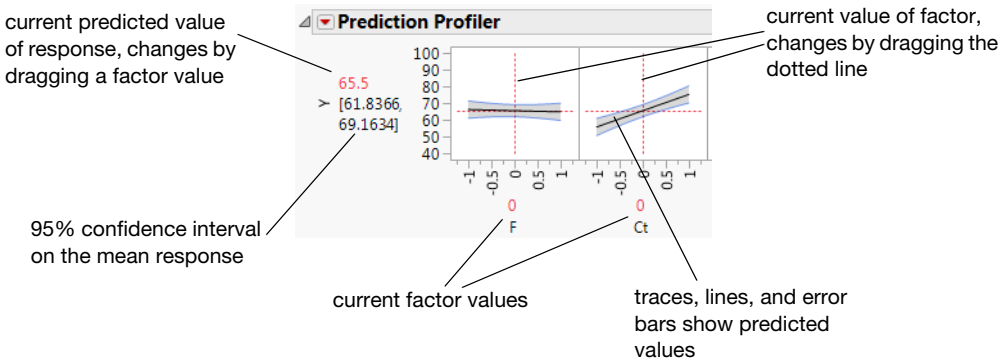
The Prediction Profiler displays profile traces (Figure 2.2) for each *X* variable. A *profile trace* is the predicted response as one variable is changed while the others are held constant at the current values. The Prediction Profiler recomputes the profiles and predicted responses (in real time) as you vary the value of an *X* variable.

- The vertical dotted line for each *X* variable shows its *current value* or *current setting*. If the variable is nominal, the X axis identifies categories.

  For each *X* variable, the value above the factor name is its current value. Change the current value by clicking in the graph or by dragging the dotted line where you want the new current value to be.
- The horizontal dotted line shows the *current predicted value* of each *Y* variable for the current values of the *X* variables.
- The black lines within the plots show how the predicted value changes when you change the current value of an individual *X* variable. In fitting platforms, the 95% confidence interval for the predicted values is shown by solid blue curves surrounding the prediction trace (for continuous variables) or the height of an error bar (for categorical variables). For continuous variables, the confidence interval region is shaded.

The Prediction Profiler is a way of changing one variable at a time and looking at the effect on the predicted response.

Figure 2.2 Illustration of Traces



The Prediction Profiler in some situations computes confidence intervals for each profiled column. If you have saved both a standard error formula and a prediction formula for the same column, the Prediction Profiler offers to use the standard errors to produce the confidence intervals rather than profiling them as a separate column.

Introduction to Profiling

It is easy to visualize a response surface with one input factor *X* and one output factor *Y*. It becomes harder as more factors and responses are added. The profilers in JMP provide a number of highly interactive cross-sectional views of any response surface. In this guide we use the following terms interchangeably.

- factor, input variable, *X* column, independent variable, setting, term
- response, output variable, *Y* column, dependent variable, outcome

Desirability profiling and optimization features are available to help find good factor settings and produce desirable responses. Most profilers also incorporate multithreading for faster computation. Simulation and defect profiling features are available for when you need to make responses that are robust and high-quality when the factors have variation.

- "Profiler Features in JMP"
- "Profiler Launch Windows"
- "Fit Group"

Profiler Features in JMP

There are several profiler facilities in JMP, accessible from a number of fitting platforms and the main menu under Graph. They are used to profile data column formulas.

Table 2.1 Profiler Features Summary

	Description	Features
Prediction Profiler	Shows vertical slices across each factor, holding other factors at current values	Desirability, Optimization, Simulator, Propagation of Error
Contour Profiler	Horizontal slices show contour lines for two factors at a time	Simulator

Table 2.1 Profiler Features Summary *(Continued)*

	Description	Features
Surface Profiler	3-D plots of responses for 2 factors at a time, or a contour surface plot for 3 factors at a time	Surface Visualization
Mixture Profiler	A contour profiler for mixture factors	Ternary Plot and Contours
Custom Profiler	A non-graphical profiler and numerical optimizer	General Optimization, Simulator
Excel Profiler	Visualize models (or formulas) stored in Excel worksheets.	Profile using Excel Models

Profiler availability is shown in Table 2.2. The Custom Profiler is available only through the Graph menu. (Model Comparison *does* have Custom Profiler available.)

Table 2.2 Where to Find JMP Profilers

Location	Profiler	Contour Profiler	Surface Profiler	Mixture Profiler
Graph Menu (as a Platform)	Yes	Yes	Yes	Yes
Fit Model: Least Squares	Yes	Yes	Yes	Yes
Fit Model: Generalized Regression	Yes			
Fit Model: Mixed Model	Yes	Yes	Yes	Yes
Fit Model: Logistic	Yes			
Fit Model: Loglinear Variance	Yes	Yes	Yes	
Fit Model: Generalized Linear Model	Yes	Yes	Yes	
Fit Model: Partial Least Squares	Yes			
Neural	Yes	Yes	Yes	
Model Comparison	Yes	Yes	Yes	Yes
Nonlinear: Factors and Response	Yes	Yes	Yes	
Nonlinear: Parameters and SSE	Yes	Yes	Yes	
Nonlinear: Fit Curve	Yes			

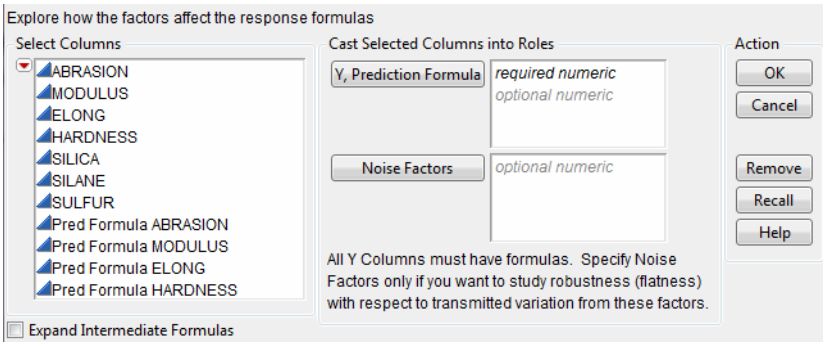
Table 2.2 Where to Find JMP Profilers *(Continued)*

Location	Profiler	Contour Profiler	Surface Profiler	Mixture Profiler
Gaussian Process	Yes	Yes	Yes	
Partial Least Squares	Yes			
Life Distribution	Yes			
Fit Life by X	Yes		Yes	
Recurrence Analysis	Yes			
Choice	Yes			
Custom Design Prediction Variance	Yes		Yes	

Profiler Launch Windows

When a profiler is invoked as a platform from the Graph menu, rather than through a fitting platform, you provide columns with formulas as the Y, Prediction Formula columns. These formulas could have been saved from the fitting platforms.

Figure 2.3 Profiler Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

The columns referenced in the formulas become the *X* columns (unless the column is also a *Y*).

Y, Prediction Formula The response columns containing formulas.

Note: You can also add columns that contain the formula and the values for the standard errors of the corresponding response columns. These columns are used to create confidence intervals. The formula columns must be named PredSE <colname>. After you click **OK**, specify whether you want to use PredSE <colname> to construct confidence intervals for Pred Formula <colname>. Otherwise, JMP creates a separate profiler plot for PredSE <colname>.

Noise Factors Used only in special cases for modeling derivatives. For more information about noise factors, see "Noise Factors".

Expand Intermediate Formulas Tells JMP that if an ingredient column to a formula is itself a formula column that refers to other columns, to substitute the original columns into the inner formula. To prevent an ingredient column from expanding, add an Other column property, name it Expand Formula, and assign a value of 0. See "Expand Intermediate Formulas".

The Surface Plot platform is discussed in "Surface Plot". The Surface Profiler is very similar to the Surface Plot platform, except Surface Plot has more modes of operation. Neither the Surface Plot platform nor the Surface Profiler have some of the capabilities common to other profilers.

Expand Intermediate Formulas

The Profiler launch window has an Expand Intermediate Formulas check box. When this box is checked, the formula that is being profiled is handled differently. If the profiled formula contains another formula column that references other columns, then the original columns are substituted into the inner formula. Therefore, the formula being profiled is profiled with respect to the original columns instead of the intermediate column references. Expand Intermediate Formula also expands formula columns that have the Vector modeling type.

For example, when Fit Model fits a logistic regression for two levels (A and B), the end formulas (Prob[A] and Prob[B]) are functions of the Lin[x] column, which itself is a function of another column x. If Expand Intermediate Formulas is selected, then when Prob[A] is profiled, it is with reference to x, not Lin[x].

In addition, using the Expand Intermediate Formulas check box enables the Save Expanded Formulas command in the platform red triangle menu. This creates a new column with a formula, which is the formula being profiled as a function of the end columns, not the intermediate columns.

Fit Group

For the REML and Stepwise personalities of the Fit Model platform, if models are fit to multiple *Y*'s, the results are combined into a Fit Group report. This enables the different *Y*'s to be profiled in the same Profiler. The Fit Group red triangle menu has options for launching the joint Profiler. Profilers for the individual *Y*'s can still be used in the respective Fit Model reports.

Fit Group reports are also created when a By variable is specified for a Stepwise analysis. This allows for the separate models to be profiled in the same Profiler.

The Fit Group scripting command can be used to fit models in different platforms, and profile the individual models in the Profiler. See the *Scripting Guide*.

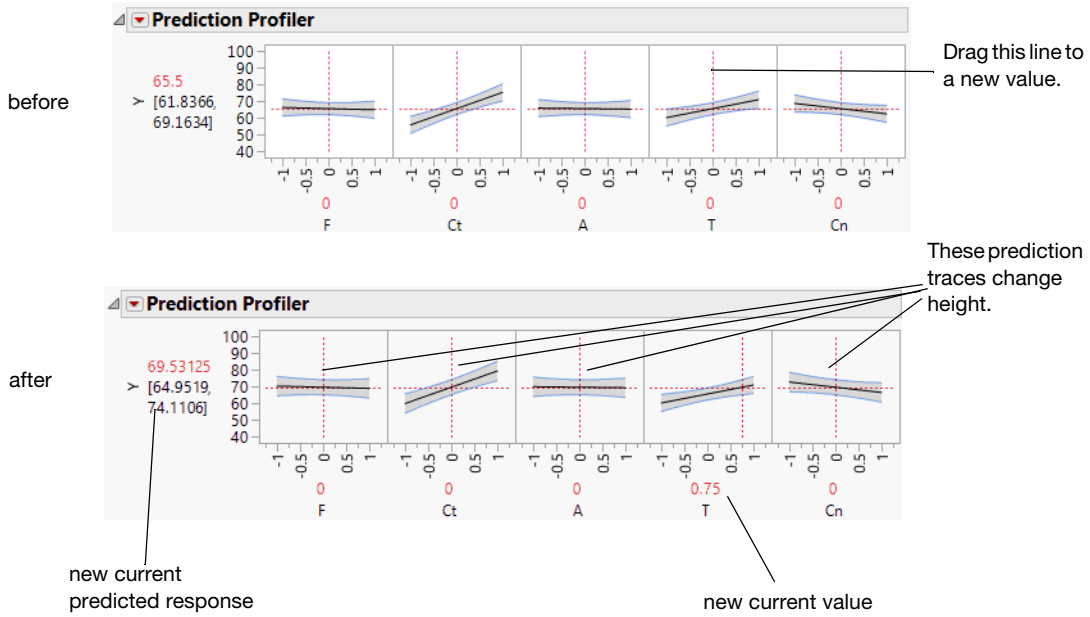
Interpret the Profiles

The illustration in Figure 2.4 describes how to use the components of the Prediction Profiler. There are several important points to note when interpreting a prediction profile:

- The importance of a factor can be assessed to some extent by the steepness of the prediction trace. If the model has curvature terms (such as squared terms), then the traces might be curved.
- If you change the value of a factor, the prediction trace for that factor is not affected, but the prediction traces of all the other factors can change. The *Y* response line must cross the intersection points of the prediction traces with their current value lines.

Note: If there are interaction effects or cross-product effects in the model, the prediction traces can shift their slope and curvature as you change current values of other terms. That is what interaction is all about. If there are no interaction effects, the traces change only in height, not slope or shape.

Figure 2.4 Changing One Factor from 0 to 0.75



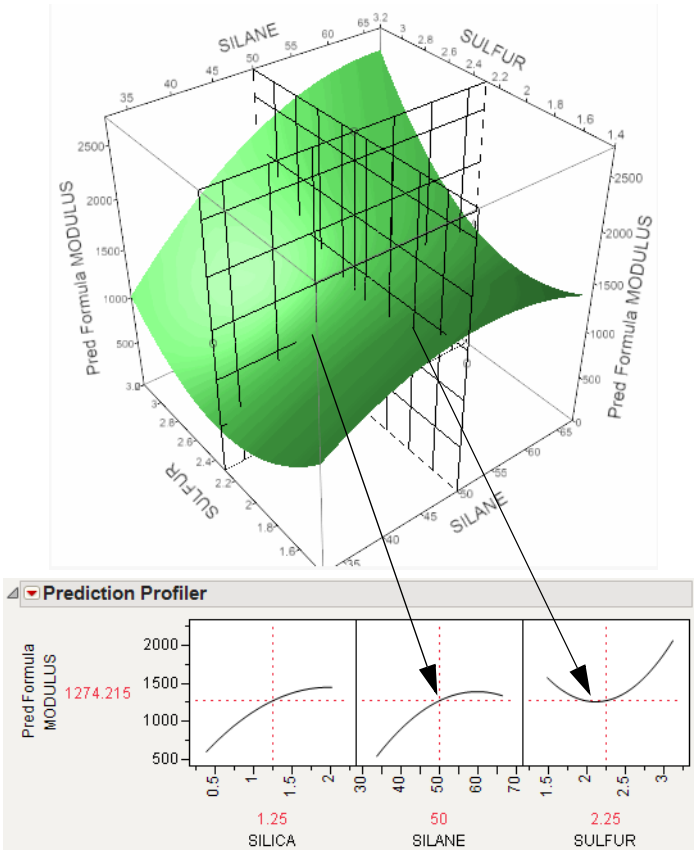
Prediction profiles are especially useful in multiple-response models to help judge which factor values can optimize a complex set of criteria.

Click a graph or drag the current value line right or left to change the factor's current value. The response values change as shown by a horizontal reference line in the body of the graph. Double-click in an axis to bring up a window that changes its settings.

The Profiler as a Cross-Section

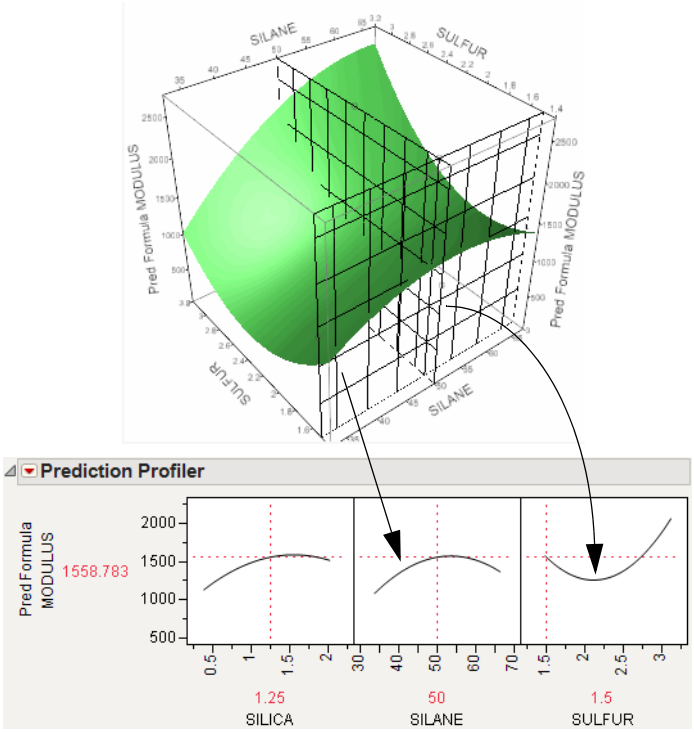
In the following example using Tiretread.jmp, look at the response surface of the expression for MODULUS as a function of SULFUR and SILANE (holding SILICA constant). Now look at how a grid that cuts across SILANE at the SULFUR value of 2.25. Note how the slice intersects the surface. If you transfer that down below, it becomes the profile for SILANE. Similarly, note the grid across SULFUR at the SILANE value of 50. The intersection when transferred down to the SULFUR graph becomes the profile for SULFUR.

Figure 2.5 Profiler as a Cross-Section



Now consider changing the current value of SULFUR from 2.25 to 1.5.

Figure 2.6 Profiler as a Cross-Section

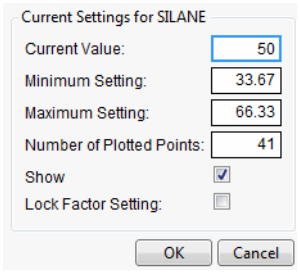


In the Prediction Profiler, note the new value just moves along the same curve for SULFUR, the SULFUR curve itself does not change. But the profile for SILANE is now taken at a different cut for SULFUR. The profile for SILANE is also a little higher and reaches its peak in the different place, closer to the current SILANE value of 50.

Set or Lock Factor Values

If you press Alt (Option on macOS) and click in a graph, a window prompts you to enter specific settings for the factor.

Figure 2.7 Continuous Factor Settings Window



For continuous variables, you can specify the following:

Current Value The value used to calculate displayed values in the profiler, equivalent to the red vertical line in the graph.

Minimum Setting The minimum value of the factor's axis.

Maximum Value The maximum value of the factor's axis.

Number of Plotted Points Specifies the number of points used in plotting the factor's prediction traces.

Show Show or hide the factor in the profiler.

Lock Factor Setting Locks the value of the factor at its current setting.

Profiler Platform Options

The Profiler red triangle menu contains the following options:

Profiler Shows or hides the Prediction Profiler.

Contour Profiler Shows or hides the Contour Profiler.

Custom Profiler Shows or hides the Custom Profiler.

Surface Profiler Shows or hides the Surface Profiler.

Mixture Profiler Shows or hides the Mixture Profiler.

Show Formulas Opens a JSL window showing all formulas being profiled.

Formulas for OPTMODEL Creates code for the OPTMODEL SAS procedure. Press Ctrl+Shift and then select **Formulas for OPTMODEL** from the red triangle menu.

The following options are available in many platforms. See *Using JMP*.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Common Profiler Topics

- "Linear Constraints"
- "Noise Factors"

Linear Constraints

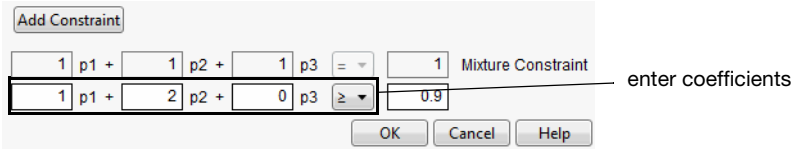
The Prediction Profiler, Custom Profiler, and Mixture Profiler can incorporate linear constraints into their operations. Linear constraints can be entered in two ways, described in the following sections.

Red Triangle Menu Item

To enter linear constraints via the red triangle menu, select **Alter Linear Constraints** from either the Prediction Profiler or Custom Profiler red triangle menu.

Choose **Add Constraint** from the resulting window, and enter the coefficients into the appropriate boxes. For example, to enter the constraint p1 + 2*p2 ≥ 0.9, enter the coefficients as shown in Figure 2.8. As shown, if you are profiling factors from a mixture design, the mixture constraint is present by default and cannot be modified.

Figure 2.8 Enter Coefficients



After you click **OK**, the Profiler updates the profile traces, and the constraint is incorporated into subsequent analyses and optimizations.

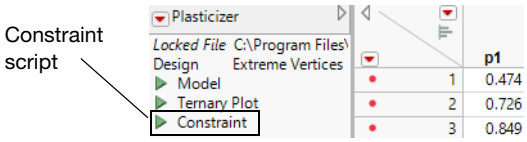
If you attempt to add a constraint for which there is no feasible solution, a message is written to the log and the constraint is not added. To delete a constraint, enter zeros for all the coefficients.

Constraints added in one profiler are not accessible by other profilers until saved. For example, if constraints are added under the Prediction Profiler, they are not accessible to the Custom Profiler. To use the constraint, you can either add it under the Custom Profiler red triangle menu, or use the **Save Linear Constraints** command described in the next section.

Constraint Table Property/Script

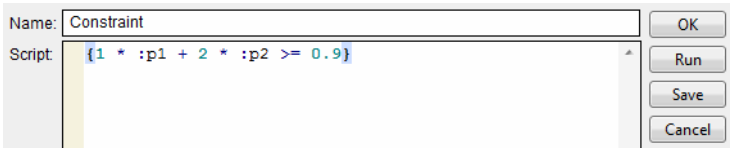
If you add constraints in one profiler and want to make them accessible by other profilers, use the **Save Linear Constraints** command, accessible through the platform red triangle menu. For example, if you created constraints in the Prediction Profiler, choose **Save Linear Constraints** under the Prediction Profiler red triangle menu. The Save Linear Constraints command creates or alters a Table Script called Constraint. An example of the Table Property is shown in Figure 2.9.

Figure 2.9 Constraint Table Script



The Constraint Table Property is a list of the constraints, and is editable. It is accessible to other profilers, and negates the need to enter the constraints in other profilers. To view or edit Constraint, right-click the green triangle and select **Edit**. The content of the constraint from Figure 2.8 is shown below in Figure 2.10.

Figure 2.10 Example Constraint



The Constraint Table Script can be created manually by choosing **New Script** from the red triangle menu beside a table name.

Note: When creating the Constraint Table Script manually, the spelling must be exactly "Constraint". Also, the constraint variables are case sensitive and must match the column name. For example, in Figure 2.10, the constraint variables are p1 and p2, not P1 and P2.

The Constraint Table Script is also created when specifying linear constraints when designing an experiment.

The Alter Linear Constraints and Save Linear Constraints commands are not available in the Mixture Profiler. To incorporate linear constraints into the operations of the Mixture Profiler, the Constraint Table Script must be created by one of the methods discussed in this section.

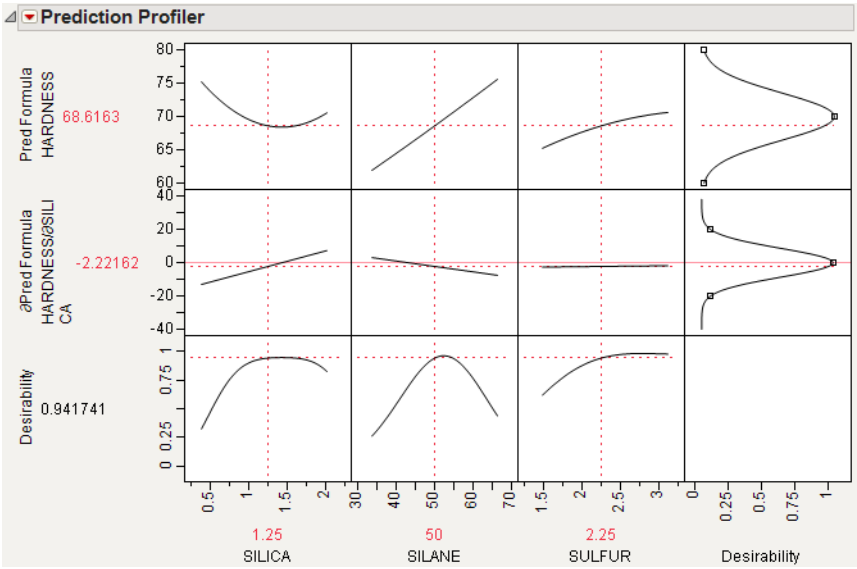
Noise Factors

Note: Noise factor optimization is also available in the Prediction Profiler, Contour Profiler, Custom Profiler, and Mixture Profiler.

Noise factors (robust process engineering) enables you to produce acceptable products reliably, despite variation in the process variables. Even when your experiment has controllable factors, there is a certain amount of uncontrollable variation in the factors that affects the response. This is called transmitted variation. Factors with this variation are called noise factors. Some factors you cannot control at all, like environmental noise factors. The mean for some factors can be controlled, but not their standard deviation. This is often the case for intermediate factors that are output from a different process or manufacturing step.

A good approach to making the process robust is to match the target at the flattest place of the noise response surface. Then, the noise has little influence on the process. Mathematically, this is the value where the first derivatives of each response with respect to each noise factor are zero. JMP computes the derivatives for you.

Figure 2.11 Noise Factor Example



To analyze a model with noise factors:

1. Fit the appropriate model (for example, using the Fit Model platform).
2. Save the model to the data table with the **Save > Prediction Formula** command.
3. Launch the **Profiler** (from the **Graph** menu).

4. Assign the prediction formula to the **Y, Prediction Formula** role and the noise factors to the **Noise Factor** role.
5. Click **OK**.

   The resulting profiler shows response functions and their appropriate derivatives with respect to the noise factors. The derivatives are set to have maximum desirability at zero.
6. Select **Optimization and Desirability > Maximize Desirability** from the **Profiler** menu.

   This finds the best settings of the factors, balanced with respect to minimizing transmitted variation from the noise factors.

Chapter 3

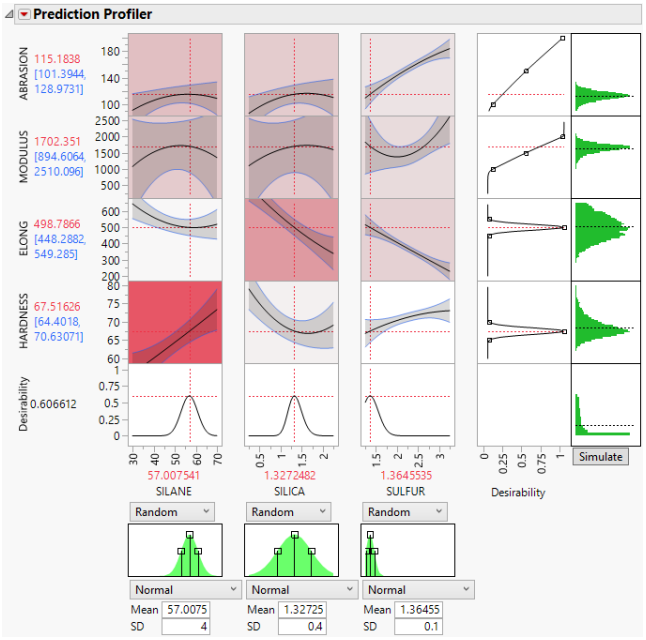
Profiler

Explore Cross Sections of Responses across Each Factor

The Prediction Profiler gives you a wealth of information about your model. Use the Prediction Profiler to do the following:

- See how your prediction model changes as you change settings of individual factors.
- Set desirability goals for your response or responses, and find optimal settings for your factors.
- Gauge your model's sensitivity to changes in the factors, where sensitivity is based on your predictive model.
- Assess the importance of your factors relative to model predictions, in a way that is independent of the model.
- Simulate your response distribution based on specified distributions for both factors and responses, and control various aspects of the appearance of the profiler.

Figure 3.1 Prediction Profiler for Four Responses with Simulator and Importance Coloring



Contents

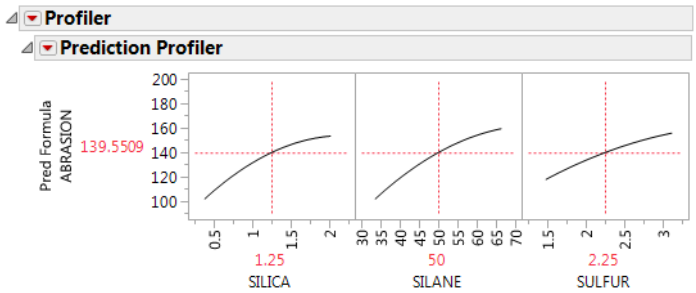
Example of the Prediction Profiler 43
Launch the Prediction Profiler Platform 44
Prediction Profiler Options 44
Maximization Options Window 52
Desirability Profiling and Optimization 53
Construction of Desirability Functions 54
How to Use the Desirability Function 55
The Desirability Profile 57
Customized Desirability Functions 57
Assess Variable Importance 59
Bagging 63
Additional Examples of the Prediction Profiler 64
Example of Desirability Profiling for Multiple Responses 65
Example of a Noise Factor in the Prediction Profiler 66
Example of Variable Importance for One Response 72
Example of Variable Importance for Multiple Responses 74
Example of Bagging to Improve Prediction 76
Example of Bagging to Indicate the Accuracy of Predictions 79
Example of Confidence Intervals Using Bagging 82
Example of Confidence Intervals Using StdErr Pred Formula Column 84
Statistical Details for the Prediction Profiler 85
Assess Variable Importance 85
Propagation of Error Bars 87
Extrapolation Control Metrics 88

Example of the Prediction Profiler

This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). The goal is to find the optimal combination of the three factors in the development of a tire tread compound. See Derringer and Suich (1980).

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Graph > Profiler**.
3. Select Pred Formula Abrasion and click **Y, Prediction Formula**.
4. Click **OK**.

Figure 3.2 Prediction Profiler Report Window

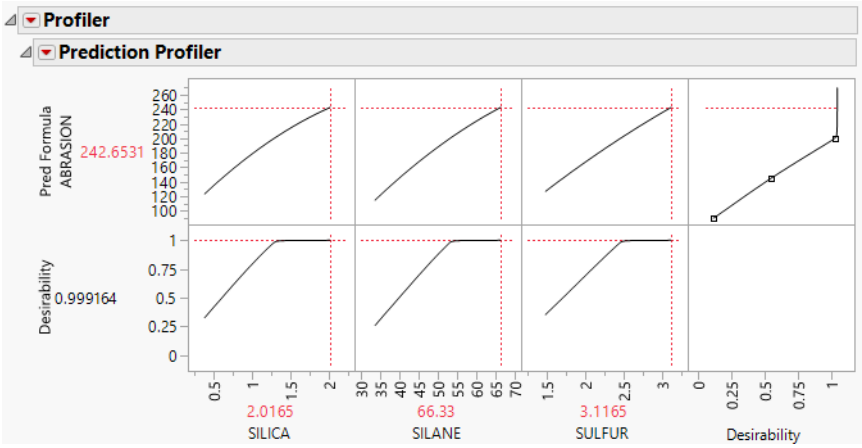


The profiler is interactive. The vertical red lines correspond to the current value of the factors. The current value of each factor is also shown in red below the horizontal axis. The red value on the vertical axis is the predicted response based on the current values of the factors. Click and drag the vertical red lines to change the current values of the factors. Click a factor's red value to set the factor to a specific value. Alternatively, you can animate the profiler to cycle through a variety of factor values.

5. Click the Prediction Profiler red triangle and select **Factor Settings > Animation**.
6. Click the play button to sequentially cycle through the factor values, one factor at a time.
7. Click the Prediction Profiler red triangle and select **Optimize and Desirability > Desirability Functions**.
8. Click the Prediction Profiler red triangle and select **Optimize and Desirability > Maximize Desirability.**

Tip: You might need to adjust the vertical axis of the prediction profiler to view the maximum value of ABRASION in the plot.

Figure 3.3 Maximize Desirability Report



By maximizing the desirability, you see that to maximize ABRASION, all factors are set to their highest levels.

Launch the Prediction Profiler Platform

The Prediction Profiler can be accessed in the following ways:

- The Prediction Profiler can be accessed directly from the Graph menu. When you access the Prediction Profiler in this way, the Prediction Profiler launch window appears. See "Profiler Launch Windows".
- The Prediction Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information about the availability of the Prediction Profiler in different platforms.
- The Prediction Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu.
- JMP PRO The Prediction Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu.

Prediction Profiler Options

The red triangle menu on the Prediction Profiler title bar has the following options:

Optimization and Desirability Submenu that consists of the following options:

Desirability Functions Shows or hides the desirability functions. Desirability is discussed in "Desirability Profiling and Optimization".

Maximize Desirability Sets the current factor values to maximize the desirability functions. Takes into account the response importance weights.

Note: In many situations, the settings that optimize the desirability function are not unique. The Maximize Desirability option gives one such setting. The Contour Profiler is a good tool for finding alternative factor combinations that optimize desirability. For an example, see "Explore Optimal Settings".

Note: If a factor has a Design Role column property value of Discrete Numeric, it is treated as continuous in the optimization of the desirability function. To account for the fact that the factor can assume only discrete levels, it is displayed in the profiler as a categorical term and an optimal allowable level is selected.

Maximize and Remember Maximizes the desirability functions and remembers the associated settings.

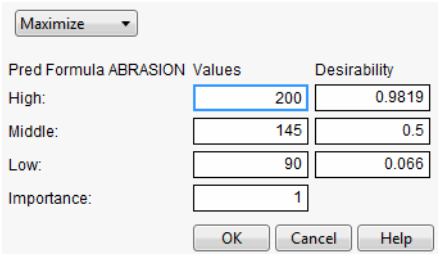
Maximization Options Opens the Maximization Options window where you can refine the optimization settings. See "Maximization Options Window".

Maximize for Each Grid Point Used only if one or more factors are locked. The ranges of the locked factors are divided into a grid, and the desirability is maximized at each grid point. This is useful if the model that you are profiling has categorical factors. Then the optimal condition can be found for each combination of the categorical factors.

Save Desirabilities Saves the three desirability function settings for each response, and the associated desirability values, as a Response Limits column property in the data table. These correspond to the coordinates of the handles in the desirability plots.

Set Desirabilities Opens the Response Goal window where you can set specific desirability values.

Figure 3.4 Response Goal Window



Save Desirability Formula Creates a column in the data table with a formula for Desirability. The formula uses the fitting formula when it can, or the response variables when it cannot access the fitting formula.

Assess Variable Importance Provides different approaches to calculating indices that measure the importance of factors to the model. These indices are independent of the model type and fitting method. See "Assess Variable Importance".

Save Bagged Predictions (Available only when the Prediction Profiler is embedded in select modeling platforms.) Launches the Bagging window. Bootstrap aggregating (bagging) enables you to create multiple training data sets by sampling with replacement from the original data. For each training set, a model is fit using the analysis platform, and predictions are made. The final prediction is a combination of the results from all of the models. This improves prediction performance by reducing the error from variance. See "Bagging".

Simulator Launches the Simulator. The Simulator enables you to create Monte Carlo simulations using random noise added to factors and predictions for the model. A typical use is to set fixed factors at their optimal settings, and uncontrolled factors and model noise to random values. You then find out the rate of responses outside the specification limits. See "Simulator".

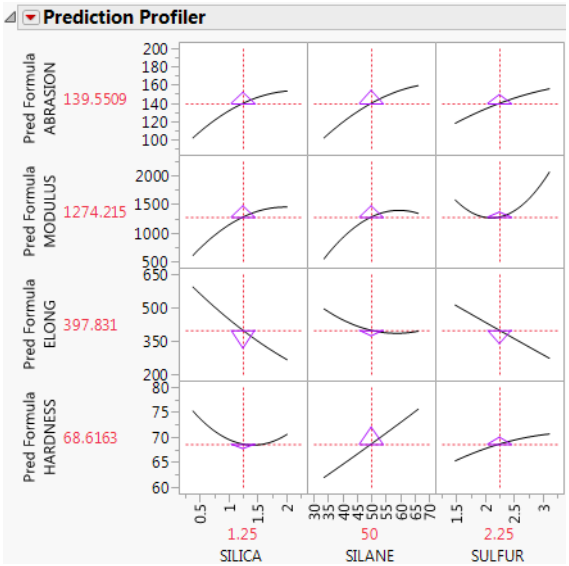
Interaction Profiler Shows or hides interaction plots that update as you update the factor values in the Prediction Profiler. Use this option to help visualize third degree interactions by seeing how the plot changes as current values for the factors change. The cells that change for a given factor are the cells that do not involve that factor directly.

Confidence Intervals Shows or hides confidence intervals in the Prediction Profiler plot. The intervals are drawn by bars for categorical factors, and curves for continuous factors. The interval values are also displayed on the vertical axis in blue. These are available when the profiler is used inside certain fitting platforms or when a standard error column of the form PredSE<colname> has been specified in the Y, Prediction Formula role of the Prediction Profiler launch dialog.

Prop of Error Bars (Appears when a Sigma column property exists in any of the factor and response variables.) Shows or hides the 3σ interval that is implied on the response due to the variation in the factor. The interval values are also displayed on the vertical axis in green. Propagation of error (POE) is important when attributing the variation of the response in terms of variation in the factor values when the factor values are not very controllable. See "Propagation of Error Bars".

Sensitivity Indicator Shows or hides a purple triangle whose height and direction correspond to the value of the partial derivative of the profile function at its current value. This is useful in large profiles to be able to quickly spot the sensitive cells.

Figure 3.5 Sensitivity Indicators



Profile at Boundary When analyzing a mixture design, JMP constrains the ranges of the factors so that settings outside the mixture constraints are not possible. This is why, in some mixture designs, the profile traces turn abruptly.

When there are mixture components that have constraints, other than the usual zero-to-one constraint, a new submenu, called Profile at Boundary, appears on the Prediction Profiler red triangle menu. It has the following two options:

Turn At Boundaries Lets the settings continue along the boundary of the restraint condition.

Stop At Boundaries Truncates the prediction traces to the region where strict proportionality is maintained.

JMP PRO **Extrapolation Control** Shows a submenu of options for extrapolation control. This feature helps identify possible extrapolated predictions. A prediction is considered an extrapolation when it is made using a combination of factor points that are not within the factor space of the original data. In the extrapolation control feature, the metric used to determine if a point is an extrapolation depends on the type of model fit. For models that are fit in the Standard Least Squares personality of the Fit Model platform, the leverage at the factor settings is used as the extrapolation metric. For all other models, the regularized Hotelling's T^2 value is used as the extrapolation metric. See "Extrapolation Control Metrics".

Extrapolation Control is available in profilers embedded in platforms and profilers launched from the Graph menu. The data used for the extrapolation control metrics depends on the type of profiler.

- When a model is built with validation, the embedded profiler and extrapolation control metrics are based on the training data.
- If you launch a profiler from the Graph menu the extrapolation control metrics are based on all data, unless you specifically exclude certain rows.
- When a model is built in a platform that ignores missing values during model fitting, those rows are excluded from the embedded profiler and extrapolation control metrics.
- When a model is built with Informative Missing, the embedded profiler and extrapolation control metrics reflect the informative missing.
- To include informative missing in the extrapolation control metrics when launching the profiler from the Graph menu use the Informative Missing column property.
- If you call Extrapolation Control from a profiler launched from the Graph menu, the regularized Hotelling's T^2 value is always used as the extrapolation metric, regardless of the type of model fit. Therefore, the extrapolation control results from a profiler embedded in the Standard Least Squares platform will not match those from the Graph menu profiler.

The extrapolation control red triangle menu includes options to either warn of possible extrapolation or to restrict the factor settings so that extrapolated predictions are not shown.

Off Turns off all extrapolation controls and warnings.

On Turns on extrapolation control. When this option is selected, it is indicated at the top of the profiler and the profile traces are restricted to factor combinations that do not lead to extrapolations.

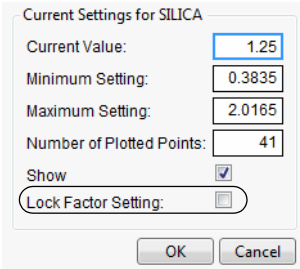
Warning On Turns on extrapolation warnings. When this option is selected, it is indicated at the top of the profiler. If a factor combination is selected that produces an extrapolation, an alert appears that reads --Possible Extrapolation--.

Extrapolation Details Shows or hides the extrapolation control details above the prediction profiler. The extrapolation control details include the value of the extrapolation metric at the current point, the value of the extrapolation threshold, the type of extrapolation metric, and the definition of the extrapolation threshold.

Set Threshold Criterion Opens a window that enables you to adjust the extrapolation threshold. When the extrapolation metric is the leverage at the factor settings, you can specify how the leverage is computed and the value of the corresponding multiplier. When the extrapolation metric is the regularized Hotelling's T^2 value, you can specify the multiplier. See "Extrapolation Control Metrics".

Reset Factor Grid Displays a window for each factor enabling you to enter a specific value for the factor's current setting, to lock that setting, and to control aspects of the grid. See the section "Set or Lock Factor Values".

Figure 3.6 Factor Settings Window



Factor Settings Submenu that consists of the following options:

Remember Settings Adds an outline node to the report that accumulates the values of the current settings each time the Remember Settings command is invoked. Each remembered setting is preceded by a radio button that is used to reset to those settings. There are options to remove selected settings or all settings in the Remember Settings red triangle menu.

Set To Data in Row Assigns the values of a data table row to the *X* variables in the Prediction Profiler.

Copy Settings Script Copies the current Prediction Profiler's settings to the clipboard.

Paste Settings Script Pastes the Prediction Profiler settings from the clipboard to a Prediction Profiler in another report.

Append Settings to Table Appends the current profiler's settings to the end of the data table. This is useful if you have a combination of settings in the Prediction Profiler that you want to add to an experiment in order to do another run.

Broadcast Factor Settings Sends the current profiler's factor settings to all other profilers, but does not link the profilers. A change in a factor in one profiler does not cause changes in any other profilers unless Broadcast Factor Settings is selected again.

Link Profilers Links all the profilers together. A change in a factor in one profiler causes that factor to change to that value in all other profilers, including Surface Plot. This is a global option, set, or unset for all profilers.

Set Script Sets a script that is called each time a factor changes. The set script receives a list of arguments of the form:

```
{factor1 = n1, factor2 = n2, ...}
```

For example, to write this list to the log, first define a function:

```
ProfileCallbackLog = Function({arg},show(arg));
```

Then enter `ProfileCallbackLog` in the Set Script dialog.

Similar functions convert the factor values to global values:

```
ProfileCallbackAssign = Function({arg},evalList(arg));
```

Or access the values one at a time:

```
ProfileCallbackAccess =
Function({arg},f1=arg["factor1"];f2=arg["factor2"]);
```

Unthreaded Enables you to change to an unthreaded analysis if multithreading does not work.

Animation Shows or hides animation controls that enable you to easily cycle through a variety of factor settings. See "Animation Controls".

Default N Levels Enables you to set the default number of levels for each continuous factor. This option is useful when the Prediction Profiler is especially large. When calculating the traces for the first time, JMP measures how long it takes. If this time is greater than three seconds, you are alerted that decreasing the Default N Levels speeds up the calculations.

Output Grid Table Produces a new data table with columns for the factors that contain grid values, columns for each of the responses with computed values at each grid point, and the desirability computation at each grid point.

If you have a lot of factors, it is impractical to use the Output Grid Table command, because it produces a large table. A memory allocation message might be displayed for large grid tables. In such cases, you should lock some of the factors, which are held at locked, constant values. To get the window to specify locked columns, ALT- or Option-click inside the profiler graph to get a window that has a Lock Factor Setting check box.

Output Random Table Prompts for a number of runs and creates an output table with that many rows, with random factor settings and predicted values over those settings. This is equivalent to (but much simpler than) opening the Simulator, resetting all the factors to a random uniform distribution, then simulating output. This command is similar to Output Grid Table, except it results in a random table rather than a sequenced one.

The prime reason to make uniform random factor tables is to explore the factor space in a multivariate way using graphical queries. This technique is called *Filtered Monte Carlo*.

Suppose you want to see the locus of all factor settings that produce a given range to desirable response settings. By selecting and hiding the points that do not qualify (using

graphical brushing or the Data Filter), you see the possibilities of what is left: the opportunity space yielding the result that you want.

Some rows might appear selected and marked with a red dot. These represent the points on the multivariate desirability Pareto Frontier - the points that are not dominated by other points with respect to the desirability of all the factors.

Alter Linear Constraints Enables you to add, change, or delete linear constraints. The constraints are incorporated into the operation of Prediction Profiler. See "Linear Constraints".

Save Linear Constraints Enables you to save existing linear constraints to a table script called Constraint. See "Linear Constraints".

Conditional Predictions Appears when random effects are included in the model. The random effects predictions are used in formulating the predicted value and profiles.

Appearance Submenu that consists of the following options:

Arrange in Rows Enter the number of plots that appear in a row. This option helps you view plots vertically rather than in one wide row.

Note: To set a default number of plots to appear in a row, go to File > Preferences > Platforms > Profiler and edit the Arrange in Rows preference.

Graph Spacing Opens a window that enables you to set the amount of horizontal space between graph panels.

Reorder X Variables Opens a window where you can reorder the model main effects by dragging them to the desired order.

Reorder Y Variables Opens a window where you can reorder the responses by dragging them to the desired order.

Adapt Y Axis Re-scales the vertical axis if the response is outside the axis range, so that the range of the response is included.

Show Creator Shows or hides the name of the platform that created the formula in the response column. The platform name appears on the vertical axis. (Available only if the response column contains a "Creator" named argument in the "Predicting" column property.)

Animation Controls

Figure 3.7 Animation Controls

Play/Pause Press play to animate the profiler. Moves through a cycle of factor settings and loops back to the beginning when the cycle is complete. Press pause to stop the animation.

Cycle Type Lists the types of cycles for factor settings.

Sequential Cycles through values for each factor, one factor after another. The name of the factor the animation is currently cycling through is displayed next to the speed slider bar.

Single Factor Cycles through values for the selected factor while all other factors are held constant. The name of the selected factor is displayed next to the speed slider bar.

Random Randomly cycles through different combinations of factor settings/values.

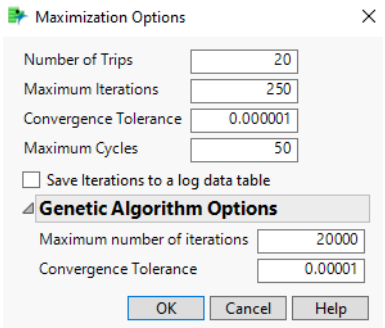
Data Sequential Sets the factor values to a row in the data table, one row at a time, starting with row 1. The row number of the current factor setting is displayed next to the speed slider bar.

Data Random Sets the factor values to a random row in the data table, one row at a time. The row number of the current factoring setting is displayed next to the speed slider bar.

Speed Slider Bar Use the slider bar to adjust the speed of the animation.

Maximization Options Window

Figure 3.8 Maximization Options Window



Number of Trips The number of random starts in the optimization algorithm. Each trip starts with a randomly chosen value for each of the factors in the model.

Maximum Iterations The maximum number of steps taken within each trip. If the algorithm does not converge before the maximum iterations is reached, a new trip is started.

Convergence Tolerance The upper limit for the convergence criterion for the optimization algorithm. If the convergence criterion is less than this value for two consecutive iterations, the algorithm stops.

Maximum Cycles (Not applicable if there are linear constraints.) The maximum number of cycles in the algorithm.

Genetic Algorithm Options (Available only if there is a disallowed combination constraint or if one of the extrapolation control features is turned on.) Provides additional options for the genetic algorithm. The genetic algorithm is used only if the solution from the regular optimization algorithm is not feasible.

Maximum number of iterations The maximum number of iterations for the genetic algorithm.

Convergence Tolerance The upper limit for the convergence criterion for the genetic algorithm.

Caution: The solution found by the genetic algorithm is not guaranteed to be the global optimum. This is true for all genetic algorithms. A genetic algorithm is needed for optimization problems that have more general constraints. For example, JMP uses the genetic algorithm when a disallowed combinations constraint is present because these types of constraints can involve categorical factors, be nonlinear, or be any general constraint entered via JSL.

Desirability Profiling and Optimization

You can define a desirability function for a single response variable or for several response variables. When you are optimizing relative to several responses, there can often be competing criteria. For example, you might want to maximize one response, minimize another, and keep a third response close to some target value.

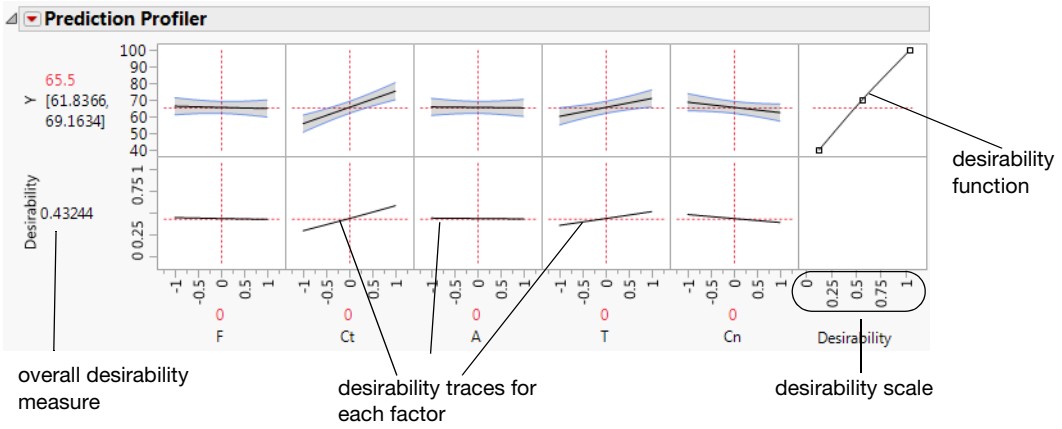
In desirability profiling, you specify a desirability function for each response. The overall desirability for all responses is defined as the geometric mean of the desirability functions for the individual responses. See Derringer and Suich (1980) for information about combining responses.

To use desirability profiling, select **Optimization and Desirability > Desirability Functions** from the Prediction Profiler red triangle menu.

Note: If the response column has a Response Limits property, desirability functions are turned on by default.

This command appends a new row to the bottom of the plot matrix, dedicated to graphing desirability. The row has a plot for each factor showing its *desirability trace*, as illustrated in Figure 3.9. It also adds a column that has an adjustable desirability function for each *Y* variable. The overall desirability measure shows on a scale of zero to one at the left of the row of desirability traces.

Figure 3.9 The Desirability Prediction Profiler



Construction of Desirability Functions

The individual desirability functions are smooth piecewise functions that pass through three defining points. These points are called control points (Low, Middle, High) and can be used to interactively control the shape of the desirability function.

- The Minimize and Maximize functions are three-part piecewise smooth functions that consist of interpolating cubics between the control points and exponentials in the tails.
- The Target function is a piecewise function that is a scale multiple of a normal density on either side of the Middle value (with different curves on each side), which is also piecewise smooth and fit to the control points. Exponential functions are fit to the tails.
- The None function enables you to specify an arbitrary desirability function. In particular, you can specify desirability to be lower at the Middle value than at the Low and High values. You can also construct custom desirability functions using formulas. See "Customized Desirability Functions".

The Low and High control points are not allowed to reach all the way to zero or one. This approach to constructing the desirability functions results in good behavior as the desirability values switch between the maximize, target, and minimize values.

Note: JMP does not use the Derringer and Suich (1980) functional forms. Because they are not smooth, they do not always work well with JMP's optimization algorithm.

Desirability Function for Multiple Optimization

When multiple responses are to be optimized, an overall desirability function is constructed and optimized. The overall desirability for all responses is defined as the geometric mean of the desirability functions for the individual responses.

Denote the individual desirability functions for k responses by $d_1, d_2,...,d_k$. Then the overall desirability function is the geometric mean of the individual desirability functions:

$$D = d_1^{1/k} d_2^{1/k} \ldots d_k^{1/k}$$

If Importance values are defined as part of the Response Limits column property or are defined in the Response Goal window, they are integrated into the overall desirability function. The Importance values are scaled so that they sum to 1. Denote the scaled importance values by $w_1, w_2,..., w_k$. Then the overall desirability is defined as a weighted geometric mean of the individual desirability functions:

$$D = d_1^{w_1} d_2^{w_2} \ldots d_k^{w_k}$$

Optimization Algorithm

The method used for optimization of the overall desirability function, or of the single desirability function if there is only one response, depends on the factor types.

- For categorical factors, a coordinate exchange algorithm is used.
- For continuous factors, a gradient descent algorithm is used.
- In the presence of constraints or mixture factors, a Wolfe reduced-gradient approach is used.
- To reduce the risk of finding local optima, JMP uses multiple random starts.

How to Use the Desirability Function

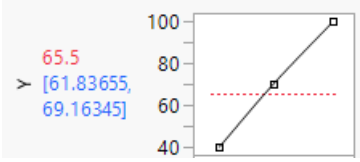
To use a variable's desirability function, drag the function handles to represent a response value.

As you drag these handles, the changing response value shows in the area labeled Desirability to the left of the plots. The dotted line is the response for the current factor settings. The overall desirability shows to the left of the row of desirability traces. Alternatively, you can select **Optimization and Desirability > Set Desirabilities** to enter specific values for the points.

Figure 3.10 shows steps to create desirability settings.

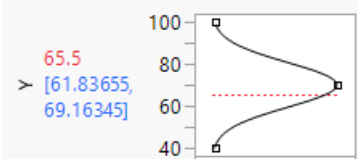
Maximize The default desirability function setting is maximize ("higher is better"). The High control point is positioned at the maximum Y value and aligned at the high desirability, close to 1. The Low control point is positioned at the minimum Y value and aligned at a low desirability, close to 0.

Figure 3.10 Maximizing Desirability



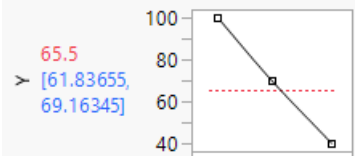
Target You can designate a target value as "best." In this example, the middle control point is positioned at Y = 70 and aligned with the maximum desirability of 1. Y becomes less desirable as its value approaches either 40 or 100. The High and Low control points at Y = 40 and Y = 100 are positioned at the minimum desirability close to 0.

Figure 3.11 Defining a Target Desirability



Minimize The minimize ("smaller is better") desirability function associates high response values with low desirability and low response values with high desirability. The curve is the maximization curve flipped around a horizontal line at the center of plot.

Figure 3.12 Minimizing Desirability



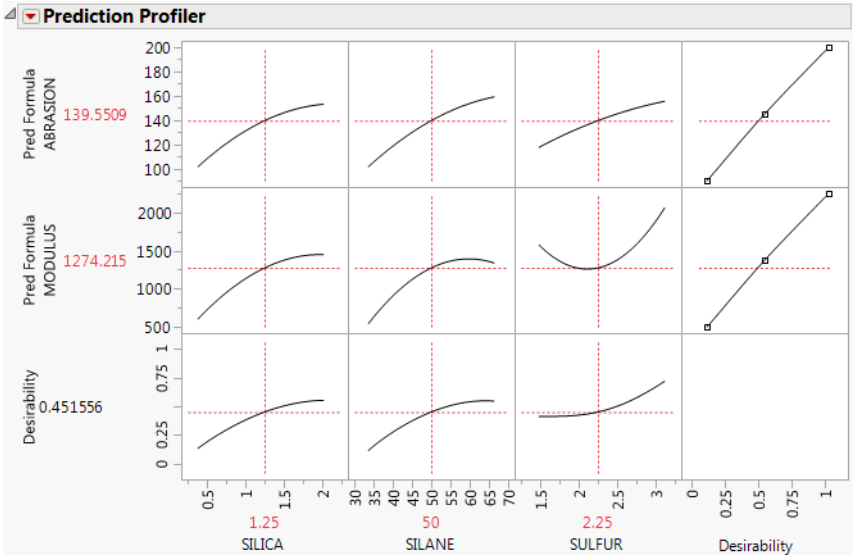
Note: Dragging the High or Low control point of a maximize or minimize desirability function across the *y*-value of the middle point results in the opposite point reflecting. A Minimize becomes a Maximize, and vice versa.

The Desirability Profile

The last row of plots shows the desirability trace for each factor. The numerical value beside the word Desirability on the vertical axis is the geometric mean of the desirability measures. This row of plots shows both the current desirability and the trace of desirabilities that result from changing one factor at a time.

For example, Figure 3.13 shows desirability functions for two responses. You want to maximize ABRASION and MODULUS. The desirability plots indicate that you could increase the desirability by increasing any of the factors.

Figure 3.13 Prediction Profile Plot with Adjusted Desirability and Factor Values



Customized Desirability Functions

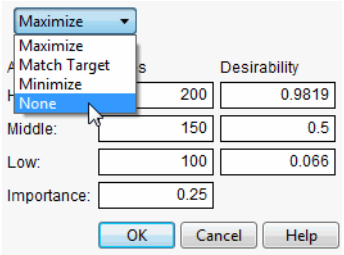
It is possible to use a customized desirability function. For example, suppose you want to maximize using the following function.

Figure 3.14 Maximizing Desirability Based on a Function

$$\frac{Pred\ Formula\ ABRASION}{96} + \frac{Pred\ Formula\ MODULUS}{700} + \frac{33}{(|Pred\ Formula\ ELONG - 450| + 1)} + \frac{2}{(|Pred\ Formula\ HARDNESS - 67| + 1)}$$

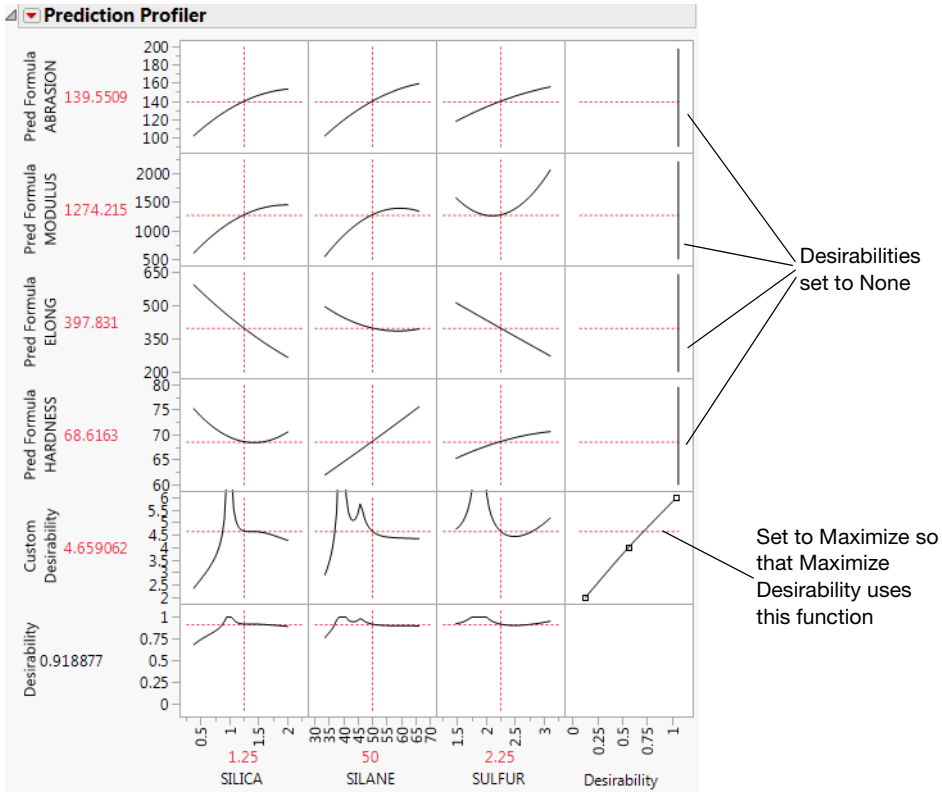
First, create a column called Custom Desirability that contains the above formula. Then, select **Graph > Profiler** to launch the platform. Select all the Pred Formula columns and the Custom Desirability column and select **Y, Prediction Formula**. Check the **Expand Intermediate Formula** option. Turn on the desirability functions by selecting **Optimization and Desirability > Desirability Functions** from the red triangle menu. All the desirability functions for the individual effects must be turned off. To do this, first double-click in a desirability plot window, and then select **None** in the window that appears. Set the desirability for Custom Desirability to be maximized.

Figure 3.15 Selecting No Desirability Goal



At this point, selecting **Optimization and Desirability > Maximize Desirability** uses only the custom Custom Desirability function.

Figure 3.16 Maximized Custom Desirability



Assess Variable Importance

The Variable Importance report calculates indices that measure the importance of factors in a model in a way that is independent of the model type and fitting method. The fitted model is used only in calculating predicted values. The method estimates the variability in the predicted response based on a range of variation for each factor. If variation in the factor causes high variability in the response, then that effect is important relative to the model.

Note: In some platforms, Assess Variable Importance is not available for categorical responses with more than two levels.

Assess Variable Importance can also be accessed in the Prediction Profiler that is obtained through the Graph menu.

For statistical details, see "Assess Variable Importance". See also Saltelli (2002).

Note: Assess Variable Importance requires that all columns reside in the same data table.

The Assess Variable Importance Report

The Assess Variable Importance menu has the following options that address the methodology used in constructing importance indices:

Independent Uniform Inputs For each factor, Monte Carlo samples are drawn from a uniform distribution defined by the minimum and maximum observed values. Use this option when you believe that your factors are uncorrelated and that their likely values are uniformly spread over the range represented in the study. This is the appropriate option for designed experiments that do not involve constraints or mixture factors.

Independent Resampled Inputs For each factor, Monte Carlo samples are obtained by resampling its set of observed values. Use this option when you believe that your factors are uncorrelated and that their likely values are not represented by a uniform distribution.

Dependent Resampled Inputs Factor values are constructed from observed combinations using a *k*-nearest neighbors approach, in order to account for correlation. This option treats observed variance and covariance as representative of the covariance structure for your factors. Use this option when you believe that your factors are correlated. Note that this option is sensitive to the number of rows in the data table. If used with a small number of rows, the results can be unreliable.

Note: The Independent Resampled Inputs and Dependent Resampled Inputs options are intended for observational studies. The Independent option is faster than the Dependent option, but the Dependent option handles multicollinearity better and does not extrapolate into regions far away from the data.

Linearly Constrained Inputs For each factor, Monte Carlo samples are drawn from a uniform distribution over a region defined by linear constraints. The linear constraints can be defined in the Prediction Profiler or constructed in connection with a designed experiment. In addition, the samples are restricted to fall within the minimum and maximum observed values. Use this option in the presence of linear constraints, when you believe that these constraints impact the distribution of the inputs.

The speed of these algorithms depends on the model evaluation speed. In general, the fastest option is Independent Uniform Inputs and the slowest is Dependent Resampled Inputs. You have the option to Accept Current Indices when the estimation process is unable to complete instantaneously.

Note: Variable importance indices are constructed using Monte Carlo sampling. For this reason, you can expect some variation in importance index values from one run to another.

Variable Importance Report

Each Assess Variable Importance option presents a Summary Report and Marginal Model Plots. When the Assess Variable Importance report opens, the factors in the Prediction Profiler are reordered according to their Total Effect importance indices. When there are multiple responses, the factors are reordered according to the Total Effect importance indices in the Overall report. When you run several Variable Importance reports, the factors in the Prediction Profiler are ordered according to their Total Effect indices in the most recent report.

Summary Report

For each response, a table displays the following elements:

Column The factor of interest.

Main Effect An importance index that reflects the relative contribution of that factor alone, not in combination with other factors.

Total Effect An importance index that reflects the relative contribution of that factor both alone and in combination with other factors. The Total Effect column is displayed as a bar chart. See "Weights".

Main Effect Std Error The Monte Carlo standard error of the Main Effect's importance index. This is a hidden column that you can access by right-clicking in the report and selecting **Columns > Main Effect Std Error**. By default, sampling continues until this error is less than 0.01. Details of the calculation are given in "Variable Importance Standard Errors". (Not available for Dependent Resampled Inputs option.)

Total Effect Std Error The Monte Carlo standard error of the Total Effect's importance index. This is a hidden column that you can access by right-clicking in the report and selecting **Columns > Total Effect Std Error**. By default, sampling continues until this error is less than 0.01. Details of the calculation are given in "Variable Importance Standard Errors". (Not available for Dependent Resampled Inputs option.)

Weights A plot that shows the Total Effect indices, located to the right of the final column. You can deselect or reselect this plot by right-clicking in the report and selecting **Columns > Weights**.

Proportion of function evaluations with missing values The proportion of Monte Carlo samples for which some combination of inputs results in an inestimable prediction. When the proportion is nonzero, this message appears as a note at the bottom of the table.

Note: When you have more than one response, the Summary Report presents an Overall table followed by tables for each response. The importance indices in the Overall report are the averages of the importance indices across all responses.

Marginal Model Plots

The Marginal Model Plots report (Figure 3.39) shows a matrix of plots, with a row for each response and columns for the factors. The factors are ordered according to the size of their overall Total Effect importance indices.

For a given response and factor, the plot shows the mean response for each factor value, where that mean is taken over all inputs to the calculation of importance indices. These plots differ from profiler plots, which show cross sections of the response. Marginal Model Plots are useful for assessing the main effects of factors.

Note that your choice of input methodology impacts the values plotted on marginal model plots. Also, because the plots are based on the generated input settings, the plotted mean responses might not follow a smooth curve.

The red triangle menu options enable you to show or hide the following aspects of the plots:

Estimate A smoothed estimate of the mean of the simulated values calculated as a function of the factor values.

> **Note:** The estimates of the mean are simulated, so the values change when you rerun the analysis.

Confidence Interval A 95% confidence band for the simulated means. This band is often narrow and might not be visible unless you expand the scale. Not available for Dependent Resampled Inputs.

> **Note:** The confidence bounds are simulated, so the bands change when you rerun the analysis.

Data The actual (unsimulated) values of the response plotted against the factor values.

Variable Importance Options

The Variable Importance red triangle menu contains the following options:

Reorder factors by main effect importance Reorders the cells in the Prediction Profiler in accordance with the importance indices for the main effects (Main Effect).

Reorder factors by total importance Reorders the cells in the Prediction Profiler in accordance with the total importance indices for the factors (Total Effect).

Colorize Profiler Colors cells in the profiler by Total Effect importance indices using a red to white intensity scale.

Note: You can click rows in the Summary Report to select columns in the data table. This can facilitate further analyses.

Bagging

Bootstrap aggregating (*bagging*) is a technique to improve predictive performance while also gaining insight into the reliability of predictions. Bagging is especially useful in unstable methods, including neural networks, classification trees, and regression trees.

Bagging creates *M* training data sets by sampling with replacement from the original data. All training data sets are of the same size as the original. For each training data set, a model is fit using the analysis platform, and predictions are made. Therefore, there are a total of *M* predictions for each observation in the original data set. The final prediction is the average of the *M* predictions.

Bagging is available in many analysis platforms. To use bagging, select **Save Bagged Predictions** from the Prediction Profiler red triangle menu. A window appears with the following options for Bagging:

Number of Bootstrap Samples Sets the number of times that you want to resample the data and build a model. A larger number results in more precise predictions. By default, the number of bootstrap samples is 100.

Random Seed Sets a random seed that you can re-enter in subsequent runs of the bagging analysis to duplicate your current results. By default, no seed is set.

Fractional Weights Performs a Bayesian bagging analysis. In each bootstrap iteration, each observation is assigned a nonzero weight. The model that makes the predictions uses the weighted observations. By default, the Fractional Weights option is not selected, and a simple bagging analysis is conducted.

Tip: Use the Fractional Weights option if the number of observations that are used in your analysis is small or if you are concerned about separation in a logistic regression setting.

Suppose that Fractional Weights is selected. For each bootstrap iteration, each observation that is used in the report is assigned a nonzero weight. These weights sum to *n,* the number of observations used in the model. For more information about how the weights are calculated and used, see *Basic Analysis*.

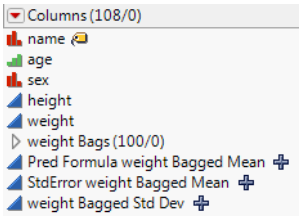
Save Prediction Formulas For each bagged prediction, this option saves the formula used to make that prediction in the column properties. This option is available in only a subset of the analysis platforms that offer bagging.

Note: If Save Prediction Formulas is not available, a note appears, stating that only the predicted values will be saved.

Bagging automatically creates new columns in the original data table. All M sets of bagged predictions are saved as hidden columns. The final prediction is saved in a column named "Pred Formula <colname> Bagged Mean". The standard deviation of the final prediction is saved in a column named "<colname> Bagged Std Dev". The standard error of the bagged mean is saved in a column named "StdError <colname> Bagged Mean." The standard error is the standard deviation divided by $\sqrt{M-1}$. Here, <colname> identifies the column in the report that was bagged.

The standard error gives insight about the precision of the prediction. A very small standard error indicates a precise prediction for that observation. For more information about bagging, see Hastie et al. (2009).

Figure 3.17 Bagging Columns



Additional Examples of the Prediction Profiler

This section contains additional examples illustrating various aspects of the Prediction Profiler platform and the Prediction Profiler embedded in many analysis platforms.

- "Example of Desirability Profiling for Multiple Responses"
- "Example of a Noise Factor in the Prediction Profiler"
- "Example of Variable Importance for One Response"
- "Example of Variable Importance for Multiple Responses"
- "Example of Bagging to Improve Prediction"
- "Example of Bagging to Indicate the Accuracy of Predictions"
- "Example of Confidence Intervals Using Bagging"
- "Example of Confidence Intervals Using StdErr Pred Formula Column"

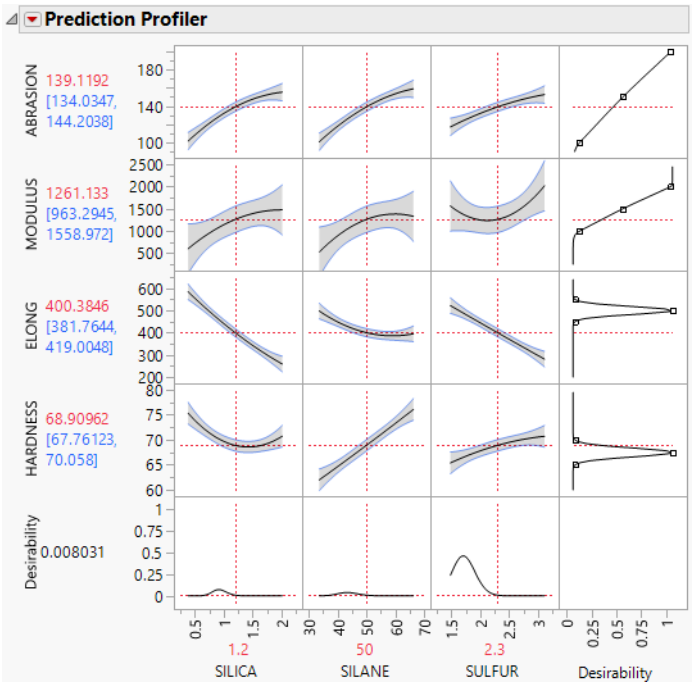
Example of Desirability Profiling for Multiple Responses

A desirability index becomes especially useful when there are multiple responses. The idea was pioneered by Derringer and Suich (1980), who give the following example. Suppose there are four responses, ABRASION, MODULUS, ELONG, and HARDNESS. Three factors, SILICA, SILANE, and SULFUR, were used in a central composite design.

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Click the green triangle next to RSM for 4 Responses to run the script.

   This script defines a model for the four responses with a full quadratic response surface. The summary tables and effect information appear for all the responses, followed by the prediction profiler shown in Figure 3.18.

Figure 3.18 Prediction Profiler for Multiple Responses before Optimization

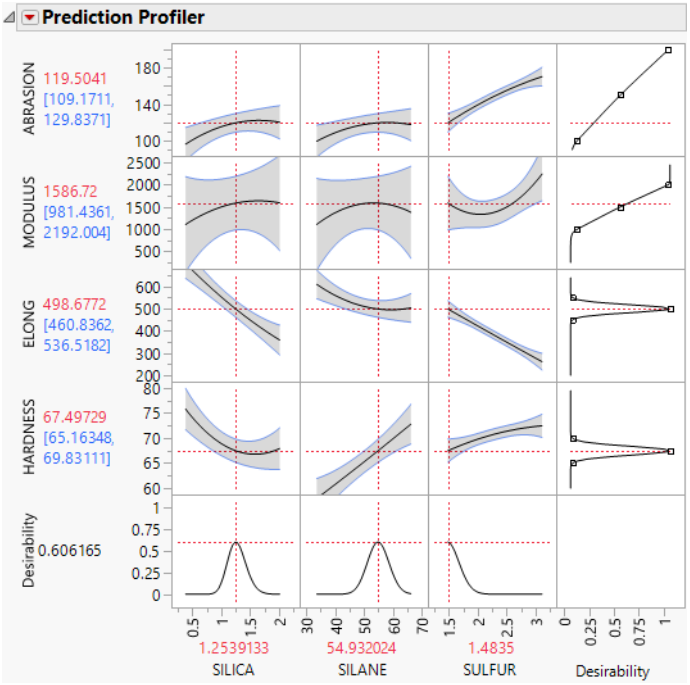


Note the following about the desirability functions:

- Maximum ABRASION and maximum MODULUS are most desirable.
- ELONG target of 500 is most desirable.
- HARDNESS target of 67.5 is most desirable.

3. Select **Optimization and Desirability > Maximize Desirability** from the Prediction Profiler red triangle menu to maximize desirability.

Figure 3.19 Prediction Profiler after Optimization



The desirability traces at the bottom decrease everywhere except the current values of the effects, which indicates that any further adjustment could decrease the overall desirability.

Example of a Noise Factor in the Prediction Profiler

This example uses the Tiretread.jmp sample data set. This data set shows the results of a tire manufacturer's experiment whose objective is to match a target value of HARDNESS= 70 based on three factors: SILICA, SILANE, and SULFUR content. Suppose the SILANE and SULFUR content are easily (and precisely) controllable, but SILICA is not easily controlled.

For comparison, first optimize the factors for hardness without considering variation from the noise factor, SILICA.

1. Select **Graph > Profiler** to launch the Prediction Profiler.
2. Assign Pred Formula HARDNESS to the **Y, Prediction Formula** role.
3. Click **OK**.
4. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Desirability Functions**.
5. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Set Desirabilities** to open the Response Goal window**.**

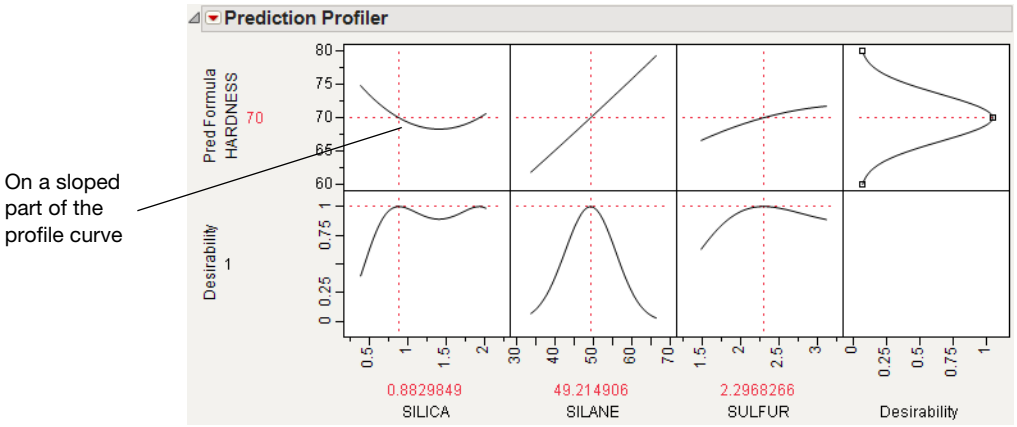
Tip: Alternatively, you can double-click in the Desirability plot to open the Response Goal window.

6. Select **Match Target** from the list and click **OK**.
7. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability** to find the optimum factor settings for the target value of HARDNESS.

   We get the following Prediction Profiler display. Notice that the SILICA factor's optimum value is on a sloped part of a profile curve. This means that variations in SILICA settings impact the response, HARDNESS.

Note: You might get different results from these because different combinations of factor values can all hit the target.

Figure 3.20 Maximizing Desirability for HARDNESS



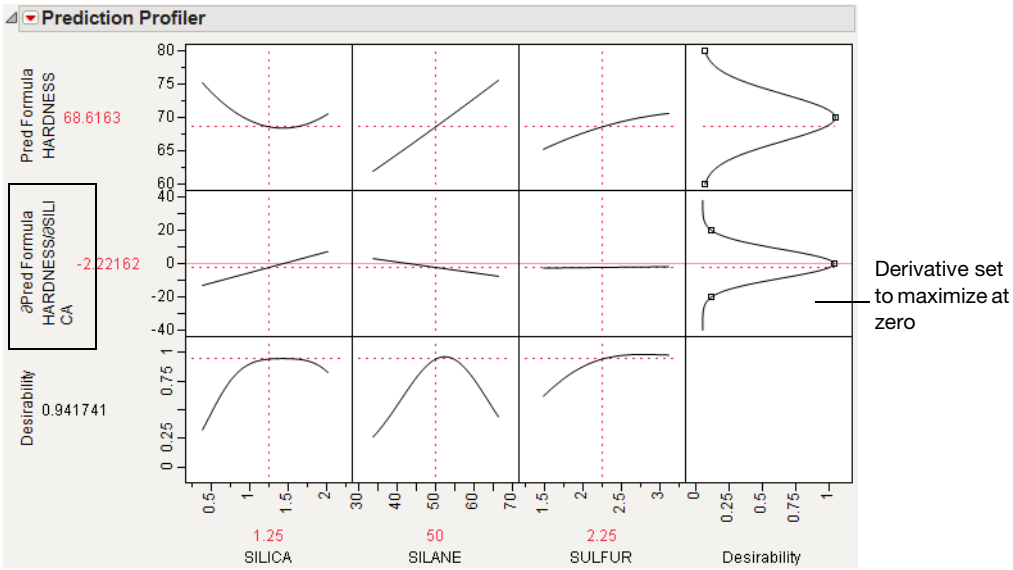
Now, we would like to not just optimize for a specific target value of HARDNESS, but also find a setting that is insensitive to the variability in the setting of Silica. Repeat the optimization process with SILICA as a noise factor.

1. Select **Graph > Profiler**.
2. Select Pred Formula HARDNESS and click **Y, Prediction Formula**.
3. Select SILICA and click **Noise Factors**.
4. Click **OK**.
5. Select **Optimization and Desirability > Desirability Functions** in the Prediction Profiler menu.

6. Double-click in the Desirability plot for the response Pred Formula HARDNESS to open the Response Goal window.
7. Select **Match Target** from the list and click **OK**.

   The resulting profiler has the appropriate derivative of the fitted model with respect to the noise factor, set to be maximized at zero, its flattest point.

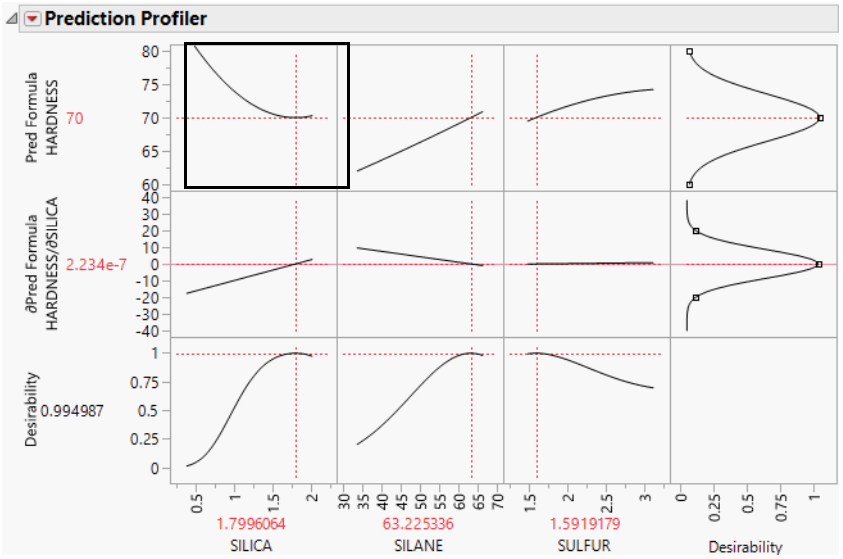
Figure 3.21 Derivative of the Prediction Formula with Respect to Silica



8. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability** to find the optimum values for the process factor, balancing for the noise factor.

   This time, we have also hit the targeted value of HARDNESS, but our value of SILICA is on its flatter region. This means variation in SILICA does not transmit as much variation to HARDNESS as when it was set at a value on the steep part of the SILICA curve.

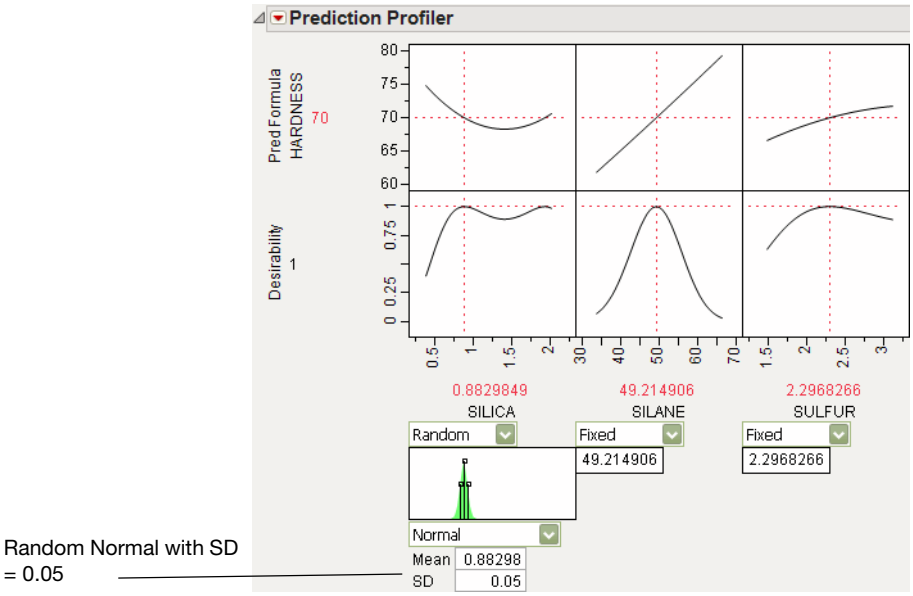
Figure 3.22 Maximize Desirability



You can see the effect of accounting for the SILICA variation on the variance of the predictions by following these steps for each profiler (one without the noise factor, and one with the noise factor):

1. In the Profiler without the noise factor, click the Prediction Profiler red triangle and select **Simulator**.
2. Assign SILICA to have a random Normal distribution with a standard deviation of 0.05. Use the default setting for the mean.

Figure 3.23 Setting a Random Normal Distribution



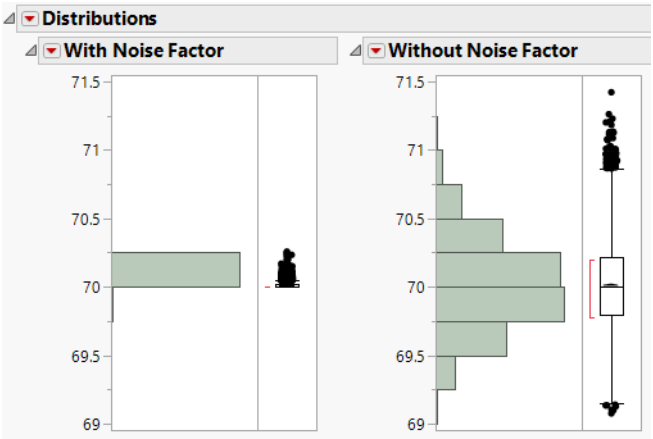
3. Click **Simulate**.
4. Click the Simulate to Table gray triangle and then click the **Make Table** button.
5. Rename the Pred Formula HARDNESS column to Without Noise Factor.
6. In the Profiler with the noise factor, click the Prediction Profiler red triangle and select **Simulator**.
7. Assign SILICA to have a random Normal distribution with a standard deviation of 0.05. Use the default setting for the mean.
8. Click **Simulate**.
9. Click the Simulate to Table gray triangle and then click the **Make Table** button.

Completing these steps for both the original and noise-factor-optimal simulations results in two similar data tables, one for each simulation. We want to compare the distributions of the predicted hardness from the two simulations. To facilitate the comparison we put the two prediction columns in a single data table.

1. Select the Pred Formula HARDNESS column in the current simulation table and select **Edit > Copy With Full Precision**.
2. In the other current simulation table, select **Cols > New Columns**.
3. Type With Noise Factor next to Column Name and click **OK**.
4. Select the With Noise Factor column and select **Edit > Paste**.
5. Select **Analyze > Distribution**.

6. Select With Noise Factor and Without Noise Factor and click **Y, Columns**.
7. Click **OK**.
8. Click the Distribution red triangle and select **Uniform Scaling**.

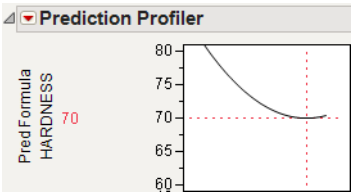
Figure 3.24 Comparison of Distributions with and without Noise Factors



The histograms show that there is much more variation in predicted Hardness when the noise factor was not included in the analysis.

It is also interesting to note the shape of the histogram when the noise factor was included. In the comparison histograms above, note that the With Noise Factor distribution has data trailing off in only one direction.The skewness in the With Noise Factor is clear if you deselect Uniform Scaling from the Distributions red triangle menu. The predictions are skewed because Hardness is at a minimum with respect to SILICA, as shown in Figure 3.25. Therefore, variation in SILICA can make only HARDNESS increase. When the non-robust solution is used, the variation could be transmitted either way.

Figure 3.25 Prediction Profiler Showing the Minima of HARDNESS by SILICA



Example of Variable Importance for One Response

The Boston Housing.jmp sample data table contains data on 13 factors that might relate to median home values. You fit a model using a neural network. Because neural networks do not accommodate formal hypothesis tests, these tests are not available to help assess which variables are important in predicting the response. However, for this purpose, you can use the Assess Variable Importance profiler option.

Note that your results might differ from, but should resemble, those shown here. There are two sources of random variability in this example. When you fit the neural network, *k*-fold cross validation is used. This partitions the data into training and validation sets at random. Also, Monte Carlo sampling is used to calculate the factor importance indices.

1. Select **Help > Sample Data Library** and open Boston Housing.jmp.
2. Select **Analyze > Predictive Modeling > Neural**.
3. Select mvalue from the Select Columns list and click **Y, Response**.
4. Select all other columns from the Select Columns list and click **X, Factor**.
5. Click **OK**.
6. In the Neural Model Launch panel, select **KFold** from the list under Validation Method.

   When you select KFold, the Number of Folds defaults to 5.
7. (Optional) Enter 123 next to Random Seed.

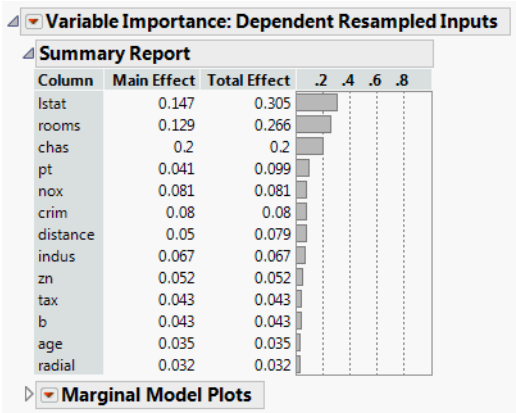
   Note: Results vary due to the random nature of choosing a validation set in the Neural Network model. Entering the seed above enables you to reproduce the results shown in this example.
8. Click **Go**.
9. Click the red triangle next to Model NTanH(3) and select **Profiler**.

   The Prediction Profiler is displayed at the very bottom of the report. Note the order of the factors for later comparison.

   Because the factors are correlated, you take this into account by choosing Dependent Resampled Inputs as the sampling method for assessing variable importance.
10. Click the Prediction Profiler red triangle and select **Assess Variable Importance > Dependent Resampled Inputs**.

    The Variable Importance: Dependent Resampled Inputs report appears. Check that the Prediction Profiler cells have been reordered by the magnitude of the Total Effect indices in the report. In Figure 3.26, check that the Total Effect importance indices identify rooms and lstat as the factors that have most impact on the predicted response.

Figure 3.26 Dependent Resampled Inputs Report

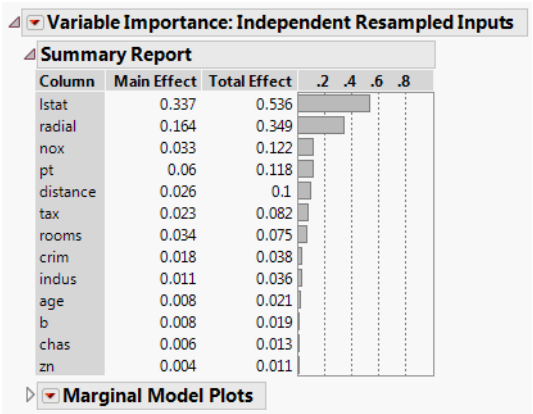


You might be interested in comparing the importance indices obtained assuming that the factors are correlated, with those obtained when the factors are assumed independent.

11. Click the Prediction Profiler red triangle and select **Assess Variable Importance > Independent Resampled Inputs**.

    The resampled inputs option makes sense in this example, because the distributions involved are not uniform. The Variable Importance: Independent Resampled Inputs report is shown in Figure 3.27. Check that the two factors identified as having the most impact on the predicted values are lstat and radial.

Figure 3.27 Independent Resampled Inputs Report



Example of Variable Importance for Multiple Responses

The data in the Tiretread.jmp sample data table are the result of a designed experiment where the factors are orthogonal. For this reason, you use importance estimates based on independent inputs. Suppose that you believe that, in practice, factor values vary throughout the design space, rather than assume only the settings defined in the experiment. Then you should choose Independent Uniform Inputs as the sampling scheme for your importance indices.

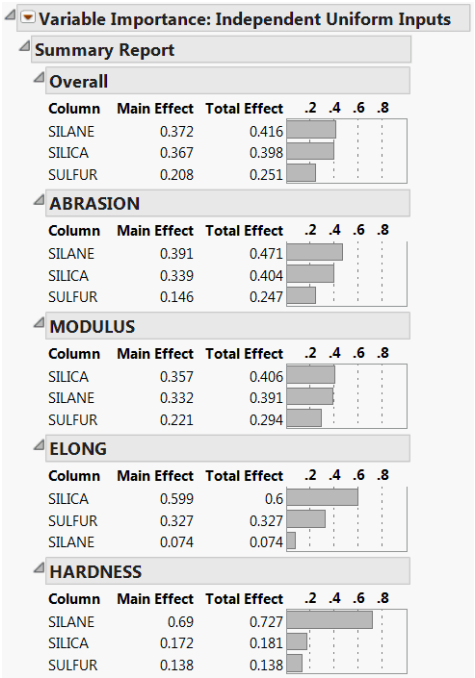
1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Run the script RSM for 4 Responses.

   The Prediction Profiler is displayed at the very bottom of the report.
3. Click the Prediction Profiler red triangle and select **Assess Variable Importance > Independent Uniform Inputs**.

   The Summary Report is shown in Figure 3.28. Because the importance indices are based on random sampling, your estimates might differ slightly from those shown in the figure.

   The report shows tables for each of the four responses. The Overall table averages the factor importance indices across responses. The factors in the Prediction Profiler (Figure 3.29) have been reordered to match their ordering on the Overall table's Total Effect importance.

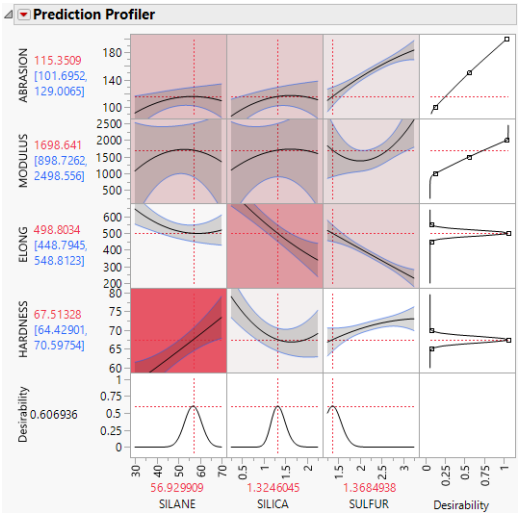
Figure 3.28 Summary Report for Four Responses



4. Click the red triangle next to Variable Importance: Independent Uniform Inputs and select **Colorize Profiler**.

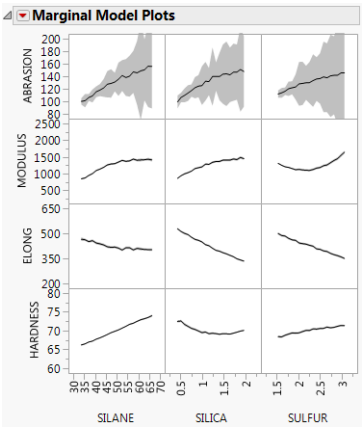
   Colors from a red to white intensity scale are overlaid on profiler panels to reflect Total Effect importance. For example, you easily see that the most important effect is that of Silane on Hardness.

Figure 3.29 Prediction Profiler for Four Responses



The Marginal Model Plots report shows mean responses for each factor across a uniform distribution of settings for the other two factors.

Figure 3.30 Marginal Model Plots for Four Responses



Example of Bagging to Improve Prediction

Bagging is used in a number of situations. One situation is to improve predictive power. Bagging is especially helpful for unstable models. This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). First, you fit a neural network model to simultaneously predict the four response variables as a function of the three factors. Then, you perform bagging on the neural network model. Last, you compare the predictions to show the improvements obtained through bagging.

Fit Neural Network Model

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Analyze > Predictive Modeling > Neural**.
3. Select ABRASION, MODULUS, ELONG, and HARDNESS and click **Y, Response**.
4. Select SILICA, SILANE, and SULFUR and click **X, Factor**.
5. Click **OK**.
6. (Optional) Enter 2121 next to Random Seed.

   Note: Results vary due to the random nature of choosing a validation set in the Neural Network model. Entering the seed above enables you to reproduce the results shown in this example.

7. Click **Go**.
8. Click the red triangle next to Model NTanH(3) and select **Save Formulas**.

   Note: This option saves the predicted values for all response variables from the neural network model to the data table. Later, these values are compared to the predictions that are obtained from bagging.

Perform Bagging

Now that the initial model has been constructed, you can perform bagging using that model. Access the Bagging feature through the Prediction Profiler.

1. Click the red triangle next to Model NTanH(3) and select **Profiler**.

   The Prediction Profiler appears at the bottom of the report.

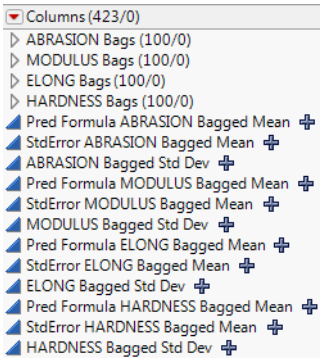
2. Click the Prediction Profiler red triangle and select **Save Bagged Predictions**.
3. Enter 100 next to Number of Bootstrap Samples.
4. (Optional) Enter 2121 next to Random Seed.

Note: Results vary due to the random nature of sampling with replacement. To reproduce the exact results in this example, set the Random Seed.

5. Click **OK**.

   Return to the data table. For each response variable, there are three new columns denoted as Pred Formula <colname> Bagged Mean, StdError <colname> Bagged Mean, <colname> Bagged Std Dev. The Pred Formula <colname> Bagged Mean columns are the final predictions.

Figure 3.31 Columns Added to Data Table After Bagging



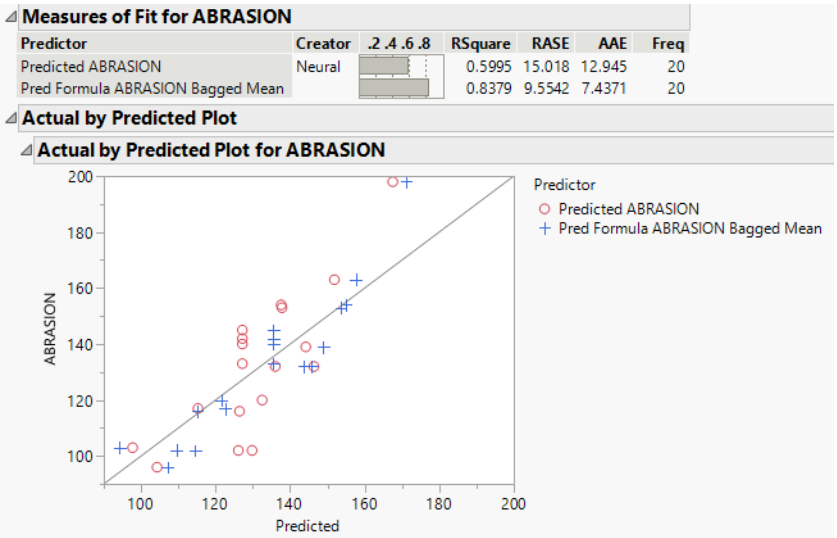
Compare the Predictions

To see how bagging improves predictive power, compare the predictions from the bagged model to the original model predictions. Use the Model Comparison platform to look at one response variable at a time.

1. Select **Analyze > Predictive Modeling > Model Comparison**.
2. Select Predicted ABRASION and click **Y, Predictors**.
3. Select Pred Formula ABRASION Bagged Mean and click **Y, Predictors.**
4. Click **OK**.

   A window that contains a list of columns appears.
5. Select ABRASION and click **OK**.
6. Click the Model Comparison red triangle and select **Plot Actual by Predicted**.

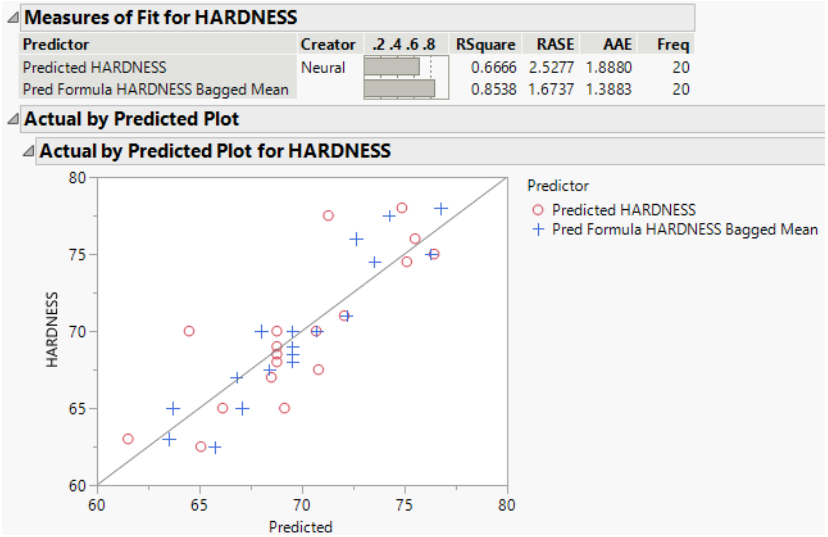
Figure 3.32 Comparison of Predictions for ABRASION



The Measures of Fit report and the Actual by Predicted Plot are shown in Figure 3.32. The predictions that were obtained from bagging are shown in blue. The predictions that were obtained from the original neural network model are shown in red. In general, the bagging predictions are closer to the line than the original model predictions. Because the bagging predictions are closer to the line, the RSquare value of 0.8379 for the bagged predictions is higher than the RSquare value for the original model predictions. You conclude that bagging improves predictions for ABRASION.

This example compared the predictions for ABRASION. To compare predictions for another response variable, follow step 2 through step 6, replacing ABRASION with the desired response variable. As another example, Figure 3.33 shows the Measures of Fit report for HARDNESS. The report shows similar findings as the Measures of Fit report for ABRASION. The RSquare value for the bagged predictions is slightly higher than the RSquare value for the original model predictions, which indicates a better fit and improved predictions.

Figure 3.33 Comparison of Predictions for HARDNESS



Example of Bagging to Indicate the Accuracy of Predictions

Bagging is also used to indicate the accuracy of the prediction through standard errors and other distributional measures. In platforms where the Save Predicted Formulas option is available in Bagging, you can make predictions on new observations and determine how accurate they are. The Save Predicted Formulas option is available in the Standard Least Squares, Generalized Regression, and Generalized Linear Models platforms.

In the Tiretread.jmp data table, suppose that you are interested in only predicting ABRASION as a function of the three factor variables. In this example, you fit a generalized regression model to predict ABRASION. Then, you perform bagging on that model. Last, you make a prediction for a new observation and investigate the accuracy of that prediction. This is done by obtaining a confidence interval for the prediction.

Fit a Generalized Regression Model

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Analyze > Fit Model**.
3. Select ABRASION and click **Y**.
4. Select **Generalized Regression** from the Personality list.
5. Select SILICA, SILANE, and SULFUR and click **Macros > Full Factorial**.

   This adds all terms, including interactions, to the model.
6. Click **Run**.

7. Click **Go**.

Figure 3.34 Parameter Estimates From Generalized Regression Report

Parameter Estimates for Original Predictors

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Intercept	-46.05865	7.526913	37.44455	<.0001*	-60.81113	-31.30617
SILICA	32.987289	2.2710897	210.97194	<.0001*	28.536036	37.438543
SILANE	1.7880765	0.1194432	224.10381	<.0001*	1.5539721	2.0221809
(SILICA-1.2)*(SILANE-50)	1.025	0.2316574	19.577413	<.0001*	0.5709598	1.4790402
SULFUR	21.813077	1.4913761	213.92398	<.0001*	18.890033	24.73612
(SILICA-1.2)*(SULFUR-2.3)	28.5	4.6331484	37.838796	<.0001*	19.419196	37.580804
(SILANE-50)*(SULFUR-2.3)	1.575	0.2316574	46.224125	<.0001*	1.1209598	2.0290402
(SILICA-1.2)*(SILANE-50)*(SULFUR-2.3)	1.55	0.4633148	11.192085	0.0008*	0.6419196	2.4580804

Normal Distribution Parameters	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Scale	5.2681724	0.7759427	46.095727	<.0001*	3.7473527	6.7889921

Perform Bagging

1. Click the red triangle next to Normal Lasso with AICc Validation and select **Profilers > Profiler**.

   The Prediction Profiler appears at the bottom of the report.
2. Click the Prediction Profiler red triangle and select **Save Bagged Predictions**.
3. Enter 500 next to Number of Bootstrap Samples.
4. (Optional) Enter 4321 next to Random Seed.

   Note: Results vary due to the random nature of sampling with replacement. To reproduce the exact results in this example, set the Random Seed.
5. Confirm that **Save Prediction Formulas** is selected.
6. Click **OK**.

   Note: This might take longer to run than the "Example of Bagging to Improve Prediction". The larger number of samples gives a better estimate of the prediction distributions.

   Return to the data table. For each response variable, there are three new columns denoted as Pred Formula <colname> Bagged Mean, StdError <colname> Bagged Mean, and <colname> Bagged Std Dev. The Pred Formula ABRASION Bagged Mean column is the final prediction.

Prediction for a New Observation

You now have predictions for ABRASION for each observation in the data table, as well as the standard errors for those predictions. Suppose that you have an observation with new values of 0.9, 43, and 2 for SILICA, SILANE, and SULFUR, respectively. You can predict the ABRASION response and obtain a confidence interval for that prediction because the Save Prediction Formulas option saves the regression equation for each bagged model. Therefore, *M* predictions are made with the new factor values to create a distribution of possible predictions. The mean is the final prediction, but analyzing the distribution tells you how accurate the prediction is.

1. In the data table, select **Rows > Add Rows**.
2. Enter 1 in the **How many rows to add** box and click **OK**.
3. Under the SILICA column, type 0.9 in the box for the new row.
4. Under the SILANE column, type 43 in the box for the new row.
5. Under the SULFUR column, type 2 in the box for the new row.

   All of the prediction columns for the new row are automatically calculated.

Figure 3.35 Values for New Row

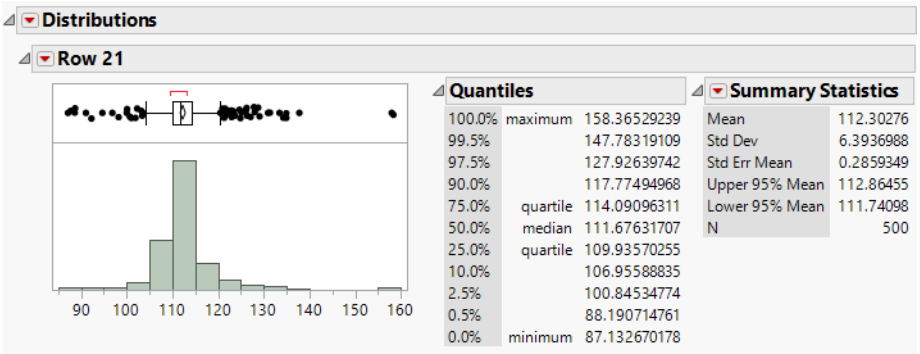
	ABRASION	MODULUS	ELONG	HARDNESS	SILICA	SILANE	SULFUR	Pred Formula ABRASION	Pred Formula MODULUS
1	102	900	470	67.5	0.7	40	2.8	96.7475395	747.638719
2	120	860	410	65	1.7	40	1.8	113.421752	1074.97186
3	117	800	570	77.5	0.7	60	1.8	114.695993	1011.176
4	198	2294	240	74.5	1.7	60	2.8	195.496359	2173.94727
5	103	490	640	62.5	0.7	40	1.8	104.934463	865.669653
6	132	1289	270	67	1.7	40	2.8	133.734829	1333.44092
7	132	1270	410	78	0.7	60	2.8	138.00907	1310.64507
8	139	1090	380	70	1.7	60	1.8	143.683282	1497.9782
9	102	770	590	76	0.3835	50	2.3	101.492758	600.398164
10	154	1690	260	70	2.0165	50	2.3	155.361002	1476.17966
11	96	700	520	63	1.2	33.67	2.3	100.727591	525.749229
12	163	1540	380	75	1.2	66.33	2.3	159.12617	1330.8286
13	116	2184	520	65	1.2	50	1.4835	117.116503	1564.51067
14	153	1784	290	71	1.2	50	3.1165	152.737257	2020.06715
15	133	1300	380	70	1.2	50	2.3	139.119239	1261.13314
16	133	1300	380	68.5	1.2	50	2.3	139.119239	1261.13314
17	140	1145	430	68	1.2	50	2.3	139.119239	1261.13314
18	142	1090	430	68	1.2	50	2.3	139.119239	1261.13314
19	145	1260	390	69	1.2	50	2.3	139.119239	1261.13314
20	142	1344	390	70	1.2	50	2.3	139.119239	1261.13314
21	•	•	•	•	0.9	43	2	114.489091	931.321632

6. Select **Tables > Transpose**.
7. Select ABRASION Bags (500/0) and click **Transpose Columns**.
8. Click **OK**.
9. Select **Analyze > Distribution**.
10. Select Row 21 and click **Y, Columns**.

Note: Row 21 corresponds to the predictions from the new observation.

11. Click **OK**.
12. Click the red triangle next to Row 21 and select **Display Options > Horizontal Layout**.

Figure 3.36 Distribution Report



Quantiles

100.0%	maximum	158.36529239
99.5%		147.78319109
97.5%		127.92639742
90.0%		117.77494968
75.0%	quartile	114.09096311
50.0%	median	111.67631707
25.0%	quartile	109.93570255
10.0%		106.95588835
2.5%		100.84534774
0.5%		88.190714761
0.0%	minimum	87.132670178

Summary Statistics

Mean	112.30276
Std Dev	6.3936988
Std Err Mean	0.2859349
Upper 95% Mean	112.86455
Lower 95% Mean	111.74098
N	500

The Distribution Report in Figure 3.36 contains information about the distribution of the predicted values of ABRASION from each bagged model. The final prediction of ABRASION for the new observation is 112.3, which is the mean of all the *M* bagged predictions. This prediction has a standard error of 6.39. You can also create confidence intervals for the new prediction using the quantiles. For example, a 95% confidence interval for the new prediction is 100.85 to 127.93.

Example of Confidence Intervals Using Bagging

Use the bagged standard errors to obtain confidence intervals around predictions for a Neural Network model.

Fit Neural Network

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Analyze > Predictive Modeling > Neural**.
3. Select ABRASION and click **Y, Response**.
4. Select SILICA, SILANE, and SULFUR and click **X, Factor**.
5. Click **OK**.
6. (Optional) Enter 123456 next to Random Seed.

Note: Results vary due to the random nature of choosing a validation set in the Neural Network model. To reproduce the results shown in this example, set the Random Seed.

7. Click **Go**.

Perform Bagging and Launch the Profiler

1. Click the red triangle next to Model NTanH(3) and select **Profiler**.
2. Click the Prediction Profiler red triangle and select **Save Bagged Predictions**.
3. Enter 100 next to Number of Bootstrap Samples.
4. (Optional) Enter 123456 next to Random Seed.

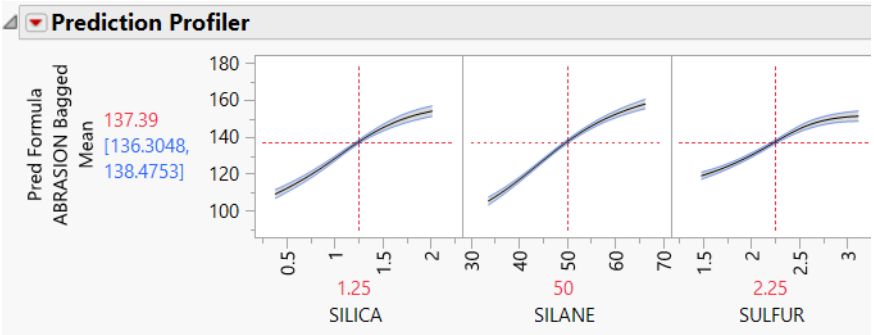
   Note: Results vary due to the random nature of sampling with replacement. To reproduce the results in this example, set the Random Seed.

5. Click **OK**.
6. Select **Graph > Profiler**.
7. Select Pred Formula ABRAISION Bagged Mean and StdError ABRASION Bagged Mean and click **Y, Prediction Formula**.
8. Select **Expand Intermediate Formulas**.
9. Click **OK**.

   A JMP Alert window appears asking if you want to use the StdError ABRASION Bagged Mean column to make intervals for the Pred Formula ABRAISION Bagged Mean column.

10. Click **Yes** in the JMP Alert window.

Figure 3.37 Profiler of Bagged Neural Predictions with Confidence Intervals



A profiler is created with confidence intervals using the values from the StdError ABRASION Bagged Mean column.

Example of Confidence Intervals Using StdErr Pred Formula Column

Use the saved Prediction Formula and StdErr Prediction Formula columns from a standard least squares model to create a prediction profiler with confidence intervals.

Fit the Model

Fit a standard least squares regression model using data on diabetes patients.

1. Select **Help > Sample Data Library** and open Diabetes.jmp.
2. Select **Analyze > Fit Model**.
3. Select Y and click **Y**.

   By default, the Personality is set to Standard Least Squares.
4. Select Age through Total Cholesterol and click **Add**.
5. Select Validation and click **Validation**.
6. Click **Run**.
7. Click the Response Y red triangle and select **Save Columns > Prediction Formula**.
8. Click the Response Y red triangle and select **Save Columns > StdErr Pred Formula**.

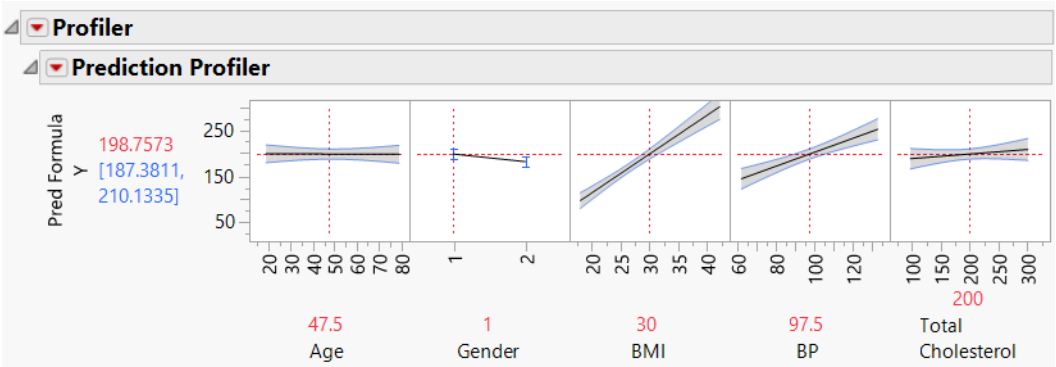
Launch the Profiler

The data table now contains two additional columns named Pred Formula Y and PredSE Y. Use these formula columns in the Profiler.

1. Select **Graph > Profiler**.
2. Select Pred Formula Y and PredSE Y and click **Y, Prediction Formula**.
3. Click **OK**.

   A JMP Alert window appears asking if you want to use the PredSE column to make intervals for the Pred Formula Y column.
4. Click **Yes** in the JMP Alert window.

Figure 3.38 Profiler with Confidence Intervals



A prediction profiler with confidence intervals constructed using the standard error of the predicted values is shown.

Statistical Details for the Prediction Profiler

- "Assess Variable Importance"
- "Propagation of Error Bars"
- "Extrapolation Control Metrics"

Assess Variable Importance

The details that follow relate to the how the variable importance indices are calculated.

Background

Denote the function that represents the predictive model by f, and suppose that $x_1, x_2, \ldots, x_n$ are the factors, or main effects, in the model. Let $y = f(x_1, x_2 \ldots, x_n)$.

- The expected value of y, $E(y)$, is defined by integrating y with respect to the joint distribution of $x_1, x_2, \ldots, x_n$.
- The variance of y, $Var(y)$, is defined by integrating $(y - E(y))^2$ with respect to the joint distribution of $x_1, x_2, \ldots, x_n$.

Main Effect

The impact of the main effect x_j on y can be described by $Var(E(y \mid x_j))$. Here the expectation is taken with respect to the conditional distribution of $x_1, x_2, \ldots, x_n$ given x_j, and the variance is taken over the distribution of x_j. In other words, $Var(E(y \mid x_j))$ measures the variation, over the distribution of x_j, in the mean of y when x_j is fixed.

It follows that the ratio $Var(E(y \mid x_j))/Var(y)$ gives a measure of the sensitivity of y to the factor x_j. The importance index in the Main Effect column in the Summary Report is an estimate of this ratio (see "Adjustment for Sampling Variation").

Total Effect

The Total Effect column represents the total contribution to the variance of $y = f(x_1, x_2 \ldots, x_n)$ from *all* terms that involve x_j. The calculation of Total Effect depends on the concept of functional decomposition. The function f is decomposed into the sum of a constant and functions that represent the effects of single variables, pairs of variables, and so on. These component functions are analogous to main effects, interaction effects, and higher-order effects. See Saltelli (2002); Sobol (1993).

Those component functions that include terms containing x_j are identified. For each of these, the variance of the conditional expected value is computed. These variances are summed. The sum represents the total contribution to $Var(y)$ due to terms that contain x_j. For each x_j, this sum is estimated using the selected methodology for generating inputs. The importance indices reported in the Total Effect column are these estimates (see "Adjustment for Sampling Variation").

Consider a simple example with two factors, x_1 and x_2. Then the Total Effect importance index for x_1 is an estimate of the following:

$$\frac{Var(E(y \mid x_1)) + Var(E(y \mid x_1, x_2))}{Var(y)}$$

Adjustment for Sampling Variation

Due to the fact that they are obtained using sampling methods, the Main Effect and Total Effect estimates shown in the Summary Table might have been adjusted. Specifically, if the Total Effect estimate is less than the Main Effect estimate, then the Total Effect importance index is set equal to the Main Effect estimate. If the sum of the Main Effect estimates exceeds one, then these estimates are normalized to sum to one.

Variable Importance Standard Errors

The standard errors that are provided for independent inputs measure the accuracy of the Monte Carlo replications. Importance indices are computed as follows:

- Latin hypercube sampling is used to generate a set of data values.
- For each set of data values, main and total effect importance estimates are calculated.
- This process is replicated until the estimated standard errors of the Main Effect and Total Effect importance indices for all factors fall below a threshold of 0.01.

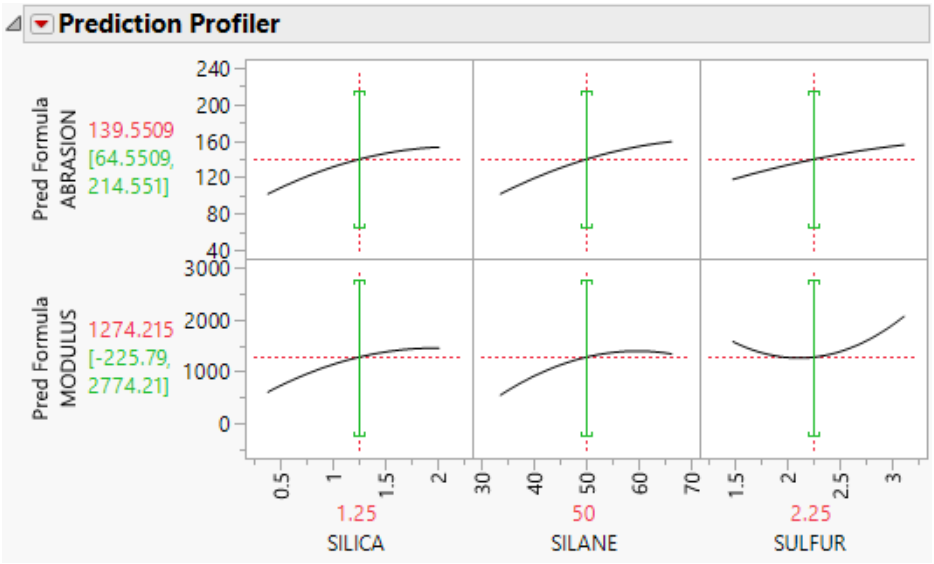
The standard errors that are reported are the standard error values in effect when the replications terminate.

Propagation of Error Bars

Propagation of error (POE) is important when attributing the variation of the response in terms of variation in the factor values when the factor values are not very controllable.

In JMP's implementation, the Prediction Profiler first looks at the factor and response variables to see whether there is a Sigma column property (a specification for the standard deviation of the column, accessed through the Cols > Column Info dialog box). If the property exists, then the Prop of Error Bars command becomes accessible in the Prediction Profiler drop-down menu. This displays the 3σ interval that is implied on the response due to the variation in the factor.

Figure 3.39 Green Propagation of Errors Bars in the Prediction Profiler



The POE is represented in the graph by a green bracket. The bracket indicates the prediction plus or minus three times the square root of the POE variance. The POE variance can be expressed as follows:

$$\sum_{i=1}^{N} \left(\sigma_{x_i}^2 \times \left(\frac{\partial f}{\partial x_i} \right)^2 \right) + \sigma_y^2$$

where σ_y is the user-specified sigma for the response column, and σ_x is the user-specified sigma for the factor column.

Currently, these partial derivatives are calculated by numerical derivatives:

centered, with δ=xrange/10000

POE limits increase dramatically in response surface models when you are over a more sloped part of the response surface. One of the goals of robust processes is to operate in flat areas of the response surface so that variations in the factors do not amplify in their effect on the response.

JMP PRO Extrapolation Control Metrics

Leverage

In models that are fit in the Standard Least Squares personality of the Fit Model platform, the leverage at the factor settings is used as the extrapolation metric.

The leverage of the i^{th} observation, h_{ii}, is the i^{th} diagonal entry of the matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, sometimes called the hat matrix. The leverage for a new prediction point is calculated as $h_{pred} = \mathbf{x}'_{pred}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{pred}$. The following two criteria can be used to determine if a prediction with leverage h_{pred} is an extrapolation:

- $h_{pred} > K \times \max(h_{ii})$, where K is a customizable multiplier
- $h_{pred} > L \times p/n$, where L is a customizable multiplier, p is the number of variables, n is the number of observations, and p/n is the average leverage

You can use the Set Threshold Criterion option to specify which criterion is used and the value of the multiplier. The default values of the multipliers are $K = 1$ and $L = 3$.

Note: Extrapolation control on profilers run from the graph menu using a saved least squares model do not implement the leverage methodology. Instead, the Regularized Hotelling's T^2 methodology is used.

Regularized Hotelling's T^2

In models other than least squares models, the Regularized Hotelling's T^2 value is used as the extrapolation metric. The T^2 value for the training data and T^2 values for the prediction points are calculated as follows:

$$T^2 = (x - \bar{x})' \hat{\Sigma}^{-1} (x - \bar{x})$$

$$T^2_{\text{pred}} = (x_{\text{pred}} - \bar{x})' \hat{\Sigma}^{-1} (x_{\text{pred}} - \bar{x})$$

where $\hat{\Sigma}$ is the Schafer and Strimmer regularized covariance matrix estimator estimated on the training data. The target matrix used for the Schafer Strimmer estimator is a diagonal covariance matrix. See Schafer and Strimmer (2005). In platforms that train models using observations with missing values, the covariance matrix is estimated with pairwise deletion.

Note: Categorical variables are converted to indicator variables for these calculations.

The calculation of the threshold depends on the number of nonmissing T^2 values computed on the training data.

- If there are ten or more nonmissing T^2 values, the threshold is set as follows:

  $$\overline{T^2} + K\hat{\sigma}_{T^2}$$

  where

  K is a customizable multiplier and is set to 3 by default

  $\hat{\sigma}_{T^2}$ is the standard deviation of the T^2 values.

- If there are less than ten nonmissing T^2 values, the threshold is set using an F distribution quantile equivalent to a $K\sigma$ limit.

  $$Q_{T^2}(q)$$

  where

  $q = \Phi(K)$

  $\Phi(\cdot)$ is the standard normal distribution

  K is a customizable multiplier and is set to 3 by default

$$T^2 \sim \frac{(n+1)(n-1)p}{n(n-p)} F(p, n-p)$$

p is the number of parameters

n is the number of nonmissing T^2 values

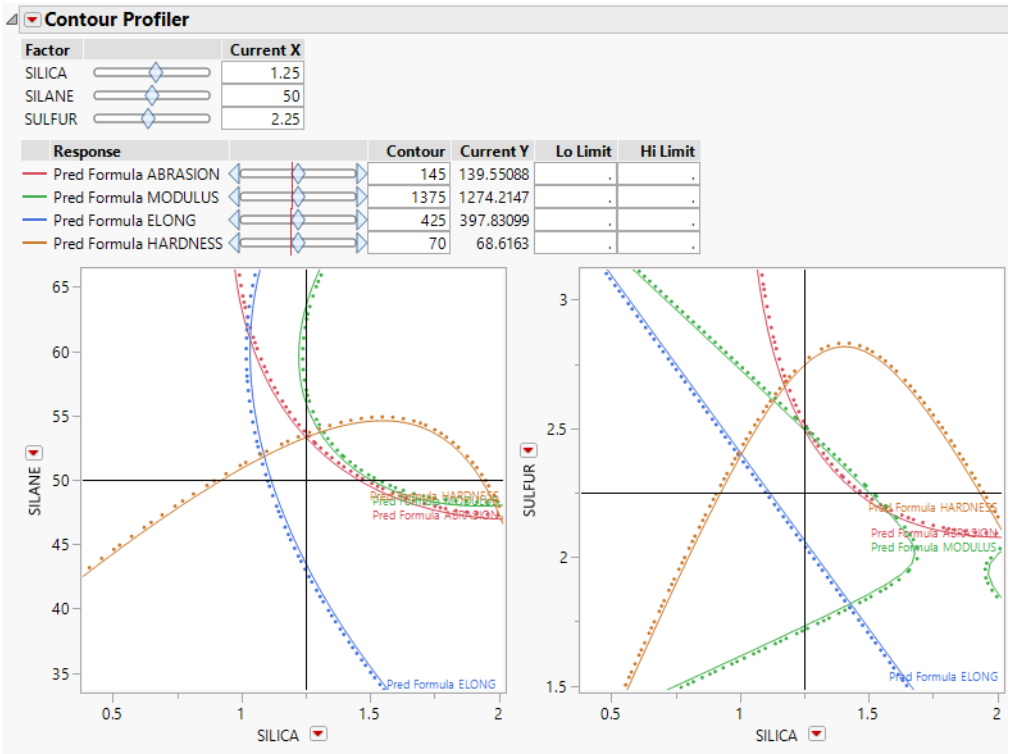
Chapter 4

Contour Profiler

Explore Contours of Responses across Two Factors

The Contour Profiler shows plots of response contours for two factors at a time. Multiple contour plots for different combinations of factors can be viewed at the same time. The interactive contour profiling facility is useful for optimizing response surfaces graphically.

Figure 4.1 Contour Profiler Example



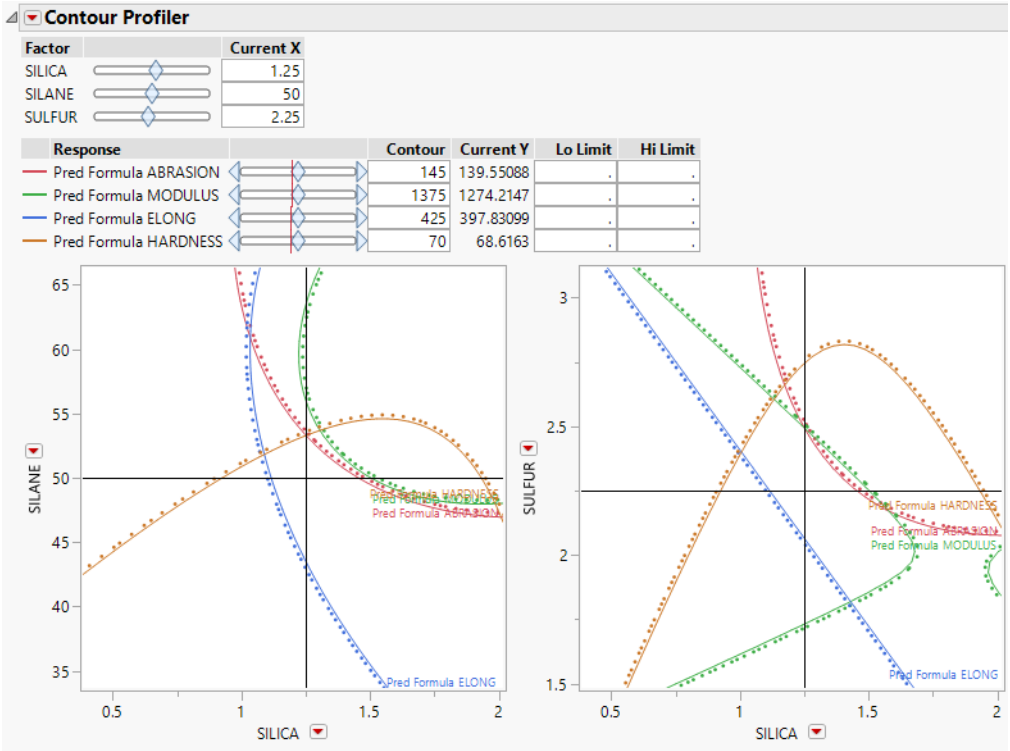
Contents

Overview of the Contour Profiler 93
Example of the Contour Profiler 93
Launch the Contour Profiler Platform 95
The Contour Profiler Report 95
Factor Settings and Controls 96
Response Settings and Controls 96
Contour Profiler Platform Options 97
Constraint Shading Settings 99
Additional Example of the Contour Profiler 99
Explore Optimal Settings 99

Overview of the Contour Profiler

The Contour Profiler shows plots of response contours for two factors at a time. Multiple contour plots for different combinations of factors can be viewed at the same time. Factors can be set to specific values to show how this affects the contours. The interactive contour profiling facility is useful for optimizing response surfaces graphically. Separate surface plots for each response are also available. Figure 4.2 shows an example of the Contour Profiler for the Tiretread.jmp sample data.

Figure 4.2 Contour Profiler



Example of the Contour Profiler

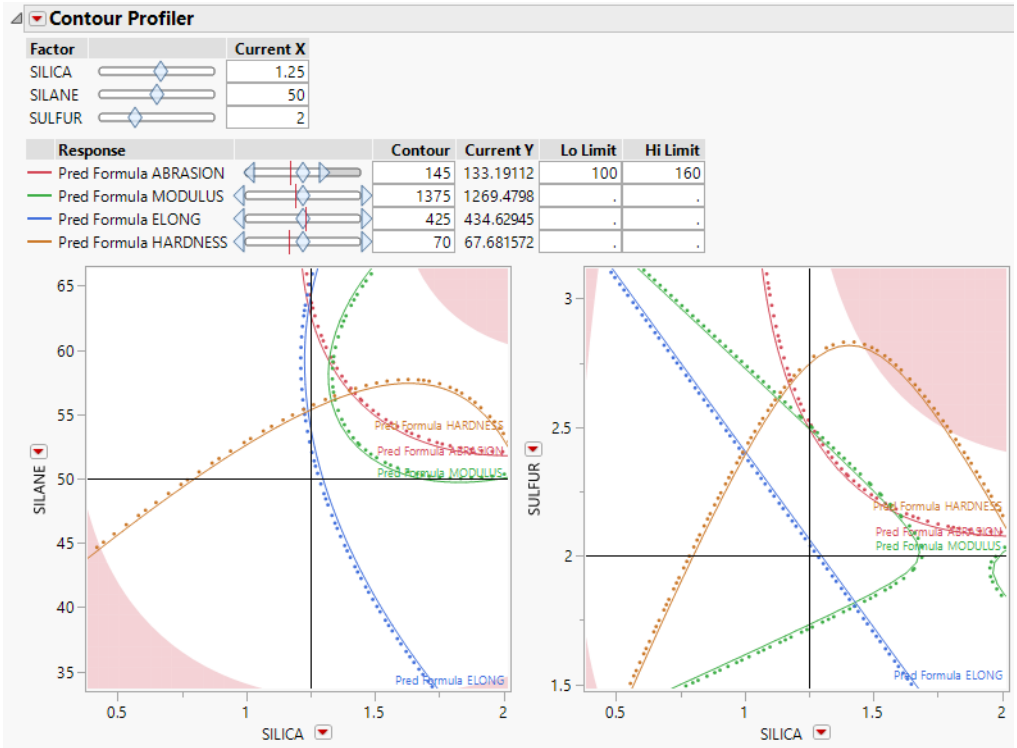
This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). This data table already contains saved prediction formula columns for the four response variables.

1. Select **Help > Sample Data Library** and open Tiretread.jmp.

2. Select **Graph > Contour Profiler**.
3. Select Pred Formula ABRASION, Pred Formula MODULUS, Pred Formula ELONG, and Pred Formula HARDNESS and click **Y, Prediction Formula**.
4. Click **OK**.
5. Click the Contour Profiler red triangle and select **Multiple Contour Frames**.
6. Under Horizontal Factor, select SILICA and under Vertical Factor, select SULFUR.
7. Click **OK**.

   On both contour plots, the values for SILICA are on the horizontal axis. The values for SILANE and SULFUR are on the vertical axes. In the first contour plot, the value for SULFUR is fixed at 2.25. In the second contour plot, the value for SILANE is fixed at 50 (Figure 4.2).
8. Click the **Current X** box for SULFUR and type 2. Click anywhere outside the box to set the value.
9. In the **Lo Limit** box for Pred Formula ABRASION, type 100.
10. In the **Hi Limit** box for Pred Formula ABRASION, type 160.
11. Click anywhere outside the box to update the plot.

Figure 4.3 Contour Profiler for Tiretread.jmp



The contour lines in the first contour plot differ from those in Figure 4.2 because SULFUR is set at a different value. This can be seen in the placement of the horizontal line at 2 in the second contour plot. Move the slider bar for SULFUR to see how different values further change the contour lines in the first contour plot. Move the slider bar for SILANE to see how different values change the contour lines in the second contour plot. The shaded area on the contour plots represent the regions of the graphs that are not feasible, based on the Lo Limit and Hi Limit values for ABRASION.

Launch the Contour Profiler Platform

The Contour Profiler can be accessed in the following ways:

- The Contour Profiler can be accessed directly from the Graph menu. When you access the Contour Profiler in this way, the Contour Profiler launch window appears. See "Profiler Launch Windows".
- The Contour Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information about the availability of the Contour Profiler in different platforms.
- The Contour Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Contour Profiler from the Profiler red triangle menu.
- JMP PRO The Contour Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Contour Profiler from the Profiler red triangle menu.

The Contour Profiler Report

The initial Contour Profiler report shows a contour profiler plot, factor settings and controls, and response settings and controls. The colored lines on the graph are the contours for the responses set by the Y slider controls or by entering values in the Contour column. There is a separately colored contour for each response. For each contour, there is a dotted line in the direction of higher response values, so that you get a sense of direction. To change the factor assignments to the horizontal and vertical axes of a contour plot, click the red triangle on the axis and select a factor. If you select the factor that is on the opposite axis, the factor assignments are interchanged.

- "Factor Settings and Controls"
- "Response Settings and Controls"

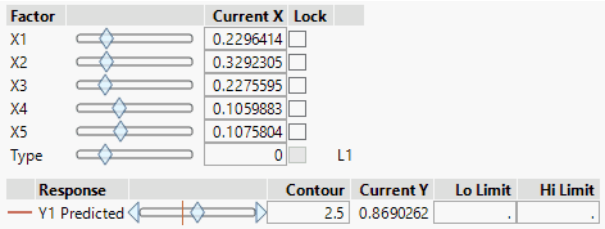
Factor Settings and Controls

Factor The list of factors.

Current X The current factor settings. Click in a box to change the value of a factor or use the slider controls. Right-click the slider control and select **Rescale Slider** to change the scale of the slider. When one value is changed, the values for the other factors remain the same, but the values in the response settings and controls report change accordingly. The Current X location is shown by the crosshair lines on the graph.

Lock A column that appears for mixture designs with more than three predictors. This column enables you to lock settings for mixture values so that they are not changed when the mixture needs to be adjusted due to other mixture effects being changed. When locked columns exist, the shaded area for a mixture recognizes the newly restricted area.

Figure 4.4 Lock Column



Note: If one or more of the factors have specification limits, a dotted rectangle is drawn on the contour plot to show the joint spec limit range.

Response Settings and Controls

Response The list of one or more responses, with a color legend. To change the color of a response, right-click the color legend for that response.

Contour The current value of the contour on the contour profiler. Click in the box to change the value. The slider control can also be used to change the value of the contour.

Current Y The predicted response based on the current X settings. The value is at the red line on the slider control. This value updates as the factor settings are changed.

Lo Limit Enables you to set a lower limit for your response. Click in a box to set the limit or click the left triangle of the slider control for that response. If a response column's Spec Limits column property has a value for Lower Spec Limit, that value is used as the initial value for Lo Limit.

High Limit Enables you to set an upper limit for your response. Click in a box to set the limit or click the right triangle of the slider control for that response. If a response column's Spec Limits column property has a value for Upper Spec Limit, that value is used as the initial value for Hi Limit.

When response limits are used, the profiler shows regions that are not feasible as shaded in the contour profiler plot.

Note: If your response has a Response Probability column property, the High Limit is automatically set to 0.5. In this case, the shaded region is the region where that response level is predicted.

Contour Profiler Platform Options

Grid Density Sets the density of the mesh plots in the surface plots.

Graph Updating Contains options to change the frequency of updates for the graph. (The difference might not be noticeable on a fast machine.) The following two options are available:

Per Mouse Move Updates the graph continuously as you drag the mouse. This is the default setting.

Per Mouse Up Updates the graph each time the mouse is released.

Surface Plot Shows or hides the mesh plots.

Contour Label Shows or hides a label for the contour lines. The label colors match the contour colors.

Contour Grid Draws contours on the Contour Profiler plot at intervals that you specify.

Remove Contour Grid Enables you to remove a contour grid once one is drawn.

Factor Settings Provides a submenu of commands that enables you to save and transfer the Contour Profiler's settings to other parts of JMP. See "Factor Settings".

Simulator Launches the Simulator. See "Simulator".

Up Dots Shows or hides dotted lines that correspond to each contour. The dotted lines show the direction of increasing response values.

Set Contours to Current Resets the contour lines to be where the current Y values are located. This means that all contour lines cross where the crosshairs are on the contour plot and the controls agree in the Y sliders.

Arrange X Controls Left Rearranges the X and Y controls horizontally with the X controls on the left or vertical with the X controls at the top.

Hide X Controls Shows or hides the X controls (Factor Settings and Controls).

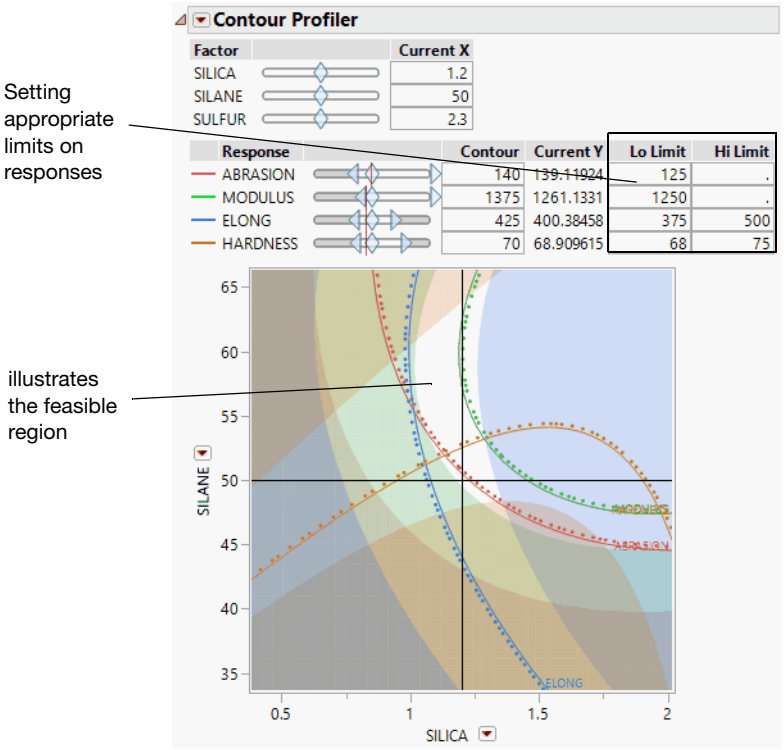
Hide Y Controls Shows or hides the Y controls (Response Settings and Controls).

Multiple Contour Frames Enables you to add one or more contour plots to the report that represent different combinations of factor settings. When you select this option, the Multiple Contour Frames dialog appears. Use the dialog to specify additional individual contour plots or a matrix of contour plots. You can specify a lower triangular, upper triangular, or full square matrix of contour frames. By default, the new contour plots will be added to the current contour plots in the report. Click the **Remove previous frames** check box to remove the current contour plots and replace them with your new selections.

Constraint Shading Settings

If you specify limits for the response columns, the areas of the plot that are outside of the feasible region are shaded. The unshaded white area represents the feasible region.

Figure 4.5 Settings for Contour Shading



Additional Example of the Contour Profiler

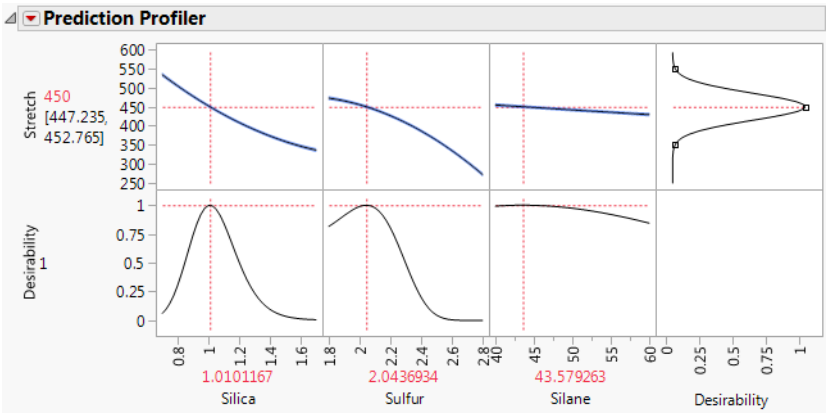
Explore Optimal Settings

Desirability functions are often used to determine the optimal settings for a response surface. The Maximize Desirability feature in the Prediction Profiler provides one combination of factor level settings that results in a predicted response that optimizes the desirability. However, there are often many factor level combinations that can optimize the desirability function. The Contour Profiler is a useful tool for finding alternative settings that optimize desirability.

This example uses the Bounce Data.jmp data table, which contains data for a tire tread experiment. The response, Stretch, is a function of three variables, Silica, Silane, and Sulfur. A model is built and the optimal settings are determined such that the value of Stretch maximizes the desirability function using the Maximize Desirability option in the Prediction Profiler. You use the Contour Profiler to explore other optimal settings for the specified value of Stretch.

1. Select **Help > Sample Data Library** and open Design Experiment/Bounce Data.jmp.
2. Click the green triangle next to the **Model** script, and then click **Run** in the launch window.
3. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability**.

Figure 4.6 Maximize Desirability Settings

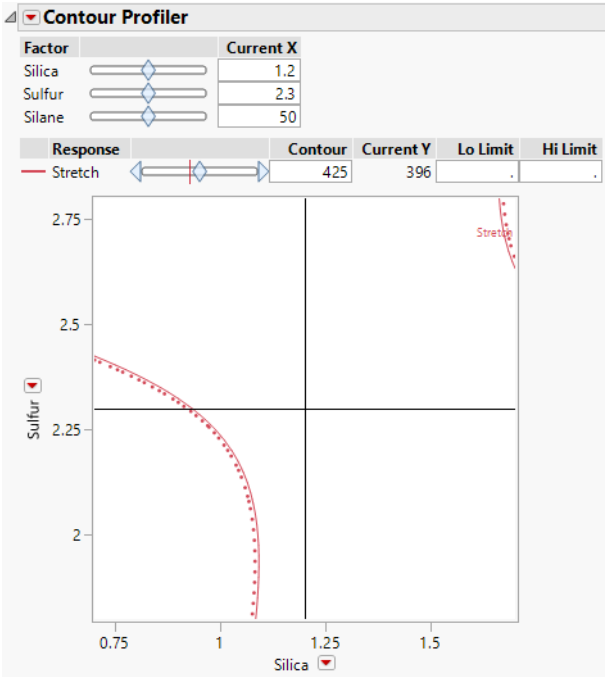


The value of Stretch that maximizes the desirability function is 450. One combination of factor level settings that results in a predicted Stretch of 450 is Silica = 1.0101, Sulfur = 2.0437, and Silane = 43.5792.

4. Click the Response Stretch red triangle and select **Factor Profiling > Contour Profiler**.

   To ensure that your new setting combinations maximize desirability, you want to make sure that the settings predict Stretch within 2 units of 450. In addition, you would like a high level of Silane, a fixed value of 60.

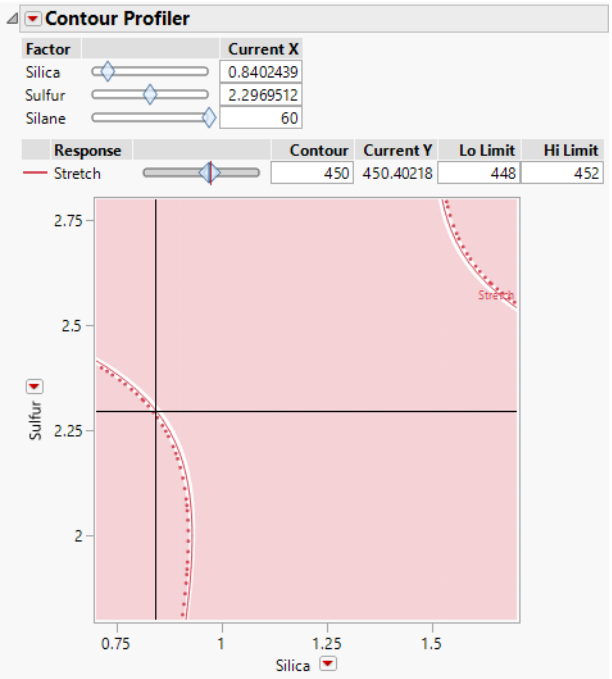
Figure 4.7 Contour Profiler for Bounce Data.jmp



The initial plot shows the contour values of Silica and Sulfur for Stretch at 425 and Silane at 50.

5. In the response controls, set the **Contour** value for Stretch to 450. Set the **Lo Limit** and **High Limit** for Stretch to 448 and 452, respectively. Press Enter.
6. Set the **Current X** for Silane to 60.

Figure 4.8 Contour Profiler for Optimal Settings



The values on the solid red curve are the Silica and Sulfur values that predict Stretch to be 450. The values within the unshaded band predict values of Stretch between 448 and 452. You can drag the crosshairs that appear on the plot to the unshaded sections to find specific settings for Silica and Sulfur. One such setting is seen in Figure 4.8.

Chapter 5

Surface Plot

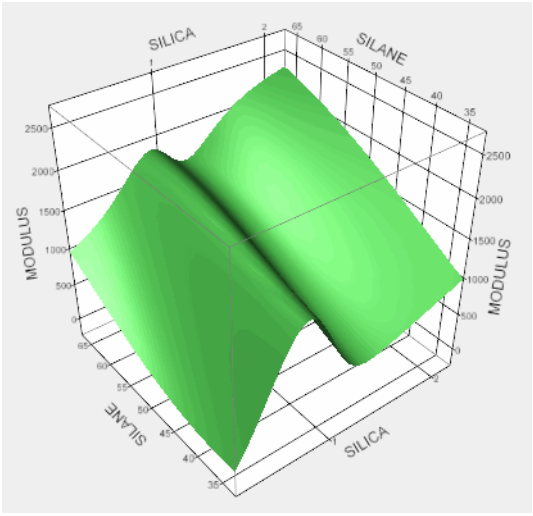
Explore Contours of Responses across Three Factors

A *surface plot* is a three-dimensional plot with one or more dependent variables represented by a smooth surface. You can display up to four dependent variables on the same surface plot.

JMP produces surface plots in the following two situations:

- The Surface Plot platform creates a stand-alone report that contains a surface plot for formulas. The formulas can be formula columns in your data table or mathematical formulas that do not involve any data points.
- The Surface Profiler option in many model fitting platforms produces a surface plot for the fitted model in the existing platform report.

Figure 5.1 Example of a Surface Plot



Contents

Overview of the Surface Plot and Surface Profiler . . . 105
Example of the Surface Plot Platform . . . 105
Launch the Surface Plot Platform . . . 108
The Surface Plot Report . . . 110
The Surface Plot . . . 110
Appearance Controls . . . 111
Independent Variables . . . 112
Dependent Variables . . . 113
Surface Plot Platform Options . . . 115
Pop-Up Menu Options . . . 115
Dependent Variables Options . . . 116
Sheet or Surface Properties . . . 117
Keyboard Shortcuts . . . 119
Additional Examples of the Surface Plot Platform . . . 119
Construct a Surface Plot for a Single Mathematical Function . . . 119
Plotting Isosurfaces . . . 120

Overview of the Surface Plot and Surface Profiler

The Surface Plot and Surface Profiler platforms are used to plot points and surfaces in three dimensions.

Surface plots are available as a separate platform (Graph > Surface Plot) and as options in many reports (known as the Surface Profiler). The surface plot functionality is similar in both situations.

The plots can be of points or surfaces or both. When the points are visible in the surface plot, the points are linked to the data table. You can click the points or use the Brush tool to select rows in the data table. The colors and markers assigned in the data table are also reflected in the surface plot.

Surfaces can be defined by a mathematical equation, or through a set of points defining a polygonal surface. Each surface can be displayed as a smooth surface or as a mesh, with or without contour lines. The labels, axes, and lighting for the surface plot are fully customizable.

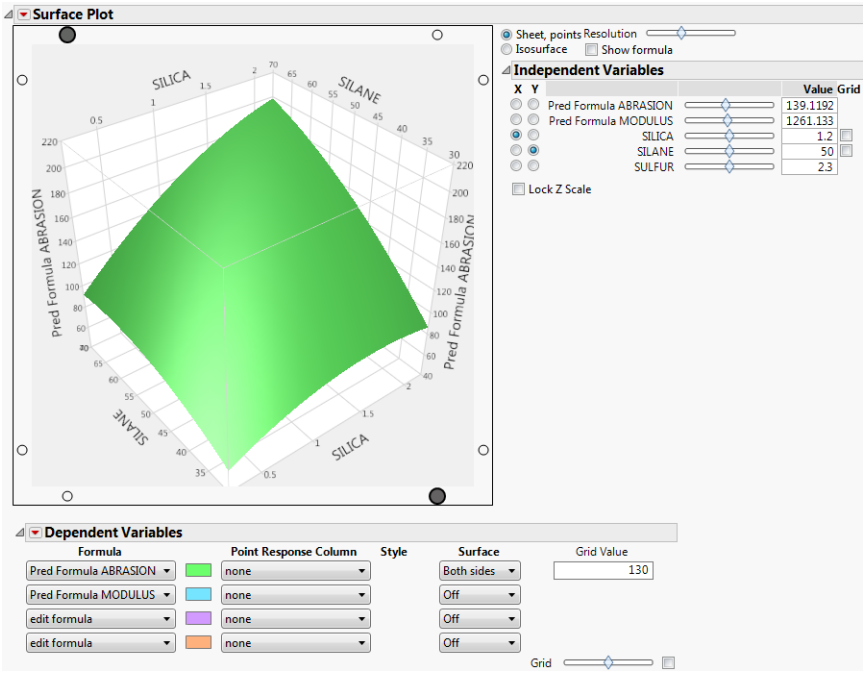
The Surface Plot is built using the 3D scene commands from the JMP Scripting Language (JSL). When you right-click in the surface plot, the pop-up menu contains Open GL Scene commands. For more information about the Open GL Scene commands, see the *Scripting Guide*.

Example of the Surface Plot Platform

This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). This data table already contains saved prediction formula columns for the four response variables. Use these prediction formula columns to examine the surfaces of two of the response variables.

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Graph > Surface Plot**.
3. Select Pred Formula ABRASION and Pred Formula MODULUS, click **Columns**, and then click **OK**.

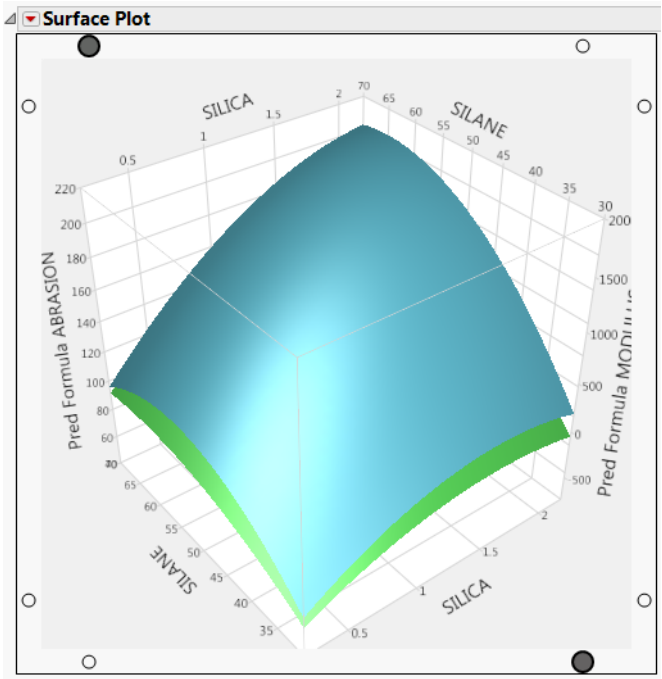
Figure 5.2 Initial Surface Plot Report



Initially, only the prediction surface for ABRASION appears in the plot. Note that in the Independent Variables report, there is a large difference between the values for ABRASION and MODULUS. The two responses are on different scales.

4. In the Dependent Variables report, select **Both Sides** from the Surface list for Pred Formula MODULUS.
5. Click the Surface Plot red triangle and select **Scale response axes independently**.

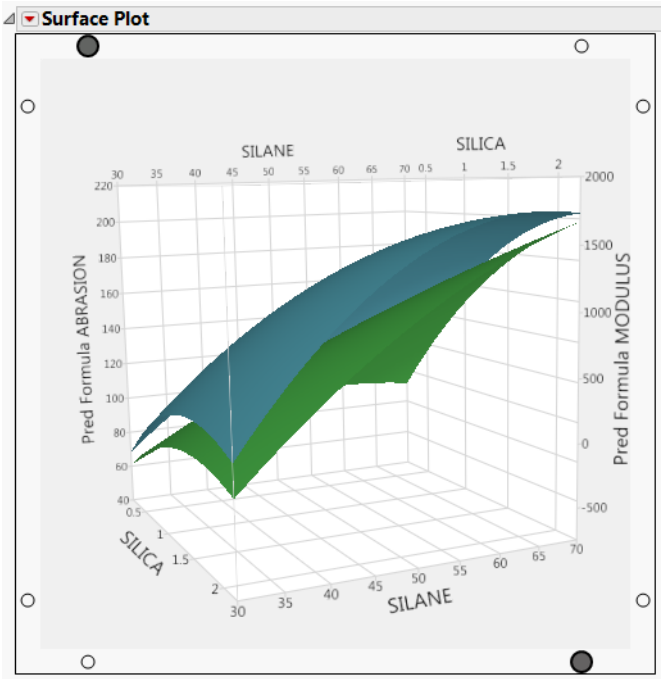
Figure 5.3 Prediction Surface for ABRASION and MODULUS



Pred Formula ABRASON is on one Z axis and Pred Formula MODULUS is on the other Z axis. Each response has its own scale. In this plot, the two factors considered are SILICA and SILANE. You can rotate the surface plot by clicking and dragging inside the graph. This enables you to view the prediction surfaces from different perspectives. For example, in Figure 5.4, you can see that predictions for ABRASION and MODULUS based on the two listed predictors have a similar surface shape, just on a different scale. Use the Independent Variables controls to see how different combinations of predictors affect the surface plot.

Tip: To see the original view after rotating, right-click anywhere on the surface plot and select **Reset**.

Figure 5.4 Rotated Prediction Surfaces

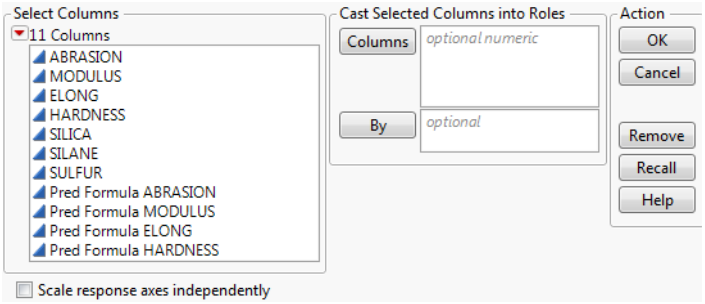


Launch the Surface Plot Platform

The Surface Plot Profiler can be accessed in the following ways:

- To launch the Surface Plot Profiler platform directly, select Surface Plot from the Graph menu. If there is a data table open, this displays the window in Figure 5.5. If you do not want to use a data table for drawing surface plots, click OK without specifying any columns. If there is no data table open, you are presented with the default surface plot shown in Figure 5.13.
- The Surface Plot Profiler can be accessed as a red triangle menu option in many modeling platforms. In the red triangle menus, it is referred to as Surface Profiler. See "Where to Find JMP Profilers" for more information about the availability of the Surface Profiler in different platforms.
- The Surface Plot Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Surface Profiler from the Profiler red triangle menu.
- JMP PRO The Surface Plot Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Surface Profiler from the Profiler red triangle menu.

Figure 5.5 Surface Plot Launch Window



Columns Specifies the columns that you want to plot. Only numeric variables can be assigned to the Columns role.

By Specifies a variable to create separate surface plots for each level of the variable.

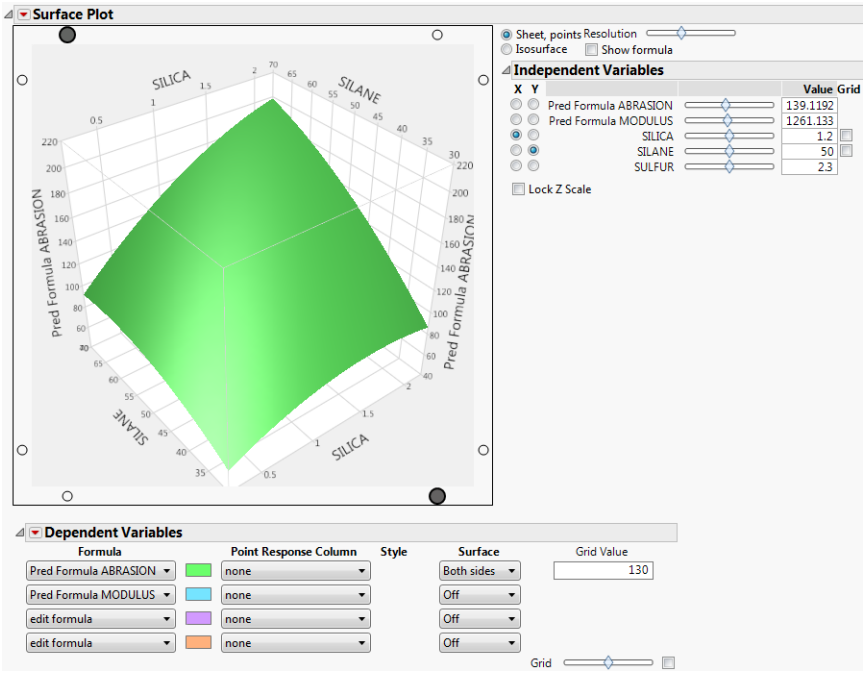
Scale response axes independently Assigns a separate scale to each response on the surface plot.

Note: When not selected, the axis scale for all responses is the same as the scale for the first item entered in the Columns role.

The Surface Plot Report

The initial Surface Plot report shows the surface plot, the appearance controls, the Independent Variables controls, and the Dependent Variables controls. If more than one prediction formula column is specified, only the first surface is displayed in the initial report. However, if more than one observed response column is specified, all points are displayed in the initial report.

Figure 5.6 Example of the Surface Plot Report



The Surface Plot

The surface plot has the following controls and settings:

Rotate Click anywhere that the cursor appears with a circular arrow icon on the surface plot and drag to rotate the plot in any direction.

Note: The Up, Down, Left, and Right arrow keys can also be used to rotate the plot.

Axis Settings Double-click an axis to reveal the axis settings window. The axis settings window enables you to change the scale, tick mark increment, and axis label settings. Like

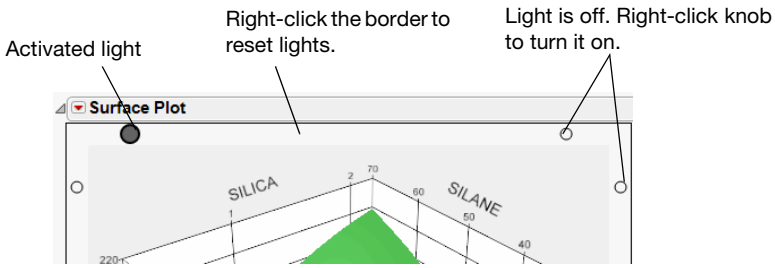
other JMP graphs, the axes can be adjusted, stretched, and compressed using the grabber tool. Hover over an axis to use the grabber tool.

Lights By default, the plot has lights shining on it. There are eight control knobs on the plot for changing the position and color of the lights. This is useful for highlighting different parts of a plot and creating contrast. Four of the eight knobs are shown in Figure 5.7.

You can perform the following actions to adjust the lights shining on the surface:

- Right-click a knob to turn that light on or off. More lights turned on brighten a plot, and fewer lights darken it.
- Drag a knob to change the position of a light.
- Change the color of a light by right-clicking on the knob. The default color is white.

Figure 5.7 Control Knobs for Lights



Appearance Controls

If you select the Surface Profiler option from the red triangle menu of a modeling platform, the Appearance Control options are different than the Appearance Controls in the stand-alone Surface Plot platform. See Figure 5.8 for an example of what the Appearance Controls look like in a modeling platform.

Surface Plot Appearance Controls

The set of controls in the top right of the report enables you to specify the overall appearance of the surface plot.

Sheet, points The setting for displaying sheets, points, and lines.

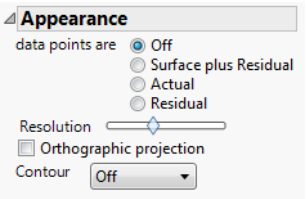
Isosurface Changes the display to show isosurfaces. See "Plotting Isosurfaces".

Show formula Shows the formula edit box, which enables you to edit the formula of the surface. The formula edit box appears only if you specified a column that contains a formula in the launch window.

Resolution A slider that affects how many points are evaluated for a formula. If the resolution is too coarse, a function with a sharp change might not be represented very well. However, if the resolution is set too high, evaluating and displaying the surface is slower than for lower resolutions.

Surface Profiler Appearance Controls

Figure 5.8 Appearance Controls



The Appearance Controls in modeling platforms contain the following controls:

data points are Contains the following options that control the appearance of the data points:

Off Turns the data points off.

Surface plus Residual Shows the difference between the predicted value and actual value on the surface.

Actual Shows the actual data points.

Residual Shows the residual values (if they are not off the plot).

Resolution A slider that affects how many points are evaluated for a formula. If the resolution is too coarse, a function with a sharp change might not be represented very well. However, if the resolution is set too high, evaluating and displaying the surface is slower than for lower resolutions.

Orthogonal Projection Changes the view of the graph, such that the axes form right angles.

Contour Determines the placement of the contour lines on the plot in relation to the surface. Options are off, below, above, and on surface.

Independent Variables

The Independent Variables report contains the following controls:

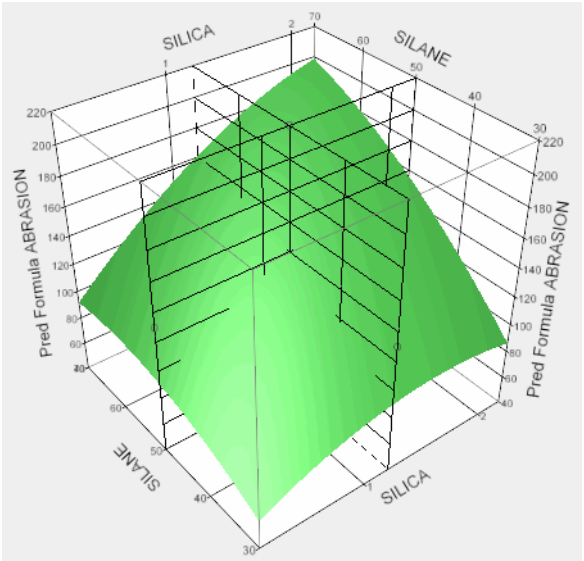
X, Y Radio buttons that enable you to select which independent variables are displayed on the *x*- and *y*-axes. This feature is useful when you have more than two independent variables.

Value The sliders and text boxes set the current values of each variable. These settings are most important for the variables that are not displayed on the axes. The plot shows the three-dimensional slice of the surface at the value shown in the text box. Move the slider to see different slices.

Lock Z Scale Locks the Z axis to its current values. This is useful when moving the sliders that are not on an axis.

Grid Selecting a check box activates a grid that is perpendicular to the selected variables axis. The sliders enable you to adjust the placement of each grid. The resolution of each grid can be controlled by adjusting axis settings. Figure 5.9 shows a surface with the X and Y grids activated.

Figure 5.9 Activated X and Y Grids



Dependent Variables

The Dependent Variables controls that are available depend on whether you have selected Sheet, points or Isosurface in the Appearance Controls.

Controls for Sheet and Points

The Dependent Variables controls that appear by default when Sheet, points is selected are shown in Figure 5.6.

Formula Enables you to select one or more formulas that are displayed as surfaces in the plot.

Point Response Column Enables you to select the column that contains values that are plotted as points.

Style (Menus appear only after you have selected a Point Response Column.) The Style menu enables you to choose how the points are displayed. The menu has the following options:

Points Shows individual points, which change according to the color and marker settings of the row in the data table.

Needles Draws lines from the *x*-*y* plane to the points, or, if a surface is also plotted, connects the surface to the points.

Mesh Connects the points into a triangular mesh.

Surface Overlays a smooth, reflective surface on the points.

Off Points are not shown at all.

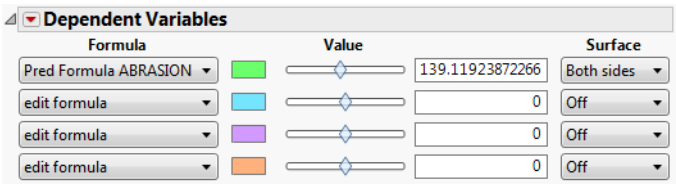
Surface Enables you to show or hide the top or bottom of a surface. If **Above only** or **Below only** is selected, the opposite side of the surface is darkened.

Grid, Grid Value Provides a slider and check box that activate a grid for the dependent variable. You can use the slider to adjust the value at which the grid is drawn. You can also enter the value into the Grid Value text box above the slider.

Controls for Isosurface

Most of the controls for Isosurface are identical to those of Sheet, points. Figure 5.10 shows the default controls, illustrating the slightly different presentation.

Figure 5.10 Dependent Variable Controls for Isosurfaces



Value Slider and text box that activate an isosurface for the dependent variable. You can use the slider to adjust the value at which the isosurface is drawn. You can also enter the value into the text box next to the slider.

Surface Plot Platform Options

The Surface Plot red triangle menu contains the following options:

Control Panel Shows or hides the Control Panel, which includes the controls for Appearance, Independent Variables, and Dependent Variables.

Scale response axes independently Scales response axes independently. See "Launch the Surface Plot Platform".

Fit to Window Determines whether the plot is resized as you resize the report window.

Auto (default) Bases the scaling on the contents of the plot. For example, a plot with By variables or used as a Surface Profiler in a Fit Model platform does not stretch to fit the resized window; the plot extends beyond the viewing area.

On Always fits the plot inside the window.

Off Prevent the plot from resizing.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Pop-Up Menu Options

You can right-click in the plot area to reveal the following pop-up menu options:

Sheet Properties (Option appears only if you right-click the sheet.) Displays the Sheet Properties window. Available only if Sheet, points is selected in the Appearance Controls. See "Sheet or Surface Properties".

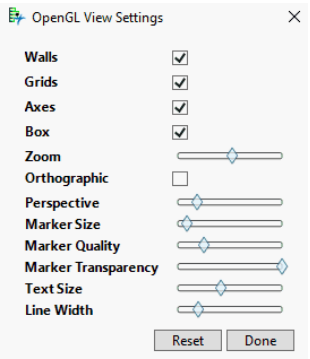
Surface Properties (Options appears only if you click the surface.) Displays the Surface Properties window. Available only if Isosurface is selected in the Appearance Controls. See "Sheet or Surface Properties".

Show Legend Shows or hides a legend.

Reset Resets the plot to the original viewpoint. Changes in wall and background color are not affected.

Settings Opens a window for changing many plot settings. For more information about the OpenGL Scene commands, see the *Scripting Guide*.

Figure 5.11 OpenGL View Settings



Hide Lights Border Shows or hides lighting controls.

Wall Color Enables you to change the plot wall color.

Background Color Enables you to change the plot background color.

Rows Enables you to change row colors or markers, and also exclude, hide, and label points.

Use Hardware Acceleration Provides for faster rendering of the display. For example, if the plot redraws slowly when rotating, this option can help it redraw faster.

Show ArcBall Provides options for using the ArcBall. The ArcBall is a sphere drawn around the plot to help visualize the directions of rotation.

Dependent Variables Options

The Dependent Variables red triangle menu provides the following options:

Formula Shows or hides the Formula column in the Dependent Variables controls.

Surface Shows or hides the Surface column in the Dependent Variables controls.

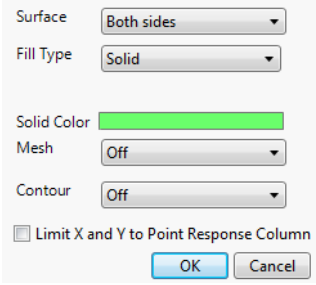
Points Shows or hides the Point Response Column column in the Dependent Variables controls.

Response Grid Shows or hides the Grid controls.

Sheet or Surface Properties

If you have selected the Sheet, points option in the Appearance Controls, you can right-click the sheet inside the plot and select **Sheet Properties** to reveal a window that enables you to change the sheet properties. If you are plotting an Isosurface, right-click the surface and select **Surface Properties** to reveal a similar window. The Sheet Properties and Surface Properties windows enable you to modify the appearance of the surface plot, including the surface color, opacity, and contours. You can also show or hide a mesh that appears on the surface plot.

Figure 5.12 Sheet Properties Window



Surface Enables you to show or hide the top or bottom of a surface. If Above only or Below only is selected, the opposite side of the surface is darkened.

Fill Type Enables you to color the surface using a solid color, or continuous or discrete gradients.

Solid Color (Available only when Solid is selected for Fill Type.) Enables you to choose a color for the surface.

Surface Fill (Available only when a gradient is selected for Fill Type.) Specifies the dependent variable's surface to which the fill type is applied. When Continuous Gradients is selected for Fill Type, you can also choose the Custom option for Surface Fill.

Custom (Available only when Continuous Gradients is selected for Fill Type and when Custom is selected for Surface Fill.) Enables you to specify an equation that defines a response value. This equation is specified using JSL and can refer to columns in the data table using column scoping.

Surface Color Theme (Available only when a gradient is selected for Fill Type.) Enables you to change the color theme for the surface or define a custom color theme. For more information about color themes, see *Using JMP*.

Gradient Levels (Available only when Discrete Gradients is selected for Fill Type.) Enables you to specify the number of different colors used on the surface plot.

Surface Color Range (Available only when a gradient is selected for Fill Type.) Enables you to choose the endpoints for the color gradient. If you choose Data, the endpoints of the color gradient are determined by the range of the data in the column selected for Surface. If you choose Axis, the endpoints of the color gradient are determined by the range of the axis for the dependent variable.

Caution: If you choose Axis for Surface Color Range and change the endpoints of the axis in the surface plot, the cutoff values for the color gradient change as well.

Lighting Enables you to change the lighting of the surface plot. Choose between None, Low Reflection, and Normal.

Mesh Enables you to turn on or off a surface mesh, for either the *X* or *Y* directions or both.

Mesh Color (Available only when a value other than Off is selected for Mesh.) Enables you to select the color for the surface mesh.

Contour Enables you to turn on or off a contour grid, either above, below, or on the surface. If turned on, the Contour Color option is revealed enabling you to change the color.

Contour Color (Available only when a value other than Off is selected for Contour.) Enables you to select the color for the contour grid.

Limit X and Y to Point Response Column Enables you to limit the range of the plot to the range of the data in the Point Response Column, if one is activated. If checked, this option essentially restricts the plot from extrapolating outside the range of the data in the Point Response Column.

Note: The equivalent JSL command for this option is `Clip Sheet( Boolean )`. You can send this message to a particular response column by appending the number of the response column. For example, `Clip Sheet2( 1 )` limits the range of the plot to the range of the data of the second response column. See the Scripting Index in the JMP Help menu for an example.

Keyboard Shortcuts

The following keyboard shortcuts can be used to manipulate the surface plot.

Table 5.1 Surface Plot Keyboard Shortcuts

Key	Function
left, right, up, and down arrows	spin
Home, End	diagonally spin
Enter (Return)	toggles ArcBall appearance
Delete	roll counterclockwise
Control	boost spin speed 10X
Shift	allows continual spinning

Additional Examples of the Surface Plot Platform

- "Construct a Surface Plot for a Single Mathematical Function"
- "Plotting Isosurfaces"

Construct a Surface Plot for a Single Mathematical Function

You can use the Surface Plot platform to produce the graph of a mathematical function without any data points.

1. Select **Graph > Surface Plot**.

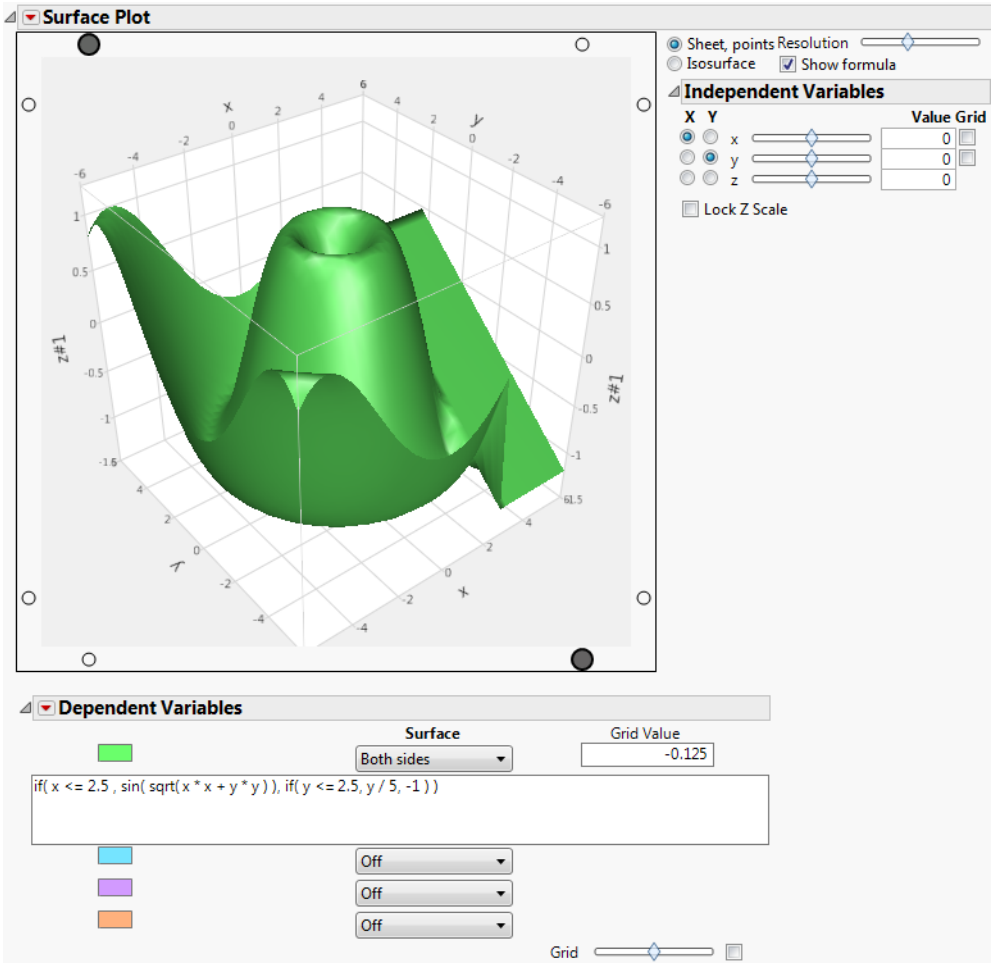
   Note: This can be done with or without a data table open.

2. Click **OK**.

   When you click OK in the Surface Plot launch window without selecting any columns, the default surface plot report appears. If no data tables are open, the launch window does not appear and this step can be skipped.

3. In the Appearance Controls, select the **Show Formula** check box to show the formula space.

Figure 5.13 Default Surface Plot with Formula



The default function shows in the box. To plot your own function, enter it in this box.

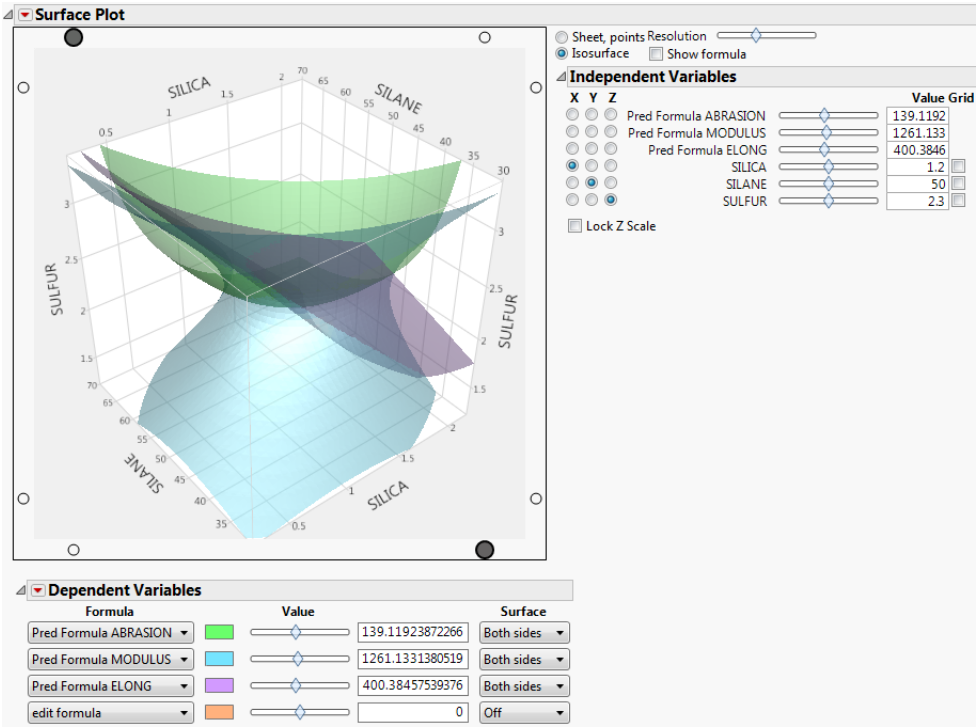
Plotting Isosurfaces

Isosurfaces are the 3-D equivalent to a 2-D contour plot. An isosurface requires a formula with three independent variables. The Resolution slider determines the $n \times n \times n$ cube of points over which the formula is evaluated. The Value slider in the Dependent Variable section selects the isosurface (that is, the contour level) value.

This example uses the Tiretread.jmp data table. The RSM for 4 Responses script produces a response surface model for dependent variables ABRASION, MODULUS, ELONG, and HARDNESS. The predicted formulas for these four responses are in the data table.

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Graph > Surface Plot**.
3. Select Pred Formula ABRASION, Pred FormulaMODULUS, and Pred FormulaELONG and click **Columns**.
4. Click **OK**.
5. Select the Isosurface radio button in the appearance controls.
6. In the Dependent Variables report, select **Both Sides** from the Surface menu for all three variables.

Figure 5.14 Isosurface of Three Variables



For the tire tread data, one might set the abrasion at a fixed minimum setting and the elongation at a fixed maximum setting. Use the MODULUS slider to see which values of MODULUS are inside the limits set by the other two surfaces.

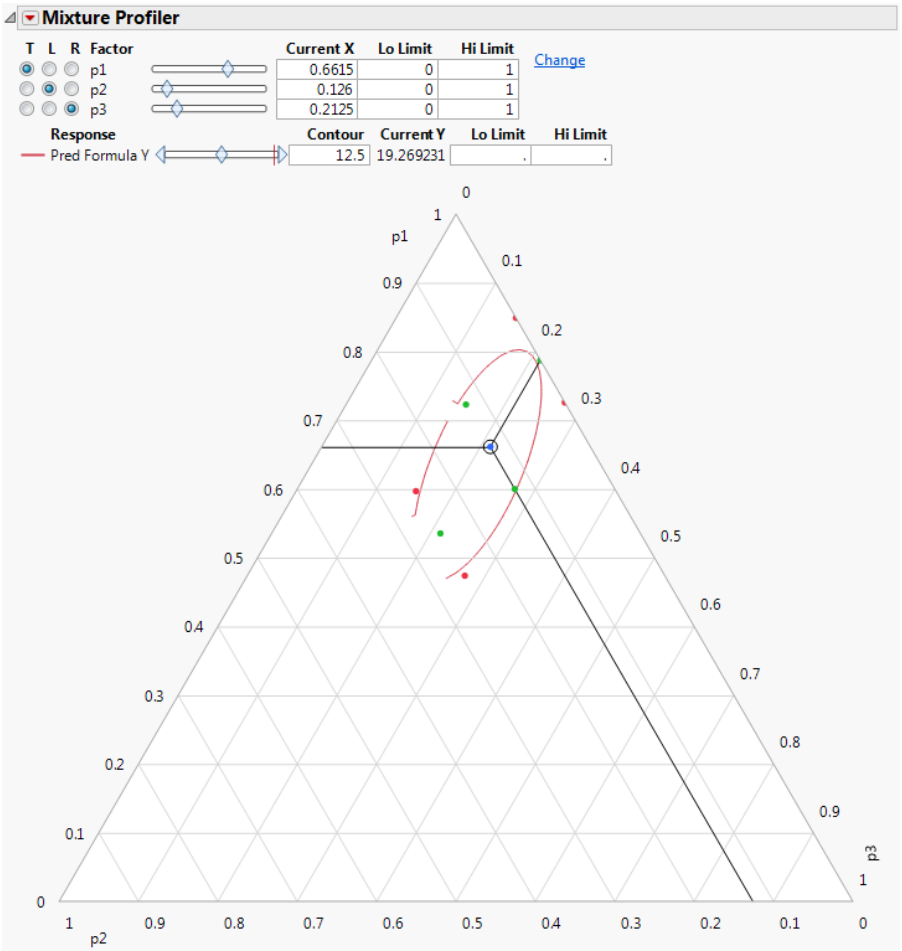
Chapter 6

Mixture Profiler

Explore Factor Effects Using Ternary Plots

The Mixture Profiler displays response contours on a ternary plot for mixture models, where three or more factors in the model are components (ingredients) in a mixture. The Mixture Profiler is useful for visualizing and optimizing response surfaces resulting from mixture experiments.

Figure 6.1 Mixture Profiler Example



Contents

Overview of the Mixture Profiler 125
Ternary Plot Overview 125
Ternary Plots with More Than Three Components 126
Example of the Mixture Profiler 127
Launch the Mixture Profiler Platform 128
The Mixture Profiler Report 129
Factor Settings and Controls 129
Response Settings and Controls 130
Mixture Profiler Platform Options 130
Customizations for the Mixture Profiler 131
Linear Constraints 131
Additional Examples of the Mixture Profiler Platform 132
Mixture Variables with a Process Variable 132
Multiple Responses with Five Mixture Variables 135

Overview of the Mixture Profiler

The Mixture Profiler shows response contours on a ternary plot for mixture models, where three or more factors in the model are components (ingredients) in a mixture. The Mixture Profiler is useful for visualizing and optimizing response surfaces resulting from mixture experiments.

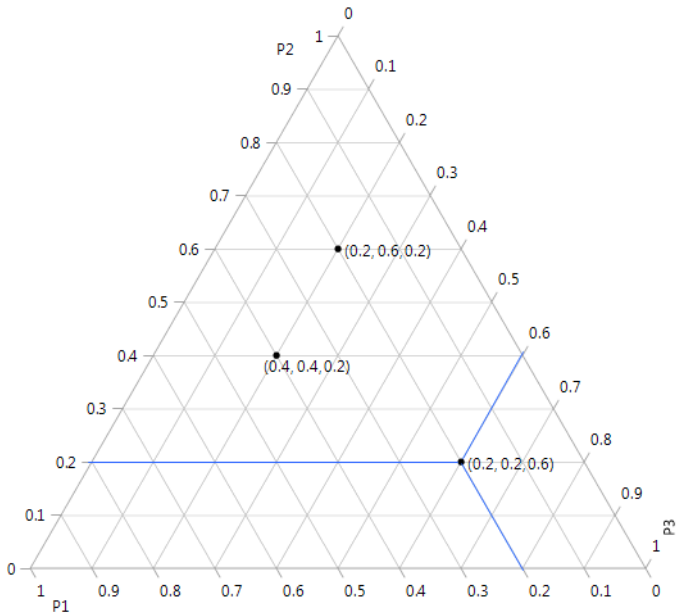
Many of the Mixture Profiler features are similar to those of the Contour Profiler as described in "Contour Profiler Platform Options". Unique to the Mixture Profiler is the use of a ternary plot.

Ternary Plot Overview

A ternary plot is a two-dimensional representation of three mixture components that sum to a constant. The plot is an equilateral triangle with an edge for each component. When unconstrained, each vertex of the triangle corresponds to a pure blend where one component is 1 (100%) and all other components are 0. When components are constrained, the feasible mixtures are represented by a portion of the ternary plot. Shading is used to exclude infeasible portions of the plot.

Figure 6.2 displays three components P1, P2, and P3. The three components are unconstrained with the range of each proportion ranging from 0 to 1 (100%). Three points are labeled with their coordinates (P1, P2, P3). One point (0.2, 0.2, 0.6) includes blue lines along the axis grid lines for each mixture component to guide you in reading the axes.

Figure 6.2 Ternary Plot Example



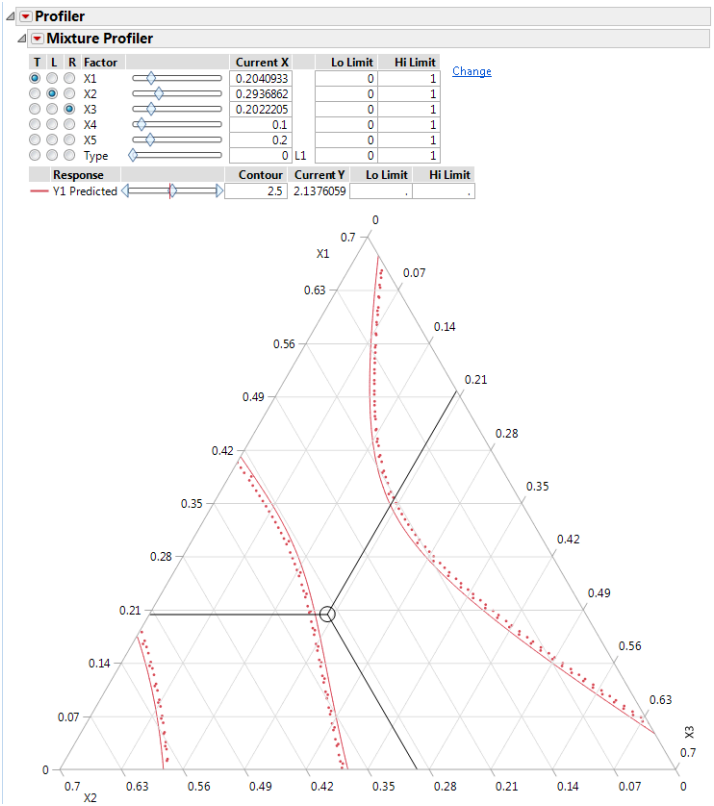
Ternary Plots with More Than Three Components

The ternary plot can display only three components at a time. For a model with more than three components, the total of the three on-axis (displayed) components is 1 minus the total of the off-axis (non-displayed) components. The plot axes are scaled such that the maximum value a component can attain is 1 minus the total for the off-axis components.

Figure 6.3 shows the Mixture Profiler Ternary Plot for an experiment with 5 factors. The Five Factor Mixture.jmp data table is being used, with the Y1 Predicted column as the formula. The on-axis factors are x1, x2, and x3, and the factors x4 and x5 are off-axis. There are no constraints on any of the mixture components. Note that all Lo Limit values are set to 0 and all Hi Limit values set to 1. The value for x4 is 0.1 and the value for x5 is 0.2, for a total of 0.3. The sum of x1, x2, and x3 is equal 1 – 0.3 = 0.7. Note that the maximum value for a plot axis is now 0.7, not 1.

If you change the value for either x4 or x5, then the values for x1, x2, and x3 change, maintaining their relative proportions, to accommodate the constraint that factor values sum to 1.

Figure 6.3 Ternary Plot with Scaled Axes to Account for Off-Axis Factors



Example of the Mixture Profiler

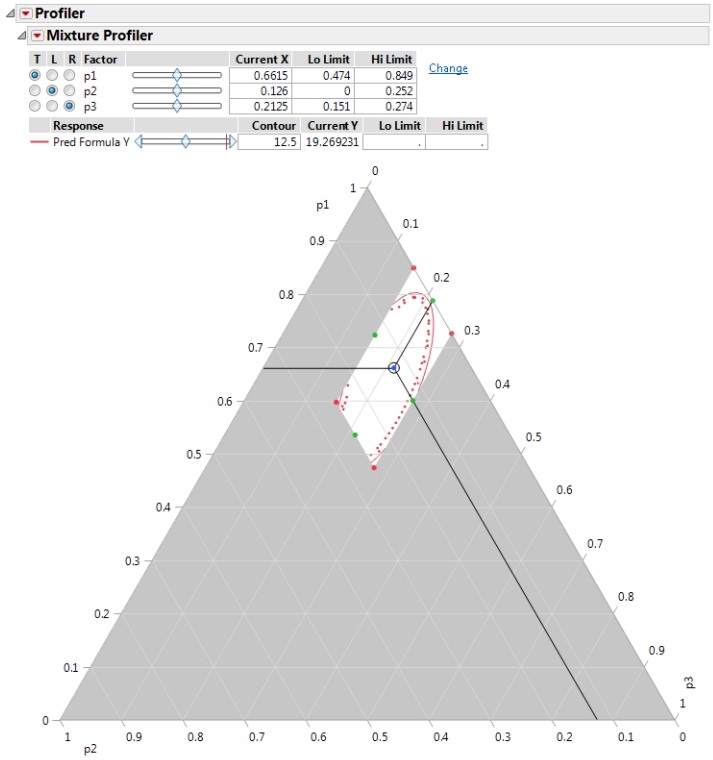
This example uses the Plasticizer.jmp sample data table There are three mixture components (p1, p2, and p3) and one response (Y). There are constraints on the component levels. This data table already contains a saved prediction formula column for the response.

1. Select **Help > Sample Data Library** and open Plasticizer.jmp.
2. Select **Graph > Mixture Profiler**.
3. Select Pred Formula Y and click **Y, Prediction Formula**.
4. Select Y from the Y, Prediction Formula role and click **Remove**.

Note: The Y column role is set to Y in the Columns list. The role automatically assigns columns in launch windows. In this case, the Y column was assigned to the Y role when the mixture profiler was launched. However, the Y column does not contain a formula, and the Mixture Profiler Y column requires a formula column.

5. Click **OK**.

Figure 6.4 Example of Mixture Profiler



Use the factor sliders to explore how changes impact the response. The unshaded region is the feasible region for this data set.

Launch the Mixture Profiler Platform

The Mixture Profiler can be accessed in the following ways:

- The Mixture Profiler can be accessed directly from the Graph menu. When you access the Mixture Profiler in this way, the Mixture Profiler launch window appears. See "Profiler Launch Windows".
- The Mixture Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information about the availability of the Mixture Profiler in different platforms.

- The Mixture Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Mixture Profiler from the Profiler red triangle menu.
- JMP PRO The Mixture Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Mixture Profiler from the Profiler red triangle menu.

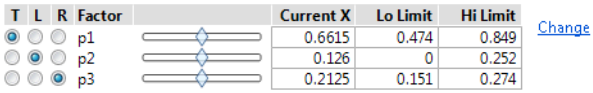
The Mixture Profiler Report

The initial Mixture Profiler report shows a mixture profiler plot, factor settings and controls, and response settings and controls.

- "Factor Settings and Controls"
- "Response Settings and Controls"

Factor Settings and Controls

Figure 6.5 Factor Settings and Controls



T, L, R Radio buttons to control factor assignments to the axes in the mixture profiler plot. **T** = top, **L** = left, and **R** = right.

Factor The list of factors.

Current X The current factor settings. Click in a box to change the value of a factor. The values for the other factors adjust proportionally to maintain the mixture sum. The slider controls can also be used to change factor settings.

Lo Limit The lower limit on each factor. Click in a box to change the value.

Hi Limit The upper limit on each factor. Click in a box to change the value.

Note: When factor limits are used, the profiler shows regions that are feasible as unshaded.

Change Opens the Factor Settings window. See "Mixture Profiler Platform Options".

Response Settings and Controls

Figure 6.6 Response Settings and Controls

Response The list of one or more responses.

Contour The current value of the contour on the mixture profiler. Click in the box to change the value. The slider control can also be used to change the value of the contour.

Current Y The predicted response based on the current X settings. The value is at the center of the cross hairs on the profiler plot. This value updates as the factor settings are changed.

Lo Limit Enables you to set a lower limit for your response.

High Limit Enables you to set an upper limit for your response.

Note: When response limits are used, the profiler shows regions that are feasible as unshaded.

Mixture Profiler Platform Options

Specify Factor Values Opens a dialog for specifying factor values. The entered values should sum to1. If needed, JMP adjusts the specified values to feasible settings. The option is also available from the **Change** button on the report window.

Show Points Shows or hides the individual points on the plot.

Show Current Value Shows or hides the three-way crosshairs on the plot. The intersection of the crosshairs represents the current factor values. The Current X values above the plot give the exact coordinates of the crosshairs.

Show Constraints Shows or hides the shading resulting from any constraints on the factors. Those constraints can be entered in the Lo Limits and Hi Limits columns above the plot, or in the Mixture Column Property for the factors.

Up Dots Shows or hides the dotted lines corresponding to each contour. The dotted lines show the direction of increasing response values, so that you get a sense of direction.

Contour Grid Draws contours on the plot at intervals that you specify.

Remove Contour Grid Removes the contour grid if one is on the plot.

Factor Settings A submenu of commands that enables you to save and transfer the Mixture Profiler settings to other parts of JMP. For more information on this submenu, see the discussion of the profiler in "Factor Settings".

Customizations for the Mixture Profiler

Customize the Mixture Profiler by right-clicking on the plot, and selecting **Customize**.

Contour Alter line color, fill color, and level of line transparency. If there are multiple contour lines on the plot, each appears individually on the list of options.

Component Constraints Alter component constraint display properties.

Linear Constraints Alter linear constraint display properties.

Grid Lines Use Axis Settings window to modify grid lines.

Reference Lines Use Axis Settings window to modify reference lines.

Double-click an axis on the plot to view the Axis Settings window. Here you can specify axis properties in detail, such as tick marks, grid lines and reference lines.

Crosshairs Alter the display properties of the crosshairs on the plot.

Marker Alter the display properties of the plot marker.

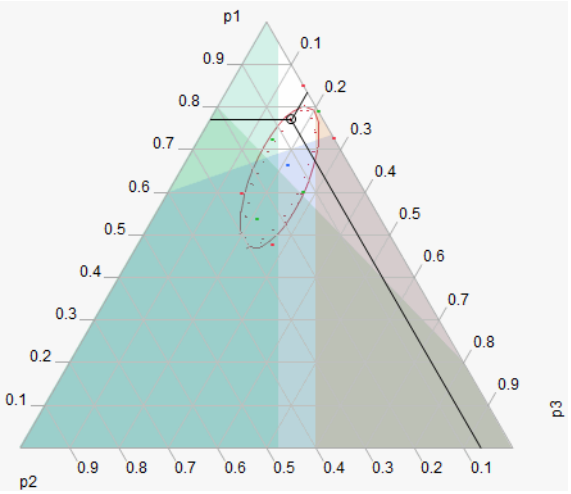
Linear Constraints

The Mixture Profiler can display linear constraints. To do this, a Constraint Table Script must be part of the data table. For more information about creating the Table Script, see "Linear Constraints".

When using constraints, unfeasible regions are shaded in the profiler. Figure 6.7 shows an example of a mixture profiler with shaded regions due to four constraints. The unshaded portion is the resulting feasible region. The constraints are below:

- $4*p2 + p3 \leq 0.8$
- $p2 + 1.5*p3 \leq 0.4$
- $p1 + 2*p2 \geq 0.8$
- $p1 + 2*p2 \leq 0.95$

Figure 6.7 Shaded Regions Due to Linear Constraints



Additional Examples of the Mixture Profiler Platform

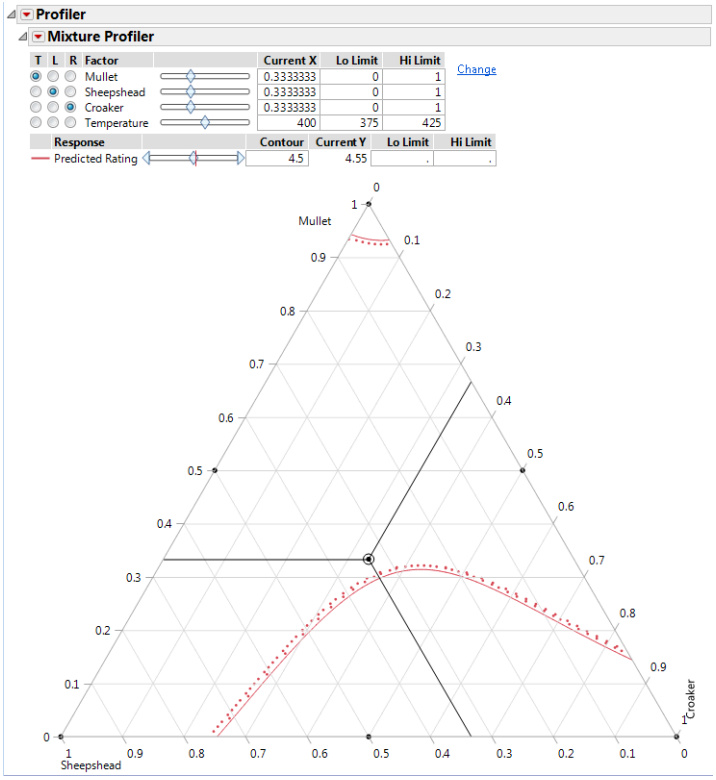
- "Mixture Variables with a Process Variable"
- "Multiple Responses with Five Mixture Variables"

Mixture Variables with a Process Variable

This example uses the Fish Patty.jmp sample data table. The data, adapted from Cornell (1990), comes from an experiment to optimize the texture of fish patties. The columns Mullet, Sheepshead, and Croaker are mixture components. Each column represents the proportion of a fish type in the fish patty. The Temperature column is a process variable. It is the oven temperature used to bake the patties. The column Rating is the response and is a measure of texture acceptability, where higher is better. A response surface model was fit to the data and the prediction formula was stored in the column Predicted Rating.

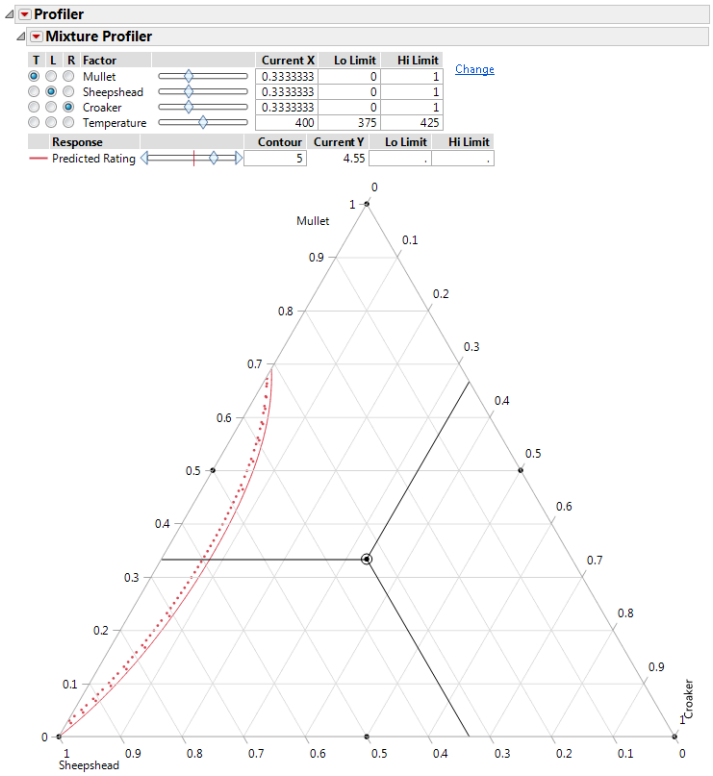
1. Select **Help > Sample Data Library** and open Fish Patty.jmp.
2. Select **Graph > Mixture Profiler**.
3. Select Predicted Rating and click **Y, Prediction Formula**.
4. Click **OK**.

Figure 6.8 Initial Output for Mixture Profiler



5. The manufacturer wants the rating to be at least 5. Use the slider control for Predicted Rating to move the contour close to 5. Alternatively, you can enter 5 in the Contour edit box to set the contour to 5.

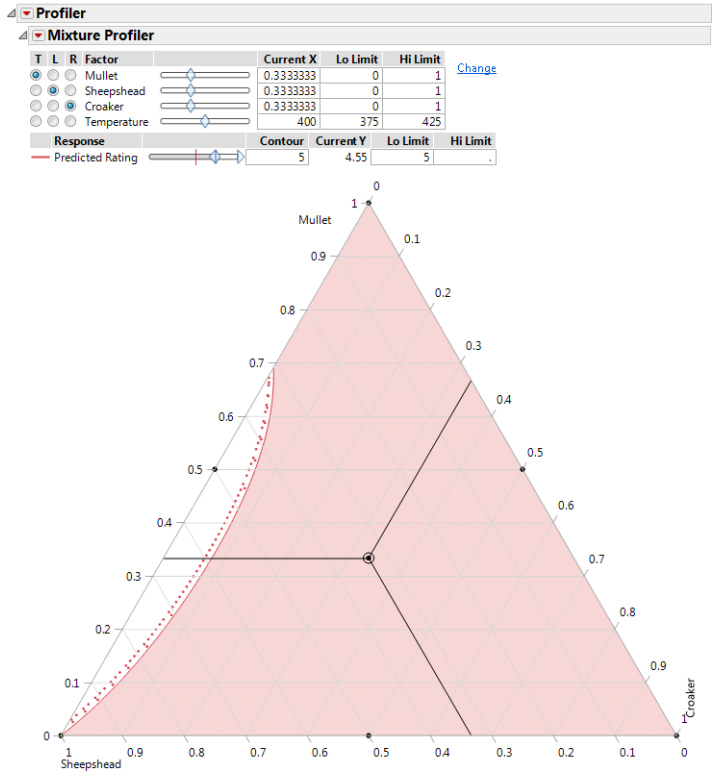
Figure 6.9 Contour Showing a Predicted Rating of 5



The Up Dots shown along the contour indicate the direction of increasing Predicted Rating.

6. Enter 5 in the Lo Limit edit box for the Predicted Rating. The resulting shaded region represents factor combinations that yield a rating less than 5. To produce patties with at least a rating of 5, the manufacturer can set the factors values anywhere in the feasible (unshaded) region. The region has small proportions of Croaker (<10%), mid to low proportions of Mullet (<70%) and mid to high proportions of Sheepshead (>30%). This region is for a baking temperature of 400 degrees.

Figure 6.10 Contour Shading Showing Predicted Rating of 5 or More.



7. Move the slide control for Temperature to observe how the feasible region changes for different temperature settings.

Additional analyses might include:

- Optimize the response across all four factors simultaneously. See "Custom Profiler" or "Desirability Profiling and Optimization".
- Simulate the response as a function of the random variation in the factors and model noise. See "Simulator".

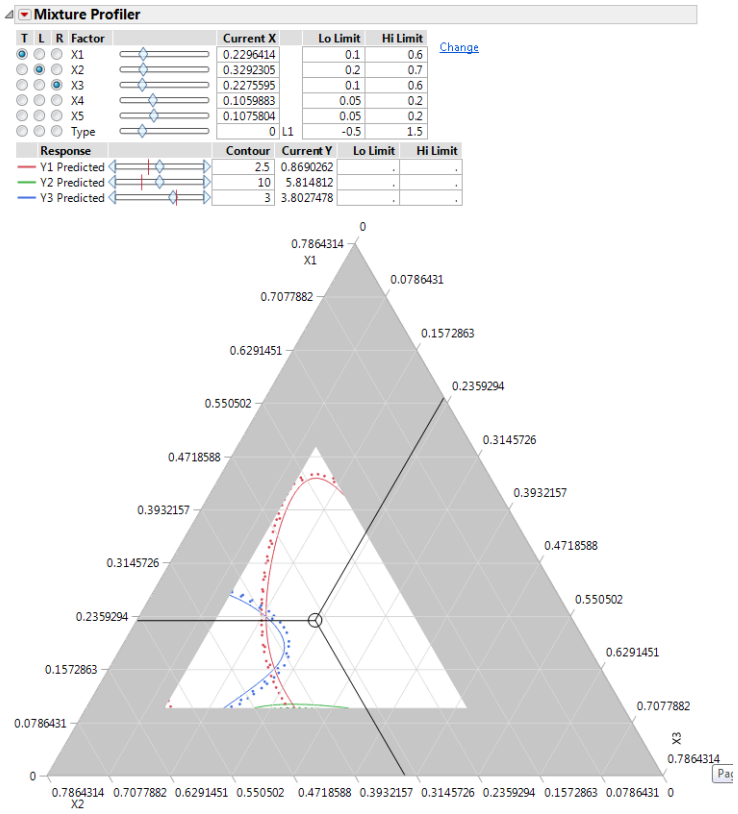
Multiple Responses with Five Mixture Variables

This example uses the Five Factor Mixture.jmp sample data table. There are five components (x1 to x5), one categorical process factor (Type), and three responses, Y1, Y2, and Y3. A response surface model is fit to each response and the prediction equations are saved in Y1 Predicted, Y2 Predicted, and Y3 Predicted.

1. Select **Help > Sample Data Library** and open Five Factor Mixture.jmp.
2. Select **Graph > Mixture Profiler**.

3. Select Y1 Predicted, Y2 Predicted, and Y3 Predicted and click **Y, Predicted**.
4. Click **OK**.
5. Enter 3 in the Contour edit box for Y3 Predicted so that the contour is visible on the plot.

Figure 6.11 Mixture Profiler with Y3 Predicted Contour Set to 3



A few items to note about the output.

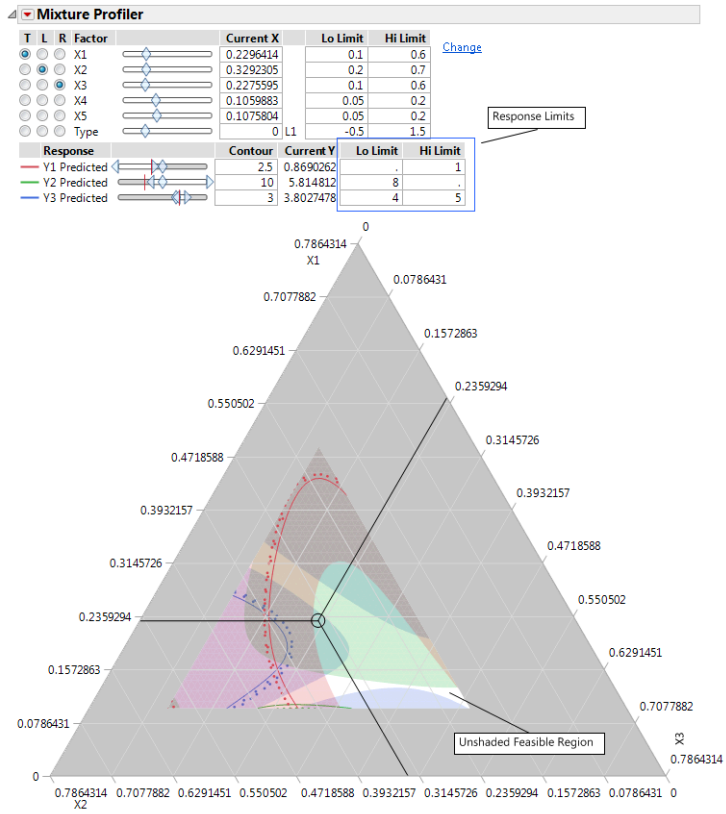
- The on axis factors are X1, X2, and X3 as indicated by the axes labels and the radio buttons in the factor settings. All of the factors have low and high limits, which were entered previously as Column Properties. See *Using JMP* for more information about entering column properties. Alternatively, you can enter the low and high limits directly by entering them in the Lo Limit and Hi Limit boxes.
- The white unshaded region is the feasible region. This is determined by the factor limits.
- The on-axis factors, x1, x2, and x3, radio buttons are selected.
- The categorical factor, Type, has a radio button, but it cannot be assigned to the plot. The current value for Type is L1, which is listed immediately to the right of the Current X box. The Current X box for Type uses a 0 to represent L1.

- All three prediction equations have contours on the plot and are differentiated by color.

Suppose the manufacturer desires to hold Y1 less than 1, Y2 greater than 8 and Y3 between 4 and 5, with a target of 4.5. The Mixture Profiler can help you investigate the response surface and find optimal factor settings.

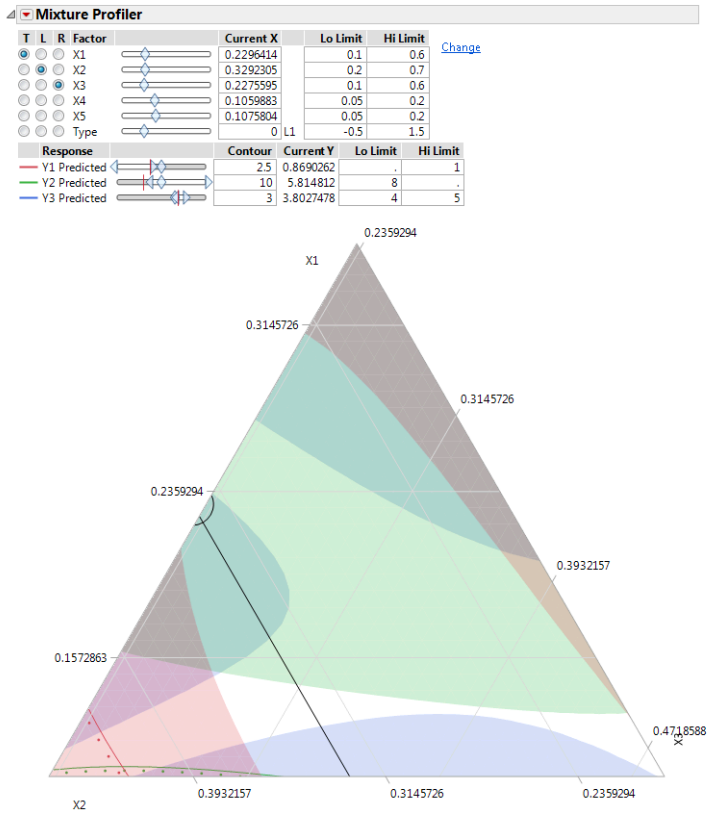
6. Enter 1 in the Y1 Predicted Hi Limit edit box. Enter 8 in the Y2 Predicted Lo Limit edit box. Enter 4 in theY3 Predicted Lo Limit edit box and 5 in the Y3 Predicted Hi Limit edit box.
7. The feasible region remains white (unshaded). Use the Response slider controls to position the contours in the feasible region. Alternatively, move the cross hair into the feasible region

Figure 6.12 Feasible Region After Setting Response Limits



8. Select the magnifier tool to zoom in on the feasible region.

Figure 6.13 Ternary Plot with Enlarged Feasible Region

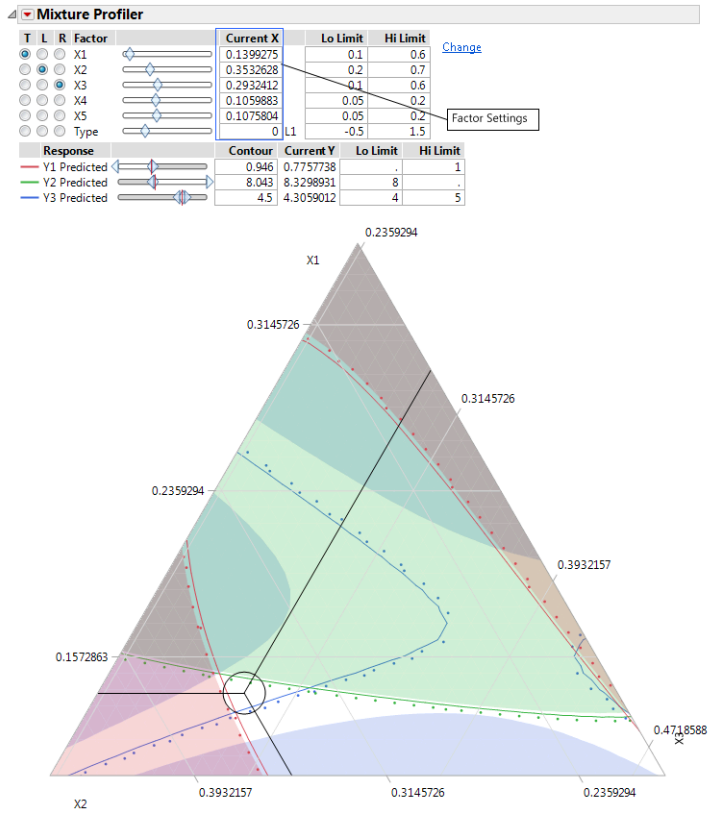


The manufacturer wants to maximize Y1, minimize Y2, and have Y3 at 4.5.

9. Use the slider controls or Contour edit boxes for Y1 Predicted to maximize the red contour within the feasible region. The Up Dots show direction of increasing predicted response.
10. Use the slider controls or Contour edit boxes for Y2 Predicted to minimize the green contour within the unshaded region.
11. Enter 4.5 in the Contour edit box for Y3 Predicted to set the blue contour to the target value.

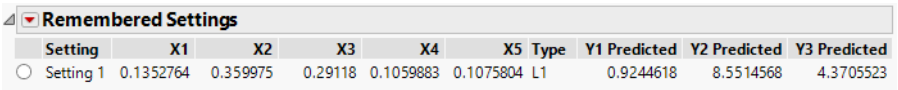
The resulting three contours do not all intersect at one spot. Y1 and Y2 cannot be optimized with Y3 on target, so you have to compromise. Position the three-way crosshairs in the middle of the contours to explore the factor levels that produce those response values. Note that these response contours are for the current settings of x4, x5, and Type.

Figure 6.14 Factor Settings for Optimal Results



12. Click the Mixture Profiler red triangle and select **Factor Settings > Remember Settings** to save the current settings. The settings are appended to the bottom of the report window.

Figure 6.15 Remembered Settings



Remembered Settings

Setting	X1	X2	X3	X4	X5	Type	Y1 Predicted	Y2 Predicted	Y3 Predicted
Setting 1	0.1352764	0.359975	0.29118	0.1059883	0.1075804	L1	0.9244618	8.5514568	4.3705523

With the current settings saved, you can now change the values of x4, x5 and Type to explore their impact on the feasible region. You can compare the factor settings and response values for each level of Type by referring to the Remembered Settings report.

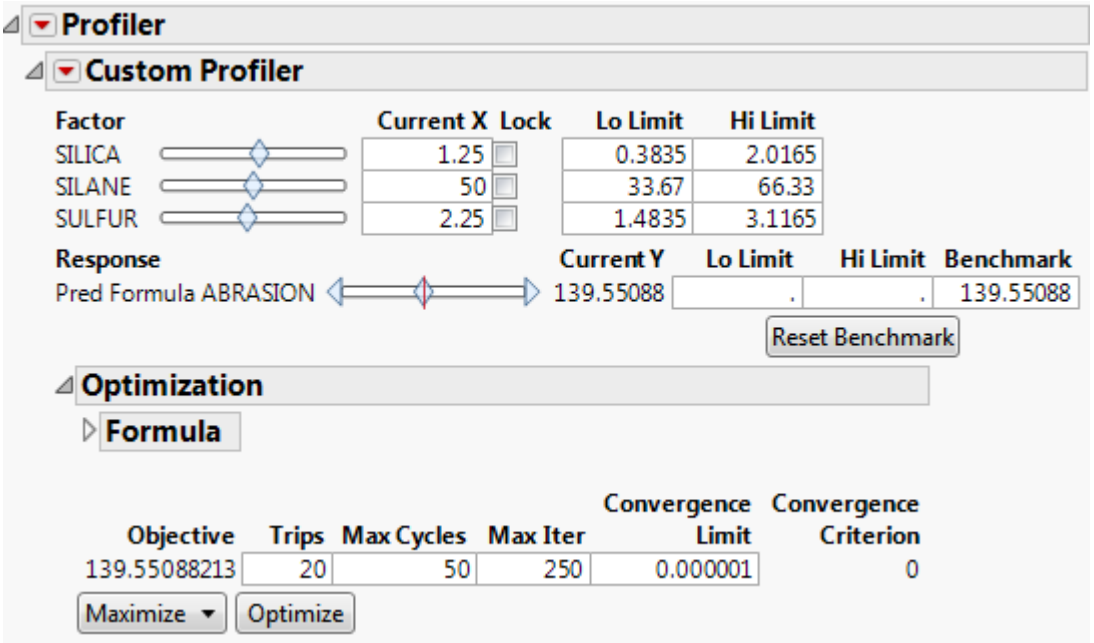
Chapter 7

Custom Profiler

Explore Response Surfaces Using a Numerical Calculator

The Custom Profiler enables you to optimize factor settings without graphical output. This is useful for large problems that would have too many graphs to visualize well.

Figure 7.1 Custom Profiler Example



Contents

Overview of the Custom Profiler 143
Example of the Custom Profiler 143
Launch the Custom Profiler Platform 145
The Custom Profiler Report 146
Factor Settings and Controls 146
Response Settings and Controls 147
Optimization Settings and Controls 147
Custom Profiler Platform Options 148

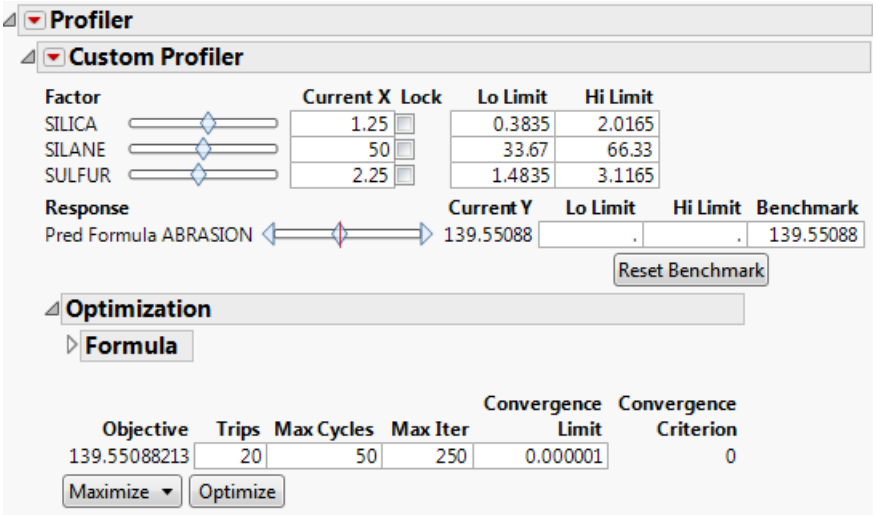
Overview of the Custom Profiler

The Custom Profiler enables you to optimize factor settings without graphical output. The Custom Profiler can be used for problems of any size. It is especially useful for large problems where the standard graphical profiler has too many graphs to visualize well.

The Custom Profiler report has many fields in common with other profilers. The Benchmark field holds a value, or benchmark, of the predicted response. You can compare new results to the benchmark value and update the value based on the current factor settings.

The Optimization report enables you to specify the formula to be optimized and specifications for the optimization.

Figure 7.2 Custom Profiler

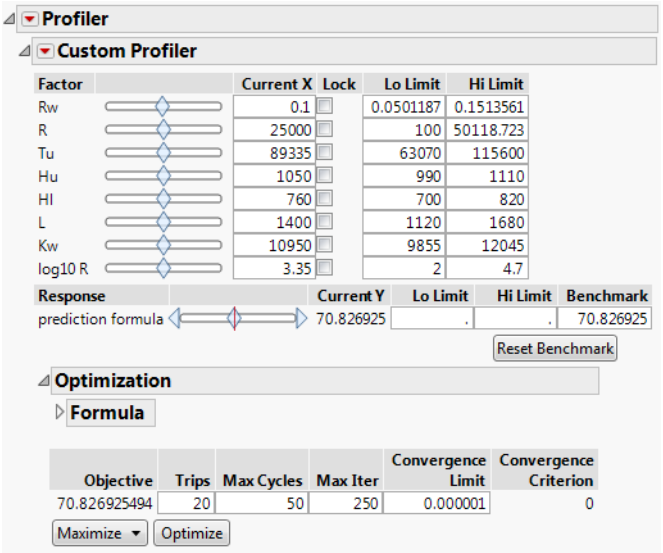


Example of the Custom Profiler

This example uses data that demonstrates the flow of water through a borehole that is drilled from the ground surface through two aquifers. You want to optimize the predicted value.

1. Select **Help > Sample Data Library > Design Experiment** and open Borehole Latin Hypercube.jmp
2. Select **Graph > Custom Profiler**.
3. Select prediction formula and click **Y, Prediction Formula.**
4. Click **OK**.

Figure 7.3 Custom Profiler Report



Note that the Benchmark value is 70.83. The Current Y value is also 70.83. This value is the predicted response with all factors set to their mean values.

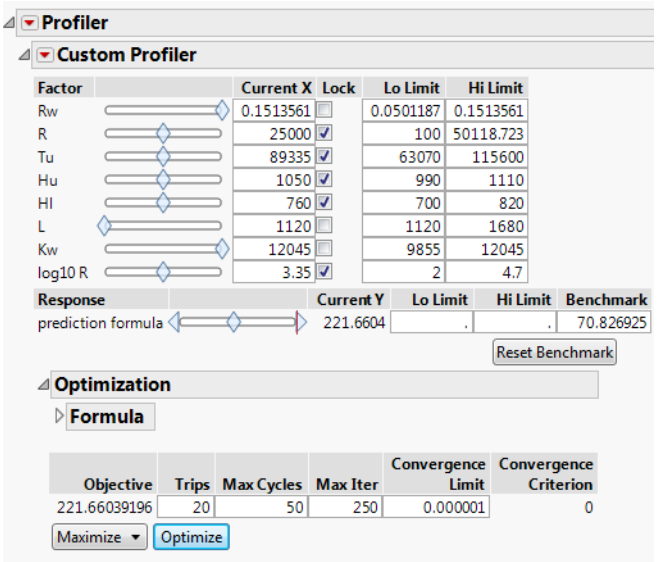
You would like to find the X values of the factors that optimize the predicted value. However, of the factors in the model, only those related to the borehole can be manipulated in practice. The other factors are out of human control. You can lock the uncontrollable factors at their mean values and optimize the factors that can be controlled.

Note: To view the description of a factor, right-click the corresponding column in the data table and select **Column Info**.

---

5. In the Custom Profiler Report, select the Lock check box next to R, Tu, Hu, HI, and log10 R.
6. In the Custom Profiler Report, click **Optimize**.

Figure 7.4 Optimized Custom Profiler Report



The optimization routine found an optimum predicted response at 221.66. In order to obtain the optimum of 221.66, Rw and Kw are set to their maximum values and L is set to its minimum value. The optimum is greater than the initial Benchmark value of 70.83.

Launch the Custom Profiler Platform

The Custom Profiler can be accessed in the following ways:

- The Custom Profiler can be accessed directly from the Graph menu. When you access the Custom Profiler in this way, the Custom Profiler launch window appears. See "Profiler Launch Windows".
- The Custom Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information about the availability of the Custom Profiler in different platforms.
- The Custom Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Custom Profiler from the Profiler red triangle menu.
- JMP PRO The Custom Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Custom Profiler from the Profiler red triangle menu.

The Custom Profiler Report

The initial Custom Profiler report shows settings and controls for the factors, responses, and optimization.

- "Factor Settings and Controls"
- "Response Settings and Controls"
- "Optimization Settings and Controls"

Factor Settings and Controls

Figure 7.5 Factor Settings and Controls

Factor		Current X	Lock	Lo Limit	Hi Limit
Rw		0.1513561		0.0501187	0.1513561
R		100		100	50118.723
Tu		115600		63070	115600
Hu		1110		990	1110
HI		700		700	820
L		1120		1120	1680
Kw		12045		9855	12045
log10 R		2		2	4.7

Factor The list of model factors.

Current X The current value of each factor. Click in a box to change the value of a factor. The slider controls can also be used to change factor settings.

Lock Enables you to lock a factor so that it is fixed when the optimization is performed. You can change a locked factor using the slider or clicking in the box in the Current X column. The lock applies only to the optimization.

Nominal Column Unlabeled column to the right of the Lock column that lists the current value of nominal factors.

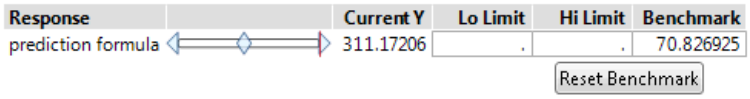
Note: The Current X column for nominal factors displays a coded (numeric) value for the current nominal factor.

Lo Limit The lower limit for each factor. Click in a box to change the value.

High Limit The upper limit for each factor. Click in a box to change the value.

Response Settings and Controls

Figure 7.6 Response Settings and Controls



Response The list of one or more responses.

Current Y The predicted response based on the current X settings. This value updates as the factor settings are changed.

Lo Limit Enables you to set a lower limit for your response.

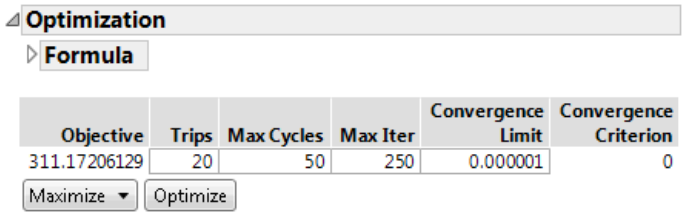
High Limit Enables you to set an upper limit for your response.

Benchmark A saved predicted value of the response. Initially, this value is set to the predicted value when all factors are at their mean value.

Reset Benchmark Updates the benchmark value to the current predicted value.

Optimization Settings and Controls

Figure 7.7 Optimization Settings and Controls



Formula The formula to be optimized. When a single response is used, the expression is the response column name. When multiple responses are used, the expression is a sum of desirability functions. You can edit the objective expression.

Objective The current value of the objective function. When a single response is used, the objective expression is the predicted response. When multiple responses are used, the objective expression is the desirability function. For more information about desirability functions, see "Desirability Profiling and Optimization".

Trips The number of random starts in the optimization algorithm. Each trip restarts the algorithm at a different starting point. This guards against finding local solutions.

Max Cycles The maximum number of cycles used in the optimization algorithm. Each cycle is single pass through the input parameters and optimizes each one individually.

Max Iter The maximum number of optimization iterations per cycle for each input parameter.

Convergence Limit The upper limit for the convergence criterion for the optimization algorithm. If the convergence criterion becomes less than this value, the algorithm stops.

Convergence Criterion The value of the convergence criterion for the optimization algorithm.

Maximize Enables you to choose to maximize or minimize the objective function.

Optimize Starts the optimization algorithm.

Custom Profiler Platform Options

Factor Settings Contains options identical to the Factor Settings submenu in the Prediction Profiler. See "Factor Settings".

Log Iterations Creates a data table that contains iterations of the optimization algorithm. The data table appears after the **Optimize** button is clicked.

Alter Linear Constraints Enables you to add, change, or delete linear constraints. The constraints are used in the Custom Profiler. See "Linear Constraints".

Save Linear Constraints Saves existing linear constraints as a data table script that is named Constraint. See "Linear Constraints".

Simulator Launches the Simulator. See "Simulator".

Chapter 8

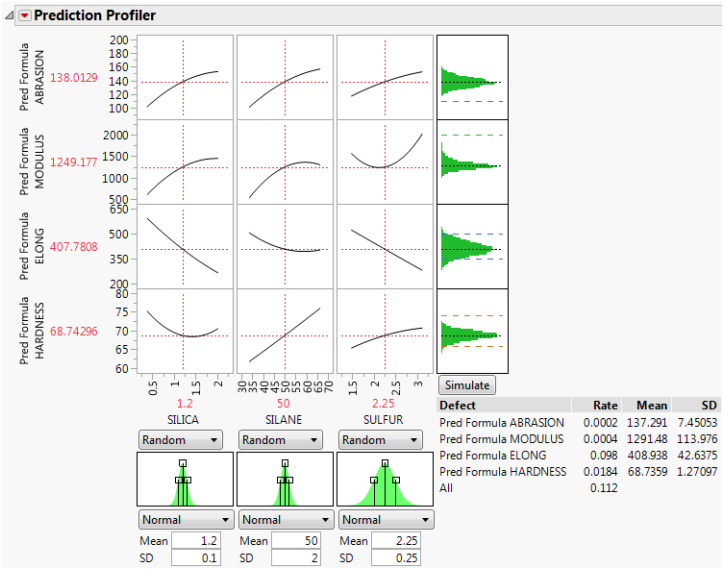
Simulator

Explore the Effects of Variation on Responses

Simulation enables you to discover the distribution of model outputs as a function of the random variation in the model inputs and noise. The Simulator in the profilers enables you to define random inputs, run simulations, and produce output tables of simulated values.

In the Prediction Profiler, the Simulator is integrated into the graphical layout. Factor specifications are displayed below each factor's profile. A simulation histogram is shown for each response.

Figure 8.1 Prediction Profiler with Simulator Example



Contents

Overview of the Simulator 151
Example of the Simulator 151
Launch the Simulator 154
Factors Simulation Settings 154
Response Simulation Settings 157
The Simulator Report 157
The Simulator Report Options 158
Defect Profiler 159
Introduction to Tolerance Design 159
Graph Scale 160
Defect Rates 160
Simulation Method and Details 161
Notes 161
Defect Parametric Profiler 162
Simulation Experiment 163
Specification Limits 163
Additional Examples of the Simulator 163
Example of the Defect Profiler 164
Example of Stochastic Optimization 167
Example of Simulating General Formulas 176
Statistical Details for the Simulator 179
Normal Weighted Distribution 179

Overview of the Simulator

Simulation enables you to discover the distribution of model outputs as a function of the random variation in the model inputs and noise. The Simulator in the profilers enables you to define random inputs, run simulations, and produce output tables of simulated values.

For example, you can use the simulator to model the defect rate of a process and to explore the robustness of the defect rate with respect to variation in the model factors. If specification limits have been set for the responses, they are displayed in the simulation output. The inclusion of specification limits enables you to perform a prospective capability analysis of the simulated model. For information about adding specification limits, see *Using JMP*.

You can access the Simulator through the Prediction, Contour, and Custom Profilers. The appearance of the Simulator depends on the profiler platform in which it appears.

Prediction Profiler

In the Prediction Profiler, the Simulator is integrated into the graphical layout. Factor specifications are displayed below each factor's profile. A simulation histogram is shown on the right for each response.

Contour Profiler and Custom Profiler

In the Contour and Custom profilers, the Simulator is not integrated into the graphical layout. There are no integrated histograms, and the interface is textual. However, the process is the same and the resulting output tables are the same as in other profilers.

Note: The Simulator supports mixture terms. The simulation assigns the random value to each mixture factor and then proportionally adjusts the values to maintain the factor constraints.

Example of the Simulator

Use sequencing to examine how the distribution of the response changes when the mean (sequencing location) and variability (sequencing spread) of the inputs change.

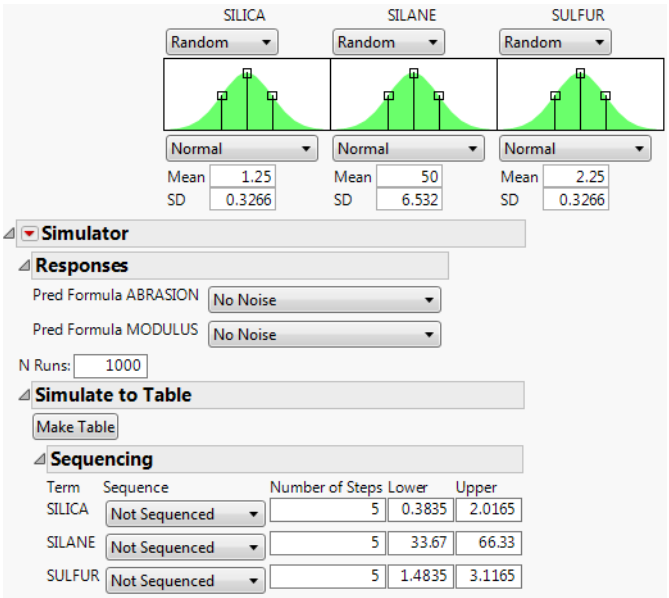
1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Graph > Profiler**.
3. Select Pred Formula ABRASION and Pred Formula MODULUS and click **Y, Prediction Formula**.
4. Click **OK**.

5. Click the Prediction Profiler red triangle and select **Simulator**.
6. Beneath each factor select **Random**.

Note: The default random distribution is a normal distribution.

7. Change the **N Runs** value to 1000.
8. Open Simulate to Table outline and then open the Sequencing outline.

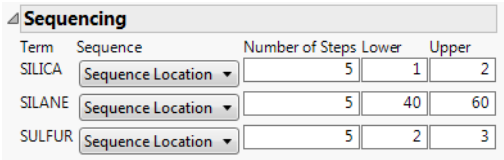
Figure 8.2 Simulator Settings



You want to examine how the responses change as the input factors change locations. To explore changes in factor mean, use sequence location. To instead explore changes in the variability of the factor, use sequence spread.

9. For SILICA, select **Sequence Location**. Leave the number of steps set to 5. Set the **Lower** limit to 1 and the **Upper** limit to 2.
10. For SILANE, select **Sequence Location**. Leave the number of steps set to 5. Set the **Lower** limit to 40 and the **Upper** limit to 60.
11. For SULFUR, select **Sequence Location**. Leave the number of steps set to 5. Set the values **Lower** limit to 2 and the **Upper** limit to 3.

Figure 8.3 Sequencing Settings



12. Click **Make Table**.

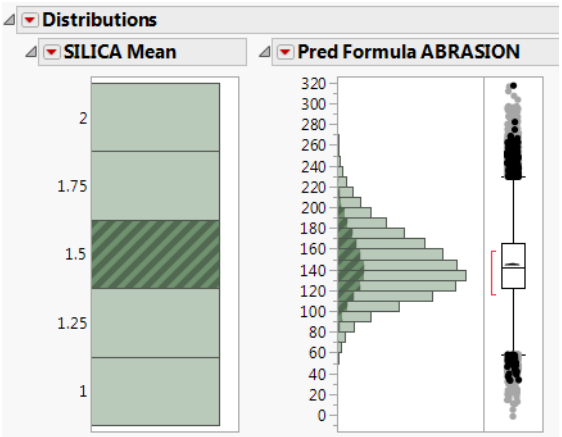
    The SILICA Mean, SILANE Mean, and SULFUR Mean columns contain five steps across their respective range of values. For example, the Silica Mean values are 1, 1.25, 1.5, 1.75, and 2. The SILICA, SILANE, and SULFUR columns are the simulated values from a normal distribution with a mean defined by the corresponding mean columns and a fixed standard deviation. Pred Formula ABRASION and Pred Formula MODULUS values are calculated for set of simulated values, so that you can explore how the responses change as the factor values change.

13. Select **Analyze > Distribution**.

14. Select SILICA Mean and Pred Formula ABRASION and click **Y, Columns**.

15. Click **OK**.

Figure 8.4 Distribution of SILICA Mean by Pred Formula ABRASION



Click a histogram bar that corresponds to a SILICA Mean to see how the predicted value for abrasion varies across the simulated silane and sulfur ranges given the selected value of silica.

Launch the Simulator

The Simulator can be accessed as a red triangle menu option in the Prediction Profiler, the Contour Profiler, and the Custom Profiler. When you launch the Simulator from the Prediction Profiler, factor specifications are displayed below each factor's profile. When you launch the Simulator from the Contour or Custom profilers, factor specifications appear in a Factors report, located below the respective profiler report. See "Factors Simulation Settings".

Once the Simulator is accessed, you can define simulation parameters for the responses in the Responses report. See "Response Simulation Settings". The Responses report also enables you to set the number of simulations.

A histogram is also added to the right of each response when the Simulator is accessed from the Prediction Profiler. Below the histograms is a **Simulate** button. Once you specify settings for the factors and responses, click **Simulate** to simulate response values. The histograms of the simulated responses are updated each time the simulate button is clicked. To simulate responses in the Contour and Custom profilers, open the Simulate to Table report and click **Make Table**. See "The Simulator Report".

Factors Simulation Settings

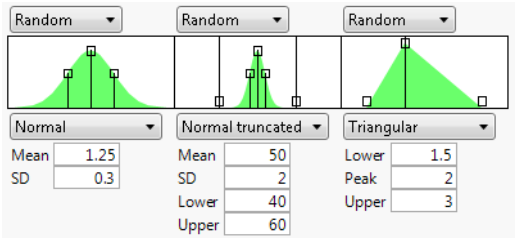
By default, factors (inputs) and responses (outputs) are defined by the functions that you select when the Profiler is launched. Simulation options enable you to modify how your factors and responses are simulated. There are controls for location, error, and noise.

Fixed Fixes the factor at the current value in the profiler for all simulation runs.

Random Draws a random value of the factor for each simulation run from the specified distribution and distributional parameters. When you select Random, distribution options appear. For more information about the options, see "Continuous Factors" and "Categorical Factors".

Note: The default random distribution is a normal distribution with a mean at the current factor setting and a standard deviation estimated by the range of the factor divided by 5. The mean of the normal distribution is automatically updated if you change the factor setting in the profiler. There are other random distributions that also update parameters automatically when the factor settings are changed. The Triangular distribution updates all three parameters, the Cauchy distribution updates the mode, and the Johnson Su distribution updates θ.

Figure 8.5 Example of Three Random Distributions



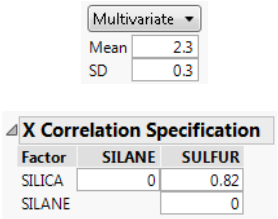
Expression Generates values for the factor based on a JSL expression. This gives you flexibility to use a random distribution of your choice. For example, you could create a censored normal distribution that guaranteed nonnegative values with an expression such as the following:

```
Max(0,RandomNormal(5,2))
```

After entering the expression, click the **Reset** button to submit the expression.

Multivariate Generates values for the factor based on a multivariate normal distribution. Specify the mean and standard deviation for each factor. The correlation matrix is defined separately in the X Correlation Specification report. You can use this option to accommodate correlated factors.

Figure 8.6 Defining a Correlation Matrix for the Multivariate Option



Continuous Factors

Many of the distributions available use standard random functions that are described in *Using JMP*. Descriptions of the specialized distributions available follow:

Normal weighted Generates values from a weighted normal distribution with the given mean and standard deviation. The weighting is a specific sampling scheme used to simulate rare events from the tails of the distribution. This is a good choice when you want to simulate very low defect rates. See "Statistical Details for the Simulator".

Normal truncated Generates values from a normal distribution limited by lower and upper limits. Any randomly generated value that exceeds a limit is discarded and another value is generated. This is a good choice when you want to simulate an inspection system where inputs that do not satisfy specification limits are discarded.

Normal censored Generates values from a normal distribution limited by lower and upper limits. Any randomly generated value that exceeds a limit is set to that limit, putting a density mass at the limits. This is a good choice when you want to simulate systems where inputs that do not satisfy specification limits are reworked until they are at the specification limit.

Sampled Generates factor values by selecting a value at random from that factor's column in the data table.

External Generates factor values by selecting a value at random from a column in another table. You are prompted to choose the table and column.

Note: The Aligned check box appears when you select Sampled or External. It is used when two or more factors are set to Sampled or External. When checked, the random draws come from the same row of the table. This maintains the correlation structure between two columns. If the Aligned option is used to associate two columns in different tables, the columns must have the same number of rows.

Categorical Factors

If a factor is categorical, then the random distribution is characterized by probabilities specified for each category. By default, the probabilities are set to the observed probabilities in the data table. You can change the probabilities using the handles in the plot or by changing the numerical values in the Prob column. The probabilities must sum to 1.

Note: If the probabilities that you select do not sum to 1, they are automatically regularized to sum to 1 once you click the Simulate button.

When the simulation for a factor is set to Random, a graphical representation of the random distribution density is shown. The graph shows the form of the density for the continuous distributions, and provides control points that can be dragged to change the distribution. The drag points for the Normal are the mean and the mean plus or minus one standard deviation. The Normal truncated and censored add points for the lower and upper limits. The Uniform and Triangular have limit control points, and the Triangular adds the mode.

Response Simulation Settings

Simulation settings for the responses enable you to simulate random noise in the responses. Often your factors explain part of the variation in the response with the rest of the variation of the response attributed to random noise. This is how the random noise is specified in your simulation.

No Noise Generates the response from the model based on the factor settings. No random noise added to the response.

Add Random Noise Adds a normal random value with mean zero and specified standard deviation to the generated response.

Add Random Weighted Noise Adds an error term to the generated response based on a weighted sampling scheme used to simulate rare events from the tails of the distribution.

Add Multivariate Noise Adds an error term to the generated response based on a multivariate normal distribution. Specify the standard deviation for the response. Specify the correlation structure between responses in the Y Correlation Specification report.

Random by Model (Available only within the Standard Least Squares and Generalized Regression personalities of the Fit Model platform.) Adds an error term to the generated response based on the model distribution specified in the Fit Model launch window.

The Simulator Report

The Simulator report contains the following items:

Simulate to Table Contains a Make Table button that saves the results of the simulation to a data table. If a response has specification limits, the table includes a Y In Spec column that indicates whether the simulated response is within the specification limits. The table also includes an Objective column, labeled Obj. If you added Desirability functions to your profiler prior to simulation, this column is the evaluation of the desirability function. If the Desirability functions were not present prior to simulation, the column contains all missing values.

Sequencing (Available only when the distribution is Normal, Uniform, or Triangular random.) Sequencing enables you to simulate across a range of locations or spreads for each factor. The simulations are saved to a data table. Select an option to sequence the location or spread of each factor. Specify the number of steps and the range for the sequencing. For each combination of the number of steps, the simulation runs N Runs. If you have two terms, each with 5 steps, and N Runs = 100, the resulting simulation table has 2,500 rows.

The Simulator Report Options

The Simulator red triangle menu contains the following options:

Automatic Histogram Update Toggles the histogram update. When this option is selected, the histograms update with new simulated values when you drag factor setting handles. By default, this option is not selected.

Defect Profiler (Available only when spec limits are specified.) Shows the defect rate as an isolated function of each factor. See "Defect Profiler".

Defect Parametric Profile (Available when the Defect Profiler is launched.) Shows the defect rate as an isolated function of the parameters of each factor's distribution. See "Defect Parametric Profiler".

N Strata (Appears only when you press Shift before clicking the Simulator red triangle menu.) Enables you to specify the number of strata for the Normal Weighted distribution. See "Statistical Details for the Simulator".

Set Random Seed (Appears only when you press Shift before clicking the Simulator red triangle menu.) Enables you to specify a seed for the simulation starting point. This enables the simulation results to be reproducible, unless the seed is set to zero. The seed is set to zero by default. If the seed is nonzero, then the latest simulation results are output when the Make Table button is clicked.

Simulation Experiment (Available for Normal, Uniform, and Triangular random distributions.) Runs a designed simulation experiment on the locations of the factor distributions. A window appears that enables you to specify the number of design points, the portion of the factor space to be used in the experiment, and which factors to vary in the experiment. Factors that are not selected for the experiment are set to their current Profiler value. See "Simulation Experiment".

Spec Limits Opens a table to set or edit specification limits for the responses. Click the **Save** button to save specification limits to the response column properties in the data table. See "Specification Limits".

Defect Profiler

To use the Defect Profiler, spec limits must be defined for at least one response. The Defect Profiler shows the probability of an out-of-spec output defect as a function of each factor while the other factors vary randomly. This is used to help visualize the changes in a factor's distribution that the process is most sensitive to, often as part of tolerance design.

Specification limits are used to define a defect, and the random distributions assigned to each factor are used to simulate responses.

Define at least one of the factors as Random in the simulation settings to obtain useful results. Otherwise, the simulation outputs are constant and the defect rate is zero if the outcome falls within spec or one if the outcome falls out of spec.

Tip: If you need to estimate very small defect rates, consider the Normal weighted random option. This random distribution estimates stable defect rates of just a few parts per million with only a few thousand simulation runs.

Introduction to Tolerance Design

Tolerance design is the investigation of how defect rates can be controlled by controlling variability in the input factors.

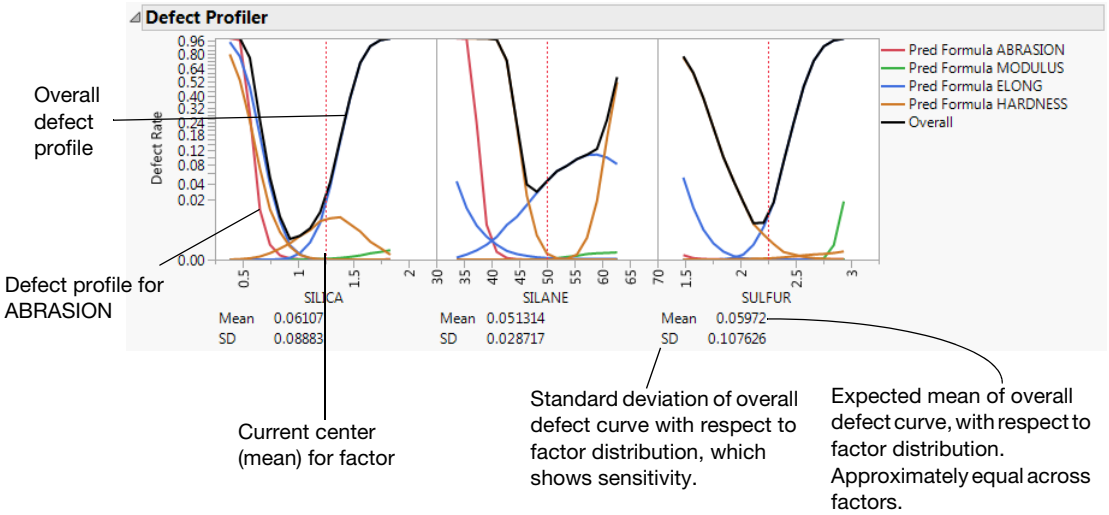
Input factors have variation. Specification limits are used to define acceptable ranges for each input. The variability in the input factors impacts the outputs. Specification limits are also used to define the acceptable range for output variables.

Sometimes, a tolerance design study shows that spec limits on an input are unnecessarily tight. Loosening such limits can result in reduced costs without sacrificing product quality. In these cases, tolerance design can save money.

In other cases, a tolerance design study might find that either tighter limits or different targets result in higher quality. In all cases, it is valuable to understand how the inputs and their variability impact that the defect rates.

The Defect Profiler shows the defect rate as a function of each factor fixed at the specified mean value while all other factors are varied according to their random specification. If there are multiple outputs with spec limits, each has a different colored defect rate curve. A black curve shows the overall defect rate.

Figure 8.7 Defect Profiler



Graph Scale

Defect rates are shown on a cubic root scale, so that both large and small defect rates are shown in some detail.

Defect Rates

The mean and standard deviation (SD) of the simulated overall defect rate are reported below each defect profile plot. This mean is calculated by integrating the defect profile curve with the specified factor distribution. Because of numerical error in the estimations, the overall mean defect rate reported under each factor can differ slightly.

The defect rates that are reported in the Defect Profiler are estimates of the simulated overall defect rate. This rate is also reported in the Rate column of the defect table that is in the Prediction Profiler outline.

Note: The defect table is added to the Prediction Profiler outline after you run a simulation and one or more of your responses has defined spec limits.

Because the rate estimates are obtained differently, by integration and simulation, they might differ slightly. If they are very different, you might consider increasing the number of simulation runs. In addition, verify that the range of the factor scale is wide enough so that the integration covers the distribution well.

The standard deviation is a measure of the sensitivity of the defect rate to the factor. It is small when either the factor profile is flat or the factor distribution has a small variance. The larger the standard deviation, the larger the impact that changes in that factor have on the defect rate variability. Comparing SDs across factors enables you to select factors for improvement to reduce defect rates.

The mean and SD values are updated when you change the factor distribution settings. This is one way to explore how to reduce defects as a function of one particular factor at a time. You can click and drag a handle point on the factor distribution and watch the mean and SD change as you drag. However, changes are not updated across all factors until you click the Rerun button to generate updated simulation runs.

Simulation Method and Details

Suppose you want a defect profile for factor X1, in the presence of random variation in X2 and X3. A series of n = N Runs simulation runs is done at each of k points in a grid of equally spaced values of X1. (By default, k is set at 17.) At each grid point, suppose that there are m defects due to the specification limits. At that grid point, the defect rate is m/n. These defect rates are connected and plotted as a continuous function of X1.

Notes

- The profile curve is not recalculated automatically when distributions change. Click **Rerun** to update the curve.
- Defect Profiling does not address the general optimization problem of optimizing quality against cost, given functions that represent all aspects of the problem. This more general problem would benefit from a surrogate model and space filling design.
- The defect profiles tend to be uneven when they are low. This is due, in part, to the use of the cubic scale where differences in low values are exaggerated. Jagged defect profiles could be due to limited simulation runs. If the overall defect curve (black line) is smooth, and the individual defect rates are fairly consistent, then your simulation probably has enough runs to provide a stable solution. If the overall defect rate curve is jagged, then consider increasing the number of runs. 20,000 runs is generally enough to stabilize the curves.

Defect Parametric Profiler

The Defect Parametric Profiler report shows the impact of process changes on the defect rate. The impacts are based on the simulation parameter settings for each factor. Four scenarios are considered.

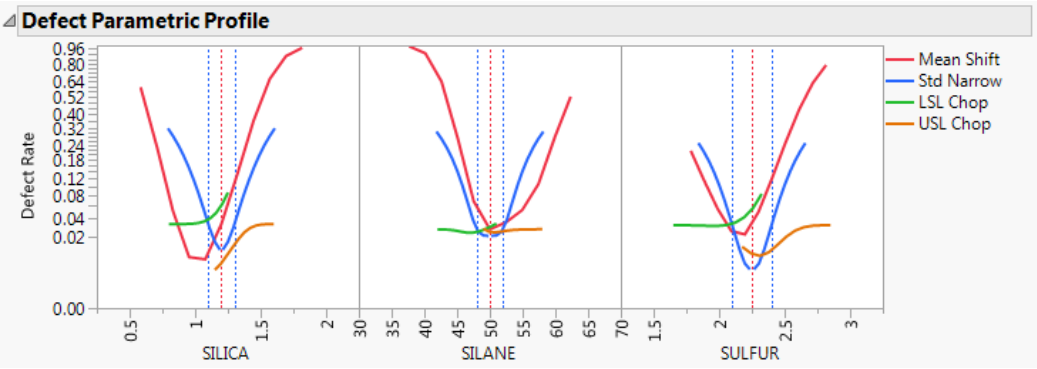
Mean Shift The impact of shifting the mean is shown by a red curve. The current mean is shown by a red dotted vertical line.

Std Narrow The impact of a reduction in variability is shown by a blue curve. The dotted blue vertical lines are set at the mean plus and minus one standard deviation. The minimum value on the curve corresponds to the defect rate when there is no variability.

LSL Chop The impact of inspection to remove all parts below the lower specification limit is shown by a green curve.

USL Chop The impact of inspection to remove all parts above the upper specification limit is shown by an orange curve.

Figure 8.8 Defect Parametric Profile



Simulation Experiment

Use the Simulation Experiment option to run a designed simulation experiment on the locations of the factor distributions. When you select Simulation Experiment, you to specify the number of design points, the portion of the factor space to be used in the experiment, and which factors to include in the experiment. For factors that are not included in the experiment, the current value set in the Profiler is used in the experiment.

The experimental design is a Latin Hypercube. At each design point, N Runs random draws are generated with the design point serving as the center of the random draws. The shape and variability come from the specified distributions. The output has one row for each design point. The responses include the defect rate for each response with spec limits and an overall defect rate. You can use a Gaussian Process model to model the overall defect rate from the simulation experiment. For more information about Gaussian Process models, see *Predictive and Specialized Modeling*.

Specification Limits

To add specification limits to a variable, right-click the desired column and select **Column Properties > Spec Limits.** In the Column Properties window, there are boxes to enter values for the Lower Spec Limit, Target, and Upper Spec Limit. Alternatively, you can define Specification Limits using the **Spec Limits** option in the Simulator red triangle menu.

The profilers support specification limits on the responses and provide a number of features:

- In the Profiler, if the Response Limits are not set up in the input data table to provide desirability coordinates, JMP looks for a Spec Limits column property and constructs desirability functions appropriate to those spec limits. For more information about Response Limits see *Using JMP*.
- If you use the Simulator to output simulation tables, JMP includes the Spec Limits column properties in the output data tables. This makes accounting for defect rates and capability indices easy.
- Adding spec limits enables the Defect Profiler.

Additional Examples of the Simulator

- "Example of the Defect Profiler"
- "Example of Stochastic Optimization"
- "Example of Simulating General Formulas"

Example of the Defect Profiler

To demonstrate a possible workflow with the Defect Profiler, we use Tiretread.jmp. The experimental data in the Tiretread.jmp sample data table comes from an experiment to study the effects of SILICA, SILANE, and SULFUR on four measures of tire tread performance.

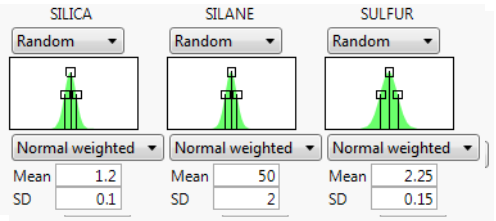
Add Specification Limits in the Simulator

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Graph > Profiler**.
3. Select Pred Formula ABRASION, Pred Formula MODULUS, Pred Formula ELONG, and Pred Formula HARDNESS and click **Y, Prediction Formula**.
4. Click **OK**.
5. Click the Prediction Profiler red triangle and select **Simulator**.
6. Click the Simulator red triangle and select **Spec Limits**.

   Note: If your columns have spec limits saved as a column property, those spec limits are shown in the Simulator Spec Limits table.

7. Set the spec limits:
   - For ABRASION set the LSL to 110.
   - For MODULUS set the USL to 2000.
   - For ELONG set the LSL to 350 and the USL to 550.
   - For HARDNESS set the LSL to 66 and the USL to 74.
8. Click **Save** to save the Spec Limits to the data table.
9. Beneath each factor, select **Random**.
10. For each factor, enter the random specifications as shown in Figure 8.9.

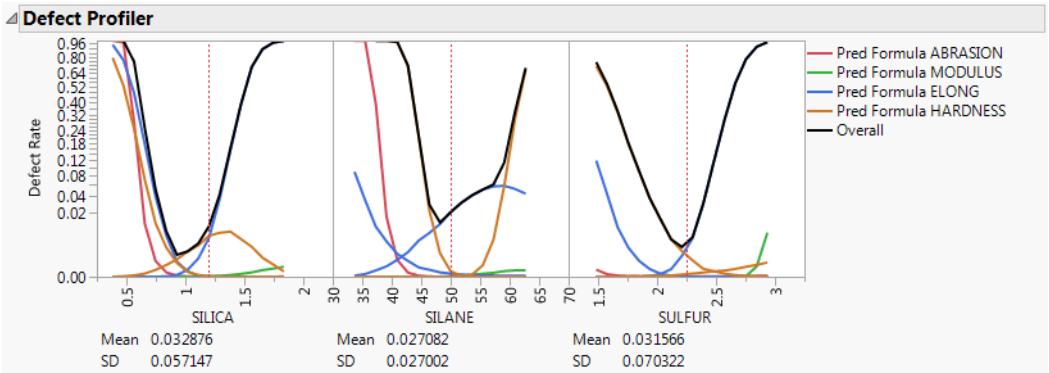
Figure 8.9 Profiler Random Specifications



Open the Defect Profiler and Defect Parametric Profile

1. Click the Simulator red triangle and select **Defect Profiler.**

Figure 8.10 Defect Profiler



Note: The defect profiler is based on simulation with random inputs. The values that you obtain might differ from those shown in Figure 8.10.

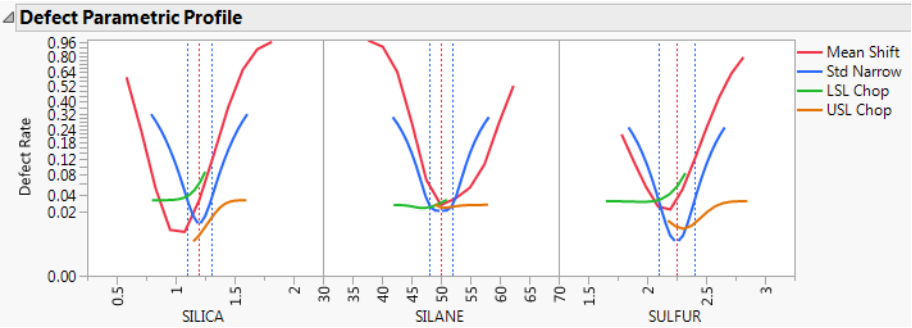
The black curves show the overall defect rate for each factor fixed at the X axis value while all other factors vary.

Consider the overall curve for SILICA. As silica varies, the defect rate goes from the lowest rate of 0.001 when silica is about 1 and increases quickly up to a defect rate of nearly 1 as silica increases or decreases from 1. However, SILICA is itself random. If you integrate the density curve of SILICA, you would estimate the average defect rate to be about 0.03, which is shown as the Mean for SILICA. This estimate of the overall defect rate estimated by the simulation is shown in the defect table under the simulation histograms. The Mean value for the overall defect rate for all factors are similar.

The Defect Profiler also includes an estimate of the standard deviation of the defect rate with respect to the variation in each factor. This value (labeled SD) is 0.057 for SILICA. The standard deviation is related to the sensitivity of the defect rate with respect to the distribution of that factor. Comparing the SD values across the three factors, the SD for SULFUR is higher than the SD values for SILICA and SILANE. This indicates that to improve the defect rate, shifting the distribution in SULFUR should have the greatest effect. A distribution can be shifted by changing its mean, changing its standard deviation, or by truncating the distribution by rejecting inputs that do not meet certain specification limits.

2. Click the Simulator red triangle and select **Defect Parametric Profile**.

Figure 8.11 Defect Parametric Profile



Consider SULFUR and note that the current defect rate (0.03) is represented in four ways corresponding to each of the four curves in the Parametric Profiler.

For the red curve, Mean Shift, the current rate is where the red curve intersects the vertical red dotted line. The Mean Shift curve represents the change in overall defect rate as the mean of SULFUR changes. One opportunity to reduce the defect rate is to shift the mean slightly to the left. If you use the crosshair tool on this plot, you see that a shift down in the mean reduces the defect rate to about 0.02.

For the blue curve, Std Narrow, the current defect rate is where the blue curve intersects the two dotted blue lines. The Std Narrow curves represent the change in defect rate as the standard deviation changes. The dotted blue lines represent one standard deviation below and above the current mean. The blue curve is drawn symmetrically around the center. At the center, the blue curve reaches a minimum, representing the defect rate for a standard deviation of zero. That is, if we totally eliminate variation in SULFUR, the defect rate is about 0.003. This is much lower than the current rate of 0.03. If you look at the other defect parametric profile curves, you can see that this is better than reducing variation in the other factors, something that we suspected by the SD value for SULFUR.

For the green curve, LSL Chop, there are no interesting opportunities for improvement in the defect rate. The green curve is above current defect rates for the entire range of the curve. This indicates that reducing the variation by rejecting parts with too-small values for SULFUR does not help reduce the defect rate.

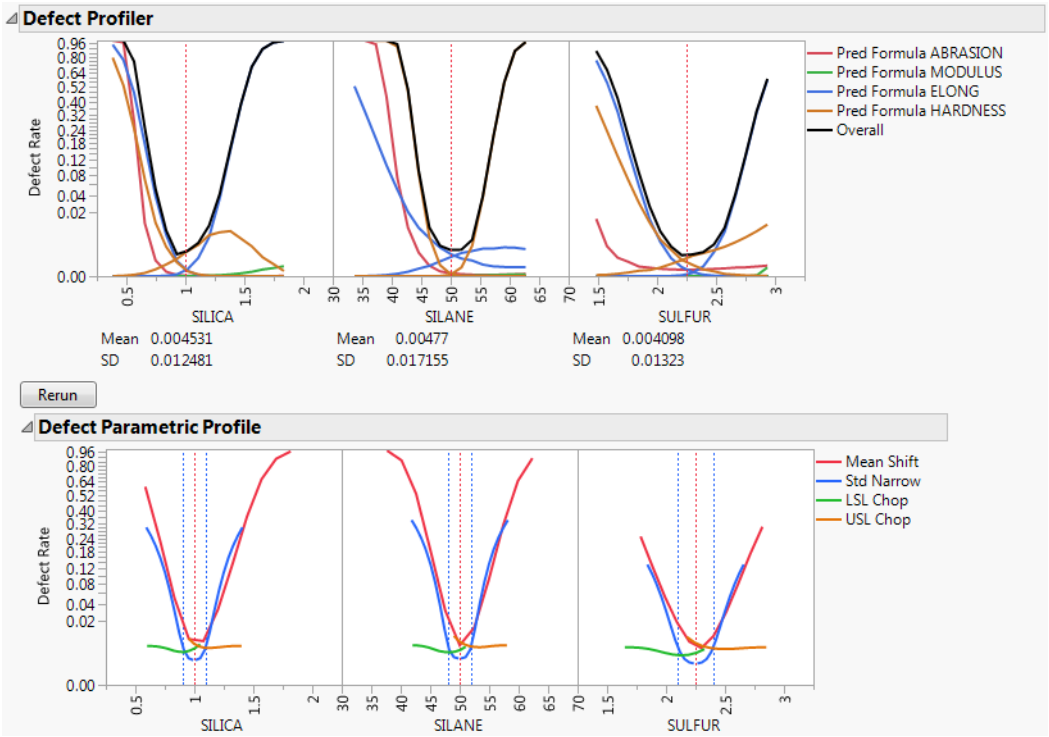
The orange curve, USL Chop, suggests a way to improve the defect rate. Reading the curve from the right, the curve starts out at the current defect rate (0.03). Then as you start rejecting more parts by decreasing the USL for SULFUR, the defect rate improves. However, moving a spec limit to the center of the distribution is equivalent to throwing away half the parts, which might not be a practical solution.

Looking at all the opportunities over all factors, it now looks like there are two good options for further investigation. You could shift the mean of SILICA to about 1 or reduce the variation in SULFUR. Because it is generally easier in practice to change a process mean than a process variation, the best first adjustment might be to shift the mean of SILICA to 1.

3. In the Prediction Profiler, adjust the Mean of SILICA from 1.2 to 1.0.
4. Below the Defect Profiler, click **Rerun**.

   The Defect Profiler updates based on the adjusted SILICA value.

Figure 8.12 Adjusted Defect Rates

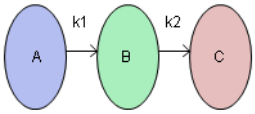


By shifting the mean of SILICA from 1.2 to 1.0, the defect rate has decreased from 0.03 to about 0.004, which is a good improvement. Investigating other defect reduction scenarios is easy to do by making changes to the distributions and rerunning the simulations.

Example of Stochastic Optimization

This example is adapted from Box and Draper (2007) and uses Stochastic Optimization.jmp. A chemical reaction converts chemical A into chemical B and B into C. The resulting amount of chemical B is a function of reaction time and reaction temperature. Use the profiler and simulator to explore factor settings to optimize the yield of B.

Figure 8.13 Chemical Reaction



For this reaction the yield of B can be computed using the Arrhenius laws. The column Yield contains the formula for yield. The formula is a function of Reaction Time (hours) and reaction rates k1 and k2. The reaction rates are a function of Reaction Temperature (degrees Kelvin) and known physical constants θ_1, θ_2, θ_3, θ_4. Therefore, Yield is a function of Reaction Time and Reaction Temperature.

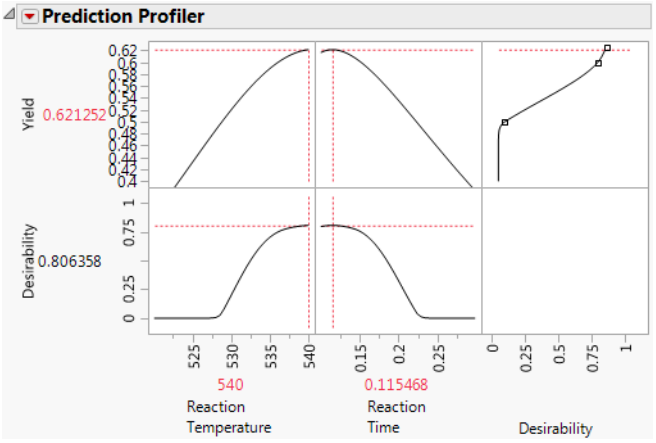
Maximize Yield Using Desirability Function

1. Select **Help > Sample Data Library** and open Stochastic Optimization.jmp.
2. Select **Graph > Profiler**.
3. Select Yield and click **Y, Prediction Formula**.
4. Click **Expand Intermediate Formulas** and then click **OK**.

   The Prediction Profiler with Desirability Functions enabled appears. For more information about Desirability Functions, see the "Desirability Profiling and Optimization".
5. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability**.

   The Prediction Profiler maximizes Yield and sets the graphs to the optimum value of Reaction Time and Reaction Temperature.

Figure 8.14 Yield Maximum



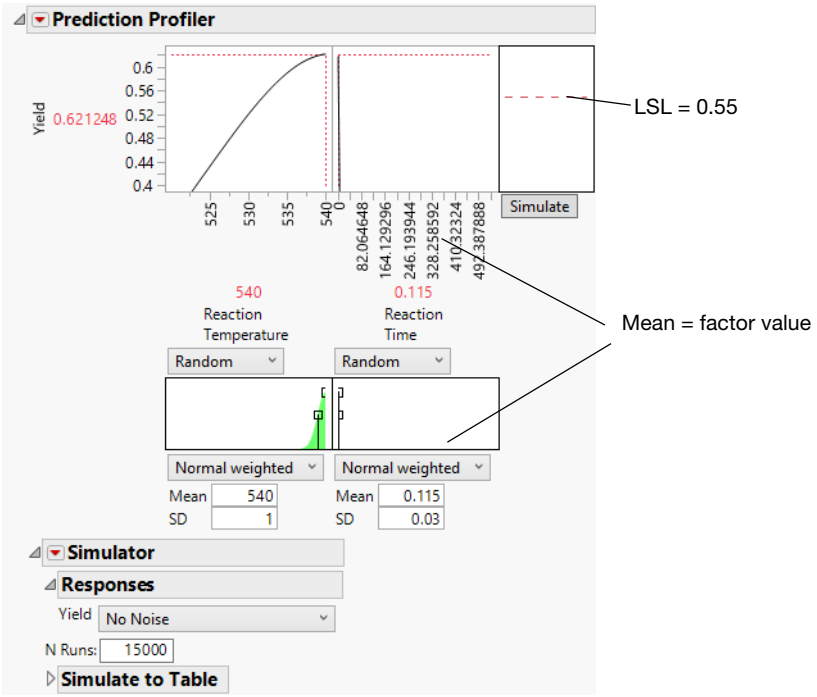
The maximum Yield is approximately 0.62 at a Reaction Time of 0.115 hours and Reaction Temperature of 540 degrees Kelvin, or hot and fast. (Your results might differ slightly due to random starting values in the optimization process.)

Simulate Defect Rate for Maximum Yield

In a production environment, process inputs cannot always be controlled exactly. What happens to Yield if the inputs (Reaction Time and Reaction Temperature) have random variation? Furthermore, if Yield has a spec limit, what percent of batches are out of spec? The Simulator can help you investigate the variation and defect rate for Yield, given variation in Reaction Time and Reaction Temperature.

1. Click the Prediction Profiler red triangle and deselect **Optimization and Desirability > Desirability Functions**.
2. Click the Prediction Profiler red triangle and select **Simulator**.
3. Set the simulation parameters for Reaction Temperature to **Random > Normal Weighted** with Mean = 540 and SD = 1.
4. Set the simulation parameters for Reaction Time to **Random > Normal Weighted** with Mean = 0.115 and SD = 0.03.
5. Set the **N Runs** value to 15,000.

Figure 8.15 Initial Simulator Settings

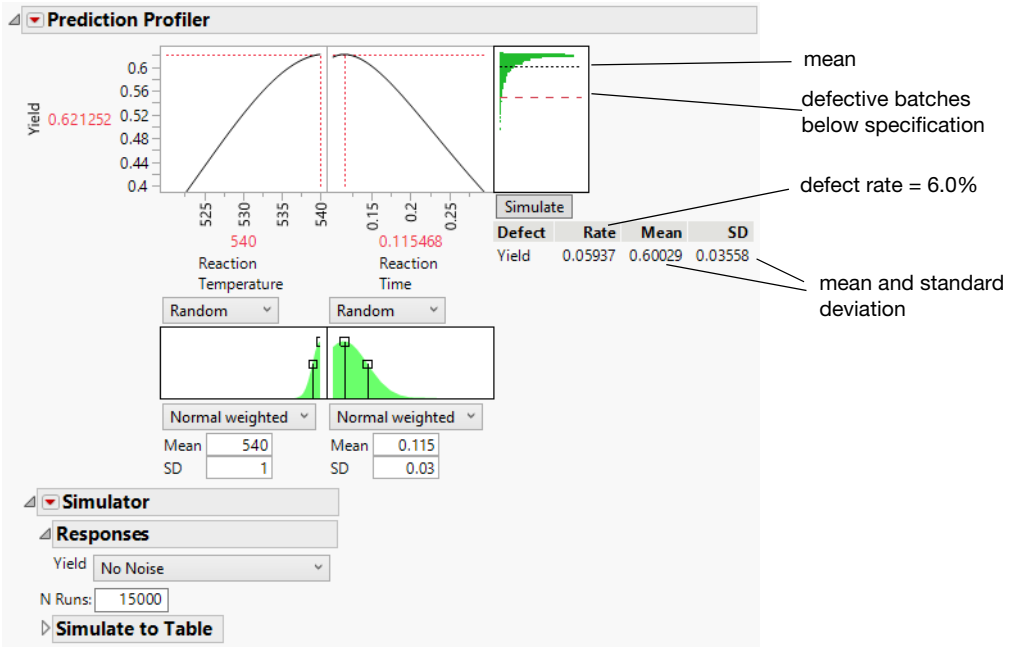


Yield has a lower spec limit of 0.55, set as a column property, and shows in Figure 8.15 as a dashed red line.

6. Click the **Simulate** button.

Note: Your numbers might differ from those shown in Figure 8.16 due to the random draws in the simulation.

Figure 8.16 Simulation Results



The predicted Yield is 0.62 for Reaction Temperature of 540 and Reaction Time of 0.115. The simulation estimates an average defect rate of about 6% with a standard deviation of 0.03 given the assumed variation in the temperature and time. The defect rate of 6.0%, indicates that about 6.0% of batches would be out of specification.

Use the simulator to explore other settings of Reaction Temperature and Reaction Time that might maintain a high Yield but with a lower defect rate. Before changing settings, save these factor settings that give us the maximum Yield for later use.

7. Click the Prediction Profiler red triangle and select **Factor Settings > Remember Settings**.
8. Type "Max Yield" and click **OK**.

The settings are appended to the report window.

Figure 8.17 Remembered Settings for Maximum Yield

Remembered Settings

	Setting	Reaction Temperature	Reaction Time	Yield
○	Max Yield	540	0.115468	0.6212525

9. Set the Mean value for Reaction Temperature to 535.

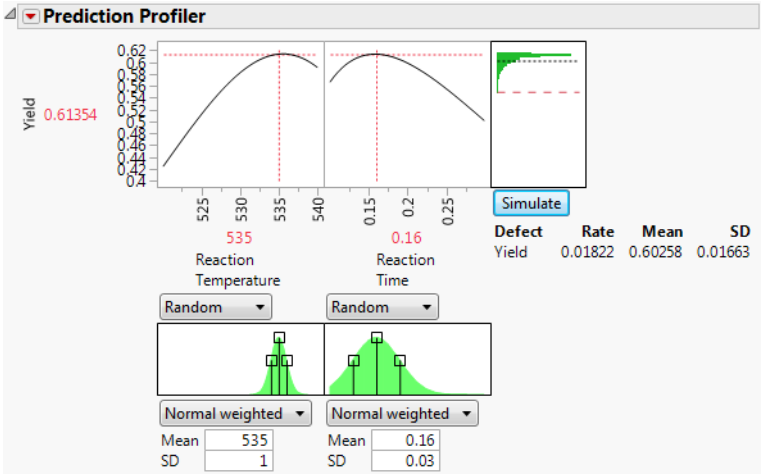
   Using the dashed red line in the Reaction Time plot, you can explore and determine the approximate value that maximized Yield. This value should be around 0.16.

10. Set the Mean value of Reaction Time to 0.16.

11. Click **Simulate**.

Note: Your numbers might differ from those shown in Figure 8.18 due to the random draws in the simulation.

Figure 8.18 Defect Rate for Temperature of 535



By making slight changes to the input factor settings, the defect rate decreases to about 1.8% with a decrease of less than 0.01 in the predicted yield. The fixed (no variability) settings that maximize Yield are not the same settings that minimize the defect rate in the presence of factor variation.

Simulation Experiment

You can run a simulation experiment to find the settings of Reaction Temperature and Reaction Time that minimize the defect rate. To do this, simulate the defect rate at each point in an experimental design for Reaction Temperature and Reaction Time. Then fit a predictive model for the defect rate and find factor settings to minimize the defect rate.

1. Click the Simulator red triangle and select **Simulation Experiment**.
2. Set the number of experimental runs to 80 and the portion of the design space to 1 to use the whole factor space in the experiment.
3. Click **OK**.

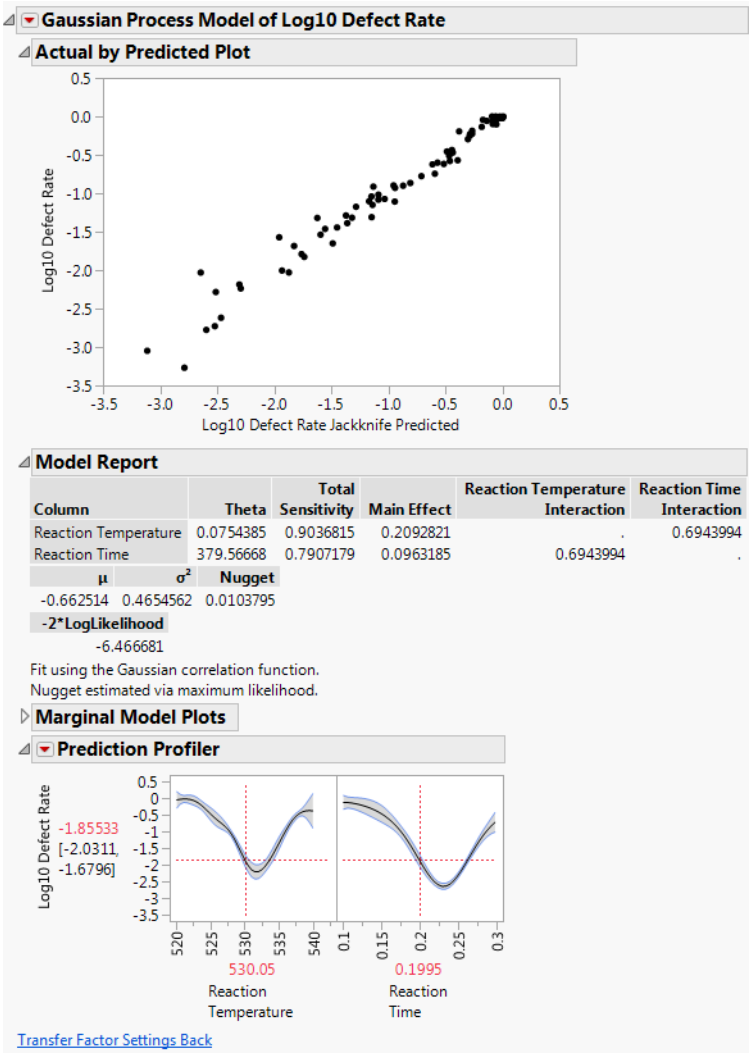
   A Latin Hypercube design with 80 design points is generated within the specified factor space, and N Runs random draws are taken at each of the design points. The design points are the center of the random draws, and the shape and variance of the random draws come from the factor distribution settings.

   A data table is created with the results of the experiment. The Overall Defect Rate is given at each design point. You can now fit a model that predicts the defect rate as a function of Reaction Temperature and Reaction Time.

   Note: Do not close the Stochastic Optimization Profiler window. You come back to it later.

4. From the new data table, click the green triangle next to the Gaussian Process script.

Figure 8.19 Results of Gaussian Process Model Fit

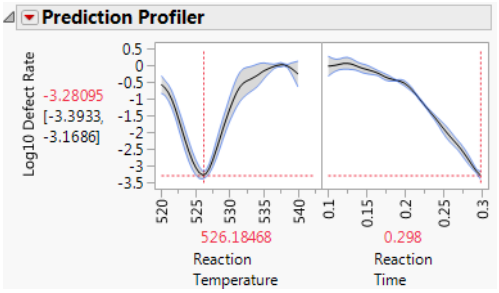


Note: Your results might be slightly different due to the random draws in the simulation.

5. To find the settings of Reaction Temperature and Reaction Time that minimizes the defect rate, click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability**.

   The desirability function is already set up to minimize the defect rate.

Figure 8.20 Settings for Minimum Defect Rate



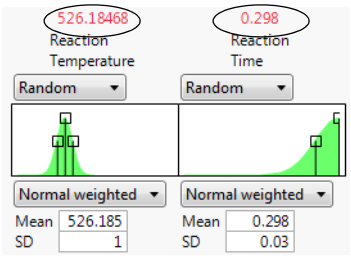
The settings that minimize the defect rate are approximately Reaction Temperature = 526 and Reaction Time = 0.3.

6. Click the **Transfer Factor Settings Back** button.

   This updates the original Profiler report to use the setting for Reaction Temperature and Reaction Time that minimize the defect rate.

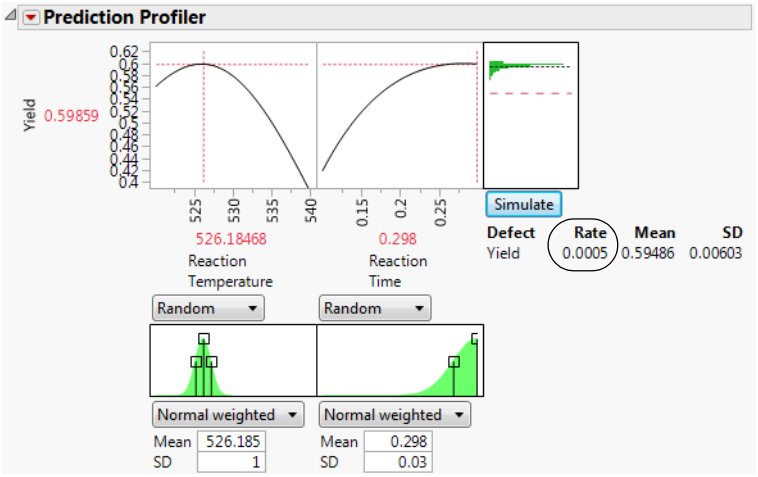
7. Return to the original Profiler report window.
8. Click the Prediction Profiler red triangle and select **Factor Settings > Remember Settings**.
9. Type "Min Defect" and click **OK**.

Figure 8.21 Minimum Defect Settings



10. With the new settings in place, click the **Simulate** button to estimate the defect rate at the new settings.

Figure 8.22 Lower Defect Rate

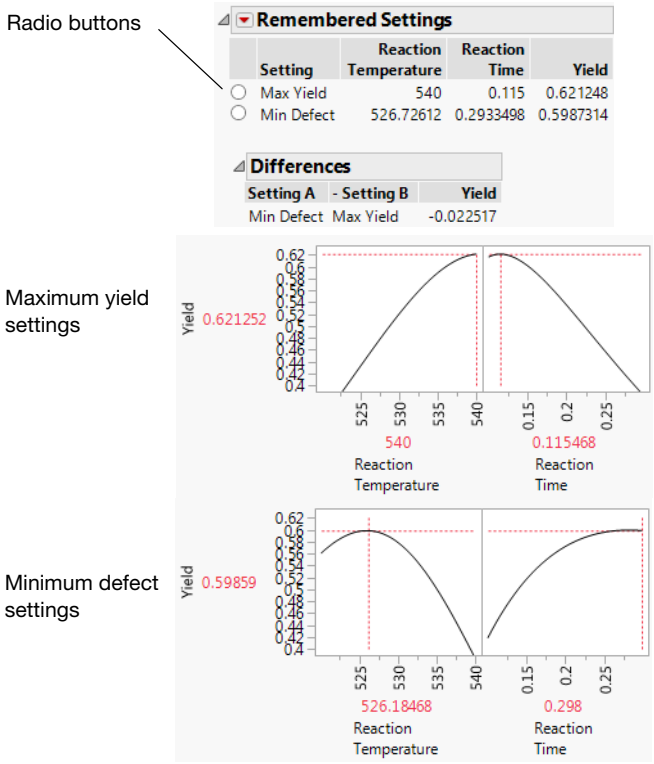


At the new settings the defect rate is 0.05%. This is much lower than the 6.0% for the settings that maximize Yield. That is a reduction of about 120x. Recall that the average Yield from the first settings is 0.62, and the new average is 0.59. The decrease in average Yield of 0.03 is the trade off for lowering the defect rate by 120x.

Because you saved the settings using Remember Settings, you can easily compare the old and new settings. The Differences report summarizes the difference.

11. Click the **Remembered Settings** radio buttons to view the profiler for each setting.

Figure 8.23 Settings Comparison



The chemist now knows what settings to use for a quality process. If the factors have no variation, the settings for maximum Yield are hot and fast. But, if the process inputs have variation similar to what we have simulated, the settings for maximum Yield produce a high defect rate. Therefore, to minimize the defect rate in the presence of factor variation, the settings should be cool and slow.

Example of Simulating General Formulas

Though the profiler and simulator are designed to work from formulas stored from a model fit, they work for any formula that is stored in a column. A typical application of simulation is to exercise financial models under certain probability scenarios to obtain the distribution of the objectives. This can be done in JMP. The key is to store the formulas into columns, set up ranges, and then conduct the simulation.

1. In the JMP Home Window, select **File > New > Script**. This opens a new script window.
2. Copy and paste the following JSL script into the new script window.

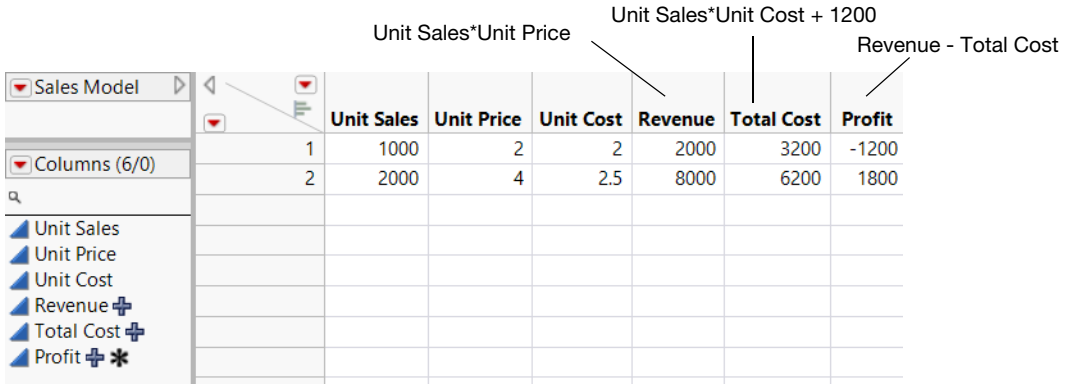
```
dt = New Table( "Sales Model" );
dt << New Column( "Unit Sales", Values( {1000, 2000} ) );
```

```
dt << New Column( "Unit Price", Values( {2, 4} ) );
dt << New Column( "Unit Cost", Values( {2, 2.5} ) );
dt << New Column( "Revenue",
   Formula( :Unit Sales * :Unit Price )
);
dt << New Column( "Total Cost",
   Formula( :Unit Sales * :Unit Cost + 1200 )
);
dt << New Column( "Profit",
   Formula( :Revenue - :Total Cost ),
   Set Property( "Spec Limits", {LSL( 0 )} )
);
Profiler(
   Y( :Revenue, :Total Cost, :Profit ),
   Objective Formula( Profit )
);
```

3. Click the Run Script icon to run the script. Alternatively, you can select Ctrl-R.

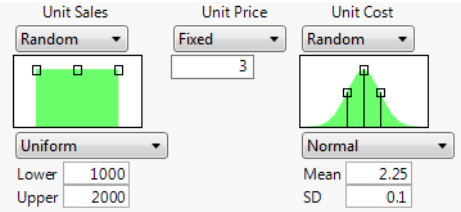
   The script creates the data table in Figure 8.24 with some initial scaling data and stores formulas into the output variables. It also launches the Prediction Profiler.

Figure 8.24 Data Table Created from Script



4. Click the Prediction Profiler red triangle and select the **Simulator**.
5. In the menus beneath Unit Sales and Unit Cost, select **Random**.
6. Fill in the factor parameters:
   - Unit Sales is **Uniform** with Lower limit 1000 and Upper limit 2000.
   - Unit Price is **Fixed** at 3.
   - Unit Cost is **Normal** with mean of 2.25 and standard deviation of 0.1.

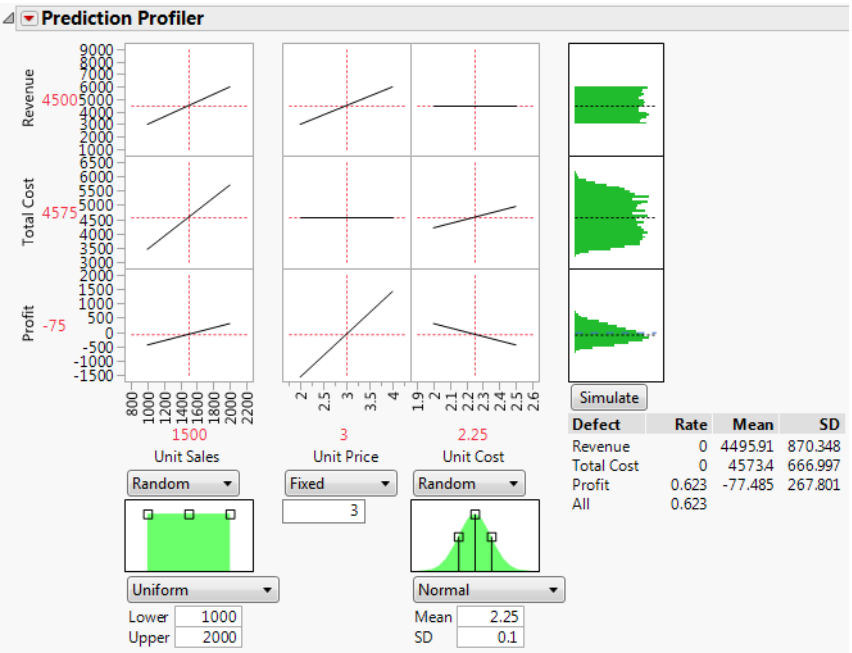
Figure 8.25 Specifications for Prediction Profiler



7. Click the **Simulate** button.

> **Note:** Your numbers might differ from those shown in Figure 8.26 due to the random draws in the simulation.

Figure 8.26 Simulator



It appears that the models are unlikely to be profitable. By putting a lower specification limit of zero on Profit, the defect report tells you that the probability of being unprofitable is 62%.

8. Change the fixed value of Unit Price to 3.25.
9. Click the **Simulate** button.

Figure 8.27 Results

Defect	Rate	Mean	SD
Revenue	0	4870.57	942.877
Total Cost	0	4573.4	666.997
Profit	0.2162	297.174	330.16
All	0.2162		

Now the probability of being unprofitable is down to about 21%.

If unit price cannot be raised anymore, you should now investigate lowering your cost or increasing sales, if you want to further decrease the probability of being unprofitable.

Statistical Details for the Simulator

Normal Weighted Distribution

JMP uses the *multivariate radial strata* method for each factor that uses the Normal Weighted distribution. This seems to work better than a number of importance sampling methods and is accurate at estimating in the extreme tails.

First, define the number of strata. The strata are a net of hyperspheres that are centered around 0. For d random factors, the strata are defined by their radial intervals. See Table 8.1.

Table 8.1 Strata Intervals

Strata Number	Inside Distance	Outside Distance
0	0	$\sqrt{d}$
1	$\sqrt{d}$	$\sqrt{d+\sqrt{2d}}$
2	$\sqrt{d+\sqrt{2d}}$	$\sqrt{d+2\sqrt{2d}}$
i	$\sqrt{d+(i-1)\sqrt{2d}}$	$\sqrt{d+i\sqrt{2d}}$
$N_{\text{Strata}}-1$	$\sqrt{d+\{(N_{\text{Strata}}-1)-1\}\sqrt{2d}}$	$\sqrt{d+(N_{\text{Strata}}-1)\sqrt{2d}}$

The default number of strata is 16. To change the number of strata, a hidden command N Strata is available if you hold the Shift key down while clicking on the red triangle next to Simulator. Increase the sample size as needed to maintain an even number of strata.

For each simulation run, the following is done:

1. Select a strata as mod($i - 1$, N_{Strata}) for run i.
2. Determine a random n-dimensional direction by scaling multivariate Normal (0,1) deviates to unit norm.
3. Determine a random distance using a chi-square quantile appropriate for the strata of a random uniform argument.
4. Scale the variates so that the norm is the random distance.
5. Scale and re-center the variates individually to be as specified for each factor.

The resulting factor distributions are multivariate normal with the appropriate means and standard deviations when estimated with the right weights. Note that you cannot use the Distribution standard deviation with weights, because it does not estimate the desired value. However, multiplying the weight by a large value, like 10^{12}, and using that as a Freq value results in the correct standard deviation.

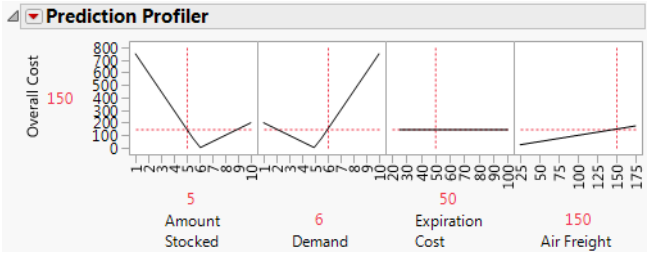
Chapter 9

Excel Profiler

Visualize Models Saved in Microsoft Excel

The JMP Add-In for Excel uses the JMP Profiler to visualize models (or formulas) stored in Excel worksheets. You can install the Excel add-in when you install JMP.

Figure 9.1 Example of a Prediction Profiler Using an Excel Model



Contents

Overview of the Excel Profiler 183
Example of an Excel Model 183
Run the JMP Profiler 185
Use Linear Constraints 186
Resolution of Profile Lines 186
Use the Excel Profiler from JMP 187

Overview of the Excel Profiler

The JMP Add-In for Excel uses the JMP Profiler to visualize models (or formulas) stored in Excel worksheets. You can choose to install the Excel Add-In when you install JMP. Profiling in the Excel Add-In is a two-step process:

1. Click the **Create/Edit Model** button (Excel 2010 through 2016) to enter information about the model that JMP needs. This needs to be done only once per model. For more information, click **Help** in the Create/Edit Model window.
2. Click the **Run Model** button (Excel 2010 through 2016) to launch the JMP Profiler and run the Excel model. See "Run the JMP Profiler".

Notes:

- The Preferences, Data Table, Graph Builder, Distribution, Fit Y by X, Fit Model, Time Series, and Control Chart buttons are not needed to profile an Excel model. For more information about these features, see *Using JMP*.
- A JMP ribbon is added to Microsoft Excel when the add-in is installed. If there is no JMP ribbon, install the add-in by double-clicking the JMP setup.exe file, selecting **Modify**, selecting **Excel Add-In**, and clicking **Next**.

Example of an Excel Model

An Excel model consists of one or more Excel formulas. Each formula must be a function of one or more other cells. This example uses the Demand.xls sample import data located in the Samples/Import Data folder.

Figure 9.2 Demand Model in Excel

Clipboard | Font | Alignment

JMP_Output_5cfcff4... | =IF(B3<=B4,(B4-B3) *B5,(B3-B4) *B6)

	A	B	C	D	E	F	G
1	JMP Profiler						
2	Label	Inputs	Min	Max	Mean		
3	Amount Stocked	5	1	10	5		
4	Demand	6	1	10	1		
5	Air Freight	150	25	175	150		
6	Expiration Cost	50	25	100	50		
7	Label	Outputs	Min	Max	Mean		
8	Overall Cost	150	0	1000	100		
9							
10							

About the Demand.xls Sample Import Data

The formula in cell B8 is a calculation of the Overall Cost associated with having different amounts of product in stock. The formula, which is shown in the Formula Bar, is a function of four cells:

- Amount Stocked is the amount of product in stock.
- Demand is the customer demand for the product.
- Air Freight is the cost per unit to ship additional product by air when the demand exceeds the amount in stock.
- Expiration Cost is the cost per unit of disposing of unused product when the demand is less than the amount in stock.

The calculations of the formula depend on the relationship between Amount Stocked and Demand.

- If Amount Stocked is less than Demand, then the company has to ship additional units, at a cost of (Demand-Amount Stocked) × Air Freight. For example, if the demand is 8, but the company has only 6 in stock, then it has to ship 8 - 6 = 2 units at a cost of 2 × 150 = 300.
- If Amount Stocked is greater than Demand, then the company has to dispose of unused product, at a cost of (Amount Stocked - Demand) × Expiration Cost. For example, if the demand is 5, but the company has 8 in stock, then it has to dispose of 8 - 5 = 3 units at a cost of 3 × 50 = 150.
- If Amount Stocked is equal to Demand, then there is no shipping cost or disposal cost.
- There is never both a shipping cost and a disposal cost at the same time.

Create the Model in Excel

1. Select **Help > Sample Data Library** and navigate up one level to the Samples/Import Data folder.
2. Double-click Demand.xls to open the file in Microsoft Excel.
3. In Microsoft Excel, click the JMP ribbon.
4. Click the **Create/Edit Model** button .

   The name of the workbook is displayed in the Model and Model Name fields.

   The Inputs and Outputs fields are populated with data from the worksheet.
5. Enter Customer Demand in the Model Name field and click **Apply**.

   The Model field is updated.
6. Select **Air Freight** in the Inputs box and then click the down arrow button .

   Air Freight is moved to the bottom of the list because you want it displayed last in the Profiler.

7. Click **OK**.

   The Excel model is saved to the worksheet.

Notes:

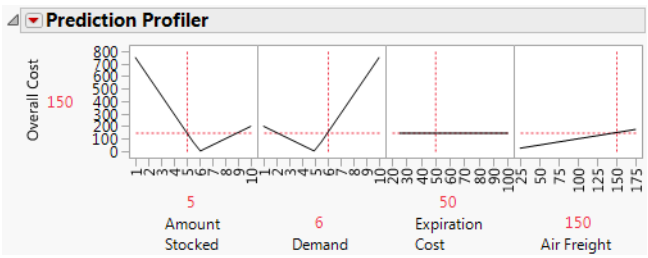
- If the fields in the Create/Edit Model window are not populated when you set up your model, click **Choose** and select the cell that contains the input or output name.
  - For inputs, specify the values and click **Apply**. Inputs must be values, not formulas.
  - For outputs, the specified cell must be a formula containing only information from the Input cells.
- You can create more than one model for a worksheet. In the Create/Edit Model window, click the plus button + next to the Model name. In the Model Name field, change the name if necessary and click the **Apply** button. You can then change the inputs and outputs as necessary and click **OK**.
- You must define the entire model on one worksheet. A model cannot reference cells on another worksheet.

Run the JMP Profiler

Once you create the model using the Excel Add-In, you can run it in the JMP Profiler. From the Excel Add-In, perform the following actions:

1. In Microsoft Excel, click the JMP ribbon.
2. Click the **Run Model** button.
3. Select the model that you want to run.
4. Click **Profile in JMP**.
5. Use the JMP Profiler to simultaneously see the effect of all inputs on the output. You can also simulate a range of input combinations to see the resulting range of output values.

Figure 9.3 Example of the Prediction Profiler Using Excel Models



Note: To ensure that your original Excel worksheet is not altered, JMP runs a hidden copy of Excel in the background that controls all of the Profiler calculations. Since the Excel worksheet and the Profiler are not dynamically linked, after you update a formula in the worksheet, you must relaunch the JMP Profiler to reflect the changes.

Use Linear Constraints

Within the JMP Profiler, you can alter the linear constraints in order to restrict the model input values. You are prompted to save the constraints to the Excel workbook. After constraints are saved to the Excel workbook, whenever the model is profiled from the Excel Add-In, the constraints are incorporated.

1. Click the Prediction Profiler red triangle and select **Alter Linear Constraints**.
2. Click **Add Constraint**.
3. Type in the constraining values.
4. Click **OK**.
5. Click the Prediction Profiler red triangle and select **Save Linear Constraints**.

   You are prompted to save the constraints to the Excel workbook.
6. Click **Yes**.

Note: When you save the .xls file, you might see a compatibility error. If so, click **Continue** to save the file.

The workbook opens in Excel. When you run the model, the constraints are reflected in the JMP Profiler. For more information about linear constraints, see "Linear Constraints".

Tip: To delete a linear constraint, set all constraint values to zero.

Resolution of Profile Lines

The Default N Levels option on the Prediction Profiler red triangle menu affects the resolution of the profile lines. Note the following information:

- This option defaults to 17 when the Profiler runs a model stored in Excel.
- This option defaults to 41 when the model is stored directly in JMP.

If the same model is stored in both Excel and JMP, then the profile lines can appear differently when the models are profiled. Increasing this value causes the Excel Profiler to run slower.

Use the Excel Profiler from JMP

After you have defined model input and outputs in an Excel file, you can profile the model from within JMP.

1. Select **Graph** > **Excel Profiler**.
2. Locate the Excel file containing the model and then click **Open**.
3. If the Excel file contains multiple models, you are prompted to select the model that you want to profile.

Use the following code to script the Excel Profiler:

```
Excel Profiler( "path to workbook", <"model name"> ) ;
```

If more than one model exists, and no model is specified, a window with the list of available models appears. For more information about scripting the Excel Profiler, see the *Scripting Guide*.

Appendix A

References

Box, G. E. P., and Draper, N. R. (1987). *Empirical Model–Building and Response Surfaces*. New York: John Wiley & Sons.

Box, G. E. P., and Draper, N. R. (2007). *Response Surfaces, Mixtures, and Ridge Analyses*. 2nd ed. New York: John Wiley & Sons.

Cornell, J. A. (1990). *Experiments with Mixtures*. 2nd ed. New York: John Wiley & Sons.

Derringer, G., and Suich, R. (1980). "Simultaneous Optimization of Several Response Variables." *Journal of Quality Technology* 12.4:214–219.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer-Verlag.

Saltelli, A. (2002). "Making best use of model evaluations to compute sensitivity indices." *Computer Physics Communications* 145:280–297.

Schafer, J., and Strimmer, K. (2005). "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics." *Statistical Applications in Genetics and Molecular Biology* 4:1–30.

Sobol, I. M. (1993). "Sensitivity Estimates for Nonlinear Mathematical Models." *MMCE* 1.4:407–414.

Appendix B

Technology License Notices

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

  - You must give any other recipients of the Work or Derivative Works a copy of this License; and
  - You must cause any modified files to carry prominent notices stating that You changed the files; and
  - You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
  - If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
  - You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

버전 16

실험 설계 가이드

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는
것이 아니라 새로운 눈을 가지는 데 있다."
Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.0

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 16 실험 설계 가이드 . Cary, NC: SAS Institute Inc.

JMP® 16 실험 설계 가이드

Copyright © 2020-2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP 를 처음 사용하든 오랫동안 사용해왔든 , JMP 와 관련해서 배워야 하는 것은 언제나 있습니다 .

JMP.com 을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다 .

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP 를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티 , 추가 기능 및 스크립트 예를 포함한 사용자용 리소스 , 포럼 , 블로그 , 컨퍼런스 정보 등

https://www.jmp.com/getstarted

목차
실험 설계 가이드

1 JMP 알아보기 21
설명서 및 추가 리소스
JMP 설명서에 사용되는 서식 규칙 23
JMP 도움말 24
JMP 설명서 라이브러리 24
JMP 학습을 위한 추가 리소스 30
JMP 자습서 30
샘플 데이터 테이블 30
통계 및 JSL 용어에 대해 알아보기 31
JMP 팁 및 힌트 알아보기 31
JMP 툴팁 31
JMP 사용자 커뮤니티 31
무료 온라인 Statistical Thinking 교육 과정 32
JMP 새로운 사용자를 위한 입문 키트 32
Statistics Knowledge 포털 32
JMP 교육 과정 32
사용자가 작성한 JMP 설명서 32
JMP 시작하기 창 32
JMP 기술 지원 33

2 DOE 소개 35
실험 설계 플랫폼 개요

3 DOE 로 시작 39
예제 및 핵심 개념
실험 설계 및 DOE 워크플로우 개요 41
커피 강도 실험 42
연구 및 목표 정의 42
설계 생성 44
실험 실행 49
데이터 분석 50
DOE 워크플로우 : 기술 , 지정 , 설계 56

반응 및 요인 정의 57
모형 지정 58
설계 생성 58
설계 평가 59
테이블 생성 60
실험 설계의 원리와 지침 60
효과 계층성 60
효과 유전성 61
효과 희소성 61
중앙점, 반복 런 및 검정 61

4 사용자 설계 63
요구 사항을 충족하는 설계 구성
사용자 설계 개요 65
사용자 설계의 예 66
사용자 설계 생성 82
반응 82
요인 84
요인 제약 조건 정의 90
별칭 항 94
설계 생성 95
설계 97
설계 평가 97
출력 옵션 98
사용자 설계 옵션 100
반응 시뮬레이션 103
행렬 X 저장 105
시작 수 107
설계 검색 시간 108
사용자 설계에 대한 기술 상세 정보 108
랜덤화 제한이 있는 설계 109
변경하기 힘든 수준의 공변량 117
주구 및 하위구 수 117
최적 기준 118
D 효율도 122
좌표 교환 알고리즘 122

5 사용자 설계의 예 123
요구 사항을 충족하는 실험 수행
선별 실험 125

주효과만추정하는 설계 125
모든 2 요인 교호작용을 추정하는 설계 128
주효과와 2 요인 교호작용의 별칭 관계를 방지하는 설계 130
과포화 선별 설계 133
고정 블록 설계 140
반응 표면 실험 144
반응 표면 설계 144
유연한 블록화 방법을 사용하는 반응 표면 설계 150
D- 최적과 I- 최적의 반응 표면 설계 비교 152
제약 조건 및 범주형 요인이 있는 반응 표면 설계 157
혼합물 실험 162
비혼합 요인이 있는 혼합물 설계 162
혼합물 혼합 설계 166
공변량이 포함된 실험 170
고정 공변량이 포함된 설계 171
변경하기 힘든 공변량이 포함된 설계 173
시간 선형 추세가 포함된 설계 177
랜덤화 제한이 있는 실험 180
분할구 실험 180
이원 분할구 실험 184
로버스트 공정 및 제품 설계를 위한 실험 190

6 설계 확대 193

기존 설계에 추가 런 추가

설계 확대 플랫폼을 사용한 예 195
설계 확대 플랫폼 시작 201
설계 확대 창 202
요인 202
요인 제약 조건 정의 203
확대 옵션 206
설계 확대 플랫폼 옵션 207
확대 옵션의 추가 예 210
설계 반복 210
설계에 중앙점 추가 212
설계 접기 생성 213
설계에 축 점 추가 214
공간 채움 확대 215
통계 상세 정보 216

7 확증적 선별 설계 219

확증적 선별 설계 개요 221
확증적 선별 설계의 예 223
확증적 선별 설계 223
부분 요인 설계와 비교 226
블록화를 사용한 확증적 선별 설계 228
확증적 선별 설계와 Plackett-Burman 설계 비교 232
확증적 선별 설계 창 235
반응 236
요인 237
설계 옵션 239
설계 241
설계 평가 241
출력 옵션 242
확증적 선별 설계 옵션 244
반응 시뮬레이션 246
기술 상세 정보 248
확증적 선별 설계의 구조 248
실험 데이터 분석 250

8 확증적 선별 적합 플랫폼 253

확증적 선별 실험의 데이터 분석

확증적 선별 적합 플랫폼 개요 255
DSD 에서 활성 효과 식별 255
DSD 를 위한 유효 모형 선택 255
확증적 선별 적합 플랫폼의 예 256
모형 적합 256
결과 검토 256
모형 축소 258
확증적 선별 적합 플랫폼 시작 260
확증적 선별 적합 보고서 260
1 단계 - 주효과 추정값 261
2 단계 - 짝수 차수 효과 추정값 262
결합 모형 모수 추정값 262
주효과 잔차 그림 263
예측 프로파일러 263
확증적 선별 적합 플랫폼 옵션 263
확증적 선별 적합 플랫폼에 대한 기술 상세 정보 264
DSD 를 위한 유효 모형 선택 알고리즘 264

9 선별 설계 267

선별 설계 개요 269
기본 원리 269
선별 설계 결과 분석 269
선별 설계의 예 270
부분 요인 설계와 주효과 선별 설계 비교 270
표준 설계가 없는 경우의 주효과 선별 설계 277
선별 설계 창 279
반응 280
요인 282
선별 유형 선택 284
부분 요인 설계 목록에서 선택 285
설계 유형 285
교락 측도를 나타내는 해상도 288
설계 표시 및 수정 289
주효과 선별 설계 291
설계 생성 292
설계 293
설계 평가 293
출력 옵션 293
테이블 생성 294
선별 설계 옵션 295
선별 설계의 추가 예 298
부분 요인 설계에서 생성 규칙 수정 298
Plackett-Burman 설계 303

10 2 수준 선별 적합 플랫폼 309
선별 실험의 데이터 분석

2 수준 선별 적합 플랫폼 개요 311
2 수준 선별 적합 플랫폼의 예 311
2 수준 선별 적합 플랫폼 시작 314
선별 보고서 315
대비 315
1/2 정규 확률도 316
모형 생성 또는 실행 316
2 수준 선별 적합 분석의 추가 예 316
Plackett-Burman 설계 분석 317
효과 유전성 예 318
과포화 설계 분석 320

2 수준 선별 적합 플랫폼에 대한 기술 상세 정보 321
효과 입력 순서 322
2 수준 선별 적합과 직교 회전 322
Lenth 유사 표준 오차 323
Lenth t- 비 323

11 반응 표면 설계 327

반응 표면 설계 개요 329
반응 표면 설계의 예 331
Box-Behnken 설계 구성 332
실험 데이터 분석 333
최적 설정 탐색 335
반응 표면 설계 창 338
반응 339
요인 340
설계 선택 342
출력 옵션 지정 343
테이블 생성 344
반응 표면 설계 옵션 345

12 완전 요인 설계 347

완전 요인 설계 개요 349
완전 요인 설계의 예 349
설계 구성 349
실험 데이터 분석 351
완전 요인 설계 생성 355
반응 355
요인 357
출력 옵션 선택 359
테이블 생성 360
완전 요인 설계 옵션 361

13 혼합물 설계 365

배합 설계

혼합물 설계 개요 367
혼합물 설계 창 367
반응 367
요인 369
삼원 그림 개요 371

혼합물 설계 유형의 예 371
최적 혼합물 설계 372
심플렉스 중심 설계 373
심플렉스 격자 설계 375
꼭지점 설계 376
ABCD 설계 380
공간 채움 설계 381
혼합물 설계 적합 383
전체 모형 검정 및 분산 분석 보고서 384
반응 표면 보고서 이해 384
화학 혼합물 예 384
설계 생성 385
혼합 모형 분석 386
예측 프로파일러 387
혼합물 프로파일러 388

14 Taguchi 설계 391
잡음 요인이 있는 설계
Taguchi 설계 개요 393
Taguchi 설계의 예 393
Taguchi 설계 생성 397
반응 397
요인 398
Taguchi 내측 및 외측 배열 설계 선택 398
코드화된 설계 표시 399
설계 테이블 생성 399
Taguchi 설계 옵션 400

15 설계 평가 403
설계 특성 탐색
설계 평가 플랫폼 개요 405
설계 평가의 예 405
손실된 런의 영향 평가 406
지정된 모형에 대한 검정력 평가 413
설계 평가 시작 창 415
설계 평가 창 415
요인 416
모형 417
별칭 항 418
설계 418

설계 평가 418
검정력 분석 419
예측 분산 프로파일 425
설계 공간 비율 그림 426
예측 분산 표면 427
추정 효율도 428
별칭 행렬 430
상관 색상 맵 431
설계 진단 432
설계 평가 옵션 434

16 설계 비교 437
설계를 동시에 비교 및 평가
설계 비교 플랫폼 개요 439
설계 비교의 예 439
런 크기가 같은 설계 439
런 크기가 다른 설계 444
주구 수가 다른 분할구 설계 447
설계 비교 플랫폼 시작 451
설계 비교 창 : 모형 및 별칭 항 지정 453
참조 설계 454
요인 454
모형 454
별칭 항 454
설계 비교 창 : 설계 평가 454
검정력 분석 455
예측 분산 프로파일 458
설계 공간 비율 그림 460
상대 추정 효율도 461
별칭 행렬 요약 463
절대 상관 464
설계 진단 466
설계 비교 옵션 468

17 예상 표본 크기 및 검정력 469

표본 크기 및 검정력 플랫폼 시작 472
1 표본 평균 계산기 473
1 표본 평균 계산기의 예 473
1 표본 평균 계산기 필드 477

1 표본 평균 계산기에 대한 통계 상세 정보 478
2 표본 평균 계산기 479
2 표본 평균 계산기의 예 479
2 표본 평균 계산기 필드 480
2 표본 평균 계산기에 대한 통계 상세 정보 481
k 표본 평균 계산기 481
k 표본 평균 계산기의 예 482
k 표본 평균 계산기 필드 484
k 표본 평균 계산기에 대한 통계 상세 정보 485
1 표본 표준편차 계산기 486
1 표본 표준편차 계산기의 예 486
1 표본 표준편차 계산기 필드 487
1 표본 표준편차 계산기에 대한 통계 상세 정보 488
1 표본 비율 계산기 489
1 표본 비율 계산기의 예 489
1 표본 비율 계산기 필드 490
단일 비율 계산기에 대한 통계 상세 정보 491
2 표본 비율 계산기 492
2 표본 비율 계산기의 예 492
2 표본 비율 계산기 필드 493
두 비율 계산기에 대한 통계 상세 정보 494
단위별 개수 계산기 494
단위별 개수 계산기의 예 495
단위별 개수 계산기 필드 496
단위별 개수 계산기에 대한 통계 상세 정보 496
시그마 품질 수준 계산기 497
시그마 품질 수준 계산기의 예 497
결함 수 계산의 예 498
시그마 품질 수준 계산기 필드 498
시그마 품질 계산기에 대한 통계 상세 정보 499
신뢰도 시험 계획 계산기 499
신뢰도 시험 계획 계산기의 예 499
신뢰도 시험 계획 계산기 필드 501
신뢰도 시험 계획 계산기에 대한 통계 상세 정보 502
신뢰도 데모 계산기 503
신뢰도 데모 계산기의 예 503
신뢰도 데모 계산기 필드 504
신뢰도 데모 계산기에 대한 통계 상세 정보 505

18 선택 설계 507

기본 제품 프로파일 선택을 위한 설계 생성

선택 설계 개요 509
선택 설계 용어 509
베이지안 D- 최적 509
선택 설계의 예 510
분석을 사용한 선택 설계의 예 512
파일럿 연구를 위한 선택 설계 생성 512
파일럿 연구 데이터 분석 514
사전 정보를 사용하여 최종 선택 실험 설계 516
설계 실행 및 결과 분석 517
선택 설계 창 520
속성 520
모형 521
설계 생성 523
설계 524
테이블 생성 525
선택 설계 옵션 525
기술 상세 정보 526
베이지안 D- 최적 및 설계 구성 526
유틸리티 중립 설계 및 로컬 D- 최적 설계 527

19 최대차이 설계 529

최선 - 최악 항목 선택을 위한 설계 생성

최대차이 설계 플랫폼 개요 531
최대차이 설계의 예 531
최대차이 설계 시작 창 535
최대차이 창 535
설계 옵션 개요 535
설계 개요 536
쌍별 따름 행렬 536
테이블 생성 536
최대차이 옵션 536

20 덮기 배열 539

성분 교호작용으로 인한 고장 감지

덮기 배열 개요 541
요인 수준 제한이 없는 덮기 배열의 예 542
설계 생성 543
실험 데이터 분석 545

요인 수준 제한이 있는 덮기 배열의 예 547
설계 생성 547
실험 데이터 분석 552
덮기 배열 창 553
요인 553
요인 수준 조합 제한 555
설계 557
지표 559
출력 옵션 560
덮기 배열 데이터 테이블 560
덮기 배열 옵션 561
덮기 배열에 대한 기술 상세 정보 562
최적화 알고리즘 562
지표 계산식 563

21 공간 채움 설계 565

공간 채움 설계 개요 567
공간 채움 설계 생성 568
반응 568
요인 570
요인 제약 조건 정의 572
공간 채움 설계 방법 575
설계 575
설계 진단 575
설계 테이블 576
공간 채움 설계 옵션 576
Sphere-Packing 설계 578
Sphere-Packing 설계 생성 578
Sphere-Packing 설계 시각화 580
라틴 초입방체 설계 581
라틴 초입방체 설계 생성 581
라틴 초입방체 설계 시각화 582
균등 설계 584
Sphere-Packing, 라틴 초입방체 및 균등 방법 비교 586
최소 잠재 설계 587
최대 엔트로피 설계 589
가우스 과정 IMSE 최적 설계 591
쾌속 유연 채움 설계 592
FFF 최적 기준 592

FFF 설계의 평균 군집 크기 설정 593
제약 조건 594
제약된 쾌속 유연 채움 설계 생성 및 보기 594
맵 쉐이프를 위한 공간 채움 설계 생성 596
Sphere-Packing 설계의 예 601
Borehole 데이터에 대한 Sphere-Packing 설계 생성 601
결정적 데이터 분석을 위한 지침 602

22 가속 수명 시험 설계 607

가속 수명 시험을 위한 실험 설계

가속 수명 시험 설계 개요 609
가속 수명 시험 설계의 예 609
가속 수명 시험 설계 확대의 예 616
가속 수명 시험 설계 생성 620
설계 구조 지정 621
가속 요인 지정 621
설계 상세 정보 지정 622
ALT 계획에 대한 규격 검토 및 업데이트 624
최적 설계 생성 및 평가 625
설계 테이블 생성 627
가속 수명 시험의 옵션 628
ALT 설계 플랫폼에 대한 통계 상세 정보 628
ALT 설계의 고장 분포 628
R 정밀도 요인 (95% CI) 630
가속 수명 시험에 대한 최적 기준 631

23 비선형 설계 633

비선형 설계 개요 635
비선형 설계의 예 636
사전 데이터가 없는 비선형 설계 생성 636
사전 데이터를 사용하여 설계 확대 640
이항 반응에 대한 설계 생성 644
비선형 설계 시작 창 647
비선형 설계 창 647
요인 648
모수 648
설계 생성 649
설계 649
테이블 생성 또는 확대 649

비선형 설계 옵션 649
비선형 설계에 대한 통계 상세 정보 650
비선형 모형 650
최적 기준의 방사형 - 구형 적분 651
최적 설계 찾기 651

24 균형 불완비 블록 설계 653

균형 불완비 블록 설계의 예 655
균형 불완비 블록 설계 생성 657
처리, 블록 지정 및 설계 생성 657
설계 확인 및 테이블 생성 658
균형 불완비 블록 설계 옵션 658
균형 불완비 블록 설계에 대한 통계 상세 정보 659

25 MSA 설계 661

측정 시스템 분석 실험에 대한 설계 생성

MSA 설계 개요 663
MSA 설계의 예 663
MSA 설계 생성 665
반응 666
요인 668
MSA 설계 668
MSA 설계 진단 668
MSA 설계 테이블 669
MSA 설계 옵션 670
MSA 설계 플랫폼의 통계 상세 정보 671

26 그룹 직교 과포화 설계 673

그룹 직교 과포화 설계 생성

그룹 직교 과포화 설계 개요 675
그룹 직교 과포화 설계의 예 675
그룹 직교 과포화 설계 생성 676
요인, 구조 지정 및 설계 생성 677
설계 확인 및 테이블 생성 678
그룹 직교 과포화 설계 옵션 678
그룹 직교 과포화 설계에 대한 통계 상세 정보 679

27 그룹 직교 과포화 설계 적합 플랫폼 681

그룹 직교 과포화 설계의 데이터 분석

그룹 직교 과포화 설계 적합 플랫폼 개요 683

그룹 직교 과포화 설계 적합의 예 683
그룹 직교 과포화 설계 적합 플랫폼 시작 685
그룹 직교 과포화 설계 적합 보고서 686
그룹 직교 과포화 설계 적합 플랫폼 옵션 687

28 표본 크기 탐색기 689

표본 크기 및 검정력 계산기

표본 크기 탐색기 플랫폼에 대한 개요 691
표본 크기 탐색기의 예 691
표본 크기 탐색기 시작 692
가설 검정에 대한 검정력 탐색기 692
1 표본 평균에 대한 검정력 693
1 표본 비율에 대한 검정력 695
1 표본 분산에 대한 검정력 697
1 표본 동등성에 대한 검정력 699
2 독립 표본 평균에 대한 검정력 701
2 독립 표본 비율에 대한 검정력 704
2 독립 표본 분산에 대한 검정력 706
2 독립 표본 동등성에 대한 검정력 708
ANOVA 에 대한 검정력 710
신뢰 구간 계산기 712
1 표본 평균에 대한 오차 한계 712
1 표본 비율에 대한 오차 한계 715
1 표본 분산에 대한 오차 한계 716
2 독립 표본 평균에 대한 오차 한계 718
2 독립 표본 비율에 대한 오차 한계 720
2 독립 표본 분산에 대한 오차 한계 722
신뢰도 데모 계산기 723
모수 신뢰도 데모 724
비모수 신뢰도 데모 725
표본 크기 탐색기의 저장된 설정 726

A 열 특성 727

DOE 에서 할당하는 열 특성 이해

열 특성 추가 및 보기 729
반응 한계 730
반응 한계의 예 731
반응 한계 편집 734
감지 한계 737
함수 반응 737

설계 역할 737
설계 역할의 예 738
코딩 739
하한값과 상한값 740
코딩 열 특성 및 중심화 다항식 741
코딩의 예 741
코딩 할당 744
혼합 747
혼합의 예 749
요인 변경 752
요인 변경의 예 753
값 순서 755
값 순서의 예 756
값 순서 할당 758
값 라벨 759
값 라벨의 예 759
블록당 런 수 761
블록당 런 수의 예 761
제약 조건 상태 762
제약 조건 상태의 예 762

B 기술 상세 정보 765

별칭 행렬 767
변경하기 힘들거나 매우 변경하기 힘든 요인이 있는 설계 767
If Possible 효과가 있는 설계 768
검정력 계산 768
단일 모수에 대한 검정력 768
범주형 효과에 대한 검정력 769
상대 예측 분산 770

C 참조 자료 773

D 기술 라이선스 표시 779

1 장

JMP 알아보기
설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 23
JMP 도움말 24
JMP 설명서 라이브러리 24
JMP 학습을 위한 추가 리소스 30
JMP 자습서 30
샘플 데이터 테이블 30
통계 및 JSL 용어에 대해 알아보기 31
JMP 팁 및 힌트 알아보기 31
JMP 툴팁 31
JMP 사용자 커뮤니티 31
무료 온라인 Statistical Thinking 교육 과정 32
JMP 새로운 사용자를 위한 입문 키트 32
Statistics Knowledge 포털 32
JMP 교육 과정 32
사용자가 작성한 JMP 설명서 32
JMP 시작하기 창 32
JMP 기술 지원 33

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 Lucida Sans Typewriter 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 "JMP 설명서 라이브러리 "(24 페이지) 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
Quality and Process Methods	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기" 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기

" 창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

2 장

DOE 소개
실험 설계 플랫폼 개요

JMP DOE 플랫폼은 실험을 설계, 평가 및 분석할 수 있도록 도와 줍니다. 대부분의 플랫폼은 설계를 구성하는 데 초점을 맞춥니다. 그 외의 다른 플랫폼은 설계 작업을 지원합니다. 이 섹션에서는 DOE 메뉴에 있는 각 플랫폼에 대해 개략적으로 살펴 봅니다.

설계 구성 플랫폼

사용자 설계 다양한 설정에 맞는 설계를 구성합니다. 사용자 설계는 전통적 설계에만 기반한 방법보다 비용 효율적이고 더 유연합니다.

사용자 설계는 다양한 유형의 요인, 제약 조건, 허용되지 않는 조합 등을 수용합니다. 주어진 런 수를 고려하여 추정이 필요한 효과 및 적합한 효과를 지정할 수 있습니다. 실험 환경의 예산에 맞도록 런 수를 지정할 수 있습니다. 또한 사용자 설계는 변경하기 힘든 요인 또는 매우 변경하기 힘든 요인을 지원하므로 분할구 설계 및 관련 설계를 구성할 수 있습니다.

사용자 설계 플랫폼은 다음과 같은 여러 가지 특수 설계 유형을 구성합니다.

- 선별
- 반응 표면
- 혼합물
- 랜덤 블록
- 분할구
- 분할 - 분할구
- 이원 분할구

다른 플랫폼을 사용하여 전통적 선별, 반응 표면 및 혼합물 설계를 구성할 수도 있습니다. 그러나 사용자 설계 플랫폼은 다른 플랫폼에서 사용할 수 없는 유연성을 제공합니다 . 분할구 설계는 사용자 설계 플랫폼을 통해서만 구성할 수 있습니다.

확증적 선별 설계 연속형 요인 및 2 수준 범주형 요인에 대한 선별 설계를 구성합니다. 확증적 선별 설계는 활성 교호작용 또는 곡률이 의심되는 경우에 유용합니다. 확증적 선별 설계를 사용하면 2 차까지의 효과 사이에 완전 교락을 방지하면서 강력한 비선형 효과의 원인을 식별할 수 있습니다.

확증적 선별 설계는 요인이 4 개 이상 있는 실험에 가장 적합합니다. 확증적 선별 설계에서는 런을 블록으로 그룹화할 수 있습니다. 블록 수는 사용자가 지정합니다.

선별 설계 임의 개수의 수준으로 연속형, 이산 수치형 및 범주형 요인에 대한 선별 설계를 구성합니다. 표준 설계가 있는 경우 다음과 같은 두 가지 옵션을 사용할 수 있습니다.

– 전통적 선별 설계 목록에서 선택합니다 . 이 설계는 2 수준 연속형 요인 , 2 수준 /3 수준 범주형 요인 또는 이산 연속형 요인을 허용합니다 .

– 주효과에 대한 직교 설계 또는 근사 직교 설계를 생성합니다 . **근사 직교 설계**는 2 수준 연속형 요인뿐 아니라 원하는 수의 수준으로 범주형 및 이산 수치형 요인을 허용합니다 . 이 설계는 무시할 수 있는 교호작용이 있는 경우 주효과를 추정하는 데 중점을 둡니다 .

대부분의 선별 상황에서 표준 설계를 사용할 수 없습니다 . 이러한 경우 근사 직교 선별 설계를 구성할 수 있습니다 .

반응 표면 설계 연속형 요인의 2 차 함수를 모델링하는 설계를 구성합니다 . 2 차 효과를 적합시키려면 반응 표면 설계에 각 요인에 대한 세 가지 설정이 필요합니다 . JMP 에서는 최대 8 개 요인에 대한 반응 표면 설계를 제공합니다 .

중심 합성 설계 또는 Box-Behnken 설계 목록에서 선택할 수 있습니다 . 필요한 경우 직교로 블록화하는 중심 합성 설계가 목록에 포함됩니다 . 중심 합성 설계에 대한 다양한 수정도 지원됩니다 .

완전 요인 설계 여러 개의 연속형 요인 또는 범주형 요인에 대한 완전 요인 설계를 구성합니다 . 두 경우 모두 수준 수에 제한이 없습니다 . 완전 요인 설계에서는 요인 설정 조합이 있을 때마다 런이 발생합니다 . 완전 요인 설계는 큰 경향이 있습니다 . 런 수는 요인 수준 수의 곱과 같습니다 .

혼합물 설계 요인이 혼합 내 성분일 때 사용하는 설계를 구성합니다 . 혼합물 실험에서 한 성분의 비율을 변경하려면 나머지 성분 중 하나 이상을 변경하여 합계를 유지해야 합니다 . 일부 전통적 혼합물 설계 방법을 포함하여 최적 , 심플렉스 중심 , 심플렉스 격자 , 꼭지점 , ABCD, 공간 채움 등 여러 설계 유형 중에서 선택할 수 있습니다 . 최적 , 꼭지점 및 공간 채움 혼합물 설계의 경우 선형 부등식 제약 조건을 지정하여 설계 공간을 제한할 수 있습니다 .

Taguchi 배열 신호 대 잡음 분석에 사용하는 설계를 구성합니다 . 이 설계는 Taguchi 의 내측 및 외측 배열 방법을 기반으로 합니다 . 제어 요인 설정은 내측 배열을 구성하고 잡음 요인 설정은 외측 배열을 구성합니다 . 평균과 신호 대 잡음 비가 관심 반응입니다 .

Taguchi 배열을 사용하는 대신 제어 요인 , 잡음 요인 및 제어 * 잡음 교호작용을 포함하는 사용자 설계를 구성할 수 있습니다 . **결합 배열**이라고 하는 이 설계는 일반적으로 Taguchi 배열보다 비용 효율성이 높고 유익합니다 .

선택 설계 잠재적 제품을 비교하는 데 사용할 수 있는 설계를 구성합니다 . 선택 설계의 요인은 제품 속성입니다 . 이 설계는 다양한 속성의 조합인 제품 프로파일을 한 쌍으로 또는 3~4 개의 그룹으로 배열합니다 . 실험은 각 응답자가 한 쌍의 프로파일 중 어떤 프로파일을 선호하는지 나타내도록 구성됩니다 . 제품 속성에 대한 사전 정보를 반영하는 선택 설계를 생성할 수 있습니다 .

최대차이 최대차이 연구 중에 응답자에게 제공할 수 있는 선택 집합으로 구성된 설계를 생성합니다 . 응답자는 소수의 선택 집합 중에서 가장 선호하는 옵션과 가장 선호하지 않는 옵션만 보고합니다 . 이렇게 하면 응답자가 선호도를 기준으로 옵션 순위를 지정하게 되고 , 이로 인해 표준 선호도 척도를 사용하는 것보다 더 확증적인 순위를 얻는 경우가 많습니다 .

JMP PRO **덮기 배열** 소프트웨어 , 네트워크 및 기타 시스템을 테스트하는 데 사용할 수 있는 조합 설계를 구성합니다 . 강도 *t* 덮기 배열은 모든 *t* 요인의 모든 수준 조합이 하나 이상의 런에 나타

나는 특성을 가지고 있습니다 . 덮기 배열에서 허용하는 범주형 요인의 수는 제한이 없으며 각 요인은 임의 개수의 수준을 사용합니다 . 허용되지 않는 조합을 지정할 수 있습니다 .

공간 채움 설계 관심 시스템이 결정적 또는 근사 결정적 시스템인 상황에 대한 설계를 구성합니다. 일반적으로 매우 복잡한 결정적 컴퓨터 시뮬레이션 모형의 단순한 대리 모형을 생성하는 작업을 포함합니다 .

결정적 시스템에서는 변동이 없습니다 . 이 설계의 목표는 적합 모형과 참모형 사이의 차이 (편향) 를 최소화하는 것입니다 . 공간 채움 설계는 각 설계점을 가능한 한 멀리 분산시키거나 설계 영역 전체에 균등하게 간격을 두어 이 목표를 달성하려고 합니다 .

JMP 에서는 7 개의 공간 채움 설계 방법을 제공합니다 . 이 중 하나인 쾌속 유연 채움 설계는 수준이 여러 개 있는 범주형 요인을 수용하고 선형 제약 조건을 지원합니다 .

가속 수명 시험 설계 고장 시간을 가속화하기 위한 극단적 조건에서 제품을 테스트하는 데 유용한 설계를 구성하고 확대합니다. 이 실험 결과를 사용하면 일반적인 작동 조건에서 신뢰도를 예측할 수 있습니다 .

수명 분포는 로그 정규 또는 Weibull 일 수 있습니다 . 설계에 하나 또는 두 개의 가속 요인이 포함될 수 있습니다 . 가속 요인이 두 개인 경우 요인 교호작용을 포함하도록 선택할 수 있습니다 . 가속 모형 모수에 대한 사전 분포를 지정할 수 있습니다 . D- 최적 설계 및 두 가지 유형의 I- 최적 설계를 사용할 수 있습니다 .

비선형 설계 모수에서 비선형인 모형을 적합시키는 데 사용하는 설계를 구성하고 확대합니다 . 기존 데이터에 적합된 모형의 추정값을 사용하여 설계를 구성할 수 있습니다 . 모형 기반 추정값이 없는 경우에는 사전 지식을 적용하여 설계를 구성할 수도 있습니다 .

균형 불완비 블록 설계 한 블록에서 k 개 처리 ($k < a$) 만 실행할 수 있는 경우 , b 개 블록에서 a 개 처리를 검정하기 위한 설계를 구성합니다 .

그룹 직교 과포화 설계 과포화 선별 설계를 구성합니다 . 이 설계는 조사할 요인의 수가 실현 가능한 런 수보다 많을 때 초기 단계 작업에 적절합니다 . 그룹 직교 과포화 설계는 모형 선택에 적합한 특성을 가진 2 수준 과포화 설계의 특수한 종류입니다 .

지원 플랫폼

설계 확대 결과 설계가 최적 상태가 되도록 기존 설계에 런을 추가합니다 . 설계 확대를 사용하면 반복적인 방식으로 실험을 수행할 수 있습니다 . 설계 반복 , 중앙점 추가 , 설계 접기 , 축점 추가, 점을 추가하여 공간 채움 설계 생성 또는 지정된 수의 런으로 설계 확대 등이 가능합니다 . 런을 블록으로 그룹화하여 원래 런을 확대된 런과 구분할 수 있습니다 . 원래 모형에 없는 모형 효과를 추가하고 해당 효과에 대한 요구 사항을 지정할 수 있습니다 .

확증적 선별 적합 설계 DSD 를 위한 유효 모형 선택이라는 방법론을 사용하여 확증적 선별 설계를 분석합니다 . 이 방법론은 확증적 선별 설계의 특수 구조를 활용합니다 .

그룹 직교 과포화 설계 적합 그룹 직교 과포화 설계를 분석합니다 . 이 분석 기법은 그룹 직교 과포화 설계의 그룹 직교 구조를 활용합니다 .

설계 평가 기존 실험 설계에 진단 기능을 제공합니다 . 설계 평가 플랫폼은 설계의 장점과 한계를 평가할 수 있는 다양한 방법을 제공합니다 . 이 플랫폼은 JMP 로 생성한 설계뿐 아니라 모든 데이터 테이블과 함께 사용할 수 있습니다 .

다음과 같은 몇 가지 진단이 제공됩니다 .

- 검정력 분석
- 예측 분산 그림
- 모수에 대한 추정 효율도
- 별칭 행렬 - 모형 효과에 대한 편향 구조 표시
- 색상 맵 - 효과 간의 절대 상관 표시
- 설계 효율도 값

설계 비교 최대 4 개의 설계를 참조 설계와 비교합니다 . 이 플랫폼을 사용하여 설계 성능을 탐색 , 평가 및 비교할 수 있습니다 . 진단 결과는 설계 작동 상태를 독립적 방식 및 서로 상대적인 방식으로 나누어 보여 줍니다 .

표본 크기 및 검정력 하나 이상의 표본 평균 , 표준편차 , 단일 또는 두 비율 , 단위별 개수 (Poisson 평균), 시그마 품질 수준 등과 같은 다양한 검정 상황에 대한 표본 크기 및 검정력 계산을 제공합니다 . 이러한 옵션에 대해 세 가지 통계량 중 두 개를 지정하여 세 번째 값을 계산합니다 . 세 가지 통계량은 감지할 차이 , 표본 크기 및 검정력입니다 . 이 값 중 하나만 제공하면 다른 두 값 사이의 관계를 보여 주는 그림이 제공됩니다 .

고장 확률 추정이 목표인 경우 신뢰도 시험 계획에 필요한 표본 크기를 계산할 수 있습니다 . 또한 제품이 특정 표준을 충족하거나 초과한다는 것을 증명하는 것이 목표인 경우 신뢰도 데모에 필요한 표본 크기도 계산할 수 있습니다 .

3 장

DOE 로 시작
예제 및 핵심 개념

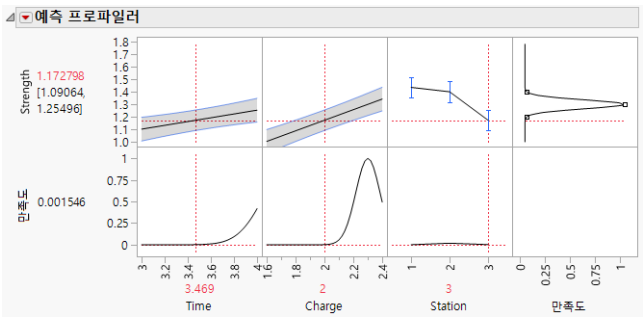
설계된 실험은 요인과 여러 반응 사이의 관계를 모델링하고 탐색하기 위한 제어된 검정 집합입니다. JMP에는 상황에 맞는 효율적인 실험 설계를 생성할 수 있도록 다양한 도구가 포함되어 있습니다. 특히 다음과 같은 종류의 설계를 사용할 수 있습니다.

- 사용자 설계 플랫폼은 고유한 상황에 맞게 사용자가 원하는 대로 설계를 만들 수 있습니다. 이 플랫폼은 모든 유형의 요인과 변경하기 어려운 요인 (분할구 상황) 을 원하는 만큼 수용하는 설계를 구성합니다. 사용자가 런 수를 제어할 수 있습니다.
- 확증적 선별 설계 플랫폼은 주효과가 다른 주효과 또는 이원 교호작용과 별칭 관계가 없는 혁신적인 선별 설계를 구성합니다. 이 설계를 사용하면 2 차 항 추정도 가능합니다.
- 선별 설계, 반응 표면 설계, 완전 요인 설계, 혼합물 설계 및 Taguchi 배열 플랫폼은 전통적 실험 설계를 구성합니다.
- 덮기 배열, 선택 설계, 공간 채움 설계, 가속 수명 시험 설계 및 비선형 설계 플랫폼은 특수한 용도의 설계 유형을 구성합니다.

설계 평가 및 설계 확대 플랫폼은 기존 설계를 평가하고 확대하기 위한 도구를 제공합니다. 표본 크기 및 검정력 플랫폼은 특수 상황에 대한 표본 크기 및 검정력 계산을 처리합니다.

이 장에서는 DOE에 대한 JMP 접근 방식을 보여 주는 예를 제공합니다. 또한 DOE의 프레임워크, 설계 생성을 지원하는 워크플로우 및 DOE의 기본 원리에 대해서도 설명합니다.

그림 3.1 프로파일러 그림의 예



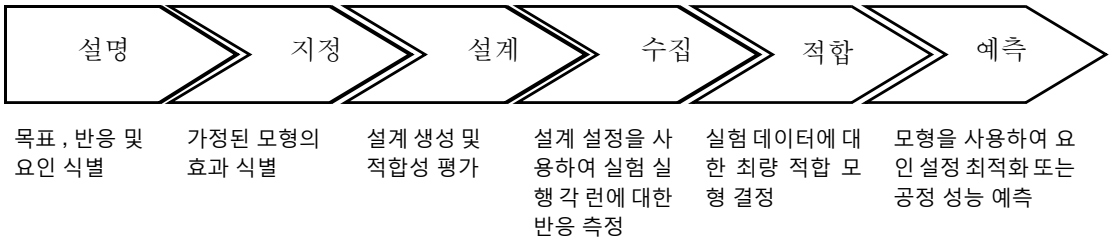
목차

실험 설계 및 DOE 워크플로우 개요 41
커피 강도 실험 42
연구 및 목표 정의 42
설계 생성 44
실험 실행 49
데이터 분석 50
DOE 워크플로우 : 기술 , 지정 , 설계 56
반응 및 요인 정의 57
모형 지정 58
설계 생성 58
설계 평가 59
테이블 생성 60
실험 설계의 원리와 지침 60
효과 계층성 60
효과 유전성 61
효과 희소성 61
중앙점, 반복 런 및 검정 61

실험 설계 및 DOE 워크플로우 개요

6단계 프레임워크는 실험 설계, 실험 실행 및 결과 분석을 위한 구조를 제공합니다. 이러한 모든 단계에서 음향 엔지니어링과 각 공정에 대한 지식이 매우 중요합니다.

그림 3.2 실험 설계 프레임워크



DOE 플랫폼에서 처음 세 단계를 수행합니다 . 최종 결과는 작업 환경에서 실행할 수 있는 설계입니다. 이 세 단계의 워크플로우에 대한 자세한 설명은 "DOE 워크플로우 : 기술 , 지정 , 설계 "(56 페이지) 에서 확인하십시오 .

설명 실험의 목표를 결정합니다 . 반응과 요인을 식별합니다 .

활성 요인 식별 , 최적의 요인 설정 찾기 , 예측 모형 구축 등이 목표일 수 있습니다 .

지정 실제 상황을 충분히 설명한다고 간주되는 가정된 모형을 결정하거나 지정합니다 .

가정된 모형은 추정할 모든 효과를 이상적으로 포함하는 초기 모형입니다 . 일부 플랫폼에서는 관심 모형을 명시적으로 구축할 수 있습니다 . 그 외 플랫폼에서는 모형이 암시적으로 선택됩니다 . 예를 들어 선별 설계 플랫폼의 경우 지정된 해상도로 모형을 선택할 수 있습니다 . 설계 해상도는 교락되는 효과를 결정합니다. 효과 교락이 발생하면 실제로 어떤 효과가 활성화되는지 모호해질 수 있습니다 .

설계 가정된 모형과 일치하는 설계를 생성합니다 . 이 설계를 평가하여 설계의 장점과 한계를 이해하고 모형과 목표에 따라 필요한 정보를 제공하도록 합니다 .

설계 생성 플랫폼의 " 설계 평가 " 또는 " 설계 진단 " 개요를 통해 설계 특성을 파악할 수 있습니다 .

다음은 제어된 조건에서 실험이 실행되는 데이터 수집 단계입니다 .

수집 각 실험을 수행하고 반응 값을 기록합니다 .

실험을 실행한 후 생성된 데이터 테이블의 스크립트를 통해 모형 적합 및 선별과 같은 플랫폼을 사용하여 모형을 적합시킬 수 있습니다. 모형은 목표에 따라 활성 효과를 식별하거나 최적 설정을 찾는 데 도움이 됩니다.

적합 가정된 모형을 실험 데이터에 적합시킵니다 .

JMP 모델링 플랫폼을 사용하여 모형을 적합시키고 구체화합니다 . 경우에 따라 모형 모호성을 해결하기 위해 설계를 확대하고 추가 런을 수행해야 할 수도 있습니다 .

예측 구체화된 모형을 사용하여 실험 목표를 처리합니다.

활성 효과를 확인하고, 반응을 최적화할 요인 수준을 찾거나, 예측 모형을 구축합니다.

설계된 실험은 일반적으로 공정 지식을 구성하기 위해 순차적으로 사용됩니다. 보통 설계 전략은 선별 설계로 시작하여 잠재적 활성 요인 목록의 범위를 좁힙니다. 그런 다음 식별된 요인을 교호작용 및 2차 효과에 대한 이해를 높이는 데 초점을 맞춘 설계에서 연구합니다. 효과의 원인이 되는 요인과 관련된 모호성을 해결하기 위해 설계를 확대해야 할 경우도 있습니다. 이 섹션에 설명된 단계는 단일 실험을 수행하고 분석하는 것과 관련이 있습니다. 그러나 목표 달성을 위해 순차적인 실험이 필요할 수도 있습니다.

"커피 강도 실험"(42 페이지)에 나오는 예에서는 DOE 워크플로우 공정의 단계를 명시적으로 보여 줍니다. 또한 데이터 테이블 스크립트를 사용하여 실험 데이터를 분석하는 방법도 보여 줍니다. 실험 설계 가이드에 나오는 대부분의 예에서는 적합한 설계를 지원하는 워크플로우와 연구를 통한 실험 데이터 분석을 둘 다 보여 줍니다.

커피 강도 실험

- "연구 및 목표 정의"
- "설계 생성"
- "실험 실행"
- "데이터 분석"

연구 및 목표 정의

당신은 지역 내 중소규모 커피 로스터 회사의 직원입니다. 개별적으로 추출한 12 온스 커피의 강도 문제를 처리해야 합니다. 커피 강도에 영향을 미치는 요인을 확인하고 해당 요인에 대한 최적 설정을 찾으려고 합니다.

반응

반응은 커피 Strength 입니다. 이 값은 굴절계를 사용하여 총 용해성 물질로 측정됩니다. 커피는 한 잔 분량의 커피 드리퍼를 사용하여 추출되며, 그라운드에서 액체가 나오고 5 분 후에 측정됩니다.

이전 연구에서는 강도 1.3 이 가장 적합하고 1.2 와 1.4 사이의 농도가 허용되는 것으로 나타났습니다.

요인

이 연구에 대해 식별된 네 가지 요인은 Grind, Temperature, Time 및 Charge 입니다. 커피는 작업장 내 세 군데 위치에서 추출됩니다. 추출 위치로 인한 변동을 고려하기 위해 Station 이 블록 요인으로 연구에 포함되었습니다. 다음은 각 요인에 대한 설명입니다.

- Grind 는 커피 입자의 굵기입니다. Grind 는 두 가지 수준 (Medium, Coarse) 으로 설정됩니다.

- Temperature 는 물을 그라운드에 따르기 직전에 측정한 온도 (화씨) 입니다 . Temperature 는 화씨 195 도와 205 도로 설정됩니다 .
- Time 은 추출 시간 (분) 입니다 . Time 은 3 분 또는 4 분으로 설정됩니다 .
- Charge 는 콘 필터에 넣은 커피의 양이며 , 물 1 온스당 커피 원두 (그램) 로 측정됩니다 . Charge 는 1.6 과 2.4 로 설정됩니다 .
- Station 은 커피를 추출하는 장소입니다 . 세 위치에 1, 2, 3 라벨이 지정됩니다 .

표 3.1 에는 요인 및 요인 설정에 대한 정보가 요약되어 있습니다. 요인과 수준은 Design Experiment 폴더의 Coffee Factors.jmp 샘플 데이터 테이블에도 제공됩니다.

표 3.1 커피 실험의 요인 및 설정 범위

요인	역할	설정 범위
Grind	범주형	Medium, Coarse
Temperature	연속형	195 ~ 205
Time	연속형	3 ~ 4
Charge	연속형	1.6 ~ 2.4
Station	블록화	1, 2, 3

다음 사항에 유의하십시오 .

- Grind 는 수준이 두 개인 범주형 요인입니다 .
- Temperature, Time 및 Charge 는 연속형 요인입니다 .
- Station 은 수준이 세 개인 블록 요인입니다 .

모든 요인은 각 런마다 변경하고 재설정할 수 있습니다. 이 실험에는 변경하기 힘든 요인이 없습니다.

커피 실험을 실행하는 데 사용되는 장비는 그림 3.3에 나와 있습니다. 이 그림은 세 군데 추출 위치 중 한 곳에 준비된 장비를 보여 줍니다. 다른 두 위치에도 같은 종류의 장비가 있습니다.

그림 3.3 커피 실험 장비

런 수

사용 가능한 리소스와 시간에 따라 총 12 개의 런을 수행할 수 있다고 결정합니다. 위치가 세 군데이므로 각 위치에서 4 번의 런을 수행합니다.

설계 생성

DOE 워크플로우 : 기술 , 지정 , 설계에 설명된 설계 워크플로우 공정 단계에 따라 설계를 생성합니다 .

- 반응 및 요인 정의
- 모형 지정
- 설계 생성
- 설계 평가
- 테이블 생성

반응 및 요인 정의

표시되는 첫 번째 개요에 반응 및 요인 정보를 입력합니다.

반응

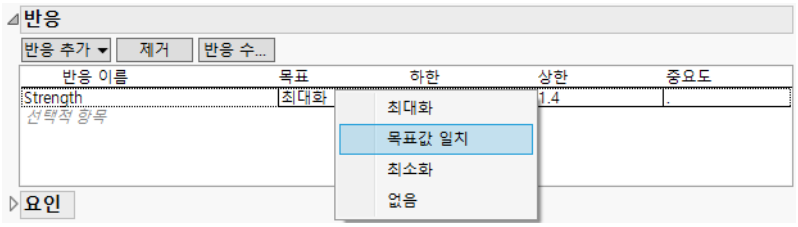
1. **DOE > 사용자 설계**를 선택합니다 .

2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Strength 를 입력합니다 .

   기본 목표는 **최대화**로 되어 있습니다. 이 실험에서는 허용 한계(1.2, 1.4) 내에서 목표 강도 1.3 의 커피를 추출할 수 있는 요인 설정을 찾으려고 합니다 .

3. 기본 목표 **최대화**를 클릭한 후 **목표값 일치**로 변경합니다 .

그림 3.4 목표값 일치를 목표로 선택



4. **하한** 아래를 클릭하고 1.2 를 입력합니다 .
5. **상한** 아래를 클릭하고 1.4 를 입력합니다 .
6. **중요도** 아래 영역은 비워 둡니다 .

   반응이 하나뿐이므로 기본적으로 반응의 중요도가 1 로 지정됩니다 .

완료된 반응 개요는 그림 3.5에 나와 있습니다.

요인

요인을 수동으로 입력하거나, 요인과 설정이 포함된 기존 테이블을 사용하여 입력합니다. 새 실험을 설계하는 경우 먼저 수동으로 요인을 입력해야 합니다. "요인 저장" 옵션을 사용하여 데이터 테이블에 요인을 저장한 후에는 저장된 테이블을 사용하여 요인을 불러올 수 있습니다.

이 예에서는 두 방법 중 하나를 선택할 수 있습니다. 자세한 내용은 "수동으로 요인 입력"(45 페이지) 또는 "요인 불러오기를 사용하여 요인 입력"(46 페이지)에서 확인하십시오.

수동으로 요인 입력

1. **요인 추가 > 범주형 > 2 수준**을 클릭합니다 .
2. 기본 이름 X1 을 Grind 로 변경합니다 .

   요청한 대로 " 역할 " 이 " 범주형 " 으로 설정되어 있습니다 . " 변경 " 속성은 기본적으로 " 쉬움 " 으로 설정되어 매 런마다 Grind 설정을 재설정할 수 있음을 나타냅니다 .

3. 기본 값 "L1" 과 "L2" 를 클릭하고 "Coarse" 와 "Medium" 으로 변경합니다 .
4. **N 개 요인 추가** 옆에 3 을 입력한 후 **요인 추가 > 연속형**을 클릭합니다 .
5. 기본 항목을 다음과 같은 요인 이름과 값으로 변경합니다 .
   - Temperature(195, 205)
   - Time(3, 4)
   - Charge(1.6, 2.4)

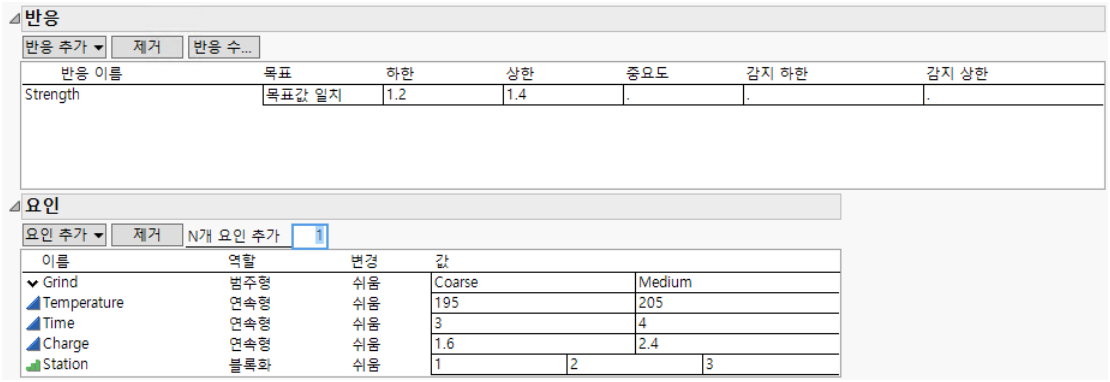
6. **요인 추가 > 블록화 > 블록당 4 회 런**을 클릭합니다 .

   예산에 따라 결정한 총 런 수는 12 회입니다 . 실행 위치 세 곳 사이에 이러한 런의 균형을 맞추려고 합니다 .

7. 기본 이름 X5 를 Station 으로 변경합니다 .

   " 역할 " 이 " 블록화 " 로 설정되어 있고 " 값 " 설정에 1 만 나타납니다 . 원하는 런 수가 지정되지 않으면 JMP 에서 블록 수를 결정할 수 없기 때문입니다 . " 설계 생성 " 개요에서 " 런 수 " 를 지정하면 Station 의 수준 수가 필요한 값으로 업데이트됩니다 .

그림 3.5 완료된 반응 및 요인 개요



8. **계속**을 클릭합니다 .

   다음 개요가 " 사용자 설계 " 창에 추가됩니다 .

   – 요인 제약 조건 정의 (이 예에서 사용되지 않음)
   – 모형
   – 별칭 항
   – 설계 생성

요인 불러오기를 사용하여 요인 입력

다음과 같이 요인 정보가 포함된 테이블을 사용하여 요인을 입력합니다 .

1. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Coffee Factors.jmp를 엽니다.

   요인을 불러오면 " 사용자 설계 " 창이 다음과 같은 추가 개요가 포함되도록 업데이트됩니다 .

   – 요인 제약 조건 정의 (이 예에서 사용되지 않음)
   – 모형
   – 별칭 항
   – 설계 생성

요인 제약 조건 정의

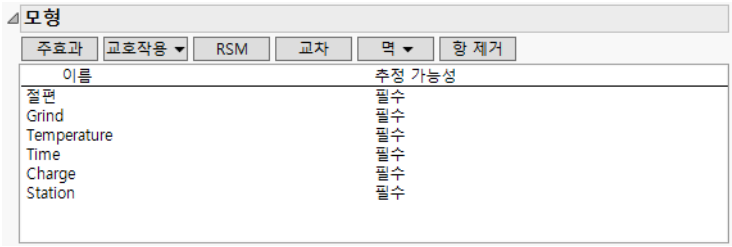
"요인 제약 조건 정의" 개요는 요인을 수동으로 입력하고 "계속"을 클릭하거나, 요인 테이블에서 요인을 불러온 후에 나타납니다. 요인 제약 조건이 있으면 반응 및 요인 단계의 일부로 제약 조건을 추가합니다. 이 설계에는 요인 설정에 대한 제약 조건이 없으므로 개요를 변경하지 말고 그대로 두십시오.

모형 지정

모형 개요

그림 3.7에서는 "모형" 개요를 보여 줍니다. "모형" 개요에서는 추정할 효과를 포함하는 가정된 모형을 지정합니다. 자세한 내용은 "지정"(41 페이지)에서 확인하십시오. 기본적으로 나타나는 목록에는 모든 주효과가 **필수**로 표시되어 설계에서 모든 주효과를 추정할 수 있음을 나타냅니다. 현재는 주요 관심이 요인의 주효과에 있으므로 "모형" 개요에 다른 효과를 추가하지 않습니다.

그림 3.6 주효과만 포함된 모형 개요



결과 복제 단계 (선택 사항)

사용자 설계 알고리즘은 랜덤 시작 설계로 시작하므로 설계가 그림 3.8에 표시된 것과 다를 수 있습니다 . 정확히 같은 런을 포함하는 설계를 얻으려면 설계를 생성하기 전에 난수 시드값과 시작 수를 다음과 같이 설정합니다 .

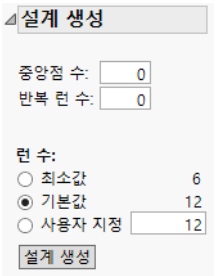
1. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 .
2. 569534903 을 입력합니다 .
3. **확인**을 클릭합니다 .
4. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택합니다 .
5. 100 을 입력합니다 .
6. **확인**을 클릭합니다 .

참고 : 난수 시드값과 시작 수를 설정하면 이 예에 나오는 것과 정확히 같은 설계가 재현됩니다 . 그러나 설계 테이블의 행 순서는 다를 수 있습니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

설계 생성

"설계 생성" 개요에서는 설계의 구조와 크기에 대한 추가 정보를 입력할 수 있습니다. 설계의 기본 런 수는 12 회로 표시됩니다. 설계 예산에 따라 결정한 총 런 수는 12 회입니다("런 수"(44 페이지) 참조).

그림 3.7 설계 생성 개요



1. **설계 생성**을 클릭합니다.

   " 설계 " 및 " 설계 평가 " 개요가 " 사용자 설계 " 창에 추가됩니다. " 출력 옵션 " 패널도 나타납니다.

   " 설계 " 개요에서는 설계를 보여 줍니다 (그림 3.8). " 결과 복제 단계 (선택 사항)"(47 페이지) 에 설명된 대로 난수 시드값과 시작 수를 설정하지 않은 경우 설계가 그림 3.8 에 나와 있는 것과 다를 수 있습니다. 이는 설계를 생성하는 데 사용된 알고리즘이 랜덤 시작 설계로 시작하기 때문입니다.

그림 3.8 커피 실험 설계

설계

런	Grind	Temperature	Time	Charge	Station
1	Coarse	195	3	1.6	1
2	Medium	195	4	2.4	2
3	Medium	195	4	1.6	3
4	Medium	205	4	2.4	1
5	Coarse	205	3	2.4	2
6	Medium	205	3	2.4	3
7	Medium	205	3	1.6	1
8	Coarse	205	4	1.6	2
9	Coarse	205	4	1.6	3
10	Coarse	195	4	2.4	1
11	Medium	195	3	1.6	2
12	Coarse	195	3	2.4	3

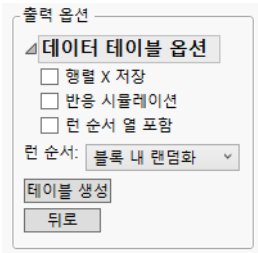
설계 평가

"설계 평가" 개요에서는 설계를 평가하는 다양한 방법을 제공합니다. 이 주제는 중요하지만 간단하게 하기 위해 이 예에서는 다루지 않습니다. 자세한 내용은 "설계 평가" 장(403 페이지)에서 확인하십시오.

테이블 생성

"출력 옵션" 패널을 사용하여 데이터 테이블에서 런 순서를 지정합니다. 기본 선택 항목인 **블록 내 랜덤화**가 적합합니다. 이 항목을 선택하면 각 Station에 대해 랜덤 순서로 런이 배열됩니다.

그림 3.9 출력 옵션



1. **테이블 생성**을 클릭합니다.

   그림 3.10 에 표시된 데이터 테이블이 열립니다. "결과 복제 단계 (선택 사항)"(47 페이지) 에 설명된 대로 난수 시드값과 시작 수를 설정하지 않은 경우 설계가 다를 수 있습니다. 이러한 설계 테이블은 또 다른 대체 최적 설계를 나타냅니다.

그림 3.10 사용자 설계 테이블

사용자 설계
설계 기준: 사용자 설계 D 최적
모형
설계 평가
일반화 회귀
DOE 대화상자
열(6/0)
Grind ✱
Temperature ✱
Time ✱
Charge ✱
Station ✱
Strength ✱

	Grind	Temperature	Time	Charge	Station	Strength
1	Coarse	195	4	2.4	1	•
2	Medium	205	3	1.6	1	•
3	Medium	205	4	2.4	1	•
4	Coarse	195	3	1.6	1	•
5	Medium	195	4	2.4	2	•
6	Medium	195	3	1.6	2	•
7	Coarse	205	4	1.6	2	•
8	Coarse	205	3	2.4	2	•
9	Medium	195	4	1.6	3	•
10	Medium	205	3	2.4	3	•
11	Coarse	205	4	1.6	3	•
12	Coarse	195	3	2.4	3	•

"열" 패널에서 요인과 반응의 오른쪽에 있는 별표에 유의하십시오. 이는 데이터 테이블의 열에 저장된 열 특성을 나타냅니다. 이러한 열 특성은 데이터 분석에 사용됩니다. 자세한 내용은 "요인"(84 페이지) 및 "요인 열 특성"(89 페이지) 에서 확인하십시오.

실험 실행

여기서는 실험을 수행합니다. 각 Station에서 설계 테이블에 표시된 순서대로 4번씩 런이 수행됩니다. 장비와 재료는 각 런이 수행될 때마다 재설정됩니다. 예를 들어 연속된 두 번의 런에 필요

한 물 온도가 195 도인 경우 가열 용기가 식은 후 12 온스의 물이 각각 195 도까지 가열됩니다. Strength 측정값이 기록됩니다.

Strength 에 대한 설계 및 실험 결과는 Design Experiment 폴더에 있는 Coffee Data.jmp 샘플 데이터 테이블(그림 3.11)에 제공됩니다.

그림 3.11 강도 결과가 포함된 커피 설계

	Grind	Temperature	Time	Charge	Station	Strength
1	Medium	205	4	2.4	1	1.78
2	Coarse	195	3	1.6	1	1.25
3	Medium	205	3	1.6	1	1.10
4	Coarse	195	4	2.4	1	1.63
5	Coarse	205	4	1.6	2	1.26
6	Medium	195	4	2.4	2	1.63
7	Medium	195	3	1.6	2	1.22
8	Coarse	205	3	2.4	2	1.51
9	Coarse	205	4	1.6	3	1.07
10	Coarse	195	3	2.4	3	1.26
11	Medium	195	4	1.6	3	1.13
12	Medium	205	3	2.4	3	1.25

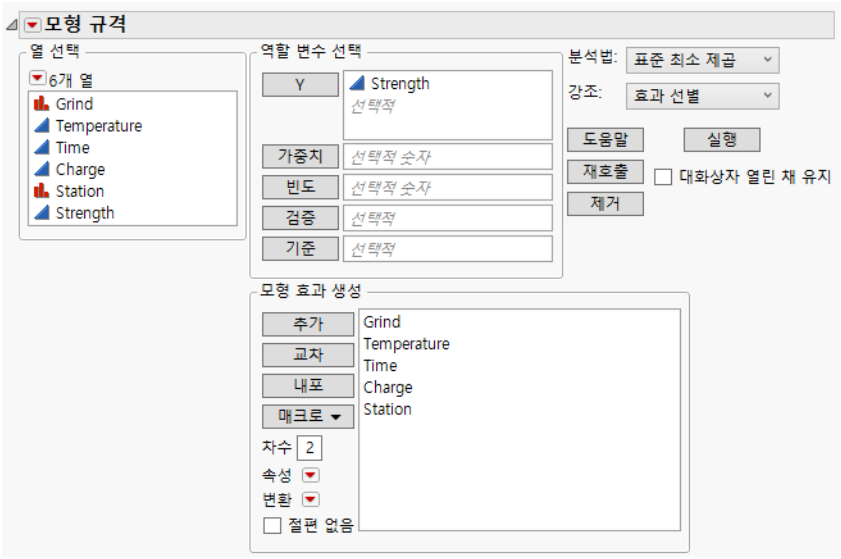
데이터 분석

사용자 설계 플랫폼을 사용하면 생성된 모형 스크립트를 설계 테이블에 저장하여 데이터 분석 작업을 쉽게 수행할 수 있습니다(그림 3.10). 실험을 수행하고 데이터를 입력한 후 이 스크립트를 실행합니다. 그러면 "사용자 설계" 창의 "모형" 개요에서 지정한 효과가 포함된 "모형 적합" 창이 열립니다.

모형 적합

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Coffee Data.jmp 를 엽니다 . 테이블 패널에 " 사용자 설계 " 로 생성한 "Model" 스크립트가 있습니다 .
2. "Model" 스크립트 옆의 녹색 삼각형을 클릭합니다 . " 모형 규격 " 창에 " 모형 " 개요에서 지정한 효과가 표시됩니다 .

그림 3.12 모형 규격 창

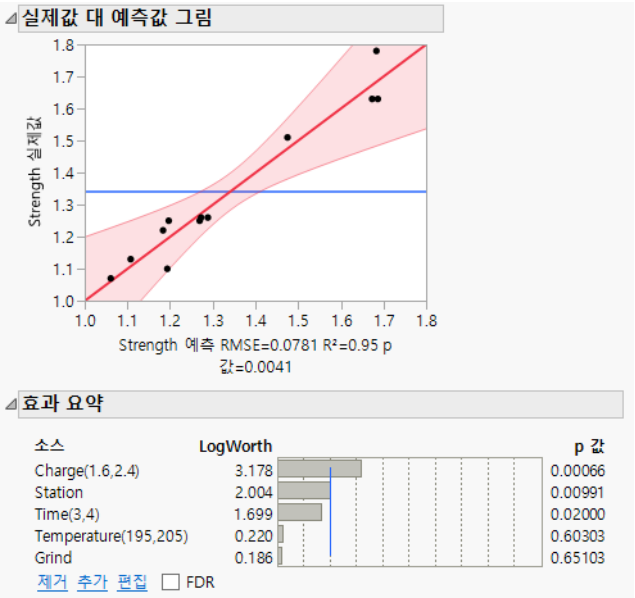


3. **대화상자 열린 채 유지** 옵션을 선택합니다.
4. **실행**을 클릭합니다.

모형 분석

"효과 요약" 및 "실제값 대 예측값 그림" 보고서는 모형에 대한 개요 정보를 제공합니다.

그림 3.13 전체 모형에 대한 효과 요약 및 실제값 대 예측값 그림



다음 사항에 유의하십시오 .

- " 실제값 대 예측값 그림 " 에 적합 결여의 증거가 없습니다 .
- " 실제값 대 예측값 " 그림에 따르면 모형이 유의합니다 . 그림 아래의 **p 값 = 0.0041** 은 전체 모형 검정의 유의 수준을 제공합니다 .
- " 효과 요약 " 보고서는 Charge, Station 및 Time 이 0.05 수준에서 유의함을 보여 줍니다 .
- " 효과 요약 " 보고서는 또한 Temperature 와 Grind 가 유의하지 않다는 것을 보여 줍니다 .

모형 축소

Temperature 와 Grind 는 활성 효과가 아니므로 모형에 랜덤 잡음을 적용합니다 . 활성 효과와 관련된 모형 모수를 더 정확하게 추정하려면 이러한 효과 없이 모형을 재적합시키십시오 .

1. " 모형 규격 " 창의 " 모형 효과 생성 " 목록에서 Temperature 와 Grind 를 선택합니다 .
2. **제거**를 클릭합니다 .
3. 모형 **강조**가 **효과 선별**로 설정되도록 합니다 .

   " 효과 선별 " 강조는 실험 설계를 분석하는 데 유용한 보고서 (예 : 예측 프로파일러) 를 제공합니다 .
4. **실행**을 클릭합니다 .

그림 3.14 축소 모형의 부분 출력

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	1.3408333	0.020047	66.88	<.0001*
Time(3,4)	0.0758333	0.020047	3.78	0.0069*
Charge(1.6,2.4)	0.1691667	0.020047	8.44	<.0001*
Station[1]	0.0991667	0.028351	3.50	0.0100*
Station[2]	0.0641667	0.028351	2.26	0.0580

효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F
Time(3,4)	1	1	0.06900833	14.3093	0.0069*
Charge(1.6,2.4)	1	1	0.34340833	71.2078	<.0001*
Station	2	2	0.16251667	16.8494	0.0021*

효과 상세 정보

척도화된 추정값

명목형 요인이 모든 수준으로 확장되었습니다.

항	척도화된 추정값	표준 오차	t 비	Prob>|t|
절편	1.3408333	0.020047	66.88	<.0001*
Time(3,4)	0.0758333	0.020047	3.78	0.0069*
Charge(1.6,2.4)	0.1691667	0.020047	8.44	<.0001*
Station[1]	0.0991667	0.028351	3.50	0.0100*
Station[2]	0.0641667	0.028351	2.26	0.0580
Station[3]	-0.163333	0.028351	-5.76	0.0007*

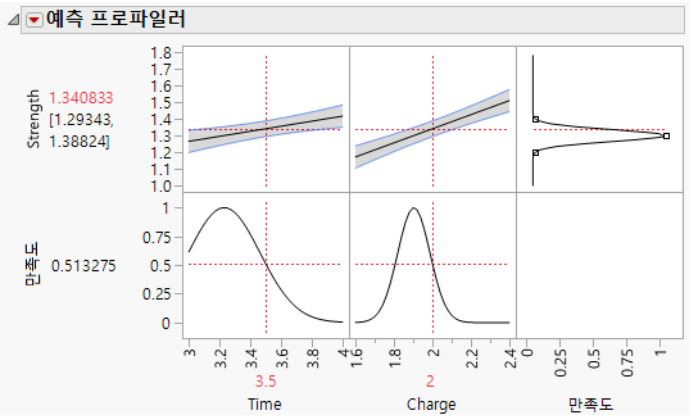
다음 사항에 유의하십시오 .

- " 효과 검정 " 보고서에 세 가지 효과가 모두 유의한 것으로 나타납니다 .
- " 척도화된 추정값 " 보고서는 Station[1] 과 Station[3] 평균이 Strength 의 평균 반응과 유의하게 다르다는 것을 나타냅니다 .
- " 모수 추정값 " 보고서에 나타나는 추정값이 " 척도화된 추정값 " 보고서의 추정값과 동일합니다 . 이는 효과가 코드화되어 있기 때문입니다 . 자세한 내용은 " 열 특성 " 부록의 " 코딩 "(739 페이지) 에서 확인하십시오 .
- Station[3] 의 추정값은 " 척도화된 추정값 " 보고서에만 나타납니다 . 이 보고서에는 명목형 요인이 확장되어 모든 수준의 추정값을 표시합니다 .
- " 모수 추정값 " 보고서는 코드화된 효과와 관련하여 모형이 지정된 모형 계수의 추정값을 제공합니다 .

모형 탐색

" 예측 프로파일러 " 가 보고서 맨 아래에 나타납니다 .

그림 3.15 예측 프로파일러



실험을 설계할 때 1.2 와 1.4 를 한계로 하여 목표값 일치를 반응 목표로 설정했습니다. JMP 에서는 이 정보를 사용하여 사용자가 지정한 사항을 반영하는 만족도 함수를 구성합니다. 자세한 내용은 "요인"(84 페이지)에서 확인하십시오.

그림 3.15 에서 다음 사항을 알 수 있습니다 .

- 그래프의 위쪽 행에 있는 처음 두 그림은 다른 요인의 설정이 주어졌을 때 한 요인에 대해 Strength 가 어떻게 변화하는지 보여 줍니다 . 예를 들어 Charge 가 2 일 때 Time 그림의 선은 예측 Strength 가 Time 에 따라 어떻게 변하는지 보여 줍니다 .
- 위쪽 행에 나오는 그림 왼쪽의 값은 선택한 요인 설정에 대한 예측 Strength(빨간색) 및 평균 Strength 의 신뢰 구간을 제공합니다 .
- 위쪽 행의 맨 오른쪽 그림은 Strength 에 대한 만족도 함수를 보여 줍니다 . 이 만족도 함수는 목표값 1.3 이 가장 적합하다는 것을 나타냅니다 . 해당 목표값에서 멀어질수록 만족도가 감소합니다 . 한계값 1.2 와 1.4 에서는 만족도가 0 에 가까워집니다 .
- 아래쪽 행의 그림은 다른 요인의 설정에 대해 각 요인의 만족도 트레이스를 보여 줍니다 .
- 아래쪽 행에 나오는 그림 왼쪽의 값은 선택한 요인 설정에 대한 반응 값의 만족도를 제공합니다 .

Time 및 Charge 열의 빨간색 수직 파선을 드래그하여 다양한 요인 설정을 탐색합니다. 모형에 교호작용이 없으므로 프로파일러는 Charge 를 증가하면 Strength 도 증가함을 나타냅니다. 또한 Strength 는 Time 의 변화보다 Charge 의 변화에 더 민감한 것으로 보입니다.

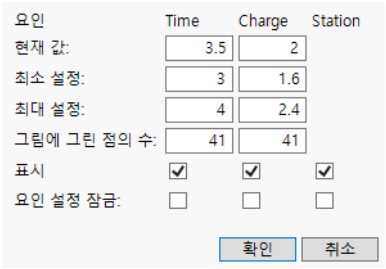
Station 은 블록 요인이므로 예측 프로파일러에 나타나지 않습니다 . 그러나 Station 에 따라 예측 Strength 가 어떻게 달라지는지 확인하고 싶을 수 있습니다 . 다음과 같이 예측 프로파일러에 Station 을 포함합니다 .

1. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **요인 격자 재설정**을 선택합니다 .

   Time, Charge 및 Station 열이 포함된 " 요인 설정 " 창이 나타납니다 . Station 아래의 **표시** 상자가 선택되어 있지 않습니다. 이는 Station 이 예측 프로파일러에 표시되지 않음을 나타냅니다.

2. Station 아래에서 **표시** 행의 상자를 선택합니다 .
3. Station 아래에서 **요인 설정 잠금** 행의 상자를 선택 취소합니다 .

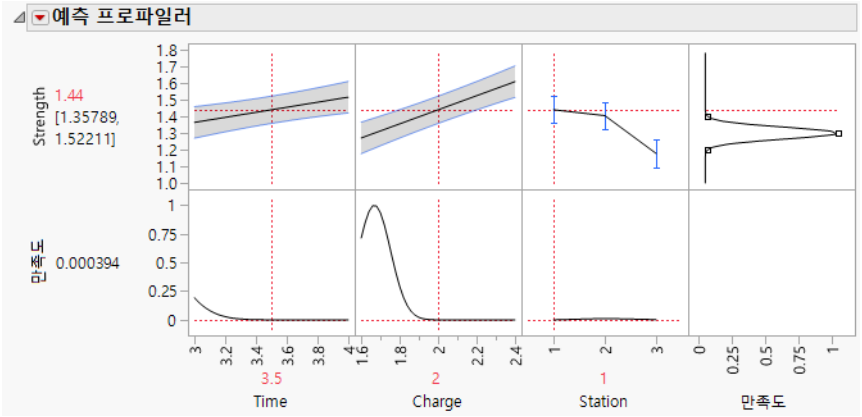
그림 3.16 요인 설정 창



4. **확인**을 클릭합니다 .

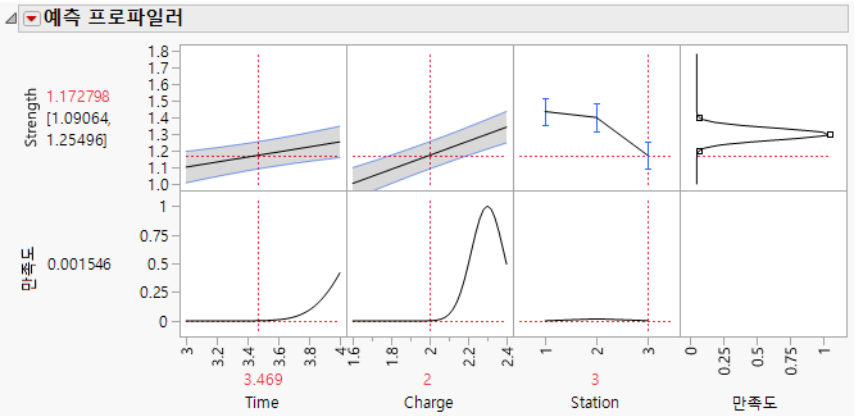
   Station 그림이 " 예측 프로파일러 " 에 나타납니다 .
5. Station 위의 두 그림 중 하나를 클릭하여 빨간색 수직 파선을 삽입합니다 .
6. 빨간색 수직 파선을 Station 1 로 이동합니다 .

그림 3.17 Station 1 의 결과를 보여 주는 예측 프로파일러



7. 빨간색 수직 파선을 Station 3 으로 이동합니다 .

그림 3.18 Station 3 의 결과를 보여 주는 예측 프로파일러



Station 1 의 설계 영역 가운데에 있는 예측 Strength 는 1.44 입니다 . Station 3 의 경우 예측 Strength 는 약 1.18 입니다 . 이 차이의 크기는 Station 변동을 해결해야 함을 나타냅니다 . Station 변동을 잘 제어하면 일관된 Strength 를 얻을 수 있습니다 . Station 일관성이 확보되면 Time 과 Charge 에 대한 공통 최적 설정을 결정할 수 있습니다 .

커피 실험 설계를 구성하는 데 사용한 공정은 DOE 워크플로우의 단계를 따릅니다. 다음 섹션에서는 DOE 워크플로우에 대해 자세히 설명합니다.

DOE 워크플로우 : 기술 , 지정 , 설계

DOE 플랫폼은 실험 설계에 내재된 워크플로우를 제공하는 일련의 단계로 구성됩니다 . 각 단계를 완료하면 **계속**을 클릭하여 다음 단계로 이동합니다 . 이 섹션에서 설명하는 요소는 9 개의 실험 설계 플랫폼에 공통으로 적용됩니다 . 이 섹션에서는 다음 플랫폼에 대해 설명합니다 .

- 사용자 설계
- 확증적 선별 설계
- 선별 설계
- 반응 표면 설계
- 완전 요인 설계
- 혼합물 설계
- JMP PRO 덮기 배열
- 공간 채움 설계
- Taguchi 배열

선택 설계, 가속 수명 시험 설계 및 비선형 설계는 위의 플랫폼과 많이 다른 특수 목적의 플랫폼입니다. 이러한 세 플랫폼은 이 섹션에서 다루지 않습니다.

이 섹션에서는 DOE 워크플로우의 단계를 설명합니다. 또한 다양한 설계 플랫폼에서의 구현에 대해서도 설명합니다.

반응 및 요인 정의

실험 설계 프레임워크의 설명 단계에서는 다음을 수행합니다 .

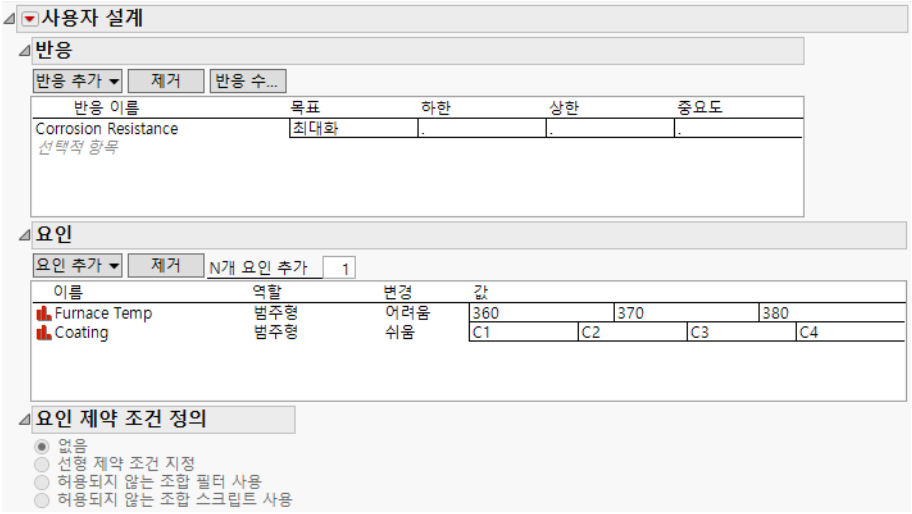
- 관심 반응 및 요인을 식별합니다 .
- 실험 목표를 결정합니다 . 반응을 최대화하거나 목표값 적중을 원할 수 있습니다 . 이때 해당 목표값을 지정합니다 . 아니면 단순히 반응에 영향을 미치는 요인을 식별할 수도 있습니다 .
- 실험 범위 또는 설계 공간을 설명하는 요인 설정을 식별합니다 .

대부분의 JMP DOE 플랫폼을 열면 반응과 요인을 나열할 수 있는 개요가 표시됩니다. "반응" 개요는 전체 플랫폼에서 공통입니다. 여기에서는 반응 및 추가 정보(예: 반응 목표, 하한, 상한, 중요도)를 삽입합니다.

"요인" 개요는 플랫폼마다 다릅니다. 이는 각 플랫폼에서 다루는 요인의 유형과 특정 설계 상황에 맞추기 위한 것입니다. 특정 플랫폼에서 반응과 요인을 입력한 후 "계속"을 클릭하면 "요인 제약 조건 정의" 개요가 나타납니다. 이 개요에서는 설계에 사용 가능한 요인의 값을 제한할 수 있습니다.

그림 3.19에서는 Design Experiment 폴더에 있는 Box Corrosion Split-Plot.jmp 샘플 데이터 테이블에서 설계를 구성하기 위해 사용자 설계 플랫폼을 사용한 "반응" 및 "요인" 개요를 보여 줍니다. 또한 "계속"을 클릭하면 나타나는 "요인 제약 조건 정의" 개요도 표시됩니다. "제약 조건 정의" 개요를 사용하면 요인 설정이 충족해야 하는 제한을 지정할 수 있습니다.

그림 3.19 Box Corrosion Split-Plot 실험의 반응 및 요인



모형 지정

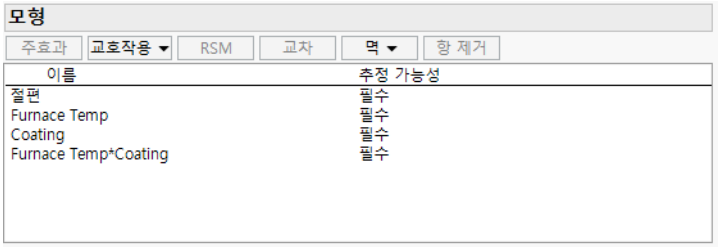
"반응" 및 "요인" 개요에 필요한 정보를 모두 입력한 후 **계속** 버튼을 클릭합니다. 그러면 설계 구성의 다음 단계로 이동하여 가정된 모형을 명시적 또는 암시적으로 선택할 수 있습니다.

사용자 설계 플랫폼을 사용하면 적합시킬 모형을 명시적으로 지정할 수 있습니다. 이 모형에 대한 최적 설계가 생성됩니다. 다른 설계 플랫폼에서는 모형을 명시적으로 지정할 수 없습니다. 예를 들어 선별 플랫폼에서 한 옵션을 사용하면 완전 요인, 부분 요인 및 Plackett-Burman 설계 목록에서 선택할 수 있습니다. 이러한 설계의 별칭 관계는 적합시킬 수 있는 모형을 암시적으로 정의합니다.

"사용자 설계"에서 "반응" 및 "요인"에 정보를 입력한 후 "계속"을 클릭하면 "모형" 개요가 나타납니다. Box Corrosion Split-Plot.jmp 샘플 데이터 테이블에 사용된 설계의 예가 그림 3.20에 나와 있습니다. 가정된 모형은 주효과(Furnace Temp, Coating) 및 교호작용을 추정할 수 있어야 합니다. 생성된 설계는 이러한 효과의 추정 가능성을 보장합니다.

대부분의 다른 플랫폼에서 "계속"을 클릭하면 선택할 수 있는 설계 모음이 제공됩니다. 완전 요인에서는 "요인" 개요가 완료된 후 설계가 결정되므로 "계속"을 클릭하면 "출력 옵션"으로 바로 이동합니다.

그림 3.20 Box Corrosion Split-Plot 실험의 모형 개요



설계 생성

대부분의 DOE 플랫폼에서 최종 설계의 크기를 어느 정도 제어할 수 있습니다. 사용자 설계에서는 런 수를 지정하고 필요한 경우 중앙점 및 반복 런 수를 지정할 수 있습니다. 다른 플랫폼에서는 다양한 수준의 유연성이 제공됩니다. 설계의 중앙점 수, 반복 런 수 또는 반복 실험 횟수를 지정할 수 있는 경우가 많습니다.

런 수와 관련된 옵션을 지정했으면 "설계 생성"을 클릭합니다. 그러면 DOE 창이 업데이트되어 "설계" 개요에 설계가 표시됩니다.

Box Corrosion Split-Plot.jmp 실험을 위한 사용자 설계(24 회 런)의 "설계" 개요가 그림 3.21에 나와 있습니다. Furnace Temp의 **변경**이 **어려움**으로 지정되었으므로 **주구** 요인은 Furnace Temp 설정의 랜덤 블록을 나타내도록 구성됩니다.

그림 3.21 Box Corrosion Split-Plot 실험의 설계 개요

설계

런	주구	Furnace Temp	Coating
1	1	370	C2
2	1	370	C3
3	1	370	C1
4	1	370	C4
5	2	370	C3
6	2	370	C4
7	2	370	C2
8	2	370	C1
9	3	360	C3
10	3	360	C2
11	3	360	C4
12	3	360	C1
13	4	380	C4
14	4	380	C3
15	4	380	C2
16	4	380	C1
17	5	380	C4
18	5	380	C2
19	5	380	C1
20	5	380	C3
21	6	360	C3
22	6	360	C1
23	6	360	C4
24	6	360	C2

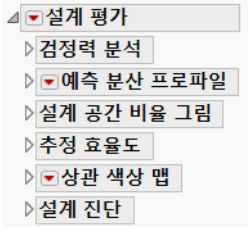
참고 : 사용자 설계에서 설계를 생성하면 " 모형 " 및 " 별칭 " 개요의 컨트롤이 비활성화됩니다 . " 뒤로 " 버튼을 사용하여 모형 또는 별칭 항을 변경할 수 있습니다 .

설계 평가

대부분의 플랫폼에서 "설계 생성"을 클릭하면 "설계 평가" 개요가 나타납니다. 여기서는 생성된 설계를 효과 감지 검정력, 예측 분산, 추정 효율도, 별칭 관계, 효과 간 상관 및 기타 설계 효율도 측도 측면에서 탐색할 수 있습니다. 그림 3.22에서는 사용자 설계의 "설계 평가" 개요를 보여 줍니다. 설계 평가에 대한 내용은 "설계 평가" 장(403 페이지)에서 확인하십시오.

일부 플랫폼의 경우 다른 유형의 설계 진단이 적합합니다. 예를 들어 공간 채움 설계는 공간 채움 설계에 특정한 지표가 포함된 "설계 진단" 개요를 제공합니다. 덮기 배열은 포함 범위에 특정한 측도가 포함된 "지표" 개요를 제공합니다.

그림 3.22 사용자 설계의 설계 평가 개요

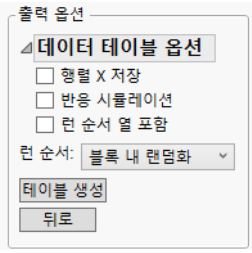


테이블 생성

대부분의 플랫폼은 "출력 옵션" 노드 또는 패널을 제공합니다. 플랫폼과 설계에 따라 "출력 옵션" 패널을 사용하여 추가 설계 구조를 지정할 수 있습니다. 예를 들어 런 수, 중앙점 수, 반복 실험 횟수 또는 생성된 데이터 테이블에 설계 런이 표시되는 순서 등을 지정할 수 있습니다.

그림 3.23에는 Design Experiment 폴더에 있는 Wine Data.jmp 샘플 데이터 테이블의 실험에 대한 "출력 옵션" 패널이 표시됩니다. 이 예에서는 다양한 "런 순서" 옵션을 선택하고 설계 데이터 테이블을 구성할 수 있습니다. 뒤로 이동하여 설계를 재구성하도록 선택할 수도 있습니다.

그림 3.23 와인 실험의 출력 옵션 패널



실험 설계의 원리와 지침

특정 원리는 실험 설계 및 실험 데이터 분석의 기초가 됩니다. 효과 계층성, 효과 유전성 및 효과 희소성 원리는 주로 모형 선택과 관련이 있습니다. 이러한 원리는 최적 모형을 검색할 때 가능한 모형 집합을 줄이는 데 도움이 됩니다. 자세한 내용은 Hamada and Wu(1992), Wu and Hamada(2009), Goos and Jones(2011) 에서 확인하십시오.

- "효과 계층성"
- "효과 유전성"
- "효과 희소성"
- "중앙점, 반복 런 및 검정"

효과 계층성

회귀 모델링에서 **효과 계층성**은 주효과(1차)가 반응의 변동을 가장 많이 설명한다는 원리입니다. 2차 효과(second-order effect), 즉 교호작용 효과와 2차 항이 변동을 두 번째로 많이 설명합니다. 그런 다음 차수가 높은 항이 계층 순서대로 나옵니다.

모델링은 주효과가 2차 효과(second-order effect)보다 중요하고 2차 효과가 3차 효과보다 중요하다는 의미를 가지고 있습니다. 즉, 차수가 높을수록 중요도가 낮습니다.

효과 유전성

효과 유전성 원리는 고차 효과의 저차 성분 모형 포함과 관련이 있습니다. 이 원리는 주효과의 크기가 작은 요인들 간에는 유의한 교호작용 효과가 없다는 관측 증거를 동기로 합니다.

강한 효과 유전성을 위해서는 모형 효과의 모든 낮은 차수 성분이 모형에 포함되어야 합니다. 삼원 교호작용(ABC)이 모형에 있다고 가정해 보겠습니다. 그러면 모든 성분 주효과와 이원 교호작용(A, B, C, AB, AC, BC)도 모형에 포함되어야 합니다.

약한 효과 유전성을 위해서는 모형 효과 중 연속된 낮은 차수 성분의 시퀀스만 포함되어야 합니다. 삼원 교호작용이 모형에 있는 경우 관련 요인 중 하나와 해당 요인을 포함하는 이원 교호작용 하나가 모형에 포함되어야 합니다. 삼원 교호작용(ABC)이 모형에 있다고 가정해 보겠습니다. B와 BC도 모형에 있으면 해당 모형은 약한 효과 유전성을 충족합니다.

연속형 요인의 경우 효과 유전성은 요인의 위치 및 척도가 달라져도 모형이 변하지 않도록 합니다.

효과 희소성

효과 희소성 원리는 반응에 대한 대부분의 변동이 상대적으로 적은 수의 효과로 설명된다고 주장합니다. 많은 효과를 연구하는 선별 설계는 효과 희소성에 크게 의존합니다. 경험에 따르면 선별 설계에 사용된 런 수는 유의하다고 할 수 있는 효과 수의 두 배 이상이어야 합니다.

중앙점 , 반복 런 및 검정

일부 DOE 플랫폼을 사용하면 중앙점(연속형 요인의 경우), 반복 런 또는 전체 반복 실험을 설계에 추가할 수 있습니다. 다음은 설계점 추가와 관련된 몇 가지 상황입니다.

중앙점 추가

연속형 요인의 중앙점을 사용하면 비선형 효과로 인한 적합 결여를 검정할 수 있습니다. 적합 결여 검정은 모형 항 누락으로 인해 오차 분산 추정값이 부풀려졌는지 여부를 판별하는 데 도움이 됩니다. 이 방법은 현명한 런 투자일 수 있습니다.

중앙점에서만 런을 반복하거나 다른 설계 런을 반복할 수 있습니다. JMP에서는 반복 런을 사용하여 모형에 독립적인 오차 추정값(순수 오차 추정값)을 생성합니다. 이 순수 오차 추정값을 사용하면 적합 결여를 검정할 수 있습니다.

중앙점은 모형 효과의 정확한 추정값을 얻는 데 도움이 되지 않습니다. 중앙점을 사용하면 곡률의 증거를 검정할 수 있지만 원인이 되는 비선형 효과를 식별할 수는 없습니다.

곡률의 원인을 식별하려면 최소 3개 수준으로 연속형 요인을 설정해야 합니다. 확증적 선별 설계는 반응에 대해 강력한 비선형 효과를 일으키는 요인을 감지하고 식별할 수 있는 3수준 설계입니다. 자세한 내용은 7장, "확증적 선별 설계"에서 확인하십시오.

반복 런 추가

런 예산으로 가능한 경우 런을 반복하거나 설계 공간 내에서 새 런을 최적으로 배포할 수 있습니다. 반복 런을 추가하면 일부 추정값에 대한 정밀도가 추가되고 적합 결여 검정의 검정력이 향상됩니다. 그러나 런 예산이 지정된 경우 반복 런을 추가하면 일반적으로 설계의 모형 효과 추정 능력이 저하됩니다. 설계 공간 내에서 최적으로 런을 배포하는 것만큼 많은 항을 추정할 수 없습니다.

적합 결여 검정

설계된 실험은 대개 실험 목표에 맞게 가능한 한 적은 수의 런이 필요하도록 구성됩니다. 런 수가 너무 적으면 극단적으로 큰 효과만 감지할 수 있습니다. 예를 들어 지정된 효과의 *t* 검정 통계량은 표준 오차에 대한 반응 평균의 변화 비율입니다. 오차 DF(자유도)가 1뿐이면 검정 임계값이 12를 초과합니다. 따라서 거의 포화된 설계에서 효과를 감지하려면 크기가 매우 커야 합니다.

유사 관측이 적합 결여 검정에 적용됩니다. 적합 결여를 감지하는 이 검정의 검정력은 분자와 분모의 자유도 값에 따라 달라집니다. 각 종류마다 df가 1뿐인 경우 0.05 수준에서 유의성을 선언하려면 *F* 값이 150을 초과해야 합니다. 각 종류의 df가 2인 경우에는 *F* 값이 19를 초과해야 합니다. 두 번째 경우에서 검정이 유의하려면 적합 결여 평균 제곱이 순수 오차 평균 제곱보다 19배 커야 합니다. 적합 결여 검정이 이상치에 민감한 것도 사실입니다.

적합 결여 검정에 대한 자세한 내용은 선형 모형 적합의 에서 확인하십시오.

런 수 결정

산업 응용 분야에서는 런을 수행할 때마다 비용이 많이 들기 때문에 런 수를 최소화해야 합니다. 관심 있는 고정 효과를 추정하려면 모형에 있는 항과 같은 수의 런만 있으면 됩니다. 효과가 활성 상태인지 여부를 판별하려면 적절한 오차 분산 추정값이 필요합니다. 이 분산의 추정값이 충분하지 않으면 모형 항을 추정하는 데 필요한 수에 4회 이상의 런을 추가하는 것이 좋습니다.

4 장

사용자 설계
요구 사항을 충족하는 설계 구성

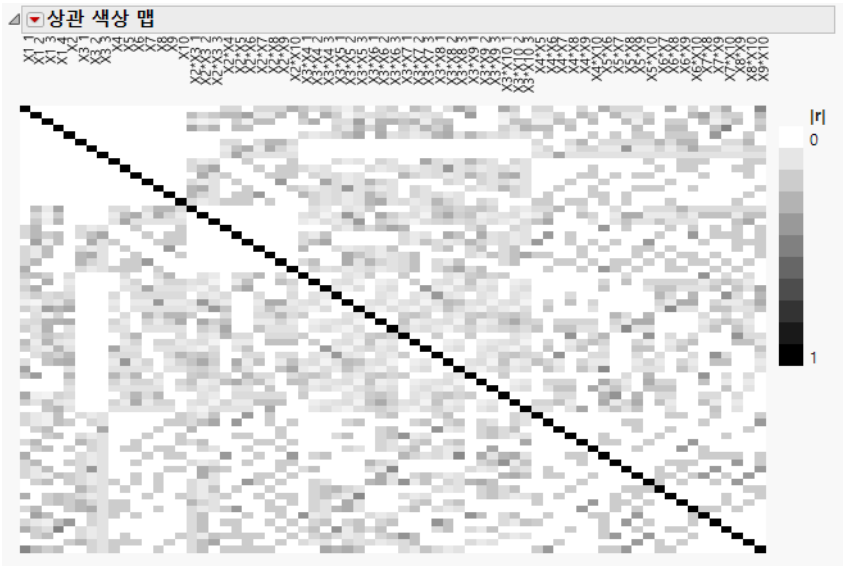
사용자 설계 플랫폼을 사용하면 특정 실험 설정에 맞게 사용자 정의된 최적 설계를 구성할 수 있습니다. 일반적으로 사용자 설계는 다른 방법을 사용하여 구성한 설계보다 비용 효율적입니다. 다음과 같은 작업을 수행할 수 있습니다.

- 다양한 유형의 요인을 입력합니다.
- 설계 공간에 제약 조건을 지정합니다.
- 주어진 런 수를 고려하여 추정이 필요한 효과 및 적합한 효과 (가능한 경우) 를 나타냅니다.
- 예산에 맞는 실험 런 수를 지정합니다.

사용자 설계 플랫폼은 선별, 반응 표면, 혼합물, 랜덤 블록, 분할구, 분할-분할구 및 이원 분할구와 같은 특수 설계 유형을 포함하여 다양한 유형의 설계를 구성합니다.

이 장에는 사용자 설계 플랫폼의 사용 방법을 보여 주는 자세한 예와 플랫폼 정보가 포함되어 있습니다. 자세한 내용은 "사용자 설계의 예" 장(123 페이지)에서 확인하십시오.

그림 4.1 절대 상관의 색상 맵



목차

사용자 설계 개요 65
사용자 설계의 예 66
사용자 설계 생성 82
반응 82
요인 84
요인 제약 조건 정의 90
별칭 항 94
설계 생성 95
설계 97
설계 평가 97
출력 옵션 98
사용자 설계 옵션 100
반응 시뮬레이션 103
행렬 X 저장 105
시작 수 107
설계 검색 시간 108
사용자 설계에 대한 기술 상세 정보 108
랜덤화 제한이 있는 설계 109
변경하기 힘든 수준의 공변량 117
주구 및 하위구 수 117
최적 기준 118
D 효율도 122
좌표 교환 알고리즘 122

사용자 설계 개요

특정 실험 요구에 맞게 사용자 정의된 최적의 설계를 구성하려면 사용자 설계 플랫폼을 사용합니다.

다음과 같은 다양한 유형의 요인을 포함할 수 있습니다.

- 연속형
- 이산 수치형 (원하는 수의 수준 포함)
- 범주형 (원하는 수의 수준 포함)
- 블록 (블록당 지정된 수의 런 포함)
- 공변량
- 혼합
- 상수
- 제어되지 않음

작동 가능 영역 지정

선형 요인 제약 조건 또는 허용되지 않는 조합을 사용하여 실험 영역이 작동 조건을 반영하도록 제한할 수 있습니다. 특히 범주형, 연속형 및 이산 수치형 요인에 대해 제한을 지정할 수 있습니다. 자세한 내용은 "요인 제약 조건 정의"(90 페이지)에서 확인하십시오.

변경하기 힘든 수준의 요인 지정

연속형, 이산 수치형, 범주형 및 혼합 요인의 경우 두 가지 수준의 변경하기 힘든 요인으로 나타낼 수 있습니다. 이러한 난이도 수준은 주구 또는 주구와 분할구로 표시됩니다. 변경하기 힘든 공변량도 지정할 수 있습니다.

기본 관심 효과 지정

가정된 모형을 명시적으로 지정할 수 있습니다. 가정된 모형은 추정할 모든 효과를 이상적으로 포함하는 초기 모형입니다. 주효과, 교호작용, 반응 표면 효과 및 다항식 효과(최대 5 제곱)의 조합이 모형에 포함될 수 있습니다. 추정이 필요한 효과와 적합한 효과를 지정할 수 있습니다. 사용자 설계는 베이지안 최적 방법을 사용하여 런 수에 따라 추정이 적합한 효과를 추정합니다. 자세한 내용은 "모형"(93 페이지)에서 확인하십시오.

런 수 지정

사용자 설계 플랫폼을 사용하면 실험 환경의 예산에 맞도록 런 수를 지정할 수 있습니다. 이 플랫폼은 필요한 효과를 추정하는 데 사용할 수 있는 최소 런 수를 나타내고 기본 런 수를 제공합니다. 이러한 값을 가이드로 사용하여 실현 가능한 런 수를 결정할 수 있습니다. 자세한 내용은 "설계 생성"(95 페이지)에서 확인하십시오.

적절한 설계 유형 구성

사용자 설계는 다양한 설계 유형을 구성할 수 있습니다. 여기에는 전통적 설계와 랜덤 블록 설계가 포함됩니다. 다양한 설계 유형의 예는 "사용자 설계의 예" 장(123 페이지)에서 확인하십시오.

최적 설계 구성

사용자 설계 플랫폼은 특정 요구 사항에 맞는 최적 설계를 구성합니다 . 이 알고리즘은 다음과 같은 몇 가지 최적 기준을 지원합니다 .

- D 최적
- I 최적
- 베이지안 D 및 I 최적 (If Possible 효과 사용)
- A 최적
- 별칭 최적

자세한 내용은 "최적 기준"(118 페이지)에서 확인하십시오.

설계는 좌표 교환 알고리즘(Meyer and Nachtsheim, 1995)을 사용하여 구성됩니다. 자세한 내용은 "좌표 교환 알고리즘"(122 페이지)에서 확인하십시오.

사용자 설계의 예

사용자 설계 도구를 사용하여 와인 맛에 영향을 주는 가공 요인의 효과를 조사하는 실험을 설계합니다. 고용주는 서로 다른 방식으로 가공할 수 있는 두 가지 품종의 Pinot Noir 포도를 재배합니다. Pinot Noir 와인 맛에 영향을 주는 요인을 판별하는 것이 목표입니다. 포도를 가공하기 전에 실험 설계를 설정합니다. 가공된 와인 샘플은 12 개월 동안 숙성된 후 여과 및 병에 담는 과정을 거칩니다. 이 시점에서 전문 와인 테스터가 와인 샘플의 품질을 평가합니다.

반응

포도밭에서 수확한 제품은 대부분 5 개의 대형 와인 유통업체에 판매됩니다. 각 업체의 와인 시음 전문가가 와인 샘플의 품질을 평가할 수 있도록 준비해야 합니다. 연구할 수 있는 요인의 수를 최대화하기 위해 각 전문가가 8 개의 서로 다른 샘플을 평가해야 한다고 결정합니다. 즉, 설계에 40 개의 와인 샘플 또는 런이 포함되어야 합니다.

등급 척도는 0 ~ 20 이며 , 0 이 최하위이고 20 이 최상위입니다. Rating 은 전문가 평가로 구성된 관심 반응 변수입니다. 반응을 최대화하는 와인 관련 요인을 식별하는 것이 목표입니다.

블록 요인

블록 요인은 직접적 관심 대상이 아닌 변동을 처리하는 데 사용됩니다. 특히 한 요인 수준의 관측값이 다른 수준의 관측값보다 더 유사할 것으로 예상될 때 블록 요인이 효과적입니다. 실험에서 한 전문가의 등급은 비슷한 특성을 가지며 다른 전문가의 등급과 다를 가능성이 있습니다. 그러나 관심 사항은 모든 전문가로부터 높은 등급을 받는 특성이 무엇인지 판별하는 것입니다.

각 평가자가 8 개 와인을 시음하므로 Rater 는 블록당 8 개의 런이 있는 블록 요인입니다. 이 실험의 경우 5 명의 평가자만 대상입니다. 평가자 모집단을 더 크게 일반화하는 것에는 관심이 없습니다.

공정 요인

연구에 필요한 9 개의 공정 요인을 식별했습니다. 여기에는 포도 품종, 포도를 재배한 밭 그리고 가공과 관련된 기타 요인 7 개가 포함됩니다. 이러한 요인을 필요에 따라 조합하여 실험할 수 있습니다. 또한 실험 중간에 원하는 대로 요인에 변화를 줄 수도 있습니다. 이러한 요인은 모두 실험에서 상대적으로 변경하기 "쉬운" 요인입니다. 요인 변경을 지정하는 방법에 대한 자세한 내용은 "변경 및 랜덤 블록"(87 페이지)에서 확인하십시오.

표 4.1 에는 요인과 요인 수준이 나와 있습니다. 이러한 모든 요인은 범주형입니다. 요인과 해당 수준은 샘플 데이터의 Design Experiment 폴더에 있는 Wine Factors.jmp 요인 테이블에도 제공됩니다.

가능한 모든 요인 조합으로 실험하기 위해 필요한 런 수는 $4 \times 2^8 = 1024$ 회입니다. 그러나 이 예에서는 40 회의 런으로 최상의 설계를 구성할 수 있습니다.

표 4.1 와인 시음 실험의 공정 요인 및 수준

요인	수준
Variety	Bernard, Dijon
Field	1, 2, 3, 4
De-Stem	No, Yes
Yeast	Cultured, Wild
Temperature	High, Low
Press	Hard, Soft
Barrel Age	New, 2 Years
Barrel Seasoning	Air, Kiln
Filtering	No, Yes

실험 목표, 요인 및 반응을 정의했으므로 사용자 설계를 생성할 수 있습니다.

반응 추가

이 사용자 설계 예의 경우 반응과 반응 목표를 추가하고 필요한 경우 하한 , 상한 및 중요도를 추가합니다 . 여기서 반응은 Rating 입니다 .

1. **DOE > 사용자 설계**를 선택합니다 .

2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Rating 을 입력합니다 .

   기본 목표는 최대화로 되어 있습니다 . 맛 등급을 최대화해야 하므로 목표를 변경하지 마십시오 .
3. " 하한 " 아래를 클릭하고 0 을 입력합니다 .

   최하위 등급이 0 입니다 .
4. " 상한 " 아래를 클릭하고 20 을 입력합니다 .

   최상위 등급이 20 입니다 .
5. " 중요도 " 아래 영역은 비워 둡니다 .

   반응이 하나뿐이므로 기본적으로 반응의 중요도가 1 로 지정됩니다 .

그림 4.2(69 페이지) 에서는 완료된 " 반응 " 개요를 보여 줍니다 .

수동 또는 자동으로 요인 추가

이 사용자 설계 예의 경우 요인을 수동으로 입력하거나 , 요인과 설정이 포함된 기존 테이블을 사용하여 자동으로 입력합니다 .

- 새 실험을 설계하는 경우 먼저 수동으로 요인을 입력해야 합니다 . 자세한 내용은 " 수동으로 요인 추가 "(68 페이지) 에서 확인하십시오 .
- " 요인 저장 " 옵션을 사용하여 요인을 저장한 후에는 저장된 테이블을 사용하여 요인을 자동으로 불러올 수 있습니다 . 자세한 내용은 " 요인 불러오기를 사용하여 자동으로 요인 추가 "(70 페이지) 에서 확인하십시오 .

두 방법 모두 네 가지 개요 , 즉 " 요인 제약 조건 정의 ", " 모형 ", " 별칭 항 " 및 " 설계 생성 " 개요를 " 사용자 설계 " 창에 추가합니다 .

수동으로 요인 추가

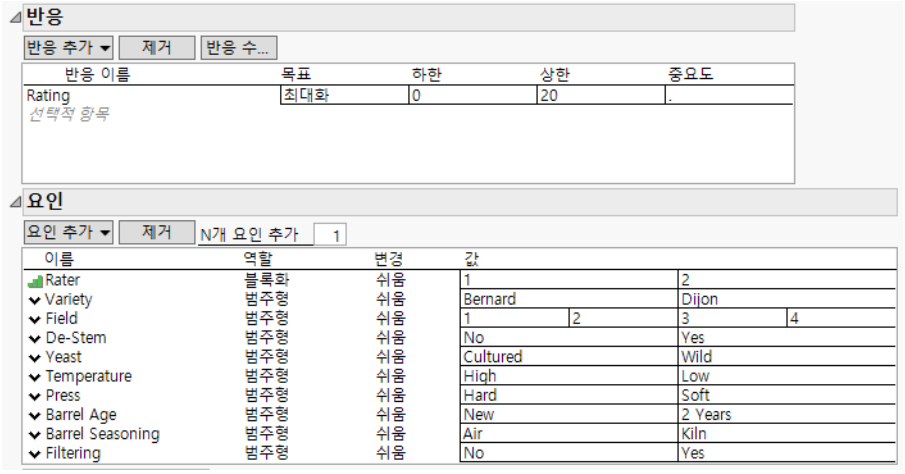
1. 먼저 블록 요인 Rater 를 추가합니다 . **요인 추가 > 블록화 > 블록당 8 회 런**을 클릭합니다 .
2. 기본 이름 X1 을 Rater 로 변경합니다 .

   " 역할 " 이 " 블록화 " 로 설정되어 있습니다 . 또한 " 값 " 설정에 1 만 나타납니다 . 원하는 런 수가 지정되지 않으면 블록 수를 결정할 수 없기 때문입니다 . " 설계 생성 " 개요에서 " 런 수 " 를 지정하면 Rater 의 수준 수가 필요한 값으로 업데이트됩니다 .
3. **요인 추가 > 범주형 > 2 수준**을 클릭합니다 .
4. 기본 이름 X2 를 Variety 로 변경합니다 .

   요청한 대로 " 역할 " 이 " 범주형 " 으로 설정되고 " 변경 " 이 " 쉬움 "(기본값) 으로 설정되어 있습니다 .
5. L1 과 L2 를 클릭하고 각각 Bernard 와 Dijon 으로 변경합니다 .
6. **요인 추가 > 범주형 > 4 수준**을 클릭합니다 .
7. 기본 이름 X3 을 Field 로 변경합니다 .

8. L1, L2, L3 및 L4 를 클릭하고 각각 1, 2, 3, 4 로 변경합니다 .
9. **요인 추가 > 범주형 > 2 수준**을 클릭합니다 .
10. 기본 이름 X4 를 De-Stem 으로 변경합니다 .
11. L1 과 L2 를 클릭하고 각각 No 와 Yes 로 변경합니다 .
12. **N 개 요인 추가** 옆에 6 을 입력한 후 **요인 추가 > 범주형 > 2 수준**을 클릭합니다 . 이렇게 하면 2 수준을 갖는 6 개의 범주형 요인이 설계에 추가됩니다 .
13. 다음과 같이 기본 요인 이름과 값을 변경합니다 .
    - Yeast(Cultured, Wild)
    - Temperature(High, Low)
    - Press(Hard, Soft)
    - Barrel Age(New, Two Years)
    - Barrel Seasoning(Air, Kiln)
    - Filtering(No, Yes)

그림 4.2 완료된 반응 및 요인 개요



14. **계속**을 클릭합니다 .

다음 개요가 " 사용자 설계 " 창에 추가됩니다 .

- 요인 제약 조건 정의 (이 예에서 사용되지 않음)
- 모형
- 별칭 항
- 설계 생성

요인 불러오기를 사용하여 자동으로 요인 추가

다음과 같이 요인 정보가 포함된 테이블을 사용하여 요인을 입력합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Wine Factors.jmp 를 엽니다 .
2. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

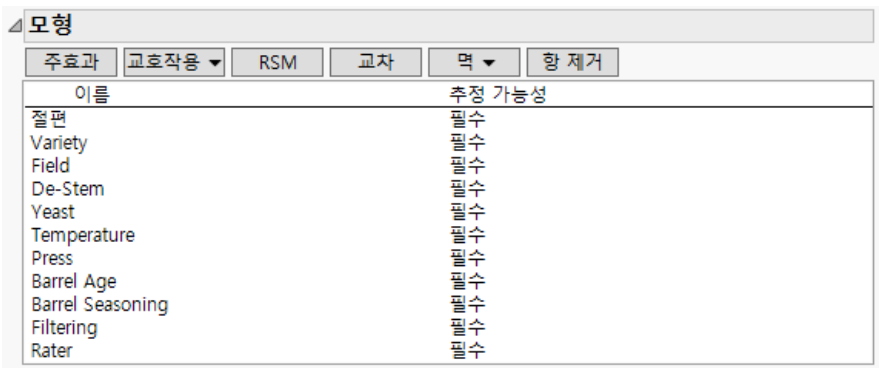
요인을 불러오면 " 사용자 설계 " 창이 자동으로 업데이트됩니다 . 다음 개요가 " 사용자 설계 " 창에 추가됩니다 .

- 요인 제약 조건 정의 (이 예에서 사용되지 않음)
- 모형
- 별칭 항
- 설계 생성

모형 정의

이 사용자 설계 예의 경우 "모형" 개요에는 모든 주효과가 "필수"로 표시되어 있으며 이는 설계에서 모든 주효과를 추정할 수 있어야 함을 나타냅니다. 이 가정된 모형에는 주효과에 대한 관심만 반영됩니다. 그러나 다른 효과를 추정하려면 "모형" 개요에 해당 효과를 추가하면 됩니다. 모형에 대한 자세한 내용은 "모형"(93 페이지)에서 확인하십시오.

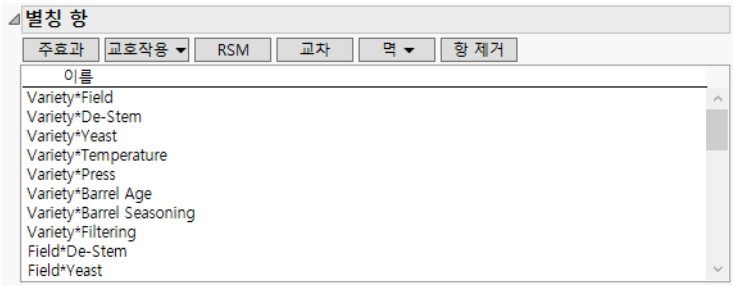
그림 4.3 모형 개요



별칭 항 정의

"별칭 항" 개요에서는 나중에 나타나는 "별칭 행렬"에 표시할 효과를 지정합니다. 별칭 행렬은 "모형" 항과 "별칭 항" 개요에 나열된 효과 사이의 별칭 관계를 보여 줍니다. "별칭 항" 개요 노드를 열어 모든 2 요인 교호작용이 나열되어 있는지 확인합니다. 별칭 행렬에 대한 자세한 내용은 "별칭 행렬"(74 페이지)에서 확인하십시오.

그림 4.4 별칭 항 개요 (일부분)



설계를 복제할 난수 시드값 설정

사용자 설계 알고리즘은 랜덤 시작 설계로 시작합니다 . 이 예제 또는 교육용 설정 등에서 사용하는 설계를 복제하려면 시작 설계를 정의하는 데 사용되는 난수 시드값을 설정합니다 . 그림 4.5 에 나와 있는 것과 똑같은 런 및 런 순서를 갖는 설계를 얻으려면 난수 시드값과 시작 수를 다음과 같이 설정합니다 .

1. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 .
2. 100526291(난수 시드값) 을 입력합니다 .
3. **확인**을 클릭합니다 .
4. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택합니다 .
5. 2 를 입력합니다 .
6. **확인**을 클릭합니다 .

참고 : 난수 시드값과 시작 수를 설정하면 이 예에 나오는 것과 정확히 같은 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

설계 생성

" 설계 생성 " 개요에서는 설계의 구조와 크기에 대한 추가 정보를 입력할 수 있습니다 . 이 예에서 설계의 기본 런 수는 16 회로 표시됩니다 . 그러나 5 명의 평가자가 각각 8 개 와인을 시음할 수 있습니다 . 즉 , 런 수가 40 회인 설계를 생성하려고 합니다 . 설계 런 수를 다음과 같이 변경합니다 .

1. **런 수** 아래의 **사용자 지정** 상자에 **40** 을 입력합니다 .

   런 반복을 원하지 않으므로 **반복 런 수** 설정을 **0** 으로 유지합니다 .
2. **설계 생성**을 클릭합니다 .

"설계" 및 "설계 평가" 개요가 "사용자 설계" 창에 추가됩니다 . "출력 옵션" 패널도 나타납니다 .

설계 확인

"설계" 개요에는 생성한 사용자 설계의 런이 표시됩니다. 나중에 "출력 옵션"에서 런 순서를 랜덤화할 수 있습니다. 지금은 이 설계가 실험에 적합한지 확인합니다. 예를 들어 5명의 평가자가 와인을 8개씩 평가하는지, 필요한 모든 요인이 표시되는지, 그리고 실현 불가능한 조합을 나타내는 설정이 없는지 확인합니다.

그림 4.5 와인 실험 설계

설계

런	Rater	Variety	Field	De-Stem	Yeast	Temperature	Press	Barrel Age	Barrel Seasoning	Filtering
1	1	Dijon	3	No	Wild	High	Soft	New	Kiln	Yes
2	2	Bernard	4	Yes	Cultured	Low	Hard	New	Kiln	No
3	3	Bernard	2	Yes	Wild	High	Hard	New	Air	No
4	4	Bernard	1	Yes	Wild	Low	Hard	2 Years	Air	Yes
5	5	Dijon	1	Yes	Wild	Low	Soft	New	Kiln	No
6	1	Dijon	2	Yes	Cultured	Low	Soft	2 Years	Air	Yes
7	2	Dijon	1	No	Cultured	High	Hard	New	Air	No
8	3	Dijon	4	No	Cultured	High	Soft	New	Air	Yes
9	4	Dijon	4	Yes	Wild	High	Hard	New	Kiln	Yes
10	5	Bernard	2	No	Wild	Low	Hard	New	Air	Yes
11	1	Bernard	3	Yes	Wild	High	Hard	New	Air	No
12	2	Dijon	1	Yes	Wild	High	Soft	2 Years	Air	Yes
13	3	Bernard	4	Yes	Wild	Low	Soft	2 Years	Kiln	No
14	4	Bernard	3	No	Cultured	Low	Soft	New	Air	No
15	5	Dijon	2	No	Cultured	High	Soft	2 Years	Kiln	No
16	1	Dijon	1	Yes	Cultured	Low	Hard	2 Years	Kiln	Yes
17	2	Bernard	2	No	Cultured	Low	Hard	2 Years	Air	No
18	3	Dijon	2	Yes	Cultured	High	Hard	New	Kiln	No
19	4	Bernard	1	No	Cultured	Low	Hard	New	Kiln	Yes
20	5	Bernard	3	Yes	Cultured	High	Hard	2 Years	Air	Yes
21	1	Bernard	4	No	Wild	High	Hard	2 Years	Kiln	No
22	2	Bernard	4	Yes	Cultured	High	Soft	2 Years	Kiln	Yes
23	3	Dijon	1	No	Wild	Low	Hard	2 Years	Kiln	No
24	4	Dijon	2	Yes	Wild	High	Soft	2 Years	Air	No
25	5	Bernard	1	No	Cultured	High	Soft	New	Kiln	No
26	1	Bernard	2	No	Wild	Low	Soft	New	Kiln	Yes
27	2	Dijon	3	No	Wild	High	Hard	New	Kiln	Yes
28	3	Bernard	1	No	Wild	High	Soft	2 Years	Air	Yes
29	4	Bernard	3	Yes	Cultured	High	Soft	2 Years	Kiln	No
30	5	Dijon	3	Yes	Wild	Low	Hard	2 Years	Kiln	No
31	1	Dijon	4	No	Cultured	Low	Hard	2 Years	Air	No
32	2	Bernard	2	Yes	Wild	Low	Soft	New	Kiln	Yes
33	3	Bernard	3	No	Cultured	Low	Soft	2 Years	Kiln	Yes
34	4	Dijon	2	No	Cultured	High	Hard	2 Years	Kiln	Yes
35	5	Dijon	4	Yes	Cultured	Low	Soft	New	Air	Yes
36	1	Bernard	1	Yes	Cultured	High	Soft	New	Air	No
37	2	Dijon	3	No	Wild	Low	Soft	2 Years	Air	No
38	3	Dijon	3	Yes	Cultured	Low	Hard	New	Air	Yes
39	4	Dijon	4	No	Wild	Low	Soft	New	Air	No
40	5	Bernard	4	No	Wild	High	Hard	2 Years	Air	Yes

설계 평가

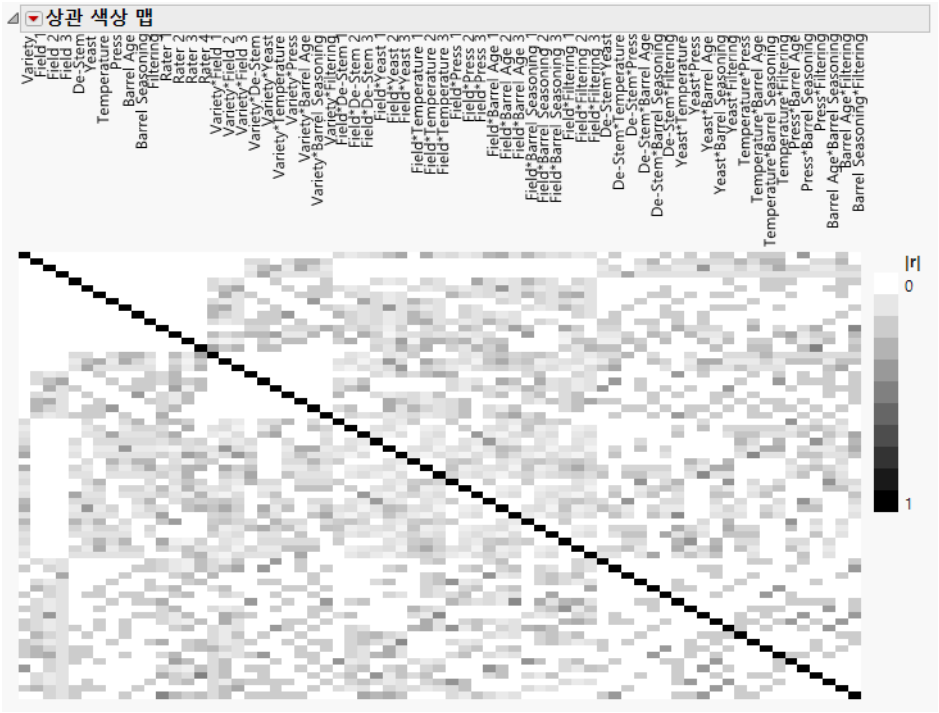
"설계 평가" 개요에서는 사용자 설계를 평가하는 다양한 방법을 제공합니다. 이 예에서는 "설계 평가" 개요를 열고 "상관 색상 맵", "별칭 행렬" 및 "설계 진단"을 검토합니다.

참고 : "설계 평가" 개요에 대한 자세한 내용은 "설계 평가" 장 (403 페이지) 에서 확인하십시오 .

상관 색상 맵

"상관 색상 맵"에는 "모형" 개요 또는 "별칭 항" 개요에 나타나는 두 효과 간의 절대 상관 값이 표시됩니다.

그림 4.6 상관 색상 맵



주효과는 맵의 왼쪽 위에 15개의 항으로 표시됩니다. 주효과와 다른 주효과의 상관에 해당하는 흰색은 상관계수 0을 나타냅니다. 이는 모든 주효과가 직교되고 독립적으로 추정할 수 있음을 의미합니다.

그림 4.6에서 검은색은 주대각 위에만 있습니다. 검은색은 절대 상관 1을 나타내며 각 항이 자신과 완벽하게 상관된다는 것을 반영합니다. 따라서 주효과는 모든 이원 교호작용과 완전히 교락되지 않습니다. 실제로 주효과와 이원 교호작용의 절대 상관 값이 매우 낮습니다. 이는 주효과의 추정값이 활성 이원 교호작용이 있을 때만 약간 편향될 수 있음을 의미합니다.

팁 : 효과 사이의 절대 상관을 보려면 커서로 색상 맵의 셀을 가리킵니다 . 범례 아래를 마우스 오른쪽 버튼으로 클릭하여 상관을 데이터 테이블에 저장하십시오 .

별칭 행렬

별칭 행렬의 왼쪽 열에 모형 효과가 나열됩니다. 주어진 모형 효과에 대해 열 항목은 열 효과(활성화된 경우)로 인해 모형 효과의 추정값이 편향되는 정도를 나타냅니다.

그림 4.7 별칭 행렬 (일부분)

별칭 행렬

효과	Variety*Field 1	Variety*Field 2	Variety*Field 3	Variety*De-Stem	Variety*Yeast	Variety*Temperature	Variety*Press	Variety*Barrel Age
절편	0	0	0	0	0	0	0	0
Variety	0	0	0	0	0	0	0	0
Field 1	0	0	0	0.141	0.141	0	-0.14	0.283
Field 2	0	0	0	0.082	-0.24	-0.33	0.082	0.327
Field 3	0	0	0	-0.12	0.346	0.115	-0.12	-0.12
De-Stem	0.141	0.082	-0.12	0	0	-0.2	0	0
Yeast	0.141	-0.24	0.346	0	0	0	-0.2	0
Temperature	0	-0.33	0.115	-0.2	0	0	0	-0.2
Press	-0.14	0.082	-0.12	0	-0.2	0	0	0
Barrel Age	0.283	0.327	-0.12	0	0	-0.2	0	0
Barrel Seasoning	0.283	0.163	0.115	0	-0.2	0.2	0.4	0
Filtering	-0.14	-0.08	0.115	-0.2	0.2	0.2	0	0.2
Rater 1	0	0	0	-0.16	-0.16	0	-0.16	0.158
Rater 2	-0.26	0.149	-0.21	-0.27	0.274	-0.37	-0.09	-0.09
Rater 3	0.091	-0.05	0.075	0	-0.19	-0.06	-0.26	-0.26
Rater 4	0.354	-0.2	0.289	0	0.25	-0.25	0	0

모형 효과 Barrel Seasoning을 예로 들어보겠습니다. Variety*Press가 활성화된 경우 Barrel Seasoning 효과의 추정 기대값이 해당 효과의 비편향 추정값과 다릅니다. 이 차이는 Variety*Press 효과의 0.4배와 같습니다. 따라서 유의한 Barrel Seasoning 추정 효과로 보이는 것이 실제로는 유의한 Variety*Press 효과일 수 있습니다.

설계 진단

"설계 진단" 개요에서는 설계 효율도 정보를 제공합니다. 효율도 측도는 설계를 이론상 최적 설계(존재하지 않을 수 있음)와 비교합니다. 효율도 값은 이 최적 설계의 효율도에 대한 설계 효율도의 비율(%)입니다. 효율도 측도에 대한 자세한 내용은 "설계 평가" 장의 "추정 효율도"(428 페이지)에서 확인하십시오.

그림 4.8 설계 진단 개요

설계 진단

D 최적 설계	
D 효율도	100
G 효율도	100
A 효율도	100
예측 평균 분산	0.4
설계 생성 시간(초)	0.05

D, G 및 A 효율도 값이 모두 100%입니다. 생성된 설계는 주효과에 대해 직교이므로 이 세 가지 효율도 기준에 따라 주효과 모형에 최적 설계입니다.

"설계 진단" 개요의 첫 번째 줄은 설계가 D 효율도 기준을 최적화하도록 구성되었음을 나타냅니다. 자세한 내용은 "사용자 설계 옵션"(100 페이지)에서 "최적 기준" 설명을 참조하십시오. 이 경우에는 설계의 D 효율도가 100% 입니다.

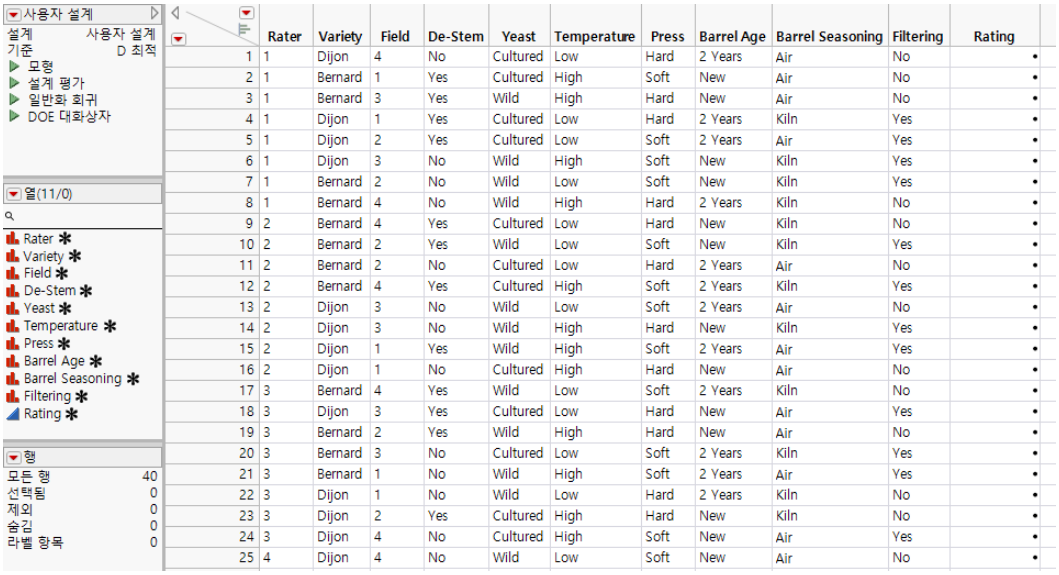
설계 테이블 옵션 지정

사용자 설계 생성의 마지막 단계는 설계 테이블의 옵션을 지정하고 테이블을 생성하는 것입니다. "출력 옵션" 패널을 사용하여 데이터 테이블에서 런 순서를 지정합니다. 기본 선택 항목인 **블록 내 랜덤화**가 이 예에 적합합니다. **테이블 생성**을 클릭하십시오.

그림 4.9의 테이블과 유사한 사용자 설계 테이블이 생성되어 열립니다.

참고 : 테이블을 생성하는 알고리즘이 랜덤 시작 설계를 사용하므로 테이블이 그림과 다르게 보일 수 있습니다 . 그림 4.9 에 나와 있는 것과 똑같은 테이블을 얻으려면 " 설계를 복제할 난수 시드값 설정 "(71 페이지) 에 설명된 대로 난수 시드값과 시작 수를 설정합니다 .

그림 4.9 사용자 설계 테이블



	Rater	Variety	Field	De-Stem	Yeast	Temperature	Press	Barrel Age	Barrel Seasoning	Filtering	Rating
1	1	Dijon	4	No	Cultured	Low	Hard	2 Years	Air	No	•
2	1	Bernard	1	Yes	Cultured	High	Soft	New	Air	No	•
3	1	Bernard	3	Yes	Wild	High	Hard	New	Air	No	•
4	1	Dijon	1	Yes	Cultured	Low	Hard	2 Years	Kiln	Yes	•
5	1	Dijon	2	Yes	Cultured	Low	Soft	2 Years	Air	Yes	•
6	1	Dijon	3	No	Wild	High	Soft	New	Kiln	Yes	•
7	1	Bernard	2	No	Wild	Low	Soft	New	Kiln	Yes	•
8	1	Bernard	4	No	Wild	High	Hard	2 Years	Kiln	No	•
9	2	Bernard	4	Yes	Cultured	Low	Hard	New	Kiln	No	•
10	2	Bernard	2	Yes	Wild	Low	Soft	New	Kiln	Yes	•
11	2	Bernard	2	No	Cultured	Low	Hard	2 Years	Air	No	•
12	2	Bernard	4	Yes	Cultured	High	Soft	2 Years	Kiln	Yes	•
13	2	Dijon	3	No	Wild	Low	Soft	2 Years	Air	No	•
14	2	Dijon	3	No	Wild	High	Hard	New	Kiln	Yes	•
15	2	Dijon	1	Yes	Wild	High	Soft	2 Years	Air	Yes	•
16	2	Dijon	1	No	Cultured	High	Hard	New	Air	No	•
17	3	Bernard	4	Yes	Wild	Low	Soft	2 Years	Kiln	No	•
18	3	Dijon	3	Yes	Cultured	Low	Hard	New	Air	Yes	•
19	3	Bernard	2	Yes	Wild	High	Hard	New	Air	No	•
20	3	Bernard	3	No	Cultured	Low	Soft	2 Years	Kiln	Yes	•
21	3	Bernard	1	No	Wild	High	Soft	2 Years	Air	Yes	•
22	3	Dijon	1	No	Wild	Low	Hard	2 Years	Kiln	No	•
23	3	Dijon	2	Yes	Cultured	High	Hard	New	Kiln	No	•
24	3	Dijon	4	No	Cultured	High	Soft	New	Air	Yes	•
25	4	Dijon	4	No	Wild	Low	Soft	New	Air	No	•

다음 사항에 유의하십시오 .

- 설계 생성 중에 테이블 패널에 " 모형 ", " 설계 평가 " 및 "DOE 대화상자 " 스크립트가 추가되었습니다 . " 모형 " 스크립트는 " 사용자 설계 " 대화상자 창에서 " 필수 " 로 지정한 효과가 포함된 " 모형 적합 " 창을 엽니다 . "DOE 대화상자 " 스크립트는 설계 테이블을 생성하는 데 사용된 창을 다시 생성합니다 .

- "열" 패널에서 요인과 반응의 오른쪽에 있는 별표는 데이터 테이블의 열에 저장된 열 특성을 나타냅니다. 이러한 열 특성은 데이터 분석에 사용됩니다. 열 특성에 대한 자세한 내용은 "요인"(84 페이지) 및 "요인 열 특성"(89 페이지)에서 확인하십시오.

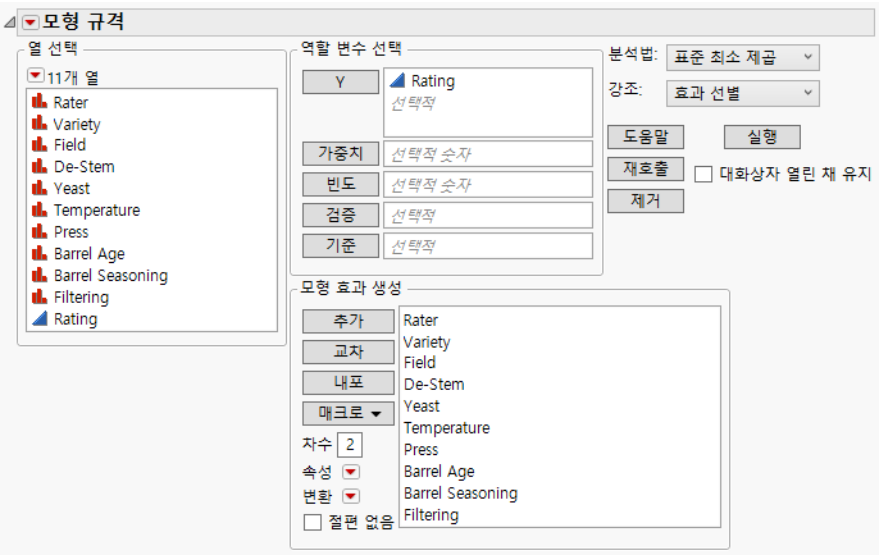
사용자 설계 분석

실험을 실행하고 데이터를 컴파일하고 나면 결과를 분석할 준비가 된 것입니다. 사용자 설계 테이블의 Rating 열을 사용하여 Rating 데이터를 기록합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Wine Data.jmp 를 엽니다.

   Wine Data.jmp 테이블은 기록한 실험 결과를 포함한다는 점만 제외하면 그림 4.9 에 표시된 사용자 설계 테이블과 정확히 동일합니다.
2. 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다.

그림 4.10 와인 실험의 모형 규격 창



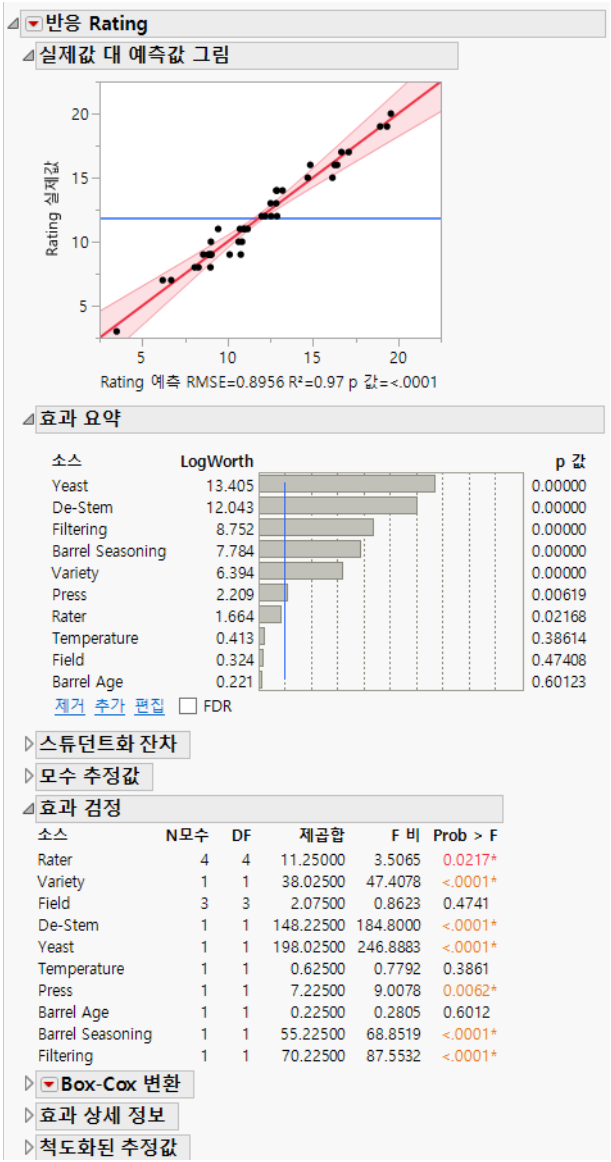
블록 요인 Rater 가 랜덤 블록 효과 대신 고정 효과로 추가되어 있습니다. 이는 5 명의 평가자가 명확하게 선택되었고 큰 모집단의 랜덤 표본이 아니므로 적합합니다.

3. **실행**을 클릭합니다.

전체 모형 결과 해석

사용자 설계 예에서 얻은 결과 모형이 아래에 나와 있습니다.

그림 4.11 모형 적합 결과 (일부분)



다음 사항에 유의하십시오 .

- " 실제값 대 예측값 그림 " 에 적합 결여의 뚜렷한 증거가 없습니다 .
- " 실제값 대 예측값 " 그림과 그 아래의 p 값에 따르면 모형이 유의합니다 .
- " 효과 검정 " 보고서는 7 개 모형 항이 0.05 수준에서 유의함을 나타냅니다 . Field, Temperature 및 Barrel Age 는 유의하지 않습니다 .

- "효과 요약" 보고서에는 이러한 효과가 유의성이 큰 것부터 순서대로 나열됩니다. LogWorth 값이 클수록 p 값이 작고 유의성이 큽니다 .

모형 축소

다음과 같이 비활성으로 식별된 효과를 제거하여 사용자 설계의 결과 모형을 축소합니다 .

1. " 효과 요약 " 보고서에서 Ctrl 키를 누르고 Temperature, Field 및 Barrel Age 를 선택합니다 .
2. **제거**를 클릭합니다 .

   보고서가 업데이트되어 이 세 효과가 제거된 모형 적합이 표시됩니다 .

프로파일러로 모형 결과 해석

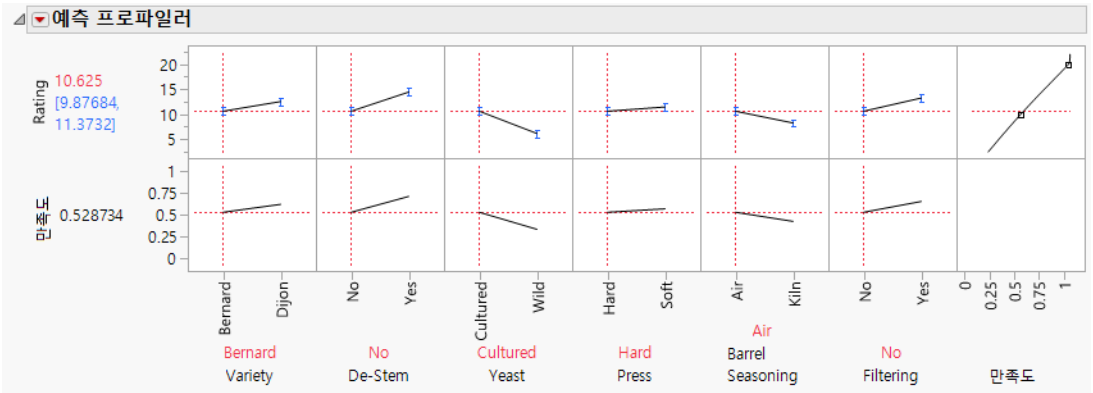
축소 모형의 "실제값 대 예측값 그림"은 적합 결여 문제가 없음을 보여 줍니다. "효과 요약" 및 "효과 검정" 보고서는 나머지 7 개 항이 0.05 수준에서 유의함을 보여 줍니다. 예측 프로파일러를 사용하여 축소된 모형을 추가적으로 탐색합니다.

그림 4.12에서는 예측 프로파일러를 보여 줍니다. 반응 목표를 "최대화"로 지정하고 하한과 상한을 각각 0 과 20 으로 설정했습니다. 이러한 제한을 설정하여 "반응 한계" 열 특성이 "사용자 설계" 테이블의 Rating 열에 저장되었습니다. 예측 프로파일러는 반응 한계 정보를 사용하여 그림 4.12 의 위쪽 행에서 맨 오른쪽 그림에 나타나는 만족도 함수를 구성합니다. 아래쪽 행은 만족도 트레이스를 표시합니다.

위쪽 행의 처음 6 개 그림은 예측 모형의 트레이스를 표시합니다. 각 요인에 대해 그림의 선은 다른 모든 요인이 빨간색 수직 파선에 의해 정의된 값으로 설정될 때 Rating 이 어떻게 변하는지 보여 줍니다. 기본적으로 프로파일러는 낮은 값으로 설정된 범주형 요인과 함께 나타납니다. 요인의 설정을 변경하면 와인에 대한 예측 Rating 이 어떻게 변하는지 확인할 수 있습니다. 평균 예측 Rating 에 대해 신뢰 구간이 제공됩니다.

프로파일러에 표시된 요인 중에 Rater 가 포함되어 있지 않습니다. Rater 는 블록 변수이기 때문입니다. 변동을 설명하기 위해 Rater 를 포함했지만 Rater 는 공정 요인 설정의 최적화에 있어서 직접적 관심 요인이 아닙니다. 지정된 설정을 사용한 와인의 예측 Rating 은 모든 평가자가 해당 와인에 대해 평가한 예측 등급의 평균입니다.

그림 4.12 축소 모형의 프로파일러



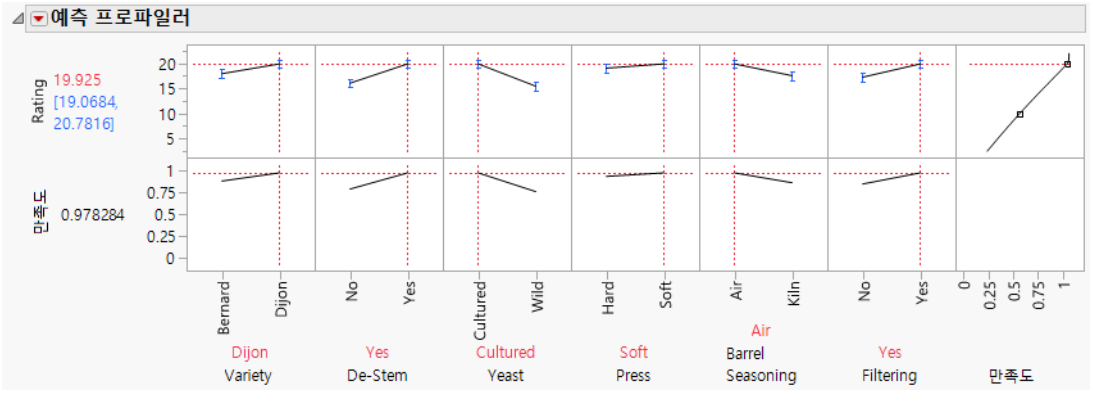
요인 설정 최적화

예측 프로파일러를 사용하여 사용자 설계 결과를 기반으로 최적 설정을 식별합니다 . 평가자 전체에서 Rating 을 최대화하는 설정을 식별하려고 합니다 .

1. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

   " 예측 프로파일러 " 의 빨간색 수직 파선이 업데이트되어 각 요인에 대한 최적 설정이 표시됩니다 . 최적 설정의 예측 등급 결과는 19.925 입니다 . 일반적으로 동일한 최적값을 얻는 요인 설정 집합이 여러 개 있을 수 있습니다 .

그림 4.13 요인 설정이 최적화된 예측 프로파일러



2. 모든 런의 예측 등급을 보려면 예측 계산식을 저장합니다 . " 반응 Rating" 의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식**을 선택합니다 .

   예측 계산식 Rating 열이 데이터 테이블에 추가됩니다 . 전체 런 중 하나 (33 행) 에 Rater 5 가 최대 등급 20 을 지정했습니다 . Rater 5 가 평가하는 이 런의 예측 등급은 19.550 입니다 . 그러나 33 행 실험은 최적 설정으로 실행되었습니다 . " 예측 프로파일러 " 에서 이러한 설정에

지정된 예측값 19.925 는 5 명의 모든 평가자가 해당 런에 대해 평가한 예측 등급의 평균을 구하여 계산됩니다 .

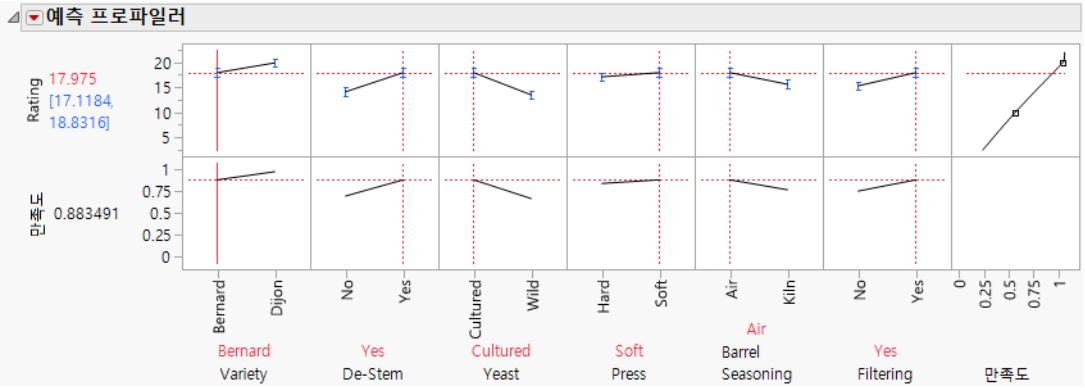
요인 수준 잠금

프로파일러에서 요인을 잠금으로써 특정 조건에서의 최적화를 탐색할 수 있습니다 . 설계 반응을 최대화한 결과 Dijon 품종의 포도에서 최적 등급을 얻는다는 사실을 알게 되었습니다 (그림 4.13). 관리자는 Bernard 포도를 재배하는 밭에 어린 Dijon 포도나무를 새로 심는 것은 비용이 많이 발생할 수 있다는 점을 지적합니다 . 따라서 Bernard 포도에 대한 최적의 공정 설정과 예측 등급을 찾아야 합니다 .

1. " 예측 프로파일러 " 의 Variety 그림에서 빨간색 수직 파선을 Bernard 로 드래그합니다 .
2. Ctrl 키를 누르고 Variety 그림 중 하나를 클릭합니다 .

   " 요인 설정 " 창이 나타납니다 .
3. **요인 설정 잠금**을 선택하고 **확인**을 클릭합니다 .
4. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 4.14 Bernard 품종에 대한 최적 설정을 사용한 예측 프로파일러



모형에 교호작용 항이 없으므로 최적 설정이 변경되지 않습니다. 이 설정에서 예측 등급은 17.975 입니다.

프로파일러에 Rater 추가

Rater 수준에 대한 프로파일러 트레이스를 보려면 다음 단계를 수행하십시오 .

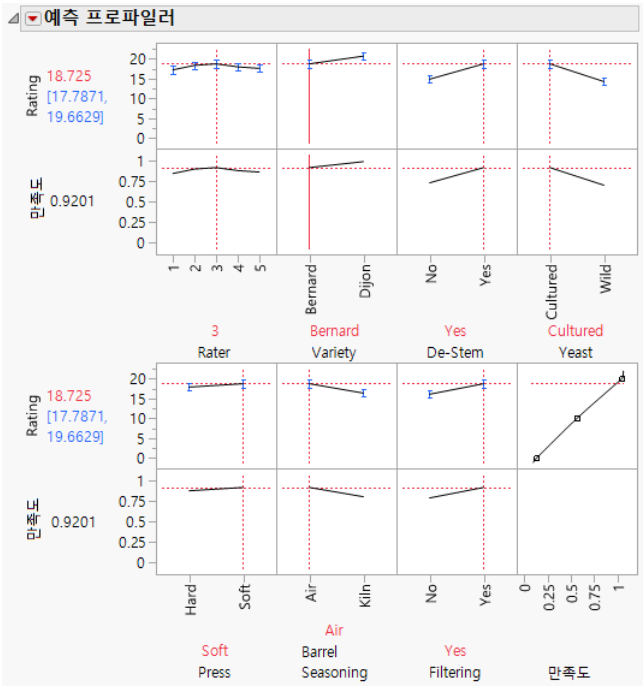
1. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **요인 격자 재설정**을 선택합니다 .

   Rater 를 포함하여 모든 요인에 대한 열이 포함된 " 요인 설정 " 창이 나타납니다 . Rater 아래의 **표시** 옆에 있는 상자는 선택되어 있지 않습니다 . 이는 Rater 가 예측 프로파일러에 표시되지 않음을 나타냅니다 .

2. Rater 아래에서 **표시** 행의 상자를 선택합니다 .
3. Rater 아래에서 **요인 설정 잠금** 행의 상자를 선택 취소합니다 .
4. **확인**을 클릭합니다 .

   프로파일러가 업데이트되어 Rater 그림이 표시됩니다 .
5. Rater 위의 그림 중 하나를 클릭합니다 .

그림 4.15 Rater 를 표시하는 축소 모형의 프로파일러



빨간색 수직 파선이 나타납니다 . 이 선을 드래그하여 각 평가자의 트레이스를 확인할 수 있습니다 . Variety 가 Bernard 에서는 계속 잠금 상태입니다 . Variety 잠금을 해제하려면 Ctrl 키를 누르고 Variety 그림 중 하나를 클릭합니다 . " 요인 설정 " 창이 나타나면 **요인 설정 잠금**을 선택 취소합니다 .

사용자 설계 예에 대한 요약

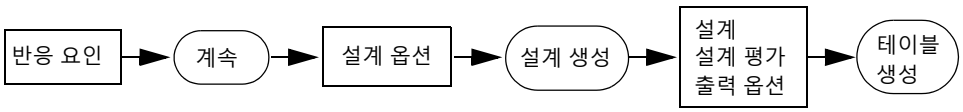
와인 시음 실험에서 런을 40 개만 사용하여 9 개 요인 중 Pinot Noir 포도 등급에 영향을 미치는 6 개의 요인을 식별했습니다. 해당 요인에 대한 최적 설정에서 최대 20 점 중 19.925 의 예측 등급을 얻을 수 있음을 확인했습니다. 또한 두 품종의 포도에 모두 맞는 최적 설정도 식별했습니다.

이 섹션에서는 "사용자 설계" 창의 개요를 사용하여 설계를 구성했습니다. 다음 섹션에서는 각 개요 및 설계 단계에 대해 자세히 설명합니다.

사용자 설계 생성

DOE > 사용자 설계를 선택하여 사용자 설계를 생성합니다. 먼저 반응 및 요인을 정의합니다. 그런 다음 계속해서 설계 옵션, 생성 및 평가를 진행합니다. 설계 과정은 그림 4.16의 흐름을 따릅니다.

그림 4.16 사용자 설계 흐름

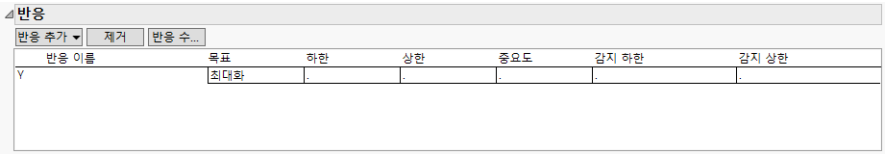


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다. 이 옵션을 사용하면 반응 이름, 목표, 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다.

그림 4.17 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다. " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다. " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다.

JMP PRO 함수 (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다. 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다.

제거 선택한 반응을 제거합니다.

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다. 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다. 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다. 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다.

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응 열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .
- " 없음 " 목표는 최적화에 관한 목표가 없음을 나타냅니다 . 만족도 함수가 구성되지 않습니다 .

참고 : 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다 . 데이터 테이블에서 반응 열의 " 열 정보 " 창을 열고 (**열** > **열 정보**) 원하는 목표값을 입력하십시오 .

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다 . 이러한 값은 반응에 대한 가중치로 처리됩니다 . 반응이 하나뿐이면 기본적으로 1 로 설정되므로 중요도를 지정할 필요가 없습니다 .

JMP PRO **감지 한계** 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다 . 감지 한계는 열 특성으로 저장됩니다 . 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오 .

반응 개요 편집

" 반응 " 개요에서 다음 사항에 유의하십시오 .

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다 .
- 목표를 클릭하여 변경할 수 있습니다 .
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다 .
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다 .

반응 한계 열 특성

반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

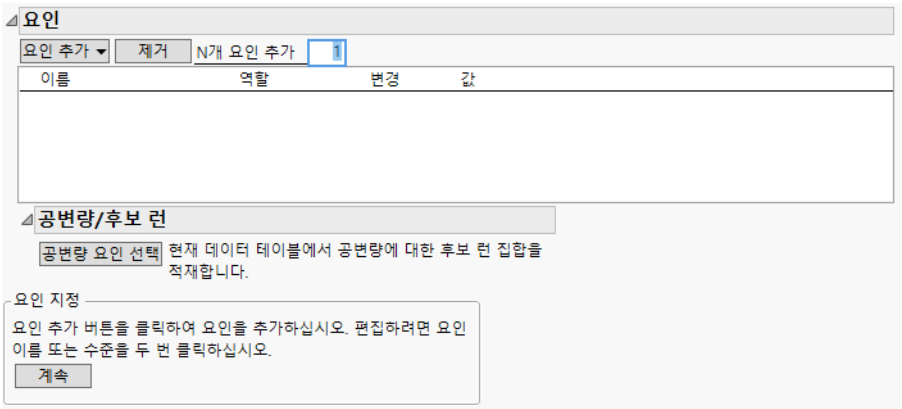
하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

요인

사용자 설계의 "요인" 개요에서 연구에 요인을 추가합니다.

팁 : "요인" 개요를 완료하면 빨간색 삼각형 메뉴에서 **요인 저장**을 선택하는 것이 좋습니다. 이렇게 하면 요인 이름, 역할, 변경 및 값을 데이터 테이블에 저장하여 나중에 다시 불러올 수 있습니다.

그림 4.18 요인 개요



요인 추가 요인 유형을 선택합니다. 자세한 내용은 "요인 유형"(85 페이지)에서 확인하십시오.

제거 선택한 요인을 제거합니다.

참고 : "계속" 또는 "뒤로" 버튼을 클릭한 후 모든 요인을 제거하려고 하면 연속형 요인 하나가 남습니다. 새 요인을 추가한 후 이 요인을 삭제할 수 있습니다.

N 개 요인 추가 여러 요인을 추가합니다. 추가할 요인 수를 입력하고 **요인 추가**를 클릭한 후 요인 유형을 선택합니다. 유형이 다른 여러 요인을 추가하려면 **N 개 요인 추가**를 반복합니다.

요인 개요

" 요인 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할을 지정합니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인 유형이 적절하게 모델링되도록 합니다 .

변경 요인 수준의 변경 정도를 " 쉬움 ", " 어려움 " 또는 " 매우 어려움 " 으로 나타냅니다 . 기본값 " 쉬움 " 을 클릭하여 변경할 수 있습니다 . 요인 변경을 " 어려움 " 또는 " 매우 어려움 " 으로 지정하면 설계에서 랜덤화할 때 이 제한이 반영됩니다 . " 요인 " 목록에 " 어려움 " 으로 지정된 요인이 없으면 요인을 " 매우 어려움 " 으로 지정할 수 없습니다 . " 요인 변경 " 열 특성은 데이터 테이블에 저장됩니다 . 자세한 내용은 " 변경 및 랜덤 블록 "(87 페이지) 에서 확인하십시오 .

값 요인에 대한 실험 설정입니다. 값을 삽입하려면 기본값을 클릭한 후 원하는 값을 입력합니다.

요인 개요 편집

" 요인 " 개요에서 다음 사항에 유의하십시오 .

- 요인 이름을 편집하려면 해당 요인 이름을 두 번 클릭합니다 .
- 범주형 요인은 요인 이름 왼쪽에 아래쪽 화살표가 있습니다 . 이 화살표를 클릭하여 수준을 추가합니다 .
- 요인 수준을 제거하려면 값을 클릭하고 **삭제**를 클릭한 후 텍스트 상자 바깥쪽을 클릭합니다 .
- " 변경 " 아래의 항목을 수정하려면 **변경** 열의 값을 클릭하고 적절한 항목을 선택합니다 .
- 값을 편집하려면 **값** 열의 값을 클릭합니다 .

요인 유형

요인 유형을 선택하려면 "사용자 설계"에서 **요인 추가**를 클릭합니다.

참고 : 각 요인의 역할을 포함하는 " 설계 역할 " 열 특성이 생성된 설계 테이블의 해당 요인 열에 추가됩니다 . " 설계 역할 " 열 특성은 요인이 올바르게 모델링되도록 합니다 .

연속형 숫자 데이터 유형만 가능합니다 . 연속형 요인은 개념적으로 공정 및 측정 시스템의 제한을 고려하여 사용자가 제공하는 하한과 상한 사이의 어떤 값이든 설정할 수 있는 요인입니다 .

이산 수치형 숫자 데이터 유형만 가능합니다 . 이산 수치형 요인은 이산형 숫자 값만 가질 수 있습니다 . 이러한 값에는 묵시적 순서가 있습니다 .

수준 수가 k 개 ($k > 2$) 인 이산 수치형 요인의 기본값은 정수 1, 2, ..., k 입니다 . $k = 2$ 수준이 있는 이산 수치형 요인의 기본값은 -1 과 1 입니다. 기본값을 실험에 사용할 설정으로 바꿉니다.

참고 : 이산 수치형 요인의 일부 수준만 설계에 나타납니다 . 나타나는 수준은 " 모형 " 개요에서 사용자가 지정한 내용에 따라 결정됩니다 . 설계에 모든 수준을 나타내야 하는 경우 선별 설계 플랫폼을 사용하는 것이 좋습니다 .

가정된 모형에서 *k* 개의 수준이 있는 이산 수치형 요인에 대한 효과는 해당 효과에 다항식 항을 포함합니다 (최대 *k-1* 차). *k* 가 6 보다 큰 경우 최대 5 제곱 수준이 포함됩니다 . 다항식 효과 (2 제곱 이상) 의 추정 가능성은 If Possible 로 설정됩니다 . 이렇게 하면 알고리즘이 런 크기에 따라 배수 수준을 사용할 수 있습니다 . 다항식 항이 포함되지 않으면 주효과만 추정하는 설계가 생성됩니다 . 가정된 모형에서 이산 수치형 요인을 처리하는 방법에 대한 자세한 내용은 " 모형 "(93 페이지) 에서 확인하십시오 .

" 모형 적합 " 에서는 이산 수치형 요인을 연속형 예측 변수로 처리합니다 . 설계 테이블에 저장된 " 모형 " 스크립트는 2 보다 큰 차수의 다항식 항을 포함하지 않습니다 .

범주형 숫자 또는 문자 데이터 유형에 해당됩니다. 결과 데이터 테이블의 데이터 유형은 범주형입니다 . 수준의 값 순서는 왼쪽에서 오른쪽으로 입력되는 값의 순서입니다 . 이 순서는 설계 데이터 테이블이 생성된 후 " 값 순서 " 열 특성에 저장됩니다 .

블록화 숫자 또는 문자 데이터 유형에 해당됩니다 . 블록 요인은 모형을 주효과로만 입력할 수 있는 범주형 요인의 특수 유형입니다 . 블록 요인을 정의할 때 블록당 런 수를 지정합니다 . " 블록당 런 수 " 열 특성은 설계 테이블에 저장됩니다 . 기본 런 크기는 항상 블록이 2 개 이상 있다고 가정합니다 . 블록당 런 수의 정수 배수가 아닌 런 크기를 지정하면 JMP 에서 가능한 최대 범위로 설계 균형을 맞추려고 합니다 . 설계 균형을 맞출 때 JMP 는 블록당 런 수가 2 회 이상이 되도록 합니다 .

공변량 숫자 또는 문자 데이터 유형에 해당됩니다 . 공변량 요인의 값은 실험 단위에 대한 측정값으로 실험 전에 공지됩니다 . 공변량 값은 최적 기준과 관련하여 결과 설계의 최적성을 보장하기 위해 선택됩니다 . 자세한 내용은 " 변경 및 랜덤 블록 "(87 페이지) 및 " 변경하기 힘든 수준의 공변량 "(117 페이지) 에서 확인하십시오 .

JMP 는 사용 가능한 실험 단위에 대해 측정된 공변량이 포함된 공개된 데이터 테이블에서 공변량 요인과 값을 가져옵니다 . 이 데이터 테이블을 현재 데이터 테이블로 지정합니다 . **요인 추가**를 사용하거나 " 공변량 / 후보 런 " 개요에서 **공변량 요인 선택**을 클릭하여 공변량을 추가합니다 . 공변량을 추가할 때는 현재 데이터 테이블의 열 목록이 열리면 이 목록에서 공변량이 포함된 열을 선택하면 됩니다 . 공변량 요인을 추가하면 후보 행이 포함된 테이블이 " 공변량 후보 런 " 개요에 로드됩니다 . 테이블이 로드되면 행을 선택하여 설계에 강제 적용할 수 있습니다 . 선택한 공변량 행을 설계에 강제 적용하려면 " 설계 생성 " 옵션인 **선택한 모든 공변량 행을 설계에 포함**을 사용합니다 .

설계에 포함된 런 수가 공변량 행 수보다 적으면 설계 테이블에 **공변량 행 인덱스** 열이 포함됩니다 . 이 열은 공변량 테이블에서 각 실험 런에 해당하는 행을 나타냅니다 .

경우에 따라 더 큰 후보 설정 집합에서 소수의 설계점 집합을 선택할 수도 있습니다 . 예를 들어 하나의 큰 단위에 대해 측정 열 (요인) 이 여러 개 있을 수 있습니다 . 각 단위에 대한 측정값을 후보 런으로 처리하려고 합니다 . 이러한 후보 런에서 반응을 측정하는 데 사용할 작지만 최적의 모음을 선택할 수 있습니다 . 이 경우 모든 후보 런이 포함된 데이터 테이블을 활성 테이블로 지정하고 " 요인 추가 " > " 공변량 " 을 선택한 다음 , 모든 측정 열을 공변량으로 입

력합니다 . 그리고 원하는 런 크기를 지정합니다 . 사용자 설계 플랫폼에서 최적의 설계 설정 모음을 식별합니다 .

참고 : 공변량이 있을 경우 " 설계 생성 " 개요에 두 개의 옵션이 포함됩니다 . 첫 번째 옵션을 사용하면 공변량 테이블에서 선택한 행을 설계에 포함시킬 수 있습니다 . 두 번째 옵션을 사용하면 설계에서 공변량 행이 반복되도록 할 수 있습니다 .

혼합 혼합 내 성분을 나타내는 연속형 요인입니다 . 혼합 요인 값의 합은 상수여야 합니다 . 기본적으로 모든 혼합 요인 값의 합은 1 입니다 . 혼합 성분의 합을 다른 양수 값으로 설정하려면 빨간색 삼각형 메뉴에서 **고급 옵션 > 혼합물 합**을 선택합니다 . " 혼합 " 열 특성은 데이터 테이블에 저장됩니다 .

상수 숫자 또는 문자 데이터 유형에 해당됩니다 . 상수 요인은 실험 중에 값이 고정되는 요인입니다 . 상수 요인은 " 모형 " 개요 또는 데이터 테이블에 저장된 " 모형 " 스크립트에 포함되지 않습니다 .

제어되지 않음 숫자 또는 문자 데이터 유형에 해당됩니다 . 제어되지 않은 요인은 생성 중에 값을 제어할 수 없지만 모형에 포함하려는 요인입니다 . 각 실험 런에 대해 요인의 값을 기록할 수 있다고 가정합니다 .

연속형 모델링 유형이 포함된 빈 열이 설계 테이블에 생성됩니다 . 필요한 경우 " 열 정보 " 창에서 열의 " 데이터 유형 " 과 " 모델링 유형 " 을 변경할 수 있습니다 . 이 열에 데이터를 입력합니다 . 제어되지 않은 요인은 " 모형 " 개요 및 데이터 테이블에 저장된 " 모형 " 스크립트에 포함됩니다 .

변경 및 랜덤 블록

런 간 요인 변경의 상대적 난이도를 지정하는 기능은 산업 실험에 유용합니다 . 변경하기 힘든 요인을 일부 설정에서 고정시키고 런을 여러 번 수행하면 편리합니다 . "변경" 값을 "어려움"으로 설정하면 분할구 설계가 생성됩니다 . "변경" 값을 "매우 어려움"으로 설정하면 분할-분할구 설계 또는 이원 분할구 설계가 생성됩니다 .

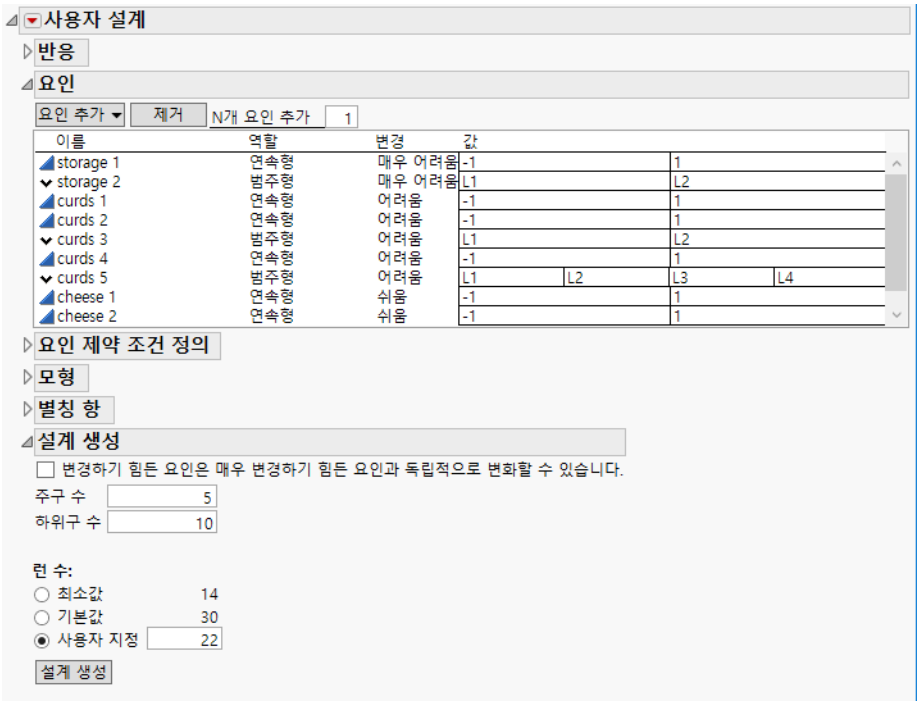
연속형 , 이산 수치형 , 범주형 및 혼합 요인의 "변경"을 "어려움" 또는 "매우 어려움"으로 설정할 수 있습니다 . 요인을 "매우 어려움"으로 설정하려면 목록에 "어려움"으로 설정된 다른 요인이 포함되어 있어야 합니다 .

공변량 요인의 "변경"을 "어려움"으로 설정할 수 있습니다 . 이 경우 다른 모든 공변량도 "어려움"으로 설정되고 나머지 요인은 "쉬움"으로 설정됩니다 . 이 알고리즘을 사용하려면 행 교환과 좌표 교환의 조합이 필요합니다 . 따라서 적당한 크기의 설계도 생성하는 데 시간이 걸릴 수 있습니다 .

변경하기 힘든 요인 또는 매우 변경하기 힘든 요인이 포함된 설계의 경우 , 사용자 설계는 지정된 최적 기준을 고려하여 최적의 설계를 찾습니다 . 자세한 내용은 "최적 기준"(118 페이지)에서 확인하십시오 . 분할구 설계를 생성하는 데 사용되는 방법론에 대한 자세한 내용은 Jones and Goos(2007) 에서 확인하십시오 . 변경하기 힘든 공변량이 포함된 설계에 대한 자세한 내용은 Jones and Goos(2015) 에서 확인하십시오 .

그림 4.19에서는 Design Experiment 폴더에 있는 Cheese Factors.jmp 샘플 데이터 테이블의 요인을 사용한 분할-분할구 시나리오를 보여 줍니다.

그림 4.19 분할 - 분할구 설계를 위한 요인 및 설계 생성 개요



하나 이상의 요인에 대해 " 변경 " 을 " 어려움 " 으로 할당하지만 " 변경 " 이 " 매우 어려움 " 으로 할당된 요인이 없는 경우 **주구**라는 범주형 요인이 설계에 추가됩니다. 이 경우 다음과 같은 분할구 설계가 생성됩니다.

- **주구**의 각 수준은 변경하기 힘든 요인의 상수 설정 블록에 해당합니다.
- 설계 테이블의 " 모형 " 스크립트는 **주구** 요인에 " 임의 효과 " 속성을 적용합니다.
- " 랜덤 블록 " 값이 지정된 " 설계 역할 " 열 특성이 **주구** 요인에 할당됩니다.

" 변경 " 을 " 어려움 " 과 " 매우 어려움 " 둘 다로 지정하면 **하위구**와 **주구**라는 범주형 요인이 설계에 추가됩니다. 이 경우 다음과 같은 분할 - 분할구 설계가 생성됩니다.

- **하위구**의 각 수준은 변경하기 힘든 요인의 상수 설정 블록에 해당합니다.
- **주구**의 각 수준은 매우 변경하기 힘든 요인의 상수 설정 블록에 해당합니다.
- 설계 테이블의 " 모형 " 스크립트는 **주구** 및 **하위구** 효과에 " 임의 효과 " 속성을 적용합니다.
- 변경하기 힘든 요인의 수준은 기본적으로 매우 변경하기 힘든 요인의 수준 내에 내포되어 있다고 가정합니다.

- " 랜덤 블록 " 값이 지정된 " 설계 역할 " 열 특성이 설계 테이블의 주구 요인과 하위구 요인 둘 다에 할당됩니다 .

이원 분할구 설계를 구성하려면 " 설계 생성 " 아래의 **변경하기 힘든 요인은 매우 변경하기 힘든 요인과 독립적으로 변화할 수 있습니다.** 옵션을 선택합니다. 이 옵션은 변경하기 힘든 요인의 수준을 매우 변경하기 힘든 요인의 수준과 교차시킵니다. 자세한 내용은 " 이원 분할구 설계 "(114 페이지) 에서 확인하십시오.

"주구 수" 및 "하위구 수" 텍스트 상자를 사용하여 주구 또는 하위구 수의 값을 지정합니다. 이 텍스트 상자는 제안된 주구 및 하위구 수로 초기화됩니다. 이러한 값을 얻는 방법에 대한 자세한 내용은 " 주구 및 하위구 수 "(117 페이지) 에서 확인하십시오.

랜덤 블록 분할구, 분할-분할구 및 이원 분할구 설계를 설명하는 자세한 내용과 시나리오는 " 랜덤화 제한이 있는 설계 "(109 페이지) 에서 확인하십시오. 변경하기 힘든 공변량이 포함된 설계에 대한 자세한 내용은 " 변경하기 힘든 수준의 공변량 "(117 페이지) 에서 확인하십시오.

요인 열 특성

각 요인에 대한 다양한 열 특성이 데이터 테이블에 저장됩니다 . 이러한 열 특성 및 관련 예에 대한 자세한 내용은 부록 A, " 열 특성 " 에서 확인할 수 있습니다 .

설계 역할 각 요인에 " 설계 역할 " 열 특성이 지정됩니다 . 요인을 정의할 때 지정하는 역할에 따라 " 설계 역할 " 열 특성의 값이 결정됩니다 . " 설계 생성 " 에서 랜덤 블록을 추가하면 해당 요인에 " 랜덤 블록 " 값이 할당됩니다 . " 설계 역할 " 특성은 실험 데이터를 모델링할 때 요인이 어떤 용도로 사용되는지를 반영합니다 . " 설계 역할 " 값은 설계 확대 플랫폼에서 사용됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 설계 역할 "(737 페이지) 에서 확인하십시오 .

요인 변경 각 요인에 " 요인 변경 " 열 특성이 할당됩니다 . " 변경 " 에 지정하는 값에 따라 " 요인 변경 " 열 특성의 값이 결정됩니다 . " 요인 변경 " 특성은 실험 데이터를 모델링할 때 요인이 사용되는 방식을 반영합니다 . " 요인 변경 " 값은 설계 확대 및 설계 평가 플랫폼에서 사용됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 요인 변경 "(752 페이지) 에서 확인하십시오 .

코딩 " 역할 " 이 " 연속형 ", " 이산 수치형 ", 연속형 " 공변량 " 또는 " 제어되지 않음 " 인 경우 요인의 " 코딩 " 열 특성이 저장됩니다 . 이 특성은 하한값과 상한값이 각각 –1 과 +1 에 해당하도록 요인 값을 변환합니다 . 자세한 내용은 " 열 특성 " 부록의 " 코딩 "(739 페이지) 에서 확인하십시오 .

값 순서 " 역할 " 이 " 범주형 " 또는 " 블록화 " 인 경우 요인의 " 값 순서 " 열 특성이 저장됩니다 . 이 특성은 요인 수준이 나타나는 순서를 결정합니다 . 자세한 내용은 " 열 특성 " 부록의 " 값 순서 "(755 페이지) 에서 확인하십시오 .

혼합 " 역할 " 이 " 혼합 " 인 경우 요인의 " 혼합 " 열 특성이 저장됩니다 . 이 특성은 요인 및 혼합물 합의 한계를 나타냅니다 . 또한 혼합 요인에 대한 코딩을 선택할 수도 있습니다 . 자세한 내용은 " 열 특성 " 부록의 " 혼합 "(747 페이지) 에서 확인하십시오 .

블록당 런 수 블록 요인에 대해 각 블록에서 허용되는 최대 런 수를 나타냅니다 . " 요인 " 개요에 " 블록화 " 요인이 지정되면 해당 요인에 대해 " 블록당 런 수 " 열 특성이 저장됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 블록당 런 수 "(761 페이지) 에서 확인하십시오 .

요인 제약 조건 정의

" 요인 제약 조건 정의 " 를 사용하여 설계 공간을 제한할 수 있습니다 . 제약 조건을 불러오거나 스크립트의 일부로 포함하지 않은 경우 **없음** 옵션이 선택됩니다 . 제약 조건을 지정하려면 다른 옵션 중 하나를 선택합니다 .

선형 제약 조건 지정 요인의 선형 조합에 대한 부등식 제약 조건을 지정합니다 . " 연속형 " 또는 " 혼합 " 역할이 설정된 요인에만 사용할 수 있습니다 . 자세한 내용은 선형 제약 조건 지정에서 확인하십시오 .

참고 : 선형 제약 조건을 포함하는 설계의 스크립트를 저장하면 해당 스크립트에 선형 제약 조건이 보다 작거나 같음 부등식 (≤) 으로 표현됩니다 .

허용되지 않는 조합 필터 사용 개별 요인의 값 제한을 기반으로 제약 조건 집합을 정의합니다 . AND 제약 조건과 OR 제약 조건을 모두 정의할 수 있습니다 . 자세한 내용은 허용되지 않는 조합 필터 사용에서 확인하십시오 .

허용되지 않는 조합 스크립트 사용 허용되지 않는 조합 및 기타 제약 조건을 스크립트 편집기 상자에 부울 JSL 표현식으로 정의합니다 . 자세한 내용은 허용되지 않는 조합 스크립트 사용에서 확인하십시오 .

참고 : 요인 제약 조건을 포함하는 설계를 분석할 때는 모형 프로파일러가 제약 조건을 적용합니다 .

선형 제약 조건 지정

설계 공간에서 연속형 요인을 독립적으로 변경할 수 없는 경우 선형 부등식 제약 조건을 지정할 수 있습니다. 선형 부등식은 허용되는 요인 수준 설정을 정의합니다.

추가를 클릭하여 하나 이상의 선형 부등식 제약 조건을 입력합니다 .

추가 설계의 모든 연속형 요인을 포함하는 선형 표현식 템플릿을 추가합니다 . 요인에 대한 계수 값을 입력한 후 선형 제약 조건을 반영하는 부등식 방향을 선택합니다 . 부등식의 오른쪽 상자에 제약 값을 지정합니다 . 제약 조건을 더 추가하려면 **추가**를 다시 클릭합니다 .

참고 : " 구 반지름 " 을 지정하여 설계 영역을 이미 제한한 경우에는 " 추가 " 옵션이 비활성화됩니다 .

마지막 제약 조건 제거 마지막 제약 조건을 제거합니다 .

제약 조건 검사 제약 조건의 일관성을 확인합니다 . 이 옵션은 중복 제약 조건을 제거하고 실행 가능성 검사를 수행합니다 . 문제가 있으면 JMP 경고가 나타납니다 . 제약 조건이 요인의 한계와 동등한 경우 " 요인 " 개요의 한계가 업데이트되었음을 알리는 JMP 경고가 나타납니다 .

허용되지 않는 조합 필터 사용

이 옵션은 허용되지 않는 조합을 쉽게 지정하기 위해 데이터 필터를 조정하여 사용합니다. 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

"필터 요인 추가" 목록에서 요인을 선택하고 **추가**를 클릭합니다. 그런 다음 슬라이더를 사용하거나(연속형 요인의 경우) 수준을 선택하여(범주형 요인의 경우) 허용되지 않는 조합을 지정합니다.

"필터 요인 추가" 메뉴의 빨간색 삼각형 옵션은 많은 플랫폼 시작 창의 "열 선택" 패널에 있는 옵션과 동일합니다. 자세한 내용은 JMP 사용의 "시작하기" 장에서 확인하십시오.

" 추가 " 를 클릭하면 " 허용되지 않는 조합 " 제어판에 선택한 요인이 표시되고 추가 제어 옵션이 제공됩니다 . 요인은 모델링 유형에 따라 다음과 같이 나타납니다 .

연속형 요인 연속형 요인의 경우 요인 설정 범위를 따라 양단 슬라이더가 나타납니다 . 슬라이더 끝을 드래그하거나 슬라이더 아래의 텍스트 값을 클릭하여 끝점을 설정함으로써 허용되지 않은 설정을 지정할 수 있습니다 . 슬라이더에서 파란색 강조 표시는 허용되지 않는 값을 나타냅니다 .

범주형 요인 범주형 요인의 경우 가능한 수준은 라벨이 지정된 블록으로 표시되거나 , 수준 수가 많으면 목록 항목으로 표시됩니다 . 허용하지 않을 수준을 선택합니다 . 여러 수준을 선택하려면 Ctrl 키를 누른 채로 선택합니다 . 해당 블록 또는 목록 항목은 강조 표시되어 허용되지 않는 수준임을 나타냅니다 . " 허용되지 않는 조합 " 패널에 범주형 요인을 추가하면 해당 요인의 수준 수가 요인 이름 뒤의 괄호 안에 표시됩니다 .

허용되지 않는 조합 옵션

제어판에는 다음과 같은 컨트롤이 있습니다 .

지우기 지정한 허용되지 않는 요인 수준 설정을 모두 지웁니다 . 선택한 요인은 지워지지 않습니다 .

다시 시작 선택한 모든 요인을 제거하고 초기 요인 목록으로 돌아갑니다 .

AND " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 AND 그룹이 됩니다 . AND 그룹 내에 지정된 요인 수준 조합은 허용되지 않습니다 .

나중에 AND 그룹에 요인을 추가하려면 그룹 개요를 클릭하여 강조 표시된 직사각형을 표시합니다 . 그런 다음 "AND" 를 선택하고 요인을 추가합니다 .

단일 요인을 제거하려면 빨간색 삼각형 메뉴에서 **삭제**를 선택합니다 .

OR " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 별도의 AND 그룹이 됩니다 . OR 로 구분된 AND 그룹의 경우 하나 이상의 AND 그룹에 지정된 조합은 허용되지 않습니다 .

요인에 대한 빨간색 삼각형 메뉴 옵션

한 요인이 여러 OR 그룹에 나타날 수 있습니다. 특정 OR 그룹에 요인이 나타나는 경우 이를 해당 요인의 인스턴스라고 합니다.

삭제 " 허용되지 않는 조합 " 패널에서 선택한 요인 인스턴스를 제거합니다 .

선택 해제 요인 인스턴스에 대한 선택을 취소합니다.

선택 반전 요인 인스턴스에 대해 선택된 값을 선택 취소하고 이전에 선택되지 않은 값을 선택합니다.

표시 옵션 범주형 요인에만 사용할 수 있습니다. 표시 모양을 변경합니다. 각 수준을 블록, 목록 또는 단일 범주로 표시하거나 각 값 옆에 체크박스를 추가하는 옵션을 사용할 수 있습니다.

찾기 범주형 요인에만 사용할 수 있습니다. 요인 이름 아래에 요인 수준에 대한 검색 문자열을 입력할 수 있는 텍스트 상자를 제공합니다. Enter 키를 누르거나 텍스트 상자 바깥쪽을 클릭하면 검색이 수행됩니다 . **찾기**가 선택되면 찾기 텍스트 상자 드롭다운 메뉴에 찾기 옵션이 나타납니다.

허용되지 않는 조합 스크립트 사용

이 옵션을 사용하면 JSL 스크립트를 사용하여 요인 수준 조합을 허용하지 않도록 설정할 수 있습니다. 이 옵션은 연속형 요인 또는 연속형과 범주형이 혼합된 요인과 함께 사용할 수 있습니다.

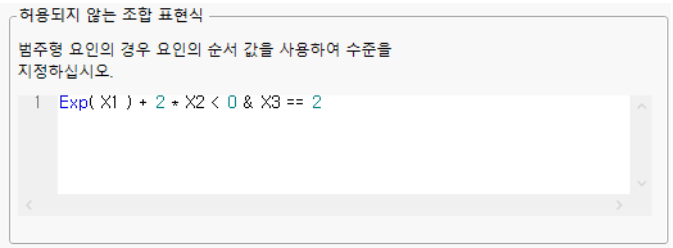
이 옵션은 허용하지 않을 조합을 식별하는 스크립트를 삽입할 수 있는 스크립트 창을 엽니다. 스크립트는 부울 표현식으로 평가되어야 합니다. 표현식이 true로 평가되면 지정된 조합이 허용되지 않습니다.

범주형 요인에 대한 표현식을 작성할 때는 수준의 순서 값 또는 수준 이름을 사용합니다 . " 요인 " 개요에서 요인 수준이 높음, 중간, 낮음으로 순서대로 지정된 경우 관련 서수 값은 1, 2, 3 입니다. 예를 들어 두 개의 연속형 요인 (X1, X2) 과 세 가지 수준 (L1, L2, L3) 이 순서대로 설정된 범주형 요인 하나가 있다고 가정해 보겠습니다. 다음 표현식이 true 인 경우 수준을 허용하지 않으려고 합니다.

$$e^{X_1} + 2X_2 < 0 \text{ 및 } X_3 = L2$$

`(Exp(X1) + 2*X2 < 0) & (X3 == 2)` 표현식을 스크립트 창에 입력합니다.

그림 4.20 스크립트 편집기의 표현식



(그림에서 불필요한 괄호는 파싱을 통해 제거됨) 함수를 부울 표현식의 일부로 입력할 수 있습니다. `(Exp(X1) + 2*X2 < 0) & (X3 == "L2")` 표현식도 유효합니다.

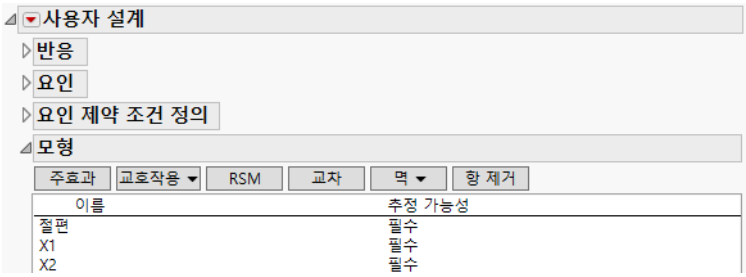
모형

"모형" 개요에서 가정된 모형, 즉 추정할 모든 효과를 포함하는 모형을 지정합니다. 지정하는 각 효과에 대해 추정 가능성을 지정할 수 있습니다. 추정 가능성 값은 해당 효과를 반드시 추정해야 하는지 여부(필수) 또는 가능한 경우 해당 효과를 추정할지 여부(If Possible)를 나타냅니다.

초기 "모형" 개요에는 모든 요인의 주효과가 포함됩니다. 수준이 3개 이상인 이산 수치형 요인을 입력한 경우 다항식 항도 초기 모형에 포함됩니다. 2차 이상 항의 추정 가능성은 If Possible로 설정됩니다. 이러한 항을 추정할 수 있게 하려면 추정 가능성을 **필수**로 변경합니다.

참고 : 이산 수치형 다항식 항의 추정 가능성이 항상 " 필수 " 로 설정되게 할 수 있습니다 . **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다 . 그런 다음 " 이산 수치형 요인의 멱 설정 필수 " 를 선택하십시오 .

그림 4.21 모형 개요



설계 테이블을 구성할 때 JMP에서 "모형" 스크립트를 데이터 테이블에 저장합니다. 이산 수치형 요인을 제외하고 "모형" 개요에 표시된 모든 효과가 "모형" 스크립트에 포함됩니다. 이산 수치형 요인의 경우 주효과와 2차 항만 "모형" 스크립트에 포함됩니다.

" 모형 " 개요에는 다음과 같은 버튼과 필드가 포함되어 있습니다 .

주효과 모형의 모든 요인에 대한 주효과와 이산 수치형 요인에 대한 다항식 항을 추가합니다 .

교호작용 교호작용 효과를 추가합니다 . " 요인 " 개요에서 요인을 선택하지 않은 경우 "2 차 ", "3 차 ", "4 차 " 또는 "5 차 " 를 선택하여 적절한 모든 교호작용을 해당 차수까지 추가합니다 . " 요인 " 개요에서 요인 이름을 선택하고 " 교호작용 " 을 선택한 후 적절한 차수를 지정하여 특정 요인에 대해 지정된 차수까지 교호작용을 추가합니다 . 비혼합 요인과 혼합 요인 간의 교호작용 및 블록 요인과 상수 요인 간의 교호작용은 추가되지 않습니다 .

RSM 모든 연속형 요인에 대해 교호작용 및 2 차 항을 최대 2 차까지 추가합니다 (반응 표면 모형 항). 범주형 요인은 RSM 항에 포함되지 않습니다 . 모든 혼합 요인과 상호 작용하는 비혼합 요인의 주효과는 제거됩니다 .

교차 특정 교호작용 항을 추가합니다 . " 요인 " 개요에서 요인 이름을 선택하고 " 모형 " 개요에서 효과 이름을 선택합니다 . " 교차 " 를 클릭하면 교차된 항이 " 모형 " 개요에 추가됩니다 .

멱 다항식 항을 추가합니다 . " 요인 " 개요에서 요인 이름을 선택하지 않으면 모든 연속형 요인에 대해 다항식 항을 추가합니다 . " 요인 " 개요에서 요인 이름을 선택하면 해당 요인에 대해

서만 다항식 항을 추가합니다 . "2 차 ", "3 차 ", "4 차 " 또는 "5 차 " 를 선택하면 해당 차수의 다항식 항이 추가됩니다 .

Scheffe 입방 (혼합 요인에 사용 가능) 모든 혼합 요인에 대해 Scheffe 입방 항을 추가합니다 . 이 항은 3 차 다항식 항을 포함하는 혼합 모형을 지정하는 데 사용됩니다 .

항 제거 선택한 효과를 제거합니다 .

이름 효과의 이름입니다 .

추정 가능성 효과에 대한 추정 필요성을 지정한 정보입니다.

- "추정 가능성"이 "필수"로 설정된 경우 알고리즘은 해당 효과를 추정할 수 있도록 합니다.
- " 추정 가능성 " 이 "If Possible" 로 설정된 경우 알고리즘은 지정된 런 수에 따라 해당 효과를 추정할 수 있도록 시도합니다 .

기본적으로 이산 수치형 요인의 다항식 항을 제외한 모든 효과가 " 필수 " 로 지정됩니다 . 효과의 추정 가능성 값을 클릭하여 변경할 수 있습니다 .

If Possible 효과의 베이지안 D- 최적 및 추정

베이지안 D-최적 설계 방법은 모든 **필수** 항에 대한 정밀한 추정값을 구하고 If Possible 항에 대한 일괄 감지 기능 및 일부 추정 가능성을 제공합니다. 자세한 내용은 "사용자 설계의 예" 장의 "반응 표면 실험"(144 페이지) 및 "베이지안 D-최적"(119 페이지)에서 확인하십시오. 주효과 모형 항이 If Possible 로 설정되면 설계 진단 계산에 베이지안 정보 행렬이 사용됩니다.

별칭 항

"별칭 항" 개요에서는 가정된 모형에 없지만 모형 항의 추정값을 편향시킬 수 있는 잠재적 활성 효과를 추가합니다. 가정된 모형에 포함되지 **않은** 효과가 활성화되었을 수 있습니다. 설계를 생성하면 "설계 평가" 아래에 "별칭 행렬" 개요가 나타납니다. 별칭 행렬 항목은 "별칭 항" 개요에서 지정한 효과에 의해 모형 모수에 전달되는 편향 정도를 나타냅니다. 자세한 내용은 "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지)에서 확인하십시오.

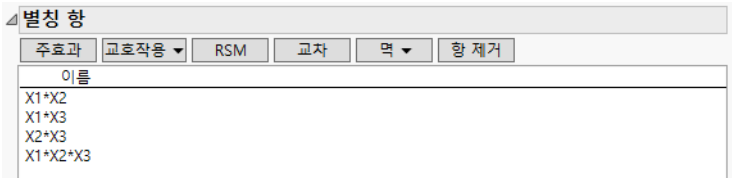
기본적으로 "별칭 항" 개요에는 "모형" 개요에 없는 모든 이원 교호작용 효과가 포함됩니다(블록 요인을 포함하는 항은 제외). 버튼을 사용하여 항을 추가합니다. 이러한 버튼을 사용하여 "별칭 항" 테이블에 효과를 추가하는 방법에 대한 설명은 "모형"(93 페이지)에서 확인하십시오.

참고 : 교호작용 항의 경우 선택한 차수의 교호작용만 추가됩니다 . 선택한 차수 및 모든 낮은 차수의 교호작용이 모형에 추가되는 " 모형 " 개요와 다릅니다 .

예를 들어 3개의 연속형 요인이 있는 설계를 지정한다고 가정해 보겠습니다. "모형" 개요에서 지정하는 가정된 모형에는 3개의 주효과만 포함됩니다. 이 경우 런을 6회만 수행할 수 있습니다. 주효과의 추정값이 활성 이원 교호작용 및 삼원 교호작용에 의해 어느 정도 편향될 수 있는지 확인하려고 합니다.

"별칭 항" 테이블에는 기본적으로 모든 이원 교호작용이 포함됩니다. **교호작용 > 3차**를 선택하여 삼원 교호작용을 추가할 수 있습니다.

그림 4.22 별칭 항 개요



"설계 생성" 개요에서 6회 런을 지정하고 **설계 생성**을 클릭하면 "설계 평가" 개요가 나타납니다. "설계 평가" 개요와 "별칭 행렬" 개요를 차례로 엽니다.

그림 4.23 별칭 행렬

효과	X1*X2	X1*X3	X2*X3	X1*X2*X3
절편	0	0	0	1
X1	0	0	1	0
X2	0	1	0	0
X3	1	0	0	0

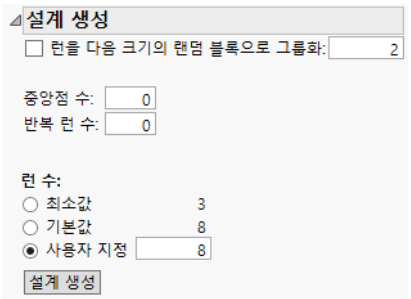
별칭 행렬은 각 주효과가 두 교호작용과 부분적으로 별칭 관계가 있음을 나타냅니다. 자세한 내용은 "설계 평가" 장의 "별칭 행렬"(430 페이지) 및 "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지)에서 확인하십시오.

설계 생성

"설계 생성" 개요에서는 설계의 블록화, 중앙점, 반복 및 런 수와 관련된 선택을 할 수 있습니다. 일반적으로 입력 영역은 다음과 같은 두 부분으로 되어 있습니다.

- 설계 구조 옵션
- 런 수 옵션

그림 4.24 설계 생성 개요



설계 구조 옵션

런을 다음 크기의 랜덤 블록으로 그룹화 (블록 요인이 지정된 경우 사용할 수 없음) 랜덤 블록 설계를 구성하려면 각 블록에 포함할 런 수를 입력합니다. 표본 크기를 지정하면 **랜덤 블록**이라는 요인이 생성됩니다. 요인 수준은 지정된 런 수를 고려하여 입력한 블록 크기와 일치하

는 크기의 블록을 정의합니다 . 런 수가 블록 크기의 정수 배수인 경우 블록 크기는 지정된 값과 동일합니다 .

주구 수 변경하기 힘든 요인 또는 매우 변경하기 힘든 요인을 지정할 때 나타납니다 . 주구 요인은 매우 변경하기 힘든 요인 (분할 - 분할구 설계) 아니면 변경하기 힘든 요인 (분할구 설계) 에 해당합니다 . JMP 에서는 모형의 계수에 대한 정보를 최대화하는 값을 주구 수로 제안합니다 . 주구 수의 값을 직접 입력할 수도 있습니다 . 자세한 내용은 " 주구 및 하위구 수 "(117 페이지) 에서 확인하십시오 .

하위구 수 매우 변경하기 힘든 요인을 지정할 때 나타납니다 . 하위구 요인은 분할 - 분할구 설계에서 매우 변경하기 힘든 요인에 해당합니다 . JMP 에서는 모형의 계수에 대한 정보를 최대화하는 값을 주구 및 하위구 수로 제안합니다 . 하위구 수의 값을 직접 입력할 수도 있습니다 . 자세한 내용은 " 주구 및 하위구 수 "(117 페이지) 에서 확인하십시오 .

변경하기 힘든 요인은 매우 변경하기 힘든 요인과 독립적으로 변화할 수 있습니다 . 조각구 (이원 분할구 또는 분할 블록이라고도 함) 설계를 생성하려면 이 옵션을 선택합니다 . 이 옵션은 변경하기 힘든 요인이 매우 변경하기 힘든 요인의 수준 내에서 랜덤화되는 설계를 생성합니다 . 이러한 요인은 매우 변경하기 힘든 요인 내에 내포되어 있지 않습니다 .

중앙점 수 설계에 연속형 또는 혼합 유형의 요인이 포함된 경우에만 나타납니다 . 설계에 중앙점으로 포함할 추가 런 수를 지정합니다 . 중앙점은 각 연속형 요인에 대한 설정이 상한 설정과 하한 설정의 중간에 있는 런입니다 . 자세한 내용은 "DOE 로 시작 " 장의 " 중앙점 , 반복 런 및 검정 "(61 페이지) 에서 확인하십시오 .

설계에 연속형 요인과 다른 유형의 요인이 모두 포함된 경우 중앙점이 다른 요인의 수준과 관련하여 균형이 맞지 않을 수 있습니다 . 사용자 설계는 설계의 D 효율도 , I 효율도 또는 별칭 효율도를 최대화하기 위해 중앙점을 선택합니다 .

반복 런 수 설계에 추가할 반복 시도 수를 지정합니다 . 전체 설계를 반복하는 것이 아니라 반복할 최적 설계점을 선택합니다 . 자세한 내용은 "DOE 로 시작 " 장의 " 중앙점 , 반복 런 및 검정 "(61 페이지) 에서 확인하십시오 .

런 수 옵션

최소값 설계 생성이 실패하지 않게 하기 위해 필요한 런 수의 하한입니다 . 최소값을 선택하면 결과 설계가 포화됩니다 . 따라서 오차 자유도가 없습니다 .

참고 : 런 수를 " 최소값 " 으로 선택하면 검정할 오차 항이 없게 됩니다 . 모수 추정값을 검정할 수 없습니다 . 이 선택은 추가 런 비용이 너무 높은 경우에만 적합합니다 .

기본값 런 수를 제안합니다 . 이 값은 런 수가 최소 런 수보다 4 회 이상 많은 균형 설계를 생성하기 위한 발견적 방법을 기반으로 합니다 .

사용자 지정 원하는 런 수를 지정합니다 . 런 수 텍스트 상자에 해당 값을 입력합니다 . 이 옵션을 사용하면 추가 런 비용을 잠재적 정보 획득으로 상쇄할 수 있습니다 .

런 수 " 변경 " 이 " 쉬움 " 으로 설정된 공변량 요인을 지정하면 이 옵션만 나타납니다 . 표시된 런 수는 데이터 테이블에서 하나 이상의 공변량과 관련된 행 수입니다 . 더 적은 수의 런을 지정할 수 있습니다 . 이 경우 선택된 공변량 런은 설계 기준을 최적화합니다 .

설계 생성

"설계 생성" 개요를 완료한 후 **설계 생성**을 클릭합니다. 사용자 설계에서 설계를 생성하여 "설계" 개요에 표시하고 "설계 평가" 개요에 평가 정보를 제공합니다. "모형" 및 "별칭" 개요의 컨트롤은 비활성화됩니다. "뒤로" 버튼을 사용하여 모형 또는 별칭 항을 변경할 수 있습니다. 설계 테이블을 생성할 수 있는 "출력 옵션" 패널도 나타납니다.

참고 : 여러 설계가 최적 기준을 최적화할 때도 있습니다 . 이 경우 **뒤로** 및 **설계 생성** 버튼을 반복해서 클릭하면 설계 알고리즘이 또 다른 설계를 생성할 수 있습니다 .

설계

"설계" 개요에는 사용자 설계의 런이 표시됩니다. 지정된 조건이 주어지면 설계는 최적이 됩니다. 런이 랜덤화되지 않을 수도 있습니다. 설계 테이블을 생성하기 전에 "출력 옵션" 패널에서 "런 순서" 옵션을 선택할 수 있습니다.

설계 평가

"설계 평가"의 빨간색 삼각형 메뉴 및 개요에는 생성된 사용자 설계의 특성을 평가하기 위한 다양한 방법이 제공됩니다.

베이지안 정보 사용 추정할 수 없는 If Possible 모형 효과를 가진 설계의 설계 진단 계산에는 베이지안 정보 행렬을 사용하도록 선택합니다 . 베이지안 정보 조정에 대한 자세한 내용은 " 베이지안 D- 최적 "(119 페이지) 에서 확인하십시오 .

참고 : 모든 모형 효과를 추정할 수 있는 경우에는 베이지안 정보 조정 없이 모든 효과에 대해 설계 진단이 제공됩니다 . 일부 항을 추정할 수 없는 경우 **베이지안 정보 사용** 옵션을 선택하지 않으면 **필수** 모형 항만 제공됩니다 .

검정력 분석 주어진 크기의 효과를 감지할 수 있는지 분석할 수 있습니다 .

예측 분산 프로파일 요인 설정 범위에 대한 예측 분산을 보여 줍니다 .

설계 공간 비율 그림 모형 예측 분산이 지정된 값보다 얼마나 아래 또는 위에 있는지 보여 줍니다 .

예측 분산 표면 두 개의 연속형 요인에 대한 예측 분산의 표면 그림을 보여 줍니다 .

추정 효율도 각 모수에 대해 이상적인 직교 설계 (존재하지 않을 수 있음) 와 비교하여 신뢰 구간 길이에 대한 증가 비율을 제공합니다 . 모수의 상대 표준 오차도 제공합니다 .

별칭 행렬 모형에 없지만 활성화 가능성이 있는 효과에 의해 모형 모수가 편향되는 정도를 나타내는 계수를 제공합니다 . " 별칭 항 " 테이블에서 잠재적 활성 효과를 나타내는 항을 지정합니다 . 자세한 내용은 " 기술 상세 정보 " 부록의 " 별칭 행렬 "(767 페이지) 에서 확인하십시오 .

상관 색상 맵 강도 척도를 사용하여 그림에서 효과 사이의 절대 상관을 표시합니다 .

> **참고 :** 기본 강도 척도는 회색 척도입니다 . 색상 척도를 변경하려면 " 상관 색상 맵 " 의 빨간색 삼각형을 클릭하고 **파랑 - 회색 - 빨강**을 선택합니다 . 사용자 정의 색상 척도를 사용하려면 그림을 마우스 오른쪽 버튼으로 클릭하고 **색상 테마**를 선택합니다 .

설계 진단 설계를 구성하는 데 사용되는 최적 기준을 나타냅니다 . 설계의 효율도 측도도 제공합니다 . 자세한 내용은 " 사용자 설계 옵션 "(100 페이지) 의 " 최적 기준 " 및 " 최적 기준 "(118 페이지) 에서 확인하십시오 .

> **참고:** "변경"이 "어려움" 또는 "매우 어려움"으로 설정되거나 "추정 가능성"이 "If Possible"로 설정된 요인이 모형에 포함된 경우 " 설계 진단 " 개요에 D 효율도 , G 효율도 및 A 효율도 통계량이 제공되지 않습니다 .

"설계 평가" 개요에 대한 자세한 내용은 "설계 평가" 장의 "설계 평가"(418 페이지)에서 확인하십시오 .

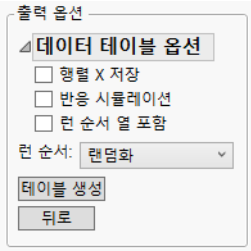
출력 옵션

사용자 설계의 행렬 X 를 저장하거나 , 반응을 시뮬레이션하거나 , 설계 데이터 테이블을 사용자 정의하려면 " 출력 옵션 " 패널을 사용합니다 . 다음과 같은 작업을 수행할 수 있습니다 .

- 사용자 설계 데이터 테이블을 표시하는 방법을 지정합니다 .
- 설계 테이블을 구성합니다 .
- " 사용자 설계 " 창의 이전 지점으로 돌아갑니다 .

대부분의 경우 "출력 옵션" 패널은 그림 4.25 와 같이 나타납니다 .

그림 4.25 출력 옵션 패널



" 출력 옵션 " 패널에는 다음 옵션이 포함되어 있습니다 .

- " 데이터 테이블 옵션 "(99 페이지)
- " 런 순서 "(99 페이지)
- " 테이블 생성 "(99 페이지)
- " 뒤로 "(100 페이지)

데이터 테이블 옵션

데이터 테이블 옵션에는 **행렬 X 저장**("행렬 X 저장"(105 페이지)), **반응 시뮬레이션**("반응 시뮬레이션"(103 페이지)) 및 **런 순서 열 포함** 옵션이 포함되어 있습니다.

런 순서 열 포함 "설계" 개요에서 설계 순서에 해당하는 열을 설계 데이터 테이블에 추가합니다.

팁 : 런 순서가 정렬된 데이터를 저장하는 경우 " 런 순서 열 포함 " 을 사용하십시오 . 그러면 설계 개요에서 런 순서가 랜덤화된 열이 제공됩니다 .

런 순서

런 순서 옵션은 설계 테이블의 런 순서를 결정합니다 . 다음과 같은 옵션을 선택할 수 있습니다 .

동일하게 유지 설계 테이블에 표시되는 행 순서가 " 설계 " 개요의 순서와 동일합니다 .

좌에서 우로 정렬 설계 테이블의 열이 왼쪽에서 오른쪽으로 정렬됩니다 .

랜덤화 설계 테이블의 행이 랜덤 순서로 나타납니다 .

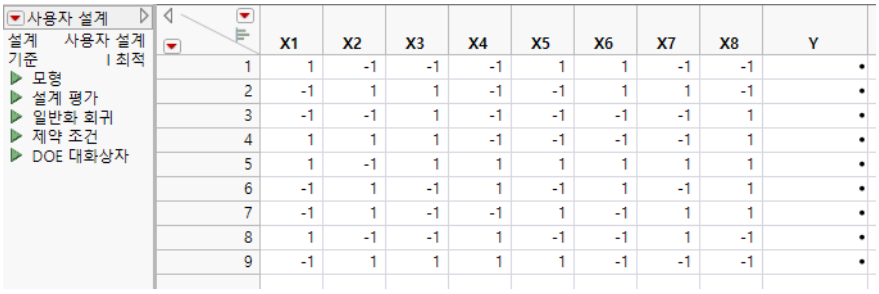
우에서 좌로 정렬 설계 테이블의 열이 오른쪽에서 왼쪽으로 정렬됩니다 .

블록 내 랜덤화 설계 테이블의 행이 블록 내에서 랜덤 순서로 나타납니다 .

테이블 생성

테이블 생성을 클릭하여 사용자 설계 데이터 테이블을 구성합니다. "사용자 설계" 테이블의 왼쪽 위에 있는 테이블 패널은 지정된 설계에 따라 필요한 스크립트를 포함할 수 있습니다. "모형", "설계 평가" 및 "DOE 대화상자" 스크립트는 항상 제공됩니다. 스크립트를 실행하려면 스크립트 이름 옆의 녹색 삼각형을 클릭합니다. 설계에 공변량이 포함된 경우 일부 공변량 행만 설계에 사용되면 설계 테이블에 공변량 행 인덱스 열이 포함됩니다.

그림 4.26 스크립트를 보여 주는 사용자 설계 테이블



	X1	X2	X3	X4	X5	X6	X7	X8	Y
1	1	-1	-1	-1	1	1	-1	-1	•
2	-1	1	1	-1	-1	1	1	-1	•
3	-1	-1	1	-1	-1	-1	-1	1	•
4	1	1	1	-1	-1	-1	-1	1	•
5	1	-1	1	1	1	1	1	1	•
6	-1	1	-1	1	-1	1	-1	1	•
7	-1	1	-1	-1	1	-1	1	1	•
8	1	-1	-1	1	-1	-1	1	-1	•
9	-1	1	1	1	1	-1	-1	-1	•

가능한 스크립트는 다음과 같습니다 .

모형 **분석 > 모형 적합** 플랫폼을 실행합니다 . 스크립트에서 설명하는 모형은 " 모형 " 개요에서 선택한 설정과 설계 유형에 따라 결정됩니다 .

설계 평가 **DOE > 설계 진단 > 설계 평가** 플랫폼을 실행합니다 . 스크립트에서 설명하는 모형은 " 모형 " 개요에서 선택한 설정과 설계 유형에 따라 결정됩니다 .

제약 조건 " 요인 제약 조건 정의 " 개요에서 **선형 제약 조건 지정** 옵션을 사용하여 입력한 모형 제약 조건을 보여 줍니다 .

허용되지 않는 조합 " 요인 제약 조건 정의 " 개요에서 **허용되지 않는 조합 필터 사용** 또는 **허용되지 않는 조합 스크립트 사용** 옵션을 사용하여 입력한 모형 제약 조건을 보여 줍니다 .

DOE 대화상자 설계 테이블을 생성하는 데 사용된 " 사용자 설계 " 창을 다시 생성합니다 . 또한 이 스크립트는 설계를 생성하는 데 사용된 난수 시드값도 포함합니다 .

뒤로

"뒤로" 버튼을 클릭하면 "설계 생성"을 클릭하기 전의 위치로 이동합니다. 이전 개요를 변경하고 설계를 다시 생성할 수 있습니다.

참고 : " 뒤로 " 버튼을 클릭한 후 모든 요인을 제거하려고 하면 연속형 요인 하나가 남습니다 . 새 요인을 추가한 후 이 연속형 요인을 삭제할 수 있습니다 .

사용자 설계 옵션

" 사용자 설계 " 의 빨간색 삼각형 메뉴에는 설계 설정 및 생성에 대한 옵션이 들어 있습니다 .

반응 저장 " 반응 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 . 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다 .

반응 불러오기 " 반응 저장 " 옵션을 사용하여 저장한 반응을 불러옵니다 .

요인 저장 " 요인 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 각 요인의 열에는 수준이 포함됩니다 . 다른 정보는 열 특성으로 저장됩니다 . 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다. 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다 . 그런 다음 " 설계 역할 " 영역에서 적절한 역할을 선택합니다 .

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 불러옵니다 .

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) " 요인 제약 조건 정의 " 또는 " 선형 제약 조건 " 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다 . 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다 . 마지막 행은 부등식 한계를 포함합니다 . 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약

조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 제약 조건 상태 "(762 페이지) 에서 확인하십시오 .

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) " 제약 조건 저장 " 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다 .

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다 . 이러한 작업에는 다음 중 하나 이상이 포함됩니다 .

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다 .

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다 .

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다 .
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다 .

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다 .

- **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다 . 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 **<Y> 시뮬레이션**이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 **<Y> 시뮬레이션**의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 **기본 분석**의 " 시뮬레이션 " 장에서 확인하십시오 .

행렬 X 저장 " 적률 행렬 " 및 " 모형 행렬 " 을 설계 데이터 테이블의 테이블 스크립트에 저장합니다 . 이 스크립트에는 적률 및 설계 행렬이 포함됩니다 . 자세한 내용은 " 행렬 X 저장 "(105 페이지) 에서 확인하십시오 .

주의 : 명목형 요인이 있는 설계의 경우 " 행렬 X 저장 " 옵션으로 저장한 모형 행렬은 선형 모형을 적합시킬 때 사용된 코딩 행렬이 아닙니다 . " 모형 " 스크립트를 실행할 때 나타나는 " 모형 적합 " 보고서에서 " 열 저장 > 코드화 테이블 저장 " 옵션을 선택하면 모형 적합에 사용되는 코딩 행렬을 가져올 수 있습니다 .

최적 기준 설계 최적 기준을 변경합니다 . 기본 기준인 **권장 사항**은 " 모형 " 개요에서 "RSM" 버튼을 사용하여 2 차 효과를 추가한 경우를 제외하고 모든 설계 유형에 대해 D- 최적을 지정합니다 . D- 최적 , I- 최적 및 별칭 - 최적 설계에 대한 자세한 내용은 " 최적 기준 "(118 페이지) 에서 확인하십시오 .

참고 : 항상 정해진 최적 기준을 사용하도록 환경 설정을 지정할 수 있습니다 . 파일 > 환경 설정 > 플랫폼 > DOE 를 선택합니다 . " 최적 기준 " 을 선택한 후 기본 기준을 선택합니다 .

시작 수 설계 구성에 사용되는 랜덤 시작 횟수를 지정할 수 있습니다. 자세한 내용은 "시작 수"(107 페이지) 에서 확인하십시오 .

설계 검색 시간 설계 검색을 위한 최대 시간 (초) 입니다 . 기본 검색 시간은 설계 복잡성을 기반으로 합니다 . 자세한 내용은 " 설계 검색 시간 "(108 페이지) 및 " 시작 수 "(107 페이지) 에서 확인하십시오 .

알고리즘 반복에 몇 초 이상이 필요한 경우 " 설계 계산 중 " 진행률 창이 나타납니다 . 진행률 창에서 **취소**를 클릭하면 계산이 중지되고 해당 시점에서 찾은 최적 설계가 제공됩니다 . 또한 " 변경 " 이 " 어려움 " 또는 " 매우 어려움 " 으로 설정되거나 " 추정 가능성 " 이 "If Possible" 로 설정된 요인이 포함되지 않은 D- 최적 설계에 대한 D 효율도가 진행률 창에 표시됩니다 .

참고 : " 설계 검색 시간 " 에 대한 환경 설정을 지정할 수 있습니다 . 파일 > 환경 설정 > 플랫폼 > DOE 를 선택합니다 . " 설계 검색 시간 " 을 선택한 후 최대 시간 (초) 을 입력합니다 . 시간이 더 필요한 특정 상황에서는 JMP 가 검색 시간을 연장합니다 .

구 반지름 설계의 연속형 요인을 초구로 제한합니다 . 반지름을 지정하고 **확인**을 클릭합니다 . 0 과의 간격이 구 반지름과 같도록 설계점이 선택됩니다 . " 설계 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

참고 : 구 반지름 제약 조건은 " 선형 제약 조건 지정 " 옵션을 사용하여 추가한 제약 조건과 결합할 수 없습니다 . 또한 변경하기 힘든 요인 (분할구 설계) 이 포함된 경우에는 이 옵션을 사용할 수 없습니다 .

고급 옵션 > 혼합물 합 혼합 요인의 합을 양수 값으로 설정합니다 . 전체 실험에서 혼합 성분을 일정하게 유지하려면 이 옵션을 사용합니다 .

고급 옵션 > 분할구 분산비 오차 분산에 대한 랜덤 주구 분산 및 하위구 분산 (있는 경우) 의 비를 지정합니다 . 이 값을 설정하려면 먼저 분할구 설계를 위한 변경하기 힘든 요인 또는 분할 - 분할구 설계를 위한 변경하기 힘든 요인과 매우 변경하기 힘든 요인을 정의해야 합니다 . 그런 다음 분할구 설계를 지정했는지 아니면 분할 - 분할구 설계를 지정했는지에 따라 분산 비율에 하나 또는 두 개의 양수를 입력할 수 있습니다 .

고급 옵션 > 사전 모수 분산 (" 모형 " 개요가 있는 경우에만 사용 가능) " 추정 가능성 " 이 "If Possible" 로 설정된 요인에 사용되는 가중치를 지정합니다 . " 설계 생성 " 을 클릭하면 옵션이 업데이트되어 기본 가중치가 표시됩니다. 가중치를 지정할 각 항에 대해 양수를 입력합니다 . 입력한 값은 사전 분산에 대한 역수의 제곱근입니다 . 값이 클수록 분산이 작으므로 효과가 활성화되지 않는다는 더 많은 사전 정보가 있습니다 .

베이지안 D- 최적 또는 I- 최적은 If Possible 요인이 있는 설계를 구성할 때 사용됩니다 . 알고리즘에 사용되는 기본값은 필수 항의 경우 0, If Possible 항이 포함된 교호작용의 경우 4, If Possible 항의 경우 1 입니다 . 자세한 내용은 " 기술 상세 정보 " 부록의 " 별칭 행렬 "(767 페이지) 및 " 최적 기준 "(118 페이지) 에서 확인하십시오 .

고급 옵션 > A- 최적 모수 가중치 (A- 최적 설계에 사용) 모형 모수에 대한 가중치를 지정합니다 . 예를 들어 2 차 효과보다 주효과의 분산에 더 가중치를 부여할 수 있습니다 . 모수 가중치에 대한 자세한 내용은 Morgan and Stallings(2017) 에서 확인하십시오 .

고급 옵션 > D 효율도 가중치 설계를 구성할 때 별칭 최적에 대한 D 효율도의 상대 중요도를 지정합니다 . 계수의 분산을 줄이는 것과 적합한 별칭 구조를 얻는 것 사이의 균형을 잡으려면 이 옵션을 선택합니다 . 값은 0 에서 1 사이여야 합니다 . 값이 클수록 D 효율도에 더 많은 가중치를 둡니다 . 기본값은 0.5 입니다 . 이 옵션은 " 최적 설계 별칭 생성 " 을 " 최적 기준 " 으로 선택한 경우에만 효과가 있습니다 .

D 효율도의 정의는 " 최적 기준 "(118 페이지) 에서 확인하십시오 . 별칭 최적에 대한 자세한 내용은 " 별칭 최적 "(121 페이지) 에서 확인하십시오 .

스크립트를 스크립트 창에 저장 " 사용자 설계 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

반응 시뮬레이션

데이터 수집 전에 또는 교육용 설정에서 설계를 탐색하려면 시뮬레이션된 반응을 사용합니다 . " 반응 시뮬레이션 " 을 선택하고 " 테이블 생성 " 을 클릭하여 설계 테이블을 생성하면 " 반응 시뮬레이션 " 옵션이 각 반응에 대해 다음을 수행합니다 .

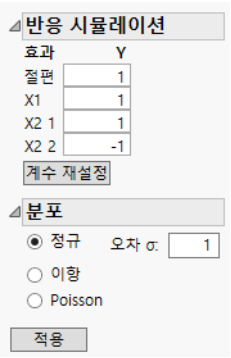
- 설계 테이블의 반응 열에 확률 반응 값을 추가합니다 .
- 시뮬레이션 모형 계산식을 포함하는 새 열을 설계 테이블에 추가합니다. 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .

모수 값을 지정하고 시뮬레이션을 위한 반응 분포를 선택할 수 있는 "모형" 창이 나타납니다. "모형" 창에서 "적용"을 클릭하면 시뮬레이션 모형 계산식을 포함하는 각 열이 업데이트됩니다.

컨트롤 창

그림 4.27에서는 연속형 요인(X1)과 3수준 범주형 요인(X2)이 각각 하나씩 있는 설계의 "모형" 창을 보여 줍니다. X2는 두 개의 모형 항으로 나타납니다.

그림 4.27 반응 시뮬레이션 컨트롤 창



초기 창에는 계수 값이 1 또는 -1로 표시되고, 오차 표준편차가 1인 정규 분포가 선택되어 있습니다. DOE 창의 "설계 평가"에서 "검정력 분석"의 일부로 "예상 계수"를 설정한 경우, "검정력 분석" 개요에서 "예상 계수" 및 "예상 RMSE"(오차 표준편차)로 지정한 값이 "반응 시뮬레이션" 개요의 기본값이 됩니다. 데이터 테이블의 모형 스크립트에 지정된 모형을 적합시킬 수 없는 경우 절편과 계수의 기본값은 0입니다.

반응 시뮬레이션

시뮬레이션된 값에 대한 모형을 지정하려면 다음을 수행하십시오.

1. "효과" 목록의 각 항에 대해 반응 값을 시뮬레이션하는 데 사용되는 선형 모형의 계수를 입력합니다. 이 값은 선형 함수 $L(\mathbf{x}, \beta) = \mathbf{x}'\beta$ 를 정의합니다. 그림 4.27 의 "반응 시뮬레이션" 개요를 참조하십시오.
   - 벡터 **x** 는 "효과" 아래에 나열된 효과를 정의하는 항으로 구성됩니다.
   - 벡터 β 는 "Y" 아래에 지정한 모형 계수의 벡터입니다.
2. "분포"에서 반응 분포를 선택합니다.
3. **적용**을 클릭합니다. 시뮬레이션된 값과 계산식을 포함하는 <Y> 시뮬레이션 열이 설계 테이블에 추가됩니다. 여기서 Y 는 반응 열의 이름입니다.

분포

"반응 시뮬레이션" 창에서 사용 가능한 다음 분포 중 하나를 선택합니다.

정규 정규 분포에서 값을 시뮬레이션합니다. "오차 σ"에 정규 분포 오차의 표준편차 값을 입력합니다. "요인" 개요에서 요인의 "변경"을 "어려움"으로 지정한 경우 "주구 σ"에 주구 오차 값을 입력할 수 있습니다. 요인의 "변경"을 "어려움" 및 "매우 어려움"으로 지정한 경우에는 하위구 오차와 주구 오차 값을 둘 다 입력할 수 있습니다. "적용"을 클릭하면 모형에 기초하여 확률 반응 벡터가 포함된 계산식과 난수 값이 <Y> 시뮬레이션 열에 입력됩니다.

이항 이항 분포에서 값을 시뮬레이션합니다. "N"에 시도 수 값을 입력합니다. 성공 확률이 $1/(1 + exp(-L(\mathbf{x}, \beta))$ 이고 시도 횟수가 N 인 이항 분포에 따라 정수인 난수가 생성됩니다. "적용"을

클릭하면 난수 값과 계산식이 <Y> 시뮬레이션 열에 입력됩니다 . N 값을 포함하는 시도 수 열도 데이터 테이블에 추가됩니다 .

Poisson 모수 *exp((L(***x,** β*))* 를 사용하여 Poisson 분포에 따라 정수인 난수를 시뮬레이션합니다 . " 적용 " 을 클릭하면 난수 값과 계산식이 <Y> 시뮬레이션 열에 입력됩니다 .

참고 : " 테이블 생성 " 을 클릭할 때마다 반응을 시뮬레이션하도록 환경 설정을 지정할 수 있습니다 . 이렇게 하려면 **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다 . 그런 다음 **반응 시뮬레이션** 을 선택합니다 .

행렬 X 저장

이 옵션은 적률 행렬과 모형 행렬을 각각 포함하는 " 적률 행렬 " 및 " 모형 행렬 " 스크립트를 설계 데이터 테이블에 저장합니다. 적률 행렬과 모형 행렬은 " 설계 평가 " 개요의 " 설계 진단 " 섹션에 나오는 " 예측 평균 분산 "을 계산하는 데 사용됩니다. 자세한 내용은 Goos and Jones(2011) 에서 확인하십시오. 분할구 설계의 경우 "V 역 " 스크립트도 저장됩니다. "V 역 " 스크립트에는 반응에 대한 공분산 행렬의 역이 포함됩니다.

주의 : 명목형 요인이 있는 설계의 경우 " 행렬 X 저장 " 옵션으로 저장한 모형 행렬은 선형 모형을 적합시킬 때 사용된 코딩 행렬이 **아닙니다** . " 모형 " 스크립트를 실행할 때 나타나는 " 모형 적합 " 보고서에서 " 열 저장 > 코드화 테이블 저장 " 옵션을 선택하면 모형 적합에 사용되는 코딩 행렬을 가져올 수 있습니다 .

참고 : 행렬 스크립트를 항상 저장하도록 환경 설정을 지정할 수 있습니다 . **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다 . 그런 다음 " 행렬 X 저장 " 을 선택합니다 .

모형 행렬

모형 행렬은 실험에 대한 설계를 설명합니다 . 모형 행렬에는 " 모형 " 개요에서 지정한 모형의 각 항에 대한 열과 각 런에 대한 행이 있습니다 . 각 런에 해당하는 모형 행렬의 행에는 모형 항의 코드화된 값이 포함됩니다 .

- 연속형 항은 -1 ~ 1 범위로 코드화됩니다 .
- 명목형 항은 선형 모형을 적합시키는 데 사용하는 코딩 벡터에 Gram-Schmidt 직교화 프로시저를 적용하여 코드화됩니다 .

Gram-Schmidt 직교화 프로시저에 대한 자세한 내용은 Horn and Johnson(2012) 에서 확인하십시오 .

참고 : 검정력 분석의 코딩은 선형 모형 적합의 코딩과 매칭됩니다 . 선형 모형 적합에서 명목형 항에 사용되는 코딩에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

적률 행렬

적률 행렬은 모형 효과에 종속되지만 설계와 독립적입니다. 이 행렬은 다음과 같이 정의됩니다.

$$\mathbf{M} = \int_R f(\mathbf{x}) f(\mathbf{x})' d\mathbf{x}$$

여기서 $f(x)$는 요인 벡터 x의 요인 조합에 해당하는 모형 효과를 나타내고 R는 설계 공간을 나타냅니다. 적률 행렬과 설계 행렬에 대한 자세한 내용은 Goos and Jones(2011) 및 Myers et al. (2009)의 내용을 기반으로 합니다. 적률 행렬을 Myers et al. (2009)에서는 영역 적률(region moments) 행렬이라고 합니다.

스크립트

" 사용자 설계 " 의 빨간색 삼각형 메뉴에서 **행렬 X 저장**을 선택합니다. 설계와 테이블이 생성된 후 " 적률 행렬 ", " 모형 행렬 " 및 "V 역 "(분할구 설계의 경우) 스크립트가 " 사용자 설계 " 테이블에 테이블 스크립트로 저장됩니다.

- " 적률 행렬 ", " 모형 행렬 " 또는 "V 역 " 스크립트에서 마우스 오른쪽 버튼을 클릭하고 **편집**을 선택합니다. 스크립트에 해당 행렬이 표시됩니다. 작성한 스크립트에 이 행렬을 복사할 수 있습니다.
- " 적률 행렬 " 스크립트를 실행하면 로그에 "Moments" 라는 적률 행렬의 행 수가 표시됩니다.
- " 모형 행렬 " 스크립트를 실행하면 로그에 "X" 라는 모형 행렬의 행 수가 표시됩니다.
- "V 역 " 스크립트를 실행하면 로그에 "V Inverse" 라는 공분산 역행렬의 행 수가 표시됩니다.

예제

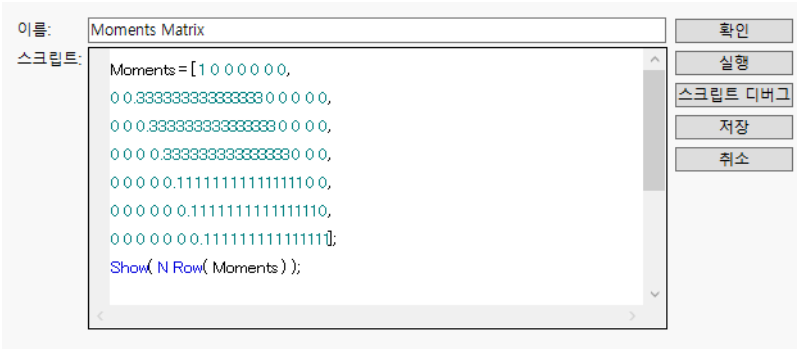
이 예에서는 모형 행렬 스크립트를 사용하는 방법을 보여 줍니다.

팁 : 로그를 보려면 **보기 > 로그** (macOS 에서는 **창 > 로그**) 를 선택합니다.

1. **DOE > 사용자 설계**를 선택합니다.
2. 연속형 요인을 3 개 추가하고 **계속**을 클릭합니다.
3. **교호작용 > 2 차**를 클릭합니다.
4. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **행렬 X 저장**을 선택합니다.
5. 기본값 런 수 (12) 를 사용하고 **설계 생성** , **테이블 생성**을 차례로 클릭합니다.
6. 테이블 패널에서 " 적률 행렬 " 을 마우스 오른쪽 버튼으로 클릭하고 **편집**을 선택합니다.

   스크립트 창에 스크립트가 나타납니다. 이 스크립트에는 **Moments** 라는 적률 행렬이 표시됩니다.

그림 4.28 적률 행렬 스크립트



7. 로그가 아직 열려 있지 않으면 **보기 > 로그** (macOS 에서는 **창 > 로그**) 를 선택합니다 .
8. 테이블 패널에서 " 적률 행렬 " 스크립트 옆의 녹색 삼각형을 클릭합니다 .
   행 수가 로그에 `N Row(Moments)=7` 로 나타납니다 .
9. 테이블 패널에서 " 모형 행렬 " 을 마우스 오른쪽 버튼으로 클릭하고 **편집**을 선택합니다 .
   스크립트 창에 스크립트가 나타납니다 . 이 스크립트에는 **X** 라는 모형 행렬이 표시됩니다 .
10. **실행**을 클릭합니다 .
    행 수가 로그에 `N Row(X)=12` 로 나타납니다 .
11. 모형 행렬을 데이터 테이블로 보려면 다음 행을 스크립트에 추가합니다 .

```
dt = New Table( "Model Matrix" );
dt << Set Matrix( X );
```

12. **실행**을 클릭합니다 .

시작 수

시작 수는 좌표 교환 알고리즘이 새 설계로 시작하는 횟수입니다. 자세한 내용은 "좌표 교환 알고리즘"(122 페이지)에서 확인하십시오. **시작 수** 옵션을 사용하여 원하는 값을 지정할 수 있습니다. 랜덤 시작 횟수가 증가하면 결과 설계의 최적성이 향상되는 경향이 있습니다.

"시작 수" 값을 지정하고 "확인"을 클릭하는 경우 외에는 설계 검색 시간에 의해 시작 수가 제어됩니다. 설계를 구성하는 데 사용된 시작 수를 확인하려면 "설계 생성"을 클릭합니다. 그런 다음 "시작 수"를 선택합니다. 텍스트 상자의 값이 특정 설계를 구성하는 데 사용된 시작 수입니다.

특수 경우에 대한 전역 최적 설계 이론이 있습니다. 좌표 교환 알고리즘은 최적 설계를 찾았음을 감지하면 검색을 중지하고 해당 설계를 반환합니다.

팁 : 특정 설계를 재현하려면 설계를 생성하는 데 사용된 원래 " 시작 수 " 와 " 난수 시드값 " 을 지정해야 합니다 . " 설계 생성 " 을 클릭한 후 빨간색 삼각형 옵션에서 이러한 값을 확인하십시오 .

이 설명서의 사용자 설계 예에서는 설명 중인 설계를 정확하게 재생성할 수 있도록 난수 시드값과 시작 수를 제공하는 경우가 많습니다.

설계 검색 시간

설계 검색 시간은 최적 설계를 찾는 데 할당된 시간입니다. 사용자 설계의 좌표 교환 알고리즘은 랜덤 시작 설계를 기반으로 하여 근사 최적 설계를 찾는 것으로 구성됩니다. 자세한 내용은 "좌표 교환 알고리즘"(122 페이지)에서 확인하십시오. 설계 검색 시간은 랜덤 시작 설계를 기반으로 생성되는 설계 수를 결정합니다.

설계 검색 시간 옵션을 사용하여 원하는 값을 지정할 수 있습니다. 검색 시간을 늘리면 결과 설계의 최적성이 향상되는 경향이 있습니다.

알고리즘을 재실행하여 생성된 설계는 다를 수 있습니다. 동일한 난수 시드값을 사용하더라도 계산 속도와 기타 요인의 변동으로 인해 최종 설계를 구성하는 데 사용되는 시작 설계의 수가 다를 수 있습니다.

참고 : 시작 설계 수는 " 시작 수 " 텍스트 상자의 값으로 지정됩니다 . 그러나 이 값은 " 설계 생성 " 을 클릭하여 설계를 구성한 후에 업데이트됩니다 .

특수 경우에 대한 전역 D- 최적 설계 이론이 있습니다 . 예를 들면 다음과 같습니다 .

- 2 수준 부분 요인 설계 또는 비정칙 직교 배열 . 이는 모든 주효과와 2 요인 교호작용 모형을 위한 전역 D- 최적 설계입니다 .
- 라틴 정방 설계 . 이는 표본 크기와 요인 수준 수가 동일하다는 가정하에 주효과 모형을 위한 D- 최적 설계입니다 .
- Plackett-Burman 설계 . 이는 주효과 모형을 위한 D- 최적 설계입니다 .

좌표 교환 알고리즘은 최적 설계를 찾았음을 감지하면 검색을 중지하고 해당 설계를 반환합니다.

사용자 설계에 대한 기술 상세 정보

- " 랜덤화 제한이 있는 설계 "
- " 변경하기 힘든 수준의 공변량 "
- " 주구 및 하위구 수 "
- " 최적 기준 "
- "D 효율도 "
- " 좌표 교환 알고리즘 "

랜덤화 제한이 있는 설계

완전히 랜덤화된 설계를 실행하는 것이 항상 실현 가능하지는 않습니다. 실험을 설계할 때는 랜덤화에 대한 논리적 또는 물리적 제한을 고려해야 합니다. 이 섹션에서는 요인 수준에 대한 실험 단위 랜덤 할당이 제한되는 다양한 유형의 설계를 사용자 설계 플랫폼에서 어떻게 처리하는지 설명합니다. 랜덤 블록 설계 및 다양한 유형의 분할구 설계가 포함됩니다.

랜덤 블록 설계

랜덤 블록 설계는 실험의 런을 더 큰 모집단에서 무작위로 선택한 것으로 간주되는 블록으로 그룹화합니다. 대개 런 블록 내의 런은 다른 블록의 런보다 동질적입니다. 이러한 경우 블록화 변수로 설명된 변동을 처리하면 다른 효과를 더 잘 식별할 수 있습니다.

랜덤 블록 설계 시나리오

Goos(2002)는 페이스트리 반죽 혼합물 실험과 관련된 예를 제공합니다. 실험의 목적은 반죽의 특정 특성이 공급 유량, 초기 수분 함량, 회전 나사 속도라는 세 가지 요인에 어떻게 의존하는지 이해하는 것입니다. 하루에 런을 4회만 수행할 수 있으므로 실험을 실행하는 데 며칠이 걸렸습니다. 환경 변수에 대한 랜덤 일별 차이가 특정 일에 수행되는 모든 런에 영향을 줄 가능성이 있습니다. 일별 변동을 처리하기 위해 런을 크기 4의 블록으로 그룹화하여 이 변동이 세 요인에 대한 정보를 손상시키지 않도록 했습니다.

블록 요인 Day는 매일 수행되는 런으로 구성됩니다. 실험이 수행된 날은 환경 조건이 다른 여러 날의 큰 모집단을 대표합니다. 따라서 Day는 랜덤 블록 요인입니다.

랜덤 블록 설계 설정

랜덤 블록 설계를 생성하려면 사용자 설계 플랫폼을 사용하여 반응과 요인을 입력하고 모형을 정의합니다. "설계 생성" 개요에서 **런을 다음 크기의 랜덤 블록으로 그룹화** 옵션을 선택하고 각 블록에 포함할 런 수를 입력합니다. 자세한 내용은 "설계 구조 옵션"(95페이지)에서 확인하십시오.

참고 : 고정 블록 요인을 정의하려면 " 요인 " 개요에 블록 요인을 입력합니다 . 랜덤 블록 요인을 정의하려면 " 요인 " 개요에 블록 요인을 입력하지 마십시오 . 대신 " 설계 생성 " 아래의 **런을 다음 크기의 랜덤 블록으로 그룹화** 옵션을 선택합니다 .

분할구 설계

분할구 설계는 특정 요인의 설정이 런 그룹에 대해 일정하게 유지되는 상황에서 사용됩니다. 산업 분야에서 이러한 요인은 대개 런 간 변경이 어렵거나 비용이 많이 듭니다. 런 그룹에 대해 설정을 일정하게 유지해야 하는 요인은 JMP에서 **변경하기 힘든 요인**으로 분류됩니다.

특정 요인은 변경하기 힘들기 때문에 실험 단위에 요인을 무작위로 할당하는 것은 실용적이지 않습니다. 대신 단위 그룹에 요인을 할당합니다. 이렇게 하면 설계 생성 및 결과 분석에서 고려해야 할 랜덤화에 대한 제한이 적용됩니다.

분할구 설계 시나리오

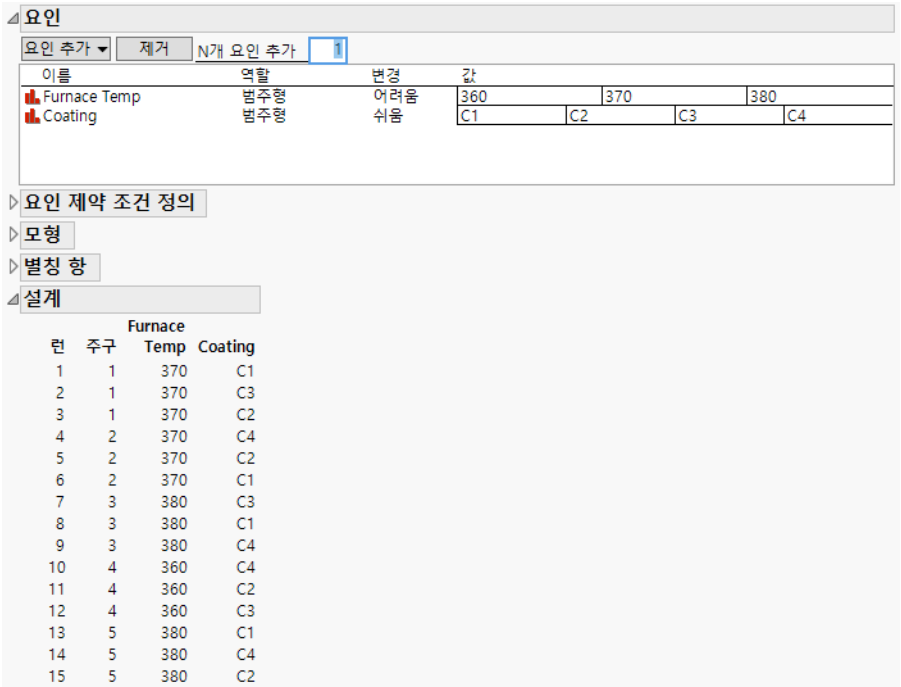
Box et al. (2005) 은 철근의 내식성을 연구하기 위한 실험을 제공합니다. 철근은 경화를 위해 용광로에 놓습니다. 그 후에 내식성 증가를 위해 코팅이 적용됩니다. 두 가지 관심 요인은 다음과 같습니다.

- Furnace 섭씨 온도 - 360, 370, 380 수준
- Coating 코팅 유형 - C1, C2, C3, C4 수준

Furnace Temp는 용광로 온도를 재설정하는 데 소요되는 시간 때문에 변경하기 힘든 요인입니다. 따라서 용광로 온도 설정마다 각각 4개의 철근이 처리됩니다. 이후 단계에서는 4개의 코팅이 무작위로 4개의 철근에 할당됩니다.

실험 단위는 철근입니다. Furnace Temp는 수준으로 주구를 정의하는 변경하기 힘든 요인입니다. 각 주구 내에서 Coating 요인은 주구 요인이 적용된 실험 단위에 무작위로 할당됩니다.

그림 4.29 분할구 설계를 위한 요인 및 설계 개요



런	주구	Furnace Temp	Coating
1	1	370	C1
2	1	370	C3
3	1	370	C2
4	2	370	C4
5	2	370	C2
6	2	370	C1
7	3	380	C3
8	3	380	C1
9	3	380	C4
10	4	360	C4
11	4	360	C2
12	4	360	C3
13	5	380	C1
14	5	380	C4
15	5	380	C2

부식 실험의 "요인" 개요에는 Furnace Temp의 "변경"이 "어려움"으로 설정되고 Coating의 "변경"이 "쉬움"으로 설정되어 있습니다. 15회 런이 5개의 주구로 구성되며, 주구 내에서 Temperature 설정이 일정하게 유지됩니다.

분할구 설계 설정

일반적으로 설정을 변경하기 힘든 몇 가지 요인을 처리 단계에 적용할 수 있습니다. 용광로 예에서는 온도뿐만 아니라 용광로 위치도 고려할 수 있습니다. "요인" 개요의 "변경" 열에서 이러한 요인의 "변경" 값을 "어려움"으로 지정합니다.

사용자 설계에 변경하기 쉬운 요인과 변경하기 힘든 요인만 포함된 경우 변경하기 힘든 요인의 런은 **주구**라는 새 요인을 사용하여 그룹화됩니다. **주구** 값은 변경하기 힘든 요인에 대해 동일한 설정으로 런 블록을 지정합니다. 설계 테이블에 저장된 "모형" 스크립트는 **주구**를 임의 효과로 처리합니다. 자세한 내용은 "변경"(85페이지) 및 "설계 구조 옵션"(95페이지)에서 확인하십시오.

분할구 설계를 생성하고 실험 데이터를 분석하는 예는 "사용자 설계의 예" 장의 "분할구 실험"(180페이지)에서 확인하십시오.

분할 - 분할구 설계

분할-분할구 설계는 변경하기 힘든 두 가지 수준의 요인이 있을 때 사용됩니다. 산업 분야에서 첫 번째 처리 단계의 재료나 실험 단위를 두 번째 처리 단계에 일괄적으로 전달할 때 이러한 설계가 종종 수행됩니다. 첫 번째 단계에서 일련의 재료에 요인이 적용됩니다. 그런 다음 추가 요인을 연구하는 두 번째 단계 처리를 위해 해당 배치가 나뉩니다. 첫 번째 단계 요인은 **매우 변경하기 힘든** 요인이고, 두 번째 단계 요인은 **변경하기 힘든** 요인으로 간주됩니다. 두 번째 처리 단계 후에 실험 단위에 추가 요인을 적용할 수 있습니다. 이러한 요인은 **변경하기 쉬운** 요인으로 간주됩니다.

분할-분할구 설계에서는 배치가 랜덤 블록으로 간주됩니다. 배치는 두 번째 단계 처리를 위해 나누어지므로 두 번째 단계 요인이 첫 번째 단계 요인 내에 내포됩니다.

분할 - 분할구 설계 시나리오

Schoen(1999) 은 치즈 품질과 관련된 분할 - 분할구 설계의 예를 제공합니다 . 요인은 Design Experiment 폴더에 있는 Cheese Factors.jmp 데이터 테이블에 제공됩니다 . 실험은 다음과 같은 세 단계의 처리로 구성됩니다 .

- 농부로부터 우유를 받아 큰 탱크에 저장합니다 .
- 이 탱크의 우유를 커드 가공에 사용되는 더 작은 탱크로 분배합니다 .
- 각 탱크의 커드를 개별 치즈 가공을 위해 프레스로 이동합니다 .

실험은 다음 요인에 대한 검정으로 구성됩니다 .

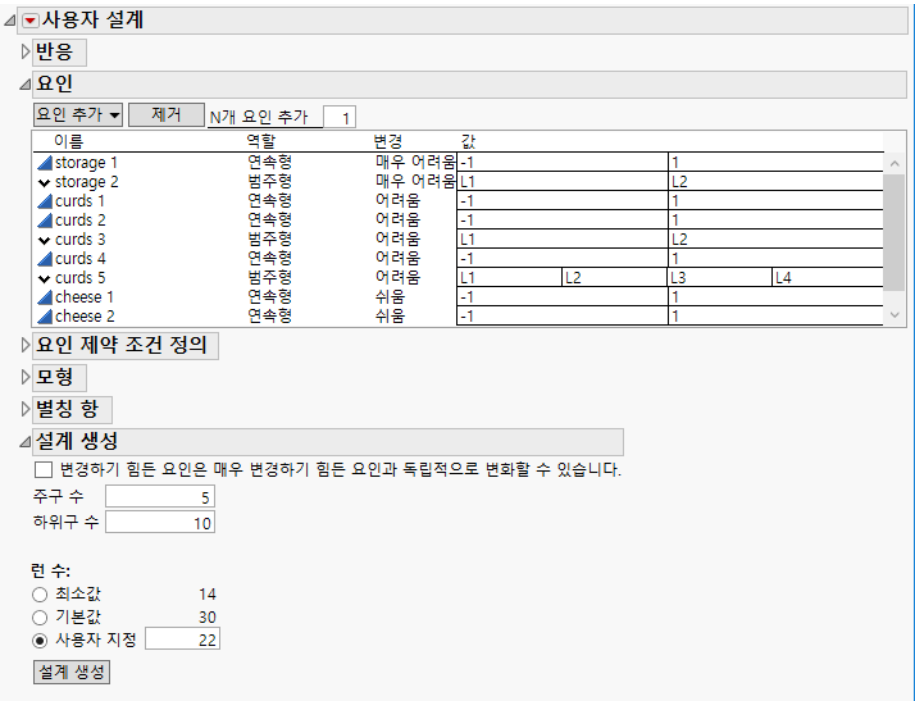
- 우유가 큰 저장 탱크에 있을 때 적용된 두 개 요인
- 더 작은 커드 가공 탱크에 적용된 5 개 요인
- 커드 가공 탱크의 개별 치즈에 적용된 3 개 요인

커드 가공 탱크에 적용된 요인의 수준(하위구)은 우유 저장 탱크에 적용된 요인의 수준(주구)에 내포됩니다.

" 요인 " 개요에는 " 변경 " 이 " 매우 어려움 ", " 어려움 " 또는 " 쉬움 " 으로 설정되어 있습니다 .

- 매우 어려움 - 저장 탱크 요인 두 개
- 어려움 - 커드 가공 탱크 요인 5 개
- 쉬움 - 치즈에 무작위로 할당할 수 있는 요인 3 개

그림 4.30 분할 - 분할구 설계를 위한 요인 및 설계 생성 개요



기본 주구 수는 5이고 기본 하위구 수는 10입니다. 런 수는 22로 설정되어 있습니다. 하위구가 10개만 사용될 경우 설계의 주구 수가 충분하지 않아 하위구 분산을 추정할 수 없습니다. 하위구 수를 11로 변경하고 **설계 생성**을 클릭하여 22개 런 설계를 확인합니다.

그림 4.31 치즈 시나리오의 분할 - 분할구 설계

설계

런	주구	하위구	storage 1	storage 2	curds 1	curds 2	curds 3	curds 4	curds 5	cheese 1	cheese 2	cheese 3
1	1	1	1	L2	-1	-1	L1	-1	L1	1	-1	L1
2	1	1	1	L2	-1	-1	L1	-1	L1	-1	1	L2
3	1	1	1	L2	-1	-1	L1	-1	L1	1	-1	L3
4	1	2	1	L2	1	1	L1	1	L2	1	1	L3
5	1	2	1	L2	1	1	L1	1	L2	-1	-1	L1
6	2	3	-1	L2	1	1	L1	-1	L3	1	1	L2
7	2	3	-1	L2	1	1	L1	-1	L3	-1	-1	L3
8	2	3	-1	L2	1	1	L1	-1	L3	1	-1	L1
9	2	4	-1	L2	-1	1	L2	1	L1	-1	-1	L2
10	2	4	-1	L2	-1	1	L2	1	L1	1	1	L1
11	3	5	1	L1	-1	-1	L2	1	L3	-1	1	L3
12	3	5	1	L1	-1	-1	L2	1	L3	1	-1	L2
13	3	6	1	L1	1	1	L2	-1	L4	1	1	L2
14	3	6	1	L1	1	1	L2	-1	L4	-1	-1	L1
15	4	7	-1	L1	-1	1	L1	-1	L2	-1	1	L1
16	4	7	-1	L1	-1	1	L1	-1	L2	1	-1	L3
17	4	8	-1	L1	1	-1	L1	1	L1	-1	1	L3
18	4	8	-1	L1	1	-1	L1	1	L1	1	-1	L2
19	5	9	-1	L2	1	-1	L2	-1	L2	1	1	L1
20	5	9	-1	L2	1	-1	L2	-1	L2	-1	-1	L2
21	5	10	-1	L2	-1	-1	L1	1	L4	-1	1	L1
22	5	10	-1	L2	-1	-1	L1	1	L4	1	-1	L3

5 개 주구는 저장 요인 **storage 1** 과 **storage 2** 에 해당합니다. 저장 요인의 설정은 주구 내에서 일정합니다. 이어지는 새 주구에서 주구 요인의 설정이 같을 경우에는 해당 주구 간에 요인 수준을 재설정해야 합니다. 예를 들어 런 10 과 11 간, 런 14 와 15 간에 **storage 1** 수준을 재설정하고, 런 18 과 19 간에 **storage 2** 수준을 재설정해야 합니다. 주구 변동을 포착하려면 지정된 설정이 같더라도 주구 간 요인을 재설정해야 합니다.

11 개 하위구는 커드 요인에 해당합니다. 하위구 내에서 커드 요인의 설정은 일정합니다. **하위구** 의 각 수준은 **주구** 의 한 수준 내에만 표시되어 **하위구** 수준이 **주구** 수준에 내포되어 있음을 나타냅니다.

치즈 요인의 수준은 각 런마다 무작위로 달라집니다.

분할 - 분할구 설계 설정

분할 - 분할구 설계의 " 요인 " 개요에는 " 변경 " 이 " 매우 어려움 " 및 " 어려움 " 으로 설정된 요인이 포함됩니다 . " 변경 " 이 " 쉬움 " 으로 설정된 요인도 설계에 포함될 수 있습니다 . 다음과 같은 두 요인이 생성됩니다 .

- **주구** 요인은 " 변경 " 이 " 매우 어려움 " 으로 설정된 요인의 상수 수준 블록을 나타냅니다 .
- **하위구** 요인은 " 변경 " 이 " 어려움 " 으로 설정된 요인의 상수 수준 블록을 나타냅니다 .
- **하위구** 요인은 " 변경 " 이 " 매우 어려움 " 으로 설정된 요인의 수준에 " 변경 " 이 " 어려움 " 으로 설정된 요인의 수준이 내포되어 있음을 반영합니다 .
- " 변경 " 이 " 쉬움 " 으로 설정된 요인의 수준은 하위구 내의 단위에 무작위로 할당됩니다 .
- **주구** 및 **하위구** 요인은 설계 테이블에 저장된 " 모형 " 스크립트에서 임의 효과로 처리됩니다 .

자세한 내용은 "요인 개요"(85 페이지) 및 "설계 구조 옵션"(95 페이지)에서 "변경"에 대한 설명을 참조하십시오.

이원 분할구 설계

이원 분할구(조각구 또는 분할 블록이라고도 함) 설계는 두 개의 분할구 성분으로 구성됩니다. 산업 분야에서 첫 번째 처리 단계의 재료나 실험 단위를 두 번째 처리 단계에 일괄적으로 전달할 때 이러한 설계가 수행됩니다. 그러나 첫 번째 처리 단계 후에 배치를 하위 배치로 나눌 수 있습니다. 두 번째 단계 처리 요인은 이러한 하위 배치에 무작위로 적용됩니다. 두 번째 단계의 특정 실험 설정이 있는 경우 해당 설정에 할당된 모든 하위 배치를 동시에 처리할 수 있습니다. 두 번째 처리 단계 후에 실험 단위에 추가 요인을 적용할 수 있습니다.

분할-분할구 설계와 달리 두 번째 단계 요인이 첫 번째 단계 요인에 **내포되지 않습니다**. 첫 번째 단계 후 배치가 나뉘어져 새 배치가 만들어집니다. 따라서 첫 번째 단계 요인과 두 번째 단계 요인 둘 다 전체 배치에 적용됩니다.

두 단계의 요인이 동등하게 변경하기 힘들 수 있지만 JMP 는 첫 번째 단계 요인을 **매우 변경하기 힘든** 것으로 나타내고 두 번째 단계 요인을 **변경하기 힘든** 것으로 나타내 이러한 요인을 구분합니다. 두 번째 처리 단계 후 실험 단위에 적용된 추가 요인은 변경하기 **쉬운** 것으로 간주됩니다.

이원 분할구 설계 시나리오

Vivacqua and Bisgaard(2004) 는 배터리 셀의 개회로 전압을 개선하기 위한 실험을 설명합니다. 다음 두 개의 처리 단계가 관심 영역입니다.

- 첫 번째 단계 : 연속 조립 공정
- 두 번째 단계 : 5 일 주기 시간의 큐어링 공정

엔지니어는 다음과 같은 2 수준 요인 6 개를 연구하려고 합니다.

- 조립 공정에 적용되는 네 개의 요인 (X1, X2, X3, X4)
- 큐어링 공정에 적용되는 두 개의 요인 (X5, X6)

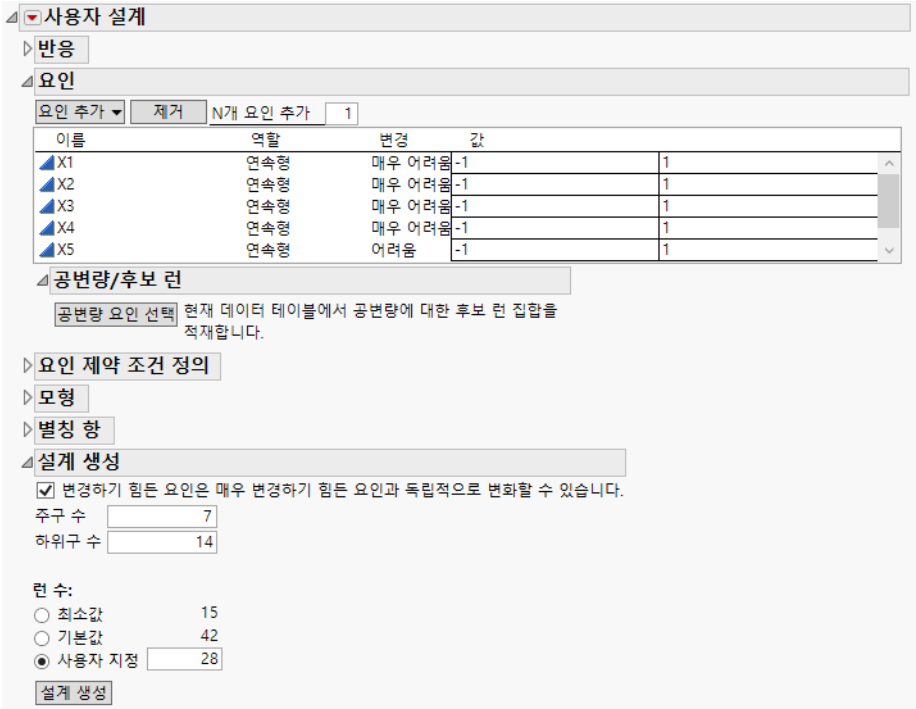
모든 요인이 2 수준인 완전 요인 설계에는 $2^6 = 64$ 회의 런이 필요하며 64*5 = 320 일이 소요됩니다. 또한 개별 배터리의 조립 조건을 변경하는 것은 실용적이지 않습니다. 그러나 2000 개 배터리 배치와 같은 큰 배치에 대해서는 조립 조건을 변경할 수 있습니다.

첫 번째 단계 요인과 두 번째 단계 요인은 둘 다 변경하기 어렵습니다. 어떤 의미에서는 두 개의 분할구 설계로 볼 수 있습니다. 그러나 첫 번째 단계 실험에서 2,000 개 배터리를 각각 500 개의 배터리로 구성된 하위 배치 4 개로 나눌 수 있습니다. 이 하위 배치를 두 번째 단계 요인 두 개에 대한 4 개의 설정에 무작위로 할당할 수 있습니다. 주어진 큐어링 조건 집합에 할당된 모든 배치를 동시에 처리할 수 있습니다. 즉, 첫 번째 단계와 두 번째 단계의 요인이 **교차**됩니다.

첫 번째 단계와 두 번째 단계의 요인을 구분하기 위해 첫 번째 단계 요인의 "변경"을 "매우 어려움"으로 지정하고 두 번째 단계 요인의 "변경"을 "어려움"으로 지정합니다(그림 4.32). 또한 "설계 생성" 아래의 "변경하기 힘든 요인은 매우 변경하기 힘든 요인과 독립적으로 변화할 수 있습

니다." 옵션에 주의하십시오. 이 옵션을 선택하지 않으면 두 수준에서 요인이 내포된 분할-분할 구 설계로 처리됩니다. 이원 분할구 설계를 생성하려면 이 옵션을 선택하십시오.

그림 4.32 이원 분할구 설계를 위한 요인 및 설계 생성 개요



기본 주구 수는 7이고 기본 하위구 수는 14입니다. **설계 생성**을 클릭하여 28회 런 설계를 확인합니다.

그림 4.33 배터리 셀의 이원 분할구 설계

설계

런	주구	하위구	X1	X2	X3	X4	X5	X6
1	1	1	1	1	1	-1	1	-1
2	1	2	1	1	1	-1	-1	1
3	1	3	1	1	1	-1	-1	-1
4	1	4	1	1	1	-1	1	1
5	2	5	-1	-1	-1	-1	-1	1
6	2	6	-1	-1	-1	-1	1	-1
7	2	7	-1	-1	-1	-1	-1	-1
8	2	8	-1	-1	-1	-1	1	1
9	3	9	1	1	-1	-1	-1	1
10	3	10	1	1	-1	-1	1	-1
11	3	11	1	1	-1	-1	-1	-1
12	3	12	1	1	-1	-1	1	1
13	4	13	1	-1	-1	1	-1	-1
14	4	14	1	-1	-1	1	1	1
15	4	1	1	-1	-1	1	1	-1
16	4	2	1	-1	-1	1	-1	1
17	5	3	-1	1	-1	1	-1	-1
18	5	4	-1	1	-1	1	1	1
19	5	5	-1	1	-1	1	-1	1
20	5	6	-1	1	-1	1	1	-1
21	6	7	1	-1	1	1	-1	-1
22	6	8	1	-1	1	1	1	1
23	6	9	1	-1	1	1	-1	1
24	6	10	1	-1	1	1	1	-1
25	7	11	-1	-1	1	-1	-1	-1
26	7	12	-1	-1	1	-1	1	1
27	7	13	-1	-1	1	-1	-1	-1
28	7	14	-1	-1	1	-1	1	1

7 개 주구는 첫 번째 단계 요인 X1, X2, X3 및 X4 에 해당합니다. 이 요인의 설정은 주구 내에서 일정합니다. 14 개 하위구는 두 번째 단계 요인 X5 와 X6 에 해당합니다. 예를 들어 주구가 다른 런 1 과 15 의 하위구에 동일한 하위구 처리가 적용됩니다. 여기서 X5 는 1 로 설정되고 X6 은 -1 로 설정됩니다.

이원 분할구 설계 설정

이원 분할구 설계에는 "변경"이 "매우 어려움" 및 "어려움"으로 설정된 요인이 필요합니다. "분할-분할구 설계 설정"(113 페이지)에 설명된 대로 **주구**와 **하위구**라는 요인이 생성됩니다. 그러나 이원 분할구 설계에서 **하위구**는 "변경"이 "매우 어려움"으로 설정된 요인의 수준에 "변경"이 "어려움"으로 설정된 요인의 수준을 내포하지 않습니다. **주구**와 **하위구**는 둘 다 설계 테이블에 저장된 "모형" 스크립트에서 임의 효과로 처리됩니다.

하위구 요인이 **주구** 요인에 **내포되지 않도록** 해야 합니다. 이렇게 하려면 "설계 생성" 개요에서 **변경하기 힘든 요인은 매우 변경하기 힘든 요인과 독립적으로 변화할 수 있습니다.** 옵션을 선택합니다(그림 4.32). 자세한 내용은 "변경"(85 페이지) 및 "설계 구조 옵션"(95 페이지)에서 확인하십시오.

분할구 설계를 생성하고 실험 데이터를 분석하는 예는 "사용자 설계의 예" 장의 "이원 분할구 실험"(184 페이지)에서 확인하십시오.

변경하기 힘든 수준의 공변량

사용자 설계 플랫폼을 사용하면 공변량을 변경하기 힘든 요인으로 지정할 수 있습니다. 실험 요인 검정에 사용할 수 있는 재료 배치에 대한 측정값이 있다고 가정해 보겠습니다. 또는 실험 요인 검정에 참여하기 위해 선택된 개인에 대한 측정값이 있다고 가정해 보겠습니다. 이러한 측정값은 실험 전에 알 수 있으므로 공변량으로 간주됩니다.

배치 또는 개인은 주구에 해당합니다. 이러한 주구 중 일부만 실험에 사용하고 싶을 수도 있습니다. 공변량 형태의 주구 정보를 사용할 수 있으므로 설계는 최적의 방식으로 주구를 선택해야 합니다.

"모형" 개요에 포함된 항에 의해 주어진 모형은 공변량 및 실험 요인을 사용하여 구성된 교호작용과 검정력을 포함할 수 있습니다.

참고 : " 공변량 " 요인의 " 변경 " 을 " 어려움 " 으로 설정하면 다른 모든 공변량도 " 어려움 " 으로 설정됩니다 . 나머지 요인은 " 쉬움 " 으로 설정해야 합니다 . 이 알고리즘을 사용하려면 행 교환과 좌표 교환의 조합이 필요하므로 적당한 크기의 설계도 생성하는 데 시간이 걸릴 수 있습니다 .

변경하기 힘든 공변량이 포함된 실험 시나리오

폴리프로필렌 판과 관련된 실험은 Goos and Jones(2011, Chapter 9) 및 Jones and Goos(2015)에 설명되어 있습니다. 여러 변수에 의해 결정된 다양한 배합에 따라 다수의 폴리프로필렌 판 배치가 생성됩니다. 일부 판은 바로 사용되고 나머지는 이후 실험용으로 보관됩니다. 보관된 배치의 구성은 알려져 있습니다.

고객은 판 배합과 관련하여 특정 요구 사항을 가지고 있습니다. 이후 실험에는 기체 플라스마 처리를 고객이 요구하는 유형의 배합으로 사용자 정의하는 것이 포함됩니다. 구성 변수는 변경하기 힘든 공변량으로 간주됩니다. 기체 플라스마 처리 요인을 주어진 배합으로 판 하위 배치에 적용할 수 있습니다.

최적 설계는 사용할 배치(공변량으로 정의됨)를 식별하고, 사용할 각 배치의 판 수를 결정하고, 기체 플라스마 수준에 대한 설정을 제공합니다. 주어진 배치에서 최적의 배치 및 판 수는 공변량에 따라 다릅니다.

"사용자 설계의 예" 장(123 페이지)에 예가 제공됩니다.

주구 및 하위구 수

JMP에서는 주구 수와 하위구 수의 기본값을 제안합니다. 이 값은 "모형" 개요에서 지정한 효과를 추정할 수 있는 균형 설계를 생성하기 위한 발견적 방법을 기반으로 합니다.

"주구 수" 또는 "하위구 수"에 결측값을 입력하면 설계의 D 효율도를 최대화하는 값이 자동으로 선택됩니다. 이 알고리즘은 "분할구 분산비" 옵션에 지정된 값을 사용합니다. 자세한 내용은 "고급 옵션 > 분할구 분산비"(102 페이지)에서 확인하십시오. D 효율도는 $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ 행렬식으로 지정됩니다. 여기서 $\mathbf{V}^{-1}$은 반응에 대한 분산 행렬의 역입니다. 자세한 내용은 Goos(2002)에서 확인하십시오.

"주구 수"와 "하위구 수"에 값을 입력하면 사용자 설계에서 결과 설계의 최적성을 최대화하려고 합니다. 분할구 설계에 대한 자세한 내용은 Jones and Goos(2007)에서 확인하십시오. 변경하기 힘든 공변량이 포함된 설계에 대한 자세한 내용은 Jones and Goos(2015)에서 확인하십시오.

최적 기준

사용자 설계는 최적 기준에 따라 달라지는 검색 루틴을 사용하여 생성됩니다. 이 섹션에서는 사용자 설계 생성을 위한 최적 기준에 대한 정보를 제공합니다.

- "D- 최적 "(118 페이지)
- " 베이지안 D- 최적 "(119 페이지)
- "I- 최적 "(120 페이지)
- " 베이지안 I- 최적 "(121 페이지)
- "A- 최적 "(120 페이지)
- " 별칭 최적 "(121 페이지)

D- 최적

기본적으로 사용자 설계 플랫폼은 "RSM" 버튼을 사용하여 완전 2 차 모형을 생성하는 경우를 제외하고 D- 최적 기준을 최적화합니다. "RSM" 버튼을 사용하는 경우에는 I- 최적 설계가 생성됩니다.

D- 최적 기준은 모형 계수 추정값의 공분산 행렬에 대한 행렬식을 최소화합니다. 결과적으로 D- 최적은 효과의 정확한 추정에 중점을 둡니다. 이 기준은 다음과 같은 경우에 적합합니다.

- 선별 설계
- 효과 추정 또는 유의성 검정에 중점을 둔 실험
- 활성 요인 식별이 실험 목표인 설계

D- 최적 기준은 가정된 모형에 종속됩니다. 참모형의 형태를 미리 알 수 없는 경우가 많아서 이 제한이 있습니다. D- 최적 설계의 런은 가정된 모형의 계수 정밀도를 최적화합니다. 극단적으로 D- 최적 설계는 런 수가 모수와 동일하고 적합 결여의 자유도가 없는 포화 상태일 수 있습니다.

특히 D- 최적 설계는 D 를 최대화합니다. 여기서 D 는 다음과 같이 정의됩니다.

$$D = \det[\mathbf{X}'\mathbf{X}]$$

$\mathbf{X}$ 는 "반응 시뮬레이션"(103 페이지)에 정의된 모형 행렬입니다.

D- 최적 분할구 설계는 D 를 최대화합니다. 여기서 D 는 다음과 같이 정의됩니다.

$$D = \det[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]$$

$\mathbf{V}^{-1}$ 은 반응에 대한 공분산 행렬의 대각 블록입니다(Goos 2002).

D-최적 설계는 계수의 표준 오차를 최소화하는 데 중점을 두기 때문에 모형이 올바른지 확인하지 못할 수 있습니다. 예를 들어 D-최적 설계에는 1차 모형의 중앙점이 포함되지 않습니다. 가정된 모형에 포함되지 않은 잠재적 활성 항이 있는 경우 If Possible 항을 지정하고 베이지안 D-최적 설계를 사용하는 것이 더 좋은 방법입니다.

베이지안 D- 최적

베이지안 D-최적은 수정된 D-최적 기준입니다. 베이지안 D-최적 기준은 잠재적 활성 교호작용 또는 비선형 효과가 있을 때 유용합니다. 자세한 내용은 DuMouchel and Jones(1994) 및 Jones et al(2008)에서 확인하십시오.

베이지안 D-최적은 지정된 모형 모수 집합을 정확하게 추정합니다. 이는 "모형" 개요에서 "추정 가능성"을 "필수"로 지정한 효과입니다. 그러나 이와 동시에 베이지안 D-최적은 런 크기에서 허용하는 대로 다른 효과(대개 고차 효과)를 추정할 수 있습니다. 이는 "모형" 개요에서 "추정 가능성"을 "If Possible"로 지정한 효과입니다. 런 크기 제한이 주어졌을 때 베이지안 D-최적 설계는 필수 효과만 포함된 모형에서 부적합성을 감지할 수 있습니다.

베이지안 D-최적 기준은 런 수가 "필수" 항의 수보다 많지만 "필수" 항과 "If Possible" 항의 합보다 작을 때 가장 효과적입니다. 이 경우 런 수는 추정할 모수의 수보다 적습니다. 베이지안 설정의 사전 정보를 사용하면 모든 필수 항을 정확하게 추정할 수 있는 동시에 일부 If Possible 항을 감지하고 추정할 수 있습니다.

의미 있는 사전 분포를 모형의 모수에 적용하기 위해 반응과 요인이 특정한 특성을 갖도록 조정됩니다(DuMouchel and Jones, 1994, Section 2.2).

다음 표기를 고려하십시오.

- **X** 는 " 반응 시뮬레이션 "(103 페이지) 에 정의된 모형 행렬입니다 .
- **K** 는 다음과 같은 값을 갖는 대각 행렬입니다 .
  - $k = 0$(필수 항의 경우)
  - $k = 1$(If Possible 주효과 , 검정력 , 3 수준 이상의 범주형 요인과 관련된 교호작용의 경우)
  - $k = 4$(다른 모든 If Possible 항의 경우)

If Possible 모수의 벡터에 설정된 사전 분포는 평균 벡터가 **0**이고, 대각 공분산 행렬의 대각 항목이 $1/k^2$인 다변량 정규 분포입니다. 따라서 k^2 값은 해당 모수의 사전 분산에 대한 역수입니다.

k의 값은 경험적으로 결정됩니다. If Possible 주효과, 검정력, 자유도가 2 이상인 교호작용의 사전 분산은 1입니다. 다른 If Possible 항의 사전 분산은 1/16입니다. DuMouchel and Jones(1994) 표기에서 $k = 1/\tau$입니다.

If Possible 항에 대한 가중치를 제어하려면 빨간색 삼각형 메뉴에서 **고급 옵션 > 사전 모수 분산**을 선택합니다. 자세한 내용은 "고급 옵션 > 사전 모수 분산"(103 페이지)에서 확인하십시오.

모수에 대한 사후 분포에는 공분산 행렬 $(\mathbf{X}'\mathbf{X}+\mathbf{K}^2)^{-1}$ 이 있습니다. 베이지안 D- 최적 설계는 사후 공분산 행렬의 역에 대한 행렬식을 최대화하여 생성됩니다.

$$\det(\mathbf{X}'\mathbf{X}+\mathbf{K}^2)$$

I- 최적

I- 최적 설계는 설계 공간에 대해 예측 평균 분산을 최소화합니다. 1 차 실험 목표가 계수 추정이 아니라 다음을 수행하는 것이라면 I- 최적 기준이 D- 최적보다 더 적합합니다.

- 반응 예측
- 최적의 작동 조건 결정
- 설계 공간에서 반응이 허용 가능한 범위 내에 속하는 영역 결정

이 경우에는 정확한 반응 예측이 정확한 모수 추정보다 우선합니다.

설계 공간의 x_0 지점에서 미지의 오차 분산에 대한 예측 분산은 다음과 같이 계산할 수 있습니다.

$$\mathrm{var}(\hat{\mathbf{Y}}|\mathbf{x}_0) = f(\mathbf{x}_0)'(\mathbf{X}'\mathbf{X})^{-1}f(\mathbf{x}_0)$$

$\mathbf{X}$는 "반응 시뮬레이션"(103 페이지)에 정의된 모형 행렬입니다.

I- 최적 설계는 전체 설계 공간에 대해 예측 분산의 적분 I 를 최소화합니다. 여기서 I 는 다음과 같이 지정됩니다.

$$I = \int_R f(\mathbf{x})'(\mathbf{X}'\mathbf{X})^{-1}f(\mathbf{x})d\mathbf{x} = \mathrm{trace}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{M}]$$

$\mathbf{M}$ 은 적률 행렬입니다.

$$\mathbf{M} = \int_R f(\mathbf{x})f(\mathbf{x})'d\mathbf{x}$$

자세한 내용은 "반응 시뮬레이션"(103 페이지)에서 확인하십시오. 반응 시뮬레이션에 대한 자세한 내용은 Goos and Jones(2011b)에서 확인하십시오.

적률 행렬은 설계에 의존하지 않으며 미리 계산할 수 있습니다. 행 벡터 $f(\mathbf{x})'$ 는 1 다음에 가정된 모형에 해당하는 효과가 나옵니다. 예를 들어 연속형 요인이 두 개 있는 완전 2 차 모형의 경우 $f(\mathbf{x})'$ 는 다음과 같이 정의됩니다.

$$f(\mathbf{x})' = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

A- 최적

A- 최적 기준은 모형 계수 추정값의 공분산 행렬에 대한 대각합을 최소화합니다. 대각합은 행렬에서 주대각 요소의 합입니다. A- 최적 설계는 회귀 계수 분산의 합을 최소화합니다.

베이지안 I- 최적

베이지안 I- 최적 설계는 설계 영역에 대해 필수 항과 If Possible 항에 대한 평균 예측 분산을 최소화합니다.

다음 표기를 고려하십시오.

- **X** 는 " 반응 시뮬레이션 "(103 페이지) 에 정의된 모형 행렬입니다.
- **K** 는 다음과 같은 값을 갖는 대각 행렬입니다.
  - $k = 0$(필수 항의 경우)
  - $k = 1$(If Possible 주효과 , 검정력 , 3 수준 이상의 범주형 요인과 관련된 교호작용의 경우)
  - $k = 4$(다른 모든 If Possible 항의 경우)

If Possible 모수의 벡터에 설정된 사전 분포는 평균 벡터가 **0** 이고, 대각 공분산 행렬의 대각 항목이 $1/k^2$인 다변량 정규 분포입니다. k 값에 대한 자세한 내용은 "베이지안 D- 최적"(119 페이지) 에서 확인하십시오.

$\mathbf{x}_0$ 지점에서 예측값의 사후 분산은 다음과 같습니다.

$$\text{var}(\hat{\mathbf{Y}}|\mathbf{x}_0) = f(\mathbf{x}_0)'(\mathbf{X}'\mathbf{X} + \mathbf{K}^2)^{-1} f(\mathbf{x}_0)$$

베이지안 I- 최적 설계는 설계 영역에 대해 다음과 같이 평균 예측 분산을 최소화합니다.

$$I_B = \text{trace}[(\mathbf{X}'\mathbf{X} + \mathbf{K}^2)^{-1}\mathbf{M}]$$

M 은 적률 행렬입니다. 자세한 내용은 "반응 시뮬레이션"(103 페이지) 에서 확인하십시오.

별칭 최적

별칭 최적은 가정된 모형에 있는 효과와 모형에 없지만 잠재적 활성 상태인 효과 사이의 별칭 관계를 최소화하려고 합니다. 모형에 없지만 잠재적 관심 대상인 효과를 **별칭 효과**라고 합니다. 별칭 - 최적 설계에 대한 자세한 내용은 Jones and Nachtsheim(2011b) 에서 확인하십시오.

구체적으로 $\mathbf{X}_1$을 "반응 시뮬레이션"(103 페이지) 에서 정의된 가정된 모형의 항에 해당하는 모형 행렬로 지정하겠습니다. 설계는 별칭 효과에 해당하는 모형을 정의합니다. $\mathbf{X}_2$로 별칭 효과에 대한 모형 항 행렬을 나타냅니다.

별칭 행렬은 다음과 같이 정의된 행렬 **A** 입니다.

$$\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$$

별칭 행렬의 항목은 모형 항 추정값과 관련된 편향 정도를 나타냅니다. 별칭 행렬의 파생에 대한 자세한 내용은 " 기술 상세 정보 " 부록의 "별칭 행렬"(767 페이지) 에서 확인하십시오.

A 에 있는 항목의 제곱합은 편향에 대한 요약 지표를 제공합니다 . 이 제곱합은 다음과 같이 대각합으로 나타낼 수 있습니다 .

$$\text{trace}(\mathbf{A}'\mathbf{A})$$

대각합 기준을 감소하는 설계는 일반적으로 D-최적 설계보다 D 효율도가 낮습니다. 결과적으로 별칭 최적은 D 효율도의 하한에 따라 $\mathbf{A}'\mathbf{A}$의 대각합을 최소화하려고 합니다. D 효율도의 정의는 "최적 기준"(118 페이지)에서 확인하십시오. D 효율도의 하한은 "고급 옵션"에서 D 효율도 가중치를 지정하여 구할 수 있습니다. 자세한 내용은 "고급 옵션 > D 효율도 가중치"(103 페이지)에서 확인하십시오.

D 효율도

X는 p개 모수를 사용하는 가정된 모형의 설계 또는 모형 행렬을 나타냅니다. 모형 행렬의 정의는 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 . $\mathbf{X}_D$ 는 가정된 모형의 D- 최적 설계에 대한 모형 행렬을 나타냅니다 . 그러면 **X** 에 의해 지정된 설계의 D 효율도가 다음과 같습니다 .

$$\text{D효율도} = \left[\frac{\det(\mathbf{X}'\mathbf{X})}{\det(\mathbf{X}_D'\mathbf{X}_D)}\right]^{1/p}$$

좌표 교환 알고리즘

사용자 설계 도구는 여러 최적 기준 중 하나를 최적화하는 설계를 구성합니다. 자세한 내용은 "최적 기준"(118 페이지)에서 확인하십시오. 기준을 최적화하기 위해 사용자 설계는 **좌표 교환 알고리즘**(Meyer and Nachtsheim, 1995)을 사용합니다. 이 알고리즘은 시작 설계를 생성하기 위해 각 요인 및 각 런에 대해 지정된 설계 영역 내에서 무작위로 값을 선택하는 것에서 시작됩니다.

요인 제약 조건이 없는 연속형 요인 및 주효과 모형이 연구에 필요하다고 가정해 보겠습니다 . 반복은 다음과 같이 모형 행렬의 각 값을 검정하는 것으로 구성됩니다 .

- 각 요인의 현재 값은 최극단값 두 개로 대체됩니다 .
- 두 대체 값에 대해 최적 기준이 계산됩니다 .
- 값 중 하나가 최적 기준을 증가시키면 이 값이 이전 값을 대체합니다 .

이 과정은 전체 반복에 대해 값 대체가 없을 때까지 계속됩니다.

다항식 항, 명목형 요인 및 요인 제약 조건을 처리하기 위해 알고리즘이 적절하게 조정됩니다.

이 과정에서 얻은 설계는 다양한 종류의 근접 설계에서 최적입니다. 그러나 이 설계는 **로컬에서만** 최적입니다. 전역 최적 설계를 찾을 확률을 높이기 위해 좌표 교환 알고리즘이 여러 번 반복됩니다. Goos and Jones(2011, p. 36)는 가장 자명한 경우를 제외한 모든 설계 상황에 대해 1,000 회 이상의 랜덤 시작을 사용하도록 권장합니다. 시작 설계 수는 "시작 수" 옵션에 의해 제어됩니다. 자세한 내용은 "시작 수"(107 페이지)에서 확인하십시오. 사용자 설계는 생성된 모든 설계 중에서 최적 기준을 최대화하는 설계를 제공합니다.

5 장

사용자 설계의 예
요구 사항을 충족하는 실험 수행

사용자 설계 플랫폼을 광범위한 실험 설계를 구성하기 위한 기본 도구로 사용합니다. 다양한 설계 유형을 구성하고 특정 실험 요구 및 리소스 예산에 맞게 세부적으로 조정할 수 있습니다.

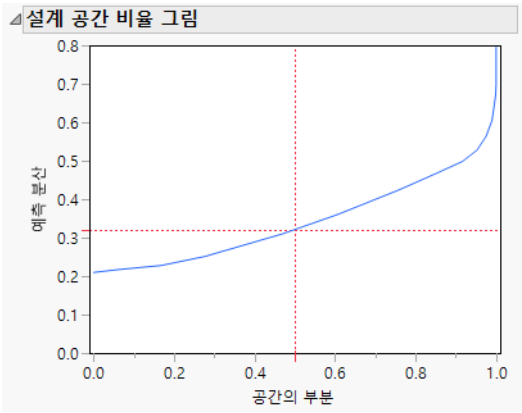
사용자 설계는 선별, 반응 표면, 완전 요인 및 혼합물 설계 플랫폼보다 많은 옵션과 컨트롤을 제공합니다. 구성한 설계는 목표에 맞게 특별히 생성됩니다. 따라서 목표를 충분히 충족시킬 수 없는 전통적 설계를 찾을 필요가 없습니다.

사용자 설계를 사용하여 구성할 수 있는 특수 목적의 유연한 설계는 다음과 같습니다.

- 선별 설계 (과포화 선별 설계 포함)
- 반응 표면 설계 (범주형 요인이 추가된 설계 포함)
- 혼합물 설계 (공정 요인이 추가된 설계 , 혼합물 혼합 설계 포함)
- 공변량을 포함하는 설계 또는 시간 선형 추세에 대한 로버스트 설계
- 고정 및 랜덤 블록 설계
- 분할구 , 분할 - 분할구 및 이원 분할구 (조각구) 설계

이 장에서는 사용자 설계 플랫폼을 사용하여 이러한 설계 유형을 구성해 봅니다. 대부분의 경우 실험 결과도 분석합니다. 사용자 설계 플랫폼에 대한 자세한 내용은 "사용자 설계" 장(63 페이지)에서 확인하십시오.

그림 5.1 설계 공간 비율 그림



목차

선별 실험 125
주효과만추정하는 설계 125
모든 2요인 교호작용을 추정하는 설계 128
주효과와 2요인 교호작용의 별칭 관계를 방지하는 설계 130
과포화 선별 설계 133
고정 블록 설계 140
반응 표면 실험 144
반응 표면 설계 144
유연한 블록화 방법을 사용하는 반응 표면 설계 150
D-최적과 I-최적의 반응 표면 설계 비교 152
제약 조건 및 범주형 요인이 있는 반응 표면 설계 157
혼합물 실험 162
비혼합 요인이 있는 혼합물 설계 162
혼합물 혼합 설계 166
공변량이 포함된 실험 170
고정 공변량이 포함된 설계 171
변경하기 힘든 공변량이 포함된 설계 173
시간 선형 추세가 포함된 설계 177
랜덤화 제한이 있는 실험 180
분할구 실험 180
이원 분할구 실험 184
로버스트 공정 및 제품 설계를 위한 실험 190

선별 실험

공정을 연구하는 초기 단계에서는 하나 이상의 반응에 영향을 줄 수 있는 요인 목록을 식별합니다. **활성** 요인, 즉 실제로 반응에 영향을 주는 요인을 식별하는 데 관심이 있습니다. **선별 설계**는 활성 가능성이 있는 요인을 판별하는 데 도움이 됩니다. 활성 요인이 식별되면 반응 표면 설계와 같은 더 정교한 설계를 구성하여 교호작용과 곡률을 모델링할 수 있습니다.

사용자 설계 플랫폼을 사용하여 구성한 선별 설계는 선별 설계 플랫폼에서 제공하는 전통적 설계와 동등한 경우가 많습니다. 그러나 사용자 설계는 전통적 선별 설계를 사용할 수 없는 경우 설계를 구성할 수 있습니다.

사용자 설계 플랫폼은 D- 최적 또는 베이지안 D- 최적 기준을 사용하여 선별 설계를 구성합니다 . D- 최적 기준은 모형 계수 추정값의 공분산 행렬에 대한 행렬식을 최소화합니다 . 결과적으로 D- 최적은 효과의 정확한 추정에 중점을 둡니다. 자세한 내용은 "사용자 설계" 장의 "최적 기준"(118 페이지) 에서 확인하십시오 .

- " 주효과만추정하는 설계 "
- " 모든 2 요인 교호작용을 추정하는 설계 "
- " 주효과와 2 요인 교호작용의 별칭 관계를 방지하는 설계 "
- " 과포화 선별 설계 "
- " 고정 블록 설계 "

주효과만추정하는 설계

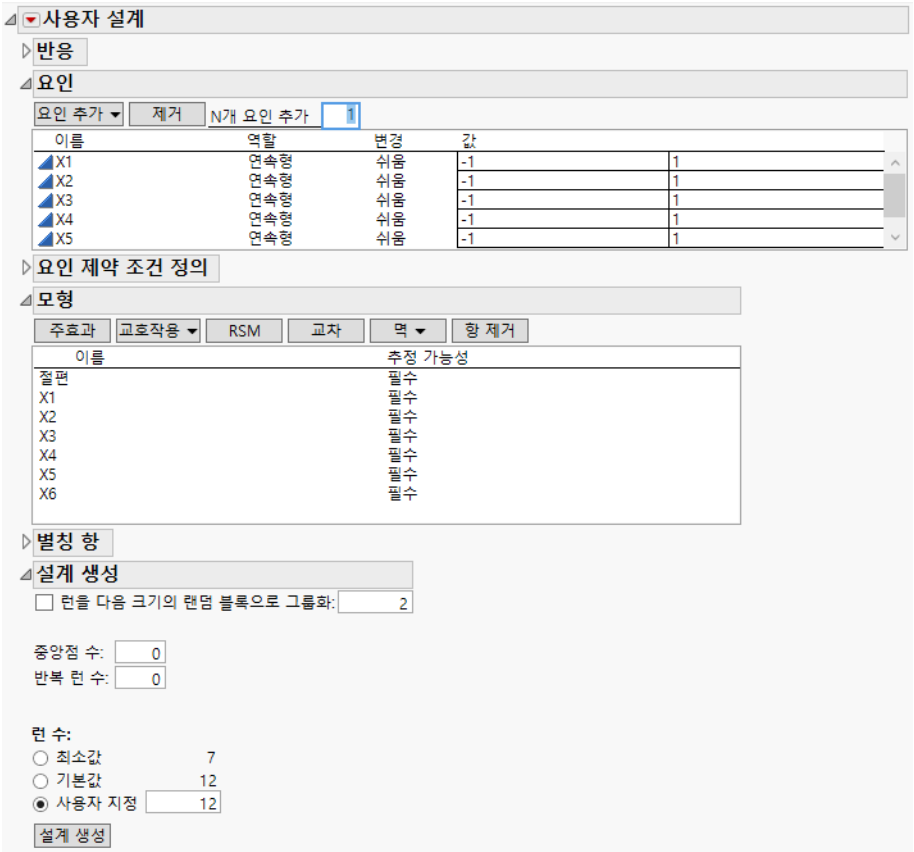
참고 : 주효과만 선별하는 설계에 대한 자세한 내용은 " 선별 설계 " 장 (267 페이지) 에서 확인하십시오 .

이 예에서는 6 개 요인의 주효과를 연구하려고 합니다 . 모든 주효과가 직교하는 선별 설계를 구성합니다 . 그러나 주효과는 2 요인 교호작용과 별칭 관계를 갖습니다 .

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 요인 " 개요에서 **N 개 요인 추가** 옆에 6 을 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. **계속**을 클릭합니다 .

   " 모형 " 개요가 나타납니다 . 여기에는 " 추정 가능성 " 이 " 필수 " 로 지정된 주효과만 포함되어 있습니다 . 즉 , 생성된 설계에서 모든 주효과를 추정할 수 있습니다 .

그림 5.2 모형 개요를 보여 주는 사용자 설계 창



기본값 런 수 12 를 그대로 유지합니다 .

참고 : 5 단계에서 " 난수 시드값 " 을 설정하고 6 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

5. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
6. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 1 을 입력하고 **확인**을 클릭합니다 .
7. **설계 생성**을 클릭합니다 .

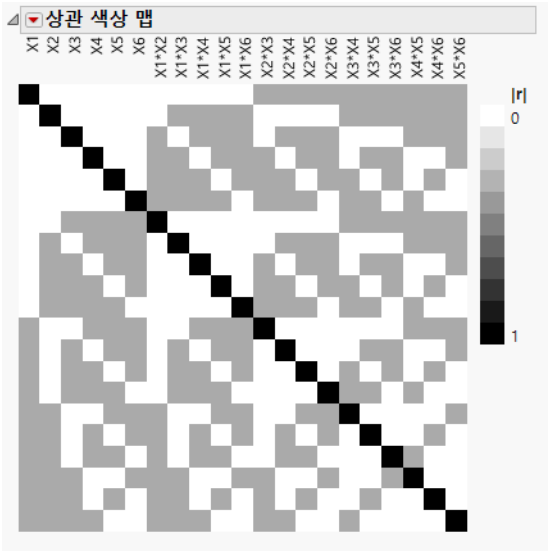
그림 5.3 주효과만 추정하는 설계

설계

런	X1	X2	X3	X4	X5	X6
1	-1	-1	-1	-1	-1	-1
2	1	-1	1	-1	-1	-1
3	1	1	-1	1	-1	-1
4	-1	1	1	-1	1	-1
5	1	-1	1	1	-1	1
6	1	1	-1	1	1	-1
7	1	1	1	-1	1	1
8	-1	1	1	1	-1	1
9	-1	-1	1	1	1	-1
10	-1	-1	-1	1	1	1
11	1	-1	-1	-1	1	1
12	-1	1	-1	-1	-1	1

8. **설계 평가 > 상관 색상 맵** 개요를 엽니다.

그림 5.4 상관 색상 맵



다음을 확인할 수 있습니다.

- 주효과는 맵의 왼쪽 위에 6 개 항으로 나타납니다.
- 6 개 주효과와 다른 주효과의 상관에 해당하는 흰색은 상관계수 0 을 나타냅니다. 이는 모든 주효과가 직교되고 서로 독립적으로 추정할 수 있음을 의미합니다.
- 일부 이원 교호작용에 해당하는 사각형의 연한 회색은 해당 효과가 서로 상관되어 있음을 나타냅니다. 이는 해당 효과를 다른 효과와 독립적으로 추정할 수 없음을 의미합니다. 정확한 상관을 보려면 이들 사각형을 커서로 가리킵니다.
- 서로 완전히 교락되는 효과는 없습니다. 절대 상관 1 을 나타내는 검은색 사각형은 주대각 위에만 있습니다.

9. **설계 평가 > 별칭 행렬** 개요를 엽니다.

그림 5.5 별칭 행렬

별칭 행렬

효과	X1*X2	X1*X3	X1*X4	X1*X5	X1*X6	X2*X3	X2*X4	X2*X5	X2*X6	X3*X4	X3*X5	X3*X6	X4*X5	X4*X6	X5*X6
절편	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X1	0	0	0	0	0	-0.33	0.333	0.333	-0.33	-0.33	-0.33	0.333	-0.33	-0.33	0.333
X2	0	-0.33	0.333	0.333	-0.33	0	0	0	0	-0.33	0.333	0.333	-0.33	-0.33	-0.33
X3	-0.33	0	-0.33	-0.33	0.333	0	-0.33	0.333	0.333	0	0	0	-0.33	0.333	-0.33
X4	0.333	-0.33	0	-0.33	-0.33	-0.33	0	-0.33	-0.33	0	-0.33	0.333	0	0	-0.33
X5	0.333	-0.33	-0.33	0	0.333	0.333	-0.33	0	-0.33	-0.33	0	-0.33	0	-0.33	0
X6	-0.33	0.333	-0.33	0.333	0	0.333	-0.33	-0.33	0	0.333	-0.33	0	-0.33	0	0

"별칭 행렬"은 모형에서 주효과 항의 계수가 잠재적 활성 상태인 2요인 교호작용 효과에 의해 어떻게 편향되는지 보여 줍니다. 열 라벨은 교호작용을 식별합니다. 예를 들어 X1 행에서 X2*X3 열의 값은 -0.33이고, X2*X4 열의 값은 0.333입니다. 즉, X1 주효과에 X2*X3 효과의 -0.33배를 더하고, X2*X4 효과의 0.333배를 더하는 식으로 X1 행의 남은 항까지 계산하여 X1 주효과의 기대값을 구할 수 있습니다. X1 주효과의 추정값이 유의하려면 X1 효과와 비교하여 이러한 교호작용의 크기를 무시할 수 있는 것으로 가정해야 합니다.

팁 : 별칭 행렬은 부분 요인 설계에서 교락 패턴을 일반화한 것입니다 .

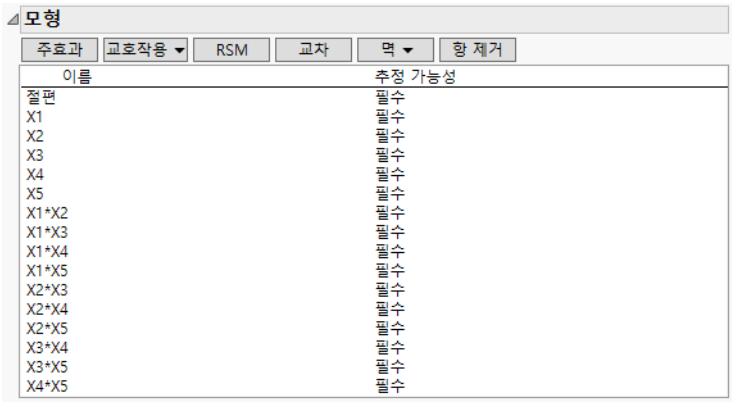
모든 2 요인 교호작용을 추정하는 설계

그림 5.5의 별칭 행렬은 효과의 부분 별칭 구조를 보여 줍니다. 다른 경우에는 주효과가 2요인 교호작용과 완전히 별칭 관계가 되거나 **교락**될 수 있습니다. 두 경우 모두 강력한 2요인 교호작용으로 인해 주효과만 추정하는 실험의 결과에 혼동을 줄 수 있습니다. 이러한 위험을 방지하기 위해 모든 2요인 교호작용을 해결하는 설계를 생성합니다.

이 예에서는 해상도 V 선별 설계를 생성합니다 . 2 요인 교호작용은 직교이지만 3 요인 교호작용과 교락됩니다 .

1. **DOE > 사용자 설계**를 선택합니다 .
2. **N 개 요인 추가** 옆에 5 를 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. **계속**을 클릭합니다 .
5. " 모형 " 개요에서 **교호작용 > 2 차**를 선택합니다 .

그림 5.6 교호작용을 보여 주는 모형 개요



모형

주효과 | 교호작용 | RSM | 교차 | 멱 | 항 제거

이름	추정 가능성
절편	필수
X1	필수
X2	필수
X3	필수
X4	필수
X5	필수
X1*X2	필수
X1*X3	필수
X1*X4	필수
X1*X5	필수
X2*X3	필수
X2*X4	필수
X2*X5	필수
X3*X4	필수
X3*X5	필수
X4*X5	필수

6. **최소값**을 클릭하여 런 수에 16 을 적용합니다 .

   참고 : 7 단계에서 " 난수 시드값 " 을 설정하고 8 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

7. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
8. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 1 을 입력하고 **확인**을 클릭합니다 .
9. **설계 생성**을 클릭합니다 .

   그림 5.7 에서는 설계의 런을 보여 줍니다 . " 모형 " 개요에서 모든 주효과와 2 요인 교호작용의 " 추정 가능성 " 이 " 필수 "(기본값) 로 지정되었으므로 효과를 추정할 수 있습니다 .

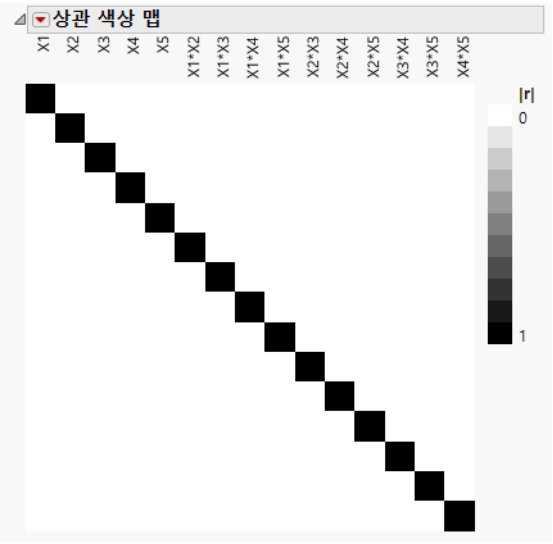
그림 5.7 모든 2 요인 교호작용을 추정하기 위한 설계

설계

런	X1	X2	X3	X4	X5
1	-1	-1	1	-1	-1
2	-1	-1	-1	1	-1
3	1	1	1	1	1
4	-1	-1	1	1	1
5	1	-1	1	-1	1
6	1	1	-1	1	-1
7	1	-1	-1	-1	-1
8	1	-1	1	1	-1
9	-1	1	1	-1	1
10	-1	-1	-1	-1	1
11	-1	1	-1	1	1
12	1	1	-1	-1	1
13	1	1	1	-1	-1
14	-1	1	1	1	-1
15	1	-1	-1	1	1
16	-1	1	-1	-1	-1

10. **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

그림 5.8 상관 색상 맵



색상 맵은 5 개의 주효과와 10 개의 이원 교호작용이 모두 상호 직교함을 나타냅니다.

주효과와 2 요인 교호작용의 별칭 관계를 방지하는 설계

6 개 연속형 요인의 주효과를 추정하는 것에 가장 관심이 있다고 가정해 보겠습니다. 그러나 주효과와 잠재적 활성 상태인 2 요인 교호작용의 별칭 관계를 최소화하는 방식을 사용하려고 합니다.

예산을 고려할 때 가능한 런 수는 16 회입니다. 요인이 6 개이고 가능한 2 요인 교호작용은 15 개입니다. 상수, 주효과 6 개 및 2 요인 교호작용 15 개를 적합시키는 데 필요한 최소 런 수는 22 입니다.

이 예에서는 주효과만 추정하는 8 회 런 설계 (" 주효과만추정하는 설계 "(125 페이지) 참조) 와 모든 2 요인 교호작용을 적합시킬 수 있는 22 회 런 설계의 절충안을 찾습니다 . 목표 달성을 위해 " 별칭 최적 " 을 최적 기준으로 사용합니다 .

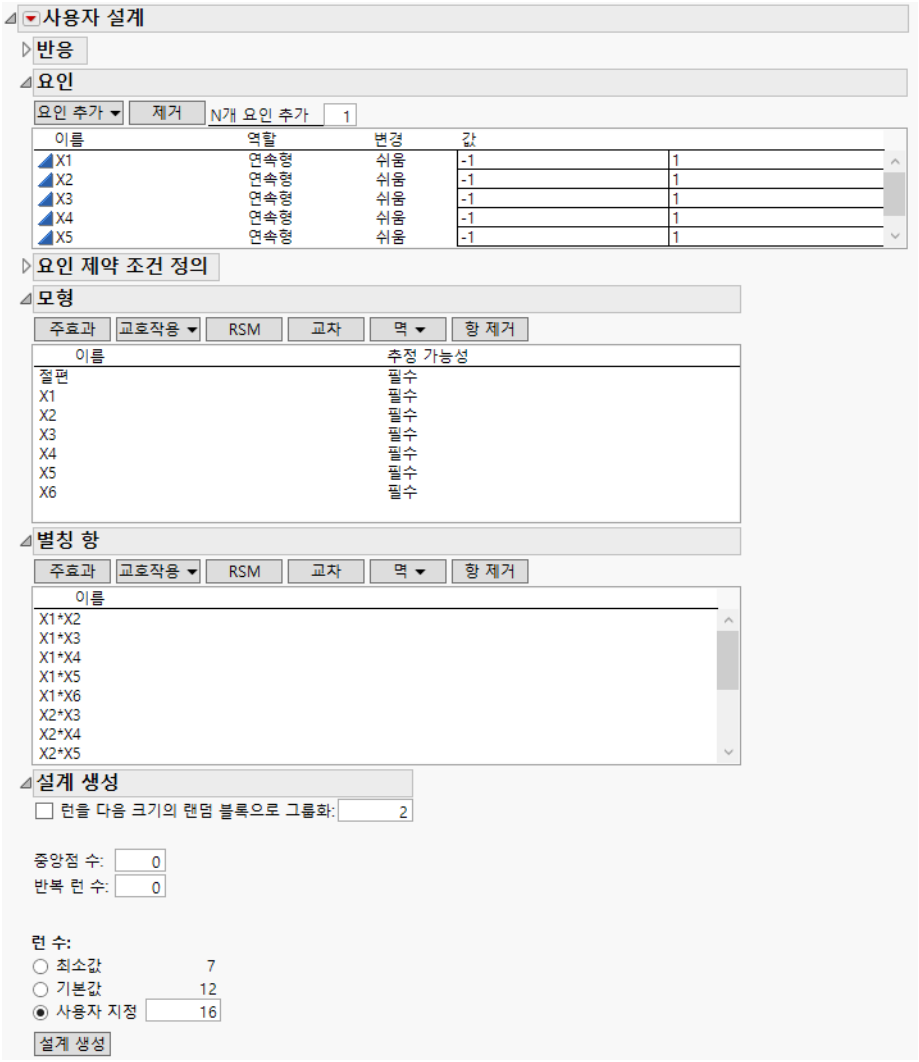
1. **DOE > 사용자 설계**를 선택합니다 .
2. **N 개 요인 추가** 옆에 6 을 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. **계속**을 클릭합니다 .

   기본적으로 모형에 주효과 항이 포함됩니다 . 이 항의 기본 추정 가능성은 " 필수 " 입니다 . " 별칭 항 " 개요에 2 차 교호작용이 추가되어 있습니다 . 기본적으로 가정된 모형에 포함되지 않은 모든 이원 교호작용은 " 별칭 항 " 목록에 추가됩니다 .
5. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **최적 기준 > 최적 설계 별칭 생성**을 선택합니다 .

" 최적 설계 별칭 생성 " 을 선택하면 JMP 에서 D 효율도를 사용하여 별칭 축소의 균형을 맞추는 설계를 생성합니다 . 자세한 내용은 " 사용자 설계 " 장의 " 별칭 최적 "(121 페이지) 에서 확인하십시오 .

6. **사용자 지정**을 클릭하고 런 수를 16 으로 변경합니다 .

그림 5.9 요인 , 모형 , 별칭 항 및 런 수



참고 : 7 단계에서 " 난수 시드값 " 을 설정하고 8 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

7. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345를 입력하고 **확인**을 클릭합니다.
8. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 10을 입력하고 **확인**을 클릭합니다.
9. **설계 생성**을 클릭합니다.
10. **설계 평가 > 별칭 행렬** 개요를 엽니다.

그림 5.10 별칭 행렬

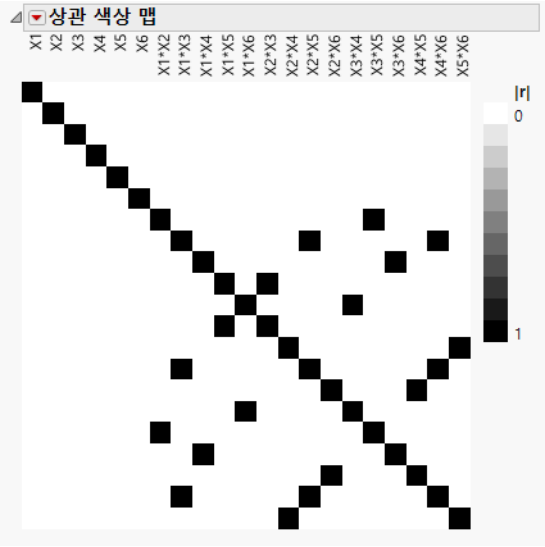
별칭 행렬

효과	X1*X2	X1*X3	X1*X4	X1*X5	X1*X6	X2*X3	X2*X4	X2*X5	X2*X6	X3*X4	X3*X5	X3*X6	X4*X5	X4*X6	X5*X6
절편	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

모든 행에 0 만 포함되어 있으며 이는 절편 및 주효과 항이 2 요인 교호작용에 의해 편향되지 않음을 의미합니다.

11. **설계 평가 > 상관 색상 맵** 개요를 엽니다.

그림 5.11 상관 색상 맵



"상관 색상 맵"에서는 주효과를 이원 교호작용과 독립적으로 추정할 수 있음을 보여 줍니다. 그러나 일부 이원 교호작용은 다른 이원 교호작용과 완전히 별칭 관계가 됩니다. 교락된 이원 교호작용을 보려면 비대각 빨간색 사각형을 커서로 가리킵니다.

이 특정 설계는 해상도 IV 직교 선별 설계입니다. 주효과는 다른 주효과 또는 이원 교호작용과 별칭 관계를 갖지 않습니다. 그러나 이원 교호작용은 다른 이원 교호작용과 완전히 별칭 관계가 됩니다.

과포화 선별 설계

일반적으로 브레인스토밍 세션에서 수십 개의 잠재적 활성 요인을 식별합니다. 데이터의 이점을 사용하지 않고 목록을 줄이는 대신 과포화 설계를 사용할 수 있습니다.

포화 설계에서는 런 수가 모형 항의 수와 같습니다. **과포화** 설계에서는 모형 항의 수가 런 수를 초과합니다(Lin, 1993). 과포화 설계는 요인 수의 절반보다 적은 런을 사용하여 수십 개의 요인을 검토할 수 있습니다. 이 설계는 요인이 많고 실험 런의 비용이 많이 드는 경우 요인 선별을 위한 유용한 선택입니다.

기존의 과포화 설계보다 우수한 활성 효과 식별을 위해 그룹 직교 과포화 설계를 사용할 수도 있습니다. 자세한 내용은 "그룹 직교 과포화 설계" 장(673 페이지)에서 확인하십시오.

과포화 설계의 제한

과포화 설계에는 다음과 같은 단점이 있습니다.

- **활성** 요인의 수가 실험 런 수의 절반보다 많으면 이러한 요인을 식별하지 못할 가능성이 있습니다. 일반적으로 런 수는 활성 요인 수보다 4 배 이상 많아야 합니다. 즉, 활성 요인이 최대 5 개 정도 있을 것으로 예상되는 경우 최소 20 회의 런을 계획해야 합니다.
- 아직은 과포화 설계 분석을 자동 프로시저로 축소할 수 없습니다. 그러나 단계별 전진 회귀를 사용하는 것은 적절합니다. 또한 선별 플랫폼 (**DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**) 은 간소화된 분석을 제공합니다.

과포화 설계 생성

이 예에서는 8 회 런을 수행하여 12 개 요인을 연구하기 위해 과포화 설계를 구성하려고 합니다. 과포화 설계를 생성하려면 절편을 제외한 모든 모형 항의 "추정 가능성"을 "If Possible"로 설정합니다.

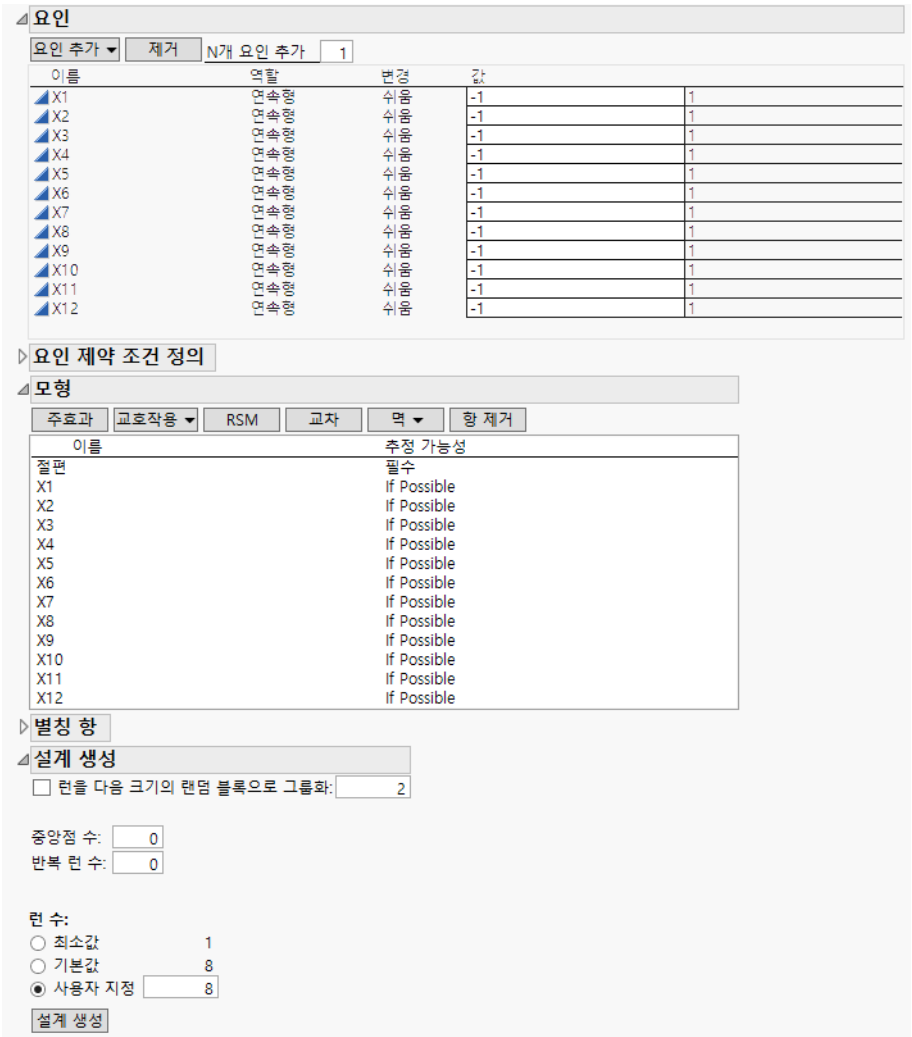
참고 : 이 예는 설명을 위한 것입니다. 과포화 설계의 런 수는 14 개 이상이어야 합니다. 활성 요인의 수가 4 개 이하인 경우 8 회 런 설계의 결과를 해석하기가 매우 어렵습니다. 자세한 내용은 " 과포화 설계의 제한 "(133 페이지) 에서 확인하십시오.

1. **DOE > 사용자 설계**를 선택합니다.
2. **N 개 요인 추가** 옆에 12 를 입력합니다.
3. **요인 추가 > 연속형**을 클릭합니다.
4. **계속**을 클릭합니다.
5. " 모형 " 개요에서 절편을 제외한 모든 항을 선택합니다.

6. 효과 옆의 필수를 클릭한 후 If Possible 로 변경합니다 .

   효과를 If Possible 로 설정하면 JMP 에서 베이지안 D- 최적 기준을 사용하여 설계를 구성합니다 .

그림 5.12 요인 , 모형 및 런 수



7. " 별칭 항 " 개요에서 모든 효과를 선택하고 **항 제거**를 클릭합니다 .

   이렇게 하면 " 상관 색상 맵 " 에 주효과만 나타납니다 . 이 그림은 설계가 생성되면 만들어집니다 .

8. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **반응 시뮬레이션**을 선택합니다 .

   이 옵션은 설계 테이블에 나타나는 확률 반응을 생성합니다 . 이 반응을 사용하여 실험 데이터 분석 방법을 확인할 수 있습니다 .

기본값 8 로 설정된 " 런 수 " 를 유지합니다 .

참고 : 9 단계에서 " 난수 시드값 " 을 설정하고 10 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 설계가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 . Y 시뮬레이션 값은 그림 5.13 에 나와 있는 값과 다를 수 있습니다 .

9. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
10. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 5 를 입력하고 **확인**을 클릭합니다 .
11. **설계 생성**을 클릭합니다 .
12. **테이블 생성**을 클릭합니다 .

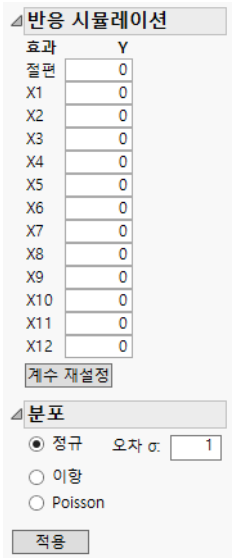
" 사용자 설계 " 창을 닫지 마십시오 . 이 예에서 나중에 다시 여기로 돌아갑니다 .

그림 5.13 시뮬레이션된 반응을 포함하는 설계 테이블

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Y	Y 시뮬레이션
1	1	-1	1	-1	-1	1	-1	-1	-1	1	1	-1	0.836971941	0.83697194
2	1	-1	-1	1	-1	-1	-1	1	1	1	-1	-1	0.0705123242	0.07051232
3	-1	1	-1	-1	1	-1	-1	-1	1	1	1	1	1.8362934786	1.83629348
4	1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	-0.287280508	-0.2872805
5	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	1.0427922172	1.04279222
6	-1	-1	1	1	1	-1	1	-1	-1	1	-1	-1	-1.212173247	-1.2121732
7	1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	-1.191896905	-1.1918969
8	1	1	-1	1	-1	-1	1	-1	-1	-1	1	1	1.1431424982	1.1431425

반응 열 Y 와 Y 시뮬레이션은 동일한 시뮬레이션 값으로 시작됩니다 . 이 값은 *N(0, s)* 분포의 난수 값입니다 . 여기서 *s* 는 검정력 분석 대화상자의 RMSE(기본값 1) 입니다 . Y 시뮬레이션 값은 " 반응 시뮬레이션 " 창의 모수 값으로 정의된 모형을 사용하여 무작위로 생성된 값으로 업데이트됩니다 . Y 열은 실험 실행 후의 실제 반응을 위한 것입니다 .

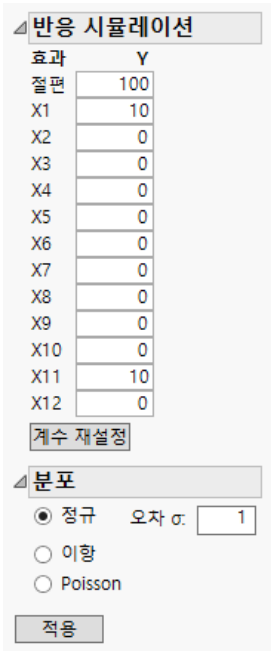
그림 5.14 반응 시뮬레이션 창



" 반응 시뮬레이션 " 창에는 모든 항에 대한 기본 계수 0, " 정규 " 분포 선택 및 " 오차 σ" 1 이 표시됩니다 . 현재 Y 및 Y 시뮬레이션 열의 값은 랜덤 변동만 반영합니다 .

13. " 반응 시뮬레이션 " 창의 계수 값을 그림 5.15 에 표시된 것과 같이 변경합니다 .

그림 5.15 시뮬레이션된 반응의 모수 값



14. **적용**을 클릭합니다 .

Y 시뮬레이션 열의 반응 값이 변경됩니다.

참고 : 반응 값은 무작위로 생성됩니다 . 따라서 그림 5.16 의 값과 정확하게 일치하지 않습니다 .

그림 5.16 X1 과 X11 이 활성화된 Y 시뮬레이션 반응 열

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Y	Y 시뮬레이션
1	1	-1	1	-1	-1	1	-1	-1	-1	1	1	-1	-0.880036237	-0.8800362
2	1	-1	-1	1	-1	-1	-1	1	1	1	-1	-1	-0.281221702	-0.2812217
3	-1	1	-1	-1	1	-1	-1	-1	1	1	1	1	-0.35913637	-0.3591364
4	1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	-0.489512232	-0.4895122
5	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	0.3489014857	0.34890149
6	-1	-1	1	1	1	-1	1	-1	-1	1	-1	-1	-1.341785751	-1.3417858
7	1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	-0.736235432	-0.7362354
8	1	1	-1	1	-1	-1	1	-1	-1	-1	1	1	-1.018953519	-1.0189535

시뮬레이션에서 X1 과 X11 을 오차 변동에 비례하여 효과가 큰 활성 요인으로 지정했습니다. 따라서 데이터 분석 결과 이러한 두 요인이 활성 요인으로 식별되어야 합니다.

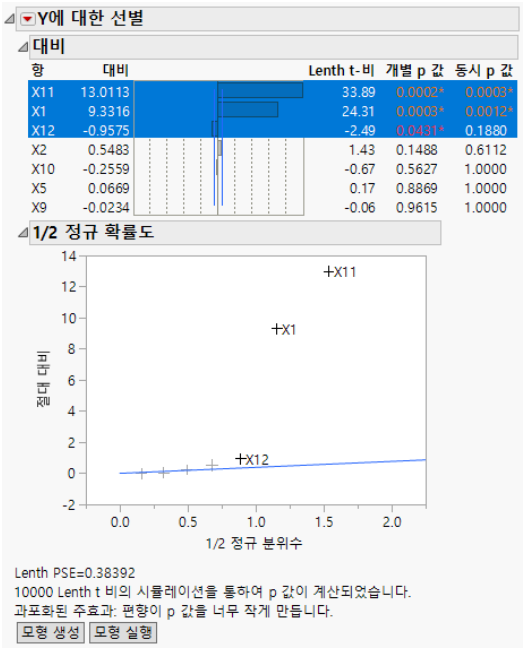
선별 플랫폼을 사용하여 과포화 설계 분석

선별 플랫폼은 활성 요인을 식별하는 방법을 제공합니다 . 선별 플랫폼을 사용하여 설계 테이블의 Y 시뮬레이션 값을 분석합니다 (그림 5.16). 선별 플랫폼은 **DOE > 전통적 설계** 메뉴에 있습니다 .

참고 : 실제 데이터 테이블에는 이 예에 나와 있는 Y 시뮬레이션 값과 약간 다른 값이 나타납니다 . 실제 보고서에 나타나는 정확한 값은 여기에 표시된 결과와는 다릅니다 .

1. **DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**을 선택합니다 .
2. Y 시뮬레이션을 선택하고 **Y** 를 클릭합니다 .
3. X1 ~ X12 를 선택하고 **X** 를 클릭한 후 **확인**을 클릭합니다 .

그림 5.17 과포화 설계에 대한 선별 보고서



X1 요인과 X11 요인의 " 대비 " 및 "Lenth t- 비 " 값이 큽니다 . 또한 이 요인의 " 동시 p 값 " 은 작습니다 . "1/2 정규 확률도 " 에서 X1 과 X11 은 둘 다 선에서 멀리 떨어져 있습니다 . " 대비 " 및 "1/2 정규 확률도 " 보고서는 X1 과 X11 이 활성 요인임을 나타냅니다 . X12 는 " 개별 p 값 " 이 0.05 미만이지만 X1 및 X11 과 비교하여 효과가 훨씬 작습니다 .

과포화된 설계이므로 모든 효과를 추정할 수 있는 모형의 경우보다 *p* 값이 작을 수 있습니다 . 효과 추정값이 다른 잠재적 활성 주효과에 의해 편향되기 때문입니다 . 그림 5.17 에서 " 모형 생성 " 버튼 바로 위에 이 가능성에 대한 경고 메시지가 있습니다 .

활성으로 보이는 효과가 다른 효과와 높은 상관관계가 있는지 여부를 확인할 수도 있습니다 . 상관관계가 높으면 한 효과가 다른 효과의 참유의성을 감출 수 있습니다 . 그림 5.19 의 색상 맵은 효과 사이의 절대 상관을 표시합니다 .

4. **모형 생성**을 클릭합니다 .

   생성된 모형에는 X1, X11 및 X12 효과만 포함됩니다 .

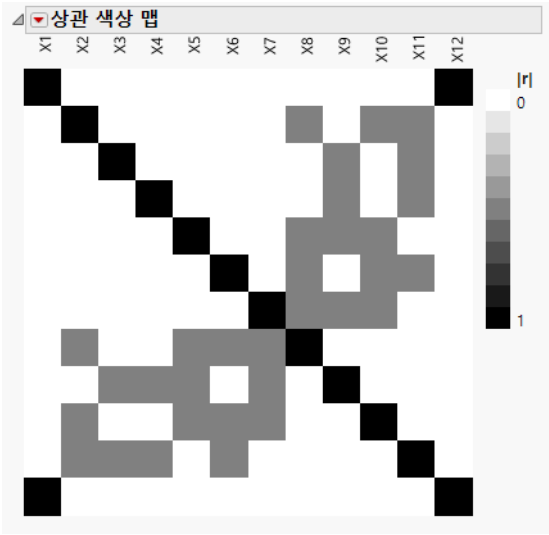
5. " 모형 규격 " 창에서 **실행**을 클릭합니다 .

그림 5.18 모형의 모수 추정값

X11 및 X1 의 모수 추정값은 모형을 시뮬레이션할 때 사용한 이론적 값에 가깝습니다 . 자세한 내용은 그림 5.15 에서 확인하십시오 . 여기서 X1 = 10 과 X11 = 10 을 사용하여 모형을 지정했습니다 . X12 요인의 유의성은 False Positive 의 예입니다 .

6. " 사용자 설계 " 창에서 **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

그림 5.19 상관 색상 맵 개요



절대 상관을 표시하려면 커서로 셀을 가리킵니다 . X1 은 다른 주효과 (X4, X5, X7) 와 최대 0.5 의 상관관계가 있습니다 . 그림 5.19 에는 JMP 기본 색상이 사용됩니다 .

단계별 회귀를 사용하여 과포화 설계 분석

단계별 회귀는 활성 요인을 식별하는 또 다른 방법입니다 . 그림 5.16 의 설계 테이블에는 3 개의 스크립트가 포함되어 있습니다 . " 모형 " 스크립트는 모형 적합 플랫폼에서 단계별 회귀를 사용하여 데이터를 분석합니다 .

참고 : 실제 데이터 테이블에는 이 예에 나와 있는 Y 시뮬레이션 값과 약간 다른 값이 나타납니다 . 실제 보고서에 나타나는 정확한 값은 여기에 표시된 결과와는 다릅니다 .

---

1. 설계 테이블의 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
2. **분석법**을 **표준 최소 제곱**에서 **단계별**로 변경합니다 .
3. **실행**을 클릭합니다 .
4. "Y 에 대한 단계별 회귀 적합 " 보고서에서 **중지 규칙**을 **최소 AICc** 로 변경합니다 .

   설계된 실험의 경우 BIC 중지 규칙은 모형에 비활성 효과를 허용하는 경향이 있으므로 대개 AICc 보다 관대합니다 .
5. **시작**을 클릭합니다 .

그림 5.20 과포화 설계에 대한 단계별 회귀

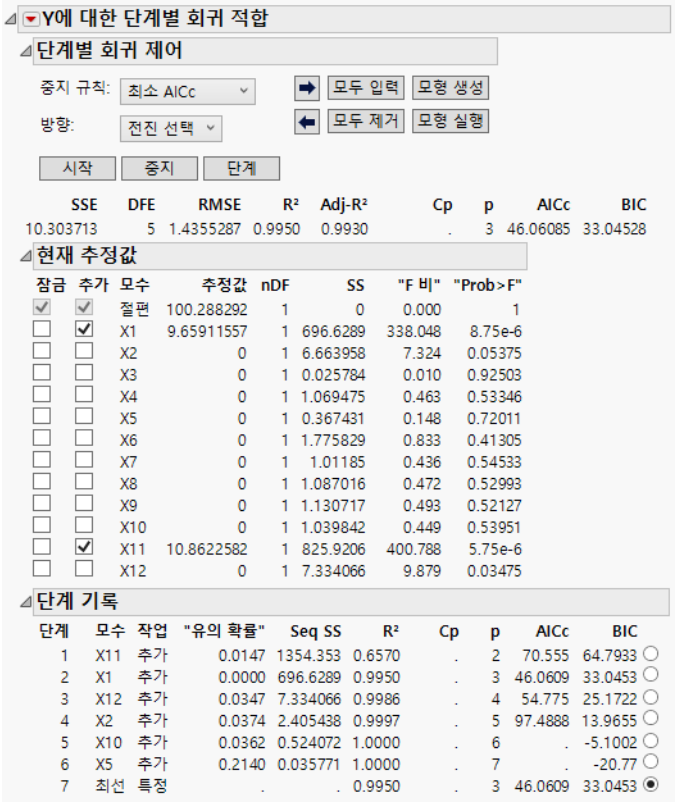


그림 5.20 에서는 선택된 모형이 두 가지 활성 요인 (X1, X11) 으로 구성되어 있음을 보여 줍니다 . 보고서 아래쪽에 단계 기록이 나타납니다 . X1 및 X11 과 다른 요인과의 상관관계는 다른 활성 요인의 효과를 감출 수 있다는 점에 유의하십시오 (그림 5.19).

참고 : 이 예에서는 큰 주효과를 두 개 정의하고 나머지는 0 으로 설정합니다 . 실제 상황에서는 이와 같이 뚜렷하게 차별화되는 효과를 가질 가능성이 낮을 수 있습니다 .

고정 블록 설계

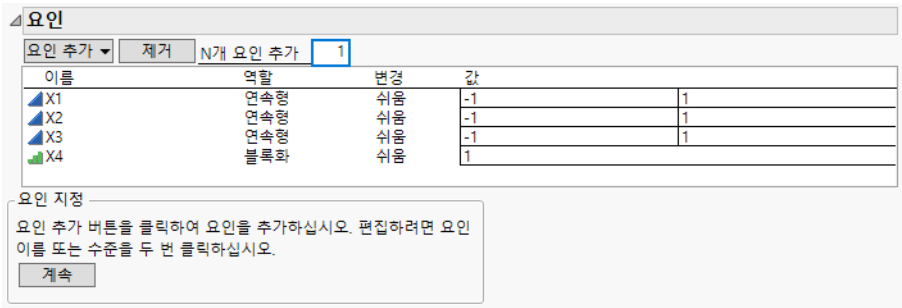
기존 선별 설계에서는 블록 크기가 2 의 거듭제곱이어야 합니다. 그러나 사용자 설계 플랫폼은 원하는 크기의 고정 블록으로 설계를 생성할 수 있습니다.

세 가지 요인을 연구하려고 한다고 가정해 보겠습니다. 하루에 세 번만 실험을 실행할 수 있고 일별 변동이 상당할 것으로 예상됩니다. 결과적으로 여러 날 동안의 설계를 블록화해야 합니다. 또한 이 연구에서 모든 2 요인 교호작용을 추정하는 데 관심이 있습니다 . 이 예에서는 블록당 3 회 런을 사용하여 설계를 구성합니다 .

1. **DOE > 사용자 설계**를 선택합니다 .

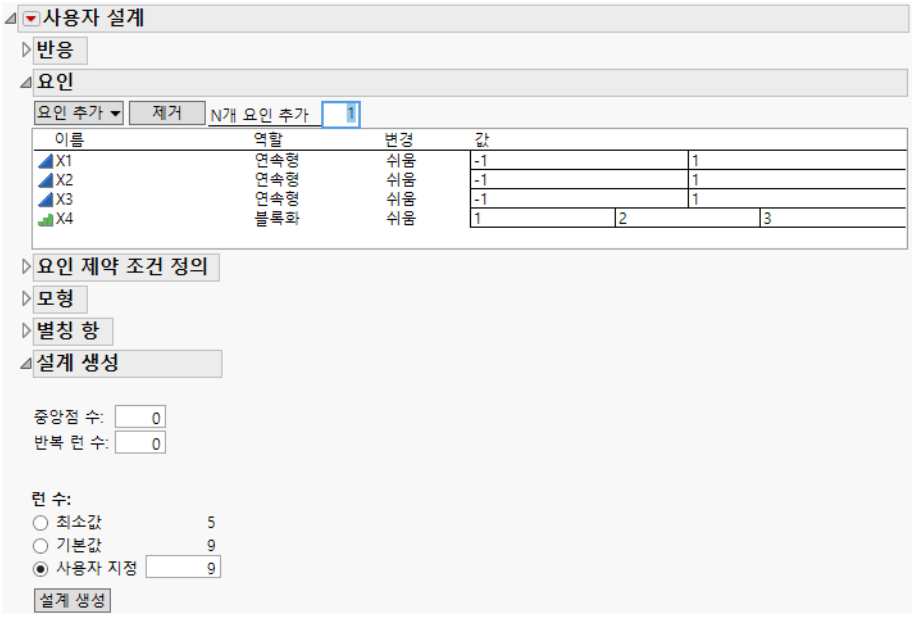
2. " 요인 " 개요에서 **N 개 요인 추가** 옆에 3 을 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. **요인 추가 > 블록화 > 블록당 3 회 런**을 클릭합니다 .
   블록 요인 X4의 값 아래에 한 수준만 표시됩니다. 여기서는 런 크기를 알 수 없기 때문입니다.

그림 5.21 X4 에 대한 블록 하나를 보여 주는 요인 개요



5. **계속**을 클릭합니다 .

그림 5.22 X4 에 대한 블록 3 개를 보여 주는 요인 개요

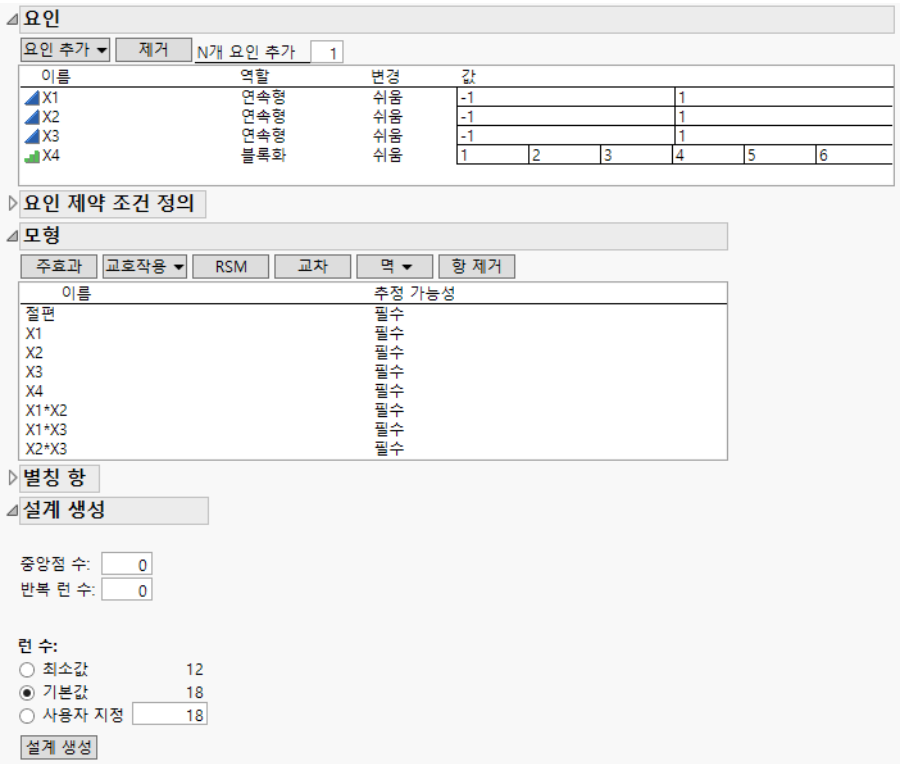


이제 " 요인 " 개요에 적절한 블록 수가 표시됩니다 . 이 값은 기본 런 크기를 블록당 런 수로 나누어 계산됩니다 . 이 예의 경우 기본 표본 크기 9 에 3 개의 블록이 필요합니다 . 이제 " 요인 " 개요에는 X4 에 대한 3 개의 값이 표시되어 블록이 3 개임을 나타냅니다 .

참고 : 다른 런 수를 지정하면 " 요인 " 개요가 업데이트되어 블록 요인에 대해 적절한 수의 값이 표시됩니다 .

6. " 요인 " 개요에서 연속형 요인 X1, X2, X3 을 선택합니다 .
7. " 모형 " 개요에서 **교호작용 > 2 차**를 클릭합니다 .

그림 5.23 X4 에 대한 블록 6 개를 보여 주는 요인 개요



이제 " 런 수 " 패널에 기본 런 크기가 18 로 표시됩니다 . " 요인 " 개요에는 X4 에 대한 6 개의 값이 표시되어 블록이 6 개임을 나타냅니다 .

참고 : 8 단계에서 " 난수 시드값 " 을 설정하고 9 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

8. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
9. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 5 를 입력하고 **확인**을 클릭합니다 .
10. **설계 생성**을 클릭합니다 .

그림 5.24 고정 블록 설계

설계

런	X1	X2	X3	X4
1	1	-1	-1	1
2	-1	-1	-1	2
3	1	1	-1	3
4	-1	-1	-1	4
5	1	1	1	5
6	-1	-1	1	6
7	-1	1	-1	1
8	1	1	-1	2
9	-1	1	1	3
10	1	-1	1	4
11	-1	-1	1	5
12	-1	1	-1	6
13	-1	-1	1	1
14	-1	1	1	2
15	1	-1	1	3
16	-1	1	1	4
17	1	-1	-1	5
18	1	1	1	6

설계에서 블록 요인 X4 를 살펴보십시오 . 6 개 블록이 표시됩니다 . 실험을 수행할 때 매일 세 번씩 시도되며 , 첫째 날의 X4 = 1, 둘째 날의 X4 = 2 등으로 지정됩니다 . 따라서 설계 테이블에서 블록 내의 시도를 랜덤화하려고 합니다 . " 출력 옵션 " 패널에서 " 런 순서 " 에 대해 " 블록 내 랜덤화 " 옵션이 이미 선택되어 있습니다 .

11. **테이블 생성**을 클릭합니다 .

그림 5.25 고정 블록 설계에 대한 설계 테이블

	X1	X2	X3	X4	Y
1	1	-1	-1	1	•
2	-1	-1	1	1	•
3	-1	1	-1	1	•
4	-1	1	1	2	•
5	1	1	-1	2	•
6	-1	-1	-1	2	•
7	-1	1	1	3	•
8	1	-1	1	3	•
9	1	1	-1	3	•
10	-1	1	1	4	•
11	1	-1	1	4	•
12	-1	-1	-1	4	•
13	-1	-1	1	5	•
14	1	1	1	5	•
15	1	-1	-1	5	•
16	-1	-1	1	6	•
17	1	1	1	6	•
18	-1	1	-1	6	•

설계 테이블의 행은 하루에 시도되는 런으로 그룹화됩니다. 이 설계를 사용하면 블록 효과, 모든 주효과 및 2 요인 교호작용을 추정할 수 있습니다.

반응 표면 실험

반응 표면 실험은 대개 활성 요인으로 식별된 소수(일반적으로 2 ~ 8 개)의 연속형 요인을 포함합니다. 반응 표면 실험의 주 목표는 요인과 반응 간의 관계에 대한 예측 모형을 개발하는 것입니다. 종종 이 예측 모형을 사용하여 공정에 대해 더 나은 작동 설정을 찾습니다. 따라서 반응 표면 실험의 가정된 모형은 일반적으로 2 차입니다.

선별 설계는 활성 효과를 식별하는 데 중점을 두기 때문에 품질을 결정하는 측도는 계수의 상대 분산 크기입니다. 이러한 상대 분산을 작게 하려고 합니다. D- 최적으로 상대 분산을 해결할 수 있습니다.

반응 표면 실험에서는 요인 범위에 대한 예측 분산이 계수 분산보다 중요합니다. 설계 공간에 대한 예측 분산은 I- 최적으로 해결됩니다. I- 최적 설계는 D- 최적 설계와 비교하여 설계 공간의 극단에 더 적은 런을 배치하는 경향이 있습니다. D- 최적 및 I- 최적에 대한 자세한 내용은 "사용자 설계" 장의 "최적 기준"(118 페이지)에서 확인하십시오.

기본적으로 사용자 설계는 " 최적 기준 " 에 대해 " 권장 사항 " 옵션을 사용합니다 . 사용자 설계는 " 모형 " 개요에서 "RSM" 버튼을 사용하여 2 차 효과를 추가할 때마다 I- 최적 기준을 " 권장 사항 " 기준으로 사용합니다 . 그렇지 않은 경우에는 D- 최적 기준을 " 권장 사항 " 기준으로 사용합니다 . 자세한 내용은 " 사용자 설계 " 장의 " 최적 기준 "(118 페이지) 에서 확인하십시오 .

- " 반응 표면 설계 "
- " 유연한 블록화 방법을 사용하는 반응 표면 설계 "
- "D- 최적과 I- 최적의 반응 표면 설계 비교 "
- " 제약 조건 및 범주형 요인이 있는 반응 표면 설계 "

반응 표면 설계

이 예에는 다음과 같은 섹션이 포함되어 있습니다 .

- " 반응 표면 설계 구성 "(144 페이지)
- " 실험 결과 분석 "(147 페이지)

반응 표면 설계 구성

활성 요인으로 식별된 3 개의 연속형 요인에 대한 반응 표면 설계를 구성합니다 . 규격 내에서 반응 (Y) 을 유지하기 위한 공정 설정을 찾으려고 합니다 . Y 에 대한 규격 하한 및 상한은 각각 54, 56 입니다 (목표값 55).

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 반응 " 개요에서 **최대화**를 클릭하고 **목표값 일치**를 선택합니다 .
3. **하한**에 54 를 입력하고 **상한**에 56 을 입력합니다 .
4. **중요도**는 비워 둡니다 .

반응이 하나뿐이므로 중요도 값은 기본적으로 1 로 설정됩니다 .

5. **N 개 요인 추가** 옆에 3 을 입력합니다 .
6. **요인 추가 > 연속형**을 클릭합니다 .

   X1, X2, X3 이라는 3 개의 연속형 요인이 추가됩니다 .
7. **계속**을 클릭합니다 .
8. " 모형 " 개요에서 **RSM** 버튼을 클릭합니다 .

   2 차 항과 교호작용 항이 모형에 추가됩니다 . 또한 " 권장 사항 " 최적 기준의 값이 I- 최적으로 설정됩니다 . " 설계 생성 " 을 클릭한 후 " 설계 진단 " 개요에서 이를 확인할 수 있습니다 . 16 으로 설정된 기본값 런 수를 그대로 둡니다 .

   참고 : 9 단계에서 " 난수 시드값 " 을 설정하고 10 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .
9. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 929281409 를 입력하고 **확인**을 클릭합니다 .
10. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 40 을 입력하고 **확인**을 클릭합니다 .
11. **설계 생성**을 클릭합니다 .

    " 설계 " 개요에 설계가 표시됩니다 .

그림 5.26 RSM 설계

설계

런	X1	X2	X3
1	0	-1	0
2	1	-1	1
3	0	0	0
4	-1	1	1
5	-1	0	0
6	1	1	1
7	-1	1	-1
8	-1	-1	-1
9	0	0	1
10	1	0	0
11	0	1	0
12	0	0	0
13	1	1	-1
14	0	0	-1
15	-1	-1	1
16	1	-1	-1

2 차 효과 추정을 위해 반응 표면 설계는 각 요인에 대해 3 개의 수준을 사용합니다 . 그림 5.26 의 설계는 중앙점이 두 개 있는 면심 중심 합성 설계입니다 .

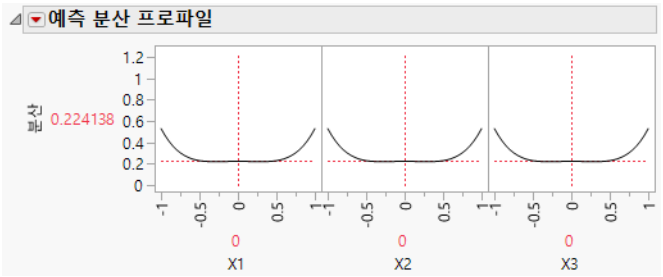
12. **설계 평가 > 설계 진단** 개요를 엽니다 .

그림 5.27 설계 진단 개요

"설계 진단" 개요의 첫 번째 행은 사용 중인 최적 기준을 나타냅니다. 이 설계는 I-최적입니다.

13. **설계 평가 > 예측 분산 프로파일** 개요를 엽니다.

그림 5.28 예측 분산 프로파일

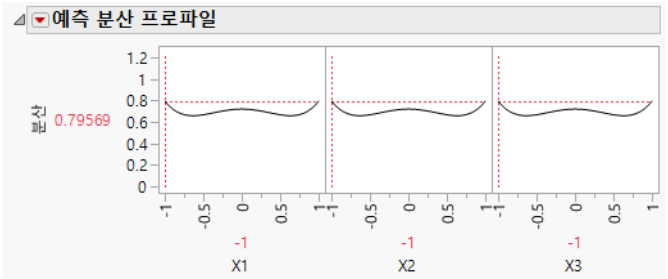


세로 축은 반응 기대값의 상대 예측 분산을 보여 줍니다. **상대 예측 분산**은 예측 분산을 오차 분산으로 나눈 값입니다. 상대 예측 분산이 1 인 경우 절대 분산은 회귀 모형의 오차 분산과 같습니다.

프로파일러는 설계 공간에 대한 상대 예측 분산 값을 보여 줍니다. 슬라이더를 움직여 예측 분산의 동작을 탐색할 수 있습니다. 예측 분산은 설계 공간의 중심에서 가장 작습니다. 이 값은 -0.5 와 0.5 사이의 요인 설정에 대해 0.2 보다 조금 큰 값으로 비교적 일정하게 유지됩니다. 설정이 설계 공간 경계에 근접할수록 예측 분산이 증가합니다.

14. " 예측 분산 프로파일 " 의 빨간색 삼각형을 클릭하고 **분산 최대화**를 선택합니다.

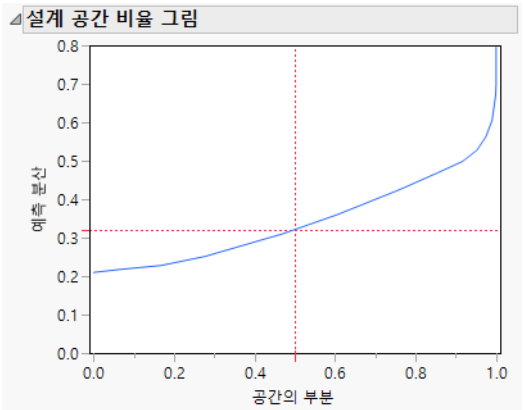
그림 5.29 상대 분산이 최대화된 예측 분산 프로파일



프로파일러는 상대 예측 분산의 최대값이 0.79569 임을 보여 줍니다.

15. **설계 평가 > 설계 공간 비율 그림** 개요를 엽니다.

그림 5.30 설계 공간 비율 그림



그림에서 파란색 곡선은 상대 예측 분산을 설계 공간 비율 함수로 표시한 것입니다. 빨간색 십자 기호 파선은 설계 공간의 50%에 대해 예측 분산이 약 0.32 이하임을 나타냅니다. 십자기호 도구를 사용하여 다른 추론을 이끌어 낼 수 있습니다. 예를 들어 "공간 비율"이 0.95일 때 예측 분산은 약 0.52입니다. 이는 설계 공간의 95%에 대해 상대 예측 분산이 0.52 미만임을 의미합니다.

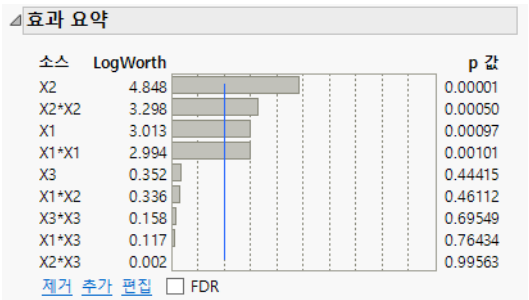
실험 결과 분석

Custom RSM.jmp 샘플 데이터 테이블에는 실험 결과가 포함되어 있습니다 . "Model" 스크립트는 DOE 창의 " 모형 " 개요에서 지정한 모든 효과를 보여 주는 " 모형 적합 " 창을 엽니다 . 이 스크립트는 사용자 설계 플랫폼에서 데이터 테이블에 저장되었습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Custom RSM.jmp를 엽니다.
2. 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
3. **실행**을 클릭합니다 .

   " 효과 요약 " 보고서에는 모형의 각 효과에 대한 LogWorth 및 p 값이 표시됩니다 . 그림의 파란색 수직선은 값 2 에서 설정됩니다 . 2 를 초과하는 LogWorth 는 0.01 수준에서 유의합니다 .

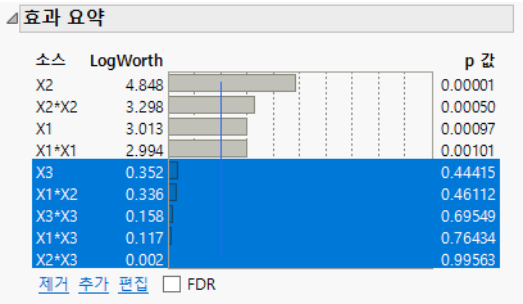
그림 5.31 효과 요약 보고서



이 보고서는 X1, X2, X1*X1 및 X2*X2 가 0.01 수준에서 유의함을 보여 줍니다 . 다른 효과는 0.10 수준에서도 유의하지 않습니다 . 유의하지 않은 이러한 효과를 제거하여 모형을 축소합니다 .

4. " 효과 요약 " 보고서에서 X3, X1*X2, X3*X3, X1*X3 및 X2*X3 을 선택합니다 .

그림 5.32 유의하지 않은 효과가 선택된 효과 요약 보고서



5. **제거**를 클릭합니다 .

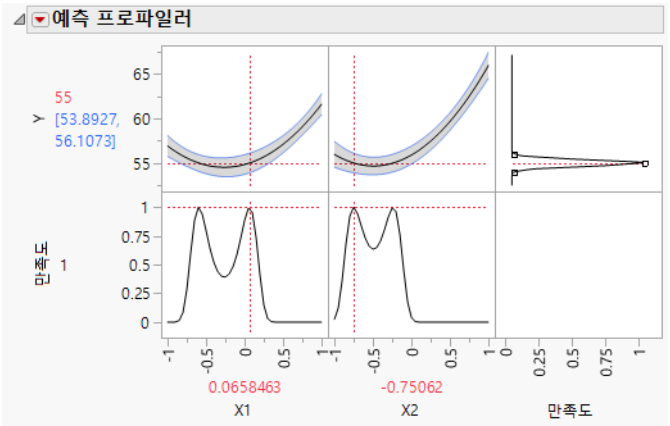
" 최소 제곱 적합 " 보고서가 업데이트되어 유의한 효과 (X1, X2, X1*X1 X2*X2) 만 포함하는 모형이 표시됩니다 .

" 최소 제곱 적합 " 창의 아래쪽에 있는 " 예측 프로파일러 " 를 사용하여 X1 및 X2 요인이 달라지면 예측 반응 (Y) 이 어떻게 변경되는지 확인합니다 . X1 및 X2 의 전체 값에서 Y 의 2 차 동작에 유의하십시오 .

" 사용자 설계 " 창의 " 반응 " 개요에서 Y 에 대한 반응 한계를 입력했습니다 . 따라서 " 반응 한계 " 열 특성이 설계 테이블의 Y 열에 연결됩니다 . Y 의 만족도 함수 (오른쪽 상단 그림) 는 " 반응 한계 " 열 특성에 포함된 정보를 기반으로 합니다 . JMP 는 이 함수를 사용하여 X1 및 X2 설정의 함수로 만족도를 계산합니다 . 만족도 함수의 트레이스는 그림의 아래쪽 행에 나타납니다 .

6. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

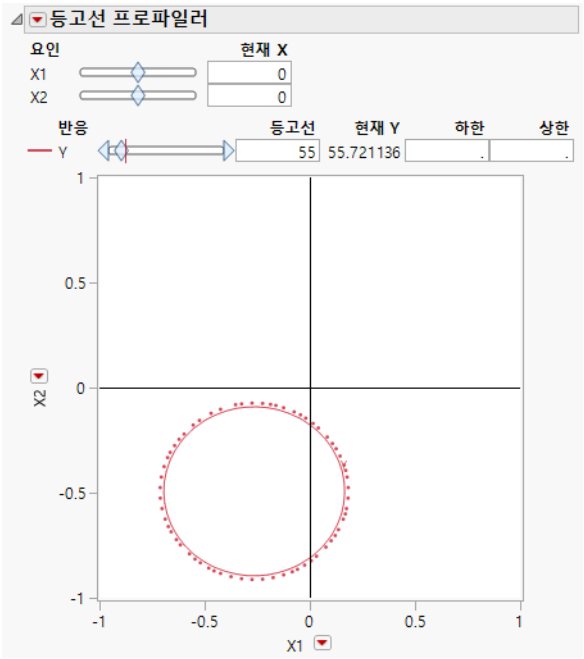
그림 5.33 만족도가 최대화된 예측 프로파일러



예측 반응은 X1 및 X2 위에 빨간색으로 표시된 공정 설정에서 목표값 55 를 달성합니다 . 그림 5.33 에서는 X2 = -0.75062 일 때 –0.65 에 가까운 X1 값도 예측값 55 를 달성한다는 것을 보여 줍니다 . 사실상 예측 프로파일러에 표시된 설정과 만족도를 최대화하는 설정이 다를 수 있습니다 . X1 과 X2 의 많은 설정에 대한 예측 반응이 55 이기 때문입니다 .

7. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 등고선 프로파일러**를 선택합니다 .
8. " 등고선 프로파일러 " 보고서에서 **등고선** 값에 55 를 입력합니다 .

그림 5.34 등고선 프로파일러



빨간색 등고선에 해당하는 X1 및 X2 설정의 예측 반응 값은 55 입니다 . 비용 효율성에 따라 이러한 공정 설정 중에서 선택할 수 있습니다 .

유연한 블록화 방법을 사용하는 반응 표면 설계

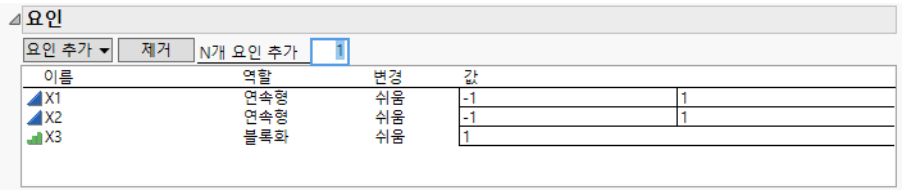
공정을 최적화할 때 연속형 요인뿐 아니라 질적 요인도 실험에 포함해야 할 수 있습니다. 배치 또는 날짜와 같은 질적 요인으로 블록화하거나, 기계 또는 배송 메커니즘과 같은 질적 요인을 포함해야 합니다. 그러나 반응 표면 설계 플랫폼은 연속형 요인만 지원합니다. 질적 요인을 포함하는 반응 표면 설계를 생성하려면 각 요인 수준에 대해 설계를 반복하면 됩니다. 그러나 이 방법은 비효율적입니다. 사용자 설계 플랫폼은 더 적은 수의 런으로 최적 설계를 구성합니다.

이 예에서는 두 개의 연속형 요인과 블록당 4 회 런이 포함된 블록 요인을 수용하는 반응 표면 설계를 구성합니다 . 비슷한 방식으로 범주형 또는 이산 수치형 요인을 포함할 수 있습니다 .

1. **DOE > 사용자 설계**를 선택합니다 .
2. **N 개 요인 추가** 옆에 2 를 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. **요인 추가 > 블록화 > 블록당 4 회 런**을 클릭합니다 .

   " 값 " 아래에 한 수준만 나타납니다 . 런 수가 지정되지 않으면 블록 수를 결정할 수 없기 때문입니다 .

그림 5.35 연속형 요인 두 개와 블록 요인이 추가된 요인 개요



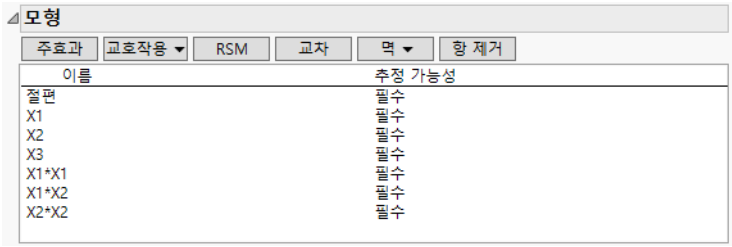
5. **계속**을 클릭합니다 .

   기본값 런 수는 12 입니다 . " 요인 " 개요가 업데이트되어 블록 요인 X3 에 대해 세 개의 수준이 표시됩니다 . X3 에 블록당 4 회 런이 필요하므로 12 회 런에는 세 개의 블록이 허용됩니다 .
6. **RSM** 을 클릭합니다 .

   블록 요인은 교호작용 또는 2 차 항에 나타날 수 없음을 알려 주는 JMP 경고 창이 표시됩니다 . JMP 는 적절한 RSM 항만 목록에 추가합니다 .
7. **확인**을 클릭하여 메시지를 해제합니다 .

   X1 과 X2 의 2 차 및 교호작용 항이 모형에 추가됩니다 . RSM 항을 추가했으므로 " 권장 사항 " 최적 기준이 "D- 최적 " 에서 "I- 최적 " 으로 변경됩니다 . 나중에 " 설계 진단 " 개요에서 이를 확인할 수 있습니다 .

그림 5.36 반응 표면 효과가 추가된 모형 개요



> **참고 :** 8 단계에서 " 난수 시드값 " 을 설정하고 9 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

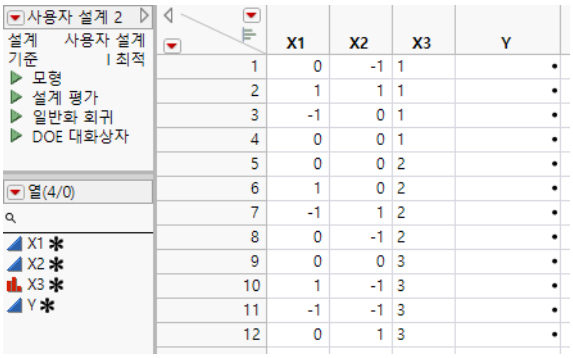
8. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
9. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 5 를 입력하고 **확인**을 클릭합니다 .
10. **설계 생성**을 클릭합니다 .
11. **설계 평가 > 설계 진단** 개요를 엽니다 .

그림 5.37 설계 진단 개요

"설계 진단" 개요의 첫 번째 행은 사용 중인 최적 기준을 나타냅니다. 이 설계는 I-최적입니다.

12. **테이블 생성**을 클릭합니다 .

그림 5.38 블록 요인이 포함된 설계 테이블

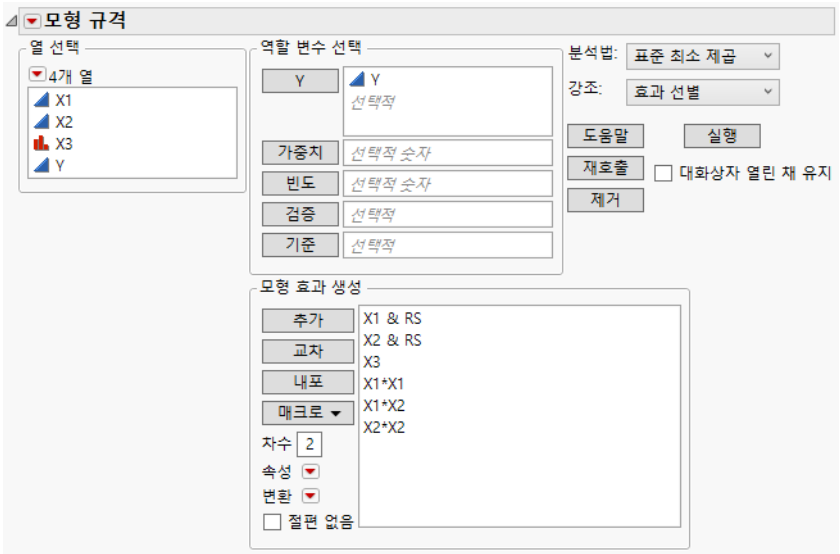


	X1	X2	X3	Y
1	0	-1	1	•
2	1	1	1	•
3	-1	0	1	•
4	0	0	1	•
5	0	0	2	•
6	1	0	2	•
7	-1	1	2	•
8	0	-1	2	•
9	0	0	3	•
10	1	-1	3	•
11	-1	-1	3	•
12	0	1	3	•

기본 " 런 순서 " 를 " 블록 내 랜덤화 " 로 지정했으므로 블록 요인 (X3) 수준이 정렬됩니다 .

13. 설계 테이블의 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

그림 5.39 모형 적합 창



다음을 확인할 수 있습니다 .

- 블록 요인 (X3) 이 효과로 입력되었습니다 .
- X3 과 관련된 교호작용이 포함되지 않았습니다 .
- 나머지 5 개 효과는 X1 과 X2 에 대한 반응 표면 모형을 정의합니다 .

D- 최적과 I- 최적의 반응 표면 설계 비교

이 예를 통해 2 요인 반응 표면 설계에서 I- 최적과 D- 최적 간 차이에 대해 알아보겠습니다 .

I- 최적 설계

1. **DOE > 사용자 설계**를 선택합니다 .
2. **N 개 요인 추가** 옆에 2 를 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. **계속**을 클릭합니다 .
5. **RSM** 을 클릭합니다 .

   X1 과 X2 의 2 차 및 교호작용 항이 모형에 추가됩니다 . RSM 항을 추가했으므로 " 권장 사항 " 최적 기준이 "D- 최적 " 에서 "I- 최적 " 으로 변경됩니다 . 나중에 " 설계 진단 " 개요에서 이를 확인할 수 있습니다 .

   참고 : 6 단계에서 " 난수 시드값 " 을 설정하고 7 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

6. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 383570403 을 입력하고 **확인**을 클릭합니다 .
7. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 8 을 입력하고 **확인**을 클릭합니다 .
8. **설계 생성**을 클릭합니다 .

그림 5.40 I- 최적 설계

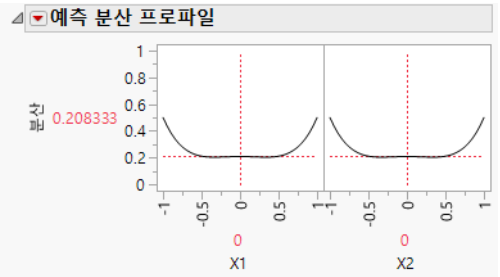
설계

런	X1	X2
1	0	0
2	1	0
3	-1	-1
4	0	0
5	0	1
6	-1	0
7	0	0
8	-1	1
9	1	-1
10	0	0
11	0	-1
12	1	1

이 I- 최적 설계에서 런 1, 4, 7, 10 은 중앙점 (X1 = 0, X2 = 0) 에 있습니다 . I- 최적 설계는 D- 최적 설계와 비교하여 설계 공간의 중심에 더 많은 런을 배치하는 경향이 있습니다 . 따라서 극단에 배치되는 런의 수가 더 적습니다 . 이 설계를 그림 5.42 에 표시된 D- 최적 설계와 비교할 수 있습니다 .

9. **설계 평가 > 예측 분산 프로파일** 개요를 엽니다 .

그림 5.41 I- 최적 모형에 대한 예측 분산 프로파일



기대 반응의 상대 예측 분산은 설계 공간의 중심에서 가장 작습니다 .

10. **설계 공간 비율 그림** 개요를 엽니다 .

설계 공간 비율 그림은 그림 5.44 의 왼쪽에 나옵니다 . "공간 비율" 이 0.95 일 때 파란색 곡선의 수직 좌표는 약 0.5 입니다 . 이는 설계 공간의 약 95% 에 대해 상대 예측 분산이 오차 분산의 50% 미만임을 의미합니다 .

현재 "사용자 설계" 창에는 I- 최적 설계가 포함되어 있습니다 . 이 창을 열어 두십시오 . 다음 섹션에서 D- 최적 설계를 생성하고 두 설계를 비교해 보겠습니다 .

D- 최적 설계

I- 최적 설계 및 D- 최적 설계에 대한 예측 분산 프로파일과 설계 공간 비율 그림을 비교하려면

1. I- 최적 설계를 포함하는 " 사용자 설계 " 창에서 " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **스크립트를 스크립트 창에 저장**을 선택합니다 .

   작업을 재현하는 스크립트가 표시된 창이 나타납니다 .
2. 이 새 스크립트 창에서 **편집 > 스크립트 실행**을 선택합니다 .

   중복 " 사용자 설계 " 창이 나타나지만 " 설계 평가 " 개요가 닫혀 있습니다 .
3. 이 새 " 사용자 설계 " 창에서 **뒤로**를 클릭합니다 .
4. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **최적 기준 > D- 최적 설계 생성**을 선택합니다 .
5. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 383570403 을 입력하고 **확인**을 클릭합니다 .
6. **설계 생성**을 클릭합니다 .

   현재 " 사용자 설계 " 창에는 D- 최적 설계가 포함되어 있습니다 .

그림 5.42 D- 최적 설계

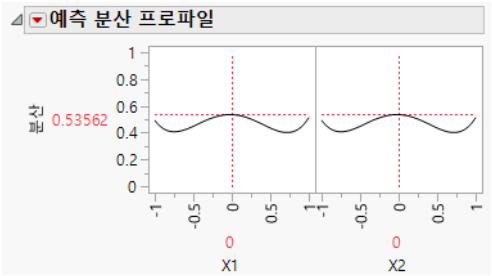
설계

런	X1	X2
1	0	0
2	1	0
3	-1	-1
4	1	-1
5	0	1
6	-1	0
7	-1	-1
8	-1	1
9	1	-1
10	-1	1
11	0	-1
12	1	1

이 D- 최적 설계에서는 런 1 만 중앙점에 있습니다 . D- 최적 설계는 I- 최적 설계와 비교하여 설계 공간의 극단에 더 많은 런을 배치하는 경향이 있습니다 . I- 최적 설계의 중앙점 런은 4 개였습니다 (그림 5.40).

7. **설계 평가 > 예측 분산 프로파일** 개요를 엽니다 .

그림 5.43 D- 최적 모형에 대한 예측 분산 프로파일



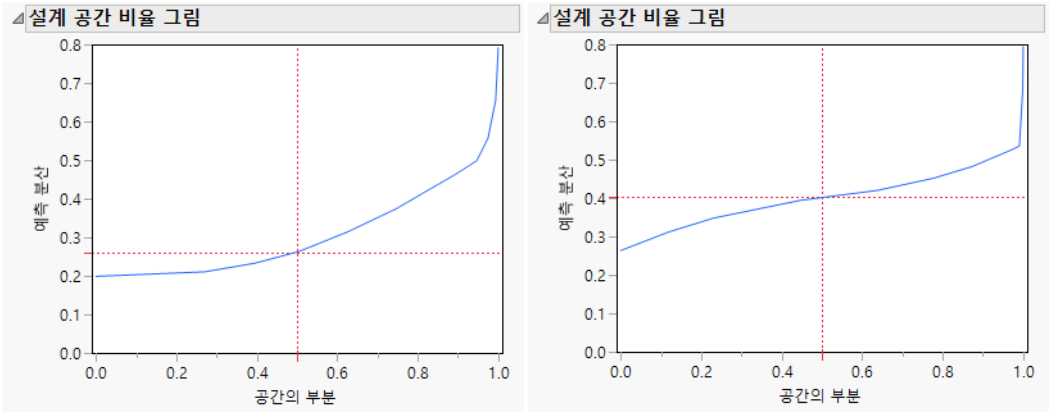
설계 영역의 중심에서 상대 예측 분산은 0.53562 이며 , I- 최적 설계의 경우 0.208333 이었습니다 (그림 5.41). 이는 상대 표준 오차가 D- 최적 설계의 경우 0.732 이고 I- 최적 설계의 경우 0.456 이라는 의미입니다 . 다른 조건이 모두 같다고 가정할 때 설계 영역의 중심에서 D- 최적 설계의 기대 반응에 대한 신뢰 구간이 I- 최적 설계의 경우보다 약 60% 더 넓습니다 .

" 설계 " 개요에서는 D- 최적 설계에 -1, 0, 1 로 설정된 X1 과 X2 의 모든 조합에 대해 하나씩 9 개의 설계점이 있음을 보여 줍니다 . D- 최적 기준은 이러한 각 설계점에서 상대 예측 분산을 낮추려고 합니다 . X1 및 X2 의 슬라이더를 -1 과 1 로 움직여 설계 영역의 극단에서 분산을 탐색합니다 . 이러한 극점에서의 분산은 대개 같은 지점에서 I- 최적 설계의 분산보다 작습니다 .

8. **설계 평가 > 설계 공간 비율 그림** 개요를 엽니다 .

설계 공간 비율 그림은 그림 5.44 의 오른쪽에 나옵니다 .

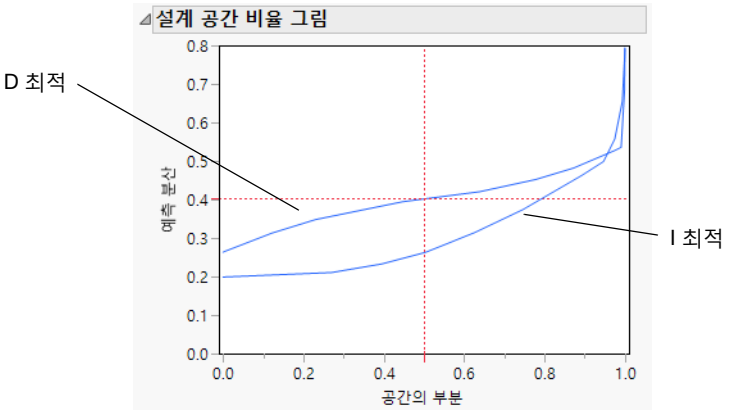
그림 5.44 설계 공간 비율 그림 (왼쪽 : I- 최적 , 오른쪽 : D- 최적)



각 그림의 빨간색 십자기호는 설계 공간의 50% 에 대한 최대 예측 분산을 나타냅니다 . 설계 공간의 50% 에 대한 I- 최적 설계의 예측 분산은 약 0.27 미만입니다 . D- 최적 설계의 경우 예측 분산은 약 0.4 입니다 .

9. I- 최적 설계에 대한 설계 공간 비율 그림에서 마우스 오른쪽 버튼을 클릭합니다 . **편집 > 프레임 내용 복사**를 선택합니다 .
10. D- 최적 설계에 대한 설계 공간 비율 그림에서 마우스 오른쪽 버튼을 클릭합니다 . **편집 > 프레임 내용 붙여넣기**를 선택합니다 .

그림 5.45 겹쳐진 설계 공간 비율 그림



I- 최적 설계의 분산 곡선은 설계 공간의 90% 이상에서 D- 최적 설계의 분산 곡선 아래에 있습니다 . 이는 I- 최적 기준이 전체 설계 공간에서 예측 분산을 최소화하려고 한다는 사실을 반영합니다 . 반면에 D- 최적 기준은 설계점에서 예측 분산을 줄이는 데 중점을 둡니다 .

제약 조건 및 범주형 요인이 있는 반응 표면 설계

이 예에서는 화학 반응의 수율을 최적화하는 설계를 생성합니다. 실험 요인에는 세 가지 수준의 범주형 요인이 포함되고 이 중 두 가지 수준이 제약 조건에 포함됩니다. 이 예에서는 허용되지 않는 조합 필터를 사용하여 제약 조건을 입력할 것입니다.

실험의 반응은 Yield 입니다 . 세 가지 요인은 다음과 같습니다 .

- Time: 관심 범위는 500 ~ 560 초입니다 .
- Temperature: 관심 범위는 350 ~ 750 도 (켈빈) 입니다 .
- Catalyst: 세 가지 촉매 (A, B, C) 를 테스트해야 합니다 .

설계 시 Catalyst 의 두 가지 수준에 대한 제약 조건을 고려해야 합니다 .

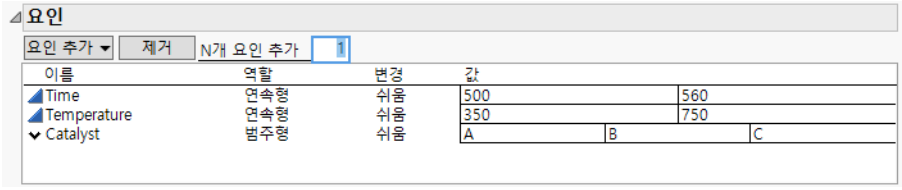
- 촉매 B 를 사용할 경우 온도는 400 도를 초과해야 합니다 .
- 촉매 C 를 사용할 경우 온도는 650 도 미만이어야 합니다 .

반응 및 요인 정의

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 반응 " 개요에서 Y 를 두 번 클릭한 후 Yield 로 변경합니다 .

   Yield 를 최대화하는 것이 목표이므로 " 최대화 " 로 설정된 **목표**를 그대로 두십시오 .
3. " 요인 " 개요에서 **N 개 요인 추가** 옆에 2 를 입력합니다 .
4. **요인 추가 > 연속형**을 클릭합니다 .
5. **요인 추가 > 범주형 > 3 수준**을 클릭합니다 .
6. 요인 이름을 Time, Temperature, Catalyst 로 각각 바꿉니다 .
7. Time 의 **값**을 500 과 560 으로 변경합니다 .
8. Temperature 의 **값**을 350 과 750 으로 변경합니다 .
9. Catalyst 의 **값**을 A, B, C 로 변경합니다 .

그림 5.46 요인 설정



10. **계속**을 클릭합니다 .

제약 조건 정의

1. " 요인 제약 조건 정의 " 개요에서 **허용되지 않는 조합 필터 사용**을 선택합니다 .

2. **필터 요인 추가** 목록에서 Temperature 와 Catalyst 를 선택합니다 .
3. **추가**를 클릭합니다 .
4. "Temperature" 슬라이더 위에 보이는 750 을 클릭한 후 400 으로 변경합니다 . Enter 키를 누르거나 다른 곳을 클릭합니다 .

   그러면 Temperature 값이 400 미만인 요인 설정이 허용되지 않습니다 .
5. Catalyst 아래의 B 블록을 클릭합니다 .

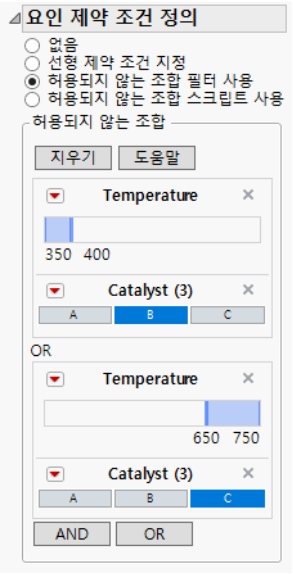
   그러면 Temperature 제약 조건을 함께 사용하여 Catalyst 가 B 이고 Temperature 가 400 미만인 요인 설정이 허용되지 않습니다 .
6. **OR** 를 클릭합니다 .

   이렇게 하면 두 번째 제약 조건을 정의할 수 있습니다 .
7. **필터 요인 추가** 목록에서 Temperature 와 Catalyst 를 선택합니다 . 이전 선택이 유지되었을 수 있습니다 .
8. **추가**를 클릭합니다 .
9. 단어 "OR" 의 아래쪽 패널에서 "Temperature" 슬라이더 위에 보이는 350 을 클릭한 후 650 으로 변경합니다 . Enter 키를 누르거나 다른 곳을 클릭합니다 .

   그러면 Temperature 값이 650 을 초과하는 요인 설정이 허용되지 않습니다 .
10. Catalyst 아래의 C 블록을 클릭합니다 .

    그러면 Temperature 제약 조건을 함께 사용하여 Catalyst 가 C 이고 Temperature 가 650 을 초과하는 요인 설정이 허용되지 않습니다 .

그림 5.47 제약 조건 정의



참고 : Catalyst 뒤의 (3) 은 catalyst 수준을 나타냅니다 .

반응 표면 항 추가 및 설계 생성

1. " 모형 " 개요에서 **RSM** 을 선택합니다 .

   연속형 요인에 대한 2 차 항만 모형에 추가됨을 알려 주는 JMP 경고가 표시됩니다 .
2. **확인**을 클릭하여 경고를 해제합니다 .

   JMP 는 적절한 반응 표면 항만 모형에 추가합니다 .

   참고 : 3 단계에서 " 난수 시드값 " 을 설정하고 4 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .
3. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 654321 을 입력하고 **확인**을 클릭합니다 .
4. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 1000 으로 설정합니다 . **확인**을 클릭합니다 .
5. **설계 생성**을 클릭합니다 .

그림 5.48 제약 조건을 충족하는 설계

설계

런	Time	Temperature	Catalyst
1	500	750	B
2	500	410	B
3	560	750	A
4	521	750	A
5	542	350	A
6	500	648.7299	C
7	500	350	A
8	533	630	C
9	506	350	C
10	560	530	A
11	560	410	B
12	530	570	B
13	530	550	A
14	530	570	B
15	557	645.1294	C
16	500	570	A
17	560	750	B
18	560	350	C

RSM 항을 " 모형 " 개요에 추가했으므로 I- 최적 설계가 생성됩니다 . 이 설계는 다음 제약 조건을 충족합니다 .

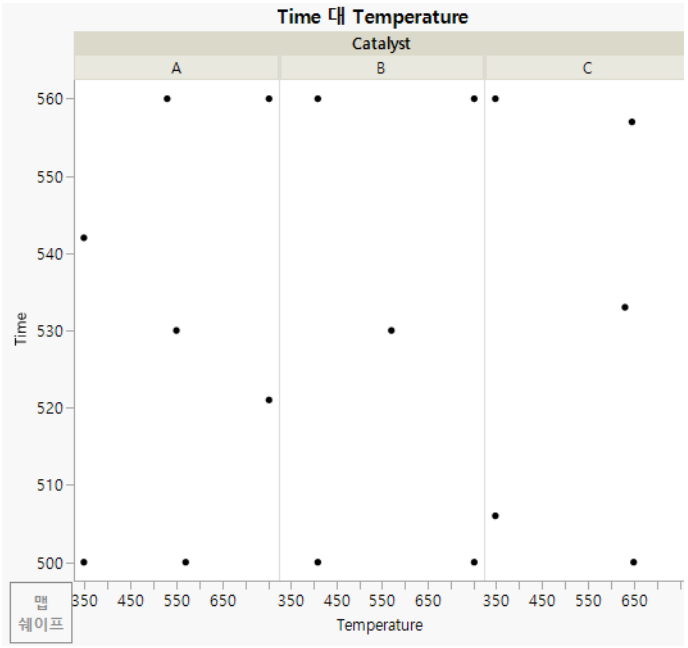
- 촉매 B 를 사용할 경우 온도는 400 도를 초과해야 합니다 .
- 촉매 C 를 사용할 경우 온도는 650 도 미만이어야 합니다 .

설계 보기

설계 영역은 500 ~ 560 사이의 Time 설정과 350 ~ 750 사이의 Temperature 설정으로 구성되었습니다 . 설계 영역 내의 설계점을 기하학적 보기로 표시하기 위해 다음과 같이 그래프를 생성합니다 .

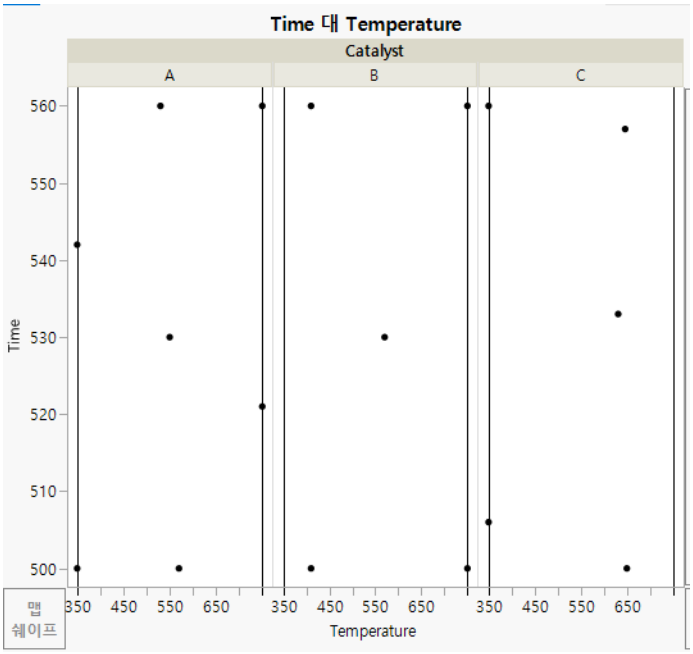
1. **테이블 생성**을 클릭합니다 .
2. **그래프 > 그래프 빌더**를 선택합니다 .
3. Time 과 Temperature 를 선택하고 템플릿의 가운데로 드래그합니다 .
4. **평활기** 아이콘을 선택 취소합니다 . 아 아이콘은 템플릿 상단의 왼쪽에서 두 번째에 있습니다 .
5. Catalyst 를 선택하고 템플릿의 맨 위에 있는 **그룹 X** 영역으로 드래그합니다 .

그림 5.49 세 가지 수준의 Catalyst 에 대한 설계점



6. Temperature 축을 두 번 클릭합니다 .
7. " 참조선 " 아래에서 다음과 같이 두 개의 참조선을 추가합니다 .
   - **값** 옆에 350 을 입력합니다 .
   - **추가**를 클릭합니다 .
   - **값** 옆에 750 을 입력합니다 .
   - **추가**를 클릭합니다 .
8. **확인**을 클릭합니다 .

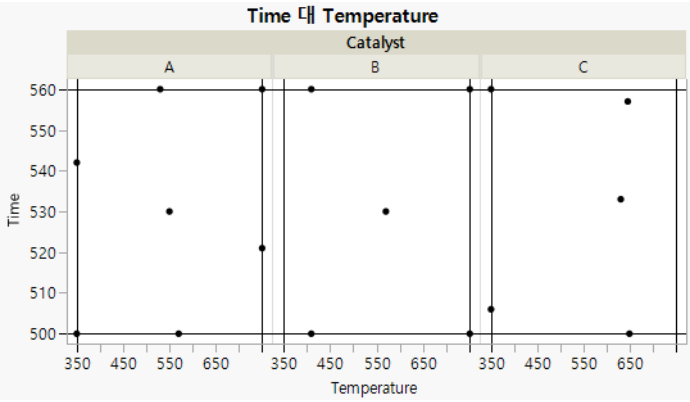
그림 5.50 세 가지 수준의 Catalyst 에 대한 Temperature 설계 경계



9. Time 축을 두 번 클릭합니다 .
10. " 참조선 " 아래에서 다음과 같이 두 개의 참조선을 추가합니다 .
    - **값** 옆에 500 을 입력합니다 .
    - **추가**를 클릭합니다 .
    - **값** 옆에 560 을 입력합니다 .
    - **추가**를 클릭합니다 .
11. **확인**을 클릭합니다 .
12. **완료**를 클릭합니다 .

참고 : 그림 5.51 의 그림은 세로로 크기가 조정됩니다 .

그림 5.51 세 가지 수준의 Catalyst 에 대한 설계 영역



Catalyst A 의 6 개 설정은 설계 영역의 가장자리에 있습니다 . Catalyst B 의 모든 Temperature 설정은 400 도 제약 조건보다 높고 , Catalyst C 의 모든 Temperature 설정은 650 도 제약 조건보다 낮습니다 .

혼합물 실험

사용자 설계 플랫폼과 혼합물 설계 플랫폼은 둘 다 모든 요인이 혼합 내 성분인 상황에 대한 설계를 구성합니다 . 그러나 혼합물 실험은 비혼합 공정 변수 (공정 요인) 를 포함할 수 있습니다 . 사용자 설계 플랫폼은 혼합 요인만 있는 설계뿐 아니라 혼합 요인과 공정 요인을 모두 수용하는 설계를 구성할 수 있습니다. 또한 사용자 설계 플랫폼에서는 혼합 성분의 합이 임의의 양수가 될 수 있습니다 . 자세한 내용은 " 사용자 설계 " 장의 " 요인 유형 "(85 페이지) 에서 " 혼합 " 에 대한 설명을 참조하십시오 .

- " 비혼합 요인이 있는 혼합물 설계 "
- " 혼합물 혼합 설계 "

비혼합 요인이 있는 혼합물 설계

Atkinson and Donev(1992)에 나오는 이 예에서는 혼합 요인과 공정 요인을 둘 다 포함하는 실험을 위한 설계를 생성합니다. 범주형 요인의 수준에 따라 균형이 조정된 18회 런 설계가 생성됩니다. 이 설계를 사용하면 전체 반응 표면을 적합시킬 수 있습니다. "설계 평가" 그림과 결과를 사용하여 설계의 상대 예측 분산을 검토합니다.

이 설계는 다음과 같이 1개의 단일 반응, 3개의 혼합 요인 및 1개의 비혼합 요인으로 구성됩니다.

- 반응은 아크릴로니트릴 파우더의 전자기 감쇠를 측정하는 Damping 입니다 .
- 세 가지 혼합 성분은 다음과 같습니다 .
  - CuSO4(황산구리), 0.2 ~ 0.8 범위

  - Na2S2O3(티오황산나트륨), 0.2 ~ 0.8 범위
  - Glyoxal(글리옥살), 0 ~ 0.6 범위
- 관심 있는 비혼합 환경 요인은 Wavelength(전자파 파장) 이며 세 가지 수준 L1, L2, L3 으로 나타냅니다 .

  Wavelength 는 연속형 변수입니다 . 그러나 연구원들은 세 가지 특정 파장에서의 예측에만 관심이 있었습니다 . 이 때문에 Wavelength 를 3 수준 범주형 요인으로 처리합니다 .

이 예는 다음과 같이 두 부분으로 이루어져 있습니다 .

- " 설계 생성 "(163 페이지)
- " 설계 평가 "(165 페이지)

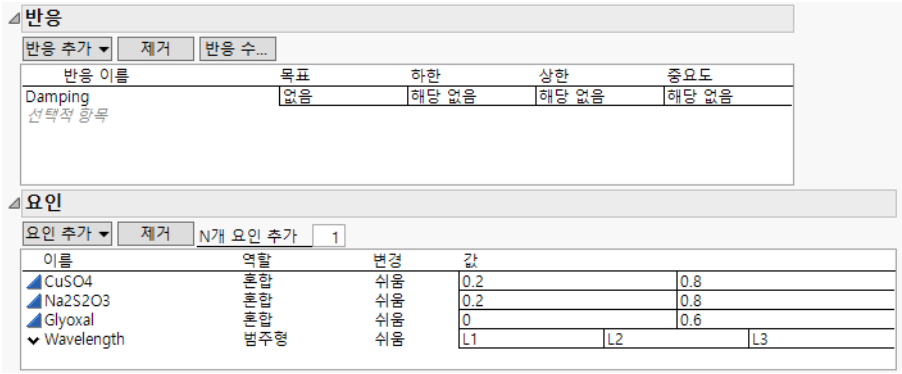
설계 생성

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Damping 을 입력합니다 .
3. " 목표 " 아래의 **최대화**를 클릭한 후 **없음**으로 변경합니다 .

   연구 작성자가 적절한 감쇠 크기에 대해 언급하지 않으므로 목표가 " 없음 " 으로 설정됩니다 . **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Donev Mixture Factors.jmp 를 엽니다 .
4. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

   그러면 세 가지 혼합 성분과 범주형 환경 요인을 불러옵니다 . 세 가지 혼합 요인의 값에 대한 한계도 불러옵니다 .

그림 5.52 반응 개요 및 요인 개요

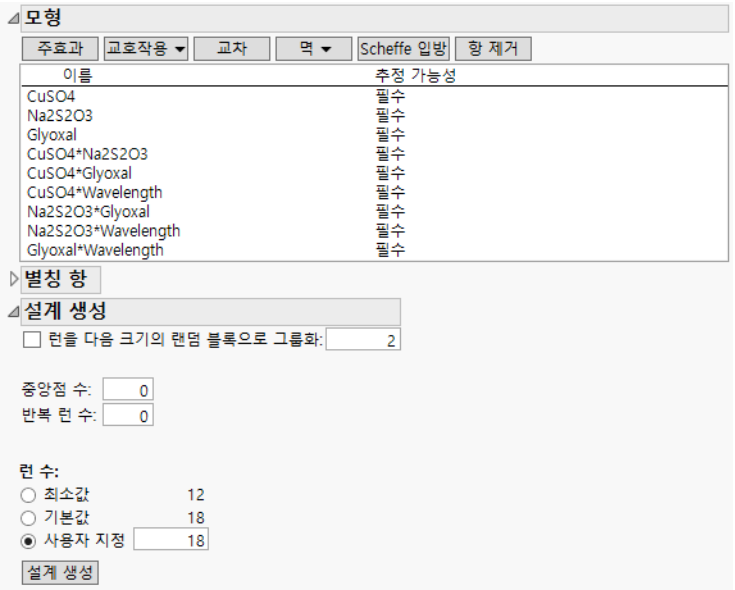


5. " 모형 " 개요에서 **교호작용 > 2 차**를 클릭합니다 .

   모든 혼합 요인과 상호 작용하는 비혼합 요인의 주효과 항이 제거됨을 알려 주는 JMP 경고 창이 표시됩니다 . 즉 , Wavelength 의 주효과는 제거되지만 혼합 요인과 Wavelength 의 모든 이원 교호작용은 추가됩니다 .
6. **확인**을 클릭하여 메시지를 해제합니다 .

"모형" 개요의 효과는 파장의 가법 효과와 함께 혼합 성분의 반응 표면 모형을 정의합니다. 자세한 내용은 Scheffé(1958) 에서 확인하십시오.

그림 5.53 모형 및 설계 생성 개요



7. 기본값 런 수 18 을 그대로 둡니다.

   18 회 런을 선택하면 파장 요인의 세 가지 수준 각각에 대해 6 회씩 런을 수행할 수 있습니다.

   참고 : 8 단계에서 "난수 시드값" 을 설정하고 9 단계에서 "시작 수" 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다. 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다.

8. (선택 사항) "사용자 설계" 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다.
9. (선택 사항) "사용자 설계" 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 5 를 입력하고 **확인**을 클릭합니다.
10. **설계 생성**을 클릭합니다.

그림 5.54 18 회 런 설계를 보여 주는 설계 개요

설계

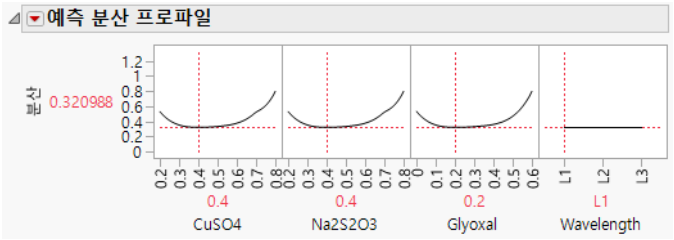
런	CuSO4	Na2S2O3	Glyoxal	Wavelength
1	0.8	0.2	0	L3
2	0.2	0.2	0.6	L3
3	0.8	0.2	0	L2
4	0.2	0.8	0	L2
5	0.5	0.5	0	L3
6	0.2	0.2	0.6	L1
7	0.5	0.2	0.3	L1
8	0.8	0.2	0	L1
9	0.2	0.2	0.6	L2
10	0.5	0.5	0	L1
11	0.2	0.8	0	L3
12	0.2	0.5	0.3	L2
13	0.5	0.5	0	L2
14	0.2	0.5	0.3	L1
15	0.5	0.2	0.3	L3
16	0.2	0.5	0.3	L3
17	0.5	0.2	0.3	L2
18	0.2	0.8	0	L1

각 Wavelength 수준에 대해 6 개의 런이 있는지 확인할 수 있습니다 .

설계 평가

1. **설계 평가** > **예측 분산 프로파일** 개요를 엽니다 .

그림 5.55 18 회 런 설계에 대한 예측 분산 프로파일



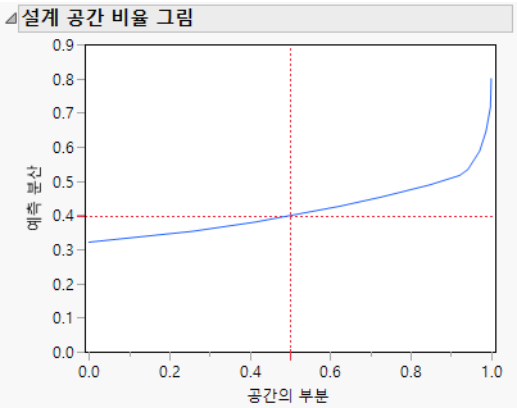
Wavelength 의 슬라이더를 움직여 혼합 요인에 대한 상대 예측 분산 프로파일이 전체 Wavelength 수준에서 변하지 않는지 확인합니다. 혼합 요인 중 하나의 슬라이더를 움직여 봅니다 . 혼합 성분의 합이 1 이 되도록 다른 두 혼합 요인의 슬라이더가 조정됩니다 . 혼합 요인의 중심 설정 가까이에서 상대 예측 분산이 가장 작습니다 .

2. " 예측 분산 프로파일 " 의 빨간색 삼각형을 클릭하고 **분산 최대화**를 선택합니다 .

   설계 공간에 대한 최대 상대 예측 분산은 오차 분산의 0.8 배입니다 .

3. **설계 공간 비율 그림** 개요를 엽니다 .

그림 5.56 18 회 런 설계에 대한 설계 공간 비율 그림



전체 설계 공간에서 상대 예측 분산은 0.8 미만입니다 . 최소 상대 예측 분산은 약 0.32 입니다 . 그림 5.55 에 표시된 것과 같이 혼합 요인의 중심 설정 가까이에서 값이 가장 작습니다 .

4. **설계 진단** 개요를 엽니다 .

그림 5.57 18 회 런 설계에 대한 설계 진단

이 설계는 D 효율도(3.6%)가 매우 낮지만 D-최적 기준에 따라 최적 설계입니다. 혼합물 설계는 혼합 제약 조건으로 인해 직교와 거리가 멀기 때문에 일반적으로 D 효율도가 매우 낮습니다. 예측 평균 분산(상대)은 0.412864 입니다. 이 값은 그림 5.56 의 설계 공간 비율 그림과 일치합니다.

혼합물 혼합 설계

이 예에서는 혼합물이 혼합되는 상황에 대한 설계를 구성합니다.

케이크에 들어가는 재료를 예로 들어보겠습니다. 마른 재료에는 밀가루, 설탕, 코코아가 있습니다. 젖은 재료에는 우유, 녹인 버터, 계란이 있습니다. 케이크의 젖은 성분과 마른 성분은 먼저 따로따로 섞은 후 하나로 혼합되는 두 가지 혼합물입니다. 표 5.1 에는 각 요인 및 실험 중에 해당 요인에 변화를 줄 수 있는 범위가 나와 있습니다.

표 5.1 마른 성분과 젖은 성분 및 실험 범위

혼합	성분	하한 및 상한 수준
Dry	Cocoa	0.1 ~ 0.2
	Sugar	0 ~ 0.15
	Flour	0.2 ~ 0.3
Wet	Butter	0.1 ~ 0.2
	Milk	0.25 ~ 0.35
	Eggs	0.05 ~ 0.20

마른 성분(코코아, 설탕 및 밀가루 혼합물)은 합친 혼합물의 45%를 차지합니다. 젖은 성분(버터, 우유, 계란)은 55%를 차지합니다.

참고 : 마른 재료와 젖은 재료에 대한 제약 조건을 사용하여 혼합 모형에 6 개 요인을 모두 포함하면 모형 특이성이 존재합니다 .

실험 목표는 Taste(맛) 등급을 최적화하는 것입니다. Taste는 1 ~ 10의 척도로 평가되며, 10이 제일 좋은 맛을 나타냅니다.

주효과 모형을 적합시키기 위해 10회 런 설계를 구성합니다. 마른 재료와 젖은 재료 비율에 대한 제약 조건 때문에 특이성을 방지하기 위해 "모형" 개요에 5개 요인만 포함해야 합니다. 포함하지 않을 요인은 임의로 선택합니다.

이 예는 다음과 같이 두 부분으로 이루어져 있습니다 .

- " 설계 생성 "(167 페이지)
- " 실험 결과 분석 "(169 페이지)

설계 생성

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Taste 를 입력합니다 .

   기본 목표는 " 최대화 " 로 되어 있습니다 . Taste 등급을 최대화해야 하므로 목표를 변경하지 마십시오 .
3. " 하한 " 아래를 클릭하고 0 을 입력합니다 .

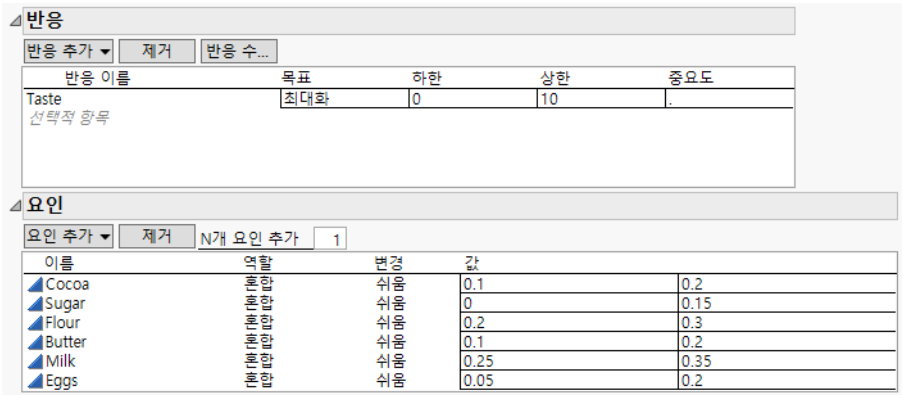
   최하위 등급이 0 입니다 .
4. " 상한 " 아래를 클릭하고 10 을 입력합니다 .

   최상위 등급이 10 입니다 .
5. " 중요도 " 아래 영역은 비워 둡니다 .

반응이 하나뿐이므로 기본적으로 반응의 중요도가 1 로 지정됩니다 .

6. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Cake Factors.jmp 를 엽니다 .
7. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

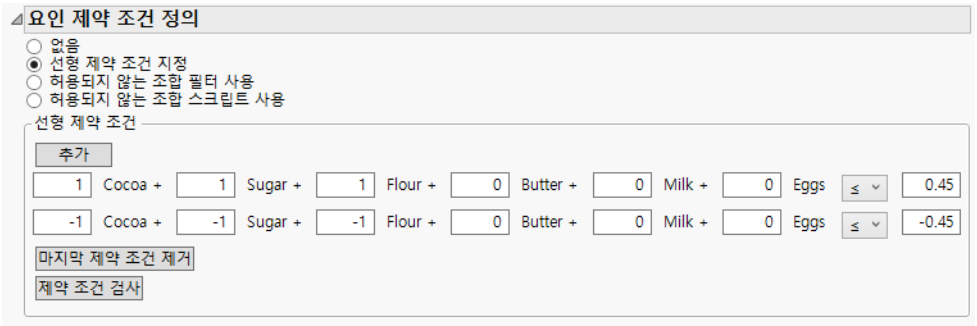
그림 5.58 완료된 반응 및 요인 개요



요인은 모두 혼합 요인입니다 . 실험의 설정 범위를 정의하는 " 값 " 은 요인마다 다릅니다 .

8. " 요인 제약 조건 정의 " 개요에서 **선형 제약 조건 지정**을 선택합니다 .
9. " 선형 제약 조건 " 패널에서 **추가**를 두 번 클릭합니다 .
10. 그림 5.59 에 표시된 제약 조건을 입력합니다 .

그림 5.59 요인 제약 조건 정의



두 번째 제약 조건은 Cocoa, Sugar, Flour 의 합이 0.45 보다 크거나 같음을 지정하는 것과 동등합니다 . 두 제약 조건은 Cocoa, Sugar, Flour 가 혼합물의 정확히 45% 를 차지하여 젖은 요인이 나머지 55% 를 구성한다는 것을 의미합니다 .

11. " 모형 " 개요에서 "Eggs" 를 선택하고 **항 제거**를 클릭합니다 .

등식 제약 조건 때문에 6 개 효과를 모두 포함하는 모형은 특이 모형입니다 .

12. **사용자 지정** 옆에 10 을 입력합니다 .

실험에서 10 개의 케이크를 구워야 합니다 .

참고 : 13 단계에서 " 난수 시드값 " 을 설정하고 14 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

13. (선택 사항) "사용자 설계 "의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
14. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 40 을 입력하고 **확인**을 클릭합니다 .
15. **설계 생성**을 클릭합니다 .

    요인 제약 조건에 등식 제약 조건이 포함되어 있음을 알려 주는 JMP 경고가 표시됩니다 . 마른 재료 비율의 합이 45% 로 제한되므로 이는 의도한 대로 된 것입니다 .
16. **확인**을 클릭하여 JMP 경고를 해제합니다 .
17. **테이블 생성**을 클릭합니다 .

그림 5.60 혼합물 혼합 설계

	Cocoa	Sugar	Flour	Butter	Milk	Eggs	Taste
1	0.2	0	0.25	0.2	0.3	0.05	•
2	0.1	0.15	0.2	0.15	0.35	0.05	•
3	0.1	0.05	0.3	0.2	0.25	0.1	•
4	0.2	0.05	0.2	0.1	0.25	0.2	•
5	0.2	0.05	0.2	0.1	0.35	0.1	•
6	0.1	0.15	0.2	0.1	0.25	0.2	•
7	0.1	0.05	0.3	0.1	0.25	0.2	•
8	0.1	0.15	0.2	0.2	0.3	0.05	•
9	0.15	0	0.3	0.1	0.35	0.1	•
10	0.2	0.05	0.2	0.2	0.25	0.1	•

마른 재료인 Cocoa, Sugar 및 Flour 설정의 합이 혼합물의 45% 이고 젖은 재료인 Butter, Milk 및 Eggs 설정의 합이 혼합물의 55% 입니다 . 또한 이 설정은 " 요인 " 개요에 지정된 상한과 하한에 맞습니다 .

실험 결과 분석

Cake Data.jmp 샘플 데이터 테이블은 실험 결과를 보여 줍니다 . 이 설계 테이블에 포함된 "Model" 스크립트는 DOE 창의 " 모형 " 개요에서 지정한 5 개 주효과를 보여 주는 " 모형 적합 " 창을 엽니다 . Eggs 의 주효과는 이 설계의 " 모형 " 개요에 포함되어 있지 않습니다 . 나머지 5 개 요인의 설정과 마른 재료 제약 조건에 의해 Eggs 수준을 결정할 수 있습니다 . 이 스크립트는 사용자 설계에서 생성되는 데이터 테이블에 저장되었습니다 .

1. Design Experiment 폴더의 Cake Data.jmp 샘플 데이터 테이블을 엽니다 .
2. 설계 테이블의 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

Eggs 로 인한 주효과는 포함되어 있지 않습니다 . 모형 특이성을 방지하기 위해 " 사용자 설계 " 창의 " 모형 " 개요에서 Eggs 효과가 제외되었습니다 . 5 개 효과는 반응 표면 효과와 혼합 효과로 지정됩니다 .

3. **실행**을 클릭합니다 .

   일부 요인에 대한 혼합물 혼합 제약 조건 때문에 프로파일러를 표시할 수 없음을 알려 주는 JMP 경고가 나타납니다 .

4. **확인**을 클릭하여 JMP 경고를 해제합니다 .

"모수 추정값" 보고서는 Sugar와 Flour가 유의 수준 0.05에서 유의함을 나타냅니다.

그림 5.61 모수 추정값 보고서

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
(Cocoa-0.1)/0.35	5.1056115	2.360215	2.16	0.0829
Sugar/0.35	16.873989	1.675197	10.07	0.0002*
(Flour-0.2)/0.35	15.750757	2.207353	7.14	0.0008*
(Butter-0.1)/0.35	5.3740398	2.654689	2.02	0.0988
(Milk-0.25)/0.35	1.5973046	2.435648	0.66	0.5409

공변량이 포함된 실험

실험에 사용할 실험 단위에 대한 측정값을 사용할 수 있는 경우가 있습니다. 이러한 측정값이 실험 반응에 영향을 줄 수 있습니다. **공변량**이라고 하는 이 변수를 설계 요인으로 포함하면 유용합니다. 이러한 값을 직접 제어할 수는 없지만 모든 효과를 가장 정확하게 추정할 수 있도록 다른 요인의 수준을 선택할 수 있습니다.

사용자 설계 플랫폼은 공변량 값을 최적의 방식으로 선택하는 설계를 구성합니다. 공변량 값은 잠재적 실험 단위에 대한 공변량 정보를 제공하는 기존 데이터 테이블에서 선택됩니다. 데이터 테이블에 나열된 실험 단위의 수보다 적은 수의 런을 지정할 수 있습니다. 변경하기 힘든 공변량도 지정할 수 있습니다. 설계를 생성할 때 런 수가 공변량 테이블의 행 수보다 적으면 설계 테이블에 **공변량 행 인덱스** 열이 포함됩니다. 이 열은 공변량 테이블에서 각 실험 런에 해당하는 행을 나타냅니다.

참고: 공변량 데이터 테이블에서 공변량 요인의 값이 비결측인 행 수는 지정된 런 수보다 크거나 같아야 합니다 .

- " 고정 공변량이 포함된 설계 "
- " 변경하기 힘든 공변량이 포함된 설계 "
- " 시간 선형 추세가 포함된 설계 "

고정 공변량이 포함된 설계

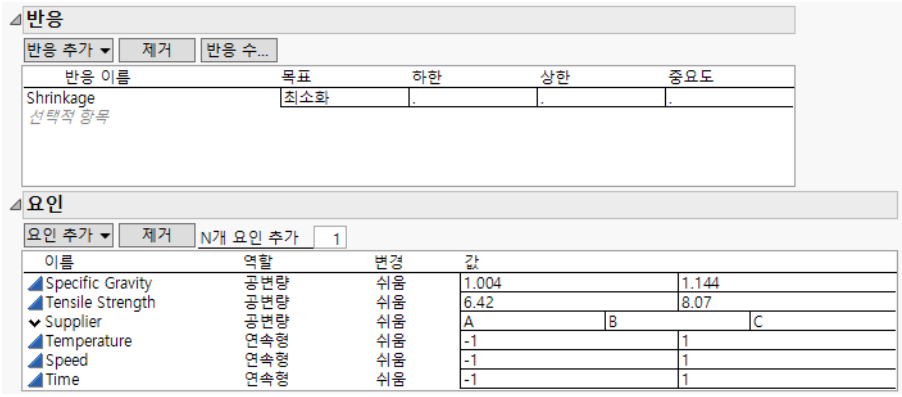
이 예에서는 사출 성형 공정에 의해 생산된 부품의 Shrinkage를 모델링하려고 합니다. Design Experiment 폴더의 Thermoplastic.jmp 샘플 데이터 테이블에는 연구에 사용할 수 있는 25개의 원료(열가소성수지) 배치가 나열됩니다. 각 배치에 대해 Specific Gravity 및 Tensile Strength 측정값을 구하기 위해 재료가 제거되었습니다. 세 번째 공변량인 Supplier도 사용할 수 있습니다.

Temperature(금형 온도), Speed(스크류 속도) 및 Time(보압 시간)이라는 세 가지 제어 가능한 요인이 Shrinkage에 미치는 효과를 연구하려고 합니다. 그러나 Specific Gravity, Tensile Strength 및 Supplier와 같은 공변량의 효과도 연구하고 싶습니다. 현재 리소스로 가능한 런 수는 12회입니다.

설계 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Thermoplastic.jmp를 엽니다.
2. **DOE > 사용자 설계**를 선택합니다.
3. "반응 이름" 아래의 Y를 두 번 클릭하고 Shrinkage를 입력합니다.
4. "목표" 아래의 **최대화**를 클릭한 후 **최소화**로 변경합니다.
5. **요인 추가**를 클릭하고 **공변량**을 선택합니다.
6. 목록에서 Specific Gravity, Tensile Strength 및 Supplier를 선택하고 **확인**을 클릭합니다.
   이 항목은 공변량이며 제어할 수 없습니다.
7. **N 개 요인 추가** 옆에 3을 입력합니다.
8. **요인 추가 > 연속형**을 클릭합니다.
9. 세 가지 연속형 요인의 이름을 Temperature, Speed, Time으로 각각 바꿉니다.
   이 항목은 요인이며 제어할 수 있습니다.

그림 5.62 반응 및 요인 개요



10. **계속**을 클릭합니다.
    런 수는 공변량 값이 있는 행 수를 나타냅니다. 공변량을 측정한 25개의 배치가 있습니다.

11. **런 수** 옆에 12 를 입력합니다 .

> **참고 :** 12 단계에서 " 난수 시드값 " 을 설정하고 13 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

12. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 84951 을 입력하고 **확인**을 클릭합니다 .
13. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 40 을 입력하고 **확인**을 클릭합니다 .
14. **설계 생성**을 클릭합니다 .

그림 5.63 12 회 런 최적 설계

설계

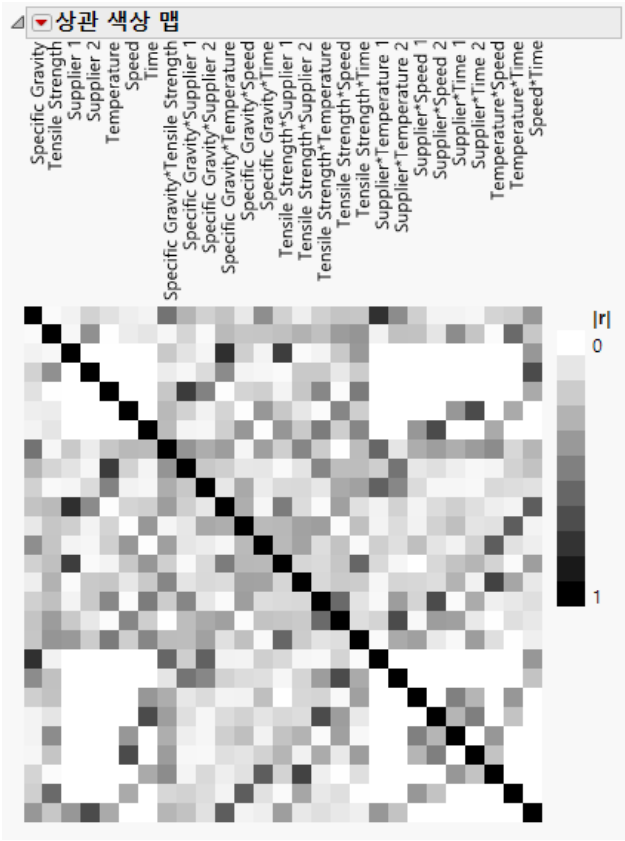
런	Specific Gravity	Tensile Strength	Supplier	Temperature	Speed	Time
1	1.107	8.07	B	1	1	1
2	1.144	6.82	B	1	-1	-1
3	1.004	6.52	C	1	1	-1
4	1.015	7.46	A	-1	1	-1
5	1.129	6.97	A	1	-1	-1
6	1.139	6.94	C	-1	1	-1
7	1.004	7.12	C	1	-1	1
8	1.067	6.7	B	-1	1	1
9	1.047	6.42	A	-1	-1	1
10	1.03	7.54	B	-1	-1	-1
11	1.094	6.48	A	1	1	1
12	1.113	6.89	C	-1	-1	1

이 설계는 잠재적 공변량 값을 고려한 D- 최적입니다 . 이 설계는 최적의 공변량 값 집합 및 제어 가능 요인의 최적 설정을 선택합니다 .

설계 평가

1. **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

그림 5.64 상관 색상 맵



주효과에 해당하는 7 개 항이 색상 맵 왼쪽 위에 나타납니다 . 이 7 개 항은 직교에 가깝습니다 . Tensile Strength 와 Supplier 2 사이의 절대 상관이 가장 큽니다 . 이 절대 상관 (약 0.43) 은 사용 가능한 공변량 값의 결과입니다 . 그림 5.64 에는 JMP 기본 색상이 사용됩니다 .

변경하기 힘든 공변량이 포함된 설계

이 예에서는 마모 (Wear) 특성이 심한 달리기 선수가 신을 운동화를 개발하기 위한 설계를 구성합니다 . 이 실험에는 다음과 같은 네 가지 요인이 포함됩니다 .

- 밑창 두께 (Thickness)
- 젤 쿠션재 양 (Gel)
- 구두창 재료 (Outsole)
- 중창 재료 (Midsole)

일반화할 수 있는 결론을 얻으려면 신발을 테스트할 대상 선수의 범위를 넓게 잡아야 합니다. 실험 예산에 맞추기 위해 각 선수가 몇 가지 실험 조합을 테스트해야 합니다.

연구에 참가할 100 명의 선수를 대상으로 데이터를 수집했습니다. 선수들에 대해 측정한 병존 변수(공변량)는 일일 평균 달린 거리(Miles), 몸무게(Weight) 및 지면에 발이 부딪히는 지점(Strike Point)입니다.

설계 생성 :

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Runners Covariates.jmp 를 엽니다 .
2. **DOE > 사용자 설계**를 선택합니다 .
3. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Wear 를 입력합니다 .
4. " 목표 " 아래의 **최대화**를 클릭한 후 **최소화**로 변경합니다 .
5. **요인 추가**를 클릭하고 **공변량**을 선택합니다 .
6. 목록에서 Miles, Weight 및 Strike Point 를 선택하고 **확인**을 클릭합니다 .

   이 항목은 달리기 선수와 관련된 변경하기 힘든 공변량입니다 .
7. Miles, Weight, Strike Point 요인 중 하나에 대해 " 변경 " 아래에서 **쉬움**을 클릭한 후 **어려움**으로 변경합니다 .

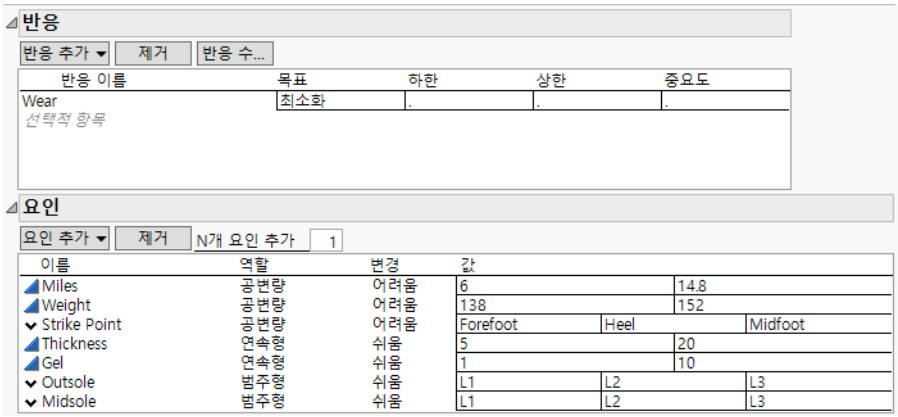
   그러면 세 공변량의 " 변경 " 값이 모두 " 어려움 " 으로 변경됩니다 .

나머지 요인을 수동으로 추가하려면 8 단계 ~ 16 단계의 작업을 수행합니다. 또는 저장된 테이블에서 요인을 불러오려면 "사용자 설계"의 빨간색 삼각형에서 **요인 불러오기**를 선택합니다. Design Experiment 폴더의 Runners Factors.jmp 샘플 데이터 테이블을 엽니다. **요인 불러오기**를 선택한 경우에는 8 단계 ~ 16 단계의 작업을 건너뛰십시오.

8. **N 개 요인 추가** 옆에 2 를 입력합니다 .
9. **요인 추가 > 연속형**을 클릭합니다 .
10. 두 요인의 이름을 Thickness 와 Gel 로 각각 바꿉니다 .
11. Thickness 의 **값**을 5 와 20 으로 변경합니다 .
12. Gel 의 **값**을 1 과 10 으로 변경합니다 .
13. **N 개 요인 추가** 옆에 2 를 입력합니다 .
14. **요인 추가 > 범주형 > 3 수준**을 클릭합니다 .
15. 두 요인의 이름을 Outsole 과 Midsole 로 각각 바꿉니다 .

    이 요인의 **값**은 기본값을 유지합니다 .

그림 5.65 반응 및 요인 개요



16. **계속**을 클릭합니다 .
17. **교호작용 > 2 차**를 선택합니다 .

    지정된 모형은 공변량과 실험 요인의 교호작용을 포함하여 모든 2요인 교호작용을 적합시킵니다 .
18. 주구 (선수) 수를 32 로 설정합니다 (이 값으로 설정되지 않은 경우).
19. " 런 수 " 의 " 사용자 지정 " 옆에 64 를 입력합니다 (이 값으로 설정되지 않은 경우).

    참고 : 20 단계에서 " 난수 시드값 " 을 설정하고 21 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .
20. (선택 사항) "사용자 설계"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
21. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 1 로 설정합니다 (이 값으로 설정되지 않은 경우). **확인**을 클릭합니다 .
22. **설계 생성**을 클릭합니다 .

그림 5.66 변경하기 힘든 공변량에 대한 설계의 처음 20 회 런

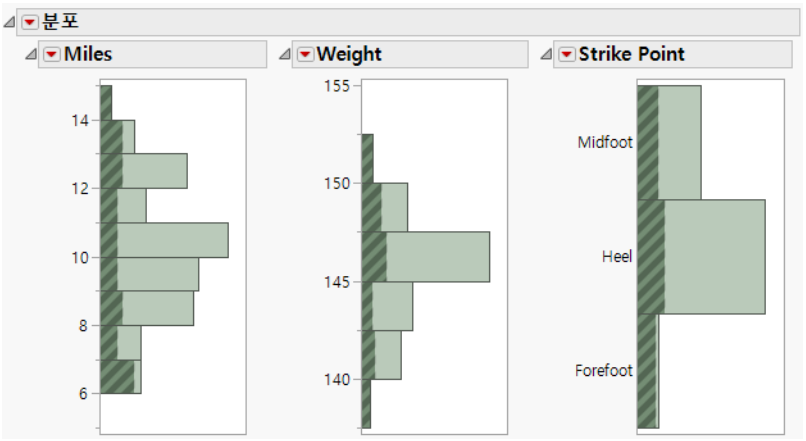
설계

런	주구	Miles	Weight	Strike Point	Thickness	Gel	Outsole	Midsole
1	1	7.6	144	Midfoot	5	10	L3	L2
2	1	7.6	144	Midfoot	20	1	L2	L1
3	2	7.6	145	Midfoot	20	10	L3	L3
4	2	7.6	145	Midfoot	5	1	L1	L1
5	3	14.8	141	Midfoot	20	1	L1	L3
6	3	14.8	141	Midfoot	5	10	L2	L1
7	4	6.4	146	Heel	20	10	L1	L2
8	4	6.4	146	Heel	5	1	L2	L3
9	5	11.2	138	Heel	5	1	L2	L3
10	5	11.2	138	Heel	20	10	L3	L2
11	6	11.2	144	Forefoot	20	1	L3	L3
12	6	11.2	144	Forefoot	5	10	L2	L1
13	7	12.8	139	Heel	5	10	L3	L1
14	7	12.8	139	Heel	20	1	L1	L2
15	8	14	139	Midfoot	5	1	L1	L2
16	8	14	139	Midfoot	20	10	L2	L3
17	9	6	150	Heel	20	10	L2	L2
18	9	6	150	Heel	5	1	L3	L3
19	10	10.8	143	Forefoot	20	1	L3	L1
20	10	10.8	143	Forefoot	5	10	L2	L2

100 명의 선수 중 공변량 값을 기준으로 32 명이 선택됩니다 . 선택된 선수에 해당하는 행은 RunnersCovariates.jmp 샘플 데이터 테이블에서 선택합니다 . 실험 요인인 Thickness, Gel, Insole 및 Outsole 의 설정은 " 모형 " 개요에 설명된 모형을 위한 최적 설계가 생성될 수 있도록 결정됩니다 .

23. RunnersCovariates.jmp 샘플 데이터 테이블을 활성 테이블로 사용하여 **분석 > 분포**를 선택합니다 .
24. 세 열을 모두 **Y, 열**로 선택합니다 .
25. **히스토그램만**을 선택합니다 .
26. **확인**을 클릭합니다 .

그림 5.67 100 명의 선수에 대한 히스토그램 (선택된 선수 데이터는 음영 처리됨)



이 히스토그램은 설계에서 공변량 간에 런 균형을 맞추기 위해 대상 선수가 선택된 방법을 보여 줍니다. 선택된 선수의 분포는 각 공변량에 대한 모든 선수의 전체 분포와 유사합니다.

시간 선형 추세가 포함된 설계

시간 순서에 따라 실험을 수행할 때 반응에서 선형 흐름이 나타나는 경우가 종종 있습니다. 일반적으로 런 순서를 랜덤화하면 추정된 요인 효과가 흐름 효과에 의해 편향되지 않습니다. 그러나 흐름을 고려하면 해당 효과의 분산을 줄일 수 있습니다.

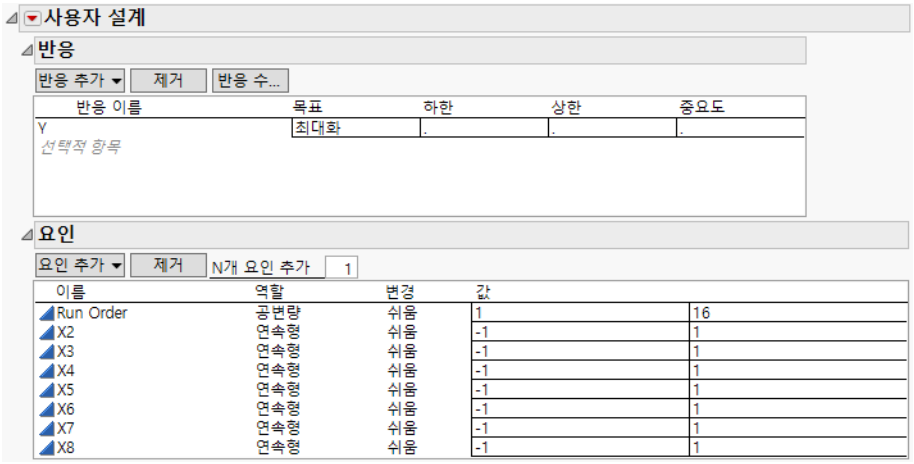
요인 변화와 독립적으로 시간에 따라 반응에서 강한 선형 추세를 의심할 이유가 있다고 가정해 보겠습니다. 추세를 설명하기 위해 선형 공변량을 포함하는 설계를 구성할 수 있습니다. 결과 설계는 이 추세 공변량을 고려한 최적입니다.

이 예에서는 7 개 요인에 대한 실험을 설계합니다. 선형 추세에 대해 로버스트한 16 회 런 설계를 구성합니다.

1. **파일 > 새로 만들기 > 데이터 테이블**을 선택합니다.
2. 열 1 을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다.
3. 열 이름을 Run Order 로 변경합니다.
4. **데이터 초기화** 옵션 목록에서 **시퀀스 데이터**를 선택합니다.
5. **끝** 옆에 16 을 입력합니다.
6. **확인**을 클릭합니다.

   1 부터 16 까지 연속된 정수가 데이터 테이블에 입력되었습니다.
7. **DOE > 사용자 설계**를 선택합니다.
8. **요인 추가 > 공변량**을 클릭합니다.
9. Run Order 를 선택하고 **확인**을 클릭합니다.
10. **N 개 요인 추가** 옆에 7 을 입력합니다.
11. **요인 추가 > 연속형**을 클릭합니다.

그림 5.68 반응 및 요인 개요



12. **계속**을 클릭합니다 .
13. **별칭 항** 개요를 엽니다 .
14. 목록에서 모든 효과를 선택하고 **항 제거**를 클릭합니다 .

    이렇게 하면 상관 색상 맵에서 교호작용 효과를 생략하고 주효과만 남습니다 .

    참고 : 15 단계에서 " 난수 시드값 " 을 설정하고 16 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

15. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 1234 를 입력하고 **확인**을 클릭합니다 .
16. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택하고 100 을 입력한 후 **확인**을 클릭합니다 .
17. **설계 생성**을 클릭합니다 .

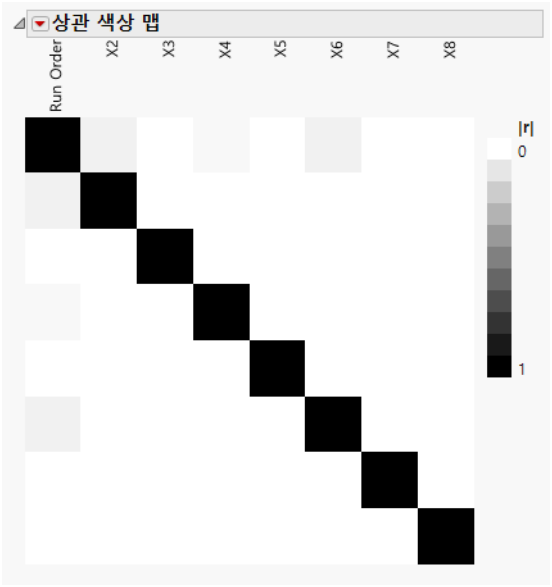
그림 5.69 설계 개요

설계

런	Run Order	X2	X3	X4	X5	X6	X7	X8
1	7	-1	-1	1	1	-1	1	1
2	1	-1	1	-1	-1	1	1	-1
3	4	1	1	1	-1	-1	-1	1
4	12	1	1	-1	1	-1	1	-1
5	15	-1	1	-1	-1	-1	-1	1
6	8	1	-1	1	-1	1	1	1
7	11	-1	1	1	-1	-1	1	-1
8	3	1	-1	1	-1	-1	-1	-1
9	5	-1	-1	-1	1	-1	1	1
10	13	-1	-1	1	1	1	-1	-1
11	16	1	1	1	1	1	1	1
12	14	-1	-1	-1	-1	1	-1	1
13	10	1	-1	-1	1	-1	-1	-1
14	9	1	-1	-1	-1	1	1	-1
15	2	1	1	-1	1	1	-1	1
16	6	-1	1	1	1	1	-1	-1

18. **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

그림 5.70 Run Order 와의 절대 상관을 보여 주는 색상 맵



색상 맵은 다음을 보여 줍니다 .

- 7 개 연속형 요인 (X2 ~ X8) 은 서로 직교합니다 .
- 시간 선형 추세 변수인 Run Order 와 X2 ~ X8 의 절대 상관은 극단적으로 낮습니다 .

19. **설계 평가 > 추정 효율도** 개요를 엽니다 .

그림 5.71 추정 효율도 개요

추정 효율도

항	CI 길이의 부분적 증가	추정값의 상대 표준 오차
절편	0	0.25
Run Order	0.65	0.412
X2	0.006	0.252
X3	4e-4	0.25
X4	0	0.25
X5	0.003	0.251
X6	4e-4	0.25
X7	4e-4	0.25
X8	0.003	0.251

X2 ~ X8 과 Run Order 의 절대 상관이 작으므로 이상적인 직교 설계에 비해 신뢰 구간 길이의 증가가 매우 작습니다 . X2 ~ X8 의 신뢰 구간 길이 증가는 모두 약 0.1% 미만입니다 .

이 예에서 런 순서 요인은 요인 효과와 거의 직교합니다. 경우에 따라 런 순서 요인과 다른 요인 간의 상관이 상당히 클 수 있습니다. 이런 상황에서도 런 순서를 요인으로 포함하면 선형 추세 효과가 설명됩니다. 또한 런 순서를 포함하면 다른 요인 효과를 더 정확하게 추정할 수 있습니다.

랜덤화 제한이 있는 실험

사용자 설계 플랫폼은 D- 최적 또는 I- 최적인 분할구 , 분할 - 분할구 및 이원 분할구 (조각구) 설계를 구성합니다 . 이러한 설계를 구성하는 방법에 대한 자세한 내용은 Goos(2002) 에서 확인하십시오 .

- " 분할구 실험 "
- " 이원 분할구 실험 "

분할구 실험

분할구 설계는 농업에서 시작되었지만 제조 및 엔지니어링 연구에서 매우 흔합니다. 분할구 실험에서 변경하기 힘든 요인은 하나의 주구와 다음 주구 사이에서만 재설정됩니다. 주구는 하위구로 나뉘며, 변경하기 쉬운 요인의 수준이 각 하위구에 무작위로 할당됩니다.

이 섹션의 예는 Kowalski, Cornell, and Vining(2002) 에서 채택되었습니다 . 자동차 시트 커버를 만드는 데 사용되는 비닐 두께에 대한 5 개 요인의 효과를 연구하려고 합니다 . 실험의 반응과 요인은 다음과 같습니다 .

- 반응은 생산되는 비닐의 **두께**입니다 . **두께**를 최대화하는 것이 목표입니다 . 두께 값의 하한은 10 입니다 .
- 주구 요인은 압출률 (extrusion rate) 과 건조 온도 (temperature) 입니다 . 이러한 요인은 공정 변수이며 변경하기가 힘듭니다 .
- 하위구 요인은 비율 (m1, m2, m3) 의 합이 1 인 세 가지 가소제입니다 . 이러한 요인은 혼합 성분입니다 .

실험 예산을 고려하면 주구 요인의 7개 설정을 실행할 수 있습니다. 각 주구마다 하위구 요인에 대한 런을 4회 수행할 수 있습니다. 따라서 총 28회 런이 가능합니다.

설계 생성

1. **DOE > 사용자 설계**를 선택합니다.
2. "반응 이름" 아래의 Y 를 두 번 클릭하고 thickness 를 입력합니다.

   "최대화" 로 설정된 기본 목표를 유지합니다.
3. "하한" 에 10 을 입력합니다.

요인을 수동으로 추가하려면 4단계 ~ 11단계의 작업을 수행합니다. 또는 저장된 테이블에서 요인을 불러오려면 "사용자 설계" 의 빨간색 삼각형에서 **요인 불러오기**를 선택합니다. Design Experiment 폴더의 Vinyl Factors.jmp 샘플 데이터 테이블을 엽니다. **요인 불러오기**를 선택한 경우에는 4단계 ~ 11단계의 작업을 건너뛰십시오.

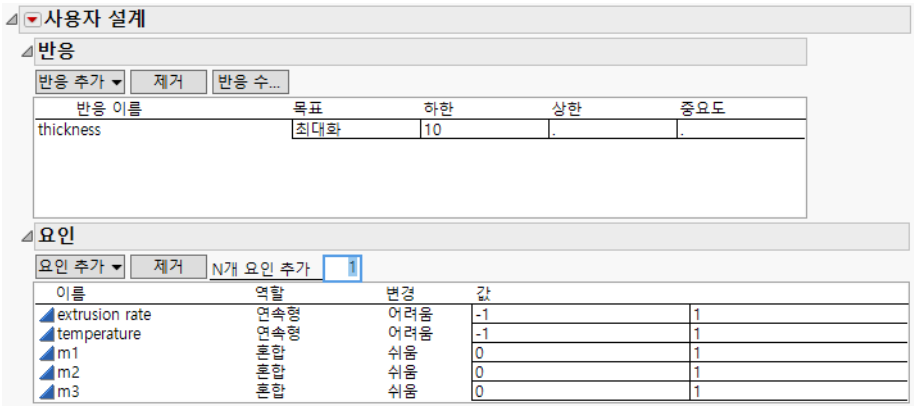
4. **N 개 요인 추가** 옆에 2 를 입력합니다.
5. **요인 추가 > 연속형**을 클릭합니다.
6. 요인 이름을 extrusion rate 와 temperature 로 각각 바꿉니다.

   이러한 두 요인의 "값" 은 기본값 –1 과 1 을 유지합니다.
7. extrusion rate 와 temperature 둘 다에 대해 **쉬움**을 클릭하고 **어려움**을 선택합니다.

   이렇게 하면 extrusion rate 와 temperature 가 주구 요인으로 정의됩니다.
8. **N 개 요인 추가** 옆에 3 을 입력합니다.
9. **요인 추가 > 혼합**을 클릭합니다.
10. 세 가지 혼합 요인의 이름을 m1, m2, m3 으로 각각 바꿉니다.

    세 요인의 "값" 은 기본값 0 과 1 을 유지합니다.

그림 5.72 반응 및 요인 개요



11. **계속**을 클릭합니다.

12. **교호작용 > 2 차**를 클릭합니다 .
13. **확인**을 클릭하여 정보 메시지를 해제합니다 .

    " 주구 수 " 의 기본값은 3 입니다 .
14. **주구 수** 옆에 7 을 입력합니다 .
15. **사용자 지정** 옆에 28 을 입력합니다 .

> **참고 :** 16 단계에서 " 난수 시드값 " 을 설정하고 17 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

16. (선택 사항) "사용자 설계 "의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
17. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 5 를 입력하고 **확인**을 클릭합니다 .
18. **설계 생성**을 클릭합니다 .

그림 5.73 설계 개요

설계

런	주구	extrusion rate	temperature	m1	m2	m3
1	1	1	1	1	0	0
2	1	1	1	0.5	0.5	0
3	1	1	1	0	0.5	0.5
4	1	1	1	0	0	1
5	2	-1	1	0	0	1
6	2	-1	1	0	1	0
7	2	-1	1	0.5	0	0.5
8	2	-1	1	1	0	0
9	3	-1	-1	0	1	0
10	3	-1	-1	0.4	0	0.6
11	3	-1	-1	0	0.5	0.5
12	3	-1	-1	1	0	0
13	4	1	-1	0	0.5	0.5
14	4	1	-1	1	0	0
15	4	1	-1	0	1	0
16	4	1	-1	0	0	1
17	5	1	1	1	0	0
18	5	1	1	0	1	0
19	5	1	1	0.5	0	0.5
20	5	1	1	0	0	1
21	6	-1	-1	1	0	0
22	6	-1	-1	0	1	0
23	6	-1	-1	0.5	0.5	0
24	6	-1	-1	0	0	1
25	7	1	-1	0	0	1
26	7	1	-1	0.6	0	0.4
27	7	1	-1	0	1	0
28	7	1	-1	0.560358	0.439642	0

주구 요인 extrusion rate 와 temperature 는 주구 요인의 수준에 따라 7 번 재설정됩니다 . 주구의 각 수준 내에서 혼합 성분 m1, m2 및 m3 에 대한 설정이 무작위로 할당됩니다 .

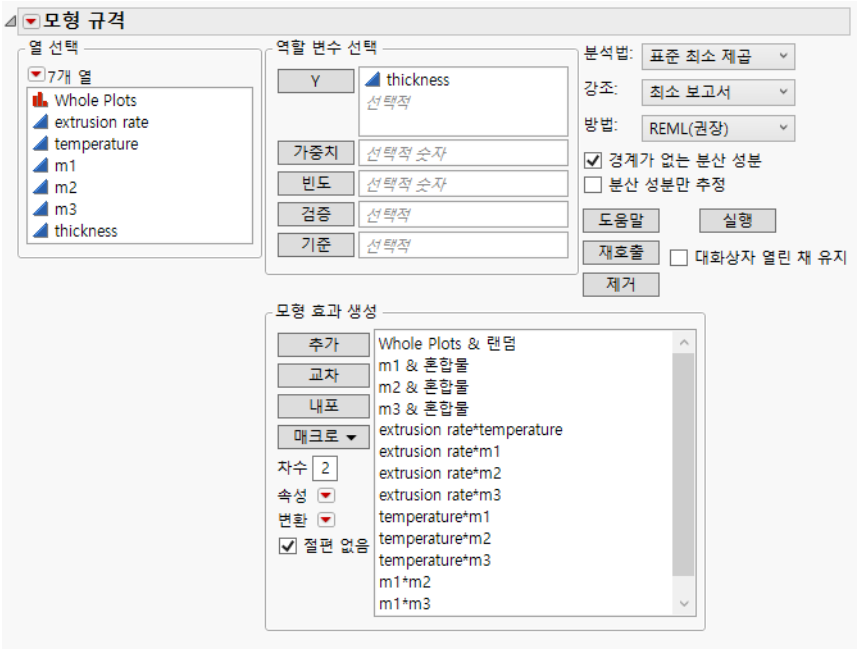
결과 분석

Vinyl Data.jmp 샘플 데이터 테이블에는 이전 버전의 JMP 에서 생성된 설계를 사용한 실험 결과가 포함되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Vinyl Data.jmp 를 엽니다 .

   이 샘플 데이터 테이블에는 28 회 런과 반응 값이 있습니다 . 사용자 설계 플랫폼을 사용하여 생성한 테이블의 설계 설정과 Vinyl Data.jmp 설계에 사용된 설정이 다를 수 있습니다 .
2. 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

그림 5.74 모형 적합 창



" 모형 적합 " 창에서 다음을 확인할 수 있습니다 .

- Whole Plots 요인에 " 임의 효과 "(**& 랜덤**) 라는 속성이 있습니다 . 이는 Whole Plots 의 수준이 랜덤하게 구현됨을 나타냅니다 . 또한 관련된 오차 항이 있습니다 .
- 분석 방법은 **REML(권장)** 입니다 . 모형에 임의 효과가 포함되어 있으므로 이 방법이 지정됩니다 . REML 모형에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

팁 : JMP Pro 사용자는 "모형 적합" 창에서 "분석법"을 "혼합 모형"으로 변경할 수 있습니다.

3. **실행**을 클릭합니다 .

그림 5.75 분할구 분석 결과

모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|
m1	9.4464782	0.989604	7.99	9.55	<.0001*
m2	5.378571	1.120267	10.35	4.80	0.0007*
m3	5.5546291	1.030955	8.786	5.39	0.0005*
extrusion rate*temperature	-1.866921	0.740748	2.858	-2.52	0.0903
extrusion rate*m1	-0.647611	0.958111	7.261	-0.68	0.5200
extrusion rate*m2	0.2944425	0.997579	8.162	0.30	0.7752
extrusion rate*m3	2.6611621	0.979087	7.733	2.72	0.0272*
temperature*m1	-1.8151	0.961061	7.328	-1.89	0.0990
temperature*m2	1.2044413	0.993319	8.065	1.21	0.2596
temperature*m3	-1.498647	0.97904	7.731	-1.53	0.1657
m1*m2	-6.458845	5.83806	13.36	-1.11	0.2881
m1*m3	8.5532871	5.607183	13.15	1.53	0.1508
m2*m3	10.575333	5.226599	13.04	2.02	0.0640

▷ 임의 효과 예측

REML 분산 성분 추정값

임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
Whole Plots	0.6350069	2.476748	2.9711536	-3.346606	8.3001021	0.4045	38.838
잔차		3.9003483	1.5898101	2.0073487	10.607283		61.162
합계		6.3770963	3.1515558	2.9308168	22.945461		100.000

-2*로그 우도 = 82.032495402
참고: 합계는 양의 분산 성분의 합입니다.
음수 추정값을 포함한 합계 = 6.3770963

"모수 추정값" 보고서는 세 가지 혼합 성분 및 extrusion rate*m3 교호작용이 0.05 수준에서 유의함을 보여 줍니다.

"REML 분산 성분 추정값" 보고서는 **주구**와 관련된 분산 성분이 2.476748임을 나타냅니다. 이는 총 변동의 38.838%입니다. 결과적으로 주구 반복과 관련된 오차 항이 잔차(내부) 오차 항보다 작습니다.

이원 분할구 실험

이원 분할구(조각구 또는 분할 블록이라고도 함) 설계는 두 개의 분할구 성분으로 구성됩니다. 산업 분야에서 첫 번째 처리 단계의 재료나 실험 단위를 두 번째 처리 단계에 일괄적으로 전달할 때 이러한 설계가 수행됩니다. 이원 분할구 설계를 사용하려면 단계 간 단위를 재정렬할 수 있어야 합니다.

첫 번째 처리 단계 후에 배치를 하위 배치로 나눌 수 있어야 합니다. 두 번째 단계 처리 요인은 이러한 하위 배치에 무작위로 적용됩니다. 두 번째 단계의 특정 실험 설정이 있는 경우 해당 설정에 할당된 모든 하위 배치를 동시에 처리할 수 있습니다. 두 번째 처리 단계 후에 실험 단위에 추가 요인을 적용할 수 있습니다.

분할-분할구 설계와 달리 두 번째 단계 요인이 첫 번째 단계 요인에 **내포되지 않습니다**. 첫 번째 단계 후 배치가 나뉘어져 새 배치가 만들어집니다. 따라서 첫 번째 단계 요인과 두 번째 단계 요인 둘 다 전체 배치에 적용됩니다.

두 단계의 요인이 동등하게 변경하기 힘들 수 있지만 JMP는 첫 번째 단계 요인을 **매우 변경하기 힘든** 것으로 나타내고 두 번째 단계 요인을 **변경하기 힘든** 것으로 나타내 이러한 요인을 구분합니다. 두 번째 처리 단계 후 실험 단위에 적용된 추가 요인은 변경하기 **쉬운** 것으로 간주됩니다.

이원 분할구 설계 시나리오

이 예는 배터리 셀의 OCV(개회로 전압)를 개선하기 위한 실험을 기반으로 합니다(Vivacqua and Bisgaard, 2004). 셀이 저절로 방전되는 것을 방지하기 위해 OCV 를 최소화해야 합니다.

배터리 셀은 다음과 같은 두 단계의 처리 과정을 거칩니다 .

- 첫 번째 단계 : 배터리를 2000 개씩 일괄 처리하는 연속 조립 공정
- 두 번째 단계 : 4000 개 배터리를 수용할 수 있는 공간에서 5 일 주기 시간의 큐어링 공정

다음과 같은 여섯 가지 2 수준 연속형 요인에 대해 연구하려고 합니다.

- 조립 공정에 적용되는 네 개의 요인 (A1, A2, A3, A4). 첫 번째 단계 요인에 대해 실행할 수 있는 시도 수는 16 회입니다 .
- 큐어링 공정에 적용되는 두 개의 요인 (C5, C6). 큐어링에는 5 일 주기 시간이 필요하므로 두 번째 단계 요인에 대해 실행할 수 있는 주기는 6 회 (30 일) 뿐입니다 . 6 회의 큐어링 주기를 사용하면 큐어링 설정을 부분 반복하여 큐어링 효과를 검정할 수 있습니다 .

첫 번째 단계 요인과 두 번째 단계 요인은 둘 다 변경하기 힘들기 때문에 분할구 두 개를 제안합니다. 그러나 첫 번째 단계 실험에서 2,000 개 배터리를 각각 500 개의 배터리로 구성된 하위 배치로 나눌 수 있습니다. 이러한 하위 배치 중 8 개를 무작위로 선택하여 큐어링 공간에서 동시에 처리할 수 있습니다.

이 실험에는 48 개의 실험 단위가 있습니다. 첫 번째 단계와 두 번째 단계의 요인이 **교차**된다는 것에 유의하십시오.

설계 생성

이원 분할구 실험을 설계하려면

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 OCV 를 입력합니다 .
3. " 목표 " 아래의 **최대화**를 클릭한 후 **최소화**를 선택합니다 .
4. 요인을 수동으로 추가하려면 5 단계 ~ 10 단계의 작업을 수행합니다 . 저장된 테이블에서 요인을 불러오려면 다음을 수행하십시오 .

   a." 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

   b. Design Experiment 폴더의 Battery Factors.jmp 샘플 데이터 테이블을 엽니다 .

   c. 11 단계으로 진행합니다 .
5. **N 개 요인 추가** 옆에 6 을 입력합니다 .
6. **요인 추가 > 연속형**을 클릭합니다 .
7. 요인 이름을 A1, A2, A3, A4, C1, C2 로 각각 바꿉니다 .

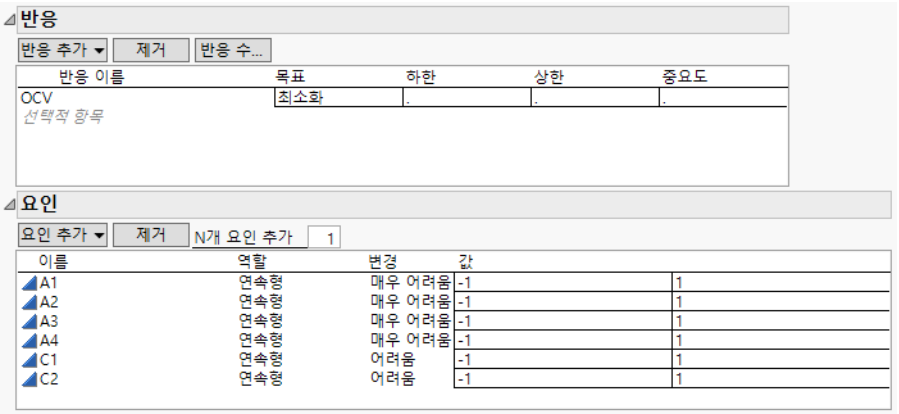
   이러한 요인의 " 값 " 은 기본값 -1 과 1 을 유지합니다 .

8. A1, A2, A3 및 A4 요인에 대해 각각 " 변경 " 아래에서 **쉬움**을 클릭한 후 **매우 어려움**으로 변경합니다 .

   첫 번째 단계와 두 번째 단계의 요인을 구분하기 위해 첫 번째 단계 요인의 " 변경 " 을 " 매우 어려움 " 으로 지정하고 두 번째 단계 요인의 " 변경 " 을 " 어려움 " 으로 지정합니다 .

9. C1 및 C2 요인에 대해 각각 " 변경 " 아래에서 **쉬움**을 클릭한 후 **어려움**으로 변경합니다 .

그림 5.76 반응 및 요인 개요



10. **계속**을 클릭합니다 .
11. " 모형 " 개요에서 **교호작용 > 2 차**를 선택합니다 .
12. " 설계 생성 " 개요에서 **변경하기 힘든 요인은 매우 변경하기 힘든 요인과 독립적으로 변화할 수 있습니다 .** 옵션을 선택합니다 .

    자세한 내용은 그림 5.77 에서 확인하십시오 . 이 옵션을 선택하면 이원 분할구 설계가 생성됩니다 . 이 옵션을 선택하지 않으면 두 수준에서 요인이 내포된 분할 - 분할구 설계로 처리됩니다 .

13. **주구 수**에 16 을 입력합니다 .

    이 값은 첫 번째 단계 요인에 대해 실행할 수 있는 시도 수입니다 .

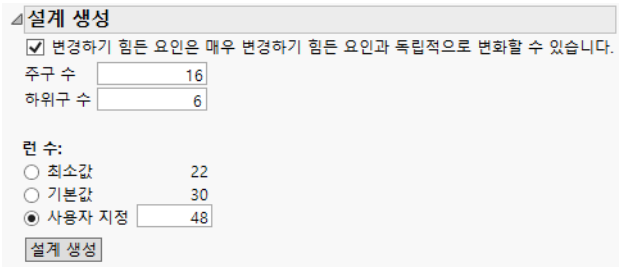
14. **하위구 수**에 6 을 입력합니다 .

    이 값은 두 번째 단계 요인에 대해 실행할 수 있는 시도 수입니다 .

15. " 런 수 " 아래에서 **사용자 지정** 옆에 48 을 입력합니다 .

    이 값은 총 실험 단위 수입니다 .

그림 5.77 설계 생성 개요



참고 : 16 단계에서 " 난수 시드값 " 을 설정하고 17 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

16. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 1866762673 을 입력하고 **확인**을 클릭합니다 .
17. (선택 사항) " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 21 을 입력하고 **확인**을 클릭합니다 .
18. **설계 생성**을 클릭합니다 .
19. **테이블 생성**을 클릭합니다 .

그림 5.78 설계 테이블 (일부분)

	주구	하위구	A1	A2	A3	A4	C1	C2	OCV
1	1	1	-1	1	-1	1	-1	1	•
2	1	2	-1	1	-1	1	1	1	•
3	1	3	-1	1	-1	1	1	-1	•
4	2	4	1	1	-1	1	1	1	•
5	2	5	1	1	-1	1	-1	1	•
6	2	6	1	1	-1	1	-1	-1	•
7	3	1	-1	-1	-1	1	-1	1	•
8	3	2	-1	-1	-1	1	1	1	•
9	3	3	-1	-1	-1	1	1	-1	•
10	4	4	-1	-1	-1	-1	1	1	•
11	4	5	-1	-1	-1	-1	-1	1	•
12	4	6	-1	-1	-1	-1	-1	-1	•
13	5	1	1	1	-1	-1	-1	1	•
14	5	2	1	1	-1	-1	1	1	•
15	5	3	1	1	-1	-1	1	-1	•
16	6	4	1	-1	-1	-1	1	1	•

설계 테이블은 **주구**의 16 개 수준을 보여 줍니다. **주구**의 각 수준에 대해 네 가지 조립 요인의 설정이 일정합니다. **주구**의 각 수준에서 500 개의 배터리로 구성된 3 개의 배치 (**하위구**) 가 큐어링 요인의 설정에 무작위로 할당됩니다. 큐어링 조건 집합 두 개가 반복됩니다 (C1 = -1, C2 = 1 과 C1 = 1, C2 = 1). 이를 확인하려면 C1 및 C2 열을 선택하고 머리글 영역을 마우스 오른쪽 버튼으로 클릭한 후 **정렬 > 오름차순**을 선택하십시오 .

결과 분석

Battery Data.jmp 샘플 데이터 테이블에는 생성한 설계의 실험 결과가 포함되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Battery Data.jmp 를 엽니다 .
2. 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   " 모형 적합 " 창에서 다음을 확인할 수 있습니다 .

   - Whole Plots 요인에 " 임의 효과 "(**& 랜덤**) 라는 속성이 있습니다 . 이는 Whole Plots 의 수준이 랜덤하게 구현됨을 나타냅니다 . 또한 관련된 오차 항이 있습니다 .
   - Subplots 요인에도 " 임의 효과 " 속성 (**& 랜덤**) 이 있습니다 .
   - 분석 방법은 **REML(권장)** 입니다 . 모형에 임의 효과가 포함되어 있으므로 이 방법이 지정됩니다 . REML 모형에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

팁 : JMP Pro 사용자는 " 모형 적합 " 창에서 " 분석법 " 을 " 혼합 모형 " 으로 변경할 수 있습니다 .

3. **대화상자 열린 채 유지** 옵션을 선택합니다 .
4. **실행**을 클릭합니다 .

그림 5.79 전체 모형에 대한 보고서

모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|
절편	38.183196	1.673404	2.075	22.82	0.0016*
A1	-4.657571	1.093074	2.916	-4.26	0.0251*
A2	0.147742	1.093074	2.916	0.14	0.9013
A3	0.9486795	1.093074	2.916	0.87	0.4510
A4	4.0716925	1.190175	3.214	3.42	0.0376*
C1	-1.827487	1.606207	0.868	-1.14	0.4804
C2	-15.73288	1.35446	0.719	-11.62	0.1065
A1*A2	0.1289012	1.080526	2.784	0.12	0.9131
A1*A3	-1.114015	1.080526	2.784	-1.03	0.3837
A1*A4	-1.92398	1.176342	3.04	-1.64	0.1993
A1*C1	-3.271232	0.480531	20.65	-6.81	<.0001*
A1*C2	3.737258	0.478635	20.09	7.81	<.0001*
A2*A3	0.3768179	1.080526	2.784	0.35	0.7520
A2*A4	1.4693535	1.176342	3.04	1.25	0.2992
A2*C1	1.9072313	0.480531	20.65	3.97	0.0007*
A2*C2	-0.161179	0.478635	20.09	-0.34	0.7398
A3*A4	0.0764369	1.176342	3.04	0.06	0.9522
A3*C1	0.1761619	0.480531	20.65	0.37	0.7176
A3*C2	-0.445242	0.478635	20.09	-0.93	0.3633
A4*C1	0.5306856	0.511896	20.84	1.04	0.3118
A4*C2	1.3506119	0.502894	20.12	2.69	0.0142*
C1*C2	-0.425326	1.606207	0.868	-0.26	0.8401

▷ 임의 효과 예측

REML 분산 성분 추정값

임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
Whole Plots	1.6095523	15.451509	15.876576	-15.66601	46.569027	0.3304	45.968
Subplots	0.8919389	8.562507	16.314188	-23.41271	40.537728	0.5997	25.473
잔차		9.5998804	3.0299487	5.6239301	19.985723		28.559
합계		33.613897	16.824615	15.325965	123.59174		100.000

-2*로그 우도 = 227.810084

참고: 합계는 양의 분산 성분의 합입니다.

음수 추정값을 포함한 합계 = 33.613897

" 모수 추정값 " 보고서는 이원 교호작용 4 개 (A1*C1, A1*C2, A2*C1, A4*C2) 와 주효과 두 개 (A1, A4) 가 0.05 수준에서 유의함을 나타냅니다 .

5. Battery Data.jmp 의 테이블 패널에서 **Reduced Model 1** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   그러면 유의하지 않은 교호작용이 제거된 " 모형 적합 " 창이 열립니다 . 남은 효과는 모든 주효과와 이원 교호작용 4 개 (A1*C1, A1*C2, A2*C1, A4*C2) 입니다 . 이 모형 축소 방법은 보수적입니다 .

6. **실행**을 클릭합니다 .

그림 5.80 예비 축소 모형에 대한 보고서

모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|
절편	38.249531	1.373127	7.723	27.86	<.0001*
A1	-4.723906	1.175023	8.5	-4.02	0.0034*
A2	0.0497917	1.165418	8.228	0.04	0.9669
A3	0.7560417	1.165418	8.228	0.65	0.5342
A4	4.2328205	1.207654	9.16	3.50	0.0065*
C1	-2.195626	0.866042	1.083	-2.54	0.2234
C2	-15.79922	0.840903	1.008	-18.79	0.0331*
A1*C1	-3.283225	0.452505	25.17	-7.26	<.0001*
A1*C2	3.8035937	0.4498	24.42	8.46	<.0001*
A2*C1	1.8797279	0.452505	25.17	4.15	0.0003*
A4*C2	1.3877006	0.482556	24.95	2.88	0.0081*

주효과 C2 는 이제 0.05 수준에서 유의합니다 (Prob>|t| = 0.0331).

7. " 모형 적합 " 창에서 A3 을 제거합니다 .

   주효과 중 A3 만 유의하지 않고 이원 교호작용에 포함되지 않습니다 .

8. **실행**을 클릭합니다 .

그림 5.81 축소 모형에 대한 보고서

모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|
절편	38.249531	1.362184	6.639	28.08	<.0001*
A1	-4.723906	1.137313	9.131	-4.15	0.0024*
A2	0.0497917	1.127389	8.818	0.04	0.9658
A4	4.2131272	1.173862	9.921	3.59	0.0050*
C1	-2.166553	0.899138	1.037	-2.41	0.2432
C2	-15.79922	0.874257	0.963	-18.07	0.0389*
A1*C1	-3.282201	0.452234	25.24	-7.26	<.0001*
A1*C2	3.8035938	0.449735	24.43	8.46	<.0001*
A2*C1	1.8843473	0.452234	25.24	4.17	0.0003*
A4*C2	1.3851686	0.4823	24.91	2.87	0.0082*

▷ 임의 효과 예측

REML 분산 성분 추정값

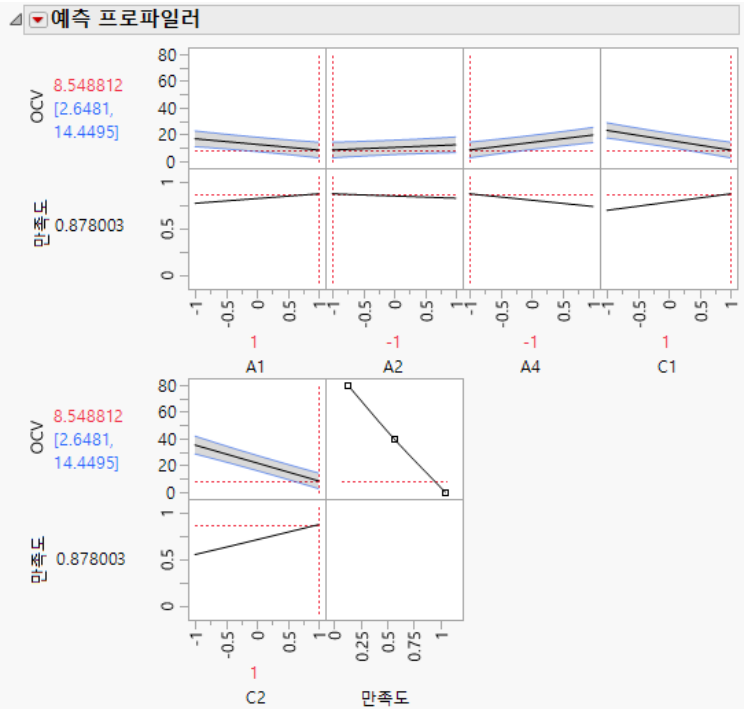
임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
Whole Plots	2.0231563	17.459497	9.7154555	-1.582446	36.50144	0.0723	60.025
Subplots	0.3473621	2.9976761	5.8732929	-8.513766	14.509119	0.6098	10.306
잔차		8.6298312	2.4692519	5.2816799	16.588508		29.669
합계		29.087005	9.1941655	17.028461	60.633509		100.000

-2*로그 우도 = 254.17861349
참고: 합계는 양의 분산 성분의 합입니다.
음수 추정값을 포함한 합계 = 29.087005

"REML 분산 성분 추정값 " 보고서는 **주구**와 관련된 분산 성분이 **하위구**의 분산 성분보다 약 6 배 정도 크다는 것을 보여 줍니다 . 즉 , 조립 공정이 큐어링 공정보다 가변성이 더 크다는 것을 나타냅니다 . 또한 내부 (잔차) 오차가 **하위구**의 오차보다 큽니다 . 변동을 줄이려면 조립 공정 및 배터리 간 차이에 중점을 두어야 합니다 .

9. " 반응 OCV" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다 .
10. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 5.82 OCV 최소화 설정을 사용한 예측 프로파일러



프로파일러는 활성으로 식별된 5 개 요인 및 OCV 를 최소화하는 설정을 보여 줍니다 .

로버스트 공정 및 제품 설계를 위한 실험

공정 또는 제품은 잡음 요인의 영향을 받는 경우가 많습니다. 이러한 요인은 제어하기가 어렵고 비용 효율성이 낮습니다. 잡음 요인의 영향을 가장 적게 받는 제어 가능한 요인의 공정 또는 제품 설정을 찾아야 합니다. 과거에는 Taguchi 설계가 잡음 요인이 있을 때 실험할 수 있는 방법을 제공했습니다. 실험 중에 잡음 요인을 제어해야 합니다.

Taguchi 설계의 대안으로 결합 배열 및 혼합 해상도 설계(Borror and Montgomery, 2000)를 사용할 수 있습니다. 혼합 해상도 설계는 제어 요인과 잡음 요인을 모두 포함하는 설계를 생성하는 데

사용됩니다. 이 설계의 목표는 제어 변수에 대한 로버스트 설정을 찾는 것입니다. 핵심은 특정 효과를 추정할 수 있는지 여부입니다. 특히 제어 변수의 주효과, 잡음 변수의 주효과, 제어 간 교호작용, 잡음*제어 교호작용 및 제어 변수의 2차 항을 추정하는 데 관심이 있습니다.

사용자 설계 도구를 사용하여 로버스트 공정 설정 연구를 위한 설계를 생성할 수 있습니다.

이 예에서는 세 가지 잡음 요인이 있을 때 네 가지 제어 설정에 대한 최적 설정을 찾으려고 합니다.

1. **DOE > 사용자 설계**를 선택합니다.
2. "요인" 개요에서 **N 개 요인 추가** 옆에 4를 입력합니다.
3. **요인 추가 > 연속형**을 클릭합니다.
4. "요인" 개요에서 **N 개 요인 추가** 옆에 3을 입력합니다.
5. **요인 추가 > 연속형**을 클릭합니다.
6. X5를 두 번 클릭한 후 Z1로 변경하여 잡음 요인을 지정합니다. 계속해서 X6과 X7도 Z2와 Z3으로 변경합니다.

그림 5.83 제어 요인 4 개와 잡음 요인 3 개가 있는 설계의 요인 개요

요인

요인 추가 | 제거 | N개 요인 추가 1

이름	역할	변경	값	
X1	연속형	쉬움	-1	1
X2	연속형	쉬움	-1	1
X3	연속형	쉬움	-1	1
X4	연속형	쉬움	-1	1
Z1	연속형	쉬움	-1	1
Z2	연속형	쉬움	-1	1
Z3	연속형	쉬움	-1	1

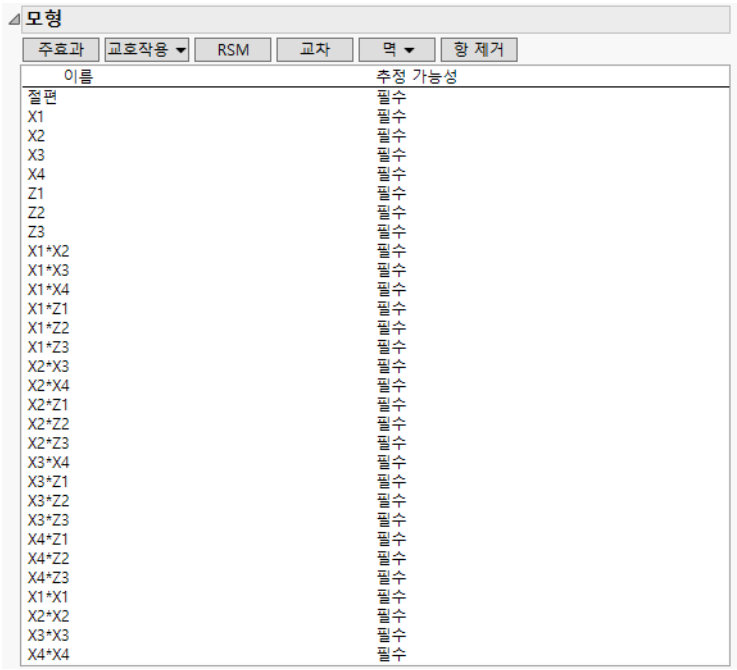
7. 계속을 클릭합니다.

이제, 잡음 요인의 영향을 가장 적게 받는 제어 요인의 최적 설정을 탐색하기 위한 모형을 설정합니다.

8. 모형에는 기본적으로 모든 주효과가 포함됩니다.
9. **교호작용 > 2 차**를 클릭하여 모든 이원 교호작용을 추가합니다.

   이렇게 하면 제어 간 교호작용, 제어 * 잡음 교호작용 및 잡음 간 교호작용이 입력됩니다. 잡음 요인 간의 교호작용은 추정하고 싶지 않습니다.
10. 목록 맨 아래로 스크롤하여 Z1*Z2, Z1*Z3 및 Z2*Z3 을 강조 표시한 후 "항 제거"를 클릭합니다. 그러면 잡음 간 교호작용이 제거됩니다.
11. "요인" 개요에서 X1, X2, X3, X4 를 선택합니다.
12. "모형" 개요에서 "멱 > 2 차"를 클릭하여 제어 변수 2 차 항을 추가합니다.

그림 5.84 잡음 요인이 있는 설계의 모형 개요



13. **중앙점 수 :** 텍스트 상자에 2 를 입력하여 중앙점 두 개를 추가합니다 .
14. **설계 생성**을 클릭합니다 .

세 가지 잡음 요인이 있을 때 제어 요인을 평가하고 최적화할 수 있는 36 회 런 설계가 생성됩니다. 분석 목표는 중요한 효과를 판별한 후 잡음 요인의 영향을 가장 적게 받는 제어 요인에 대한 설정을 선택하는 것입니다.

6 장

설계 확대
기존 설계에 추가 런 추가

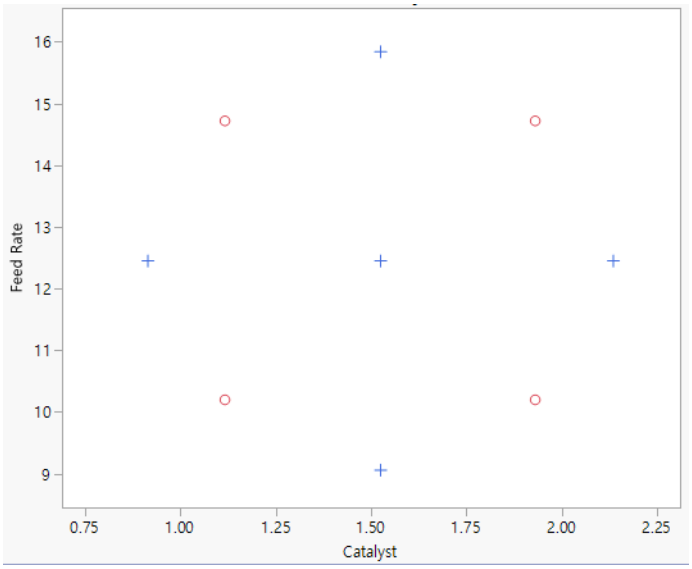
설계 확대 플랫폼을 사용하면 기존 실험 설계에 런을 추가할 수 있습니다. 실험은 대개 순차적 공정이며 설계된 실험을 확대하면 단일 설계의 모호성을 해결하는 데 도움이 됩니다.

설계 확대 플랫폼은 기존 설계 테이블에 대해 전체 설계를 최적화하는 추가 런을 구성합니다. 다음과 같은 목표를 달성하기 위해 런을 추가할 수 있습니다.

- 설계를 지정된 횟수만큼 반복합니다.
- 중앙점을 추가합니다.
- 설계 접기를 생성합니다.
- 중앙점과 함께 축 점을 추가하여 선별 설계를 반응 표면 설계로 변환합니다.
- 설계에 공간 채움 점을 추가합니다.
- 특정 모형 적합을 위해 설계에 런을 추가합니다.

설계 확대에 대한 자세한 내용은 Goos and Jones(2011)에서 확인하십시오.

그림 6.1 원래 설계 런 (+) 과 확대된 런 (o)



목차

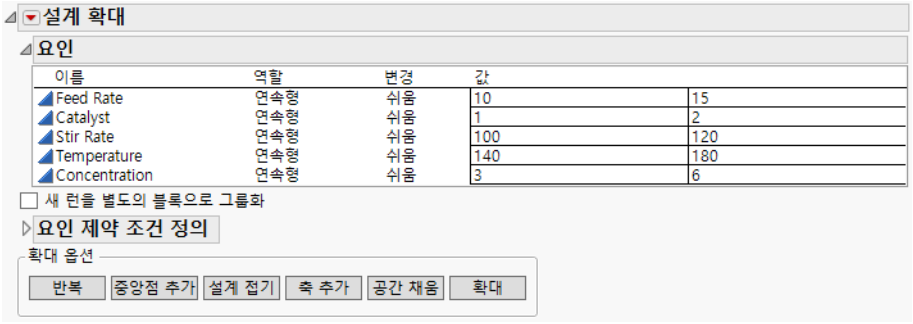
설계 확대 플랫폼을 사용한 예 195
설계 확대 플랫폼 시작 201
설계 확대 창 202
요인 202
요인 제약 조건 정의 203
확대 옵션 206
설계 확대 플랫폼 옵션 207
확대 옵션의 추가 예 210
설계 반복 210
설계에 중앙점 추가 212
설계 접기 생성 213
설계에 축 점 추가 214
공간 채움 확대 215
통계 상세 정보 216

설계 확대 플랫폼을 사용한 예

" 설계 확대 플랫폼 " 을 사용하여 설계에 추가 런을 추가합니다 . 이 예에서는 설계 확대 플랫폼을 사용하여 선별 설계 결과의 모호성을 해결하는 방법을 보여 줍니다 . 이 연구에서 화학 엔지니어는 화학 공정의 반응 백분율에 대한 5 개 요인의 효과를 조사합니다 . 원래 설계는 주효과를 적합시키기 위한 8 회 런 설계였습니다 . 이 예에서는 총 16 회 런을 사용하여 모든 2 요인 교호작용을 추정할 수 있도록 8 회 런 설계를 확대합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 8 Runs.jmp 를 엽니다 .
2. **DOE > 설계 확대**를 선택합니다 .
3. Percent Reacted 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Feed Rate, Catalyst, Stir Rate, Temperature 및 Concentration 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 6.2 반응기 예제의 요인

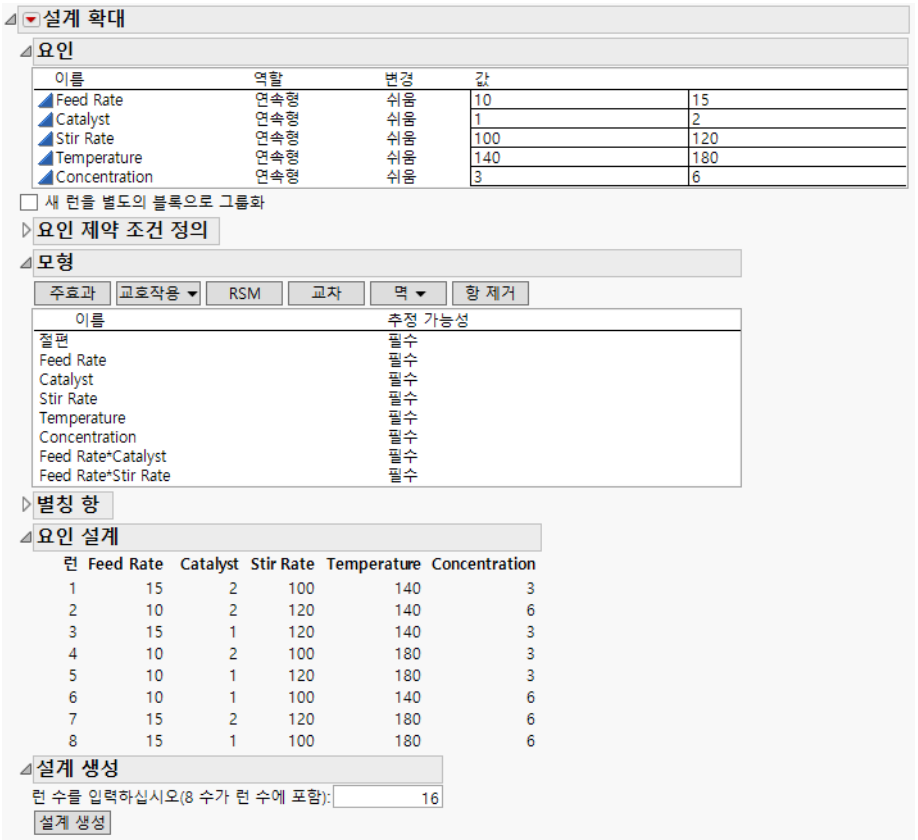


참고 : 새 런을 별도의 블록으로 그룹화를 선택하여 설계에 블록 요인을 추가할 수 있습니다 . 그러나 이 예에서는 16 회 런으로 모든 2 요인 교호작용을 추정하며 모형에 추가 블록 요인이 있으면 이를 수행할 수 없습니다 .

6. **확대**를 클릭합니다 .

그림 6.3 에 표시된 모형은 데이터 테이블의 "Model" 스크립트를 사용하여 정의됩니다 . 8 회 런을 사용하여 최대 7 개 효과를 추정할 수 있습니다 . 스크립트의 모형은 5 개의 모든 주효과 및 2 개의 교호작용 항을 포함합니다 . 모든 항은 모형에 대해 필수로 정의됩니다 . 효과 추정 가능성에 대한 자세한 내용은 " 사용자 설계 " 장의 " 모형 "(93 페이지) 에서 확인하십시오 .

그림 6.3 초기 설계 확대 플랫폼



7. " 모형 " 개요에서 **교호작용 > 2 차**를 선택합니다 .

   모든 2 요인 교호작용이 모형에 추가됩니다 . 지정된 모형에 대한 최소 런 수는 " 설계 생성 " 텍스트 편집 상자에 표시된 것과 같이 16 회입니다 .

   참고 : 8 단계에서 " 난수 시드값 " 을 설정하고 9 단계에서 " 시작 수 " 를 설정하면 그림 6.4 에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 선택 사항입니다 .

8. (선택 사항) " 설계 확대 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 282322901 을 입력하고 **확인**을 클릭합니다 .
9. (선택 사항) " 설계 확대 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 800 을 입력하고 **확인**을 클릭합니다 .
10. **설계 생성**을 클릭합니다 .
11. **설계** 표시 아이콘을 클릭하여 설계를 확인합니다 .

그림 6.4 16 회 런으로 확대된 설계

설계

런	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration
1	15	2	100	140	3
2	10	2	120	140	6
3	15	1	120	140	3
4	10	2	100	180	3
5	10	1	120	180	3
6	10	1	100	140	6
7	15	2	120	180	6
8	15	1	100	180	6
9	15	1	120	180	6
10	10	2	100	140	6
11	10	1	100	180	6
12	15	2	120	140	6
13	15	2	120	180	3
14	10	2	120	180	3
15	10	1	100	140	3
16	10	1	120	140	3

12. **테이블 생성**을 클릭하여 원래 설계와 결과 및 확대된 런을 포함하는 설계 테이블을 생성합니다 .

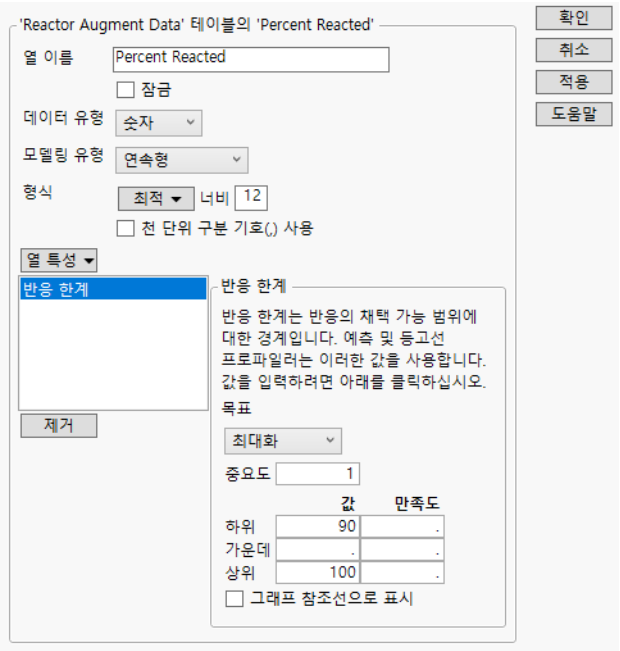
확대된 설계 분석

확대된 예제 설계를 계속 진행하면서 , 추가 실험 런을 수행하고 분석을 위해 결과를 기록했다고 가정해 보겠습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor Augment Data.jmp 를 엽니다 .

   Percent Reacted 를 최대화하는 것이 목표입니다 . 이 샘플 데이터 테이블에서 이 열의 " 반응 한계 " 열 특성이 **최대화**로 설정되어 있습니다 . 하한과 상한은 각각 90 과 100 으로 설정되어 있습니다 .

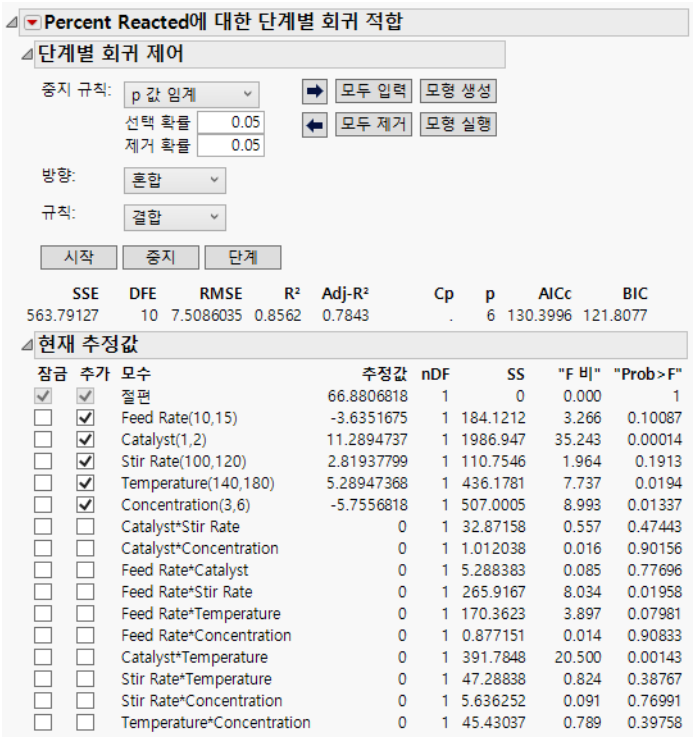
그림 6.5 "Percent Reacted" 열의 " 반응 한계 " 열 특성



2. 데이터 테이블의 테이블 패널에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   Model 스크립트는 모든 주효과 및 2요인 교호작용이 효과로 추가된 "모형 적합" 창을 엽니다.
3. 적합 분석법을 **표준 최소 제곱**에서 **단계별**로 변경한 다음 **실행**을 클릭합니다 .
4. " 단계별 회귀 제어 " 패널의 **중지 규칙** 메뉴에서 **p 값 임계**를 선택하고 **방향** 메뉴에서 **혼합**을 선택합니다 . **선택 확률**에 0.05 를 입력하고 **제거 확률**에 0.05 를 입력합니다 .
5. " 현재 추정값 " 패널에서 모든 주효과 항 (Feed Rate ~ Concentration) 에 대해 **추가** 체크박스를 선택합니다 .

그림 6.6 완료된 단계별 모형

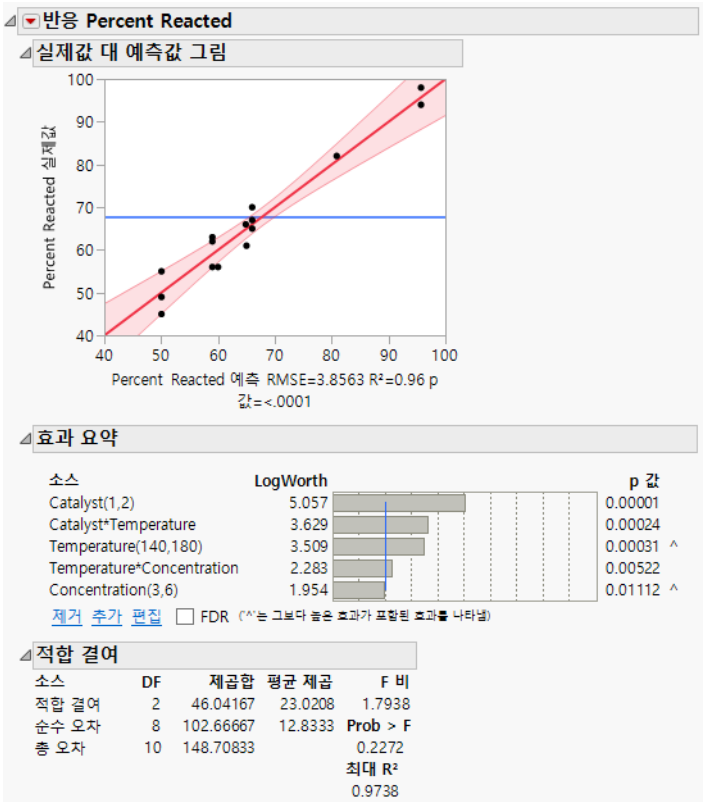


6. **시작**을 클릭합니다 .

   단계별 회귀는 주효과로 시작하고 p 값 기반 검색을 사용하여 3 개의 주효과 항과 2 개의 교호작용 항이 있는 모형을 생성합니다 . " 단계별 회귀 제어 " 패널에서 **모형 생성**을 클릭합니다 . 단계별 절차를 사용하여 생성된 모형에 대한 모형 적합 플랫폼이 시작됩니다 .

7. **실행**을 클릭합니다 .

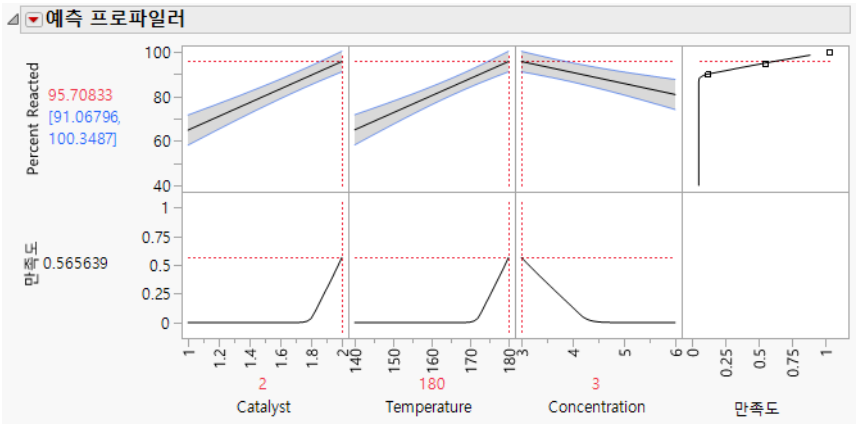
그림 6.7 예측 모형 분산 분석 및 적합 결여 검정



" 실제값 대 예측값 그림 " 은 전체 모형이 유의함을 나타냅니다 (p 값 < .0001). " 실제값 대 예측값 그림 " 과 " 적합 결여 " 검정 둘 다 모형 오지정 증거를 나타내지 않습니다 . " 효과 요약 " 보고서는 Catalyst 가 가장 유의한 효과임을 보여 줍니다 .

8. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 6.8 Percent Reacted 최대값

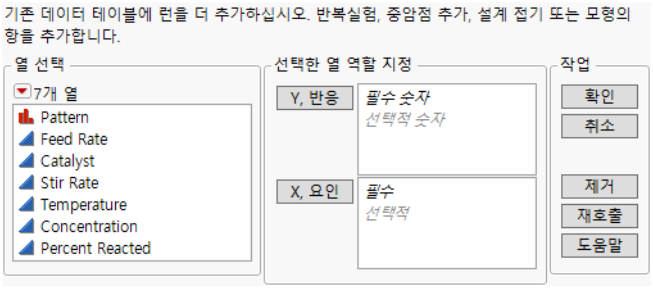


이 예측 프로파일 그림은 Catalyst 와 Temperature 설정이 높고 Concentration 설정이 낮을 때 반응 백분율이 최대가 된다는 것을 보여 줍니다. 이러한 극단적 설정에서 Percent Reacted 의 추정값은 98.38 입니다 . 만족도 프로파일은 예측 반응이 반응 한계를 벗어나는 영역에서 평탄화됩니다 .

설계 확대 플랫폼 시작

확대할 설계가 포함된 데이터 테이블에서 **DOE** > **설계 확대**를 선택하여 설계 확대 플랫폼을 시작합니다 .

그림 6.9 설계 확대 시작 창



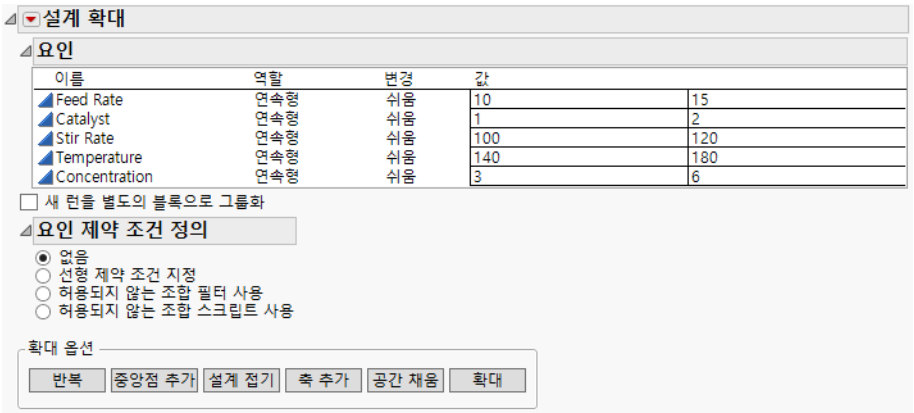
Y, 반응 하나 이상의 수치형 반응 열을 입력합니다 . 반응 입력은 필수입니다 .

X, 요인 요인 열을 입력합니다 . 요인은 모든 데이터 유형 또는 모델링 유형이 될 수 있습니다 .

설계 확대 창

초기 " 설계 확대 " 창은 " 요인 " 개요 , " 요인 제약 조건 정의 " 개요 및 " 확대 옵션 " 패널로 구성됩니다 .

그림 6.10 Reactor 8 Runs.jmp 를 사용한 초기 설계 확대 창



요인

요인은 " 설계 확대 " 시작 창에서 지정한 요인입니다 .

이름 설계 역할 열 특성이 " 랜덤 블록 " 인 요인을 제외하고 " 설계 확대 " 시작 창에 **X, 요인**으로 나열된 모든 요인을 보여 줍니다 .

역할 열 특성에 지정된 요인의 설계 역할을 지정합니다 . 요인에 " 설계 역할 " 열 특성이 없고 상수이면 " 역할 " 열에 **상수**가 나타납니다 . 그렇지 않으면 요인의 모델링 유형이 " 역할 " 열에 나타납니다 .

변경 데이터 테이블에서 " 요인 변경 " 열 특성에 지정한 요인 변경 수준을 " 쉬움 ", " 어려움 " 또는 " 매우 어려움 " 으로 나타냅니다 . 요인에 " 요인 변경 " 열 특성이 없으면 " 변경 " 이 " 쉬움 " 으로 지정됩니다 .

참고 : 요인의 " 요인 변경 " 열 특성이 " 어려움 " 또는 " 매우 어려움 " 으로 설정된 경우에는 해당 주구 요인이 " 설계 확대 " 시작 창의 **X, 요인** 목록에 포함되어야 합니다 .

값 연속형 요인의 경우 최소값과 최대값을 표시합니다 . 범주형 요인의 경우 수준을 표시합니다 .

팁 : 역할이 " 범주형 " 또는 " 상수 " 인 요인은 아래쪽 화살표 아이콘과 함께 " 이름 " 열에 나타납니다 . 아래쪽 화살표를 클릭하여 수준을 추가할 수 있습니다 . 요인이 " 상수 " 이고 범주형 모델링 유형을 사용하는 경우 여러 수준을 추가할 수 있습니다 . 요인이 " 상수 " 이고 연속형 모델링 유형을 사용하는 경우에는 수준을 하나만 추가할 수 있습니다 .

새 런을 별도의 블록으로 그룹화 원래 설계에 대한 블록과 확대된 런에 대한 블록을 사용하여 설계에 블록 요인을 추가합니다 .

요인 제약 조건 정의

"공간 채움" 또는 "확대" 옵션을 사용하여 설계를 확대하는 경우 추가된 런의 설계 공간에 대한 제한을 정의할 수 있습니다.

" 요인 제약 조건 정의 " 를 사용하여 설계 공간을 제한할 수 있습니다 . 제약 조건을 불러오거나 스크립트의 일부로 포함하지 않은 경우 **없음** 옵션이 선택됩니다 . 제약 조건을 지정하려면 다른 옵션 중 하나를 선택합니다 .

선형 제약 조건 지정 요인의 선형 조합에 대한 부등식 제약 조건을 지정합니다 . " 연속형 " 또는 " 혼합 " 역할이 설정된 요인에만 사용할 수 있습니다 . 자세한 내용은 선형 제약 조건 지정에서 확인하십시오 .

참고 : 선형 제약 조건을 포함하는 설계의 스크립트를 저장하면 해당 스크립트에 선형 제약 조건이 보다 작거나 같음 부등식 (≤) 으로 표현됩니다 .

허용되지 않는 조합 필터 사용 개별 요인의 값 제한을 기반으로 제약 조건 집합을 정의합니다 . AND 제약 조건과 OR 제약 조건을 모두 정의할 수 있습니다 . 자세한 내용은 허용되지 않는 조합 필터 사용에서 확인하십시오 .

허용되지 않는 조합 스크립트 사용 허용되지 않는 조합 및 기타 제약 조건을 스크립트 편집기 상자에 부울 JSL 표현식으로 정의합니다 . 자세한 내용은 허용되지 않는 조합 스크립트 사용에서 확인하십시오 .

참고 : 요인 제약 조건을 포함하는 설계를 분석할 때는 모형 프로파일러가 제약 조건을 적용합니다 .

선형 제약 조건 지정

설계 공간에서 연속형 요인을 독립적으로 변경할 수 없는 경우 선형 부등식 제약 조건을 지정할 수 있습니다. 선형 부등식은 허용되는 요인 수준 설정을 정의합니다.

추가를 클릭하여 하나 이상의 선형 부등식 제약 조건을 입력합니다 .

추가 설계의 모든 연속형 요인을 포함하는 선형 표현식 템플릿을 추가합니다 . 요인에 대한 계수 값을 입력한 후 선형 제약 조건을 반영하는 부등식 방향을 선택합니다 . 부등식의 오른쪽 상자에 제약 값을 지정합니다 . 제약 조건을 더 추가하려면 **추가**를 다시 클릭합니다 .

참고 : " 구 반지름 " 을 지정하여 설계 영역을 이미 제한한 경우에는 " 추가 " 옵션이 비활성화됩니다 .

마지막 제약 조건 제거 마지막 제약 조건을 제거합니다 .

제약 조건 검사 제약 조건의 일관성을 확인합니다 . 이 옵션은 중복 제약 조건을 제거하고 실행 가능성 검사를 수행합니다 . 문제가 있으면 JMP 경고가 나타납니다 . 제약 조건이 요인의 한계와 동등한 경우 " 요인 " 개요의 한계가 업데이트되었음을 알리는 JMP 경고가 나타납니다 .

허용되지 않는 조합 필터 사용

이 옵션은 허용되지 않는 조합을 쉽게 지정하기 위해 데이터 필터를 조정하여 사용합니다. 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

"필터 요인 추가" 목록에서 요인을 선택하고 **추가**를 클릭합니다. 그런 다음 슬라이더를 사용하거나(연속형 요인의 경우) 수준을 선택하여(범주형 요인의 경우) 허용되지 않는 조합을 지정합니다.

"필터 요인 추가" 메뉴의 빨간색 삼각형 옵션은 많은 플랫폼 시작 창의 "열 선택" 패널에 있는 옵션과 동일합니다. 자세한 내용은 JMP 사용의 "시작하기" 장에서 확인하십시오.

" 추가 " 를 클릭하면 " 허용되지 않는 조합 " 제어판에 선택한 요인이 표시되고 추가 제어 옵션이 제공됩니다 . 요인은 모델링 유형에 따라 다음과 같이 나타납니다 .

연속형 요인 연속형 요인의 경우 요인 설정 범위를 따라 양단 슬라이더가 나타납니다 . 슬라이더 끝을 드래그하거나 슬라이더 아래의 텍스트 값을 클릭하여 끝점을 설정함으로써 허용되지 않은 설정을 지정할 수 있습니다 . 슬라이더에서 파란색 강조 표시는 허용되지 않는 값을 나타냅니다 .

범주형 요인 범주형 요인의 경우 가능한 수준은 라벨이 지정된 블록으로 표시되거나 , 수준 수가 많으면 목록 항목으로 표시됩니다 . 허용하지 않을 수준을 선택합니다 . 여러 수준을 선택하려면 Ctrl 키를 누른 채로 선택합니다 . 해당 블록 또는 목록 항목은 강조 표시되어 허용되지 않는 수준임을 나타냅니다 . " 허용되지 않는 조합 " 패널에 범주형 요인을 추가하면 해당 요인의 수준 수가 요인 이름 뒤의 괄호 안에 표시됩니다 .

허용되지 않는 조합 옵션

제어판에는 다음과 같은 컨트롤이 있습니다 .

지우기 지정한 허용되지 않는 요인 수준 설정을 모두 지웁니다 . 선택한 요인은 지워지지 않습니다 .

다시 시작 선택한 모든 요인을 제거하고 초기 요인 목록으로 돌아갑니다 .

AND " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 AND 그룹이 됩니다 . AND 그룹 내에 지정된 요인 수준 조합은 허용되지 않습니다 .

나중에 AND 그룹에 요인을 추가하려면 그룹 개요를 클릭하여 강조 표시된 직사각형을 표시합니다 . 그런 다음 "AND" 를 선택하고 요인을 추가합니다 .

단일 요인을 제거하려면 빨간색 삼각형 메뉴에서 **삭제**를 선택합니다 .

OR " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 별도의 AND 그룹이 됩니다 . OR 로 구분된 AND 그룹의 경우 하나 이상의 AND 그룹에 지정된 조합은 허용되지 않습니다 .

요인에 대한 빨간색 삼각형 메뉴 옵션

한 요인이 여러 OR 그룹에 나타날 수 있습니다. 특정 OR 그룹에 요인이 나타나는 경우 이를 해당 요인의 인스턴스라고 합니다.

삭제 " 허용되지 않는 조합 " 패널에서 선택한 요인 인스턴스를 제거합니다 .

선택 해제 요인 인스턴스에 대한 선택을 취소합니다 .

선택 반전 요인 인스턴스에 대해 선택된 값을 선택 취소하고 이전에 선택되지 않은 값을 선택합니다 .

표시 옵션 범주형 요인에만 사용할 수 있습니다 . 표시 모양을 변경합니다 . 각 수준을 블록 , 목록 또는 단일 범주로 표시하거나 각 값 옆에 체크박스를 추가하는 옵션을 사용할 수 있습니다 .

찾기 범주형 요인에만 사용할 수 있습니다 . 요인 이름 아래에 요인 수준에 대한 검색 문자열을 입력할 수 있는 텍스트 상자를 제공합니다 . Enter 키를 누르거나 텍스트 상자 바깥쪽을 클릭하면 검색이 수행됩니다 . **찾기**가 선택되면 찾기 텍스트 상자 드롭다운 메뉴에 찾기 옵션이 나타납니다 .

허용되지 않는 조합 스크립트 사용

이 옵션을 사용하면 JSL 스크립트를 사용하여 요인 수준 조합을 허용하지 않도록 설정할 수 있습니다. 이 옵션은 연속형 요인 또는 연속형과 범주형이 혼합된 요인과 함께 사용할 수 있습니다.

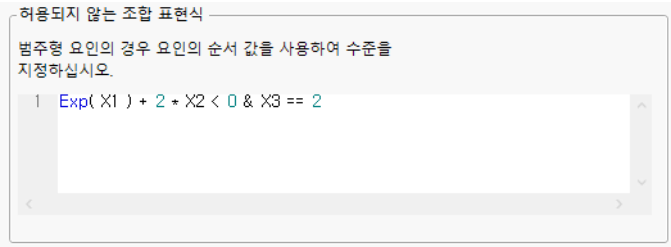
이 옵션은 허용하지 않을 조합을 식별하는 스크립트를 삽입할 수 있는 스크립트 창을 엽니다. 스크립트는 부울 표현식으로 평가되어야 합니다. 표현식이 true로 평가되면 지정된 조합이 허용되지 않습니다.

범주형 요인에 대한 표현식을 작성할 때는 수준의 순서 값 또는 수준 이름을 사용합니다 . " 요인 " 개요에서 요인 수준이 높음 , 중간 , 낮음으로 순서대로 지정된 경우 관련 서수 값은 1, 2, 3 입니다 . 예를 들어 두 개의 연속형 요인 (X1, X2) 과 세 가지 수준 (L1, L2, L3) 이 순서대로 설정된 범주형 요인 하나가 있다고 가정해 보겠습니다 . 다음 표현식이 true 인 경우 수준을 허용하지 않으려고 합니다 .

$$e^{X_1} + 2X_2 < 0 \text{ 및 } X_3 = L2$$

`(Exp(X1) + 2*X2 < 0) & (X3 == 2)` 표현식을 스크립트 창에 입력합니다.

그림 6.11 스크립트 편집기의 표현식



(그림에서 불필요한 괄호는 파싱을 통해 제거됨) 함수를 부울 표현식의 일부로 입력할 수 있습니다. `(Exp(X1) + 2*X2 < 0) & (X3 == "L2")` 표현식도 유효합니다.

확대 옵션

설계 확대 플랫폼에는 다음과 같은 확대 옵션이 있습니다.

반복 설계를 지정된 횟수만큼 반복합니다. 자세한 내용은 " 설계 반복 "(210 페이지) 에서 확인하십시오.

중앙점 추가 중앙점을 추가합니다. 설계에 중앙점으로 포함할 추가 런 수를 지정합니다. 중앙점은 각 연속형 요인에 대한 설정이 상한 설정과 하한 설정의 중간에 있는 런입니다. 자세한 내용은 "DOE 로 시작 " 장의 " 중앙점 , 반복 런 및 검정 "(61 페이지) 에서 확인하십시오.

설계에 연속형 요인과 다른 유형의 요인이 모두 포함된 경우 중앙점이 다른 요인의 수준과 관련하여 균형이 맞지 않을 수 있습니다. 설계 확대는 설계의 D 효율도 , I 효율도 또는 별칭 효율도를 최대화하기 위해 중앙점을 선택합니다.

자세한 내용은 " 설계에 중앙점 추가 "(212 페이지) 에서 확인하십시오.

설계 접기 설계 접기를 생성합니다. 자세한 내용은 " 설계 접기 생성 "(213 페이지) 에서 확인하십시오.

축 추가 축 점과 중앙점을 설계에 추가합니다. " 축 추가 " 를 사용하여 선별 설계에서 중심 합성 설계를 생성할 수 있습니다. 자세한 내용은 " 설계에 축 점 추가 "(214 페이지) 에서 확인하십시오.

공간 채움 연속형 요인으로 구성된 설계에 런을 추가합니다. 추가 런은 쾌속 유연 채움 방법론을 사용하여 구성됩니다. 자세한 내용은 " 공간 채움 확대 "(215 페이지) 에서 확인하십시오. 공간 채움은 새 런을 별도의 블록으로 그룹화하는 것을 지원하지 않습니다.

확대 모형을 사용하여 설계에 런을 추가합니다. 대개 이 모형에는 원래 모형보다 항이 더 많습니다. 자세한 내용은 " 설계 확대 플랫폼을 사용한 예 "(195 페이지) 에서 확인하십시오.

설계 확대 플랫폼 옵션

" 가속 수명 시험 계획 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

반응 저장 " 반응 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 . 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다 .

반응 불러오기 " 반응 저장 " 옵션을 사용하여 저장한 반응을 불러옵니다 .

요인 저장 " 요인 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 각 요인의 열에는 수준이 포함됩니다 . 다른 정보는 열 특성으로 저장됩니다 . 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다 . 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다 . 그런 다음 " 설계 역할 " 영역에서 적절한 역할을 선택합니다 .

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 불러옵니다 .

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) " 요인 제약 조건 정의 " 또는 " 선형 제약 조건 " 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다 . 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다 . 마지막 행은 부등식 한계를 포함합니다 . 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 제약 조건 상태 "(762 페이지) 에서 확인하십시오 .

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) " 제약 조건 저장 " 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다 .

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다 . 이러한 작업에는 다음 중 하나 이상이 포함됩니다 .

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다 .

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다 .

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다 .
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다 .

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다 .

- **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다. 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 **기본 분석**의 " 시뮬레이션 " 장에서 확인하십시오 .

행렬 X 저장 " 적률 행렬 " 과 " 모형 행렬 " 이라는 스크립트를 설계 데이터 테이블에 저장합니다 . 이 스크립트에는 적률 및 설계 행렬이 포함됩니다 . 자세한 내용은 " 사용자 설계 " 장의 " 행렬 X 저장 "(105 페이지) 에서 확인하십시오 .

주의 : 명목형 요인이 있는 설계의 경우 " 행렬 X 저장 " 옵션으로 저장한 " 모형 행렬 " 스크립트의 행렬은 선형 모형을 적합시킬 때 사용된 코딩 행렬이 **아닙니다** . " 모형 " 스크립트를 실행할 때 나타나는 " 모형 적합 " 보고서에서 " 열 저장 > 코드화 테이블 저장 " 옵션을 선택하면 모형 적합에 사용되는 코딩 행렬을 가져올 수 있습니다 .

최적 기준 설계 최적 기준을 변경합니다 . 기본 기준인 **권장 사항**은 " 모형 " 개요에서 "RSM" 버튼을 사용하여 2 차 효과를 추가한 경우를 제외하고 모든 설계 유형에 대해 D- 최적을 지정합니다 . D- 최적 , I- 최적 및 별칭 - 최적 설계에 대한 자세한 내용은 " 사용자 설계 " 장의 " 최적 기준 "(118 페이지) 에서 확인하십시오 .

참고 : 항상 정해진 최적 기준을 사용하도록 환경 설정을 지정할 수 있습니다 . **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다 . " 최적 기준 " 을 선택한 후 기본 기준을 선택합니다 .

시작 수 설계 구성에 사용되는 랜덤 시작 횟수를 지정할 수 있습니다. 자세한 내용은 " 사용자 설계 " 장의 " 시작 수 "(107 페이지) 에서 확인하십시오 .

설계 검색 시간 설계 검색을 위한 최대 시간 (초) 입니다 . 기본 검색 시간은 설계 복잡성을 기반으로 합니다 . 자세한 내용은 " 사용자 설계 " 장의 " 설계 검색 시간 "(108 페이지) 및 " 사용자 설계 " 장의 " 시작 수 "(107 페이지) 에서 확인하십시오 .

알고리즘 반복에 몇 초 이상이 필요한 경우 " 설계 계산 중 " 진행률 창이 나타납니다 . 진행률 창에서 **취소**를 클릭하면 계산이 중지되고 해당 시점에서 찾은 최적 설계가 제공됩니다 . 또한 " 변경 " 이 " 어려움 " 또는 " 매우 어려움 " 으로 설정되거나 " 추정 가능성 " 이 "If Possible" 로 설정된 요인이 포함되지 않은 D- 최적 설계에 대한 D 효율도가 진행률 창에 표시됩니다 .

참고 : " 설계 검색 시간 " 에 대한 환경 설정을 지정할 수 있습니다 . 파일 > 환경 설정 > 플랫폼 > DOE 를 선택합니다 . 설계 검색 시간을 선택한 후 최대 시간 (초) 을 입력합니다 . 시간이 더 필요한 특정 상황에서는 JMP 가 검색 시간을 연장합니다 .

구 반지름 설계의 연속형 요인을 초구로 제한합니다 . 반지름을 지정하고 **확인**을 클릭합니다 . 0 과의 간격이 구 반지름과 같도록 설계점이 선택됩니다 . " 설계 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

참고 : 구 반지름 제약 조건은 " 선형 제약 조건 지정 " 옵션을 사용하여 추가한 제약 조건과 결합할 수 없습니다 . 또한 변경하기 힘든 요인 (분할구 설계) 이 포함된 경우에는 이 옵션을 사용할 수 없습니다 .

고급 옵션

혼합물 합 혼합 요인의 합을 양수 값으로 설정합니다 . 전체 실험에서 혼합 성분을 일정하게 유지하려면 이 옵션을 사용합니다 .

분할구 분산비 오차 분산에 대한 랜덤 주구 분산 및 하위구 분산 (있는 경우) 의 비를 지정합니다 . 이 값을 설정하려면 먼저 분할구 설계를 위한 변경하기 힘든 요인 또는 분할 - 분할구 설계를 위한 변경하기 힘든 요인과 매우 변경하기 힘든 요인을 정의해야 합니다 . 그런 다음 분할구 설계를 지정했는지 아니면 분할 - 분할구 설계를 지정했는지에 따라 분산 비율에 하나 또는 두 개의 양수를 입력할 수 있습니다 .

사전 모수 분산 (" 모형 " 개요가 있는 경우에만 사용 가능) " 추정 가능성 " 이 "If Possible" 로 설정된 요인에 사용되는 가중치를 지정합니다 . " 설계 생성 " 을 클릭하면 옵션이 업데이트되어 기본 가중치가 표시됩니다 . 가중치를 지정할 각 항에 대해 양수를 입력합니다 . 입력한 값은 사전 분산에 대한 역수의 제곱근입니다 . 값이 클수록 분산이 작으므로 효과가 활성화되지 않는다는 더 많은 사전 정보가 있습니다 .

베이지안 D- 최적 또는 I- 최적은 If Possible 요인이 있는 설계를 구성할 때 사용됩니다 . 알고리즘에 사용되는 기본값은 필수 항의 경우 0, If Possible 항이 포함된 교호작용의 경우 4, If Possible 항의 경우 1 입니다 . 자세한 내용은 " 기술 상세 정보 " 부록의 " 별칭 행렬 "(767 페이지) 및 " 사용자 설계 " 장의 " 최적 기준 "(118 페이지) 에서 확인하십시오 .

A- 최적 모수 가중치 (A- 최적 설계에 사용) 모형 모수에 대한 가중치를 지정합니다 . 예를 들어 2 차 효과보다 주효과의 분산에 더 가중치를 부여할 수 있습니다 . 모수 가중치에 대한 자세한 내용은 Morgan and Stallings(2017) 에서 확인하십시오 .

D 효율도 가중치 설계를 구성할 때 별칭 최적에 대한 D 효율도의 상대 중요도를 지정합니다. 계수의 분산을 줄이는 것과 적합한 별칭 구조를 얻는 것 사이의 균형을 잡으려면 이 옵션을 선택합니다. 값은 0 에서 1 사이여야 합니다. 값이 클수록 D 효율도에 더 많은 가중치를 둡니다. 기본값은 0.5 입니다. 이 옵션은 " 최적 설계 별칭 생성 " 을 " 최적 기준 " 으로 선택한 경우에만 효과가 있습니다.

D 효율도의 정의는 " 사용자 설계 " 장의 " 최적 기준 "(118 페이지) 에서 확인하십시오. 별칭 최적에 대한 자세한 내용은 " 사용자 설계 " 장의 " 별칭 최적 "(121 페이지) 에서 확인하십시오.

스크립트를 스크립트 창에 저장 "사용자 설계" 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다.

확대 옵션의 추가 예

- " 설계 반복 "
- " 설계에 중앙점 추가 "
- " 설계 접기 생성 "
- " 설계에 축 점 추가 "
- " 공간 채움 확대 "

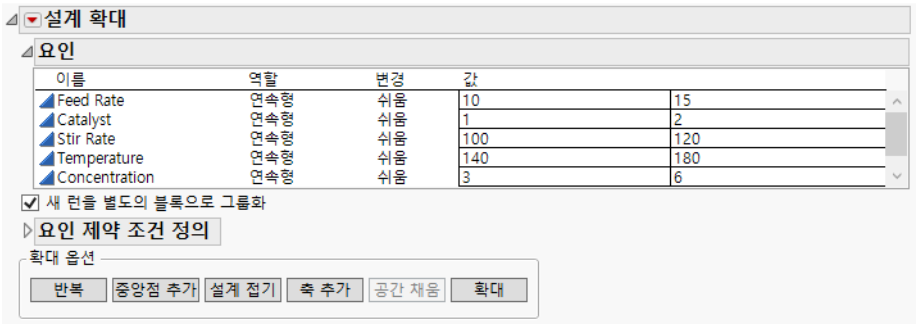
설계 반복

오차 분산이 일정하다는 가정을 직접 확인하려면 반복을 사용하여 설계를 확대합니다. 반복은 공정이나 측정의 변동이 클 때 회귀 계수의 변동도 줄여 줍니다.

이 예에서는 8 회 런 설계를 반복합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 8 Runs.jmp 를 엽니다.
2. **DOE > 설계 확대**를 선택합니다.
3. Percent Reacted 를 선택하고 **Y, 반응**을 클릭합니다.
4. Feed Rate, Catalyst, Stir Rate, Temperature 및 Concentration 을 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.

그림 6.12 확대 유형 선택



6. **새 런을 별도의 블록으로 그룹화**를 선택하여 설계에 블록 변수를 추가합니다 .
7. **반복**을 클릭하고 2 를 입력한 후 **확인**을 클릭합니다 .

참고 : 반복 대화상자가 표시되어 각 런을 수행할 횟수를 입력하도록 요청합니다 . 2 를 입력하면 각 런이 결과 설계에 두 번 나타나도록 지정됩니다 . 이렇게 되면 기존 설계가 한 번 반복됩니다 .

그림 6.13 반복을 사용하여 확대한 반응기 데이터 설계

요인 설계

런	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	블록
1	15	2	100	140	3	1
2	10	2	120	140	6	1
3	15	1	120	140	3	1
4	10	2	100	180	3	1
5	10	1	120	180	3	1
6	10	1	100	140	6	1
7	15	2	120	180	6	1
8	15	1	100	180	6	1
9	15	2	100	140	3	2
10	10	2	120	140	6	2
11	15	1	120	140	3	2
12	10	2	100	180	3	2
13	10	1	120	180	3	2
14	10	1	100	140	6	2
15	15	2	120	180	6	2
16	15	1	100	180	6	2

원래 설계의 8 회 런은 블록 1 에 있고 반복된 런은 블록 2 에 있습니다 .

참고 : 설계 확대 플랫폼에서 전체 설계를 평가할 수 있습니다 . 설계 평가에 대한 자세한 내용은 15 장 , " 설계 평가 " 에서 확인하십시오 .

8. **테이블 생성**을 클릭합니다 .

그림 6.14 반복된 설계

	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	블록	Percent Reacted
1	15	2	100	140	3	1	61
2	10	2	120	140	6	1	67
3	15	1	120	140	3	1	56
4	10	2	100	180	3	1	94
5	10	1	120	180	3	1	66
6	10	1	100	140	6	1	56
7	15	2	120	180	6	1	82
8	15	1	100	180	6	1	45
9	15	2	100	140	3	2	•
10	10	2	120	140	6	2	•
11	15	1	120	140	3	2	•
12	10	2	100	180	3	2	•
13	10	1	120	180	3	2	•
14	10	1	100	140	6	2	•
15	15	2	120	180	6	2	•
16	15	1	100	180	6	2	•

이 테이블에는 원래 설계 및 결과가 블록 1 에 있습니다 . 확대된 런은 블록 2 에 있습니다 . 반응 백분율 값은 확대된 런을 측정한 후 테이블에 추가될 수 있습니다 .

설계에 중앙점 추가

곡률을 확인하고 요인 영역의 중심에서 예측 오차를 줄이려면 중앙점 추가를 사용하여 설계를 확대합니다. 일반적으로 중앙점은 적합 결여 검정에 사용할 수 있는 순수 오차를 독립적으로 추정할 수 있는 반복된 점입니다.

이 예에서는 8 회 런 설계에 중앙점 두 개를 추가합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 8 Runs.jmp 를 엽니다 .
2. **DOE > 설계 확대**를 선택합니다 .
3. Percent Reacted 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Feed Rate, Catalyst, Stir Rate, Temperature 및 Concentration 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. **새 런을 별도의 블록으로 그룹화**를 선택하여 설계에 블록 변수를 추가합니다 .
7. **중앙점 추가**를 클릭하고 2 를 입력한 후 **확인**을 클릭합니다 .
8. **테이블 생성**을 클릭합니다 .

그림 6.15 중앙점이 두 개 추가된 설계

	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	블록	Percent Reacted
1	15	2	100	140	3	1	61
2	10	2	120	140	6	1	67
3	15	1	120	140	3	1	56
4	10	2	100	180	3	1	94
5	10	1	120	180	3	1	66
6	10	1	100	140	6	1	56
7	15	2	120	180	6	1	82
8	15	1	100	180	6	1	45
9	12.5	1.5	110	160	4.5	2	•
10	12.5	1.5	110	160	4.5	2	•

2 개의 중앙점 런에 대한 데이터가 수집되면 모형에 곡률이 있는지 검정할 수 있습니다.

설계 접기 생성

주효과 및 2 요인 교호작용의 교락을 제거하려면 설계 접기를 사용하여 설계를 확대합니다. 설계 접기는 특히 포화 또는 근사 포화 부분 요인 설계나 Plackett-Burman 설계에 대한 추가 작업으로 유용합니다.

이 예에서는 8 회 런 설계의 설계 접기를 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 8 Runs.jmp 를 엽니다.
2. **DOE > 설계 확대**를 선택합니다 .
3. Percent Reacted 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Feed Rate, Catalyst, Stir Rate, Temperature 및 Concentration 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. **새 런을 별도의 블록으로 그룹화**를 선택하여 설계에 블록 변수를 추가합니다 .
7. **설계 접기**를 클릭한 후 **확인**을 클릭합니다 .

   설계 접기 대화상자에서 요인을 선택하지 않으면 모든 설계 요인이 설계 접기에 포함됩니다 . 일부 요인만 설계 접기하도록 선택하면 나머지 요인은 원래 런에서 반복됩니다 .

그림 6.16 모든 요인에 대한 설계 접기

요인 설계

런	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	블록
1	15	2	100	140	3	1
2	10	2	120	140	6	1
3	15	1	120	140	3	1
4	10	2	100	180	3	1
5	10	1	120	180	3	1
6	10	1	100	140	6	1
7	15	2	120	180	6	1
8	15	1	100	180	6	1
9	10	1	120	180	6	2
10	15	1	100	180	3	2
11	10	2	100	180	6	2
12	15	1	120	140	6	2
13	15	2	100	140	6	2
14	15	2	120	180	3	2
15	10	1	100	140	3	2
16	10	2	120	140	3	2

8. **테이블 생성**을 클릭합니다 .

블록 1 의 런에 대한 요인 수준을 낮음에서 높음으로 또는 그 반대로 변경하여 블록 2 의 각 런을 얻을 수 있습니다 .

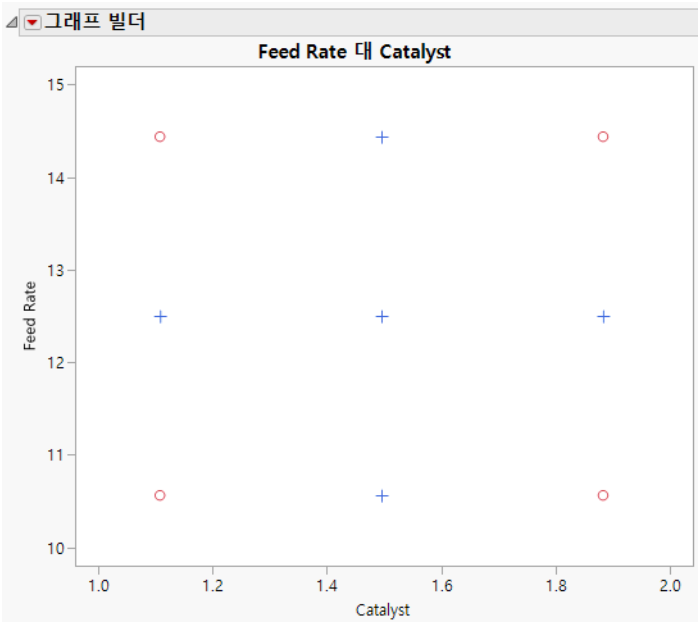
설계에 축 점 추가

반응 표면 설계를 얻기 위해 선별 설계를 확대하려면 축 점과 중앙점을 사용하여 설계를 확대합니다 . 이렇게 하면 반응 표면 모형을 데이터에 적합시켜 반응에 대한 요인 설정을 최적화할 수 있습니다 .

이 예에서는 축 점과 중앙점을 사용하여 8 회 런 선별 설계를 확대합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 8 Runs.jmp 를 엽니다 .
2. **DOE > 설계 확대**를 선택합니다 .
3. Percent Reacted 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Feed Rate, Catalyst, Stir Rate, Temperature 및 Concentration 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. **새 런을 별도의 블록으로 그룹화**를 선택하여 설계에 블록 변수를 추가합니다 .
7. **축 추가**를 클릭하고 축 값과 중앙점 수에 각각 1 과 2 를 입력한 후 **확인**을 클릭합니다 .
8. **테이블 생성**을 클릭합니다 .
9. " 설계 확대 " 테이블에서 **행 > 열 값에 따른 색상 또는 표식**을 선택합니다 .
10. 블록을 선택하고 " 표식 " 에서 **표준**을 선택한 후 **확인**을 클릭합니다 .
11. **그래프 > 그래프 빌더**를 선택합니다 .
12. Feed Rate 를 Y 로 드래그하고 Catalyst 를 X 로 드래그합니다 . 그런 다음 평활기 아이콘을 클릭하여 제거합니다 .

그림 6.17 두 가지 요인에 대한 설계 확대의 그래픽 보기



Feed Rate 대 Catalyst 그래프의 "+" 점은 추가된 축 점과 중앙점입니다. 축 점은 설계 공간의 표면에 있습니다. 설계를 확대할 때 축 값을 늘리거나 줄이면 축 점이 설계 공간을 벗어나거나 내부로 이동합니다.

공간 채움 확대

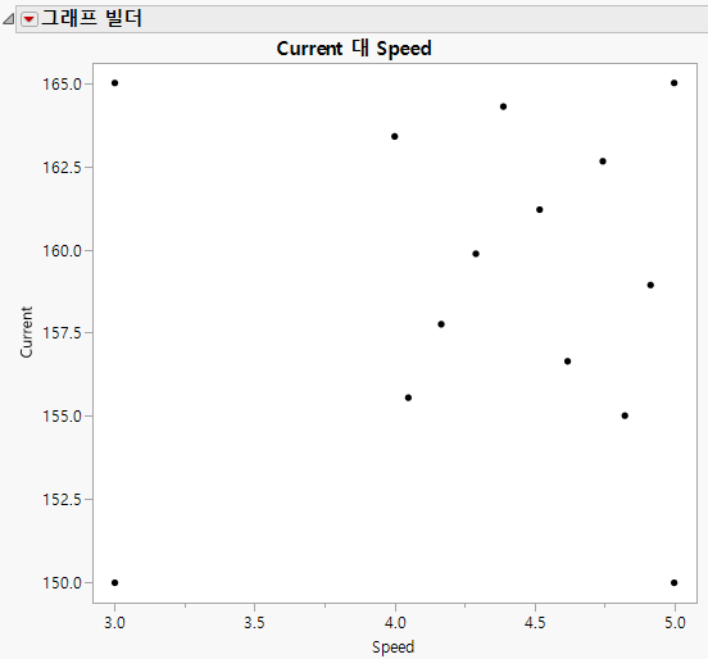
연속 요인으로 이루어진 설계에 점을 추가하려면 공간 채움을 사용하여 설계를 확대합니다. "공간 채움" 옵션은 설계 공간에 대한 제약 조건을 수용합니다. 선형 제약 조건 또는 허용되지 않는 조합을 지정할 수 있습니다. 사용된 알고리즘에 대한 자세한 내용은 "통계 상세 정보"(216 페이지)에서 확인하십시오.

이 예에서는 설계 공간에서 두 요인에 대한 제약 조건이 추가되어 반응 (Depth) 이 최대화된 제한 영역을 탐색하기 위해 설계를 확대합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/DOE Example 1.jmp를 엽니다.
2. **DOE > 설계 확대**를 선택합니다.
3. Depth 를 선택하고 **Y, 반응**을 클릭합니다.
4. Speed 와 Current 를 선택하고 **X, 요인**을 클릭한 후 **확인**을 클릭합니다.
5. " 요인 제약 조건 정의 " 개요를 엽니다.
6. **선형 제약 조건 지정**을 선택하고 **추가**를 클릭합니다.
7. Speed 에 1 을 입력하고 " 크거나 같음 " 부호를 선택한 후 Speed 하한에 4 를 입력합니다.

8. **추가**를 클릭합니다 .
9. Current 에 1 을 입력하고 " 크거나 같음 " 부호를 선택한 후 Current 하한에 155 를 입력합니다 .
10. **공간 채움**을 클릭합니다 .
11. 10 을 입력하고 **확인**을 클릭합니다 .
12. **테이블 생성**을 클릭합니다 .
13. " 설계 확대 " 테이블에서 " 그래프 빌더 " 를 선택합니다 .
14. Speed 를 X 놓기 영역으로 드래그하고 Current 를 Y 놓기 영역으로 드래그합니다 .

그림 6.18 공간 채움 확대 런



확대된 런이 설계 공간에서 오른쪽 위의 공간을 채웁니다. 이 공간이 제한된 설계 공간입니다.

통계 상세 정보

이 섹션에는 공간 채움 확대 옵션을 위한 점을 선택하는 데 사용되는 알고리즘에 대한 통계 상세 정보가 포함되어 있습니다.

설계를 확대하는 데 사용되는 알고리즘은 지정된 설계 영역에서 다수의 랜덤 점을 생성하는 것으로 시작합니다. 그런 다음 빠른 Ward 알고리즘을 사용하여 이러한 점을 여러 군집으로 군집화합니다. 이때 군집 수는 사용자가 지정하는 추가 런 수와 같습니다.

최종 설계점은 기존 런과 추가 런에 대해 MaxPro(최대 투영) 기준을 최적화하여 결정됩니다 . p 개 요인이 있고 n 이 기존 런과 추가 런의 수와 같은 경우 MaxPro 기준은 군집에서 다음 기준을 최소화하는 점을 찾습니다 .

$$C_{MaxPro} = \sum_{i}^{n-1} \sum_{j=i+1}^{n} \left[1 / \prod_{k=1}^{p} (x_{ik} - x_{jk})^2 \right]$$

MaxPro 기준은 모든 요인을 포함하는 방식으로 설계점 간 거리의 곱을 최대화합니다. 이는 요인 투영에 적합한 공간 채움 특성을 제공한다는 목표를 지원합니다. 자세한 내용은 Joseph et al. (2015)에서 확인하십시오.

7 장

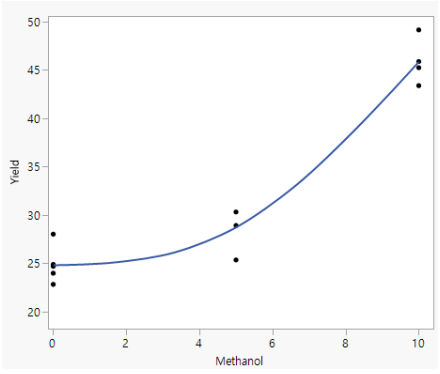
확증적 선별 설계

DSD(확증적 선별 설계)는 선별 설계입니다. 이 설계는 일반적으로 4개 이상의 요인이 있는 초기 단계 실험 작업에 적합합니다. DSD는 연속형 요인 또는 2수준 범주형 요인의 조합에 사용될 수 있습니다. 대부분의 요인이 연속형일 때 가장 잘 작동합니다. 각 연속형 요인은 연속형 요인에 대한 2차 모형 항을 조사할 수 있는 세 가지 수준을 사용합니다.

다음과 같은 면에서 확증적 선별 설계가 표준 선별 설계보다 우수하다고 할 수 있습니다.

- 각 연속형 요인을 세 가지 수준으로 처리하여 비선형 효과의 원인을 식별하는 데 도움이 됩니다. 표준 선별 설계의 연속형 요인에는 두 가지 수준만 있습니다. 선별 설계에 중앙점을 추가할 수 있지만 이 점은 곡률이 있는 경우에만 설정됩니다. 2 차 효과의 원인이 되는 요인은 식별할 수 없습니다.
- 2 차까지의 모든 효과 사이에 교락을 방지합니다. 연속형 요인을 사용하는 확증적 선별 설계의 주효과는 다른 주효과와 서로 직교하고 2 요인 교호작용 및 2 차 효과와 직교합니다. 2 요인 교호작용은 서로 완전히 교락되지 않습니다. 런 수가 비슷한 많은 표준 선별 설계에서는 교락이 발생합니다.
- 표준 선별 설계의 초기 결과에서 모호성을 해결하기 위해 비용이 많이 드는 추가 실험이 필요하지 않습니다.

그림 7.1 곡률을 보여 주는 요인 값 대 반응 그림



목차

확증적 선별 설계 개요 221
확증적 선별 설계의 예 223
확증적 선별 설계 223
부분 요인 설계와 비교 226
블록화를 사용한 확증적 선별 설계 228
확증적 선별 설계와 Plackett-Burman 설계 비교 232
확증적 선별 설계 창 235
반응 236
요인 237
설계 옵션 239
설계 241
설계 평가 241
출력 옵션 242
확증적 선별 설계 옵션 244
반응 시뮬레이션 246
기술 상세 정보 248
확증적 선별 설계의 구조 248
실험 데이터 분석 250

확증적 선별 설계 개요

조사자는 반응에 가장 큰 영향을 미치는 요인을 식별하려고 할 때 선별 설계를 사용합니다. 선별 설계를 사용하면 상당히 작은 실험에서 많은 요인을 연구할 수 있습니다.

많은 표준 선별 설계는 주효과를 추정하는 데 중점을 둡니다. 확증적 선별 설계는 표준 선별 설계에 비해 몇 가지 장점을 제공합니다. 이 설계는 효과 교락을 방지하고, 반응에 대해 비선형 효과를 가지는 요인을 식별할 수 있습니다. 확증적 선별 설계의 장점과 구성에 대한 자세한 내용은 Jones and Nachtsheim(2011a)에서 확인하십시오.

연속형 요인만 포함하는 설계의 경우 확증적 선별 설계와 표준 선별 설계에서 다음 특성을 비교해 보십시오.

참고 : 2 차 효과가 언급되면 표준 선별 설계에 중앙점이 있는 것으로 가정합니다 .

- 주효과가 2 요인 교호작용과 직교합니다 .
  - 확증적 선별 설계 : 항상
  - 표준 선별 설계 : 해상도 IV 이상에만 해당
- 2 요인 교호작용이 다른 2 요인 교호작용과 완전히 교락되지 않습니다 .
  - 확증적 선별 설계 : 항상
  - 표준 선별 설계 : 해상도 V 이상에만 해당
- 주효과와 2 차 효과만 포함하는 모형에서 모든 2 차 효과를 추정할 수 있습니다 .
  - 확증적 선별 설계 : 항상
  - 표준 선별 설계 : 아님

이러한 특성은 이 섹션의 나머지 부분에서 더 자세히 설명됩니다.

표준 선별 설계

부분 요인 설계 또는 Plackett-Burman 설계와 같은 표준 선별 설계는 비교적 적은 리소스 할당으로 많은 요인을 연구하려고 합니다. 그러나 표준 선별 설계에는 다음과 같은 원치 않는 기능이 있습니다 .

- 일부 주효과와 2 요인 교호작용의 별칭 관계를 지정할 수 있습니다 . 예를 들어 Plackett-Burman 설계에서 주효과는 몇 가지 2 요인 교호작용과 상관관계가 있습니다 . 하나 이상의 2 요인 교호작용 효과가 큰 경우 실험자는 추가 런을 수행하여 모호성을 해결해야 합니다 .
- 일부 2 요인 교호작용이 서로 교락될 수도 있습니다 . 따라서 2 요인 교호작용 효과가 큰 경우 실험자는 추가 런을 수행하여 모호성을 해결해야 합니다 .
- 일반적으로 연속형 요인은 두 가지 수준 (낮음 , 높음) 으로 설정됩니다 . 그러나 엔지니어와 과학자들은 연속형 요인이 세 가지 수준 (낮음 , 중간 , 높음) 으로 설정된 설계를 선호하는 경우가 많습니다 . 두 가지 수준으로는 실제 시스템에서 흔히 볼 수 있는 비선형성을 감지하

기가 어렵기 때문입니다. 중앙점이 추가된 기존 선별 설계를 사용하여 비선형성을 감지할 수는 있지만 이 설계는 원인이 되는 요인을 식별하지 못합니다.

확증적 선별 설계

확증적 선별 설계를 사용하여 다음을 수행할 수 있습니다.

- 중요한 요인을 더 빠르고 효율적으로 식별할 수 있도록 모형 모호성을 방지합니다.
- 최대 2 차 항까지의 교락을 방지하면서 비선형 효과의 원인을 식별합니다. 따라서 기존 선별 설계에서 중앙점을 사용할 때와 같이 비선형성을 **감지**할 수 있을 뿐만 아니라 원인이 되는 요인을 **식별**할 수 있습니다.

확증적 선별 설계는 다음과 같은 장점을 제공합니다.

- 확증적 선별 설계에는 적은 수의 런만 있으면 됩니다. 요인이 6 개 이상인 경우 필요한 최소 런 수는 보통 요인 수의 두 배 정도면 됩니다. 런 수에 대한 자세한 내용은 " 컨퍼런스 행렬 및 런 수 "(249 페이지) 에서 확인하십시오.
- 주효과가 2 요인 교호작용과 직교합니다. 이는 활성 2 요인 교호작용이 모형에 포함되는지 여부에 관계없이 주효과의 추정값이 이러한 교호작용에 의해 편향되지 않음을 의미합니다. 해상도 III 선별 설계에서는 일부 주효과와 교호작용 효과가 교락됩니다. 또한 Plackett-Burman 설계에서는 활성 2 요인 교호작용이 있을 때 주효과 추정값이 편향됩니다.
- 2 요인 교호작용이 다른 2 요인 교호작용과 완전히 교락되지 않습니다. 그러나 2 요인 교호작용이 다른 2 요인 교호작용과 상관될 수 있습니다. 해상도 IV 선별 설계에서는 일부 2 요인 교호작용 효과가 완전히 교락됩니다.
- 주효과와 2 차 항으로만 구성된 모형에서 모든 2 차 효과를 추정할 수 있습니다. 이를 통해 비선형성을 설명하는 요인을 식별할 수 있습니다. 중앙점이 추가된 기존 선별 설계에서는 주효과와 2 차 효과로 구성된 모형에서 모든 2 차 효과를 추정할 수 **없습니다**.
- 2 차 효과는 주효과와 직교하며 다른 2 요인 교호작용과 완전히 교락되지 않습니다. 2 차 효과는 교호작용 효과와 상관될 수 있습니다.
- 요인이 6 개에서 30 개 이상인 경우 , 3 개 이하의 요인을 포함하는 완전 2 차 모형의 모수를 높은 정밀도로 추정할 수 있습니다.
- 요인이 18 개 이상인 경우 4 개의 요인으로 완전 2 차 모형을 적합시킬 수 있습니다. 요인이 24 개 이상인 경우에는 5 개의 요인으로 완전 2 차 모형을 적합시킬 수 있습니다.

확증적 선별 설계 플랫폼

확증적 선별 설계 플랫폼을 사용하면 연속형 요인 및 2수준 범주형 요인에 대한 확증적 선별 설계를 구성할 수 있습니다. 또한 블록화된 설계를 구성할 수 있습니다. 많은 효과가 활성화되어 있을 때 설계의 효과 감지 능력에 대한 신뢰성을 향상시키는 비중앙점 런을 더 추가할 수 있습니다.

효과 간의 절대 상관 값을 보려면 "확증적 선별 설계" 창의 "설계 평가" 개요에 포함된 "상관 색상 맵"을 사용합니다. 확증적 선별 설계와 다른 설계의 상관 색상 맵을 비교하면 해당 설계의 별칭 구조를 비교할 수 있습니다. 자세한 내용은 "설계 평가" 장의 " 상관 색상 맵"(431 페이지) 에서 확인하십시오.

확증적 선별 설계의 구조에 대한 자세한 내용은 "확증적 선별 설계의 구조"(248 페이지)에서 확인하십시오. 블록을 사용한 확증적 선별 설계에 대한 자세한 내용은 "확증적 선별 설계의 블록화"(240 페이지)에서 확인하십시오. 확증적 선별 설계를 사용하여 얻은 데이터를 분석하는 방법에 대한 제안은 "실험 데이터 분석"(250 페이지)에서 확인하십시오.

확증적 선별 적합 플랫폼

DSD(확증적 선별 설계)를 실행한 후 확증적 선별 적합 플랫폼을 사용하여 결과를 분석합니다. DSD에 적용된 표준 모형 선택 방법을 사용하면 활성 효과를 식별하지 못할 수 있습니다. 확증적 선별 적합 플랫폼은 **DSD를 위한 유효 모형 선택**이라는 알고리즘을 사용하여 활성 주효과와 2차 효과(second-order effect)를 식별합니다. 이 알고리즘은 DSD의 특수 구조를 활용합니다. 자세한 내용은 8 장, "확증적 선별 적합 플랫폼"에서 확인하십시오.

JMP에서 DSD를 생성하면 "확증적 선별 적합"이라는 스크립트가 설계 테이블에 포함됩니다. 이 스크립트는 DSD를 위한 유효 모형 선택 방법론을 사용하여 자동으로 분석을 실행합니다.

확증적 선별 설계의 예

- " 확증적 선별 설계 "
- " 부분 요인 설계와 비교 "
- " 블록화를 사용한 확증적 선별 설계 "
- " 확증적 선별 설계와 Plackett-Burman 설계 비교 "

확증적 선별 설계

6 개 요인 중 추출 공정의 수율에 영향을 미치는 요인을 판별해야 한다고 가정해 보겠습니다.

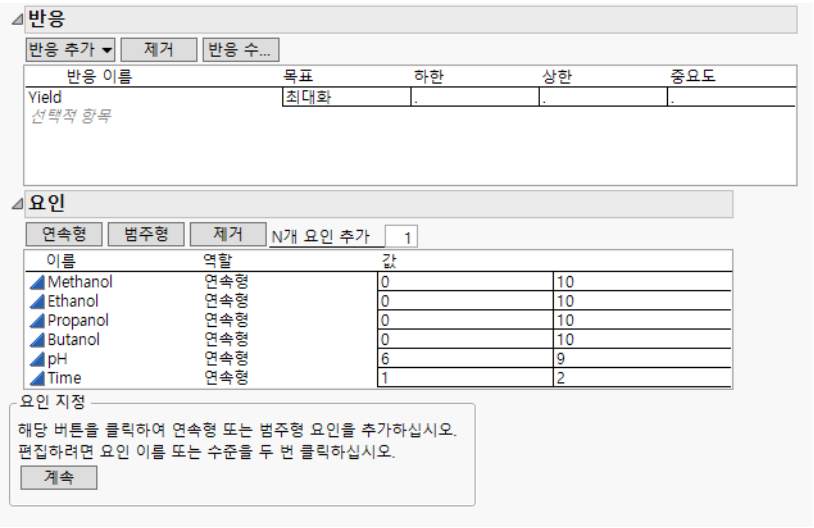
설계 생성

요인과 해당 설정은 Extraction Factors.jmp 데이터 테이블에 제공됩니다 . 조사를 위해 확증적 선별 설계를 생성합니다 .

1. **DOE > 확증적 선별 > 확증적 선별 설계**를 선택합니다 .
2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Yield 를 입력합니다 .
3. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Extraction Factors.jmp 를 엽니다 .
4. " 확증적 선별 설계 " 옆의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

   요인 이름과 범위가 " 요인 " 개요에 추가됩니다 .

그림 7.2 Extraction 설계의 반응 및 요인



5. **계속**을 클릭합니다.

   " 설계 옵션 " 개요가 열립니다. 여기서는 블록화 구조를 지정할 수 있습니다. 이 예에서는 블록화할 필요가 없으므로 기본 선택 **블록이 필요하지 않음**을 적용합니다.

   추가 런을 포함하도록 선택하여 2 차 효과 (second-order effect) 감지 능력을 크게 향상시킬 수도 있습니다. 최소 4 회의 추가 런이 권장되며 이 값이 기본으로 설정됩니다.

6. **설계 생성**을 클릭합니다.

   " 확증적 선별 설계 " 창이 업데이트되어 " 설계 " 개요와 " 설계 평가 " 개요가 표시됩니다.

그림 7.3 설계 개요

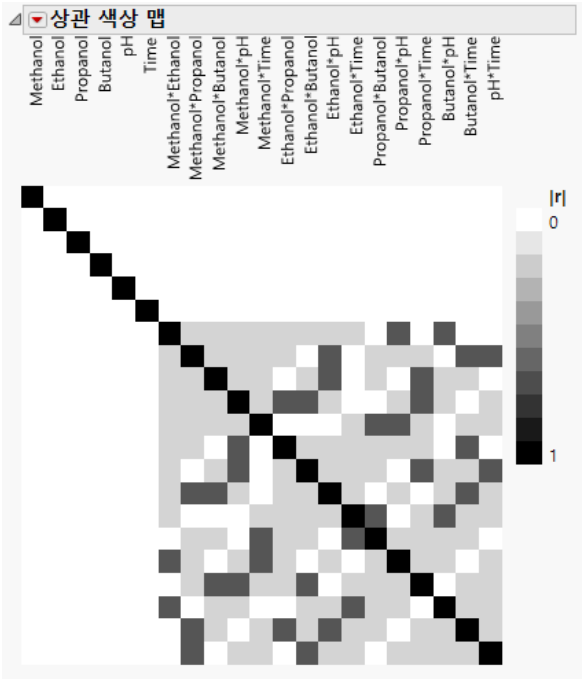
설계

런	Methanol	Ethanol	Propanol	Butanol	pH	Time
1	5	10	10	10	9	2
2	5	0	0	0	6	1
3	10	5	10	10	6	2
4	0	5	0	0	9	1
5	10	0	5	10	9	1
6	0	10	5	0	6	2
7	10	0	0	5	9	2
8	0	10	10	5	6	1
9	10	10	0	0	7.5	2
10	0	0	10	10	7.5	1
11	10	0	10	0	6	1.5
12	0	10	0	10	9	1.5
13	10	10	0	10	6	1
14	0	0	10	0	9	2
15	10	10	10	0	9	1
16	0	0	0	10	6	2
17	5	5	5	5	7.5	1.5

7. **설계 평가 > 상관 색상 맵** 개요를 엽니다.

" 상관 색상 맵 " 은 모든 주효과 및 2 요인 교호작용의 절대 상관 값에 색상 강도 척도를 할당합니다 .

그림 7.4 Extraction 설계의 상관 색상 맵



다음 사항에 유의하십시오 .

- 단색의 흰색 영역은 주효과 간 또는 주효과와 2 요인 교호작용 간에 상관관계가 없음을 나타냅니다 .
- 음영이 있는 회색 영역은 2 요인 교호작용 간의 절대 상관이 작음을 나타냅니다 .
- 검은색 사각형은 절대 상관 1 을 나타냅니다 . 이 사각형은 모두 주대각 위에 나타나며 효과 자체의 기대 상관을 반영합니다 .

" 출력 옵션 " 패널에서 " 런 순서 " 가 **랜덤화**로 설정되어 있습니다 .

8. **테이블 생성**을 클릭하여 그림 7.5 에 표시된 데이터 테이블을 생성합니다 .

참고 : 생성된 설계의 런은 그림 7.5 에 표시된 순서와 다르게 나타날 수 있습니다 .

그림 7.5 추출 공정의 확증적 선별 설계

	Methanol	Ethanol	Propanol	Butanol	pH	Time	Yield
1	5	0	0	0	6	1	•
2	0	0	10	0	9	2	•
3	5	10	10	10	9	2	•
4	0	5	0	0	9	1	•
5	10	0	10	0	6	1.5	•
6	10	10	10	0	9	1	•
7	10	5	10	10	6	2	•
8	0	10	0	10	9	1.5	•
9	0	10	5	0	6	2	•
10	5	5	5	5	7.5	1.5	•
11	10	10	0	10	6	1	•
12	10	10	0	0	7.5	2	•
13	10	0	5	10	9	1	•
14	0	10	10	5	6	1	•
15	10	0	0	5	9	2	•
16	0	0	0	10	6	2	•
17	0	0	10	10	7.5	1	•

부분 요인 설계와 비교

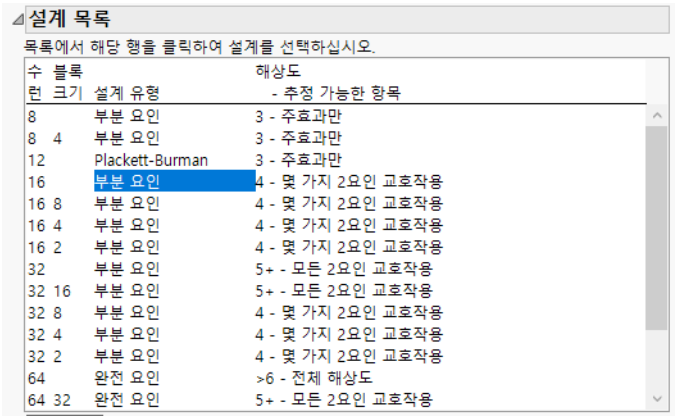
" 확증적 선별 설계 "(223 페이지) 에서 확증적 선별 설계 대신 기존 선별 설계를 선택했다고 가정해 보겠습니다 . 이 예에서는 교락에 대해 두 설계를 비교합니다 .

1. **DOE > 전통적 설계 > 2 수준 선별 > 선별 설계**를 선택합니다 .
2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Yield 를 입력합니다 .
3. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Extraction Factors.jmp 를 엽니다 .
4. " 선별 설계 " 옆의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

   요인 이름과 범위가 " 요인 " 개요에 추가됩니다 .
5. **계속**을 클릭합니다 .
6. **부분 요인 설계 목록에서 선택**을 선택합니다 .
7. **계속**을 클릭합니다 .

   가능한 설계가 " 설계 목록 " 에 나타납니다 .

그림 7.6 6 개 연속형 요인에 대한 선별 설계 목록



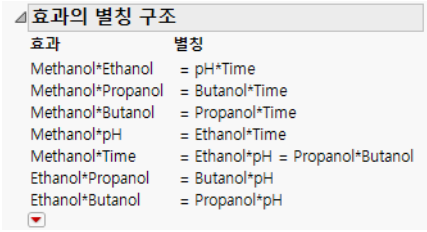
설계 목록

목록에서 해당 행을 클릭하여 설계를 선택하십시오.

수 런	블록 크기	설계 유형	해상도 - 추정 가능한 항목
8		부분 요인	3 - 주효과만
8	4	부분 요인	3 - 주효과만
12		Plackett-Burman	3 - 주효과만
16		부분 요인	4 - 몇 가지 2요인 교호작용
16	8	부분 요인	4 - 몇 가지 2요인 교호작용
16	4	부분 요인	4 - 몇 가지 2요인 교호작용
16	2	부분 요인	4 - 몇 가지 2요인 교호작용
32		부분 요인	5+ - 모든 2요인 교호작용
32	16	부분 요인	5+ - 모든 2요인 교호작용
32	8	부분 요인	4 - 몇 가지 2요인 교호작용
32	4	부분 요인	4 - 몇 가지 2요인 교호작용
32	2	부분 요인	4 - 몇 가지 2요인 교호작용
64		완전 요인	>6 - 전체 해상도
64	32	완전 요인	5+ - 모든 2요인 교호작용

8. 그림 7.6 에 강조 표시된 16 회 런 부분 요인 설계 (블록 없음) 를 선택합니다 .
9. **계속**을 클릭합니다 .
10. **설계 표시 및 수정 > 효과의 별칭 구조** 개요를 엽니다 .

그림 7.7 부분 요인 설계에 대한 효과의 별칭 구조



효과의 별칭 구조

효과	별칭
Methanol*Ethanol	= pH*Time
Methanol*Propanol	= Butanol*Time
Methanol*Butanol	= Propanol*Time
Methanol*pH	= Ethanol*Time
Methanol*Time	= Ethanol*pH = Propanol*Butanol
Ethanol*Propanol	= Butanol*pH
Ethanol*Butanol	= Propanol*pH

16 회 런 부분 요인 설계에 대한 "효과의 별칭 구조" 개요에서는 모든 2 요인 교호작용이 하나 이상의 다른 2 요인 교호작용과 교락된다는 것을 보여 줍니다. 이 부분 요인 설계에서 Ethanol*Time 교호작용은 Methanol*pH 교호작용과 교락됩니다. 활성 교호작용을 판별하려면 추가 시도를 실행해야 합니다. 요인이 다른 순서로 입력된 경우 Ethanol*Time 교호작용이 다른 2 요인 교호작용 두 개와 별칭 구조를 가질 수 있습니다.

"확증적 선별 설계"(223 페이지) 섹션에서 17 회 런 확증적 선별 설계를 구성했습니다. 이 DSD 의 "상관 색상 맵"(그림 7.4) 은 다른 2 요인 교호작용과 교락되는 2 요인 교호작용이 없음을 보여 줍니다. 부분 요인 설계의 경우 교락된 2 요인 교호작용이 7 개 있습니다. 활성 2 요인 효과가 있다고 의심되면 DSD 를 선택하는 것이 더 좋습니다.

설계 비교 플랫폼 (DOE > 설계 진단 > 설계 비교) 을 사용하여 두 설계를 더 철저하게 비교할 수 있습니다. 자세한 내용은 16 장, "설계 비교"에서 확인하십시오.

블록화를 사용한 확증적 선별 설계

원료 제약 조건으로 인해 두 개의 개별 로트에서 재료를 사용하여 추출 실험을 실행해야 한다고 가정해 보겠습니다. 잠재적 로트 변동을 설명하기 위해 블록 변수를 사용하여 확증적 선별 설계를 생성할 수 있습니다.

설계 생성

추출 요인과 해당 설정은 Extraction Factors.jmp 데이터 테이블에 제공됩니다. 다음과 같이 블록을 사용하여 확증적 선별 설계를 생성합니다.

1. **DOE > 확증적 선별 > 확증적 선별 설계**를 선택합니다.
2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Yield 를 입력합니다.
3. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Extraction Factors.jmp 를 엽니다.
4. " 확증적 선별 설계 " 옆의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다.

   요인 이름과 범위가 " 요인 " 개요에 추가됩니다.
5. **계속**을 클릭합니다.

   " 설계 옵션 " 개요가 열립니다. 여기서는 블록화 구조를 지정할 수 있습니다.
6. **2 차 효과 추정을 위해 중앙점 런과 블록 추가**를 선택합니다.

   2 로 설정된 **블록 수**를 그대로 둡니다.

   추가 런 없이 생성된 Extraction2 Data.jmp 샘플 데이터 테이블에 대한 설계를 다시 생성하는 중입니다. 추가 런을 4 회 실행하는 것이 좋지만 이 예에서는 추가 런을 지정하지 않습니다.
7. **추가 런 수** 옆에서 0 을 선택합니다.
8. **설계 생성**을 클릭합니다.

   " 확증적 선별 설계 " 창이 업데이트되어 " 설계 " 개요와 " 설계 평가 " 개요가 표시됩니다.

   " 요인 " 및 " 설계 " 개요에 **블록**이 추가되었는지 확인합니다.
9. " 요인 " 개요에서 **블록**을 두 번 클릭하고 Lot 를 입력합니다.

   " 출력 옵션 " 패널에서 " 런 순서 " 가 " 블록 내 랜덤화 " 로 설정되어 있습니다.
10. **테이블 생성**을 클릭합니다.

참고 : 생성된 설계의 런은 그림 7.8 에 표시된 순서와 다르게 나타날 수 있습니다.

그림 7.8 추출 공정에 대해 블록을 사용한 확증적 선별 설계

	블록	Methanol	Ethanol	Propanol	Butanol	pH	Time	Yield
1	1	5	5	5	5	7.5	1.5	•
2	1	5	10	10	10	9	2	•
3	1	5	0	0	0	6	1	•
4	1	10	0	5	0	9	2	•
5	1	0	0	0	10	7.5	2	•
6	1	0	10	5	10	6	1	•
7	1	10	10	10	0	7.5	1	•
8	2	10	5	0	10	9	1	•
9	2	0	10	0	0	9	1.5	•
10	2	0	0	10	5	9	1	•
11	2	10	0	10	10	6	1.5	•
12	2	0	5	10	0	6	2	•
13	2	10	10	0	5	6	2	•
14	2	5	5	5	5	7.5	1.5	•

런 1 은 Lot 1 의 중앙점 런이고 런 14 는 Lot 2 의 중앙점 런입니다 .

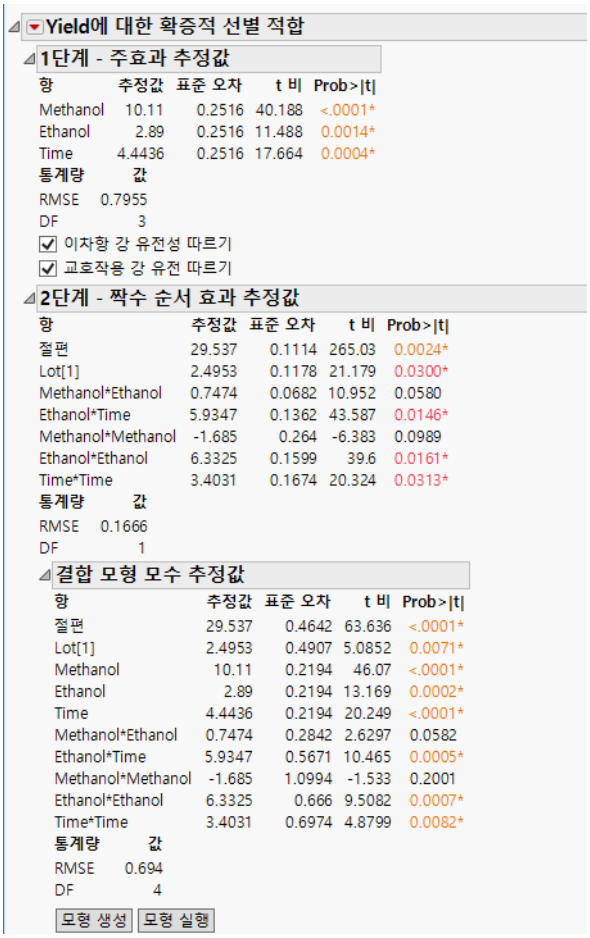
실험 데이터 분석

이제 실험을 수행하고 설계 테이블의 Yield 열에 데이터를 기록합니다(그림 7.8). Extraction2 Data.jmp 샘플 데이터 테이블에 실험 결과가 포함되어 있습니다. Extraction2 Data.jmp 샘플 데이터 테이블의 런은 그림 7.8 의 런과 순서가 다릅니다.

모든 2 차 효과 (second-order effect) 를 탐색하려면 모든 가능 모형 회귀를 사용하는 것이 한 가지 옵션입니다 . 다른 옵션은 단계별 전진 회귀를 사용하는 것입니다 . 그러나 이와 같은 표준 방법은 활성 효과를 식별하지 못하는 경우가 많습니다 . 이러한 이유로 확증적 선별 적합 플랫폼을 사용합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Extraction2 Data.jmp 를 엽니다 .
2. 설계 테이블의 테이블 패널에서 **Fit Definitive Screening** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

그림 7.9 확증적 선별 적합 보고서



Yield에 대한 확증적 선별 적합

1단계 - 주효과 추정값

항	추정값	표준 오차	t 비	Prob>|t|
Methanol	10.11	0.2516	40.188	<.0001*
Ethanol	2.89	0.2516	11.488	0.0014*
Time	4.4436	0.2516	17.664	0.0004*

통계량	값
RMSE	0.7955
DF	3

☑ 이차항 강 유전성 따르기
☑ 교호작용 강 유전 따르기

2단계 - 짝수 순서 효과 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	29.537	0.1114	265.03	0.0024*
Lot[1]	2.4953	0.1178	21.179	0.0300*
Methanol*Ethanol	0.7474	0.0682	10.952	0.0580
Ethanol*Time	5.9347	0.1362	43.587	0.0146*
Methanol*Methanol	-1.685	0.264	-6.383	0.0989
Ethanol*Ethanol	6.3325	0.1599	39.6	0.0161*
Time*Time	3.4031	0.1674	20.324	0.0313*

통계량	값
RMSE	0.1666
DF	1

결합 모형 모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	29.537	0.4642	63.636	<.0001*
Lot[1]	2.4953	0.4907	5.0852	0.0071*
Methanol	10.11	0.2194	46.07	<.0001*
Ethanol	2.89	0.2194	13.169	0.0002*
Time	4.4436	0.2194	20.249	<.0001*
Methanol*Ethanol	0.7474	0.2842	2.6297	0.0582
Ethanol*Time	5.9347	0.5671	10.465	0.0005*
Methanol*Methanol	-1.685	1.0994	-1.533	0.2001
Ethanol*Ethanol	6.3325	0.666	9.5082	0.0007*
Time*Time	3.4031	0.6974	4.8799	0.0082*

통계량	값
RMSE	0.694
DF	4

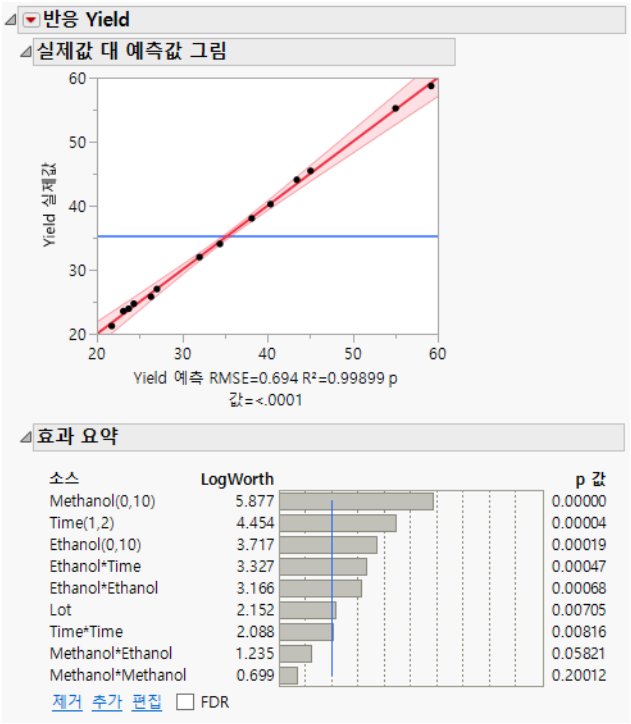
모형 생성 모형 실행

확증적 선별 적합에 의해 잠재적 활성 상태인 것으로 식별된 효과가 " 결합 모형 모수 추정값 " 보고서에 나열됩니다 .

3. " 결합 모형 모수 추정값 " 보고서 맨 아래의 **모형 실행** 버튼을 클릭합니다 .

   그러면 잠재적 활성 상태인 것으로 식별된 효과에 대해 표준 최소 제곱 모형을 적합시킵니다 .

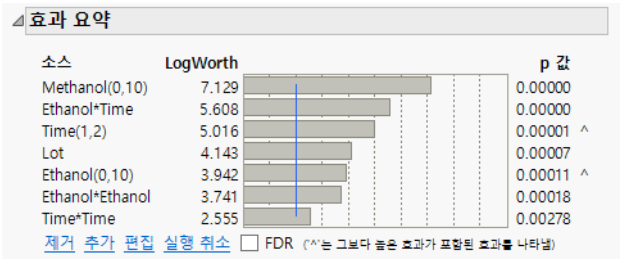
그림 7.10 표준 최소 제곱 보고서



"실제값 대 예측값 그림"에 적합 결여의 증거가 없습니다. "효과 요약" 보고서는 Methanol*Ethonal 과 Methanol*Methanol 이 유의하지 않음을 보여 줍니다. 따라서 모형에서 이들 효과를 제거하기로 결정했습니다.

4. "효과 요약" 보고서에서 Methanol*Ethonal 과 Methanol*Methanol 을 선택하고 **제거**를 클릭합니다.

그림 7.11 최종 활성 효과 집합



나머지 효과는 모두 유의합니다. 이러한 효과가 활성 효과인 것으로 결론을 내릴 수 있습니다.

확증적 선별 설계와 Plackett-Burman 설계 비교

Plackett-Burman 설계는 선별을 위한 부분 요인 설계 대신 사용할 수 있습니다. 그러나 Plackett-Burman 설계에는 2요인 교호작용에 의한 주효과의 복잡한 별칭 구조가 있습니다.

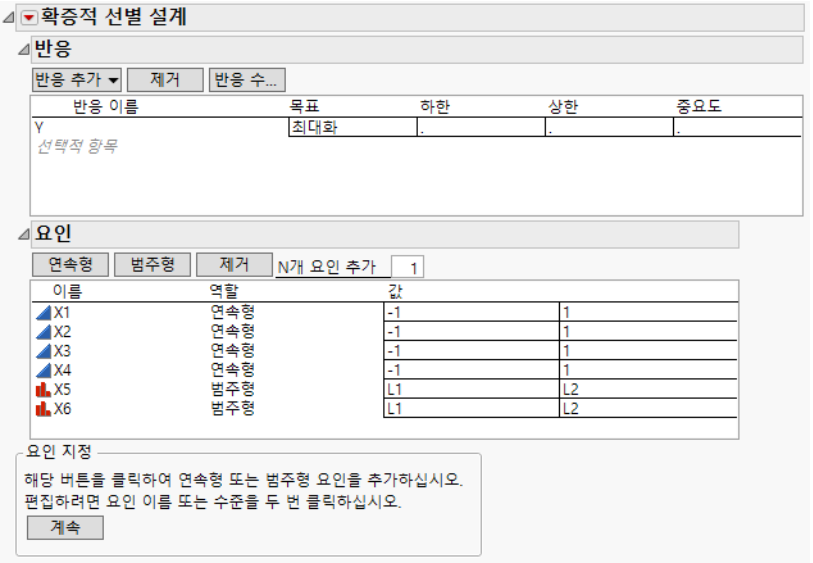
이 예에서는 설계 평가 플랫폼을 사용하여 확증적 선별 설계와 Plackett-Burman 설계를 비교하는 방법을 보여 줍니다. 설계 비교 플랫폼을 사용한 자세한 예는 "설계 비교" 장의 "런 크기가 같은 설계"(439 페이지)에서 확인하십시오.

확증적 선별 설계

1. **DOE > 확증적 선별 > 확증적 선별 설계**를 선택합니다.
2. **N 개 요인 추가** 상자에 **4** 를 입력하고 **연속형**을 클릭합니다.
3. **N 개 요인 추가** 상자에 **2** 를 입력하고 **범주형**을 클릭합니다.

   창이 그림 7.12 에 표시된 것처럼 보입니다.

그림 7.12 연속형 요인 4 개와 범주형 요인 2 개가 포함된 확증적 선별 대화상자



4. **계속**을 클릭합니다.

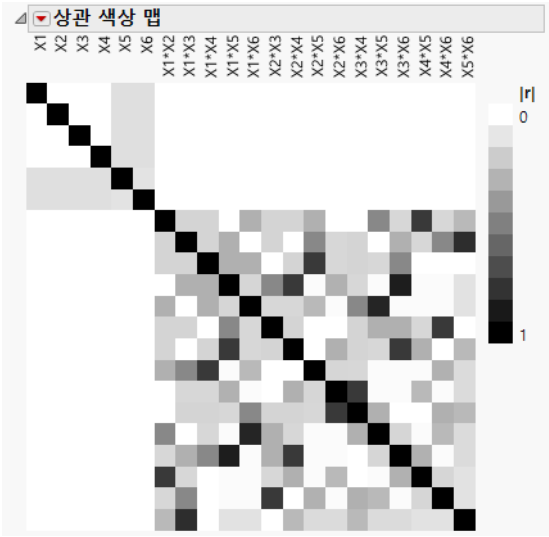
   이 예에서는 블록이 필요하지 않습니다. "설계 옵션" 개요에서 "블록이 필요하지 않음" 옵션이 선택되어 있는지 확인하십시오.

   거의 같은 크기의 설계를 비교하기 위해 추가 런을 지정하지 않습니다.
5. **추가 런 수** 옆에서 0 을 선택합니다.
6. **설계 생성**을 클릭합니다.

   14 회 런 설계가 생성됩니다.

7. **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

그림 7.13 확증적 선별 설계의 색상 맵



범주형 주효과는 다른 범주형 주효과 및 연속형 요인의 주효과와 약간 상관되어 있습니다 . 이로 인해 추정값의 정밀도가 조금 감소합니다 .

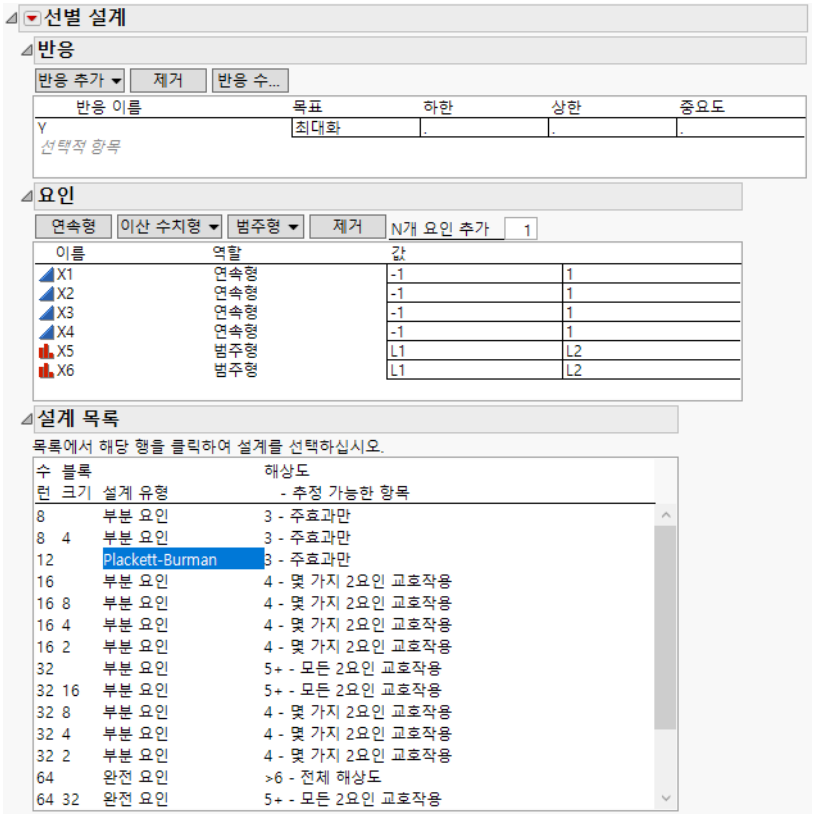
8. 아래의 Plackett-Burman 설계와 색상 맵을 비교할 때까지 " 확증적 선별 설계 " 창을 닫지 않습니다 .

Plackett-Burman 설계

이제 같은 요인 구조를 사용하여 Plackett-Burman 설계를 생성합니다 .

1. **DOE > 전통적 설계 > 2 수준 선별 > 선별 설계**를 선택합니다 .
2. **N 개 요인 추가** 상자에 **4** 를 입력하고 **연속형**을 클릭합니다 .
3. **N 개 요인 추가** 상자에 **2** 를 입력하고 **범주형 > 2 수준**을 클릭합니다 .
4. **계속**을 클릭합니다 .
5. **부분 요인 설계 목록에서 선택**을 선택하고 **계속**을 클릭합니다 .
   가능한 설계가 " 설계 목록 " 에 나타납니다 .

그림 7.14 Plackett-Burman 설계



6. 12 회 런 Plackett-Burman 설계를 선택합니다 (그림 7.14).
7. **계속**을 클릭합니다 .
8. **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

   12 회 런 Plackett-Burman 설계의 색상 맵을 14 회 런 확증적 선별 설계의 색상 맵과 비교합니다 .

그림 7.15 Plackett-Burman 상관 (왼쪽) 과 확증적 선별 상관 (오른쪽)

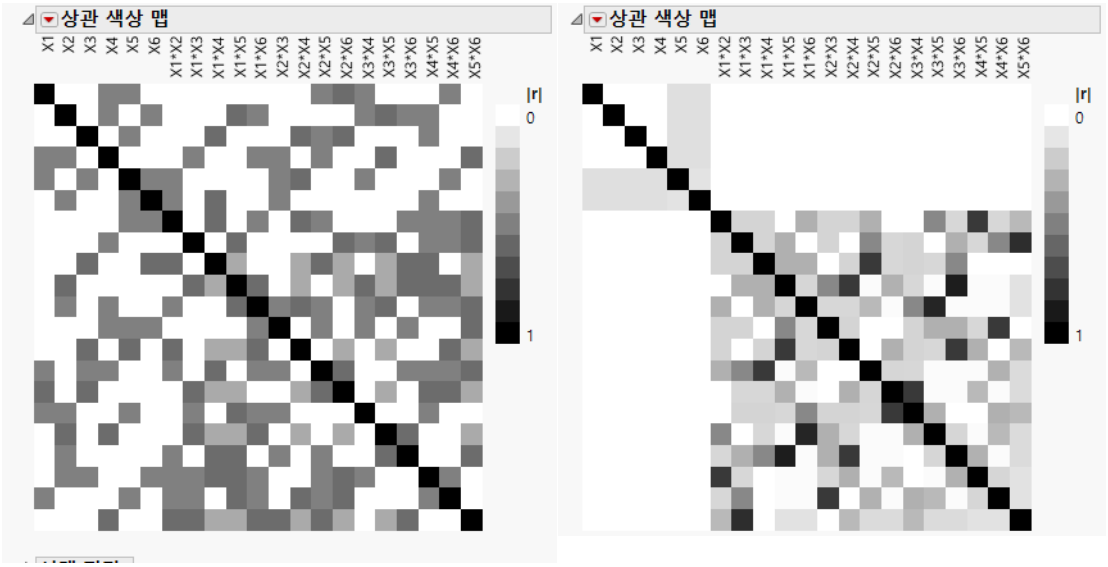


그림 7.15에는 두 가지 색상 맵이 모두 표시되지만 Plackett-Burman 색상 맵에서 주효과와 이원 교호작용이 포함된 부분만 표시됩니다. 삼원 교호작용 없이 Plackett-Burman 설계의 색상 맵을 구성하려면 설계를 구성한 다음, "설계 평가"를 사용하여 색상 맵을 확인하십시오.

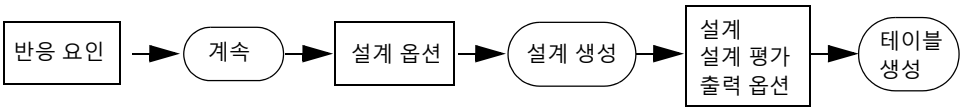
왼쪽의 Plackett-Burman 설계 색상 맵에서 대부분의 2요인 교호작용이 주효과와 상관되어 있음을 확인할 수 있습니다. 이는 무시할 수 없는 2요인 교호작용이 몇 가지 주효과를 편향시킨다는 것을 의미합니다. 이로 인해 활성 주효과를 식별하지 못하거나, 비활성 주효과가 활성 효과라는 잘못된 결론을 내릴 수 있습니다.

이를 오른쪽의 확증적 선별 설계 색상 맵과 대조해 봅니다. 추가 런이 두 개 포함된 확증적 선별 설계는 주효과의 분산이 약간 증가하는 대신 주효과 및 2요인 교호작용에 대한 완전한 독립성을 얻습니다.

확증적 선별 설계 창

설계 단계를 진행하는 동안 "확증적 선별 설계" 창이 업데이트됩니다. 자세한 내용은 "DOE로 시작" 장의 "DOE 워크플로우: 기술, 지정, 설계"(56페이지)에서 확인하십시오. 개요는 개요를 업데이트하는 버튼을 클릭할 때 구분되어 나타나며 그림 7.16의 흐름을 따릅니다.

그림 7.16 확증적 선별 설계 흐름

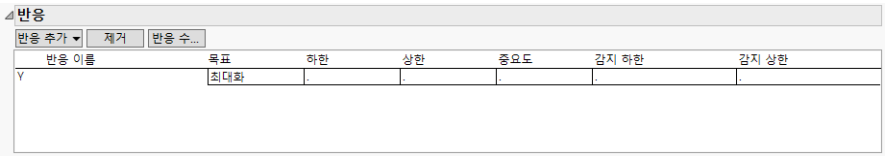


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다 . 이 옵션을 사용하면 반응 이름 , 목표 , 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다 .

그림 7.17 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다 . " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다 . " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다 .

JMP PRO 함수 (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다 . 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다 .

제거 선택한 반응을 제거합니다 .

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다 . 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다 . 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다 . 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다 .

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응 열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .

- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .
- " 없음 " 목표는 최적화에 관한 목표가 없음을 나타냅니다 . 만족도 함수가 구성되지 않습니다 .

참고 : 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다 . 데이터 테이블에서 반응 열의 " 열 정보 " 창을 열고 (**열** > **열 정보**) 원하는 목표값을 입력하십시오 .

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다 . 이러한 값은 반응에 대한 가중치로 처리됩니다 . 반응이 하나뿐이면 기본적으로 1 로 설정되므로 중요도를 지정할 필요가 없습니다 .

JMP PRO **감지 한계** 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다 . 감지 한계는 열 특성으로 저장됩니다 . 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오 .

반응 개요 편집

" 반응 " 개요에서 다음 사항에 유의하십시오 .

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다 .
- 목표를 클릭하여 변경할 수 있습니다 .
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다 .
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다 .

반응 한계 열 특성

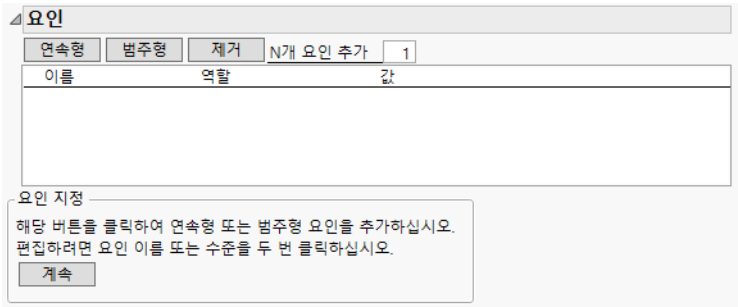
반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

요인

"요인" 개요에서 요인을 추가합니다.

그림 7.18 요인 개요



" 요인 " 개요에는 다음과 같은 버튼이 포함되어 있습니다 .

연속형 **N 개 요인 추가**에 지정된 수의 연속형 요인을 입력합니다 .

범주형 **N 개 요인 추가**에 지정된 수의 명목형 요인을 입력합니다 .

제거 선택한 요인을 제거합니다 .

N 개 요인 추가 지정된 유형의 여러 요인을 추가합니다 . 추가할 요인 수를 입력하고 " 연속형 " 또는 " 범주형 " 을 클릭합니다 . 유형이 다른 여러 요인을 추가하려면 **N 개 요인 추가**를 반복합니다 .

팁 : " 요인 " 패널을 완료하면 빨간색 삼각형 메뉴에서 요인 저장을 선택하십시오 . 이렇게 하면 요인 이름과 값을 데이터 테이블에 저장하여 나중에 다시 불러올 수 있습니다 . 자세한 내용은 " 확증적 선별 설계 옵션 "(244 페이지) 에서 확인하십시오 .

" 요인 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할을 지정합니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인 유형이 적절하게 모델링되도록 합니다 .

값 요인에 대한 실험 설정입니다 . 값을 삽입하려면 기본값을 클릭한 후 원하는 값을 입력합니다 .

요인 개요 편집

" 요인 " 개요에서 다음 사항에 유의하십시오 .

- 요인 이름을 편집하려면 해당 요인 이름을 두 번 클릭합니다 .
- 값을 편집하려면 **값** 열의 값을 클릭합니다 .

요인 유형

연속형 숫자 데이터 유형만 가능합니다 . 연속형 요인은 개념적으로 측정 시스템의 제한을 고려하여 사용자가 제공하는 하한과 상한 사이의 어떤 값이든 설정할 수 있는 요인입니다 .

범주형 두 가지 수준의 숫자 또는 문자 데이터 유형에 해당됩니다 . 범주형 요인의 경우 값 순서는 왼쪽에서 오른쪽으로 입력되는 값의 순서입니다 . 이 순서는 설계 데이터 테이블이 생성된 후 " 값 순서 " 열 특성에 저장됩니다 .

요인 열 특성

각 요인에 대한 다양한 열 특성이 데이터 테이블에 저장됩니다 .

설계 역할 각 요인에 " 설계 역할 " 열 특성이 할당됩니다 . 요인을 정의할 때 지정하는 역할에 따라 " 설계 역할 " 열 특성의 값이 결정됩니다 ." 설계 옵션 " 에서 블록을 추가하면 해당 요인에 " 블록화 " 값이 할당됩니다 ." 설계 역할 " 특성은 실험 데이터를 모델링할 때 요인이 어떤 용도로 사용되는지를 반영합니다 ." 설계 역할 " 값은 설계 확대 플랫폼에서 사용됩니다 .

요인 변경 각 요인에 " 쉬움 " 으로 설정된 " 요인 변경 " 열 특성이 할당됩니다 . 확증적 선별 설계에서는 각 실험 런에 대해 요인 수준을 변경할 수 있다고 가정합니다 ." 요인 변경 " 값은 설계 평가 및 설계 확대 플랫폼에서 사용됩니다 .

코딩 " 설계 역할 " 이 " 연속형 " 인 경우 요인의 " 코딩 " 열 특성이 저장됩니다 . 이 특성은 하한값과 상한값이 각각 –1 과 +1 에 해당하도록 요인 값을 변환합니다 ." 최소 제곱 적합 " 보고서의 추정값과 검정은 변환된 값을 기반으로 합니다 .

값 순서 " 설계 역할 " 이 " 범주형 " 또는 " 블록화 " 인 경우 요인의 " 값 순서 " 열 특성이 저장됩니다 . 이 특성은 요인 수준이 나타나는 순서를 결정합니다 .

블록당 런 수 각 블록의 런 수를 나타냅니다 ." 설계 옵션 " 개요에서 블록을 선택하고 " 설계 생성 " 을 클릭하면 기본 이름이 " 블록 " 인 요인이 " 요인 " 목록에 추가됩니다 ." 블록당 런 수 " 열 특성은 해당 요인에 대해 저장됩니다 .

설계 옵션

"설계 옵션" 개요에서는 블록화 구조, 블록 수 및 추가 런 수를 지정할 수 있습니다. 블록 효과는 주효과와 직교합니다. 블록 크기가 같을 필요는 없습니다.

이 개요에는 다음 옵션이 포함되어 있습니다 .

블록이 필요하지 않음 설계에 블록 요인이 포함되지 않음을 나타냅니다 . 기본적으로 이 항목이 선택됩니다 .

2 차 효과 추정을 위해 중앙점 런과 블록 추가 " 블록 수 " 텍스트 상자에 지정된 수의 블록을 추가합니다 . 블록 효과가 주효과와 직교하고 모든 주효과 및 2 차 효과로 구성된 모형을 추정할 수 있는 설계를 구성합니다 . 자세한 내용은 "2 차 효과 추정을 위해 중앙점 런과 블록 추가 "(240 페이지) 에서 확인하십시오 .

추가적인 중앙점 런 없이 블록 추가 " 블록 수 " 텍스트 상자에 지정된 수의 블록을 추가합니다 . 설계 구조에 필요한 만큼만 중앙점 런을 추가합니다 . 블록 효과가 주효과와 직교하지만 모든 주효과 및 2 차 효과로 구성된 모형을 추정할 수 없는 설계를 구성합니다 . 자세한 내용은 " 추가적인 중앙점 런 없이 블록 추가 "(241 페이지) 에서 확인하십시오 .

참고 : 일부 2 차 효과는 중요하지 않다고 가정할 수 있는 경우에만 **추가적인 중앙점 런 없이 블록 추가** 옵션을 사용하십시오 .

블록 수 추가할 블록 수를 나타냅니다 . 추가할 수 있는 블록 수는 2 개에서 요인 수까지입니다 .

추가 런 수 **유효 모형 선택**을 수행할 수 있는 비중앙점 런을 추가합니다 . 자세한 내용은 " 추가 런 "(250 페이지) 및 " 확증적 선별 적합 플랫폼 " 장의 "DSD 를 위한 유효 모형 선택 "(255 페이지) 에서 확인하십시오 .

팁 : 추가 런 옵션을 사용하여 설계에 런을 추가하면 활성 효과가 많이 있을 때 효과를 감지할 수 있는 능력이 향상됩니다 . 권장되는 추가 런 수는 4 회이며 , 이렇게 하면 활성 2 차 효과 (second-order effect) 를 식별할 수 있도록 설계 성능이 크게 향상됩니다 .

설계 생성 설계를 생성하여 " 설계 " 개요에 표시하고 " 설계 평가 " 개요에 평가 정보를 제공합니다 . 설계 테이블을 생성할 수 있는 " 출력 옵션 " 패널도 나타납니다 .

확증적 선별 설계의 블록화

이 섹션에서는 다음과 같은 두 가지 블록화 옵션에 대해 설명합니다 .

- "2 차 효과 추정을 위해 중앙점 런과 블록 추가 "(240 페이지)
- " 추가적인 중앙점 런 없이 블록 추가 "(241 페이지)

2 차 효과 추정을 위해 중앙점 런과 블록 추가

참고 : 2 차 효과를 추정하는 블록화된 설계의 구성 및 특성에 대한 자세한 내용은 Jones and Nachtsheim(2016) 에서 확인하십시오 . 이 연구 자료에는 블록을 임의 효과로 처리하는 데 대한 내용도 포함되어 있습니다 .

2차 효과 추정을 위해 중앙점 런과 블록 추가 옵션은 다음과 같은 특성이 있는 설계를 구성합니다.

- 블록 효과가 주효과와 직교합니다 .
- 모든 주효과와 2 차 효과로 구성된 모형을 추정할 수 있습니다 .

설계에 연속형 요인만 포함된 경우 k 요인에 대해 이러한 특성이 있는 블록화된 설계를 다음과 같이 구성할 수 있습니다 .

- k 요인에 대한 DSD 설계에서 중앙점 런을 제거합니다 .
- 같은 블록에 컨퍼런스 행렬 접기 쌍을 할당합니다 .
- 각 블록에 중앙점 런을 하나씩 추가합니다 .

일부 요인이 범주형인 경우 **2차 효과 추정을 위해 중앙점 런과 블록 추가** 옵션은 특정 블록 내에 중앙점 런 쌍을 추가합니다. 이 구조는 직교성 및 모든 주효과와 2차 효과의 추정 가능성을 보장합니다.

블록에 접기 쌍이 포함되어야 한다는 블록 크기 요구 사항이 있으므로 블록 수의 범위는 *k*가 짝수이면 2 ~ *k*개, *k*가 홀수이면 2 ~ *k+1*개일 수 있습니다. 자세한 내용은 "컨퍼런스 행렬 및 런 수"(249 페이지)에서 확인하십시오. JMP는 같은 크기의 블록을 구성하려고 시도합니다.

추가적인 중앙점 런 없이 블록 추가

추가적인 중앙점 런 없이 블록 추가 옵션은 모든 요인이 연속형일 때 중앙점 런이 하나 있고 일부 요인이 범주형일 때는 중앙점 런이 두 개 있는 설계를 구성합니다. 결과 설계의 특성은 다음과 같습니다.

- 블록 효과가 주효과와 직교합니다.
- 블록 효과가 2 차 효과의 선형 조합과 교락될 가능성이 있습니다. 즉, 모든 주효과와 2 차 효과로 구성된 모형을 추정하지 **못할 수** 있습니다.

따라서 일부 2차 효과를 무시할 수 있다고 가정할 수 있는 경우에만 이 옵션을 사용하십시오.

추가 중앙점 런 없이 *k* 요인에 대한 블록화된 설계는 다음과 같이 구성합니다.

- 같은 블록에 컨퍼런스 행렬 접기 쌍을 할당합니다.
- 모든 요인이 연속형인 경우 단일 블록에 단일 중앙점 런을 할당합니다.
- 범주형 요인이 있는 경우 블록화되지 않은 확증적 선별 설계에는 컨퍼런스 행렬에 의해 정의된 접기 쌍에 두 개의 중앙점 런을 추가해야 합니다. 자세한 내용은 "컨퍼런스 행렬 및 런 수"(249 페이지)에서 확인하십시오. 추가 중앙점 런 없이 블록화된 설계를 구성하기 위해 이러한 두 개의 중앙점 런이 단일 블록에 추가됩니다.

블록에 접기 쌍이 포함되어야 한다는 블록 크기 요구 사항이 있으므로 블록 수의 범위는 *k*가 짝수이면 2 ~ *k*개, *k*가 홀수이면 2 ~ *k+1*개일 수 있습니다. 자세한 내용은 "컨퍼런스 행렬 및 런 수"(249 페이지)에서 확인하십시오. JMP는 같은 크기의 블록을 구성하려고 시도합니다.

설계

"설계" 개요에서는 확증적 선별 설계를 위한 런을 보여 줍니다. 런은 표준 순서로 제공됩니다. 설계 테이블의 런 순서를 변경하려면 테이블을 생성하기 전에 "출력 옵션" 패널에서 "런 순서" 옵션을 선택할 수 있습니다.

참고 : 4 개 이하의 요인에 대한 확증적 선별 설계는 5 개 요인 설계를 기반으로 합니다. 자세한 내용은 "4 개 이하의 요인에 대한 확증적 선별 설계"(250 페이지)에서 확인하십시오.

설계 평가

참고 : Cotter 설계에는 "설계 평가" 개요가 표시되지 않습니다.

"설계 평가" 개요에서는 생성된 설계의 특성을 평가하는 다양한 방법을 제공합니다. "설계 평가" 개요를 열고 다음 옵션을 확인합니다.

검정력 분석 주어진 크기의 효과를 감지할 수 있는지 분석할 수 있습니다.

예측 분산 프로파일 요인 설정 범위에 대한 예측 분산을 보여 줍니다.

설계 공간 비율 그림 모형 예측 분산이 지정된 값보다 얼마나 아래 또는 위에 있는지 보여 줍니다.

예측 분산 표면 두 개의 연속형 요인에 대한 예측 분산의 표면 그림을 보여 줍니다.

추정 효율도 각 모수에 대해 이상적인 직교 설계 (존재하지 않을 수 있음) 와 비교하여 신뢰 구간 길이에 대한 증가 비율을 제공합니다. 모수의 상대 표준 오차도 제공합니다.

별칭 행렬 모형에 없지만 활성화 가능성이 있는 효과에 의해 모형 모수가 편향되는 정도를 나타내는 계수를 제공합니다.

상관 색상 맵 강도 척도를 사용하여 그림에서 효과 사이의 절대 상관을 표시합니다.

설계 진단 설계를 구성하는 데 사용되는 최적 기준을 나타냅니다. 설계의 효율도 측도도 제공합니다.

> **참고 :** 설계 진단에 사용된 모형에는 모든 주효과 및 2 요인 교호작용이 포함됩니다 (모든 2 요인 교호작용을 추정할 수 있는 경우). 그렇지 않은 경우에는 모형에 모든 주효과가 포함됩니다.

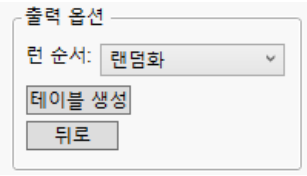
"설계 평가" 패널에 대한 자세한 내용은 "설계 평가" 장의 "설계 평가"(418 페이지)에서 확인하십시오.

출력 옵션

"출력 옵션" 패널을 사용하여 다음 작업을 수행할 수 있습니다.

- 설계 데이터 테이블의 런 순서를 지정합니다.
- 설계 테이블을 구성합니다.
- "확증적 선별 설계" 창의 이전 지점으로 돌아갑니다.

그림 7.19 출력 옵션 패널



"출력 옵션" 패널에는 다음 옵션이 포함되어 있습니다.

런 순서

런 순서 옵션은 설계 테이블의 런 순서를 결정합니다. 다음과 같은 옵션을 선택할 수 있습니다.

동일하게 유지 설계 테이블에 표시되는 행 순서가 " 설계 " 개요의 순서와 동일합니다.

좌에서 우로 정렬 설계 테이블의 열이 왼쪽에서 오른쪽으로 정렬됩니다.

랜덤화 설계 테이블의 행이 랜덤 순서로 나타납니다.

우에서 좌로 정렬 설계 테이블의 열이 오른쪽에서 왼쪽으로 정렬됩니다.

블록 내 랜덤화 설계 테이블의 행이 블록 내에서 랜덤 순서로 나타납니다.

테이블 생성

테이블 생성을 클릭하여 확증적 선별 설계 데이터 테이블을 구성합니다.

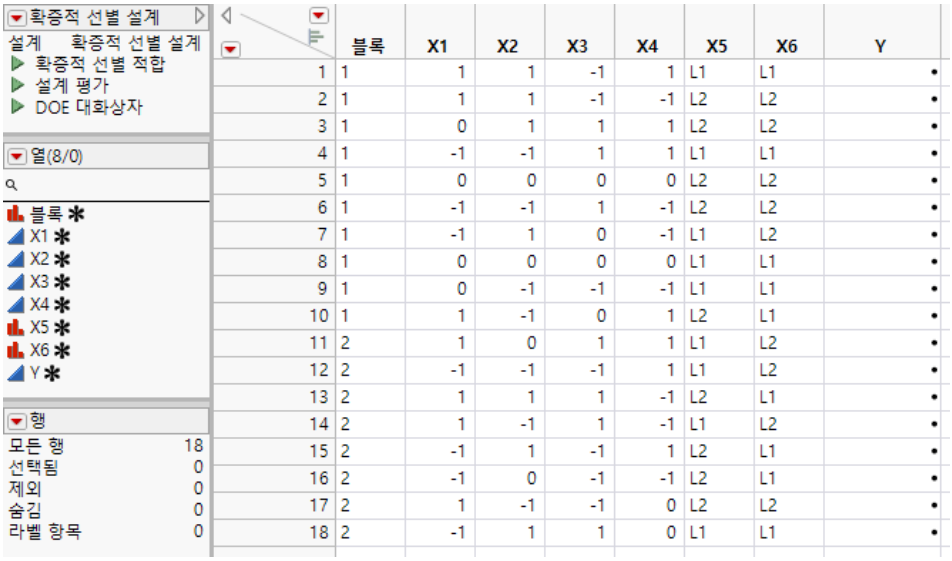
" 확증적 선별 설계 " 테이블의 왼쪽 위에 있는 테이블 패널에는 다음 스크립트가 포함되어 있습니다. 스크립트를 실행하려면 스크립트 이름 옆의 녹색 삼각형을 클릭합니다.

확증적 선별 적합 **DOE > 확증적 선별 > 확증적 선별 적합** 플랫폼을 실행합니다. 자세한 내용은 8 장 , " 확증적 선별 적합 플랫폼 " 에서 확인하십시오.

설계 평가 **DOE > 설계 진단 > 설계 평가** 플랫폼을 실행합니다. 자세한 내용은 15 장 , " 설계 평가 " 에서 확인하십시오.

DOE 대화상자 설계 테이블을 생성하는 데 사용된 " 확증적 선별 설계 " 창을 다시 생성합니다.

그림 7.20 스크립트를 보여 주는 확증적 선별 설계 테이블



	블록	X1	X2	X3	X4	X5	X6	Y
1	1	1	1	-1	1	L1	L1	•
2	1	1	1	-1	-1	L2	L2	•
3	1	0	1	1	1	L2	L2	•
4	1	-1	-1	1	1	L1	L1	•
5	1	0	0	0	0	L2	L2	•
6	1	-1	-1	1	-1	L2	L2	•
7	1	-1	1	0	-1	L1	L2	•
8	1	0	0	0	0	L1	L1	•
9	1	0	-1	-1	-1	L1	L1	•
10	1	1	-1	0	1	L2	L1	•
11	2	1	0	1	1	L1	L2	•
12	2	-1	-1	-1	1	L1	L2	•
13	2	1	1	1	-1	L2	L1	•
14	2	1	-1	1	-1	L1	L2	•
15	2	-1	1	-1	1	L2	L1	•
16	2	-1	0	-1	-1	L2	L1	•
17	2	1	-1	-1	0	L2	L2	•
18	2	-1	1	1	0	L1	L1	•

뒤로

"뒤로" 버튼을 클릭하면 "설계 생성"을 클릭하기 전의 위치로 이동합니다. 이전 개요를 변경하고 설계를 다시 생성할 수 있습니다.

확증적 선별 설계 옵션

확증적 선별 설계 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

반응 저장 "반응" 패널의 정보를 새 데이터 테이블에 저장합니다. 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다. 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다.

반응 불러오기 "반응 저장" 옵션을 사용하여 저장한 반응을 불러옵니다.

요인 저장 "요인" 패널의 정보를 새 데이터 테이블에 저장합니다. 각 요인의 열에는 수준이 포함됩니다. 다른 정보는 열 특성으로 저장됩니다. 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다. 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다. 그런 다음 "설계 역할" 영역에서 적절한 역할을 선택합니다.

요인 불러오기 "요인 저장" 옵션을 사용하여 저장한 요인을 불러옵니다.

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) "요인 제약 조건 정의" 또는 "선형 제약 조건" 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다. 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다. 마지막 행은 부등식 한계를 포함합니다. 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다. 자세한 내용은 "열 특성" 부록의 "제약 조건 상태"(762 페이지)에서 확인하십시오.

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) "제약 조건 저장" 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다.

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다. 이러한 작업에는 다음 중 하나 이상이 포함됩니다.

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다.

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다 .

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다 .
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다 .

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다 .

- **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다 . 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 **기본 분석**의 " 시뮬레이션 " 장에서 확인하십시오 .

행렬 X 저장 " 적률 행렬 " 및 " 모형 행렬 " 을 설계 데이터 테이블의 테이블 스크립트에 저장합니다 . 이 스크립트에는 적률 및 설계 행렬이 포함됩니다 . 자세한 내용은 " 사용자 설계 " 장의 " 행렬 X 저장 "(105 페이지) 에서 확인하십시오 .

주의 : 명목형 요인이 있는 설계의 경우 " 행렬 X 저장 " 옵션으로 저장한 모형 행렬은 선형 모형을 적합시킬 때 사용된 코딩 행렬이 **아닙니다** . 확증적 모형 적합에서 **모형 실행**을 선택할 때 나타나는 " 모형 적합 " 보고서에서 " 열 저장 > 코드화 테이블 저장 " 옵션을 선택하면 모형 적합에 사용되는 코딩 행렬을 가져올 수 있습니다 .

스크립트를 스크립트 창에 저장 " 확증적 선별 설계 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

반응 시뮬레이션

" 테이블 생성 " 을 클릭하여 설계 테이블을 생성할 때 " 반응 시뮬레이션 " 옵션이 각 반응에 대해 다음을 수행합니다 .

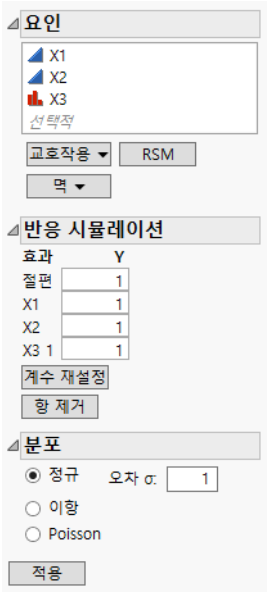
- 설계 테이블의 반응 열에 확률 반응 값을 추가합니다 .
- 시뮬레이션 모형 계산식을 포함하는 새 열을 설계 테이블에 추가합니다. 계산식과 값은 주효과 모형에 기초합니다 .

모형을 정의하기 위해 효과를 추가 및 제거하고, 모수 값을 지정하고, 시뮬레이션을 위한 반응 분포를 선택할 수 있는 "모형" 창이 열립니다. "모형" 창에서 "적용"을 클릭하면 시뮬레이션 모형 계산식을 포함하는 각 열이 업데이트됩니다.

컨트롤 창

그림 7.21에서는 연속형 요인이 두 개(X1, X2), 2수준 범주형 요인이 한 개(X3) 있는 설계의 "모형 "창을 보여 줍니다.

그림 7.21 반응 시뮬레이션 컨트롤 창



이 창에는 다음과 같은 세 가지 개요가 있습니다 .

- 요인
- 반응 시뮬레이션
- 분포

초기 "반응 시뮬레이션" 개요에는 모든 계수의 값이 1인 주효과 모형의 항이 표시됩니다. "분포" 개요에는 오차 표준편차가 1인 정규 분포가 표시됩니다. DOE 창의 "설계 평가"에서 "검정력 분석"의 일부로 "예상 계수"를 설정한 경우, "검정력 분석" 개요에서 "예상 계수" 및 "예상 RMSE"(오차 표준편차)로 지정한 값이 "반응 시뮬레이션" 개요의 초기값이 됩니다.

요인

" 요인 " 개요를 사용하여 시뮬레이션 모형에 항을 추가합니다 .

교호작용 목록에서 요인을 선택합니다 . 그런 다음 " 교호작용 " 메뉴에서 교호작용 차수를 선택합니다 . 해당 교호작용은 " 반응 시뮬레이션 " 개요의 " 효과 " 목록에 추가됩니다 .

RSM 가능한 모든 반응 표면 항을 " 반응 시뮬레이션 " 개요의 " 효과 " 목록에 추가합니다 .

멱 목록에서 요인을 선택합니다 . 그런 다음 " 멱 " 메뉴에서 차수를 선택합니다 . 해당 멱은 " 반응 시뮬레이션 " 개요의 " 효과 " 목록에 추가됩니다 .

반응 시뮬레이션

시뮬레이션된 값에 대한 모형을 지정하려면 다음을 수행하십시오 .

1. " 효과 " 목록의 각 항에 대해 반응 값을 시뮬레이션하는 데 사용되는 선형 모형의 계수를 입력합니다 . 이 값은 선형 함수 $L(\mathbf{x}, \beta) = \mathbf{x}'\beta$ 를 정의합니다 . 그림 7.21 의 " 반응 시뮬레이션 " 개요를 참조하십시오 .
   - 벡터 **x** 는 " 효과 " 아래에 나열된 효과를 정의하는 항으로 구성됩니다 .
   - 벡터 β 는 "Y" 아래에 지정한 모형 계수의 벡터입니다 .
2. " 분포 " 에서 반응 분포를 선택합니다 .
3. **적용**을 클릭합니다 . 시뮬레이션된 값과 계산식을 포함하는 <Y> 시뮬레이션 열이 설계 테이블에 추가됩니다 . 여기서 Y 는 반응 열의 이름입니다 .

계수 재설정 모든 계수를 0 으로 설정합니다 .

항 제거 " 효과 " 목록에서 항을 제거합니다 . 제거할 효과를 선택하고 " 항 제거 " 를 클릭합니다 . 이때 주효과는 제거할 수 없습니다 .

분포

"반응 시뮬레이션" 창에서 사용 가능한 다음 분포 중 하나를 선택합니다.

정규 정규 분포에서 값을 시뮬레이션합니다 . " 오차 σ" 에 정규 분포 오차의 표준편차 값을 입력합니다 . " 요인 " 개요에서 요인의 " 변경 " 을 " 어려움 " 으로 지정한 경우 " 주구 σ" 에 주구 오차 값을 입력할 수 있습니다 . 요인의 " 변경 " 을 " 어려움 " 및 " 매우 어려움 " 으로 지정한 경우에는 하위구 오차와 주구 오차 값을 둘 다 입력할 수 있습니다 . " 적용 " 을 클릭하면 모형에 기초하여 확률 반응 벡터가 포함된 계산식과 난수 값이 <Y> 시뮬레이션 열에 입력됩니다 .

이항 이항 분포에서 값을 시뮬레이션합니다 . "N" 에 시도 수 값을 입력합니다 . 성공 확률이 $1/(1 + exp(-L(\mathbf{x}, \beta)))$ 이고 시도 횟수가 N 인 이항 분포에 따라 정수인 난수가 생성됩니다 . "적용 " 을

클릭하면 난수 값과 계산식이 <Y> 시뮬레이션 열에 입력됩니다 . N 값을 포함하는 시도 수 열도 데이터 테이블에 추가됩니다 .

Poisson 모수 *exp(L(***x,** β*))* 를 사용하여 Poisson 분포에 따라 정수인 난수를 시뮬레이션합니다 . " 적용 " 을 클릭하면 난수 값과 계산식이 <Y> 시뮬레이션 열에 입력됩니다 .

참고 : " 테이블 생성 " 을 클릭할 때마다 반응을 시뮬레이션하도록 환경 설정을 지정할 수 있습니다 . 이렇게 하려면 **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다 . 그런 다음 **반응 시뮬레이션** 을 선택합니다 .

기술 상세 정보

- " 확증적 선별 설계의 구조 "
- " 실험 데이터 분석 "

확증적 선별 설계의 구조

그림 7.22 에서는 8 개의 연속형 요인과 가짜 요인에 해당하는 4 개의 추가 런을 사용한 확증적 선별 설계의 예를 보여 줍니다 . 다음을 확인할 수 있습니다 .

- 각 행 쌍은 접기 쌍이며 짝수 번호 행의 값은 이전 행의 -1 배입니다 . 설계 접기를 수행하면 2 요인 교호작용과 주효과의 교락이 제거됩니다 .
- 각 요인은 3 회 런의 중앙값으로 설정되며 설계 구성과 함께 모든 2 차 효과를 추정할 수 있게 합니다 .
- 17 ~ 20 행은 가짜 요인에 해당하는 추가 런입니다 .
- 마지막 행에 중앙점 런을 추가하면 절편 및 모든 주효과와 2 차 효과를 포함하는 모형을 적합시킬 수 있습니다 .

이 구조는 연속형 요인에 대한 확증적 선별 설계의 대표적인 예입니다 .

그림 7.22 8 개의 연속형 요인에 대한 확증적 선별 설계

설계

런	X1	X2	X3	X4	X5	X6	X7	X8
1	0	1	1	1	1	1	1	1
2	0	-1	-1	-1	-1	-1	-1	-1
3	1	0	-1	-1	-1	-1	1	1
4	-1	0	1	1	1	1	-1	-1
5	1	-1	0	-1	1	1	-1	-1
6	-1	1	0	1	-1	-1	1	1
7	1	-1	-1	0	1	1	1	1
8	-1	1	1	0	-1	-1	-1	-1
9	1	-1	1	1	0	-1	-1	1
10	-1	1	-1	-1	0	1	1	-1
11	1	-1	1	1	-1	0	1	-1
12	-1	1	-1	-1	1	0	-1	1
13	1	1	-1	1	-1	1	0	-1
14	-1	-1	1	-1	1	-1	0	1
15	1	1	-1	1	1	-1	-1	0
16	-1	-1	1	-1	-1	1	1	0
17	1	1	1	-1	-1	1	-1	1
18	-1	-1	-1	1	1	-1	1	-1
19	1	1	1	-1	1	-1	1	-1
20	-1	-1	-1	1	-1	1	-1	1
21	0	0	0	0	0	0	0	0

컨퍼런스 행렬 및 런 수

JMP 의 확증적 선별 설계는 **컨퍼런스 행렬** (Xiao et al., 2012) 을 사용하여 구성됩니다 . 컨퍼런스 행렬은 *m* x *m* 행렬 **C** 입니다 . 여기서 *m* 은 짝수입니다 . **C** 행렬은 대각 항목이 *0*, 비대각 항목이 *1* 또는 *–1* 이고 다음을 충족합니다 .

$$\mathbf{C}'\mathbf{C} = (m-1)\mathbf{I}_{mxm}.$$

참고 : *m* 의 특정 짝수 값에 대해 컨퍼런스 행렬이 있는지 여부는 알 수 없습니다 .

요인 수 *k* 가 5 이상이라고 가정해 보겠습니다 . $k \leq 4$ 요인인 경우에는 "4 개 이하의 요인에 대한 확증적 선별 설계"(250 페이지) 의 설명을 참조하십시오 .

k 개의 연속형 요인의 경우 컨퍼런스 행렬을 사용할 수 있다고 가정해 보겠습니다 .

- *k* 가 짝수인 경우 *k* x *k* 컨퍼런스 행렬을 사용하여 설계의 *k* 회 런을 정의합니다 . 음의 **–C** 는 설계 접기한 런을 정의합니다 . 절편 , 주효과 및 2 차 효과를 포함하는 모형을 추정할 있도록 설계에 중앙점이 추가됩니다 . 따라서 짝수 *k* 에 대해 확증적 선별 설계의 최소 런 수는 $2k + 1$ 입니다 .
- *k* 가 홀수인 경우 *(k+1)* x *(k+1)* 컨퍼런스 행렬이 사용되고 마지막 열은 삭제됩니다 . 중앙점이 추가됩니다 . 따라서 홀수 *k* 에 대해 선별 설계의 최소 런 수는 $2k + 3$ 입니다 .

일부 요인이 범주형이고 컨퍼런스 행렬을 사용할 수 있는 경우 비슷한 프로시저가 사용됩니다. 자세한 내용은 Jones and Nachtsheim(2013) 에서 확인하십시오 .

- 단일 중앙점 대신 두 개의 추가 런이 필요합니다. 이 두 런은 모든 연속형 요인이 중간 값으로 설정된 중앙점 런입니다 .
- k 개 요인이 있고 k 가 짝수인 경우 설계의 런 수는 $2k+2$ 입니다 .
- k 가 홀수인 경우 런 수는 $2k+4$ 입니다 .

컨퍼런스 행렬을 사용할 수 없는 m 값에 대해서는 그 다음으로 큰 컨퍼런스 행렬을 사용하여 확증적 선별 설계를 구성할 수 있습니다. 결과적으로 필요한 런 수는 연속형 요인의 경우 $2k+3$, 범주형 요인의 경우 $2k+4$를 초과할 수 있습니다.

추가 런

추가 런은 가상(가짜) 요인을 사용하여 구성됩니다. f개의 가짜 요인을 설계에 추가하면 2^f개의 추가 런이 생깁니다.

실험 연구에서 요인 수는 k로 나타냅니다. 4개 또는 8개의 추가 런을 설계에 추가할 수 있습니다. 추가 런은 "컨퍼런스 행렬 및 런 수"(249 페이지)에서 설명된 것처럼 $k+f$개의 요인에 대한 설계를 생성한 후 마지막 f개 열을 삭제하여 구성합니다. 추가 런이 4개만 있어도 모형 선택에 매우 유용할 수 있습니다.

추가 런을 사용하는 방법에 대한 자세한 내용은 "확증적 선별 적합 플랫폼" 장의 "DSD를 위한 유효 모형 선택"(255 페이지)에서 확인하십시오.

4 개 이하의 요인에 대한 확증적 선별 설계

4개 이하의 요인에 대한 확증적 선별 설계는 5개 요인의 확증적 선별 설계를 기준으로 사용하여 구성됩니다. 컨퍼런스 행렬 방법에 따라 엄격하게 구성된 $k \leq 4$ 요인 설계의 경우 바람직하지 않은 특성이 있기 때문입니다. 특히 2차 효과(second-order effect)를 분리하기가 어렵습니다.

$k \leq 4$ 요인을 지정하면 5개 요인에 대한 확증적 선별 설계가 구성되고 불필요한 열이 삭제됩니다. 따라서 $k \leq 4$ 요인이 포함된 블록화되지 않은 설계의 런 수는 모든 요인이 연속형인 경우 13이고 일부 요인이 범주형인 경우에는 14입니다.

실험 데이터 분석

일반적으로 이원 교호작용이 활성일 가능성을 허용하는 모형을 적합시키려고 합니다. 순수 2차 항을 모형에 포함할 수도 있습니다. 완전 2차 모형을 가정하거나, 특정 2차 항만 포함하는 사전 모형을 지정할 수 있습니다.

이원 교호작용

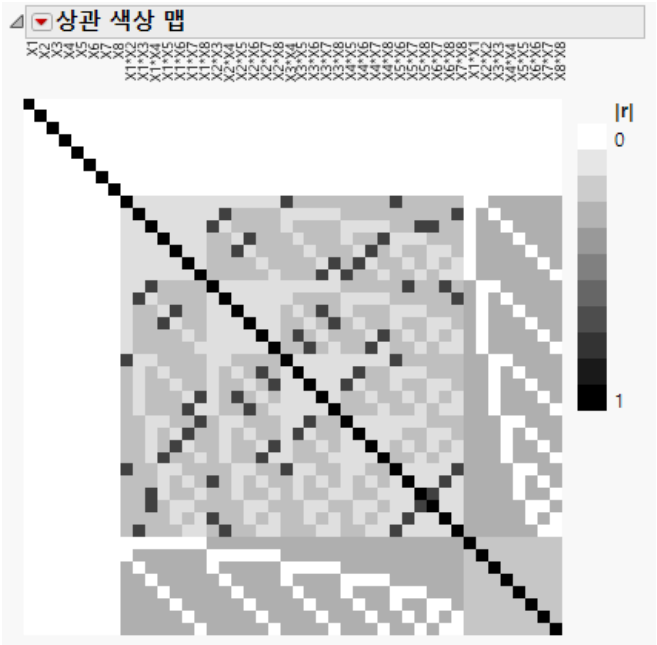
이러한 모형을 적합시킬 때 다음 두 가지 사실을 염두에 두어야 합니다 .

- 이원 교호작용 효과와 2 차 효과는 상관되는 경우가 종종 있습니다 .

- 이원 교호작용 효과와 2 차 효과를 **모두** 동시에 추정할 수는 없습니다 .

그림 7.23에서는 그림 7.22에 나오는 8개의 연속형 요인을 사용한 설계의 상관 색상 맵을 보여 줍니다. 이 색상 맵은 완전 2차 설계에 대한 것입니다. 8개의 순수 2차 효과는 맨 오른쪽에 나열됩니다. "DOE > 설계 진단 > 설계 평가"를 사용하고 "별칭 항" 목록에 적절한 항을 입력하여 이 그림을 구성할 수 있습니다. 자세한 내용은 "설계 평가" 장의 "별칭 항"(418 페이지)에서 확인하십시오.

그림 7.23 완전 2 차 모형의 상관 색상 맵



효과 간 절대 상관을 보려면 커서로 색상 맵을 가리킵니다. 주효과는 모든 이원 교호작용 및 순수 2차 효과와 상관관계가 없음을 알 수 있습니다. 또한 주대각 위에만 검은색 셀만 있으므로 다른 효과와 완전히 교락되는 효과가 없다는 것도 알 수 있습니다. 그러나 2요인 교호작용 간의 절대 상관 중 일부는 0.75로 상당히 큰 편입니다. 또한 2요인 교호작용과 순수 2차 효과 간의 절대 상관은 0 또는 0.3118입니다.

주효과와 순수 2차 효과만 활성화된 경우 주효과와 2차 효과가 포함된 포화 모형을 적합시킬 수 있습니다. 이 모형은 3차 이상의 활성 효과가 없다고 가정할 때 비편향 효과 추정값을 산출합니다.

2 차 효과 (second-order effect) 와 관련된 상관관계 때문에 이원 교호작용이 포함된 모형을 적합시키는 데 주의해야 합니다 . 분석 방법론에는 다음이 포함되며 첫 번째 방법이 선호됩니다 .

- 유효 모형 선택 방법은 특히 많은 효과가 활성일 가능성이 있는 경우에 적합합니다 . 자세한 내용은 " 확증적 선별 적합 플랫폼 " 장의 "DSD 를 위한 유효 모형 선택 "(255 페이지) 에서 확인하십시오 .

- 단계별 전진 회귀 또는 모든 가능한 부분집합 회귀는 다음 조건이 충족될 경우에 적합합니다 .
  - 활성 효과의 수가 런 수의 절반 이하입니다 .
  - 활성 이원 교호작용이 최대 2 개이거나 활성 2 차 효과가 최대 1 개입니다 .

  자세한 내용은 " 단계별 전진 회귀 또는 모든 가능한 부분집합 회귀 "(252 페이지) 에서 확인하십시오 .

단계별 전진 회귀 또는 모든 가능한 부분집합 회귀

이 방법은 먼저 전체 반응 표면 모형을 지정하는 것으로 구성됩니다 . 그 후 다음 중 하나를 수행합니다 .

- 모형 유전성을 보장하기 위해 " 중지 규칙 " 을 " 최소 AICc" 로 설정하고 " 규칙 " 을 " 결합 " 으로 설정한 상태로 단계별 전진 회귀를 사용합니다 .
- 모든 가능 모형 회귀를 사용합니다 . 이때 유전성 제한을 적용하는 옵션을 선택하고 모형 선택에 AICc 기준을 사용합니다 .

런 수가 모수의 수보다 적어서 전체 반응 표면 모형을 적합시킬 수 없습니다 . 따라서 효과 희소성 , 즉 활성 효과의 수가 런 수보다 적다는 가정하에 분석을 수행합니다 . 이 방법에는 다음과 같은 몇 가지 제한이 있습니다 .

- 활성 효과의 수가 런 수의 절반을 초과하면 단계별 회귀와 모든 가능 모형 회귀 둘 다 적절한 모형을 찾는 데 어려움이 있습니다 .
- 중간 2 차 효과를 감지하기 위한 검정력이 낮습니다 . 검정력이 0.9 를 초과하려면 2 차 효과가 오차 표준편차의 3 배보다 많아야 합니다 .
- 효과 교락 때문에 여러 모형이 동등해질 수 있습니다 . 교락을 해결하려면 추가 런이 필요합니다 .

8 장

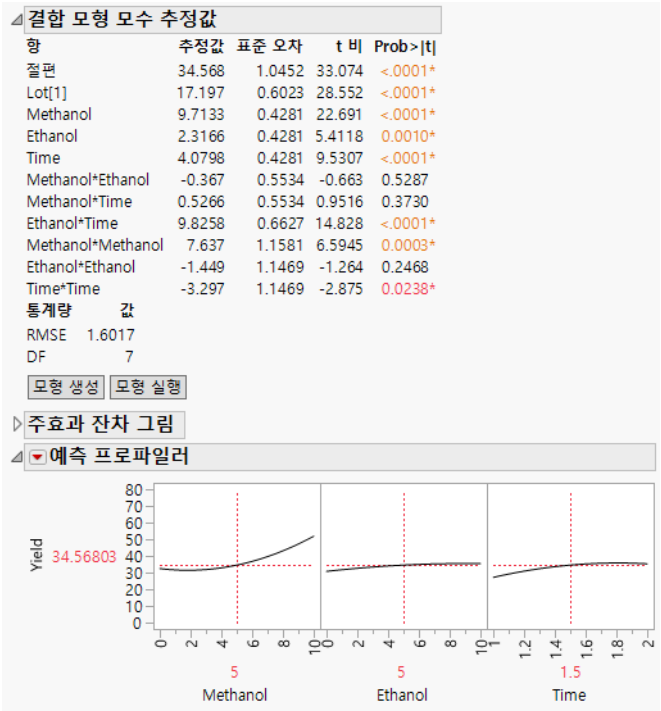
확증적 선별 적합 플랫폼

확증적 선별 실험의 데이터 분석

확증적 선별 적합 플랫폼을 사용하면 DSD를 위한 유효 모형 선택이라는 방법론을 사용하여 DSD(확증적 선별 설계)를 분석할 수 있습니다. 이 방법론은 확증적 선별 설계의 고유 구조를 활용합니다.

DSD에 적용된 표준 모형 선택 방법을 사용하면 활성 효과를 식별하지 못할 수 있습니다. DSD를 위한 유효 모형 선택 알고리즘은 DSD 구조를 활용하여 활성 주효과와 2차 효과(second-order effect)를 식별합니다.

그림 8.1 확증적 선별 적합 결과



목차

확증적 선별 적합 플랫폼 개요 255
DSD에서 활성 효과 식별 255
DSD를 위한 유효 모형 선택 255
확증적 선별 적합 플랫폼의 예 256
모형 적합 256
결과 검토 256
모형 축소 258
확증적 선별 적합 플랫폼 시작 260
확증적 선별 적합 보고서 260
1단계 - 주효과 추정값 261
2단계 - 짝수 차수 효과 추정값 262
결합 모형 모수 추정값 262
주효과 잔차 그림 263
예측 프로파일러 263
확증적 선별 적합 플랫폼 옵션 263
확증적 선별 적합 플랫폼에 대한 기술 상세 정보 264
DSD를 위한 유효 모형 선택 알고리즘 264

확증적 선별 적합 플랫폼 개요

확증적 선별 적합 플랫폼은 DSD의 특수 구조를 활용하는 방법론을 사용하여 DSD(확증적 선별 설계)를 분석합니다. 이 방법론을 **DSD를 위한 유효 모형 선택**이라고 합니다. JMP에서 설계를 생성하면 "확증적 선별 적합"이라는 스크립트가 설계 테이블에 포함됩니다. 이 스크립트는 DSD를 위한 유효 모형 선택 방법론을 사용하여 자동으로 분석을 실행합니다.

DSD 에서 활성 효과 식별

DSD는 단일 실험에서 주효과와 2차 효과(second-order effect)를 식별하는 데 유용한 3수준 설계입니다. 최소 런 크기 DSD는 활성 효과의 수가 런 수의 약 절반 미만인 경우와 효과 크기가 표준편차의 2배를 초과하는 경우에 높은 확률로 활성 항을 정확하게 식별할 수 있습니다.

그러나 4개 이상의 적절하게 선택된 런으로 최소 런 크기 DSD를 확대하면 높은 확률로 훨씬 더 많은 효과를 식별할 수 있습니다. 이러한 런은 **추가 런**이라고 하며 **가짜 요인**이라고 하는 가상의 비활성 요인에 해당합니다. 추가 런에 대한 자세한 내용은 "확증적 선별 설계" 장의 "확증적 선별 설계의 구조"(248 페이지)에서 확인하십시오.

추가 런은 설계의 2차 효과(second-order effect) 감지 능력을 크게 향상시킵니다. 이러한 이유로 Jones and Nachtsheim(2016)에서는 최소 4회의 추가 런을 포함할 것을 권장합니다.

DSD 를 위한 유효 모형 선택

DSD에 표준 모형 선택 방법이 적용되면 활성 효과를 식별하지 못할 수 있습니다. 자세한 내용은 Errore et al. (2017)에서 확인하십시오. 또한 표준 선택 방법은 DSD의 구조를 활용하지 않습니다. 확증적 선별 적합 플랫폼은 DSD의 구조를 최대한 활용하는 DSD를 위한 유효 모형 선택 방법을 사용합니다.

Jones and Nachtsheim(2016)에서는 표준 방법과 DSD를 위한 유효 모형 선택 방법을 사용한 시뮬레이션 연구 결과를 보고합니다. DSD에서 요인 수와 가짜 요인 수의 합을 c로 나타냅니다. 여러 상황에서 활성 주효과 수가 3을 초과하면 최대 $c/2$개의 활성 2차 효과(second-order effect)를 명확하게 식별할 수 있습니다. 강한 효과 유전성을 가정할 때 활성 주효과가 3개 이하이면 모든 활성 2차 효과(second-order effect)를 명확하게 식별할 수 있습니다. **명확한 식별**은 오차 표준편차에 대한 계수 절대값의 비율이 3을 초과하고 효과 감지를 위한 검정력이 0.80을 초과한다는 의미입니다.

확증적 선별 적합 플랫폼 기본 설정은 **강한 효과 유전성**을 가정합니다. 강한 효과 유전성은 A와 B가 모두 포함된 경우에만 A*B 교호작용이 모형에 포함될 수 있음을 의미합니다. 강한 효과 유전성을 위해서는 모형 효과의 모든 낮은 차수 성분이 모형에 포함되어야 합니다. 활성 2차 효과(second-order effect)를 식별할 때 이 알고리즘은 강한 효과 유전성 및 앞에서 명확하게 식별할 수 있는 활성 2차 효과(second-order effect) 수에 대해 언급한 결과를 사용합니다.

DSD에서 주효과와 2차 효과(second-order effect)는 서로 직교합니다. DSD를 위한 유효 모형 선택 방법은 이 사실을 활용합니다. 반응의 선형 공간은 주효과가 차지하는 부분 공간과 이 부분

공간의 직교 여공간으로 구분됩니다. Miller and Sitter(2005)에서는 주효과가 차지하는 선형 부분 공간을 **홀수 공간**이라고 합니다. 주효과, 3요인 효과, 5요인 효과 등 홀수 효과에 대한 모든 정보가 이 공간에 포함되기 때문입니다. 또한 이 공간의 직교 여공간을 **짝수 공간**이라고 합니다. 절편, 2요인 효과, 4요인 효과 등 짝수 효과에 대한 모든 정보가 이 공간에 포함되기 때문입니다.

확증적 선별 적합은 이 개념을 따릅니다. 주효과가 차지하는 부분 공간은 홀수 공간입니다. 이 공간의 직교 여공간, 즉 짝수 공간에는 2차 효과(second-order effect)와 블록 변수(있는 경우)가 포함됩니다. 이 알고리즘에 대한 자세한 내용은 Jones and Nachtsheim(2016) 및 "DSD를 위한 유효 모형 선택 알고리즘"(264페이지)에서 확인하십시오.

확증적 선별 적합 플랫폼의 예

Extraction 3 Data.jmp 샘플 데이터 테이블의 설계는 두 블록의 6 개 요인에 대한 확증적 선별 설계입니다. **2 차 효과 추정을 위해 중앙점 런과 블록 추가** 옵션을 선택하고 추가 런 수를 4 회로 지정하여 설계를 생성했습니다. 결과 설계에 포함된 런 수는 18 회입니다. 이 설계는 확증적 선별 설계의 분석을 탐색하는 데 사용됩니다.

- "모형 적합"
- "결과 검토"
- "모형 축소"

참고 : 연속형 요인과 범주형 요인이 있는 설계의 예를 보려면 Peanut Data.jmp 샘플 데이터 파일을 열고 "Fit Definitive Screening" 테이블 스크립트를 실행하십시오.

모형 적합

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Extraction 3 Data.jmp 를 엽니다.
2. **DOE > 확증적 선별 > 확증적 선별 적합**을 선택합니다.
3. Yield 를 선택하고 **Y** 를 클릭합니다.
4. Lot ~ Time 을 선택하고 **X** 를 클릭합니다.
5. 확인을 클릭합니다.

적합은 2단계 분석을 수행합니다. 알고리즘에 대한 자세한 내용은 "확증적 선별 적합 플랫폼에 대한 기술 상세 정보"(264페이지)에서 확인하십시오.

결과 검토

1 단계 : 주효과 추정값

1단계에서는 활성 가능성이 있는 주효과를 판별합니다.

그림 8.2 주효과에 대한 1 단계 보고서

1단계 - 주효과 추정값

항	추정값	표준 오차	t 비	Prob>|t|
Methanol	9.7133	0.3674	26.438	<.0001*
Ethanol	2.3166	0.3674	6.3055	0.0015*
Time	4.0798	0.3674	11.104	0.0001*

통계량	값
RMSE	1.3747
DF	5

☑ 이차항 강 유전성 따르기
☑ 교호작용 강 유전 따르기

참고 : 가짜 요인은 설계 또는 분석에서 요인으로 나타나지 않습니다.

자유도 2의 오차 제곱합은 두 개의 가짜 요인에 해당하는 4회 런에서 계산됩니다. 가짜 요인은 비활성이므로 이 오차 분산 추정값은 비편향입니다. 각 주효과에 대해 주효과 반응 Y_{ME}를 이 추정값과 대조하여 검정합니다. 이 예에서는 세 가지 요인 Methanol, Ethanol 및 Time의 p 값이 임계값보다 작고 활성으로 유지됩니다. 임계값에 대한 자세한 내용은 "1 단계 방법론"(264 페이지)에서 확인하십시오.

세 가지 비활성 요인인 Propanol, Butanol 및 pH의 변동은 가짜 요인 제곱합과 합동하여 그림 8.2에 표시된 자유도 5의 RMSE 통계량을 산출합니다.

2 단계 : 짝수 차수 효과 추정값

2 단계에서는 단계별 부분집합 선택 방법을 사용하여 활성 가능성이 있는 2차 효과(second-order effect) 목록을 생성합니다. 교호작용 및 2차 항은 2차(second-order effect) 또는 짝수 차수 효과입니다.

그림 8.3 짝수 차수 효과에 대한 2 단계 보고서

2단계 - 짝수 순서 효과 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	34.568	1.3459	25.683	0.0015*
Lot[1]	17.197	0.7757	22.171	0.0020*
Methanol*Ethanol	-0.367	0.7127	-0.515	0.6581
Methanol*Time	0.5266	0.7127	0.7389	0.5369
Ethanol*Time	9.8258	0.8534	11.514	0.0075*
Methanol*Methanol	7.637	1.4914	5.1208	0.0361*
Ethanol*Ethanol	-1.449	1.477	-0.981	0.4299
Time*Time	-3.297	1.477	-2.232	0.1552

통계량	값
RMSE	2.0626
DF	2

1단계에서 세 가지 주효과가 활성으로 식별되었으므로 활성 2차 효과(second-order effect)에 대한 단계별 부분집합 선택 절차는 모든 2차 효과(second-order effect)가 포함될 때까지 계속될 수 있습니다. 2단계에서 6개의 2차 효과(second-order effect)가 모두 보고되므로 2단계 RMSE가 1단계 RMSE보다 더 큽니다. 자세한 내용은 "2 단계 방법론"(266 페이지)에서 확인하십시오.

2단계 보고서에 제공된 자유도 2의 RMSE는 모든 2차 효과(second-order effect) 6개의 최종 부분집합에서 얻은 오차 추정값입니다.

결합 결과

모형에 대해 선택한 효과가 "결합 모형 모수 추정값" 보고서에 나열됩니다.

그림 8.4 결합 모형 모수 추정값 보고서

결합 모형 모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	34.568	1.0452	33.074	<.0001*
Lot[1]	17.197	0.6023	28.552	<.0001*
Methanol	9.7133	0.4281	22.691	<.0001*
Ethanol	2.3166	0.4281	5.4118	0.0010*
Time	4.0798	0.4281	9.5307	<.0001*
Methanol*Ethanol	-0.367	0.5534	-0.663	0.5287
Methanol*Time	0.5266	0.5534	0.9516	0.3730
Ethanol*Time	9.8258	0.6627	14.828	<.0001*
Methanol*Methanol	7.637	1.1581	6.5945	0.0003*
Ethanol*Ethanol	-1.449	1.1469	-1.264	0.2468
Time*Time	-3.297	1.1469	-2.875	0.0238*

통계량	값
RMSE	1.6017
DF	7

보고서 맨 아래에 제공된 RMSE와 자유도는 일반적인 표준 최소 제곱 통계량입니다. 이러한 효과를 최종 모형에 대한 잠재 요인으로 사용합니다.

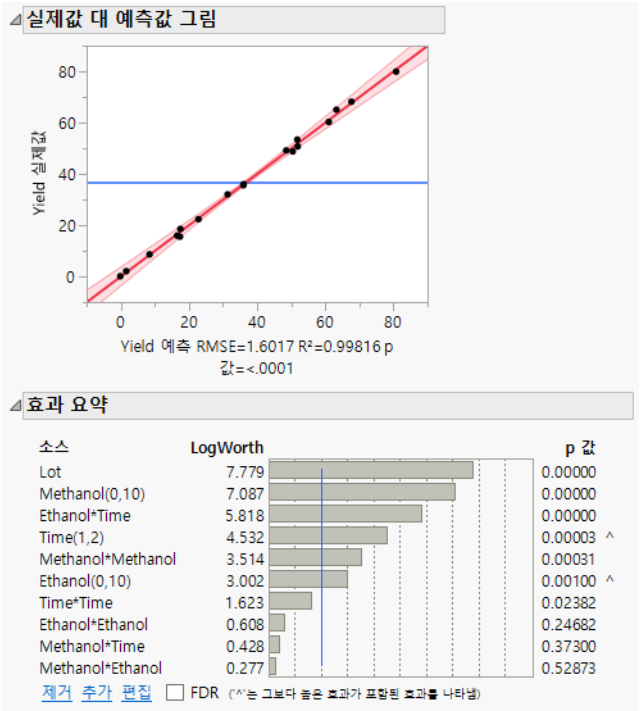
모형 축소

"모형 생성" 버튼은 "모형 적합" 규격 창에 나열된 항에 맞는 모형을 생성합니다. 표준 최소 제곱을 사용하여 모형을 직접 실행하려면 "모형 실행" 버튼을 클릭합니다.

1. **모형 실행**을 클릭합니다.

   "실제값 대 예측값 그림"은 적합 결여가 없음을 보여 줍니다. "효과 요약" 보고서는 모형을 더 축소할 수 있음을 나타냅니다.

그림 8.5 실제값 대 예측값 그림 및 효과 요약 보고서

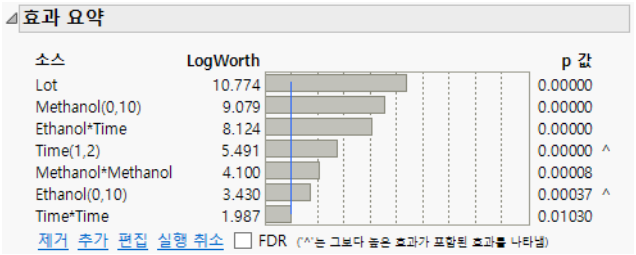


2. " 효과 요약 " 보고서에서 Methanol*Ethanol 을 선택하고 **제거**를 클릭합니다 .

   Methanol*Time 의 p 값은 0.33750 입니다 . 다음에 이 효과를 제거합니다 .
3. " 효과 요약 " 보고서에서 Methanol*Time 을 선택하고 **제거**를 클릭합니다 .

   Ethanol*Ethanol 의 p 값은 0.15885 입니다 . 다음에 이 효과를 제거합니다 .
4. " 효과 요약 " 보고서에서 Ethanol*Ethanol 을 선택하고 **제거**를 클릭합니다 .

그림 8.6 최종 모형의 효과를 보여 주는 효과 요약 보고서



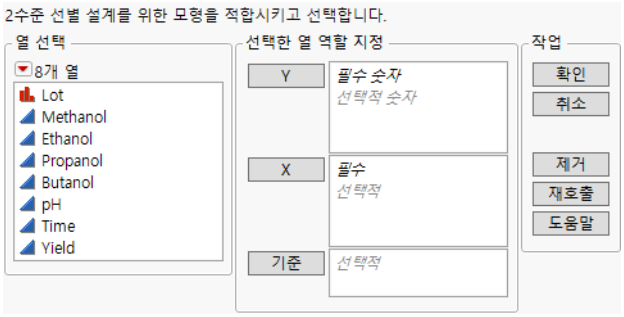
나머지 효과는 유의합니다 . 이러한 효과가 활성 효과인 것으로 결론을 내릴 수 있습니다 .

확증적 선별 적합 플랫폼 시작

확증적 선별 적합 플랫폼을 시작하려면 **DOE** > **확증적 선별** > **확증적 선별 적합**을 선택합니다 . 그림 8.7 의 시작 창은 Extraction3 Data.jmp 를 사용합니다 .

참고 : JMP 에서 설계를 생성하면 " 확증적 선별 적합 " 이라는 스크립트가 설계 테이블에 포함됩니다 . 분석을 바로 시작하려면 이 스크립트를 실행하십시오 .

그림 8.7 확증적 선별 적합 시작 창



Y 하나 이상의 수치형 반응 변수입니다 .

X 연속형 또는 2 수준 범주형 요인입니다 . 이 플랫폼은 분석을 수행할 때 DSD 의 고유 기능을 사용하므로 이러한 요인으로 DSD 또는 설계 접기를 정의해야 합니다 .

기준 각 수준으로 개별 분석을 정의하는 열입니다 . 지정된 열의 각 수준에 대해 해당하는 행이 분석됩니다 . 결과는 별도의 보고서에 나타납니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

확증적 선별 적합 보고서

- "1 단계 - 주효과 추정값 "
- "2 단계 - 짝수 차수 효과 추정값 "
- " 결합 모형 모수 추정값 "
- " 주효과 잔차 그림 "
- " 예측 프로파일러 "

1 단계 - 주효과 추정값

" 주효과 추정값 " 보고서에는 활성으로 식별된 주효과가 나열됩니다 . *p* 값이 *p* 값 임계보다 작은 주효과가 활성으로 간주됩니다 .

- 가짜 요인 또는 중앙점 반복 실험이 있는 경우 모형과 독립적인 오차 분산 추정량이 생성됩니다 . 이 추정값과 대조하여 주효과를 검정합니다 .
- 가짜 요인 또는 중앙점 반복 실험이 없는 경우 비활성 주효과에서 생성된 오차 분산 추정값과 대조하여 주효과 부분집합을 순차적으로 검정합니다 . 이 절차를 실행하려면 하나 이상의 주효과가 비활성이어야 합니다 .

두 경우 모두 비활성 주효과의 변동은 주효과를 검정하는 데 사용된 오차 분산에 통합됩니다.

그림 8.8 1 단계 보고서

1단계 - 주효과 추정값

항	추정값	표준 오차	t 비	Prob>|t|
Methanol	9.7133	0.3674	26.438	<.0001*
Ethanol	2.3166	0.3674	6.3055	0.0015*
Time	4.0798	0.3674	11.104	0.0001*

통계량	값
RMSE	1.3747
DF	5

☑ 이차항 강 유전성 따르기
☑ 교호작용 강 유전 따르기

항 활성으로 식별된 주효과입니다 . "1 단계 방법론 "(264 페이지) 에 설명된 대로 검정했을 때 이 효과의 *p* 값은 임계값보다 작습니다 .

추정값 주효과에 대한 Y 회귀 적합의 모수 추정값입니다 .

표준 오차 1 단계 RMSE 를 사용하여 계산된 추정값의 표준 오차입니다 .

t 비 추정값을 표준 오차로 나눈 값입니다 .

Prob>|t| t 비와 오차 DF(자유도) 를 사용하여 계산된 *p* 값입니다 .

RMSE 1 단계 분석 결과로 구한 평균 제곱 오차의 제곱근입니다 .

- 가짜 요인 또는 중앙점 반복 실험이 있는 경우 , 평균 제곱 오차는 활성으로 식별되지 않은 주효과에서 생성되는 분산 추정값과 합동된 가짜 요인 또는 중앙점의 분산 추정값입니다 .
- 가짜 요인 또는 중앙점 반복 실험이 없는 경우 , 평균 제곱 오차는 활성으로 식별되지 않은 주효과에서 생성되는 분산 추정값입니다 .

DF RMSE 를 생성하는 데 사용되는 오차 추정값과 관련된 자유도입니다 .

- 가짜 요인 또는 중앙점 반복 실험이 있는 경우 DF 는 가짜 요인 수 , 중앙점 반복 실험 수 및 활성으로 식별되지 않은 주효과 수의 합입니다 .
- 가짜 요인 또는 중앙점 반복 실험이 없는 경우 DF 는 활성으로 식별되지 않은 주효과 수입니다 .

2 차 항 강 유전성 따르기 2 단계에서 2 차 항에 적용되는 유전성 규칙을 제어하려면 선택하거나 선택 해제합니다 .

교호작용 강 유전성 따르기 2단계에서 교호작용에 적용되는 유전성 규칙을 제어하려면 선택하거나 선택 해제합니다.

참고 : 강한 효과 유전성은 A 와 B 가 모두 포함된 경우에만 A*B 교호작용이 모형에 포함될 수 있음을 의미합니다. 강한 효과 유전성을 위해서는 모형 효과의 모든 낮은 차수 성분이 모형에 포함되어야 합니다.

2 단계 - 짝수 차수 효과 추정값

"짝수 차수 효과 추정값" 보고서에는 활성으로 식별된 2차 효과(second-order effect)가 나열됩니다. 활성 2차 효과(second-order effect)는 "2 단계 방법론"(266 페이지)에 설명된 대로 단계별 변수 선택 절차 또는 전진 선택을 사용하여 식별됩니다. 블록 효과(포함된 경우)도 유의성과 관계없이 나열됩니다.

그림 8.9 2 단계 보고서

2단계 - 짝수 순서 효과 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	34.568	1.3459	25.683	0.0015*
Lot[1]	17.197	0.7757	22.171	0.0020*
Methanol*Ethanol	-0.367	0.7127	-0.515	0.6581
Methanol*Time	0.5266	0.7127	0.7389	0.5369
Ethanol*Time	9.8258	0.8534	11.514	0.0075*
Methanol*Methanol	7.637	1.4914	5.1208	0.0361*
Ethanol*Ethanol	-1.449	1.477	-0.981	0.4299
Time*Time	-3.297	1.477	-2.232	0.1552

통계량	값
RMSE	2.0626
DF	2

항 활성으로 식별된 블록 요인 및 2 차 효과 (second-order effect) 입니다.

추정값 Y_{2nd} 로 정의된 2 단계 2 차 효과 (second-order effect) 에 대한 Y 회귀 적합의 모수 추정값입니다. 자세한 내용은 " 반응 분해 "(264 페이지) 에서 확인하십시오.

표준 오차 2 단계 RMSE 를 사용하여 계산된 추정값의 표준 오차입니다.

t 비 추정값을 표준 오차로 나눈 값입니다.

Prob>|t| t 비와 오차 DF(자유도) 를 사용하여 계산된 p 값입니다.

RMSE 2 단계 분석 결과로 구한 평균 제곱 오차의 제곱근입니다. RMSE 는 활성으로 식별된 2 차 효과 (second-order effect) 를 적합시킨 후 Y_{2nd} 의 잔차 분산으로 추정됩니다. 자세한 내용은 " 반응 분해 "(264 페이지) 에서 확인하십시오.

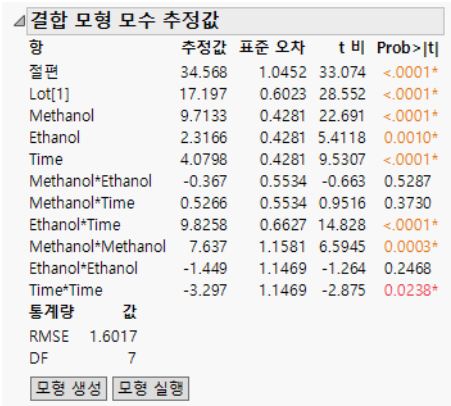
DF RMSE 를 생성하는 데 사용되는 오차 추정값과 관련된 자유도입니다.

결합 모형 모수 추정값

"결합 모형 모수 추정값" 보고서에는 최종 모형의 항과 각 항의 표준 최소 제곱 추정값, 표준 오차, t 비, p 값, RMSE 및 모형 자유도가 나열됩니다.

보고서 아래에는 결합 모형을 생성하거나 실행하는 버튼이 있습니다.

그림 8.10 결합 모형 모수 추정값 보고서



결합 모형 모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	34.568	1.0452	33.074	<.0001*
Lot[1]	17.197	0.6023	28.552	<.0001*
Methanol	9.7133	0.4281	22.691	<.0001*
Ethanol	2.3166	0.4281	5.4118	0.0010*
Time	4.0798	0.4281	9.5307	<.0001*
Methanol*Ethanol	-0.367	0.5534	-0.663	0.5287
Methanol*Time	0.5266	0.5534	0.9516	0.3730
Ethanol*Time	9.8258	0.6627	14.828	<.0001*
Methanol*Methanol	7.637	1.1581	6.5945	0.0003*
Ethanol*Ethanol	-1.449	1.1469	-1.264	0.2468
Time*Time	-3.297	1.1469	-2.875	0.0238*

통계량	값
RMSE	1.6017
DF	7

모형 생성 모형 실행

모형 생성 "결합 모형 모수 추정값" 보고서의 모형 항과 "확증적 선별 적합" 분석에 지정된 반응을 포함하는 "모형 적합" 창에 대한 모형을 생성합니다. "표준 최소 제곱" 분석법이 지정됩니다.

모형 실행 "결합 모형 모수 추정값" 보고서의 모형 항과 "확증적 선별 적합" 분석에 지정된 반응에 대한 표준 최소 제곱 적합을 실행합니다.

주효과 잔차 그림

다른 모든 요인에 적합된 주효과 모형의 잔차 대 표시된 요인의 수준에 대한 그림을 보여 줍니다. 수준에 따라 잔차가 다른 요인은 중요한 주효과를 나타냅니다.

참고 : 블록 요인에 대한 그림은 없습니다.

예측 프로파일러

결합 모형에 대한 예측 프로파일러를 보여 줍니다. 예측 프로파일러에 대한 자세한 내용은 Profilers의 에서 확인하십시오.

확증적 선별 적합 플랫폼 옵션

1 단계 p 값 설정 1 단계에 사용되는 p 값 임계를 설정합니다. 자세한 내용은 "1 단계 방법론"(264 페이지) 에서 확인하십시오.

2 단계 비율 설정 1 단계 MSE 에 대한 2 단계 MSE 의 비율을 지정합니다. 자세한 내용은 "2 단계 방법론"(266 페이지) 에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

확증적 선별 적합 플랫폼에 대한 기술 상세 정보

DSD 를 위한 유효 모형 선택 알고리즘

이 섹션에서는 확증적 선별 적합 플랫폼에 사용되는 알고리즘의 요약 정보를 제공합니다. 자세한 내용은 Jones and Nachtsheim(2016)에서 확인하십시오.

반응 분해

유효 모형 선택 알고리즘은 반응 Y 를 Y_{ME} 와 Y_{2nd} 라는 두 가지 반응으로 표현하여 $Y = Y_{ME} + Y_{2nd}$ 가 되도록 합니다.

- Y_{ME} 는 주효과 및 가짜 요인에 대한 Y 회귀에서 구한 예측값입니다.

  설계의 접기 구조 때문에 Y_{ME} 에 블록 요인을 포함할 필요가 없습니다. 블록 요인은 Y_{2nd} 에 포함됩니다.
- Y_{2nd} 는 $Y_{2nd} = Y - Y_{ME}$ 로 지정됩니다.

참고 : DSD 에서 Y_{ME} 열과 Y_{2nd} 열은 직교입니다.

분석은 다음 두 단계로 진행됩니다.

- 1 단계 : 반응 Y_{ME} 를 사용하여 주효과를 식별합니다. 1 단계에서는 활성으로 간주되는 주효과를 식별합니다.
- 2 단계 : 반응 Y_{2nd} 를 사용하여 2 차 효과 (second-order effect) 를 식별합니다. 2 단계에서는 1 단계의 활성 주효과에 있는 모든 2 차 항을 고려하고 이러한 포함 효과 중에서 활성으로 간주되는 부분집합을 결정합니다.

참고 : 블록 요인이 있으면 유의하지 않더라도 2 단계의 효과 목록에 포함됩니다.

1 단계 방법론

1 단계 방법론은 설계에 가짜 요인 또는 중앙점 반복 실험이 포함되는지 여부에 따라 달라집니다.

사례 1: 가짜 요인 또는 중앙점 반복 실험이 있는 경우

1. 가짜 요인 또는 중앙점 반복 실험을 사용하면 모형과 독립적인 오차 분산 추정량이 생성됩니다 . 3 차 이상의 홀수 차수 활성 효과가 없다고 가정할 경우 이 추정값은 비편향입니다 .
2. Y_{ME} 를 사용하여 주효과를 이 추정값과 대조하여 검정합니다 . *p* 값이 *p* 값 임계보다 작은 주효과가 활성으로 간주됩니다 . 임계값은 다음과 같습니다 .
   - 오차 자유도가 1 인 경우 임계값은 0.20 입니다 .
   - 오차 자유도가 2 인 경우 임계값은 0.10 입니다 .
   - 오차 자유도가 3 이상인 경우 임계값은 0.05 입니다 .
   - 사용자 지정 *p* 값이 임계값입니다 .

   참고 : 다른 *p* 값 임계를 지정하려면 " 확증적 선별 적합 " 의 빨간색 삼각형 메뉴에서 **1 단계 p 값 설정**을 선택하십시오 .
3. 주효과의 *p* 값이 임계값보다 작지 않으면 활성 주효과 및 활성 2 요인 효과가 없다는 결론을 내립니다 . 절차가 종료됩니다 .
4. 활성 주효과가 발견되면 비활성 주효과의 변동이 (1) 에서 생성된 오차 분산과 통합됩니다 .

참고 : 범주형 요인이 설계에 있는 경우 주효과가 활성으로 선택될 때마다 추정 계수가 다시 계산됩니다 .

사례 2: 가짜 요인 또는 중앙점 반복 실험이 없는 경우

이 경우에는 모형에 독립적인 오차 분산 추정량이 제공되지 않습니다 . 비활성 주효과에서 생성된 오차 분산 추정값과 대조하여 주효과 부분집합을 순차적으로 검정합니다 . *m* 개의 주효과가 있다고 가정해 보겠습니다 .

1. Y_{ME} 를 반응으로 사용한 추정 효과의 절대값이 내림차순으로 정렬됩니다 .
2. 각 $1 \le i < m$ 에 대해 절대값이 i^{th} 번째로 큰 효과를 해당 효과 및 절대값이 더 큰 모든 효과가 포함된 모형의 조정된 잔차 제곱합과 대조하여 검정합니다 .
3. 모형에서 *p* 값이 가장 작은 효과가 활성 효과로 간주됩니다 .
4. 활성 주효과가 발견되면 비활성 주효과의 변동을 사용하여 Y_{ME} 를 반응으로 사용한 오차 분산의 추정값을 생성합니다 .

참고 : 사례 2 에서 확증적 선별 적합 절차가 제대로 작동하려면 주효과 중 하나 이상이 활성이어야 하고 하나 이상이 비활성이어야 합니다 . 활성 주효과가 없거나 모든 주효과가 활성이면 절차에서 주효과 집합이 식별되지만 해당 부분집합을 얻기 위한 절차의 정확성이 떨어집니다 .

2 단계 방법론

2 단계에서 고려되는 요인은 강 유전성 옵션에 따라 달라집니다. 강 유전성이 선택되면 1 단계에서 주효과가 활성으로 식별된 요인과 관련된 2 차 효과(second-order effect)만 고려됩니다. 2 단계 방법론은 1 단계에서 식별된 활성 주효과의 수에 따라 달라집니다.

사례 1: 활성 주효과가 7 개 이하인 경우

2 단계에서는 단계별 부분집합 선택 절차를 사용합니다. 목표는 1 단계의 RMSE 에 대한 2 단계의 RMSE 비율이 지정된 임계값보다 큰 경우 계속해서 모형에 2 차 효과(second-order effect)를 추가하는 것입니다. 비율이 임계값보다 작거나 같으면 모형에 추가할 2 차 효과(second-order effect)가 더 없음을 나타냅니다. 기본 임계값은 1 입니다. 임계값이 작을수록 더 큰 임계값에 비해 활성으로 식별될 수 있는 항 수가 증가합니다.

참고 : RMSE 비율 임계값을 1 이 아닌 다른 값으로 지정하려면 " 확증적 선별 적합 " 의 빨간색 삼각형 메뉴에서 **2 단계 비율 설정**을 선택하십시오 .

2 단계에서

- 오차 자유도가 1 인 경우 임계값은 0.20 입니다 .
- 오차 자유도가 2 인 경우 임계값은 0.10 입니다 .
- 오차 자유도가 3 이상인 경우 임계값은 0.05 입니다 .
- 사용자 지정 *p* 값이 임계값입니다 .

1. Y_{2nd} 의 변동을 1 단계의 오차 추정값과 대조하여 검정한 후 2 차 효과 (second-order effect) 로 인한 추가 변동이 있는지 여부를 판별합니다 .
   – 이 검정의 *p* 값이 임계값을 초과하면 절차가 종료되고 활성 2 차 효과 (second-order effect) 가 식별되지 않습니다 .
2. 이 검정의 *p* 값이 임계값보다 작거나 같으면 *k* 크기 (*k = 1,2,3,...*) 의 부분집합이 *k = 1* 부터 시작하여 계속 검정됩니다 .
3. 각 *k* 에 대해 해당 크기의 각 부분집합에 대한 잔차 제곱합을 1 단계의 오차 추정값과 대조하여 검정합니다 . RMSE 가 가장 작은 부분집합이 식별됩니다 .
4. 1 단계 RMSE 에 대한 RMSE 의 비율이 2 단계 비율보다 작은 *k* 가 발견될 때까지 절차가 계속됩니다 .
5. 최종값 *k* 에 해당하는 앞선 부분집합의 효과는 활성 2 요인 효과로 간주됩니다 .

사례 2: 활성 주효과가 8 개 이상인 경우

1 단계에서 활성 주효과가 8 개 이상 식별된 경우 2 단계에서는 2 차 항에 대해 전진 선택을 사용합니다.

9 장

선별 설계

선별 설계는 산업 실험에 가장 많이 사용되는 설계 중 하나입니다. 대개 실험 초기 단계에서 사용되며 여러 요인을 검토하여 반응에 가장 큰 영향을 미치는 요인을 식별합니다. 요인이 식별되면 좀 더 민감한 설계를 사용하여 해당 요인을 연구합니다. 일반적으로 선별 설계에 필요한 실험 런 수가 다른 설계보다 적기 때문에 공정 개선을 시작하기 위한 비교적 저렴하고 효율적인 방법입니다.

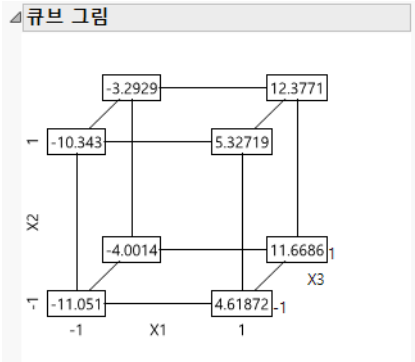
실험 상황에 맞는 표준 선별 설계가 있는 경우 몇 가지 표준 선별 설계 중에서 선택할 수 있습니다. 해당되는 경우 블록화된 설계도 목록에 포함됩니다. 요인은 2수준 연속형 요인, 3수준 범주형 요인 또는 이산형 값만 가정할 수 있는 연속형 요인(**이산 수치형** 요인)일 수 있습니다.

그러나 표준 선별 설계를 사용할 수 없는 상황이 있습니다. 이러한 경우 선별 설계 플랫폼은 **주효과 선별 설계**를 구성합니다. 주효과 선별 설계는 직교 또는 근사 직교 설계입니다. 이 설계는 무시할 수 있는 교호작용이 있는 경우 주효과를 추정하는 데 중점을 둡니다.

또한 JMP는 다음과 같이 선별 설계를 대신하는 두 가지 방법을 제공합니다.

- 확증적 선별 설계는 활성 2 요인 교호작용이 있다고 생각되거나, 반응에 대한 연속형 요인의 효과 그림에 강한 곡률이 나타난다고 의심되는 경우에 특히 유용합니다. 자세한 내용은 7 장, "확증적 선별 설계"에서 확인하십시오.
- 사용자 설계는 다른 방법을 사용하여 구성한 설계보다 훨씬 유연하고 비용 효율적입니다. 자세한 내용은 4 장, "사용자 설계"에서 확인하십시오.

그림 9.1 부분 요인 설계의 결과



목차

선별 설계 개요 269
기본 원리 269
선별 설계 결과 분석 269
선별 설계의 예 270
부분 요인 설계와 주효과 선별 설계 비교 270
표준 설계가 없는 경우의 주효과 선별 설계 277
선별 설계 창 279
반응 280
요인 282
선별 유형 선택 284
부분 요인 설계 목록에서 선택 285
설계 유형 285
교락 측도를 나타내는 해상도 288
설계 표시 및 수정 289
주효과 선별 설계 291
설계 생성 292
설계 293
설계 평가 293
출력 옵션 293
테이블 생성 294
선별 설계 옵션 295
선별 설계의 추가 예 298
부분 요인 설계에서 생성 규칙 수정 298
Plackett-Burman 설계 303

선별 설계 개요

선별 실험은 대개 규모가 작고 반응에 영향을 미치는 요인을 식별하는 데 목적이 있습니다. 정교한 모델링 대신 식별이 목표이므로 선별 설계의 연속형 요인은 보통 두 가지 수준으로 설정됩니다. 그러나 선별 상황에 이산 수치형 요인이나 범주형 요인도 포함될 수 있으며 이 경우 전통적 선별 설계는 상황에 적합하지 않을 수 있습니다. 선별 설계 플랫폼은 세 가지 유형의 요인, 즉 2 수준 연속형 요인, 범주형 요인 및 이산 수치형 요인을 모두 처리할 수 있습니다.

이 설계에는 다음과 같은 두 가지 유형이 있습니다 .

- 전통적 설계 : 표준 선별 설계가 있는 경우 부분 요인 설계 , Plackett-Burman, Cotter 및 혼합 수준 설계가 포함된 목록에서 선택할 수 있습니다 .
- 주효과 선별 설계 : 표준 설계가 가능한지 여부에 관계없이 JMP 에 주효과 선별 설계를 구성하도록 요청할 수 있습니다 . 이 설계는 직교 또는 근사 직교이며 , 무시할 수 있는 교호작용이 있는 경우 주효과를 추정하는 데 중점을 둡니다 . 자세한 내용은 " 주효과 선별 설계 "(291 페이지) 에서 확인하십시오 .

기본 원리

실험 공정의 초기에 주효과 연구에 중점을 두는 것은 효과 계층성이라는 경험적 원리에 의해 뒷받침됩니다. 이 원리는 낮은 차수 효과가 높은 차수 효과보다 중요할 가능성이 높다고 주장합니다. 이러한 이유로 선별 설계는 활성 주효과를 식별하는 데 중점을 둡니다. 차수가 높은 교호작용에 관심이 있는 경우 선별 설계는 2 요인 교호작용이 3 요인 교호작용보다 중요하다고 가정합니다. 자세한 내용은 Wu and Hamada(2009) 및 "DOE 로 시작" 장의 " 효과 계층성 "(60 페이지) 에서 확인하십시오 .

선별 설계의 효율성은 효과 희소성 원리에도 영향을 받습니다 . 효과 희소성 원리는 반응에 대한 대부분의 변동이 상대적으로 적은 수의 효과로 설명된다고 주장합니다 . 자세한 내용은 "DOE 로 시작 " 장의 " 효과 희소성 "(61 페이지) 에서 확인하십시오 .

효과 희소성의 중요성을 제대로 이해하기 위해 7 개의 2 수준 요인이 있다고 가정해 보겠습니다 . 완전 요인 설계와 선별 설계를 비교해 보십시오 .

- 완전 요인 설계는 요인 수준의 모든 조합으로 구성됩니다 . 런 수는 각 요인에 대한 수준 수의 곱입니다 . 이 예에서 완전 요인 설계의 런 수는 $2^7 = 128$ 회입니다 .
- 반면에 선별 설계에는 완전 요인 설계의 런 중 일부만 필요합니다 . 7 개 요인의 주효과를 8 회 런 선별 설계로 연구할 수 있습니다 .

선별 설계 결과 분석

선별 설계는 많은 수의 요인 또는 교호작용을 검정하는 데 종종 사용됩니다. 오차 자유도가 있어 오차 추정값을 생성할 수 있는 경우 일반적인 회귀 기법을 사용하여 실험 결과를 분석할 수 있습니다("분석 > 모형 적합").

그러나 오차 자유도가 없는 경우가 있습니다. 이 경우 효과 희소성을 가정하면 선별 플랫폼("DOE > 전통적 설계 > 2수준 선별 > 2수준 선별 적합")에서 2수준 설계의 결과를 분석하는 방법을 제공합니다. 선별 플랫폼은 여러 개의 반응과 요인을 허용합니다. 그림과 통계량으로 유의한 효과를 자동으로 표시합니다. 자세한 내용은 "2수준 선별 적합 플랫폼" 장(309페이지)에서 확인하십시오. 현재 장에 나오는 예는 "부분 요인 설계에서 생성 규칙 수정"(298 페이지) 및 "Plackett-Burman 설계"(303페이지)에서 확인하십시오.

선별 설계의 예

- "부분 요인 설계와 주효과 선별 설계 비교"
- "표준 설계가 없는 경우의 주효과 선별 설계"

부분 요인 설계와 주효과 선별 설계 비교

이 예에서는 엔지니어가 전자 빔 용접기를 사용하여 두 부품을 결합하는 공정을 조사하려는 경우를 가정해 보겠습니다. 엔지니어는 두 부품을 용접 고정기에 꼭 맞게 끼웁니다. 빔 발전기에 전압이 가해지면 두 부품을 가열하는 전자 흐름이 생성되어 융합을 일으킵니다. 융합 영역의 이상적인 깊이는 0.17인치입니다. 엔지니어는 용접 공정을 연구하여 빔 발전기에서 원하는 융합 영역 깊이를 만들 수 있는 최적의 설정을 판별하려고 합니다.

이 연구를 위해 엔지니어는 다음과 같은 7 개 요인을 탐색하려고 합니다.

- Operator 는 용접기를 조작하는 기술자입니다. 일반적으로 두 명의 기술자가 기계를 작동합니다.
- Speed(rpm) 는 부품이 빔 아래에서 회전하는 속도입니다.
- Current(암페어) 는 빔 강도에 영향을 주는 전류입니다.
- Mode 는 사용되는 용접 방법입니다.
- Wall Size(mm) 는 부품 벽의 두께입니다.
- Geometry 는 조인트가 V 형 조인트인지 아니면 K 형 조인트인지를 나타냅니다.
- Material 은 용접되는 재료 유형입니다.

여기서 세 가지 요인 Speed, Current, Wall Size는 연속형입니다. 네 가지 요인 Operator, Mode, Geometry 및 Material은 범주형입니다. 각 범주형 요인의 수준은 2개입니다.

엔지니어는 각 공정이 실행된 후 부품을 반으로 자릅니다. 이로써 두 부품이 융합된 부분이 드러납니다. 이 융합 영역의 길이(인치)는 용입 깊이입니다. 용입 깊이가 이 연구의 반응입니다.

연구 목표는 다음과 같습니다.

- 용접 깊이에 영향을 주는 요인을 찾습니다.
- 요인 효과를 계량화합니다.
- 용접 깊이 0.17 인치 (공차 ±0.05 인치) 를 예측하는 특정 요인 설정을 찾습니다.

실험 예산으로 최대 12회 런이 가능합니다. 실험 상황에 대한 두 가지 설계를 구성하고 비교해 봅니다. 첫 번째 방법은 8회 런을 사용한 전통적 부분 요인 설계입니다. 두 번째 방법은 12회 런을 사용한 주효과 선별 설계입니다.

표준 선별 설계 구성

이 섹션에서는 현재 실험 상황에 대한 표준 선별 설계를 구성합니다.

반응 지정

1. **DOE > 전통적 설계 > 2 수준 선별 > 선별 설계**를 선택합니다.
2. "반응" 패널에서 "반응 이름" 아래의 Y 를 두 번 클릭하고 Depth 를 입력합니다.

   기본 목표는 최대화로 되어 있습니다. 이 실험에서는 허용 한계 (0.12, 0.22) 내에서 목표 깊이 0.17 을 얻을 수 있는 요인 설정을 찾으려고 합니다.
3. 기본 목표 최대화를 클릭한 후 **목표값 일치**로 변경합니다.
4. **하한** 아래를 클릭하고 0.12 를 입력합니다.
5. **상한** 아래를 클릭하고 0.22 를 입력합니다.
6. **중요도** 아래 영역은 비워 둡니다.

   반응이 하나뿐이므로 기본적으로 반응의 중요도가 1 로 지정됩니다.

완료된 반응 개요는 그림 9.2에 나와 있습니다. 이제 요인을 지정합니다.

요인 지정

수동 또는 자동으로 요인을 입력할 수 있습니다.

- 요인을 수동으로 입력하려면 "수동으로 요인 지정"(271 페이지)의 내용을 참조하십시오.
- 요인을 자동으로 입력하려면 Weld Factors.jmp 데이터 테이블을 사용하십시오.
  1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Weld Factors.jmp 를 엽니다.
  2. "선별 설계"의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다. "설계 선택"(272 페이지)으로 진행합니다.

수동으로 요인 지정

1. **N 개 요인 추가** 상자에 **3** 을 입력하고 **연속형**을 클릭합니다.
2. X1 을 두 번 클릭하고 Speed 를 입력합니다.
3. Tab 키를 사용하여 나머지 값 및 요인에서 이동합니다. 다음과 같이 변경합니다.

   a.Speed 값을 3 과 5 로 변경합니다.

   b. X2 를 Current 로 변경하고 값을 150 과 165 로 변경합니다.

c. X3 을 Wall Size 로 변경하고 값을 20 과 30 으로 변경합니다 .

4. **N 개 요인 추가** 상자에 **4** 를 입력하고 **범주형 > 2 수준**을 선택합니다 .
5. X4 를 두 번 클릭하고 Operator 를 입력합니다 .
6. Tab 키를 사용하여 나머지 값 및 요인에서 이동합니다 . 다음과 같이 변경합니다 .

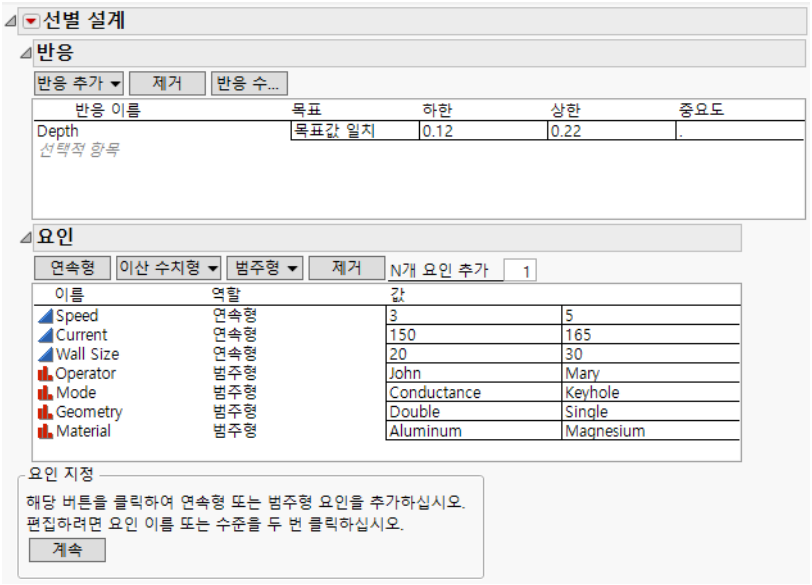
a.Operator 값을 John 과 Mary 로 변경합니다 .

b. X5 를 Mode 로 변경하고 값을 Conductance 와 Keyhole 로 변경합니다 .

c. X6 을 Geometry 로 변경하고 값을 Double 과 Single 로 변경합니다 .

d. X7 을 Material 로 변경하고 값을 Aluminum 과 Magnesium 으로 변경합니다 .

그림 9.2 용접 실험을 위한 반응 및 요인 개요



설계 선택

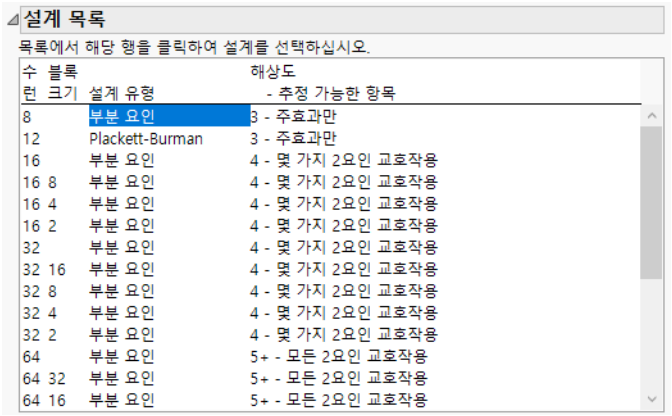
1. **계속**을 클릭합니다 .

   지정한 요인과 수준 조합을 표준 부분 요인 설계에서 수용할 수 있으므로 " 선별 유형 선택 " 패널이 나타납니다 . 목록에서 표준 설계를 선택하거나 주효과 설계를 구성할 수 있습니다 .

   참고 : 다음 단계에서 난수 시드값을 설정하면 이 예에 나오는 결과가 재현됩니다 . 설계를 직접 구성할 때는 이 단계가 필요하지 않습니다 .

2. (선택 사항) " 선별 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
3. 기본 선택인 **부분 요인 설계 목록에서 선택**을 적용하고 **계속**을 클릭합니다 .
4. 첫 번째 부분 요인 설계를 선택합니다 .

그림 9.3 연속형 요인 3 개와 범주형 요인 4 개에 대한 설계 목록

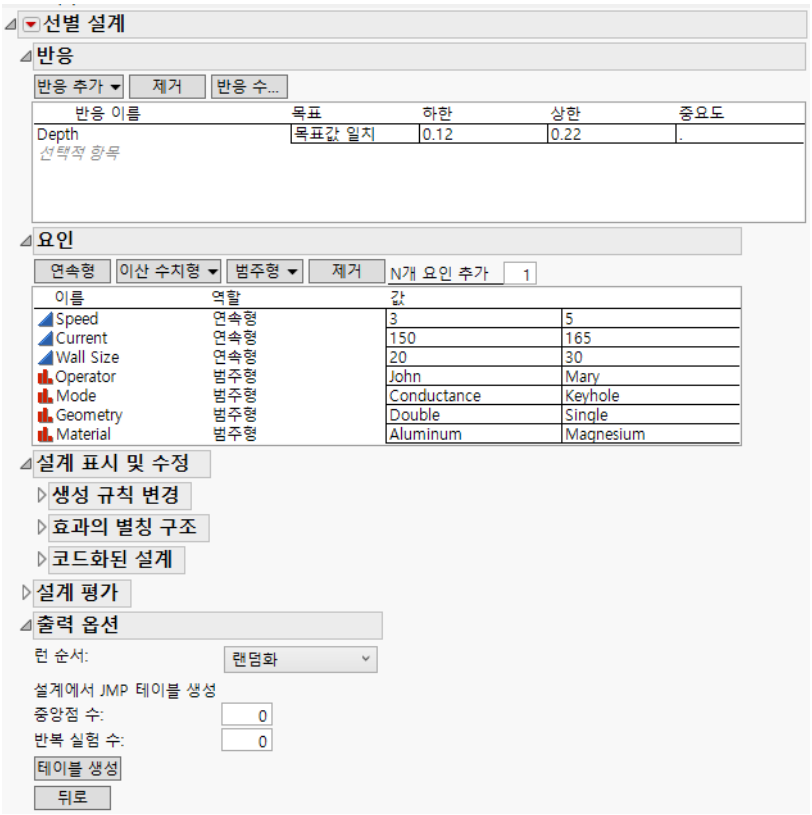


이렇게 하면 8 회 런의 해상도 3 부분 요인 설계가 지정됩니다 . 해상도에 대한 자세한 내용은 " 교락 측도를 나타내는 해상도 "(288 페이지) 에서 확인하십시오 .

5. **계속**을 클릭합니다 .

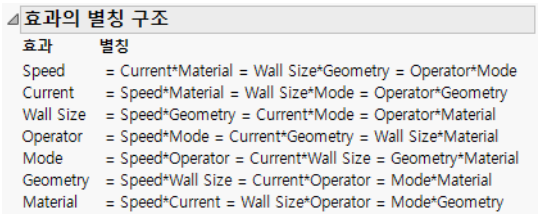
" 출력 옵션 " 개요에서 " 런 순서 " 가 **랜덤화**로 설정되어 있습니다 . 이는 설계 런이 랜덤 순서로 나타난다는 의미입니다 . 이 순서대로 실험 런을 구성해야 합니다 .

그림 9.4 완료된 선별 설계 창



6. " 설계 표시 및 수정 " 에서 " 효과의 별칭 구조 " 개요를 엽니다 .

그림 9.5 8 회 런 부분 요인 설계의 별칭 구조



해상도 3 설계를 선택했습니다 (그림 9.3). 해상도 3 설계에서는 일부 주효과가 이원 교호작용과 교락됩니다 . " 효과의 별칭 구조 " 개요에서는 이 해상도 3 설계에서 모든 주효과가 3 개의 이원 교호작용과 완전히 교락되어 있음을 나타냅니다 . 이원 교호작용의 활성이 의심된다면 이것은 잘못된 설계입니다 . 교락에 대한 설명은 "2 수준 정칙 부분 요인 "(286 페이지) 에서 확인하십시오 .

7. **테이블 생성**을 클릭합니다 .

그림 9.6 설계 데이터 테이블

	패턴	Speed	Current	Wall Size	Operator	Mode	Geometry	Material	Depth
1	+++++++	5	165	30	Mary	Keyhole	Single	Magnesium	•
2	−+−+−+−	3	165	20	Mary	Conductance	Single	Aluminum	•
3	+−+−−+−	5	150	30	John	Conductance	Single	Aluminum	•
4	−−−−+++	3	150	20	John	Keyhole	Single	Magnesium	•
5	−++−+−−	3	165	30	John	Keyhole	Double	Aluminum	•
6	++−−−−+	5	165	20	John	Conductance	Double	Magnesium	•
7	+−−++−−	5	150	20	Mary	Keyhole	Double	Aluminum	•
8	−−++−−+	3	150	30	Mary	Conductance	Double	Magnesium	•

다음을 확인할 수 있습니다 .

- 이 테이블에는 앞에서 지정한 반응 , 요인 및 수준 이름이 사용됩니다 .
- **패턴** 열에는 설계 런에 대한 상한 및 하한 설정 할당이 표시됩니다 .
- 이 부분 요인 설계는 해상도 3 설계입니다 . 이 설계를 사용하여 8 회 런으로 7 개 요인의 주효과를 연구할 수 있습니다 .

주효과 선별 설계 생성

주효과 선별 설계는 직교 또는 근사 직교 설계입니다 . 이 섹션에서는 7 개 요인에 대한 주효과 선별 설계를 구성합니다 .

1. " 선별 설계 " 창을 엽니다 . 창을 닫았으면 설계 데이터 테이블에서 **DOE 대화상자** 스크립트를 실행합니다 .
2. **뒤로**를 클릭합니다 .
3. **계속**을 클릭합니다 .

참고 : 다음 두 단계에서 난수 시드값과 시작 수를 설정하면 이 예에 나오는 것과 정확히 같은 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

4. (선택 사항) " 선별 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 12345 를 입력하고 **확인**을 클릭합니다 .
5. (선택 사항) " 선별 설계 " 의 빨간색 삼각형을 클릭하고 **시작 수**를 선택한 후 50 을 입력하고 **확인**을 클릭합니다 .
6. " 선별 유형 선택 " 패널에서 **주효과 선별 설계 생성** 옵션을 선택합니다 .
7. **계속**을 클릭합니다 .

   " 런 수 " 아래에 런 수가 12 로 설정된 **기본값** 옵션이 선택되어 있습니다 . 이 설정을 그대로 유지합니다 .
8. **설계 생성**을 클릭합니다 .

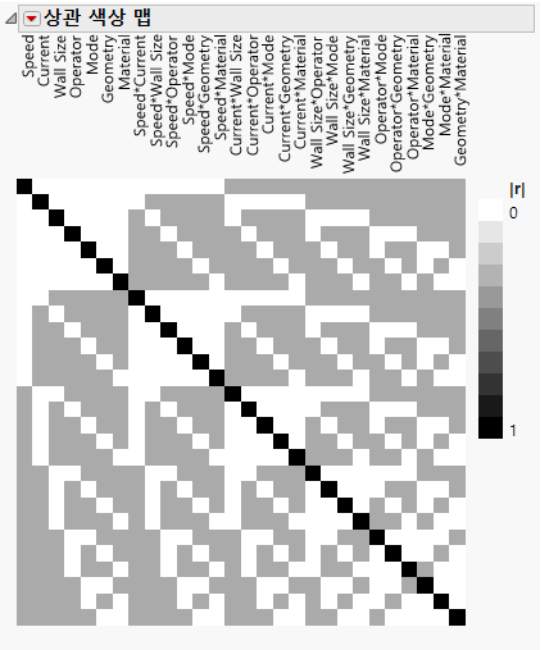
그림 9.7 주효과 선별 설계

설계

런	Speed	Current	Wall Size	Operator	Mode	Geometry	Material
1	3	165	20	John	Conductance	Single	Aluminum
2	5	150	30	Mary	Conductance	Double	Aluminum
3	3	150	30	John	Keyhole	Double	Aluminum
4	5	165	20	Mary	Keyhole	Double	Magnesium
5	3	165	20	Mary	Conductance	Double	Aluminum
6	5	150	20	John	Conductance	Double	Magnesium
7	3	150	30	Mary	Conductance	Single	Magnesium
8	5	150	20	John	Keyhole	Single	Aluminum
9	3	165	30	John	Keyhole	Double	Magnesium
10	5	165	30	John	Conductance	Single	Magnesium
11	3	150	20	Mary	Keyhole	Single	Magnesium
12	5	165	30	Mary	Keyhole	Single	Aluminum

9. " 설계 평가 " 개요를 열고 " 상관 색상 맵 " 개요를 엽니다 .

그림 9.8 12 회 런 주효과 선별 설계의 상관 색상 맵



" 상관 색상 맵 " 에서는 주효과 간에 상관관계가 없음을 보여 줍니다 . 이는 색상 맵 왼쪽 위의 흰색 비대각 셀로 표시됩니다 . 각 주효과는 일부 이원 교호작용과 부분적으로 별칭 관계가 있습니다 (회색 셀로 표시). 회색 셀 중 하나를 커서로 가리키면 절대 상관이 0.333 으로 표시됩니다 .

이 경우 12 회 런 주효과 선별 설계는 Plackett-Burman 설계이며 " 설계 목록 " 에서 선택할 수 있습니다 . 그러나 많은 설계 상황에서는 주효과 설계에서 발생하는 부분 별칭 구조가 실험 상황에 맞게 조정한 부분 요인 설계에서 발생하는 완전 교락보다 낫습니다 .

다음 섹션에서는 표준 설계가 없는 상황에 대한 예를 보여 줍니다 . 이 경우 JMP 는 주효과 선별 설계를 구성합니다 .

표준 설계가 없는 경우의 주효과 선별 설계

주효과 선별 설계는 주효과에 대한 직교 또는 근사 직교 설계입니다 . 표준 설계가 없는 상황에서 표준 설계 대신 이 설계를 사용할 수 있습니다 . 주효과 선별 설계는 교호작용을 무시할 수 있는 경우 주효과를 추정하는 데 탁월합니다 .

이 실험 상황에서는 표준 설계가 없습니다 . 각각 수준이 4 개 , 6 개인 범주형 요인 2 개와 연속형 요인이 11 개인 총 13 개 요인을 연구하기 위한 설계가 필요합니다 .

1. **DOE > 전통적 설계 > 2 수준 선별 > 선별 설계**를 선택합니다 .

   " 반응 " 패널에 Y 라는 기본 반응이 하나 있습니다 . 기본 반응을 그대로 유지합니다 .

2. " 요인 " 패널에서 **범주형**을 클릭하고 **4 수준**을 선택합니다 .

   L1 ~ L4 수준이 지정된 X1 변수가 추가됩니다 .

3. **범주형**을 클릭하고 **6 수준**을 선택합니다 .

   L1 ~ L6 수준이 지정된 X2 변수가 추가됩니다 .

4. **N 개 요인 추가** 옆에 11 을 입력합니다 .

5. **연속형**을 클릭합니다 .

   각각 두 개의 수준 (-1 과 1) 이 지정된 11 개 요인 X3 ~ X13 이 추가됩니다 .

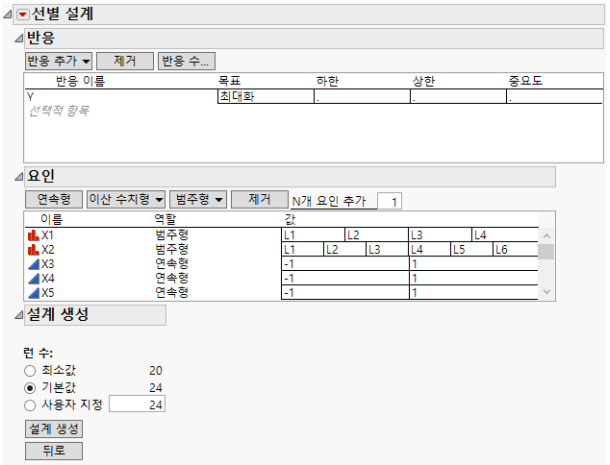
6. **계속**을 클릭합니다 .

   " 설계 생성 " 패널이 나타납니다 .

   이 상황에서 사용할 수 있는 표준 설계가 없으므로 " 설계 목록 " 에서 설계를 선택하는 옵션이 없습니다 .

   기본 런 수 , 24 를 그대로 유지합니다 .

그림 9.9 설계 생성 패널이 있는 13 개 요인 설계를 위한 선별 설계 창



7. **설계 생성**을 클릭합니다 .

   " 설계 " 및 " 설계 평가 " 개요가 나타납니다 .

8. " 설계 " 개요를 열어 랜덤화된 설계를 확인합니다 .

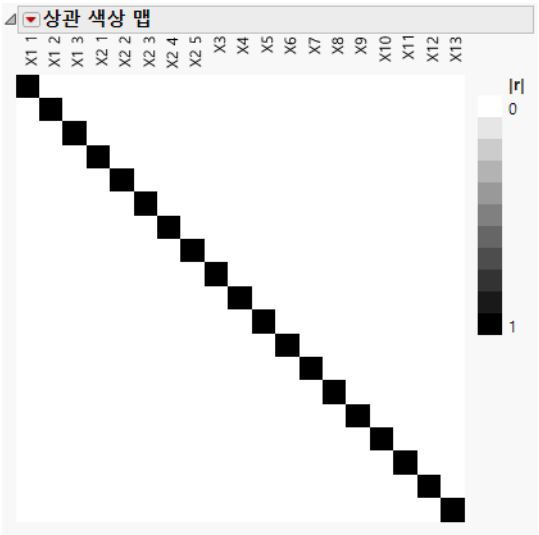
   참고 : 설계를 생성하는 알고리즘은 랜덤 시작 설계를 사용합니다 . 이 설계를 재현하려면 보고서 제목 옆의 빨간색 삼각형 메뉴에서 **스크립트를 스크립트 창에 저장**을 선택하여 난수 시드값과 함께 스크립트를 저장하십시오 .

   이제 " 상관 색상 맵 " 을 검토하여 이 특정 설계가 직교인지 확인합니다 .

9. **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

   색상 맵 (그림 9.10) 의 주대각 위에 검은색 항목 (JMP 기본 색상 사용) 이 표시되어 상관계수가 1 인 상관관계를 나타냅니다 . 각 주대각 셀은 항 자체의 상관관계 (1) 에 해당하기 때문입니다 . 비대각 상관은 모두 흰색이며 개별 항 간의 상관관계가 0 임을 나타냅니다 . 관련 항 및 해당 항의 절대 상관을 표시하려면 셀을 커서로 가리킵니다 .

그림 9.10 상관 색상 맵



10. **테이블 생성**을 클릭하여 설계 테이블을 구성합니다.

    이 테이블에는 실험 런이 랜덤 순서로 포함되어 있습니다. 랜덤화된 이 순서로 실험을 수행하고 결과를 Y 열에 삽입합니다. 데이터 테이블에서 **모형** 스크립트를 실행하여 결과를 분석합니다.

선별 설계 창

설계 단계를 진행하는 동안 "선별 설계" 창이 업데이트됩니다. 자세한 내용은 "DOE 워크플로우: 기술, 지정, 설계"(56 페이지)에서 확인하십시오. 개요는 창을 업데이트하는 버튼을 클릭할 때 구분되어 나타납니다. 이러한 개요는 아래 그림의 흐름을 따릅니다.

그림 9.11 표준 설계가 있을 때 선별 설계 흐름

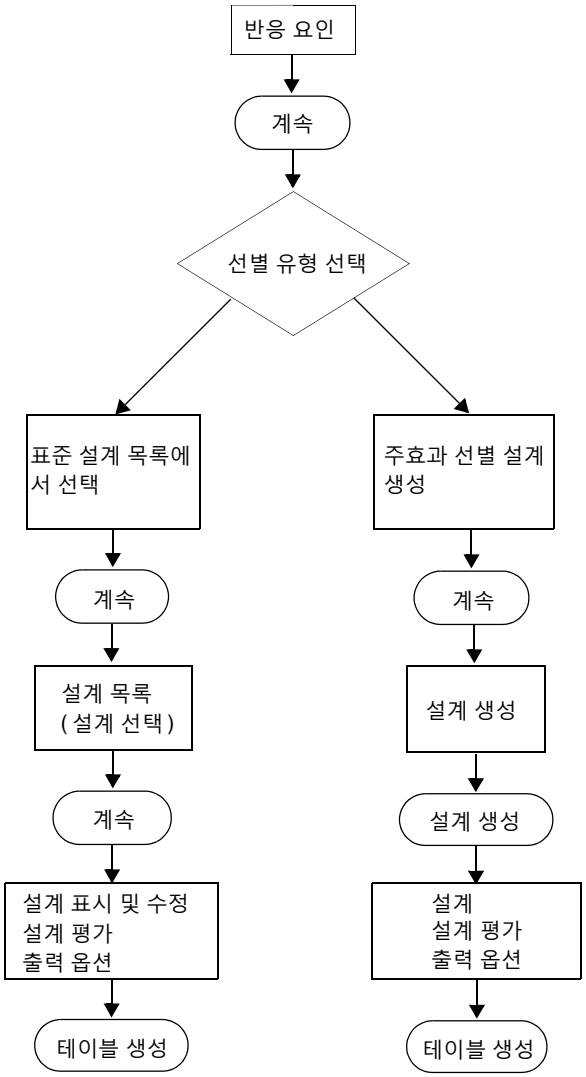
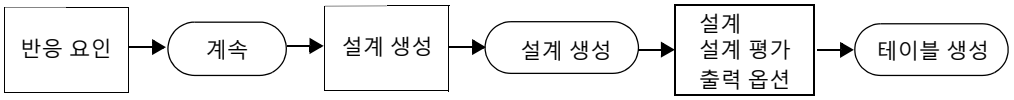


그림 9.12 표준 설계가 없을 때 선별 설계 흐름

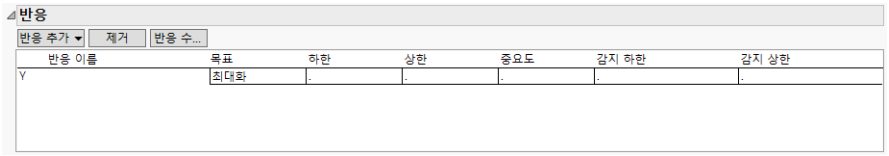


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다 . 이 옵션을 사용하면 반응 이름 , 목표 , 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다 .

그림 9.13 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다 . " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다 . " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다 .

JMP PRO 함수 (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다 . 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다 .

제거 선택한 반응을 제거합니다 .

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다 . 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다 . 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다 . 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다 .

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응 열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .

– "없음" 목표는 최적화에 관한 목표가 없음을 나타냅니다. 만족도 함수가 구성되지 않습니다.

참고: 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다. 데이터 테이블에서 반응 열의 "열 정보" 창을 열고 (**열** > **열 정보**) 원하는 목표값을 입력하십시오.

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다. 이러한 값은 반응에 대한 가중치로 처리됩니다. 반응이 하나뿐이면 기본적으로 1 로 설정되므로 중요도를 지정할 필요가 없습니다.

JMP PRO **감지 한계** 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다. 감지 한계는 열 특성으로 저장됩니다. 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오.

반응 개요 편집

"반응" 개요에서 다음 사항에 유의하십시오.

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다.
- 목표를 클릭하여 변경할 수 있습니다.
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다.
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다.

반응 한계 열 특성

반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

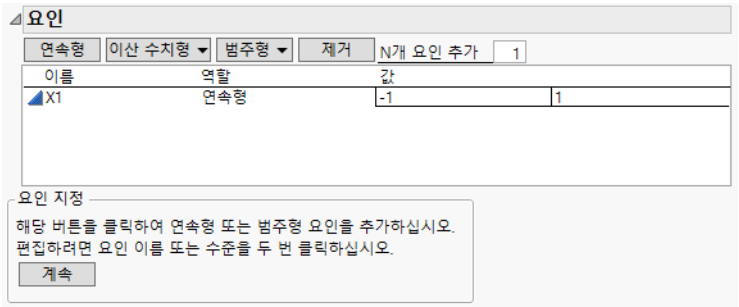
하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

요인

"요인" 개요에서 요인을 추가합니다.

팁: "요인" 개요를 완료하면 빨간색 삼각형 메뉴에서 **요인 저장**을 선택하는 것이 좋습니다. 이 옵션을 사용하면 요인 이름, 역할, 변경 및 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다.

그림 9.14 요인 개요



연속형 연속형 요인을 추가합니다 . 결과 데이터 테이블의 데이터 유형은 숫자입니다 . 연속형 요인은 개념적으로 공정 및 측정 시스템의 제한을 고려하여 사용자가 제공하는 하한과 상한 사이의 어떤 값이든 설정할 수 있는 요인입니다 .

이산 수치형 이산 수치형 요인을 추가합니다 . 이산 수치형 요인은 이산형 숫자 값만 가질 수 있습니다 . 이러한 값에는 묵시적 순서가 있습니다 . 결과 데이터 테이블의 데이터 유형은 숫자입니다 .

선별 설계는 이산 수치형 요인의 모든 수준을 포함하며 수준의 균형을 맞추려고 시도합니다 . " 모형 적합 " 에서는 이산 수치형 요인을 연속형 예측 변수로 처리합니다 .

수준 수가 k 개 ($k > 2$) 인 이산 수치형 요인의 기본값은 정수 1, 2, ..., k 입니다 . $k = 2$ 수준이 있는 이산 수치형 요인의 기본값은 -1 과 1 입니다 . 기본값을 실험에 사용할 설정으로 바꿉니다 .

범주형 범주형 요인을 추가합니다 . 클릭하여 수준 수를 선택하거나 지정합니다 . 결과 데이터 테이블의 데이터 유형은 문자입니다 . 수준의 값 순서는 왼쪽에서 오른쪽으로 입력되는 값의 순서입니다 . 이 순서는 설계 데이터 테이블이 생성된 후 " 값 순서 " 열 특성에 저장됩니다 .

범주형 요인의 기본값은 L1, L2,..., Lk 이며 여기서 k 는 지정한 수준 수입니다 . 기본값을 실험과 관련된 수준 이름으로 바꿉니다 .

제거 선택한 요인을 제거합니다 .

N 개 요인 추가 여러 요인을 추가합니다 . 추가할 요인 수를 입력하고 **요인 추가**를 클릭한 후 요인 유형을 선택합니다 . 유형이 다른 여러 요인을 추가하려면 **N 개 요인 추가**를 반복합니다 .

요인 개요

" 요인 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할입니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인 유형이 적절하게 모델링되도록 합니다 . 요인의 역할에 따라 데이터 테이블에 저장되는 다른 요인 특성이 결정됩니다 . 자세한 내용은 " 요인 열 특성 "(284 페이지) 에서 확인하십시오 .

값 요인에 대한 실험 설정입니다 .

요인 개요 편집

" 요인 " 개요에서 다음 사항에 유의하십시오 .

- 요인 이름을 편집하려면 해당 요인 이름을 두 번 클릭합니다 .
- 값을 편집하려면 " 값 " 열의 값을 클릭합니다 .

요인 열 특성

각 요인에 대해 " 선별 설계 " 창에서 " 테이블 생성 " 을 선택하여 설계를 생성하면 다양한 열 특성이 설계 테이블에 저장됩니다 . 또한 " 요인 저장 " 옵션을 선택할 때 생성되는 데이터 테이블에도 이러한 특성이 자동으로 저장됩니다 . 이러한 열 특성 및 관련 예에 대한 자세한 내용은 부록 A, " 열 특성 " 에서 확인하십시오 .

코딩 " 역할 " 이 " 연속형 " 또는 " 이산 수치형 " 인 경우 요인의 " 코딩 " 열 특성이 저장됩니다 . 이 특성은 하한값과 상한값이 각각 –1 과 +1 에 해당하도록 요인 값을 변환합니다 . 자세한 내용은 " 열 특성 " 부록의 " 코딩 "(739 페이지) 에서 확인하십시오 .

값 순서 " 역할 " 이 " 범주형 " 이거나 블록 변수가 구성된 경우 요인의 " 값 순서 " 열 특성이 저장됩니다 . 이 특성은 요인 수준이 나타나는 순서를 결정합니다 . 자세한 내용은 " 열 특성 " 부록의 " 값 순서 "(755 페이지) 에서 확인하십시오 .

설계 역할 각 요인에 " 설계 역할 " 열 특성이 지정됩니다 . 요인을 정의할 때 지정하는 역할에 따라 " 설계 역할 " 열 특성의 값이 결정됩니다 . 블록이 있는 설계를 선택하면 해당 블록 요인에 " 블록화 " 값이 할당됩니다 . " 설계 역할 " 특성은 실험 데이터를 모델링할 때 요인이 어떤 용도로 사용되는지를 반영합니다 . " 설계 역할 " 값은 설계 확대 플랫폼에서 사용됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 설계 역할 "(737 페이지) 에서 확인하십시오 .

요인 변경 각 요인에 값이 " 쉬움 " 인 " 요인 변경 " 열 특성이 할당됩니다 . " 요인 변경 " 특성은 실험 데이터를 모델링할 때 요인이 사용되는 방식을 반영합니다 . " 요인 변경 " 값은 설계 확대 및 설계 평가 플랫폼에서 사용됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 요인 변경 "(752 페이지) 에서 확인하십시오 .

블록당 런 수 블록 요인에 대해 각 블록에서 허용되는 최대 런 수를 나타냅니다 . " 요인 " 개요에 " 블록화 " 요인이 지정되면 해당 요인에 대해 " 블록당 런 수 " 열 특성이 저장됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 블록당 런 수 "(761 페이지) 에서 확인하십시오 .

선별 유형 선택

반응과 요인을 입력한 후 **계속**을 클릭하면 다음 결과 중 하나가 발생합니다 .

- 표준 설계에서 요인과 수준을 수용할 수 있는 경우 " 선별 유형 선택 " 패널에 두 가지 옵션이 나타납니다 . 자세한 내용은 " 선별 유형 선택 옵션 "(285 페이지) 에서 확인하십시오 .
- 요인과 수준에 맞는 표준 설계가 나열되지 않으면 " 선별 유형 선택 " 패널이 나타나지 않습니다 . 주효과 선별 설계를 구성하기 위한 " 설계 생성 " 개요가 열립니다 . 자세한 내용은 " 설계 생성 "(292 페이지) 에서 확인하십시오 .

선별 유형 선택 옵션

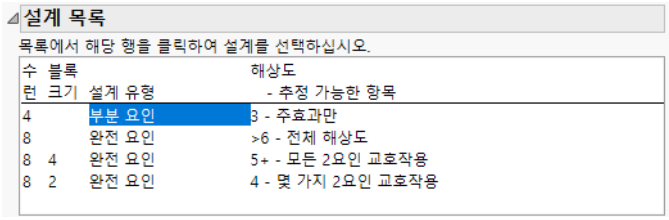
부분 요인 설계 목록에서 선택 설계 목록에서 선택할 수 있습니다. 이 옵션이 기본값입니다. 자세한 내용은 " 부분 요인 설계 목록에서 선택 "(285 페이지) 에서 확인하십시오.

주효과 선별 설계 생성 주효과 선별 설계의 런 수를 지정할 수 있는 " 설계 생성 " 개요를 엽니다. 주효과 선별 설계에 대한 자세한 내용은 "주효과 선별 설계"(291 페이지) 에서 확인하십시오.

부분 요인 설계 목록에서 선택

선택할 수 있는 선별 설계 목록에는 실험 런을 같은 크기 (2 의 거듭제곱) 의 블록으로 그룹화하는 설계가 포함됩니다. 사용할 선별 설계 유형을 선택하고 **계속**을 클릭합니다.

그림 9.15 부분 요인 설계 유형 선택



" 설계 목록 " 에는 다음과 같은 열이 포함되어 있습니다.

런 수 설계의 총 런 수입니다.

블록 크기 블록 내 런 수입니다. 블록 수는 런 수를 블록 크기로 나눈 값입니다.

설계 유형 설계 유형에 대한 설명입니다. 자세한 내용은 " 설계 유형 "(285 페이지) 에서 확인하십시오.

해상도 설계 해상도 및 별칭 구조 유형에 대한 간략한 설명을 제공합니다. 자세한 내용은 "교락 측도를 나타내는 해상도"(288 페이지) 에서 확인하십시오.

설계 유형

" 설계 목록 " 은 다음과 같은 유형의 설계를 제공합니다.

- "2 수준 완전 요인 "(286 페이지)
- "2 수준 정칙 부분 요인 "(286 페이지)
- "Plackett-Burman 설계 "(286 페이지)
- " 혼합수준 설계 "(286 페이지)
- "Cotter 설계 "(287 페이지)

2 수준 완전 요인

완전 요인 설계는 요인 수준의 모든 조합에 대한 런을 포함합니다. 표본 크기는 요인 수준의 곱입니다. 2 수준 설계의 경우 표본 크기는 2^k이며 여기서 k는 요인 수입니다.

완전 요인 설계는 모든 효과에 대해 직교입니다. 따라서 효과 추정값은 상관관계가 없습니다. 또한 분석에서 효과를 제거해도 다른 추정값이 변경되지 않습니다. 오차 분산의 추정값과 자유도가 다르기 때문에 p 값은 약간 변경됩니다.

완전 요인 설계를 사용하면 최대 요인 수까지 모든 차수의 교호작용을 추정할 수 있습니다. 그러나 대부분의 경험적 모델링은 요인과 반응 간의 실제 함수 관계에 대한 1 차 또는 2 차 근사만 포함합니다. 이러한 관점에서 완전 요인 설계는 실험 런을 비효율적으로 사용하는 것입니다.

2 수준 정칙 부분 요인

정칙 부분 요인 설계의 표본 크기도 2 의 거듭제곱입니다. 2 수준 설계의 경우 k가 요인 수이면 정칙 부분 요인 설계의 런 수는 2^{k-p}이며 여기서 $p<k$입니다. 2^{k-p} 부분 요인 설계는 요인 수가 k개인 완전 요인 설계의 $1/2^{-p}$ 배입니다. 완전 요인 설계와 마찬가지로 정칙 부분 요인 설계도 직교입니다.

k개 요인에 대한 완전 요인 설계는 최대 k차까지 모든 교호작용 효과의 추정값을 제공합니다. 그러나 실험 런은 대개 비용이 많이 들기 때문에 작은 규모의 설계가 선호됩니다. 규모가 작은 설계에서는 일부 고차 효과가 다른 효과와 **교락**되므로 효과를 서로 구별할 수 없습니다. 교락된 효과의 선형 조합을 추정할 수는 있지만 변동의 원인을 하나 이상의 특정 효과로 지정할 수는 없습니다.

실제로 부분 요인 설계는 다른 교호작용 효과와 교락되는 교호작용 효과를 미리 결정하여 구성됩니다. 일반적으로 실험자는 3 개 이상의 요인을 포함하는 교호작용에는 관심이 없습니다. 3 차 이상의 교호작용 효과는 무시할 수 있는 것으로 가정하는 경우가 많습니다.

Plackett-Burman 설계

Plackett-Burman 설계는 선별을 위한 정칙 부분 요인 설계 대신 사용할 수 있습니다. Plackett-Burman 설계의 런 수는 2 의 거듭제곱이 아니라 4 의 배수입니다. 런 크기가 16 ~ 32 인 2 수준 부분 요인 설계는 없습니다. 그러나 20 회, 24 회 및 28 회 런 Plackett-Burman 설계는 있습니다.

Plackett-Burman 설계에서 주효과는 서로 직교하고 2 요인 교호작용은 주효과와 부분적으로 교락됩니다. 이에 반해 해상도 3 정칙 부분 요인 설계에서는 일부 2 요인 교호작용을 주효과와 구별할 수 없습니다. Plackett-Burman 설계는 여러 요인 중에서 큰 주효과를 감지하는 데 관심이 있고 교호작용을 무시할 수 있을 때 유용합니다.

혼합수준 설계

수준이 3 개 이상인 범주형 또는 이산 수치형 요인을 포함하는 대부분의 설계에는 사용 가능한 표준 설계가 없습니다. 이 경우 선별 플랫폼은 **주효과 선별 설계**를 생성합니다. 이러한 설계는 주효과에 대한 직교 또는 근사 직교 설계입니다.

표준 혼합수준 설계가 있는 경우 가능한 설계가 "설계 목록"에 제공됩니다. "설계 목록"은 최대 13개 요인으로 순수 3수준 요인에 대한 부분 요인 설계를 제공합니다. 2수준 설계와 3수준 설계가 혼합된 경우에는 표 9.1에 나열된 완전 요인 및 직교 배열 설계가 "설계 목록"에 포함됩니다.

요인 수가 표에 나열된 설계의 수를 초과하지 않으면 해당 열의 적절한 부분집합을 사용하여 설계를 조정할 수 있습니다.

표 9.1 혼합수준 설계 테이블

	요인 수	
설계	2 수준	3 수준
L18 John 및 L18 Taguchi	1	7
L18 Chakravarty	3	6
L18 Hunter	8	4
L36 Taguchi	11	12

Cotter 설계

참고 : 기본적으로 Cotter 설계는 " 설계 목록 " 에 포함되지 않습니다 . Cotter 설계를 포함하려면 " 선별 설계 " 의 빨간색 삼각형 메뉴에서 **Cotter 설계 제한**을 선택 취소합니다 . Cotter 설계를 항상 표시하려면 **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택한 후 **Cotter 설계 제한**을 선택 취소하십시오 .

Cotter 설계는 일부 요인이 상호 작용할 수도 있는 많은 요인을 매우 적은 수의 런으로 검정해야 할 때 유용합니다. Cotter 설계는 효과 희소성 원리에 의존합니다. 이 설계는 효과를 합할 때 합성분 중 하나에 활성 효과가 있을 경우 해당 효과를 효과 합에 표시한다고 가정합니다. 단점은 부호가 혼합된 여러 활성 효과의 합이 0에 가까워져 효과를 나타내지 못할 수 있다는 것입니다. 이 False Negative 위험 때문에 많은 통계학자들이 이 설계의 사용을 권장하지 않습니다.

*k*개 요인의 경우 Cotter 설계의 런 수는 $2k + 2$ 회입니다. 설계 구조는 "한 번에 하나의 요인 변경 (OFAT)" 방법과 비슷합니다.

Cotter 설계는 다음과 같이 구성됩니다 .

- 모든 요인이 높은 수준으로 설정된 런이 정의됩니다 .
- 다음 *k* 런에 대해 각각 차례로 한 요인은 낮은 수준으로 설정되고 나머지 요인은 높은 수준으로 설정됩니다 .
- 다음 런은 모든 요인이 낮은 수준으로 설정됩니다 .
- 다음 *k* 런에 대해 각각 차례로 한 요인은 높은 수준으로 설정되고 나머지 요인은 낮은 수준으로 설정됩니다 .

- 런이 랜덤화됩니다 .

Cotter 설계를 구성할 때 설계 데이터 테이블에는 회귀변수로 사용할 열 집합이 포함됩니다. 열 이름은 <요인 이름> Odd 및 <요인 이름> Even 형식을 사용합니다. 이 설계는 제공된 요인을 포함하는 홀수 차수 교호작용 항과 짝수 차수 교호작용 항을 각각 합하여 구성됩니다.

예를 들어 A, B, C라는 세 가지 요인이 있다고 가정해 보겠습니다. 표 9.2에서는 회귀변수 열의 값이 계산되는 방법을 보여 줍니다.

표 9.2 Cotter 설계 테이블

홀수 회귀변수 열과 짝수 회귀변수 열의 합산 효과	
AOdd = A + ABC	AEven = AB + AC
BOdd = B + ABC	BEven = AB + BC
COdd = C + ABC	CEven = BC + AC

홀수(Odd) 및 짝수(Even) 열은 직교 변환을 정의합니다. 따라서 홀수 및 짝수 열의 모수 검정은 원래 효과에 대한 조합을 검정하는 것과 동등합니다.

교락 측도를 나타내는 해상도

설계 **해상도**는 설계의 교락 정도에 대한 측도입니다. 선별 설계에서는 런 수와 설계 해상도 사이의 균형을 맞추어야 합니다.

실험은 **해상도 번호**에 따라 다음과 같은 그룹으로 분류됩니다 .

- 해상도 3 은 일부 주효과가 하나 이상의 2 요인 교호작용과 교락됨을 의미합니다 . 주효과가 유의하려면 이러한 교호작용을 무시할 수 있는 것으로 가정해야 합니다 .
- 해상도 4 는 주효과가 다른 주효과 또는 2 요인 교호작용과 교락되지 않음을 의미합니다 . 그러나 일부 2 요인 교호작용은 다른 2 요인 교호작용과 교락됩니다 .
- 해상도 5 는 주효과와 다른 주효과 , 주효과와 2 요인 교호작용 , 2 요인 교호작용 쌍 사이에 각각 교락이 없음을 의미합니다 . 일부 2 요인 교호작용은 3 요인 교호작용과 교락됩니다 .
- 해상도 5+ 는 설계의 해상도가 5 보다 크지만 완전 요인 설계는 아님을 의미합니다 .
- 해상도 6 은 모든 차수의 효과 사이에 교락이 없음을 의미합니다 . 이 설계는 완전 요인 설계입니다 .

최소 일탈 설계는 주어진 해상도에 대한 교락 수를 최소화합니다. 주어진 해상도의 최소 일탈 설계는 관계 정의에서 최소 길이의 **단어** 수를 최소화합니다. 단어에 대한 설명은 "생성 규칙 변경"(289 페이지)에서 확인하십시오. 최소 일탈 설계에 대한 자세한 내용은 Fries and Hunter(1984)에서 확인하십시오.

설계 표시 및 수정

"설계 목록"에서 모든 연속형 요인 또는 2수준 범주형 요인과 블록 요인(가능한 경우)이 포함된 부분 요인 설계를 선택하는 경우 "계속"을 클릭하면 "설계 표시 및 수정" 개요가 열립니다. 이 개요의 보고서를 사용하여 설계를 수정합니다. 생성 규칙을 변경하여 설계를 구성하는 예는 "부분 요인 설계에서 생성 규칙 수정"(298 페이지)에서 확인하십시오.

참고 : 교호작용이 주효과와 동일하지 않으므로 Plackett-Burman 설계 또는 Cotter 설계의 경우 " 생성 규칙 변경 " 및 " 효과의 별칭 구조 " 개요가 나타나지 않습니다 .

생성 규칙 변경 설계의 관계 정의를 지정합니다 . 관계 정의는 JMP 에서 제공할 완전 요인 설계의 부분을 결정합니다 . 자세한 내용은 " 생성 규칙 변경 "(289 페이지) 에서 확인하십시오 .

효과의 별칭 구조 부분 요인 설계의 교락 패턴을 표시합니다 . 패널 아래쪽의 빨간색 화살표를 클릭하여 지정된 차수에 대한 교호작용을 확인합니다. 교호작용 및 별칭 구조가 데이터 테이블에 제공됩니다 .

코드화된 설계 각 런에서 요인의 상한값과 하한값 패턴을 표시합니다 .

참고 : Cotter 설계의 경우 " 생성 규칙 변경 " 및 " 효과의 별칭 구조 " 개요가 적용되지 않으므로 표시되지 않습니다 .

생성 규칙 변경

생성 규칙은 특정 부분 요인 설계를 구성하는 데 사용되는 관계를 정의합니다. 선별 플랫폼의 기본 생성자는 최소 일탈 설계를 생성합니다. 최소 일탈 설계는 낮은 차수 효과의 별칭 구조를 최소화합니다. 실험 상황에 따라 표준 부분과 다른 코딩 또는 별칭 구조를 제공하는 설계 부분을 정의해야 할 수 있습니다. "설계 표시 및 수정" 개요에서 생성 규칙을 변경하여 이 작업을 수행할 수 있습니다. 관계 정의 및 생성 규칙에 대한 자세한 내용은 Montgomery(2009)에서 확인하십시오.

팁 : 특정 효과 추정을 위한 설계를 생성하려면 사용자 설계 도구를 사용하는 것이 좋습니다 . 사용자 설계에 대한 자세한 내용은 " 사용자 설계 " 장 (63 페이지) 에서 확인하십시오 .

설계의 관계 정의는 생성 규칙의 단어에 의해 결정됩니다. 단어는 요인의 곱으로 나타내지만 설계 행렬에서 해당 열에 대한 항목의 요소별 곱으로 해석됩니다. 관계 정의는 곱이 1의 열(항등)인 단어로 구성됩니다.

그림 9.16에서는 2^{5-2} 부분 요인 설계(5개 요인과 8회 런)에 대한 기본 생성 규칙을 보여 줍니다.

그림 9.16 표준 2^{5-2} 설계의 생성 규칙

" 생성 규칙 변경 " 패널의 각 열에서 맨 위에 나열된 요인과 상자가 선택된 열의 요인이 관계 정의의 단어를 구성합니다 . 예를 들어 첫 번째 열은 Temperature = Feed Rate*Catalyst*Stir Rate 가 관계 정의의 단어임을 나타냅니다 .

- +/- 상자가 선택된 경우 생성 규칙과 관련된 부호가 양수이고 해당 단어가 항등과 같습니다 .
- +/- 상자가 선택되지 않은 경우 생성 규칙과 관련된 부호가 음수이고 해당 단어가 마이너스 항등과 같습니다 .

자세한 내용은 "2^{5-3} 설계에서 관계 정의 얻기"(290 페이지) 에서 확인하십시오 .

완전 요인 설계의 **주 부분**은 모든 관계 정의를 항등과 같게 설정하여 얻은 부분 요인 설계입니다 . 기본적으로 JMP 에서 제공하는 요인 설계는 주 부분입니다 . +/- 상자는 기본적으로 모든 생성 규칙에 대해 선택되므로 관계 정의의 각 단어는 항등과 같습니다 .

생성 규칙은 설계에 대한 코딩 및 효과의 별칭 구조를 결정합니다 . 경우에 따라 표준 부분과 다른 코딩 또는 별칭 구조를 생성하는 부분을 사용할 수도 있습니다 .

- 생성 규칙을 변경하려면 해당하는 상자를 선택합니다 .
- 선택 항목에 따라 " 효과의 별칭 구조 " 결과 및 " 코드화된 설계 " 가 어떻게 변하는지 확인하려면 **적용**을 클릭합니다 .

예는 " 다른 부분을 얻기 위해 생성 규칙 변경 "(301 페이지) 에서 확인하십시오 .

2^{5-3} 설계에서 관계 정의 얻기

그림 9.16 에서는 다음과 같은 두 개의 체크박스 열을 보여 줍니다 .

- 첫 번째 열은 Temperature = Feed Rate*Catalyst*Stir Rate 라는 단어를 나타냅니다 .
- 두 번째 열은 Concentration = Catalyst*Stir Rate 라는 단어를 나타냅니다 .

+1 값으로 구성된 열을 나타내도록 I 를 정의합니다 . 모든 요인 수준이 -1 또는 +1 이므로 첫 번째 열의 단어는 Temperature*Feed Rate*Catalyst*Stir Rate = I 와 동등합니다 . 두 번째 열의 단어는 Concentration*Catalyst*Stir Rate = I 와 동등합니다 . 이 둘을 함께 사용하여 다음과 같이 2^{5-2} 설계에 대한 관계 정의가 제공됩니다 .

I = Temperature*Feed Rate*Catalyst*Stir Rate = Concentration*Catalyst*Stir Rate

2^{5-3} 설계에서 효과의 별칭 구조 관계 얻기

"효과의 별칭 구조" 개요의 별칭 구조는 관계 정의 및 요인 수준이 +1 과 -1 이라는 사실에 의해 결정됩니다 . 첫 번째 생성 규칙은 Temperature = Feed Rate*Catalyst*Stir Rate 이고 , 두 번째 생성 규칙은 Concentration = Catalyst*Stir Rate 입니다 .

" 효과의 별칭 구조 " 개요에서 첫 번째 관계를 얻기 위해 이 두 생성 규칙을 적용하면 다음과 같은 표현식이 제공됩니다 .

Temperature = Feed Rate*Catalyst*Stir Rate = Feed Rate*Concentration

– 두 번째 등식은 두 번째 생성 규칙을 사용하여 Catalyst*Stir Rate 를 Concentration 으로 대체한 결과입니다 .

이제 첫 번째 표현식과 세 번째 표현식에 Concentration 으로 사후 곱셈을 적용하여 다음 표현식을 얻습니다 .

Temperature*Concentration = Feed Rate*Concentration*Concentration

설계 행렬의 Concentration 열에 -1 및 +1 값이 포함되어 있으므로 Concentration*Concentration 항은 +1 값의 열을 나타냅니다 . 다음 표현식은 " 효과의 별칭 구조 " 개요에 표시되는 첫 번째 별칭 관계가 됩니다 .

Temperature*Concentration = Feed Rate*I = Feed Rate

비슷한 계산을 사용하여 다른 별칭 관계도 구할 수 있습니다 .

주효과 선별 설계

실험에 범주형 또는 이산 수치형 요인이 포함되거나 런 수가 제한되는 경우 주효과 선별을 위한 직교 설계를 구성하지 못할 수 있습니다 . 그러나 주효과 선별 설계를 구성할 수 있습니다 . 자세한 내용은 Lekivetz et al. (2015) 에서 확인하십시오 .

주효과 선별 설계는 카이제곱 기준에 설명된 대로 균형 특성이 좋은 설계입니다 . 자세한 내용은 "카이제곱 효율도"(292 페이지)에서 확인하십시오 . 이러한 설계에는 주효과 모형에 대해 적합한 통계 특성이 있습니다 .

설계를 생성하는 데 사용되는 알고리즘은 강도 2 의 직교 배열을 구성하려고 합니다 . 강도 2 직교 배열은 교호작용을 무시할 수 있을 때 주효과의 직교 추정을 허용합니다 . 이러한 배열은 선별 설계에 이상적입니다 . 해상도 3 의 정칙 부분 요인 설계와 Plackett-Burman 설계는 강도 2 직교 배열의 예입니다 .

설계에서 요인 수준에 대해 가능한 모든 쌍을 고려하십시오. 알고리즘은 수준 쌍 수의 균형을 가능한 한 맞추려고 합니다. 고정된 수의 열이 생성되었다고 가정하면 균형된 새 열이 무작위로 구성됩니다. 새 열을 포함하는 쌍에 대해 달성한 균형 정도를 반영하는 측도가 정의됩니다. 알고리즘은 새 열 내에서 수준을 교환하여 이 측도를 최소화하려고 합니다.

카이제곱 효율도

설계에 설계 행렬의 열에 해당하는 n 회 런과 p 개 요인이 있다고 가정해 보겠습니다.

- 요인 k 와 l 의 수준을 각각 $a = 0, 1, \ldots, s_k - 1$ 과 $b = 0, 1, \ldots, s_l - 1$ 로 나타냅니다.
- 수준 조합 (a,b) 가 k 및 l 열에 나타나는 횟수를 $n_{kl}(a,b)$ 로 나타냅니다.

k 및 l 열에 의해 증명되는 직교성 결여 측도는 다음 표현식으로 지정됩니다.

$$\chi^2_{kl} = \sum_{a=0}^{s_k-1} \sum_{b=0}^{s_l-1} \frac{[n_{kl}(a,b) - n/(s_k s_l)]^2}{n/(s_k s_l)}$$

설계의 평균 비직교성 측도는 다음 표현식으로 지정됩니다.

$$\chi^2 = \sum_{1 \le k < l \le p} \chi^2_{kl} / [p(p-1)/2]$$

χ^2 의 가능한 최대값 (χ^2_{max}) 이 계산됩니다. 설계의 카이제곱 효율도는 다음과 같이 정의됩니다.

$$\text{카이제곱 효율도} = 100(1 - \chi^2/\chi^2_{max})$$

카이제곱 효율도는 수준 쌍이 극단적 균형 부족을 나타내는 설계와 비교하여 χ^2가 0에 얼마나 가까운지 나타냅니다.

설계 생성

주효과 선별 설계를 구성할 때 " 설계 생성 " 개요를 사용하여 런 수를 지정할 수 있습니다. 설계를 생성하려면 **설계 생성**을 클릭합니다.

최소값 설계 생성이 실패하지 않게 하기 위해 필요한 런 수의 하한입니다. **최소값**을 선택하면 결과 설계가 포화됩니다. 따라서 오차 자유도가 없습니다.

참고 : 런 수를 " 최소값 " 으로 선택하면 검정할 오차 항이 없습니다. 모수 추정값을 검정할 수 없습니다. 이 선택은 추가 런 비용이 너무 높은 경우에만 적합합니다.

기본값 런 수를 제안합니다. 이 값은 런 수가 최소 런 수보다 4 회 이상 많은 균형 설계를 생성하기 위한 발견적 방법을 기반으로 합니다.

사용자 지정 원하는 런 수를 지정합니다. **런 수** 텍스트 상자에 해당 값을 입력합니다. 이 옵션을 사용하면 추가 런 비용을 잠재적 정보 획득으로 상쇄할 수 있습니다.

설계

"설계" 개요에서는 주효과 선별 설계를 위한 런을 보여 줍니다. 설계 테이블의 런 순서를 변경하려면 테이블을 생성하기 전에 "출력 옵션" 패널에서 "런 순서" 옵션을 선택할 수 있습니다.

설계 평가

참고 : Cotter 설계에는 " 설계 평가 " 개요가 표시되지 않습니다 .

" 설계 평가 " 개요에서는 생성된 설계의 특성을 평가하는 다양한 방법을 제공합니다 . " 설계 평가 " 개요를 열고 다음 옵션을 확인합니다 .

검정력 분석 주어진 크기의 효과를 감지할 수 있는지 분석할 수 있습니다 .

예측 분산 프로파일 요인 설정 범위에 대한 예측 분산을 보여 줍니다 .

설계 공간 비율 그림 모형 예측 분산이 지정된 값보다 얼마나 아래 또는 위에 있는지 보여 줍니다 .

예측 분산 표면 두 개의 연속형 요인에 대한 예측 분산의 표면 그림을 보여 줍니다 .

추정 효율도 각 모수에 대해 이상적인 직교 설계 (존재하지 않을 수 있음) 와 비교하여 신뢰 구간 길이에 대한 증가 비율을 제공합니다 . 모수의 상대 표준 오차도 제공합니다 .

별칭 행렬 모형에 없지만 활성화 가능성이 있는 효과에 의해 모형 모수가 편향되는 정도를 나타내는 계수를 제공합니다 .

상관 색상 맵 강도 척도를 사용하여 그림에서 효과 사이의 절대 상관을 표시합니다 .

설계 진단 설계를 구성하는 데 사용되는 최적 기준을 나타냅니다 . 설계의 효율도 측도도 제공합니다 .

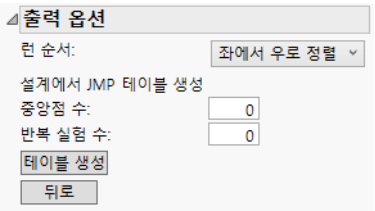
> **참고 :** 설계 진단에 사용된 모형에는 모든 주효과 및 2 요인 교호작용이 포함됩니다 (모든 2 요인 교호작용을 추정할 수 있는 경우). 그렇지 않은 경우에는 모형에 모든 주효과가 포함됩니다 .

"설계 평가" 패널에 대한 자세한 내용은 "설계 평가" 장의 "설계 평가"(418 페이지)에서 확인하십시오 .

출력 옵션

"출력 옵션" 패널에서 출력 데이터 테이블에 대한 상세 정보를 지정합니다 . 끝나면 **테이블 생성** 을 클릭하여 설계에 대한 데이터 테이블을 구성합니다 . 그림 9.17 에서는 "설계 목록" 에서 선택한 표준 설계의 "출력 옵션" 패널을 보여 줍니다 . 주효과 선별 설계의 경우 "런 순서" 만 있습니다 .

그림 9.17 출력 옵션 선택



런 순서

런 순서 옵션은 설계 테이블의 런 순서를 결정합니다 .

동일하게 유지 설계 테이블에 표시되는 행 순서가 " 코드화된 설계 " 또는 " 설계 " 개요의 순서와 동일합니다 .

좌에서 우로 정렬 설계 테이블의 열이 왼쪽에서 오른쪽으로 정렬됩니다 .

랜덤화 설계 테이블의 행이 랜덤 순서로 나타납니다 .

우에서 좌로 정렬 설계 테이블의 열이 오른쪽에서 왼쪽으로 정렬됩니다 .

블록 내 랜덤화 설계 테이블의 행이 블록 내에서 랜덤 순서로 나타납니다 . **주효과 선별 설계 생성** 을 선택한 경우에는 사용할 수 없습니다 .

중앙점 및 반복 실험

중앙점 수 설계에 중앙점으로 포함할 추가 런 수를 지정합니다 . 중앙점은 모든 연속형 요인이 요인 범위의 가운데에 설정되는 런입니다 . **주효과 선별 설계 생성**을 선택한 경우에는 이 옵션을 사용할 수 없습니다 .

설계에 연속형 요인과 범주형 요인이 모두 포함되어 있다고 가정해 보겠습니다 . " 출력 옵션 " 패널에서 중앙점을 요청할 경우 다음과 같이 중앙점이 분포됩니다 .

1. 범주형 요인의 설정은 " 요인 " 개요에 지정된 값 순서를 사용하여 정렬됩니다 .
2. 범주형 요인의 각 설정 조합에 순서대로 중앙점 하나가 할당되고 모든 중앙점이 할당될 때까지 반복됩니다 .

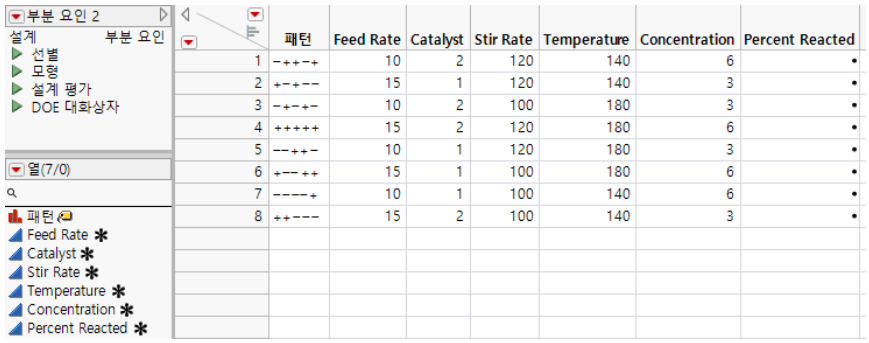
반복 실험 수 " 설계 목록 " 의 설계에 대해 중앙점을 포함하여 전체 설계를 반복할 횟수를 지정합니다 . 한 번 반복하면 런 수가 두 배가 됩니다 . **주효과 선별 설계 생성**을 선택한 경우에는 이 옵션을 사용할 수 없습니다 .

참고 : 중앙점 또는 반복 실험을 요청하고 " 테이블 생성 " 을 반복해서 클릭하면 가장 최근에 구성된 설계 테이블에 이러한 작업이 적용됩니다 .

테이블 생성

테이블 생성을 클릭하여 실험 런이 포함된 데이터 테이블을 생성합니다. 이 테이블에는 지정된 상한값과 하한값이 각 런에 대해 나타납니다.

그림 9.18 설계 데이터 테이블



	패턴	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	Percent Reacted
1	−++−+	10	2	120	140	6	•
2	+−+−−	15	1	120	140	3	•
3	−+−+−	10	2	100	180	3	•
4	+++++	15	2	120	180	6	•
5	−−++−	10	1	120	180	3	•
6	+−−++	15	1	100	180	6	•
7	−−−−+	10	1	100	140	6	•
8	++−−−	15	2	100	140	3	•

테이블 이름은 해당 테이블을 생성한 설계 유형입니다.

설계 테이블에는 다음과 같은 스크립트가 포함되어 있습니다.

선별 "DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합" 플랫폼을 실행합니다. 모든 요인이 2 수준인 경우에만 제공됩니다.

모형 "분석 > 모형 적합" 플랫폼을 실행합니다.

설계 평가 "DOE > 설계 진단 > 설계 평가" 플랫폼을 실행합니다.

DOE 대화상자 설계 테이블을 생성하는 데 사용된 "선별 설계" 창을 다시 생성합니다. 또한 이 스크립트는 설계를 생성하는 데 사용된 난수 시드값도 포함합니다.

데이터를 분석하려면 **선별** 또는 **모형** 스크립트를 실행합니다.

"설계 목록"에서 설계를 선택한 경우 설계 테이블에 **패턴** 열이 포함됩니다. **패턴** 열에는 지정된 행의 런을 요약하는 항목이 포함되어 있습니다. 하한 설정은 "–", 상한 설정은 "+", 중앙점은 "0"으로 각각 나타냅니다. **패턴**은 그림에서 라벨 변수로 유용할 수 있습니다.

선별 설계 옵션

"선별 설계"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

반응 저장 "반응" 패널의 정보를 새 데이터 테이블에 저장합니다. 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다. 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다.

반응 불러오기 "반응 저장" 옵션을 사용하여 저장한 반응을 불러옵니다.

요인 저장 "요인" 패널의 정보를 새 데이터 테이블에 저장합니다. 각 요인의 열에는 수준이 포함됩니다. 다른 정보는 열 특성으로 저장됩니다. 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다. 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다. 그런 다음 " 설계 역할 " 영역에서 적절한 역할을 선택합니다.

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 불러옵니다.

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) " 요인 제약 조건 정의 " 또는 " 선형 제약 조건 " 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다. 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다. 마지막 행은 부등식 한계를 포함합니다. 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다. 자세한 내용은 " 열 특성 " 부록의 " 제약 조건 상태 "(762 페이지) 에서 확인하십시오.

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) " 제약 조건 저장 " 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다.

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다. 이러한 작업에는 다음 중 하나 이상이 포함됩니다.

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다.

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다.

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다. " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오.

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다.

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다.
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다. 계산식과 값은 설계 창에 지정된 모형에 기초합니다.
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다.

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다.

– **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다. 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 **기본 분석**의 " 시뮬레이션 " 장에서 확인하십시오 .

행렬 X 저장 " 적률 행렬 " 및 " 모형 행렬 " 을 설계 데이터 테이블의 테이블 스크립트에 저장합니다 . 이 스크립트에는 적률 및 설계 행렬이 포함됩니다 . 자세한 내용은 " 사용자 설계 " 장의 " 행렬 X 저장 "(105 페이지) 에서 확인하십시오 .

주의 : 명목형 요인이 있는 설계의 경우 " 행렬 X 저장 " 옵션으로 저장한 모형 행렬은 선형 모형을 적합시킬 때 사용된 코딩 행렬이 **아닙니다** . " 모형 " 스크립트를 실행할 때 나타나는 " 모형 적합 " 보고서에서 " 열 저장 > 코드화 테이블 저장 " 옵션을 선택하면 모형 적합에 사용되는 코딩 행렬을 가져올 수 있습니다 .

Cotter 설계 제한 설계 목록에서 Cotter 설계를 제외합니다 . 이 옵션은 기본적으로 선택되어 있습니다 . 설계 목록에 Cotter 설계를 표시하려면 이 옵션을 선택 취소합니다 .

참고 : Cotter 설계를 항상 표시하도록 환경 설정을 지정할 수 있습니다 . **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택한 후 "Cotter 설계 제한 " 을 선택 취소하십시오 .

시작 수 (주효과 선별 설계에만 사용 가능) 최종 설계를 최적화하기 위해 알고리즘이 전체 설계를 처음부터 다시 생성하는 최대 횟수를 지정합니다 .

설계 검색 시간 (주효과 선별 설계에만 사용 가능) 설계 검색을 위한 최대 시간 (초) 을 지정합니다 . 기본 검색 시간은 15 초입니다 .

알고리즘 반복에 몇 초 이상이 필요한 경우 " 설계 계산 중 " 진행률 창이 나타납니다 . 진행률 표시줄에 **카이제곱 효율도**가 표시됩니다 . 자세한 내용은 " 카이제곱 효율도 "(292 페이지) 에서 확인하십시오 . 진행률 창에서 **취소**를 클릭하면 계산이 중지되고 해당 시점에서 찾은 최적 설계가 제공됩니다 .

참고 : " 설계 검색 시간 " 에 대한 환경 설정을 지정할 수 있습니다 . **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다 . **설계 검색 시간**을 선택한 후 최대 시간 (초) 을 입력합니다 . 직교 배열이 발견되면 검색이 종료됩니다 . 시간이 더 필요한 특정 상황에서는 JMP 가 검색 시간을 자동으로 연장합니다 .

열 시작 수 (주효과 선별 설계에만 사용 가능) 다음 열을 구성하기 전에 알고리즘이 지정된 열을 최적화하려고 시도하는 최대 횟수를 지정합니다 . 기본 열 시작 수는 50 입니다 . 자세한 내용은 " 주효과 선별 설계 "(291 페이지) 에서 확인하십시오 .

스크립트를 스크립트 창에 저장 "선별 설계" 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

선별 설계의 추가 예

- " 부분 요인 설계에서 생성 규칙 수정 "
- "Plackett-Burman 설계 "

부분 요인 설계에서 생성 규칙 수정

Meyer, et al. (1996) 에서 채택된 다음 예에서는 많은 요인이 있을 때 선별 설계 플랫폼을 사용하는 방법을 보여 줍니다 . 이 예에서 화학 엔지니어는 화학 공정의 반응 백분율에 대한 5 개 요인의 효과를 조사합니다 . 요인은 다음과 같습니다 .

- Feed Rate - 반응실에 추가되는 원료의 양 (리터 / 분)
- Catalyst(%)
- Stir Rate - 반응실의 프로펠러 RPM
- Temperature(섭씨)
- Concentration(반응물 농도)

생산 제약 조건에 따라 실험 크기가 12 회 런 이하로 제한됩니다 . 8 회 런 부분 요인 설계와 12 회 런 Plackett-Burman 설계를 고려하기로 결정했습니다 . 또한 다음 내용이 참이라고 가정합니다 .

- Temperature*Concentration 교호작용이 활성이므로 이 교호작용이 주효과와 별칭 관계를 갖지 않는 설계를 원합니다 .
- Catalyst*Temperature* 교호작용이 활성일 가능성이 낮습니다 .
- Stir Rate*Concentration 교호작용이 활성일 가능성이 낮습니다 .

이 정보를 사용하여 설계를 구성합니다.

표준 부분 요인 설계 생성

표준 부분 요인 설계를 생성하려면 다음을 수행하십시오.

- " 반응 지정 "(299 페이지)
- " 요인 지정 "(299 페이지) 또는 " 수동으로 요인 지정 "(299 페이지)
- " 설계 선택 "(300 페이지)

반응 지정

1. **DOE > 전통적 설계 > 2 수준 선별 > 선별 설계**를 선택합니다 .
2. " 반응 이름 " 아래의 Y 를 두 번 클릭하고 Percent Reacted 를 입력합니다 .

   기본 목표는 최대화로 되어 있습니다 . **목표**는 반응을 최대화하는 것이지만 허용 가능한 최소 반응 백분율은 90(하한) 이고 상한은 100(상한) 입니다 .
3. " 하한 " 아래를 클릭하고 90 을 입력합니다 .
4. " 상한 " 아래를 클릭하고 100 을 입력합니다 .
5. " 중요도 " 아래 영역은 비워 둡니다 .

   반응이 하나뿐이므로 기본적으로 반응의 중요도가 1 로 지정됩니다 .

   그림 9.19 에서는 완료된 " 반응 " 개요를 보여 줍니다 . 이제 요인을 지정합니다 .

요인 지정

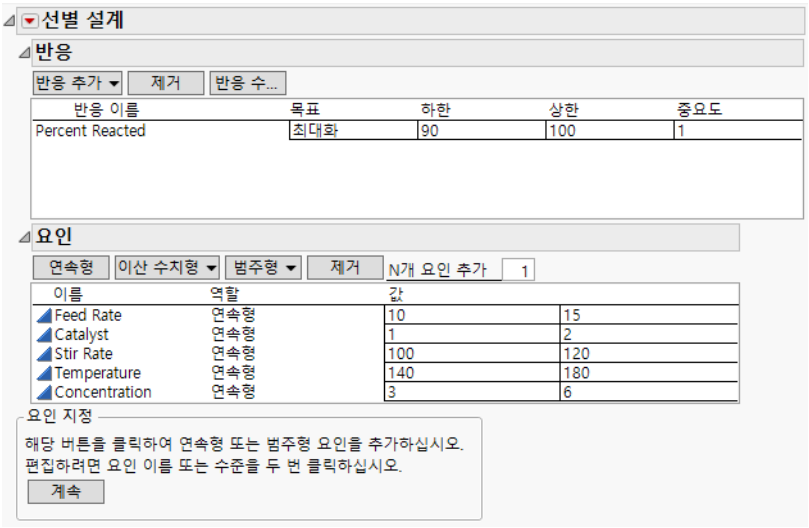
수동 또는 자동으로 요인을 입력할 수 있습니다 .

- 요인을 자동으로 입력하려면 Reactor Factors.jmp 데이터 테이블을 사용하십시오 .
  1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor Factors.jmp 를 엽니다 .
  2. " 선별 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 . " 설계 선택 "(300 페이지) 으로 진행합니다 .
- 요인을 수동으로 입력하려면 아래 단계를 수행하십시오 .

수동으로 요인 지정

1. **N 개 요인 추가** 상자에 5 를 입력하고 **연속형**을 클릭하여 5 개의 연속형 요인을 추가합니다 .
2. 기본 요인 이름 (X1 ~ X5) 을 Feed Rate, Catalyst, Stir Rate, Temperature 및 Concentration 으로 변경합니다 .
3. 각 요인의 하한값과 상한값을 입력합니다 .
   - Feed Rate: 10, 15
   - Catalyst: 1, 2
   - Stir Rate: 100, 120
   - Temperature: 140, 180
   - Concentration: 3, 6

그림 9.19 반응 및 요인 개요

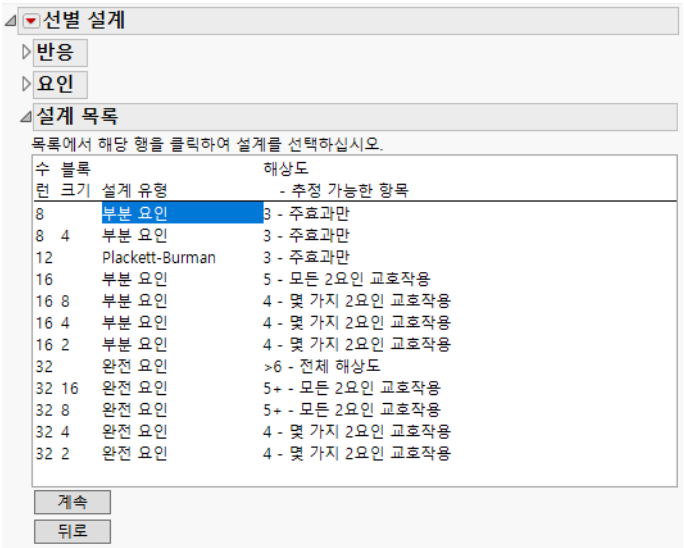


설계 선택

1. **계속**을 클릭합니다 .
2. " 선별 유형 선택 " 패널에서 기본 선택인 **부분 요인 설계 목록에서 선택**을 적용하고 **계속**을 클릭합니다 .

   지정한 요인과 수준에 대한 설계가 " 설계 목록 " 에 나열됩니다 (그림 9.20).

그림 9.20 5 개 연속형 요인에 대한 부분 요인 설계



3. 원하는 설계가 목록에서 맨 처음에 나오고 기본으로 선택되어 있습니다 (그림 9.20). 해당 선택을 적용하고 **계속**을 클릭합니다 .

   8 회 런으로 제한되고 블록 요인이 없으므로 최적 설계 옵션은 블록이 없는 8 회 런 부분 요인 설계입니다 . 이 설계는 2^{5-2} 부분 요인 설계입니다 . 즉 , 5 개 요인에 대한 완전 요인 설계의 1/4 입니다 .

다른 부분을 얻기 위해 생성 규칙 변경

이 예에서는 Temperature*Concentration 교호작용이 주효과와 교락되는지 여부를 알아보려고 합니다 . " 설계 표시 및 수정 " 개요를 사용하여 선택한 설계의 별칭 구조를 검토하고 해당하는 경우 변경하십시오 .

1. " 효과의 별칭 구조 " 개요를 엽니다 .

그림 9.21 효과의 별칭 구조 개요

효과의 별칭 구조

효과	별칭
Feed Rate	= Temperature*Concentration
Catalyst	= Stir Rate*Concentration
Stir Rate	= Catalyst*Concentration
Temperature	= Feed Rate*Concentration
Concentration	= Feed Rate*Temperature = Catalyst*Stir Rate
Feed Rate*Catalyst	= Stir Rate*Temperature
Feed Rate*Stir Rate	= Catalyst*Temperature

   활성으로 의심되는 Temperature*Concentration 교호작용은 주효과 Feed Rate 와 교락됩니다 . Feed Rate 가 비활성으로 의심되는 효과와 별칭 관계가 있고 Temperature*Concentration 교호작용이 주효과와 별칭 관계가 없는 설계를 구성하기 위해 생성 규칙을 변경하려고 합니다 .

2. " 생성 규칙 변경 " 개요를 엽니다 .

   기본 생성 규칙은 완전 요인 설계의 표준 (주) 1/4 부분을 제공합니다 . Catalyst*Temperature 및 Stir Rate*Concentration 교호작용이 활성일 가능성이 낮다고 의심했습니다 . 이러한 두 교호작용이 Feed Rate 와 교락되도록 생성 규칙을 재정의합니다 . 재정의된 생성 규칙은 완전 요인 설계의 다른 1/4 부분을 제공합니다 .

3. 다음을 수행합니다 .
   - Temperature 열에서 Stir Rate 를 선택 취소합니다 .
   - Concentration 열에서 Catalyst 를 선택 취소합니다 .
   - Concentration 열에서 Feed Rate 를 선택합니다 .

그림 9.22 새 생성 규칙

생성 규칙 변경

요인	Temperature	Concentration
Feed Rate	☑	☑
Catalyst	☑	☐
Stir Rate	☐	☑
+/-	☑	☑

적용

4. **적용**을 클릭합니다 .

그림 9.23 수정된 생성 규칙에 대한 효과의 별칭 구조 개요

효과의 별칭 구조

효과	별칭
Feed Rate	= Catalyst*Temperature = Stir Rate*Concentration
Catalyst	= Feed Rate*Temperature
Stir Rate	= Feed Rate*Concentration
Temperature	= Feed Rate*Catalyst
Concentration	= Feed Rate*Stir Rate
Catalyst*Stir Rate	= Temperature*Concentration
Catalyst*Concentration	= Stir Rate*Temperature

정의한 설계에서는 Feed Rate 가 Catalyst*Temperature 및 Stir Rate*Concentration 과 교락됩니다 . 또한 Temperature*Concentration 교호작용이 이제 이원 교호작용 Catalyst*Stir Rate 와 교락됩니다 .

5. " 출력 옵션 " 개요에서 기본 런 순서 설정인 **랜덤화**를 적용하고 **테이블 생성**을 클릭합니다 .

그림 9.24 8 회 런 부분 요인 설계 테이블

부분 요인
설계 부분 요인
선별
모형
설계 평가
DOE 대화상자

열(7/0)
패턴
Feed Rate ✱
Catalyst ✱
Stir Rate ✱
Temperature ✱
Concentration ✱
Percent Reacted ✱

	패턴	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	Percent Reacted
1	−−−++	10	1	100	180	6	•
2	−+++−	10	2	120	140	3	•
3	+−+−+	15	1	120	140	6	•
4	−−++−	10	1	120	180	3	•
5	++−+−	15	2	100	180	3	•
6	+++++	15	2	120	180	6	•
7	−+−−+	10	2	100	140	6	•
8	+−−−−	15	1	100	140	3	•

설계 테이블에서는 구성된 설계를 보여 줍니다 . 이 테이블에는 " 선별 " 창에서 정의한 반응에 대한 열 (Percent Reacted) 이 포함되어 있으며 , 여기에 실험 결과를 기록할 수 있습니다 .

선별 , 모형 및 DOE 대화상자 스크립트도 포함되어 있습니다 . 이러한 스크립트에 대한 자세한 내용은 " 테이블 생성 "(294 페이지) 에서 확인하십시오 .

결과 분석

이제 실험을 수행하고 데이터를 기록한 후 결과 분석을 진행합니다 .

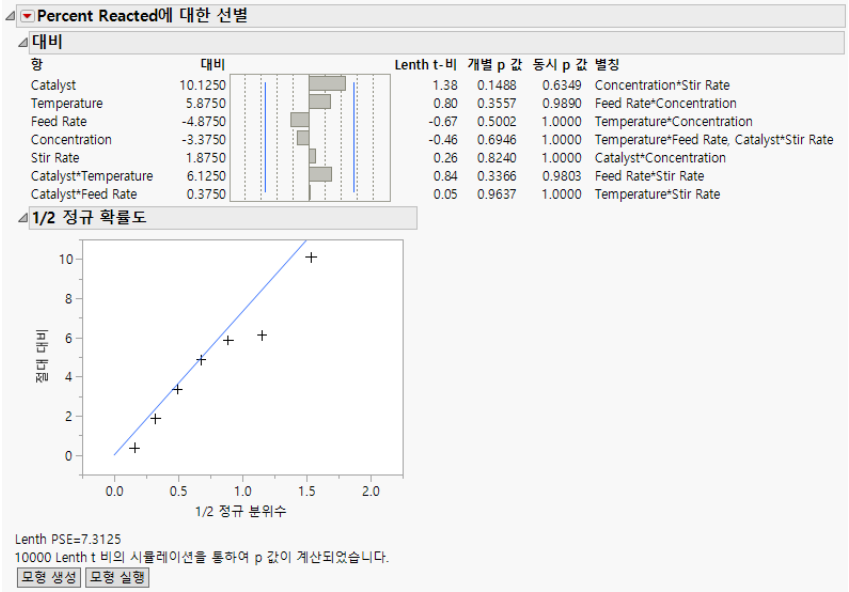
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 8 Runs.jmp 를 엽니다 .

   8 회 런으로 7 개 효과를 추정할 수 있습니다 . 이 중 몇 개만 활성일 것으로 예상합니다 . 7 개 효과를 추정하려고 하므로 오차에 대한 자유도가 없습니다 . 이러한 이유로 선별 플랫폼을 사용하여 결과를 분석합니다 .

2. 데이터 테이블에서 **선별** 스크립트를 실행합니다 .

선별 스크립트는 반응 및 요인에 대한 선별 플랫폼 (**DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**) 을 시작합니다 .

그림 9.25 선별 보고서의 예



참고 : *p* 값은 시뮬레이션 기반 기법을 사용하여 계산되므로 여기에 표시된 값과 *p* 값이 정확하게 일치하지 않을 수 있습니다 .

보고서에는 Lenth t- 비를 기반으로 개별 및 동시 *p* 값이 모두 표시됩니다 . 개별 *p* 값과 관련해서도 유의한 효과가 없습니다 . "1/2 정규 확률도 " 는 효과에 랜덤 잡음만 반영됨을 나타냅니다 .

선별 보고서에 대한 자세한 내용은 "2 수준 선별 적합 플랫폼 " 장의 " 선별 보고서 "(315 페이지) 에서 확인하십시오 .

Plackett-Burman 설계

부분 요인 예에서는 5 개 연속형 요인에 대한 8 회 런 부분 요인 설계를 보여 줍니다 . 그러나 4 회 런을 더 추가할 수 있다고 가정해 보겠습니다 . 이 예에서는 12 회 런 Plackett-Burman 설계를 구성합니다 . "선별 " 창을 쉽게 완료하기 위해 "반응 불러오기 " 및 "요인 불러오기 " 명령을 사용합니다 .

Plackett-Burman 설계 생성

1. **DOE > 전통적 설계 > 2 수준 선별 > 선별 설계** 를 선택합니다 .

2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor Response.jmp 를 엽니다 .
3. " 선별 설계 " 의 빨간색 삼각형을 클릭하고 **반응 불러오기**를 선택합니다 .
4. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor Factors.jmp 를 엽니다 .
5. " 선별 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

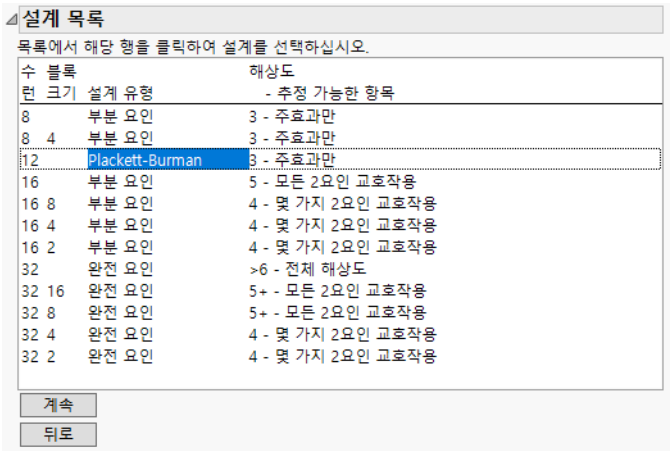
   반응 불러오기 및 **요인 불러오기** 명령을 사용하면 " 반응 " 및 " 요인 " 개요에 반응 / 요인 이름 , 반응의 목표와 한계 , 요인 값 등이 채워집니다 . 완료된 반응 및 요인 개요는 그림 9.19 에서 확인하십시오 .
6. **계속**을 클릭합니다 .

참고 : 다음 단계에서 난수 시드값을 설정하면 이 예에 나오는 런 순서가 재현됩니다 . 설계를 직접 구성할 때는 이 단계가 필요하지 않습니다 .

7. (선택 사항) " 선별 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 후 34567 을 입력하고 **확인**을 클릭합니다 .
8. " 선별 유형 선택 " 패널에서 기본 선택인 **부분 요인 설계 목록에서 선택**을 적용하고 **계속**을 클릭합니다 .
9. Plackett-Burman 설계를 선택합니다 .

   런 크기가 2 의 거듭제곱이 아닌 Plackett-Burman 설계는 복잡한 별칭 구조를 갖는 경향이 있습니다 . 특히 주효과는 여러 가지 이원 교호작용과 부분적으로 별칭 관계를 가질 수 있습니다 . 자세한 내용은 " 설계 평가 "(305 페이지) 에서 확인하십시오 . 12 회 런 Plackett-Burman 설계는 해상도 3 으로 지정됩니다 .

그림 9.26 Plackett-Burman 선별 설계를 보여 주는 설계 목록



10. **계속**을 클릭합니다 .
11. **테이블 생성**을 클릭합니다 .

그림 9.27 Placket-Burman 설계에 대한 설계 테이블

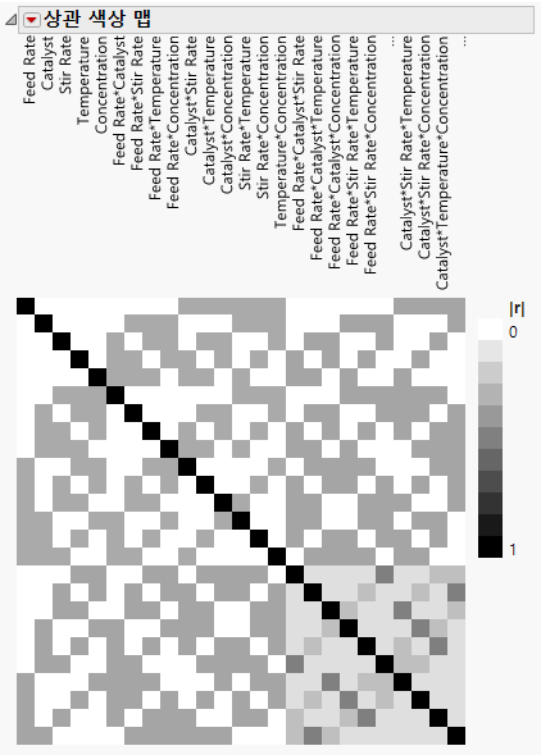
	패턴	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	Percent Reacted
1	+++++	15	2	120	180	6	•
2	+−−−+	15	1	100	140	6	•
3	+−+++	15	1	120	180	6	•
4	−+++−	10	2	120	180	3	•
5	++−−−	15	2	100	140	3	•
6	−−−+−	10	1	100	180	3	•
7	−+−++	10	2	100	180	6	•
8	+++−−	15	2	120	140	3	•
9	−−+−+	10	1	120	140	6	•
10	−−+−−	10	1	120	140	3	•
11	−+−−+	10	2	100	140	6	•
12	+−−+−	15	1	100	180	3	•

Percent Reacted 열이 설계 테이블에 포함되어 있습니다. 테이블에 표시된 순서대로 실험 런을 수행하고 결과를 Percent Reacted 열에 기록해야 합니다.

설계 평가

1. " 선별 설계 " 창으로 돌아갑니다 . 창을 닫았으면 설계 테이블에서 "DOE 대화상자 " 스크립트를 실행합니다 .
2. **설계 평가 > 상관 색상 맵** 개요를 엽니다 .

그림 9.28 절대 상관의 색상 맵



주대각 셀은 예상대로 상관계수가 1 인 상관관계를 가집니다 . 흰색 셀은 상관관계가 0 인 효과에 해당합니다 . 회색 및 검은색 셀은 상관관계가 0 보다 큰 효과에 해당합니다 . 관련된 효과와 절대 상관을 표시하려면 커서로 셀을 가리킵니다 . 예를 들어 Feed Rate 는 몇 가지 이원 및 삼원 교호작용과 상관관계가 있습니다 .

참고 : 선별 설계의 색상 맵에는 요인이 8 개 이하인 설계의 삼원 교호작용이 포함됩니다 . 설계 평가 스크립트의 색상 맵에는 이원 교호작용만 포함됩니다 .

3. " 별칭 행렬 " 개요를 엽니다 .

그림 9.29 최대 이원 교호작용을 보여 주는 별칭 행렬 (일부분)

별칭 행렬

효과	Feed Rate*Catalyst	Feed Rate*Stir Rate	Feed Rate*Temperature	Feed Rate*Concentration	Catalyst*Stir Rate	Catalyst*Temperature	Catalyst*Concentration	Stir Rate*Temperature	Stir Rate*Concentration	Temperature*Concentration
절편	0	0	0	0	0	0	0	0	0	0
Feed Rate	0	0	0	0	0.333	-0.33	-0.33	0.333	0.333	0.333
Catalyst	0	0.333	-0.33	-0.33	0	0	0	0.333	-0.33	0.333
Stir Rate	0.333	0	0.333	0.333	0	0.333	-0.33	0	0	0.333
Temperature	-0.33	0.333	0	0.333	0.333	0	0.333	0	0.333	0
Concentration	-0.33	0.333	0.333	0	-0.33	0.333	0	0.333	0	0

설계가 주효과에 대해 직교이므로 " 별칭 행렬 " 에서 효과 간 상관관계를 나타내는 숫자 값을 제공합니다 . 자세한 내용은 " 설계 평가 " 장의 " 별칭 행렬 "(430 페이지) 에서 확인하십시오 . 예를 들어 Feed Rate 는 6 개의 이원 교호작용 및 4 개의 삼원 교호작용과 부분적으로 별칭 관계를 갖습니다 . 이러한 교호작용은 Feed Rate 행에서 0.333 및 -0.33 항목에 해당합니다 .

결과 분석

Plackett-Burman.jmp 데이터 테이블에는 설계된 실험의 결과가 포함되어 있습니다 . Temperature* Concentration 교호작용이 활성일 것으로 의심했습니다 . 이 교호작용만 잠재적 활성 교호작용이라는 가정하에 진행합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Plackett-Burman.jmp 를 엽니다 .
2. 왼쪽의 아이콘을 클릭하여 **Model** 스크립트를 실행합니다 .
3. " 열 선택 " 목록에서 Temperature 를 선택하고 " 모형 효과 생성 " 목록에서 Concentration 을 선택합니다 .
4. **교차**를 클릭합니다 .
5. **실행**을 클릭합니다 .

그림 9.30 전체 모형의 모수 추정값

모수 추정값

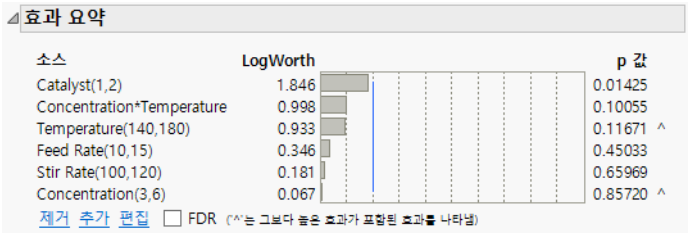
항	추정값	표준 오차	t 비	Prob>|t|
절편	66.166667	2.639444	25.07	<.0001*
Feed Rate(10,15)	-2.333333	2.850926	-0.82	0.4503
Catalyst(1,2)	10.5	2.850926	3.68	0.0142*
Stir Rate(100,120)	1.3333333	2.850926	0.47	0.6597
Temperature(140,180)	5	2.639444	1.89	0.1167
Concentration(3,6)	-0.5	2.639444	-0.19	0.8572
Concentration*Temperature	-6.5	3.232646	-2.01	0.1006

" 실제값 대 예측값 그림 " 은 모형 적합 결여가 없음을 나타냅니다 . " 모수 추정값 " 보고서는 Catalyst 가 0.05 수준에서 유의하고 Concentration*Temperature 교호작용이 0.10 수준에서 거의 유의하다는 것을 보여 줍니다 .

모형 축소

반응에 가장 큰 영향을 미치는 효과를 식별하려고 합니다 . 이러한 활성 효과를 더 명확하게 보려면 " 효과 요약 " 개요를 사용하여 유의하지 않은 효과를 제거합니다 .

그림 9.31 전체 모형의 효과 요약 개요



Concentration 은 최소 유의 효과이지만 p 값 오른쪽의 캐럿에 표시된 대로 고차 교호작용 (Concentration*Temperature) 에 포함됩니다 . 효과 유전성 원리에 따라 Concentration*Temperature 교호작용이 모형에 남아 있는 동안에는 Concentration 을 모형에서 제거하면 안 됩니다 . 자세한

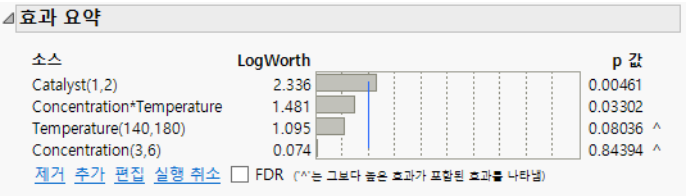
내용은 "DOE 로 시작 " 장의 " 효과 유전성 "(61 페이지) 에서 확인하십시오 . 그 다음 최소 유의 효과는 Stir Rate 입니다 .

1. " 효과 요약 " 개요에서 Stir Rate 를 선택하고 **제거**를 클릭합니다 .

   제거할 수 있는 다음 최소 유의 효과는 Feed Rate 입니다 .

2. " 효과 요약 " 개요에서 Feed Rate 를 선택하고 **제거**를 클릭합니다 .

그림 9.32 축소 모형의 효과 요약 개요



"p 값" 열은 Catalyst 주효과와 Concentration*Temperature 교호작용 둘 다 0.05 수준에서 유의함을 나타냅니다. 모형을 더 이상 축소하면 안 됩니다. 다른 모든 교호작용이 비활성이거나 무시할 수 있는 경우 Catalyst 및 Concentration*Temperature 교호작용이 활성 효과라는 결론을 내릴 수 있습니다.

10 장

2 수준 선별 적합 플랫폼

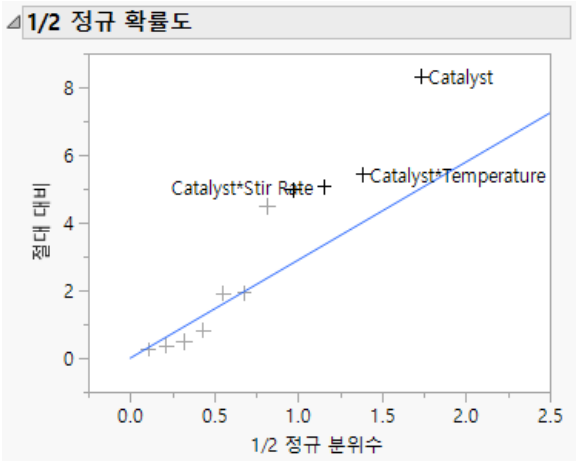
선별 실험의 데이터 분석

2수준 선별 적합 플랫폼은 선별 설계의 실험 데이터 결과를 분석하는 데 사용할 수 있는 모델링 플랫폼입니다. 2수준 선별 적합 플랫폼은 반응에 큰 영향을 미치는 효과를 식별하는 데 유용합니다.

2수준 선별 적합 플랫폼은 효과 희소성 원리를 기반으로 합니다(Box and Meyer 1986). 이 원리는 선별 설계에서 연구하는 효과 중 활성 효과는 상대적으로 적다고 주장합니다. 대부분은 비활성이므로 실제 효과를 무시할 수 있으며 추정값을 랜덤 오차로 처리할 수 있습니다.

관심 모형에 교호작용 항이 포함되어 있으면 선별 설계에서 오차 자유도를 제공하지 않는 경우가 종종 있습니다. 결과적으로 효과에 대해 전통적 검정을 사용할 수 없습니다. 이러한 경우 2수준 선별 적합 플랫폼이 특히 유용합니다.

그림 10.1 2 수준 선별 적합 보고서의 1/2 정규 확률도



목차

2 수준 선별 적합 플랫폼 개요 311
2 수준 선별 적합 플랫폼의 예 311
2 수준 선별 적합 플랫폼 시작 314
선별 보고서 315
대비 315
1/2 정규 확률도 316
모형 생성 또는 실행 316
2 수준 선별 적합 분석의 추가 예 316
Plackett-Burman 설계 분석 317
효과 유전성 예 318
과포화 설계 분석 320
2 수준 선별 적합 플랫폼에 대한 기술 상세 정보 321
효과 입력 순서 322
2수준 선별 적합과 직교 회전 322
Lenth 유사 표준 오차 323
Lenth t-비 323

2 수준 선별 적합 플랫폼 개요

선별 설계 분석은 반응의 변동이 대부분 적은 수의 효과로 설명된다는 **효과 희소성** 원리를 기반으로 합니다. 이 원리에 따라 추정값이 작은 효과를 사용하여 모형의 오차를 추정합니다. 그런 다음 더 큰 효과가 활성인지 여부를 검정할 수 있습니다.

" 모형 적합 "(" 분석 > 모형 적합 ") 또는 "2 수준 선별 적합 "("DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합 ") 을 사용하여 선별 실험의 데이터를 분석할 수 있습니다 . 다음 지침을 사용하여 적절한 모델링 플랫폼을 선택하십시오 .

- 요인이 모두 2 수준 및 직교인 경우 2 수준 선별 적합 플랫폼의 모든 통계량이 적합합니다 .
- 과포화 주효과 설계의 데이터가 있는 경우 2 수준 선별 적합 플랫폼은 활성 요인을 선택하는 데 효과적이지만 오차 또는 유의성을 추정하는 데는 효과적이지 않습니다 . *p* 값 생성을 위한 몬테카를로 시뮬레이션은 이 경우에 유효하지 않은 가정을 사용합니다 .
- 수준이 3 개 이상인 범주형 또는 이산 수치형 요인이 있는 경우 2 수준 선별 적합 플랫폼이 적합하지 않습니다 . JMP 는 연관된 모형 항을 연속형으로 처리합니다 . 이러한 요인의 경우 주효과와 다항식 효과 전체에 변동이 분산됩니다. 이 상황에서는 모형 적합 플랫폼을 사용하는 것이 좋습니다 .
- 데이터가 직교하지 않는 경우 2수준 선별 적합 플랫폼에서 생성된 추정값은 표준 회귀 추정값과 다릅니다 . JMP 는 큰 효과를 식별할 수 있지만 각 효과를 효율적으로 검정하지 못합니다 . 이는 효과를 모형에 입력할 때 인위적으로 직교화하여 초기에 모형에 입력되는 효과가 표준 회귀에서보다 덜 유의하게 보이게 하기 때문입니다 . 자세한 내용은 " 효과 입력 순서 "(322 페이지) 에서 확인하십시오 .
- 혼합물 설계의 경우 2 수준 선별 적합 플랫폼이 적합하지 않습니다 . 자세한 내용은 " 혼합물 설계 " 장의 " 혼합물 설계 적합 "(383 페이지) 에서 확인하십시오 .

2 수준 선별 적합 플랫폼의 예

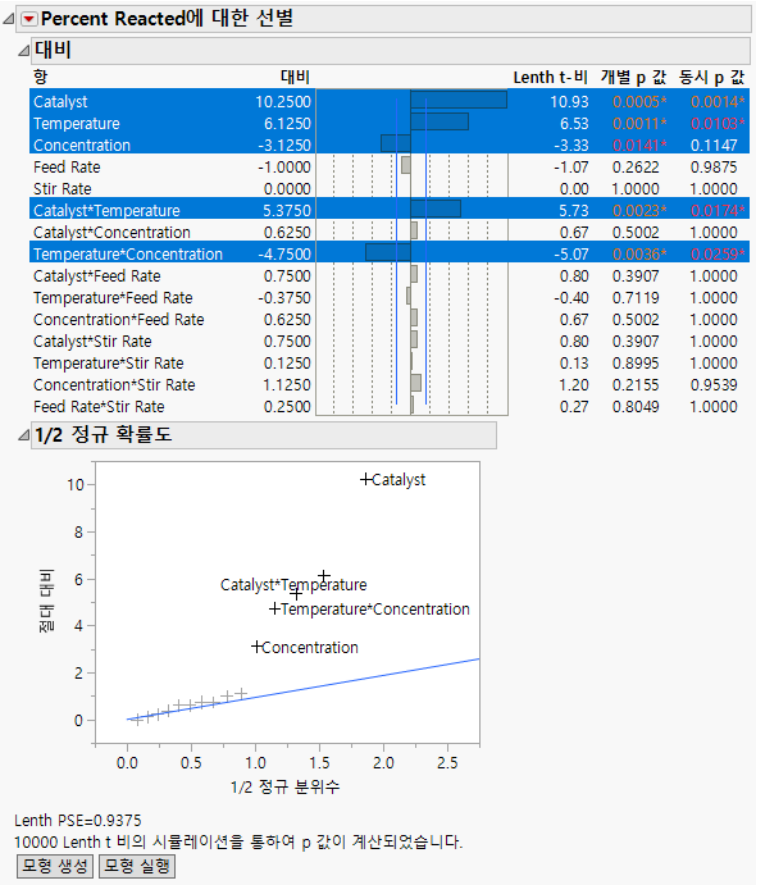
Reactor Half Fraction.jmp 샘플 데이터 테이블에는 Box et al.(1978) 에서 논의한 설계에서 파생된 실험 결과가 포함되어 있습니다. 주효과와 이원 교호작용이 포함된 모형에서 유의한 효과를 식별하려고 합니다. 이 예에서는 16 회 런 설계에 대해 15 개의 모수가 있는 모형을 사용합니다. 또한 이 예에서는 2수준 선별 적합 플랫폼과 모형 적합 플랫폼을 둘 다 사용하여 분석합니다.

2 수준 선별 적합

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor Half Fraction.jmp 를 엽니다 .
2. **DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**을 선택합니다 .
3. Percent Reacted 를 선택하고 **Y** 를 클릭합니다 .
4. Feed Rate ~ Concentration 을 선택하고 **X** 를 클릭한 후 **확인**을 클릭합니다 .

참고 : 2 수준 선별 적합 플랫폼은 교호작용 항이 포함된 모형을 자동으로 생성합니다 . 이는 모형에 포함할 교호작용을 수동으로 지정하는 모형 적합 플랫폼과 대조적입니다 .

그림 10.2 Reactor Half Fraction.jmp 2 수준 선별 적합 설계 보고서



선별 보고서에서 다음과 같은 특징을 알 수 있습니다 .

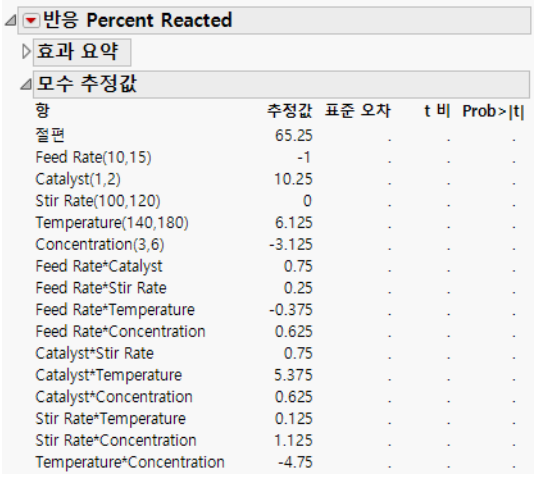
- 개별 *p* 값이 0.10 미만인 효과가 선택되었습니다 .
- Lenth PSE(유사 표준 오차) 를 사용하여 *t* 비를 계산합니다 . Lenth PSE 값은 "1/2 정규 확률도 " 아래에 표시됩니다 .
- 개별 *p* 값과 동시 p 값이 둘 다 표시됩니다 . 0.05 미만인 값에는 별표가 표시됩니다 .
- "1/2 정규 확률도 " 를 사용하면 효과를 빠르게 검토할 수 있습니다 . 효과 목록에서 처음에 선택한 효과도 이 그림에 라벨이 지정되어 있습니다 .

이 예에서는 Catalyst, Temperature 및 Concentration 과 2 요인 교호작용 두 개가 함께 선택됩니다 . 또는 모형 적합 플랫폼에서 같은 모형을 적합시킬 수도 있습니다 .

모형 적합

1. Reactor Half Fraction.jmp 테이블에서 **분석 > 모형 적합**을 선택합니다 .
2. Percent Reacted 를 선택하고 **Y** 를 클릭합니다 .
3. Feed Rate ~ Concentration 을 선택한 후 **매크로 > 특정 차수까지의 요인**을 선택합니다 .
4. **실행**을 클릭합니다 .
5. " 모수 추정값 " 개요를 엽니다 .

그림 10.3 Reactor Half Fraction.jmp 에 대한 포화 설계 모수 추정값



반응 Percent Reacted

효과 요약

모수 추정값

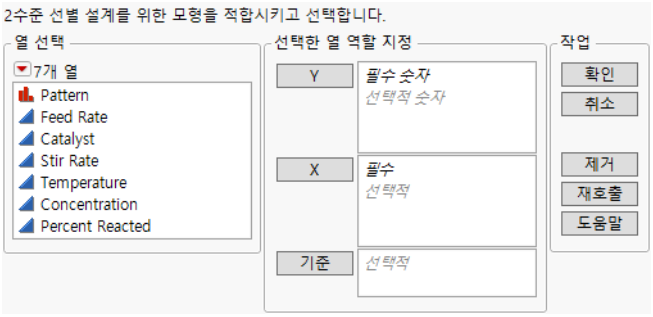
항	추정값	표준 오차	t 비	Prob>|t|
절편	65.25	.	.	.
Feed Rate(10,15)	-1	.	.	.
Catalyst(1,2)	10.25	.	.	.
Stir Rate(100,120)	0	.	.	.
Temperature(140,180)	6.125	.	.	.
Concentration(3,6)	-3.125	.	.	.
Feed Rate*Catalyst	0.75	.	.	.
Feed Rate*Stir Rate	0.25	.	.	.
Feed Rate*Temperature	-0.375	.	.	.
Feed Rate*Concentration	0.625	.	.	.
Catalyst*Stir Rate	0.75	.	.	.
Catalyst*Temperature	5.375	.	.	.
Catalyst*Concentration	0.625	.	.	.
Stir Rate*Temperature	0.125	.	.	.
Stir Rate*Concentration	1.125	.	.	.
Temperature*Concentration	-4.75	.	.	.

관측값이 16 개이고 모형 항이 16 개이므로 오차 항을 추정하기에 관측값이 부족합니다. 오차 추정값이 없으면 표준 검정을 수행할 수 없습니다. 모수 추정값이 제공되지만 오차 자유도가 없으므로 표준 오차, *t* 비, *p* 값이 모두 결측입니다. 이는 선별 설계의 정보를 최대한 활용할 수 있는 2 수준 선별 적합 플랫폼의 장점을 보여 줍니다.

2 수준 선별 적합 플랫폼 시작

DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합을 선택하여 2 수준 선별 적합 플랫폼을 시작합니다 .

그림 10.4 2 수준 선별 적합 플랫폼의 시작 창



Y 분석할 열입니다 . 데이터 유형이 숫자여야 합니다 .

X 연속형 또는 이산형 2 수준 직교 요인입니다 .

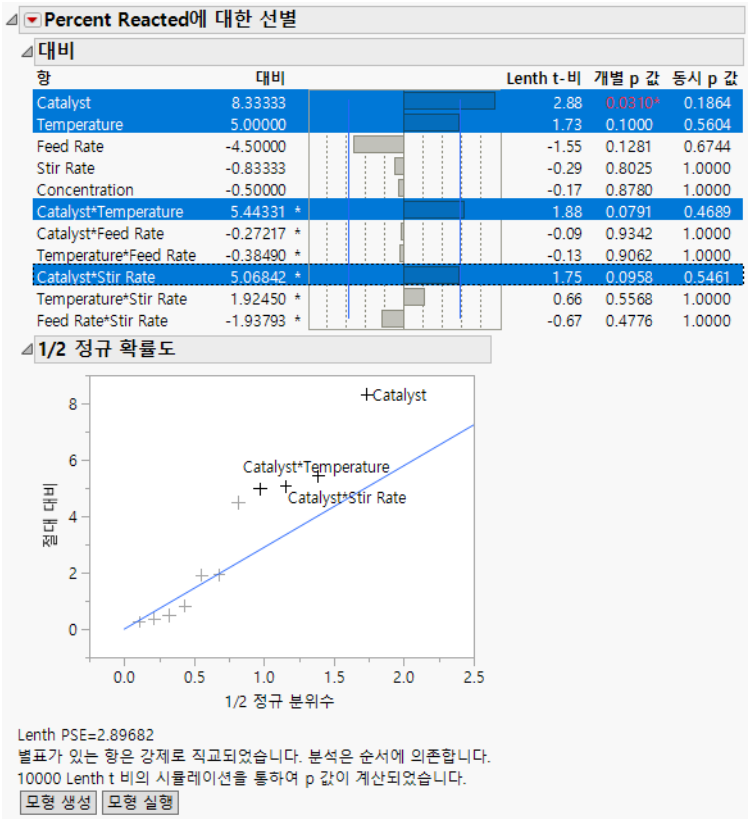
기준 각 수준으로 개별 분석을 정의하는 열입니다 . 지정된 열의 각 수준에 대해 해당하는 행이 분석됩니다 . 결과는 별도의 보고서에 나타납니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

참고 : 선별 설계 플랫폼의 부분 요인 설계 목록에서 설계를 생성한 경우 설계 테이블에 **선별**이라는 스크립트가 포함됩니다 . 분석을 바로 시작하려면 이 스크립트를 실행하십시오 .

선별 보고서

"2 수준 선별 적합 " 보고서는 설계 대비 (후보 모형 항) 요약 테이블과 1/2 정규 확률도를 보여 줍니다 . 보고서에서 선택한 모형을 시작하거나 실행할 수 있습니다 .

그림 10.5 선별 보고서



- " 대비 "(315 페이지)
- "1/2 정규 확률도 "(316 페이지)
- " 모형 생성 또는 실행 "(316 페이지)

대비

" 대비 " 보고서에는 모형 효과 , 대비 값 , Lenth *t*- 비 , 개별 / 동시 *p* 값 및 별칭 (있는 경우) 이 나열됩니다 . 계층 순서에 따라 효과가 분석에 입력됩니다 . 자세한 내용은 " 효과 입력 순서 "(322 페이지) 에서 확인하십시오 . 개별 *p* 값이 0.10 미만인 효과가 선택되었습니다 .

항 요인의 이름입니다 .

대비 요인의 추정값입니다 . 직교 설계의 경우 이 값은 회귀 모수 추정값과 동일합니다 . 비직교 설계의 경우에는 해당하지 않습니다 . 대비 옆에 별표가 표시되어 직교성 결여를 나타낼 수 있습니다 .

막대 차트 0.10 수준에서 유의한 값을 나타내는 파란색 수직선으로 Lenth *t*- 비를 표시합니다 .

Lenth t- 비 대비 /PSE 로 계산된 Lenth *t*- 비이며 , 여기서 PSE 는 Lenth 유사 표준 오차입니다 . 자세한 내용은 "Lenth 유사 표준 오차 "(323 페이지) 에서 확인하십시오 .

개별 p 값 선형 모형의 표준 *p* 값과 유사합니다 . 개별 *p* 값이 작으면 유의한 효과를 나타냅니다 . 자세한 내용은 "Lenth 유사 표준 오차 "(323 페이지) 에서 확인하십시오 .

참고 : 분석을 다시 실행하는 경우 *p* 값이 정확히 동일할 것으로 예상하지 마십시오 . 몬테카를로 방법에 따르면 같은 분석이 반복되는 경우 유사하지만 동일하지 않은 값을 제공해야 합니다 .

동시 p 값 다중 비교를 위한 조정 *p* 값입니다 .

별칭 초기 효과에 대한 후기 효과의 정확한 별칭이 있는 경우에만 나타납니다 .

1/2 정규 확률도

1/2 정규 확률도는 1/2 정규 분포의 분위수 절대값에 대해 그림에 표시된 대비 절대값을 보여 줍니다. 파란색 선은 Lenth 추정값 σ의 기울기로 원점을 통과합니다. 작은 효과는 오차 항으로 간주되며 평균 0, 표준편차 σ인 정규 분포가 있는 것으로 가정합니다. 이러한 항은 파란색 선에 있습니다. 유의한 효과는 평균이 0이 아니고 파란색 선을 벗어납니다.

1/2 정규 확률도는 대화식입니다. 관심 효과 주위를 직사각형 형태로 드래그하여 모형 효과를 선택할 수 있습니다. 보고서에서 Ctrl 키를 누르고 효과 이름을 클릭할 수도 있습니다.

모형 생성 또는 실행

모형 생성 버튼은 선택한 효과가 채워진 " 모형 적합 " 창을 시작합니다 .

모형 실행 버튼은 선택한 효과에 대한 효과 선별 모형을 실행합니다.

참고 : 선택한 모형이 효과 유전성을 따르지 않을 경우 JMP 경고가 나타납니다 . **취소**를 클릭하고 **모형 생성** 버튼을 사용하여 " 모형 적합 " 창을 시작한 후 결측 항을 추가할 수 있습니다 . 아니면 **계속**을 클릭한 후 효과 유전성 규칙을 따르지 않는 모형을 생성할 수 있습니다. 자세한 내용은 "DOE 로 시작 " 장의 " 효과 유전성 "(61 페이지) 에서 확인하십시오 .

2 수준 선별 적합 분석의 추가 예

- "Plackett-Burman 설계 분석 "
- " 효과 유전성 예 "

- " 과포화 설계 분석 "

Plackett-Burman 설계 분석

Plackett-Burman 설계는 부분 요인 선별 설계 대신 사용할 수 있습니다. 2 수준 부분 요인 설계는 특성상 2 의 거듭제곱과 동일한 수의 여러 런이 있어야 합니다. 그러나 12, 24 및 28 회 런 설계의 경우 Plackett-Burman 설계를 사용할 수 있습니다.

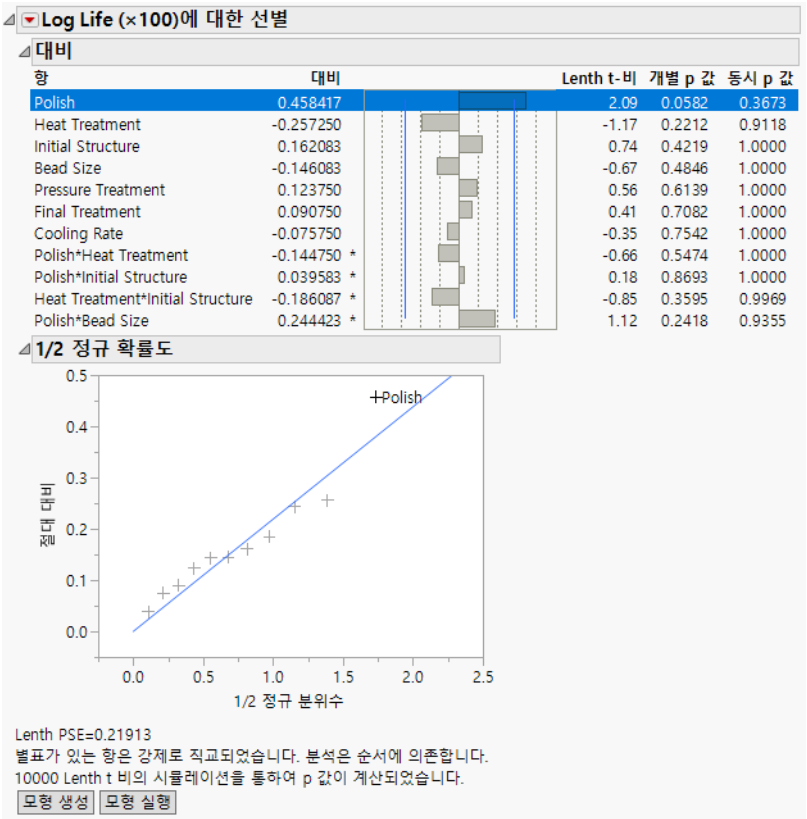
Weld-Repaired Castings.jmp 샘플 데이터 테이블은 Plackett-Burman 설계를 사용하며 Box et al. Box et al.(1978) 에서 논의한 설계에서 파생된 실험 결과가 포함되어 있습니다 . 7 개 요인이 용접 품질에 영향을 줄 것으로 추정됩니다 . 7 개 요인에는 Initial Structure, Bead Size, Pressure Treatment, Heat Treatment, Cooling Rate, Polish 및 Final Treatment 가 포함됩니다 . 12 회 런이 포함된 Plackett-Burman 설계를 사용하여 7 개 요인의 중요도를 조사합니다 . 반응은 100 × log(수명) 입니다 . 오차를 모델링하는 데 사용된 4 개의 항도 샘플 데이터 테이블에 포함되어 있지만 이 분석에는 사용되지 않습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Weld-Repaired Castings.jmp 를 엽니다 .
2. **DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**을 선택합니다 .

   열 특성을 기반으로 시작 창이 채워져 있습니다 . Log Life (x100) 가 반응 **Y** 입니다 .

   7 개 요인 Initial Structure, Bead Size, Pressure Treatment, Heat Treatment, Cooling Rate, Polish 및 Final Treatment 가 **X** 에 대해 지정되어 있습니다 .
3. **X** 창 맨 아래로 스크롤합니다 . ε1, ε2, ε3 및 ε4 를 선택하고 **제거**를 클릭합니다 .
4. **확인**을 클릭합니다 .

그림 10.6 Weld-Repaired Castings.jmp 에 대한 선별 보고서



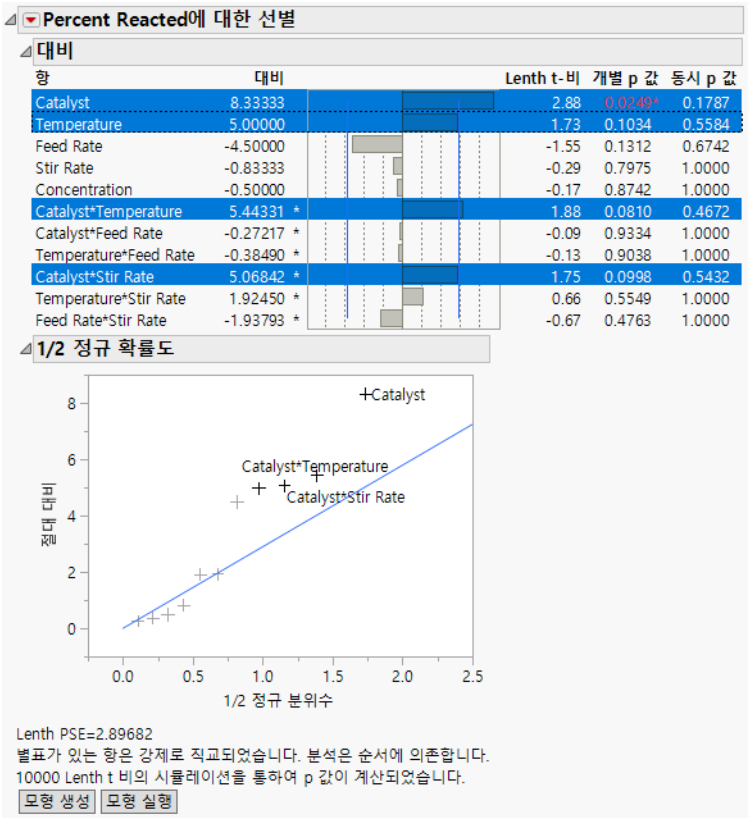
유의한 것으로 식별된 효과는 Polish 뿐입니다. 별표가 표시된 4 개 항은 앞의 효과와 직교하지 않음을 나타냅니다. 이러한 효과에 대해 계산된 대비 값은 직교화 후의 값입니다. 따라서 이러한 추정값은 해당하는 회귀 분석에서 구한 추정값과 일치하지 않습니다. **모형 실행** 버튼을 사용하여 단일 요인 모형을 적합시킬 수 있습니다.

효과 유전성 예

이 예에서는 2 수준 선별 적합 플랫폼에서 효과 유전성을 따르지 않는 모형을 식별할 때 모형을 생성하는 것을 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Plackett-Burman.jmp 를 엽니다 .
2. **DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**을 선택합니다 .
3. Percent Reacted 를 선택하고 **Y** 를 클릭합니다 .
4. Feed Rate ~ Concentration 을 선택하고 **X** 를 클릭합니다 .
5. **확인**을 클릭합니다 .

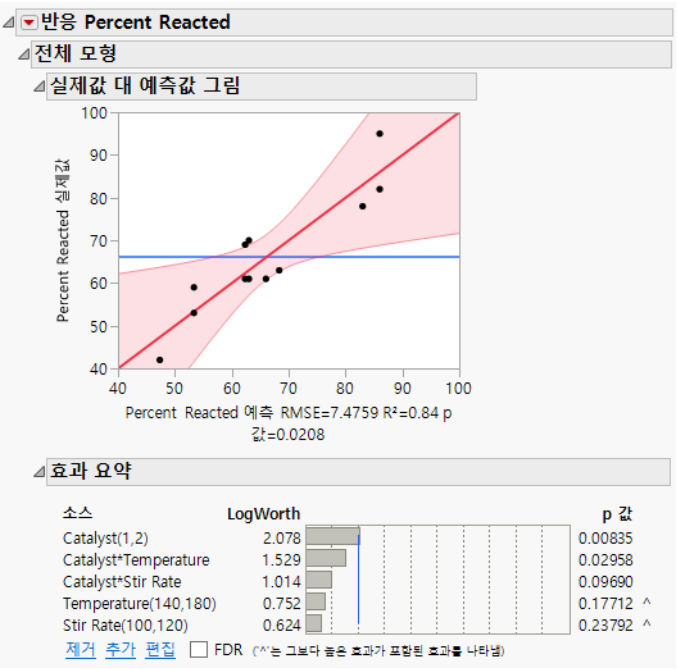
그림 10.7 모형 대비



Catalyst*Stir Rate 교호작용은 강조 표시되어 있는데 Stir Rate 주효과는 강조 표시되지 않습니다 . 효과 유전성 원리에 따라 Stir Rate 주효과를 모형에 추가해야 합니다 . 자세한 내용은 "DOE 로 시작 " 장의 " 효과 유전성 "(61 페이지) 에서 확인하십시오 .

6. **모형 생성**을 클릭합니다 .
7. Stir Rate 를 선택하고 " 모형 효과 생성 " 섹션에서 **추가**를 클릭합니다 .
8. **실행**을 클릭합니다 .

그림 10.8 모형 적합 보고서 (일부분)



Stir Rate 및 Temperature 의 p 값 오른쪽에 있는 ^ 기호는 이러한 효과가 유의하지는 않지만 유의한 고차 효과에 포함되어 있음을 나타냅니다.

과포화 설계 분석

과포화 설계는 런보다 요인이 더 많습니다. 이 설계는 분석할 때 효과 희소성에 크게 의존하므로 2수준 선별 적합 플랫폼이 분석에 이상적입니다. 목표는 활성 효과를 판별하는 것입니다.

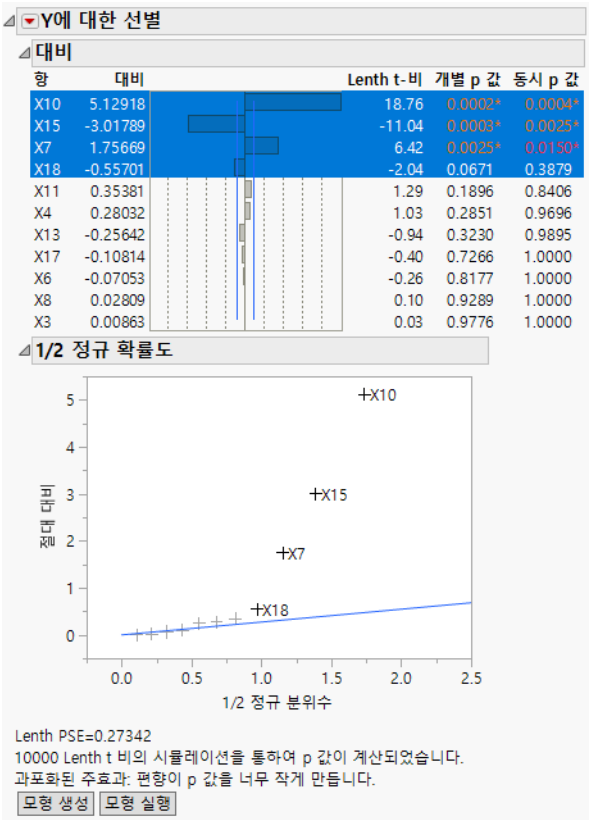
이 예에서는 요인이 18 개이지만 런이 12 회뿐인 시뮬레이션된 설계가 고려됩니다 . Y 는 다음에 의해 생성됩니다 .

$$Y = 2(X7) + 5(X10) - 3(X15) + \varepsilon$$

여기서 $\varepsilon \sim N(0,1)$ 입니다 . 따라서 3 개의 활성 요인으로 Y 가 생성되었습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Supersaturated.jmp 를 엽니다 .
2. **DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. X1 ~ X18 을 선택하고 **X** 를 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 10.9 Supersaturated.jmp 에 대한 선별 보고서



4 개 요인이 강조 표시되어 있습니다. 0.05 임계값 기준에 따라 X10, X15, X7 요인이 활성입니다. "1/2 정규 확률도"에서 X18 은 파란색 선에 가까우며, 이는 유의성에 대한 0.1 경계 값에 가깝다는 것을 나타냅니다. 0.1 임계값은 활성 가능성이 있는 요인이 누락되지 않도록 요인 선택에 관대합니다.

X10, X15, X7 에 대한 대비 추정값 5.1, –3, 1.8 은 시뮬레이션된 값 (5, –3, 2) 에 가깝습니다.

p 값은 직교 설계를 가정하는 시뮬레이션을 기반으로 하므로 통계적으로 완전히 유효하지는 않습니다 (과포화 설계의 경우 해당하지 않음). 그러나 위에 표시된 것처럼 추가 조사를 위한 효과를 식별하는 데 유용할 수 있습니다.

2 수준 선별 적합 플랫폼에 대한 기술 상세 정보

- " 효과 입력 순서 "
- "2 수준 선별 적합과 직교 회전 "
- "Lenth 유사 표준 오차 "

- "Lenth t- 비 "

효과 입력 순서

2 수준 선별 적합 플랫폼에는 다음과 같이 작업 순서가 신중하게 정의되어 있습니다 .

1. 맨 먼저 대비의 절대 크기에 따라 주효과 항이 모형에 입력됩니다 . 모든 효과는 모형에서 앞의 효과와 직교됩니다 . 이 방법을 사용하면 효과 순서가 단계별 전진 회귀의 경우와 동일하게 됩니다 . 또한 주효과를 기준으로 정렬하면 과정 후반에서 기본 별칭 항을 선택하는 데 도움이 됩니다 .
2. 주효과 후에 모든 2 차 교호작용 , 3 차 교호작용 등이 순서대로 모형에 입력됩니다 . B 와 C 가 모두 D 보다 크기 때문에 2 차 교호작용은 이전의 모든 항과 교차한 후 새 항을 가져옵니다 . 전체 입력 순서는 AB, AC, BC, AD, BD, CD 입니다 .
3. 모형에 이미 있는 효과에 대한 정확한 별칭 효과가 " 별칭 " 열에 나타납니다 . 몇 가지 이전 효과의 선형 조합인 효과는 나타나지 않습니다 . 직교성 결여로 인한 부분 별칭 구조가 있는 경우 관련된 효과에 별표가 표시됩니다 .
4. 이 과정은 *n* 개 효과를 얻을 때까지 계속됩니다 . 여기서 *n* 은 데이터 테이블의 행 수이므로 완전히 포화된 모형을 생성합니다 . 요인으로 완전 포화가 불가능한 경우 JMP 는 임의의 직교된 효과를 생성하여 나머지 변동을 흡수합니다 . 이러한 효과에는 Null n 이라는 라벨이 지정되며 여기서 *n* 은 숫자입니다 . 예를 들어 설계에 정확한 반복 행이 있는 경우 이 상황이 발생합니다 .

2 수준 선별 적합과 직교 회전

수학적 개념으로 보면 2수준 선별 적합 플랫폼은 반응 벡터에서 *n* 개 값을 가져와 *n* 개의 새로운 값으로 회전시킵니다. 그런 다음 회전된 값은 요인 및 요인 교호작용의 공간에 의해 매핑됩니다.

대비 = $\mathbf{T}'$ × 반응

위의 방정식에서 **T** 는 *n* 개 값을 얻을 때까지 절편, 요인의 주효과, 이원 교호작용, 삼원 교호작용 등으로 시작하는 *n* x *n* 개의 정규직교화된 값 집합입니다. **T** 의 첫 번째 열이 절편이고 다른 모든 열이 여기에 직교하므로 이러한 다른 열은 모두 대비입니다. 즉, 합이 0 입니다. **T** 는 직교이므로 선형 모형에서 **X** 로 사용할 수 있습니다. $\mathbf{T}'$ 는 $\mathbf{T}^{-1}$ 및 $(\mathbf{T}'\mathbf{T})\mathbf{T}'$ 와 동등하므로 역변환이 필요하지 않습니다. 대비는 선형 모형에서 추정된 모수입니다.

절편을 제외하고 활성인 효과가 모형에 없는 경우 대비는 랜덤 독립 변량을 다른 랜덤 독립 변량으로 직교 회전하는 것에 불과합니다. 직교 회전한 이 랜덤 변량은 원래의 랜덤 독립 변량과 분산이 같습니다. 일부 효과가 활성인 경우에도 비활성 효과는 모형의 오차와 동일한 변동을 나타냅니다. 효과 및 설계가 활성 효과와 랜덤 오차 효과를 분리할 수 있을 만큼 강력해야 합니다.

Lenth 유사 표준 오차

Lenth PSE(유사 표준 오차)라고 하는 Lenth 방법(Lenth 1989)은 비활성으로 보이는 효과를 사용하여 잔차 표준 오차의 추정값을 생성합니다. Lenth PSE를 사용하면 대비가 독립적이고 공통 분산을 갖는 실험의 표준 오차를 추정할 수 있습니다.

n개 행이 있는 경우 이 플랫폼은 $n - 1$개 대비를 생성합니다. 이러한 대비는 $\hat{C}_i$로 나타내고, 여기서 $i = 1, \ldots, n - 1$입니다.

Lenth PSE 를 구하려면 먼저 다음을 계산합니다 .

$$v = 1.5[median_{i = 1, \ldots, n-1}|\hat{C}_i|]$$

Lenth PSE 는 비활성 가능성이 있는 효과를 기반으로 하며 다음과 같이 정의됩니다 .

$$PSE = 1.5\left[median_{|\hat{C}_i| < 2.5v}|\hat{C}_i|\right]$$

Lenth PSE 값은 "선별" 보고서 맨 아래에 표시됩니다.

Lenth t- 비

각 대비에 대해 t 비는 대비를 PSE로 나눈 값으로 계산됩니다. 귀무가설 하에서 이러한 t 비의 기준 분포는 계산적으로 처리하기가 어렵습니다. 따라서 시뮬레이션을 통해 계산됩니다. 아래에 설명된 이 방법은 Ye and Hamada(2000)의 논의를 기반으로 합니다.

i^{th} 대비의 t 비는 다음과 같이 t_i로 나타냅니다 .

$$t_i = \hat{C}_i/(PSE)$$

선별 실험에서 중요한 것은 **개별 오차율**, 즉 특정 효과가 활성이 아닐 때 활성으로 선언할 확률입니다. 개별 오차는 i^{th} 효과의 경우 $|t_i|$가 크고 기준 분포의 상위 꼬리에 속할 때 발생합니다.

이 플랫폼은 비교적 많은 수의 효과를 생성하므로 **실험별 오차율**도 중요합니다. 실험별 오차율은 활성 효과가 없을 때 **어떤 한** 효과를 활성으로 선언할 확률입니다. 실험별 오차는 활성 효과가 없을 경우 절대 t 비의 최대값$(\max|t_i|)$이 크고 기준 분포의 상위 꼬리에 속할 때 발생합니다.

2수준 선별 적합 플랫폼은 몬테카를로 시뮬레이션을 사용하여 두 가지 유형의 오차율에 대한 기준 분포를 얻습니다. 평균이 0이고 표준편차가 PSE와 동일한 정규 분포에서 시뮬레이션된 $n - 1$개 값의 집합을 고려하십시오. 이 값 집합은 활성 효과가 없다는 귀무가설 하에서 실험에 대한 잠재적 대비 값을 나타냅니다. $n - 1$개의 랜덤 대비 값 집합이 총 10,000개 생성됩니다.

개별 p 값

i^{th} 대비를 가정해 보겠습니다. 10,000*$(n - 1)$개의 시뮬레이션 값을 각각 사용하여 Lenth t-비가 생성됩니다. 개별 오차율에 대한 기준 분포는 이러한 t 비의 절대값에 의해 근사값이 계산됩니다. 개별 p 값은 내림차순으로 정렬되어 있는 시뮬레이션된 절대 t 비 10,000*$(n - 1)$개 중 관측된 절대

Lenth t-비의 보간 부분 위치입니다. 이는 기준 분포에 대해 관측된 절대 Lenth t-비 절대값의 오른쪽 영역에 근사한 값을 산출합니다.

동시 p 값

모든 효과가 비활성일 때 t 비로 인해 귀무가설이 기각될 경우 실험별 오차가 발생합니다. 이와 동등하게 절대 t 비의 최대값($\max|t_i|$)으로 인해 귀무가설이 기각될 경우 실험별 오차가 발생합니다.

이 경우 기준 분포를 얻으려면 10,000개의 각 시뮬레이션에서 계산된 절대 t 비의 최대값을 고려하십시오. 10,000개의 이 최대값이 기준 분포를 구성합니다. 동시 p 값은 내림차순으로 정렬되어 있는 시뮬레이션된 절대 t 비 최대값 10,000개 중 관측된 절대 Lenth t-비의 보간 부분 위치입니다. 이는 시뮬레이션된 절대 t 비의 최대값을 기반으로 기준 분포에 대한 절대 Lenth t-비 절대값의 오른쪽 영역에 근사한 값을 산출합니다.

몬테카를로 시뮬레이션 옵션

기본 시뮬레이션 집합 수(10,000)를 변경하려면 `LenthSimN`이라는 전역 JSL 변수에 값을 할당해야 합니다. p 값을 도출하는 데 사용된 시뮬레이션 수가 보고서 창에 보고됩니다.

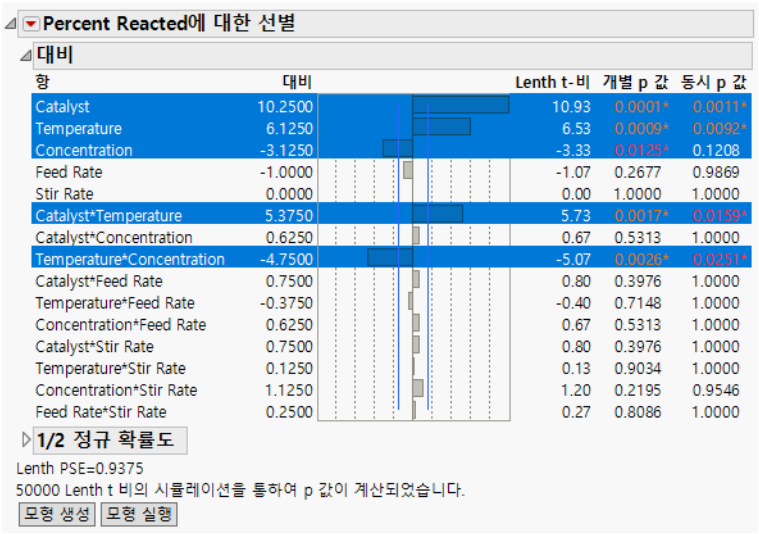
다음 예에서는 시뮬레이션 수를 50,000 으로 설정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor Half Fraction.jmp 를 엽니다 .
2. **DOE > 전통적 설계 > 2 수준 선별 > 2 수준 선별 적합**을 선택합니다 .
3. Percent Reacted 를 선택하고 **Y** 를 클릭합니다 .
4. Feed Rate ~ Concentration 을 선택하고 **X** 를 클릭한 후 **확인**을 클릭합니다 .
5. "Percent Reacted 에 대한 선별 " 옆의 빨간색 삼각형을 클릭하고 **스크립트 저장 > 스크립트 창에**를 선택합니다 .
6. 스크립트의 첫 번째 줄에 **LenthSimN=50000;** 을 삽입합니다 .

```
LenthSimN=50000;
Fit Two Level Screening(
   Y( :Percent Reacted),
   X( :Feed Rate, :Catalyst, :Stir Rate, :Temperature, :Concentration )
);
```

7. 스크립트 창을 마우스 오른쪽 버튼으로 클릭하고 **스크립트 실행**을 선택합니다 .

그림 10.10 *p* 값 도출을 위해 N = 50,000 시뮬레이션을 사용한 2 수준 적합



시뮬레이션 수가 보고서 창에 명시됩니다 .

참고 : LenthSimN=0 이면 표준 *t* 분포가 사용되고 동시 *p* 값은 제공되지 않습니다 . 이 방법은 권장되지 않습니다 .

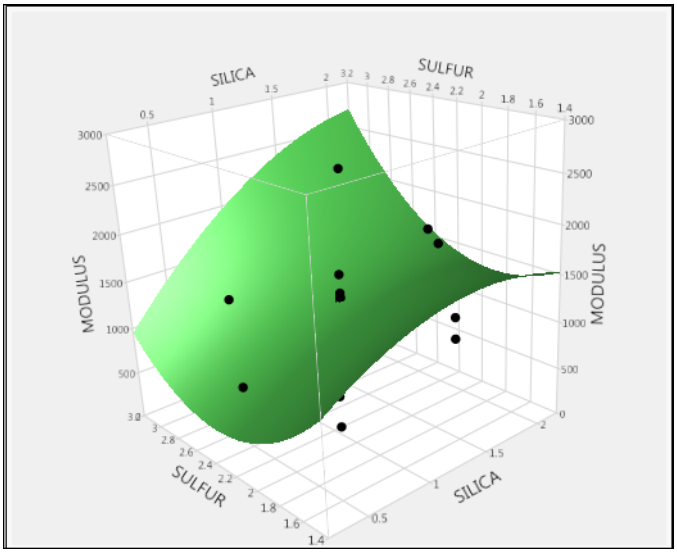
11 장

반응 표면 설계

반응 표면 설계는 2차 표면을 모델링하는 데 유용합니다. 반응 표면 모형은 설계 영역 내에서 반응의 최소값 또는 최대값이 발생하는 점을 식별할 수 있습니다. 2차 함수를 적합시키려면 각 요인에 대해 3개의 구분되는 값이 필요하므로 표준 2수준 설계는 곡면을 적합시키는 데 적절하지 않습니다.

반응 표면 설계는 반응에 대한 2차 예측 방정식을 적합시킬 수 있습니다. 이러한 방정식의 2차 항은 실제 반응 함수에서 곡률을 모델링합니다. 설계 영역 내에 최대값 또는 최소값이 있는 경우 해당 값이 발생하는 점을 추정할 수 있습니다.

그림 11.1 반응 표면 설계 결과 모형



목차

반응 표면 설계 개요 329
반응 표면 설계의 예 331
Box-Behnken 설계 구성 332
실험 데이터 분석 333
최적 설정 탐색 335
반응 표면 설계 창 338
반응 339
요인 340
설계 선택 342
출력 옵션 지정 343
테이블 생성 344
반응 표면 설계 옵션 345

반응 표면 설계 개요

반응 표면 설계 플랫폼은 중심 합성 설계와 Box-Behnken 설계라는 전통적 설계를 제공합니다(블록화된 설계 포함). 중심 합성 설계의 경우 축 점 배치 및 설계의 여러 측면을 제어할 수 있습니다. 반응 표면 설계는 연속형 요인에만 사용할 수 있으며 최대 8개 요인에 대해 제공됩니다.

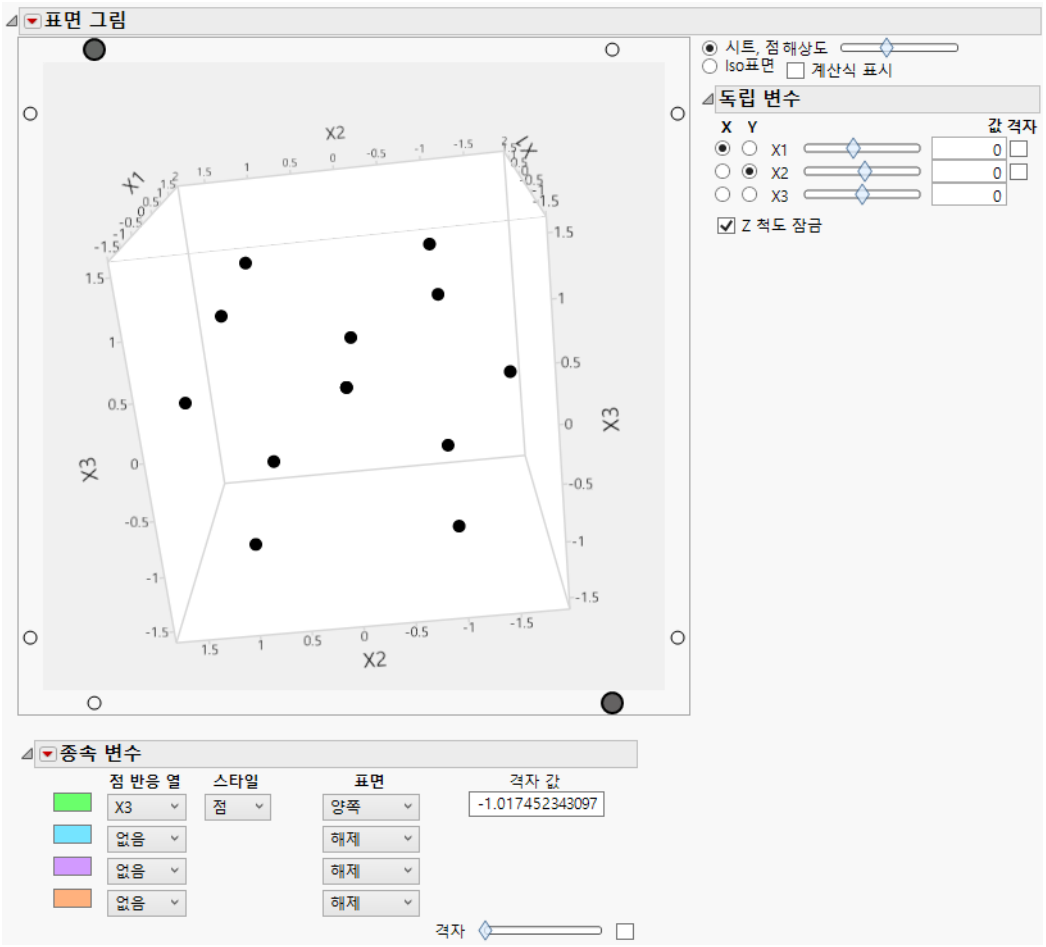
팁 : "DOE > 사용자 설계 " 를 사용하여 특정 실험 상황에 맞는 최적의 반응 표면 설계를 구성할 수 있습니다 . 사용자 설계는 전통적 반응 표면 설계보다 훨씬 유연한 반응 표면 설계를 구성합니다 . 특히 사용자 설계 플랫폼을 사용하면 범주형 요인 또는 9 개 이상의 연속형 요인이 포함된 반응 표면 설계를 생성할 수 있습니다 . 런 수 및 설계 공간에 대한 제한을 지정할 수도 있습니다 . 자세한 내용은 " 사용자 설계의 예 " 장의 " 반응 표면 실험 "(144 페이지) 에서 확인하십시오 .

중심 합성 설계 (그림 11.2) 는 2 수준 부분 요인 설계와 다음과 같은 두 가지 다른 유형의 점을 결합합니다 .

- **중앙점** . 모든 요인 값이 범위의 중앙 값으로 설정됩니다 .
- **축 점** . 한 요인은 상한 또는 하한값 (**축** 값) 으로 설정되고 다른 모든 요인은 범위의 중앙 값으로 설정됩니다 .

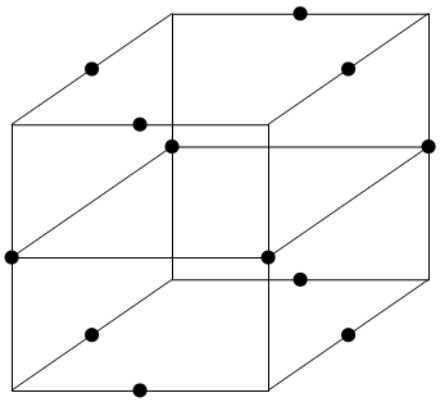
축 점을 기준으로 한 선택에 따라 중심 합성 설계는 각 요인에 대해 최대 5개의 고유 설정을 가질 수 있으며 축 점은 지정된 요인 범위를 넘어 확장될 수 있습니다.

그림 11.2 세 가지 요인에 대한 중심 합성 설계



Box-Behnken 설계(그림 11.3)에 요인당 수준이 세 개뿐이며 요인 범위에 의해 정의된 입방체의 꼭지점에 설계점이 없습니다. 이 유형의 설계는 엔지니어링 고려 사항으로 인해 설계점을 사용하지 않아야 할 때 유용할 수 있습니다. 그러나 입방체 꼭지점에 설계점이 없다는 것은 Box-Behnken 설계가 중심 합성 설계에 비해 꼭지점 근처에서 예측 분산이 높고 정밀도가 낮다는 의미입니다.

그림 11.3 세 가지 요인에 대한 Box-Behnken 설계



JMP 에서 다음 두 가지 방법으로 반응 표면 설계를 구성할 수 있습니다 .

- 반응 표면 설계 플랫폼 사용 (최대 8 개의 연속형 요인)
- 사용자 설계 플랫폼 사용 후 " 모형 " 개요에서 "RSM" 버튼 클릭

두 경우 모두 모형 적합을 위해 실행할 수 있는 "모형" 스크립트가 설계 테이블에 포함됩니다 . "모형" 스크립트는 "반응 표면 효과" 속성을 각 주효과에 적용하여 "모형 적합" 창에서 주효과 뒤에 **&RS** 가 표시되도록 합니다 . 이 속성을 사용하면 "최소 제곱 적합" 보고서에 "반응 표면" 보고서가 포함됩니다 . 이 보고서에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

참고 : 연속형 요인이 20 개 넘게 포함된 반응 표면 설계의 경우 " 표준 최소 제곱 " 보고서에 " 반응 표면 " 개요가 표시되지 않습니다 .

반응 표면 설계의 예

이 예에서는 타이어 트레드 실험을 위한 Box-Behnken 설계를 구성합니다 . 연장 (Stretch) 측정을 위해 목표값 450 과 일치시키는 것이 실험 목표입니다 . 연장 값은 타이어 트레드 컴파운드를 제조하는 데 사용되는 Silica, Silane 및 Sulfur 양에 따라 달라집니다 . 목표값을 달성하는 설정을 찾기 위해 광범위한 요인 설정을 대상으로 실험하려고 합니다 .

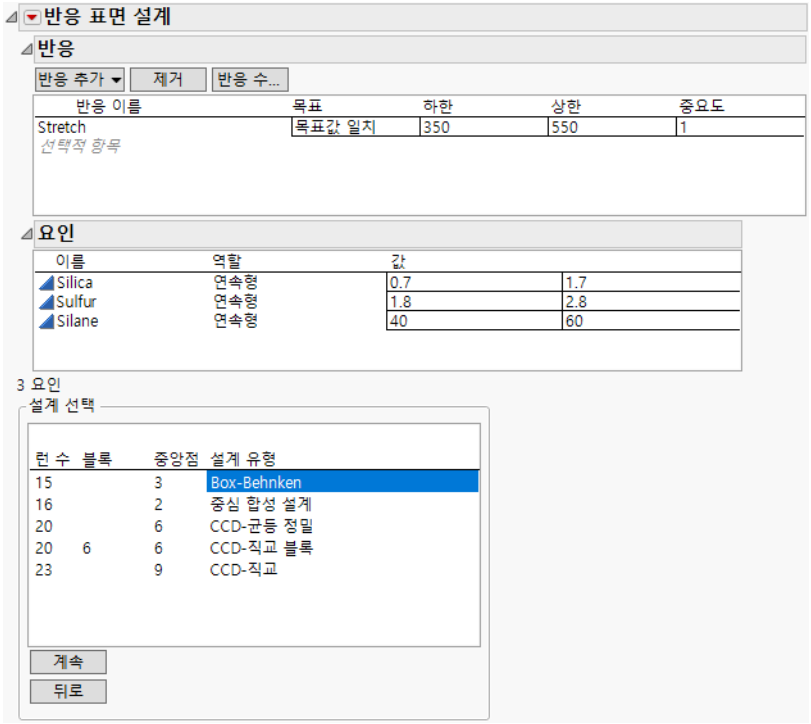
- "Box-Behnken 설계 구성 "
- " 실험 데이터 분석 "
- " 최적 설정 탐색 "

Box-Behnken 설계 구성

편의상 이 예에서는 기존 테이블의 반응과 요인을 불러옵니다 . 새 실험을 직접 설계할 때는 반응과 요인을 수동으로 입력합니다 . 자세한 내용은 " 반응 "(339 페이지) 및 " 요인 "(340 페이지) 에서 확인하십시오 .

1. **DOE > 전통적 설계 > 반응 표면 설계**를 선택합니다 .
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Bounce Response.jmp 를 엽니다 .
3. " 반응 표면 설계 " 의 빨간색 삼각형을 클릭하고 **반응 불러오기**를 선택합니다 .
4. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Bounce Factors.jmp 를 엽니다 .
5. " 반응 표면 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

그림 11.4 타이어 트레드 설계를 위한 반응 및 요인 개요



" 반응 " 개요에 Stretch 의 목표가 " 목표값 일치 " 로 설정되어 있습니다 .

" 설계 선택 " 패널에 가능한 설계가 나타납니다 .

참고 : 6 단계에서 " 난수 시드값 " 을 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이 단계가 필요하지 않습니다 .

6. (선택 사항) " 반응 표면 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 . "12345" 를 입력하고 **확인**을 클릭합니다 .
7. **계속**을 클릭하여 Box-Behnken 설계 선택을 유지합니다 .
8. **테이블 생성**을 클릭합니다 .

그림 11.5 Box-Behnken 설계 테이블

	패턴	Silica	Sulfur	Silane	Stretch
1	−0+	0.7	2.3	60	•
2	0+−	1.2	2.8	40	•
3	000	1.2	2.3	50	•
4	000	1.2	2.3	50	•
5	+0−	1.7	2.3	40	•
6	+0+	1.7	2.3	60	•
7	0−+	1.2	1.8	60	•
8	0++	1.2	2.8	60	•
9	−0−	0.7	2.3	40	•
10	−+0	0.7	2.8	50	•
11	000	1.2	2.3	50	•
12	−−0	0.7	1.8	50	•
13	0−−	1.2	1.8	40	•
14	+−0	1.7	1.8	50	•
15	++0	1.7	2.8	50	•

이제 실험을 수행하고 데이터 테이블에 반응을 입력합니다.

실험 데이터 분석

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Bounce Data.jmp 를 엽니다 .

   Bounce Data.jmp 파일에 실험 결과가 포함되어 있습니다 .
2. **Model** 스크립트를 실행합니다 .

   " 모형 효과 생성 " 목록의 주효과 뒤에 **& RS** 가 표시됩니다 . 이 접미사는 해당 주효과가 " 표준 최소 제곱 " 보고서에 " 반응 표면 " 보고서를 생성하는 반응 표면 효과임을 나타냅니다 .
3. **실행**을 클릭합니다 .

그림 11.6 적합 결여 및 효과 검정 보고서

적합 결여

소스	DF	제곱합	평균 제곱	F 비
적합 결여	3	11.750000	3.91667	0.9792
순수 오차	2	8.000000	4.00000	Prob > F
총 오차	5	19.750000		0.5411
				최대 R²
				0.9999

모수 추정값

효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F
Silica(0.7,1.7)	1	1	12880.125	3260.791	<.0001*
Sulfur(1.8,2.8)	1	1	5778.125	1462.816	<.0001*
Silane(40,60)	1	1	968.000	245.0633	<.0001*
Silica*Sulfur	1	1	52441.000	13276.20	<.0001*
Silica*Silane	1	1	0.250	0.0633	0.8114
Sulfur*Silane	1	1	8556.250	2166.139	<.0001*
Silica*Silica	1	1	2592.923	656.4362	<.0001*
Sulfur*Sulfur	1	1	4653.231	1178.033	<.0001*
Silane*Silane	1	1	0.231	0.0584	0.8186

Box-Cox 변환

적합 결여를 나타내는 징후가 없고 " 효과 검정 " 보고서에 따르면 두 개의 고차항 (Silica* Silane, Silane*Silane) 을 제외한 모든 항의 p 값이 0.0001 미만입니다 . 그림 11.6 에 있는 테이블의 해석에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

4. " 반응 표면 " 옆의 표시 아이콘을 클릭하여 보고서를 엽니다 .
5. " 정준 곡률 " 옆의 표시 아이콘을 클릭합니다 .

그림 11.7 반응 표면 보고서

반응 표면

계수

	Silica(0.7,1.7)	Sulfur(1.8,2.8)	Silane(40,60)	Stretch
Silica(0.7,1.7)	26.5	114.5	-0.25	-40.125
Sulfur(1.8,2.8)	.	-35.5	46.25	-26.875
Silane(40,60)	.	.	0.25	11

해

변수	임계값
Silica(0.7,1.7)	1.7912411
Sulfur(1.8,2.8)	2.1986422
Silane(40,60)	23.424426

해: 안장점

임계값이 데이터 범위를 벗어납니다.

해의 예측값 360.38388

정준 곡률

고유값 및 고유 벡터

고유값	62.9095	3.2989	-74.9584
Silica(0.7,1.7)	0.82779	-0.29879	-0.47486
Sulfur(1.8,2.8)	0.52687	0.12315	0.84097
Silane(40,60)	0.19280	0.94634	-0.25937

보고서의 첫 번째 부분으로 표시된 " 계수 " 테이블에는 추정된 모형 모수가 간략하게 요약되어 있습니다 . 처음 세 열은 2 차 항의 계수를 제공합니다 . 마지막 열은 선형 항의 계수를 제공합니다 . 예측 표현식을 전체적으로 표시하려면 " 반응 Stretch" 의 빨간색 삼각형에서 **추정값** > **예측 표현식 표시**를 선택합니다 .

" 해 " 보고서는 단일 임계값이 발생하는 점의 좌표를 제공합니다 . 이 경우 해당 점은 안장점 (최대값도 아니고 최소값도 아닌 점) 이며 설계 공간 범위를 벗어납니다 .

" 정준 곡률 " 보고서는 효과의 고유값 및 고유 벡터를 표시합니다 . 이러한 값은 표면 곡률의 특성과 방향에 대한 정보를 제공합니다 . 큰 양의 고유값 62.9095 는 양의 곡률을 나타내고 고유 벡터 값은 곡률이 주로 Silica 방향에 있음을 나타냅니다 . 큰 음의 고유값 -74.9584 는 음의 곡률을 나타내고 고유 벡터 값은 곡률이 주로 Sulfur 방향에 있음을 나타냅니다 .

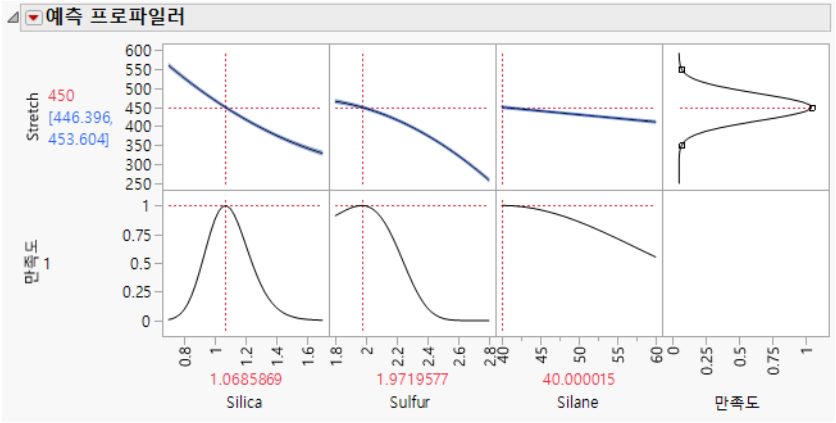
그림 11.7 에 있는 반응 표면 분석 테이블에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

이제 예측 프로파일러 및 등고선 프로파일러를 사용하여 최적 설정을 찾습니다.

최적 설정 탐색

1. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 11.8 만족도가 최대화된 Bounce Data.jmp 의 예측 프로파일러



참고 : 그림에 표시된 것과 최적 설정이 다를 수 있습니다 . 예측 Stretch 가 450 인 점이 많이 있기 때문입니다 .

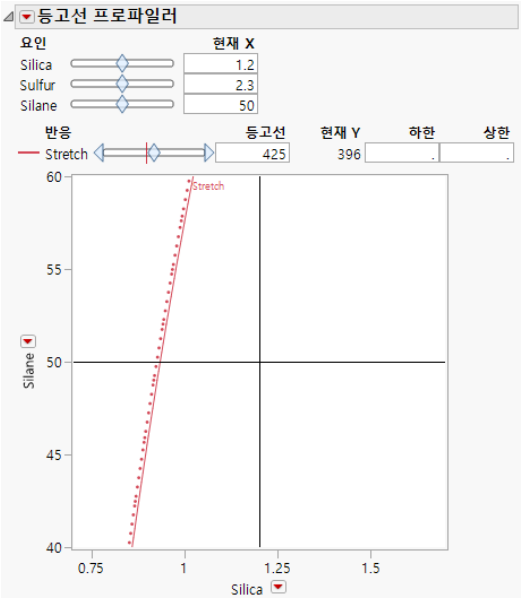
반응을 지정할 때 목표는 하한과 상한이 각각 350, 550 인 상태에서 목표값 (450) 일치로 설정되었습니다 . 이 목표는 설계 테이블로 이어지고 한계 값은 Stretch 의 " 반응 한계 " 열 특성에 입력되었습니다 . 이러한 반응 한계에서 만족도 함수가 구성됩니다 (그림 11.8 의 오른쪽 위 셀). 자세한 내용은 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

만족도 함수를 최대화할 때 JMP 는 예측 Stretch 가 450 이 되는 가능한 여러 요인 수준 설정 조합 중에서 하나를 식별합니다 . 그림 11.8 에서는 이러한 설정을 Silica = 1.069, Sulfur = 1.972, Silane = 40.000 으로 나타냅니다 . 이제 " 등고선 프로파일러 " 를 사용하여 만족도 함수를 최대화하는 다른 점을 식별합니다 .

예측 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

2. " 반응 Stretch" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 등고선 프로파일러** 를 선택합니다 . Sulfur 값을 2.0 으로 설정하고 목표값을 달성하려는 경우를 가정해 보겠습니다 . 또한 Silane 및 Silica 에 대해 선택한 설정이 예측 Stretch 를 450(5 단위) 이내에 유지하도록 해야 합니다 .
3. 세로 축에서 Sulfur 옆의 빨간색 삼각형을 클릭하고 Silane 을 선택합니다 .

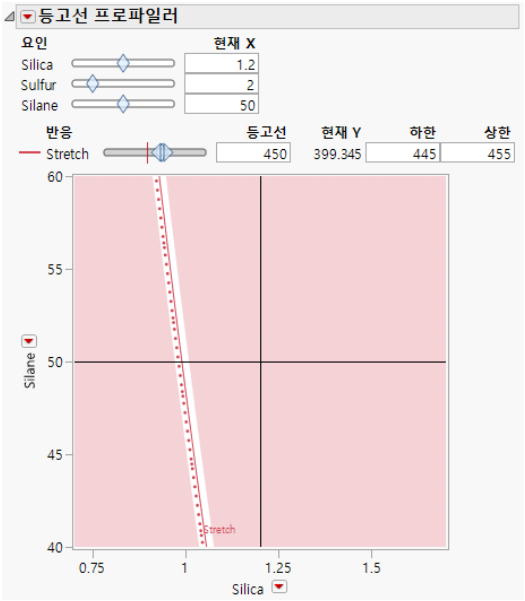
그림 11.9 Bounce Data.jmp 의 등고선 프로파일러



이 그림은 Stretch 가 425 이고 Sulfur 가 2.3 인 경우 Silane 및 Silica 값의 등고선을 보여 줍니다.

4. Sulfur 의 **현재 X** 를 2 로 설정합니다 .
5. Stretch 의 **등고선**을 450 으로 설정합니다 .
6. Stretch 의 **하한** 및 **상한**을 각각 445 와 455 로 설정합니다 . Enter 키를 누릅니다 .

그림 11.10 Silica 및 Silane 의 최적 설정을 보여 주는 등고선 프로파일러

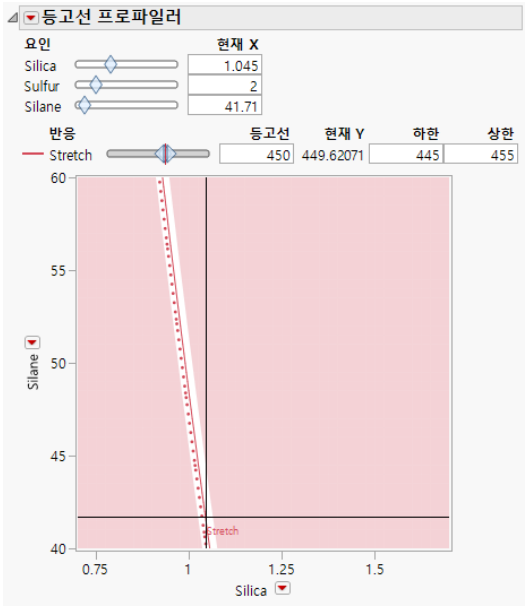


Silica 및 Silane 값의 음영 처리되지 않은 띠는 Sulfur 가 2.0 으로 설정된 경우 445 에서 455 사이의 예측 Stretch 를 제공합니다 . 단색의 빨간색 곡선 위의 값은 예측 Stretch 값 450 을 나타냅니다 .

7. 그림에 나타나는 십자기호를 음영 처리되지 않은 띠로 드래그하여 실제적 관점에서 공정에 가장 적합한 Silica 및 Silane 설정을 찾습니다 .

   공정은 높은 수준보다 낮은 수준의 Silane 에서 더 로버스트하다고 가정해 보겠습니다 . 그러면 그림 11.11 의 설정을 고려할 수 있습니다 .

그림 11.11 Silica 및 Silane 의 특정 설정을 보여 주는 등고선 프로파일러



Sulfur = 2.0 의 경우 십자기호로 식별된 요인 설정은 Silica = 1.045, Silane = 41.71 입니다. 이 설정은 "현재 X" 아래에 표시됩니다. 이 설정에서 예측 Stretch 는 449.62071 이며 "현재 Y" 옆에 표시됩니다.

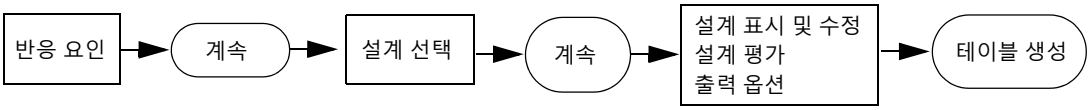
등고선 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오.

반응 표면 설계 창

"반응 표면 설계" 창은 2 차 표면을 모델링하기 위한 설계를 구성하는 단계를 안내합니다. 중심 합성 설계, Box-Behnken 설계 또는 이 두 설계 유형 중 하나의 블록화된 버전을 선택할 수 있습니다. 중심 합성 설계를 선택하면 축 점을 조정할 수 있습니다.

설계 단계를 진행하는 동안 "반응 표면 설계" 창이 업데이트됩니다. 개요는 개요를 업데이트하는 버튼을 클릭할 때 구분되어 나타나며 그림 11.12 의 흐름을 따릅니다.

그림 11.12 반응 표면 설계 흐름

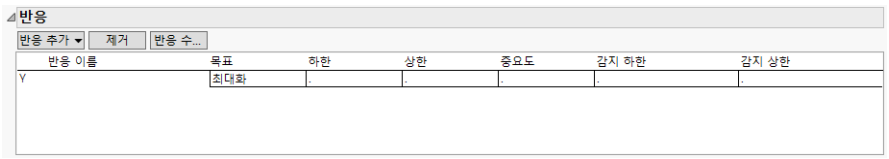


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다 . 이 옵션을 사용하면 반응 이름 , 목표 , 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다 .

그림 11.13 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다 . " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다 . " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다 .

JMP PRO 함수 (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다 . 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다 .

제거 선택한 반응을 제거합니다 .

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다 . 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다 . 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다 . 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다 .

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응 열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .

- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .
- " 없음 " 목표는 최적화에 관한 목표가 없음을 나타냅니다 . 만족도 함수가 구성되지 않습니다 .

참고 : 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다 . 데이터 테이블에서 반응 열의 " 열 정보 " 창을 열고 (**열 > 열 정보**) 원하는 목표값을 입력하십시오 .

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다 . 이러한 값은 반응에 대한 가중치로 처리됩니다 . 반응이 하나뿐이면 기본적으로 1 로 설정되므로 중요도를 지정할 필요가 없습니다 .

JMP PRO **감지 한계** 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다 . 감지 한계는 열 특성으로 저장됩니다 . 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오 .

반응 개요 편집

" 반응 " 개요에서 다음 사항에 유의하십시오 .

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다 .
- 목표를 클릭하여 변경할 수 있습니다 .
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다 .
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다 .

반응 한계 열 특성

반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

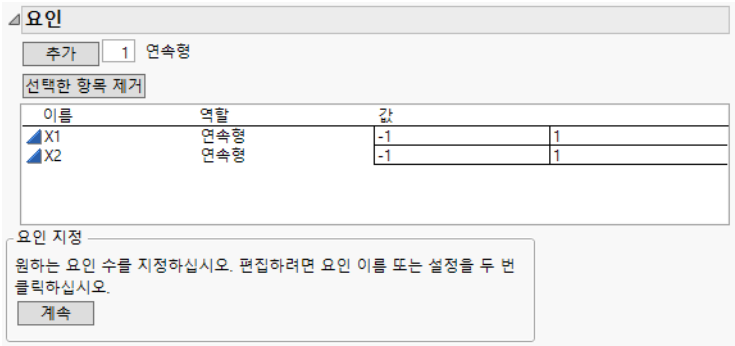
요인

반응 표면 설계의 요인은 연속형만 가능합니다.

팁 : 범주형 요인을 포함하는 반응 표면 설계를 생성하려면 **DOE > 사용자 설계**를 사용하십시오 .

반응 표면 설계의 초기 "요인" 패널에는 두 가지 연속형 요인이 나타납니다.

그림 11.14 요인 개요



" 요인 " 개요에는 다음과 같은 버튼이 포함되어 있습니다 .

추가 지정된 수의 연속형 요인을 입력합니다 .

선택한 항목 제거 선택한 요인을 제거합니다 .

팁 : " 요인 " 패널을 완료하면 빨간색 삼각형 메뉴에서 " 요인 저장 " 을 선택하십시오 . 이렇게 하면 요인 이름과 값을 데이터 테이블에 저장하여 나중에 다시 불러올 수 있습니다 . 자세한 내용은 " 반응 표면 설계 옵션 "(345 페이지) 에서 확인하십시오 .

" 요인 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할을 " 연속형 " 으로 지정합니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인 유형이 적절하게 모델링되도록 합니다 .

값 요인에 대한 실험 설정입니다. 값을 삽입하려면 기본값을 클릭한 후 원하는 값을 입력합니다.

요인 열 특성

각 요인에 대한 다양한 열 특성이 완료된 설계의 데이터 테이블에 저장됩니다 .

설계 역할 각 요인에 " 설계 역할 " 열 특성이 할당됩니다 . 요인을 정의할 때 지정하는 역할에 따라 " 설계 역할 " 열 특성의 값이 결정됩니다 . 블록이 있는 설계를 선택하면 해당 요인에 " 블록화 " 값이 할당됩니다 . " 설계 역할 " 특성은 실험 데이터를 모델링할 때 요인이 어떤 용도로 사용되는지를 반영합니다 . " 설계 역할 " 값은 설계 확대 플랫폼에서 사용됩니다 .

요인 변경 각 요인에 " 쉬움 " 으로 설정된 " 요인 변경 " 열 특성이 할당됩니다 . 반응 표면 설계 플랫폼에서는 각 실험 런에 대해 요인 수준을 변경할 수 있다고 가정합니다 . " 요인 변경 " 값은 설계 평가 및 설계 확대 플랫폼에서 사용됩니다 .

코딩 "설계 역할"이 "연속형"인 경우 요인의 "코딩" 열 특성이 저장됩니다. 이 특성은 하한값과 상한값이 각각 –1 과 +1 에 해당하도록 요인 값을 변환합니다. "최소 제곱 적합" 보고서의 추정값과 검정은 변환된 값을 기반으로 합니다.

블록당 런 수 각 블록의 런 수를 나타냅니다. 블록이 있는 설계를 선택하고 "테이블 생성"을 클릭하면 기본 이름이 "블록"인 요인이 "요인" 목록에 추가됩니다. "블록당 런 수" 열 특성은 해당 요인에 대해 저장됩니다.

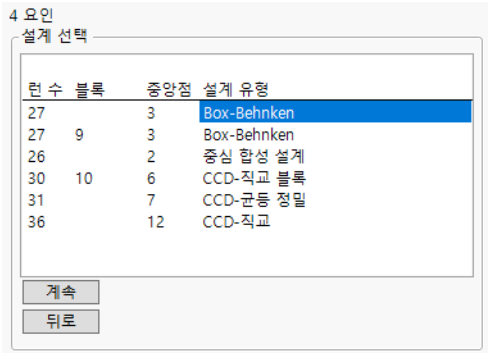
설계 선택

반응과 요인을 입력하고 **계속**을 클릭한 후 설계 목록에서 선택합니다. 설계에는 다음 두 가지 유형이 포함됩니다.

- "Box-Behnken 설계"(342 페이지)
- "중심 합성 설계"(342 페이지)

사용할 설계를 선택하고 **계속**을 클릭합니다.

그림 11.15 네 가지 요인에 대한 설계 선택 패널



Box-Behnken 설계

Box-Behnken 설계에 요인당 수준이 세 개뿐이며 요인 범위에 의해 정의된 입방체의 꼭지점에 설계점이 없습니다. 이러한 설계는 엔지니어링 고려 사항으로 인해 극단적 설정을 피해야 할 때 유용할 수 있습니다. 그러나 이 설계는 중심 합성 설계에 비해 꼭지점 근처에서 예측 분산이 더 높습니다.

중심 합성 설계

중심 합성 설계에는 중앙점과 축 점이 있습니다. **축 점**은 한 요인은 상한 또는 하한값(**축 값**)으로 설정되고 다른 모든 요인은 범위의 중앙 값으로 설정되는 점입니다.

중심 합성 설계에는 지정된 요인 범위에 의해 정의된 초입방체의 면을 벗어나는 축 점이 있을 수 있습니다. 즉, 각 요인에는 "요인" 개요에 지정된 값 범위를 벗어나는 두 가지 설정을 포함하여 5 개의 고유 설정이 필요할 수 있습니다. 그러나 JMP를 사용하면 면에 설계점을 놓을 수 있습니다.

다음과 같은 유형의 중심 합성 설계를 사용할 수 있습니다.

중심 합성 설계 지정된 수의 요인에 대한 일반적인 중심 합성 설계입니다.

CCD- 균등 정밀 설계 공간의 중심 근처에서 예측 분산이 매우 고르게 되도록 중앙점 수가 선택됩니다.

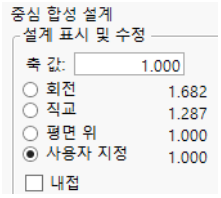
CCD- 직교 2 차 모수 추정값이 다른 모수 추정값과 최소로 상관되도록 중앙점 수와 축 값이 선택됩니다.

CCD- 직교 블록 2 차 모수 추정값 및 블록 효과가 다른 모수 추정값과 최소로 상관됩니다.

축 값

중심 합성 설계를 선택한 후 **계속**을 클릭하면 축 크기 조정 정보를 제공하는 옵션을 사용할 수 있습니다. 표시된 값은 축 값을 놓을 때 지정된 요인 범위의 1/2을 곱하는 데 사용됩니다. "축 값" 옆에 1.0을 지정하면 결과 설계의 축 점은 요인에 의해 정의된 입방체의 면에 놓입니다. 다음 옵션에 따라 축 값을 설정할 수 있습니다.

그림 11.16 축 값 패널



회전 예측 분산은 설계 중심과의 척도화된 거리에만 의존합니다. 축 점은 요인 범위보다 더 극단적입니다. 이 요인 범위를 실제로 달성할 수 없는 경우 **평면 위**를 선택하거나 값을 직접 지정하십시오.

직교 효과가 직교합니다. 축 점은 요인 범위보다 더 극단적입니다. 이 요인 범위를 실제로 달성할 수 없는 경우 **평면 위**를 선택하거나 값을 직접 지정하십시오.

평면 위 축 점을 지정된 요인 범위의 극단에 놓습니다.

사용자 지정 "축 값" 텍스트 상자에 입력한 값으로 지정된 거리에 축 점을 놓습니다.

내접 축 점이 요인 범위의 아래쪽 및 위쪽 끝에 놓이도록 설계의 척도를 재조정합니다. 요인 설계점은 이 척도에 따라 축소됩니다.

출력 옵션 지정

"출력 옵션" 패널에서 출력 데이터 테이블에 대한 상세 정보를 지정할 수 있습니다. 끝나면 **테이블 생성**을 클릭하여 설계에 대한 데이터 테이블을 구성합니다.

런 순서

런 순서 옵션은 설계 테이블의 런 순서를 결정합니다. 다음과 같은 옵션을 선택할 수 있습니다.

동일하게 유지 설계 테이블에 표시되는 행 순서가 " 설계 " 및 " 예상 계수 " 개요의 순서와 동일합니다.

좌에서 우로 정렬 설계 테이블의 열이 왼쪽에서 오른쪽으로 정렬됩니다.

랜덤화 설계 테이블의 행이 랜덤 순서로 나타납니다.

우에서 좌로 정렬 설계 테이블의 열이 오른쪽에서 왼쪽으로 정렬됩니다.

블록 내 랜덤화 설계 테이블의 행이 블록 내에서 랜덤 순서로 나타납니다.

중앙점 및 반복 실험

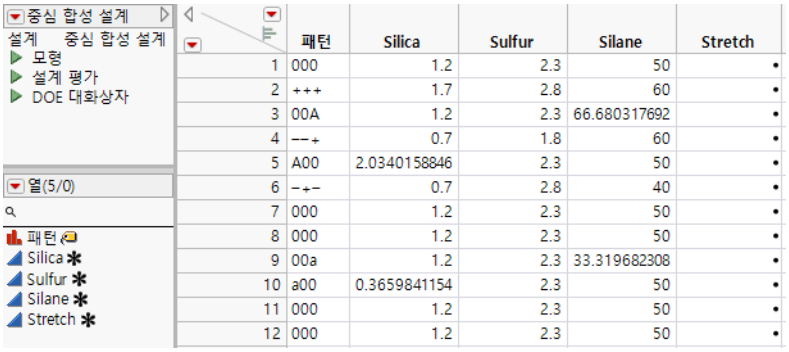
중앙점 수 설계에 나타나는 중앙점의 수입니다. 중앙점은 모든 연속형 요인이 요인 범위의 가운데에 설정되는 런입니다. 표시된 초기값은 선택한 설계의 중앙점 수입니다.

반복 실험 수 중앙점을 포함하여 전체 설계를 반복할 횟수입니다. 한 번 반복하면 런 수가 두 배가 됩니다.

테이블 생성

테이블 생성을 클릭하여 실험 런이 포함된 설계 테이블을 생성합니다.

그림 11.17 Bounce 요인 및 반응에 대한 직교 중심 합성 설계


중심 합성 설계
설계 중심 합성 설계
모형
설계 평가
DOE 대화상자
열(5/0)
패턴
Silica ✱
Sulfur ✱
Silane ✱
Stretch ✱

	패턴	Silica	Sulfur	Silane	Stretch
1	000	1.2	2.3	50	•
2	+++	1.7	2.8	60	•
3	00A	1.2	2.3	66.680317692	•
4	−−+	0.7	1.8	60	•
5	A00	2.0340158846	2.3	50	•
6	−+−	0.7	2.8	40	•
7	000	1.2	2.3	50	•
8	000	1.2	2.3	50	•
9	00a	1.2	2.3	33.319682308	•
10	a00	0.3659841154	2.3	50	•
11	000	1.2	2.3	50	•
12	000	1.2	2.3	50	•

왼쪽 위의 테이블 패널에 있는 **설계** 노트는 테이블을 생성한 설계 유형을 나타냅니다(중심 합성 설계). 이 정보는 여러 설계를 비교하는 경우 유용할 수 있습니다.

패턴 열

패턴 열은 요인 값에 관한 각 행의 런에 대한 설명을 기호로 표시합니다.

팁 : 패턴은 그림에서 라벨 변수로 유용할 수 있습니다.

표 11.1 패턴 열 설명

-	하한값
+	상한값
0	범위의 중앙 값
a	축 하한값
A	축 상한값

설계 테이블 스크립트

설계 테이블에는 다음과 같은 스크립트가 포함되어 있습니다 .

모형 **분석 > 모형 적합** 플랫폼을 실행합니다 .

설계 평가 "DOE > 설계 진단 > 설계 평가 " 플랫폼을 실행합니다 .

DOE 대화상자 설계 테이블을 생성하는 데 사용된 " 반응 표면 설계 " 창을 다시 생성합니다 . 또한 이 스크립트는 설계를 생성하는 데 사용된 난수 시드값도 포함합니다 .

반응 표면 설계 옵션

반응 표면 설계 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

반응 저장 " 반응 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 . 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다 .

반응 불러오기 " 반응 저장 " 옵션을 사용하여 저장한 반응을 불러옵니다 .

요인 저장 " 요인 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 각 요인의 열에는 수준이 포함됩니다 . 다른 정보는 열 특성으로 저장됩니다 . 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다 . 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다 . 그런 다음 " 설계 역할 " 영역에서 적절한 역할을 선택합니다 .

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 불러옵니다 .

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) " 요인 제약 조건 정의 " 또는 " 선형 제약 조건 " 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다 . 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다 . 마지막 행은 부등식 한계를 포함합니다 . 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 제약 조건 상태 "(762 페이지) 에서 확인하십시오 .

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) " 제약 조건 저장 " 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다 .

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다 . 이러한 작업에는 다음 중 하나 이상이 포함됩니다 .

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다 .

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

행렬 X 저장 " 적률 행렬 " 및 " 모형 행렬 " 을 설계 데이터 테이블의 테이블 스크립트에 저장합니다 . 이 스크립트에는 적률 및 설계 행렬이 포함됩니다 . 자세한 내용은 " 사용자 설계 " 장의 " 행렬 X 저장 "(105 페이지) 에서 확인하십시오 .

주의 : 명목형 요인이 있는 설계의 경우 " 행렬 X 저장 " 옵션으로 저장한 모형 행렬은 선형 모형을 적합시킬 때 사용된 코딩 행렬이 **아닙니다** . " 모형 " 스크립트를 실행할 때 나타나는 " 모형 적합 " 보고서에서 " 열 저장 > 코드화 테이블 저장 " 옵션을 선택하면 모형 적합에 사용되는 코딩 행렬을 가져올 수 있습니다 .

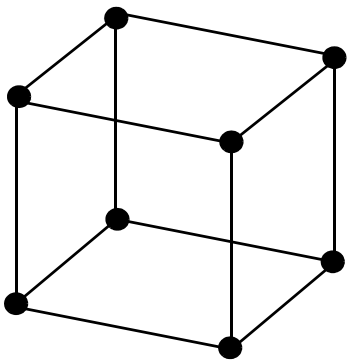
스크립트를 스크립트 창에 저장 " 반응 표면 설계 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

12 장

완전 요인 설계

완전 요인 설계는 가능한 모든 요인 설정 조합으로 실행되는 실험을 정의합니다. 완전 요인 설계를 사용하면 가능한 모든 교호작용을 추정할 수 있습니다. 완전 요인 설계는 선별 설계에 비해 규모가 크고, 높은 수준의 교호작용이 활성화되지 않는 경우가 종종 있으므로 비효율적일 수 있습니다. 일반적으로 이 설계는 요인과 수준의 수가 적고 가능한 모든 교호작용에 대한 정보가 필요할 때 사용됩니다. 예를 들어 완전 요인 설계는 MSA(측정 시스템 분석)의 기초를 형성하는 경우가 많습니다.

그림 12.1 세 가지 2 수준 요인에 대한 완전 요인 설계



목차

완전 요인 설계 개요 349
완전 요인 설계의 예 349
설계 구성 349
실험 데이터 분석 351
완전 요인 설계 생성 355
반응 355
요인 357
출력 옵션 선택 359
테이블 생성 360
완전 요인 설계 옵션 361

완전 요인 설계 개요

완전 요인 설계에서는 요인 수준의 모든 조합으로 실험 런을 수행합니다. 표본 크기는 요인 수준 수의 곱입니다. 예를 들어 2수준 요인, 3수준 요인 및 4수준 요인이 포함된 요인 실험의 런 수는 2 x 3 x 4 = 24입니다.

완전 요인 설계 플랫폼은 수준 수에 제한이 없는 연속형 요인과 범주형 요인을 모두 지원합니다. 완전히 랜덤한 방식으로 실험을 실행할 수 있다고 가정합니다.

완전 요인 설계는 모든 설계 유형 중에서 가장 보수적입니다. 그러나 표본 크기는 요인 수에 따라 기하급수적으로 증가하므로 완전 요인 설계를 실행하는 데 비용이 너무 많이 들 수 있습니다. 사용자 설계, 확증적 선별 설계 및 선별 설계는 덜 보수적이지만 더 유용하고 비용 효율적입니다.

완전 요인 설계의 예

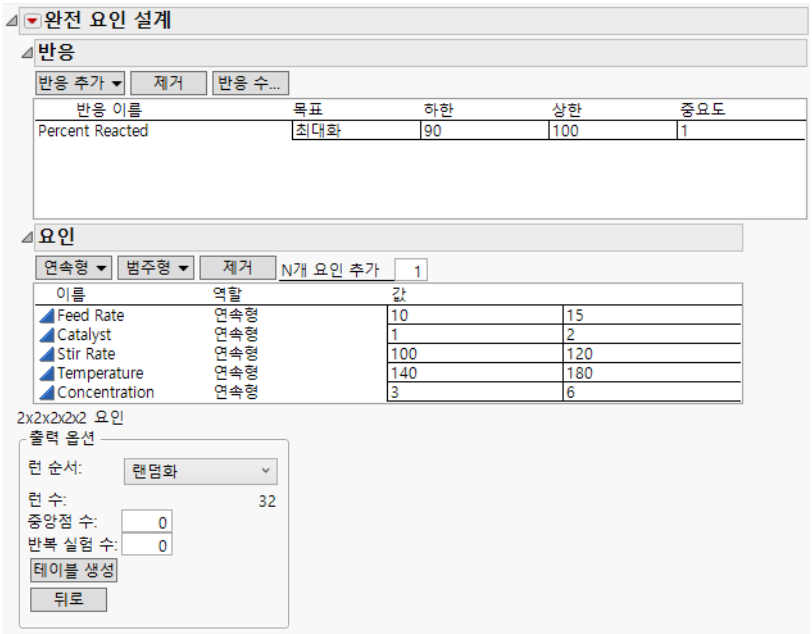
이 예에서는 반응기 수율에 영향을 미치는 2수준 요인 5개(Feed Rate, Catalyst, Stir Rate, Temperature, Concentration)의 효과를 연구하기 위해 완전 요인 설계를 구성합니다. 각각 수준이 2개인 요인이 5개 있으므로 완전 요인 설계의 런은 최소 $2^5 = 32$회입니다. 이 실험 상황과 관련하여 규모가 더 작은 선별 설계에 대한 내용은 "선별 설계" 장의 "선별 설계의 추가 예"(298페이지)에서 확인하십시오.

이 예에서는 기존 테이블의 반응과 요인을 불러옵니다. 새 실험을 직접 설계할 때는 반응과 요인을 수동으로 입력합니다. 자세한 내용은 "반응"(355페이지) 및 "요인"(357페이지)에서 확인하십시오.

설계 구성

1. **DOE > 전통적 설계 > 완전 요인 설계**를 선택합니다.
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor Response.jmp 를 엽니다.
3. " 완전 요인 설계 " 의 빨간색 삼각형을 클릭하고 **반응 불러오기**를 선택합니다.
4. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor Factors.jmp 를 엽니다.
5. " 완전 요인 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다.
6. **계속**을 클릭합니다.

그림 12.2 완전 요인 예의 반응 및 요인 패널



참고 : 7 단계에서 난수 시드값을 설정하면 설계 테이블의 런이 이 예와 같은 순서로 나타납니다. 설계를 직접 구성할 때는 이 단계가 필요하지 않습니다.

7. (선택 사항) " 완전 요인 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다. "12345" 를 입력하고 **확인**을 클릭합니다.

   " 출력 옵션 " 패널의 " 런 순서 " 가 랜덤화로 설정되어 있습니다. 설계 테이블의 런 순서는 난수 시드값에 따라 랜덤으로 결정됩니다.

   " 런 수 " 는 32 로 설정되어 있습니다. 이 값은 가능한 모든 요인 수준 조합의 수입니다.

8. **테이블 생성**을 클릭합니다.

   설계 데이터 테이블의 첫 번째 열에는 각 런의 요인 수준 조합이 + 및 - 기호로 표시되어 상한 및 하한 요인 설정을 나타냅니다. 또한 이 테이블에는 실험을 수행할 때 반응 값을 입력하기 위한 빈 Y 열 (Percent Reacted) 이 있습니다.

그림 12.3 반응기 실험에 대한 완전 요인 설계

	패턴	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	Percent Reacted
1	++−+−	15	2	100	180	3	•
2	+++++	15	2	120	180	6	•
3	−++−−	10	2	120	140	3	•
4	−+−−+	10	2	100	140	6	•
5	−++++	10	2	120	180	6	•
6	−+−−−	10	2	100	140	3	•
7	++−−+	15	2	100	140	6	•
8	−−+−+	10	1	120	140	6	•
9	−−−−−	10	1	100	140	3	•
10	++−++	15	2	100	180	6	•
11	+−++−	15	1	120	180	3	•
12	−+++−	10	2	120	180	3	•
13	+−−+−	15	1	100	180	3	•
14	−−++−	10	1	120	180	3	•
15	−++−+	10	2	120	140	6	•
16	+++−+	15	2	120	140	6	•
17	+−−++	15	1	100	180	6	•
18	++++−	15	2	120	180	3	•
19	−−−+−	10	1	100	180	3	•
20	+−−−+	15	1	100	140	6	•
21	−+−++	10	2	100	180	6	•
22	−−−++	10	1	100	180	6	•
23	++−−−	15	2	100	140	3	•
24	+−+−+	15	1	120	140	6	•
25	−−+++	10	1	120	180	6	•
26	+++−−	15	2	120	140	3	•
27	+−+−−	15	1	120	140	3	•
28	−−+−−	10	1	120	140	3	•
29	−−−−+	10	1	100	140	6	•
30	−+−+−	10	2	100	180	3	•
31	+−−−−	15	1	100	140	3	•
32	+−+++	15	1	120	180	6	•

실험 데이터 분석

이제 완료된 실험의 데이터 분석을 진행합니다. 선별과 단계별 회귀라는 두 가지 방법을 사용하여 결과를 분석합니다. 그런 다음 예측 프로파일러를 사용하여 최적 설정을 찾을 수 있습니다.

선별 플랫폼을 사용한 분석

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 32 Runs.jmp 를 엽니다 .
2. **Screening** 스크립트를 실행합니다 .

   " 선별 " 보고서는 " 대비 " 보고서와 "1/2 정규 확률도 " 를 보여 줍니다 . " 대비 " 보고서는 최대 5 원 교호작용까지 총 31 개의 잠재적 효과에 대한 추정값을 보여 줍니다 .

그림 12.4 Reactor 32 Runs.jmp 에 대한 대비 보고서

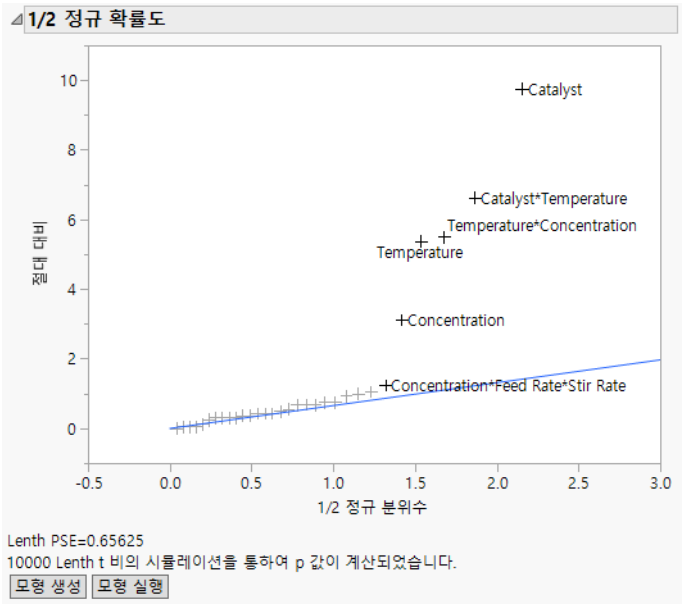
대비

항	대비	Lenth t-비	개별 p 값	동시 p 값
Catalyst	9.75000	14.86	<.0001*	<.0001*
Temperature	5.37500	8.19	<.0001*	0.0005*
Concentration	-3.12500	-4.76	0.0011*	0.0152*
Feed Rate	-0.68750	-1.05	0.2883	0.9998
Stir Rate	-0.31250	-0.48	0.6494	1.0000
Catalyst*Temperature	6.62500	10.10	<.0001*	0.0002*
Catalyst*Concentration	1.00000	1.52	0.1335	0.9553
Temperature*Concentration	-5.50000	-8.38	<.0001*	0.0004*
Catalyst*Feed Rate	0.68750	1.05	0.2883	0.9998
Temperature*Feed Rate	-0.43750	-0.67	0.4941	1.0000
Concentration*Feed Rate	0.06250	0.10	0.9266	1.0000
Catalyst*Stir Rate	0.43750	0.67	0.5212	1.0000
Temperature*Stir Rate	1.06250	1.62	0.1139	0.9205
Concentration*Stir Rate	0.43750	0.67	0.4941	1.0000
Feed Rate*Stir Rate	0.37500	0.57	0.5859	1.0000
Catalyst*Temperature*Concentration	-0.12500	-0.19	0.8557	1.0000
Catalyst*Temperature*Feed Rate	0.68750	1.05	0.2883	0.9998
Catalyst*Concentration*Feed Rate	-0.93750	-1.43	0.1587	0.9784
Temperature*Concentration*Feed Rate	0.31250	0.48	0.6494	1.0000
Catalyst*Temperature*Stir Rate	0.56250	0.86	0.3772	1.0000
Catalyst*Concentration*Stir Rate	0.06250	0.10	0.9266	1.0000
Temperature*Concentration*Stir Rate	0.06250	0.10	0.9266	1.0000
Catalyst*Feed Rate*Stir Rate	0.75000	1.14	0.2478	0.9989
Temperature*Feed Rate*Stir Rate	-0.37500	-0.57	0.5859	1.0000
Concentration*Feed Rate*Stir Rate	-1.25000	-1.90	0.0705	0.7592
Catalyst*Temperature*Concentration*Feed Rate	0.31250	0.48	0.6494	1.0000
Catalyst*Temperature*Concentration*Stir Rate	-0.31250	-0.48	0.6494	1.0000
Catalyst*Temperature*Feed Rate*Stir Rate	0.00000	0.00	1.0000	1.0000
Catalyst*Concentration*Feed Rate*Stir Rate	0.75000	1.14	0.2478	0.9989
Temperature*Concentration*Feed Rate*Stir Rate	0.50000	0.76	0.4328	1.0000
Catalyst*Temperature*Concentration*Feed Rate*Stir Rate	-0.25000	-0.38	0.7150	1.0000

참고 : " 대비 " 보고서의 *p* 값은 몬테카를로 시뮬레이션을 사용하여 계산되므로 그림 12.4 에 표시된 값과 동일하지 않을 수 있습니다 . 자세한 내용은 "2 수준 선별 적합 플랫폼 " 장의 "Lenth 유사 표준 오차 "(323 페이지) 에서 확인하십시오 .

" 대비 " 개요에 강조 표시된 6 개 효과는 "1/2 정규 확률도 " 에서 라벨이 지정되어 있습니다 .

그림 12.5 Reactor 32 Runs.jmp 에 대한 1/2 정규 확률도



"1/2 정규 확률도 " 는 라벨이 지정된 효과 중 5 개 이상이 랜덤 변형의 결과일 경우 예상할 수 있는 것보다 더 크다는 강력한 증거를 제시합니다 . 즉 , 이러한 효과가 활성임을 나타냅니다 . 이 그림에는 삼원 교호작용 Concentration*Feed Rate*Stir Rate 가 활성이라는 징후가 명확하게 나타나지 **않습니다** .

그림 12.4 의 " 대비 " 개요에서 삼원 교호작용 Concentration*Feed Rate*Stir Rate 의 " 개별 p 값 " 은 0.0705 이고 " 동시 p 값 " 은 0.7592 입니다 . 이 효과는 "1/2 정규 확률도 " 에서 강조되지 않고 *p* 값이 크기 때문에 모형에 포함하지 않기로 결정합니다 .

3. "1/2 정규 확률도 " 에서 직사각형 형태로 드래그하여 Concentration*Feed Rate*Stir Rate 를 제외하고 라벨이 지정된 모든 효과를 선택합니다 .
4. **모형 생성**을 클릭하여 5 개 효과가 포함된 " 모형 적합 " 창을 엽니다 .
5. **실행**을 클릭합니다 .

   " 실제값 대 예측값 그림 " 에 적합 결여의 증거가 없으며 " 효과 요약 " 개요에서는 5 개 효과가 모두 유의함을 보여 줍니다 .

단계별 회귀를 사용한 분석

1. Reactor 32 Runs.jmp 데이터 테이블로 돌아가거나 , **도움말 > 샘플 데이터 라이브러리**를 선택한 후 Design Experiment/Reactor 32 Runs.jmp 를 열어 데이터 테이블을 다시 엽니다 .
2. **Model** 스크립트를 실행합니다 .

   " 모형 효과 생성 " 목록에 이원 교호작용까지만 포함되어 있습니다 . 하지만 모든 교호작용을 고려하려고 합니다 .
3. " 열 선택 " 목록에서 Feed Rate ~ Concentration 을 선택합니다 .

4. **매크로 > 완전 요인**을 클릭합니다 .

   가능한 모든 효과가 " 모형 효과 생성 " 목록에 추가됩니다 .

5. **분석법**을 **단계별**로 변경합니다 .
6. **실행**을 클릭합니다 .
7. **중지 규칙**을 **최소 AICc** 로 변경합니다 .

   설계된 실험의 경우 AICc 가 BIC 보다 선호됩니다 . BIC 중지 규칙은 모형에 비활성 효과를 허용하는 경향이 있어서 대개 AICc 보다 관대하기 때문입니다 .

8. **시작**을 클릭합니다 .

   단계별 절차에서는 6 개 효과를 잠재적 활성 효과로 선택합니다 .

9. **모형 실행**을 클릭합니다 .

   그러면 6 개 효과를 사용하여 모형을 적합시킵니다 . " 효과 요약 " 개요에서는 Catalyst* Concentration 의 p 값이 0.0896 이므로 0.05 유의 수준에서 유의하지 않음을 나타냅니다 .

10. " 효과 요약 " 개요에서 Catalyst*Concentration 을 선택하고 **제거**를 클릭합니다 .

    나머지 5 개 효과는 모두 매우 유의합니다 . 해당 효과는 선별 플랫폼을 사용하여 식별한 5 개 효과와 동일합니다 (" 선별 플랫폼을 사용한 분석 "(351 페이지) 참조).

예측 프로파일러를 사용한 최적 설정

이제 모형에 유지한 5 개의 유의 효과와 관련된 세 가지 활성 요인에 대한 최적 설정을 찾습니다 .

1. Reactor 32 Runs.jmp 데이터 테이블에서 **Reduced Model** 스크립트를 실행합니다 .

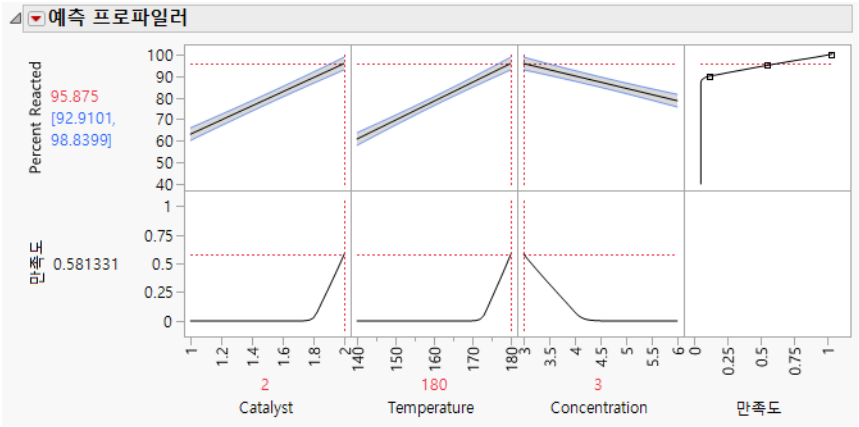
   "Reduced Model" 스크립트는 " 선별 플랫폼을 사용한 분석 "(351 페이지) 및 " 단계별 회귀를 사용한 분석 "(353 페이지) 에서 식별한 5 개 효과 모형에 대한 " 모형 적합 " 창을 엽니다 .

2. **실행**을 클릭합니다 .

   " 완전 요인 " 창에서 반응을 정의할 때 " 목표 " 를 " 최대화 " 로 지정했으므로 " 예측 프로파일러 " 보고서에 " 만족도 " 가 표시됩니다 . 프로파일러의 위쪽 행에서 맨 오른쪽에 표시된 만족도 함수는 값이 100 일 때 가장 적합하고 값이 90 이하일 때 가장 적합하지 않음을 나타냅니다 .

3. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 12.6 만족도 최대화 설정을 보여 주는 예측 프로파일러

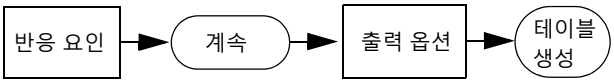


표시된 설정에서 Percent Reacted 예측 평균은 95.875 이고 신뢰 구간은 92.91 ~ 98.84 입니다 . 세 가지 요인 모두 식별된 설정이 실험에 사용된 범위의 극단에 있습니다 . 이후 실험에서는 이러한 설정 이외에서 공정 동작을 분석해야 합니다 .

완전 요인 설계 생성

DOE > 전통적 설계 > 완전 요인 설계를 선택하여 완전 요인 설계를 생성합니다. 먼저 반응 및 요인을 정의합니다. 그런 다음 출력 옵션을 설정합니다. 설계 과정은 그림 12.7의 흐름을 따릅니다. 작업 흐름에 대한 자세한 내용은 "DOE 워크플로우: 기술, 지정, 설계"(56 페이지)에서 확인하십시오.

그림 12.7 완전 요인 설계 흐름

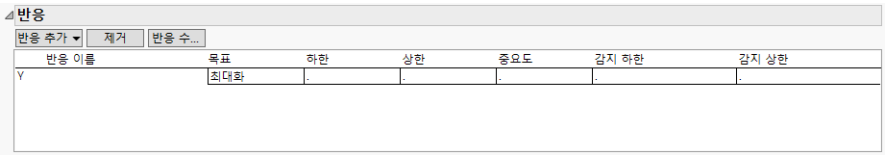


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다 . 이 옵션을 사용하면 반응 이름 , 목표 , 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다 .

그림 12.8 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다 . " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다 . " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다 .

JMP PRO 함수 (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다 . 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다 .

제거 선택한 반응을 제거합니다 .

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다 . 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다 . 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다 . 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다 .

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응 열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .
- " 없음 " 목표는 최적화에 관한 목표가 없음을 나타냅니다 . 만족도 함수가 구성되지 않습니다 .

참고 : 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다 . 데이터 테이블에서 반응 열의 " 열 정보 " 창을 열고 (**열** > **열 정보**) 원하는 목표값을 입력하십시오 .

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다 . 이러한 값은 반응에 대한 가중치로 처리됩니다 . 반응이 하나뿐이면 기본적으로 1 로 설정되므로 중요도를 지정할 필요가 없습니다 .

JMP PRO 감지 한계 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다 . 감지 한계는 열 특성으로 저장됩니다 . 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오 .

반응 개요 편집

" 반응 " 개요에서 다음 사항에 유의하십시오 .

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다 .
- 목표를 클릭하여 변경할 수 있습니다 .
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다 .
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다 .

반응 한계 열 특성

반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

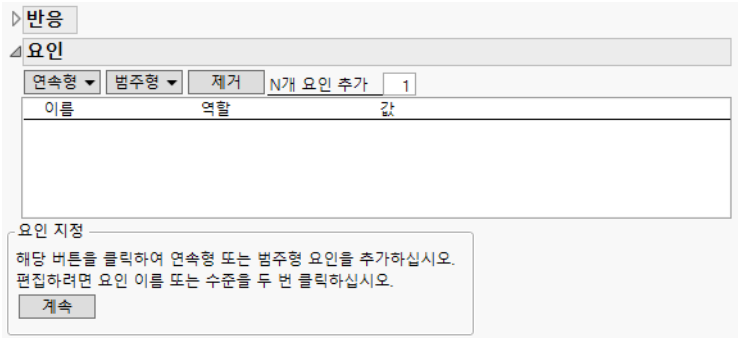
하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

요인

완전 요인 설계의 요인은 연속형 또는 범주형일 수 있습니다.

팁 : " 요인 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **요인 저장**을 선택하는 것이 좋습니다 . 이 옵션을 사용하면 요인 이름 , 역할 , 변경 및 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다 .

그림 12.9 요인 개요



연속형 연속형 요인을 추가합니다 . 결과 데이터 테이블의 데이터 유형은 숫자입니다 . 연속형 요인은 개념적으로 공정 및 측정 시스템의 제한을 고려하여 사용자가 제공하는 하한과 상한 사이의 어떤 값이든 설정할 수 있는 요인입니다 .

범주형 범주형 요인을 추가합니다 . 클릭하여 수준 수를 선택하거나 지정합니다 . 결과 데이터 테이블의 데이터 유형은 문자입니다 . 수준의 값 순서는 왼쪽에서 오른쪽으로 입력되는 값의 순서입니다 . 이 순서는 설계 데이터 테이블이 생성된 후 " 값 순서 " 열 특성에 저장됩니다 .

범주형 요인의 기본값은 L1, L2,..., Lk 이며 여기서 k 는 지정한 수준 수입니다 . 기본값을 실험과 관련된 수준 이름으로 바꿉니다 .

제거 선택한 요인을 제거합니다 .

N 개 요인 추가 여러 요인을 추가합니다 . 추가할 요인 수를 입력하고 **요인 추가**를 클릭한 후 요인 유형을 선택합니다 . 유형이 다른 여러 요인을 추가하려면 **N 개 요인 추가**를 반복합니다 .

요인 개요

" 요인 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할입니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인 유형이 적절하게 모델링되도록 합니다 . 요인의 역할에 따라 데이터 테이블에 저장되는 다른 요인 특성이 결정됩니다 . 자세한 내용은 " 요인 열 특성 "(359 페이지) 에서 확인하십시오 .

값 요인에 대한 실험 설정입니다 .

요인 개요 편집

" 요인 " 개요에서 다음 사항에 유의하십시오 .

- 요인 이름을 편집하려면 해당 요인 이름을 두 번 클릭합니다 .
- 값을 편집하려면 " 값 " 열의 값을 클릭합니다 .

요인 열 특성

각 요인에 대해 "선별 설계" 창에서 "테이블 생성"을 선택하여 설계를 생성하면 다양한 열 특성이 설계 테이블에 저장됩니다. 또한 "요인 저장" 옵션을 선택할 때 생성되는 데이터 테이블에도 이러한 특성이 자동으로 저장됩니다. 이러한 열 특성 및 관련 예에 대한 자세한 내용은 부록 A, "열 특성"에서 확인하십시오.

코딩 "역할" 이 "연속형" 인 경우 요인의 "코딩" 열 특성이 저장됩니다. 이 특성은 하한값과 상한값이 각각 –1 과 +1 에 해당하도록 요인 값을 변환합니다. 자세한 내용은 "열 특성" 부록의 " 코딩 "(739 페이지) 에서 확인하십시오 .

값 순서 "역할" 이 "범주형" 이거나 블록 변수가 구성된 경우 요인의 "값 순서" 열 특성이 저장됩니다. 이 특성은 요인 수준이 나타나는 순서를 결정합니다. 자세한 내용은 "열 특성" 부록의 " 값 순서 "(755 페이지) 에서 확인하십시오 .

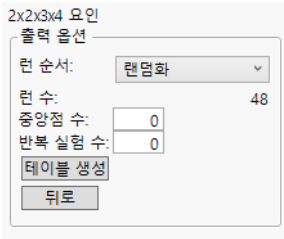
설계 역할 각 요인에 " 설계 역할 " 열 특성이 지정됩니다 . 요인을 정의할 때 지정하는 역할에 따라 " 설계 역할 " 열 특성의 값이 결정됩니다 . " 설계 역할 " 특성은 실험 데이터를 모델링할 때 요인이 어떤 용도로 사용되는지를 반영합니다 . " 설계 역할 " 값은 설계 확대 플랫폼에서 사용됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 설계 역할 "(737 페이지) 에서 확인하십시오 .

요인 변경 각 요인에 값이 " 쉬움 " 인 " 요인 변경 " 열 특성이 할당됩니다 . " 요인 변경 " 특성은 실험 데이터를 모델링할 때 요인이 사용되는 방식을 반영합니다 . " 요인 변경 " 값은 설계 확대 및 설계 평가 플랫폼에서 사용됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 요인 변경 "(752 페이지) 에서 확인하십시오 .

출력 옵션 선택

반응과 요인을 입력하고 **계속**을 클릭한 후 "출력 옵션" 개요에서 설계 테이블에 대한 옵션을 선택할 수 있습니다. 완전 요인 설계의 구조가 개요 맨 위에 나타납니다.

그림 12.10 출력 옵션 패널



런 순서

런 순서 옵션은 설계 테이블의 런 순서를 결정합니다 .

동일하게 유지 설계 테이블의 행이 왼쪽에서 오른쪽으로 정렬됩니다 .

좌에서 우로 정렬 설계 테이블의 열이 왼쪽에서 오른쪽으로 정렬됩니다 .

랜덤화 설계 테이블의 행이 랜덤 순서로 나타납니다.

우에서 좌로 정렬 설계 테이블의 열이 오른쪽에서 왼쪽으로 정렬됩니다.

중앙점 및 반복 실험

런 수 중앙점 또는 반복 실험을 추가하기 전 설계의 런 수를 표시합니다.

중앙점 수 설계에 중앙점으로 포함할 추가 런 수를 지정합니다. 중앙점은 모든 연속형 요인이 요인 범위의 가운데에 설정되는 런입니다.

설계에 연속형 요인과 범주형 요인이 모두 포함되어 있다고 가정해 보겠습니다. "출력 옵션" 패널에서 중앙점을 요청할 경우 다음과 같이 중앙점이 분포됩니다.

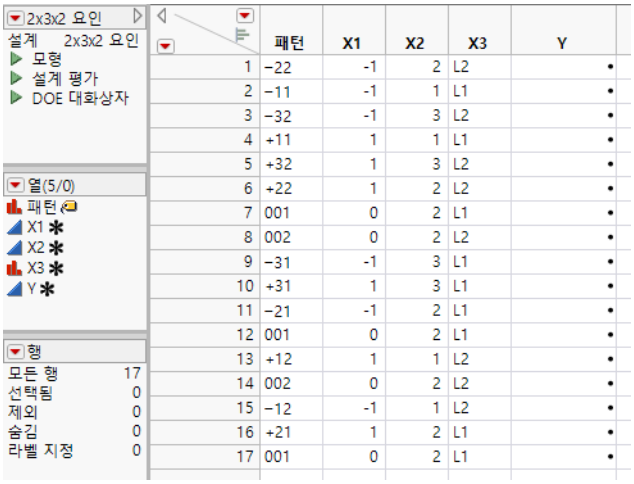
1. 범주형 요인의 설정은 "요인" 개요에 지정된 값 순서를 사용하여 정렬됩니다.
2. 범주형 요인의 각 설정 조합에 순서대로 중앙점 하나가 할당되고 모든 중앙점이 할당될 때까지 반복됩니다.

반복 실험 수 중앙점을 포함하여 전체 설계를 반복할 횟수입니다. 한 번 반복하면 런 수가 두 배가 됩니다.

테이블 생성

"테이블 생성"을 클릭하면 실험 런이 포함된 데이터 테이블이 생성됩니다. 그림 12.11의 예에서는 X1(2수준 연속형 요인), X2(3수준 연속형 요인), X3(2수준 범주형 요인)이라는 세 가지 요인에 대한 5개 중앙점이 포함된 완전 요인 설계를 보여 줍니다. 이 설계는 요인 수준의 기본값을 사용합니다. 중앙점은 7, 8, 12, 14 및 17 행에 있습니다. 자세한 내용은 "패턴 열"(361 페이지)에서 확인하십시오.

그림 12.11 설계 데이터 테이블



	패턴	X1	X2	X3	Y
1	−22	-1	2	L2	•
2	−11	-1	1	L1	•
3	−32	-1	3	L2	•
4	+11	1	1	L1	•
5	+32	1	3	L2	•
6	+22	1	2	L2	•
7	001	0	2	L1	•
8	002	0	2	L2	•
9	−31	-1	3	L1	•
10	+31	1	3	L1	•
11	−21	-1	2	L1	•
12	001	0	2	L1	•
13	+12	1	1	L2	•
14	002	0	2	L2	•
15	−12	-1	1	L2	•
16	+21	1	2	L1	•
17	001	0	2	L1	•

왼쪽 위에 표시된 테이블 이름은 해당 테이블을 생성한 설계 유형입니다.

설계 테이블 스크립트

설계 테이블에는 다음과 같은 스크립트가 포함되어 있습니다 .

모형 " 분석 > 모형 적합 " 플랫폼을 실행합니다 .

설계 평가 "DOE > 설계 진단 > 설계 평가 " 플랫폼을 실행합니다 .

DOE 대화상자 설계 테이블을 생성하는 데 사용된 " 완전 요인 설계 " 창을 다시 생성합니다 . 또한 이 스크립트는 설계를 생성하는 데 사용된 난수 시드값도 포함합니다 .

데이터를 분석하려면 **선별** 또는 **모형** 스크립트를 실행합니다.

패턴 열

패턴 열에는 지정된 행의 런을 요약하는 항목이 포함되어 있습니다. 패턴을 그림의 라벨 변수로 사용할 수 있습니다.

- 2 수준 연속형 요인의 경우 하한 설정은 "–", 상한 설정은 "+", 중앙점은 "0" 으로 각각 나타냅니다 .
- 3 수준 이상 연속형 요인의 경우 다음과 같습니다 .
  - 중앙점이 아닌 경우 요인 설정은 런의 값 수준에 해당하는 정수로 나타냅니다 .
  - 중앙점인 경우 요인 설정은 "0" 으로 나타냅니다 .
- 범주형 요인의 경우 요인 설정은 런의 값 수준에 해당하는 정수로 나타냅니다 .

완전 요인 설계 옵션

완전 요인 설계 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

반응 저장 " 반응 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 . 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다 .

반응 불러오기 " 반응 저장 " 옵션을 사용하여 저장한 반응을 불러옵니다 .

요인 저장 " 요인 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 각 요인의 열에는 수준이 포함됩니다 . 다른 정보는 열 특성으로 저장됩니다 . 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다. 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다 . 그런 다음 " 설계 역할 " 영역에서 적절한 역할을 선택합니다 .

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 불러옵니다 .

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) " 요인 제약 조건 정의 " 또는 " 선형 제약 조건 " 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다 . 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다 . 마지막 행은 부등식 한계를 포함합니다 . 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 제약 조건 상태 "(762 페이지) 에서 확인하십시오 .

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) " 제약 조건 저장 " 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다 .

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다 . 이러한 작업에는 다음 중 하나 이상이 포함됩니다 .

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다 .

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다 .

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다 .
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다 .

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다 .

- **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다 . 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 **기본 분석**의 " 시뮬레이션 " 장에서 확인하십시오 .

고급 옵션 사용할 수 없음

스크립트를 스크립트 창에 저장 " 완전 요인 설계 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

13 장

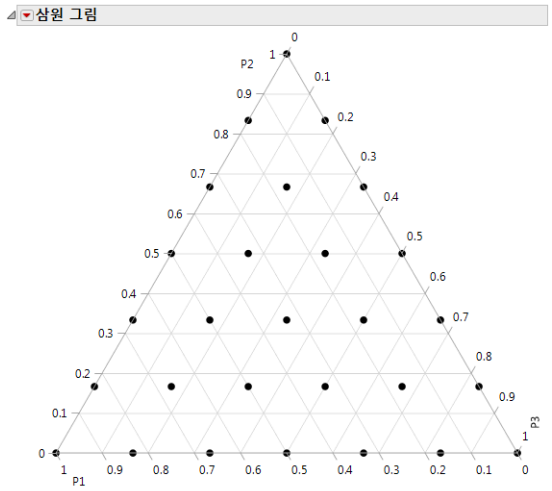
혼합물 설계
배합 설계

혼합물 설계 플랫폼을 사용하면 혼합물의 성분인 요인으로 실험을 구축할 수 있습니다. 심플렉스, 꼭지점 및 격자와 같은 몇 가지 전통적 혼합물 설계 방법 중에서 선택합니다. 꼭지점 방법의 경우 혼합 요인 공간의 형상을 제한하는 일련의 선형 부등식 제약 조건을 제공할 수 있습니다.

혼합물 실험에서 요인 값은 혼합물의 비율이며 0과 1 사이입니다. 혼합물 실험에는 요인 비율의 합이 1(100%)인 3개 이상의 요인이 포함됩니다. 혼합물 실험은 요인을 서로 독립적으로 변경할 수 없다는 점에서 다른 실험 유형과 다릅니다. 한 요인의 비율을 변경하면 하나 이상의 다른 요인의 비율도 변경되어야 합니다. 이 단순한 사실이 요인 공간, 설계 특성, 결과 해석 등 혼합물 실험의 모든 측면에 큰 영향을 미칩니다.

혼합 성분은 합이 1이 되는 비율이므로 혼합물 설계의 실현 가능 영역은 심플렉스 형태가 됩니다. 예를 들어 3요인 설계는 삼각형 2D 그래프로 시각화하고 4요인 설계는 3D 4면체로 시각화할 수 있습니다.

그림 13.1 삼원 그림



목차

혼합물 설계 개요 367
혼합물 설계 창 367
반응 367
요인 369
삼원 그림 개요 371
혼합물 설계 유형의 예 371
최적 혼합물 설계 372
심플렉스 중심 설계 373
심플렉스 격자 설계 375
꼭지점 설계 376
ABCD 설계 380
공간 채움 설계 381
혼합물 설계 적합 383
전체 모형 검정 및 분산 분석 보고서 384
반응 표면 보고서 이해 384
화학 혼합물 예 384
설계 생성 385
혼합 모형 분석 386
예측 프로파일러 387
혼합물 프로파일러 388

혼합물 설계 개요

다음과 같은 혼합물 설계 유형을 선택할 수 있습니다 .

최적 사용자 설계 도구를 사용하여 최적 혼합물 설계를 생성합니다 . 자세한 내용은 " 최적 혼합물 설계 "(372 페이지) 에서 확인하십시오 .

심플렉스 중심 (무제약 요인 또는 하한 제약 조건이 있는 요인에 사용할 수 있음) 순수 혼합물 및 지정된 차수까지의 요인 혼합으로 설계를 생성합니다 . 순수 혼합물에서는 한 성분이 100% 이고 다른 모든 성분은 0% 입니다 . 자세한 내용은 " 심플렉스 중심 설계 "(373 페이지) 에서 확인하십시오 .

심플렉스 격자 (무제약 요인 또는 하한 제약 조건이 있는 요인에 사용할 수 있음) 순수 혼합물 및 런에 대한 삼각 격자를 생성하는 공간 채움 혼합으로 설계를 생성합니다 . 격자 수준의 수를 지정할 수 있습니다 . 자세한 내용은 " 심플렉스 격자 설계 "(375 페이지) 에서 확인하십시오 .

꼭지점 (하나 이상의 요인에 대해 선형 제약 조건 또는 제한된 상한 / 하한이 있는 경우에 사용할 수 있음) 제약이 있는 요인 공간의 꼭지점 및 꼭지점 조합을 기반으로 설계를 생성합니다 . 자세한 내용은 " 꼭지점 설계 "(376 페이지) 에서 확인하십시오 .

ABCD 설계 Snee(1975) 가 고안한 혼합에 대한 선별 설계를 생성합니다 . 자세한 내용은 "ABCD 설계 "(380 페이지) 에서 확인하십시오 .

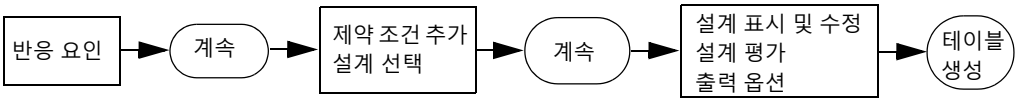
공간 채움 선형 제약 조건을 수용하는 공간 채움 설계를 생성합니다 . 자세한 내용은 " 공간 채움 설계 "(381 페이지) 에서 확인하십시오 .

참고 : 최적 , 꼭지점 및 공간 채움 설계는 성분에 대한 선형 제약 조건과 제한을 통합할 수 있습니다 .

혼합물 설계 창

설계 단계를 진행하는 동안 "혼합물 설계" 창이 업데이트됩니다. 자세한 내용은 "DOE 워크플로우: 기술, 지정, 설계"(56 페이지)에서 확인하십시오. 워크플로우와 개요는 설계 유형에 따라 다릅니다.

그림 13.2 혼합물 설계 워크플로우

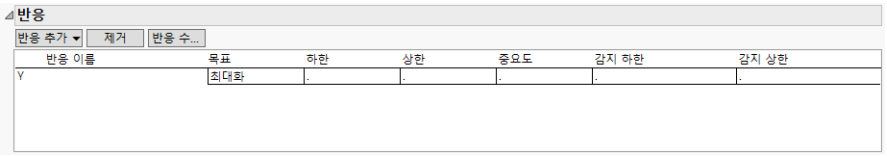


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다 . 이 옵션을 사용하면 반응 이름 , 목표 , 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다 .

그림 13.3 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다 . " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다 . " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다 .

JMP PRO **함수** (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다 . 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다 .

제거 선택한 반응을 제거합니다 .

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다 . 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다 . 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다 . 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다 .

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응 열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .

– "없음" 목표는 최적화에 관한 목표가 없음을 나타냅니다. 만족도 함수가 구성되지 않습니다.

참고: 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다. 데이터 테이블에서 반응 열의 "열 정보" 창을 열고 (**열** > **열 정보**) 원하는 목표값을 입력하십시오.

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다. 이러한 값은 반응에 대한 가중치로 처리됩니다. 반응이 하나뿐이면 기본적으로 1로 설정되므로 중요도를 지정할 필요가 없습니다.

JMP PRO **감지 한계** 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다. 감지 한계는 열 특성으로 저장됩니다. 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오.

반응 개요 편집

"반응" 개요에서 다음 사항에 유의하십시오.

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다.
- 목표를 클릭하여 변경할 수 있습니다.
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다.
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다.

반응 한계 열 특성

반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

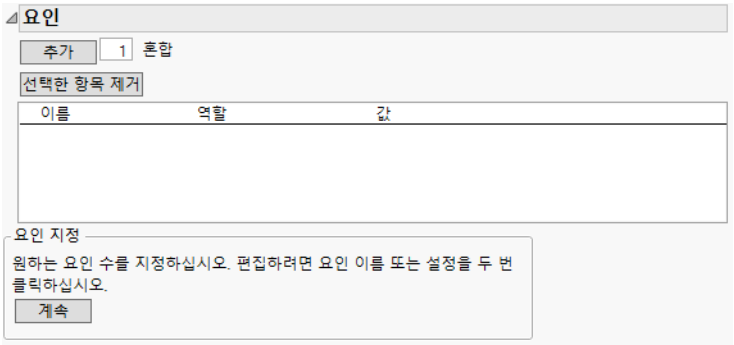
하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

요인

"요인" 개요에서 요인을 추가합니다.

팁: "요인" 개요를 완료하면 빨간색 삼각형 메뉴에서 **요인 저장**을 선택하는 것이 좋습니다. 이렇게 하면 요인 이름, 역할, 변경 및 값을 데이터 테이블에 저장하여 나중에 다시 불러올 수 있습니다.

그림 13.4 요인 개요



추가 추가할 요인 수를 입력하고 " 추가 " 를 클릭합니다 .

선택한 항목 제거 선택한 요인을 제거합니다 .

요인 목록

" 요인 " 목록에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할을 지정합니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인 유형이 적절하게 모델링되도록 합니다 .

값 요인에 대한 실험 설정입니다. 값을 삽입하려면 기본값을 클릭한 후 원하는 값을 입력합니다.

요인 목록 편집

" 요인 " 목록에서 다음을 수행합니다 .

- 요인 이름을 편집하려면 해당 요인 이름을 두 번 클릭합니다 .
- 값을 편집하려면 **값** 열의 값을 클릭합니다 .

선형 제약 조건

하나 이상의 선형 부등식 제약 조건을 입력하려면 **선형 제약 조건** 버튼을 클릭합니다. 설계의 모든 연속형 요인을 포함하는 선형 표현식 템플릿이 나타납니다. 요인에 대한 계수 값을 입력한 후 선형 제약 조건을 반영하는 부등식 방향을 선택합니다. 부등식의 오른쪽 상자에 제약 값을 지정합니다. 제약 조건을 더 추가하려면 **선형 제약 조건**을 다시 클릭합니다.

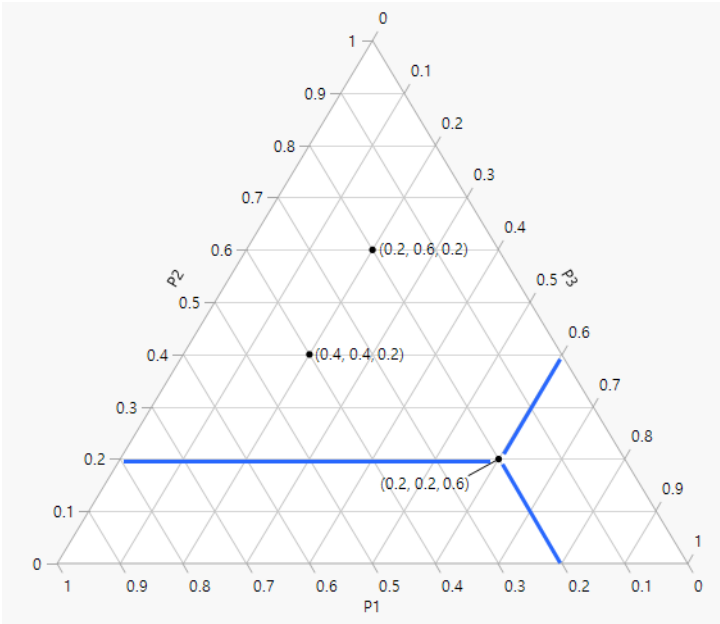
참고 : 선형 제약 조건을 포함하는 설계의 스크립트를 저장하면 해당 스크립트에 선형 제약 조건이 보다 작거나 같음 부등식 (≤) 으로 표현됩니다 .

삼원 그림 개요

삼원 그림은 합이 상수가 되는 세 가지 혼합 성분을 2차원으로 표현한 것입니다. 이 그림은 각 성분에 대한 모서리가 있는 정삼각형입니다. 제약이 없는 경우 삼각형의 각 꼭지점은 순수 혼합에 해당합니다. 즉, 한 성분이 1(100%)이고 다른 모든 성분은 0입니다. 성분에 제약이 있는 경우 실현 가능한 혼합은 삼원 그림의 부분으로 표현됩니다. 음영은 그림에서 실현 불가능한 부분을 제외하는 데 사용됩니다.

그림 13.5에는 세 가지 성분(P1, P2, P3)이 표시됩니다. 세 성분은 0에서 1(100%) 사이의 각 비율 범위로 제약되지 않습니다. 세 점에는 좌표, 즉 (P1, P2, P3) 형식으로 라벨이 지정됩니다. 한 점 (0.2, 0.2, 0.6)에는 각 혼합 성분의 축 격자선을 따라 파란색 선이 포함되어 축 값을 읽을 수 있습니다.

그림 13.5 삼원 그림의 예



혼합물 설계 유형의 예

- "최적 혼합물 설계"
- "심플렉스 중심 설계"
- "심플렉스 격자 설계"
- "꼭지점 설계"
- "ABCD 설계"

- " 공간 채움 설계 "

최적 혼합물 설계

최적 혼합물 설계는 혼합물 설계 플랫폼 또는 사용자 설계 플랫폼을 사용하여 생성할 수 있습니다.

여기서는 혼합물 설계 플랫폼을 사용하여 세 가지 요인이 있는 최적 혼합물 설계의 예를 생성합니다.

1. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다.
2. 이 예에서는 세 가지 요인을 사용합니다. " 요인 " 섹션을 변경할 필요가 없습니다.
3. **계속**을 클릭합니다.
4. " 혼합물 설계 유형 선택 " 패널에서 **최적**을 클릭합니다.
5. (선택 사항) 이 예의 출력과 일치시키려면 " 혼합물 설계 " 의 빨간색 삼각형을 클릭하고 " 난수 시드값 설정 " 을 선택한 후 1409 를 입력합니다.
6. (선택 사항) 이 예의 출력과 일치시키려면 " 혼합물 설계 " 의 빨간색 삼각형을 클릭하고 " 시작 수 " 를 선택한 후 2 를 입력합니다.
7. **요인 제약 조건 정의** 패널에서 **선형 제약 조건 지정**을 클릭한 후 **추가**를 클릭합니다.
8. "X1" 및 "X2" 상자에 1 을 입력하고 제약 조건 상자에 0.8 을 입력합니다.

   이렇게 하면 X1+X2 가 혼합물의 80% 미만으로 제한됩니다.
9. **모형** 패널에서 **교호작용 > 2 차**를 클릭합니다.

   이렇게 하면 모형에 교호작용 효과가 추가됩니다.
10. **설계 생성** 패널에서 **중앙점 수**에 2 를 입력합니다.
11. **설계 생성**을 클릭합니다.

그림 13.6 12 회 런 최적 설계

설계

런	X1	X2	X3
1	0.8	0	0.2
2	0	0	1
3	0	0.8	0.2
4	0	0	1
5	0	0.4	0.6
6	0.4	0	0.6
7	0	0.8	0.2
8	0.4	0	0.6
9	0.4	0.4	0.2
10	0.8	0	0.2
11	0.264183	0.269687	0.46613
12	0.264183	0.269687	0.46613

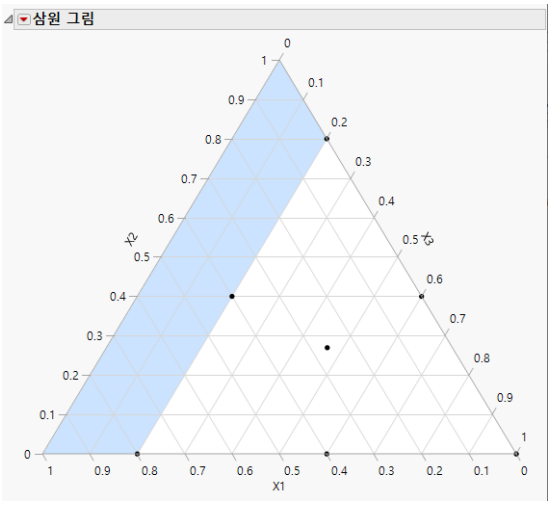
12. **테이블 생성**을 클릭합니다.

참고 : 실제로는 혼합물 양을 적절한 정밀도 수준으로 반올림합니다 .

설계 시각화

1. 설계 테이블에서 **그래프 > 삼원 그림**을 선택합니다 .
2. X1, X2, X3 을 선택하고 **X, 그림에 그리기**를 클릭한 후 **확인**을 클릭합니다 .

그림 13.7 최적 설계의 삼원 그림



삼원 그림은 설계를 구성하는 7 개의 고유 혼합물을 보여 줍니다 . 음영은 제약된 영역을 나타냅니다 . 삼원 그림에 대한 자세한 내용은 "삼원 그림 개요"(371 페이지)에서 확인하십시오 .

심플렉스 중심 설계

n 개 요인이 포함된 *k* 차 심플렉스 중심 설계는 다음과 같은 특성을 가진 혼합물로 구성됩니다 .

- 단일 요인 혼합물 (순수 혼합물)
- 동일한 수준으로 두 가지 요인 모두 혼합
- 동일한 *k* 수준으로 한 번에 최대 *k* 개 요인을 추가 혼합

k 가 요인 수와 같을 경우 모든 성분의 양이 동일한 중앙점 런이 포함됩니다 .

참고 : 이 설계는 하나 이상의 요인에 하한이 있을 경우 사용할 수 있습니다 .

설계 생성

심플렉스 중심 설계 예를 생성하려면

1. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다 .
2. 이 예에서는 세 가지 요인을 사용합니다 . " 요인 " 섹션을 변경할 필요가 없습니다 .
3. **계속**을 클릭합니다 .
4. **심플렉스 중심**을 클릭하여 설계를 생성합니다 .

참고 : 기본 설정 $k = 2$ 가 사용되었습니다 . k 값을 조정할 수 있는 텍스트 상자가 있습니다 .

그림 13.8 세 가지 요인의 2 차 심플렉스 중심 설계

설계

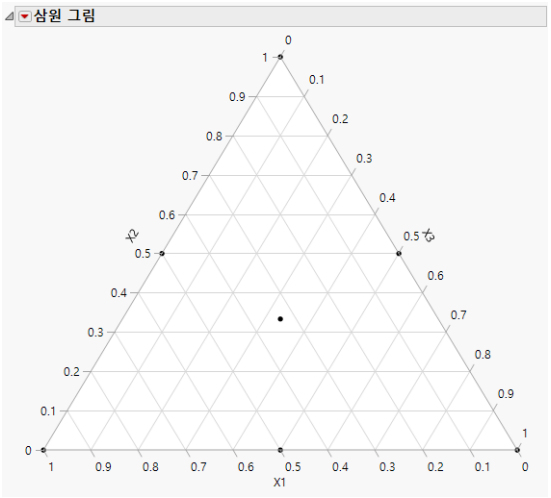
런	X1	X2	X3
1	1.00000	0.00000	0.00000
2	0.00000	1.00000	0.00000
3	0.00000	0.00000	1.00000
4	0.50000	0.50000	0.00000
5	0.50000	0.00000	0.50000
6	0.00000	0.50000	0.50000
7	0.33333	0.33333	0.33333

5. **테이블 생성**을 클릭합니다 .

설계 시각화

6. 설계 테이블에서 **그래프 > 삼원 그림**을 선택합니다 .
7. X1, X2, X3 을 선택하고 **X, 그림에 그리기**를 클릭한 후 **확인**을 클릭합니다 .

그림 13.9 심플렉스 중심 설계의 삼원 그림



삼원 그림에 대한 자세한 내용은 "삼원 그림 개요"(371 페이지)에서 확인하십시오.

심플렉스 격자 설계

심플렉스 격자 설계는 런에 대한 삼각 격자를 생성하는 공간 채움 설계입니다. 이 설계는 요인의 값이 i/m인 혼합물 집합입니다. 여기서 i는 0 ~ m 사이에서 달라지며 요인의 합은 1입니다. 3수준 설계의 경우 요인을 0, 0.33, 0.66 또는 1로 설정할 수 있습니다.

참고 : 이 설계는 하나 이상의 요인에 하한이 있을 경우 사용할 수 있습니다.

심플렉스 격자 설계 예를 생성하려면

1. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다.
2. 이 예에서는 세 가지 요인을 사용합니다. " 요인 " 섹션을 변경할 필요가 없습니다.
3. **계속**을 클릭합니다.
4. **심플렉스 격자**를 클릭하여 설계를 생성합니다.

   참고 : 기본 설정 5 수준이 사용되었습니다. 수준 수를 조정할 수 있는 텍스트 상자가 있습니다.

그림 13.10 세 가지 요인의 5 수준 심플렉스 격자 설계

설계

런	X1	X2	X3
1	0.00000	0.00000	1.00000
2	0.00000	0.20000	0.80000
3	0.00000	0.40000	0.60000
4	0.00000	0.60000	0.40000
5	0.00000	0.80000	0.20000
6	0.00000	1.00000	0.00000
7	0.20000	0.00000	0.80000
8	0.20000	0.20000	0.60000
9	0.20000	0.40000	0.40000
10	0.20000	0.60000	0.20000
11	0.20000	0.80000	0.00000
12	0.40000	0.00000	0.60000
13	0.40000	0.20000	0.40000
14	0.40000	0.40000	0.20000
15	0.40000	0.60000	0.00000
16	0.60000	0.00000	0.40000
17	0.60000	0.20000	0.20000
18	0.60000	0.40000	0.00000
19	0.80000	0.00000	0.20000
20	0.80000	0.20000	0.00000
21	1.00000	0.00000	0.00000

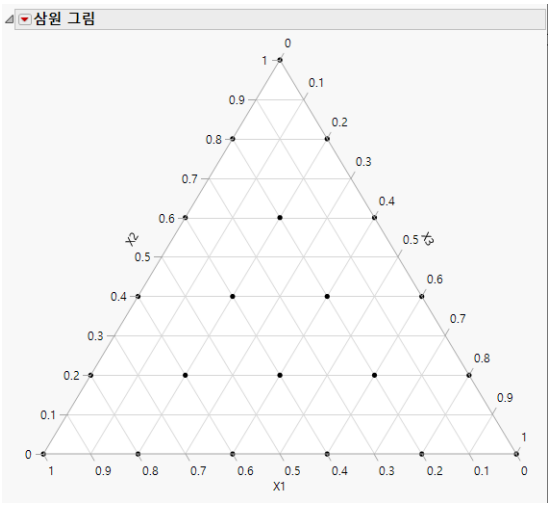
5. **테이블 생성**을 클릭합니다.

설계 시각화

6. 설계 테이블에서 **그래프 > 삼원 그림**을 선택합니다.

7. X1, X2, X3 을 선택하고 **X, 그림에 그리기**를 클릭한 후 **확인**을 클릭합니다 .

그림 13.11 심플렉스 격자 설계의 삼원 그림



삼원 그림에 대한 자세한 내용은 "삼원 그림 개요"(371 페이지)에서 확인하십시오.

수준 수가 증가하면 설계점 수도 증가하여 설계점 사이 간격이 줄어듭니다. 심플렉스 중심 설계와 달리 심플렉스 격자 설계는 중심을 반드시 포함하지는 않습니다.

꼭지점 설계

꼭지점 설계를 사용하려면 " 요인 " 패널에 정의한 한계 또는 선형 제약 조건에 따라 하나 이상의 요인에 대해 제한된 범위가 필요합니다. 꼭지점 설계는 요인 공간의 꼭지점과 지정된 차수까지의 꼭지점 평균의 혼합으로 구성됩니다. 선으로 연결된 두 인접 꼭지점의 중심점은 2 차 중심입니다. 평면을 공유하는 꼭지점의 중심점은 3 차 중심입니다.

범위 제약 조건이 있는 꼭지점의 예

꼭지점 설계 예를 생성하려면

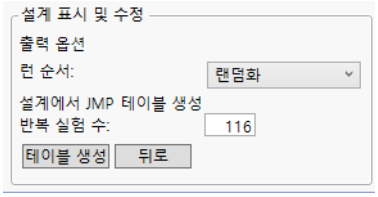
1. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다 .
2. " 요인 " 에서 추가할 요인 수에 2 를 입력하고 **추가**를 클릭합니다 .
3. 그림 13.12 에 표시된 대로 5 개 요인의 범위를 설정하고 **계속**을 클릭합니다 .

그림 13.12 5 개 요인의 범위

이름	역할	값	
X1	혼합	0.05	0.25
X2	혼합	0.1	0.3
X3	혼합	0.1	0.3
X4	혼합	0.1	0.4
X5	혼합	0.05	0.25

4. **차수** 텍스트 상자에 4 를 입력하고 **꼭지점**을 클릭합니다 .

그림 13.13 꼭지점 예의 표시 및 수정 패널



이 5 개 요인의 4 차 꼭지점 설계에 사용할 기본 런 수는 116 입니다 . 이 값은 5 개 요인에 대한 꼭지점 수이며 요인 범위에 제약 조건이 지정될 경우 요인의 최대 4 차까지의 꼭지점 평균입니다 . JMP 는 기본 혼합 집합을 후보 집합으로 사용하여 더 작은 D 최적 설계를 생성합니다 .

5. (선택 사항) 이 예의 출력과 일치시키려면 " 혼합물 설계 " 의 빨간색 삼각형을 클릭하고 " 난수 시드값 설정 " 을 선택한 후 1409 를 입력합니다 .
6. " 원하는 표본 크기 선택 :" 상자에 10 을 입력하고 **부분집합 찾기**를 클릭하여 설계를 생성합니다 .

참고 : 부분집합 찾기 옵션은 행 교환 방법 (좌표 교환 아님) 을 사용하여 행의 최적 부분집합을 찾습니다 .

그림 13.14 10 회 런 D- 최적 꼭지점 설계

설계

런	X1	X2	X3	X4	X5
1	0.05000	0.30000	0.20000	0.40000	0.05000
2	0.05000	0.10000	0.30000	0.40000	0.15000
3	0.25000	0.30000	0.10000	0.10000	0.25000
4	0.05000	0.30000	0.30000	0.10000	0.25000
5	0.25000	0.10000	0.30000	0.10000	0.25000
6	0.25000	0.30000	0.30000	0.10000	0.05000
7	0.25000	0.10000	0.30000	0.30000	0.05000
8	0.05000	0.20000	0.10000	0.40000	0.25000
9	0.25000	0.30000	0.10000	0.30000	0.05000
10	0.25000	0.10000	0.10000	0.30000	0.25000

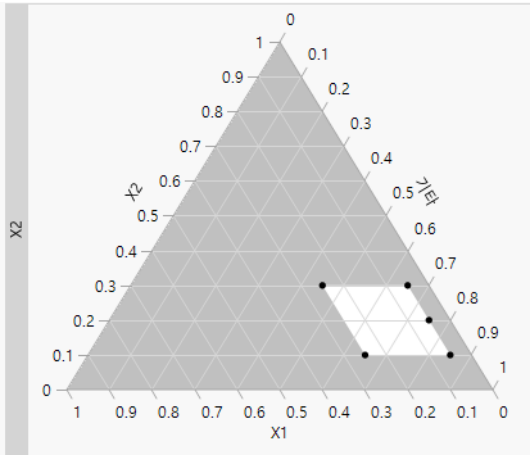
7. **테이블 생성**을 클릭합니다 .

설계 시각화

8. 설계 테이블에서 **그래프 > 삼원 그림**을 선택합니다 .

9. X1, X2, X3, X4, X5 를 선택하고 **X, 그림에 그리기**를 클릭한 후 **확인**을 클릭합니다 .

그림 13.15 5 개 요인 설계의 삼원 그림 출력 (일부분)



혼합 성분이 4 개 이상 있는 경우 삼원 그림에 일련의 그림의 표시됩니다. 각 그림에는 두 개의 성분 축이 있고 세 번째 축은 다른 모든 성분의 합입니다. 제약된 영역이 있으면 그림에 음영이 적용됩니다. 음영 처리되지 않은 부분은 실현 가능 영역을 나타냅니다. 삼원 그림에 대한 자세한 내용은 "삼원 그림 개요"(371 페이지)에서 확인하십시오.

선형 제약 조건이 있는 꼭지점의 예

Snee(1979) 및 Piepel(1988)이 제안한 전통적 예를 고려해 보겠습니다. 이 예에서는 요인 한계와 3 개의 선형 제약 조건이 있는 세 가지 요인(X1, X2, X3)을 사용합니다.

이 예에 대한 꼭지점 설계를 생성하려면

1. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다 .
2. X1, X2 및 X3 에 대해 그림 13.16 의 값을 입력하고 **계속**을 클릭합니다 .

그림 13.16 Snee 및 Piepel 예에 대한 값과 선형 제약 조건

요인

이름	역할	값	
X1	혼합	0.1	0.5
X2	혼합	0.1	0.7
X3	혼합	0	0.7

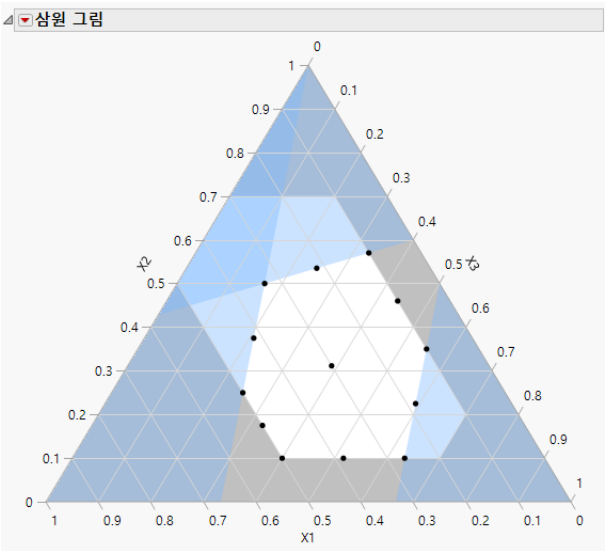
-0.85	X1 +	-0.9	X2 +	-1	X3	≤	-0.9
0.85	X1 +	0.9	X2 +	1	X3	≤	0.95
-0.7	X1 +	0	X2 +	-1	X3	≤	-0.4

3. **선형 제약 조건**을 세 번 클릭합니다. 그림 13.16 에 표시된 대로 제약 조건을 입력합니다.
4. **꼭지점** 버튼을 클릭합니다.
5. **테이블 생성**을 클릭합니다.

설계 시각화

6. 설계 테이블에서 **그래프 > 삼원 그림**을 선택합니다.
7. X1, X2, X3 을 선택하고 **X, 그림에 그리기**를 클릭한 후 **확인**을 클릭합니다.

그림 13.17 제약 조건이 있는 Piepel 예를 보여 주는 삼원 그림



삼원 그림은 요인 한계 및 선형 제약 조건의 정의에 따라 실현 가능 영역을 보여 줍니다. 설계점은 실현 가능 영역의 6 개 꼭지점, 6 개의 모서리 중간점 및 전체 중심에 있습니다. 삼원 그림에 대한 자세한 내용은 "삼원 그림 개요"(371 페이지)에서 확인하십시오.

꼭지점 방법에 대한 통계 상세 정보

선형 제약 조건이 있는 경우 JMP 는 R.E. Wheeler 가 개발한 CONSIM 알고리즘을 사용합니다. 이 알고리즘은 Snee(1979)에 설명되어 있고, Piepel(1988)에 의해 CONVRT 로 제안되었습니다. 이 방법은 Cornell(1990, Appendix 10a)에도 설명되어 있습니다. 이 방법은 제약 조건을 결합하고 꼭지점이 제약 조건을 위반하는지 확인합니다. 위반할 경우 해당 꼭지점을 삭제하고 새로 계산합니다. 중심점 수행을 위한 CONAEV 방법은 Piepel (1988)에 의해 제안되었습니다.

선형 제약 조건이 없는 경우(범위 제약 조건만 있음)에는 Snee and Marquardt(1974) 및 Snee(1975)에 의해 개발된 XVERT 방법을 사용하여 꼭지점 설계가 구성됩니다. 꼭지점을 찾은 후 심플렉스 중심 방법으로 지정된 차수까지의 꼭지점 조합을 생성합니다.

XVERT 방법은 먼저 범위가 가장 작은 $nf-1$개 요인의 주어진 하한 및 상한 값을 사용하여 전체 2^{nf-1} 설계를 생성합니다. 그런 다음 요인 값의 합이 1이 되어야 한다는 제한에 따라 제외된 한 요인의 값을 계산합니다. 요인 범위에 속하지 않는 점을 유지합니다. 그렇지 않으면 값을 증가하거나 감소하여 범위 내로 가져오고, 다른 요인을 각각 차례로 동일한 양만큼 감소하거나 증가합니다. 이 방법은 초기 제한을 충족하는 점을 유지합니다.

위의 알고리즘은 요인 제약 조건으로 정의된 심플렉스에서 실현 가능 영역의 꼭지점을 생성합니다. 그러나 Snee(1975)에서는 실현 가능 영역의 모서리와 면에 대한 중심을 갖는 것도 유용할 수 있음을 보여 주었습니다. 실현 가능 영역의 일반화된 n차원 면은 경계의 $nf-n$ 및 면 위에 있는 꼭지점의 평균으로 정의된 중심에 의해 정의됩니다. 이 알고리즘은 경계 조건의 가능한 모든 조합을 생성한 후 첫 번째 단계에서 생성된 꼭지점에 대해 평균을 계산합니다.

ABCD 설계

ABCD 설계는 혼합에 대한 선별 설계입니다. 자세한 내용은 Snee(1975)에서 확인하십시오.

ABCD 설계 예를 생성하려면

1. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다.
2. "요인"에서 추가할 요인 수에 3을 입력하고 **추가**를 클릭합니다.
3. **계속**을 클릭합니다.
4. **ABCD 설계**를 클릭합니다.

그림 13.18 28 회 런의 6 개 요인 ABCD 설계 중 14 회 런

설계

런	X1	X2	X3	X4	X5	X6
1	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000
3	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000
4	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000
5	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000
6	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
7	0.50000	0.50000	0.00000	0.00000	0.00000	0.00000
8	0.50000	0.00000	0.50000	0.00000	0.00000	0.00000
9	0.50000	0.00000	0.00000	0.50000	0.00000	0.00000
10	0.50000	0.00000	0.00000	0.00000	0.50000	0.00000
11	0.50000	0.00000	0.00000	0.00000	0.00000	0.50000
12	0.00000	0.50000	0.50000	0.00000	0.00000	0.00000
13	0.00000	0.50000	0.00000	0.50000	0.00000	0.00000
14	0.00000	0.50000	0.00000	0.00000	0.50000	0.00000

5. **테이블 생성**을 클릭합니다.

공간 채움 설계

공간 채움 혼합물 설계는 설계 영역 전체에 설계점을 분산시킵니다. 이 설계는 선형 제약 조건을 수용합니다. "DOE > 특수 용도 > 공간 채움 설계"에 있는 쾌속 유연 채움 설계 방법과 비슷한 방식으로 설계가 생성됩니다(" 쾌속 유연 채움 설계"(592 페이지) 참조).

" 혼합물 설계 " 의 빨간색 삼각형 옵션 중 다음 두 가지는 공간 채움 설계와 관련이 있습니다 .

FFF 최적 기준 쾌속 유연 채움 혼합물 설계 유형의 경우 MaxPro 기준 (기본값) 과 Centroid 기준 중 하나를 선택할 수 있습니다 . 자세한 내용은 "FFF 최적 기준 "(382 페이지) 에서 확인하십시오 .

고급 옵션 > 평균 군집 크기 설정 쾌속 유연 채움 혼합물 설계 유형의 경우 각 군집 또는 각 설계점을 정의하는 데 사용하기 위해 무작위로 생성되는 평균 점 개수를 지정할 수 있습니다 . 자세한 내용은 " 평균 군집 크기 설정 "(383 페이지) 에서 확인하십시오 .

공간 채움 예

공간 채움 설계 예를 생성하려면

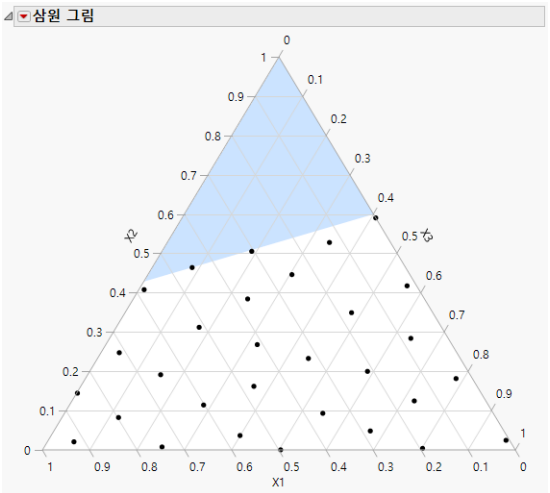
1. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다 .
2. (선택 사항) 이 예의 출력과 일치시키려면 " 혼합물 설계 " 의 빨간색 삼각형을 클릭하고 " 난수 시드값 설정 " 을 선택한 후 1409 를 입력합니다 .
3. 이 예에서는 세 가지 요인을 사용합니다 . " 요인 " 섹션을 변경할 필요가 없습니다 .
4. **계속**을 클릭합니다 .
5. **선형 제약 조건**을 클릭합니다 .
6. X1 및 X3 상자에 각각 0.7 과 1 을 입력하고 부등식 방향을 ≥ 으로 변경한 후 제약 조건 상자에 0.4 를 입력합니다 .

   이렇게 하면 $0.7X1+X3$ 이 혼합물의 40% 이상으로 제한됩니다 .
7. **런** 상자에 30 을 입력합니다 .
8. **공간 채움**을 클릭합니다 .
9. **테이블 생성**을 클릭합니다 .

설계 시각화

10. 설계 테이블에서 **그래프 > 삼원 그림**을 선택합니다 .
11. X1, X2, X3 을 선택하고 **X, 그림에 그리기**를 클릭한 후 **확인**을 클릭합니다 .

그림 13.19 선형 제약 조건이 1 개 있는 공간 채움 설계



이 설계는 Centroid FFF 최적 기준을 사용하여 구성됩니다. 점은 제약된 설계 영역에 속해 있고 이 영역 전체에 대해 상당히 잘 분산되어 있습니다. 삼원 그림에 대한 자세한 내용은 "삼원 그림 개요"(371 페이지)에서 확인하십시오.

혼합물 공간 채움 설계에 대한 통계 상세 정보

FFF 최적 기준

쾌속 유연 채움 설계를 위한 알고리즘은 지정된 설계 영역에서 다수의 랜덤 점을 생성하는 것으로 시작합니다. 그런 다음 빠른 Ward 알고리즘을 사용하여 이러한 점을 여러 군집으로 군집화합니다. 이때 군집 수는 사용자가 지정한 런 수와 같습니다.

최종 설계점은 기본 MaxPro(최대 투영) 최적 기준을 사용하거나 Centroid 기준을 선택하여 얻을 수 있습니다. 이러한 옵션은 보고서의 빨간색 삼각형 메뉴에서 "FFF 최적 기준" 아래에 있습니다.

MaxPro p 개 요인이 있고 n 이 지정된 런 수와 같은 경우 MaxPro 기준은 군집에서 다음 기준을 최소화하는 점을 찾습니다.

$$C_{MaxPro} = \sum_{i}^{n-1} \sum_{j=i+1}^{n} \left[1 / \prod_{k=1}^{p} (x_{ik} - x_{jk})^2 \right]$$

MaxPro 기준은 모든 요인을 포함하는 방식으로 잠재적 설계점 간 거리의 곱을 최대화합니다. 이는 요인 투영에 적합한 공간 채움 특성을 제공한다는 목표를 지원합니다. 자세한 내용은 Joseph et al. (2015) 에서 확인하십시오. MaxPro 옵션이 기본값입니다.

Centroid 이 방법은 각 군집의 중심에 설계점을 놓습니다. 설계 공간의 임의 점에서 가장 가까운 인접 설계점까지의 평균 거리가 다른 설계보다 작다는 특성이 있습니다.

참고 : 항상 정해진 최적 기준을 사용하도록 환경 설정을 지정할 수 있습니다 . **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다 . FFF 최적 기준을 선택한 후 기본 기준을 선택합니다 .

평균 군집 크기 설정

평균 군집 크기 설정 옵션은 "혼합물 설계"의 빨간색 삼각형 메뉴에서 "고급 옵션" 아래에 있습니다. 이 옵션을 사용하면 각 군집 또는 각 설계점을 정의하는 데 사용하기 위해 균등하게 생성되는 평균 점 개수를 지정할 수 있습니다.

기본적으로 공간 채움 설계 유형의 런 수가 200 이하인 경우 균등하게 생성된 총 10,000개의 랜덤 점이 군집화 알고리즘의 기초로 사용됩니다. 런 수가 200을 초과하는 경우 기본값은 50 입니다. 이 값을 늘리면 요인 수가 많은 설계에 특히 유용할 수 있습니다.

참고 : 요인 수와 지정된 런 값에 따라 **고급 옵션 > 평균 군집 크기 설정**을 선택하여 설계점당 평균 초기점 수를 늘릴 수 있습니다 .

선형 제약 조건

"선형 제약 조건" 개요에서 **선형 제약 조건** 옵션을 선택하여 설계 영역을 제한할 수 있습니다.

선형 제약 조건을 지정하면 군집화 알고리즘의 기초를 구성하는 랜덤 점이 제약된 설계 영역 내에 무작위로 분포됩니다. 군집화 알고리즘에 이러한 점이 사용됩니다.

혼합물 설계 적합

혼합물 설계에 대한 모형을 적합시킬 때 요인의 합이 상수라는 점을 고려해야 합니다. 기존의 완전 선형 모형은 전체를 추정할 수 없습니다.

혼합물 반응에 적합한 반응 표면 모형은 Scheffé 다항식 (Scheffé 1958) 입니다 . **선형 모형 적합**의 에서 Cox 혼합 및 Scheffé 입방 매크로에 대한 설명을 참조하십시오 . Scheffé 다항식 모형은 다음을 수행합니다 .

- 절편 표시 안 함
- 모든 선형 주효과 항 포함
- 모든 제곱 항 제외 (예 : X1*X1)
- 모든 교차 항 포함 (예 : X1*X2)

이 모형에서는 모수를 쉽게 해석할 수 있습니다(Cornell 1990). 선형 항의 계수는 혼합물이 단일 요인으로 구성된 극점에서 적합된 반응입니다. 교차 항의 계수는 요인 공간의 각 모서리에서의 곡률을 나타냅니다.

3차 다항식 항이 있는 혼합 모형의 경우 Scheffé 입방 모형을 사용할 수 있습니다. Scheffé 입방 모형은 X1*X2*(X1-X2) 형식의 항을 포함합니다.

- Scheffé 다항식 모형을 적합시키려면 " 모형 적합 " 플랫폼에서 " 혼합물 반응 표면 " 매크로를 사용합니다 .
- Scheffé 입방 모형을 적합시키려면 " 모형 적합 " 플랫폼에서 **Scheffe 입방** 매크로를 사용합니다 .
- 다른 혼합 모형을 적합시키도록 선택한 경우 " 모형 적합 " 플랫폼에 " 절편 없음 " 모형 및 " 혼합 효과 " 에 대한 옵션이 효과 속성으로 포함됩니다 .

팁 : 사용자 설계 모형 개요에는 Scheffé 입방 모형 항을 추가할 수 있는 옵션이 있습니다 . 생성된 설계 테이블에 Scheffé 입방 모형에 대한 스크립트가 포함됩니다 .

전체 모형 검정 및 분산 분석 보고서

일반적으로 JMP는 전체 모형 분산 분석 테이블에서 절편을 제외한 모든 모수가 0인지 검정합니다. 절편이 없는 혼합 모형의 경우 JMP는 선형 효과 조합이 상수라는 점에서 숨겨진 절편을 찾습니다. 숨겨진 절편을 찾으면 0 절편 모형이 아니라 절편 모형에 대한 전체 모형 검정을 수행합니다. 이 검정은 선형 모수를 제외한 모든 모수가 0인지 그리고 모수가 같은지 검정하는 것과 동등합니다.

또한 숨겨진 절편 특성을 사용하면 R^2이 결측으로 보고되지 않고 절편 모형과 관련하여 보고됩니다.

반응 표면 보고서 이해

반응 표면 효과 "&RS"로 표시된 효과가 있는 경우 JMP는 적합 반응 표면을 분석하는 추가 보고서를 생성합니다. 이 보고서는 원래 혼합 모형이 아니라 전체 반응 모형을 위해 만들어졌습니다. 그러나 JMP에서 절편이 없는 모형이 나타나면 숨긴 절편을 갖는 선형 반응 표면 항을 찾게 됩니다. 단, 제곱 항은 없습니다. 이 경우 계산을 중단하고 마지막 반응 표면 항에서 축소하여 최적의 임계값을 계산합니다. 상수를 산출하고 마지막 반응 표면 항을 포함하는 모든 조합에 대해 이를 수행할 수 있습니다.

화학 혼합물 예

세 가지 가소제 (p1, p2, p3) 가 자동차 시트 커버에 사용되는 비닐의 79.5% 를 차지합니다 (Cornell, 1990). 이 79.5% 내에서 각 가소제는 $0.409 \le x1 \le 0.849$, $0 \le x2 \le 0.252$, $0.151 \le x3 \le 0.274$ 제약 조건으로 제한됩니다 . Cornell 은 14 회 런 설계를 사용하여 2 차 혼합 모형을 수집된 반응에 적합시킵니다 . 이 설계는 4 개의 꼭지점과 전체 중심에서 반복 측정을 수행하는 3 차 꼭지점 설계입니다 .

- " 설계 생성 "
- " 혼합 모형 분석 "
- " 예측 프로파일러 "

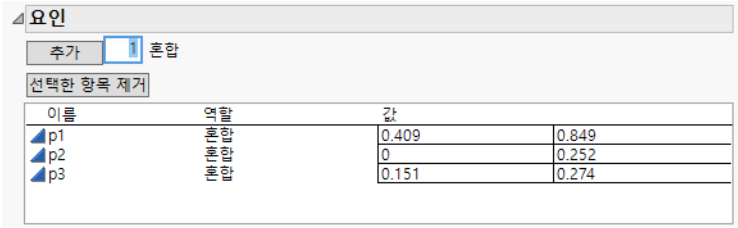
- " 혼합물 프로파일러 "

설계 생성

JMP 에서 Cornell 의 혼합물 설계를 생성하려면

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Plastifactors.jmp 를 엽니다 .
2. **DOE > 전통적 설계 > 혼합물 설계**를 선택합니다 .
3. " 혼합물 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

그림 13.20 가소제 실험에 대한 요인 및 요인 제약 조건



4. **계속**을 클릭합니다 .
5. **차수** 상자에 3 을 입력합니다 .
6. **꼭지점**을 클릭합니다 .
7. **테이블 생성**을 클릭합니다 . JMP 에서 9 개 요인 설정을 사용하여 JMP 테이블을 생성합니다 .

그림 13.21 꼭지점 혼합물 설계

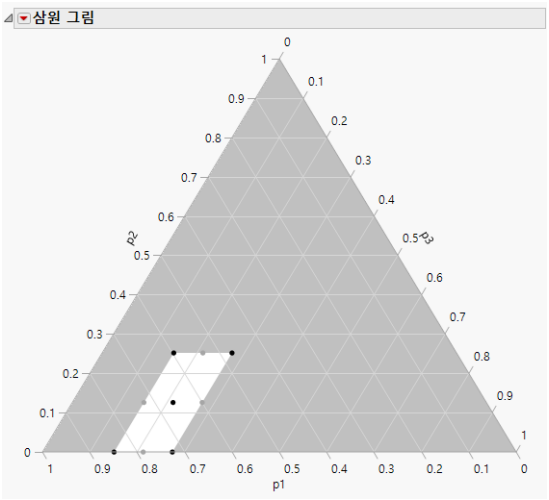
	p1	p2	p3	Y
1	0.597	0.252	0.151	•
2	0.726	0	0.274	•
3	0.6615	0.126	0.2125	•
4	0.6	0.126	0.274	•
5	0.7875	0	0.2125	•
6	0.849	0	0.151	•
7	0.5355	0.252	0.2125	•
8	0.474	0.252	0.274	•
9	0.723	0.126	0.151	•

참고 : 설계 생성 시 사용된 난수 시드값 때문에 테이블이 다를 수 있습니다 .

이제 꼭지점과 전체 중심을 복제하여 추가 설계 런을 5 회 지정한 후 14 회 런 설계를 생성합니다 .

8. 꼭지점과 중심을 식별하려면 삼원 그림을 사용합니다 . 설계 테이블에서 **그래프 > 삼원 그림**을 선택합니다 .
9. p1, p2, p3 을 선택하고 **X, 그림에 그리기**를 클릭한 후 **확인**을 클릭합니다 .

그림 13.22 설계의 삼원 그림



10. 꼭지점과 중심점을 강조 표시합니다 .

그림 13.23 꼭지점 및 중앙점 식별

	p1	p2	p3	Y
1	0.597	0.252	0.151	•
2	0.726	0	0.274	•
3	0.6615	0.126	0.2125	•
4	0.6	0.126	0.274	•
5	0.7875	0	0.2125	•
6	0.849	0	0.151	•
7	0.5355	0.252	0.2125	•
8	0.474	0.252	0.274	•
9	0.723	0.126	0.151	•

11. **편집 > 복사**를 선택하여 선택된 행을 클립보드에 복사합니다 .
12. 행 10 의 첫 번째 셀을 클릭하고 **편집 > 붙여넣기**를 선택하여 중복 행을 테이블에 추가합니다 .

참고 : 2 차 모형 , 중앙점 2 개 , 설계 반복 4 회 및 런 14 회를 포함하는 사용자 설계 도구를 사용하여 유사한 설계를 생성할 수 있습니다 .

Cornell 이 얻은 결과 (*Y* 값) 가 포함된 가소제 설계는 Plasticizer.jmp 샘플 데이터 파일에 제공됩니다 .

혼합 모형 분석

가소제 실험을 분석하려면 Plasticizer.jmp 샘플 데이터 테이블을 사용하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Plasticizer.jmp 를 엽니다 .

2. 데이터 테이블 왼쪽 위에 있는 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   완료된 " 모형 적합 " 시작 창이 나타납니다 .

참고 : 설계를 생성하면 모형 스크립트가 설계 테이블에 저장됩니다 .

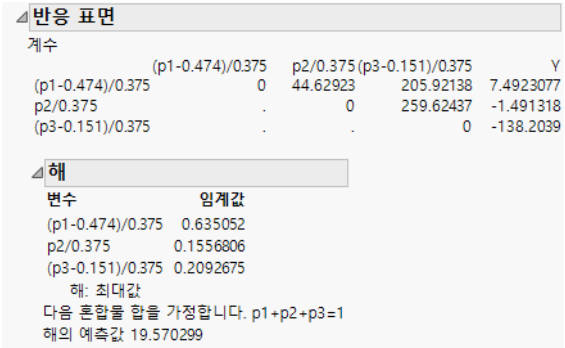
3. **실행**을 클릭하여 반응 표면 분석을 표시합니다 .
4. Plasticizer.jmp 에는 **예측 계산식** Y 열이 포함되어 있습니다 . 이 열은 분석 후 " 반응 Y" 의 빨간색 삼각형에서 **열 저장 > 예측 계산식**을 선택하여 추가되었습니다 .
5. 예측 계산식을 보려면 열 목록에서 더하기 기호를 클릭합니다 .

   0−50.1465**p1* − 282.198**p2* − 911.648**p3* + *p2*p1*317.363 + *p3*p1*1464.330 + *p3*p2*1846.218

참고 : 이 결과는 Cornell(1990) 에 보고된 계수를 수정합니다 .

반응 표면 해 보고서에 따르면 (0.63505, 0.15568, 0.20927) 점에서 최대 예측값 19.570299 가 발생합니다 .

그림 13.24 혼합물 반응 표면 분석


반응 표면

계수

	(p1-0.474)/0.375	p2/0.375	(p3-0.151)/0.375	Y
(p1-0.474)/0.375	0	44.62923	205.92138	7.4923077
p2/0.375	.	0	259.62437	-1.491318
(p3-0.151)/0.375	.	.	0	-138.2039

해

변수	임계값
(p1-0.474)/0.375	0.635052
p2/0.375	0.1556806
(p3-0.151)/0.375	0.2092675

해: 최대값

다음 혼합물 합을 가정합니다. p1+p2+p3=1

해의 예측값 19.570299

예측 프로파일러

이 보고서에는 예측 프로파일러가 포함되어 있습니다 .

1. 예측 프로파일러에서는 교차효과가 예측 트레이스에서 곡률로 표시되는 방식을 확인할 수 있습니다 . 한 성분의 값을 변경하면 (빨간색 선 드래그) 다른 두 성분의 값이 반대 방향으로 이동하여 비율과 전체 혼합 제약 조건 (성분의 합이 100% 여야 함) 을 유지합니다 .

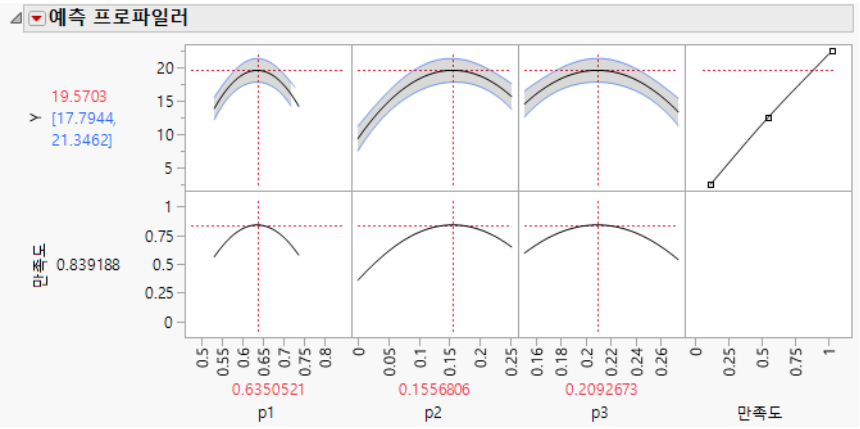
   참고 : 예측 프로파일러 트레이스의 축은 요인 (p1, p2, p3) 에 대해 정의된 상한 및 하한에서 시작합니다 . 성분 효과를 탐색할 때 두 번째 성분의 한계에 도달하면 성분을 더 이상 이동할 수 없으며 세 번째 변수만 변경되어 혼합 제약 조건을 유지합니다 .

2. 프로파일 곡선을 세 가지 모든 성분의 수준을 고려한 경계까지 표시하려면 " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **경계에서 프로파일 > 경계에서 중지**를 선택합니다 .

3. 혼합 성분을 최적화하려면 " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 함수**를 선택합니다 .
4. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택하여 최대 Y 를 위한 최적 요인 설정을 얻습니다 .

프로파일러에 표시된 최적 설정 (반올림) 은 0.6350(p1), 0.1557(p2), 0.2093(p3) 이고 이 설정은 반응 추정값 19.5703 을 제공합니다 .

그림 13.25 혼합물 분석 예에 대한 프로파일러의 최대 만족도



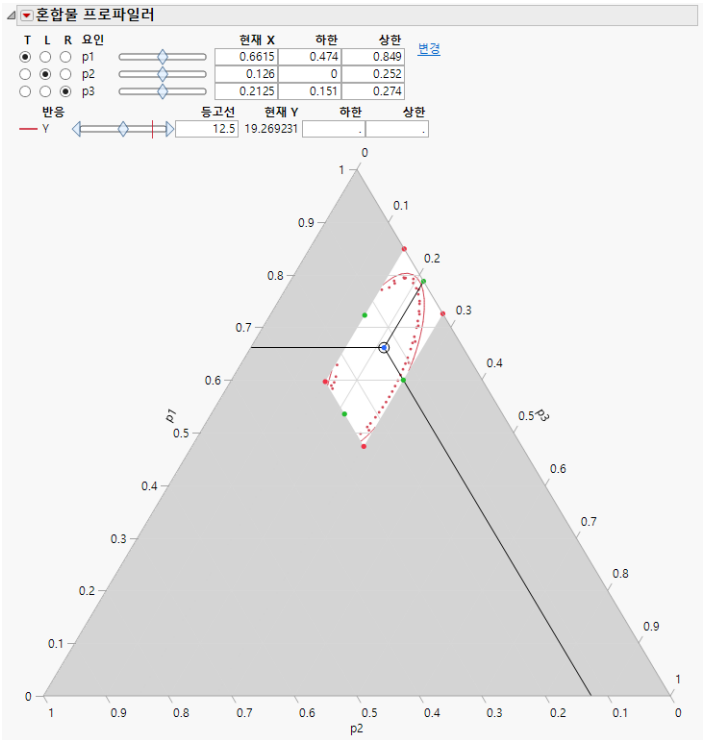
혼합물 프로파일러

" 모형 적합 " 보고서에는 혼합물 실험에서 반응 표면을 시각화하고 최적화하는 데 유용한 **혼합물 프로파일러**도 있습니다 . 대부분의 기능이 등고선 프로파일러의 기능과 동일합니다 . 그러나 다음과 같은 몇몇 기능은 혼합물 프로파일러에 고유합니다 .

- Cartesian 그림 대신 삼원 그림을 사용하여 세 가지 혼합 요인을 한 번에 볼 수 있습니다 .
- 요인이 4 개 이상인 경우 라디오 버튼을 사용하여 그림에 그릴 요인을 선택할 수 있습니다 .
- 요인에 제약 조건이 있는 경우 " 하한 " 및 " 상한 " 열에 요인의 하한과 상한을 입력할 수 있습니다 . 이 설정은 프로파일러에서 실현 불가능한 영역에 음영을 적용합니다 .

" 반응 Y" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 혼합물 프로파일러**를 선택하여 가소제 데이터의 혼합물 프로파일러를 표시합니다 .

그림 13.26 가소제 예에 대한 혼합물 프로파일러



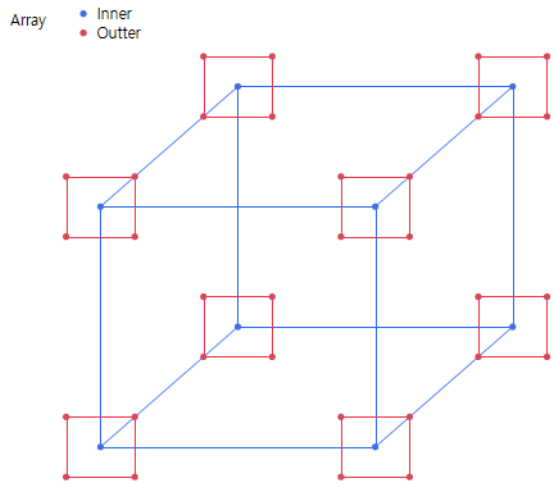
14 장

Taguchi 설계
잡음 요인이 있는 설계

Taguchi 설계는 교차 배열을 사용하여 잡음이 있는 요인을 탐색합니다. Taguchi 설계는 제어 요인을 위한 내측 배열 설계와 잡음 요인을 위한 외측 배열 설계라는 두 부분으로 나뉩니다. 전체 설계는 두 배열의 교차 곱입니다.

생산 단계에서 제어하기 어렵거나 비용이 많이 드는 잡음 요인을 실험 중에 제어해야 합니다. Taguchi 설계의 대안으로 결합 배열 및 혼합 해상도 설계(Borror and Montgomery, 2000)를 사용할 수 있습니다. 사용자 설계 도구를 사용하여 잡음 요인이 있는 설계를 생성하는 예는 "사용자 설계의 예" 장의 "로버스트 공정 및 제품 설계를 위한 실험"(190 페이지)에서 확인하십시오.

그림 14.1 제어 요인 3 개와 잡음 요인 2 개가 있는 Taguchi 설계



목차

Taguchi 설계 개요 393
Taguchi 설계의 예 393
Taguchi 설계 생성 397
    반응 397
    요인 398
    Taguchi 내측 및 외측 배열 설계 선택 398
    코드화된 설계 표시 399
    설계 테이블 생성 399
Taguchi 설계 옵션 400

Taguchi 설계 개요

Taguchi 설계는 제어 요인과 잡음 요인이라는 두 가지 유형의 요인을 사용합니다 .

- 내측 설계는 제어 요인을 위해 구성됩니다 .
- 외측 설계는 잡음 요인을 위해 구성됩니다 .

내측 및 외측 설계 런의 모든 조합에 대해 실험이 수행됩니다. 실험 중에 잡음 요인을 제어해야 합니다. 잡음 설정 전체에서 내측 배열의 각 설정에 대한 평균과 표준편차는 분석을 위해 신호 대 잡음 측도로 결합됩니다. 표 14.1에는 신호 대 잡음 성능 통계량이 나열됩니다.

표 14.1 권장 성능 통계량

목표	S/N 비 계산식
망목 (NTB)	$\frac{S}{N} = 10\log\left(\frac{\bar{Y}^2}{s^2}\right)$
망대 (LTB) - 최대화	$\frac{S}{N} = -10\log\left(\frac{1}{n}\sum_i \frac{1}{Y_i^2}\right)$
망소 (STB) - 최소화	$\frac{S}{N} = -10\log\left(\frac{1}{n}\sum_i Y_i^2\right)$

Taguchi 설계의 예

3 개의 잡음 요인에 대해 평가되는 4 개의 제어 요인을 연구하려면 Taguchi 설계를 사용합니다. 이 예는 Byrne and Taguchi(1986)에 설명된 실험입니다. 실험 목표는 나일론 튜브의 접착력 (이탈력) 을 최대화하기 위한 제어 요인의 설정을 찾는 것입니다.

이 설계에는 다음과 같은 4 개의 신호 요인이 있습니다 .

Interfer 튜브 및 커넥터 간섭 . 수준이 3 개인 신호 요인입니다 .

Wall 커넥터의 벽 두께 . 수준이 3 개인 신호 요인입니다 .

Depth 커넥터에 튜브를 삽입하는 깊이 . 수준이 3 개인 신호 요인입니다 .

Adhesive 접착제 비율 . 수준이 3 개인 신호 요인입니다 .

이 설계에는 다음과 같은 3 개의 잡음 요인이 있습니다 .

시간 처리 시간 . 수준이 2 개인 잡음 요인입니다 .

Temperature 온도 . 수준이 2 개인 잡음 요인입니다 .

Humidity 상대 습도 . 수준이 2 개인 잡음 요인입니다 .

설계 생성

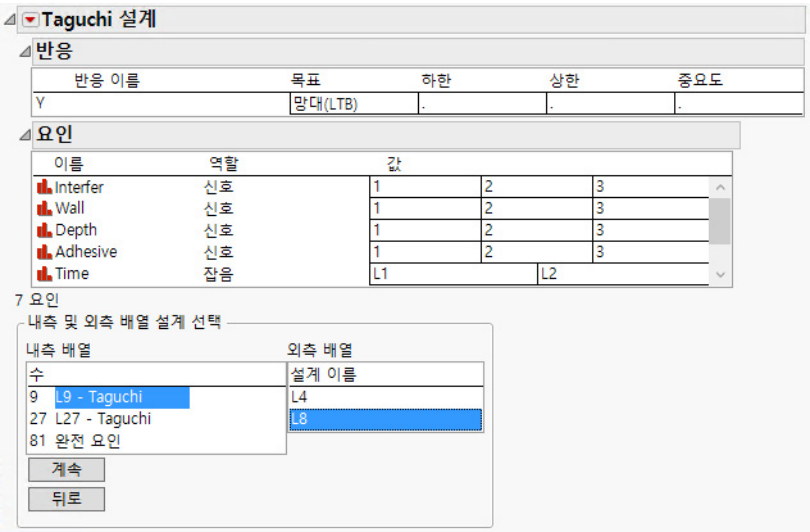
1. **DOE > 전통적 설계 > Taguchi 배열**을 선택합니다 .
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Byrne Taguchi Factors.jmp 를 엽니다 .
3. "Taguchi 배열 " 창에서 "Taguchi 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

   " 요인 " 패널에 3 수준 제어 (신호) 요인 4 개와 잡음 요인 3 개가 표시됩니다 .

   참고 : 반응에 대해 선택한 " 목표 " 에 따라 설계 데이터 테이블에 추가되는 SN 비율 계산식이 결정됩니다 .

4. 내측 배열에 **L9 - Taguchi** 가 선택되어 있는지 확인합니다 .
5. 외측 배열 설계에 대해 **L8** 을 클릭합니다 .

그림 14.2 완료된 Taguchi 설계 창



6. **계속**을 클릭합니다 .
7. **테이블 생성**을 클릭하여 그림 14.3 에 표시된 설계 테이블을 생성합니다 .

   내측 배열에는 각각 수준이 3 개인 신호 요인 4 개에 대한 9 개의 런이 있습니다 . 신호 요인 설정은 테이블의 처음 네 열에 제공됩니다 .

   외측 설계는 2 수준 잡음 요인 3 개에 대한 완전 요인 설계입니다 . 외측 배열은 데이터 테이블의 8 개 열에 제공됩니다 . 열 이름은 외측 배열 런의 패턴입니다 . 예를 들어 "---" 열은 모든 잡음 수준이 낮을 때 수집된 결과를 나타냅니다 . 9 회의 각 시도가 잡음 요인의 8 개 조합에서 수행되어 실험 시도는 총 72 회입니다 .

그림 14.3 데이터 입력 전 Taguchi 설계

	Interfer	Wall	Depth	Adhesive	패턴	−−−	−−+	−+−	−++	+−−	+−+	++−	+++	평균	SN 비
1	1	1	1	1	----	•	•	•	•	•	•	•	•	•	•
2	1	2	2	2	-000	•	•	•	•	•	•	•	•	•	•
3	1	3	3	3	-+++	•	•	•	•	•	•	•	•	•	•
4	2	1	2	3	0-0+	•	•	•	•	•	•	•	•	•	•
5	2	2	3	1	00+-	•	•	•	•	•	•	•	•	•	•
6	2	3	1	2	0+-0	•	•	•	•	•	•	•	•	•	•
7	3	1	3	2	+-+0	•	•	•	•	•	•	•	•	•	•
8	3	2	1	3	+0-+	•	•	•	•	•	•	•	•	•	•
9	3	3	2	1	++0-	•	•	•	•	•	•	•	•	•	•

이 설계 테이블에 72 개의 실험 결과를 입력할 수 있습니다 .

8. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Byrne Taguchi Data.jmp 를 엽니다 .

그림 14.4 완전한 Taguchi 설계 테이블 (Byrne Taguchi Data.jmp)

	Interfer	Wall	Depth	Adhesive	Pattern	Y---	Y--+	Y-+-	Y-++	Y+--	Y+-+	Y++-	Y+++	Mean Y	SN Ratio Y
1	1	1	1	1	----	15.6	9.5	16.9	19.9	19.6	19.6	20	19.1	17.525	24.02534
2	1	2	2	2	-000	15	16.2	19.4	19.6	19.7	19.8	24.2	21.9	19.475	25.52164
3	1	3	3	3	-+++	16.3	16.7	19.1	15.6	22.6	18.2	23.3	20.4	19.025	25.33476
4	2	1	2	3	0-0+	18.3	17.4	18.9	18.6	21	18.9	23.2	24.7	20.125	25.90425
5	2	2	3	1	00+-	19.7	18.6	19.4	25.1	25.6	21.4	27.5	25.3	22.825	26.90753
6	2	3	1	2	0+-0	16.2	16.3	20	19.8	14.7	19.6	22.5	24.7	19.225	25.32574
7	3	1	3	2	+-+0	16.4	19.1	18.4	23.6	16.8	18.6	24.3	21.6	19.85	25.71081
8	3	2	1	3	+0-+	14.2	15.6	15.1	16.8	17.8	19.6	23.2	24.4	18.3375	24.83231
9	3	3	2	1	++0-	16.1	19.9	19.3	17.3	23.1	22.7	22.6	28.6	21.2	26.15198

SN 비 Y 열은 반응을 최대화하기 위한 성능 통계량인 망대 (LTB) 입니다 . 이 값은 다음과 같이 반응의 역수 제곱 평균에 대한 상용로그의 –10 배로 계산됩니다 .

$$-10Log10\left[Mean\left[\frac{1}{(Y---)^2}, \frac{1}{(Y--+)^2}, \frac{1}{(Y-+-)^2}, \frac{1}{(Y-++)^2}, \frac{1}{(Y+--)^2}, \frac{1}{(Y+-+)^2}, \frac{1}{(Y++-)^2}, \frac{1}{(Y+++)^2}\right]\right]$$

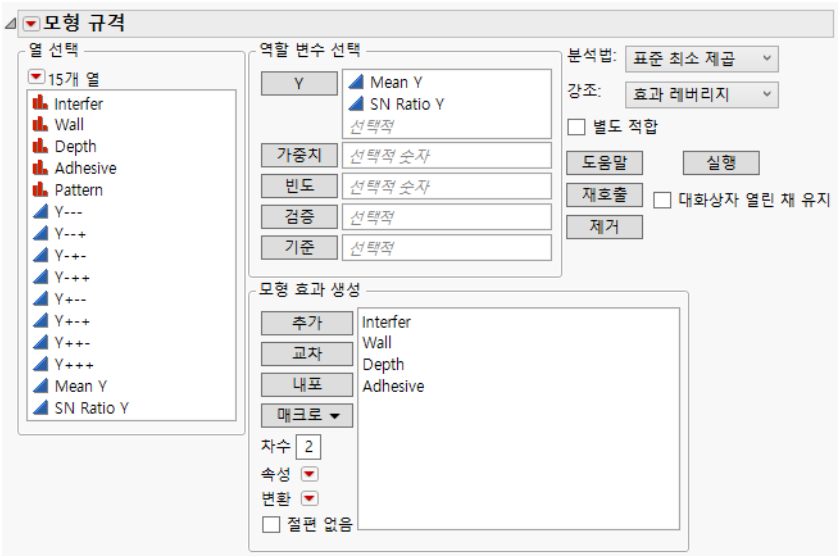
모든 개별 y 반응 값이 클 때 이 표현식도 큽니다 . 즉 , 전체 잡음 설정에서 반응이 가장 큰 신호 설정을 찾으려고 합니다 .

데이터 분석

이제 데이터를 분석할 수 있습니다 . 분석 목표는 평균과 신호 대 잡음 비를 모두 최대화하는 요인 설정을 찾는 것입니다 .

1. Byrne Taguchi Data.jmp 데이터 테이블에서 녹색 화살표를 클릭하여 "Model" 스크립트를 실행합니다 .

그림 14.5 Taguchi 데이터에 대한 모형 적합 시작 창

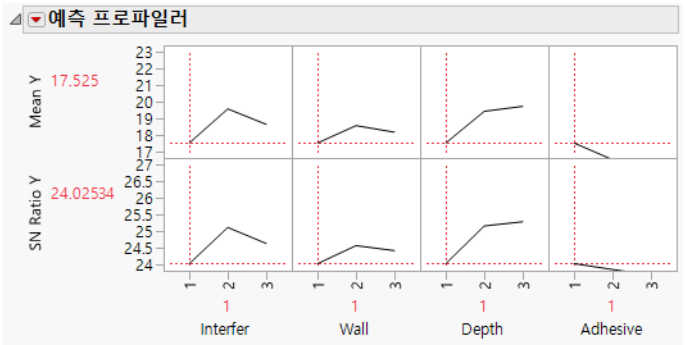


이 스크립트는 " 모형 적합 " 창을 시작합니다 . 모형에는 평균 (**평균 Y**) 및 신호 대 잡음 비 (SN 비 Y) 반응을 모델링하기 위해 네 가지 신호 요인의 주효과가 포함됩니다 .

2. **실행**을 클릭합니다 .

   보고서 맨 아래의 " 예측 프로파일러 " 를 검토하면 이 실험에서 신호 대 잡음 비의 최대값 및 해당 최대값에서의 평균 값을 제공하는 설정을 빠르게 찾을 수 있습니다 .

그림 14.6 예측 프로파일러

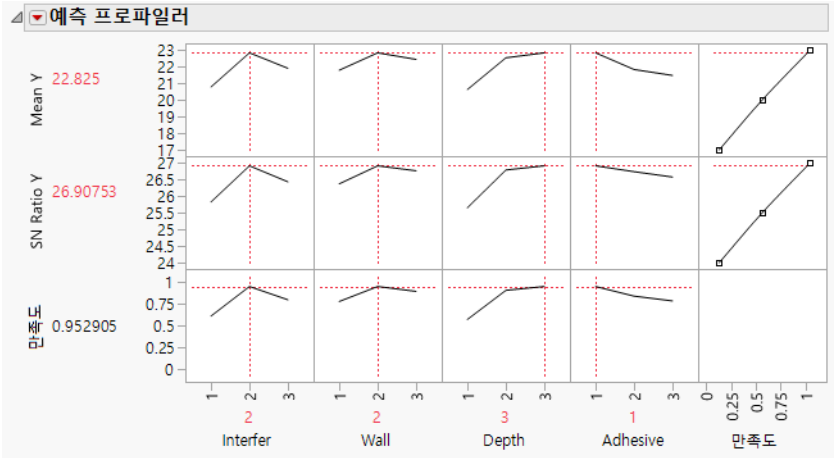


3. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 함수**를 선택합니다 .

   이렇게 하면 그림 14.7에 표시된 것과 같이 트레이스 행과 함수 설정 열이 프로파일러에 추가됩니다 . 기본 만족도 함수는 이 실험에서 원하는 " 망대 (LTB)" 로 설정됩니다 . 예측 프로파일러의 만족도 함수에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

4. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다.

그림 14.7 Byrne Taguchi 데이터에 대한 최적 요인 설정

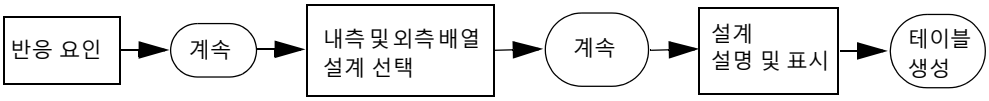


이 예에서 Interfer 및 Wall의 최적 설정은 2, Depth는 3, Adhesive는 1입니다. 이 설정은 예측 평균값 22.8과 SN 비 26.9를 산출합니다.

Taguchi 설계 생성

DOE > 전통적 설계 > Taguchi 배열을 선택하여 Taguchi 설계를 생성합니다. 먼저 반응과 요인을 정의합니다. 그런 다음 계속해서 설계 옵션, 생성 및 평가를 진행합니다. 설계 과정은 그림 14.8의 흐름을 따릅니다. 자세한 내용은 "DOE 워크플로우: 기술, 지정, 설계"(56페이지)에서 확인하십시오.

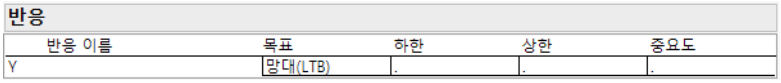
그림 14.8 Taguchi 설계 흐름



반응

"반응" 개요를 사용하여 Taguchi 설계의 반응을 지정합니다.

그림 14.9 반응 개요



" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 (기본값 Y). 이 이름을 변경하려면 이름을 두 번 클릭한 후 다른 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 에 따라 신호 대 잡음 함수의 형식이 결정됩니다 (표 14.1(393 페이지) 참조). " 망대 (LTB)", " 망목 (NTB)", " 망소 (STB)" 또는 " 없음 " 중 하나를 목표로 선택합니다 . 설계를 생성할 때 JMP 는 선택한 목표를 반영하는 SN 비 계산식을 데이터 테이블에 저장합니다 . 목표를 변경하려면 해당 목표를 클릭한 후 다른 목표를 입력합니다 . 반응에 대한 제한이 있는 경우 " 하한 " 및 " 상한 " 을 사용합니다 .

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다 . 이러한 값은 반응에 대한 가중치로 처리됩니다 . Taguchi 설계의 경우 반응이 하나뿐이므로 중요도 값은 기본적으로 1 로 설정됩니다 . 중요도 값이 필요하지 않습니다 .

요인

" 요인 " 개요를 사용하여 Taguchi 설계의 요인을 지정합니다 .

신호 두 개 이상의 2 수준 또는 3 수준 신호 요인을 지정합니다 . 신호 요인은 시스템 제어 입력 변수입니다 . 이 요인은 생산 단계에서 제어할 수 있습니다 .

잡음 하나 이상의 잡음 요인을 지정합니다 . 잡음 요인은 생산 단계에서 제어하기 어렵거나 비용이 많이 드는 변수입니다 . 그러나 실험 중에 잡음 요인을 제어할 수 있어야 합니다 . " 잡음 " 을 클릭하여 잡음 요인을 추가합니다 .

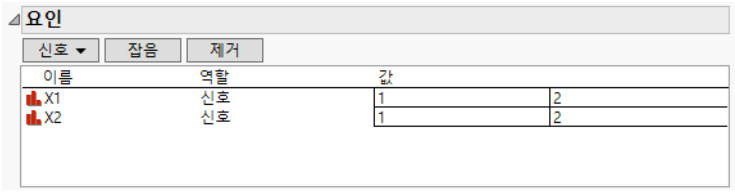
제거 선택한 요인을 제거합니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 다른 이름을 입력합니다 .

역할 요인의 설계 역할 (신호 또는 잡음) 을 지정합니다 . 역할은 요인을 추가할 때 설정됩니다 .

값 요인에 대한 실험 설정입니다 . 값을 삽입하려면 기본값을 클릭한 후 원하는 값을 입력합니다 .

그림 14.10 요인 개요



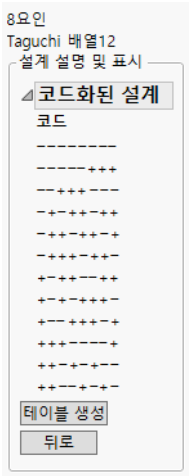
Taguchi 내측 및 외측 배열 설계 선택

"요인 지정" 패널에서 "계속"을 클릭하면 "내측 및 외측 배열 설계 선택" 옵션이 나타납니다. 신호 요인에 따라 내측 배열에 대해 직교 설계를 위한 하나 이상의 선택 옵션이 있습니다. 직교 "L" 설계를 위한 외측 배열 선택 옵션은 설계의 요인 수와 유형에 따라 다릅니다.

코드화된 설계 표시

내측 및 외측 배열에 대한 설계 유형을 선택한 후 내측 배열에 2수준 요인이 있으면 "코드화된 설계"가 표시됩니다.

그림 14.11 8 개 요인 L12 설계를 위한 코딩



"코드화된 설계"는 각 런에서 요인의 상한값과 하한값 패턴을 표시합니다.

설계 테이블 생성

테이블 생성을 Taguchi 설계Taguchi 설계클릭하면 그림 14.12에 표시된 것과 유사한 Taguchi 설계 테이블이 나타납니다. 데이터 테이블에서 각 행은 내측 배열 런을 나타냅니다. **패턴** 변수의 값에서 더하기 기호는 상한을 나타내고 빼기 기호는 하한을 나타냅니다. 또한 각 요인에 대한 열과 "패턴" 열에 나타나는 각 패턴에 대한 열도 테이블에 포함됩니다. 패턴 열은 실험에 대한 반응을 포함합니다. "Taguchi 설계"의 빨간색 삼각형 메뉴에서 "반응 시뮬레이션"을 선택한 경우에는 시뮬레이션된 반응 값이 패턴 열에 채워집니다.

그림 14.12 8 개 요인 L12 설계에 대한 Taguchi 설계 테이블

	X1	X2	X3	X4	X5	X6	X7	X8	패턴	run 1	run 2	평균	SN 비
1	1	1	1	1	1	1	1	1	--------	•	•	•	•
2	1	1	1	1	1	2	2	2	-----+++	•	•	•	•
3	1	1	2	2	2	1	1	1	--+++---	•	•	•	•
4	1	2	1	2	2	1	2	2	-+-++-++	•	•	•	•
5	1	2	2	1	2	2	1	2	-++-++-+	•	•	•	•
6	1	2	2	2	1	2	2	1	-+++-++-	•	•	•	•
7	2	1	2	2	1	1	2	2	+--++--++	•	•	•	•
8	2	1	2	1	2	2	2	1	+-+-+++-	•	•	•	•
9	2	1	1	2	2	2	1	2	+--+++-+	•	•	•	•
10	2	2	2	1	1	1	1	2	+++----+	•	•	•	•
11	2	2	1	2	1	2	1	1	++-+-+--	•	•	•	•
12	2	2	1	1	2	1	2	1	++--+-+-	•	•	•	•

Taguchi 설계 옵션

"Taguchi 설계"의 빨간색 삼각형 메뉴에는 설계 설정 및 생성에 대한 옵션이 포함되어 있습니다.

반응 저장 "반응" 패널의 정보를 새 데이터 테이블에 저장합니다. 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다. 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다.

반응 불러오기 "반응 저장" 옵션을 사용하여 저장한 반응을 불러옵니다.

요인 저장 "요인" 패널의 정보를 새 데이터 테이블에 저장합니다. 각 요인의 열에는 수준이 포함됩니다. 다른 정보는 열 특성으로 저장됩니다. 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다. 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다. 그런 다음 "설계 역할" 영역에서 적절한 역할을 선택합니다.

요인 불러오기 "요인 저장" 옵션을 사용하여 저장한 요인을 불러옵니다.

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) "요인 제약 조건 정의" 또는 "선형 제약 조건" 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다. 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다. 마지막 행은 부등식 한계를 포함합니다. 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다. 자세한 내용은 "열 특성" 부록의 "제약 조건 상태"(762 페이지)에서 확인하십시오.

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) "제약 조건 저장" 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다.

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다. 이러한 작업에는 다음 중 하나 이상이 포함됩니다.

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다.

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자" 스크립트에 포함됩니다.

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다. "테이블 생성"을 클릭하기 전에 이 옵션을 선택하십시오.

"테이블 생성"을 클릭하면 다음과 같은 결과가 나타납니다.

– 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다 .
– 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
– 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다 .

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다 .

– **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다. 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 **기본 분석**의 " 시뮬레이션 " 장에서 확인하십시오 .

15 장

설계 평가
설계 특성 탐색

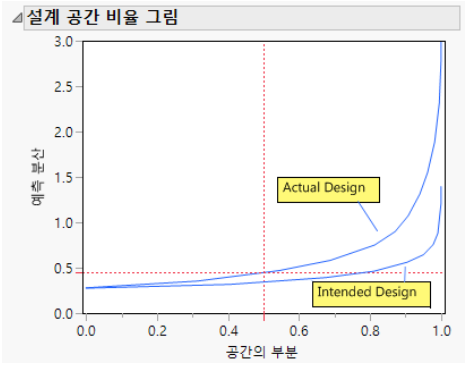
"설계 평가" 플랫폼을 사용하면 다음을 수행할 수 있습니다.

- 기존 실험 설계의 장점과 한계를 확인합니다.
- 설계에서 반응에 대한 중요한 변화와 관련된 효과를 감지할 수 있는지 판별합니다.
- 추정값의 정밀도와 예측 분산을 처리합니다.
- 별칭 구조를 파악합니다.
- 효율도 측도를 얻습니다.

설계 평가 플랫폼은 여러 DOE 플랫폼에서 제공한 "설계 평가" 개요에 나타나는 결과를 생성합니다. "설계 평가" 개요에 제공되는 설계 진단은 다음과 같습니다.

- 검정력 분석
- 예측 분산 프로파일러 및 표면 그림
- 설계 공간 비율 그림 - 설계 공간에서 지정된 값을 초과하는 예측 분산의 정도 표시
- 모수 추정값을 위한 추정 효율도 측도
- 별칭 행렬 - 모형 효과에 대한 편향 구조 표시
- 색상 맵 - 효과 간의 절대 상관 표시
- 설계 효율도 값

그림 15.1 두 개의 설계 공간 비율 그림 비교



목차

설계 평가 플랫폼 개요 405
설계 평가의 예 405
손실된 런의 영향 평가 406
지정된 모형에 대한 검정력 평가 413
설계 평가 시작 창 415
설계 평가 창 415
요인 416
모형 417
별칭 항 418
설계 418
설계 평가 418
검정력 분석 419
예측 분산 프로파일 425
설계 공간 비율 그림 426
예측 분산 표면 427
추정 효율도 428
별칭 행렬 430
상관 색상 맵 431
설계 진단 432
설계 평가 옵션 434

설계 평가 플랫폼 개요

설계 평가 플랫폼은 설계의 특성을 평가할 수 있는 강력한 도구를 제공합니다. 설계를 JMP로 생성했는지 다른 도구로 생성했는지는 상관없습니다. 실험을 수행하기 전에 플랫폼을 사용하여 여러 설계 중에서 선택할 수 있습니다. 수행한 설계에서 올바르지 않은 설정이나 손실된 런으로 인한 영향도 평가할 수 있습니다. 가정된 모형의 항을 수정하여 수정된 모형을 추정할 때의 영향을 확인할 수 있습니다. 또한 "별칭 항" 개요에 나타나는 항을 수정하여 별칭 행렬에 미치는 영향을 확인할 수도 있습니다.

시작 창에 설계 정보를 입력하는 것으로 시작합니다 . 그런 다음 가정된 모형을 수정하고 , 모형에 없는 효과 중 잠재적 관심 대상인 효과를 지정할 수 있습니다 . 설계 평가 플랫폼은 사용자가 지정한 내용에 따라 생성된 설계의 특성을 평가하는 다양한 방법을 제공합니다 .

검정력 분석 주어진 크기의 효과를 감지할 수 있는지 분석할 수 있습니다 .

예측 분산 프로파일 요인 설정 범위에 대한 예측 분산을 보여 줍니다 .

설계 공간 비율 그림 모형 예측 분산이 지정된 값보다 얼마나 아래 또는 위에 있는지 보여 줍니다 .

예측 분산 표면 두 개의 연속형 요인에 대한 예측 분산의 표면 그림을 보여 줍니다 .

추정 효율도 각 모수에 대해 이상적인 직교 설계 (존재하지 않을 수 있음) 와 비교하여 신뢰 구간 길이에 대한 증가 비율을 제공합니다 . 모수의 상대 표준 오차도 제공합니다 .

별칭 행렬 모형에 없지만 활성화 가능성이 있는 효과에 의해 모형 모수가 편향되는 정도를 나타내는 계수를 제공합니다 .

상관 색상 맵 강도 척도를 사용하여 그림에서 효과 사이의 절대 상관을 표시합니다 .

참고 : 기본 강도 척도는 회색 척도입니다 . 색상 척도를 변경하려면 " 상관 색상 맵 " 의 빨간색 삼각형을 클릭하고 **파랑 - 회색 - 빨강**을 선택합니다 . 사용자 정의 색상 척도를 사용하려면 그림을 마우스 오른쪽 버튼으로 클릭하고 **색상 테마**를 선택합니다 .

설계 진단 설계의 효율도 측도를 제공합니다.

참고 : 몇 가지 DOE 플랫폼에서는 설계를 구성할 때 " 설계 평가 " 개요가 나타납니다 . 이 개요는 " 설계 평가 " 플랫폼에서 제공한 결과를 보여 줍니다 . " 설계 평가 " 개요를 제공하는 플랫폼은 사용자 설계 , 확증적 선별 설계 , 선별 설계 , 반응 표면 설계 , 최적 설계 유형을 사용한 혼합물 설계입니다 .

설계 평가의 예

- " 손실된 런의 영향 평가 "
- " 지정된 모형에 대한 검정력 평가 "

손실된 런의 영향 평가

테니스 공의 탄력성(Stretch)에 대한 세 가지 요인(Silica, Sulfur, Silane)의 효과를 탐색하기 위한 실험이 수행되었습니다. 실험 목표는 Stretch에 대한 예측 모형을 개발하는 것입니다. 반응 표면 설계 플랫폼을 사용하여 15회 런 Box-Behnken 설계가 선택되었습니다. 실험 후 연구원은 Silica = 0.7, Silane = 50인 두 번의 런이 올바르게 처리되지 않았음을 알게 되었습니다. 이러한 런은 데이터 분석에 포함할 수 없습니다.

설계 평가 플랫폼을 사용하여 이 두 런이 포함되지 않을 경우에 대한 영향을 평가합니다. 의도한 15회 런 설계의 진단 결과를 얻은 후 이를 두 런이 없는 실제 13회 런 설계와 비교합니다.

의도한 설계와 실제 설계 구성

의도한 설계

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Bounce Data.jmp 를 엽니다.
2. **DOE > 설계 진단 > 설계 평가**를 선택합니다.
3. Silica, Sulfur, Silane 을 선택하고 **X, 요인**을 클릭합니다.

   원할 경우 Stretch 를 **Y, 반응**으로 추가할 수 있습니다. 그러나 반응을 지정해도 설계 특성에 영향을 미치지 않습니다.
4. **확인**을 클릭합니다.

의도한 설계의 "설계 평가" 창을 열어 두십시오.

팁 : 의도한 설계의 " 설계 평가 " 창을 화면 왼쪽 영역에 두십시오. 다음 단계 후 실제 설계에 대한 평가 창을 오른쪽에 둡니다.

누락된 런이 있는 실제 설계

이 섹션에서는 Silica = 0.7, Silane = 50 인 두 런을 제외합니다. 해당 런은 데이터 테이블의 3 행과 7 행입니다.

1. Bounce Data.jmp 에서 3 행과 7 행을 선택하고 강조 표시된 영역에서 마우스 오른쪽 버튼을 클릭한 후 **숨기고 제외하기**를 선택합니다.
2. **DOE > 설계 진단 > 설계 평가**를 선택합니다.
3. **재호출**을 클릭합니다.
4. **확인**을 클릭합니다.

실제 설계의 "설계 평가" 창을 열어 두십시오.

팁 : 두 설계를 쉽게 비교할 수 있도록 실제 설계의 " 설계 평가 " 창을 의도한 설계의 " 설계 평가 " 창 오른쪽에 두십시오.

의도한 설계와 실제 설계 비교

이제 다음 방법을 사용하여 두 설계를 비교할 수 있습니다 .

- " 검정력 분석 "(407 페이지)
- " 예측 분산 프로파일 "(408 페이지)
- " 설계 공간 비율 그림 "(410 페이지)
- " 추정 효율도 "(411 페이지)
- " 상관 색상 맵 "(412 페이지)
- " 설계 진단 "(412 페이지)

검정력 분석

각 창에서 다음을 수행합니다 .

1. " 검정력 분석 " 개요를 엽니다 .

   이 개요에는 모든 " 예상 계수 " 의 값이 기본값 1 로 표시됩니다 . 이 값은 교호작용 항과 2 차 항이 비활성이라는 가정하에 주효과 항의 전체 수준에서 2 단위의 예상 반응 변화를 감지하는 것에 해당합니다 .

   검정력 계산은 오차 항 (예상 RMSE) 을 1 로 가정합니다 . 이전 연구를 통해 RMSE 를 대략 2 로 간주합니다 .

2. **예상 RMSE** 옆에 2 를 입력합니다 .

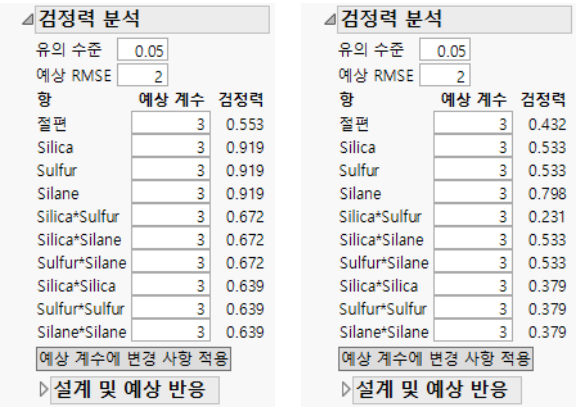
   텍스트 상자 바깥쪽을 클릭하면 검정력 값이 업데이트됩니다 .

   교호작용 항과 2 차 항이 비활성이라는 가정하에 주효과의 전체 수준에서 약 6 단위의 예상 반응 차이를 감지하려고 합니다. 이를 균등하게 설정하기 위해 빨간색 삼각형 옵션을 사용합니다 .

3. " 검정력 분석 " 의 빨간색 삼각형을 클릭하고 **검정력을 위한 델타 설정**을 선택합니다 .
4. 델타 값을 6 으로 입력합니다 .
5. **확인**을 클릭합니다 .

그림 15.2에서는 "설계 및 예상 반응" 개요가 닫혀 있는 두 개요를 보여 줍니다.

그림 15.2 검정력 분석 개요 - 의도한 설계 (왼쪽), 실제 설계 (오른쪽)



검정력 분석

유의 수준 0.05
예상 RMSE 2

항	예상 계수	검정력
절편	3	0.553
Silica	3	0.919
Sulfur	3	0.919
Silane	3	0.919
Silica*Sulfur	3	0.672
Silica*Silane	3	0.672
Sulfur*Silane	3	0.672
Silica*Silica	3	0.639
Sulfur*Sulfur	3	0.639
Silane*Silane	3	0.639

예상 계수에 변경 사항 적용
설계 및 예상 반응

검정력 분석

유의 수준 0.05
예상 RMSE 2

항	예상 계수	검정력
절편	3	0.432
Silica	3	0.533
Sulfur	3	0.533
Silane	3	0.798
Silica*Sulfur	3	0.231
Silica*Silane	3	0.533
Sulfur*Silane	3	0.533
Silica*Silica	3	0.379
Sulfur*Sulfur	3	0.379
Silane*Silane	3	0.379

예상 계수에 변경 사항 적용
설계 및 예상 반응

실제 설계의 검정력 값은 의도한 설계의 값과 비교할 때 균등하게 작습니다. Silica 와 Sulfur 의 경우 의도한 설계의 검정력이 실제 설계의 검정력에 비해 거의 두 배입니다. Silica*Sulfur 교호작용의 경우 의도한 설계의 검정력은 0.672 인데 비해 실제 설계의 검정력은 0.231 입니다. 실제 설계는 의도한 설계와 비교하여 검정력 손실이 상당히 큽니다.

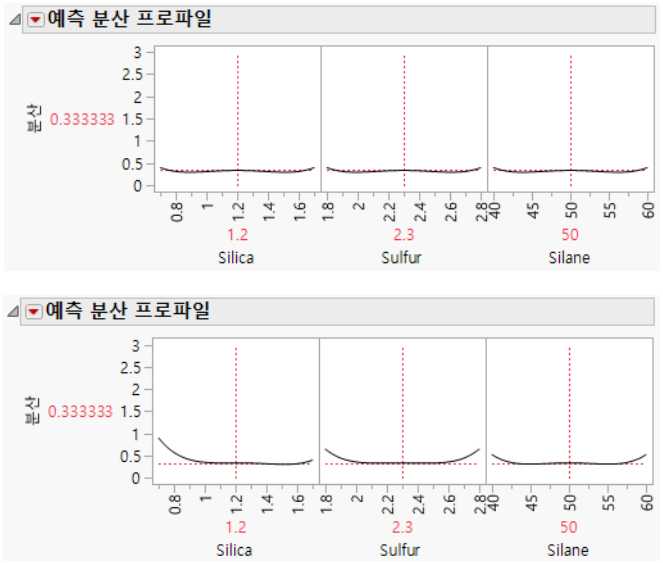
예측 분산 프로파일

1. 각 창에서 " 예측 분산 프로파일 " 개요를 엽니다 .
2. 실제 설계 창에서 수직 축을 마우스 오른쪽 버튼으로 클릭하고 메뉴에서 **편집 > 축 설정 복사** 를 선택합니다 .

   이렇게 하면 축 설정을 포함하는 스크립트가 생성됩니다 . 이제 이 축 설정을 의도한 설계의 " 예측 분산 프로파일 " 그림에 적용합니다 .
3. 의도한 설계의 " 설계 평가 " 창에서 " 예측 분산 프로파일 " 개요를 찾은 후 세로 축에 대한 척도에 커서를 놓습니다 . 수직 축을 마우스 오른쪽 버튼으로 클릭하고 메뉴에서 **편집 > 축 설정 붙여넣기** 를 선택합니다 .

   그림 15.5 에 표시된 그림을 참조하십시오 . 의도한 설계에 대한 그림이 위에 있고 실제 설계에 대한 그림이 아래에 있습니다 .

그림 15.3 예측 분산 프로파일 - 의도한 설계 (위), 실제 설계 (아래)



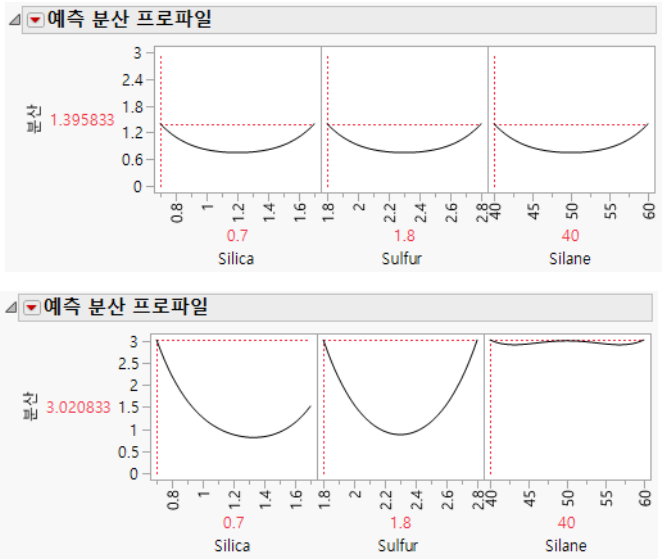
" 예측 분산 프로파일 " 그림은 상대 예측 분산의 프로파일러 보기입니다 . 설계 공간의 다양한 영역에서 상대 예측 분산을 탐색할 수 있습니다 .

두 그림 모두 설계 공간의 중심에서 동일한 상대 예측 분산 값을 보여 줍니다 . 그러나 설계 공간 가장자리 근처의 점에 대한 분산은 의도한 설계의 동일한 점에 비해 크게 나타납니다 . 세 개의 수직선을 모두 요인 설정 가장자리 근처의 점으로 움직여 이 현상을 살펴보십시오 .

4. 두 창에서 " 예측 분산 프로파일 " 의 빨간색 삼각형을 클릭하고 **분산 최대화**를 선택합니다 .

   그림 15.4 에서는 의도한 설계와 실제 설계에 대한 최대 상대 예측 분산을 보여 줍니다 .

그림 15.4 최대화된 예측 분산 프로파일 - 의도한 설계 (위), 실제 설계 (아래)



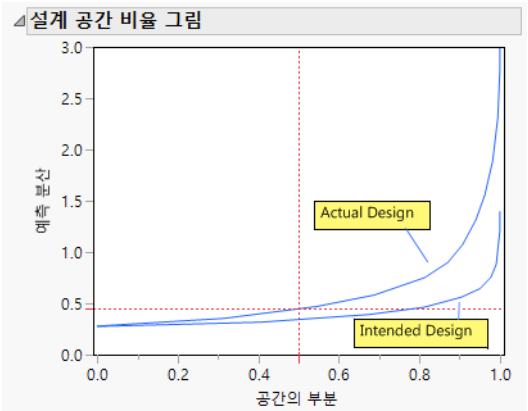
두 설계 모두 동일한 점 (Silica=0.7, Sulfur=1.8, Silane=40) 을 최대 예측 분산이 발생하는 설계 점 중 하나로 식별합니다. 최대 예측 분산은 의도한 설계의 경우 1.396 이고 실제 설계의 경우 3.021 입니다. 예측 분산이 최대화되는 다른 실험점도 있습니다. 실제 설계의 최대 예측 분산이 크면 설계 공간의 일부에서 의도한 설계를 수행했을 때보다 예측 정확도가 낮아진다는 것을 의미합니다.

설계 공간 비율 그림

1. 각 창에서 " 설계 공간 비율 그림 " 개요를 엽니다.
2. 의도한 설계 창에서 그림을 마우스 오른쪽 버튼으로 클릭하고 **편집 > 프레임 내용 복사**를 선택합니다.
3. 실제 설계 창에서 " 설계 공간 비율 그림 " 개요를 찾습니다.
4. 그림을 마우스 오른쪽 버튼으로 클릭하고 **편집 > 프레임 내용 붙여넣기**를 선택합니다.

그림 15.5 에서는 주석이 추가된 그림을 보여 줍니다. 각 " 설계 공간 비율 그림 " 은 설계 공간에서 상대 예측 분산이 특정 값 미만이 되는 비율을 보여 줍니다.

그림 15.5 설계 공간 비율 그림



실제 설계의 상대 예측 분산은 전체 설계 공간에 대해 의도한 설계의 분산보다 큽니다. 설계 공간 범위가 클수록 차이가 증가합니다.

추정 효율도

각 창에서 "추정 효율도" 개요를 엽니다.

그림 15.6 추정 효율도 개요 - 의도한 설계 (왼쪽), 실제 설계 (오른쪽)

추정 효율도

항	CI 길이의 부분적 증가	추정값의 상대 표준 오차
절편	1.236	0.577
Silica	0.369	0.354
Silane	0.369	0.354
Sulfur	0.369	0.354
Silica*Silane	0.936	0.5
Silica*Sulfur	0.936	0.5
Sulfur*Silane	0.936	0.5
Silica*Silica	1.016	0.52
Silane*Silane	1.016	0.52
Sulfur*Sulfur	1.016	0.52

추정 효율도

항	CI 길이의 부분적 증가	추정값의 상대 표준 오차
절편	1.082	0.577
Silica	0.803	0.5
Sulfur	0.803	0.5
Silane	0.275	0.354
Silica*Sulfur	2.122	0.866
Silica*Silane	0.803	0.5
Sulfur*Silane	0.803	0.5
Silica*Silica	1.268	0.629
Sulfur*Sulfur	1.268	0.629
Silane*Silane	1.268	0.629

실제 설계(오른쪽)에서 모든 모수의 상대 표준 오차가 의도한 설계(왼쪽)의 표준 오차를 초과하거나 같습니다. 비절편 모수 3개를 제외한 모든 경우에 대해 실제 설계의 상대 표준 오차가 의도한 설계의 표준 오차를 초과합니다.

"CI 길이의 증가 비율"은 현재 설계에서 제공한 모수 신뢰 구간의 길이를 런 크기가 동일한 이상적 설계에서 제공한 신뢰 구간의 길이와 비교합니다. 신뢰 구간 길이와 이에 따른 CI 길이의 증가 비율은 런 수의 영향을 받습니다. 자세한 내용은 "CI 길이의 증가 비율"(428 페이지)에서 확인하십시오. 실제 설계의 경우 런 수가 감소했지만 Silane, Silica*Silane 및 Sulfur*Silane 항은 의도한 설계에 비해 증가 정도가 적습니다. 실제 설계를 정의하기 위해 제거된 두 런에서 Silane이 중앙점으로 설정되었기 때문입니다. 이러한 런을 제거한 결과, 해당 모수에 대한 신뢰 구간의 너비가 중앙점이 없는 이상적 직교 설계의 너비와 더 유사합니다.

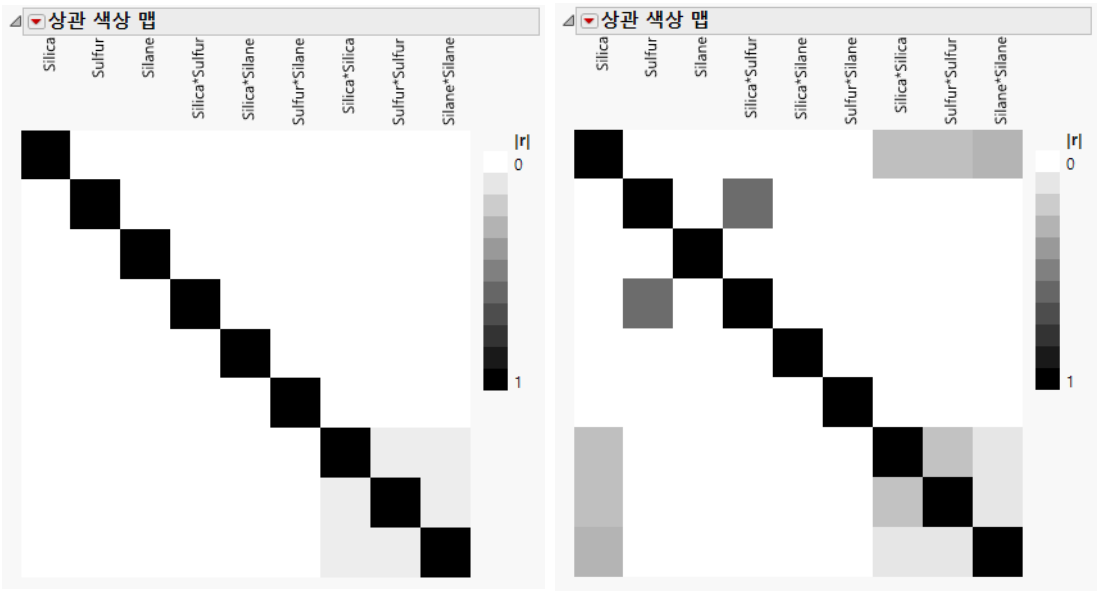
상관 색상 맵

각 보고서에서 다음을 수행합니다 .

1. **상관 색상 맵** 개요를 엽니다 .

   두 색상 맵은 " 모형 " 개요의 효과를 보여 줍니다 . 각 그림은 JMP 기본 강도 척도 (흰색 ~ 검은색) 를 사용하여 색상이 적용된 효과 간 절대 상관을 보여 줍니다 . 효과 간 상관이 0 이거나 매우 작으면 이상적입니다 .

그림 15.7 상관 색상 맵 - 의도한 설계 (왼쪽), 실제 설계 (오른쪽)



절대 상관 값의 범위는 0(흰색)에서 1(검은색)까지입니다. 셀을 커서로 가리키면 절대 상관 값이 표시됩니다. 실제 설계의 색상 맵은 의도한 설계의 색상 맵보다 큰 절대 상관을 보여 줍니다. 예를 들어 Sulfur 와 Silica*Sulfur 간 상관이 의도한 설계의 경우 < .0001 이고 실제 설계의 경우 0.5774 입니다.

설계 진단

각 보고서에서 "설계 진단" 개요를 엽니다.

그림 15.8 설계 진단 - 의도한 설계 (왼쪽), 실제 설계 (오른쪽)

의도한 설계(왼쪽)는 실제 설계(오른쪽)보다 효율도 값이 높고 평균 예측 분산이 낮습니다. 설계 평가를 분석한 결과, 손실된 두 런이 설계에 좋지 않은 영향을 미쳤음을 알 수 있습니다.

효율도 측도를 계산할 때 런 수와 모형 행렬 둘 다 고려된다는 점에 유의하십시오. 특히 *D* 효율도, *G* 효율도 및 *A* 효율도는 주어진 설계의 런 크기에 이상적인 설계를 기준으로 계산됩니다. 규모가 큰 설계가 작은 설계보다 효율적이지만 반드시 그런 것은 아닙니다. 그러나 주어진 수의 요인에 대해 설계가 크면 작은 설계보다 예측 평균 분산 값이 더 작은 경향이 있습니다. 효율도 측도를 정의하는 방법에 대한 자세한 내용은 "설계 진단"(432 페이지)에서 확인하십시오.

지정된 모형에 대한 검정력 평가

이 예에서는 6 개 요인 중 추출 공정의 수율에 영향을 미치는 요인을 판별하기 위해 확증적 선별 설계를 구성했습니다. 데이터는 Design Experiment 폴더에 있는 Extraction Data.jmp 샘플 데이터 테이블에 제공됩니다. 설계가 **확증적** 선별 설계이므로 각 요인의 수준은 3 개입니다. 자세한 내용은 "확증적 선별 설계" 장(219 페이지)에서 확인하십시오.

강력한 2 차 효과를 감지하기 위한 검정력에 관심이 있습니다. 강한 효과는 크기가 오차 변동의 3 배 이상인 것으로 간주합니다.

실험에서 6 개 요인을 연구하지만 효과 희소성에 따라 몇 가지 요인만 활성이라는 것을 알 수 있습니다. 결과적으로 모형에서 더 적은 수의 요인을 기반으로 검정력을 조사할 수 있습니다. 또한 관련 공정에 대한 과거 연구에서는 세 가지 요인(Propanol, Butanol, pH)에 무시할 수 있는 주효과가 있고, 다른 요인과의 교호작용이 없으며, 2 차 효과가 없다는 강력한 증거를 제시합니다. 이를 통해 예상 모형에 Methanol, Ethanol 및 Time 에 대해서만 주효과, 교호작용 효과 및 2 차 효과가 포함된다고 간주할 수 있습니다. 3 요인 반응 표면 모형의 컨텍스트에서 검정력을 조사하기로 결정했습니다.

설계 평가 플랫폼을 사용하여 Methanol, Ethanol 또는 Time 의 강력한 2 차 효과를 감지하기 위한 설계 검정력을 확인할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Extraction Data.jmp 를 엽니다 .
2. **DOE > 설계 진단 > 설계 평가**를 선택합니다 .
3. Methanol, Ethanol 및 Time 을 선택하고 **X, 요인**을 클릭합니다 .

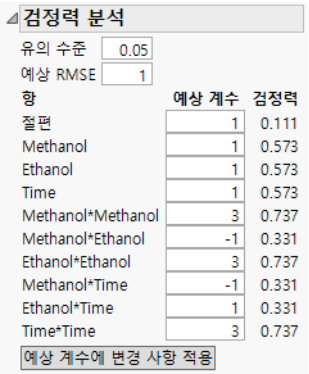
   원할 경우 Yield 를 **Y, 반응**으로 추가할 수 있습니다 . 그러나 반응을 지정해도 설계 특성에 영향을 미치지 않습니다 .
4. **확인**을 클릭합니다 .
5. " 모형 " 개요에서 **RSM** 을 클릭합니다 .

   이렇게 하면 세 요인에 대한 교호작용 항과 2 차 항이 추가됩니다 .
6. " 검정력 분석 " 개요를 엽니다 .

   " 예상 RMSE" 가 기본적으로 1 로 설정되어 있습니다 . 과거 연구에서 얻은 RMSE 추정값이 있지만 입력하지 않아도 됩니다 . 관심 효과의 크기가 오차 변동의 3 배이기 때문입니다 .
7. **예상 계수**에서 Methanol*Methanol, Ethanol*Ethanol 및 Time*Time 옆에 3 을 입력합니다 .

8. **예상 계수에 변경 사항 적용**을 클릭합니다 .

그림 15.9 계수에 변경 사항을 적용한 후의 검정력 분석 개요

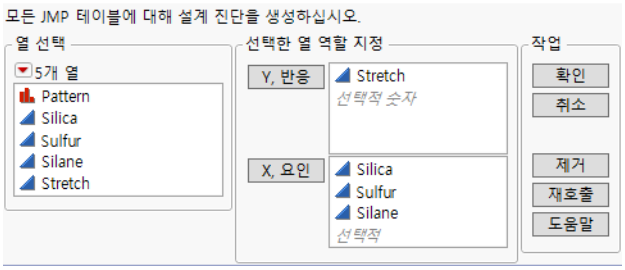


크기가 오차 변동의 3 배인 2 차 효과를 감지하는 검정력은 0.737 입니다 . 이는 세 가지 요인의 반응 표면인 최종 모형을 가정합니다 . 또한 검정의 유의 수준을 0.05 로 가정합니다 .

설계 평가 시작 창

설계 평가 플랫폼을 시작하려면 대상 데이터 테이블을 열고 **DOE > 설계 진단 > 설계 평가**를 선택합니다 . 그림 15.10 의 예에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 샘플 데이터 테이블을 사용합니다 .

그림 15.10 설계 평가 시작 창



시작 창에는 다음과 같은 버튼이 포함되어 있습니다 .

Y, 반응 하나 이상의 반응 열을 입력합니다 . 반응 입력은 선택 사항입니다 . 반응 값은 설계 평가에 사용되지 않습니다 . 반응은 숫자여야 합니다 .

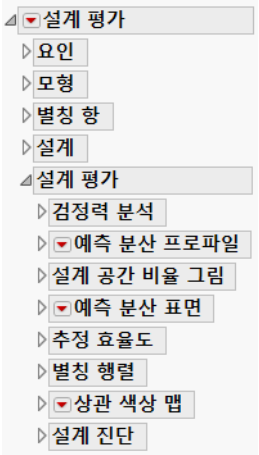
X, 요인 요인 열을 입력합니다 . 요인은 모든 데이터 유형 또는 모델링 유형이 될 수 있습니다 .

설계 평가 창

" 설계 평가 " 창은 두 부분으로 구성됩니다 . 그림 15.11 을 참조하십시오 . 이 그림에는 모든 개요 노드가 닫혀 있습니다 .

- " 요인 ", " 모형 ", " 별칭 항 " 및 " 설계 " 개요는 모형과 설계를 정의합니다 .
- " 설계 평가 " 개요는 설계의 특성을 설명하는 결과를 제공합니다 .

그림 15.11 가능한 모든 개요를 보여 주는 설계 평가 창



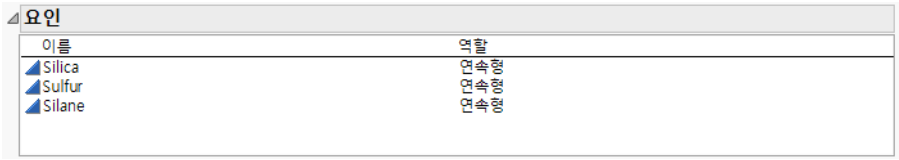
"요인","모형","별칭 항" 및 "설계" 개요에는 요인, 가정된 모형, 잠재적으로 별칭 구조를 가지는 관심 효과 및 실제 설계에 대해 입력하는 정보가 포함됩니다. JMP는 시작 창에서 사용자가 선택한 항목과 설계 테이블을 사용하여 이러한 개요를 채웁니다. 그러나 "모형" 및 "별칭 항" 개요에서 효과를 수정할 수 있습니다. 내용을 지정하면 "설계 평가" 개요가 업데이트됩니다. 이러한 개요를 열어 설계 정보를 제공하는 보고서 또는 컨트롤 창을 확인할 수 있습니다.

- "요인"
- "모형"
- "별칭 항"
- "설계"
- "설계 평가"
- "검정력 분석"
- "예측 분산 프로파일"
- "설계 공간 비율 그림"
- "예측 분산 표면"
- "추정 효율도"
- "별칭 행렬"
- "상관 색상 맵"
- "설계 진단"

요인

"요인" 개요에는 시작 창에 입력한 요인이 나열됩니다. "모형" 개요에서 효과를 구성할 요인을 선택할 수 있습니다.

그림 15.12 요인 개요



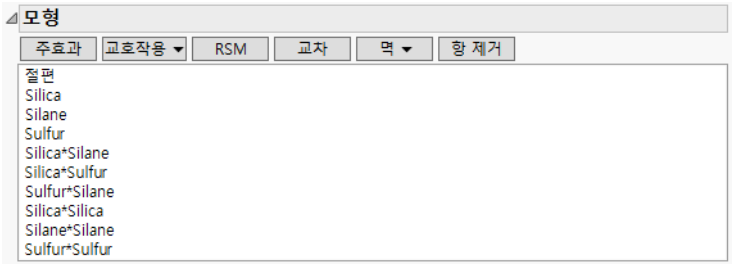
모형

"모형" 개요에 나타나는 항목은 관련 데이터 테이블에 "모형" 또는 "모형 적합" 스크립트가 포함되어 있는지 여부에 따라 다릅니다.

- 스크립트가 있으면 해당 스크립트에 지정된 추정 가능 효과가 "모형" 개요에 포함됩니다. 효과는 스크립트에 지정된 순서대로 모형에 입력됩니다. "모형" 스크립트에 있는 효과가 추정 가능 효과보다 많을 경우 추정할 수 없는 효과는 평가를 위해 모형에 포함되지 않습니다.
- 스크립트가 없으면 주효과만 "모형" 개요에 포함됩니다.

그림 15.13에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 데이터 테이블에 대한 "모형" 개요를 보여 줍니다. 데이터 테이블의 "모형" 스크립트에는 세 가지 요인(Silica, Silane, Sulfur)에 대한 반응 표면 효과가 포함되어 있습니다. 따라서 "모형" 개요에 이 세 요인에 대한 주효과, 이원 교호작용 및 2차 효과가 포함됩니다.

그림 15.13 Bounce Data.jmp 에 대한 모형 개요



다음 버튼을 사용하여 "모형" 개요에 효과를 추가할 수 있습니다.

주효과 모형의 모든 요인에 대한 주효과를 추가합니다.

교호작용 교호작용 효과를 추가합니다. "요인" 개요에서 요인을 선택하지 않은 경우 "2 차", "3 차", "4 차" 또는 "5 차" 를 선택하여 적절한 모든 교호작용을 해당 차수까지 추가합니다. "요인" 개요에서 요인 이름을 선택하고 "교호작용" 을 선택한 후 적절한 차수를 지정하여 특정 요인에 대해 지정된 차수까지 교호작용을 추가합니다. 비혼합 요인과 혼합 요인 간의 교호작용 및 블록 요인과 상수 요인 간의 교호작용은 추가되지 않습니다.

RSM 연속형 요인에 대해 교호작용 및 2 차 항을 최대 2 차까지 추가합니다 (반응 표면 모형 항). 범주형 요인은 RSM 항에 포함되지 않습니다. 모든 혼합 요인과 상호 작용하는 비혼합 요인의 주효과는 제거됩니다.

교차 특정 교호작용 항을 추가합니다. " 요인 " 개요에서 요인 이름을 선택하고 " 모형 " 개요에서 효과 이름을 선택합니다. " 교차 " 를 클릭하면 교차된 항이 " 모형 " 개요에 추가됩니다.

멱 다항식 항을 추가합니다. " 요인 " 개요에서 요인 이름을 선택하지 않으면 모든 연속형 요인에 대해 다항식 항을 추가합니다. " 요인 " 개요에서 요인 이름을 선택하면 해당 요인에 대해서만 다항식 항을 추가합니다. "2 차 ", "3 차 ", "4 차 " 또는 "5 차 " 를 선택하면 해당 차수의 다항식 항이 추가됩니다.

Scheffe 입방 모든 혼합 요인에 대해 Scheffe 입방 항을 추가합니다. 이 항은 3 차 다항식 항을 포함하는 혼합 모형을 지정하는 데 사용됩니다.

항 제거 선택한 효과를 제거합니다.

별칭 항

가정된 모형에 포함되지 **않은** 효과가 활성화되었을 수 있습니다. "별칭 항" 개요에서는 가정된 모형에 없지만 모형 항의 추정값을 편향시킬 수 있는 잠재적 활성 효과를 나열합니다. 별칭 행렬 항목은 "별칭 항" 개요에서 지정한 효과에 의해 모형 모수에 전달되는 편향 정도를 나타냅니다. 자세한 내용은 "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지)에서 확인하십시오.

기본적으로 "별칭 항" 개요에는 "모형" 개요에 없는 모든 이원 교호작용 효과가 포함됩니다(블록 요인을 포함하는 항은 제외). 버튼을 사용하여 항을 추가하거나 제거합니다. 이러한 버튼을 사용하여 "별칭 항" 테이블에 효과를 추가하는 방법에 대한 설명은 "모형"(417 페이지)에서 확인하십시오.

설계 평가 플랫폼에서 "별칭 행렬" 개요는 별칭 행렬 효과에 대한 변경 사항을 반영하기 위해 즉시 업데이트됩니다. 사용자 설계 플랫폼의 "별칭 항" 개요에서 효과를 수정한 후에는 "설계 생성" 을 클릭해야 합니다. 설계를 구성하는 다른 DOE 플랫폼에는 "별칭 항" 개요가 없습니다. 그러나 설계를 구성한 후 적절한 효과가 포함된 "별칭 행렬" 개요가 "설계 평가"에 나타납니다.

설계

"설계" 개요에서는 시작 창에서 지정한 요인에 대한 설계 런을 보여 줍니다. "설계 평가" 개요에서 특성을 탐색하면서 쉽게 설계를 볼 수 있습니다.

설계 평가

설계 평가 플랫폼 내의 "설계 평가"는 선택한 설계 및 "모형" 개요와 "별칭 항" 개요에서 지정한 내용을 기반으로 합니다. "설계 평가" 개요를 제공하는 몇 가지 DOE 설계 플랫폼은 사용자 설계, 확증적 선별 설계, 선별 설계, 반응 표면 설계, 최적 설계 유형을 사용한 혼합물 설계입니다. 이러한 플랫폼의 "설계 평가"는 구성한 설계를 기반으로 합니다.

검정력 분석

"검정력 분석" 개요에서는 모형의 모수에 대한 검정력을 계산합니다. 검정력은 주어진 크기의 활성 효과를 감지할 확률입니다. "검정력 분석" 개요는 설계에서 실제적으로 중요한 효과를 감지하는 능력을 평가하는 데 도움이 됩니다. 계수와 RMSE 가정이 정확하다고 간주할 때 검정력이 높을수록 유의 효과를 감지할 확률이 큽니다. 검정력은 런 수, 유의 수준 및 추정 오차 변동에 따라 달라집니다. 특히 추가 런이 필요한지 여부를 판별할 수 있습니다.

이 섹션에서는 다음 항목을 다룹니다 .

- " 검정력 분석 개요 "(419 페이지)
- " 검정력 분석 상세 정보 "(420 페이지)
- " 개별 모수 검정 "(421 페이지)
- " 수준이 세 개 이상인 범주형 효과 검정 "(422 페이지)
- " 설계 및 예상 반응 개요 "(422 페이지)
- " 커피 실험에 대한 검정력 분석 "(423 페이지)

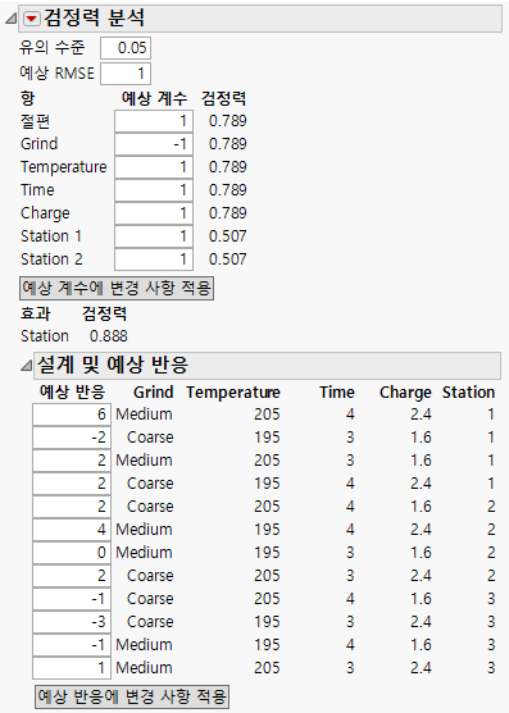
검정력 분석 개요

"모형" 개요에 나열된 효과에 대해 검정력이 계산됩니다. 여기에는 연속형, 이산 수치형, 범주형, 블록화, 혼합 및 공변량 요인이 포함됩니다. 검정은 개별 모형 모수 및 전체 효과에 대해 수행됩니다. 검정력을 계산하는 방법에 대한 자세한 내용은 "기술 상세 정보" 부록의 "검정력 계산"(768 페이지)에서 확인하십시오.

검정력은 모형 모수의 지정된 값에서 효과가 없다는 귀무가설을 기각할 확률입니다. 사실 모형 모수의 값이 아니라 실제적으로 중요한 평균 반응의 차이를 감지하는 데 관심이 있습니다. "검정력 분석" 개요에서는 예상 계수의 지정된 값에 대한 예상 반응을 계산할 수 있습니다. 이 기능은 평균 반응에서 감지할 차이와 관련된 계수 값을 결정하는 데 도움이 됩니다.

그림 15.14에서는 Design Experiment 폴더에 있는 Coffee Data.jmp 샘플 데이터 테이블의 설계에 대한 "검정력 분석" 개요를 보여 줍니다. "모형" 스크립트에 지정된 모형은 주효과만 포함된 모형입니다.

그림 15.14 Coffee Data.jmp 에 대한 검정력 분석



" 검정력 분석 " 개요에서 다음을 수행할 수 있습니다 .

- 감지할 차이를 반영하는 계수 값을 지정합니다 . 개요의 위쪽에 있는 " 예상 계수 " 에 이 값을 입력합니다 .
- 예상 반응 값을 지정하고 이를 적용하여 해당하는 예상 계수를 결정합니다 . " 설계 및 예상 반응 " 패널에서 예상 반응을 지정합니다 .

"검정력 분석"의 빨간색 삼각형 메뉴에서 다음을 수행할 수 있습니다.

- 검정력을 지정하여 해당 검정력을 달성할 예상 계수를 결정합니다 . 지정한 검정력을 충족할 수 있는 계수가 없는 경우 계수가 0 으로 설정됩니다 .
- 델타 값을 지정합니다 . 이 경우 예상 계수는 지정된 델타의 1/2 이 됩니다 . 그러면 예상 계수에 따라 검정력이 업데이트됩니다 .

검정력 분석 상세 정보

유의 수준 및 예상 RMSE 값을 지정합니다 . 이러한 값은 모형 모수에 대한 검정력을 계산하는 데 사용됩니다 .

유의 수준 효과가 없다는 가설이 참인데도 이를 기각할 확률입니다 . 값을 입력하면 검정력 계산이 즉시 업데이트됩니다 .

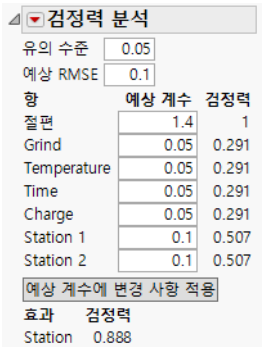
예상 RMSE 오차 변동의 제곱근에 대한 추정값입니다. 값을 입력하면 검정력 계산이 즉시 업데이트됩니다.

"검정력 분석" 보고서의 위쪽 부분은 "예상 계수"의 기본값과 함께 열립니다(그림 15.14 참조). 기본값은 델타를 기반으로 합니다. 자세한 내용은 "고급 옵션 > 검정력을 위한 델타 설정"(434 페이지)에서 확인하십시오.

참고 : 설계가 과포화되어 추정할 모수의 수가 런 수를 초과하면 예상 계수가 0 으로 설정됩니다.

그림 15.15에서는 예상 계수의 값이 지정된 "검정력 분석" 보고서의 위쪽 부분을 보여 줍니다. 이러한 값은 감지할 차이를 반영합니다.

그림 15.15 Coffee Data.jmp 에 대해 지정된 예상 계수



개별 모수 검정

" 항 " 열에는 모형 항이 나열됩니다. 각 항의 " 예상 계수 " 열에는 해당 항의 값이 포함됩니다. " 검정력 " 열의 값은 예상 계수에 의해 계수의 참 값이 주어진 경우 해당 항의 계수가 0 이라는 검정의 검정력입니다.

항 검정 대상 계수와 관련된 모형 항입니다.

참고 : " 검정력 분석 " 보고서에 나타나는 모형 항의 순서는 " 표준 최소 제곱 " 을 사용하여 얻은 " 모수 추정값 " 보고서의 모수 순서와 동일하지 않을 수 있습니다. 이 차이는 모형에 포함되는 교호작용의 자유도가 2 이상인 경우에만 발생할 수 있습니다.

예상 계수 모형 항과 관련된 계수의 값입니다. 이 값은 검정력 계산에 사용됩니다. 또한 " 설계 및 예상 반응 " 개요에서 " 예상 반응 " 열을 계산할 때도 사용됩니다. " 예상 계수 " 열에 새 값을 설정하는 경우 **예상 계수에 변경 사항 적용**을 클릭하여 " 검정력 " 및 " 예상 반응 " 열을 업데이트합니다.

참고 : 연속형 효과에 대한 예상 계수의 기본값은 1 입니다 . 범주형 효과의 경우에는 1 과 –1 값이 교대로 나타납니다 . 빨간색 삼각형 메뉴에서 **고급 옵션 > 검정력을 위한 델타 설정**을 선택하여 델타 값을 지정할 수 있습니다 . 델타 값을 변경하면 예상 계수의 절대값이 델타의 1/2 이 되도록 계수가 업데이트됩니다 . 자세한 내용은 " 고급 옵션 > 검정력을 위한 델타 설정 "(434 페이지) 에서 확인하십시오 .

검정력 지정된 예상 계수에 의해 계수의 참 값이 주어진 경우 효과가 없다는 귀무가설을 기각할 확률입니다 . 수치형 요인과 관련된 계수의 경우 평균 반응의 변화 (모형 기반) 는 계수 값의 두 배입니다 . 범주형 요인과 관련된 계수의 경우 전체 요인 수준에서 평균 반응의 변화 (모형 기반) 는 예상 계수 절대값의 두 배와 같습니다 .

지정된 유의 수준과 예상 RMSE 가 계산에 사용됩니다 . 검정력 계산에 대한 자세한 내용은 " 기술 상세 정보 " 부록의 " 단일 모수에 대한 검정력 "(768 페이지) 에서 확인하십시오 .

예상 계수에 변경 사항 적용 " 예상 계수 " 열에 새 값을 설정하는 경우 **예상 계수에 변경 사항 적용**을 클릭하여 " 검정력 " 및 " 예상 반응 " 열을 업데이트합니다 .

수준이 세 개 이상인 범주형 효과 검정

수준이 세 개 이상인 범주형 효과가 모형에 포함된 경우 " 예상 계수에 변경 사항 적용 " 버튼 아래에 다음 열이 나타납니다 .

효과 범주형 효과입니다 .

검정력 효과 없음 검정에 대한 검정력 계산입니다. 검정의 귀무가설은 효과에 해당하는 모든 모형 모수가 0 이라는 것입니다 . 감지할 차이는 " 예상 계수 " 열에서 효과의 모형 항에 해당하는 값으로 정의됩니다 . 검정력 계산은 예상 계수에 의해 결정된 반응 평균의 차이를 반영합니다 .

지정된 유의 수준과 예상 RMSE 가 계산에 사용됩니다 . 검정력 계산에 대한 자세한 내용은 " 기술 상세 정보 " 부록의 " 범주형 효과에 대한 검정력 "(769 페이지) 에서 확인하십시오 .

설계 및 예상 반응 개요

"설계 및 예상 반응" 개요에는 설계 앞에 "예상 반응" 열이 표시됩니다. 첫 번째 열의 각 항목은 설계 설정에 해당하는 예상 반응입니다. 예상 반응은 예상 계수를 사용하여 계산됩니다.

그림 15.16에서는 그림 15.15에 제공된 예상 계수 지정 값에 해당하는 "설계 및 예상 반응" 개요를 보여 줍니다.

그림 15.16 Coffee Data.jmp 에 대한 예상 반응

설계 및 예상 반응

예상 반응	Grind	Temperature	Time	Charge	Station
1.6	Medium	205	4	2.4	1
1.4	Coarse	195	3	1.6	1
1.4	Medium	205	3	1.6	1
1.6	Coarse	195	4	2.4	1
1.6	Coarse	205	4	1.6	2
1.5	Medium	195	4	2.4	2
1.3	Medium	195	3	1.6	2
1.6	Coarse	205	3	2.4	2
1.3	Coarse	205	4	1.6	3
1.2	Coarse	195	3	2.4	3
1.1	Medium	195	4	1.6	3
1.2	Medium	205	3	2.4	3

예상 반응에 변경 사항 적용

"예상 반응" 열에서 각 요인 설정의 값을 지정할 수 있습니다. 이러한 값은 감지할 차이를 반영합니다.

예상 반응에 변경 사항 적용을 클릭하면 " 예상 계수 " 열과 " 검정력 " 열이 모두 업데이트됩니다 .

예상 반응 예상 계수 값을 모형의 계수로 사용하여 구한 반응 값입니다 . 개요가 처음 나타날 때는 " 예상 계수 " 열의 기본값을 기반으로 예상 반응 값이 계산됩니다 . " 예상 반응 " 열에 새 값을 설정하는 경우 **예상 반응에 변경 사항 적용**을 클릭하여 " 예상 계수 " 및 " 검정력 " 열을 업데이트합니다 .

설계 " 예상 반응 " 열의 오른쪽 열에 설계의 모든 런에 대한 요인 설정이 표시됩니다 .

예상 반응에 변경 사항 적용 " 예상 반응 " 열에 새 값을 설정하는 경우 **예상 반응에 변경 사항 적용**을 클릭하여 " 예상 계수 " 및 " 검정력 " 열을 업데이트합니다 .

커피 실험에 대한 검정력 분석

Coffee Data.jmp 데이터 테이블의 설계를 예로 사용합니다. Strength 에 대한 다양한 크기의 효과를 감지할 설계 검정력에 관심이 있다고 가정해 보겠습니다. Grind 는 2 수준 범주형 요인, Temperature, Time 및 Charge 는 연속형 요인, Station 은 3 수준 범주형 (블록) 요인입니다.

이 예에서는 블록 요인 역할의 Station 을 무시합니다. 대신 Station 이 Strength 에 미치는 영향에 관심이 있습니다. Station 은 3 수준 범주형 요인이므로 " 모수 " 목록에 두 개의 항 (Station 1, Station 2) 으로 표시됩니다.

특히 평균 Strength 에서 다음과 같은 변화를 감지할 확률에 대해 알아보려고 합니다 .

- Grind 를 Coarse 에서 Medium 으로 바꿀 때 0.10 단위 변경
- Temperature, Time 및 Charge 를 낮은 수준에서 높은 수준으로 바꿀 때 0.10 단위 이상 변경
- 전체 예상 평균을 초과하는 0.10 단위의 각 Station 1 과 Station 2 로 인한 증가 . 이는 전체 예상 평균에서 0.20 단위의 Station 3 으로 인한 감소에 해당합니다 .

0.05 를 유의 수준으로 설정합니다. 고정 설계 설정에 대한 Strength 의 표준편차 추정값은 0.1 이며 이 값을 예상 RMSE 로 입력합니다.

그림 15.17에서는 이러한 값이 입력된 "검정력 분석" 노드를 보여 줍니다. 구체적으로 유의 수준, 예상 RMSE 및 각 예상 계수 값을 지정합니다.

"예상 계수에 변경 사항 적용"을 클릭하면 지정한 모형을 반영하여 "예상 반응" 값이 업데이트 됩니다.

그림 15.17 예상 계수 패널의 사용자 지정 내용이 포함된 검정력 분석 개요

검정력 분석

유의 수준 0.05

예상 RMSE 0.1

항	예상 계수	검정력
절편	1.4	1
Grind	0.05	0.291
Temperature	0.05	0.291
Time	0.05	0.291
Charge	0.05	0.291
Station 1	0.1	0.507
Station 2	0.1	0.507

예상 계수에 변경 사항 적용

효과	검정력
Station	0.888

설계 및 예상 반응

예상 반응	Grind	Temperature	Time	Charge	Station
1.6	Medium	205	4	2.4	1
1.4	Coarse	195	3	1.6	1
1.4	Medium	205	3	1.6	1
1.6	Coarse	195	4	2.4	1
1.6	Coarse	205	4	1.6	2
1.5	Medium	195	4	2.4	2
1.3	Medium	195	3	1.6	2
1.6	Coarse	205	3	2.4	2
1.3	Coarse	205	4	1.6	3
1.2	Coarse	195	3	2.4	3
1.1	Medium	195	4	1.6	3
1.2	Medium	205	3	2.4	3

예상 반응에 변경 사항 적용

Temperature는 수준이 -1과 1로 코드화된 연속형 요인입니다. Temperature가 Strength에 영향을 미치지 않는다는 귀무가설이 있는 검정을 가정해 보겠습니다. 그림 15.17에서는 Temperature의 전체 수준에서 0.10(=2*0.05) 단위의 차이를 감지할 검정력이 0.291에 불과하다는 것을 보여 줍니다.

이제 전체 Station 효과에 대한 검정을 고려하십시오. Station은 3수준 범주형 요인입니다. Station이 Strength에 영향을 미치지 않는다는 귀무가설이 있는 검정을 수행하려고 합니다. 이는 범주형 요인에 대한 일반적인 *F* 검정이며, **분석 > 모형 적합**을 실행할 때 "효과 검정" 보고서에 제공됩니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

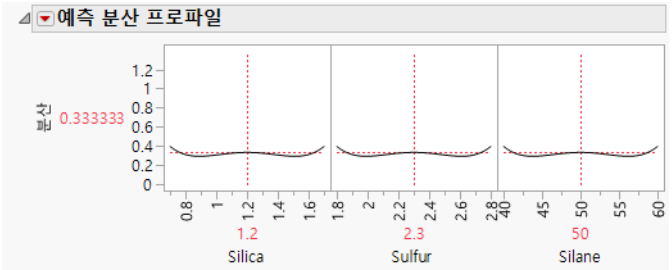
검정력은 "예상 계수에 변경 사항 적용" 버튼 바로 아래에 표시됩니다. Station 1과 Station 2 모형 항에 대한 "예상 계수"의 항목은 둘 다 0.10입니다. 이 설정은 두 스테이션이 전체 예상 평균보다 0.10 단위만큼 Strength를 증가시키는 효과가 있다는 의미입니다. Station 1 및 Station 2 계수의 이 설정에 대해 Station 3은 전체 예상 평균에서 0.20 단위만큼 Strength를 감소시키는 효과가 있습니다. 그림 15.17에서는 이 크기 이상의 차이를 감지할 검정력이 0.888임을 보여 줍니다.

예측 분산 프로파일

예측 분산 프로파일은 설계 공간에서 대략적인 예측 변동을 이해하는 데 도움이 됩니다. 예측 분산이 낮아야 좋습니다. 최대 분산을 찾으려면 **분산 최대화** 옵션을 사용합니다. 자세한 내용은 "분산 최대화"(425 페이지)에서 확인하십시오.

예측 분산 프로파일은 상대 예측 분산을 다른 요인의 값이 고정된 상태에서 각 요인의 함수로 그림에 나타냅니다. 그림 15.18에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 데이터 테이블의 예측 분산 프로파일을 보여 줍니다.

그림 15.18 예측 분산 프로파일러



상대 예측 분산

주어진 요인 설정에 대한 예측 분산은 설계 및 요인 설정에 따라 달라지는 양과 오차 분산의 곱입니다. 실험을 실행하기 전에는 오차 분산을 알 수 없으므로 예측 분산도 알 수 없습니다. 그러나 오차 분산에 대한 예측 분산의 비는 오차 분산의 함수가 아닙니다. **상대 예측 분산**이라고 하는 이 비율은 설계와 요인 설정에만 영향을 받습니다. 따라서 데이터를 얻기 전에 상대 예측 분산을 계산할 수 있습니다. 자세한 내용은 "기술 상세 정보" 부록의 "상대 예측 분산"(770 페이지)에서 확인하십시오.

실험을 실행하고 최소 제곱 모형을 적합시킨 후 모형 적합의 MSE(평균 제곱 오차)를 사용하여 오차 분산을 추정할 수 있습니다. 설정에 대한 상대 예측 분산을 곱해 설정에서 실제 예측 분산을 추정할 수 있습니다.

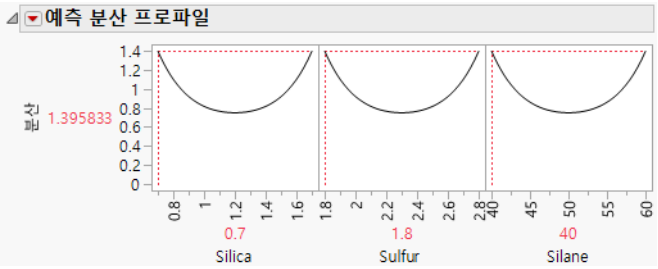
설계 공간 전체에서 예측 분산이 작은 것이 이상적입니다. 일반적으로 표본 크기가 증가하면 오차 분산이 감소합니다. 설계를 비교할 때 두 설계의 예측 분산 프로파일러를 나란히 놓을 수 있습니다. 평균적으로 예측 분산이 낮은 설계가 선호됩니다.

분산 최대화

최대 상대 예측 분산의 측면에서 설계를 평가하거나 비교할 수 있습니다. "예측 분산 프로파일"의 빨간색 삼각형 메뉴에서 **분산 최대화** 옵션을 선택합니다. JMP는 상대 예측 분산을 최대화하는 만족도 함수를 사용합니다. "예측 분산 프로파일"에서 최대 분산 값은 상대 예측 분산의 최악값(설계 관점에서 가장 적합하지 않음)입니다. 최대 분산은 둘 이상의 요인 설정 조합에서 발생할 수 있습니다.

그림 15.19 에서는 "분산 최대화"가 선택된 "예측 분산 프로파일"을 보여 줍니다. Design Experiment 폴더에 있는 Bounce Data.jmp 샘플 데이터 테이블을 대상으로 한 그림이 표시됩니다. 상대 예측 분산의 최대값은 Silica = 0.7, Sulfur = 1.8, Silane = 40 일 때 1.3958 입니다. 그러나 여러 요인 설정에서 이와 동일한 상대 분산을 가질 수 있다는 점에 유의하십시오. Silica = 1.7, Sulfur = 2.8, Silane = 60 인 설계점의 상대 예측 분산도 1.3958 입니다. 표면을 사용하여 예측 분산을 평가하려면 "예측 분산 표면"(427 페이지)에 나오는 내용을 참조하십시오.

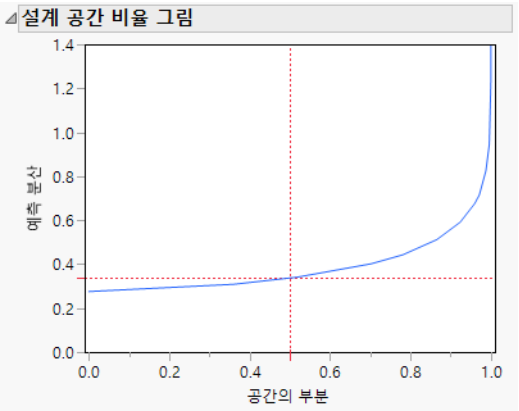
그림 15.19 최대 분산을 보여 주는 예측 분산 프로파일



설계 공간 비율 그림

"설계 공간 비율 그림"은 설계 공간에서 상대 예측 분산이 특정 값 미만이 되는 비율을 보여 줍니다. 예측 분산 값이 낮은 설계 공간의 비율이 클수록 바람직합니다. 그림 15.20 에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 샘플 데이터 테이블에 대한 "설계 공간 비율 그림"을 보여 줍니다.

그림 15.20 설계 공간 비율 그림



그림의 *X* 축은 설계 공간의 비율을 나타냅니다(0 ~ 100% 범위). *Y* 축은 상대 예측 분산 값을 나타냅니다. 파란색 곡선에 있는 *(x, y)* 점의 경우 *x* 값은 분산이 *y* 보다 작거나 같은 설계 공간의 비율입니다. 빨간색 점선 십자기호는 설계 공간의 50%에 대해 상대 예측 분산의 경계를 이루는 값을 표시합니다.

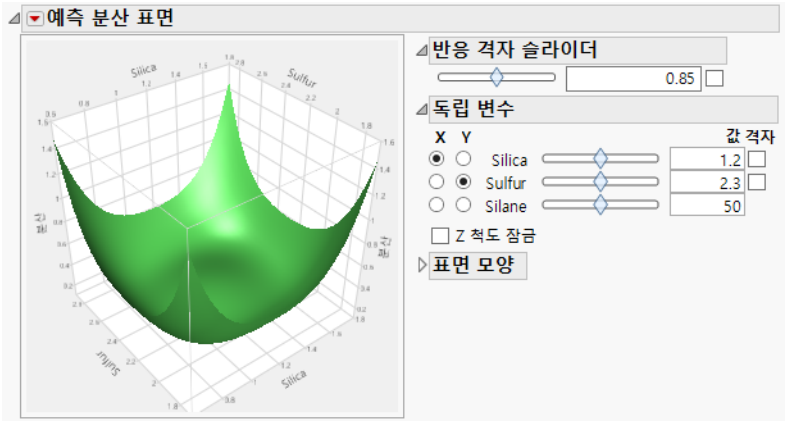
그림 15.20에서는 상대 예측 분산의 최소값이 0.3보다 약간 작고 최대값이 1.4 미만임을 보여 줍니다. 실제 최대값은 그림 15.19에 표시된 대로 1.395833입니다. 빨간색 점선 십자기호는 설계 공간의 약 50%에 대해 상대 예측 분산이 0.34 미만임을 나타냅니다. 십자기호 도구를 사용하여 "공간 비율" 값에 해당하는 최대 상대 예측 분산을 찾을 수 있습니다. 그림 15.20의 십자기호 도구를 사용하여 예측 분산 값의 90%가 약 0.55 미만인지 확인하십시오.

참고 : 설계 공간 비율 그림을 구성할 때 설계 공간의 몬테카를로 표집이 사용됩니다. 따라서 같은 설계에 대한 그림이 약간 다를 수 있습니다.

예측 분산 표면

"예측 분산 표면" 보고서는 상대 예측 분산 표면을 두 설계 요인의 함수로 그림에 나타냅니다. 그림 15.21에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 샘플 데이터 테이블에 대한 "예측 분산 표면" 개요를 보여 줍니다. 빨간색 삼각형 메뉴에서 **제어판**을 선택하여 컨트롤을 표시하거나 숨길 수 있습니다. 자세한 내용은 "제어판"(427 페이지)에서 확인하십시오.

그림 15.21 예측 분산 표면



요인이 두 개 이상이면 "예측 분산 표면" 개요에 두 변수에 대한 상대 예측 분산 그림이 표시됩니다. "예측 분산 표면" 개요에는 상대 예측 분산 계산식이 그림에 표시됩니다. 그림을 드래그하여 투시도를 회전하고 변경할 수 있습니다.

제어판

제어판에는 그림을 조정할 수 있는 다음과 같은 옵션이 들어 있습니다.

반응 격자 슬라이더 **격자** 체크박스는 분산의 상수 값을 표시하는 격자를 겹쳐 놓습니다. 분산 값은 텍스트 상자에 표시됩니다. 슬라이더를 사용하여 격자 배치를 조정할 수 있습니다. 또는 텍스트 상자에 분산 값을 입력할 수도 있습니다. 상자 바깥쪽을 클릭하면 그림이 업데이트됩니다.

독립 변수 이 패널을 사용하면 그림의 축으로 사용할 두 요인을 선택하고 축으로 사용되지 않는 요인의 설정을 지정할 수 있습니다. 해당 열을 클릭하여 X 축과 Y 축에 대한 요인을 각각 선택합니다. 슬라이더와 텍스트 상자를 사용하여 축에 대해 선택되지 **않은** 각 요인의 값을 지정할 수 있습니다. 그림에는 축으로 사용되지 않는 요인의 지정된 값에서 표면의 3 차원 조각이 표시됩니다. 슬라이더를 움직여 다른 조각을 볼 수 있습니다.

각 격자 체크박스는 해당 요인에 대한 격자를 활성화합니다. 슬라이더를 사용하여 각 격자의 배치를 조정할 수 있습니다.

"Z 척도 잠금 " 은 z 축을 현재 값으로 잠급니다. 이 기능은 축에 해당되지 않는 슬라이더를 움직일 때 유용합니다.

표면 모양 **해상도** 슬라이더는 계산식에 대해 평가되는 점 수에 영향을 줍니다. 해상도가 정밀하지 않고 너무 거칠면 급격한 변화가 있는 함수가 잘 표현되지 않을 수 있습니다. 그러나 해상도를 높게 설정하면 표면 평가 및 표시 속도가 느려질 수 있습니다.

직교 투영 체크박스는 2 차원으로 투영된 그림을 표시합니다.

등고선 메뉴는 등고선 곡선의 배치를 제어합니다. 등고선 곡선은 반응 값이 일정한 점의 집합입니다. 등고선을 " 해제 "(기본값) 하여 표시하지 않거나 " 아래쪽 ", " 위쪽 " 또는 " 표면상 " 에 배치하도록 선택할 수 있습니다.

추정 효율도

이 보고서는 모형의 각 모수 추정값에 대한 CI(신뢰 구간) 길이의 증가 비율 및 추정값의 상대 표준 오차를 제공합니다. 두 값 모두 작을수록 좋습니다. 그림 15.22 에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 샘플 데이터 테이블의 "추정 효율도" 개요를 보여 줍니다.

그림 15.22 추정 효율도 개요

추정 효율도

항	CI 길이의 부분적 증가	추정값의 상대 표준 오차
절편	1.236	0.577
Silica	0.369	0.354
Silane	0.369	0.354
Sulfur	0.369	0.354
Silica*Silane	0.936	0.5
Silica*Sulfur	0.936	0.5
Sulfur*Silane	0.936	0.5
Silica*Silica	1.016	0.52
Silane*Silane	1.016	0.52
Sulfur*Sulfur	1.016	0.52

CI 길이의 증가 비율

"CI 길이의 증가 비율 " 은 현재 설계에서 제공한 모수 신뢰 구간의 길이를 이상적 설계에서 제공한 신뢰 구간의 길이와 비교합니다.

- 모수에 대한 이상적인 신뢰 구간의 길이를 실제 신뢰 구간의 길이에서 뺍니다.
- 이 차이를 이상적 신뢰 구간의 길이로 나눕니다.

직교 D-최적 설계의 경우 증가 비율이 0입니다. 설계를 선택할 때 신뢰 구간 길이의 증가 비율을 가능한 한 작게 하려고 합니다.

이상적 설계

보통 최소 제곱 추정량의 공분산 행렬은 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$입니다. $(\mathbf{X}'\mathbf{X})^{-1}$의 주대각 요소는 모수 추정값의 **상대 분산**(분산을 σ^2으로 나눈 값)입니다. 2수준 설계에서 효과 코드화 규칙("열 특성" 부록의 "코딩"(739 페이지) 참조)을 사용할 경우 모수 추정값에 대한 상대 분산의 최소값은 $1/n$입니다. 여기서 n은 런 수입니다. 이는 설계의 모든 효과가 직교이고 설계가 D-최적일 때 해당됩니다.

$\hat{\beta}$가 모수 추정값의 벡터라고 하면 **이상적 설계**(존재하지 않을 수 있음)는 공분산 행렬이 다음과 같이 지정된 설계입니다.

$$\text{Var}(\hat{\beta}) = (\sigma^2/n)\mathbf{I}_n$$

여기서 $\mathbf{I}_n$은 n x n 단위 행렬이고 σ는 반응의 표준편차입니다.

직교 D-최적 설계가 존재하면 해당 설계가 이상적입니다. 그러나 위의 정의는 이상적 설계의 개념을 직교와 D-최적을 모두 충족하는 설계가 존재하지 않는 상황으로 확장합니다.

이 정의는 여러 수준의 범주형 요인이 있는 설계에도 적합합니다. 범주형 요인에 사용되는 직교 코드화를 통해 이러한 설계가 이상적 공분산 행렬을 가질 수 있습니다. 사용자 설계의 경우 "사용자 설계" 창의 옵션에서 "행렬 X 저장"을 선택하고, 설계 테이블을 생성한 다음, 설계 테이블에 저장된 "모형 행렬" 스크립트를 확인하면 코딩 행렬을 볼 수 있습니다.

신뢰 구간 길이의 증가 비율

이상적 설계에서 모수 추정값의 표준 오차는 다음과 같이 지정됩니다.

$$SE_{Ideal}(\hat{\beta}) = (\sigma/\sqrt{n})\mathbf{I}_n$$

신뢰 구간 길이는 표준 오차에 의해 결정됩니다. 신뢰 구간 길이의 증가 비율은 주어진 설계의 표준 오차와 이상적 설계의 표준 오차 간 차이를 이상적 설계의 표준 오차로 나눈 값입니다.

i^{th} 모수 추정값의 경우 신뢰 구간 길이의 증가 비율은 다음과 같이 정의됩니다.

$$FI = \frac{\sigma\sqrt{(\mathbf{X}'\mathbf{X}_{ii})^{-1}} - (\sigma/(\sqrt{n}))}{(\sigma/\sqrt{n})} = \sqrt{n(\mathbf{X}'\mathbf{X}_{ii})^{-1}} - 1$$

다음은 각 요소에 대한 설명입니다.

σ^2은 알려지지 않은 반응 분산입니다.

$\mathbf{X}$는 주어진 설계의 모형 행렬이며, "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지)에 정의되어 있습니다.

$(\mathbf{X}'\mathbf{X})^{-1}_{ii}$은 $(\mathbf{X}'\mathbf{X})^{-1}$의 i^{th} 주대각 항목입니다.

n은 런 수입니다.

추정값의 상대 표준 오차

추정값의 상대 표준 오차는 오차 표준편차에 대한 모수 추정값의 표준편차 비율을 제공합니다. 이 값은 오차 표준편차를 기준으로 모형 모수 추정값의 표준 오차 크기를 나타냅니다. i^{th} 모수 추정값의 경우 추정값의 상대 표준 오차는 다음과 같이 정의됩니다.

$$SE = \sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{ii}}$$

다음은 각 요소에 대한 설명입니다.

X 는 "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지) 에 정의된 모형 행렬입니다.

$(\mathbf{X}'\mathbf{X})^{-1}_{ii}$ 은 $(\mathbf{X}'\mathbf{X})^{-1}$ 의 i^{th} 주대각 항목입니다.

별칭 행렬

별칭 행렬은 실제로 활성이지만 모형에 포함되지 않은 항이 모형 항 추정에 영향을 미치는 방식에 대한 문제를 처리합니다. "별칭 항" 개요에서는 가정된 모형에 없지만 모형 항의 추정값을 편향시킬 수 있는 잠재적 활성 효과를 나열합니다. 별칭 행렬 항목은 "별칭 항" 효과에 의해 모형 모수에 전달되는 편향 정도를 나타냅니다. 자세한 내용은 "별칭 항"(418 페이지) 에서 확인하십시오.

별칭 행렬의 행은 "모형" 개요에 나열된 모형 효과에 해당하는 항입니다. 열은 "별칭 항" 개요에 나열된 효과에 해당하는 항입니다. 주어진 행과 열의 항목은 별칭 항이 모형 항에 해당하는 모수 추정값에 영향을 미치는 정도를 나타냅니다.

설계를 평가할 때 별칭 행렬의 항목과 관련하여 다음 두 가지 상황 중 하나가 발생하면 가장 이상적입니다. 항목이 작거나, 그렇지 않은 경우 별칭 항의 효과가 작아서 편향도 작아집니다. 별칭 항의 효과가 상당히 클 것으로 생각되면 해당 항을 모형에 포함하거나 별칭 최적 설계를 고려해야 합니다.

별칭 행렬 계산에 대한 자세한 내용은 "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지) 에서 확인하십시오. Lekivetz, R.(2014) 도 참조하십시오.

다음 사항에 유의하십시오.

- 가정된 모형에 대한 직교 설계의 경우 별칭 행렬의 상관은 상관 색상 맵의 절대 상관에 해당합니다.
- 설계 복잡성에 따라 별칭 행렬 항목이 1 보다 크거나 -1 보다 작을 수 있습니다.

별칭 행렬의 예

Design Experiment 폴더의 Cake Data.jmp 샘플 데이터 테이블을 예로 들어보겠습니다. 이 설계는 주효과 모형을 가정합니다. 데이터 테이블에서 "모형" 스크립트를 실행하여 이를 확인할 수 있습니다. 결과적으로 "설계 평가" 창의 "모형" 개요에 절편과 5 개 주효과만 나타납니다. "별칭 항" 개요에는 이원 교호작용이 포함되어 있습니다.

그림 15.23 Coffee Data.jmp 에 대한 별칭 행렬

별칭 행렬

효과	Grind*Temperature	Grind*Time	Grind*Charge	Temperature*Time	Temperature*Charge	Time*Charge
절편	0	0	0	0	0	0
Grind	0	0	0	0.333	-0.33	-0.33
Temperature	0	0.333	-0.33	0	0	-0.33
Time	0.333	0	-0.33	0	-0.33	0
Charge	-0.33	-0.33	0	-0.33	0	0
Station 1	-0.41	0	0	0	0	0.816
Station 2	0.707	0	0	0	0	0

별칭 행렬의 첫 번째 열에는 모형 항이 표시되어 행을 정의합니다. 별칭 항의 이원 교호작용이 위쪽에 나열되어 열을 정의합니다. 모형 효과 Temperature를 예로 들어보겠습니다. Grind*Time 교호작용만 활성 이원 교호작용인 경우, Temperature의 계수 추정값은 Grind*Time 효과의 참 값보다 0.333배만큼 편향됩니다. 다른 교호작용이 활성인 경우 "별칭 행렬"의 값은 Temperature 계수 추정값에 의해 발생하는 추가 편향의 크기를 나타냅니다.

상관 색상 맵

"상관 색상 맵"에는 "모형" 개요 또는 "별칭 항" 개요에 나타나는 두 효과 간의 절대 상관 값이 표시됩니다. 색상 맵의 셀은 맵 위에서 식별됩니다. "모형" 개요의 각 효과에 대한 셀과 "별칭 항" 개요의 각 효과에 대한 셀이 있습니다. 값이 작을수록 좋습니다.

기본적으로 절대 상관 크기는 흰색에서 회색, 검은색까지의 강도 색상 테마를 사용하여 나타냅니다. 일반적으로 좋은 설계의 색상 맵에는 흰색 비대각이 많이 표시되어 직교성 또는 개별 항 간의 상관이 적음을 나타냅니다. 효과 간의 절대 상관이 크면 추정값 표준 오차가 증가합니다.

두 효과 간의 절대 상관을 보려면 해당 셀을 커서로 가리킵니다. 색상 척도를 변경하려면 "상관 색상 맵"의 빨간색 삼각형을 클릭하고 **파랑-회색-빨강**을 선택합니다. 사용자 정의 색상 척도를 사용하려면 그림을 마우스 오른쪽 버튼으로 클릭하고 **색상 테마**를 선택합니다. 상관계수표를 저장하려면 범례 아래의 그림 오른쪽을 마우스 오른쪽 버튼으로 클릭하고 **상관계수표**를 선택합니다.

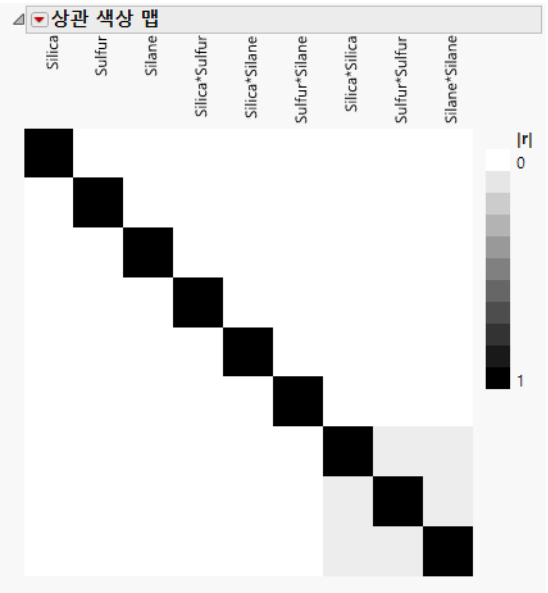
색상 맵의 예

그림 15.24에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 샘플 데이터 테이블에 대한 "상관 색상 맵"을 보여 줍니다. 검은색은 절대 상관 1을 나타냅니다. 주대각의 검은색 셀은 모형 항 자체의 상관관계를 보여 줍니다.

다른 모든 셀은 흰색 또는 회색입니다. 회색 사각형은 2차 항 간의 상관에 해당합니다. 이 값을 보려면 각 회색 사각형을 커서로 가리킵니다. 다른 2차 항과의 절대 상관은 0.0714로 작습니다.

상관 측면에서 볼 때 이 설계는 양호합니다. 효과 상관관계가 높으면 반응에 영향을 주는 효과를 판별하기가 어렵습니다.

그림 15.24 상관 색상 맵



팁 : 상관계수표를 저장하려면 범례 아래의 그림 오른쪽을 마우스 오른쪽 버튼으로 클릭하고 **상관계수표**를 선택합니다 .

설계 진단

"설계 진단" 개요에서는 *D* 효율도, *G* 효율도, *A* 효율도 및 평균 예측 분산을 보여 줍니다. "변경"이 "어려움"이나 "매우 어려움"으로 설정된 요인 또는 "추정 가능성"이 "If Possible"로 지정된 효과를 포함하는 설계의 경우 이러한 진단이 표시되지 않습니다.

"설계 평가"가 아닌 다른 DOE 플랫폼에서 "설계 진단"에 액세스한 경우 설계를 생성하는 데 필요한 시간이 "설계 생성 시간"에 표시됩니다. 설계 평가 플랫폼에서 "설계 진단"에 액세스한 경우에는 설계 평가 플랫폼에서 결과를 계산하는 데 필요한 시간이 "설계 생성 시간"에 표시됩니다.

그림 15.25에서는 Design Experiment 폴더에 있는 Bounce Data.jmp 샘플 데이터 테이블에 대한 "설계 진단" 개요를 보여 줍니다.

그림 15.25 설계 진단 개요

주의 : 효율도 측도를 독자적으로 해석하면 안 됩니다 . 그러나 설계를 비교할 때 사용할 수 있습니다 . 두 가지 설계가 주어진 경우 효율도 측도가 높은 설계가 더 좋습니다 . 모든 기준에 대한 최대 효율도는 100 이지만 많은 설계 문제로 인해 100% 효율도는 불가능합니다 .

표기

아래에 나오는 효율도 측도에 대한 설명에 다음 표기가 사용됩니다 .

- **X** 는 모형 행렬입니다 .
- n 은 설계의 런 수입니다 .
- p 는 모형의 항 수입니다 (절편 포함).
- $Var(\hat{y}|\underset{\sim}{x})$ 는 점 $\underset{\sim}{x}$ 에서의 상대 예측 분산입니다 . 자세한 내용은 " 기술 상세 정보 " 부록의 " 상대 예측 분산 "(770 페이지) 에서 확인하십시오 .
- $Var(\hat{y}|\underset{\sim}{x})_{max}$ 는 설계 영역에 대한 최대 상대 예측 분산입니다 .

D 효율도

D- 최적 기준 관점에서 이상적 직교 설계의 효율도에 대한 설계 효율도입니다 . D- 최적 설계는 회귀 계수 벡터에 대한 결합 신뢰 영역의 크기를 최소화합니다 .

$$D\text{효율도} = 100\left(\frac{1}{n}|\mathbf{X}'\mathbf{X}|^{1/p}\right)$$

G 효율도

G- 최적 기준 관점에서 이상적 직교 설계의 효율도에 대한 설계 효율도입니다 . G- 최적 설계는 설계 영역에 대한 최대 예측 분산을 최소화합니다 .

$$G\text{효율도} = 100p/(nVar(\hat{y}|\underset{\sim}{x})_{max})$$

D 가 설계 영역이라고 하면

$$Var(\hat{y}|\underset{\sim}{x})_{max} = \underset{\underset{\sim}{x} \text{ in D}}{maximum}[\underset{\sim}{x}'(\mathbf{X}'\mathbf{X})^{-1}\underset{\sim}{x}]$$

참고 : G 효율도는 설계 공간의 몬테카를로 표집을 사용하여 계산됩니다 . 따라서 같은 설계에 대한 계산이 약간 다를 수 있습니다 .

A 효율도

A- 최적 기준 관점에서 이상적 직교 설계의 효율도에 대한 설계 효율도입니다 . A- 최적 설계는 회귀 계수 분산의 합을 최소화합니다 .

$$A\text{-efficiency} = 100p/(n\text{Trace}(\mathbf{X}'\mathbf{X})^{-1})$$

예측 평균 분산

설계 공간의 점 $\underset{\sim}{\mathbf{x}}$ 에서 상대 예측 분산은 다음과 같이 정의됩니다 .

$$Var(\hat{y}|\underset{\sim}{\mathbf{x}}) = \underset{\sim}{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\underset{\sim}{\mathbf{x}}$$

이것은 예측 분산을 오차 분산으로 나눈 값입니다. 계산에 대한 자세한 내용은 Goos and Jones(2011) 의 "Section 4.3.5" 에서 확인하십시오.

참고 : 설계 영역이 선형 제약 조건 또는 허용되지 않는 조합에 의해 제약을 받는 경우 평균 예측 분산 값은 설계 공간 비율 그림의 예측 분산 값을 사용하여 계산됩니다 .

설계 생성 시간 (초)

설계 평가 플랫폼에서 결과를 계산하는 데 필요한 시간이 " 설계 생성 시간 " 에 표시됩니다.

설계 평가 옵션

" 설계 평가 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

고급 옵션 > 분할구 분산비 오차 분산에 대한 랜덤 주구 분산 및 하위구 분산 (있는 경우) 의 비를 지정합니다 . 이 값을 설정하려면 먼저 분할구 설계를 위한 변경하기 힘든 요인 또는 분할 - 분할구 설계를 위한 변경하기 힘든 요인과 매우 변경하기 힘든 요인을 정의해야 합니다 . 그런 다음 분할구 설계를 지정했는지 아니면 분할 - 분할구 설계를 지정했는지에 따라 분산 비율에 하나 또는 두 개의 양수를 입력할 수 있습니다 .

고급 옵션 > 검정력을 위한 델타 설정 감지할 차이에 대한 값을 지정합니다 . 이 값은 " 검정력 분석 " 보고서의 " 예상 계수 " 에 적용됩니다 . 연속형 효과에 대한 " 예상 계수 " 값은 델타 /2 로 설정됩니다 . 범주형 효과의 경우 델타 /2 와 – 델타 /2 값이 교대로 나타납니다 . 검정력 분석에 대한 자세한 내용은 " 검정력 분석 "(419 페이지) 에서 확인하십시오 .

기본적으로 델타는 2 로 설정됩니다 . 따라서 " 예상 계수 " 기본값은 연속형 효과의 경우 1 이고 범주형 효과의 경우 1 과 –1 이 교대로 나타납니다 . 델타가 2 일 때 예상 계수로 입력되는 기본값의 특성은 다음과 같습니다 .

– 수치형 효과에 대한 검정력 계산에서는 요인이 설계 영역의 가장 낮은 설정에서 가장 높은 설정으로 변경될 때 선형 주효과로 인한 반응 평균의 델타 변화를 가정합니다 .

- 2 수준 범주형 요인과 관련된 모수에 대한 검정력 계산에서는 전체 요인 수준에서 반응 평균의 델타 변화를 가정합니다 .
- 수준이 3 개 이상인 범주형 효과에 대한 검정력 계산은 모든 수준의 반응 평균이 동일하다는 귀무가설에 대한 다중 자유도 *F* 검정을 기반으로 합니다 . 예상 계수에 의해 결정된 반응 평균의 값으로 검정력이 계산됩니다 . 예상 계수의 다양한 구성으로 수준의 차이를 델타 값만큼 정의할 수 있습니다 . 그러나 이러한 구성의 검정력 값은 다른 수준의 예상 계수에 따라 달라집니다 .

스크립트를 스크립트 창에 저장 " 설계 평가 " 창을 재현하는 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

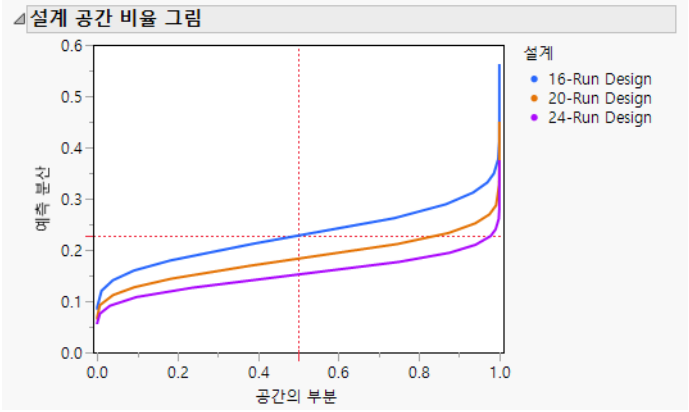
16 장

설계 비교

설계를 동시에 비교 및 평가

설계 비교 플랫폼은 최대 4개의 설계를 참조 설계와 비교합니다. 이 플랫폼을 사용하여 설계 성능을 탐색, 평가 및 비교할 수 있습니다. 진단 결과는 설계 작동 상태를 독립적 방식 및 서로 상대적인 방식으로 나누어 보여 줍니다. 특정 요구 사항에 관하여 설계를 비교하기 위해 가정된 모형 및 별칭 항 목록에서 항을 변경할 수 있습니다.

그림 16.1 런 크기가 다른 세 가지 설계 비교



목차

설계 비교 플랫폼 개요 439
설계 비교의 예 439
런 크기가 같은 설계 439
런 크기가 다른 설계 444
주구 수가 다른 분할구 설계 447
설계 비교 플랫폼 시작 451
설계 비교 창 : 모형 및 별칭 항 지정 453
참조 설계 454
요인 454
모형 454
별칭 항 454
설계 비교 창 : 설계 평가 454
검정력 분석 455
예측 분산 프로파일 458
설계 공간 비율 그림 460
상대 추정 효율도 461
별칭 행렬 요약 463
절대 상관 464
설계 진단 466
설계 비교 옵션 468

설계 비교 플랫폼 개요

설계 평가 플랫폼의 확장인 설계 비교 플랫폼을 사용하면 최대 4 개의 설계를 참조 설계와 쉽게 비교할 수 있습니다. 하나 또는 두 설계의 성능을 다른 설계와 비교하려면 기준 설계로 간주되는 참조 설계를 선택합니다. "모형" 개요에서 효과를 지정하고 "별칭 항" 개요에서 관심 효과를 지정할 수 있습니다.

" 설계 평가 " 보고서는 다음 항목에 대한 진단 결과 및 그림을 보여 줍니다 .

- 검정력 분석
- 예측 분산
- 설계 공간 비율
- 상대 추정 효율도
- 별칭 행렬 진단
- 효과 간 상관 (교락 포함)
- 전체 설계에 대한 상대 효율도 측도

설계 비교의 예

- " 런 크기가 같은 설계 "
- " 런 크기가 다른 설계 "
- " 주구 수가 다른 분할구 설계 "

런 크기가 같은 설계

이 예에서는 각각 13 회 런을 수행하는 6 개 요인에 대한 두 설계를 비교합니다. 하나는 단일 중앙점을 사용하여 확대된 12 회 런 PB(Plackett-Burman) 설계이고 다른 하나는 DSD(확증적 선별 설계) 입니다.

주효과만 비교

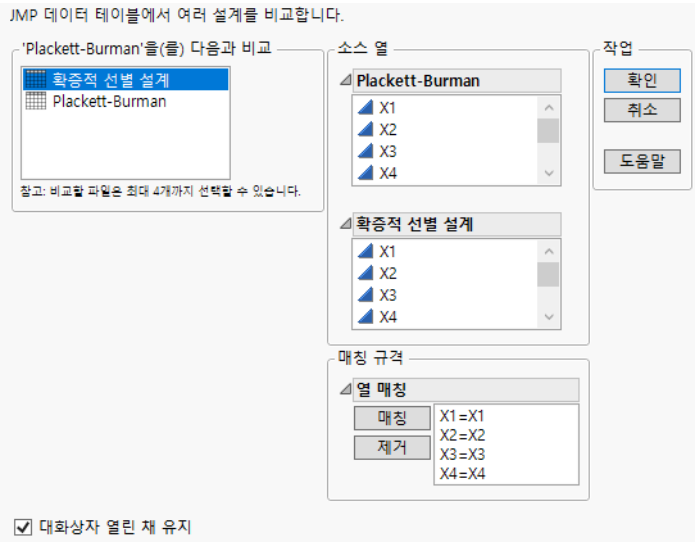
먼저 추정할 모형에 주효과만 포함된 경우를 가정하고 두 설계를 비교합니다 .

1. **도움말 > 샘플 데이터**를 선택하고 **샘플 스크립트 디렉터리 열기**를 클릭한 후 Compare Same Run Size.jsl 을 선택합니다 .
2. 스크립트 창을 마우스 오른쪽 버튼으로 클릭하고 **스크립트 실행**을 선택합니다 .

   " 확증적 선별 설계 " 와 "Plackett-Burman" 이라는 13 회 런 설계 테이블 두 개가 구성됩니다 . 이 두 설계를 비교하려고 합니다 . Plackett-Burman 테이블이 활성화되어 있으므로 이 설계를 참조 설계로 사용하여 DSD 를 비교합니다 .

3. "Plackett-Burman" 데이터 테이블에서 **DOE > 설계 진단 > 설계 비교**를 선택합니다 .
4. **'Plackett-Burman' 을 (를) 다음과 비교** 목록에서 확증적 선별 설계를 선택합니다 .
5. "Plackett-Burman" 패널과 " 확증적 선별 설계 " 패널에서 X1 ~ X6 을 선택합니다 .
6. " 열 매칭 " 개요를 열고 **매칭**을 클릭합니다 .

그림 16.2 매칭된 열이 있는 시작 창

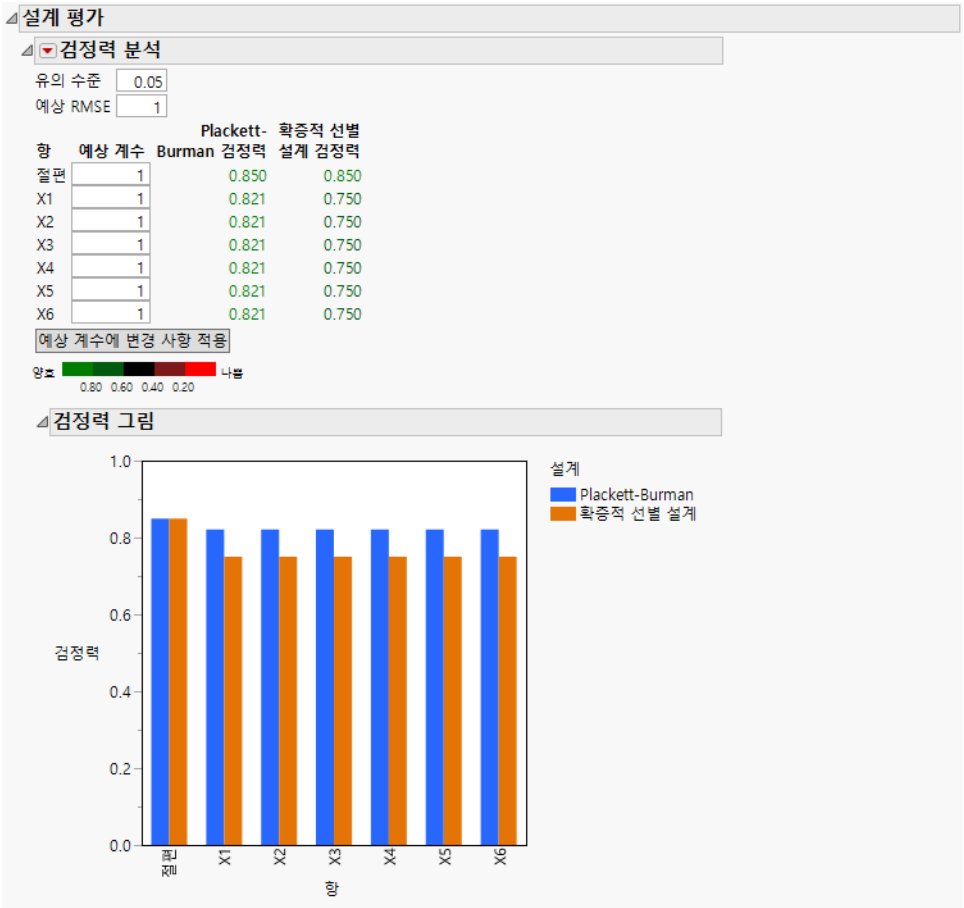


이렇게 하면 두 설계의 요인 간에 대응 관계가 정의됩니다 .

7. **확인**을 클릭합니다 .

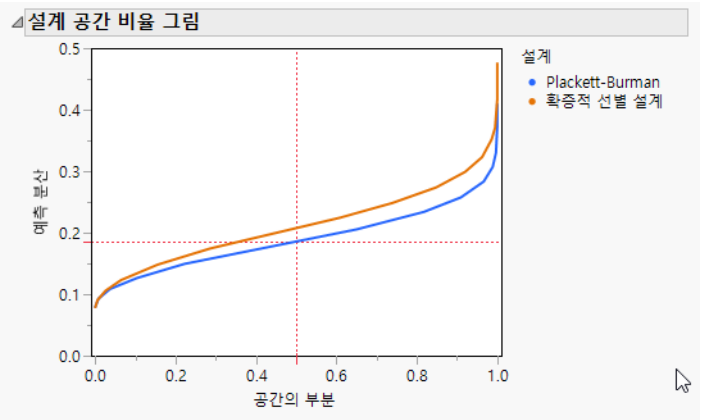
참조 설계는 Plackett-Burman 설계입니다 . " 설계 평가 " 개요에서 비교 지표로 PB 와 DSD 를 비교합니다 . 검정력 , 예측 분산 , 추정 효율도 , 별칭 구조 및 설계 효율도 측도를 기준으로 설계가 비교됩니다 .

그림 16.3 PB 와 DSD 비교를 위한 검정력 분석



검정력 , 예측 분산 및 추정 효율도 측면에서 PB 설계가 DSD 보다 우수합니다 . 그림 16.3 에서는 유의 수준 , 예상 RMSE 및 계수가 기본 설정으로 지정된 " 검정력 분석 " 보고서를 보여줍니다 . 주효과 검정의 경우 PB 설계가 DSD 보다 검정력이 높습니다 .

그림 16.4 PB 와 DSD 비교를 위한 설계 공간 비율 그림



" 설계 공간 비율 그림 " 은 전체 설계 공간에서 PB 설계가 DSD 보다 예측 분산이 작음을 나타냅니다 .

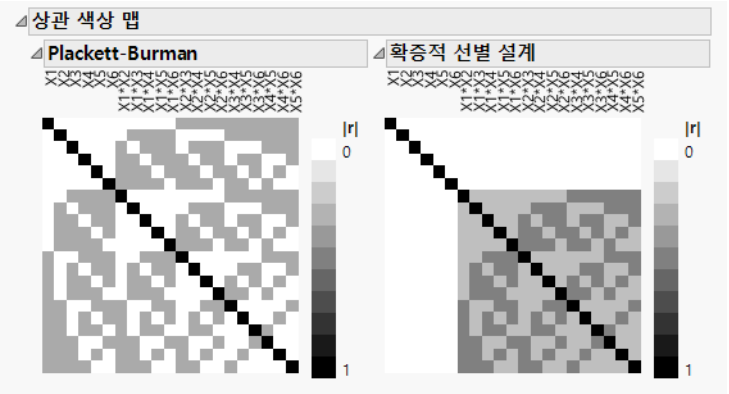
주효과만 활성인 것으로 의심되는 경우 PB 설계가 적합하다고 결론을 내릴 수 있습니다.

이원 교호작용 비교

이제 일부 이원 교호작용이 활성 효과일 수 있다고 의심하는 경우를 가정해 보겠습니다 . 아래 분석에서는 해당 이원 교호작용이 실제로 활성일 경우 PB 가 DSD 보다 적합하지 않다는 것을 보여줍니다 .

1. " 절대 상관 " 보고서에서 " 상관 색상 맵 " 보고서를 열고 아래의 색상 맵 보고서를 엽니다 .

그림 16.5 PB 와 DSD 비교를 위한 색상 맵

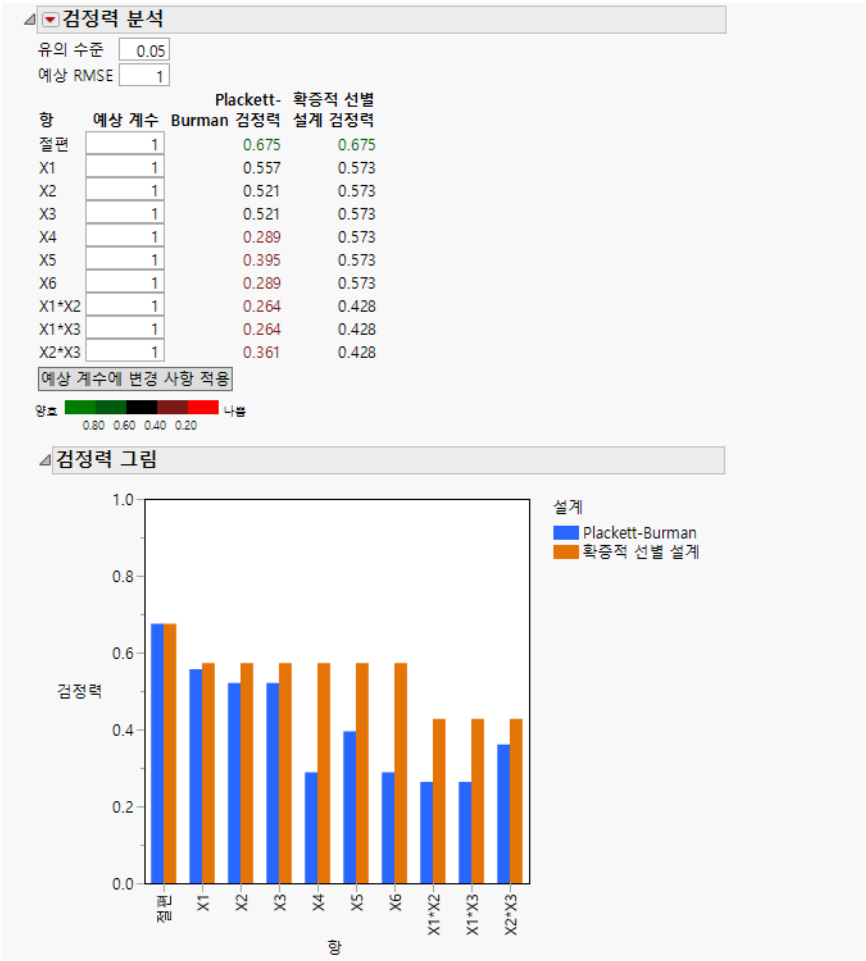


" 상관 색상 맵 " 그림을 보면 PB 설계에서는 주효과와 이원 교호작용이 별칭 관계를 가집니다 . 반면에 DSD 에서는 주효과와 이원 교호작용이 별칭 관계를 갖지 않습니다 .

일부 이원 교호작용이 활성인 경우 설계를 비교하는 방법에 대해 더 자세히 알아보려면 " 모형 " 개요에서 이원 교호작용을 추가합니다 .

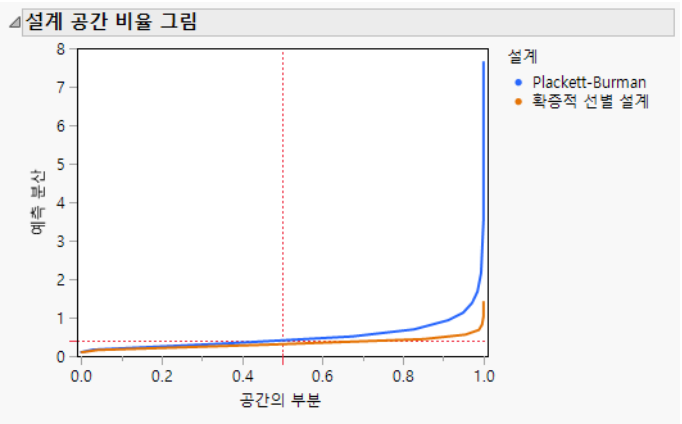
2. " 요인 " 개요에서 X1 ~ X3 을 선택합니다 .
3. " 모형 " 개요에서 **교호작용 > 2 차**를 선택합니다 .

그림 16.6 교호작용이 있는 PB 와 DSD 비교를 위한 검정력 분석



" 항 " 목록에 세 가지 이원 교호작용이 표시됩니다 . 이러한 이원 교호작용이 활성이면 모든 효과에서 검정력을 비교할 때 DSD 가 PB 보다 성능이 우수합니다 .

그림 16.7 교호작용이 있는 PB 와 DSD 비교를 위한 예측 분산



또한 모형의 세 가지 교호작용에 대한 예측 분산을 보면 DSD 가 PB 설계보다 우수합니다 . 다른 보고서를 탐색하여 잠재적 활성 교호작용이 있을 때 DSD 가 더 적합하다는 것을 확인할 수 있습니다 .

런 크기가 다른 설계

이 예에서는 런 크기가 각각 16, 20, 24 인 세 가지 설계를 비교합니다 . 주효과 모형에 대한 설계가 구성됩니다 . 설계 비교 플랫폼을 사용하면 런 크기가 클 때 얻을 수 있는 잠재적 이점이 추가 리소스 비용을 감수할 만큼 가치가 있는지 확인할 수 있습니다 .

1. **도움말 > 샘플 데이터**를 선택하고 **샘플 스크립트 디렉터리 열기**를 클릭한 후 Compare Three Run Sizes.jsl 을 선택합니다 .
2. 스크립트 창을 마우스 오른쪽 버튼으로 클릭하고 **스크립트 실행**을 선택합니다 .

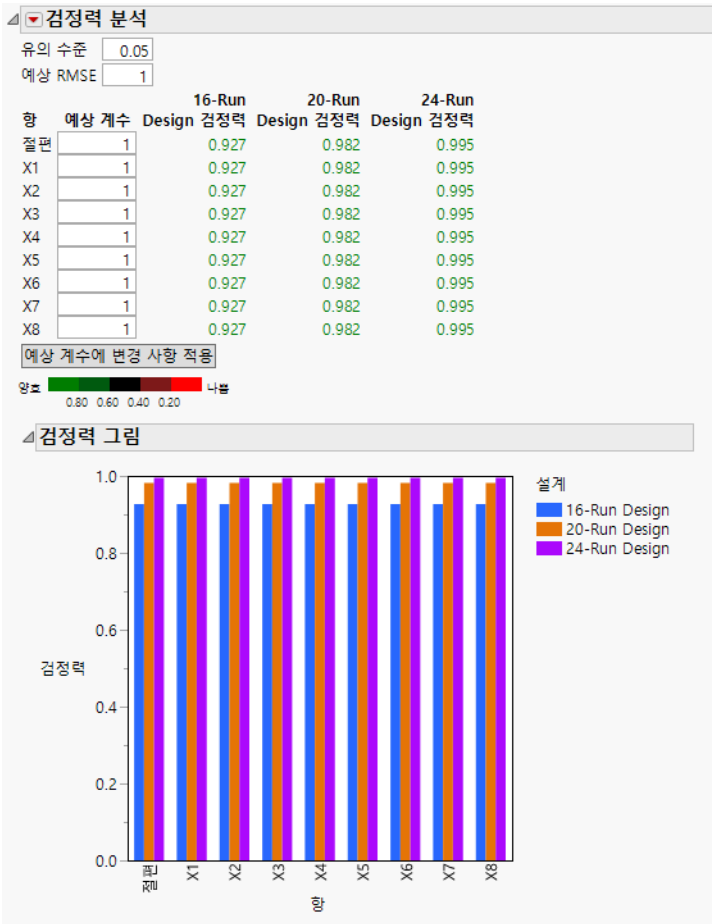
   사용자 설계를 사용하여 다음 세 가지 설계 테이블이 구성됩니다 . 여기에는 주효과만 " 모형 " 개요의 항목으로 포함됩니다 .

   – 16-Run Design
   – 20-Run Design
   – 24-Run Design

   이 세 가지 설계를 비교하려고 합니다 . 16-Run Design 테이블이 활성화되어 있습니다 .
3. 16-Run Design 데이터 테이블에서 **DOE > 설계 진단 > 설계 비교**를 선택합니다 .
4. **'16-Run Design' 을 (를) 다음과 비교** 목록에서 20-Run Design 과 24-Run Design 을 선택합니다 .

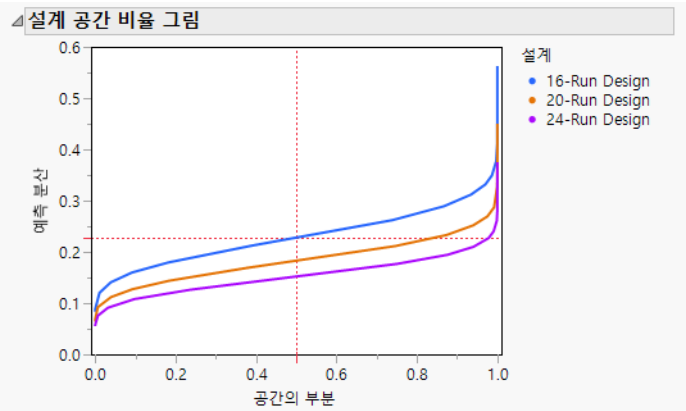
   각 설계에 대한 패널이 시작 창에 추가됩니다 . JMP 는 세 가지 설계 테이블에 나타나는 순서에 따라 자동으로 열을 매칭합니다 .
5. **확인**을 클릭합니다 .

그림 16.8 검정력 분석 비교



계수가 " 예상 RMSE" 의 차수인 경우 세 가지 설계 모두 주효과를 감지할 수 있는 검정력이 높습니다 .

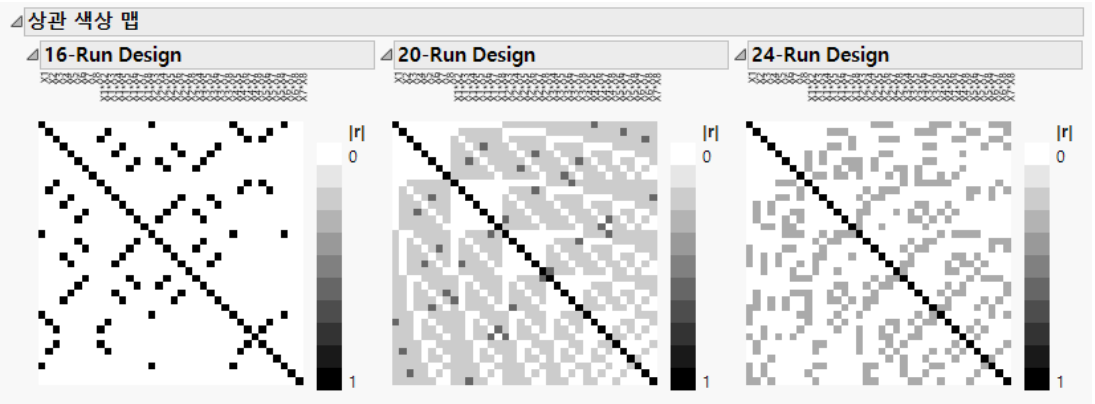
그림 16.9 설계 공간 비율 비교



전체 설계 공간에 대한 예측 분산을 보면 예상대로 24 회 런 설계가 다른 두 설계보다 우수합니다 . 20 회 런 설계는 16 회 런 설계보다 우수합니다 .

6. "절대 상관" 보고서에서 "상관 색상 맵" 보고서를 열고 아래 3 개의 색상 맵 보고서를 엽니다.

그림 16.10 상관 색상 맵 비교

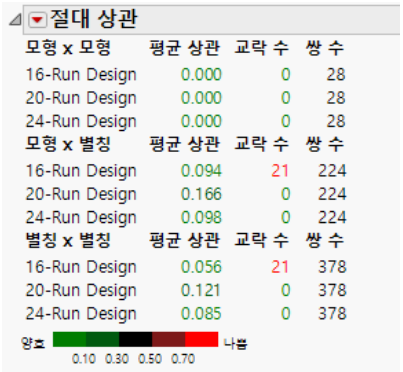


16 회 런 설계의 " 상관 색상 맵 " 은 일부 주효과와 일부 2 요인 교호작용 간의 교락 및 2 요인 교호작용의 교락이 있음을 나타냅니다 .

20 회 런 설계의 " 상관 색상 맵 " 은 일부 주효과와 일부 2 요인 교호작용 간의 상관 및 일부 2 요인 교호작용 간의 상관이 크다는 것을 나타냅니다 .

24 회 런 설계는 주효과와 2 요인 교호작용 간의 상관 및 2 요인 교호작용 간의 상관이 중간 정도로 보통이라는 것을 나타냅니다 .

그림 16.11 절대 상관 비교



절대 상관

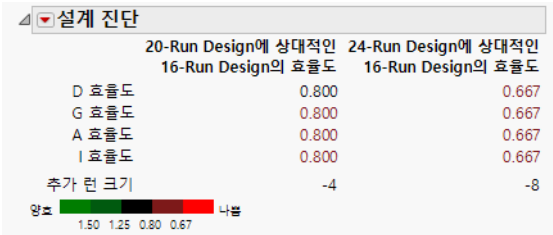
모형 x 모형	평균 상관	교락 수	쌍 수
16-Run Design	0.000	0	28
20-Run Design	0.000	0	28
24-Run Design	0.000	0	28
모형 x 별칭	**평균 상관**	**교락 수**	**쌍 수**
16-Run Design	0.094	21	224
20-Run Design	0.166	0	224
24-Run Design	0.098	0	224
별칭 x 별칭	**평균 상관**	**교락 수**	**쌍 수**
16-Run Design	0.056	21	378
20-Run Design	0.121	0	378
24-Run Design	0.085	0	378

양호 나쁨
0.10 0.30 0.50 0.70

" 절대 상관 " 테이블에는 " 상관 색상 맵 " 에 표시된 정보가 요약됩니다 . 세 가지 설계 모두 주효과만으로 모형이 구성되고 별칭 행렬에 2 요인 교호작용이 포함됩니다 .

16 회 런 설계의 경우 테이블의 " 모형 x 별칭 " 부분은 주효과와 2 요인 교호작용 간의 교락이 9 개 있음을 나타냅니다 . " 별칭 x 별칭 " 부분은 6 개의 2 요인 교호작용이 교락되어 있음을 나타냅니다 .

그림 16.12 설계 진단 비교



설계 진단

	20-Run Design에 상대적인 16-Run Design의 효율도	24-Run Design에 상대적인 16-Run Design의 효율도
D 효율도	0.800	0.667
G 효율도	0.800	0.667
A 효율도	0.800	0.667
I 효율도	0.800	0.667
추가 런 크기	-4	-8

양호 나쁨
1.50 1.25 0.80 0.67

" 설계 진단 " 보고서는 여러 가지 효율도 측도에 관해 16 회 런 설계의 효율도를 20 회 런 설계 및 24 회 런 설계와 모두 비교합니다 . 상대 효율도 값이 1 을 초과하면 주어진 측도에 대해 참조 설계가 바람직하다는 것을 나타냅니다 . 값이 1 보다 작으면 참조 설계와 비교되는 설계가 바람직하다는 것을 나타냅니다 . 16 회 런 설계는 모든 지표에서 다른 두 설계보다 효율도가 낮으므로 더 큰 설계가 바람직하다는 것을 나타냅니다 .

주구 수가 다른 분할구 설계

이 예에서는 주구 수가 다른 두 분할구 설계를 비교합니다 . 다음과 같은 세 요인에 대한 설계입니다 .

- 변경하기 힘든 연속형 요인
- 변경하기 쉬운 연속형 요인
- 변경하기 쉬운 3 수준 범주형 요인

설계는 모든 2 요인 교호작용을 가정된 모형에 포함합니다. 20 회 런이 가능하고 4 개 또는 8 개의 주구를 사용하여 비교하려고 합니다.

설계 비교 시작

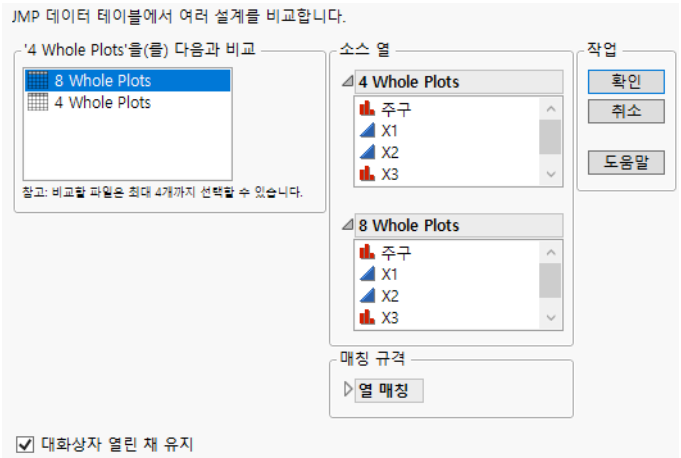
1. **도움말 > 샘플 데이터**를 선택하고 **샘플 스크립트 디렉터리 열기**를 클릭한 후 Compare Split Plots.jsl 을 선택합니다.
2. 스크립트 창을 마우스 오른쪽 버튼으로 클릭하고 **스크립트 실행**을 선택합니다.

   사용자 설계를 사용하여 다음 두 설계 테이블이 구성됩니다.
   - 4 Whole Plots
   - 8 Whole Plots

   이 두 설계를 비교하려고 합니다. 4 Whole Plots 테이블이 활성화되어 있습니다.
3. 4 Whole Plots 테이블에서 **DOE > 설계 진단 > 설계 비교**를 선택합니다.
4. **'4 Whole Plots' 을 (를) 다음과 비교** 목록에서 8 Whole Plots 를 선택합니다.

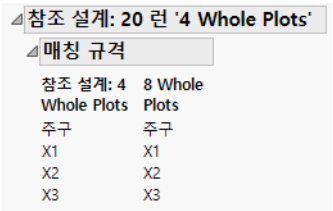
   이 설계에 대한 패널이 시작 창에 추가됩니다. JMP 는 두 설계 테이블에 나타나는 순서에 따라 자동으로 열을 매칭합니다.

그림 16.13 완료된 시작 창



5. **확인**을 클릭합니다.
6. " 참조 설계 : 20 런 '4 Whole Plots' " 아래의 **매칭 규격** 개요를 엽니다.

그림 16.14 분할구 설계를 위한 매칭 규격

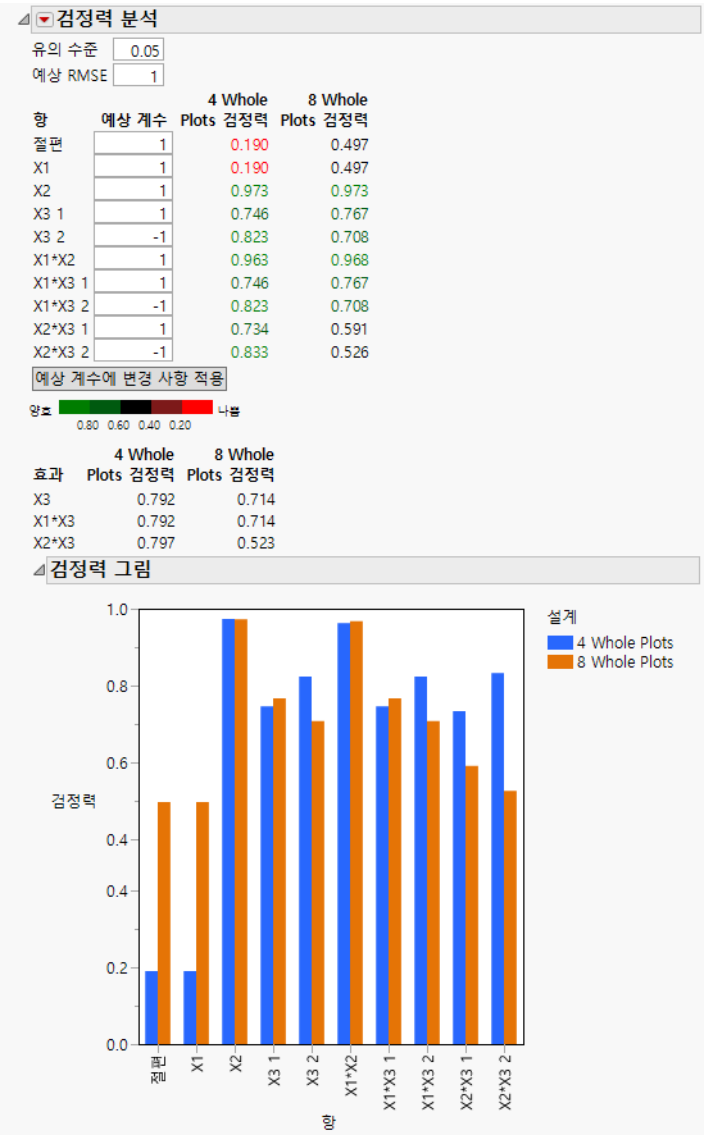


"주구" 열은 설계의 일부로 입력됩니다. 설계 비교를 수행하려면 주구 구조를 알아야 하므로 이 열이 필요합니다.

보고서 검토

"설계 평가" 보고서는 두 설계를 비교하는 다양한 진단 결과를 제공합니다.

그림 16.15 두 분할구 설계에 대한 검정력 분석



"검정력 분석" 보고서는 주구 요인 X1 에 대한 검정력의 경우 4 주구 설계(0.19)가 8 주구 설계(0.497)보다 훨씬 작습니다. 그러나 4 주구 설계는 분할구 효과, 특히 두 분할구 요인의 교호작용 X2*X3 을 감지할 수 있는 검정력(0.797)이 다른 설계(0.523)보다 높습니다. 결합 효과 X2*X3 에 대한 검정력은 색상 막대 및 범례 아래에 제공됩니다.

그림 16.16 분할구 설계를 비교하는 상대 추정 효율도

상대 추정 효율도

항	8 Whole Plots에 상대적인 4 Whole Plots의 효율도
절편	0.778
X1	0.778
X2	1.004
X3 1	0.924
X3 2	1.154
X1*X2	0.982
X1*X3 1	0.924
X1*X3 2	1.154
X2*X3 1	1.107
X2*X3 2	1.449

양호 나쁨
1.50 1.25 0.80 0.67

추정값의 상대 표준 오차

항	4 Whole Plots	8 Whole Plots
절편	0.553	0.43
X1	0.553	0.43
X2	0.23	0.231
X3 1	0.25	0.231
X3 2	0.22	0.254
X1*X2	0.24	0.235
X1*X3 1	0.25	0.231
X1*X3 2	0.22	0.254
X2*X3 1	0.256	0.283
X2*X3 2	0.218	0.315

"상대 추정 효율도" 보고서는 X1 의 상대 추정 효율도가 0.778 임을 보여 줍니다. 이는 X1 의 표준 오차가 8 주구 설계보다 4 주구 설계에서 현저히 크다는 것을 나타냅니다.

"추정값의 상대 표준 오차" 보고서를 엽니다. 4 주구 설계에서 X1 의 상대 표준 오차가 0.553 이고 8 주구 설계의 오차는 0.43 임을 확인할 수 있습니다.

"상대 추정 효율도" 보고서에서 X2*X3 2 의 상대 추정 효율도는 1.449 이며, 이는 X2*X3 2 와 관련된 모수의 표준 오차가 4 주구 설계보다 8 주구 설계에서 현저히 크다는 것을 나타냅니다.

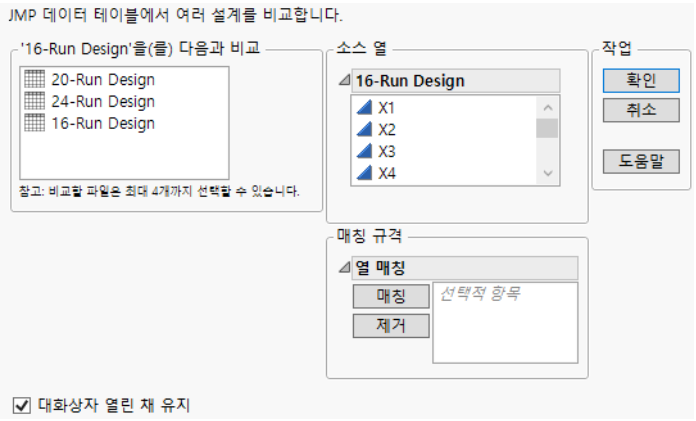
"검정력 분석" 및 "상대 추정 효율도" 보고서는 설계 선택이 주구 효과 X1 감지의 중요성과 관련이 있음을 나타냅니다. 8 주구 설계를 사용하면 주구 효과를 더 잘 감지할 수 있습니다. 4 주구 설계는 범주형 변수를 포함하는 분할구 효과를 감지하는 데 약간 더 좋습니다.

설계 비교 플랫폼 시작

DOE > 설계 진단 > 설계 비교를 선택하여 설계 비교 플랫폼을 시작합니다. 열려 있는 모든 데이터 테이블이 왼쪽 목록에 나타납니다. 활성 데이터 테이블과 열은 "소스 열" 패널에 나타납니다. 초기 "소스 열" 패널의 설계는 **참조 설계**, 즉 다른 설계와 비교되는 설계입니다. 참조 설계와 비교할 설계를 추가하면 해당 열이 참조 설계 패널 아래의 패널에 나타납니다.

그림 16.17 에서는 " 런 크기가 다른 설계 "(444 페이지) 의 세 가지 설계에 대한 시작 창을 보여 줍니다 .

그림 16.17 설계 비교 시작 창



설계 테이블 선택

왼쪽 목록에서 최대 4 개의 설계 테이블을 선택합니다 .

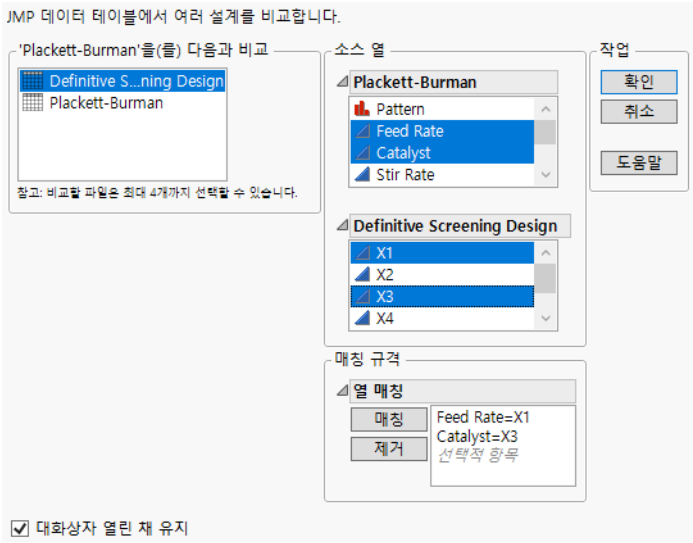
- 여러 설계를 참조 설계와 비교하려면 왼쪽 목록에서 해당 설계 테이블을 동시에 선택해야 합니다 .
- " 소스 열 " 목록에서 하나 이상의 설계를 바꾸려면 왼쪽의 목록에서 원하는 테이블을 선택합니다 . 그러면 참조 설계 테이블 아래의 설계 테이블이 바뀝니다 .

참고 : 참조 설계 테이블을 자체적으로 비교할 수 있으며 이렇게 하면 요인에 대한 설계 열 할당을 탐색할 때 유용합니다 .

열 매칭

" 열 매칭 " 패널에서 각 설계 테이블의 열 중 서로 일치하는 열을 지정합니다 . 열을 매칭하려면 " 소스 열 " 목록의 각 설계 테이블에서 매칭할 열을 선택한 후 " 매칭 "을 클릭합니다 .

그림 16.18 매칭할 열 선택



- 각 목록의 단일 열을 매칭하려면 각 목록에서 단일 열을 선택한 후 " 매칭 " 을 클릭합니다 .
- 각 목록에서 올바른 매칭 순서로 나타나는 여러 열을 매칭하려면 각 목록에서 해당 열을 선택합니다 . " 매칭 " 버튼을 클릭합니다 . 그러면 목록 순서대로 매칭됩니다 (그림 16.18 참조). 이 예에서 Feed Rate 는 X1 과 매칭되고 Catalyst 는 X3 과 매칭됩니다 .
- 목록에 같은 수의 열이 있는 경우 목록에 나오는 순서대로 매칭하려면 " 매칭 " 을 클릭할 필요가 없습니다 . " 확인 " 을 클릭하여 시작 창이 실행되면 JMP 에서 열을 목록에 나오는 순서대로 자동 매칭합니다 . 보고서의 " 매칭 규격 " 개요에서 매칭을 검토할 수 있습니다 .

설계 비교 창 : 모형 및 별칭 항 지정

" 설계 비교 " 창은 다음과 같은 두 가지 개요 집합으로 구성됩니다 .

- " 요인 ", " 모형 " 및 " 별칭 항 " 개요를 사용하여 모형에 있는 효과와 잠재적 활성 효과를 지정합니다 .
- " 설계 평가 " 개요의 진단을 사용하여 설계를 비교합니다 . " 모형 " 및 " 별칭 항 " 개요에서 변경한 내용은 " 설계 평가 " 보고서에 업데이트됩니다 .

"설계 비교" 보고서는 참조 설계의 열 이름을 사용합니다.

이 섹션에서는 "참조 설계", "요인", "모형" 및 "별칭 항" 개요에 대해 설명합니다. "설계 평가" 개요에 대한 설명은 "설계 비교 창: 설계 평가"(454 페이지)에서 확인하십시오.

참조 설계

참조 설계에 대한 창 이름이 개요 제목에 나타납니다. "매칭 규격" 개요에는 시작 창에 입력한 규격이 나열됩니다.

요인

"요인" 개요를 사용하여 "모형" 및 "별칭 항" 목록에 효과를 추가합니다.

"요인" 개요에는 참조 설계의 열 이름과 코드화된 값을 사용하여 요인이 나열됩니다. 주구 및 하위구 열은 요인이 아니므로 "요인" 개요에 나타나지 않습니다. 그러나 이 열은 분석을 위해 필요합니다.

모형

모형에 있어야 한다고 생각되는 효과에 대해 설계를 비교하기 위해 효과를 추가하거나 제거합니다. 초기에는 참조 설계 테이블의 "모형" 스크립트에 있는 효과 및 비교되는 모든 설계에서 추정하는 효과가 "모형" 개요에 나열됩니다. 참조 설계 테이블에 "모형" 스크립트가 없는 경우에는 비교되는 모든 설계에서 추정할 수 있는 주효과만 "모형" 개요에 표시됩니다. 효과를 추가하고 제거하는 방법에 대한 자세한 내용은 "설계 평가" 장의 "모형"(417페이지)에서 확인하십시오.

참고 : 설계가 과포화되어 추정할 모수의 수가 런 수를 초과하면 추정할 수 있는 효과 집합만 " 모형 " 개요에 나열됩니다 .

별칭 항

활성 가능성이 있는 효과에 대해 설계를 비교하기 위해 효과를 추가하거나 제거합니다. 초기에는 "모형" 개요에 없는 모든 2 요인 교호작용이 "별칭 항" 개요에 포함됩니다. 이 개요의 효과는 "별칭 행렬 요약" 및 "절대 상관" 개요의 계산에 영향을 줍니다. 자세한 내용은 "별칭 행렬 요약"(463 페이지) 및 "절대 상관"(464 페이지)에서 확인하십시오.

효과를 추가하고 제거하는 방법에 대한 자세한 내용은 "설계 평가" 장의 "별칭 항"(418 페이지)에서 확인하십시오.

설계 비교 창 : 설계 평가

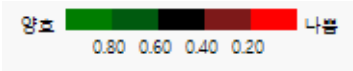
- " 검정력 분석 "
- " 예측 분산 프로파일 "
- " 설계 공간 비율 그림 "
- " 상대 추정 효율도 "

- "별칭 행렬 요약"
- "절대 상관"
- "설계 진단"

색상 대시보드

"설계 평가"의 몇몇 개요는 색상 막대에 따라 색상이 적용된 값을 보여 줍니다. 색상은 진단 측도에 적용되며 양호한 동작 또는 나쁜 동작을 반영하는 값(설계)을 확인하는 데 도움이 됩니다. 범례 값을 편집하여 양호한 동작과 나쁜 동작의 정의를 반영하는 색상을 적용할 수 있습니다.

그림 16.19 색상 대시보드



개요의 빨간색 삼각형 메뉴에서 다음 두 옵션을 선택하거나 색상 막대를 마우스 오른쪽 버튼으로 클릭하여 색상 막대를 수정할 수 있습니다.

범례 값 표시 색상 막대 아래에 나타나는 값을 표시하거나 숨깁니다.

범례 값 편집 색상을 정의하는 값을 지정합니다.

검정력 분석

검정력은 주어진 크기의 활성 효과를 감지할 확률입니다. "검정력 분석" 보고서는 설계에서 실제적으로 중요한 효과를 감지하는 능력을 평가하고 비교하는 데 도움이 됩니다. 각 설계의 "검정력 분석" 보고서는 "모형" 개요에 있는 효과에 대한 검정력을 계산합니다.

"검정력 분석" 보고서는 개별 모형 모수 및 전체 효과에 대한 검정력을 제공합니다. 또한 "검정력 그림"과 "검정력과 표본 크기 비교" 그림도 제공합니다.

검정력은 런 수, 유의 수준 및 추정 오차 변동에 따라 달라집니다. 검정력을 계산하는 방법에 대한 자세한 내용은 "기술 상세 정보" 부록의 "검정력 계산"(768 페이지)에서 확인하십시오.

그림 16.20 세 가지 설계에 대한 검정력 분석 개요

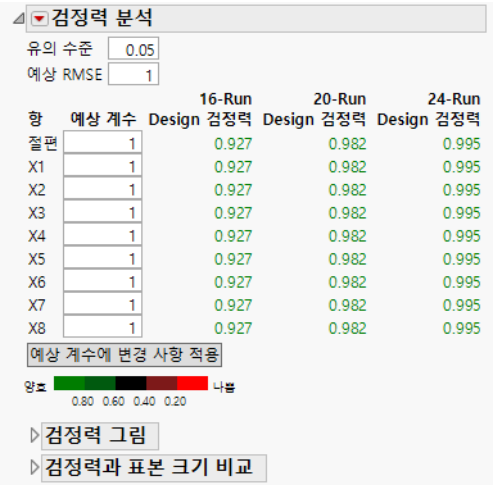


그림 16.20에서는 "런 크기가 다른 설계"(444 페이지)에서 구성한 세 가지 설계의 "검정력 분석" 개요를 보여 줍니다. 두 가지 이원 교호작용이 "모형" 개요에 추가되었습니다.

검정력 분석 보고서

"유의 수준" 및 "예상 RMSE"에 값을 지정하면 모형 모수에 대한 검정력을 계산할 때 이 값이 사용됩니다. 감지할 차이를 반영하는 계수 값을 "예상 계수"로 입력합니다. 모든 설계의 결과를 업데이트하려면 "예상 계수에 변경 사항 적용"을 클릭합니다.

유의 수준 효과가 없다는 가설이 참인데도 이를 기각할 확률입니다. 값을 입력하면 검정력 계산이 즉시 업데이트됩니다.

예상 RMSE 오차 변동의 제곱근에 대한 추정값입니다. 값을 입력하면 검정력 계산이 즉시 업데이트됩니다.

"예상 계수에 변경 사항 적용" 버튼 아래에 나오는 색상 그래디언트에 따라 검정력 값에 색상이 적용됩니다. "검정력 분석"의 빨간색 삼각형 메뉴 옵션을 사용하여 색상 범례를 제어할 수 있습니다. 자세한 내용은 "색상 대시보드"(455 페이지)에서 확인하십시오.

검정력 그림에 대한 자세한 내용은 "검정력 그림"(457 페이지)에서 확인하십시오.

참고 : 설계가 과포화되어 추정할 모수의 수가 런 수를 초과하면 추정할 수 있는 효과 집합만 "검정력 분석" 개요에 나열됩니다.

개별 모수 검정

" 항 " 열에는 모형 항이 나열됩니다 . 각 항의 " 예상 계수 " 열에는 해당 항의 값이 포함됩니다 . " 검정력 " 값은 예상 계수에 의해 계수의 참 값이 주어진 경우 해당 항의 계수가 0 이라는 검정의 검정력입니다 . 이때 설계 및 " 모형 " 개요의 항이 제공됩니다 .

항 검정 대상 계수와 관련된 모형 항입니다 .

예상 계수 모형 항과 관련된 계수의 값입니다 . 이 값은 검정력 계산에 사용됩니다 . " 예상 계수 " 열에 새 값을 설정하는 경우 **예상 계수에 변경 사항 적용**을 클릭하여 " 검정력 " 계산을 업데이트합니다 .

참고 : 연속형 효과에 대한 예상 계수의 기본값은 1 입니다 . 범주형 효과의 경우에는 1 과 –1 값이 교대로 나타납니다 .

검정력 지정된 예상 계수에 의해 계수의 참 값이 주어진 경우 효과가 없다는 귀무가설을 기각할 확률입니다 .

- 수치형 요인과 관련된 계수의 경우 상한 설정에서 하한 설정까지 평균 반응의 변화 (모형 기반) 는 계수 값의 두 배입니다 .
- 범주형 요인과 관련된 계수의 경우 전체 요인 수준에서 평균 반응의 변화 (모형 기반) 는 예상 계수 절대값의 두 배와 같습니다 . 범주형 요인의 검정력에 대한 자세한 내용은 " 수준이 세 개 이상인 범주형 효과 검정 "(457 페이지) 에서 확인하십시오 .

지정된 유의 수준과 예상 RMSE 가 계산에 사용됩니다 . 검정력 계산에 대한 자세한 내용은 " 기술 상세 정보 " 부록의 " 단일 모수에 대한 검정력 "(768 페이지) 에서 확인하십시오 .

예상 계수에 변경 사항 적용 " 예상 계수 " 열에 새 값을 설정하는 경우 **예상 계수에 변경 사항 적용**을 클릭하여 " 검정력 " 값을 업데이트합니다 .

수준이 세 개 이상인 범주형 효과 검정

수준이 세 개 이상인 범주형 효과가 모형에 포함된 경우 " 예상 계수에 변경 사항 적용 " 버튼 아래에 다음 열이 나타납니다 .

효과 범주형 효과입니다 .

검정력 효과 없음 검정에 대한 검정력 계산입니다 . 검정의 귀무가설은 효과에 해당하는 모든 모형 모수가 0 이라는 것입니다 . 감지할 차이는 " 예상 계수 " 열에서 효과의 모형 항에 해당하는 값으로 정의됩니다 . 검정력 계산은 예상 계수에 의해 결정된 반응 평균의 차이를 반영합니다 .

지정된 유의 수준과 예상 RMSE 가 계산에 사용됩니다 . 검정력 계산에 대한 자세한 내용은 " 기술 상세 정보 " 부록의 " 범주형 효과에 대한 검정력 "(769 페이지) 에서 확인하십시오 .

검정력 그림

"검정력 그림"은 검정력 분석의 검정력 값을 그래픽 형식으로 보여 줍니다 . 이 그림은 각 효과 및 각 설계에 대한 검정력을 병렬 막대 차트로 표시합니다 .

그림 16.21 세 가지 설계에 대한 검정력 그림

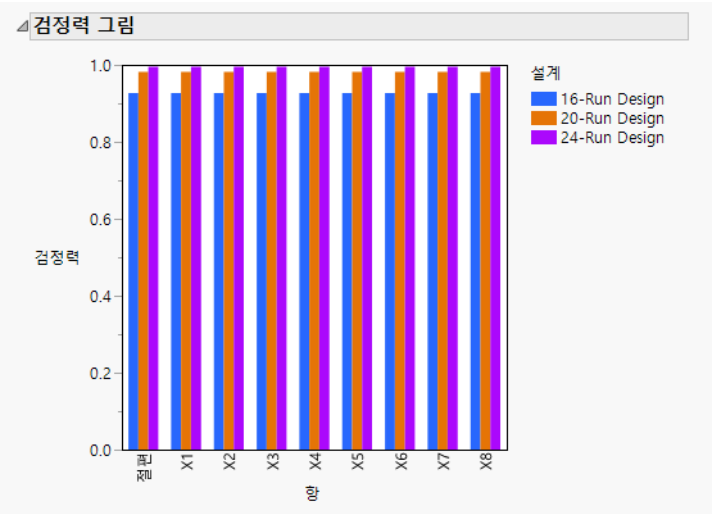
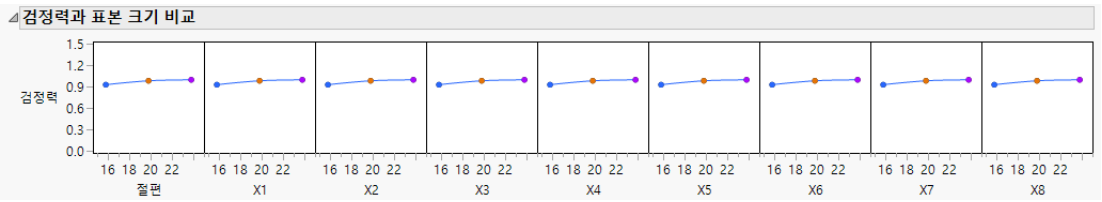


그림 16.21의 "검정력 그림"은 "런 크기가 다른 설계"(444페이지)에서 구성한 세 가지 설계의 주 효과를 대상으로 합니다.

검정력과 표본 크기 비교

"검정력과 표본 크기 비교" 그림은 비교하는 설계의 런 크기가 다른 경우에만 나타납니다. 이 그림을 사용하면 표본 크기가 모형의 각 효과에 대한 검정력에 미치는 영향을 확인할 수 있습니다. "검정력 그림" 그래프와 동일한 정보를 제공하지만 형식이 다릅니다. 정수 표본 크기의 검정력 값이 선분으로 연결됩니다.

그림 16.22 세 가지 설계에 대한 검정력과 표본 크기 비교 프로파일러



예측 분산 프로파일

"예측 분산 프로파일" 개요에서는 비교 중인 각 설계에 대한 상대 예측 분산 프로파일러를 보여줍니다. 각 그림은 상대 예측 분산을 다른 요인의 값이 고정된 상태에서 각 요인의 함수로 표시합니다.

설계 공간에서 참조 설계에 대한 상대 예측 분산의 최대값을 찾으려면 "예측 분산 프로파일" 옆의 빨간색 삼각형에서 **분산 최대화** 옵션을 선택합니다. 자세한 내용은 "분산 최대화"(459 페이지)에서 확인하십시오.

그림 16.23 세 가지 설계에 대한 예측 분산 프로파일

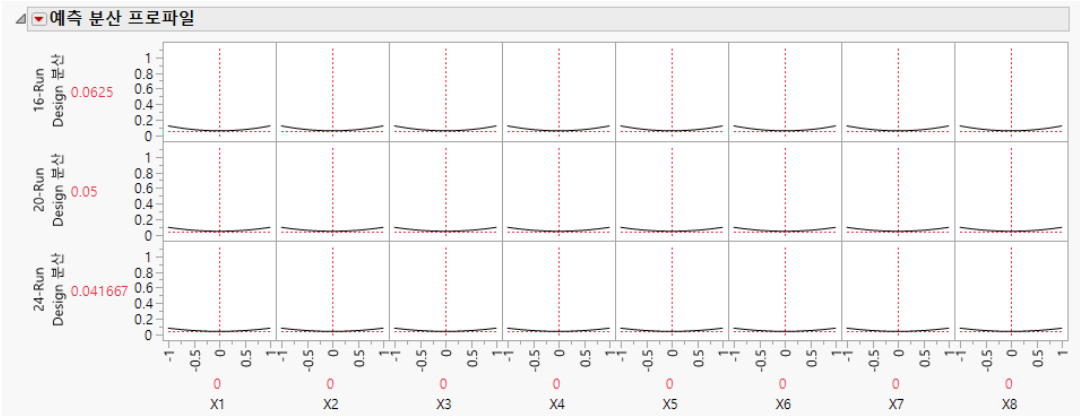


그림 16.23 의 "예측 분산 프로파일" 그림은 "런 크기가 다른 설계"(444 페이지)에서 구성한 세 가지 설계를 대상으로 합니다. 두 가지 이원 교호작용(X1*X3, X2*X3)이 "모형" 개요에 추가되었습니다. 그림에서 각 연속형 요인의 초기값은 설계 설정의 중간점입니다. 왼쪽의 "분산" 값은 런 수가 증가할수록 중앙점에서 분산이 감소함을 나타냅니다.

상대 예측 분산

주어진 요인 설정에 대한 예측 분산은 설계 및 요인 설정에 따라 달라지는 양과 오차 분산의 곱입니다. 실험을 실행하기 전에는 오차 분산을 알 수 없으므로 예측 분산도 알 수 없습니다. 그러나 오차 분산에 대한 예측 분산의 비는 오차 분산의 함수가 아닙니다. **상대 예측 분산**이라고 하는 이 비율은 설계와 요인 설정에만 영향을 받습니다. 따라서 데이터를 얻기 전에 상대 예측 분산을 계산할 수 있습니다. 자세한 내용은 "기술 상세 정보" 부록의 "상대 예측 분산"(770 페이지)에서 확인하십시오.

실험을 실행하고 최소 제곱 모형을 적합시킨 후 모형 적합의 MSE(평균 제곱 오차)를 사용하여 오차 분산을 추정할 수 있습니다. 설정에 대한 상대 예측 분산을 곱해 설정에서 실제 예측 분산을 추정할 수 있습니다.

설계 공간 전체에서 예측 분산이 작은 것이 이상적입니다. 일반적으로 표본 크기가 증가하면 오차 분산이 감소합니다. 설계를 비교할 때 평균적으로 예측 분산이 낮은 설계가 선호됩니다.

분산 최대화

최대 상대 예측 분산의 측면에서 설계를 평가하거나 비교할 수도 있습니다. "예측 분산 프로파일" 옆의 빨간색 삼각형에서 **분산 최대화** 옵션을 선택합니다. JMP는 상대 예측 분산을 최대화하는 만족도 함수를 사용합니다. "예측 분산 프로파일"의 "분산" 값은 상대 예측 분산의 최악 값(설계 관점에서 가장 적합하지 않음)입니다.

그림 16.24 세 가지 설계에 대한 최대 분산을 보여 주는 예측 분산 프로파일

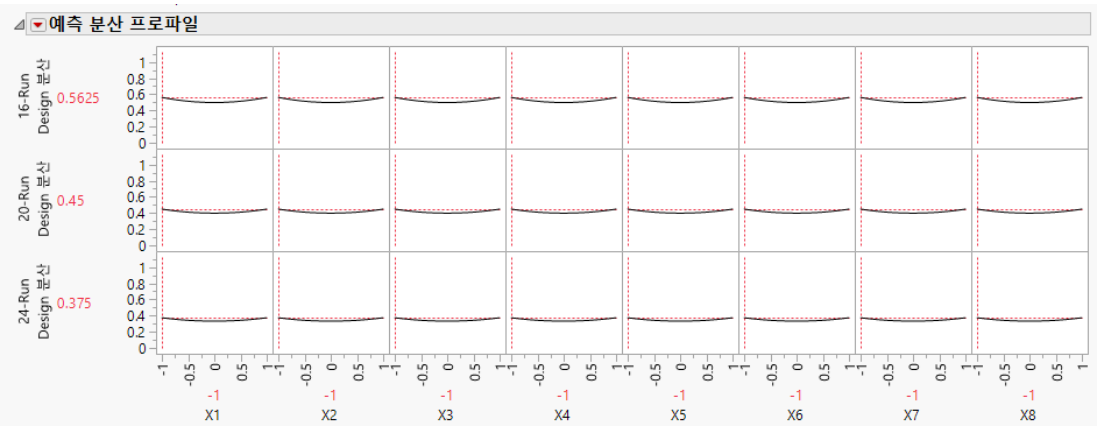


그림 16.24에서는 "런 크기가 다른 설계"(444 페이지)에서 구성한 세 가지 설계에 대해 "분산 최대화"를 선택한 후의 "예측 분산 프로파일"을 보여 줍니다. 예상대로 런 크기가 증가하면 최대 상대 예측 분산이 감소합니다. 이 최악 상대 분산 값을 제공하는 요인의 값도 그림에 표시됩니다. 그러나 여러 설정에서 동일한 최대 상대 분산을 산출할 수 있음을 염두에 두십시오.

설계 공간 비율 그림

"설계 공간 비율 그림"은 설계 공간에서 상대 예측 분산이 특정 값 미만이 되는 비율을 보여 줍니다.

그림 16.25 세 가지 설계에 대한 설계 공간 비율 그림

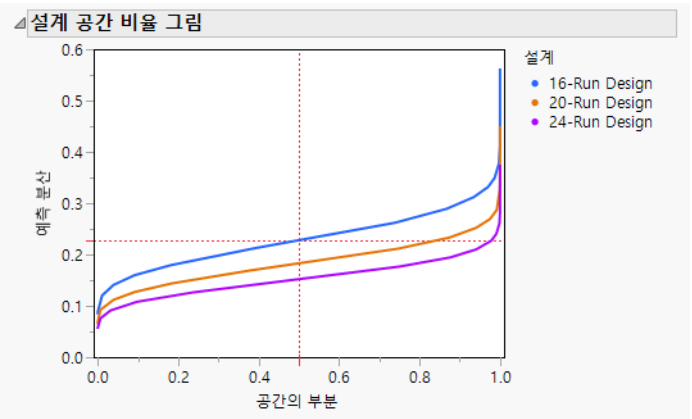


그림 16.25 에서는 " 런 크기가 다른 설계 "(444 페이지) 에서 구성한 세 가지 설계에 대한 " 설계 공간 비율 그림 " 을 보여 줍니다 . 다음 사항에 유의하십시오 .

- 그림의 *X* 축은 설계 공간의 비율을 나타냅니다 (0 ~ 100% 범위).
- *Y* 축은 상대 예측 분산 값을 나타냅니다 .

- 주어진 곡선에 있는 *(x, y)* 점의 경우 *x* 값은 분산이 *y* 보다 작거나 같은 설계 공간의 비율입니다.
- 빨간색 점선 십자기호는 설계 공간의 50%에서 참조 설계에 대한 상대 예측 분산의 경계를 이루는 값을 표시합니다.

그림 16.25에서는 24회 런 설계의 상대 예측 분산이 다른 두 설계에 비해 균등하게 작다는 것을 보여 줍니다. 20회 런 설계의 예측 분산은 16회 런 설계보다 균등하게 작습니다. 빨간색 점선 십자기호는 설계 공간의 약 50%에 대해 20회 런 설계의 상대 예측 분산이 약 0.23 미만임을 나타냅니다.

십자기호 도구를 사용하여 "공간 비율" 값에 해당하는 최대 상대 예측 분산을 찾을 수 있습니다. 예를 들어 십자기호 도구를 사용하여 24회 런 설계에 대한 예측 분산 값의 90%가 약 0.20 미만인지 확인하십시오.

참고: 설계 공간 비율 그림을 구성할 때 설계 공간의 몬테카를로 표집이 사용되므로 같은 설계에 대한 그림이 약간 다를 수 있습니다.

상대 추정 효율도

"상대 추정 효율도" 보고서는 가정된 모형에 있는 모수에 대한 모수 추정값의 표준 오차 측면에서 설계를 비교합니다. 표준 오차는 모수 추정값의 신뢰 구간 길이를 제어합니다. 이 보고서는 효율도 비율과 상대 표준 오차를 제공합니다.

상대 추정 효율도 값 테이블 아래에 나오는 색상 그래디언트에 따라 상대 추정 효율도 값에 색상이 적용됩니다. "상대 추정 효율도"의 빨간색 삼각형 메뉴 옵션을 사용하여 색상 범례를 제어할 수 있습니다. 자세한 내용은 "색상 대시보드"(455 페이지)에서 확인하십시오.

그림 16.26 두 가지 분할구 설계를 비교하는 상대 추정 효율도

상대 추정 효율도

항	8 Whole Plots에 상대적인 4 Whole Plots의 효율도
절편	0.778
X1	0.778
X2	1.004
X3 1	0.924
X3 2	1.154
X1*X2	0.982
X1*X3 1	0.924
X1*X3 2	1.154
X2*X3 1	1.107
X2*X3 2	1.449

양호 나쁨
1.50 1.25 0.80 0.67

추정값의 상대 표준 오차

항	4 Whole Plots	8 Whole Plots
절편	0.553	0.43
X1	0.553	0.43
X2	0.23	0.231
X3 1	0.25	0.231
X3 2	0.22	0.254
X1*X2	0.24	0.235
X1*X3 1	0.25	0.231
X1*X3 2	0.22	0.254
X2*X3 1	0.256	0.283
X2*X3 2	0.218	0.315

그림 16.26에서는 "주구 수가 다른 분할구 설계"(447페이지)에서 비교한 분할구 설계의 "상대 추정 효율도" 개요를 보여 줍니다.

상대 추정 효율도

주어진 항에 대해 비교 설계에 대한 참조 설계의 추정 효율도는 비교 설계 항의 상대 표준 오차를 참조 설계 항의 상대 표준 오차로 나눈 값입니다. 값이 1보다 작으면 참조 설계가 비교 설계만큼 효율적이지 않음을 나타냅니다. 값이 1보다 크면 참조 설계가 더 효율적임을 나타냅니다.

추정값의 상대 표준 오차

"추정값의 상대 표준 오차" 보고서는 오차 표준편차에 대한 모수 추정값의 표준편차 비율을 제공합니다. 이 값은 오차 표준편차를 기준으로 모형 모수 추정값의 표준 오차 크기를 나타냅니다. i^{th} 모수 추정값의 경우 추정값의 상대 표준 오차는 다음과 같이 정의됩니다.

$$SE = \sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{ii}}$$

다음은 각 요소에 대한 설명입니다.

X 는 "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지) 에 정의된 모형 행렬입니다.

$(\mathbf{X}'\mathbf{X})^{-1}_{ii}$ 은 $(\mathbf{X}'\mathbf{X})^{-1}$ 의 i^{th} 주대각 항목입니다.

별칭 행렬 요약

별칭 행렬은 실제로 활성이지만 모형에 포함되지 않은 항이 모형 항 추정에 영향을 미치는 방식에 대한 문제를 처리합니다. "별칭 항" 개요에서는 가정된 모형에 없지만 모형 항의 추정값을 편향시킬 수 있는 잠재적 활성 효과를 나열합니다. 별칭 행렬 항목은 "별칭 항" 효과에 의해 모형 모수에 전달되는 편향 정도를 나타냅니다. 자세한 내용은 "별칭 항"(454 페이지) 및 "별칭 행렬"(463 페이지)에서 확인하십시오.

"별칭 행렬 요약" 테이블에는 가정된 모형의 항이 나열됩니다. 이러한 항은 "모형" 개요에 나열된 효과에 해당합니다. 설계가 주어지면 "항" 열의 각 항목에 대해 "별칭 항" 개요의 효과에 해당하는 항에 대한 별칭 행렬 항목 제곱합의 제곱근이 계산됩니다. 이 값은 주어진 설계의 "제곱근 평균 제곱 값" 열에 보고됩니다. 예는 "별칭 행렬 요약 값 계산의 예"(464 페이지)에서 확인하십시오.

참고 : "별칭 행렬 요약" 보고서는 "별칭 항" 목록에 효과가 있는 경우에만 나타납니다.

그림 16.27 두 설계에 대한 별칭 행렬 요약

별칭 행렬 요약

항	루트 평균 제곱 값 확증적 선별 설계	루트 평균 제곱 값 Plackett-Burman
절편	0.0000	0.0000
X1	0.0000	0.2722
X2	0.0000	0.2722
X3	0.0000	0.2722
X4	0.0000	0.2722
X5	0.0000	0.2722
X6	0.0000	0.2722
합계	0.0000	0.2520

양호 나쁨
0.10 0.30 0.50 0.70

그림 16.27에서는 "런 크기가 같은 설계"(439 페이지)에서 구성한 Plackett-Burman 설계와 확증적 선별 설계의 "별칭 행렬 요약" 보고서를 보여 줍니다. "모형" 개요에는 주효과만 있습니다. 모든 2 요인 교호작용은 "별칭 항" 목록에 있습니다. 이 테이블에서는 확증적 선별 설계의 경우 주효과가 2 요인 교호작용과 상관 관계가 없음을 보여 줍니다.

"별칭 행렬 요약" 테이블 아래에 나오는 색상 그래디언트에 따라 제곱근 평균 제곱 값에 색상이 적용됩니다. "별칭 행렬 요약"의 빨간색 삼각형 메뉴 옵션을 사용하여 색상 범례를 제어할 수 있습니다. 자세한 내용은 "색상 대시보드"(455 페이지)에서 확인하십시오.

별칭 행렬

별칭 행렬의 행은 "모형" 개요에 나열된 모형 효과에 해당하는 항입니다. 열은 "별칭 항" 개요에 나열된 효과에 해당하는 항입니다. 주어진 행과 열의 항목은 별칭 항이 모형 항에 해당하는 모수 추정값에 영향을 미치는 정도를 나타냅니다.

설계를 평가할 때 별칭 행렬의 항목과 관련하여 다음 두 가지 상황 중 하나가 발생하면 가장 이상적입니다. 항목이 작거나, 그렇지 않은 경우 별칭 항의 효과가 작아서 편향도 작습니다. 별칭 항

의 효과가 상당히 클 것으로 생각되면 해당 항을 모형에 포함하거나 별칭 최적 설계를 고려해야 합니다. 별칭 최적은 별칭 행렬의 제곱 값을 기반으로 합니다.

별칭 행렬에 대한 자세한 배경 정보는 "기술 상세 정보" 부록의 "별칭 행렬"(767 페이지)에서 확인하십시오. Lekivetz, R.(2014) 도 참조하십시오.

별칭 행렬 요약 값 계산의 예

이 예에서는 "별칭 행렬 요약" 개요에 나타나는 값의 계산을 보여 줍니다. 여기서는 주효과만 활성이라는 가정하에 두 설계를 비교합니다 .

1. **도움말 > 샘플 데이터**를 선택하고 **샘플 스크립트 디렉터리 열기**를 클릭한 후 Compare Same Run Size.jsl 을 선택합니다 .
2. 스크립트 창을 마우스 오른쪽 버튼으로 클릭하고 **스크립트 실행**을 선택합니다 .

   다음과 같은 두 개의 13 회 런 설계 테이블이 구성됩니다 .
   - 확증적 선별 설계
   - Plackett-Burman

   Plackett-Burman 설계에만 관심이 있습니다 . 이 테이블이 활성 테이블입니다 .
3. Plackett-Burman 테이블에서 **DOE > 설계 진단 > 설계 평가**를 선택합니다 .
4. X1 ~ X6 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 별칭 항 " 개요를 열어 모든 2 요인 교호작용이 " 별칭 항 " 목록에 있는지 확인합니다 .
7. " 별칭 행렬 " 개요를 엽니다 .

   " 효과 " 열에 나열된 각 모형 항에 대해 주어진 열의 해당 행에 있는 항목은 별칭 항이 모형 항에 해당하는 모수 추정값에 영향을 미치는 정도를 나타냅니다 .

예를 들어 그림 16.27 의 X1 에 해당하는 별칭 행렬 요약 항목을 얻으려면 별칭 행렬에서 X1 에 대한 행의 항을 제곱하여 평균을 구한 후 제곱근을 취합니다. 결과 값은 0.2722 입니다.

절대 상관

"절대 상관" 보고서에는 모형 항과 별칭 항 간의 상관관계 정보가 요약됩니다.

그림 16.28 세 가지 설계에 대한 절대 상관 보고서

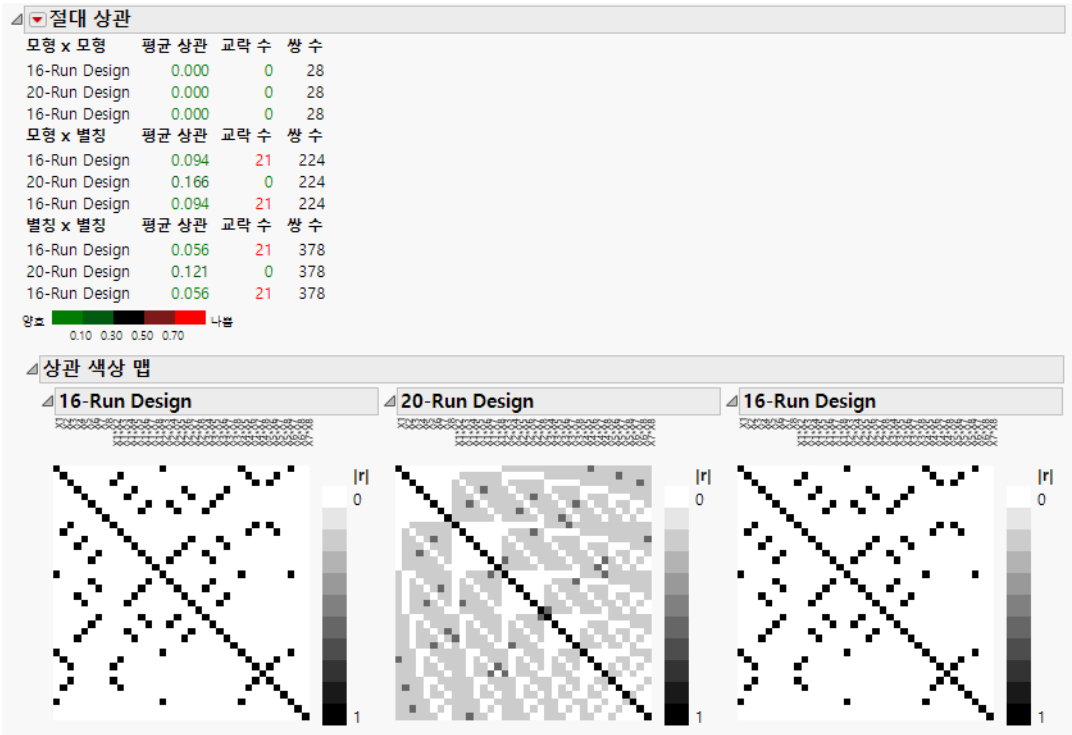


그림 16.28에서는 "런 크기가 다른 설계"(444 페이지)에서 구성한 세 가지 설계의 "절대 상관" 보고서를 보여 줍니다. "모형" 개요에는 주효과만 있습니다.

절대 상관 테이블

"절대 상관" 보고서의 테이블은 다음 세 섹션으로 구분됩니다.

- "모형 x 모형" 은 "모형" 목록의 효과에 해당하는 항 간의 상관을 고려합니다.
- "모형 x 별칭" 은 "모형" 목록의 효과에 해당하는 항과 "별칭" 목록의 효과에 해당하는 항 간의 상관을 고려합니다.
- "별칭 x 별칭" 은 "별칭" 목록의 효과에 해당하는 항 간의 상관을 고려합니다.

참고 : 별칭 항이 없으면 "모형 x 모형" 섹션만 나타납니다.

보고서의 각 섹션에 대해 다음 항목이 제공됩니다.

평균 상관 보고서의 이 섹션에서 고려된 모든 항 쌍에 대한 상관의 평균입니다.

교락 수 교락된 항으로 구성된 항 쌍의 수입니다.

항 수 보고서의 이 섹션에서 고려된 항 쌍의 총 수입니다.

"절대 상관" 테이블 아래에 나오는 색상 그래디언트에 따라 이 테이블의 값에 색상이 적용됩니다. "절대 상관"의 빨간색 삼각형 메뉴 옵션을 사용하여 색상 범례를 제어할 수 있습니다. 자세한 내용은 "색상 대시보드"(455 페이지)에서 확인하십시오.

상관 색상 맵

"상관 색상 맵" 개요에는 각 설계에 대한 그림이 표시됩니다. 색상 맵의 셀은 맵 위에서 식별됩니다. "모형" 개요 또는 "별칭 항" 개요에 나타나는 효과에 해당하는 모든 항에 대한 셀이 있습니다. 각 셀은 두 항 간의 절대 상관 값에 따라 색상이 적용됩니다.

기본적으로 절대 상관 크기는 흰색에서 회색, 검은색까지의 강도 색상 테마를 사용하여 나타냅니다. 일반적으로 좋은 설계의 색상 맵에는 흰색 비대각이 많이 표시되어 직교성 또는 개별 항 간의 상관이 적음을 나타냅니다. 효과 간의 절대 상관이 크면 추정값 표준 오차가 증가합니다.

두 효과 간의 절대 상관 값을 보려면 커서로 해당 셀을 가리킵니다. 색상 척도를 변경하려면 "상관 색상 맵"의 빨간색 삼각형을 클릭하고 **파랑-회색-빨강**을 선택합니다. 사용자 정의 색상 척도를 사용하려면 그림을 마우스 오른쪽 버튼으로 클릭하고 **색상 테마**를 선택합니다. 상관계수표를 저장하려면 범례 아래의 그림 오른쪽을 마우스 오른쪽 버튼으로 클릭하고 **상관계수표**를 선택합니다.

절대 상관 및 상관 색상 맵의 예

그림 16.28에서는 "런 크기가 다른 설계"(444 페이지)에서 구성한 Plackett-Burman 설계와 확증적 선별 설계의 "절대 상관" 보고서를 보여 줍니다. "모형" 개요에 주효과만 포함되므로 "별칭 항" 개요에 모든 2 요인 교호작용이 포함됩니다. 모든 주효과 및 이원 교호작용이 색상 맵에 표시됩니다.

16 회 런 설계의 "상관 색상 맵"에서 비대각의 검은색 셀은 해당 항의 상관관계가 1 이므로 완전히 교락됨을 나타냅니다. 모형 항(주효과)이 별칭 항(2 요인 교호작용)과 교락되는 경우가 9 개 있고 별칭 항이 서로 교락되는 경우가 6 개 있습니다. 이는 보고서의 "쌍별 교락" 아래에 표시됩니다.

20 회 런 및 24 회 런 설계의 색상 맵에는 단색의 검은색 비대각 셀이 없습니다. 따라서 이러한 설계는 주효과 또는 이원 교호작용 효과 쌍 간의 교락이 없습니다. 그러나 20 회 런 설계와 24 회 런 설계 모두 16 회 런 설계보다 "모형 x 별칭" 항의 "평균 상관"이 높다는 점에 주목하십시오. 16 회 런 설계는 교락을 나타내지만 평균 상관관계가 20 회 런 및 24 회 런 설계보다 낮습니다.

설계 진단

"설계 진단" 개요에서는 비교 설계를 기준으로 참조 설계의 *D* 효율도, *G* 효율도, *A* 효율도 및 *I* 효율도를 보여 줍니다. "추가 런 크기"도 표시됩니다. 두 가지 설계가 주어진 경우 상대 효율도 측도가 높은 설계가 더 좋습니다.

그림 16.29 세 가지 설계에 대한 설계 진단

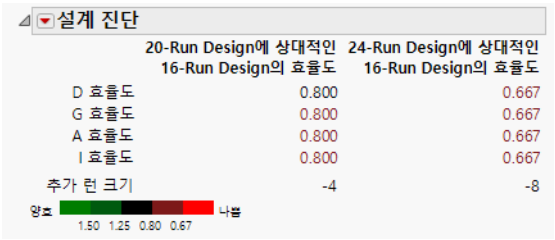


그림 16.29에서는 "런 크기가 다른 설계"(444 페이지)에서 구성한 세 가지 설계의 "설계 진단" 보고서를 보여 줍니다. "모형" 개요에는 주효과만 있습니다.

"설계 진단" 테이블 아래에 나오는 색상 그래디언트에 따라 이 테이블의 값에 색상이 적용됩니다. "설계 진단"의 빨간색 삼각형 메뉴 옵션을 사용하여 색상 범례를 제어할 수 있습니다. 자세한 내용은 "색상 대시보드"(455 페이지)에서 확인하십시오.

효율도 및 추가 런 크기

D 효율도, *G* 효율도, *A* 효율도 및 *I* 효율도 각각에 대한 상대 효율도가 "설계 진단" 보고서에 표시됩니다. 이 값은 각 설계의 효율도 값을 계산한 후 적절한 비율을 적용하여 구할 수 있습니다. 상대 효율도 측도에 대한 설명은 "상대 효율도 측도"(467 페이지)에서 확인하십시오.

추가 런 크기는 참조 설계의 런 수에서 비교 설계의 런 수를 뺀 값입니다. 참조 설계의 런 수가 비교 설계보다 많은 경우 참조 설계의 효율도를 달성하기 위해 필요한 추가 런 수가 "추가 런 크기"에 표시됩니다.

상대 효율도 측도

표기

- $\mathbf{X}$ 는 모형 행렬입니다 .
- p 는 모형의 항 수입니다 (절편 포함).
- $Var(\hat{y}|\mathbf{x})$ 는 점 $\mathbf{x}$ 에서의 상대 예측 분산입니다 . 자세한 내용은 " 기술 상세 정보 " 부록의 " 상대 예측 분산 "(770 페이지) 에서 확인하십시오 .

상대 효율도

비교 설계 (*Comp*) 에 대한 참조 설계 (*Ref*) 의 상대 효율도는 다음 표현식을 사용하여 구합니다 .

D 효율도 Eff_{Ref}/Eff_{Comp}. 여기서 각 설계의 *Eff* 는 다음과 같이 지정됩니다 .

$$Eff = |\mathbf{X}'\mathbf{X}|^{1/p}$$

G 효율도 Eff_{Comp}/Eff_{Ref}. 여기서 각 설계의 *Eff* 는 다음과 같이 지정됩니다 .

$$Eff = Var(\hat{y}|\underset{\sim}{\mathbf{x}})_{max} = \underset{\underset{\sim}{\mathbf{x}} \text{ in D}}{maximum}[\underset{\sim}{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\underset{\sim}{\mathbf{x}}]$$

D 가 설계 영역이라고 하면

참고 : G 효율도는 설계 공간의 몬테카를로 표집을 사용하여 계산됩니다 . 보고된 값은 더 큰 $Var(\hat{y}|\underset{\sim}{\mathbf{x}})_{max}$ 또는 몬테카를로 표집의 예측 분산을 기반으로 합니다 . 따라서 같은 설계에 대한 계산이 약간 다를 수 있습니다 .

A 효율도 Eff_{Comp}/Eff_{Ref}. 여기서 각 설계의 Eff 는 다음과 같이 지정됩니다 .

$$Eff = \text{Trace}[(\mathbf{X}'\mathbf{X})^{-1}]$$

I 효율도 Eff_{Comp}/Eff_{Ref}. 여기서 각 설계의 Eff 는 다음과 같이 지정됩니다 .

$$Eff = \frac{\int \underset{\sim}{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\underset{\sim}{\mathbf{x}}d\mathbf{x}}{\int d\mathbf{x}}$$

계산에 대한 자세한 내용은 Goos and Jones(2011) 의 "Section 4.3.5" 에서 확인하십시오 .

설계 비교 옵션

고급 옵션 > 분할구 분산비 오차 분산에 대한 랜덤 주구 분산 및 하위구 분산 (있는 경우) 의 비를 지정합니다 . 이 값을 설정하려면 먼저 분할구 설계를 위한 변경하기 힘든 요인 또는 분할 - 분할구 설계를 위한 변경하기 힘든 요인과 매우 변경하기 힘든 요인을 정의해야 합니다 . 그런 다음 분할구 설계를 지정했는지 아니면 분할 - 분할구 설계를 지정했는지에 따라 분산 비율에 하나 또는 두 개의 양수를 입력할 수 있습니다 .

17 장

예상 표본 크기 및 검정력

표본 크기 및 검정력 플랫폼을 사용하면 실험이나 연구를 계획할 때 특정 질문에 답하는 능력에 표본 크기가 미치는 영향을 조사할 수 있습니다.

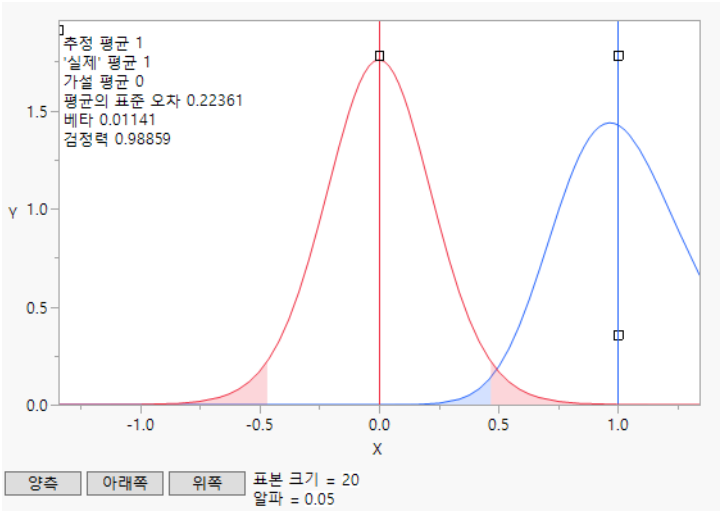
- 테스트할 단위 수는 몇 개입니까?
- 처리 평균의 차이를 감지할 수 있습니까?
- 고장 시간을 추정하기 위해 테스트해야 할 단위 수는 몇 개입니까?

표본 크기 및 검정력 플랫폼은 계산기 모음입니다. 계산기에는 데이터 입력 및 계산된 값을 위한 필드가 있습니다. 지정할 값과 계산할 값을 선택할 수 있습니다. 예를 들어 검정력 및 감지할 효과 크기를 지정한 후 필요한 표본 크기를 계산할 수 있습니다. 아니면 표본 크기와 검정력을 지정하여 감지할 수 있는 효과 크기를 계산할 수도 있습니다.

가설 검정을 위한 표본 크기 계산은 제1종 오류와 제2종 오류 간의 균형을 기반으로 합니다. 제1종 오류(α)는 귀무가설이 참일 때 귀무가설을 기각할 확률입니다. 제2종 오류(β)는 귀무가설이 거짓일 때 귀무가설을 채택할 확률입니다. 일반적으로 α와 β를 작게 하려고 합니다. 검정력(1 - β)은 귀무가설이 거짓일 때 기각할 확률입니다.

JMP의 검정력과 표본 크기에 대한 자세한 내용은 Barker(2011)에서 확인하십시오.

그림 17.1 1 표본 평균에 대한 검정력 애니메이션



목차

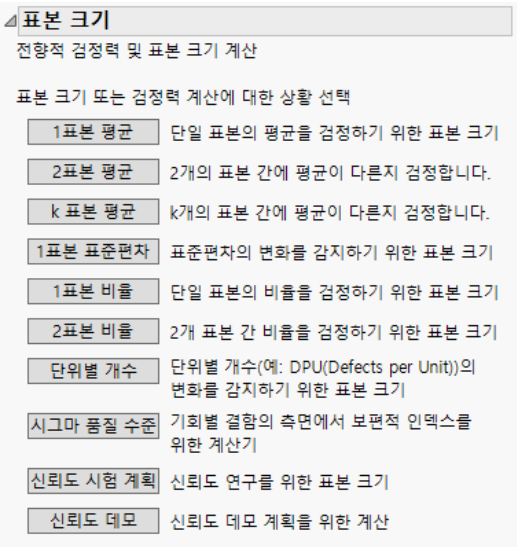
표본 크기 및 검정력 플랫폼 시작 472
1 표본 평균 계산기 473
1표본 평균 계산기의 예 473
1표본 평균 계산기 필드 477
1표본 평균 계산기에 대한 통계 상세 정보 478
2 표본 평균 계산기 479
2표본 평균 계산기의 예 479
2표본 평균 계산기 필드 480
2표본 평균 계산기에 대한 통계 상세 정보 481
k 표본 평균 계산기 481
k 표본 평균 계산기의 예 482
k 표본 평균 계산기 필드 484
k 표본 평균 계산기에 대한 통계 상세 정보 485
1 표본 표준편차 계산기 486
1표본 표준편차 계산기의 예 486
1표본 표준편차 계산기 필드 487
1표본 표준편차 계산기에 대한 통계 상세 정보 488
1 표본 비율 계산기 489
1표본 비율 계산기의 예 489
1표본 비율 계산기 필드 490
단일 비율 계산기에 대한 통계 상세 정보 491
2 표본 비율 계산기 492
2표본 비율 계산기의 예 492
2표본 비율 계산기 필드 493
두 비율 계산기에 대한 통계 상세 정보 494
단위별 개수 계산기 494
단위별 개수 계산기의 예 495
단위별 개수 계산기 필드 496
단위별 개수 계산기에 대한 통계 상세 정보 496
시그마 품질 수준 계산기 497
시그마 품질 수준 계산기의 예 497
결함 수 계산의 예 498
시그마 품질 수준 계산기 필드 498
시그마 품질 계산기에 대한 통계 상세 정보 499
신뢰도 시험 계획 계산기 499
신뢰도 시험 계획 계산기의 예 499
신뢰도 시험 계획 계산기 필드 501
신뢰도 시험 계획 계산기에 대한 통계 상세 정보 502

신뢰도 데모 계산기 503
신뢰도 데모 계산기의 예 503
신뢰도 데모 계산기 필드 504
신뢰도 데모 계산기에 대한 통계 상세 정보 505

표본 크기 및 검정력 플랫폼 시작

표본 크기 및 검정력 플랫폼은 실험이나 연구를 계획하기 위한 계산기 모음입니다 . **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택하여 표본 크기 및 검정력 플랫폼을 시작합니다 .

그림 17.2 표본 크기 및 검정력 시작 창



다음 계산기 중에서 선택합니다 .

- "1 표본 평균 계산기 "(473 페이지)
- "2 표본 평균 계산기 "(479 페이지)
- "k 표본 평균 계산기 "(481 페이지)
- "1 표본 표준편차 계산기 "(486 페이지)
- "1 표본 비율 계산기 "(489 페이지)
- "2 표본 비율 계산기 "(492 페이지)
- " 단위별 개수 계산기 "(494 페이지)
- " 시그마 품질 수준 계산기 "(497 페이지)
- " 신뢰도 시험 계획 계산기 "(499 페이지)
- " 신뢰도 데모 계산기 "(503 페이지)

1 표본 평균 계산기

단일 평균에 대한 가설 검정을 위해 표본 크기를 계산하려면 "1 표본 평균 " 계산기를 사용합니다 . 표본 크기 , 검정력 , 유의성과 감지할 가설 차이 간의 균형을 연구해야 합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

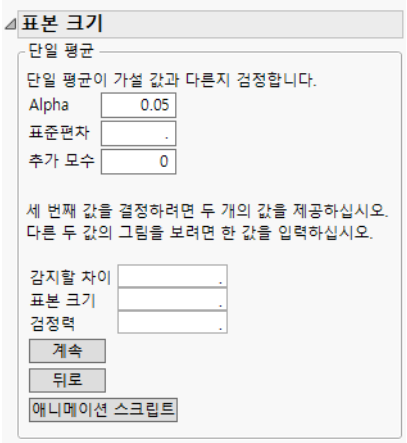
H_0: $\mu = \mu_0$

다음 양측 대립가설과 비교

H_a: $\mu \neq \mu_0$

여기서 μ는 실제 평균이고 μ_0은 귀무가설 평균 또는 참조 값입니다. 감지할 차이는 표본 집합을 기반으로 감지해야 한다고 간주되는 크기 (δ)로, μ_0과의 차이입니다. 동일한 유의 수준과 검정력이 주어진 경우 큰 차이를 감지하는 것보다 작은 차이를 감지하기 위해 더 큰 표본 크기가 필요합니다. 관심 모집단은 평균 μ와 표준편차 σ의 정규 분포를 따른다고 가정합니다.

그림 17.3 단일 평균에 대한 초기 표본 크기 및 검정력 계산기



1 표본 평균 계산기의 예

다음 예에서는 회사에서 개발 중인 새 직물의 가연성이 현재 소재에 비해 성능이 개선되었음을 입증하려고 합니다 . 이전 테스트에 따르면 이 직물의 연소 시간에 대한 표준편차는 2 초입니다 .

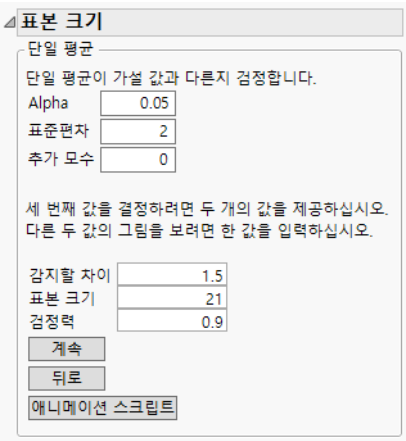
표본 크기 계산의 예

이 초기 예에서는 유의 수준 $\alpha = 0.05$ 에서 1.5 초 차이를 감지할 수 있는 검정력이 90% 인 실험을 설계하려고 합니다 . "1 표본 평균 " 계산기를 사용하여 테스트할 직물 표본 수를 계산하십시오 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .

2. **1 표본 평균** 버튼을 클릭합니다 .
3. **알파** 설정값 0.05 를 그대로 둡니다 .
4. **표준편차**에 2 를 입력합니다 .
5. **추가 모수** 설정값 0 을 그대로 둡니다 .
6. **감지할 차이**에 1.5 를 입력합니다 .
7. **표본 크기**는 비워 둡니다 .
8. **검정력**에 0.9 를 입력합니다 .
9. **계속**을 클릭합니다 .

그림 17.4 1 표본 평균 계산기



유의 수준이 0.05 일 때 연소 시간에서 1.5 초의 유의차를 감지할 확률이 90% 가 되려면 21 개의 직물 표본이 필요합니다.

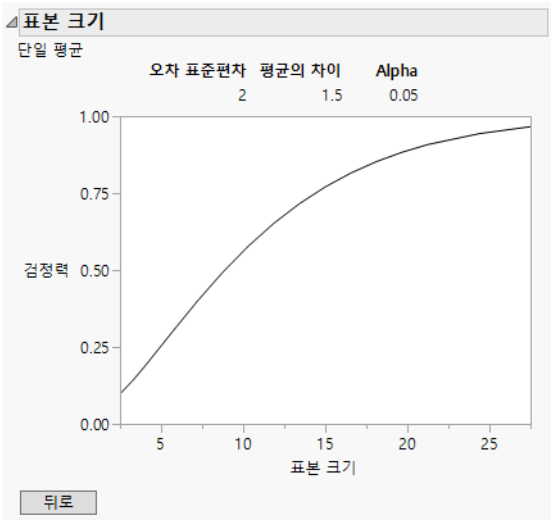
검정력과 표본 크기 비교 그림의 예

직물 실험에서 표본 크기와 검정력 간의 균형을 탐색하려면 검정력과 표본 크기 비교 그림을 사용합니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **1 표본 평균**을 클릭합니다 .
3. **알파** 설정값 0.05 를 그대로 둡니다 .
4. **표준편차**에 2 를 입력합니다 .
5. **추가 모수** 설정값 0 을 그대로 둡니다 .
6. **감지할 차이**에 1.5 를 입력합니다 .
7. **표본 크기**는 비워 둡니다 .

8. **검정력**은 비워 둡니다 .
9. **계속**을 클릭하여 표본 크기별 검정력 그림을 시작합니다 .

그림 17.5 표본 크기별 검정력



이 그림은 약 0.1 ~ 0.95 사이에서 변화하는 검정력에 대한 표본 크기 범위를 보여 줍니다. 실험의 테스트 횟수를 약 15 회로 줄이고 검정력이 75% 를 초과하도록 유지할 수 있습니다. 그러나 테스트를 10 회만 실행하면 유의차 1.5 를 감지할 수 있는 검정력이 약 50% 로 떨어집니다.

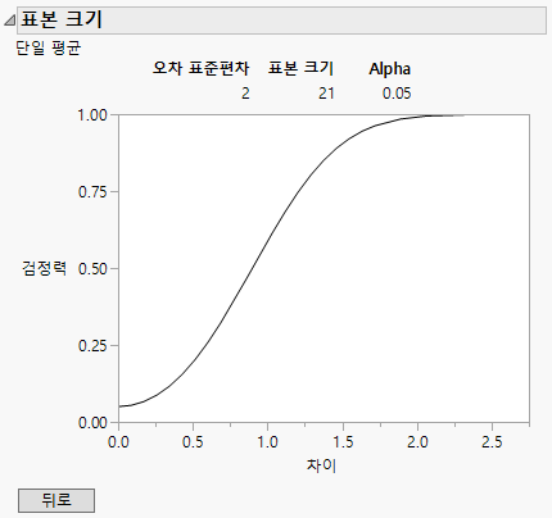
팁 : 그림에서 표본 크기와 검정력 조합을 얻으려면 십자도구 기호를 사용하십시오 .

검정력과 차이 비교 그림의 예

직물 실험에서 21 개 관측값으로 감지할 수 있는 차이 크기와 검정력 간의 균형을 탐색하려면 검정력과 차이 비교 그림을 사용합니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **1 표본 평균**을 클릭합니다 .
3. **알파** 설정값 0.05 를 그대로 둡니다 .
4. **표준편차**에 2 를 입력합니다 .
5. **추가 모수** 설정값 0 을 그대로 둡니다 .
6. **감지할 차이**는 비워 둡니다 .
7. **표본 크기**에 21 을 입력합니다 .
8. **검정력**은 비워 둡니다 .
9. **계속**을 클릭합니다 .

그림 17.6 표본 크기 21 에 대해 감지할 차이별 검정력 그림



유의 수준이 0.05이고 관측값 수가 21일 때 90%의 검정력으로 1.5초 차이를 감지할 수 있습니다. 차이가 1초로 작아지면 21개 직물 표본을 사용하여 차이를 감지할 수 있는 검정력은 약 50%입니다.

애니메이션 스크립트의 예

표본 크기를 변경할 때 검정력에 어떤 영향을 주는지 살펴보려면 애니메이션 스크립트를 사용합니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **1 표본 평균**을 클릭합니다 .
3. **알파** 설정값 0.05 를 그대로 둡니다 .

   팁 : 애니메이션이 시작된 후 애니메이션 창에서 " 알파 " 값을 클릭하여 유의 수준을 변경할 수 있습니다 .

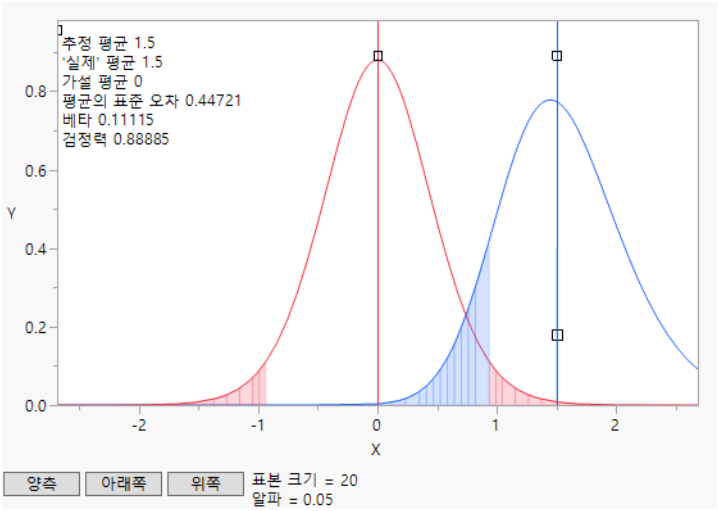
4. **표준편차**에 2 를 입력합니다 .
5. **추가 모수** 설정값 0 을 그대로 둡니다 .
6. **감지할 차이**에 1.5 를 입력합니다 .
7. **표본 크기**는 비워 둡니다 .

   팁 : 표본 크기를 비워 두면 기본 표본 크기가 20 으로 설정됩니다 . 애니메이션 스크립트가 시작된 후 기본값 20 이 표시됩니다 . 애니메이션이 시작되기 전이나 후에 표본 크기를 변경할 수 있습니다 . 애니메이션이 시작된 후 변경하려면 애니메이션 창에서 " 표본 크기 " 값을 클릭하십시오 .

8. **검정력**은 비워 둡니다 .

9. **애니메이션 스크립트**를 클릭합니다 .

그림 17.7 검정력을 보여 주는 초기 애니메이션 스크립트



초기 애니메이션 그림에는 다음과 같은 두 개의 *t* 밀도 곡선이 표시됩니다 .

- 빨간색 곡선은 실제 평균이 0 일 때 *t* 분포를 표시합니다 .
- 파란색 곡선은 실제 평균이 1.5, 즉 감지할 차이일 때 *t* 분포를 표시합니다 .
- 파란색 음영은 제 2 종 오류가 발생할 확률을 나타냅니다 . 제 2 종 오류는 차이가 있는데 차이를 감지하지 못할 확률입니다 . 제 2 종 오류의 확률은 흔히 β 로 표시됩니다 .
- 빨간색 음영은 제 1 종 오류가 발생할 확률을 나타냅니다 . 제 1 종 오류는 평균의 차이가 없는데 차이가 유의하다는 결론을 내릴 확률입니다 . 제 1 종 오류의 확률은 흔히 α 로 표시됩니다 .

사각형 핸들을 선택하여 드래그하면 곡선 위치를 기준으로 통계량의 변화를 볼 수 있습니다 . 표본 크기 및 알파 값을 변경하려면 그림 아래의 해당 값을 클릭합니다 .

기본적으로 애니메이션은 양측 검정을 보여 줍니다. **양측**, **아래쪽** 및 **위쪽** 버튼을 사용하여 대립 가설을 같지 않음, 보다 작음 또는 보다 큼으로 전환할 수 있습니다.

1 표본 평균 계산기 필드

다음 통계량을 지정합니다 .

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

팁 : 단측 검정의 경우 $\alpha = \alpha^*2$ 를 사용합니다 . $\alpha = 0.05$ 인 단측 검정의 경우 $\alpha = 0.10$ 인 양측 계산기를 사용합니다 . 결과 값은 $\alpha = 0.05$ 인 단측 검정에 필요한 값입니다 .

표준편차 가정된 표준편차입니다 . 오차 표준편차의 추정값은 이전 모형 적합에서 구한 RMSE(제곱근 평균 제곱 오차) 일 수 있습니다 .

팁 : 표준편차 단위로 측정된 차이를 감지하는 데 필요한 표본 크기를 추정하려면 표준편차 1 을 사용하십시오 .

추가 모수 가설 검정에서 μ 이외의 모수 수입니다 . 이 옵션은 다요인 설계에 사용할 수 있습니다 . 간단한 사례의 경우 이 필드의 기본값 0 을 그대로 유지합니다 .

효과가 직교인 다요인 설계의 경우 여기에 추가 모형 모수의 수를 지정할 수 있습니다 . 예를 들어 3 개의 2 요인 교호작용이 모두 포함된 3 요인 2 수준 설계의 경우 추가 모수의 수는 다른 주효과에 대한 모수 2 개와 교호작용에 대한 모수 3 개를 합해 모두 5 개입니다 .

다음 값 중 두 개를 지정하여 세 번째 값을 계산하거나 , 한 값을 지정하여 다른 두 값 사이의 관계 그림을 볼 수 있습니다 .

감지할 차이 실제 평균과 가설 평균 또는 참조 평균 사이에 통계적으로 유의하다고 선언할 수 있는 최소 차이입니다 .

표본 크기 실험의 총 관측값 (런 , 실험 단위 또는 표본) 수입니다 .

검정력 귀무가설이 거짓일 때 기각할 확률입니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

계산기 버튼

계속 두 개의 모수가 지정된 경우 결측값을 계산하거나 , 모수가 하나만 지정된 경우 두 개의 결측 모수를 비교하는 그림을 시작합니다 .

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다 .

애니메이션 스크립트 검정력과 감지할 차이의 관계를 설명하고 탐색하기 위한 대화식 그림을 시작합니다 . 자세한 내용은 " 애니메이션 스크립트의 예 "(476 페이지) 에서 확인하십시오 .

1 표본 평균 계산기에 대한 통계 상세 정보

1 표본 평균 계산은 다음 가설 검정을 위한 기존 전체 *F*- 검정과 축소 F- 검정의 비교를 기반으로 합니다 .

H_0: μ = μ_0 대 H_a: μ ≠ μ_0

JMP 는 다음과 같이 검정력을 계산합니다 .

$$\Pr(\text{reject } H_0 \mid \mu = \mu_0 + \delta) = 1 - F\left(f_{1-\alpha}, 1, n-p-1, \frac{n\delta^2}{\sigma^2}\right)$$

다음은 각 요소에 대한 설명입니다 .

α 는 유의 수준입니다 .

n 은 표본 크기입니다 .

p 는 추가 모수의 수입니다 .

δ 는 감지할 차이입니다 .

$f_{1-\alpha}$ 는 $F(1, n- p -1)$ 분포의 $(1 - \alpha)^{th}$ 분위수입니다 .

$F(x, df_1, df_2, nc)$ 는 자유도가 df_1 및 df_2 이고 , x 에서 계산된 비중심성 모수 nc 가 있는 비중심 F 분포의 누적 분포 함수입니다 .

δ 및 n에 대한 분석 해가 존재하지 않으므로 수치 해를 사용하여 계산합니다.

JMP의 계산에 대한 자세한 내용은 Barker(2011, Section 2.1)에서 확인하십시오.

2 표본 평균 계산기

두 평균에 대한 가설 검정을 위해 표본 크기를 계산하려면 "2 표본 평균 " 계산기를 사용합니다 . 표본 크기 , 유의 수준 , 원하는 검정력이 주어지고 데이터 변동이 가정된 경우 감지 가능한 처리 효과의 크기와 같은 질문에 대한 답을 구할 수 있습니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

$$H_0: \mu_1 - \mu_2 = 0$$

다음 양측 대립가설과 비교

$$H_a: \mu_1 - \mu_2 \neq 0$$

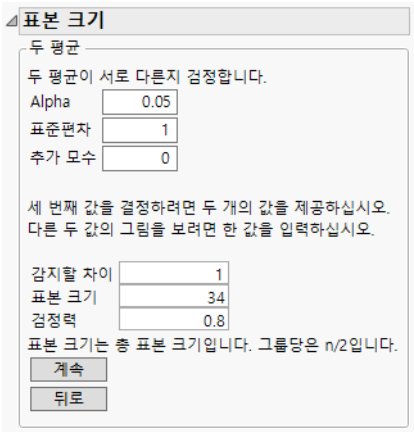
여기서 μ_1과 μ_2는 두 모집단의 실제 평균입니다. 관심 모집단은 정규 분포를 따르고, 평균 간의 차이(δ)를 감지하려고 한다고 가정합니다.

2 표본 평균 계산기의 예

두 그룹을 비교 중이고 0.05 유의 수준에서 80% 검정력으로 그룹 평균 간의 1 표준편차 차이를 감지하려는 경우를 가정해 보겠습니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **2 표본 평균**을 클릭합니다 .
3. **알파** 값 0.05 를 그대로 둡니다 .
4. **표준편차**에 1 을 입력합니다 .
5. **추가 모수** 값 0 을 그대로 둡니다 .
6. **감지할 차이**에 1 을 입력합니다 .
7. **표본 크기**는 비워 둡니다 .
8. **검정력**에 0.8 을 입력합니다 .
9. **계속**을 클릭합니다 .

그림 17.8 2 표본 평균 계산기 결과



표본 크기는 34로 계산됩니다. 각 그룹에서 17개의 표본을 사용하면 $\alpha = 0.05$ 유의 수준에서 두 표본 평균 간의 1 표준편차 차이를 감지할 확률이 80%입니다.

2 표본 평균 계산기 필드

다음 통계량을 지정합니다.

알파 제 1 종 오류의 확률, 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다. 일반적으로 검정의 유의 수준이라고 합니다. 기본 유의 수준은 0.05 입니다.

표준편차 두 그룹의 표준편차가 동일하다는 전제하에 가정된 표준편차입니다.

추가 모수 가설 검정에서 μ_1 과 μ_2 이외의 모수 수입니다. 이 옵션은 다요인 설계에 사용할 수 있습니다. 간단한 사례의 경우 이 필드의 기본값 0 을 그대로 유지합니다.

다요인 균형 설계의 경우 여기에 추가 모형 모수의 수를 지정할 수 있습니다. 예를 들어 3 개의 2 요인 교호작용이 모두 포함된 3 요인 2 수준 설계의 경우 추가 모수의 수는 다른 주효과에 대한 모수 1 개와 교호작용에 대한 모수 3 개를 합해 모두 4 개입니다.

다음 모수 중 두 개를 지정하여 세 번째 값을 계산하거나, 한 값을 지정하여 다른 두 모수 사이의 관계 그림을 볼 수 있습니다.

감지할 차이 두 평균 사이에 통계적으로 유의하다고 선언할 수 있는 최소 차이입니다.

표본 크기 실험의 총 관측값 (런, 실험 단위 또는 표본) 수입니다. 각 그룹에서 이 표본 크기의 1/2 씩 사용합니다.

검정력 귀무가설이 거짓일 때 기각할 확률입니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다.

계산기 버튼

계속 두 개의 모수가 지정된 경우 결측값을 계산하거나, 모수가 하나만 지정된 경우 두 개의 결측 모수를 비교하는 그림을 시작합니다.

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다.

2 표본 평균 계산기에 대한 통계 상세 정보

2 표본 평균 계산은 다음 가설 검정을 위한 기존 전체 *F*- 검정과 축소 F- 검정의 비교를 기반으로 합니다.

H_0: $\mu_1 - \mu_2 = 0$ 대 H_a: $\mu_1 - \mu_2 \neq 0$

JMP 는 다음과 같이 검정력을 계산합니다.

$$\text{검정력} = \Pr(\text{reject } H_0 \mid (\mu_1 - \mu_2 = \delta)) = 1 - F\left(f_{1-\alpha}, 1, 2n - p - 2, \frac{n\delta^2}{2\sigma^2}\right)$$

다음은 각 요소에 대한 설명입니다.

α 는 유의 수준입니다.

n 은 그룹당 표본 크기입니다.

p 는 추가 모수의 수입니다.

δ 는 감지할 차이입니다.

$f_{1-\alpha}$ 는 $F(1, 2n - p - 2)$ 분포의 $(1 - \alpha)^{\text{th}}$ 분위수입니다.

$F(x, \text{df}_1, \text{df}_2, nc)$ 는 자유도가 df_1 및 df_2 이고, x 에서 계산된 비중심성 모수 nc 가 있는 비중심 F 분포의 누적 분포 함수입니다.

δ 및 n 에 대한 분석 해가 존재하지 않으므로 수치 해를 사용하여 계산합니다.

JMP 의 계산에 대한 자세한 내용은 Barker(2011, Section 2.2) 에서 확인하십시오.

k 표본 평균 계산기

약 2 ~ 10 개 평균에 대한 가설 검정에 적합한 표본 크기를 결정하려면 k 표본 평균 계산기를 사용합니다. 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다.

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

다음 양측 대립가설과 비교

H_a: 모든 평균이 동일하지는 않음

다음은 각 요소에 대한 설명입니다 .

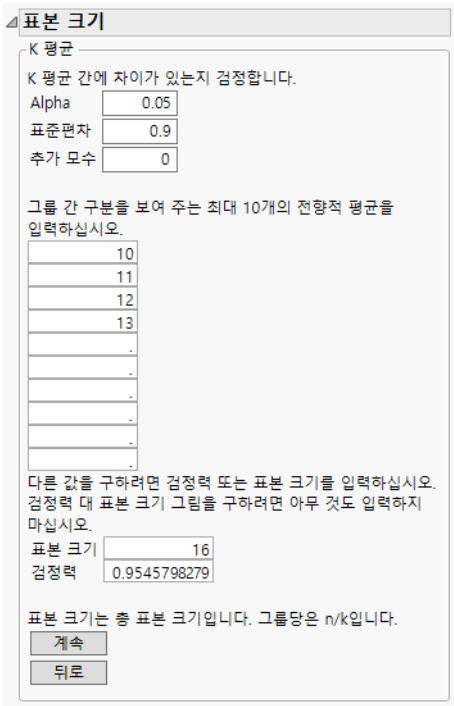
$$X_{ij} \sim N(\mu_j, \sigma^2) \text{ for } i = 1, \ldots, n, \; j = 1, \ldots, k$$

k 표본 평균 계산기의 예

4 개의 각 그룹에 관측값이 4 개씩 있고 모든 그룹 평균이 동일한지 여부를 테스트하려는 경우를 가정해 보겠습니다 . 표준편차가 0.9 일 때 그룹 평균은 10, 11, 12, 13 일 것으로 예상합니다 . 0.05 유의 수준에서 하나 이상의 평균에 대한 차이를 감지할 검정력을 계산합니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **k 표본 평균**을 클릭합니다 .
3. **알파** 값 0.05 를 그대로 둡니다 .
4. **표준편차**에 0.9 를 입력합니다 .
5. **추가 모수** 값 0 을 그대로 둡니다 .
6. 예상 평균의 4 개 수준에 10, 11, 12, 13 을 입력합니다 .
7. **표본 크기**에 16 을 입력합니다 .
8. **검정력**은 비워 둡니다 .
9. **계속**을 클릭합니다 .

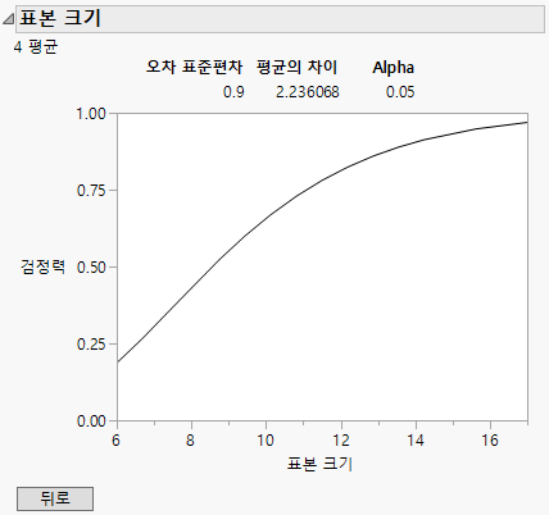
그림 17.9 4 개 평균에 대한 예상 검정력



검정력은 약 0.95 로 계산됩니다 . 즉 , 모평균이 10, 11, 12, 13 이고 표준편차는 0.9, 총 표본 크기는 16 으로 가정할 때 0.05 유의 수준에서 하나 이상의 평균이 다르다는 것을 감지할 확률이 95% 입니다 .

10. " 표본 크기 " 16 을 클릭하고 삭제합니다 .
11. " 검정력 " 0.95 를 클릭하고 삭제합니다 .
12. **계속**을 클릭하여 검정력과 표본 크기 비교 그림을 시작합니다 .

그림 17.10 K 평균에 대한 예상 검정력 및 표본 크기별 검정력 그림



검정력과 표본 크기 비교 그림에서 표본 크기 16이 허용 가능한지 확인할 수 있습니다. "표본 크기" 축을 커서로 가리키고 드래그하여 축을 16 이상으로 확장합니다. 표본 크기가 16을 초과할 경우 검정력이 약간 증가하는 것을 확인할 수 있습니다.

또한 평균의 차이도 그림에 보고되며 이 값은 총 평균으로부터의 편차에 대한 제곱합의 제곱근으로 계산됩니다. 이 경우에는 $(-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2$의 제곱근, 즉 5의 제곱근입니다. 따라서 이 예에서 평균의 차이는 약 2.236입니다.

k 표본 평균 계산기 필드

다음 통계량을 지정합니다.

알파 제 1 종 오류의 확률, 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다. 일반적으로 검정의 유의 수준이라고 합니다. 기본 유의 수준은 0.05 입니다.

표준편차 모든 그룹의 표준편차가 동일하다는 전제하에 가정된 표준편차입니다.

추가 모수 가설 검정에서 μ 이외의 모수 수입니다. 이 옵션은 다요인 설계에 사용할 수 있습니다. 간단한 사례의 경우 이 필드의 기본값 0 을 그대로 유지합니다.

예상 평균 평균에 대해 가정된 값입니다. 두 개 이상의 평균을 지정해야 합니다.

다음 값 중 하나를 지정하여 두 번째 값을 계산하거나, 두 값을 모두 비워 두어 두 값 사이의 관계 그림을 볼 수 있습니다.

표본 크기 실험의 총 관측값 (런, 실험 단위 또는 표본) 수입니다. 각 그룹에서 n/k 표본 크기를 사용합니다. 여기서 n 은 총 관측값 수입니다.

검정력 귀무가설이 거짓일 때 기각할 확률입니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다.

계산기 버튼

계속 하나의 모수가 지정된 경우 결측값을 계산하거나 , 모수가 지정되지 않은 경우 두 모수를 비교하는 그림을 시작합니다 .

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다 .

k 표본 평균 계산기에 대한 통계 상세 정보

k 표본 평균 계산은 다음을 기반으로 합니다 .

$$\text{검정력} = 1 - F\left(f_{1-\alpha}, 1k-1, 2kn-p-2k, \frac{n\sum_{j=1}^{k}(\mu_j-\bar{\mu})^2}{\sigma^2}\right)$$

다음은 각 요소에 대한 설명입니다 .

α 는 유의 수준입니다 .

σ 는 k 그룹 사이에서 공통 표준편차입니다 .

n 은 그룹당 표본 크기입니다 .

p 는 추가 모수의 수입니다 .

$\bar{\mu}$ 는 *k* 그룹 예상 평균의 평균입니다 .

$f_{1-\alpha}$ 는 $F(k - 1, kn - p - k)$ 분포의 $(1 - \alpha)^{\text{th}}$ 분위수입니다 .

$F(x, df_1, df_2, nc)$ 는 자유도가 df_1 및 df_2 이고, x 에서 계산된 비중심성 모수 nc 가 있는 비중심 *F* 분포의 누적 분포 함수입니다 .

수치 해를 사용하여 *n* 을 계산합니다.

평균의 차이

평균의 차이는 다음과 같이 계산됩니다 .

$$\sqrt{\sum_{j=1}^{k}(\mu_j-\bar{\mu})^2}$$

JMP 의 계산에 대한 자세한 내용은 Barker(2011, Section 2.3) 에서 확인하십시오 .

1 표본 표준편차 계산기

단일 표준편차에 대한 가설 검정을 위해 표본 크기를 계산하려면 "1 표본 표준편차 " 계산기를 사용합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

H_0: $\sigma = \sigma_0$

다음 단측 대립가설과 비교

H_a: $\sigma > \sigma_0$ 또는 H_a: $\sigma < \sigma_0$

여기서 σ는 모표준편차이고 σ_0는 가설 값입니다. 감지할 차이는 표본 집합을 기반으로 감지해야 한다고 간주되는 크기(δ)로, σ_0과의 차이입니다.

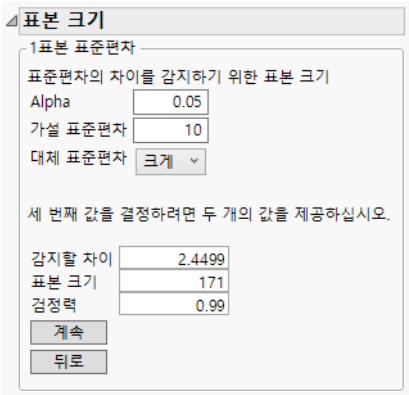
1 표본 표준편차 계산기의 예

참고 : 이 예는 NIST(미국 국립표준기술연구소) 의 온라인 설명서에서 발췌했습니다 . https://www.itl.nist.gov/div898/handbook/prc/section2/prc232.htm 에서 NIST 설명서 예제에 액세스할 수 있습니다 .

실리콘 웨이퍼 로트의 저항력 측정에 대한 분산은 100 ohm-cm2 여야 합니다 . 구매자는 분산이 155 ohm-cm2 보다 크면 로트를 수락하지 않으려고 합니다 . 목표값 100 ohm-cm2 에서 55 ohm-cm2 증가를 감지하기 위해 로트 분산을 정확하게 추정하려면 몇 개의 웨이퍼를 테스트해야 합니까 ? 표준편차 측면에서 가설 표준편차 σ_0 은 10(100 의 제곱근) 이고 σ 는 12.4499(100 + 55 = 155 의 제곱근) 입니다 . 감지할 차이는 12.4499 – 10 = 2.4499 입니다 . 원하는 검정력은 0.99 이고 유의 수준은 0.05 입니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **1 표본 표준편차**를 클릭합니다 .
3. **알파** 값 0.05 를 그대로 둡니다 .
4. **가설 표준편차**에 10 을 입력합니다 .
5. **대립가설 표준편차**에서 " 크게 " 를 선택합니다 .
6. **감지할 차이**에 2.4499 를 입력합니다 .
7. **표본 크기**는 비워 둡니다 .
8. **검정력**에 0.99 를 입력합니다 .
9. **계속**을 클릭합니다 .

그림 17.11 1 표본 표준편차 계산기



유의 수준이 0.05 일 때 표준편차 10 에서 2.4499 증가를 감지할 확률이 99% 가 되려면 171 개의 웨이퍼를 테스트해야 합니다.

1 표본 표준편차 계산기 필드

다음을 지정합니다.

알파 제 1 종 오류의 확률, 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다. 일반적으로 검정의 유의 수준이라고 합니다. 기본 유의 수준은 0.05 입니다.

가설 표준편차 표본 표준편차를 비교할 가설 또는 기준 표준편차입니다.

대립가설 표준편차 감지할 변화의 방향이며 " 크게 " 또는 " 작게 " 를 선택합니다.

> **참고 :** " 대립가설 표준편차 " 메뉴에서 " 작게 " 를 선택한 경우 " 감지할 차이 " 에 음수를 입력하십시오.

다음 값 중 두 개를 지정하여 세 번째 값을 계산합니다.

감지할 차이 감지 가능한 최소 차이입니다 (통계적으로 유의하다고 선언할 수 있는 최소 차이). 이 값은 가설 값과 실제 값의 차이입니다.

표본 크기 실험의 총 관측값 (런, 실험 단위 또는 표본) 수입니다.

검정력 귀무가설이 거짓일 때 기각할 확률입니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다.

계산기 버튼

계속 결측값을 계산합니다.

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다.

1 표본 표준편차 계산기에 대한 통계 상세 정보

JMP에서 검정력을 계산하는 데 사용하는 계산식은 "대립가설 표준편차" 옵션의 설정에 따라 달라집니다.

대립가설 표준편차가 "크게" 로 설정된 경우 JMP 는 다음과 같이 검정력을 계산합니다.

$$\text{검정력} = 1 - F\left(\frac{\sigma_0^2 \cdot \chi_{1-\alpha}}{(\sigma_0 + \delta)^2}, n-1\right)$$

$$\delta = \sigma_0\sqrt{\frac{\chi_{1-\alpha}}{\chi_\beta}} - \sigma_0$$

다음은 각 요소에 대한 설명입니다.

σ_0 은 가설 표준편차입니다.
σ 는 실제 표준편차입니다.
X_p 는 $X_{n\text{-}1}{}^2$ 분포의 $(1 - p)^{th}$ 분위수입니다.
$F(x, n\text{-}1)$ 는 x 에서 계산된 $X_{n\text{-}1}{}^2$ 분포의 누적 분포 함수입니다.

대립가설 표준편차가 "작게" 로 설정된 경우 JMP 는 다음과 같이 검정력을 계산합니다.

$$\text{검정력} = 1 - F\left(\frac{\sigma_0^2 \cdot \chi_{\alpha}}{(\sigma_0 + \delta)^2}, n-1\right)$$

$$\delta = \sigma_0\sqrt{\frac{\chi_{\alpha}}{\chi_{1-\beta}}} - \sigma_0$$

다음은 각 요소에 대한 설명입니다.

σ_0 은 가설 표준편차입니다.
σ 는 실제 표준편차입니다.
X_p 는 $X_{n\text{-}1}{}^2$ 분포의 $(1 - p)^{th}$ 분위수입니다.
$F(x, n\text{-}1)$ 는 x 에서 계산된 $X_{n\text{-}1}{}^2$ 분포의 누적 분포 함수입니다.

수치 해를 사용하여 n 을 계산합니다.

JMP의 계산에 대한 자세한 내용은 Barker(2011, Section 2.4)에서 확인하십시오.

1 표본 비율 계산기

단일 비율에 대한 가설 검정을 위해 표본 크기를 계산하려면 "1 표본 비율 " 계산기를 사용합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

$H_0: p = p_0$

다음 양측 대립가설과 비교

$H_a: p \neq p_0$

또는 다음 단측 대립가설과 비교

$H_a: p < p_0$ 또는 $H_a: p > p_0$

여기서 p는 모비율이고 p_0은 귀무가설 비율입니다.

단일 비율 계산기는 실제 검정 크기도 제공합니다. 이 값은 지정된 가정에 대한 실제 제1종 오류 비율입니다. 이항 분포는 이산형이므로 표본이 작거나 비율이 0 또는 1에 가까운 경우 실제 검정 크기는 명시된 유의 수준과 크게 다를 수 있습니다. 유의 수준이 명시된 수준 이상이 되게 하려면 "정확 Clopper-Pearson" 방법을 사용하십시오.

1 표본 비율 계산기의 예

조립 라인의 과거 결함 비율이 0.1 과 같다고 가정해 보겠습니다 . 표본 크기가 100 이고 유의 수준이 0.05 인 경우 과거 비율과 0.1 이상 차이가 나는 결함 비율을 감지할 수 있는 검정력을 계산하려고 합니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **1 표본 비율**을 클릭합니다 .
3. **알파** 값 0.05 를 그대로 둡니다 .
4. **비율** 값 0.1 을 그대로 둡니다 .
5. **방법** 값 " 정확 Agresti-Coull" 을 그대로 둡니다 .
6. 검정 유형 값 **양측**을 그대로 둡니다 .
7. **귀무가설 비율** 값 0.2 를 그대로 둡니다 .
8. **표본 크기**에 100 을 입력합니다 .
9. **계속**을 클릭합니다 .

그림 17.12 1 표본 비율 계산기

표본 크기가 100 인 경우 검정력은 약 70% 입니다. 실제 검정 크기는 원하는 값 0.05 보다 약간 작은 0.0467 정도입니다. 표본 크기가 100 인 경우 결함 비율이 0.2 이면 귀무가설을 기각하는 제 1 종 오류의 확률은 5% 입니다. 또는 귀무가설을 기각하지 않는 제 2 종 오류의 확률은 30%(1 - 검정력) 입니다.

1 표본 비율 계산기 필드

다음 통계량 및 검정 설정을 지정합니다 .

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

비율 알려져 있거나 가설에 지정된 실제 비율입니다 . 기본값은 0.1 입니다 .

팁 : 다른 모든 모수가 고정된 경우 비율이 0.5 일 때 표본 크기가 가장 큽니다 .

방법 방법을 선택합니다 . " 정확 Agresti-Coull" 또는 " 정확 Clopper-Pearson" 을 선택할 수 있습니다 . Clopper-Pearson 방법은 Agresti-Coull 방법보다 더 보수적인 경향이 있습니다 (표본 크기가 더 큼).

단측 또는 양측 단측 또는 양측 검정 중 하나를 선택합니다 .

다음 모수 중 두 개를 지정하여 세 번째 값을 계산하거나 , 한 값을 지정하여 다른 두 모수 사이의 관계 그림을 볼 수 있습니다 .

참고 : 이 그림은 정확 방법 대신 정규 근사 방법을 사용합니다 .

귀무가설 비율 (p_0) 에 대한 검정 비율입니다 . 기본값은 0.2 입니다 .

표본 크기 실험의 총 관측값 (런 , 실험 단위 또는 표본) 수입니다 .

검정력 귀무가설이 거짓일 때 기각할 확률입니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다.

계산기 버튼

계속 두 개의 모수가 지정된 경우 결측값을 계산하거나, 모수가 하나만 지정된 경우 두 개의 결측 모수를 비교하는 그림을 시작합니다.

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다.

단일 비율 계산기에 대한 통계 상세 정보

단일 비율 표본 크기 계산은 이항 분포를 기반으로 한 정확 방법을 사용합니다. 정확 계산을 사용하면 명시된 검정력 수준을 얻을 수 있습니다.

Agresti-Coull 방법

정확 Agresti-Coull 방법은 수정 Wald 기반 검정 통계량을 사용합니다. JMP는 양측 귀무가설 하에서 다음과 같이 검정력을 계산합니다.

$$\text{검정력} = \sum_{y=0}^{n} \Pr\{Y = y | Y \sim \text{binomial}(n, p)\} I\{T(y) \geq \chi_{1-\alpha}\}$$

다음은 각 요소에 대한 설명입니다.

$$T(y) = \frac{(\hat{p} - p_0)^2}{\frac{\hat{p}(1-\hat{p})}{n+4}}$$

$$\hat{p} = \frac{y+2}{n+4}$$

$I\{T(y_1,y_2) \geq \chi_{1-\alpha}\} = 1$ 단, $(T(y_1,y_2) \geq \chi_{1-\alpha})$ 그렇지 않은 경우 0

$\chi_{1-\alpha}$ 는 χ_1^2 분포의 $(1 - \alpha)^{\text{th}}$ 분위수입니다.

n 또는 p_0에 대해 닫힌 형식의 표현식이 없으므로 수치적 기법을 사용하여 n 또는 p_0을 계산합니다.

수정 Wald 검정 통계량에 대한 자세한 내용은 Agresti and Coull(1998)에서 확인하십시오. JMP의 계산에 대한 자세한 내용은 Barker(2011, Section 3.3)에서 확인하십시오.

Clopper-Pearson 방법

정확 Clopper-Pearson 방법은 이항 분포를 기반으로 합니다. 이 방법을 사용하면 유의 수준이 명시된 수준 이상이 됩니다. Clopper-Pearson 방법은 Agresti-Coull 방법보다 더 보수적입니다(표본 크기가 더 큼).

정확 Clopper-Pearson 방법은 이항 분포를 직접 사용합니다. 수치적 기법을 사용하여 미지의 모수를 계산합니다.

정확 Clopper-Pearson 방법에 대한 자세한 내용은 Clopper and Pearson(1934) 또는 Agresti and Coull(1998, Section 1)에서 확인하십시오.

2 표본 비율 계산기

두 비율에 대한 가설 검정을 위해 표본 크기를 계산하려면 "2 표본 비율 " 계산기를 사용합니다. 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다.

$$H_0: p_1 - p_2 = D_0$$

다음 양측 대립가설과 비교

$$H_a: p_1 - p_2 \neq D_0$$

또는 다음 단측 대립가설 중 하나와 비교

$$H_a: (p_1 - p_2) < D_0 \text{ 또는 } H_a: (p_1 - p_2) > D_0$$

여기서 p_1과 p_2는 두 모집단의 모비율이고, D_0은 가설 비율 차이입니다.

두 비율 계산기는 실제 검정 크기도 제공합니다. 이 값은 지정된 가정에 대한 실제 제1종 오류 비율입니다. 이항 분포는 이산형이므로 표본이 작거나 비율이 0 또는 1에 가까운 경우 실제 검정 크기는 명시된 유의 수준과 크게 다를 수 있습니다.

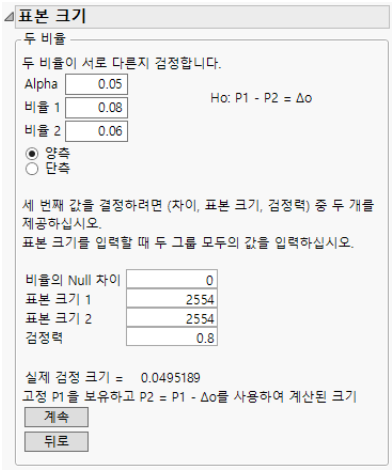
2 표본 비율 계산기의 예

두 개의 실리콘 웨이퍼 조립 라인에 대한 책임자라고 가정해 보겠습니다. 많은 런을 통해 얻은 지식을 바탕으로 조립 라인 중 하나의 결함 비율은 8% 이고 다른 라인의 결함 비율은 6% 입니다. 80% 의 검정력으로 결함 비율에 차이가 있다는 결론을 내리기 위해 필요한 표본 크기를 알아보려고 합니다.

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다.
2. **2 표본 비율**을 클릭합니다.
3. **알파** 설정값 0.05 를 그대로 둡니다.
4. **비율 1** 에 0.08 을 입력합니다.
5. **비율 2** 에 0.06 을 입력합니다.
6. **양측** 검정을 선택합니다.
7. **귀무가설의 비율 차이**에 0 을 입력합니다.
8. **표본 크기 1** 과 **표본 크기 2** 는 비워 둡니다.

9. **검정력**에 0.8 을 입력합니다 .
10. **계속**을 클릭합니다 .

그림 17.13 양측 검정에 대한 두 비율의 차이



이 계산기는 결함 비율의 차이를 감지하려면 각 생산 라인에서 2,554 개의 웨이퍼 표본을 테스트해야 한다는 것을 보여 줍니다 .

2 표본 비율 계산기 필드

다음 통계량 및 검정 설정을 지정합니다 .

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

비율 1 알려져 있거나 가설에 지정된 모집단 1 의 비율입니다 . 기본값은 0.5 입니다 .

비율 2 알려져 있거나 가설에 지정된 모집단 2 의 비율입니다 . 기본값은 0.1 입니다 .

단측 또는 양측 단측 또는 양측 검정 중 하나를 선택합니다 .

표본 크기가 하나의 모수로 간주되는 경우 다음 모수 중 두 개를 지정하여 세 번째 값을 계산합니다 . 표본 크기를 계산하려면 둘 다 입력하거나 둘 다 비워 두어야 합니다 .

귀무가설의 비율 차이 귀무가설 하의 비율 차이입니다 . 기본값은 0.2 입니다 .

표본 크기 1 비율 1 을 추정하기 위한 실험의 총 관측값 (런 , 실험 단위 또는 표본) 수입니다 .

표본 크기 2 비율 2 를 추정하기 위한 실험의 총 관측값 (런 , 실험 단위 또는 표본) 수입니다 .

참고 : 표본 크기가 계산되는 경우 표본 크기 1 과 표본 크기 2 의 값이 같습니다 . 표본 크기를 직접 입력하는 경우에는 두 값이 같을 필요가 없습니다 .

검정력 귀무가설이 거짓일 때 기각할 확률입니다 . 검정력이 높아야 좋습니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

계산기 버튼

계속 결측값을 계산합니다 .

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다 .

두 비율 계산기에 대한 통계 상세 정보

두 비율 계산기는 단일 비율 계산기에 사용된 것과 유사한 정확 방법을 기반으로 합니다. 양측 가설이 다음과 같은 경우

$$H_0: p_1 - p_2 = D_0 \text{대 } H_a: p_1 - p_2 \neq D_0$$

JMP 는 귀무가설 하에서 다음과 같이 검정력을 계산합니다 .

$$\text{검정력} = \sum_{y_1=0}^{n_1} \sum_{y_2=0}^{n_2} \Pr(Y_1 = y_1)\Pr(Y_2 = y_2) I\{T(y_1,y_2) \geq \chi_{1-\alpha}\}$$

수정 Wald 통계량은 다음과 같이 정의됩니다 .

$$T(y_1,y_2) = \frac{(\hat{p}_1 - \hat{p}_2 - \delta_0)^2}{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1+2} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2+2}} \text{ where } \hat{p}_j = \frac{y_j+1}{n_j+2} \text{ 의 경우 } j = 1, 2$$

및

$$I\{T(y_1,y_2) \geq \chi_{1-\alpha}\} = 1 \text{ 단, } (T(y_1,y_2) \geq \chi_{1-\alpha}) \text{ 그렇지 않은 경우 } 0$$

$\chi_{1-\alpha}$는 χ^2 분포의 $(1 - \alpha)^{th}$ 분위수입니다.

n_i 또는 δ_0에 대해 닫힌 형식의 표현식이 없으므로 수치적 기법을 사용하여 n_i 및 δ_0을 계산합니다.

수정 Wald 검정 통계량에 대한 자세한 내용은 Agresti and Coull(1998)에서 확인하십시오. JMP 의 계산에 대한 자세한 내용은 Barker(2011, Section 3.3)에서 확인하십시오.

단위별 개수 계산기

단위별 결함 수에 대한 단측 가설 검정을 위해 표본 크기를 계산하려면 " 단위별 개수 " 계산기를 사용합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

$$H_0: \lambda \leq \lambda_0$$

다음 단측 대립가설과 비교

$$H_a: \lambda > \lambda_0$$

여기서 λ는 Poisson 분포의 비율입니다.

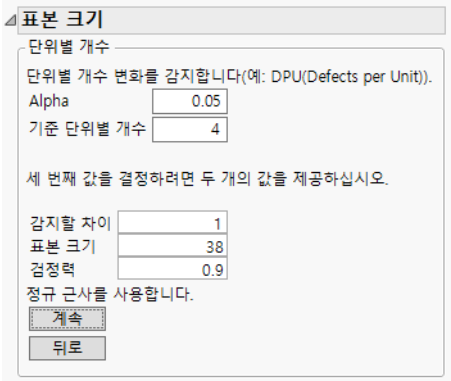
단위별 개수 계산기의 예

웨이퍼당 결함 수가 4 개 (목표값) 이하인 웨이퍼 제조 공정을 가정해 보겠습니다 . 새 공정이 목표값을 충족하는지 확인하려고 합니다 . 0.05 유의 수준에서 90% 의 검정력으로 목표값에 대해 웨이퍼당 결함 수 차이 1 을 감지하려면 몇 개의 웨이퍼를 테스트해야 합니까 ?

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **단위별 개수**를 클릭합니다 .
3. **알파** 값 0.05 를 그대로 둡니다 . 이 값은 새 공정이 목표와 같을 때 검정에 실패할 확률입니다 .
4. **기준 단위별 개수**에 4 를 입력합니다 . 이 값은 웨이퍼당 목표 결함 수가 4 개임을 나타냅니다 .
5. **감지할 차이**에 1 을 입력합니다 .
6. " 검정력 " 에 0.9 를 입력합니다 .

   이 값은 1 보다 큰 변화 , 즉 웨이퍼당 결함 수가 5 이상임을 감지할 확률입니다 . 이러한 경우 유의 수준 (알파) 을 **생산자 위험**이라고 하고 검정력 (베타) 을 **소비자 위험**이라고 합니다 .
7. **계속**을 클릭합니다 .

그림 17.14 단위별 개수 계산기



38 개 웨이퍼를 테스트해야 합니다 . 결함이 173 개 미만일 경우 공정이 목표값을 충족합니다 .

참고 : 173 개의 결함은 귀무가설을 기각하지 않고 38 개 웨이퍼에서 발생할 수 있는 최대 결함 수입니다 .

단위별 개수 계산기 필드

다음 통계량을 지정합니다 .

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

기준 단위별 개수 단위별 목표 결함 수입니다 . 기본값은 0.1 입니다 .

다음 통계량 중 두 개를 지정하여 세 번째 값을 계산합니다 .

감지할 차이 지정된 단위별 결함 수에 대해 검정할 최소 감지 가능 차이입니다 .

표본 크기 단위 수입니다 .

검정력 귀무가설이 거짓일 때 기각할 확률입니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

계산기 버튼

계속 결측값을 계산합니다 .

뒤로 이전 "표본 크기 및 검정력" 시작 창으로 돌아갑니다.

단위별 개수 계산기에 대한 통계 상세 정보

단위별 개수 계산기의 계산은 Poisson 분포에 정규 근사 적용을 기반으로 합니다 . 이 검정에는 다음 가설이 사용됩니다 .

$H_0: \lambda \le \lambda_0$ 대 $H_a: (\lambda > \lambda_o)$

위의 가설에는 다음 검정 통계량이 사용됩니다 .

$$T = \frac{\bar{d} - \lambda_0}{\sqrt{\lambda_0 / n}}$$

여기서 n은 표본 크기이고 $\bar{d}$는 표본의 단위별 평균 결함 수입니다. T는 대략적으로 정규 분포되었다고 가정합니다. 검정력 계산은 귀무가설과 대립가설 하에서 T 분포를 기반으로 합니다.

$$\text{검정력} = \Pr(\text{reject } H_0 \mid \lambda = \lambda_0 + \delta)$$

$$= \Pr[T > Z_{1-\alpha} \mid (\lambda = \lambda_0 + \delta)] = 1 - \Phi\left(\frac{Z_{1-\alpha} - \delta\sqrt{n/\lambda_0}}{\sqrt{(\lambda_0 - \delta)/\lambda_0}}\right)$$

여기서 $\Phi()$는 표준 정규 누적 분포 함수이고 Z_{1-p}는 표준 정규 분포의 $(1 - p)^{\text{th}}$ 분위수입니다.

1 - β 를 사용하여 귀무가설 기각을 위한 원하는 검정력을 나타내면 표본 크기가 다음과 같이 계산됩니다 .

$$n = \frac{\lambda_0}{\delta^2}\left(Z_{1-\alpha} - Z_{\beta}\sqrt{\frac{\lambda_0 + \delta}{\lambda_0}}\right)^2$$

δ에 대한 분석 해가 존재하지 않으므로 검정력과 n이 주어지면 수치적 방법을 사용하여 δ 값을 계산합니다.

JMP 의 계산에 대한 자세한 내용은 Barker(2011, Section 2.5) 에서 확인하십시오 .

시그마 품질 수준 계산기

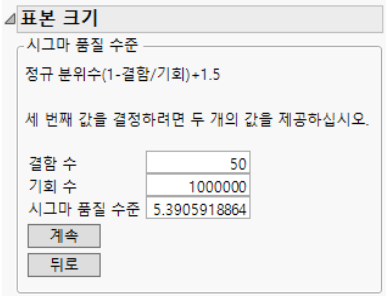
결함 수, 결함 기회 수 및 시그마 품질 수준 간의 관계를 탐색하려면 "시그마 품질 수준" 계산기를 사용합니다. 시그마 품질 수준은 주어진 결함 비율을 6 시그마 척도와 관련짓는 통계량입니다. 예를 들어 백만 개의 기회별 3.4 결함 비율은 6 시그마 공정이 됩니다.

시그마 품질 수준 계산기의 예

플랜트에서 1,000,000 개의 기회 중 50 개 결함에 대한 시그마 품질 수준을 추정하려고 합니다 . 기회는 결함 수에 대한 단위입니다 . 단일 항목일 수도 있고 , 각 잠재적 결함이 기회인 단일 항목에 여러 결함이 있을 수도 있습니다 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **시그마 품질 수준**을 클릭합니다 .
3. **결함 수**에 50 을 입력합니다 .
4. **기회 수**에 1000000 을 입력합니다 .
5. **계속**을 클릭합니다 .

그림 17.15 시그마 품질 수준 계산기



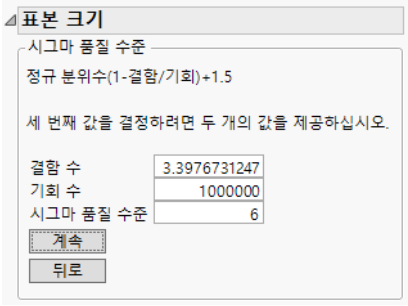
시그마 품질 수준은 약 5.4 입니다 .

결함 수 계산의 예

백만 개의 기회 중 발생할 수 있는 최대 결함 수를 계산하고 6 시그마 품질 수준 공정을 계속 사용하려면 다음 단계를 수행하십시오 .

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **시그마 품질 수준** 버튼을 클릭합니다 .
3. **결함 수**는 비워 둡니다 .
4. **기회 수**에 1000000 을 입력합니다 .
5. **시그마 품질 수준**에 6 을 입력합니다 .
6. **계속**을 클릭합니다 .

그림 17.16 1,000,000 개 기회에 대한 6 시그마 품질 수준



계산 결과, 6 시그마 품질 수준 공정을 달성하려면 1,000,000 개 기회 중 **결함 수**가 3 을 초과할 수 없음을 알 수 있습니다.

시그마 품질 수준 계산기 필드

다음 통계량 중 두 개를 지정하여 세 번째 값을 계산합니다 .

결함 수 결함 수입니다 .

기회 수 결함에 대한 기회 수입니다 .

시그마 품질 수준 표준편차 단위의 결함 비율입니다.

계산기 버튼

계속 결측값을 계산합니다 .

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다 .

시그마 품질 계산기에 대한 통계 상세 정보

시그마 품질 계산기의 계산은 n 개 기회 중 d 개 결함에 대한 시그마 품질 수준 σ_q 의 정의를 기반으로 합니다 .

$$\sigma_q = Z_{1-\frac{d}{n}} + 1.5$$

이 정의에서 d 와 n 은 다음과 같이 계산할 수 있습니다 .

$$d = n[1 - \Phi(\sigma_q - 1.5)]$$

$$n = d[1 - \Phi(\sigma_q - 1.5)]^{-1}$$

여기서 $\Phi()$는 표준 정규 누적 분포 함수이고 $Z_{1-d/n}$는 표준 정규 분포의 $(1 - d/n)^{th}$ 분위수입니다.

신뢰도 시험 계획 계산기

신뢰성 연구에 필요한 표본 크기, 연구 기간 또는 정밀도를 결정하려면 "신뢰도 시험 계획" 계산기를 사용합니다. 신뢰도 연구는 제품의 고장 시간 및 고장 확률을 추정하는 데 사용됩니다. 시험 계획 계산기는 기대 신뢰 구간을 사용하여 추정값의 정밀도를 정의합니다. 이러한 추정값은 적합 고장 분포를 기반으로 합니다. 고장 분포를 적합시키려면 최소한의 고장 수가 관측되어야 합니다.

신뢰도 시험 계획 계산기의 예

회사에서 신제품을 개발한 후 양측 절대 정밀 200 시간과 95% 신뢰 구간을 사용하여 단위의 20%가 고장 날 때까지의 시간을 추정하기 위해 시험할 단위 수를 추정하려고 합니다. 즉, 추정 시간에 대한 신뢰 구간이 생성될 때 상한과 하한의 차이가 약 200 시간이어야 합니다. 회사는 2,500 시간 동안 시험을 실행할 수 있습니다. 또한 유사 제품에 대한 연구를 통해 대략적인 고장 분포를 위치 모수 $\alpha = 2000$, 척도 모수 $\beta = 3$ 인 Weibull 분포로 가정합니다.

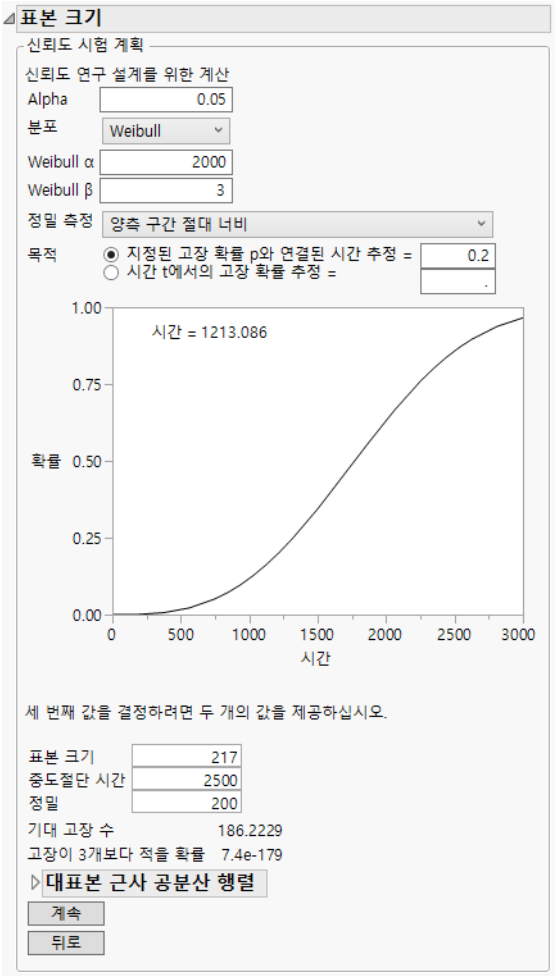
필요한 표본 크기를 계산하려면

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **신뢰도 시험 계획**을 클릭합니다 .
3. 알파 설정값 0.05 를 그대로 둡니다 .
4. " 분포 " 목록에서 **Weibull** 을 선택합니다 .
5. **Weibull** α 모수에 2000 을 입력합니다 .
6. **Weibull** β 모수에 3 을 입력합니다 .
7. " 정밀 측정 " 목록에서 **양측 구간 절대 너비**를 선택합니다 .
8. **지정된 고장 확률 p 와 연결된 시간 추정**을 선택하고 **p** 에 0.2 를 입력합니다 .

참고 : 연구 목적에 따라 누적 분포 함수 그림에 라벨이 지정됩니다 . 여기서 시간 = 1213 은 20% 의 단위 고장에 대한 시간 추정값입니다 .

9. **표본 크기**는 비워 둡니다 .
10. **중도절단 시간**에 2500 을 입력합니다 .
11. **정밀**에 200 을 입력합니다 .
12. **계속**을 클릭합니다 .

그림 17.17 신뢰도 시험 계획 계산기



200 시간의 정밀도로 단위의 20% 가 고장 날 때까지 걸린 시간을 추정하려면 2,500 시간 동안 217 개 단위를 시험해야 합니다. 이 시험의 기대 고장 수는 약 186 개로, 고장 분포 추정에 필요한 최소값 3 개보다 훨씬 큽니다.

신뢰도 시험 계획 계산기 필드

다음 통계량 및 검정 설정을 지정합니다 .

알파 정밀 측정을 정의하는 데 사용되는 신뢰 구간의 유의 수준입니다 .

분포 가정된 고장 분포입니다 . 사용 가능한 분포는 Weibull, 로그 정규 , Frechet, 로그로지스틱 , SEV, 정규 , LEV 및 로지스틱입니다 . 이러한 분포에 대한 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

위치 고장 분포의 위치 모수입니다 .

참고 : " 분포 " 가 "Weibull" 로 설정되면 " 위치 " 필드가 "Weibull α" 로 표시됩니다 .

척도 고장 분포의 척도 모수입니다 .

참고 : " 분포 " 가 "Weibull" 로 설정되면 " 척도 " 필드가 "Weibull β" 로 표시됩니다 .

정밀 측정 정밀 측정의 정의입니다 . 정의는 추정되는 수량 (고장 시간 또는 확률) 의 기대 신뢰 구간을 기반으로 합니다 . 정밀 측정에 선택할 수 있는 옵션은 다음과 같습니다 .

구간 비율 정밀 값을 하한에 대한 상한 비율의 제곱근으로 정의합니다 . 상한이 하한보다 크거나 같으므로 이 비율은 항상 1 보다 큽니다 . 추정값의 정밀도가 증가하면 구간 비율이 감소합니다 .

양측 구간 절대 너비 정밀 값을 신뢰 구간의 너비 또는 상한과 하한의 차이로 정의합니다 .

하위 단측 구간 절대 너비 정밀 값을 하위 구간의 너비 또는 추정값 신뢰 구간의 하한과 추정값의 차이로 정의합니다 .

양측 구간 상대 너비 정밀 값을 추정값에 상대적인 신뢰 구간 너비로 정의합니다 . 이 값은 상한과 하한 간의 차이를 추정값으로 나눈 것입니다 .

하위 단측 구간 상대 너비 정밀 값을 추정값에 상대적인 하위 구간 너비로 정의합니다 . 이 값은 추정값과 하한 간의 차이를 추정값으로 나눈 것입니다 .

목적 연구 목적입니다 . 다음 목적 중 하나를 선택하고 해당 값을 입력합니다 .

- 지정된 고장 확률 p 와 연결된 시간 추정
- 시간 t 에서의 고장 확률 추정

참고 : 그림은 고장 분포의 누적 분포 함수입니다 . 연구 목적에 따라 시간 또는 확률 추정값 라벨이 그림에 지정됩니다 .

다음 통계량 중 두 개를 지정하여 세 번째 값을 계산합니다 .

표본 크기 신뢰도 시험에 포함할 단위 수입니다 .

중도절단 시간 신뢰도 시험을 실행하는 데 사용 가능한 시간입니다 .

정밀 정밀 수준입니다 . 이 값에 대한 단위의 정의는 선택한 정밀 측정에 해당합니다 .

계산기 버튼

계속 결측값을 계산합니다 .

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다 .

추가 신뢰도 시험 계획 계산

표본 크기 , 중도절단 시간 또는 정밀뿐만 아니라 다음 통계량도 계산됩니다 .

기대 고장 수 지정된 신뢰도 시험에 대해 기대되는 고장 수입니다 .

고장이 3 개보다 적을 확률 지정된 신뢰도 시험으로 인한 고장 수가 3 개보다 적을 확률입니다 . 이 값은 고장 분포의 위치 및 척도 모수에 대해 안정적인 추정값을 얻기 위해 필요한 최소 고장 수가 3 개이므로 중요합니다 . 고장 수가 한 개 또는 두 개뿐이면 추정값이 불안정합니다 . 이 확률이 크면 분포 모수의 추정을 신뢰할 수 있을 만큼 충분한 수의 고장을 관측하지 못할 위험이 있습니다 . 표본 크기 또는 중도절단 시간을 늘리면 고장이 3 개보다 적을 확률을 낮출 수 있습니다 .

대표본 근사 공분산 행렬 고장 분포의 위치 모수와 척도 모수에 대한 근사 분산 및 공분산을 제공합니다 .

신뢰도 시험 계획 계산기에 대한 통계 상세 정보

신뢰도 시험 계획의 목적은 다음 통계량

$$q(p) = \exp[\sigma\Phi^{-1}(p) + \mu]$$

또는 다음 고장 확률을 추정하는 것입니다 .

$$p(t) = \Phi\left[\frac{\log(t) - \mu}{\sigma}\right]$$

이때 가정된 고장 시간 분포 및 정밀 측정과 함께 지정된 유의 수준이 사용됩니다 . $\Phi()$ 는 위치 모수가 μ 이고 척도 모수가 σ 인 가정된 고장 시간 분포의 표준 누적 분포 함수입니다 . Wald 신뢰 구간 및 정밀 측정은 t(시간)와 n(표본 크기)의 함수로 정밀 수준을 정의합니다 . 수치적 방법을 사용하여 원하는 수량을 계산합니다 .

JMP 의 계산에 대한 자세한 내용은 Barker(2011, Section 6) 에서 확인하십시오 .

신뢰도 데모 계산기

지정된 기간 동안 지정된 수의 단위를 시험하려면 " 신뢰도 데모 " 계산기를 사용합니다 . k 개 미만의 단위에서 고장이 발생하면 데모를 통과하고 , 제품 신뢰도가 신뢰도 표준을 충족하거나 초과한다는 결론을 내릴 수 있습니다 . 신뢰도 데모를 다음과 같은 가설 검정으로 제시해 보겠습니다 .

$H_o: p < p^*$ 대 $H_a: p \geq p^*$

여기서 p는 신제품에 대한 t^*에서의 생존 시간 확률이고, p^*는 시간 t^*에서의 표준 생존 확률입니다.

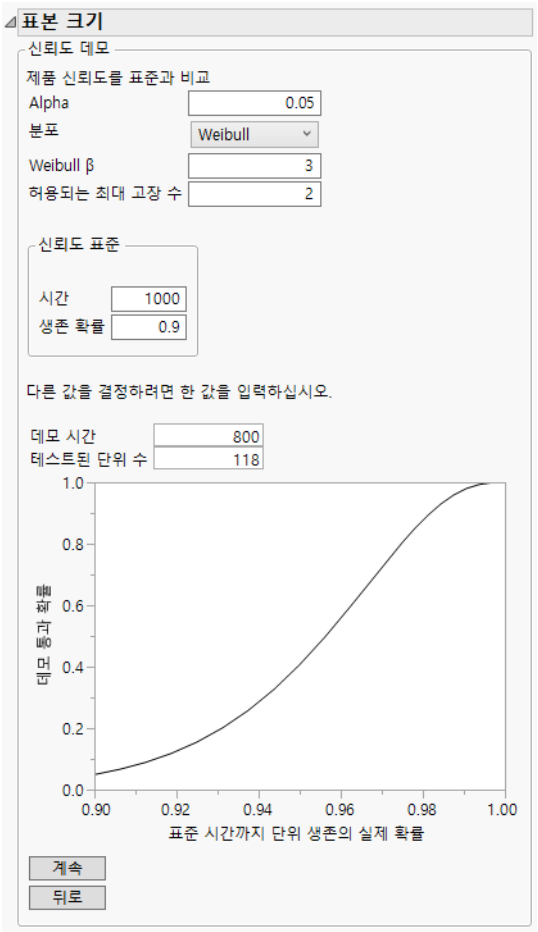
신뢰도 데모 계산기의 예

회사에서 1,000 시간 후의 생존 확률에 대한 과거 신뢰도 표준 (90%) 과 비교하여 신제품의 신뢰도를 평가하는 데 필요한 표본 크기를 추정하려고 합니다 . 유사 제품에 대한 이전 연구를 통해 고장 시간 분포를 β 모수가 3 인 Weibull 로 가정합니다 . 800 시간 동안 데모를 실행할 수 있으며 실험 결과로 허용되는 최대 고장 수는 2 개입니다 .

필요한 표본 크기를 계산하려면

1. **DOE > 설계 진단 > 표본 크기 및 검정력**을 선택합니다 .
2. **신뢰도 데모**를 클릭합니다 .
3. 알파 설정값 0.05 를 그대로 둡니다 .
4. " 분포 " 목록에서 **Weibull** 을 선택합니다 .
5. **Weibull** β 에 3 을 입력합니다 .
6. **허용되는 최대 고장 수**에 2 를 입력합니다 .
7. **시간**에 1000 을 입력합니다 .
8. **생존 확률**에 0.9 를 입력합니다 .
9. **데모 시간**에 800 을 입력합니다 .
10. **테스트된 단위 수**는 비워 둡니다 .
11. **계속**을 클릭합니다 .

그림 17.18 신뢰도 데모 계산기



데모에서 118 개의 단위를 실행해야 합니다. 또한 800 시간 동안 관측된 고장 수가 2 개 이하이면 신제품 신뢰도가 표준 이상이라는 결론을 내릴 수 있습니다.

이 그림은 신뢰도 표준의 시간까지 단위 생존의 실제 확률이 증가할수록 데모 통과 확률이 증가함을 보여 줍니다. 이 예에서 표준 시간까지 단위 생존 확률이 92% 를 넘지 못하면 신제품의 데모 통과 확률은 약 12.5% 에 불과합니다.

신뢰도 데모 계산기 필드

다음 통계량 및 검정 설정을 지정합니다 .

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

분포 가정된 고장 시간 분포입니다 . 사용 가능한 분포는 Weibull, 로그 정규 , Frechet, 로그로지스틱 , SEV, 정규 , LEV 및 로지스틱입니다 . 이러한 분포에 대한 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

척도 고장 시간 분포의 척도 모수입니다 .

참고 : " 분포 " 가 "Weibull" 로 지정되면 " 척도 " 필드가 "Weibull β" 로 표시됩니다 .

허용되는 최대 고장 수 테스트 데모 성공을 위해 허용되는 최대 고장 수입니다 .

신뢰도 표준 테스트 중인 표준의 정의입니다 . 이 표준은 시간 및 해당 시간까지 항목이 생존할 확률을 기반으로 합니다 .

시간 테스트 대상 항목이 생존해야 하는 최소 시간입니다 .

생존 확률 테스트 대상 항목이 신뢰도 표준에 정의된 시간까지 생존할 확률입니다 .

다음 두 값 중 하나를 입력하여 다른 값을 계산합니다 .

데모 시간 데모 실행 시간입니다 .

테스트된 단위 수 데모에 필요한 단위 수입니다 .

계산기 버튼

계속 결측값을 계산합니다 .

뒤로 이전 " 표본 크기 및 검정력 " 시작 창으로 돌아갑니다 .

신뢰도 데모 계산기에 대한 통계 상세 정보

신뢰도 데모는 척도 모수가 σ 인 가정된 고장 시간 분포에 따라 달라집니다 . 신뢰도 표준 , 즉 시간 t 와 위치 μ 에서의 생존 확률은 다음과 같이 명시됩니다 .

$$p = 1 - \Phi\left[\frac{\log t - \mu}{\sigma}\right]$$

여기서 μ 는 다음 식을 사용하여 계산됩니다 .

$$\mu = \log(t) - \sigma\Phi^{-1}(1-p)\cdot$$

표본 크기와 검정 크기를 계산하기 위해 시간 t 에서의 생존 확률이 다음과 같은 가설 검정으로 제시됩니다 .

$H_o: p < p^*$ 대 $H_a: p \geq p^*$

여기서 p^*는 시간 t^*에서의 표준 생존 확률입니다.

다음과 같은 α 수준에서 가설을 검정하려고 합니다 .

$\alpha = \Pr(k$ 개 이하의 고장 수 $\mid H_0$ true).

n개의 독립 단위에 대한 검정이므로 고장 수는 (n, p) 이항 분포를 갖습니다. 여기서 p는 시간 t 전에 단위가 고장 날 확률입니다 . 따라서 α 를 t 와 n 의 함수로 표현할 수 있습니다 .

$$\alpha = \Pr\left(x \le k | x \sim \text{Bin}\left\{n, \Phi\left[\frac{\log t - \mu^*}{\sigma^*}\right]\right\}\right)$$

여기서 μ^*와 σ^*는 가정된 신뢰도 표준의 값입니다.

이항 분포와 베타 분포의 특성으로 인해 다음 식을 사용하여 t 를 계산할 수 있습니다 .

$$t = \frac{t^* \exp\{\sigma^* \Phi^{-1}[1 - \text{B}^{-1}(\alpha; n-k, k+1)]\}}{\exp[\sigma^* \Phi^{-1}(1-p^*)]}$$

n 의 경우 Brent 방법을 사용하여 다음 식의 근을 찾습니다 .

$$\text{B}^{-1}(\alpha; n-k, k+1) - 1 + \Phi\left[\frac{\log t - (\log t^*) + \sigma^* \Phi^{-1}(1-p^*)}{\sigma^*}\right] = 0$$

다음은 각 요소에 대한 설명입니다 .

$\text{B}^{-1}(\alpha; n-k, k+1)$ 은 $\text{Beta}(n-k; k+1)$ 분포의 α 분위수입니다 .

$\Phi()$는 가정된 고장 시간 분포의 누적 분포 함수입니다.

JMP의 계산에 대한 자세한 내용은 Barker(2011, Section 5)에서 확인하십시오.

18 장

선택 설계

기본 제품 프로파일 선택을 위한 설계 생성

선택(또는 이산 선택) 실험은 이산 선호도를 모델링하기 위한 데이터를 제공합니다. 연구 참가자에게는 다양한 속성을 가진 잠재적 제품 집합, 즉 제품 프로파일이 제공됩니다. 참가자는 각 프로파일 집합에서 선호하는 프로파일을 선택합니다. 예를 들어 컴퓨터 회사에서 고급 랩톱을 설계할 때 프로세서 속도, 하드 디스크 크기, 화면 크기, 배터리 수명 및 가격과 같은 주요 특징의 상대적 중요도에 관심이 있을 수 있습니다. 선택 실험은 고객에게 이러한 특징의 상대 값을 제시하고 제품 특징 간에 최적의 절충안을 나타냅니다.

선택 실험 결과는 컨조인트 분석 방법을 사용하여 분석됩니다. 자세한 내용은 Consumer Research의 에서 확인하십시오.

그림 18.1 8 개의 선택 집합이 있는 설문 조사

	선택 집합	선택 ID	Disk Size	Speed	Battery Life	Price
1	1	1	40 GB	2.0 GHz	6 Hrs	$1000
2	1	2	80 GB	1.5 GHz	4 Hrs	$1500
3	2	1	80 GB	1.5 GHz	4 Hrs	$1200
4	2	2	40 GB	1.5 GHz	6 Hrs	$1500
5	3	1	40 GB	1.5 GHz	4 Hrs	$1200
6	3	2	80 GB	2.0 GHz	6 Hrs	$1500
7	4	1	40 GB	2.0 GHz	4 Hrs	$1000
8	4	2	80 GB	1.5 GHz	6 Hrs	$1500
9	5	1	80 GB	2.0 GHz	6 Hrs	$1200
10	5	2	40 GB	1.5 GHz	6 Hrs	$1000
11	6	1	40 GB	1.5 GHz	4 Hrs	$1500
12	6	2	80 GB	1.5 GHz	6 Hrs	$1000
13	7	1	80 GB	1.5 GHz	4 Hrs	$1000
14	7	2	40 GB	2.0 GHz	6 Hrs	$1200
15	8	1	80 GB	1.5 GHz	6 Hrs	$1200
16	8	2	80 GB	2.0 GHz	4 Hrs	$1500

목차

선택 설계 개요 509
선택 설계 용어 509
베이지안 D-최적 509
선택 설계의 예 510
분석을 사용한 선택 설계의 예 512
파일럿 연구를 위한 선택 설계 생성 512
파일럿 연구 데이터 분석 514
사전 정보를 사용하여 최종 선택 실험 설계 516
설계 실행 및 결과 분석 517
선택 설계 창 520
속성 520
모형 521
설계 생성 523
설계 524
테이블 생성 525
선택 설계 옵션 525
기술 상세 정보 526
베이지안 D-최적 및 설계 구성 526
유틸리티 중립 설계 및 로컬 D-최적 설계 527

선택 설계 개요

이산 선택 실험은 제품 설계 공정을 뒷받침합니다. 이러한 실험을 통해 사람들이 사고 싶어 하는 제품을 설계할 수 있도록 회사 시장에 맞는 제품 특징의 우선 순위를 정할 수 있습니다. 선택 설계 플랫폼은 제품 속성을 요인으로 사용하여 실험을 생성합니다. 연구할 속성과 해당 값을 선택하는 것은 매우 중요합니다. 소비자의 제품 구매 결정에 영향을 미칠 수 있는 모든 속성을 포함해야 합니다. 선택 실험 설계에 대한 자세한 내용과 지침은 Sall(2008)에서 확인하십시오.

선택 설계 용어

다음은 선택 설계와 관련된 용어입니다.

- **속성**은 제품의 특징입니다.
- **프로파일**은 제품 속성의 규격입니다.
- **선택 집합**은 프로파일 모음입니다.
- **설문 조사**는 선택 집합 모음입니다.
- **부분 프로파일**은 각 선택 집합 내에서 지정된 수의 속성만 변경되는 선택 설계의 프로파일입니다. 나머지 속성은 변하지 않습니다.

이산 선택 실험에서는 여러 선택 집합을 포함하는 설문 조사를 응답자에게 제공합니다. 일반적으로 선택 집합은 의사결정 과정을 쉽게 완료하기 위해 소수의 프로파일만 포함합니다. 각 응답자는 각 선택 집합 내에서 자신이 선호하는 프로파일을 지정합니다. 예를 들어 랩톱 실험의 속성에는 속도, 저장 용량, 화면 크기, 배터리 수명 및 가격이 포함될 수 있습니다. 이러한 속성의 다양한 조합으로 제품 프로파일이 구성됩니다. 선택 집합은 두 개의 프로파일로 구성될 수 있습니다. 응답자는 각 선택 집합에서 원하는 프로파일을 선택합니다.

관련된 속성이 많은 경우 각 선택 집합에 **부분 프로파일**이 포함되는 설문 조사를 구성할 수 있습니다. 부분 프로파일이 있는 선택 집합에서는 지정된 수의 속성만 변하고 나머지 속성은 일정하게 유지됩니다. 이렇게 하면 선택 작업의 복잡성을 줄일 수 있습니다.

베이지안 D- 최적

이산 선택 모형은 모수에서 비선형이므로 선택 설계의 효율도는 미지의 모수에 따라 달라집니다. 선택 설계 플랫폼은 사용자가 지정하는 모수 확률 값의 사전 분포에 대해 설계를 최적화하는 베이지안 방법을 사용합니다. 베이지안 D-최적 기준은 사전 분포와 관련한 정보 행렬의 행렬식에 대한 로그 기대값입니다. 선택 설계 플랫폼은 사전 확률 분포와 관련하여 이 기대를 최대화합니다. 자세한 내용은 "베이지안 D-최적 및 설계 구성"(526페이지) 및 Kessels et al. (2011)에서 확인하십시오.

다음과 같은 유형의 설계도 생성할 수 있습니다.

- 유틸리티 중립 설계 - 유틸리티 중립 설계에서는 선택 집합 내의 모든 선택 항목에 대한 확률이 동등합니다. 사전 평균은 0 으로 설정됩니다.

- 로컬 D- 최적 설계 - 로컬 D- 최적 설계는 사전 평균을 고려하지만 사전 공분산 행렬의 정보는 포함하지 않습니다 .

유틸리티 중립 설계와 로컬 D-최적 설계에 대한 자세한 내용은 "유틸리티 중립 설계 및 로컬 D-최적 설계"(527 페이지)에서 확인하십시오 .

선택 설계의 예

실험 정보

이 예의 경우 커피숍에서 대부분의 고객을 만족시킬 수 있는 이상적인 커피를 만들고 싶어 합니다 . 관리자는 고객 선호도에 영향을 주는 요인을 판별해 줄 것을 요청했습니다 . 특히 다음 요인 (속성) 에 대해 이상적인 커피를 만드는 설정을 판별해야 합니다 .

- 커피 입자 크기 (중간 또는 굵게)
- 온도 (195°, 200°, 205°)
- 추출 시간 (3 분 , 3.5 분 , 4 분)
- 1 온스당 커피 양 (1.6g, 2g, 2.4g)

요인 수준의 각 조합을 **프로파일**이라고 합니다. 가능한 모든 프로파일을 모든 응답자 표본에 제공하여 선호도 정보를 얻는 것은 실용적이지 않습니다. 그러나 소수의 프로파일로 구성된 **선택 집합**에서 선호하는 프로파일을 선택하도록 응답자에게 요청할 수 있습니다.

이 예에서는 각 응답자가 여러 선택 집합에서 자신의 선호도를 나타내는 실험을 설계합니다 . 설계 구조는 다음과 같습니다 .

- 응답자 10 명
- 선택 집합 12 개
- 선택 집합당 프로파일 2 개
- 응답자당 12 개의 선택 집합이 모두 포함된 설문 조사 1 개

실험 결과 각 응답자마다 12 개의 반응이 나옵니다. 이러한 선호도를 분석하면 대부분의 고객을 만족시키는 커피를 만드는 방법에 대한 결론을 내릴 수 있습니다.

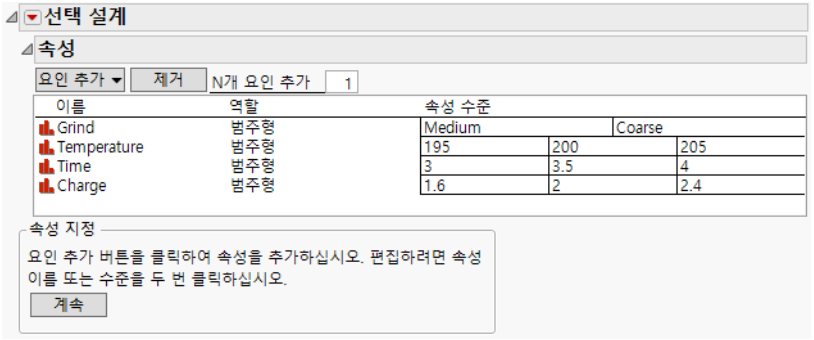
설계 생성

요인을 수동으로 입력하거나 , 요인과 설정이 포함된 기존 테이블을 사용하여 자동으로 입력합니다 . 편의상 이 예에서는 기존 테이블을 사용합니다 . 그러나 새 실험을 설계하는 경우 먼저 수동으로 요인을 입력해야 합니다 . 요인을 수동으로 입력하는 방법에 대한 자세한 내용은 " 속성 "(520 페이지) 에서 확인하십시오 .

1. **DOE > 소비자 연구 > 선택 설계**를 선택합니다 .
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Coffee Choice Factors.jmp 를 엽니다 .

3. " 선택 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

그림 18.2 속성이 정의된 선택 설계 창



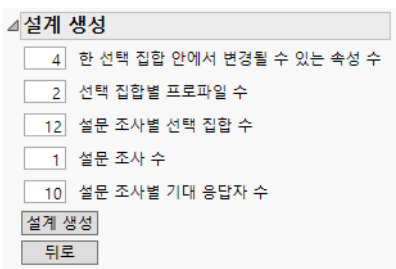
4. **계속**을 클릭합니다 .
5. (선택 사항) "DOE 모형 제어 " 개요를 엽니다 .

   주효과만 있습니다. 이 예에서는 네 가지 요인의 주효과를 포함하는 모형에만 관심이 있습니다 . 그러나 추가 효과를 추정할 수 있는 설계를 생성하려면 이 개요에서 원하는 효과를 추가합니다 .
6. " 설계 생성 " 패널에서 다음을 수행합니다 .
   - **한 선택 집합 안에서 변경될 수 있는 속성 수** 4 를 그대로 유지합니다 .
   - **선택 집합별 프로파일 수** 2 를 그대로 유지합니다 .
   - **설문 조사별 선택 집합 수**에 12 를 입력합니다 .

     이 예에서는 응답자가 12 개의 선택 집합을 평가합니다 .
   - **설문 조사 수** 1 을 그대로 유지합니다 .
   - **설문 조사별 기대 응답자 수**에 10 을 입력합니다 .

     이 예에서는 응답자가 10 명입니다 .

그림 18.3 완료된 설계 생성 패널



참고 : 7 단계에서 " 난수 시드값 " 을 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이 단계가 필요하지 않습니다 .

7. (선택 사항) " 선택 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 . "12345" 를 입력하고 **확인**을 클릭합니다 .
8. **설계 생성**을 클릭합니다 .

   12 개의 선택 집합이 있고 각 선택 집합은 두 개의 커피 프로파일로 구성됩니다 .
9. **프로파일 및 반응에 대해 별도의 테이블 출력**을 선택합니다 .

   이렇게 하면 선택 집합에 대한 설명이 " 선택 프로파일 " 이라는 데이터 테이블에 추가됩니다 . 두 번째 데이터 테이블("선택 런")은 반응 정보를 쉽게 입력할 수 있도록 하기 위해 생성됩니다.
10. **테이블 생성**을 클릭합니다 .

    " 선택 프로파일 " 테이블은 각각 두 개의 프로파일로 구성된 선택 집합 12 개를 보여 줍니다 . " 선택 런 " 테이블에서는 **프로파일 ID** 를 사용하여 **반응** 열에 선호하는 프로파일을 기록할 수 있습니다 . 선호하는 프로파일이 선택 1 이면 1 을 입력하고 선택 2 이면 2 를 입력합니다 . 응답자가 선호하는 항목이 없을 경우에는 반응을 결측 상태로 둘 수도 있습니다 .

    " 선택 프로파일 " 테이블의 **선택** 스크립트를 사용하면 실험 결과를 쉽게 분석할 수 있습니다 . 이 스크립트는 " 선택 모형 " 의 완료된 시작 창을 엽니다 . 선택 모형에 대한 자세한 내용은 Consumer Research 의 에서 확인하십시오 .

    " 선택 프로파일 " 테이블의 **DOE 대화상자** 스크립트는 설계 대화상자를 다시 시작합니다 .

분석을 사용한 선택 설계의 예

이 예에서는 컴퓨터 제조업체가 신형 랩톱을 제조하려고 하는데 비용이 많이 드는 개발 공정을 시작하기 전에 고객 선호도 정보를 얻으려고 합니다 . 제조업체는 두 개의 프로파일 집합으로 구성된 설계를 구성하여 10 명의 응답자에게 제공하기로 결정합니다 . 이 선택 설계의 목표는 잠재적 랩톱 구매자가 다음과 같은 네 가지 속성 모음의 이점에 대해 어떻게 생각하는지 이해하는 것입니다 .

- 하드 드라이브 디스크 크기 (40GB 또는 80GB)
- 프로세서 속도 (1.5GHz 또는 2.0GHz)
- 배터리 수명 (4 시간 또는 6 시간)
- 컴퓨터 가격 (1000 달러 , 1200 달러 또는 1500 달러)

10 명의 응답자에 대한 설계를 구성하려면 먼저 응답자 한 명만 대상으로 소규모 파일럿 연구를 수행합니다 . 그런 다음 결과를 분석하고 해당 모수 추정값을 10 명의 응답자에 대한 최종 연구를 설계할 때 사전 정보로 사용합니다 .

파일럿 연구를 위한 선택 설계 생성

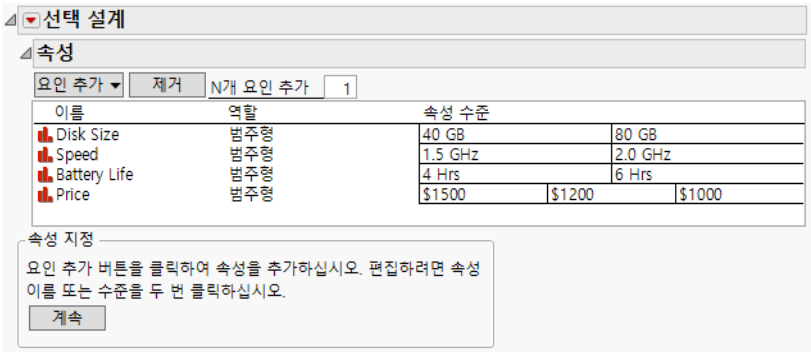
이 섹션에서는 응답자가 한 명인 연구를 위한 선택 설계를 구성합니다.

요인 및 수준 정의

이 예에서는 기존 테이블의 요인을 불러옵니다 . 새 실험을 직접 설계할 때는 요인을 수동으로 입력합니다 .

1. **DOE > 소비자 연구 > 선택 설계**를 선택합니다 .
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Laptop Factors.jmp 를 엽니다 .
3. " 선택 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

그림 18.4 속성이 정의된 선택 설계 창



설계 생성

1. **계속**을 클릭합니다 .
2. 이 파일럿 설문 조사는 단일 응답자에게 제공됩니다 . "DOE 모형 제어 "," 사전 규격 " 및 " 설계 생성 " 패널의 기본값은 그대로 유지합니다 .
3. (선택 사항) " 선택 설계 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 . "12345" 를 입력하고 **확인**을 클릭합니다 .
4. **설계 생성**을 클릭합니다 .

참고 : 3 단계에서 " 난수 시드값 " 을 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이 단계가 필요하지 않습니다 .

그림 18.5 파일럿 설계

단일 설문 조사에는 8 개의 선택 집합이 포함되고 각 선택 집합은 두 개의 랩톱 프로파일로 구성됩니다 .

5. **프로파일과 반응을 한 테이블에 결합** 옵션이 선택되어 있는지 확인합니다 .

   이렇게 하면 선택 집합과 설문 조사 결과가 동일한 테이블에 추가됩니다 .

6. **테이블 생성**을 클릭합니다 .

이 설문 조사는 사전 정보가 없다는 가정하에 설계되었습니다. 이로 인해 일부 선택 집합에서 유용한 정보를 끌어내지 못할 수 있습니다. 계획은 단일 응답자로부터 설문 조사 결과를 얻고 결과를 분석한 후 파일럿 설문 조사 결과를 최종 연구 설계의 사전 정보로 사용하는 것입니다.

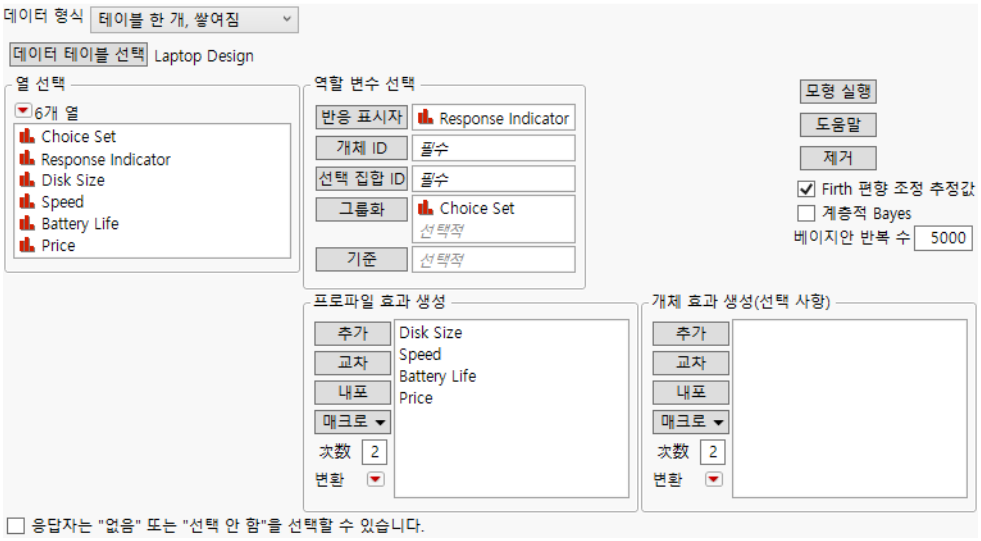
파일럿 연구 데이터 분석

파일럿 설문 조사 설계가 완료되었으므로 단일 응답자에게 제공합니다. 응답자는 각 집합에서 하나의 프로파일을 선택한 후 선택한 프로파일에는 1 을 입력하고 거부한 프로파일에는 0 을 입력합니다. 그러면 선택 플랫폼을 사용하여 결과를 분석합니다.

참고 : 선택 플랫폼에 대한 자세한 내용은 Consumer Research 의 에서 확인하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Laptop Design.jmp 를 엽니다.
2. **Choice** 스크립트를 실행합니다 .

그림 18.6 선택 모형 시작 창



단일 설문 조사 및 단일 응답자가 있으므로 그룹화 변수는 Choice Set 뿐입니다 .

3. **모형 실행**을 클릭합니다 .

그림 18.7 파일럿 연구의 모수 추정값

모수 추정값

항	추정값	표준 오차
Disk Size[40 GB]	-0.007125177	0.4991763970
Speed[1.5 GHz]	0.214880936	0.5265848733
Battery Life[4 Hrs]	0.052925888	0.4895109018
Price[$1500]	-0.978517706	0.7062009731
Price[$1200]	0.076206148	0.7442825255

AICc	43.487115
BIC	13.884322
-2*로그 우도	3.4871145
-2*Firth 로그 우도	-2.203446

그래디언트에 수렴됨
Firth 편향 조정 추정값

10명의 응답자에게 제공할 최종 선택 설계를 구성하려면 모수 추정값의 사전 평균 및 분산이 필요합니다 . 그림 18.7 의 분석은 모수 평균 추정값 (" 추정값 ") 과 표준 오차 추정값 (" 표준 오차 ") 을 제공합니다 . 표준 오차를 표준편차의 사전 추정값으로 처리합니다 . 속성의 분산 추정값을 계산하려면 JMP 테이블을 생성하고 표준 오차 제곱을 계산합니다 .

4. " 모수 추정값 " 보고서를 마우스 오른쪽 버튼으로 클릭하고 **데이터 테이블로 만들기**를 선택합니다 .
5. 새 데이터 테이블에서 **표준 오차** 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **새 계산식 열** > **변환** > **제곱**을 선택합니다 .

   표준 오차 ^2 열이 데이터 테이블에 추가됩니다 . 이 값은 선택 모형 모수에 대한 사전 분산 추정값으로 사용됩니다 .

그림 18.8 마지막 열에 분산 추정값이 추가된 제목 없음 데이터 테이블

	항	추정값	표준 오차	표준 오차^2
1	Disk Size[40 GB]	-0.007125177	0.4991763970	0.2491770753
2	Speed[1.5 GHz]	0.214880936	0.5265848733	0.2772916288
3	Battery Life[4 Hrs]	0.052925888	0.4895109018	0.2396209229
4	Price[$1500]	-0.978517706	0.7062009731	0.4987198144
5	Price[$1200]	0.076206148	0.7442825255	0.5539564778

참고 : 여기서 " 제목 없음 " 데이터 테이블을 닫지 마십시오 .

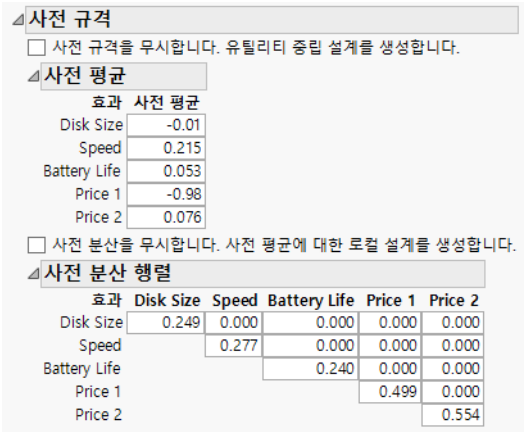
사전 정보를 사용하여 최종 선택 실험 설계

이 섹션에서는 파일럿 랩톱 연구를 통해 얻은 사전 정보를 사용하여 최종 설계를 구성합니다 . 최종 설계는 10 명의 참가자 집합에 제공됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Laptop Factors.jmp 를 엽니다 .
2. **DOE > 소비자 연구 > 선택 설계**를 선택합니다 .
3. " 선택 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .
4. **계속**을 클릭합니다 .
5. " 제목 없음 " 테이블에서 **추정값** 열의 값을 " 선택 설계 " 창의 " 사전 평균 " 개요에 입력합니다 (그림 18.9 참조).

   " 제목 없음 " 테이블에서 전체 열을 복사한 다음 , " 사전 평균 " 개요의 **사전 평균** 아래에 있는 Disk Size 텍스트 상자를 클릭하고 마우스 오른쪽 버튼을 클릭한 후 " 붙여넣기 " 를 선택할 수 있습니다 .
6. " 제목 없음 " 테이블에서 **표준 오차 ^2** 열의 값을 " 선택 설계 " 창에 있는 " 사전 분산 행렬 " 개요의 주대각 항목에 입력합니다 (그림 18.9 참조). 이 값은 소수점 이하 세 자리로 반올림하여 하나씩 입력합니다 .

그림 18.9 파일럿 연구의 사전 평균 및 분산 정보



사전 규격

☐ 사전 규격을 무시합니다. 유틸리티 중립 설계를 생성합니다.

사전 평균

효과	사전 평균
Disk Size	-0.01
Speed	0.215
Battery Life	0.053
Price 1	-0.98
Price 2	0.076

☐ 사전 분산을 무시합니다. 사전 평균에 대한 로컬 설계를 생성합니다.

사전 분산 행렬

효과	Disk Size	Speed	Battery Life	Price 1	Price 2
Disk Size	0.249	0.000	0.000	0.000	0.000
Speed		0.277	0.000	0.000	0.000
Battery Life			0.240	0.000	0.000
Price 1				0.499	0.000
Price 2					0.554

7. " 설계 생성 " 패널에서 **설문 조사 수**에 2 를 입력하고 **설문 조사별 기대 응답자 수**에 5 를 입력합니다 .

   이렇게 하면 총 10 명의 응답자에 대한 설문 조사 도구가 제공되고 두 가지 서로 다른 프로파일 집합이 허용됩니다 .
8. **설계 생성**을 클릭합니다 .
9. **테이블 생성**을 클릭합니다 .

   설계 테이블에 160 개의 행이 포함됩니다 . 10 명의 연구 응답자에 대해 각각 16 개의 행이 있습니다 . 각 응답자는 각각 2 개의 프로파일로 구성된 8 개의 선택 집합을 갖습니다 . 2 개의 설문 조사가 각각 5 명의 응답자에게 제공됩니다 . 2(프로파일 수) * 8(선택 집합 수) * 2(설문 조사 수) * 5(응답자 수) = 160 계산을 통해 160 개 행이 생성됩니다 .

   이제 최종 설계를 10 명의 응답자에게 제공할 수 있습니다 .

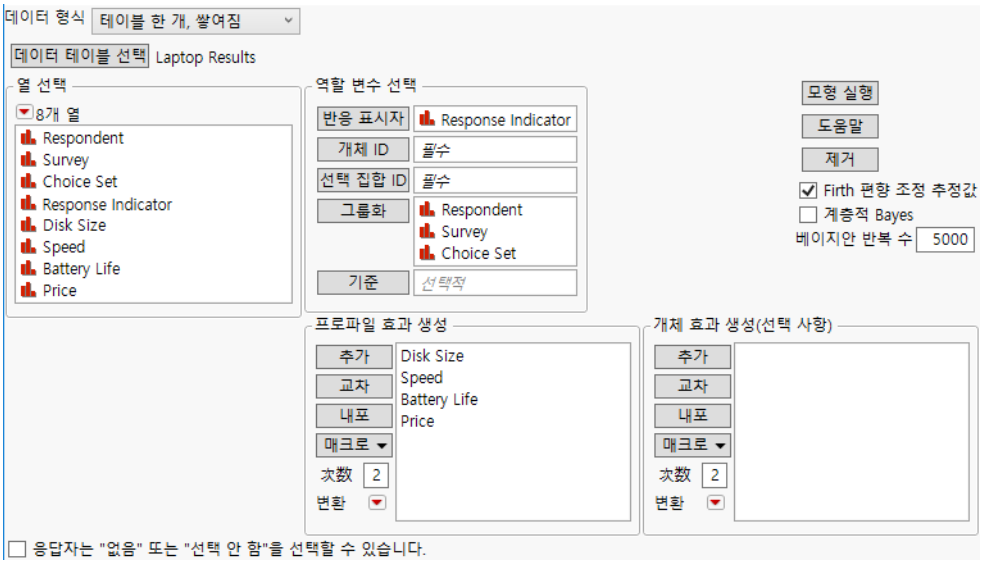
설계 실행 및 결과 분석

이 섹션에서는 최종 설계를 통해 얻은 결과를 분석합니다. 특히 랩톱의 가격이나 다른 특성을 바꾸는 것이 잠재적 구매자가 인식하는 만족도에 어떻게 영향을 미치는지 알고 싶습니다. 이 만족도를 랩톱 속성의 **유틸리티 값**이라고 합니다.

유의한 속성 판별

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Laptop Results.jmp 를 엽니다 .
2. **Choice** 스크립트를 실행합니다 .

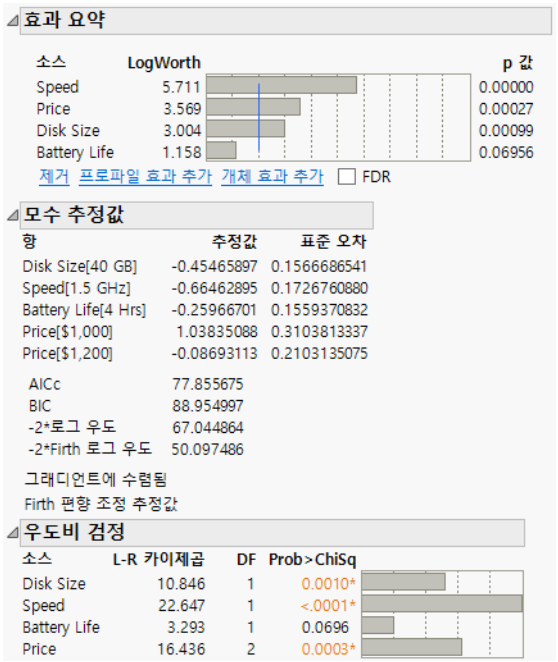
그림 18.10 선택 모형 시작 창



여러 설문 조사 및 응답자가 있으므로 Respondent, Survey 및 Choice Set라는 세 가지 그룹화 변수가 있습니다.

3. **모형 실행**을 클릭합니다.

그림 18.11 최종 랩톱 설계의 초기 분석



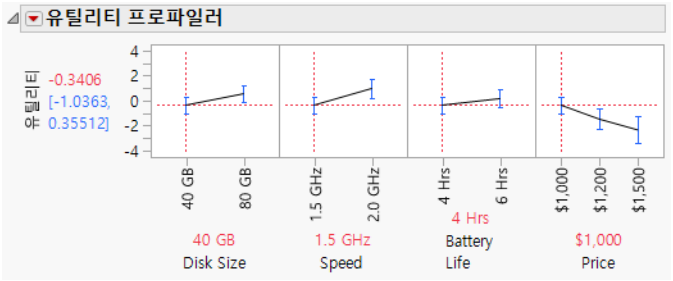
" 효과 요약 " 및 " 가능도비 검정 " 개요에서는 Disk Size, Speed 및 Price 가 0.05 수준에서 유의하고 "Battery Life" 는 아주 약간 유의함을 나타냅니다 .

단위 비용 및 기회 비용 찾기

이제 프로파일러를 사용하여 유틸리티 값을 보고 랩톱 속성이 변경될 때 이 값이 어떻게 변하는지 확인합니다 .

1. " 선택 모형 " 의 빨간색 삼각형을 클릭하고 **유틸리티 프로파일러**를 선택합니다 .

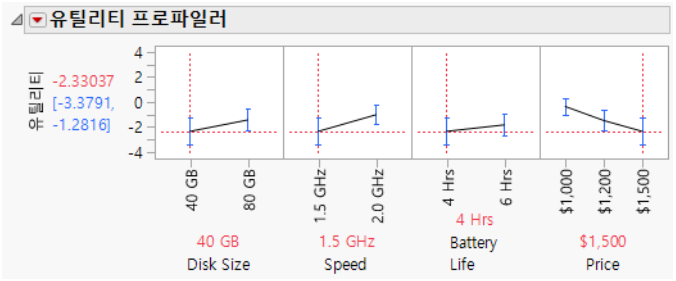
그림 18.12 Price = $1000 의 유틸리티 프로파일러



각 속성 값이 가장 낮은 값으로 설정된 경우 " 유틸리티 " 값은 –0.3406 입니다 . 가장 먼저 단위 유틸리티 비용을 확인하려고 합니다 .

2. Price 슬라이더를 $1,500 로 움직입니다 .

그림 18.13 Price = $1500 의 유틸리티 프로파일러

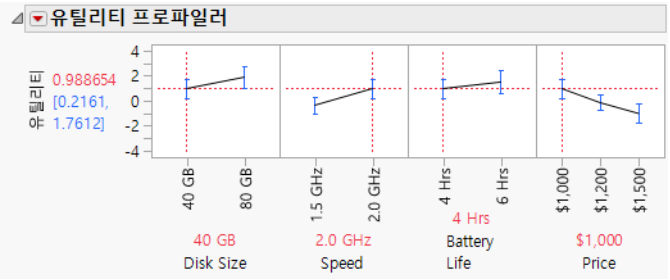


Price 가 $1,000 에서 $1,500 로 변경되면 " 유틸리티 " 가 –0.3406 에서 –2.3303 으로 변합니다 . 즉 , 랩톱 가격을 500 달러 인상하면 유틸리티 (만족도) 가 약 2 단위 낮아집니다 . 따라서 단위 유틸리티 비용을 약 250 달러로 추정할 수 있습니다 .

이제 이 단위 유틸리티 비용 추정값으로 다른 속성을 변경하고 , 유틸리티 변화를 확인하고 , 속성 변경과 관련된 금전적 가치 근사값을 찾을 수 있습니다 . 예를 들어 가장 유의한 속성은 Speed 입니다 (그림 18.11 참조).

3. " 유틸리티 프로파일러 " 에서 "Price" 를 다시 $1,000(가장 낮은 값) 로 설정하고 "Speed" 를 2.0GHz(더 높은 값) 로 변경합니다 .

그림 18.14 Speed 값을 증가할 때의 유틸리티 값



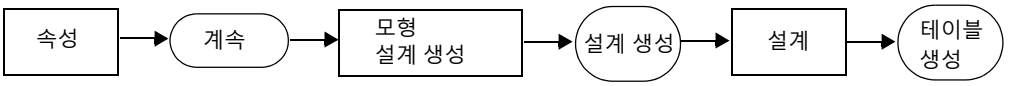
" 유틸리티 " 값은 그림 18.12 에 표시된 원래 값 –0.3406 에서 0.9886 으로 바뀌어 총 1.3292 단위의 변화를 보입니다 . 유틸리티 비용 추정값 또는 250 달러를 사용하면 1.5GHz 랩톱에 비해 2.0GHz 랩톱의 가격 인상은 1.3292*$250.00 = 332.30 달러로 계산될 수 있습니다 . 이것은 선택 연구에서 제조업체가 이 랩톱 속성의 가격을 결정하는 기준으로 사용할 수 있는 달러 값입니다 . 다른 속성에 대해서도 유사한 계산을 할 수 있습니다 .

선택 설계 창

"선택 설계" 창은 속성 선호도를 모델링하기 위한 선택 설계를 구성하는 단계를 안내합니다. 가정된 모형과 사전 정보를 지정할 수 있습니다.

설계 단계를 진행하는 동안 "선택 설계" 창이 업데이트됩니다. 개요는 개요를 업데이트하는 버튼을 클릭할 때 구분되어 나타나며 그림 18.15의 흐름을 따릅니다.

그림 18.15 선택 설계 흐름

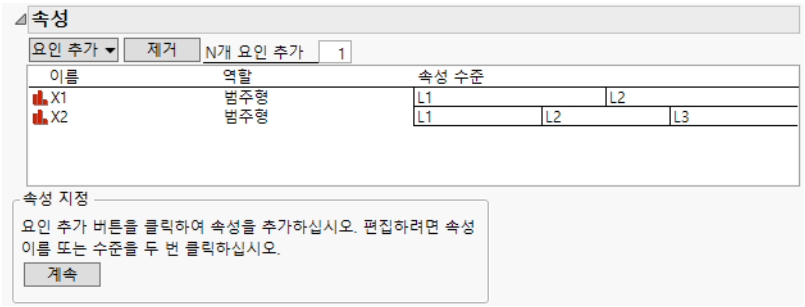


속성

선택 설계의 속성은 범주형만 가능합니다.

팁 : " 속성 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **요인 저장**을 선택하는 것이 좋습니다 . 이렇게 하면 속성 이름 , 역할 및 수준을 데이터 테이블에 저장하여 나중에 다시 불러와 사용할 수 있습니다 .

그림 18.16 속성 개요



" 속성 " 개요에는 다음과 같은 버튼이 포함되어 있습니다 .

요인 추가 선택한 수준 수를 사용하는 속성을 추가합니다 .

제거 선택한 속성을 제거합니다 .

N 개 요인 추가 여러 속성을 추가합니다 . 추가할 속성 수를 입력하고 **요인 추가**를 클릭한 후 수준 수를 선택합니다 . 수준 수가 다른 여러 속성을 추가하려면 **N 개 요인 추가**를 반복합니다 .

" 속성 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

이름 속성의 이름입니다 . X1, X2 등의 기본 이름이 속성에 지정됩니다 . 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 속성의 설계 역할을 " 범주형 " 으로 지정합니다 .

속성 수준 속성 이름 또는 설명입니다 . " 속성 수준 " 을 삽입하려면 기본 수준을 클릭한 후 원하는 이름을 입력합니다 .

속성 개요 편집

- 속성 이름을 편집하려면 해당 속성 이름을 두 번 클릭합니다 .
- 속성 수준을 편집하려면 해당 수준을 클릭합니다 .

속성 열 특성

각 속성에 대해 JMP 는 선택 플랫폼에서 구성된 데이터 테이블에 " 값 순서 " 열 특성을 저장합니다 . " 값 순서 " 열 특성은 " 속성 " 개요에 지정된 순서대로 보고서에 수준이 나타나도록 지정합니다 . 자세한 내용은 " 열 특성 " 부록의 " 값 순서 "(755 페이지) 에서 확인하십시오 .

모형

모형 개요는 다음 두 부분으로 구성됩니다 .

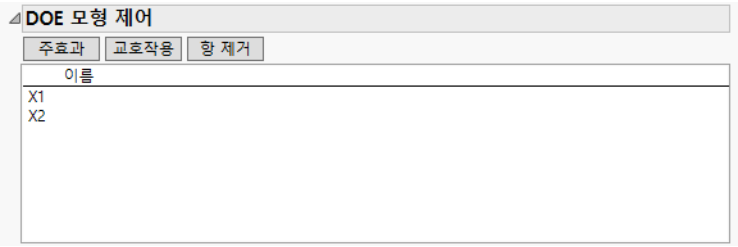
- 가정된 모형 , 즉 추정할 모든 효과를 포함하는 모형을 지정할 수 있습니다 . 자세한 내용은 "DOE 모형 제어 "(522 페이지) 에서 확인하십시오 .

- 속성 수준에 대한 사전 지식을 지정하여 더 적합한 설계를 생성할 수 있습니다. 자세한 내용은 " 사전 규격 "(522 페이지) 에서 확인하십시오 .

DOE 모형 제어

"DOE 모형 제어" 개요에서 가정된 모형을 지정합니다. 모든 주효과는 기본적으로 포함됩니다. 모든 이원 교호작용을 추가하려면 "교호작용" 버튼을 클릭합니다.

그림 18.17 DOE 모형 제어 개요



설계 테이블을 구성할 때 JMP에서 "선택" 스크립트를 데이터 테이블에 저장합니다. "선택" 스크립트는 "DOE 모형 제어" 개요에서 선택한 모형이 포함된 선택 플랫폼을 시작합니다.

"DOE 모형 제어 " 개요에는 다음과 같은 버튼이 포함되어 있습니다 .

주효과 모형의 모든 속성에 대한 주효과를 추가합니다 .

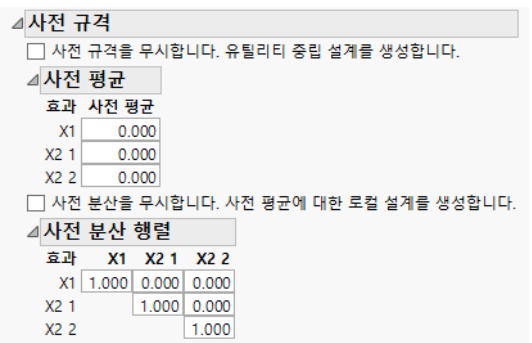
교호작용 모든 2 차 교호작용을 추가합니다 . 일부 교호작용을 포함하지 않으려면 제거할 교호작용을 선택하고 **항 제거**를 클릭합니다 .

항 제거 선택한 효과를 제거합니다 .

사전 규격

모형 모수에 대해 다변량 정규 사전 분포의 규격을 입력합니다. "사전 평균" 개요에 사전 분포의 평균을 입력하고 "사전 분산 행렬" 개요에 공분산 행렬을 입력합니다.

그림 18.18 사전 규격 개요



다음 옵션을 사용하여 사전 규격을 무시할 수 있습니다 .

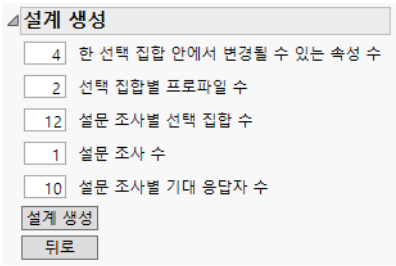
사전 규격을 무시합니다 . 유틸리티 중립 설계를 생성합니다 . 사전 평균을 0 으로 설정하고 로컬 D- 최적 설계를 생성합니다 . 이 설계를 유틸리티 중립 설계라고 합니다 . 자세한 내용은 Huber and Zwerina(1996) 에서 확인하십시오 .

사전 분산을 무시합니다 . 사전 평균에 대한 로컬 설계를 생성합니다 . 로컬 D- 최적 설계를 생성합니다 . 로컬 설계는 사전 평균을 고려하지만 공분산 행렬은 무시합니다 . 자세한 내용은 Huber and Zwerina(1996) 에서 확인하십시오 .

설계 생성

"설계 생성" 패널에서 설계 구조를 정의하는 규격을 입력합니다.

그림 18.19 커피 예에 대한 설계 생성 패널



참고 : 그림 18.19 는 커피 예에서 가져왔습니다 . 자세한 내용은 " 선택 설계의 예 "(510 페이지) 에서 확인하십시오 .

다음 각 항목에 대한 수를 입력합니다 .

한 선택 집합 안에서 변경될 수 있는 속성 수 총 속성 수보다 작거나 같은 수를 입력합니다 . 이 값은 주로 총 속성 수로 설정됩니다 . 그러나 여러 속성을 비교하는 경우 응답자의 선택 과정을 간소화하려면 총 속성 수보다 작은 값으로 설정하십시오 .

선택 집합별 프로파일 수 응답자가 선호도를 지정할 때 선택해야 하는 프로파일의 수를 입력합니다 .

설문 조사별 선택 집합 수 각 응답자로부터 얻으려는 선호도의 수를 입력합니다 .

설문 조사 수 고유한 선택 집합 모음의 수를 입력합니다. 여러 응답자에게 설문 조사를 제공하려는 경우에 유용합니다 .

설문 조사별 기대 응답자 수 총 응답자 수를 총 설문 조사 수로 나눈 값을 입력합니다 .

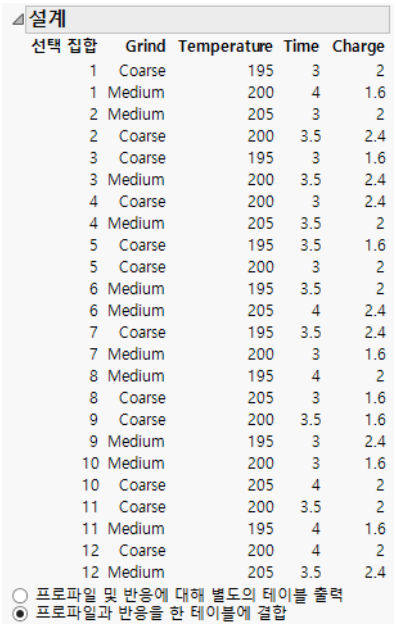
설계 생성

"설계 생성" 개요를 완료한 후 **설계 생성**을 클릭하여 설계를 생성합니다. 설계가 "설계" 개요에 나타납니다.

설계

"설계" 개요에서는 선택 사항에 따라 최적 설계를 위한 런을 보여 줍니다. 설계를 검토하여 요구 사항을 충족하는지 확인합니다.

그림 18.20 커피 예에 대한 설계 개요



설계

선택 집합	Grind	Temperature	Time	Charge
1	Coarse	195	3	2
1	Medium	200	4	1.6
2	Medium	205	3	2
2	Coarse	200	3.5	2.4
3	Coarse	195	3	1.6
3	Medium	200	3.5	2.4
4	Coarse	200	3	2.4
4	Medium	205	3.5	2
5	Coarse	195	3.5	1.6
5	Coarse	200	3	2
6	Medium	195	3.5	2
6	Medium	205	4	2.4
7	Coarse	195	3.5	2.4
7	Medium	200	3	1.6
8	Medium	195	4	2
8	Coarse	205	3	1.6
9	Coarse	200	3.5	1.6
9	Medium	195	3	2.4
10	Medium	200	3	1.6
10	Coarse	205	4	2
11	Coarse	200	3.5	2
11	Medium	195	4	1.6
12	Coarse	200	4	2
12	Medium	205	3.5	2.4

○ 프로파일 및 반응에 대해 별도의 테이블 출력
◉ 프로파일과 반응을 한 테이블에 결합

참고 : 그림 18.20 는 커피 예에서 가져왔습니다 . 자세한 내용은 " 선택 설계의 예 "(510 페이지) 에서 확인하십시오 .

참고 : 최적 설계를 찾기 위한 알고리즘은 랜덤 시작 설계를 기반으로 합니다 . 이 때문에 생성된 설계가 고유하지 않습니다 . **뒤로** 및 **설계 생성** 버튼을 반복해서 클릭하면 설계 알고리즘이 또 다른 설계를 생성합니다 .

출력 옵션

다음 출력 옵션 중 하나를 선택합니다 .

프로파일 및 반응에 대해 별도의 테이블 출력 다음과 같은 두 개의 데이터 테이블을 표시합니다 .

- " 선택 프로파일 " 테이블에는 **설문 조사** 및 **선택 집합** 열로 식별되는 각 행에 프로파일이 나열됩니다 . 선택 집합 내에서 프로파일은 **선택 ID** 로 식별됩니다 . 이 테이블은 설문 조사 도구를 구성하는 데 유용합니다 .

– " 선택 런 " 테이블에는 응답자 선호도를 입력할 수 있는 빈 **반응** 열이 제공됩니다 . 각 행은 단일 선택 집합에 해당합니다 . 행은 **응답자** , **설문 조사** 및 **선택 집합**별로 정렬됩니다 . 다음 열에 선택 집합 ID 가 제공되고 그 뒤에 반응 열이 있습니다 . 응답자의 선호도에 대한 선택 집합 ID 를 " 반응 " 열에 입력합니다 .

프로파일과 반응을 한 테이블에 결합 응답자 선호도를 입력할 수 있는 빈 **반응 표시자** 열이 포함된 단일 " 선택 프로파일 " 테이블을 제공합니다 . 각 행은 단일 프로파일에 해당합니다 . 테이블은 **응답자** , **설문 조사** 및 **선택 집합**별로 정렬됩니다 . 응답자가 선호하는 프로파일에 1(또는 0 이 아닌 다른 수치 표시자) 을 입력하고 선택 집합의 다른 프로파일에 0 표시자를 입력합니다 .

참고 : **반응 표시자** 열에 입력하는 값은 숫자여야 합니다 .

테이블 생성

테이블 생성을 클릭하여 "출력 옵션"(524 페이지)에서 선택한 하나 이상의 테이블을 구성합니다 . "선택 프로파일" 테이블의 테이블 패널에 **선택** 스크립트가 있습니다 . 이 스크립트를 실행한 후 **모형 실행**을 클릭하여 실험 결과를 분석할 수 있습니다 .

선택 모형 보고서에 대한 자세한 내용은 Consumer Research 의 에서 확인하십시오 .

선택 설계 옵션

요인 저장 " 요인 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 각 요인의 열에는 수준이 포함됩니다 . 다른 정보는 열 특성으로 저장됩니다 . 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다 . 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다 . 그런 다음 " 설계 역할 " 영역에서 적절한 역할을 선택합니다 .

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 불러옵니다 .

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다 . 이러한 작업에는 다음 중 하나 이상이 포함됩니다 .

– 설계 생성을 위한 검색 알고리즘 초기화
– 설계 생성을 위한 런 순서 랜덤화
– 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 생성된 설계의 난수 시드값을 입력합니다 .

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

반응 시뮬레이션 선택 모형 시뮬레이션 계산식을 포함하는 열과 반응 확률을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다 .

- 시뮬레이션된 선택 모형 반응 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
- 모형 효과에 대한 주변 유틸리티 값을 설정할 수 있는 " 모형 " 창이 나타납니다 .
- **선택 모형 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다 . 이 스크립트는 값을 다시 시뮬레이션하거나 주변 유틸리티를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 적용 " 을 클릭하면 시뮬레이션된 선택 모형 반응 값의 계산식이 " 선택 모형 시뮬레이션 " 열에서 업데이트됩니다 . " 적용 " 을 다시 클릭하면 이 열의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . DOE 의 예에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오 .

시작 수 설계 구성에 사용되는 랜덤 시작 횟수를 지정할 수 있습니다 . 자세한 내용은 " 베이지안 D- 최적 및 설계 구성 "(526 페이지) 에서 확인하십시오 .

고급 옵션 선택 설계에는 사용할 수 없습니다 .

스크립트를 스크립트 창에 저장 " 선택 설계 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

기술 상세 정보

베이지안 D- 최적 및 설계 구성

베이지안 D- 최적 기준은 사전 분포와 관련한 다항 로짓 모형에서 모수 추정량 최대 가능도의 정보 행렬 행렬식에 대한 로그 기대값입니다 . 선택 설계 플랫폼은 사전 확률 분포를 나타내는 모수 벡터 표본과 관련하여 이 기대를 최대화합니다 . 자세한 내용은 Kessels et al. (2011) 에서 확인하십시오 .

부분 프로파일 설계의 경우 JMP 는 다음과 같은 2 단계 설계 알고리즘을 사용합니다 .

1. 각 선택 집합의 상수 속성은 **속성 균형** 방식을 사용하여 결정됩니다 .
2. 비상수 속성의 수준은 베이지안 D- 최적을 사용하여 결정됩니다 .

속성 균형은 알고리즘이 전체 설계에서 각 속성이 상수로 고정되는 횟수의 균형을 맞추려고 시도하는 것을 의미합니다 . 둘 이상의 속성이 상수로 고정되면 알고리즘은 설계에서 상수로 고정된 속성 쌍의 발생이 균형을 이루도록 합니다 .

비상수 속성의 수준은 베이지안 D-최적 기준을 최적화하도록 결정됩니다. 랜덤 시작 설계가 발견됩니다. 그러면 좌표 교환 알고리즘을 사용하여 비상수 속성의 수준이 생성되고 베이지안 D-최적 기준이 최적화될 때까지 평가됩니다. 다변량 정규 사전 분포에 대한 적분을 포함하는 계산은 Gotwalt et al. (2009)에 설명된 구적법을 사용합니다.

참고 : 베이지안 D- 최적 기준으로 인해 일부 비상수 속성의 수준이 동일한 선택 집합이 생성될 수 있습니다. 프로파일 내에서 비상수 수준을 변경하여 모든 프로파일의 확률이 매우 높거나 낮은 무정보적 선택 집합이 생성되는 경우가 여기에 해당합니다.

유틸리티 중립 설계 및 로컬 D- 최적 설계

선택 설계 플랫폼을 사용하는 경우 사전 평균을 0으로 설정하여 유틸리티 중립 설계를 생성할 수 있습니다. 유틸리티 중립 설계에서는 선택 집합 내의 모든 선택 항목에 대한 확률이 동등합니다. 자세한 내용은 Huber and Zwerina(1996)에서 확인하십시오.

로컬 D-최적 설계도 생성할 수 있습니다. 로컬 D-최적 설계는 사전 평균을 고려하지만 사전 공분산 행렬의 정보는 포함하지 않습니다. 자세한 내용은 Huber and Zwerina(1996)에서 확인하십시오.

19 장

최대차이 설계

최선 - 최악 항목 선택을 위한 설계 생성

최대차이(최대 차이 척도화) 연구는 표준 선호도 척도를 사용하여 평가 항목의 상대 중요도를 결정하는 연구의 대안입니다. 최대차이 연구에서 응답자는 소수의 선택 집합 중에서 가장 선호하는 옵션과 가장 선호하지 않는 옵션만 보고합니다. 이렇게 하면 응답자가 선호도를 기준으로 옵션 순위를 지정하게 되고, 이로 인해 표준 선호도 척도를 사용하는 것보다 더 확증적인 순위를 얻는 경우가 많습니다.

최대차이 연구 중에 응답자에게 제공할 수 있는 한 요인에 대한 선택 집합으로 구성된 설계를 생성해야 할 때 최대차이 플랫폼을 사용합니다. 연구를 수행한 후 최대차이 분석 플랫폼을 사용하여 데이터를 분석합니다.

그림 19.1 최대차이 설계 테이블

	개체	선택 집합	Candy	선택
1	1	1	Reese's Cups	•
2	1	1	Hershey Bar	•
3	1	1	Snickers	•
4	1	1	Butterfinger	•
5	1	2	Butterfinger	•
6	1	2	Heath Bars	•
7	1	2	Plain M&Ms	•
8	1	2	Snickers	•
9	1	3	Plain M&Ms	•
10	1	3	Snickers	•
11	1	3	Peanut M&Ms	•
12	1	3	Hershey Bar	•
13	1	4	Snickers	•
14	1	4	Peanut M&Ms	•
15	1	4	Reese's Cups	•
16	1	4	Heath Bars	•

목차

최대차이 설계 플랫폼 개요 531
최대차이 설계의 예 531
최대차이 설계 시작 창 535
최대차이 창 535
설계 옵션 개요 535
설계 개요 536
쌍별 따름 행렬 536
테이블 생성 536
최대차이 옵션 536

최대차이 설계 플랫폼 개요

선택 집합은 응답자가 가장 선호하는 항목(최선 항목)과 가장 선호하지 않는 항목(최악 항목)을 선택해야 하는 항목 모음입니다. 최대차이 플랫폼을 사용하면 선택 집합에 나타나는 항목 수와 응답자에게 제공할 선택 집합 수를 지정할 수 있습니다.

최대차이 플랫폼은 단일 설문 조사를 제공합니다. 응답자가 여러 명이면 동일한 설문 조사를 모든 응답자에게 제공하거나 각 응답자에 대한 설문 조사를 구성할 수 있습니다.

최대차이 설계는 균형 불완비 블록 설계의 따름 행렬을 가능한 한 근사하게 매칭하도록 구성됩니다. 이것은 설계에서 각 항목 쌍의 발생 횟수가 동일하다는 의미입니다. 선택 집합의 수가 선택 집합에서 지정된 수의 항목에 대해 균형 불완비 블록을 구성할 수 있을 정도이면 균형 불완비 블록 설계가 구성됩니다.

항목 또는 프로파일은 처리 요인의 수준으로 간주됩니다. 선택 집합은 블록으로 간주됩니다. 최대차이 연구에서 블록 크기는 처리 수보다 작습니다.

최대차이 설계의 예

회사의 사무용품 구매자로서 명절 파티를 위해 사탕을 사야 합니다. 먼저 사람들이 좋아하는 사탕 종류를 결정하려고 합니다. 이를 확인하기 위해 최대차이 연구를 수행합니다.

동료 5명을 무작위로 선택하여 7개 종류의 사탕을 평가하도록 요청합니다. 이전 연구의 경험을 바탕으로 평가자가 7개 종류의 항목을 선호도 순으로 순위를 지정하기가 어렵다는 것을 알고 있습니다. 대신 크기가 4인 선택 집합으로 구성된 설계를 생성합니다. 관리 가능한 연구를 위해 설문 조사를 최대차이 연구로 구성합니다. 이 연구에서는 각 동료에게 7개의 각 선택 집합에서 가장 선호하는 사탕과 가장 선호하지 않는 사탕을 지정하도록 요청합니다. 이렇게 선택하면 균형 불완비 블록 설계가 생성됩니다. 자세한 내용은 Cochran and Cox(1957)에서 확인하십시오. 각 동료에게 동일한 설문 조사를 제공합니다.

설계 생성

선택 집합에 대한 항목 또는 프로파일이 나열된 테이블을 구성합니다 . 이 예에서는 항목 테이블 Candy Profiles.jmp 가 이미 구성되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Candy Profiles.jmp 를 엽니다 . 이 테이블에는 7 개 사탕 종류가 나열되어 있습니다 .
2. **DOE > 소비자 연구 > 최대차이 설계**를 선택합니다 .
3. " 열 선택 " 목록에서 Candy 를 선택하고 **X, 요인**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 설계 옵션 " 개요에서 다음을 수행합니다 .
   - **선택 집합별 프로파일 수**를 **4** 로 설정합니다 .

– **선택 집합 수**를 **7** 로 설정합니다 .

참고 : 6 단계에서 " 난수 시드값 " 을 설정하면 이 예에 표시된 것과 동일한 결과가 재현됩니다 . 설계를 직접 구성할 때는 이 단계가 필요하지 않습니다 .

6. (선택 사항) " 최대차이 연구 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 . "12345" 를 입력하고 **확인**을 클릭합니다 .
7. **설계 생성**을 클릭합니다 .

   " 설계 " 개요에는 각각 4 개의 사탕 종류로 구성된 선택 집합 7 개가 표시됩니다 . " 쌍별 따름 행렬 " 은 각 사탕 종류 쌍이 함께 나타나는 선택 집합 수의 행렬을 제공합니다 . 행렬의 주대각은 각 사탕 종류가 나타나는 선택 집합의 수입니다 . 이 예의 경우 각 사탕은 4 개의 선택 집합에 나타나고 4 개의 선택 집합 중 2 개에서 서로 다른 종류의 사탕과 함께 나타납니다 .
8. **테이블 생성**을 클릭합니다 .

그림 19.2 사탕 선호도 설문 조사에 대한 설계 테이블 (일부분)

	개체	선택 집합	Candy	선택
1	1	1	Reese's Cups	•
2	1	1	Hershey Bar	•
3	1	1	Snickers	•
4	1	1	Butterfinger	•
5	1	2	Butterfinger	•
6	1	2	Heath Bars	•
7	1	2	Plain M&Ms	•
8	1	2	Snickers	•
9	1	3	Plain M&Ms	•
10	1	3	Snickers	•
11	1	3	Peanut M&Ms	•
12	1	3	Hershey Bar	•
13	1	4	Snickers	•
14	1	4	Peanut M&Ms	•
15	1	4	Reese's Cups	•
16	1	4	Heath Bars	•

설계 테이블에는 선호도 기록을 위한 **선택** 열이 포함되어 있습니다 . 각 선택 집합에 대해 가장 선호하는 사탕에는 1, 가장 선호하지 않는 사탕에는 -1, 다른 두 사탕에는 0 을 각각 기록합니다 .

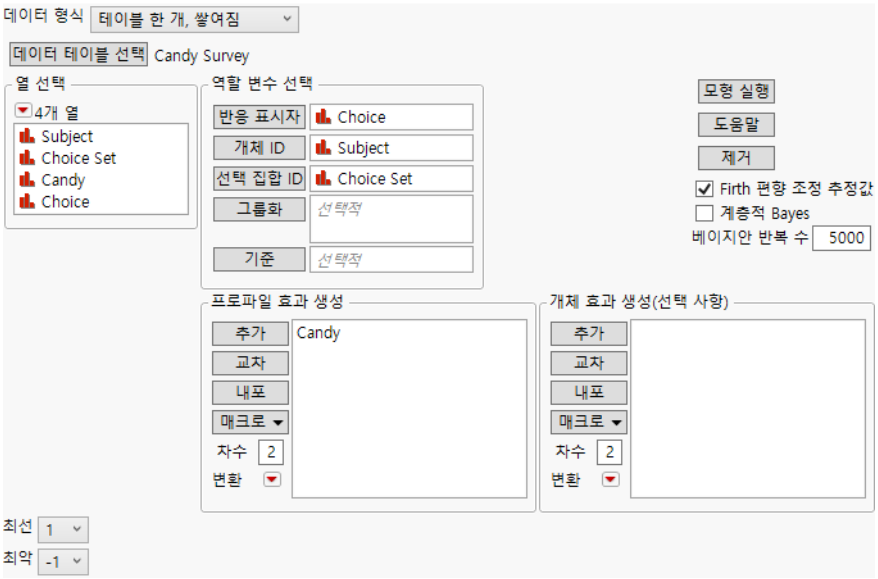
연구 결과 분석

연구를 수행하고 Candy Survey.jmp 에 데이터를 기록합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Candy Survey.jmp 를 엽니다 . 테이블에는 Subject 열에 나열된 5 명의 각 응답자에게 설문 조사를 실시한 결과가 표시됩니다 .
2. **분석 > 소비자 조사 > 최대차이**를 선택합니다 .
3. " 데이터 테이블 선택 " 을 클릭하고 Candy Survey 를 선택한 후 **확인**을 클릭합니다 .
4. 다음과 같이 열 역할을 할당합니다 .

- Choice 를 선택하고 **반응 표시자**를 클릭합니다 .
- Subject 를 선택하고 **개체 ID** 를 클릭합니다 .
- Choice Set 를 선택하고 **선택 집합 ID** 를 클릭합니다 .
- Candy 를 선택하고 " 프로파일 효과 생성 " 패널에서 **추가**를 클릭합니다 .

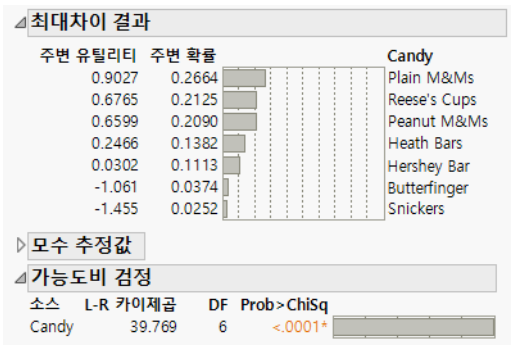
그림 19.3 완료된 최대차이 분석 시작 창



최선 선택을 1 로 지정하고 최악 선택을 -1 로 지정했으므로 시작 창 왼쪽 아래의 " 최선 " 및 " 최악 " 선택을 변경하지 마십시오 .

5. **모형 실행**을 클릭합니다 .

그림 19.4 최대차이 보고서



이 보고서는 Candy 가 유의함을 나타냅니다 . 유틸리티가 가장 높은 세 가지 사탕 종류는 Plain M&Ms, Reese's Cups, Peanut M&Ms 입니다 .

6. " 최대차이 모형 " 의 빨간색 삼각형을 클릭하고 **모든 수준 비교 보고서**를 선택합니다 .

그림 19.5 사탕 종류에 대한 모든 수준 비교 보고서

모든 수준 비교 보고서

차이(행-열) 차이의 표준 오차 Wald p 값	Butterfinger	Heath Bars	Hershey Bar	Peanut M&Ms	Plain M&Ms	Reese's Cups	Snickers
Butterfinger	0	-1.3079	-1.0915	-1.7212	-1.964	-1.7378	0.39343
	0	0.52954	0.52038	0.54692	0.56495	0.54938	0.5059
		0.01964	0.04478	0.0038	0.00162	0.00364	0.44306
Heath Bars	1.30792	0	0.21642	-0.4133	-0.656	-0.4299	1.70135
	0.52954	0	0.47619	0.48741	0.49094	0.46959	0.54758
	0.01964		0.65286	0.40343	0.19185	0.36752	0.0042
Hershey Bar	1.0915	-0.2164	0	-0.6297	-0.8725	-0.6463	1.48493
	0.52038	0.47619	0	0.47406	0.49984	0.49096	0.53808
	0.04478	0.65286		0.19443	0.09149	0.19835	0.00992
Peanut M&Ms	1.72121	0.41329	0.62971	0	-0.2427	-0.0166	2.11463
	0.54692	0.48741	0.47406	0	0.49706	0.49054	0.56901
	0.0038	0.40343	0.19443		0.62896	0.97325	0.00086
Plain M&Ms	1.96396	0.65604	0.87246	0.24275	0	0.22616	2.35738
	0.56495	0.49094	0.49984	0.49706	0	0.49327	0.58562
	0.00162	0.19185	0.09149	0.62896		0.65002	0.00037
Reese's Cups	1.7378	0.42988	0.6463	0.01659	-0.2262	0	2.13123
	0.54938	0.46959	0.49096	0.49054	0.49327	0	0.5714
	0.00364	0.36752	0.19835	0.97325	0.65002		0.00083
Snickers	-0.3934	-1.7013	-1.4849	-2.1146	-2.3574	-2.1312	0
	0.5059	0.54758	0.53808	0.56901	0.58562	0.5714	0
	0.44306	0.0042	0.00992	0.00086	0.00037	0.00083	

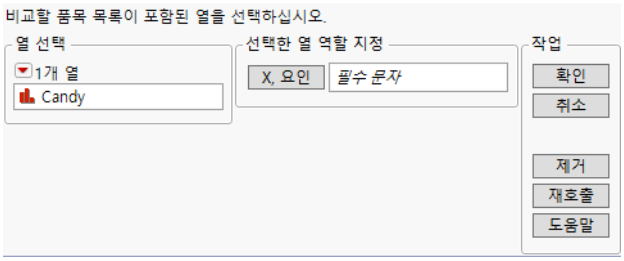
비교 보고서는 유틸리티 측면에서 유의한 차이가 있는 사탕 종류의 쌍을 나타냅니다 . 각 셀의 세 번째 항목은 행 항목의 유틸리티에서 열 항목의 유틸리티를 뺀 값으로 정의된 차이의 p 값입니다 . p 값의 색상 강도는 차이가 얼마나 유의한지를 나타냅니다 . 파란색 또는 빨간색 음영은 차이 (행 - 열) 가 음수인지 아니면 양수인지를 나타냅니다 . p 값은 다중 비교 오차율을 제어하기 위해 조정되지 않으며 지침으로만 사용해야 합니다 . 모든 수준 비교 보고서에 대한 자세한 내용은 Consumer Research 의 에서 확인하십시오 .

최대차이 설계 시작 창

최대차이 설계 플랫폼을 사용하려면 시작 데이터 테이블이 필요합니다 . 시작 데이터 테이블에는 평가를 위해 응답자에게 제공할 항목을 나열하는 문자 데이터 열이 포함되어야 합니다 .

JMP 에서 시작 데이터 테이블을 활성화하고 **DOE > 소비자 연구 > 최대차이 설계**를 선택합니다 . 활성 데이터 테이블이 없으면 시작 데이터 테이블을 찾기 위한 대화상자가 나타납니다 .

그림 19.6 Candy Profiles.jmp 를 사용한 최대차이 시작 창

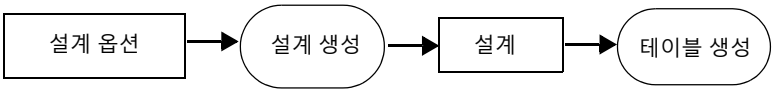


X, 요인 응답자가 평가할 항목이 포함된 문자 열입니다 .

최대차이 창

설계 단계를 진행하는 동안 "최대차이 연구" 창이 업데이트됩니다. 개요는 개요를 업데이트하는 버튼을 클릭할 때 구분되어 나타나며 그림 19.7의 흐름을 따릅니다.

그림 19.7 최대차이 설계 흐름



"설계 옵션" 개요를 보여 주는 "최대차이 연구" 창이 열립니다. "설계 생성"을 클릭하면 "설계" 개요가 나타납니다. 설계 테이블을 구성하려면 "테이블 생성"을 클릭합니다.

설계 옵션 개요

다음을 지정합니다 .

선택 집합별 프로파일 수 각 선택 집합에 포함할 항목 수입니다 .

선택 집합 수 응답자가 받아서 평가할 항목 집합의 총 수입니다 .

설계 개요

"설계" 개요는 연속된 양의 정수를 사용하여 선택 집합을 식별합니다. 각 선택 집합을 구성하는 항목이 나열됩니다.

쌍별 따름 행렬

"쌍별 따름 행렬" 개요는 각 요인 수준 쌍이 함께 나타나는 선택 집합 수의 행렬을 제공합니다. 행렬의 주대각은 각 요인 수준이 나타나는 선택 집합의 수입니다.

테이블 생성

" 테이블 생성 " 버튼은 설계 테이블을 생성합니다 . 설계 테이블은 다음과 같은 네 개의 열로 구성됩니다 .

개체 처음에는 1 로 채워져 있습니다 . 이 값을 응답자를 나타내는 적절한 식별자로 바꿉니다 .

팁 : 응답자 식별자를 쉽게 추가하는 방법은 JMP 사용의 에서 확인하십시오 . 식별자를 이름으로 변경하려면 " 열 > 재코딩 " 을 사용합니다 .

선택 집합 각 선택 집합에 대한 지정자입니다 .

요인 지정한 요인의 수준입니다 . 이 수준은 선택 집합의 항목입니다 .

선택 결과를 입력할 수 있는 열입니다 . 숫자 값을 사용하여 최선 선택 , 최악 선택 및 중간 선택을 나타냅니다 . 일반적으로 1, -1, 0 값이 사용되고 이 값은 최대차이 분석 플랫폼 (" 분석 > 소비자 조사 > 최대차이 ") 에서 분석할 때 필요합니다 .

최대차이 옵션

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다 . 이러한 작업에는 다음 중 하나 이상이 포함됩니다 .

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다 .

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

반응 시뮬레이션 반응 값을 시뮬레이션하려면 이 옵션을 선택합니다 . " 테이블 생성 " 을 클릭하면 설계 테이블과 함께 " 선택 모형 시뮬레이션 " 창이 열리고 " 확률 " 및 " 선택 모형 시뮬레

이션 " 열이 설계 테이블에 추가됩니다 . " 선택 모형 시뮬레이션 " 계산식 열에는 확률 반응이 포함되고 " 확률 " 계산식 열에는 확률이 포함됩니다 .

주의 : 반응 시뮬레이션에 사용된 모형을 변경하려면 " 선택 모형 시뮬레이션 " 창의 " 주변 유틸리티 " 열에 마지막 수준을 제외한 모든 요인 수준의 값을 입력합니다 . 주변 유틸리티의 합이 0 이어야 하므로 마지막 수준의 주변 유틸리티를 편집할 수 없습니다 . 나머지 주변 유틸리티 값을 지정하면 요인의 마지막 수준이 적절하게 조정됩니다 .

시작 수 설계 구성에 사용되는 랜덤 시작 횟수입니다 . 이 값은 기본적으로 10 으로 설정됩니다 .

고급 옵션 최대차이 설계에는 사용할 수 없습니다 .

스크립트를 스크립트 창에 저장 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

20 장

JMP PRO 덮기 배열
성분 교호작용으로 인한 고장 감지

덮기 배열은 성분 또는 부시스템 간의 교호작용으로 인해 고장이 발생하는 결정적 시스템을 테스트하는 데 사용됩니다. 설계 목표는 교호작용으로 인해 시스템 고장이 유발되는지 여부를 알아내는 것입니다. 응용 분야에는 소프트웨어, 회로 및 네트워크 설계가 포함됩니다.

결정적 테스트가 수행되므로 설계의 주안점은 필요한 모든 교호작용을 포함해야 한다는 것입니다. 덮기 배열 플랫폼은 고효율의 덮기 배열을 생성합니다. 테스트 프로토콜에 실현 가능하지 않은 요인 수준 조합을 제외할 수도 있습니다.

그림 20.1 강도 3 덮기 배열

설계

런	X1	X2	X3	X4	X5
1	L1	L1	L1	L1	L1
2	L1	L1	L1	L2	L2
3	L2	L2	L2	L2	L2
4	L2	L1	L2	L2	L1
5	L2	L2	L2	L1	L2
6	L2	L2	L1	L1	L1
7	L2	L1	L1	L1	L2
8	L1	L2	L2	L1	L1
9	L1	L2	L1	L2	L2
10	L1	L2	L2	L2	L1
11	L1	L1	L2	L1	L2
12	L2	L2	L1	L2	L1

목차

덮기 배열 개요 541
요인 수준 제한이 없는 덮기 배열의 예 542
설계 생성 543
실험 데이터 분석 545
요인 수준 제한이 있는 덮기 배열의 예 547
설계 생성 547
실험 데이터 분석 552
덮기 배열 창 553
요인 553
요인 수준 조합 제한 555
설계 557
지표 559
출력 옵션 560
덮기 배열 데이터 테이블 560
덮기 배열 옵션 561
덮기 배열에 대한 기술 상세 정보 562
최적화 알고리즘 562
지표 계산식 563

JMP PRO 덮기 배열 개요

덮기 배열을 사용하면 성분 또는 부시스템 간의 교호작용으로 인해 고장이 발생하는 시스템을 테스트할 수 있습니다 . 덮기 배열은 다음 조건에 해당되는 소프트웨어 , 회로 및 네트워크 설계 등의 분야에 주로 사용됩니다 .

- 대개 테스트 비용이 많이 듭니다 .
- 고장이 발생하는 교호작용을 밝히는 데 중점을 두고 테스트합니다 .
- 테스트 런은 대개 결정론적이므로 결과가 성공 또는 실패로 나타납니다 .
- 반복 런은 동일한 결과를 생성하므로 낭비입니다 .
- 설계 효율도는 중복 런을 포함하지 않고 적용할 수 있는 가능 조건의 수를 기반으로 합니다 .

시스템 테스트는 비용이 많이 들기 때문에 테스트 크기를 줄이는 것이 중요합니다. 일반적으로 가능한 모든 교호작용을 테스트하려면 비용이 너무 많이 들고 불필요한 경우도 종종 있습니다. 경험에 따르면 대부분의 고장은 소수 성분 간의 교호작용으로 인해 발생합니다. 고장을 유발할 가능성이 있는 성분 조합의 가장 큰 크기를 **강도**라고 하며 이로 인해 설계 크기가 결정됩니다.

덮기 배열 플랫폼에서 필요한 설계 강도를 지정합니다. 필요하다면 허용되지 않는 요인 수준 조합을 정의합니다. 덮기 배열 플랫폼은 요구 사항에 맞는 고효율의 설계를 생성합니다. 포함 범위 측면에서 설계 품질을 평가하는 데 사용할 수 있는 지표를 제공합니다. 또한 결과를 분석하는 데 사용할 수 있는 스크립트를 설계 데이터 테이블에 제공합니다.

덮기 배열은 요인 수준 설정의 특정 조합을 실현할 수 없는 상황에서 주로 사용됩니다. 덮기 배열 플랫폼은 요인 수준 조합에 제한이 있는 경우에도 매우 효율적인 덮기 배열을 찾을 수 있습니다.

덮기 배열의 구조와 배열 계산 알고리즘에 대한 배경 지식은 Colbourn(2004), Colbourn et al. (2011), Hartman and Raskin(2004) 및 Martirosyan(2003)에서 확인하십시오. 요인 수준에 제한이 있는 덮기 배열에 대한 자세한 내용은 Cohen et al. (2007) 및 Morgan(2009)에서 확인하십시오.

JMP PRO 덮기 배열 및 강도

강도 t 덮기 배열은 *t* 요인 수준 설정의 모든 조합을 테스트하는 설계입니다. *k* 요인에 대한 특정 설정으로 정의된 교호작용을 가정해 보겠습니다. 해당 교호작용을 포함하는 모든 테스트에서 고장이 발생하면 해당 교호작용이 고장을 **감지**합니다. 이 용어 설명에 따르면 강도 *t* 설계를 생성하여 최대 *t* 요인의 교호작용과 관련된 고장을 감지할 수 있습니다.

문헌에서는 덮기 배열(covering array)을 **요인 덮기 설계**(factor covering design)라고도 합니다. 배경 지식 및 자세한 내용은 Yilmaz et al. (2014), Cohen et al. (2003) 및 Dalal and Mallows(1998)에서 확인하십시오.

덮기 배열의 특성을 설명하기 위해 각각 수준이 두 개인 7개의 범주형 요인에 대한 상황을 가정해 보겠습니다. 요인 수준의 모든 쌍별 조합을 테스트하려고 합니다.

이 경우 사용할 수 있는 설계는 8회 런 해상도 III 주효과 설계입니다.

그림 20.2 해상도 III 설계 (강도 2)

설계

런	X1	X2	X3	X4	X5	X6	X7
1	L2	L2	L2	L2	L2	L2	L2
2	L1	L2	L1	L2	L1	L2	L1
3	L2	L1	L1	L2	L2	L1	L1
4	L1	L1	L2	L2	L1	L1	L2
5	L2	L2	L2	L1	L1	L1	L1
6	L1	L2	L1	L1	L2	L1	L2
7	L2	L1	L1	L1	L1	L2	L2
8	L1	L1	L2	L1	L2	L2	L1

두 요인의 수준에 대한 모든 쌍별 조합이 나타나므로 이 요인 설계는 강도 2 덮기 배열입니다 . 예를 들어 X1 과 X2 의 경우 다음 조합이 각각 두 번 나타납니다 .

- L2, L1
- L2, L2
- L1, L1
- L1, L2

그러나 강도 2 덮기 배열에는 다음과 같이 런이 6 개만 있으면 됩니다.

그림 20.3 강도 2 덮기 배열

설계

런	X1	X2	X3	X4	X5	X6	X7
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	1
3	2	2	2	1	1	1	2
4	2	1	1	2	2	1	2
5	1	2	1	2	1	2	2
6	1	1	2	1	2	2	1

두 요인의 수준에 대한 모든 쌍별 조합이 6회 런에서 한 번 이상 나타납니다. 덮기 배열 설계는 더 적은 수의 런으로 강도 2 포함 범위를 달성하므로 해상도 III 설계보다 효율적입니다.

덮기 배열의 효율도는 필요한 포함 범위를 달성하기 위해 사용된 런 수로 측정됩니다. 런 수가 적을수록 설계 효율도가 높아집니다.

JMP PRO 요인 수준 제한이 없는 덮기 배열의 예

이 예의 데이터는 소프트웨어 테스트 분야의 상호 운용성과 관련이 있습니다 . 관심 요인은 다음 4 개입니다 .

- Web Browser: Safari, IE, Firefox, Chrome, Other
- Operating System: Windows 또는 macOS
- RAM: 4, 8 또는 16 MB
- Connection Speed: 0-1, 1-5 또는 >5 Mbps

이러한 요인 조합 중 고장을 일으킬 가능성이 있는 조합을 알아내려고 합니다.

반응은 각 요인 설정 조합에서 시스템이 제대로 작동하는지 여부입니다.

각 설정 조합을 테스트하려면 실험을 90(5x2x3x3)번 시도해야 합니다. 런 크기를 관리할 수 있도록 하기 위해 세 요인의 모든 조합이 테스트됨을 나타내는 강도 3 포함 범위가 필요하다고 결정합니다.

JMP PRO 설계 생성

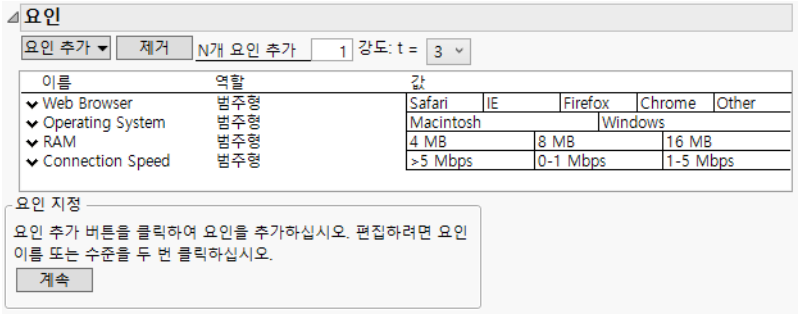
다음 단계를 수행하여 강도 3 덮기 배열을 생성합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Software Factors.jmp 를 엽니다.

   Software Factors.jmp 데이터 테이블에는 요인과 해당 설정이 포함되어 있습니다.
2. **DOE > 특수 용도 > 덮기 배열**을 선택합니다.
3. **강도 : t =** 옆의 메뉴에서 3 을 선택합니다.
4. " 덮기 배열 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다.

   " 요인 " 개요에 네 가지 요인과 요인 수준이 채워져 있습니다.

그림 20.4 소프트웨어 요인에 대한 요인 개요



5. **계속**을 클릭합니다.

   설계 설정에 대한 제한을 입력할 수 있는 " 요인 수준 조합 제한 " 개요가 열립니다. 이 설계에는 제한이 없으므로 기본 선택인 **없음**을 변경하지 마십시오.
6. **설계 생성**을 클릭합니다.

   " 설계 " 개요가 열리고 45 회 런 설계가 표시됩니다.

그림 20.5 소프트웨어 요인에 대한 설계 및 지표 개요

설계

런	Web Browser	Operating System	RAM	Connection Speed
1	Safari	Macintosh	4 MB	>5 Mbps
2	Safari	Windows	4 MB	0-1 Mbps
3	Safari	Macintosh	4 MB	1-5 Mbps
4	Safari	Windows	8 MB	>5 Mbps
5	Safari	Macintosh	8 MB	0-1 Mbps
6	Safari	Macintosh	8 MB	1-5 Mbps
7	Safari	Macintosh	16 MB	>5 Mbps
8	Safari	Macintosh	16 MB	0-1 Mbps
9	Safari	Windows	16 MB	1-5 Mbps
10	IE	Windows	4 MB	>5 Mbps
11	IE	Macintosh	4 MB	0-1 Mbps
12	IE	Macintosh	4 MB	1-5 Mbps
13	IE	Macintosh	8 MB	>5 Mbps
14	IE	Macintosh	8 MB	0-1 Mbps

최적화 최대 반복 수: 250

측정 기준

런 수: 45

t	포함 범위	다양성
3	100.00	68.33
4	50.00	100.00

" 지표 " 개요에서 $t = 3$ 에 해당하는 행을 예로 들어보겠습니다 . " 포함 범위 " 는 100% 로 , 설계에 3 요인 교호작용이 100% 포함됨을 나타냅니다 . 강도 3 설계를 요청했으므로 이는 예상된 결과입니다 . $t = 3$ 의 " 다양성 " 열은 표시되는 3 요인 교호작용의 68.33% 가 고유함을 나타냅니다 . 즉 , 일부 3 요인 조합이 반복됩니다 .

$t = 4$의 "포함 범위"는 50%로, 설계에 4요인 교호작용의 1/2만 포함됨을 나타냅니다. 4요인 설정의 가능한 고유 조합은 90 개입니다 . 설계의 45 회 런은 이 고유 조합의 1/2 을 구성합니다 . " 다양성 " 값 100% 는 4 원 교호작용이 반복되지 않는다는 사실을 뒷받침합니다 .

7. **테이블 생성**을 클릭합니다 .

그림 20.6 소프트웨어 요인에 대한 설계 테이블 (일부분)

덮기 배열
설계 덮기 배열, 강도 3
DOE 대화상자
분석

열(5/0)
반응
Web Browser ✱
Operating System ✱
RAM ✱
Connection Speed ✱

행
모든 행 45
선택됨 0
제외 0
숨김 0
라벨 항목 0

	반응	Web Browser	Operating System	RAM	Connection Speed
1	•	Safari	Macintosh	4 MB	>5 Mbps
2	•	Safari	Windows	4 MB	0-1 Mbps
3	•	Safari	Macintosh	4 MB	1-5 Mbps
4	•	Safari	Windows	8 MB	>5 Mbps
5	•	Safari	Macintosh	8 MB	0-1 Mbps
6	•	Safari	Macintosh	8 MB	1-5 Mbps
7	•	Safari	Macintosh	16 MB	>5 Mbps
8	•	Safari	Macintosh	16 MB	0-1 Mbps
9	•	Safari	Windows	16 MB	1-5 Mbps
10	•	IE	Windows	4 MB	>5 Mbps
11	•	IE	Macintosh	4 MB	0-1 Mbps
12	•	IE	Macintosh	4 MB	1-5 Mbps
13	•	IE	Macintosh	8 MB	>5 Mbps
14	•	IE	Macintosh	8 MB	0-1 Mbps
15	•	IE	Windows	8 MB	1-5 Mbps
16	•	IE	Macintosh	16 MB	>5 Mbps
17	•	IE	Windows	16 MB	0-1 Mbps
18	•	IE	Macintosh	16 MB	1-5 Mbps

설계가 데이터 테이블에 표시됩니다 . 왼쪽 위의 테이블 패널에서 다음을 확인할 수 있습니다 .

- " 설계 " 노트는 현재 설계가 강도 3 덮기 배열임을 나타냅니다 .
- "DOE 대화상자 " 스크립트는 " 덮기 배열 " 창 설정을 재현합니다 .
- " 분석 " 스크립트는 실험 데이터를 분석합니다 .

JMP PRO 실험 데이터 분석

설계 테이블이 준비되었으므로 실험을 수행하고 설계 테이블의 **반응** 열에 데이터를 기록할 수 있습니다 (그림 20.6 참조). 실험 결과는 Software Data.jmp 샘플 데이터 테이블에 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Software Data.jmp 를 엽니다 .
2. 테이블 패널에서 **Analysis** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

그림 20.7 소프트웨어 실험 데이터 분석

Summary

Success Runs	38
Failure Runs	3
Missing	4

Failure Analysis Details

3 Factor Interactions

Factors	Failure Levels	Failure Count
Web Browser, Operating System, RAM	Firefox, Macintosh, 8 MB	2
Web Browser, RAM, Connection Speed	Firefox, 4 MB, 0-1 Mbps	1
	Firefox, 8 MB, 1-5 Mbps	1
	Firefox, 8 MB, >5 Mbps	1

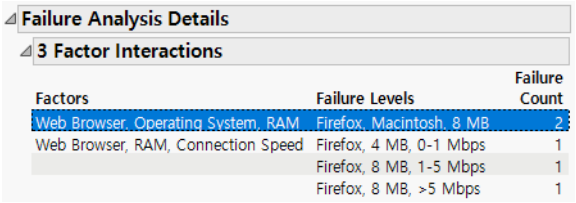
"Summary" 개요에서는 실패한 테스트가 3 개이고 통과 또는 실패 결과가 없는 테스트가 4 개임을 나타냅니다 .

"Failure Analysis Details" 개요에서는 관련된 삼원 교호작용의 측면에서 고장 분석 결과를 제공합니다 . 이 개요에는 모든 테스트가 실패한 요인 수준 조합만 나열됩니다 . 주어진 삼원 설정 조합이 포함된 테스트가 성공하면 해당 삼원 설정 조합은 시스템 고장의 원인이 될 수 없습니다 .

Web Browser 가 Firefox, Operating System 이 macOS, RAM 이 8MB 로 각각 설정된 경우 두 번의 고장이 발생했습니다 . 이 조합을 사용하면 Connection Speed 설정에 관계없이 고장이 발생했습니다 .

3. "3 Factor Interactions" 보고서에서 첫 번째 행을 선택합니다 .

그림 20.8 분석 보고서에서 교호작용 선택



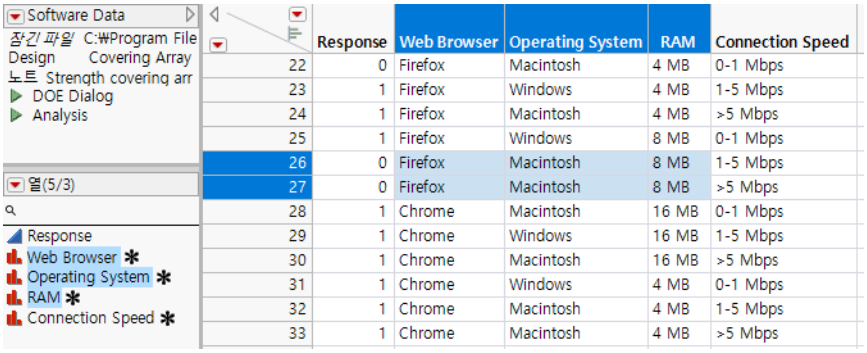
Failure Analysis Details

3 Factor Interactions

Factors	Failure Levels	Failure Count
Web Browser, Operating System, RAM	Firefox, Macintosh, 8 MB	2
Web Browser, RAM, Connection Speed	Firefox, 4 MB, 0-1 Mbps	1
	Firefox, 8 MB, 1-5 Mbps	1
	Firefox, 8 MB, >5 Mbps	1

이렇게 하면 데이터 테이블에서 해당 행과 열이 선택됩니다 .

그림 20.9 데이터 테이블에서 해당 행과 열 선택



	Response	Web Browser	Operating System	RAM	Connection Speed
22	0	Firefox	Macintosh	4 MB	0-1 Mbps
23	1	Firefox	Windows	4 MB	1-5 Mbps
24	1	Firefox	Macintosh	4 MB	>5 Mbps
25	1	Firefox	Windows	8 MB	0-1 Mbps
26	0	Firefox	Macintosh	8 MB	1-5 Mbps
27	0	Firefox	Macintosh	8 MB	>5 Mbps
28	1	Chrome	Macintosh	16 MB	0-1 Mbps
29	1	Chrome	Windows	16 MB	1-5 Mbps
30	1	Chrome	Macintosh	16 MB	>5 Mbps
31	1	Chrome	Windows	4 MB	0-1 Mbps
32	1	Chrome	Macintosh	4 MB	1-5 Mbps
33	1	Chrome	Macintosh	4 MB	>5 Mbps

Web Browser, RAM, 및 Connection Speed 조합과 관련하여 세 번의 고장이 발생했습니다 . 이 중 두 개 (Firefox, 8MB, 1-5Mbps 와 Firefox, 8MB, >5Mbps) 는 Web Browser, Operating System, RAM 교호작용에 대한 두 가지 고장에 해당합니다 . 보고서에서 이러한 행을 선택하면 데이터 테이블에서도 해당 행과 열이 선택됩니다 .

JMP PRO 요인 수준 제한이 있는 덮기 배열의 예

이 예는 Dalal and Mallows(1998) 에 설명된 예를 참조하여 새로 구성되었습니다 . 발신 전화기 (Near Phone) 로 수신 전화기 (Far Phone) 에 전화를 겁니다 . 각 전화 통화는 A 형 또는 B 형 인터페이스를 통과합니다 . 관심 요인은 다음 5 개입니다 .

- Market: USA, UK, Canada, France, Mexico
- Near Phone: ISDN, Bus(Business), Coin, Res(Residential)
- Near Interface: A 또는 B
- Far Phone: ISDN, Bus(Business), Coin, Res(Residential)
- Far Interface: A 또는 B

이러한 요인 쌍 조합 중 고장을 일으킬 가능성이 있는 조합을 알아내려고 합니다 . 그러나 다음과 같은 특정 조합은 불가능합니다 .

- 한 전화기 (Near 또는 Far) 의 ISDN 회선은 A 인터페이스를 사용할 수 없습니다 .
- 발신 전화기 (Near) 의 Business 및 Residential 회선은 B 인터페이스를 사용할 수 없습니다 .

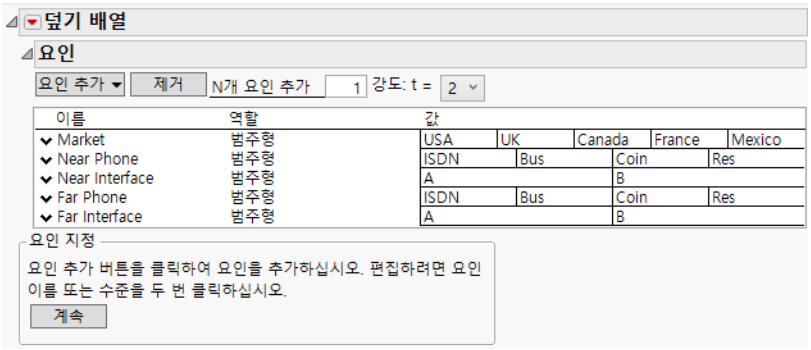
JMP PRO 설계 생성

요인과 해당 설정은 Phone Factors.jmp 데이터 테이블에 제공됩니다. 다음 단계를 수행하여 강도 2 덮기 배열을 생성합니다.

JMP PRO 요인 불러오기

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Phone Factors.jmp 를 엽니다 .
   Phone Factors.jmp 데이터 테이블에는 요인과 해당 설정이 포함되어 있습니다 .
2. **DOE > 특수 용도 > 덮기 배열**을 선택합니다 .
   강도 : t = 옆의 메뉴는 기본적으로 2 로 설정되어 있습니다 .
3. " 덮기 배열 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .
   " 요인 " 개요에 다섯 가지 요인과 요인 수준이 채워져 있습니다 .

그림 20.10 전화기 요인에 대한 요인 개요



4. **계속**을 클릭합니다 .

   " 요인 수준 조합 제한 " 개요가 열립니다 .

JMP PRO 요인 수준 조합 제한

다음 두 가지 방법으로 허용되지 않는 조합을 지정할 수 있습니다 .

- 허용되지 않는 조합 필터 사용
- 허용되지 않는 조합 스크립트 사용

필터를 사용하면 허용되지 않는 조합을 직관적으로 지정할 수 있습니다. 스크립트를 사용하면 허용되지 않는 조합을 쉽고 빠르게 지정할 수 있지만 먼저 스크립트를 작성하거나 저장해야 합니다. 이 예에서 필터를 사용하여 조합을 지정하지 않으려면 "스크립트를 사용하여 허용되지 않는 조합 지정"(550 페이지)으로 건너뜁니다.

다음과 같은 제한이 있습니다

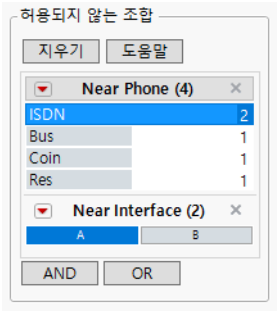
- 한 전화기 (Near 또는 Far) 의 ISDN 회선은 A 인터페이스를 사용할 수 없습니다 .
- 발신 전화기 (Near) 의 Business 및 Residential 전화는 B 인터페이스를 사용할 수 없습니다 .

JMP PRO 필터를 사용하여 허용되지 않는 조합 지정

필터 인터페이스를 사용하여 허용되지 않는 조합을 입력하려면 이 방법을 사용합니다 . 아니면 " 스크립트를 사용하여 허용되지 않는 조합 지정 "(550 페이지) 에 표시된 스크립트를 붙여넣을 수 있습니다 .

1. **허용되지 않는 조합 필터 사용**을 선택합니다 .
2. " 필터 요인 추가 " 목록에서 Near Phone 과 Near Interface 를 선택하고 **추가** 버튼을 클릭합니다 .
3. Ctrl 키를 누르고 Near Phone 에서 ISDN 을 클릭하고 Near Interface 에서 A 를 클릭합니다 .

그림 20.11 첫 번째 제약 조건을 보여 주는 허용되지 않는 조합 패널



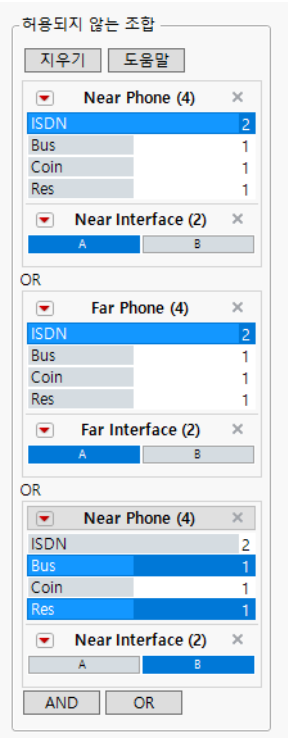
두 블록 모두 진하게 바뀝니다 . 발신 전화기 (Near) 의 ISDN 회선은 A 인터페이스를 사용할 수 없다는 제약 조건을 추가했습니다 .

4. **OR** 를 클릭합니다 .
5. " 필터 요인 추가 " 목록에서 Far Phone 과 Far Interface 를 선택하고 **추가** 버튼을 클릭합니다 .
6. Ctrl 키를 누르고 Far Phone 에서 ISDN 을 클릭하고 Far Interface 에서 A 를 클릭합니다 .

   수신 전화기 (Far) 의 ISDN 회선은 A 인터페이스를 사용할 수 없다는 제약 조건을 추가했습니다 .
7. **OR** 를 클릭합니다 .
8. "필터 요인 추가" 목록에서 Near Phone 과 Near Interface 를 선택하고 **추가** 버튼을 클릭합니다 .
9. Ctrl 키를 누르고 Near Phone에서 Bus와 Res를 클릭하고 Near Interface에서 B를 클릭합니다.

   발신 전화기 (Near) 의 Business 및 Residential 회선은 B 인터페이스를 사용할 수 없다는 제한을 추가했습니다 .

그림 20.12 완료된 허용되지 않는 조합 필터



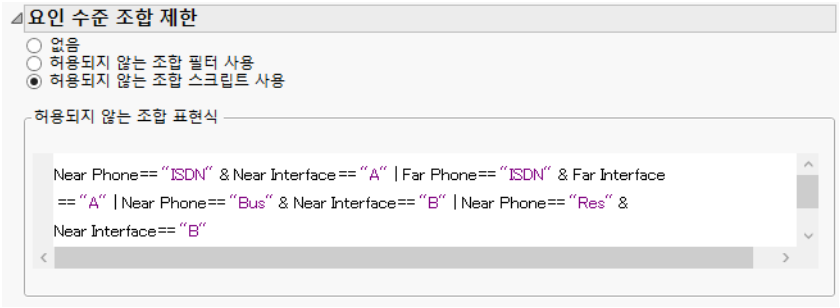
JMP PRO 스크립트를 사용하여 허용되지 않는 조합 지정

스크립트를 작성하여 허용되지 않는 조합을 지정할 수도 있습니다 . 요인을 불러온 후 (" 요인 불러오기 "(547 페이지) 참조) 다음을 수행하십시오 .

1. **계속**을 클릭합니다 .
2. **허용되지 않는 조합 스크립트 사용**을 선택합니다 .
3. 다음 스크립트를 복사하여 " 허용되지 않는 조합 표현식 " 스크립트 상자에 붙여넣습니다 .

```
(Near Phone == "ISDN" & Near Interface == "A") |
(Far Phone == "ISDN" & Far Interface == "A")  |
(Near Phone == "Bus" & Near Interface == "B") |
(Near Phone == "Res" & Near Interface == "B")
```

그림 20.13 완료된 허용되지 않는 스크립트 창



JMP PRO 설계 테이블 구성

참고 : 다음 두 단계에서 난수 시드값을 설정하면 이 예에 나오는 것과 정확히 같은 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

1. " 덮기 배열 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 .
2. 632 를 입력하고 **확인**을 클릭합니다 .
3. **설계 생성**을 클릭합니다 .

   " 설계 " 개요가 열리고 20 회 런 설계가 표시됩니다 . " 지표 " 개요가 창에 추가됩니다 .

그림 20.14 전화기 설계에 대한 지표 개요

" 지표 " 개요에서는 강도 2 포함 범위가 100% 임을 나타냅니다 . 즉 , 허용 가능한 모든 2 요인 조합이 설계에 표시됩니다 . 또한 모든 3 요인 조합의 65% 가 설계에 포함됩니다 .

4. **테이블 생성**을 클릭합니다 .

   설계 테이블에 설계가 추가됩니다 . 다양한 스크립트와 반응 열이 제공됩니다 .

그림 20.15 덮기 배열 설계 테이블

	Response	Market	Near Phone	Near Interface	Far Phone	Far Interface
1	•	Mexico	Res	A	Res	A
2	•	Mexico	Bus	A	Bus	A
3	•	France	Coin	B	Bus	B
4	•	UK	Res	A	Coin	A
5	•	UK	ISDN	B	Bus	B
6	•	France	ISDN	B	Res	A
7	•	Canada	Bus	A	Res	A
8	•	Mexico	ISDN	B	ISDN	B
9	•	Canada	ISDN	B	Coin	B
10	•	UK	Bus	A	ISDN	B
11	•	Canada	Res	A	Bus	B
12	•	USA	ISDN	B	ISDN	B
13	•	UK	Coin	B	Res	B
14	•	Mexico	Coin	B	Coin	B
15	•	USA	Bus	A	Bus	A
16	•	USA	Coin	A	Coin	A
17	•	USA	Res	A	Res	A
18	•	France	Res	A	ISDN	B
19	•	France	Bus	A	Coin	B
20	•	Canada	Coin	B	ISDN	B

JMP PRO 실험 데이터 분석

설계 테이블이 준비되었으므로 실험을 수행하고 설계 테이블의 **반응** 열에 데이터를 기록할 수 있습니다. 실험 결과는 Phone Data.jmp 샘플 데이터 테이블에 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Phone Data.jmp 를 엽니다.
2. 테이블 패널에서 **Analysis** 스크립트 옆의 녹색 삼각형을 클릭합니다.

그림 20.16 전화기 실험 데이터 분석

Summary

Success Runs	17
Failure Runs	3
Missing	0

Failure Analysis Details

2 Factor Interactions

Factors	Failure Levels	Failure Count
Near Interface, Far Phone	A, Coin	3
Far Phone, Far Interface	Coin, A	2
Market, Near Phone	USA, Coin	1
Market, Near Phone	UK, Res	1
Market, Near Phone	France, Bus	1
Market, Far Phone	USA, Coin	1
Market, Far Phone	UK, Coin	1
Market, Far Phone	France, Coin	1
Market, Far Interface	UK, A	1
Near Phone, Near Interface	Coin, A	1
Near Phone, Far Phone	Bus, Coin	1
Near Phone, Far Phone	Res, Coin	1
Near Phone, Far Interface	Coin, A	1

"Summary" 개요에서는 3 개의 테스트가 실패했음을 나타냅니다 .

이원 교호작용이 고장을 감지하는 가장 낮은 수준의 교호작용이므로 "Failure Analysis Details" 개요에 "2 Factor Interactions" 보고서가 포함됩니다 .

"2 Factor Interactions" 보고서에는 세 가지 고장의 원인이 될 수 있는 조합이 표시됩니다 . Near Interface 가 A 로 설정되고 Far Phone 이 Coin 으로 설정된 조합 하나가 세 가지 실패 모두의 원인이 될 수 있습니다 . 아니면 두 개 또는 세 개의 다른 조합으로 인해 세 가지 고장이 발생했을 수 있습니다 .

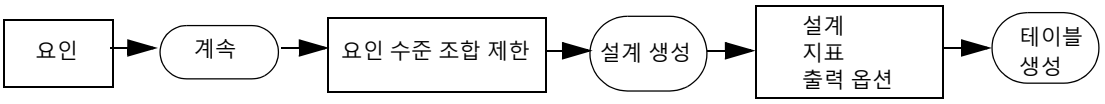
3. "2 Factor Interactions" 보고서에서 첫 번째 행을 선택합니다 .

데이터 테이블에서 4, 16 및 19 행이 선택됩니다 . 이 조합의 경우 Market, Near Phone 및 Far Interface 설정에 관계없이 고장이 발생합니다 . 그러나 요인 설정의 다른 조합도 이러한 고장의 원인이 될 수 있다는 점에 유의하십시오 .

JMP PRO 덮기 배열 창

설계 단계를 진행하는 동안 덮기 배열 창이 업데이트됩니다 . DOE 창의 일반적인 흐름에 대한 자세한 내용은 "DOE 워크플로우 : 기술 , 지정 , 설계 "(56 페이지) 에서 확인하십시오 . 개요는 개요를 업데이트하는 버튼을 클릭할 때 구분되어 나타나며 그림 20.17 의 흐름을 따릅니다 .

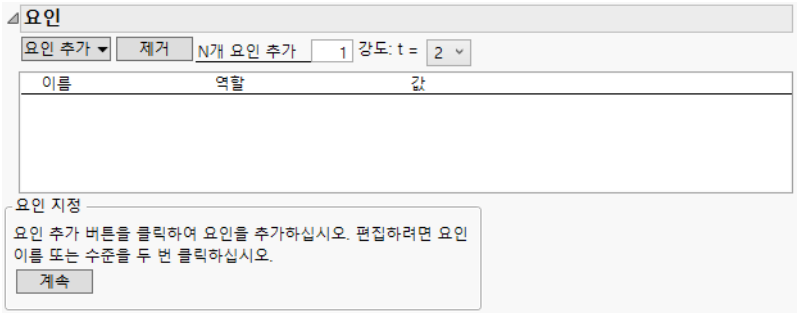
그림 20.17 덮기 배열 흐름



JMP PRO 요인

" 요인 " 개요에서 요인을 추가합니다 .

그림 20.18 요인 개요



요인 추가 **N 개 요인 추가**에 지정된 요인 수를 입력합니다 . 모든 요인은 범주형입니다 . 수준 수를 선택하거나 지정합니다 .

제거 선택한 요인을 제거합니다 .

N 개 요인 추가 특정 수준 수를 가진 여러 요인을 추가합니다 . 추가할 요인 수를 입력하고 **요인 추가**를 클릭한 후 수준 수를 선택하거나 지정합니다 . 수준 수가 다른 여러 요인을 추가하려면 **N 개 요인 추가**를 반복합니다 .

강도 : t = 배열 강도를 지정하는 값을 선택합니다 .

팁 : " 요인 " 패널을 완료하면 빨간색 삼각형 메뉴에서 **요인 저장**을 선택하십시오 . 이렇게 하면 요인 이름과 값을 데이터 테이블에 저장하여 나중에 다시 불러올 수 있습니다 . 자세한 내용은 " 덮기 배열 옵션 "(561 페이지) 에서 확인하십시오 .

JMP PRO 요인 테이블

" 요인 " 테이블에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할을 지정합니다 . 모든 덮기 배열 요인의 설계 역할은 " 범주형 " 입니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인이 적절하게 모델링되도록 합니다 .

값 요인에 대한 설정입니다 . 값을 삽입하려면 기본값을 클릭한 후 원하는 값을 입력합니다 . 설계 테이블의 값 순서는 왼쪽에서 오른쪽으로 입력되는 값의 순서입니다 .

JMP PRO 요인 테이블 편집

다음 내용을 참조하여 " 요인 " 개요에서 작업을 수행합니다 .

- 요인 이름을 편집하려면 해당 요인 이름을 두 번 클릭합니다 .
- 범주형 요인은 요인 이름 왼쪽에 아래쪽 화살표가 있습니다 . 이 화살표를 클릭하여 수준을 추가합니다 .
- 요인 수준을 제거하려면 값을 클릭하고 **삭제**를 클릭한 후 텍스트 상자 바깥쪽을 클릭합니다 .
- 값을 편집하려면 **값** 열의 값을 클릭합니다 .

JMP PRO 요인 열 특성

각 요인의 " 값 라벨 " 열 특성은 설계 테이블에 저장됩니다 . " 값 라벨 " 열 특성은 라벨이 지정된 열에 값을 나타냅니다 . 이러한 라벨은 데이터 테이블에 표시되고 그림과 보고서에 사용됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 값 라벨 "(759 페이지) 에서 확인하십시오 .

JMP PRO 요인 수준 조합 제한

" 요인 " 개요를 완료하고 " 계속 " 을 클릭하면 " 요인 수준 조합 제한 " 개요가 나타납니다 . 이 개요를 사용하면 허용되지 않는 요인 수준 조합을 지정할 수 있습니다 . 제약 조건을 불러오거나 스크립트의 일부로 포함하지 않은 경우 **없음** 옵션이 선택됩니다 . 제약 조건을 지정하려면 다른 옵션 중 하나를 선택합니다 .

허용되지 않는 조합 필터 사용 개별 요인의 값 제한을 기반으로 제약 조건 집합을 정의합니다 . AND 제약 조건과 OR 제약 조건을 모두 정의할 수 있습니다 . 자세한 내용은 " 허용되지 않는 조합 필터 사용 "(555 페이지) 에서 확인하십시오 .

허용되지 않는 조합 스크립트 사용 허용되지 않는 조합 및 기타 제약 조건을 스크립트 편집기 상자에 부울 JSL 표현식으로 정의합니다 . 자세한 내용은 " 허용되지 않는 조합 스크립트 사용 "(556 페이지) 에서 확인하십시오 .

JMP PRO 허용되지 않는 조합 필터 사용

이 옵션은 허용되지 않는 조합을 쉽게 지정하기 위해 데이터 필터를 조정하여 사용합니다. 데이터 필터 사용 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

허용되지 않는 조합을 추가하려면

1. " 필터 요인 추가 " 목록에서 요인을 선택하고 **추가**를 클릭합니다 .
2. 수준을 선택하여 허용되지 않는 조합을 지정합니다 .

참고 : " 필터 요인 추가 " 메뉴의 빨간색 삼각형 옵션은 많은 플랫폼 시작 창의 " 열 선택 " 패널에 있는 옵션과 동일합니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

"추가"를 클릭하면 초기 패널이 업데이트됩니다. "허용되지 않는 조합" 제어판에 선택한 요인이 표시되고 추가 제어 옵션이 제공됩니다.

덮기 배열 플랫폼은 범주형 요인만 허용합니다. 범주형 요인의 경우 가능한 수준은 라벨이 지정된 블록으로 표시되거나, 수준 수가 많으면 목록 항목으로 표시됩니다. 허용하지 않을 수준을 선택합니다. 여러 수준을 선택하려면 Ctrl 키를 누릅니다. 해당 블록 또는 목록 항목은 강조 표시되어 허용되지 않는 수준임을 나타냅니다. "허용되지 않는 조합" 패널에 요인을 추가하면 해당 범주형 요인의 수준 수가 요인 이름 뒤의 괄호 안에 표시됩니다.

허용되지 않는 조합 옵션

지우기 지정한 허용되지 않는 요인 수준 설정을 모두 지웁니다 . 선택한 요인은 지워지지 않습니다 .

AND " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 AND 그룹이 됩니다 . AND 그룹 내에 지정된 요인 수준 조합은 허용되지 않습니다 .

나중에 AND 그룹에 요인을 추가하려면 그룹 개요를 클릭하여 강조 표시된 직사각형을 표시합니다 . 그런 다음 "AND" 를 선택하고 요인을 추가합니다 .

단일 요인을 제거하려면 빨간색 삼각형 메뉴에서 **삭제**를 선택합니다 .

OR " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 별도의 AND 그룹이 됩니다 . OR 로 구분된 AND 그룹의 경우 하나 이상의 AND 그룹에 지정된 조합은 허용되지 않습니다 .

요인에 대한 빨간색 삼각형 메뉴 옵션

한 요인이 여러 OR 그룹에 나타날 수 있습니다 . 특정 OR 그룹에 요인이 나타나는 경우 이를 해당 요인의 인스턴스라고 합니다 .

삭제 " 허용되지 않는 조합 " 패널에서 선택한 요인 인스턴스를 제거합니다 .

선택 해제 요인 인스턴스에 대한 선택을 취소합니다 .

선택 반전 요인 인스턴스에 대해 선택된 값을 선택 취소하고 이전에 선택되지 않은 값을 선택합니다 .

표시 옵션 표시 모양을 변경합니다 .

블록 표시 각 수준을 블록으로 표시합니다 .

목록 표시 각 수준을 목록의 멤버로 표시합니다 .

단일 범주 표시 각 수준을 표시합니다 .

체크박스 표시 각 값 옆에 체크박스를 추가합니다 .

라디오 상자 표시 단일 선택을 위해 각 값 옆에 라디오 선택 항목을 추가합니다 .

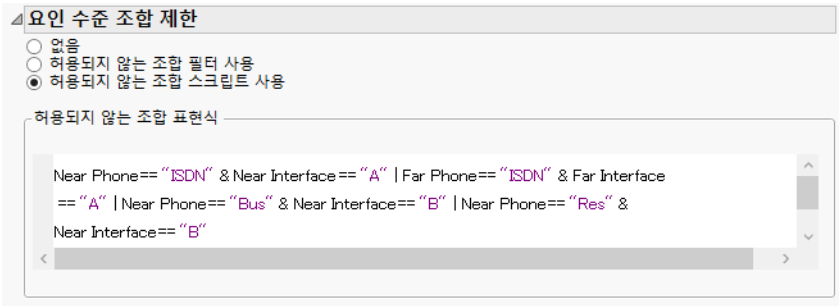
찾기 (범주형 요인에 사용할 수 있음) 요인 이름 아래에 요인 수준에 대한 검색 문자열을 입력할 수 있는 텍스트 상자를 제공합니다 . Enter 키를 누르거나 텍스트 상자 바깥쪽을 클릭하면 검색이 수행됩니다 . **찾기**를 선택하면 " 검색 " 패널의 검은색 삼각형 메뉴에 " 찾기 " 옵션이 나타납니다 . " 옵션 " 에는 " 용어 포함 "," 구 포함 "," 구로 시작 " 또는 " 구로 끝남 " 및 " 결과 반전 " 이 포함되어 있습니다 .

JMP PRO 허용되지 않는 조합 스크립트 사용

이 옵션은 허용하지 않을 조합을 식별하는 스크립트를 삽입할 수 있는 스크립트 창을 엽니다. 스크립트는 부울 표현식으로 평가되어야 합니다. 표현식이 true로 평가되면 지정된 조합이 허용되지 않습니다.

표현식을 작성할 때 수준 이름을 인용 부호로 묶습니다. 수준에 순서 값을 사용하지 마십시오. 예를 들어 그림 20.19에서는 전화기 인터페이스 예, "스크립트를 사용하여 허용되지 않는 조합 지정"(550 페이지)에서 입력한 스크립트를 보여 줍니다.

그림 20.19 따옴표로 묶은 수준 이름을 보여 주는 스크립트 창



JMP PRO 설계

"설계 생성"을 클릭하면 "설계" 및 "지표" 개요가 나타납니다. 계산이 많이 필요한 설계의 경우 진행률 표시줄이 나타납니다.

"설계" 개요에는 생성한 설계가 표시됩니다. 첫 번째 열은 런 순서 목록입니다. 모든 런을 보려면 스크롤 막대를 사용해야 할 수도 있습니다. 나머지 열은 각 런의 요인 설정을 보여 줍니다.

최적화

덮기 배열 플랫폼에서 구성했거나 "설계 불러오기" 빨간색 삼각형 옵션을 사용하여 불러온 설계의 크기를 줄이려면 "최적화"를 선택합니다. 덮기 배열 플랫폼에서 구성한 설계가 최적으로 확인되면 "최적화"를 사용할 수 없습니다. 특히 2수준 요인에 대해 이 플랫폼에서 구성한 모든 무제약 강도 2 설계는 최적입니다. 또한 *t+1* 요인에 대한 무제약 강도 *t* 설계는 모든 *t*에 최적입니다.

알고리즘에 대한 자세한 내용은 "최적화 알고리즘"(562 페이지)에서 확인하십시오.

참고 : 최적화는 시간이 많이 걸리는 작업이지만 반복적으로 실행하여 점진적으로 더 좋은 설계를 생성할 수 있습니다 .

최대 반복 수 옵션을 사용하여 설계를 최적화하는 데 사용할 최대 반복 수를 지정합니다.

JMP PRO 불충족 제약 조건

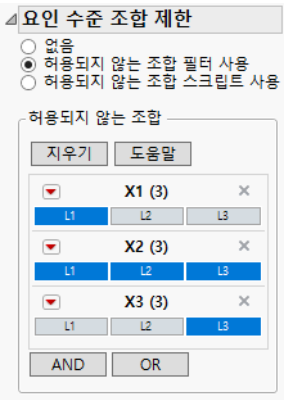
필요한 모든 요인 수준을 나타내는 덮기 배열을 생성할 수 없게 하는 제약 조건 집합을 **불충족** 제약 조건이라고 합니다.

불충족 제약 조건의 예

각각 수준이 세 개인 세 가지 요인에 대한 강도 2 설계를 가정해 보겠습니다 .

1. **DOE > 특수 용도 > 덮기 배열**을 선택합니다 .
2. **N 개 요인 추가** 옆에 3 을 입력합니다 .
3. **요인 추가** 메뉴에서 **3 수준**을 선택합니다 .
4. **계속**을 클릭합니다 .
5. **허용되지 않는 조합 필터 사용**을 선택합니다 .
6. **필터 요인 추가** 목록에서 세 요인을 모두 선택하고 **추가** 버튼을 클릭합니다 .
7. Ctrl 키를 누르고 다음 수준을 선택합니다 .
   - X1 에서 L1 을 선택합니다 .
   - X2 에서 L1, L2, L3 을 선택합니다 .
   - X3 에서 L3 을 선택합니다 .

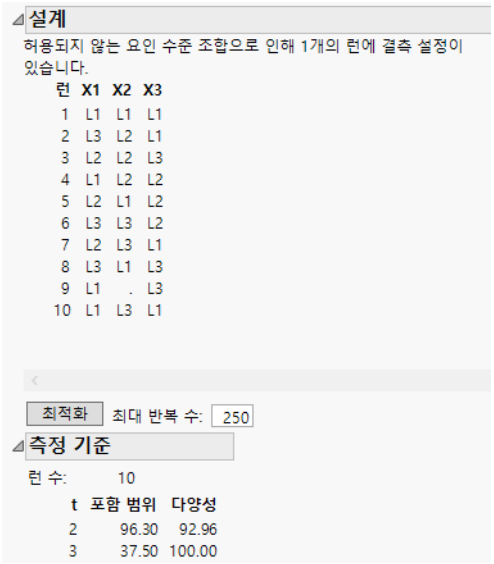
그림 20.20 완료된 요인 수준 조합 제한 패널



참고 : 다음 두 단계에서 난수 시드값을 설정하면 이 예에 나오는 것과 정확히 같은 결과가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

8. " 덮기 배열 " 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 .
9. 12345 를 입력하고 **확인**을 클릭합니다 .
10. **설계 생성**을 클릭합니다 .

그림 20.21 설계 및 지표 개요



" 설계 " 개요 아래의 노트는 제약 조건으로 인해 1 개의 런에 결측 설정이 있음을 나타냅니다 . 해당 런은 런 9 입니다 . 덮기 배열의 강도가 2 가 되게 하려면 X1 = L1, X3 = L3 조합이 필요합니다 . 그러나 이러한 설정의 경우 제약 조건으로 인해 X2 에 대한 모든 설정이 허용되지 않습니다 .

JMP PRO 지표

" 지표 " 개요는 설계가 강도 요구 사항을 얼마나 잘 충족하는지에 대한 정보를 제공합니다 . 무제약 설계의 이러한 지표에 대한 배경 지식은 Dalal and Mallows(1998) 에서 확인하십시오 .

t 요인 수입니다 .

포함 범위 설계에 나오는 고유 *t* 요인 설정의 수를 가능한 *t* 요인 설정의 총 수에 대한 비율 (백분율) 로 나타낸 값입니다 . *t* 포함 범위가 100% 이면 가능한 모든 *t* 요인 설정이 설계에 포함됩니다 . 각 *t* 요인 설정은 여러 번 나타날 수 있습니다 .

제약 및 불충족 설계의 경우 모든 *t* 요인 조합에 제약 조건이 적용된 후 가능한 *t* 요인 설정의 수에 맞게 포함 범위 정의가 조정됩니다 . 자세한 내용은 " 지표 계산식 "(563 페이지) 에서 확인하십시오 .

다양성 설계에서 고유 *t* 요인 설정의 수를 *t* 요인 설정의 총 수에 대한 비율 (백분율) 로 나타낸 값입니다 . *t* 다양성은 설계에서 반복이 방지되는 정도를 측정합니다 . *t* 다양성이 100% 이면 *t* 요인 설정이 반복되지 않습니다 . *t* 다양성이 50% 이면 고유 *t* 요인 설정이 나타나는 평균 횟수가 2 회입니다 .

제약 및 불충족 설계의 경우 결측 설정이 있는 런 수에 맞게 다양성 정의가 조정됩니다 . 자세한 내용은 " 지표 계산식 "(563 페이지) 에서 확인하십시오 .

JMP PRO 출력 옵션

테이블 생성 " 덮기 배열 " 데이터 테이블을 생성합니다 .

뒤로 " 요인 " 개요로 돌아갑니다 . 이전 개요를 변경하고 설계를 다시 생성할 수 있습니다 .

참고 : " 요인 수준 조합 제한 " 개요에서 허용되지 않는 조합을 정의하면 해당 조합이 스크립트로 유지됩니다 . " 계속 " 을 클릭하면 " 허용되지 않는 조합 스크립트 사용 " 패널에 스크립트가 표시됩니다 .

JMP PRO 덮기 배열 데이터 테이블

"덮기 배열" 데이터 테이블에 포함된 첫 번째 열에 반응을 입력할 수 있습니다. 나머지 열은 요인 설정을 제공합니다.

왼쪽 위의 테이블 패널에는 설계가 덮기 배열임을 나타내고 설계 강도를 제공하는 " 설계 " 노트가 포함되어 있습니다 . 테이블 패널에는 다음 스크립트도 있습니다 .

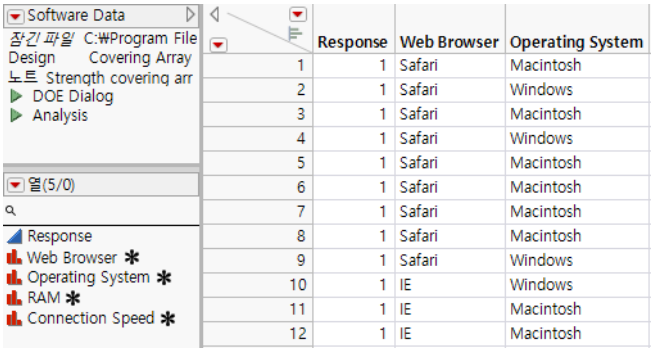
DOE 대화상자 설계 테이블을 생성하는 데 사용된 " 덮기 배열 " 창을 다시 생성합니다 .

허용되지 않는 조합 " 요인 수준 조합 제한 " 개요에 입력한 요인 수준 제한을 표시합니다 .

분석 실험 데이터 분석 결과를 제공합니다. 스크립트에 대한 자세한 내용은 "분석 스크립트"(560 페이지) 에서 확인하십시오 . 배경 지식은 Zhang and Zhang(2011) 에서 확인하십시오 .

팁 : 스크립트를 실행하려면 스크립트 이름 옆의 녹색 삼각형을 클릭합니다 .

그림 20.22 스크립트를 보여 주는 Software Data.jmp 의 덮기 배열 테이블 (일부분)



	Response	Web Browser	Operating System
1	1	Safari	Macintosh
2	1	Safari	Windows
3	1	Safari	Macintosh
4	1	Safari	Windows
5	1	Safari	Macintosh
6	1	Safari	Macintosh
7	1	Safari	Macintosh
8	1	Safari	Macintosh
9	1	Safari	Windows
10	1	IE	Windows
11	1	IE	Macintosh
12	1	IE	Macintosh

JMP PRO 분석 스크립트

" 분석 " 스크립트는 반응 열에 대해 다음과 같이 가정합니다 .

- 반응은 고장이 발생할 경우 0, 성공할 경우 1 로 기록됩니다 .

- 결측값이 허용됩니다 .
- 반응 열은 연속형입니다 .
- 반응 열을 이동하거나 이름을 바꿀 수 있습니다 .

" 분석 " 스크립트는 다음과 같은 두 가지 개요가 포함된 보고서를 생성합니다 .

- " 요약 " 개요는 성공한 런 수 , 고장이 발생한 런 수 및 반응이 없는 런 수를 제공합니다 .
- " 고장 분석 상세 정보 " 보고서에는 "<k> 요인 교호 작용 " 보고서가 포함됩니다 . *k* 값은 고장을 감지하는 최소 교호작용 개수입니다 . **감지**의 정의는 " 덮기 배열 및 강도 "(541 페이지) 에서 확인하십시오 . 세 열에는 다음 항목이 포함됩니다 .
  - **요인** 열에는 고장을 감지하는 모든 *k* 요인 조합이 나열됩니다 .
  - **고장 수준** 열에는 **요인** 열에서 고장을 감지하는 *k* 요인의 값이 나열됩니다 .
  - **고장 개수** 열에는 **고장 수준**의 *k* 요인 조합에 해당하는 고장 수가 표시됩니다 .

  참고 : 고장 관측값은 **고장 수준** 열에 나열된 둘 이상의 *k* 요인 조합에 나타날 수 있습니다 .

"<k> 요인 교호작용" 보고서의 행은 데이터 테이블에 동적으로 연결되어 있습니다. 보고서에서 하나 이상의 행을 선택하면 데이터 테이블에서도 해당 행이 선택됩니다.

JMP PRO 덮기 배열 옵션

덮기 배열 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

요인 저장 각 요인에 대해 요인 수준을 포함하는 열로 구성된 데이터 테이블을 생성합니다 . 각 요인 열에는 " 설계 역할 ", " 값 순서 " 및 " 요인 변경 " 이라는 열 특성이 포함됩니다 . 요인을 저장하면 DOE 창에 요인을 빠르게 불러올 수 있습니다 .

참고 : 빈 테이블에 데이터를 입력하여 덮기 배열의 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할 (" 범주형 ") 을 할당해야 합니다 . 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다 . 그런 다음 " 설계 역할 " 영역에서 " 범주형 " 을 선택합니다 .

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 " 요인 " 개요에 불러옵니다 .

설계 불러오기 활성 데이터 테이블에서 설계를 불러옵니다 . 활성 데이터 테이블이 없으면 테이블을 열기 위한 대화상자가 나타납니다 . " 설계 불러오기 " 를 선택하면 설계에서 요인으로 지정할 열을 선택할 수 있는 메뉴가 나타납니다 . 모든 열을 범주형으로 가져옵니다 . 열과 열 값이 " 요인 " 개요에 나열됩니다 . " 설계 " 개요에는 데이터 테이블의 각 행에 대한 런이 표시되고 각 런에 대한 요인 값이 제공됩니다 .

" 설계 불러오기 " 옵션을 사용하면 다음과 같이 기존 설계의 지표를 가져오고 , 설계를 수정하거나 , " 분석 " 스크립트를 구성할 수 있습니다 .

- " 지표 " 개요에는 지정된 설계의 *t* 포함 범위 및 *t* 다양성이 표시됩니다 .

- **뒤로**를 클릭하여 요인 수준 제한을 적용한 후 새 설계를 구성할 수 있습니다.
- **테이블 생성**을 클릭하면 반응을 입력할 수 있는 설계 테이블이 생성됩니다. 이 테이블에는 설계에 대한 "분석" 스크립트가 포함되어 있습니다.

난수 시드값 설정 JMP 가 랜덤 성분이 있는 특정 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다. 시드값은 덮기 배열을 위한 시작 설계 및 반복 횟수를 선택합니다. 대부분의 설계에서 난수 시드값은 설계를 그대로 재현하지만 런 순서는 다를 수 있습니다.

참고 : 시간 상한 및 반복 횟수는 설계 생성 시간을 제한하는 데 사용됩니다. 일부 대형 설계와 고강도 설계의 경우 시스템에 따라 시간 제한이 반복 제한보다 우선할 수 있습니다. 이러한 설계에서는 난수 시드값으로 설계가 재현되지 않을 수 있습니다.

설계를 재현하려면 "설계 생성"을 클릭하기 전에 설계 생성에 사용되는 난수 시드값을 입력합니다.

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자" 스크립트에 포함됩니다.

고급 옵션 덮기 배열에는 사용할 수 없습니다.

스크립트를 스크립트 창에 저장 "덮기 배열" 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다.

JMP PRO 덮기 배열에 대한 기술 상세 정보

- "최적화 알고리즘"
- "지표 계산식"

JMP PRO 최적화 알고리즘

"최적화" 버튼은 **생성 후 랜덤화된 최적화 도구**(Nayeri et al., 2013)라고도 하는 덮기 배열 최적화 도구 클래스와 개념적으로 유사한 알고리즘을 호출합니다. 그러나 JMP 의 알고리즘은 제약조건이 있는 설계도 처리한다는 점에서 완전히 다릅니다. 특히 제약된 덮기 배열은 물론 불충족 제약 덮기 배열도 최적화합니다.

이 알고리즘은 최적화할 설계를 강도가 지정된 덮기 배열이라고 가정합니다. 강도 t 의 K 요인 설계인 경우, 모든 ${}_{K}C_{t}$ 요인 투영을 반복적으로 검사하여 런을 제거하거나 병합할 수 있는지 여부를 결정합니다. 따라서 K 또는 t 가 커지면 알고리즘 실행 시간이 빠르게 증가합니다. 성능을 개선하기 위해 워크스테이션에서 사용 가능한 최대 CPU 코어를 사용하도록 JMP 구현이 스레드됩니다.

JMP PRO 지표 계산식

포함 범위 및 다양성 계산식은 제약 조건이 있는지 여부에 따라 달라집니다. 다음 표기가 사용됩니다.

${}_uC_v$ 는 한 번에 v 씩 적용된 u 개 항목의 조합 수입니다.
t 는 설계 강도입니다.
K 는 요인 수입니다.
$M = {}_KC_t$
$i = 1, 2, ..., M$ 은 t 요인의 모든 조합 또는 **투영**을 정렬하는 인덱스입니다.
v_{ik} 는 k^{th} 요인의 수준 수입니다.
n_i 는 설계에서 i^{th} 투영에 대한 고유 t 투플의 수입니다.
p_i 는 i^{th} 투영에서 요인에 대한 v_{ik} 의 곱입니다.
r 는 설계의 런 수입니다.

JMP PRO 무제약 설계

포함 범위 및 다양성은 다음 식을 사용하여 구합니다.

$$Coverage = \frac{1}{M}\sum_{i=1}^{M} n_i/p_i$$

$$Diversity = \frac{1}{M}\sum_{i=1}^{M} n_i/r$$

JMP PRO 제약 설계

제약된 설계에서는 특정 t 투플이 허용되지 않습니다. 이로 인해 일부 t 투플에 대한 결측값이 발생할 수 있습니다. t 요인의 일부 조합에서 유효한 t 투플이 없을 수 있습니다. 포함 범위와 다양성은 가능한 유효 조합의 관점에서 정의해야 합니다. 따라서 제약된 설계의 계산식에는 다음과 같은 추가 표기가 필요합니다.

a_i 는 i^{th} 투영의 요인으로 인해 유효하지 않은 t 투플의 수입니다.
m 은 유효한 t 투플이 없는 투영의 수입니다.
q_i 는 설계에서 i^{th} 투영의 요인에 대한 결측값이 있는 런 수입니다.
$r_i = r - q_i$
$M' = M - m$

포함 범위 및 다양성은 다음 식을 사용하여 구합니다 .

$$Coverage = \frac{1}{M'} \sum_{i=1}^{M'} n_i/(p_i - a_i)$$

$$Diversity = \frac{1}{M'} \sum_{i=1}^{M'} n_i/r_i$$

유효하지 않은 t 투플이 없고($M' = M$) 결측값도 없는 경우(모든 i에 대해 $r_i = r$), 제약된 설계의 포함 범위 및 다양성에 대한 정의가 무제약 설계에 대한 정의로 축소됩니다. 자세한 내용은 Morgan(2014)에서 확인하십시오.

21 장

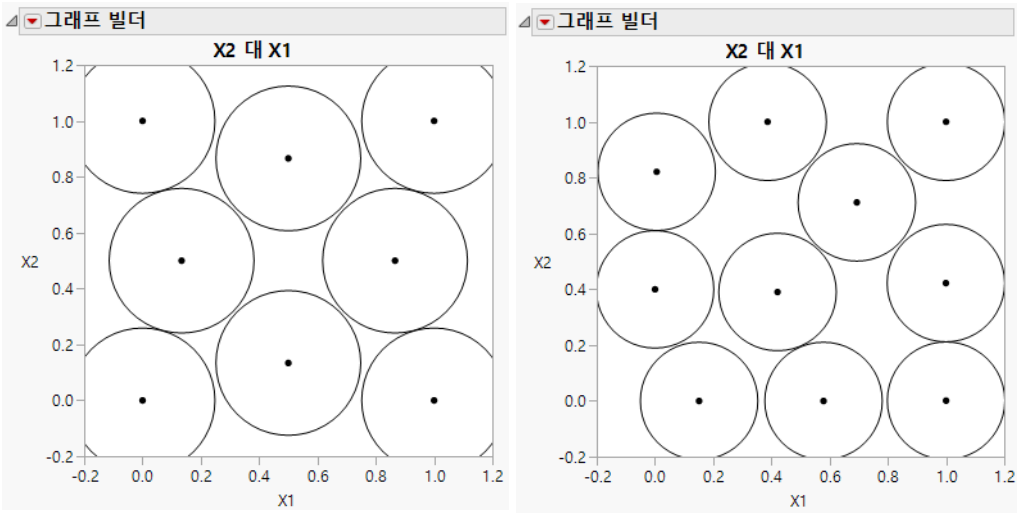
공간 채움 설계

공간 채움 설계는 런 간의 변동보다 모형 형태에 대한 관심이 훨씬 큰 상황에서 유용합니다. 컴퓨터 시뮬레이션 모형의 민감도 연구를 예로 들겠습니다. 이 상황에서 기계적 또는 결정적 모델링 문제에 대한 변동은 무시할 정도로 작습니다. 변동이 없는 시스템의 경우 반복, 랜덤화, 블록화 등은 관련이 없습니다.

공간 채움 플랫폼은 연속형 요인과 범주형 요인이 모두 있는 상황에 적합한 설계를 제공합니다. 연속형 요인의 경우 공간 채움 설계를 사용하는 목적은 다음과 같습니다.

- 두 설계점 사이의 거리 최대화
- 점 사이의 간격을 균등하게 설정

그림 21.1 공간 채움 설계



목차

공간 채움 설계 개요 567
공간 채움 설계 생성 568
반응 568
요인 570
요인 제약 조건 정의 572
공간 채움 설계 방법 575
설계 575
설계 진단 575
설계 테이블 576
공간 채움 설계 옵션 576
Sphere-Packing 설계 578
Sphere-Packing 설계 생성 578
Sphere-Packing 설계 시각화 580
라틴 초입방체 설계 581
라틴 초입방체 설계 생성 581
라틴 초입방체 설계 시각화 582
균등 설계 584
Sphere-Packing, 라틴 초입방체 및 균등 방법 비교 586
최소 잠재 설계 587
최대 엔트로피 설계 589
가우스 과정 IMSE 최적 설계 591
쾌속 유연 채움 설계 592
FFF 최적 기준 592
FFF 설계의 평균 군집 크기 설정 593
제약 조건 594
제약된 쾌속 유연 채움 설계 생성 및 보기 594
맵 쉐이프를 위한 공간 채움 설계 생성 596
Sphere-Packing 설계의 예 601
Borehole 데이터에 대한 Sphere-Packing 설계 생성 601
결정적 데이터 분석을 위한 지침 602

공간 채움 설계 개요

공간 채움 설계는 결정적 또는 근사 결정적 시스템을 모델링하는 데 유용합니다. 컴퓨터 시뮬레이션을 결정적 시스템의 한 예로 들 수 있습니다. 이러한 시뮬레이션은 복잡하게 연관된 많은 변수를 포함하여 매우 복합적일 수 있습니다. 이러한 시스템에 대해 설계된 실험의 목표는 제한된 요인 범위에서 시스템 동작을 적절하게 예측하는 단순한 경험적 모형을 찾는 것입니다.

랜덤 잡음이 상당히 큰 시스템에 대한 실험 목표는 예측 분산을 최소화하는 것입니다. 결정적 시스템에 대한 실험의 경우 분산은 없지만 **편향**이 있습니다. 편향은 근사 모형과 실제 수학 함수 간의 차이입니다. 공간 채움 설계 목표는 이러한 편향의 경계를 나타내는 것입니다.

편향의 경계를 나타내는 한 가지 방법은 실험 경계를 벗어나지 않고 설계점을 가능한 한 서로 멀리 퍼뜨리는 것입니다. 또한 관심 영역에서 설계점 간격을 균등하게 두는 것도 방법이 될 수 있습니다.

공간 채움 설계 도구는 다음 설계 방법을 지원합니다 .

Sphere Packing 설계점 쌍 사이의 최소 거리를 최대화합니다 . 자세한 내용은 "Sphere-Packing 설계 "(578 페이지) 및 "Borehole 데이터에 대한 Sphere-Packing 설계 생성 "(601 페이지) 에서 확인하십시오 .

라틴 초입방체 설계점 사이의 최소 거리를 최대화하지만 각 요인의 수준 간격이 균등해야 합니다 . 이 방법은 균등 분포와 비슷한 설계를 생성합니다 . 라틴 초입방체 방법은 Sphere-Packing 방법과 균등 설계 방법을 절충한 것입니다 . 자세한 내용은 " 라틴 초입방체 설계 "(581 페이지) 에서 확인하십시오 .

균등 경험적 균등 분포를 갖는 설계점과 이론적 균등 분포 사이의 불일치를 최소화합니다 . 자세한 내용은 " 균등 설계 "(584 페이지) 에서 확인하십시오 .

최소 잠재 구형 안에서 중심 가까이에 점을 펼쳐 놓습니다. 자세한 내용은 " 최소 잠재 설계 "(587 페이지) 에서 확인하십시오 .

최대 엔트로피 데이터 집합 분포에 포함된 정보의 양을 측정합니다 . 자세한 내용은 " 최대 엔트로피 설계 "(589 페이지) 에서 확인하십시오 .

가우스 과정 IMSE 최적 실험 영역에 대한 가우스 과정의 통합 평균 제곱 오차를 최소화하는 설계를 생성합니다 . 자세한 내용은 " 가우스 과정 IMSE 최적 설계 "(591 페이지) 에서 확인하십시오 .

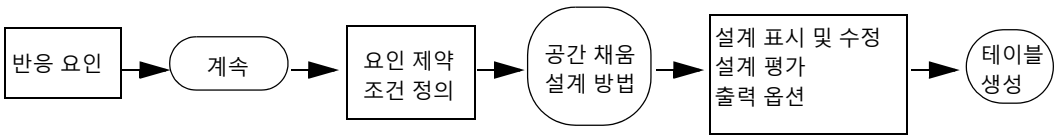
쾌속 유연 채움 쾌속 유연 채움 방법은 설계 공간의 랜덤 점으로 군집을 이룹니다 . 이러한 군집은 최적화 기준에 따라 설계점을 선택하는 데 사용됩니다 . 이 방법만 범주형 요인을 수용하고 설계 공간에 제약 조건을 지정할 수 있습니다. 선형 제약 조건 및 허용되지 않는 조합을 지정할 수 있습니다 . 자세한 내용은 " 쾌속 유연 채움 설계 "(592 페이지) 및 " 제약된 쾌속 유연 채움 설계 생성 및 보기 "(594 페이지) 에서 확인하십시오 .

참고 : 런 수가 500 이하이면 가우스 과정 모형이 데이터 테이블에 저장됩니다 . 런 수가 500 을 초과하면 신경망 모형이 데이터 테이블에 저장됩니다 .

공간 채움 설계 생성

DOE > 특수 용도 > 공간 채움 설계를 선택하여 공간 채움 설계를 생성합니다. 먼저 반응 및 요인을 정의합니다. 그런 다음 계속해서 설계 옵션, 생성 및 평가를 진행합니다. 설계 과정은 그림 21.2의 흐름을 따릅니다.

그림 21.2 공간 채움 설계 흐름

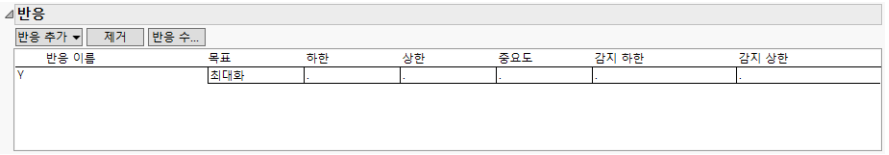


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다. 이 옵션을 사용하면 반응 이름, 목표, 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다.

그림 21.3 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다. " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다. " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다.

JMP PRO 함수 (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다. 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다.

제거 선택한 반응을 제거합니다.

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다. 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다. 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다. 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다.

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응 열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .
- " 없음 " 목표는 최적화에 관한 목표가 없음을 나타냅니다 . 만족도 함수가 구성되지 않습니다 .

참고 : 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다 . 데이터 테이블에서 반응 열의 " 열 정보 " 창을 열고 (**열** > **열 정보**) 원하는 목표값을 입력하십시오 .

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다 . 이러한 값은 반응에 대한 가중치로 처리됩니다 . 반응이 하나뿐이면 기본적으로 1 로 설정되므로 중요도를 지정할 필요가 없습니다 .

JMP PRO **감지 한계** 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다 . 감지 한계는 열 특성으로 저장됩니다 . 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오 .

반응 개요 편집

" 반응 " 개요에서 다음 사항에 유의하십시오 .

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다 .
- 목표를 클릭하여 변경할 수 있습니다 .
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다 .
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다 .

반응 한계 열 특성

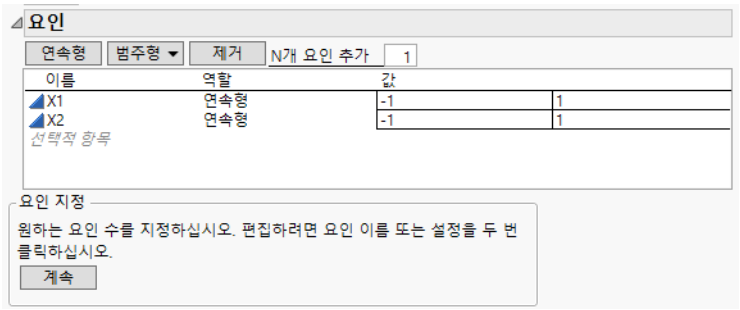
반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

요인

"요인" 개요에서 공간 채움 설계의 요인을 추가합니다.

그림 21.4 요인 개요



"요인" 개요에는 다음 옵션이 포함되어 있습니다.

연속형 **N 개 요인 추가**에 지정된 수의 연속형 요인을 입력합니다.

범주형 **N 개 요인 추가**에 지정된 수의 명목형 요인을 입력합니다.

제거 선택한 요인을 제거합니다.

N 개 요인 추가 지정된 유형의 여러 요인을 추가합니다. 추가할 요인 수를 입력하고 "연속형" 또는 "범주형"을 클릭합니다. 유형이 다른 여러 요인을 추가하려면 **N 개 요인 추가**를 반복합니다.

팁 : "요인" 패널을 완료하면 빨간색 삼각형 메뉴에서 **요인 저장**을 선택하십시오. 이렇게 하면 요인 이름과 값을 데이터 테이블에 저장하여 나중에 다시 불러올 수 있습니다. 자세한 내용은 "공간 채움 설계 옵션"(576 페이지)에서 확인하십시오.

요인 개요

" 요인 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

이름 요인의 이름입니다 . 요인을 추가하면 X1, X2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

역할 요인의 설계 역할을 지정합니다 . 요인의 " 설계 역할 " 열 특성은 데이터 테이블에 저장됩니다 . 이 특성은 요인 유형이 적절하게 모델링되도록 합니다 .

값 요인에 대한 실험 설정입니다 . 값을 삽입하려면 기본값을 클릭한 후 원하는 값을 입력합니다 .

요인 개요 편집

" 요인 " 개요에서 다음 사항에 유의하십시오 .

- 요인 이름을 편집하려면 해당 요인 이름을 두 번 클릭합니다 .
- 범주형 요인은 요인 이름 왼쪽에 아래쪽 화살표가 있습니다 . 이 화살표를 클릭하여 수준을 추가합니다 .
- 요인 수준을 제거하려면 값을 클릭하고 **삭제**를 클릭한 후 텍스트 상자 바깥쪽을 클릭합니다 .
- 값을 편집하려면 **값** 열의 값을 클릭합니다 .

요인 유형

연속형 숫자 데이터 유형만 가능합니다 . 연속형 요인은 개념적으로 공정 및 측정 시스템의 제한을 고려하여 사용자가 제공하는 하한과 상한 사이의 어떤 값이든 설정할 수 있는 요인입니다 .

범주형 숫자 또는 문자 데이터 유형에 해당됩니다 . 범주형 요인의 경우 값 순서는 왼쪽에서 오른쪽으로 입력되는 값의 순서입니다 . 이 순서는 설계 데이터 테이블이 생성된 후 " 값 순서 " 열 특성에 저장됩니다 .

요인 열 특성

각 요인에 대한 다양한 열 특성이 데이터 테이블에 저장됩니다 .

설계 역할 각 요인에 " 설계 역할 " 열 특성이 할당됩니다 . 요인을 정의할 때 지정하는 역할에 따라 " 설계 역할 " 열 특성의 값이 결정됩니다 . " 설계 역할 " 특성은 실험 데이터를 모델링할 때 요인이 어떤 용도로 사용되는지를 반영합니다 . " 설계 역할 " 값은 설계 확대 플랫폼에서 사용됩니다 .

요인 변경 각 요인에 " 쉬움 " 으로 설정된 " 요인 변경 " 열 특성이 할당됩니다 . 공간 채움 설계에서는 각 실험 런에 대해 요인 수준을 변경할 수 있다고 가정합니다 . " 요인 변경 " 값은 설계 평가 및 설계 확대 플랫폼에서 사용됩니다 .

코딩 " 역할 " 이 " 연속형 " 인 경우 요인의 " 코딩 " 열 특성이 저장됩니다 . 이 특성은 하한값과 상한값이 각각 –1 과 +1 에 해당하도록 요인 값을 변환합니다 . " 최소 제곱 적합 " 보고서의 추정값과 검정은 변환된 값을 기반으로 합니다 .

값 순서 " 역할 " 이 " 범주형 " 또는 " 블록화 " 인 경우 요인의 " 값 순서 " 열 특성이 저장됩니다 . 이 특성은 요인 수준이 나타나는 순서를 결정합니다 .

요인 제약 조건 정의

쾌속 유연 채움 방법을 사용하여 구성한 공간 채움 설계에만 요인 제약 조건을 지정할 수 있습니다.

" 요인 제약 조건 정의 " 를 사용하여 설계 공간을 제한할 수 있습니다 . 제약 조건을 불러오거나 스크립트의 일부로 포함하지 않은 경우 **없음** 옵션이 선택됩니다 . 제약 조건을 지정하려면 다른 옵션 중 하나를 선택합니다 .

선형 제약 조건 지정 요인의 선형 조합에 대한 부등식 제약 조건을 지정합니다 . " 연속형 " 또는 " 혼합 " 역할이 설정된 요인에만 사용할 수 있습니다 . 자세한 내용은 " 선형 제약 조건 지정 " 에서 확인하십시오 .

참고 : 선형 제약 조건을 포함하는 설계의 스크립트를 저장하면 해당 스크립트에 선형 제약 조건이 보다 작거나 같음 부등식 (≤) 으로 표현됩니다 .

허용되지 않는 조합 필터 사용 개별 요인의 값 제한을 기반으로 제약 조건 집합을 정의합니다 . AND 제약 조건과 OR 제약 조건을 모두 정의할 수 있습니다 . 자세한 내용은 " 허용되지 않는 조합 필터 사용 " 에서 확인하십시오 .

허용되지 않는 조합 스크립트 사용 허용되지 않는 조합 및 기타 제약 조건을 스크립트 편집기 상자에 부울 JSL 표현식으로 정의합니다 . 자세한 내용은 " 허용되지 않는 조합 스크립트 사용 " 에서 확인하십시오 .

참고 : 요인 제약 조건을 포함하는 설계를 분석할 때는 모형 프로파일러가 제약 조건을 적용합니다 .

선형 제약 조건 지정

설계 공간에서 연속형 요인을 독립적으로 변경할 수 없는 경우 선형 부등식 제약 조건을 지정할 수 있습니다. 선형 부등식은 허용되는 요인 수준 설정을 정의합니다.

추가를 클릭하여 하나 이상의 선형 부등식 제약 조건을 입력합니다 .

추가 설계의 모든 연속형 요인을 포함하는 선형 표현식 템플릿을 추가합니다 . 요인에 대한 계수 값을 입력한 후 선형 제약 조건을 반영하는 부등식 방향을 선택합니다 . 부등식의 오른쪽 상자에 제약 값을 지정합니다 . 제약 조건을 더 추가하려면 **추가**를 다시 클릭합니다 .

참고 : " 구 반지름 " 을 지정하여 설계 영역을 이미 제한한 경우에는 " 추가 " 옵션이 비활성화됩니다 .

마지막 제약 조건 제거 마지막 제약 조건을 제거합니다 .

제약 조건 검사 제약 조건의 일관성을 확인합니다 . 이 옵션은 중복 제약 조건을 제거하고 실행 가능성 검사를 수행합니다 . 문제가 있으면 JMP 경고가 나타납니다 . 제약 조건이 요인의 한계와 동등한 경우 " 요인 " 개요의 한계가 업데이트되었음을 알리는 JMP 경고가 나타납니다 .

허용되지 않는 조합 필터 사용

이 옵션은 허용되지 않는 조합을 쉽게 지정하기 위해 데이터 필터를 조정하여 사용합니다. 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

"필터 요인 추가" 목록에서 요인을 선택하고 **추가**를 클릭합니다. 그런 다음 슬라이더를 사용하거나(연속형 요인의 경우) 수준을 선택하여(범주형 요인의 경우) 허용되지 않는 조합을 지정합니다.

"필터 요인 추가" 메뉴의 빨간색 삼각형 옵션은 많은 플랫폼 시작 창의 "열 선택" 패널에 있는 옵션과 동일합니다. 자세한 내용은 JMP 사용의 "시작하기" 장에서 확인하십시오.

" 추가 " 를 클릭하면 " 허용되지 않는 조합 " 제어판에 선택한 요인이 표시되고 추가 제어 옵션이 제공됩니다 . 요인은 모델링 유형에 따라 다음과 같이 나타납니다 .

연속형 요인 연속형 요인의 경우 요인 설정 범위를 따라 양단 슬라이더가 나타납니다 . 슬라이더 끝을 드래그하거나 슬라이더 아래의 텍스트 값을 클릭하여 끝점을 설정함으로써 허용되지 않은 설정을 지정할 수 있습니다 . 슬라이더에서 파란색 강조 표시는 허용되지 않는 값을 나타냅니다 .

범주형 요인 범주형 요인의 경우 가능한 수준은 라벨이 지정된 블록으로 표시되거나 , 수준 수가 많으면 목록 항목으로 표시됩니다 . 허용하지 않을 수준을 선택합니다 . 여러 수준을 선택하려면 Ctrl 키를 누른 채로 선택합니다 . 해당 블록 또는 목록 항목은 강조 표시되어 허용되지 않는 수준임을 나타냅니다 . " 허용되지 않는 조합 " 패널에 범주형 요인을 추가하면 해당 요인의 수준 수가 요인 이름 뒤의 괄호 안에 표시됩니다 .

허용되지 않는 조합 옵션

제어판에는 다음과 같은 컨트롤이 있습니다 .

지우기 지정한 허용되지 않는 요인 수준 설정을 모두 지웁니다 . 선택한 요인은 지워지지 않습니다 .

다시 시작 선택한 모든 요인을 제거하고 초기 요인 목록으로 돌아갑니다 .

AND " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 AND 그룹이 됩니다 . AND 그룹 내에 지정된 요인 수준 조합은 허용되지 않습니다 .

나중에 AND 그룹에 요인을 추가하려면 그룹 개요를 클릭하여 강조 표시된 직사각형을 표시합니다 . 그런 다음 "AND" 를 선택하고 요인을 추가합니다 .

단일 요인을 제거하려면 빨간색 삼각형 메뉴에서 **삭제**를 선택합니다 .

OR " 필터 요인 추가 " 목록을 엽니다 . 선택한 요인은 별도의 AND 그룹이 됩니다 . OR 로 구분된 AND 그룹의 경우 하나 이상의 AND 그룹에 지정된 조합은 허용되지 않습니다 .

요인에 대한 빨간색 삼각형 메뉴 옵션

한 요인이 여러 OR 그룹에 나타날 수 있습니다. 특정 OR 그룹에 요인이 나타나는 경우 이를 해당 요인의 인스턴스라고 합니다.

삭제 "허용되지 않는 조합" 패널에서 선택한 요인 인스턴스를 제거합니다.

선택 해제 요인 인스턴스에 대한 선택을 취소합니다.

선택 반전 요인 인스턴스에 대해 선택된 값을 선택 취소하고 이전에 선택되지 않은 값을 선택합니다.

표시 옵션 범주형 요인에만 사용할 수 있습니다. 표시 모양을 변경합니다. 각 수준을 블록, 목록 또는 단일 범주로 표시하거나 각 값 옆에 체크박스를 추가하는 옵션을 사용할 수 있습니다.

찾기 범주형 요인에만 사용할 수 있습니다. 요인 이름 아래에 요인 수준에 대한 검색 문자열을 입력할 수 있는 텍스트 상자를 제공합니다. Enter 키를 누르거나 텍스트 상자 바깥쪽을 클릭하면 검색이 수행됩니다. **찾기**가 선택되면 찾기 텍스트 상자 드롭다운 메뉴에 찾기 옵션이 나타납니다.

허용되지 않는 조합 스크립트 사용

이 옵션을 사용하면 JSL 스크립트를 사용하여 요인 수준 조합을 허용하지 않도록 설정할 수 있습니다. 이 옵션은 연속형 요인 또는 연속형과 범주형이 혼합된 요인과 함께 사용할 수 있습니다.

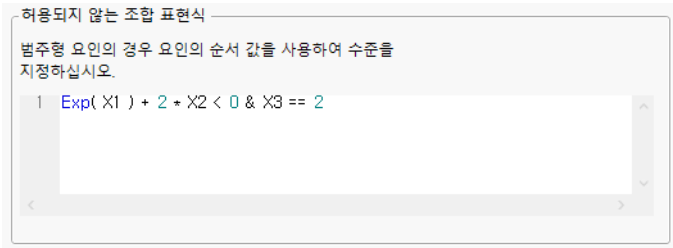
이 옵션은 허용하지 않을 조합을 식별하는 스크립트를 삽입할 수 있는 스크립트 창을 엽니다. 스크립트는 부울 표현식으로 평가되어야 합니다. 표현식이 true로 평가되면 지정된 조합이 허용되지 않습니다.

범주형 요인에 대한 표현식을 작성할 때는 수준의 순서 값 또는 수준 이름을 사용합니다. "요인" 개요에서 요인 수준이 높음, 중간, 낮음으로 순서대로 지정된 경우 관련 서수 값은 1, 2, 3 입니다. 예를 들어 두 개의 연속형 요인 (X1, X2) 과 세 가지 수준 (L1, L2, L3) 이 순서대로 설정된 범주형 요인 하나가 있다고 가정해 보겠습니다. 다음 표현식이 true 인 경우 수준을 허용하지 않으려고 합니다.

$$e^{X_1} + 2X_2 < 0 \text{ 및 } X_3 = L2$$

`(Exp(X1) + 2*X2 < 0) & (X3 == 2)` 표현식을 스크립트 창에 입력합니다.

그림 21.5 스크립트 편집기의 표현식



(그림에서 불필요한 괄호는 파싱을 통해 제거됨) 함수를 부울 표현식의 일부로 입력할 수 있습니다. (Exp(X1) + 2*X2 < 0) & (X3 == "L2") 표현식도 유효합니다.

공간 채움 설계 방법

다음과 같은 방법을 사용하여 공간 채움 설계를 구성할 수 있습니다 .

- "Sphere-Packing 설계 "(578 페이지)
- " 라틴 초입방체 설계 "(581 페이지)
- " 균등 설계 "(584 페이지)
- " 최소 잠재 설계 "(587 페이지)
- " 최대 엔트로피 설계 "(589 페이지)
- " 가우스 과정 IMSE 최적 설계 "(591 페이지)
- " 쾌속 유연 채움 설계 "(592 페이지)

설계

"설계" 개요에서는 공간 채움 설계를 위한 런을 보여 줍니다.

설계 진단

"설계 진단" 개요에는 공간 채움 설계 요인의 값이 0 ~ 1 척도로 표시됩니다. "최소 거리"는 척도화된 값을 기반으로 하며 각 점에서 최근접 이웃까지의 최소 거리입니다. 최근접 이웃의 행 번호는 "최근접 점" 열에 제공됩니다. 테이블 아래에 표시된 불일치 값은 설계점과 균등 분포 사이의 통합 차이입니다.

MaxPro(최대 투영) 기준은 전체 설계 및 사용된 범주형 변수의 각 수준에 대해 제공됩니다. 값이 작을수록 좋습니다. MaxPro 기준에 대한 자세한 내용은 "MaxPro"(592페이지)에서 확인하십시오.

참고 : 임의의 차원에 대해 두 점이 동일한 값 ($x_{ik} = x_{jk}$) 을 가질 수 있으면 MaxPro 기준이 정의되지 않습니다 . 이 경우에는 " 설계 진단 " 에 MaxPro 대한 결측값이 표시됩니다 .

설계 테이블

"설계 테이블" 버튼을 사용하여 공간 채움 설계 구성을 완료합니다.

테이블 생성 " 공간 채움 설계 " 데이터 테이블을 생성합니다 .

뒤로 **설계 생성**을 클릭하기 전의 위치로 이동합니다 . 이전 개요를 변경하고 설계를 다시 생성할 수 있습니다 .

공간 채움 설계 옵션

공간 채움 설계 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

반응 저장 " 반응 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 . 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다 .

반응 불러오기 " 반응 저장 " 옵션을 사용하여 저장한 반응을 불러옵니다 .

요인 저장 " 요인 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 각 요인의 열에는 수준이 포함됩니다 . 다른 정보는 열 특성으로 저장됩니다 . 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다 . 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다 . 그런 다음 " 설계 역할 " 영역에서 적절한 역할을 선택합니다 .

요인 불러오기 " 요인 저장 " 옵션을 사용하여 저장한 요인을 불러옵니다 .

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) " 요인 제약 조건 정의 " 또는 " 선형 제약 조건 " 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다 . 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다 . 마지막 행은 부등식 한계를 포함합니다 . 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다 . 자세한 내용은 " 열 특성 " 부록의 " 제약 조건 상태 "(762 페이지) 에서 확인하십시오 .

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) " 제약 조건 저장 " 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다 .

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다 . 이러한 작업에는 다음 중 하나 이상이 포함됩니다 .

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다 .

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자 " 스크립트에 포함됩니다 .

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다 .

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다 .
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다 .

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다 .

- **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다. 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 기본 분석의 " 시뮬레이션 " 장에서 확인하십시오 .

FFF 최적 기준 쾌속 유연 채움 설계 방법의 경우 MaxPro 기준 (기본값) 과 Centroid 기준 중 하나를 선택할 수 있습니다 . 자세한 내용은 "FFF 최적 기준 "(592 페이지) 에서 확인하십시오 .

시작 수 새 설계를 생성하기 위해 선택한 설계 유형의 알고리즘이 시작되는 횟수를 지정합니다 . 지정된 설계 유형에 대한 기준에 따라 최선의 설계가 반환됩니다 . 기본적으로 모든 설계 유형에 대해 1 로 설정합니다 . 쾌속 유연 채움 설계에는 사용되지 않습니다 .

고급 옵션 > 평균 군집 크기 설정 쾌속 유연 채움 설계 방법의 경우 각 군집 또는 각 설계점을 정의하는 데 사용하기 위해 무작위로 생성되는 평균 점 개수를 지정할 수 있습니다 .

고급 옵션 > MaxPro 범주형 가중치 쾌속 유연 채움 설계 방법의 경우 MaxPro 계산에서 범주형 요인에 사용되는 가중치를 지정할 수 있습니다. 범주형 수준이 동일한 점에 가중치를 곱합니다. 이렇게 하면 범주형 요인의 각 수준에 대한 하위 설계에서 더 좋은 공간 채움 특성이 제공됩니다.

스크립트를 스크립트 창에 저장 공간 채움 설계 플랫폼에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다.

Sphere-Packing 설계

Sphere-Packing 설계 방법은 설계점 쌍 사이의 최소 거리를 최대화합니다. 이 최대화를 통해 설계 영역 내에서 점을 가능한 한 많이 퍼뜨릴 수 있습니다.

- "Sphere-Packing 설계 생성 "
- "Sphere-Packing 설계 시각화 "

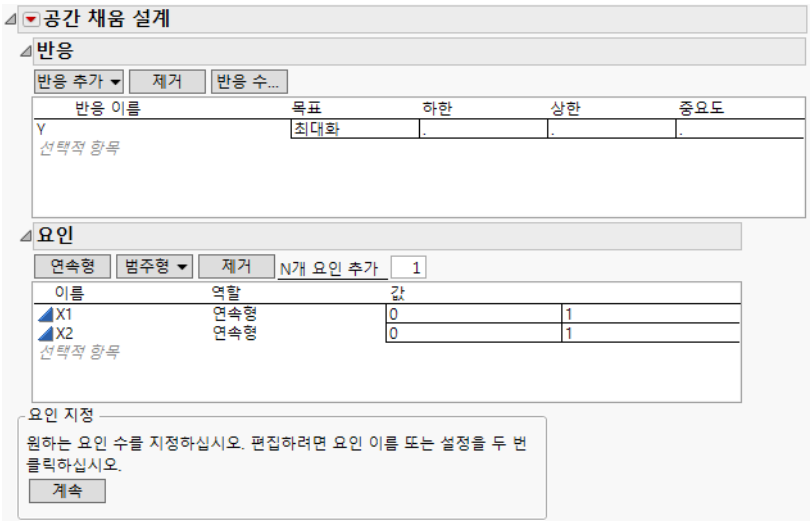
Sphere-Packing 설계 생성

Sphere-Packing 설계를 생성하려면 다음 단계를 수행하십시오.

1. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다.
2. 반응과 요인을 입력합니다.

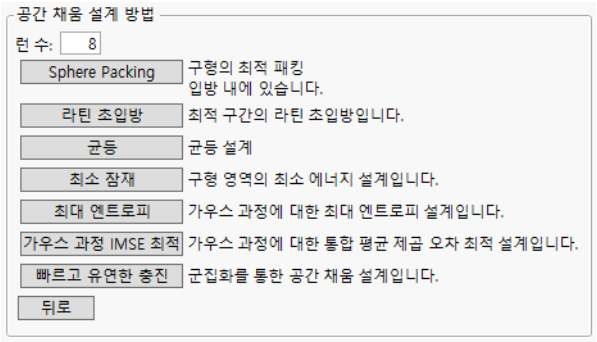
   자세한 내용은 " 반응 "(568 페이지) 에서 확인하십시오.
3. 필요한 경우 요인 수준 값을 변경합니다. 예를 들어 그림 21.6 에서는 값이 –1 ~ 1(기본값) 대신 0 ~ 1 범위로 설정된 기존의 두 요인 X1 과 X2 를 보여 줍니다.

그림 21.6 두 요인에 대한 공간 채움 설계 대화상자



4. **계속**을 클릭합니다 .
5. 설계 규격 대화상자에서 표본 크기 (**런 수**) 를 지정합니다 . 그림 21.7 에는 표본 크기가 8 로 지정되어 있습니다 .

그림 21.7 공간 채움 설계 대화상자



6. **Sphere Packing** 을 클릭합니다 .

   설계가 생성되고 설계 런과 설계 진단이 표시됩니다 . 그림 21.8 에서는 **최소 거리**가 0.518 인 " 설계 진단 " 패널을 보여 줍니다 . 결과가 아래 그림과 약간 다를 수 있지만 최소 거리는 같습니다 . 한 차원에 대해 점의 값이 동일할 수 있으므로 Sphere Packing 에 대한 MaxPro 통계량은 정의되지 않았습니다 .

그림 21.8 Sphere-Packing 설계 진단

7. **테이블 생성**을 클릭합니다 . 이 테이블을 사용하여 다음 섹션에서 설명하는 시각화 예를 완료합니다 .

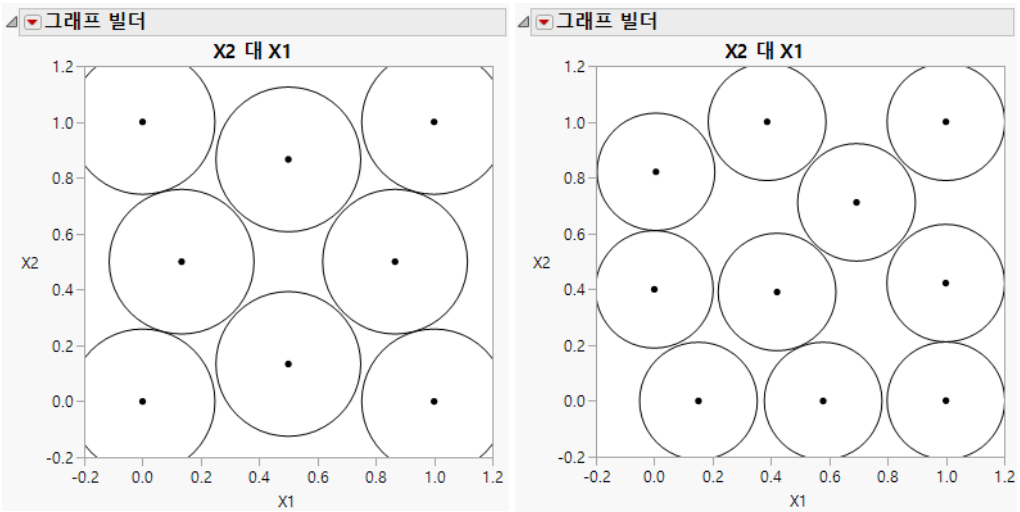
Sphere-Packing 설계 시각화

Sphere-Packing 설계를 시각화하려면 Sphere-Packing 설계 테이블을 사용하여 다음과 같이 그래프 빌더로 그림을 생성합니다 .

1. **그래프 > 그래프 빌더**를 선택합니다 .
2. X1 을 **X** 로 지정하고 X2 를 **Y** 로 지정합니다 .
3. 평활기 아이콘을 **클릭**하여 평활기 (파란색 선) 를 그래프에서 제거합니다 .
4. 그림을 마우스 오른쪽 버튼으로 클릭하고 **그래프 > 크기 / 척도 > 등간격 가로 - 세로 축**을 선택하여 프레임이 정사각형이 되도록 프레임 크기를 조정합니다 .
5. 그림을 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다 . "사용자 정의 " 패널이 나타나면 더하기 기호를 클릭하여 텍스트 편집 영역을 표시한 후 다음 스크립트를 입력합니다 .
   ```
   For Each Row(Circle({:X1, :X2}, 0.518/2))
   ```
   여기서 0.518 은 " 설계 진단 " 패널에서 확인한 최소 거리 값입니다 . 이 스크립트는 그림 21.9 의 왼쪽에 표시된 것과 같이 각 설계점이 중심에 있고 반지름이 0.259(지름 0.518 의 1/2) 인 원을 그립니다 . 이 그림은 JMP 가 설계점을 효율적으로 채우는 방법을 보여 줍니다 .
6. 이제 이전 섹션에서 설명한 절차를 표본 크기만 8 대신 10 으로 지정하고 똑같이 반복합니다 .

   그래픽 스크립트에서 0.518 을 10 회 런으로 생성된 최소 거리로 변경해야 합니다 . 그림이 나타나면 다시 프레임 크기를 설정하고 , 진단 보고서의 최소 거리를 원의 지름으로 사용하여 그래픽 스크립트를 생성합니다 . 그림 21.9 의 오른쪽 그래프와 유사한 그래프가 표시됩니다 . Sphere Packing 의 불규칙 특성에 유의하십시오 . 배열은 랜덤 시작점에 기반하므로 실제로 이 과정을 세 번 정도 반복하면 그림이 약간 다를 수 있습니다 .

그림 21.9 8 회 런 (왼쪽) 과 10 회 런 (오른쪽) 을 사용한 Sphere-Packing 예



라틴 초입방체 설계

라틴 초입방체 설계에서 각 요인의 수준 수는 설계의 런 수와 동일합니다 . 수준은 요인의 하한에서 상한까지 간격이 균등합니다 . Sphere Packing 방법과 마찬가지로 라틴 초입방체 방법도 설계 점 사이의 최소 거리를 최대화하도록 점을 선택하지만 제약 조건이 있습니다 . 이 제약 조건으로 인해 요인 수준 사이 간격이 균등하게 유지됩니다 .

- " 라틴 초입방체 설계 생성 "
- " 라틴 초입방체 설계 시각화 "

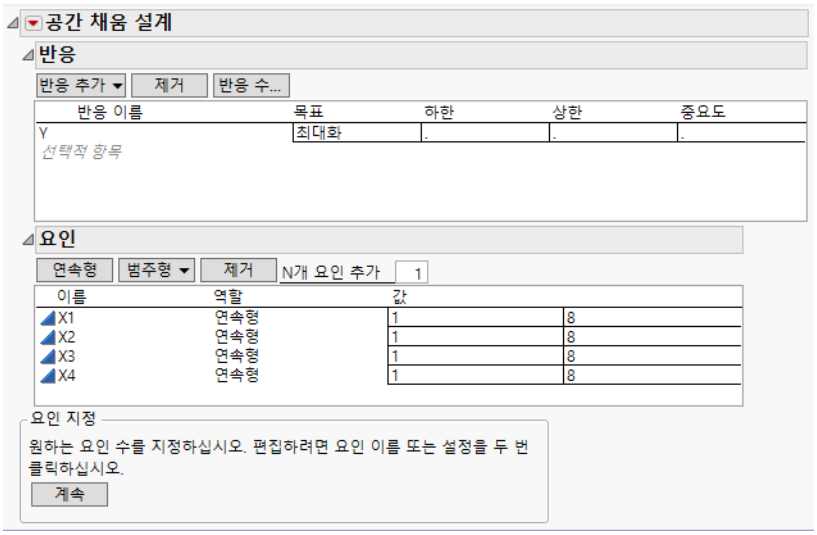
라틴 초입방체 설계 생성

라틴 초입방체 설계를 생성하려면 다음 단계를 수행하십시오 .

1. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다 .
2. 반응과 요인 (필요한 경우) 을 입력합니다 .

   자세한 내용은 " 반응 "(568 페이지) 에서 확인하십시오 .
3. 필요한 경우 요인 수준 값을 변경합니다 . 그림 21.10 에서는 요인 네 개를 추가하고 값을 기본 값 –1 과 1 대신 1 과 8 로 변경합니다 .

그림 21.10 네 요인에 대한 공간 채움 설계 대화상자



4. **계속**을 클릭합니다 .
5. 설계 규격 대화상자에서 표본 크기 (**런 수**) 를 지정합니다 . 이 예에서는 표본 크기 8 을 사용합니다 .

6. **라틴 초입방체**를 클릭합니다 (그림 21.7 참조). 그림 21.11 에 표시된 것과 유사한 요인 설정 및 설계 진단 결과가 나타납니다 . 이 그림은 요인이 4 개이고 런 수가 8 인 라틴 초입방체 설계를 보여 줍니다 .

참고 : 이 예의 목적은 각 열 (요인) 에 각 수준이 한 번만 할당되고 각 열의 수준 순열이 다름을 보여 주는 것입니다 .

그림 21.11 수준이 8 개인 4 개 요인과 8 회 런에 대한 라틴 초입방체 설계

4 요인
공간 채움 라틴 초입방

설계

런	X1	X2	X3	X4
1	4.00000	7.00000	1.00000	4.00000
2	2.00000	6.00000	5.00000	7.00000
3	1.00000	4.00000	6.00000	1.00000
4	7.00000	5.00000	3.00000	8.00000
5	5.00000	2.00000	8.00000	6.00000
6	6.00000	8.00000	7.00000	3.00000
7	8.00000	3.00000	4.00000	2.00000
8	3.00000	1.00000	2.00000	5.00000

설계 진단

런	척도화X1	척도화X2	척도화X3	척도화X4	최소 거리	최근접 점
1	0.42857	0.85714	0.00000	0.42857	0.782	2
2	0.14286	0.71429	0.57143	0.85714	0.782	1
3	0.00000	0.42857	0.71429	0.00000	0.926	2
4	0.85714	0.57143	0.28571	1.00000	0.795	2
5	0.57143	0.14286	1.00000	0.71429	0.845	2
6	0.71429	1.00000	0.85714	0.28571	0.892	7
7	1.00000	0.28571	0.42857	0.14286	0.892	6
8	0.28571	0.00000	0.14286	0.57143	0.892	1

불일치 = 0.039317
MaxPro
전체 설계 11.40992

라틴 초입방체 설계 시각화

라틴 초입방체 기법의 특성을 시각화하려면 그래프 빌더를 사용하여 그림을 생성합니다 .

1. 두 개 요인을 사용하는 또 다른 라틴 초입방체 설계를 생성합니다 .
2. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다 .
3. 두 개 요인을 입력하고 요인 값을 기본값 –1 과 1 대신 0 과 1 이 되도록 변경합니다 .
4. **계속**을 클릭합니다 .
5. 표본 크기 (**런 수**) 를 8 로 지정합니다 .
6. **라틴 초입방체**를 클릭합니다 .

그림 21.12 2 개 요인과 8 회 런을 사용한 라틴 초입방체 설계

2 요인
공간 채움 라틴 초입방

설계

런	X1	X2
1	0.28571	0.00000
2	0.00000	0.85714
3	0.57143	0.28571
4	0.42857	0.71429
5	0.14286	0.42857
6	1.00000	0.14286
7	0.85714	0.57143
8	0.71429	1.00000

설계 진단

런	척도화X1	척도화X2	최소 거리	최근접 점
1	0.28571	0.00000	0.404	3
2	0.00000	0.85714	0.452	4
3	0.57143	0.28571	0.404	1
4	0.42857	0.71429	0.404	5
5	0.14286	0.42857	0.404	4
6	1.00000	0.14286	0.452	7
7	0.85714	0.57143	0.404	3
8	0.71429	1.00000	0.404	4

불일치 = 0.009279
MaxPro
전체 설계 11.32801

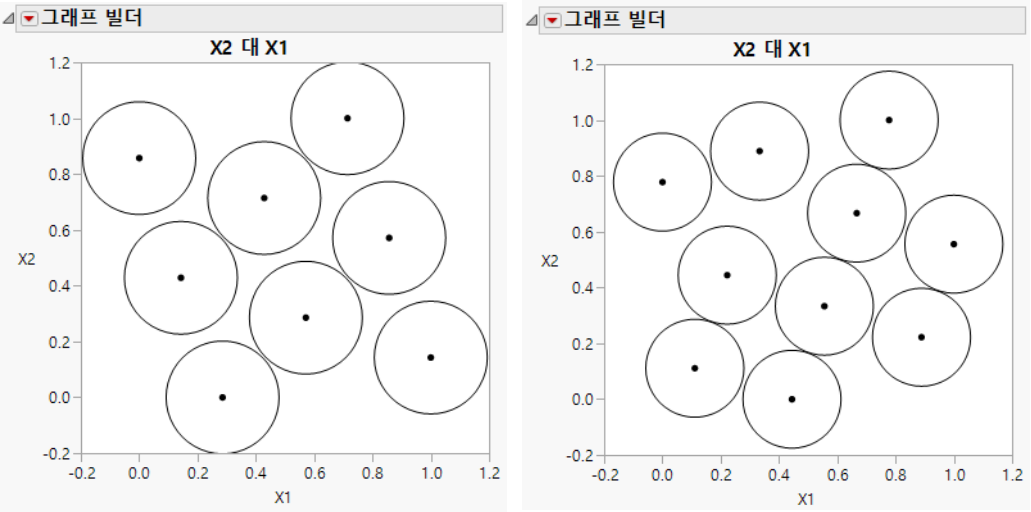
7. **테이블 생성**을 클릭합니다 .
8. **그래프 > 그래프 빌더**를 선택합니다 .
9. X1 을 **X** 로 지정하고 X2 를 **Y** 로 지정합니다 .
10. 평활기 아이콘을 클릭하여 평활기 (파란색 선) 를 그래프에서 제거합니다 .
11. 그림을 마우스 오른쪽 버튼으로 클릭하고 **그래프 > 크기 / 척도 > 등간격 가로 - 세로 축**을 선택하여 프레임이 정사각형이 되도록 프레임 크기를 조정합니다 .
12. 그림을 마우스 오른쪽 버튼으로 클릭하고 메뉴에서 **사용자 정의**를 선택합니다 . " 사용자 정의 " 패널에서 큰 더하기 기호를 클릭하여 텍스트 편집 영역을 표시한 후 다음 스크립트를 입력합니다 .

    ```
    For Each Row(Circle({:X1, :X2}, 0.404/2))
    ```

    여기서 `0.404` 는 " 설계 진단 " 패널에서 확인한 최소 거리 값입니다 (그림 21.12 참조). 이 스크립트는 그림 21.13 의 왼쪽에 표시된 것과 같이 각 설계점이 중심에 있고 반지름이 0.202(지름 0.404 의 1/2) 인 원을 그립니다 . 이 그림은 JMP 가 설계점을 효율적으로 채우는 방법을 보여 줍니다 .
13. 위의 절차를 런 수만 8 대신 10 으로 지정하고 똑같이 반복합니다 (5 단계). 그래픽 스크립트에서 `0.404` 를 10 회 런으로 생성된 최소 거리로 변경해야 합니다 .

그림 21.13 의 오른쪽 그래프와 유사한 그래프가 표시됩니다 . Sphere Packing 의 불규칙 특성에 유의하십시오 . 배열은 랜덤 시작점에 기반하므로 실제로 이 과정을 반복하면 그림이 약간 다를 수 있습니다 .

그림 21.13 8 회 런 (왼쪽) 과 10 회 런 (오른쪽) 을 사용한 라틴 초입방체 설계 비교



라틴 초입방체 설계에서 각 점 쌍 사이의 최소 거리는 Sphere-Packing 설계의 경우보다 작습니다. 라틴 초입방체 설계는 각 요인의 수준 간격이 균등하도록 제한하기 때문입니다. Sphere-Packing 설계는 제약 조건 없이 최소 거리를 최대화합니다.

균등 설계

균등 설계는 설계점(경험적 균등 분포)과 이론적 균등 분포 사이의 불일치를 최소화합니다.

참고 : 이 설계는 미지의 함수 적분에 대해 정확한 단순 추정값을 얻는 데 가장 유용합니다 . 추정값은 실험에서 관측한 반응의 평균입니다 .

1. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다 .
2. 반응과 요인 (필요한 경우) 을 입력합니다 .

   자세한 내용은 " 반응 "(568 페이지) 에서 확인하십시오 .
3. 요인 수준 값을 0 과 1 로 변경합니다 .
4. **계속**을 클릭합니다 .
5. 설계 규격 대화상자에서 표본 크기를 지정합니다 . 이 예에서는 표본 크기 (**런 수**) 8 을 사용합니다 .
6. **균등** 버튼을 클릭합니다. 설계가 생성되고 설계 런과 설계 진단이 표시됩니다(그림 21.14 참조).

참고 : 균등 분포의 주안점은 점을 퍼뜨리는 것이 아닙니다 . 그림 21.14 의 최소 거리는 서로 상당히 다릅니다 .

그림 21.14 8 회 런을 사용한 균등 공간 채움 설계의 요인 설정 및 진단

2 요인
공간 채움 균등 설계

설계

런	X1	X2
1	0.93092	0.69041
2	0.69276	0.18514
3	0.06860	0.32218
4	0.18696	0.81304
5	0.56496	0.93540
6	0.81410	0.43443
7	0.32218	0.06860
8	0.43643	0.56357

설계 진단

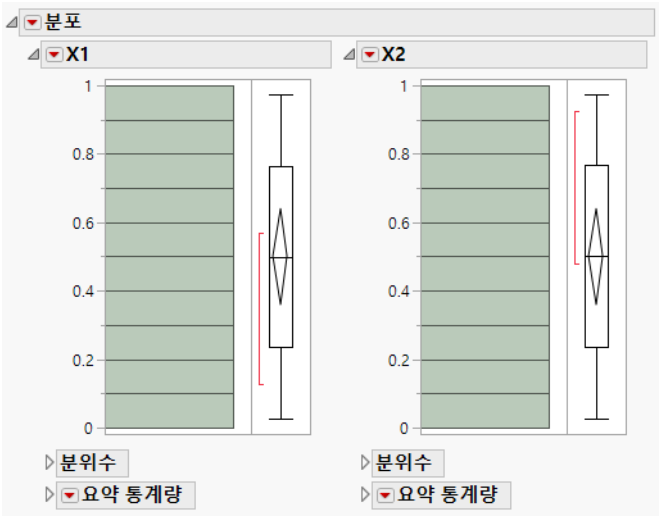
런	척도화X1	척도화X2	최소 거리	최근접 점
1	0.93092	0.69041	0.281	6
2	0.69276	0.18514	0.277	6
3	0.06860	0.32218	0.359	7
4	0.18696	0.81304	0.353	8
5	0.56496	0.93540	0.393	8
6	0.81410	0.43443	0.277	2
7	0.32218	0.06860	0.359	3
8	0.43643	0.56357	0.353	4

불일치 = 0.004663
MaxPro
전체 설계 15.40271

7. **테이블 생성**을 클릭합니다.

균등 설계를 생성해도 요인 수준 간격이 균등하지 않을 수 있습니다. 그러나 런 수를 늘리고 각 요인에 대해 분포를 실행하면(**분석** > **분포** 사용) 평평한 히스토그램이 표시됩니다.

그림 21.15 런 수가 20 으로 증가한 경우 각 요인에 대한 평평한 히스토그램



Sphere-Packing, 라틴 초입방체 및 균등 방법 비교

공간 채움 설계 방법을 비교하기 위해 이전 예에 나와 있는 대로 **Sphere Packing**, **라틴 초입방체** 및 **균등** 설계를 생성합니다. "설계 진단" 테이블에 요인의 값이 0 ~ 1 척도로 표시됩니다. "최소 거리"는 척도화된 값을 기반으로 하며 각 점에서 최근접 이웃까지의 최소 거리입니다. 불일치 값은 설계점과 균등 분포 사이의 통합 차이입니다.

그림 21.16에서는 세 가지 8회 런 공간 채움 설계의 설계 진단 비교를 보여 줍니다. 균등 설계의 불일치가 가장 작습니다(최선). Sphere-Packing 설계의 불일치가 가장 큽니다(최악). 라틴 초입방체 설계의 불일치는 최적값에 더 가까운 중간 값입니다. MaxPro 기준은 균등 설계에 비해 라틴 초입방체 설계에서 더 작습니다.

또한 점 쌍 사이의 최소 거리는 Sphere-Packing 방법을 사용할 때 가장 큽니다(최선). 균등 설계의 점 쌍은 약 절반 씩만 떨어져 있습니다. 라틴 초입방체 설계는 점을 퍼트리는 데 있어서 Sphere-Packing 설계와 더 비슷하게 동작합니다.

퍼트리기와 불일치 둘 다를 기준으로 볼 때 라틴 초입방체 설계에서 양호한 절충안을 나타냅니다.

그림 21.16 세 가지 8 회 런 공간 채움 방법의 진단 비교

공간 채움 Sphere Packing

▷ 설계

◢ 설계 진단

런	척도화X1	척도화X2	최소 거리	최근접 점
1	0.50000	0.86602	0.518	3
2	0.86603	0.50000	0.518	6
3	0.13397	0.50000	0.518	1
4	0.50000	0.13398	0.518	7
5	0.00000	1.00000	0.518	1
6	1.00000	1.00000	0.518	2
7	1.00000	0.00000	0.518	4
8	0.00000	0.00000	0.518	3

불일치 = 0.046969
MaxPro
전체 설계 .

공간 채움 라틴 초입방

▷ 설계

◢ 설계 진단

런	척도화X1	척도화X2	최소 거리	최근접 점
1	0.28571	0.00000	0.404	3
2	0.00000	0.85714	0.452	4
3	0.57143	0.28571	0.404	1
4	0.42857	0.71429	0.404	5
5	0.14286	0.42857	0.404	4
6	1.00000	0.14286	0.452	7
7	0.85714	0.57143	0.404	3
8	0.71429	1.00000	0.404	4

불일치 = 0.009279
MaxPro
전체 설계 11.32801

공간 채움 균등 설계

▷ 설계

◢ 설계 진단

런	척도화X1	척도화X2	최소 거리	최근접 점
1	0.93092	0.69041	0.281	6
2	0.69276	0.18514	0.277	6
3	0.06860	0.32218	0.359	7
4	0.18696	0.81304	0.353	8
5	0.56496	0.93540	0.393	8
6	0.81410	0.43443	0.277	2
7	0.32218	0.06860	0.359	3
8	0.43643	0.56357	0.353	4

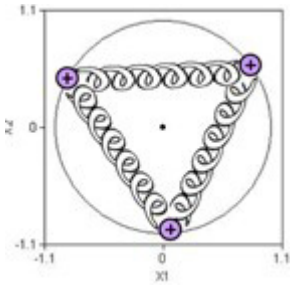
불일치 = 0.004663
MaxPro
전체 설계 15.40271

또 다른 비교 항목은 설계를 계산하는 데 걸리는 시간입니다. 균등 설계 방법을 사용할 때 계산 시간이 가장 오래 걸립니다. 또한 설계 계산 시간은 런 수에 따라 빠르게 증가합니다. 비슷한 문제의 경우, 모든 공간 채움 설계 방법이 사용자 설계 도구의 *D*-최적 설계보다 계산 시간이 더 오래 걸립니다.

최소 잠재 설계

최소 잠재 설계는 구형 안에 점을 펼쳐 놓습니다. 이 설계가 어떻게 생성되는지 이해하기 위해 그림 21.17에 표시된 것과 같이 점을 다른 모든 점에 스프링으로 연결된 전자로 가정해 보겠습니다. 쿨롱 힘은 점을 밀어내지만 스프링은 점을 서로 끌어당깁니다. 이 설계는 시스템의 잠재적 에너지를 최소화하는 점 간격입니다.

그림 21.17 최소 잠재 설계



최소 잠재 설계의 특징

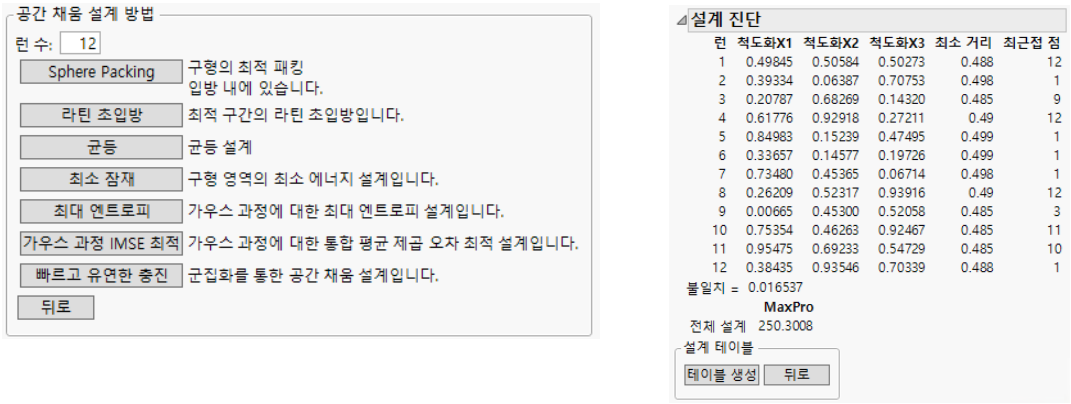
- 구형 대칭을 갖습니다 .
- 거의 직교입니다 .
- 간격이 균등합니다 .

최소 잠재 설계 예를 생성하려면

1. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다 .
2. 연속형 요인을 3 개 추가합니다 .

   자세한 내용은 " 요인 "(570 페이지) 에서 확인하십시오 .
3. 요인 수준 값을 0 과 1 로 변경합니다 .
4. **계속**을 클릭합니다 .
5. 그림 21.18 왼쪽에 표시된 설계 규격 대화상자에서 표본 크기 (**런 수**) 를 입력합니다 . 이 예에서는 표본 크기 12 를 사용합니다 .
6. **최소 잠재** 버튼을 클릭합니다 . 설계가 생성되고 설계 런과 설계 진단이 표시됩니다 (그림 21.18 의 오른쪽 참조).

그림 21.18 최소 잠재 설계를 위한 공간 채움 방법 및 설계 진단



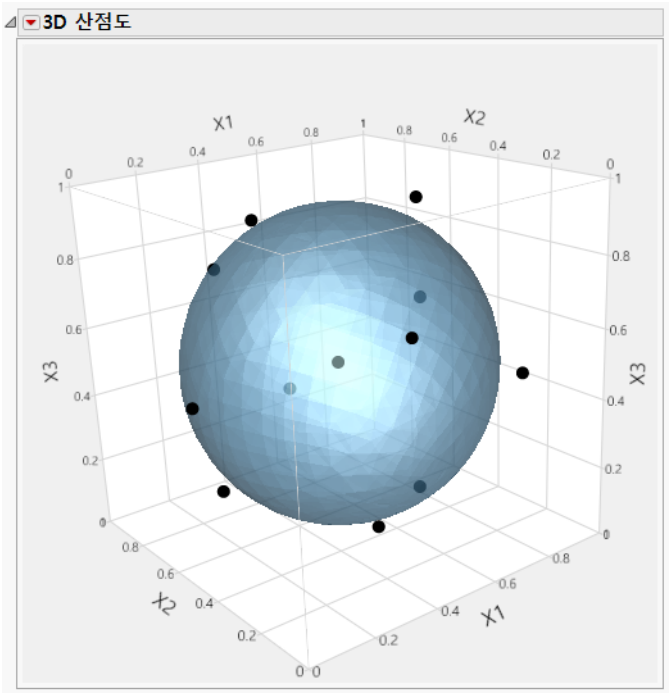
7. **테이블 생성**을 클릭합니다 .

3D 산점도 그래픽 플랫폼을 사용하여 최소 잠재 설계의 구형 대칭을 확인할 수 있습니다 .

1. JMP 설계 테이블을 생성한 후 **그래프** > **3D 산점도** 명령을 선택합니다 .
2. "3D 산점도 " 시작 대화상자에서 X1, X2, X3 을 **Y, 열**로 선택하고 **확인**을 클릭하여 설계점의 초기 3 차원 산점도를 표시합니다 .
3. 그림 21.19 에 나오는 것과 유사한 결과를 표시하려면 다음을 수행합니다 .
   - "3D 산점도 " 의 빨간색 삼각형을 클릭하고 **정규 등고선 타원**을 선택합니다 .
   - 그림을 마우스 오른쪽 버튼으로 클릭하고 **설정**을 선택한 후 **표식 크기** 슬라이더를 사용하여 표식 크기를 늘립니다 .

이제 점이 타원 표면에 균등하게 펼쳐지는 것을 시각화할 수 있습니다.

그림 21.19 구면의 최소 잠재 설계점

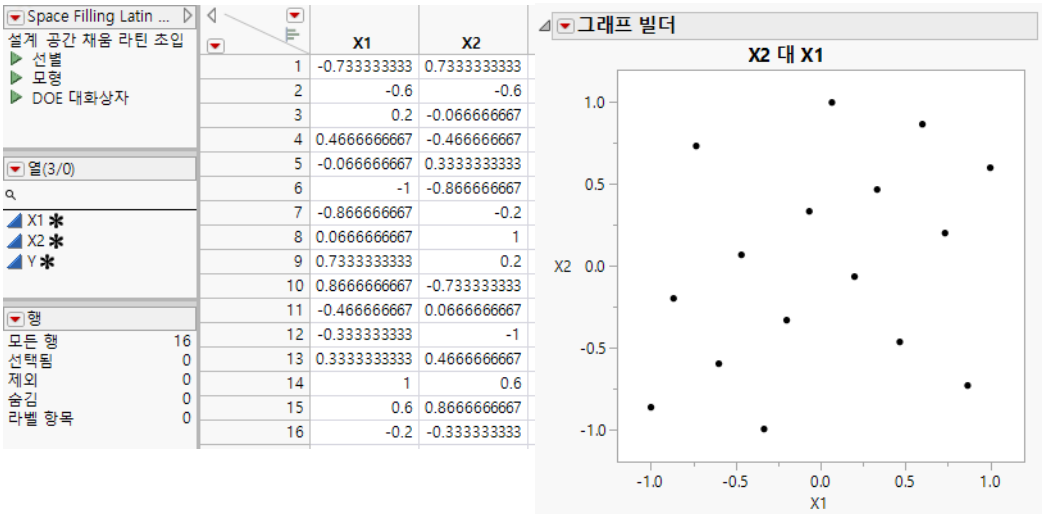


최대 엔트로피 설계

최대 엔트로피 설계는 컴퓨터 실험에서 라틴 초입방 설계 대신 사용할 수 있습니다. 라틴 초입방 설계는 가우스 과정 모형과 함께 많이 사용되는 설계입니다. 컴퓨터 시뮬레이션 전문가는 좌표축에 대한 모든 투영이 균등한 라틴 초입방체 설계를 선호합니다.

그러나 그림 21.20의 예에서 볼 수 있듯이 라틴 초입방체 설계는 최적의 공간 채움 특성을 제공하지 못할 수 있습니다. 이 예는 요인 수준이 -1 ~ 1 사이의 값으로 설정된 2요인 16회 런 라틴 초입방체 설계입니다. 두 요인에 대한 그림은 이 설계에 포함 범위가 없는 영역이 있음을 보여 줍니다. 특히 0에 가까운 X1과 -1에 가까운 X2의 포함 범위가 적절하지 않습니다.

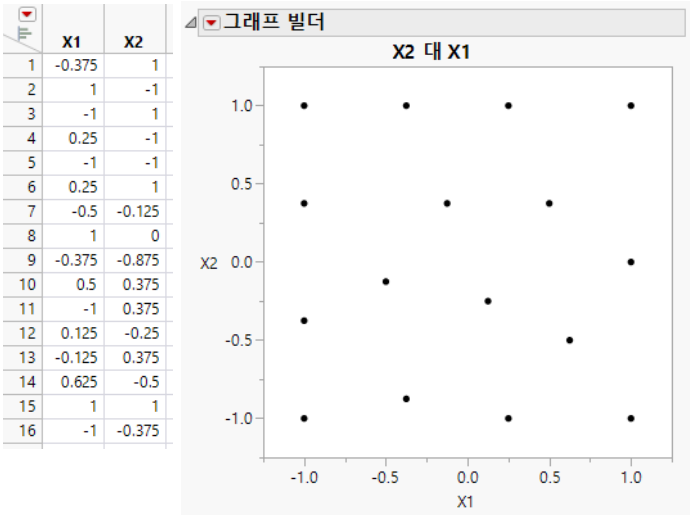
그림 21.20 2 요인 라틴 초입방체 설계



	X1	X2
1	-0.733333333	0.7333333333
2	-0.6	-0.6
3	0.2	-0.066666667
4	0.4666666667	-0.466666667
5	-0.066666667	0.3333333333
6	-1	-0.866666667
7	-0.866666667	-0.2
8	0.0666666667	1
9	0.7333333333	0.2
10	0.8666666667	-0.733333333
11	-0.466666667	0.0666666667
12	-0.333333333	-1
13	0.3333333333	0.4666666667
14	1	0.6
15	0.6	0.8666666667
16	-0.2	-0.333333333

최대 엔트로피 설계는 실험에 포함된 정보량을 최적화합니다. 자세한 내용은 아래 기술 정보를 참조하십시오. 요인 수준이 -1 ~ 1 사이로 설정된 경우, 그림 21.21에 표시된 2요인 최대 엔트로피 설계가 그림 21.20의 라틴 초입방체 설계보다 영역을 더 잘 포함합니다. 런 수가 증가할수록 설계의 공간 채움 특성이 향상됩니다.

그림 21.21 2 요인 최대 엔트로피 설계



	X1	X2
1	-0.375	1
2	1	-1
3	-1	1
4	0.25	-1
5	-1	-1
6	0.25	1
7	-0.5	-0.125
8	1	0
9	-0.375	-0.875
10	0.5	0.375
11	-1	0.375
12	0.125	-0.25
13	-0.125	0.375
14	0.625	-0.5
15	1	1
16	-1	-0.375

기술 정보 최대 엔트로피 설계는 데이터를 정규 분포 (m, s^2 R) 에서 수집한 것으로 가정하고 실험의 샤논 정보 (Shewry and Wynn(1987)) 를 최대화합니다 .

$$R_{ij} = \exp\left(-\sum_k \theta_k (x_{ik} - x_{jk})^2\right)$$

위의 식은 서로 다른 두 설계점 (x_i, x_j)에서 반응 값의 상관입니다. 계산상으로 이 설계는 |**R**|, 즉 표본 상관 행렬의 행렬식을 최대화합니다. x_i와 x_j가 멀리 떨어져 있으면 $\mathbf{R}_{ij}$는 0에 근접합니다. x_i와 x_j가 서로 가까이 있으면 $\mathbf{R}_{ij}$는 1에 가깝습니다.

가우스 과정 IMSE 최적 설계

가우스 과정 IMSE 최적 설계 방법은 가우스 과정 모형에 적합한 설계를 구성합니다. 가우스 과정 모형은 다양한 표면을 적합시킵니다. 가우스 과정 IMSE 최적 설계는 실험 영역에 대한 가우스 과정 모형의 통합 평균 제곱 오차를 최소화합니다. 가우스 과정 IMSE 최적 설계 방법은 kriging 모형과 유사한 상관 구조를 사용합니다. 자세한 내용은 Jones and Johnson(2009)에서 확인하십시오.

공분산 모수 벡터

가우스 과정 모형의 가우스 과정 IMSE 최적 설계 배합에서 공분산 모수 벡터는 상관 구조를 결정합니다. 각 요인에 대한 세타가 있습니다. 세타 0은 상관 1에 해당하므로 적합된 표면이 해당 요인의 방향으로 편평하게 됩니다. 세타가 증가하면 상관이 감소하여 요인 방향으로 유연한 표면을 생성할 수 있습니다.

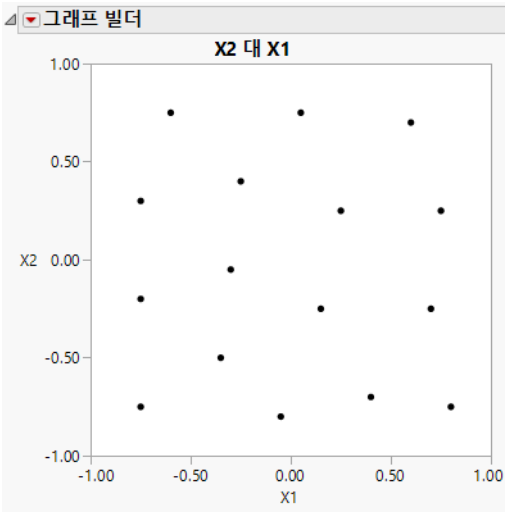
"공분산 모수 벡터" 개요에서 "세타" 아래의 값 목록에 표면 사전 지식을 반영하는 값을 입력할 수 있습니다.

라틴 초입방체 설계와 가우스 과정 IMSE 최적 설계 비교

가우스 과정 IMSE 최적 설계는 라틴 초입방체 설계의 대안입니다. IMSE 최적 설계를 앞에서 그림 21.20에 표시된 라틴 초입방체와 비교할 수 있습니다. 그림 21.22의 테이블과 중첩 그림은 가우스 IMSE 최적 설계를 보여 줍니다. 이 설계에서 요인 영역에 대해 균등한 포함 범위를 제공하는 것을 확인할 수 있습니다.

그림 21.22 2 요인 라틴 초입방체 설계와 가우스 IMSE 최적 설계 비교

	X1	X2
1	-0.05	-0.8
2	-0.75	0.3
3	0.25	0.25
4	0.7	-0.25
5	-0.3	-0.05
6	-0.75	-0.2
7	-0.6	0.75
8	0.6	0.7
9	-0.75	-0.75
10	0.8	-0.75
11	0.4	-0.7
12	-0.35	-0.5
13	-0.25	0.4
14	0.15	-0.25
15	0.75	0.25
16	0.05	0.75



참고 : 최대 엔트로피 설계와 가우스 과정 IMSE 최적 설계 둘 다 100 회의 랜덤 시작을 사용하여 생성되었습니다 .

쾌속 유연 채움 설계

쾌속 유연 채움 설계를 위한 알고리즘은 지정된 설계 영역에서 다수의 랜덤 점을 생성하는 것으로 시작합니다. 그런 다음 빠른 Ward 알고리즘을 사용하여 이러한 점을 여러 군집으로 군집화합니다. 이때 군집 수는 사용자가 지정한 런 수와 같습니다.

최종 설계점은 기본 MaxPro(최대 투영) 최적 기준을 사용하거나 Centroid 기준을 선택하여 얻을 수 있습니다 .

참고 : 범주형 요인 또는 요인 제약 조건이 있으면 공간 채움 설계에서 쾌속 유연 채움 방법만 사용할 수 있습니다 .

FFF 최적 기준

" 공간 채움 설계 " 의 빨간색 삼각형 메뉴에는 FFF 최적 기준 아래에 최적 기준이 포함되어 있습니다 .

MaxPro p 개 요인이 있고 n 이 지정된 런 수와 같은 경우 MaxPro 기준은 군집에서 다음 기준을 최소화하는 점을 찾습니다 .

$$C_{MaxPro} = \sum_{i}^{n-1} \sum_{j=i+1}^{n} \left[1 / \prod_{k=1}^{p} (x_{ik} - x_{jk})^2 \right]$$

MaxPro 기준은 모든 요인을 포함하는 방식으로 잠재적 설계점 간 거리의 곱을 최대화합니다. 이는 요인 투영에 적합한 공간 채움 특성을 제공한다는 목표를 지원합니다. 자세한 내용은 Joseph et al. (2015) 에서 확인하십시오. MaxPro 옵션이 기본값입니다.

Centroid 이 방법은 각 군집의 중심에 설계점을 놓습니다. 설계 공간의 임의 점에서 가장 가까운 인접 설계점까지의 평균 거리가 다른 설계보다 작다는 특성이 있습니다.

참고 : 항상 정해진 최적 기준을 사용하도록 환경 설정을 지정할 수 있습니다. **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택합니다. **FFF 최적 기준**을 선택한 후 기본 기준을 선택합니다.

범주형 요인

범주형 요인이 있는 경우 알고리즘은 다음과 같이 진행됩니다.

- 총 설계점 수는 범주형 요인의 총 수준 조합 수에서 균형을 이룹니다. *m* 개의 수준 조합이 있고 *k* 개의 설계점이 각 조합에 할당된다고 가정해 보겠습니다.
- 연속형 변수에 의해 정의된 설계 공간 내에 다수의 점이 생성됩니다. 이러한 점은 *k* 개의 기본 군집으로 그룹화됩니다.
- 점을 포함하는 *k* 개의 기본 군집은 각각 *m* 개의 하위 군집으로 다시 군집화됩니다.
- 각 기본 군집 내에서 지정된 FFF 최적 기준을 사용하여 *m* 개의 하위 군집 각각에 대해 설계점이 계산됩니다.
- *k* 개의 각 기본 군집에 대해 *m* 개의 수준 조합 중 하나가 *m* 개의 각 하위 군집 설계점에 무작위로 할당됩니다. 이를 통해 총 *km* 개의 설계점이 계산됩니다.
- *k* 개의 각 기본 군집에서 MaxPro 기준에 따라 *m* 개의 각 수준 조합에 대한 설계점이 선택됩니다. 범주형 요인의 수준이 동일한 점에는 MaxPro 범주형 가중치를 곱합니다. MaxPro 범주형 가중치에 대한 설명은 " 공간 채움 설계 옵션 "(576 페이지) 에서 확인하십시오. 이 과정은 *k* 개의 모든 기본 군집에서 10 회 반복하거나, 지정된 설계점을 변경해도 더 이상 개선되지 않을 때까지 계속됩니다.

FFF 설계의 평균 군집 크기 설정

평균 군집 크기 설정 옵션을 사용하면 각 군집 또는 각 설계점을 정의하는 데 사용하기 위해 무작위로 생성되는 평균 점 개수를 지정할 수 있습니다. 이 옵션은 "공간 채움 설계"의 빨간색 삼각형 메뉴에서 "고급 옵션" 아래에 있습니다.

기본적으로 "런 수"를 200 이하로 설정하면 무작위로 생성된 총 10,000 개의 점이 군집화 알고리즘의 기초로 사용됩니다. 런 수가 200 을 초과하면 기본값 50 이 사용됩니다. 이 값을 늘리면 요인 수가 많거나, 허용되지 않는 조합이 군집화 알고리즘에 사용되는 점의 분포를 제한하는 설계에서 특히 유용할 수 있습니다.

참고 : 요인 수와 지정된 런 수 값에 따라 **고급 옵션 > 평균 군집 크기 설정**을 선택하여 설계점당 평균 초기점 수를 늘릴 수 있습니다 .

제약 조건

"요인" 개요가 완료되면 **계속**을 클릭합니다. 그러면 "요인 제약 조건 정의" 개요가 나타납니다. 이 개요를 사용하여 설계 영역을 제한할 수 있습니다. 이 개요에 대한 자세한 내용은 "요인 제약 조건 정의"(572 페이지)에서 확인하십시오.

"허용되지 않는 조합 필터 사용" 및 "허용되지 않는 조합 스크립트 사용" 옵션을 사용하여 허용되지 않는 요인 수준 조합을 지정할 수 있습니다. 또는 "선형 제약 조건 지정" 옵션을 사용하여 선형 부등식에 대한 한계를 지정할 수 있습니다. 그러나 설계는 이 두 방법에 대해 다르게 생성됩니다.

허용되지 않는 조합 필터 사용 및 허용되지 않는 조합 스크립트 사용

허용되지 않는 조합을 지정하면 군집화 알고리즘의 기초를 구성하는 랜덤 점이 무제약 설계 영역 내에 무작위로 분포됩니다. 그런 다음 허용되지 않는 점이 제거되고 나머지 점으로 군집화가 진행됩니다.

참고 : 무작위로 생성된 초기점에 의한 무제약 설계 공간의 기본 포함 범위는 제약 조건의 특성과 지정된 런 수에 따라 필요한 런 수를 산출하기에 충분하지 않을 수 있습니다 . 이 경우 " 충분한 개수의 점을 찾을 수 없습니다 ." 라는 JMP 경고가 나타날 수 있습니다 . 군집화 알고리즘의 기초를 구성하는 초기점 수를 늘리려면 **고급 옵션 > 평균 군집 크기 설정**을 선택하여 설계점당 평균 초기점 수를 더 큰 값으로 지정합니다 . 자세한 내용은 "FFF 설계의 평균 군집 크기 설정 "(593 페이지)에서 확인하십시오 .

선형 제약 조건 지정

선형 제약 조건 지정 옵션을 사용하면 군집화 알고리즘의 기초를 구성하는 랜덤 점이 제약된 설계 영역 내에 무작위로 분포됩니다. 군집화 알고리즘에 이러한 점이 사용됩니다.

제약된 쾌속 유연 채움 설계 생성 및 보기

이 예에서는 제약된 쾌속 유연 채움 설계를 생성하고 표시합니다.

설계 구성

1. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다 .
2. 연속형 요인을 3 개 추가합니다 .
3. 요인 수준 값을 0 과 1 로 변경하고 **계속**을 클릭합니다 .
4. " 요인 제약 조건 정의 " 개요에서 **선형 제약 조건 지정**을 선택합니다 .

공간 채움 설계 방법 중 " 쾌속 유연 채움 " 만 사용할 수 있습니다 .

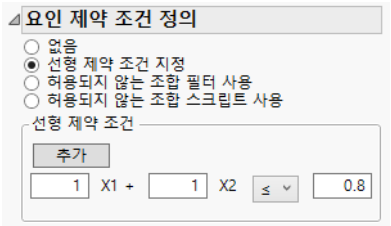
5. **추가**를 선택합니다 .
6. 다음 계수와 한계를 입력합니다 .

   X1 - 1

   X2 - 1

   한계 - 0.8

그림 21.23 선형 제약 조건



7. **런 수** 옆에 200 을 입력합니다 .
8. **쾌속 유연 채움**을 선택합니다 .

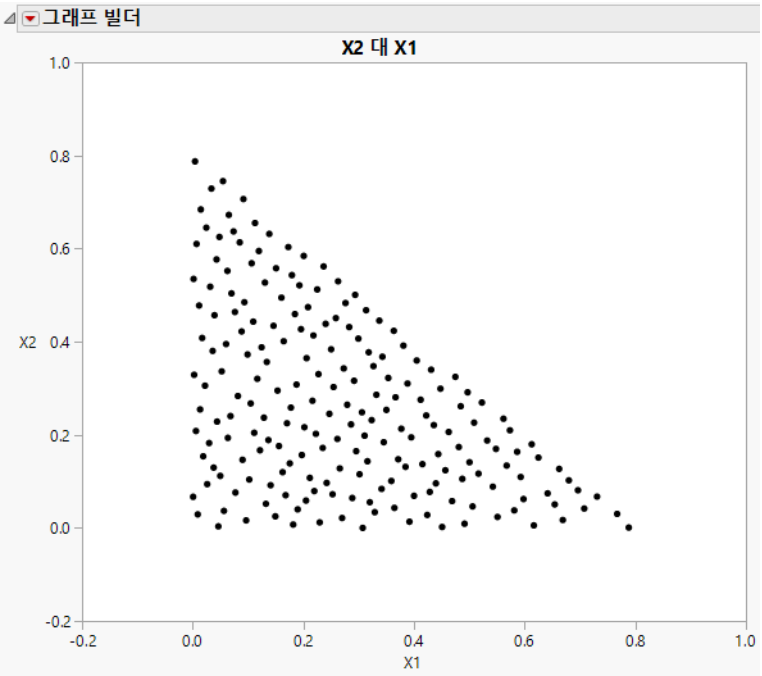
   제약 조건을 충족하는 설계가 생성됩니다 . " 설계 " 개요를 열어 설계를 확인합니다 .
9. **테이블 생성**을 선택하여 데이터 테이블을 구성합니다 .

그림 구성

1. 데이터 테이블을 활성화하고 **그래프 > 그래프 빌더**를 선택합니다 .
2. X1 을 **X** 라벨이 지정된 놓기 영역으로 드래그합니다 .
3. X2 를 **Y** 라벨이 지정된 놓기 영역으로 드래그합니다 .
4. 평활기 아이콘을 클릭하여 **평활기**를 제거합니다 .
5. " 그래프 빌더 " 의 빨간색 삼각형을 클릭하고 **제어판 표시**를 클릭하여 선택 취소합니다 .

그림 21.24 의 그래프와 유사한 그래프가 표시됩니다 . 점은 선형 제약 조건 $X1 + X2 \leq 0.8$ 을 충족합니다 .

그림 21.24 선형 제약 조건이 1 개 있는 쾌속 유연 채움 설계



맵 쉐이프를 위한 공간 채움 설계 생성

FFF(쾌속 유연 채움) 설계를 사용하면 맵 쉐이프와 같은 영역에서 실행할 실험을 설계할 수 있습니다. 예를 들어 미국의 주 전역에 퍼져 있는 위치에서 두 가지 표본 수집 방법(예: 공기 또는 토양 표본)을 비교하려는 경우 FFF 설계를 고려하십시오.

미국 조지아 주를 대상으로 가정해 보겠습니다. 주 전역에 퍼져 있는 40 개의 표본(각 표본 추출 방법에 대해 20 개씩)이 있어야 합니다. 맵 쉐이프를 위한 FFF 설계를 생성하려면 맵 경계의 위도와 경도가 포함된 파일이 필요합니다. 자세한 내용은 Essential Graphing 에서 확인하십시오.

다음 예에서는 미국 조지아 주의 경도와 위도를 확인하고 주 쉐이프를 정의한 후 설계를 생성합니다. 이 설계는 40 개 표본 각각에 대한 위치 및 표본 추출 방법을 제공합니다.

조지아 주의 경도 및 위도 찾기

1. 일반적으로 다음 위치에 있는 디렉터리로 이동합니다.

   Windows: C:/Program Files/SAS/JMP/< 버전 >/Maps

   macOS: /Library/Application Support/JMP/< 버전 >/Maps

2. US-State-XY.jmp 와 US-State-Name.jmp 를 엽니다.

   US-State-Name.jmp 에서 Shape ID 11 이 미국 조지아 주에 해당합니다. 이 예에서는 조지아 주를 대상으로 합니다.

3. US-State-XY.jmp 에서 **행 > 행 선택 > 선택 조건**을 선택합니다 .
4. Shape ID 를 선택하고 텍스트 상자에 11 을 입력한 후 **확인**을 클릭합니다 .

   이렇게 하면 조지아 주를 나타내는 Shape ID = 11 인 모든 행이 선택됩니다 .
5. **테이블 > 부분집합**을 선택하고 **확인**을 클릭하여 US-State-XY.jmp 테이블의 부분집합을 가져옵니다 . 여기에는 조지아 주의 위도 및 경도 데이터 (선택한 행) 가 포함됩니다 .
6. US-State-XY.jmp 와 US-State-Name.jmp 를 닫습니다 .

값 범위 가져오기

공간 채움 설계를 위한 공간을 정의하려면 조지아 주의 위도와 경도에 대한 값 범위를 알아야 합니다. 그런 다음 이러한 값을 사용하여 설계 요인을 생성할 수 있습니다.

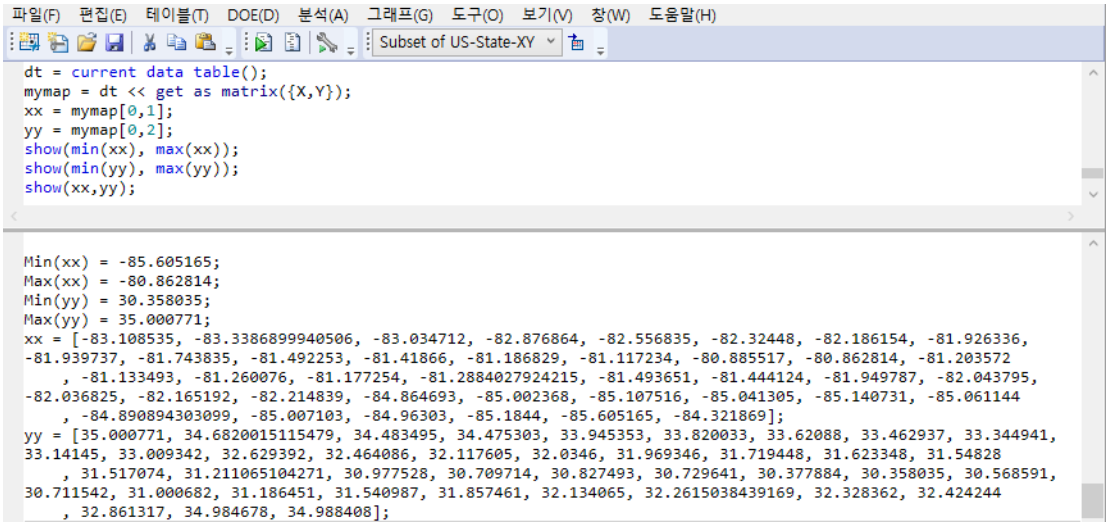
1. **파일 > 새로 만들기 > 스크립트**를 선택합니다 . 스크립트 창에서 마우스 오른쪽 버튼을 클릭하고 **중첩된 로그 표시**를 선택합니다 .
2. 다음 내용을 복사하여 새 스크립트에 붙여 넣습니다 .

```
dt = current data table();
mymap = dt << get as matrix({X,Y});
xx = mymap[0,1];
yy = mymap[0,2];
show(min(xx), max(xx));
show(min(yy), max(yy));
show(xx,yy);
```

3. US-State-XY.jmp 테이블의 부분집합을 활성화하고 스크립트를 실행합니다 .

   조지아 주의 위도와 경도에 대한 값 범위 및 **xx** 행렬과 **yy** 행렬이 로그에 나타납니다 .

그림 21.25 범위 찾기 스크립트와 로그 결과



요인 추가

1. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다 .
2. 연속형 요인 두 개를 설계에 추가합니다 . **N 개 요인 추가** 텍스트 상자에 2 를 입력하고 **연속형**을 클릭합니다 .
3. 요인 이름을 다음과 같이 longitude 와 latitude 로 변경하고 " 요인 " 개요에서 최소값과 최대값을 입력합니다 .
   - longitude 값을 -85.6 과 -80.9 로 입력합니다 . 이 값은 로그에서 x 의 최소 및 최대 경도 값입니다 .
   - latitude 값을 30.4 와 35.0 으로 입력합니다 . 이 값은 로그에서 y 의 최소 및 최대 위도 값입니다 .
4. 범주형 요인을 추가합니다. **범주형 > 2 수준**을 클릭합니다. 요인 이름을 Method 로 바꿉니다.

그림 21.26 요인

요인

이름	역할	값	
longitude	연속형	-85.6	-80.7
latitude	연속형	30.3	35
Method	범주형	L1	L2

5. **계속**을 클릭합니다 .

쉐이프 정의

공간 채움 설계를 위한 공간을 좀 더 자세히 정의하려면 다각형을 사용하여 해당 주의 쉐이프를 정의해야 합니다.

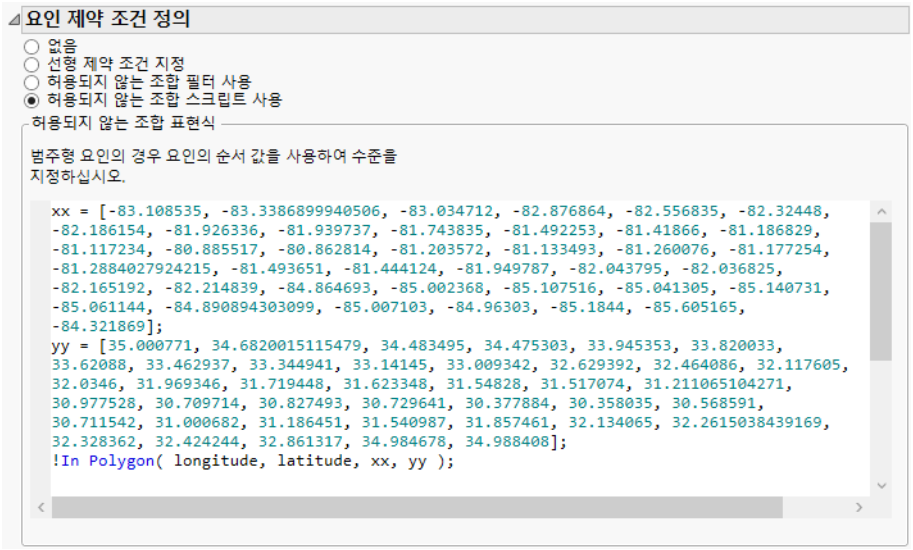
1. **허용되지 않는 조합 스크립트 사용**을 선택하여 " 허용되지 않는 조합 표현식 " 텍스트 상자를 엽니다 .

   경도 및 위도 범위는 정사각형을 정의합니다 . 조지아 주는 정사각형이 아니라 다각형입니다 . 조지아 주 다각형을 벗어나는 점을 허용하지 않으려고 합니다 .
2. 스크립트 창으로 이동한 후 스크립트 로그에서 **xx** 및 **yy** 행렬을 복사합니다 . 그런 다음 " 허용되지 않는 조합 표현식 " 텍스트 상자에 붙여 넣습니다 .
3. 끝에 다음 행을 추가합니다 .

   ```
   !In Polygon(longitude, latitude , xx, yy);
   ```

   이렇게 하면 **xx** 및 **yy** 좌표로 정의된 다각형에 없는 점이 제외됩니다 .

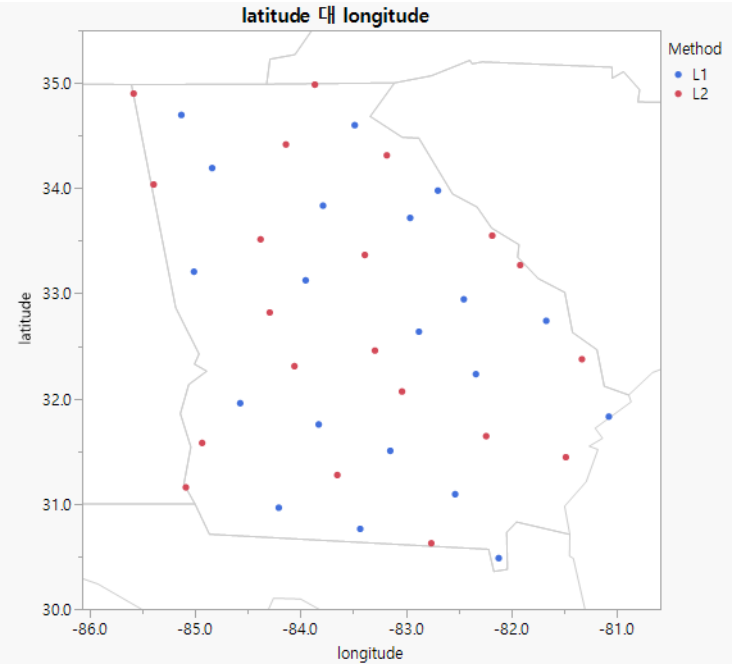
그림 21.27 허용되지 않는 조합 표현식



설계 생성

1. **런 수**를 40 으로 변경하고 **쾌속 유연 채움**을 클릭합니다 .
2. **테이블 생성**을 클릭합니다 .
3. 설계 테이블에서 **그래프 > 그래프 빌더**를 선택합니다 .
4. 다음과 같이 열을 해당 영역으로 드래그합니다 .
   - longitude 열 - **X** 영역
   - latitude 열 - **Y** 영역
   - Method 열 - **색상** 영역
5. 그래프 위에서 " 평활기 " 요소를 클릭하여 기능을 해제합니다 .
6. 주 쉐이프를 추가하려면 그래프를 마우스 오른쪽 버튼으로 클릭하고 **그래프 > 배경 맵**을 선택합니다 .
7. **경계** 아래에서 **미국 주**를 선택하고 **확인**을 클릭합니다 .

그림 21.28 미국 조지아 주에 대한 FFF 설계의 예



참고 : FFF 알고리즘의 랜덤 시작점으로 인해 맵에서 정확한 점 위치가 표시된 것과 다를 수 있습니다 .

이 맵은 표본을 수집할 40 개 위치를 나타냅니다. 표본 추출 방법 (L1 또는 L2) 은 점 색상으로 표시됩니다.

"그래프 > 배경 맵" 창에는 다음과 같은 추가 옵션이 있습니다.

- 검정할 자치주를 확인하려면 **경계** 아래에서 **미국 자치주**를 선택합니다 .
- 검정 위치를 거리 수준으로 맵에 나타내려면 **이미지** 아래에서 **Street Map Service** 를 선택합니다 .

Sphere-Packing 설계의 예

Worley(1987) 는 지표면에서 두 개의 대수층을 통과하여 뚫은 보어홀을 통한 물의 흐름 모형을 제시했습니다 . 반응 변수 y 는 보어홀을 통한 흐름 속도 (m^3/ 년) 이며 다음 방정식에 의해 결정됩니다 .

$$y = \frac{2\pi T_u (H_u - H_l)}{\ln(r/r_w)\left[1 + \frac{2LT_u}{\ln(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}\right]}$$

이 모형에는 다음과 같은 8 개 요인이 있습니다 .

r_w = 보어홀 반지름 (0.05 ~ 0.15m)

r = 영향 반지름 (100 ~ 50,000m)

T_u = 상부 대수층의 투과율 (63,070 ~ 115,600m^2/ 년)

H_u = 상부 대수층의 등수압면 수두 (990 ~ 1100m)

T_l = 하부 대수층의 투과율 (63.1 ~ 116m^2/ 년)

H_l = 하부 대수층의 등수압면 수두 (700 ~ 820m)

L = 보어홀 길이 (1120 ~ 1680m)

K_w = 보어홀의 수리 전도도 (9855 ~ 12,045m/ 년)

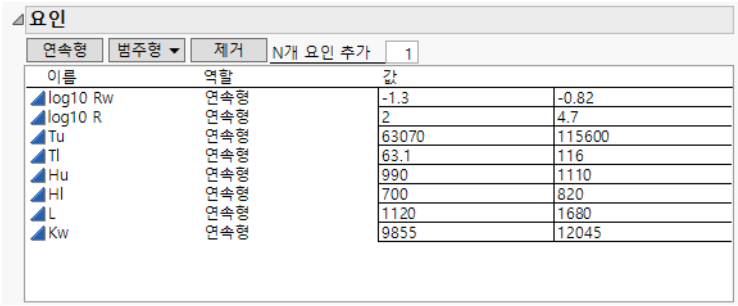
Sphere Packing 설계를 사용하여 반응 y 계산을 위한 조건 집합을 얻을 수 있습니다. 그런 다음 설계에 사용된 입력 범위에서 참모형을 추정하기 위한 모형을 생성할 수 있습니다. 추정된 모형을 평가하면 8 개의 각 요인이 반응에 미치는 영향을 이해하는 데 도움이 됩니다.

Borehole 데이터에 대한 Sphere-Packing 설계 생성

저장된 요인 설정이 포함된 데이터 테이블을 사용하여 보어홀 모형에 대한 Sphere Packing 설계를 생성할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Borehole Factors.jmp 를 엽니다 .
2. **DOE > 특수 용도 > 공간 채움 설계**를 선택합니다 .
3. " 공간 채움 설계 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

그림 21.29 보어홀 예의 요인 값을 불러온 요인 패널



참고 : r 및 r_w 의 로그가 요인으로 사용됩니다 .

4. **계속**을 클릭합니다 .
5. " 런 수 " 를 32 로 설정하고 Enter 키를 누릅니다 .
6. **Sphere Packing** 을 클릭하여 설계를 생성합니다 .
7. **테이블 생성**을 클릭하여 실험의 설계 설정을 보여 주는 테이블을 생성합니다 .

참고 : 설계 테이블에 " 모형 " 테이블 스크립트가 포함되어 있습니다 . 이 스크립트는 반응 y 에 대해 가우스 과정 모형을 실행합니다 .

8. (선택 사항) 이 예의 완료된 데이터 테이블을 보려면 **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Borehole Sphere Packing.jmp 를 엽니다 .

   설계가 난수 시드값에서 생성되므로 결과 설정은 완료된 테이블에 표시된 설정과 다를 수 있습니다 .

결정적 데이터 분석을 위한 지침

결정적 데이터에는 랜덤 성분이 없음을 기억해야 합니다. 입력 값이 동일하면 생성되는 출력도 동일합니다. 따라서 적합 통계 모형의 p 값은 일반적인 의미를 갖지 않습니다. 큰 F 통계량, 즉 낮은 p 값은 모형 항으로 인한 효과를 나타냅니다. 그러나 효과 또는 모형 예측을 위한 유효한 신뢰 구간을 구성할 수 없습니다.

결정적 데이터에 대한 모형 적합의 잔차는 잡음 측도가 아닙니다. 대신 잔차는 모형 편향의 측도입니다. 편향은 실제 값과 예측값의 차이입니다. 잔차에 고유한 패턴이 있으면 편향을 줄이기 위해 모형에 추가 항을 고려해야 한다는 것을 나타냅니다.

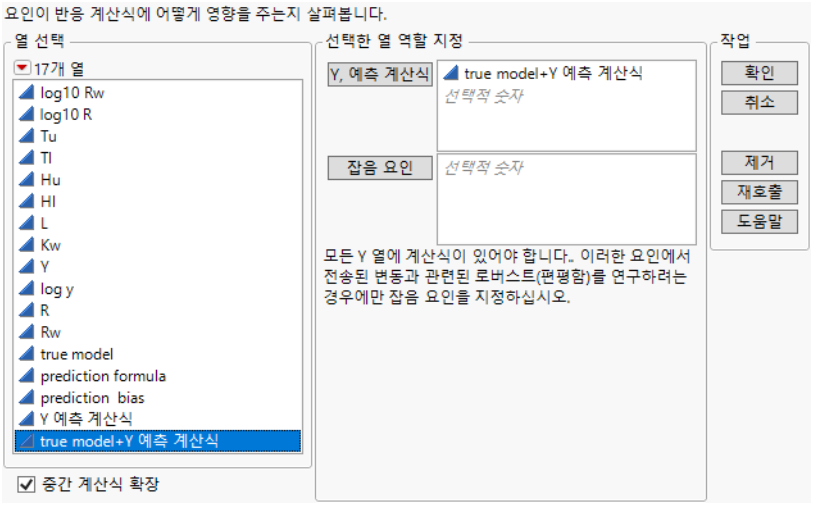
Borehole Sphere-Packing 설계 분석

참모형을 단순 분석 형태로 사용할 수 없는 경우가 종종 있습니다 . 따라서 예측 편향은 관측된 데이터 점에서만 알 수 있습니다 . 그러나 이 예에서는 참모형의 함수 형태가 알려져 있습니다 .

Borehole Sphere Packing.jmp 데이터 테이블의 true model 열에 알려진 함수의 계산식이 포함되어 있습니다. 이 계산식을 사용하면 요인 입력 영역에 대한 예측 편향을 프로파일링할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Borehole Sphere Packing.jmp를 엽니다.
2. "Model (GP from DOE)" 스크립트 옆의 녹색 삼각형을 클릭합니다.
   " 가우스 과정 모형 " 보고서를 사용하여 요인과 결과 Y 간의 관계를 탐색합니다.
3. "Y 의 가우스 과정 모형 " 옆의 빨간색 삼각형을 클릭하고 **예측 계산식 저장**을 선택합니다.
4. Borehole Sphere Packing.jmp 데이터 테이블로 돌아갑니다.
5. 데이터 격자에서 true model 및 Y 예측 계산식의 열 머리글을 선택합니다.
6. 마우스 오른쪽 버튼을 클릭하고 **새 계산식 열 > 결합 > 차**를 선택합니다.
   이렇게 하면 편향을 포함하는 새 열이 생성됩니다.
7. Borehole Sphere Packing.jmp 데이터 테이블에서 **그래프 > 프로파일러**를 선택합니다.
8. true model-Y 예측 계산식을 선택한 후 **Y, 예측 계산식**을 클릭합니다.
9. **중간 계산식 확장**을 선택합니다.
   이 옵션은 편향을 8 개 설계 요인의 함수로 표시합니다.

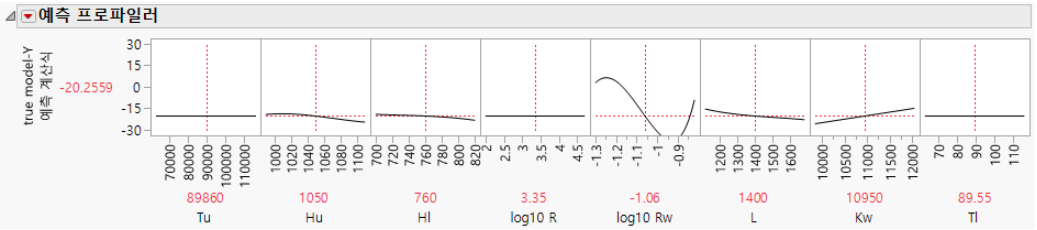
그림 21.30 Borehole Sphere-Packing 데이터의 프로파일러 대화상자



10. **확인**을 클릭합니다.

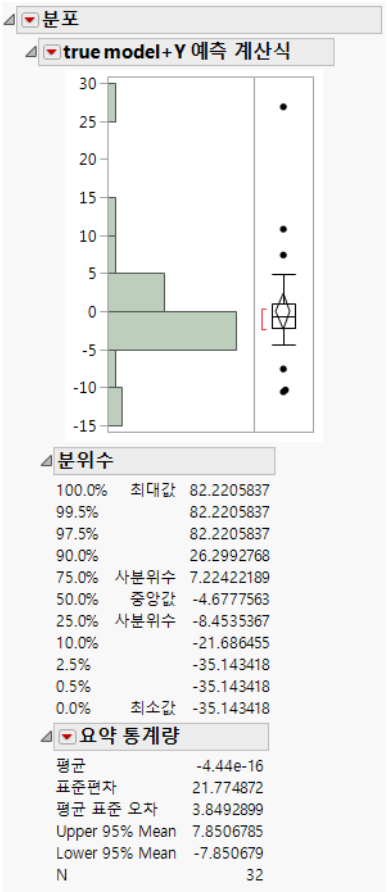
    프로파일러는 기본적으로 설계 영역의 중심으로 설정됩니다. 편향이 없으면 모든 프로파일 트레이스가 각 요인의 값 범위에서 일정합니다. 이 예에서는 logRw, Hu 및 Hl 변수가 편향에 가장 큰 영향을 미칩니다.

그림 21.31 Y 축이 -30 ~ 30 으로 설정된 Borehole GP 모형의 편향을 보여 주는 프로파일러



프로파일러를 사용하여 전체 도메인에 대한 예측 편향 범위를 탐색할 수 있습니다. 최소 및 최대 편향 점을 찾으려면 "예측 프로파일러"의 빨간색 삼각형 메뉴에서 **최적화 및 만족도 > 만족도 함수**를 선택합니다. 자세한 내용은 Profilers의 에서 확인하십시오. 설계점에 대한 예측 편향을 평가하려면 **분석 > 분포**를 선택하여 분포 분석을 확인합니다.

그림 21.32 예측 편향 분포



이 예에서는 참모형이 알려져 있습니다. 그러나 많은 적용 분야에서 요인 설정에 대한 반응을 알 수 없습니다. 실험 데이터에 대한 예측 편향은 설계 도메인 전체의 편향을 과소 추정할 수 있습니다.

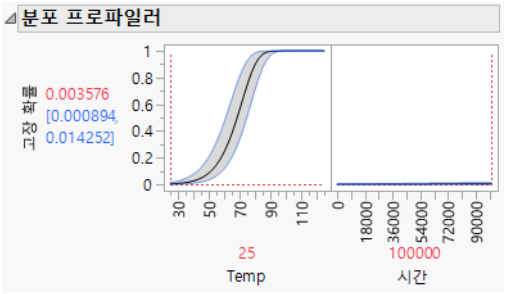
22 장

가속 수명 시험 설계

가속 수명 시험을 위한 실험 설계

가속 수명 시험(ALT) 설계 플랫폼을 사용하면 가속 수명 시험 실험을 위한 계획을 설계할 수 있습니다. 정상적인 사용 조건에서는 보통 제품 신뢰도가 높기 때문에 고장 날 때까지 제품을 시험하는 데 필요한 시간이 너무 깁니다. 정상적인 사용 조건에서 제품을 시험하는 대신 더 열악한 조건에서 제품을 시험할 수 있습니다. 조건이 열악하면 제품의 성능이 더 빨리 저하되고 더 빨리 고장날 수 있습니다. 그러면 이 가속 고장 데이터를 사용하여 정상적인 사용 조건에서의 제품 신뢰도를 예측할 수 있습니다. 초기 실험을 설계하거나 기존 실험을 확대할 수 있습니다.

그림 22.1 ALT 실험의 고장 확률을 보여 주는 프로파일러



목차

가속 수명 시험 설계 개요 609
가속 수명 시험 설계의 예 609
가속 수명 시험 설계 확대의 예 616
가속 수명 시험 설계 생성 620
설계 구조 지정 621
가속 요인 지정 621
설계 상세 정보 지정 622
ALT 계획에 대한 규격 검토 및 업데이트 624
최적 설계 생성 및 평가 625
설계 테이블 생성 627
가속 수명 시험의 옵션 628
ALT 설계 플랫폼에 대한 통계 상세 정보 628
ALT 설계의 고장 분포 628
R 정밀도 요인(95% CI) 630
가속 수명 시험에 대한 최적 기준 631

가속 수명 시험 설계 개요

정상적인 사용 조건에서 제품 신뢰도가 높을 때는 가속 수명 시험을 사용합니다. 정상적일 때보다 더 열악한 조건에서 제품을 시험하여 고장이 발생할 때까지의 시간을 가속화할 수 있습니다. 이 가속 고장 데이터를 사용하여 정상적인 사용 조건에서의 제품 신뢰도를 예측합니다.

이러한 시험을 **가속 수명 시험**(ALT)이라고 합니다. 고장날 수준으로 시간을 가속화하는 요인을 설정할 수 있는데 이러한 요인을 **가속 요인**이라고 합니다. 일반적으로 가속 수명 시험 모형은 비선형 모형입니다. 비선형 모형에 대한 자세한 내용은 에서 확인하십시오. "비선형 설계" 장의 "비선형 모형"(650 페이지)에서 확인하십시오.

ALT 설계 플랫폼은 하나 또는 두 개의 가속 요인을 포함하는 상황에 대한 설계를 생성하고 평가합니다. D-최적, 분위수 추정 또는 고장 확률을 사용하여 설계를 최적화할 수 있습니다.

ALT 설계를 생성하려면 가속 모형 모수의 초기 추정값이 필요합니다. 주제 관련 지식을 사용하여 시작 추정값을 정의할 수 있습니다. 시작 추정값을 알리는 정보가 없으면 다변량 정규 사전 분포를 지정하여 불확도를 설명할 수 있습니다.

또한 가속 수명 시험(ALT) 설계 플랫폼을 사용하여 현재 실험을 확대할 수 있습니다. 추정값 분산을 줄이기 위해 현재 설계를 확대할 수도 있습니다.

가속 수명 시험 설계의 예

최적 고장 확률 ALT를 생성하려면 ALT 설계 도구를 사용합니다. 이 예에서는 기계 부품에 대한 가속 수명 시험을 설계해야 한다고 가정해 보겠습니다. 단일 가속 요인은 토크이며 정상 사용 스트레스는 35Nm(뉴턴 미터)입니다. 정상 사용 스트레스에서 단위의 10%가 고장 나는 시간을 추정하려고 합니다.

다음은 시험 계획에 대한 특성입니다.

- 시험에 사용할 수 있는 총 단위 수는 100 입니다.
- 수명 분포는 Weibull 로 가정합니다.
- 수명 - 스트레스 관계는 로그 변환에 의해 제공됩니다.
- 세 가지 스트레스 수준 (50, 75 및 100Nm) 에서 토크를 시험할 계획입니다.
- 시험 수준에서 초기 고장 시간을 예측할 수 있는 사전 지식이 있습니다. 자세한 내용은 " 사전 추정값 구하기 "(610 페이지) 에서 확인하십시오.
- 시험은 5,000 회 반복 수행됩니다.
- 연속적으로 고장에 대한 공정을 모니터링합니다.

사전 추정값 구하기

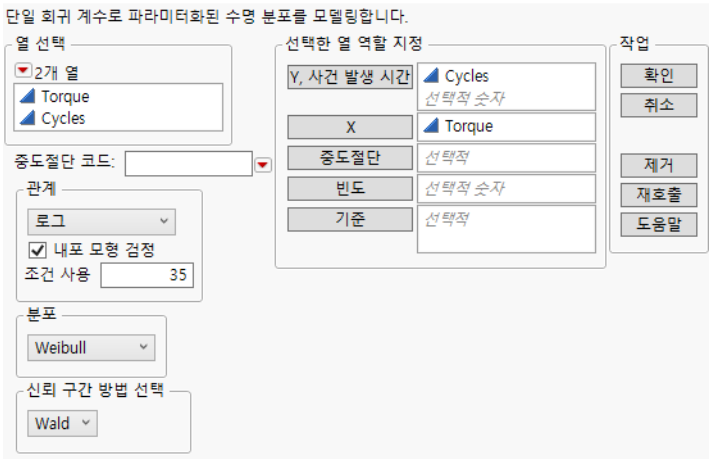
가속 수명 시험 설계를 생성하려면 모수의 사전 추정값을 제공해야 합니다 . 사전 추정값을 구하는 방법은 다음과 같습니다 .

1. 공정 지식을 사용하여 적은 수의 스트레스 수준에서 가설적이지만 가능성 있는 고장 시간 테이블을 생성합니다 .
2. 수명 분포 적합 플랫폼을 사용하여 모형을 적합시키고 모형 모수의 추정값을 구합니다 .
3. ALT 설계 플랫폼을 사용하여 설계를 생성할 때 이 추정값을 사전 값으로 사용합니다 .

위에서 설명한 방법에 따라 균형 설계의 고장 주기 수 추정값을 포함하는 데이터 테이블을 생성합니다 . 테이블은 설계에 사용할 세 가지 스트레스 수준에서 각각 5 개의 단위로 구성됩니다 . 이 예에서는 샘플 데이터 라이브러리에 있는 Torque Prior.jmp 데이터 테이블을 사용합니다 .

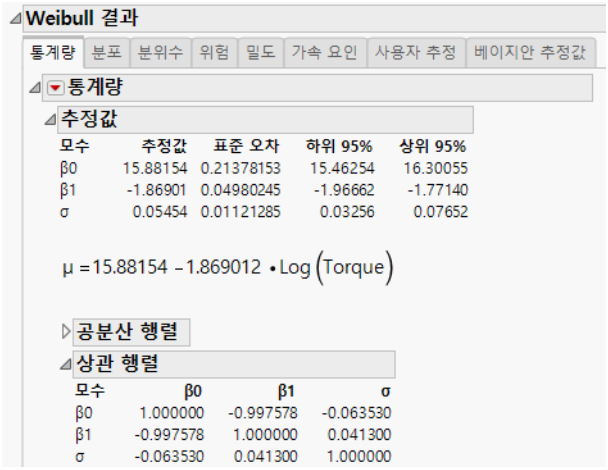
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Torque Prior.jmp 를 엽니다 .
2. **분석 > 신뢰성 및 생존 > 수명 분포 적합**을 선택합니다 .
3. Cycles 를 선택하고 **Y, 사건 발생 시간**을 클릭합니다 .
4. Torque 를 선택하고 **X** 를 클릭합니다 .
5. **관계** 목록에서 **로그**를 선택합니다 .
6. **사용 조건**에 35 를 입력합니다 . 35Nm 가 정상적인 사용 조건입니다 .
7. **분포** 목록에서 **Weibull** 을 선택합니다 .

그림 22.2 수명 분포 적합 시작 창



8. **확인**을 클릭합니다 .
9. "Weibull 결과 " 개요가 나올 때까지 아래로 스크롤하여 " 상관 행렬 " 개요를 엽니다 .

그림 22.3 사전 데이터의 수명 분포 적합 모형



평균에 대한 모형은 " 추정값 " 개요에 제공됩니다 . " 추정값 " 열에는 절편 (β0), 선형 계수 (β1) 및 척도 (σ) 에 대한 모수 추정값이 포함되어 있습니다 . " 표준 오차 " 열에는 추정값에 대한 표준 오차가 포함되어 있습니다 . 모수 추정값에 대한 상관 추정값은 " 상관 행렬 " 개요에 제공됩니다 . ALT 설계를 생성하기 위한 사전 값으로 이들 모수 추정값 , 표준 오차 및 상관을 사용합니다 .

기본 규격 입력

가속 수명 시험 설계를 생성하려면 가속 요인, 사용 조건 및 시험 조건을 정의합니다.

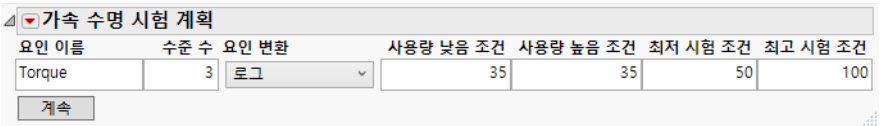
1. **DOE** > **특수 용도** > **가속 수명 시험 설계**를 선택합니다 .

   한 개의 가속 요인과 **연속 모니터링**이 기본적으로 선택되어 있습니다 .
2. **계속**을 클릭합니다 .
3. **요인 이름**에서 X1 을 클릭하고 Torque 를 입력합니다 .

   수준 수는 기본적으로 3 으로 설정되어 있습니다 .
4. " 요인 변환 " 에서 **로그**를 선택합니다 .
5. **낮은 사용 조건**과 **높은 사용 조건** 둘 다에 35 를 입력합니다 .

   낮은 사용 조건과 높은 사용 조건을 모두 35 로 설정한다는 것은 35 가 정상적인 사용 조건임을 나타냅니다 .
6. **최저 시험 조건**에 50 을 입력하고 **최고 시험 조건**에 100 을 입력합니다 .

그림 22.4 완료된 ALT 규격 창



7. **계속**을 클릭합니다 .
8. **Torque 수준 값**이 50, 75, 100 으로 설정되어 있습니다 .

   참고 : JMP 는 최저 시험 조건과 최고 시험 조건 사이에서 균등한 간격으로 수준 값을 설정합니다 . 수준은 변경할 수 있습니다 .

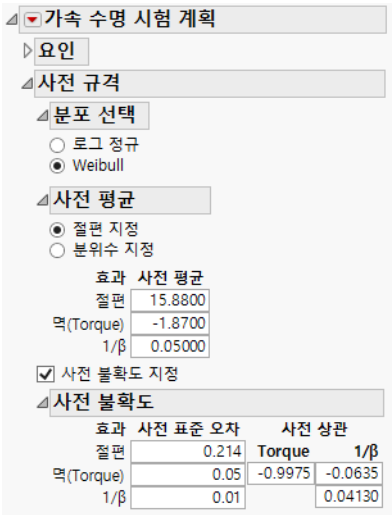
9. **분포 선택**에 **Weibull** 이 선택되어 있는지 확인합니다 .

사전 정보 및 나머지 규격 입력

사전 정보를 입력하여 예제 ALT 설계 생성을 계속합니다.

1. **사전 평균**에서 **절편 지정**을 선택하고 수명 분포 적합 플랫폼에서 가상 데이터를 사용하여 구한 가속 모형 모수의 사전 추정값을 다음과 같이 입력합니다 .
   - **절편**에 15.88 을 입력합니다 .
   - **멱 (Torque)** 에 -1.87 을 입력합니다 .
   - **1/**β(Weibull 모형의 척도 모수) 에 0.05 를 입력합니다 .
2. **사전 불확도 지정**을 선택합니다 .
3. 사전 가속 모형의 표준 오차 추정값과 상관을 입력합니다 . 그림 22.3 에서는 완료된 사전 불확도 개요를 보여 줍니다 .

그림 22.5 완료된 사전 규격 개요

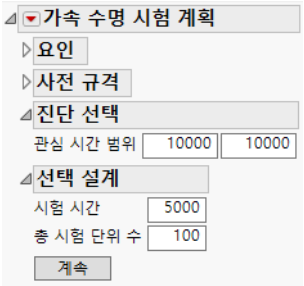


" 가속 수명 시험 계획 " 의 빨간색 삼각형을 클릭하고 **ALT 최적 기준 > 최적 고장 확률 생성**을 선택합니다 .

5,000 회 반복 시험할 계획입니다 . 최대 10,000 회 시행에 대해 예측하려고 합니다 .

4. " 진단 선택 " 개요에서 **관심 시간 범위**의 두 상자에 모두 10,000 을 입력합니다 .
5. " 선택 설계 " 에서 **시험 시간**에 5,000 을 입력하고 **시험 단위 수**에 100 을 입력합니다 .

그림 22.6 완료된 진단 및 선택 설계 개요

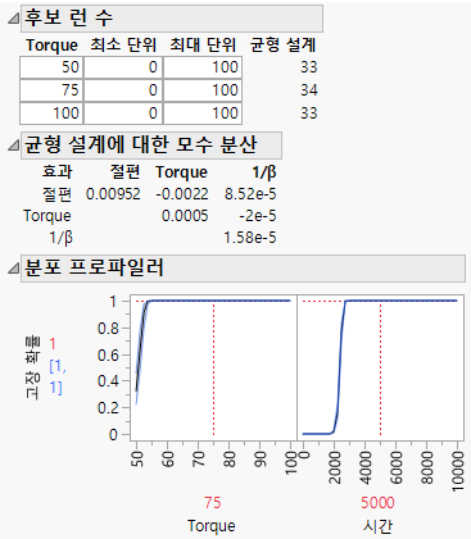


설계 생성

계속해서 ALT 예제에 대해 설계를 생성합니다.

1. **계속**을 클릭합니다 .

그림 22.7 균형 설계 진단

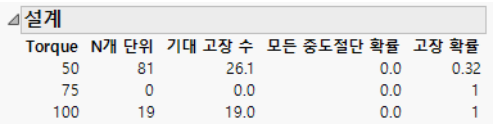


균형 설계의 런 수가 " 후보 런 수 " 개요에 나타납니다 . " 균형 설계에 대한 모수 분산 " 개요에는 이 설계의 모수에 대한 공분산 행렬이 표시됩니다 . " 분포 프로파일러 " 도 나타납니다 . 최적 설계를 구하면 모수 분산 및 분포 프로파일러 결과를 균형 설계의 결과와 비교하여 불확도의 감소를 확인할 수 있습니다 .

2. **설계 생성**을 클릭합니다 .

   다른 결과와 함께 최적 실험 설계가 나타납니다 .

그림 22.8 최적 설계



최적 설계는 시험 런의 수준 , 시험할 총 단위 수 및 지정한 사전 정보를 기반으로 계산됩니다 . 최적 설계는 다음과 같이 각 토크 수준에서 지정된 단위 수를 시험하는 것으로 구성됩니다 .

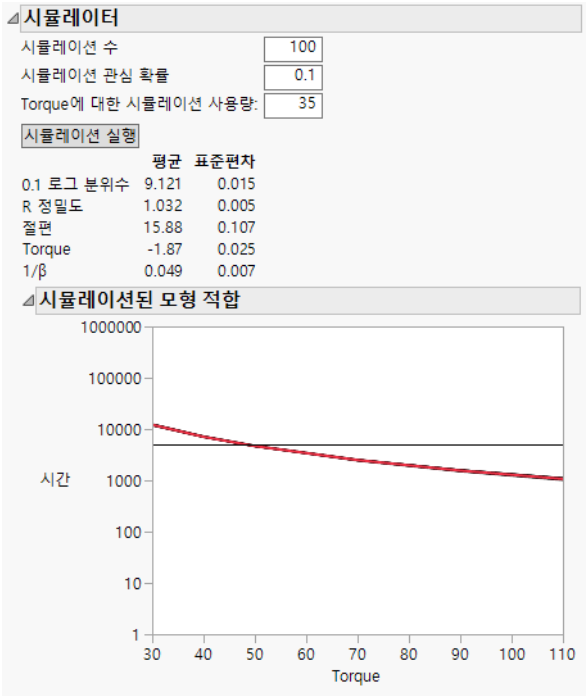
- 50Nm 에서 81 개 단위
- 75Nm 에서 0 개 단위
- 100Nm 에서 19 개 단위

시뮬레이션 실행

예제 ALT 설계를 시뮬레이터로 평가합니다 . 시뮬레이터를 사용하면 설계를 생성하는 데 사용된 가정을 기반으로 시험 계획 결과를 시뮬레이션할 수 있습니다 . 이렇게 하면 시험 계획을 실행하기 전에 해당 계획이 충분한지 평가할 수 있습니다 .

1. **시뮬레이션 수**에 100 을 입력합니다 .
2. **시뮬레이션 관심 확률**에 0.1 을 입력합니다 .
3. **Torque 에 대한 시뮬레이션 사용량**에 35 를 입력합니다 .
4. **시뮬레이션 실행**을 클릭하여 실험 시뮬레이션을 100 회 실행합니다 . 이렇게 하면 설계의 사전 가정을 고려해 볼 때 시험할 단위 수 100 개가 충분한지 여부를 확인할 수 있습니다 .

그림 22.9 시뮬레이션 결과



시뮬레이션된 모형 적합을 보면 100 회의 모형 시뮬레이션이 모두 빨간색 곡선으로 표시된 평균에 가깝기 때문에 모두 유사하다는 것을 알 수 있습니다 . Y 축은 로그 척도 위에 있습니다 .

5. Y 축을 두 번 클릭하고 " 유형 " 메뉴에서 **선형**을 선택한 후 **확인**을 클릭합니다 .

   이제 정상 사용 조건 (35) 주위의 시뮬레이션된 적합에서 변동을 쉽게 볼 수 있습니다 . 정상 사용 조건의 추정에서 변동이 예상됩니다 . 시뮬레이션은 ALT 계획이 허용 가능한지 여부를 평가하는 데 도움이 됩니다 . 변동이 너무 크면 시험 단위 수 또는 시험 조건을 조정해야 할 수도 있습니다 .

설계 저장

마지막으로 시험 계획을 생성하여 데이터 테이블에 저장함으로써 예제 ALT 설계를 완료합니다.

1. **시험 계획 생성**을 클릭하여 설계를 데이터 테이블에 저장합니다. 이 테이블은 설계를 요약한 것입니다.
2. **테이블 생성**을 클릭하여 설계를 데이터 테이블에 저장합니다. 이 테이블에는 각 시험 런에 대해 하나씩 총 100 개의 행이 있습니다.

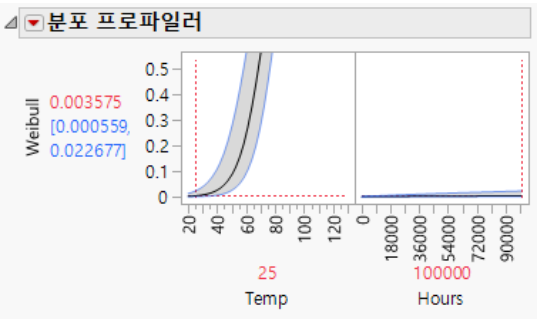
가속 수명 시험 설계 확대의 예

이 예에서는 가속 수명 시험 설계 플랫폼을 사용하여 기존 설계를 확대하는 방법을 보여 줍니다. 이 예에서는 150 개 단위의 Capacitor 를 세 가지 온도(섭씨 85°, 105° 및 125°)에서 1500 시간 동안 시험합니다. 실험 결과는 Capacitor ALT.jmp 샘플 데이터 테이블에 기록됩니다. 생성된 결과 모형은 정상 사용 온도(섭씨 25°)일 때 100,000 시간에서 고장 나는 모집단의 비율을 예측하는 데 사용됩니다.

현재 예측 검토

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Capacitor ALT.jmp 를 엽니다.
2. 녹색 삼각형을 클릭하여 **Fit Life by X** 테이블 스크립트를 실행합니다.
3. " 비교 " 개요의 " 분포 " 탭에 있는 " 분포 프로파일러 " 에서 다음과 같이 요인 설정을 변경합니다.
   - Temp 위의 105 를 클릭하고 25 로 변경합니다.
   - Hours 위의 750.5 를 클릭하고 100,000 으로 변경합니다.

그림 22.10 Capacitor 모형의 분포 프로파일러



현재 연구에 따르면 Capacitor 모집단에서 25°일 때 100,000 시간에서 고장 나는 예측 비율은 0.00358 이고 95% 신뢰 구간은 0.00056 ~ 0.02268 입니다. 고장 비율을 보다 정확히 예측하려고 합니다. 추정값 정밀도의 측도인 신뢰 구간 너비를 줄이기 위해 추가 시험을 통해 연구를 확대할 수 있습니다.

설계 확대

예측 고장 비율의 더 정확한 추정값을 얻기 위해 설계를 확대하려고 합니다 . 원래 설계는 85, 105 및 125 온도 설정을 사용했습니다 . 확대된 설계에서는 온도 값을 90, 110, 125 로 설정하여 시험할 계획입니다 . 이 중에서 두 설정이 새 값입니다 . 최적으로 선택된 런을 사용하여 다음과 같이 설계를 확대합니다 .

1. **DOE** > **특수 용도** > **가속 수명 시험 설계**를 선택합니다 .
2. **한 개의 가속 요인**을 선택하고 **연속 모니터링**을 선택한 후 **계속**을 클릭합니다 .
3. **요인 이름**에 Temp 를 입력합니다 .
4. **수준 수**에 5 를 입력합니다 .

   확대된 런은 세 수준 (90, 110, 125) 만 포함하지만 원래 실험에 사용된 수준도 지정하여 총 5 개 수준이 되어야 합니다 . " 요인 변환 " 은 기본적으로 " 아레니우스 섭씨 " 로 설정되어 있습니다 .
5. **낮은 사용 조건**과 **높은 사용 조건** 둘 다에 25 를 입력합니다 .
6. **최저 시험 조건**에 85 를 입력하고 **최고 시험 조건**에 125 를 입력합니다 .
7. **계속**을 클릭합니다 .
8. **Temp 수준 값**에 85, 90, 105, 110 및 125 를 입력합니다 .

   원래 실험의 수준이 세 개 있습니다 (85, 105, 125). 확대된 설계에는 새 수준 두 개 (90, 110) 와 첫 번째 실험의 수준 중 하나 (125) 가 있습니다 .

   참고 : 모든 수준이 나열되어야 합니다 .
9. **분포 선택**에 **Weibull** 이 선택되어 있는지 확인합니다 .
10. **사전 평균**에서 **절편 지정**을 선택합니다 . "Weibull 결과 " 개요의 " 통계량 " 탭에 있는 " 수명 분포 적합 추정값 " 개요에서 현재 가속 모형 모수 추정값을 입력합니다 .

그림 22.11 Weibull 결과 개요의 모수 추정값 및 적합 모형

추정값

모수	추정값	표준 오차	하위 95%	상위 95%
β0	-35.19979	4.6912686	-44.39451	-26.00508
β1	1.38896	0.1566362	1.08196	1.69597
σ	1.30471	0.1119128	1.08536	1.52405

$$\mu = -35.19979 + \frac{\left(1.388964 \bullet 11604.518122\right)}{\left(\text{Temp} + 273.15\right)}$$

   - **절편**에 -35.200 을 입력합니다 .
   - **활성화 에너지 (Temp)** 에 1.389 를 입력합니다 .

     이 값은 활성화 에너지의 추정값이며 , 볼츠만 상수를 곱한 역온도 계수입니다 (켈빈 온도로 측정).

– **1/β** 에 1.305 를 입력합니다 .

Weibull 분포의 경우 JMP 는 위치 모수 (μ) 와 척도 모수 (σ) 에 따라 달라지는 파라미터화를 사용합니다 . 일반적인 α 및 β 파라미터화 측면에서 척도 모수는 σ = 1/β 입니다 . 자세한 내용은 "Weibull"(629 페이지) 에서 확인하십시오 .

" 가속 수명 시험 계획 " 창에서 **사전 불확도 지정** 옵션을 선택하여 사전 평균에 대한 불확도를 지정할 수 있습니다 . 이 예제에서는 사전 불확도를 지정하지 않습니다 . " 사전 평균 " 개요에 지정된 값이 실제 모수 값이라고 가정하여 설계가 생성됩니다 .

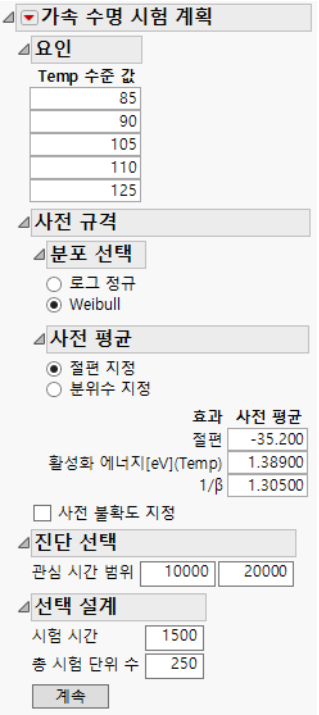
11. " 선택 설계 " 에서 **시험 시간**에 1500 을 입력합니다 .

시험은 1500 시간 (원래 설계의 시간) 동안 수행됩니다 .

12. **시험 단위 수**에 250 을 입력합니다 .

이전 실험에서 150 개 단위를 시험했고 확대된 실험에서 추가 100 개 단위를 시험하여 총 시험 단위 수가 250 개입니다 .

그림 22.12 완료된 설계 상세 정보 창



13. **계속**을 클릭합니다 .

14. **후보 런 수**에서 온도 85, 105 및 125 에 대한 최소 단위 수로 50 을 입력합니다 . 이 숫자는 이미 완료된 런 수를 나타냅니다 .

15. " 가속 수명 시험 계획 " 의 빨간색 삼각형을 클릭하고 **ALT 최적 기준 > 최적 고장 확률 생성**을 선택합니다 .

이 기준에 대한 자세한 내용은 " 최적 고장 확률 생성 "(628 페이지) 에서 확인하십시오 .

16. **설계 생성**을 클릭합니다 .

다른 결과와 함께 최적 실험 설계가 나타납니다 .

그림 22.13 최적 설계

설계

Temp	N개 단위	기대 고장 수	모든 중도절단 확률	고장 확률
85	50	6.9	0.0	0.14
90	58	12.3	0.0	0.21
105	50	30.0	0.0	0.6
110	0	0.0	0.0	0.75
125	92	91.2	0.0	0.99

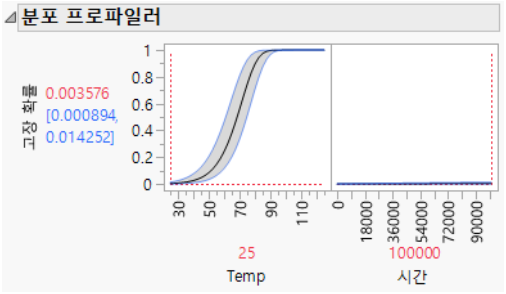
최적 설계는 시험 런의 수준 , 최소 시험 단위 수 , 시험할 총 단위 수 (" 후보 런 수 " 개요의 정보) 및 지정한 기타 정보를 기반으로 계산됩니다 . 최적 설계는 다음과 같이 각 온도 수준에서 지정된 단위 수를 시험하는 것으로 구성됩니다 .

- 85°C 에서 50 개 단위 . 이전 실험을 수행할 때 이미 85°C 에서 50 개 단위를 시험했으므로 추가 단위가 필요하지 않습니다 .
- 90°C 에서 50 개 단위 .
- 105°C 에서 50 개 단위 . 이전 실험을 수행할 때 105°C 에서 50 개 단위를 시험했으므로 추가 단위가 필요하지 않습니다 .
- 110°C 에서 0 개 단위 . 이 수준은 필요하지 않습니다 .
- 125°C 에서 50 개 단위 . 이전 실험을 수행할 때 125°C 에서 50 개 단위를 시험했으므로 추가 단위가 필요하지 않습니다 .

확대된 설계를 원래 연구와 비교

1. " 분포 프로파일러 " 에서 "Temp" 에 정상적인 사용 조건인 25 를 입력하고 " 시간 " 에 100,000 을 입력합니다 . 고장 나는 모집단 비율의 추정값은 0.00357 이고 95% 신뢰 구간은 0.00093 ~ 0.01361 입니다 . 이 구간은 초기 실험의 구간 , 즉 0.00056 ~ 0.02268(그림 22.10 참조) 보다 좁습니다 .

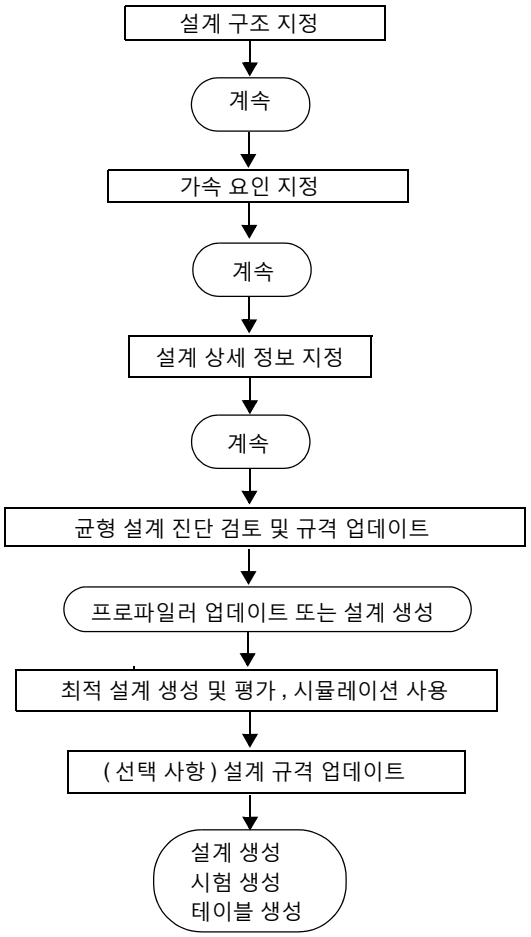
그림 22.14 Temp = 25, 시간 = 100000 에 대한 분포 프로파일러



가속 수명 시험 설계 생성

DOE > 특수 용도 > 가속 수명 시험 설계를 선택하여 가속 수명 시험(ALT) 설계를 생성합니다. 설계를 진행하는 동안 가속 수명 시험 계획이 업데이트됩니다. 가속 수명 시험 설계를 생성하는 단계는 그림 22.15의 흐름을 따릅니다.

그림 22.15 가속 수명 시험 계획 흐름

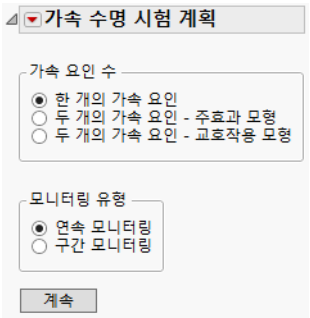


이 섹션에서는 "가속 수명 시험 계획" 흐름의 개요를 설명합니다.

설계 구조 지정

" 가속 수명 시험 계획 " 개요를 사용하여 설계의 모형 구조 및 검사 유형을 정의합니다 .

그림 22.16 초기 ALT 설계 창



가속 요인 개수

한 개의 가속 요인 요인이 하나인 설계를 생성할 때 선택합니다 .

두 개의 가속 요인 - 주효과 모형 요인이 두 개인 경우 , 주효과 모형에 대한 설계를 생성할 때 선택합니다 .

두 개의 가속 요인 - 교호작용 모형 요인이 두 개인 경우, 모형에 주효과 및 교호작용 항이 포함되는 설계를 생성할 때 선택합니다 .

모니터링 유형

연속 모니터링 정확한 고장 시간이 기록되는 설계에 대해 선택합니다 . 시험 시간을 초과하는 고장 시간은 오른쪽 중도절단됩니다 .

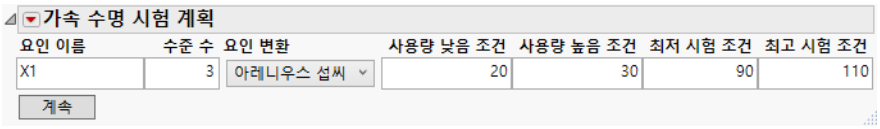
구간 모니터링 특정 시간 간격에 따라 단위 고장을 검사하는 설계에 대해 선택합니다 . 고장 시간은 구간 중도절단됩니다 . 검사 수 , 1 차 검사 시간 및 검사 구간을 입력합니다 . 검사 구간이 불규칙한 경우 나중에 " 선택 설계 " 개요에서 검사 시간을 변경할 수 있습니다 .

가속 수명 시험 계획 하나 이상의 가속 요인에 대한 상세 정보를 지정합니다.

가속 요인 지정

하나 이상의 가속 요인에 대한 상세 정보를 지정합니다.

그림 22.17 ALT 규격 창



요인 이름 각 가속 요인의 이름을 입력합니다 .

수준 수 각 가속 요인에 대해 실험에 포함할 수준 수를 입력합니다 .

요인 변환 각 가속 요인에 대해 변환 함수를 선택합니다 . 이 변환은 수명 - 스트레스 관계 , 즉 전체 스트레스 수준에서 수명 분포가 변하는 방식을 설명합니다 . " 아레니우스 섭씨 ", " 역수 ", " 로그 ", " 제곱근 " 및 " 선형 " 변환 중에서 하나를 사용할 수 있습니다 .

낮은 사용 조건 각 가속 요인에 대해 일반적인 사용 조건의 하한 값을 입력합니다 .

높은 사용 조건 각 가속 요인에 대해 일반적인 사용 조건의 상한 값을 입력합니다 .

참고 : " 낮은 사용 조건 " 값과 " 높은 사용 조건 " 값은 같을 수 있습니다 . 사용 조건이 단일 값인 경우 동일한 값을 사용합니다 .

최저 시험 조건 각 가속 요인에 대해 최저 시험 조건을 입력합니다 .

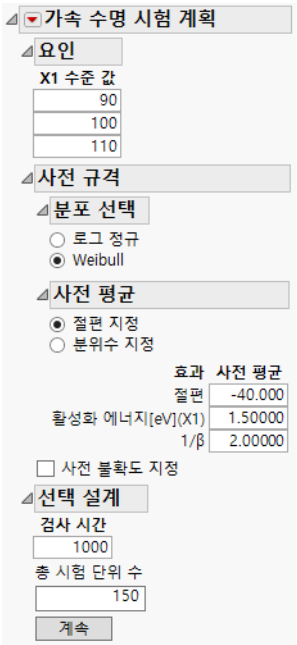
최고 시험 조건 각 가속 요인에 대해 최고 시험 조건을 입력합니다 .

참고 : 최저 시험 조건과 최고 시험 조건은 설계에서 가속 요인의 끝점을 정의합니다 . 초기 후보는 최저 시험 조건과 최고 시험 조건 사이에서 균등한 간격의 수준을 사용합니다 . 설계를 구성하기 전에 수준을 조정할 수 있습니다 .

설계 상세 정보 지정

요인 수준, 사전 분포의 세부 정보, 관심 시간 범위와 확률, 시험 시간 및 시험 단위 수를 지정합니다.

그림 22.18 분포 상세 정보



요인 가속 요인의 수준을 입력합니다. 기본적으로 수준은 최저 시험 조건과 최고 시험 조건 사이에서 간격이 균등합니다.

분포 선택 각 가속 요인에 대한 수명 분포 (로그 정규 또는 Weibull) 를 선택합니다. 자세한 내용은 "ALT 설계 플랫폼에 대한 통계 상세 정보 "(628 페이지) 에서 확인하십시오.

사전 평균 가속 모형 모수의 사전 추정값을 입력합니다 . 사전 추정값은 베이지안 사전 분포의 초모수입니다. 사전 평균 값은 주제 관련 지식을 활용한 최적 추측이거나 이전 연구에 기반할 수 있습니다.

가속 요인이 하나뿐인 경우 절편 또는 분위수를 지정할 수 있는 선택 옵션이 있습니다.

절편 지정 절편을 포함한 모형 모수의 사전 평균을 입력합니다.

분위수 지정 가속 요인의 사전 평균 , 지정된 시간의 기대 고장 비율 및 가속 요인 값을 입력합니다. 이 정보는 절편을 계산하는 데 사용됩니다.

사전 불확도 지정 가속 모형 모수의 사전 분포에 대한 표준 오차 및 상관 값을 선택하여 입력합니다. 표준 오차는 가속 모형 모수의 사전 추정값에 대한 불확도를 반영합니다.

참고 : 사전 불확도 측도를 지정하지 않으면 " 사전 평균 " 에서 입력된 값을 실제 모수 값으로 처리하여 설계가 생성됩니다. 사전 평균 모수가 실제 값에 가까우면 이 설계가 최적에 가깝습니다. 그러나 이 설계는 모수 추정값 오지정에 로버스트하지 않습니다. 사전 추정값이 확실하지 않은 경우 사전 불확도 지정을 사용하여 불확도를 반영해야 합니다.

진단 선택 사용 가능한 선택 옵션은 설계 최적 기준에 따라 달라집니다. 설계 최적 기준은 " 가속 수명 시험 계획 " 의 빨간색 삼각형 메뉴에 있습니다. D- 최적 설계의 경우 진단 선택 옵션이 없습니다.

관심 시간 범위 (고장 확률 최적 설계에만 사용 가능) 고장 나는 모집단의 비율을 추정할 시간 구간을 지정합니다. 왼쪽 상자에 하한 값을 입력하고 오른쪽 상자에 상한 값을 입력합니다. 특정 시점에 관심이 있는 경우 해당 값을 두 상자에 모두 입력합니다.

관심 확률 (분위수 추정값 최적 설계에만 사용 가능) 시간을 추정할 고장 비율을 지정합니다. 예를 들어 단위의 10% 가 고장 나는 시간을 추정하려면 0.10 을 입력합니다.

선택 설계 시험 시간, 검사 구간 및 시험 단위 수와 관련된 값을 입력합니다.

시험 시간 (연속 모니터링에만 사용 가능) 단위를 시험하는 시간을 지정합니다. 설계 테이블을 생성할 때 각 단위의 고장 시간 또는 우측 중도절단 여부를 기록합니다.

검사 시간 (구간 모니터링에만 사용 가능) 검사를 수행하는 시간을 지정합니다. 설계 테이블을 생성할 때 이 시간을 사용하여 **시작 시간** 및 **종료 시간** 열을 구성합니다. 각 구간에서 고장 나는 단위 수가 기록됩니다.

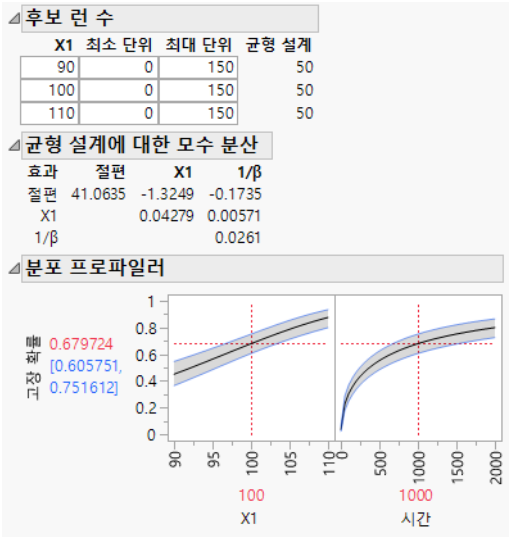
시험 단위 수 실험의 단위 수입니다.

- 초기 실험을 설계하는 경우 시험할 단위 수를 입력합니다.
- 이전 실험을 확대하는 경우 이전 실험에서 시험한 단위 수와 다음 실험의 단위 수를 더한 값을 입력합니다.

ALT 계획에 대한 규격 검토 및 업데이트

가속 수명 시험 계획의 설계 상세 정보를 지정하고 **계속**을 클릭하면 세 가지 새 개요가 창에 추가됩니다.

그림 22.19 추가 개요 노드



후보 런 수 가속 요인의 각 수준에서 허용되는 최소 및 최대 단위 수를 입력합니다. 이전 실험을 확대하는 경우에는 각 설정에 대해 각 수준에서 이미 실행한 단위 수를 " 최소 단위 " 열에 포함합니다.

균형 설계에 대한 모수 분산 균형 설계를 위해 가속 모형 모수 추정값의 공분산 행렬에 비례하는 행렬을 제공합니다.

θ 모수와 관련하여 모형의 첫 번째 편도함수 행렬을 **X** 로 나타내고, 오차 분산을 σ^2 으로 나타냅니다. 일반적인 조건에서 θ 의 최소 제곱 추정량은 다음과 같은 점근 공분산 행렬을 사용하여 점근적 비편향입니다.

$$Cov(\hat{\theta}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

" 균형 설계에 대한 모수 분산 " 개요는 **(X'X)-1** 을 제공하며 여기서 도함수가 수치로 계산됩니다. 이때 " 사전 평균 " 개요에 지정된 값이 실제 모수 값이라고 가정하여 계산이 수행됩니다.

자세한 내용은 " 비선형 설계 " 장의 " 비선형 모형 "(650 페이지) 에서 확인하십시오.

분포 프로파일러 가속 요인과 시간의 변화에 따라 균형 설계에 기반한 고장 확률을 탐색할 수 있습니다. 이 확률은 " 사전 평균 " 에 지정된 값이 실제 모수 값이라는 가정을 기반으로 합니다.

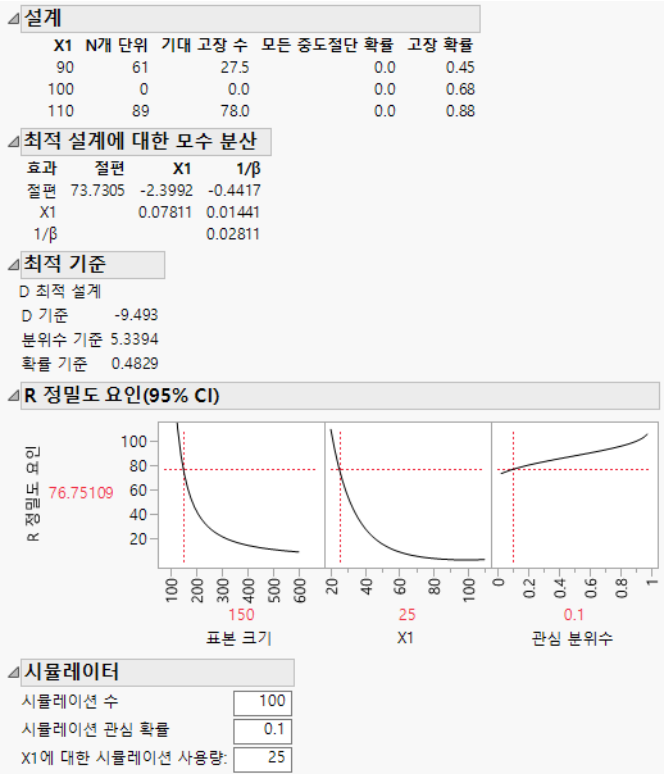
팁 : " 분포 선택 ", " 사전 평균 " 또는 " 선택 설계 " 를 변경한 경우 **프로파일러 업데이트**를 클릭하여 프로파일러를 업데이트하십시오.

최적 설계 생성 및 평가

설계 생성을 클릭하여 최적 ALT 설계를 생성하고 설계 품질에 대한 결과를 확인합니다.

참고 : 기본적으로 최적 설계는 D- 최적입니다 . 최적 기준을 변경하려면 " 가속 수명 시험 계획 " 의 빨간색 삼각형을 클릭하고 **ALT 최적 기준**을 선택합니다 .

그림 22.20 설계 개요



설계 가속 수준의 각 조합에서 시험할 단위 수를 표시합니다. 단일 요인의 경우 첫 번째 열은 요인 수준을 나타냅니다. 요인이 두 개인 경우 처음 두 열은 요인 수준을 나타냅니다.

N 개 단위 가속 수준의 각 조합에서 시험할 단위 수입니다 .

기대 고장 수 설계 설정에 대한 기대 고장 수입니다 . 기대 수는 사전 모형 규격을 사용하여 계산됩니다 .

모든 중도절단 확률 가속 수준 조합에서 시험한 단위가 하나도 고장 나지 않을 확률입니다 . 이 확률은 사전 모형 규격을 사용하여 계산됩니다 .

고장 확률 설계 설정에 대한 고장 확률 기대값입니다 . 고장 확률은 고장 기대값을 시험 단위 수로 나누어서 계산합니다 .

참고 : 기대 고장 수와 중도절단 확률을 통해 사전 규격이 적절한지 판단할 수 있습니다 .

최적 설계에 대한 모수 분산 최적 설계에 대한 가속 모형 모수 추정값의 공분산 행렬에 비례하는 행렬을 제공합니다. 이때 " 사전 평균 " 개요에 지정된 값이 실제 모수 값이라고 가정하여 계산이 수행됩니다. 자세한 내용은 " 균형 설계에 대한 모수 분산 "(625 페이지) 에서 확인하십시오.

참고 : " 최적 설계에 대한 모수 분산 " 행렬의 값을 " 균형 설계에 대한 모수 분산 " 행렬의 값과 비교하여 최적 설계에 의해 추정값 분산이 어느 정도 감소되는지 확인할 수 있습니다.

최적 기준 다음과 같은 세 가지 최적 기준의 값이 제공됩니다.

D 기준 설계의 D- 최적. 자세한 내용은 "D- 최적 설계 생성 "(628 페이지) 에서 확인하십시오.

분위수 기준 설계의 시간 I- 최적. 자세한 내용은 " 최적 분위수 추정값 생성 "(628 페이지) 에서 확인하십시오.

확률 기준 설계의 확률 I- 최적. 자세한 내용은 " 최적 고장 확률 생성 "(628 페이지) 에서 확인하십시오.

R 정밀도 요인 (95% CI) 표본 크기, 요인 수준 및 관심 분위수가 R 정밀도 요인에 미치는 영향을 평가하기 위한 프로파일러입니다. R 정밀도 요인 (95% CI) 은 관심 분위수의 95% 신뢰 구간에 대한 정밀도의 측도입니다.

시뮬레이터 데이터를 수집하기 전에 시뮬레이터를 사용하여 가속 모형을 탐색합니다.

시뮬레이션 수 실행할 시뮬레이션 시도 수입니다.

시뮬레이션 관심 확률 관심 있는 고장 확률입니다.

시뮬레이션 사용량 가속 요인의 사용 조건 값입니다.

" 시뮬레이션 실행 " 을 클릭하면 요약 통계량 테이블과 모형 적합 그림이 나타납니다.

설계 테이블 생성

이 버튼을 사용하여 가속 수명 시험 설계, 시험 계획 및 테이블을 생성합니다.

설계 생성 " 분포 선택 ", " 사전 평균 ", " 사전 분산 행렬 ", " 선택 설계 " 또는 " 후보 런 수 " 를 변경한 경우 최적 설계를 업데이트합니다.

시험 계획 생성 각 행이 개별 설계 설정에 해당하는 데이터 테이블을 생성합니다. 이 테이블에는 가속 요인 설계 설정 및 해당 설계 설정에서 시험할 단위 수가 나타납니다.

테이블 생성 고장 시간 데이터를 기록하는 데 사용할 수 있는 테이블을 생성합니다.

- 연속 모니터링의 경우, 시험할 각 단위와 해당 단위의 설계 설정에 해당하는 행이 테이블에 포함됩니다. 실험 중에 단위가 고장 나면 **고장 시간** 열에 해당 단위의 고장 시간을 기록하고 **중도절단** 열에 0 을 기록합니다. 실험 중에 단위가 고장 나지 않으면 **고장 시간** 열에 최종 검사 시간을 기록하고 **중도절단** 열에 1 을 기록합니다.
- 구간 모니터링의 경우, 각 설계 설정과 시간 구간 조합에 해당하는 행이 테이블에 포함됩니다. 시간 구간은 " 선택 설계 " 개요에 입력한 " 검사 시간 " 을 기반으로 하는 **시작 시간**

및 종료 시간 열에 의해 정의됩니다 . 각 설정 및 시간 구간에 대해 고장 난 단위의 수를 고장 수 열에 기록합니다 .

가속 수명 시험의 옵션

" 가속 수명 시험 계획 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

반응 시뮬레이션 테이블 생성을 클릭하면 시뮬레이션된 반응을 테이블에 추가합니다 . 시뮬레이션된 반응은 " 사전 평균 " 개요에 지정한 모수 값에 대해 선택한 분포에서 무작위로 추출하여 생성됩니다 . 시뮬레이션된 반응이 지정된 시험 시간을 초과하면 시험 시간 값에서 관측이 중도절단됩니다 .

ALT 최적 기준

D- 최적 설계 생성 몬테카를로 구의 수가 0 보다 크면 베이지안 D- 최적 설계를 생성합니다 . 최적 기준은 사전 분포와 관련한 정보 행렬의 행렬식에 대한 로그 기대값입니다 . 몬테카를로 구의 수가 0 이면 로컬 D- 최적 설계를 생성합니다 . 결과적으로 D- 최적은 계수의 정확한 추정에 중점을 둡니다 .

최적 분위수 추정값 생성 최적 기준 기본값입니다 . 지정된 사용 조건에서 지정된 백분위수의 분산을 최소화하는 설계를 생성합니다 .

최적 고장 확률 생성 지정된 사용 조건에서 관심 시간 범위 동안 고장 확률 추정값의 분산을 최소화하는 설계를 생성합니다 .

고급 옵션

N 몬테카를로 구 적분에 사용되는 0 이 아닌 반지름 값의 수를 설정할 수 있습니다 . JMP 는 주어진 최적 기준을 최적화하는 비선형 설계를 찾기 위해 모수 사전 분포에 대한 Fisher 정보 행렬의 행렬식에 대한 로그 적분을 최소화합니다 . 이러한 적분은 수치로 계산됩니다 . 적분 수행 방법에 대한 자세한 내용은 " 비선형 설계 " 장의 " 비선형 설계 옵션 "(649 페이지) 에서 확인하십시오 .

팁 : 기본적으로 "N 몬테카를로 구 " 는 4 로 설정되어 있습니다 . 값이 클수록 수치 정확도는 향상되지만 계산 시간은 길어집니다 .

스크립트를 스크립트 창에 저장 " 가속 수명 시험 계획 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

ALT 설계 플랫폼에 대한 통계 상세 정보

ALT 설계의 고장 분포

가속 수명 시험 설계 플랫폼에서는 로그 정규 또는 Weibull 고장 분포를 선택할 수 있습니다 .

이 섹션에는 각 분포의 pdf(확률 밀도 함수)와 cdf(누적 분포 함수)에 대한 파라미터화가 포함되어 있습니다. Weibull 분포에 대한 자세한 내용은 Reliability and Survival Methods의 에서 확인하십시오.

로그 정규

로그 정규 분포는 데이터 범위가 10의 여러 거듭제곱일 경우 고장 시간을 나타내는 데 흔히 사용됩니다. 보통 이 분포는 독립적이며 같은 분포를 따르는 다수의 작은 양수 확률 변수의 곱으로 개념화됩니다. 이 분포는 데이터 값의 로그가 정규 분포로 나타날 때 적절합니다. 일반적으로 확률 분포 함수는 강한 오른쪽 왜도가 특징입니다.

로그 정규 계열은 위치 모수 (μ) 와 형태 모수 (σ) 로 파라미터화됩니다 . 로그 정규 pdf 및 cdf 는 다음과 같이 지정되며 , 여기서 로그는 밑이 e 입니다 .

$$f(x;\mu,\sigma) = \frac{1}{x\sigma}\phi_{nor}\left[\frac{\log(x)-\mu}{\sigma}\right], \quad x>0$$

$$F(x;\mu,\sigma) = \Phi_{nor}\left[\frac{\log(x)-\mu}{\sigma}\right],$$

아래 함수는

$$\phi_{nor}(z)= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$$

및

$$\Phi_{nor}(z) = \int_{-\infty}^{z}\phi_{nor}(w)dw$$

각각 표준 정규 분포(N(0,1))에 대한 pdf와 cdf입니다.

Weibull

Weibull 분포를 사용하면 위험 비율이 증가하거나 감소하는 고장 시간 데이터를 모델링할 수 있습니다. 이 분포는 형태 모수 값을 기반으로 다양한 유형의 데이터를 모델링할 때 엄청난 유연성을 제공하므로 신뢰성 분석에 자주 사용됩니다.

일반적으로 Weibull pdf 와 cdf 는 다음과 같이 나타냅니다 .

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha^{\beta}}x^{(\beta-1)}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]; \qquad x>0,\ \alpha>0,\ \beta>0$$

$$F(x;\alpha,\beta) = 1-\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]$$

여기서 α는 척도 모수이고 β는 형태 모수입니다. Weibull 분포는 β = 1 일 때 지수 분포로 축소됩니다.

Weibull 파라미터화

JMP 에서 Weibull 파라미터화는 σ 를 척도 모수로 , μ 를 위치 모수로 정의합니다 . 이 모수는 다음과 같이 α 및 β Weibull 파라미터화와 관련됩니다 .

$$\alpha = \exp(\mu)$$

및

$$\beta = \frac{1}{\sigma}$$

이러한 모수를 사용할 경우 Weibull 분포의 pdf 와 cdf 는 μ = log(α) 및 σ = 1/β 인 위치 - 척도 파라미터화를 사용하여 로그 변환된 SEV(최소 극단값 분포) 로 표현됩니다 .

$$f(x;\mu,\sigma) = \frac{1}{x\sigma}\phi_{sev}\left[\frac{\log(x)-\mu}{\sigma}\right], \quad x > 0, \sigma > 0$$

$$F(x;\mu,\sigma) = \Phi_{sev}\left[\frac{\log(x)-\mu}{\sigma}\right]$$

여기서 ,

$$\phi_{sev}(z) = \exp[z - \exp(z)]$$

및

$$\Phi_{sev}(z) = 1 - \exp[-\exp(z)]$$

각각 표준화된 최소 극단값(μ = 0, σ = 1) 분포에 대한 pdf 와 cdf 입니다.

R 정밀도 요인 (95% CI)

R 정밀도 요인은 고장 시간 분포의 p 번째 분위수에 대한 t_p 추정값 정밀도의 측도입니다. 로그 (t_p) 에 대한 대략적인 95% 신뢰 구간은 다음과 같이 계산됩니다.

$$\log(\hat{t}_p) \pm z_{0.975}\sqrt{\widehat{Var}[\log(\hat{t}_p)]} = \log(\hat{t}_p) \pm \log(\hat{R})$$

t_p 에 대한 신뢰 구간은 다음과 같아집니다.

$$[\hat{t}_p/\hat{R}, \hat{t}_p\hat{R}]$$

다음은 각 요소에 대한 설명입니다.

$$\hat{R} = \exp\left[z_{0.975}\sqrt{\widehat{Var}[\log(\hat{t}_p)]}\right]$$

R을 구하려면 다음과 같이 로그(t_p)의 분산에 대한 대표본 근사치를 사용합니다.

$$R = \exp z_{0.975}\sqrt{AVar[\log(\hat{t}_p)]}$$

가속 수명 시험에 대한 최적 기준

가속 수명 시험(ALT)이 최대 두 개의 가속 요소 x_1과 x_2(수준이 각각 J_1과 J_2)를 사용하여 수행된다고 가정해 보겠습니다. 시험 계획의 목표는 각 요인 수준 조합에 할당할 표본 수를 결정하는 것입니다.

최적 기준 계산은 각 수준별 정보 행렬에서 가져온 총 Fisher 정보 행렬과

$$I(\theta) = \sum_{i,j} n_{ij} I_{ij}(\theta)$$

사전 불확도 S(해당하는 경우)의 포함 여부에 따라 달라집니다.

D- 최적

D-최적 설계 $\mathbf{D}_d$는 가능한 모든 n_{jj} 값에 대해 다음 표현식을 최소화하는 설계입니다.

$$\log\left|\int_{\theta}[\mathbf{S}^{-1} + \mathbf{I}(\theta)]^{-1}\pi(\theta)d\theta\right|$$

여기서 |·| 부분은 행렬식 연산이고 $\pi(\theta)$는 모수에 대한 사전 분포입니다. 이 분포는 보통 다변량 정규 분포입니다. 행렬식의 로그는 수치 안정성을 위해 사용됩니다.

분위수 최적

분위수 최적 설계 $\mathbf{D}_q$는 다음 표현식을 최소화하는 설계입니다.

$$\int_{\theta}[\mathbf{c}'\mathbf{I}(\theta)c]\pi(\theta)d\theta$$

여기서 벡터 **c**는 사용 조건에 따라 달라집니다.

단일 사용 조건의 경우

$$\mathbf{c}' = \{1, 0, 0, 0, z_p\}$$

여기서 z_p는 고장 분포의 p번째 분위수입니다.

사용 조건이 범위인 경우에는 다음과 같습니다.

$$\varepsilon_i = \frac{G(x_{1iUh}) - G(x_{1iUl})}{g(x_{iUh}) - g(x_{iUl})}$$

여기서 $G(\cdot)$는 $g(\cdot)$의 부정적분이고, x_{iUh}는 i번째 요인에 대해 고려되는 가장 높은 사용 수준이고, x_{iUl}는 동일한 요인에 대해 고려되는 가장 낮은 사용 수준입니다. 결과

$$\mathbf{c}' = \{1, \varepsilon_1, \varepsilon_2, \varepsilon_1, \varepsilon_2, z_p\}$$

여기서 z_p는 고장 분포의 p번째 분위수입니다.

고장 확률 최적

고장 확률 최적 설계 $\mathbf{D}_f$는 다음 표현식을 최소화하는 설계입니다.

$$\int_{\theta} \left\{ \frac{\frac{1}{\sigma^2}\left[\frac{\phi(z_u(\theta))}{\Phi(z_u(\theta))}\right]^2 c'_{fp} I(\theta) c_{fp}}{-\log(\Phi(z_u(\theta)))} \right\} (\pi(\theta) d\theta)$$

여기서 $z_U(\boldsymbol{\theta})$는 "분위수 최적"에 정의된 사용 조건에서의 표준화된 로그 시간입니다. $\mathbf{c}_{fp}$는 "분위수 최적"에 정의된 $\mathbf{c}$와 유사하지만 마지막 항목이 $z_U(\boldsymbol{\theta})$라는 점이 다릅니다. 분자의 값은 $\log\Phi(z_U(\boldsymbol{\theta}))$의 점근 분산을 기반으로 하며, 여기서 로그 변환은 수치 안정성을 위해 사용됩니다. 그런 다음 점근 분산은 일관성을 위해 원래 $\log\Phi(z_U(\boldsymbol{\theta}))$에 따라 표준화되며 더 작은 값일수록 좋습니다. 사용 조건이 범위인 경우 계산 편의를 위해 점근 분산에 대해서가 아니라 $\log\Phi(\cdot)$ 안에서 적분이 계산됩니다. 합계의 분산이 분산의 합계를 제한하고 로그 함수와 cdf가 둘 다 단조 함수라는 특성이 있으므로 적합 측정 기준으로서 여전히 적합합니다. 따라서 위의 표현식을 최소화하면 위로부터 서서히 제한됨으로써 고장 확률의 실제 점근 분산이 최소화됩니다.

23 장

비선형 설계

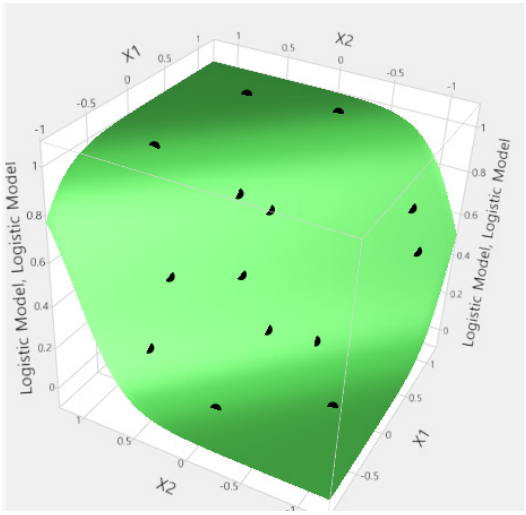
미지의 모수에 비선형 모형을 적합시키는 것이 실험 목표인 경우 비선형 모형 적합에 중요한 영역에 설계점을 배치하려면 비선형 설계를 사용합니다. 선형 모형에 최적인 직교 설계를 사용할 수 있지만 일반적으로 이 설계는 적합 모수 추정값의 불확도를 최소화하거나 정밀도를 최대화하는 위치에 설계점을 놓지 않습니다.

비선형 설계의 효율도는 미지의 모수 값에 따라 달라집니다. 이는 최선의 설계를 찾기 위해 모수를 미리 알아야 한다는 점에서 순환 문제의 원인이 됩니다. JMP는 베이지안 방법을 사용하여 모수 값의 지정된 범위에서 평균 효율도를 최대화하는 비선형 설계를 구성합니다. 이러한 범위를 적절하게 지정하려면 관심 시스템에 대한 정보가 있어야 합니다.

비선형 설계는 선형 모형 설계와 비교하여 다음과 같은 장점을 제공합니다.

- 잘 선택된 모형을 사용한 예측은 요인 설정이 다양한 경우에 적합합니다.
- 복잡한 곡률과 점근적 모습을 갖는 반응 표면을 모델링할 수 있습니다.

그림 23.1 비선형 모형의 설계점



목차

비선형 설계 개요 635
비선형 설계의 예 636
사전 데이터가 없는 비선형 설계 생성 636
사전 데이터를 사용하여 설계 확대 640
이항 반응에 대한 설계 생성 644
비선형 설계 시작 창 647
비선형 설계 창 647
요인 648
모수 648
설계 생성 649
설계 649
테이블 생성 또는 확대 649
비선형 설계 옵션 649
비선형 설계에 대한 통계 상세 정보 650
비선형 모형 650
최적 기준의 방사형-구형 적분 651
최적 설계 찾기 651

비선형 설계 개요

비선형 설계 플랫폼을 사용하여 모수에서 비선형인 모형을 적합시키는 설계를 구성합니다. 비선형 모형에 대해 최적 설계를 구성하거나 기존 데이터를 최적으로 확대할 수 있습니다. 기본 공정을 기술하는 정보를 기반으로 한 비선형 설계는 다항식 모형에 대한 표준 설계를 사용할 때보다 더 정확하게 모형 모수의 값을 추정하고 공정 동작을 예측할 수 있습니다. 비선형 모형에 대한 배경 지식은 "비선형 모형"(650 페이지)에서 확인하십시오.

비선형 모형의 설계 효율도는 설계에서 추정하려는 모수의 알려지지 않은 값에 따라 달라집니다. 이러한 이유로 JMP는 베이지안 방법을 사용하여 광범위한 모수 확률 값에 대해 효율적인 설계를 구성합니다. 미지의 모수 값 범위와 사전 분포를 지정할 수 있습니다. 사전 분포는 "균등", "정규", "로그 정규" 및 "지수" 중에서 선택할 수 있습니다.

비선형 설계 플랫폼은 사용자가 지정하는 모수 확률 값의 사전 분포에 대해 설계를 최적화하는 베이지안 방법을 사용합니다. 베이지안 D-최적 기준은 이 사전 확률 분포를 나타내는 모수 벡터 표본과 관련하여 정보 행렬의 행렬식에 대한 로그 기대값입니다. 정보 행렬 항목은 설계점에서의 예측 분산에 따라 달라집니다. 분산이 작은 관측값, 즉 반응이 거의 확실한 경우에는 제공되는 정보가 매우 적습니다. 따라서 최적 설계는 일부 설계 설정을 분산이 높은 점에 배치합니다. 자세한 내용은 Gotwalt et al. (2009)에서 확인하십시오.

최적 설계의 주요 내용은 실현 가능한 실험 영역에서 최적 설계가 예측 분산이 가장 높은 위치에 점을 배치한다는 것입니다. 이 사실은 직관에 반대되는 것 같지만 대체 설계가 다른 위치에 점을 배치하면 최적 설계의 설계점에 대한 예측 분산이 훨씬 높아집니다. 모수에서 선형인 모형의 경우 분산이 높은 점은 실험 영역의 꼭지점에 있는 경향이 있습니다. 그러나 모수에서 비선형인 모형의 경우에는 그렇지 않을 수 있습니다.

참고 : 비선형 설계는 랜덤 시작 설계를 사용하여 계산됩니다 . 따라서 동일한 규격에 대해 생성된 비선형 설계가 대부분 다릅니다 .

비선형 설계 플랫폼을 사용하려면 기존 데이터 테이블이 있어야 합니다 . 해당 데이터 테이블에는 다음 요소가 포함되어야 합니다 .

- 반응에 대한 열
- 각 요인에 대한 열
- 요인과 반응 간의 관계를 보여 주는 계산식이 포함된 열 . 이 계산식은 미지의 모수를 포함해야 합니다 .

참고 : 이것은 모델링을 위해 비선형 플랫폼에서 사용되는 데이터 테이블에 필요한 형식과 동일합니다 .

테이블은 다음 두 형식 중 하나일 수 있습니다 .

- 열 정보만 포함하고 행은 없는 템플릿일 수 있습니다. 자세한 내용은 "사전 데이터가 없는 비선형 설계 생성 "(636 페이지) 에서 확인하십시오 .

- 예측 변수 정보가 있는 행을 포함할 수 있습니다. 이 경우 예측 변수 값이 비선형 설계에 포함됩니다. 자세한 내용은 "사전 데이터를 사용하여 설계 확대"(640 페이지)에서 확인하십시오.

비선형 설계의 예

- "사전 데이터가 없는 비선형 설계 생성"
- "사전 데이터를 사용하여 설계 확대"
- "이항 반응에 대한 설계 생성"

사전 데이터가 없는 비선형 설계 생성

이 예에서는 아직 데이터를 수집하지 않았지만 미지의 모수에 대해 추측할 수 있을 때 설계를 생성하는 방법을 보여 줍니다. 여기에서는 화학 반응에서 중간 생성물의 분별 수율(Observed Yield)을 모델링합니다. 분별 수율은 반응 시간과 온도의 함수입니다. 자세한 내용은 Box and Draper(1987)에서 확인하십시오.

설계 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reaction Kinetics Start.jmp를 엽니다.

   다음을 확인할 수 있습니다.

   – 데이터가 수집되지 않았으므로 데이터 테이블에 행이 없습니다.

   – 예측 변수 Reaction Temperature 와 Reaction Time 의 열에는 "코딩", "설계 역할" 및 "요인 변경" 특성이 있습니다. 특성을 확인하려면 "열" 패널에서 ✱을 클릭합니다. 이러한 특성은 설계를 구성할 때 예측 변수를 처리하는 방법을 JMP 에 알려 줍니다. 열 특성을 저장하는 방법에 대한 자세한 내용은 "열 특성" 장 (727 페이지)에서 확인하십시오.

   – Observed Yield 열에는 실험을 실행하여 얻은 반응 데이터가 포함됩니다.

   – Yield Model 열은 예측 변수를 반응 (Observed Yield) 과 관련짓는 계산식을 포함합니다. 계산식을 보려면 "열" 패널에서 ✙을 클릭합니다. 계산식은 t1 및 t3 모수에서 비선형입니다.

2. **DOE > 특수 용도 > 비선형 설계**를 선택합니다.
3. Observed Yield 를 선택하고 **Y, 반응**을 클릭합니다.
4. Yield Model 을 선택하고 **X, 예측변수 계산식**을 클릭합니다.
5. **확인**을 클릭합니다.

   이 예에서 Reaction Temperature 값 (510, 540) 및 Reaction Time 값 (0.1, 0.3) 은 "코딩" 열 특성을 사용하여 지정되었습니다. "요인" 개요에서 적절한 값 범위를 직접 지정할 수도 있습니다.

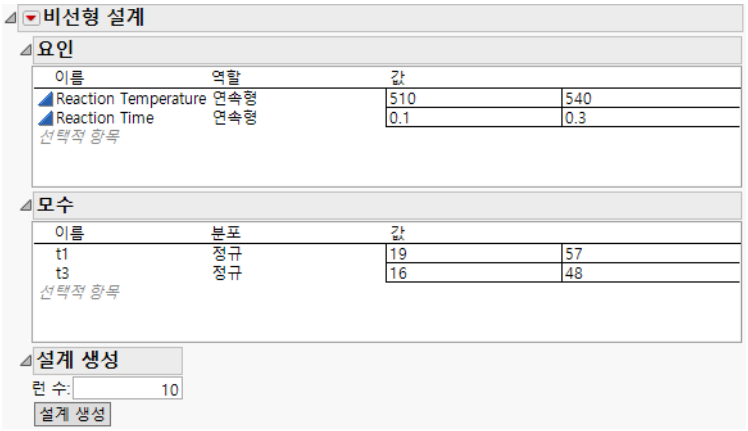
6. 모수 t1 의 값을 25 와 50 으로 변경하고 t3 의 값을 30 과 35 로 변경합니다.

이 새 값은 실험 상황에 대해 적절한 범위의 모수 값을 나타냅니다. 기본값은 예측 계산식 정의에 지정된 초기 모수 값을 기반으로 구성되었습니다 . 계산식 생성에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

t1 과 t3 의 각 " 분포 " 아래에 표시된 사전 분포는 기본적으로 " 정규 " 로 설정되어 있습니다 .

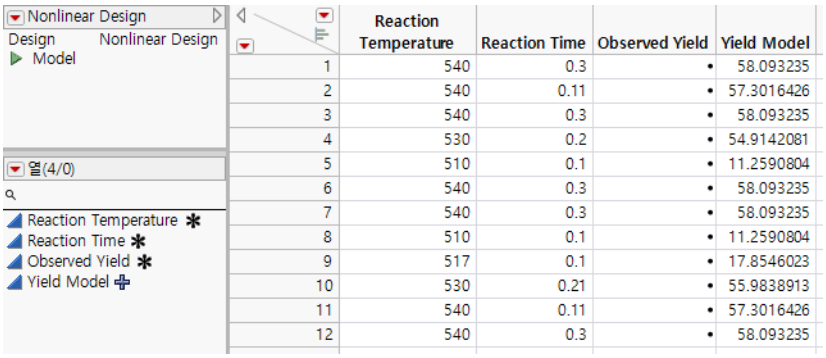
7. " 설계 생성 " 패널에서 런 수를 12 로 변경합니다 .

그림 23.2 Reaction Kinetics 실험의 완료된 개요



8. **설계 생성**을 클릭합니다 .
9. **테이블 생성**을 클릭합니다 .

그림 23.3 설계 테이블



	Reaction Temperature	Reaction Time	Observed Yield	Yield Model
1	540	0.3	•	58.093235
2	540	0.11	•	57.3016426
3	540	0.3	•	58.093235
4	530	0.2	•	54.9142081
5	510	0.1	•	11.2590804
6	540	0.3	•	58.093235
7	540	0.3	•	58.093235
8	510	0.1	•	11.2590804
9	517	0.1	•	17.8546023
10	530	0.21	•	55.9838913
11	540	0.11	•	57.3016426
12	540	0.3	•	58.093235

그림 23.3의 설계와 유사한 설계가 표시됩니다. 런 순서는 다를 수 있으며 이에 따라 Reaction Temperature 및 Reaction Time 의 값과 Yield Model 에 대해 계산된 값도 약간 다를 수 있습니다 . Yield Model 열에는 모형의 계산식이 포함되므로 값이 나타납니다 . 또한 비선형 모형을 관측값에 적합시키는 데 사용할 수 있는 **모형** 스크립트가 테이블에 포함되어 있습니다 .

설계 테이블을 생성했으므로 실험을 실행하고 Observed Yield 열에 반응을 기록합니다 . Design Experiment 폴더의 Reaction Kinetics.jmp 데이터 테이블에 설계에 대한 관측 결과가 포함되어 있습니다 .

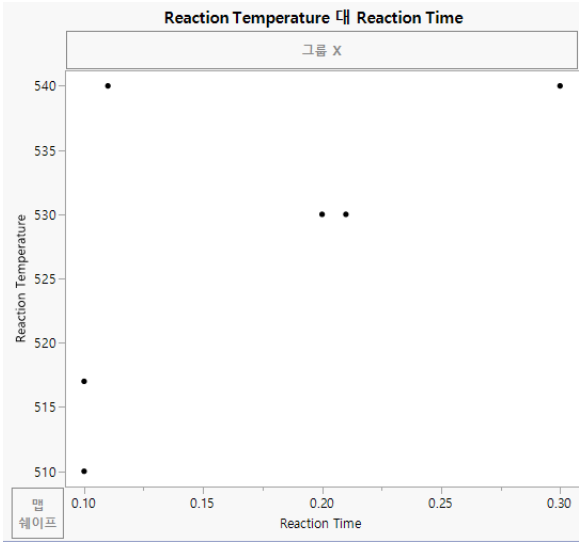
설계 탐색

결과를 분석하기 전에 설계 설정을 볼 수 있는 그림을 구성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reaction Kinetics.jmp 를 엽니다 .

1. **그래프 > 그래프 빌더**를 선택합니다 .
2. Reaction Temperature 를 **Y** 놓기 영역으로 드래그합니다 .
3. Reaction Time 을 **X** 놓기 영역으로 드래그합니다 .
4. 그래프 위의 두 번째 아이콘을 클릭하여 평활기를 선택 취소합니다 .

그림 23.4 설계 설정

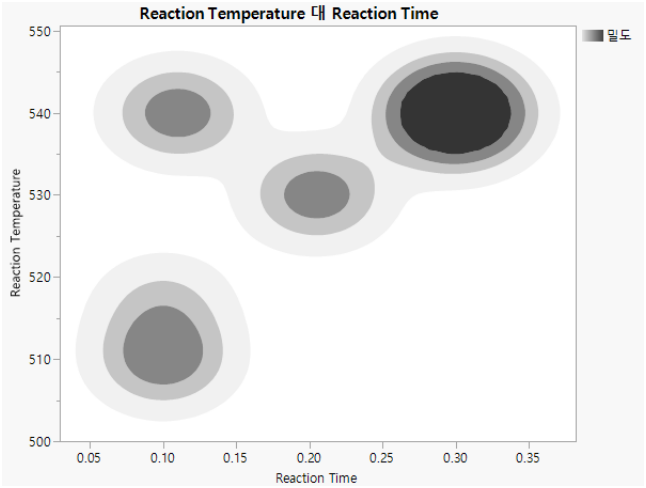


점이 세 영역에 있습니다 . 낮은 온도와 높은 시간 (그래프의 오른쪽 아래 부분) 에는 점이 없습니다 . 직교 설계와 달리 비선형 설계는 설계점을 설계 영역 모서리에 놓지 않을 수도 있습니다 . 이 예에서 낮은 온도와 높은 시간의 설계점은 비효율적입니다 .

나머지 세 모서리에서 설계점의 밀도를 보려면 등고선 도구를 사용합니다 .

5. 을 클릭하여 등고선 도구를 설정합니다 .
6. **완료**를 클릭합니다 .

그림 23.5 밀도 등고선을 사용한 설계 설정



낮은 시간과 높은 온도에서는 상대적으로 점 수가 적습니다. 설계 테이블에서 해당 점이 두 개뿐임을 확인할 수 있습니다. 모형 및 모수 규격 때문에 최적 설계는 높은 시간과 높은 온도에 설계점을 더 놓습니다.

결과 분석

설계를 시각적으로 탐색했으므로 이제 결과를 분석합니다.

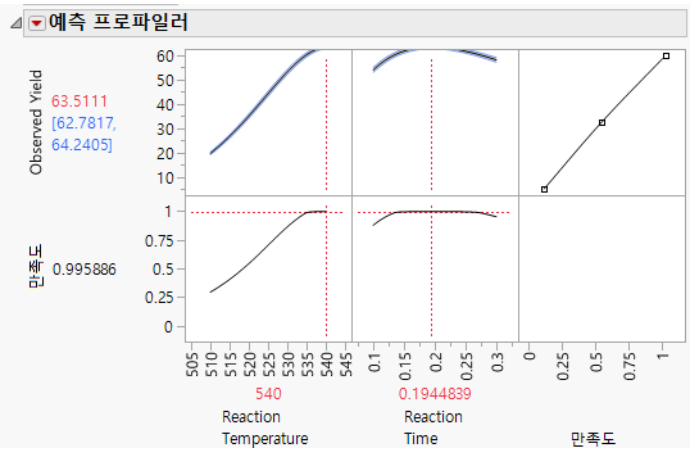
참고 : 1 단계 ~ 4 단계를 수행하는 대신 **모형** 스크립트를 실행할 수 있습니다.

1. **분석 > 전문 모델링 > 비선형**을 선택합니다.
2. Observed Yield 를 선택하고 **Y, 반응**을 클릭합니다.
3. Yield Model 을 선택하고 **X, 예측변수 계산식**을 클릭합니다.

   " 사용자 계산식 적합 옵션 " 패널에 모형이 나타납니다.
4. **확인**을 클릭합니다.
5. " 제어판 " 에서 **시작**을 클릭합니다.

   " 중지 한계 " 값 중 하나에 도달할 때까지 해 반복 검색 작업이 진행됩니다. 그런 다음 " 해 " 및 " 추정값 상관계수 " 보고서가 나타납니다.
6. " 비선형 적합 " 의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다.
7. 수율을 최대화하려면 " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다.

그림 23.6 최대 수율을 위한 시간 및 온도 설정



추정된 최대 수율은 반응 온도 540 도 (켈빈) 와 반응 시간 0.1945 분에서 약 63.5% 입니다 .

사전 데이터를 사용하여 설계 확대

이 예에서는 사전 데이터가 있을 때 비선형 설계를 생성하는 방법을 보여 줍니다. 여기에서는 화학 반응 데이터를 사용합니다. 사용 가능한 유기 기질의 흡수율(속도)을 해당 기질의 농도 함수로 모델링하려고 합니다. 자세한 내용은 Meyers(1986)에서 확인하십시오. 이미 실험을 실행했지만 결과를 활용하여 더 정확한 모수 추정값을 얻는 것이 목표입니다.

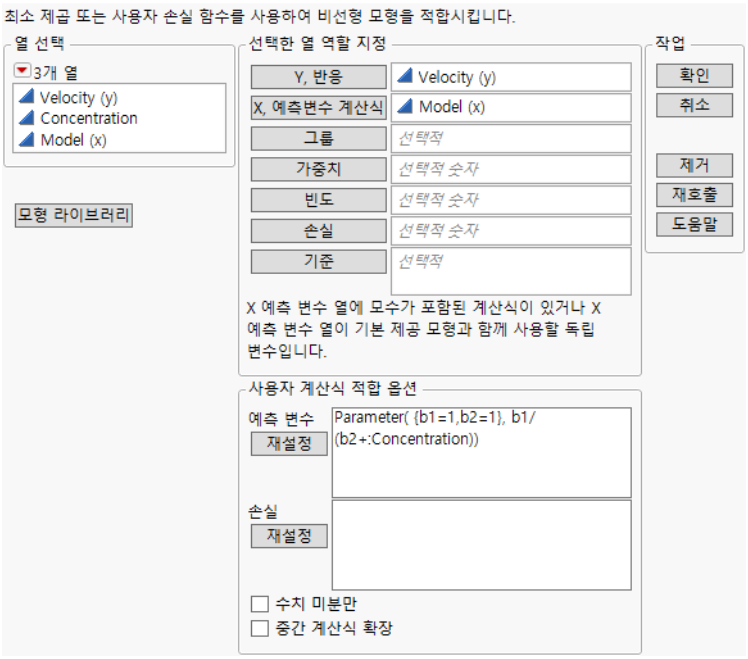
사전 모수 추정값 구하기

기존 실험 데이터를 사용하여 더 적합한 모수 추정값을 구합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nonlinear Examples/Chemical Kinetics.jmp를 엽니다 .
2. " 열 " 패널에서 Model (x) 옆의 더하기 기호를 클릭합니다 . 계산식 편집기가 열립니다 .
3. 계산식 편집기의 아래쪽 중간에 있는 **모수** 개요에는 모형 모수의 현재 값이 표시됩니다 . 값 (b1 = 1, b2 = 1) 은 초기 추정값입니다 . 이 값은 데이터 테이블에서 Model (x) 값을 계산하는 데 사용됩니다 . 다음 실험을 위해 이 값을 더 적합한 추정값으로 바꾸려고 합니다 .
4. **취소**를 클릭하여 계산식 편집기 창을 닫습니다 .
5. **분석 > 전문 모델링 > 비선형**을 선택합니다 .
6. Velocity (y) 를 선택하고 **Y, 반응**을 클릭합니다 .
7. Model (x) 를 선택하고 **X, 예측변수 계산식**을 클릭합니다 .

   Model (x) 에 의해 제공된 계산식이 " 사용자 계산식 적합 옵션 " 패널에 나타납니다 .

그림 23.7 비선형 분석 시작 창

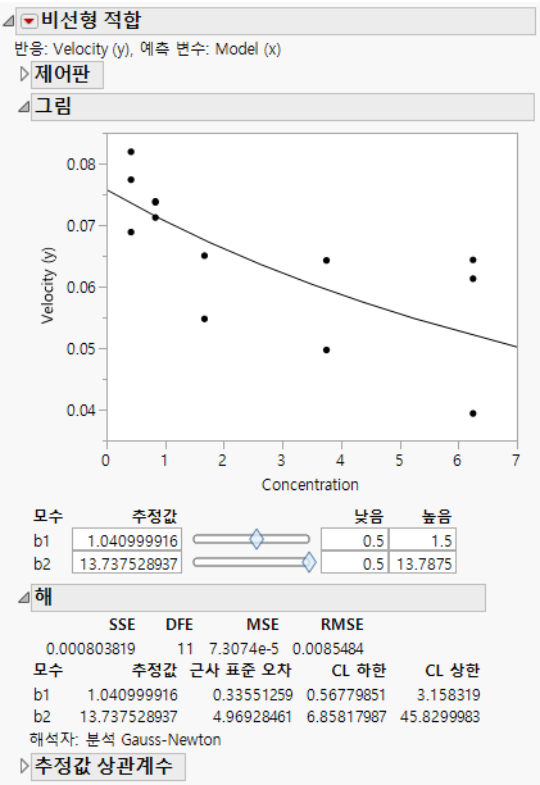


8. **확인**을 클릭합니다 .
9. " 제어판 " 에서 **시작**을 클릭합니다 .

   " 중지 한계 " 값 중 하나에 도달할 때까지 해 반복 검색 작업이 진행됩니다 . 그런 다음 " 해 " 및 " 추정값 상관계수 " 보고서가 나타납니다 . 또한 " 해 " 보고서에 신뢰 한계를 추가할 수 있는 옵션이 " 제어판 " 에 나타납니다 .
10. " 제어판 " 에서 **신뢰 한계**를 클릭합니다 .

    모수 b1 및 b2 에 대한 신뢰 구간이 " 해 " 보고서에 나타납니다 .

그림 23.8 비선형 적합 결과



b1 과 b2 의 "CL 하한" 및 "CL 상한" 값은 모수의 값 범위를 정의합니다. 이제 확대된 비선형 설계에서 이 구간을 사용하여 사전 값의 범위를 정의합니다.

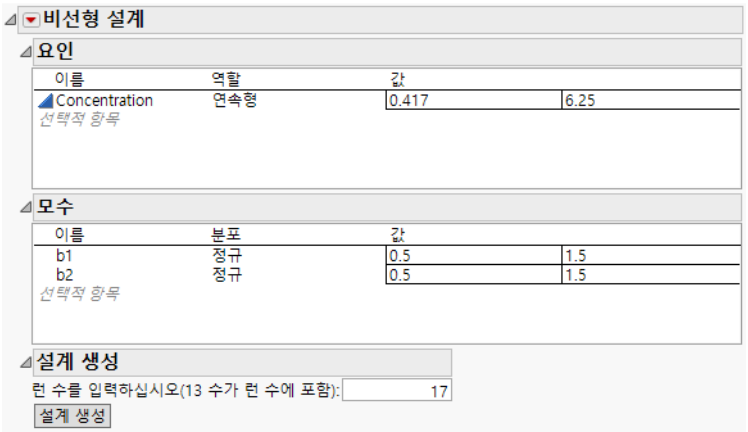
참고 : 이 결과는 다음 단계에서 필요하므로 "비선형 적합" 보고서를 닫지 마십시오.

설계 확대

이제 비선형 모수를 더 정확하게 추정하기 위한 설계를 생성합니다.

1. Chemical Kinetics.jmp 데이터 테이블을 활성화하고 **DOE > 특수 용도 > 비선형 설계**를 선택합니다.
2. Velocity (y) 를 선택하고 **Y, 반응**을 클릭합니다.
3. Model (x) 를 선택하고 **X, 예측변수 계산식**을 클릭합니다.
4. **확인**을 클릭합니다.

그림 23.9 비선형 설계의 요인 및 모수 개요

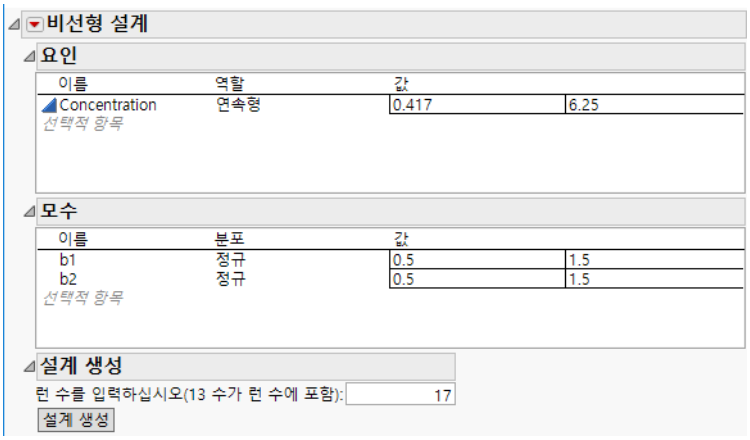


Chemical Kinetics.jmp 데이터에서 Concentration 값 범위는 0.417 ~ 6.25 입니다 . 따라서 이 값은 처음에 " 요인 " 개요에 하한값과 상한값으로 나타납니다 . 더 넓은 구간을 포함하도록 이 값을 변경하려고 합니다 .

5. 0.417 을 클릭하고 0.1 을 입력합니다 . 6.25 를 클릭하고 7 을 입력합니다 .
6. 각 모수의 사전 분포는 설정값 " 정규 " 를 그대로 둡니다 .

   모수의 값 범위는 해당 모수에 대한 지식의 불확도를 반영합니다 . 가능한 모수 값의 95% 를 포함하는 범위를 지정할 수 있습니다 . 그림 23.8 에 표시된 " 비선형 적합 " 보고서의 신뢰 한계는 이러한 범위를 제공합니다 . " 모수 " 개요에서 모수 값을 신뢰 한계 (소수점 이하 세 자리로 반올림) 로 바꿉니다 .
7. DOE " 비선형 설계 " 창의 " 모수 " 에서 다음 값을 b1 및 b2 에 입력합니다 .
   - b1: 0.568, 3.158
   - b2: 6.858, 45.830

그림 23.10 업데이트된 요인 및 모수 값



8. " 설계 생성 " 패널의 **런 수**에 40 을 입력합니다 .
9. **설계 생성**을 클릭합니다 .

   " 설계 " 개요가 열리고 원래 13 회 런에 대한 Concentration 및 Velocity (y) 값과 추가 27 회 런에 대한 새 Concentration 설정을 보여 줍니다 .
10. **테이블 생성**을 클릭합니다 .

    이렇게 하면 원래 13 회 런 설계에 대한 설정 및 결과와 27 회 새 런에 대한 설정이 포함된 JMP 설계 테이블이 새로 생성됩니다 . " 테이블 생성 " 대신 " 테이블 확대 " 를 클릭하면 새 데이터 테이블을 생성하지 않고 기존 데이터 테이블에 새 런을 추가할 수 있습니다 .

새 런은 더 넓은 Concentration 값 구간과 원래 실험에서 구해 사전 분포를 정의하는 데 사용되는 b1 및 b2 값 범위를 반영합니다. 둘 다 b1 및 b2 의 더 정확한 추정값을 얻을 수 있습니다.

이항 반응에 대한 설계 생성

일부 적용 분야에서는 성공/실패(이항) 측정에 관심이 있을 수 있습니다. 이 예에서는 단일 설계 요인으로 이항 반응에 대한 비선형 설계를 구성합니다. 비선형 설정의 설계 및 특히 이항 사례에 대한 자세한 내용은 Gotwalt et al. (2009) 의 내용을 기반으로 합니다.

로지스틱 연결 함수와 함께 일반화 선형 모형을 사용하여 이항 반응의 성공 확률을 단일 요인 x 의 함수로 모델링하려고 합니다 .

$$Y(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Y 의 분산은 가중치로 사용됩니다.

이 모형은 미지의 모수 β_0 및 β_1 에서 비선형입니다. 비선형 설계 도구를 사용하여 실험 설계를 계획합니다. 목표는 이항 반응 Y 에 대한 x 의 효과를 모델링하는 것입니다.

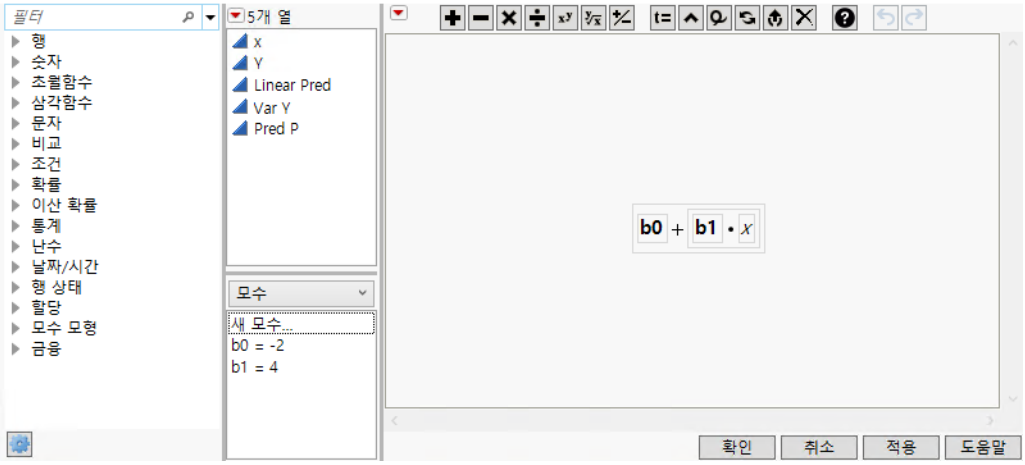
원하는 비선형 설계를 생성하려면 예측 변수 열, 적합시킬 비선형 반응 모형의 계산식 열 및 가중치 계산식 열을 포함하는 데이터 테이블이 있어야 합니다.

설계 구성을 위한 데이터 테이블

Design Experiment 폴더의 One Factor Logistic Design.jmp 데이터 테이블에는 다음 열이 포함되어 있습니다 .

- 예측 변수 x 열 . 이 열에 정의된 " 코딩 " 특성은 하한값과 상한값을 각각 0 과 1 로 설정합니다 .
- 반응 Y 열 . 이 열은 시험에서 수집된 반응 (0/1) 을 포함합니다 .
- 연결 함수의 선형 부분에 대한 계산식을 포함하는 Linear Pred 열 . 계산식을 보려면 " 열 " 패널에서 Linear Pred 오른쪽의 더하기 기호를 클릭합니다 . 모형 계산식에는 **b0** 및 **b1** 모수의 초기 추정값이 포함됩니다 . 계산식을 정의할 때 초기 모수 값이 설정됩니다 . 이러한 값은 계산식 편집기 창의 아래쪽 중간에 있는 계산식 요소 패널에 표시됩니다 .

그림 23.11 모수 추정값을 사용한 Linear Predictor 계산식

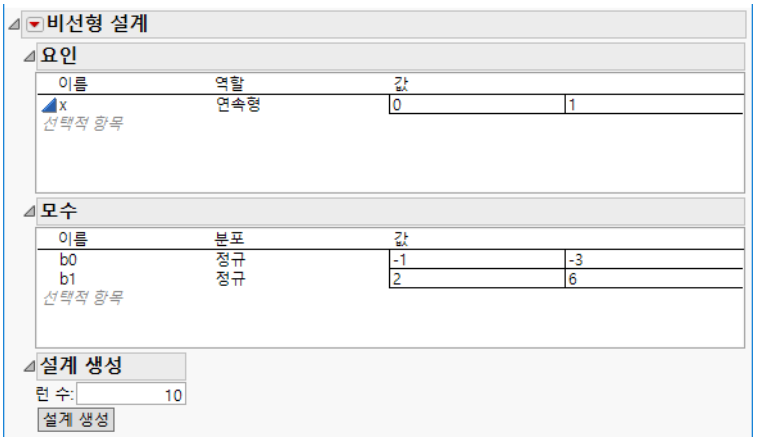


- 가정된 로지스틱 모형을 기반으로 하는 반응의 분산 계산식을 포함하는 Var Y 열 . 이 계산식은 *p(1-p)* 이며 , 여기서 *p* 는 로지스틱 함수입니다 . 이 열은 가중치 열로 사용됩니다 .
- 로지스틱 연결 함수를 포함하는 Pred P 열 .

설계 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/One Factor Logistic Design.jmp 를 엽니다 .
2. **DOE > 특수 용도 > 비선형 설계**를 선택합니다 .
3. Y 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Linear Pred 를 선택하고 **X, 예측 변수 계산식**을 클릭합니다 .
5. Var Y 를 선택하고 **가중치**를 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 23.12 비선형 설계 창

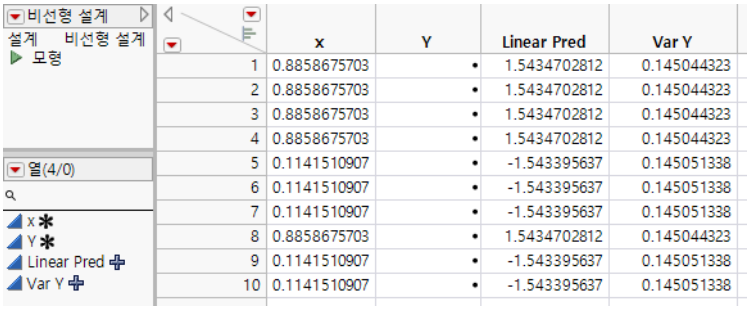


" 요인 " 개요에는 값이 0 과 1 로 지정된 x 요인이 표시됩니다 . " 모수 " 개요에는 각 사전 분포가 정규 분포로 설정된 두 모형 모수가 표시됩니다 . JMP 는 모수 값의 초기 설정을 기반으로 기본값을 계산합니다 . 이 개요에서 모수의 값 범위와 가정된 분포를 변경할 수 있습니다 . 이 예에서는 모든 설정을 그대로 두십시오 . 기본 런 수는 10 입니다 .

7. " 비선형 설계 " 의 빨간색 삼각형을 클릭하고 **고급 옵션 > N 몬테카를로 구**를 선택합니다 . 로컬 최적 설계를 생성하기 위해 구 수를 0 으로 설정합니다 . **확인**을 클릭합니다 .
8. **설계 생성**을 클릭합니다 .
9. **테이블 확대**를 클릭합니다 .

   이렇게 하면 10 회 런 설계가 One Factor Logistic Design.jmp 에 추가됩니다 (그림 23.13 참조). 최적화 알고리즘에 랜덤 성분이 있으므로 설계 테이블이 그림과 다릅니다 .

그림 23.13 One Factor Logistic Design.jmp 확대



	x	Y	Linear Pred	Var Y
1	0.8858675703	•	1.5434702812	0.145044323
2	0.8858675703	•	1.5434702812	0.145044323
3	0.8858675703	•	1.5434702812	0.145044323
4	0.8858675703	•	1.5434702812	0.145044323
5	0.1141510907	•	-1.543395637	0.145051338
6	0.1141510907	•	-1.543395637	0.145051338
7	0.1141510907	•	-1.543395637	0.145051338
8	0.8858675703	•	1.5434702812	0.145044323
9	0.1141510907	•	-1.543395637	0.145051338
10	0.1141510907	•	-1.543395637	0.145051338

데이터 테이블에 저장된 모형은 설계 생성에 사용된 "Linear Pred"를 적합시키기 위한 비선형 모형입니다. 데이터를 수집한 후, Y에 대한 x 효과를 모델링하려면 로짓 연결 함수와 함께 GLM(일반화 선형 모형)을 사용합니다.

비선형 설계 시작 창

비선형 설계 플랫폼을 사용하려면 다음 열이 포함된 기존 데이터 테이블이 있어야 합니다 .

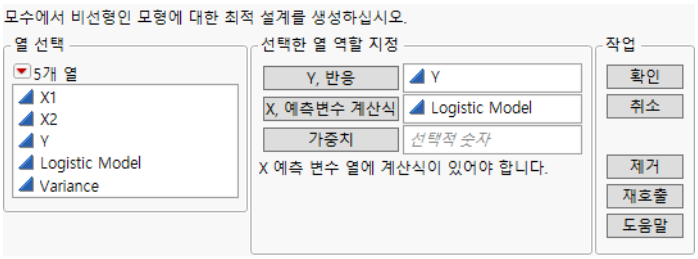
- 반응에 대한 열
- 각 요인에 대한 열
- 요인과 반응 간의 관계를 보여 주는 계산식이 포함된 열 . 이 계산식은 미지의 모수를 포함해야 합니다 .

계산식에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

테이블에 예측 변수 및 반응에 대한 값이 포함될 수 있습니다 . 이러한 값이 포함된 경우 테이블에 내포된 설계를 확대하는 설계를 구성합니다 . 예측 변수에 대해 결측값을 포함하는 행은 없습니다 .

시작 데이터 테이블을 활성화하고 **DOE > 특수 용도 > 비선형 설계**를 선택합니다 .

그림 23.14 비선형 시작 창



Y, 반응 반응 값에 대한 숫자 열입니다 .

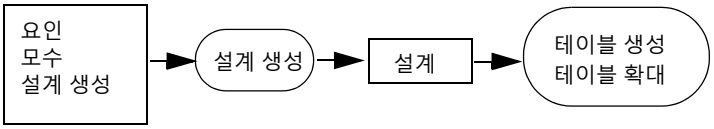
X, 예측변수 계산식 비선형 모형에 대한 계산식을 포함하는 숫자 열입니다 . 이 계산식은 모수를 포함해야 합니다 .

가중치 (선택 사항) 관측값에 가중치를 할당하는 숫자 열입니다 .

비선형 설계 창

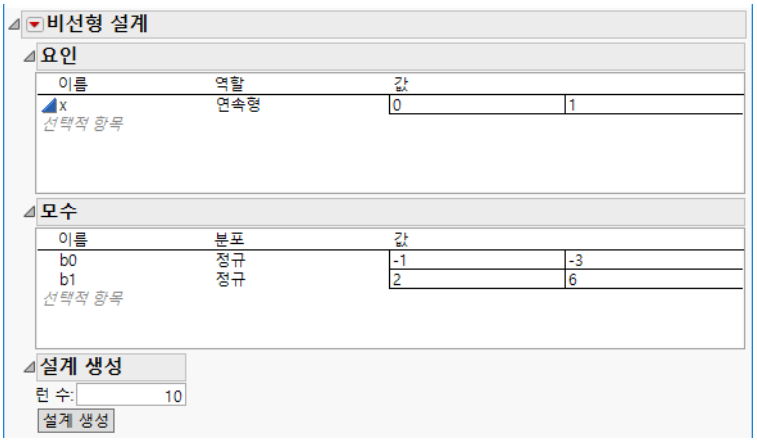
설계 단계를 진행하는 동안 "비선형 설계" 창이 업데이트됩니다. 개요는 개요를 업데이트하는 버튼을 클릭할 때 구분되어 나타나며 그림 23.15의 흐름을 따릅니다.

그림 23.15 비선형 설계 흐름



초기 설계 창에는 "요인", "모수" 및 "설계 생성" 개요가 표시됩니다.

그림 23.16 초기 설계 창



요인

모형 계산식에 사용된 열 이름이 "요인" 개요의 "이름" 열에 자동으로 삽입됩니다. 각 요인의 역할은 "연속형"으로 설정됩니다.

처음에는 값 요인의 값이 -1과 1로 설정됩니다. "코딩" 열 특성을 사용하여 값을 정의한 경우에는 해당 값이 대신 사용됩니다. 요인 개요에서 이 값을 변경할 수 있습니다.

모수

모형 계산식에 사용된 모수 이름이 "모수" 개요의 "이름" 열에 자동으로 삽입됩니다.

처음에는 각 모수의 값이 모수 정의에 지정된 초기값을 중심으로 하여 대칭 간격으로 설정됩니다. 이 간격은 초기값의 거리를 0으로 하고 초기값을 중심으로 이 너비의 간격을 구성하여 얻습니다. 이러한 값은 모형 모수에 대한 사전 분포를 정의하는 데 사용됩니다.

참고 : 모형 모수에 대한 불확도를 반영하기 위해 분포 선택과 함께 모수 값을 조정하십시오 .

"분포" 아래에 네 가지 계열의 사전 분포가 나열됩니다 . 모수에 지정하는 값에 따라 사용되는 사전 분포 계열의 멤버가 결정됩니다 . 하한값을 *low* 로 , 상한값을 *high* 로 나타냅니다 . 그러면 다음과 같이 분포가 결정됩니다 .

- 균등 : 간격 (*low*, *high*) 에서 균등한 분포입니다 .
- 정규 , 로그 정규 , 지수 : *low* 가 0.025 분위수이고 *high* 가 0.975 분위수인 분포입니다 .

설계 생성

JMP 는 다음과 같이 결정된 런 수를 제안합니다 .

- 데이터 테이블을 확대하지 않는 경우 런 수는 모수 수의 4 배에 2 를 더한 값입니다 .
- 데이터 테이블을 확대하는 경우 런 수는 데이터 테이블의 런 수에 모수 수의 2 배를 더한 값입니다 .

참고 : 설계를 확대하는 경우 JMP 에서 제안하거나 사용자가 지정하는 런 수에는 데이터 테이블의 관측값에 해당하는 런이 포함됩니다 . 런 수를 적절하게 조정하십시오 .

설계

"설계 생성"을 클릭하면 설계가 구성되고 "비선형 설계" 창에 "설계" 개요가 추가됩니다. "설계" 개요에서 요인 수준 설정을 검토할 수 있습니다.

테이블 생성 또는 확대

"테이블 생성" 버튼은 새 설계 테이블을 생성합니다. 원래 테이블에 기존 런이 포함된 경우 새 테이블에도 기존 런이 포함됩니다. **테이블 확대** 버튼은 기존 테이블에 새 런을 추가합니다.

비선형 설계 옵션

비선형 설계 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

반응 저장 " 반응 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 . 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다 .

반응 불러오기 사용할 수 없음

요인 저장 " 요인 " 패널의 정보를 새 데이터 테이블에 저장합니다 . 각 요인의 열에는 수준이 포함됩니다 . 다른 정보는 열 특성으로 저장됩니다 . 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다 .

요인 불러오기 사용할 수 없음

제약 조건 저장 사용할 수 없음

제약 조건 불러오기 사용할 수 없음

반응 시뮬레이션 반응 값을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 . 그러면 결과 설계 테이블의 반응 열에 시뮬레이션된 값이 포함됩니다 .

> **참고 :** 반응을 항상 시뮬레이션하도록 환경 설정을 지정하려면 **파일 > 환경 설정 > 플랫폼 > DOE** 를 선택한 후 **반응 시뮬레이션**을 선택하십시오 .

시작 수 구적법을 사용하여 비선형 설계가 생성되는 횟수를 설정합니다. 생성된 설계 중에서 최적 기준을 최대화하는 설계가 선택됩니다 .

고급 옵션 > 몬테카를로 표본 수 최적 기준 계산에 사용되는 구당 팔각면 수를 설정합니다 . 기본값은 팔각면 1 개입니다 . 자세한 내용은 " 최적 기준의 방사형 - 구형 적분 "(651 페이지) 에서 확인하십시오 .

고급 옵션 > N 몬테카를로 구 최적 기준 계산에 사용되는 0 이 아닌 반지름 값의 수를 설정합니다 . 기본값은 2 입니다 . 자세한 내용은 " 최적 기준의 방사형 - 구형 적분 "(651 페이지) 에서 확인하십시오 .

> **참고 :** "N 몬테카를로 구 "(반지름 수) 를 0 으로 설정하면 중앙점만 계산에 사용됩니다 . 이렇게 하면 모수의 초기값에 최적인 로컬 설계가 제공됩니다 . 상황에 따라 이 방법이 적절할 수 있습니다 .

비선형 설계에 대한 통계 상세 정보

- " 비선형 모형 "
- " 최적 기준의 방사형 - 구형 적분 "
- " 최적 설계 찾기 "

비선형 모형

n 개 반응의 벡터를 $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ ‘으로 나타냅니다 . 비선형 모형은 다음 특성에 의해 정의됩니다 .

- Y_i 는 독립적이며 지수 계열 분포와 같은 분포를 따릅니다 .
- 예측 변수 값 $\mathbf{x}_i$ 의 벡터가 제공된 경우 각 Y_i 의 기대값은 모수 θ 의 비선형 함수입니다 . 이 함수를 다음과 같이 나타냅니다 .

  $E(Y_i(\mathbf{x})) = f(\theta, \mathbf{x}_i)$

- 각 Y_i 는 다음과 같이 나타냅니다 .

  $Y_i = f(\theta, \mathbf{x}_i) + \varepsilon_i$

- 오차 벡터 $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ ‘의 평균은 $\mathbf{0}$ 이고 공분산 행렬은 $\sigma^2\mathbf{I}$ 입니다 . 여기서 $\mathbf{I}$ 는 $n \times n$ 단위 행렬입니다 .

모수 θ 와 관련하여 함수 f 의 첫 번째 편도함수 행렬을 X 로 나타냅니다 . 일반적인 조건에서 θ 의 최소 제곱 추정량은 다음과 같이 지정된 점근 공분산 행렬을 사용하여 점근적 비편향입니다 .

$$Cov(\hat{\theta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

이 결과에 대한 증명은 Wu(1981) 및 Jennrich(1969)에서 확인하십시오.

최적 기준의 방사형 - 구형 적분

최적 기준은 사전 분포와 관련한 정보 행렬의 행렬식에 대한 로그 기대값입니다. 따라서 최적의 비선형 설계를 찾으려면 모수 사전 분포에 대한 Fisher 정보 행렬의 행렬식에 대한 로그 적분을 최대화해야 합니다. 이 적분은 수치로 계산되어야 합니다. 비선형 설계 플랫폼에 사용되는 방법은 Gotwalt et al. (2009)의 내용을 기반으로 합니다.

사전 분포가 정규 분포인 경우 적분은 방사형 방향 하나와 모수 수에서 1을 뺀 값과 동일한 여러 각 방향으로 다시 파라미터화됩니다. 적분의 방사형 부분은 반지름 = 0에 대한 결과와 함께 Radau-Gauss-Laguerre 구적을 사용하여 계산됩니다. 이는 특정 수의 초팔각면 (hyperoctahedra)을 구성하여 각각 무작위로 회전하는 방법으로 수행됩니다.

사전 분포가 정규 분포가 아닌 경우에는 새 모수가 정규 분포를 따르도록 적분이 다시 파라미터화됩니다. 그런 다음 방사형-구형 적분 방법이 적용됩니다.

참고: 모수에 대한 사전 분포가 해에 적합하지 않고 프로세스가 실패하면 모수 공간의 영역에서 Fisher 정보 행렬이 특이하다는 메시지가 창에 추가됩니다 . 이 경우에는 사전 분포 또는 모수 범위를 변경해 보십시오 .

최적 설계 찾기

최적 설계를 찾는 데 사용되는 방법은 Meyer and Nachtsheim(1995) 에 설명된 좌표 교환 알고리즘과 유사합니다 . 비선형 최적 설계를 구하는 방법에 대한 자세한 내용은 Gotwalt et al. (2009) 에서 확인하십시오 . 일반적 방법은 다음과 같이 진행됩니다 .

- 비특이 시작 설계를 찾을 때까지 랜덤 설계를 시험합니다 .
- 반복이 수행됩니다 . 각 반복은 모든 런을 통과하도록 구성됩니다 .
- 각 런에 대해 요인이 한 번에 하나씩 최적화됩니다 .
- 목적 함수는 베이지안 D- 최적 기준입니다 . 이 기준은 사전 분포와 관련한 정보 행렬의 행렬식에 대한 로그 기대값입니다 .
- 목적 함수의 변화가 작으면 반복이 종료됩니다 .

24 장

균형 불완비 블록 설계

모든 처리 또는 요인 조합을 모든 블록에 포함할 수 없을 때 균형 불완비 블록 설계(BIBD)를 사용할 수 있습니다. 한 블록에서 k개 처리만 실행할 수 있는 경우, b개 블록에서 a개 처리($k < a$)를 시험하기 위해 BIBD를 사용합니다. 예를 들어 동시에 시험할 수 있는 처리 수에 제한이 있을 수 있습니다. BIBD 설계는 각 처리 및 처리 쌍이 블록에 동일한 횟수로 함께 나타날 수 있게 구성됩니다. 이렇게 하면 **균형** 설계가 됩니다. 모든 블록에 모든 처리를 포함하지 않으면 **불완비** 설계가 됩니다.

균형 불완비 블록 설계 플랫폼을 사용하면 BIBD를 구성할 수 있습니다. 또한 각 처리가 모든 블록에 나타나는 완비 블록 설계도 구성할 수 있습니다.

그림 24.1 크기 3 의 4 개 블록에서 4 개 처리에 대한 BIBD

설계

블록			
1	L2	L3	L4
2	L3	L4	L1
3	L4	L1	L2
4	L1	L2	L3

목차

균형 불완비 블록 설계의 예 655
균형 불완비 블록 설계 생성 657
처리, 블록 지정 및 설계 생성 657
설계 확인 및 테이블 생성 658
균형 불완비 블록 설계 옵션 658
균형 불완비 블록 설계에 대한 통계 상세 정보 659

균형 불완비 블록 설계의 예

이 예에서는 플랜트의 세 가지 새 제품 배합을 시험하고 일별로 블록화하려고 합니다 . 하루에 실행할 수 있는 배합 수는 두 개로 제한됩니다 . 또한 일주일 (작업 주) 안에 실험을 완료해야 합니다 . 최대 5 개의 블록이 있는 설계가 필요합니다 .

1. **DOE > 특수 용도 > 균형 불완비 블록 설계**를 선택합니다 .
2. " 처리 크기 " 를 3 으로 설정합니다 .

   팁 : 처리 크기는 단일 요인의 수준 수 또는 둘 이상 요인의 수준 조합을 나타낼 수 있습니다 .

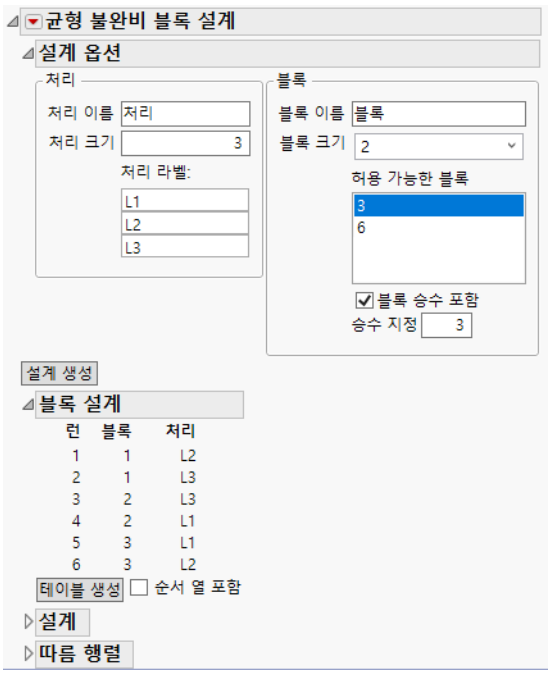
3. **블록 크기**에 2 를 선택합니다 .
4. **허용 가능한 블록**에 3 을 선택합니다 .

   3 개 또는 6 개 블록을 선택할 수 있습니다 . 이 예에서는 각 블록이 하루이므로 5 일 또는 5 개 블록을 초과할 수 없습니다 . 크기가 2 인 블록이 3 개이므로 런 수는 6 입니다 .

참고 : " 처리 이름 ", " 블록 이름 " 및 " 처리 라벨 " 을 입력하여 설계 테이블을 사용자 정의할 수 있습니다 .

5. **설계 생성**을 클릭합니다 .

그림 24.2 크기 2 의 3 개 블록에서 3 개 처리에 대한 BIBD



"블록 설계" 개요에서 볼 수 있듯이 설계의 런 수는 6 이며 각 블록에 처리가 2 개씩 포함된 3 개의 블록으로 구성됩니다. 다른 설계 정보를 더 보려면 "설계" 및 "따름 행렬" 개요를 엽니다. 자세한 내용은 "설계 확인 및 테이블 생성"(658 페이지) 에서 확인하십시오.

원래 블록이 3 개 또는 6 개인 설계를 선택해야 했습니다. 플랜트 시간 제약 조건 때문에 3 개의 블록을 선택했습니다. 그러나 실험을 6 일 동안 실행할 수 있다고 가정해 보겠습니다. 그러면 6 개의 블록을 사용하도록 설계를 업데이트할 수 있습니다.

6. **허용 가능한 블록** 메뉴에서 6 을 선택합니다.
7. **설계 생성**을 클릭하여 설계를 업데이트합니다.
8. **따름 행렬** 개요를 엽니다.

그림 24.3 크기 2 의 6 개 블록에서 3 개 처리에 대한 BIBD 설계 상세 정보

따름 행렬

	처리		
블록	L1	L2	L3
1	1	0	1
2	1	0	1
3	0	1	1
4	0	1	1
5	1	1	0
6	1	1	0

쌍별 처리 빈도

처리	L1	L2	L3
L1	4	2	2
L2		4	2
L3			4

위치 빈도

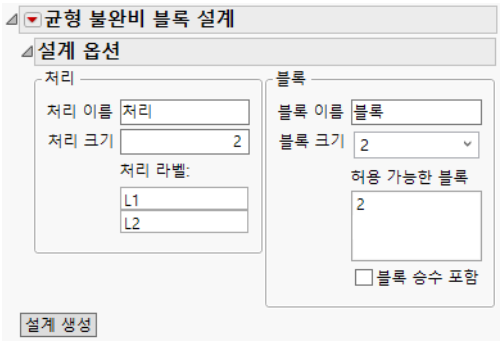
	위치	
처리	1	2
L1	4	0
L2	2	2
L3	0	4

이제 설계에 6개의 블록이 있습니다. 따름 행렬은 각 블록의 처리를 보여 줍니다. "쌍별 처리 빈도"를 보면 설계에서 각 처리가 네 번(주대각 항목) 나오고 각 처리 쌍이 두 번(비대각 항목) 나온다는 것을 알 수 있습니다. 각 블록에는 두 개의 위치가 있습니다. "위치 빈도"를 보면 처리 L1 은 항상 위치 1에 나오고, L2는 두 위치 사이에 나뉘고, L3은 위치 2에 나온다는 것을 알 수 있습니다.

균형 불완비 블록 설계 생성

DOE > 특수 용도 > 균형 불완비 블록 설계를 선택하여 균형 불완비 블록 설계를 생성합니다 .

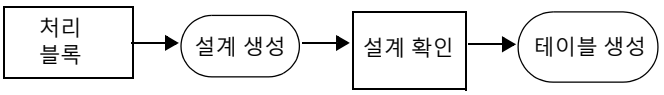
그림 24.4 균형 불완비 블록 설계 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

설계를 진행하는 동안 균형 불완비 블록 설계 창이 업데이트됩니다 . BIBD 를 생성하는 단계는 그림 24.5 의 흐름을 따릅니다 .

그림 24.5 균형 불완비 블록 설계 흐름



처리 , 블록 지정 및 설계 생성

" 설계 옵션 " 개요에는 다음 옵션이 포함되어 있습니다 .

처리 이름 처리 요인의 이름입니다 .

처리 크기 처리 수입니다 .

> **팁 :** 처리는 요인 수준 조합일 수 있습니다 .

처리 라벨 처리에 사용할 라벨입니다 . 기본적으로 L1 ~ L*a* 라벨이 지정되며 , 여기서 *a* 는 총 처리 수입니다 .

블록 이름 블록 요인의 이름입니다 .

블록 크기 한 블록에서 실행할 수 있는 처리 수입니다 .

참고 : 블록 크기 값은 처리 수에 따라 결정됩니다 . 목록에 값이 없으면 BIBD 플랫폼에서 지정된 처리 수에 대해 해당 블록 크기의 설계를 생성할 수 없기 때문입니다 .

허용 가능한 블록 처리 수와 블록 크기에 따라 설계에 허용되는 블록 수입니다 .

블록 승수 포함 허용 가능한 초기 블록 크기의 배수 (사용자 지정) 와 동일한 수의 블록을 포함하는 설계를 선택할 수 있습니다 . 추가 (배수) 블록은 원래 설계의 순환 순열입니다 .

승수 지정 (" 블록 승수 포함 " 을 선택한 경우에만 사용 가능) 허용 가능한 초기 블록 크기에 적용할 승수입니다 .

설계 생성 설계를 생성합니다 .

설계 확인 및 테이블 생성

블록 설계 BIBD 를 기존 설계 레이아웃으로 표시합니다 .

테이블 생성 BIBD 데이터 테이블을 생성합니다 .

순서 열 포함 각 블록 내의 런 순서에 대한 열을 설계 테이블에 포함합니다 .

설계 BIBD 를 각 행이 블록 및 해당 블록의 처리를 나타내는 블록 레이아웃으로 표시합니다 .

따름 행렬 각 블록에 나오는 처리를 보여 줍니다 . 행렬의 각 행은 설계의 블록에 해당하며 해당 블록에 처리가 있으면 1, 없으면 0 을 포함합니다 .

쌍별 처리 빈도 각 처리가 다른 모든 처리와 함께 같은 블록에 나오는 횟수를 비대각 항목에 표시합니다 . 주대각 항목은 각 처리의 총 발생 횟수입니다 .

위치 빈도 각 처리가 각 블록 위치에 나오는 횟수를 표시합니다 .

균형 불완비 블록 설계 옵션

균형 불완비 블록 설계 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다 . " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오 .

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다 .

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다 .
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다 . 계산식과 값은 설계 창에 지정된 모형에 기초합니다 .
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다 .

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다 .

– **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다. 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다 .

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다 . " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다 . 여기서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 **기본 분석**의 " 시뮬레이션 " 장에서 확인하십시오 .

스크립트를 스크립트 창에 저장 " 균형 불완비 블록 설계 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

균형 불완비 블록 설계에 대한 통계 상세 정보

BIBD 의 총 관측값 수는 다음 방정식에 의해 결정됩니다 .

$$N = ar = bk$$

다음은 각 요소에 대한 설명입니다 .

a 는 처리 수입니다 .
r 는 설계에서 각 처리가 나타나는 횟수입니다 .
b 는 블록 수입니다 .
k 는 각 블록의 처리 수입니다 .

각 처리 쌍이 나타나는 블록 수는 다음과 같이 구합니다 .

$$\lambda = \frac{r(k-1)}{a-1}$$

여기서 λ는 정수여야 합니다.

이러한 조건은 BIBD가 존재하는 데 필요하지만 충분하지는 않습니다. BIBD 플랫폼은 BIBD를 생성할 수 있는 값을 제공합니다.

"블록 승수 포함"을 선택하면 원래 BIBD가 지정된 횟수만큼 복사됩니다. 추가(배수) 블록은 원래 설계의 순환 순열입니다. 복사된 설계에서 처리 위치는 원래 설계와 다릅니다.

BIBD에 대한 자세한 내용은 SAS Institute Inc. (2013)에서 확인하십시오.

25 장

MSA 설계

측정 시스템 분석 실험에 대한 설계 생성

Gage R&R 연구라고도 하는 측정 시스템 분석 실험은 완전 요인 설계를 기반으로 합니다. MSA 플랫폼은 MSA(측정 시스템 분석) 실험에 대한 완전 요인 설계를 생성하고 설계 진단 측도를 제공합니다. MSA 실험 분석에 대한 내용은 Quality and Process Methods의 에서 확인하십시오.

그림 25.1 MSA 요인 설정

이름	MSA 역할	수준 수	분산	랜덤화
연산자	연산자	5	1	예
부품	부품	3	1	예
게이지	게이지	3	1	예

목차

MSA 설계 개요 663
MSA 설계의 예 663
MSA 설계 생성 665
반응 666
요인 668
MSA 설계 668
MSA 설계 진단 668
MSA 설계 테이블 669
MSA 설계 옵션 670
MSA 설계 플랫폼의 통계 상세 정보 671

MSA 설계 개요

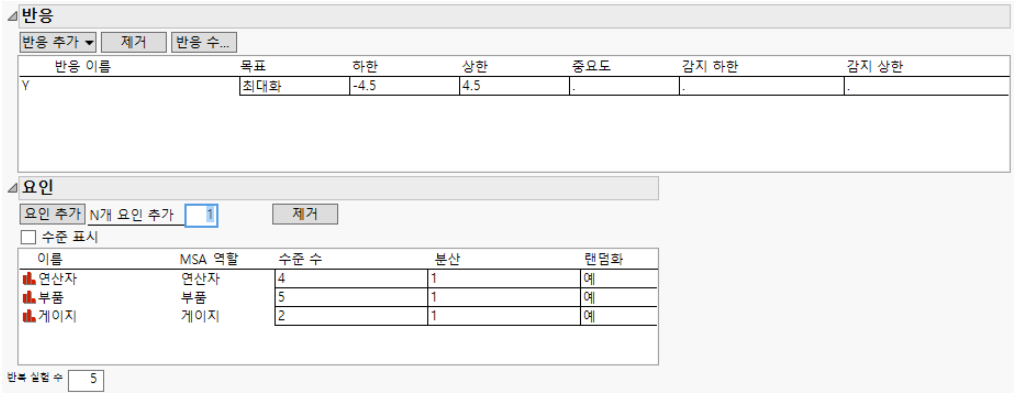
MSA(측정 시스템 분석) 설계는 요인이 임의 효과인 완전 요인 설계입니다. 이 연구의 초점은 반응에서 각 요인으로 인해 발생한 변동을 추정하는 것입니다. MSA 설계 플랫폼은 MSA 연구와 관련된 진단 측도가 포함된 완전 요인 설계를 생성합니다.

MSA 설계의 예

이 예에서는 기기와 작업자로 인해 발생하는 측정 시스템의 변동을 연구하려고 합니다. 연구실에 2 개의 저울이 있고 4 명의 작업자가 이 연구에 참여합니다. 표준 부품 5 개로 이루어진 한 세트의 무게를 측정하려고 하며, 각 작업자는 각 기기에서 각 부품을 6 번씩 측정합니다. 이전의 변동 추정값이 없으므로 분산 값 1 을 기반으로 실험을 설계합니다. 이 분산 값은 진단에서 실제 변동으로 척도화됩니다.

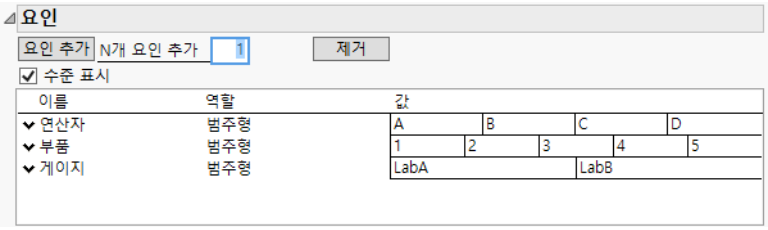
1. **DOE > 특수 용도 > MSA 설계**를 선택합니다.
2. 반응 Y 에 대한 목표를 **목표값 일치로 설정합니다**.
3. 작업자의 **수준 수**를 4 로 설정합니다.
4. 부품의 **수준 수**를 5 로 설정합니다.
5. 게이지의 **수준 수**를 2 로 설정합니다.
6. **반복 실험 수**를 5 로 설정하여 각 설계 런에 6 개의 측정값을 포함시킵니다.

그림 25.2 요인 설정



7. **수준 표시**를 선택합니다.
8. 그림 25.3 에 표시된 대로 요인 값을 편집합니다.

그림 25.3 요인 값



참고 : 9 단계에서 " 난수 시드값 " 을 설정하면 이 예에 표시된 것과 같은 설계와 진단이 재현됩니다 .

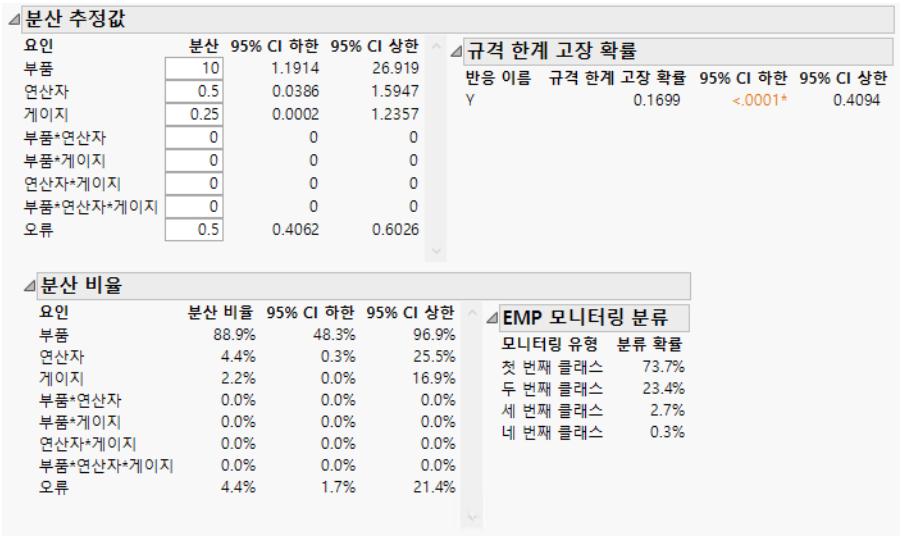
9. (선택 사항) MSA 의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택한 다음 123 을 입력합니다 .
10. **설계 생성**을 클릭합니다 .

    참고 : 테이블 생성을 클릭하여 데이터를 수집하기 위한 테이블을 생성합니다 . 설계 테이블에 포함된 분석 스크립트에 대한 내용은 "MSA 설계 테이블"(669 페이지)에서 확인하십시오.

11. "설계 진단" 개요가 나타날 때까지 아래로 스크롤한 후 회색 표시 아이콘을 클릭하여 엽니다.

    " 설계 진단 " 보고서에는 가정된 분산에 대한 설정이 포함되며 , 이 설정은 신뢰 구간 , 분산 비율 및 EMP 모니터링 분류를 추정하는 데 사용됩니다 . 이러한 추정값은 MSA 연구 설계의 강도를 평가하는 데 사용됩니다 . 다른 분산 가정을 탐색하거나 **뒤로** 버튼을 사용하여 설계를 조정합니다 .
12. 요인의 기대 분산을 조정하여 설계 진단에 미치는 영향을 탐색합니다 . 작업자 분산을 0.5, 부품 분산을 10 그리고 게이지 분산을 0.25 로 설정합니다 .

그림 25.4 MSA 설계 진단

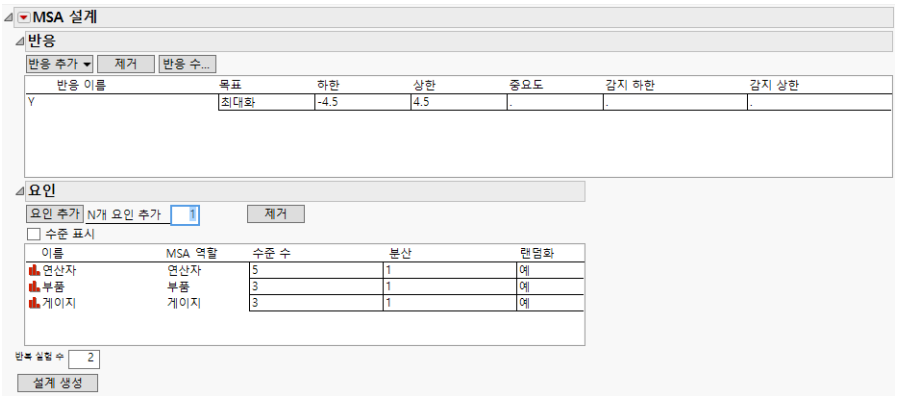


모든 진단은 적합한 평균 제곱의 표집 분산을 사용하여 계산된 시뮬레이션을 기반으로 합니다. 분산 추정 가정을 기반으로 하는 이 연구는 첫 번째 클래스에서 진행한 결과 78%의 확률이 나왔습니다.

MSA 설계 생성

DOE > 특수 용도 > MSA 설계를 선택하여 MSA 설계를 생성합니다.

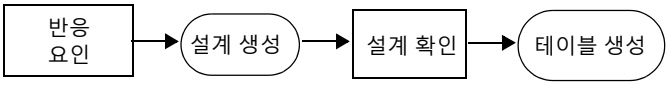
그림 25.5 MSA 설계 창



"열 선택"의 빨간색 삼각형에 있는 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

MSA 설계 창은 설계 단계를 진행하는 동안 업데이트됩니다. MSA 설계를 생성하는 단계는 그림 25.6의 흐름을 따릅니다.

그림 25.6 MSA 설계 흐름

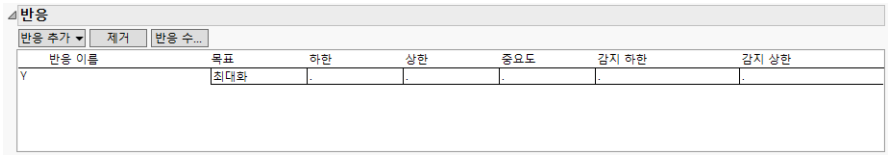


반응

"반응" 개요를 사용하여 하나 이상의 반응을 지정합니다.

팁 : " 반응 " 개요를 완료하면 빨간색 삼각형 메뉴에서 **반응 저장**을 선택하는 것이 좋습니다 . 이 옵션을 사용하면 반응 이름 , 목표 , 한계 및 중요도 값을 데이터 테이블에 저장하여 나중에 DOE 플랫폼에서 다시 불러올 수 있습니다 .

그림 25.7 반응 개요



반응 추가 " 최대화 ", " 목표값 일치 ", " 최소화 ", " 없음 " 또는 " 함수 " 목표 유형을 사용하여 단일 반응을 입력합니다 . " 목표값 일치 " 를 선택하는 경우 목표값 한계를 입력해야 합니다 . " 최대화 " 또는 " 최소화 " 를 선택하는 경우에는 한계를 입력할 필요가 없지만 만족도 함수를 사용하려면 입력하는 것이 유용할 수 있습니다 .

JMP PRO 함수 (JMP Pro 에서만 사용할 수 있음) 설계 데이터 테이블에 FDE 열 특성을 가진 여러 개의 Y 열을 추가합니다 . 이러한 열은 DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에 사용할 수 있습니다 .

제거 선택한 반응을 제거합니다 .

반응 수 입력한 수가 총 반응 수가 되도록 추가 반응을 입력합니다 . 기본값 Y 이외의 반응을 입력한 경우 각 추가 반응의 목표는 마지막에 입력한 반응과 관련된 목표입니다 . 그렇지 않은 경우에는 기본적으로 " 목표값 일치 " 로 설정됩니다 . 테이블에서 " 목표 " 유형을 클릭하여 목표를 변경할 수 있습니다 .

" 반응 " 개요에는 다음과 같은 열이 포함되어 있습니다 .

반응 이름 반응의 이름입니다 . 반응을 추가하면 Y, Y2 등의 기본 이름이 지정됩니다 . 이 이름을 변경하려면 이름을 두 번 클릭한 후 원하는 이름을 입력합니다 .

목표 , 하한 , 상한 " 목표 " 는 반응 최대화 , 반응 최소화 , 목표값 일치 또는 반응 목표 없음 중에서 무엇을 원하는지 JMP 에 알려 줍니다 . JMP 는 이러한 지정에 따라 설계 테이블의 각 반응

열에 " 반응 한계 " 열 특성을 할당합니다 . 이 정보를 사용하여 각 반응에 대한 만족도 함수를 정의합니다 . " 프로파일러 " 및 " 등고선 프로파일러 " 는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다 . 자세한 내용은 Profilers 의 "Profiler" 장 및 " 열 특성 " 부록의 " 반응 한계 "(730 페이지) 에서 확인하십시오 .

- " 최대화 " 목표는 가장 큰 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 최소화 " 목표는 가장 작은 값이 가장 좋음을 나타냅니다 . 자연 하한 또는 상한이 있는 경우 이를 " 하한 " 또는 " 상한 " 으로 지정할 수 있습니다 .
- " 목표값 일치 " 목표는 특정 목표값이 가장 좋음을 나타냅니다 . 기본 목표값은 하한과 상한의 중간으로 가정됩니다 .
- " 없음 " 목표는 최적화에 관한 목표가 없음을 나타냅니다 . 만족도 함수가 구성되지 않습니다 .

참고 : 목표 반응이 하한과 상한의 중간이 아닌 경우에는 설계 테이블을 생성한 후 목표값을 변경할 수 있습니다 . 데이터 테이블에서 반응 열의 " 열 정보 " 창을 열고 (**열 > 열 정보**) 원하는 목표값을 입력하십시오 .

중요도 반응이 여러 개인 경우 지정한 중요도 값을 사용하여 전체 만족도 함수를 계산합니다 . 이러한 값은 반응에 대한 가중치로 처리됩니다 . 반응이 하나뿐이면 기본적으로 1 로 설정되므로 중요도를 지정할 필요가 없습니다 .

JMP PRO **감지 한계** 하한 및 상한 감지 한계는 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다 . 감지 한계는 열 특성으로 저장됩니다 . 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 일반화 회귀 장에서 확인하십시오 .

반응 개요 편집

" 반응 " 개요에서 다음 사항에 유의하십시오 .

- 반응 이름을 편집하려면 해당 반응을 두 번 클릭합니다 .
- 목표를 클릭하여 변경할 수 있습니다 .
- 한계 또는 중요도 가중치를 클릭하여 변경할 수 있습니다 .
- 반응이 여러 개인 경우 중요도 가중치 값을 입력할 수 있습니다 .

반응 한계 열 특성

반응을 입력할 때 지정하는 "목표", "하한", "상한" 및 "중요도"를 사용하여 최적 요인 설정을 찾습니다. 각 반응에 대한 정보는 생성된 설계 데이터 테이블에 "반응 한계" 열 특성으로 저장됩니다. JMP에서는 이 정보를 사용하여 만족도 함수를 정의합니다. 만족도 함수는 최적 요인 설정을 찾기 위해 "예측 프로파일러"에서 사용됩니다. "반응 한계" 열 특성에 대한 자세한 내용과 사용 예는 "열 특성" 부록의 "반응 한계"(730 페이지)에서 확인하십시오.

하한과 상한을 지정하지 않으면 JMP는 반응에 대해 관측된 데이터 범위를 사용하여 만족도 함수의 한계를 정의합니다. 하한과 상한을 지정하면 만족도 함수의 규격을 제어할 수 있습니다. 만족도 함수 구성에 대한 자세한 내용은 Profilers의 "Profiler" 장에서 확인하십시오.

요인

"요인" 개요는 기본적으로 5명의 작업자, 3개의 부품 및 3개의 게이지 연구로 설정됩니다.

참고 : MSA 설계 프로그램에는 부품 역할을 하는 최소 1 개의 요인이 필요합니다 .

요인 추가 하나 이상의 요인을 추가하려면 클릭합니다 .

제거 선택한 요인을 제거합니다 .

N 개 요인 추가 여러 요인을 추가합니다 . 추가할 요인 수를 입력하고 **요인 추가**를 클릭합니다 .

수준 표시 요인 수준을 표시하려면 선택합니다 . 이 옵션을 선택하면 다음 옵션을 사용할 수 있습니다 .

이름 요인 이름이며 , 수준을 추가할 수 있는 메뉴가 있습니다 .

역할 범주형 요인에 고정됩니다 . MSA 요인은 기본적으로 범주형입니다 .

값 요인 수준을 지정하며 , 클릭하여 편집할 수 있습니다 .

이름 요인 이름을 지정하며 , 클릭하여 편집할 수 있습니다 . 처음에는 3 개의 요인 즉 , 작업자 , 부품 및 게이지가 목록에 나타납니다 . 이 목록에서 요인을 추가하거나 제거할 수 있습니다 .

MSA 역할 요인의 MSA 역할을 지정합니다 . 클릭하여 " 작업자 ","부품 ","게이지 " 또는 " 없음 " 을 선택할 수 있습니다 .

수준 수 요인의 수준 수를 지정합니다 .

분산 요인으로 인한 사전 확률을 지정합니다 . 기본값은 1 이며 , 이 값은 진단에서 실제 변동으로 척도화됩니다 .

랜덤화 설계 테이블에서 요인을 랜덤화하도록 지정합니다 .

반복 실험 수 설계에 포함할 0~10 의 반복 실험 횟수를 지정합니다 . 반복 실험 수는 초기 측정 이외에 추가되는 측정 횟수입니다 . 기본값은 2 번 반복이므로 결국 각 런에 대해 3 번 측정하게 됩니다 . 런당 한 번만 측정하도록 지정하려면 반복 횟수를 0 으로 설정합니다 .

MSA 설계

" 설계 " 개요는 MSA 설계에 대한 런을 보여 줍니다 .

MSA 설계 진단

"설계 진단" 보고서의 모든 진단은 적절한 평균 제곱의 표집 분산을 사용하여 계산된 시뮬레이션을 기반으로 합니다. 각 모형 요소의 분산 추정값을 변경할 수 있습니다. 분산 추정값이 변경

되면 설계 진단이 업데이트됩니다. 따라서 해당 설계가 다른 가정된 모형 하에서 시스템을 어떻게 측정하는지 그 성능을 탐색할 수 있습니다.

MSA 설계 분산 추정값

분산 추정값 보고서를 사용하면 각 MSA 모형 요인의 분산 추정값을 지정할 수 있습니다. 진단은 분산 추정값을 기반으로 하며 시뮬레이션된 CI(신뢰 구간) 경계를 포함합니다. 또한 시뮬레이션된 추정값 및 CI 경계는 규격을 벗어난 반응의 확률에도 제공됩니다. 이러한 규격은 반응에 대해 설정된 하한 및 상한에 의해 결정됩니다.

MSA 설계 분산 비율

"분산 비율" 보고서에는 추정된 분산을 기반으로 하는 추정 분산 비율에 대한 진단이 들어 있습니다. 이 보고서에는 또한 시뮬레이션된 근사 CI 경계와 측정 시스템 EMP 분류 확률이 포함되어 있습니다. EMP 분류에 대한 자세한 내용은 Quality and Process Methods의 에서 확인하십시오.

MSA 설계 시뮬레이션 결과 및 옵션

"시뮬레이션" 보고서에는 분산 비율 CI 경계를 추정하는 데 사용되는 시뮬레이션 분산의 분포에 대한 상자 그림이 포함되어 있습니다. 시뮬레이션은 다수의 시도 횟수를 기반으로 합니다.

"시뮬레이션 결과"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

통계량 표시 각 모형 요인에 대한 요약 통계를 엽니다.

점 표시 상자 그림에서 수염을 벗어난 점을 표시하거나 숨깁니다.

데이터 테이블 생성 시뮬레이션 결과에 대한 데이터 테이블을 엽니다.

MSA 설계 테이블

다음 버튼을 사용하여 MSA 설계 생성을 완료합니다.

테이블 생성 MSA 설계 데이터 테이블을 생성합니다. 테이블에는 다음과 같은 분석용 스크립트가 포함되어 있습니다.

EMP 측정 시스템 분석 MSA 분석 플랫폼을 시작합니다. 자세한 내용은 Quality and Process Methods 에서 확인하십시오.

계량형 차트 계량형 게이지 차트 플랫폼을 시작합니다. 자세한 내용은 Quality and Process Methods 에서 확인하십시오.

DOE 대화상자 현재 설계 테이블을 생성하는 데 사용된 규격을 사용하여 MSA 설계 플랫폼을 시작합니다.

연산자 워크시트 데이터 수집용 워크시트로 사용할 각 작업자의 데이터 테이블을 엽니다.

뒤로 설계 규격을 조정할 수 있도록 뒤로 이동합니다.

MSA 설계 옵션

"MSA 설계"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

반응 저장 "반응" 패널의 정보를 새 데이터 테이블에 저장합니다. 그러면 반응 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다. 이 옵션은 반응을 재사용할 것으로 예상하는 경우 유용합니다.

반응 불러오기 "반응 저장" 옵션을 사용하여 저장한 반응을 불러옵니다.

요인 저장 "요인" 패널의 정보를 새 데이터 테이블에 저장합니다. 각 요인의 열에는 수준이 포함됩니다. 다른 정보는 열 특성으로 저장됩니다. 그러면 요인 및 관련 정보를 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

참고 : 빈 테이블에 데이터를 입력하여 요인 테이블을 생성할 수 있지만 각 열에 적절한 설계 역할을 할당해야 합니다. 이렇게 하려면 데이터 격자에서 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 설계 역할**을 선택합니다. 그런 다음 "설계 역할" 영역에서 적절한 역할을 선택합니다.

요인 불러오기 "요인 저장" 옵션을 사용하여 저장한 요인을 불러옵니다.

제약 조건 저장 (일부 플랫폼에서는 사용할 수 없음) "요인 제약 조건 정의" 또는 "선형 제약 조건" 개요에서 정의한 요인 제약 조건을 각 제약 조건에 대한 열과 함께 데이터 테이블에 저장합니다. 그러면 제약 조건을 대부분의 DOE 창에 빠르게 불러올 수 있습니다.

제약 조건 테이블에서 처음에 나오는 행은 각 요인의 계수를 포함합니다. 마지막 행은 부등식 한계를 포함합니다. 각 제약 조건 열에는 제약 조건을 "less than" 또는 "greater than" 제약 조건으로 식별하는 **제약 조건 상태** 열 특성이 포함됩니다. 자세한 내용은 "열 특성" 부록의 "제약 조건 상태"(762 페이지)에서 확인하십시오.

제약 조건 불러오기 (일부 플랫폼에서는 사용할 수 없음) "제약 조건 저장" 옵션을 사용하여 저장한 요인 제약 조건을 불러옵니다.

난수 시드값 설정 JMP 가 랜덤 성분이 있는 작업을 제어하는 데 사용하는 난수 시드값을 설정합니다. 이러한 작업에는 다음 중 하나 이상이 포함됩니다.

- 설계 생성을 위한 검색 알고리즘 초기화
- 설계 생성을 위한 런 순서 랜덤화
- 랜덤 시작에 기반한 설계의 시작 설계 선택

설계를 재현하려면 **설계 생성**을 클릭하기 전에 설계를 생성한 난수 시드값을 입력합니다.

참고 : 설계와 관련된 난수 시드값은 설계 데이터 테이블에 저장된 "DOE 대화상자" 스크립트에 포함됩니다.

MSA 설계 플랫폼의 통계 상세 정보

분산 및 분산 비율 시뮬레이션은 시뮬레이션된 평균 제곱 항을 사용하여 계산됩니다. 평균 제곱 항은 연관된 기대 평균 제곱으로 척도화된 카이제곱 분포로부터 시뮬레이션됩니다. 시뮬레이션된 분산은 적률 추정량 방법을 사용하여 계산되며 그런 다음 분산 비율의 신뢰 경계 계산에 사용됩니다.

부품의 분산을 예로 들어 보겠습니다. 시뮬레이션 계산은 다음 단계를 따릅니다.

1. 부품의 평균 제곱 즉, 부품과 관련된 각 이원 교호 작용에 대한 평균 제곱과 부품과 관련된 차수가 높은 교호 작용의 평균 제곱을 시뮬레이션합니다 .
2. 포함 및 제외를 사용하여 분산을 계산합니다 .

   부품의 분산 = 부품의 평균 제곱 – 이원 교호 작용의 평균 제곱 + 삼원 교호 작용의 평균 제곱 -…+/- 차수가 높은 교호 작용의 평균 제곱
3. 그런 다음 부품의 시뮬레이션된 분산 비율이 부품의 분산 / 총 분산으로 계산됩니다 .

참고 : 평균 제곱 항의 분산만 직접적으로 알 수 있습니다 . 즉 , 평균 제곱 = 카이제곱 분산 (df)* 기대 평균 제곱입니다 . 분산 및 분산 비율의 분포는 시뮬레이션된 값에 따라 근사값이 계산됩니다 . 이런 이유 때문에 분산 항이 업데이트되거나 플랫폼이 실행될 때마다 신뢰 경계가 매번 달라질 수 있습니다 .

26 장

그룹 직교 과포화 설계

그룹 직교 과포화 설계 생성

그룹 직교 과포화 설계는 선별 설계의 특수한 경우입니다. 이 설계는 조사할 요인의 수가 실현 가능한 런 수보다 많을 때 초기 단계 작업에 적절합니다. 과포화 설계는 모형 모수보다 관측값이 적은 설계입니다. 즉, 설계가 모든 주효과를 동시에 추정할 수 없습니다. 이로 인해 유의한 효과를 감지하기가 어렵습니다. 그룹 직교 과포화 설계는 모형 선택에 적합한 특성을 가진 2수준 과포화 설계의 특수한 종류입니다.

목차

그룹 직교 과포화 설계 개요 675
그룹 직교 과포화 설계의 예 675
그룹 직교 과포화 설계 생성 676
요인, 구조 지정 및 설계 생성 677
설계 확인 및 테이블 생성 678
그룹 직교 과포화 설계 옵션 678
그룹 직교 과포화 설계에 대한 통계 상세 정보 679

그룹 직교 과포화 설계 개요

GOSSD(그룹 직교 과포화 설계)는 모형을 쉽게 선택할 수 있는 특정 구조를 갖도록 구성됩니다. 설계 요인은 그룹으로 분할되며, 같은 그룹의 다른 요인과는 상관되지만 다른 그룹의 요인과는 직교하는 특징을 갖습니다. 한 그룹은 세 개 이상의 "가짜" 요인으로 구성됩니다. 이러한 가짜 요인은 설계 요인에 할당되지 않습니다. 대신 설계 분석에만 사용됩니다. 가짜 요인은 2차 이상 효과를 무시할 수 있다는 가정하에 분산의 비편향 추정값을 제공합니다.

실제로 GOSSD를 설정하려면 다음과 같은 내용을 고려해야 합니다.

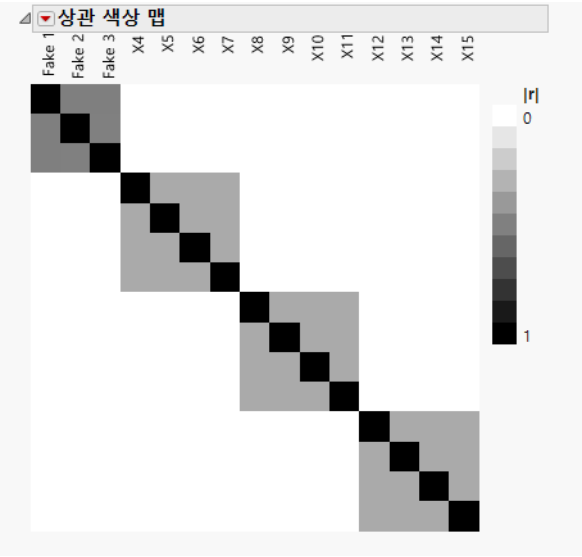
- 그룹은 많을수록 좋습니다.
- 활성으로 간주되는 요인은 별도의 그룹에 두어야 합니다.
- 예상 계수가 양수가 되도록 요인 수준을 선택합니다.

그룹 직교 과포화 설계의 예

이 예에서는 최대 12회 런으로 12개의 요인을 연구하는 실험을 설계하려고 합니다. 설계에 사용할 모수의 수는 최소 16개입니다. 절편에 대한 모수 하나가 필요하고 세 개 이상의 가짜 요인이 있어야 합니다.

1. **DOE > 특수 용도 > 그룹 직교 과포화 설계 > 그룹 직교 과포화 설계**를 선택합니다.
2. **런 수**를 선택합니다.
3. **런 수**에 12 를 입력하고 텍스트 상자 바깥쪽을 클릭합니다.
4. " 구조 " 창에서 그룹 수 4, 그룹 크기 4, 모수 수가 16 인 두 번째 옵션을 선택합니다.
5. X1 요인 이름을 클릭하고 이름을 Fake 1 로 변경합니다.
6. X2 와 X3 에 대해 이 과정을 반복하여 각각 Fake 2 와 Fake 3 으로 이름을 지정합니다.
7. **설계 생성**을 클릭합니다.

그림 26.1 GOSSD 상관 색상 맵

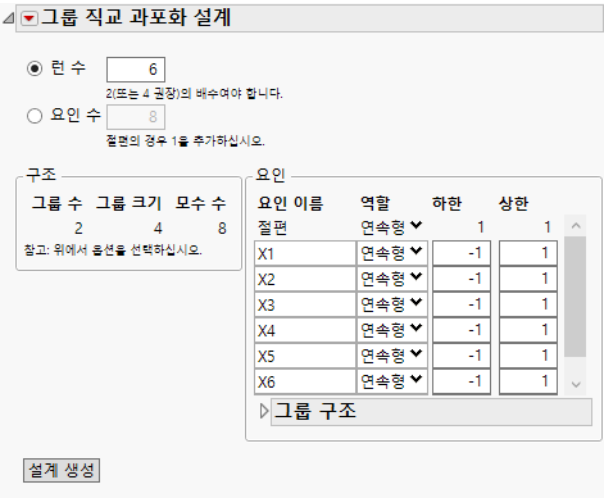


설계 구조가 색상 맵에 분명히 나타나 있습니다. 그룹 내의 요인은 상관되어 있습니다. 서로 다른 그룹의 요인은 상관관계가 없으며 직교 상태임을 나타냅니다.

그룹 직교 과포화 설계 생성

DOE > 특수 용도 > 그룹 직교 과포화 설계 > 그룹 직교 과포화 설계를 선택하여 그룹 직교 과포화 설계를 생성합니다.

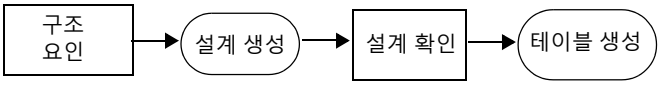
그림 26.2 그룹 직교 과포화 설계 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

설계를 진행하는 동안 GOSSD 창이 업데이트됩니다. GOSSD를 생성하는 단계는 그림 26.3의 흐름을 따릅니다.

그림 26.3 그룹 직교 과포화 설계 흐름



요인 , 구조 지정 및 설계 생성

" 설계 옵션 " 개요에는 다음 옵션이 포함되어 있습니다 .

팁 : 가능한 설계 구조를 정의하기 위한 런 수 또는 요인 수를 제공하십시오 .

런 수 설계 런 수를 정의합니다 . 이 수는 2 의 배수여야 하며 , 권장되는 수는 4 의 배수입니다 .

요인 수 요인 수를 정의합니다 . 절편에 대한 요인과 3 ~ 5 개의 가짜 요인을 포함합니다 . 요인 수는 8 의 배수여야 합니다 .

구조 설계에 사용 가능한 구조 목록을 제공합니다 . 이 목록의 열은 런 수를 설정했는지 아니면 요인 수를 설정했는지에 따라 다릅니다 .

그룹 수 설계의 그룹 수입니다 . 최소값은 2 입니다 .

그룹 크기 각 그룹의 요인 수입니다 .

런 수 (요인 수를 설정한 경우 사용 가능) 설계의 런 수입니다 .

모수 수 (런 수를 설정한 경우 사용 가능) 설계의 모수 수입니다 . 모수에는 절편 하나와 설계 요인이 포함되며 나머지 모수는 가짜 요인에 할당됩니다 .

요인

요인 이름 절편 이름은 고정되어 있습니다 . 나머지 요인은 이름을 바꿀 수 있습니다 .

역할 요인을 연속형 또는 범주형으로 정의할 수 있습니다 .

하한 각 요인의 하한을 설정할 수 있습니다 .

상한 각 요인의 상한을 설정할 수 있습니다 .

그룹 구조 각 그룹에 나타나는 요인을 표시합니다 .

바꾸기 그룹 간에 요인을 바꿀 수 있습니다 .

설계 생성 설계를 생성합니다 .

설계 확인 및 테이블 생성

설계 설계 행렬을 표시합니다.

테이블 생성 GOSSD 데이터 테이블을 생성합니다. 설계 테이블에는 다음과 같은 스크립트가 포함되어 있습니다.

모형 (" 응답 " 열이 채워진 경우에만 실행됨) **DOE > 특수 용도 > 그룹 직교 과포화 설계 > 그룹 직교 과포화 설계 적합** 플랫폼을 실행합니다.

DOE 대화상자 설계 테이블을 생성하는 데 사용된 "GOSSD 설계 " 창을 다시 생성합니다.

GOSSD 구조 설계 구조를 표시합니다 (런 수 , 모수 수 , 그룹 수 , 그룹 크기).

다변량 설계 요인에 대해 **분석 > 다변량** 플랫폼을 실행합니다.

모형 행렬 설계 행렬을 포함합니다.

DOE 시뮬레이션 (반응을 시뮬레이션하는 경우 사용 가능) 시뮬레이션 창을 시작합니다.

설계 평가

상관 행렬 효과 사이의 상관 행렬을 표시합니다.

상관 색상 맵 강도 척도를 사용하여 그림에서 효과 사이의 절대 상관을 표시합니다 . 빨간색 삼각형 메뉴에는 상관계수를 데이터 테이블에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹 직교 과포화 설계 옵션

"그룹 직교 과포화 설계" 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

반응 시뮬레이션 시뮬레이션 계산식을 포함하는 열과 반응 값을 설계 테이블에 추가합니다. " 테이블 생성 " 을 클릭하기 전에 이 옵션을 선택하십시오.

" 테이블 생성 " 을 클릭하면 다음과 같은 결과가 나타납니다.

- 시뮬레이션된 반응 값 집합이 각 반응 열에 추가됩니다.
- 각 반응에 대해 시뮬레이션 모형 계산식을 포함하는 새 열이 설계 테이블에 추가됩니다. 계산식과 값은 설계 창에 지정된 모형에 기초합니다.
- 모형 효과에 대한 계수 값을 설정하고 정규 , 이항 또는 Poisson 분포 중 하나를 지정할 수 있는 " 모형 " 창이 나타납니다.

참고 : 모든 설계 유형에 모든 분포를 사용할 수 있는 것은 아닙니다.

- **DOE 시뮬레이션** 스크립트가 설계 테이블에 저장됩니다. 이 스크립트는 값을 다시 시뮬레이션하거나 시뮬레이션된 반응 분포를 변경할 수 있도록 " 모형 " 창을 다시 엽니다.

" 모형 " 창에서 시뮬레이션된 반응 값의 분포를 제어하는 옵션을 선택합니다. " 적용 " 을 클릭하면 시뮬레이션된 반응 값의 계산식이 <Y> 시뮬레이션이라는 새 열에 저장됩니다. 여기

서 Y 는 반응 이름입니다 . " 적용 " 을 다시 클릭하면 <Y> 시뮬레이션의 계산식과 값이 업데이트됩니다 .

자세한 내용은 " 사용자 설계 " 장의 " 반응 시뮬레이션 "(103 페이지) 에서 확인하십시오 .

참고 : JMP PRO " 반응 시뮬레이션 " 을 사용하면 JMP Pro 시뮬레이션 기능을 사용하여 시뮬레이션 분석을 수행할 수 있습니다 . 자세한 내용과 DOE 예는 기본 분석의 " 시뮬레이션 " 장에서 확인하십시오 .

스크립트를 스크립트 창에 저장 " 그룹 직교 과포화 설계 " 창에서 지정한 설계의 스크립트를 생성한 후 열린 스크립트 창에 저장합니다 .

그룹 직교 과포화 설계에 대한 통계 상세 정보

설계 구성에 대한 통계 상세 정보는 Jones, et.al. (2019) 에서 확인하십시오 .

27 장

그룹 직교 과포화 설계 적합 플랫폼

그룹 직교 과포화 설계의 데이터 분석

그룹 직교 과포화 설계 적합 플랫폼을 사용하면 그룹 직교 과포화 설계를 분석할 수 있습니다. 이 분석 기법은 그룹 직교 과포화 설계의 그룹 직교 구조를 활용합니다.

그림 27.1 그룹 직교 과포화 설계 적합 결과

목차

그룹 직교 과포화 설계 적합 플랫폼 개요 683
그룹 직교 과포화 설계 적합의 예 683
그룹 직교 과포화 설계 적합 플랫폼 시작 685
그룹 직교 과포화 설계 적합 보고서 686
그룹 직교 과포화 설계 적합 플랫폼 옵션 687

그룹 직교 과포화 설계 적합 플랫폼 개요

GOSSD(그룹 직교 과포화 설계) 는 활성 효과를 쉽게 식별할 수 있도록 구성됩니다 . 이 설계는 같은 그룹의 요인은 서로 상관되지만 그룹 간의 요인은 직교하도록 요인을 그룹화합니다 . 이 구조로 인해 다음과 같은 두 가지 분석 기능을 사용할 수 있습니다 .

- 첫 번째 그룹에 할당된 가짜 요인은 2 차 이상 효과를 무시할 수 있다는 가정하에 분산을 추정하는 데 사용됩니다 .
- 그룹 간 직교성을 통해 2 단계 모형 선택 절차가 가능합니다 .

그룹 직교 과포화 설계 적합 플랫폼은 2 단계 모형 선택 절차를 수행하고, 모든 요인의 회귀 그림을 표시한 다음, 모형 적합 플랫폼에서 이 모형을 검사하는 옵션과 함께 활성 효과에 대한 모형을 제공합니다.

그룹 직교 과포화 설계 적합의 예

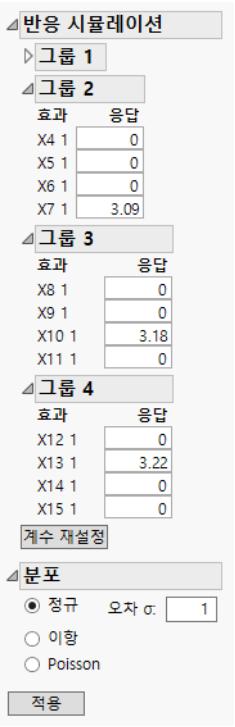
이 예에서는 최대 12 회 런으로 12 개의 요인을 연구하는 실험을 설계하려고 합니다 . GOSSD 에서 설계에 사용할 모수의 수는 최소 16 개입니다 . 절편에 대한 모수 1 개 , 가짜 요인에 대한 3 개 이상의 모수 , 설계 요인에 대한 12 개의 모수가 필요합니다 . 시뮬레이션된 반응을 사용하여 모형을 적합시킵니다 .

팁 : 가짜 요인은 3 개로 제한되지 않습니다 . 더 사용할 수 있습니다 .

설계 생성

1. **DOE > 특수 용도 > 그룹 직교 과포화 설계 > 그룹 직교 과포화 설계**를 선택합니다 .
2. **런 수**를 선택합니다 .
3. **런 수**에 12 를 입력하고 텍스트 상자 바깥쪽을 클릭합니다 .
4. " 구조 " 섹션에서 그룹 수 4, 그룹 크기 4, 모수 수가 16 인 두 번째 옵션을 선택합니다 .
5. X1 요인 이름을 클릭하고 이름을 Fake 1 로 변경합니다 .
6. X2 와 X3 에 대해 이 과정을 반복하여 각각 Fake 2 와 Fake 3 으로 이름을 지정합니다 .
7. " 그룹 직교 과포화 설계 " 의 빨간색 삼각형을 클릭하고 **반응 시뮬레이션**을 선택합니다 .
8. **설계 생성**을 클릭합니다 .
9. **테이블 생성**을 클릭합니다 .

그림 27.2 반응 시뮬레이션 창



"반응 시뮬레이션" 창에는 설계 테이블을 생성하는 데 사용된 설정이 표시됩니다. "그룹 1"의 개요 상자는 닫혀 있습니다. "그룹 1"은 "가짜" 요인을 포함합니다. 이 요인은 모형에 영향을 주지 않는다고 가정되므로 모두 0으로 설정됩니다. 값은 그림 27.2에 표시된 것과 다를 수 있습니다. 이 시뮬레이션의 경우 그룹당 활성 효과가 하나씩 있습니다.

팁 : " 반응 시뮬레이션 " 창을 사용하여 다른 시뮬레이션에 대한 새 값을 입력할 수 있습니다 .

시뮬레이션된 반응 분석

1. **DOE > 특수 용도 > 그룹 직교 과포화 설계 > 그룹 직교 과포화 설계 적합**을 선택합니다 .
2. 응답을 선택하고 **Y** 를 클릭합니다 .
3. Fake 1 ~ X15 를 선택하고 **X** 를 클릭합니다 .
4. **확인**을 클릭합니다 .

팁 : 그룹 직교 과포화 설계 플랫폼을 사용하여 생성한 설계 테이블에 " 모형 " 스크립트가 포함되어 있습니다 . 분석을 바로 시작하려면 이 스크립트를 실행하십시오 .

그림 27.3 그룹 직교 과포화 설계 적합 결과

시뮬레이션에서 X7, X10, X13을 효과가 있는 요인으로 설정했습니다. GOSSD 적합 절차에서 이 세 요인이 유의하다고 식별되었습니다. 또한 X14도 식별되었습니다. 효과의 크기와 유의성에 따라 추가적인 주의가 필요한지 여부를 결정해야 합니다.

참고 : 시뮬레이션 결과는 랜덤 결과 생성으로 인해 그림 27.3 에 표시된 것과 다를 수 있습니다 . 그룹에 활성 상태일 수 있는 추가 효과가 한 개 이상 있다는 경고가 나타날 수 있습니다 . 이는 과포화 설계의 한계이며 결과가 판단에 결정적이라고 할 수 없습니다 .

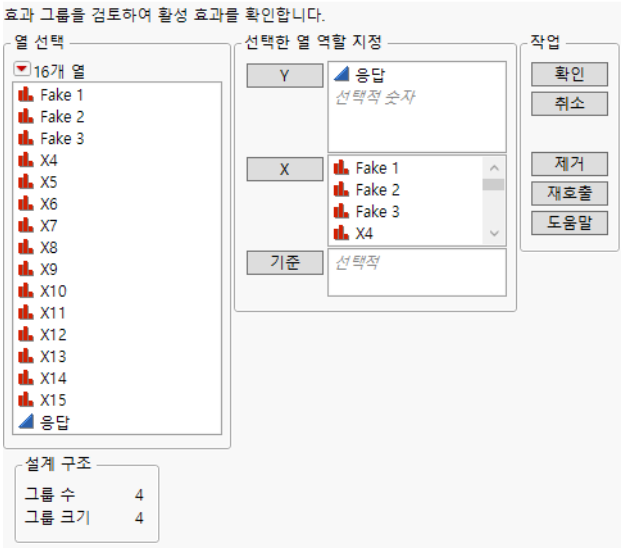
5. **모형 실행**을 클릭하여 식별된 모형을 실행합니다 .

그룹 직교 과포화 설계 적합 플랫폼 시작

DOE > 특수 용도 > 그룹 직교 과포화 설계 > 그룹 직교 과포화 설계 적합을 선택하여 그룹 직교 과포화 설계 적합 플랫폼을 시작합니다 .

참고 : JMP 에서 그룹 직교 과포화 설계 플랫폼을 사용하여 설계를 생성한 경우 설계 테이블에 " 모형 " 스크립트가 포함됩니다 . 분석을 바로 시작하려면 이 스크립트를 실행하십시오 .

그림 27.4 그룹 직교 과포화 설계 적합 시작 창



Y 하나 이상의 수치형 반응 변수입니다 .

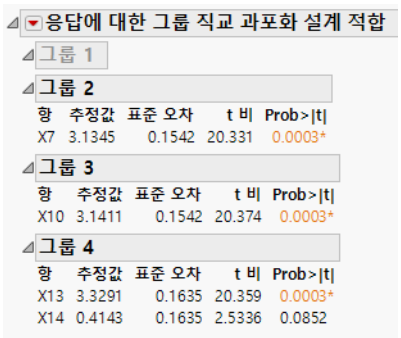
X 2 수준 연속형 또는 범주형 요인입니다 . 이 플랫폼은 분석을 수행할 때 GOSSD 의 고유 구조를 사용하므로 이러한 요인으로 GOSSD 를 정의해야 합니다 .

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

그룹 직교 과포화 설계 적합 보고서

그룹 직교 과포화 설계 적합 보고서에는 두 섹션이 있습니다 . 첫 번째 섹션에서는 그룹을 한 번에 하나씩 적합시켜 각 그룹 내에서 활성 효과의 추정값을 보고합니다 . 활성 그룹만 표시됩니다 . 이것이 분석의 1 단계입니다 . 2 단계에서는 결합 모형이 구성됩니다 .

그림 27.5 1 단계 결과



항 활성으로 식별된 주효과입니다 .

추정값 주효과에 대한 Y 회귀 적합의 모수 추정값입니다 .

표준 오차 그룹 1 의 가짜 요인을 사용하여 계산된 추정값의 표준 오차입니다 .

t 비 추정값을 표준 오차로 나눈 값입니다 .

Prob>|t| t 비와 오차 DF(자유도) 를 사용하여 계산된 *p* 값입니다 .

보고서의 두 번째 섹션에는 2 단계 분석의 결과가 포함되어 있습니다 . 2 단계에서는 설계 결과를 평가하는 데 사용되는 모형을 구성합니다 .

그림 27.6 2 단계 결과

항 활성으로 식별된 주효과입니다 .

추정값 주효과에 대한 Y 회귀 적합의 모수 추정값입니다 .

표준 오차 전체 데이터 집합을 사용하여 지정된 모형과 절편을 적합시켜 구한 추정값의 표준 오차입니다 .

t 비 추정값을 표준 오차로 나눈 값입니다 .

Prob>|t| t 비와 오차 DF(자유도) 를 사용하여 계산된 *p* 값입니다 .

RMSE 평균 제곱 오차의 제곱근입니다 .

DF RMSE 를 생성하는 데 사용되는 오차 추정값과 관련된 자유도입니다 .

그룹 직교 과포화 설계 적합 플랫폼 옵션

"그룹 직교 과포화 설계 적합" 플랫폼의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹 직교 과포화 설계 적합 플랫폼의 통계 상세 정보

2 단계 전략을 사용하여 그룹 직교 과포화 설계를 적합시킵니다 . 첫 번째 단계에서는 활성 효과를 포함하는 그룹을 식별합니다 .

- 가짜 요인으로 σ 추정값을 구한 후 추가 분석에서 가짜 요인을 제거합니다 .
- 각 요인 그룹의 유의성을 검정합니다 .

두 번째 단계에서는 활성 그룹의 개별 요인에 대한 유의성을 고려합니다. 이 단계에서는 단계별 부분집합 절차를 사용합니다. 자세한 내용은 Jones and Nachtsheim(2016) 에서 확인하십시오. 자세한 통계 정보는 Jones, et al. (2019) 에서 확인하십시오.

28 장

표본 크기 탐색기

표본 크기 및 검정력 계산기

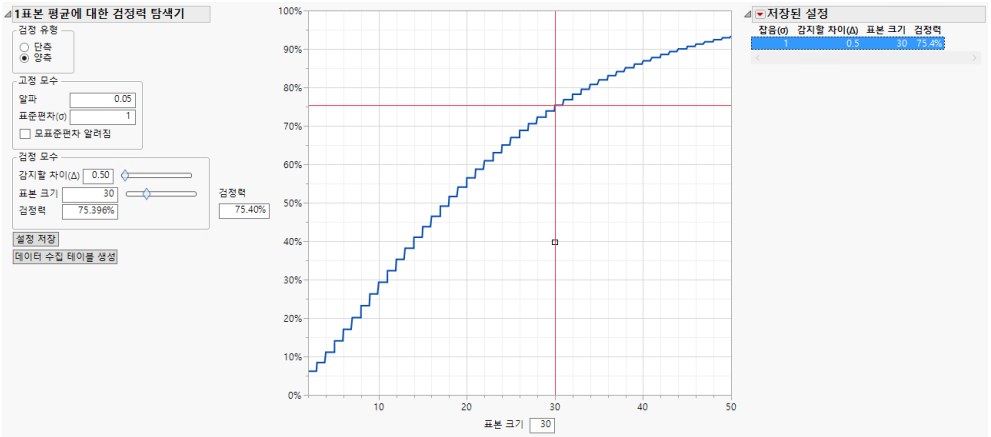
실험 또는 연구를 계획할 때는 표본 크기 탐색기를 사용하여 다음과 같은 질문에 답변합니다.

- 테스트할 단위 수는 몇 개입니까?
- 처리 평균의 차이를 감지할 수 있습니까?
- 지정된 너비를 가진 구간을 생성하려면 표본이 몇 개가 필요합니까?
- 고장 시간을 추정하기 위해 테스트해야 할 단위 수는 몇 개입니까?

표본 크기 탐색기는 연구 모수 및 가정을 조정하여 서로 다른 연구 설계를 탐색해 볼 수 있도록 해주는 대화식 유틸리티 모음입니다. 도구는 용도에 따라 다음과 같이 그룹화됩니다.

- 가설 검정을 위한 검정력
- 구간 추정을 위한 신뢰 구간
- 신뢰도 데모

그림 28.1 1 표본 평균에 대한 표본 크기 탐색기



목차

표본 크기 탐색기 플랫폼에 대한 개요 691
표본 크기 탐색기의 예 691
표본 크기 탐색기 시작 692
가설 검정에 대한 검정력 탐색기 692
1표본 평균에 대한 검정력 693
1표본 비율에 대한 검정력 695
1표본 분산에 대한 검정력 697
1표본 동등성에 대한 검정력 699
2독립 표본 평균에 대한 검정력 701
2독립 표본 비율에 대한 검정력 704
2독립 표본 분산에 대한 검정력 706
2독립 표본 동등성에 대한 검정력 708
ANOVA에 대한 검정력 710
신뢰 구간 계산기 712
1표본 평균에 대한 오차 한계 712
1표본 비율에 대한 오차 한계 715
1표본 분산에 대한 오차 한계 716
2독립 표본 평균에 대한 오차 한계 718
2독립 표본 비율에 대한 오차 한계 720
2독립 표본 분산에 대한 오차 한계 722
신뢰도 데모 계산기 723
모수 신뢰도 데모 724
비모수 신뢰도 데모 725
표본 크기 탐색기의 저장된 설정 726

표본 크기 탐색기 플랫폼에 대한 개요

표본 크기 탐색기는 표본 크기 및 검정력 계산에 사용됩니다. 표본 크기 탐색기를 사용하면 연구 가정이 연구에 필요한 관측값(런, 실험 단위 또는 표본)의 개수에 어떤 영향을 주는지 평가 및 비교할 수 있습니다. 가설 검정, 구간 추정 및 신뢰도 데모에 대한 탐색기가 있습니다.

검정력과 표본 크기에 대한 자세한 내용은 Barker(2011)에서 확인하십시오.

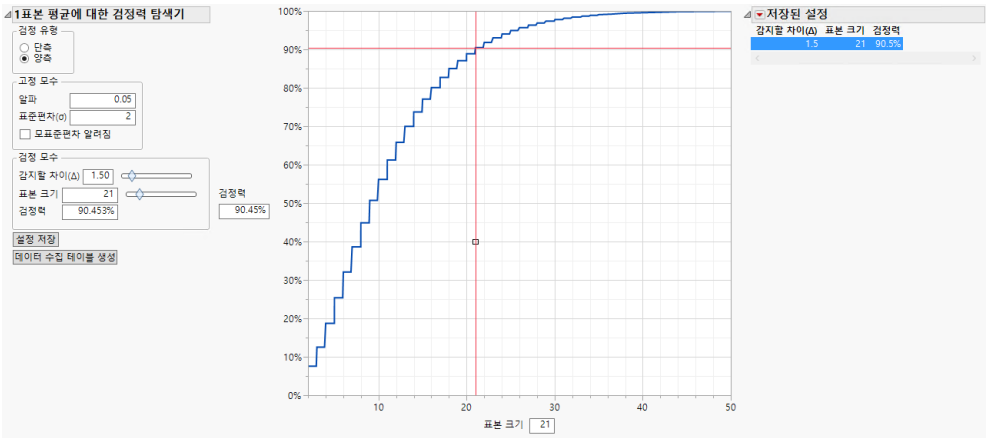
표본 크기 탐색기의 예

표본 크기 탐색기를 사용하면 새 소재 연구에서 테스트할 단위 수를 결정하는 데 도움이 됩니다. 회사에서 개발 중인 새 직물의 가연성이 현재 소재에 비해 성능이 개선되었음을 보여주고자 한다고 가정해 보겠습니다. 이전 테스트에 따르면 이 직물의 연소 시간에 대한 표준편차는 2초입니다.

테스트할 직물 표본 수를 계산하려면 1 표본 평균에 대한 검정력 탐색기를 사용합니다. 유의 수준 $\alpha = 0.05$ 일 때 1.5 초 차이를 감지할 수 있는 검정력이 90% 인 실험을 설계하려고 합니다.

1. **DOE > 표본 크기 탐색기 > 검정력 > 1 표본 평균에 대한 검정력**을 선택합니다.
2. **알파** 설정값 0.05 를 그대로 둡니다.
3. **표준편차**에 2 를 입력합니다.
4. **감지할 차이**에 1.5 를 입력합니다.
5. **검정력**에 0.9 를 입력합니다.
6. " 설정 저장 " 을 클릭하여 이러한 설정을 저장합니다.
7. 연구를 위해 슬라이더와 텍스트 상자를 사용하여 다른 표본 크기 및 검정력의 가능성을 탐색합니다.

그림 28.2 1 표본 평균 검정력 탐색기



유의 수준이 0.05일 때 연소 시간에서 1.5초의 유의차를 감지할 확률이 90%가 되려면 21개의 직물 표본이 필요합니다.

표본 크기 탐색기 시작

DOE > 표본 크기 탐색기를 선택하여 표본 크기 탐색기를 시작합니다. 선택할 수 있는 계산기가 여러 개 있습니다.

가설을 검정하려면 **DOE > 표본 크기 탐색기 > 검정력**을 선택합니다.

- "1 표본 평균에 대한 검정력 "(693 페이지)
- "1 표본 비율에 대한 검정력 "(695 페이지)
- "1 표본 분산에 대한 검정력 "(697 페이지)
- "1 표본 동등성에 대한 검정력 "(699 페이지)
- "2 독립 표본 평균에 대한 검정력 "(701 페이지)
- "2 독립 표본 평균에 대한 검정력 "(701 페이지)
- "2 독립 표본 분산에 대한 검정력 "(706 페이지)
- "2 독립 표본 동등성에 대한 검정력 "(708 페이지)
- "ANOVA 에 대한 검정력 "(710 페이지)

구간을 추정하려는 경우 **DOE > 표본 크기 탐색기 > 신뢰 구간**을 선택합니다 .

- "1 표본 평균에 대한 오차 한계 "(712 페이지)
- "1 표본 비율에 대한 오차 한계 "(715 페이지)
- "1 표본 분산에 대한 오차 한계 "(716 페이지)
- "2 독립 표본 평균에 대한 오차 한계 "(718 페이지)
- "2 독립 표본 비율에 대한 오차 한계 "(720 페이지)
- "2 독립 표본 분산에 대한 오차 한계 "(722 페이지)

신뢰도 데모를 하려는 경우 **DOE > 표본 크기 탐색기 > 신뢰도**를 선택합니다.

- " 모수 신뢰도 데모 "(724 페이지)
- " 비모수 신뢰도 데모 "(725 페이지)

가설 검정에 대한 검정력 탐색기

가설 검정에 사용할 표본 크기를 계산하려면 검정력 탐색기를 사용합니다. 상수(1표본 사례) 또는 두 그룹(2표본 사례) 사이에서 비롯될 수 있는 지정된 차이를 감지하려면 표본 크기가 가설 검정의 검정력에 미치는 영향을 탐색합니다. 1표본 사례의 경우 가설은 단일 모수(평균, 비율 또는

분산)와 관련됩니다. 2표본 사례의 경우 가설은 두 그룹을 비교하는 것이 목적입니다. 그룹이 2개를 초과하는 경우 ANOVA에 대한 검정력 탐색기를 사용할 수 있으며, 이에 대한 자세한 내용은 "ANOVA에 대한 검정력"에서 확인하십시오.

가설을 검정하려면 **DOE > 표본 크기 탐색기 > 검정력**을 선택합니다.

- "1 표본 평균에 대한 검정력 "(693 페이지)
- "1 표본 비율에 대한 검정력 "(695 페이지)
- "1 표본 분산에 대한 검정력 "(697 페이지)
- "1 표본 동등성에 대한 검정력 "(699 페이지)
- "2 독립 표본 평균에 대한 검정력 "(701 페이지)
- "2 독립 표본 평균에 대한 검정력 "(701 페이지)
- "2 독립 표본 분산에 대한 검정력 "(706 페이지)
- "2 독립 표본 동등성에 대한 검정력 "(708 페이지)
- "ANOVA 에 대한 검정력 "(710 페이지)

1 표본 평균에 대한 검정력

단일 평균에 대한 가설 검정에 사용할 표본 크기를 결정하려면 1표본 평균에 대한 검정력 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 검정력 > 1표본 평균에 대한 검정력**을 선택합니다. 변동 가정, 표본 크기, 검정력, 유의성 및 감지할 가설 차이 간의 균형을 탐색합니다. 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다.

H_0: $\mu = \mu_0$

다음 양측 대립가설과 비교

H_a: $\mu \neq \mu_0$

또는 다음 단측 대립가설과 비교

H_a: $\mu < \mu_0$ 또는 H_a: $\mu > \mu_0$

여기서 μ 는 실제 평균이고 μ_0 은 귀무가설 평균 또는 참조 값입니다. 감지할 차이는 중요성 때문에 감지해야 하는 크기 (Δ) 로, μ_0 와의 차이입니다. 동일한 유의 수준과 검정력이 주어진 경우 큰 차이를 감지하는 것보다 작은 차이를 감지하기 위해 더 큰 표본 크기가 필요합니다. 관심 모집단은 평균 μ 와 표준편차 σ 의 정규 분포를 따른다고 가정합니다.

1 표본 평균에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

검정 유형 단측 또는 양측 가설 검정을 지정합니다 .

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 **유의 수준**이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

표준편차 (σ) 가정된 표준편차입니다 .

팁 : 표준편차 단위로 측정된 차이를 감지하는 데 필요한 표본 크기를 추정하려면 표준편차 1 을 사용하십시오 .

모표준편차 알려짐 알려진 모집단 표준편차를 계산에 사용하도록 지정합니다 .

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

감지할 차이 (Δ) 실제 평균과 가설 평균 또는 참조 평균 간에 통계적으로 유의하다고 선언할 수 있는 최소 차이를 지정합니다 .

표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 .

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

1 표본 검정력에 대한 통계 상세 정보

1 표본 평균 계산은 σ 를 알 수 없는 경우 기존의 *t* - 검정을 기반으로 합니다 . σ 를 알고 있는 사례에서는 계산에 *z* 검정이 사용됩니다 . σ 를 알 수 없는 사례에서는 검정력이 대립가설에 따라 계산됩니다 .

상단측 대립가설의 경우 :

$$\Pr(\mu > \mu_0 \mid \mu = \mu_0 + \delta) = 1 - T\left(t_{1-\alpha,\, n-1}; n-1, \frac{\delta}{s/(\sqrt{n})}\right)$$

하단측 대립가설의 경우 :

$$\Pr(\mu < \mu_0 \mid \mu = \mu_0 - \delta) = T\left(-t_{1-\alpha,\, n-1}; n-1, \frac{\delta}{s/(\sqrt{n})}\right)$$

양측 대립가설의 경우 :

$$\Pr(\mu \neq \mu_0 \mid \mu = \mu_0 \pm \delta) = 1 - T\left(t_{1-\alpha/2,\, n-1}; n-1, \frac{\delta}{s/(\sqrt{n})}\right) + T\left(-t_{1-\alpha/2,\, n-1}; n-1, \frac{\delta}{s/(\sqrt{n})}\right)$$

다음은 각 요소에 대한 설명입니다.

α 는 유의 수준입니다 .
n 은 표본 크기입니다 .
s - 표준편차입니다 .
δ 는 감지할 차이입니다 .
$t_{1-\alpha,\nu}$ 는 자유도가 ν 인 중심 *t*- 분포의 (1 - α) 번째 분위수입니다 .
$T(t; \nu, \lambda)$ 는 자유도가 ν 이고 비중심성 모수 λ 가 있는 비중심 *t* 분포의 누적 분포 함수입니다 .

σ를 알고 있는 경우 검정력 계산에는 위의 방정식에 *z* 분포와 σ가 사용됩니다. δ와 *n*에 대해 닫힌 형식의 해가 없기 때문에 이들을 계산하는 데에는 수치 루틴이 사용됩니다.

1 표본 비율에 대한 검정력

단일 비율에 대한 가설 검정에 사용할 표본 크기를 결정하려면 1 표본 비율에 대한 검정력 탐색기를 사용합니다 . **DOE > 표본 크기 탐색기 > 검정력 > 1 표본 비율에 대한 검정력**을 선택합니다 . 표본 크기 , 검정력 , 유의성 및 감지할 가설 차이 간의 균형을 탐색합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

$H_0: p = p_0$

다음 양측 대립가설과 비교

$H_a: p \neq p_0$

또는 다음 단측 대립가설과 비교

$H_a: p < p_0$ 또는 $H_a: p > p_0$

여기서 p는 모비율이고 p_0은 귀무가설 비율입니다.

1 표본 비율에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

검정 유형 단측 또는 양측 가설 검정을 지정합니다 .

모형 유형

정확 검정 Clopper-Pearson 방법론을 기반으로 계산하도록 지정합니다 .

정규 근사 정규 근사 방법론을 기반으로 계산하도록 지정합니다 .

팁 : 이항 분포는 이산형이므로 표본이 작거나 비율이 0 또는 1 에 가까운 경우 실제 검정 크기는 명시된 유의 수준과 크게 다를 수 있습니다 . 유의 수준이 명시된 수준 이상이 되게 하려면 정확 검정을 사용합니다 .

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

가정된 비율 (p0) 연구에서 예상하거나 가정하는 비율 즉 , 귀무가설 값을 지정합니다 .

대립가설 비율 (pA) 검정할 비율 즉 , 대립가설 값을 지정합니다 .

표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 .

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

1 표본 비율 탐색기에 대한 통계 상세 정보

정확 검정의 경우 검정력은 대립가설의 양식을 기반으로 계산됩니다.

상단측 대립가설의 경우:

$$\Pr(p > p_0 | p = p_a) = \sum_{x = x_{1-\alpha, n, p_o}}^{n} \binom{n}{x} p_a^x (1-p_a)^{n-x}$$

하단측 대립가설의 경우:

$$\Pr(p < p_0 | p = p_a) = \sum_{x = x_0}^{x = x_{\alpha, n, p_o} - 1} \binom{n}{x} p_a^x (1-p_a)^{n-x}$$

양측 대립가설의 경우:

$$\Pr(p \neq p_0 | p = p_a) = \sum_{x = x_{1-\alpha/2, n, p_o}}^{n} \binom{n}{x} p_a^x (1-p_a)^{n-x} + \sum_{x = x_0}^{x = x_{\alpha/2, n, p_o} - 1} \binom{n}{x} p_a^x (1-p_a)^{n-x}$$

여기서 $x_{q,n,p}$는 시도 횟수가 n이고 확률이 p인 이항 분포의 q번째 분위수입니다.

정규 근사의 경우 검정력은 대립가설의 양식을 기반으로 계산됩니다.

상단측 대립가설의 경우:

$$\Pr(p > p_0 \mid p = p_a) = 1 - \Phi\left(z_{1-\alpha}\sqrt{\left(\frac{p_0(1-p_0)}{p_a(1-p_a)}\right)} - \frac{(p_a - p_0)}{\sqrt{\frac{p_a(1-p_a)}{n}}} \right)$$

하단측 대립가설의 경우:

$$\Pr(p < p_0 \mid p = p_a) = \Phi\left(-z_{1-\alpha}\sqrt{\left(\frac{p_0(1-p_0)}{p_a(1-p_a)}\right)} - \frac{(p_a - p_0)}{\sqrt{\frac{p_a(1-p_a)}{n}}} \right)$$

양측 대립가설의 경우:

$$\Pr(p \neq p_0 \mid p = p_a) =$$

$$1-\Phi\left(z_{1-\alpha/2}\sqrt{\left(\frac{p_0(1-p_0)}{p_a(1-p_a)}\right)} - \frac{(p_a - p_0)}{\sqrt{\frac{p_a(1-p_a)}{n}}} \right) + \Phi\left(-z_{1-\alpha/2}\sqrt{\left(\frac{p_0(1-p_0)}{p_a(1-p_a)}\right)} - \frac{(p_a - p_0)}{\sqrt{\frac{p_a(1-p_a)}{n}}} \right)$$

1 표본 분산에 대한 검정력

단일 분산에 대한 가설 검정에 사용할 표본 크기를 결정하려면 1표본 분산에 대한 검정력 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 검정력 > 1표본 분산에 대한 검정력**을 선택합니다. 표본 크기, 검정력, 유의성 및 감지할 가설 차이(귀무가설 값과 대립가설 값 간의 비율로 정의됨) 간의 균형을 탐색합니다. 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다.

H_0: $\sigma = \sigma_0$

다음 양측 대립가설과 비교

H_a: $\sigma \neq \sigma_0$

또는 다음 단측 대립가설과 비교

H_a: $\sigma < \sigma_0$ 또는 H_a: $\sigma > \sigma_0$

여기서 σ는 실제 분산이고 σ_0는 귀무가설 분산 또는 참조 값입니다. 감지할 차이는 표본 집합을 기반으로 감지해야 한다고 간주되는 크기로, σ_0와의 차이입니다. 이 차이는 σ_0/σ_a 비율로 표현되거나 귀무가설 분산과 대립가설 가정 분산의 비율로 표현됩니다. 동일한 유의 수준과 검정력이 주어진 경우 분산의 큰 차이를 감지하는 것보다 작은 차이를 감지하는 데 더 큰 표본 크기가 필요합니다. 관심 모집단은 평균 μ와 표준편차 σ의 정규 분포를 따른다고 가정합니다.

1 표본 분산에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

검정 유형 단측 또는 양측 가설 검정을 지정합니다.

고정 모수

알파 제 1 종 오류의 확률, 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다. 일반적으로 검정의 유의 수준이라고 합니다. 기본 유의 수준은 0.05 입니다.

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다.

분산 비율 (귀무가설 / 대립가설) 대립가설 분산 (기대 분산) 에 대한 귀무가설 분산 (참조 값) 의 비율을 지정합니다.

표본 크기 실험에 필요한 총 관측값 (런, 실험 단위 또는 표본) 수를 지정합니다.

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다.

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다. 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다. 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오.

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다.

1 표본 분산 탐색기에 대한 통계 상세 정보

단일 표본 그룹의 분산을 검정하기 위한 검정력 계산은 χ^2 검정을 기반으로 합니다. 대립가설의 양식을 기반으로 계산됩니다.

상단측 대립가설의 경우:

$$\Pr(\sigma > \sigma_0 \mid \sigma = \sigma_0/\rho) = 1 - \chi^2(\rho x_{1-\alpha,\, n-1}; n-1)$$

하단측 대립가설의 경우 :

$$\Pr(\sigma < \sigma_0 \mid \sigma = \sigma_0/\rho) = \chi^2(\rho x_{\alpha,\, n-1}; n-1)$$

양측 대립가설의 경우:

$$\Pr(\sigma \neq \sigma_0 \mid \sigma = \sigma_0/\rho) = 1 - \chi^2(\rho x_{1-\alpha/2,\, n-1}; n-1) + \chi^2(\rho x_{\alpha/2,\, n-1}; n-1)$$

다음은 각 요소에 대한 설명입니다.

α 는 유의 수준입니다 .
n 은 표본 크기입니다 .
$\rho = \sigma_a/\sigma_0$
$x_{1-\alpha,\nu}$ 는 자유도가 ν 인 중심 χ^2 분포의 (1 - α) 번째 분위수입니다 .
$\chi^2(x, \nu)$ 은 자유도가 ν 인 중심 χ^2 분포의 누적 분포 함수입니다 .

1 표본 동등성에 대한 검정력

단일 평균에 대한 동등성 검정에 사용할 표본 크기를 결정하려면 1 표본 동등성에 대한 검정력 탐색기를 사용합니다 . **DOE > 표본 크기 탐색기 > 검정력 > 1 표본 동등성에 대한 검정력**을 선택합니다 . 변동 가정 , 표본 크기 , 검정력 , 유의성 및 동등성 범위 간의 균형을 탐색합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

H_0: $\mu - \mu_0 \geq \delta_M$ 또는 H_0: $\mu - \mu_0 \leq \delta_m$

다음 대립가설과 비교

H_a: $\delta_m < \mu - \mu_0 < \delta_M$

여기서 μ 는 실제 평균 , μ_0 는 참조 값 , (δ_m, δ_M) 는 동등성 범위입니다 . 동일한 유의 수준과 검정력이 주어진 경우 큰 차이를 감지하는 것보다 작은 차이를 감지하기 위해 더 큰 표본 크기가 필요합니다 . 관심 모집단은 평균 μ 와 표준편차 σ 의 정규 분포를 따른다고 가정합니다 .

1 표본 동등성에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

동등성 범위

최대 차이 평균이 해당 값을 초과할 경우 참조 평균과 다르다고 간주되는 최대 값을 지정합니다 .

최소 차이 평균이 해당 값 미만일 경우 참조 평균과 다르다고 간주되는 최소 값을 지정합니다 .

참고 : 일반적으로 동등성 범위는 대칭입니다 .

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률입니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

표준편차 (σ) 가정된 표준편차입니다 .

모표준편차 알려짐 알려진 모집단 표준편차를 기반으로 계산하도록 지정합니다 .

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

평균의 차이 실제 평균과 가설 또는 참조 평균 (두 평균이 동등하다고 간주되는 경우) 간의 차이를 지정합니다 .

표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 .

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

1 표본 동등성 탐색기에 대한 통계 상세 정보

단일 표본 그룹에서 동등성을 검정하기 위한 검정력 계산은 Chow et al. 에 설명된 방법을 기반으로 합니다 . Box et al.(2008) 에서 논의한 설계에서 파생된 실험 결과가 포함되어 있습니다 .

σ 를 알 수 없는 경우 검정력 (1-β) 는 다음과 같이 계산됩니다 .

$$1-\beta = T\left(-t_{1-\alpha,\,n-1};n-1,\frac{\delta-\delta_M}{s/(\sqrt{n})}\right) - T\left(t_{1-\alpha,\,n-1};n-1,\frac{\delta-\delta_m}{s/(\sqrt{n})}\right)$$

다음은 각 요소에 대한 설명입니다 .

α 는 유의 수준입니다 .
n 은 표본 크기입니다 .
s - 표준편차입니다 .
δ 는 감지할 차이입니다 .
(δ_m, δ_M) 는 동등성 범위입니다 .
$t_{1-\alpha,\nu}$ 는 자유도가 ν 인 중심 *t*- 분포의 (1 - α) 번째 분위수입니다 .
$T(t; \nu, \lambda)$ 는 자유도가 ν 이고 비중심성 모수 λ 가 있는 비중심 *t* 분포의 누적 분포 함수입니다 .

σ 를 알고 있는 경우 검정력 (1-β) 는 다음과 같이 계산됩니다 .

$$1-\beta = \Phi\left(\frac{\delta-\delta_M}{\sigma/\sqrt{n}} - z_{1-\alpha}\right) + \Phi\left(\frac{\delta-\delta_m}{\sigma/\sqrt{n}} - z_{1-\alpha}\right) - 1$$

2 독립 표본 평균에 대한 검정력

두 그룹의 평균에 대한 가설 검정에 사용할 표본 크기를 결정하려면 2 독립 표본 평균에 대한 검정력 탐색기를 사용합니다 . **DOE > 표본 크기 탐색기 > 검정력 > 2 독립 표본 평균에 대한 검정력**을 선택합니다 . 표본 크기 , 검정력 , 유의성 및 감지할 가설 차이 간의 균형을 탐색합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

$$H_0: \mu_1 - \mu_2 = 0$$

다음 양측 대립가설과 비교

$$H_a: \mu_1 - \mu_2 \neq 0$$

또는 다음 단측 대립가설과 비교

$$H_a: \mu_1 - \mu_2 < 0 \quad \text{또는} \quad H_a: \mu_1 - \mu_2 > 0$$

여기서 μ_1과 μ_2는 두 모집단의 실제 평균입니다. 관심 모집단은 정규 분포를 따르고, 평균 간의 차이 (δ)를 감지하려고 한다고 가정합니다.

2 독립 표본 평균에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

검정 유형 단측 또는 양측 가설 검정을 지정합니다 .

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

그룹 1 표준편차 (σ_1) 그룹 중 하나인 그룹 1 의 가정된 표준편차를 지정합니다 . 오차 표준편차의 추정값은 이전 모형 적합에서 구한 RMSE(제곱근 평균 제곱 오차) 일 수 있습니다 .

그룹 2 표준편차 (σ_1) 두 번째 그룹인 그룹 2 의 가정된 표준편차를 지정합니다 . 오차 표준편차의 추정값은 이전 모형 적합에서 구한 RMSE(제곱근 평균 제곱 오차) 일 수 있습니다 .

모표준편차 알려짐 알려진 모집단 표준편차를 기반으로 계산하도록 지정합니다 .

검정 모수 감지할 차이 (Δ) 그룹 평균 간에 통계적으로 유의하다고 선언할 수 있는 최소 차이를 지정합니다 .

그룹 1 표본 크기 실험에서 그룹 1 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 1 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

그룹 2 표본 크기 실험에서 그룹 2 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 2 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

총 표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 검정력 곡선은 총 표본 크기를 기반으로 합니다 .

팁 : 총 표본 크기를 변경하지 않고 그대로 둔 채로 한쪽 그룹의 표본 크기를 조정하면 나머지 다른 그룹의 표본 크기에 영향을 주게 됩니다 . 총 표본 크기를 조정하면 두 그룹이 그에 비례하여 조정되며 그렇지 않을 경우 한쪽 그룹이 잠기게 됩니다 . 한쪽 그룹 표본 크기만 잠글 수 있습니다 .

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

2 독립 평균에 대한 통계 상세 정보

두 표본 그룹의 평균 차이를 검정하는 검정력 계산은 기존의 t 검정을 기반으로 하며, σ_1과 σ_2를 알고 있는 경우에는 *z* 검정을 기반으로 합니다.

σ를 알 수 없는 사례에서는 $\sigma_1=\sigma_2=\sigma$이면 대립가설의 양식을 기반으로 검정력이 계산됩니다. 상단측 대립가설의 경우:

$$\Pr(\mu_1 > \mu_2 \mid \mu_1 = \mu_2 + \delta) = 1 - T\left(t_{1-\alpha,\, n_1+n_2-2;n_1+n_2-2},\ \frac{\delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)$$

하단측 대립가설의 경우 :

$$\Pr(\mu_1 < \mu_2 \mid \mu_1 = \mu_2 - \delta = T\left(-t_{1-\alpha,\, n_1+n_2-2;n_1+n_2-2},\ \frac{\delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)$$

양측 대립가설의 경우:

$$\Pr(\mu_1 \neq \mu_2 \mid \mu_1 = \mu_2 \pm \delta) =$$

$$1-T\left(t_{1-\frac{\alpha}{2},\,n_1+n_2-2};n_1+n_2-2,\frac{\delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)+T\left(-t_{1-\frac{\alpha}{2},\,n_1+n_2-2};n_1+n_2-2,\frac{\delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)$$

다음은 각 요소에 대한 설명입니다.

α 는 유의 수준입니다 .

n_1 과 n_2 는 그룹 표본 크기입니다 .

s 는 합동 표준 편차입니다 .

δ 는 감지할 차이입니다 .

$t_{1-\alpha,\nu}$ 는 자유도가 ν 인 중심 t- 분포의 $(1 - \alpha)$ 번째 분위수입니다 .

$T(t; \nu, \lambda)$ 는 자유도가 ν 이고 비중심성 모수 λ 가 있는 비중심 t 분포의 누적 분포 함수입니다 .

σ_1과 σ_2가 같지 않다고 가정하는 경우 위의 방정식에서 자유도 n_1+ n_2-2는 다음과 같이 설정됩니다.

$$\mathrm{DF}=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$$

σ_1과 σ_2를 알고 있는 경우 검정력 계산에는 z 분포가 사용됩니다. 이 검정력은 대립가설의 양식을 기반으로 계산됩니다. 상단측 대립가설의 경우:

$$\Pr(\mu_1>\mu_2\,|\,\mu_1=\mu_2+\delta)=1-\Phi\left(z_{1-\alpha}-\frac{\delta}{\sqrt{\frac{\sigma_1}{n_1}+\frac{\sigma_2}{n_2}}}\right)$$

하단측 대립가설의 경우 :

$$\Pr(\mu_1<\mu_2\,|\,\mu_1=\mu_2-\delta=\Phi\left(-z_{1-\alpha}-\frac{\delta}{\sqrt{\frac{\sigma_1}{n_1}+\frac{\sigma_2}{n_2}}}\right)$$

양측 대립가설의 경우:

$$\Pr(\mu_1 \neq \mu_2 \mid \mu_1 = \mu_2 \pm \delta) = 1 - \Phi\left(z_{1-\frac{\alpha}{2}} - \frac{\delta}{\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}}\right) + \Phi\left(-z_{1-\alpha} - \frac{\delta}{\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}}\right)$$

다음은 각 요소에 대한 설명입니다.

α 는 유의 수준입니다 .
n_1 과 n_2 는 그룹 표본 크기입니다 .
σ_1 과 σ_2 는 알고 있는 그룹 표준편차입니다 .
δ 는 감지할 차이입니다 .
$z_{1-\alpha}$ 는 z- 분포의 $(1 - \alpha)$ 번째 분위수입니다 .
$\Phi(x)$ 는 정규 분포의 누적 분포 함수입니다 .

2 독립 표본 비율에 대한 검정력

두 그룹의 비율에 대한 가설 검정에 사용할 표본 크기를 결정하려면 2 독립 표본 비율에 대한 검정력 탐색기를 사용합니다 . **DOE > 표본 크기 탐색기 > 검정력 > 2 독립 표본 비율에 대한 검정력**을 선택합니다 . 표본 크기 , 검정력 , 유의성 및 감지할 가설 차이 간의 균형을 탐색합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

$$H_0: p_1 - p_2 = D_0$$

다음 양측 대립가설과 비교

$$H_a: p_1 - p_2 \neq D_0$$

또는 다음 단측 대립가설 중 하나와 비교

$$H_a: (p_1 - p_2) < D_0 \quad \text{또는} \quad H_a: (p_1 - p_2) > D_0$$

여기서 p_1과 p_2는 두 모집단의 모비율이고, D_0은 가설 비율 차이입니다.

2 독립 표본 비율에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

검정 유형 단측 또는 양측 가설 검정을 지정합니다 .

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

그룹 1 비율 (p1) 그룹 1 에 대해 예상하거나 가정하는 비율을 지정합니다 .

그룹 2 비율 (p2) 그룹 2 에 대해 예상하거나 가정하는 비율을 지정합니다 .

그룹 1 표본 크기 실험에서 그룹 1 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 1 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

그룹 2 표본 크기 실험에서 그룹 2 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 2 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

총 표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 검정력 곡선은 총 표본 크기를 기반으로 합니다 .

팁 : 총 표본 크기를 변경하지 않고 그대로 둔 채로 한쪽 그룹의 표본 크기를 조정하면 나머지 다른 그룹의 표본 크기에 영향을 주게 됩니다 . 총 표본 크기를 조정하면 두 그룹이 그에 비례하여 조정되며 그렇지 않을 경우 한쪽 그룹이 잠기게 됩니다 . 한쪽 그룹 표본 크기만 잠글 수 있습니다 .

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

2 독립 표본 비율에 대한 통계 상세 정보

두 표본 그룹의 비율 차이를 검정하기 위한 검정력 계산은 정규 근사를 기반으로 합니다. 이 계산은 대립가설의 양식에 따라 달라집니다. 상단측 대립가설의 경우:

$$\Pr(p_1 > p_2 \mid p_1 = p_2 + \delta) = 1 - \Phi\left(z_{1-\alpha} - \frac{\delta}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}\right)$$

하단측 대립가설의 경우 :

$$\Pr(p_1 > p_2 \mid p_1 = p_2 - \delta) = \Phi\left(-z_{1-\alpha} - \frac{\delta}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}\right)$$

양측 대립가설의 경우:

$$\Pr(p_1 \neq p_2 \mid p_1 = p_2 \pm \delta) =$$

$$1 - \Phi\left(z_{1-\frac{\alpha}{2}} - \frac{\delta}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \right) + \Phi\left(-z_{1-\alpha} - \frac{\delta}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \right)$$

다음은 각 요소에 대한 설명입니다.

α 는 유의 수준입니다 .
n_1 과 n_2 는 그룹 표본 크기입니다 .
p_1 과 p_2 는 알고 있는 그룹 표준편차입니다 .
δ 는 감지할 차이입니다 .
$z_{1-\alpha}$ 는 z- 분포의 (1 - α) 번째 분위수입니다 .
$\Phi(x)$ 는 정규 분포의 누적 분포 함수입니다 .

2 독립 표본 분산에 대한 검정력

두 그룹의 분산에 대한 가설 검정에 사용할 표본 크기를 결정하려면 2독립 표본 분산에 대한 검정력 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 검정력 > 2독립 표본 분산에 대한 검정력**을 선택합니다. 표본 크기, 검정력, 유의성 및 감지할 가설 차이 간의 균형을 탐색합니다. 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다.

$$H_0: \sigma_1 = \sigma_2$$

다음 양측 대립가설과 비교

$$H_a: \sigma_1 \neq \sigma_2$$

또는 다음 단측 대립가설과 비교

$$H_a: \sigma_1 < \sigma_2 \quad \text{또는} \quad H_a: \sigma_1 > \sigma_2$$

여기서 σ_1은 그룹 1의 분산이고 σ_2는 그룹 2의 분산입니다. 감지할 차이는 표본 집합을 기반으로 감지해야 한다고 간주되는 크기로, σ_0와의 차이입니다. 이 차이는 σ_2/σ_1 비율 즉, 분산의 비율로 표현됩니다. 동일한 유의 수준과 검정력이 주어진 경우 분산의 큰 차이를 감지하는 것보다 작은 차이를 감지하는 데 더 큰 표본 크기가 필요합니다. 관심 모집단은 정규 분포라고 가정합니다.

2 독립 표본 분산에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

검정 유형 단측 또는 양측 가설 검정을 지정합니다.

고정 모수

알파 제 1 종 오류의 확률, 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다. 일반적으로 검정의 유의 수준이라고 합니다. 기본 유의 수준은 0.05 입니다.

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다.

분산 비율 (그룹 2/ 그룹 1) 분산의 비율을 지정합니다.

그룹 1 표본 크기 실험에서 그룹 1 에 필요한 관측값 (런, 실험 단위 또는 표본) 수를 지정합니다. 그룹 1 표본 크기를 잠그려면 " 잠금 " 을 선택합니다.

그룹 2 표본 크기 실험에서 그룹 2 에 필요한 관측값 (런, 실험 단위 또는 표본) 수를 지정합니다. 그룹 2 표본 크기를 잠그려면 " 잠금 " 을 선택합니다.

총 표본 크기 실험에 필요한 총 관측값 (런, 실험 단위 또는 표본) 수를 지정합니다. 검정력 곡선은 총 표본 크기를 기반으로 합니다.

팁 : 총 표본 크기를 변경하지 않고 그대로 둔 채로 한쪽 그룹의 표본 크기를 조정하면 나머지 다른 그룹의 표본 크기에 영향을 주게 됩니다. 총 표본 크기를 조정하면 두 그룹이 그에 비례하여 조정되며 그렇지 않을 경우 한쪽 그룹이 잠기게 됩니다. 한쪽 그룹 표본 크기만 잠글 수 있습니다.

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다.

2 독립 표본 분산 탐색기에 대한 통계 상세 정보

두 표본 그룹의 분산 비율을 검정하기 위한 검정력 계산은 표준 F 검정을 기반으로 합니다. 이 계산은 대립가설의 양식에 따라 달라집니다. 상단측 대립가설의 경우:

$$\Pr(\sigma_0 > \sigma_1 \mid \sigma_1 = \sigma_0 \rho) = 1 - F(\rho f_{1-\alpha, n_1-1, n_2-1}; n_1 - 1, n_2 - 1)$$

하단측 대립가설의 경우 :

$$\Pr(\sigma_0 < \sigma_1 \mid \sigma_1 = \sigma_0 \rho) = F(\rho f_{\alpha, n_1-1, n_2-1}; n_1 - 1, n_2 - 1)$$

양측 대립가설의 경우:

$$\Pr(\sigma_0 \neq \sigma_1 \mid \sigma_1 = \sigma_0 \rho) =$$

$$1 - F\left(\rho f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}; n_1-1, n_2-1\right) + F\left(\rho f_{\frac{\alpha}{2}, n_1-1, n_2-1}; n_1-1, n_2-1\right)$$

다음은 각 요소에 대한 설명입니다.

α 는 유의 수준입니다 .
n_1 과 n_2 는 그룹 표본 크기입니다 .
$\rho=\sigma_1/\sigma_0$
$f_{1-\alpha,\nu1,\nu2}$ 는 자유도가 ν_1 및 ν_2 인 F 분포의 (1 - α) 번째 분위수입니다 .
$F(x, \nu)$ 는 자유도가 ν 인 F 분포의 누적 분포 함수입니다 .

2 독립 표본 동등성에 대한 검정력

두 그룹의 동등성 검정에 사용할 표본 크기를 결정하려면 2 표본 동등성에 대한 검정력 탐색기를 사용합니다 . **DOE > 표본 크기 탐색기 > 검정력 > 2 독립 표본 동등성에 대한 검정력**을 선택합니다 . 변동 가정 , 표본 크기 , 검정력 , 유의성 및 동등성 범위 간의 균형을 탐색합니다 . 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다 .

$$H_0: \mu_1 - \mu_2 \geq \delta_M \text{ 또는 } H_0: \mu_1 - \mu_2 \leq \delta_m$$

다음 대립가설과 비교

$$H_a: \delta_m < \mu_1 - \mu_2 < \delta_M$$

여기서 μ_1 및 μ_2 는 실제 그룹 평균이고 (δ_m, δ_M) 는 동등성 범위입니다 . 동일한 유의 수준과 검정력이 주어진 경우 큰 차이를 감지하는 것보다 작은 차이를 감지하기 위해 더 큰 표본 크기가 필요합니다 . 관심 모집단은 정규 분포라고 가정합니다 .

2 독립 표본 동등성에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

동등성 범위

최대 차이 이 차이 값보다 크면 참조 평균과 다르다고 간주하게 되는 차이를 지정합니다 .

최소 차이 이 차이 값보다 작으면 참조 평균과 다르다고 간주하게 되는 차이를 지정합니다 .

참고 : 일반적으로 동등성 차이 한계는 대칭입니다 . 하지만 반드시 대칭일 필요는 없습니다 .

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

그룹 1 표준편차 (σ_1) 그룹 중 하나인 그룹 1 의 가정된 표준편차를 지정합니다 .

그룹 2 표준편차 (σ_1) 두 번째 그룹인 그룹 2 의 가정된 표준편차를 지정합니다 .

모표준편차 알려짐 알려진 모집단 표준편차를 기반으로 계산하도록 지정합니다 .

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

평균의 차이 동등성 범위를 정의하는 그룹 평균 간의 차이를 지정합니다 .

그룹 1 표본 크기 실험에서 그룹 1 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 1 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

그룹 2 표본 크기 실험에서 그룹 2 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 2 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

총 표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 검정력 곡선은 총 표본 크기를 기반으로 합니다 .

팁 : 총 표본 크기를 변경하지 않고 그대로 둔 채로 한쪽 그룹의 표본 크기를 조정하면 나머지 다른 그룹의 표본 크기에 영향을 주게 됩니다 . 총 표본 크기를 조정하면 두 그룹이 그에 비례하여 조정되며 그렇지 않을 경우 한쪽 그룹이 잠기게 됩니다 . 한쪽 그룹 표본 크기만 잠글 수 있습니다 .

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

2 독립 표본 동등성 탐색기에 대한 통계 상세 정보

두 그룹 평균의 동등성을 검정하기 위한 검정력 계산은 Chow et al. 에 설명된 방법을 기반으로 합니다 . Box et al.(2008) 에서 논의한 설계에서 파생된 실험 결과가 포함되어 있습니다 .

σ_1 과 σ_2 를 알 수 없는 경우 검정력 (1-β) 는 다음과 같이 계산됩니다 .

$$1-\beta = \left(-t_{1-\alpha,\, df};df,\ \frac{\delta - \delta_M}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right) - T\left(t_{1-\alpha,\, df};df,\ \frac{\delta - \delta_m}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right)$$

다음은 각 요소에 대한 설명입니다.

$$df = \begin{cases} n_1 + n_2 - 2 & \text{단, } \sigma_1 = \sigma_2 \\ \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2}\right)^2} & \text{단, } \sigma_1 \neq \sigma_2 \end{cases}$$

α 는 유의 수준입니다.
n_1 과 n_2 는 그룹 표본 크기입니다.
s_1 과 s_2 는 그룹 표준편차입니다.
δ 는 감지할 차이입니다.
(δ_m, δ_M) 는 동등성 범위입니다.
$t_{1-\alpha,\nu}$ 는 자유도가 ν 인 중심 t- 분포의 (1 - α) 번째 분위수입니다.
$T(t; \nu, \lambda)$ 는 자유도가 ν 이고 비중심성 모수 λ 가 있는 비중심 t 분포의 누적 분포 함수입니다.

σ 를 알고 있는 경우 검정력 (1-β) 는 다음과 같이 계산됩니다.

$$1 - \beta = \Phi\left(\left|\frac{\delta - \delta_M}{\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}}\right| - z_{1-\alpha}\right) + \Phi\left(\left|\frac{\delta - \delta_m}{\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}}\right| - z_{1-\alpha}\right) - 1$$

ANOVA 에 대한 검정력

ANOVA 를 사용하여 분석할 k 개의 그룹 또는 처리에 대한 연구에 사용할 표본 크기를 결정하려면 ANOVA 에 대한 검정력 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 검정력 > ANOVA 에 대한 검정력**을 선택합니다. 변동 가정, 표본 크기, 검정력 및 유의성 간의 균형을 탐색합니다. 표본 크기 및 검정력은 다음 가설 검정과 관련됩니다.

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

다음 양측 대립가설과 비교

H_a: 모든 평균이 동일하지는 않음

다음은 각 요소에 대한 설명입니다 .

$$X_{ij} \sim N(\mu_j, \sigma^2) \text{ for } i = 1, \ldots, n, \; j = 1, \ldots, k$$

ANOVA 에 대한 검정력 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

그룹 수 실험에 사용되는 그룹 또는 처리의 수를 지정합니다 .

그룹 내 분산 (σ^2) 표준편차가 모든 그룹에서 같다고 가정하는 경우 각 그룹의 가정된 분산을 지정합니다 .

검정 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

그룹 간 분산 총 평균을 이루는 개별 그룹 평균의 분산을 지정합니다 .

표본 크기 (그룹당) 실험의 각 그룹에 필요한 관측값 (런 , 실험 단위 또는 처리) 수를 지정합니다 .

검정력 귀무가설이 거짓일 때 기각할 확률을 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 검정력도 증가합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

ANOVA 탐색기에 대한 통계 상세 정보

각 그룹의 표준편차가 같다고 가정하고 여러 그룹의 평균 중에서 차이를 검정하기 위한 검정력 계산은 표준 F 검정을 기반으로 합니다. 검정력 (1-β)는 다음과 같이 계산됩니다.

$$1 - \beta = F\left(f_{1-\alpha, N-K, n-1}, \frac{\sigma_{BG}^2}{\sigma^2}\right)$$

다음은 각 요소에 대한 설명입니다.

$F(df_1, df_2, \lambda)$는 비중심 모수 λ가 있는 비중심 F-분산입니다.

K는 그룹 수입니다.

n은 각 그룹 내의 표본 수입니다(모든 그룹이 같다고 가정).

$N = nK$

σ^2_{BG}는 총 평균을 이루는 그룹 평균의 변동입니다.

σ^2는 그룹 내 분산입니다(모든 그룹이 같다고 가정).

신뢰 구간 계산기

사전에 지정된 특성으로 구간 추정값을 구하기 위해 연구 표본 크기를 선택하려면 신뢰 구간 계산기를 사용합니다. 표본 크기가 구간 추정값의 오차 한계에 미치는 영향을 탐색합니다. 하나의 평균에 대해 신뢰 구간, 예측 구간 또는 공차 구간에 대한 표본 크기를 탐색합니다. 두 개의 평균에 대한 신뢰 구간의 표본 크기를 탐색하거나 하나 또는 두 개의 비율이나 분산에 대한 신뢰 구간의 표본 크기를 탐색합니다.

구간을 추정하려면 **DOE > 표본 크기 탐색기 > 신뢰 구간**을 선택합니다 .

- "1 표본 평균에 대한 오차 한계 "(712 페이지)
- "1 표본 비율에 대한 오차 한계 "(715 페이지)
- "1 표본 분산에 대한 오차 한계 "(716 페이지)
- "2 독립 표본 평균에 대한 오차 한계 "(718 페이지)
- "2 독립 표본 비율에 대한 오차 한계 "(720 페이지)
- "2 독립 표본 분산에 대한 오차 한계 "(722 페이지)

1 표본 평균에 대한 오차 한계

신뢰 구간, 예측 구간 또는 공차 구간에 대한 표본 크기를 결정하려면 1 표본 평균에 대한 구간 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰 구간 > 1 표본 평균에 대한 오차 한계**를 선택합니다. 변동 가정, 표본 크기, 유의성 및 오차 한계 간의 균형을 탐색합니다.

1 표본 평균에 대한 구간 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

구간 용도

신뢰도 평균에 대한 신뢰 구간을 지정합니다 .

예측 미래 관측값 하나에 대한 예측 구간을 지정합니다 .

공차 모집단 비율을 포함하는 공차 구간을 지정합니다 .

구간 유형

한계 단측 구간을 지정합니다 (상한 또는 하한).

구간 양측 구간을 지정합니다 .

고정 모수

알파 신뢰 수준 1 - 알파를 지정합니다 . 유의 수준 기본값은 95% 신뢰 구간에 대해 0.05 입니다 .

표준편차 (σ) 가정된 표준편차입니다 .

비율 (" 구간 용도 " 가 " 공차 " 로 선택된 경우에만 사용할 수 있음) 해당 공차 구간에 포함될 모집단의 비율을 지정합니다 .

모표준편차 알려짐 알려진 모집단 표준편차를 기반으로 계산하도록 지정합니다 .

구간 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

표본 크기 구간을 생성하는 데 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 .

오차 한계 (" 구간 유형 " 이 " 구간 " 으로 선택된 경우에만 사용할 수 있음) 구간 너비의 1/2 을 지정합니다 . 다른 모든 모수가 고정된 경우 표본 크기가 커질수록 오차 한계는 감소합니다 .

한계 크기 (" 구간 유형 " 이 " 한계 " 로 선택된 경우에만 사용할 수 있음) 한계에서 추정값까지의 거리를 지정합니다 . 다른 모든 모수가 고정된 경우 표본 크기가 커질수록 한계는 감소합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

1 표본 구간 탐색기에 대한 통계 상세 정보

σ가 알려져 있을 경우 표준 정규 기반 프로시저를 사용하여 각 구간 유형을 계산하며 이외의 경우에는 *t* 분포 프로시저 또는 근사 방식을 사용합니다.

신뢰 구간

신뢰 구간의 MOE(오차 한계)는 다음과 같이 계산됩니다.

$$\text{MOE} = \begin{cases} z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}} & \text{단, } \sigma \text{ 알 수 있음} \\ t_{1-\alpha/2,\,n-1}\dfrac{s}{\sqrt{n}} & \text{단, } \sigma \text{ 알 수 없음} \end{cases}$$

한계는 다음과 같이 계산됩니다.

$$\text{한계} = \begin{cases} \pm z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}} & \text{단, } \sigma \text{ known} \\ t_{1-\alpha,\,n-1}\dfrac{s}{\sqrt{n}} & \text{단, } \sigma \text{ unknown} \end{cases}$$

예측 구간

예측 구간의 MOE(오차 한계)는 다음과 같이 계산됩니다.

$$\text{MOE} = \begin{cases} z_{1-\alpha/2}\sigma\sqrt{\left(1+\dfrac{1}{n}\right)} & \text{단, } \sigma \text{ 알 수 있음} \\ t_{1-\alpha/2,\,n-1}s\sqrt{\left(1+\dfrac{1}{n}\right)} & \text{if } \sigma \text{ 알 수 없음} \end{cases}$$

한계는 다음과 같이 계산됩니다.

$$\text{한계} = \begin{cases} \pm z_{1-\alpha}\sigma\sqrt{\left(1+\dfrac{1}{n}\right)} & \text{단, } \sigma \text{ 알 수 있음} \\ \pm t_{1-\alpha,\,n-1}s\sqrt{\left(1+\dfrac{1}{n}\right)} & \text{단, } \sigma \text{ 알 수 없음} \end{cases}$$

공차 구간

모집단의 비율 q에 대한 공차 구간의 경우 오차 한계 또는 한계는 Krishnamoorthy and Mathew(2009)에 설명된 근사 프로시저를 기반으로 계산됩니다.

공차 구간의 MOE(오차 한계)는 다음과 같이 계산됩니다.

$$\text{MOE} = \begin{cases} \sigma\sqrt{\left(\dfrac{N\chi^2_{q,\,1,\,\frac{1}{n}}}{\chi^2_{\alpha,\,N}}\right)} & \text{단, } \sigma \text{ 알 수 있음} \\ s\sqrt{\left(\dfrac{(n-1)\chi^2_{q,\,1,\,\frac{1}{n}}}{\chi^2_{\alpha,\,n-1}}\right)} & \text{단, } \sigma \text{ 알 수 없음} \end{cases}$$

한계는 다음과 같이 계산됩니다.

$$\text{한계} = \begin{cases} \pm\dfrac{t_{1-\alpha,\,N,\,z_q\sqrt{n}}}{\sqrt{n}} & \text{단, } \sigma \text{ 알 수 있음} \\ \pm\dfrac{t_{1-\alpha,\,n-1,\,z_q\sqrt{n}}}{\sqrt{n}} & \text{if } \sigma \text{ 알 수 없음} \end{cases}$$

여기서 N은 점근적 동작을 나타내는 적당히 큰 정수값입니다. JMP는 N을 2000으로 설정합니다.

1 표본 비율에 대한 오차 한계

신뢰 구간에 대한 표본 크기를 결정하려면 1표본 비율에 대한 구간 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰 구간 > 1표본 비율에 대한 오차 한계**를 선택합니다. 가정된 비율, 표본 크기, 유의성 및 구간에 대한 오차 한계 간의 균형을 탐색합니다. 계산에는 정규 근사가 사용됩니다.

1 표본 비율에 대한 구간 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

구간 유형

한계 단측 구간을 지정합니다 (상한 또는 하한).

구간 양측 구간을 지정합니다 .

고정 모수

알파 신뢰 수준 1 - 알파를 지정합니다. 유의 수준 기본값은 95% 신뢰 구간에 대해 0.05입니다.

구간 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다.

비율 구간의 가정된 비율을 지정합니다.

표본 크기 구간을 생성하는 데 필요한 총 관측값(런, 실험 단위 또는 표본) 수를 지정합니다.

오차 한계 구간 너비의 1/2 을 지정합니다. 다른 모든 모수가 고정된 경우 표본 크기가 커질수록 오차 한계는 감소합니다.

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다. 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다. 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오.

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다.

1 표본 비율 구간 탐색기에 대한 통계 상세 정보

모집단 비율을 알아내기 위한 구간 계산은 Agresti and Coull 방법을 기반으로 합니다. 자세한 내용은 Agresti and Coull(1998)에서 확인하십시오.

신뢰 구간의 MOE(오차 한계)는 다음과 같이 계산됩니다.

$$\text{MOE} = z_{1-\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

한계는 다음과 같이 계산됩니다.

$$\text{한계} = \pm z_{1-\alpha}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

다음은 각 요소에 대한 설명입니다.

$$\tilde{n} = n + z_q^2$$

및

$$\tilde{p} = \frac{np + z_q^2/2}{\tilde{n}}$$

구간의 경우 $q = 1 - \alpha/2$, 한계의 경우 $q = 1 - \alpha$

1 표본 분산에 대한 오차 한계

신뢰 구간에 대한 표본 크기를 결정하려면 1표본 분산에 대한 구간 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰 구간 > 1표본 분산에 대한 오차 한계**를 선택합니다. 표본 크기, 유의성 및 구간에 대한 오차 한계 간의 균형을 탐색합니다.

1 표본 분산에 대한 구간 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

구간 유형

하한 단측 하한 구간을 지정합니다.

상한 단측 상한 구간을 지정합니다.

구간 양측 구간을 지정합니다.

고정 모수

알파 신뢰 수준 1 - 알파를 지정합니다. 유의 수준 기본값은 95% 신뢰 구간에 대해 0.05입니다.

구간 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다.

표본 크기 구간을 생성하는 데 필요한 총 관측값(런, 실험 단위 또는 표본) 수를 지정합니다.

구간 너비 구간의 전체 너비를 지정합니다. 다른 모든 모수가 고정되면 표본 크기가 커질수록 구간 너비는 줄어듭니다.

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다. 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다. 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오.

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다.

1 표본 분산 구간 탐색기에 대한 통계 상세 정보

모집단 분산을 알아내기 위한 구간 계산은 X^2 분산을 기반으로 합니다. 구간 한계는 표본 추정값을 기준으로 대칭이 아니라는 점에 유의합니다.

구간 너비는 다음과 같이 계산됩니다.

$$\text{너비} = (n-1)\left(\frac{1}{\chi^2_{\alpha/2,\,n-1}} - \frac{1}{\chi^2_{1-\alpha/2,\,n-1}}\right)$$

하한은 다음과 같이 계산됩니다

$$\text{하한} = \frac{n-1}{\chi^2_{1-\alpha,\,n-1}}$$

상한은 다음과 같이 계산됩니다.

$$\text{상한} = \frac{n-1}{\chi^2_{\alpha,\, n-1}}$$

2 독립 표본 평균에 대한 오차 한계

신뢰 구간에 대한 표본 크기를 결정하려면 2독립 표본 평균에 대한 구간 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰 구간 > 2독립 표본 평균에 대한 오차 한계**를 선택합니다. 변동 가정, 표본 크기, 유의성 및 구간에 대한 오차 한계 간의 균형을 탐색합니다.

2 독립 표본 평균에 대한 구간 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

구간 유형

한계 단측 구간을 지정합니다 (상한 또는 하한).

구간 양측 구간을 지정합니다 .

고정 모수

알파 신뢰 수준 1 - 알파를 지정합니다. 유의 수준 기본값은 95% 신뢰 구간에 대해 0.05 입니다.

그룹 1 표준편차 (σ_1) 그룹 중 하나인 그룹 1 의 가정된 표준편차를 지정합니다 .

그룹 2 표준편차 ($\sigma 2$) 두 번째 그룹인 그룹 2 의 가정된 표준편차를 지정합니다 .

모표준편차 알려짐 알려진 모집단 표준편차를 기반으로 계산하도록 지정합니다 .

구간 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

그룹 1 표본 크기 실험에서 그룹 1 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 1 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

그룹 2 표본 크기 실험에서 그룹 2 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 2 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

총 표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 오차 한계 곡선은 총 표본 크기를 기반으로 합니다 .

팁 : 총 표본 크기를 변경하지 않고 그대로 둔 채로 한쪽 그룹의 표본 크기를 조정하면 나머지 다른 그룹의 표본 크기에 영향을 주게 됩니다 . 총 표본 크기를 조정하면 두 그룹이 그에 비례하여 조정되며 그렇지 않을 경우 한쪽 그룹이 잠기게 됩니다 . 한쪽 그룹 표본 크기만 잠글 수 있습니다 .

오차 한계 구간 너비의 1/2 을 지정합니다 . 다른 모든 모수가 고정된 경우 표본 크기가 커질수록 오차 한계는 감소합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

2 독립 표본 구간 탐색기에 대한 통계 상세 정보

모집단 평균의 차이를 알아내기 위한 구간 계산은 σ_1과 σ_2를 알고 있는 경우에는 표준 정규를 기반으로 하고 알 수 없는 경우에는 t 분포를 기반으로 합니다.

MOE(오차 한계)는 다음과 같이 계산됩니다.

$$\text{MOE} = \begin{cases} z_{1-\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}} & \text{단, } \sigma_1 \text{및} \sigma_2 \text{ 알 수 있음} \\ t_{1-\alpha/2,\,n-1}\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}} & \text{단, } \sigma_1 \text{및} \sigma_2 \text{ 알 수 없음} \end{cases}$$

σ_1과 σ_2를 알 수 없는 경우 한계는 다음과 같이 계산됩니다.

$$\text{한계} = \pm t_{1-\alpha,\,df}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

다음은 각 요소에 대한 설명입니다.

$$df = \begin{cases} n_1+n_2-2 & \text{단, } \sigma_1 = \sigma_2 \\ \dfrac{\left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^2}{\dfrac{1}{n_1-1}\left(\dfrac{s_1^2}{n_1}\right)^2+\dfrac{1}{n_2-1}\left(\dfrac{s_2^2}{n_2}\right)^2} & \text{단, } \sigma_1 \neq \sigma_2 \end{cases}$$

σ를 알고 있는 경우 한계는 다음과 같이 계산됩니다.

$$\text{한계} = \pm z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

2 독립 표본 비율에 대한 오차 한계

두 비율의 차이, 로그 상대 위험도 또는 로그 승산의 신뢰 구간에 대한 표본 크기를 결정하려면 2 독립 표본 비율에 대한 구간 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰 구간 > 2 독립 표본 비율에 대한 오차 한계**를 선택합니다. 변동 가정, 표본 크기, 유의성 및 구간에 대한 오차 한계 간의 균형을 탐색합니다.

2 독립 표본 비율에 대한 구간 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

구간 유형

한계 단측 구간을 지정합니다 (상한 또는 하한).

구간 양측 구간을 지정합니다 .

비교 방법

비율의 차이 두 비율의 차이에 대한 신뢰 구간을 지정합니다 (p_1 -p_2).

로그 상대 위험도 로그 척도의 상대 위험도 (p_1/p_2) 에 대한 신뢰 구간을 지정합니다 .

로그 승산 로그 척도의 승산 (p_1/p_2) 에 대한 신뢰 구간을 지정합니다 .

고정 모수

알파 신뢰 수준 1 - 알파를 지정합니다. 유의 수준 기본값은 95% 신뢰 구간에 대해 0.05 입니다.

구간 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

그룹 1 비율 그룹 1 의 가정된 비율을 지정합니다 .

그룹 2 비율 그룹 2 의 가정된 비율을 지정합니다 .

그룹 1 표본 크기 실험에서 그룹 1 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 1 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

그룹 2 표본 크기 실험에서 그룹 2 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 그룹 2 표본 크기를 잠그려면 " 잠금 " 을 선택합니다 .

총 표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다 . 오차 한계 곡선은 총 표본 크기를 기반으로 합니다 .

팁 : 총 표본 크기를 변경하지 않고 그대로 둔 채로 한쪽 그룹의 표본 크기를 조정하면 나머지 다른 그룹의 표본 크기에 영향을 주게 됩니다 . 총 표본 크기를 조정하면 두 그룹이 그에 비례하여 조정되며 그렇지 않을 경우 한쪽 그룹이 잠기게 됩니다 . 한쪽 그룹 표본 크기만 잠글 수 있습니다 .

오차 한계 구간 너비의 1/2 을 지정합니다 . 다른 모든 모수가 고정된 경우 표본 크기가 커질수록 오차 한계는 감소합니다 .

한계 크기 구간의 한계를 지정합니다 . 다른 모든 모수가 고정된 경우 표본 크기가 커질수록 오차 한계는 감소합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

2 독립 비율 구간 탐색기에 대한 통계 상세 정보

각 구간 유형에 대한 계산은 정규 근사를 기반으로 합니다.

비율의 차이

비율의 차이에 대한 구간의 경우 MOE(오차 한계)는 다음과 같이 계산됩니다.

$$\text{MOE} = z_{1-\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$$

한계는 다음과 같이 계산됩니다.

$$\text{한계} = \pm z_{1-\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$$

로그 상대 위험도

상대 위험도 로그의 경우 MOE(오차 한계)는 다음과 같이 계산됩니다.

$$\text{MOE} = z_{1-\alpha/2}\sqrt{\frac{(1-p_1)}{n_1 p_1}+\frac{(1-p_2)}{n_2 p_2}}$$

한계는 다음과 같이 계산됩니다.

$$\text{한계} = \pm z_{1-\alpha}\sqrt{\frac{(1-p_1)}{n_1 p_1}+\frac{(1-p_2)}{n_2 p_2}}$$

로그 승산

승산비 로그의 경우 오차 한계는 다음과 같이 계산됩니다.

$$\text{MOE} = z_{1-\alpha/2}\sqrt{\frac{1}{n_1 p_1} + \frac{1}{n_1(1-p_1)} + \frac{1}{n_2 p_2} + \frac{1}{n_2(1-p_2)}}$$

한계는 다음과 같이 계산됩니다.

$$\text{한계} = \pm z_{1-\alpha}\sqrt{\frac{1}{n_1 p_1} + \frac{1}{n_1(1-p_1)} + \frac{1}{n_2 p_2} + \frac{1}{n_2(1-p_2)}}$$

2 독립 표본 분산에 대한 오차 한계

신뢰 구간에 대한 표본 크기를 결정하려면 2독립 표본 분산에 대한 구간 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰 구간 > 2독립 표본 분산에 대한 오차 한계**를 선택합니다. 표본 크기, 유의성 및 구간에 대한 오차 한계 간의 균형을 탐색합니다.

2 독립 표본 분산에 대한 구간 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

구간 유형

하한 단측 하한 구간을 지정합니다.

상한 단측 상한 구간을 지정합니다.

구간 양측 구간을 지정합니다.

고정 모수

알파 신뢰 수준 1 - 알파를 지정합니다. 유의 수준 기본값은 95% 신뢰 구간에 대해 0.05 입니다.

구간 모수

그룹 1 표본 크기 실험에서 그룹 1 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다. 그룹 1 표본 크기를 잠그려면 " 잠금 " 을 선택합니다.

그룹 2 표본 크기 실험에서 그룹 2 에 필요한 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다. 그룹 2 표본 크기를 잠그려면 " 잠금 " 을 선택합니다.

총 표본 크기 실험에 필요한 총 관측값 (런 , 실험 단위 또는 표본) 수를 지정합니다. 오차 한계 곡선은 총 표본 크기를 기반으로 합니다.

팁 : 총 표본 크기를 변경하지 않고 그대로 둔 채로 한쪽 그룹의 표본 크기를 조정하면 나머지 다른 그룹의 표본 크기에 영향을 주게 됩니다 . 총 표본 크기를 조정하면 두 그룹이 그에 비례하여 조정되며 그렇지 않을 경우 한쪽 그룹이 잠기게 됩니다 . 한쪽 그룹 표본 크기만 잠글 수 있습니다 .

구간 너비 구간의 전체 너비를 지정합니다 . 다른 모든 모수가 고정되면 표본 크기가 커질수록 구간 너비는 줄어듭니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

2 독립 분산 비율 구간 탐색기에 대한 통계 상세 정보

두 모집단 분산의 비율을 알아내기 위한 구간 계산은 *F*-분포를 기반으로 합니다. MOE(오차 한계) 는 다음과 같이 계산됩니다.

$$\text{MOE} = F_{1-\alpha/2,\, n_1-1,\, n_2-1} - F_{\alpha/2,\, n_1-1,\, n_2-1}$$

하한은 다음과 같이 계산됩니다.

$$\text{하한} = F_{\alpha,\, n_1-1,\, n_2-1}$$

상한은 다음과 같이 계산됩니다.

$$\text{상한} = F_{1-\alpha,\, n_1-1,\, n_2-1}$$

신뢰도 데모 계산기

지정된 기간 동안 지정된 단위 수를 검정하려면 " 신뢰도 데모 " 계산기를 사용합니다 . *k* 개 미만의 단위에서 고장이 발생하면 데모를 통과하고 , 제품 신뢰도가 신뢰도 표준을 충족하거나 초과한다는 결론을 내릴 수 있습니다 .

신뢰도 데모의 경우 **DOE > 표본 크기 탐색기 > 신뢰도**를 선택합니다.

- " 모수 신뢰도 데모 "(724 페이지)
- " 비모수 신뢰도 데모 "(725 페이지)

모수 신뢰도 데모

신뢰도 데모를 위해 검정할 단위 수를 결정하려면 모수 신뢰도에 대한 데모 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰도 > 모수 신뢰도 데모**를 선택합니다. 시간, 유의성, 고장 수 및 검정할 단위 수 간의 균형을 탐색합니다.

모수 신뢰도에 대한 데모 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

분포 가정된 고장 시간 분포를 지정합니다 . 사용 가능한 분포 : Weibull 및 로그 정규 . 이러한 분산에 대한 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

데모 시간 검정 대상 항목이 의도한 신뢰도로 생존해야 하는 최소 시간을 지정합니다 .

데모 신뢰도 검정 대상 항목이 신뢰도 표준의 정의된 데모 시간까지 생존할 확률을 지정합니다 .

척도 모수 고장 시간 분포의 척도 모수를 지정합니다 .

참고 : " 분포 " 가 "Weibull" 로 지정되면 " 척도 " 필드가 "Weibull β" 로 표시됩니다 .

데모 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

허용된 최대 고장 수 성공적인 검정 데모에 허용되는 최대 고장 수를 지정합니다 .

표본 크기 데모에 필요한 검정의 총 단위 수를 지정합니다 .

검정 시간 데모 검정 시간을 지정합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다 .

모수 신뢰도 데모 탐색기에 대한 통계 상세 정보

모수 신뢰도 데모 검정의 경우 제품이 신뢰도 R_d로 데모 시간 T_d를 달성함을 데모하는 데 필요한 검정 시간은 가정된 수명 분포를 사용하여 계산된 성공 확률을 가진 이항 분포를 사용하여 계산됩니다.

Weibull 분포에 대한 검정 시간은 다음과 같이 계산됩니다.

$$\text{검정 시간} = T_d\left(\frac{-\log(P)}{\log(R_d)}\right)^{1/\beta}$$

로그 정규 분포에 대한 검정 시간은 다음과 같이 계산됩니다.

$$\text{검정 시간} = \exp\left(\frac{z_{1-e^{-\log P}} - z_{1-R_d}}{\sigma}\right)$$

다음은 각 요소에 대한 설명입니다.

$$\log P = -\log\left(\frac{Sf}{Sf+1}\right)$$

$$Sf = \frac{n-c}{c+1} f_{\alpha,\, 2(n-c),\, 2(c+1)}$$

c는 데모 동안 허용되는 최대 고장 수입니다.

비모수 신뢰도 데모

가정된 고장 비율 분포가 없을 경우 신뢰도 데모를 위해 검정할 단위 수를 결정하려면 비모수 신뢰도에 대한 데모 탐색기를 사용합니다. **DOE > 표본 크기 탐색기 > 신뢰도 > 비모수 신뢰도 데모**를 선택합니다. 시간, 유의성, 고장 수 및 검정할 단위 수 간의 균형을 탐색합니다.

비모수 신뢰도에 대한 데모 탐색기의 옵션

라디오 버튼, 텍스트 상자 및 슬라이더를 사용하여 연구 가정을 설정하고 표본 크기를 탐색합니다. 설정을 변경하면 그에 따라 곡선이 업데이트됩니다. 또는 곡선 위의 십자 기호를 드래그하거나 축 텍스트 상자의 값을 조정하여 설정을 변경합니다.

고정 모수

알파 제 1 종 오류의 확률 , 즉 귀무가설이 참일 때 귀무가설을 기각할 확률을 지정합니다 . 일반적으로 검정의 유의 수준이라고 합니다 . 기본 유의 수준은 0.05 입니다 .

데모 모수 상호 연관되어 있으며 변경 시 업데이트되는 모수입니다 .

허용된 최대 고장 수 성공적인 검정 데모에 허용되는 최대 고장 수를 지정합니다 .

표본 크기 데모에 필요한 검정의 총 단위 수를 지정합니다 .

데모 신뢰도 검정 대상 항목이 신뢰도 표준의 정의된 데모 시간까지 생존할 확률을 지정합니다 .

설정 저장 현재 설정을 " 저장된 설정 " 테이블에 저장합니다 . 이렇게 하면 일련의 대체 연구 계획을 저장할 수 있습니다 . 자세한 내용은 " 표본 크기 탐색기의 저장된 설정 "(726 페이지) 에서 확인하십시오 .

데이터 수집 테이블 생성 데이터 수집에 사용할 수 있는 새 데이터 테이블을 생성합니다.

비모수 신뢰도 데모 탐색기에 대한 통계 상세 정보

비모수 신뢰도 데모의 경우 데모 신뢰도는 이항 분포를 기반으로 계산됩니다. 이항 cdf의 α분위수는 F-분포의 α분위수로 변환함으로써 계산될 수 있습니다. 자세한 내용은 Jowett(1963)에서 확인하십시오.

신뢰도는 다음과 같이 계산됩니다.

$$R = \frac{Sf}{Sf+1}$$

다음은 각 요소에 대한 설명입니다.

$$Sf = \frac{n-c}{c+1} f_{\alpha,\, 2(n-c),\, 2(c+1)}$$

c는 데모 동안 허용되는 최대 고장 수입니다.

표본 크기 탐색기의 저장된 설정

"저장된 설정" 보고서는 연구 시나리오를 비교할 수 있도록 저장된 설정에 대한 테이블을 제공합니다. **설정 저장** 버튼을 클릭할 때마다 이 테이블에 새 행이 추가됩니다. 테이블에 있는 행을 클릭하면 탐색기가 해당 설정으로 재설정됩니다. 테이블에 추가로 모수를 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열**을 선택합니다.

"저장된 설정"의 빨간색 삼각형에는 다음 옵션이 포함되어 있습니다.

선택한 행 지우기 선택된 행을 " 저장된 설정 " 테이블에서 제거합니다.

모두 지우기 모든 행을 " 저장된 설정 " 테이블에서 지웁니다.

테이블 생성 숨겨진 열을 포함하여 " 저장된 설정 " 테이블의 모든 열을 포함하는 새로운 데이터 테이블을 생성합니다.

부록 A

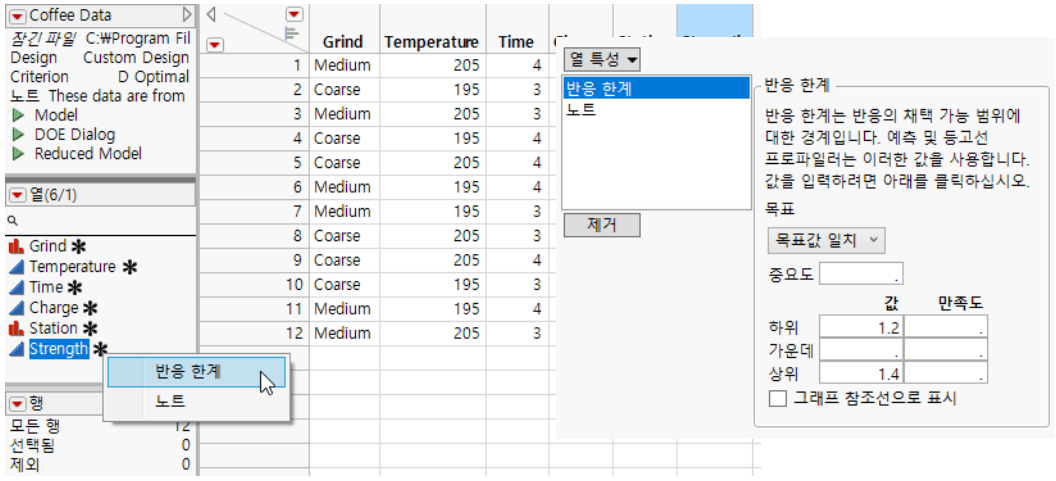
열 특성

DOE 에서 할당하는 열 특성 이해

DOE 플랫폼을 사용하여 설계를 구성할 경우 결과 설계를 포함하는 데이터 테이블에 열 특성이 저장됩니다. 이 부록에서는 DOE 플랫폼으로 구성한 설계에 저장되는 열 특성에 대해서만 설명합니다. DOE 플랫폼에서 각 열 특성을 할당하고 사용하는 방법을 예를 통해 보여 줍니다. DOE 플랫폼에서 할당하지 않는 열 특성에 대한 설명은 JMP 사용의 에서 확인하십시오.

이 부록에 설명된 일부 열 특성은 일반적인 모델링 상황에서 유용합니다. 특성을 보다 일반적으로 사용하려는 경우 특성을 직접 지정할 수 있습니다. 이 기능은 설계가 DOE 플랫폼에서 생성되지 않은 경우에 특히 유용합니다. 이 부록의 일부 예에서는 열 특성을 직접 추가하는 상황을 보여 줍니다.

그림 A.1 열 특성 별표 및 열 정보 창



목차

열 특성 추가 및 보기 729
반응 한계 730
반응 한계의 예 731
반응 한계 편집 734
감지 한계 737
함수 반응 737
설계 역할 737
설계 역할의 예 738
코딩 739
하한값과 상한값 740
코딩 열 특성 및 중심화 다항식 741
코딩의 예 741
코딩 할당 744
혼합 747
혼합의 예 749
요인 변경 752
요인 변경의 예 753
값 순서 755
값 순서의 예 756
값 순서 할당 758
값 라벨 759
값 라벨의 예 759
블록당 런 수 761
블록당 런 수의 예 761
제약 조건 상태 762
제약 조건 상태의 예 762

열 특성 추가 및 보기

DOE 플랫폼은 설계 테이블을 구성할 때 특정 열 특성을 자동으로 테이블에 저장합니다. 그러나 설계된 실험과 관련된 일부 열 특성은 일반적인 모델링 상황에서 유용합니다. DOE 플랫폼을 사용하여 생성하지 않은 데이터 테이블에서 이러한 열 특성을 사용하려면 특성을 직접 추가하면 됩니다.

열 특성 추가

하나 이상의 열에 열 특성을 할당하려면 다음을 수행하십시오 .

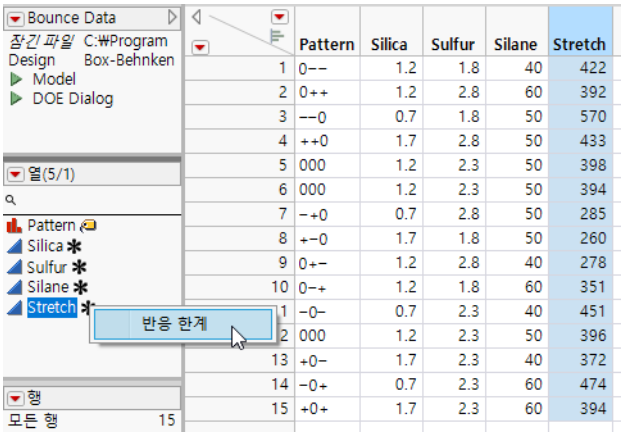
1. 특성을 할당할 열을 하나 이상 선택합니다 .
2. 다음 중 하나를 수행합니다 .
   - 머리글 영역을 마우스 오른쪽 버튼으로 클릭하고 **열 특성**을 선택한 후 특성을 선택합니다 .
   - 머리글 영역을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 후 " 열 특성 " 메뉴에서 특성을 선택합니다 .
   - **열 > 열 정보**를 선택한 후 " 열 특성 " 메뉴에서 특성을 선택합니다 .
3. 표시된 열 특성 패널에서 값을 지정하고 적절한 옵션을 선택합니다 .
4. **적용**을 클릭하여 열 특성을 추가하거나 , **확인**을 클릭하여 열 특성을 추가하고 열 특성 창을 닫습니다 .

팁 : 열에 이미 포함된 특성을 다른 열에 적용하려는 경우가 있습니다 . 해당 특성을 다른 열에 적용하려면 " 속성 표준화 " 명령을 사용합니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

열 특성 보기

특정 열에 할당된 특성을 보려면 " 열 " 패널에서 열 특성 별표 아이콘(✱)을 클릭합니다. 특성을 클릭하여 설정을 보거나 편집할 수 있습니다. 그림 A.2 에서는 Design Experiment 폴더의 Bounce Data.jmp 샘플 데이터 테이블에서 Stretch 에 할당된 열 특성을 보여 줍니다.

그림 A.2 두 개의 열 특성을 표시하는 Stretch 별표 아이콘



"열 정보" 메뉴의 "열 특성" 목록에 액세스하여 열 특성을 볼 수도 있습니다. 열 특성을 표시할 열을 하나 이상 선택하고 다음 중 하나를 수행합니다.

- 머리글 영역을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택한 후 "열 특성" 목록에서 특성을 선택합니다.
- **열 > 열 정보**를 선택한 후 "열 특성" 목록에서 특성을 선택합니다.

반응 한계

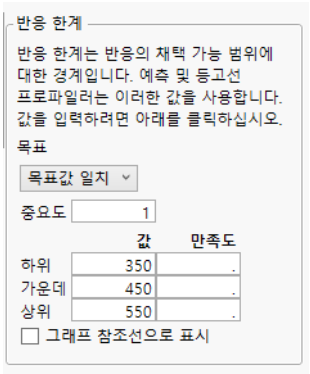
"반응 한계" 열 특성을 사용하여 다음을 지정할 수 있습니다.

- 반응의 변동 범위에 대한 경계
- 만족도 목표
- 반응의 중요도 측도
- 만족도 값

"반응 한계" 열 특성은 반응의 만족도 함수를 정의합니다. "프로파일러" 및 "등고선 프로파일러"는 만족도 함수를 사용하여 최적 요인 설정을 찾습니다. 자세한 내용은 Profilers의 에서 확인하십시오.

그림 A.3에서는 Design Experiment 폴더의 Bounce Data.jmp 샘플 데이터 테이블에서 반응(Stretch)에 대한 "열 정보" 창의 "반응 한계" 패널을 보여 줍니다.

그림 A.3 반응 한계 패널의 예



" 반응 한계 " 패널은 다음 영역으로 구성됩니다 .

목표 메뉴에서 반응 목표를 선택합니다 . " 최대화 ", " 목표값 일치 ", " 최소화 " 및 " 없음 " 중에서 선택할 수 있습니다 . JMP 는 선택한 목표에 맞도록 반응의 만족도 함수를 정의합니다 . 한계를 지정하면 해당 한계를 사용하여 만족도 함수가 정의됩니다 . 한계를 지정하지 않으면 반응 분포에서 도출된 보수적 한계 값을 기반으로 만족도 함수가 정의됩니다 . " 없음 " 을 목표로 선택하면 모든 반응 값의 만족도가 동일하다고 간주됩니다 . 자세한 내용은 " 사용자 설계 " 장의 " 반응 "(82 페이지) 에서 확인하십시오 .

중요도 반응이 여러 개인 경우 전체 만족도 함수를 계산할 때 각 반응에 대한 상대적 가중치를 입력합니다 . 중요도 값은 모든 양수가 될 수 있습니다 . 중요도 값을 지정하지 않으면 주어진 분석에서 모든 반응의 중요도 값이 동일한 것으로 간주됩니다 . 반응이 하나뿐이면 중요도가 1 로 지정됩니다 .

값 반응의 하한 및 상한과 중간 값을 지정합니다 . JMP 는 이러한 값을 사용하여 반응의 만족도 함수를 구성합니다 . 한계를 지정하지 않으면 보수적 한계 값을 기반으로 만족도 함수가 정의됩니다 . 목표가 " 목표값 일치 " 이고 " 중간 " 값을 지정하지 않으면 목표값이 하한과 상한의 중간점으로 정의됩니다 .

만족도 " 하한 ", " 중간 " 및 " 상한 " 값의 만족도를 반영하는 값을 지정합니다 . 만족도 값은 0 에서 1 사이여야 합니다 . 만족도 값을 지정하지 않으면 선택된 목표에 따라 JMP 에서 값을 할당합니다 .

그래프 참조선으로 표시 " 실제값 대 예측값 그림 " 과 " 예측 프로파일러 " 에서 " 하한 ", " 중간 " 및 " 상한 " 값에 대한 가로 참조선을 표시합니다 . 이 옵션은 한계가 지정된 경우에만 적용됩니다 .

반응 한계의 예

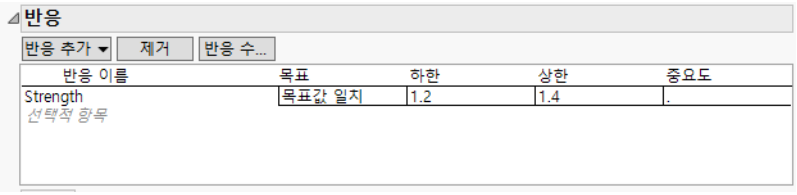
Design Experiment 폴더의 Coffee Data.jmp 샘플 데이터 테이블에는 커피 강도(Strength)를 최적화하기 위해 수행된 실험 결과가 포함되어 있습니다. 실험 설계 및 분석에 대한 자세한 설명은 "DOE로 시작" 장의 "커피 강도 실험"(42페이지)에서 확인하십시오.

가장 적합한 값으로 간주되는 목표 강도 1.3 의 커피를 추출할 수 있는 요인 설정을 찾는 것이 목표입니다 . 1.2 보다 작은 값과 1.4 보다 큰 값은 허용되지 않습니다 . 1.2 ~ 1.4 사이의 값은 1.3 에서 멀어질수록 만족도가 감소합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Coffee Data.jmp 를 엽니다 .
2. 테이블 패널에서 **DOE Dialog** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   이 스크립트는 Coffee Data.jmp 에서 실험 설계를 생성하는 데 사용한 " 사용자 설계 " 대화상자를 다시 생성합니다 .
3. " 반응 " 개요를 엽니다 .

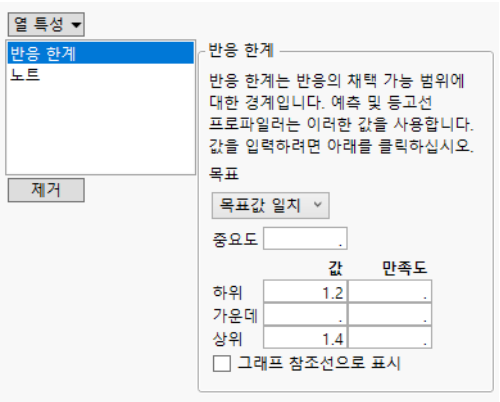
그림 A.4 사용자 설계 창의 반응 개요



이 실험을 설계할 때 반응 목표를 " 목표값 일치 " 로 지정하고 하한과 상한을 각각 1.2 와 1.4 로 지정했습니다 . 중요도 값은 지정하지 않았습니다 . 반응이 하나뿐이므로 기본적으로 1 로 지정됩니다 . 설계 테이블을 구성할 때 " 반응 한계 " 열 특성이 Strength 에 할당되었습니다 .

4. " 사용자 설계 " 창을 닫습니다 .
5. Coffee Data.jmp 샘플 데이터 테이블에서 Strength 열을 선택한 후 **열 > 열 정보**를 선택합니다 .
6. " 열 특성 " 목록에서 " 반응 한계 " 를 선택합니다 .

그림 A.5 Strength 에 대한 반응 한계 열 특성



다음을 확인할 수 있습니다 .

– " 목표 " 가 **목표값 일치**로 설정되어 있습니다 .

- " 중요도 " 값이 없습니다 . 중요도 값이 없으면 주어진 분석에서 모든 반응의 중요도 값이 동일한 것으로 간주됩니다 . 따라서 JMP 는 Strength 에 중요도 값 1 을 할당합니다 .
- 하한은 1.2 입니다 .
- 상한은 1.4 입니다 .
- 중간 값은 지정되지 않았습니다 .

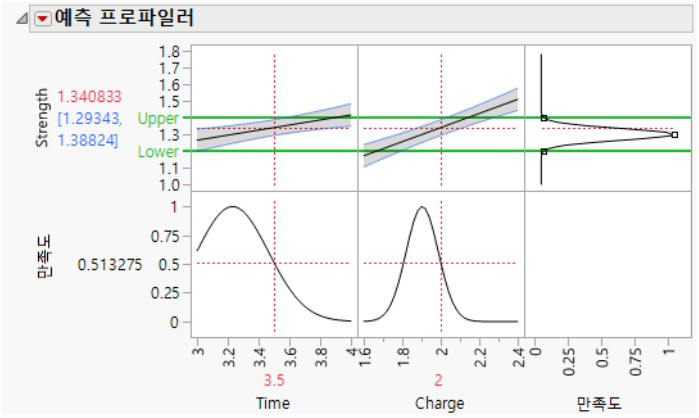
  중간 값이 지정되지 않았으므로 목표값이 하한과 상한의 중간점인 1.3 으로 정의됩니다 .
- " 만족도 " 값이 지정되지 않았습니다 .

7. **그래프 참조선으로 표시** 옵션을 선택합니다 .

   이 옵션은 " 실제값 대 예측값 그림 " 과 " 예측 프로파일러 " 에서 " 하한 ", " 중간 " 및 " 상한 " 값에 대한 가로 참조선을 표시합니다 .
8. **확인**을 클릭합니다 .
9. Coffee Data.jmp 데이터 테이블에서 **Reduced Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
10. **실행**을 클릭합니다 .

    " 예측 프로파일러 " 가 보고서 맨 아래에 나타납니다 .

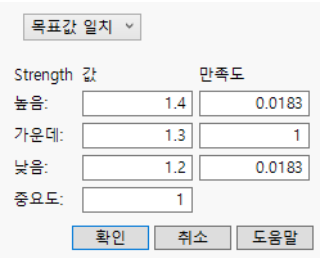
그림 A.6 Strength 의 만족도 함수를 보여 주는 프로파일러



Strength 의 만족도 함수는 " 만족도 " 위의 오른쪽 그림에 나옵니다 . 이 그림은 데이터 테이블에 Strength 에 대한 " 반응 한계 " 열 특성이 포함되어 있기 때문에 나타납니다 . Strength 의 상한 및 하한에 대한 참조선도 " 예측 프로파일러 " 에 표시됩니다 .

11. Ctrl 키를 누르고 만족도에 대한 Strength 그림을 클릭합니다 .

그림 A.7 Strength 의 반응 목표 창



다음을 확인할 수 있습니다 .

- " 중간 " 값은 " 반응 한계 " 열 특성에 지정한 " 상한 " 과 " 하한 " 의 중간점으로 결정됩니다 .
- " 목표 " 가 " 목표값 일치 " 로 설정되었으므로 " 중간 " 값의 " 만족도 " 는 1 로 설정됩니다 .
- " 상한 " 및 " 하한 " 값의 " 만족도 " 는 매우 작은 값인 0.0183 으로 설정됩니다 .
- 그림 A.6 의 만족도 그림은그림 A.7 에 표시된 만족도 값이 JMP 에서 어떻게 사용되는지 보여 줍니다 . 기본적으로 Strength 의 만족도 함수는 " 하한 " 및 " 상한 " 값을 벗어나면 0 이고 , Strength 가 목표값 1.3 에 가까워질수록 서서히 1 로 증가합니다 . 모형에 반응이 하나뿐이므로 " 중요도 " 값은 1 로 설정됩니다 .

12. **취소**를 클릭하여 창을 닫습니다 .
13. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

    Time 및 Charge 설정이 업데이트되어 Strength 의 만족도 함수를 최대화하는 요인 설정을 보여 줍니다 . 그러나 다른 많은 설정도 만족도 함수를 최대화합니다 . 만족도 함수를 최대화하는 다른 설정을 확인하는 방법에 대한 자세한 내용은 Profilers 의 을 참조하십시오 .

반응 한계 편집

Vinyl Data.jmp 샘플 데이터 테이블에는 반응 thickness(두께) 에 " 반응 한계 " 열 특성이 이미 할당되어 있습니다 . 이 특성의 목표는 thickness 를 최대화하는 것입니다 . thickness 를 최대화하는 대신 thickness 목표값을 8.5 로 지정하고 비닐 thickness 가 6 ~ 10 사이의 값이 되게 하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Vinyl Data.jmp 를 엽니다 .
2. thickness 열을 선택하고 **열 > 열 정보**를 선택합니다 .

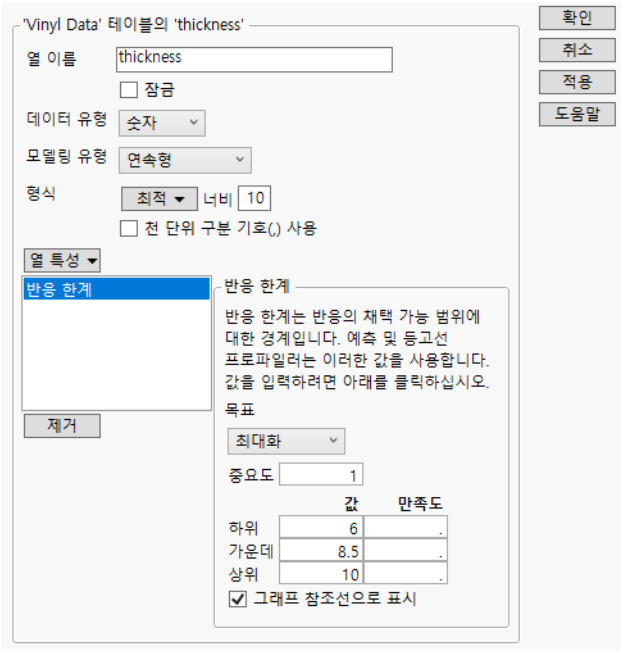
   " 열 특성 " 목록에 " 반응 한계 " 특성이 표시됩니다 . thickness 에 이 특성만 할당되었습니다 . 목록 오른쪽에는 " 반응 한계 " 패널이 있습니다 .
3. **최대화**를 클릭하고 **목표값 일치**를 선택합니다 .
4. **중요도** 값에 1 을 입력합니다 .
5. " 값 " 에서 " 하한 " 에 6, " 중간 " 에 8.5, " 상한 " 에 10 을 각각 입력합니다 .

   이 예는 비대칭 반응 한계를 보여 줍니다 . thickness 값은 최소 6 에서 최대 10 까지 허용됩니다 . 그러나 thickness 목표값은 8.5 입니다 .

6. **그래프 참조선으로 표시**를 선택합니다 .

   이 옵션은 " 실제값 대 예측값 그림 " 과 " 예측 프로파일러 " 에서 " 하한 ", " 중간 " 및 " 상한 " 값에 대한 가로 참조선을 표시합니다 .

그림 A.8 완료된 반응 한계 패널

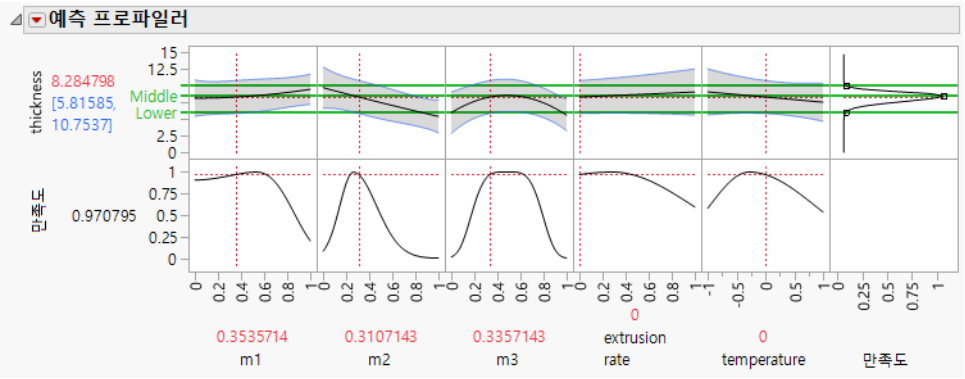


7. **확인**을 클릭합니다 .
8. Vinyl Data.jmp 데이터 테이블에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   m1, m2 및 m3 은 혼합 요인입니다 . 또한 Whole Plots 라는 랜덤 요인이 설계에 포함되어 있습니다 . 이 때문에 기본 " 방법 " 이 "REML(권장)" 입니다 .
9. **실행**을 클릭합니다 .
10. " 반응 thickness" 의 빨간색 삼각형을 클릭하고 **행 진단 > 실제값 대 예측값 그림**을 선택합니다 .

    하한 , 중간 및 상한에 대한 참조선이 " 실제값 대 예측값 그림 " 에 나타납니다 .
11. " 반응 thickness" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다 .

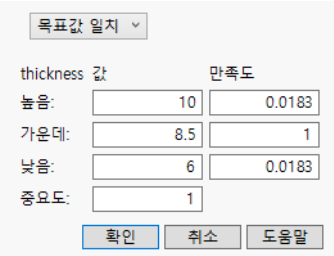
그림 A.9 비대칭 만족도 함수를 보여 주는 예측 프로파일러



" 만족도 " 위의 오른쪽 그림은 JMP 가 thickness 에 대해 구성한 만족도 함수를 보여 줍니다 . 만족도는 " 중간 " 한계 8.5 에서 1 입니다 . 기본적으로 6 미만이거나 10 을 초과하는 thickness 값의 만족도는 0 입니다 .

12. Ctrl 키를 누르고 만족도에 대한 thickness 그림을 클릭합니다 .

그림 A.10 Thickness 의 반응 목표 창



이 창에는 " 상한 ", " 중간 " 및 " 하한 " 값 설정이 표시됩니다 . 또한 실험 목표인 " 목표값 일치 " 에 따라 JMP 에서 할당하는 " 만족도 " 값도 표시됩니다 . 그림 A.9 에 표시된 만족도 함수는그림 A.10 에서 " 상한 ", " 중간 " 및 " 하한 " 값의 만족도 설정과 일치하는 연속 곡선입니다 . 다른 값에 대해 만족도 함수는 선택한 목표에 맞는 만족도를 할당합니다 .

13. **취소**를 클릭합니다 .
14. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

    요인 설정이 업데이트되어 thickness 의 만족도 함수를 최대화하는 값이 표시됩니다 . 만족도 함수를 최대화하는 다른 설정도 많습니다 . 이러한 최적 설정에서 예측 반응은 8.5 입니다 . 앞에서 한계값을 6 과 10 으로 지정하고 8.5 를 목표값으로 설정했습니다 .
15. 변경 사항을 저장하지 않고 Vinyl Data.jmp 샘플 데이터 테이블을 닫습니다 .

JMP PRO 감지 한계

"감지 한계" 열 특성은 반응을 측정할 수 있는 경계를 정의하며 이 경계를 벗어나면 반응을 측정할 수 없습니다. 이러한 한계를 사용하여 일반화 회귀 플랫폼에서 중도절단 반응을 지정할 수 있습니다.

JMP PRO 함수 반응

"FDE X" 열 특성은 열을 함수 반응으로 지정합니다. DOE 요인을 보조 요인으로 사용하는 함수 데이터 탐색기 플랫폼에서는 FDE 열 특성을 가진 여러 개의 Y 열을 사용할 수 있습니다.

설계 역할

일반 모형 및 설계된 실험의 요인은 다양한 방식으로 동작할 수 있습니다. JMP 는 설계 역할 열 특성을 사용하여 이러한 동작을 설명합니다. 다음과 같은 설계 역할을 사용할 수 있습니다.

- 연속형
- 이산 수치형
- 범주형
- 블록화
- 공변량
- 혼합
- 상수
- 제어되지 않음
- 랜덤 블록
- 신호
- 잡음

많은 JMP DOE 플랫폼에서 다양한 설계 역할을 사용하여 요인을 지정할 수 있습니다. 일부 플랫폼의 경우 설계 요구 사항으로 인해 JMP 에서 자동으로 요인을 정의합니다. 예를 들어 매우 변경하기 힘든 요인과 변경하기 힘든 요인을 지정하면 JMP 에서 주구 및 하위구 요인을 생성합니다. 다양한 설계 역할이 나타날 수 있는 플랫폼에서 JMP 가 설계 테이블을 생성하면 각 요인에 " 설계 역할 " 열 특성이 할당됩니다.

- " 랜덤 블록 ", " 신호 " 및 " 잡음 " 이외의 설계 역할에 대한 설명은 " 사용자 설계 " 장의 " 요인 유형 "(85 페이지) 에서 확인하십시오.
- " 랜덤 블록 " 설계 역할에 대한 설명은 " 사용자 설계 " 장의 " 변경 및 랜덤 블록 "(87 페이지) 에서 확인하십시오.

- " 신호 " 및 " 잡음 " 설계 역할에 대한 설명은 "Taguchi 설계 " 장의 " 요인 "(398 페이지) 에서 확인하십시오 .

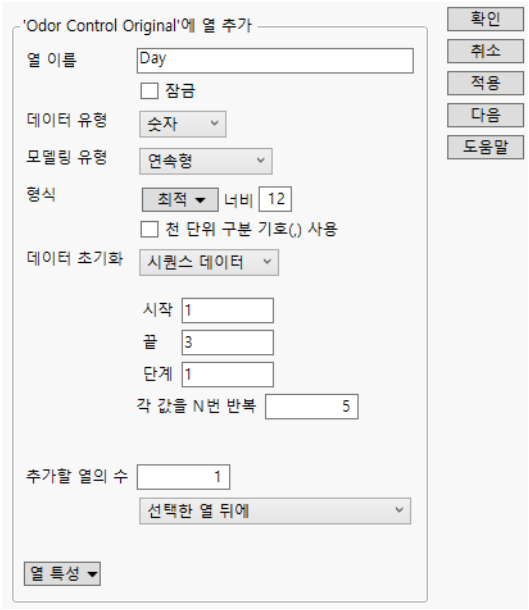
설계 역할의 예

Odor Control Original.jmp 샘플 데이터 테이블의 실험은 세 가지 요인이 odor(냄새) 에 미치는 효과를 연구합니다 . 15 회 런 실험을 설계하고 수행했습니다 . 그러나 결과가 보고되었을 때 실험이 3 일 동안 수행되었음을 알게 되었습니다 . 처음 5 회 런은 첫째 날에 , 두 번째 5 회 런은 둘째 날에 그리고 나머지 5 회 런은 셋째 날에 각각 수행되었습니다 .

온도와 습도의 변화가 반응에 영향을 줄 수 있으므로 Day 를 랜덤 블록 요인으로 포함하려고 합니다 . 설계 테이블에 Day 열을 추가하는 것은 쉽습니다 . 그러나 설계 평가 플랫폼을 사용하여 예기치 않은 블록이 포함된 설계를 원래 설계와 비교하려고 합니다 . 또한 설계를 확대해야 할 경우를 위해 설계 확대 플랫폼도 사용하기를 원합니다 . 설계 평가 및 설계 확대 플랫폼을 사용하려면 새 Day 열에 " 설계 역할 " 열 특성을 추가해야 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Odor Control Original.jmp 를 엽니다 .
2. 첫 번째 열 Run 을 선택합니다 .
3. **열 > 새 열**을 선택합니다 .
4. **열 이름** 옆에 Day 를 입력합니다 .
5. **데이터 초기화** 목록에서 **시퀀스 데이터**를 선택합니다 .
6. 다음을 입력합니다 .
   - **시작** : 1
   - **끝** : 3
   - **단계** : 1
   - **각 값을 N 번 반복** : 5
7. **추가할 열의 수** 옆에 1 을 입력합니다 .

그림 A.11 완료된 새 열 창



8. **확인**을 클릭합니다.

   Day 열이 데이터 테이블의 두 번째 열로 추가됩니다.
9. Day 열을 선택합니다.
10. **열 > 열 정보**를 선택합니다.
11. "열 특성" 목록에서 **설계 역할**을 선택합니다.
12. "설계 역할" 패널에서 **연속형**을 클릭하고 **랜덤 블록**을 선택합니다.
13. **확인**을 클릭합니다.

    "열" 패널에서 Day 옆에 별표가 표시됩니다.
14. Day 옆의 별표를 클릭하여 "설계 역할" 열 특성이 할당되었는지 확인합니다.
15. 변경 사항을 저장하지 않고 Odor Control Original.jmp 샘플 데이터 테이블을 닫습니다.

코딩

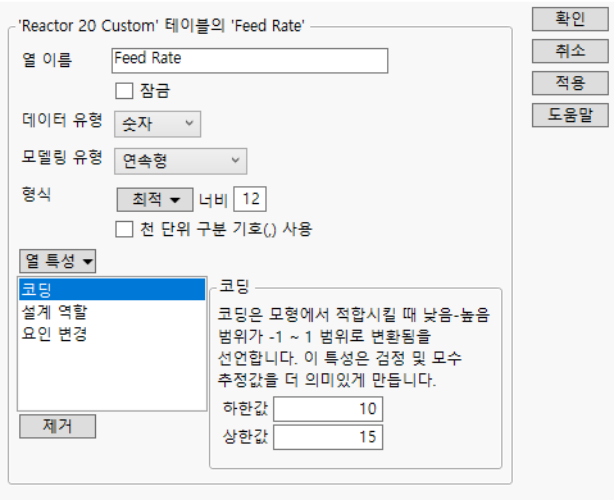
"코딩" 열 특성은 숫자 데이터 유형의 열에만 적용됩니다. 이 특성은 열의 데이터에 선형 변환을 적용합니다. "코딩" 열 특성 창에서 하한값과 상한값을 지정합니다. 원래 데이터의 하한값과 상한값이 각각 –1과 +1로 변환됩니다. JMP의 모형 적합 플랫폼에서 해당 열이 모형 효과로 입력될 때마다 이 변환된 데이터 값이 사용됩니다.

코딩 특성은 다음과 같은 이유로 유용합니다 .

- 코드화된 예측 변수를 사용하여 구한 모수 추정값은 더 쉽게 해석하고 비교할 수 있습니다 .
- 코드화된 예측 변수를 사용하면 모형에서 교호작용 항 및 고차 항과의 다중공선성을 줄일 수 있습니다 .

가속 수명 시험 설계 이외의 DOE 플랫폼에서 설계를 생성하는 경우 JMP는 숫자 데이터 유형을 사용하여 각 비혼합 요인의 "코딩" 열 특성을 정의합니다. 그림 A.12에서는 Design Experiment 폴더의 Reactor 20 Custom.jmp 샘플 데이터 테이블에 포함된 Feed Rate 열의 "코딩" 열 특성 패널을 보여 줍니다.

그림 A.12 Feed Rate 의 코딩 특성 패널



하한값과 상한값

DOE 플랫폼에서 설계 구성의 일부로 "코딩" 특성이 적용되는 경우 처음에는 요인의 최소 설정과 최대 설정이 "하한값"과 "상한값"으로 각각 지정됩니다.

"코딩" 특성을 이 특성이 포함되지 않은 열에 적용하면 JMP에서는 최소값을 "하한값"으로 삽입하고 최대값을 "상한값"으로 삽입합니다. 필요에 따라 이 값을 변경할 수 있습니다.

주의 : " 코딩 " 열 특성이 열에 할당된 후 해당 열의 값을 변경하면 값이 자동으로 업데이트되지 않습니다 ." 코딩 " 열 특성이 있는 열의 값을 변경할 경우 상한값과 하한값이 적절한지 검토하십시오 .

" 코딩 " 열 특성은 상한값과 하한값의 중간점을 빼서 열의 각 값을 중심화합니다 . 그런 다음 범위의 1/2 로 나눕니다 . H 가 상한값이고 L 이 하한값이라고 가정하면 열의 모든 X 가 다음 값으로 변환됩니다 .

$$\frac{X-(H+L)/2}{(H-L)/2}$$

각 요인에 대해 변환된 값은 중간점이 0 이고 -1 ~ +1 범위를 가집니다.

코딩 열 특성 및 중심화 다항식

"중심화 다항식" 옵션은 "모형 적합" 시작 창에서 "모형 규격"의 빨간색 삼각형 메뉴에 있습니다. 중심화 다항식을 사용하면 다항식 항과 관련된 연속형 열을 각 값의 평균을 빼는 방법으로 중심화합니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

"중심화 다항식" 옵션은 "코딩" 열 특성이 할당된 열에는 영향을 주지 않습니다. 해당 열과 관련된 다항식 항의 값은 "코딩" 특성에 지정된 대로 중심화 및 척도화됩니다. 모형의 다른 열에 "코딩" 특성이 없고 "중심화 다항식" 옵션을 선택한다고 가정해 보겠습니다. 이 경우 해당 열은 표시되는 모든 다항식 항에서 평균을 기준으로 중심화됩니다.

코딩의 예

Reactor 20 Custom.jmp 샘플 데이터 테이블에는 사용자 설계 플랫폼을 사용하여 구성한 20 회 런 설계의 데이터가 포함되어 있습니다 . 이 실험에서는 화학 공정의 반응 백분율 (Percent Reacted) 에 대한 5 개 요인의 효과를 조사합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Reactor 20 Custom.jmp 를 엽니다 .
2. 테이블 패널에서 **DOE Dialog** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
3. **요인** 개요를 엽니다 .

그림 A.13 Reactor 20 Custom.jmp 에 사용된 설계의 요인 개요

요인

요인 추가 | 제거 | N개 요인 추가 1

이름	역할	변경	값	
Feed Rate	연속형	쉬움	10	15
Catalyst	연속형	쉬움	1	2
Stir Rate	연속형	쉬움	100	120
Temperature	연속형	쉬움	140	180
Concentration	연속형	쉬움	3	6

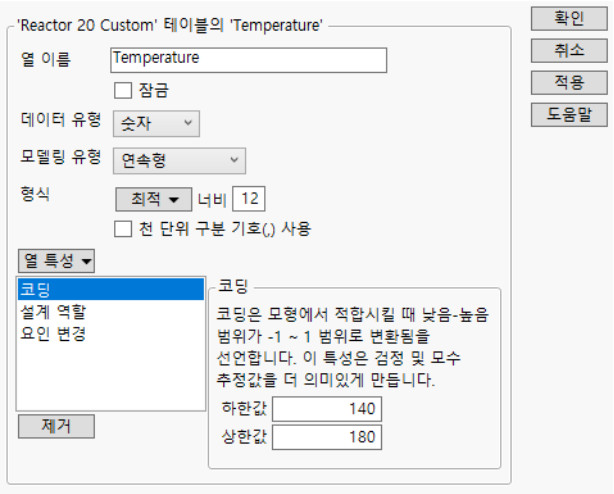
Temperature 설정의 범위는 140 ~ 180 입니다 . 설계가 생성될 때 Temperature 에 " 코딩 " 열 특성이 할당되었습니다 . 하한값이 140 으로 설정되고 상한값이 180 으로 설정되었습니다 .

4. " 사용자 설계 " 창을 닫습니다 .

5. Reactor 20 Custom.jmp 샘플 데이터 테이블의 " 열 " 패널에서 Temperature 옆의 별표를 클릭하고 **코딩**을 선택합니다 .

   열 정보 창이 나타나고 " 코딩 " 열 특성 패널이 표시됩니다 . JMP 가 설계 테이블을 구성할 때 하한값과 상한값을 지정하는 열 특성을 추가했음을 알 수 있습니다 . 실제로 이 단계를 반복하면 JMP 가 5 개 요인 모두에 대해 " 코딩 " 특성을 추가했음을 확인할 수 있습니다 .

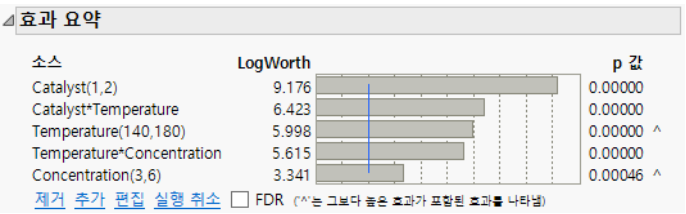
그림 A.14 Temperature 의 코딩 패널



6. **취소**를 클릭하여 열 정보 창을 닫습니다 .
7. Reactor 20 Custom.jmp 샘플 데이터 테이블에서 **Reduced Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   이 스크립트는 전체 모형을 분석한 결과 유의한 것으로 판별된 5 개 효과만 포함하는 모형을 적합시킵니다 .
8. **실행**을 클릭합니다 .

그림 A.15 축소 모형에 대한 효과 요약 보고서

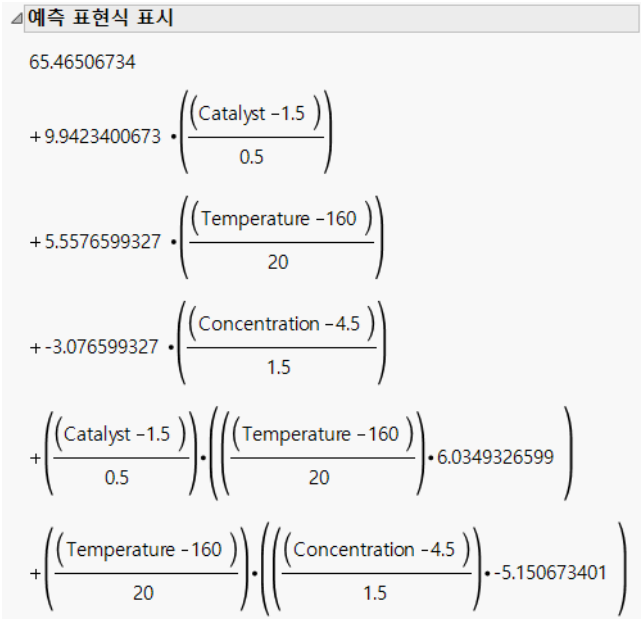


" 소스 " 목록에서 주효과 , 즉 Catalyst, Temperature 및 Concentration 오른쪽의 괄호 안에 " 코딩 " 열 특성에 사용된 상한값과 하한값이 표시됩니다 . 교호작용 효과에는 " 코딩 " 특성에 의해 설정된 범위가 표시되지 않습니다 .

팁 : Temperature 와 Concentration 의 p 값 오른쪽에 "^" 기호가 있습니다 . 이 기호는 해당 주 효과가 *p* 값이 더 작은 교호작용 효과의 성분임을 나타냅니다 . 교호작용 효과가 모형에 포함된 경우 효과 유전성 원리에 따라 모든 성분 효과도 모형에 포함되어야 합니다 . 자세한 내용은 "DOE 로 시작 " 장의 " 효과 유전성 "(61 페이지) 에서 확인하십시오 .

9. "반응 Percent Reacted" 의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다.
   " 예측 표현식 " 개요를 보고 코딩이 예측 계산식에 어떤 영향을 주는지 확인할 수 있습니다 .

그림 A.16 축소 모형에 대한 예측 표현식



" 코딩 " 열 특성에 지정된 대로 각 요인이 변환됩니다 . 예를 들어 Temperature 의 경우 다음을 확인할 수 있습니다 .

- " 코딩 " 특성의 하한값은 140 으로 설정되었습니다 . Temperature 값 140 은 -1 로 변환됩니다 .
- " 코딩 " 특성의 상한값은 180 으로 설정되었습니다 . Temperature 값 180 은 +1 로 변환됩니다 .
- 하한값과 상한값의 중간점은 160 입니다 . Temperature 값 160 은 0 으로 변환됩니다 .

변환된 값은 효과를 비교하는 데 도움이 됩니다. Catalyst 의 추정 계수는 9.942 이고 Concentration 의 추정 계수는 -3.077 입니다. 결과적으로 Percent Reacted 에 대한 Catalyst 의 예측 효과가 Percent Reacted 에 대한 Concentration 효과의 3 배가 넘습니다. 또한 이 계수를 보면 예측 Percent Reacted 는 Catalyst 가 커질수록 증가하고 Concentration 이 커질수록 감소한다는 것을 알 수 있습니다.

변환된 값은 계수를 해석하는 데 도움이 됩니다.

- 모든 요인이 중간점에 있을 때 변환된 값은 0입니다. 예측 Percent Reacted는 절편, 즉 65.465입니다.
- Catalyst 와 Concentration 이 중간점에 있을 때 Temperature 가 20 단위 증가하면 Percent Reacted 가 5.558 단위 증가합니다.
- Concentration 이 중간점에 있고 변환된 값이 0 이라고 가정해 보겠습니다.
  - Catalyst 가 중간점에 있을 때 Temperature 가 20 단위 증가하면 Percent Reacted 가 5.558 단위 증가합니다.
  - Catalyst 가 상한 설정일 때 Temperature 가 20 단위 증가하면 Percent Reacted 가 11.593(5.558 + 6.035) 단위 증가합니다.

결과적으로 교호작용 항의 계수 6.035는 Catalyst가 0.5 단위 변경될 때 예측 Percent Reacted에 대한 모형에서 기울기의 증가입니다.

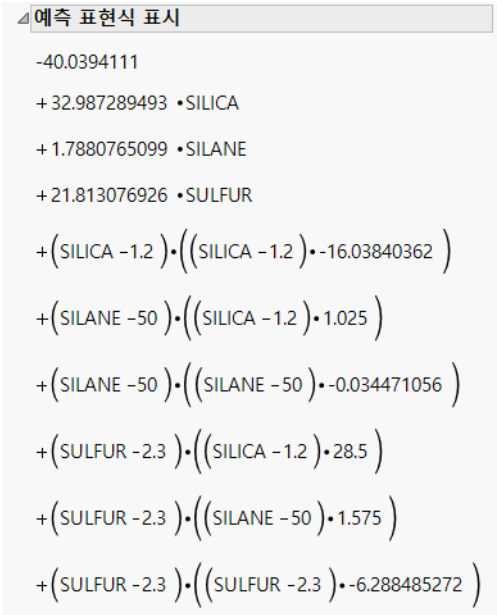
코딩 할당

Tiretread.jmp 샘플 데이터 테이블의 실험 데이터는 타이어 트레드 성능의 네 가지 측도에 대한 SILICA, SILANE 및 SULFUR 효과를 연구하는 실험 결과입니다. 이 예에서는 응답 중 하나인 ABRASION만 고려합니다.

먼저 코드화되지 않은 요인을 사용하여 모형을 적합시킵니다. 그런 다음 코딩 특성을 요인에 할당하고 모형을 다시 실행하여 의미 있는 모수 추정값을 구합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Tiretread.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. ABRASION 을 선택하고 **Y** 를 클릭합니다.
4. SILICA, SILANE, SULFUR 를 선택하고 **매크로 > 반응 표면**을 클릭합니다.
5. **대화상자 열린 채 유지**를 선택합니다.
6. **실행**을 클릭합니다.
7. " 반응 ABRASION" 의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다.

그림 A.17 코드화되지 않은 요인을 사용하는 모형의 예측 표현식



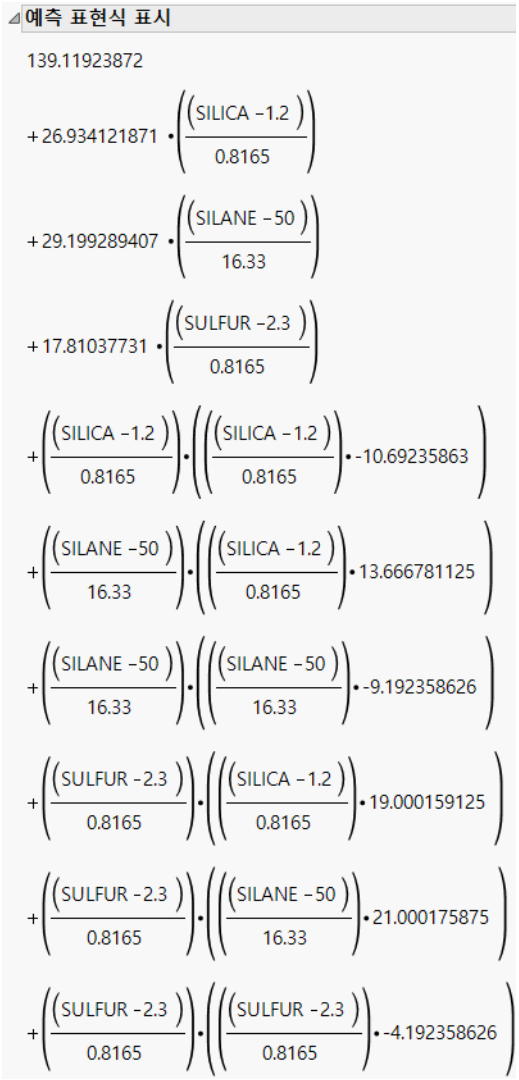
계수는 효과 크기를 비교하는 데 도움이 되지 않습니다. 계수 크기는 설정 범위에서 효과가 ABRASION 에 미치는 영향을 반영하지 않습니다. 또한 계수를 쉽게 해석할 수 없습니다. 예를 들어 SILICA 가 범위의 중간점에 있을 때 계수로 예측 반응을 해석하기가 쉽지 않습니다.

" 코딩 " 열 특성을 세 요인에 적용하여 코딩이 어떻게 계수를 더 의미 있게 만드는지 살펴봅니다.

8. Tiretread.jmp 데이터 테이블의 " 열 " 패널에서 SILICA, SILANE 및 SULFUR 를 선택합니다. " 열 " 의 빨간색 삼각형을 클릭하고 **속성 표준화**를 선택합니다.
9. " 특성 표준화 " 패널에서 **열 특성 > 코딩**을 선택합니다.
10. **확인**을 클릭합니다.

    " 열 " 패널에서 SILICA, SILANE 및 SULFUR 옆에 별표가 나타나 해당 열에 열 특성이 할당되었음을 나타냅니다.
11. " 모형 적합 " 창에서 **실행**을 클릭합니다.
12. " 반응 ABRASION" 의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다.

그림 A.18 코드화된 요인을 사용하는 모형의 예측 표현식



코드화된 요인에 대한 계수를 사용하면 효과 크기를 비교할 수 있습니다. 설계 설정 범위에서 ABRASION 에 가장 큰 영향을 주는 효과는 SILANE 입니다. SILICA 효과와 SILANE*SULFUR 교호작용의 효과도 큽니다.

코드화된 요인에 대한 계수도 쉽게 해석됩니다. 예를 들어 모든 요인이 범위의 중심에 있을 때 ABRASION 예측값은 절편, 즉 139.12 입니다.

13. 변경 사항을 저장하지 않고 Tiretread.jmp 샘플 데이터 테이블을 닫습니다.

혼합

"혼합" 열 특성은 데이터 테이블의 열이 혼합 성분을 나타내는 경우에 유용합니다. 혼합 성분의 합은 특정 상수가 되어야 합니다. 이 때문에 비혼합 요인과 차이가 있습니다. "혼합" 열 특성은 다음 두 가지 용도로 사용됩니다.

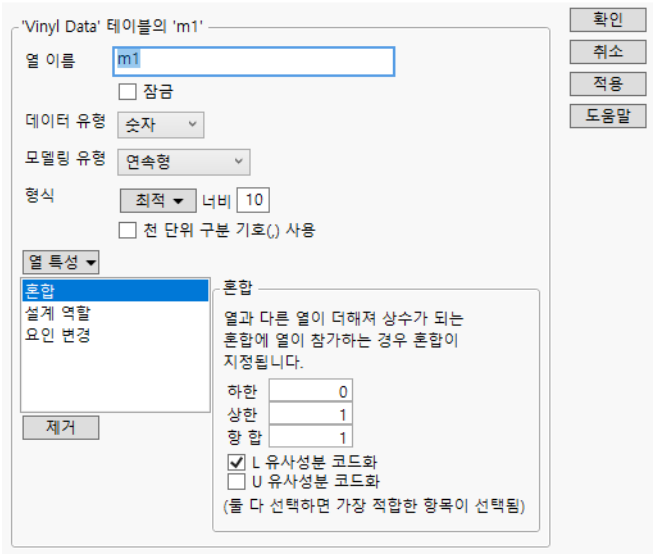
- 열을 혼합 성분으로 식별합니다.

  "분석 > 모형 적합" 창에서 "혼합" 열 특성이 할당된 열을 모형 효과로 추가하면 절편이 없는 모형이 자동으로 생성됩니다.
- 혼합 성분에 대한 코딩을 정의합니다.

  혼합 성분의 코딩은 비혼합 요인의 코딩과 다릅니다. 그러나 비혼합 요인과 마찬가지로 혼합 요인에 대한 코딩도 모수 추정값을 해석하는 데 도움이 된다는 장점이 있습니다. 자세한 내용은 "유사성분 코드화"(748 페이지)에서 확인하십시오.

그림 A.19에서는 Design Experiment 폴더의 Vinyl Data.jmp 샘플 데이터 테이블에서 m1 요인에 대한 "혼합" 열 특성 패널을 보여 줍니다.

그림 A.19 혼합 열 특성 패널



"혼합" 열 특성 패널에서 다음을 지정할 수 있습니다.

하한 유사성분 코드화에 사용되는 하한값을 지정합니다. DOE 플랫폼에서 설계 구성 중에 "혼합" 특성이 적용되는 경우 "하한"은 요인의 최소 설정으로 지정됩니다. "혼합" 특성을 해당 특성이 포함되지 않은 열에 적용하면 최소값이 "하한"으로 삽입됩니다. 필요에 따라 이 값을 변경할 수 있습니다.

상한 유사성분 코드화에 사용되는 상한값을 지정합니다. DOE 플랫폼에서 설계 구성 중에 "혼합" 특성이 적용되는 경우 "상한"은 요인의 최대 설정으로 지정됩니다. "혼합" 특성을 해

당 특성이 포함되지 않은 열에 적용하면 최대값이 " 상한 " 으로 삽입됩니다 . 필요에 따라 이 값을 변경할 수 있습니다 .

항 합 혼합 성분의 합을 지정합니다 . " 혼합 " 특성을 해당 특성이 포함되지 않은 열에 적용하면 기본값 1 이 " 항 합 " 에 삽입됩니다 .

L 유사성분 코드화 하한이 0 에 해당하도록 데이터 값을 변환합니다 .

U 유사성분 코드화 상한이 0 에 해당하도록 데이터 값을 변환합니다 .

유사성분 코드화

유사성분은 선형 변환입니다. S를 혼합 성분의 합으로 나타내고 i 열 $X_1, X_2, \ldots, X_q$에 "혼합" 열 특성이 할당되었다고 가정합니다. 이러한 열과 이 열에서 구성된 효과가 "모형 적합" 창에 효과로 입력되는 경우를 가정해 보겠습니다.

다음을 정의합니다 .

$L = \sum_{i=1}^{q} L_i$ 여기서 L_i 는 X_i 의 하한입니다 .

$U = \sum_{i=1}^{q} U_i$ 여기서 U_i 는 X_i 의 상한입니다 .

x_i 가 X_i 열의 값을 나타낸다고 하면 x_i 에서 L 유사성분은 다음과 같이 정의됩니다 .

$$x_i^L = (X_i - L_i)/(S - L)$$

x_i 에서 U 유사성분은 다음과 같이 정의됩니다 .

$$x_i^U = (U_i - X_i)/(U - S)$$

L 유사성분 코드화와 **U 유사성분 코드화**를 둘 다 선택하는 경우, $(S - L) < (U - S)$이면 모형 적합 플랫폼에서 L 코드화를 사용합니다. 그렇지 않은 경우 U 코드화가 사용됩니다.

모형 적합에서는 모수 추정값을 계산하기 전에 유사성분을 사용하여 혼합 요인이 변환됩니다. 이렇게 하면 모수 추정값을 더 의미 있게 만드는 데 도움이 됩니다. 모수 추정값과 관련된 보고서에서 혼합 주효과는 코딩 변환에 의해 제공됩니다. 프로파일러와 같은 다른 보고서는 코드화되지 않은 값을 기반으로 합니다.

혼합의 예

Design Experiment 폴더에 있는 Donev Mixture Data.jmp 샘플 데이터 테이블의 데이터는 Atkinson and Donev(1992) 의 예를 기반으로 합니다 . 혼합 요인 3 개와 비혼합 요인 1 개가 설계에 포함되어 있습니다 .

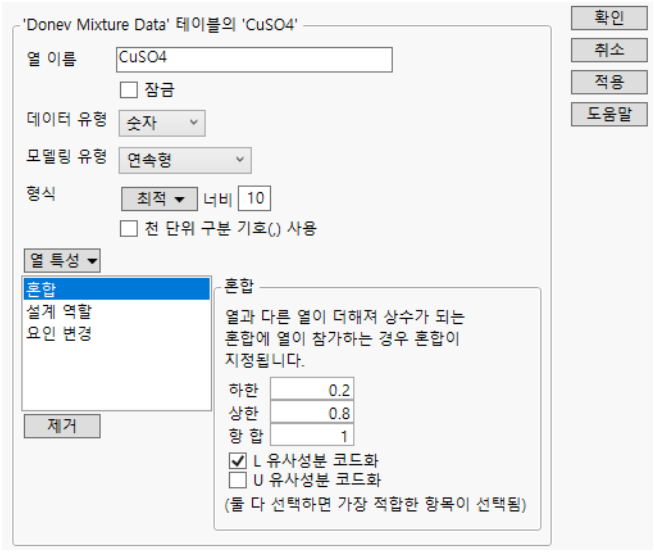
- 반응은 아크릴로니트릴 파우더의 전자기 감쇠를 측정하는 Damping 입니다 .
- 세 가지 혼합 성분은 황산구리 (CuSO4), 티오황산나트륨 (Na2S2O3), 글리옥살 (Glyoxal) 입니다 .
- 관심 있는 비혼합 환경 요인은 빛의 파장 (Wavelength) 입니다 .

이론적으로 Wavelength 는 연속형이지만 연구원들은 세 가지 개별 파장에서의 예측에만 관심이 있었습니다 . 따라서 Wavelength 는 3 수준 범주형 요인으로 간주됩니다 .

사용자 설계를 사용하여 이 상황에 맞는 설계를 구성하는 방법에 대한 자세한 내용은 " 사용자 설계의 예 " 장의 " 혼합물 실험 "(162 페이지) 에서 확인하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Donev Mixture Data.jmp 를 엽니다 .
2. " 열 " 패널에서 CuSO4 옆의 별표를 클릭하고 **혼합**을 선택합니다 .

그림 A.20 CuSO4 의 혼합 열 특성 패널



다음을 확인할 수 있습니다 .

- " 하한 " 은 0.2 이며 CuSO4 의 최소 설계 설정입니다 .
- " 상한 " 은 0.8 이며 CuSO4 의 최대 설계 설정입니다 .
- " 항 합 " 은 1 로 설정되어 있습니다 . 이 값은 세 가지 혼합 요인의 합입니다 .

- **L 유사성분 코드화** 옵션이 선택되어 있습니다 . 자세한 내용은 " 유사성분 코드화 "(748 페이지) 에서 확인하십시오 .

3. **취소**를 클릭합니다 .
4. " 열 " 패널에서 Glyoxal 옆의 별표를 클릭하고 **혼합**을 선택합니다 .

   이 요인에 대해 다음을 확인할 수 있습니다 .

   - " 하한 " 은 0 이며 Glyoxal 의 최소 설계 설정입니다 .
   - " 상한 " 은 0.6 이며 Glyoxal 의 최대 설계 설정입니다 .
5. **취소**를 클릭합니다 .
6. Donev Mixture Data.jmp 데이터 테이블에서 **Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   이 모형에는 혼합 요인의 주효과와 네 가지 모든 요인에 대한 이원 교호작용이 포함됩니다 .
7. **실행**을 클릭합니다 .

   " 모수 추정값 " 보고서에서 혼합 요인은 유사성분 코드화된 형식으로 나타납니다 . 교호작용에 나타나는 혼합 요인은 코드화된 형식으로 표시되지 않습니다 . 그러나 모형 적합은 유사성분을 기반으로 합니다 . " 모수 추정값 " 보고서의 처음 세 항 (그림 A.21 참조) 은 혼합 요인에 대한 코드화된 형식을 보여 줍니다 .

그림 A.21 모수 추정값 보고서

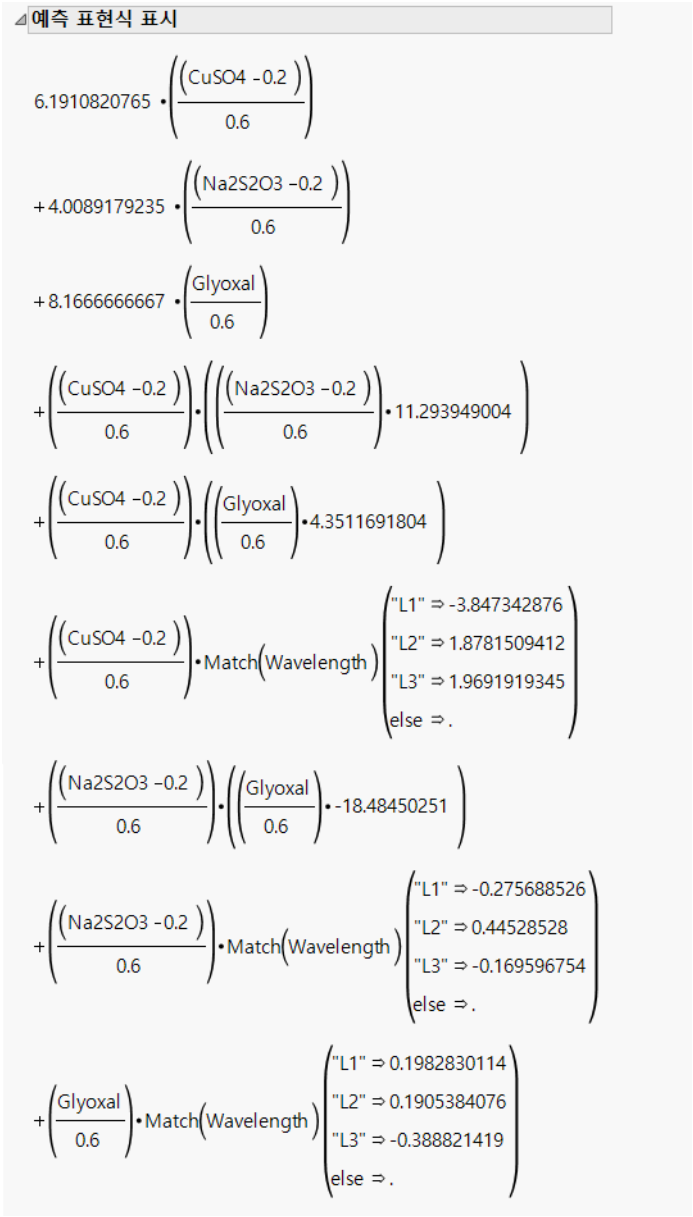
모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
(CuSO4-0.2)/0.6	6.1910821	0.918805	6.74	0.0005*
(Na2S2O3-0.2)/0.6	4.0089179	0.918805	4.36	0.0048*
Glyoxal/0.6	8.1666667	0.921638	8.86	0.0001*
CuSO4*Na2S2O3	11.293949	4.728922	2.39	0.0542
CuSO4*Glyoxal	4.3511692	4.512775	0.96	0.3722
CuSO4*Wavelength[L1]	-3.847343	1.113399	-3.46	0.0135*
CuSO4*Wavelength[L2]	1.8781509	1.079113	1.74	0.1324
Na2S2O3*Glyoxal	-18.4845	4.512775	-4.10	0.0064*
Na2S2O3*Wavelength[L1]	-0.275689	1.064004	-0.26	0.8042
Na2S2O3*Wavelength[L2]	0.4452853	1.103681	0.40	0.7006
Glyoxal*Wavelength[L1]	0.198283	1.090627	0.18	0.8617
Glyoxal*Wavelength[L2]	0.1905384	1.090647	0.17	0.8671

8. " 반응 Damping" 의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다 .

   " 예측 표현식 " 보고서는 적합된 모형을 보여 줍니다 . 혼합 요인은 "L 유사성분 코드화 " 를 사용하여 변환되었습니다 .

그림 A.22 Damping 모형의 예측 표현식



Wavelength L2 에서의 예측에 관심이 있다고 가정해 보겠습니다 . 또한 Na2S2O3 와 Glyoxal 은 각각 하한값 0.2 와 0 으로 설정되고 CuSO4 는 상한값 0.8 로 설정된다고 가정합니다 . 이 경우 예측 Damping 은 CuSO4 의 모수 추정값 (6.191) 과 CuSO4*Wavelength[L2] 의 모수 추정값 (1.878) 을 더한 값과 동일합니다 . " 예측 프로파일러 " 에서 이를 확인할 수 있습니다 .

9. " 반응 Damping" 의 빨간색 삼각형을 클릭하고 **열 저장 > 코드화 테이블 저장**을 선택합니다 .

그림 A.23 코드화된 혼합 요인을 보여 주는 코드화 테이블의 처음 세 열

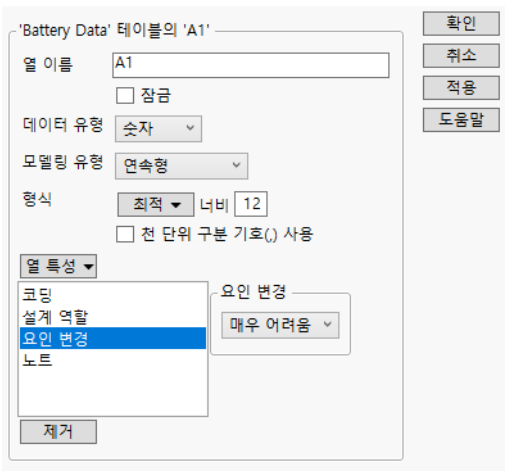
	(CuSO4-0.2)/0.6	(Na2S2O3-0.2)/0.6	Glyoxal/0.6
1	1	0	0
2	0.4	0.6	0
3	0	1	0
4	0.5	0	0.5
5	0	0.5	0.5
6	0	0	1
7	1	0	0
8	0.6	0.4	0
9	0	1	0
10	0.5	0	0.5
11	0	0.5	0.5
12	0	0	1
13	1	0	0
14	0.4	0.6	0
15	0	1	0
16	0.5	0	0.5
17	0	0.5	0.5
18	0	0	1

이 특정 설계의 경우 "L 유사성분 코드화 " 를 통해 혼합 요인이 0 ~ 1 사이의 범위로 변환됩니다 . 일반적으로는 일어나지 않는 일입니다 .

요인 변경

"요인 변경" 열 특성은 설계된 실험에서 요인 설정을 변경하기가 얼마나 어려운지를 나타냅니다. "요인 변경"에 사용 가능한 값은 "쉬움", "어려움" 및 "매우 어려움"입니다. 예를 들어 그림 A.24에서는 Design Experiment 폴더에 있는 Battery Data.jmp 샘플 데이터 테이블에서 A1 요인에 대한 "요인 변경" 열 특성 패널을 보여 줍니다.

그림 A.24 요인 변경 열 특성 패널



- 변경하기 힘든 요인과 매우 변경하기 힘든 요인이 설계에 포함된 경우 하위구 및 주구 요인도 설계에 포함되어야 합니다 .
  - 주구 요인의 수준은 매우 변경하기 힘든 요인의 수준이 일정하게 유지되는 런 그룹을 정의합니다 .
  - 하위구 요인의 수준은 변경하기 힘든 요인의 수준이 일정하게 유지되는 런 그룹을 정의합니다 .
- 변경하기 힘든 요인만 설계에 포함되고 매우 변경하기 힘든 요인은 없는 경우 주구 요인을 포함해야 합니다 . 주구 요인의 수준은 변경하기 힘든 요인의 수준이 일정하게 유지되는 런 그룹을 정의합니다 . 자세한 내용은 " 사용자 설계 " 장의 " 변경 및 랜덤 블록 "(87 페이지) 에서 확인하십시오 .

설계 확대 및 평가

설계 평가 및 설계 확대 플랫폼의 경우 "요인 변경" 열 특성은 "변경"이 "어려움" 또는 "매우 어려움"으로 지정된 요인을 식별합니다. 그러나 이러한 플랫폼에서는 시작 창에서 주구 요인과 하위구 요인도 모형 효과로 입력해야 합니다. 주구 요인과 하위구 요인이 설계 구조의 일부이기 때문입니다.

사용자 설계

사용자 설계 플랫폼을 사용하면 모든 요인 변경이 "쉬움"으로 지정된 설계를 생성할 수 있습니다. 분할구, 분할-분할구 또는 이원 분할구(조각구) 설계를 구성할 수도 있습니다. 이러한 설계를 구성할 때는 값을 변경하기 힘들거나 매우 변경하기 힘든 요인을 식별해야 합니다. "요인" 개요에서 "변경"이 "쉬움", "어려움" 또는 "매우 어려움"으로 지정되는 요인을 식별할 수 있습니다. 사용자 설계 플랫폼에서 설계 테이블을 구성할 때 "요인" 개요에 나타나는 모든 요인에 "요인 변경" 특성이 할당됩니다.

"변경"이 "어려움" 또는 "매우 어려움"으로 지정된 요인에 대한 설계를 구성하려면 사용자 설계 플랫폼을 사용해야 합니다. 다른 DOE 플랫폼에서도 구성하는 요인에 "요인 변경" 열 특성을 할당하지만 해당 요인의 열 특성 값이 "쉬움"으로 설정됩니다.

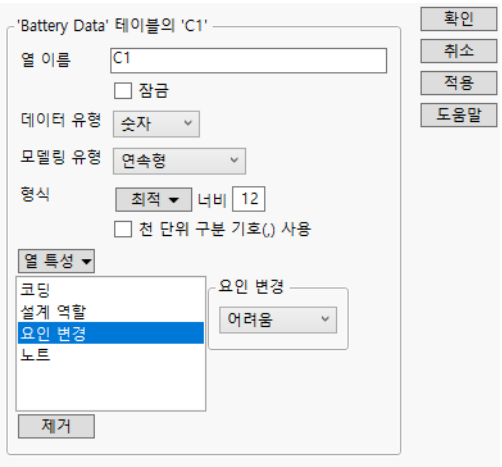
"사용자 설계" 창에서 요인 테이블을 사용하여 요인을 불러오는 경우 "요인 변경" 열 특성을 해당 테이블의 열에 할당할 수 있습니다. 테이블을 사용하여 요인을 불러오면 지정된 요인 변경 설정이 "요인" 개요에 나타납니다.

요인 변경의 예

Design Experiment 폴더의 Battery Data.jmp 샘플 데이터 테이블에는 배터리의 OCV(개회로 전압) 를 연구하는 실험 데이터가 포함되어 있습니다 . 이 설계는 이원 분할구 설계입니다 . 자세한 배경 정보는 " 사용자 설계의 예 " 장 (184 페이지) 에서 확인하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Battery Data.jmp 를 엽니다 .
2. " 열 " 패널에서 C1 요인 오른쪽의 별표를 클릭합니다 .
3. **요인 변경**을 선택합니다 .

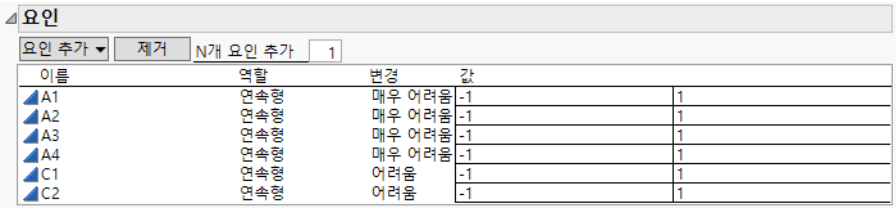
그림 A.25 C1 의 요인 변경 패널



C1 의 " 요인 변경 " 값은 " 어려움 " 입니다 . 그림 A.24 에서는 A1 의 " 요인 변경 " 값이 " 매우 어려움 " 으로 지정된 것을 보여 줍니다 .

4. **확인**을 클릭합니다 .
5. 데이터 테이블에서 **DOE Dialog** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
6. **요인** 개요를 엽니다 .

그림 A.26 배터리 실험의 요인 개요



A1, A2, A3 및 A4 요인은 " 변경 " 이 " 매우 어려움 " 으로 설정되고 C1 및 C2 요인은 " 변경 " 이 " 어려움 " 으로 설정되어 있습니다 . 사용자 설계 플랫폼은 설계 테이블을 구성할 때 이러한 설정을 해당 열에 " 요인 변경 " 열 특성으로 저장합니다 .

" 설계 " 개요에서 **주구** 및 **하위구** 요인을 확인하십시오 .

그림 A.27 설계 개요 (일부분)

설계

런	주구	하위구	A1	A2	A3	A4	C1	C2
1	1	1	-1	-1	1	-1	-1	-1
2	1	2	-1	-1	1	-1	1	-1
3	1	3	-1	-1	1	-1	-1	1
4	2	4	-1	1	1	1	1	1
5	2	5	-1	1	1	1	-1	1
6	2	6	-1	1	1	1	1	-1
7	3	7	1	1	1	-1	-1	1
8	3	8	1	1	1	-1	1	1
9	3	9	1	1	1	-1	-1	-1
10	4	10	1	-1	1	1	1	1
11	4	11	1	-1	1	1	1	-1
12	4	12	1	-1	1	1	-1	-1

사용자 설계 플랫폼에서는 " 어려움 " 또는 " 매우 어려움 " 으로 지정된 요인 변경을 설명하기 위해 두 개의 요인이 생성됩니다 . **주구** 요인은 " 매우 어려움 " 요인 수준이 일정하게 유지되는 런을 그룹화하고 , **하위구** 요인은 " 어려움 " 요인 수준이 일정하게 유지되는 런을 그룹화합니다 . 설계 평가 및 설계 확대 플랫폼에서 " 요인 변경 " 열 특성을 포함하는 열을 입력할 때 이러한 요인을 모형 효과로 포함해야 합니다 .

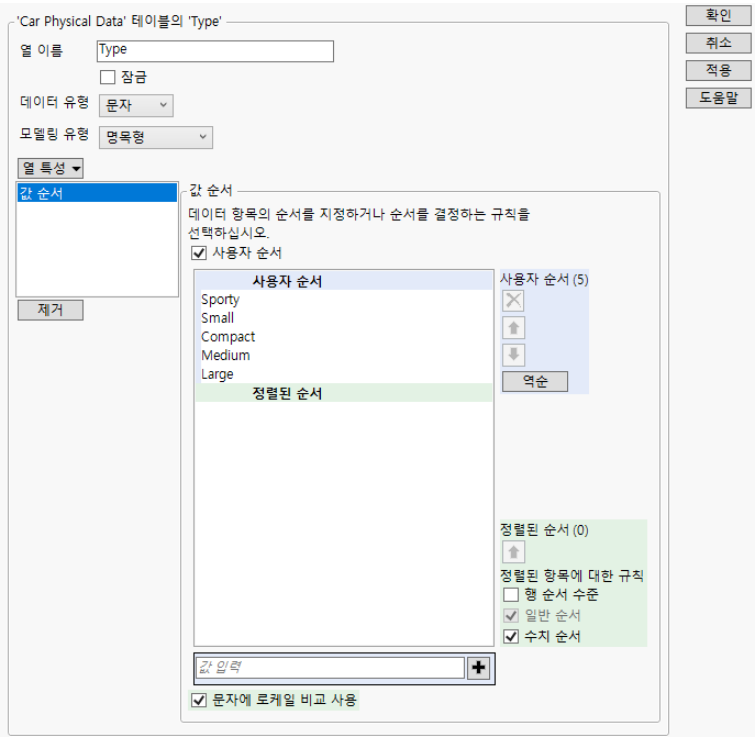
값 순서

" 값 순서 " 열 특성은 열의 값에 순서를 할당합니다 . 이 순서는 그림과 분석에 사용됩니다 . 보고서에 값이 나타나는 순서를 지정할 수 있습니다 .

참고 : JMP 는 요일과 같이 일반 순서가 있는 특정 값을 보고서에 나타낼 때 적절한 방식을 사용하여 자동으로 정렬합니다 . 자세한 내용은 **JMP 사용**의 에서 확인하십시오 .

그림 A.28에서는 Car Physical Data.jmp 샘플 데이터 테이블의 Type 열에 대한 "값 순서" 패널을 보여 줍니다. Type 값을 포함하는 보고서는 값 수준을 Sporty, Small, Compact, Medium, Large 순으로 배치합니다. "값 순서" 목록 오른쪽의 버튼을 사용하여 원하는 값 순서를 지정할 수 있습니다.

그림 A.28 Type 에 대한 값 순서 열 특성



대부분의 DOE 플랫폼을 사용하여 생성된 설계에서는 구성된 요인(주구, 하위구)을 포함하는 범주형 요인 및 블록 요인에 "값 순서" 특성이 할당됩니다. 이 특성은 수준을 "요인" 개요에 나타나는 순서에 따라 정렬합니다. 구성된 요인의 수준은 연속된 정수이며 "값 순서" 특성은 이러한 일반 순서를 지정합니다. 필요에 따라 요인의 "값 순서" 설정을 수정할 수 있습니다.

덮기 배열 또는 Taguchi 배열 플랫폼에서는 "값 순서" 특성이 할당되지 않습니다. 덮기 배열 플랫폼은 "값 라벨" 열 특성을 할당합니다. 자세한 내용은 "값 라벨"(759페이지)에서 확인하십시오.

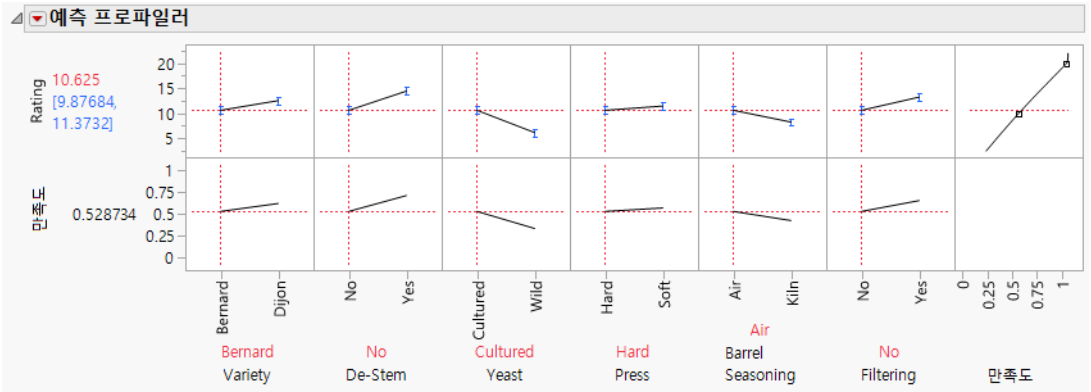
값 순서의 예

" 예측 프로파일러 " 에서 요인의 값을 다른 순서로 표시하려는 경우를 가정해 보겠습니다 . 사용자 설계를 사용하여 구성한 와인 시음 실험을 예로 들겠습니다 . 와인은 Wine Data.jmp 샘플 데이터 테이블에 각각 Rater 로 나열된 5 명의 전문가에 의해 평가됩니다 . Rater 는 고정 블록 요인입니다 . 9 개 요인이 연구 대상입니다 . 응답은 Rating 입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Wine Data.jmp 를 엽니다 .
2. 테이블 패널에서 **Reduced Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
3. **실행**을 클릭합니다 .

" 예측 프로파일러 " 가 보고서 맨 아래에 나타납니다 .

그림 A.29 원래 값 순서로 표시된 프로파일러

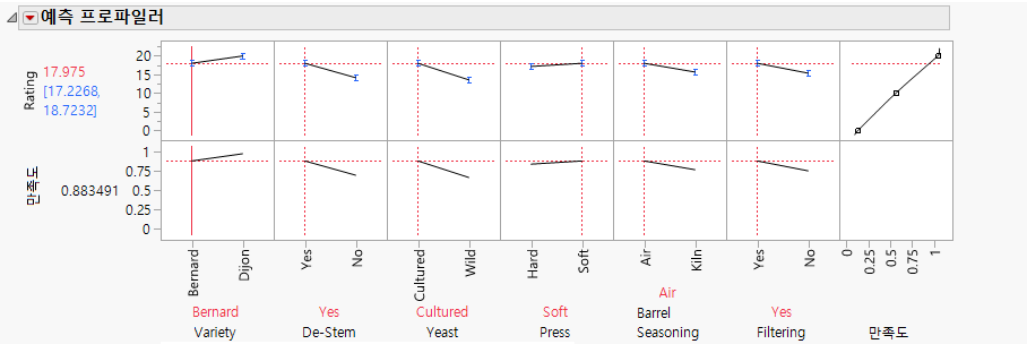


De-Stem 및 Filtering 값은 No 다음에 Yes 가 나오는 순서로 되어 있습니다 . Yes 수준이 먼저 나오도록 순서를 바꾸려고 합니다 .

팁 : 프로파일러를 정렬하려면 " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **모양 > 행으로 배열**을 선택합니다 . 이 예에서는 7 을 입력하고 **확인**을 클릭합니다 .

4. " 반응 Rating" 보고서를 닫습니다 .
5. 데이터 테이블의 " 열 " 패널에서 De-Stem 옆의 별표를 클릭하고 " 값 순서 " 를 선택합니다 .
6. **역순**을 클릭합니다 .
7. **확인**을 클릭합니다 .
8. " 열 " 패널에서 Filtering 옆의 별표를 클릭하고 " 값 순서 " 를 선택합니다 .
9. **역순**을 클릭합니다 .
10. **확인**을 클릭합니다 .
11. 다시 **Reduced Model** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
12. **실행**을 클릭합니다 .

그림 A.30 새 값 순서로 표시된 프로파일러



이제 De-Stem 및 Filtering 수준이 Yes 다음에 No 가 나오는 순서로 되어 있습니다 .

13. 변경 사항을 저장하지 않고 Wine Data.jmp 샘플 데이터 테이블을 닫습니다 .

값 순서 할당

Candy Bars.jmp 샘플 데이터 테이블을 예로 들겠습니다. Brand 아래에 나열된 18 개 브랜드 중 Hershey 와 M&M/Mars 가 Name 열에 나열된 사탕 종류가 가장 많습니다. 이 두 브랜드를 보고서에 맨 먼저 나타내려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Candy Bars.jmp 를 엽니다 .
2. Brand 열을 선택합니다 .
3. **열 > 열 정보**를 선택합니다 .
4. " 열 특성 " 에서 " 값 순서 " 를 선택합니다 .
5. " 값 순서 " 목록에서 "Hershey" 를 선택합니다 .
6. **위쪽 화살표**를 5 번 클릭합니다 .

   이제 Hershey 가 " 값 순서 " 목록에 맨 먼저 나옵니다 .
7. " 값 순서 " 목록에서 "M&M/Mars" 를 선택합니다 .
8. **위쪽 화살표**를 7 번 클릭합니다 .

   이제 M&M/Mars 가 " 값 순서 " 목록에서 두 번째 위치에 나옵니다 .
9. **확인**을 클릭합니다 .

   " 값 순서 " 열 특성을 나타내는 별표가 " 열 " 패널의 Brand 옆에 나타납니다 . 이제 JMP 는 Brand 를 포함하는 보고서에서 Hershey 와 M&M/Mars 를 먼저 나열합니다 .
10. **분석 > 분포**를 선택합니다 .
11. Calories 를 선택하고 **Y, 열**을 클릭합니다 .
12. Brand 를 선택하고 **기준**을 클릭합니다 .
13. **확인**을 클릭합니다 .

14. Ctrl 키를 누르고 "Calories" 의 빨간색 삼각형을 클릭한 후 **표시 옵션 > 가로 레이아웃**을 선택합니다 .

    18 개 브랜드 중 Hershey 와 M&M/Mars 에 대한 " 분포 " 보고서가 맨 먼저 나옵니다 .

15. 변경 사항을 저장하지 않고 Candy Bars.jmp 샘플 데이터 테이블을 닫습니다 .

값 라벨

"값 라벨" 열 특성은 라벨이 지정된 열에 값을 나타냅니다. 이러한 라벨은 데이터 테이블에 표시되고 그림과 보고서에 사용됩니다. 데이터 테이블에서 셀 내부를 두 번 클릭하면 원래 값을 볼 수 있습니다. "값 라벨" 열 특성을 할당하고 사용하는 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

DOE 플랫폼 중 덮기 배열 플랫폼에서만 "값 라벨" 열 특성을 할당할 수 있습니다. 덮기 배열 플랫폼은 "명목형" 모델링 유형을 사용하여 요인을 데이터 테이블에 저장합니다. 기본 값은 1 ~ 덮기 배열 "요인" 개요에 지정한 수준 수 범위의 연속된 정수입니다. "요인" 개요에서 지정한 값은 기본 정수에 할당되는 값 라벨입니다.

값 라벨의 예

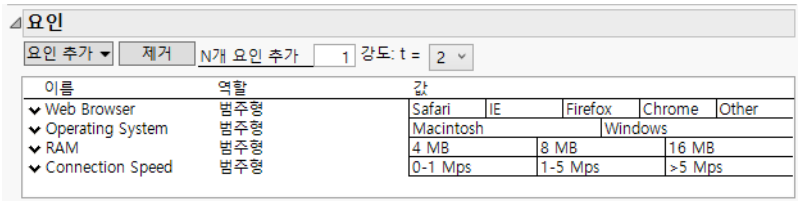
인터넷 기반 소프트웨어 응용 프로그램을 테스트하여 운영 환경의 구성 요소에서 발생하는 문제를 감지하려고 합니다. 관심 있는 네 가지 구성 요소는 브라우저, 운영 체제, 컴퓨터 RAM 및 연결 속도입니다. 테스트 시간을 최소화하기 위해 이원 교호작용 검정으로 제한합니다.

필요한 요인 수준 조합을 검정하기 위해 강도 2 덮기 배열을 구성합니다 .

1. **DOE > 특수 용도 > 덮기 배열**을 선택합니다 .
2. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Software Factors.jmp 를 엽니다 .
3. " 덮기 배열 " 의 빨간색 삼각형을 클릭하고 **요인 불러오기**를 선택합니다 .

   요인과 요인 수준이 " 요인 " 개요에 나타납니다 .

그림 A.31 소프트웨어 요인에 대한 요인 개요



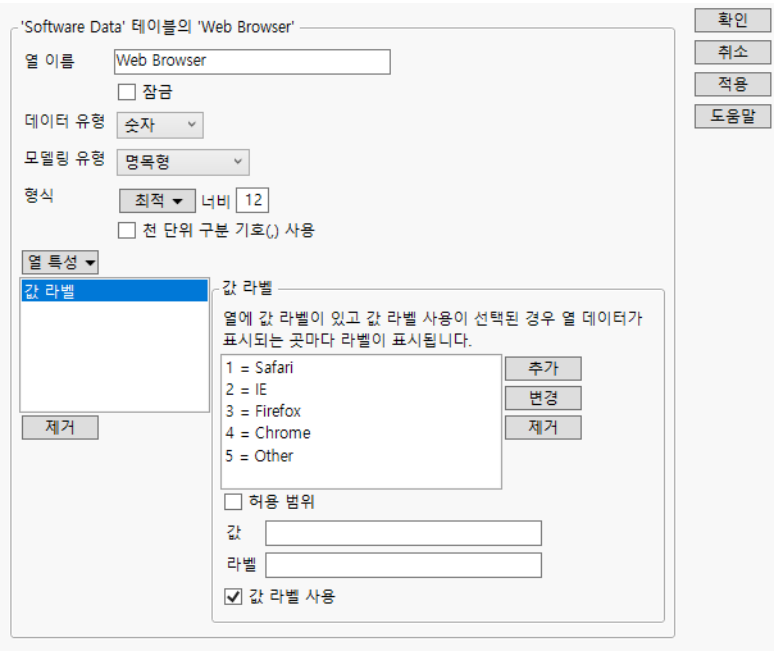
   네 가지 요인의 " 역할 " 이 " 범주형 " 으로 기술되어 있습니다 .

4. **계속**을 클릭합니다 .

" 요인 수준 조합 제한 " 개요가 열립니다 . 모든 설정 조합이 가능하므로 이 설정을 **없음**으로 그대로 둡니다 .

5. **설계 생성**을 클릭합니다 .
6. **테이블 생성**을 클릭합니다 .
7. " 열 " 패널에서 Web Browser 옆의 별표를 클릭하고 **값 라벨**을 선택합니다 .

그림 A.32 요인 A 의 열 정보 창

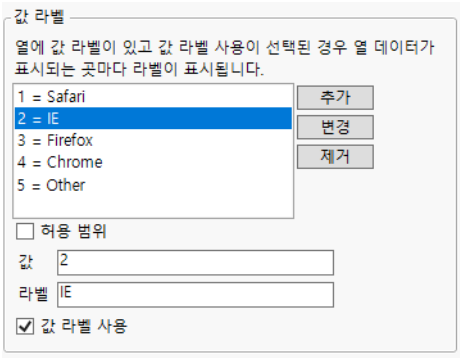


Web Browser 의 데이터 유형은 " 숫자 " 이고 모델링 유형은 " 명목형 " 입니다 . Web Browser 의 기본 값은 정수 1, 2, 3, 4, 5 입니다 . 설계를 구성할 때 입력한 값에 해당하는 값 라벨이 이 값에 할당됩니다 .

2 에 대한 값 라벨을 "IE" 에서 "Internet Explorer" 로 변경합니다 .

8. " 값 라벨 " 목록에서 **2 = IE** 를 선택합니다 .

그림 A.33 선택 항목이 포함된 값 라벨 패널



9. " 라벨 " 옆에 **Internet Explorer** 를 입력합니다 .
10. **변경**을 클릭합니다 .

    변경 사항이 데이터 테이블에 나타납니다 .

참고 : 라벨이 아니라 숫자 값을 사용하려면 **값 라벨 사용**을 선택 취소하십시오 .

블록당 런 수

DOE 플랫폼을 사용하여 블록 요인이 포함된 설계를 구성할 때 "블록화" 값이 지정된 "설계 역할" 열 특성이 요인에 할당됩니다. 또한 JMP는 각 블록 요인에 "블록당 런 수" 열 특성을 할당합니다. "블록당 런 수" 특성은 각 블록에서 허용되는 최대 런 수를 나타냅니다. 이 특성은 설계 평가 및 설계 확대 플랫폼에서 요인의 블록화 구조를 나타내는 데 사용됩니다. 자세한 내용은 "사용자 설계" 장의 "블록화"(86 페이지)에서 확인하십시오.

참고 : " 블록당 런 수 " 열 특성은 설계 구성 중에 JMP 에서 할당합니다 . 이 열 특성을 직접 할당할 수 없습니다 .

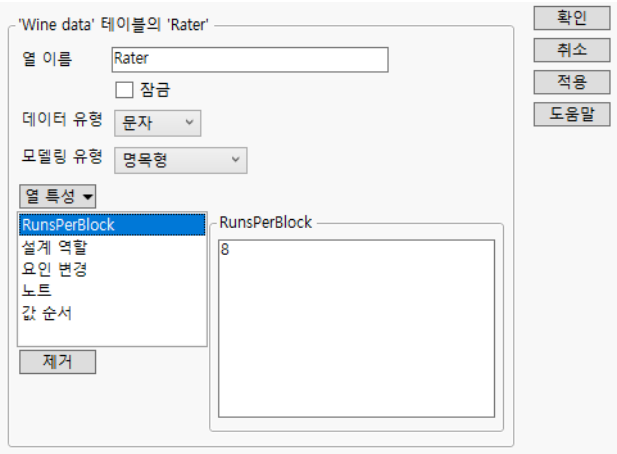
블록당 런 수의 예

" 사용자 설계 " 장의 " 사용자 설계의 예 "(66 페이지) 에 설명된 와인 시음 실험을 예로 들겠습니다 . 5 명의 평가자 (Rater) 가 와인 샘플을 시음하고 각 평가자는 8 개의 샘플을 시음합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Wine Data.jmp 를 엽니다 .
2. " 열 " 패널에서 Rater 옆의 별표를 클릭하고 **설계 역할**을 선택합니다 .

   " 설계 역할 " 이 " 블록화 " 로 설정되어 있습니다 .
3. **취소**를 클릭합니다 .
4. " 열 " 패널에서 Rater 옆의 별표를 클릭하고 **블록당 런 수**를 선택합니다 .

그림 A.34 Rater 의 블록당 런 수 열 특성 패널



" 블록당 런 수 " 의 값은 8 입니다 . "DOE 대화상자 " 스크립트로 구성된 설계에는 40 회의 런이 있습니다 . 평가자가 5 명이므로 JMP 는 각 평가자마다 8(40/5) 회 런을 실행하는 설계를 구성합니다 .

제약 조건 상태

사용자 설계 및 혼합물 설계 플랫폼에서는 설계에 지정한 제약 조건을 저장할 수 있습니다 . "제약 조건 저장 "을 선택하면 각 선형 제약 조건의 계수가 데이터 테이블의 열에 나타납니다 . 부등식을 제한하는 값은 테이블의 마지막 행에 제공됩니다 .

각 제약 조건 열에는 " 제약 조건 상태 " 열 특성이 할당됩니다 . 이 특성은 제약 조건을 정의하는 부등식의 방향을 지정합니다 . 설계 플랫폼에서 " 제약 조건 불러오기 "를 선택하면 " 제약 조건 상태 " 열 특성이 JMP 에 부등식의 방향을 알려 줍니다 .

참고 : " 제약 조건 상태 " 열 특성은 설계 구성 중에 JMP 에서 할당합니다 . 이 열 특성을 직접 할당할 수 없습니다 .

제약 조건 상태의 예

Design Experiment 폴더의 Piepel.jmp 샘플 데이터 테이블에는 세 가지 연속형 요인을 사용하는 혼합물 설계가 포함되어 있습니다 . 이 설계는 Snee(1979) 및 Piepel(1988) 에 제시된 실험을 기반으로 하며 , 각 요인에 대한 경계 제약 조건과 세 가지 추가 선형 제약 조건이 있습니다 . 이 예에서는 다음을 수행합니다 .

1. 세 가지 추가 제약 조건 중 하나를 변경합니다 .
2. 제약 조건을 테이블에 저장합니다 .

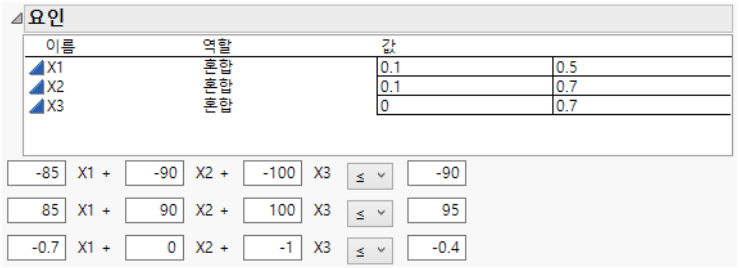
3. " 제약 조건 상태 " 열 특성이 제약 조건의 부등식 방향을 설명하는 방식을 관찰합니다 .

여기에서 좀 더 발전된 예에 대한 자세한 내용은 " 혼합물 설계 " 장의 " 선형 제약 조건이 있는 꼭지점의 예 "(378 페이지) 에서 확인하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Piepel.jmp 를 엽니다 .
2. 테이블 패널에서 **DOE Dialog** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   " 요인 " 개요 아래에 세 가지 선형 제약 조건이 있습니다 . 제약 조건을 더 쉽게 해석할 수 있도록 하기 위해 첫 번째 제약 조건을 " 보다 크거나 같음 " 부등식 측면에서 다시 작성하려고 합니다 .

그림 A.35 요인 개요 아래의 선형 제약 조건



3. 첫 번째 제약 조건에서 다음을 수행합니다 .
   - X1 옆에 85 를 입력합니다 .
   - X2 옆에 90 을 입력합니다 .
   - X3 옆에 100 을 입력합니다 .
   - 부등식 메뉴에서 ≥ 을 선택합니다 .
   - 부등식 부호 오른쪽에 90 을 입력합니다 .
4. " 혼합물 설계 " 의 빨간색 삼각형을 클릭하고 **제약 조건 저장**을 선택합니다 .

   제약 조건 정보를 포함하는 테이블이 나타납니다 .

그림 A.36 제약 조건 테이블

제목 없음 7

	제약 조건 1	제약 조건 2	제약 조건 3
1	-85	85	-0.7
2	-90	90	0
3	-100	100	-1
4	-90	95	-0.4

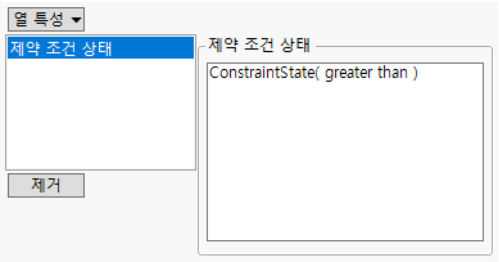
열(3/0)
제약 조건 1 ✱
제약 조건 2 ✱
제약 조건 3 ✱

각 열의 1 ~ 3 행에는 X1, X2, X3 요인의 계수가 포함되어 있습니다 . 4 행에는 부등식 부호의 오른쪽에 나오는 값이 있습니다 .

5. 제약 조건 1 옆의 별표를 클릭합니다 .

6. **제약 조건 상태**를 클릭합니다 .

그림 A.37 제약 조건 상태 열 특성 패널



X1 의 " 제약 조건 상태 " 패널을 보면 부등식 방향이 " 보다 크거나 같음 "(≥) 을 나타내는 "greater than" 임을 알 수 있습니다 .

7. **취소**를 클릭합니다 .

부록 B

기술 상세 정보

이 부록에는 별칭 행렬, 검정력 계산 및 상대 예측 분산에 대한 정보가 포함되어 있습니다.

- " 별칭 행렬 "(767 페이지)
- " 검정력 계산 "(768 페이지)
- " 상대 예측 분산 "(770 페이지)

목차

별칭 행렬 767
변경하기 힘들거나 매우 변경하기 힘든 요인이 있는 설계 767
If Possible 효과가 있는 설계 768
검정력 계산 768
단일 모수에 대한 검정력 768
범주형 효과에 대한 검정력 769
상대 예측 분산 770

별칭 행렬

별칭 행렬 항목은 "별칭 항" 개요에서 지정한 효과에 의해 모형 모수에 전달되는 편향 정도를 나타냅니다. 별칭 행렬은 별칭 최적을 정의하는 데도 사용됩니다. 자세한 내용은 "사용자 설계" 장의 "별칭 최적"(121 페이지)에서 확인하십시오.

별칭 행렬의 계산은 모형 행렬을 기반으로 합니다. 자세한 내용은 "사용자 설계" 장의 "모형 행렬"(105 페이지)에서 확인하십시오.

$\mathbf{X}_1$을 "모형" 개요의 항에 해당하는 모형 행렬로 지정하겠습니다. "별칭 항" 개요에 지정된 효과에 대한 모형 항 행렬을 $\mathbf{X}_2$로 나타냅니다.

가정된 모형은 다음과 같습니다.

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \varepsilon$$

일부 별칭 항이 활성이고 참모형이 다음과 같이 제공되는 경우를 가정해 보겠습니다.

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$$

β_1의 최소 제곱 추정량은 다음과 같이 구합니다.

$$\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$$

일반적인 회귀 가정하에서 $\hat{\beta}_1$의 기대값은 다음과 같이 구합니다.

$$E[\hat{\beta}_1] = \beta_1 + \mathbf{A}\beta_2$$

여기서 $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ 입니다.

행렬 **A**를 **별칭 행렬**이라고 합니다.

변경하기 힘들거나 매우 변경하기 힘든 요인이 있는 설계

변경하기 힘들거나 매우 변경하기 힘든 요인이 있는 설계에 대한 별칭 행렬은 다음과 같이 제공됩니다.

$$\mathbf{A} = (\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{V}^{-1}\mathbf{X}_2)$$

여기서 **V**는 반응에 대한 공분산 행렬의 대각 블록입니다.

If Possible 효과가 있는 설계

If Possible 효과가 있는 설계 (베이지안 D- 최적 및 I- 최적 설계) 에 대한 별칭 행렬은 다음과 같이 제공됩니다 .

$$\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1 + \mathbf{K}^2)^{-1}\mathbf{X}_1'\mathbf{X}_2$$

여기서 **K** 는 다음 값을 갖는 대각 행렬입니다 .

- $k = 0$(필수 항의 경우)
- $k = 1$(If Possible 주효과 , 검정력 , 3 수준 이상의 범주형 요인과 관련된 교호작용의 경우)
- $k = 4$(다른 모든 If Possible 항의 경우)

베이지안 설계의 경우 별칭 행렬은 $\mathbf{K}^{-1}$ 의 사전 분산을 가진 능형 회귀에 해당하는 효과 별칭 구조를 제공합니다. 베이지안 설계에 대한 자세한 내용은 "사용자 설계" 장의 "베이지안 D- 최적 "(119 페이지) 및 "사용자 설계" 장의 "베이지안 I- 최적 "(121 페이지) 에서 확인하십시오 .

사용자 설계 플랫폼에서는 빨간색 삼각형 메뉴의 **고급 옵션 > 사전 모수 분산**을 선택하여 If Possible 항에 사용되는 가중치를 제어할 수 있습니다. 여기서 **K**의 주대각 요소를 지정하여 모든 모형 항에 대한 사전 분산을 설정할 수 있습니다. "설계 생성"을 클릭하면 옵션이 업데이트되어 기본 가중치가 표시됩니다.

검정력 계산

" 검정력 분석 " 보고서는 단일 모수 값에 대한 검정력 계산 및 설계에 3 수준 이상의 범주형 요인이 포함된 경우 전체 효과에 대한 검정력 계산을 제공합니다 . 이 섹션에서는 다음 두 가지 경우에 대한 계산을 설명합니다 .

- " 단일 모수에 대한 검정력 "(768 페이지)
- " 범주형 효과에 대한 검정력 "(769 페이지)

단일 모수에 대한 검정력

이 섹션에서는 단일 모수에 대한 검정력을 계산하는 방법을 설명합니다. 다음 표기가 사용됩니다.

X 모형 행렬입니다 . 명목형 효과에 대한 코딩에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 . " 사용자 설계 " 장의 " 모형 행렬 "(105 페이지) 에 나오는 내용도 참조하십시오 .

> **참고 :** 모형 적합을 실행하여 모형 행렬을 볼 수 있습니다 . 그런 다음 기본 보고서의 빨간색 삼각형 메뉴에서 **열 저장 > 코드화 테이블 저장**을 선택하십시오 .

β_t 관심 항에 해당하는 모수입니다 .

$\hat{\beta}_i$ β_i 의 최소 제곱 추정량입니다 .

β_i^A 예상 계수 값입니다. 감지할 차이는 $2\beta_i^A$ 입니다.

$\hat{\beta}_i$의 분산은 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$의 i^{th} 주대각 항목으로 구합니다. 여기서 σ^2은 오차 분산입니다. $(\mathbf{X}'\mathbf{X})^{-1}$의 i^{th} 주대각 항목은 $(\mathbf{X}'\mathbf{X})_{ii}^{-1}$로 나타냅니다.

오차 분산 σ^2은 MSE를 사용하여 추정되고 자유도가 $n-p-1$입니다. 여기서 n은 관측값 수이고, p는 모형에서 절편을 제외한 항 수입니다. $n-p-1=0$이면 JMP는 오차의 자유도를 1로 설정합니다. 이렇게 하면 포화 설계의 모수에 대해 검정력이 추정됩니다.

$H_0:\beta_i = 0$ 검정은 다음과 같이 지정됩니다.

$$\frac{\hat{\beta}_i}{\sqrt{MSE(\mathbf{X}'\mathbf{X})_{ii}^{-1}}}$$

또는 다음과 같이 지정될 수도 있습니다.

$$F_0 = \frac{\hat{\beta}_i^2}{MSE(\mathbf{X}'\mathbf{X})_{ii}^{-1}}$$

귀무가설 하에서 검정 통계량 F_0은 *1* 및 $n-p-1$ 자유도에서 F 분포를 따릅니다.

β_i의 실제 값이 β_i^A인 경우 F_0은 다음과 같이 계산된 비중심성 모수를 갖는 비중심 F 분포를 따릅니다.

$$\lambda = \frac{(\beta_i^A)^2}{\sigma^2(\mathbf{X}'\mathbf{X})_{ii}^{-1}}$$

검정력을 계산하려면 먼저 α 수준 임계값 (F_c)을 구합니다.

$$\alpha = 1 - FDist(F_c, 1, n-p-1)$$

그런 다음 아래와 같이 검정력을 계산합니다.

$$Power = 1 - FDist(F_c, 1, n-p-1, \lambda)$$

범주형 효과에 대한 검정력

이 섹션에서는 전체 범주형 효과에 대한 검정력을 계산하는 방법을 설명합니다. 다음 표기가 사용됩니다.

X 모형 행렬입니다. 자세한 내용은 " 기술 상세 정보 " 부록의 " 별칭 행렬 "(767 페이지) 에서 확인하십시오.

β 모수 벡터입니다.

$\hat{\beta}$ β의 최소 제곱 추정값입니다.

β^A 예상 계수 값의 벡터입니다.

L 범주형 효과에 대한 검정을 정의하는 행렬입니다. 행렬 **L** 은 β 에서 범주형 효과에 해당하는 모수의 값을 식별하여 0 으로 설정합니다. 범주형 효과 검정에 대한 귀무가설은 다음과 같이 지정됩니다.

$H_0: \mathrm{L}\beta = \mathbf{0}$

r **L** 의 순위입니다. *r* 는 범주형 효과의 수준 수에서 1 을 뺀 값입니다.

참고 : 모형 적합을 실행하여 설계 행렬을 볼 수 있습니다. 그런 다음 기본 보고서의 빨간색 삼각형 메뉴에서 **열 저장 > 코드화 테이블 저장**을 선택하십시오.

$\hat{\beta}$ 의 공분산 행렬은 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ 로 구합니다. 여기서 σ^2은 오차 분산입니다.

오차 분산 σ^2은 MSE를 사용하여 추정되고 자유도가 $n-p-1$입니다. 여기서 n은 관측값 수이고, p는 모형에서 절편을 제외한 항 수입니다. $n-p-1=0$이면 JMP는 오차의 자유도를 1로 설정합니다. 이렇게 하면 포화 설계의 모수에 대해 검정력이 추정됩니다.

$H_0: \mathrm{L}\beta = \mathbf{0}$ 검정은 다음과 같이 지정됩니다.

$$F_0 = \left((\mathbf{L}\hat{\beta})'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\mathbf{L}\hat{\beta})\right)/(rMSE)$$

귀무가설 하에서 검정 통계량 F_0은 r 및 $n-p-1$ 자유도에서 F 분포를 따릅니다.

β 의 실제 값이 β^A 인 경우 F_0 은 다음과 같이 계산된 비중심성 모수를 갖는 비중심 F 분포를 따릅니다.

$$\lambda = \left((\mathbf{L}\beta^A)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\mathbf{L}\beta^A)\right)/\sigma^2$$

검정력을 계산하려면 먼저 α 수준 임계값 (F_c) 을 구합니다.

$$\alpha = 1 - FDist(F_c, r, n-p-1)$$

그런 다음 아래와 같이 검정력을 계산합니다.

$$Power = 1 - FDist(F_c, r, n-p-1, \lambda)$$

상대 예측 분산

다음 표기를 고려하십시오.

X 모형 행렬입니다. 자세한 내용은 " 사용자 설계 " 장의 " 모형 행렬 "(105 페이지) 에서 확인하십시오. 사용자 설계는 모형 행렬을 보여 주는 스크립트를 제공합니다. 자세한 내용은 " 사용자 설계 " 장의 " 행렬 X 저장 "(105 페이지) 에서 확인하십시오.

σ^2 오차 분산입니다 .

$\hat{\beta}$ 모수의 최소 제곱 추정값 벡터입니다 .

$\mathbf{x}_i'$ $\mathbf{X}$ 의 i^{th} 행입니다 .

이 표기를 사용할 경우 $\mathbf{X}$의 i^{th} 행에 대한 예측 반응은 다음과 같이 지정됩니다.

$$\hat{\mathbf{Y}} = \mathbf{x}_i'\hat{\beta}$$

$\mathbf{x}_i$로 정의된 설정에서 상대 예측 분산은 다음과 같이 지정됩니다.

$$\frac{\mathbf{x}_i' var(\hat{\mathbf{Y}})\mathbf{x}_i}{\sigma^2} = \frac{\mathbf{x}_i' var(\mathbf{x}_i'\hat{\beta})\mathbf{x}_i}{\sigma^2} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$$

부록 C

참조 자료

Agresti, A., and Coull, B. A. (1998). “Approximate Is Better than ‘Exact’ for Interval Estimation of Binomial Proportions.” American Statistician 52:119–126.

Agresti, A., and Caffo, B. N. (2000). “Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures.” American Statistician 54:280–288.

Atkinson, A. C., and Donev, A. N.(1992). Optimum Experimental Designs. New York: Oxford University Press.

Barker, C. (2011). “Power and Sample Size Calculations in JMP.” Technical Report, Cary NC: SAS Institute Inc. Accessed May 6, 2018. https://www.jmp.com/content/dam/jmp/documents/en/technical-reports/powerAndSampleSize.pdf.

Borror, C. M., and Montgomery, D. C. (2000). “Mixed Resolution Designs as Alternatives to Taguchi Inner-Outer Array Designs for Robust Design Problems.” Quality and Reliability Engineering International 16:117–128.

Box, G. E. P. (1988). “Signal–to–Noise Ratio, Performance Criteria, and Transformations.” Technometrics 30:1–40.

Box, G. E. P., and Behnken, D. W. (1960). “Some New Three-Level Designs for the Study of Quantitative Variables.” Technometrics 2:455–475.

Box, G. E. P., and Draper, N. R. (1987). Empirical Model–Building and Response Surfaces. New York: John Wiley & Sons.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). Statistics for Experimenters. New York: John Wiley & Sons.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (2005). Statistics for Experimenters: Design, Innovation, and Discovery. 2nd ed. New York: John Wiley & Sons.

Box, G. E. P., and Meyer, R. D. (1986). “An Analysis of Unreplicated Fractional Factorials.” Technometrics 28:11–18.

Box, G. E. P., and Wilson, K. B. (1951). “On the Experimental Attainment of Optimum Conditions.” Journal of the Royal Statistical Society, Series B 13:1–45.

Byrne, D. M., and Taguchi, G. (1986). “The Taguchi Approach to Parameter Designs.” Quality Congress Transactions 177:168–177.

Chow, S. -C., Shao, J., and Wang, H. (2008). Sample Size Calculations in Clinical Research. 2nd ed. Boac Raton, FL: Chapman & Hall/CRC.

Clopper, C. J., and Pearson, E. S. (1934). "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial." Biometrika 26:404–413.

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs. 2nd ed. New York: John Wiley & Sons.

Cohen, M. B., Gibbons, P. B., Mugridge, W. B., and Colbourn, C. J. (2003). "Constructing Test Suites for Interaction Testing." In Proceedings 25th International Conference on Software Engineering, 38–48.

Cohen, M. B., Dwyer, M. B., and Shi, J. (2007). "Interaction Testing of Highly-Configurable Systems in the Presence of Constraints." In Proceedings of the 2007 International Symposium on Software Testing and Analysis, 129–139.

Colbourn, C. (2004). "Combinatorial Aspects of Covering Arrays." Le Matematiche 58:121–167.

Colbourn, C., and Dinitz, J., eds. (2010). Handbook of Combinatorial Designs. 2nd ed. Boca Raton, FL: CRC Press.

Cornell, J. A. (1990). Experiments with Mixtures. 2nd ed. New York: John Wiley & Sons.

Dalal, S. R., and Mallows, C. L. (1998). "Factor-Covering Designs for Testing Software." Technometrics 40:234–243.

DuMouchel, W., and Jones, B. (1994). "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model." Technometrics 36:37–47.

Errore, A., Jones, B., Li, W., and Nachtsheim, C. (2017). "Using Definitive Screening Designs to Identify Active First- and Second-Order Factor Effects." Journal of Quality Technology 49:244–264.

Fries, A., and Hunter, W. G. (1980). "Minimum Aberration 2^{k-p} Designs." Technometrics 22:601–608.

Gotwalt, C. M., Jones, B. A., and Steinberg, D. M., (2009). "Fast Computation of Designs Robust to Parameter Uncertainty for Nonlinear Settings." Technometrics 51:88–95.

Goos, P. (2002). The Optimal Design of Blocked and Split Plot Experiments. New York: Springer.

Goos, P., and Jones, B. (2011). Optimal Design of Experiments: A Case Study Approach. New York: John Wiley & Sons.

Hamada, M., and Wu, C. F. J., (1992). "Analysis of Designed Experiments with Complex Aliasing." Technometrics 24:130–137.

Hartman, A., and Raskin, L. (2004). "Problems and Algorithms for Covering Arrays." Discrete Mathematics 284:149–156.

Horn, R., and Johnson, C. (2012). Matrix Analysis. 2nd ed. Cambridge: Cambridge University Press.

Huber, J., and Zwerina, K. (1996). "The Importance of Utility Balance in Efficient Choice Designs." Journal of Marketing Research 33:307–317.

Jennrich, R. I. (1969). "Asymptotic Properties of Non-Linear Least Squares Estimators." Annals of Mathematical Statistics 40:633–643.

Johnson, M. E., and Nachtsheim, C. J. (1983). “Some Guidelines for Constructing Exact D–Optimal Designs on Convex Design Spaces.” Technometrics 25:271–277.

Jones, B. (1991). “An Interactive Graph For Exploring Multidimensional Response Surfaces.” In Proceedings of the Annual Joint Statistical Meetings of the American Statistical Association. Alexandria. VA: American Statistical Association.

Jones, B., and Goos, P. (2007). “A Candidate-set-free Algorithm for Generating D-optimal Designs.” Applied Statistics 56:1–18. Accessed January 23, 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3001117/.

Jones, B., and Goos, P. (2015). “Optimal Design of Blocked Experiments in the Presence of Supplementary Information about the Blocks.” Technometrics 47:01–317.

Jones, B., and Johnson, R. T. (2009). “Design and Analysis for the Gaussian Process Model.” Quality and Reliability Engineering International 25:515–524. Accessed January 23, 2017. https://www.stat.osu.edu/~comp_exp/jour.club/Jones_Johnson_QREI09.pdf.

Jones, B., Lekivetz, R., Majumdar, D., Nachtsheim, C. J., and Stallings, J. W. (2019). “Construction, Properties, and Analysis of Group-Orthogonal Supersaturated Designs.” Technometrics 62: 403 - 414. Accessed December 15, 2020. https://doi.org/10.1080/00401706.2019.1654926

Jones, B., Lin, D. K. J., and Nachtsheim, C. J. (2008). “Bayesian D-Optimal Supersaturated Designs.” Journal of Statistical Planning and Inference 138:86–92.

Jones, B., and Nachtsheim, C. J. (2011a). “A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects.” Journal of Quality Technology 43:1–15. Accessed January 23, 2017. https://www.jmp.com/en_us/whitepapers/jmp/class-three-level-definitive-screening.html.

Jones, B., and Nachtsheim, C. J. (2011b). “Efficient Designs with Minimal Aliasing.” Technometrics 53:62–71.

Jones, B., and Nachtsheim, C. J. (2013). “Definitive Screening Designs with Added Two-Level Categorical Factors.” Journal of Quality Technology 45:121–129.

Jones, B., and Nachtsheim, C. J. (2016a). “Blocking Schemes for Definitive Screening Designs.” Technometrics 58:74–83.

Jones, B., and Nachtsheim, C. J. (2016b). “Effective Model Selection for Definitive Screening Designs.” Technometrics. Accessed September 28, 2017. https://www.tandfonline.com/doi/full/10.1080/00401706.2016.1234979.

Joseph, V. R., Gul, E., and Ba, S. (2015). “Maximum Projection Designs for Computer Experiments.” Biometrika 102:371–380.

Jowett, G. (1963). “The Relationship Between the Binomial and F Distributions.” Journal of the Royal Statistical Society 13:55–57.

Kessels, R., Jones, B., Goos, P. (2011). “Bayesian Optimal Designs for Discrete Choice Experiments with Partial Profiles.” Journal of Choice Modelling 4:52–74.

Kowalski, S. M., Cornell, J. A., and Vining, G. G. (2002). “Split Plot Designs and Estimation Methods for Mixture Experiments with Process Variables.” Technometrics 44:72–79.

Krishnamoorthy, K., and Mathew, T. (2009). Statistical Tolerance Regions: Theory,

Applications, and Computations. New Your: John Wiley & Sons.

Lenth, R. V. (1989). "Quick and Easy Analysis of Unreplicated Fractional Factorials." Technometrics 31:469–473.

Lekivetz, R. (2014). "What is an Alias Matrix?" Accessed January 23, 2017. https://blogs.sas.com/content/jmp/2014/06/05/what-is-an-alias-matrix/.

Lekivetz, R., Sitter, R., Bingham, D., Hamada, M. S., Moore, L. M., and Wendelberger, J. R. (2015). "On Algorithms for Obtaining Orthogonal and Near-Orthogonal Arrays for Main-Effects Screening." Journal of Quality Technology 47:2–13.

Lin, D. K. J. (1993). "A New Class of Supersaturated Design." Technometrics 35:28–31.

Martirosyan, S. (2003). "Perfect Hash Families, Identifiable Parent Property Codes and Covering Arrays." Ph.D. diss., Universitat Duisburg-Essen.

Meyer, R. K., and Nachtsheim, C. J. (1995). The Coordinate Exchange Algorithm for Constructing Exact Optimal Designs." Technometrics 37:60–69.

Meyer, R. D., Steinberg, D. M., and Box, G.(1996). Follow-up Designs to Resolve Confounding in Multifactor Experiments. Technometrics 38:307.

Miller, A., and Sitter, R. R. (2005). "Using Folded-Over Nonorthogonal Designs." Technometrics 47:502–513.

Montgomery, D. C. (2013). Design and Analysis of Experiments. 8th ed. New York: John Wiley & Sons.

Morgan, J. (2009). "Constrained Covering Arrays: Resolving Invalid Level Combination Constraints." Technical Report, Cary NC: SAS Institute Inc., also in Supplemental Proceedings of the 20th International Symposium on Software Reliability Engineering. Accessed May 6, 2018. https://www.jmp.com/content/dam/jmp/documents/en/technical-reports/ConstrainedCovArray09.pdf.

Morgan, J. (2014). "Computing Coverage and Diversity Metrics for Constrained Covering Arrays," Technical Report, Cary NC: SAS Institute Inc. Accessed May 6, 2018. https://www.jmp.com/content/dam/jmp/documents/en/technical-reports/CoverageDiversity2014.pdf.

Morgan, J. P., and Stallings, J. (2017). "Optimal experimental design that targets meaningful information." WIREs Comput Stat 2017, 9:e1393. doi: 10.1002/wics.1393.

Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2009). Response Surface Methodology, Process and Product Optimization Using Designed Experiments. 3rd ed. New York: John Wiley & Sons.

Mylona, K., Goos, P., and Jones, B. (2014). "Optimal Design of Blocked and Split-plot Experiments for Fixed Effects and Variance Component Estimation." Technometrics 56:132–144.

Nayeri, P., Colbourn, C., and Konjevod, G. (2013). "Randomized Postoptimization of Covering Arrays." European Journal of Combinatorics 34:91–103.

Piepel, G. F. (1988). “Programs for Generating Extreme Vertices and Centroids of Linearly Constrained Experimental Regions.” Journal of Quality Technology 20:125–139.

Sall, J. (2008). “Choice Experimental Designs Are Different.” Accessed January 22, 2016. https://blogs.sas.com/content/jmp/2008/12/16/choice-experimental-designs-are-different/.

SAS Institute Inc. (2013). “The %MktBIBD Macro.” Accessed September 27, 2017. https://support.sas.com/rnd/app/macros/mktbibd/mktbibd.htm.

Schoen, E. D. (1999). “Designing Fractional Two-level Experiments with Nested Error Structures.” Journal of Applied Statistics 26: 495–508.

Shewry, M. C., and Wynn, H. P. (1987). “Maximum Entropy Sampling.” Journal of Applied Statistics 14:165–170.

Snee, R. D., and Marquardt, D. W. (1974). “Extreme Vertices Designs for Linear Mixture Models.” Technometrics 16:391–408.

Snee, R. D. (1975). “Experimental Designs for Quadratic Models in Constrained Mixture Spaces.” Technometrics 17:149–159.

Snee, R. D. (1979). “Experimental Designs for Mixture Systems with Multicomponent Constraints.” Communications in Statistics -Theory and Methods 8:303–326.

Snee, R. D. (1985). “Computer Aided Design of Experiments - Some Practical Experiences.” Journal of Quality Technology 17:222–236.

Vivacqua, C. A., and Bisgaard, S. (2004). “Strip-Block Experiments for Process Improvement and Robustness.” Quality Engineering 16:495–500.

Worley, B. A. (1987). Deterministic Uncertainty Analysis. Technical Report ORN-0628, Oak Ridge National Laboratory.

Wu, C. F. J. (1981). “Asymptotic Theory of Nonlinear Least Squares Estimation.” Annals of Mathematical Statistics 9:501–513.

Wu, C. F. J., and Hamada, M. S. (2009). Experiments: Planning, Analysis, and Optimization. 2nd ed. New York: John Wiley & Sons.

Xiao, L., Lin, D. K. J., and Bai, F. (2012). “Constructing Definitive Screening Designs Using Conference Matrices.” Journal of Quality Technology 44:2–8.

Ye, K. Q., and Hamada, M. (2000). “Critical Values of the Lenth Method for Unreplicated Factorial Designs.” Journal of Quality Technology 32:57–66.

Yilmaz, C., Fouche, S., Cohen, M. B., Porter, A., Demiroz, G., and Koc, U. (2014). “Moving Forward with Combinatorial Interaction Testing.” Computer 47:37–45.

Zahran, A., Anderson-Cook, C. M., and Myers, R. H. (2003). “Fraction of Design Space to Assess the Prediction Capability of Response Surface Designs.” Journal of Quality Technology 35:377–386.

Zhang, Z., and Zhang, J. (2011). “Characterizing Failure-Causing Parameter Interactions by Adaptive Testing.” In Proceedings of the 2011 International Symposium on Software Testing and Analysis, 331–341.

부록 D

기술 라이선스 표시

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS “AS IS”AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

버전 16

선형 모형 적합

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는 것이 아니라 새로운 눈을 가지는 데 있다."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.0

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 16 선형 모형 적합. Cary, NC: SAS Institute Inc.

JMP® 16 선형 모형 적합

Copyright © 2020-2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP를 처음 사용하든 오랫동안 사용해왔든, JMP와 관련해서 배워야 하는 것은 언제나 있습니다.

JMP.com을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다.

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티, 추가 기능 및 스크립트 예를 포함한 사용자용 리소스, 포럼, 블로그, 컨퍼런스 정보 등

https://www.jmp.com/getstarted

목차
선형 모형 적합

1 JMP 알아보기 15

설명서 및 추가 리소스

JMP 설명서에 사용되는 서식 규칙 17
JMP 도움말 18
JMP 설명서 라이브러리 18
JMP 학습을 위한 추가 리소스 24
JMP 자습서 24
샘플 데이터 테이블 24
통계 및 JSL 용어에 대해 알아보기 25
JMP 팁 및 힌트 알아보기 25
JMP 툴팁 25
JMP 사용자 커뮤니티 25
무료 온라인 Statistical Thinking 교육 과정 26
JMP 새로운 사용자를 위한 입문 키트 26
Statistics Knowledge 포털 26
JMP 교육 과정 26
사용자가 작성한 JMP 설명서 26
JMP 시작하기 창 26
JMP 기술 지원 27

2 모형 규격 29

선형 모형 지정

모형 적합 플랫폼 개요 31
모형 적합을 사용한 회귀 분석의 예 32
모형 적합 플랫폼 시작 34
모형 적합 시작 창 요소 35
모형 효과 생성 37
적합 분석법 43
모형 규격 옵션 45
결측값 정보화 46
타당도 검사 47
모형 규격 및 모형 적합의 예 47

단순 선형 회귀 48
X 다항식 (k 차) 49
X 및 Z 다항식 (k 차) 50
다중 선형 회귀 50
일원 분산 분석 51
이원 분산 분석 52
교호작용이 있는 이원 분산 분석 53
삼원 완전 요인 55
공분산 분석 , 동일한 기울기 57
공분산 분석 , 서로 다른 기울기 58
2 요인 내포 임의 효과 모형 60
3 요인 완전 내포 임의 효과 모형 62
단순 분할구 또는 반복 측정 모형 63
2 요인 반응 표면 모형 64
매듭이 지정된 스플라인 효과 66

3 표준 최소 제곱 모형 69

모형의 공통 클래스 분석

예 : 표준 최소 제곱 사용 72
표준 최소 제곱 분석법 시작 74
모형 적합 시작 창 75
모형 적합 시작 창의 표준 최소 제곱 옵션 76
표준 최소 제곱의 검증 78
결측값 78
최소 제곱 적합 보고서 79
단일 반응과 다중 반응의 비교 79
강조와 관련된 보고서 구조 79
특수 보고서 79
최소 제곱 적합 옵션 83
적합 그룹 옵션 83
반응 옵션 84
회귀 보고서 85
적합 요약 85
분산 분석 86
모수 추정값 87
효과 검정 88
효과 상세 정보 89
적합 결여 104
추정값 105
예측 표현식 표시 107

정렬된 추정값 108
확장 추정값 111
표시자 파라미터화 추정값 113
순차 검정 114
사용자 검정 115
다중 비교 117
기울기 비교 130
결합 요인 검정 130
역추정 예측 131
Cox 혼합물 135
모수 검정력 136
추정값 상관계수 138
명목형 효과에 대한 코딩 139
효과 선별 140
척도화된 추정값 및 연속형 항 코딩 141
그림 옵션 142
정규 확률도 보고서 146
Bayes 그림 보고서 147
파레토도 보고서 149
요인 프로파일링 150
프로파일러 151
교호작용 그림 151
등고선 프로파일러 153
혼합물 프로파일러 154
입방체도 155
Y 의 Box-Cox 변환 156
표면 프로파일러 158
행 진단 159
레버리지 그림 161
Press 164
열 저장 165
예측 계산식 167
효과 요약 보고서 168
혼합 및 임의 효과 모형 보고서 및 옵션 172
혼합 모형 및 임의 효과 모형 172
REML(제한 최대 가능도) 방법 176
EMS(전통적) 모형 적합 보고서 181
모형 항 사이에 선형 종속성이 있는 모형 182
특이성 상세 정보 183
모수 추정값 보고서 184

효과 검정 보고서 ... 185
예 ... 185
표준 최소 제곱 분석법에 대한 통계 상세 정보 ... 187
강조 규칙 ... 187
사용자 검정 예에 대한 상세 정보 ... 188
추정값 상관계수 ... 189
레버리지 그림 상세 정보 ... 189
Kackar-Harville 수정 ... 192
검정력 분석 ... 193

4 표준 최소 제곱 예 ... 203
모형의 공통 클래스 분석
일원 분산 분석 예 ... 205
동일한 기울기의 공분산 분석 예 ... 207
서로 다른 기울기의 공분산 분석 예 ... 209
반응 표면 모형 예 ... 212
전체 반응 표면 모형 적합 ... 212
모형 축소 ... 213
반응 표면 보고서 검토 ... 214
예측 프로파일러를 사용하여 임계점 찾기 ... 216
등고선 프로파일러를 사용하여 표면 보기 ... 217
분할구 설계 예 ... 219
임의 효과 모수 추정 예 ... 223
매듭이 지정된 스플라인 효과 예 ... 225
활성 요인에 대한 Bayes 그림 예 ... 227

5 단계별 회귀 모형 ... 229
변수 선택을 사용하여 모형 찾기
단계별 회귀 개요 ... 231
예 : 단계별 회귀 사용 ... 231
단계별 보고서 ... 233
단계별 플랫폼 옵션 ... 233
단계별 회귀 제어 패널 ... 234
현재 추정값 보고서 ... 240
단계 기록 보고서 ... 241
교차 , 교호작용 또는 다항식 항이 있는 모형 ... 242
결합 규칙의 예 ... 242
범주형 및 순서형 효과가 있는 모형 ... 244
계층 항 생성 ... 244
명목형 항이 있는 모형의 예 ... 245

계층 항에 대한 제한 규칙의 예 249
이항 및 순서형 로지스틱 단계별 회귀 수행 251
예 : 로지스틱 단계별 회귀 사용 252
모든 가능 모형 옵션 253
예 : 모든 가능 모형 옵션 사용 254
모형 평균화 옵션 255
예 : 모형 평균화 옵션 사용 256
단계별 회귀의 검증 옵션 257
단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합 257
단계별 회귀의 K 폴드 교차 검증 260

6 일반화 회귀 모형 263
변수 선택 기법을 사용하여 모형 생성

일반화 회귀 분석법 개요 265
일반화 회귀의 예 266
일반화 회귀 분석법 시작 269
분포 270
일반화 회귀 보고서 창 277
일반화 회귀 보고서 옵션 278
모형 시작 제어판 278
반응 분포 279
추정 방법 옵션 279
고급 제어 283
검증 방법 옵션 286
조기 중지 287
시작 287
모형 적합 보고서 287
회귀 그림 288
모형 요약 288
추정 상세 정보 291
해 경로 291
중심화 및 척도화된 예측 변수에 대한 모수 추정값 294
원래 예측 변수에 대한 모수 추정값 295
활성 모수 추정값 296
효과 검정 297
모형 적합 옵션 297
일반화 회귀 분석법에 대한 통계 상세 정보 305
추정 방법에 대한 통계 상세 정보 306
고급 제어에 대한 통계 상세 정보 308
분포에 대한 통계 상세 정보 309

7 일반화 회귀 예 315
정규화 기법을 사용하여 모형 생성
Poisson 일반화 회귀 예 317
이항 일반화 회귀 예 319
영과잉 Poisson 회귀 예 321
가로형 데이터에 대한 일반화 회귀 예 323

8 혼합 모형 327
평균 및 공분산 공동 모델링
혼합 모형 분석법 개요 330
예 : 혼합 모형 사용 330
혼합 모형 분석법 시작 335
모형 적합 시작 창 335
데이터 형식 340
혼합 적합 보고서 341
적합 통계량 343
임의 효과 공분산 모수 추정값 345
고정 효과 모수 추정값 345
반복 효과 공분산 모수 추정값 346
랜덤 계수 347
임의 효과 예측 347
고정 효과 검정 347
순차 검정 348
다중 비교 348
기울기 비교 349
주변 모형 추론 349
실제값 대 예측값 그림 349
잔차 그림 349
Profilers 350
조건부 모형 추론 351
실제값 대 조건부 예측값 그림 351
조건부 잔차 그림 351
조건부 프로파일러 352
변동도 352
공분산 및 상관 행렬 353
열 저장 354
혼합 모형 분석법의 추가 예 354
반복 측정 예 355
분할구 예 370
공간 예 : 균일도 시험 376

상관관계가 있는 반응의 예 384
혼합 모형 분석법에 대한 통계 상세 정보 390
수렴 스코어 검정 391
랜덤 계수 모형 392
반복 측정 393
반복 공분산 구조 393
공간 및 시간 변동 398
Kackar-Harville 수정 401

9 다변량 반응 모형 403
MANOVA 를 사용하여 관계 적합
다중 반응 모형의 예 405
MANOVA 보고서 407
MANOVA 적합 옵션 407
반응 지정 408
반응 선택 옵션 408
사용자 검정 옵션 409
모형 효과 옵션 409
다변량 검정 413
확장된 다변량 보고서 414
다변량 검정 비교 415
단변량 검정 및 구형성 검정 415
반복 측정을 사용하는 다변량 모형 416
반복 측정 예 417
복합 다변량 모형의 예 418
판별 분석 421
판별 저장 옵션의 예 421
MANOVA 분석법에 대한 통계 상세 정보 422
다변량 검정 422
근사 F- 검정 423
정준 상세 정보 424

10 로그 선형 분산 모형 427
반응의 분산 및 평균 모델링
로그 선형 분산 모형 개요 429
산포 효과 429
모형 규격 429
참고 430
로그 선형 분산 분석법 시작 430
예 : 로그 선형 분산 사용 431

로그 선형 보고서 433
로그 선형 플랫폼 옵션 434
열 저장 435
행 진단 435
잔차 검토 435
적합 모형 프로파일링 436
적합 모형 프로파일링의 예 436

11 로지스틱 회귀 모형 439
명목형 또는 순서형 반응에 대한 회귀 모형 적합

명목형 및 순서형 로지스틱 분석법 개요 441
명목형 로지스틱 회귀 441
순서형 로지스틱 회귀 441
로지스틱 회귀 모형을 적합시키는 기타 JMP 플랫폼 442
로지스틱 회귀의 예 442
명목형 로지스틱 회귀의 예 442
순서형 로지스틱 회귀의 예 444
명목형 및 순서형 로지스틱 분석법 시작 446
로지스틱 회귀 모형의 검증 447
로지스틱 적합 보고서 448
전체 모형 검정 449
적합 상세 정보 450
적합 결여 검정 450
로지스틱 적합 플랫폼 옵션 451
명목형 및 순서형 적합 옵션 451
명목형 적합 옵션 453
순서형 적합 옵션 454
로지스틱 회귀의 추가 예 455
역추정 예측의 예 455
명목형 로지스틱 모형에 대한 효과 요약의 사용 예 457
2 차 순서형 로지스틱 모형의 예 459
예 : 여러 열의 개수 쌓기 462
명목형 및 순서형 로지스틱 분석법에 대한 통계 상세 정보 463
로지스틱 회귀 모형 463
승산비 464
통계적 검정의 관계 465

12 일반화 선형 모형 467
비정규 반응 분포에 대한 모형 적합

일반화 선형 모형 분석법 개요 469

일반화 선형 모형의 예 469
일반화 선형 모형 분석법 시작 472
일반화 선형 모형 적합 보고서 473
전체 모형 검정 474
일반화 선형 모형 적합 보고서 옵션 475
일반화 선형 모형 분석법의 추가 예 477
대비를 사용하여 변수 수준의 차이 비교 477
오프셋을 사용하는 Poisson 회귀 478
로그 연결을 사용하는 정규 회귀 480
일반화 선형 모형 분석법에 대한 통계 상세 정보 482
모형 선택 및 이탈도 484

A 통계 상세 정보 487

선형 모형 적합

반응 모형 489
연속형 반응 489
명목형 반응 489
순서형 반응 491
요인 모형 492
연속형 요인 492
명목형 요인 493
순서형 요인 503
빈도 508
일반적인 가정 509
가정된 모형 509
상대적 유의성 509
다중 추론 509
타당도 평가 509
대체 방법 510
핵심 통계적 개념 510
불확도, 통합 개념 511
두 개의 기본 적합 기계 511
가능도, AICc 및 BIC 514
검정력 계산 515
LSN 계산 516
LSV 계산 516
검정력 계산 517
조정 검정력 계산 518
신뢰 한계를 사용한 역추정 예측 519

B 참조 자료 521
C 기술 라이선스 표시 527

1 장

JMP 알아보기
설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 17
JMP 도움말 18
JMP 설명서 라이브러리 18
JMP 학습을 위한 추가 리소스 24
JMP 자습서 24
샘플 데이터 테이블 24
통계 및 JSL 용어에 대해 알아보기 25
JMP 팁 및 힌트 알아보기 25
JMP 툴팁 25
JMP 사용자 커뮤니티 25
무료 온라인 Statistical Thinking 교육 과정 26
JMP 새로운 사용자를 위한 입문 키트 26
Statistics Knowledge 포털 26
JMP 교육 과정 26
사용자가 작성한 JMP 설명서 26
JMP 시작하기 창 26
JMP 기술 지원 27

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 Lucida Sans Typewriter 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 "JMP 설명서 라이브러리 "(18 페이지) 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
Quality and Process Methods	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기" 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기

" 창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

2 장

모형 규격
선형 모형 지정

모형 적합 플랫폼을 사용하면 복잡한 모형을 효율적으로 지정할 수 있습니다. 매크로, 속성 및 변환을 통해 작업이 간소화됩니다. 모형 적합은 다양한 모형과 효과 구조를 적합시키는 방법입니다.

여기에는 다음과 같은 내용이 포함됩니다 .

- 단순 및 다중 선형 회귀
- 분산 및 공분산 분석
- 임의 효과 , 내포 효과 , 혼합 효과 , 반복 측정 및 분할구 모형
- 명목형 및 순서형 로지스틱 회귀
- MANOVA(다변량 분산 분석)
- 정준 상관 및 판별 분석
- 로그 선형 분산 (평균 및 분산 모델링)
- GLM(일반화 선형 모형)
- 모수 생존 및 비례 위험
- 반응 변수 선별 (많은 수의 반응 연구)

JMP PRO JMP Pro에서는 다음 항목도 적합시킬 수 있습니다.

- 다양한 공분산 구조를 가진 혼합 모형
- Elastic Net, Lasso 및 능형 회귀를 포함한 일반화 회귀 모형
- 부분 최소 제곱

모형 적합 플랫폼을 사용하면 원하는 분석법을 선택하여 다양한 유형의 모형을 적합시킬 수 있습니다. 이 장에서는 대부분의 분석법에 공통된 "모형 규격" 창의 요소에 대해 중점적으로 다룹니다.

목차

모형 적합 플랫폼 개요 31
모형 적합을 사용한 회귀 분석의 예 32
모형 적합 플랫폼 시작 34
모형 적합 시작 창 요소 35
모형 효과 생성 37
적합 분석법 43
모형 규격 옵션 45
결측값 정보화 46
타당도 검사 47
모형 규격 및 모형 적합의 예 47
단순 선형 회귀 48
X 다항식(k차) 49
X 및 Z 다항식(k차) 50
다중 선형 회귀 50
일원 분산 분석 51
이원 분산 분석 52
교호작용이 있는 이원 분산 분석 53
삼원 완전 요인 55
공분산 분석, 동일한 기울기 57
공분산 분석, 서로 다른 기울기 58
2요인 내포 임의 효과 모형 60
3요인 완전 내포 임의 효과 모형 62
단순 분할구 또는 반복 측정 모형 63
2요인 반응 표면 모형 64
매듭이 지정된 스플라인 효과 66

모형 적합 플랫폼 개요

모형 적합 플랫폼을 사용하면 효과 구조가 복잡한 모형을 효율적으로 지정할 수 있습니다. 이러한 효과 구조는 모형 모수에 대해 선형입니다. 모형을 지정한 후에는 여러 적합 **분석법** 중에서 적절한 적합 기법을 선택할 수 있습니다. 분석법을 선택하면 해당 분석법과 관련된 선택 항목이 "모형 적합" 창에 제공됩니다. 이 장에서는 대부분의 분석법에 공통된 "모형 규격" 창의 요소에 대해 중점적으로 다룹니다. 모든 분석법에 대한 설명은 "모형 적합 시작 창 요소"(35 페이지)에서 확인하십시오.

모형 적합은 여러 방법을 사용하여 적합시킬 수 있는 다양한 모형을 지정하는 데 사용할 수 있습니다. 표 2.1에는 모형 적합을 사용하여 정의할 수 있는 몇 가지 일반적인 모형이 나열되어 있습니다. 표에서 X 및 Z 효과는 연속형 모델링 유형의 열을 나타내고 A, B, C 효과는 명목형 또는 순서형 모델링 유형의 열을 나타냅니다.

이러한 모형 효과를 생성하는 클릭 시퀀스, 모형 적합 그림 및 몇 가지 예는 "모형 규격 및 모형 적합의 예"(47 페이지)에서 확인하십시오.

표 2.1 표준 모형 유형

모형 유형	모형 효과
단순 선형 회귀	X
X 다항식(k 차)	X, X*X,..., X^k
X 및 Z 다항식(k 차)	X, X*X,..., X^k, Z, Z*Z,..., Z^k
다중 선형 회귀	X, Z 및 기타 연속형 열
일원 분산 분석	A
이원 분산 분석	A, B
교호작용이 있는 이원 분산 분석	A, B, A*B
삼원 완전 요인	A, B, C, A*B, A*C, B*C, A*B*C
공분산 분석, 동일한 기울기	A, X
공분산 분석, 서로 다른 기울기	A, X, A*X
2 요인 내포 임의 효과 모형	A, B[A]& 랜덤
3 요인 완전 내포 임의 효과 모형	A, B[A]& 랜덤, C[A,B]& 랜덤
단순 분할구 또는 반복 측정 모형	A, B[A]& 랜덤, C, C*A
2 요인 반응 표면 모형	X&RS, Z&RS, X*X, X*Z, Z*Z

모형 적합을 사용한 회귀 분석의 예

유산소 피트니스 연구의 결과 데이터가 있으며 몇 가지 연속형 변수에서 산소 섭취량을 예측하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 . " 분석법 " 상자가 비어 있습니다 .
3. Oxy 를 선택하고 **Y** 를 클릭합니다 .

   연속형 반응을 지정하면 " 분석법 " 이 기본적으로 " 표준 최소 제곱 " 으로 설정되지만 다른 분석법을 선택할 수 있습니다 . " 강조 " 는 기본적으로 " 효과 레버리지 " 로 설정됩니다 .
4. Ctrl 키를 누르고 Sex, Age, Weight, Runtime, RunPulse, RstPulse 및 MaxPulse 를 선택합니다 . **추가**를 클릭하여 " 모형 효과 생성 " 목록에 추가합니다 . 이 창을 나중에 사용할 수 있게 하려면 **대화상자 열린 채 유지**를 선택할 수 있습니다 . " 모형 규격 " 창이 그림 2.1 에 표시된 것처럼 보입니다 .
5. **실행**을 클릭합니다 . 그림 2.2 에 보고서의 일부분이 나와 있습니다 .

그림 2.1 Fitness 회귀 모형에 대한 모형 규격 창

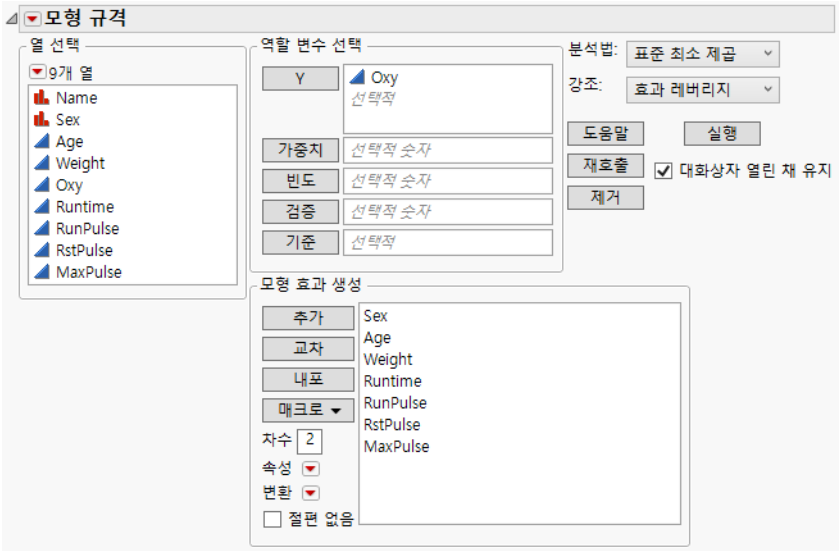
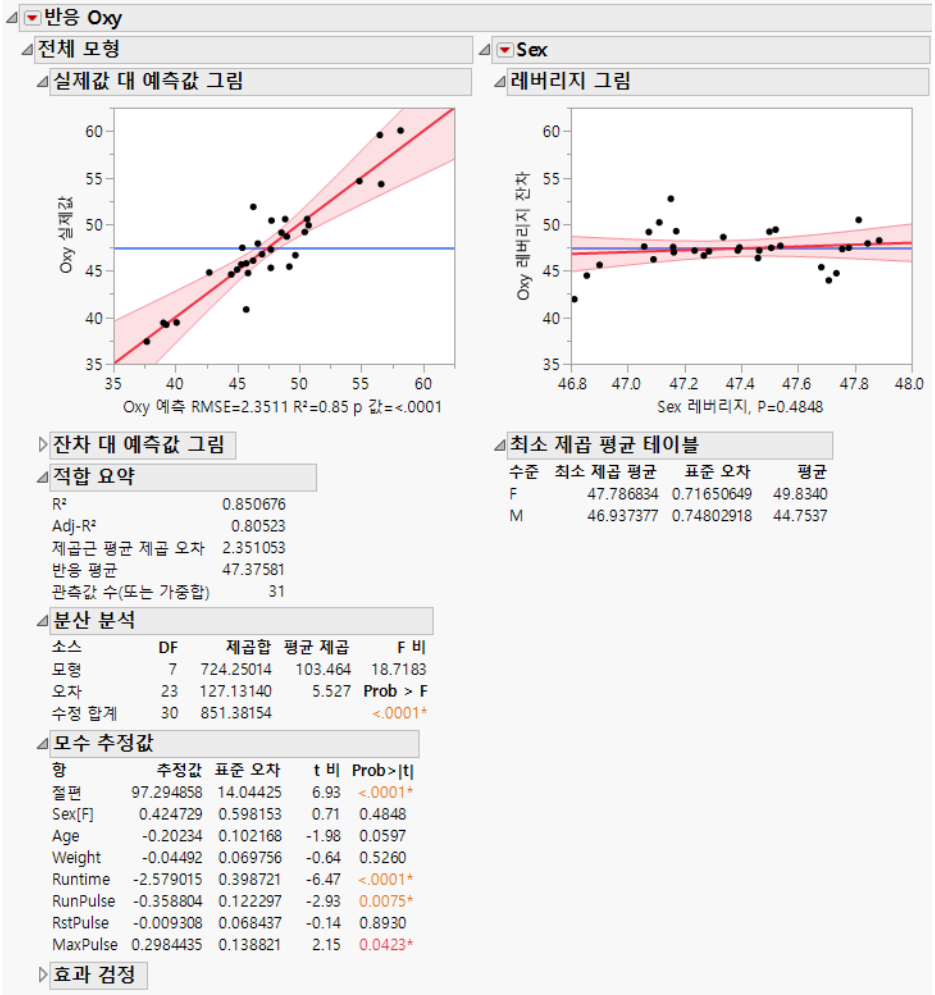


그림 2.2 Fitness 데이터에 대한 표준 최소 제곱 보고서 (일부분)



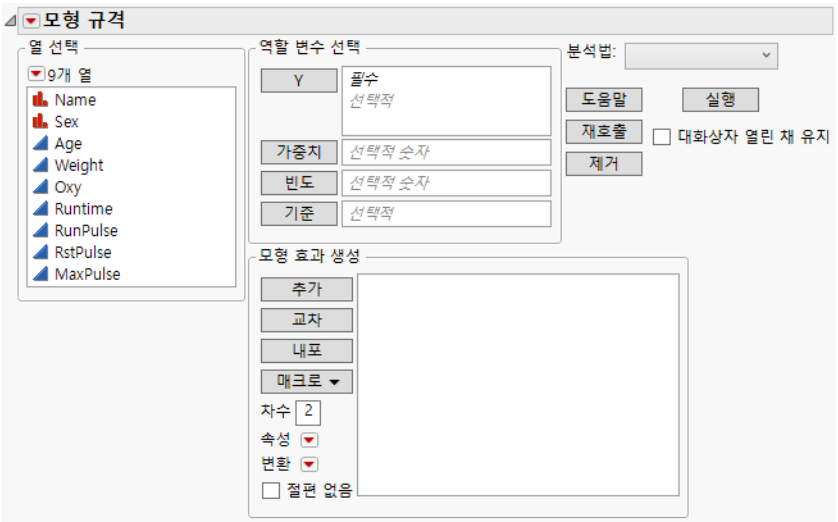
전체 모형에 대한 그림과 보고서는 맨 왼쪽 보고서 열에 나타납니다. 오른쪽 열에는 모형에 지정한 각 효과의 레버리지 그림이 표시됩니다. 공간이 제한되어 그림 2.2에는 Sex 열만 표시되지만 다른 6가지 효과에 대한 열도 보고서에 표시됩니다. 빨간색 삼각형 메뉴에는 보고서 창에 보고서와 그림을 추가하는 기타 옵션이 포함되어 있습니다. 표준 최소 제곱 보고서 창에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "최소 제곱 적합 보고서"(79 페이지)에서 확인하십시오.

"모수 추정값" 보고서의 p 값을 살펴보면 Runtime, RunPulse 및 MaxPulse가 산소 섭취량의 유의한 예측 변수임을 알 수 있습니다. 다음 단계에서는 유의하지 않은 예측 변수를 제거하여 모형을 축소할 수 있습니다. 자세한 내용은 "단계별 회귀 모형" 장(229 페이지)에서 확인하십시오.

모형 적합 플랫폼 시작

분석 > 모형 적합을 선택하여 모형 적합 플랫폼을 시작합니다. 그림 2.3에서는 Fitness.jmp 샘플 데이터 테이블에 대한 시작 창의 예를 보여 줍니다. 모형 적합 시작 창은 분석법이라고 하는 많은 플랫폼의 공통 시작 창입니다. 자세한 내용은 "적합 분석법"(43 페이지)에서 확인하십시오.

그림 2.3 모형 적합 시작 창



참고 : Model 또는 model 스크립트가 있는 데이터 테이블에서 **분석 > 모형 적합**을 선택하면 해당 스크립트를 기반으로 시작 창이 자동으로 채워집니다.

모형 적합 시작 창에는 다음과 같은 주 영역이 있습니다.

- 열 선택 - 현재 데이터 테이블의 열 목록입니다. 현재 데이터 테이블에 제외된 열이 포함되어 있으면 해당 열은 목록에 나타나지 않습니다. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.
- 역할 변수 선택 - 표준 JMP 시작 창 버튼이 포함되어 있습니다. 이 버튼에 대한 설명은 "모형 적합 시작 창 요소"(35 페이지)에서 확인하십시오.
- 모형 효과 생성 - 모형에 효과를 입력하는 데 사용하는 옵션이 포함되어 있습니다. 자세한 내용은 "모형 효과 생성"(37 페이지)에서 확인하십시오.
- 분석법 - 모형 유형을 선택할 수 있는 목록입니다. 분석법을 선택하면 선택한 분석법에 따라 서로 다른 옵션이 나타납니다. 자세한 내용은 "적합 분석법"(43 페이지)에서 확인하십시오.

모형 적합 시작 창 요소

다음 모형 적합 시작 창 요소는 대부분의 분석법에 공통적입니다 .

모형 규격 " 모형 규격 " 메뉴에는 다음 옵션이 포함되어 있습니다 .

중심화 다항식 모형에 다항식이 포함된 경우 효과를 중심화합니다 .

결측값 정보화 효과에 결측값을 수용하는 코딩 시스템을 생성합니다 .

유의 수준 설정 모형 보고서에 신뢰 구간의 유의 수준을 설정합니다 .

데이터 테이블에 저장 모형 규격을 현재 데이터 테이블의 스크립트에 저장합니다 .

스크립트 창에 저장 모형 규격을 스크립트 창에 저장합니다 .

SAS 작업 생성 현재 모형 규격에 대한 SAS 코드를 SAS 프로그램 창에 저장합니다 .

SAS 로 전송 현재 모형 규격에 대한 SAS 코드를 SAS 세션으로 전송합니다 .

수렴 설정 일부 분석법에서 모형 수렴에 사용되는 최대 반복 수 및 수렴 한계를 설정하는 옵션이 포함되어 있습니다 .

이러한 옵션에 대한 자세한 내용은 " 모형 규격 옵션 "(45 페이지) 에서 확인하십시오 .

열 선택 현재 데이터 테이블에서 제외되지 않은 열을 나열합니다 .

Y 모형에 대한 하나 이상의 반응 변수 (종속 변수) 를 식별합니다 .

참고 : 벡터 값이 있는 반응 열은 모형 적합에서 지원되지 않습니다 .

가중치 분석을 위해 각 행에 가중치를 할당하는 값이 들어 있는 열입니다 . 자세한 내용은 " 가중치 "(36 페이지) 에서 확인하십시오 .

빈도 분석을 위해 각 행에 빈도를 할당하는 값이 들어 있는 열입니다 . 일반적으로 빈도 열의 효과는 데이터 테이블을 확장하여 정수 빈도가 *k* 인 행을 *k* 개의 행으로 확장하는 것입니다 . 빈도 비율로 지정할 수 있습니다 . 자세한 내용은 " 빈도 "(36 페이지) 에서 확인하십시오 .

JMP PRO 검증 JMP Pro 에서는 일부 분석법에 대해 검증 열을 입력할 수 있습니다 . 각 분석법에서 검증 열을 처리하는 방법에 대한 자세한 내용은 해당 분석법 관련 장에서 확인하십시오 ." 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다 ." 검증 열 생성 " 유틸리티에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 . JMP 모델링 플랫폼에서 검증 열이 사용되는 방법에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

추가 모형에 효과를 추가합니다 . 자세한 내용은 " 추가 "(37 페이지) 에서 확인하십시오 .

교차 둘 이상의 변수를 교차시켜 교호작용 및 다항식 효과를 생성합니다 . 자세한 내용은 " 교차 "(37 페이지) 에서 확인하십시오 .

내포 내포 효과를 생성합니다 . 자세한 내용은 " 내포 "(38 페이지) 에서 확인하십시오 .

매크로 일반적으로 사용되는 모형에 대한 효과를 생성합니다 . 자세한 내용은 " 매크로 "(39 페이지) 에서 확인하십시오 .

차수 매크로를 사용하여 생성된 요인 또는 다항식 효과가 있는 모형에 지정된 차수를 적용합니다 . 자세한 내용은 " 매크로 "(39 페이지) 의 **특정 차수까지의 요인** 및 **특정 차수까지의 다항식** 에서 확인하십시오 .

속성 모형 효과에 속성을 적용합니다 . 이러한 속성은 효과를 처리하는 방법을 결정합니다 . 자세한 내용은 " 속성 "(40 페이지) 에서 확인하십시오 .

변환 선택한 연속형 효과 또는 *Y* 열을 변환합니다 . 자세한 내용은 " 변환 "(41 페이지) 에서 확인하십시오 .

절편 없음 모형에서 절편 항을 제외합니다 .

분석법 적합 방법론을 지정합니다 . 자세한 내용은 " 모형 적합 시작 창 요소 "(35 페이지) 에서 확인하십시오 . 선택한 분석법에 따라 다른 옵션이 나타납니다 .

목표 수준 (특정 분석법에서 *Y* 가 이항 변수이고 명목형 모델링 유형일 때만 사용 가능) 확률을 모델링할 수준을 지정합니다 . 기본값은 수준 순서에 따라 두 수준 중 더 높은 값입니다 .

도움말 모형 적합 시작 창의 도움말 항목으로 이동합니다 .

재호출 최근에 실행한 모형 규격으로 시작 창을 채웁니다 .

제거 선택한 변수를 할당된 역할에서 제거합니다 . 또는 효과를 두 번 클릭하거나 효과를 선택하고 Backspace 키를 누릅니다 .

실행 지정된 모형 및 분석법에 대한 보고서 창을 생성합니다 .

대화상자 열린 채 유지 분석을 실행한 후 언제든지 분석을 변경하고 다시 실행할 수 있도록 시작 창을 열린 채로 유지합니다 .

빈도

빈도 텍스트 상자에 입력된 빈도 변수는 대부분의 모형 적합 분석법에서 지원됩니다. 일반적으로 빈도는 다음과 같은 방식으로 해석됩니다. 빈도가 f인 행이 있다고 가정해 보겠습니다. 그러면 계산된 결과는 각각 빈도가 1인 해당 행의 f 개 복사본을 포함하는 데이터 테이블의 결과와 동일합니다.

빈도 값이 0이거나 결측인 행은 분석에서 제외됩니다. 빈도 값이 음수인 행은 중도절단된 관측값에 대해서만 허용되며 그렇지 않으면 분석에서 제외됩니다. 음수 빈도 값을 중도절단된 관측값과 함께 사용하면 절단된 분포를 적합시킬 수 있습니다.

빈도 값은 정수일 필요가 없습니다. 값이 정수가 아닌 열을 포함하여 빈도 열을 처리하는 방법을 설명하는 기술 상세 정보는 "빈도"(508 페이지)에서 확인하십시오.

가중치

가중치 변수는 관측값의 분산이 서로 다른 경우에 유용할 수 있습니다. 예를 들어 각 행이 사전 요약된 평균으로 구성된 데이터에 대해 회귀 모델링을 수행할 때 이러한 현상이 발생할 수 있습니다. 여기서 관측값을 더 많이 포함하는 행(분산이 더 작음)이 관측값을 더 적게 포함하는 행(분산

이 더 큼)보다 손실 함수에 더 크게 기여해야 합니다. 적절하게 정의된 가중치를 사용하여 이를 확인할 수 있습니다.

가중치 변수는 많은 모형 적합 분석법에서 지원됩니다. 가중치 변수를 지원하는 각 분석법은 다음 방법 중 하나를 사용합니다.

- 분산 척도
- 빈도 대칭

분산 척도

최소 제곱 또는 최대 가능도 정규 이론을 사용하여 추정을 수행할 때 주어진 행의 가중치 w는 손실 함수에 대한 해당 행의 기여도를 $w^{-1/2}$로 조정합니다.

가중치 변수는 추정값과 표준 오차에 영향을 줍니다. 그러나 빈도 변수와 달리 가설 검정에 사용되는 자유도에는 영향을 주지 않습니다.

가중치 값이 음수이거나 0인 행은 분석에서 제외됩니다.

빈도 대칭

명목형 로지스틱 및 순서형 로지스틱 분석법에서 가중치 변수는 빈도 변수인 것처럼 처리됩니다. 이러한 분석법에서 가중치 변수와 빈도 변수가 모두 지정되면 각 관측값의 빈도 값이 가중치 값과 빈도 값의 곱과 동일한 것으로 처리합니다.

모형 효과 생성

이 섹션에서는 모형에 효과를 쉽게 입력하기 위해 사용할 수 있는 옵션에 대해 설명합니다. 이러한 옵션을 사용하여 특정 유형의 모형을 생성하는 방법에 대한 예는 "모형 규격 및 모형 적합의 예"(47 페이지)에서 확인하십시오.

추가

모형에 효과를 추가합니다. 이러한 효과는 "열 선택" 목록에서 직접 추가하거나 해당 목록에서 선택한 후 "매크로" 또는 "속성"을 사용하여 수정할 수 있습니다. "교차" 또는 "내포"를 사용하여 효과를 생성하고 추가하거나 수정할 수도 있습니다. 효과에 포함된 변수의 모델링 유형과 효과에 할당된 속성에 따라 해당 효과가 모형에서 처리되는 방식이 결정됩니다.

참고 : "모형 효과 생성" 목록에서 효과를 제거하려면 효과를 두 번 클릭하거나 선택한 후 **제거**를 클릭하거나 Backspace 키 또는 Delete 키를 누르십시오.

교차

교호작용 또는 다항식 효과를 생성합니다. "열 선택" 목록에서 두 개 이상의 변수를 선택하고 **교차**를 클릭합니다. 아니면 "열 선택" 목록에서 하나 이상의 변수를 선택하고 "모형 효과 생성" 목록에서 하나 이상의 효과를 선택한 후 **교차**를 클릭합니다.

교차효과를 파라미터화하고 코드화하는 방법에 대한 자세한 내용은 "통계 상세 정보"(487 페이지)에서 확인하십시오.

참고 : 최대 10 개의 열을 결합하는 효과를 교차 및 내포로 생성할 수 있습니다 .

교차효과의 예

제품 코팅에 염료를 도포해야 한다고 가정해 보겠습니다. 염료 pH(Dye pH)와 염료 농도(Dye Concentration) 둘 다 코팅 색상에 영향을 주는 것 같습니다. 이러한 요인 효과를 알아보기 위해 염료 pH 및 염료 농도가 각각 높은 수준과 낮은 수준으로 설정된 실험을 설계합니다. 염료 pH가 색상에 미치는 영향은 염료 농도의 수준이 낮을 때보다 높을 때 더 뚜렷하게 나타날 수 있습니다. 이를 **교호작용**이라고 합니다. 가능한 이 교호작용을 모델링하려면 "모형 효과 생성" 목록에서 교차 항(Dye pH * Dye Concentration)을 포함합니다. 이렇게 하면 JMP에서 교호작용을 검정할 수 있습니다.

내포

내포 효과를 생성합니다. 한 효과(B)의 수준이 다른 효과(A)의 단일 수준 내에서만 발생하는 경우 B가 A 안에 **내포**된다고 합니다. 일반적으로 B[A] 표기가 사용되며 이는 "B가 A 안에 내포되어 있다"는 의미입니다. 내포는 계층 관계를 정의합니다. A를 **외부** 효과라고 하고 B를 **내부** 효과라고 합니다. 내포된 항은 범주형이어야 합니다.

참고 : 내포 항은 외부에서 내부 순으로 지정해야 합니다 . 예를 들어 B 가 A 안에 내포되고 C 가 B 안에 내포된 경우 모형은 A, B[A], C[B,A]((A, B[A], C[A,B] 와 같음) 로 지정됩니다 . 최대 10 개의 열을 결합하는 효과를 교차 및 내포로 생성할 수 있습니다 .

내포 효과의 예

두 학교의 각 수학 교사를 내포의 예로 들어보겠습니다 . 한 학교에는 세 명의 수학 교사가 있고 다른 학교에는 두 명의 수학 교사가 있습니다 . 각 학교의 각 교사는 중복되지 않는 학생 그룹으로 구성된 2 ~ 3 개의 학급을 가르칩니다 . 이 예에서 학급 (C) 은 교사 (B) 에 내포되고 교사 (B) 는 학교 (A) 에 내포됩니다 . " 모형 적합 " 시작 창에서 다음과 같이 효과를 입력합니다 .

1. A 와 B 를 모두 " 모형 효과 생성 " 패널에 추가합니다 .
2. " 모형 효과 생성 " 패널에서 B 를 선택합니다 .
3. " 열 선택 " 목록에서 A 를 선택합니다 .
4. **내포**를 클릭합니다 . 그러면 B 가 B[A] 효과로 변환됩니다 .
5. C 를 " 모형 효과 생성 " 패널에 추가합니다 .
6. " 모형 효과 생성 " 패널에서 C 를 선택합니다 .
7. " 열 선택 " 목록에서 A 와 B 를 선택합니다 .
8. **내포**를 클릭합니다 . 그러면 C 가 C[A, B] 효과로 변환됩니다 .

매크로

" 매크로 " 목록에서 일반적으로 사용되는 모형에 대한 효과를 자동으로 생성하는 옵션을 선택하고 " 모형 효과 생성 " 목록에 입력합니다 .

완전 요인 " 열 선택 " 목록에서 선택한 열에 대한 모든 주효과 및 교호작용을 생성합니다 . 이러한 효과는 " 열 선택 " 목록에서 주효과가 나열된 순서에 따라 입력됩니다 . 대체 순서에 대한 내용은 이 테이블의 **정렬된 요인 배치**를 참조하십시오 .

특정 차수까지의 요인 모든 주효과 및 지정된 차수까지의 교호작용을 생성합니다 . **매크로** 버튼 아래의 **차수** 상자에 차수를 지정합니다 .

정렬된 요인 배치 **완전 요인** 옵션과 동일한 효과 집합을 생성하지만 차수 순으로 나열됩니다 . 모든 주효과가 먼저 나열되고 그 뒤에 모든 이원 교호작용 , 모든 삼원 교호작용 등이 나열됩니다 .

반응 표면 주효과 , 이원 교호작용 및 2 차 항을 생성합니다 . 선택한 주효과에는 **RS** 로 표시된 반응 표면 속성이 제공됩니다 . RS 속성이 주효과에 적용되고 " 표준 최소 제곱 " 분석법을 선택하면 반응 표면 보고서가 제공됩니다 . 이 보고서에는 반응 표면의 임계값 및 형태에 대한 정보가 포함됩니다 .

" 속성 "(40 페이지) 의 " 반응 표면 효과 " 및 실험 설계 가이드의 도 확인하십시오 .

혼합물 반응 표면 주효과 및 이원 교호작용을 생성합니다 . 주효과에는 반응 표면 (**RS**) 및 혼합물 (**혼합물**) 속성이 표시됩니다 . " 표준 최소 제곱 " 분석법을 사용할 경우 **혼합물** 속성은 혼합물 모형을 적합시킵니다 . **RS** 속성은 혼합물 모형과 관련된 반응 표면 보고서를 생성합니다 .

" 속성 "(40 페이지) 의 " 혼합물 효과 " 및 실험 설계 가이드의 도 확인하십시오 .

특정 차수까지의 다항식 주효과 및 지정된 차수까지의 다항식 항을 생성합니다 . **매크로** 버튼 아래의 **차수** 상자에 차수를 지정합니다 .

Scheffe 입방 혼합물 실험에 대한 반응 표면을 지정하는 데 유용한 주효과 , 교호작용 및 Scheffe 입방 항을 생성합니다 . 이 매크로는 완전 입방 모형을 생성합니다 .

혼합물에 대해 3 차 다항식 모형을 적합시키는 경우 X1*X1*X2 와 같은 짝수 거듭제곱 항은 추정할 수 없으므로 포함하지 않아야 합니다 . 그러나 표면의 완전 다항식 규격에는 Scheffe 입방 항이라고 하는 X1*X2*(X1 – X2) 형식의 항이 포함될 수 있습니다 .

차수 상자에 3 을 입력한 후 **혼합물 반응 표면** 매크로 명령을 선택하면 Scheffe 입방 항도 포함됩니다 .

JMP PRO **그룹화된 회귀변수** 보고서의 여러 부분에서 하나의 효과로 처리되는 연속형 요인 집합에 대한 단일 효과를 생성합니다 . " 효과 요약 " 에서 그룹화된 효과를 추가하거나 제거할 수 있습니다 . 그룹의 각 요인에 대한 개별 그림과 검정이 아니라 각 그룹 효과에 대해 하나의 레버리지 그림과 하나의 효과 검정을 나타냅니다 . 이 옵션은 데이터 테이블에 범주형 효과 수준을 나타내는 표시자 열이 포함되어 있을 때 유용할 수 있습니다 .

참고 : 회귀변수 그룹에 포함된 비연속형 요인은 분석에 포함되지 않습니다 .

속성

"속성" 목록에서는 "모형 효과 생성" 목록에서 선택한 효과에 할당할 수 있는 속성을 선택합니다.

임의 효과 효과에 " 랜덤 " 속성을 할당합니다 . 임의 효과에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 임의 효과 및 적합 방법 지정 "(175 페이지) 에서 확인하십시오 .

반응 표면 효과 효과에 **RS** 속성을 할당합니다 . 관련된 모형 항이 " 모형 효과 생성 " 목록에 포함되어 있어야 합니다 . " 매크로 " 목록의 " 반응 표면 " 옵션은 이러한 항을 자동으로 생성하고 주효과에 RS 속성을 할당합니다 . 반응 표면 보고서를 생성하려는 경우 교호작용 및 다항식 항에는 반응 표면 속성을 할당할 필요가 없습니다 . 주효과에만 이 속성을 할당하면 됩니다 .

로그 분산 효과 효과에 " 로그 분산 " 속성을 할당합니다 . 이 속성은 효과가 반응의 분산 모형에 포함된다는 것을 나타냅니다 .

모형의 효과를 반응의 평균과 분산 **모두**에 포함하려면 효과를 두 번 지정해야 합니다 . 탭 형식 인터페이스에서 " 평균 효과 " 탭과 " 분산 효과 " 탭에 모두 표시되어야 합니다 . 아니면 " 평균 효과 " 탭에서 두 번 입력할 수 있으며 , 한 번은 " 로그 분산 효과 " 속성을 사용하고 또 한 번은 " 로그 분산 효과 " 속성을 사용하지 않습니다 .

혼합물 효과 주효과에 " 혼합물 " 속성을 할당합니다 . 이 속성은 혼합물과 관련된 주효과를 지정하는 데 사용됩니다 . " 매크로 " 목록의 " 혼합물 반응 표면 " 옵션은 선택한 효과에 혼합물 속성을 자동으로 할당하고 가능한 경우 반응 표면 보고서를 제공합니다 .

제외된 효과 효과에 " 제외됨 " 속성을 할당합니다 . 이 속성은 모형 적합에서 해당 효과를 제외합니다 . 그러나 이 효과는 적합 결여 검정을 위해 관측값을 그룹화하는 데 사용됩니다 . " 표준 최소 제곱 " 분석법을 사용할 경우 이 효과에 대한 최소 제곱 평균 테이블이 제공됩니다 .

매듭이 지정된 스플라인 효과 연속형 주효과에 " 매듭 지정 " 속성을 할당합니다 . 이 속성은 효과에 대한 3 차 스플라인을 모형 규격에 암시적으로 추가합니다 . 자세한 내용은 " 매듭이 지정된 스플라인 효과 "(40 페이지) 에서 확인하십시오 .

매듭이 지정된 스플라인 효과

매듭이 지정된 스플라인은 예측 변수의 유연한 함수를 사용하여 반응 Y를 적합시키는 데 사용됩니다. 단일 예측 변수 X를 예로 들겠습니다. "매듭이 지정된 스플라인 효과"가 X에 할당되고 k 개의 매듭이 지정되면 k - 2 개의 추가 효과가 예측 변수 집합에 암시적으로 추가됩니다. 이러한 각 효과는 매듭에 의해 세그먼트가 정의되는 조각별 3 차 다항식 스플라인입니다. 자세한 내용은 Stone and Koo(1985)를 참조하십시오.

스플라인 수는 사용자가 직접 지정한 매듭 수에 의해 결정됩니다. 스플라인과 관련된 계수는 분석법에 사용되는 방법을 기반으로 추정됩니다.

매듭 배치는 문헌의 지침을 따릅니다. 특히 100 개 이하의 점이 있는 경우 첫 번째 매듭과 마지막 매듭은 각각 다섯 번째로 작은 점과 가장 큰 점입니다. 그렇지 않은 경우 첫 번째 매듭과 마지막 매듭은 각각 매듭이 5 개 이하일 때는 0.05 분위수와 0.95 분위수에 배치되고, 매듭이 5 개보다 많을 때는 0.025 분위수와 0.975 분위수에 배치됩니다. 기본 매듭 수는 관측값이 30 개보다 많으면 5 개이고 관측값이 30 개 미만이면 3 개입니다.

참고 : 매듭이 지정된 스플라인은 주효과 연속형 항에 대해서만 구현됩니다 .

변환

"변환" 옵션은 선택한 Y 열 또는 "모형 효과 생성" 텍스트 상자에서 선택한 주효과를 변환합니다.

참고 : " 열 선택 " 목록에서 마우스 오른쪽 버튼을 클릭하고 " 변환 " 을 선택하여 열을 변환할 수도 있습니다 . 변환된 열에 대한 참조가 " 열 선택 " 목록에 나타납니다 . 그런 다음 데이터 테이블 열과 마찬가지로 모형 적합 창의 열을 사용할 수 있습니다 . 자세한 내용은 **JMP 사용**의 에서 확인하십시오 .

없음 적용된 변환 옵션을 제거합니다 .

로그 선택한 변수에 자연 로그 변환을 적용합니다 .

Sqrt 선택한 변수의 값 제곱근을 적용합니다 .

제곱 선택한 변수의 값 제곱을 적용합니다 .

역수 변수 X 에 1/X 변환을 적용합니다 .

지수 선택한 변수에 지수 변환을 적용합니다 .

Arrhenius 변수 T(섭씨 온도) 에 아레니우스 변환을 적용합니다 .

$$X = \frac{11604.5181215503}{T + 273.15}$$

이 식은 아레니우스 관계식의 구성 요소로 , 여기에 활성화 에너지를 곱하게 됩니다 .

아레니우스 역 변수 X 에 아레니우스 역 변환을 적용합니다 .

$$T = \frac{11604.5181215503}{X} - 273.15$$

Logit 선택한 열에 대한 로지스틱 함수의 역함수를 계산합니다 (여기서 p 는 0 ~ 1 사이).

$$\text{로짓}(p) = \log\left(\frac{p}{1-p}\right)$$

로지스틱 선택한 열에 대한 로지스틱 (Squish 및 Logist 라고도 함) 함수를 계산합니다 (결과 범위 0 ~ 1).

$$\text{로지스틱}(x) = \frac{1}{(1 + e^{-x})}$$

로지스틱 백분율 선택한 열에 대한 로짓을 백분율로 계산합니다 (여기서 *pct* 는 0 ~ 100 범위의 백분율).

$$\text{LogitPct}(pct) = \log\left(\frac{\left(\frac{pct}{100}\right)}{1-\left(\frac{pct}{100}\right)}\right)$$

LogisticPct 선택한 열에 대한 로지스틱 (또는 Logist) 을 백분율로 계산합니다 (결과 범위 0 ~ 100).

$$\text{LogisticPct}(x) = \frac{100}{(1+e^{-x})}$$

절편 없음

절편 항이 없는 모형을 적합시키려면 "절편 없음"을 선택합니다. 특정 모델링 구조에는 절편 모형이 필요하지 않습니다. 이러한 경우 "절편 없음" 상자가 기본으로 선택되어 있습니다.

모형 효과 생성 탭

다음 분석법의 경우 탭 형식 인터페이스를 사용하여 모형 효과를 입력할 수 있습니다 .

참고 : 첫 번째 (주) 탭에서 효과에 적용된 속성은 모형에서 효과가 처리되는 방법을 결정합니다 . 모형을 실행한 후 보고서의 빨간색 삼각형 메뉴에서 " 모형 대화상자 " 를 요청하면 해당 효과가 적절한 탭에 나타납니다 .

표준 최소 제곱 다음과 같이 모형 효과를 입력합니다 .

고정 효과 탭 고정 효과로 모델링할 효과를 입력합니다 . 고정 효과는 특정 처리 수준이 관심 대상인 효과입니다 . 전체 처리 수준에 대해 평균 반응을 비교할 수 있습니다 .

임의 효과 탭 임의 효과로 모델링할 효과를 입력합니다 . 임의 효과는 수준이 더 큰 모집단의 랜덤 표본으로 간주되는 효과입니다 . 이 효과로 인한 반응의 변동을 추정할 수 있습니다 .

혼합 모형 다음과 같이 모형 효과를 입력합니다 .

고정 효과 탭 고정 효과로 모델링할 효과를 입력합니다 . 이 테이블에서 " 표준 최소 제곱 " 을 참조하십시오 .

임의 효과 탭 임의 효과로 모델링할 효과를 입력합니다 . 분산 성분 모형 및 랜덤 계수 모형에 사용합니다 .

반복 구조 탭 반복 효과에 대한 공분산 구조를 선택하려면 사용합니다 .

로그 선형 분산 다음과 같이 모형 효과를 입력합니다 .

평균 효과 탭 기대값을 모델링할 효과를 입력합니다 .

분산 효과 탭 분산을 모델링할 효과를 입력합니다 .

효과의 기대값과 분산을 모두 모델링하려면 두 탭에 모두 입력해야 합니다 .

모수 생존 다음과 같이 모형 효과를 입력합니다 .

위치 효과 탭 위치 모수 mu 또는 형상 모수 (Weibull 분포의 경우) 를 모델링하는 데 사용할 효과를 입력합니다 .

척도 효과 탭 척도 모수를 모델링하는 데 사용할 효과를 입력합니다 .

적합 분석법

모형 적합 시작 창에서 분석법을 지정하여 적합 및 분석 방법을 선택합니다. JMP 는 하나 이상의 반응과 입력된 요인을 기반으로 원하는 분석법에 대한 초기 컨텍스트 기반 추측을 수행하지만 "분석법" 메뉴에서 이 선택을 변경할 수 있습니다.

다음과 같은 적합 분석법을 사용할 수 있습니다 .

표준 최소 제곱 반응이 연속형인 모형을 적합시킵니다 . 회귀 , 분산 분석 , 공분산 분석 , 혼합 모형 및 설계된 실험 분석 기법이 포함됩니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장 (69 페이지) 및 " 표준 최소 제곱에 대한 강조 옵션 "(77 페이지) 에서 확인하십시오 .

단계별 표준 최소 제곱 및 순서형 로지스틱 분석 (또는 이항 반응을 포함하는 명목형) 에 대한 변수 선택을 용이하게 합니다 . 연속형 반응의 경우 교차 검증 , p 값 , BIC 및 AICc 기준이 제공됩니다 . 모든 가능 모형을 적합시키고 모형 평균화를 위한 옵션도 제공됩니다 . 로지스틱 적합의 경우 *p* 값 , BIC 및 AICc 기준이 제공됩니다 . 자세한 내용은 " 단계별 회귀 모형 " 장 (229 페이지) 에서 확인하십시오 .

JMP PRO **일반화 회귀** 정규화 (벌점이라고도 함) 회귀 기법을 사용하여 일반화 선형 모형을 적합시킵니다 . 정규화 기법에는 능형 회귀 , Lasso 회귀 , 적응형 Lasso 회귀 , Elastic Net 및 적응형 Elastic Net 이 포함됩니다 . 반응 분포에는 정규 , 이항 , Poisson, 영과잉 Poisson, 음이항 , 영과잉 음이항 및 감마 분포가 포함됩니다 . 자세한 내용은 " 일반화 회귀 모형 " 장 (263 페이지) 및 " 분포 "(270 페이지) 에서 확인하십시오 .

JMP PRO **혼합 모형** 복소 공분산 구조를 가진 연속형 반응에 대해 다양한 선형 모형을 적합시킵니다 . 다음과 같은 상황에서 사용됩니다 .

- 분할구 실험
- 랜덤 계수 모형
- 반복 측정 설계
- 공간 데이터
- 상관관계가 있는 반응 데이터

자세한 내용은 " 혼합 모형 " 장 (327 페이지) 에서 확인하십시오 .

MANOVA 여러 개의 연속형 Y 변수를 포함하는 모형을 적합시킵니다 . 다변량 분산 분석 , 반복 측정 , 판별 분석 및 정준 상관 기법이 포함됩니다 . 자세한 내용은 " 다변량 반응 모형 " 장 (403 페이지) 에서 확인하십시오 .

로그 선형 분산 연속형 Y 변수의 경우 평균과 분산 모두에 대한 모형을 생성합니다 . 두 모형에 대해 서로 다른 효과 집합을 지정할 수 있습니다 . 자세한 내용은 " 로그 선형 분산 모형 " 장 (427 페이지) 에서 확인하십시오 .

명목형 로지스틱 명목형 반응에 대해 로지스틱 회귀 모형을 적합시킵니다 . 자세한 내용은 " 로지스틱 회귀 모형 " 장 (439 페이지) 에서 확인하십시오 .

순서형 로지스틱 순서형 반응에 대해 로지스틱 회귀 모형을 적합시킵니다 . 자세한 내용은 " 로지스틱 회귀 모형 " 장 (439 페이지) 에서 확인하십시오 .

비례 위험 중도절단을 고려하여 설명 변수가 생존 시간에 미치는 영향을 평가하기 위해 준모수 회귀 모형 (Cox 비례 위험 모형) 을 적합시킵니다 .

분석 > 신뢰성 및 생존 > 비례 위험 모형 적합을 선택하여 이 분석법을 시작할 수도 있습니다 . 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

모수 생존 생존 시간에 대해 일반 선형 회귀 모형을 적합시킵니다 . 생존 시간을 하나 이상의 설명 변수에 대한 함수로 표현할 수 있는 경우 이 옵션을 사용합니다 . 다양한 생존 분포 및 중도절단을 고려합니다 .

분석 > 신뢰성 및 생존 > 모수 생존 모형 적합을 선택하여 이 분석법을 시작할 수도 있습니다 . 자세한 내용은 Reliability and Survival Methods 의 에서 확인하십시오 .

일반화 선형 모형 다양한 분포 및 연결 함수를 사용하여 일반화 선형 모형을 적합시킵니다 . 로지스틱 , Poisson 및 지수 회귀 기법이 포함됩니다 . 자세한 내용은 " 일반화 선형 모형 " 장 (467 페이지) 에서 확인하십시오 .

JMP PRO **부분 최소 제곱** 잠재 요인을 사용하여 하나 이상의 *Y* 변수에 대해 모형을 적합시킵니다 . 설명 변수 (*X* 변수) 의 상관관계가 높거나 관측값보다 *X* 변수가 더 많은 경우에 이 방법으로 모형을 적합시킬 수 있습니다 .

분석 > 다변량 방법 > 부분 최소 제곱을 선택하여 부분 최소 제곱 분석을 시작할 수도 있습니다 . 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

반응 변수 선별 여러 반응에 대해 선형 모형 효과의 검정 수행 과정을 자동화합니다 . 검정 결과와 요약 통계량은 데이터 테이블과 그림에 제공됩니다 . FDR(False Discovery Rate) 방법은 유의성이 잘못 선언되지 않도록 보호합니다 . 로버스트 추정 방법은 이상치에 대한 검정 민감도를 줄여 줍니다 .

참고 : 이 분석법은 연속형 반응만 허용합니다 . **분석 > 선별 > 반응 변수 선별**을 선택하여 개별 요인에 대한 반응 변수 선별을 사용할 수도 있습니다 . 이 플랫폼은 범주형 반응을 지원하고 동등성 검정 및 실제적 유의성 검정도 제공합니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

모형 규격 옵션

" 모형 규격 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

중심화 다항식 차수가 1 보다 큰 효과에 포함된 연속형 항을 평균으로 중심화합니다 . 이 옵션은 효과에 포함된 항에 " 혼합물 효과 " 속성이 할당되거나 " 혼합물 " 열 특성이 있는 경우를 제외하고 기본적으로 선택됩니다 . " 코딩 " 열 특성이 있는 항은 지정된 상한값과 하한값 사이의 중간으로 중심화됩니다 .

중심화는 회귀 계수를 더 쉽게 해석하고 낮은 차수 효과와 높은 차수 효과 사이의 공선성을 줄이는 데 유용합니다 .

결측값 정보화 결측값에 대한 코딩 시스템을 제공합니다 . 이 시스템을 사용하면 결측값이 있어도 예측 모형을 추정할 수 있습니다 . 이 옵션은 결측 데이터에 정보가 포함된 경우 유용합니다 . 자세한 내용은 " 결측값 정보화 "(46 페이지) 에서 확인하십시오 .

이 옵션은 표준 최소 제곱 , 단계별 , 일반화 회귀 , MANOVA, 로그 선형 분산 , 명목형 로지스틱 , 순서형 로지스틱 , 비례 위험 , 모수 생존 , 일반화 선형 모형 및 반응 변수 선별 분석법에서 사용할 수 있습니다 .

유의 수준 설정 모형 적합 분석에 신뢰 구간의 유의 수준을 설정합니다 . 기본 유의 수준은 0.05 입니다 .

오차 규격 (임의 효과가 없는 경우 표준 최소 제곱 분석법에만 사용 가능) " 최소 제곱 적합 " 보고서에서 표준 오차 및 검정에 사용되는 오차 분산과 오차 자유도를 지정합니다 . " 오차 규격 " 을 변경해도 " 스튜던트화 잔차 " 그림과 "Box Cox 변환 " 보고서는 영향을 받지 않습니다 . " 오차 규격 " 이 " 순수 오차 " 또는 " 지정됨 " 인 경우 " 분산 분석 " 보고서에 추가 열이 나타납니다 . 자세한 내용은 " 분산 분석 "(86 페이지) 에서 확인하십시오 .

기본 추정값 모형의 표준 제곱근 평균 제곱 오차와 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다 .

순수 오차 " 적합 결여 " 보고서의 순수 오차 평균 제곱과 관련 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다 . 자세한 내용은 " 적합 결여 "(104 페이지) 에서 확인하십시오 .

주의 : 순수 오차 자유도가 1 이면 검정력이 약하고 신뢰 한계가 크다는 경고 메시지가 표시됩니다 .

지정됨 사용자가 지정한 오차 분산 값과 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다 .

데이터 테이블에 저장 모형 적합 시작 창 규격을 데이터 테이블에 연결된 스크립트로 저장합니다 . 스크립트 이름은 "Model" 입니다 . 테이블에 "Model" 이라는 스크립트가 포함된 경우 **분석 > 모형 적합**을 선택하면 이 스크립트가 시작 창을 자동으로 채웁니다 . 필요한 경우 스크립트 이름을 바꿀 수 있습니다 .

JSL 스크립트에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

스크립트 창에 저장 모형 적합 시작 창 규격을 스크립트 창에 복사합니다 . 스크립트 창을 저장한 후 언제든지 스크립트를 실행하여 모형을 다시 생성할 수 있습니다 .

SAS 작업 생성 스크립트 창에서 SAS 로 현재 분석 및 데이터 테이블을 다시 생성할 수 있는 SAS 프로그램을 생성합니다 . 프로그램이 생성되면 SAS 로 코드를 전송하는 몇 가지 옵션을 사용할 수 있습니다 .

1. 코드를 복사하여 SAS 프로그램 편집기에 붙여 넣습니다 . 이 방법은 이전 버전의 SAS(8.2 이전 버전) 를 실행 중인 경우에 유용합니다 .
2. **편집 > SAS 로 전송**을 선택합니다 .
3. 파일을 저장한 후 두 번 클릭하여 SAS 의 로컬 복사본에서 엽니다 . 이 방법은 SAS ODS 옵션을 활용하려는 경우에 유용합니다 (예 : SAS 코드에서 HTML 또는 PDF 출력 생성).

자세한 내용은 JMP 사용의 에서 확인하십시오 .

SAS 로 전송 코드를 SAS 로 전송하고 JMP 에서 결과를 표시합니다 . SAS 서버에 연결되어 있지 않으면 연결 과정을 안내하는 메시지가 표시됩니다 .

자세한 내용은 JMP 사용의 에서 확인하십시오 .

수렴 설정 " 수렴 설정 " 메뉴에는 다음 옵션이 포함되어 있습니다 .

최대 반복 수 모형 적합에 사용되는 최대 반복 수를 지정합니다 . 최대 반복 수의 기본값은 100 입니다 . 모형이 쉽게 수렴되지 않는 경우 반복 수를 늘릴 수 있습니다 . 데이터 집합이 매우 크거나 모형이 복잡한 경우에는 반복 수를 제한할 수 있습니다 .

수렴 한계 모형 적합에 대한 수렴 한계를 지정합니다 . 모형이 쉽게 수렴되지 않는 경우 수렴 한계를 늘릴 수 있습니다 . 수렴 한계의 기본값은 0.00000001 입니다 .

참고 : " 수렴 설정 " 메뉴는 특정 분석법에 대해서만 나타납니다 . " 표준 최소 제곱 " 분석법을 사용할 경우 임의 효과가 존재하고 시작 창에서 " 방법 " 에 "REML" 이 선택된 경우에만 " 수렴 설정 " 메뉴가 나타납니다 .

결측값 정보화

"결측값 정보화" 옵션은 결측값이 있어도 예측 모형을 추정할 수 있는 코딩 시스템을 생성합니다. 연속형 모형 효과와 범주형 모형 효과를 모두 코드화합니다.

연속형 효과

연속형 주효과에 결측값이 있으면 새 설계 행렬 열이 생성됩니다. 이 열은 주효과 열이 결측이면 값이 1이고 결측이 아니면 값이 0인 표시 변수입니다. 또한 연속형 주효과의 결측값은 분석에 포함된 행의 비결측값 평균으로 대체됩니다. 이 평균은 모수 추정값의 해석력을 유지하는 중립 값입니다.

표시 변수와 관련된 모수는 결측값 그룹화를 통해 예측된 반응과 공변량이 평균으로 설정된 경우 예측된 반응 간의 차이를 추정합니다.

고차 효과의 경우 공변량의 결측값이 공변량 평균으로 대체됩니다. 이렇게 하면 "중심화 다항식"이 선택되었다는(기본 설정) 가정하에 결측값이 있는 행의 고차 효과가 0이 됩니다. "중심화 다항식"은 다항식에 포함된 개별 항을 평균으로 중심화하기 때문입니다.

"효과 검정" 보고서에서 결측값이 있는 각 연속형 주효과의 모수 수 = 2입니다. "모수 추정값" 보고서에서 결측값이 있는 연속형 주효과의 모수에는 <열 이름> 또는 평균(결측된 경우)이라는 라벨이 지정되고, 표시자 모수에는 <열 이름> 결측됨이라는 라벨이 지정됩니다. 데이터 테이블에 저장하는 예측 계산식은 이러한 모형 모수에 해당하는 표현식으로 제공됩니다.

범주형 효과

명목형 또는 순서형 주효과에 결측값이 있는 경우 결측값은 해당 효과의 개별 수준으로 코드화됩니다. 따라서 "효과 검정" 보고서에서 결측값이 있는 각 범주형 주효과에 하나의 추가 모수가 있습니다.

"모수 추정값" 보고서에서 명목형 효과의 모수에는 <열 이름>[]이라는 라벨이 지정됩니다. 순서형 효과의 모수에는 <열 이름>[-x]라는 라벨이 지정되며, 여기서 x는 순서 값이 가장 높은 수준을 나타냅니다.

연속형 효과와 마찬가지로 데이터 테이블에 저장하는 예측 계산식은 모형 모수에 해당하는 표현식으로 제공됩니다.

코딩 테이블

"표준 최소 제곱" 분석법을 사용하는 경우 **열 저장 > 코딩 테이블 저장**을 선택하여 결측값 정보화 모형에 사용된 설계 행렬 열을 볼 수 있습니다.

타당도 검사

모형 적합은 모형을 검사하여 계층에 중복 효과 또는 결측 효과와 같은 오류가 있는지 확인합니다. 경고 메시지가 표시되면 **계속**을 클릭하여 적합을 진행하거나 **취소**를 클릭하여 적합 과정을 중지할 수 있습니다.

모형 규격 및 모형 적합의 예

이 섹션에서는 모형 적합 플랫폼의 " 모형 효과 생성 " 패널을 사용하여 지정할 수 있는 다양한 모형 유형의 효과를 입력하기 위한 템플릿을 제공합니다 .

- 모형 효과 X 와 Z 는 연속형 열을 나타냅니다 .
- 모형 효과 A, B, C 는 명목형 또는 순서형 열을 나타냅니다 .

또한 대부분의 모형에 대해 모형 적합 보기가 시각적으로 표시됩니다 .

- " 단순 선형 회귀 "

- "X 다항식 (k 차)"
- "X 및 Z 다항식 (k 차)"
- " 다중 선형 회귀 "
- " 일원 분산 분석 "
- " 이원 분산 분석 "
- " 교호작용이 있는 이원 분산 분석 "
- " 삼원 완전 요인 "
- " 공분산 분석 , 동일한 기울기 "
- " 공분산 분석 , 서로 다른 기울기 "
- "2 요인 내포 임의 효과 모형 "
- "3 요인 완전 내포 임의 효과 모형 "
- " 단순 분할구 또는 반복 측정 모형 "
- "2 요인 반응 표면 모형 "
- " 매듭이 지정된 스플라인 효과 "

단순 선형 회귀

입력할 효과 : X

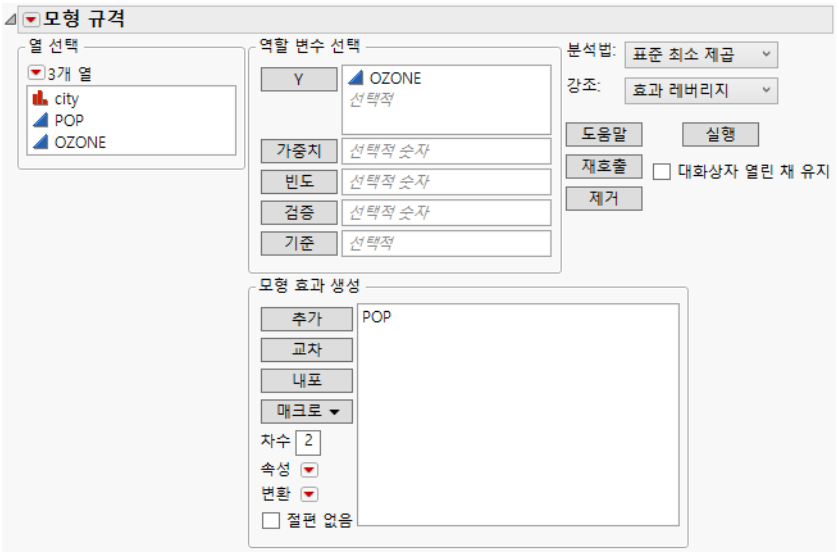
1. " 열 선택 " 목록에서 X 를 선택합니다 .
2. **추가**를 클릭합니다 .

단순 선형 회귀 모형의 예

지정된 도시의 인구 (천 단위) 와 오존 수준 사이의 관계에 관심이 있습니다 . Ozone 은 관심 반응 변수이고 POP 는 연속형 모형 효과입니다 .

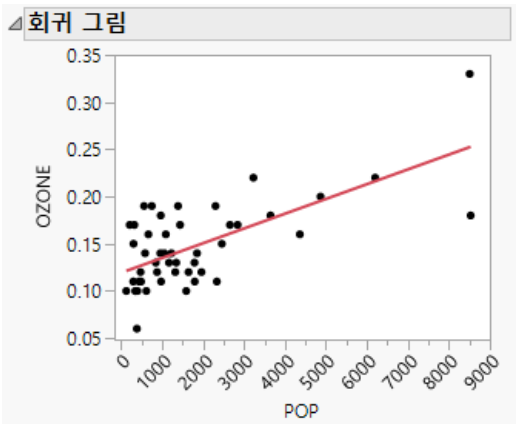
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Polycity.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Ozone 을 선택하고 Y 를 클릭합니다 .
4. " 열 선택 " 목록에서 POP 를 선택합니다 .
5. **추가**를 클릭합니다 .

그림 2.4 단순 선형 회귀에 대한 모형 적합 창



실행을 클릭하면 회귀 그림을 포함한 다양한 결과를 보여 주는 "최소 제곱 적합" 보고서가 나타납니다. "회귀 그림"에는 데이터와 해당 데이터에 대한 단순 선형 회귀 모형 적합이 표시됩니다.

그림 2.5 단순 선형 회귀에 대한 모형 적합



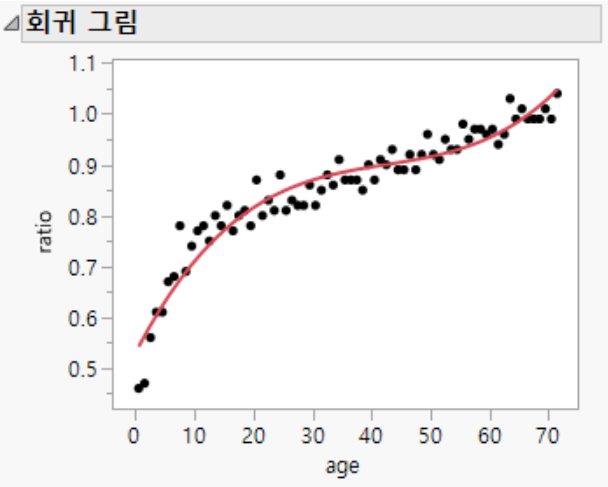
X 다항식 (k 차)

입력할 효과 : X, X*X,..., X^k

1. **차수** 텍스트 상자에 *k* 를 입력합니다 .
2. " 열 선택 " 목록에서 X 를 선택합니다 .
3. **매크로 > 특정 차수까지의 다항식**을 선택합니다 .

그림 2.6에서는 Growth.jmp 샘플 데이터 테이블의 데이터와 해당 데이터에 대한 3차 다항식 모형 적합 그림을 보여 줍니다. 이것은 **Bivariate** 데이터 테이블 스크립트를 실행할 때 생성되는 적합 중 하나입니다.

그림 2.6 한 변수에서 3 차 다항식에 대한 모형 적합



X 및 Z 다항식 (k 차)

입력할 효과 : X, X*X,..., X^k, Z, Z*Z,..., Z^k

1. **차수** 텍스트 상자에 *k* 를 입력합니다 .
2. " 열 선택 " 목록에서 X 와 Z 를 선택합니다 .
3. **매크로 > 특정 차수까지의 다항식**을 선택합니다 .

다중 선형 회귀

입력할 효과 : 선택한 열

1. " 열 선택 " 목록에서 관심 있는 연속형 효과를 선택합니다 .
2. **추가**를 클릭합니다 .

그림 2.7에서는 Grandfather Clocks.jmp 샘플 데이터 테이블의 데이터와 해당 데이터에 대한 다중 선형 회귀 적합이 포함된 표면 프로파일러 그림을 보여 줍니다. 모형 효과는 Age 와 Bidders 입니다. 반응은 Price 입니다. **Fit Model with Surface Profiler Plot** 데이터 테이블 스크립트를 실행하여 그림을 확인할 수 있습니다.

그림 2.7 예측 변수가 두 개인 다중 선형 회귀 모형에 대한 모형 적합

다중 선형 회귀 모형의 예

여러 예측 변수가 있는 예는 "모형 적합을 사용한 회귀 분석의 예"(32페이지)에서 확인하십시오.

일원 분산 분석

입력할 효과 : A

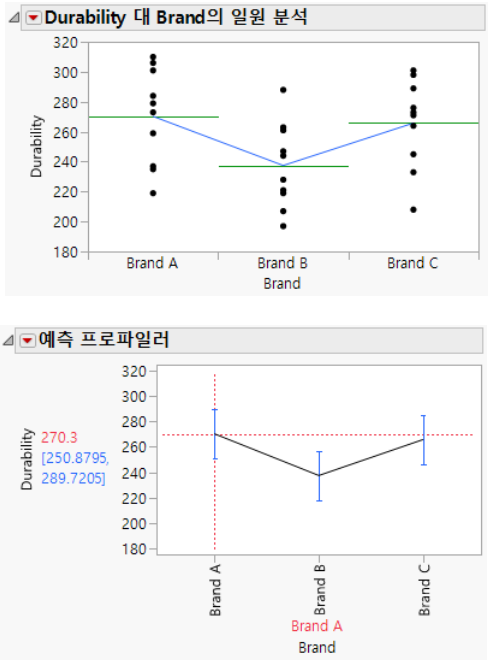
1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 하나 (A) 를 선택합니다 .
2. **추가**를 클릭합니다 .

Golf Balls.jmp 샘플 데이터 테이블을 예로 들어보겠습니다. Brand(브랜드)에 따라 Durability(내구성)가 다른지 여부에 관심이 있습니다. 그림 2.8에서는 두 개의 그림을 보여 줍니다.

첫 번째 그림은 "X로 Y 적합"을 사용하여 생성되며 브랜드별 데이터를 보여 줍니다. 각 브랜드의 평균에 수평선이 표시되고 선분이 평균을 연결합니다. 이 그림을 생성하려면 Golf Balls.jmp 샘플 데이터 테이블에서 **Oneway: Durability by Brand** 스크립트를 실행합니다.

두 번째 그림은 "모형 적합"을 사용하여 생성되는 프로파일러 그림입니다. 이 그림은 각 브랜드에 대한 예측 반응을 선분으로 연결하여 보여 줍니다. 이 그림을 생성하려면 Golf Balls.jmp 샘플 데이터 테이블에서 Fit Model: Durability by Brand 스크립트를 실행합니다. 빨간색 수직 파선을 관심 브랜드로 드래그합니다. 그러면 빨간색 수평 파선이 업데이트되어 예측 반응에서 세로 축과 교차합니다.

그림 2.8 일원 분산 분석에 대한 모형 적합



이원 분산 분석

입력할 효과 : A, B

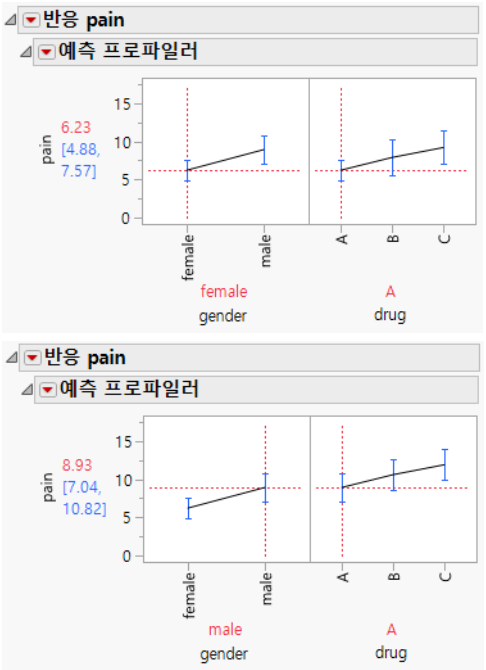
1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다 .
2. **추가**를 클릭합니다 .

그림 2.9에서는 Analgesics.jmp 샘플 데이터 테이블의 데이터 적합에 대한 두 개의 프로파일러 그림을 보여 줍니다. 모형 효과는 gender와 drug입니다. 반응은 pain입니다. 이 그림을 생성하려면 **분석 > 모형 적합**을 선택하고 pain을 Y로, gender와 drug를 모형 효과로 선택한 후 **실행**을 클릭합니다. 그런 다음 "반응 pain"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다.

각 그림의 선분은 빨간색 수직 파선으로 정의된 설정의 예측값을 연결합니다. 다른 설정에 대한 예측을 확인하려면 이 선분을 이동합니다.

그림 2.9의 위쪽 그림에는 여성에 대한 예측이 표시되고 아래쪽 그림에는 남성에 대한 예측이 표시됩니다. 세 가지 약물의 상대적 효과는 전체 gender 수준에서 일관적입니다. 이는 모형에 교호작용 항이 없기 때문입니다. 교호작용이 있는 예는 그림 2.11에서 확인하십시오.

그림 2.9 교호작용이 없는 이원 분산 분석에 대한 모형 적합



교호작용이 있는 이원 분산 분석

입력할 효과 : A, B, A*B

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다 .
2. **매크로 > 완전 요인**을 선택합니다 .

또는

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다 .
2. **추가**를 클릭합니다 .
3. " 열 선택 " 목록에서 A 와 B 를 다시 선택하고 **교차**를 클릭합니다 .

교호작용이 있는 이원 분산 분석의 예

팝콘 산출량 데이터가 있고 팝콘 유형과 배치 크기가 산출량에 영향을 주는지 여부에 관심이 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Popcorn.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 yield 를 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 popcorn 과 batch 를 선택합니다 .

5. **매크로 > 완전 요인**을 선택합니다 .

그림 2.10 교호작용이 있는 이원 분산 분석에 대한 모형 적합 창

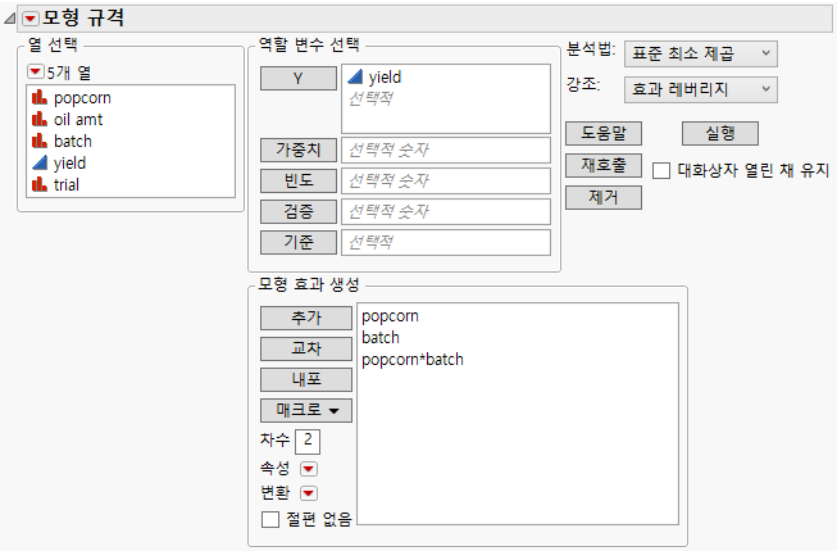
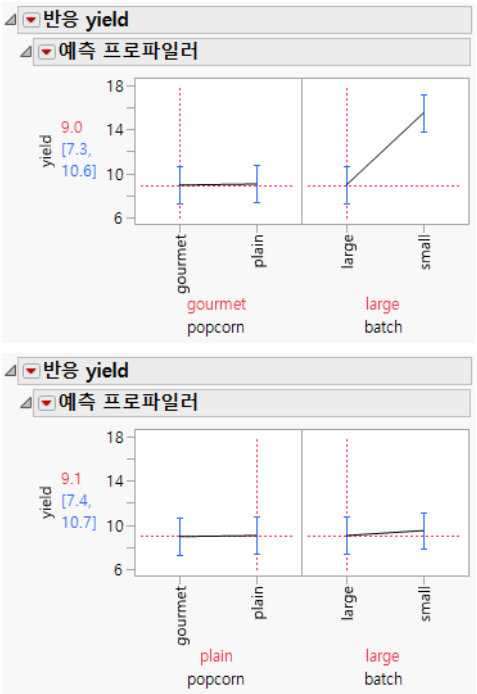


그림 2.11 에서는 이 예에 대한 적합 프로파일러 그림을 보여 줍니다. 이 그림을 생성하려면 그림 2.10 에 표시된 "모형 적합" 창에서 **실행**을 클릭합니다. 그런 다음 "반응 yield"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다.

popcorn 이 위쪽 그림에는 gourmet 로 설정되고 아래쪽 그림에는 plain 으로 설정되어 있습니다. batch 설정에 대한 예측값이 popcorn 유형에 따라 어떻게 달라지는지 확인하십시오. 이것은 popcorn 과 batch 간 교호작용의 결과입니다.

그림 2.11 교호작용이 있는 이원 분산 분석에 대한 모형 적합



삼원 완전 요인

입력할 효과 : A, B, C, A*B, A*C, B*C, A*B*C

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 세 개 (A, B, C) 를 선택합니다 .
2. **매크로 > 완전 요인**을 선택합니다 .

삼원 완전 요인의 예

Speed(속도), Angle(각도), Material(재료) 또는 이러한 요인의 교호작용이 절삭 공구의 Wear(마모) 에 영향을 주는지 여부에 관심이 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Tool Wear.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Wear 를 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 Speed, Angle, Material 을 선택합니다 .
5. **매크로 > 완전 요인**을 선택합니다 .

그림 2.12 삼원 완전 요인에 대한 모형 적합 창

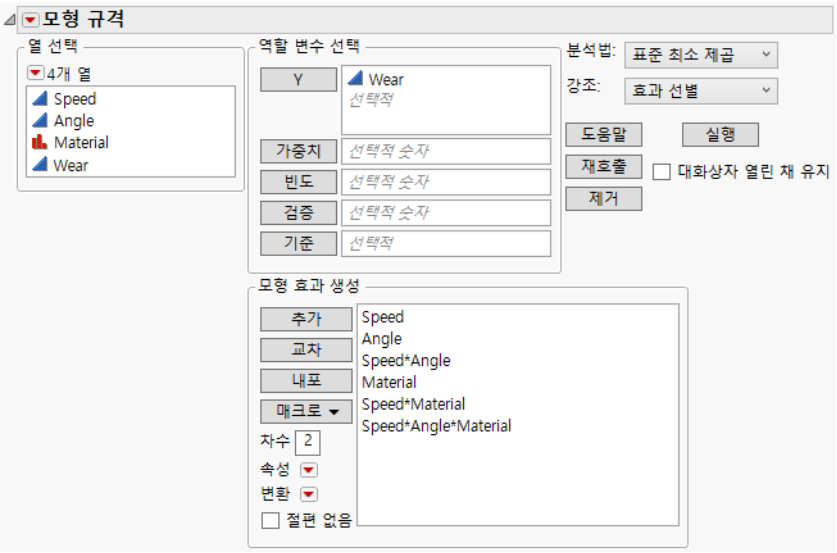
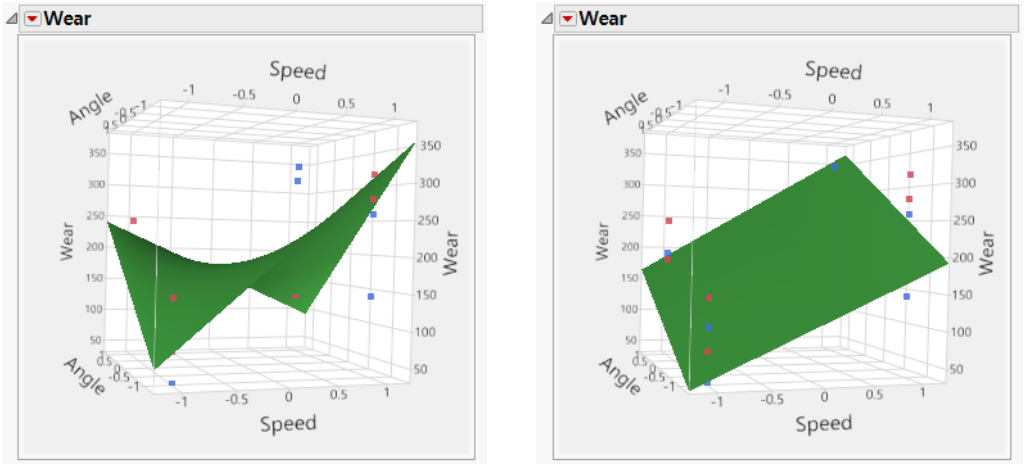


그림 2.13의 표면 프로파일러 그림에서는 두 가지 연속형 효과, 즉 Speed와 Angle 측면에서 Wear에 대한 예측 반응을 보여 줍니다. 왼쪽 그림에는 Material이 A일 때의 예측 반응이 표시되고 오른쪽 그림에는 Material이 B일 때의 예측 반응이 표시됩니다. Material이 A인 점은 빨간색으로 표시되고 Material이 B인 점은 파란색으로 표시됩니다. Material 수준에 따른 반응 표면 형태의 차이는 삼원 교호작용의 결과입니다.

표면 프로파일러 그림을 생성하려면 그림 2.12에 표시된 "모형 적합" 창에서 **실행**을 클릭합니다. 그런 다음 "반응 Wear"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 표면 프로파일러**를 선택합니다. 그림에 점을 추가하려면 "표면 모양" 패널을 열고 **실제값**을 클릭합니다. 점을 더 크게 표시하려면 그림에서 마우스 오른쪽 버튼을 클릭하고 **설정**을 선택한 후 **표식 크기**를 조정합니다.

Material A와 B 둘 다에 대한 그림을 표시하려면 "독립 변수" 패널에서 Material로 표시된 슬라이더를 사용하여 Material A에는 0을 설정하고 Material B에는 1을 설정합니다. 테이블에는 표면 프로파일러 그림을 생성하는 두 개의 데이터 테이블 스크립트(**Prediction and Surface Profilers**, **Surface Profilers for Two Materials**)가 포함되어 있습니다.

그림 2.13 삼원 완전 요인 설계에 대한 모형 적합 - Material A(왼쪽), Material B(오른쪽)



공분산 분석 , 동일한 기울기

X를 공변량으로 사용하여 A의 효과를 검정하는 데 관심이 있습니다. 반응에 대한 X의 효과가 A의 수준에 따라 달라지지 않는다고 믿을 만한 이유가 있다고 가정해 보겠습니다.

입력할 효과 : A, X

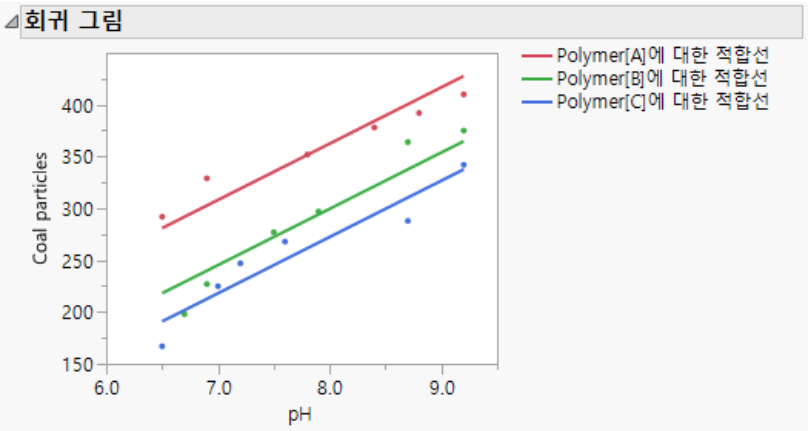
1. "열 선택" 목록에서 명목형 또는 순서형 효과 하나(A)와 연속형 효과 하나(X)를 선택합니다.
2. **추가**를 클릭합니다 .

그림 2.14에서는 Cleansing.jmp 샘플 데이터 테이블의 데이터를 보여 줍니다. 세척 탱크에서 어떤 Polymer(폴리머)가 Coal particles(석탄 입자)를 가장 많이 제거하는지에 관심이 있습니다. 그러나 탱크의 pH도 제거에 영향을 주는 것으로 의심됩니다. 이 그림은 공분산 분석 적합을 보여 줍니다. 여기서 pH와 Coal particles를 연결하는 기울기는 전체 Polymer 수준에서 동일하다고 가정합니다.

다음 단계에 따라 그림 2.14의 그림을 생성할 수 있습니다. "모형 적합" 창에서 Coal particles를 **Y**로 입력하고 pH와 Polymer 둘 다 **모형 효과 생성** 상자에 입력합니다. **실행**을 클릭합니다. "최소 제곱 적합" 보고서에 "회귀 그림"이 나타납니다. 점에 색상을 지정하려면 **행 > 열 값에 따른 색상 또는 표식**을 선택하고, **열 값에 따른 표식** 목록에서 Polymer를 선택한 후 **확인**을 클릭합니다.

그러나 더 완전한 분석은 pH와 Polymer가 Coal particles에 미치는 영향에서 상호 작용한다는 것을 나타냅니다. 해당 모형 적합은 "공분산 분석, 서로 다른 기울기"(58 페이지)에 나와 있습니다.

그림 2.14 공분산 분석 , 동일한 기울기에 대한 모형 적합



공분산 분석 , 서로 다른 기울기

여기서도 X를 공변량으로 사용하여 A의 효과를 검정하는 데 관심이 있습니다. 그러나 반응에 대한 X의 효과가 A의 수준에 따라 달라질 가능성을 허용하도록 모형을 생성합니다.

입력할 효과 : A, X, A*X

1. "열 선택" 목록에서 명목형 또는 순서형 효과 하나(A)와 연속형 효과 하나(X)를 선택합니다.
2. **매크로 > 완전 요인**을 선택합니다 .

또는

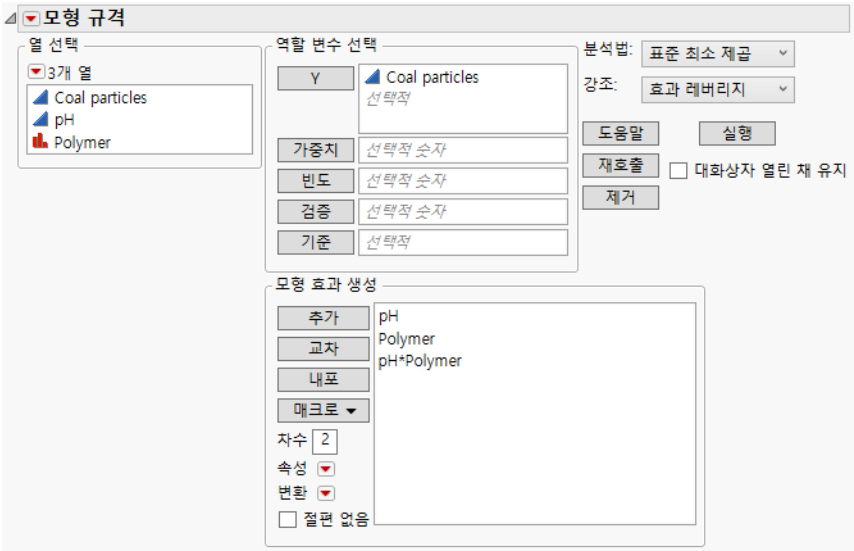
1. "열 선택" 목록에서 명목형 또는 순서형 효과 하나(A)와 연속형 효과 하나(X)를 선택합니다.
2. **추가**를 클릭합니다 .
3. " 열 선택 " 목록에서 A 와 X 를 다시 선택하고 **교차**를 클릭합니다 .

공분산 분석 , 서로 다른 기울기의 예

세 가지 폴리머 (Polymer) 중 석탄 입자 제거 (Coal particles) 에 영향을 미치는 종류가 있는지 여부에 관심이 있습니다 . 탱크 pH 는 폴리머의 탱크 세척력에 영향을 줄 수 있으므로 공변량으로 포함됩니다 . 세 가지 Polymer 유형에 대해 pH 와 Coal particles 간의 관계를 모델링할 때 기울기가 다를 수 있다는 점을 고려합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cleansing.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Coal particles 를 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 pH 와 Polymer 를 선택합니다 .
5. **매크로 > 완전 요인**을 선택합니다 .

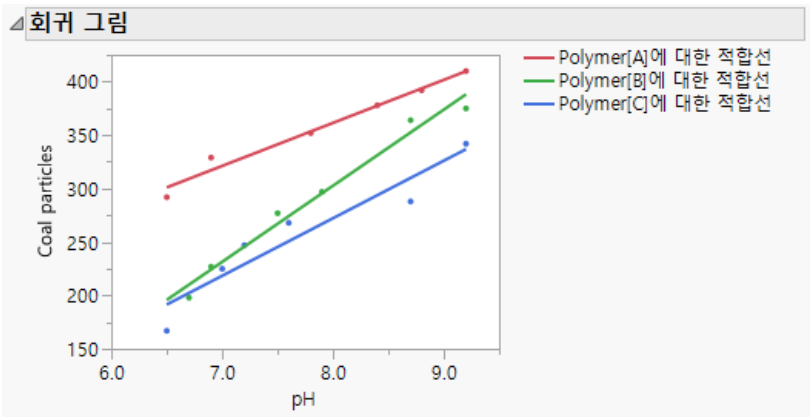
그림 2.15 공분산 분석 , 서로 다른 기울기에 대한 모형 적합 창



실행을 클릭하면 "최소 제곱 적합" 보고서가 나타납니다. "효과 검정" 보고서는 pH와 Polymer 간의 교호작용이 유의하므로 모형에 포함되어야 함을 나타냅니다.

보고서에 제공된 회귀 그림이 그림 2.16에 나와 있습니다. 이 그림은 점과 모형 적합을 보여 줍니다. 교호작용으로 인해 pH와 Coal particles를 연결하는 선의 기울기가 Polymer에 따라 달라질 수 있습니다. 이 교호작용에 상관없이 관심 범위에 대해 Polymer A가 일관되게 가장 높은 제거율을 보인다는 점에 유의하십시오. 그림 2.16에 표시된 것처럼 점에 색상을 지정하려면 **행 > 열 값에 따른 색상 또는 표식**을 선택하고, **열 값에 따른 표식** 목록에서 Polymer를 선택한 후 **확인**을 클릭합니다.

그림 2.16 공분산 분석 , 서로 다른 기울기에 대한 모형 적합



2 요인 내포 임의 효과 모형

두 가지 요인 A와 B가 있지만 B가 A에 내포된 모형을 예로 들어보겠습니다. 내포 효과가 고정 효과로 처리되는 경우도 있지만 대부분의 상황에서 내포 효과는 임의 효과로 처리됩니다. 따라서 아래 설명된 모형에서는 내포 효과가 임의 효과로 입력됩니다.

입력할 효과 : A, B[A]& 랜덤

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다 .
2. **추가**를 클릭합니다 .
3. A 안에 B 내포 : " 모형 효과 생성 " 목록에서 B 를 선택하고 " 열 선택 " 목록에서 A 를 선택합니다 . 두 효과가 강조 표시되어야 합니다 .
4. **내포**를 클릭합니다 .
5. " 모형 효과 생성 " 목록에서 B[A] 가 강조 표시된 상태로 **속성 > 임의 효과**를 선택합니다 .

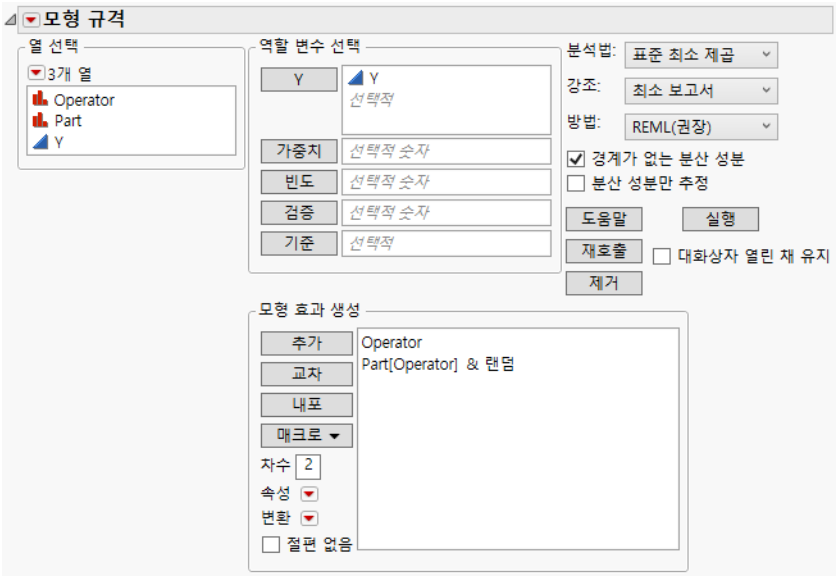
2 요인 내포 임의 효과 모형의 예

측정 시스템 분석 연구의 일부로 무작위 선택된 24개의 부품을 측정합니다. 이 부품은 일반적으로 이러한 부품을 측정하는 6명의 작업자에게 균등하게 분배됩니다. 각 작업자는 할당된 4개 부품 각각에 대해 세 가지 측정을 독립적으로 수행합니다.

한 명의 작업자가 측정한 부품은 해당 작업자에 의해서만 측정되므로 Part 는 Operator 에 내포됩니다 . 부품은 랜덤 생산 표본이므로 Part 가 임의 효과로 간주됩니다 . 이러한 특정 작업자 6 명에게 관심이 있으므로 Operator 는 고정 효과로 처리됩니다 . 적절한 모형을 지정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/2 Factors Nested.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Y 를 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 Operator 와 Part 를 선택합니다 .
5. **추가**를 클릭합니다 .
6. Operator 안에 Part 내포 : " 모형 효과 생성 " 목록에서 Part 를 선택합니다 . " 열 선택 " 목록에서 Operator 를 선택합니다 . 두 효과가 강조 표시되어야 합니다 .
7. **내포**를 클릭합니다 .
8. " 모형 효과 생성 " 목록에서 Part[Operator] 가 강조 표시된 상태로 **속성 > 임의 효과**를 선택합니다 .

그림 2.17 2 요인 내포 임의 효과 모형에 대한 모형 적합 창



9. **실행**을 클릭하여 " 최소 제곱 적합 " 보고서를 생성합니다 .

그림 2.18에서는 두 개의 그림을 보여 줍니다. 첫 번째 그림은 4개 부품 각각에 대한 각 Operator의 세 가지 측정값을 보여 주는 계량형 차트입니다. 수평 선분은 각 Operator의 평균 측정값을 표시합니다.

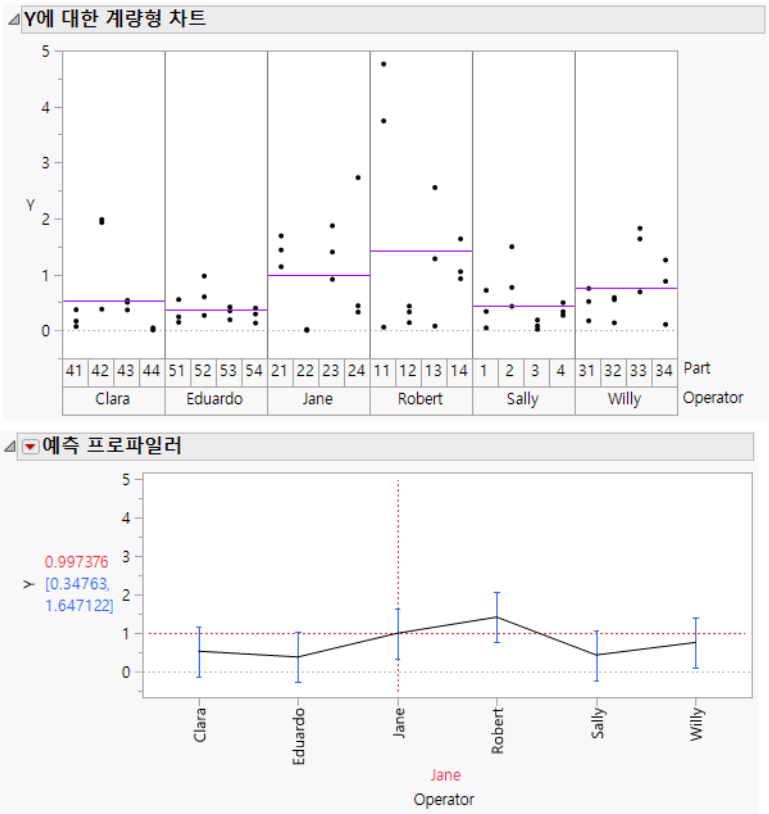
그림 2.18의 계량형 차트를 생성하려면 2 Factors Nested.jmp 샘플 데이터 테이블에서 **Variability Chart - Nested** 데이터 테이블 스크립트를 실행합니다. "계량형 게이지"의 빨간색 삼각형을 클릭한 후 **범위 막대 표시**를 선택 취소하고 **그룹 평균 표시**를 선택합니다.

두 번째 그림은 "최소 제곱 적합" 보고서에서 Operator에 대한 "예측 프로파일러" 그림입니다. 이 그림은 각 작업자에 대한 예측 반응을 보여 줍니다. Jane에 설정된 빨간색 수직 파선은 Jane의 예측 반응이 0.997임을 나타냅니다. "예측 프로파일러" 그림에 제공된 모형 예측과 "계량형 차트"의 원시 데이터 간 대응 관계를 확인할 수 있습니다.

예측 프로파일러 그림을 생성하려면 "반응 Y"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다.

이러한 그림은 각 Operator에 대한 예측 측정값을 모델링하는 방법을 보여 줍니다. 그러나 작업자의 부품 측정 방식이 다른지 여부에만 관심이 있는 것은 아닙니다. 부품 측정 자체의 변동성에도 관심이 있으므로 Part와 관련된 분산 성분을 추정해야 합니다.

그림 2.18 2 요인 내포 임의 효과 모형에 대한 모형 적합



3 요인 완전 내포 임의 효과 모형

세 가지 요인 A, B, C가 있지만 B는 A에 내포되고 C는 A와 B 둘 다에 내포된 모형을 예로 들어 보겠습니다. 또한 B와 C를 임의 효과로 간주합니다.

입력할 효과 : A, B[A]& 랜덤 , C[A,B]& 랜덤

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 세 개 (A, B, C) 를 선택합니다 .
2. **추가**를 클릭합니다 .
3. A 안에 B 내포 : " 모형 효과 생성 " 목록에서 B 를 선택하고 " 열 선택 " 목록에서 A 를 선택합니다 . 두 효과가 강조 표시되어야 합니다 .
4. **내포**를 클릭합니다 .
5. A 와 B 안에 C 내포 : " 모형 효과 생성 " 목록에서 C 를 선택하고 " 열 선택 " 목록에서 A 와 B 를 선택합니다 . 세 효과가 강조 표시되어야 합니다 .
6. **내포**를 클릭합니다 .

7. " 모형 효과 생성 " 목록에서 B[A] 와 C[A,B] 가 강조 표시된 상태로 **속성 > 임의 효과**를 선택합니다 .

단순 분할구 또는 반복 측정 모형

여기서 A 는 주구 변수, B[A] 는 주구 ID, C 는 분할구 또는 반복 측정 변수입니다.

입력할 효과 : A, B[A]& 랜덤 , C, C*A

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다 .
2. **추가**를 클릭합니다 .
3. A 안에 B 내포 : " 모형 효과 생성 " 목록에서 B 를 선택하고 " 열 선택 " 목록에서 A 를 선택합니다 . 두 효과가 강조 표시되어야 합니다 .
4. **내포**를 클릭합니다 .
5. " 모형 효과 생성 " 목록에서 B[A] 를 선택합니다 .
6. **속성 > 임의 효과**를 선택합니다 .
7. " 열 선택 " 목록에서 세 번째 명목형 또는 순서형 효과 (C) 를 선택합니다 .
8. **추가**를 클릭합니다 .
9. " 모형 효과 생성 " 목록에서 C 를 선택하고 " 열 선택 " 목록에서 A 를 클릭합니다 . 두 효과 모두 강조 표시되어야 합니다 .
10. **교차**를 클릭합니다 .

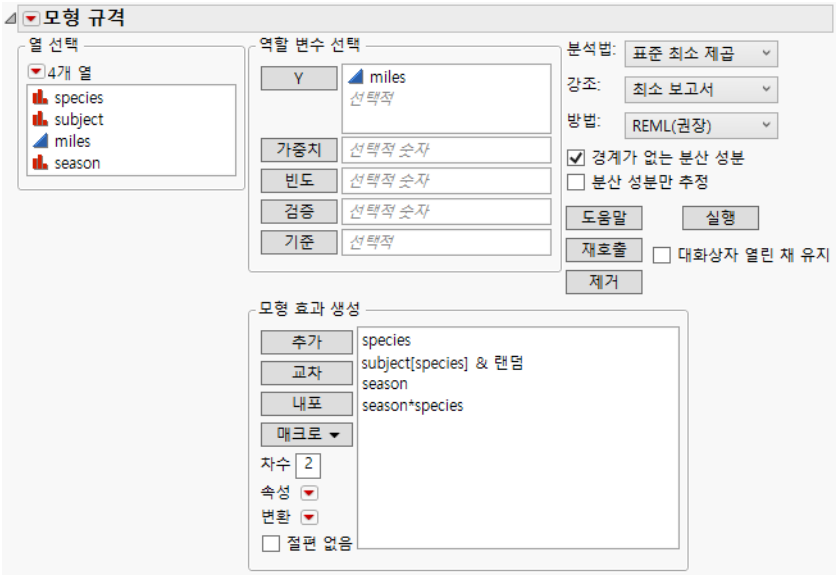
단순 반복 측정 모형의 예

사계절 중 각 계절마다 6 마리 동물 각각의 이동 거리에 대한 데이터가 있습니다 . 두 가지 종 (species) 이 있으며 동물 식별자는 species 에 내포되어 있습니다 . 6 마리의 동물은 더 큰 종 개체군을 대표하므로 subject 를 임의 효과로 처리하기로 합니다 . 각 동물에 대해 반복 측정이 있다는 사실을 고려하여 반응 (miles) 을 species 와 season 의 함수로 모델링하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Animals.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 miles 를 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 species 와 subject 를 선택합니다 .
5. **추가**를 클릭합니다 .
6. species 안에 subject 내포 : " 모형 효과 생성 " 목록에서 subject 를 선택합니다 . " 열 선택 " 목록에서 species 를 선택합니다 . 두 효과가 강조 표시되어야 합니다 .
7. **내포**를 클릭합니다 .
8. " 모형 효과 생성 " 목록에서 subject[species] 를 선택합니다 .
9. **속성 > 임의 효과**를 선택합니다 .

10. " 열 선택 " 목록에서 season 을 선택합니다 .
11. **추가**를 클릭합니다 .
12. " 모형 효과 생성 " 목록에서 season 을 선택합니다 . " 열 선택 " 목록에서 species 를 클릭합니다 . 두 효과 모두 강조 표시되어야 합니다 .
13. **교차**를 클릭합니다 .

그림 2.19 단순 반복 측정 모형에 대한 모형 적합 창



2 요인 반응 표면 모형

입력할 효과 : X&RS, Z&RS, X*X, X*Z, Z*Z

1. " 열 선택 " 목록에서 연속형 효과 두 개 (X, Z) 를 선택합니다 .
2. **매크로 > 반응 표면**을 선택합니다 .

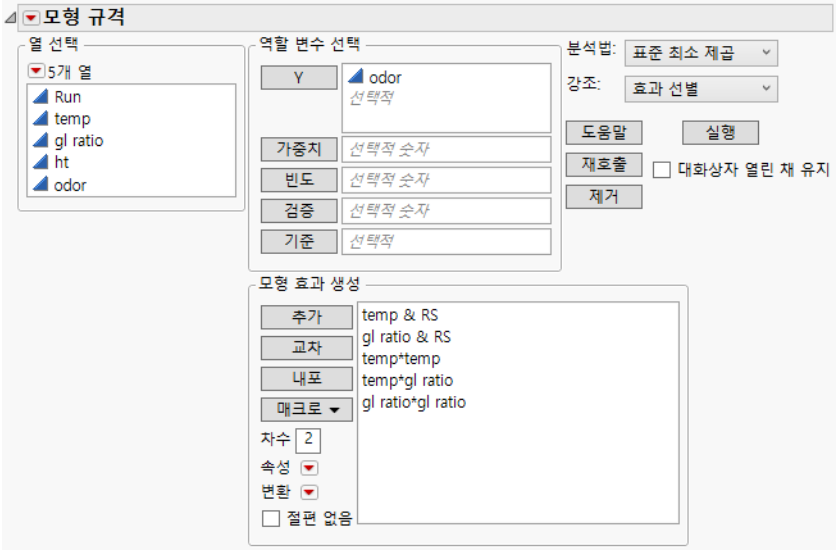
2 요인 반응 표면 모형의 예

Odor Control Original.jmp 샘플 데이터 테이블에서 반응 표면을 적합시켜 response, odor 을 temp 및 gl ratio 의 함수로 모델링하려고 합니다 . Response Surface 데이터 테이블 스크립트와 같이 ht 를 포함할 수도 있지만 이 예에서는 포함하지 않습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Odor Control Original.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 odor 을 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 temp 와 gl ratio 를 선택합니다 .

5. **매크로 > 반응 표면**을 선택합니다 .

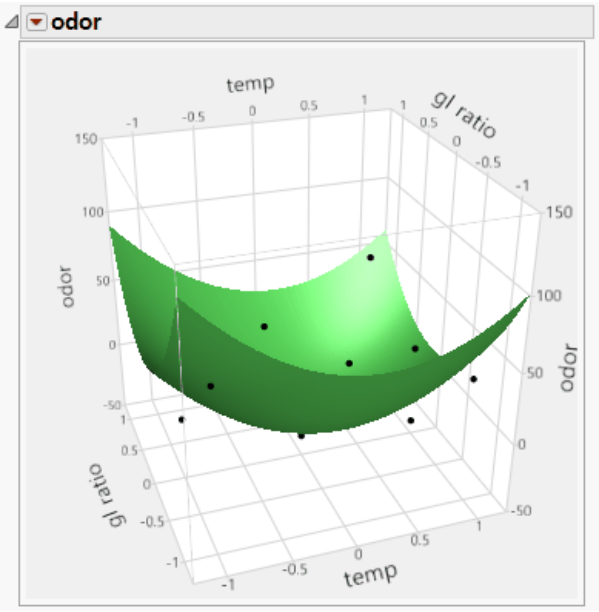
그림 2.20 2 요인 반응 표면 모형에 대한 모형 적합 창



6. **실행**을 클릭합니다 .

그림 2.21에서는 Odor Control Original.jmp 샘플 데이터 테이블의 데이터와 해당 데이터에 대한 2 차 반응 표면 적합이 포함된 표면 프로파일러 그림을 보여 줍니다. 이 그림을 생성하려면 "반응 odor"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 표면 프로파일러**를 선택합니다. 점을 표시하려면 표시 아이콘을 클릭하여 **표면 모양** 패널을 열고 **실제값**을 클릭합니다. 점을 더 크게 표시하려면 그림에서 마우스 오른쪽 버튼을 클릭하고 **설정**을 선택한 후 **표식 크기**를 조정합니다.

그림 2.21 두 변수에서 2 차 다항식에 대한 모형 적합



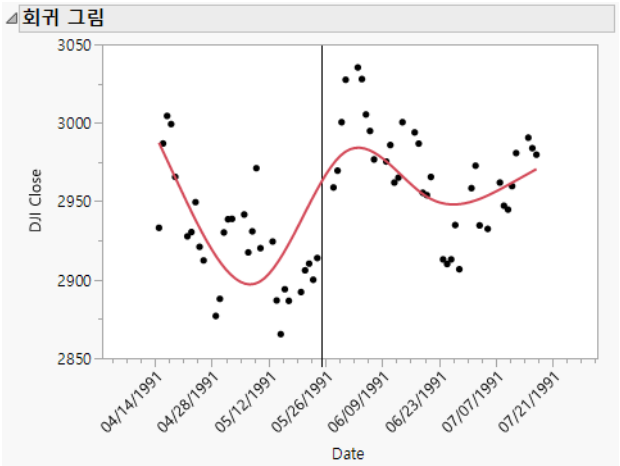
매듭이 지정된 스플라인 효과

입력할 효과 : X& 매듭 지정

1. " 열 선택 " 목록에서 연속형 효과 X 를 선택합니다 .
2. **추가**를 클릭합니다 .
3. " 모형 효과 생성 " 목록에서 X 를 선택합니다 .
4. **속성 > 매듭이 지정된 스플라인 효과**를 선택합니다 .
5. 메뉴가 나타나면 매듭 수를 지정하거나 기본값을 적용합니다 .
6. **확인**을 클릭합니다 .

그림 2.22 에서는 XYZ Stock Averages (plots).jmp 샘플 데이터 테이블의 데이터에 대한 모형 적합이 포함된 회귀 그림을 보여 줍니다 . 여기서 Date 에는 매듭이 지정된 스플라인 효과가 할당되고 5 개의 매듭이 지정되었습니다 . DJI Close 가 반응입니다 .

그림 2.22 5 개의 매듭이 있는 매듭 지정 스플라인에 대한 모형 적합



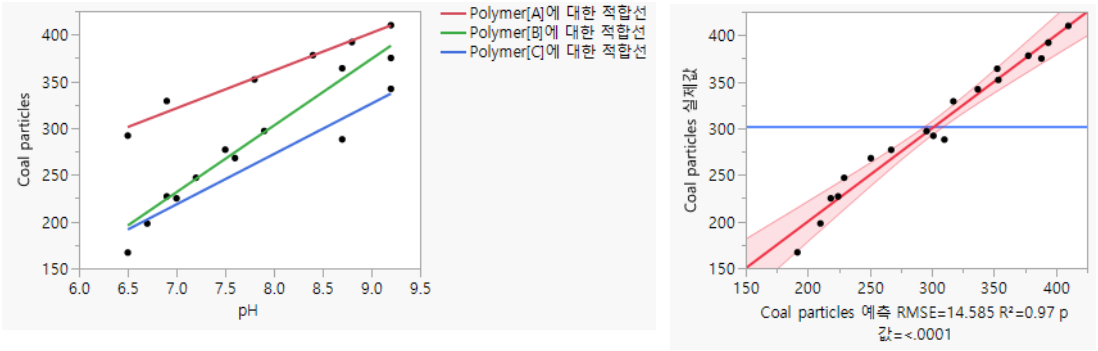
3 장

표준 최소 제곱 모형
모형의 공통 클래스 분석

모형 적합 플랫폼 내의 표준 최소 제곱 분석법은 광범위한 표준 모형을 적합시킵니다. 이러한 모형에는 설계된 실험을 분석하는 데 일반적으로 사용되는 모형뿐만 아니라 회귀, 분산 분석, 공분산 분석 및 혼합 모형이 포함됩니다. 표준 최소 제곱 분석법을 사용하면 최소 제곱을 사용하거나, 임의 효과의 경우 REML(제한 최대 가능도)을 사용하여 연속형 반응 데이터에 대한 선형 모형을 생성할 수 있습니다.

분석 결과는 프로파일러, 등고선 그림 및 표면 그림과 같은 강력한 동적 시각화 도구로 지원됩니다 (Profilers 참조). 이러한 시각적 표시는 모형에 대한 관심을 가지게 하거나 모형에 대한 이해를 보완하고 지원합니다. 이를 통해 여러 반응을 동시에 최적화하고 잡음 효과를 탐색할 수 있습니다.

그림 3.1 표준 최소 제곱 그림의 예



목차

예 : 표준 최소 제곱 사용 72
표준 최소 제곱 분석법 시작 74
모형 적합 시작 창 75
모형 적합 시작 창의 표준 최소 제곱 옵션 76
표준 최소 제곱의 검증 78
결측값 78
최소 제곱 적합 보고서 79
단일 반응과 다중 반응의 비교 79
강조와 관련된 보고서 구조 79
특수 보고서 79
최소 제곱 적합 옵션 83
적합 그룹 옵션 83
반응 옵션 84
회귀 보고서 85
적합 요약 85
분산 분석 86
모수 추정값 87
효과 검정 88
효과 상세 정보 89
적합 결여 104
추정값 105
예측 표현식 표시 107
정렬된 추정값 108
확장 추정값 111
표시자 파라미터화 추정값 113
순차 검정 114
사용자 검정 115
다중 비교 117
기울기 비교 130
결합 요인 검정 130
역추정 예측 131
Cox 혼합물 135
모수 검정력 136
추정값 상관계수 138
명목형 효과에 대한 코딩 139
효과 선별 140
척도화된 추정값 및 연속형 항 코딩 141
그림 옵션 142

정규 확률도 보고서 146
Bayes 그림 보고서 147
파레토도 보고서 149
요인 프로파일링 150
프로파일러 151
교호작용 그림 151
등고선 프로파일러 153
혼합물 프로파일러 154
입방체도 155
Y의 Box-Cox 변환 156
표면 프로파일러 158
행 진단 159
레버리지 그림 161
Press 164
열 저장 165
예측 계산식 167
효과 요약 보고서 168
혼합 및 임의 효과 모형 보고서 및 옵션 172
혼합 모형 및 임의 효과 모형 172
REML(제한 최대 가능도) 방법 176
EMS(전통적) 모형 적합 보고서 181
모형 항 사이에 선형 종속성이 있는 모형 182
특이성 상세 정보 183
모수 추정값 보고서 184
효과 검정 보고서 185
예 185
표준 최소 제곱 분석법에 대한 통계 상세 정보 187
강조 규칙 187
사용자 검정 예에 대한 상세 정보 188
추정값 상관계수 189
레버리지 그림 상세 정보 189
Kackar-Harville 수정 192
검정력 분석 193

예 : 표준 최소 제곱 사용

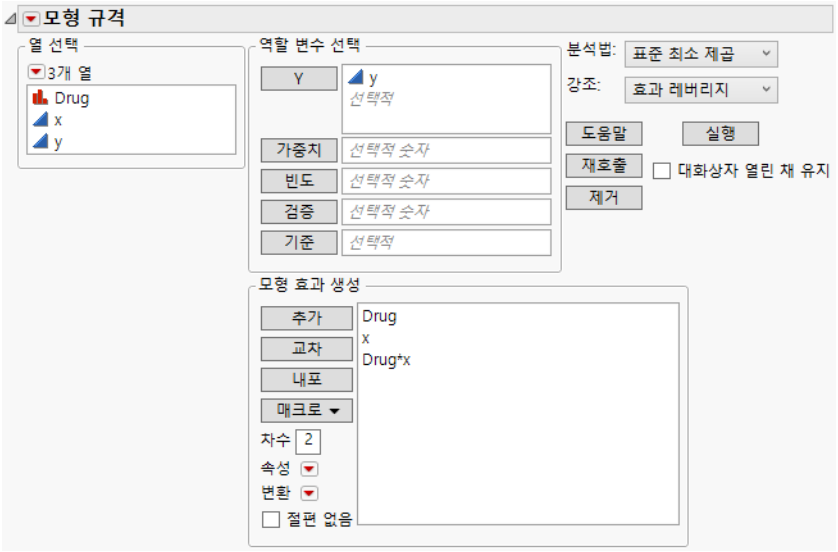
질병 치료에 약물이 미치는 영향에 대한 연구에서 30명의 환자를 무작위로 10명씩 세 그룹으로 나누었습니다. 이 중 두 그룹에는 약물(Drug a, Drug d)이 투여되고 세 번째 그룹에는 플라시보(Drug f)가 처리되었습니다. 각 환자에 대해 전처리 측정(x) 및 후처리 측정(y)이 수행되었습니다. 전처리 점수 x는 환자 간의 질병 단계 차이를 설명하기 위해 공변량으로 포함됩니다. 이 예는 Snedecor and Cochran 연구 자료(1967, p. 422)에서 가져왔습니다.

세 가지 약물(Drug) 그룹에 차이가 있는지 확인하려고 합니다. y 반응, Drug, x 모형 효과 및 Drug 와 x 의 교호작용으로 모형을 생성합니다. 교호작용은 질병 단계에 따라 약물 효과에 차이가 있는 상황을 설명할 수 있습니다. 모형 적합 창과 다양한 분석법에 대한 배경 정보는 " 모형 규격 " 장 (29 페이지) 에서 확인하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .

   이 열을 **Y** 로 추가하면 적합 **분석법**이 " 표준 최소 제곱 " 으로 설정됩니다 . " 효과 레버리지 " 가 선택된 **강조** 옵션이 추가되며 원하는 경우 변경할 수 있습니다 .
4. Drug 와 x 를 선택합니다 . " 열 선택 " 목록에 이 두 효과가 강조 표시된 상태에서 **매크로**를 클릭하고 **완전 요인**을 선택합니다 . 매크로는 이 두 효과 및 효과의 이원 교호작용을 " 모형 효과 생성 " 목록에 추가합니다 .

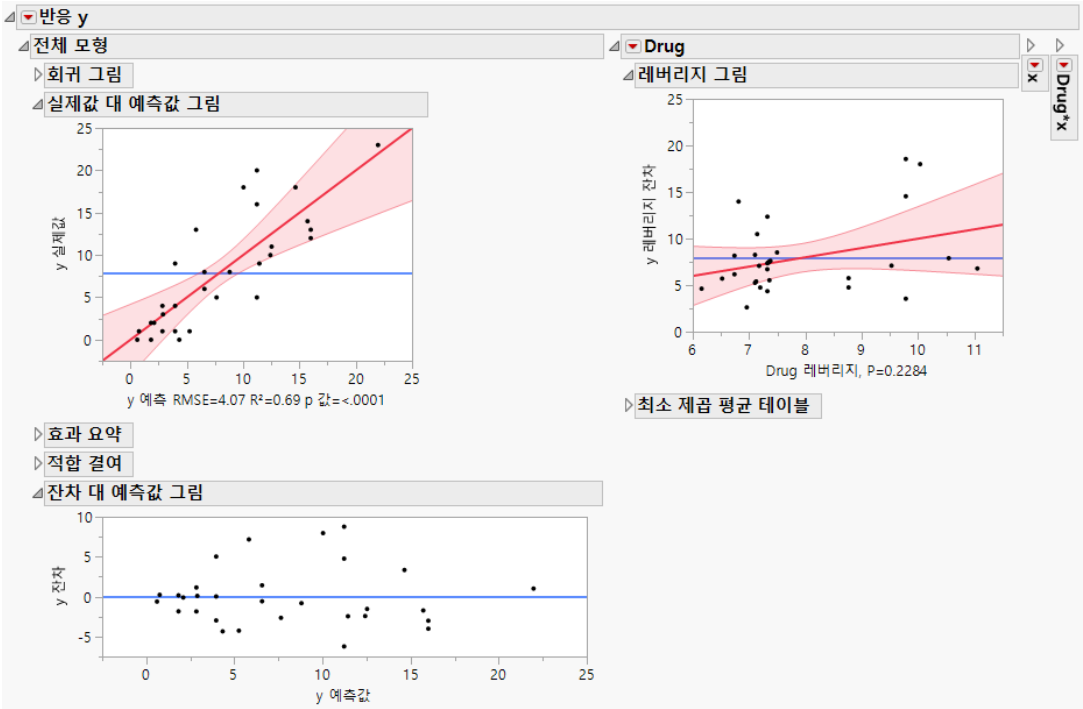
그림 3.2 완료된 모형 적합 시작 창



5. **실행**을 클릭합니다 .

그림 3.3에 "최소 제곱 적합" 보고서가 표시되어 있습니다. 포함된 보고서 중 일부는 공간을 고려하여 닫혀 있습니다. "실제값 대 예측값 그림", "잔차 대 예측값 그림" 및 "레버리지 그림"은 모형 적합 및 기본 가정에 대해 불일치를 나타내지 않습니다.

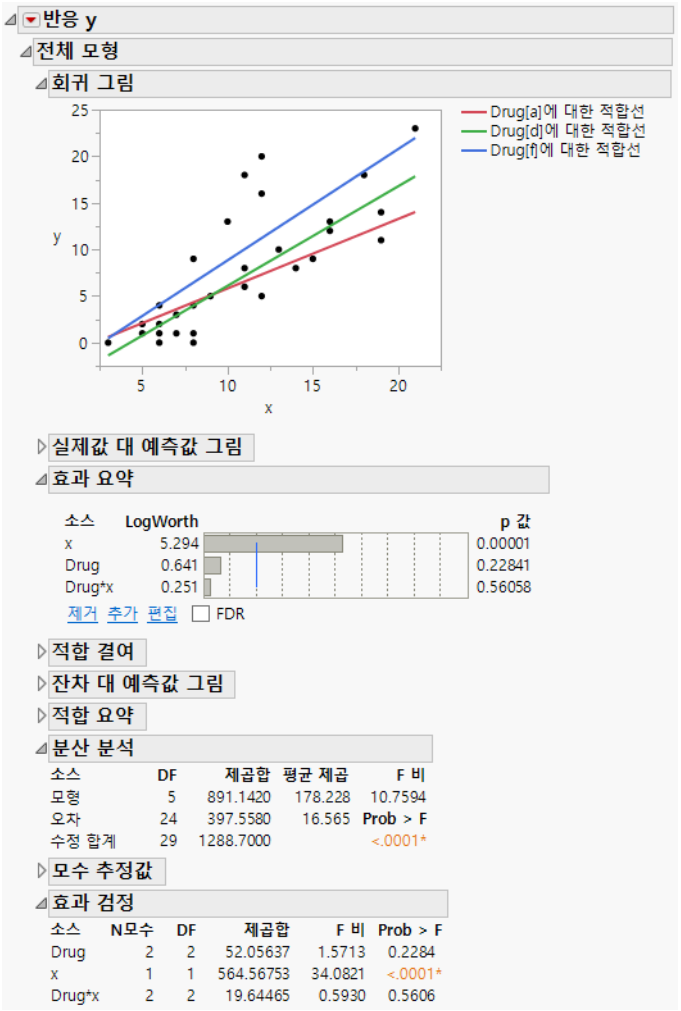
그림 3.3 모형 적합 평가를 위한 그림을 보여 주는 최소 제곱 적합 보고서



모형 적합에 명백한 문제가 없으므로 통계 검정을 해석할 수 있습니다. 그림 3.4에서는 관련 보고서를 보여 줍니다. 전체 모형은 "분산 분석" 보고서에 표시된 대로 유의합니다.

"회귀 그림"에 따르면 Drug와 전처리 측정 x가 상호 작용하지만 "효과 검정" 보고서의 "Prob > F" 값은 이러한 결론을 지지하지 않습니다. 또한 "효과 검정" 보고서에서는 x가 y를 설명하는 데 유의하지만 Drug는 유의하지 않음을 보여 줍니다. 이 연구를 수행하여 세 그룹 간의 차이를 감지하지 못했습니다. 그러나 Drug에 효과가 없다는 결론을 내릴 수는 없습니다. 약물의 효과는 서로 다를 수 있지만 연구 규모가 그 차이를 감지할 만큼 크지 않았습니다.

그림 3.4 유의성 평가를 위한 보고서를 표시하는 최소 제곱 적합 보고서



표준 최소 제곱 분석법 시작

표준 최소 제곱은 모형 적합 시작 창에서 선택할 수 있는 여러 분석 기법 중 하나입니다.

이 섹션에서는 모형 적합 시작 창에서 표준 최소 제곱을 적합 방법으로 선택하는 방법을 설명합니다. 이 선택 항목과 관련된 옵션에 대한 내용도 다룹니다. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

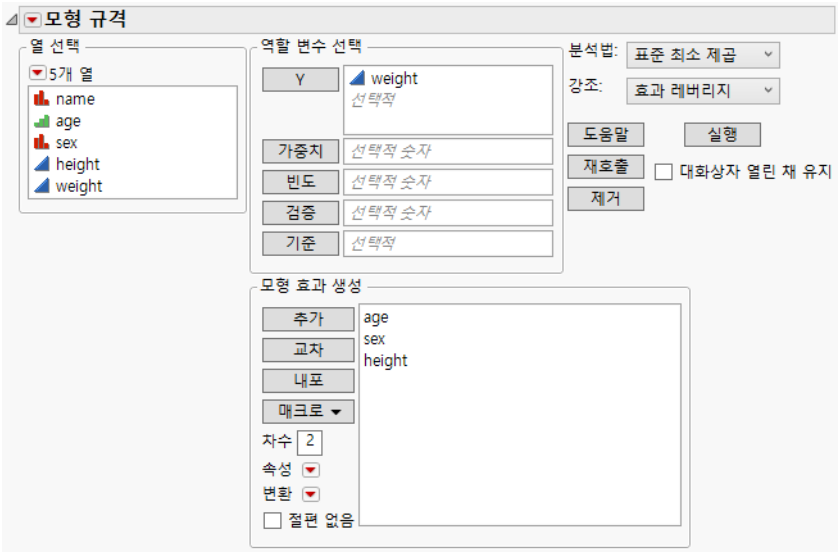
모형 적합 시작 창

모형 적합 시작 창에서 고정 효과와 임의 효과가 모두 있는 모형을 지정할 수 있습니다. 옵션은 사용자가 지정한 모형의 특성에 따라 다릅니다.

고정 효과만

표준 최소 제곱 분석법을 사용하여 모형을 적합시키려면 **분석 > 모형 적합**을 선택한 후 " 분석법 " 목록에서 **표준 최소 제곱**을 선택합니다. Y 목록에 하나 이상의 연속형 변수를 입력하면 분석법이 기본적으로 " 표준 최소 제곱 " 으로 설정됩니다 . 그러나 연속형 Y 변수에 대해 다른 항목을 선택할 수 있습니다 . 표준 최소 제곱 적합에 대해 고정 효과만 지정하면 " 모형 적합 " 시작 창이 그림 3.5 에 표시된 것과 같이 나타납니다 . 이 예에서는 Big Class.jmp 샘플 데이터 테이블을 사용한 시작 창을 보여 줍니다 .

그림 3.5 고정 효과 모형에 대한 모형 적합 시작 창



"분석법" 목록에서 "표준 최소 제곱" 분석법이 선택되면 "강조" 옵션도 나타납니다. "강조" 옵션은 초기 보고서 창에 제공되는 보고서를 제어합니다. JMP는 포함된 모형 효과를 기반으로 사용자에게 필요한 보고서를 추론합니다. 그러나 초기 보고서의 일부로 표시되지 않은 보고서도 기본 보고서의 빨간색 삼각형 메뉴에서 적절한 옵션을 선택하여 표시할 수 있습니다.

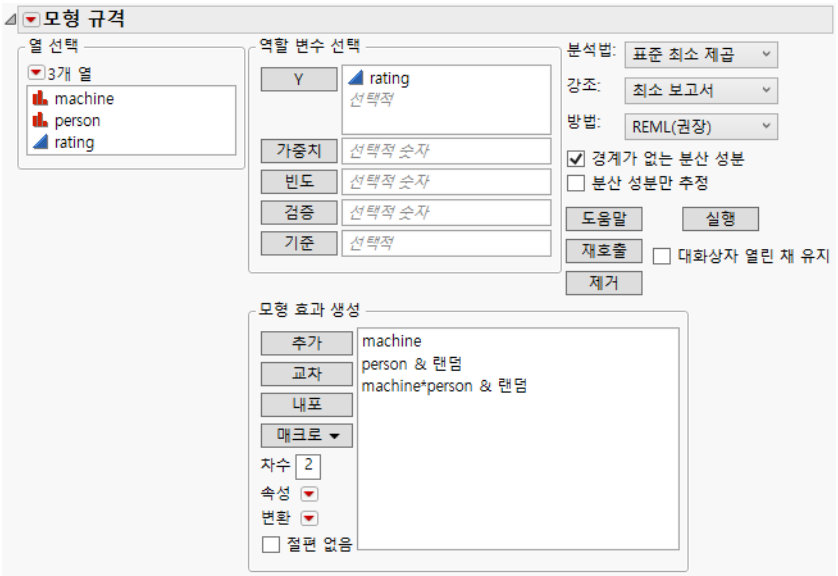
각 "강조" 옵션에서 사용할 수 있는 보고서에 대한 자세한 내용은 "표준 최소 제곱에 대한 강조 옵션"(77 페이지)에서 확인하십시오.

임의 효과

지정된 모형에 하나 이상의 임의 효과가 포함된 경우 "모형 적합" 시작 창에서 추가 옵션을 사용할 수 있습니다. Machine.jmp 샘플 데이터 테이블을 예로 들어보겠습니다. 무작위로 선택된 6 명의 작업자가 각각 3 대의 기계에서 작업을 수행하고 출력이 평가됩니다. 6 명의 특정 작업자에 대한 평가가 다른지 여부를 판단하는 것보다 직원 전체의 평가에서 변동을 추정하는 데 관심이 있습니다. 이를 위해서는 모형을 지정할 때 person 과 machine*person 을 임의 효과로 처리해야 합니다.

이 모형에 대한 모형 적합 시작 창이 그림 3.6 에 나와 있습니다 . person 에 임의 효과 속성이 적용되면 " 방법 " 옵션 및 분산 성분과 관련된 두 가지 옵션이 " 모형 적합 " 시작 창에 나타납니다 .

그림 3.6 임의 효과를 포함하는 모형에 대한 모형 적합 시작 창



모형 적합 시작 창의 표준 최소 제곱 옵션

다음 모형 적합 시작 창 옵션은 표준 최소 제곱 분석법과 관련이 있습니다 .

강조 초기 보고서 창에 나타나는 보고서 및 그림 유형을 제어합니다 . 자세한 내용은 " 표준 최소 제곱에 대한 강조 옵션 "(77 페이지) 에서 확인하십시오 .

방법 (임의 효과가 지정된 경우에만 표시됨) 다음 방법 중 하나를 사용하여 모형을 추정합니다.

REML 자세한 내용은 "REML 분산 성분 추정값 "(178 페이지) 에서 확인하십시오 .

EMS 기대 평균 제곱을 나타내며 적률 방법이라고도 합니다 . 자세한 내용은 "EMS(전통적) 모형 적합 보고서 "(181 페이지) 에서 확인하십시오 .

비유계 분산 성분 (" 방법 " 에 "REML" 이 선택된 경우에만 표시됨) 분산 성분 추정값이 음수가 될 수 있도록 허용합니다 . 이 옵션은 기본적으로 선택되어 있습니다 . 분산 추정값을 0 으로

제한하면 고정 효과 검정에서 편향이 발생할 수 있으므로 고정 효과에 관심이 있다면 이 옵션이 선택된 상태여야 합니다. 자세한 내용은 "음의 분산"(175 페이지)에서 확인하십시오.

분산 성분만 추정 ("방법"에 "REML"이 선택된 경우에만 표시됨) 분산 성분 추정값만 표시하는 보고서를 제공합니다. 자세한 내용은 "분산 성분만 추정"(176 페이지)에서 확인하십시오.

별도 적합 (Y 변수가 여러 개이고 임의 효과가 없는 모형에만 표시됨) 비결측인 모든 행을 사용하여 각 Y 변수에 대해 별도의 모형을 적합시킵니다. 자세한 내용은 "결측값"(78 페이지)에서 확인하십시오.

표준 최소 제곱에 대한 강조 옵션

"강조" 목록의 세 옵션은 표준 최소 제곱 분석법에 대한 초기 보고서의 일부로 표시되는 그림 및 보고서 유형을 제어합니다. 아래 설명을 참조하십시오. JMP에서는 데이터 테이블의 행 수, "모형 효과 생성" 목록에 입력된 효과 수 및 효과에 적용된 속성을 기반으로 기본 강조 옵션이 선택됩니다. 사용자가 필요에 따라 이 강조 선택을 변경할 수 있습니다. JMP에서 강조가 선택되는 방법에 대한 자세한 내용은 "강조 규칙"(187 페이지)에서 확인하십시오.

초기 보고서가 열리면 플랫폼 보고서 창의 빨간색 삼각형 메뉴에서 다른 보고서와 그림을 추가할 수 있습니다.

다음 강조 옵션을 사용할 수 있습니다.

효과 레버리지 모형 적합 상세 정보가 포함된 보고서 및 레버리지 그림과 잔차 그림을 표시합니다. 이 옵션은 모형 적합에 중점을 두는 경우에 유용합니다.

효과 레버리지는 가장 포괄적인 옵션입니다. 이 강조 옵션은 보고서를 전체 모형 관련 보고서와 개별 모형 효과 관련 보고서로 나눕니다. 전체 모형 보고서는 보고서 창의 왼쪽에 "전체 모형"이라는 제목 아래에 표시되고, 효과 보고서는 오른쪽에 표시됩니다.

효과 선별 정렬되거나 척도화된 모수 추정값 보고서와 그래프(해당하는 경우), 예측 프로파일러 및 모형 적합 상세 정보가 포함된 보고서를 표시합니다. 이 옵션은 효과가 많이 있고 선별 설계와 같이 초기에 활성 효과를 찾는 데 중점을 두는 경우에 유용합니다.

"효과 선별"을 선택하면 Box-Cox 변환이 계산됩니다. 추정된 λ에 대한 신뢰 구간에 1이 포함되지 않으면 "Box-Cox 변환" 보고서가 나타납니다. 자세한 내용은 "Y의 Box-Cox 변환"(156 페이지)에서 확인하십시오.

최소 보고서 회귀 그림 및 모형 적합 상세 정보가 포함된 보고서만 표시합니다. 이 강조 옵션은 임의 효과 속성이 모형 효과에 적용될 때 기본값입니다.

이 옵션은 최소한의 상세 정보와 가장 간략한 정보를 제공합니다. 빨간색 삼각형 메뉴에서 관심 있는 특정 보고서를 요청할 수 있습니다.

모든 강조 옵션에 대해 표시되는 보고서 또는 그림을 변경하려면 플랫폼 환경 설정을 사용하십시오. **파일 > 환경 설정 > 플랫폼 > 최소 제곱 적합**으로 이동하여 **설정** 체크박스를 사용합니다.

- 옵션이 보고서에 표시되지 않도록 하려면 옵션 옆의 **설정**을 선택하고 해당 옵션은 선택하지 않습니다.
- 옵션이 보고서에 표시되도록 하려면 **설정**을 선택하고 해당 옵션을 선택합니다.

JMP PRO 표준 최소 제곱의 검증

JMP Pro 에서는 " 모형 적합 " 창에 검증 열을 지정할 수 있습니다 . 검증 열은 숫자 데이터 유형이어야 하고 두 개 이상의 구분되는 값을 포함해야 합니다 .

- 열에 두 개의 값이 포함된 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 열에 세 개의 값이 포함된 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 열에 네 개 이상의 구분되는 값이 포함된 경우에는 검증 열이 무시됩니다 .

JMP Pro 의 표준 최소 제곱 분석법은 검증 열 사용을 지원합니다 . 검증 열을 입력하면 " 교차 검증 " 보고서가 제공됩니다 . 자세한 내용은 " 교차 검증 보고서 "(82 페이지) 에서 확인하십시오 .

검증 열을 입력하면 검증 데이터 집합과 테스트 데이터 집합의 관측값이 보고서의 그림에서 각각 'v' 와 't' 로 표시됩니다 .

JMP 모델링 플랫폼에서 검증 열이 사용되는 방법에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

결측값

기본적으로 Y 또는 모형 효과에 대한 결측값이 있는 행은 분석에서 제외됩니다 .

참고 : JMP 에서는 " 모형 적합 " 창의 " 모형 규격 " 빨간색 삼각형 메뉴에 " 결측값 정보화 " 옵션을 제공합니다 . " 결측값 정보화 " 옵션을 사용하면 모형 효과가 결측인 행을 사용하여 모형을 적합시킬 수 있습니다 . 자세한 내용은 " 모형 규격 " 장의 " 결측값 정보화 "(46 페이지) 에서 확인하십시오 .

모형에 임의 효과가 포함된 경우 Y 값은 기본적으로 별도로 적합됩니다 . 개별 보고서가 " 적합 그룹 " 보고서에 나타납니다 .

모형에 고정 효과만 포함되어 있고 다음 조건에 해당되는 경우를 가정해 보겠습니다 .

- 둘 이상의 Y 반응을 지정했습니다 .
- 이 중 일부 Y 반응에 결측값이 있습니다 .
- " 모형 적합 " 시작 창에서 " 별도 적합 " 옵션을 선택하지 않았습니다 .

그러면 다음 옵션 중 하나를 선택하라는 메시지가 표시됩니다 .

- 별도 적합 - 해당하는 특정 Y 에 대해 비결측인 모든 행을 사용하여 각 Y 를 적합시킵니다 .
- 함께 적합 - 모든 Y 변수에 대해 비결측인 행만 사용하여 각 Y 를 적합시킵니다 .

" 별도 적합 " 을 선택하면 " 적합 그룹 " 보고서에 Y 변수에 대한 개별 보고서가 포함됩니다 . " 적합 그룹 " 의 빨간색 삼각형 메뉴에서 프로파일러를 선택하여 동일한 프로파일러에서 모든 Y 변수를

볼 수 있습니다. 아니면 개별 Y 변수 보고서에서 프로파일러를 선택하여 프로파일러에서 해당 변수만 볼 수도 있습니다.

"함께 적합"을 선택하면 "최소 제곱 적합" 보고서에 각 Y 변수에 대한 개별 보고서가 포함됩니다. 그러나 보고서의 일부는 모든 Y 변수에 대해 결합됩니다(예: 효과 요약 및 프로파일러).

최소 제곱 적합 보고서

표준 최소 제곱 분석법을 사용하여 모형을 적합시키면 " 최소 제곱 적합 " 보고서가 생성됩니다 . 이 보고서의 내용은 데이터 특성과 " 모형 적합 " 시작 창에서 선택한 옵션에 따라 결정됩니다 .

팁 : 기본적으로 나타나지 않는 보고서를 항상 표시하려면 **파일 > 환경 설정 > 플랫폼 > 최소 제곱 적합**을 사용하여 해당 보고서를 선택하십시오 .

단일 반응과 다중 반응의 비교

단일 반응 변수 Y를 적합시키는 경우 "최소 제곱 적합" 창에서 "반응 Y"라는 제목의 보고서에 여러 상세 보고서가 구성됩니다. 여러 반응을 적합시키는 경우에는 일반적으로 "최소 제곱 적합"이라는 제목의 보고서에 개별 반응에 대한 보고서가 구성됩니다. 그러나 결측 반응 데이터가 있을 때 "별도 적합" 옵션을 선택하면 개별 반응에 대한 보고서가 "적합 그룹"이라는 제목의 보고서에 구성됩니다.

강조와 관련된 보고서 구조

"모형 적합" 시작 창에서 "효과 레버리지" 강조를 선택하면 지정된 반응에 대한 보고서가 열 형태로 배열됩니다. 왼쪽 열은 모형과 관련된 추가 보고서를 포함하는 "전체 모형" 보고서로 구성됩니다. 모형의 각 효과에 대한 보고서는 "전체 모형" 보고서의 오른쪽 열에 표시됩니다.

"모형 적합" 시작 창에서 "효과 선별" 또는 "최소 보고서" 강조를 선택하면 각 반응에 대한 모든 보고서가 왼쪽 열에 배열됩니다.

특수 보고서

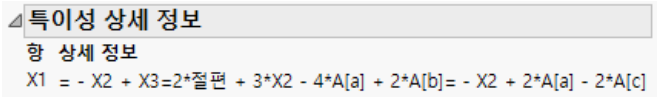
이 섹션에서는 효과 속성과 관련하여 사용자가 선택한 옵션 또는 데이터 구조에 따라 사용할 수 있는 보고서에 대해 설명합니다.

특이성 상세 정보

모형 효과 사이에 선형 종속성이 있는 경우 " 특이성 상세 정보 " 보고서가 반응 보고서 제목 아래에 첫 번째 보고서로 나타납니다 . 이 보고서에는 모형 항이 충족하는 선형 함수 테이블이 포함됩

니다 . 이러한 함수는 모형 항 사이의 별칭 관계를 정의합니다 . 그림 3.7 에서는 Singularity.jmp 샘플 데이터 테이블에 대한 예를 보여 줍니다 .

그림 3.7 특이성 상세 정보 보고서

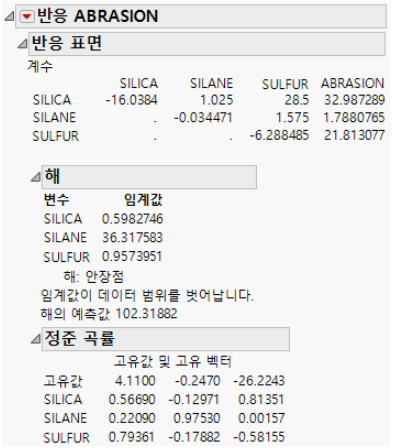


효과 사이에 선형 종속성이 있는 경우 일부 모형 항의 추정값이 고유하지 않습니다. 자세한 내용은 "모형 항 사이에 선형 종속성이 있는 모형"(182 페이지)에서 확인하십시오.

반응 표면 보고서

모형 효과에 반응 표면 (&RS) 또는 혼합물 반응 표면 (&RS& 혼합물) 속성이 있는 경우 " 반응 표면 " 보고서가 제공됩니다 . Tiretread.jmp 샘플 데이터 테이블에 대한 " 반응 표면 " 보고서의 예는 그림 3.8 에서 확인하십시오 .

그림 3.8 반응 표면 보고서



계수 테이블

"반응 표면" 보고서의 첫 번째 부분으로 표시된 "계수" 테이블에는 추정된 모형 모수가 간략하게 요약되어 있습니다. 첫 번째 열은 2 차 항의 계수를 제공합니다. 마지막 열은 선형 항의 계수를 제공합니다. 예측 표현식을 전체적으로 표시하려면 보고서의 빨간색 삼각형에서 **추정값** > **예측 표현식 표시**를 선택합니다.

해 보고서

"해" 보고서에는 임계값(최대값, 최소값 또는 안장점) 및 해당 점에서의 예측값이 제공됩니다. 또한 해가 데이터 범위를 벗어나는 경우 사용자에게 알려 줍니다.

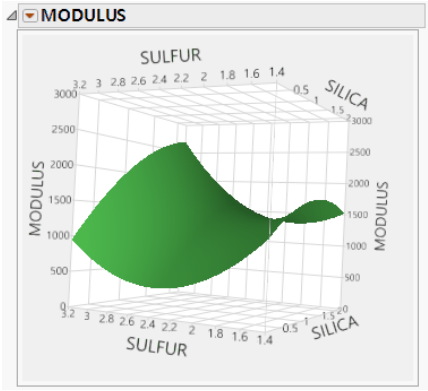
정준 곡률 보고서

2 차 모수 추정값 행렬의 고유값과 고유 벡터는 곡률 유형을 결정합니다. 고유 벡터는 가장 큰 곡률과 가장 작은 곡률의 방향을 포함하여 표면의 주곡률 방향을 보여 줍니다.

고유값은 " 정준 곡률 " 테이블의 첫 번째 행에 제공됩니다 .

- 고유값이 음수이면 반응 표면이 최대값에서 아래쪽으로 곡선을 이룹니다 .
- 고유값이 양수이면 표면 모양이 최소값에서 위쪽으로 곡선을 이룹니다 .
- 양수 고유값과 음수 고유값이 모두 있으면 표면이 안장 모양이 되어 한쪽은 위로 , 다른 쪽은 아래로 각각 곡선을 이룹니다 . Tiretread.jmp 샘플 데이터 테이블을 사용한 예는 그림 3.9 에서 확인하십시오 .

그림 3.9 안장 모양 표면을 사용한 표면 프로파일러 그림



고유값 아래에 나열된 고유 벡터는 주 축의 방향을 보여 줍니다. 고유값의 절대값이 클수록 관련된 방향으로 반응 표면의 곡률이 커집니다. 경우에 따라 0 고유값이 발생합니다. 이 고유값은 적합된 표면이 해당 고유 벡터가 설명하는 방향을 따라 편평하다는 의미입니다.

참고 : 20 개가 넘는 요인으로 구성된 반응 표면 모형의 경우 반응 표면 보고서가 표시되지 않습니다 . 오류 메시지 또는 경고가 제공되지 않습니다 . 반응 표면 설계에 대한 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오 .

혼합 및 임의 효과 모형 보고서

"모형 적합" 시작 창에서 임의 효과를 지정하면 "방법" 목록이 나타납니다. 이 목록에는 "REML(권장)" 과 "EMS(전통적)" 라는 두 가지 적합 방법이 제공됩니다. 선택한 모형과 방법에 따라 " 열 저장 " 및 " 프로파일러 " 옵션과 추가 보고서가 표시됩니다.

REML 방법 보고서에 대한 자세한 내용은 "REML(제한 최대 가능도) 방법 "(176 페이지) 에서 확인하십시오. EMS 방법 보고서에 대한 자세한 내용은 "EMS(전통적) 모형 적합 보고서 "(181 페이지) 에서 확인하십시오.

교차 검증 보고서

JMP PRO " 모형 적합 " 시작 창에 검증 열을 입력하면 " 교차 검증 " 보고서가 표시됩니다 . 이 보고서에는 검증에 사용된 각 데이터 집합에 대한 다음 정보가 제공됩니다 .

소스 데이터 집합을 훈련 데이터 집합 , 검증 데이터 집합 또는 테스트 데이터 집합으로 식별합니다 .

R² 훈련 데이터 집합을 사용하여 파생된 모형과 관련하여 주어진 데이터 집합의 관측값에 대해 계산된 R² 값입니다 . 훈련 데이터 집합의 경우 이 값은 일반적인 R² 값입니다 .

각 훈련 , 검증 및 테스트 데이터 집합의 R² 값은 다음과 같이 계산됩니다 .

- 주어진 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다 . 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다 .
- 예측 오차를 제곱하고 합하여 SSE_{Source} 를 구합니다 . 여기서 서브스크립트 *Source* 는 훈련 , 검증 또는 테스트 데이터 집합을 나타냅니다 .
- *Source* 데이터 집합의 관측값에 대한 실제 반응과 평균 간의 차이를 제곱하고 합합니다 . 이 값을 SST_{Source} 로 나타냅니다 .
- *Source* 데이터 집합의 R² 은 다음과 같습니다 .

$$R^2_{Source} = 1 - \frac{SSE_{Source}}{SST_{Source}}$$

참고 : 검증 데이터 집합과 테스트 데이터 집합의 R² 값은 음수일 수 있습니다 .

RASE 평균 제곱 예측 오차의 제곱근입니다 . 각 훈련 , 검증 및 테스트 데이터 집합의 RASE 는 다음과 같이 계산됩니다 .

- 주어진 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다 . 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다 .
- 예측 오차를 제곱하고 합하여 SSE_{Source} 를 구합니다 . 여기서 서브스크립트 *Source* 는 훈련 , 검증 또는 테스트 데이터 집합을 나타냅니다 .
- 관측값 수를 n 으로 나타냅니다 .
- RASE 는 다음과 같습니다 .

$$RASE_{Source} = \sqrt{\frac{SSE_{Source}}{n}}$$

빈도 소스 데이터 집합의 관측값 수입니다 .

최소 제곱 적합 옵션

Y 가 두 개 이상 있고 결측 반응 값이 없거나 , 결측값이 있는 Y 가 두 개 이상 있을 수 있습니다 . " 함께 적합 " 을 선택하면 반응이 " 최소 제곱 적합 " 보고서에 그룹화됩니다 . 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

프로파일러 모든 반응을 하나의 프로파일러에 표시합니다 . 모든 반응에 대한 모형 항의 효과를 동시에 보고 여러 최적화를 수행할 수 있습니다 . 자세한 내용은 " 요인 프로파일링 "(150 페이지) 에서 확인하십시오 .

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다 .

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 " 효과 요약 " 보고서를 표시합니다 . 자세한 내용은 " 효과 요약 보고서 "(168 페이지) 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

적합 그룹 옵션

Y 가 두 개 이상 있고 " 별도 적합 " 을 선택하면 반응이 " 적합 그룹 " 보고서에 그룹화됩니다 . 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

프로파일러 모든 반응을 하나의 프로파일러에 표시합니다 . 모든 반응에 대한 모형 항의 효과를 동시에 보고 여러 최적화를 수행할 수 있습니다 . 자세한 내용은 " 프로파일러 "(151 페이지) 에서 확인하십시오 .

등고선 프로파일러 모든 반응을 하나의 등고선 프로파일러에 표시합니다 . 모든 반응에 대한 모형 항의 효과를 동시에 탐색할 수 있습니다 .

표면 프로파일러 각 반응에 대한 별도의 표면 프로파일러 보고서를 표시합니다 .

행으로 배열 플랫폼 분석 보고서를 지정된 수의 행으로 재배열합니다 . 이 옵션은 창 또는 출력 페이지에 더 많은 보고서가 들어가도록 보고서를 배열하는 데 주로 사용됩니다 .

적합도를 기준으로 정렬 적합 유의성 (R^2) 을 기준으로 보고서를 정렬합니다 . 이 옵션은 R^2 이 플랫폼 수준으로 표시되는 플랫폼에만 사용할 수 있습니다 . 예를 들어 하나의 시작 창에서 수

백 개의 일원 분석을 생성하면 **적합 그룹**에 분석이 나타나고, 가장 강한 관계가 먼저 나타나도록 정렬할 수 있습니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

반응 옵션

반응에 대한 빨간색 삼각형 메뉴를 사용하면 필요에 따라 보고서를 사용자 정의할 수 있습니다.

회귀 보고서 기본 보고서와 보고서 옵션을 제공합니다. 자세한 내용은 "회귀 보고서"(85 페이지)에서 확인하십시오.

추정값 모수 추정값과 관련된 추가 분석을 위한 옵션을 제공합니다. 자세한 내용은 "추정값"(105 페이지)에서 확인하십시오.

효과 선별 유의한 효과를 식별하기 위한 보고서 및 그림을 제공합니다. 자세한 내용은 "효과 선별"(140 페이지)에서 확인하십시오.

요인 프로파일링 반응이 모형 항과 어떻게 관련되어 있는지 검토하기 위한 프로파일러, 교호작용 및 입방체도를 제공합니다. Box-Cox 변환 적합을 위한 그림과 보고서도 제공합니다. 자세한 내용은 "요인 프로파일링"(150 페이지)에서 확인하십시오.

행 진단 잔차를 검토하기 위한 그림 및 보고서를 제공합니다. 또한 PRESS 통계량을 보고하고 Durbin-Watson 검정을 제공합니다. 자세한 내용은 "행 진단"(159 페이지)에서 확인하십시오.

열 저장 결과를 별도의 데이터 테이블에 저장하는 "코딩 테이블 저장"을 제외하고 모형 결과를 데이터 테이블의 열로 저장합니다. 자세한 내용은 "열 저장"(165 페이지)에서 확인하십시오.

모형 대화상자 현재 분석을 위해 완료된 "모형 적합" 시작 창을 표시합니다.

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 "효과 요약" 보고서를 표시합니다. 자세한 내용은 "효과 요약 보고서"(168 페이지)에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

회귀 보고서

" 회귀 보고서 " 메뉴는 모형 적합 , 효과 유의성 및 모형 모수에 대한 요약 정보를 제공합니다 .

적합 요약 모형 적합 요약을 표시하거나 숨깁니다 . 자세한 내용은 " 적합 요약 "(85 페이지) 에서 확인하십시오 .

분산 분석 적합 모형을 단순 평균 모형과 비교하기 위한 계산을 표시하거나 숨깁니다 . 자세한 내용은 " 분산 분석 "(86 페이지) 에서 확인하십시오 .

모수 추정값 모수 추정값 및 각 모수가 0 이라는 가설에 대한 *t*- 검정이 포함된 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 모수 추정값 "(87 페이지) 에서 확인하십시오 .

효과 검정 모형의 고정 효과에 대한 검정을 표시하거나 숨깁니다. 자세한 내용은 "효과 검정"(88 페이지) 에서 확인하십시오 .

효과 상세 정보 개별 효과에 대한 상세 정보 , 그림 및 검정이 포함된 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 효과 상세 정보 "(89 페이지) 에서 확인하십시오 .

" 효과 레버리지 " 강조 옵션을 선택하면 " 최소 제곱 적합 " 보고서 창 맨 위에 각 효과의 개별 보고서가 제공됩니다 . 이 보고서에는 레버리지 그림과 효과 상세 정보 옵션이 포함됩니다 . 자세한 내용은 " 레버리지 그림 "(161 페이지) 에서 확인하십시오 .

적합 결여 검정을 수행할 수 있을 때 모형에 적절한 효과가 있는지 평가하는 검정을 표시하거나 숨깁니다 . 자세한 내용은 " 적합 결여 "(104 페이지) 에서 확인하십시오 .

모든 신뢰 구간 표시 다음 통계량에 대한 신뢰 구간을 표시하거나 숨깁니다 .

- " 모수 추정값 " 보고서의 모수 추정값
- " 최소 제곱 평균 " 테이블의 최소 제곱 평균

AICc " 적합 요약 " 보고서에 AICc 및 BIC 값을 표시하거나 숨깁니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

적합 요약

" 적합 요약 " 보고서에는 R^2 계산과 AICc 및 BIC 값과 같은 상세 정보가 제공됩니다 .

R^2 랜덤 오차가 아니라 모형에 기인할 수 있는 반응의 변동 비율을 추정합니다 . R^2(다중 결정계수라고도 함) 은 해당 " 분산 분석 " 테이블의 값을 사용하여 다음과 같이 계산됩니다 .

$$\frac{\text{Sum of Squares(Model)}}{\text{Sum of Squares(C. Total)}}$$

R^2 이 1 에 가까울수록 R^2 이 0 에 가까울 때보다 데이터에 더 잘 적합된다는 것을 나타냅니다 . R^2 이 0 에 가까우면 모형이 반응 평균보다 더 나은 반응 예측 변수가 아님을 나타냅니다 .

참고 : R^2 값이 낮으면 설명되지 않은 변동을 설명하는 변수가 모형에 없는 것일 수 있습니다 . 하지만 데이터의 내재 변동 범위가 큰 경우에는 유용한 회귀 모형이라도 R^2 값이 낮을 수 있습니다 . 일반적인 R^2 값에 대해 알아보려면 연구 영역의 문헌을 읽어 보십시오 .

Adj-R^2 모형의 모수 수에 대한 R^2 통계량을 조정합니다 . Adj-R^2 을 사용하면 모수 수가 다른 모형을 쉽게 비교할 수 있습니다 . 계산에 자유도가 사용됩니다 . Adj-R^2 은 해당 " 분산 분석 " 테이블의 값을 사용하여 다음과 같이 계산됩니다 .

$$1 - \frac{\text{Mean Square(Error)}}{\text{제곱합(수정 합계)/DF(수정 합계)}}$$

제곱근 평균 제곱 오차 랜덤 오차의 표준편차를 추정합니다 . 이 값은 " 분산 분석 " 보고서에 표시된 오차 평균 제곱의 제곱근입니다 .

참고 : 제곱근 평균 제곱 오차는 일반적으로 *RMSE* 라고 합니다 .

반응 평균 반응 값의 전체 평균을 표시합니다 .

관측값 수 (또는 가중합) 모형에 사용된 관측값 수를 표시합니다 .

- 이 값은 결측값 , 제외된 행 , 가중치 또는 빈도 역할에 할당된 열이 없는 경우 데이터 테이블의 행 수와 동일합니다 .
- 이 값은 가중치 역할에 할당된 열이 있는 경우 가중치 열에 포함된 양수 값의 합입니다 .
- 이 값은 빈도 역할에 할당된 열이 있는 경우 빈도 열에 포함된 양수 값의 합입니다 .

AICc (" 회귀 보고서 " 메뉴에서 "AICc" 옵션을 선택했거나 , " 최소 제곱 적합 " 환경 설정에서 "AICc" 를 설정한 경우에만 표시됨) 수정된 AICc(Akaike 정보 기준) 값을 표시합니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

BIC (" 회귀 보고서 " 메뉴에서 "AICc" 옵션을 선택했거나 , " 최소 제곱 적합 " 환경 설정에서 "AICc" 를 설정한 경우에만 표시됨) BIC(Bayes 정보 기준) 값을 표시합니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

분산 분석

"분산 분석" 보고서에는 적합 모형을 모든 예측값이 반응 평균과 동일한 모형과 비교하기 위한 계산이 제공됩니다.

참고 : " 모형 적합 " 시작 창에 빈도 또는 가중치 변수를 입력하면 " 빈도 "(36 페이지) 및 " 가중치 "(36 페이지) 에 나오는 설명과 일치하도록 " 분산 분석 " 보고서의 항목이 조정됩니다 .

" 분산 분석 " 보고서에는 다음과 같은 열이 포함됩니다 .

소스 세 가지 변동 소스이며 모형 , 오차 및 수정 합계 중 하나입니다 .

DF 각 변동 소스에 대한 관련 DF(자유도) 입니다 . 수정 합계 DF 는 항상 관측값 수에서 1 을 뺀 값이며 다음과 같이 모형 자유도와 오차 자유도로 분할됩니다 .

- 모형 DF 는 모형 적합에 사용되는 모수 (절편 제외) 의 수입니다 .
- 오차 DF 는 수정 합계 DF 와 모형 DF 의 차이입니다 .

제곱합 각 변동 소스에 대한 관련 SS(제곱합) 입니다 .

- 총 (수정 합계) SS 는 반응 값과 표본 평균 간의 차이에 대한 제곱합입니다 . 이 값은 반응 값의 총 변동을 나타냅니다 .
- 오차 SS 는 적합된 값과 실제값 간의 차이에 대한 제곱합입니다 . 이 값은 적합 모형에 의해 설명되지 않은 변동을 나타냅니다 .
- 모형 SS 는 수정 합계 SS 와 오차 SS 간의 차이입니다 . 이 값은 모형에 의해 설명된 변동을 나타냅니다 .

평균 제곱 모형 및 오차 변동 소스에 대한 평균 제곱 통계량입니다 . 각 평균 제곱은 제곱합을 해당 자유도로 나눈 것입니다 .

참고 : 오차에 대한 평균 제곱의 제곱근은 " 적합 요약 " 보고서의 RMSE 와 동일합니다 .

사용된 MSE (" 오차 규격 " 이 " 순수 오차 " 또는 " 지정됨 " 인 경우에만 표시됨) 기본 오차 규격이 선택되지 않은 경우 사용되는 평균 제곱 오차입니다 . 이 값은 평균 제곱 열의 평균 제곱 오차 대신 F 비를 계산하는 데 사용됩니다 .

사용된 DFE (" 오차 규격 " 이 " 순수 오차 " 또는 " 지정됨 " 인 경우에만 표시됨) 기본 오차 규격이 선택되지 않은 경우 사용되는 오차 자유도입니다 . 이 값은 DF 열의 오차 DF 대신 Prob > F 값을 계산하는 데 사용됩니다 .

F 비 모형 평균 제곱을 오차 평균 제곱으로 나눈 것입니다 . F 비는 모형이 모든 예측값과 반응 평균이 동일한 모형과 유의하게 다른지 여부를 검정하기 위한 검정 통계량입니다 .

Prob > F 검정의 *p* 값입니다 . Prob > F 값은 절편을 제외한 모든 모수가 0 일 때 관측된 것과 동일한 F 비를 얻을 확률을 측정합니다 . Prob > F 값이 작으면 관측된 *F* 비의 발생 확률이 낮음을 나타냅니다 . 이러한 값은 모형에 하나 이상의 유의 효과가 있다는 증거로 간주됩니다 .

모수 추정값

"모수 추정값" 보고서에는 모형 모수의 추정값이 표시되고, 각 모수에 대해 모수가 0이라는 가설에 대한 *t*-검정이 제공됩니다.

참고 : 모형 항 사이에 선형 종속성이 있는 경우에도 가능하면 추정값을 구하고 검정합니다 . 이러한 추정값에는 " 편향 " 또는 " 영점화됨 " 이라는 라벨이 지정됩니다 . 자세한 내용은 " 모형 항 사이에 선형 종속성이 있는 모형 "(182 페이지) 에서 확인하십시오 .

항 추정된 모수에 해당하는 모형 항입니다 . " 모형 적합 " 시작 창에서 " 절편 없음 " 옵션을 선택한 경우를 제외하고 첫 번째 항은 항상 절편입니다 . 연속형 효과는 데이터 테이블 열 이름과 함께 나타납니다 . 고차 항의 일부인 연속형 열은 중심화될 수 있습니다 . 명목형 또는 순서형

효과는 대괄호 안에 수준 값이 포함된 상태로 나타납니다. 명목형 및 순서형 항의 코딩에 대한 자세한 내용은 " 명목형 효과에 대한 코딩 "(139 페이지) 및 " 통계 상세 정보 " 부록의 " 요인 모형 "(492 페이지) 에서 확인하십시오.

추정값 각 항에 대한 모수 추정값입니다. 이 값은 모형 계수의 추정값입니다. 모형 항 사이에 선형 종속성이 있으면 " 편향 " 또는 " 영점화됨 " 이라는 라벨이 지정될 수 있습니다. 자세한 내용은 " 모형 항 사이에 선형 종속성이 있는 모형 "(182 페이지) 에서 확인하십시오.

표준 오차 추정된 각 모수에 대한 표준 오차 추정값입니다.

***t* 비** 각 모수의 실제 값이 0 인지 여부를 검정합니다. *t* 비는 표준 오차에 대한 추정값의 비율입니다. 모형에 대한 일반적인 가정이 주어지면 *t* 비는 귀무가설 하에서 스튜던트 *t* 분포를 따릅니다.

Prob>|t| 실제 모수 값이 0 이라는 검정에 반대되는 양측 대립가설에 대한 *p* 값입니다.

95% 하한 모수 추정값에 대한 95% 신뢰 하한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 하한을 선택한 경우에만 나타납니다.

95% 상한 모수 추정값에 대한 95% 신뢰 상한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 상한을 선택한 경우에만 나타납니다.

표준화 베타 모든 항이 평균 0, 분산 1 로 표준화된 회귀 모형의 모수 추정값입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 표준화 베타를 선택한 경우에만 나타납니다.

VIF 모형의 각 항에 대한 분산 팽창 계수입니다. VIF 값이 높으면 모형의 항 사이에 공선성 문제가 있음을 나타냅니다.

i 번째 항 x_i 의 VIF 는 다음과 같이 정의됩니다.

$$VIF_i = \frac{1}{1 - R_i^2}$$

여기서 R_i^2 은 다른 설명 변수의 함수로 나타낸 x_i 회귀에 대한 R^2(다중 결정 계수) 입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > VIF 를 선택한 경우에만 나타납니다.

설계 표준 오차 모수 추정값의 상대 분산 제곱근입니다 (Goos and Jones 2011, p. 25).

$$\sqrt{diag(X'X)^{-1}}$$

이러한 값은 RMSE 로 나눈 표준 오차입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 설계 표준 오차를 선택한 경우에만 나타납니다.

효과 검정

"효과 검정" 보고서는 모형에 고정 효과가 있는 경우에만 나타납니다. 주어진 효과에 대한 효과 검정은 해당 효과와 관련된 모든 모수가 0 이라는 귀무가설을 검정합니다. 단일 연속형 설명 변수의 경우 효과에 하나의 모수만 있을 수 있습니다. 이 경우 검정은 "모수 추정값" 보고서의 해당 항에 대한 *t*-검정과 동등합니다. 명목형 또는 순서형 효과에는 수준 수에 따라 관련된 모수가 여

러 개 있을 수 있습니다. 이러한 효과에 대한 효과 검정은 관련된 모든 모수가 0 인지 여부를 검정합니다.

다음 사항에 유의하십시오.

- 가능한 경우 항이 선형 종속성과 관련된 효과에 대해 효과 검정이 수행됩니다. 자세한 내용은 " 모형 항 사이에 선형 종속성이 있는 모형 "(182 페이지) 에서 확인하십시오.
- 파라미터화 및 특이성 처리는 SAS GLM 프로시저와 다릅니다. 파라미터화 및 특이성 처리에 대한 자세한 내용은 "통계 상세 정보" 부록의 " 요인 모형 "(492 페이지) 에서 확인하십시오.

" 효과 검정 " 보고서에는 다음 열이 포함됩니다.

소스 모형의 효과입니다.

모수 수 효과와 관련된 모수의 수입니다. 연속형 효과에는 하나의 모수가 있습니다. 명목형 또는 순서형 효과에 대한 모수의 수는 수준 수에서 1 을 뺀 값입니다. 교차 효과에 대한 모수의 수는 각 개별 효과에 대한 모수의 수를 곱한 값입니다.

DF 효과 검정에 대한 자유도입니다. 일반적으로 모수 수와 DF 는 동일합니다. 예측 변수 사이에 선형 종속성이 있는 경우에는 값이 다를 수 있습니다. 이러한 경우 DF 는 모수 수보다 작을 수 있으며 이는 효과와 관련된 하나 이상의 모수를 검정할 수 없음을 나타냅니다. DF 가 모수 수보다 작은 경우에는 항상 LostDFs 가 보고서 줄 오른쪽에 나타납니다. 오차의 자유도가 있는 경우 검정이 수행됩니다. 자세한 내용은 " 효과 검정 보고서 "(185 페이지) 에서 확인하십시오.

제곱합 효과가 0 이라는 가설에 대한 제곱합입니다.

F 비 효과가 0 인지 검정하기 위한 F 통계량입니다. "F 비 " 는 효과의 평균 제곱을 오차 평균 제곱으로 나눈 비율입니다. 효과의 평균 제곱은 효과 제곱합을 자유도로 나눈 값입니다.

Prob > F 효과 검정의 *p* 값입니다.

평균 제곱 효과의 평균 제곱, 즉 효과에 대한 제곱합을 DF 로 나눈 값입니다.

참고 : 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 평균 제곱**을 선택한 경우에만 나타납니다.

효과 상세 정보

" 효과 상세 정보 " 보고서에는 개별 효과에 대한 상세 정보, 그림 및 검정이 제공됩니다. 이 보고서는 " 모형 적합 " 시작 창에서 선택한 강조에 따라 별도의 보고서로 구성됩니다.

- 효과 레버리지 강조 : " 전체 모형 " 보고서 오른쪽에 있는 " 최소 제곱 적합 " 보고서의 위에 각 효과의 개별 보고서가 있습니다. 이 경우 보고서에는 효과에 대한 레버리지 그림이 포함됩니다.
- 효과 선별 또는 최소 보고서 강조 : " 효과 상세 정보 " 보고서가 제공되지만 처음에 닫혀 있습니다. 표시 아이콘을 클릭하면 보고서가 표시됩니다.

보고서의 초기 콘텐츠는 최소 제곱 평균 테이블입니다. 효과의 특성에 따라 이 테이블이 적절하지 않을 수 있으며 초기 기본 보고서에 콘텐츠가 아무 것도 표시되지 않을 수 있습니다. 그러나 특정 빨간색 삼각형 옵션을 사용할 수 있습니다.

효과 옵션 테이블

효과 이름 옆의 빨간색 삼각형 메뉴는 다음 옵션을 제공합니다. 특정 모델링 유형의 경우 일부 옵션은 적절하지 않을 수 있으므로 사용할 수 없습니다 .

최소 제곱 평균 테이블 효과를 검정할 때 비교되는 통계량을 표시합니다 . 자세한 내용은 " 최소 제곱 평균 테이블 "(91 페이지) 에서 확인하십시오 .

연속형 효과에는 이 옵션을 사용할 수 없습니다 .

최소 제곱 평균 그림 명목형 및 순서형 효과에 대한 최소 제곱 평균 그림을 표시합니다. 교호작용 효과의 경우 이 옵션은 " 최소 제곱 평균 그림 옵션 " 창을 표시합니다 . 자세한 내용은 " 최소 제곱 평균 그림 "(92 페이지) 에서 확인하십시오 .

최소 제곱 평균 대비 ... 명목형 효과 , 순서형 효과 및 교호작용의 수준을 비교하기 위해 대비를 지정하고 검정할 수 있는 " 대비 지정 " 창을 표시합니다 . 자세한 내용은 " 최소 제곱 평균 대비 ..."(95 페이지) 에서 확인하십시오 .

최소 제곱 평균 스튜던트 t 스튜던트 *t* 검정을 사용하여 최소 제곱 평균의 쌍별 비교를 위한 검정 및 신뢰 구간을 표시합니다 . 자세한 내용은 " 최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD"(98 페이지) 에서 확인하십시오 .

참고 : 유의 수준은 개별 비교에 적용되고 모든 비교에 전체적으로 적용되지 않습니다 . 비교 모음의 오차율은 개별 검정의 오차율보다 큽니다 .

최소 제곱 평균 Tukey HSD Tukey-Kramer HSD(Honestly Significant Difference) 검정 (Tukey 1953, Kramer 1956) 을 사용하여 최소 제곱 평균의 쌍별 비교를 위한 검정 및 신뢰 구간을 표시합니다 . 자세한 내용은 " 최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD"(98 페이지) 에서 확인하십시오 .

참고 : 유의 수준은 쌍별 비교 모음에 적용됩니다 . 유의 수준은 표본 크기가 같으면 정확하고 표본 크기가 다르면 약간 작습니다 (Hayter 1984).

최소 제곱 평균 Dunnett 지정하는 대조 수준에 대해 쌍별 비교를 위한 검정 및 신뢰 구간을 표시합니다 . 또한 검정 결과 그림도 제공합니다 . 자세한 내용은 " 최소 제곱 평균 Dunnett"(101 페이지) 에서 확인하십시오 .

조각 검정 교호작용에 있는 각 열의 각 수준에 대해 교호작용에 있는 다른 분류 열의 모든 수준 간 쌍별 비교를 공동으로 검정합니다 . 자세한 내용은 " 조각 검정 "(102 페이지) 에서 확인하십시오 .

참고 : 명목형 및 순서형 효과를 포함하는 교호작용에만 사용할 수 있습니다 .

검정력 분석 효과 검정에 대한 검정력을 분석할 수 있는 " 검정력 상세 정보 " 보고서를 표시합니다 . 자세한 내용은 " 검정력 분석 "(103 페이지) 에서 확인하십시오 .

최소 제곱 평균 테이블

최소 제곱 평균은 다른 모형 요인이 **중립** 값으로 설정된 범주형 효과의 수준에 대해 모형에서 예측한 값입니다. 연속형 효과의 중립 값은 표본 평균으로 정의됩니다. 관심 효과와 연관이 없는 명목형 효과의 중립 값은 해당 효과에 대한 계수의 평균입니다. 연관이 없는 순서형 효과의 중립 값은 값 순서화에서 효과의 첫 번째 수준으로 정의됩니다.

최소 제곱 평균은 **조정 평균** 또는 **모집단 주변 평균**이라고도 합니다. 모형에 다른 효과가 있는 경우 최소 제곱 평균이 단순 평균과 다를 수 있습니다. 사실 최소 제곱 평균은 표본 평균보다 서로 더 가까운 것이 일반적입니다. 이러한 예측이 수행되는 중립 값의 특성 때문에 이런 상황이 발생합니다.

최소 제곱 평균은 다른 모형 요인의 특정 값에 대한 예측이므로 서로 비교할 수 있습니다. 효과를 검정할 때 최소 제곱 평균을 사용하여 비교합니다. 최소 제곱 평균에 대한 자세한 내용은 "통계 상세 정보" 부록의 "명목형 요인에 대한 최소 제곱 평균"(496 페이지) 및 "순서형 최소 제곱 평균 "(506 페이지) 에서 확인하십시오 .

주효과의 경우 최소 제곱 평균 테이블에 표본 평균도 포함됩니다 (그림 3.10).

최소 제곱 평균 테이블의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. weight 를 선택하고 **Y** 를 클릭합니다 .
4. age, sex 및 height 를 선택하고 **추가**를 클릭합니다 .
5. **강조** 목록에서 **효과 선별**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. " 효과 상세 정보 " 보고서는 " 최소 제곱 적합 " 보고서의 아래쪽 근처에 나타나고 처음에는 닫혀 있습니다 . "효과 상세 정보" 보고서 옆의 표시 아이콘을 클릭하여 보고서를 표시합니다 .

그림 3.10 에 표시된 것처럼 " 효과 상세 정보 " 보고서에는 세 가지 효과에 대한 보고서가 각각 표시됩니다 . " 최소 제곱 평균 테이블 " 은 age 및 sex 에 제공되지만 연속형 효과인 height 에는 제공되지 않습니다 . 최소 제곱 평균이 표본 평균과 어떻게 다른지 확인하십시오 .

그림 3.10 최소 제곱 평균 테이블

반응 weight

효과 상세 정보

age

최소 제곱 평균 테이블

수준	최소 제곱 평균	표준 오차	평균
12	119.38671	5.5302002	99.000
13	105.70112	5.2695143	94.714
14	93.53945	3.9461891	100.833
15	99.61686	5.1729356	108.286
16	109.22769	7.7788438	118.333
17	121.90696	8.0768318	140.667

sex

최소 제곱 평균 테이블

수준	최소 제곱 평균	표준 오차	평균
F	121.43592	6.1844278	100.944
M	117.33750	5.7895918	108.318

height

"최소 제곱 평균 테이블" 보고서에는 다음 열이 포함됩니다.

수준 범주형 수준 또는 수준 조합입니다.

최소 제곱 평균 각 수준에 대한 최소 제곱 평균의 추정값입니다.

추정 가능성 (최소 제곱 평균을 추정할 수 없는 경우에만 표시됨) 최소 제곱 평균을 추정할 수 없는 경우 경고가 표시됩니다.

표준 오차 각 수준에 대한 최소 제곱 평균의 표준 오차를 표시합니다.

95% 하한 최소 제곱 평균에 대한 95% 신뢰 하한을 표시합니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 하한을 선택한 경우에만 나타납니다.

95% 상한 최소 제곱 평균에 대한 95% 신뢰 상한을 표시합니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 상한을 선택한 경우에만 나타납니다.

평균 지정된 수준에 대한 반응 표본 평균을 표시합니다. 이 평균은 모형의 다른 효과에 대한 값이 이 효과에서 균형을 이루지 못하는 경우 최소 제곱 평균과 다릅니다.

최소 제곱 평균 그림

"최소 제곱 평균 그림" 옵션은 명목형 효과, 순서형 효과 및 교호작용에 대한 최소 제곱 평균 그림을 생성합니다. 교호작용 효과의 경우 이 옵션은 "최소 제곱 평균 그림 옵션" 창을 표시합니다. 자세한 내용은 "최소 제곱 평균 그림 옵션"(93 페이지)에서 확인하십시오.

"최소 제곱 평균 그림" 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

신뢰 한계 표시 그림에 각 추정값의 신뢰 한계를 표시하거나 숨깁니다.

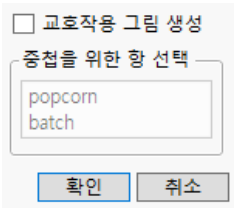
연결된 점 표시 그림에 각 수준에 대한 최소 제곱 평균을 연결하는 하나 이상의 선을 표시하거나 숨깁니다.

제거 지정된 효과에 대한 최소 제곱 평균 그림 보고서를 제거합니다 .

최소 제곱 평균 그림 옵션

교호작용 효과의 빨간색 삼각형 메뉴에서 " 최소 제곱 평균 그림 " 옵션을 선택하면 " 최소 제곱 평균 그림 옵션 " 창이 나타납니다 .

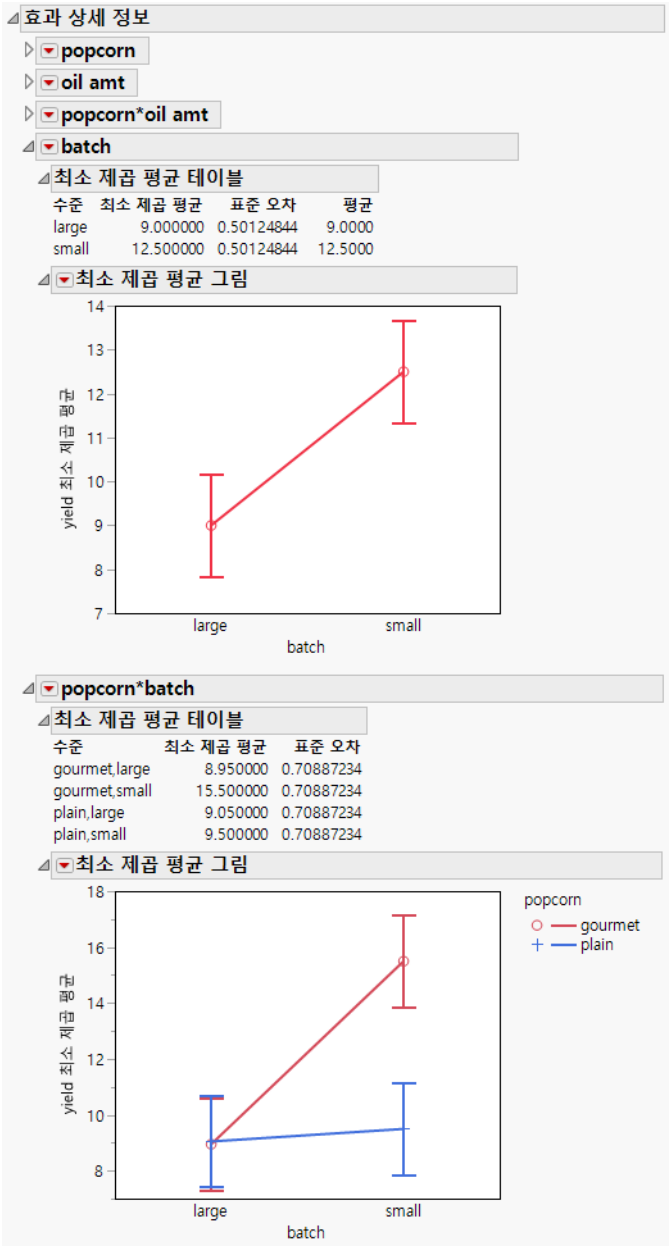
그림 3.11 최소 제곱 평균 그림 옵션 창



창에서 아무 것도 선택하지 않고 "확인"을 클릭하면 최소 제곱 평균 그림 하나가 나타납니다. 그림의 가로 축은 각 조합에 대해 별도의 효과를 얻기 위해 내포된 요인의 수준으로 구성됩니다. 교호작용 그림을 생성하려면 "교호작용 그림 생성" 옆의 상자를 선택합니다. "중첩을 위한 항 선택" 옵션을 사용하면 교호작용 그림에서 중첩 변수로 표시되는 효과를 선택할 수 있습니다.

삼원 교호작용 항의 경우 교호작용 그림에 대한 중첩 변수를 선택하면 두 번째 옵션 패널이 나타납니다. 두 번째 패널에서 아무 것도 선택하지 않고 "확인"을 클릭하면 교호작용 그림 하나가 나타납니다. 아니면 두 번째 옵션 패널을 사용하여 "별도 그림을 위한 항 선택"에서 선택한 효과의 각 수준에 대한 별도의 교호작용 그림을 생성할 수도 있습니다.

그림 3.12 두 효과에 대한 최소 제곱 평균 테이블 및 최소 제곱 평균 그림



최소 제곱 평균 그림의 예

그림 3.12 에 나오는 보고서를 생성하려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Popcorn.jmp 를 엽니다 .

2. **분석 > 모형 적합**을 선택합니다 .
3. Yield 를 선택하고 **Y** 를 클릭합니다 .
4. popcorn, oil amt 및 batch 를 선택하고 **매크로 > 완전 요인**을 클릭합니다 . 그러면 " 강조 " 가 " 효과 선별 " 로 변경됩니다 .
5. **실행**을 클릭합니다 .
6. " 효과 상세 정보 " 표시 아이콘을 클릭하여 7 개 모형 효과의 상세 정보를 표시합니다 .
7. "batch" 의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 그림**을 선택합니다 .
8. "popcorn*batch" 의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 그림**을 선택합니다 . " 최소 제곱 평균 그림 옵션 " 창이 나타납니다 .
9. " 최소 제곱 평균 그림 옵션 " 창에서 " 교호작용 그림 생성 " 옆의 상자를 클릭합니다 .
10. " 중첩을 위한 항 선택 " 에서 popcorn 을 선택합니다 .
11. **확인**을 클릭합니다 .
12. popcorn*batch 에 대한 그림에서 요인을 전치하려면 8 단계 및 9 단계를 반복합니다 .
13. " 중첩을 위한 항 선택 " 에서 batch 를 선택하고 **확인**을 클릭합니다 .

그림 3.13 요인이 전치된 교호작용의 최소 제곱 평균 그림

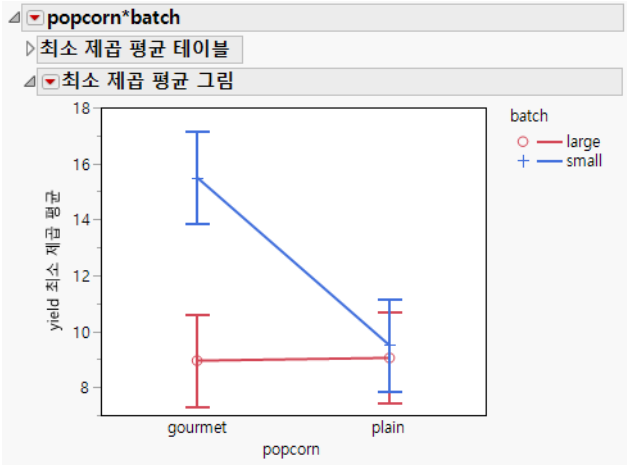


그림 3.13 에서는 요인이 전치된 popcorn*batch 교호작용 그림을 보여 줍니다. 이 그림을 그림 3.12 의 그림과 비교해 봅니다. 이러한 그림은 동일한 정보를 나타내지만 관심 대상에 따라 하나가 다른 것보다 직관적일 수 있습니다.

최소 제곱 평균 대비 ...

대비는 모수 값의 선형 결합입니다. "대비 지정" 창에서 여러 대비를 지정하고 값이 0 인지 여부를 공동으로 검정할 수 있습니다(그림 3.14).

JMP에서는 효과의 최소 제곱 평균에 의해 대비를 구성합니다. 대비의 각 열은 합이 0이 되어 절대값의 합이 2가 되도록 정규화됩니다. 대비에 공변량이 포함된 경우 대비를 검정할 공변량 값을 지정할 수 있습니다.

"대비 지정" 상자에는 효과의 이름과 효과의 수준 이름이 표시됩니다. 대비 값은 처음에 0으로 설정되고 + 및 - 기호를 포함하는 셀 옆에 나타납니다. 수준을 비교하려면 이러한 버튼을 클릭합니다.

+ 또는 - 버튼을 클릭할 때마다 대비 계수는 합이 0이 되고 가능한 경우 절대값의 합이 2가 되도록 정규화됩니다. 추가 수준을 비교하려면 **새 열** 버튼을 클릭합니다. 그러면 새 대비를 정의할 새 열이 나타납니다. 끝나면 **완료**를 클릭합니다. "대비" 보고서가 나타납니다(그림 3.15). 전체 검정은 모든 대비에 대한 결합 *F*-검정입니다.

참고 : 가능한 최대 대비 수보다 많이 지정하려고 하면 검정이 자동으로 평가됩니다 .

" 대비 " 보고서에는 결합 *F*- 검정에 대한 다음 상세 정보가 제공됩니다 .

SS 결합 검정의 제곱합입니다 .

NumDF 분자 자유도입니다 .

DenDF 분모 자유도입니다 .

F 비 SS 를 NumDF 로 나눈 값을 평균 제곱 오차로 나눈 값의 비율입니다 .

Prob > F 유의성 검정의 *p* 값입니다 .

검정 상세 정보 보고서

"검정 상세 정보" 보고서(그림 3.15)에는 검정한 각 대비에 대한 열이 표시됩니다. 이 보고서에는 각 대비에 대해 추정값, 표준 오차, 해당 단일 대비 검정에 대한 *t* 비, 해당 *p* 값, 제곱합, 대비 추정값의 신뢰 구간 등이 제공됩니다.

모수 함수 보고서

"모수 함수" 보고서(그림 3.15)에는 지정한 대비가 모형 항의 선형 결합으로 표시됩니다.

최소 제곱 평균 대비의 예

최소 제곱 평균 대비 옵션을 설명하기 위해 처음 두 age 수준을 다음 두 수준과 비교하는 대비를 구성합니다.

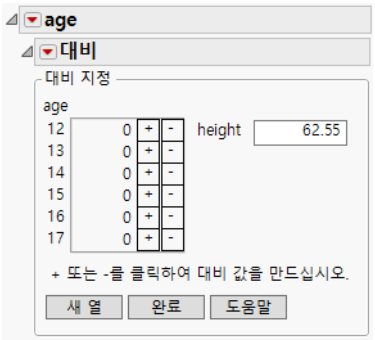
다음 단계에 따라 그림 3.14 에 표시된 보고서를 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. weight 를 선택하고 **Y** 를 클릭합니다 .
4. age, sex 및 height 를 선택하고 **추가**를 클릭합니다 .

5. " 열 선택 " 목록에서 age 를 선택하고 " 모형 효과 생성 " 목록에서 height 를 선택한 후 **교차**를 클릭합니다 .
6. **실행**을 클릭합니다 .

   " 최소 제곱 적합 " 보고서가 나타납니다 .
7. "age" 의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 대비 ...** 를 선택합니다 .

그림 3.14 age 에 대한 최소 제곱 평균 대비 지정

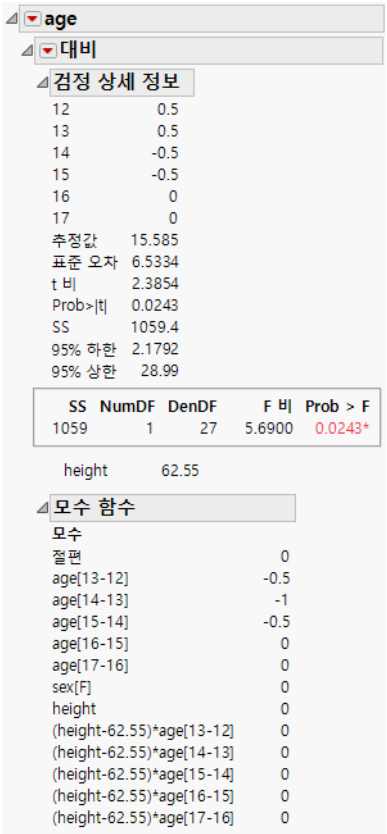


8. age 12 와 13 에 대해 "+" 를 클릭합니다 .
9. age 14 와 15 에 대해 "-" 를 클릭합니다 .

   이 대비는 키 (height) 62.55 에서 예측값을 기반으로 두 연령 (age) 그룹의 평균 체중 (weight) 이 다른지 여부를 검정합니다 .
10. 연속형 효과 height 옆에 텍스트 상자가 있습니다 . 기본값은 연속형 효과의 평균입니다 .
11. **완료**를 클릭합니다 .
12. " 검정 상세 정보 " 및 " 모수 함수 " 보고서를 엽니다 .

그림 3.15에 "대비" 보고서가 표시되어 있습니다. 대비 검정은 0.05 수준에서 유의합니다. 평균 키 (height) 62.55에서 12세 및 13세 (age) 어린이의 예측 체중 (weight) 이 14세 및 15세 (age) 어린이의 예측 체중 (weight) 과 통계적으로 다르다는 결론을 내립니다.

그림 3.15 최소 제곱 평균 대비 보고서



age

대비

검정 상세 정보

12	0.5
13	0.5
14	-0.5
15	-0.5
16	0
17	0
추정값	15.585
표준 오차	6.5334
t 비	2.3854
Prob>|t|	0.0243
SS	1059.4
95% 하한	2.1792
95% 상한	28.99

SS	NumDF	DenDF	F 비	Prob > F
1059	1	27	5.6900	0.0243*

height 62.55

모수 함수

모수	
절편	0
age[13-12]	-0.5
age[14-13]	-1
age[15-14]	-0.5
age[16-15]	0
age[17-16]	0
sex[F]	0
height	0
(height-62.55)*age[13-12]	0
(height-62.55)*age[14-13]	0
(height-62.55)*age[15-14]	0
(height-62.55)*age[16-15]	0
(height-62.55)*age[17-16]	0

최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD

" 최소 제곱 평균 스튜던트 t" 및 " 최소 제곱 평균 Tukey HSD"(Honestly Significant Difference) 옵션은 모형 효과의 쌍별 비교를 검정합니다 .

- 최소 제곱 평균 스튜던트 t 옵션은 일반적인 독립 표본 , 등분산 *t*- 검정을 기반으로 합니다 . 각 비교는 지정된 유의 수준을 기반으로 합니다 . 여러 비교를 수행하여 구한 전체 오차율이 지정된 유의 수준을 초과합니다 .
- 최소 제곱 평균 Tukey HSD 옵션은 Tukey HSD 검정을 수행합니다 . 이러한 비교를 위한 유의 수준은 전체 쌍별 비교 모음에 적용됩니다 . 따라서 최소 제곱 평균 Tukey HSD 의 신뢰 구간이 최소 제곱 평균 스튜던트 t 의 신뢰 구간보다 넓습니다 . 유의 수준은 표본 크기가 같으면 정확하고 표본 크기가 다르면 약간 작습니다 (Hayter 1984).

그림 3.16에서는 Big Class.jmp 샘플 데이터 테이블의 age 효과에 대한 최소 제곱 평균 Tukey 보고서를 보여 줍니다. 이 보고서를 생성하려면 **Fit Model** 데이터 테이블 스크립트를 실행하고 "age"의 빨간색 삼각형을 클릭한 후 **최소 제곱 평균 Tukey HSD**를 선택합니다. 기본적으로 이 보고서에는 교차표 보고서와 연결 문자 보고서가 표시됩니다.

그림 3.16 최소 제곱 평균 Tukey HSD 보고서

최소 제곱 평균 차이 Tukey HSD

α= 0.050 Q= 3.02917

최소 제곱 평균[j]

최소 제곱 평균[i]	평균[i]-평균[j] 차이 / 표준 오차 / 차이 CL 하한 / 차이 CL 상한	12	13	14	15	16	17
12		0	13.6856	25.8473	19.7698	10.159	-2.5203
		0	6.96659	7.24776	8.04937	9.96733	10.5955
		0	-7.4174	3.89261	-4.613	-20.034	-34.616
		0	34.7886	47.8019	44.1527	40.3517	29.5754
13		-13.686	0	12.1617	6.08426	-3.5266	-16.206
		6.96659	0	6.82786	7.59798	9.69566	10.0859
		-34.789	0	-8.521	-16.931	-32.896	-46.758
		7.41737	0	32.8444	29.0998	25.8432	14.346
14		-25.847	-12.162	0	-6.0774	-15.688	-28.368
		7.24776	6.82786	0	6.28014	8.57237	8.65355
		-47.802	-32.844	0	-25.101	-41.655	-54.581
		-3.8926	8.52105	0	12.9462	10.2789	-2.1545
15		-19.77	-6.0843	6.07741	0	-9.6108	-22.29
		8.04937	7.59798	6.28014	0	9.23282	9.21068
		-44.153	-29.1	-12.946	0	-37.579	-50.191
		4.61302	16.9313	25.101	0	18.3569	5.61058
16		-10.159	3.52657	15.6882	9.61083	0	-12.679
		9.96733	9.69566	8.57237	9.23282	0	10.8916
		-40.352	-25.843	-10.279	-18.357	0	-45.672
		20.0337	32.8963	41.6554	37.5786	0	20.3131
17		2.52026	16.2058	28.3675	22.2901	12.6793	0
		10.5955	10.0859	8.65355	9.21068	10.8916	0
		-29.575	-14.346	2.15448	-5.6106	-20.313	0
		34.6159	46.7577	54.5806	50.1908	45.6716	0

수준		최소 제곱 평균
17	A	121.90696
12	A	119.38671
16	A B	109.22769
13	A B	105.70112
15	A B	99.61686
14	B	93.53945

교차표 보고서

두 옵션 모두 교차표 보고서라는 행렬을 표시합니다. 각 셀에는 평균 차이, 차이 표준 오차, 신뢰하한 및 상한이 포함됩니다. 유의 수준 및 해당 임계값이 행렬 위에 표시됩니다. 기본 유의 수준은 0.05이지만 "모형 적합" 시작 창에서 다른 유의 수준을 지정할 수 있습니다. 통계적으로 차이가 있는 평균 쌍에 해당하는 셀은 빨간색으로 표시됩니다.

연결 문자 보고서

기본적으로 연결 문자 보고서는 교차표 행렬 아래에 나타납니다. 동일한 문자를 공유하거나 이 문자로 연결되는 수준은 통계적으로 차이가 없습니다. 공통 문자로 연결되지 않는 수준은 통계적으로 차이가 있습니다.

그림 3.16에서 17, 12, 16, 13 및 15 수준은 문자 A로 연결됩니다. 연결은 이러한 수준이 0.05 유의 수준에서 차이가 없음을 나타냅니다. 또한 16, 13, 15 및 14 수준은 문자 B로 연결되어 해당 수준이 통계적으로 차이가 없음을 나타냅니다. 그러나 age 17과 14 및 age 12와 14는 공통 문자로 연결되지 않으므로 이 두 쌍의 수준이 통계적으로 차이가 있음을 나타냅니다.

팁 : 연결 문자 보고서를 마우스 오른쪽 버튼으로 클릭하고 " 열 " 을 선택하여 연결 문자 (" 문자 "), 표준 오차 (" 표준 오차 ") 및 신뢰 구간 한계 ("X% 하한 ", "X% 상한 ") 가 포함된 열을 추가할 수 있습니다 . " 문자 " 열에서 연결 문자는 단일 열로 연결됩니다 . 유의 수준과 신뢰 수준은 " 모형 적합 " 시작 창에서 " 유의 수준 설정 " 옵션을 사용하여 지정한 유의 수준에 의해 결정됩니다 .

최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD 옵션

각 보고서 창에 나타나는 빨간색 삼각형 옵션은 선택적 보고서를 표시하거나 숨깁니다 . 최소 제곱 평균 스튜던트 t 에는 아래 옵션을 모두 사용할 수 있습니다 . 최소 제곱 평균 Tukey HSD 에는 처음 네 개 옵션을 사용할 수 있습니다 . 최소 제곱 평균 스튜던트 t 와 최소 제곱 평균 Tukey HSD 둘 다 교차표 보고서 및 연결 문자 보고서가 기본적으로 표시됩니다 .

교차표 보고서 각 수준 쌍에 대해 평균 차이 , 차이 표준 오차 및 차이의 신뢰 한계를 제공하는 이원 테이블을 표시합니다. 유의한 차이를 포함하는 셀의 내용이 빨간색으로 강조 표시됩니다.

연결 문자 보고서 유의한 비교와 유의하지 않은 비교를 연결 문자로 나타냅니다 . 동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다. 동일한 문자로 연결된 수준에는 유의한 차이가 없습니다 .

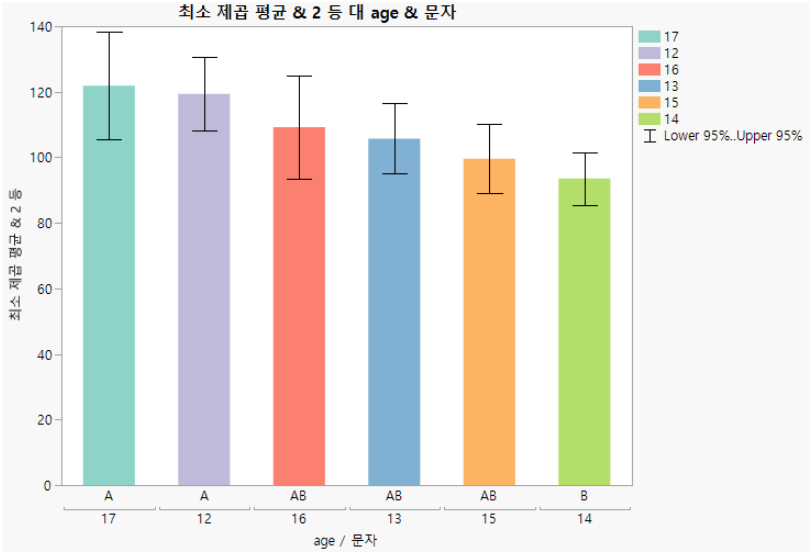
연결 문자 테이블 저장 효과 수준 , 연결 문자 , 최소 제곱 평균 , 표준 오차 및 신뢰 구간을 제공하는 열이 포함된 데이터 테이블을 생성합니다 . 이 테이블에는 신뢰 구간이 위에 겹쳐진 최소 제곱 평균의 막대 차트 (색상으로 구분됨) 를 생성하는 " 막대 차트 " 스크립트가 포함되어 있습니다 . 수준은 최소 제곱 평균이 큰 것부터 순서대로 배열됩니다 .

그림 3.17 에서는 Big Class.jmp 예의 막대 차트를 보여 줍니다 . **Fit Model** 데이터 테이블 스크립트를 실행하고 "age" 의 빨간색 삼각형을 클릭한 후 **최소 제곱 평균 Tukey HSD** 를 선택합니다 . " 최소 제곱 평균 차이 Tukey HSD" 보고서에서 **연결 문자 테이블 저장**을 선택합니다 . 데이터 테이블이 나타나면 **막대 차트** 스크립트를 실행합니다 .

정렬된 차이 보고서 차이를 가장 큰 것부터 순서대로 정렬하고 표준 오차, 신뢰 한계 및 p 값을 제공합니다 . 또한 신뢰 구간이 중첩된 막대 차트에 차이를 표시합니다 .

상세 비교 각 비교에 대한 개별 상세 보고서를 제공합니다 . 이 보고서에는 지정된 비교에 대한 추정 차이 , 표준 오차 , 신뢰 구간 , t 비 , 자유도 및 단측 / 양측 검정의 p 값이 표시됩니다 . 또한 비교에 대한 유의성 검정을 보여 주는 t 분포 그림도 표시됩니다 . 음영 부분의 영역은 양측 검정에 대한 p 값입니다 .

그림 3.17 최소 제곱 평균 차이 HSD 연결 문자 테이블의 막대 차트



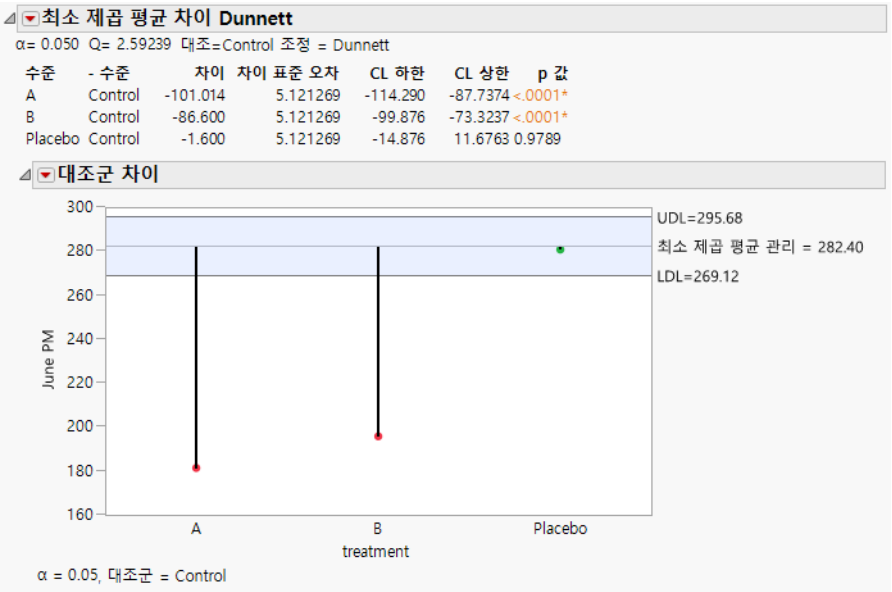
최소 제곱 평균 Dunnett

Dunnett 검정(Dunnett 1955)에서는 일련의 평균을 대조군 평균과 비교합니다. 오차율은 쌍별 비교 모음에 적용됩니다. 최소 제곱 평균 Dunnett 옵션은 지정된 효과의 수준에 대해 Dunnett 검정을 수행합니다. *p* 값 및 신뢰 구간 계산에 Hsu 요인 분석 근사가 사용됩니다(Hsu 1992).

"최소 제곱 평균 Dunnett"을 선택하면 효과의 대조 수준을 입력하는 대화상자가 나타납니다. "최소 제곱 평균 차이 Hsu-Dunnett" 보고서에는 유의 수준, 검정 통계량(Q) 값 및 대조 수준이 표시됩니다.

그림 3.18에서는 Cholesterol.jmp 샘플 데이터 테이블의 treatment 효과에 대해 "최소 제곱 평균 Dunnett" 옵션을 사용하여 나타나는 보고서를 보여 줍니다. 여기서 반응은 June PM이고, Control이라는 treatment 수준이 대조 수준으로 지정되었습니다.

그림 3.18 최소 제곱 평균 Dunnett 보고서



이 보고서에는 다음 두 가지 옵션이 있습니다 .

대조군 차이 보고서 " 대조군 차이 " 보고서는 기본적으로 표시됩니다 . 효과의 각 수준에 대해 대조 수준과 비교되는 수준 , 추정 차이 , 차이 표준 오차 , 신뢰 구간 및 비교를 위한 *p* 값 등의 정보가 테이블에 표시됩니다 .

대조군 차이 차트 대조 수준을 제외한 각 수준에서 점은 해당 수준의 최소 제곱 평균과 대조 수준의 최소 제곱 평균 간 차이를 표시합니다 . 결정 상한 및 하한 (UDL, LDL) 이 그림에 표시됩니다 . 이 보고서에는 " 요약 보고서 표시 " 옵션과 " 표시 옵션 " 이 있습니다 . " 요약 보고서 표시 " 옵션은 그림 상세 정보를 제공합니다 . " 표시 옵션 " 을 사용하면 그림 모양을 수정할 수 있습니다 .

조각 검정

" 조각 검정 " 옵션은 명목형 또는 순서형 열로 구성된 교호작용 효과에 대해 활성화됩니다. 이 옵션은 교호작용에 포함된 각 명목형 또는 순서형 열의 각 수준에 대해 해당 수준과 관련된 설정의 전체 쌍별 비교를 공동으로 검정하는 보고서를 생성합니다. 검정은 사실상 교호작용의 지정된 " 조각 " 내의 차이를 검정하는 것입니다.

A 의 수준 중 하나가 "Small" 인 경우 A*B*C 교호작용에 관심이 있다고 가정해 보겠습니다. A = Small 조각에 대한 " 조각 검정 " 보고서에서는 A = Small 일 때 B*C 수준의 전체 쌍별 비교를 공동으로 검정합니다. 이렇게 하면 교호작용 내에서 수준의 차이를 감지할 수 있습니다.

" 조각 검정 " 보고서는 " 최소 제곱 평균 대비 " 보고서와 동일한 형식을 따릅니다. 자세한 내용은 " 최소 제곱 평균 대비 ..."(95 페이지) 에서 확인하십시오.

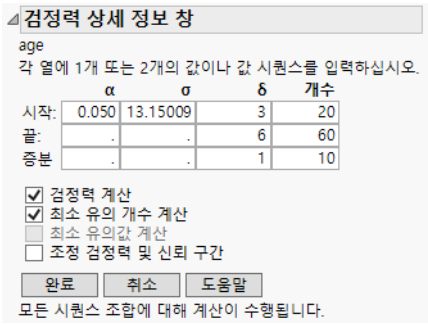
검정력 분석

특정 효과의 *F*-검정에 대한 후향적 또는 전향적 상세 정보를 얻기 위해 정보를 입력할 수 있는 "검정력 상세 정보 창"을 엽니다.

참고: 필요한 차이를 감지할 수 있도록 충분한 수의 관측값이 연구에 포함되게 하려면 실험을 **설계**할 때 검정력에 대한 정보를 사용하십시오. 이러한 분석을 **전향적** 검정력 분석이라고 합니다. DOE 플랫폼을 사용하여 연구를 설계하는 것이 좋습니다. DOE > 표본 크기 및 검정력과 DOE > 설계 평가 둘 다 전향적 검정력 분석에 유용합니다. 표준 최소 제곱을 사용한 전향적 검정력 분석의 예는 " 전향적 검정력 분석 "(198 페이지) 에서 확인하십시오.

그림 3.19에서는 Big Class.jmp 샘플 데이터 테이블에 대한 "검정력 상세 정보 창"의 예를 보여 줍니다. "검정력 상세 정보 창"을 사용하면 알파(α), 시그마(σ), 델타(δ) 및 개수(연구 크기) 값에 대한 검정력을 탐색할 수 있습니다. 단일 값("시작"만), 두 개의 값("시작", "끝") 또는 시작("시작"), 끝("끝") 및 값 시퀀스의 증분("증분")을 입력합니다. 지정한 값의 가능한 모든 조합에 대해 검정력 계산이 보고됩니다.

그림 3.19 검정력 상세 정보 창



자세한 내용은 "검정력 분석"(193 페이지)에서 확인하십시오.

" 검정력 상세 정보 창 " 보고서에는 다음 열과 옵션이 포함됩니다.

알파 (α) 검정의 유의 수준입니다. 이 값은 0 에서 1 사이이며 대개 0.05, 0.01 또는 0.10 입니다. " 모형 적합 " 시작 창에서 " 유의 수준 설정 " 을 선택한 후 다른 값을 설정한 경우를 제외하고 첫 번째 행에 표시된 알파의 초기값은 0.05 입니다.

시그마 (σ) 모형의 잔차 오차에 대한 추정값입니다. 지침을 위해 첫 번째 행에 표시된 초기값은 RMSE(평균 제곱 오차의 제곱근) 입니다.

델타 (δ) 관심 효과 크기입니다. 자세한 내용은 " 효과 크기 "(193 페이지) 에서 확인하십시오. 첫 번째 행에 표시된 초기값은 가설에 대한 제곱합의 제곱근을 연구에 사용된 관측값 수의 제곱근으로 나눈 값 (즉 , $\delta = \sqrt{SS/n}$) 입니다.

개수 (n) 표본 크기입니다. 첫 번째 행에 표시된 초기값은 현재 연구의 관측값 수입니다.

검정력 계산 검정력을 α, σ, δ 및 *n* 의 함수로 계산합니다 . 검정력은 지정된 σ 및 *n* 에 대해 α 수준에서 유의한 검정 결과를 확인하여 δ 크기의 차이를 감지할 확률입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 검정력 계산 "(517 페이지) 에서 확인하십시오 .

최소 유의 개수 계산 지정된 δ 및 σ 에 대해 α 수준에서 유의한 검정 결과를 얻기 위해 필요한 최소 관측값 수를 계산합니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 "LSN 계산 "(516 페이지) 에서 확인하십시오 .

최소 유의값 계산 *p* 값 α 를 생성하는 모수의 선형 함수 또는 모수의 최소 양수 값을 계산합니다 . 최소 유의값은 α, σ 및 *n* 의 함수입니다 . 이 옵션은 1 자유도 검정에만 사용할 수 있습니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 "LSV 계산 "(516 페이지) 에서 확인하십시오 .

조정 검정력 및 신뢰 구간 후향적 검정력 계산은 *F* 분포의 비중심성 모수를 추정할 때 표준 오차 및 검정 모수의 추정값을 사용합니다 . **조정 검정력**은 양의 편향이 제거된 비중심성 모수의 추정값에 기반한 후향적 검정력 계산입니다 (Wright and O'Brien 1988).

조정 검정력의 신뢰 구간은 비중심성 추정값의 신뢰 구간을 기반으로 합니다 .

조정 검정력은 표본 추정값을 처리하므로 검정력 및 신뢰 한계는 현재 연구에서 추정된 δ 에 대해서만 계산됩니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 조정 검정력 계산 "(518 페이지) 에서 확인하십시오 .

적합 결여

"적합 결여" 보고서에는 모형이 데이터를 잘 적합시키는지 여부를 평가하는 검정의 상세 정보를 제공합니다. "적합 결여" 보고서는 이 검정을 수행할 수 있을 때만 나타납니다. 이 검정은 모형에 독립적인 추정값을 사용하여 반응의 분산을 추정할 수 있는지가 관건입니다. 이 추정값을 생성하려면 모형 효과의 반복된 값에서 반응 값을 사용할 수 있어야 합니다. 검정에는 이러한 반복된 관측값을 사용하여 제곱합을 기반으로 **순수 오차** 추정값을 계산하는 과정이 포함됩니다.

다음 상황에서는 검정 통계량을 계산할 수 없으므로 " 적합 결여 " 보고서가 나타나지 않습니다 .

- X 변수에 대해 반복된 점이 없어 순수 오차 제곱합을 계산할 수 없는 경우
- 모형이 **포화**되어 추정 모수가 관측값 수만큼 있는 경우. 이러한 모형은 완벽하게 적합되므로 적합 결여를 평가할 수 없습니다 .

모형의 오차 제곱합과 순수 오차 제곱합 사이의 차이를 **적합 결여** 제곱합이라고 합니다. 모형이 적절하지 않은 경우 적합 결여 변동이 순수 오차 변동보다 매우 클 수 있습니다. 예를 들어 예측 변수의 함수 형태가 잘못되었거나, 모형의 교호작용이 충분하지 않거나 올바르지 않을 수 있습니다.

" 적합 결여 " 보고서에는 다음 열이 포함됩니다 .

소스 변동의 소스로는 적합 결여 , 순수 오차 및 총 오차의 세 가지가 있습니다 .

DF 각 오차 소스의 DF(자유도) 입니다 .

– 총 오차의 DF 는 분산 분석 테이블의 오차 행에 나오는 DF 값과 같습니다 . 총 오차 DF 는 제곱합 분해를 기반으로 적합 결여의 자유도와 순수 오차의 자유도로 분할됩니다 .

– 순수 오차 DF 는 반복된 각 관측값 그룹에서 풀링된 것입니다 . 일반적으로 g 개의 그룹이 있고 각 그룹이 각 효과에 대해 동일한 설정을 갖는 경우 순수 오차 DF, 즉 DF_{PE} 는 다음과 같이 정의됩니다 .

$$DF_{PE} = \sum_{i=1}^{g} (n_i - 1)$$

여기서 n_i 는 i 번째 그룹의 반복 실험 수입니다 .

– 적합 결여 DF 는 총 오차 DF 와 순수 오차 DF 사이의 차이입니다 .

제곱합 각 오차 소스에 대한 관련 SS(제곱합) 입니다 .

– 총 오차 SS 는 해당 분산 분석 테이블의 오차 행 (참조) 에 나오는 제곱합입니다 .

– 순수 오차 SS 는 반복된 각 관측값 그룹에 대한 제곱합 값의 총계입니다 . 순수 오차 SS 를 DF 로 나눈 값은 지정된 예측 변수 설정에서 반응의 분산을 추정합니다 . 이 추정값은 모형의 영향을 받지 않습니다 . 일반적으로 g 개의 그룹이 있고 각 그룹이 각 효과에 대해 동일한 설정을 갖는 경우 순수 오차 SS, 즉 SS_{PE} 는 다음과 같이 정의됩니다 .

$$SS_{PE} = \sum_{i=1}^{g} SS_i$$

여기서 SS_i 는 관측된 각 반응과 i 번째 그룹의 평균 반응 간의 차이에 대한 제곱합입니다 .

– 적합 결여 SS 는 총 오차 제곱합과 순수 오차 제곱합 사이의 차이입니다 .

평균 제곱 소스의 평균 제곱 , 즉 제곱합을 DF 로 나눈 값입니다 . 순수 오차 평균 제곱에 비해 적합 결여 평균 제곱이 크면 모형이 잘 적합되지 않음을 나타냅니다 . F 비는 공식 검정을 제공합니다 .

F 비 적합 결여의 평균 제곱 대비 순수 오차의 평균 제곱에 대한 비율입니다 . F 비는 적합 결여 및 순수 오차 평균 제곱에 의해 추정된 분산이 동일하다는 가설 , 즉 " 적합 결여 없음 " 을 나타내는 것으로 해석되는 가설을 검정합니다 .

Prob > F 적합 결여 검정의 p 값입니다 . p 값이 작으면 적합 결여가 유의함을 나타냅니다 .

최대 R^2 이러한 효과만 기반으로 한 모형에서 얻을 수 있는 최대 R^2 입니다 . 순수 오차 제곱합은 모형 형태에 따라 변동되지 않습니다 . 따라서 이러한 반복된 효과가 있는 모형에서 설명할 수 있는 최대 변동은 다음과 같습니다 .

$$\frac{SS(\text{수정 합계}) - SS(\text{순수 오차})}{SS(\text{수정 합계})} = 1 - \frac{SS(\text{순수 오차})}{SS(\text{수정 합계})}$$

이 계산식은 최대 R^2 을 정의합니다 .

추정값

"추정값" 메뉴에서는 모형 모수에 대한 추가 상세 정보를 제공합니다. JMP에서 명목형 및 순서형 효과가 코드화되는 방식을 검토하면 추정값을 이해하는 데 도움이 됩니다. 자세한 내용은 "사용

자 검정 예에 대한 상세 정보"(188 페이지), "통계 상세 정보" 부록의 "명목형 요인"(493 페이지) 및 "통계 상세 정보" 부록의 "순서형 요인"(503 페이지)에서 확인하십시오.

모형에 임의 효과가 포함된 경우 다음 중 적절한 옵션만 "추정값" 메뉴에서 사용할 수 있습니다.

"추정값" 메뉴에서는 다음 옵션을 제공합니다.

예측 표현식 표시 추정된 모형에 대한 방정식을 포함하는 "예측 표현식" 보고서를 표시하거나 숨깁니다. 예는 "예측 표현식 표시"(107 페이지)에서 확인하십시오.

정렬된 추정값 선별 상황에서 유용할 수 있는 "정렬된 모수 추정값" 보고서를 표시하거나 숨깁니다. 설계가 포화되지 않은 경우 이 보고서는 절편을 제외한 항이 유의성이 큰 것부터 순서대로 정렬된 "모수 추정값" 보고서입니다. 설계가 포화된 경우에는 유사 *t*- 검정이 제공됩니다. 자세한 내용은 "정렬된 추정값"(108 페이지)에서 확인하십시오.

확장 추정값 (하나 이상의 효과가 연속형이 아닌 경우에만 사용 가능) 명목형 효과의 모든 수준에 대한 모수 추정값을 제공하여 "모수 추정값" 보고서를 확장하는 "확장 추정값" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "확장 추정값"(111 페이지)에서 확인하십시오.

표시자 파라미터화 추정값 (모형 효과 중 명목형 열이 있는 경우에만 사용 가능) 전통적 표시 함수를 사용하여 모형의 명목형 효과가 파라미터화된 모수 추정값을 포함하는 "표시 함수 파라미터화" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "표시자 파라미터화 추정값"(113 페이지)에서 확인하십시오.

순차 검정 모형에 효과가 순차적으로 추가될 때 제곱합이 포함된 "순차 (유형 1) 검정" 보고서를 표시하거나 숨깁니다. 순차 제곱합을 기반으로 *F*- 검정을 수행합니다. 자세한 내용은 "순차 검정"(114 페이지)에서 확인하십시오.

사용자 검정 사용자 가설을 검정할 수 있습니다. 자세한 내용은 "사용자 검정"(115 페이지)에서 확인하십시오.

다중 비교 효과 수준 간의 비교를 지정할 수 있습니다. 이러한 비교에는 단일 효과가 포함되거나, 유연한 사용자 비교를 정의할 수 있습니다. 전체 평균 및 대조군 평균과 비교하거나, Tukey HSD 또는 스튜던트 *t* 를 사용하여 전체 쌍별 비교를 수행할 수 있습니다. 스튜던트 *t* 방법을 지정하는 경우 동등성 검정도 수행하여 실제적으로 중요한 쌍별 차이를 식별할 수 있습니다. 자세한 내용은 "다중 비교"(117 페이지)에서 확인하십시오.

기울기 비교 (명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있는 경우에만 사용 가능) ANCOVA(공분산 분석) 모형에서 교호작용 효과의 각 수준에 대한 기울기를 비교할 수 있는 보고서를 생성합니다. 자세한 내용은 "기울기 비교"(130 페이지)에서 확인하십시오.

결합 요인 검정 (모형에 교호작용이 포함된 경우에만 사용 가능) 모형의 각 주효과에 대해 해당 주효과와 관련된 모든 모수의 결합 검정을 표시하거나 숨깁니다. 자세한 내용은 "결합 요인 검정"(130 페이지)에서 확인하십시오.

역추정 예측 하나 이상의 반응 값에 대해 설명 변수 값을 예측할 수 있습니다. 자세한 내용은 "역추정 예측"(131 페이지)에서 확인하십시오.

Cox 혼합물 (모형에 혼합물 효과가 포함된 경우에만 사용 가능) Cox 혼합물 모형에 대한 모수 추정값을 생성합니다. 이를 사용하여 요인 효과를 도출하고 설계 공간의 참조점을 기준으로

반응 표면 모양을 추정할 수 있습니다 . 자세한 내용은 "Cox 혼합물 "(135 페이지) 에서 확인하십시오 .

모수 검정력 해당 가설 검정과 관련된 검정력 및 기타 상세 정보를 제공하는 열을 " 모수 추정값 " 보고서에 추가합니다 . 자세한 내용은 " 모수 검정력 "(136 페이지) 에서 확인하십시오 .

추정값 상관계수 모형의 모든 모수 추정값에 대한 상관 행렬을 표시하거나 숨깁니다 . 자세한 내용은 " 추정값 상관계수 "(138 페이지) 에서 확인하십시오 .

오차 규격 (임의 효과가 없는 경우에만 사용 가능) " 최소 제곱 적합 " 보고서에서 표준 오차 및 검정에 사용되는 오차 분산과 오차 자유도를 지정합니다 . " 오차 규격 " 을 변경해도 " 스튜던트화 잔차 " 그림과 "Box-Cox 변환 " 보고서는 영향을 받지 않습니다 . " 오차 규격 " 이 " 순수 오차 " 또는 " 지정됨 " 인 경우 " 분산 분석 " 보고서에 추가 열이 나타납니다 . 자세한 내용은 " 분산 분석 "(86 페이지) 에서 확인하십시오 .

기본 추정값 모형의 표준 제곱근 평균 제곱 오차와 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다 .

순수 오차 " 적합 결여 " 보고서의 순수 오차 평균 제곱과 관련 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다 . 자세한 내용은 " 적합 결여 "(104 페이지) 에서 확인하십시오 .

주의 : 순수 오차 자유도가 1 이면 검정력이 약하고 신뢰 한계가 크다는 경고 메시지가 표시됩니다 .

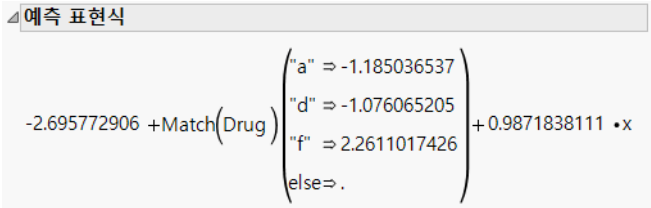
지정됨 사용자가 지정한 오차 분산 값과 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다 .

예측 표현식 표시

"예측 표현식 표시" 옵션은 반응을 예측하는 데 사용되는 방정식을 표시합니다. 그림 3.20에서는 Drug.jmp 샘플 데이터 테이블의 예를 보여 줍니다. 이 표현식은 일반적인 JMP 계산식으로 제공됩니다. 예를 들어 x = 10인 Drug a에 대한 반응을 예측하려면 반올림을 사용하여 –2.696 - 1.185 + 0.987(10) = 5.99를 계산합니다.

팁 : 예측 계산식에 자릿수를 지정하려면 **파일 > 환경 설정 > 테이블**로 이동하여 **기본 필드 너비** 값을 변경합니다 .

그림 3.20 예측 표현식



위의 보고서를 생성하려면 다음 단계를 수행하십시오 .

예측 표현식의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Drug 와 x 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 y" 의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다 . 그러면 그림 3.20 의 보고서가 나타납니다 .

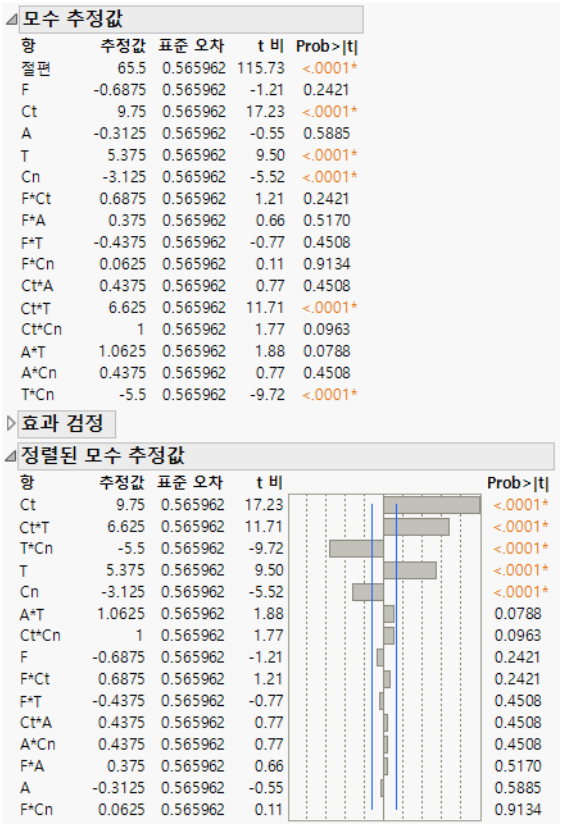
정렬된 추정값

"정렬된 추정값" 옵션은 선별 상황에서 유용한 "모수 추정값" 보고서를 생성합니다. 설계가 포화되지 않은 경우 "정렬된 추정값" 보고서에는 "모수 추정값" 보고서에서 절편을 제외한 항이 유의성이 큰 것부터 순서대로 정렬된 정보가 제공됩니다(그림 3.21의 두 번째 보고서). 설계가 포화된 경우에는 유사 *t*-검정이 제공됩니다. 이 검정은 Lenth **유사 표준 오차**를 기반으로 합니다 (Lenth 1989). 자세한 내용은 "Lenth PSE"(110 페이지)에서 확인하십시오.

정렬된 추정값 보고서의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. 창 아래쪽 근처의 **차수** 상자에 2 가 표시되어 있는지 확인합니다 .
5. F, Ct, A, T 및 Cn 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **추정값 > 정렬된 추정값**을 선택합니다 .

그림 3.21 정렬된 모수 추정값



모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	65.5	0.565962	115.73	<.0001*
F	-0.6875	0.565962	-1.21	0.2421
Ct	9.75	0.565962	17.23	<.0001*
A	-0.3125	0.565962	-0.55	0.5885
T	5.375	0.565962	9.50	<.0001*
Cn	-3.125	0.565962	-5.52	<.0001*
F*Ct	0.6875	0.565962	1.21	0.2421
F*A	0.375	0.565962	0.66	0.5170
F*T	-0.4375	0.565962	-0.77	0.4508
F*Cn	0.0625	0.565962	0.11	0.9134
Ct*A	0.4375	0.565962	0.77	0.4508
Ct*T	6.625	0.565962	11.71	<.0001*
Ct*Cn	1	0.565962	1.77	0.0963
A*T	1.0625	0.565962	1.88	0.0788
A*Cn	0.4375	0.565962	0.77	0.4508
T*Cn	-5.5	0.565962	-9.72	<.0001*

효과 검정

정렬된 모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
Ct	9.75	0.565962	17.23	<.0001*
Ct*T	6.625	0.565962	11.71	<.0001*
T*Cn	-5.5	0.565962	-9.72	<.0001*
T	5.375	0.565962	9.50	<.0001*
Cn	-3.125	0.565962	-5.52	<.0001*
A*T	1.0625	0.565962	1.88	0.0788
Ct*Cn	1	0.565962	1.77	0.0963
F	-0.6875	0.565962	-1.21	0.2421
F*Ct	0.6875	0.565962	1.21	0.2421
F*T	-0.4375	0.565962	-0.77	0.4508
Ct*A	0.4375	0.565962	0.77	0.4508
A*Cn	0.4375	0.565962	0.77	0.4508
F*A	0.375	0.565962	0.66	0.5170
A	-0.3125	0.565962	-0.55	0.5885
F*Cn	0.0625	0.565962	0.11	0.9134

"강조"가 "효과 선별"로 설정되고 모든 효과에 모수가 하나만 있는 경우 "정렬된 모수 추정값" 보고서도 자동으로 나타납니다.

"모수 추정값" 보고서와 "정렬된 모수 추정값" 보고서 (둘 다 그림 3.21 에 표시됨) 에는 다음과 같은 차이가 있습니다 .

- "정렬된 모수 추정값" 보고서에는 절편이 표시되지 않습니다 .
- 효과가 *t* 비의 절대값을 기준으로 정렬되며 가장 유의한 효과가 맨 위에 표시됩니다 .
- 막대 차트에는 0.05 유의 수준에 대한 임계값을 나타내는 수직선과 함께 *t* 비가 표시됩니다 .

포화 모형의 정렬된 추정값 보고서

선별 실험에는 자유도가 충분하지 않아 오차를 추정할 수 없는 완전 포화 모형이 포함되는 경우가 많습니다. 이러한 경우 "정렬된 추정값" 보고서(그림 3.21)에서는 상대 표준 오차를 제공하고, Lenth PSE(유사 표준 오차)를 사용하여 *t* 비와 *p* 값을 생성합니다. 이러한 값은 이름에 **유사**라는 라벨이 지정됩니다. 자세한 내용은 "Lenth PSE"(110 페이지) 및 "유사 t 비"(110 페이지)에서 확인하십시오. 노트에 변경 사항이 설명되고 PSE가 표시됩니다.

보고서에는 다음 열이 포함됩니다.

항 관심 계수가 있는 모형 항입니다.

추정값 *p* 값이 작은 것부터 순서대로 정렬된 모수 추정값이 제공됩니다.

상대 표준 오차 잔차 오차에 대한 자유도가 없는 경우 보고서에 상대 표준 오차가 제공됩니다. 상대 표준 오차는 제곱근 평균 제곱 오차를 1 로 설정하여 계산됩니다.

유사 t 비 유사 표준 오차를 사용하여 계산된 추정값의 *t* 비입니다. Lenth PSE 값이 보고서 하단에 표시됩니다.

유사 p 값 *m*/3 의 DFE(오차 자유도) 값을 사용하여 계산된 *p* 값입니다. 여기서 *m* 은 절편을 제외한 모수의 수입니다. DFE 값이 보고서 하단에 표시됩니다.

Lenth PSE

Lenth PSE(유사 표준 오차)는 Lenth(1989)로 인한 잔차 오차의 추정값입니다. 이는 선별 실험에서 비교적 적은 수의 효과가 활성화되는 효과 희소성 원리를 기반으로 합니다. 비활성 효과는 랜덤 잡음을 나타내고 Lenth 추정값의 기반을 만듭니다.

값은 다음과 같이 계산됩니다.

1. 모든 비절편 모수의 절대값을 고려합니다.
2. 절대값이 3.75 x 추정값 절대값의 중앙값을 초과하는 모든 모수 추정값을 제거합니다.
3. 모수 추정값에 대한 나머지 절대값의 중앙값에 1.5 를 곱합니다.

유사 t 비

상대 표준 오차가 같으면 Lenth PSE 가 보고서 하단에 표시됩니다. 유사 *t* 비는 다음과 같이 계산됩니다.

$$\text{유사 } t\text{비율} = \frac{\text{추정값}}{\text{PSE}}$$

상대 표준 오차가 같지 않으면 T 척도 Lenth PSE 가 계산됩니다. 이 값은 추정값의 PSE 를 상대 표준 오차로 나눈 값입니다. 유사 *t* 비는 다음과 같이 계산됩니다.

$$\text{유사 } t\text{비율} = \frac{\text{추정값}}{\text{T척도Lenth PSE} \times \text{상대 표준 오차}}$$

지정된 추정값의 표준 오차를 추정하기 위해 T 척도 Lenth PSE 는 추정값의 상대 표준 오차를 곱하여 조정됩니다.

포화 모형의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Y 를 선택하고 **Y** 를 클릭합니다.

4. 5 개 열 (F, Ct, A, T 및 Cn) 을 선택합니다 .
5. **매크로** 버튼을 클릭하고 **완전 요인**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **추정값 > 정렬된 추정값**을 선택합니다 .

   사용된 Lenth PSE 및 자유도가 보고서 하단에 제공됩니다 . 보고서에 따르면 유사 p 값을 기반으로 Ct, Ct*T, T*Cn, T 및 Cn 효과가 매우 유의합니다 .

그림 3.22 포화 모형의 정렬된 모수 추정값 보고서

정렬된 모수 추정값

항	추정값	상대 표준 오차	유사 t 비	유사 p 값
Ct	9.75	0.176777	14.86	<.0001*
Ct*T	6.625	0.176777	10.10	<.0001*
T*Cn	-5.5	0.176777	-8.38	<.0001*
T	5.375	0.176777	8.19	<.0001*
Cn	-3.125	0.176777	-4.76	0.0007*
F*A*Cn	-1.25	0.176777	-1.90	0.0850
A*T	1.0625	0.176777	1.62	0.1355
Ct*Cn	1	0.176777	1.52	0.1576
F*Ct*Cn	-0.9375	0.176777	-1.43	0.1826
F*Ct*A	0.75	0.176777	1.14	0.2789
F*Ct*A*Cn	0.75	0.176777	1.14	0.2789
F	-0.6875	0.176777	-1.05	0.3187
F*Ct	0.6875	0.176777	1.05	0.3187
F*Ct*T	0.6875	0.176777	1.05	0.3187
Ct*A*T	0.5625	0.176777	0.86	0.4108
F*A*T*Cn	0.5	0.176777	0.76	0.4632
Ct*A	0.4375	0.176777	0.67	0.5196
F*T	-0.4375	0.176777	-0.67	0.5196
A*Cn	0.4375	0.176777	0.67	0.5196
F*A	0.375	0.176777	0.57	0.5799
F*A*T	-0.375	0.176777	-0.57	0.5799
A	-0.3125	0.176777	-0.48	0.6438
F*T*Cn	0.3125	0.176777	0.48	0.6438
F*Ct*T*Cn	0.3125	0.176777	0.48	0.6438
Ct*A*T*Cn	-0.3125	0.176777	-0.48	0.6438
F*Ct*A*T*Cn	-0.25	0.176777	-0.38	0.7110
Ct*T*Cn	-0.125	0.176777	-0.19	0.8526
F*Cn	0.0625	0.176777	0.10	0.9259
Ct*A*Cn	0.0625	0.176777	0.10	0.9259
A*T*Cn	0.0625	0.176777	0.10	0.9259
F*Ct*A*T	0	0.176777	0.00	1.0000

오차 자유도가 없으므로 일반적인 검정을 수행할 수 없습니다.
상대 표준 오차는 잔차 표준 오차 1에 해당합니다. Lenth PSE = 0.65625 및 DFE=10.333을(를) 사용하여 유사 t 비 및 p 값이 계산되었습니다.

확장 추정값

모수 추정값을 처리하는 경우 JMP에서 명목형 및 순서형 열이 코드화되는 방식을 이해해야 합니다. 명목형 열을 코드화하는 방법에 대한 자세한 내용은 "사용자 검정 예에 대한 상세 정보"(188 페이지)에서 확인하십시오. 순서형 열을 코드화하고 모델링하는 방법에 대한 자세한 내용은 "통계 상세 정보" 부록의 "명목형 요인"(493 페이지) 및 "통계 상세 정보" 부록의 "순서형 요인"(503 페이지)에서 확인하십시오.

모형에 명목형 항이 있고 전체 추정값 집합에 대한 상세 정보를 확인하려는 경우 "확장 추정값" 옵션을 사용합니다. "확장 추정값" 옵션은 추정값, 표준 오차, *t* 비 및 *p* 값을 제공합니다.

확장 추정값 보고서의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Drug 와 x 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 y" 의 빨간색 삼각형을 클릭하고 **추정값 > 확장 추정값**을 선택합니다 .

그림 3.23에 "확장 추정값" 보고서와 "모수 추정값" 보고서가 표시되어 있습니다. Drug[f] 항의 추정값은 "확장 추정값" 보고서에 나타납니다. 검정의 귀무가설은 Drug f 그룹의 평균이 전체 평균과 다르지 않다는 것입니다. Drug[f]에 대한 검정이 0.05 수준에서 유의하므로 Drug f 그룹의 평균 반응이 전체 반응과 차이가 있음을 나타냅니다. 자세한 내용은 "확장 추정값에 대한 검정 해석"(112 페이지)에서 확인하십시오.

그림 3.23 모수 추정값과 확장 추정값의 비교

반응 y

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	-2.695773	1.911085	-1.41	0.1702
Drug[a]	-1.185037	1.060822	-1.12	0.2742
Drug[d]	-1.076065	1.041298	-1.03	0.3109
x	0.9871838	0.164498	6.00	<.0001*

확장 추정값

명목형 요인이 모든 수준으로 확장되었습니다.

항	추정값	표준 오차	t 비	Prob>|t|
절편	-2.695773	1.911085	-1.41	0.1702
Drug[a]	-1.185037	1.060822	-1.12	0.2742
Drug[d]	-1.076065	1.041298	-1.03	0.3109
Drug[f]	2.2611017	1.093974	2.07	0.0488*
x	0.9871838	0.164498	6.00	<.0001*

확장 추정값에 대한 검정 해석

모형이 *n* 개 수준을 가진 단일 명목형 요인으로 구성되어 있다고 가정해 보겠습니다. 해당 요인은 *n-1* 개 수준에 대해 각각 하나씩 *n-1* 개 표시 변수로 나타냅니다. 이러한 *n-1* 개 표시 변수 중 하나에 해당하는 모수 추정값은 해당 수준에 대한 평균 반응과 모든 수준에 대한 평균 반응 간의 차이입니다. JMP에서 명목형 변수를 코드화하는 방식 때문에 이렇게 표현됩니다("사용자 검정 예에 대한 상세 정보"(188 페이지) 참조). 모수 추정값은 대개 해당 수준의 **효과**로 해석됩니다.

Cholesterol.jmp 샘플 데이터 테이블에서 단일 요인 treatment 와 반응 June PM 을 예로 들어보겠습니다. 항 또는 표시 변수인 treatment[A]와 관련된 모수 추정값은 treatment A에 대한 June PM 평균과 June PM 전체 평균 간의 차이입니다.

명목형 변수의 **모든** 수준에 대한 효과는 합이 0이 되어야 합니다. 수준 순서에서 마지막 수준, 즉 –1s로 코드화된 수준의 효과를 가정해 보겠습니다. 이 수준의 효과는 다른 *n-1* 개 수준에 대한 효과 합의 반대 부호 값입니다. 따라서 마지막 수준의 효과는 다른 *n-1* 개 수준에 대한 모수 추정값 합의 반대 부호 값입니다.

"추정값" 메뉴의 "확장 추정값" 옵션은 결측 추정값을 계산하고 명목형 열을 포함하는 모든 효과를 검정한 후 텍스트 보고서에 표시합니다. 이러한 효과의 모든 수준에서 추정값의 평균 또는 합이 0인지 확인할 수 있습니다. 특히 이 관계는 이러한 추정값 및 관련 검정이 서로 독립적이지 않음을 나타냅니다.

그림 3.23에 표시된 Drug.jmp 보고서에서 Drug와 관련된 항의 추정값은 공변량 x를 포함하는 모형을 기반으로 합니다.

참고 :

- Drug[a] 에 대한 추정값은 Drug a 에 대한 최소 제곱 평균과 y 의 전체 평균 간의 차이입니다 .
- " 확장 추정값 " 보고서에 제공된 Drug[f] 에 대한 추정값은 Drug[a] 및 Drug[d] 에 대한 추정값 합의 반대 부호 값입니다 .
- " 확장 추정값 " 보고서에 제공된 Drug [f] 에 대한 *t*- 검정은 Drug f 그룹에 대한 반응이 전체 평균 반응과 다른지 여부를 검정합니다 .
- 명목형 요인이 고차 교호작용에 포함된 경우 " 확장 추정값 " 보고서가 길어질 수 있습니다 . 예를 들어 2 수준 명목형 요인의 5 원 교호작용은 단일 모수 추정값을 생성하지만 부호 변화까지 모두 동일한 $2^5 = 32$ 개의 확장된 효과를 가집니다 .

표시자 파라미터화 추정값

이 옵션은 명목형 열이 표시자(SAS GLM) 파라미터화를 사용하여 코드화되고 연속형으로 처리되는 모형에 대한 모수 추정값을 제공하는 "표시 함수 파라미터화" 보고서를 표시합니다. 순서형 열은 일반적인 JMP 코딩 체계를 사용하여 코드화됩니다. SAS GLM 및 JMP 코딩 체계에 대한 설명은 "통계 상세 정보" 부록의 "요인 모형"(492 페이지)에서 확인하십시오.

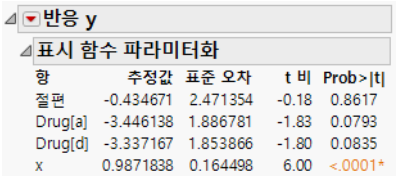
JMP 코딩 체계에서 명목형 변수 수준의 표시자에 해당하는 추정값은 같은 수준의 평균 반응과 모든 수준의 평균 반응 간 차이에 대한 추정값입니다. JMP 코딩을 보려면 "표준 최소 제곱"의 빨간색 삼각형 메뉴에서 **열 저장 > 코딩 테이블 저장**을 선택합니다.

표시자 코딩 체계에서 명목형 변수 수준의 표시자에 해당하는 추정값은 같은 수준의 평균 반응과 마지막 수준의 평균 반응 간 차이에 대한 추정값입니다. 마지막 수준은 값 순서 코딩이 가장 높은 수준이며 표시 함수가 모형에 포함되지 않은 수준입니다.

주의 : " 표시 함수 파라미터화 " 보고서에 제공된 표준 오차 및 t 비는 " 모수 추정값 " 보고서의 값과 다릅니다 . 이는 추정하는 모수가 서로 다르기 때문입니다 .

그림 3.24 의 보고서를 생성하려면 " 확장 추정값 보고서의 예 "(112 페이지) 의 단계를 수행하십시오 . 여기서 " 확장 추정값 " 을 선택하는 대신 **표시자 파라미터화 추정값**을 선택합니다 .

그림 3.24 표시자 파라미터화 추정값



반응 y

표시 함수 파라미터화

항	추정값	표준 오차	t 비	Prob>|t|
절편	-0.434671	2.471354	-0.18	0.8617
Drug[a]	-3.446138	1.886781	-1.83	0.0793
Drug[d]	-3.337167	1.853866	-1.80	0.0835
x	0.9871838	0.164498	6.00	<.0001*

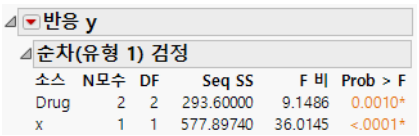
JMP 코딩 체계는 모형 적합의 일반화 회귀 분석법을 제외하고 JMP 전체에서 명목형 변수에 사용됩니다. 일반화 회귀 분석법의 명목형 변수에 사용되는 코딩 체계에 대한 자세한 내용은 "일반화 회귀 모형" 장의 "일반화 회귀 분석법 시작"(269 페이지)에서 확인하십시오.

JMP를 사용하는 경우와 SAS GLM 파라미터화를 사용하는 경우 각각 파생된 모형에 차이가 있을 수 있습니다. 일부 모형은 동등합니다. 다른 모형(예: 절편이 없는 모형, 결측 셀이 있는 모형, 명목형 또는 순서형 효과가 있는 모형, 혼합물 모형)은 차이를 나타낼 수 있습니다.

순차 검정

" 순차 검정 " 보고서에는 모형에 효과가 순차적으로 추가될 때 제곱합 및 검정이 표시됩니다 . 입력 순서는 " 모형 적합 " 시작 창의 " 모형 효과 생성 " 목록에 나타나는 효과의 순서로 정의됩니다 . 그림 3.25 에는 Drug.jmp 샘플 데이터 테이블에 대한 보고서가 표시되어 있습니다 .

그림 3.25 순차 검정 보고서



반응 y

순차(유형 1) 검정

소스	N모수	DF	Seq SS	F 비	Prob > F
Drug	2	2	293.60000	9.1486	0.0010*
x	1	1	577.89740	36.0145	<.0001*

순차 검정의 기반이 되는 제곱합을 **제1종 제곱합**이라고도 합니다. 이 값은 지정된 효과 입력 순서에 따라 모형을 적합시켜 계산됩니다. 특정 효과를 예로 들겠습니다. 해당 효과 **전에** 입력한 모든 효과를 포함하는 모형에 대한 모형 제곱합을 계산합니다. 그런 다음 이러한 효과와 지정된 효과를 **모두** 포함하는 모형에 대한 모형 제곱합을 계산합니다. 지정된 효과에 대한 순차 제곱합은 모형 제곱합의 증가입니다.

그림 3.25에서는 Drug.jmp 샘플 데이터 테이블에 대한 순차 제곱합을 보여 줍니다. "모형 적합" 시작 창에서 Drug를 먼저 입력한 후 x를 입력했습니다. Drug로만 구성된 모형의 제곱합은 293.6입니다. x가 모형에 추가되면 모형 제곱합이 871.4974가 됩니다. 증가한 값 577.8974가 x에 대한 순차 제곱합입니다.

"순차(유형 1) 검정" 보고서에 표시된 검정은 순차 제곱합에 기반한 *F*-검정이며 **제1종 검정**이라고도 합니다. *F* 비는 지정된 효과를 검정합니다. 여기서 모형은 해당 효과 및 위의 "소스" 열에 나열된 효과만 포함합니다.

순차 제곱합을 합하면 모형 제곱합이 됩니다. 또한 일반적인 모형을 가정했을 때 각 값이 통계적으로 독립적이라는 좋은 특징이 있습니다. 그러나 모형의 항 순서에 따라 값이 달라지므로 적절하지 않은 경우가 많습니다.

다음과 같은 경우 순차 검정이 적절합니다 .

- 적절한 순서로 지정된 분산 모형의 균형 분석 (즉 , 효과 목록에서 주효과 다음에 이원 교호작용이 나오는 순서)
- 적절한 순서로 지정된 순수 내포 모형
- 적절한 순서로 지정된 다항 회귀 모형

"모수 추정값" 및 "효과 검정" 보고서에 제공된 검정은 **제3종 제곱합**을 기반으로 합니다. 여기서 효과에 대한 제곱합은 다른 모든 효과가 모형에 입력된 후 해당 효과에 의해 설명되는 추가 제곱합입니다.

사용자 검정

모형 모수와 관련된 하나 이상의 사용자 가설을 검정하려면 "추정값" 메뉴에서 **사용자 검정**을 선택합니다. 이 창에서 모형 모수에 대한 하나 이상의 선형 함수 또는 **대비**를 지정할 수 있습니다.

결과에는 각 대비에 대한 개별 검정과 모든 대비에 대한 결합 검정이 포함됩니다. 자세한 내용은 그림 3.26에서 확인하십시오. 개별 대비 보고서에는 모수의 지정된 선형 함수에 대한 추정값과 표준 오차가 제공됩니다. *t* 비, *p* 값 및 관련 제곱합도 제공됩니다. 개별 대비 결과 아래에 나오는 모든 대비의 결합 검정에서는 제곱합, 분자 자유도, *F* 비 및 *p* 값을 제공합니다.

주의 : 이러한 검정은 잔차 오차를 사용하여 수행됩니다 . 모형에 임의 효과가 있고 REML 대신 EMS 를 사용할 경우에는 이러한 검정이 적절하지 않을 수 있습니다 .

참고 : 고차 효과에 포함된 효과를 검정하는 경우 사용자 검정 대신 최소 제곱 평균에 대한 검정을 사용하는 것이 좋습니다 . 최소 제곱 평균은 다른 모형 효과에 대해 조정됩니다 . " 효과 상세정보 " 에서 최소 제곱 평균 대비를 검정할 수 있습니다 .

사용자 검정 보고서 구성 요소

" 사용자 검정 " 규격 창에는 다음과 같은 구성 요소가 있습니다 .

편집 가능한 텍스트 상자 " 사용자 검정 " 제목 표시줄 아래의 공간은 검정 이름을 입력할 수 있는 편집 가능한 영역입니다 .

모수 모형 항을 나열합니다 . 항 목록의 오른쪽에는 해당 모수에 대한 0 열이 있습니다 . 이 셀에 값을 입력하여 검정을 위한 선형 함수를 지정합니다 .

"=" 기호 " 모수 " 목록의 마지막 행에는 = 라벨이 지정됩니다 . 이 셀에 상수를 입력하여 각 대비에 대한 규격을 완료합니다 .

열 추가 모수의 여러 선형 함수를 공동으로 검정할 수 있도록 0 열을 추가합니다 .

완료 " 완료 " 버튼을 클릭하여 검정을 수행합니다 . 그러면 보고서가 변경되어 각 검정 열에 대한 검정 통계량 값 , 표준 오차 및 기타 통계량이 표시됩니다 . 모든 열에 대한 결합 *F*- 검정은 보고서 하단의 상자에 제공됩니다 .

사용자 검정 보고서 옵션

" 사용자 검정 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

검정력 분석 결합 검정에 대한 검정력 분석을 제공합니다. 이 옵션은 검정이 수행된 후에만 사용할 수 있습니다 . 자세한 내용은 " 모수 검정력 "(136 페이지) 에서 확인하십시오 .

제거 " 사용자 검정 " 보고서를 제거합니다 .

참고 : 결합 사용자 검정을 여러 번 수행하려면 **추정값 > 사용자 검정**을 반복해서 선택하십시오 .

그림 3.26 에서는 Cholesterol.jmp 샘플 데이터 테이블을 사용하여 세 가지 대비가 포함된 규격 창의 예를 보여 줍니다. 세 가지 검정 모두 상수가 0 으로 설정됩니다. 이 검정에 대한 보고서가 그림 3.27 에 표시되어 있습니다.

사용자 검정의 예

Cholesterol.jmp 샘플 데이터 테이블은 6 개 기간에서 20 명의 환자를 대상으로 한 반복 측정을 제공합니다. 네 개의 treatment 그룹이 연구됩니다. 일반적으로 이 데이터는 모든 반복 측정을 반응으로 사용하여 적절하게 분석해야 합니다. 이 예에서는 June PM 에 대한 반응만 고려합니다.

세 개의 대비를 검정하려는 경우를 가정해 보겠습니다 . 다음에 대한 평균 반응을 비교하려고 합니다 .

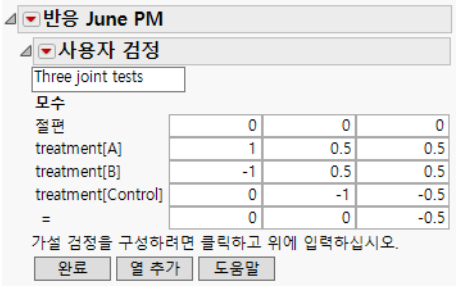
- treatment A 대 treatment B
- treatment A/B 대 대조군
- treatment A/B 대 대조군 / 플라시보 그룹

사용자 검정을 사용하여 이러한 대비를 검정하려면 다음을 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cholesterol.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. June PM 을 선택하고 **Y** 를 클릭합니다 .
4. treatment 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 June PM" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 사용자 검정**을 선택합니다 .
7. " 사용자 검정 " 규격 보고서에서 **열 추가**를 두 번 클릭하여 두 개의 추가 열을 생성합니다 .
8. 그림 3.26 에 표시된 것과 같이 편집 가능 영역에 검정 이름을 입력하고 세 열에 값을 입력합니다 .

이러한 값, 특히 세 번째 열의 값을 구하는 방법은 " 통계 상세 정보 " 부록의 " 모수 해석 "(493 페이지) 에서 확인하십시오 .

그림 3.26 세 가지 대비에 대한 사용자 검정 규격 보고서


반응 June PM
사용자 검정
Three joint tests
모수

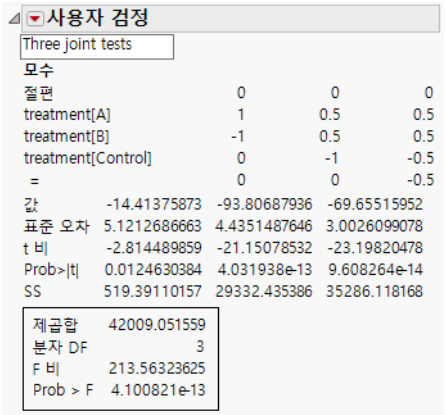
절편	0	0	0
treatment[A]	1	0.5	0.5
treatment[B]	-1	0.5	0.5
treatment[Control]	0	-1	-0.5
=	0	0	-0.5

가설 검정을 구성하려면 클릭하고 위에 입력하십시오.
완료 열 추가 도움말

9. **완료**를 클릭합니다 .

그림 3.27에 표시된 결과는 세 가설 모두 개별적으로 유의할 뿐만 아니라 공동으로도 유의하다는 것을 나타냅니다.

그림 3.27 세 가지 대비에 대한 검정을 보여 주는 사용자 검정 보고서


사용자 검정
Three joint tests
모수

절편	0	0	0
treatment[A]	1	0.5	0.5
treatment[B]	-1	0.5	0.5
treatment[Control]	0	-1	-0.5
=	0	0	-0.5
값	-14.41375873	-93.80687936	-69.65515952
표준 오차	5.1212686663	4.4351487646	3.0026099078
t 비	-2.814489859	-21.15078532	-23.19820478
Prob>|t|	0.0124630384	4.031938e-13	9.608264e-14
SS	519.39110157	29332.435386	35286.118168

제곱합	42009.051559
분자 DF	3
F 비	213.56323625
Prob > F	4.100821e-13

다중 비교

모형 효과 수준으로 정의된 평균을 비교하는 검정 및 신뢰 수준을 얻으려면 이 옵션을 사용합니다. 다중 비교 방법의 목표는 잘못된 결론에 도달할 확률을 제어하면서 그룹 평균이 다른지 여부를 확인하는 것입니다. "다중 비교" 옵션을 사용하면 그룹 평균을 전체 평균(평균 분석) 및 대조군 평균과 비교할 수 있습니다. Tukey HSD 또는 스튜던트 *t*를 사용하여 쌍별 비교를 수행할 수도 있습니다. 스튜던트 *t* 방법을 지정하는 경우 동등성 검정도 수행하여 실제적으로 중요한 쌍별 차이를 식별할 수 있습니다.

스튜던트 *t* 방법은 개별 비교에 대한 오차율만 제어합니다. 따라서 정확한 다중 비교 절차는 아닙니다. 제공된 다른 모든 방법은 관심 있는 모든 비교에 대한 전체 오차율을 제어합니다. 이러한 각 방법은 *p* 값 및 신뢰 한계를 계산할 때 다중 비교 **조정**을 사용합니다.

모형에 명목형 및 순서형 효과가 포함된 경우 최소 제곱 평균 추정값을 사용하여 비교를 수행하거나, 사용자 정의 추정값을 사용하여 특정 비교를 정의할 수 있습니다. 모형에 연속형 효과만 포함된 경우에는 사용자 정의 추정값을 사용하여 평균을 비교할 수 있습니다.

팁 : 연속형 효과가 비교적 적은 수준으로 구성되어 있다고 가정해 보겠습니다 . 최소 제곱 평균 추정값을 사용하여 비교하려는 경우 해당 효과에 순서형 또는 명목형 모델링 유형을 할당하는 것이 좋습니다 .

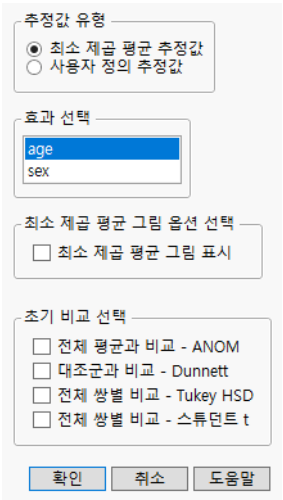
옵션 시작

그림 3.28에서는 "다중 비교" 옵션에 대한 컨트롤 창의 예를 보여 줍니다. 이 예는 Big Class.jmp 데이터 테이블을 기반으로 하며 weight를 Y로 지정하고 age, sex 및 height를 모형 효과로 지정합니다. 비교를 위해 최소 제곱 평균 추정값과 사용자 정의 추정값의 두 가지 추정값 유형을 사용할 수 있습니다.

최소 제곱 평균 추정값

이 옵션은 모형에 명목형 또는 순서형 효과가 있는 경우에만 사용할 수 있으며 최소 제곱 평균을 비교합니다 . 최소 제곱 평균은 모형에 있는 다른 효과의 일부 중립 값에서 계산된 평균입니다 . 최소 제곱 평균의 정의는 " 최소 제곱 평균 테이블 "(91 페이지) 에서 확인하십시오 . 관심 효과를 선택해야 합니다 . 그림 3.28 에는 age 에 대한 최소 제곱 평균 추정값이 지정되어 있습니다 . 최소 제곱 평균 그림을 표시하는 옵션이 있습니다 . 자세한 내용은 " 최소 제곱 평균 그림 옵션 "(93 페이지) 에서 확인하십시오 .

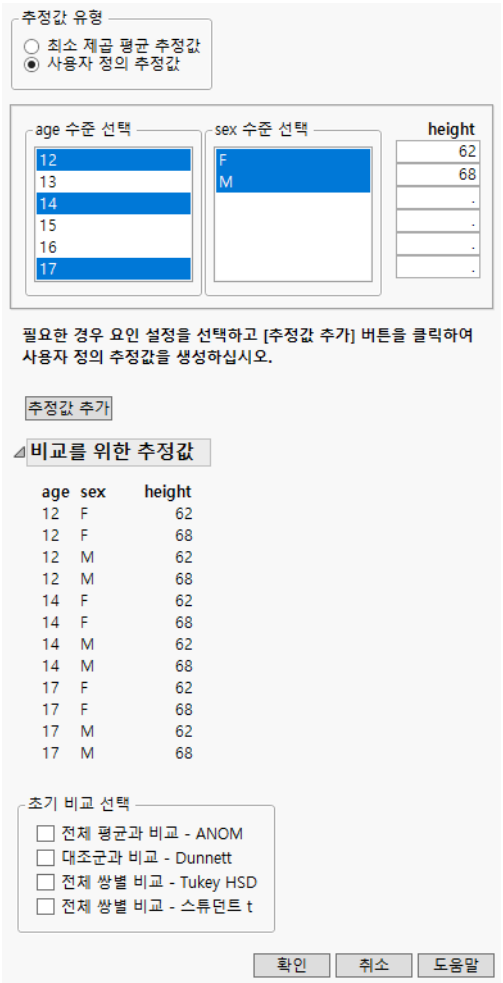
그림 3.28 최소 제곱 평균 추정값에 대한 시작 창



사용자 정의 추정값

그림 3.29에서는 사용자 정의 추정값 규격을 보여 줍니다. age의 세 가지 수준과 sex의 두 가지 수준 모두 선택되었습니다. 또한 height의 두 값이 수동으로 입력되었습니다. "추정값 추가" 버튼을 클릭하여 지정된 수준의 가능한 모든 조합이 나열되었습니다. 여기서 추정값을 더 지정하고 "추정값 추가" 버튼을 다시 클릭하여 "비교를 위한 추정값" 목록에 추가할 수 있습니다.

그림 3.29 사용자 정의 추정값에 대한 시작 창



"사용자 정의 추정값"을 사용하는 경우 다음과 같이 모델링 유형에 따라 수준이 지정되지 않은 효과가 설정됩니다.

- 연속형 효과는 효과의 평균으로 설정됩니다.
- 명목형 및 순서형 효과는 값 순서화에서 첫 번째 수준으로 설정됩니다.

참고: 이 섹션에서는 **평균**이라는 용어를 사용하여 최소 제곱 평균 추정값 또는 사용자 정의 추정값을 나타냅니다.

최소 제곱 평균 그림 옵션 선택

최소 제곱 평균 그림 표시를 선택하면 최소 제곱 평균 그림을 나타낼 수 있습니다. 효과가 교호작용 항인 경우 교호작용 그림을 생성할 수도 있습니다. 중첩을 위한 항을 선택합니다. 교호작용 그림을 선택하지 않으면 최소 제곱 그림에 효과 항이 내포됩니다. 자세한 내용은 "최소 제곱 평균 그림 옵션"(93 페이지)에서 확인하십시오.

초기 비교 선택

추정값을 지정한 후에는 " 초기 비교 선택 " 에서 옵션을 선택하여 초기 보고서에 표시할 비교 유형을 선택할 수 있습니다. 아니면 옵션을 선택하지 않고 " 확인 " 을 클릭합니다.

전체 평균과 비교 - ANOM 각 효과 최소 제곱 평균을 전체 최소 제곱 평균과 비교합니다. 평균 분석

대조군과 비교 - Dunnett 각 효과 최소 제곱 평균을 대조 수준의 최소 제곱 평균과 비교합니다.

전체 쌍별 비교 - Tukey HSD 중복도를 위한 Tukey HSD 조정을 사용하여 효과 최소 제곱 평균의 전체 쌍별 비교를 검정합니다.

전체 쌍별 비교 - 스튜던트 t 중복도 조정 없이 효과 최소 제곱 평균의 전체 쌍별 비교를 검정합니다.

이러한 각 옵션을 선택하면 맨 위 영역에 해당 보고서에 대한 상세 정보가 표시된 보고서가 열립니다. 이 정보에는 분위수 또는 임계값이 포함됩니다. 실제 다중 비교 절차의 경우 다중 비교 조정에 사용된 방법이 표시됩니다. "사용자 정의 추정값"을 지정한 경우 보고서에는 지정된 추정값에 따라 변하지 않는 효과 목록 및 이러한 효과가 설정된 수준이 표시됩니다. 별도로 지정하지 않으면 연속형 효과는 평균으로 설정됩니다. 명목형 또는 순서형 효과는 값 순서화에서 첫 번째 수준으로 설정됩니다.

"초기 비교 선택" 목록에서 선택하지 않고 "확인"을 클릭하면 "최소 제곱 평균 추정값" 테이블 또는 "사용자 정의 추정값" 테이블이 표시된 "다중 비교" 보고서가 열립니다. "다중 비교"의 빨간색 삼각형 메뉴에서 위에 나열된 모든 옵션을 사용할 수 있습니다. 사용 가능한 보고서 및 옵션은 아래에서 설명합니다.

최소 제곱 평균 또는 사용자 정의 추정값 보고서

기본적으로 " 다중 비교 " 옵션은 시작 창에서 선택한 추정값 유형에 따라 " 최소 제곱 평균 추정값 " 보고서 또는 " 사용자 정의 추정값 " 보고서를 표시합니다. 이 테이블에는 관심 수준의 각 조합에 대해 평균 추정값과 검정 및 신뢰 구간이 제공됩니다. 구체적으로 다음 항목이 이 테이블에 포함됩니다.

범주형 효과의 수준 보고서의 첫 번째 열은 하나 이상의 관심 효과를 식별합니다. 열의 값은 분석 대상 그룹을 지정합니다.

추정값 각 그룹에 대한 평균 추정값입니다.

표준 오차 각 그룹에 대한 평균의 표준 오차입니다.

DF 평균이 0 인지 여부를 검정하는 데 사용할 자유도입니다 .

95% 하한 평균의 신뢰 하한입니다 . " 모형 적합 " 창에서 " 유의 수준 설정 " 을 선택하여 신뢰 수준을 변경할 수 있습니다 .

95% 상한 평균의 신뢰 상한입니다 .

t 비 유의성 검정의 t 비입니다 . 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > t 비를 선택한 경우에만 나타납니다 .

Prob>|t| 유의성 검정의 p 값입니다 . 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > Prob>|t| 를 선택한 경우에만 나타납니다 .

산술평균 추정값 (" 최소 제곱 평균 추정값 " 보고서에만 표시됨) 각 그룹에 대한 산술평균의 추정값입니다 .

N (" 최소 제곱 평균 추정값 " 보고서에만 표시됨) 각 그룹에 대한 평균을 계산하는 데 사용되는 관측값 수입니다 .

참고 : 테이블을 마우스 오른쪽 버튼으로 클릭하고 " 열 " 을 선택하여 t 비 및 p 값을 구할 수 있습니다 .

전체 평균과 비교

이 옵션은 지정된 수준에 대한 평균을 해당 수준에 대한 전체 평균과 비교합니다. 전체 평균과의 차이에 대한 신뢰 구간을 보여 주는 테이블과 결정 한계를 보여 주는 차트가 표시됩니다. 비교에 사용된 방법을 ANOM(평균 분석) 이라고 합니다 (Nelson et al. 2005). ANOM 은 전체 평균에 대한 전체 쌍별 비교의 결합 오차율을 제어하는 다중 비교 절차입니다. Lipid Data.jmp 샘플 데이터 테이블에 기반한 보고서는 그림 3.30 에서 확인하십시오.

ANOM 은 분석 분석과 비슷해 보일 수 있습니다. 그러나 모든 수준에 대한 전체 평균과 다른 평균을 갖는 수준을 식별한다는 점에서 근본적으로 다릅니다. 반면에 분산 분석은 평균 자체의 차이를 검정합니다.

" 전체 평균과 비교 " 보고서 맨 위에는 다음 내용이 표시됩니다 .

분위수 결정 한계를 구성하는 데 사용된 Nelson h 통계량의 값입니다 .

조정 DF 결정 한계를 구성하는 데 사용된 자유도입니다 .

평균 평균의 평균입니다 . 최소 제곱 추정값의 경우 평균의 평균은 그룹 최소 제곱 평균의 가중 평균입니다 . 이 가중 평균은 그룹 최소 제곱 평균이 계산되는 중립 설정에서 전체 평균을 나타냅니다 .

명확히 말해 평균 최소 제곱 평균은 $L(X'X)^{-1}L'$ 행렬의 대각 항목에 반비례하는 가중치를 사용한 가중 평균입니다 . 여기서 L 은 그룹 최소 제곱 평균을 계산하는 데 사용되는 계수의 행렬입니다 . 최소 제곱 평균의 기술적 정의에 대한 내용은 SAS Institute Inc. (2020b).

사용자 정의 추정값에 대한 평균의 평균도 비슷하게 정의됩니다 . 그러나 이 경우 L 은 추정값을 정의하는 데 사용되는 계수의 행렬입니다 .

조정 다음 중 임계값을 구하는 데 사용되는 방법을 설명합니다 .

Nelson 정확 임계값과 p 값을 제공합니다 . 가능할 때마다 특히 추정값이 상관관계가 없을 때 사용됩니다 .

Nelson-Hsu 사용된 Hsu 요인 분석 근사 (Hsu 1992) 를 기반으로 근사 임계값과 p 값을 제공합니다 . 정확 값을 구할 수 없을 때 사용됩니다 .

Sidak Nelson 과 Nelson-Hsu 둘 다 실패하면 사용됩니다 .

기술 상세 정보는 SAS Institute Inc. (2020b).

"전체 평균과 비교" 보고서 메뉴에서 다음 세 가지 옵션을 사용할 수 있습니다.

전체 평균과 차이

이 보고서에는 그룹 평균과 전체 평균의 각 비교에 대해 다음과 같은 상세 정보가 제공됩니다 .

- 비교되는 수준
- 차이 - 추정된 차이
- 표준 오차 - 차이의 표준 오차
- 신뢰 구간의 하한 및 상한
- t 비 - " 차이 " 및 " 표준 오차 " 열의 비율

전체 평균과 비교 의사결정 차트

이 의사결정 차트에는 각 그룹의 평균에 점이 표시됩니다. 평균의 평균에 수평선을 그립니다. 결정 상한 및 하한이 그림에 표시됩니다. 그룹 평균에 해당하는 점이 이 한계를 벗어난다고 가정해 보겠습니다. 이는 지정된 유의 수준에서 평균 분석 검정을 기반으로 그룹 평균이 전체 평균과 다르다는 것을 나타냅니다. 유의 수준은 차트 아래에 표시됩니다.

" 전체 평균과 비교 의사결정 차트 " 보고서 메뉴에는 다음 옵션이 있습니다 .

요약 보고서 표시 각 그룹에 대한 추정값 , 결정 한계 및 한계 초과를 보여 주는 테이블을 생성합니다 .

표시 옵션 차트 표시를 제어하는 몇 가지 옵션을 제공합니다 .

조정 p 값 계산

p 값이 포함된 열 (Prob>|t|) 을 " 전체 평균과 비교 " 보고서에 추가합니다. 불균형 설계에 대한 정확 임계값 및 p 값을 계산하려면 복소 적분이 필요하며 계산이 어려울 수 있습니다. 이러한 분위수 계산이 실패하면 Sidak 분위수가 계산되지만 p 값은 사용할 수 없습니다.

전체 평균과 비교의 예

Lipid Data.jmp 샘플 데이터 테이블을 예로 들어보겠습니다 . 음주 (Alcohol use) 및 심장 기록 (Heart History) 을 제어하면서 네 가지 흡연 이력 (Smoking History) 범주 중 평균 하루 커피 섭취량 (Coffee intake (cups/day)) 이 전체 평균 커피 섭취량과 다르다는 점에서 특이한 것이 있는지 확

인하려고 합니다 . Coffee intake (cups/day) 를 반응으로 사용하고 Smoking History, Alcohol Use 및 Heart History 를 모형 효과로 사용하여 모형을 지정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Lipid Data.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Coffee intake (cups/day) 를 선택하고 **Y** 를 클릭합니다 .
4. Smoking History, Alcohol Use 및 Heart History 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 Coffee intake (cups/day)" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 다중 비교**를 선택합니다 .
7. " 효과 선택 " 목록에서 Smoking History 를 선택합니다 .
8. " 초기 비교 선택 " 목록에서 **전체 평균과 비교 - ANOM** 을 선택합니다 .
9. **확인**을 클릭합니다 .

그림 3.30에 표시된 결과는 비흡연자와 cigarette 흡연자의 최소 제곱 평균이 커피 섭취량 측면에서 전체 평균과 유의하게 다르다는 것을 나타냅니다.

그림 3.30 평가를 위한 전체 평균과 비교

반응 Coffee intake (cups/day)

Smoking History에 대한 다중 비교

최소 제곱 평균 추정값

전체 평균과 비교

분위수 = 2.50786, 조정 DF = 85.0, 평균 = 1.66719, 조정 = Nelson-Hsu

전체 평균과 차이

Smoking History	-Smoking History	차이	표준 오차	조정된 95% 하한	조정된 95% 상한	t 비
no	평균	-0.59761	0.165469	-1.01259	-0.18264	-3.61
quit	평균	0.85689	0.634351	-0.73397	2.44775	1.35
cigarettes	평균	2.79463	0.704828	1.02702	4.56223	3.96
pipes	평균	-1.18007	1.502993	-4.94936	2.58922	-0.79

전체 평균과 비교 의사결정 차트

대조군과 비교

"대조군과 비교 - Dunnett"을 선택하면 대조군을 지정할 수 있는 창이 열립니다. "최소 제곱 평균 추정값"을 선택한 경우 선택한 효과의 모든 수준으로 목록이 구성됩니다. "사용자 정의 추정값"을 선택한 경우에는 지정한 효과 수준의 조합으로 목록이 구성됩니다.

대조군을 선택한 후 "확인"을 클릭하면 "최소 제곱 적합" 보고서에 "대조군과 비교" 보고서가 나타납니다. 이 옵션은 지정된 설정에 대한 평균을 대조군 평균과 비교합니다. 대조군과의 차이에 대한 신뢰 구간을 보여 주는 테이블과 결정 한계를 보여 주는 차트가 표시됩니다. Dunnett 방법이 비교에 사용됩니다. Dunnett 방법은 모든 비교에 대해 오차율을 제어하는 다중 비교 절차입니다(Hsu 1996, Westfall et al. 2011).

p 값과 신뢰 구간을 정확하게 계산할 수 없는 경우 Hsu 요인 분석 근사가 사용됩니다(Hsu 1992). 불균형 설계에 대한 정확 임계값 및 p 값을 계산하려면 복소 적분이 필요하며 계산량이 많을 수 있습니다. 이러한 분위수 계산이 실패하면 Sidak 분위수가 계산됩니다.

지정된 추정값에 대해 변하지 않는 효과 목록 외에도 다음 내용이 " 대조군과 비교 " 보고서 맨 위에 표시됩니다 .

분위수 Dunnett 검정의 임계값입니다 .

조정 DF 신뢰 구간을 구성하는 데 사용된 자유도입니다 .

대조군 대조군을 정의하는 설정입니다 . 이 설정은 단일 효과를 선택한 경우 단일 수준이며 , 둘 이상의 효과에 대한 사용자 정의 조합을 지정한 경우에는 수준 조합입니다 .

조정 다음 중 임계값을 구하는 데 사용되는 방법입니다 .

Dunnett 정확 임계값과 p 값을 제공합니다 . 가능할 때마다 특히 추정값이 상관관계가 없을 때 사용됩니다 .

Dunnett-Hsu Hsu 요인 분석 근사 (Hsu 1992) 를 기반으로 근사 임계값과 p 값을 제공합니다 . 정확 값을 구할 수 없을 때 사용됩니다 .

Sidak Dunnett 과 Dunnett-Hsu 둘 다 실패하면 사용됩니다 .

기술 상세 정보는 SAS Institute Inc. (2020b).

"대조군과 비교" 보고서 메뉴에서 다음 세 가지 옵션을 사용할 수 있습니다.

대조군과 차이

이 보고서에는 그룹 평균과 대조군 평균의 각 비교에 대해 다음과 같은 상세 정보가 제공됩니다 .

- 비교되는 수준
- 차이 - 추정된 차이
- 표준 오차 - 차이의 표준 오차
- 신뢰 구간의 하한 및 상한
- t 비 - " 차이 " 및 " 표준 오차 " 열의 비율

대조군과 비교 의사결정 차트

이 의사결정 차트에는 대조군과 비교되는 각 그룹의 평균에 점이 표시됩니다. 수평선은 대조군의 평균을 나타냅니다. 결정 상한 및 하한이 그림에 표시됩니다. 이 한계를 벗어나는 점은 지정된 유의 수준에서 Dunnett 검정을 기반으로 평균이 대조군 평균과 다른 그룹에 해당합니다. 해당 수준은 차트 아래에 표시됩니다.

" 대조군과 비교 의사결정 차트 " 보고서 메뉴에는 다음 옵션이 있습니다 .

요약 보고서 표시 각 그룹에 대한 추정값 , 결정 한계 및 한계 초과를 보여 주는 테이블을 생성합니다 .

표시 옵션 차트 표시를 제어하는 몇 가지 옵션을 제공합니다 .

조정 p 값 계산

p 값이 포함된 열 (Prob>|t|) 을 "대조군과 비교" 보고서에 추가합니다. 불균형 설계에 대한 정확 임계값 및 p 값을 계산하려면 복소 적분이 필요하며 계산이 어려울 수 있습니다. 이러한 분위수 계산이 실패하면 Sidak 분위수가 계산되지만 p 값은 사용할 수 없습니다.

전체 쌍별 비교

" 전체 쌍별 비교 " 옵션은 "Tukey HSD 전체 쌍별 비교 " 또는 " 스튜던트 t 전체 쌍별 비교 " 보고서를 표시합니다 (Hsu 1996, Westfall et al. 2011). Tukey HSD 비교는 유의 수준이 전체 쌍별 비교에 공동으로 적용되도록 구성됩니다. 반면에 스튜던트 t 비교는 유의 수준이 각 개별 비교에 적용됩니다. 스튜던트 t 검정을 사용하여 여러 쌍별 비교를 수행할 때 비교 중 하나가 차이를 잘못 나타낼 위험이 명시된 유의 수준을 훨씬 초과할 수 있습니다.

"Tukey HSD 전체 쌍별 비교 " 보고서 맨 위에는 다음 내용이 표시됩니다 .

분위수 검정의 임계값입니다 . Tukey HSD 의 경우 분위수는 $q/(\sqrt{2})$ 입니다 . 여기서 q 는 스튜던트화 범위 통계량의 적절한 백분율입니다 .

조정 DF 신뢰 구간을 구성하는 데 사용된 자유도입니다 .

조정 다음 중 임계값을 구하는 데 사용되는 방법을 설명합니다 .

Tukey 정확 임계값과 p 값을 제공합니다 . 평균이 상관관계가 없고 분산이 같을 때 또는 분산 균형 설계일 때 사용됩니다 .

Tukey-Kramer 근사 임계값과 p 값을 제공합니다 . 정확 값을 구할 수 없을 때 사용됩니다 .

기술 상세 정보는 SAS Institute Inc. (2020b).

" 스튜던트 t 전체 쌍별 비교 " 보고서 맨 위에는 분위수 (t 검정에 대한 임계값) 및 DF(t 검정에 사용된 자유도) 가 표시됩니다 .

전체 쌍별 차이 보고서

Tukey HSD 와 스튜던트 *t* 는 둘 다 전체 수준 쌍을 비교합니다 . " 전체 쌍별 차이 " 보고서에는 각 쌍별 비교에 대해 다음과 같은 정보가 제공됩니다 .

- 비교되는 수준
- 차이 - 추정된 평균 차이
- 표준 오차 - 차이의 표준 오차
- t 비 - 차이가 0 인지 여부에 대한 검정의 *t* 비
- Prob > |t| - 검정의 *p* 값
- 평균 차이에 대한 신뢰 구간의 하한 및 상한

전체 쌍별 비교 산점도

이 그림은 **차이 그림**(diffogram) 또는 **평균-평균 산점도**라고도 하며, 전체 평균 쌍별 차이에 대한 신뢰 구간을 표시합니다. 예는 그림 3.32에서 확인하십시오. 색상은 유의한 차이를 나타냅니다.

이 그림에는 참조선이 대각선에서 위쪽으로 기울어진 선으로 표시됩니다. 이 선은 두 평균이 동일한 점을 나타냅니다. 각 선분은 쌍별 비교에 대한 신뢰 구간에 해당합니다. 선분에 표시된 점의 좌표는 해당 그룹에 대한 평균입니다. 이러한 점 중 하나를 커서로 가리키면 비교되는 그룹을 식별하고 추정된 차이를 보여 주는 툴팁이 표시됩니다. 선분이 대각선에서 선과 교차하는 경우 평균이 동일할 수 있고 비교가 유의하지 않습니다.

쌍별 비교 산점도에는 다음 옵션이 있습니다 .

참조선 표시 산점도의 점에 대한 참조 격자선을 표시합니다. 산점도에 점이 많이 있는 경우 권장되지 않습니다 . 점이 많은 경우에는 점을 커서로 가리켜 툴팁 라벨을 보는 것이 더 좋습니다 .

전체 쌍별 차이 연결 문자

유의한 비교와 유의하지 않은 비교를 연결 문자로 나타내는 보고서를 표시하려면 이 옵션을 사용합니다. 동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다. 동일한 문자로 연결된 수준에는 유의한 차이가 없습니다.

전체 쌍별 차이 연결 문자 테이블 저장

이 옵션은 효과 수준, 연결 문자, 최소 제곱 평균, 표준 오차 및 신뢰 구간을 포함하는 열이 있는 데이터 테이블을 생성합니다. 데이터 테이블에는 신뢰 구간이 위에 겹쳐진 최소 제곱 평균의 막대 차트(색상으로 구분됨)를 생성하는 "막대 차트" 스크립트가 포함되어 있습니다. 수준은 최소 제곱 평균이 큰 것부터 순서대로 배열됩니다.

동등성 검정

하나 이상의 동등성 검정을 수행하려면 이 옵션을 사용합니다. 동등성 검정은 관심 있는 **실제적** 차이를 감지하려는 경우에 유용합니다. 지정한 값보다 작은 차이는 실제적으로 동등한 것으로

간주되도록 그룹 평균에 대한 임계 차이를 지정해야 합니다. 즉, 두 그룹 평균의 차이가 이 값보다 작거나 같으면 동등하다고 간주할 수 있습니다.

이 값을 지정하면 "동등성 검정" 보고서가 나타납니다. 지정한 경계는 보고서 맨 위에 표시됩니다. 이 보고서는 동등성 검정을 제공하는 테이블과 이를 표시하는 산점도로 구성됩니다. 동등성 검정 및 신뢰 구간은 스튜던트 *t* 임계값을 기반으로 합니다.

참고 : 동등성 검정은 스튜던트 *t* 방법에만 사용할 수 있습니다 .

동등성 TOST 검정

TOST(Two One-Sided Tests: 두 개의 단측 검정) 방법은 평균에 실제적인 차이가 있는지 검정하는 데 사용됩니다(Schuirmann 1987). 실제 차이가 임계값을 초과한다는 귀무가설에 대해 두 개의 단측 합동 분산 *t*-검정이 구성됩니다. 두 검정이 모두 기각되면 평균의 차이가 통계적으로 두 임계값 중 어느 하나도 초과하지 않습니다. 따라서 그룹은 실제적으로 동등하다고 간주됩니다. 검정이 하나만 기각되거나 둘 다 기각되지 않으면 그룹이 실제적으로 동등하지 않을 수 있습니다.

" 동등성 TOST 검정 " 보고서에는 각 비교에 대해 다음과 같은 정보가 제공됩니다 .

- 차이 - 추정된 평균 차이
- 하한 t 비 , 상한 t 비 - 두 개의 단측 합동 분산 유의성 검정에 대한 하한 및 상한 *t* 비
- 하한 p 값 , 상한 p 값 - 하한 및 상한 *t* 비에 해당하는 *p* 값
- 최대 p 값 - 하한 및 상한 *p* 값의 최대값
- 평균 차이에 대한 1–2α 신뢰 구간의 하한 및 상한

참고 : 동등성 TOST 검정은 스튜던트 *t* 방법에만 사용할 수 있습니다 .

동등성 검정 산점도

이 산점도에서는 동등성 검정에 따라 실제적으로 동등한 평균과 동등하지 않은 평균을 나타내기 위해 색상을 사용합니다. 이 그림을 **차이 그림**(diffogram) 또는 평균-평균 산점도라고도 합니다.

이 그림에는 대각선 위의 실선 참조선 및 음영 참조 대역이 표시됩니다. 대역 너비는 실제 차이의 두 배입니다. 선분에 있는 점의 좌표는 해당 그룹에 대한 평균입니다. 각 선분이 쌍별 비교를 위한 1–2α 신뢰 구간에 해당하는 세 번째 축이 대각선에 내포되어 있습니다. 이러한 점 중 하나를 커서로 가리키면 비교되는 그룹 및 추정된 차이를 나타내는 툴팁이 표시됩니다. 선분이 대각선 대역 내에 완전히 포함되면 평균이 실제적으로 동등하다고 할 수 있습니다.

참고 : 동등성 검정 산점도는 스튜던트 *t* 방법에만 사용할 수 있습니다 .

동등성 검정 산점도에는 다음 옵션이 있습니다 .

참조선 표시 산점도의 점에 대한 참조선을 표시합니다 . 산점도에 점이 많이 있는 경우 권장되지 않습니다 . 점이 많은 경우에는 점을 커서로 가리켜 툴팁 라벨을 보는 것이 더 좋습니다 .

제거

이 옵션은 "스튜던트 t 전체 쌍별 비교" 보고서에서 "동등성 검정" 보고서를 제거합니다.

Tukey HSD 전체 쌍별 비교의 예

Lipid Data.jmp 샘플 데이터 테이블을 예로 들어보겠습니다. 두 연령 (25 세, 35 세) 과 평균 키를 대상으로 성별 및 비흡연자와 과거 흡연자 (Smoking History 는 각각 no 와 quit) 의 Cholesterol 차이에 관심이 있습니다.

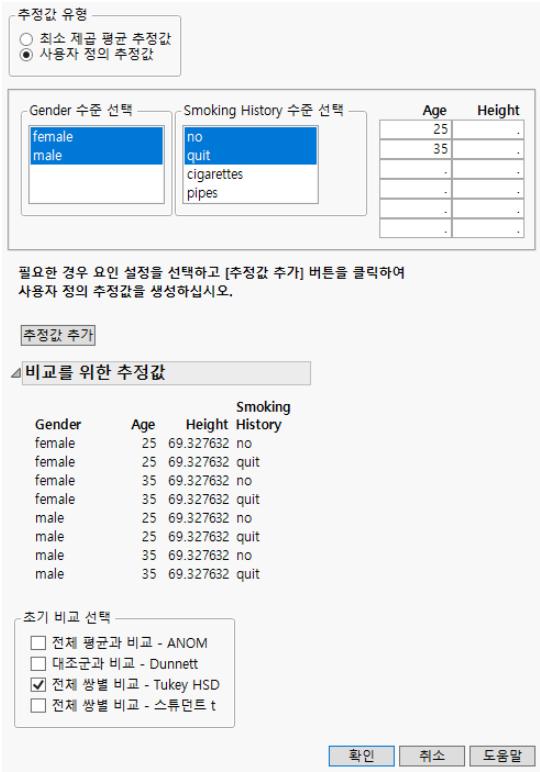
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Lipid Data.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Cholesterol 을 선택하고 **Y** 를 클릭합니다.
4. Gender, Age, Height 및 Smoking History 를 선택하고 **추가**를 클릭합니다.
5. **실행**을 클릭합니다.
6. " 반응 Cholesterol" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 다중 비교**를 선택합니다.
7. " 추정값 유형 " 목록에서 **사용자 정의 추정값**을 클릭합니다.
8. "Gender 수준 선택" 목록에서 female(기본적으로 선택되어 있어야 함)과 male을 선택합니다.
9. "Smoking History 수준 선택 " 목록에서 no 와 quit 를 선택합니다.
10. Age 목록에서 처음 두 행에 25 와 35 를 입력합니다.

    Height 목록에는 값을 입력하지 마십시오. Height 값이 지정되지 않았으므로 Height 열의 평균 값이 다중 비교 보고서에 사용됩니다.
11. **추정값 추가**를 클릭합니다.

    지정한 수준의 가능한 모든 조합이 " 비교를 위한 추정값 " 보고서에 나타납니다.
12. " 초기 비교 선택 " 목록에서 **전체 쌍별 비교 - Tukey HSD** 를 선택합니다.

    창이 그림 3.31 에 표시된 것과 같이 완료되었는지 확인합니다.

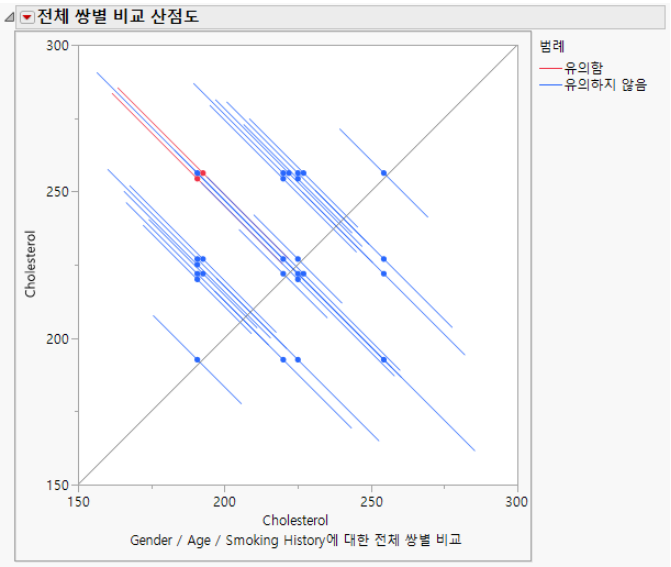
그림 3.31 항목이 채워진 사용자 정의 추정값 창



13. **확인**을 클릭합니다 .

" 전체 쌍별 차이 " 보고서에 따르면 28 개의 쌍별 비교 중 2 개가 유의합니다 . 그림 3.32 에 표시된 " 전체 쌍별 비교 산점도 " 에서는 이러한 비교에 대한 신뢰 구간이 빨간색으로 표시됩니다 . 점을 커서로 가리키면 해당 점이 나타내는 쌍별 비교를 확인할 수 있습니다 . 툴팁에는 비교 대상인 두 수준 간의 차이도 포함됩니다 . 그림 3.32 의 빨간색 점 두 개는 여성과 남성 둘 다에 대해 35 세 과거 흡연자와 25 세 비흡연자를 비교하는 점을 나타냅니다 .

그림 3.32 사용자 정의 비교를 위한 전체 쌍별 비교 산점도



기울기 비교

"기울기 비교" 옵션은 명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있을 때 나타납니다. 이 옵션은 ANCOVA(공분산 분석) 모형에서 기울기를 비교할 수 있는 보고서를 생성합니다. 이 보고서에서는 교호작용 효과의 각 수준에 대한 기울기를 전체 기울기와 비교합니다. 비교에는 ANOM(평균 분석)과 전체 평균이 함께 사용됩니다. ANOM(평균 분석) 보고서에 대한 자세한 내용은 "전체 평균과 비교"(121 페이지)에서 확인하십시오.

전체 평균 기울기는 기울기의 가중 평균이며 이때 가중치는 기울기 추정값의 분산에 반비례합니다. 이러한 분산은 "전체 평균 기울기와의 차이" 테이블에서 "표준 오차" 열의 제곱 값입니다.

결합 요인 검정

"결합 요인 검정" 옵션은 교호작용 효과가 있을 때 나타납니다. JMP는 모형의 각 주효과에 대해 해당 주효과를 포함하는 항의 모든 계수가 0인지 여부를 확인하는 결합 검정을 생성합니다. 이 검정은 모형에 있는 다른 모든 효과를 조건으로 합니다. 구체적으로 말해 결합 검정은 제한된 모형의 일반 선형 가설 검정입니다. 이 모형에서는 지정된 효과에 해당하는 모든 모수와 이 효과를 포함하는 교호작용이 0으로 설정됩니다.

결합 요인 검정 보고서의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. weight 를 선택하고 **Y** 를 클릭합니다 .

4. **차수** 상자에 2 가 표시되어 있는지 확인합니다 .
5. age, sex 및 height 를 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 weight" 의 빨간색 삼각형을 클릭하고 **추정값 > 결합 요인 검정**을 선택합니다 .

그림 3.33 결합 요인 검정 보고서

결합 요인 검정

항	DF	제곱합	F 비	Prob > F
age	15	6116.2127	2.8488	0.0139*
sex	7	2113.1080	2.1091	0.0879
height	7	9217.8156	9.2002	<.0001*

age에 대한 검정의 자유도는 15입니다. 이 검정에는 age에 대한 5개 모수, age*sex에 대한 5개 모수, height*age에 대한 5개 모수가 포함됩니다. 이 검정의 귀무가설은 15개 모수가 모두 0이라는 것입니다.

역추정 예측

역추정 예측은 반응 변수의 값이 주어진 경우 통계 모형을 사용하여 설명 변수의 값을 추론할 때 발생합니다. 역추정 예측은 **보정**이라고도 합니다.

"추정값" 메뉴에서 "역추정 예측"을 선택하면 지정된 반응 값에 해당하는 독립 변수 X의 값을 추정할 수 있습니다(그림 3.37). 또한 모형의 다른 설명 변수에 대한 값을 지정할 수 있습니다(그림 3.37). 역추정 예측 계산은 지정된 반응 값에 해당하는 X 값에 대한 신뢰 한계를 제공합니다. 반응 값을 평균 반응 또는 간단히 개별 반응으로 지정할 수 있습니다. 예는 "역추정 예측의 예"(132 페이지)에서 확인하십시오.

여러 설명 변수 분석

모형에 여러 설명 변수가 포함된 경우 다른 변수의 지정된 값에 대해 X 값을 예측할 수 있습니다. 심장 박동 수가 60일 때 산소 섭취량이 50이 되는 달리기 시간을 예측하고 싶을 수 있습니다. 남성과 여성 둘 다에 대해 별도의 역추정 예측을 원할 수도 있습니다. 역추정 예측 옵션을 사용하여 이러한 요구 사항을 지정합니다.

역추정 예측 창의 왼쪽에는 설명 변수 목록이 표시됩니다. 예는 그림 3.37에서 확인하십시오. 각 연속형 변수는 처음에 평균으로 설정됩니다. 각 명목형 또는 순서형 변수는 값 순서화에서 가장 낮은 수준으로 설정됩니다. 예측할 변수의 값을 제거하고 결측으로 설정해야 합니다. 또한 역추정 예측을 포함할 다른 변수의 값을 지정해야 합니다(기본 설정과 다른 경우). 창의 오른쪽에 있는 목록에서 관심 있는 하나 이상의 반응 값을 제공할 수 있습니다. 예는 "여러 모형 효과를 사용하여 단일 X 값 예측의 예"(134 페이지)에서 확인하십시오.

참고 : 역추정 예측의 신뢰 한계는 때때로 단측 구간이 되거나 무한 구간도 될 수 있습니다 . 기술 상세 정보는 " 통계 상세 정보 " 부록의 " 신뢰 한계를 사용한 역추정 예측 "(519 페이지) 에서 확인하십시오 .

역추정 예측의 예

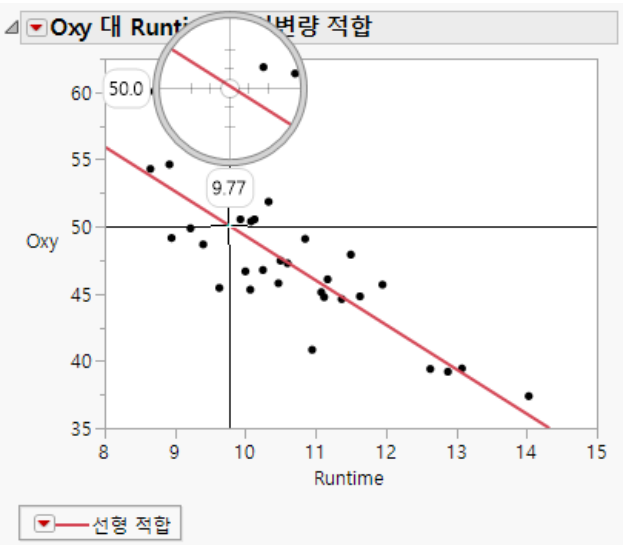
이 예에서는 Runtime 에서 산소 섭취량을 예측하는 회귀 모형을 적합시킵니다 . 그런 다음 지정된 산소 섭취량 값을 도출하는 Runtime 값을 추정합니다 . *X* 변수 (Runtime) 가 하나뿐이므로 X 로 Y 적합 플랫폼을 사용하여 역추정 예측값의 시각적 근사를 구하는 것으로 시작합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Oxy 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Runtime 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Oxy 대 Runtime 의 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 .

   아래 설명된 대로 십자기호 도구를 사용하여 평균 Oxy 값이 50 이 되는 Runtime 의 근사값을 산출합니다 .
7. **도구 > 십자기호**를 선택합니다 .
8. 예측선을 클릭한 후 십자기호 도구를 드래그하여 Oxy = 50 에 대한 역추정 예측을 찾습니다 .

그림 3.34에서는 Runtime이 약 9.77일 때 Oxy가 약 50이 되는 역추정 예측을 제공합니다.

그림 3.34 Fitness.jmp 에 대한 이변량 적합

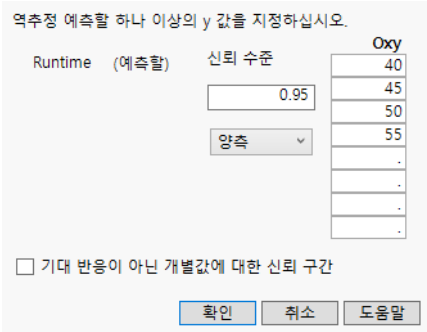


신뢰 구간과 함께 Runtime 에 대한 정확 예측을 구하려면 모형 적합 시작 창을 사용합니다 .

1. Fitness.jmp 샘플 데이터 테이블에서 **분석 > 모형 적합**을 선택합니다 .
2. Oxy 를 선택하고 **Y** 를 클릭합니다 .
3. Runtime 을 선택하고 **추가**를 클릭합니다 .

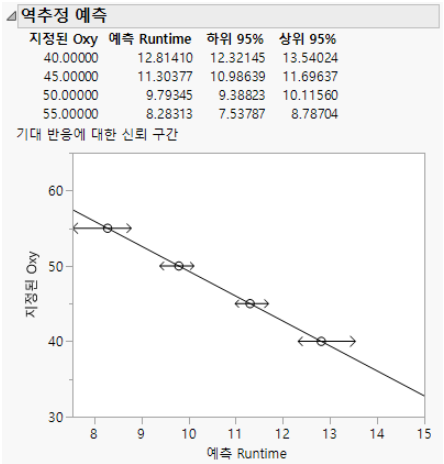
4. **실행**을 클릭합니다.
5. " 반응 Oxy" 의 빨간색 삼각형을 클릭하고 **추정값 > 역추정 예측**을 선택합니다.
6. 그림 3.35 에 표시된 대로 Oxy 값 4 개를 입력합니다.
7. **확인**을 클릭합니다.

그림 3.35 완료된 역추정 예측 규격 창



"역추정 예측" 보고서에는 지정된 각 Oxy 값에 해당하는 예측 Runtime 값이 제공됩니다. 또한 평균 반응을 얻는 것과 관련하여 Runtime 값에 대한 95% 신뢰 상한 및 하한도 보고서에 표시됩니다.

그림 3.36 역추정 예측 보고서



Oxy 값이 50 이 되는 정확 예측 Runtime 은 9.7935 입니다. 이 값은 그림 3.34 에 표시된 이변량 적합 보고서의 Runtime 근사값 9.77 에 가깝습니다. "역추정 예측" 보고서에는 Oxy 와 Runtime 간의 선형 관계 및 신뢰 구간을 보여 주는 그림도 제공됩니다.

여러 모형 효과를 사용하여 단일 X 값 예측의 예

이 예에서는 RstPulse 가 60 일 때 산소 섭취량이 50 이 되는 Runtime 을 예측합니다 . 남성과 여성 둘 다에 대해 Runtime 이 예측됩니다 .

1. Fitness.jmp 샘플 데이터 테이블에서 **분석 > 모형 적합**을 선택합니다 .
2. Oxy 를 선택하고 **Y** 를 클릭합니다 .
3. Sex, Runtime 및 RstPulse 를 선택한 후 **추가**를 선택합니다 .
4. **실행**을 클릭합니다 .
5. " 반응 Oxy" 의 빨간색 삼각형을 클릭하고 **추정값 > 역추정 예측**을 선택합니다 .
6. Runtime 값을 예측해야 하므로 해당 값을 삭제합니다 .
7. Sex 옆의 **모두**를 선택하여 Sex 의 모든 수준에 대한 Runtime 을 추정합니다 .
8. RstPulse 의 평균을 60 으로 바꿉니다 .
9. 그림 3.37 에 표시된 대로 Oxy 값 50 을 입력합니다 .
10. **확인**을 클릭합니다 .

그림 3.37 다중 회귀 모형에 대한 역추정 예측 규격

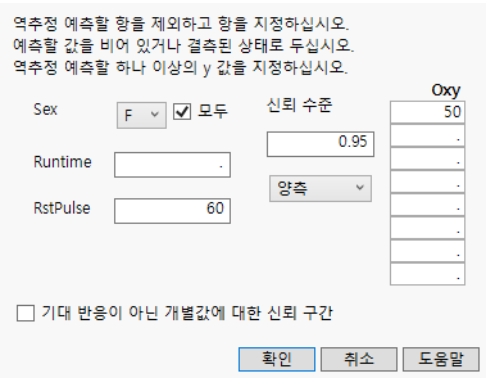
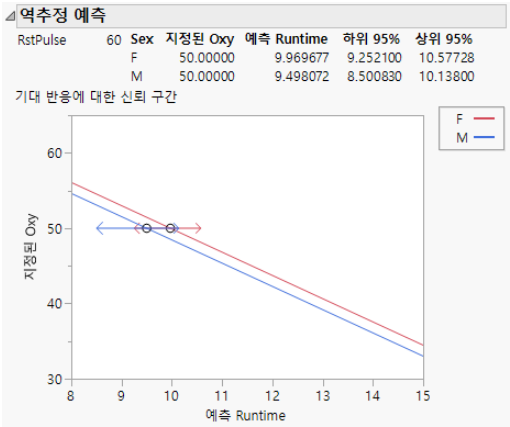


그림 3.38에 표시된 보고서에는 여성과 남성 둘 다에 대한 예측 Runtime 값이 제공됩니다. 또한 평균 반응 50을 제공하는 Runtime 값에 대한 95% 신뢰 구간도 보고서에 포함됩니다.

그림 3.38 다중 회귀 모형에 대한 역추정 예측 보고서



이 그림에서는 RstPulse가 60일 때 여성 및 남성에 대한 선형 적합을 보여 줍니다. 두 신뢰 구간은 각각 빨간색과 파란색으로 표시됩니다.

Cox 혼합물

참고 : 이 옵션은 혼합물 모형에만 사용할 수 있습니다 .

표준 최소 제곱은 Scheffe(1958)에서 제안된 파라미터화를 사용하여 혼합물 모형을 적합시킵니다. 이 모형의 모수를 사용하여 혼합물 성분의 효과를 쉽게 판단할 수 없습니다. Cox 혼합물 모형은 재파라미터화되고 제약된 버전의 Scheffe 모형입니다. 모수 추정값을 사용하여 요인 효과 및 설계 공간의 참조점을 기준으로 한 반응 표면 모양을 도출할 수 있습니다. 자세한 내용은 Cornell 연구 자료(1990)에서 확인하십시오.

"Cox 혼합물" 옵션은 참조 혼합물을 입력하는 창을 엽니다. 참조 혼합물에 대해 합이 1이 되지 않는 성분을 입력하면 합이 1이 되도록 성분의 크기가 비례적으로 조정됩니다. 재조정된 혼합물은 보고서에 "참조 혼합물"로 표시됩니다. 성분 효과도 보고서에 나타납니다. Cox 성분 효과는 요인이 Cox 효과 방향을 따라 최소값에서 최대값으로 이동할 때 예측 반응의 차이입니다.

Cox 혼합물의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Five Factor Mixture.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y1 을 선택하고 **Y** 를 클릭합니다 .
4. X1 ~ X5 를 선택합니다 .
5. **매크로 > 혼합물 반응 표면**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y1" 의 빨간색 삼각형을 클릭하고 **추정값 > Cox 혼합물**을 선택합니다 .

8. " 참조 혼합물 " 규격 창에서 다음 값을 입력합니다 .
   - X1 옆에 0.15 입력
   - X2 옆에 0.25 입력
   - X3 옆에 0.4 입력
   - X4 옆에 0.1 입력
   - X5 옆에 0.1 입력
9. **확인**을 클릭합니다 .

그림 3.39 Cox 혼합물

Cox 참조 혼합 모형

Cox 모수 추정값

모수	추정값	표준 오차	t 비	Prob>|t|
절편	2.071	1.2541	1.651	0.1035
X1	3.120	7.7538	0.402	0.6887
X2	-5.828	7.0104	-0.831	0.4089
X3	-0.696	3.5822	-0.194	0.8465
X4	3.935	12.6619	0.311	0.7570
X5	8.740	12.5405	0.697	0.4883
X1^2	-15.897	16.3627	-0.972	0.3349
X2^2	17.877	13.5932	1.315	0.1931
X3^2	5.234	8.2498	0.634	0.5280
X4^2	-91.239	159.9347	-0.570	0.5703
X5^2	-145.693	158.3991	-0.920	0.3611
X1*X2	-2.938	24.9012	-0.118	0.9065
X1*X3	-7.617	19.1697	-0.397	0.6924
X2*X3	-16.250	18.7489	-0.867	0.3893
X1*X4	53.690	50.5711	1.062	0.2923
X2*X4	-17.702	49.9082	-0.355	0.7240
X3*X4	-7.808	44.0229	-0.177	0.8598
X1*X5	31.813	50.6186	0.628	0.5319
X2*X5	-2.279	49.3761	-0.046	0.9633
X3*X5	17.983	43.5145	0.413	0.6808
X4*X5	177.431	97.3862	1.822	0.0731

Cox 참조 혼합

요인	참조 혼합
X1	0.1500000
X2	0.2500000
X3	0.4000000
X4	0.1000000
X5	0.1000000

구성 요소 효과

성분	Cox 효과	구성 요소 범위
X1	0.80746	0.2200000
X2	-2.33103	0.3000000
X3	-0.48747	0.4200000
X4	0.65585	0.1500000
X5	1.45663	0.1500000

보고서에는 Cox 혼합물 모형에 대한 모수 추정값과 표준 오차 및 가설 검정이 표시됩니다. 참조 혼합물이 오른쪽에 나타납니다. 아래쪽에는 성분 효과와 성분 범위가 나타납니다.

모수 검정력

통계적 검정의 **검정력**은 실제로 차이가 있을 때 검정이 유의할 확률입니다. 검정력은 연구에서 실제 효과가 유의하다고 선언할 가능성을 나타냅니다. "모수 검정력" 옵션은 **후향적** 검정력 분석을 처리합니다.

참고: 필요한 차이를 감지할 수 있도록 충분한 수의 관측값이 연구에 포함되게 하려면 실험을 **설계**할 때 검정력에 대한 정보를 사용하십시오 . 이 유형의 분석을 **전향적** 검정력 분석이라고 합니다 . DOE 플랫폼을 사용하여 연구를 설계하는 것이 좋습니다 . DOE > 표본 크기 및 검정력과 DOE > 설계 평가 둘 다 전향적 검정력 분석에 유용합니다 . 표준 최소 제곱을 사용한 전향적 검정력 분석의 예는 " 전향적 검정력 분석 "(198 페이지) 에서 확인하십시오 .

차이를 감지하기 위한 검정력은 다음 요인의 영향을 받습니다 .

- 표본 크기
- 알 수 없는 잔차 오차 분산
- 검정의 유의 수준
- 감지할 효과의 크기

이미 연구를 수행하고 데이터를 분석하여 관심 효과가 유의하지 않음을 확인했다고 가정해 보겠습니다. 감지할 수 있는 차이의 크기 또는 수행한 검정의 검정력에 관심이 있을 수 있습니다. 아니면 주어진 크기의 차이를 높은 확률로 감지하기 위해 필요한 관측값 수를 알고 싶을 수도 있습니다.

"모수 검정력" 옵션은 "모수 추정값" 보고서에 후향적 검정력 분석과 관련된 세 개의 값 열을 삽입합니다. 최소 유의값(LSV0.05), 최소 유의 개수(LSN0.05) 및 검정력 계산(조정 검정력 0.05)이 제공됩니다.

모수 검정력 계산은 관측된 표본과 변동 프로파일이 동일한 새 표본에 적용됩니다.

주의 : LSV0.05, LSN 및 조정 검정력 0.05 에서 제공하는 결과를 전향적 검정력 분석에 사용하면 안 됩니다 . 이 값은 향후 연구에 내재된 불확실성을 반영하지 않습니다 .

- LSV0.05 는 **최소 유의값**입니다 . 이 값은 유의 수준 0.05 에서 이 검정을 유의하게 만드는 추정값의 가장 작은 절대값입니다 . 더 구체적으로 설계에 대한 관측값 수 , 평균 제곱 오차 , 제곱합 및 교차곱 행렬은 변경되지 않는다고 가정해 보겠습니다 . 그런 다음 추정값의 절대값이 LSV0.05 보다 작으면 Prob>|t| 값이 0.05 를 초과합니다. 자세한 내용은 "LSV(최소 유의값)"(196 페이지) 에서 확인하십시오 .
- LSN 은 **최소 유의 개수**입니다 . 이 값은 유의 수준 0.05 에서 이 검정을 유의하게 만드는 관측값 수입니다 . 구체적으로 설계에 대한 모수 추정값 , 평균 제곱 오차 , 제곱합 및 교차곱 행렬은 변경되지 않는다고 가정해 보겠습니다 . 그런 다음 관측값 수가 LSN 보다 작으면 Prob>|t| 값이 0.05 를 초과합니다 . 자세한 내용은 "LSN(최소 유의 개수)"(195 페이지) 에서 확인하십시오 .
- 조정 검정력 0.05 는 **조정 검정력** 값입니다 . 이 값은 이 검정이 유의할 확률의 추정값입니다 . 현재 연구의 표본 값은 일반적으로 검정력 계산에 사용되는 모수 값으로 대체됩니다 . **조정** 검정력 계산은 표본 추정값을 비중심성 모수에 대한 계산식에 직접 대체하여 발생하는 편향을 조정합니다 (Wright and O'Brien 1988). 자세한 내용은 " 조정 검정력 및 신뢰 구간 "(196 페이지) 에서 확인하십시오 .

LSV, LSN 및 조정 검정력은 검정의 민감도를 평가하는 데 유용합니다. 이러한 후향적 계산은 이해를 위한 교육용 도구로도 사용됩니다. 그러나 값을 해석할 때 주의해야 합니다(Hoenig and Heisey 2001).

LSV, LSN 및 조정 검정력에 대한 자세한 내용은 "검정력 분석"(193 페이지)에서 확인하십시오. 후향적 분석의 예는 "후향적 검정력 분석의 예"(197 페이지)에서 확인하십시오.

추정값 상관계수

"추정값" 메뉴의 "추정값 상관계수" 명령은 모수 추정값에 대한 상관 행렬을 계산합니다. 이러한 상관은 공선성이 있는지 여부를 나타냅니다.

이 행렬의 생성에 대한 이해를 위해 일반적인 최소 제곱 회귀 공식을 고려합니다. 여기서 반응 (Y) 은 예측 변수 ($x's$) + 오차 (ε) 의 선형 함수입니다.

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon$$

데이터 테이블의 각 행에는 반응 값 및 p 예측 변수의 값이 포함됩니다. 각 관측값에 대해 예측 변수 값은 고정 값으로 간주됩니다. 그러나 반응 값은 확률 변수의 실현값으로 간주됩니다.

고정된 예측 변수 값을 고려하면 Y 값 집합에 대해 β_0, β_1, ..., β_p 계수를 추정할 수 있습니다. 일반적으로 Y 값 집합이 다르면 계수 추정값이 다릅니다. "추정값 상관계수" 옵션은 이러한 모수 추정값의 이론적 상관계수를 계산합니다. 기술 상세 정보는 "사용자 검정 예에 대한 상세 정보"(188 페이지)에서 확인하십시오.

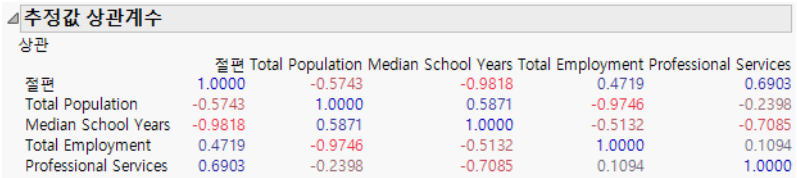
모수 추정값의 상관계수는 예측 변수 값과 절편을 나타내는 항에만 의존합니다. 두 모수 추정값 간의 상관계수는 반응 값의 영향을 받지 않습니다.

두 추정값 간에 높은 양의 상관관계가 있으면 해당하는 두 예측 변수 간에 공선성 관계가 존재할 수 있음을 나타냅니다. 그러나 이러한 상관관계를 주의하여 해석해야 합니다(Belsley et al. 1980, p. 185, 92–94). 또한 예측 변수를 재조정하여 평균이 변화되면 모수 추정값과 절편 값의 상관관계가 변경됩니다.

추정값 상관계수의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Socioeconomic.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Median House Value 를 선택하고 **Y** 를 클릭합니다.
4. Total Population, Median School Years, Total Employment 및 Professional Services 를 선택하고 **추가**를 클릭합니다.
5. " 강조 " 목록에서 **최소 보고서**를 선택합니다.
6. **실행**을 클릭합니다.
7. " 반응 Median House Value" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 추정값 상관계수**를 선택합니다.

그림 3.40 추정값 상관계수 보고서



추정값 상관계수

상관

	절편	Total Population	Median School Years	Total Employment	Professional Services
절편	1.0000	-0.5743	-0.9818	0.4719	0.6903
Total Population	-0.5743	1.0000	0.5871	-0.9746	-0.2398
Median School Years	-0.9818	0.5871	1.0000	-0.5132	-0.7085
Total Employment	0.4719	-0.9746	-0.5132	1.0000	0.1094
Professional Services	0.6903	-0.2398	-0.7085	0.1094	1.0000

보고서에서는 절편과 Median School Years의 모수 추정값 간에 높은 음의 상관관계가 있음을 보여 줍니다(–0.9818). Total Population과 Total Employment 간에도 높은 음의 상관관계가 있습니다(–0.9746).

명목형 효과에 대한 코딩

명목형 모델링 유형의 열을 모형에 입력하면 JMP에서는 이 열을 내부적으로 연속형 표시 변수 집합으로 나타냅니다. 각 변수는 –1, 0, 1 값만 가정합니다. 이 코딩은 표시 변수를 사용하여 명목형 변수를 코드화하는 여러 방법 중 하나입니다. 명목형 열에 *n* 개 수준이 있는 경우 열을 나타내려면 *n-1* 개의 표시 변수가 필요합니다. *n-1* 개의 표시 변수가 필요하다는 것은 명목형 열과 관련된 주효과가 *n-1* 의 자유도를 갖는다는 사실과 직접 관련이 있습니다. 자세한 내용은 "통계 상세 정보" 부록의 "명목형 요인"(493 페이지)에서 확인하십시오.

팁 : 기본 보고서의 빨간색 삼각형 메뉴에서 열 저장 > 코딩 테이블 저장을 선택하여 코딩을 볼 수 있습니다 . 자세한 내용은 " 코딩 테이블 저장 "(167 페이지) 에서 확인하십시오 .

수준이 4 개인 명목형 열이 있다고 가정해 보겠습니다. Cholesterol.jmp 샘플 데이터 테이블의 treatment 열을 예로 사용합니다. treatment 열에는 A, B, Control 및 Placebo라는 4 개의 수준이 있습니다. 처음 세 수준은 각각 표시 변수로 나타냅니다. 이러한 표시 변수 이름은 treatment[A], treatment[B] 및 treatment[Control]로 지정됩니다.

지정된 수준의 표시 변수는 해당 수준에 1, 마지막 수준에 –1, 나머지 수준에 0 값을 각각 할당합니다 . 표 3.1 에서는 이 예에 대한 treatment[A], treatment[B] 및 treatment[Control] 표시 변수의 정의를 보여 줍니다 . 표시 변수 treatment[A] 를 예로 들어보겠습니다 . 표 3.1 에 표시된 대로 이 변수는 다음 값을 할당합니다 .

- treatment = A 인 행에 값 1 이 할당됩니다 .
- treatment = B 또는 Control 인 행에 값 0 이 할당됩니다 .
- treatment = Placebo 인 행에 값 –1 이 할당됩니다 .

표 3.1 Cholesterol.jmp 의 treatment 에 대한 표시 변수 예시

행에 할당된 treatment	treatment[A]	treatment[B]	treatment[Control]
A	1	0	0
B	0	1	0

표 3.1 Cholesterol.jmp 의 treatment 에 대한 표시 변수 예시 (계속)

행에 할당된 treatment	treatment[A]	treatment[B]	treatment[Control]
Control	0	0	1
Placebo	-1	-1	-1

수준 순서는 "값 순서" 열 특성 (할당된 경우) 또는 JMP 에서 할당한 기본 정렬 방식에 의해 결정됩니다. 일반적으로 기본 정렬 방식은 숫자 정렬 순서 (숫자) 와 영숫자 정렬 순서 (문자 데이터) 입니다. 그러나 월 이름과 같은 특정 범주형 값은 기본적으로 적절하게 정렬됩니다. 값 순서화에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

이러한 변수는 모형을 파라미터화하는 데 사용됩니다. 변수는 일반적으로 데이터 테이블에 표시되지 않지만 이러한 변수에 대한 추정 계수가 "모수 추정값" 및 다른 보고서에 제공됩니다. 사용 가능한 다른 코딩도 많지만 이 코딩이 실용적이고 해석 가능한 것으로 입증되었습니다.

순서형 효과의 코딩에 대한 자세한 내용은 "통계 상세 정보" 부록의 "순서형 요인"(503 페이지) 에서 확인하십시오.

효과 선별

선별 설계에서 오차 자유도를 제공하지 않는 경우가 종종 있습니다 . 따라서 효과에 대해 전통적 검정을 사용할 수 없습니다 . 이러한 경우 효과 선별이 특히 유용합니다 .

이러한 설계의 경우 효과 크기에 대한 대부분의 추론은 비절편 모수의 추정값이 상관관계가 없고 분산이 동일하다고 가정합니다. 이 가정은 많은 전통적 실험 설계와 관련된 모형에 적용됩니다. 그러나 이러한 가정이 해당되지 않는 경우가 있습니다. 두 경우 모두 효과 선별 플랫폼을 사용하면 유의한 효과를 확인하는 데 도움이 됩니다.

효과 선별 플랫폼은 **효과 희소성** 원리를 사용합니다 (Box and Meyer 1986). 이 원리는 선별 설계에서 연구하는 효과 중 활성 효과는 상대적으로 적다고 주장합니다. 대부분은 비활성 상태이며, 이는 실제 효과가 0 이고 추정값이 랜덤 오차임을 의미합니다.

" 효과 선별 " 에서 다음 옵션을 사용할 수 있습니다 .

척도화된 추정값 평균이 0 이고 범위가 2 가 되도록 척도화된 요인에 해당하는 모수 추정값을 제공합니다 . 자세한 내용은 " 척도화된 추정값 및 연속형 항 코딩 "(141 페이지) 에서 확인하십시오 .

정규 확률도 정규성에서 벗어난 모수 추정값을 식별하여 활성 효과를 확인하는 데 도움이 됩니다 . 자세한 내용은 " 정규 확률도 보고서 "(146 페이지) 에서 확인하십시오 .

Bayes 그림 Bayes 방법을 사용하여 모든 모형 항에 대한 사후 확률을 계산합니다 . 자세한 내용은 "Bayes 그림 보고서 "(147 페이지) 에서 확인하십시오 .

파레토도 직교화 및 표준화된 모수 추정값의 절대값을 해당 절대값의 합과 연결하여 그림에 나타냅니다 . 자세한 내용은 " 파레토도 보고서 "(149 페이지) 에서 확인하십시오 .

척도화된 추정값 및 연속형 항 코딩

모수 추정값은 해당 요인의 척도에 따라 크게 좌우됩니다. 요인을 그램(g)에서 킬로그램(kg)으로 변환하면 모수 추정값이 1,000의 배수로 변경됩니다. 동일한 변경을 제곱(2차) 항에 적용하면 모수 추정값이 100만의 배수로 변경됩니다.

효과 크기를 더 잘 이해하고 비교하려면 모수 추정값을 척도에 따라 변하지 않는 방식으로 검토해야 합니다. 모수 추정값의 크기를 반응에 대한 해당 항의 효과 크기와 연관시키는 척도를 사용하는 것이 적절합니다. 이를 수행하는 방법에는 여러 가지가 있습니다.

보고서의 빨간색 삼각형 메뉴에 있는 효과 선별 > 척도화된 추정값 옵션은 척도화된 요인에 해당하는 계수를 제공합니다. 요인은 평균이 0이고 범위가 2가 되도록 척도화됩니다. 그림 3.41에서는 Drug.jmp에 대한 보고서를 보여 줍니다.

요인에 대한 표본 값의 최대값과 최소값이 표본 평균에서 등거리에 있는 경우 척도화된 요인의 범위는 –1 ~ 1입니다. 이 척도화는 실험 설계에 사용된 기존의 코딩에 해당합니다. 단일 요인을 사용한 회귀의 경우 척도화된 모수 추정값은 요인이 전체 범위에서 이동할 때 예측된 반응 변동의 절반입니다.

척도화된 추정값은 데이터에 코드화되지 않은 값이 포함된 경우 모형 항의 영향을 평가하는 데 중요합니다. 직교 설계의 경우 척도화된 추정값은 코드화되지 않은 데이터의 추정값과 동일합니다.

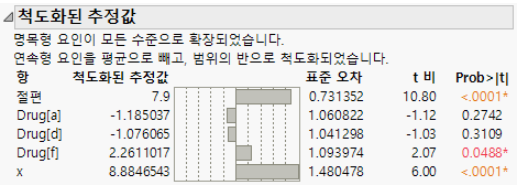
참고 : " 코딩 " 열 특성은 –1 에서 1 사이의 범위로 값이 코드화되도록 요인 값을 선형으로 척도화합니다 . 모수 추정값은 이러한 코드화된 값으로 제공되므로 이 경우에는 척도화된 추정값이 필요하지 않습니다 . 척도화된 추정값을 구하는 데 사용되는 변환과 달리 코드화된 값은 평균이 0이 아닐 수 있습니다 .

척도화된 추정값의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Drug 와 x 를 선택하여 **모형 효과 생성** 목록에 추가합니다 .
5. " 강조 " 를 **최소 보고서**로 변경합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 y" 의 빨간색 삼각형을 클릭하고 **효과 선별 > 척도화된 추정값**을 선택합니다 .

보고서에서는 연속형 요인 x가 평균을 기준으로 중심화되고 범위의 반으로 척도화되었음을 나타냅니다.

그림 3.41 척도화된 추정값 보고서



척도화된 추정값

명목형 요인이 모든 수준으로 확장되었습니다.
연속형 요인을 평균으로 빼고, 범위의 반으로 척도화되었습니다.

항	척도화된 추정값	표준 오차	t 비	Prob>|t|
절편	7.9	0.731352	10.80	<.0001*
Drug[a]	-1.185037	1.060822	-1.12	0.2742
Drug[d]	-1.076065	1.041298	-1.03	0.3109
Drug[f]	2.2611017	1.093974	2.07	0.0488*
x	8.8846543	1.480478	6.00	<.0001*

모형은 각 Drug 그룹에 대해 하나씩 세 개의 평행선을 적합시킵니다. x 값의 범위는 3 ~ 21 입니다. x 의 척도화된 추정값 8.8846543 은 Drug 그룹 중 하나에 대해 x = 21 에 대한 예측값과 x = 3 에 대한 예측값 간 차이의 절반입니다. 보고서의 빨간색 삼각형 메뉴에서 **열 저장 > 예측 계산식**을 선택하여 이 사실을 확인할 수 있습니다. 그런 다음 x = 21 및 x = 3 에서 각 Drug 그룹에 대한 예측값을 구하기 위해 행을 추가합니다.

이 특정 데이터 집합의 x 값 범위에서 x 의 영향은 약 17.8 단위의 범위에서 반응에 변화를 주는 것입니다. 원시 데이터에 기반한 x 의 모수 추정값은 0.9871838 이므로 반응에 대한 직접적인 해석을 허용하지 않습니다.

그림 옵션

"정규 확률도", "Bayes 그림" 및 "파레토도" 옵션은 "효과 선별" 보고서의 일부로 나타나는 보고서를 제공합니다. 이러한 보고서를 추정값 간의 이분산 및 상관관계를 수정하도록 구성할 수 있습니다.

참고 : " 정규 확률도 ", "Bayes 그림 " 및 " 파레토도 " 옵션을 사용하려면 모형에 4 개 이상의 모수가 포함되어야 합니다 . 이러한 모수 중 하나가 절편일 수 있습니다 .

변환

그림 옵션 중 하나를 선택하면 " 효과 선별 " 보고서 제목 바로 아래에 다음 정보가 나타납니다 .

- 추정값의 분산이 동일하고 상관관계가 없는 경우 다음과 같은 두 가지 노트가 나타납니다 .
  - 모수 추정값들은 등분산을 가집니다 .
  - 모수 추정값 간에 상관 관계가 없습니다 .
- 추정값의 분산이 동일하지 않거나 상관관계가 있는 경우 옵션 목록이 관련 노트를 대체합니다 . 기본적으로 선택된 목록 항목은 JMP 에서 추정값을 변환했음을 나타냅니다 . 변환 중 하나 또는 둘 다 실행 취소하려면 해당 목록 항목을 선택합니다 .

Lenth PSE 값

Lenth PSE(유사 표준 오차) 테이블은 노트 또는 옵션 목록 바로 아래에 나타납니다 . PSE 에 대한 설명은 "Lenth PSE"(110 페이지) 에서 확인하십시오 . Lenth 테이블에 나타나는 통계량은 모수 추정값의 분산 및 상관관계에 따라 달라집니다 .

모수 추정값의 분산이 동일하고 상관관계가 없는 경우 Lenth 테이블에 다음 통계량이 제공됩니다.

Lenth PSE 추정값에 대한 Lenth 유사 표준 오차입니다.

모수 추정값의 분산이 다르거나 상관관계가 있는 경우 또는 둘 다 해당하는 경우 Lenth 테이블에 다음 통계량이 제공됩니다.

t- 검정 척도 Lenth PSE 변환된 모수 추정값에 대해 계산된 Lenth 유사 표준 오차를 변환된 척도의 표준 오차로 나눈 값입니다.

코드화된 척도 Lenth PSE 변환된 모수 추정값에 대한 Lenth 유사 표준 오차입니다.

모수 추정값 모집단 보고서

"모수 추정값 모집단" 보고서에는 모수 추정값에 대한 검정이 제공됩니다. 검정은 옵션 목록에 지정된 변환을 기반으로 합니다.

- **등분산을 가지도록 표준화된 추정값 사용** 옵션은 정규화 변환을 적용하여 분산을 표준화합니다. 이 옵션은 분산이 동일하지 않을 때 기본적으로 선택됩니다.
- **상관 관계가 없도록 직교된 추정값 사용** 옵션은 변환을 적용하여 상관관계를 제거합니다. 이 옵션은 추정값이 상관되어 있을 때 기본적으로 선택됩니다. 적용되는 변환은 순차 제곱합을 계산하는 데 사용되는 변환과 동일합니다. 추정값은 모든 이전 변수가 모형에 입력된 후 변수의 추가 기여도를 측정합니다.
- 추정값의 분산이 동일하고 상관관계가 없음을 나타내는 노트가 있으면 변환이 적용되지 않습니다.

테이블에 나타나는 열은 처음에 노트 또는 옵션 목록에 설명된 선택 항목에 따라 다릅니다. 보고서에서는 *p* 값이 0.20 이하인 추정값에 해당하는 행이 강조 표시됩니다. 보고서의 모든 버전에 "항", "추정값" 및 *t* 비와 Prob>|t| 또는 유사 *t* 비와 유사 *p* 값 중 하나가 제공됩니다.

항 모수 추정값에 관심이 있는 모형 항을 제공합니다.

추정값 모수에 대한 추정값을 제공합니다. 모형 효과의 척도가 동일한 경우에만 추정값 크기를 비교할 수 있습니다.

t 비 오차의 자유도가 있는 경우 나타납니다. 이 값은 모수 추정값을 표준 오차로 나눈 값입니다.

Prob>|t| 검정의 *p* 값을 제공합니다. 변환이 적용되면 이 옵션은 변환된 데이터를 사용하여 검정의 *p* 값을 제공합니다.

유사 t 비 오차의 자유도가 없는 경우 나타납니다. 모수의 상대 표준 오차가 동일한 경우 유사 t 비는 모수 추정값을 Lenth PSE 로 나눈 값입니다. 상대 표준 오차가 다른 경우에는 "유사 t 비"(110 페이지)에 표시된 대로 계산됩니다.

유사 p 값 오차의 자유도가 없는 경우 나타납니다. 이 *p* 값은 *t* 분포를 사용하여 파생됩니다. 자유도는 *m*/3 이며 가장 가까운 정수로 내림합니다. 여기서 *m* 은 모수의 수입니다.

등분산을 가지도록 표준화된 추정값 사용을 선택하고 모수 추정값이 상관관계가 없다는 노트가 나타나면 보고서에 "표준화 추정값"이라는 열이 표시됩니다. 이 열은 추정값이 등분산을 가지도록 변환한 결과로 생성된 모수 추정값을 제공합니다.

등분산을 가지도록 표준화된 추정값 사용과 **상관 관계가 없도록 직교된 추정값 사용**을 둘 다 선택하면 보고서에 " 직교 코드화됨 " 이라는 열이 표시됩니다 . 다음 정보가 제공됩니다 .

직교 코드화됨 추정값을 직교화하는 데 사용되는 변환 결과로 생성된 모수 추정값을 제공합니다 .

직교 t 비 오차의 자유도가 있는 경우 나타납니다 . 변환된 추정값의 *t* 비를 제공합니다 .

유사 직교 t 비 오차의 자유도가 없는 경우 나타납니다 . 이 값은 직교화된 추정값 (직교 코드화됨) 을 코드화된 척도 Lenth PSE 로 나누어 계산된 *t* 비입니다 .

효과 선별 보고서

그림 3.42에서는 Bicycle.jmp 샘플 데이터 테이블의 "Fit Model" 스크립트를 실행하여 생성되는 "효과 선별" 보고서를 보여 줍니다. 이 형식의 보고서를 생성하려면 **효과 선별 > 정규 확률도**를 선택해야 합니다. 보고서 제목 바로 아래의 노트는 변환이 필요하지 않음을 나타냅니다. 따라서 Lenth PSE가 표시됩니다. 오차에 대한 자유도가 없으므로 잔차 오차의 추정값을 생성할 수 없습니다. 따라서 Lenth PSE를 잔차 오차의 추정값으로 사용하여 유사 *t* 비를 구합니다. 이러한 *t* 비에 대해 유사 *p* 값이 제공됩니다. *p* 값이 0.20 이하인 세 개의 추정값에 해당하는 비절편 항의 행이 강조 표시됩니다.

그림 3.42 분산이 동일하고 상관관계가 없는 추정값에 대한 효과 선별 보고서

효과 선별

모수 추정값들은 등분산을 가집니다.
모수 추정값 간에 상관 관계가 없습니다.

Lenth PSE
0.75

모수 추정값 모집단

항	추정값	유사 t 비	유사 p 값
절편	66.5000	88.6667	<.0001*
HBars[down]	0.5000	0.6667	0.5649
Dynamo[off]	-6.0000	-8.0000	0.0096*
Seat[down]	1.7500	2.3333	0.1266
Tires[hard]	-1.2500	-1.6667	0.2195
Gear[low]	-11.2500	-15.0000	0.0023*
Raincoat[off]	0.2500	0.3333	0.7665
Brkfast[no]	0.5000	0.6667	0.5649

직교 t-검정에 유사 표준 오차가 사용되었습니다.

정규 확률도

분산이 동일하지 않고 상관관계가 있는 추정값에 대한 효과 선별 보고서

Odor.jmp 샘플 데이터 테이블에서 "Model" 스크립트를 실행하고 **실행**을 클릭합니다. 그림 3.43에 표시된 보고서를 생성하려면 " 반응 odor" 의 빨간색 삼각형을 클릭하고 **효과 선별 > 정규 확률도**를 선택합니다. " 반응 odor" 의 빨간색 삼각형을 클릭하고 "Bayes 그림 " 또는 " 파레토도 " 옵션을 선택하여 보고서를 생성할 수도 있습니다.

보고서 제목 바로 아래에 있는 노트는 추정값을 표준화 및 직교화하기 위해 변환이 필요했음을 나타냅니다. " 추정값 상관 " 보고서에는 상관 행렬이 표시됩니다.

이 보고서에는 *t*-검정 척도 및 코드화된 척도 Lenth PSE가 표시됩니다. 그러나 오차에 대한 자유도가 있으므로 " 모수 추정값 모집단 " 보고서의 검정에 Lenth PSE가 사용되지 않습니다. *p* 값이 0.20

이하인 세 개의 추정값에 해당하는 비절편 항의 행이 강조 표시됩니다. "모수 추정값 모집단" 보고서 아래의 노트는 직교 추정값이 모형에 대한 입력 순서에 따라 달라진다는 것을 나타냅니다.

그림 3.43 분산이 동일하지 않고 상관관계가 있는 추정값에 대한 효과 선별 보고서

효과 선별

등분산을 가지도록 표준화된 추정값 사용

상관 관계가 없도록 직교된 추정값 사용

	Lenth PSE
t-검정 척도	1.7815489
코드화된 척도	14.253752

모수 추정값 모집단

항	추정값	t 비	직교 코드화됨	직교 t 비	Prob>|t|
절편	-4.3333	-0.2422	15.2000	1.8998	0.1159
temp	-25.8750	-2.3618	-18.8964	-2.3618	0.0646
gl ratio	-23.5000	-2.1450	-17.1620	-2.1450	0.0848
ht	-10.6250	-0.9698	-7.7594	-0.9698	0.3767
temp*gl ratio	-22.0000	-1.4200	-11.3608	-1.4200	0.2149
temp*ht	-5.2500	-0.3389	-2.7111	-0.3389	0.7485
gl ratio*ht	-5.0000	-0.3227	-2.5820	-0.3227	0.7600
temp*temp	25.2917	1.5684	12.2138	1.5266	0.1874
gl ratio*gl ratio	18.5417	1.1498	9.5025	1.1877	0.2883
ht*ht	-7.2083	-0.4470	-3.5763	-0.4470	0.6736

각 직교 추정값은 이전의 효과에 의해 조건부가 됩니다.

추정값 상관

상관										
절편	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.5547	-0.5547	-0.5547
temp	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
gl ratio	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ht	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
temp*gl ratio	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
temp*ht	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
gl ratio*ht	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
temp*temp	-0.5547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0769	0.0769
gl ratio*gl ratio	-0.5547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0769	1.0000	0.0769
ht*ht	-0.5547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0769	0.0769	1.0000

상관 관계가 없도록 만들기 위한 변환

추정값 상관 보고서

"추정값 상관" 보고서는 추정값이 상관된 경우에만 나타납니다(그림 3.43). 이 보고서는 모수 추정값에 대한 상관 행렬을 제공합니다. 이 행렬은 빨간색 삼각형의 추정값 > 추정값 상관계수 옵션을 선택하여 얻은 행렬과 비슷합니다. 그러나 더 간결한 형태로 표현하기 위해 보고서에 열 머리글이 표시되지 않습니다. 자세한 내용은 "추정값 상관계수"(138 페이지)에서 확인하십시오.

" 상관 관계가 없도록 만들기 위한 변환 " 보고서

" 상관 관계가 없도록 만들기 위한 변환 " 보고서는 추정값이 상관된 경우에만 나타납니다 . 이 보고서는 상관관계가 없는 모수 추정값을 생성하기 위해 설계 행렬을 변환하는 데 사용되는 행렬을 제공합니다 . 변환 또는 **직교 코드화된** 추정값은 이 행렬을 사용하여 원래 추정값을 앞에 곱하고 결과에 배율을 곱하여 구합니다 . 배율은 RMSE(제곱근 평균 제곱 오차) 와 행 수의 함수입니다 . 이 행렬은 다음과 같이 정의됩니다 .

RMSE/sqrt(NRows)

변환 행렬은 Cholesky 분해를 사용하여 얻을 수 있습니다. X′X를 LL′로 표현합니다. 여기서 L은 Cholesky 분해의 하삼각 행렬입니다. 그러면 변환 행렬이 L′로 제공됩니다.

이 변환은 모형 항 순서에서 각 회귀변수를 앞에 오는 회귀변수에 대해 직교정규화합니다. 변환을 통해 대각 요소가 동일한 대각 공분산 행렬이 생성됩니다. 코드화된 추정값은 이 반복 과정의 결과입니다.

참고 : 직교 코드화된 추정값은 모형의 항 순서에 따라 다릅니다 . 각 효과의 기여도는 해당 효과가 이전에 입력한 효과와 직교된 후에만 추정됩니다 . 주효과를 먼저 입력한 후 이원 교호작용 , 삼원 교호작용 등의 순서로 입력하는 것이 좋습니다 .

정규 확률도 보고서

"정규 확률도" 보고서 제목 아래에서 정규 확률도 또는 1/2 정규 확률도(Daniel 1959)를 선택합니다. 두 그림 모두 효과 희소성 원리, 즉 비교적 적은 수의 효과가 활성화된다는 개념에 근거합니다. 비활성 효과는 랜덤 잡음을 나타냅니다. 해당 추정값은 평균이 0이고 분산이 σ^2인 정규 분포를 따른다고 가정할 수 있습니다. 여기서 σ^2은 잔차 오차 분산을 나타냅니다. 따라서 정규 확률도에서 비활성 효과를 나타내는 추정값은 기울기가 σ인 선에 가깝게 분포됩니다.

정규 확률도

변환이 필요하지 않은 경우 정규 확률도의 수직 좌표는 추정값을 나타내고 수평 좌표는 정규 분위수를 나타냅니다. 비활성 효과를 나타내는 점은 기울기가 σ인 선을 따라야 합니다. Lenth PSE를 사용하여 σ를 추정하고 이 기울기의 파란색 선이 그림에 표시됩니다.

직교성 변환이 적용된 경우 세로 축은 정규 추정값을 나타냅니다. 이러한 값은 "모수 추정값 모집단" 보고서에 있는 직교 *t* 비 값입니다. 직교 *t* 비 값은 직교 코드화된 추정값을 코드화된 척도 Lenth PSE로 나눈 값입니다.

추정값은 σ의 추정값에 의해 정규화되므로 비활성 효과에 해당하는 점은 기울기 1의 선을 따라 분포되어야 합니다. 기울기가 1인 빨간색 선과 기울기가 t-검정 척도 Lenth PSE와 동일한 파란색 선이 그림에 표시됩니다.

모든 경우에 "모수 추정값 모집단" 보고서의 *p* 값을 기반으로 0.20 수준에서 정규성을 벗어난 추정값은 그림에 라벨이 지정됩니다.

1/2 정규 확률도

1/2 정규 확률도에는 효과의 절대값이 표시됩니다. 축 구성 및 표시된 선은 정규 확률도의 양상을 반영합니다.

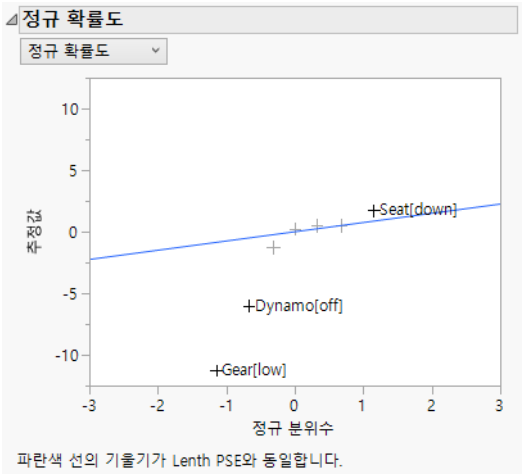
그림 3.44에서는 Bicycle.jmp 샘플 데이터 테이블에 대한 "정규 확률도" 보고서를 보여 줍니다. 이 모형에는 변환이 필요하지 않으므로 정규 분위수에 대한 원시 추정값이 그림에 표시됩니다. 기울기가 Lenth PSE와 동일한 선이 그림에 표시됩니다. 이 그림에서는 Gear, Dynamo 및 Seat가 활성 요인임을 나타냅니다.

정규 확률도의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Bicycle.jmp 를 엽니다 .

2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. HBars ~ Brkfast 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **효과 선별 > 정규 확률도**를 선택합니다 .
   다음과 같은 정규 확률도가 나타납니다 .

그림 3.44 정규 확률도



Bayes 그림 보고서

Bayes 그림 보고서에는 활성 효과를 확인하는 또 다른 방법이 제공됩니다. 이 보고서는 Bayes 방법을 사용하여 사후 확률을 계산하는 데 도움이 됩니다. Box and Meyer(1986) 를 기반으로 한 이 방법은 추정값이 두 분포의 혼합물이라고 가정합니다 . 비활성 효과에 해당하는 대부분의 추정값은 분산이 σ^2 인 순수 랜덤 정규 잡음으로 간주됩니다. 활성 효과에 해당하는 나머지 추정값은 분산이 σ^2 보다 K 배 큰 **오염** 분포에서 오는 것으로 간주됩니다 .

항 모수 추정값에 해당하는 모형 항을 제공합니다 .

추정값 모수 추정값을 제공합니다. Bayes 그림은 추정된 표준편차가 1 인 추정값과 관련하여 생성됩니다 . 추정값 간에 상관관계가 없으면 t 비가 사용됩니다 . 추정값 간에 상관관계가 있으면 직교 t 비가 사용됩니다 .

사전 확률 추정값이 0 이 아닐 확률 , 즉 추정값이 오염 분포에 있을 확률을 지정할 수 있습니다 . 추정값에 대한 사전 확률은 대개 동일한 값으로 설정됩니다 . 처음에는 일반적으로 권장되는 값 0.2 가 표시되지만 값을 변경할 수 있습니다 .

K 오염 오차 분산에 대한 오염 분포 분산의 비율을 나타내는 오염 계수 값입니다. 일반적으로 K 는 기본값 10 으로 설정됩니다 .

표준 오차 척도 표준 오차 추정값에 대한 자유도가 있는 경우 이 값은 1 로 설정됩니다. 보고서에 사용된 추정값은 단위 분산으로 변환 및 척도화되므로 JMP 에서 이 값을 사용합니다. 오차 추정값이 없는 포화 모형의 경우 이 값은 0 으로 설정됩니다 . 다른 값을 지정할 경우에는 RMSE 추정값의 배율로 간주하십시오 .

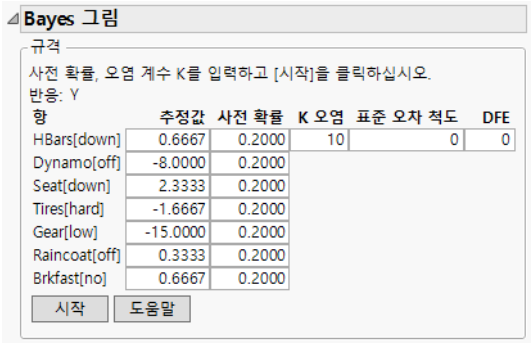
DFE 오차 자유도를 제공합니다 .

그림 3.45에서는 Bicycle.jmp 샘플 데이터 테이블에 대한 Bayes 그림의 기본 설정이 표시된 규격 창을 보여 줍니다. 이 창에서 "시작"을 클릭하면 보고서가 업데이트되어 각 항에 대한 사후 확률 및 막대 차트가 표시됩니다(그림 3.46).

Bayes 그림의 예

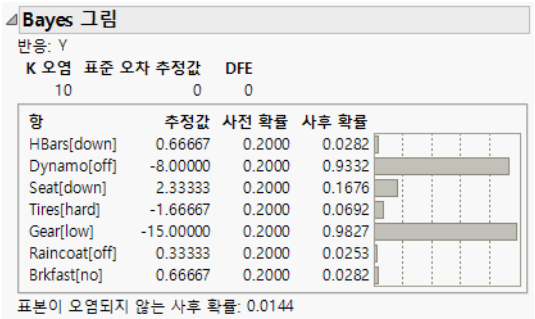
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Bicycle.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. HBars ~ Brkfast 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **효과 선별 > Bayes 그림**을 선택합니다 .

그림 3.45 Bayes 그림 규격



7. **시작**을 클릭하여 사후 확률을 계산합니다 .

그림 3.46 Bayes 그림 보고서



"Bayes 그림" 보고서에서 그림 아래의 노트에는 표본이 오염되지 않는 사후 확률이 제공됩니다. 사후 확률은 사전 확률 및 데이터를 기반으로 활성 효과가 전혀 없을 확률입니다. 사후 확률 값 (0.0144) 이 작으므로 활성 효과가 존재할 가능성이 있음을 나타냅니다. 사후 확률 열은 적어도 Dynamo 와 Gear 가 활성 효과임을 나타냅니다.

파레토도 보고서

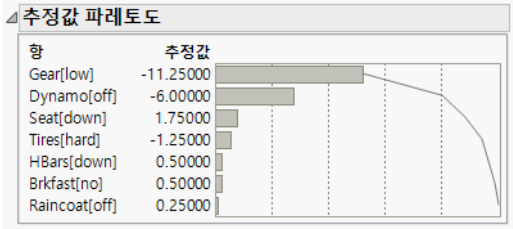
파레토도 보고서에는 추정값의 절대값에 대한 파레토 차트가 제공됩니다 . 그림 3.47 에서는 Bicycle.jmp 샘플 데이터 테이블에 대한 파레토도 보고서를 보여 줍니다 .

- 원래 추정값의 분산이 동일하고 상관관계가 없으면 원래 추정값이 그림에 표시됩니다 .
- 원래 추정값의 분산이 동일하지 않고 상관관계가 없으면 *t* 비가 그림에 표시됩니다 .
- 원래 추정값의 분산이 동일하지 않고 상관관계가 있으면 직교 코드화된 추정값이 그림에 표시됩니다 .

그림의 누적합 선은 추정값의 절대값을 합합니다. 직교 추정값은 모형에 항을 입력하는 순서에 따라 달라집니다.

참고 : 그림에 나타나는 추정값은 동일한 분산을 가지도록 표준화되고 직교하도록 변환됩니다 . 이러한 변환을 실행 취소하는 옵션을 사용할 수 있습니다 . 자세한 내용은 " 변환 "(142 페이지) 에서 확인하십시오 .

그림 3.47 파레토도



요인 프로파일링

"요인 프로파일링" 메뉴는 추정된 모형을 탐색하고 시각화하는 데 도움이 됩니다. 반응 표면의 모양을 탐색하고, 최적 설정을 찾고, 지정된 잡음 가정을 기반으로 반응 데이터를 시뮬레이션하고, 필요한 경우 반응을 변환할 수 있습니다.

프로파일러, 등고선 프로파일러, 혼합물 프로파일러 및 표면 프로파일러는 표준 최소 제곱 모형을 넘어 사용 범위가 확장되는 다목적 도구입니다. 해석 및 사용에 대한 자세한 내용은 Profilers 에서 확인하십시오.

모형의 반응 열에 한 변수의 가역 변환이 포함되어 있으면 프로파일러 및 교호작용 그림에 변환되지 않은 척도의 반응이 표시됩니다.

"요인 프로파일링"에서 다음 옵션을 사용할 수 있습니다.

참고 : 모형에 표현식 또는 벡터가 효과로 포함된 경우 이러한 옵션 중 대부분을 사용할 수 없습니다 .

프로파일러 각 X 변수에 대한 예측 트레이스를 표시합니다 . 하나 이상의 반응에 대한 최적 설정을 찾고 시뮬레이션을 사용하여 반응 분포를 탐색할 수 있습니다 . 자세한 내용은 " 프로파일러 "(151 페이지) 및 Profilers 의 에서 확인하십시오 .

교호작용 그림 모형에 교호작용 효과가 있는 경우 교호작용 그림의 행렬을 표시합니다 . 자세한 내용은 " 교호작용 그림 "(151 페이지) 에서 확인하십시오 .

등고선 프로파일러 반응 표면을 그래픽으로 최적화하는 데 유용한 대화식 등고선 프로파일러를 제공합니다 . 자세한 내용은 " 등고선 프로파일러 "(153 페이지) 및 Profilers 의 에서 확인하십시오 .

혼합물 프로파일러 혼합물 실험 모형의 반응 등고선을 삼원 그림에 표시합니다 . 자세한 내용은 " 혼합물 프로파일러 "(154 페이지) 및 Profilers 의 에서 확인하십시오 .

참고 : 이 옵션은 " 혼합물 효과 " 속성을 3 개 이상의 효과에 적용하거나 " 혼합물 " 특성을 3 개 이상의 열에 적용한 경우에만 나타납니다 .

입방체도 입방체 꼭지점에 요인 범위의 극단에 대한 예측값을 표시합니다 . 자세한 내용은 " 입방체도 "(155 페이지) 에서 확인하십시오 .

Y 의 Box-Cox 변환 분산의 정규성 및 동질성 가정 충족과 관련하여 가장 적합한 반응의 Box-Cox 멱 변환을 찾습니다 . 자세한 내용은 "Y 의 Box-Cox 변환 "(156 페이지) 에서 확인하십시오 .

표면 프로파일러 반응 표면의 3 차원 표면 그림을 표시합니다 . 자세한 내용은 " 표면 프로파일러 "(158 페이지) 및 Profilers 의 에서 확인하십시오 .

프로파일러

참고 : 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

프로파일러 또는 "예측 프로파일러"에서는 각 X 변수에 대한 예측 트레이스를 보여 줍니다.

그림 3.48에서는 Reactor.jmp 샘플 데이터 테이블에 대한 프로파일러의 일부를 보여 줍니다. 각 X 변수에 대한 수직 점선은 **현재 값** 또는 **현재 설정**을 표시합니다. 프로파일러를 사용하면 한 번에 하나의 변수를 변경하여 예측 반응에 대한 영향을 검토할 수 있습니다.

그림 3.48 예측 트레이스 그림

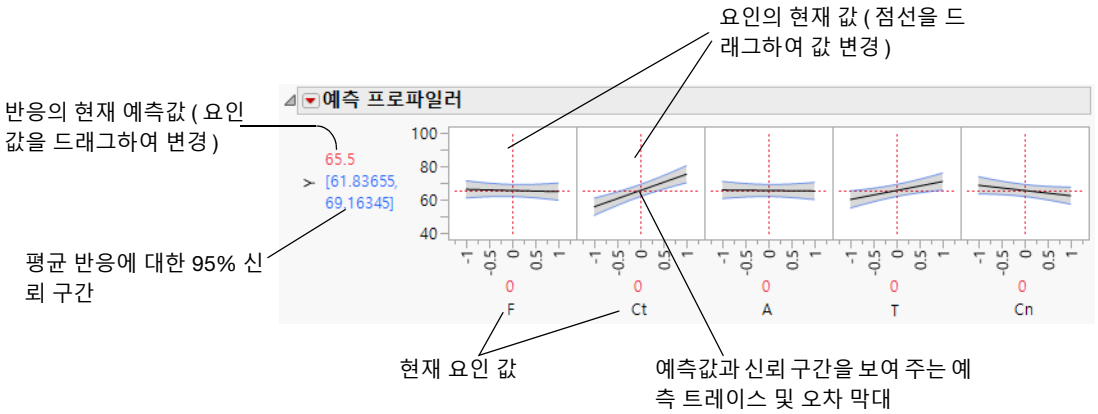


그림 3.48 의 F 및 Ct 요인은 연속형입니다. 요인이 명목형이면 가로 축에 수준이 표시됩니다.

각 X 변수에 대해 요인 이름 위의 값이 현재 값입니다. 그래프를 클릭하거나, 새 현재 값을 지정할 점선을 드래그하여 현재 값을 변경합니다.

- 수평 점선은 X 변수의 현재 값에 대한 각 Y 변수의 **현재 예측값**을 나타냅니다 .
- 그림 안의 검은색 선은 개별 X 변수의 현재 값을 변경할 때 예측값이 어떻게 변하는지 보여 줍니다 . 예측값에 대한 95% 신뢰 구간은 예측 트레이스 (연속형 변수의 경우) 또는 오차 막대 (범주형 변수의 경우) 주위의 점선 곡선으로 표시됩니다 .

교호작용 그림

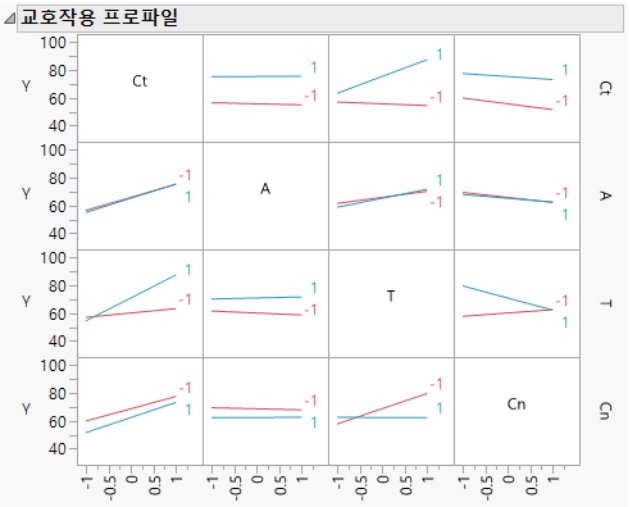
모형에 교호작용 효과가 있는 경우 "교호작용 그림" 옵션은 교호작용 그림의 행렬을 표시합니다. 행렬의 각 셀에는 그림이 나타나는 열에 표시되는 효과에 대해 가로 축이 척도화된 그림이 포함됩니다. 해당하는 행에 표시된 효과와 해당 효과의 교호작용에 대해 선분이 표시됩니다. 따라서 교호작용 그림은 행 효과와 열 효과의 교호작용을 보여 줍니다.

행 효과의 각 수준에 대해 선분이 그림에 표시됩니다. 모형에서 예측한 반응 값이 선분으로 결합됩니다. 평행하지 않는 선분은 교호작용이 있을 수 있다는 시각적 증거를 제공합니다. 그러나 교호작용이 있다는 결론을 내리기 전에 제안된 교호작용의 p 값을 확인해야 합니다. 그림 3.49에서는 Reactor.jmp 샘플 데이터 테이블에 대한 교호작용 그림 행렬을 제공합니다.

교호작용 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. **차수** 상자에 2 가 있는지 확인합니다 .
5. Ct, A, T 및 Cn 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 교호작용 그림**을 선택합니다 .

그림 3.49 교호작용 그림



T*Cn 교호작용에 해당하는 그림은 그림 맨 아래 행의 세 번째 그림 또는 마지막 그림 열의 세 번째 그림입니다. 두 그림 모두 낮은 수준의 T에서 Y에 대한 Cn의 효과가 상당히 일정하다는 것을 보여 줍니다. 이때 Cn이 설정된 수준이 높은지 아니면 낮은지는 상관이 없습니다. 그러나 높은 수준의 T에서 Y에 대한 Cn의 효과는 해당 수준에 따라 다릅니다. –1의 Cn은 1의 Cn보다 예측 Y 값이 더 높습니다. 이 교호작용은 p 값 < 0.0001이므로 유의합니다.

특정 설계에서는 이원 교호작용이 다른 이원 교호작용과 별칭 관계를 갖습니다. 별칭 관계가 있으면 " 교호작용 프로파일 " 그림에서 이러한 교호작용에 해당하는 셀이 흐리게 표시됩니다.

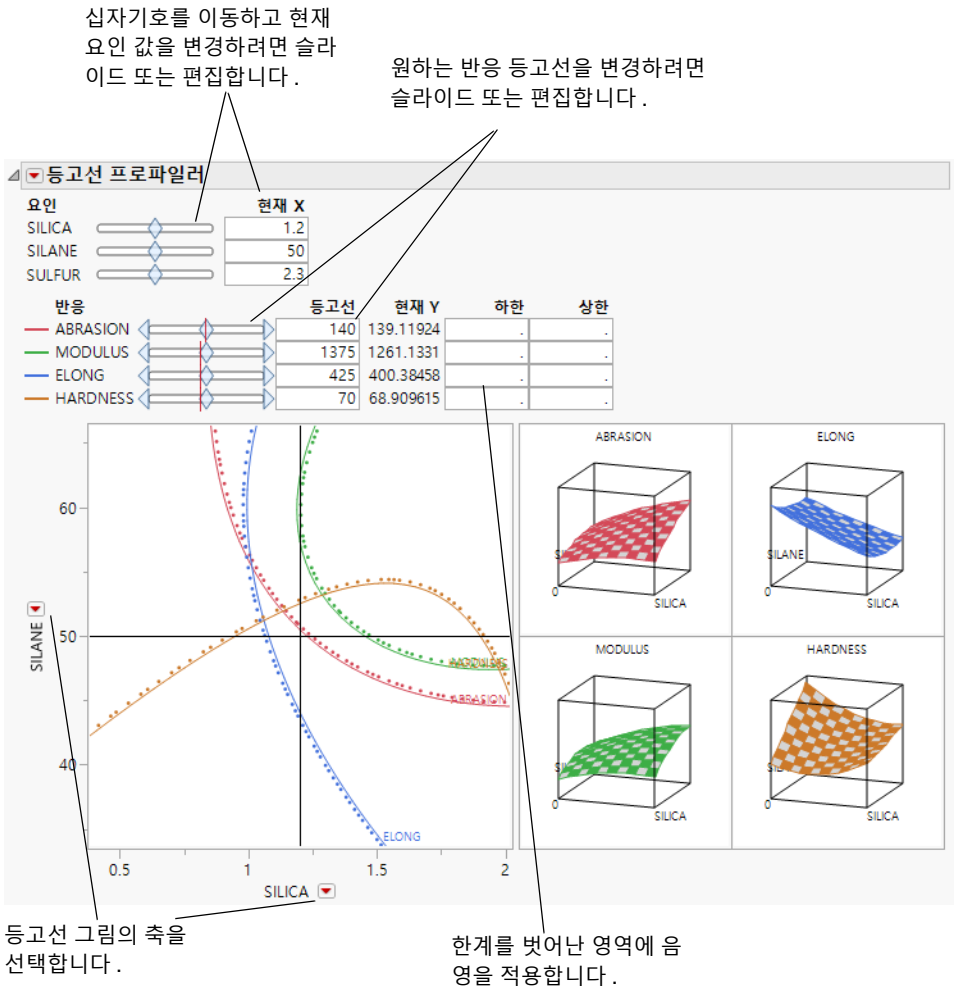
등고선 프로파일러

참고 : 등고선 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

반응 표면을 그래픽으로 최적화하려면 대화식 등고선 프로파일러를 사용합니다. 등고선 프로파일러에는 적합 모형에 대해 한 번에 두 요인의 등고선이 표시됩니다. 표면 그림도 보고서에 포함됩니다.

그림 3.50에서는 Tiretread.jmp 샘플 데이터 테이블에 대한 등고선 프로파일러 보기를 보여 줍니다. **RSM for 4 Responses** 데이터 테이블 스크립트를 실행하고 "최소 제곱 적합" 보고서 메뉴에서 **프로파일러 > 등고선 프로파일러**를 선택합니다.

그림 3.50 등고선 프로파일러



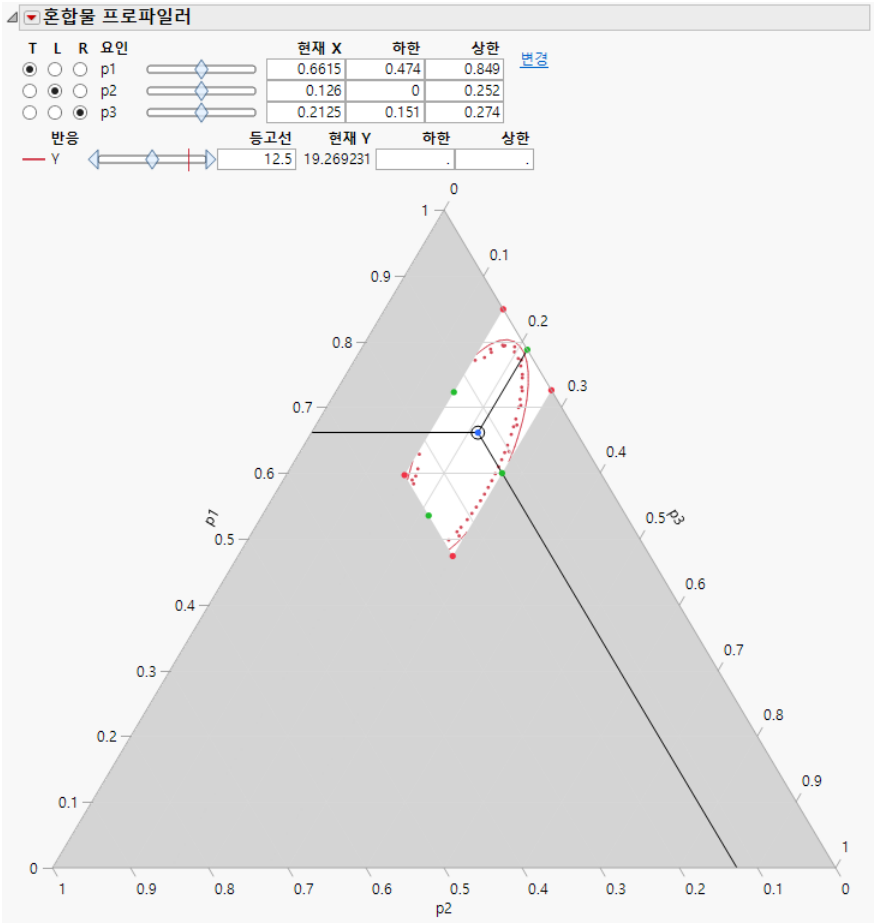
혼합물 프로파일러

참고 : 이 옵션은 효과에 대해 **매크로 > 혼합물 반응 표면** 옵션을 지정한 경우에만 나타납니다. 혼합물 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오.

혼합물 프로파일러에는 혼합물 실험 모형의 반응 등고선이 삼원 그림에 표시됩니다. 실험에서 세 개 이상의 요인이 혼합물 성분인 경우 혼합물 프로파일러를 사용합니다. 혼합물 프로파일러를 사용하여 실험의 반응 표면을 시각화하고 최적화할 수 있습니다.

그림 3.51 에서는 Plasticizer.jmp 샘플 데이터 테이블의 모형에 대한 혼합물 프로파일러를 보여 줍니다. **Model** 데이터 테이블 스크립트를 실행한 후 보고서의 빨간색 삼각형 메뉴에서 **요인 프로파일링 > 혼합물 프로파일러**를 선택합니다. 그림의 왼쪽 위에 있는 여러 라디오 버튼을 선택하여 요인에 대한 그림 축을 수정합니다. 그림의 오른쪽 위에 있는 "하한" 및 "상한" 열을 사용하면 요인과 반응 모두에 대한 제약 조건을 입력할 수 있습니다.

그림 3.51 혼합물 프로파일러



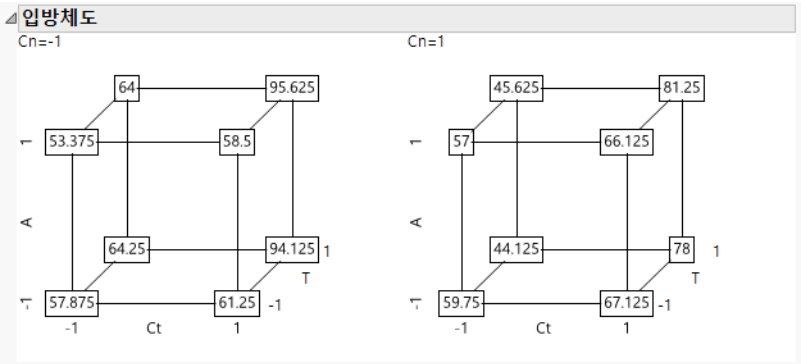
입방체도

"입방체도" 옵션은 요인 범위의 극단에 대한 예측값을 표시합니다. 이러한 값은 입방체의 꼭지점에 나타납니다(그림 3.52). 꼭지점은 요인의 최소 및 최대 관측값으로 정의됩니다. 반응이 여러 개 있으면 여러 반응이 각 꼭지점에 누적되어 표시됩니다.

입방체도의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. **차수** 상자에 2 가 있는지 확인합니다 .
5. Ct, A, T 및 Cn 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 입방체도**를 선택합니다 .

그림 3.52 입방체도



Cn = –1 에 대한 입방체와 Cn = 1 에 대한 입방체가 하나씩 있습니다. 요인이 다른 입방체 좌표에 매핑되도록 레이아웃을 변경하려면 첫 번째 입방체에서 요인 이름을 클릭합니다. 그런 다음 원하는 축의 요인 이름을 포함하도록 드래그합니다. 예를 들어 그림 3.52 에서 T 를 클릭하여 Ct 위로 드래그하면 T 와 Ct 및 해당 좌표의 위치가 서로 바뀝니다. 단일 입방체에서 Cn 수준을 보려면 첫 번째 입방체에서 해당 요인 위로 드래그하여 다른 요인과 교환합니다.

Y 의 Box-Cox 변환

Box-Cox 멱 변환은 분산의 정규성 및 동질성에 대한 일반적인 회귀 가정이 더 상세하게 충족되도록 반응을 변환하는 데 사용됩니다. 그런 다음 회귀 모형을 사용하여 변환된 반응을 적합시킬 수 있습니다. 그러나 다른 이유로 변수를 변환할 때도 Box-Cox 멱 변환을 사용할 수 있습니다. 이 변환은 반응 Y 가 순양수인 경우에만 적합합니다.

일반적으로 사용되는 변환은 반응을 지정된 횟수만큼 거듭제곱합니다 . Box and Cox(1964) 에서는 이러한 멱 변환 계열을 공식화하고 기술했습니다 . 변환 계산식은 모수 λ 로 표현된 연속 정의

를 제공하고 오차 제곱합을 비교할 수 있도록 구성되었습니다. 구체적으로, 다음 방정식을 사용하여 변환 계열을 제공합니다.

$$Y_\lambda = \begin{cases} \dfrac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \text{단, } \lambda \neq 0 \\ \dot{y}\ln(y) & \text{단, } \lambda = 0 \end{cases}$$

여기서 $\dot{y}$ 는 기하평균을 나타냅니다.

"Y 의 Box-Cox 변환" 옵션은 λ = –2 에서 2 까지 0.2 단위로 증가하여 변환을 적합시킵니다. λ 값을 선택하기 위해 이러한 각 변환에 대한 가능도 함수가 계산됩니다. 이 함수는 오차가 독립적이며 평균이 0, 분산이 σ^2 인 정규 분포를 따른다고 가정하여 계산됩니다. 가능도를 최대화하는 λ 값이 선택됩니다. 또한 이 값은 λ 값에 대해 SSE 를 최소화합니다. SSE 를 최소화하는 λ 값은 SSE 가 가장 작은 격자 점을 둘러싼 두 증분 격자점 사이의 2 차 보간을 사용하여 구합니다. 이 보간법으로 인해 SSE 값이 음수가 되면 SSE 를 최소화하는 λ 의 격자 값이 최선의 λ 로 보고됩니다.

"Box-Cox 변환" 보고서에는 λ 값에 대한 SSE(오차 제곱합) 값을 보여 주는 그림이 표시됩니다. 그림의 빨간색 수평선은 λ 에 대한 단측 95% 신뢰 구간을 나타냅니다. 이 신뢰 구간은 Box and Cox(1964, p. 216) 에 정의된 신뢰 영역을 기반으로 합니다. 신뢰 영역은 다음 부등식으로 정의됩니다.

$$SSE(\lambda) < SSE(\lambda_{best}) * \exp(ChiSquareQuantile(0.95,1) / dfe)$$

다음은 각 요소에 대한 설명입니다.

$SSE(\lambda_{best})$ 는 보고된 최선의 λ 를 사용하여 계산된 SSE 입니다.

ChiSquareQuantile(0.95,1) 은 자유도가 1 인 χ^2 분포의 0.95 번째 분위수입니다.

dfe 는 회귀 모형에 대한 분산 분석 테이블의 오차 자유도입니다.

"Box-Cox 변환" 보고서에서는 다음 옵션을 제공합니다.

변환을 사용한 재적합 변환된 Y 변수를 정의하기 위한 람다 값을 지정한 다음, 변환된 변수에 대한 최소 제곱 적합을 제공할 수 있습니다.

변환으로 바꾸기 변환된 Y 변수를 정의하기 위한 람다 값을 지정한 다음, 기존 최소 제곱 적합을 변환된 변수에 대한 적합으로 바꿀 수 있습니다. 반응이 여러 개인 경우 "변환으로 바꾸기" 는 변환 대상 반응에 대한 보고서만 바꿉니다.

최적 변환 저장 데이터 테이블에 새 열을 생성하고 최적 변환을 위한 계산식을 저장합니다.

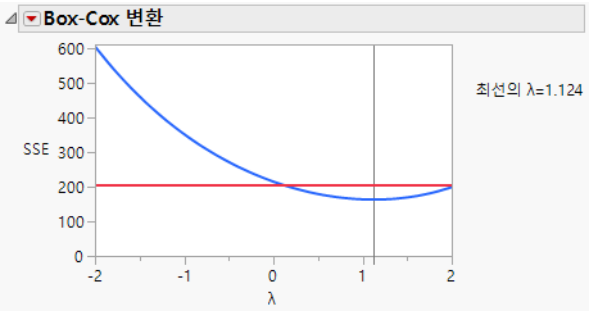
특정 변환 저장 람다 값을 지정하고, 지정된 변환을 위한 계산식이 포함된 열을 데이터 테이블에 생성할 수 있습니다.

추정값 테이블 –2 에서 2 까지 0.2 단위로 증가하는 모든 λ 에 대한 모수 추정값 및 SSE 값을 포함하는 새 데이터 테이블을 생성합니다.

Y 의 Box-Cox 변환의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. **차수** 상자에 2 가 있는지 확인합니다 .
5. F, Ct, A, T 및 Cn 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > Y 의 Box-Cox 변환**을 선택합니다 .

그림 3.53 Y 의 Box-Cox 변환



이 그림에서는 최선의 λ 값이 0.1 과 2.0 사이임을 보여 줍니다 . JMP 가 λ 값의 0.2 단위 격자에서 최적 값 두 개 사이의 보간을 사용하여 선택한 값은 1.124 입니다 .

8. (선택 사항) 그래프를 생성하는 데 사용된 SSE 값을 보려면 **추정값 테이블**을 선택합니다 .

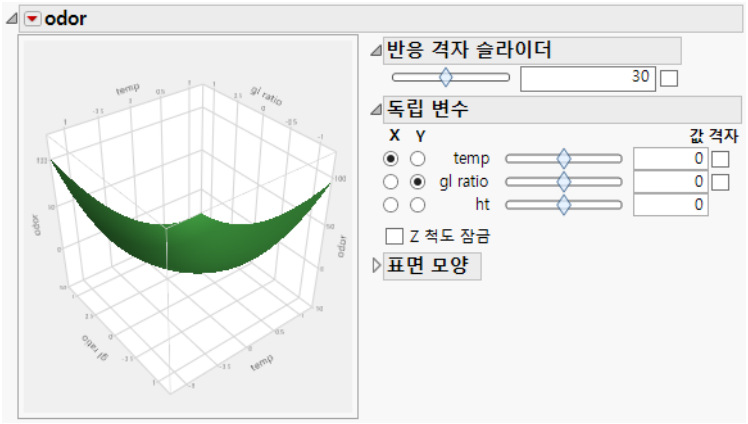
표면 프로파일러

참고 : 표면 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

표면 프로파일러에는 반응 표면의 3 차원 표면 그림이 표시됩니다.

그림 3.54 에서는 Odor.jmp 샘플 데이터 테이블의 모형에 대한 표면 프로파일러를 보여 줍니다. **Model** 데이터 테이블 스크립트를 실행한 후 보고서의 빨간색 삼각형 메뉴에서 **요인 프로파일링 > 표면 프로파일러**를 선택합니다. "독립 변수" 아래의 라디오 버튼을 사용하여 축의 변수를 변경할 수 있습니다. 또한 "표면 모양" 아래에서 "실제값"을 클릭하여 그림에 점을 표시할 수도 있습니다.

그림 3.54 표면 그림



행 진단

행 진단 메뉴는 행 또는 관측값과 관련된 문제를 해결합니다 .

회귀 그림 범주형 효과의 각 수준에 대한 데이터 산점도와 회귀선을 보여 주는 " 회귀 그림 " 보고서를 표시합니다 .

이 옵션은 모형에 연속형 효과가 정확히 하나만 있고 범주형 효과가 최대 한 개인 경우에만 나타납니다 . 이 경우 " 회귀 그림 " 보고서가 기본적으로 제공됩니다 .

실제값 대 예측값 그림 Y 의 예측값에 대한 Y 의 관측값을 보여 주는 " 실제값 대 예측값 그림 " 을 표시합니다 . 자세한 내용은 " 레버리지 그림 "(161 페이지) 에서 확인하십시오 .

참고 : " 모형 적합 " 시작 창의 " 강조 " 에서 " 효과 레버리지 " 또는 " 효과 선별 " 을 선택하고 R^2 값이 0.999 미만인 경우 " 실제값 대 예측값 그림 " 이 기본적으로 표시됩니다 .

효과 레버리지 그림 모형의 각 효과에 대한 " 레버리지 그림 " 보고서를 표시합니다 . 이 그림에서는 관측값이 해당 효과에 대한 검정에 어떤 영향을 미치는지 보여 주고 다중공선성에 대한 정보를 제공합니다 . 자세한 내용은 " 레버리지 그림 "(161 페이지) 에서 확인하십시오 .

참고 : " 모형 적합 " 시작 창의 " 강조 " 에서 " 효과 레버리지 " 를 선택하고 R^2 값이 0.999 미만인 경우 효과 레버리지 그림이 기본적으로 표시됩니다 . 이 그림은 " 전체 모형 " 보고서의 오른쪽에 나타납니다 . " 강조 " 에서 다른 옵션을 선택하면 " 효과 상세 정보 " 보고서에 효과 레버리지 그림이 나타납니다 . 모든 경우에 효과 레버리지 그림을 표시하려면 회귀 보고서 > 효과 상세 정보 옵션을 선택해야 합니다 .

잔차 대 예측값 그림 " 잔차 대 예측값 그림 " 보고서를 표시합니다 . 이 그림에서는 Y 의 예측값에 대해 그림에 표시된 잔차를 보여 줍니다 . 일반적으로 0 주위에 무작위로 흩어진 잔차 값을 원합니다 .

참고 : " 모형 적합 " 시작 창의 " 강조 " 에서 " 효과 레버리지 " 또는 " 효과 선별 " 을 선택하고 R^2 값이 0.999 미만인 경우 " 잔차 대 예측값 그림 " 이 기본적으로 표시됩니다 .

잔차 대 행 번호그림 " 잔차 대 행 번호 그림 " 보고서를 표시합니다 . 행 번호에 대한 잔차 값이 그림에 표시됩니다 . 이 그림은 관측값의 행 순서로 인한 패턴을 탐지하는 데 도움이 될 수 있습니다 .

스튜던트화 잔차 그림 " 스튜던트화 잔차 " 그림을 표시합니다 . 그림의 각 점은 현재 관측값이 삭제된 상태에서 구한 표준편차 추정값을 사용하여 계산됩니다. 이러한 잔차를 RStudent 또는 외부 스튜던트화 잔차라고도 합니다 .

이 그림은 다음과 같은 두 가지 한계 집합을 포함합니다 .

- 그림에 빨간색으로 나타나는 외부 한계는 95% Bonferroni 한계입니다 . 이러한 한계는 ± tQuantile(0.025/*n*, *n*–*p*–1) 에 배치됩니다 . 여기서 *n* 은 관측값 수이고 *p* 는 예측 변수의 수입니다 .
- 그림에 녹색으로 나타나는 내부 한계는 95% 개별 *t* 분포 한계입니다 . 이러한 한계는 ± tQuantile(0.025, *n*–*p*–1) 에 배치됩니다 . 여기서 *n* 은 관측값 수이고 *p* 는 예측 변수의 수입니다 .

빨간색 한계를 벗어나는 점은 가능성 있는 이상치로 처리해야 합니다. 녹색 한계를 벗어나지만 빨간색 한계 내에 있는 점은 가능하지만 확신도가 적은 이상치로 처리해야 합니다 . 이러한 한계에 대한 95% 신뢰 수준은 " 모형 규격 " 창의 " 유의 수준 설정 " 옵션에서 선택한 항목의 영향을 받지 않습니다 .

주의 : 열 저장 > 스튜던트화 잔차를 사용하여 저장된 잔차는 외부 스튜던트화 잔차가 아닙니다 .

참고 : 모형에 임의 효과가 포함되고 시작 창에서 " 방법 " 에 "REML" 을 지정한 경우 , " 스튜던트화 잔차 " 그림에 한계가 포함되지 않으며 그림에 표시된 점은 외부 스튜던트화되지 않습니다 .

잔차 대 정규 분위수 그림 (시작 창에서 " 방법 " 에 "REML" 을 지정한 경우에는 사용 불가능) " 잔차 정규 분위수 그림 " 을 표시합니다 . 정규 분포의 분위수에 대한 잔차 값이 그림에 표시됩니다 . 이 그림은 잔차의 정규성 가정을 평가하는 데 도움이 될 수 있습니다 .

Press Press 통계량 및 RMSE(제곱근 평균 제곱 오차) 를 제공하는 "Press" 보고서를 표시합니다 . Press 통계량은 여러 모형을 비교할 때 유용합니다 . Press 통계량이 낮은 모형이 선호됩니다 . 자세한 내용은 "Press"(164 페이지) 에서 확인하십시오 .

Durbin-Watson 검정 (빈도 열을 지정한 경우에는 사용 불가능) 잔차에 1 차 자기상관이 있는지 여부를 검정하기 위한 통계량을 제공하는 "Durbin-Watson" 보고서를 표시합니다 . 또한 잔차와 확률 <DW(통계량과 관련된 정확 확률) 의 자기상관도 보고서에 표시됩니다 . 이 옵션은 시계열 데이터에만 사용할 수 있으며 관측값이 시간순으로 정렬되어 있다고 가정합니다 .

레버리지 그림

X 에 대한 효과 레버리지 그림은 다음과 같은 방식으로 유용합니다 .

- X 에 대한 가설 검정에 영향을 미칠 수 있는 점을 확인할 수 있습니다 .
- 일반적이지 않은 패턴과 모형 가정 위반을 발견할 수 있습니다 .
- 다중공선성 문제를 발견할 수 있습니다 .

생성

효과에 대한 레버리지 그림에서는 다른 효과가 이미 모형에 있는 경우 이 효과를 모형에 추가할 때 미치는 영향을 보여 줍니다. 설명을 위해 단일 연속형 효과 X에 대한 효과 레버리지 그림을 생성하는 경우를 가정해 보겠습니다. 더 일반적인 상황에서 가로 축 척도에 대한 자세한 내용은 "가로 축 척도화"(162 페이지)에서 확인하십시오.

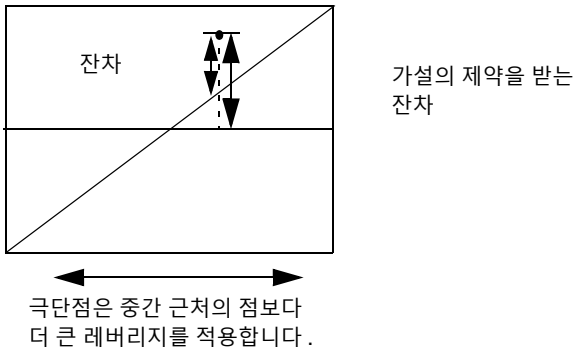
X를 제외한 모든 예측 변수에 대해 반응 Y가 회귀 분석되고 잔차가 계산됩니다. 이러한 잔차를 Y 잔차라고 합니다. 그런 다음 모형의 다른 모든 예측 변수에 대해 X가 회귀 분석되고 잔차가 계산됩니다. 이러한 잔차를 X 잔차라고 합니다. X 잔차에는 모형에서 X를 제외하고 계산된 Y 잔차에 있는 정보 외에 추가 정보가 포함될 수 있습니다.

기본적으로 X에 대한 효과 레버리지 그림은 Y 잔차에 대한 X 잔차의 산점도입니다(그림 3.57). JMP는 그림을 해석하고 다른 그림과 비교하는 데 도움이 되도록 Y의 평균을 Y 잔차에 추가하고 X의 평균을 X 잔차에 추가합니다. 변환된 Y 잔차를 Y 레버리지 잔차라고 하고 변환된 X 잔차를 X 레버리지 값이라고 합니다. 효과 레버리지 그림의 점은 이러한 X 레버리지와 Y 레버리지 잔차 쌍입니다.

JMP는 이러한 점과 평균에 대한 신뢰 대역에 최소 제곱선을 적합시킵니다. 적합선은 빨간색 실선이고 신뢰 대역은 빨간색으로 음영 처리됩니다. 최소 제곱선의 기울기는 Y가 X 및 다른 예측 변수에 대해 회귀 분석되는 모형에서 X에 대한 계수의 추정값입니다. 파란색 수평 파선은 Y 레버리지 잔차의 평균에 설정됩니다. 이 선은 X 잔차가 Y 잔차와 선형 관계가 없는 상황을 설명합니다. 적합선의 기울기가 0이 아닌 경우 모형에 X를 추가하면 변동을 설명하는 데 유용할 수 있습니다.

그림 3.55에서는 잔차가 레버리지 그림에 어떻게 표시되는지 보여 줍니다. 점에서 적합선까지의 거리는 효과를 포함하는 모형의 잔차입니다. 점에서 수평선까지의 거리는 모형에 효과가 없을 때의 잔차 오차입니다. 즉, 레버리지 그림의 평균 선은 모수(효과)의 가설 값이 0으로 제한된 모형을 나타냅니다.

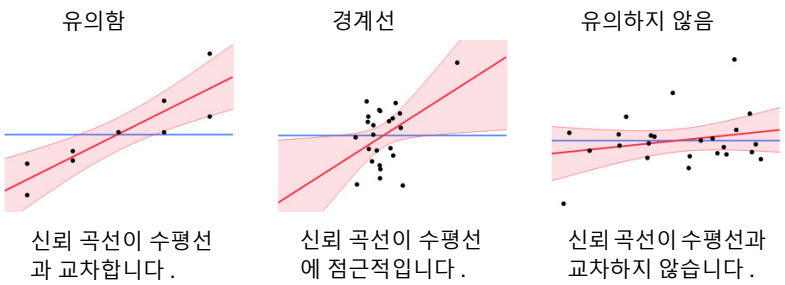
그림 3.55 일반 레버리지 그림의 예



신뢰 곡선

적합선에 대한 신뢰 곡선이 레버리지 그림에 표시됩니다. 이러한 곡선은 관심 검정이 5% 수준(또는 "모형 적합" 시작 창의 "유의 수준 설정"에 지정한 수준)에서 유의한지 여부를 시각적으로 보여 줍니다. 곡선 사이의 신뢰 영역에 가설을 나타내는 수평선이 포함되어 있으면 효과가 유의하지 않습니다. 곡선이 선과 교차하면 효과가 유의합니다. 그림 3.56의 예를 참조하십시오.

그림 3.56 레버리지 그림에 표시된 유의성 비교



가로 축 척도화

예측 변수 X의 모델링 유형이 연속형이면 가로 축이 X 단위로 척도화됩니다. 가로 축 범위는 X 값 범위를 반영합니다. 레버리지 그림에서 적합선 기울기는 X에 대한 모수 추정값입니다. 그림 3.57의 왼쪽 그림을 참조하십시오.

효과가 명목형 또는 순서형이거나, 교호작용과 같은 복합 효과인 경우 가로 축에 효과의 값을 직접 나타낼 수 없습니다. 이 경우 가로 축은 반응 단위로 척도화되고 적합선은 기울기가 1인 대각선입니다. "전체 모형" 레버리지 그림(모든 모수 값이 0이라는 가설이 지정됨)에는 이 척도화가 사용됩니다. 자세한 내용은 "레버리지 그림 상세 정보"(189 페이지)에서 확인하십시오. 이 그림의 경우 가로 축은 그림 3.57의 오른쪽 그림에 표시된 것처럼 전체 모형에 대한 예측 반응 값을 기준으로 척도화됩니다.

단순 회귀 분석에서 선형 효과에 대한 레버리지 그림은 기존의 예측 변수에 대한 실제 반응 값 그림과 동일합니다.

레버리지

이러한 그림은 모형에 효과를 포함하기 위한 검정에서 점의 영향을 시각화하는 데 도움이 되므로 **레버리지**라는 용어가 사용됩니다. 그림의 중심에서 수평으로 멀리 떨어진 점은 중심에 가까운 점보다 효과 검정에 더 많은 영향을 미칩니다. 효과 검정에는 잔차 제곱합을 해당 효과가 제거된 모형의 잔차 제곱합과 비교하는 과정이 포함됩니다. 극단에서는 가설의 제약을 받기 전과 후의 잔차 차이가 비교적 큰 경향이 있습니다. 따라서 이러한 잔차는 해당 효과의 가설 검정에 대한 제곱합에 더 크게 기여합니다.

다중공선성

다중공선성은 두 개 이상의 예측 변수가 밀접하게 관련되어 있거나, 좀 더 엄밀히 말하면 거의 선형 종속 관계에 관련된 상태를 의미합니다. 다중공선성이 존재하면 표준 오차가 팽창되고 모수 추정값이 불안정해질 수 있습니다. 효과가 다른 예측 변수와 동일 선상에 있으면 점의 가로 값이 그림의 가운데 쪽으로 모이는 경향이 있습니다. 이 상황은 적합선의 기울기가 불안정함을 나타냅니다.

전체 모형 실제값 대 예측값 그림

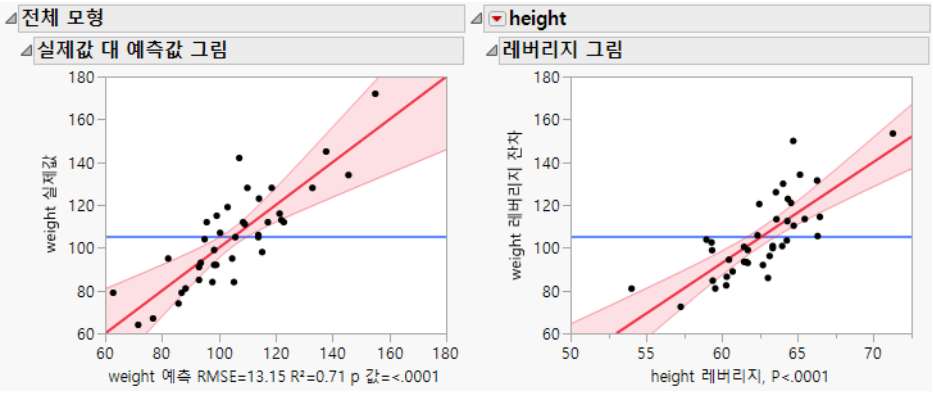
"효과 레버리지 그림" 옵션은 모형의 각 효과에 대한 레버리지 그림을 생성합니다. 실제값 대 예측값 그림도 레버리지 그림으로 간주할 수 있습니다. 이 그림을 사용하면 모형의 모든 모수(절편 제외)가 0이라는 검정을 시각화할 수 있습니다. 동일한 검정이 "분산 분석" 보고서에서 분석적으로 수행됩니다. 이 그림에 대한 자세한 내용은 "레버리지 그림 상세 정보"(189 페이지)에서 확인하십시오.

선형 효과에 대한 레버리지 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. weight 를 선택하고 **Y** 를 클릭합니다 .
4. height, age 및 sex 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .

그림 3.57에서는 전체 모형에 대한 "실제값 대 예측값 그림"과 height 효과에 대한 "레버리지 그림"을 보여 줍니다. 왼쪽의 전체 모형 그림에서는 모든 효과를 검정합니다. 신뢰 곡선이 반응(weight)의 평균에서 수평선과 교차하므로 모형이 유의하다는 것을 추론할 수 있습니다. 또한 오른쪽의 height에 대한 "레버리지 그림"에서는 모형에 age와 sex가 있더라도 height가 유의함을 보여 줍니다. 두 그림 모두 영향력 있는 점 또는 다중공선성과 관련된 문제를 나타내지 않습니다.

그림 3.57 전체 모형 및 효과 레버리지 그림



Press

Press 또는 **예측 오차 제곱합** 통계량은 leave-one-out 교차 검증을 사용하여 계산된 예측 오차의 추정값입니다. leave-one-out 교차 검증에서는 각 관측값이 차례로 제거됩니다. 특정 관측값을 예로 들겠습니다. 관측값을 주지 않고 모형을 적합시킨 후 이 관측값에 대한 예측값을 계산합니다. 해당 관측값에 대한 잔차가 계산됩니다. 이 절차가 모든 관측값에 적용되며 잔차의 제곱합이 계산되어 Press 값을 제공합니다.

Press 통계량은 다음과 같이 구합니다.

$$\text{Press} = \sum_{i=1}^{n} (\hat{y}_{(i)} - y_i)^2$$

여기서 n은 관측값 수, y_i는 i번째 관측값의 반응 관측값, $\hat{y}_{(i)}$는 i번째 관측값의 반응 예측값입니다. 이러한 값은 해당 관측값을 포함하지 않은 모형 적합을 기반으로 합니다.

Press RMSE는 $\sqrt{Press/n}$으로 정의됩니다.

Press R^2은 1 - $Press/SS_{Total}$로 정의됩니다.

열 저장

각 " 열 저장 " 옵션은 하나 이상의 새 열을 현재 데이터 테이블에 추가합니다 . 적합 모형이 REML 인 경우 추가 " 열 저장 " 옵션이 나타납니다 . 자세한 내용은 "REML 열 저장 옵션 "(180 페이지) 에서 확인하십시오 .

다음 사항에 유의하십시오 .

- 계산식이 생성되면 " 계산식 " 열 특성으로 입력됩니다 .
- 대부분의 새 열에는 해당 열을 설명하고 최소 제곱 적합으로 열이 생성되었음을 나타내는 " 노트 " 열 특성이 추가됩니다 .
- " 예측 계산식 " 및 " 예측값 " 옵션의 경우 " 예측 " 열 특성이 추가됩니다 . 이 특성은 모형 비교 (분석 > 예측 모델링 > 모형 비교) 를 수행할 때 JMP 에서 내부적으로 사용됩니다 . 또한 열의 출처를 기록하므로 여러 모형을 적합시키는 경우에도 유용합니다 .

" 열 저장 " 에서 다음 옵션을 사용할 수 있습니다 .

예측 계산식 계산식과 예측값이 모두 포함된 **예측 계산식 < 열 이름 >** 이라는 새 열을 생성합니다 . " 예측 " 열 특성이 추가되어 예측 소스를 나타냅니다 .

> **참고 :** **예측 계산식 < 열 이름 >** 은 **< 열 이름 >** 에서 특정한 특성을 상속합니다 . 여기에는 반응 한계 , 규격 한계 및 관리 한계가 포함됩니다 . **예측 계산식 < 열 이름 >** 을 저장한 후에는 **< 열 이름 >** 에 대한 이러한 특성을 변경해도 **예측 계산식 < 열 이름 >** 에서 업데이트되지 않습니다 .

자세한 내용은 " 예측 계산식 "(167 페이지) 에서 확인하십시오 .

예측값 지정된 모형에서 계산한 예측값이 포함된 **예측값 < 열 이름 >** 이라는 새 열을 생성합니다 . " 노트 " 와 " 예측 " 열 특성이 모두 추가되어 예측 소스를 나타냅니다 .

> **참고 :** **예측값 < 열 이름 >** 은 **< 열 이름 >** 에서 특정한 특성을 상속합니다 . 여기에는 반응 한계 , 규격 한계 및 관리 한계가 포함됩니다 . **예측값 < 열 이름 >** 을 저장한 후에는 **< 열 이름 >** 에 대한 이러한 특성을 변경해도 **예측값 < 열 이름 >** 에서 업데이트되지 않습니다 .

잔차 반응 관측값에서 예측값을 뺀 값이 포함된 **잔차 < 열 이름 >** 이라는 새 열을 생성합니다 .

평균 신뢰 구간 **평균 < 열 이름 > 의 95% 하한** 및 **평균 < 열 이름 > 의 95% 상한**이라는 두 개의 새 열을 생성합니다 . 이 열에는 평균 반응에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다 .

> **참고 :** Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다 .

개별값 신뢰 구간 **개별 < 열 이름 > 의 95% 하한**과 **개별 < 열 이름 > 의 95% 상한**이라는 두 개의 새 열을 생성합니다 . 이 열에는 개별 반응 값에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다 .

참고 : Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다 .

스튜던트화 잔차 잔차를 표준 오차로 나눈 값이 포함된 스튜던트화 잔차 < 열 이름 > 이라는 새 열을 생성합니다 .

외부 스튜던트화 잔차 (적합 방법이 REML 인 경우에는 사용 불가능) 잔차를 현재 행을 제외하는 표준 오차 추정값으로 나눈 값이 포함된 외부 스튜던트화 잔차 < 열 이름 > 이라는 새 열을 생성합니다 . 자세한 내용은 " 스튜던트화 잔차 그림 "(160 페이지) 에서 확인하십시오 .

해트 값 h < 열 이름 > 이라는 새 열을 생성합니다 . 열 값은 $X(X'X)^{-1}X'$ 행렬의 대각 값이며 hat 값이라고도 합니다 .

예측값의 표준 오차 예측 평균 반응의 표준 오차가 포함된 예측 표준 오차 < 열 이름 > 이라는 새 열을 생성합니다 .

잔차의 표준 오차 잔차 값의 표준 오차가 포함된 표준 오차 잔차 < 열 이름 > 이라는 새 열을 생성합니다 .

개별값의 표준 오차 개별 예측 반응 값의 표준 오차가 포함된 표준 오차 개별 < 열 이름 > 이라는 새 열을 생성합니다 .

효과 레버리지 쌍 각 레버리지 그림에 대한 X 레버리지 값과 Y 레버리지 잔차가 포함된 새 열 집합을 생성합니다 . 모형의 각 효과에 대해 두 개의 열이 추가됩니다 . 반응 열 이름이 R 이고 효과가 X 인 경우 새 열 이름은 다음과 같습니다 .

- R 에 대한 X 의 X 레버리지
- R 에 대한 X 의 Y 레버리지

열 패널에서 이러한 열은 레버리지라는 열 그룹으로 구성됩니다 .

Cook 거리 영향력 Cook 거리 영향력 통계량의 값이 포함된 Cook 거리 영향력 < 열 이름 > 이라는 새 열을 생성합니다 .

예측 표준 오차 계산식 예측값의 표준 오차에 대한 계산식과 값이 모두 포함된 예측값 표준 오차 < 열 이름 > 이라는 새 열을 생성합니다 .

참고 : 저장된 계산식이 클 수 있습니다 . 계산식이 필요하지 않으면 " 예측값의 표준 오차 " 옵션을 사용합니다 .

평균 신뢰 한계 계산식 평균 < 열 이름 > 의 95% 하한 및 평균 < 열 이름 > 의 95% 상한이라는 두 개의 새 열을 생성합니다 . 이 열은 평균 반응에 대한 95% 신뢰 하한 및 상한의 계산식과 값이 모두 포함됩니다 .

참고 : Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다 .

개별값 신뢰 한계 계산식 개별 < 열 이름 > 의 95% 하한과 개별 < 열 이름 > 의 95% 상한이라는 두 개의 새 열을 생성합니다 . 이 열은 개별 반응 값에 대한 95% 신뢰 하한 및 상한의 계산식과 값이 모두 포함됩니다 .

참고 : Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다 .

코딩 테이블 저장 첫 번째 열에 모든 모형 모수에 대한 JMP 코딩이 표시되는 새 데이터 테이블을 생성합니다 . 마지막 열은 반응 값을 제공합니다 . 반응 열을 두 개 이상 입력한 경우 이러한 모든 열이 코딩 테이블의 마지막 열로 나타납니다 .

참고 : 코딩 데이터 테이블에는 분석에 사용된 데이터 테이블 이름을 제공하는 " 원래 데이터 " 라는 테이블 변수가 포함되어 있습니다 . 기준 변수가 지정된 경우에는 " 원래 데이터 " 테이블 변수에 기준 변수와 해당 수준이 제공됩니다 .

JMP PRO **예측 계산식 게시** 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

JMP PRO **표준 오차 계산식 게시** 표준 오차 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

JMP PRO **평균 신뢰 한계 계산식 게시** 평균 반응에 대한 신뢰 한계 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

JMP PRO **개별 신뢰 한계 계산식 게시** 개별 반응 값에 대한 신뢰 한계 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

예측 계산식

예측 계산식 <열 이름>은 예측 계산식을 포함하므로 예측값 <열 이름>과 다릅니다. 예측 계산식을 보려면 예측 계산식 <열 이름> 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택합니다. 모형이 크면 예측 계산식에 상당한 공간이 필요할 수 있습니다. 예측값 열에 계산식이 필요하지 않으면 **열 저장 > 예측값** 옵션을 사용합니다. 계산식에 대한 자세한 내용은 JMP 사용 의 에서 확인하십시오.

"예측 계산식" 옵션은 새 행의 값을 예측하거나 프로파일러와 함께 사용하는 데 유용합니다. 프로파일러는 "최소 제곱 적합" 보고서(요인 프로파일링)에서 사용할 수 있습니다. 그러나 데이터 테이블에 계산식 열이 포함된 경우 **그래프** 메뉴에 제공된 프로파일러를 사용할 수도 있습니다. 여러 반응을 분석하는 경우 **그래프** 메뉴에서 프로파일러에 액세스하는 것이 유용할 수 있습니다.

참고 : 그래프 > 프로파일러를 선택하여 프로파일러에 액세스하는 경우 먼저 " 예측 계산식 " 및 " 예측 표준 오차 계산식 " 을 사용하여 계산식 열을 데이터 테이블에 저장하십시오 . 그런 다음 이 두 계산식을 " 프로파일러 " 창의 "Y, 예측 계산식 " 역할에 배치합니다 . **확인**을 클릭한 후 " 예측값 표준 오차 < **열 이름** >" 을 사용하여 " 예측 계산식 < **열 이름** >" 에 대한 신뢰 구간을 생성할지 여부를 지정합니다 . 그렇지 않으면 JMP 에서 " 예측값 표준 오차 < **열 이름** >" 에 대한 별도의 프로파일러 그림을 생성합니다 .

효과 요약 보고서

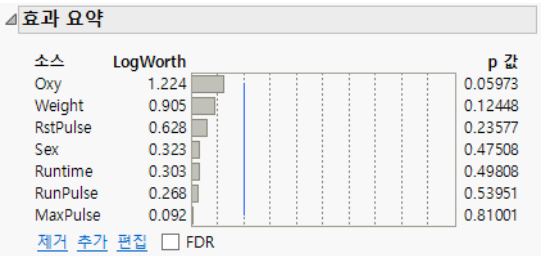
" 효과 요약 " 옵션은 대화식 보고서를 표시합니다. 이 보고서에는 모형의 효과에 대한 LogWorth 또는 FDR LogWorth 값 그림이 제공됩니다. 모형에서 효과를 추가하거나 제거할 수 있는 컨트롤도 보고서에 제공됩니다. 모형 적합 보고서는 " 효과 요약 " 보고서의 변경 사항에 따라 자동으로 업데이트됩니다.

" 효과 요약 " 보고서는 다음 분석법에서 사용할 수 있습니다 .

- 표준 최소 제곱
- 명목형 로지스틱
- 순서형 로지스틱
- 비례 위험
- 모수 생존
- 일반화 선형 모형

그림 3.58 에서는 Fitness.jmp 데이터 테이블에 대한 " 효과 요약 " 보고서의 초기 보기를 표시합니다. **FDR** 라벨이 지정된 체크박스는 요약 테이블에 나타나는 열을 제어합니다.

그림 3.58 효과 요약 보고서



효과 요약 테이블 열

" 효과 요약 " 테이블에는 다음과 같은 열이 포함되어 있습니다 .

소스 모형 효과를 *p* 값이 작은 것부터 순서대로 나열합니다 .

LogWorth 각 모형 효과에 대해 -$\log_{10}$(p 값) 으로 정의된 LogWorth 를 표시합니다 . 이 변환은 그래프 작성에 적절한 척도를 제공하도록 p 값을 조정합니다 . –$\log_{10}$(0.01) = 2 이므로 2 를 초과하는 값은 0.01 수준에서 유의합니다 .

FDR LogWorth 각 모형 효과에 대해 -$\log_{10}$(FDR p 값) 으로 정의된 FDR(False Discovery Rate) LogWorth 를 표시합니다 . 이 통계량은 유의성을 그림에 표시하고 평가하는 데 가장 적합합니다 . 그러나 유의성 순서에 따라 크게 좌우되고 , 양의 상관관계가 있는 검정에 대해 보수적이며 , 유의 수준에서 실험별 보호를 제공하지 않습니다 . LogWorth 열을 **FDR LogWorth** 열로 바꾸려면 **FDR** 체크박스를 선택합니다 .

막대 그래프 LogWorth 또는 FDR LogWorth 값의 막대 그래프를 표시합니다 . 이 그래프에는 정수 값에 수직 파선이 있고 2 에 파란색 참조선이 있습니다 .

p 값 각 모형 효과에 대한 p 값을 표시합니다 . 일반적으로 이 값은 모형 보고서의 " 효과 검정 " 테이블 또는 " 효과 가능도비 검정 " 테이블에 표시된 유의성 검정에 해당하는 p 값입니다 .

FDR p 값 각 모형 효과에 대해 Benjamini-Hochberg 기법을 사용하여 계산된 FDR(False Discovery Rate) p 값을 표시합니다 . 이 기법은 여러 검정에 대한 FDR(False Discovery Rate) 을 제어하기 위해 p 값을 조정합니다 . **p 값** 열을 **FDR p 값** 열로 바꾸려면 **FDR** 체크박스를 선택합니다 .

FDR 수정에 대한 자세한 내용은 Benjamini and Hochberg 연구 자료 (1995) 에서 확인하십시오 . FDR(False Discovery Rate) 에 대한 자세한 내용은 Predictive and Specialized Modeling 의 또는 Westfall et al. 연구 자료 (2011) 에서 확인하십시오 .

효과 유전성 열 유의성이 높은 고차 효과의 성분인 저차 효과를 식별합니다 . 저차 효과는 캐럿으로 식별됩니다 . 자세한 내용은 " 효과 유전성 "(171 페이지) 에서 확인하십시오 .

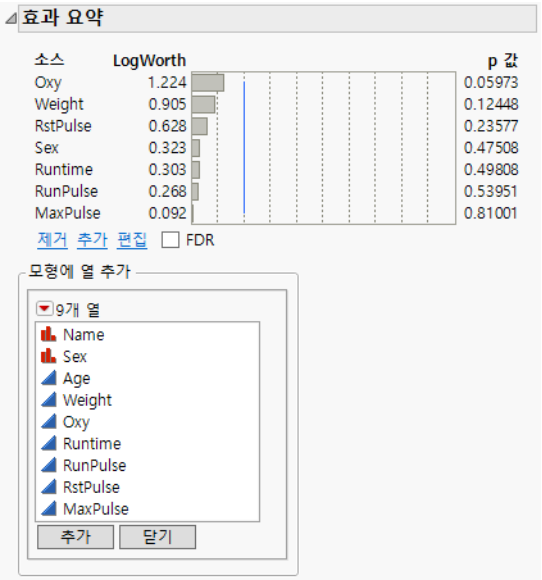
효과 요약 테이블 옵션

요약 테이블 아래의 옵션을 사용하여 효과를 추가 및 제거할 수 있습니다 .

제거 선택한 효과를 모형에서 제거합니다 . 하나 이상의 효과를 제거하려면 대상 효과에 해당하는 행을 선택하고 **제거** 버튼을 클릭합니다 .

추가 데이터 테이블의 모든 열 목록이 포함된 패널을 엽니다 . 모형에 추가할 열을 선택한 후 열 선택 목록 아래의 **추가**를 클릭하면 해당 열이 모형에 추가됩니다 . 패널을 닫으려면 **닫기**를 클릭합니다 . 그림 3.59 에서는 열 추가 패널을 보여 줍니다 .

그림 3.59 효과 요약 열 추가 패널



편집 " 열 선택 " 목록과 " 효과 " 지정 패널이 포함된 " 모형 편집 " 패널을 엽니다 . " 효과 " 패널은 " 모형 적합 " 시작 창의 " 모형 효과 생성 " 패널과 비슷합니다 . " 모형 편집 " 패널을 사용하여 개별 , 교차 , 내포 및 변환 효과를 추가할 수 있습니다 . " 매크로 " 메뉴를 사용하여 여러 효과를 추가할 수도 있습니다 . " 추가 ", " 교차 ", " 내포 ", " 매크로 " 및 " 변환 " 을 사용하여 효과를 생성하는 방법에 대한 자세한 내용은 "모형 효과 생성"(37페이지)에서 확인하십시오.

" 효과 " 패널 오른쪽의 다음 옵션을 사용할 수 있습니다 .

커밋 업데이트를 모형에 적용합니다 .

닫기 모형을 변경하지 않고 패널을 닫습니다 .

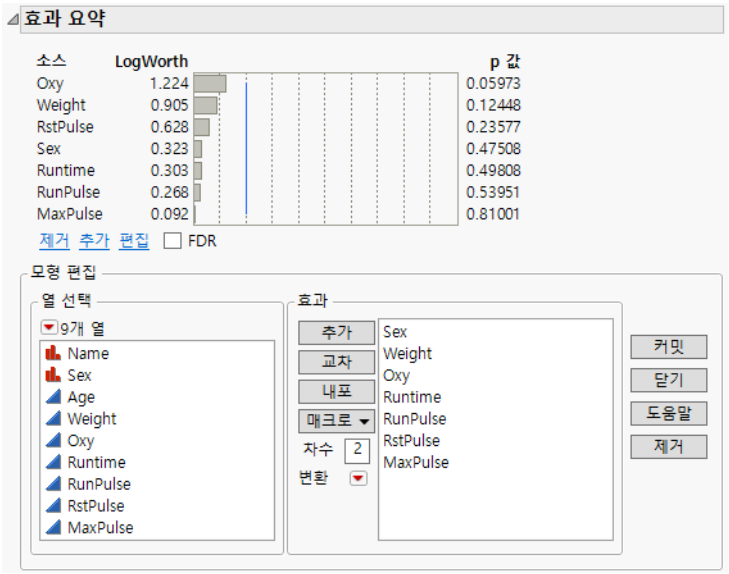
제거 " 효과 " 목록에서 선택한 하나 이상의 효과를 제거합니다 .

그림 3.60 에서는 " 모형 편집 " 패널을 보여 줍니다 .

팁 : " 편집 " 버튼을 사용하면 모형에 대한 업데이트를 최대한 제어할 수 있습니다 . " 제거 " 및 " 추가 " 버튼의 기능이 포함되고 모형에 추가할 효과를 생성할 수 있습니다 .

실행 취소 모형의 효과에 대한 변경 사항을 취소할 수 있습니다 .

그림 3.60 효과 요약 모형 편집 패널



효과 유전성

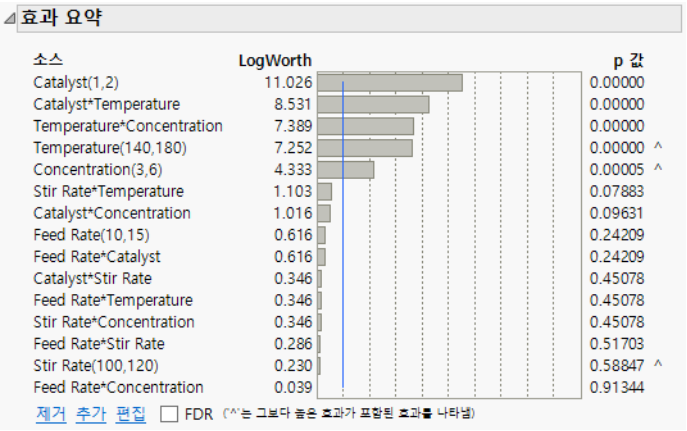
모형에 유의한 고차 효과가 포함된 경우 일부 또는 모든 저차 성분이 유의하지 않더라도 유지할 수 있습니다. **강한 효과 유전성** 원리에 따르면 모형에 고차 효과가 포함되면 모든 저차 성분도 포함되어야 합니다. **약한 효과 유전성** 원리는 성분 체인이 포함되어야 함을 나타냅니다.

고차 효과의 저차 성분이 "효과 요약" 테이블에서 고차 효과 아래에 나타나면 맨 오른쪽 열에 캐럿이 표시됩니다. 캐럿은 포함하는 고차 효과가 저차 효과보다 유의하다는 것을 나타냅니다. 저차 효과를 포함하는 모든 고차 효과가 저차 효과보다 유의성이 낮으면 저차 효과 행에 캐럿이 나타나지 않습니다.

캐럿으로 표시된 효과를 제거하는 경우 두 가지 효과 제거 방법 중 하나를 선택할 수 있습니다. 캐럿으로 표시된 효과를 포함하여 선택한 모든 효과를 제거하려면 **선택한 모든 효과 제거**를 선택합니다. 모형에 계속 유지되는 고차 효과를 제외하고 선택한 효과만 제거하려면 **비포함 효과만 제거**를 선택합니다.

그림 3.61에서는 저차 효과를 포함하는 고차 효과 아래에 세 가지 저차 효과가 표시된 "효과 요약" 테이블의 예를 보여 줍니다. 예를 들어 Stir Rate(100,120)가 Stir Rate*Temperature 아래에 나타납니다.

그림 3.61 Reactor 32 Runs.jmp 에 대해 효과 유전성이 포함된 효과 요약 테이블



다중 반응

다중 반응의 경우 각 반응 모형에 각 효과가 나타나지만 "효과 요약" 보고서는 하나만 나타납니다. 테이블에는 각 효과에 대해 해당 효과의 p 값 중 최소 p 값이 표시됩니다. 효과를 추가하거나 제거하면 모든 반응에 대한 모형에 적용됩니다.

혼합 및 임의 효과 모형 보고서 및 옵션

혼합 및 임의 효과 모형은 모형 적합 시작 창에서 지정할 수 있습니다 . 표준 최소 제곱 분석법은 분산 성분 공분산 구조를 적합시킵니다 . 이러한 모형을 적합시키기 위해 REML 과 EMS 라는 두 가지 방법이 제공됩니다 .

참고 : JMP PRO JMP Pro 사용자는 모형 적합 창에서 혼합 모형 분석법을 사용하는 것이 좋습니다 . 혼합 모형 분석법은 표준 최소 제곱 분석법에서 제공하는 것보다 광범위한 공분산 구조 집합을 제공합니다 .

- " 혼합 모형 및 임의 효과 모형 "
- "REML(제한 최대 가능도) 방법 "
- "EMS(전통적) 모형 적합 보고서 "

혼합 모형 및 임의 효과 모형

임의 효과 모형은 모든 요인이 임의 효과를 나타내는 모형입니다. 자세한 내용은 "임의 효과"(173 페이지)에서 확인하십시오. 이러한 모형을 **분산 성분 모형**이라고도 합니다. 대개 임의 효과 모형은 계층적 모형입니다. 고정 효과와 임의 효과가 모두 포함된 모형을 **혼합 모형**이라고 합니다.

반복 측정 및 분할구 모형은 혼합 모형의 특수한 경우입니다. **혼합 모형**이라는 용어는 임의 효과 모형을 포함하기 위해 사용되는 경우가 많습니다.

혼합 모형을 적합시키려면 모형 적합 시작 창에서 임의 효과를 지정해야 합니다. 그러나 모든 모형 효과가 임의 효과인 경우에는 계량형/계수형 게이지 차트 플랫폼에서 모형을 적합시킬 수도 있습니다. 이 플랫폼에서는 특정 모형만 적합시킬 수 있습니다. 계량형/계수형 게이지 차트 플랫폼에서 사용되는 적합 방법은 분산 성분 추정값에 음수를 허용하지 않습니다. 계량형/계수형 게이지 차트 플랫폼에서 분산 성분 모형을 적합시키는 방법에 대한 자세한 내용은 Quality and Process Methods의 에서 확인하십시오.

임의 효과

임의 효과는 수준이 일부 모집단의 랜덤 표본으로 간주되는 요인입니다. 대개 관심 대상은 임의 효과의 정확한 수준이 아니라 수준에 의해 반영되는 변동(**분산 성분**)입니다. 그러나 임의 효과의 특정 수준에 대한 반응을 예측하려는 상황도 있습니다. 기술적으로 임의 효과는 평균이 0이고 분산이 0이 아닌 정규 분포를 따르는 것으로 간주됩니다.

두 개의 특정 오븐이 금형 수축에 미치는 영향이 다른지 여부에 관심이 있다고 가정해 보겠습니다. 오븐은 한 번에 하나의 배치(50개 금형)만 처리할 수 있습니다. 한 배치당 50개씩 무작위로 선택한 세 개의 금형 배치를 두 오븐에 각각 연속적으로 넣어 연구를 설계합니다. 배치가 처리되면 각 배치에서 무작위로 선택한 5개 부품에 대해 수축이 측정됩니다.

Batch는 각 배치에 대해 하나씩 6개 수준이 있는 요인입니다. 따라서 모형에는 Oven과 Batch라는 두 가지 요인이 포함됩니다. 특히 각 오븐이 수축에 미치는 영향을 비교하는 데 관심이 있으므로 Oven은 고정 효과입니다. 그러나 특정 6개의 배치가 평균 수축에 미치는 영향에는 관심이 없습니다. 이러한 배치는 이 실험에서 선택할 수 있고 분석 결과를 일반화해야 하는 배치의 전체 모집단을 나타냅니다. Batch는 임의 효과로 간주됩니다. 가능한 모든 배치에서 수축 변동을 연구하는 이 실험에서 Batch 요인은 관심 대상입니다. 요인이 설명하는 수축의 변동량을 추정하는 데 관심이 있습니다. 또한 한 오븐에서 하나의 배치만 처리할 수 있으므로 Batch가 Oven에 내포됩니다.

이제 두 가지 사료 체계가 적용된 암탉의 알 무게에 관심이 있다고 가정해 보겠습니다. 10마리 암탉을 무작위로 선택하여 5마리는 사료 체계 A에 할당하고 5마리는 사료 체계 B에 할당합니다. 그러나 10마리의 암탉은 연구 설계에서 고려되지 않은 유전적 차이가 있습니다. 이 경우 특정 암탉이 낳은 달걀의 예측 무게 및 전체 암탉 간의 달걀 무게에 대한 분산에 관심이 있습니다.

전통적 선형 혼합 모형

JMP 에서는 다음과 같이 전통적 선형 혼합 효과 모형을 적합시킵니다 .

$$Y = X\beta + Z\gamma + \varepsilon$$
$$\gamma \sim N(0, G)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

다음은 각 요소에 대한 설명입니다 .

- Y - 반응의 n x 1 벡터
- $\mathbf{X}$ - 고정 효과에 대한 n x p 설계 행렬
- β - 설계 행렬 X 를 사용하는 알려지지 않은 고정 효과의 p x 1 벡터
- $\mathbf{Z}$ - 임의 효과에 대한 n x s 설계 행렬
- γ - 설계 행렬 Z 를 사용하는 알려지지 않은 임의 효과의 s x 1 벡터
- ε - 알 수 없는 랜덤 오차의 n x 1 벡터
- $\mathbf{G}$ - 임의 효과의 각 수준에 대해 동일한 항목이 있는 s x s 대각 행렬
- I_n - n x n 항등 행렬
- γ 와 ε 는 독립적 관계

G 의 대각 요소와 σ^2 을 **분산 성분**이라고 합니다. 이러한 분산 성분은 고정 효과 벡터 (β) 및 임의 효과 벡터 (γ) 와 함께 추정 대상인 모형 모수입니다.

경우에 따라 이 모형의 공분산 구조를 **분산 성분** 구조라고도 합니다 (SAS Institute Inc. 2020d, ch. 83 이 공분산 구조만 표준 최소 제곱 분석법에서 사용할 수 있습니다.

JMP PRO 혼합 모형 분석법은 잔차, AR(1)(**1차 자기회귀**), 비정형 및 공간을 포함한 다양한 공분산 구조를 적합시킵니다. 자세한 내용은 "혼합 모형" 장의 "반복 구조 탭"(339 페이지) 에서 확인하십시오.

임의 효과가 있는 모형의 적합 방법 (REML, EMS)

JMP 에서는 임의 효과가 있는 모형을 적합시키기 위해 다음 두 가지 방법을 제공합니다 .

- REML - **제한 최대 가능도** (항상 권장되는 방법)
- EMS - **기대 평균 제곱** (이전 교재로 교육할 때만 사용)

REML 방법은 기존의 EMS 방법을 대체하는 주요 적합 방법입니다. REML 은 적용 가능성 면에서 EMS 방법보다 훨씬 더 일반적입니다. REML 방법은 Patterson and Thompson 연구 자료 (1974) 에서 최초로 공개되었습니다. Wolfinger et al. 연구 자료 (1994) 및 Searle et al. 연구 자료 (1992) 도 참조하십시오.

EMS 방법은 강력한 성능을 가진 컴퓨터가 나오기 전에 개발되었으며 **적률 방법**이라고도 합니다. 연구원들은 균형 잡힌 상황이 되도록 제한했고, 임의 효과 및 혼합 모형에 대한 추정값을 쉽게 계산할 수 있는 EMS 방법을 사용했습니다. 요즘도 임의 효과가 포함된 모형을 시작할 때 EMS 방법을 사용하는 교재가 많기 때문에 JMP 에서 EMS 옵션을 제공합니다. 자세한 내용은 McCulloch et al., 2008, Poduri, 1997, Searle et al., 1992 등에서 확인하십시오.

REML 방법은 고정 효과 모수에 종속되지 않는 제한 가능도 함수의 최대 가능도 추정을 수행합니다. 이렇게 하면 분산 성분의 추정값이 생성되어 고정 효과의 추정값을 구하는 데 사용됩니다. 정밀도 추정값은 모수에 대한 공분산 행렬의 추정값을 기반으로 합니다. 데이터가 불균형한 경우에도 REML 은 유용한 추정값, 검정 및 신뢰 구간을 제공합니다.

EMS 방법은 관측 평균 제곱을 기대 평균 제곱에 일치시켜 분산 성분의 추정값을 계산합니다. 균형 설계의 경우 추정값을 구하는 방법을 지정하는 복잡한 규칙 집합을 사용합니다. 이 기법을 불균형 데이터에 적용하면 문제가 있습니다.

균형 데이터의 경우 REML 추정값은 EMS 추정값과 동일합니다. 그러나 EMS와 달리 REML은 불균형 데이터에서도 정상적으로 수행됩니다.

임의 효과 및 적합 방법 지정

임의 효과가 있는 모형은 모형 적합 시작 창에서 지정됩니다. 임의 효과를 지정하려면 "모형 효과 생성" 목록에서 해당 효과를 강조 표시하고 **속성 > 임의 효과**를 선택합니다. 그러면 모형 효과 목록의 효과 이름에 "& 랜덤"이 추가됩니다. 임의 효과의 정의는 "임의 효과"(173 페이지)에서 확인하십시오. 임의 효과를 별도의 효과 탭에서 지정할 수도 있습니다. 자세한 내용은 "모형 규격" 장의 "모형 효과 생성 탭"(42 페이지)에서 확인하십시오.

"모형 적합" 시작 창에서 효과에 "& 랜덤" 속성이 추가되면 "REML(권장)" 또는 "EMS(전통적)" 적합 방법을 선택할 수 있습니다.

주의 : 교차 및 내포 관계를 명시적으로 선언해야 합니다 . 예를 들어 각 개체가 하나의 그룹에만 있는 경우 개체 ID 는 해당 개체를 포함하는 그룹을 식별할 수도 있습니다 . 이러한 경우 개체 ID 를 그룹에 내포된 것으로 선언해야 합니다 . 설계 구조를 명시적으로 정의할 수 있도록 주의하십시오 .

분산 성분에 대한 무제한 파라미터화

분산 성분을 파라미터화하는 방식에는 두 가지가 있습니다(무제한 및 제한 방식). 고정 효과와 임의 효과의 교호작용 같은 혼합 효과가 모형에 있을 때 이 문제가 발생합니다. 이러한 교호작용 항은 임의 효과로 간주됩니다.

제한된 방식의 경우 임의 효과의 각 수준에 대해 고정 효과의 전체 수준에서 교호작용 효과의 합이 0으로 간주됩니다. 무제한 방식의 경우 혼합 항은 평균 0과 공통 분산을 사용하는 정규 분포의 독립된 랜덤 구현으로 간주됩니다. 이 가정은 일반적으로 잔차 오차에 적용되는 가정과 비슷합니다.

JMP와 SAS에서는 무제한 방식을 사용합니다. 대부분의 통계학 교재에는 제한된 방식이 사용되므로 이를 구분하는 것이 중요합니다. 두 가지 방식 모두 60년 동안 광범위하게 교육에 사용되었습니다. 두 가지 방식에 대한 자세한 내용은 Cobb 연구 자료(1998, Section 13.3)에서 확인하십시오.

음의 분산

분산은 항상 양수이지만 분산의 비편향 추정값이 음수인 상황이 있을 수 있습니다. 효과가 매우 약하거나, 분산 성분에 해당하는 수준이 매우 적으면 실험에서 음수 추정값이 발생할 수 있습니다. 의도하지 않았지만 관측된 데이터에서 음수 추정값이 생성될 수 있습니다.

비유계 분산 성분

JMP에서는 REML과 EMS 둘 다에 대해 음수 추정값을 생성할 수 있습니다. REML의 경우 "모형 적합" 시작 창에 "비유계 분산 성분"과 "분산 성분만 추정"이라는 두 가지 옵션이 있습니다. 기본적으로 "비유계 분산 성분" 옵션이 선택되어 있습니다. 이 옵션을 선택 취소하면 분산 성분 추정값이 음수가 아닌 값으로 제한됩니다.

고정 효과에 관심이 있는 경우 "비유계 분산 성분" 옵션을 선택된 상태로 두어야 합니다. 분산 추정값을 음수가 아닌 값으로 제한하면 고정 효과 검정에서 편향이 발생합니다.

분산 성분만 추정

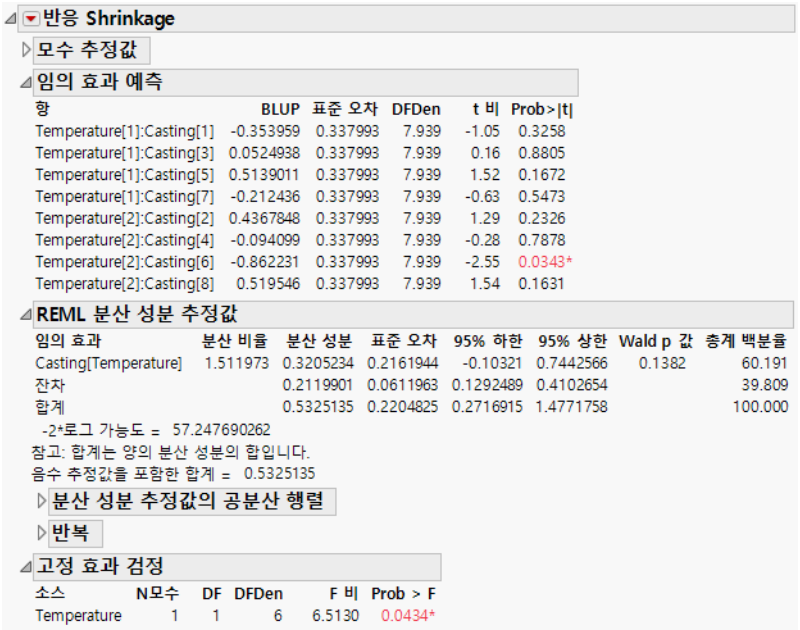
"REML 분산 성분 추정값" 보고서만 표시하려면 이 옵션을 선택합니다. 분산 성분에만 관심이 있는 경우 분산 성분을 음수가 아닌 값으로 제한할 수 있습니다. "비유계 분산 성분" 옵션을 선택 취소하고 "분산 성분만 추정" 옵션을 선택하는 것이 적절할 수 있습니다.

REML(제한 최대 가능도) 방법

선택한 적합 방법에 따라 "최소 제곱 적합" 보고서에 서로 다른 분석 결과가 제공되고 "열 저장" 및 "프로파일러"에 대한 추가 메뉴 옵션이 제공됩니다. 특히 분산과 자유도의 경우 일반적인 방식으로 분할되지 않으므로 분산 분석 보고서가 표시되지 않습니다. "REML 분산 성분 추정값" 보고서에서 잔차 분산 추정값을 얻을 수 있습니다. 자세한 내용은 "REML 분산 성분 추정값"(178 페이지)에서 확인하십시오. "효과 검정" 보고서는 고정 효과가 검정되는 "고정 효과 검정" 보고서로 대체됩니다. 추가 보고서는 임의 효과에 대한 예측값과 분산 성분에 대한 상세 정보를 제공합니다.

그림 3.62에서는 REML 방법을 사용하여 생성된 Investment Castings.jmp 샘플 데이터 적합 보고서를 보여 줍니다. **Model: REML** 스크립트를 실행한 후 모형을 적합시킵니다. Casting은 임의 효과이며 Temperature에 내포되어 있습니다.

그림 3.62 REML 방법에 대한 최소 제곱 적합 보고서



반응 Shrinkage

모수 추정값

임의 효과 예측

항	BLUP	표준 오차	DFDen	t 비	Prob>|t|
Temperature[1]:Casting[1]	-0.353959	0.337993	7.939	-1.05	0.3258
Temperature[1]:Casting[3]	0.0524938	0.337993	7.939	0.16	0.8805
Temperature[1]:Casting[5]	0.5139011	0.337993	7.939	1.52	0.1672
Temperature[1]:Casting[7]	-0.212436	0.337993	7.939	-0.63	0.5473
Temperature[2]:Casting[2]	0.4367848	0.337993	7.939	1.29	0.2326
Temperature[2]:Casting[4]	-0.094099	0.337993	7.939	-0.28	0.7878
Temperature[2]:Casting[6]	-0.862231	0.337993	7.939	-2.55	0.0343*
Temperature[2]:Casting[8]	0.519546	0.337993	7.939	1.54	0.1631

REML 분산 성분 추정값

임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
Casting[Temperature]	1.511973	0.3205234	0.2161944	-0.10321	0.7442566	0.1382	60.191
잔차		0.2119901	0.0611963	0.1292489	0.4102654		39.809
합계		0.5325135	0.2204825	0.2716915	1.4771758		100.000

-2*로그 가능도 = 57.247690262
참고: 합계는 양의 분산 성분의 합입니다.
음수 추정값을 포함한 합계 = 0.5325135

분산 성분 추정값의 공분산 행렬

반복

고정 효과 검정

소스	N모수	DF	DFDen	F 비	Prob > F
Temperature	1	1	6	6.5130	0.0434*

임의 효과 예측

이 보고서에는 모형의 각 항에 대해 BLUP(**최량 선형 비편향 예측 변수**)의 경험적 추정값 및 해당 계수가 0인지 여부에 대한 검정이 제공됩니다.

참고: "임의 효과 예측" 보고서를 표시하려면 회귀 보고서 > 모수 추정값 옵션을 선택해야 합니다.

항 모형에서 임의 효과에 해당하는 항입니다.

BLUP 각 임의 효과에 대한 BLUP(**최량 선형 비편향 예측 변수**) 의 경험적 추정값입니다. 자세한 내용은 " 최량 선형 비편향 예측 변수 "(178 페이지) 에서 확인하십시오.

표준 오차 BLUP 의 표준 오차입니다.

DFDen 효과가 0 인지에 대한 검정의 분모 자유도입니다. 대부분의 경우 *t*- 검정의 자유도는 분수입니다.

t 비 효과가 0 인지 검정하기 위한 *t* 비입니다. *t* 비는 BLUP 를 표준 오차로 나눈 값입니다.

Prob>|t| 검정의 *p* 값입니다.

95% 하한 BLUP 에 대한 95% 신뢰 하한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 하한을 선택한 경우에만 나타납니다.

95% 상한 BLUP 에 대한 95% 신뢰 상한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 상한을 선택한 경우에만 나타납니다.

최량 선형 비편향 예측 변수

BLUP(최량 선형 비편향 예측 변수)는 임의 효과의 추정량을 나타냅니다. 구체적으로 모든 비편향 추정량 중에서 평균 제곱 예측 오차를 최소화하는 추정량입니다. "임의 효과 예측" 보고서에는 BLUP 또는 경험적 BLUP의 추정값이 제공됩니다. BLUP는 알려지지 않은 분산 성분의 값에 의존하므로 경험적입니다. 분산 성분의 추정값이 BLUP 계산식으로 대체되어 보고서에 표시된 추정값이 생성됩니다.

REML 분산 성분 추정값

"모형 적합" 시작 창에서 "REML"을 적합 방법으로 선택하면 "REML 분산 성분 추정값" 보고서가 제공됩니다. 이 보고서에는 다음 열이 포함됩니다.

임의 효과 모형의 임의 효과입니다.

분산 비율 잔차의 분산 성분에 대한 효과의 분산 성분 비율입니다. 이 값은 효과의 추정 분산을 모형의 추정 오차 분산과 비교합니다.

분산 성분 효과에 대한 추정 분산 성분입니다. "합계"의 분산 성분은 양의 분산 성분만 합한 값입니다. 모든 분산 성분의 합은 테이블 아래에 제공됩니다.

표준 오차 분산 성분 추정값에 대한 표준 오차입니다.

95% 하한 분산 성분에 대한 95% 신뢰 하한입니다. 자세한 내용은 "분산 성분에 대한 신뢰 구간"(178 페이지)에서 확인하십시오.

95% 상한 분산 성분에 대한 95% 신뢰 상한입니다. 자세한 내용은 "분산 성분에 대한 신뢰 구간"(178 페이지)에서 확인하십시오.

Wald p 값 공분산 모수가 0이라는 검정에 대한 p 값입니다. 이 열은 "모형 적합" 시작 창에서 "비유계 분산 성분"을 선택한 경우에만 나타납니다.

총계 백분율 합계의 분산 성분에 대한 효과의 분산 성분 비율을 백분율로 나타낸 값입니다.

분산 성분 제곱근(표준편차) 해당 분산 성분의 제곱근입니다. 이 값은 효과에 대한 표준편차의 추정값입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 분산 성분 제곱근(표준편차)을 선택한 경우에만 나타납니다.

CV 분산 성분에 대한 변동 계수입니다. 이 값은 분산 성분의 제곱근에 100을 곱하여 평균 반응으로 나눈 값입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > CV를 선택한 경우에만 나타납니다.

표준 KHC Kackar-Harville 수정입니다. 자세한 내용은 "Kackar-Harville 수정"(179 페이지)에서 확인하십시오. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 표준 KHC를 선택한 경우에만 나타납니다.

분산 성분에 대한 신뢰 구간

신뢰 한계를 계산하는 데 사용되는 방법은 "모형 적합" 시작 창에서 "비유계 분산 성분"을 선택했는지 여부에 따라 다릅니다. "비유계 분산 성분"은 기본적으로 선택되어 있습니다.

- " 비유계 분산 성분 " 을 선택한 경우 Wald 기반 신뢰 구간이 계산됩니다 . 이러한 구간은 점근적으로 유효하지만 표본이 작은 경우에는 신뢰할 수 없습니다 .
- "비유계 분산 성분"을 선택하지 않은 경우, 즉 모수의 하한 제약 조건이 0 인 경우에는 Satterthwaite 근사가 사용됩니다 (Satterthwaite 1946).

Kackar-Harville 수정

REML 방법에서 고정 효과의 표준 오차는 분산 성분 추정값을 사용하여 추정됩니다. 그러나 이러한 추정값의 변동성을 고려하지 않으면 표준 오차가 과소추정됩니다. 증가한 변동을 설명하기 위해 Kackar-Harville 수정(Kackar and Harville 1984, Kenward and Roger 1997)을 사용하여 고정 효과의 공분산 행렬이 조정됩니다. 고정 효과의 공분산 행렬을 포함하는 모든 계산에 이 수정이 사용됩니다. 여기에는 최소 제곱 평균, 고정 효과 검정, 신뢰 구간 및 예측 분산이 포함됩니다. 통계 상세 정보는 "Kackar-Harville 수정 "(192 페이지)에서 확인하십시오.

표준 KHC 는 Kackar-Harville 수정의 Frobenius(행렬) norm 입니다. 균형 잡힌 설계의 경우 표준 KHC 가 작은 경향이 있습니다.

분산 성분 추정값의 공분산 행렬

이 보고서에는 분산 성분에 대한 점근 공분산 행렬의 추정값이 제공됩니다. 이 값은 관측된 Fisher 정보 행렬의 역입니다.

반복

G 의 분산 성분 및 σ^2 의 추정값은 이러한 모수에만 의존하는 잔차 로그 가능도 함수를 최대화하여 구합니다 . 반복 절차에서는 잔차 로그 가능도 함수를 최대화하거나 , 이와 동등하게 음의 잔차 로그 가능도의 두 배 (-2* 로그 가능도) 를 최소화하려고 합니다 . " 반복 " 보고서에는 이 절차에 대한 상세 정보가 제공됩니다 .

반복 반복 번호입니다 .

-2* 로그 가능도 음의 로그 가능도의 두 배입니다 . 이것은 목적 함수입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

그래디언트 norm 목적 함수의 그래디언트 (1 차 도함수) norm 입니다 .

모수 " 모수 " 라벨이 지정된 열과 나머지 열은 각각 임의 효과에 해당합니다 . 열 순서는 "REML 분산 성분 추정값 " 보고서에서 임의 효과가 나열된 순서를 따릅니다 . 각 반복에서 열의 값은 해당 시점에서 분산 성분의 추정값입니다 .

수렴 기준은 기본 공차가 10^{-8} 인 그래디언트를 기반으로 합니다 . " 모형 적합 " 시작 창에서 수렴 설정 > 수렴 한계 옵션을 선택한 후 원하는 공차를 지정하여 기준을 변경할 수 있습니다 .

고정 효과 검정

REML 을 사용하는 경우 " 효과 검정 " 보고서에 고정 효과에 대한 검정이 제공됩니다 . 이 보고서에는 다음 열이 포함됩니다 .

소스 모형의 고정 효과입니다 .

모수 수 효과와 관련된 모수의 수입니다 .

DF 효과와 관련된 자유도입니다 .

DFDen 분모 자유도입니다 . 이 값은 Kenward-Roger 수정을 사용하여 공분산 행렬을 조정할 때 얻은 통계량 분포에 대한 근사를 기반으로 합니다 . 자세한 내용은 "Kackar-Harville 수정 "(179 페이지) 및 " 임의 효과 "(76 페이지) 에서 확인하십시오 .

F 비 계산된 *F* 비입니다 .

Prob > F 효과 검정의 *p* 값입니다 .

REML 열 저장 옵션

REML 방법을 사용하는 경우 " 열 저장 " 메뉴에 6 개의 추가 옵션이 나타납니다 . 이러한 옵션 이름은 **조건부**로 시작합니다 . 이 접두사는 해당 열을 계산할 때 기대값 0 대신 임의 효과와 관련된 항의 예측값을 사용한다는 것을 나타냅니다 .

조건부 예측 계산식 예측 계산식을 데이터 테이블의 새 열에 저장합니다 .

조건부 예측값 예측값을 데이터 테이블의 새 열에 저장합니다 .

조건부 잔차 잔차를 데이터 테이블의 새 열에 저장합니다 .

조건부 평균 CI 평균에 대한 신뢰 구간을 저장합니다 .

조건부 개별값 CI 개별값에 대한 신뢰 구간을 저장합니다 .

JMP PRO **조건부 계산식 게시** 조건부 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

REML 프로파일러 옵션

REML 방법을 사용하는 경우 요인 프로파일링 > 프로파일러를 선택하면 " 예측 프로파일러 " 옆의 빨간색 삼각형 메뉴에 "조건부 예측"이라는 새 옵션이 나타납니다. 조건부 값은 기대값 0 대신 임의 효과에 대한 예측값을 사용합니다.

참고: 프로파일러에는 모든 요인 수준 조합에 대한 조건부 예측값과 조건부 평균 신뢰 구간이 표시됩니다 . 일부 조합은 내포로 인해 의미가 없을 수도 있습니다 .

EMS(전통적) 모형 적합 보고서

주의 : EMS 는 사용하지 않는 것이 좋습니다 . REML 이 권장되는 방법입니다 .

"EMS" 를 적합 방법으로 선택하면 4 개의 새 보고서가 표시됩니다. 고정 효과와 임의 효과 둘 다에 대한 검정은 "임의 효과에 대해 검정" 보고서에서 수행되므로 "효과 검정" 보고서는 표시되지 않습니다.

기대 평균 제곱

모형 효과에 대한 기대 평균 제곱은 잔차 오차 분산을 포함하여 분산 성분과 고정 효과 값의 선형 결합입니다. 이 테이블에는 각 모형 효과의 기대 평균 제곱을 정의하는 계수가 제공됩니다. 행렬의 행은 왼쪽에 나열된 효과에 해당합니다. 열은 상단에 표시된 분산 성분에 해당합니다. 각 기대 평균 제곱에는 계수가 1 인 잔차 분산이 포함됩니다. 이 정보는 테이블 아래에 제공됩니다.

그림 3.63 에서는 Investment Castings.jmp 샘플 데이터 테이블에 대한 "기대 평균 제곱" 보고서를 보여 줍니다. "Model: EMS" 스크립트를 실행한 후 모형을 실행합니다.

그림 3.63 기대 평균 제곱 보고서

기대 평균 제곱

행 평균 제곱 대 열 분산 성분
EMS

	절편	Temperature	Casting[Temperature]&랜덤
절편	0	0	0
Temperature	0	16	4
Casting[Temperature]&랜덤	0	0	4

+ 1.0 x 잔차 분산

테이블에서 알 수 있듯이 Casting[Temperature] 에 대한 기대 평균 제곱은 다음과 같습니다 .

$$4\sigma^2_{Casting[Temperature]} + \sigma^2_{Error}$$

분산 성분 추정값

분산 성분의 추정값은 기대 평균 제곱을 해당하는 관측 평균 제곱에 일치시킨 후 계산하여 구합니다 . " 분산 성분 추정값 " 보고서에는 추정된 분산 성분이 제공됩니다 .

성분 임의 효과를 나열합니다 .

분산 성분 추정값 분산 성분의 추정값을 제공합니다 .

총계 백분율 분산 성분 합계에 대한 분산 성분의 비율을 제공합니다 .

CV 분산 성분에 대한 변동 계수를 제공합니다 . 이 값은 분산 성분의 제곱근에 100 을 곱하여 평균 반응으로 나눈 값입니다 .

참고 : 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > CV 를 선택한 경우에만 나타납니다 .

분모 합성 검정

검정할 각 효과에 대해 F 통계량이 생성됩니다 . 이 통계량의 분모는 귀무가설 하에서 분자 평균 제곱의 기대값을 갖는 평균 제곱입니다 . 이 분모는 고정 효과와 관련된 값 및 분산 성분에서 생성되거나 **합성됩니다** .

소스 검정할 효과를 표시합니다 .

MS Den *F*- 검정의 분모에 대한 추정 평균 제곱을 제공합니다 .

DF Den 합성 분모의 자유도를 제공합니다 . 이 값은 Satterthwaite 방법 (Satterthwaite 1946) 을 사용하여 생성됩니다 .

MS 합성 (분모) 분모 합성에 사용된 분산 성분을 제공합니다 . 잔차 오차 분산은 항상 이 합성의 일부입니다 .

임의 효과에 대해 검정

이 보고서에는 고정 효과와 임의 효과에 대한 검정이 제공됩니다 .

소스 검정할 효과를 나열합니다 . 여기에는 고정 효과와 임의 효과가 포함됩니다 .

SS 효과에 대한 제곱합을 제공합니다 .

MS Num 분자 평균 제곱을 제공합니다 .

DF Num 분자 자유도를 제공합니다 .

F 비 검정에 대한 *F* 비를 제공합니다 . 이 값은 분모 평균 제곱에 대한 분자 평균 제곱의 비율입니다 . 분모 평균 제곱은 " 분모 합성 검정 " 보고서에서 얻을 수 있습니다 .

Prob > F 효과 검정을 위한 *p* 값을 제공합니다 .

주의 : 최소 제곱 평균에 대한 표준 오차와 대비 *F*- 검정에 대한 분모는 합성 분모를 사용합니다 . 공통 수준에서 비교되는 교차 효과가 포함된 검정과 같은 특정 상황에서는 이러한 검정이 적절하지 않을 수 있습니다 . 사용자 검정은 잔차 오류를 사용하여 수행되고 레버리지 그림은 잔차 오류를 사용하여 생성되므로 역시 적절하지 않을 수 있습니다 .

EMS 프로파일러

EMS 방법을 사용하는 경우 요인 프로파일링 > 프로파일러를 선택하면 프로파일러에서 고정 효과 모형을 기반으로 예측 및 조건부 평균 신뢰 구간을 제공합니다 . 이러한 값은 임의 효과에 대한 예측값을 기반으로 하지 않습니다 .

모형 항 사이에 선형 종속성이 있는 모형

예측 변수 행렬의 열 사이에 선형 종속성이 있으면 몇 가지 표준 최소 제곱 보고서에 영향을 줍니다 .

- " 특이성 상세 정보 "

- "모수 추정값 보고서"
- "효과 검정 보고서"
- "예"

특이성 상세 정보

선형 회귀 모형은 $Y = X\beta + \varepsilon$ 로 표현됩니다. 여기서 X 는 첫 번째 열이 *1* 로 구성되고 나머지 열이 모형의 비절편 항에 대한 값으로 구성된 행렬입니다. 모형이 절편을 포함하여 p 개 항으로 구성된 경우 X 는 $n \times p$ 행렬이며, 여기서 n 은 관측값 수입니다. 벡터 b 로 나타내는 모수 추정값은 일반적으로 다음 계산식에 의해 제공됩니다.

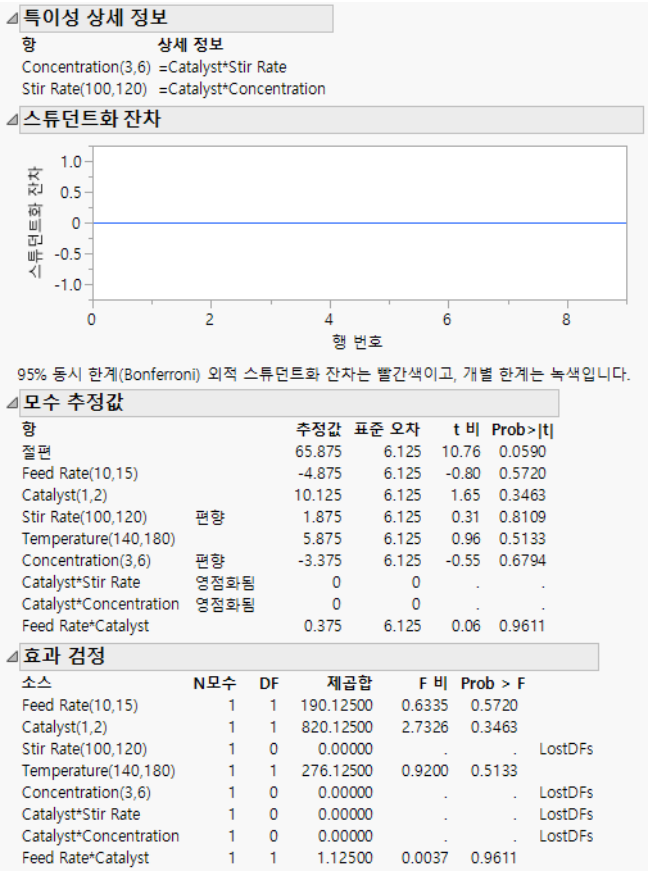
$$b = (X'X)^{-1}X'Y$$

그러나 이 계산식은 $X'X^{-1}$이 있다고 가정합니다. 즉, $p \times p$ 행렬 $X'X$를 반전할 수 있거나 완전 계수라고 가정합니다. X 열 사이에 선형 종속성이 있으므로 $X'X$를 반전할 수 없는 상황이 종종 발생합니다.

이 경우 $X'X$ 행렬은 특이 행렬이고 "최소 제곱 적합" 보고서에 "특이성 상세 정보" 보고서가 포함됩니다. 이 보고서에는 선형 종속성을 설명하는 표현식 테이블이 포함되어 있습니다. 이러한 선형 종속성과 관련된 항은 별칭이 지정됩니다(교락).

그림 3.64에서는 Reactor 8 Runs.jmp 샘플 데이터 테이블에 대한 보고서를 보여 줍니다. 이러한 보고서를 생성하려면 Percent Reacted 를 Y 로 지정하여 모형을 적합시킵니다. Feed Rate, Catalyst, Stir Rate, Temperature, Concentration, Catalyst*Stir Rate, Catalyst*Concentration 및 Feed Rate*Catalyst 를 모형 효과로 입력합니다.

그림 3.64 선형 종속성이 있는 모형에 대한 특이성 및 모수 추정값 보고서



특이성 상세 정보

항	상세 정보
Concentration(3,6)	=Catalyst*Stir Rate
Stir Rate(100,120)	=Catalyst*Concentration

스튜던트화 잔차

95% 동시 한계(Bonferroni) 외적 스튜던트화 잔차는 빨간색이고, 개별 한계는 녹색입니다.

모수 추정값

항		추정값	표준 오차	t 비	Prob>|t|
절편		65.875	6.125	10.76	0.0590
Feed Rate(10,15)		-4.875	6.125	-0.80	0.5720
Catalyst(1,2)		10.125	6.125	1.65	0.3463
Stir Rate(100,120)	편향	1.875	6.125	0.31	0.8109
Temperature(140,180)		5.875	6.125	0.96	0.5133
Concentration(3,6)	편향	-3.375	6.125	-0.55	0.6794
Catalyst*Stir Rate	영점화됨	0	0	.	.
Catalyst*Concentration	영점화됨	0	0	.	.
Feed Rate*Catalyst		0.375	6.125	0.06	0.9611

효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F	
Feed Rate(10,15)	1	1	190.12500	0.6335	0.5720	
Catalyst(1,2)	1	1	820.12500	2.7326	0.3463	
Stir Rate(100,120)	1	0	0.00000	.	.	LostDFs
Temperature(140,180)	1	1	276.12500	0.9200	0.5133	
Concentration(3,6)	1	0	0.00000	.	.	LostDFs
Catalyst*Stir Rate	1	0	0.00000	.	.	LostDFs
Catalyst*Concentration	1	0	0.00000	.	.	LostDFs
Feed Rate*Catalyst	1	1	1.12500	0.0037	0.9611	

모수 추정값 보고서

$X'X$가 특이 행렬인 경우 일반화 역행렬을 사용하여 추정값을 구합니다. 이 방법을 사용하면 선형 종속성과 관련된 모수의 일부(전부는 아님)를 추정할 수 있습니다. 모수는 관련된 항이 모형에 입력되는 순서에 따라 추정되므로 마지막으로 입력한 항은 모수가 추정되지 않는 항입니다. "모수 추정값" 보고서에 추정값이 제공되고 추정할 수 없는 모수의 추정값은 0으로 표시됩니다.

그러나 선형 종속성과 관련된 항의 모수 추정값은 고유하지 않습니다. 관련된 항은 별칭 관계이므로 최소 제곱 기준을 충족하는 추정값 벡터가 무한히 많습니다. 이러한 경우 "모수 추정값" 보고서에서 해당 추정값의 왼쪽에 "편향"이 표시됩니다. 모수를 추정할 수 없는 선형 종속성 관련 항의 경우 "모수 추정값" 보고서에서 추정값 0의 왼쪽에 "영점화됨"이 표시됩니다. 예는 그림 3.64에서 확인하십시오.

오차 추정값에 사용할 수 있는 자유도가 있는 경우 편향 추정값을 사용하여 추정된 모수에 대한 *t*-검정이 수행됩니다. 그러나 이러한 검정은 추정값이 고유하지 않다는 점을 고려하여 신중하게 해석해야 합니다.

효과 검정 보고서

표준 최소 제곱 적합에서는 존재하는 모형 자유도와 같은 수의 모수만 추정할 수 있습니다 . " 효과 검정 " 보고서의 검정을 수행할 때 각 효과는 모형에 입력된 마지막 효과로 간주됩니다 .

- 모든 모형 자유도가 다른 효과에 사용되는 경우 효과의 DF 가 0 으로 표시됩니다 . DF 가 0 이면 제곱합을 계산할 수 없습니다 . 따라서 해당 효과를 검정할 수 없습니다 .
- 일부 모형 자유도가 다른 효과에 사용되는 경우 나머지 효과의 DF 는 0 이 아닙니다 . 그러나 DF 가 모수의 수 (" 모수 수 ") 보다 작을 수 있으며 , 이는 관련 모수 중 일부만 검정할 수 있다는 의미입니다 .

오차에 대한 자유도가 있다고 가정하고 효과에 대한 자유도가 0이 아니면 *F*-검정이 수행됩니다. DF가 모수 수보다 작으면 효과에 해당하는 행의 오른쪽 끝에 LostDFs라는 설명이 표시됩니다(그림 3.64). 이러한 효과는 손실된 자유도를 흡수한 별칭 관계의 효과에 기인하지 않은 모형 제곱합만 설명할 수 있습니다. 따라서 "효과 검정" 보고서에 제공된 제곱합은 효과와 관련된 "실제" 제곱합보다 적게 표시될 가능성이 높습니다. 검정이 유의하면 유의성이 의미가 있습니다. 그러나 유의성 부족은 신중하게 해석해야 합니다.

통계 상세 정보는 SAS Institute Inc. (2020c).

예

Singularity.jmp 샘플 데이터 테이블을 엽니다. 반응 Y, 4개의 예측 변수 X1, X2, X3, A 및 5개의 관측값이 있습니다. 예측 변수는 4개 수준의 명목형 변수 A를 제외하고 모두 연속형입니다. 또한 연속형 효과 사이에 선형 종속성이 있습니다. 즉, X3 = X1 + X2 입니다.

고유하지 않은 추정값

선형 종속성이 있을 때 추정값이 고유하지 않은지 확인하려면 다음을 수행하십시오 .

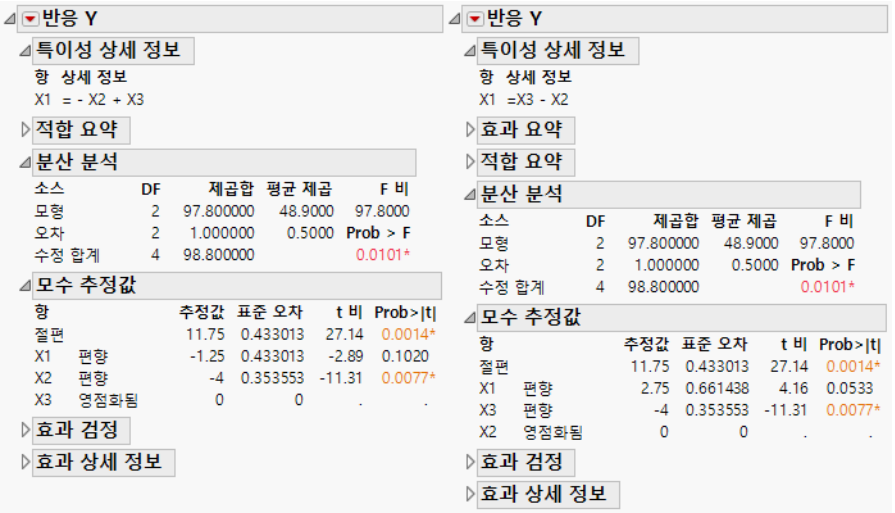
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Singularity.jmp 를 엽니다 .
2. **Model 1** 스크립트를 실행합니다 . 그러면 X1, X2, X3 순서로 효과가 입력된 " 모형 적합 " 시작 창이 열립니다 .
3. **실행**을 클릭하고 보고서 창은 열린 상태로 둡니다 .
4. **Model 2** 스크립트를 실행합니다 . 그러면 X1, X3, X2 순서로 효과가 입력된 " 모형 적합 " 시작 창이 열립니다 .
5. **실행**을 클릭하고 보고서 창은 열린 상태로 둡니다 .

두 보고서(그림 3.65)를 비교합니다. 두 보고서의 맨 위에 있는 "특이성 상세 정보" 보고서에는 선형 종속성이 표시됩니다(X1 = X3 - X2).

이제 두 모형의 "모수 추정값" 보고서를 비교합니다. 예를 들어 X1에 대한 추정값이 Model 1은 –1.25이고 Model 2는 2.75입니다. 두 모형 모두 모형 자유도가 2이므로 효과와 관련된 항 중 두 개만 추정됩니다. "분산 분석" 보고서를 참조하십시오. 추정된 두 항의 추정값은 "편향"이라는 라벨이 지정되며, 나머지 추정값은 0으로 설정되고 "영점화됨"이라는 라벨이 지정됩니다.

"효과 검정" 보고서에서는 검정이 수행되지 않았음을 보여 줍니다. 각 행에 LostDFs 라벨이 지정되어 있습니다. 이런 결과가 발생하는 이유는 이러한 효과 중 하나에 대한 효과 검정을 수행하려면 해당 효과를 모형에 마지막으로 입력해야 하기 때문입니다. 그러나 다른 두 효과가 모형 자유도 2와 관련된 모형 제곱합을 완전히 설명합니다. 따라서 관심 효과에 대한 자유도 또는 관련 제곱합이 남아 있지 않습니다.

그림 3.65 Model 1(왼쪽) 및 Model 2(오른쪽) 에 대한 최소 제곱 적합 보고서



반응 Y

특이성 상세 정보

항 상세 정보
X1 = - X2 + X3

적합 요약

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	2	97.800000	48.9000	97.8000
오차	2	1.000000	0.5000	Prob > F
수정 합계	4	98.800000		0.0101*

모수 추정값

항		추정값	표준 오차	t 비	Prob>|t|
절편		11.75	0.433013	27.14	0.0014*
X1	편향	-1.25	0.433013	-2.89	0.1020
X2	편향	-4	0.353553	-11.31	0.0077*
X3	영점화됨	0	0	.	.

효과 검정

효과 상세 정보

반응 Y

특이성 상세 정보

항 상세 정보
X1 =X3 - X2

효과 요약

적합 요약

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	2	97.800000	48.9000	97.8000
오차	2	1.000000	0.5000	Prob > F
수정 합계	4	98.800000		0.0101*

모수 추정값

항		추정값	표준 오차	t 비	Prob>|t|
절편		11.75	0.433013	27.14	0.0014*
X1	편향	2.75	0.661438	4.16	0.0533
X3	편향	-4	0.353553	-11.31	0.0077*
X2	영점화됨	0	0	.	.

효과 검정

효과 상세 정보

LostDFs

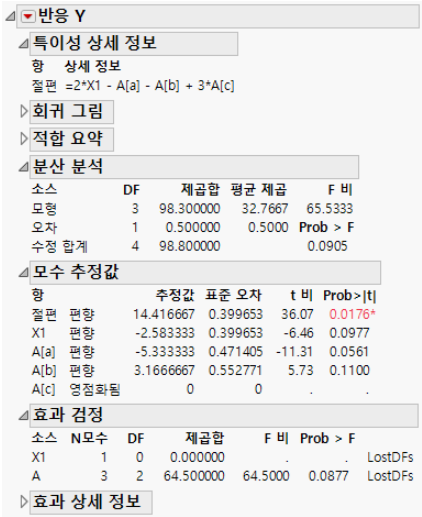
LostDFs 에 대해 더 자세히 알아보려면 아래 단계를 수행하거나 Fit Model Report 데이터 테이블 스크립트를 실행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Singularity.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 클릭합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. X1 과 A 를 선택하고 **추가**를 클릭합니다 .
5. **강조**를 **최소 보고서**로 설정합니다 .
6. **실행**을 클릭합니다 .

그림 3.66에 보고서의 일부가 표시되어 있습니다. "특이성 상세 정보" 보고서에서는 X1 및 A 효과와 연결된 3개 항과 관련하여 선형 종속성이 있음을 보여 줍니다. 명목형 효과를 코드화하는 방법에 대한 자세한 내용은 "사용자 검정 예에 대한 상세 정보"(188페이지)에서 확인하십시오. "분산 분석" 보고서에서는 모형 자유도가 3이라는 것을 보여 줍니다. "모수 추정값" 보고서에서는 3개의 항 X1, A[a], A[b]에 대한 "편향" 추정값과 4번째 항 A[c]에 대한 "영점화됨" 추정값을 보여 줍니다.

"효과 검정" 보고서에서는 X1을 검정할 수 없음을 보여 줍니다. A를 먼저 입력하고 A가 세 개의 모형 자유도를 설명해야 하기 때문입니다. 그러나 A는 검정할 수 있지만 자유도가 두 개뿐입니다. X1을 먼저 입력하고 모형 자유도 중 하나를 설명해야 합니다. A에 대한 검정은 부분적이므로 신중하게 해석해야 합니다.

그림 3.66 X1 및 A 를 사용한 모형의 최소 제곱 적합 보고서


반응 Y

특이성 상세 정보

항	상세 정보
절편	=2*X1 - A[a] - A[b] + 3*A[c]

회귀 그림

적합 요약

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	3	98.300000	32.7667	65.5333
오차	1	0.500000	0.5000	Prob > F
수정 합계	4	98.800000		0.0905

모수 추정값

항		추정값	표준 오차	t 비	Prob>|t|
절편	편향	14.416667	0.399653	36.07	0.0176*
X1	편향	-2.583333	0.399653	-6.46	0.0977
A[a]	편향	-5.333333	0.471405	-11.31	0.0561
A[b]	편향	3.1666667	0.552771	5.73	0.1100
A[c]	영점화됨	0	0	.	.

효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F	
X1	1	0	0.000000	.	.	LostDFs
A	3	2	64.500000	64.5000	0.0877	LostDFs

효과 상세 정보

표준 최소 제곱 분석법에 대한 통계 상세 정보

- " 강조 규칙 "
- " 사용자 검정 에에 대한 상세 정보 "
- " 추정값 상관계수 "
- " 레버리지 그림 상세 정보 "
- "Kackar-Harville 수정 "
- " 검정력 분석 "

강조 규칙

" 모형 적합 " 시작 창의 기본 강조는 행 수 (n), " 모형 효과 생성 " 목록에 입력된 효과 수 (k) 및 효과에 적용된 속성을 기반으로 합니다 .

- $n > 1000$ 이면 " 강조 " 가 " 최소 보고서 " 로 설정됩니다 .
- $n \leq 1000, k \leq 4$ 이면 " 강조 " 가 " 효과 레버리지 " 로 설정됩니다 .
- $n \leq 1000, k \geq 10$ 이면 " 강조 " 가 " 효과 선별 " 로 설정됩니다 .

- $n \leq 1000, 4 < k < 10$ 및 $n - k > 20$ 이면 " 강조 " 가 " 효과 레버리지 " 로 설정됩니다 .
- 효과에 " 임의 효과 " 속성이 있으면 " 강조 " 가 " 최소 보고서 " 로 설정됩니다 .
- 이러한 조건 중 어느 것도 해당되지 않으면 " 강조 " 가 " 효과 선별 " 로 설정됩니다 .

사용자 검정 예에 대한 상세 정보

" 사용자 검정의 예 "(116 페이지) 에서는 Cholesterol.jmp 샘플 데이터 테이블을 사용하여 세 가지 대비를 검정합니다 . 구체적으로 다음 내용을 비교하려고 합니다 .

- treatment A 의 평균 반응 대 treatment B 의 평균 반응
- 결합된 treatment A/B 의 평균 반응 대 대조군의 평균 반응
- 결합된 treatment A/B 의 평균 반응 대 결합된 대조군 / 플라시보 그룹의 평균 반응

"사용자 검정" 열에 입력하는 대비 계수를 유도하기 위해 다음을 수행하십시오. 4개의 treatment 그룹에 대한 이론적 효과를 α_A, α_B, $\alpha_{Control}$ 및 $\alpha_{Placebo}$ 로 나타냅니다. 이것은 treatment 효과이므로 합이 0으로 제한됩니다. 표시 변수와 관련된 모수는 처음 세 효과만 나타내므로 처음 세 효과를 기준으로 대비를 공식화해야 합니다. 자세한 내용은 "사용자 검정 예에 대한 상세 정보"(188 페이지) 및 "통계 상세 정보" 부록의 "모수 해석"(493 페이지)에서 확인하십시오.

검정할 가설은 다음과 같이 모형 효과를 기준으로 작성할 수 있습니다 .

- treatment A 와 treatment B 비교 :

  $$\alpha_A - \alpha_B = 0$$

- treatment A/B 와 대조군 비교 :

  $$0.5(\alpha_A + \alpha_B) - \alpha_{Control} = 0$$

- treatment A/B 와 대조군 / 플라시보 그룹 비교 :

  $$0.5(\alpha_A + \alpha_B) - 0.5(\alpha_{Control} + \alpha_{Placebo}) = \alpha_A + \alpha_B = 0$$

  이 대비를 위한 대비 계수를 얻으려면 모형 효과를 기준으로 placebo 효과를 작성해야 합니다 . 구체적으로 $\alpha_A + \alpha_B + \alpha_{Control} + \alpha_{Placebo} = 0$ 이라는 사실을 사용합니다 . 그러면 $\alpha_{Placebo} = -\alpha_A - \alpha_B - \alpha_{Control}$ 입니다 . 결론은 다음과 같습니다 .

$$\begin{aligned} 0.5(\alpha_A + \alpha_B) - 0.5(\alpha_{Control} + \alpha_{Placebo}) &= 0.5(\alpha_A + \alpha_B) - 0.5(\alpha_{Control} - \alpha_A - \alpha_B - \alpha_{Control}) \\ &= 0.5(\alpha_A + \alpha_B) - 0.5\alpha_{Control} + 0.5(\alpha_A + \alpha_B + \alpha_{Control}) \\ &= \alpha_A + \alpha_B \\ &= 0 \end{aligned}$$

추정값 상관계수

관측값 n개와 예측 변수 $p - 1$개가 있는 데이터 집합을 예로 들겠습니다. 행렬 **X**를 설계 행렬로 정의합니다. 즉, **X**는 첫 번째 열이 *1*로 구성되고, 나머지 $p - 1$개 열이 $p - 1$개의 예측 변수 값으로 구성된 $n \times p$ 행렬입니다. 명목형 열은 표시자 예측 변수를 기준으로 코드화됩니다. 각 열은 행렬 X의 열입니다.

회귀 계수 벡터의 추정값은 다음과 같습니다.

$$\hat{\beta} = (X'X)^{-1}X'Y$$

여기서 Y는 반응 값의 벡터를 나타냅니다.

일반적인 회귀 가정에서 $\hat{\beta}$의 공분산 행렬은 다음과 같습니다.

$$Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

여기서 σ^2은 반응의 분산을 나타냅니다.

추정값에 대한 상관 행렬은 공분산 행렬의 각 항목을 대각 항목의 제곱근의 곱으로 나누어 얻습니다. 다음과 같이 V를 대각 행렬이 되도록 정의합니다. 이 행렬의 각 항목은 공분산 행렬의 대각 항목에 대한 제곱근입니다.

$$V = Sqrt(Diag(Cov(\hat{\beta})))$$

그러면 모수 추정값에 대한 상관 행렬이 다음과 같이 지정됩니다.

$$Corr(\hat{\beta}) = \sigma^2 V^{-1}(X'X)^{-1}V^{-1}$$

레버리지 그림 상세 정보

효과 레버리지 그림은 **편회귀 잔차 레버리지 그림**(Belsley et al. 1980) 또는 **추가 변수 그림**(Cook and Weisberg 1982)이라고도 합니다. Sall(1990)에서는 이러한 그림을 선형 가설에 적용하도록 일반화했습니다.

JMP 에서는 다음과 같은 두 가지 유형의 레버리지 그림을 제공합니다.

- 효과 레버리지 그림에는 다른 모든 효과가 모형에 있을 때 해당 효과가 모형에 없다는 가설을 기준으로 관측값이 표시됩니다.
- "실제값 대 예측값 그림" 보고서에 나오는 전체 모형 레버리지 그림에는 요인 효과가 없다는 가설을 기준으로 관측값이 표시됩니다.

효과 레버리지 그림에서는 하나의 효과만 0으로 가정됩니다. 그러나 전체 모형에 대한 실제값 대 예측값 그림에서는 모든 효과가 0으로 가정됩니다. Sall(1990)에서는 레버리지 그림의 개념을 임의의 선형 가설로 일반화하며, 전체 모형 레버리지 그림이 그 예입니다. 이 섹션에 요약된 해당 연구 자료의 상세 정보는 JMP에 있는 두 가지 유형의 그림에 특화되어 있습니다.

생성

추정 가능한 관심 가설이 다음과 같다고 가정해 보겠습니다 .

$$L\beta = 0$$

레버리지 그림은 각 점에서 경사진 회귀선까지의 거리가 무제약 잔차를 표시하도록 점을 그려 이 검정의 특징을 나타냅니다. 0에서 수평선까지의 거리는 적합이 가설의 제약을 받을 때 잔차를 표시합니다. 이 두 잔차 집합의 제곱합 간의 차이는 가설로 인한 제곱합입니다. 이 값은*F*-검정의 주성분이 됩니다.

가설의 제약을 받는 모수 추정값을 다음과 같이 작성할 수 있습니다 .

$$b_0 = b - (X'X)^{-1}L'\lambda$$

여기서 b 는 최소 제곱 추정값입니다 .

$$b = (X'X)^{-1}X'y$$

람다는 가설 제약 조건에 대한 라그랑즈 승수이며 다음과 같이 계산됩니다 .

$$\lambda = (L(X'X)^{-1}L')^{-1}Lb)$$

무제약 잔차와 가설의 제약을 받는 잔차는 각각 다음과 같습니다 .

$$r = y - Xb$$

$$r_0 = r + X(X'X)^{-1}L'\lambda$$

각 관측값에 대해 가로 축 값이 v_x 이고 세로 축 값이 v_y 인 점을 가정해 보겠습니다 . 다음은 각 요소에 대한 설명입니다 .

- v_x 는 제약된 잔차에서 무제약 잔차를 뺀 $r_0 - r$ 이며 , 제약 조건이 적용된 후 남은 정보를 반영합니다 .
- v_y 는 가로 축 값에 무제약 잔차를 더한 값입니다 .

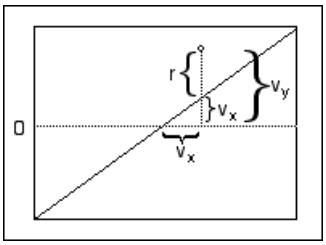
따라서 이 점의 x 및 y 좌표는 다음과 같습니다 .

$$v_x = X(X'X)^{-1}\mathbf{L}'\lambda \text{ 및 } v_y = r + v_x$$

이러한 점을 기반으로 레버리지 그림이 생성됩니다. 그림 3.67에서는 평균 반응이 0이고 실선 기울기가 1인 레버리지 그림 생성 구조를 보여 줍니다.

JMP의 레버리지 그림에는 반응 평균 $\bar{y}$ 에 수평 점선이 있습니다. 그림에 표시된 점은 $(v_x + \bar{y}, v_y)$ 로 지정됩니다.

그림 3.67 레버리지 그림 생성



레버리지 그림에 검정 중첩

단순 선형 회귀에서 반응 기대값에 대한 신뢰 한계를 예측 변수 **x** 의 평활 함수로 그림에 표시할 수 있습니다 .

$$\text{Upper(x)} = \mathbf{xb} + t_{\alpha/2} s \sqrt{\text{x}(\text{X}'\text{X})^{-1}\text{x}'}$$

$$\text{Lower(x)} = \mathbf{xb} - t_{\alpha/2} s \sqrt{\text{x}(\text{X}'\text{X})^{-1}\text{x}'}$$

여기서 $\mathbf{x} = [1\ x]$ 는 예측 변수의 2- 벡터입니다 .

그림 3.56 에 표시된 것처럼 이러한 신뢰 곡선을 통해 해당 가설 검정의 유의성을 시각적으로 평가할 수 있습니다 .

- 유의함 : 기울기 모수가 0 과 유의한 차이가 있으면 신뢰 곡선이 반응 평균에서 수평선과 교차합니다 .
- 경계선 : 기울기 모수에 대한 *t*- 검정이 유의성 한계에 있으면 신뢰 곡선이 반응 평균에서 수평선에 점근적입니다 .
- 유의하지 않음 : 기울기 모수가 0 과 유의한 차이가 없으면 신뢰 곡선이 반응 평균에서 수평선과 교차하지 않습니다 .

레버리지 그림은 신뢰 곡선을 표시하여 이 개념을 반영합니다 . 그림이 적절하게 중앙에 오도록 조정됩니다 . 가로 축의 점을 *z* 로 나타냅니다 . 함수를 다음과 같이 정의합니다 .

$$\text{Upper}(z) = \text{z} + \sqrt{s^2 t^2_{\alpha/2} \bar{h} + (F_\alpha / F) z^2}$$

및

$$\text{Lower}(z) = \text{z} - \sqrt{s^2 t^2_{\alpha/2} \bar{h} + (F_\alpha / F) z^2}$$

여기서 *F* 는 가설에 대한 *F* 통계량이고 F_α 는 유의 수준 α 에 대한 참조 값입니다 .

또한 $\bar{h} = \bar{\text{x}}(\text{X}'\text{X})^{-1}\bar{\text{x}}'$ 입니다 . 여기서 $\bar{\text{x}}$ 는 예측 변수에 대한 적절한 중간 값 (예 : 평균) 으로 구성된 행 벡터입니다 .

이러한 함수는 다음과 같이 단순 선형 회귀에 대한 신뢰 곡선과 동일한 방식으로 작동합니다 .

- *F* 통계량이 참조 값보다 크면 신뢰도 함수가 가로 축과 교차합니다 .
- *F* 통계량이 참조 값과 같으면 신뢰도 함수의 가로 축이 점근선으로 표시됩니다 .
- *F* 통계량이 참조 값보다 작으면 신뢰도 함수가 교차하지 않습니다 .

또한 Upper(z) - Lower(z)가 z에서의 예측값에 대한 유효한 신뢰 구간이어야 합니다.

Kackar-Harville 수정

고정 효과의 분산 행렬은 항상 Kackar-Harville 수정을 포함하도록 변경됩니다. BLUP의 분산 행렬 및 BLUP와 고정 효과 간의 공분산은 Kackar-Harville 수정 대상이 아닙니다. 이 방법의 근거는 임의 효과에 수준이 많이 있을 때 BLUP에 대한 수정은 계산 및 메모리 사용량이 많을 수 있다는 것입니다.

SAS에서는 DDFM=KENWARDROGER가 설정된 경우에만 Kackar-Harville 수정이 고정 효과와 BLUP에 대해 모두 수행됩니다 .

- 고정 효과 모수만 포함하는 선형 결합의 표준 오차는 PROC MIXED DDFM=KENWARDROGER 와 일치합니다 . 이 경우 PROC MIXED 와 JMP 에서 사용하는 서로 다른 파라미터화 간의 변환에 주의를 기울여야 합니다 .
- BLUP 모수만 포함하는 선형 결합의 표준 오차는 PROC MIXED DDFM=SATTERTHWAITE 와 일치합니다 .
- 불균형 데이터의 경우 고정 효과와 BLUP를 모두 포함하는 선형 결합의 표준 오차는 DDFM 옵션에 대한 PROC MIXED 와 일치하지 않습니다 . 그러나 이러한 표준 오차는 DDFM=SATTERTHWAITE 및 DDFM=KENWARDROGER 옵션을 사용한 결과 사이에 있습니다 . 균형 데이터의 경우 Kackar-Harville 수정이 null 이므로 JMP 는 DDFM 옵션에 관계없이 SAS 와 일치합니다 .

자유도

고정 효과 모수의 선형 결합만 포함하는 검정의 자유도는 Kenward-Roger 수정을 사용하여 계산됩니다. 따라서 이러한 검정의 JMP 결과는 DDFM=KENWARDROGER 옵션을 사용하는 PROC MIXED 와 일치합니다. 선형 결합에 BLUP가 있는 경우 JMP는 Satterthwaite 근사를 사용하여 자유도를 구합니다. 결과는 이전 단락에서 표준 오차에 대해 설명한 것과 유사한 패턴을 따릅니다.

Kackar-Harville 수정 및 Kenward-Roger DF 방법에 대한 자세한 내용은 Kenward and Roger 연구 자료(1997)에서 확인하십시오. Satterthwaite 방법은 SAS Institute Inc. (2020d, ch. 83).

검정력 분석

검정력 계산과 관련된 옵션은 연속형 반응 모형에만 사용할 수 있습니다 . 다음과 같은 컨텍스트에서 검정력 및 관련된 검정 상세 정보를 사용할 수 있습니다 .

모수 추정값

각 모수 추정값에 대한 후향적 검정 상세 정보를 얻으려면 보고서의 빨간색 삼각형 메뉴에서 **추정값 > 모수 검정력**을 선택합니다. 이 옵션은 현재 연구 데이터를 기반으로 각 모수에 대한 0.05 유의 수준 검정의 최소 유의값, 최소 유의 개수 및 조정 검정력을 표시합니다.

효과 또는 효과 상세 정보

특정 효과의 *F*-검정에 대한 후향적 또는 전향적 상세 정보를 얻으려면 효과의 빨간색 삼각형 메뉴에서 **검정력 분석**을 선택합니다. 효과 선별 및 최소 보고서 분석법의 경우 각 효과에 대한 보고서는 "효과 상세 정보" 아래에 있습니다. 효과 레버리지 분석법의 경우 효과에 대한 보고서는 보고서의 첫 번째(전체 모형) 열 오른쪽에 있습니다.

최소 제곱 평균 대비

하나 이상의 대비 검정에 대한 후향적 또는 전향적 상세 정보를 얻으려면 효과의 빨간색 삼각형 메뉴에서 **최소 제곱 평균 대비**를 선택합니다. 원하는 대비를 정의하고 "완료"를 클릭합니다. "대비" 빨간색 삼각형 메뉴에서 **검정력 분석**을 선택합니다.

사용자 검정

사용자 검정에 대한 후향적 또는 전향적 상세 정보를 얻으려면 반응의 빨간색 삼각형 메뉴에서 **추정값 > 사용자 검정**을 선택합니다. 원하는 대비를 정의하고 **완료**를 클릭합니다. "사용자 검정" 빨간색 삼각형 메뉴에서 **검정력 분석**을 선택합니다.

첫 번째를 제외한 모든 경우에 "검정력 분석"을 선택하면 "검정력 상세 정보 창"이 열립니다. 그러면 "검정력 상세 정보 창"에 정보를 입력하여 필요에 따라 계산을 수정할 수 있습니다.

효과 크기

효과 크기(δ)는 귀무가설과 관련 모수의 실제 값 간의 차이를 측정한 것입니다. 귀무가설은 0으로 설정된 단일 선형 대비 또는 이러한 대비 여러 개로 공식화할 수 있습니다. δ의 값은 대비의 실제 값과 가설 값 0 간의 차이를 반영합니다.

일반적으로 효과 크기는 다음과 같이 지정됩니다 .

$$\delta = \sqrt{SS_{Hyp(Pop)}/n}$$

여기서 $SS_{Hyp(Pop)}$ 는 모집단 모수로 주어진 검정 대상 가설의 제곱합이고 n은 총 관측값 수입니다.

가능한 관측값이 있는 경우 추정된 효과 크기는 가설에 대해 계산된 제곱합을 δ의 계산식에 대입하여 계산됩니다.

균형 일원 배치

*i*번째 그룹의 평균 반응이 α_i인 *k*개 수준의 균형 일원 배치에 대한 특수한 경우를 예로 들겠습니다.

$$\delta^2 = \frac{\sum(\alpha_i - \bar{\alpha})^2}{k}$$

JMP 모수를 코드화하여 i=1, 2,..., k-1 에 대해 다음과 같이 됩니다 .

$$\beta_i = (\alpha_i - \bar{\alpha})$$

및

$$\beta_k = -\sum_{m=1}^{k-1} \alpha_m$$

따라서 이러한 모수에 관해 2 수준 균형 배치에 대한 δ 는 다음과 같이 제공됩니다 .

$$\delta^2 = \frac{\beta_1^2 + (-\beta_1)^2}{2} = \beta_1^2$$

또는 $\delta = |\beta_1|$

불균형 일원 배치

*k*개 수준이 있는 불균형 일원 배치에서 *i*번째 그룹의 평균 반응이 α_i, n_i는 관측값, $n = \sum n_i$ 인 경우:

$$\delta^2 = \sum \frac{n_i}{n}(\alpha_i - \bar{\alpha})^2$$

효과 크기 및 검정력

검정력은 실제 효과 크기가 지정된 값일 때 α 유의 수준에서 가설에 대한 *F*- 검정이 유의할 확률입니다 . 실제 효과 크기가 δ 인 경우 검정 통계량은 다음과 같이 계산된 비중심성 모수를 갖는 비중심 *F* 분포를 따릅니다 .

$$\lambda = (n\delta^2)/\sigma^2$$

귀무가설이 참인 경우, 즉 효과 크기가 0 이면 비중심성 모수는 0 이고 검정 통계량은 중심 *F* 분포를 따릅니다.

검정력은 λ에 따라 증가합니다. 특히 검정력은 표본 크기 *n* 과 효과 크기 δ, 에 따라 증가하고 오차 분산 σ^2에 따라 감소합니다.

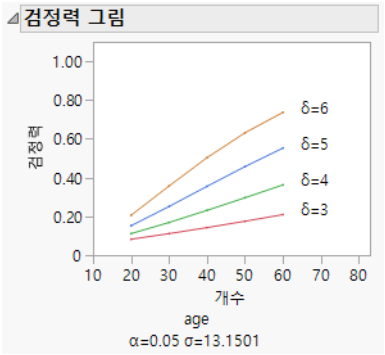
Cohen(1977) 과 같은 일부 참고 도서에는 JMP 에서 사용하는 원시 효과 크기 대신 표준화된 효과 크기 $\Delta = \delta/\sigma$ 가 사용됩니다. 표준화된 효과 크기의 경우 비중심성 모수는 $\lambda = n\Delta^2$ 입니다.

"검정력 상세 정보 창"에서 δ는 처음에 $\sqrt{SS_{Hyp}/n}$ 로 설정되어 있습니다. SS_{Hyp}는 가설에 대한 제곱합이고 n은 현재 연구의 관측값 수입니다. SS_{Hyp}는 데이터에서 계산된 δ의 추정값이지만 이러한 추정값은 편향되어 있습니다(Wright and O'Brien 1988). δ에 대한 표본 추정값을 사용하여 검정력을 계산하려면 "검정력 계산" 대신 "조정 검정력 및 신뢰 구간" 계산을 사용할 수 있습니다. 조정 검정력 계산은 편향이 부분적으로 수정된 δ의 추정값을 사용합니다. 자세한 내용은 "통계 상세 정보" 부록의 "조정 검정력 계산"(518 페이지)에서 확인하십시오.

표본 크기별 검정력 그림

표본 크기별 검정력 그림을 보려면 "검정력" 보고서 맨 아래의 빨간색 삼각형 메뉴에서 "검정력 그림" 옵션을 선택합니다. 그러면 "검정력" 테이블의 "검정력" 및 "개수" 열이 그림으로 표시됩니다. 그림 3.68에 표시된 그림은 "후향적 검정력 분석의 예"(197 페이지)에서 얻은 "검정력" 테이블을 그림으로 나타낸 결과입니다.

그림 3.68 표본 크기별 검정력 그림



LSN(최소 유의 개수)

LSN(최소 유의 개수)은 델타, 시그마 및 알파 값이 주어졌을 때 유의한 검정 결과를 생성하는 최소 관측값 수입니다. 델타, 시그마 및 알파는 각각 효과 크기, 오차 표준편차 및 유의 수준을 나타냅니다.

참고 : LSN 은 유의 확률을 고려하지 않으므로 권장되는 표본 크기가 아닙니다 . 이 값은 지정된 델타 및 시그마 값을 기반으로 계산됩니다 .

LSN 에는 다음과 같은 특성이 있습니다 .

- LSN 이 실제 표본 크기 n 보다 작으면 효과가 유의합니다 .
- LSN 이 n 보다 크면 효과가 유의하지 않습니다 . 추가 데이터가 현재 표본과 동일한 구조 결과를 나타낼 것이라고 생각하는 경우 LSN 은 유의성을 달성하는 데 필요한 데이터 양을 제안합니다 .
- LSN 이 n 과 같으면 p 값이 유의 수준 알파와 같습니다 . 이 경우 검정은 유의성의 경계에 있습니다 .

- n = LSN 일 때 계산된 효과 크기 검정의 검정력은 항상 0.5 보다 크거나 같습니다 . 그러나 검정력은 0.5 에 가까울 수 있으며 이는 계획 목적상 낮은 것으로 간주됩니다 .

LSV(최소 유의값)

LSV(최소 유의값) 는 단일 자유도 가설 검정에 대해 계산됩니다. 여기에는 좀 더 일반적인 선형 대비는 물론 개별 모형 모수의 유의성에 대한 검정이 포함됩니다. LSV 는 지정된 유의 수준에서 유의한 것으로 간주될 최소 효과 크기 (절대값) 입니다. LSV 는 확률 척도가 아니라 모수 척도에 따라 검정의 민감도를 측정합니다.

LSV 에는 다음과 같은 특성이 있습니다 .

- 모수 추정값 또는 대비의 절대값이 LSV 보다 크거나 같으면 유의성 검정의 p 값이 알파보다 작거나 같습니다 .
- 모수 추정값 또는 대비의 절대값은 유의성 검정의 p 값이 알파와 같을 때만 LSV 와 같습니다.
- LSV 는 모수 또는 모수의 선형 결합에 대한 $1 - \alpha$ 신뢰 구간의 반지름입니다 . $1 - \alpha$ 신뢰 구간은 모수 또는 대비의 추정값에서 중심화됩니다 .

검정력

검정력은 해당 검정이 유의한 결과를 제공할 확률입니다. 검정력은 효과 크기 δ, 유의 수준 α, 오차 표준편차 σ 및 표본 크기 n 의 함수입니다. 검정력은 지정된 유의 수준에서 지정된 효과 크기를 감지할 확률입니다. 일반적으로 실제적 또는 과학적 중요도의 차이를 감지할 검정력이 높은 연구를 설계하기를 원합니다.

검정력에는 다음과 같은 특성이 있습니다 .

- 모수의 실제 값이 가설 값이면 검정력이 검정의 유의 수준과 같습니다 . 일반적으로 유의 수준은 0.05 와 같이 작은 값입니다 . 가정된 가설이 참일 때 유의한 결과를 얻을 확률이 낮아야 하므로 작은 값이 적절합니다 .
- 일반적으로 모수의 실제 값이 가설 값이 아니면 검정력이 가능한 한 커야 합니다 .
- 표본 크기가 증가하고 , 오차 분산이 감소하고 , 실제 모수 값과 가설 값의 차이가 증가할수록 검정력이 증가합니다 .

조정 검정력 및 신뢰 구간

후향적 검정력 분석에서는 일반적으로 표본 추정값이 검정력 계산과 관련된 모집단 모수를 대체합니다. 이로 인해 비중심성 모수 추정값이 양의 편향을 갖게 됩니다 (Wright and O'Brien 1988). 조정 검정력 계산은 이 편향이 부분적으로 수정된 비중심성 모수의 추정값을 기반으로 합니다.

조정 검정력에 대한 신뢰 구간을 생성할 수도 있습니다. 이 신뢰 구간은 넓은 경향이 있습니다. 자세한 내용은 Wright and O'Brien 연구 자료 (1988) 에서 확인하십시오.

조정 검정력 및 신뢰 구간 계산은 데이터에서 추정된 δ 값(기본적으로 제공되는 값)에만 관련됩니다. 다른 델타 값의 경우 조정 검정력 및 신뢰 구간이 제공되지 않습니다.

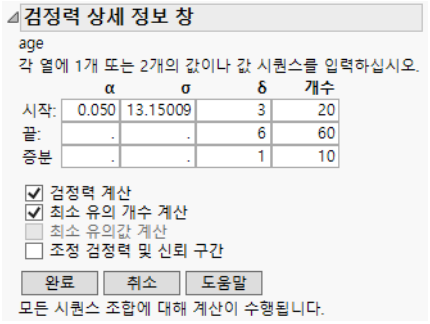
자세한 내용은 "통계 상세 정보" 부록의 "조정 검정력 계산"(518 페이지)에서 확인하십시오.

후향적 검정력 분석의 예

이 예에서는 Big Class.jmp 샘플 데이터 테이블을 사용한 후향적 검정력 분석을 보여 줍니다. "검정력 상세 정보 창"(그림 3.69)에서는 α, σ, δ 및 개수(연구 크기)에 대한 값 범위에서 다양한 통계량을 탐색할 수 있습니다. "완료"를 클릭하면 창이 계산 결과로 바뀝니다.

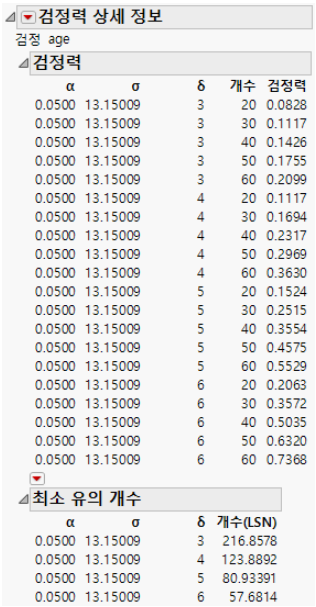
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. weight 를 선택하고 **Y** 를 클릭합니다.
4. age, sex 및 height 를 효과로 추가합니다.
5. **실행**을 클릭합니다.
6. "age" 의 빨간색 삼각형을 클릭하고 **검정력 분석**을 선택합니다.

그림 3.69 age 에 대한 검정력 상세 정보 창



7. 그림 3.69 에 표시된 대로 **시작** 상자의 δ 값을 3 으로 바꾸고 **끝** 및 **증분** 상자에 각각 6 과 1 을 입력합니다.
8. 그림 3.69 에 표시된 대로 **시작** 상자의 **개수** 값을 20 으로 바꾸고 **끝** 및 **증분** 상자에 각각 60 과 10 을 입력합니다.
9. **검정력 계산** 및 **최소 유의 개수 계산**을 선택합니다.
10. **완료**를 클릭합니다.
11. "검정력 상세 정보 창" 이 "검정력 상세 정보" 보고서로 바뀝니다.

그림 3.70 age 에 대한 검정력 상세 정보 보고서



검정력 상세 정보

검정 age

검정력

α	σ	δ	개수	검정력
0.0500	13.15009	3	20	0.0828
0.0500	13.15009	3	30	0.1117
0.0500	13.15009	3	40	0.1426
0.0500	13.15009	3	50	0.1755
0.0500	13.15009	3	60	0.2099
0.0500	13.15009	4	20	0.1117
0.0500	13.15009	4	30	0.1694
0.0500	13.15009	4	40	0.2317
0.0500	13.15009	4	50	0.2969
0.0500	13.15009	4	60	0.3630
0.0500	13.15009	5	20	0.1524
0.0500	13.15009	5	30	0.2515
0.0500	13.15009	5	40	0.3554
0.0500	13.15009	5	50	0.4575
0.0500	13.15009	5	60	0.5529
0.0500	13.15009	6	20	0.2063
0.0500	13.15009	6	30	0.3572
0.0500	13.15009	6	40	0.5035
0.0500	13.15009	6	50	0.6320
0.0500	13.15009	6	60	0.7368

최소 유의 개수

α	σ	δ	개수(LSN)
0.0500	13.15009	3	216.8578
0.0500	13.15009	4	123.8892
0.0500	13.15009	5	80.93391
0.0500	13.15009	6	57.6814

이 분석은 계산 시 Big Class.jmp 샘플 데이터 테이블의 구조와 동일한 구조의 연구를 가정하므로 후향적 검정력 분석입니다. 예를 들어 이 예의 검정력 계산은 모형에 입력된 효과와 각 연령(age) 및 성별(sex) 그룹의 참가자 수에 따라 달라집니다. 또한 σ 값은 향후 연구의 대표 값으로 대체할 수 있지만 현재 연구에서 도출되었습니다.

그림 3.70에 표시된 검정력 결과에 대한 자세한 내용은 "검정력"(196 페이지)에서 확인하십시오. LSN(최소 유의 개수)에 대한 자세한 내용은 "LSN(최소 유의 개수)"(195 페이지)에서 확인하십시오.

전향적 검정력 분석

전향적 분석은 "지정된 크기의 차이가 존재하는 경우 제안된 표본 크기, 유의 수준 및 오차 분산 추정값이 주어지면 차이를 감지할 수 있습니까?"라는 질문에 답하는 데 도움이 됩니다. 전향적 검정력 분석에서는 그룹 평균 및 표본 크기의 추정값을 데이터 테이블에 제공해야 합니다. 또한 "검정력 상세 정보 창"에서 오차 표준편차 σ의 추정값도 제공해야 합니다.

동일한 그룹 크기

독립된 세 개 그룹의 평균을 비교하는 경우를 가정해 보겠습니다. 지정된 검정력을 달성하기 위한 표본 크기를 계산하려면 **DOE > 표본 크기 및 검정력**을 선택한 후 **k 표본 평균**을 선택합니다. 그런 다음 "표준편차" 옆에 오차 표준편차의 추정값을 입력합니다. 감지할 최소 차이를 반영하는 평균을 전향적 평균 목록에 입력합니다. 예를 들어 두 평균 사이의 8 단위 차이를 감지하려면 평균의 극단값을 입력합니다(예: 40, 40, 48). 검정력은 총 평균으로부터의 편차를 기반으로 하므로 원하는 차이를 반영하는 값만 입력할 수 있습니다(예: 0, 0, 8).

계속을 클릭하면 검정력 대 표본 크기를 보여 주는 그래프가 표시됩니다. 대신 "표본 크기" 창에서 검정력 또는 표본 크기를 지정하면 다른 통계량이 계산되어 "표본 크기" 창에 표시됩니다. 특히 검정력을 지정할 경우 제공되는 표본 크기는 필요한 총 표본 크기입니다. k 표본 평균 계산은 동일한 그룹 크기를 가정합니다. 세 그룹의 경우 표본 크기를 3으로 나누어 개별 그룹 크기를 구합니다. K 표본 평균에 대한 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

동일하지 않은 그룹 크기

크기가 다른 그룹을 사용하는 연구를 설계한다고 가정해 보겠습니다. 박테리아 수를 줄이는 것으로 알려진 두 가지 처리를 연구하기 위한 실험을 계획해야 합니다. 이러한 처리 효과를 처리가 적용되지 않은 대조군의 결과와 비교하려고 합니다. 또한 둘 중 한 처리 그룹과 대조군의 평균 간에 최소 8 단위의 차이를 감지하려고 합니다. 그러나 대조군은 각 처리 그룹의 두 배여야 합니다. 또한 두 처리 그룹은 크기가 같아야 합니다. 이전 연구에 따르면 오차 표준편차는 약 5 또는 6 입니다.

이 상황에 대한 전향적 검정력 분석을 수행하려면 Bacteria.jmp 샘플 데이터 테이블에 표시된 몇 가지 기본 정보를 포함하는 데이터 테이블을 생성합니다.

그림 3.71 Bacteria.jmp 데이터 테이블

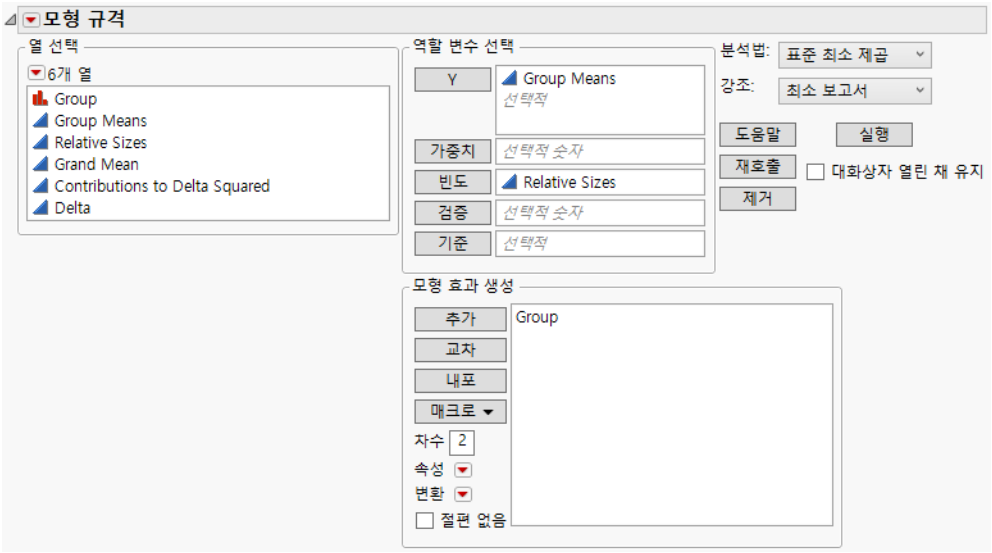
	Group	Group Means	Relative Sizes
1	Control	40	2
2	Treatment A	40	1
3	Treatment B	48	1

- Group 열은 그룹을 식별합니다 .
- Means 열은 열 간에 감지해야 할 최소 차이를 반영합니다 . 여기서 대조군의 평균은 약 40 이라고 가정합니다 . 둘 중 한 처리 그룹의 평균이 대조군의 평균보다 8 단위 이상 크면 검정이 유의하다는 결론을 내리려고 합니다 . 이러한 이유로 두 처리 그룹 중 하나에 평균 48 을 할당하고 다른 처리 그룹의 평균은 대조군의 평균과 동일하게 설정합니다. 아니면 대조군과 처리 그룹 중 하나에 평균 0 을 할당하고 나머지 처리 그룹에 평균 8 을 할당할 수 있습니다 . 그룹 평균의 차이는 모집단 값입니다 .
- Relative Sizes 열에는 처리 그룹의 원하는 상대 크기가 표시됩니다 . 이 열은 대조군이 각 처리 그룹의 두 배여야 함을 나타냅니다 . 아니면 상대 크기 기준을 준수하는 처리 크기에 대한 초기 추측 값으로 시작할 수 있습니다 .

> **참고 :** Relative Sizes 열에는 " 빈도 " 역할이 할당되어야 합니다 . " 열 " 패널에서 열 이름 오른쪽에 있는 기호를 확인하십시오 .

그런 다음 모형 적합을 사용하여 일원 분산 분석 모형을 적합시킵니다(그림 3.72). 시작 창에서 Relative Sizes가 **빈도**로 선언되어 있습니다. 또한 "최소 보고서" 강조 옵션이 선택되어 있습니다.

그림 3.72 박테리아 연구에 대한 모형 적합 시작 창



실행을 클릭하여 " 최소 제곱 적합 " 보고서를 생성합니다 . 그룹 내에 오차 변동이 없는 데이터 테이블을 지정했으므로 이 보고서에는 " 제곱근 평균 제곱 오차 " 및 오차의 제곱합이 0.0 으로 표시됩니다 . 검정력 분석을 얻으려면 오차 변동에 대해 제안된 값 범위를 입력해야 합니다 . 특히 오차 변동이 약 5 에서 최대 6 이 될 수 있다는 정보가 있습니다 .

1. " 효과 상세 정보 " 옆의 표시 아이콘을 클릭하여 이 보고서를 엽니다 .
2. "Group" 의 빨간색 삼각형을 클릭하고 **검정력 분석**을 선택합니다 .
3. 의심되는 오차 변동 범위를 탐색하려면 σ 아래의 첫 번째 상자에 5 를 입력하고 두 번째 상자에 6 을 입력합니다 (그림 3.73).
4. δ 는 3.464102 로 입력되어 있습니다 . 이 값은 그룹 평균의 지정된 차이에 해당하는 효과 크기입니다 . 데이터 테이블에는 효과 크기 계산을 보여 주는 세 개의 숨겨진 열이 포함되어 있습니다 . 자세한 내용은 " 불균형 일원 배치 "(194 페이지) 에서 확인하십시오 .
5. 해당 연구 크기 범위에 대해 검정력을 탐색하려면 **개수** 아래의 첫 번째 상자에 16, 두 번째 상자에 64, 세 번째 상자에 증분 4 를 각각 입력합니다 (그림 3.73).
6. **검정력 계산**을 선택합니다 .
7. **완료**를 클릭합니다 .

그림 3.73 박테리아 연구에 대한 검정력 상세 정보 창

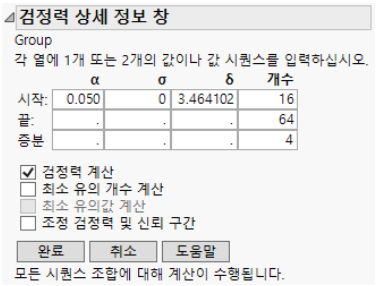
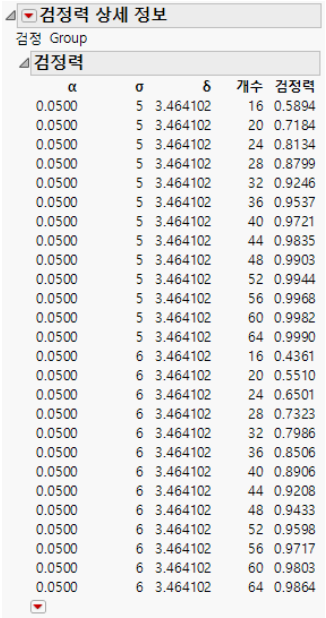


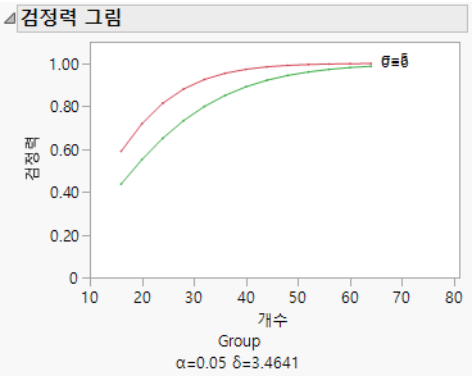
그림 3.74에 표시된 "검정력 상세 정보" 보고서가 "검정력 상세 정보 창"을 대체합니다. 이 보고서에서는 $\alpha = 0.05$, $\sigma = 5$와 6의 모든 조합 및 4씩 증가하는 16 ~ 64의 표본 크기에 대한 검정력 계산을 제공합니다. σ가 5일 때 약 90% 검정력을 얻으려면 약 32개의 표본 크기가 필요합니다. 대조군에 16명, 각 처리 그룹에 8명의 참가자가 필요합니다. 반면에 σ가 6일 때는 총 44명의 참가자가 필요합니다.

그림 3.74 박테리아 연구에 대한 검정력 상세 정보 보고서



"검정력 상세 정보" 보고서의 테이블 맨 아래에 있는 화살표를 클릭하면 그림 3.75에 표시된 것과 같이 σ의 두 값에 대한 검정력 대 표본 크기 그림이 표시됩니다. 여기서 빨간색 표식은 $\sigma = 5$에 해당하고 녹색 표식은 $\sigma = 6$에 해당합니다.

그림 3.75 박테리아 연구에 대한 검정력 그림



4 장

표준 최소 제곱 예
모형의 공통 클래스 분석

이 장에서는 몇 가지 표준 최소 제곱 항목에 대해 지침 자료를 이용한 예를 제공합니다. 여기에는 분산 분석, 공분산 분석, 반응 표면 모형, 분할구 설계, 임의 효과 모수 추정, 매듭이 지정된 스플라인 적합, Bayes 방법을 제공하는 스크립트를 사용한 활성 요인 식별 등이 포함됩니다.

목차

일원 분산 분석 예 205
동일한 기울기의 공분산 분석 예 207
서로 다른 기울기의 공분산 분석 예 209
반응 표면 모형 예 212
전체 반응 표면 모형 적합 212
모형 축소 213
반응 표면 보고서 검토 214
예측 프로파일러를 사용하여 임계점 찾기 216
등고선 프로파일러를 사용하여 표면 보기 217
분할구 설계 예 219
임의 효과 모수 추정 예 223
매듭이 지정된 스플라인 효과 예 225
활성 요인에 대한 Bayes 그림 예 227

일원 분산 분석 예

일원 분산 분석에서는 명목형 변수에 의해 식별된 서로 다른 각 그룹에 서로 다른 평균이 적합됩니다 . JMP 모형을 지정하려면 연속형 반응 열과 명목형 효과 열을 선택합니다 . 이 예에서는 Drug.jmp 의 데이터를 사용합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Drug 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .

이 예에서 Drug 에는 a, d, f 라는 세 가지 수준이 있습니다 . 표준 최소 제곱 적합 방법은 이 규격을 선형 모형으로 변환합니다 . 명목형 변수는 1, 0, –1 값만 가정하는 일련의 표시 변수를 정의합니다 . 선형 모형은 다음과 같이 지정됩니다 .

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

다음은 각 요소에 대한 설명입니다 .

- y_i 는 i 번째 관측값에 대한 관측된 반응입니다 .
- x_{1i} 는 i 번째 관측값에 대한 첫 번째 표시 변수의 값입니다 .
- x_{2i} 는 i 번째 관측값에 대한 두 번째 표시 변수의 값입니다 .
- β_0, β_1, β_2 는 각각 절편 , 첫 번째 표시 변수 , 두 번째 표시 변수에 대한 모수입니다 .
- ε_i 는 독립적이며 정규 분포를 따르는 오차 항입니다 .

첫 번째 표시 변수 x_1 은 다음과 같이 정의됩니다 . 여기서 Drug = a 는 값 1, Drug = d 는 값 0, Drug = f 는 값 –1 을 각각 표시 변수에 적용합니다 .

$$x_{1i} = \begin{cases} 1, & \text{Drug} = \text{a} \\ 0, & \text{Drug} = \text{d} \\ -1, & \text{Drug} = \text{f} \end{cases}$$

두 번째 표시 변수 x_2 에는 다음 값이 제공됩니다 .

$$x_{2i} = \begin{cases} 0, & \text{Drug} = \text{a} \\ 1 & \text{Drug} = \text{d} \\ -1, & \text{Drug} = \text{f} \end{cases}$$

이 파라미터화와 관련하여 세 가지 수준의 평균 추정값은 다음과 같이 정의됩니다 .

$$\mu_a = \beta_0 + \beta_1$$

$$\mu_d = \beta_0 + \beta_2$$

$$\mu_f = \beta_0 - \beta_1 - \beta_2$$

β_i 를 계산하면 다음과 같이 됩니다 .

$$\beta_0 = \frac{(\mu_a + \mu_d + \mu_f)}{3} = \mu \text{ (전체 수준의 평균)}$$

$$\beta_1 = \mu_a - \mu$$

$$\beta_2 = \mu_d - \mu$$

따라서 회귀변수가 각 수준의 표시자에서 마지막 수준의 표시자를 뺀 값으로 코드화되면 한 수준의 모수는 해당 수준의 반응과 모든 수준의 평균 반응 간의 차이로 해석됩니다. 명목형 요인의 파라미터화 해석에 대한 자세한 내용은 부록 "통계 상세 정보"(487 페이지)에서 확인하십시오.

그림 4.1에서는 Drug 효과에 대한 레버리지 그림과 최소 제곱 평균 테이블을 보여 줍니다. 그림 4.2에서는 약물 데이터 일원 분석에 대한 모수 추정값 보고서와 효과 검정 보고서를 보여 줍니다.

그림 4.1 Drug 에 대한 레버리지 그림 및 최소 제곱 평균 테이블

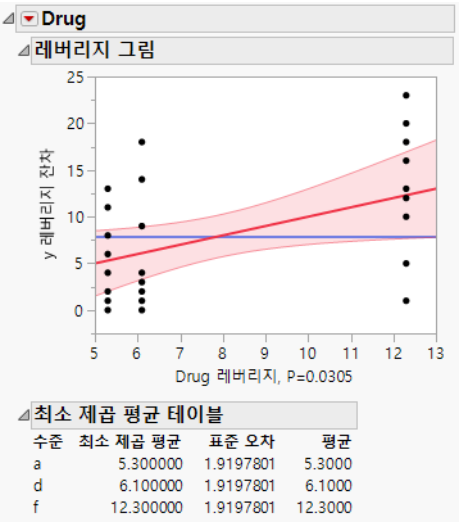
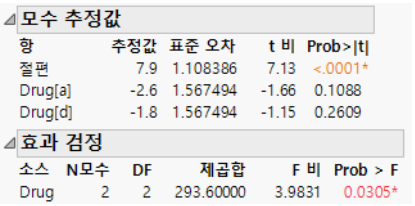


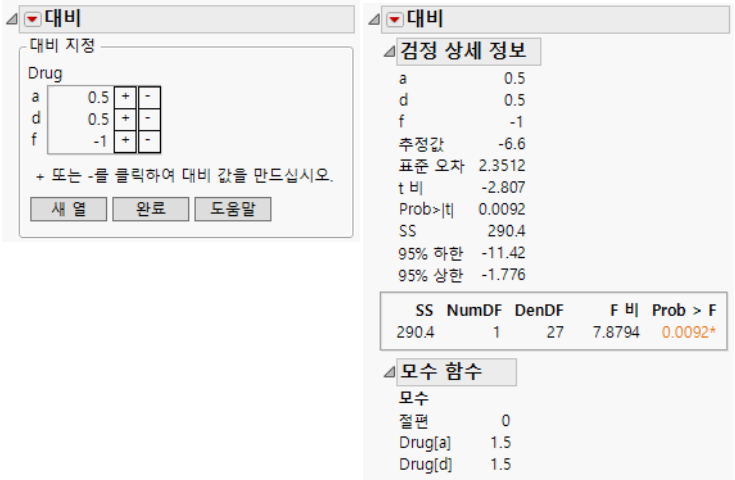
그림 4.2 Drug.jmp 에 대한 모수 추정값 및 효과 검정



다음과 같이 최소 제곱 평균의 대비를 사용하여 Drug 효과를 더 자세히 연구할 수 있습니다 .

1. "Drug" 의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 대비 ...** 를 선택합니다 .
2. Drug a 및 d 의 + 상자를 클릭하고 Drug f 의 - 상자를 클릭하여 약물 a 와 d 의 평균을 f 와 비교하는 대비를 정의합니다 (그림 4.3 참조).
3. **완료**를 클릭합니다 .

그림 4.3 Drug 실험에 대한 대비 예



"대비" 보고서에서는 약물 f의 최소 제곱 평균이 다른 두 약물에 대한 최소 제곱 평균의 평균과 유의하게 다르다는 것을 보여 줍니다.

동일한 기울기의 공분산 분석 예

연속형 회귀변수 항이 있는 분산 분석 모형을 **공분산 분석**이라고 합니다 . Drug.jmp 샘플 데이터 테이블에서 x 열은 공변량입니다 .

공변량은 모형 방정식에 x_{3i} 항을 추가합니다 . 공분산 분석 모형은 다음과 같이 작성됩니다 .

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

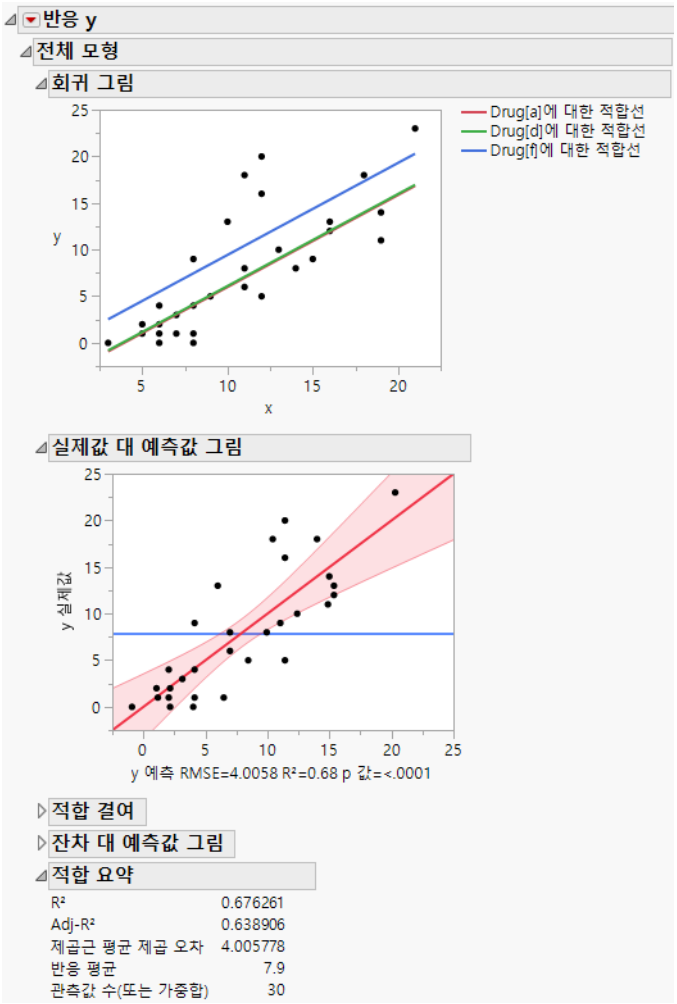
두 가지 모형 효과가 있습니다. 하나는 두 모수를 포함하는 명목형 주효과이고 다른 하나는 한 모수와 관련된 연속형 공변량입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Drug 와 x 를 모두 선택하고 **추가**를 클릭합니다 .

5. **실행**을 클릭합니다.

보고서의 "회귀 그림"에서는 기울기가 동일한 모형이 적합되었음을 보여 줍니다(그림 4.4). 주효과 모형(Drug 효과만)과 비교하여 R^2은 22.8%에서 67.6%로 증가합니다. 제곱근 평균 제곱 오차는 6.07에서 4.0으로 감소합니다. 그림 4.4에서 볼 수 있듯이 전체 모형에 대한 *F*-검정 유의 확률은 0.03에서 0.0001 미만으로 감소합니다.

그림 4.4 약물 데이터에 대한 공분산 분석



약물 데이터 테이블에는 반복된 관측값이 포함되어 있습니다. 예를 들어 행 1과 9는 둘 다 Drug = a이고 x = 11입니다. 모형을 적합시킬 때 반복된 관측값을 사용하여 변동의 **순수 오차** 추정값을 생성할 수 있습니다. 공변량의 지정되지 않은 함수 형식 또는 명목형 효과의 교호작용에 대한 또 다른 오차 추정값을 생성할 수 있습니다. 이러한 추정값은 적합 결여 검정의 기반이 됩니다. 적합 결여 오차가 유의하면 데이터의 일부 효과가 모형에 의해 설명되지 않음을 나타냅니다.

"적합 결여" 보고서에는 약물 데이터에 대한 이 검정의 결과가 표시됩니다. Prob > F 값 = 0.7507에서 알 수 있듯이 적합 결여 오차는 유의하지 않습니다.

공변량 x는 이전에 Drug 변수에 의해 설명된 반응 변동의 대부분을 설명합니다. 따라서 모형이 훨씬 적은 오차로 적합되더라도 Drug 효과는 더 이상 유의하지 않습니다. 주효과 모형에서 관측된 Drug의 유의성은 공변량에 의해 설명되는 것으로 보입니다.

공분산 모형의 최소 제곱 평균은 일반 평균과 다릅니다. 이 최소 제곱 평균은 공변량 x가 반응 y에 미치는 효과에 대해 조정되기 때문입니다. 최소 제곱 평균은 공변량 x가 어떤 중립 값으로 고정될 때 Drug의 세 가지 각 수준에 대한 예측값입니다. 중립 값은 공변량의 평균인 10.7333으로 선택됩니다.

최소 제곱 평균은 "모수 추정값" 보고서에 제공된 모수 추정값을 사용하여 계산됩니다.

예측 표현식 -2.696 - 1.185*Drug[a] - 1.0761*Drug[d] + 0.98718*x

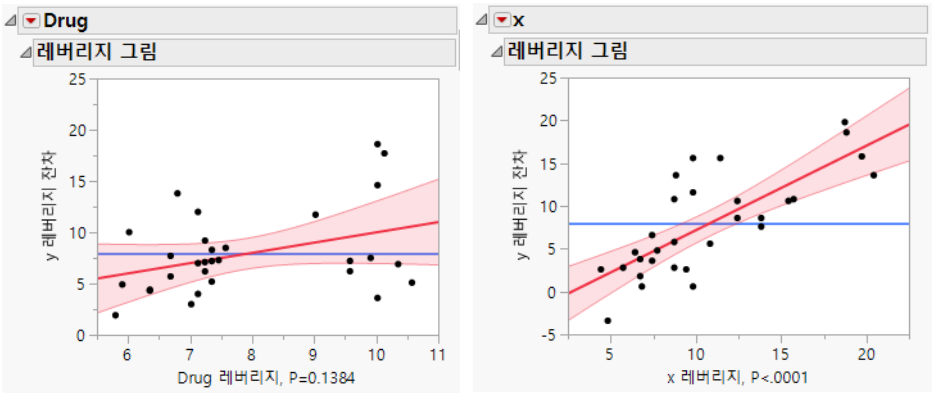
a: -2.696 - 1.185*(1) -1.0761*(0) + 0.98718*(10.7333) = 6.71

d: -2.696 - 1.185*(0) -1.0761*(1) + 0.98718*(10.7333) = 6.82

f: -2.696 - 1.185*(-1) -1.0761*(-1) + 0.98718*(10.7333) = 10.16

그림 4.5에서는 각 효과에 대한 레버리지 그림을 보여 줍니다. 공변량이 유의하므로 Drug의 레버리지 값은 최소 제곱 평균에서 어느 정도 분산되어 있습니다.

그림 4.5 약물 검정 데이터에 대한 레버리지 그림 비교



서로 다른 기울기의 공분산 분석 예

이 예에서는 Drug.jmp 샘플 데이터 테이블을 계속 사용하여 공변량의 기울기가 Drug 수준에 따라 달라지는 모형을 적합시킵니다. 모형을 적합시킨 후 공변량 x의 특정 값에서 Drug 수준에 대한 최소 제곱 평균을 비교합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.

3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Drug 와 x 를 모두 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .

   이렇게 하면 **차수** 상자에 지정된 차수까지의 항이 모형에 추가됩니다 . 기본 **차수** 값은 2 입니다 . 따라서 주효과 Drug, x 및 교호작용 Drug*x 가 모형 효과 목록에 추가됩니다 .
5. **실행**을 클릭합니다 .

이 규격은 공변량의 기울기를 Drug 수준에 따라 달라지게 하는 두 개의 열을 선형 모형에 추가합니다(x_{4i} 및 x_{5i}). Drug에 대한 표시 변수에 공변량 값을 곱하여 새 변수가 생성되고 다음 계산식을 제공합니다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i$$

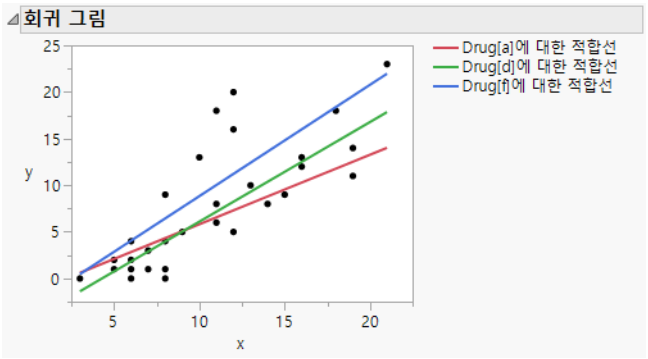
표 4.1에서는 이 모형의 코딩을 보여 줍니다. X의 평균은 10.7333입니다. 이 값은 연속형 항을 중심화하는 데 사용됩니다.

표 4.1 서로 다른 기울기의 공분산 분석 코딩

회귀변수	효과	값
X_1	Drug[a]	a: +1, d: 0, f: –1
X_2	Drug[d]	a: 0, d: +1, f: –1
X_3	X	X의 값
X_4	Drug[a]*(X - 10.733)	a: X – 10.7333, d: 0, f: –(X – 10.7333)
X_5	Drug[d]*(X - 10.733)	a: 0, d: X – 10.7333, f: –(X – 10.7333)

그림 4.6에 보고서의 일부가 표시되어 있습니다. "회귀 그림"에서는 기울기가 다른 적합선을 보여 줍니다. "효과 검정" 보고서에서는 교호작용의 p 값 0.56을 제공합니다. 이것은 유의하지 않으므로 모형에 서로 다른 기울기를 포함할 필요가 없음을 나타냅니다.

그림 4.6 교호작용 그림



효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F
Drug	2	2	52.05637	1.5713	0.2284
x	1	1	564.56753	34.0821	<.0001*
Drug*x	2	2	19.64465	0.5930	0.5606

스포트라이트 분석 수행

이제 공변량 x 의 특정 값에서 Drug 수준에 대한 최소 제곱 평균을 비교하려고 합니다 . 공분산 분석 모형에서 이러한 유형의 비교를 **스포트라이트 분석**이라고도 합니다 . 스포트라이트 분석에 대한 자세한 내용은 Spiller et al. 연구 자료 (2013) 에서 확인하십시오 .

1. " 반응 y" 의 빨간색 삼각형을 클릭하고 **추정값 > 다중 비교**를 선택합니다 .
2. " 다중 비교 " 창에서 **사용자 정의 추정값**을 선택합니다 .
3. **Drug 수준 선택** 아래의 세 값을 모두 선택합니다 .
4. x 아래의 첫 번째 상자에 12.5 를 입력합니다 .
5. **추정값 추가**를 클릭합니다 .

   그러면 x = 12.5 에서 Drug 의 세 수준 비교가 추가됩니다 .
6. **확인**을 클릭합니다 .

   " 사용자 정의 추정값 " 보고서에는 공변량 x 가 12.5 로 설정된 각 Drug 수준의 최소 제곱 평균 추정값이 표시됩니다 . " 사용자 정의 추정값에 대한 다중 비교 " 옆의 빨간색 삼각형에는 추정값 간의 차이를 검정할 수 있는 옵션이 포함되어 있습니다 .

그림 4.7 사용자 정의 추정값 보고서

사용자 정의 추정값에 대한 다중 비교

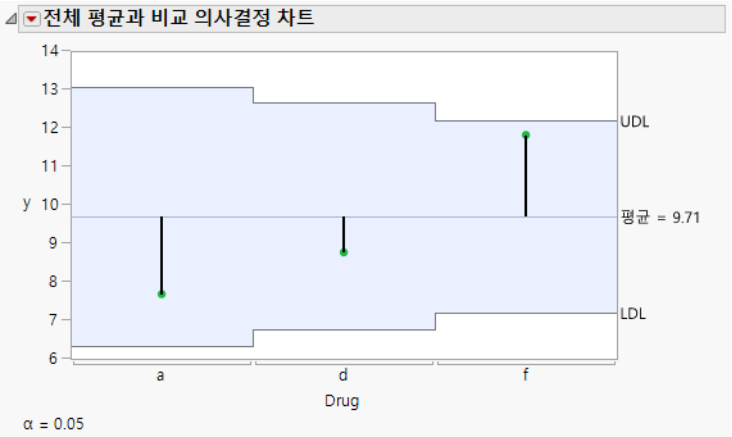
사용자 정의 추정값

x = 12.5

Drug	추정값	표준 오차	DF	95% 하한	95% 상한
a	7.684713	1.5772057	24	4.4295208	10.939906
d	8.771371	1.4401229	24	5.7991033	11.743639
f	11.822214	1.2943345	24	9.1508392	14.493590

7. " 사용자 정의 추정값에 대한 다중 비교 " 옆의 빨간색 삼각형을 클릭하고 **전체 평균과 비교**를 선택합니다 .

그림 4.8 전체 평균과 비교 의사결정 차트



"전체 평균과 비교" 옵션은 평균과 세 가지 최소 제곱 평균 간의 차이에 대한 ANOM(평균 분석) 차트를 생성합니다. ANOM 차트를 참조하여 x = 12.5에서 반응에 대한 Drug 효과는 유의하지 않다고 결론을 내릴 수 있습니다.

반응 표면 모형 예

이 예에서는 반응 표면 모형을 적합시킵니다 . 반응을 최소화하는 것이 목적입니다 .

- " 전체 반응 표면 모형 적합 "
- " 모형 축소 "
- " 반응 표면 보고서 검토 "
- " 예측 프로파일러를 사용하여 임계점 찾기 "
- " 등고선 프로파일러를 사용하여 표면 보기 "

전체 반응 표면 모형 적합

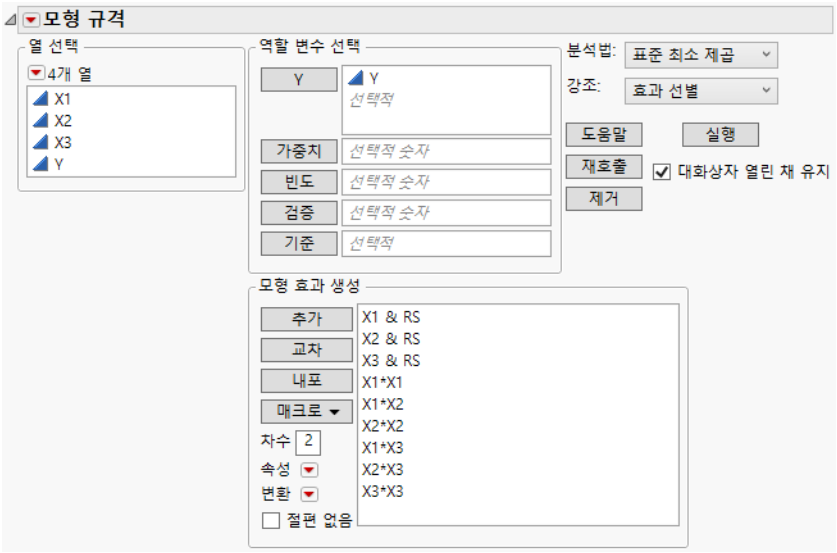
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Custom RSM.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다 .

   데이터 테이블에 **Model** 스크립트가 포함되어 있어서 " 모형 규격 " 창이 **Model** 스크립트에 지정된 대로 채워집니다 . 다음 사항에 유의하십시오 .

   – 주효과는 **모형 효과 생성** 목록에 **&RS** 접미사가 추가된 상태로 표시되며 이는 반응 표면 매크로가 적용되었음을 나타냅니다 .

– 세 가지 예측 변수 X1, X2, X3 의 전체 반응 표면에 대한 효과가 표시됩니다 .
– **&RS** 접미사가 추가된 항이 모형에 포함되어 있으므로 분석 결과에 " 반응 표면 " 보고서가 포함됩니다 .

그림 4.9 반응 표면 분석을 위한 모형 적합 시작 창

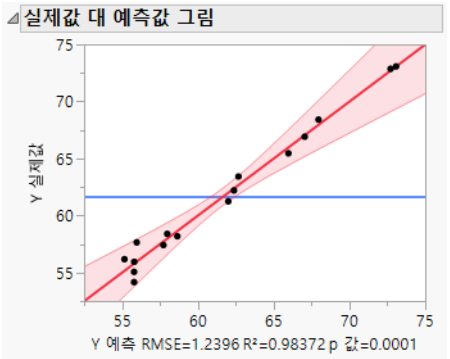


3. **실행**을 클릭합니다 .

모형 축소

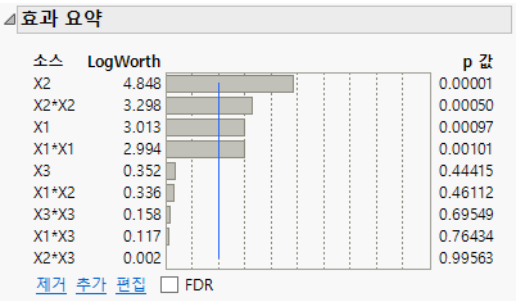
" 실제값 대 예측값 그림 " 에서는 모형이 유의함을 보여 줍니다 . 적합 결여를 나타내는 증거가 없습니다 .

그림 4.10 전체 모형에 대한 실제값 대 예측값 그림



"효과 요약" 보고서에서는 여러 효과가 유의하지 않음을 나타냅니다. 특히 X2*X3은 최소 유의 효과입니다(p 값 = 0.99563). 이제 대화식 "효과 요약" 보고서를 사용하여 모형을 체계적으로 축소합니다.

그림 4.11 효과 요약 보고서



1. " 효과 요약 " 보고서에서 X2*X3 을 클릭하고 **제거**를 클릭합니다 .

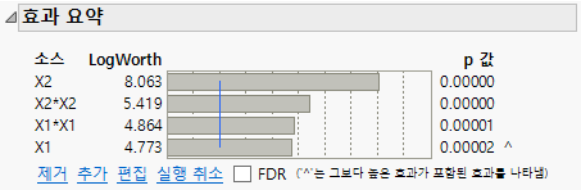
   모형이 업데이트됩니다 .

   " 효과 요약 " 보고서의 "p 값 " 열에 따르면 X1*X3 이 유의하지 않습니다 .
2. X1*X3 을 클릭하고 **제거**를 클릭합니다 .

   X3*X3 의 p 값은 이 효과가 유의하지 않음을 나타냅니다 .
3. X3*X3 을 클릭하고 **제거**를 클릭합니다 .
4. X1*X2 를 클릭하고 **제거**를 클릭합니다 .

   X3 이 유의하지 않습니다 . 이 효과는 고차 효과에 포함되지 않으므로 효과 유전성 원리를 위반하지 않고 제거할 수 있습니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 효과 유전성 "(171 페이지) 에서 확인하십시오 .
5. X3 을 클릭하고 **제거**를 클릭합니다 .

그림 4.12 모형 축소 후 효과 요약 보고서

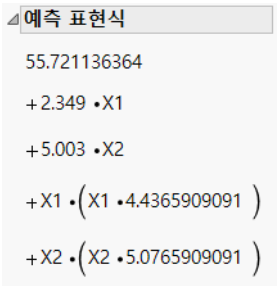


나머지 효과는 모두 유의합니다 .

반응 표면 보고서 검토

1. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다 .

그림 4.13 예측 표현식

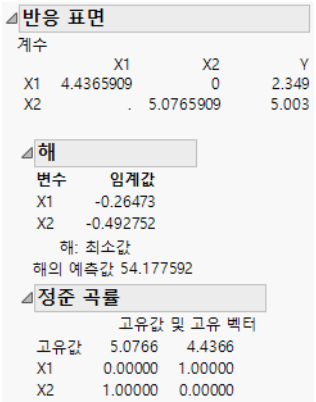


팁 : 예측 표현식은 예측 계산식입니다 . **열 저장 > 예측 계산식**을 선택하여 이 계산식을 얻을 수도 있습니다 .

다음 단계에서는 예측 표현식을 참조하여 모형 계수를 확인할 수 있습니다 .

2. " 반응 표면 " 보고서와 " 정준 곡률 " 보고서를 차례로 엽니다 .

그림 4.14 반응 표면 보고서



첫 번째 테이블에는 2 차 모형 계수가 행렬 형식으로 제공됩니다 . X1*X1 의 계수는 4.4365909, X2*X2 의 계수는 5.0765909, X1*X2 의 계수는 0 입니다 . 선형 효과 계수 (X1 의 2.349, X2 의 5.003) 는 Y 라벨이 지정된 열에 제공됩니다 .

" 해 " 보고서에는 임계값이 표시됩니다 . 이러한 값은 최대값 , 최소값 또는 안장점이 발생하는 지점을 나타냅니다 . 이 예의 " 해 " 보고서에서는 X1 = -0.265, X2 = -0.493 의 임계값에서 반응 표면의 최소값이 54.18 임을 보여 줍니다 .

" 정준 곡률 " 보고서에는 2 차 모수 추정값 행렬의 고유 구조가 표시됩니다 . 고유 구조는 곡률의 형태와 방향을 식별하는 데 유용합니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 정준 곡률 보고서 "(81 페이지) 에서 확인하십시오 .

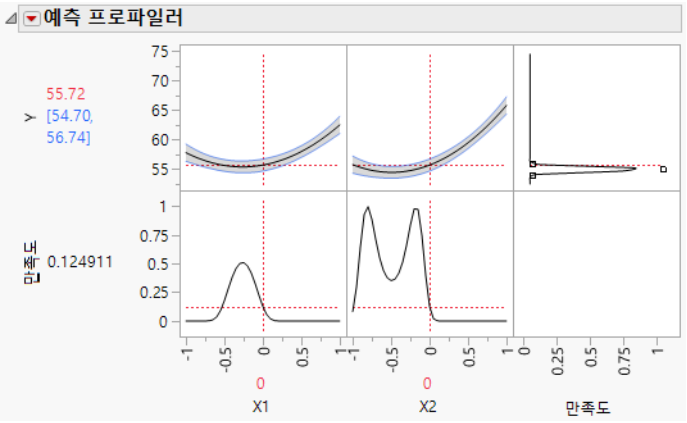
이 예에서는 두 고유값이 모두 양수이므로 표면이 최소값을 얻는다는 것을 나타냅니다. 가장 큰 곡률의 방향은 최대 고유값 (5.0766) 에 해당합니다 . 방향은 해당하는 고유 벡터 성분에 의

해 정의됩니다 . 첫 번째 방향의 경우 고유 벡터 값이 1.00 인 X2 가 방향을 결정합니다 . 두 번째 방향은 고유 벡터 값이 1.00 인 X1 에 의해 결정됩니다 .

예측 프로파일러를 사용하여 임계점 찾기

" 예측 프로파일러 " 보고서에는 X1 및 X2에 대한 트레이스를 따라 반응 표면의 2차 동작이 표시됩니다. Y에 "반응 한계" 열 특성이 설정되어 있으므로 프로파일러에 만족도 함수도 표시됩니다.

그림 4.15 목표값 일치가 목표인 예측 프로파일러

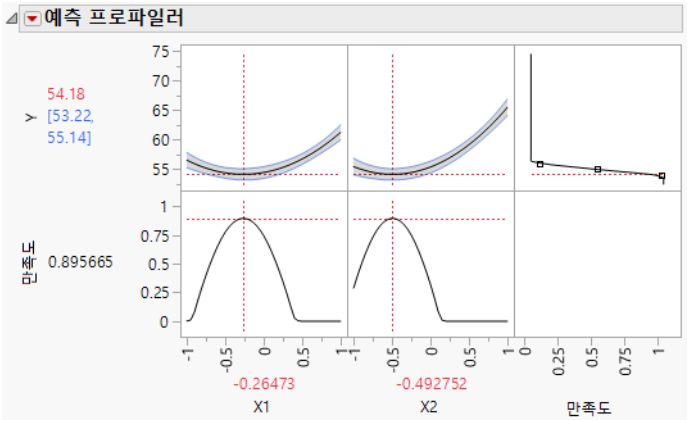


" 반응 한계 " 열 특성의 목표는 " 목표값 일치 " 로 설정됩니다 . 이 예에서는 목표값 일치 대신 Y를 최소화하는 데 관심이 있습니다 . 따라서 다음과 같이 설정을 변경해야 합니다 .

1. Ctrl 키를 누르고 " 예측 프로파일러 " 의 오른쪽 위 셀을 클릭합니다 .
2. " 반응 목표 " 대화상자의 옵션 목록에서 **최소화**를 선택합니다 .
3. **확인**을 클릭합니다 .

   이제 만족도 함수에 Y 최소화 목표가 반영됩니다 .
4. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 4.16 최소화가 목표이고 만족도가 최대화된 예측 프로파일러



설계 영역 내에서 Y 를 최소화하는 설정이 프로파일러 아래에 나타납니다 . 이 설정은 " 해 " 보고서에 제공된 " 임계값 " 과 정확히 일치합니다 .

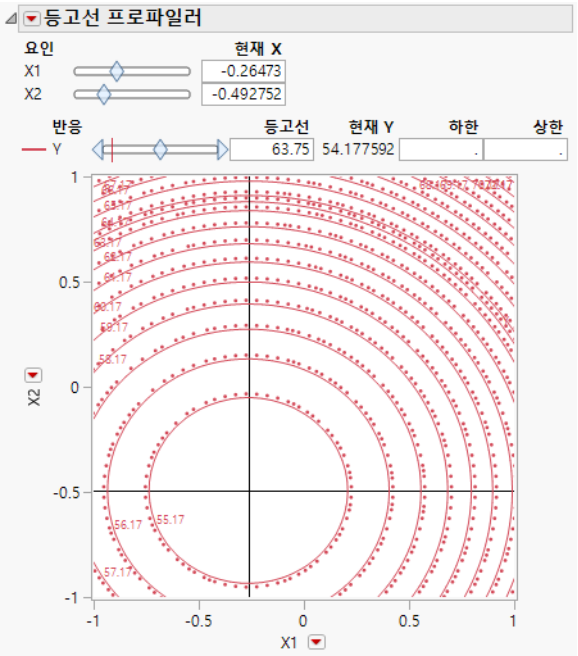
등고선 프로파일러를 사용하여 표면 보기

" 등고선 프로파일러 " 에는 반응 표면의 등고선이 표시됩니다 . 이는 임계점 영역에서 예측 반응에 대한 대체 프로파일러 시각화를 제공합니다 .

1. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 등고선 프로파일러**를 선택합니다 .
2. " 등고선 프로파일러 " 의 빨간색 삼각형을 클릭하고 **등고선 격자**를 선택합니다 .
3. **증분**에 1 을 입력합니다 .
4. **확인**을 클릭합니다 .

   등고선이 1 단위 간격으로 표시됩니다 . 자세한 내용은 그림 4.17 에서 확인하십시오 .
5. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **요인 설정 > 연결 프로파일러**를 선택합니다 .

그림 4.17 임계점에 십자기호가 있는 등고선 프로파일러

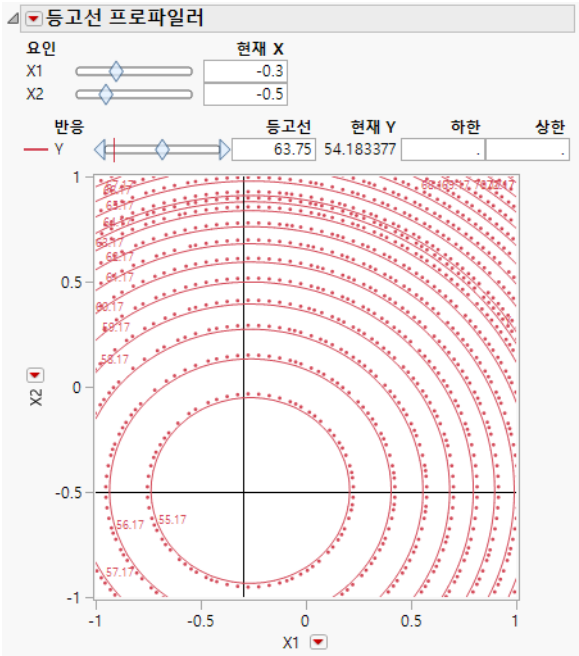


등고선 프로파일러를 예측 프로파일러에 연결하면 등고선 프로파일러의 **현재 X** 값이 예측 프로파일러에 표시된 X 값에 연결됩니다 . 예측 프로파일러의 X 값은 Y 가 최소화되는 임계점을 제공합니다 . 등고선 프로파일러의 십자기호는 임계점을 표시합니다 . **현재 Y** 값은 예측 프로파일러에 따라 예측된 최소값인 54.177592 입니다 .

공정 요인의 값을 반응을 최적화하는 값으로 정확하게 설정할 수 없는 경우가 종종 있습니다 . 등고선 프로파일러를 사용하면 공정 요인의 대체 설정을 식별하는 데 도움이 됩니다 . 다음 단계에서는 공정을 소수점 이하 한 자리만 사용하여 X1 및 X2 값으로 설정할 수 있으며 , 공정 설정은 방향에 관계없이 소수점 이하 한 자리씩 변할 수 있다고 가정합니다 .

6. " 등고선 프로파일러 " 보고서의 **현재 X** 아래에서 X1 옆에 -0.3 을 입력하고 X2 옆에 -0.5 를 입력합니다 .

그림 4.18 X1 = -0.3, X2 = -0.5 를 표시하는 등고선 프로파일러



십자기호는 가장 안쪽 등고선 내에 있으며 현재 Y(현재 X 설정의 예측 Y 값) 는 예측 최소값 54.177592 와 약간 다른 54.183377 입니다 .

7. 등고선 프로파일러에서 십자기호를 클릭하고 드래그하여 X1 = -0.3 및 X2 = -0.5의 0.1 단위 반지름 내에서 현재 X 값을 탐색합니다 .

   예측 Y 값은 모두 54.4 미만입니다 . 실제로 설정이 가장 안쪽 등고선 내의 어떤 한 지점까지 이동하면 예측 Y 가 등고선 값 55.17 보다 작습니다 .

분할구 설계 예

임의 효과 수준은 더 많은 수준의 모집단에서 무작위로 선택됩니다 . 추론을 위해 임의 효과의 분포를 평균 0 과 특정 분산 (**분산 성분**이라고 함) 의 정규 분포로 가정합니다 .

어떤 의미에서는 잔차 오차를 구성하는 임의 효과가 모든 모형에 하나 이상 있다고 할 수 있습니다. 개별 관측값은 훨씬 더 큰 모집단에서 무작위로 선택된다고 가정하고 오차 항은 평균이 0 이고 분산이 σ^2 인 것으로 가정합니다.

가장 일반적인 임의 효과 모형은 반복 측정 또는 분할구 모형입니다. 표 4.2에는 분할구 모형의 효과 유형이 나열되어 있습니다. 이러한 모형을 사용하는 경우 실험은 두 가지 계층을 포함합니다. 일부 효과를 실험의 주구 또는 개체에 적용합니다. 그런 다음 이러한 주구를 분할하거나 서로 다른 시간에 개체를 측정하고, 나머지 효과를 해당 하위 단위 내에 적용합니다. 주구 또는 개

체를 설명하는 효과는 주구 효과이고 하위구 또는 반복 측정을 설명하는 효과는 하위구 효과입니다. 일반적으로 하위 단위 효과는 모형에서 생략되고 잔차 오차로 흡수됩니다.

표 4.2 분할구 모형의 효과 유형

분할구 모형	효과 유형	반복 측정 모형
주구 처리	고정 효과	개체 간 처리
주구 ID	임의 효과	개체 ID
하위구 처리	고정 효과	개체 내 처리
하위구 ID	임의 효과	반복 측정 ID

이러한 각 사례는 계층화된 모형으로 처리할 수 있으며 몇 가지 기존 방식을 사용하여 타당하게 모형을 적합시킬 수 있습니다 . 다음과 같이 두 가지 다른 실험으로 처리됩니다 .

1. 주구 실험은 주구 또는 개체를 실험 단위로 사용하여 오차 항을 구성합니다 .
2. 하위구 처리는 개별 측정을 실험 단위로 사용하여 오차 항을 구성합니다 (잔차 오차로 유지).

주구를 검정하는 기존 방식은 다음 중 하나를 수행하는 것입니다.

- 전체 측정에서 평균을 구하여 주구 효과에 적합시킵니다 .
- 주구 평균 제곱을 주구 ID 평균 제곱으로 나누어 *F* 비를 계산합니다 .
- 분할 또는 반복 측정으로 서로 다른 열이 생성되도록 데이터를 구성합니다 . 그런 다음 MANOVA 모형을 적합시키고 단변량 통계량을 사용합니다 .

이러한 방식은 단순한 구조이거나 데이터가 완전하고 균형을 이룰 때 사용할 수 있습니다. 그러나 임의 효과의 구조에 작동하는 더 일반적인 모형이 있습니다. 이 일반화 모형은 고정 효과와 임의 효과를 모두 포함하므로 **혼합 모형**이라고 합니다.

계층화된 설계의 가장 일반적인 유형은 균형 분할구로, 대개 시간에 따른 반복 측정 형식입니다. 일부 효과에 대한 하나의 실험 단위가 세분화되고 (대개 시간별) 다른 효과가 이 하위 단위에 적용됩니다.

Animals.jmp 샘플 데이터 테이블의 가상 데이터를 예로 들겠습니다 . 이 연구에서는 여우와 코요테의 계절별 사냥 습관 차이에 대한 정보를 수집했습니다 . 1 년 동안 각 계절마다 여우 3 마리와 코요테 3 마리를 표적으로 삼아 주기적으로 관찰했습니다 . 1 년 중 다른 계절마다 굴을 돌아다니는 평균 거리 (마일) 를 기록하여 가장 근사한 값으로 반올림했습니다 . 모형은 다음과 같은 특성으로 정의됩니다 .

- 연속형 반응 변수 miles
- fox 또는 coyote 값을 갖는 species 효과
- fall, winter, spring 및 summer 값을 갖는 season 효과
- 여우와 코요테 둘 다에 대해 명목형 값 1, 2, 3 을 갖는 동물 식별 코드 subject

모형에는 다음과 같은 두 개의 계층이 있습니다 .

1. 상위 계층은 개체 간 계층으로 , 개체 변화에 따라 여우 또는 코요테가 되는 효과 (species 효과) 가 검정됩니다 .
2. 하위 계층은 개체 내 계층으로 , 한 개체 내에서 계절 변화에 따라 4 가지 계절 (season 효과) 에 대한 반복 측정 요인이 검정됩니다 . 개체 내 변동은 잔차 오차에 반영됩니다 .

season 효과는 *F* 통계량의 분모에 잔차 오차를 사용할 수 있습니다. 그러나 개체 간 변동은 잔차 오차에 의해 측정되지 않으며 모형에서 species 내의 subject(subject[species]) 효과로 포착되어야 합니다. 개체 간 효과 species에 대한 *F* 통계량은 잔차 오차 대신 이 내포 효과를 *F* 비 분모에 사용합니다.

참고 : JMP Pro 사용자는 혼합 모형 분석법을 사용하여 이 모형을 생성할 수 있습니다 .

이 데이터에 대한 분할구 모형을 지정하려면 다음 단계를 수행하십시오 .

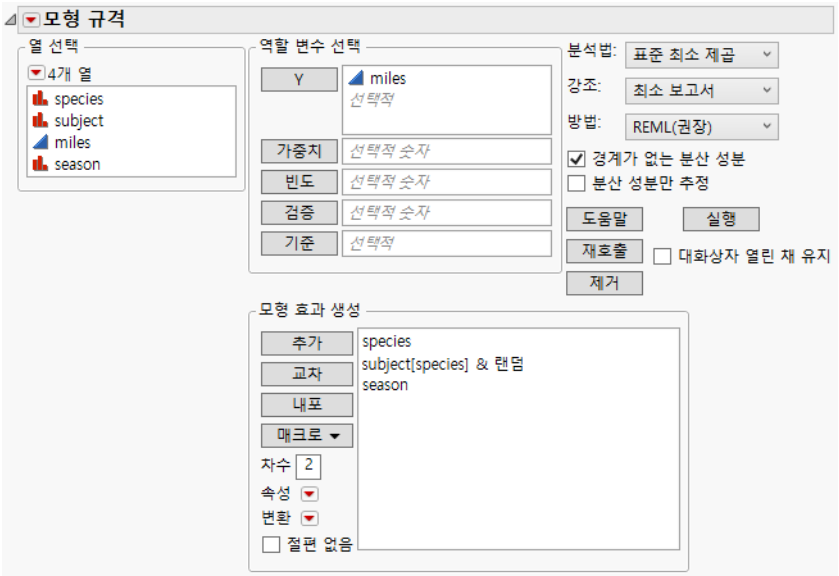
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Animals.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. miles 를 선택하고 **Y** 를 클릭합니다 .
4. species 와 subject 를 선택하고 **추가**를 클릭합니다 .
5. " 열 선택 " 목록에서 species 를 선택합니다 .
6. " 모형 효과 생성 " 목록에서 subject 를 선택합니다 .
7. **내포**를 클릭합니다 .

   그러면 species 내의 subject(subject[species]) 효과가 모형에 추가됩니다 .
8. 내포 효과 subject[species] 를 선택합니다 .
9. **속성 > 임의 효과**를 선택합니다 .

   이제 이 내포 효과는 species 효과의 오차 항으로 식별되고 "subject[species]& 랜덤 " 으로 표시됩니다 .
10. " 열 선택 " 목록에서 season 을 선택하고 **추가**를 클릭합니다 .

    속성 메뉴를 사용하여 효과를 임의 효과로 정의하면 대화상자의 오른쪽 위에 " 방법 " 옵션 (**REML** 및 **EMS**) 이 나타납니다 . **REML** 옵션이 기본적으로 선택되어 있습니다 . 완료된 시작 창이 그림 4.19 에 표시되어 있습니다 .

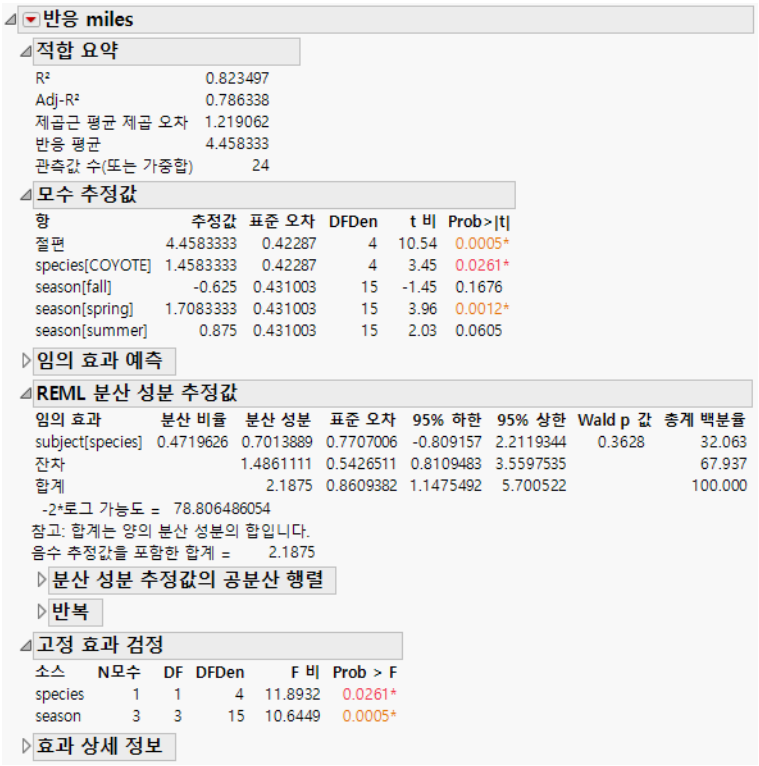
그림 4.19 모형 적합 대화상자



11. **실행**을 클릭합니다 .

    그림 4.20 에 보고서가 표시되어 있습니다 . 두 고정 효과 (species 와 season) 가 모두 유의합니다 . "REML 분산 성분 추정값 " 보고서에는 species 에 내포된 subject 의 추정값 및 잔차 분산이 제공됩니다 .

그림 4.20 REML 분석 보고서 (일부분)



반응 miles

적합 요약

R^2	0.823497
Adj-R^2	0.786338
제곱근 평균 제곱 오차	1.219062
반응 평균	4.458333
관측값 수(또는 가중합)	24

모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|
절편	4.4583333	0.42287	4	10.54	0.0005*
species[COYOTE]	1.4583333	0.42287	4	3.45	0.0261*
season[fall]	-0.625	0.431003	15	-1.45	0.1676
season[spring]	1.7083333	0.431003	15	3.96	0.0012*
season[summer]	0.875	0.431003	15	2.03	0.0605

임의 효과 예측

REML 분산 성분 추정값

임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
subject[species]	0.4719626	0.7013889	0.7707006	-0.809157	2.2119344	0.3628	32.063
잔차		1.4861111	0.5426511	0.8109483	3.5597535		67.937
합계		2.1875	0.8609382	1.1475492	5.700522		100.000

-2*로그 가능도 = 78.806486054

참고: 합계는 양의 분산 성분의 합입니다.
음수 추정값을 포함한 합계 = 2.1875

분산 성분 추정값의 공분산 행렬

반복

고정 효과 검정

소스	N모수	DF	DFDen	F 비	Prob > F
species	1	1	4	11.8932	0.0261*
season	3	3	15	10.6449	0.0005*

효과 상세 정보

임의 효과 모수 추정 예

임의 효과에는 이중 특성이 있습니다. 임의 효과는 잔차 오차(예: 주구 실험 단위와 관련된 오차) 를 나타낸다는 것이 한 가지 특성입니다. 다른 특성은 모수를 임의 효과의 각 수준과 연관시켜 고정 효과처럼 사용된다는 것입니다. 모수는 평균 0 및 분산 성분에 의해 추정된 분산을 사용하는 정규 분포에서 파생된다는 추가 정보가 있습니다. 이 추가 정보에 따라 모수 추정값은 0 으로 축소됩니다.

임의 효과와 관련된 모수 추정값을 BLUP(최량 선형 비편향 예측 변수)라고 합니다. 이러한 BLUP 는 연구원에 따라 관심 모수로 간주되거나 방법론의 흥미 없는 부산물로 간주되기도 합니다.

BLUP 모수 추정값은 임의 효과 최소 제곱 평균을 추정하는 데 사용되며 이러한 평균은 결국 총 평균으로 축소됩니다. 효과의 분산 및 효과의 수준당 관측값 수에 따라 축소 정도가 다릅니다. 분산 추정값이 크면 축소가 거의 없습니다. 분산 성분이 작으면 더 크게 축소됩니다. 분산 성분이 0 이면 효과 수준이 정확히 0 으로 축소됩니다. 축소가 반전되는 매우 큰 음의 분산 성분을 얻을 수도 있습니다. 고정 효과를 분산 성분이 매우 큰 특수 임의 효과로 간주할 수 있습니다.

REML 방법은 각 개별 수준에 대한 정보와 수준 간 분산에 대한 정보의 균형을 조정합니다. 수준당 관측값 수가 많으면 추정값이 덜 축소됩니다. 수준당 관측값 수가 매우 적으면 추정값이 더 축소됩니다. 관측값이 무한히 많은 경우에는 축소되지 않고 추정값이 고정 효과와 동일합니다.

여러 야구 선수의 타율 정보가 있다고 가정해 보겠습니다. 선수 간 타율 실적에 대한 분산 성분은 타율에서 선수 간의 일반적인 변동 크기를 설명합니다. 예를 들어 선수가 경기에 참여한 횟수가 몇 번뿐이면 타율이 비정상적으로 작거나 큽니다. 그러면 해당 추정값은 단지 몇 타석에만 근거하므로 신뢰할 수 없는 경우가 많습니다. 그러나 이 추정값을 총 평균과 혼합하면 추정값이 총 평균에 가까이 축소되므로 추정값을 더 신뢰할 수 있습니다. 오랜 타율 기록이 있는 선수의 경우 기록이 부족한 선수보다 훨씬 적게 축소할 수 있습니다.

이를 직접 확인할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Baseball.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Batting 을 선택하고 **Y** 를 클릭합니다 .
4. Player 를 선택하고 **추가**를 클릭합니다 .
5. " 모형 효과 생성 " 상자에서 "Player" 를 선택하고 " 속성 " 목록에서 **임의 효과**를 선택합니다 .
6. **실행**을 클릭합니다 .

표 4.3 에서는 REML(권장) 적합에 대한 Player 보고서의 최소 제곱 평균을 보여 줍니다 . 또한 EMS 방법을 사용하여 계산된 적률 방법 추정값도 표시됩니다 . 적률 방법 추정값은 일반적인 Player 평균입니다 . 3타석 기록만 있는 Suarez에 대한 REML 추정값은 타석 수가 더 많은 다른 선수에 비해 총 평균에 더 가까워지도록 많이 축소됩니다 .

표 4.3 적률 방법과 REML 간 추정값 비교

	적률 방법	REML	N
분산 성분	0.01765	0.019648	
Anderson	0.29500000	0.29640407	6
Jones	0.20227273	0.20389793	11
Mitchell	0.32333333	0.32426295	6
Rodriguez	0.55000000	0.54713393	6
Smith	0.35681818	0.35702094	11
Suarez	0.55000000	0.54436227	3
최소 제곱 평균	일반 평균과 같음	평균에서 축소됨	

매듭이 지정된 스플라인 효과 예

평활 다항식의 분할을 지정된 효과에 적합시키려면 **매듭이 지정된 스플라인 효과** 옵션을 사용합니다. 이 속성을 선택하면 매듭 점 수를 지정할 수 있는 창이 나타납니다. 매듭이 지정된 스플라인은 주효과 연속형 항에 대해서만 구현됩니다.

JMP는 점을 배치할 때 문헌의 권장 사항을 따릅니다. 매듭이 지정된 스플라인은 Stone 스플라인 또는 Stone-Koo 스플라인이라고도 합니다. 자세한 내용은 Stone and Koo 연구 자료(1985)에서 확인하십시오. 점이 100개 이하인 경우 첫 번째 매듭과 마지막 매듭은 각각 최소값과 최대값 내의 5번째 점입니다. 그렇지 않은 경우 첫 번째 매듭과 마지막 매듭은 각각 매듭이 5개 이하일 때는 0.05 분위수와 0.95 분위수에 배치되고, 매듭이 5개보다 많을 때는 0.025 분위수와 0.975 분위수에 배치됩니다. 점이 30개 이하인 경우를 제외하고 기본 매듭 수는 5개입니다. 30개 이하인 경우 기본값은 3개 매듭입니다.

매듭이 지정된 스플라인은 평활 스플라인과 대조적으로 다음과 같은 특성을 나타냅니다.

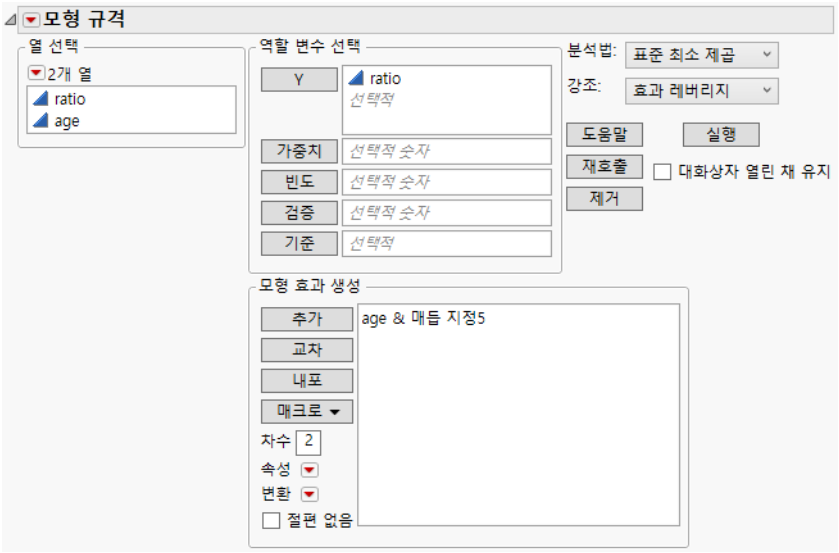
- 매듭이 지정된 스플라인은 항이 많은 일반 모형 내에서 사용되지만 평활 스플라인은 이변량 회귀에 사용됩니다.
- 회귀 기저가 반응의 함수가 아닙니다.
- 매듭이 지정된 스플라인은 절약성 원리에 따라 k 개 매듭 점의 곡률을 위해 $k-2$ 개의 항만 추가합니다.
- 매듭이 지정된 스플라인은 데이터 범위를 벗어나는 외삽이 다항식이 아니라 직선이라는 점에서 순수 다항식에 비해 보수적입니다.
- 간단한 방법으로 곡률을 검정할 수 있습니다.

매듭이 지정된 스플라인 효과를 사용한 곡률 검정의 예

곡률을 검정하려면 다음을 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Growth.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. ratio 열을 선택하고 **Y** 를 클릭합니다.
4. age 열을 선택하고 **추가**를 클릭합니다.
5. **효과** 페인에서 age 를 선택하고 **속성 > 매듭이 지정된 스플라인 효과**를 선택합니다.
6. 매듭 수에 5 를 입력하고 **확인**을 클릭합니다.

그림 4.21 모형 적합 시작 창

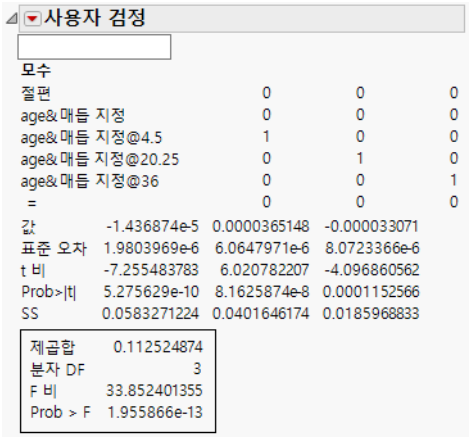


7. **실행**을 클릭합니다 .
8. " 반응 ratio" 의 빨간색 삼각형을 클릭하고 **추정값 > 사용자 검정**을 선택합니다 .
   " 사용자 검정 " 보고서에 열이 하나만 있습니다 . 이 예에서는 세 개의 열이 필요합니다 .
9. **열 추가** 버튼을 두 번 클릭하여 총 세 개의 열을 생성합니다 .
10. 첫 번째 열의 **age& 매듭 지정 @4.5** 에 1 을 입력합니다 .
11. 두 번째 열의 **age& 매듭 지정 @20.25** 에 1 을 입력합니다 .
12. 세 번째 열의 **age& 매듭 지정 @36** 에 1 을 입력합니다 .

그림 4.22 곡률에 대한 사용자 검정의 값

13. **완료**를 클릭합니다 .
    " 사용자 검정 " 보고서의 "Prob > F" 값이 작으므로 곡률이 있음을 의미합니다 .

그림 4.23 곡률 보고서



사용자 검정

모수			
절편	0	0	0
age&매듭 지정	0	0	0
age&매듭 지정@4.5	1	0	0
age&매듭 지정@20.25	0	1	0
age&매듭 지정@36	0	0	1
=	0	0	0
값	-1.436874e-5	0.0000365148	-0.000033071
표준 오차	1.9803969e-6	6.0647971e-6	8.0723366e-6
t 비	-7.255483783	6.020782207	-4.096860562
Prob>|t|	5.275629e-10	8.1625874e-8	0.0001152566
SS	0.0583271224	0.0401646174	0.0185968833

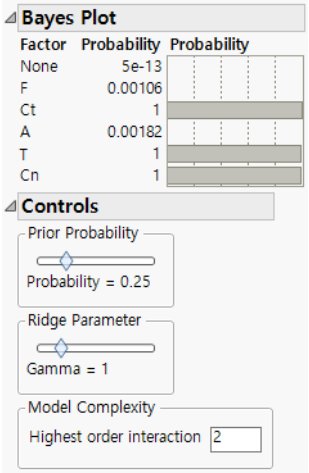
제곱합	0.112524874
분자 DF	3
F 비	33.852401355
Prob > F	1.955866e-13

활성 요인에 대한 Bayes 그림 예

실험 설계를 수행하고 활성 요인을 판별하려는 경우를 가정해 보겠습니다 . JMP 를 사용하여 여러 가지 방법으로 이를 처리할 수 있습니다 . 이 예에서는 Bayes 방법으로 활성 요인을 식별하는 데 사용할 수 있는 스크립트를 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다 .
2. Samples > Scripts 폴더에서 BayesPlotforFactors.jsl 샘플 스크립트를 엽니다 .
3. **편집 > 스크립트 실행**을 선택합니다 .
4. Y 를 선택하고 **Y, Response** 를 클릭합니다 .
5. F, Ct, A, T 및 Cn 을 선택하고 **X, Factor** 를 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 4.24 요인 활동에 대한 Bayes 그림



"Model Complexity"에서는 고려할 최고 차수 교호작용이 2임을 나타냅니다. 따라서 최대 2차 교호작용을 포함하는 가능한 모든 모형이 생성됩니다. "Prior Probability"에 할당된 값을 기반으로 가능한 각 모형의 사후 확률이 계산됩니다. 요인에 대한 확률은 해당 요인이 포함된 각 모형에 대한 확률의 합입니다.

이 방법으로 Ct, T 및 Cn이 활성 요인으로 식별되고 A와 F가 비활성 요인으로 식별됩니다.

능형 회귀 모수가 0(허용되지 않음)이면 모든 모형이 최소 제곱으로 적합됩니다. 능형 회귀 모수가 증가하면 모형의 모수 추정값은 0에 가까워지도록 축소됩니다. 능형 회귀 모수에 대한 자세한 내용 및 이 값이 0이 될 수 없는 이유는 Box and Meyer 연구 자료(1993)에서 확인하십시오.

5 장

단계별 회귀 모형
변수 선택을 사용하여 모형 찾기

단계별 회귀는 회귀 모형에 대한 효과의 부분집합을 선택하기 위한 방법입니다 . 이 방법은 다음과 같은 상황에서 유용하게 사용됩니다 .

- 모형에 대한 항 선택을 설명하는 이론이 거의 없는 경우
- 적합도가 높은 예측 변수를 대화식으로 탐색하려는 경우
- 불필요한 항을 추정함으로써 발생하는 분산을 줄여 모형의 예측 성능을 향상시키려는 경우

범주형 예측 변수에 대해 다음을 수행할 수 있습니다 .

- 연관된 항이 모형에 추가되는 방식을 결정하기 위해 여러 규칙 중에서 선택합니다 .
- 효과 유전성을 적용합니다 .

단계별 플랫폼을 사용하면 가능한 모든 모형을 탐색하고 모형 평균화를 수행할 수도 있습니다.

목차

단계별 회귀 개요 231
예 : 단계별 회귀 사용 231
단계별 보고서 233
단계별 플랫폼 옵션 233
단계별 회귀 제어 패널 234
현재 추정값 보고서 240
단계 기록 보고서 241
교차 , 교호작용 또는 다항식 항이 있는 모형 242
결합 규칙의 예 242
범주형 및 순서형 효과가 있는 모형 244
계층 항 생성 244
명목형 항이 있는 모형의 예 245
계층 항에 대한 제한 규칙의 예 249
이항 및 순서형 로지스틱 단계별 회귀 수행 251
예: 로지스틱 단계별 회귀 사용 252
모든 가능 모형 옵션 253
예: 모든 가능 모형 옵션 사용 254
모형 평균화 옵션 255
예: 모형 평균화 옵션 사용 256
단계별 회귀의 검증 옵션 257
단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합 257
단계별 회귀의 K 폴드 교차 검증 260

단계별 회귀 개요

JMP에서 단계별 회귀는 모형 적합 플랫폼의 분석법입니다. **단계별** 기능은 다른 최소 제곱 플랫폼에서 제공하는 것과 동일한 추정값을 계산하지만 여러 모형 중에서 검색 및 선택을 쉽게 할 수 있습니다.

이 방식에는 주의해야 할 부작용이 있습니다. 선택된 모형은 고정 모형 내에서 검정된 것이 아니므로 모형에 대한 통계량의 유의 수준이 일반적인 통계적 가정을 위반합니다. 긍정적인 면도 있습니다. 이 방식은 30년 동안 항 수를 줄이는 데 유용하게 사용되고 있습니다. Subset Selection in Regression(Miller 1990) 설명서에서 모형 선택 통계량에 통계적 의미를 부여합니다.

이 장에서는 고정 모형에서 계산이 유효함을 나타내기 위해 **유의 확률**이라는 용어를 기계적으로 사용하며, 실제 유의 확률은 보고된 확률과 다르다는 것을 인식하고 있습니다.

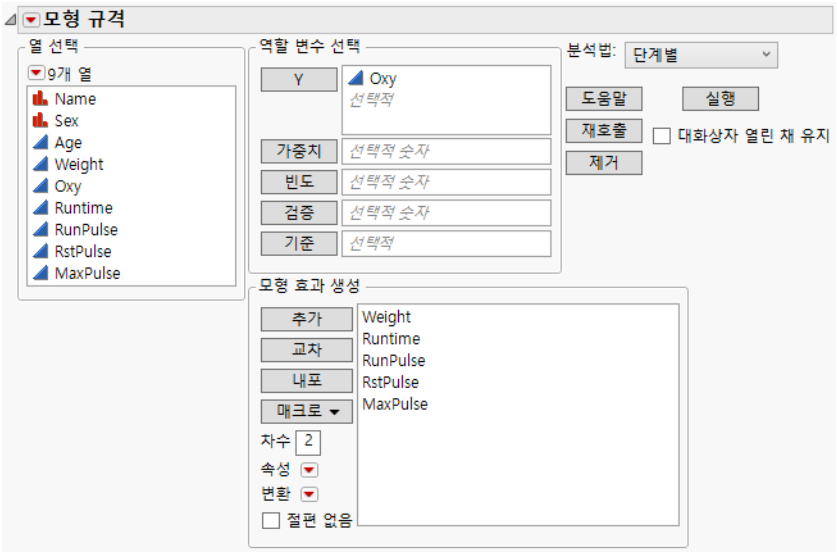
예 : 단계별 회귀 사용

Fitness.jmp 데이터 테이블에는 유산소 피트니스 연구 결과가 포함되어 있습니다. 미리 정해진 거리 동안 러닝머신에서 달리는 사람의 산소 섭취량을 측정하는 특수 검정을 사용하여 유산소 피트니스를 평가할 수 있습니다. 그러나 피트니스를 평가하고 산소 섭취량을 예측하는 간단한 측정을 사용한 계산식을 찾는 것이 더 경제적일 수 있습니다. 이러한 방정식을 식별하기 위해 1.5마일을 달린 31명의 참가자를 대상으로 연령, 체중, 달리기 시간 및 맥박을 측정했습니다.

참고 : 설명을 위해 MaxPulse 및 RunPulse 의 특정 값은 Rawlings(1988, p.105) 에 보고된 데이터에서 변경되었습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Oxy 를 선택하고 **Y** 를 클릭합니다 .
4. Weight, Runtime, RunPulse, RstPulse, MaxPulse 를 선택하고 **추가**를 클릭합니다 .
5. **분석법**에서 **단계별**을 선택합니다 .

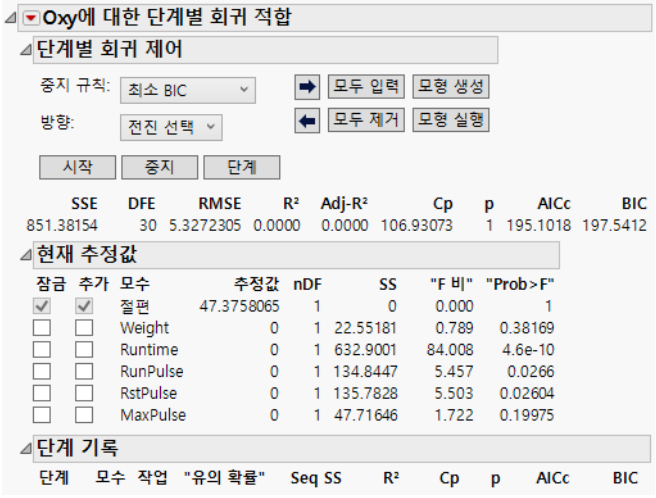
그림 5.1 완료된 모형 적합 시작 창



JMP PRO **검증**은 JMP Pro에서만 사용할 수 있습니다.

6. **실행**을 클릭합니다 .

그림 5.2 단계별 보고서 창



산소 섭취량을 예측하는 적합한 방정식을 찾으려면 서로 다른 여러 회귀 모형을 비교해야 합니다. 단계별 보고서 창의 옵션을 사용하여 효과 조합이 포함된 모형을 검색하고 원하는 모형을 선택합니다.

단계별 보고서

단계별 보고서 창에는 다음과 같은 요소가 포함되어 있습니다 .

플랫폼 옵션 " 단계별 회귀 적합 " 의 빨간색 삼각형 메뉴에는 모든 변수에 영향을 주는 옵션이 포함되어 있습니다 . 자세한 내용은 " 단계별 플랫폼 옵션 "(233 페이지) 에서 확인하십시오 .

단계별 회귀 제어 회귀변수 효과 확률 제한 , 효과 선택 방법 결정 , 선택 프로세스 시작 또는 중지 , 모형 생성 등을 수행하는 컨트롤이 포함되어 있습니다 . 자세한 내용은 " 단계별 회귀 제어 패널 "(234 페이지) 에서 확인하십시오 .

현재 추정값 모형 효과를 추가 , 제거 및 잠글 수 있는 보고서입니다 . 자세한 내용은 " 현재 추정값 보고서 "(240 페이지) 에서 확인하십시오 .

단계 기록 모형에 항을 추가한 효과를 기록하는 보고서입니다 . 자세한 내용은 " 단계 기록 보고서 "(241 페이지) 에서 확인하십시오 .

단계별 플랫폼 옵션

" 단계별 회귀 적합 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

K 폴드 교차 검증 (연속형 반응에만 사용 가능) 선택 프로세스에서 K 폴드 교차 검증을 수행합니다 . 이 옵션을 선택하면 최대 K 폴드 R^2 중지 규칙을 사용할 수 있습니다 (" 단계별 회귀 제어 패널 "(234 페이지)). 검증에 대한 자세한 내용은 " 단계별 회귀의 검증 옵션 "(257 페이지) 에서 확인하십시오 .

모든 가능 모형 (연속형 반응에만 사용 가능) 지정된 한계까지의 가능한 모든 모형을 적합시키고 각 항 수에 대한 최적 모형을 표시합니다 . 하나의 모형에 적합시킬 최대 항 수를 입력합니다 . 또한 모형의 각 항 수에 대해 표시할 최적 모형 결과의 최대 수를 입력합니다 . 범주형 변수는 표시 변수를 사용하여 나타냅니다 . 자세한 내용은 " 범주형 및 순서형 효과가 있는 모형 "(244 페이지) 에서 확인하십시오 . 강한 효과 유전성을 충족하는 모형으로 표시되는 모형을 제한할 수 있습니다 . 자세한 내용은 " 모든 가능 모형 옵션 "(253 페이지) 에서 확인하십시오 .

모형 평균화 (연속형 반응에만 사용 가능) 단일 최적 모형을 선택하는 대신 여러 모형의 적합을 평균화할 수 있습니다 . 자세한 내용은 " 모형 평균화 옵션 "(255 페이지) 에서 확인하십시오 .

기준 기록 그림 AICc 와 BIC 대 모수의 수 그림을 생성합니다 . " 기준 기록 " 그림에는 두 개의 음영 영역이 포함됩니다 . 최소 AICc 값을 V^{best} 로 정의합니다 . 그러면 녹색 영역은 [V^{best}, V^{best}+4] 범위로 정의됩니다 . 노란색 영역은 (V^{best}+4, V^{best}+10] 범위로 정의됩니다 .

R^2 기록 그림 (연속형 반응에만 사용 가능) 훈련 및 검증 R^2 대 모수의 수 그림을 생성합니다 .

기록 지우기 단계 기록을 지우고 재설정합니다 .

JMP PRO 검증과 함께 모형 내보내기 (단계별 시작 창에서 검증 열을 입력한 경우에만 사용 가능) " 모형 생성 " 을 선택하면 " 모형 규격 " 창에 검증 열을 추가합니다 . " 모형 실행 " 을 선택하면 검증 열을 사용하여 모형을 실행합니다 .

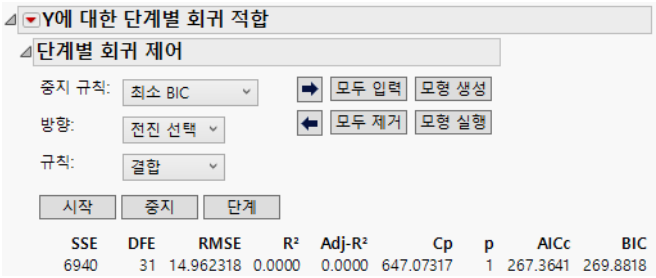
이 옵션은 기본적으로 선택되어 있습니다 .

모형 대화상자 현재 모형에 대해 완료된 " 모형 적합 " 시작 창을 표시합니다 .

단계별 회귀 제어 패널

"단계별 회귀 제어" 패널을 사용하여 회귀변수 효과 확률을 제한하고, 효과 선택 방법을 결정하고, 선택 프로세스를 시작 또는 중지하고, 모형을 실행할 수 있습니다. 제외되거나 결측된 행이 있는지 나타내는 노트가 "시작" 버튼 아래에 표시됩니다.

그림 5.3 단계별 회귀 제어 패널



중지 규칙

중지 규칙은 선택되는 모형을 결정합니다 . "p 값 임계 " 를 제외한 모든 중지 규칙에 대해 " 전진 선택 " 및 " 후진 제거 " 방향만 허용됩니다 . 검증을 사용하는 중지 규칙은 " 최대 검증 R^2" 및 " 최대 K 폴드 R^2" 뿐입니다 . 자세한 내용은 " 단계별 회귀의 검증 옵션 "(257 페이지) 에서 확인하십시오 .

p 값 임계 *p* 값(유의 수준)을 사용하여 모형에서 효과를 입력하고 제거합니다. "p 값 임계"를 선택하면 다음과 같은 두 가지 옵션이 나타납니다 .

선택 유의확률 전진 선택 단계 중에 모형에 효과를 입력해야 하는 최대 *p* 값을 지정합니다 .

제거 유의확률 후진 제거 단계 중에 모형에서 효과를 제거해야 하는 최소 *p* 값을 지정합니다.

최소 AICc 수정된 Akaike 정보 기준 최소값을 사용하여 최적 모형을 선택합니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

최소 BIC Bayes 정보 기준 최소값을 사용하여 최적 모형을 선택합니다. 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

JMP PRO **최대 검증 R^2** 검증 데이터 집합의 최대 R^2 을 사용하여 최적 모형을 선택합니다 . 이 옵션은 두 개 또는 세 개의 구분되는 값을 포함하는 검증 열을 사용하는 경우에만 선택할 수 있습니다 . 검증에 대한 자세한 내용은 " 단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합 "(257 페이지) 에서 확인하십시오 .

최대 K 폴드 R^2 K 폴드 교차 검증의 최대 R^2 을 사용하여 최적 모형을 선택합니다 . 단계별 회귀의 빨간색 삼각형 메뉴에서 "K 폴드 교차 검증 " 옵션을 선택하여 " 최대 K 폴드 R^2" 중지 규칙에 액세스할 수 있습니다 . JMP Pro 사용자는 4 개 이상의 값이 있는 검증 데이터 집합을 사용

하여 옵션에 액세스할 수 있습니다 . 이 옵션을 선택하면 폴드 수를 지정해야 합니다 . 검증에 대한 자세한 내용은 " 단계별 회귀의 K 폴드 교차 검증 "(260 페이지) 에서 확인하십시오 .

방향

선택한 방향은 모형에서 효과가 입력 및 제거되는 방식을 제어합니다 . 다음 옵션 중 하나를 선택합니다 .

전진 선택 *p* 값이 가장 작은 항을 입력합니다 . "p 값 임계 " 중지 규칙을 선택한 경우 해당 항은 **선택 유의확률**에 지정된 수준에서 유의해야 합니다 . 자세한 내용은 " 전진 선택 예 "(238 페이지) 에서 확인하십시오 .

후진 제거 *p* 값이 가장 큰 항을 제거합니다 . "p 값 임계 " 중지 규칙을 선택한 경우 해당 항은 **제거 유의확률**에 지정된 수준에서 유의하지 않아야 합니다 . 자세한 내용은 " 후진 제거 예 "(239 페이지) 에서 확인하십시오 .

참고 : " 방향 " 에서 " 후진 제거 " 를 선택한 경우 " 시작 " 또는 " 단계 " 를 클릭하기 전에 " 모두 입력 " 을 클릭해야 합니다 .

혼합 P 값 중지 규칙을 선택한 경우에만 사용할 수 있습니다 . 이 옵션은 전진 선택 단계와 후진 제거 단계를 교대로 실행합니다 . **선택 유의확률**을 충족하는 가장 유의한 항을 포함하고 **제거 유의확률**을 충족하는 가장 유의하지 않은 항을 제거합니다 . 나머지 항이 유의할 때까지 항을 계속 제거한 후 전진 선택 방향으로 변경됩니다 .

시작 , 중지 , 단계 버튼

"시작", "중지" 및 "단계" 버튼을 사용하면 모형에서 항이 입력 또는 제거되는 방식을 제어할 수 있습니다.

참고 : 모든 중지 규칙은 *p* 값에 따른 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향) 로 정의된 모형만 고려합니다 . 중지 규칙은 모든 가능 모형을 고려하지 않습니다 .

시작 항 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향) 를 자동화합니다 . 적합 모형 중 선택한 중지 규칙에 따라 최적으로 간주되는 모형이 마지막에 나열됩니다 . "p 값 임계 " 중지 규칙을 제외하고 중지 규칙 통계량 동작에서 국소적 감소를 무시하는 모형이 최적 모형으로 선택됩니다 . 최적 모형 오른쪽에 있는 버튼을 클릭하면 " 모형 생성 " 및 " 모형 실행 " 옵션에 사용할 모형이 선택되지만 선택을 변경해도 됩니다 .

- "p 값 임계 " 의 경우 최적 모형은 " 선택 유의확률 " 및 " 제거 유의확률 " 기준을 기반으로 합니다 . 자세한 내용은 "p 값 임계 "(234 페이지) 에서 확인하십시오 .
- " 최소 AICc" 및 " 최소 BIC" 의 경우 최적 모형을 찾을 때까지 자동 적합이 계속됩니다 . 최적 모형은 최소 AICc 또는 BIC 값을 사용한 모형이며 , 각각 AICc 또는 BIC 값이 더 큰 모형이 10 개까지 있을 수 있습니다 . 이 모형은 " 모수 " 열의 " 최선 " 항과 " 작업 " 열의 " 특정 " 항으로 지정됩니다 .

– " 최대 검증 R^2"(JMP Pro 만 해당) 및 " 최대 K 폴드 R^2" 의 경우 최적 모형을 찾을 때까지 자동 적합이 계속됩니다 . 최적 모형은 검증 R^2 또는 R^2 K 폴드 값을 사용한 모형이며 , 각각 검증 R^2 또는 R^2 K 폴드 값이 더 작은 모형이 10 개까지 있을 수 있습니다 . 이 모형은 " 모수 " 열의 " 최선 " 항과 " 작업 " 열의 " 특정 " 항으로 지정됩니다 .

중지 " 시작 " 버튼을 사용하여 시작된 자동 선택 프로세스를 중지합니다 .

단계 항을 하나씩 입력하거나 (전진 선택 방향) 하나씩 제거합니다 (후진 제거 방향). 언제든지 " 단계 기록 " 보고서의 오른쪽에 있는 버튼을 클릭하여 모형을 선택할 수 있습니다 . 선택한 모형 항은 " 현재 추정값 " 보고서에서 업데이트됩니다 . " 모형 생성 " 또는 " 모형 실행 " 을 클릭하면 이 모형이 사용됩니다 .

규칙

참고 : 모형에 관련 항이 포함된 경우에만 나타납니다 . 명목형 또는 순서형 변수가 있으면 관련 항이 생성되고 " 현재 추정값 " 테이블에 나타납니다 .

모형에 항 계층이 있을 때 적용되는 규칙을 변경하려면 " 규칙 " 을 사용합니다 . 계층은 다음과 같은 방식으로 나타날 수 있습니다 .

- 변수가 다른 변수의 성분이면 계층이 생성됩니다 . 예를 들어 모형에 A, B 및 A*B 변수가 포함된 경우 계층에서 A 와 B 는 A*B 에 대한 **상위** 항입니다 .
- 명목형 또는 순서형 변수를 포함하는 경우에도 계층이 생성됩니다 . 트리 구조에서 다른 항 위에 있는 항이 **상위** 항입니다 . 자세한 내용은 "계층 항 생성"(244페이지)에서 확인하십시오.

다음 옵션 중 하나를 선택합니다 .

결합 상위 항목이 있는 항을 입력할 때 두 개의 개별 검정에 대한 p 값을 계산합니다 . 첫 번째 p 값 , p_1 은 항을 상위 항과 그룹화한 후 입력에 대한 해당 그룹의 유의 확률을 결합 F- 검정으로 검정하여 계산됩니다 . 두 번째 p 값 , p_2 는 상위 항이 이미 모형에 추가된 후 해당 항의 입력에 대한 유의 확률을 검정한 결과입니다 . 상위 항목이 있는 항의 입력에 대한 최종 유의 확률은 $\max(p_1, p_2)$ 입니다 .

팁 : 결합 규칙을 사용하면 상위 항에 특히 강한 효과가 있을 수 있는 유의하지 않은 교호작용 항이 포함되지 않습니다 . 이 시나리오에서는 강한 주효과로 인해 입력에 대한 그룹의 유의 확률 p_1 이 매우 작아질 수 있습니다 . 그러나 두 번째 검정에서 교호작용 자체는 유의하지 않습니다 . 결과적으로 p_2 가 더 크고 입력에 대한 최종 유의 확률로 사용됩니다 .

주의 : 상위 항목이 있는 항의 자유도 값은 입력에 대한 두 가지 유의 확률 중 어느 것이 더 큰지에 따라 다릅니다 . 입력에 대한 최종 유의 확률에 사용되는 검정은 " 현재 추정값 " 테이블의 자유도 nDF 를 결정합니다 . 따라서 p_1 이 사용되면 nDF 는 결합 검정을 위한 그룹의 항 수와 같고 , p_2 가 사용되면 nDF 는 1 입니다 .

결합 옵션은 기본 규칙입니다 . 자세한 내용은 "교차, 교호작용 또는 다항식 항이 있는 모형"(242페이지) 에서 확인하십시오 .

제한 상위 항목이 추가될 때까지 상위 항목을 포함한 항을 추가할 수 없도록 제한합니다. 자세한 내용은 " 범주형 및 순서형 효과가 있는 모형 "(244 페이지) 및 " 계층 항에 대한 제한 규칙의 예 "(249 페이지) 에서 확인하십시오.

규칙 없음 선택 루틴으로 인해 계층이 깨지는지 여부에 관계없이 루틴에서 완전히 자유롭게 항을 선택할 수 있습니다.

전체 효과 해당 효과가 포함된 항이 유의한 경우 전체 효과만 입력합니다. 이 규칙은 3 개 이상의 수준이 있는 범주형 변수가 가능한 모형 효과로 입력된 경우에만 적용됩니다. 자세한 내용은 " 규칙 "(249 페이지) 에서 확인하십시오.

버튼

" 단계별 회귀 제어 " 패널에는 다음 버튼이 포함되어 있습니다.

시작 완료될 때까지 선택 프로세스를 자동화합니다.

중지 선택 프로세스를 중지합니다.

단계 선택 프로세스를 한 번에 한 단계씩 증가시킵니다.

화살표 버튼 선택 프로세스에서 한 단계씩 앞뒤로 이동합니다.

모두 입력 잠기지 않은 모든 항을 모형에 입력합니다.

모두 제거 잠기지 않은 모든 항을 모형에서 제거합니다.

모형 생성 " 현재 추정값 " 테이블에 표시된 모형에서 " 모형 적합 " 창에 대한 모형을 생성합니다. 명목형 또는 순서형 항이 있는 경우 **모형 생성**은 모형에 필요한 항을 포함하는 임시 변환 열을 생성합니다.

모형 실행 " 현재 추정값 " 테이블에 표시된 모형을 실행합니다. 명목형 또는 순서형 항이 있는 경우 **모형 실행**은 모형에 필요한 항을 포함하는 임시 변환 열을 생성합니다.

통계량

" 단계별 회귀 제어 " 패널 아래에 다음 통계량이 표시됩니다.

SSE 현재 모형에 대한 오차 제곱합입니다.

DFE 현재 모형에 대한 오차 자유도입니다.

RMSE 현재 모형에 대한 제곱근 평균 제곱 오차 (잔차) 입니다.

R² 랜덤 오차가 아니라 모형의 항에 기인할 수 있는 반응의 변동 비율입니다.

Adj-R² 모수의 개수가 서로 다른 모형들을 비교하기 좋도록 계산에 자유도를 사용하여 R^2 값을 조정합니다. 여러 모형을 검토 중이고 모형의 항 수에 맞게 조정해야 하므로 조정 R^2 은 단계별 절차에서 유용합니다.

Cp 모형 선택을 위한 Mallow C_p 기준입니다. 이 값은 총 제곱 오차의 대체 측도이며 다음과 같이 정의할 수 있습니다.

$$C_p = \left(\frac{\text{SSE}_p}{s^2}\right) - (N - 2p)$$

여기서 s^2 은 전체 모형에 대한 MSE 이고 , SSE_p 는 절편을 포함하여 p 개 변수가 있는 모형에 대한 오차 제곱합입니다 . p 는 x 변수의 수 +1 입니다 . C_p 가 p 로 그래프 표시된 경우 Mallows(1973) 에서는 C_p 가 p 에 처음 접근하는 모형을 선택할 것을 권장합니다 .

p 절편을 포함한 모형의 모수 수입니다 .

AICc 수정된 Akaike 정보 기준입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

BIC Bayes 정보 기준 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

전진 선택 예

전진 선택에서는 모형에 항이 입력되고 모든 항이 유의할 때까지 가장 유의한 항이 추가됩니다 .

1. " 예 : 단계별 회귀 사용 "(231 페이지) 의 단계를 완료합니다 .

   방향은 기본적으로 " 전진 선택 " 으로 되어 있습니다 .

2. **단계**를 클릭합니다 .

   그림 5.4 에서는 한 단계 후 가장 유의한 항인 Runtime 이 모형에 입력되었음을 확인할 수 있습니다 .

3. **시작**을 클릭합니다 .

   그림 5.5 에서는 RstPulse 와 Weight 를 제외한 모든 항이 추가되었음을 확인할 수 있습니다 .

그림 5.4 전진 선택의 한 단계 후 현재 추정값 테이블

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	82.4217727	1	0	0.000	1
☐	☐	Weight	0	1	1.323628	0.171	0.68267
☐	☑	Runtime	-3.3105554	1	632.9001	84.008	4.6e-10
☐	☐	RunPulse	0	1	15.36208	2.118	0.15673
☐	☐	RstPulse	0	1	0.130138	0.017	0.89814
☐	☐	MaxPulse	0	1	1.567361	0.202	0.65632

그림 5.5 전진 선택의 세 단계 후 현재 추정값 테이블

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	80.9007896	1	0	0.000	1
☐	☐	Weight	0	1	4.989591	0.827	0.37137
☐	☑	Runtime	-2.9701867	1	443.2028	73.971	3.25e-9
☐	☑	RunPulse	-0.3751142	1	55.14175	9.203	0.00529
☐	☐	RstPulse	0	1	0.350744	0.056	0.81399
☐	☑	MaxPulse	0.35421891	1	41.34703	6.901	0.01403

후진 제거 예

후진 제거에서는 모든 항이 모형에 입력된 후 나머지 항이 모두 유의할 때까지 가장 유의하지 않은 항이 제거됩니다 .

1. " 예 : 단계별 회귀 사용 "(231 페이지) 의 단계를 완료합니다 .
2. **모두 입력**을 클릭합니다 .

그림 5.6 모형에 입력된 모든 효과

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	82.3936054	1	0	0.000	1
☐	☑	Weight	-0.0509071	1	4.83788	0.772	0.38784
☐	☑	Runtime	-2.9518165	1	366.3375	58.489	5.29e-8
☐	☑	RunPulse	-0.3970425	1	59.51519	9.502	0.00495
☐	☑	RstPulse	0.01239004	1	0.199033	0.032	0.85995
☐	☑	MaxPulse	0.38479281	1	45.83023	7.317	0.01212

3. **방향**에서 " 후진 제거 " 를 선택합니다 .
4. **단계**를 두 번 클릭합니다 .

   첫 번째 후진 제거 단계에서 RstPulse 가 제거되고 두 번째 후진 제거 단계에서 Weight 가 제거됩니다 .

그림 5.7 항이 제거된 현재 추정값 및 단계 기록 테이블

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	80.9007896	1	0	0.000	1
☐	☐	Weight	0	1	4.989591	0.827	0.37137
☐	☑	Runtime	-2.9701867	1	443.2028	73.971	3.25e-9
☐	☑	RunPulse	-0.3751142	1	55.14175	9.203	0.00529
☐	☐	RstPulse	0	1	0.350744	0.056	0.81399
☐	☑	MaxPulse	0.35421891	1	41.34703	6.901	0.01403

단계 기록

단계	모수	작업	"유의 확률"	Seq SS	R^2	Cp	p	AICc	BIC	
1	모두	추가	.	.	0.8161	6	6	157.051	162.22	○
2	RstPulse	제거	0.8600	0.199033	0.8158	4.0318	5	153.721	158.825	○
3	Weight	제거	0.3714	4.989591	0.8100	2.8284	4	151.592	156.362	◉

그림 5.7에 표시된 "현재 추정값" 및 "단계 기록" 테이블에는 후진 제거 단계별 선택 프로세스가 요약되어 있습니다. "단계 기록" 테이블에서 세 번째 단계의 BIC 값이 156.362 입니다. "단계"를 다시 클릭하여 모형에서 다른 모수를 제거하면 BIC 값이 159.984 로 증가합니다. 이러한 이유로 단계 3 모형을 선택합니다. 이 모형은 "시작" 버튼으로 생성하는 모형과 동일합니다.

현재 추정값 보고서

" 현재 추정값 " 보고서를 사용하여 모형 효과를 입력 , 제거 및 잠글 수 있습니다 . 절편은 모형에 영구적으로 잠겨 있습니다 . 이 보고서에는 다음 열이 포함됩니다 .

그림 5.8 현재 추정값 테이블

SSE	DFE	RMSE	R^2	Adj-R^2	Cp	p	AICc	BIC
851.38154	30	5.3272305	0.0000	0.0000	106.93073	1	195.1018	197.5412

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	47.3758065	1	0	0.000	1
☐	☐	Weight	0	1	22.55181	0.789	0.38169
☐	☐	Runtime	0	1	632.9001	84.008	4.6e-10
☐	☐	RunPulse	0	1	134.8447	5.457	0.0266
☐	☐	RstPulse	0	1	135.7828	5.503	0.02604
☐	☐	MaxPulse	0	1	47.71646	1.722	0.19975

잠금 모형에서 항을 잠그거나 잠그지 않습니다 . 선택된 항은 모형에서 입력하거나 제거할 수 없습니다 .

추가 항이 현재 모형에 있는지 여부를 나타냅니다 . 항의 체크박스를 클릭하여 효과를 수동으로 모형에 가져오거나 모형에서 내보낼 수 있습니다 .

모수 효과 이름입니다 .

추정값 현재 모수 추정값이며 , 효과가 현재 모형에 없으면 값이 0 입니다 .

nDF 항의 자유도 수입니다 . 항이 모형에 입력되면 다른 항도 모형에 추가되는 경우 항의 자유도는 두 개 이상입니다 .

Wald/ 점수 카이제곱 (반응이 범주형인 경우에만 표시됨) 모형의 각 항에 대한 검정 통계량입니다 . 모형에 아직 포함되지 않은 항에 대한 검정 통계량은 모형에 항을 포함하는 스코어 검정을 기반으로 합니다 . 모형에 이미 포함된 항에 대한 검정 통계량은 모형에서 항을 제거하는 Wald 검정을 기반으로 합니다 .

" 유의 확률 " (반응이 범주형인 경우에만 표시됨) nDF 자유도에 기반한 Wald/ 점수 카이제곱 검정 통계량과 관련된 *p* 값입니다 . " 유의 확률 " 은 모형에 포함할 다음 항을 결정하는 데 사용됩니다 .

SS (반응이 연속형인 경우에만 표시됨) 항이 모형에 입력될 경우 오차 (잔차) SS(제곱합) 의 감소 또는 항이 모형에서 제거될 경우 오차 SS 의 증가입니다 . 항이 특정 방식으로 제한되면 SS 가 0 으로 보고될 수 있습니다 .

F 비 (반응이 연속형인 경우에만 표시됨) 항 효과가 0 임을 검정하기 위한 기존의 검정 통계량입니다 . 이 값은 *t* 비의 제곱입니다 . 모형이 적합한 것으로 선택되었지만 항을 검정하기 위한 *F* 분포가 없으므로 인용 부호로 묶여 있습니다 .

"Prob>F" (반응이 연속형인 경우에만 표시됨) *F* 통계량과 관련된 *p* 값입니다 . 실제 유의 확률로 신뢰할 수 없으므로 "F 비 " 와 마찬가지로 인용 부호로 묶여 있습니다 .

R 모형의 다른 효과와의 다중 상관입니다 . 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > R 을 선택한 경우에만 나타납니다 .

단계 기록 보고서

각 단계를 수행할 때 모형에 항을 추가한 효과가 "단계 기록" 보고서에 기록됩니다. 예를 들어 Fitness.jmp 예의 "단계 기록" 보고서에는 모형에 항이 입력된 순서 및 각 모형에 대한 통계량이 표시됩니다. 자세한 내용은 "예: 단계별 회귀 사용"(231 페이지)에서 확인하십시오.

오른쪽의 라디오 버튼을 사용하여 모형을 선택합니다.

그림 5.9 단계 기록 보고서

단계 기록

단계	모수	작업	"유의 확률"	Seq SS	R^2	Cp	p	AICc	BIC	
1	Runtime	추가	0.0000	632.9001	0.7434	7.8825	2	155.397	158.81	○
2	RunPulse	추가	0.1567	15.36208	0.7614	7.4298	3	155.787	159.984	○
3	MaxPulse	추가	0.0140	41.34703	0.8100	2.8284	4	151.592	156.362	◉

범주형 반응에 대한 단계 기록

범주형 반응이 있는 모형의 "단계 기록" 보고서에는 L-R 카이제곱, "유의 확률", 카이제곱 입력 및 "유의 확률" 입력의 네 가지 추가 열이 포함됩니다.

L-R 카이제곱 및 "유의 확률" 열에는 전체 모형 대 축소 모형의 가능도비 검정 통계량과 p 값이 포함됩니다. 여기서 전체 모형은 지정된 항을 포함하는 모형이고 축소 모형은 지정된 항을 포함하지 않습니다.

카이제곱 입력 및 "유의 확률" 입력 열에는 모형에 포함할 가장 최근 항을 선택하는 데 사용된 Wald/점수 카이제곱 및 "유의 확률" 값이 포함됩니다.

그림 5.10 범주형 반응에 대한 단계 기록 (첫 번째 단계 전)

현재 추정값

잠금	추가	모수	추정값	nDF	Wald/점수 카이제곱	"유의 확률"
☑	☑	절편[M]	-0.0645385	1	0	1
☐	☐	Oxy	0	1	7.275453	0.00699
☐	☐	Weight	0	1	10.13425	0.00146
☐	☐	Runtime	0	1	5.764797	0.01635
☐	☐	RunPulse	0	1	1.865024	0.17205
☐	☐	RstPulse	0	1	2.693096	0.10078
☐	☐	MaxPulse	0	1	1.467541	0.22573

단계 기록

단계	모수	작업	L-R 카이제곱	"유의 확률"	카이제곱 입력	"유의 확률" 입력	R^2	p	AICc	BIC

그림 5.11 범주형 반응에 대한 단계 기록 (첫 번째 단계 후)

현재 추정값

잠금	추가	모수	추정값	nDF	Wald/ 점수 카이제곱	"유의 확률"
☑	☑	절편[M]	-16.228294	1	0	1
☐	☐	Oxy	0	1	8.750599	0.0031
☐	☑	Weight	0.20792272	1	6.994297	0.00818
☐	☐	Runtime	0	1	6.371395	0.0116
☐	☐	RunPulse	0	1	1.428737	0.23197
☐	☐	RstPulse	0	1	3.510878	0.06097
☐	☐	MaxPulse	0	1	0.365726	0.54534

단계 기록

단계	모수	작업	L-R 카이제곱	"유의 확률"	카이제곱 입력	"유의 확률" 입력	R^2	p	AICc	BIC
1	Weight	추가	12.16669	0.0005	10.1342	0.00146	0.2833	2	35.2047	37.6441

그림 5.11 의 "단계 기록" 테이블에서 Weight 에 대한 카이제곱 입력 및 "유의 확률" 입력 값은 그림 5.10 의 Weight 에 대한 Wald/ 점수 카이제곱 및 "유의 확률" 값과 일치합니다.

교차 , 교호작용 또는 다항식 항이 있는 모형

일부 모형, 특히 실험 설계와 관련된 모형은 교호작용 항을 포함합니다. 연속형 요인의 경우 교호작용은 효과를 나타내는 열의 곱입니다. 명목형 및 순서형 요인의 경우 교호작용은 범주형 수준을 나타내는 항의 곱이 포함된 모형 항으로 정의됩니다.

교호작용 항이 있으면 고차 효과의 저차 성분이 모형에 포함되도록 모형 선택 프로세스에 제한을 적용하려는 경우가 종종 있습니다. 이 방법은 효과 유전성 원리에 의해 제안됩니다. 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오. 예를 들어 이원 교호작용이 모형에 포함된 경우 성분 주효과(**상위**)도 포함되어야 합니다.

결합 규칙의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. " 차수 " 상자에 2 를 입력합니다 .
5. F, Ct, A, T 및 Cn 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **분석법**에서 **단계별**을 선택합니다 .
7. **실행**을 클릭합니다 .

그림 5.12 결합 규칙을 사용한 초기 현재 추정값 보고서

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	65.5	1	0	0.000	1
☐	☐	F	0	1	15.125	0.066	0.79972
☐	☐	Ct	0	1	3042	23.412	3.67e-5
☐	☐	A	0	1	3.125	0.014	0.90823
☐	☐	T	0	1	924.5	4.611	0.03998
☐	☐	Cn	0	1	312.5	1.415	0.24363
☐	☐	F*Ct	0	1	15.125	0.066	0.79972
☐	☐	F*A	0	3	22.75	0.031	0.9926
☐	☐	F*T	0	1	6.125	0.027	0.87178
☐	☐	F*Cn	0	1	0.125	0.001	0.98161
☐	☐	Ct*A	0	1	6.125	0.027	0.87178
☐	☐	Ct*T	0	1	1404.5	7.612	0.00979
☐	☐	Ct*Cn	0	1	32	0.139	0.71193
☐	☐	A*T	0	1	36.125	0.157	0.69476
☐	☐	A*Cn	0	1	6.125	0.027	0.87178
☐	☐	T*Cn	0	1	968	4.863	0.03525

그림 5.12의 모형에는 5개의 연속형 요인에 대한 최대 2요인 교호작용의 모든 항이 포함되어 있습니다. "규칙"(236페이지)에 설명된 결합, 제한 및 전체 효과 규칙을 사용하여 교호작용 항 입력을 제어할 수 있습니다.

결합 규칙은 두 가지 검정을 기반으로 교호작용 항의 입력을 결정합니다. 자세한 내용은 "결합"(236페이지)에서 확인하십시오. 자유도(nDF)를 기반으로 두 검정 중 *p* 값에 사용된 검정을 확인할 수 있습니다. 예를 들어 교호작용 항 F*A의 nDF 값은 3입니다. 이는 F*A가 상위 항 F 및 A로 그룹화되고, 3 자유도 결합 *F*-검정을 기반으로 하여 입력 대상으로 고려된다는 의미입니다. 반면에 교호작용 항 Ct*T의 nDF 값은 1입니다. 이는 상위 항 Ct와 T가 이미 모형에 포함된 후 Ct*T의 유의성을 검정하는 1 자유도 *F*-검정을 기반으로 하여 Ct*T가 입력 대상으로 고려된다는 의미입니다. **단계**를 한 번 클릭하여 Ct가 단독으로 입력되는지 확인합니다.

8. **단계**를 다시 클릭하여 Ct*T 가 T 와 함께 입력되는지 확인합니다 (Ct 는 이미 모형에 있음).

그림 5.13 결합 규칙을 사용한 현재 추정값 보고서 (한 단계)

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	65.5	1	0	0.000	1
☐	☐	F	0	1	15.125	0.263	0.61236
☐	☑	Ct	9.75	2	4446.5	39.676	6.74e-9
☐	☐	A	0	1	3.125	0.054	0.81819
☐	☑	T	5.375	2	2329	20.781	2.93e-6
☐	☐	Cn	0	1	312.5	6.715	0.01524
☐	☐	F*Ct	0	2	30.25	0.256	0.7764
☐	☐	F*A	0	3	22.75	0.123	0.9459
☐	☐	F*T	0	2	21.25	0.178	0.83755
☐	☐	F*Cn	0	1	0.125	0.002	0.96335
☐	☐	Ct*A	0	2	9.25	0.077	0.92601
☐	☑	Ct*T	6.625	1	1404.5	25.064	2.72e-5
☐	☐	Ct*Cn	0	1	32	0.562	0.45989
☐	☐	A*T	0	2	39.25	0.334	0.7194
☐	☐	A*Cn	0	1	6.125	0.106	0.74747
☐	☐	T*Cn	0	1	968	43.488	4.49e-7

유의한 교호작용 항이 있는 경우 같은 단계에서 여러 항을 입력할 수 있습니다. **단계** 버튼을 두 번 클릭하면 Ct*T는 포함된 두 효과 Ct 및 T와 함께 입력됩니다. 그러나 두 성분 항을 제거하지 않으

면 교차 항을 제거할 수 없으므로 한 단계 뒤로는 대칭이 아닙니다. 이제 Ct와 T의 자유도는 각각 2입니다. 이는 단계별 방법에서 Ct 또는 T를 제거하는 경우 Ct*T도 제거해야 하기 때문입니다. "방향"을 **후진 제거**로 변경하고 **단계**를 클릭하면 Ct*T가 제거되고 Ct 및 T의 자유도가 1로 변경됩니다.

범주형 및 순서형 효과가 있는 모형

일반적으로 단계별 회귀에서는 모형에 범주형 효과가 있는 상황을 다루지 않았습니다 . 다음 사항에 유의하십시오 .

- 회귀 모형에 명목형 또는 순서형 효과가 포함된 경우 해당 효과는 표시자 열 집합으로 나타냅니다 .
- 범주형 효과에 수준이 두 개뿐인 경우 해당 효과는 단일 열로 나타냅니다 .
- 범주형 효과에 *k* 개 수준이 있는 경우 (*k* > 2) 이 효과는 k-1 개 열로 나타내야 합니다 .

JMP에서 사용되는 표준 플랫폼에 대한 규칙은 모수 추정값 평균이 전체 수준에서 0인 항으로 명목형 변수를 나타내는 것입니다.

단계별 플랫폼에서 범주형 변수(명목형 및 순서형)는 **계층적** 방식으로 코드화됩니다. 이것은 다른 최소 제곱 적합 플랫폼의 코딩과 다릅니다. 계층적 코딩에서 범주형 변수의 수준은 반응의 평균을 가장 많이 분리하는 수준 그룹으로 연속하여 분할됩니다. 분할을 통해 *k* - 1개 항으로 *k* 수준 범주형 변수를 나타낼 수 있습니다.

참고 : 계층적 코딩에서 생성된 초기 항은 가장 큰 분리를 담당하는 그룹을 나타냅니다 . 이 코딩 체계의 장점은 이러한 정보화 항이 모형에 일찍 입력될 가능성이 있다는 것입니다 .

계층 항 생성

계층 항은 파티션 분석과 유사한 트리 구조를 사용하여 생성됩니다. 그러나 최대화되는 기준은 그룹 간 제곱합(SSB) 입니다.

수준이 *k* 개인 명목형 변수의 경우 *k* 개 수준은 SSB가 가장 큰 두 개의 수준 그룹으로 분할됩니다. 이러한 두 수준 그룹을 A1과 A2라고 합니다. 여기서 A1의 평균이 더 작고 A2의 평균이 더 큽니다. A1과 A2의 두 수준 그룹은 A1의 수준 값이 1이고 A2의 수준 값이 -1인 표시 변수를 정의하는 데 사용됩니다. 이 변수는 명목형 변수에 대한 첫 번째 계층 항입니다.

초기 두 그룹 A1과 A2 내의 각 수준에 대해 SSB가 가장 큰 두 개의 수준 그룹으로 분할됩니다. SSB가 가장 큰 수준 그룹이 A1의 수준에 속한다고 가정해 보겠습니다. 이러한 두 그룹을 B1과 B2라고 합니다. 여기서 B1의 평균이 더 작고 B2의 평균이 더 큽니다. B1과 B2의 두 수준 그룹은 B1의 수준 값은 1, B2의 수준 값은 -1, A2의 수준 값은 0인 계층 변수를 정의하는 데 사용됩니다. 다음 변수를 생성하려면 B1, B2 및 A2의 수준 분할을 고려해야 합니다. SSB를 최대화하는 분할이 다음 계층 변수를 정의합니다. 이 프로세스는 *k-1* 개 계층 항이 생성될 때까지 계속됩니다.

순서형 변수의 경우 분할에서 고려되는 수준 그룹에는 순서가 인접한 수준만 포함됩니다. 이렇게 하면 생성된 항이 수준 순서를 준수합니다.

규칙 및 계층 항

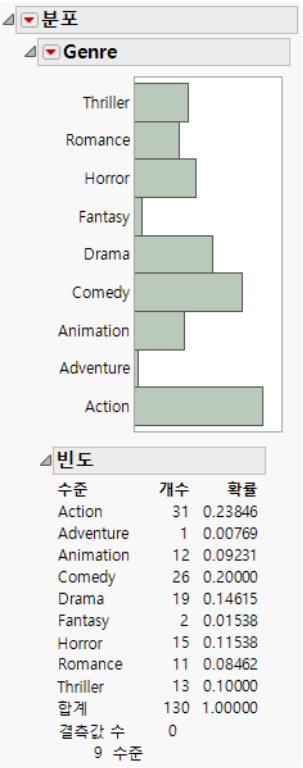
결합 규칙 또는 **제한** 규칙을 사용하는 경우 계층에서 위에 있는 모든 항이 입력되지 않으면 해당 항을 모형에 입력할 수 없습니다. **전체 효과** 규칙을 사용하고 범주형 변수에 대한 항을 입력하면 관련된 모든 항이 입력됩니다. 예는 "예제의 계층 항 생성"(248 페이지)에서 확인하십시오.

명목형 항이 있는 모형의 예

이 예에서는 2011 년에 개봉한 영화 관련 데이터를 사용합니다 . 특히 총 수입액을 나타내는 World Gross 값에 관심이 있습니다 . 잠재적 예측 변수는 Rotten Tomatoes Score, Audience Score 및 Genre 입니다 . 두 점수 (Score) 변수는 연속형이지만 Genre 는 명목형입니다 . 단계별 회귀를 사용하여 모형을 축소하기 전에 관심 변수를 탐색하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hollywood Movies.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Genre 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

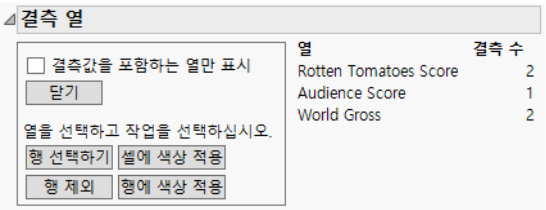
그림 5.14 Genre 분포



Genre 에 9 개 수준이 있으므로 8 개의 모형 항으로 표시됩니다 . 결측 데이터 때문에 추가 데이터 탐색 결과 8 개 수준만 단계별 회귀에 고려됩니다 .

5. 데이터 테이블의 " 열 " 패널에서 관심 열을 선택합니다 (Rotten Tomatoes Score, Audience Score, World Gross).
6. **분석 > 선별 > 결측값 탐색**을 선택합니다 .
7. **Y, 열**을 클릭한 후 **확인**을 클릭합니다 .

그림 5.15 결측 열 보고서



Rotten Tomatoes Score 의 2 개 행 , Audience Score 의 1 개 행 , World Gross 의 2 개 행에 각각 결측값이 있습니다 .

8. " 결측 열 " 보고서에서 **열** 아래에 나열된 세 열을 선택합니다 .
9. **행 선택하기**를 클릭합니다 .

   데이터 테이블의 " 행 " 패널에서 3 개 행이 선택되어 있음을 확인할 수 있습니다 . 이 세 행에는 예측 변수 또는 반응에 대한 결측 데이터가 포함되어 있으므로 단계별 분석에서 자동으로 제외됩니다 . Adventure 범주에 포함된 항목은 128 번 행뿐이므로 해당 범주는 분석에서 완전히 제거됩니다 . 결과적으로 단계별 분석에 사용할 Genre 에는 8 개 범주만 있습니다 . 결측 데이터의 효과를 확인했으므로 이제 단계별 분석을 수행합니다 .
10. **분석 > 모형 적합**을 선택합니다 .
11. Rotten Tomatoes Score, Audience Score 및 Genre 를 선택하고 **추가**를 클릭합니다 .

    Rotten Tomatoes Score, Audience Score 및 Genre 를 예측 변수로 사용하여 표준 최소 제곱 모형을 World Gross 에 적합시키면 잔차의 이분산성이 매우 커집니다 . 이것은 재무 데이터의 대표적인 예입니다 . 로그 변환을 사용하면 등분산 회귀 가정을 더 잘 충족시킬 수 있습니다 .
12. " 열 선택 " 목록에서 World Gross 를 마우스 오른쪽 버튼으로 클릭하고 **변환 > 로그**를 선택합니다 .

    변환된 변수 로그 [World Gross] 가 " 열 선택 " 목록 아래쪽에 나타납니다 .
13. 로그 [World Gross] 를 선택하고 **Y** 를 클릭합니다 .
14. " 분석법 " 목록에서 **단계별**을 선택합니다 .
15. **실행**을 클릭합니다 .

그림 5.16 모형 항 목록을 보여 주는 현재 추정값 테이블

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	4.09998703	1	0	0.000	1
☐	☐	Rotten Tomatoes Score	0	1	10.78976	3.303	0.07154
☐	☐	Audience Score	0	1	8.715114	2.655	0.10577
☐	☐	Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation}	0	1	49.78148	16.850	7.25e-5
☐	☐	Genre{Drama&Horror&Thriller-Fantasy&Romance&Comedy}	0	1	9.560063	2.918	0.09008
☐	☐	Genre{Drama-Horror&Thriller}	0	1	2.027141	0.608	0.43718
☐	☐	Genre{Horror-Thriller}	0	1	0.013011	0.004	0.95043
☐	☐	Genre{Fantasy&Romance-Comedy}	0	1	1.428612	0.428	0.51439
☐	☐	Genre{Fantasy-Romance}	0	1	0.000376	0.000	0.99157
☐	☐	Genre{Action-Animation}	0	1	1.362765	0.408	0.52426

" 현재 추정값 " 테이블에서 Genre 는 7 개 항으로 표시됩니다 . 이 중 두 개의 항으로 모형을 생성하여 이러한 항이 어떻게 정의되는지 확인합니다 .

16. **추가** 아래에서 Genre 의 처음 두 항 옆에 있는 상자를 선택합니다 .
    - Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation}
    - Genre{Drama&Horror&Thriller-Fantasy&Romance&Comedy}
17. **모형 생성**을 클릭합니다 .

    이 두 항이 " 모형 규격 " 창의 모형 효과 목록에 임시 변환 열로 추가됩니다 . 다음 섹션에서 이 열에 대해 설명합니다 .

예제의 계층 항 생성

결측값으로 인해 Genre는 8개 수준의 명목형 변수입니다. "현재 추정값" 테이블에는 Genre가 7개 항으로 표시됩니다. Genre의 수준이 8개이므로 이것은 적절합니다. 아래에서는 Genre를 나타내는 처음 두 항에 대해 설명합니다. 후속 항은 이와 비슷한 방식으로 정의됩니다.

첫 번째 항

처음 나타나는 행은 Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation} 입니다. 이 변수의 형식은 Genre{A1 - A2} 이며 , A1 과 A2 는 빼기 기호 (-) 로 구분됩니다. 이 표기는 다음 두 수준 집합 사이에 그룹 간 제곱합을 기준으로 최대 분리가 발생한다는 것을 나타냅니다.

- Drama, Horror, Thriller, Fantasy, Romance, Comedy(A1 로 표시)
- Action, Animation(A2 로 표시)

Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation} 항을 모형에 포함하면 해당 항을 나타내는 임시 변환 열이 모형에 사용됩니다. 열에 다음 값이 포함됩니다.

- 1 - Drama, Horror, Thriller, Fantasy, Romance, Comedy
- -1 - Action, Animation

두 번째 항

두 번째로 나타나는 항은 Genre{Drama&Horror&Thriller-Fantasy&Romance&Comedy} 입니다. 이 수준 집합은 첫 번째 항의 첫 번째 분할 (A1) 에 완전히 포함됩니다. 이 표기는 다음 수준을 대조합니다.

- Drama, Horror, Thriller
- Fantasy, Romance, Comedy

알고리즘은 Drama, Horror, Thriller, Fantasy, Romance, Comedy 수준(A1)과 Action, Animation 수준(A2)의 모든 분할 중에서 이 분할이 그룹 간 제곱합이 가장 큰 것으로 결정합니다.

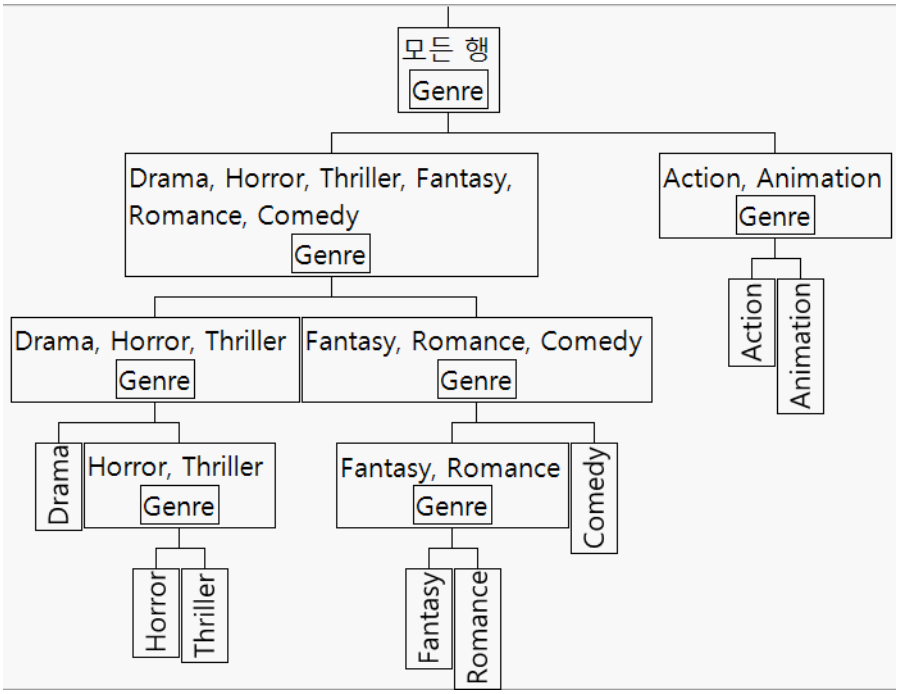
이 항을 모형에 포함하면 해당 항을 나타내는 임시 변환 열이 모형에 사용됩니다. 열에 다음 값이 포함됩니다.

- 1 - Drama, Horror, Thriller
- -1 - Fantasy, Romance, Comedy
- 0 - Action, Animation

항 계층

그룹 간 제곱합 기준에 따라 항 분할이 계속됩니다 . 그림 5.17 에서는 항 정의를 생성하는 계층을 보여 줍니다 .

그림 5.17 계층적 코딩에 사용되는 분할을 보여 주는 트리



규칙

결합 규칙 또는 **제한** 규칙을 사용하는 경우 계층에서 위에 있는 모든 항이 입력되지 않으면 해당 항을 모형에 입력할 수 없습니다. 예를 들어 Genre{Action-Animation} 을 입력하면 JMP 에서 Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation}도 입력합니다.

전체 효과 규칙을 사용하고 Genre 항 중 하나를 입력하면 모든 Genre 항이 입력됩니다.

계층 항에 대한 제한 규칙의 예

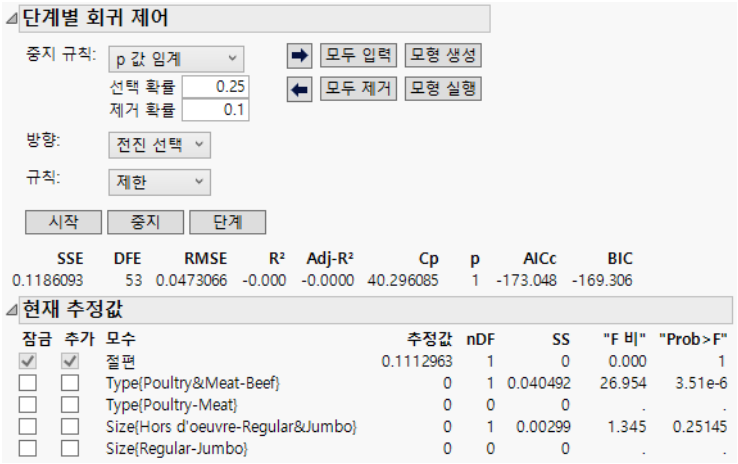
명목형 또는 순서형 항이 있는 모형의 경우 모형을 생성하거나 실행하면 모형에 포함된 계층 항을 포함하는 임시 변환 열이 모형 적합에 사용됩니다. 모형 자체는 새 "모형 적합" 창에 나타납니다. 이 예에서는 **단계별** 회귀에서 계층적 효과가 있는 모형을 생성하는 방법을 자세히 보여 줍니다.

핫도그 종류 Type(Meat, Beef, Poultry) 과 핫도그 크기 Size(Jumbo, Regular, Hors d'oeuvre) 의 함수로 핫도그의 온스당 비용 ($/oz) 을 조사하는 간단한 모형이 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hot Dogs2.jmp 를 엽니다 .

2. **분석 > 모형 적합**을 선택합니다 .
3. $/oz 를 선택하고 **Y** 를 클릭합니다 .
4. Type 과 Size 를 선택하고 **추가**를 클릭합니다 .
5. **분석법**에서 **단계별**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. **중지 규칙**에서 **p 값 임계**를 선택합니다 .
8. **규칙**에서 **제한**을 선택합니다 .

그림 5.18 P 값 임계 및 제한 규칙이 설정된 단계별 회귀 제어 패널



기본 규칙으로 선택된 " 결합 " 을 " 제한 " 으로 변경하면 두 항에 대한 "F 비 " 및 "Prob > F" 값이 결측으로 표시됩니다 . 해당 항은 Type{Poultry-Meat} 와 Size{Regular-Jumbo} 입니다 . 이러한 두 항은 상위 항이 입력될 때까지 모형에 입력될 수 없기 때문입니다 .

9. **단계**를 클릭합니다 .

Type{Poultry&Meat-Beef} 항이 모형에 입력됩니다 . 이 항은 "Prob>F" 값이 가장 작으며 해당 값은 " 선택 유의확률 " 임계값 0.25 미만입니다 .

그림 5.19 하나의 항이 추가된 단계별 제어판

Type{Poultry-Meat} 항에 대한 "F 비 " 및 "Prob > F" 값이 나타납니다 . 상위 항이 모형에 입력되었으므로 이제 Type{Poultry-Meat} 를 입력할 수 있습니다 .

10. **단계**를 클릭합니다 .

Type{Poultry-Meat} 는 나머지 항 중 "Prob>F" 값이 가장 작고 해당 값이 " 선택 유의확률 " 임계값 미만이므로 모형에 입력할 다음 항입니다 .

11. **단계**를 클릭합니다 .

"Prob>F" 값이 0.1577 인 Size{Hors d'oeuvre-Regular&Jumbo} 항이 모형에 입력됩니다 . 이제 상위 항이 모형에 있으므로 Size{Regular-Jumbo} 항을 모형에 입력할 수 있으며 "Prob>F" 값이 나타납니다 .

그러나 Size{Regular-Jumbo} 항의 "Prob>F" 값은 " 선택 유의확률 " 값 0.25 를 초과하는 0.7566 입니다 . 따라서 " 단계 " 를 다시 클릭하면 모형에 항이 입력되지 않습니다 .

그림 5.20 최종 모형에 대한 현재 추정값 보고서

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☑	☑	절편	0.10849381	1	0	0.000	1
☐	☑	Type{Poultry&Meat-Beef}	-0.0275278	0	0	.	.
☐	☑	Type{Poultry-Meat}	-0.0205527	1	0.013985	11.083	0.00164
☐	☑	Size{Hors d'oeuvre-Regular&Jumbo}	-0.0121982	1	0.002596	2.057	0.1577
☐	☐	Size{Regular-Jumbo}	0	1	0.000125	0.097	0.7566

팁 : 전체 단계별 프로세스를 자동으로 실행하려면 " 시작 " 버튼을 사용하십시오 . 이 작업을 확인하려면 **모두 제거**를 클릭합니다 . 그런 다음 **시작**을 클릭합니다 .

12. **모형 생성**을 클릭합니다 .

" 모형 생성 " 을 클릭하면 단계별 프로세스에서 선택된 세 가지 모형 효과만 포함하는 모형 적합 시작 창이 나타납니다 . 시작 창에서 모형에 입력된 세 가지 계층적 효과를 정의하는 임시 변환 열이 모형에 포함됩니다 .

이항 및 순서형 로지스틱 단계별 회귀 수행

모형 적합의 단계별 분석법은 반응이 순서형 또는 명목형일 때 순서형 로지스틱 단계별 회귀를 수행합니다. 명목형 반응은 로지스틱 단계별 회귀 적합 절차에서 순서형 반응으로 처리됩니다. 반응의 수준이 두 개뿐인 경우 순서형 로지스틱 회귀 모형은 명목형 로지스틱 회귀 모형과 동일합니다. 로지스틱 단계별 회귀를 실행하려면 순서형 또는 명목형 반응을 지정하고, 모형에 항을 추가한 후 "분석법" 메뉴에서 "단계별"을 선택합니다.

로지스틱 모형에 대한 단계별 보고서는 반응이 연속형일 때 제공되는 보고서와 비슷합니다 . 다음 요소는 로지스틱 회귀 결과에만 해당됩니다 .

- 반응이 범주형이면 모형의 전체 적합이 음의 로그 가능도 (-1* 로그 가능도) 로 제공됩니다 . 이 값은 전체 반복 최대 가능도 적합을 기반으로 계산됩니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .
- " 현재 추정값 " 보고서에는 카이제곱 통계량 ("Wald/ 점수 카이제곱 ") 및 해당 *p* 값 (" 유의 확률 ") 이 표시됩니다 . 검정 통계량 열에는 현재 모형에 없는 모수에 대한 점수 통계량 및 현재 모형에 있는 모수에 대한 Wald 통계량이 표시됩니다 . 회귀 추정값 (" 추정값 ") 은 전체 반복 최대 가능도 적합을 기반으로 합니다 .

- " 단계 기록 " 보고서에는 "L-R 카이제곱 " 이 표시됩니다 . 이 값은 모형의 다른 항이 주어졌을 때 해당 회귀 모수가 0 이라는 가설의 가능도비 검정에 대한 검정 통계량입니다 . " 유의 확률 " 은 이 검정의 *p* 값입니다 .

참고 : 반응이 명목형인 경우 " 모형 생성 " 버튼을 클릭하면 모형 적합의 명목형 로지스틱 분석법을 사용하여 현재 모형을 적합시킬 수 있습니다 . " 모형 적합 " 시작 창이 나타나면 " 실행 " 을 클릭합니다 .

예 : 로지스틱 단계별 회귀 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Sex 를 선택하고 **Y** 를 클릭합니다 .
4. Weight, Runtime, RunPulse, RstPulse 및 MaxPulse 를 선택하고 **추가**를 클릭합니다 .
5. **분석법**에서 **단계별**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. **시작**을 클릭합니다 .

그림 5.21 로지스틱 단계별 보고서

현재 추정값

잠금	추가	모수	추정값	nDF	Wald/ 점수 카이제곱	"유의 확률"
☑	☑	절편[M]	-35.807846	1	0	1
☐	☑	Weight	0.28229655	1	5.471429	0.01933
☐	☑	Runtime	1.32680366	1	2.545732	0.11059
☐	☐	RunPulse	0	1	0.470408	0.4928
☐	☐	RstPulse	0	1	0.007216	0.9323
☐	☐	MaxPulse	0	1	0.076497	0.7821

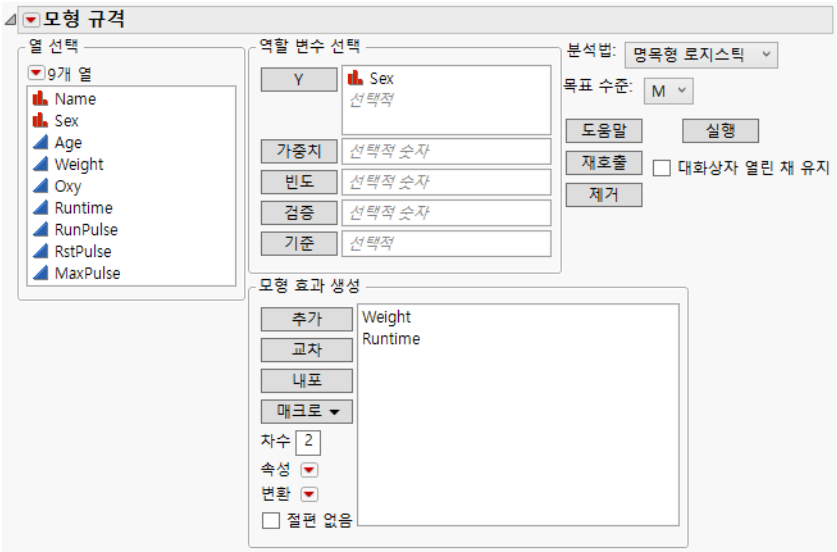
단계 기록

단계	모수	작업	L-R 카이제곱	"유의 확률"	카이제곱 입력	"유의 확률" 입력	R^2	p	AICc	BIC
1	Weight	추가	12.16669	0.0005	10.1342	0.00146	0.2833	2	35.2047	37.6441
2	Runtime	추가	8.017826	0.0046	6.37139	0.0116	0.4700	3	29.6472	33.0603
3	RunPulse	추가	1.440088	0.2301	1.44684	0.22903	0.5036	4	30.8567	35.0542
4	MaxPulse	추가	0.071809	0.7887	0.06994	0.79142	0.5052	5	33.6464	38.4164
5	RstPulse	추가	0.007156	0.9326	0.00722	0.93228	0.5054	6	36.7393	41.8432
6	최선	특정	.	.	0.00722	0.93228	0.4700	3	29.6472	33.0603

중지 규칙에 따라 Weight 와 Runtime 두 변수가 모형에 입력됩니다 .

8. **모형 생성**을 클릭합니다 .

그림 5.22 축소 모형에 대한 모형 규격 창



두 변수가 모형 효과로 포함된 모형 규격 창이 나타납니다 . " 분석법 " 이 " 명목형 로지스틱 " 으로 선택되어 있습니다 . 반응이 순서형인 경우 " 분석법 " 은 " 순서형 로지스틱 " 입니다 .

모든 가능 모형 옵션

"모든 가능 모형" 옵션을 사용하면 예측 변수를 사용하여 생성할 수 있는 모든 모형을 조사할 수 있습니다. 이 옵션은 단계별 회귀의 빨간색 삼각형에서 액세스합니다.

다음 사항에 유의하십시오 .

- 이 옵션은 모형 수가 5 백만 개를 초과하는 큰 문제에는 실용적이지 않습니다 .
- 범주형 예측 변수는 표시 변수로 나타냅니다 . 자세한 내용은 " 범주형 및 순서형 효과가 있는 모형 "(244 페이지) 에서 확인하십시오 .

다음 옵션은 표시되는 모형 수를 제한합니다 .

모형의 최대 항 수 : 모형의 최대 항 수 값을 입력합니다 .

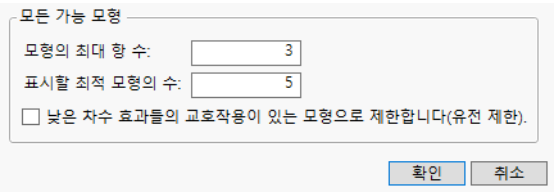
표시할 최적 모형의 수 : 표시할 각 크기의 최대 모형 수를 입력합니다 . R^2 값에 따라 최적 모형이 나타납니다 .

낮은 차수 효과들의 교호작용이 있는 모형으로 제한합니다 (유전 제한). 고차 효과가 포함된 경우 모든 저차 효과를 포함하는 모형만 표시합니다 . 이러한 모형은 강한 효과 유전성을 충족합니다 . 이 옵션은 예측 변수에 교호작용 또는 다항식 항이 포함된 경우에 유용합니다 .

예 : 모든 가능 모형 옵션 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Oxy 를 선택하고 **Y** 를 클릭합니다 .
4. Runtime, RunPulse, RstPulse 및 MaxPulse 를 선택하고 **추가**를 클릭합니다 .
5. **분석법**에서 **단계별**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. 단계별 회귀의 빨간색 삼각형 메뉴를 클릭하고 **모든 가능 모형**을 선택합니다 .
8. 최대 항 수에 3 을 입력하고 최적 모형의 수에 5 를 입력합니다 .

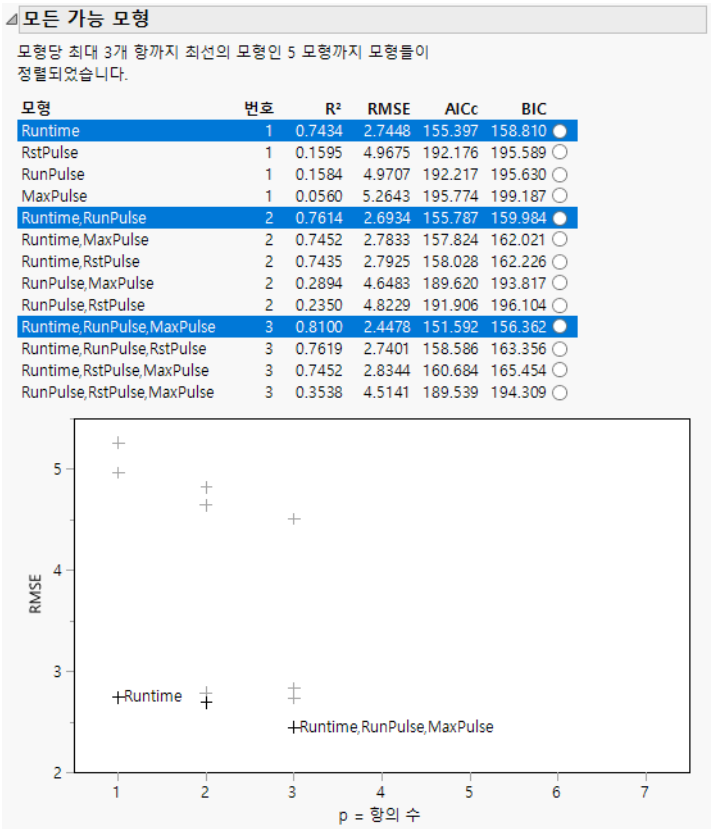
그림 5.23 모든 가능 모형 팝업 대화상자



9. **확인**을 클릭합니다 .

가능한 모든 모형(모형에 최대 3개 항 포함)이 적합됩니다.

그림 5.24 모든 가능 모형 보고서



포함된 모수의 수가 작은 것부터 순서대로 모형이 나열됩니다. 각 모수 수에 대해 R^2이 가장 높은 모형이 강조 표시됩니다. 테이블 오른쪽의 라디오 버튼 열을 사용하면 한 번에 하나의 모형을 선택하여 결과를 확인할 수 있습니다.

참고 : 모형 선택에 권장되는 기준은 BIC 또는 AICc 값이 가장 작은 모형을 선택하는 것입니다 . 일부 분석가는 C_p 통계량도 확인하려고 합니다 . Mallow C_p 통계량이 계산되지만 처음에는 테이블에 숨겨져 있습니다 . 값을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열 > Cp** 를 선택하십시오 .

모형 평균화 옵션

모형 평균화 기법을 사용하면 단일 모형을 선택하는 대신 여러 모형의 적합을 평균화할 수 있습니다 . 평균화된 각 단일 모형보다 결과 평균 모형의 예측 성능이 더 우수한 경우가 종종 있습니다 . 모형 평균화 기능은 데이터 과대적합 모형을 방지하는 데 유용합니다 . 모형에 입력할 항이 많이 선택되면 적합 시 모수 추정값이 부풀려지는 경향이 있습니다 . 모형을 평균화하면 더 약한

항에 대한 추정값이 축소되므로 우수한 예측이 생성됩니다. 모형은 다음과 같이 계산되는 각 모형의 AICc 가중치를 고려하여 평균화됩니다.

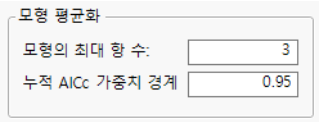
AICcWeight = exp[-0.5(AICc - AICcBest)]

AICcBest는 적합된 모형 중에서 가장 작은 AICc 값입니다. AICc 가중치 값은 각 모형에 대해 계산되어 내림차순으로 정렬되며 합이 1이 되도록 척도화됩니다. 합이 1보다 작은 척도화된 AICc 가중치 값에서 지정된 누적 AICc 가중치 경계 값을 뺀 값은 0으로 설정됩니다. 이로 인해 약한 모형은 평균화된 모형에 사용되지 않습니다. 평균화된 모형의 모수는 AICcWeight가 0이 아닌 모형에 대해 모수 추정값의 가중 평균입니다.

예 : 모형 평균화 옵션 사용

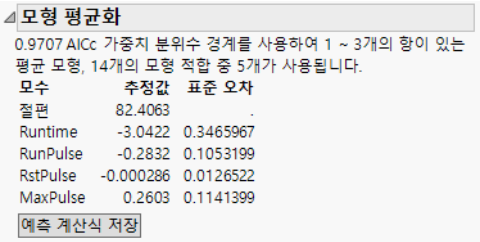
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Oxy 를 선택하고 **Y** 를 클릭합니다.
4. Runtime, RunPulse, RstPulse 및 MaxPulse 를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **단계별**을 선택합니다.
6. **실행**을 클릭합니다.
7. 단계별 회귀의 빨간색 삼각형 메뉴를 클릭하고 **모형 평균화**를 선택합니다.
8. " 모형의 최대 항 수 " 에 3 을 입력합니다.
9. " 누적 AICc 가중치 경계 " 의 0.95 를 그대로 유지합니다.

그림 5.25 모형 평균화 창



10. **확인**을 클릭합니다.

그림 5.26 모형 평균화 보고서



" 모형 평균화 " 보고서에는 각 모수에 대한 평균 추정값 및 표준 오차가 나타납니다 . 표시된 표준 오차는 추정값의 편향을 0 에 가깝게 반영합니다 .

11. **예측 계산식 저장**을 클릭하여 예측 계산식을 원래 데이터 테이블에 저장합니다 .

단계별 회귀의 검증 옵션

JMP에서 단계별 회귀 모형에 대한 교차 검증을 수행하려면 "단계별 회귀 적합"의 빨간색 삼각형을 클릭하고 **K 폴드 교차 검증**을 선택합니다. 자세한 내용은 "단계별 회귀의 K 폴드 교차 검증"(260 페이지)에서 확인하십시오.

JMP PRO JMP Pro 에서는 " 모형 적합 " 창에 검증 열을 지정할 수 있습니다 . 검증 열은 숫자 데이터 유형이어야 하고 두 개 이상의 구분되는 값을 포함해야 합니다 .

- 열에 두 개의 값이 포함된 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 열에 세 개의 값이 포함된 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 열에 네 개 이상의 구분되는 값이 포함되고 반응이 연속형인 경우 해당 값은 k 폴드 검증을 위한 폴드를 정의합니다 .

검증 열 사용에 대한 자세한 내용은 "단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합"(257 페이지)에서 확인하십시오.

단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합

JMP PRO 두 개 또는 세 개의 값이 있는 검증 열을 지정하는 경우 단계별 회귀는 훈련 데이터 집합을 기반으로 모형을 적합시킵니다. 검증 데이터 집합 및 테스트 데이터 집합에 대한 모형 적합 통계량이 보고됩니다. 이러한 통계량이 정의되는 방식에 대한 자세한 내용은 "검증 및 테스트 데이터 집합 통계량 정의"(258 페이지)에서 확인하십시오.

반응이 연속형인 경우 다음 통계량이 " 단계별 회귀 제어 " 패널에 나타납니다 .

- 검증 R^2(" 단계 기록 " 보고서에도 표시됨)
- RASE 검증
- 테스트 R^2(테스트 데이터 집합이 있는 경우)
- RASE 테스트 (테스트 데이터 집합이 있는 경우)

반응이 이항 명목형 또는 순서형인 경우 다음 통계량이 " 단계별 회귀 제어 " 패널에 나타납니다 .

- 검증 R^2(" 단계 기록 " 보고서에도 표시됨)
- 검증 평균 로그 오차
- 테스트 R^2(테스트 데이터 집합이 있는 경우)

- 테스트 평균 로그 오차 (테스트 데이터 집합이 있는 경우)

최대 검증 R^2

모형 적합 창에서 두 개 또는 세 개의 값이 있는 검증 열을 지정하는 경우 "중지 규칙"은 기본적으로 "최대 검증 R^2"으로 설정됩니다. 이 규칙은 검증 데이터 집합에 대해 R^2 통계량을 최대화하는 모형을 찾습니다. "전진 선택" 또는 "후진 제거"로 설정된 방향과 함께 이 규칙을 적용할 수 있습니다.

참고 : 최대 검증 R^2 은 p 값에 따른 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향) 로 정의된 모형만 고려합니다 . 이 규칙은 모든 가능 모형을 고려하지 않습니다 .

"단계" 버튼을 사용하여 항을 하나씩 입력하거나(전진 선택 방향) 하나씩 제거할 수 있습니다(후진 제거 방향). 언제든지 "단계 기록" 보고서의 "검증 R^2" 오른쪽에 있는 버튼을 클릭하여 모형을 선택할 수 있습니다. 선택한 모형 항은 "현재 추정값" 보고서에서 업데이트됩니다. "모형 생성" 또는 "모형 실행"을 클릭하면 이 모형이 사용됩니다.

전진 선택 방향

전진 선택 방향을 사용하는 경우 단계별 회귀는 가장 작은 다음 p 값을 기준으로 항을 추가하여 연속 모형을 생성합니다.

"단계" 대신 "시작"을 클릭하면 항 입력 프로세스가 자동으로 진행됩니다. 적합 모형 중 최적으로 간주되는 모형이 마지막에 나열됩니다. 이 모형은 검증 R^2의 국소적 감소를 무시하여 얻습니다. 특히 검증 R^2이 가장 큰 모형이며, 검증 R^2 값이 더 작은 모형이 10개까지 있을 수 있습니다. 이 모형은 "모수" 열의 "최선" 항과 "작업" 열의 "특정" 항으로 지정됩니다. "검증 R^2" 오른쪽의 버튼을 클릭하면 이 최적 모형이 선택되지만 선택을 변경해도 됩니다.

후진 제거 방향

후진 제거 방향을 사용하는 경우 단계별 회귀는 가장 큰 다음 p 값을 기준으로 항을 제거하여 연속 모형을 생성합니다.

"후진 제거" 방향을 사용하려면 먼저 "모두 입력"을 클릭하여 모든 항을 모형에 입력해야 합니다. 후진 제거 방향은 전진 선택 방향과 비슷한 방식으로 작동합니다. "단계" 대신 "시작"을 클릭하면 항 제거 프로세스가 자동으로 진행됩니다. 최적 모형은 검증 R^2이 가장 큰 모형이며, 검증 R^2 값이 더 작은 모형이 10개까지 있을 수 있습니다.

검증 및 테스트 데이터 집합 통계량 정의

이 섹션에는 검증 R^2 및 RASE 검증이 정의되어 있습니다. 테스트 R^2 및 RASE 테스트는 테스트 데이터 집합에 대해 완전히 유사한 방식으로 계산됩니다.

연속형 반응

검증 R^2 검증 데이터 집합에 대한 R^2 측도이며 다음과 같이 계산됩니다 .

- 검증 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다 . 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다 .
- 예측 오차의 제곱합을 계산하여 $SSE_{Validation}$ 을 구합니다 .
- 검증 데이터 집합의 실제 반응과 평균 간의 차이에 대한 제곱합을 계산합니다 . 이 값이 $SST_{Validation}$ 입니다 .
- 검증 R^2 은 다음과 같습니다 .

$$\text{검증 } R^2 = 1 - \frac{SSE_{Validation}}{SST_{Validation}}$$

참고 : 검증 R^2 은 음수일 수 있습니다 .

RASE 검증 검증 데이터 집합에 대한 평균 제곱 예측 오차의 제곱근입니다 . 이 값은 다음과 같이 계산됩니다 .

- 검증 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다 . 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다 .
- 예측 오차의 제곱합을 계산하여 $SSE_{Validation}$ 을 구합니다 .
- 검증 데이터 집합의 관측값 수를 $n_{Validation}$ 으로 나타냅니다 .
- RASE 검증은 다음과 같습니다 .

$$\text{RASE 검증} = \sqrt{\frac{SSE_{Validation}}{n_{Validation}}}$$

이항 명목형 또는 순서형 반응

검증 R^2 검증 데이터 집합에 대한 엔트로피 R^2 측도 (McFadden R^2 이라고도 함) 이며 다음과 같이 계산됩니다 .

- 훈련 데이터 집합을 사용하여 모형이 적합됩니다 .
- 모든 관측값에 대한 예측 확률을 구합니다 .
- 훈련 데이터 집합 모형에 기반한 예측 확률을 사용하여 검증 데이터 집합의 관측값에 대해 모형의 가능도가 계산됩니다 . 이 통계량을 $Likelihood_Full_{Validation}$ 이라고 합니다 .
- 검증 데이터 집합의 데이터를 사용하여 축소 모형 (예측 변수 없음) 의 가능도가 계산됩니다 . 이 통계량을 $Likelihood_Reduced_{Validation}$ 이라고 합니다 .
- 검증 R^2 은 다음과 같습니다 .

$$\text{검증 } R^2 = 1 - \frac{\log(Likelihood_Full_{Validation})}{\log(Likelihood_Reduced_{Validation})}$$

참고 : 검증 R^2 은 음수일 수 있습니다 .

검증 평균 로그 오차 검증 데이터 집합의 평균 로그 오차는 다음과 같이 계산됩니다 .

- 검증 데이터 집합의 각 관측값에 대해 훈련 데이터 집합에 기반한 모형에 의해 결정된 예측 확률의 로그를 계산합니다 .
- 이 로그의 합을 검증 데이터 집합의 관측값 수로 나눈 후 결과 값의 음수를 구합니다 .

팁 : 검증 평균 로그 오차 값이 작을수록 좋습니다 .

단계별 회귀의 K 폴드 교차 검증

K 폴드 교차 검증은 데이터를 무작위로 *k* 개의 부분집합으로 나눕니다. 결과적으로 *k* 개 집합이 각각 검증 데이터 집합으로 사용되고 나머지 데이터는 모형 적합을 위한 훈련 데이터 집합으로 사용됩니다. 총 *k* 개 모형이 적합되고 *k* 개의 검증 통계량이 계산됩니다. 최종 모형에서는 최량 검증 통계량을 제공하는 모형이 선택됩니다. 이 방법은 제한된 양의 데이터를 효율적으로 사용하기 때문에 소규모 데이터 집합에 유용합니다.

참고 : K 폴드 교차 검증은 연속형 반응에만 사용할 수 있습니다 .

JMP에서 "단계별 회귀 적합"의 빨간색 삼각형을 클릭하고 **K 폴드 교차 검증**을 선택합니다.

JMP Pro 에서는 다음 두 가지 방법으로 *k* 폴드 교차 검증에 액세스할 수 있습니다 .

- " 단계별 회귀 적합 " 의 빨간색 삼각형을 클릭하고 **K 폴드 교차 검증**을 선택합니다 .
- 네 개 이상의 구분되는 값이 포함된 검증 열을 지정합니다 .

R^2 K 폴드 통계량

K 폴드 교차 검증을 수행하면 "단계별 회귀 제어" 패널의 다른 통계량 오른쪽에 "R^2 K 폴드" 통계량이 나타납니다. R^2 K 폴드는 *k* 폴드에 대한 검증 R^2 값의 평균입니다.

최대 K 폴드 R^2

k 폴드 교차 검증을 사용하는 경우 "중지 규칙"은 기본적으로 "최대 K 폴드 R^2"으로 설정됩니다. 이 규칙은 R^2 K 폴드 통계량을 최대화합니다.

참고 : 최대 K 폴드 R^2 은 *p* 값에 따른 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향) 로 정의된 모형만 고려합니다 . 이 규칙은 모든 가능 모형을 고려하지 않습니다 .

최대 K 폴드 R^2 중지 규칙은 최대 검증 R^2 중지 규칙과 비슷한 방식으로 작동합니다. 자세한 내용은 "최대 검증 R^2"(258 페이지)에서 확인하십시오. "R^2 검증"을 "R^2 K 폴드"로 바꾸어 참조하면 됩니다.

6 장

JMP PRO 일반화 회귀 모형
변수 선택 기법을 사용하여 모형 생성

모형 적합 플랫폼의 일반화 회귀 분석법은 JMP Pro 에서만 사용할 수 있습니다 .

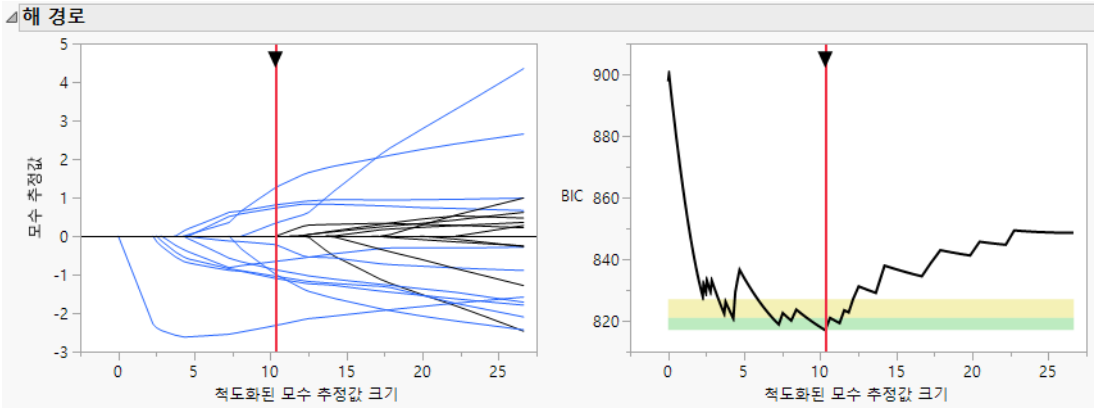
JMP Pro에서 모형 적합 플랫폼의 일반화 회귀 분석법은 축소 방법을 포함한 다양한 변수 선택 기법을 제공하며 특히 상관관계가 있는 고차원 데이터 모델링을 처리합니다. 이 중에서 Lasso 회귀와 Elastic Net이라는 두 가지 기법은 모델링 절차의 일부로 변수 선택을 수행합니다.

많은 변수를 포함하는 대규모 데이터 집합은 대개 다중공선성 문제를 나타냅니다. 최신 데이터 집합에는 관측값보다 변수가 더 많이 포함될 수 있으므로 기존 모델링 기법을 사용하려면 변수 선택이 필요합니다. 다중공선성 및 많은 예측 변수가 있으면 전통적 설계 기법의 단점이 드러납니다.

설계된 실험을 포함하여 상관관계가 거의 없는 소규모 데이터 집합에서도 Lasso 회귀와 Elastic Net이 유용합니다. 예측 모형을 생성하거나, 모형 축소 또는 향후 연구를 위해 변수를 선택할 때 이 기법을 사용할 수 있습니다.

일반화 회귀 분석법은 많은 모델링 상황에 유용합니다. 이 분석법을 사용하면 반응 변수에 대한 다양한 분포를 지정할 수 있습니다. 반응이 연속형, 이항, 개수 또는 영과잉일 때 사용합니다. 또한 변수 선택에 관심이 있거나 예측 변수의 공선성이 의심될 때 사용합니다. 보다 일반적으로는 다른 기법을 사용하여 생성한 모형과 비교하는 모형을 적합시킬 때 사용합니다.

그림 6.1 Elastic Net 적합에 대한 해 경로



목차

일반화 회귀 분석법 개요 265
일반화 회귀의 예 266
일반화 회귀 분석법 시작 269
분포 270
일반화 회귀 보고서 창 277
일반화 회귀 보고서 옵션 278
모형 시작 제어판 278
반응 분포 279
추정 방법 옵션 279
고급 제어 283
검증 방법 옵션 286
조기 중지 287
시작 287
모형 적합 보고서 287
회귀 그림 288
모형 요약 288
추정 상세 정보 291
해 경로 291
중심화 및 척도화된 예측 변수에 대한 모수 추정값 294
원래 예측 변수에 대한 모수 추정값 295
활성 모수 추정값 296
효과 검정 297
모형 적합 옵션 297
일반화 회귀 분석법에 대한 통계 상세 정보 305
추정 방법에 대한 통계 상세 정보 306
고급 제어에 대한 통계 상세 정보 308
분포에 대한 통계 상세 정보 309

JMP PRO 일반화 회귀 분석법 개요

일반화 회귀 분석법은 정규화(벌점) 회귀 기법을 특징으로 합니다. 이 기법은 모형 계수를 0에 가까워지도록 축소하여 더 나은 모형을 적합시키려고 합니다. 결과 추정값은 편향되어 있습니다. 편향이 증가하면 예측 분산이 감소되어 벌점 없는 모형에 비해 전체 예측 오차를 낮출 수 있습니다. 이 중에서 Elastic Net과 Lasso 회귀라는 두 가지 기법은 모델링 절차의 일부로 변수 선택을 포함합니다.

Elastic Net 및 Lasso 회귀와 같은 모델링 기법은 일반적으로 공선성이 문제가 되는 대규모 데이터 집합에 특히 유용합니다. 또한 최신 데이터 집합에는 관측값보다 변수가 더 많이 포함될 때가 종종 있습니다. 이 상황을 $p > n$ 문제라고도 하며, 여기서 n은 관측값 수이고 p는 예측 변수의 수입니다. 기존 모델링 기법을 사용할 경우 이러한 데이터 집합에 대해 변수 선택이 필요합니다.

설계된 실험을 포함하여 상관관계가 거의 없는 소규모 데이터 집합에도 Elastic Net과 Lasso 회귀를 사용할 수 있습니다. 예측 모형을 생성하거나, 모형 축소 또는 향후 연구를 위해 변수를 선택할 때 이 기법을 사용할 수 있습니다.

이 분석법은 다음과 같은 종류의 모델링 기법을 제공합니다.

- 최대 가능도
- 단계 기반 추정
- 벌점 회귀 방법

Elastic Net과 Lasso 회귀는 비교적 최근의 기법입니다(Tibshirani 1996, Zou and Hastie 2005). 두 기법 모두 모형 계수의 크기에 벌점을 적용하여 지속적 축소가 발생합니다. 축소 크기는 **조정 모수**에 의해 결정됩니다. 최적의 축소 수준은 여러 검증 방법 중 하나에 의해 결정됩니다. 두 기법 모두 계수를 0으로 축소할 수 있습니다. 이러한 방식으로 변수 선택이 모델링 절차에 기본 제공됩니다. Elastic Net 모형은 Lasso 회귀와 능형 회귀 둘 다 특수 사례로 포함합니다. 자세한 내용은 "추정 방법에 대한 통계 상세 정보"(306 페이지)에서 확인하십시오.

일반화 회귀 모델링 기법에 대한 상세 정보

- 최대 가능도 방법은 전통적 설계 방식입니다. 이 방법은 다른 기법과 비교할 수 있는 기준을 제공하며 가설 검정과 같은 기존 추론 기법에 대해 가장 적합합니다.
- 전진 선택은 단계별 회귀 방법입니다. 전진 선택에서는 항이 모형에 입력됩니다. 모든 항이 모형에 포함되거나 자유도가 남아 있지 않을 때까지 유의성이 가장 높은 항이 추가됩니다.
- Lasso 회귀에는 두 가지 단점이 있습니다. 상관관계가 높은 여러 변수가 있을 때 해당 그룹에서 하나의 변수만 선택하려는 경향이 있습니다. 변수 수 (p)가 관측값 수 (n)를 초과할 경우 Lasso 회귀는 최대 n개의 예측 변수를 선택합니다.
- 반면, Elastic Net은 상관관계가 있는 그룹에서 모든 변수를 선택하여 적절한 계수를 적합시키는 경향이 있습니다. 또한 $p > n$일 때 n개를 초과하는 예측 변수를 선택할 수 있습니다.
- 능형 회귀는 첫 번째로 제안된 벌점 회귀 방법입니다 (Hoerl 1962, Hoerl and Kennard 1970). 능형 회귀는 계수를 0으로 축소하지 않으므로 변수 선택을 수행하지 않습니다.

- 이중 Lasso 회귀는 초기 Lasso 회귀 모형에서 변수 선택을 수행하여 선택 단계와 축소 단계를 구분하려고 합니다 . 그런 다음 초기 모형에서 선택한 변수가 두 번째 Lasso 회귀 모형의 입력 변수로 사용됩니다 .
- 2 단계 전진 선택은 두 단계의 단계별 전진 회귀를 수행합니다 . 첫 번째 단계에서 주효과에 대한 변수 선택을 수행합니다 . 그런 다음 두 번째 단계에서 고차 효과를 모형에 입력할 수 있습니다 .

일반화 회귀 분석법은 Lasso 회귀 및 Elastic Net 의 **적응형** 버전에도 적합합니다. 이러한 적응형 버전은 실제 활성 집합에 포함되지 않은 변수보다 실제 활성 집합의 변수에 벌점을 적게 줍니다. **실제 활성 집합**은 모형에서 반응에 실제 영향을 미치는 항 집합을 나타냅니다. 적응형 Lasso 회귀 및 Elastic Net 은 oracle 특성을 유지하기 위해 개발되었습니다. oracle 특성이 유지됨으로써 추정값은 점진적으로 예측 변수의 실제 활성 집합에 모형을 적합시켰을 때 얻을 수 있는 값이 됩니다. 더 구체적으로 말하면 모형이 계수가 0 이어야 하는 예측 변수를 올바르게 식별할 수 있습니다. 추정값은 실제 활성 집합만 포함하여 시작했을 때 얻을 수 있는 값에 수렴합니다. 자세한 내용은 " 적응형 방법 "(307 페이지) 에서 확인하십시오 .

일반화 회귀 분석법을 사용하면 반응 변수에 대한 다양한 분포를 지정할 수 있습니다. 분포 적합에는 정규, Cauchy, 스튜던트 *t*, 지수, 감마, Weibull, 로그 정규, 베타, 이항, 베타 이항, Poisson, 음이항, 영과잉 이항, 영과잉 베타 이항, 영과잉 Poisson, 영과잉 음이항 및 영과잉 감마가 포함됩니다. 이러한 유연성을 통해 범주형 반응과 개수 반응, 연속형 반응 및 오른쪽으로 치우친 연속형 반응을 적합시킬 수 있습니다. 또한 분위수 회귀 모형과 Cox 비례 위험 모형도 적합시킬 수 있습니다. 일부 분포의 경우 중도절단 데이터에 모형을 적합시킬 수 있습니다. 이 분석법은 모형 선택을 위한 다양한 검증 기준을 제공하고 훈련, 검증 및 테스트 열을 지원합니다. 자세한 내용은 " 분포 "(270 페이지) 에서 확인하십시오 .

JMP PRO 일반화 회귀의 예

Diabetes.jmp 샘플 데이터 테이블의 데이터는 442 명의 당뇨병 환자에 대한 측정값으로 구성됩니다 . 관심 반응은 Y 이며 기준 측정을 수행하고 1 년 후 측정된 질병 진행률입니다 . 질병 진행과 관련이 있다고 간주되는 10 가지 변수도 기준에서 측정됩니다 . 이 예에서는 일반화 회귀 기법을 사용하여 예측 모형을 개발하는 방법을 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Y 를 선택하고 **Y** 를 클릭합니다 .
4. Age ~ Glucose 를 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .

   이렇게 하면 차수 2(**차수** 상자의 기본값) 까지의 모든 항이 모형에 추가됩니다 .
5. " 열 선택 " 목록에서 Validation 을 선택하고 **검증**을 클릭합니다 .
6. " 분석법 " 목록에서 **일반화 회귀**를 선택합니다 .
7. **실행**을 클릭합니다 .

" 모형 비교 " 보고서 , " 모형 시작 " 제어판 및 " 검증 열을 사용한 정규 표준 최소 제곱 " 보고서가 포함된 " 일반화 회귀 " 보고서가 나타납니다 .

" 모형 시작 " 제어판에서 다음 사항에 유의하십시오 .

– " 모형 적합 " 시작 창에서 " 분포 " 에 " 정규 " 를 지정했으므로 " 반응 분포 " 가 " 정규 " 로 설정됩니다 .

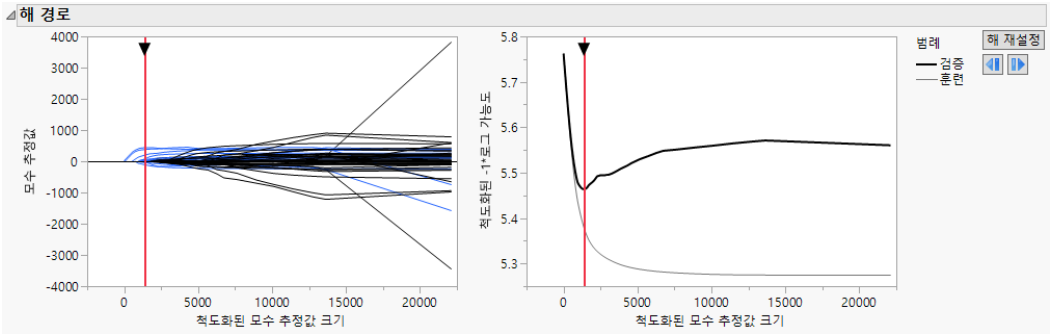
– 기본 추정 방법은 Lasso 회귀입니다 .

– " 모형 적합 " 창에서 검증 열을 지정했으므로 " 검증 방법 " 이 " 검증 열 " 로 설정됩니다 .

8. **시작**을 클릭합니다 .

" 검증 열을 사용한 정규 Lasso 회귀 " 보고서가 나타납니다 . " 해 경로 " 보고서 (그림 6.2) 에서는 모수 추정값 및 척도화된 -1* 로그 가능도 그림을 보여 줍니다 . 척도화된 모수 추정값의 크기가 감소하면 축소량이 증가합니다 . 그림 맨 오른쪽의 추정값은 최대 가능도 추정값입니다 . 빨간색 수직선은 검증 기준 (이 예에서는 Validation 열에 정의된 홀드백 표본) 에 의해 선택된 모수 값을 나타냅니다 .

그림 6.2 해 경로 그림



9. " 검증 열을 사용한 정규 Lasso 회귀 " 옆의 빨간색 삼각형을 클릭하고 **0 이 아닌 항 선택**을 선택합니다 .

이 옵션은 " 원래 예측 변수에 대한 모수 추정값 " 보고서 (그림 6.3) 에서 0 이 아닌 항을 강조 표시하고 해 경로 그림에서 해당 경로를 강조 표시합니다 . 데이터 테이블의 해당 열도 선택됩니다 . 55 개의 모수 추정값 중 11 개만 0 이 아닙니다 . 정규 분포에 대한 척도 모수 (시그마) 도 추정되어 " 원래 예측 변수에 대한 모수 추정값 " 보고서 하단에 있는 별도의 테이블에 표시됩니다 . 55 개의 모수 추정값 중 일부만 그림 6.3 에 나와 있습니다 .

그림 6.3 원래 예측 변수에 대한 모수 추정값 보고서의 일부

원래 예측 변수에 대한 모수 추정값

항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	-247.8422	42.926975	33.334183	<.0001*	-331.9775	-163.7068
Age	0	0	0	1.0000	0	0
Gender[1-2]	8.7286488	6.9678967	1.5692446	0.2103	-4.928178	22.385475
BMI	5.702121	0.9303784	37.562422	<.0001*	3.8786128	7.5256293
BP	0.8156913	0.2526706	10.421787	0.0012*	0.320466	1.3109166
Total Cholesterol	0	0	0	1.0000	0	0
LDL	0	0	0	1.0000	0	0
HDL	-0.47425	0.3986141	1.4155011	0.2341	-1.25552	0.3070188
TCH	0	0	0	1.0000	0	0
LTG	41.286654	7.0844177	33.963392	<.0001*	27.401451	55.171858
Glucose	0	0	0	1.0000	0	0
(Age-48.5181)*Gender[1-2]	-0.082735	0.3621512	0.0521915	0.8193	-0.792538	0.6270682
(Age-48.5181)*(BMI-26.3758)	0	0	0	1.0000	0	0
(Age-48.5181)*(BP-94.647)	0.0043544	0.0171735	0.0642885	0.7998	-0.029305	0.0380137
(Age-48.5181)*(Total Cholesterol-189.14)	0	0	0	1.0000	0	0
(Age-48.5181)*(LDL-115.439)	0	0	0	1.0000	0	0
(Age-48.5181)*(HDL-49.7885)	0	0	0	1.0000	0	0
(Age-48.5181)*(TCH-4.07025)	0	0	0	1.0000	0	0
(Age-48.5181)*(LTG-4.64141)	0	0	0	1.0000	0	0
(Age-48.5181)*(Glucose-91.2602)	0.0231484	0.0233144	0.9858084	0.3208	-0.022547	0.0688439
Gender[1-2]*(BMI-26.3758)	0	0	0	1.0000	0	0
Gender[1-2]*(BP-94.647)	0	0	0	1.0000	0	0
Gender[1-2]*(Total Cholesterol-189.14)	0	0	0	1.0000	0	0
Gender[1-2]*(LDL-115.439)	0	0	0	1.0000	0	0
Gender[1-2]*(HDL-49.7885)	-0.221009	0.4728172	0.2184913	0.6402	-1.147714	0.7056955
Gender[1-2]*(TCH-4.07025)	0	0	0	1.0000	0	0
Gender[1-2]*(LTG-4.64141)	0	0	0	1.0000	0	0
Gender[1-2]*(Glucose-91.2602)	0	0	0	1.0000	0	0
(BMI-26.3758)*(BP-94.647)	0.0254032	0.0525392	0.2337819	0.6287	-0.077572	0.1283782
(BMI-26.3758)*(Total Cholesterol-189.14)	0	0	0	1.0000	0	0
(BMI-26.3758)*(LDL-115.439)	0	0	0	1.0000	0	0
(BMI-26.3758)*(HDL-49.7885)	0	0	0	1.0000	0	0
(BMI-26.3758)*(TCH-4.07025)	0	0	0	1.0000	0	0
(BMI-26.3758)*(LTG-4.64141)	0	0	0	1.0000	0	0
(BMI-26.3758)*(Glucose-91.2602)	0.0210214	0.0597655	0.1237153	0.7250	-0.096117	0.1381596

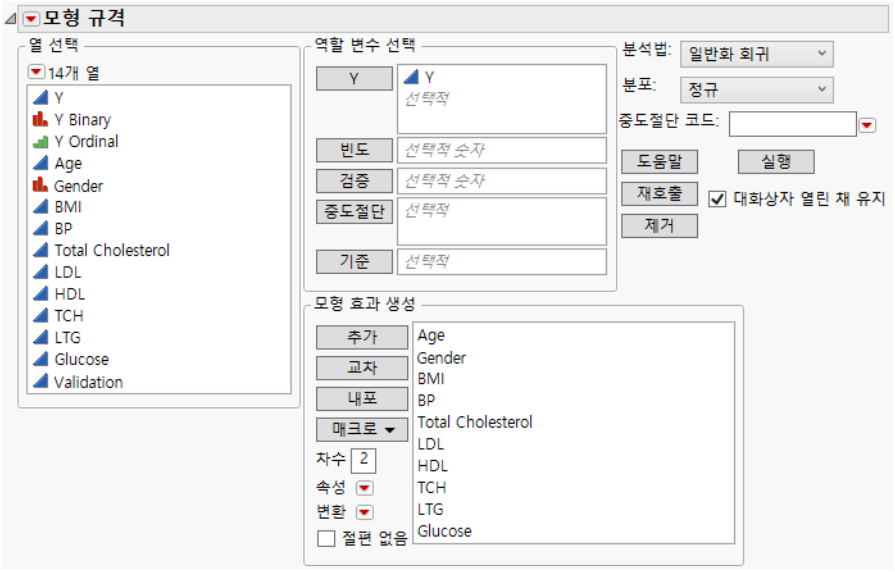
정규 분포 모수	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
척도	52.512586	2.3230199	511.00007	<.0001*	47.959551	57.065621

예측 계산식을 저장하려면 " 검증 열을 사용한 정규 Lasso 회귀 " 보고서 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다 .

JMP PRO 일반화 회귀 분석법 시작

분석 > 모형 적합을 선택하고 **Y** 에 대해 하나 이상의 열을 입력한 후 **분석법** 메뉴에서 **일반화 회귀**를 선택하여 일반화 회귀 분석법을 시작합니다 .

그림 6.4 일반화 회귀가 선택된 모형 적합 시작 창



"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 "모형 규격" 장(29 페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 일반화 회귀 분석법에 대한 정보를 제공합니다.

일반화 회귀 분석법에 사용되는 명목형 변수의 파라미터화는 다른 모형 적합 분석법을 사용한 파라미터화와 다릅니다. 일반화 회귀 분석법은 표시 함수 파라미터화를 사용합니다. 이 파라미터화 방법에서 명목형 변수 수준의 표시자에 해당하는 추정값은 같은 수준의 평균 반응과 마지막 수준의 평균 반응 간 차이에 대한 추정값입니다. 마지막 수준은 값 순서 코딩이 가장 높은 수준이며 표시 함수가 모형에 포함되지 않은 수준입니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격"의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

절편 항이 없는 모형을 지정하려면 " 모형 적합 " 창의 " 모형 효과 생성 " 패널에서 " 절편 없음 " 옵션을 선택합니다 . 이 옵션을 선택하는 경우 다음 사항에 유의하십시오 .

- 예측 변수가 중심화되거나 척도화되지 않습니다 .
- 보고서 창에서 승산비 , 위험 비율 및 발생률 비율을 사용할 수 없습니다 .
- 순서형 로지스틱 분포에서는 " 절편 없음 " 옵션을 사용할 수 없습니다 .

주의 : 결과는 모형 효과의 척도에 민감하므로 Lasso 회귀 또는 Elastic Net 에서 " 절편 없음 " 옵션을 사용하는 것은 권장되지 않습니다 . 대신 이 추정 방법의 적응형 버전을 사용하는 것이 좋습니다 .

중도절단

다음 방법 중 하나를 사용하여 반응 변수에 대한 중도절단을 지정할 수 있습니다 .

- 구간 중도절단 , 오른쪽 중도절단 및 왼쪽 중도절단 반응의 경우 " 감지 한계 " 열 특성을 포함하는 열을 지정합니다 . 열 특성에 지정된 한계는 중도절단되지 않은 것으로 간주되는 반응 값 범위를 정의합니다 .
  - 구간 중도절단 반응의 경우 " 감지 한계 " 열 특성에서 " 감지 하한 " 과 " 감지 상한 " 둘 다에 비결측값을 지정합니다 .
  - 오른쪽 중도절단 반응의 경우 " 감지 한계 " 열 특성에서 " 감지 하한 " 에는 결측값을 지정하고 " 감지 상한 " 에는 비결측값을 지정합니다 .
  - 왼쪽 중도절단 반응의 경우 " 감지 한계 " 열 특성에서 " 감지 하한 " 에는 비결측값을 지정하고 " 감지 상한 " 에는 결측값을 지정합니다 .
- 오른쪽 중도절단 반응의 경우 시작 창에서 오른쪽 중도절단된 관측값에 대한 표시자가 포함된 열을 " 중도절단 " 열로 지정합니다 . 그런 다음 " 중도절단 코드 " 목록에서 해당 열의 값 중 오른쪽 중도절단된 관측값을 지정하는 값을 선택합니다 .
- 구간 중도절단 및 왼쪽 중도절단 반응의 경우 "Y" 열 역할에 중도절단 구간을 정의하는 두 열을 지정합니다 .
  - 구간 중도절단 반응의 경우 첫 번째 Y 변수는 하한을 제공합니다 .
  - 왼쪽 중도절단 반응의 경우 첫 번째 Y 변수는 결측값을 포함합니다 .
  - 구간 중도절단 반응과 왼쪽 중도절단 반응 모두 두 번째 Y 변수는 각 반응의 상한을 제공합니다 .

  Y 및 중도절단을 지원하는 분포에 대한 두 열을 지정하면 열이 중도절단을 나타내는지 여부를 묻는 경고가 표시됩니다 . "아니요"를 선택할 경우 해당 열은 별도의 반응으로 처리됩니다 .

  참고 : 일반화 회귀 플랫폼 환경 설정의 " 두 개의 반응 열에 대한 처리 " 설정을 사용하여 두 반응에 대한 기본 동작을 지정할 수 있습니다 .

정규, 지수, 감마, Weibull, 로그 정규 또는 Cox 비례 위험 분포가 지정된 경우 중도절단을 사용할 수 있습니다 .

JMP PRO 분포

분석법 메뉴에서 **일반화 회귀**를 선택하면 **분포** 옵션이 나타납니다 . 여기서 Y 의 분포를 지정할 수 있습니다 . 약어 ZI는 영과잉(Zero-Inflated)을 의미합니다 . 분포는 반응에 따라 연속형 , 이산형 및 영과잉의 세 가지 범주로 구분됩니다 . 아래에 옵션에 대한 설명이 나옵니다 .

참고 : " 모형 규격 " 창에서 Y 변수를 여러 개 지정하는 경우 지정된 모든 Y 변수에 대해 동일한 반응 분포를 사용해야 합니다 . 동일한 " 일반화 회귀 " 보고서에서 각 반응 변수마다 개별 분포를 적합시키려면 스크립트를 사용해야 합니다 .

JMP PRO 연속형

정규 Y 는 평균이 μ 이고 표준편차가 σ 인 정규 분포를 가집니다 . 정규 분포는 대칭적이며 , 표본 크기가 충분히 크면 중심 극한 정리에 의해 다양한 다른 분포가 정규 분포와 가까워질 수 있습니다 . μ 의 연결 함수는 항등입니다 . 즉 , Y 의 평균은 선형 모형으로 표현됩니다 .

참고 : 정규 분포가 지정된 경우 " 표준 최소 제곱 " 이 " 최대 가능도 " 추정 방법을 대체합니다 .

정규 분포의 척도 모수는 σ 입니다 . 추정 방법에 벌점이 없는 경우 척도 모수 σ 의 추정값은 RMSE(제곱근 평균 제곱 오차) 입니다 . RMSE 는 σ^2 의 일반적인 비편향 추정량에 대한 제곱근입니다 . 표시된 결과는 중도절단된 관측값이 포함된 경우가 아니면 표준 최소 제곱 적합과 동등합니다 .

참고 : 일반화 회귀 분석법에 사용되는 명목형 변수의 파라미터화는 표준 최소 제곱 분석법을 사용한 파라미터화와 다릅니다 . 이 차이 때문에 명목형 또는 순서형 효과를 포함하는 모형에 대한 모수 추정값이 달라집니다 .

자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

Cauchy Y 는 위치 모수가 μ 이고 척도 모수가 σ 인 Cauchy 분포를 가집니다 . Cauchy 분포는 평균과 표준편차가 정의되어 있지 않습니다 . 중앙값과 최빈값은 모두 μ 입니다 . 대부분의 데이터는 본질적으로 Cauchy 분포를 따르지 않습니다 . 그러나 이상치 비율이 큰 (최대 50%) 데이터에 대해 로버스트 회귀를 수행하는 데 유용합니다 . μ 의 연결 함수는 항등입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

t(5) Y 는 자유도가 5 이고 위치 모수가 μ, 척도 모수가 σ 인 스튜던트 *t* 분포를 가집니다 . 스튜던트 *t* 분포는 대칭적이며 , 정규 분포와 Cauchy 분포 사이의 범위를 포괄하는 로버스트 옵션입니다 . 스튜던트 *t* 분포의 자유도가 무한대에 가까워지면 정규 분포와 동등합니다 . 스튜던트 *t* 분포의 자유도가 1 이면 Cauchy 분포와 동등합니다 . μ 의 연결 함수는 항등입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

지수 Y 는 평균 모수가 μ 인 지수 분포를 가집니다 . 지수 분포는 오른쪽으로 치우친 분포로 , 대개 수명이나 연속된 이벤트 사이의 시간을 모델링하는 데 사용됩니다 . μ 의 연결 함수는 로그입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

감마 Y 는 평균 모수가 μ 이고 산포 모수가 σ 인 감마 분포를 가집니다 . 감마는 널리 사용되는 다른 분포 계열을 포함하는 유연한 분포입니다 . 예를 들어 지수 분포는 감마 분포의 특수한 경우로 , σ = μ 입니다 . 카이제곱 분포도 감마 분포에서 파생될 수 있습니다 . μ 의 연결 함수는 로그입니다 . 자세한 내용은 "분포에 대한 통계 상세 정보"(309페이지)에서 확인하십시오 .

Weibull Y 는 평균 모수가 μ 이고 척도 모수가 σ 인 Weibull 분포를 가집니다 . Weibull 분포는 유연한 분포로 , 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다 . μ 의 연결 함

수는 항등입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

로그 정규 Y 는 평균 모수가 μ 이고 척도 모수가 σ 인 로그 정규 분포를 가집니다 . 로그 정규 분포는 오른쪽으로 치우친 분포로 , 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다 . μ 의 연결 함수는 항등입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

베타 Y 는 평균 모수가 μ 이고 산포 모수가 σ 인 베타 분포를 가집니다 . 베타 분포에 대한 반응은 0 에서 1 사이이며 대개 비율을 모델링하는 데 사용됩니다 . μ 의 연결 함수는 로짓입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

분위수 회귀 분위수 회귀는 반응에 대한 지정된 조건부 분위수를 모델링합니다 . 기본 분포의 형태에 대해서는 가정하지 않습니다 . 분위수 회귀를 선택하면 " 분포 " 메뉴 아래에 " 분위수 " 상자가 나타납니다 . 원하는 분위수를 지정합니다 .

" 모형 대화상자 " 창에서 " 분위수 " 에 0.5(기본값) 를 지정할 경우 분위수 회귀는 반응의 조건부 중앙값을 모델링합니다. 분위수 회귀는 회귀 계수로 표현되는 조건부 분위수의 변화 비율이 분위수에 따라 달라지는 경우에 특히 유용합니다 . 최소 제곱 회귀와 비교할 때 분위수 회귀의 장점은 이질적 조건부 분포를 사용하여 데이터를 모델링할 수 있는 유연성입니다 .

분위수 회귀는 반복 방식을 통해 목적 함수를 최소화하여 적합됩니다 . 분위수 회귀에 대한 자세한 내용은 Koenker and Hallock 연구 자료 (2001) 및 Portnoy and Koenker 연구 자료 (1997) 에서 확인하십시오 .

분위수 회귀를 선택할 경우 " 추정 방법 " 에서 " 최대 가능도 " 만 사용할 수 있으며 " 검증 방법 " 에서는 " 없음 " 만 사용할 수 있습니다 .

참고 : 분위수 회귀 적합에 시간이 많이 소요되는 경우 진행률 표시줄이 나타납니다 . 진행률 표시줄에는 목적 함수의 상대적 변동이 표시됩니다 . " 현재 추정값 채택 " 을 클릭하면 계산이 중지되며 보고된 모수 추정값은 해당 지점에서 최적 모형 적합에 해당합니다 .

Cox 비례 위험 Cox 비례 위험 모형은 예측 변수가 있는 사건 발생 시간 데이터에 대한 회귀 모형입니다 . 이 모형은 예측 변수와 위험 함수 간의 승법 관계를 기반으로 합니다 . 예측 변수가 생존 시간에 미치는 영향을 조사하는 데 이 모형을 사용할 수 있습니다 . 모형에는 일반적인 위험 함수를 제공하기 위해 예측 변수로 척도화된 임의의 기준 위험 함수가 포함됩니다 . 비례 위험 모형은 각 예측 변수에 대한 모수 추정값과 표준 오차를 생성합니다 . Cox 비례 위험 모형은 D. R. Cox(1972) 에 의해 처음 제안되었습니다 . 비례 위험 모형에 대한 자세한 내용은 Kalbfleisch and Prentice 연구 자료 (2002) 에서 확인하십시오 .

Cox 비례 위험을 선택할 경우 " 검증 방법 " 에서 "BIC" 와 "AICc" 만 사용할 수 있습니다 . 또한 능형 회귀 추정 방법을 사용할 수 없습니다 .

참고 : 반응에 동일 값이 있으면 Efron 가능도가 사용됩니다 . 자세한 내용은 Efron 연구 자료 (1981) 에서 확인하십시오 . 이 방법은 모형 적합 플랫폼의 비례 위험 분석법 또는 비례 위험 모형 적합 플랫폼에서 사용되는 동일 값 처리 방법과 다릅니다 .

JMP PRO 이산형

이항 Y 는 p 와 n 모수가 있는 이항 분포를 가집니다 . 반응 Y 는 n 번의 독립 시행 (모든 시행에 대한 확률 p 는 고정됨) 에서 총 성공 횟수를 나타냅니다 . 이 분포에서는 표본 크기 열을 사용할 수 있습니다 . 나열된 열이 없으면 표본 크기가 1 이라고 가정합니다 . p 의 연결 함수는 로짓입니다 . 명목형 모델링 유형이 있는 이진 반응 변수를 선택하면 이항 반응 분포만 사용할 수 있습니다 . 자세한 내용은 "분포에 대한 통계 상세 정보"(309페이지)에서 확인하십시오.

"분포"에서 "이항"을 선택하면 다음 방법 중 하나를 사용하여 반응 변수를 지정해야 합니다.

- 요약되지 않음 : 데이터가 사건 빈도로 요약되지 않은 경우 단일 이진 열을 반응으로 지정합니다 . 이 열의 모델링 유형이 명목형이면 수준 중 하나를 목표 수준으로 지정할 수 있습니다 . 기본 목표 수준 값은 수준 순서에 따라 두 수준 중 더 높은 값입니다 .
- 빈도 열로 요약됨 : 데이터가 성공 및 실패 빈도로 요약되는 경우 단일 이진 열을 반응으로 지정합니다 . 이 열의 모델링 유형이 명목형이면 수준 중 하나를 목표 수준으로 지정할 수 있습니다 . 기본 목표 수준 값은 수준 순서에 따라 두 수준 중 더 높은 값입니다 . 빈도 열을 **빈도** 역할에 할당합니다 .
- 두 번째 Y 로 입력된 표본 크기 열로 요약됨 : 데이터가 사건 (성공) 빈도 및 시행 횟수로 요약되는 경우 , 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 Y 로 지정합니다 .

참고 : 이항 분포가 지정된 경우 " 로지스틱 회귀 " 가 " 최대 가능도 " 추정 방법을 대체합니다 .

베타 이항 Y 는 성공 확률 p, 시행 횟수 n 및 과대산포 모수 δ 가 있는 베타 이항 분포를 가집니다 . 이 분포는 이항 분포의 과대산포 버전입니다 .

산포 모수 δ 를 갖는 베타 이항 분포를 모수 p 및 $n = 20$ 을 갖는 이항 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoBetaBinomial.jsl 을 실행하십시오 .

베타 이항 분포를 사용하려면 각 관측값에 대해 표본 크기가 1 보다 커야 합니다 . 따라서 사용자는 표본 크기 열을 지정해야 합니다 . 표본 크기 열을 삽입하려면 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 Y 로 지정합니다 . p 의 연결 함수는 로짓입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

다항 Y 는 이산형 수준이 세 개 이상인 다항 분포를 가집니다 . 반응 변수의 모델링 유형은 명목형 또는 순서형이어야 합니다 . 이 모형은 반응 변수의 각 수준에 대해 별도의 절편 및 효과 모수를 적합시킵니다 . 반응 변수의 수준이 k 개이면 모형에 $k - 1$ 개의 절편 및 효과 모수가 포함됩니다 . 다항 분포의 연결 함수는 다항 로짓입니다 . 자세한 내용은 " 통계 상세 정보 " 장의 " 명목형 반응 "(489 페이지) 에서 확인하십시오 .

순서형 로지스틱 Y 는 순서형 수준이 있는 다항 분포를 가집니다 . 반응 변수의 모델링 유형은 순서형이어야 합니다 . 이 모형은 반응 변수의 각 수준에 대해 절편을 적합시킵니다 . 효과 모수는 반응 변수의 모든 수준에서 공통입니다 . 순서형 로지스틱 분포의 연결 함수는 순서형 로짓입니다 . 자세한 내용은 " 통계 상세 정보 " 장의 " 순서형 반응 "(491 페이지) 에서 확인하십시오 .

참고 : 일반화 회귀에서 순서형 로지스틱에 대한 절편 파라미터화는 모형 적합의 순서형 로지스틱 분석법에서와 다릅니다. 일반화 회귀의 첫 번째 절편 항은 순서형 로지스틱 분석법의 첫 번째 절편 항에 해당합니다. 일반화 회귀의 후속 절편 항은 반응 변수의 순서형 수준에 대한 절편 항 간의 연속 차이입니다.

Poisson Y 는 평균이 λ 인 Poisson 분포를 가집니다. 일반적으로 Poisson 분포는 특정 구간 내의 사건 수를 모델링하며 대개 개수 데이터로 표현됩니다. λ 의 연결 함수는 로그입니다. Poisson 회귀는 Y 가 정수가 아닌 값이더라도 허용됩니다. 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오.

음이항 Y 는 평균이 μ 이고 산포 모수가 σ 인 음이항 분포를 가집니다. 일반적으로 음이항 분포는 지정된 실패 횟수에 도달하기 전의 성공 횟수를 모델링합니다. 또한 음이항 분포는 특정 조건에서 감마 Poisson 분포와 동등합니다. 음이항 분포와 감마 Poisson 분포 간 연결에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오.

평균 λ 및 산포 모수 σ 를 갖는 감마 Poisson 분포를 평균 λ 를 갖는 Poisson 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoGammaPoisson.jsl 을 실행하십시오.

μ 의 연결 함수는 로그입니다. 음이항 회귀는 Y 가 정수가 아닌 값이더라도 허용됩니다. 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오.

JMP PRO 영과잉

ZI 이항 Y 는 모수 p, n 및 영과잉 모수 π 가 있는 영과잉 이항 분포를 가집니다. 반응 Y 는 n 번의 독립 시행 (모든 시행에 대한 확률 p 는 고정됨) 에서 총 성공 횟수를 나타냅니다. 이 분포에서는 표본 크기 열을 사용할 수 있습니다. 나열된 열이 없으면 표본 크기가 1 이라고 가정합니다. p 의 연결 함수는 로짓입니다. 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오.

ZI 베타 이항 Y 는 성공 확률 p, 시행 횟수 n, 과대산포 모수 δ 및 영과잉 모수 π 가 있는 베타 이항 분포를 가집니다. 이 분포는 ZI 이항 분포의 과대산포 버전입니다. ZI 베타 이항 분포를 사용하려면 각 관측값에 대해 표본 크기가 1 보다 커야 합니다. 따라서 사용자는 표본 크기 열을 지정해야 합니다. 표본 크기 열을 삽입하려면 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 Y 로 지정합니다. p 의 연결 함수는 로짓입니다. 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오.

ZI Poisson Y 는 평균 모수가 λ 이고 영과잉 모수가 π 인 영과잉 Poisson 분포를 가집니다. λ 모수는 0 팽창 없이 Poisson 분포의 관측값을 기반으로 하는 조건부 평균입니다. λ 의 연결 함수는 로그입니다. ZI Poisson 회귀는 Y 가 정수가 아닌 값이거나 0 값이 관측되지 않더라도 허용됩니다. 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오.

ZI 음이항 Y 는 위치 모수 μ, 산포 모수 σ 및 영과잉 모수 π 가 있는 영과잉 음이항 분포를 가집니다. μ 모수는 0 팽창 없이 음이항 분포의 관측값을 기반으로 하는 조건부 평균입니다. μ 의 연결 함수는 로그입니다. ZI 음이항 회귀는 Y 가 정수가 아닌 값이거나 0 값이 관측되지 않더라도 허용됩니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309페이지)에서 확인하십시오.

ZI 감마 Y 는 평균 모수가 μ 이고 영과잉 모수가 π 인 영과잉 감마 분포를 가집니다 . 많은 경우 0 이 아닌 반응이 감마 분포를 따른다고 간주할 수 있습니다 . 보험금 청구의 경우 여기에 해당합니다 . 청구 값은 감마 분포에 근사하지만 청구가 없는 보험 증권에 대한 데이터에는 0 도 있습니다 . 영과잉 감마 분포는 데이터를 0 반응과 0 이 아닌 반응으로 나누지 않고도 이러한 데이터를 직접 처리할 수 있습니다 . μ 모수는 0 팽창 없이 감마 분포의 관측값을 기반으로 하는 조건부 평균입니다 . μ 의 연결 함수는 로그입니다 . 자세한 내용은 " 분포에 대한 통계 상세 정보 "(309 페이지) 에서 확인하십시오 .

표 6.1에는 다양한 분포가 할당된 Y 변수의 데이터 유형, 모델링 유형 및 기타 요구 사항이 나와 있습니다.

표 6.1 분포에 대한 Y 요구 사항

분포	데이터 유형	모델링 유형	기타
정규	숫자	연속형	
Cauchy	숫자	연속형	
t(5)	숫자	연속형	
지수	숫자	연속형	양수
감마	숫자	연속형	양수
Weibull	숫자	연속형	양수
로그 정규	숫자	연속형	양수
베타	숫자	연속형	0에서 1 사이
분위수 회귀	숫자	연속형	
Cox 비례 위험	숫자	연속형	음수 아님
이항, 요약되지 않음	임의	임의	이진
이항, 빈도 열로 요약됨	임의	임의	이진
이항, 두 번째 Y로 입력된 개수 열로 요약됨	숫자	연속형	음수 아님
베타 이항	숫자	연속형	음수 아님
다항	임의	명목형 또는 순서형	
순서형 로지스틱	임의	순서형	
Poisson	숫자	임의	음수 아님

표 6.1 분포에 대한 Y 요구 사항 (계속)

분포	데이터 유형	모델링 유형	기타
음이항	숫자	임의	음수 아님
영과잉 이항	숫자	임의	음수 아님
영과잉 베타 이항	숫자	임의	음수 아님
영과잉 Poisson	숫자	임의	음수 아님
영과잉 음이항	숫자	임의	음수 아님
영과잉 감마	숫자	연속형	음수 아님

이러한 분포를 파라미터화하는 방법에 대한 자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오. 표 6.2에는 상세 정보가 요약되어 있습니다.

표 6.2 분포 , 모수 및 연결 함수

분포	모수	평균 모형 연결 함수
정규	μ, σ	항등 (μ)
Cauchy	μ, σ	항등 (μ)
t(5)	μ, σ	항등 (μ)
지수	μ	로그 (μ)
감마	μ, σ	로그 (μ)
Weibull	μ, σ	항등 (μ)
로그 정규	μ, σ	항등 (μ)
베타	μ	로짓 (μ)
이항	*n*, *p*	로짓 (*p*)
베타 이항	*n*, *p*, δ	로짓 (*p*)
Poisson	λ	로그 (μ)
음이항	μ, σ	로그 (μ)
영과잉 이항	*n*, *p*, π(영과잉)	로짓 (*p*)
영과잉 베타 이항	*n*, *p*, δ, π(영과잉)	로짓 (*p*)

표 6.2 분포 , 모수 및 연결 함수 (계속)

분포	모수	평균 모형 연결 함수
영과잉 Poisson	λ, π(영과잉)	로그 (μ)
영과잉 음이항	μ, σ, π(영과잉)	로그 (μ)
영과잉 감마	μ, σ, π(영과잉)	로그 (μ)

적절한 분포를 선택한 후 **실행**을 클릭합니다. 그러면 " 일반화 회귀 " 보고서 창이 나타납니다.

JMP PRO 일반화 회귀 보고서 창

" 모형 적합 " 시작 창에서 **실행**을 클릭하면 다음 항목이 포함된 " 일반화 회귀 " 보고서 창이 나타납니다 .

- 모형 비교 보고서 - 보고서의 적합된 모든 모형을 비교할 수 있습니다 . " 모형 시작 " 제어판을 사용하여 새 모형이 적합될 때마다 " 모형 비교 " 보고서에 추가됩니다 . " 표시 " 열의 체크 박스를 사용하여 적합 모형에 대한 보고서를 표시하거나 숨길 수 있습니다 . " 모형 비교 " 보고서의 다른 열에는 적합된 각 모형에 대한 정보 및 적합 통계량이 포함됩니다 . " 모형 비교 " 보고서의 열을 기준으로 모형을 정렬하려면 원하는 열 머리글을 클릭합니다 . 처음 클릭하면 오름차순으로 정렬되고 해당 열 머리글을 두 번째 클릭하면 내림차순으로 정렬됩니다 .

  참고 : 하나 이상의 모형이 적합되면 " 모형 비교 " 보고서가 나타납니다 .

- 모형 시작 제어판 - 모형을 적합시킵니다 . 자세한 내용은 " 모형 시작 제어판 "(278 페이지) 에서 확인하십시오 . 모형을 적합시키면 적합된 모형의 유형을 설명하는 개요가 제목에 추가됩니다 . 자세한 내용은 " 모형 적합 보고서 "(287 페이지) 및 " 모형 적합 옵션 "(297 페이지) 에서 확인하십시오 .
  - 모형 항 사이에 선형 종속성이 있는 경우 모형 항이 충족하는 선형 함수를 보여 주는 " 특이성 상세 정보 " 보고서가 " 모형 시작 " 제어판에 포함됩니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 모형 항 사이에 선형 종속성이 있는 모형 "(182 페이지) 에서 확인하십시오 .
- 최대 가능도 보고서 - 최대 가능도 추정을 사용하여 적합된 모형의 결과를 보여 줍니다 . 자세한 내용은 " 최대 가능도 "(280 페이지) 에서 확인하십시오 . " 최대 가능도 " 보고서는 다음 조건이 충족될 경우에만 나타납니다 .
  - 예측 변수 사이에 선형 종속성이 없습니다 .
  - 관측값이 예측 변수보다 더 많습니다 .
  - 예측 변수가 250 개 이하입니다 .

참고 : 정규 분포가 지정된 경우 이 보고서에 " 표준 최소 제곱 " 라벨이 지정됩니다 . 이항 분포가 지정된 경우 이 보고서에 " 로지스틱 회귀 " 라벨이 지정됩니다 .

JMP PRO 일반화 회귀 보고서 옵션

모형 대화상자 현재 분석을 위해 완료된 " 모형 적합 " 시작 창을 표시합니다 .

난수 시드값 설정 K 폴드 및 홀드백 검증에 사용되는 난수화 과정을 위한 시드값을 설정합니다 . 분석을 재현하려는 경우에 유용합니다 . 시드를 양수 값으로 설정하고 스크립트를 저장하면 시드값이 스크립트에 자동으로 저장됩니다 . 스크립트를 실행하면 항상 동일한 교차 검증 분석이 생성됩니다 .

코딩 테이블 저장 첫 번째 열에 모든 모형 모수에 대한 JMP 코딩이 표시되는 새 데이터 테이블을 생성합니다 . 마지막 열에는 반응 변수의 값이 표시됩니다 . 이항 반응에 대한 표본 크기 열을 지정한 경우 반응 열과 표본 크기 열이 모두 테이블에 표시됩니다 . 두 개의 반응 열을 사용하여 중도절단을 지정한 경우 두 반응 열이 모두 테이블에 표시됩니다 . " 감지 한계 " 열 특성이 포함된 반응 열을 사용하여 중도절단을 지정한 경우에는 코딩 테이블의 반응 열에 " 감지 한계 " 열 특성이 포함됩니다 .

참고 : 코딩 데이터 테이블에는 분석에 사용된 데이터 테이블 이름을 제공하는 " 원래 데이터 " 라는 테이블 변수가 포함되어 있습니다 . 기준 변수가 지정된 경우에는 " 원래 데이터 " 테이블 변수에 기준 변수와 해당 수준도 제공됩니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO 모형 시작 제어판

" 모형 시작 " 제어판에는 다음과 같은 옵션이 있습니다 .

- " 반응 분포 "(279 페이지)
- " 추정 방법 옵션 "(279 페이지)
- " 고급 제어 "(283 페이지)

- " 검증 방법 옵션 "(286 페이지)
- " 조기 중지 "(287 페이지)
- " 시작 "(287 페이지)

JMP PRO 반응 분포

제어판에서 사용할 수 있는 반응 분포는 "모형 적합" 시작 창에 지정된 반응 변수 속성에 따라 다릅니다. 적절한 반응 분포가 "반응 분포" 옵션에 제공됩니다. 반응 분포에 대한 자세한 내용은 "분포"(270 페이지)에서 확인하십시오.

JMP PRO 추정 방법 옵션

사용 가능한 추정 방법은 선택 및 벌점이 없는 기법, 단계별 모형 선택 기법 및 벌점 회귀 기법으로 그룹화할 수 있습니다.

최대 가능도, 표준 최소 제곱 및 로지스틱 회귀 방법은 "모형 적합" 시작 창에 지정된 전체 모형을 적합시킵니다. 변수 선택은 수행되지 않습니다. 이러한 모형은 다른 방법과의 비교를 위한 기준으로 사용될 수 있습니다.

참고 : 제공된 보고서에 대해 최대 가능도 , 표준 최소 제곱 및 로지스틱 회귀 중 하나만 사용할 수 있습니다 . 이 추정 방법의 이름은 " 모형 적합 " 시작 창에 지정된 분포에 따라 다릅니다 .

후진 제거, 전진 선택, 가지 치기 전진 선택, 최량 부분집합 및 2 단계 전진 선택 방법은 각 단계에서 모형에 추가되거나 제거되는 변수를 기반으로 합니다. 그러나 회귀 계수에 벌점을 부과하지 않습니다.

Dantzig 선택기, Lasso 회귀, Elastic Net, 능형 회귀 및 이중 Lasso 회귀 방법은 벌점 회귀 기법입니다. 모형의 예측 능력을 향상시키기 위해 회귀 계수 크기를 축소하고 추정값 분산을 줄입니다.

참고 : 데이터의 공선성이 높은 경우 Lasso 회귀와 Elastic Net 의 적응형 버전에서 적절한 해를 제공하지 못할 수 있습니다 . 적응형 버전은 MLE 가 적절한 추정값을 제공한다고 가정하기 때문입니다 . 이러한 경우에는 " 적응형 " 옵션을 사용하지 않는 것이 좋습니다 .

이러한 기법에는 다음과 같은 두 가지 유형의 벌점이 사용됩니다 .

- l_1 벌점 - 회귀 계수의 **절대값** 합에 벌점을 부과합니다 .
- l_2 벌점 - 회귀 계수의 **제곱**합에 벌점을 부과합니다 .

관측 데이터에 대한 기본 추정 방법은 Lasso 회귀입니다. 데이터 테이블에 DOE 스크립트가 포함되어 있고 특이성이 없는 경우 기본 추정 방법은 효과 유전성 옵션이 활성화된 전진 선택입니다. 데이터 테이블에 DOE 스크립트가 포함되어 있고 설계 행렬에 특이성이 있는 경우 기본 추정 방법은 효과 유전성 옵션이 활성화된 2 단계 전진 선택입니다.

모형 적합에 다음 방법을 사용할 수 있습니다 .

선택 및 벌점이 없는 추정 방법

최대 가능도 모형 모수에 대한 MLE(최대 가능도 추정값) 를 계산합니다 . 벌점이 부과되지 않습니다 . 분위수 회귀의 경우 최대 가능도 추정 방법만 사용할 수 있습니다 . " 모형 적합 " 시작 창에서 " 검증 " 열을 지정한 경우 최대 가능도 모형이 훈련 데이터 집합에 적합됩니다 . 다음 조건이 충족되면 최대 가능도 모형 보고서가 기본적으로 나타납니다 .

- 예측 변수 사이에 선형 종속성이 없습니다 .
- 관측값이 예측 변수보다 더 많습니다 .
- 예측 변수가 250 개 이하입니다 .

" 최대 가능도 " 옵션을 사용하면 일반화 회귀 분석법에서 지원하는 반응 분포에 대한 전통적 모형을 생성할 수 있습니다 . 또한 최대 가능도 기반 모형을 모형 비교를 위한 기준으로 사용할 수 있습니다 .

정규 또는 이항 분포가 지정된 경우 " 최대 가능도 " 방법을 각각 " 표준 최소 제곱 " 또는 " 로지스틱 회귀 " 라고 합니다 .

표준 최소 제곱 정규 분포를 지정하면 " 최대 가능도 " 추정 방법이 " 표준 최소 제곱 " 추정 방법으로 대체됩니다 . 기본 보고서는 일반적인 표준 최소 제곱 결과를 제공하는 " 표준 최소 제곱 " 보고서입니다 .

로지스틱 회귀 이항 분포를 지정하면 " 최대 가능도 " 추정 방법이 " 로지스틱 회귀 " 추정 방법으로 대체됩니다 . 기본 보고서는 " 로지스틱 회귀 " 보고서입니다 . 로지스틱 결과는 최대 가능도 결과와 동일합니다 .

단계별 추정 방법

참고 : 다항 분포가 지정된 경우에는 단계별 추정 방법을 사용할 수 없습니다 .

후진 제거 후진 제거 회귀를 사용하여 모수 추정값을 계산합니다 . 선택한 모형은 선택한 검증 방법과 관련된 최적 해를 제공합니다 . 후진 제거는 모형에 모든 모수를 포함하고 모형에 절편만 남을 때까지 각 단계에서 하나의 효과를 제거하는 것으로 시작됩니다 . 각 단계마다 각 모수에 대해 Wald 검정을 사용하여 제거할 모수를 결정합니다 .

주의 : 후진 제거를 위한 해 경로의 가로 축은 다른 추정 방법에서 동일한 축의 반대입니다 . 따라서 후진 제거 추정 방법에 대한 해 경로에서 왼쪽에서 오른쪽으로 이동하면 모형에 항이 추가되는 것이 아니라 제거됩니다 .

전진 선택 단계별 전진 회귀를 사용하여 모수 추정값을 계산합니다 . 각 단계에서 스코어 검정이 가장 유의한 효과가 모형에 추가됩니다 . 선택한 모형은 선택한 검증 방법과 관련된 최적 해를 제공합니다 .

교호작용이 있고 효과 유전성 옵션이 활성화된 경우 교락 효과는 다음과 같은 방식으로 처리됩니다 . 주어진 단계에서 스코어 검정이 가장 유의한 효과가 효과 유전성을 위반하는 경우

교락 효과가 생성됩니다 . 교락 효과는 스코어 검정이 가장 유의한 효과 및 효과 유전성을 충족하는 데 필요한 기타 비활성 효과를 포함합니다 . 교락 효과에 스코어 검정이 가장 유의한 효과가 있는 경우 교락 효과의 모든 효과가 모형에 추가됩니다 .

가지 치기 전진 선택 전진 단계와 후진 단계를 혼합하여 모수 추정값을 계산합니다 . 절편만 있는 모형으로 알고리즘이 시작됩니다. 첫 번째 단계에서는 스코어 검정이 가장 유의한 효과가 모형에 추가됩니다 . 첫 번째 단계 이후 알고리즘은 각 단계에서 다음과 같은 세 가지 가능성을 고려합니다 .

1. 모형에 없는 효과 중 스코어 검정이 가장 유의한 효과를 추가합니다 .
2. 모형에 있는 효과 중 Wald 검정이 가장 유의하지 않은 효과를 제거합니다 .
3. 한 단계에서 위의 두 작업을 모두 수행합니다 .

각 단계에서 수행되는 작업을 선택하기 위해 알고리즘은 지정된 검증 방법을 사용합니다. 예를 들어 BIC 검증 방법이 지정된 경우 알고리즘은 결과 BIC 값이 가장 작은 작업을 선택합니다 . 교호작용이 있고 효과 유전성 옵션이 활성화된 경우 교락 효과는 효과 추가에만 고려되고 효과 제거에는 고려되지 않습니다 .

모형이 포화 상태가 되면 알고리즘은 후진 단계를 시도하여 모형이 개선되는지 확인합니다 . 알고리즘의 최대 단계 수는 모수 수의 5 배입니다 . 선택한 모형은 선택한 검증 방법과 관련된 최적 해를 제공합니다 .

가지 치기 전진 선택은 단계별 회귀 분석법의 혼합 단계 옵션 대신 사용할 수 있습니다 . 그러나 모형에 추가되거나 제거되는 변수를 결정할 때 p 값을 사용하지 않습니다 .

팁 : 가지 치기 전진 선택 추정 방법에는 조기 중지 옵션을 사용하지 않는 것이 좋습니다 .

최량 부분집합 각 단계에서 모형의 활성 효과 수를 늘려 모수 추정값을 계산합니다 . 각 단계마다 단계 번호에 지정된 수의 효과를 포함하는 모든 가능 모형 중에서 적절한 모형이 선택됩니다 . 해 경로 그림의 가로 축에 있는 값은 모형의 활성 효과 수를 나타냅니다 . 단계 0 은 절편만 있는 모형에 해당합니다 . 단계 1 은 활성 효과가 하나만 포함된 모형 중 최적 모형에 해당합니다 . 단계는 " 모형 시작 " 보고서의 " 고급 제어 " 에 지정된 " 최대 효과 수 " 값까지 계속됩니다 . 자세한 내용은 " 고급 제어 "(283 페이지) 에서 확인하십시오 .

팁 : 최량 부분집합 추정 방법은 계산량이 많습니다 . 따라서 큰 문제에는 권장되지 않습니다 .

2 단계 전진 선택 (모형에 2 차 이상 효과가 있는 경우에만 사용 가능) 모수 추정값을 두 단계로 계산합니다 . 첫 번째 단계에서는 주효과에 대해 단계별 전진 회귀 모형을 실행하여 모형에 유지할 효과를 결정합니다 . 두 번째 단계에서는 첫 번째 단계에서 선택된 주효과로만 구성된 모든 고차 효과에 대해 단계별 전진 회귀 모형을 실행합니다 . 이 방법은 강한 효과 유전성을 가정합니다 .

첫 번째 단계에서 유지되지 않은 항은 " 모수 추정값 " 보고서에 0 인 항으로 나타납니다 . 그러나 두 번째 단계 모형을 적합시킬 때 무시됩니다 . 첫 번째 단계에서 선택된 항은 두 번째 단계에 강제 적용되지 않으며 두 번째 단계에서 선택할 수 있습니다 .

벌점 추정 방법

Dantzig 선택기 (정규 분포가 지정되고 " 절편 없음 " 옵션이 선택되지 않은 경우에만 사용 가능) 선형 프로그래밍 방식으로 l_1 벌점을 적용하여 모수 추정값을 계산합니다 . 자세한 내용은 Candes and Tao 연구 자료 (2007) 에서 확인하십시오 . Dantzig 선택기는 설계된 실험의 결과를 분석하는 데 유용합니다 . 직교 문제의 경우 Dantzig 선택기와 Lasso 회귀에서 동일한 결과를 제공합니다 . 자세한 내용은 "Dantzig 선택기 "(306 페이지) 에서 확인하십시오 .

올가미 l_1 벌점을 적용하여 모수 추정값을 계산합니다. l_1 벌점으로 인해 일부 계수가 0으로 추정될 수 있습니다 . 따라서 적합 절차의 일부로 변수 선택이 수행됩니다 . 일반 Lasso 회귀의 경우 모든 계수에 동일하게 벌점이 적용됩니다 .

적응형 Lasso 회귀 회귀 계수 절대값의 가중 합에 벌점을 적용하여 모수 추정값을 계산합니다 . l_1 벌점의 가중치는 oracle 특성 (Zou 2006) 을 유지하는 것과 같은 방식으로 데이터에 의해 결정됩니다 . 이 옵션은 MLE 를 사용하여 l_1 벌점에 가중치를 적용합니다 . 예측 변수의 수가 관측값 수를 초과하거나 예측 변수 사이에 엄격한 선형 종속성이 있으면 MLE 를 계산할 수 없습니다 . 회귀 모수의 MLE 를 계산할 수 없는 경우 일반화 역 해 또는 능형 회귀 해가 l_1 벌점 가중치에 사용됩니다 . 자세한 내용은 " 적응형 방법 "(307 페이지) 에서 확인하십시오 .

일반적으로 "Lasso 회귀 " 및 " 적응형 Lasso 회귀 " 옵션은 예측 변수의 상관관계가 높을 때 간결한 모형을 선택합니다 . 이러한 기법은 상관관계가 높은 예측 변수 그룹 중 하나만 선택하는 경향이 있습니다 . 고차원 데이터의 예측 변수는 대개 상관관계가 높습니다 . 이러한 유형의 데이터에는 Lasso 회귀보다 Elastic Net 을 사용하는 것이 좋습니다 . 자세한 내용은 "Lasso 회귀 "(306 페이지) 에서 확인하십시오 .

Elastic Net l_1 벌점과 l_2 벌점을 모두 적용하여 모수 추정값을 계산합니다 . l_1 벌점을 적용하면 변수 선택이 수행됩니다 . l_2 벌점은 능형 회귀와 같이 계수를 축소하여 예측 능력을 향상시킵니다 .

적응형 Elastic Net 적응형 l_1 벌점과 l_2 벌점을 사용하여 모수 추정값을 계산합니다 . 이 옵션은 MLE 를 사용하여 l_1 벌점에 가중치를 적용합니다 . 예측 변수의 수가 관측값 수를 초과하거나 예측 변수 사이에 엄격한 선형 종속성이 있으면 MLE 를 계산할 수 없습니다 . 회귀 모수의 MLE 를 계산할 수 없는 경우 일반화 역 해 또는 능형 회귀 해가 l_1 벌점 가중치에 사용됩니다 . " 고급 제어 " 패널에서 "Elastic Net 알파 " 값을 설정할 수 있습니다 . 자세한 내용은 " 적응형 방법 "(307 페이지) 에서 확인하십시오 .

예측 변수의 상관관계가 높을 때 Elastic Net 의 예측 정확성이 Lasso 회귀보다 높은 경향이 있습니다 . 실제로 능형 회귀와 Lasso 회귀 모두 Elastic Net 의 특수한 사례입니다 . 예측 능력 면에서는 적응형 Elastic Net 이 Elastic Net 및 적응형 Lasso 회귀를 능가하는 경우가 많습니다 . Elastic Net 에는 상관관계가 높은 예측 변수 그룹을 선택하고 관련 예측 변수에 적절한 모수 추정값을 할당하는 기능이 있습니다 . 자세한 내용은 "Elastic Net"(307 페이지) 에서 확인하십시오 .

참고 : "Elastic Net" 적합을 선택하고 "Elastic Net 알파 " 를 결측으로 설정하면 알고리즘이 Lasso 회귀 , Elastic Net, 능형 회귀 적합을 순서대로 계산합니다 . 적합에 시간이 많이 소요되는 경우 진행률 표시줄이 나타납니다 . " 현재 추정값 채택 " 을 클릭하면 계산이 중지되며 보고된 모수 추정값은 해당 지점에서 최적 모형 적합에 해당합니다 . 진행률 표시줄은 알고리즘이 각각 Lasso 회귀 , Elastic Net 및 능형 회귀를 적합시키고 있음을 나타냅니다 . 이 정보를 사용하여 " 현재 추정값 채택 " 을 클릭할 시기를 결정할 수 있습니다 .

능형 회귀 능형 회귀를 사용하여 모수 추정값을 계산합니다 . 능형 회귀는 l_2 벌점을 적용하고 모수 추정값이 0 이 되지 않는 편향 회귀 기법입니다 . 모형의 모든 예측 변수를 유지하려는 경우에 유용합니다 . 자세한 내용은 " 능형 회귀 "(306 페이지) 에서 확인하십시오 .

이중 Lasso 회귀 모수 추정값을 두 단계로 계산합니다 . 첫 번째 단계에서는 Lasso 회귀 모형을 적합시켜 두 번째 단계에 사용할 항을 결정합니다 . 두 번째 단계에서는 첫 번째 단계에서 결정된 항을 사용하여 Lasso 회귀 모형을 적합시킵니다 . 표시된 해 경로 결과 및 모수 추정값 보고서는 두 번째 단계 적합에 대한 것입니다 . 첫 번째 단계에서 모형에 추가되는 변수가 없으면 두 번째 단계 없이 첫 번째 단계의 결과가 보고서에 나타납니다 .

이중 Lasso 회귀는 관측값 수가 예측 변수의 수보다 적을 때 특히 유용합니다 . 변수 선택 및 축소 작업을 두 단계로 나누면 두 번째 단계의 Lasso 회귀에서 모형에 포함되어야 할 항에 과도하게 벌점을 부과할 가능성이 적습니다 . 이중 Lasso 회귀는 완화 Lasso 회귀와 비슷합니다 . 완화 Lasso 회귀는 Hastie et al. 연구 자료 (2009, p. 91) 에 설명되어 있습니다 .

적응형 이중 Lasso 회귀 모수 추정값을 두 단계로 계산합니다 . 첫 번째 단계에서는 적응형 Lasso 회귀 모형을 적합시켜 두 번째 단계에 사용할 항을 결정합니다 . 두 번째 단계에서는 첫 번째 단계에서 결정된 항을 사용하여 적응형 Lasso 회귀 모형을 적합시킵니다 . 두 번째 단계에서는 첫 번째 단계 모형에 포함된 항만 고려하고 첫 번째 단계의 모수 추정값을 기반으로 한 가중치를 사용합니다 . " 고급 제어 " 의 " 적응형 벌점 가중치 " 옵션을 사용하여 가중치 계산 방법을 선택할 수 있습니다 . 자세한 내용은 " 고급 제어 옵션 "(284 페이지) 에서 확인하십시오 . 표시된 결과는 두 번째 단계 적합에 대한 것입니다 . 첫 번째 단계에서 모형에 추가되는 변수가 없으면 두 번째 단계 없이 첫 번째 단계의 결과가 보고서에 나타납니다 . 자세한 내용은 " 적응형 방법 "(307 페이지) 에서 확인하십시오 .

JMP PRO 고급 제어

"고급 제어" 옵션을 사용하여 모형 적합 과정의 다양한 요소를 조정할 수 있습니다. 여러 컨트롤이 조정 모수의 격자와 관련이 있습니다.

조정 모수

Lasso 회귀 및 능형 회귀 추정 방법의 해 경로는 단일 조정 모수에 따라 다릅니다. Elastic Net 의 해 경로는 가능도 벌점에 대한 조정 모수 및 Elastic Net 알파에 따라 다릅니다. Elastic Net 의 가능도에 대한 벌점은 Lasso 회귀 및 능형 회귀 추정 방법과 관련된 벌점의 가중 합입니다. Elastic Net 알파는 이러한 두 벌점의 가중치를 결정합니다. 자세한 내용은 "추정 방법에 대한 통계 상세 정보"(306 페이지) 및 "고급 제어에 대한 통계 상세 정보"(308 페이지) 에서 확인하십시오.

조정 모수가 0 이면 해에 벌점이 부과되지 않고 최대 가능도 추정값이 계산됩니다. 조정 모수가 증가하면 벌점이 증가합니다.

해는 선택한 검증 방법과 관련하여 벌점이 부과된 음의 로그 가능도 함수를 최소화하는 모수 추정값 집합입니다. 현재 해는 해 경로 그림에서 빨간색 수직 실선으로 지정됩니다.

참고 : 해 경로 그림에서 척도화된 모수 추정값 크기가 감소하면 조정 모수 값이 증가합니다 . MLE 에 가까운 추정값은 큰 크기와 관련이 있고 벌점이 크게 부과된 추정값은 작은 크기와 관련이 있습니다 .

다음 사항에 유의하십시오 .

- 조정 모수가 너무 작으면 일반적으로 데이터가 과대적합되어 분산이 큰 모형이 생성됩니다 .
- 조정 모수가 너무 크면 일반적으로 데이터가 과소적합됩니다 .

조정 모수 격자

해를 얻기 위해 미세 격자에서 조정 모수가 증가합니다 .

- Lasso 회귀 , Elastic Net 알파가 지정된 Elastic Net 및 능형 회귀의 경우 해를 제공하는 조정 모수의 값은 조정 모수 격자에서 최량 적합을 제공하는 값입니다 .

  참고 : "Elastic Net 알파 " 는 기본적으로 0.99 로 설정되어 있습니다 .

- "Elastic Net 알파 " 에 값을 설정하지 않으면 알파 값도 미세 격자에서 증가합니다 . 조정 모수의 값이 고정된 경우 알파는 연속된 10 개의 알파 값이 검증 방법에 의해 결정된 최량 적합을 개선하지 못할 때까지 변경됩니다 . 이 과정은 조정 모수 값의 전체 격자에 대해 반복됩니다 . 조정 모수 및 알파의 최종 값은 조정 모수 격자에서 최량 적합을 제공하는 값입니다 .

조정 모수 값 격자의 범위는 대부분의 경우 0 부터 모든 비절편 항이 0 인 가장 작은 값까지입니다 . 모든 비절편 항이 0 인 조정 모수의 가장 작은 값을 **상한**으로 정의합니다 . 조정 모수의 **하한**은 0.0001 로 설정되는 다음 두 경우를 제외하고 0 입니다 .

- 설계 행렬이 특이 행렬인 경우 최대 가능도 추정값을 계산할 수 없습니다 . 0.0001 하한을 사용하면 MLE 에 가까운 추정값을 계산할 수 있습니다 .
- 이항 또는 다항 분포를 선택한 경우 0.0001 하한을 사용하면 분리를 방지할 수 있습니다 .

고급 제어 옵션

효과 유전성 적용 관련 고차 효과보다 저차 효과를 먼저 모형에 입력해야 합니다 . 즉 , 대부분의 경우 X 가 모형에 없으면 X^2 도 모형에 없다는 의미입니다 . 그러나 " 전진 선택 " 을 제외한 다른 추정 방법의 경우 모형에 X^2 이 추가되고 동일한 단계에서 X 가 제거될 수 있습니다 . 이 옵션은 데이터 테이블에 DOE 스크립트가 포함된 경우 활성화되지만 기본적으로 해제되어 있습니다 .

Elastic Net 알파 Elastic Net 에 대한 α 모수를 설정합니다 . 이 α 모수는 Elastic Net 계수를 추정할 때 l_1 및 l_2 벌점 조정 모수의 혼합을 결정합니다 . 기본값은 $\alpha = 0.99$ 이며 , 이는 l_1 벌점의 계

수를 0.99 로 설정하고 l_2 벌점의 계수를 0.01 로 설정합니다. 이 옵션은 "추정 방법"에 "Elastic Net" 을 선택한 경우에만 사용할 수 있습니다 . 자세한 내용은 " 추정 방법에 대한 통계 상세 정보 "(306 페이지) 에서 확인하십시오 .

격자 점 수 조정 모수에 대한 하한과 상한 사이의 격자점 수를 지정합니다 . 각 격자점 값에서 해당 조정 모수 값에 대한 모수 추정값을 얻습니다 . 기본 격자점 수는 150 개입니다 .

최소 벌점률 조정 모수의 상한에 대한 하한 비율의 최소값을 나타냅니다 . 조정 모수의 하한이 0 이면 해가 MLE 를 제공합니다 . MLE 또는 MLE 에 매우 가까운 해를 포함하지 않으려는 경우 " 최소 벌점률 " 을 0 이 아닌 값으로 설정할 수 있습니다 . 이중 Lasso 회귀 추정 방법의 경우 이 옵션에 지정된 값은 적합의 첫 번째 단계에서만 사용됩니다 . 설계 행렬에 특이성이 있으면 기본값이 0.0001 입니다 . 그렇지 않으면 기본값이 0 입니다 .

격자 척도 격자 척도의 분포를 선택하는 옵션을 제공합니다 . 선형 , 제곱근 또는 로그 척도 중에서 선택할 수 있습니다 . 선택한 척도에 따라 지정된 격자점 수와 같은 수의 격자점이 조정 모수의 하한과 상한 사이에 분포됩니다 . 기본 격자 척도는 제곱근입니다 . 자세한 내용은 " 고급 제어에 대한 통계 상세 정보 "(308 페이지) 에서 확인하십시오 .

첫 번째 단계 해 이중 Lasso 회귀 및 2 단계 전진 선택의 첫 번째 단계에서 해를 선택하는 옵션을 제공합니다. 기본적으로 지정된 검증 방법에 따라 최량 적합 해가 선택되고 초기에 표시됩니다 (최량 적합). 녹색 또는 노란색 영역에 있는 l_1 노름 값이 더 크거나 더 작은 모형을 초기에 표시하도록 선택할 수 있습니다 . 예를 들어 " 노란색 영역에서 가장 작음 " 을 선택할 경우 초기에 표시되는 해는 노란색 영역에서 l_1 노름이 가장 작은 모형입니다 . 자세한 내용은 " 비교 가능 모형 영역 "(293 페이지) 에서 확인하십시오 .

최대 효과 수 최량 부분집합 추정 방법에 대해 모형에서 고려할 최대 효과 수를 지정합니다 . 이 옵션을 사용하여 모형 적합에 필요한 계산 수를 제한할 수 있습니다 . 기본값은 10 입니다 .

초기에 표시된 해 " 해 경로 " 보고서에서 현재 모형으로 초기에 표시되는 해를 선택하는 옵션을 제공합니다. 현재 모형은 수직 실선으로 식별됩니다. 자세한 내용은 " 현재 모형 표시자 "(291 페이지) 에서 확인하십시오 . 최량 적합 해는 수직 점선으로 식별됩니다 . 기본적으로 지정된 검증 방법에 따라 최량 적합으로 간주되는 해가 표시됩니다 .

녹색 또는 노란색 영역에 있는 l_1 노름 값이 더 크거나 더 작은 모형을 초기에 표시하도록 선택할 수 있습니다 . 예를 들어 " 노란색 영역에서 가장 작음 " 을 선택할 경우 초기에 표시되는 해는 노란색 영역에서 l_1 노름이 가장 작은 모형입니다 . 자세한 내용은 " 비교 가능 모형 영역 "(293 페이지) 에서 확인하십시오 .

적응형 벌점 가중치 적응형 이중 Lasso 회귀의 두 번째 단계에서 사용되는 벌점 가중치를 계산하는 옵션을 제공합니다 . 기본적으로 " 역 해 " 옵션이 선택되어 있습니다 . 이 옵션은 첫 번째 단계 적합의 모수 추정값을 사용하여 벌점 가중치를 계산합니다 .

" 역 모형 평균 " 옵션은 AICc 또는 BIC 모형의 가중 평균인 해의 모수 추정값을 사용하여 벌점 가중치를 계산합니다 . AICc 검증 방법을 선택하는 경우 AICc 모형이 사용됩니다 . 그렇지 않은 경우 BIC 모형이 사용됩니다 . " 역 모형 평균 " 옵션을 사용하면 최대 가능도 해 (있는 경우) 가 적응형 이중 Lasso 회귀 모형의 해 경로에서 맨 오른쪽 점으로 나타납니다 .

항 강제 적용 모형에 강제로 적용할 항 (있는 경우) 을 선택할 수 있습니다 . 모형에 강제 적용된 항은 벌점에 포함되지 않습니다 .

JMP PRO 검증 방법 옵션

다음 방법을 사용하여 모형 적합을 검증할 수 있습니다 .

참고 : " 분위수 회귀 " 분포에는 " 없음 " 검증 방법 옵션만 허용됩니다 . " 최대 가능도 " 추정 방법에는 " 없음 " 과 " 검증 열 " 검증 방법만 허용됩니다 . "Cox 비례 위험 " 분포에는 "BIC", "AICc" 및 " 없음 " 검증 방법만 허용됩니다 . "Dantzig 선택기 " 추정 방법에는 "BIC" 와 "AICc" 검증 방법만 허용됩니다 .

K 폴드 조정 모수의 각 값에 대해 다음 단계가 수행됩니다 .

- 관측값이 *k* 개의 부분집합 또는 **폴드**로 분할됩니다 .
- 그러면 각 폴드가 검증 데이터 집합으로 사용됩니다 . 모형은 폴드에 **없는** 관측값에 적합됩니다 . 폴드에 **있는** 관측값에 대해 해당 모형에 기반한 로그 가능도가 계산되어 **검증** 로그 가능도를 제공합니다 .
- *k* 폴드에 대한 검증 로그 가능도의 평균이 계산됩니다 . 이 값은 조정 모수 값에 대한 검증 로그 가능도로 사용됩니다 .

최대 검증 로그 가능도를 갖는 조정 모수의 값은 최종 해를 생성하는 데 사용됩니다 . 최종 모형을 생성하기 위해 조정 모수의 최적값에 대해 파생된 모든 *k* 개 모형이 전체 데이터 집합에 적합됩니다 . 이 중에서 검증 로그 가능도가 가장 높은 모형이 최종 모형으로 선택됩니다 . 이 최종 모형에 사용된 훈련 데이터 집합이 훈련 데이터 집합으로 지정되고 이 모형의 홀드아웃 폴드가 검증 데이터 집합입니다 . 이러한 훈련 데이터 집합과 검증 데이터 집합이 최종 해에 대해 보고된 결과 및 그림에 사용됩니다 .

홀드백 검증 데이터 집합에 대해 지정된 비율의 데이터를 무작위로 선택하고 데이터의 다른 부분을 사용하여 모형을 적합시킵니다 . 최종 해는 검증 데이터 집합에 대한 음의 로그 가능도를 최소화하는 값입니다 . 이 방법은 큰 데이터 집합에 유용합니다 . 랜덤 선택은 단순 랜덤 표집을 기반으로 할 때보다 더 균형 있는 훈련 데이터 집합 및 검증 데이터 집합을 생성하기 위해 모형 요인에 대한 층화 표집을 기반으로 합니다 .

Leave-One-Out leave-one-out 교차 검증을 수행합니다 . 이 방법은 폴드 수가 행 수와 같은 K 폴드와 동등합니다 . 중간 또는 큰 데이터 집합에는 이 옵션을 사용하면 안 됩니다 . 심지어 중간 크기 정도의 관측값 수 일 때도 처리 시간이 오래 걸릴 수 있습니다 . 최종 해에 대해 보고된 결과 및 그림에 사용되는 훈련 데이터 집합과 검증 데이터 집합 (한 행) 은 K 폴드 검증에 대해 수행된 대로 결정됩니다 .

BIC 해 경로에서 BIC(Bayes 정보 기준) 를 최소화합니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

AICc 해 경로에서 수정 AICc(Akaike 정보 기준) 를 최소화합니다 . AICc 는 검증 방법의 기본 설정입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

참고 : 모수 수가 표본 크기에 근접하거나 초과하는 경우 AICc 가 정의되지 않습니다 .

ERIC 해 경로에서 ERIC(확장된 정규화 정보 기준) 를 최소화합니다 . 자세한 내용은 " 모형 적합 상세 정보 "(289 페이지) 에서 확인하십시오 . 지수 계열 분포와 Lasso 회귀 및 적응형 Lasso 회귀 추정 방법에만 사용할 수 있습니다 .

없음 검증을 사용하지 않습니다 . 최대 가능도 추정 방법 및 분위수 회귀에만 사용할 수 있습니다 .

검증 열 " 모형 적합 " 창에서 지정한 열을 " 검증 " 역할이 할당된 것으로 사용합니다 . 최종 해는 검증 데이터 집합에 대한 음의 로그 가능도를 최소화하는 값입니다 . Dantzig 선택기 추정 방법이 지정되거나 분위수 회귀 또는 Cox 비례 위험 분포가 지정된 경우에는 이 옵션을 사용할 수 없습니다 .

JMP PRO 조기 중지

조기 중지 옵션은 다음과 같은 조기 중지 규칙을 추가합니다 .

- 전진 선택의 경우 모형에 변수를 추가하여 검증 측도를 개선하지 못하는 단계가 연속으로 10 번이 되면 알고리즘이 종료됩니다 . 해는 10 번의 연속된 단계 앞에 오는 단계의 모형입니다 .
- Lasso 회귀 , Elastic Net 및 능형 회귀의 경우 연속된 10 개의 조정 모수 값이 검증 방법에 의해 결정된 최량 적합을 개선하지 못하면 알고리즘이 종료됩니다 . 해는 10 개의 연속된 값 앞에 오는 조정 모수 값에 해당하는 추정값입니다 .

참고 : AICc 및 BIC 검증 방법의 경우 4 개 이상의 예측 변수가 모형에 입력될 때까지 조기 중지가 발생하지 않습니다 .

JMP PRO 시작

시작을 클릭하면 보고서가 열립니다. 보고서 제목은 선택한 반응 분포, 추정 방법 및 검증 방법을 지정합니다. "모형 시작" 제어판으로 돌아가 추가 분석을 수행하고 다른 반응 분포, 추정 방법 및 검증 방법을 선택할 수 있습니다.

JMP PRO 모형 적합 보고서

" 모형 시작 " 패널에서 지정하는 각 추정 방법 및 검증 방법에 대해 보고서가 생성됩니다 . 보고서는 선택한 분포 , 추정 방법 및 검증 방법을 제목에 지정합니다 .

기본적으로 다음 보고서가 제공됩니다 .

- 회귀 그림
- 모형 요약
- 추정 상세 정보 (Lasso 회귀 , Elastic Net 및 능형 회귀에만 표시됨)
- 해 경로 (최대 가능도 추정 방법 및 분위수 회귀를 제외하고 모두 표시됨)
- 원래 예측 변수에 대한 모수 추정값

- 효과 검정

JMP PRO 회귀 그림

참고 : " 회귀 그림 " 보고서는 연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개일 때만 나타납니다 . " 분포 " 옵션이 " 다항 ", " 순서형 로지스틱 " 또는 "Cox 비례 위험 " 으로 설정된 경우에는 사용할 수 없습니다 . 반응은 연속형이어야 합니다 .

"회귀 그림" 보고서는 세로 축에 반응 값이 표시되고 가로 축에 연속형 예측 변수가 표시된 그림을 보여 줍니다. 점 위에 회귀선이 표시됩니다. 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 회귀선이 있고 그림 옆에 범례가 나타납니다. 반응이 성공 횟수 및 시행 횟수로 지정된 경우에는 성공 횟수를 시행 횟수로 나눈 값이 그림의 세로 축에 표시됩니다.

JMP PRO 모형 요약

" 모형 요약 " 보고서에서는 적합된 모형을 설명하고 적합 자체에 대한 요약 정보를 제공합니다 .

JMP PRO 모형 설명 상세 정보

" 모형 요약 " 보고서의 첫 번째 부분에서는 적합된 모형을 설명하는 정보를 제공합니다 .

반응 " 모형 적합 " 창에서 "Y" 역할에 할당된 열입니다 . 두 개의 열을 사용하여 구간 중도절단을 지정한 경우 두 열 이름이 모두 나열됩니다 .

분포 " 모형 적합 " 창에서 선택한 분포입니다 . 분위수 회귀의 경우 반응에 대해 지정된 분위수 값도 표시됩니다 .

추정 방법 " 모형 시작 " 패널에서 선택한 추정 방법입니다 .

검증 방법 " 모형 시작 " 패널에서 선택한 검증 방법입니다 .

평균 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 평균에 대한 모형 연결 함수입니다.

위치 모형 연결 위치 모수에 대한 모형 연결 함수이며 " 모형 적합 " 창에서 "Cauchy" 또는 "t(5)" 분포를 선택한 경우 표시됩니다 .

척도 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 척도 모수에 대한 모형 연결 함수입니다 .

확률 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 확률에 대한 모형 연결 함수입니다.

산포 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 산포 모수에 대한 모형 연결 함수입니다 .

영과잉 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 영과잉 모수에 대한 모형 연결 함수입니다 .

감지 하한 반응 열에 할당된 " 감지 한계 " 열 특성에 지정한 감지 하한입니다 .

감지 상한 반응 열에 할당된 " 감지 한계 " 열 특성에 지정한 감지 상한입니다 .

중도절단 열 " 모형 적합 " 창에서 " 중도절단 " 역할에 할당된 열입니다 .

중도절단 코드 오른쪽 중도절단된 관측값을 지정하는 중도절단 열의 값입니다 . 이 값은 " 모형 적합 " 창의 " 중도절단 코드 " 목록에 지정된 값입니다 .

JMP PRO 모형 적합 상세 정보

" 모형 요약 " 보고서의 두 번째 부분에서는 모형 적합과 관련된 통계량을 제공합니다 . " 검증 방법 " 에서 " 홀드백 " 또는 " 검증 열 " 을 선택하는 경우 이러한 통계량은 훈련 데이터 집합과 검증 데이터 집합에 대해 별도로 계산됩니다 . " 검증 방법 " 에서 "K 폴드 " 또는 "Leave-One-Out" 을 선택하면 " 모형 요약 " 보고서의 이 부분을 사용할 수 없습니다 .

행 수 행 수입니다 .

빈도 합 " 모형 적합 " 창에서 " 빈도 " 또는 " 가중치 " 역할에 할당된 열 값의 합입니다 .

참고 : -1* 로그 가능도 , BIC, AICc 및 ERIC 의 경우 값이 작을수록 좋습니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

-1* 로그 가능도 현재 모형에 대한 가능도 함수의 음의 자연 로그입니다. 자세한 내용은 "통계 상세 정보 " 장의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

참고 : 분위수 회귀에는 -1* 로그 가능도를 사용할 수 없습니다 .

목적 함수 (분위수 회귀에만 사용 가능) 지정된 분위수 회귀 모형을 적합시키기 위해 최소화되는 함수의 값입니다 . 최소화되는 함수는 점검 - 손실 함수입니다 .

모수 수 현재 모형에서 0 이 아닌 모수의 수입니다 .

BIC 다음과 같이 정의되는 Bayes 정보 기준입니다 .

$$\text{BIC} = -2LogLikelihood + k\ln(n)$$

자세한 내용은 "통계 상세 정보" 장의 "가능도, AICc 및 BIC"(514페이지)에서 확인하십시오.

AICc 다음과 같이 정의되는 수정된 Akaike 정보 기준입니다 .

$$\text{AICc} = -2LogLikelihood + 2k + 2k(k+1)/(n-k-1)$$

자세한 내용은 "통계 상세 정보" 장의 "가능도, AICc 및 BIC"(514페이지)에서 확인하십시오.

ERIC (지수 계열 분포와 Lasso 회귀 또는 적응형 Lasso 회귀 추정 방법이 지정된 경우에만 사용 가능) 확장된 정규화 정보 기준입니다 . 자세한 내용은 Hui et al. 연구 자료 (2015) 에서 확인하십시오 . ERIC 는 다음과 같이 정의됩니다 .

$$\text{ERIC} = -2LogLikelihood + (k-2)\ln(n\phi/\lambda)$$

여기서 λ 는 조정 모수의 값이고 ϕ 는 장애 모수입니다 .

일반화 R^2 (분위수 회귀에는 사용 불가능) 일반 회귀 모형에 적용할 수 있는 R^2 측도의 확장입니다 . 일반화 R^2 은 적합 모형의 가능도 (L_M) 를 절편만 있는 (상수) 모형의 가능도 (L_0) 와 비

교합니다 . 이 값은 최대값이 1 이 되도록 척도화됩니다 . 이항 분포를 제외한 다른 분포의 일반화 R^2 은 다음과 같이 정의됩니다 .

$$\text{일반화 } R^2 = 1 - (L_0/L_M)^{2/n}$$

이항 분포가 지정된 경우 일반화 R^2 은 다음과 같이 정의됩니다 .

$$\text{일반화 } R^2 = \frac{1 - (L_0/L_M)^{2/n}}{1 - L_0^{2/n}}$$

일반화 R^2 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다 . 일반화 R^2 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2 로 단순화됩니다 . 일반화 R^2 를 Nagelkerke 또는 Craig and Uhler R^2 라고도 하는데 , 이는 Cox-Snell 유사 R^2 를 정규화한 것입니다 . 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오 .

참고 : 정규 분포가 지정되면 일반화 R^2 이 R^2 으로 바뀝니다 .

주의 : 서로 다른 반응 분포를 사용하는 모형에 대해 일반화 R^2 값을 비교하면 안 됩니다 . 비교는 주어진 반응 분포를 사용하고 절편만 있는 모형을 대상으로 합니다 .

R^2 (정규 분포가 지정된 경우에만 사용 가능) 랜덤 오차가 아니라 모형에 기인할 수 있는 반응의 변동 비율을 추정합니다 . R^2 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다 . R^2 값은 다음과 같이 계산됩니다 .

$$1 - \frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}$$

Adj-R^2 (정규 분포가 지정되고 추정 방법에 벌점이 포함되지 않은 경우에만 사용 가능) 모형의 모수 수에 대한 R^2 통계량을 조정합니다 . Adj-R^2 을 사용하면 모수 수가 다른 모형을 쉽게 비교할 수 있습니다 . 계산에 자유도가 사용됩니다 . Adj-R^2 값은 다음과 같이 계산됩니다 .

$$1 - \frac{\frac{1}{N-(p-1)} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{\frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2}$$

여기서 N 은 관측값 수이고 p 는 모수 수입니다 .

참고 : 검증 데이터 집합이 있는 경우 조정 R^2 통계량은 훈련 데이터 집합에 대해서만 보고됩니다 .

RASE (정규 분포가 지정된 경우에만 사용 가능) RASE(제곱근 평균 제곱 오차) 는 현재 모형에서 평균 제곱 예측 오차의 제곱근입니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장의 "RASE"(82 페이지) 에서 확인하십시오 .

람다 벌점 (Dantzig 선택기 , Lasso 회귀 , Elastic Net, 능형 회귀 및 이중 Lasso 회귀 추정 방법에만 사용 가능) 현재 모형에 대한 조정 모수 λ 의 값입니다 . 자세한 내용은 " 추정 방법에 대한 통계 상세 정보 "(306 페이지) 에서 확인하십시오 .

JMP PRO 추정 상세 정보

" 추정 상세 정보 " 보고서에서는 Dantzig 선택기, Lasso 회귀, Elastic Net, 능형 회귀 및 이중 Lasso 회귀 추정 방법에 대한 고급 제어 설정을 보여 줍니다. 이러한 컨트롤에 대한 자세한 내용은 " 고급 제어 "(283 페이지) 에서 확인하십시오.

JMP PRO 해 경로

참고 : " 해 경로 " 보고서는 최대 가능도를 제외한 모든 추정 방법과 분위수 회귀를 제외한 모든 분포에 대해 나타납니다 .

" 해 경로 " 보고서에서는 다음 두 가지 그림을 보여 줍니다 .

- 해 경로 그림 - 추정된 모수의 값을 표시합니다 .
- 검증 그림 - 선택한 검증 방법에 해당하는 검증 통계량의 값을 표시합니다 .

두 그림에 대한 수평 척도는 척도화된 모수 추정값 크기로 제공됩니다 . 이 값은 평균 모형에 대한 척도화된 모수 추정값의 절대값을 합하여 정의되는 l_1 노름입니다 . 절편 , 산포 모수 및 영과잉 모수에 해당하는 추정값은 l_1 노름 계산에서 제외됩니다 . 다음 사항에 유의하십시오 .

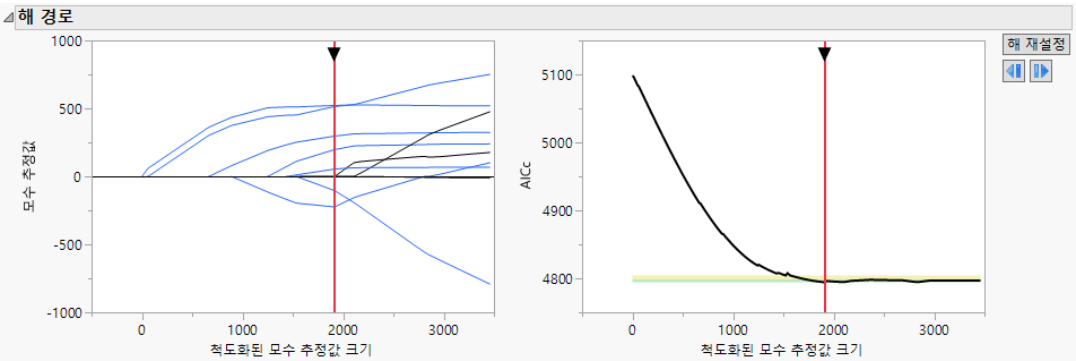
- l_1 노름 값이 큰 추정값은 MLE 에 가깝습니다 .
- l_1 노름 값이 작은 추정값은 벌점이 크게 부과됩니다 .
- l_1 노름이 감소하면 조정 모수 값이 증가합니다 .

JMP PRO 현재 모형 표시자

두 그림 모두 " 원래 예측 변수에 대한 모수 추정값 " 보고서에 표시된 해의 l_1 노름 값에서 빨간색 수직 실선이 추가됩니다. 두 그림 중 하나에서 빨간색 수직선 맨 위의 화살표를 드래그하면 벌점 크기가 변경되어 새 현재 모형을 나타낼 수 있습니다. 검증 그림에서도 그림의 아무 곳이나 클릭하여 모형을 변경할 수 있습니다. 빨간색 수직선을 드래그하여 새 모형을 나타내면 보고서의 결과가 현재 선택된 모형을 반영하도록 업데이트됩니다. 수직 파선은 최량 적합 모형에 유지됩니

다. 검증 그림 옆의 **해 재설정** 버튼을 클릭하면 빨간색 수직선 및 해당 결과가 초기 해로 돌아갑니다. 일부 검증 방법의 경우 비교 가능한 모형을 식별하는 영역이 검증 그림에 제공됩니다. 자세한 내용은 "비교 가능 모형 영역"(293 페이지)에서 확인하십시오.

그림 6.5 Diabetes.jmp 에 대한 해 경로 보고서 (AICc 검증을 사용한 Lasso 회귀)



해 경로 그림에 대한 자세한 내용은 "해 경로 그림"(292 페이지)에서 확인하십시오. 검증 그림에 대한 자세한 내용은 "검증 그림"(293 페이지)에서 확인하십시오.

JMP PRO 해 경로 그림

해 경로 그림에서 경로를 선택하여 "모수 추정값" 보고서에서 해당 항을 강조 표시할 수 있습니다. 이렇게 하면 데이터 테이블의 해당 열도 선택됩니다. 하나의 보고서에서 행을 선택하면 다른 보고서의 해당 행과 해 경로 그림의 해당 경로가 강조 표시됩니다. 여러 경로 또는 행을 선택하려면 Shift 키를 누르고 클릭합니다.

모수 추정값은 해 경로 그림의 세로 축을 사용하여 그림에 표시됩니다. 이러한 값은 **척도화된** 모수 추정값입니다. 중심화 및 척도화된 예측 변수로 표현된 모형에 대해 값이 파생됩니다("중심화 및 척도화된 예측 변수에 대한 모수 추정값"(294 페이지) 참조).

예측 변수의 수가 관측값 수보다 적으면 일반적으로 해 경로 그림에 추정값의 전체 범위(0 ~ MLE에 의해 주어진 벌점 없는 적합)가 표시됩니다. 그렇지 않으면 벌점이 부과되지 않은 해에 가까운 크기로 그림이 확장됩니다. 이 현상은 끝에서 두 번째 격자점에서 MLE 해까지의 이동이 너무 커서 끝에서 두 번째 격자점까지의 해에 대한 상세 정보가 분명하지 않을 때 발생합니다. 이 경우 MLE가 최종 해가 아니면 축이 끝에서 두 번째 격자점까지만 확장되도록 해 경로 그림의 척도가 재조정됩니다.

JMP PRO 해 ID

내부적으로 해 경로의 각 해에는 해 ID가 할당됩니다. 초기에 제공된 해가 아닌 다른 해를 선택하기 위해 조정 모수를 변경하면 해당 해 ID가 "스크립트 저장" 옵션으로 생성된 스크립트에 나타납니다. 해 ID는 `해 ID 설정(N)` 명령의 N 값입니다. 해 ID를 저장하면 스크립트를 실행할 때 선택한 해를 다시 생성할 수 있습니다.

JMP PRO 검증 그림

검증 그림에는 전체 조정 모수 값 또는 척도화된 모수 추정값 크기 값에 대한 모형의 적합 능력을 나타내는 통계량 그림이 표시됩니다. 표시된 통계량은 선택한 검증 방법에 따라 다릅니다. 표 6.3에는 각 검증 방법에 대해 표시되는 통계량이 나열되어 있습니다. 모든 검증 방법에서 값이 작을수록 좋습니다. K 폴드 및 Leave-One-Out 검증 방법과 값이 세 개 이상인 검증 열의 경우 표시되는 통계량은 전체 폴드에서 척도화된 음의 로그 가능도 값의 평균입니다.

표 6.3에서 **척도화된 -1*로그 가능도**는 음의 로그 가능도를 음의 로그 가능도가 계산되는 집합의 관측값 수로 나눈 값입니다.

표 6.3 검증 방법과 해당 검증 통계량 및 영역

검증 방법	검증 통계량	조정 모수 영역
K 폴드	전체 *K* 폴드에 대해 척도화된 -1* 로그 가능도 값의 평균	두 개
홀드백	척도화된 -1*로그 가능도	없음
Leave-One-Out	모든 폴드에 대해 척도화된 -1* 로그 가능도 값의 평균	두 개
BIC	훈련 데이터의 BIC	두 개
AICc	훈련 데이터의 AICc	두 개
ERIC	훈련 데이터의 ERIC	두 개
두 개 또는 세 개의 값이 있는 검증 열	척도화된 -1*로그 가능도	없음
K > 3개의 값이 있는 검증 열	전체 *K* 폴드에 대해 척도화된 -1* 로그 가능도 값의 평균	두 개

JMP PRO 비교 가능 모형 영역

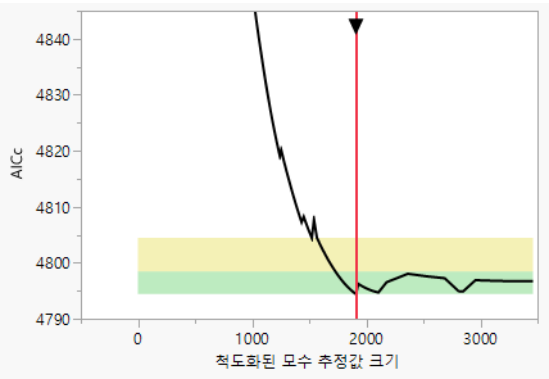
모형이 최적 모형으로 추정되지만 이 선택과 관련하여 불확실성이 존재할 수 있습니다. 경쟁 모형도 거의 적합할 수 있으며 유용한 정보를 포함할 수 있습니다. AICc, BIC, K 폴드 및 Leave-One-Out 검증 방법과 4개 이상의 값이 있는 검증 열의 경우 고려할 만한 경쟁 모형을 식별하는 영역이 검증 그림에 제공됩니다. 이 영역을 벗어나는 모형은 권장되지 않습니다. 자세한 내용은 Burnham and Anderson 연구 자료(2004) 및 Burnham et al. 연구 자료 (2011)에서 확인하십시오.

영역은 검증 통계량 값의 구간입니다. 이러한 영역은 전체 가로 축에 걸쳐 녹색 또는 노란색 직사각형으로 표시됩니다. 검증 통계량 값이 영역에 포함되면 모형이 해당 영역에 포함됩니다. 빨간

색 수직 실선을 드래그하여 영역 내의 해를 탐색할 수 있습니다. 자세한 내용은 "현재 모형 표시자"(291 페이지)에서 확인하십시오.

그림 6.6에서는 두 영역을 표시하기 위해 세로 축이 확장된 Diabetes.jmp에 대한 검증 그림을 보여 줍니다.

그림 6.6 Diabetes.jmp 에 대한 검증 그림 (AICc 검증을 사용한 Lasso 회귀)



BIC, AICc 및 ERIC 검증에 대한 영역

이러한 검증 방법의 경우 두 영역이 그림에 표시됩니다 . 최적 해를 위한 검증 BIC, AICc 및 ERIC 값을 V^{best} 로 나타냅니다 .

- 녹색 영역은 모형이 최적 모형과 유사하다는 강력한 증거가 있는 모형을 식별합니다 . 녹색 영역 구간은 [V^{best}, V^{best}+4] 입니다 .
- 노란색 영역은 모형이 최적 모형과 유사하다는 약한 증거가 있는 모형을 식별합니다. 노란색 영역 구간은 (V^{best}+4, V^{best}+10] 입니다 .

K 폴드 검증 , Leave-One-Out 검증 및 4 개 이상의 값이 있는 검증 열에 대한 영역

이러한 검증 방법의 경우 두 영역이 그림에 표시됩니다 . 최적 모형을 위한 해에서 각 검증 데이터 집합에 대해 척도화된 음의 로그 가능도 함수가 실행됩니다 . 이러한 값의 표준 오차를 L^{SE} 로 나타냅니다 . 최적 해에 대한 척도화된 음의 로그 가능도는 L^{best} 로 나타냅니다 .

- 녹색 영역은 모형이 최적 모형과 유사하다는 강력한 증거가 있는 모형을 식별합니다 . 녹색 영역 구간은 [L^{best}, L^{best}+L^{SE}] 입니다 .
- 노란색 영역은 모형이 최적 모형과 유사하다는 약한 증거가 있는 모형을 식별합니다. 노란색 영역 구간은 (L^{best}+L^{SE}, L^{best}+2.5*L^{SE}] 입니다 .

JMP PRO 중심화 및 척도화된 예측 변수에 대한 모수 추정값

"중심화 및 척도화된 예측 변수에 대한 모수 추정값" 보고서에서는 모형의 모든 모수에 대한 추정값 및 기타 결과를 제공합니다. 초기 테이블에는 모형의 예측 변수에 대한 계수가 포함됩니다.

추가 테이블에는 척도, 산포 또는 영과잉 모수와 같은 다른 모형 모수가 포함됩니다. 자세한 내용은 "분포"(270 페이지)에서 확인하십시오. 두 테이블 모두 동일한 결과 열을 포함합니다.

팁 : "중심화 및 척도화된 예측 변수에 대한 모수 추정값" 보고서에서 항을 클릭하여 해 경로 그림에서 해당 경로를 강조 표시할 수 있습니다. 데이터 테이블의 해당 열도 선택됩니다. 이 방법은 추가 분석을 실행할 때 유용합니다. 여러 행을 선택하려면 Shift 키를 누르고 항을 클릭합니다.

일반화 회귀 분석법의 모든 적합에 대해 모든 예측 변수는 평균 0 으로 중심화되고 제곱합이 1 과 같도록 척도화됩니다.

- 각 관측값에서 평균을 뺍니다.
- 그런 다음 각 차이를 평균과의 차이 제곱합의 제곱근으로 나눕니다.

이렇게 하면 적용된 벌점과 관련하여 모든 예측 변수가 동등한 관계가 됩니다.

참고 : 시작 창에서 "절편 없음" 옵션을 선택하면 예측 변수가 중심화 및 척도화되지 않습니다.

"중심화 및 척도화된 예측 변수에 대한 모수 추정값" 보고서에는 중심화 및 척도화된 예측 변수로 표현된 모형의 모수 추정값이 제공됩니다. 추정값은 지정한 검증 방법에 의해 결정됩니다. 이 추정값은 해 경로 그림에 빨간 수직선으로 표시됩니다.

이 보고서는 다음 정보를 제공합니다.

항 모형 항 목록입니다. "고급 제어" 옵션을 사용하여 모형에 강제로 적용한 항 옆에는 "강제 적용됨" 이 표시됩니다.

추정값 중심화 및 척도화된 모형 항에 해당하는 모수 추정값입니다.

표준 오차 추정값의 표준 오차입니다. 이 값은 M- 추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009) 을 사용하여 구합니다.

Wald 카이제곱 모수가 0 인지 여부에 대한 Wald 검정의 카이제곱 값입니다.

Prob > ChiSquare Wald 검정의 p 값입니다.

95% 하한 모수에 대한 95% 신뢰 구간의 하한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정" 을 선택하여 α 수준을 변경할 수 있습니다.

95% 상한 모수에 대한 95% 신뢰 구간의 상한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정" 을 선택하여 α 수준을 변경할 수 있습니다.

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 사용 가능) 모형 항이 충족하는 선형 함수입니다.

JMP PRO 원래 예측 변수에 대한 모수 추정값

"원래 예측 변수에 대한 모수 추정값" 보고서에서는 모형의 모든 모수에 대한 추정값 및 기타 결과를 제공합니다. 초기 테이블에는 모형의 예측 변수에 대한 계수가 포함됩니다. 추가 테이블에

는 척도, 산포 또는 영과잉 모수와 같은 다른 모형 모수가 포함됩니다. 자세한 내용은 "분포"(270 페이지)에서 확인하십시오. 두 테이블 모두 동일한 결과 열을 포함합니다.

팁 : "원래 예측 변수에 대한 모수 추정값" 보고서에서 항을 클릭하여 해 경로 그림에서 해당 경로를 강조 표시할 수 있습니다. 데이터 테이블의 해당 열도 선택됩니다. 이 방법은 추가 분석을 실행할 때 유용합니다. 여러 행을 선택하려면 Shift 키를 누르고 항을 클릭합니다.

"원래 예측 변수에 대한 모수 추정값" 보고서에는 원래(중심화 및 척도화되지 않음) 예측 변수로 표현된 모형의 모수 추정값이 제공됩니다.

이 보고서는 다음 정보를 제공합니다.

항 모형 항 목록입니다. "고급 제어" 옵션을 사용하여 모형에 강제로 적용한 항 옆에는 "강제 적용됨" 이 표시됩니다.

추정값 원래 측정값으로 표현된 주어진 모형 항에 해당하는 모수 추정값입니다.

표준 오차 추정값의 표준 오차입니다. 이 값은 M- 추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009) 을 사용하여 구합니다.

Wald 카이제곱 모수가 0 인지 여부에 대한 Wald 검정의 카이제곱 값입니다.

Prob > ChiSquare Wald 검정의 p 값입니다.

95% 하한 모수에 대한 95% 신뢰 구간의 하한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정" 을 선택하여 α 수준을 변경할 수 있습니다.

95% 상한 모수에 대한 95% 신뢰 구간의 상한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정" 을 선택하여 α 수준을 변경할 수 있습니다.

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 사용 가능) 모형 항이 충족하는 선형 함수입니다.

VIF (정규 분포인 경우에만 사용 가능하며, 모수 추정값 테이블에서 마우스 오른쪽 버튼을 클릭하고 열 > VIF 를 선택한 경우에만 표시됨) 모형의 각 항에 대한 VIF(분산 팽창 계수) 입니다. VIF 값이 높으면 모형의 항 사이에 공선성 문제가 있음을 나타냅니다.

i^{th} 항 x_i 의 VIF 는 다음과 같이 계산됩니다.

$$VIF_i = \frac{1}{1 - R_i^2}$$

여기서 $R_i{}^2$ 은 다른 설명 변수의 함수로 나타낸 x_i 회귀에 대한 R^2 입니다.

JMP PRO 활성 모수 추정값

"활성 모수 추정값" 보고서에서는 "원래 예측 변수에 대한 모수 추정값" 보고서의 일부를 제공합니다. "활성 모수 추정값" 보고서에는 0 이 아닌 모수만 표시됩니다.

JMP PRO 효과 검정

" 효과 검정 " 보고서에서는 다음 정보를 제공합니다 .

소스 모형의 효과 목록입니다 .

모수 수 효과와 관련된 모수의 수입니다 .

DF Wald 카이제곱 검정의 자유도입니다 . 이 값은 모형의 효과와 관련된 0 이 아닌 모수 추정값의 수입니다 .

Wald 카이제곱 효과와 관련된 모든 모수가 0 인지 여부에 대한 Wald 검정의 카이제곱 값입니다 .

Prob > ChiSquare Wald 카이제곱 검정의 *p* 값입니다 .

정규 분포를 사용하고 추정 방법에 벌점이 없으면 보고서에 "Wald 카이제곱 " 과 "Prob > ChiSquare" 대신 다음 열이 나타납니다 .

제곱합 효과가 0 이라는 가설을 위한 제곱합입니다 .

F 비 효과가 0 인지 검정하기 위한 F 통계량입니다 . "F 비 " 는 효과의 평균 제곱을 오차 평균 제곱으로 나눈 비율입니다 . 오차의 평균 제곱은 효과 제곱합을 자유도로 나눈 값입니다 .

Prob > F 효과 검정의 *p* 값입니다 .

효과 계수가 0 으로 추정된 경우 다음 사항에 유의하십시오 .

- 효과의 자유도가 1 이면 해당 효과에 대한 행의 맨 오른쪽에 " 제거 " 라고 표시됩니다 .
- 효과의 자유도가 1 을 초과하는 경우 " 제거된 수준 " 이라고 표시되고 그 뒤에 모수 추정값이 0 인 항에 해당하는 수준 수가 나옵니다 .

JMP PRO 모형 적합 옵션

각 모형 적합 보고서에는 다음 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다 .

주의 : 시작 창에서 표현식 데이터 유형 또는 벡터 모델링 유형이 있는 열을 지정하면 플랫폼의 많은 옵션을 사용할 수 없습니다 .

회귀 보고서 지정된 모형 적합에 대해 표시되는 보고서를 사용자 정의할 수 있습니다 . 다음 보고서 중 " 중심화 및 척도화된 예측 변수에 대한 모수 추정값 " 보고서와 " 활성 모수 추정값 " 보고서를 제외한 모든 보고서가 기본적으로 표시됩니다 .

모형 요약 모형에 대한 규격 및 적합도 통계량 정보를 포함하는 " 모형 요약 " 보고서를 표시하거나 숨깁니다 . 이 옵션은 적용 가능한 모형에 대한 " 추정 상세 정보 " 보고서도 표시합니다 . 자세한 내용은 " 모형 요약 "(288 페이지) 및 " 추정 상세 정보 "(291 페이지) 에서 확인하십시오 .

해 경로 (최대 가능도 모형에는 사용 불가능) 해 경로 및 검증 경로 그림을 표시하거나 숨깁니다 . 자세한 내용은 " 해 경로 "(291 페이지) 에서 확인하십시오 .

중심화 및 척도화된 예측 변수에 대한 모수 추정값 중심화 및 척도화된 모수 추정값 테이블을 표시하거나 숨깁니다 . 자세한 내용은 " 중심화 및 척도화된 예측 변수에 대한 모수 추정값 "(294 페이지) 에서 확인하십시오 .

원래 예측 변수에 대한 모수 추정값 데이터의 원래 척도로 표현된 모수 추정값 테이블을 표시하거나 숨깁니다 . 자세한 내용은 " 원래 예측 변수에 대한 모수 추정값 "(295 페이지) 에서 확인하십시오 .

활성 모수 추정값 (최대 가능도 또는 능형 회귀 모형에는 사용 불가능) 현재 선택된 모형에 대한 활성 또는 0 이 아닌 모수 추정값 테이블을 표시하거나 숨깁니다 .

해 경로 요약 표시 (최대 가능도 또는 능형 회귀 모형에는 사용 불가능) 해 경로 그림과 검증 경로 그림에서 활성 집합이 변경되는 점에 대한 적합 통계량 테이블이 포함된 보고서를 표시하거나 숨깁니다 . 사용 가능한 통계량은 추정 방법에 따라 다릅니다 . BIC 검증을 사용한 정규 Lasso 회귀 모형에 사용할 수 있는 조건부 모형 확률에 대한 자세한 내용은 Hu et al. 연구 자료 (2019) 에서 확인하십시오 . BIC, AICc 또는 ERIC 검증을 지정하면 BIC 및 AICc 열의 셀에 검증 그림과 동일한 방식으로 색상이 적용됩니다 . 자세한 내용은 " 비교 가능 모형 영역 "(293 페이지) 에서 확인하십시오 .

효과 검정 각 효과에 대한 검정을 표시하거나 숨깁니다 . 주어진 효과에 대한 효과 검정은 해당 효과와 관련된 모든 모수가 0 이라는 귀무가설을 검정합니다 . 명목형 또는 순서형 효과에는 수준 수에 따라 관련된 모수가 여러 개 있을 수 있습니다 . 이러한 효과에 대한 효과 검정은 관련된 모든 모수가 0 인지 여부를 검정합니다 . 다항 분포를 지정한 경우에는 전체 반응 수준에 대해 효과가 결합됩니다 . 자세한 내용은 " 효과 검정 "(297 페이지) 에서 확인하십시오 .

예측 표현식 표시 추정된 모형에 대한 방정식을 포함하는 " 예측 표현식 " 보고서를 표시하거나 숨깁니다 . 예는 " 표준 최소 제곱 모형 " 장의 " 예측 표현식 표시 "(107 페이지) 에서 확인하십시오 .

0 이 아닌 항 선택 보고서에서 계수가 0 이 아닌 항을 강조 표시합니다 . 또한 데이터 테이블에서 관련된 모든 열을 선택합니다 . " 추정 방법 " 에서 " 능형 회귀 " 를 선택한 경우에는 이 옵션을 사용할 수 없습니다 .

0 인 항 선택 보고서에서 계수가 0 인 항을 강조 표시합니다 . 또한 데이터 테이블에서 관련된 모든 열을 선택합니다 . " 추정 방법 " 에서 " 능형 회귀 " 를 선택한 경우에는 이 옵션을 사용할 수 없습니다 .

활성 집합 다시 시작 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 모수 추정값이 0 이 아닌 항을 기반으로 한 항 집합이 " 모형 효과 생성 " 목록에 포함된 " 모형 적합 " 시작 창을 여는 옵션이 포함되어 있습니다 . 이러한 항은 **활성** 효과입니다 . 시작 창의 다른 모든 규격은 원래 분석에 사용된 것과 같습니다 .

참고 : 기준 변수가 포함된 보고서에서 " 활성 집합 다시 시작 " 옵션 중 하나를 선택하는 경우 기준 변수는 " 모형 적합 " 시작 창에 추가되지 않습니다 .

활성 효과를 사용하여 다시 시작 활성 효과만 사용하여 " 모형 효과 생성 " 목록을 채웁니다 .

활성 주효과 및 2 차 요인 적합 다시 시작 활성 효과를 사용하여 생성된 2 차 요인으로 " 모형 효과 생성 " 목록을 채웁니다 .

활성 주효과 및 3 차 요인 적합 다시 시작 활성 효과를 사용하여 생성된 3 차 요인으로 " 모형 효과 생성 " 목록을 채웁니다 .

활성 주효과 및 완전 요인 적합 다시 시작 활성 효과를 사용하여 생성된 완전 요인으로 " 모형 효과 생성 " 목록을 채웁니다 .

활성 주효과 및 2 차 다항식 적합 다시 시작 활성 효과를 사용하여 생성된 2 차 다항식으로 " 모형 효과 생성 " 목록을 채웁니다 .

활성 주효과 및 3 차 다항식 적합 다시 시작 활성 효과를 사용하여 생성된 3 차 다항식으로 " 모형 효과 생성 " 목록을 채웁니다 .

활성 주효과 및 반응 표면 모형 다시 시작 활성 효과를 사용하여 생성된 반응 표면 모형으로 "모형 효과 생성 " 목록을 채웁니다 .

비활성 경로 숨기기 현재 활성화되지 않은 경로는 흐리게 나타나도록 해 경로 모수 추정값 그림에서 비활성 경로의 투명도를 조정합니다 .

승산비 (이항 분포가 지정되고 모형에 절편이 포함된 경우에만 사용 가능하며 , 벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 범주형 예측 변수에 대한 승산비와 연속형 예측 변수에 대한 단위 승산비 및 범위 승산비가 포함된 보고서를 표시합니다 . **승산비**는 두 사건에 대한 승산 비율입니다 . 한 사건의 **승산**은 관심 사건이 발생할 확률 대 발생하지 않을 확률의 비율입니다 . 관심 사건은 " 모형 적합 " 시작 창의 " 목표 수준 " 에 정의됩니다 .

각 범주형 예측 변수에 대해 승산비 보고서가 나타납니다 . 범주형 모형 항의 모든 수준 조합에 대해 승산비가 표시됩니다 .

연속형 예측 변수가 있는 경우 다음과 같은 두 개의 추가 보고서가 나타납니다 .

– 단위 승산비 보고서 . **단위 승산비**는 연속형 모형 항의 1 단위 변화에 대해 계산됩니다 .

– 범위 승산비 보고서 . **범위 승산비**는 연속형 모형 항의 전체 범위에 대해 계산됩니다 .

승산비 보고서의 신뢰 구간은 Wald 기반 구간입니다 . 모형 항에 대한 승산비는 해당 모형 항이 고차 효과에 포함되지 않은 경우에만 의미가 있습니다 .

참고 : 모형에 교호작용이 있으면 " 다중 비교 " 옵션을 사용하여 승산비를 구할 수 있습니다 . 자세한 내용은 " 다중 비교 "(300 페이지) 에서 확인하십시오 .

발생률 비율 (Poisson 또는 음이항 분포가 지정되고 모형에 절편이 포함된 경우에만 사용 가능) 범주형 예측 변수에 대한 발생률 비율과 연속형 예측 변수에 대한 단위 발생률 비율 및 범위 발생률 비율이 포함된 보고서를 표시합니다 . **발생률 비율**은 두 사건에 대한 발생률 비율입니다 . 모형 항의 **발생률**은 주어진 기간 동안 새로 발생하는 사건 수입니다 .

각 범주형 예측 변수에 대해 발생률 비율 보고서가 나타납니다 . 범주형 모형 항의 모든 수준 조합에 대해 발생률 비율이 표시됩니다 .

연속형 예측 변수가 있는 경우 다음과 같은 두 개의 추가 보고서가 나타납니다 .

- 단위 발생률 비율 보고서 . **단위 발생률 비율**은 연속형 모형 항의 1 단위 변화에 대해 계산됩니다 .
- 범위 발생률 비율 보고서 . **범위 발생률 비율**은 연속형 모형 항의 전체 범위에 대해 계산됩니다 .

발생률 비율 보고서의 신뢰 구간은 Wald 기반 구간입니다 . 모형 항에 대한 발생률 비율은 해당 모형 항이 고차 효과에 포함되지 않은 경우에만 의미가 있습니다 .

위험 비율 (Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 범주형 예측 변수에 대한 위험 비율과 연속형 예측 변수에 대한 단위 위험 비율 및 범위 위험 비율이 포함된 보고서를 표시합니다 . **위험 비율**은 두 사건에 대한 위험 비율입니다 . 시간 t 에서 사건의 **위험 비율**은 사건이 시간 t 까지 생존했다고 가정할 때 추가 시간 동안 생존하지 못할 조건부 확률입니다 .

각 범주형 예측 변수에 대해 위험 비율 보고서가 나타납니다 . 범주형 모형 항의 모든 수준 조합에 대해 위험 비율이 표시됩니다 .

연속형 예측 변수가 있는 경우 다음과 같은 두 개의 추가 보고서가 나타납니다 .

- 단위 위험 비율 보고서 . **단위 위험 비율**은 연속형 모형 항의 1 단위 변화에 대해 계산됩니다 .
- 범위 위험 비율 보고서 . **범위 위험 비율**은 연속형 모형 항의 전체 범위에 대해 계산됩니다 .

위험 비율 보고서의 신뢰 구간은 Wald 기반 구간입니다 . 모형 항에 대한 위험 비율은 해당 모형 항이 고차 효과에 포함되지 않은 경우에만 의미가 있습니다 .

추정값 공분산 모수 추정값의 공분산을 보여 주는 행렬을 표시합니다 . 이러한 값은 M- 추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009) 을 사용하여 계산됩니다 . 공분산 행렬은 0 인 항을 포함하지 않습니다 .

추정값 상관계수 모수 추정값의 상관계수를 보여 주는 행렬을 표시합니다 . 이러한 값은 M- 추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009) 을 사용하여 계산됩니다 . 상관 행렬은 0 인 항을 포함하지 않습니다 .

역추정 예측 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) Y 및 다른 X 변수에 대한 특정 값이 주어졌을 때 X 값을 예측합니다 . 이 옵션은 연속형 변수를 예측하는 데만 사용할 수 있습니다 . 역추정 예측에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 역추정 예측 "(131 페이지) 에서 확인하십시오 .

다중 비교 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형 또는 범주형 예측 변수가 포함되지 않은 모형에는 사용 불가능) " 다중 비교 " 시작 창을 표시합니다 . " 다중 비교 " 시작 창과 보고서에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 다중 비교 "(117 페이지) 에서 확인하십시오 . 다중 비교는 선형 예측 변수 척도에서 수행됩니다 . 이항 분포가 지정되면 승산비에 대해 다중 비교가 수행됩니다 . Poisson 분포가 지정되면 발생률 비율에 대해 다중 비교가 수행됩니다 . Cox 비례 위험 분포가 지정되면 위험 비율에 대해 다중 비교가 수행됩니다 .

혼동 행렬 (이항 , 다항 또는 순서형 로지스틱 분포가 지정된 경우에만 사용 가능) 실제 반응 수준과 예측 반응 수준을 테이블 형식으로 생성하는 행렬을 표시합니다 . 적합한 모형은 예측 반응 수준이 실제 반응 수준과 동일해야 합니다 . 혼동 행렬을 사용하면 예측 반응이 실제 반응에 잘 맞는지 평가할 수 있습니다 . 검증을 사용한 경우 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 혼동 행렬이 표시됩니다 .

확률 임계 설정 (이항 분포가 지정된 경우에만 사용 가능) 반응을 분류하기 위한 경계 확률을 지정합니다 . 기본적으로 관측값은 예측 확률이 0.5 를 초과할 경우 목표 수준으로 분류됩니다 . 임계값을 변경하여 목표 수준으로 분류하기 위한 경계를 0.5 이외의 값으로 지정합니다 . 혼동 행렬의 예측 비율은 지정된 임계값에 따른 분류를 반영하도록 업데이트됩니다 .

반응에 " 수익 행렬 " 열 특성이 있는 경우 수익 행렬에 의해 확률 임계 초기값이 결정됩니다 .

프로파일러 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 적합 모형을 탐색할 수 있는 다양한 프로파일러를 제공합니다 .

참고 : 행 수가 500 보다 작거나 같고 예측 변수의 수가 30 보다 작거나 같은 경우 해 경로 그림에서 현재 모형 표시자를 드래그하면 프로파일러 그림이 계속 업데이트됩니다 . 그렇지 않은 경우에는 마우스 버튼을 놓으면 그림이 업데이트됩니다 .

프로파일러 예측 프로파일러를 표시합니다 . 모수 추정값이 0 이고 , 계수가 0 이 아닌 교호작용 항에 포함되지 않은 예측 변수는 프로파일러에 나타나지 않습니다 . 예측 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

분포 프로파일러 (이항 또는 분위수 회귀 분포가 지정된 경우에는 사용 불가능) 예측 변수 및 반응의 누적 분포 함수 프로파일러를 표시합니다 . 반응은 맨 오른쪽 셀에 표시됩니다 .

분위수 프로파일러 (이항 또는 분위수 회귀 분포가 지정된 경우에는 사용 불가능) 예측 반응을 예측 변수의 함수 및 누적 분포 함수의 분위수로 보여 주는 프로파일러를 표시합니다 . 분위수는 " 확률 " 이라고 하며 맨 오른쪽 셀에 표시됩니다 .

생존 프로파일러 (정규 , 지수 , Weibull, 로그 정규 또는 Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 생존 함수를 예측 변수 및 반응의 함수로 보여 주는 프로파일러를 표시합니다 . 반응은 맨 오른쪽 셀에 표시됩니다 .

위험 프로파일러 (정규 , 지수 , Weibull, 로그 정규 또는 Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 위험 비율을 예측 변수 및 반응의 함수로 보여 주는 프로파일러를 표시합니다 . 반응은 맨 오른쪽 셀에 표시됩니다 .

사용자 검정 사용자 가설을 검정할 수 있는 " 사용자 검정 " 보고서를 표시합니다 . 모형에 해 경로가 있는 경우 해를 업데이트하면 사용자 검정 결과가 업데이트됩니다 . 사용자 검정에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 사용자 검정 "(115 페이지) 에서 확인하십시오 . " 사용자 검정 " 의 빨간색 삼각형 메뉴에는 사용자 검정 보고서를 제거하는 옵션이 포함되어 있습니다 .

진단 그림 현재 모형이 얼마나 적합한지 평가하는 데 도움이 되는 다양한 그림을 제공합니다 . 검증 열을 지정하거나 " 검증 방법 " 에서 "K 폴드 ", " 홀드백 " 또는 "Leave-One-Out" 을 선택한 경우 아래 옵션을 사용하여 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합 (해당하는 경우) 을 보거나 , 이러한 데이터 집합에 대해 별도의 그림을 생성할 수 있습니다 . "K 폴드 " 또는 "Leave-One-Out" 을 선택한 경우 그림은 예측 오차를 최적화하는 검증 데이터 집합 및 이에 대한 훈련 데이터 집합에 해당합니다 . 자세한 내용은 "K 폴드 "(286 페이지) 에서 확인하십시오 .

참고 : 해 경로 그림에서 현재 모형 표시자를 드래그하면 모든 진단 그림이 계속 업데이트됩니다 .

진단 번들 (이항 , 다항 , 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정된 경우에는 사용 불가능) 잔차 대 예측값 그림 , 잔차 대 행 번호 그림 , 잔차 히스토그램 및 관측된 반응보다 더 큰 반응이 관측될 확률 히스토그램을 포함하는 네 개의 그래프를 표시합니다 .

모든 관측값을 사용하여 그래프가 생성됩니다 . 검증 열을 사용하거나 " 검증 방법 " 에서 "K 폴드 ", " 홀드백 " 또는 "Leave-One-Out" 을 선택한 경우 체크박스를 사용하여 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합 (해당하는 경우) 을 선택할 수 있습니다 . 데이터 테이블에서 이러한 데이터 집합에 해당하는 행이 선택되고 그래프에서 해당 점과 영역이 강조 표시됩니다 . 이 옵션을 사용하여 모형 적합이 전체 데이터 집합에서 유사한지 여부를 판별할 수 있습니다 .

" 더 큰 반응이 관측될 적합 확률 " 히스토그램은 모형의 적합도를 평가하는 데 도움이 됩니다 . 다음과 같이 분포에 따라 다른 기준이 적용됩니다 .

- 영과잉 분포와 분위수 회귀를 제외한 다른 분포의 경우 " 올바른 " 모형은 거의 균등한 값 분포를 표시해야 합니다 .
- 영과잉 분포의 경우 히스토그램은 0 에 점질량을 표시하고 다른 곳에서는 거의 균등한 분포를 표시해야 합니다 .
- 분위수 회귀의 경우 히스토그램은 지정된 분위수 왼쪽에는 거의 균등한 값 분포를 표시하고 지정된 분위수 오른쪽에는 약간 더 높은 값을 거의 균등한 분포로 표시해야 합니다 .

기준 생존 및 위험 그림 (Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 반응 변수에 대한 기준 비례 위험 함수의 생존 및 위험 함수를 보여 주는 " 기준 생존 및 위험 " 그림을 표시합니다 . 아래에는 그림에 표시된 값이 포함된 테이블이 있습니다 .

참고 : Cox 비례 위험 분포가 지정된 경우 " 진단 그림 " 에 " 기준 생존 및 위험 그림 " 옵션만 있습니다 .

ROC 곡선 (이항 , 다항 또는 순서형 로지스틱 분포가 지정된 경우에만 사용 가능) ROC(Receiver Operating Characteristic) 곡선을 표시합니다 . 검증을 사용한 경우 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 ROC 곡선이 표시됩니다 .

ROC 곡선은 반응 수준을 올바르게 분류하기 위한 적합 확률 능력을 측정합니다 . 주대각과 곡선의 거리가 멀수록 적합도가 높습니다 . ROC 곡선에 대한 소개는 **기본 분석**의 에서 확인하십시오 .

반응 수준이 세 개 이상이면 ROC 곡선 그림에 각 반응 수준에 대한 ROC 곡선이 표시됩니다 . 주어진 반응 수준에 대해 이 곡선은 해당 수준으로 올바르게 분류하기 위한 ROC 곡선입니다 . ROC 곡선에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

Lift 곡선 (이항 , 다항 또는 순서형 로지스틱 분포가 지정된 경우에만 사용 가능) 모형에 대한 Lift 곡선을 표시합니다 . 검증을 사용한 경우 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 Lift 곡선이 표시됩니다 .

Lift 곡선은 적합 확률이 감소할 때 반응 수준이 얼마나 효과적으로 분류되는지 보여 줍니다 . 적합 확률은 가로 축을 따라 내림차순으로 표시됩니다 . 적합 확률의 수직 좌표는 해당 확률 이상에 대한 올바른 분류 비율을 전체 올바른 분류 비율로 나눈 값입니다 . 적합 확률이 임계값을 초과하는 관측값만 선택한 경우 Lift 곡선을 사용하여 대부분의 관측값을 올바르게 분류할 수 있는지 확인할 수 있습니다 .

반응 수준이 세 개 이상이면 Lift 곡선 그림에 각 반응 수준에 대한 Lift 곡선이 표시됩니다 . 주어진 반응 수준에 대해 이 곡선은 해당 수준으로 올바르게 분류하기 위한 Lift 곡선입니다 . Lift 곡선에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

실제값 대 예측값 그림 (이항 , 다항 , 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정된 경우에는 사용 불가능) 세로 축에 실제 Y 값을 표시하고 가로 축에 예측 Y 값을 표시합니다 . 검증을 사용한 경우 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다 .

잔차 대 예측값 그림 (이항 , 다항 , 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정된 경우에는 사용 불가능) 세로 축에 잔차를 표시하고 가로 축에 예측 Y 값을 표시합니다 . 검증을 사용한 경우 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다 .

예측 변수별 잔차 도표화 (이항 , 다항 , 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정되거나 , 벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 모형의 각 예측 변수에 대해 세로 축에 잔차를 표시하고 가로 축에 예측 변수 값을 표시합니다 . 모형의 예측 변수에 대해 각각 하나씩 그림이 있습니다 . 검증을 사용한 경우 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다 .

정규 분위수 그림 (정규 분포가 지정되고 중도절단이 없는 경우에만 사용 가능) 세로 축에 정규 분위수를 표시하고 가로 축에 표준화 잔차를 표시합니다 . 검증을 사용한 경우 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다 .

열 저장 적합 모형을 기반으로 한 열을 데이터 테이블에 저장할 수 있습니다 . " 분포 " 에서 "Cox 비례 위험 " 을 선택한 경우 사용할 수 있는 옵션에 대한 내용은 "Cox 비례 위험 모형의 열 저장 옵션 "(305 페이지) 에서 확인하십시오 . 다른 모든 분포의 경우 다음 열을 데이터 테이블에 저장할 수 있습니다 .

함수 예측 계산식 저장 (반응 열에 "FDE FPC 수 " 열 특성이 포함된 경우에만 사용 가능) 각 함수 주성분에 대한 예측 계산식이 포함된 각 FDE 주성분 반응 열을 데이터 테이블에 저장합니다 . 이 옵션은 함수 데이터 탐색기 플랫폼의 고유 함수 열과 예측 계산식의 선형 조합인 모형 예측 계산식도 추가합니다 . 모형 예측 계산식을 사용하여 원래 반응을 프로파일링할 수 있는 스크립트가 데이터 테이블에 추가됩니다 . 반응은 FDE 주성분 반응 열의 "FDE 출력 " 열 특성에 지정됩니다 . 함수 주성분에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

참고 : " 함수 예측 계산식 저장 " 옵션은 보고서 창에서 모든 FDE 주성분 반응에 대한 계산식 열을 저장합니다 . 단일 반응에 여러 모형이 적합된 경우 각 반응의 최종 모형을 사용하여 해당 반응에 대한 예측 계산식을 생성합니다 .

예측 계산식 저장 관측된 (표준화되지 않음) 데이터 값이 주어졌을 때 예측 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 예측 계산식은 0 인 항을 포함하지 않습니다 . 평균 계산식에 대한 자세한 내용은 "분포에 대한 통계 상세 정보"(309페이지)에서 확인하십시오.

반응 열이 범주형인 경우 이 옵션은 각 반응 수준에 대한 확률 열과 가장 확률이 높은 반응이 포함된 열을 생성합니다 . 가장 확률이 높은 반응 열에는 모형에 따른 확률이 가장 높은 수준이 포함됩니다 . 확률 임계값이 0.5 가 아닌 경우 이 옵션은 확률 임계값을 기반으로 가장 확률이 높은 반응을 포함하는 추가 열을 생성합니다 .

평균 신뢰 구간 평균 반응의 95% 신뢰 하한 및 상한이 포함된 두 열을 데이터 테이블에 저장합니다 .

참고 : " 모형 적합 " 창의 " 모형 규격 " 빨간색 삼각형 메뉴에서 " 유의 수준 설정 " 을 선택하여 신뢰 구간의 α 수준을 변경할 수 있습니다 .

예측값의 표준 오차 예측 평균 반응의 표준 오차가 포함된 열을 데이터 테이블에 저장합니다 .

잔차 계산식 저장 Y 에서 예측 계산식을 뺀 형식으로 제공되는 잔차의 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 잔차 계산식은 0 인 항을 포함하지 않습니다 . " 분포 " 에서 " 이항 " 을 선택한 경우에는 사용할 수 없습니다 .

분산 계산식 저장 예측 분산에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 예측 분산은 선택한 분포의 분산 계산식을 사용하여 계산됩니다 . 연결 함수에 관련된 모수의 값은 추정된 선형 성분에 역 연결 함수를 적용하여 추정됩니다 . 다른 모수는 모수 추정값으로 대체됩니다 . 분산 계산식에 대한 자세한 내용은 "분포에 대한 통계 상세 정보"(309페이지)에서 확인하십시오 . " 분포 " 에서 " 이항 " 을 선택한 경우에는 사용할 수 없습니다 .

선형 예측 변수 저장 설계 행렬과 모수 추정값 벡터의 곱을 구하는 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 일반적으로 이것을 $\mathbf{X}\beta$ 라고 합니다 . 이 계산식은 0 인 항을 포함하지 않습니다 .

검증 열 저장 (K 폴드 , 홀드백 또는 Leave-One-Out 검증 방법이 지정된 경우에만 사용 가능) 폴드에 행 할당을 설명하는 열을 저장합니다 . K 폴드의 경우 행이 할당된 폴드가 열에 나열됩니다 . 홀드백의 경우 각 행은 훈련 데이터 집합 또는 검증 데이터 집합에 속하는 것으로 식별됩니다 . Leave-One-Out 의 경우 행 값은 해당 행이 제외되는 순서를 나타냅니다 .

참고 : 시작 창에서 검증 열을 선택한 경우 " 검증 열 저장 " 옵션이 나타나지 않습니다 .

분포 계산식 저장 (이항 또는 분위수 회귀 분포가 지정된 경우에는 사용 불가능) 누적 분포 함수에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다 .

생존 계산식 저장 (연속형 분포가 지정된 경우에만 사용 가능) 관측된 시간에서의 생존 확률에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다. 생존 함수는 1 에서 누적 분포 함수를 뺀 값과 같습니다.

시뮬레이션 계산식 저장 적합 모형에 대한 추정 모수를 사용하여 시뮬레이션 값을 생성하는 계산식이 포함된 열을 데이터 테이블에 저장합니다. 이 열은 " 시뮬레이션 " 유틸리티에서 " 스위치 인할 열 " 로 사용할 수 있습니다. 자세한 내용은 **기본 분석**의 에서 확인하십시오.

Cook 거리 영향력 (정규 분포 및 표준 최소 제곱 추정 방법이 지정된 경우에만 사용 가능) Cook 거리 영향력 통계량의 값이 포함된 열을 데이터 테이블에 저장합니다.

해트 값 (정규 분포 및 표준 최소 제곱 추정 방법이 지정된 경우에만 사용 가능) $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ 의 주대각 요소가 포함된 열을 데이터 테이블에 저장합니다. 이러한 값을 hat 값이라고도 합니다.

예측 계산식 게시 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다. 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오.

Cox 비례 위험 모형의 열 저장 옵션

생존 계산식 저장 관측된 시간에서의 생존 확률에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다.

Cox Snell 잔차 계산식 저장 Cox-Snell 잔차에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다. Cox-Snell 잔차는 양수입니다. Cox-Snell 잔차에 대한 자세한 내용은 Meeker and Escobar 연구 자료 (1998, sec. 17.6.1) 에서 확인하십시오.

Martingale 잔차 계산식 저장 Martingale 잔차에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다. Martingale 잔차는 개인에 대해 관측된 사건 수와 조건부 기대 사건 수 간의 차이로 정의됩니다. Martingale 잔차의 평균은 0 이고 범위는 음의 무한대와 1 사이입니다. 자세한 내용은 Fleming and Harrington 연구 자료 (1991) 에서 확인하십시오.

선형 예측 변수 저장 설계 행렬과 모수 추정값 벡터의 곱을 구하는 계산식이 포함된 열을 데이터 테이블에 저장합니다. 일반적으로 이것을 $\mathbf{X}\beta$ 라고 합니다. 이 계산식은 0 인 항을 포함하지 않습니다.

적합 제거 적합 보고서를 제거합니다.

JMP PRO 일반화 회귀 분석법에 대한 통계 상세 정보

- " 추정 방법에 대한 통계 상세 정보 "
- " 고급 제어에 대한 통계 상세 정보 "
- " 분포에 대한 통계 상세 정보 "

JMP PRO 추정 방법에 대한 통계 상세 정보

벌점 회귀 방법은 벌점을 적용하여 회귀 계수에 편향을 도입합니다.

JMP PRO 능형 회귀

능형 회귀에서 회귀 계수에 l_2 벌점이 적용됩니다. 능형 회귀 계수 추정값은 다음과 같이 제공됩니다.

$$\hat{\beta}^{ridge} = \text{argmin}_{\beta}\left\{\sum_{i=1}^{N} -LogLikelihood(\beta;y_i) + \frac{\lambda}{2}\sum_{j=1}^{p}\beta_j^2\right\},$$

여기서 $\sum_{j=1}^{p}\beta_j^2$ 은 l_2 벌점, λ 는 조정 모수, N 은 행 수, p 는 변수 수입니다.

JMP PRO Dantzig 선택기

Dantzig 선택기에서 회귀 계수에 l_∞ 벌점이 적용됩니다. Dantzig 선택기에 대한 계수 추정값은 다음 기준을 충족합니다.

$$\min_{\beta}\left\|X^T(y - X\beta)\right\|_\infty \quad \text{subject to } \|\beta\|_1 \le t$$

여기서 $\|v\|_\infty$ 는 벡터 v 성분의 최대 절대값인 l_∞ 노름을 나타냅니다.

JMP PRO Lasso 회귀

Lasso 회귀에서 회귀 계수에 l_1 벌점이 적용됩니다. Lasso 회귀의 계수 추정값은 다음과 같이 제공됩니다.

$$\hat{\beta}^{lasso} = \text{argmin}_{\beta}\left\{\frac{1}{2}\sum_{i=1}^{N} -LogLikelihood(\beta;y_i) + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\right\},$$

여기서 $\sum_{j=1}^{p}|\beta_j|$ 는 l_1 벌점 , λ 는 조정 모수 , N 은 행 수 , p 는 변수 수입니다 .

JMP PRO Elastic Net

Elastic Net 은 l_1 벌점과 l_2 벌점을 결합합니다 . Elastic Net 의 계수 추정값은 다음과 같이 제공됩니다 .

$$\hat{\beta}^{enet} = \text{argmin}_{\beta}\left\{\sum_{i=1}^{N} -LogLikelihood(\beta;y_i) + \lambda\sum_{j=1}^{p}\left(\alpha|\beta_j| + \frac{(1-\alpha)}{2}\beta_j^2\right)\right\},$$

다음은 방정식에 사용되는 표기입니다 .

- $\sum_{j=1}^{p}|\beta_j|$ - l_1 벌점
- $\sum_{j=1}^{p}\beta_j^2$ - l_2 벌점
- λ - 조정 모수
- α - l_1 벌점과 l_2 벌점의 결합을 결정하는 모수
- N - 행 수
- p - 변수 수

팁 : Elastic Net 에서 단일 예측 변수에 대한 가변 α 및 λ 의 축소 효과를 보여 주는 두 개의 샘플 스크립트가 있습니다 . **도움말 > 샘플 데이터**를 선택하고 **샘플 스크립트 디렉터리 열기**를 클릭한 후 demoElasticNetAlphaLambda.jsl 또는 demoElasticNetAlphaLambda2.jsl 을 선택합니다 . 각 스크립트에는 스크립트 사용 방법 및 목적에 대한 설명이 포함되어 있습니다 .

JMP PRO 적응형 방법

적응형 Lasso 회귀 방법은 가중 벌점을 사용하여 일관된 계수 추정값을 제공합니다. l_1 벌점의 가중 형식은 다음과 같습니다 .

$$\sum_{j=1}^{p}\frac{|\beta_j|}{\tilde{\beta}_j},$$

여기서 $\tilde{\beta}_j$ 는 MLE 입니다 (MLE 가 있는 경우). MLE 가 없고 반응이 정규 분포인 경우 최소 제곱을 사용하여 추정이 수행되고 $\tilde{\beta}_j$ 는 일반화 역행렬을 사용하여 구한 해입니다 . 반응이 정규 분포가 아닌 경우 $\tilde{\beta}_j$ 는 능형 회귀 해입니다 .

적응형 Lasso 회귀의 경우 이 가중 l_1 벌점을 사용하여 $\hat{\beta}^{lasso}$ 계수를 결정합니다.

적응형 Elastic Net 은 이 가중 l_1 벌점을 사용하고 가중 l_2 벌점도 부과합니다 . 적응형 Elastic Net 의 가중 l_2 벌점은 다음과 같습니다 .

$$\sum_{j=1}^{p}\left(\frac{\beta_j}{\tilde{\beta}_j}\right)^2,$$

여기서 $\tilde{\beta}_j$ 는 MLE 입니다 (MLE 가 있는 경우). MLE 가 없고 반응이 정규 분포인 경우 최소 제곱을 사용하여 추정이 수행되고 $\tilde{\beta}_j$ 는 일반화 역행렬을 사용하여 구한 해입니다 . 반응이 정규 분포가 아닌 경우 $\tilde{\beta}_j$ 는 능형 회귀 해입니다 .

JMP PRO 고급 제어에 대한 통계 상세 정보

JMP PRO 격자

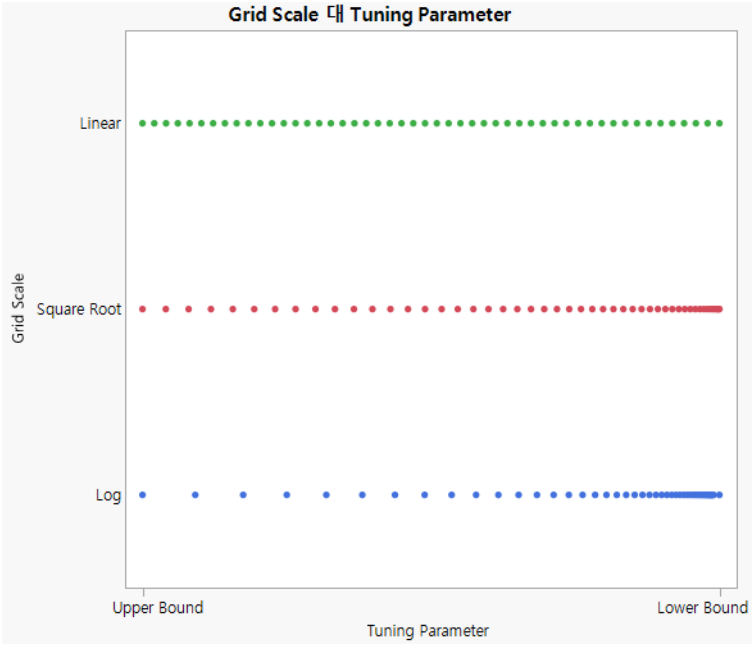
능형 회귀 및 Lasso 회귀에서 벌점 가능도를 가장 잘 최소화하는 조정 모수는 조정 모수 값의 격자를 검색하여 찾습니다. 이 값 격자는 조정 모수의 하한과 상한 사이에 있습니다. "고급 제어" 에서 격자점 수를 지정할 수 있습니다.

하한은 0.0001 로 설정되는 특수한 경우 외에는 0 입니다. 자세한 내용은 "조정 모수"(283 페이지) 에서 확인하십시오. 조정 모수의 하한이 0 이면 해에 벌점이 부과되지 않고 계수가 MLE 입니다. 상한은 모든 비절편 항이 0 인 가장 작은 값입니다.

하한과 상한 사이의 값 격자를 반복 검색하여 최적의 조정 모수 값을 결정합니다. 가능한 조정 모수 격자를 설정할 때 선형, 로그 및 제곱근이라는 세 가지 척도를 사용할 수 있습니다. 기본 격자 척도는 제곱근입니다.

벌점이 부과되지 않은 추정값과 이전 단계 사이의 차이가 큰 경우가 있습니다. 이러한 큰 차이로 인해 해 경로가 왜곡될 수 있습니다. 로그 척도는 큰 값이 거의 없는 작은 조정 모수 값을 검색하는 데 중점을 두는 반면, 선형 척도는 최소값에서 최대값까지 검색을 균등하게 분산시킵니다. 제곱근 척도는 다른 두 척도를 절충한 방법입니다. 그림 6.7 에서는 다양한 격자 척도를 보여 줍니다.

그림 6.7 조정 모수 격자 척도 옵션



JMP PRO 분포에 대한 통계 상세 정보

아래에서는 일반화 회귀 분석법에 의한 분포 적합을 모형 적합에 사용된 모수를 기준으로 구분하여 보여 줍니다. 설명을 통해 구체적으로 언급되어 있지는 않지만 일반화 회귀 분석법을 사용하면 이산형 분포에 대해 정수가 아닌 값을 지정할 수 있습니다.

JMP PRO 연속형 분포

정규 분포

$$f(y|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right], -\infty < y < \infty$$

$$E(Y) = \mu$$

$$Var(Y) = \sigma^2$$

Cauchy 분포

$$f(y|\mu,\sigma) = \left\{\pi\sigma\left[1+\left(\frac{y-\mu}{\sigma}\right)^2\right]\right\}^{-1}, -\infty < y < \infty$$

$E(Y) = \text{정의되지 않음}$

$Var(Y) = \text{정의되지 않음}$

t(5) 분포

$$f(y|\mu, \sigma) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi\sigma^2}} \left[1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-\left(\frac{\nu+1}{2}\right)}, -\infty < y < \infty, \nu = 5$$

$E(Y) = \mu$

$Var(Y) = \sigma^2 \frac{\nu}{\nu-2}$

지수 분포

$$f(y|\theta) = \frac{1}{\mu} \exp\left[-\frac{y}{\mu}\right], y > 0$$

$E(Y) = \mu$

$Var(Y) = \mu^2$

감마 분포

$$f(y|\mu, \sigma) = \frac{y^{(\mu/\sigma)-1} \exp[-y/\sigma]}{\Gamma[\mu/\sigma]\sigma^{\mu/\sigma}}, y > 0$$

$E(Y) = \mu$

$Var(Y) = \mu\sigma$

Weibull 분포

$$f(y|\mu, \sigma) = \frac{1}{y\sigma} \exp\left[\frac{\log(y)-\mu}{\sigma}\right] \exp\left\{-\exp\left[\frac{\log(y)-\mu}{\sigma}\right]\right\}, y > 0$$

$E(Y) = \exp(\mu)\Gamma[1+\sigma]$

$Var(Y) = \exp(2\mu)\{\Gamma[1+2\sigma] - (\Gamma[1+\sigma])^2\}$

로그 정규 분포

$$f(y|\mu,\sigma) = \frac{1}{y\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}[\log(y)-\mu]^2\right\}, y>0$$

$$E(Y) = \exp\left(\mu+\frac{\sigma^2}{2}\right)$$

$$Var(Y) = [\exp(\sigma^2)-1]\exp(2\mu+\sigma^2)$$

베타 분포

$$f(y|\mu,\sigma) = \frac{\Gamma[1/\sigma]}{\Gamma[\mu/\sigma]\Gamma[(1-\mu)/\sigma]}y^{\mu/\sigma-1}(1-y)^{\frac{(1-\mu)}{\sigma}-1}, y\in(0,1)$$

$$E(Y) = \mu$$

$$Var(Y) = \mu(1-\mu)\frac{\sigma}{\sigma+1}$$

JMP PRO 이산형 분포

이항 분포

$$f(y|n,p) = \binom{n}{y}p^y(1-p)^{n-y}, y = 0, 1, 2, \ldots, n$$

$$E(Y) = np$$

$$Var(Y) = np(1-p)$$

베타 이항 분포

$$f(y|n,p,\delta) = \binom{n}{y}\frac{\Gamma\left[\frac{1}{\delta}-1\right]\Gamma\left[y+p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-y+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-1+\frac{1}{\delta}\right]}, y = 0, 1, 2, \ldots, n$$

$$E(Y) = np$$

$$Var(Y) = np(1-p)[1+(n-1)\delta]$$

Poisson 분포

$$f(y|\lambda) = \frac{\lambda^y}{y!}\exp(-\lambda)\ , y = 0, 1, 2, \ldots$$

$$E(Y) = \lambda$$

$$Var(Y) = \lambda$$

음이항 분포

$$f(y|\mu, \sigma) = \frac{\Gamma[y + (1/\sigma)]}{\Gamma[y+1]\Gamma[1/\sigma]}\left[\frac{(\mu\sigma)^y}{(1+\mu\sigma)^{y+(1/\sigma)}}\right], y = 0, 1, 2, \ldots$$

$$E(Y) = \mu$$

$$Var(Y) = \mu + \sigma\mu^2$$

JMP PRO 영과잉 분포

영과잉 이항 분포

$$f(y|n, p, \pi) = \begin{cases} \pi + (1-\pi)(1-p)^n, \text{ for } y = 0 \\ (1-\pi)\binom{n}{y}p^y(1-p)^{n-y}, \text{ for } y = 1, 2, \ldots, n \end{cases}$$

$$E(Y) = np(1-\pi)$$

$$Var(Y) = (1-\pi)np(1-p) + n^2p^2(1-\pi)^2$$

영과잉 베타 이항 분포

$$f(y|n, p, \delta, \pi) = \begin{cases} \pi + (1-\pi)\dfrac{\Gamma\left[\frac{1}{\delta}-1\right]\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-1+\frac{1}{\delta}\right]}, \text{ for } y = 0 \\ (1-\pi)\binom{n}{y}\dfrac{\Gamma\left[\frac{1}{\delta}-1\right]\Gamma\left[y+p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-y+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right]\Gamma\left[n-1+\frac{1}{\delta}\right]}, y = 1, 2, \ldots, n \end{cases}$$

$$E(Y) = np(1-\pi)$$

$$Var(Y) = (1-\pi)np\{1+(1-p)[1+(n-1)\delta]\} - [np(1-\pi)]^2$$

영과잉 Poisson 분포

$$f(y|\lambda,\pi) = \begin{cases} \pi + (1-\pi)\exp[-\lambda] \text{ 단, } y = 0 \\ (1-\pi)\dfrac{\lambda^y}{y!}\exp[-\lambda], \text{ for } y = 1, 2, \ldots \end{cases}$$

$$E(Y) = (1-\pi)\lambda$$

$$Var(Y) = \lambda(1-\pi)(1+\lambda\pi)$$

영과잉 음이항 분포

$$f(y|\mu,\sigma,\pi) = \begin{cases} \pi + (1-\pi)(1+\mu\sigma)^{-(1/\sigma)}, \text{ for } y = 0 \\ (1-\pi)\dfrac{\Gamma[y+(1/\sigma)]}{\Gamma[y+1]\Gamma[1/\sigma]}\left[\dfrac{(\mu\sigma)^y}{(1+\mu\sigma)^{y+(1/\sigma)}}\right], \text{ for } y = 1, 2, \ldots \end{cases}$$

$$E(Y) = (1-\pi)\mu$$

$$Var(Y) = \mu(1-\pi)[1+\mu(\sigma+\pi)]$$

영과잉 감마 분포

$$f(y|\mu,\sigma,\pi) = \begin{cases} \pi \text{ , for } y = 0 \\ (1-\pi)\dfrac{\exp(-y/\sigma)}{\Gamma[\mu/\sigma]\sigma^{\mu/\sigma}y^{1-\mu/\sigma}} \text{ , for } y > 0 \end{cases}$$

$$E(Y) = \mu(1-\mu)$$

$$Var(Y) = \mu(1-\pi)(\sigma+\mu) - (1-\pi)^2\mu^2$$

7 장

일반화 회귀 예
정규화 기법을 사용하여 모형 생성

이 장에서는 일반화 회귀 플랫폼을 사용하는 몇 가지 모형에 대해 지침 자료를 이용한 예를 제공합니다.

목차

Poisson 일반화 회귀 예 317
이항 일반화 회귀 예 319
영과잉 Poisson 회귀 예 321
가로형 데이터에 대한 일반화 회귀 예 323

Poisson 일반화 회귀 예

Liver Cancer.jmp 샘플 데이터 테이블에는 136 명의 환자에 대한 간암 Node Count(결절 수) 값이 포함되어 있습니다 . 또한 BMI, Age(연령), Time(시간), Markers(표지자), Hepatitis(간염) 및 Jaundice(황달) 라는 6 가지 잠재적 관련 변수에 대한 측정값도 포함됩니다 . 이러한 변수는 데이터 테이블의 각 열에서 " 노트 " 열 특성에 설명되어 있습니다 . " 노트 " 열 특성을 보려면 열 이름을 마우스 오른쪽 버튼으로 클릭하고 " 열 정보 " 를 선택한 후 " 열 특성 " 아래의 " 노트 " 를 선택합니다 .

이 예에서는 6 개의 예측 변수를 사용하여 Node Count 에 대한 예측 모형을 개발합니다 . Node Count 는 Poisson 분포를 사용하여 모델링됩니다 .

1. **도움말 > 샘플 데이터 라이브러리** 를 선택하고 Liver Cancer.jmp 를 엽니다 .
2. **분석 > 모형 적합** 을 선택합니다 .
3. " 열 선택 " 목록에서 Node Count 를 선택하고 **Y** 를 클릭합니다 .
4. BMI ~ Jaundice 를 선택하고 **매크로 > 특정 차수까지의 요인** 을 클릭합니다 .

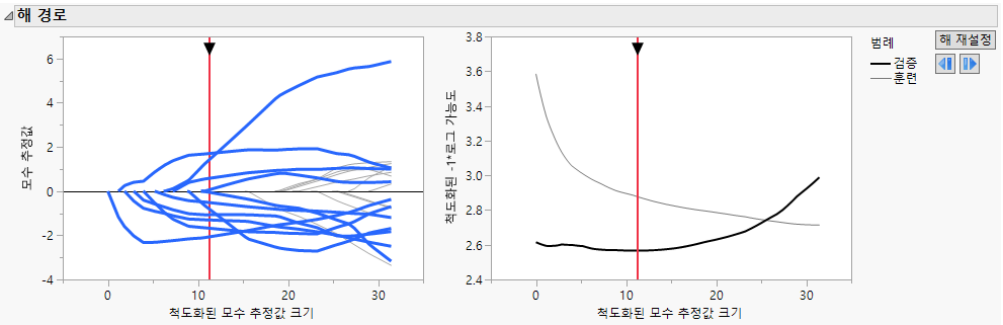
   이렇게 하면 차수 2(**차수** 상자의 기본값) 까지의 모든 항이 모형에 추가됩니다 .
5. " 열 선택 " 목록에서 Validation 을 선택하고 **검증** 을 클릭합니다 .
6. " 분석법 " 목록에서 **일반화 회귀** 를 선택합니다 .
7. " 분포 " 목록에서 **Poisson** 을 선택합니다 .
8. **실행** 을 클릭합니다 .

   " 모형 비교 " 보고서 , " 모형 시작 " 제어판 및 " 검증 열을 사용한 Poisson 최대 가능도 " 보고서가 포함된 " 일반화 회귀 " 보고서가 나타납니다 . 기본 추정 방법은 "Lasso 회귀 " 입니다 .
9. **적응형** 상자를 선택합니다 .
10. **시작** 을 클릭합니다 .
11. " 검증 열을 사용한 Poisson 적응형 Lasso 회귀 " 옆의 빨간색 삼각형을 클릭하고 **0 이 아닌 항 선택** 을 선택합니다 .

    그림 7.1 에는 해 경로가 나와 있습니다 . 계수가 0 이 아닌 항의 경로가 강조 표시됩니다 . 해가 MLE 에서 멀어질수록 해 경로가 그림의 오른쪽에서 왼쪽으로 움직인다고 생각할 수 있습니다 . 여러 항에 상당히 일찍 0 으로 축소하는 경로가 있습니다 .

    해 경로 그림의 세로 축은 표준화된 예측 변수의 모수 추정값을 나타냅니다 . 빨간색 수직선은 교차 검증에 의해 결정된 최적 축소에서의 값을 나타냅니다 . 이 점에서 11 개 항의 계수가 0 이 아닙니다 . 빨간색 수직선은 검증 데이터 집합의 최소 척도화된 –1* 로그 가능도 값을 나타냅니다 .

그림 7.1 0 이 아닌 항이 강조 표시된 적응형 Lasso 회귀 적합의 해 경로



" 원래 예측 변수에 대한 모수 추정값 " 보고서 (그림 7.2) 에서는 비중심화 및 비척도화 데이터에 대한 모수 추정값을 보여 줍니다 . 모수 추정값이 0 이 아닌 11 개 항이 강조 표시됩니다 . 여기에 교호작용 효과가 포함되어 있습니다 . 모든 예측 변수 열이 계수가 0 이 아닌 항에 나타나므로 데이터 테이블에서 6 개의 모든 예측 변수 열이 선택됩니다 .

" 효과 검정 " 보고서에서 계수 추정값이 0 인 10 개 효과가 " 제거 " 로 지정됩니다 . " 효과 검정 " 보고서는 0.05 수준에서 하나의 효과 (Age*Markers 교호작용) 만 유의함을 나타냅니다 .

그림 7.2 0 이 아닌 항이 강조 표시된 모수 추정값 보고서

원래 예측 변수에 대한 모수 추정값

항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	1.9658064	1.5268652	1.6576014	0.1979	-1.026794	4.9584071
BMI	-0.014557	0.0655906	0.0492589	0.8244	-0.143113	0.1139977
Age	-0.000344	0.0052439	0.0043119	0.9476	-0.010622	0.0099335
Time	0	0	0	1.0000	0	0
Markers[0-1]	0.3853241	0.3347727	1.3248061	0.2497	-0.270818	1.0414665
Hepatitis[0-1]	0	0	0	1.0000	0	0
Jaundice[0-1]	-0.220764	0.1645347	1.8002793	0.1797	-0.543246	0.1017185
(BMI-23.1737)*(Age-56.3994)	0.0007336	0.0014188	0.267356	0.6051	-0.002047	0.0035144
(BMI-23.1737)*(Time-9.00945)	0	0	0	1.0000	0	0
(BMI-23.1737)*Markers[0-1]	-0.066574	0.0511526	1.6938495	0.1931	-0.166831	0.0336832
(BMI-23.1737)*Hepatitis[0-1]	0.0269878	0.0827496	0.1063662	0.7443	-0.135198	0.1891741
(BMI-23.1737)*Jaundice[0-1]	0	0	0	1.0000	0	0
(Age-56.3994)*(Time-9.00945)	0	0	0	1.0000	0	0
(Age-56.3994)*Markers[0-1]	-0.023498	0.0083023	8.0106599	0.0047*	-0.03977	-0.007226
(Age-56.3994)*Hepatitis[0-1]	0	0	0	1.0000	0	0
(Age-56.3994)*Jaundice[0-1]	0	0	0	1.0000	0	0
(Time-9.00945)*Markers[0-1]	-0.007529	0.0104633	0.5178189	0.4718	-0.028037	0.0129784
(Time-9.00945)*Hepatitis[0-1]	0	0	0	1.0000	0	0
(Time-9.00945)*Jaundice[0-1]	0.0009619	0.0093476	0.0105895	0.9180	-0.017359	0.0192828
Markers[0-1]*Hepatitis[0-1]	-0.422619	0.397399	1.1309549	0.2876	-1.201507	0.3562684
Markers[0-1]*Jaundice[0-1]	0	0	0	1.0000	0	0
Hepatitis[0-1]*Jaundice[0-1]	0	0	0	1.0000	0	0

12. " 원래 예측 변수에 대한 모수 추정값 " 보고서에서 (Age - 56.3994)*Markers[0-1] 행을 클릭합니다 .

    이렇게 하면 해 경로 그림에서 해당 효과의 경로가 강조 표시되고 데이터 테이블에서 Age 및 Markers 열이 선택됩니다 .

13. " 검증 열을 사용한 Poisson 적응형 Lasso 회귀 " 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장** 및 **열 저장 > 분산 계산식 저장**을 선택합니다 .

Node Count 예측 계산식과 Node Count 분산이라는 두 열이 데이터 테이블에 추가됩니다 .

14. 데이터 테이블에서 열 머리글 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식을 확인합니다. 아니면 "열" 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다.

    예측 계산식 저장 열의 예측 계산식은 모형의 추정된 선형 부분에 지수 함수를 적용합니다 . Poisson 분포의 분산이 평균과 같으므로 Node Count 분산의 예측 분산 계산식은 동일한 계산식으로 제공됩니다 .

이항 일반화 회귀 예

이 예에서는 Liver Cancer.jmp 샘플 데이터 테이블의 이항 반응 Severity(심각도) 에 대한 예측 모형을 개발하는 방법을 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Liver Cancer.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Severity 를 선택하고 **Y** 를 클릭합니다 .
4. BMI ~ Jaundice 를 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .

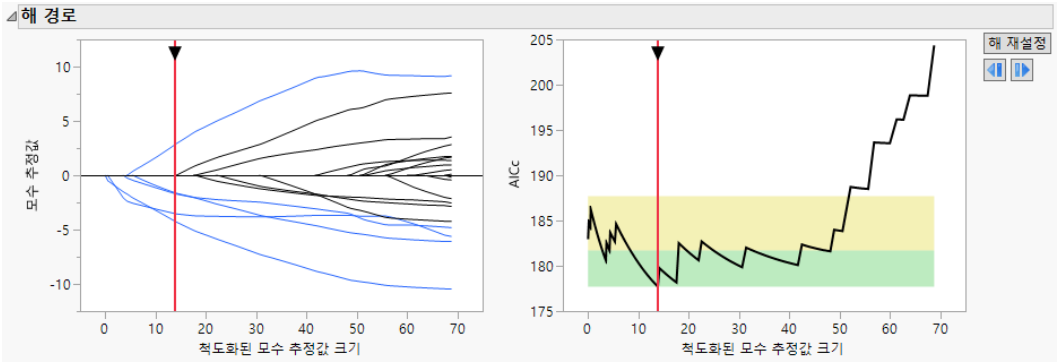
   차수 2(**차수** 상자의 기본값) 까지의 모든 항이 모형에 추가됩니다 .
5. " 분석법 " 목록에서 **일반화 회귀**를 선택합니다 .

   " 분포 " 목록에 " 이항 " 분포가 자동으로 표시됩니다 . Y 가 이항 변수이고 명목형 모델링 유형일 때는 이 분포만 사용할 수 있습니다 .
6. **실행**을 클릭합니다 .

   " 모형 비교 " 보고서 , " 모형 시작 " 제어판 및 " 로지스틱 회귀 " 보고서가 포함된 " 일반화 회귀 " 보고서가 나타납니다 . 기본 추정 방법은 "Lasso 회귀 " 입니다 .
7. " 추정 방법 " 에서 **Elastic Net** 을 선택합니다 .
8. **적응형** 상자를 선택합니다 .
9. **시작**을 클릭합니다 .

   "AICc 검증을 사용한 이항 적응형 Elastic Net" 보고서가 나타납니다 . 그림 7.3 에는 해 경로가 나와 있습니다 .

그림 7.3 해 경로 그림



계수가 0 이 아닌 항의 경로가 파란색으로 표시됩니다 . 최적 모수 값은 MLE 에서 상당히 축소됩니다 . 오른쪽의 검증 그림은 몇 가지 모형을 최적 모형으로 고려할 수 있음을 나타냅니다 . 해당 모형을 보려면 녹색 부분에 검은색 선이 있는 영역에서 빨간색 세로 막대를 움직입니다 .

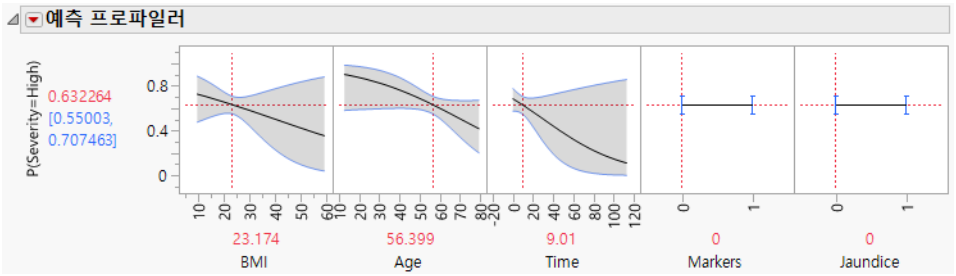
10. "AICc 검증을 사용한 이항 적응형 Elastic Net" 옆의 빨간색 삼각형을 클릭하고 **0 인 항 선택** 옵션을 선택합니다 .

    " 원래 예측 변수에 대한 모수 추정값 " 보고서에서 계수 추정값이 0 인 16 개 항이 강조 표시됩니다 . 이 항은 " 효과 검정 " 보고서에 " 제거 " 로 지정됩니다 .

    또한 " 효과 검정 " 보고서는 0.05 수준에서 유의한 항이 없음을 보여 줍니다 . 그러나 Time*Markers 교호작용의 p 값 (0.0626) 이 작고 Time 효과의 p 값 (0.1458) 도 작습니다 .

11. "AICc 검증을 사용한 이항 적응형 Elastic Net" 옆의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다 .

그림 7.4 Severity = High 일 때 Time 이 낮을 확률에 대한 프로파일러

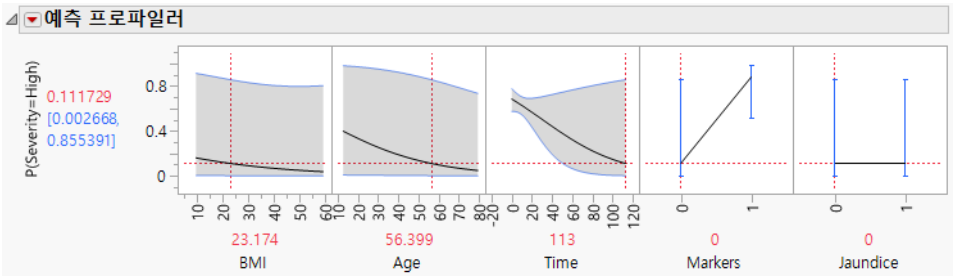


예측 프로파일러를 검토하여 Time 및 Time*Markers 교호작용이 Severity 에 미치는 영향을 확인합니다 .

참고 : Hepatitis 예측 변수는 활성 (0 이 아님) 항에 나타나지 않으므로 프로파일러에 표시되지 않습니다 . Markers 와 Jaundice 는 활성 교호작용 항에 나타나므로 주효화가 활성이 아니더라도 프로파일러에 표시됩니다 .

12. Time 의 빨간색 파선을 왼쪽에서 오른쪽으로 움직여 Markers 와의 교호작용을 확인합니다 (그림 7.4 및 그림 7.5). 진단 이후 Time 값이 작은 환자의 경우 Markers 는 Severity 에 거의 영향을 주지 않습니다 . 그러나 장기간 진단을 받고 연구에 참여한 환자의 경우 Markers 가 중요합니다 . 이러한 환자에게 정상 표지자는 Severity 가 높을 확률이 낮음을 나타냅니다 .

그림 7.5 Severity = High 일 때 Time 이 높을 확률에 대한 프로파일러



영과잉 Poisson 회귀 예

Fishing.jmp 샘플 데이터 테이블에는 공원을 방문하는 그룹이 잡은 물고기 수에 영향을 미치는 다양한 요인 연구를 위한 가상의 데이터가 포함되어 있습니다. 여행 중인 가족 또는 그룹에 대한 250 개의 반응이 데이터 테이블에 포함되어 있습니다. 이 예에서는 Fish Caught(잡힌 물고기 수) 를 Live Bait(산 미끼), Fishing Poles(낚싯대), Camper(캠퍼), People(사람) 및 Children(어린이) 의 함수로 모델링합니다. 이러한 변수는 데이터 테이블의 각 열에서 "노트" 열 특성에 설명되어 있습니다. "노트" 열 특성을 보려면 열 이름을 마우스 오른쪽 버튼으로 클릭하고 "열 정보"를 선택한 후 "열 특성" 아래의 "노트"를 선택합니다.

데이터 테이블에는 Fished(낚시 여부) 라는 숨겨진 열이 있습니다. 데이터 수집 중 그룹의 누군가가 실제로 낚시를 했는지 여부는 확인되지 않았습니다. 그러나 두 가지 경우 (그룹에서 낚시를 한 사람이 아무도 없거나 그룹에서 낚시를 한 모든 사람이 운이 없었음) 중 하나를 통해 물고기를 잡지 못하는 상황이 발생할 수 있다는 점을 강조하기 위해 Fished 열이 테이블에 포함되었습니다.

따라서 두 가지 원인을 통해 반응이 0 일 수 있습니다. 이 문제를 해결하기 위해 영과잉 분포를 적합시킬 수 있습니다 . Poisson 분포는 낚시를 한 사람에 대한 개수 데이터에 적합하므로 영과잉 Poisson 분포를 적합시킵니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fishing.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Fish Caught 를 선택하고 **Y** 를 클릭합니다 .
4. Live Bait ~ Children 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .

   차수 2(**차수** 상자의 기본값) 까지의 항이 모형에 추가됩니다 .
5. " 열 선택 " 목록에서 "Validation" 을 선택하고 **검증**을 클릭합니다 .
6. " 분석법 " 목록에서 **일반화 회귀**를 선택합니다 .

7. " 분포 " 목록에서 **ZI Poisson** 을 선택합니다 .
8. **실행**을 클릭합니다 .

   " 모형 비교 " 보고서 , " 모형 시작 " 제어판 및 " 검증 열을 사용한 ZI Poisson 최대 가능도 " 보고서가 포함된 " 일반화 회귀 " 보고서가 나타납니다 . 기본 추정 방법은 "Lasso 회귀 " 입니다 .
9. " 추정 방법 " 목록에서 **Elastic Net** 을 선택합니다 .
10. **시작**을 클릭합니다 .

    " 검증 열을 사용한 ZI Poisson Elastic Net" 보고서가 나타납니다 . " 해 경로 ", " 원래 예측 변수에 대한 모수 추정값 " 보고서 및 " 효과 검정 " 보고서에서 하나의 항이 0 임을 나타냅니다 . 두 개의 모수 추정값 보고서에서 마지막 줄에 추정값이 표시된 영과잉 모수는 매우 유의합니다 . 이것은 반응 (Fish Caught) 에서 나타난 일부 변동은 일부 그룹이 낚시를 하지 않았기 때문일 수 있음을 나타냅니다 .

그림 7.6 원래 예측 변수에 대한 모수 추정값 보고서

원래 예측 변수에 대한 모수 추정값

항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	1.3599776	0.3237335	17.647701	<.0001*	0.7254716	1.9944837
Live Bait[0-1]	-0.379555	0.1919937	3.9081951	0.0481*	-0.755856	-0.003255
Fishing Poles	0.3034685	0.0465239	42.547553	<.0001*	0.2122833	0.3946537
Camper[0-1]	-0.363094	0.1899392	3.654349	0.0559	-0.735368	0.0091795
People	0.0776017	0.121377	0.4087617	0.5226	-0.160293	0.3154962
Children	-0.192656	0.1028943	3.5057658	0.0612	-0.394325	0.009013
Live Bait[0-1]*(Fishing Poles-0.856)	-0.026766	0.0668764	0.160184	0.6890	-0.157841	0.1043095
Live Bait[0-1]*Camper[0-1]	-0.336382	0.48759	0.4759452	0.4903	-1.292041	0.6192764
Live Bait[0-1]*(People-2.516)	-0.164164	0.1994702	0.6773258	0.4105	-0.555118	0.2267908
Live Bait[0-1]*(Children-0.936)	0.1025463	0.2362161	0.1884607	0.6642	-0.360429	0.5655213
(Fishing Poles-0.856)*Camper[0-1]	-0.446199	0.1120088	15.869111	<.0001*	-0.665732	-0.226666
(Fishing Poles-0.856)*(People-2.516)	-0.00611	0.0853344	0.0051267	0.9429	-0.173362	0.1611423
(Fishing Poles-0.856)*(Children-0.936)	-0.17587	0.080167	4.8127458	0.0282*	-0.332995	-0.018746
Camper[0-1]*(People-2.516)	-0.195261	0.3035369	0.4138148	0.5200	-0.790182	0.3996608
Camper[0-1]*(Children-0.936)	0.1899805	0.2998548	0.4014171	0.5264	-0.397724	0.7776852
(People-2.516)*(Children-0.936)	0	0	0	1.0000	0	0

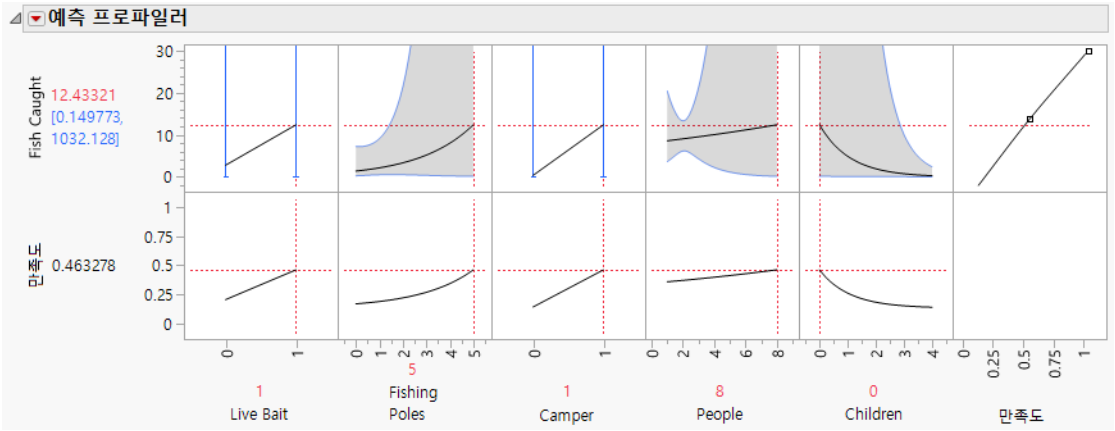
ZI Poisson 분포 모수	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
영과잉	0.7815222	0.0354433	486.19905	<.0001*	0.7120545	0.8509898

" 효과 검정 " 보고서는 0.05 수준에서 네 개의 항 , 즉 Live Bait, Fishing Poles, Fishing Poles* Camper 및 Fishing Poles*Children 이 유의함을 나타냅니다 .

11. " 검증 열을 사용한 ZI Poisson Elastic Net" 옆의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다 .
12. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 함수**를 선택합니다 .

    반응에 함수가 적용되어 Fish Caught 수를 최대화하는 것이 만족도가 높음을 나타냅니다 . 만족도 함수에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .
13. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 7.7 Fish Caught 가 최대화된 예측 프로파일러



예측 변수의 설정을 변경하여 유의한 효과 , 즉 Live Bait, Fishing Poles, Fishing Poles*Camper 및 Fishing Poles*Children 의 영향을 확인할 수 있습니다 . 예를 들어 Live Bait 는 더 많은 물고기와 관련이 있고 , Camper 는 캠핑을 하지 않는 사람보다 낚싯대를 더 많이 가져와 더 많은 물고기를 잡는 경향이 있습니다 .

14. " 검증 열을 사용한 ZI Poisson Elastic Net" 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장** 및 **열 저장 > 분산 계산식 저장**을 선택합니다 .

    Fish Caught 예측 계산식과 Fish Caught 분산이라는 두 열이 데이터 테이블에 추가됩니다 .

15. 데이터 테이블에서 열 머리글 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식을 확인합니다 . 아니면 " 열 " 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다 . 두 계산식에서 모두 추정된 영과잉 모수가 0.781522155 로 나타납니다 .

가로형 데이터에 대한 일반화 회귀 예

예측 변수가 관측값 수보다 많은 경우 기존 회귀 방법은 실용적이지 않습니다. 이러한 경우 변수 선택을 포함하는 회귀 방법을 사용하여 회귀 모형을 적합시킬 수 있습니다. 이 예에서는 변수 선택 수준이 서로 다른 세 모형을 비교합니다. 세 모형이 모두 검증 그림의 녹색 영역 내에 있으므로 이러한 모형 모두 최적 모형과 유사하다는 강력한 증거가 있습니다.

Prostate Cancer.jmp 샘플 데이터 테이블에는 165 명의 남성으로부터 수집한 혈청 표본의 결과가 포함되어 있습니다 . 이 중 약 절반이 전립선암을 앓고 있습니다 . 혈청 표본에서 측정한 667 개의 단백질이 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Prostate Cancer.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Status 를 선택하고 **Y** 를 클릭합니다 .

이 열은 명목형 반응 열이므로 " 분석법 " 이 " 명목형 로지스틱 " 으로 변경되고 " 목표 수준 " 옵션이 나타납니다 . 이 옵션의 기본값은 데이터 테이블의 " 목표 수준 " 열 특성에 지정된 값인 "CCD" 입니다 .

4. " 분석법 " 목록에서 **일반화 회귀**를 선택합니다 .

   " 분포 " 목록에 " 이항 " 분포가 자동으로 표시됩니다 . Y 가 이항 변수이고 명목형 모델링 유형일 때는 이 분포만 사용할 수 있습니다 .

5. " 열 선택 " 목록에서 Proteins 열 그룹을 선택하고 **추가**를 클릭합니다 .

   그러면 열 그룹에 포함된 667 개 열이 모두 모형에 추가됩니다 .

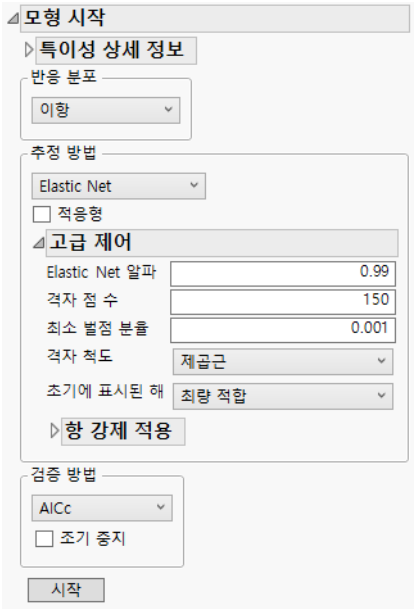
6. **실행**을 클릭합니다 .

   " 모형 시작 " 제어판이 포함된 " 일반화 회귀 " 보고서가 나타납니다 . 예측 변수의 수가 관측값 수보다 많으므로 초기 로지스틱 회귀 모형 적합이 없습니다 .

7. " 추정 방법 " 에서 **Elastic Net** 을 선택합니다 .

8. " 고급 제어 " 옆의 회색 표시 아이콘을 클릭합니다 .

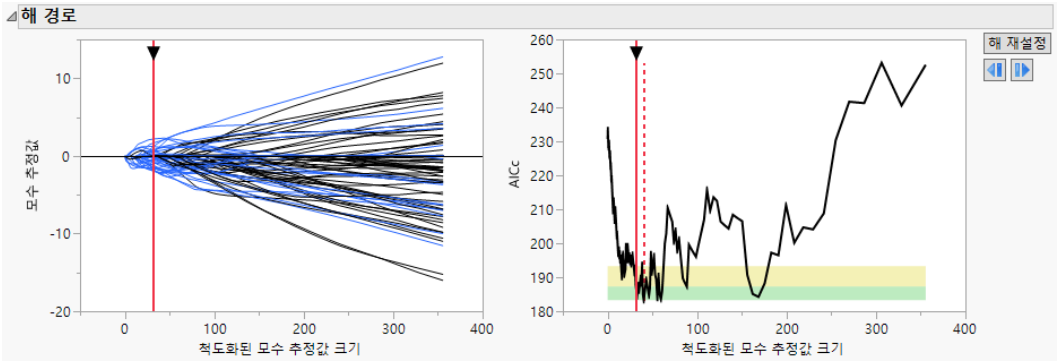
그림 7.8 고급 제어



9. " 초기에 표시된 해 " 에서 **녹색 영역에서 가장 작음**을 선택합니다 .

10. **시작**을 클릭합니다 .

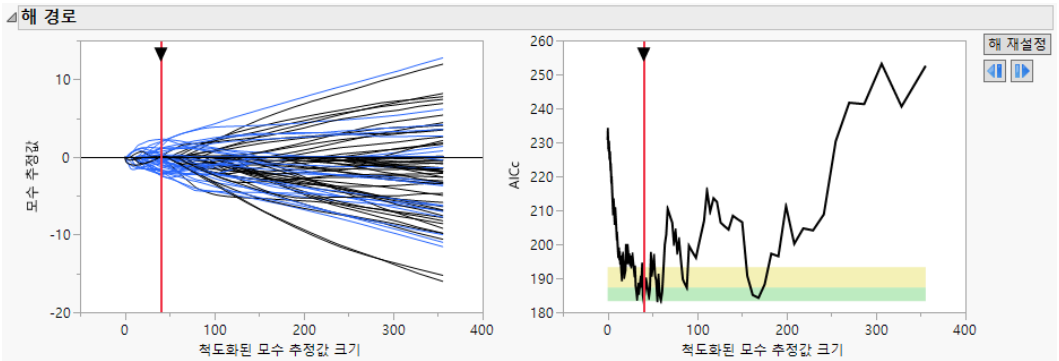
그림 7.9 녹색 영역에서 가장 작은 모형



" 해 경로 " 에서는 최소 AICc 모형과 유사한 것으로 간주되는 가장 작은 모형을 보여 줍니다 . 여기서 가장 작은 모형은 모수의 수가 가장 적은 모형을 의미합니다 .

11. "AICc 검증을 사용한 이항 Elastic Net" 옆의 회색 표시 아이콘을 클릭합니다 .
12. " 모형 시작 " 옆의 회색 표시 아이콘을 클릭합니다 .
13. " 초기에 표시된 해 " 에서 **최량 적합**을 선택합니다 .
14. **시작**을 클릭합니다 .

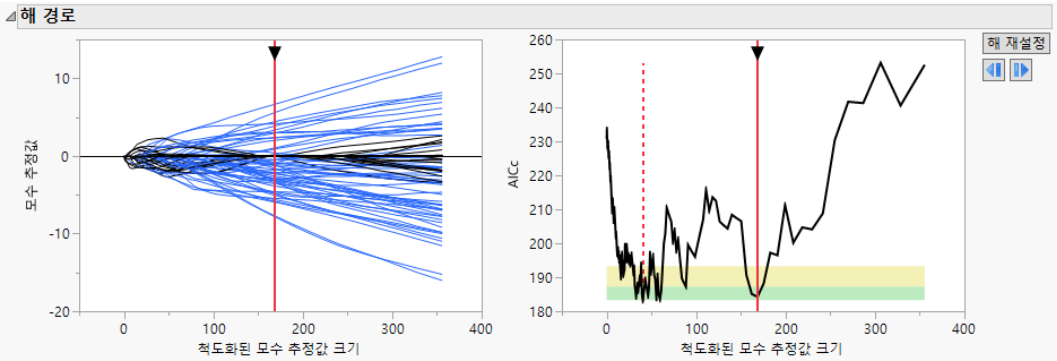
그림 7.10 최량 적합 모형



" 해 경로 " 에서는 최량 적합 모형을 보여 줍니다 . 여기서 최량 적합 모형은 최소 AICc 값을 가진 모형을 의미합니다 .

15. "AICc 검증을 사용한 이항 Elastic Net" 옆의 회색 표시 아이콘을 클릭합니다 .
16. " 모형 시작 " 옆의 회색 표시 아이콘을 클릭합니다 .
17. " 초기에 표시된 해 " 에서 **녹색 영역에서 가장 큼**을 선택합니다 .
18. **시작**을 클릭합니다 .

그림 7.11 녹색 영역에서 가장 큰 모형



" 해 경로 " 에서는 최소 AICc 모형과 유사한 것으로 간주되는 가장 큰 모형을 보여 줍니다 . 여기서 가장 큰 모형은 모수의 수가 가장 많은 모형을 의미합니다 .

19. "AICc 검증을 사용한 이항 Elastic Net" 옆의 회색 표시 아이콘을 클릭합니다 .

그림 7.12 모형 비교 테이블

모형 비교

표시	반응 분포	추정 방법	검증 방법	0이 아닌 모수	AICc	BIC	일반화 R^2
☑	이항	Elastic Net	AICc	36	186.07102	277.07256	0.7466199
☑	이항	Elastic Net	AICc	40	183.29544	281.08164	0.8021925
☑	이항	Elastic Net	AICc	57	184.25399	299.49849	0.9824565

" 모형 비교 " 보고서에 세 개의 모형이 표시됩니다 . "0 이 아닌 모수 " 열을 사용하여 각 모형의 크기를 식별할 수 있습니다 . 모형의 모수 수가 증가하면 " 일반화 R^2" 값이 증가합니다 . 이러한 모형이 모두 녹색 영역에 있으므로 모든 모형이 최적 모형과 유사하다는 강력한 증거가 있습니다 .

8 장

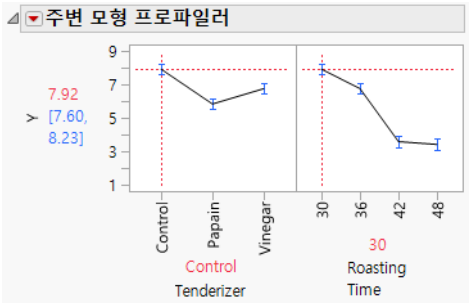
JMP PRO 혼합 모형
평균 및 공분산 공동 모델링

모형 적합 플랫폼의 혼합 모형 분석법은 JMP Pro 에서만 사용할 수 있습니다 .

JMP Pro에서 모형 적합 플랫폼의 혼합 모형 분석법은 복소 공분산 구조의 연속형 반응에 대해 다양한 선형 모형을 적합시킵니다. 이러한 모형에는 랜덤 계수, 반복 측정, 공간 데이터 및 상관된 여러 반응이 있는 데이터가 포함됩니다. 혼합 모형 분석법을 사용하여 직관적인 인터페이스로 선형 혼합 모형 및 모형의 공분산 구조를 간편하게 지정하고, 최대 가능도 방법으로 이러한 모형을 적합시킬 수 있습니다.

모델링 결과는 프로파일러, 표면 그림 및 등고선 그림과 같은 대화식 시각화 도구에 의해 지원됩니다. 이러한 도구를 사용하여 모형에 대한 이해를 보완할 수 있습니다.

그림 8.1 분할구 실험에 대한 주변 모형 프로파일러



목차

혼합 모형 분석법 개요 330
예 : 혼합 모형 사용 330
혼합 모형 분석법 시작 335
모형 적합 시작 창 335
데이터 형식 340
혼합 적합 보고서 341
적합 통계량 343
임의 효과 공분산 모수 추정값 345
고정 효과 모수 추정값 345
반복 효과 공분산 모수 추정값 346
랜덤 계수 347
임의 효과 예측 347
고정 효과 검정 347
순차 검정 348
다중 비교 348
기울기 비교 349
주변 모형 추론 349
실제값 대 예측값 그림 349
잔차 그림 349
Profilers 350
조건부 모형 추론 351
실제값 대 조건부 예측값 그림 351
조건부 잔차 그림 351
조건부 프로파일러 352
변동도 352
공분산 및 상관 행렬 353
열 저장 354
혼합 모형 분석법의 추가 예 354
반복 측정 예 355
분할구 예 370
공간 예: 균일도 시험 376
상관관계가 있는 반응의 예 384
혼합 모형 분석법에 대한 통계 상세 정보 390
수렴 스코어 검정 391
랜덤 계수 모형 392
반복 측정 393
반복 공분산 구조 393

공간 및 시간 변동 398
Kackar-Harville 수정 401

JMP PRO 혼합 모형 분석법 개요

JMP Pro 의 혼합 모형 분석법을 사용하면 복소 공분산 구조를 가진 모형을 분석할 수 있습니다 . 다음과 같은 경우 이 방법으로 분석할 수 있습니다 .

- 분할구 실험
- 랜덤 계수 모형
- 반복 측정 설계
- 공간 데이터
- 상관관계가 있는 반응 데이터

분할구 실험은 실험 단위의 수준 또는 크기가 두 개 이상이고 결과적으로 여러 오차 항을 생성하는 실험입니다. 보통 일부 요인은 변경하기 쉽고 다른 요인은 변경하기 어려울 때 이러한 설계가 필요합니다. 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

랜덤 계수 모형은 계층적 또는 다수준 모형이라고도 합니다(Singer 1998, Sullivan et al. 1999). 배치 또는 개체의 절편 및 기울기가 무작위로 달라진다고 추정될 때 이러한 모형이 사용됩니다. 제약 산업의 약물 안정성 시험과 교육 리서치의 개별 성장 연구에서 랜덤 계수 모형이 필요한 경우가 많습니다.

반복 측정 설계 , 공간 데이터 및 상관된 반응 데이터는 관측값이 독립적이지 않다는 특성을 공유하므로 상관 구조를 모델링해야 합니다 .

- 반복 측정 설계 (개체 내 설계라고도 함) 는 오차의 상관관계를 허용하면서 시간 또는 공간에 따른 반응의 변화를 모델링합니다 .
- 공간 데이터는 일반적으로 위도와 경도 같은 둘 이상의 차원에서 측정되는 값입니다 . 공간 측정값은 주로 공간 근접성의 함수로 상관관계를 보입니다 .
- 상관된 반응 데이터는 동일한 실험 단위에 대해 여러 가지 측정을 수행한 결과입니다 . 예를 들어 의학 연구에서 개인의 키 , 체중 및 혈압 측정값 또는 제조품에 대해 측정한 경도 , 강도 및 탄성은 상관관계가 있을 가능성이 큽니다. 이러한 측정값을 개별적으로 연구할 수 있지만 상관된 반응으로 처리하면 유용한 정보를 끌어낼 수 있습니다 .

관측값 간의 상관관계를 설명하지 못하면 처리 효과에 대한 잘못된 결론을 내릴 수 있습니다. 그러나 공분산 구조 모수를 추정할 때 데이터의 정보가 사용됩니다. 추정되는 모수의 수는 검정력과 제1종 오류 비율에 영향을 줍니다. 따라서 공분산 모형을 신중하게 선택해야 합니다. 자세한 내용은 "반복 측정 예"(355 페이지)에서 확인하십시오.

JMP PRO 예 : 혼합 모형 사용

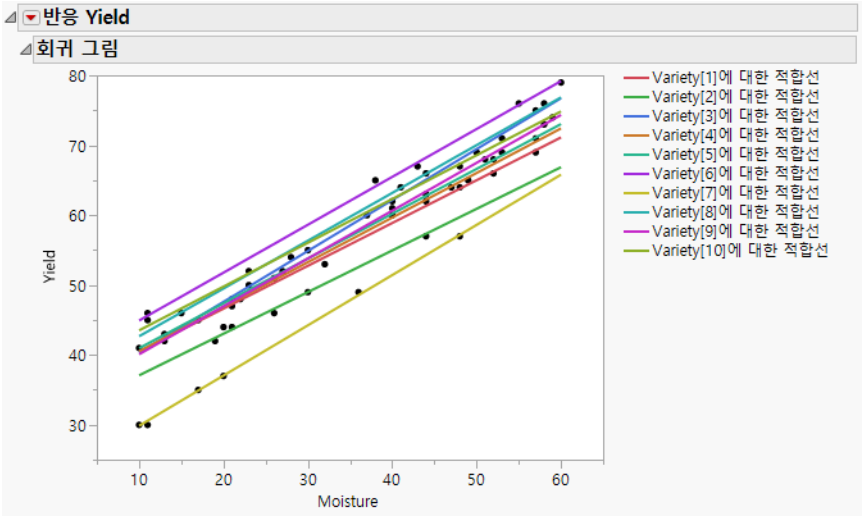
밀 수확량 연구에서 건조 기후 조건에 맞게 적응된 경질 붉은 겨울밀 품종 모집단에서 10종의 밀을 무작위로 선택합니다. 그런 다음 6개의 1에이커 대지에 무작위로 할당합니다. 예정지의 수분 함량은 발아율에 영향을 줄 수 있으며 결과적으로 최종 수확량이 달라질 수 있습니다. 따라서 각

대지에 대해 상위 36 인치 토양의 예정지 수분 양이 결정됩니다. 이 연구에서 수분 함량이 수확량에 영향을 주는지 여부를 확인하는 데 관심이 있습니다.

품종이 무작위로 선택되므로 각 품종에 대한 회귀 모형은 품종 모형 모집단에서 선택한 랜덤 모형입니다. 절편과 기울기는 각 품종에 대해 랜덤하며 상관관계가 있을 수 있습니다. 랜덤 계수는 고정 효과에서 중심화됩니다. 고정 효과는 모집단 절편과 기울기이며, 이는 품종의 절편과 기울기 모집단에 대한 기대값입니다. 이 예는 Littell et al. 연구 자료 (2006, p. 320)에서 가져왔습니다.

표준 최소 제곱 분석법에서 REML 을 사용하여 모형을 적합시키면 절편과 기울기의 변동을 볼 수 있습니다(그림 8.2). 기울기의 변동성은 크지 않지만 절편은 상당한 차이가 있습니다. 절편과 기울기는 음의 상관관계가 있을 수 있으며 절편이 낮은 품종의 기울기가 더 높은 것으로 보입니다.

그림 8.2 표준 최소 제곱 회귀

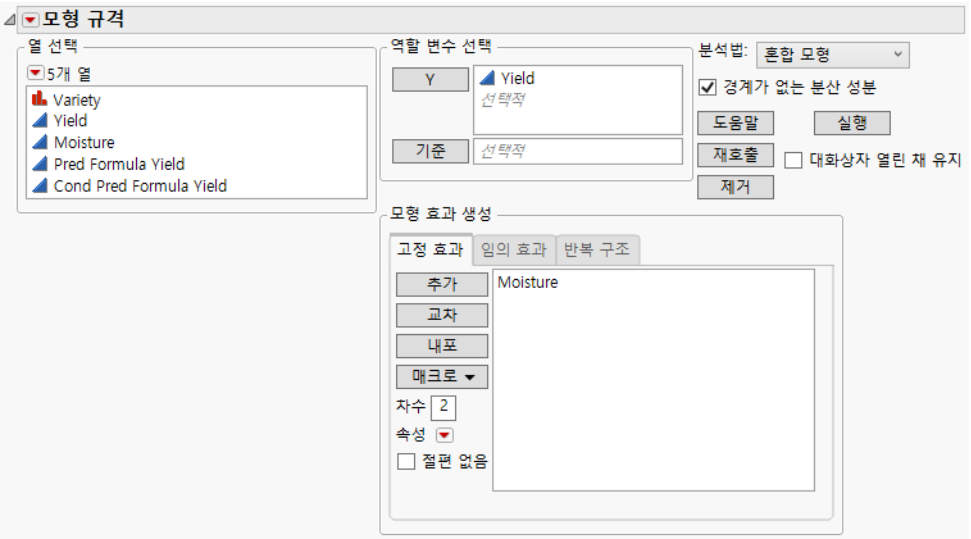


절편과 기울기 간의 상관관계를 모델링하기 위해 혼합 모형 분석법을 사용합니다 . 이를 통해 모집단 회귀 방정식과 품종별 방정식을 결정하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Wheat.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Yield 를 선택하고 **Y** 를 클릭합니다 .

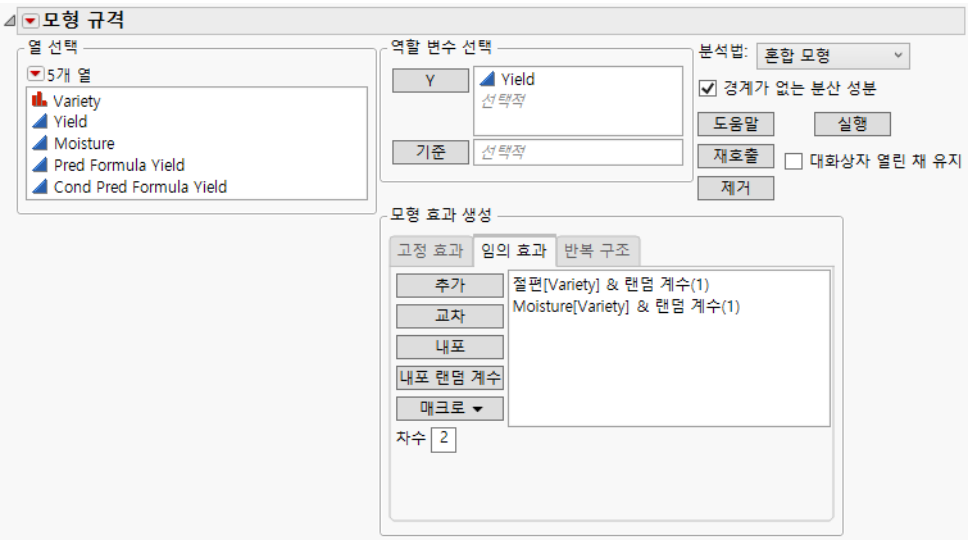
   이 열을 Y 로 추가하면 적합 분석법이 " 표준 최소 제곱 " 으로 설정됩니다 .
4. " 분석법 " 목록에서 **혼합 모형**을 선택합니다 . 아니면 " 혼합 모형 " 분석법을 먼저 선택한 후 **Y** 를 클릭하여 Yield 를 추가할 수도 있습니다 .
5. Moisture 를 선택하고 " 고정 효과 " 탭에서 **추가**를 클릭합니다 .

그림 8.3 고정 효과를 보여 주는 완료된 모형 적합 시작 창



6. **임의 효과** 탭을 선택합니다 .
7. Moisture 를 선택하고 **추가**를 클릭합니다 .
8. " 열 선택 " 목록에서 **Variety** 를 선택하고 " 임의 효과 " 탭에서 Moisture 를 선택한 후 **내포 랜덤 계수**를 클릭합니다 .

그림 8.4 임의 효과 탭을 보여 주는 완료된 모형 적합 시작 창



임의 효과는 품종별로 그룹화되고 절편은 랜덤 성분으로 포함됩니다 .

9. (선택 사항 .) " 모형 규격 " 의 빨간색 삼각형을 클릭하고 **중심화 다항식** 옵션의 설정을 확인합니다 .

   Moisture 효과는 임의 효과로 포함되어 있으므로 " 중심화 다항식 " 옵션이 선택되어 있더라도 평균으로 중심화되지 않습니다 .

10. **실행**을 클릭합니다 .

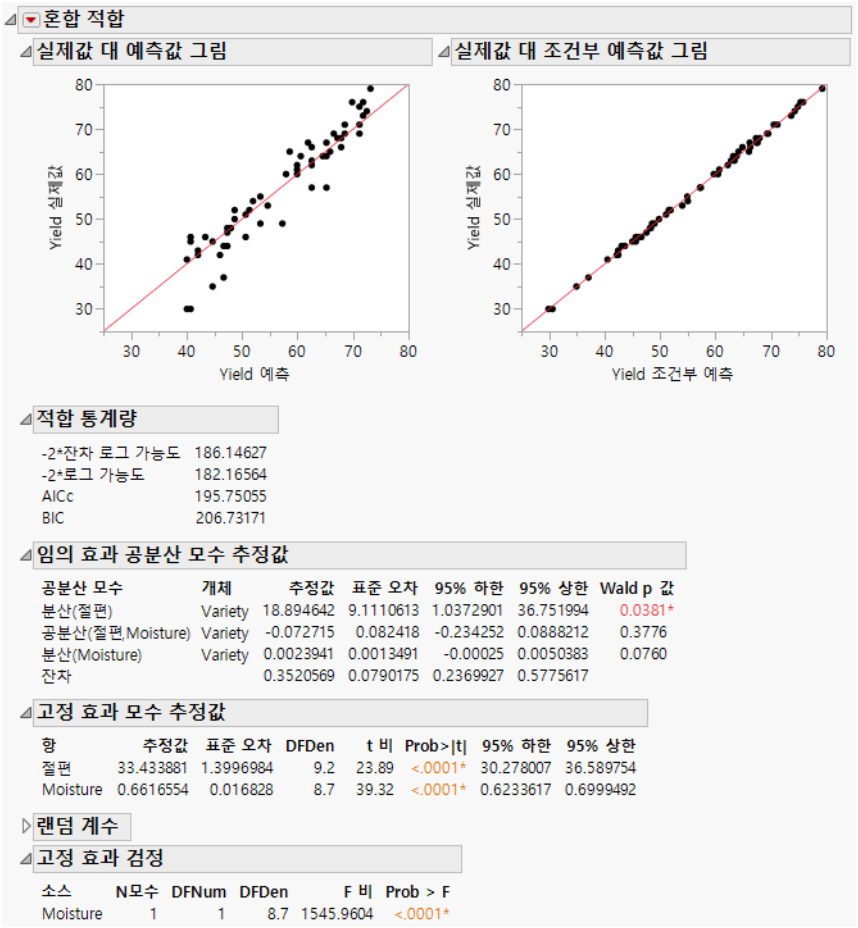
    그림 8.5 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . " 실제값 대 예측값 그림 " 은 모형 적합 및 기본 가정에 대해 불일치를 나타내지 않습니다 .

    모형 적합에 명백한 문제가 없으므로 통계 검정을 해석하고 회귀 방정식을 구할 수 있습니다 . " 고정 효과 검정 " 보고서에서 볼 수 있듯이 수확량에 대한 수분의 효과는 유의합니다 . " 고정 효과 모수 추정값 " 에 제공된 추정값은 다음 방정식이 추정된 모집단 회귀 방정식임을 나타냅니다 .

    $$Yield = 33.43 + 0.66 * Moisture$$

    " 임의 효과 공분산 모수 추정값 " 보고서에는 품종의 절편 분산을 나타내는 " 분산 (절편)", 기울기를 나타내는 " 분산 (Moisture)" 및 공분산을 나타내는 " 공분산 (절편 , Moisture)" 의 추정값이 제공됩니다 . 이 경우 절편과 기울기는 추정값의 신뢰 구간에 0 이 포함되므로 유의한 상관관계가 없습니다 . 이 보고서에는 잔차 분산의 추정값도 제공됩니다 .

그림 8.5 혼합 적합 보고서



모집단 회귀 방정식의 추정값이 있지만 품종 (Variety) 2 의 추정 수확량에도 관심이 있습니다 .

11. " 랜덤 계수 " 보고서를 열어 " 절편 " 및 "Moisture" 에 대한 품종 효과의 추정값을 확인합니다 . 이러한 계수는 각 품종이 모집단과 어떻게 다른지 추정합니다 .

그림 8.6 랜덤 계수 보고서

랜덤 계수

Variety

Variety	절편	Moisture
1	0.9577924	-0.049211
2	-2.284289	-0.066697
3	-0.408109	0.0672225
4	0.6960205	-0.023306
5	1.115904	-0.019904
6	4.639151	0.0238887
7	-10.73004	0.0564233
8	2.4011709	0.0224336
9	-0.176207	0.0233566
10	3.7886051	-0.034207

공분산 행렬

임의 효과	절편	Moisture
절편	18.89464	-0.07272
Moisture	-0.07272	0.002394

" 고정 효과 모수 추정값 " 및 " 랜덤 계수 " 보고서에서 Variety 2 에 대해 다음과 같은 예측 방정식을 구합니다 .

$$Yield = 33.433 + 0.662 * Moisture - 2.284 - 0.067 * Moisture$$

$$Yield = 31.149 + 0.595 * Moisture$$

Variety 2 는 모집단 평균보다 낮은 수확량으로 시작하여 모집단 평균보다 낮은 비율로 Moisture 와 함께 증가합니다 .

JMP PRO 혼합 모형 분석법 시작

혼합 모형은 모형 적합 시작 창에서 선택할 수 있는 여러 분석법 중 하나입니다 . 이 섹션에서는 모형 적합 시작 창에서 혼합 모형을 적합 방법으로 선택하는 방법을 설명합니다 . 이 선택 항목과 관련된 옵션에 대한 내용도 다룹니다 .

JMP PRO 모형 적합 시작 창

고정 효과 , 임의 효과 , 반복 구조 또는 이들의 조합을 사용하여 모형을 지정할 수 있습니다 . 옵션은 사용자가 지정한 모형의 특성에 따라 다릅니다 .

혼합 모형 분석법을 사용하여 모형을 적합시키려면 **분석 > 모형 적합**을 선택한 후 "분석법" 목록에서 **혼합 모형**을 선택합니다 . 분석법을 선택하기 **전에** Y 목록에 연속형 변수를 입력하면 분석법이 기본적으로 "표준 최소 제곱"으로 설정됩니다 . "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

혼합 모형 분석법을 사용하여 모형을 적합시킬 때 비유계 분산 성분을 허용할 수 있습니다 . 이렇게 하면 추정값이 음수인 분산 성분이 0 으로 보고되지 않습니다 . 이 옵션은 기본적으로 선택되어 있습니다 . 분산 추정값을 0 으로 제한하면 고정 효과 검정에서 편향이 발생할 수 있으므로 고

정 효과에 관심이 있다면 이 옵션이 선택된 상태여야 합니다. "비유계 분산 성분" 옵션에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "음의 분산"(175 페이지)에서 확인하십시오.

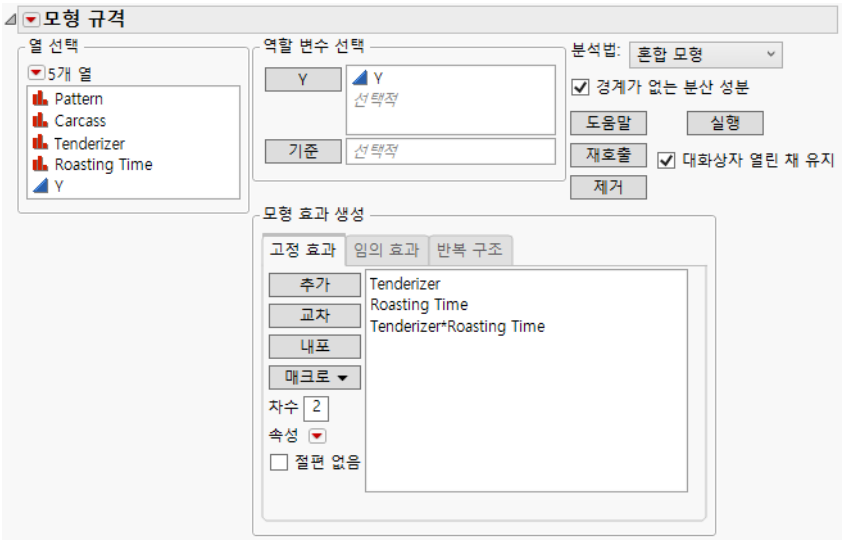
JMP PRO 고정 효과 탭

"고정 효과" 탭에서 모든 고정 효과를 추가합니다. 필요에 따라 "추가", "교차", "내포", "매크로" 및 "속성" 옵션을 사용합니다. 이러한 옵션에 대한 자세한 내용은 "모형 규격" 장(29 페이지)에서 확인하십시오.

참고 : 연속형 열이 임의 효과에 포함된 경우에는 " 모형 규격 " 의 빨간색 삼각형 메뉴에서 " 중심화 다항식 " 옵션이 선택되어 있더라도 해당 열이 중심화되지 않습니다 .

그림 8.7 에는 Split Plot.jmp 샘플 데이터 테이블 분석의 고정 효과가 나와 있습니다 . 모형에 고정 효과가 없을 수도 있습니다 . 예는 " 공간 예 : 균일도 시험 "(376 페이지) 에서 확인하십시오 .

그림 8.7 완료된 고정 효과를 보여 주는 모형 적합 시작 창



JMP PRO 임의 효과 탭

"임의 효과" 탭을 사용하여 기존의 분산 성분 모형 및 랜덤 계수 모형을 지정합니다.

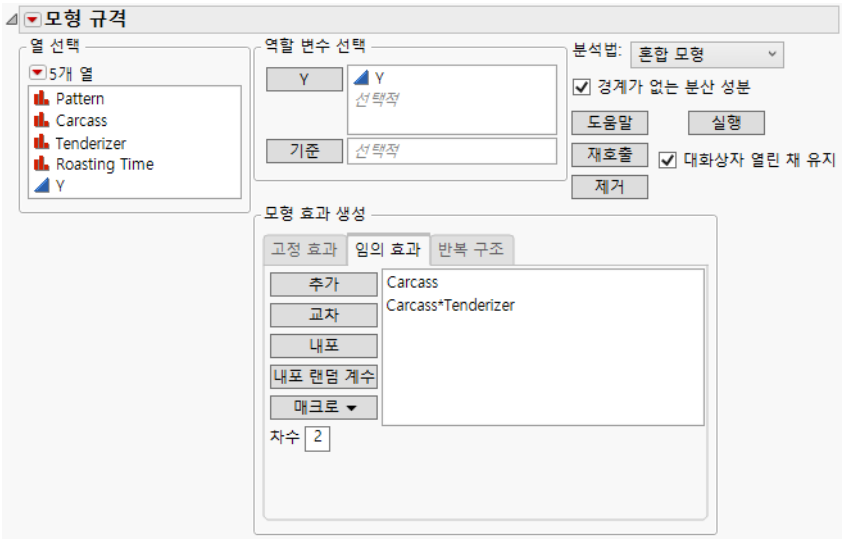
참고 : 연속형 열이 임의 효과에 포함된 경우에는 " 모형 규격 " 의 빨간색 삼각형 메뉴에서 " 중심화 다항식 " 옵션이 선택되어 있더라도 해당 열이 중심화되지 않습니다 .

분산 성분

기존 분산 성분 모형의 경우 "추가", "교차" 또는 "내포" 옵션을 사용하여 랜덤 블록, 주구 오차 항 및 하위구 오차 항과 같은 항을 지정합니다. 이러한 옵션에 대한 자세한 내용은 "모형 규격" 장(29 페이지)에서 확인하십시오.

그림 8.8 에서는 Carcass 가 랜덤 블록인 Split Plot.jmp 샘플 데이터의 임의 효과 지정을 보여 줍니다. " 분할구 예 "(370 페이지) 에서 이 예를 자세히 설명합니다 .

그림 8.8 완료된 임의 효과 탭을 보여 주는 모형 적합 시작 창



랜덤 계수

랜덤 계수 모형을 생성하려면 " 내포 랜덤 계수 " 버튼을 사용하여 랜덤 계수 그룹을 생성합니다 .

1. " 열 선택 " 목록에서 연속형 예측 변수 열을 선택합니다 .
2. **임의 효과** 탭을 선택한 후 **추가**를 선택합니다 .
3. " 임의 효과 " 탭에서 이러한 효과를 선택합니다 . 수준이 개별 회귀 모형을 정의하는 임의 효과가 포함된 열도 선택합니다 . 이 열은 기본적으로 SAS PROC MIXED 에서 random 문의 개체입니다 .
4. **내포 랜덤 계수** 버튼을 클릭합니다 .

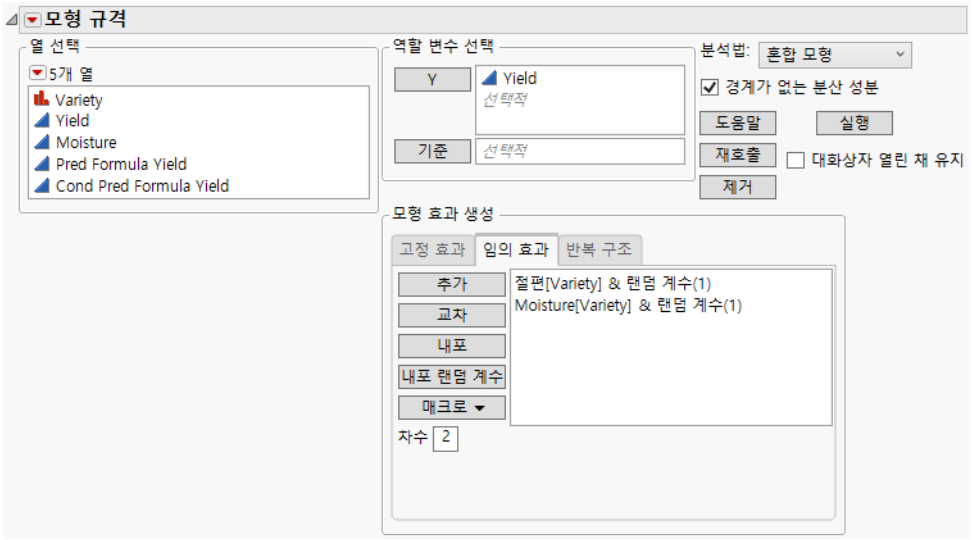
이 마지막 단계에서는 임의 효과의 수준 내에서 상관되는 랜덤 절편 및 랜덤 기울기 효과를 생성합니다. 개체는 개체 간의 변동성으로 인해 다른 개체에 내포됩니다. 절편이 모든 그룹에 대해 고정되어 있다고 간주하는 경우 **절편[그룹]&랜덤 계수(1)**을 선택한 후 **제거**를 클릭합니다.

계층적 선형 모형에서와 마찬가지로 이러한 방식으로 여러 랜덤 계수 그룹을 정의할 수 있습니다. 기울기 및 절편 계수에 대해 처리 효과별 랜덤 배치와 랜덤 배치 효과가 모두 있을 때 이 작업

이 필요할 수 있습니다. 또한 성취도 점수에 대해 랜덤 학생 효과와 랜덤 학교 효과가 있고 학생이 학교에 내포된 경우와 같은 계층적 선형 모형에서도 필요할 수 있습니다.

랜덤 계수는 비정형 공분산 구조를 사용하여 모델링됩니다. 그림 8.9에서는 Wheat.jmp 샘플 데이터의 랜덤 계수 지정을 보여 줍니다. 자세한 내용은 "예: 혼합 모형 사용"(330페이지)에서 확인하십시오.

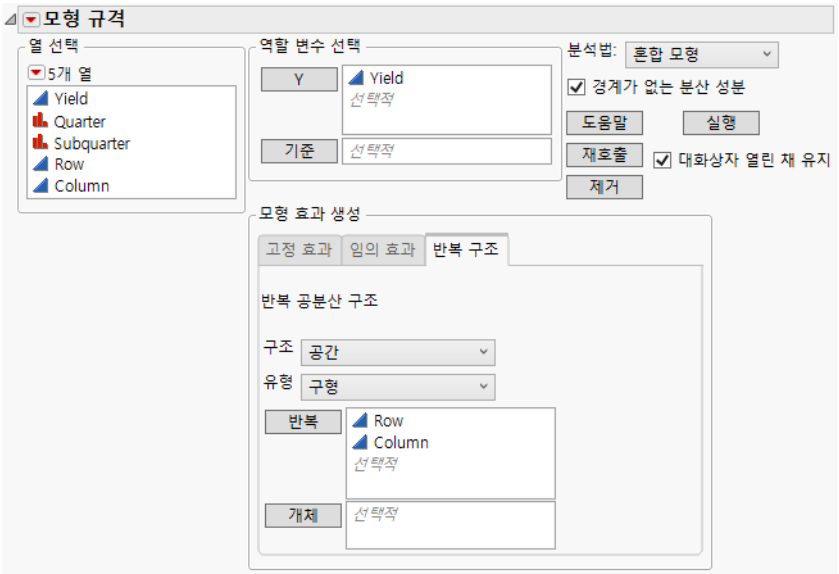
그림 8.9 랜덤 계수를 보여 주는 완료된 모형 적합 시작 창



JMP PRO 반복 구조 탭

"반복 구조" 탭을 사용하여 모형의 반복 효과에 대한 공분산 구조를 선택합니다.

그림 8.10 반복 구조 탭을 보여 주는 완료된 모형 적합 시작 창



구조

반복 구조는 기본적으로 "잔차"로 설정됩니다. 잔차 구조는 관측값 간에 공분산이 없음을 지정합니다. 즉, 오차가 독립적이라는 의미입니다. 잔차 구조와 이분산 구조 외에 다른 모든 공분산 구조도 관측값 간의 공분산을 모델링합니다. 구조에 대한 자세한 내용은 통계 상세 정보 섹션의 "반복 측정"(393 페이지) 및 "공간 및 시간 변동"(398 페이지)에서 확인하십시오.

표 8.1에는 사용 가능한 공분산 구조, 각 구조를 사용하기 위한 요구 사항 및 주어진 구조에 대한 공분산 모수의 수가 나열되어 있습니다. 관측 시간 수는 J로 표시됩니다.

표 8.1 반복 공분산 구조 요구 사항

구조	반복 열 유형	필요한 반복 열 수	개체	모수 수
잔차	해당 없음	0	해당 없음	0
이분산	범주형	1	선택적	J
비정형	범주형	1	필수	$J(J+1)/2$
AR(1)	연속형	1	선택적	2
복합 대칭	범주형	1	필수	2

표 8.1 반복 공분산 구조 요구 사항 (계속)

구조	반복 열 유형	필요한 반복 열 수	개체	모수 수
Antedependent 등분산	범주형		필수	*J*
Toeplitz	범주형	1	필수	*J*
복합 대칭 이분산	범주형	1	필수	*J*+1
Antedependent	범주형		필수	2*J*-1
Toeplitz 이분산	범주형	1	필수	2*J*-1
공간	연속형	2+	선택적	
공간 이방성	연속형	2+	선택적	
공간(너겟 사용)	연속형	2+	선택적	
공간 이방성(너겟 사용)	연속형	2+	선택적	

잔차 구조를 사용하여 "반복" 또는 "개체" 열을 입력하면 해당 열이 무시됩니다. 이 경우 "잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다"라는 경고가 표시됩니다.

유형

공간 공분산 구조 중 하나를 선택하면 공간 구조 유형을 선택할 수 있는 "유형" 목록이 나타납니다. 여기서 "거듭제곱", "지수","가우스" 및 "구형"의 네 가지 유형을 사용할 수 있습니다. 그림 8.10에서는 Uniformity Trial.jmp 샘플 데이터에 대해 공간 구조 및 구형 유형이 선택된 것을 보여줍니다.

반복

반복 측정 구조를 정의하는 열을 입력합니다. 반복 열의 모델링 유형은 공분산 구조에 따라 다릅니다. 각 반복 측정 공분산 구조의 요구 사항에 대한 자세한 내용은 표 8.1에서 확인하십시오.

개체

개체를 정의하는 하나 이상의 열을 입력합니다. 개체 열은 범주형이어야 합니다.

JMP PRO 데이터 형식

모형 적합 플랫폼에서 혼합 모형 분석법을 사용하려면 모든 반응 측정값이 하나의 반응 열에 포함되어야 합니다. 반복 측정 데이터가 여러 열에 기록되는 경우도 있습니다. 이때 각 행은 개체이고 반복 측정값은 별도의 반응 열에 기록됩니다. 혼합 모형 분석법을 실행하기 전에 이 형식의

데이터를 쌓아야 합니다. Cholesterol.jmp 및 Cholesterol Stacked.jmp 샘플 데이터 테이블에서는 각각 가로형 형식과 쌓인 형식을 보여 줍니다. 가로형 테이블의 각 행은 쌓인 테이블에서 Patient의 한 수준에 해당합니다.

JMP PRO 혼합 적합 보고서

"혼합 적합" 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

모형 보고서 혼합 모형 적합과 관련된 보고서를 생성합니다. 이러한 보고서에는 적합 통계량 및 모형 모수에 대한 추정값과 검정이 제공됩니다. 자세한 내용은 "모형 보고서"(342 페이지)에서 확인하십시오.

분산 성분의 선형 결합 분산 성분의 선형 결합에 대한 신뢰 구간을 계산할 수 있는 보고서를 표시합니다. 처음에는 이 보고서에 편집 가능한 텍스트 상자 및 모형의 분산 성분 테이블이 포함됩니다. 텍스트 상자를 사용하여 선형 결합에 라벨을 지정할 수 있습니다. 테이블에서 오른쪽 열의 셀에 값을 입력하여 신뢰 구간에 대한 선형 함수를 지정할 수 있습니다. 모수의 선형 결합을 지정하고 "완료" 를 클릭하면 지정된 선형 결합에 대한 신뢰 구간이 포함된 테이블이 나타납니다.

이 테이블에는 추정값, 표준 오차 및 두 가지 유형의 신뢰 구간(Satterthwaite, Wald)과 Wald *p* 값이 포함되어 있습니다. Wald *p* 값은 추정값이 0 과 다르다는 가설 검정에 해당합니다.

팁: Satterthwaite 신뢰 구간은 양수 값으로 제한되므로 지정된 계수가 음수인 경우에는 권장되지 않습니다. 추정값이 음수이면 Satterthwaite 신뢰 구간을 생성할 수 없으며 결측값으로 보고됩니다.

다중 비교 하나 이상의 효과 및 초기 비교를 선택할 수 있는 "다중 비교" 대화상자 창을 엽니다. 이 보고서는 범주형 고정 효과에 사용할 수 있습니다. 자세한 내용은 "다중 비교"(348 페이지)에서 확인하십시오.

기울기 비교 (명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있는 경우에만 사용 가능) ANCOVA(공분산 분석) 모형에서 교호작용 효과의 각 수준에 대한 기울기를 비교할 수 있는 보고서를 생성합니다. 자세한 내용은 "기울기 비교"(349 페이지)에서 확인하십시오.

역추정 예측 하나 이상의 반응 값에 대해 설명 변수 값을 예측합니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "역추정 예측"(131 페이지)에서 확인하십시오.

주변 모형 추론 주변 예측값과 주변 잔차를 기반으로 그림을 생성합니다. 이러한 그림에는 임의 효과로 인한 변동이 표시됩니다. 자세한 내용은 "주변 모형 추론"(349 페이지)에서 확인하십시오.

조건부 모형 추론 조건부 예측값과 조건부 잔차를 기반으로 그림을 생성합니다. 이러한 그림에는 임의 효과가 고려된 후 남아 있는 변동이 표시됩니다. 자세한 내용은 "조건부 모형 추론"(351 페이지)에서 확인하십시오.

열 저장 다양한 모형 결과를 데이터 테이블의 열로 저장하는 옵션이 포함되어 있습니다. 자세한 내용은 " 열 저장 "(354 페이지) 에서 확인하십시오 .

모형 대화상자 현재 분석을 위해 완료된 " 모형 적합 " 시작 창을 엽니다 . 자세한 내용은 " 모형 적합 시작 창 "(335 페이지) 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

모형 보고서

" 모형 보고서 " 에서 사용할 수 있는 보고서는 사용자가 수행하는 분석 유형에 따라 결정됩니다 . 이러한 보고서 중 일부는 기본적으로 표시됩니다 .

적합 통계량 모형 적합 통계량 보고서를 표시합니다 . 자세한 내용은 " 적합 통계량 "(343 페이지) 에서 확인하십시오 .

임의 효과 공분산 모수 추정값 임의 효과 공분산 모수 추정값 보고서를 표시합니다 . 이 보고서는 시작 창에서 임의 효과를 지정한 경우 나타납니다 . 자세한 내용은 " 임의 효과 공분산 모수 추정값 "(345 페이지) 에서 확인하십시오 .

고정 효과 모수 추정값 고정 효과 모수 추정값 보고서를 표시합니다 . 자세한 내용은 " 고정 효과 모수 추정값 "(345 페이지) 에서 확인하십시오 .

반복 효과 공분산 모수 추정값 반복 효과 공분산 모수 추정값 보고서를 표시합니다 . 이 보고서는 시작 창에서 반복 효과를 지정한 경우 나타납니다 . 자세한 내용은 " 반복 효과 공분산 모수 추정값 "(346 페이지) 에서 확인하십시오 .

표시자 파라미터화 추정값 (고정 효과 중 명목형 열이 있는 경우에만 사용 가능) " 표시 함수 파라미터화 " 보고서를 표시합니다 . 이 보고서는 명목형 고정 효과 열이 표시자 (SAS GLM) 파라미터화를 사용하여 코드화되고 연속형으로 처리되는 모형을 기반으로 고정 효과에 대한 모수 추정값을 제공합니다 . 순서형 열은 일반적인 JMP 코딩 체계를 사용하여 코드화됩니다 . SAS GLM 및 JMP 코딩 체계에 대한 설명은 " 통계 상세 정보 " 부록의 " 요인 모형 "(492 페이지) 에서 확인하십시오 .

주의 : " 표시 함수 파라미터화 " 보고서에 제공된 표준 오차 , t 비 및 기타 결과는 " 모수 추정값 " 보고서의 값과 다릅니다 . 이는 추정하는 모수가 서로 다르기 때문입니다 .

랜덤 계수 랜덤 계수 보고서를 표시합니다 . 이 보고서는 시작 창에서 임의 효과를 지정한 경우 나타납니다 . 자세한 내용은 " 랜덤 계수 "(347 페이지) 에서 확인하십시오 .

임의 효과 예측 임의 효과 예측 보고서를 표시합니다 . 이 보고서는 시작 창에서 임의 효과를 지정한 경우 나타납니다 . 자세한 내용은 " 임의 효과 예측 "(347 페이지) 에서 확인하십시오 .

고정 효과 검정 (고정 효과를 하나 이상 포함하는 모형에만 사용 가능) 고정 효과 검정을 표시합니다 . 이 보고서는 시작 창에서 고정 효과를 지정한 경우 나타납니다 . 자세한 내용은 " 고정 효과 검정 "(347 페이지) 에서 확인하십시오 .

순차 검정 (고정 효과를 하나 이상 포함하는 모형에만 사용 가능) 모형에 효과가 순차적으로 추가될 때 제곱합이 포함된 " 순차 (유형 1) 검정 " 보고서를 표시합니다 . 순차 제곱합을 기반으로 *F*- 검정을 수행합니다 . 자세한 내용은 " 순차 검정 "(348 페이지) 에서 확인하십시오 .

JMP PRO 적합 통계량

" 적합 통계량 " 보고서에는 모형 비교에 사용되는 통계량이 제공됩니다. 모든 적합 통계량은 값이 작을수록 좋습니다. 한 모형이 다른 모형에 포함된 경우 두 모형 간의 가능도비 검정을 수행할 수 있습니다. 그렇지 않은 경우 가능도를 비교하면 유용할 수 있습니다. 예는 " 공간 구조 모형 적합 "(376 페이지) 에서 확인하십시오 .

적합 통계량 보고서 설명에서는 다음과 같은 표기를 사용합니다 .

- 혼합 모형 지정

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon$$

여기서 $\mathbf{y}$ 는 관측값의 nx1 벡터 , β 는 고정 효과 모수의 벡터 , γ 는 임의 효과 모수의 벡터 , ε 는 오차의 벡터입니다 .

- γ 및 ε 벡터는 다음과 같은 다변량 정규 분포를 따른다고 가정합니다 .

$$E\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

및

$$Var\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

- 이러한 가정을 통해 $\mathbf{y}$ 의 분산은 다음과 같이 계산됩니다 .

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

적합 통계량 보고서 설명

-2* 잔차 로그 가능도 음의 잔차 로그 가능도 (목적 함수) 의 두 배인 최종 계산 결과입니다 .

$$-2 loglikelihood_{\mathbf{R}}(\mathbf{G},\mathbf{R}) = \log|\mathbf{V}| + \log\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right| + \mathbf{r}'\mathbf{V}^{-1}\mathbf{r} + (n-p)\log(2\pi)$$

다음은 각 요소에 대한 설명입니다 .

$$\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})$$

p 는 X 의 계수입니다 . 모형의 고정 효과 부분이 동일한 모형을 비교하는 경우에만 잔차 가능도를 사용합니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

-2* 로그 가능도 음의 로그 가능도 함수의 두 배인 계산 결과입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

모형에서 고정 효과 , 임의 효과 및 반복 효과가 서로 다른 모형을 비교하는 경우 로그 가능도를 사용합니다 .

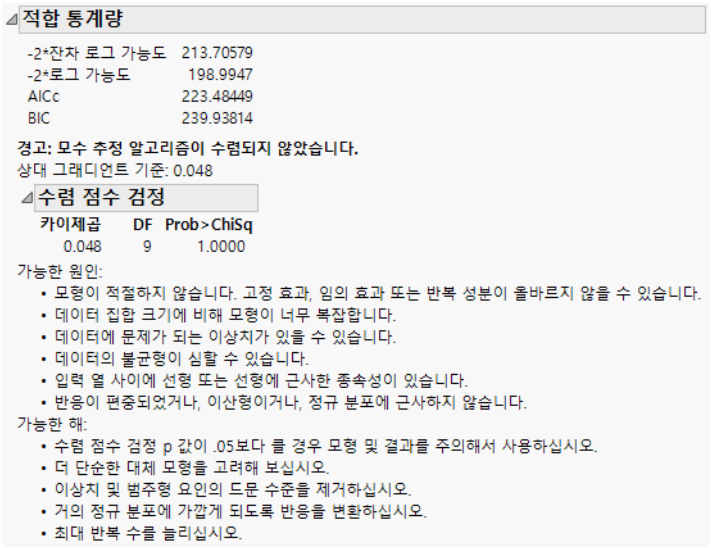
AICc 수정된 Akaike 정보 기준입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

BIC Bayes 정보 기준 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

JMP PRO 수렴 스코어 검정

모형 수렴에 문제가 있는 경우 적합 통계량 아래에 경고 메시지가 표시됩니다. 그림 8.11 에서는 수렴 문제의 원인과 가능한 해결 방법을 제안하는 경고를 보여 줍니다. 또한 최종 반복에서 상대 그래디언트 검정도 포함됩니다. 이 검정이 유의하지 않으면 모형이 올바르지만 수렴 기준에 완전히 도달하지 못할 수 있습니다. 이 경우 모형과 결과를 신중하게 사용해야 합니다. 자세한 내용은 "수렴 스코어 검정"(391 페이지) 에서 확인하십시오.

그림 8.11 수렴 스코어 검정



JMP PRO 임의 효과 공분산 모수 추정값

" 임의 효과 공분산 모수 추정값 " 보고서에는 모형에 지정한 임의 효과의 공분산 모수에 대한 상세 정보가 제공됩니다 .

공분산 모수 모형에 지정한 임의 효과의 모든 공분산 모수를 나열합니다 .

참고 : 모형에 분산 성분만 포함된 경우에는 이 열에 " 분산 성분 " 라벨이 지정됩니다 .

개체 공분산 행렬의 대각 블록을 구성한 개체를 나열합니다 .

추정값 효과에 대한 추정 분산 또는 공분산 성분을 제공합니다 .

참고 : 모형이 REML 모형과 동등한 경우 총 공분산 모수에 대한 행이 테이블에 추가됩니다 . 총 공분산 성분의 추정값은 양의 분산 성분만 합한 값입니다 .

표준 오차 공분산 성분 추정값에 대한 표준 오차를 제공합니다 .

95% 하한 공분산 성분에 대한 95% 신뢰 하한을 제공합니다 . 자세한 내용은 " 분산 성분에 대한 신뢰 구간 "(345 페이지) 에서 확인하십시오 .

95% 상한 공분산 성분에 대한 95% 신뢰 상한을 제공합니다 . 자세한 내용은 " 분산 성분에 대한 신뢰 구간 "(345 페이지) 에서 확인하십시오 .

Wald p 값 공분산 모수가 0 이라는 검정에 대한 p 값을 제공합니다 . 이 열은 " 모형 적합 " 시작 창에서 " 비유계 분산 성분 " 을 선택한 경우에만 나타납니다 .

JMP PRO 분산 성분에 대한 신뢰 구간

신뢰 한계를 계산하는 데 사용되는 방법은 " 모형 적합 " 시작 창에서 " 비유계 분산 성분 " 을 선택했는지 여부에 따라 다릅니다 . " 비유계 분산 성분 " 은 기본적으로 선택되어 있습니다 .

- " 비유계 분산 성분 " 을 선택한 경우 Wald 기반 신뢰 구간이 계산됩니다 . 이러한 구간은 점근적으로 유효하지만 표본이 작은 경우에는 신뢰할 수 없습니다 . 구간이 더 넓기 때문에 추정값이 0 과 유의한 차이가 없다고 잘못 생각할 수 있습니다 .
- " 비유계 분산 성분 " 을 선택하지 않은 경우 , 즉 모수의 하한 제약 조건이 0 인 경우에는 Satterthwaite 근사가 사용됩니다 (Satterthwaite 1946). 신뢰 구간의 하한도 0 으로 제한됩니다 .

JMP PRO 고정 효과 모수 추정값

" 고정 효과 모수 추정값 " 보고서에는 모형에 지정한 고정 효과 모수에 대한 상세 정보가 제공됩니다 . 각 모수에 대해 다음과 같은 상세 정보가 제공됩니다 .

- 추정값
- 표준 오차
- 추정값이 0 이라는 가설에 대한 t- 검정

- 추정값에 대한 95% 신뢰 구간

"고정 효과 모수 추정값" 보고서에는 다음 열이 포함됩니다.

항 추정된 모수에 해당하는 모형 항을 제공합니다. " 모형 적합 " 시작 창에서 " 절편 없음 " 옵션을 선택한 경우를 제외하고 첫 번째 항은 항상 절편입니다. 고차항의 일부인 연속형 열은 기본적으로 중심화됩니다 . 명목형 또는 순서형 효과는 대괄호 안에 수준 값이 포함된 상태로 나타납니다. 명목형 및 순서형 항의 코딩에 대한 자세한 내용은 " 요인 모형 "(492 페이지) 에서 확인하십시오 .

참고 : 연속형 열이 임의 효과에 포함된 경우에는 " 모형 규격 " 의 빨간색 삼각형 메뉴에서 " 중심화 다항식 " 옵션이 선택되어 있더라도 해당 열이 중심화되지 않습니다 .

추정값 각 항에 대한 모수 추정값을 제공합니다 . 이 값은 모형에 있는 항 계수의 추정값입니다 .

표준 오차 모수 추정값에 대한 표준 오차 추정값을 제공합니다 .

DFDen 효과 검정에 대한 분모 자유도, 즉 오차 자유도를 제공합니다. DFDen은 Kenward-Roger 1 차 근사를 사용하여 계산됩니다 . 자세한 내용은 "Kackar-Harville 수정 "(401 페이지) 에서 확인하십시오 .

t 비 모수의 실제 값이 0 인지 여부를 검정합니다 . *t* 비는 표준 오차에 대한 추정값의 비율입니다 . 모형에 대한 일반적인 가정이 주어지면 *t* 비는 귀무가설 하에서 스튜던트 *t* 분포를 따릅니다 .

Prob>|t| *t* 비의 양측 검정에 대한 *p* 값을 나열합니다 .

95% 하한 모수에 대한 95% 신뢰 하한을 표시합니다 .

95% 상한 모수에 대한 95% 신뢰 상한을 표시합니다 .

JMP PRO 반복 효과 공분산 모수 추정값

"반복 효과 공분산 모수 추정값" 보고서에는 모형에 지정한 반복 효과의 공분산 모수에 대한 상세 정보가 제공됩니다. 여기에는 각 모수에 대한 추정값, 표준 오차 및 95% 신뢰경계가 포함됩니다. 등방성 공간 모형의 경우 범위, 너겟 및 문턱 (sill) 값을 사용하여 공분산 모수 추정값을 해석합니다. 자세한 내용은 " 변동도 "(399 페이지) 에서 확인하십시오.

참고 : 분산은 효과 자체의 공분산입니다 .

" 반복 효과 공분산 모수 추정값 " 보고서에는 다음 열이 포함됩니다 .

공분산 모수 모형의 반복 효과에 대한 모든 공분산 모수를 나열합니다 .

추정값 효과에 대한 추정 분산 또는 공분산 성분을 제공합니다 .

표준 오차 분산 또는 공분산 성분 추정값에 대한 표준 오차를 제공합니다 .

95% 하한 공분산 성분에 대한 95% 신뢰 하한을 제공합니다 . 자세한 내용은 " 분산 성분에 대한 신뢰 구간 "(345 페이지) 에서 확인하십시오 .

95% 상한 공분산 성분에 대한 95% 신뢰 상한을 제공합니다 . 자세한 내용은 " 분산 성분에 대한 신뢰 구간 "(345 페이지) 에서 확인하십시오 .

JMP PRO 랜덤 계수

혼합 모형 분석법을 사용하면 각 임의 효과에 대해 추정된 계수를 보여 주는 보고서와 공분산 추정값 행렬을 보여 주는 보고서가 제공됩니다. 계수 보고서의 각 행은 임의 효과의 한 수준에 해당합니다. 임의 효과의 수준과 관련된 모든 계수 추정값이 행에 표시됩니다. 랜덤 계수 추정값은 고정 효과 추정값과 함께 사용되어 임의 효과의 특정 수준에 대한 예측을 생성합니다.

JMP PRO 임의 효과 예측

이 보고서에는 모형의 각 임의 효과에 대해 BLUP(**최량 선형 비편향 예측 변수**)라는 추정값, 표준 오차, 자유도 및 Satterthwaite 기반 신뢰 구간이 제공됩니다.

표준 오차를 추정하려면 BLUP 공분산 행렬을 계산해야 하며 이 작업은 시간이 많이 걸릴 수 있습니다. 계산 시간이 길어지면 진행률 표시줄이 나타납니다.

JMP PRO 고정 효과 검정

이 보고서에는 모형의 각 고정 효과에 대한 유의성 검정이 표시됩니다. 주어진 효과에 대한 검정은 해당 효과와 관련된 모든 모수가 0 이라는 귀무가설을 검정합니다. 단일 연속형 설명 변수의 경우 효과에 하나의 모수만 있을 수 있습니다. 이 경우 검정은 " 고정 효과 모수 추정값 " 보고서의 해당 항에 대한 *t*-검정과 동등합니다. 명목형 또는 순서형 효과에는 수준 수에 따라 관련된 모수가 여러 개 있을 수 있습니다. 이러한 효과에 대한 효과 검정은 관련된 모든 모수가 0 인지 여부를 검정합니다.

" 고정 효과 검정 " 보고서에는 다음 열이 포함됩니다 .

소스 모형의 고정 효과를 나열합니다 .

모수 수 효과와 관련된 모수의 수를 표시합니다 . 연속형 효과에는 하나의 모수가 있습니다 . 명목형 또는 순서형 효과에 대한 모수의 수는 수준 수에서 1 을 뺀 값입니다 . 교차 효과에 대한 모수의 수는 각 개별 효과에 대한 모수의 수를 곱한 값입니다 .

DFNum 효과 검정에 대한 분자 자유도를 표시합니다 .

DFDen 효과 검정에 대한 분모 자유도, 즉 오차 자유도를 표시합니다. DFDen 은 Kenward-Roger 1 차 근사를 사용하여 계산됩니다 . 자세한 내용은 "Kackar-Harville 수정 "(401 페이지) 에서 확인하십시오 .

F 비 효과가 0 인지 검정하기 위한 계산된 *F* 비를 제공합니다 .

Prob > F 효과 검정을 위한 *p* 값을 제공합니다 .

JMP PRO 순차 검정

이 보고서에는 고정 효과에 대한 순차(유형 1) 검정이 표시됩니다. 모형에 효과가 순차적으로 추가될 때 제곱합이 보고서에 포함됩니다. 입력 순서는 "모형 적합" 시작 창의 "모형 효과 생성" 목록에 나타나는 효과의 순서로 정의됩니다.

순차 검정의 기반이 되는 제곱합을 **제1종 제곱합**이라고도 합니다. 이 값은 지정된 효과 입력 순서에 따라 모형을 적합시켜 계산됩니다. 특정 효과를 예로 들겠습니다. 해당 효과 **전에** 입력한 모든 효과를 포함하는 모형에 대한 모형 제곱합을 계산합니다. 그런 다음 이러한 효과와 지정된 효과를 **모두** 포함하는 모형에 대한 모형 제곱합을 계산합니다. 지정된 효과에 대한 순차 제곱합은 모형 제곱합의 증가입니다.

다음과 같은 경우 순차 검정이 적절합니다.

- 적절한 순서로 지정된 분산 모형의 균형 분석 (즉 , 효과 목록에서 주효과 다음에 이원 교호작용이 나오는 순서)
- 적절한 순서로 지정된 순수 내포 모형
- 적절한 순서로 지정된 다항 회귀 모형

" 순차 (유형 1) 검정 " 보고서에는 다음 열이 포함됩니다 .

소스 모형의 고정 효과를 나열합니다 .

모수 수 효과와 관련된 모수의 수를 표시합니다 . 연속형 효과에는 하나의 모수가 있습니다 . 명목형 또는 순서형 효과에 대한 모수의 수는 수준 수에서 1 을 뺀 값입니다 . 교차 효과에 대한 모수의 수는 각 개별 효과에 대한 모수의 수를 곱한 값입니다 .

DFNum 효과 검정에 대한 분자 자유도를 표시합니다 .

DFDen 효과 검정에 대한 분모 자유도, 즉 오차 자유도를 표시합니다. DFDen은 Kenward-Roger 1 차 근사를 사용하여 계산됩니다 . 자세한 내용은 "Kackar-Harville 수정 "(401 페이지) 에서 확인하십시오 .

F 비 효과가 0 인지 검정하기 위한 계산된 *F* 비를 제공합니다 .

Prob > F 효과 검정을 위한 *p* 값을 제공합니다 .

JMP PRO 다중 비교

"다중 비교" 옵션은 주효과와 교호작용 효과의 최소 제곱 평균을 비교하는 다양한 방법을 제공합니다. 다중 비교 옵션에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "다중 비교"(117 페이지)에서 확인하십시오. 혼합 모형 예는 "6 월 (June) 의 모든 치료 (Treatment) 비교"(367 페이지) 및 "분할구 예"(370 페이지)에서 확인하십시오.

모형의 고정 효과 부분만 다중 비교에 사용됩니다. "다중 비교" 보고서에는 최소 제곱 평균 추정값, 표준 오차, 효과 없음에 대한 *t*-검정 및 95% 신뢰 구간이 표시됩니다. 이 보고서 다음에는 사용자가 선택한 여러 비교 검정이 나옵니다. "전체 쌍별 비교" 보고서에는 동등성 검정이 제공됩니다.

JMP PRO 기울기 비교

"기울기 비교" 옵션은 명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있을 때 나타납니다. 이 옵션은 ANCOVA(공분산 분석) 모형에서 기울기를 비교할 수 있는 보고서를 생성합니다. 이 보고서에서는 교호작용 효과의 각 수준에 대한 기울기를 전체 기울기와 비교합니다. 비교에는 ANOM(평균 분석)과 전체 평균이 함께 사용됩니다. ANOM(평균 분석) 보고서에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "전체 평균과 비교"(121 페이지)에서 확인하십시오.

전체 평균 기울기는 기울기의 가중 평균이며 이때 가중치는 기울기 추정값의 분산에 반비례합니다. 이러한 분산은 "전체 평균 기울기와의 차이" 테이블에서 "표준 오차" 열의 제곱 값입니다.

JMP PRO 주변 모형 추론

주변 모형 그림은 주변 예측값과 주변 잔차를 기반으로 합니다.

실제값 대 예측값 그림 임의 효과를 고려하지 않고 실제값 대 모형에서 예측한 값을 그림에 표시합니다. "실제값 대 예측값 그림"은 기본적으로 나타납니다. 자세한 내용은 "실제값 대 예측값 그림"(349 페이지)에서 확인하십시오.

잔차 그림 임의 효과를 고려하지 않고 모형 적합을 평가하는 잔차 그림을 제공합니다. 자세한 내용은 "잔차 그림"(349 페이지)에서 확인하십시오.

프로파일러, 등고선 프로파일러, 혼합물 프로파일러, 표면 프로파일러 임의 효과를 고려하지 않고 반응과 모형 항 간의 관계를 검토할 수 있는 프로파일러를 제공합니다. 자세한 내용은 "Profilers"(350 페이지)에서 확인하십시오.

JMP PRO 실제값 대 예측값 그림

"실제값 대 예측값 그림"은 기본적으로 나타납니다. 이 그림을 통해 임의 효과로 인한 변동을 반영하는 모형 적합을 시각적으로 평가할 수 있습니다. 여기에는 Y의 주변 예측값에 대해 Y의 관측값이 표시됩니다. 이러한 값은 열 저장 > 예측 계산식을 선택하여 구한 예측값입니다.

선형 혼합 모형은 $E[\mathbf{Y}|\gamma] = \mathbf{X}\beta + \mathbf{Z}\gamma$로 나타냅니다. 여기서 β는 고정 효과 계수의 벡터이고 γ는 임의 효과 계수의 벡터입니다. **주변 예측**은 $\mathbf{X}\hat{\beta}$로 지정된 예측 모형의 고정 효과 부분에서 구한 예측입니다.

JMP PRO 잔차 그림

주변 잔차는 고정 효과의 적합만 기반으로 하는 예측 오차를 반영합니다. 주변 잔차는 실제값과 열 저장 > 예측 계산식을 선택하여 구한 예측값 간의 차이입니다.

선형 혼합 모형은 $E[\mathbf{Y}|\gamma] = \mathbf{X}\beta + \mathbf{Z}\gamma$ 로 나타냅니다 . 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임의 효과 계수의 벡터입니다 . **주변 잔차**는 예측 모형의 고정 효과 부분에서 구한 잔차입니다 .

$$\mathbf{r} = \mathbf{Y} - \mathbf{X}\hat{\beta}$$

" 잔차 그림 " 옵션은 모형 적합을 평가할 수 있는 세 가지 시각적 방법을 제공합니다 .

잔차 대 예측값 그림 *Y* 의 예측값에 대해 그림에 표시된 잔차를 보여 줍니다 . 일반적으로 0 주위에 무작위로 흩어진 잔차 값을 원합니다 .

잔차 분위수 그림 표준 정규 분포의 분위수에 대해 그림에 표시된 잔차의 분위수를 보여 줍니다 . 잔차의 막대 차트도 표시됩니다 . 잔차가 정규 분포를 따르는 경우 정규 분위수 그림의 점은 거의 빨간색 주대각선을 따라 분포되어야 합니다 . 이 유형의 그림을 분위수 - 분위수 그림 또는 Q-Q 그림이라고도 합니다 . 정규 분위수 그림에는 Lilliefors 신뢰경계도 표시됩니다 (Conover 1999).

잔차 대 행 번호 그림 행 번호에 대해 그림에 표시된 잔차를 보여 줍니다 . 이 그림은 관측값의 행 순서로 인한 패턴을 탐지하는 데 도움이 될 수 있습니다 .

JMP PRO Profilers

"주변 모형 프로파일러"의 그림은 주변 예측값을 기반으로 합니다 . 이러한 값은 열 저장 > 예측 계산식을 선택하여 구한 예측값입니다 .

선형 혼합 모형은 $E[\mathbf{Y}|\gamma] = \mathbf{X}\beta + \mathbf{Z}\gamma$ 로 나타냅니다 . 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임의 효과 계수의 벡터입니다 . **주변 예측**은 $\mathbf{X}\hat{\beta}$ 로 지정된 예측 모형의 고정 효과 부분에서 구한 예측입니다 .

참고 : 주변 모형 프로파일러 그림에서는 고정 효과의 개별 설정에 대한 예측만 보여 줍니다 . " 프로파일러 " 에는 고정 효과에 해당하는 셀만 표시됩니다 . 임의 효과를 표시할 수 있는 다른 프로파일러에서는 고정 효과의 설정만으로 예측값을 결정합니다 .

다음과 같은 네 가지 유형의 프로파일러가 제공됩니다 .

- 프로파일러
- 등고선 프로파일러
- 혼합물 프로파일러
- 표면 프로파일러

적합 중인 모형에 적절한 옵션이 활성화됩니다 . 프로파일러의 예는 그림 8.21 에서 확인하십시오 . 표면 프로파일러의 예는 그림 8.42 에서 확인하십시오 . 표면 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

JMP PRO 조건부 모형 추론

조건부 모형 진단 그림은 조건부 잔차를 기반으로 합니다 . 조건부 잔차는 고정 효과와 임의 효과가 모두 적합된 후의 예측 오차를 반영합니다 .

실제값 대 조건부 예측값 그림 임의 효과를 고려하여 실제값 대 모형에서 예측한 값을 그림에 표시합니다 . 임의 효과가 있는 경우 " 실제값 대 조건부 예측값 그림 " 이 기본적으로 나타납니다 . 자세한 내용은 " 실제값 대 조건부 예측값 그림 "(351 페이지) 에서 확인하십시오 .

조건부 잔차 그림 임의 효과를 고려하여 모형 적합을 평가하는 잔차 그림을 제공합니다 . 자세한 내용은 " 조건부 잔차 그림 "(351 페이지) 에서 확인하십시오 .

조건부 프로파일러 , 조건부 등고선 프로파일러 , 조건부 혼합물 프로파일러 , 조건부 표면 프로파일러

임의 효과를 고려하여 반응과 모형 항 간의 관계를 검토할 수 있는 프로파일러를 제공합니다 . 자세한 내용은 " 조건부 프로파일러 "(352 페이지) 에서 확인하십시오 .

변동도 관측값 간의 거리가 증가할 때 공분산의 변화를 보여 주는 변동도 그림을 제공합니다 . " 잔차 " 구조를 선택하면 시간 또는 공간 좌표로 사용할 열을 선택할 수 있습니다 . 자세한 내용은 " 변동도 "(352 페이지) 에서 확인하십시오 .

JMP PRO 실제값 대 조건부 예측값 그림

"실제값 대 조건부 예측값 그림"은 기본적으로 나타납니다. 이 그림을 통해 임의 효과로 인한 변동을 설명하는 모형 적합을 시각적으로 평가할 수 있습니다. 여기에는 Y의 조건부 예측값에 대해 Y의 관측값이 표시됩니다. 이러한 값은 열 저장 > 조건부 예측 계산식을 선택하여 구한 예측값입니다.

선형 혼합 모형은 $E[\mathbf{Y}|\gamma] = \mathbf{X}\beta + \mathbf{Z}\gamma$로 나타냅니다. 여기서 β는 고정 효과 계수의 벡터이고 γ는 임의 효과 계수의 벡터입니다. **조건부 예측**은 $\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma}$로 지정된 모형에서 구한 예측입니다.

JMP PRO 조건부 잔차 그림

조건부 잔차는 고정 효과와 임의 효과 둘 다의 적합을 기반으로 하는 예측 오차를 반영합니다. 조건부 잔차는 실제값과 열 저장 > 조건부 예측 계산식을 선택하여 구한 조건부 예측값 간의 차이입니다.

선형 혼합 모형은 $E[\mathbf{Y}|\gamma] = \mathbf{X}\beta + \mathbf{Z}\gamma$ 로 나타냅니다 . 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임의 효과 계수의 벡터입니다 . **조건부 잔차**는 다음과 같이 계산됩니다 .

$$\mathbf{r} = \mathbf{Y} - (\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma})$$

" 조건부 잔차 그림 " 옵션은 모형 적합을 평가할 수 있는 세 가지 시각적 방법을 제공합니다 .

조건부 잔차 대 예측값 그림 Y 의 조건부 예측값에 대해 그림에 표시된 조건부 잔차를 보여 줍니다 . 일반적으로 0 주위에 무작위로 흩어진 조건부 잔차를 원합니다 .

조건부 잔차 분위수 그림 표준 정규 분포의 분위수에 대해 그림에 표시된 조건부 잔차의 분위수를 보여 줍니다. 조건부 잔차의 막대 차트도 표시됩니다. 조건부 잔차가 정규 분포를 따르는 경우 정규 분위수 그림의 점은 거의 빨간색 주대각선을 따라 분포되어야 합니다. 이 유형의 그림을 분위수-분위수 그림 또는 Q-Q 그림이라고도 합니다. 정규 분위수 그림에는 Lilliefors 신뢰경계도 표시됩니다 (Conover 1999).

조건부 잔차 대 행 번호 그림 행 번호에 대해 그림에 표시된 조건부 잔차를 보여 줍니다. 이 그림은 관측값의 행 순서로 인한 패턴을 탐지하는 데 도움이 될 수 있습니다.

JMP PRO 조건부 프로파일러

조건부 모형 프로파일러 그림은 조건부 예측값을 기반으로 합니다. 이러한 값은 열 저장 > 조건부 예측 계산식을 선택하여 구한 예측값입니다.

선형 혼합 모형은 $E[\mathbf{Y}|\gamma] = \mathbf{X}\beta + \mathbf{Z}\gamma$로 나타냅니다. 여기서 β는 고정 효과 계수의 벡터이고 γ는 임의 효과 계수의 벡터입니다. **조건부 예측**은 $\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma}$로 지정된 모형에서 구한 예측입니다.

다음과 같은 네 가지 유형의 프로파일러가 제공됩니다.

- 조건부 프로파일러
- 조건부 등고선 프로파일러
- 조건부 혼합물 프로파일러
- 조건부 표면 프로파일러

적합 중인 모형에 적절한 옵션이 활성화됩니다. 프로파일러의 예는 그림 8.21에서 확인하십시오. 표면 프로파일러의 예는 그림 8.42에서 확인하십시오. 표면 프로파일러에 대한 자세한 내용은 Profilers의 에서 확인하십시오.

JMP PRO 변동도

변동도 그림은 관측값의 거리를 기반으로 관측값의 공간 또는 시간 상관관계를 설명합니다. 이 그림에서는 반분산이 거리 또는 시간의 함수로 표시됩니다. 이론적으로 반분산은 지정된 거리만큼 떨어진 위치에서 반응 값 간의 차이에 대한 분산의 2분의 1입니다. 지정된 거리에서 반분산과 상관관계는 반비례 관계를 나타냅니다. 지정된 거리에서 값 간의 상관관계가 작으면 해당 거리에서 관측값 간의 반분산이 큽니다.

모형 적합 창에서 등방성 반복 구조(AR(1), 공간 또는 공간(너겟 사용))를 지정하면 변동도 그림이 기본적으로 표시됩니다. 잔차 구조를 지정하는 경우 빨간색 삼각형 메뉴에서 "변동도" 옵션을 선택하면 변동도 계산에 사용할 연속형 열을 선택할 수 있습니다. 데이터의 공간 또는 시간 구조를 설명하는 열을 개수 제한 없이 포함할 수 있습니다.

초기 변동도 보고서에는 거리에 대한 경험적 반분산 그림이 표시됩니다. 배경 정보를 포함한 자세한 내용은 "Antedependent 공분산 구조"(397 페이지)에서 확인하십시오.

등방성 구조에 대한 반분산 곡선

반분산 곡선은 AR(1), 거듭제곱, 지수, 가우스 및 구형과 같은 등방성 공분산 구조에 대해 제공됩니다. 공간 구조의 경우 너겟을 사용한 모형과 사용하지 않은 모형에 대해 곡선이 제공됩니다.

이론적 모형에 대한 곡선은 공분산 모수 추정값을 사용하여 적합됩니다. 기본 계산식은 Chilès and Delfiner 연구 자료 (2012) 및 Cressie 연구 자료 (1993) 에서 확인하십시오.

이론적 모형을 사용하여 데이터가 선택된 등방성 구조에 맞는지 확인할 수 있습니다. 잔차 구조를 선택한 경우 경험적 변동도를 사용하여 데이터가 시간 또는 공간 구조를 나타내는지 확인할 수 있습니다. 점이 수평선을 따르면 거리에 따라 상관관계가 변하지 않으며 잔차 구조가 적절하다는 것을 의미합니다. 점이 패턴을 표시하는 경우 다양한 등방성 모형을 적합시키면 모형을 재적합하는 데 적절한 반복 구조가 제안될 수 있습니다.

JMP PRO 너겟

너겟은 변동도 원점의 값 0 에서 매우 작은 분리 거리의 반분산 값으로 수직 이동을 나타냅니다. 너겟을 사용한 변동도 모형은 원점에서 불연속성을 가집니다. 0 바로 위의 거리에 대한 이론적 곡선의 값이 너겟입니다.

JMP PRO 변동도 옵션

AR(1) AR(1) 공분산 구조에 대한 변동도를 그림에 표시합니다 .

공간 지수 , 가우스 , 거듭제곱 또는 구형 공분산 구조에 대한 변동도를 그림에 표시합니다 .

공간 (너겟 사용) 너겟을 사용하여 지수 , 가우스 , 거듭제곱 또는 구형 공분산 구조에 대한 변동도를 그림에 표시합니다 .

JMP PRO 공분산 및 상관 행렬

" 공분산 및 상관 행렬 " 메뉴에는 모형과 관련된 공분산 행렬을 볼 수 있는 옵션이 포함되어 있습니다 .

고정 효과의 공분산 모형의 고정 효과에 대한 공분산 행렬을 표시하거나 숨깁니다 .

공분산 모수의 공분산 모형의 임의 효과에 대한 공분산 행렬을 표시하거나 숨깁니다 . 행렬의 효과는 G 측 임의 효과 , R 측 임의 효과 , 잔차 효과의 순서로 정렬됩니다 .

모든 모수의 공분산 모형의 모든 효과에 대한 공분산 행렬을 표시하거나 숨깁니다 . 행렬의 효과는 고정 효과 , G 측 임의 효과 , R 측 임의 효과 , 잔차 효과의 순서로 정렬됩니다 .

고정 효과의 상관관계 모형의 고정 효과에 대한 상관 행렬을 표시하거나 숨깁니다 .

JMP PRO 열 저장

" 열 저장 " 메뉴의 각 옵션은 하나 이상의 새 열을 데이터 테이블에 추가합니다 .

예측 계산식 계산식과 주변 평균 예측값이 모두 포함된 **예측 계산식 < 열 이름 >** 이라는 새 열을 생성합니다 . " 예측 " 열 특성이 추가되어 예측 소스를 나타냅니다 . 자세한 내용은 " 주변 모형 추론 "(349 페이지) 에서 확인하십시오 .

예측값의 표준 오차 주변 예측 평균 반응에 대한 표준 오차가 포함된 **예측 표준 오차 < 열 이름 >** 이라는 새 열을 생성합니다 .

평균 신뢰 구간 **평균 < 열 이름 > 의 95% 하한** 및 **평균 < 열 이름 > 의 95% 상한**이라는 두 개의 새 열을 생성합니다 . 이 열에는 평균 반응에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다 . 이 구간에는 추정의 변동이 포함되고 반응의 변동은 포함되지 않습니다 .

개별값 신뢰 구간 (G 측 효과만 포함된 모형에 사용 가능) **개별값 < 열 이름 > 의 95% 하한**과 **개별값 < 열 이름 > 의 95% 상한**이라는 두 개의 새 열을 생성합니다 . 이 열에는 개별 반응 값에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다 . 이 구간에는 반응과 추정의 변동이 모두 포함됩니다 .

잔차 반응 관측값에서 주변 평균 예측값을 뺀 값이 포함된 **잔차 < 열 이름 >** 이라는 새 열을 생성합니다 . 자세한 내용은 " 주변 모형 추론 "(349 페이지) 에서 확인하십시오 .

조건부 예측 계산식 계산식과 조건부 평균 예측값이 모두 포함된 **조건부 예측 계산식 < 열 이름 >** 이라는 새 열을 생성합니다 . " 예측 " 열 특성이 추가되어 예측 소스를 나타냅니다 . 자세한 내용은 " 조건부 프로파일러 "(352 페이지) 에서 확인하십시오 .

조건부 예측 표준 오차 조건부 예측 평균 반응에 대한 표준 오차가 포함된 **조건부 예측 표준 오차 < 열 이름 >** 이라는 새 열을 생성합니다 .

조건부 평균 CI (G 측 효과가 포함된 모형에 사용 가능) **조건부 평균 < 열 이름 > 의 95% 하한** 및 **조건부 평균 < 열 이름 > 의 95% 상한**이라는 두 개의 새 열을 생성합니다 . 이 열에는 조건부 예측의 기대값에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다 . 이 신뢰 구간에는 임의 효과가 있는 모형의 임의 효과 추정값이 포함됩니다 . 자세한 내용은 " 조건부 모형 추론 "(351 페이지) 에서 확인하십시오 .

조건부 잔차 반응 관측값에서 조건부 평균 예측값을 뺀 값이 포함된 **조건부 잔차 < 열 이름 >** 이라는 새 열을 생성합니다 . 자세한 내용은 "조건부 모형 추론"(351 페이지) 에서 확인하십시오 .

시뮬레이션 계산식 저장 (분산 성분 및 랜덤 계수 모형에만 사용 가능) 적합 모형에 대한 추정 모수를 사용하여 시뮬레이션 값을 생성하는 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 이 열은 " 시뮬레이션 " 유틸리티에서 " 스위치 인할 열 " 로 사용할 수 있습니다 . 자세한 내용은 **기본 분석**의 에서 확인하십시오 .

JMP PRO 혼합 모형 분석법의 추가 예

- " 반복 측정 예 "

- "분할구 예"
- "공간 예 : 균일도 시험"
- "상관관계가 있는 반응의 예"

JMP PRO 반복 측정 예

Cholesterol Stacked.jmp 샘플 데이터 테이블을 예로 들겠습니다. 대조 약물에 대해 두 가지 새 콜레스테롤 약물을 테스트하기 위한 연구가 수행되었습니다. 콜레스테롤 수치가 높은 20명의 환자를 각각 네 가지 치료(두 가지 실험 약물, 대조 약물, 플라시보)에 무작위로 할당했습니다. 연구 기간 동안 각 환자의 총 콜레스테롤 수치를 6번 측정했습니다(4월 1일, 5월 1일, 6월 1일 아침과 오후 한 번씩). 이 연구를 통해 신약이 콜레스테롤 수치를 낮추는 데 효과가 있는지, 그리고 시간과 치료 간에 상관관계가 있는지 여부에 관심이 있습니다.

JMP PRO 배경

일반적으로 이러한 설계를 분석하기 위해 다음과 같은 두 가지 방법이 사용되었습니다 .

- MANOVA(다변량 분산 분석)
- 수정된 Huynh-Feldt(1976) 또는 Greenhouse-Geisser(1959) 연구 자료를 사용한 시간 단변량 ANOVA(분산 분석) 의 분할구

두 옵션 모두 모형 적합의 MANOVA 분석법을 통해 사용할 수 있습니다. 이러한 양극단 두 옵션은 공분산 구조를 모델링하기 위한 것입니다. MANOVA 분석에서는 모든 변수와 공분산이 개별적으로 추정되는 비정형 공분산 구조를 가정합니다. 시간 분석의 독립 분할구에서는 모든 오차가 독립적이라고 가정합니다. 이는 가우스 데이터에 대해 복합 대칭 공분산 구조를 가정하는 것에 해당합니다.

이 두 모형은 처리 효과에 대해 상당히 다른 결론을 내릴 수 있습니다 . 복소 공분산 구조를 가정하면 데이터의 정보를 사용하여 공분산 모수를 추정합니다 . 너무 많은 공분산 모수를 적합시키면 모형 과대적합이 발생할 위험이 있습니다 . 반복 측정 데이터를 모델링하는 경우 이러한 문제를 상쇄하는 공분산 구조를 찾아야 합니다 .

- 모형이 과대적합되면 덜 복잡한 공분산 구조를 가정할 때보다 차이를 감지하는 검정력이 더 작습니다 .
- 모형이 과소적합되면 제 1 종 오류 제어가 손실됩니다 . 경우에 따라 기각률이 부풀려질 수도 있습니다 . 그렇지 않은 경우 분산 팽창으로 인해 기각률이 감소합니다 .

JMP PRO 공분산 구조

혼합 모형 분석법은 다양한 공분산 구조를 적합시킵니다. 시간에 따른 반복 측정의 경우 대개 Toeplitz 공분산 구조와 AR(1)(1차 자기회귀) 공분산 구조 모두 적절한 상관 구조를 제공합니다. 이러한 구조는 모형을 과대적합시키지 않고도 상관관계가 있는 관측값을 허용합니다. AR(1)은 공통 분산 모수를 가정하는 반면, 이분산을 사용한 Toeplitz 공분산 행렬은 반복 측정 변수의 각

단위에 대해 고유한 분산을 추정합니다. 자세한 내용은 "반복 공분산 구조 요구 사항"(339 페이지)에서 확인하십시오.

이 예에서는 네 가지 공분산 구조를 적합시킵니다. 관측 시간 수 *J* 는 6 입니다.

- " 공분산 구조 : 비정형 "(356 페이지). 비정형 모형은 모든 공분산 모수 (*J*(*J*+1)/2) 를 합하여 적합시킵니다. 이 예에서 모형은 21 개의 분산을 적합시킵니다.
- " 공분산 구조 : 잔차 "(359 페이지). 잔차 모형은 일반적인 분산 성분 구조와 동등합니다. 이 예에서 모형은 두 개의 분산을 적합시킵니다.
- " 공분산 구조 : Toeplitz"(361 페이지). Toeplitz 모형은 2*J*-1 개의 공분산 모수를 적합시킵니다. 이 예에서 모형은 11 개의 분산을 적합시킵니다.
- " 공분산 구조 : AR(1)"(363 페이지). 이 모형은 두 개의 공분산 모수를 적합시킵니다. 한 모수는 분산을 결정하고 다른 모수는 시간에 따라 공분산이 어떻게 변하는지 결정합니다.

AICc 를 사용하여 모형 적합을 평가합니다. BIC 기준도 사용할 수 있습니다. 이 경우 두 기준에 의해 선택된 모형이 동일합니다. 최적의 공분산 구조를 선택한 후 추가 분석을 계속합니다.

- "AR(1) 구조를 사용한 추가 분석 "(365 페이지)
- "AR(1) 에 대한 회귀 모형의 예 "(368 페이지)

팁 : 이 예를 진행하는 동안 모형 적합 시작 창을 열어 두십시오.

JMP PRO 데이터 구조

Cholesterol.jmp 데이터 테이블은 반복 측정 데이터를 기록하는 데 일반적으로 사용되는 형식으로 되어 있습니다. 혼합 모형 분석법을 사용하여 이러한 데이터를 분석하려면 각 콜레스테롤 측정값이 Cholesterol Stacked.jmp 와 같이 고유의 행에 포함되어야 합니다. Cholesterol Stacked.jmp 를 구성하기 위해 테이블 > 쌓기를 사용하여 Cholesterol.jmp 의 데이터를 쌓았습니다.

쌓인 테이블의 Days 열은 계산식을 사용하여 생성되었습니다. Days 열은 콜레스테롤 수치를 측정한 시간(일 수)을 연구에 제공합니다. 모델링 유형은 연속형입니다. AR(1) 공분산 구조에는 반복 효과가 연속형이어야 하므로 이 유형이 필요합니다.

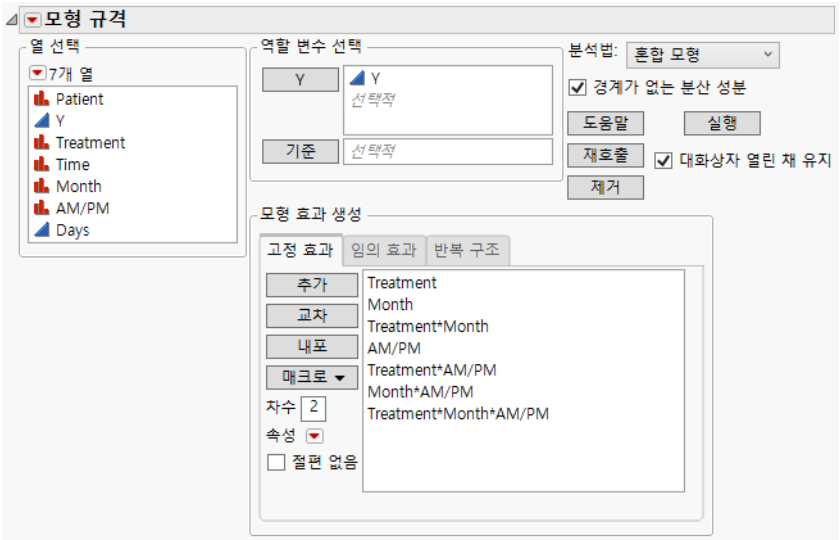
JMP PRO 공분산 구조 : 비정형

먼저 비정형 공분산 구조를 사용하여 모형을 적합시킵니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cholesterol Stacked.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. 다음 예에서 시작 창으로 돌아갈 수 있도록 **대화상자 열린 채 유지**를 선택합니다.
4. Y 를 선택하고 **Y** 를 클릭합니다.
5. " 분석법 " 목록에서 **혼합 모형**을 선택합니다.

6. Treatment, Month, 및 AM/PM 을 선택한 후 **매크로 > 완전 요인**을 선택합니다 .

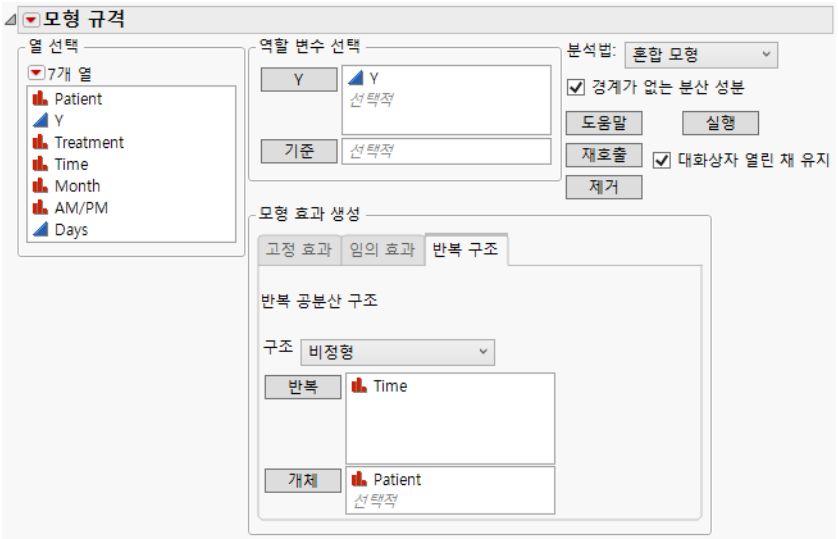
그림 8.12 완료된 고정 효과 탭을 보여 주는 모형 적합 시작 창



7. **반복 구조** 탭을 선택합니다 .
8. " 구조 " 목록에서 **비정형**을 선택합니다 .
9. Time 을 선택하고 **반복**을 클릭합니다 . **반복** 열은 개체 내에서 반복 측정을 정의합니다 .
10. Patient 를 선택하고 **개체**를 클릭합니다 .

참고 : 비정형 공분산 모형에서는 반복 구조 변수가 중복 값을 가질 수 없습니다 . 이 예에서는 개체가 치료 (treatment) 에 내포되고 , 각 치료 내에서 1, 2, 3, 4, 5 값을 사용하여 환자 (patient) 에 번호가 매겨져 있다고 가정합니다 . 이 분석을 실행하면 경고가 표시됩니다 . 반복 변수의 각 값에 서로 다른 식별자를 지정하기 위해 환자 번호를 다시 매겨야 합니다 . 또는 데이터 테이블에 치료 내의 내포를 나타내는 열을 생성하고 이 효과를 개체로 입력해야 합니다 .

그림 8.13 완료된 반복 구조 탭을 보여 주는 모형 적합 시작 창



11. **실행**을 클릭합니다 .

    그림 8.14 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . AICc 또는 BIC 를 사용하여 세 가지 모형을 비교하려고 하므로 " 적합 통계량 " 보고서를 확인해야 합니다 . 비정형 모형의 AICc 는 703.84 입니다 .

    " 반복 효과 공분산 모수 추정값 " 보고서에서는 21 개의 모든 공분산 모수에 대한 추정값을 보여 줍니다 . 예상대로 더 가까운 시간의 관측값이 더 멀리 떨어진 관측값보다 공분산이 큽니다 . 또한 시간에 따라 분산이 증가합니다 .

그림 8.14 혼합 적합 보고서 - 비정형 공분산 구조

혼합 적합

실제값 대 예측값 그림

적합 통계량

-2*잔차 로그 가능도	554.98818
-2*로그 가능도	557.89101
AICc	703.83696
BIC	773.32814

반복 효과 공분산 모수 추정값

반복 효과: Time
개체: Patient

공분산 모수	추정값	표준 오차	95% 하한	95% 상한
분산(April AM)	18.725	6.6202872	5.7494754	31.700525
공분산(April PM,April AM)	18.354884	6.6035621	5.4121399	31.297628
분산(April PM)	19.268932	6.8125963	5.9164888	32.621376
공분산(May AM,April AM)	9.2756709	8.4629182	-7.311344	25.862686
공분산(May AM,April PM)	5.5074058	8.3704001	-10.89828	21.913089
분산(May AM)	56.603347	20.012305	17.379949	95.826744
공분산(May PM,April AM)	9.4147226	8.5038584	-7.252534	26.081979
공분산(May PM,April PM)	6.6230805	8.4532303	-9.944946	23.191108
공분산(May PM,May AM)	55.365523	19.83529	16.489068	94.241978
분산(May PM)	57.058089	20.173081	17.519577	96.596602
공분산(June AM,April AM)	1.1945478	8.6266098	-15.7133	18.102392
공분산(June AM,April PM)	0.3183447	8.7461245	-16.82374	17.460434
공분산(June AM,May AM)	1.106725	14.992149	-28.27735	30.490796
공분산(June AM,May PM)	0.6455101	15.050552	-28.85303	30.14405
분산(June AM)	63.512277	22.454981	19.501323	107.52323
공분산(June PM,April AM)	1.5810194	8.7687993	-15.60551	18.76755
공분산(June PM,April PM)	0.7647113	8.8882627	-16.65596	18.185386
공분산(June PM,May AM)	0.9262595	15.232066	-28.92804	30.780561
공분산(June PM,May PM)	0.6543895	15.292238	-29.31785	30.626625
공분산(June PM,June AM)	63.878686	22.700344	19.386829	108.37054
분산(June PM)	65.568482	23.181959	20.132677	111.00429

고정 효과 모수 추정값

고정 효과 검정

소스	N모수	DFNum	DFDen	F 비	Prob > F
Treatment	3	3	16.0	274.96713	<.0001*
Month	2	2	15.0	340.48166	<.0001*
Treatment*Month	6	6	18.3	123.47461	<.0001*
AM/PM	1	1	16.0	360.93593	<.0001*
Treatment*AM/PM	3	3	16.0	0.6339843	0.6038
Month*AM/PM	2	2	15.0	1.1988247	0.3289
Treatment*Month*AM/PM	6	6	18.3	1.1642781	0.3671

JMP PRO 공분산 구조 : 잔차

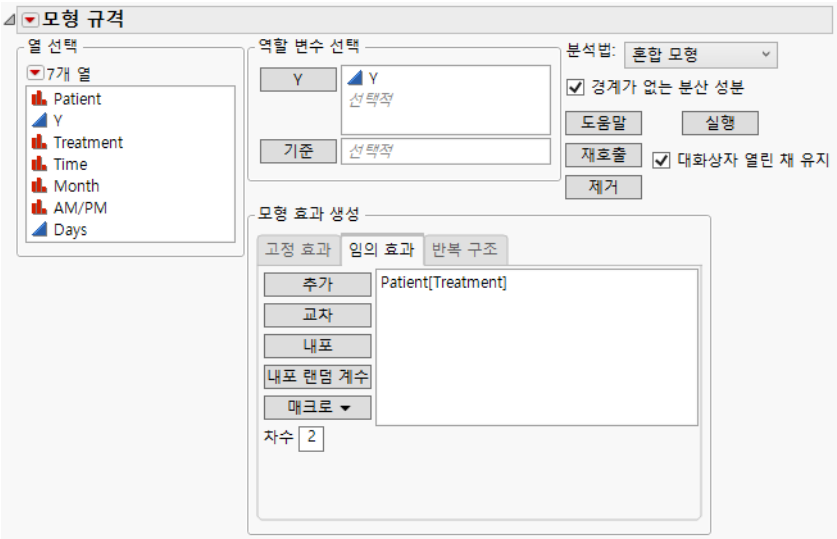
잔차 공분산 구조는 분할구 모형을 적합시킬 때 적절합니다 .

1. " 공분산 구조 : 비정형 "(356 페이지) 에 나오는 1 단계 ~ 7 단계를 완료합니다 .
2. **반복 구조** 탭의 " 구조 " 목록에서 **잔차**를 선택합니다 .
3. 이전 예에서 계속하는 경우 Time 과 Patient 를 제거합니다 .

   그렇지 않으면 " 잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다 " 라는 경고가 표시됩니다 . **확인**을 클릭하여 분석을 계속할 수도 있습니다 .
4. **임의 효과** 탭을 선택합니다 .

5. Patient 를 선택하고 **추가**를 클릭합니다 .
6. " 임의 효과 " 영역에서 Patient 를 선택하고 Treatment 열을 선택한 후 **내포**를 클릭합니다 .

그림 8.15 완료된 임의 효과 탭을 보여 주는 모형 적합 시작 창



7. **실행**을 클릭합니다 .

   그림 8.16 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . " 적합 통계량 " 보고서에 따르면 잔차 모형의 AICc 는 832.55 이며 이는 비정형 모형의 703.84 와 비교됩니다 .

   " 임의 효과 공분산 모수 추정값 " 보고서에는 두 공분산 모수의 추정값이 표시됩니다 . 이러한 값은 Treatment 에 내포된 Patient 의 분산 및 잔차 분산의 추정값입니다 .

그림 8.16 혼합 적합 보고서 - 잔차 공분산 구조

혼합 적합

실제값 대 예측값 그림 | 실제값 대 조건부 예측값 그림

적합 통계량

-2*잔차 로그 가능도	721.03949
-2*로그 가능도	765.45515
AICc	832.55192
BIC	889.92993

임의 효과 공분산 모수 추정값

분산 성분	추정값	표준 오차	95% 하한	95% 상한	Wald p 값
Patient[Treatment]	11.707432	6.2749012	-0.591148	24.006012	0.0621
잔차	35.081923	5.546939	26.320843	49.105827	
합계	46.789355	7.7386523	34.678063	66.607248	

고정 효과 모수 추정값

랜덤 계수

고정 효과 검정

소스	N모수	DFNum	DFDen	F 비	Prob > F
Treatment	3	3	16.0	274.96713	<.0001*
Month	2	2	80.0	622.88066	<.0001*
Treatment*Month	6	6	80.0	226.49464	<.0001*
AM/PM	1	1	80.0	13.700114	0.0004*
Treatment*AM/PM	3	3	80.0	0.0240643	0.9949
Month*AM/PM	2	2	80.0	0.0324977	0.9680
Treatment*Month*AM/PM	6	6	80.0	0.030708	0.9999

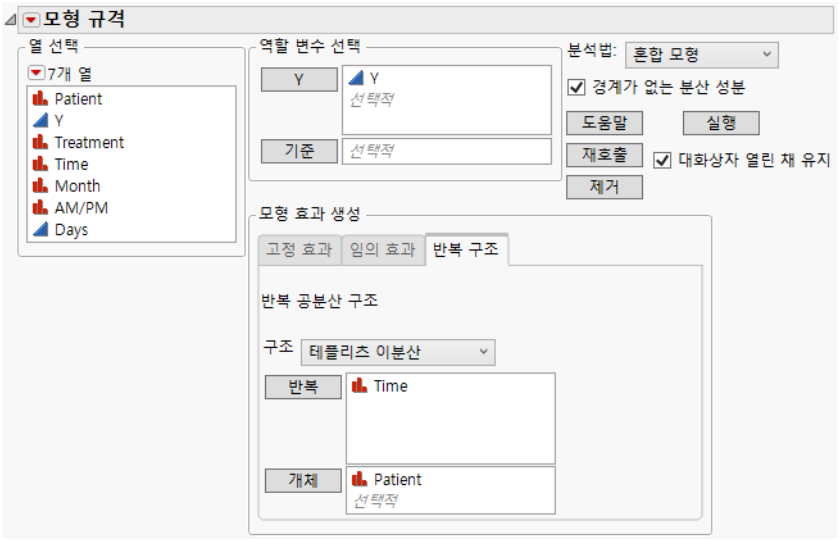
JMP PRO 공분산 구조 : Toeplitz

Toeplitz 이분산 구조를 사용하여 모형을 적합시킵니다 .

1. " 공분산 구조 : 비정형 "(356 페이지) 에 나오는 1 단계 ~ 6 단계를 완료합니다 .
2. 이전 예에서 계속하는 경우 **임의 효과** 탭의 Patient[Treatment] 를 선택한 후 **제거**를 클릭합니다 .

   임의 효과와 반복 효과를 모두 포함하면 데이터가 부족하여 효과를 둘 다 추정하지 못하는 경우가 많습니다 .
3. **반복 구조** 탭을 선택합니다 .
4. " 구조 " 목록에서 **Toeplitz 이분산**을 선택합니다 .
5. Time 을 선택하고 **반복**을 클릭합니다 .
6. Patient 를 선택하고 **개체**를 클릭합니다 .

그림 8.17 완료된 반복 구조 탭을 보여 주는 모형 적합 시작 창



7. **실행**을 클릭합니다 .

그림 8.18 혼합 적합 보고서 - Toeplitz 이분산 구조

혼합 적합

실제값 대 예측값 그림

적합 통계량

-2*잔차 로그 가능도	659.09915
-2*로그 가능도	688.02972
AICc	788.02972
BIC	855.59193

반복 효과 공분산 모수 추정값

개체: Patient

공분산 모수	추정값	표준 오차	95% 하한	95% 상한
테플 상관(1)	0.6230225	0.0856377	0.4551756	0.7908693
테플 상관(2)	0.2555783	0.1463264	-0.031216	0.5423728
테플 상관(3)	0.0962006	0.1887883	-0.273818	0.466219
테플 상관(4)	-0.1112	0.2248683	-0.551934	0.3295338
테플 상관(5)	0.2427218	0.1955609	-0.140571	0.6260141
분산(April AM)	17.552396	6.9782256	3.8753253	31.229467
분산(April PM)	24.652754	8.3215651	8.3427857	40.962721
분산(May AM)	73.950898	25.930509	23.128035	124.77376
분산(May PM)	63.668022	21.766899	21.005684	106.33036
분산(June AM)	69.45679	22.578117	25.204495	113.70909
분산(June PM)	52.484909	19.171565	14.909332	90.060486

고정 효과 모수 추정값

고정 효과 검정

소스	N모수	DFNum	DFDen	F 비	Prob > F
Treatment	3	3	14.1	230.25344	<.0001*
Month	2	2	23.7	363.8599	<.0001*
Treatment*Month	6	6	26.2	134.01838	<.0001*
AM/PM	1	1	5.2	106.73459	0.0001*
Treatment*AM/PM	3	3	5.2	0.1874794	0.9007
Month*AM/PM	2	2	26.9	0.036706	0.9640
Treatment*Month*AM/PM	6	6	34.6	0.0347301	0.9998

참고 : JMP 의 혼합 모형 분석법은 상관관계를 보고하는 반면 SAS 의 PROC MIXED 는 공분산을 보고합니다 .

"적합 통계량" 보고서에 따르면 이분산을 사용한 Toeplitz 모형의 AICc는 788.03입니다. 이 값을 잔차 모형의 832.55 및 비정형 모형의 703.84와 비교합니다.

Toeplitz 이분산 구조를 사용하는 경우 11 개의 공분산 모수를 추정해야 합니다. 이러한 추정값은 "반복 효과 공분산 모수 추정값" 보고서에 표시됩니다. Toeplitz 상관 추정값 다음에 각 시점에 대한 분산 추정값이 나옵니다. 이 행렬이 파라미터화되는 방법에 대한 자세한 내용은 "반복 측정"(393 페이지)에서 확인하십시오.

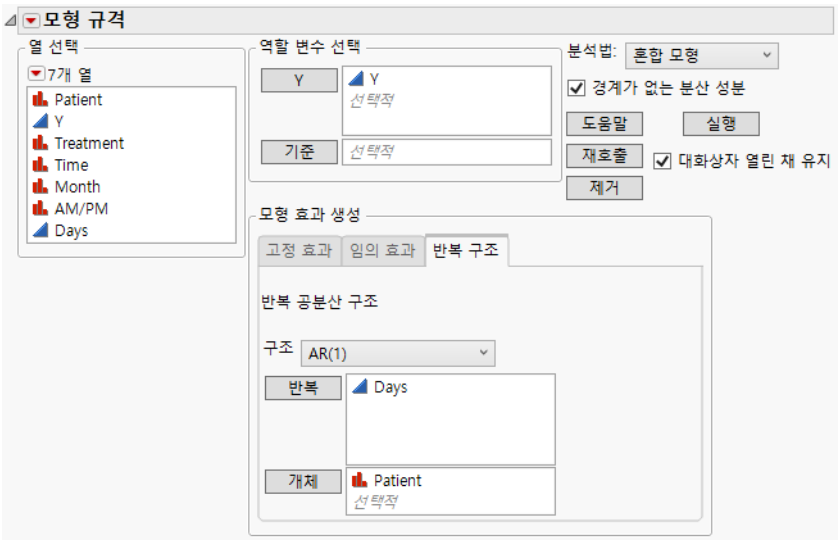
JMP PRO 공분산 구조 : AR(1)

마지막으로 AR(1) 구조를 적합시킵니다 .

1. " 공분산 구조 : 비정형 "(356 페이지) 에 나오는 1 단계 ~ 6 단계를 완료합니다 .
2. 이전 예에서 계속하는 경우 **반복** 상자의 Time 을 선택한 후 **제거**를 클릭합니다 .
   AR(1) 을 사용하는 경우 반복 값에 대한 변수가 연속형이어야 합니다 .

3. " 구조 " 목록에서 **AR(1)** 을 선택합니다 .
4. Days 를 선택하고 **반복**을 클릭합니다 .

그림 8.19 완료된 반복 구조 탭을 보여 주는 모형 적합 시작 창



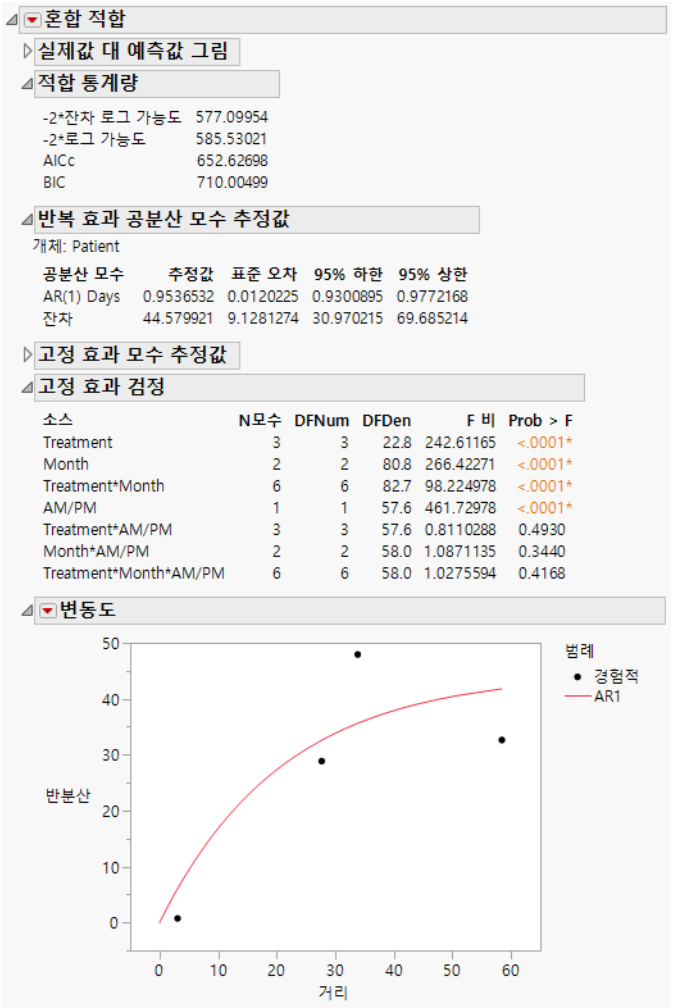
5. **실행**을 클릭합니다 .

   그림 8.20 에 " 혼합 적합 " 보고서가 표시되어 있습니다 ." 적합 통계량 " 보고서에 따르면 AR(1) 모형의 AICc 는 652.63 입니다 . 이 값을 잔차 모형의 832.55, 비정형 모형의 703.84 및 Toeplitz 이분산 모형의 788.03 과 비교합니다 . AICc 기준에 따르면 AR(1) 모형이 네 개 모형 중 가장 적합합니다 .

   AR(1) 구조를 사용하는 경우 두 개의 공분산 모수를 추정해야 합니다 . 이러한 추정값은 " 반복 효과 공분산 모수 추정값 " 보고서에 표시됩니다 . AR(1) Days 모수 추정값은 AR(1) 구조의 상관 모수인 ρ 의 추정값입니다 .

   변동도 그림에는 AR(1) 모형에 대한 경험적 반분산과 곡선이 표시됩니다 . Days 에는 0 이 아닌 값이 5 개뿐이므로 4 개의 거리 클래스만 가능하고 4 개의 점만 표시됩니다 . AR(1) 구조는 적절해 보입니다 . 다른 구조를 탐색하려면 " 변동도 " 의 빨간색 삼각형 메뉴에서 옵션을 선택합니다 ." 변동도 " 옵션에 대한 자세한 내용은 " 변동도 "(352 페이지) 에서 확인하십시오 .

그림 8.20 혼합 적합 보고서 - AR(1) 공분산 구조



JMP PRO AR(1) 구조를 사용한 추가 분석

AR(1) 모형이 최량 적합을 제공하므로 이 모형을 채택하여 분석을 계속합니다 . " 고정 효과 검정 " 보고서에서는 AM/PM 의 주효과뿐만 아니라 Treatment 와 Month 간에 유의한 교호작용이 있음을 나타냅니다 . 이제 이러한 유의 효과에 대해 살펴보겠습니다 .

1. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **주변 모형 추론 > 프로파일러**를 선택합니다 .

   " 주변 모형 프로파일러 " 보고서 (그림 8.21) 를 사용하면 Treatment, Month 및 AM/PM 의 다양한 설정이 콜레스테롤 수준 (Y) 에 미치는 영향을 확인할 수 있습니다 .

2. Month 그림에서 빨간색 수직 점선을 April에서 May로 드래그한 후 June으로 드래그합니다.

3 개월 동안 Y 에 대한 예측 AM 측정값이 277.4(April 평균) 에서 177.7(June 평균) 로 감소합니다 .

3. Treatment 그림에서 빨간색 수직 점선을 A 에서 B 로 드래그합니다 .

   Month 그림의 선을 April 에서 June 으로 드래그하면 Treatment B 의 경우 Y 에 대한 예측 AM 평균이 276.8(April) 에서 191.2(June) 로 감소하는 것을 볼 수 있습니다 .

4. Treatment 그림에서 수직 점선을 Control 로 드래그한 후 Placebo 로 드래그합니다 .

   Treatment 를 Control 또는 Placebo 로 설정하면 3 개월 동안 거의 변화가 없습니다 (그림 8.22). 다음으로 AM/PM 효과를 살펴봅니다 .

5. 빨간색 수직선을 드래그하여 Treatment 와 Month 를 12 개의 모든 수준 조합으로 설정합니다 .

   12 개의 모든 조합에 대해 예측 콜레스테롤 수준은 일관되게 아침보다 오후에 더 높아 주효과를 설명합니다 .

Treatment A 는 Treatment B 보다 5 월 (May) 에 콜레스테롤 수치가 더 낮은 것으로 보입니다 . 이 효과가 유의하면 Treatment A 가 B 보다 빨리 반응한다고 할 수 있습니다 . 다음 섹션 "6 월 (June) 의 모든 치료 (Treatment) 비교 "(367 페이지) 에서는 치료 (Treatment) 를 평가하는 방법을 보여 줍니다 .

그림 8.21 Treatment A 에 대한 주변 프로파일러 그림

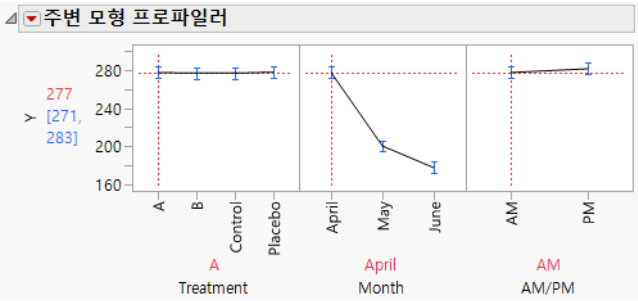
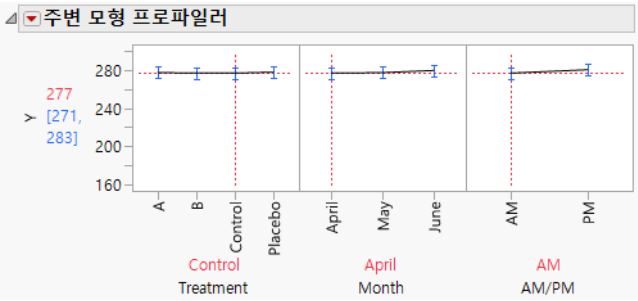


그림 8.22 Control 에 대한 주변 프로파일러 그림

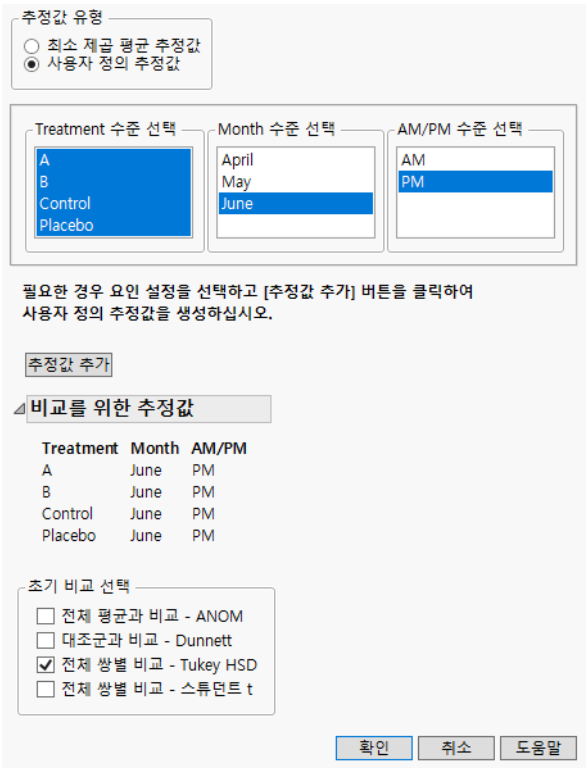


6 월 (June) 의 모든 치료 (Treatment) 비교

연구는 4 월 (April), 5 월 (May), 6 월 (June) 에 걸쳐 수행됩니다 . 6 월 (June) 의 PM 측정에서 어떤 치료법이 차이가 있는지 알아보려고 합니다 .

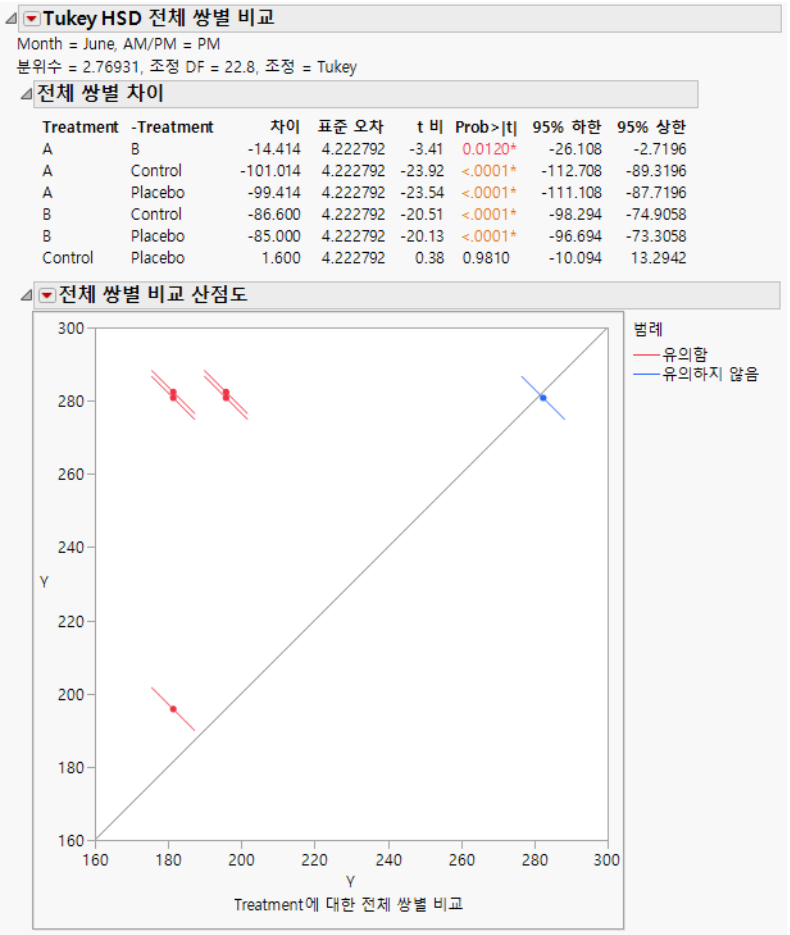
1. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **다중 비교**를 선택합니다 .
2. " 추정값 유형 " 에서 **사용자 정의 추정값**을 선택합니다 .
3. "Treatment 수준 선택 " 패널에서 네 가지 치료 유형을 모두 선택합니다 .
4. "Month 수준 선택 " 패널에서 "June" 을 선택합니다 .
5. "AM/PM 수준 선택 " 패널에서 "PM" 을 선택합니다 .
6. **추정값 추가**를 클릭합니다 .
7. " 초기 비교 선택 " 목록에서 **전체 쌍별 비교 - Tukey HSD** 를 선택합니다 .

그림 8.23 완료된 다중 비교 창



8. **확인**을 클릭합니다 .

그림 8.24 June PM 의 모든 Treatment 에 대한 Tukey HSD 전체 쌍별 비교 보고서



Tukey HSD 전체 쌍별 비교

Month = June, AM/PM = PM
분위수 = 2.76931, 조정 DF = 22.8, 조정 = Tukey

전체 쌍별 차이

Treatment	-Treatment	차이	표준 오차	t 비	Prob>|t|	95% 하한	95% 상한
A	B	-14.414	4.222792	-3.41	0.0120*	-26.108	-2.7196
A	Control	-101.014	4.222792	-23.92	<.0001*	-112.708	-89.3196
A	Placebo	-99.414	4.222792	-23.54	<.0001*	-111.108	-87.7196
B	Control	-86.600	4.222792	-20.51	<.0001*	-98.294	-74.9058
B	Placebo	-85.000	4.222792	-20.13	<.0001*	-96.694	-73.3058
Control	Placebo	1.600	4.222792	0.38	0.9810	-10.094	13.2942

전체 쌍별 비교 산점도

"Tukey HSD 전체 쌍별 비교" 보고서에는 "전체 쌍별 차이" 보고서와 "전체 쌍별 비교 산점도"가 표시됩니다. June PM 측정에 대해 Control과 Placebo를 제외한 모든 치료가 유의하게 다릅니다.

치료 A 와 B 간의 차이를 검토합니다. 평균 차이는 -14.414 이고 신뢰 구간 범위는 -26.108 ~ -2.7196 입니다. 치료 A 에 의한 콜레스테롤 측정값 감소가 치료 B 에 의한 감소보다 2.7 ~ 26.1 점 정도 높다는 결론을 내릴 수 있습니다. 치료 A 와 B 둘 다 Control 과 Placebo 에 비해 매우 효과적입니다.

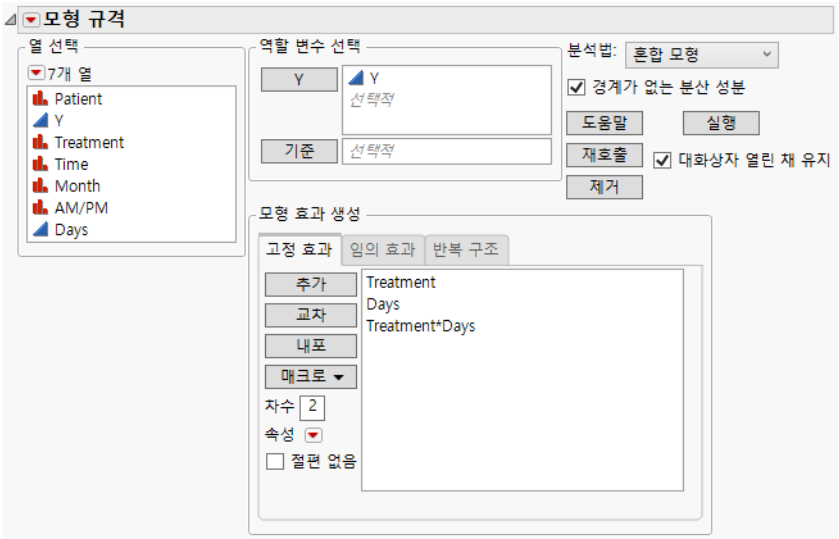
JMP PRO AR(1) 에 대한 회귀 모형의 예

Month 및 AM/PM 범주형 효과를 사용하여 콜레스테롤 데이터에 대한 네 가지 공분산 구조를 비교했습니다. 비정형 적합을 위해 범주형 효과가 필요했습니다. 결과적으로 AR(1) 공분산 구조를 사용하기로 결정했습니다.

이제 범주형 효과 대신 연속형 효과 Days 를 사용하여 치료 효과를 모델링한다고 가정해 보겠습니다 . 그런 다음 치료 기간 동안 임의의 시간에서 콜레스테롤 수준을 예측할 수 있습니다 .

1. " 공분산 구조 : AR(1)"(363 페이지) 에서 1 단계 ~ 4 단계를 수행한 후 모형 적합 시작 창으로 돌아갑니다 .
2. " 고정 효과 " 탭에서 기존 고정 효과를 선택하고 **제거**를 클릭합니다 .
3. Treatment 와 Days 를 선택한 후 **매크로 > 완전 요인**을 선택합니다 .

그림 8.25 고정 효과 탭을 보여 주는 모형 적합 시작 창



4. " 모형 규격 " 의 빨간색 삼각형을 클릭하고 **중심화 다항식**을 선택 취소합니다 .

참고 : 기본 설정에서는 교호작용 항에 사용되는 연속형 효과가 중심화됩니다 . " 중심화 다항식 " 옵션을 해제하면 교호작용 항의 연속형 효과가 중심화되지 않습니다 .

5. **실행**을 클릭합니다 .

그림 8.26 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . Treatment 와 Days 의 교호작용이 매우 유의하며 약물에 대해 다양한 회귀를 나타낸다는 것을 확인할 수 있습니다 .

참고 : 다른 날짜의 약물에 대한 결과를 예측하려면 프로파일러를 사용합니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

그림 8.26 혼합 적합 보고서 - 연속형 고정 효과를 사용한 AR(1) 공분산 구조

혼합 적합

적합 통계량

-2*잔차 로그 가능도	791.54856
-2*로그 가능도	789.76821
AICc	811.78655
BIC	837.64312

반복 효과 공분산 모수 추정값

개체: Patient

공분산 모수	추정값	표준 오차	95% 하한	95% 상한
AR(1) Days	0.7983061	0.0473103	0.7055796	0.8910327
잔차	100.5539	18.519406	72.239307	149.60244

고정 효과 모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|	95% 하한	95% 상한
절편	276.17698	1.9946144	52.7	138.46	<.0001*	272.17582	280.17815
Treatment[A]	-6.13599	3.4547735	52.7	-1.78	0.0815	-13.06621	0.7942322
Treatment[B]	-0.352827	3.4547735	52.7	-0.10	0.9190	-7.283049	6.5773954
Treatment[Control]	1.9453265	3.4547735	52.7	0.56	0.5758	-4.984896	8.8755487
Days	-0.735803	0.0505395	53.3	-14.56	<.0001*	-0.83716	-0.634446
Treatment[A]*Days	-0.877856	0.087537	53.3	-10.03	<.0001*	-1.053411	-0.7023
Treatment[B]*Days	-0.643666	0.087537	53.3	-7.35	<.0001*	-0.819221	-0.46811
Treatment[Control]*Days	0.788161	0.087537	53.3	9.00	<.0001*	0.6126056	0.9637164

고정 효과 검정

소스	N모수	DFNum	DFDen	F 비	Prob > F
Treatment	3	3	52.7	1.302892	0.2832
Days	1	1	53.3	211.9633	<.0001*
Treatment*Days	3	3	53.3	76.472494	<.0001*

JMP PRO 분할구 예

혼합 모형 분석법은 분할구 실험을 지정하고 분석하는 간단한 방법을 제공합니다. 혼합 모형은 효과를 지정할 수 있는 탭을 제공합니다. 결과 분석은 임의 효과를 대상으로 합니다. 그러나 분할구 실험은 표준 최소 제곱 분석법을 사용하여 분석할 수도 있습니다.

Split Plot.jmp 샘플 데이터 테이블에는 육류에 대한 연화제 및 조리 시간의 효과 연구에서 가져온 데이터가 포함되어 있습니다. 육류 포장 공장의 도체에서 6개의 소도체가 무작위로 선택되었습니다. 각 도체의 오른쪽 갈비살에서 균일한 조건으로 3개의 로스트용 고기말이가 준비되었습니다. 3개의 로스트에는 각각 무작위로 연화제 처리가 할당되었습니다. 처리 후 코어링 장치를 사용하여 각각 중앙 근처에 4개의 코어를 표시했습니다.

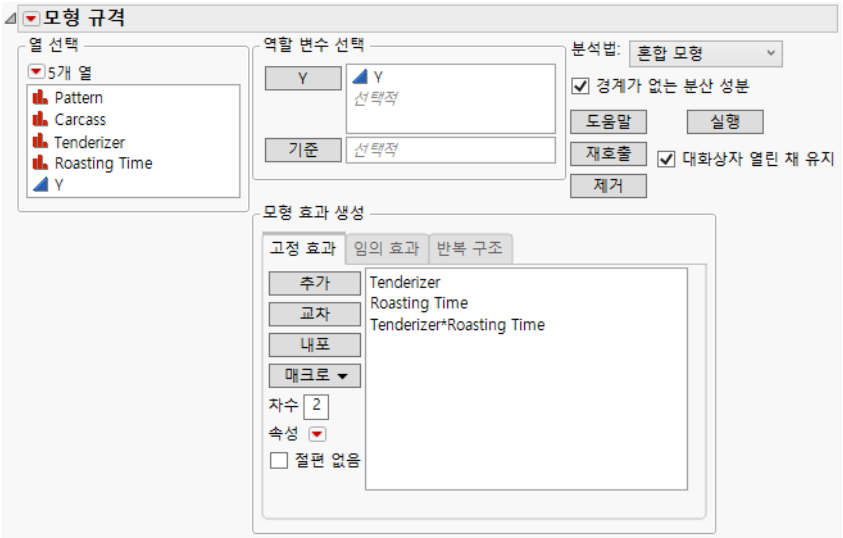
같은 도체에서 나온 3개의 로스트를 예열된 오븐에 함께 넣고 조리할 수 있도록 했습니다. 30분 후 각 로스트에서 코어 중 하나를 무작위로 없앴습니다. 36분, 42분, 48분 후 다시 같은 방식으로 코어를 제거했습니다. 각 세트가 서빙 온도로 냉각되면 Warner-Bratzler 장치를 사용하여 코어의 연질을 측정했습니다. 측정값이 클수록 고기가 질기다는 것을 나타냅니다.

이 연구에서는 연화제 및 로스팅 시간의 효과에 초점을 맞추고 특히 연화제와 로스팅 시간 사이의 교호작용이 있는지 여부에 관심이 있습니다 . 이 설계를 사용하여 목표를 해결합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Split Plot.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .

3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. " 분석법 " 목록에서 **혼합 모형**을 선택합니다 .
5. Tenderizer 와 Roasting Time 을 선택한 후 **매크로 > 완전 요인**을 선택합니다 .

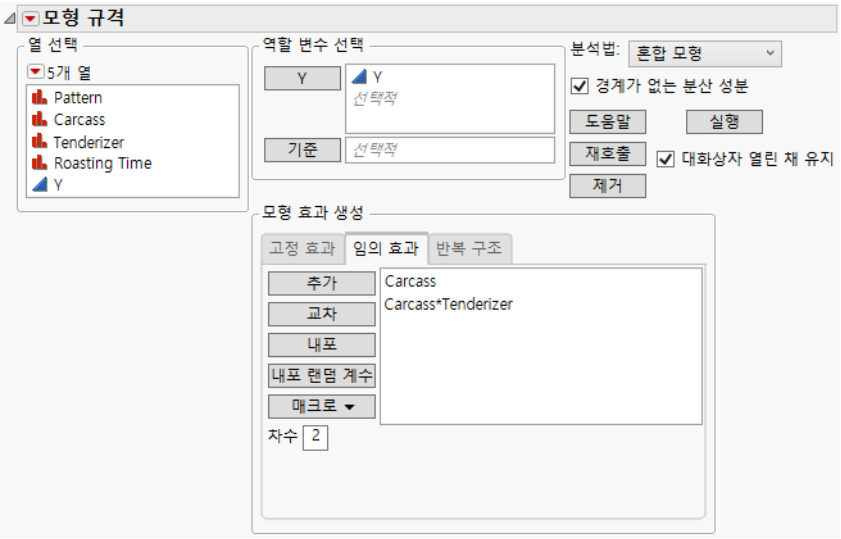
그림 8.27 완료된 고정 효과 탭을 보여 주는 모형 적합 시작 창



6. **임의 효과** 탭을 선택합니다 .
7. Carcass 를 선택하고 **추가**를 클릭하여 임의 carcass 효과를 생성합니다 .
8. Carcass 와 Tenderizer 를 선택하고 **교차**를 클릭합니다 .

   Carcass*Tenderizer 교호작용은 주구 요인 Tenderizer 의 오차 항입니다 . 이것은 표준 최소 제곱의 Carcass*Tenderizer& 랜덤 항과 동등합니다 .

그림 8.28 완료된 임의 효과 탭을 보여 주는 모형 적합 시작 창

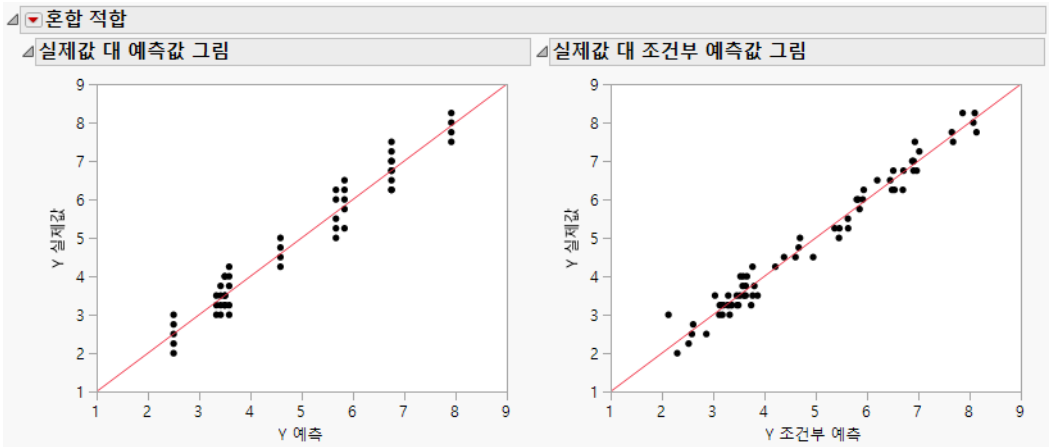


9. **실행**을 클릭합니다 .

   그림 8.29 에 " 혼합 적합 " 보고서가 표시되어 있습니다 .

   " 실제값 대 예측값 그림 " 및 " 실제값 대 조건부 예측값 그림 " 에서 모형 적합 문제가 보이지 않으므로 결과를 해석할 수 있습니다 . " 고정 효과 검정 " 보고서에서는 연화제와 로스팅 시간 사이에 유의한 교호작용이 있음을 나타냅니다 .

그림 8.29 혼합 적합 보고서



혼합 적합

실제값 대 예측값 그림

실제값 대 조건부 예측값 그림

적합 통계량

-2*잔차 로그 가능도	68.324206
-2*로그 가능도	23.659713
AICc	62.231142
BIC	87.809705

임의 효과 공분산 모수 추정값

분산 성분	추정값	표준 오차	95% 하한	95% 상한	Wald p 값
Carcass	0.0605042	0.0411897	-0.020226	0.1412346	0.1419
Carcass*Tenderizer	-0.007958	0.0076058	-0.022865	0.0069492	0.2954
잔차	0.0863433	0.0182028	0.0594013	0.1369749	
합계	0.1468476	0.0450325	0.0871689	0.2983101	

고정 효과 모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|	95% 하한	95% 상한
절편	4.7777778	0.1041208	5.0	45.89	<.0001*	4.5101253	5.0454302
Tenderizer[Control]	0.6388889	0.0389128	10.0	16.42	<.0001*	0.5521867	0.7255911
Tenderizer[Papain]	-0.673611	0.0389128	10.0	-17.31	<.0001*	-0.760313	-0.586909
Roasting Time[30]	2.0555556	0.0599803	45.0	34.27	<.0001*	1.934749	2.1763621
Roasting Time[36]	0.8888889	0.0599803	45.0	14.82	<.0001*	0.7680823	1.0096955
Roasting Time[42]	-1.25	0.0599803	45.0	-20.84	<.0001*	-1.370807	-1.129193
Tenderizer[Control]*Roasting Time[30]	0.4444444	0.084825	45.0	5.24	<.0001*	0.2735982	0.6152907
Tenderizer[Control]*Roasting Time[36]	0.4444444	0.084825	45.0	5.24	<.0001*	0.2735982	0.6152907
Tenderizer[Control]*Roasting Time[42]	-0.583333	0.084825	45.0	-6.88	<.0001*	-0.75418	-0.412487
Tenderizer[Papain]*Roasting Time[30]	-0.326389	0.084825	45.0	-3.85	0.0004*	-0.497235	-0.155543
Tenderizer[Papain]*Roasting Time[36]	-0.409722	0.084825	45.0	-4.83	<.0001*	-0.580569	-0.238876
Tenderizer[Papain]*Roasting Time[42]	0.6458333	0.084825	45.0	7.61	<.0001*	0.474987	0.8166796

랜덤 계수

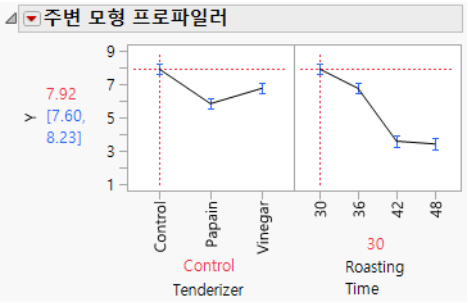
고정 효과 검정

소스	N모수	DFNum	DFDen	F 비	Prob > F
Tenderizer	2	2	10.0	190.00818	<.0001*
Roasting Time	3	3	45.0	656.61634	<.0001*
Tenderizer*Roasting Time	6	6	45.0	18.458286	<.0001*

연화제와 로스팅 시간 사이의 교호작용 탐색

1. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **주변 모형 추론 > 프로파일러**를 선택합니다 .

그림 8.30 Roasting Time 이 30 분으로 설정된 주변 모형 프로파일러



2. "Roasting Time" 패널에서 빨간색 수직 파선을 36, 42, 48 로 움직입니다 .

   그림 8.30 에서 로스팅 시간이 30 분 또는 36 분일 때 파파인 (Papain) 과 식초 (Vinegar) 연화제 모두 대조군 (Control) 보다 연질 점수가 유의하게 낮습니다 . 그러나 42 분에서는 유의한 차이가 없습니다 . 48 분일 때는 파파인 (Papain) 값이 대조군 (Control) 보다 낮지만 식초 (Vinegar) 값은 그렇지 않습니다 . 42 분을 제외한 모든 시간에서 파파인 (Papain) 의 연질 점수가 식초 (Vinegar) 보다 낮습니다 .

3. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **다중 비교**를 선택합니다 .

4. **Tenderizer*Roasting Time** 을 선택합니다 .

5. **전체 쌍별 비교 - Tukey HSD** 를 선택한 후 **확인**을 클릭합니다 .

   그림 8.31 에서는 쌍별 비교 목록의 일부를 보여 줍니다 . 프로파일러에서 관측한 파파인 (Papain)과 식초(Vinegar)의 차이는 대부분 통계적으로 유의합니다. 따라서 파파인(Papain) 이 더 나은 연화제입니다 .

그림 8.31 다중 비교 (일부분)

혼합 적합

Tenderizer*Roasting Time에 대한 다중 비교

최소 제곱 평균 추정값

Tenderizer	Roasting Time	추정값	표준 오차	DF	95% 하한	95% 상한
Control	30	7.9166667	0.15214556	20.355	7.5996515	8.2336819
Control	36	6.7500000	0.15214556	20.355	6.4329848	7.0670152
Control	42	3.5833333	0.15214556	20.355	3.2663181	3.9003485
Control	48	3.4166667	0.15214556	20.355	3.0996515	3.7336819
Papain	30	5.8333333	0.15214556	20.355	5.5163181	6.1503485
Papain	36	4.5833333	0.15214556	20.355	4.2663181	4.9003485
Papain	42	3.5000000	0.15214556	20.355	3.1829848	3.8170152
Papain	48	2.5000000	0.15214556	20.355	2.1829848	2.8170152
Vinegar	30	6.7500000	0.15214556	20.355	6.4329848	7.0670152
Vinegar	36	5.6666667	0.15214556	20.355	5.3496515	5.9836819
Vinegar	42	3.5000000	0.15214556	20.355	3.1829848	3.8170152
Vinegar	48	3.3333333	0.15214556	20.355	3.0163181	3.6503485

Tukey HSD 전체 쌍별 비교

분위수 = 3.44501, 조정 DF = 45.0, 조정 = Tukey-Kramer

전체 쌍별 차이

Tenderizer	Roasting Time	-Tenderizer	-Roasting Time	차이	표준 오차	t 비	Prob>|t|	95% 하한	95% 상한
Control	30	Control	36	1.16667	0.1696500	6.88	<.0001*	0.58222	1.75111
Control	30	Control	42	4.33333	0.1696500	25.54	<.0001*	3.74889	4.91778
Control	30	Control	48	4.50000	0.1696500	26.53	<.0001*	3.91555	5.08445
Control	30	Papain	30	2.08333	0.1616430	12.89	<.0001*	1.52647	2.64019
Control	30	Papain	36	3.33333	0.1616430	20.62	<.0001*	2.77647	3.89019
Control	30	Papain	42	4.41667	0.1616430	27.32	<.0001*	3.85981	4.97353
Control	30	Papain	48	5.41667	0.1616430	33.51	<.0001*	4.85981	5.97353
Control	30	Vinegar	30	1.16667	0.1616430	7.22	<.0001*	0.60981	1.72353
Control	30	Vinegar	36	2.25000	0.1616430	13.92	<.0001*	1.69314	2.80686
Control	30	Vinegar	42	4.41667	0.1616430	27.32	<.0001*	3.85981	4.97353
Control	30	Vinegar	48	4.58333	0.1616430	28.35	<.0001*	4.02647	5.14019
Control	36	Control	42	3.16667	0.1696500	18.67	<.0001*	2.58222	3.75111
Control	36	Control	48	3.33333	0.1696500	19.65	<.0001*	2.74889	3.91778
Control	36	Papain	30	0.91667	0.1616430	5.67	<.0001*	0.35981	1.47353
Control	36	Papain	36	2.16667	0.1616430	13.40	<.0001*	1.60981	2.72353
Control	36	Papain	42	3.25000	0.1616430	20.11	<.0001*	2.69314	3.80686
Control	36	Papain	48	4.25000	0.1616430	26.29	<.0001*	3.69314	4.80686
Control	36	Vinegar	30	8.882e-16	0.1616430	0.00	1.0000	-0.55686	0.55686
Control	36	Vinegar	36	1.08333	0.1616430	6.70	<.0001*	0.52647	1.64019
Control	36	Vinegar	42	3.25000	0.1616430	20.11	<.0001*	2.69314	3.80686
Control	36	Vinegar	48	3.41667	0.1616430	21.14	<.0001*	2.85981	3.97353
Control	42	Control	48	0.16667	0.1696500	0.98	0.9974	-0.41778	0.75111
Control	42	Papain	30	-2.25000	0.1616430	-13.92	<.0001*	-2.80686	-1.69314
Control	42	Papain	36	-1.00000	0.1616430	-6.19	<.0001*	-1.55686	-0.44314
Control	42	Papain	42	0.08333	0.1616430	0.52	1.0000	-0.47353	0.64019
Control	42	Papain	48	1.08333	0.1616430	6.70	<.0001*	0.52647	1.64019
Control	42	Vinegar	30	-3.16667	0.1616430	-19.59	<.0001*	-3.72353	-2.60981
Control	42	Vinegar	36	-2.08333	0.1616430	-12.89	<.0001*	-2.64019	-1.52647
Control	42	Vinegar	42	0.08333	0.1616430	0.52	1.0000	-0.47353	0.64019
Control	42	Vinegar	48	0.25000	0.1616430	1.55	0.9187	-0.30686	0.80686
Control	48	Papain	30	-2.41667	0.1616430	-14.95	<.0001*	-2.97353	-1.85981
Control	48	Papain	36	-1.16667	0.1616430	-7.22	<.0001*	-1.72353	-0.60981
Control	48	Papain	42	-0.08333	0.1616430	-0.52	1.0000	-0.64019	0.47353
Control	48	Papain	48	0.91667	0.1616430	5.67	<.0001*	0.35981	1.47353
Control	48	Vinegar	30	-3.33333	0.1616430	-20.62	<.0001*	-3.89019	-2.77647
Control	48	Vinegar	36	-2.25000	0.1616430	-13.92	<.0001*	-2.80686	-1.69314
Control	48	Vinegar	42	-0.08333	0.1616430	-0.52	1.0000	-0.64019	0.47353
Control	48	Vinegar	48	0.08333	0.1616430	0.52	1.0000	-0.47353	0.64019

JMP PRO 공간 예 : 균일도 시험

Uniformity Trial.jmp 샘플 데이터 테이블을 예로 들겠습니다. 8 x 8 격자의 대지에 대해 작물 균일도 시험이 수행되었습니다. 균일도 시험에서는 실험 처리가 적용되지 않은 구획에서 테스트 작물을 재배합니다. 일반적으로 측정되는 반응 변수는 수확량입니다. 구획에서 수행할 설계된 실험을 계획하기 위한 배경으로 해당 구획의 변동을 특성화하는 것이 이 연구의 목적입니다. 자세한 내용은 Littell et al. 연구 자료 (2006, p. 447) 에서 확인하십시오.

목적은 이러한 데이터의 정보를 사용하여 16 개의 처리로 수확량 시험을 설계하는 것입니다 . 특히 해당 구획에서 다음과 같은 향후 실험을 수행할지 여부를 결정하려고 합니다 .

- 4 개의 블록이 있는 완비 블록 설계 (데이터에 Quarter 로 표시됨)
- 16 개의 블록이 있는 불완비 블록 설계 (데이터에 Subquarter 로 표시됨)
- 공간 상관 오차가 있는 완전히 랜덤화된 설계

이 목적을 위해 공간 데이터를 둘 이상의 차원이 반복 효과로 포함된 반복 측정으로 처리할 수 있습니다. 따라서 " 적합 통계량 " 보고서의 값을 사용하여 적절한 모형을 비교하고 선택할 수 있습니다. 유의한 공간 변동이 있는지 확인하는 것으로 시작한 후 너겟 효과가 있는지 여부를 확인합니다.

너겟 효과가 있는지 여부를 확인한 후에는 가장 적합한 공간 공분산 구조를 결정합니다. 마지막으로 블록 모형을 적합시키고 이러한 모형을 최적 공간 구조와 비교합니다. 이 예에서는 AICc 와 BIC 를 모두 사용하여 최적 모형을 선택합니다. "공간 상관 구조"(399 페이지) 에서는 너겟 효과 및 기타 공간 용어에 대한 자세한 내용을 제공합니다.

팁 : 이 섹션에서는 모형을 적합하는 것부터 최적 공분산 구조를 결정하는 것까지 공간 데이터 적합에 대한 다양한 내용을 설명합니다. 각 예를 진행하는 동안 모형 적합 시작 창을 열어 두십시오.

JMP PRO 공간 구조 모형 적합

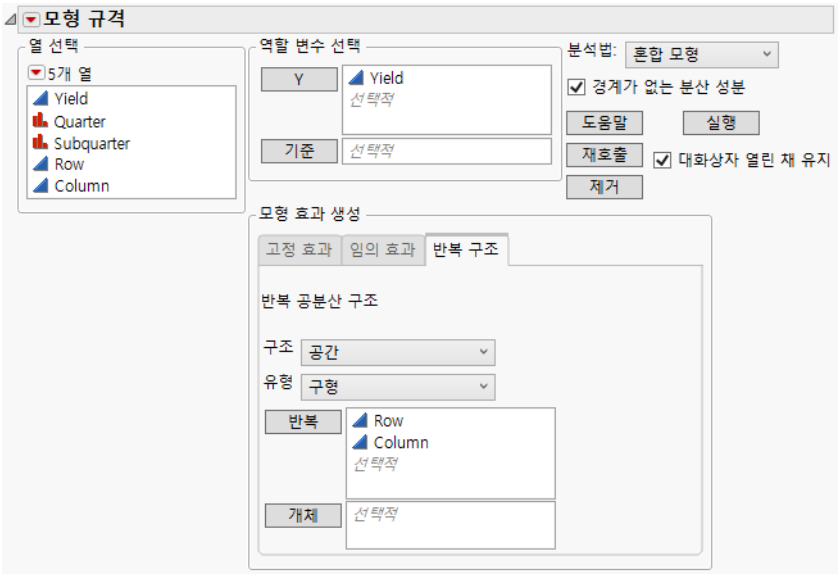
유의한 공간 변동이 있는지 확인하기 위해 공간 변동을 설명하는 모형을 적합시킬 수 있습니다. 그런 다음 이 공간 모형의 가능도를 공간 변동이 고려되지 않은 모형의 가능도와 비교할 수 있습니다. 독립 오차 모형이 공간 모형 계열에 내포되어 있으므로 이를 수행할 수 있습니다. 독립 오차 모형은 공간 상관 (ρ) 이 0 인 공간 모형입니다. 이는 두 모형에 대한 공식 가능도비 검정을 수행할 수 있다는 의미입니다.

먼저 공간 구조를 설명하는 모형을 적합시킵니다 .

1. **도움말 > 샘플 데이터 라이브러리** 를 선택하고 Uniformity Trial.jmp 를 엽니다 .
2. **분석 > 모형 적합** 을 선택합니다 .
3. 다음 예에서 시작 창으로 돌아갈 수 있도록 **대화상자 열린 채 유지** 를 선택합니다 .
4. Yield 를 선택하고 **Y** 를 클릭합니다 .
5. " 분석법 " 목록에서 **혼합 모형** 을 선택합니다 .

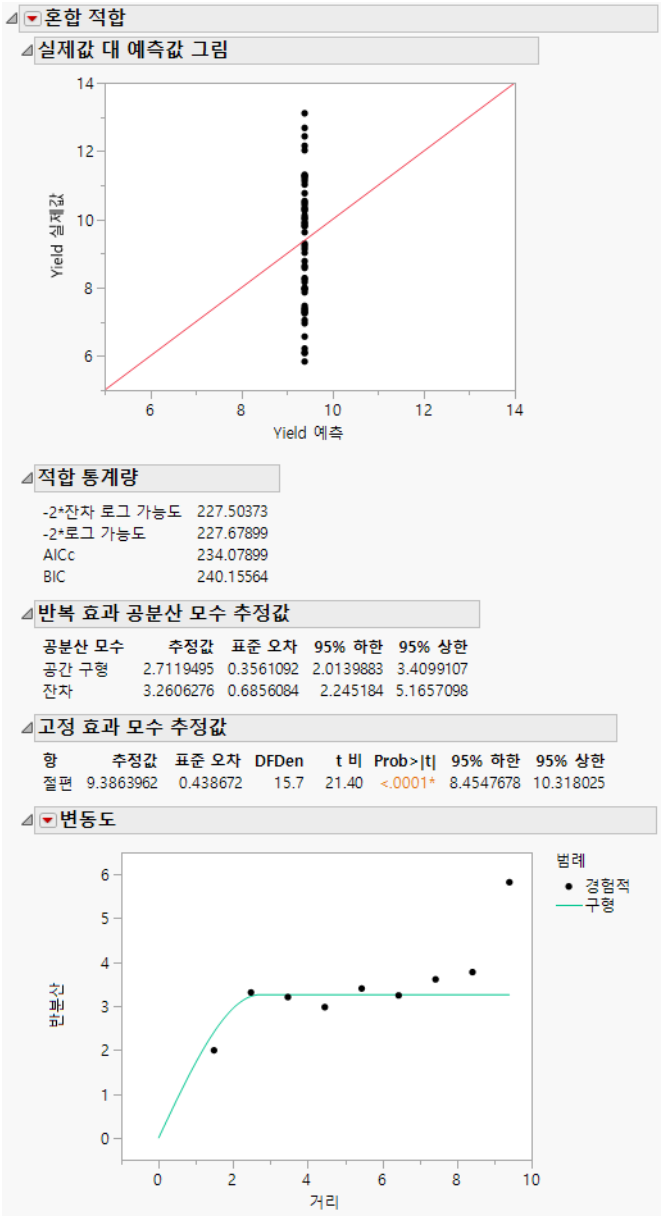
6. **반복 구조** 탭을 선택합니다 .
7. " 구조 " 옆의 목록에서 **공간**을 선택합니다 .
8. " 유형 " 옆의 목록에서 **구형**을 선택합니다 .
9. Row 와 Column 을 선택하고 **반복**을 클릭합니다 .

그림 8.32 반복 구조 탭을 보여 주는 완료된 모형 적합 시작 창



10. **실행**을 클릭합니다 .

그림 8.33 혼합 적합 보고서 - 공간 구형



"혼합 적합" 보고서의 "실제값 대 예측값 그림"에서는 예측 수확량(Yield)이 단일 값임을 나타냅니다. 이는 공간 공분산만 적합되었기 때문입니다. "적합 통계량" 보고서에 따르면 -2*로그 가능도는 227.68이고 AICc는 234.08입니다.

등방성 공간 구조가 적합되었으므로 변동도 그림이 표시됩니다. 이 시험은 8 x 8 격자로 설계되므로 매우 먼 거리보다 짧은 거리에 더 많은 점 쌍이 있습니다. 레이아웃은 그림 8.37에서 확인하

십시오. "변동도"에서는 최대 약 8.4 거리에 대해 구형 공간 구조가 가장 적합하다는 것을 보여 줍니다. 최종 거리에 대한 거리 클래스는 두 개의 주대각 점 쌍으로만 구성됩니다.

"반복 효과 공분산 모수 추정값" 보고서에는 범위(공간 구형 = 2.71) 및 문턱(sill)(잔차 = 3.26)의 추정값이 제공됩니다. 자세한 내용은 "변동도"(399 페이지)에서 확인하십시오.

JMP PRO 독립 오차 모형 적합

다음으로 독립 오차 모형을 적합시킵니다.

1. "모형 적합" 시작 창으로 돌아갑니다.
2. **반복 구조** 탭을 선택합니다.
3. "구조" 목록에서 **잔차**를 선택합니다.
4. "반복" 효과 목록에서 Row 와 Column 을 제거합니다.

   그렇지 않으면 "잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다" 라는 경고가 표시됩니다.
5. **실행**을 클릭합니다.

   독립 오차 모형에 대한 적합 통계량은 -2* 로그 가능도 = 254.22 이고 AICc = 258.41 입니다. 이 값은 각각 공간 상관 모형의 해당 값 (-2* 로그 가능도 = 227.68, AICc = 234.08) 을 초과합니다 . 이러한 통계량의 값이 작을수록 더 적합하다는 것을 나타내므로 공간 모형의 적합도가 더 좋다고 할 수 있습니다.

JMP PRO 가능도비 검정 수행 (선택 사항)

공식 가능도비 검정에서는 공간 상관 모형이 유의한 변동을 설명하는지 여부를 보여 줍니다. 유효한 가능도비 검정을 생성하려면 한 모형이 다른 모형에 내포되어야 합니다.

일반적으로 공간 모형은 공식 가능도비 검정이 아니라 AICc 또는 BIC 를 사용하여 비교됩니다. AICc 또는 BIC 를 평가하는 것이 더 빠르고, 많은 공간 모형이 내포되지 않습니다.

독립 오차 모형이 공간 모형 계열에 내포되어 있으므로 이 예에서 가능도비 검정을 수행할 수 있습니다. 독립 오차 모형은 공간 상관(ρ)이 0 인 공간 모형입니다. 이는 두 모형에 대한 공식 가능도비 검정을 수행할 수 있다는 의미입니다.

이 예에서 가능도비 검정 통계량은 254.22 - 227.68 = 26.54 입니다. 이 값을 자유도 1 에 대한 카이 제곱 분포와 비교하면 공간 상관이 없다는 귀무가설이 p 값 < 0.0001 로 기각됩니다. 즉, 이러한 데이터에 유의한 공간 변동이 포함되어 있다는 결론을 내릴 수 있습니다.

JMP PRO 공간 공분산 유형 선택

다음으로 데이터에 가장 적합한 공간 공분산 구조를 결정합니다.

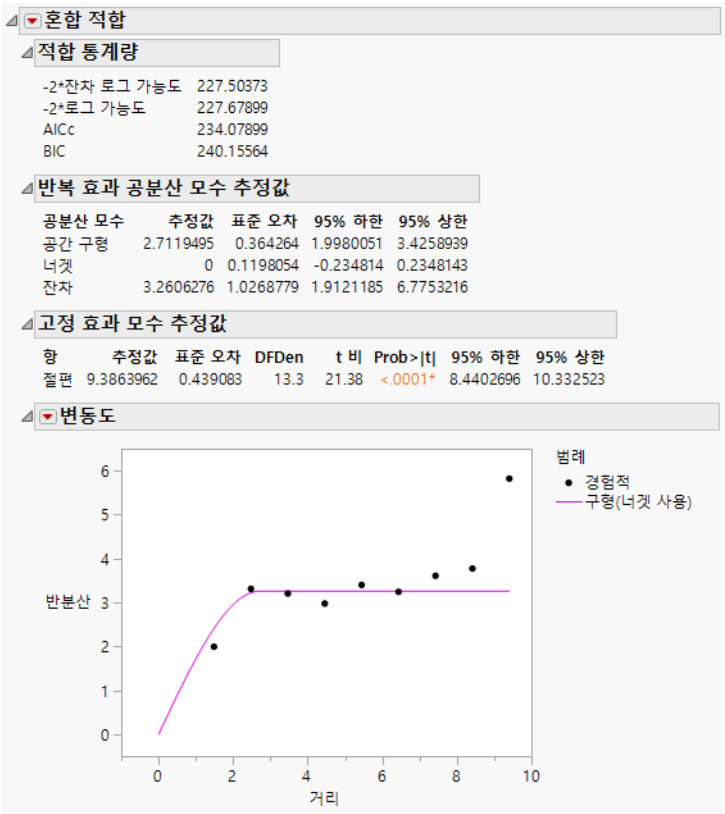
- 너겟 효과 (상대적으로 짧은 거리에 대한 변동) 사용 또는 사용 안 함
- 등방성 (공간 상관이 모든 방향으로 같음) 또는 이방성 (공간 상관이 두 방향으로 다름)

- 구조 유형 (구형 , 가우스 , 지수 또는 거듭제곱)

1. " 모형 적합 " 시작 창으로 돌아갑니다 .
2. **반복 구조** 탭을 선택합니다 .
3. Row 와 Column 을 선택하고 **반복**을 클릭합니다 .
4. " 구조 " 목록에서 **공간 (너겟 사용)** 을 선택합니다 .
5. " 유형 " 목록에서 **구형**을 선택합니다 .
6. **실행**을 클릭합니다 .

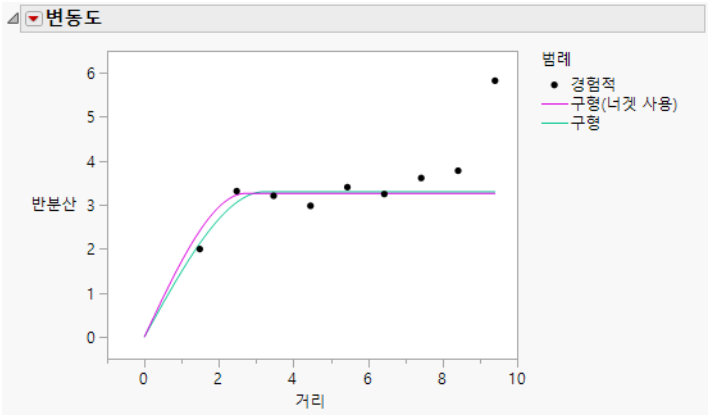
   그림 8.34 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . 로그 가능도는 기본적으로 너겟이 없는 구형 모형과 같고 AICc 는 약간 높습니다 (236.36 대 234.08). " 반복 효과 공분산 모수 추정값 " 보고서에서는 너겟 공분산 모수의 추정값이 0 임을 보여 줍니다 . 즉 , 너겟 효과에 대한 증거가 없습니다 .

그림 8.34 혼합 적합 보고서 - 공간 구형 (너겟 사용)



7. " 변동도 " 의 빨간색 삼각형을 클릭하고 **공간 > 구형**을 선택합니다 .

그림 8.35 혼합 적합 보고서 - 변동도



사실상 두 변동도는 동일합니다 . 이는 너겟 효과의 증거가 없다는 의미이기도 합니다 .

8. " 모형 적합 " 시작 창으로 돌아갑니다 .
9. **반복 구조** 탭을 선택합니다 .
10. 이방성을 검정하기 위해 " 구조 " 목록에서 **공간 이방성**을 선택합니다 .
11. " 유형 " 목록에서 **구형**을 선택합니다 .
12. **실행**을 클릭합니다 .

그림 8.36 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . 적합 통계량은 등방성 (공간 구조) 구형 모형만큼 적합하지 않음을 나타냅니다 (AICc 240.54 대 234.08). " 반복 효과 공분산 모수 추정값 " 보고서에서는 Row(공간 구형 Row) 및 Column(공간 구형 Column) 공분산의 추정값이 매우 근사하다는 것을 보여 줍니다 . 즉 , 격자의 행과 열 내에서 공간 상관이 다름을 나타내는 증거가 없습니다 .

그림 8.36 혼합 적합 보고서 - 공간 이방성 구형

혼합 적합

적합 통계량

-2*잔차 로그 가능도	232.09009
-2*로그 가능도	231.86638
AICc	240.54434
BIC	248.50191

반복 효과 공분산 모수 추정값

공분산 모수	추정값	표준 오차	95% 하한	95% 상한
공간 구형 Row	2.2453516	0.4114883	1.4388492	3.0518539
공간 구형 Column	2.2285839	0.3316821	1.5784989	2.8786689
잔차	3.0050408	0.6112702	2.0921756	4.6820483

고정 효과 모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob>|t|	95% 하한	95% 상한
절편	9.4039776	0.3585807	21.3	26.23	<.0001*	8.658977	10.148978

JMP PRO 공간 구조 유형 결정

너겟이 없는 등방성 공간 구조가 적절합니다. 사용 가능한 여러 유형을 비교하여 공간 구조 유형을 결정할 수 있습니다.

1. "모형 적합" 시작 창으로 돌아갑니다.
2. **반복 구조** 탭을 선택합니다.
3. Row 와 Column 을 선택하고 **반복**을 클릭합니다.
4. "구조" 목록에서 **공간**을 선택합니다.
5. "유형" 목록에서 **거듭제곱**을 선택합니다.
6. **실행**을 클릭합니다.

   모형 적합에 대한 AICc 및 BIC 값에 유의하십시오.
7. "모형 적합" 시작 창으로 돌아갑니다.
8. "유형" 목록에서 **지수**를 선택합니다.
9. **실행**을 클릭합니다.

   모형 적합에 대한 AICc 및 BIC 값에 유의하십시오.
10. "모형 적합" 시작 창으로 돌아갑니다.
11. "유형" 목록에서 **가우스**를 선택합니다.
12. **실행**을 클릭합니다.

    모형 적합에 대한 AICc 및 BIC 값에 유의하십시오.

표 8.2에는 이러한 유형 및 수행한 다른 적합에 대해 관측된 AICc 값이 요약되어 있습니다.

표 8.2 공간 모형 적합에 대한 적합 통계량

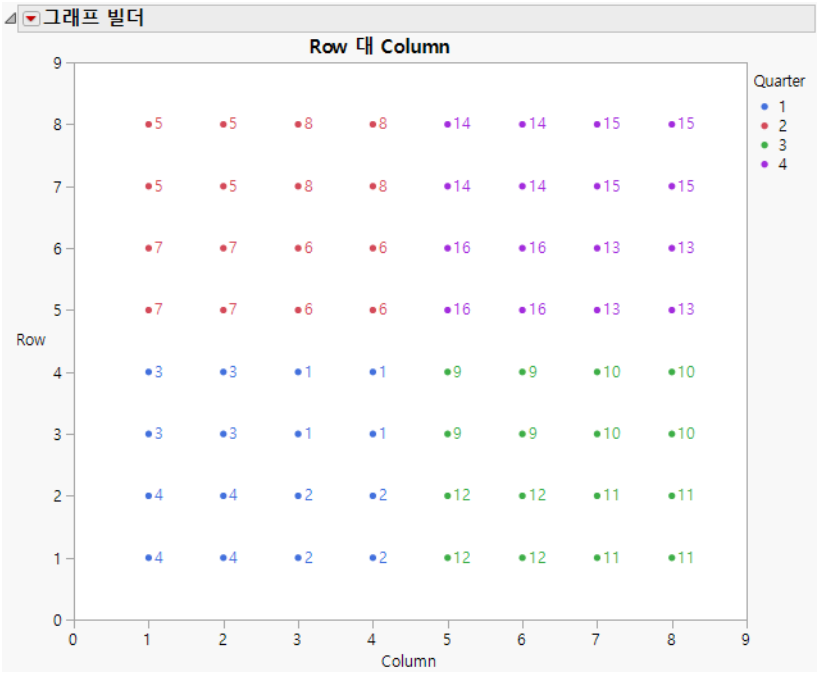
구조	유형	AICc	BIC
공간	구형	234.08	240.16
잔차		258.41	262.53
공간(너겟 사용)	구형	236.36	244.31
공간 이방성	구형	240.54	248.50
공간	거듭제곱	240.24	246.32
공간	지수	240.24	246.32
공간	가우스	238.37	244.44

AICc 값이 가장 작은 최량 적합 모형은 구형 공분산 구조를 사용한 공간 구조 모형입니다. 이제 이 모형을 완비 블록 모형 및 불완비 블록 모형과 비교하여 균일도 시험의 목표를 완료합니다.

JMP PRO 모형을 블록 설계와 비교

1. Uniformity Trial 데이터 테이블로 돌아갑니다 .
2. 테이블 패널에서 "Graph Builder" 스크립트를 실행합니다 .

그림 8.37 제안된 완비 / 불완비 블록 설계에 대한 그래프 빌더



이 그림에서는 해당 구획에 대해 제안된 완비 블록 설계 및 불완비 블록 설계를 보여 줍니다 . 색상은 완비 블록으로 사용되는 4 등분 구획을 나타냅니다 . 번호가 매겨진 점은 불완비 블록으로 사용되는 16 등분 구획을 나타냅니다 .

완비 블록 모형을 적합시키려면 다음 단계를 수행하십시오 .

1. " 모형 적합 " 시작 창으로 돌아갑니다 .
2. **반복 구조** 탭을 선택합니다 .
3. " 구조 " 목록에서 **잔차**를 선택합니다 .
4. 효과에서 Row 와 Column 을 제거합니다 .

   그렇지 않으면 " 잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다 " 라는 팝업 창이 표시됩니다 .
5. **임의 효과** 탭을 선택합니다 .

6. Quarter 를 선택하고 **추가**를 클릭합니다 .
7. **실행**을 클릭합니다 .

불완비 블록 모형을 적합시키려면 다음 단계를 수행하십시오.

1. " 모형 적합 " 시작 창으로 돌아갑니다 .
2. **임의 효과** 탭을 선택합니다 .
3. Quarter 를 선택하고 **제거**를 클릭합니다 .
4. Subquarter 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .

다음 목록에서는 경쟁 모형에 대한 AICc 및 BIC 값을 모두 보여 줍니다 . 구형 공분산 구조를 사용할 때 최량 모형 적합 결과가 생성됩니다. 이는 향후에 이 구획을 사용하여 연구를 수행하는 경우 공간 상관 오차가 있는 완전히 랜덤화된 설계가 선호된다는 것을 나타냅니다 .

- 구형 모형 (" 공간 구조 유형 결정 "(382 페이지) 참조)
  - AICc: 234.08
  - BIC: 240.16
- 완비 블록 (RCBD) 모형
  - AICc: 259.90
  - BIC: 265.97
- 불완비 블록 모형
  - AICc: 248.77
  - BIC: 254.85

JMP PRO 상관관계가 있는 반응의 예

이 예에서는 관심 있는 특성에 대해 웨이퍼 생산을 처리하는 두 레이아웃의 효과를 연구합니다. 50개의 각 웨이퍼는 4개의 사분면으로 분할되고 각 사분면에 대해 특성이 측정됩니다. 일반적으로 이 유형의 데이터는 각 행에 관심 단위 중 하나에 대한 모든 반복 측정값이 포함되는 형식으로 제공됩니다. 이 유형의 데이터는 각 반응에 대해 별도의 모형을 사용하여 분석되는 경우가 많습니다. 그러나 단일 단위에 대해 반복 측정을 수행하는 경우 단위 내 상관관계가 있을 가능성이 큽니다. 이 상관관계를 설명하지 못하면 잘못된 결정 및 예측이 생성될 수 있습니다. 혼합 모형 분석법을 사용하여 가능한 상관관계를 설명하고 모델링할 수 있습니다.

반복 측정 데이터에 대한 혼합 모형 분석의 경우 각 반복 측정값이 고유의 행에 포함되어야 합니다. 모든 반복 측정값이 동일한 행에 포함되는 일반적인 형식의 데이터가 제공되면 테이블 > 쌓기를 사용하여 혼합 모형 분석에 적합한 데이터 테이블을 생성할 수 있습니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

이 예에서는 먼저 일반적인 데이터 테이블 형식을 사용하여 단변량 모형을 적합시킵니다 . 그런 다음 혼합 모형 분석법을 사용하여 네 개의 반응에 대한 모형을 적합시키는 동시에 반응 간의 가능한 상관관계를 설명합니다 .

JMP PRO 단변량 모형 적합

표준 최소 제곱을 사용하여 네 개의 각 사분면에 대해 단변량 모형을 적합시킵니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Wafer Quadrants.jmp 를 엽니다 .

   이 데이터 테이블은 혼합 모형 분석에 사용할 수 있도록 각 Quadrant 에서 측정한 각 Y 값이 하나의 행에 포함된 구조입니다 . 단변량 분석을 수행하기 위해 저장된 스크립트를 사용하여 테이블을 분할합니다 .

2. Wafer Quadrants.jmp 데이터 테이블에서 **Split Y by Quadrant** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

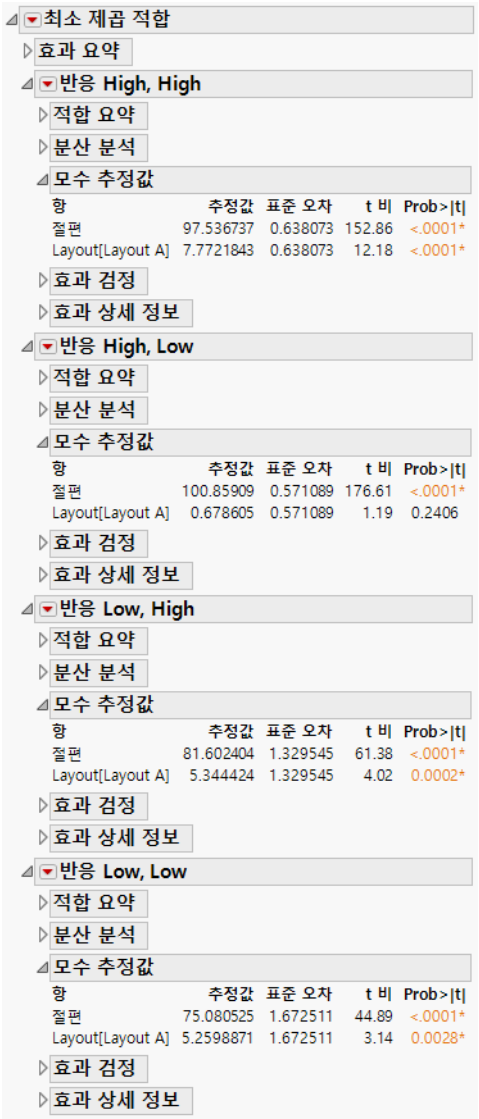
   새 데이터 테이블은 반복 측정 데이터를 기록하는 데 일반적으로 사용되는 형식으로 되어 있습니다 . Wafer ID 의 각 값은 행을 정의하고 해당 웨이퍼에 대한 네 개의 측정값이 단일 행에 제공됩니다 .

3. **분석 > 모형 적합**을 선택합니다 .

   데이터 테이블에 "Model" 스크립트가 있으므로 " 모형 규격 " 창이 이미 채워져 있습니다 . High, High ~ Low, Low 열은 Y 로 입력되고 Layout 은 단일 모형 효과입니다 .

4. **실행**을 클릭합니다 .

그림 8.38 네 가지 단변량 모형



보고서에 따르면 Layout 은 High, Low 사분면을 제외한 모든 사분면에 대해 통계적으로 유의한 효과입니다 .

JMP PRO 혼합 모형 분석 수행

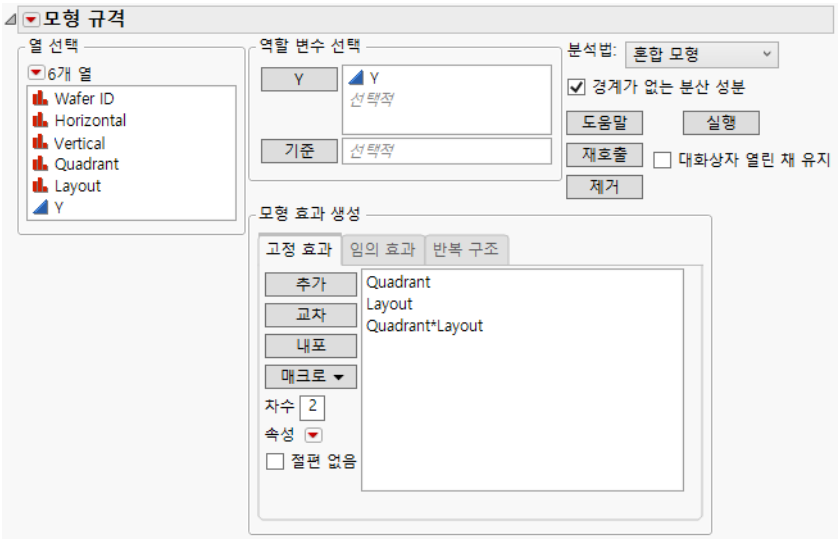
혼합 모형 분석을 사용하여 더 많은 정보를 얻을 수 있습니다 .

1. Wafer Quadrants.jmp 로 돌아갑니다 . 데이터 테이블을 닫았으면 **도움말 > 샘플 데이터 라이브러리**를 선택하고 Wafer Quadrants.jmp 를 엽니다 .

2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. **분석법** 목록에서 **혼합 모형**을 선택합니다 .
5. **열 선택** 목록에서 Quadrant 와 Layout 을 선택하고 **매크로 > 완전 요인**을 클릭합니다 .

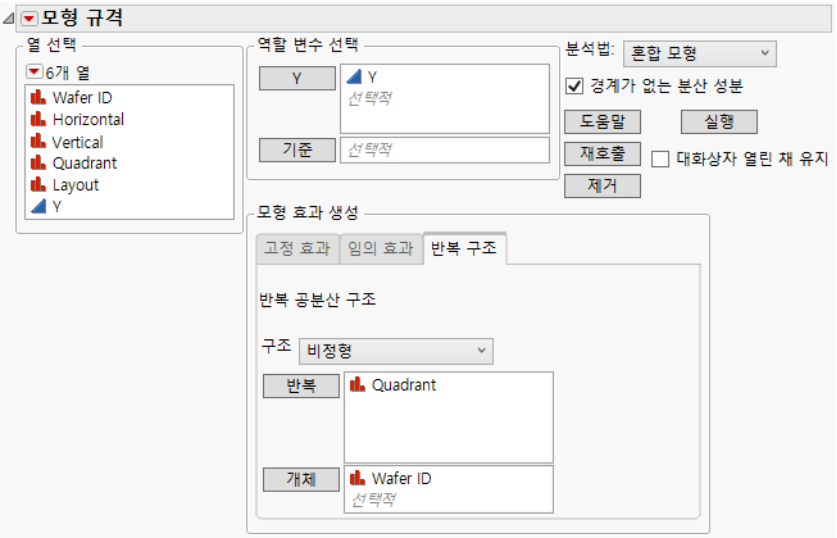
   이 모형 규격을 사용하면 반복 측정값에 대한 Layout 의 효과 및 Layout 과 Quadrant 의 가능한 교호작용을 탐색할 수 있습니다 .

그림 8.39 완료된 고정 효과 탭을 보여 주는 모형 적합 시작 창



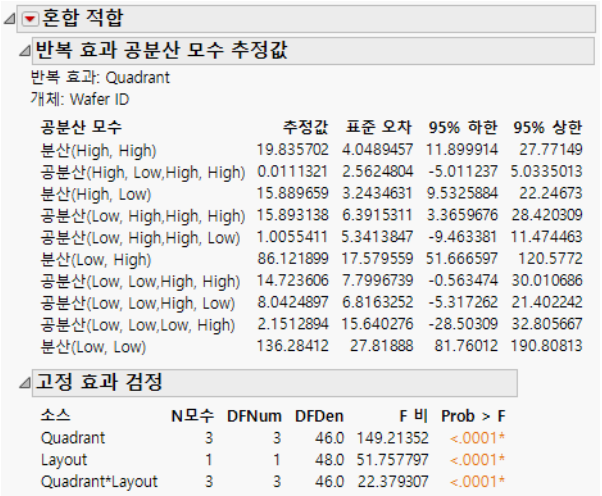
6. **반복 구조** 탭을 선택합니다 .
7. " 구조 " 목록에서 **비정형**을 선택합니다 .
8. Quadrant 를 선택하고 **반복**을 클릭합니다 .
9. Wafer ID 를 선택하고 **개체**를 클릭합니다 .

그림 8.40 반복 구조 탭을 보여 주는 모형 적합 시작 창



10. **실행**을 클릭합니다 .

그림 8.41 고정 효과 모수 추정값 보고서가 닫혀 있는 혼합 적합 보고서



혼합 적합

반복 효과 공분산 모수 추정값

반복 효과: Quadrant
개체: Wafer ID

공분산 모수	추정값	표준 오차	95% 하한	95% 상한
분산(High, High)	19.835702	4.0489457	11.899914	27.77149
공분산(High, Low,High, High)	0.0111321	2.5624804	-5.011237	5.0335013
분산(High, Low)	15.889659	3.2434631	9.5325884	22.24673
공분산(Low, High,High, High)	15.893138	6.3915311	3.3659676	28.420309
공분산(Low, High,High, Low)	1.0055411	5.3413847	-9.463381	11.474463
분산(Low, High)	86.121899	17.579559	51.666597	120.5772
공분산(Low, Low,High, High)	14.723606	7.7996739	-0.563474	30.010686
공분산(Low, Low,High, Low)	8.0424897	6.8163252	-5.317262	21.402242
공분산(Low, Low,Low, High)	2.1512894	15.640276	-28.50309	32.805667
분산(Low, Low)	136.28412	27.81888	81.76012	190.80813

고정 효과 검정

소스	N모수	DFNum	DFDen	F 비	Prob > F
Quadrant	3	3	46.0	149.21352	<.0001*
Layout	1	1	48.0	51.757797	<.0001*
Quadrant*Layout	3	3	46.0	22.379307	<.0001*

" 반복 효과 공분산 모수 추정값 " 보고서에는 네 가지 반응에 대한 추정 분산 및 공분산이 제공됩니다 . Low, High, High, High 의 공분산에 대한 신뢰 구간에 0 이 포함되지 않습니다 . 이는 두 사분면의 측정값 간에 양의 공분산이 있음을 나타냅니다 . 이 정보는 혼합 모형 분석에 사용되며 반응이 독립적으로 모델링되는 경우 사용할 수 없습니다 .

" 고정 효과 검정 " 보고서에서는 유의한 레이아웃 (Layout)* 특성 (Characteristic) 교호작용이 있음을 나타냅니다 .

JMP PRO 프로파일러를 사용하여 레이아웃 * 특성 교호작용 탐색

1. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **주변 모형 추론 > 프로파일러**를 선택합니다 .
2. 프로파일러 그림에서 먼저 Layout A 에 빨간색 수직 점선을 설정한 후 Layout B 에 설정하여 전체 사분면에서 Y 의 예측값을 비교합니다 .

그림 8.42 레이아웃 A 의 사분면에 대한 프로파일러

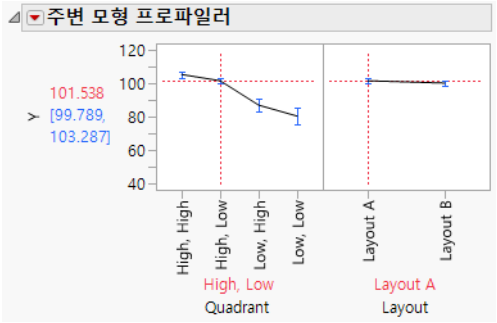
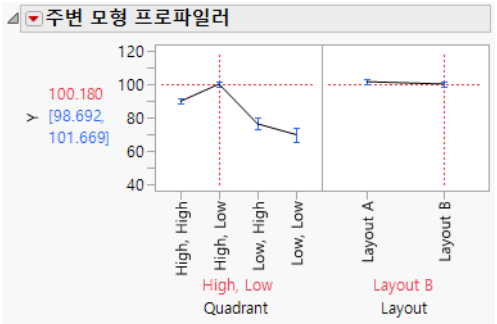


그림 8.43 레이아웃 B 의 사분면에 대한 프로파일러



각 Layout 설정에서 프로파일의 차이는 유의한 교호작용에 대한 정보를 제공합니다 . 교호작용은 부분적으로 High, High 사분면의 차이로 인해 발생하는 것으로 보입니다 .

JMP PRO 레이아웃 및 사분면별 Y 그림

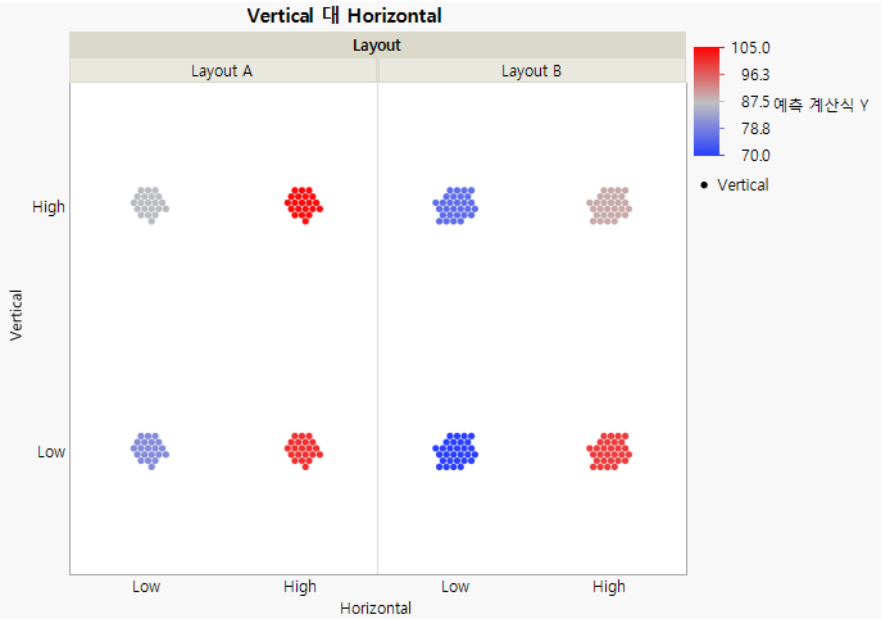
그래프 빌더를 사용하여 교호작용의 특성을 탐색합니다 .

1. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식**을 선택합니다 .

   예측 계산식이 데이터 테이블의 **예측 계산식** Y 열에 저장됩니다 .
2. **그래프 > 그래프 빌더**를 선택합니다 .
3. Horizontal 을 **X** 영역으로 드래그합니다 .
4. Vertical 을 **Y** 영역으로 드래그합니다 .

5. 예측 계산식 Y 를 **색상** 영역으로 드래그합니다 .
6. Layout 을 **펼침** 영역으로 드래그합니다 .
7. **완료**를 클릭하여 제어판을 숨깁니다 .

그림 8.44 완료된 그래프 빌더 그림



참고 : 그래프 빌더의 점은 무작위로 지터링되므로 그림 8.44 의 그림과 정확히 일치하지 않을 수 있습니다 .

이 그림에서는 색상 강도 척도를 사용하여 8 개의 Layout 및 Quadrant 조합에 대한 예측된 차이를 보여 줍니다 . High, Low 사분면에 대한 예측값은 오른쪽 하단에 있습니다 . 색상 그래디언트는 이러한 예측값에 대해 상대적으로 작은 차이를 보여 줍니다. 다른 차이는 명확하게 표시됩니다 .

JMP PRO 혼합 모형 분석법에 대한 통계 상세 정보

- " 수렴 스코어 검정 "
- " 랜덤 계수 모형 "
- " 반복 측정 "
- " 반복 공분산 구조 "
- " 공간 및 시간 변동 "
- "Kackar-Harville 수정 "

JMP PRO 수렴 스코어 검정

수렴 실패 경고는 알 수 없는 MLE(최대 가능도 추정값)가 모형 적합 알고리즘의 최종 반복에서 제공된 모수와 일치한다는 가설에 대한 스코어 검정을 보여 줍니다. 이 가설 검정은 상대 그래디언트 기준이 스코어 검정 통계량과 대수적으로 동등하므로 가능합니다. 또한 스코어 검정에는 실제 MLE에 대한 정보가 필요 없습니다.

JMP PRO 스코어 검정

먼저 단일 모수 θ 를 가정해 보겠습니다. l 을 θ 의 로그 가능도 함수로 하고 $\mathbf{x}$ 를 데이터로 합니다. 점수는 다음과 같이 θ 에 대한 로그 가능도 함수의 도함수입니다.

$$U(\theta) = \frac{\partial l(\theta|\mathbf{x})}{\partial \theta}$$

관측된 정보:

$$I(\theta) = -\frac{\partial^2}{\partial \theta^2} l(\theta|\mathbf{x})$$

$H_0: \theta = \theta_0$ 의 스코어 검정에 대한 통계량:

$$S(\theta_0) = \frac{U(\theta_0)^2}{I(\theta_0)}$$

이 통계량은 귀무가설 하에서 자유도가 1인 점근적 카이제곱 분포를 따릅니다.

스코어 검정을 여러 모수로 일반화할 수 있습니다. 모수 θ의 벡터를 가정해 보겠습니다. $H_0: \theta = \theta_0$의 스코어 검정에 대한 통계량은 다음과 같습니다.

$$S(\theta_0) = \mathbf{U}'(\theta_0)\mathbf{I}^{-1}(\theta_0)\mathbf{U}(\theta_0)$$

다음은 각 요소에 대한 설명입니다.

$$\mathbf{U}(\theta) = \frac{\partial l(\theta|\mathbf{x})}{\partial \theta}$$

및

$$\mathbf{I}(\theta) = -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \theta(\partial \theta')}$$

$\mathbf{U}'$는 $\mathbf{U}$ 행렬의 전치를 나타냅니다.

검정 통계량은 자유도가 k인 점근적 카이제곱 분포입니다. 여기서 k는 비유계 모수의 수입니다.

JMP PRO 상대 그래디언트

혼합 모형 적합 절차의 수렴 기준은 상대 그래디언트 $g'\mathbf{H}^{-1}g$를 기반으로 합니다. 여기서 $g(\theta) = \mathbf{U}(\theta)$는 로그 가능도 함수의 그래디언트이고, $\mathbf{H}(\theta) = -\mathbf{I}(\theta)$는 헤시안 행렬입니다.

θ_0을 알고리즘이 종료되는 θ의 값이라고 하겠습니다. θ_0에서 평가된 상대 그래디언트가 스코어 검정 통계량입니다. p 값은 자유도가 k인 카이제곱 분포를 사용하여 계산됩니다. 이 p 값은 알 수 없는 MLE 값이 θ_0와 일치하는지 여부를 나타냅니다. "임의 효과 공분산 모수 추정값" 보고서에 나열된 비유계 모수의 수는 k입니다.

JMP PRO 랜덤 계수 모형

표준 랜덤 계수 모형은 각 개체에 대한 랜덤 절편과 기울기를 지정합니다. y_{ij}가 i 번째 개체에 대한 j 번째 관측값의 측정값을 나타낸다고 하겠습니다. 그러면 랜덤 계수 모형을 다음과 같이 지정할 수 있습니다.

$$y_{ij} = a_i + b_i x_{ij} + e_{ij}$$

다음은 각 요소에 대한 설명입니다.

$$i = 1, 2, \ldots, t$$

$$j = 1, 2, \ldots, n_i$$

$$\begin{bmatrix} a_i \\ b_i \end{bmatrix} \sim iid\ N\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{G}\right)$$

$$\mathbf{G} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$$

및

$$e_{ij} \sim iid\ N(0, \sigma^2)$$

추정되는 고정 성분 및 랜덤 성분을 반영하도록 모형을 재구성할 수 있습니다.

$$y_{ij} = (\alpha + a_i^*) + (\beta + b_i^*)x_{ij} + e_{ij}$$

다음은 각 요소에 대한 설명입니다.

$$a_i^* = \alpha_i - \alpha$$

$$b_i^* = \beta_i - \beta$$

및

$$\begin{bmatrix} a_i^* \\ b_i^* \end{bmatrix} \sim iid\ N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{G}\right)$$

G 및 e_{ij}는 위와 같이 정의됩니다.

JMP PRO 반복 측정

반복 측정 모형의 형식은 $y_{ijk} = \alpha_{ij} + s_{ik} + e_{ijk}$ 이고 , 다음은 각 요소에 대한 설명입니다 .

α_{ij} 를 처리 및 시간 요인으로 작성할 수 있습니다 .
s_{ik} 는 i 번째 처리에 할당된 k 번째 개체의 임의 효과입니다 .
$j = 1,\ldots,m$ 은 시간에 따른 반복 측정을 나타냅니다 .

s_{ik}를 독립적이며 동일한 분포를 따르는 $N(0, \sigma_s^2)$ 변수로 가정합니다. 처리 요인의 수는 t 로 나타내고 개체 수는 s 로 나타냅니다 . 그러면 e_{ijk} 의 분포는 $N(0, \Sigma)$ 이고 , 다음과 같습니다 .

$$\Sigma = \mathbf{I}_{ts} \otimes Var(\mathbf{y}_{ik} | s_{ik})$$

및

$$\mathbf{y}'_{ik} | s_{ik} = \left(\left[y_{i1k}\ y_{i2k}\ \cdots\ y_{imk} \right] \middle| s_{ik} \right)$$

처리 내에서 ik번째 개체에 해당하는 공분산 행렬 Σ의 대각 성분 블록을 Σ_{ik}로 나타냅니다. 즉, $\Sigma_{ik} = Var(\mathbf{y}_{ik} | s_{ik})$입니다. 일반적으로 개체 내에서 시간에 따른 관측값은 독립적이지 않으므로 $y_{ijk} | s_{ik}$의 분산을 추정해야 합니다. 상관관계를 설명하지 못하면 추론이 왜곡될 수 있습니다.

Σ_{ik}에 사용할 수 있는 공분산 구조에 대한 자세한 내용은 "반복 공분산 구조"(393페이지) 및 "공간 및 시간 변동"(398 페이지)에서 확인하십시오.

JMP PRO 반복 공분산 구조

이 섹션에서는 다음 공분산 구조에 대한 파라미터화를 제공합니다.

- " 이분산 공분산 구조 "(394 페이지)
- " 비정형 공분산 구조 "(394 페이지)
- " 복합 대칭 공분산 구조 "(395 페이지)
- "AR(1) 공분산 구조 "(396 페이지)
- "Toeplitz 공분산 구조 "(397 페이지)
- "Antedependent 공분산 구조 "(397 페이지)

JMP PRO 이분산 공분산 구조

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ & \sigma_2^2 & 0 & \dots & 0 \\ & & \sigma_3^2 & \dots & 0 \\ & & & \ddots & \vdots \\ & & & & \sigma_m^2 \end{bmatrix}$$

시간 j 에서 관측값 간의 분산은 다음과 같습니다 .

$$\sigma_j^2 = Var(y_{ijk} | s_{ik})$$

분산은 반복 열의 수준마다 다를 수 있습니다. 서로 다른 수준에서 관측값 간의 공분산은 0 입니다.

JMP PRO 비정형 공분산 구조

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1[m-1]} & \sigma_{1m} \\ & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2[m-1]} & \sigma_{2m} \\ & & \sigma_3^2 & \dots & \sigma_{3[m-1]} & \sigma_{3m} \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \sigma_{[m-1]m} \\ & & & & & \sigma_m^2 \end{bmatrix}$$

시간 j에서 관측값 간의 분산은 다음과 같습니다.

$$\sigma_j^2 = Var(y_{ijk} | s_{ik})$$

시간 j 및 j'에서 관측값 간의 공분산은 다음과 같습니다.

$$\sigma_{jj'} = Cov(y_{ijk}, y_{ij'k} | s_{ik})$$

분산은 반복 열의 수준마다 다를 수 있습니다. 서로 다른 수준에서 관측값 간의 공분산은 고유합니다.

JMP PRO 복합 대칭 공분산 구조

JMP에서 복합 대칭 공분산 구조는 독립 오차 방법의 혼합 모형을 사용하여 구현됩니다. 임의 효과는 두 가지 범주(G측 또는 R측)로 분류됩니다. 자세한 내용은 Searle et al. 연구 자료(1992)에서 확인하십시오.

G 측 임의 효과는 임의 효과에 대한 설계 행렬과 관련됩니다. R 측 임의 효과는 잔차 오차와 관련됩니다. 개체 내 분산은 설계 구조의 일부이며 G 측에서 모델링됩니다. 개체 간 분산은 잔차 구조에 속하며 R 측에서 모델링됩니다. 독립 구조의 분산은 다음과 같은 방식으로 모델링됩니다.

- 임의 효과 G 측 분산은 $s_{ik} \sim \text{iid } N(0, \sigma_s^2)$ 으로 모델링됩니다.
- R 측 분산은 $e_{ijk} \sim \text{iid } N(0, \sigma^2)$ 으로 모델링됩니다.

그러면 공분산 행렬이 다음과 같이 정의됩니다.

$$\Sigma_{ik} = \sigma_s^2 \mathbf{J} + \sigma^2 \mathbf{I} = \begin{bmatrix} \sigma_s^2 + \sigma^2 & \sigma_s^2 & \cdots & \sigma_s^2 \\ & \sigma_s^2 + \sigma^2 & \cdots & \sigma_s^2 \\ & & \ddots & \vdots \\ & & & \sigma_s^2 + \sigma^2 \end{bmatrix}$$

여기서 **J** 는 *1* 로 구성된 행렬이고 **I** 는 항등 행렬입니다.

모든 분산을 R 측에서 모델링할 수도 있습니다. 가우스 가정을 고려할 때 이 복합 대칭 공분산 구조는 독립 모형 (SAS 의 Type=CS) 과 동등합니다. 이 구조는 JMP 에서 반복 구조 탭의 복합 대칭 구조를 통해 사용할 수 있습니다. 여기서 반복 관측값 쌍 간의 상관관계는 관측값 간의 시간 차이에 관계없이 동일합니다. 따라서 상관 행렬을 다음과 같이 지정할 수 있습니다.

$$C = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ & 1 & \cdots & \rho \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

또한 JMP에서 복합 대칭 구조를 사용하면 모든 시점에서 관측값 간에 공통 분산(σ_e^2)을 가정할 수 있습니다. 그러면 공분산 구조는 $\Sigma = \sigma_e^2 C$ 이고, 다음과 같습니다.

$$\sigma_e^2 = \sigma_s^2 + \sigma^2$$

및

$$\rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2}.$$

여기서 ρ 은 클래스 간 상관계수이고 σ_e^2 는 잔차 분산입니다 . JMP 에서 복합 대칭 이분산 구조를 사용할 수도 있습니다 . 이 방법은 시점에 따라 분산이 다를 수 있습니다 . 그러면 다음과 같은 공분산 행렬이 생성됩니다 .

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \cdots & \rho\sigma_1\sigma_{m-1} & \rho\sigma_1\sigma_m \\ & \sigma_2^2 & \rho\sigma_2\sigma_3 & \cdots & \rho\sigma_2\sigma_{m-1} & \rho\sigma_2\sigma_m \\ & & \sigma_3^2 & \cdots & \rho\sigma_3\sigma_{m-1} & \rho\sigma_3\sigma_m \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \rho\sigma_{m-1}\sigma_m \\ & & & & & \sigma_m^2 \end{bmatrix}$$

JMP PRO AR(1) 공분산 구조

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho^{|t_1-t_2|}\sigma^2 & \rho^{|t_1-t_3|}\sigma^2 & \cdots & \rho^{|t_1-t_{m-1}|}\sigma^2 & \rho^{|t_1-t_m|}\sigma^2 \\ & \sigma^2 & \rho^{|t_2-t_3|}\sigma^2 & \cdots & \rho^{|t_2-t_{m-1}|}\sigma^2 & \rho^{|t_2-t_m|}\sigma^2 \\ & & \sigma^2 & \cdots & \rho^{|t_3-t_{m-1}|}\sigma^2 & \rho^{|t_3-t_m|}\sigma^2 \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma^2 & \rho^{|t_{m-1}-t_m|}\sigma^2 \\ & & & & & \sigma^2 \end{bmatrix}$$

여기서 t_j는 관측값 j의 시간입니다. 이 구조에서 지정된 시간의 관측값은 동일한 분산(σ^2)을 가집니다. ρ(-1 < ρ < 1) 모수는 한 단위 시간 간격을 갖는 두 관측값 간의 상관입니다. 관측값 간의 시간 차이가 증가하면 ρ 모수가 더 크게 거듭제곱되므로 공분산이 감소합니다. 대부분 AR(1)은 제1종 오류 제어를 유지하면서 더 높은 검정력으로 개체 내 상관을 위한 적절한 모형을 제공합니다.

JMP PRO Toeplitz 공분산 구조

Toeplitz 구조에서 고정된 수의 시간 단위로 구분된 관측값은 동일한 상관을 가집니다. AR(1) 상관 구조와 달리 고정된 시간 차이에서 Toeplitz 상관은 임의적입니다. d 단위 간격의 관측값에 대한 상관을 ρ_d 로 나타냅니다. 상관 행렬은 다음과 같이 정의됩니다.

$$C = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{m-1} & \rho_m \\ & 1 & \rho_1 & \cdots & \rho_{m-2} & \rho_{m-1} \\ & & 1 & \cdots & \rho_{m-3} & \rho_{m-2} \\ & & & \ddots & \vdots & \vdots \\ & & & & 1 & \rho_1 \\ & & & & & 1 \end{bmatrix}$$

JMP의 다음 두 옵션은 이 상관 행렬을 사용합니다.

- Toeplitz 구조는 모든 시점의 관측값에 대해 공통 분산 (σ^2) 을 가정합니다. 공분산 구조는 $\Sigma = \sigma^2 C$ 입니다.
- Toeplitz 이분산 구조를 사용하면 시점에 따라 분산이 다를 수 있습니다.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_3 & \cdots & \rho_{m-1}\sigma_1\sigma_{m-1} & \rho_m\sigma_1\sigma_m \\ & \sigma_2^2 & \rho_1\sigma_2\sigma_3 & \cdots & \rho_{m-2}\sigma_2\sigma_{m-1} & \rho_{m-1}\sigma_2\sigma_m \\ & & \sigma_3^2 & \cdots & \rho_{m-3}\sigma_3\sigma_{m-1} & \rho_{m-2}\sigma_3\sigma_m \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \rho_1\sigma_{m-1}\sigma_m \\ & & & & & \sigma_m^2 \end{bmatrix}$$

JMP PRO Antedependent 공분산 구조

Antedependence 모형은 유연하고 시간에 따라 상관 구조가 달라질 수 있는 일반 모형입니다. 이 모형에서 인접한 시점 j - 1과 j에서 두 관측값 간의 상관은 고유하며 $\rho_{j[j-1]}$로 나타냅니다.

인접하지 않은 시점 j와 j'에서 관측값 쌍 간의 상관은 그 사이에 있는 모든 인접 상관의 곱입니다.

$$Corr(y_{ijk}, y_{ij'k} | s_{ik}) = \prod_{k=j}^{j'-1} \rho_{k[k-1]}$$

예를 들어 시점 j=2와 j'=6에서 관측값 쌍 간의 상관은 $\rho_{21}\rho_{32}\rho_{43}\rho_{54}$가 됩니다.

상관 행렬은 다음과 같이 정의됩니다 .

$$C = \begin{bmatrix} 1 & \rho_{10} & \rho_{10}\rho_{21} & \cdots & \rho_{10}\cdots\rho_{[m-1][m-2]} & \rho_{10}\cdots\rho_{m[m-1]} \\ & 1 & \rho_{21} & \cdots & \rho_{21}\cdots\rho_{[m-1][m-2]} & \rho_{21}\cdots\rho_{m[m-1]} \\ & & 1 & \cdots & \rho_{32}\cdots\rho_{[m-1][m-2]} & \rho_{32}\cdots\rho_{m[m-1]} \\ & & & \ddots & \vdots & \vdots \\ & & & & 1 & \rho_{m[m-1]} \\ & & & & & 1 \end{bmatrix}$$

JMP 의 다음 두 옵션은 이 상관 행렬을 사용합니다 .

- Antedependent 등분산 구조는 상관이 달라질 수 있다는 것은 허용하지만 전체 관측 시간에서 동일한 분산을 가정합니다 . 관측값 간의 분산은 항상 σ^2 이고 공분산 행렬은 $\Sigma = \sigma^2 C$ 입니다 .
- Antedependent 구조를 사용하면 지정된 시간에서 관측값 간의 분산이 달라질 수 있습니다 . 시간 j 에서 관측값 간의 분산은 σ_j^2 로 나타냅니다 . 그러면 공분산 행렬이 다음과 같이 정의됩니다 .

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{10}\sigma_1\sigma_2 & \rho_{10}\rho_{21}\sigma_1\sigma_3 & \cdots & \rho_{10}\cdots\rho_{[m-1][m-2]}\sigma_1\sigma_{m-1} & \rho_{10}\cdots\rho_{m[m-1]}\sigma_1\sigma_m \\ & \sigma_2^2 & \rho_{21}\sigma_2\sigma_3 & \cdots & \rho_{21}\cdots\rho_{[m-1][m-2]}\sigma_2\sigma_{m-1} & \rho_{21}\cdots\rho_{m[m-1]}\sigma_2\sigma_m \\ & & \sigma_3^2 & \cdots & \rho_{32}\cdots\rho_{[m-1][m-2]}\sigma_3\sigma_{m-1} & \rho_{32}\cdots\rho_{m[m-1]}\sigma_3\sigma_m \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \rho_{m[m-1]}\sigma_{m-1}\sigma_m \\ & & & & & \sigma_m^2 \end{bmatrix}$$

JMP PRO 공간 및 시간 변동

단순 모형 $y_i = \mu + e_i$를 가정해 보겠습니다. 공간 또는 시간 구조는 오차 항 e_i를 통해 모델링됩니다. 일반적으로 공간 상관 모형은 $Var(e_i) = \sigma_i^2$ 및 $Cov(e_i, e_j) = \sigma_{ij}$로 정의할 수 있습니다.

$\mathbf{s}_i$가 y_i의 위치를 나타낸다고 하겠습니다. 이때 $\mathbf{s}_i$는 공간 또는 시간을 반영하는 좌표로 지정됩니다. 일반적으로 공간 또는 시간 구조는 공분산이 $\mathbf{s}_i$와 $\mathbf{s}_j$ 사이의 유클리드 거리 d_{ij}에 대한 함수라고 가정하여 제한됩니다. 공분산은 $Cov(e_i, e_j) = \sigma^2[f(d_{ij})]$로 작성할 수 있으며, 여기서 $f(d_{ij})$는 관측값 y_i와 y_j 간의 상관을 나타냅니다.

둘 이상의 위치 좌표가 있는 경우 $f(d_{ij})$가 방향에 종속되지 않으면 공분산 구조가 **등방성**입니다. 그렇지 않으면 구조가 **이방성**입니다.

JMP PRO 공간 상관 구조

JMP에서 사용할 수 있는 공간 모형의 상관 구조는 아래와 같습니다. 이러한 구조는 제약된 경우 외에는 양수인 ρ에 의해 파라미터화됩니다.

- 구형

$$f(d_{ij}) = [1 - 1.5(d_{ij}/\rho) + 0.5(d_{ij}/\rho)^3] \times 1_{\{d_{ij} < \rho\}}$$

여기서, $1_{\{d_{ij} < \rho\}} = \begin{cases} 1\text{단, } d_{ij} < \rho \\ 0, \text{if } d_{ij} \geq \rho \end{cases}$

- 지수

$$f(d_{ij}) = \exp(-d_{ij}/\rho)$$

- 가우스

$$f(d_{ij}) = \exp(-d_{ij}^2/\rho^2)$$

- 거듭제곱

$$f(d_{ij}) = \rho^{d_{ij}}$$

이방성 모형의 경우 상관 함수는 각 방향에 대해 ρ_κ 모수를 포함합니다.

JMP PRO 변동도

공간 과정이 2차 정상 상태인 경우 "공간 상관 구조"(399페이지)에 나열된 구조는 **변동도**를 정의합니다. 지리 통계학에서 차용된 변동도는 공간 변동을 설명하고 추정하는 표준 도구입니다. 여기에서는 반분산을 사용하여 관측값 간의 거리 d_{ij}에 대한 함수로 공간 변동을 측정합니다.

$Z(\mathbf{s})$가 위치 $\mathbf{s}$에서 반응의 값을 나타낸다고 하겠습니다. $\mathbf{s}_i$와 $\mathbf{s}_j$에서 관측값 간의 **반분산**은 다음과 같이 정의됩니다.

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = (Var(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)))/2$$

반응에 상수 평균이 있으면 표현식을 다음과 같이 단순화할 수 있습니다.

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = E[(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2]/2$$

등방성 과정의 경우 반분산은 점 사이의 거리 (h)에만 종속되며 함수를 다음과 같이 지정할 수 있습니다.

$$\gamma(h) = E[(Z(\mathbf{s}_i) - Z((\mathbf{s}_i + h)))^2]/2$$

다음은 변동도와 관련된 용어입니다 .

너겟 절편으로 정의됩니다 . 너겟은 $h = 0$ 에서 불연속 이동을 나타냅니다 .

문턱 (Sill) 더 먼 거리에 대해 도달한 평탄 시작 구간에서의 반변동도 값으로 정의됩니다 . 이 값은 관측값의 분산에 해당합니다 . 너겟 효과가 없는 모형에서 문턱 (sill) 은 σ^2 입니다 . 너겟 효과가 있는 모형에서 문턱 (sill) 은 $\sigma^2 + c_1$ 이며 , 여기서 c_1 은 너겟을 나타냅니다 . **부분 문턱 (sill)** 은 σ^2 으로 정의됩니다 .

범위 반변동도가 문턱 (sill) 에 도달하는 거리로 정의됩니다 . 범위 미만의 거리에서는 관측값이 공간적으로 상관됩니다 . 범위를 초과하거나 같은 거리의 경우 공간 상관은 사실상 0 입니다 . 구형 모형에서 ρ 는 범위입니다 . 지수 모형에서는 3ρ 가 실제적 범위입니다 . 가우스 모형에서는 $\rho\sqrt{3}$ 이 실제적 범위입니다 . 실제적 범위는 공분산이 문턱 (sill) 의 95% 로 축소되는 거리로 정의됩니다 .

그림 8.34(380 페이지) 에서 반복 효과 공분산 모수 추정값은 다양한 반변동도 특징을 나타냅니다 .

공간 구형 범위 (ρ) 의 추정값입니다 .

너겟 c_1 의 척도화된 추정값입니다 . 잔차 x 너겟이 c_1 입니다 .

잔차 너겟이 없는 모형의 문턱 (sill) 또는 부분 문턱 (sill) 입니다 .

JMP PRO 변동도 추정값

지정된 등방성 공간 구조에 대해 추정된 변동도는 "공간 상관 구조"(399 페이지) 에 나오는 적절한 함수에 대한 관측 데이터의 비선형 최소 제곱 적합을 사용하여 구할 수 있습니다 .

JMP PRO 경험적 반분산

경험적 반분산을 계산하려면 변동도 공분산을 위해 선택된 변수에 대한 모든 점 쌍 사이의 거리가 계산됩니다 . 거리 범위는 10 개의 동일 구간으로 구분됩니다 . 데이터에 10 개의 구간이 허용되지 않으면 가능한 한 많은 구간이 생성됩니다 .

점 쌍으로 구성된 거리 클래스가 생성됩니다 . h 번째 거리 클래스는 거리가 h 번째 구간에 속하는 모든 점 쌍으로 구성됩니다 .

다음 표기를 고려하십시오 .

n 점 쌍의 총 수

C_h 거리가 h 번째로 큰 구간에 속하는 점으로 구성된 거리 클래스

$Z(\mathbf{x})$ $\mathbf{x}$ 에서 반응의 값 ($\mathbf{x}$ = 시간 또는 공간 좌표의 벡터)

$\gamma(h)$ 거리 클래스 C_h 에 대한 반분산

반분산 함수 (γ) 는 다음과 같이 정의됩니다 .

$$\gamma(h) = \begin{cases} \frac{1}{2n}\left\{ \sum_{(x,y) \text{ in } C_h} [Z(\mathbf{x}) - Z(\mathbf{y})]^2 \right\} & \text{의 경우 } h > 0 \\ \hat{c}_1 & \text{for } h = 0 \end{cases}$$

여기서 $\hat{c}_1$ 은 너겟 효과의 추정값입니다.

JMP PRO Kackar-Harville 수정

고정 효과의 분산 행렬은 항상 Kackar-Harville 수정을 포함하도록 변경됩니다. BLUP 의 분산 행렬 및 BLUP 와 고정 효과 간의 공분산은 Kackar-Harville 수정 대상이 아닙니다. 이 방법의 근거는 임의 효과에 수준이 많이 있을 때 BLUP 에 대한 수정은 계산 및 메모리 사용량이 많을 수 있다는 것입니다. SAS 에서는 DDFM=KENWARDROGER 가 설정된 경우에만 Kackar-Harville 수정이 고정 효과와 BLUP 에 대해 모두 수행됩니다.

공분산 모수와 관련하여 0 이 아닌 2 차 도함수가 있는 공분산 구조의 경우 Kenward-Roger 공분산 행렬 조정에 2 차 항이 포함됩니다. 이 항으로 인해 표준 오차 축소가 발생할 수 있습니다. 또한 조정된 결과 공분산 행렬은 부정부호 행렬이 될 수 있으며 재파라미터화에서 불변하지 않습니다. 1 차 Kenward-Roger 공분산 행렬 조정은 계산에서 2 차 도함수를 제거합니다. 모든 공간 구조 및 AR(1) 구조는 일반적으로 0 이 아닌 2 차 도함수를 유도하는 공분산 구조입니다.

JMP 에서는 Kenward-Roger 1 차 조정을 구현하므로 다음 사항에 유의하십시오 .

- 고정 효과 모수만 포함하는 선형 결합의 표준 오차는 PROC MIXED DDFM=KENWARDROGER (FIRSTORDER) 와 일치합니다 . 이때 PROC MIXED 와 JMP 에서 사용하는 서로 다른 파라미터화 간의 변환에 주의를 기울여야 합니다 .
- BLUP 모수만 포함하는 선형 결합의 표준 오차는 PROC MIXED DDFM=SATTERTHWAITE 와 일치합니다 .
- 불균형 데이터의 경우 고정 효과와 BLUP 를 모두 포함하는 선형 결합의 표준 오차는 DDFM 옵션에 대한 PROC MIXED 와 일치하지 않습니다 . 그러나 이러한 표준 오차는 DDFM=SATTERTHWAITE 옵션과 DDFM=KENWARDROGER 옵션을 사용하여 구한 값 사이에 있습니다 . 균형 데이터의 경우 Kackar-Harville 수정이 null 이므로 JMP 는 DDFM 옵션에 관계없이 SAS 와 일치합니다 .

JMP PRO 자유도

고정 효과 모수의 선형 결합만 포함하는 검정의 자유도는 1 차 Kenward-Roger 수정을 사용하여 계산됩니다. 따라서 이러한 검정의 JMP 결과는 DDFM=KENWARDROGER(FIRSTORDER) 옵션을 사용하여 PROC MIXED와 일치합니다. 선형 결합에 BLUP가 있는 경우 JMP는 Satterthwaite 근사를 사용하여 자유도를 구합니다. 결과는 이전 단락에서 표준 오차에 대해 설명한 것과 유사한 패턴을 따릅니다.

Kackar-Harville 수정 및 Kenward-Roger DF 방법에 대한 자세한 내용은 Kenward and Roger 연구 자료(1997)에서 확인하십시오. Satterthwaite 방법은 SAS Institute Inc. (2020d, ch. 83)

9 장

다변량 반응 모형
MANOVA 를 사용하여 관계 적합

다변량 모형은 여러 반응(Y 변수)을 효과 집합에 적합시킵니다. 적절한 반응 설계를 사용하여 전체 Y 변수에 대한 함수를 검정할 수 있습니다.

표준 MANOVA(다변량 분산 분석) 모형을 생성하는 것 외에 다음 기법을 사용할 수도 있습니다.

- 반복 측정 분석 - 각 개체에 대해 반복 측정을 수행하고, 전체 측정에서 개체 사이 효과와 개체 내 효과를 모두 분석하려는 경우에 사용할 수 있습니다. 이 다변량 접근 방식은 전체 측정의 상관 구조가 임의적일 때 특히 중요합니다.
- 정준 상관 - 상관관계가 가장 높은 X 및 Y 변수의 선형 결합을 찾을 수 있습니다.
- 판별 분석 - 여러 그룹의 점과 다변량 평균 사이의 거리 계산식을 찾아 점을 해당 점이 속할 가능성이 가장 높은 그룹으로 분류할 수 있습니다. 판별 분석을 더 완벽하게 구현하려면 **판별** 플랫폼을 사용합니다.

다변량 적합은 모수 추정값과 최소 제곱 평균을 보여 주는 가장 기본적인 예비 분석으로 시작됩니다. 그런 다음 Y 변수에 대한 반응 설계를 지정하고 다변량 검정을 수행할 수 있습니다.

목차

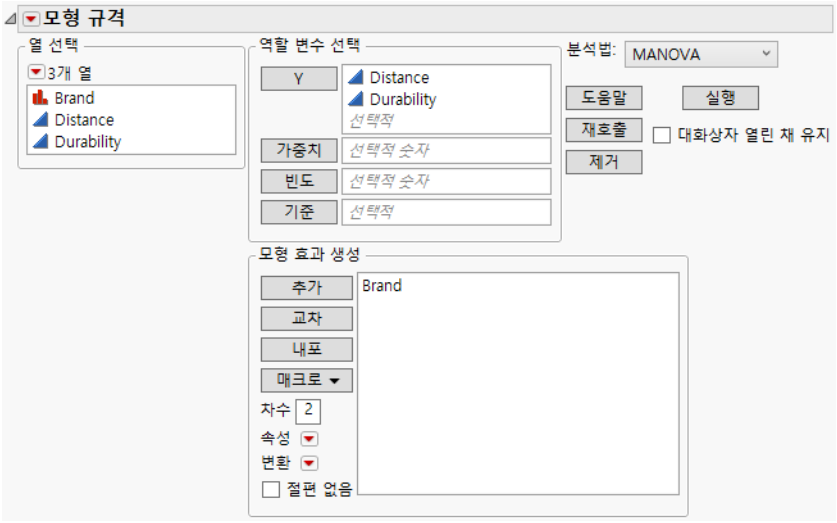
다중 반응 모형의 예 405
MANOVA 보고서 407
MANOVA 적합 옵션 407
반응 지정 408
반응 선택 옵션 408
사용자 검정 옵션 409
모형 효과 옵션 409
다변량 검정 413
확장된 다변량 보고서 414
다변량 검정 비교 415
단변량 검정 및 구형성 검정 415
반복 측정을 사용하는 다변량 모형 416
반복 측정 예 417
복합 다변량 모형의 예 418
판별 분석 421
판별 저장 옵션의 예 421
MANOVA 분석법에 대한 통계 상세 정보 422
다변량 검정 422
근사 F-검정 423
정준 상세 정보 424

다중 반응 모형의 예

이 예에서는 Golf Balls.jmp 샘플 데이터 테이블을 사용합니다 (McClave and Dietrich 1988). 세 가지 브랜드의 골프공에 대해 이동 거리 및 내구성 측정값을 비교한 데이터가 포함되어 있습니다. 랜덤으로 표집한 각 브랜드의 골프공 10 개를 로봇 골퍼가 무작위 순서로 쳤습니다. 세 가지 골프공 브랜드에 대해 거리 및 내구성이 동일하다는 가설을 검정하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Golf Balls.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Distance 와 Durability 를 선택하고 **Y** 를 클릭합니다.
4. Brand 를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **MANOVA** 를 선택합니다.

그림 9.1 MANOVA 설정



6. **실행**을 클릭합니다.

그림 9.2 MANOVA 보고서 창

반응 설계를 아직 지정하지 않았으므로 초기 결과 자체는 흥미롭지 않을 수 있습니다. 반응 설계를 지정하면 다변량 플랫폼에 다변량 추정값 및 검정 테이블이 표시됩니다. 반응 설계를 지정하는 방법에 대한 자세한 내용은 "반응 지정"(408 페이지)에서 확인하십시오.

MANOVA 보고서

MANOVA 보고서 창에는 다음과 같은 요소가 포함되어 있습니다 .

MANOVA 적합 빨간색 삼각형 메뉴 저장 옵션을 포함합니다 . 자세한 내용은 "MANOVA 적합 옵션 "(407 페이지) 에서 확인하십시오 .

반응 지정 다양한 검정에 대한 반응 설계를 지정할 수 있습니다 . 자세한 내용은 " 반응 지정 "(408 페이지) 에서 확인하십시오 .

모수 추정값 각 반응 변수에 대한 모수 추정값을 포함합니다 (표준 오차 또는 *t*- 검정과 같은 상세 정보 제외). 각 반응 변수마다 하나의 열이 있습니다 .

최소 제곱 평균 모든 반응 열의 전체 최소 제곱 평균 , 각 명목형 수준의 최소 제곱 평균 , 평균의 최소 제곱 평균 그림을 보고합니다 .

부분 상관 *X* 효과에 대해 조정된 초기 적합에서 구한 잔차의 공분산 행렬 및 부분 상관 행렬을 표시합니다 .

전체 E&H 행렬 E 행렬 및 H 행렬을 표시합니다 .

- **E** 행렬의 요소는 잔차의 교차곱입니다 .
- **H** 행렬은 가설 제곱합 및 교차곱에 해당합니다 .

전체 모형과 모형의 각 효과에 대한 **H** 행렬이 있습니다 . **E** 및 **H** 행렬의 주대각 요소는 단변량 *F* 검정에 대한 가설 (분자) 및 오차 (분모) 제곱합에 해당합니다 . 주어진 반응 설계에 대한 새 **E** 행렬과 **H** 행렬이 이 초기 행렬에서 구성되고 , 새 행렬에서 다변량 검정 통계량이 계산됩니다 .

MANOVA 적합 옵션

다음과 같은 MANOVA 적합 옵션을 사용할 수 있습니다 .

판별 저장 판별 분석을 수행하고 결과를 데이터 테이블에 저장합니다 . 자세한 내용은 " 판별 분석 "(421 페이지) 에서 확인하십시오 .

예측값 저장 예측 반응을 데이터 테이블에 저장합니다 .

잔차 저장 잔차를 데이터 테이블에 저장합니다 .

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

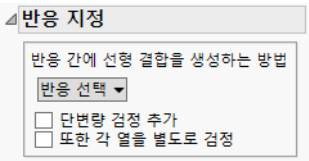
스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

반응 지정

"반응 지정" 패널을 사용하여 다양한 검정에 대한 반응 설계를 지정합니다.

그림 9.3 반응 지정 패널



반응 선택 M 행렬에 대한 선택 옵션을 제공합니다 . 자세한 내용은 " 반응 선택 옵션 "(408 페이지) 에서 확인하십시오 .

단변량 검정 추가 조정 및 조정되지 않은 단변량 반복 측정 검정과 다변량 검정을 수행합니다 . 반복 측정 모형에 사용됩니다 .

또한 각 열을 별도로 검정 각 반응에 대한 단변량 ANOVA 검정과 다변량 검정을 수행합니다 .

다음 버튼은 반응 옵션을 선택한 후에만 사용할 수 있습니다.

실행 분석을 수행하고 다변량 추정값 및 검정을 표시합니다 . 자세한 내용은 " 다변량 검정 "(413 페이지) 에서 확인하십시오 .

도움말 " 반응 지정 " 패널에 대한 도움말을 표시합니다 .

직교화 행렬을 직교정규화합니다 . 직교정규화는 " 합 " 을 제외한 모든 반응 유형에 대한 열 대비 (합이 0) 후에 수행됩니다 .

마지막 열 삭제 변환 차원을 축소합니다 .

반응 선택 옵션

반응 설계는 M 행렬을 구성합니다 . M 행렬의 열은 다변량 분석을 위한 변환 변수 집합을 정의합니다 . **반응 선택** 버튼에는 M 행렬을 위한 다음 옵션이 포함되어 있습니다 .

반복 측정 합 반응과 대비 반응을 모두 구성하고 실행합니다 .

합 단일 값을 제공하는 반응 합입니다 .

항등 각각의 개별 반응을 사용합니다 (항등 행렬).

대비 각 반응과 첫 번째 반응을 비교합니다 .

다항식 직교 다항식 행렬을 구성합니다 .

Helmert 각 반응을 그 아래에 나열된 결합 반응과 비교합니다 .

프로파일 각 반응을 그 다음 반응과 비교합니다 .

평균 각 반응을 나머지 반응의 평균과 비교합니다 .

복합 반응이 두 가지 효과로 구성된 경우 적절한 여러 반응 함수를 생성하고 실행합니다 .

사용자 정의 입력한 사용자 M 행렬을 사용합니다 .

다변량 회귀에 대해 가장 일반적인 반응 설계는 **반복 측정** 및 **항등**입니다 . **대비**, **Helmert**, **프로파일** 및 **평균** 옵션은 동일한 공간을 생성하므로 제공된 검정에 거의 차이가 없습니다 . 그러나 " 최소 제곱 평균 " 및 " 모수 추정값 " 테이블의 검정과 상세 정보는 각 옵션에 따라 서로 다른 결과를 강조합니다 .

반복 측정 및 **복합** 옵션은 반응 효과 이름을 지정하는 대화상자를 표시합니다 . 그런 다음 추가 사용자 입력을 기다리지 않고 여러 반응 함수를 적합시킵니다 . 그렇지 않은 경우 선택에 따라 제어판이 확장되어 지정을 구체화할 수 있습니다 .

사용자 검정 옵션

사용자 검정 옵션을 사용하여 효과 수준의 사용자 검정을 설정합니다 .

참고 : 사용자 검정을 생성하는 방법에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 사용자 검정 "(115 페이지) 에서 확인하십시오 .

모형 효과 옵션

각 효과 이름 옆의 메뉴 아이콘은 다변량 적합에 대한 추가 정보를 요청하는 다음과 같은 옵션을 제공합니다 .

검정 상세 정보 다변량 검정 통계량을 생성하는 데 사용되는 $\mathbf{E}^{-1}\mathbf{H}$ 행렬의 고유값 및 고유 벡터를 표시합니다 . 자세한 내용은 " 검정 상세 정보 "(410 페이지) 에서 확인하십시오 .

중심 그림 검정 공간에서 구성된 처음 두 정준 변수에 대한 중심 (다변량 최소 제곱 평균) 을 그림에 표시합니다 . 자세한 내용은 " 중심 그림 "(411 페이지) 에서 확인하십시오 .

정준 점수 저장 Canon[1], Canon[2] 등의 변수를 현재 데이터 테이블의 열로 저장합니다 . 이 열에는 값과 계산식이 둘 다 있습니다 . 예는 " 정준 점수 저장 "(411 페이지) 에서 확인하십시오 . 기술 상세 정보는 " 정준 상세 정보 "(424 페이지) 에서 확인하십시오 .

대비 대비 대화상자에서 지정한 처리 수준의 통계적 대비를 수행합니다 .

참고 : 대비 명령은 단일 반응의 회귀와 동일합니다 . **최소 제곱 평균 대비** 명령에 대한 설명과 예는 " 표준 최소 제곱 모형 " 장의 " 최소 제곱 평균 대비 ..."(95 페이지) 에서 확인하십시오 .

검정 상세 정보

" 검정 상세 정보 " 보고서에는 전체 모형 또는 지정된 효과의 검정에 대한 정준 상세 정보가 제공됩니다 .

고유값 다변량 검정 통계량을 계산하는 데 사용되는 $\mathbf{E}^{-1}\mathbf{H}$ 행렬의 고유값입니다 .

정준 상관 각 고유값과 관련된 정준 상관입니다 . 이것은 모형의 다른 모든 효과에 대해 수정된 효과와 변환된 반응의 정준 상관관계입니다 .

고유 벡터 $\mathbf{E}^{-1}\mathbf{H}$ 행렬 또는 이와 동등한 $(\mathbf{E}+\mathbf{H})^{-1}\mathbf{H}$ 의 고유 벡터입니다 .

검정 상세 정보의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다 .

   Iris 데이터 (Mardia et al. 1979) 의 Species 에는 Virginica, Setosa 및 Versicolor 라는 세 가지 수준이 있습니다 . 또한 각 표본에 대해 측정된 네 가지 값 (Petal length, Petal width, Sepal length, Sepal width) 이 있습니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Petal length, Petal width, Sepal length 및 Sepal width 를 선택하고 **Y** 를 클릭합니다 .
4. Species 를 선택하고 **추가**를 클릭합니다 .
5. **분석법**에서 **MANOVA** 를 선택합니다 .
6. **실행**을 클릭합니다 .
7. **반응 선택** 버튼을 클릭하고 **항등**을 선택합니다 .
8. **실행**을 클릭합니다 .
9. "Species" 의 빨간색 삼각형을 클릭하고 **검정 상세 정보**를 선택합니다 .

   고유값 , 고유 벡터 및 정준 상관이 나타납니다 .

그림 9.4 검정 상세 정보

Species

검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.0234386	199.1453	8	288	<.0001*
Pillai 대각합	1.1918988	53.4665	8	290	<.0001*
Hotelling-Lawley	32.47732	582.1970	8	203.4	<.0001*
Roy 최대근	32.191929	1166.9574	4	145	<.0001*

고유값	정준 상관
32.1919292	0.98482089
0.28539104	0.47119702
1.235e-15	0
-6.174e-16	0

고유 벡터				
Sepal length	-0.0684059	0.00198791	-0.2350196	0.1176771
Sepal width	-0.1265612	0.1785267	0.21657608	0.04510419
Petal length	0.18155288	-0.0768636	0.23964446	0.06563465
Petal width	0.23180286	0.23417227	-0.2865277	-0.2438953

중심 그림

중심 그림 명령은 "Species"의 빨간색 삼각형에서 액세스할 수 있으며, 검정 공간에서 구성된 처음 두 정준 변수에 대한 중심(다변량 최소 제곱 평균)을 그림에 표시합니다(그림 9.5 참조). 첫 번째 정준 축은 세로 축이므로 검정 공간이 1차원일 경우 중심이 세로 축에 정렬됩니다. 중심점은 95% 신뢰 영역에 해당하는 원과 함께 나타납니다(Mardia et al. 1979). 중심 그림이 효과 검정에서 생성되면 검정 중인 효과에 해당하는 원이 빨간색으로 나타납니다. 다른 원은 파란색입니다. 행렬도 반지름은 검정 공간에서 원래 반응 변수의 방향을 보여 줍니다. 자세한 내용은 "중심 그림 옵션에 대한 상세 정보"(424 페이지)에서 확인하십시오.

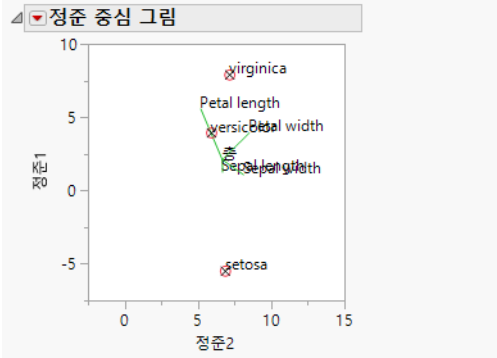
중심값 표시 아이콘을 클릭하면 그림 9.5에 표시된 추가 정보가 나타납니다.

고유값이 있는 첫 번째 정준 축은 두 번째 축보다 훨씬 더 많은 분리를 설명합니다. 평균이 잘 분리(구분)되어 있으며, 첫 번째 그룹이 다른 두 그룹보다 더 멀리 떨어져 있습니다. 첫 번째 정준 변수에서는 petal width 변수에 비해 petal length 변수의 기여도가 높은 것으로 보입니다. 행렬도 반지름 및 관련 고유 벡터를 사용하여 변수 그룹 간의 관계를 확인할 수 있습니다.

그림 9.5 중심 그림 및 중심값

중심값

층	2.10510645	6.66147254	0	0
Sepal length	1.25222977	6.68625758	0	0
Sepal width	1.07496353	8.11458779	0	0
Petal length	5.56320996	5.19742431	0	0
Petal width	3.93284247	8.50789104	0	0
setosa	-5.5024935	6.87660555	0	0
versicolor	3.93015594	5.93357291	0	0
virginica	7.88765689	7.17423914	0	0



정준 점수 저장

Canon[i] 열을 데이터 테이블에 저장합니다. 여기서 i는 Y 변수에 대한 i번째 정준 점수를 나타냅니다. 정준 점수는 다변량 검정 통계량을 생성하는 데 사용되는 $\mathbf{E}^{-1}\mathbf{H}$ 행렬을 기반으로 계산됩니다. 0이 아닌 고유값에 해당하는 고유 벡터에 대해 정준 점수가 저장됩니다.

정준 상관

정준 상관 분석은 특정 명령은 아니지만 다변량 적합 플랫폼에서 일련의 명령을 사용하여 수행할 수 있습니다.

1. " 검정 상세 정보의 예 "(410 페이지) 에서 1 단계 ~ 8 단계를 수행합니다.
2. " 전체 모형 " 의 빨간색 삼각형을 클릭하고 **검정 상세 정보**를 선택합니다.
3. " 전체 모형 " 의 빨간색 삼각형을 클릭하고 **정준 점수 저장**을 선택합니다.

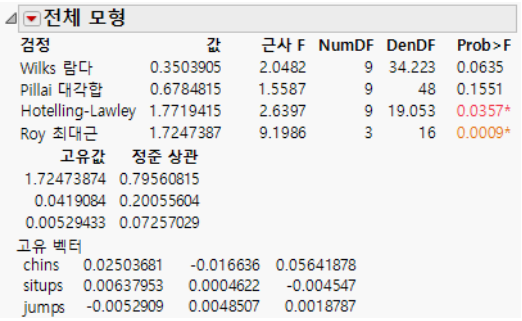
상세 정보에서 고유값 옆에 정준 상관 값(**정준 상관**)이 나열됩니다. 저장된 변수는 Canon[1], Canon[2] 등입니다. 이 열에는 값과 계산식이 둘 다 있습니다.

X 측에 대한 정준 변수를 구하려면 *X* 변수와 *Y* 변수를 교환하여 동일한 단계를 반복합니다. Canon[n] 열이 데이터 테이블에 이미 추가된 경우에는 Canon[n] 2(또는 다른 숫자)라는 고유한 이름의 새 열이 저장됩니다.

다음은 정준 상관 분석의 추가 예입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Exercise.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. chins, situps 및 jumps 를 선택하고 Y 를 클릭합니다.
4. weight, waist 및 pulse 를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **MANOVA** 를 선택합니다.
6. **실행**을 클릭합니다.
7. **반응 선택** 버튼을 클릭하고 **항등**을 선택합니다.
8. **실행**을 클릭합니다.
9. " 전체 모형 " 의 빨간색 삼각형을 클릭하고 **검정 상세 정보**를 선택합니다.
10. " 전체 모형 " 의 빨간색 삼각형을 클릭하고 **정준 점수 저장**을 선택합니다.

그림 9.6 정준 상관



전체 모형

검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.3503905	2.0482	9	34.223	0.0635
Pillai 대각합	0.6784815	1.5587	9	48	0.1551
Hotelling-Lawley	1.7719415	2.6397	9	19.053	0.0357*
Roy 최대근	1.7247387	9.1986	3	16	0.0009*

고유값	정준 상관
1.72473874	0.79560815
0.0419084	0.20055604
0.00529433	0.07257029

고유 벡터

chins	0.02503681	-0.016636	0.05641878
situps	0.00637953	0.0004622	-0.004547
jumps	-0.0052909	0.0048507	0.0018787

출력 정준 변수는 *Y* 변수의 선형 결합으로 표시된 고유 벡터를 사용합니다 . 예를 들어 Canon[1] 은 다음과 같이 계산됩니다 .

```
0.02503681*chins + 0.00637953*situps + -0.0052909*jumps
```

이 정준 분석은 평균이 0 이고 표준편차가 1 인 표준화 변수를 생성하지는 않지만 이러한 기능이 있는 계산기를 사용하여 새 표준화 변수를 쉽게 정의할 수 있습니다 .

다변량 검정

반응 선택의 예

1. " 다중 반응 모형의 예 "(405 페이지) 에 나오는 1 단계 ~ 6 단계를 완료합니다 .
2. **반응 선택** 버튼을 클릭하고 **항등**을 선택합니다 .
3. **실행**을 클릭합니다 .

그림 9.7 다변량 검정 보고서

항등

M 행렬

M-변환 모수 추정값

전체 모형

검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.2117852	15.2485	4	52	<.0001*
Pillai 대각합	0.936487	11.8876	4	54	<.0001*
Hotelling-Lawley	3.0216575	19.4162	4	30.19	<.0001*
Roy 최대근	2.7688021	37.3788	2	27	<.0001*

절편

검정	값	정확한 F	NumDF	DenDF	Prob>F
F-검정	3226.6839	41946.890	2	26	<.0001*

Brand

검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.2117852	15.2485	4	52	<.0001*
Pillai 대각합	0.936487	11.8876	4	54	<.0001*
Hotelling-Lawley	3.0216575	19.4162	4	30.19	<.0001*
Roy 최대근	2.7688021	37.3788	2	27	<.0001*

M 행렬 보고서에는 사용자가 지정한 반응 설계가 제공됩니다 . **M- 변환 모수 추정값** 보고서에는 원래 모수 추정값 행렬에 M 행렬의 전치를 곱한 값이 제공됩니다 .

참고 : 이 장의 처음에 나왔듯이 행렬 이름 **E** 와 **H** 는 오차 및 가설 교차곱을 나타냅니다 . 반응 설계가 지정된 후 **E** 와 **H** 는 반응 설계에 의해 변환된 행렬을 나타내며 실제로는 **M′EM** 및 **M′HM** 입니다 .

확장된 다변량 보고서

다변량 적합에서 가설과 오차로 인한 제곱합은 단일 숫자가 아니라 제곱과 교차곱의 행렬입니다. 그리고 가설 제곱합 및 교차곱 행렬(**H** 또는 **SSCP**)이 잔차 행렬(**E**)과 비교되는 값의 크기를 측정하는 방법이 많이 있습니다. JMP는 문헌에 일반적으로 설명된 네 가지 다변량 검정을 보고합니다. 정확 유의 수준으로 검정을 원하는 경우 참고 도서에서 표를 검색해야 할 수도 있습니다. 다행히 네 가지 검정 모두 근사 *F* 검정으로 변환할 수 있습니다. 반응 설계에서 단일 값이 산출되거나 가설이 단일 자유도인 경우 다변량 검정이 동등하며 정확 *F* 검정을 동일하게 수행합니다. JMP는 검정에 **정확 F**라는 라벨을 지정하고, 그렇지 않은 경우에는 **근사 F** 라벨을 지정합니다.

골프공 예에서는 효과가 하나뿐이므로 전체 모형 검정과 Brand 검정이 동일하며, 근사 *F* 검정으로 네 가지 다변량 검정을 보여 줍니다. DF가 두 개(각 반응에 대해 하나씩)인 단일 절편만 있으므로 이에 대한 *F* 검정이 정확하며 **정확 F** 라벨이 지정됩니다.

"전체 모형", "절편" 및 "Brand" 보고서의 빨간색 삼각형 메뉴에는 고유값, 정준 상관, 중심값 목록, 중심 그림 등의 추가 정보를 생성하는 옵션과 정준 변량을 저장할 수 있는 **저장** 명령이 포함되어 있습니다.

효과(이 예에서는 Brand) 팝업 메뉴에는 대비를 지정하는 옵션도 포함되어 있습니다.

사용자 검정 및 대비 기능은 단일 반응이 있는 회귀의 기능과 동일합니다. 자세한 내용은 "표준 최소 제곱 모형" 장(69페이지)에서 확인하십시오.

MANOVA 테이블 검정에 대한 계산식은 "다변량 검정"(422페이지)에서 확인하십시오.

확장된 다변량 보고서에는 다음 열이 포함됩니다.

검정 테이블의 각 통계 검정에 라벨을 지정합니다. 반응 함수 값(**M** 행렬에 지정된 열)의 수가 1이거나 효과에 반응 함수당 자유도가 하나뿐인 경우 정확 *F* 검정이 제공됩니다. 그렇지 않은 경우 Wilks 람다 (Λ), Pillai 대각합, Hotelling-Lawley 대각합 및 Roy 최대근이라는 네 가지 표준 다변량 검정 통계량에 근사 *F* 검정이 제공됩니다.

값 보고서의 각 다변량 통계 검정에 대한 값입니다.

근사 F 또는 정확 F 다변량 검정에 해당하는 *F* 값입니다. 반응 설계에서 단일 값이 산출되거나 검정 자유도가 1인 경우 정확 *F* 검정입니다.

NumDF 분자 자유도입니다.

DenDF 분모 자유도입니다.

Prob>F *F* 값에 해당하는 유의 확률입니다.

참고 : 구형성 검정 테이블에 대한 자세한 내용은 "단변량 검정 및 구형성 검정"(415페이지)에서 확인하십시오.

다변량 검정 비교

네 가지 표준 다변량 검정은 대개 비슷한 결과를 제공하지만 결과가 다른 경우도 있고 특정 검정이 다른 검정보다 이점이 있을 수도 있습니다. 확실하게 어떤 검정이 가장 우수하다고 할 수는 없습니다 . 일반적으로 검정력 측면에서 선호하는 순서는 다음과 같습니다 .

1. Pillai 대각합
2. Wilks 람다
3. Hotelling-Lawley 대각합
4. Roy 최대근

귀무가설에서 큰 편차가 있고 고유값이 크게 다를 경우 선호 순서는 반대입니다(Seber 1984).

단변량 검정 및 구형성 검정

반복 측정 모형과 같이 다변량 문제를 단변량 문제로 변환할 수 있는 경우가 있습니다 (Huynh and Feldt 1970). 다음과 같은 경우 다변량 컨텍스트에서 단변량 검정을 사용할 수 있습니다 .

- 반응 설계 행렬 **M** 이 직교정규 (**M**′**M** = 항등) 인 경우
- **M** 이 각 변환의 계수 합이 0 인 둘 이상의 반응을 생성하는 경우
- **구형성** 조건이 충족되는 경우 . 구형성 조건은 M- 변환 반응 간에 상관관계가 없고 분산이 동일하다는 의미입니다. **M**′**ΣM** 은 항등 행렬에 비례하며 , 여기서 **Σ** 는 *Y* 변수의 공분산입니다.

이러한 조건이 유지되면 **E** 및 **H** 검정 행렬의 주대각 요소가 합산되어 *F* 검정의 분모와 분자에 대한 단변량 제곱합을 생성합니다. 위의 조건이 유지되지 않으면 오류 메시지가 나타납니다. Golf Balls.jmp 의 경우 항등 행렬이 **M** 행렬로 지정됩니다. 열 벡터 중심화 및 직교정규화 후에는 항등 행렬을 완전 계수 행렬로 변환할 수 없습니다. 따라서 단변량 요청이 무시됩니다.

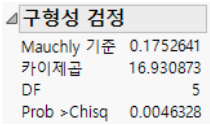
단변량 및 구형성 검정의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Dogs.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. LogHist0, LogHist1, LogHist3 및 LogHist5 를 선택하고 **Y** 를 클릭합니다 .
4. drug 와 dep1 을 선택하고 **추가**를 클릭합니다 .
5. " 모형 효과 생성 " 패널에서 drug 를 선택합니다 . " 열 선택 " 패널에서 dep1 을 선택합니다 . **교차**를 클릭합니다 .
6. " 분석법 " 에서 **MANOVA** 를 선택합니다 .
7. **실행**을 클릭합니다 .
8. **단변량 검정 추가** 옆의 체크박스를 선택합니다 .
9. **반응 선택** 메뉴에서 **반복 측정**을 선택합니다 .

   "Y 이름 " 에 " 시간 " 을 입력하고 **단변량 검정 추가**를 선택해야 합니다 .

10. **확인**을 클릭합니다 .

그림 9.8 구형성 검정



구형성 검정은 구형성 가정을 검정하기 위해 Mauchly 기준을 사용하여 개체 내 효과에 대한 비조정 단변량 *F* 검정의 적합성을 확인합니다(Anderson 1958). 구형성 검정과 단변량 검정은 항상 직교정규화된 **M** 행렬을 사용하여 수행됩니다 . 다음 지침을 사용하여 구형성 검정을 해석하십시오 .

- 실제 공분산 구조가 구형인 경우 비조정 단변량 *F* 검정을 사용할 수 있습니다 .
- 구형성 검정이 유의하면 실제 공분산 구조가 구형이 아님을 나타냅니다. 따라서 다변량 검정 또는 조정된 단변량 검정을 사용할 수 있습니다 .

단변량 *F* 통계량은 구형성이 없어도 근사 *F* 분포를 갖지만 분자와 분모의 자유도가 일부 부분 입실론(ε) 만큼 감소합니다. Box 연구 자료(1954), Greenhouse and Geisser 연구 자료(1959) 및 Huynh-Feldt 연구 자료(1976) 에서는 입실론 자유도 조정을 추정하는 기법을 제공합니다. Muller and Barton 연구 자료(1989)에서는 검정력 연구를 기반으로 Greenhouse-Geisser 버전을 권장합니다.

다변량 보고서의 입실론 조정 검정에는 G-G(Greenhouse-Geisser) 또는 H-F(Huynh-Feldt) 라벨이 지정됩니다. 입실론 조정은 **값** 열에 표시됩니다.

반복 측정을 사용하는 다변량 모형

다변량 적합의 일반적인 용도 중 하나는 **종단 데이터**라고도 하는 반복 측정을 사용한 데이터 분석입니다. 시간에 따라 반복해서 개체를 측정하고, 각 시간 측정값이 변수를 구성하도록 데이터가 배열됩니다. 이 상관관계가 **구형성**이라는 패턴을 만드는 경우 외에는 측정값 간의 상관관계 때문에 데이터를 단일 열에 쌓아 단변량 모형으로 분석하면 안 됩니다. 이 항목에 대한 자세한 내용은 이전에 나온 "단변량 검정 및 구형성 검정"(415 페이지) 섹션에서 확인하십시오.

반복 측정을 사용하면 분석이 다음과 같은 두 계층으로 구분됩니다 .

- 개체 사이 (또는 개체 전체) 효과는 반복 측정 열의 합을 모형 효과에 적합시켜 모델링됩니다 . 이 방법은 **합** 반응 함수를 사용하는 것과 같습니다 . 이 반응 함수는 1 로 구성된 단일 벡터인 **M** 행렬입니다 .
- 개체 내 효과 (반복 효과 또는 시간 효과) 는 반복 측정 열의 차이를 적합시키는 반응 함수로 모델링됩니다 . 이 분석은 **대비** 반응 함수 또는 기타 유사한 차분 함수 (**다항식** , **Helmert**, **프로파일** 또는 **평균**) 를 사용하여 수행할 수 있습니다 . 반복 측정 간의 차이를 모델링하는 경우 차이를 새로운 개체 내 효과 (대개 시간) 로 간주하십시오 . 모형의 효과를 적합시키는 경우 효과를 개체 내 효과와의 교호작용으로 해석할 수 있습니다 . 예를 들어 절편의 효과가 시간

(개체 내) 효과가 되어 반복 측정 간의 전체적인 차이를 보여 줍니다 . A 효과가 있는 경우 개체 내 검정은 A* 시간 교호작용에 대한 검정으로 해석되어 , 반복 측정 간의 차이가 A 효과에서 어떻게 달라지는지 모델링합니다 .

표 9.1(418 페이지) 에서는 반응 함수와 모형 효과 사이의 관계를 단변량 모형 지정을 사용할 때와 비교하여 보여 줍니다 . **합** (개체 사이) 모형과 **대비** (개체 내) 모형을 모두 사용하여 반응을 단일 열에 쌓고 표준 단변량 적합을 구하여 얻은 검정을 재생성할 수 있어야 합니다 .

다음과 같이 직접 / 간접 방법을 사용하여 반복 측정 분석을 수행할 수 있습니다 .

- 직접적인 방법은 팝업 메뉴의 " 반복 측정 " 항목을 사용하는 것입니다 . 이 항목을 선택하면 반복 측정에 대해 개체 내 효과를 나타내는 효과의 이름을 지정하도록 요청됩니다 . 그런 다음 **대비**와 **합** 반응 함수를 모두 적합시킵니다 . 이 방법을 사용하면 개체 내 효과 이름을 사용하여 효과에 적절한 라벨이 지정된다는 이점이 있습니다 .
- 간접적인 방법은 두 반응 함수를 개별적으로 지정하는 것입니다 . 먼저 **합** 반응 함수를 완료하고 두 번째로 **대비** 또는 차이를 모델링하는 다른 함수 중 하나를 완료합니다 . 대비 적합에서 개체 내 효과를 모형 효과와 연결해야 합니다 .

반복 측정 예

Cole and Grizzle(1966) 연구를 예로 들어보겠습니다 . 결과는 샘플 데이터 폴더의 Dogs.jmp 테이블에 있습니다 . 16 마리의 개가 각각 두 개의 수준을 갖는 drug 및 dep1 변수로 정의된 4 개 그룹에 할당됩니다 . 종속 변수는 약물을 주입하고 0, 1, 3 및 5 분 후 히스타민의 혈중 농도입니다 . 농도 로그는 데이터의 평균과 분산 간의 상관관계를 최소화하는 데 사용됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Dogs.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. LogHist0, LogHist1, LogHist3 및 LogHist5 를 선택하고 **Y** 를 클릭합니다 .
4. drug 와 dep1 을 선택한 후 **매크로** 메뉴에서 **완전 요인**을 선택합니다 .
5. " 분석법 " 에서 **MANOVA** 를 선택합니다 .
6. **실행**을 클릭합니다 .
7. **반응 선택** 메뉴에서 **반복 측정**을 선택합니다 .

   "Y 이름 " 에 " 시간 " 을 입력해야 합니다 . **단변량 검정 추가** 체크박스를 선택하면 반응이 단일 열에 쌓인 것처럼 계산되는 단변량 검정이 보고서에 포함됩니다 .
8. **확인**을 클릭합니다 .

그림 9.9 반복 측정 창

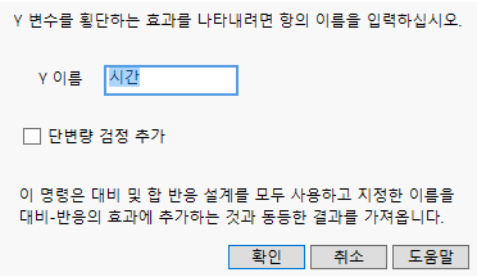


표 9.1에서는 LogHist0, LogHist1, LogHist3 및 LogHist5 열의 데이터가 단일 *Y* 열에 쌓인 경우, **합** 및 **대비** 반응 설계에 대한 다변량 검정이 단변량 검정의 라벨 지정 방법에 어떻게 대응하는지 보여 줍니다. 새 행은 명목형 그룹화 변수인 시간으로 식별됩니다.

표 9.1 대응하는 다변량 검정과 단변량 검정

합 M 행렬 개체 사이		대비 M 행렬 개체 내	
다변량 검정	단변량 검정	다변량 검정	단변량 검정
절편	절편	절편	시간
drug	drug	drug	시간*drug
depl	depl	depl	시간*depl

개체 사이 분석이 먼저 생성됩니다. 이 분석은 팝업 메뉴에서 **합**을 선택한 경우와 동일합니다(제목 제외).

다음으로 개체 내 분석이 생성됩니다. 이 분석은 팝업 메뉴에서 **대비**를 선택한 경우와 동일(제목 제외)하지만 보고서의 효과 이름에 개체 내 효과 이름(시간)이 추가되었습니다. 절편 항은 반복 측정값의 전체적인 차이를 추정하므로 이전의 **절편** 위치는 시간입니다.

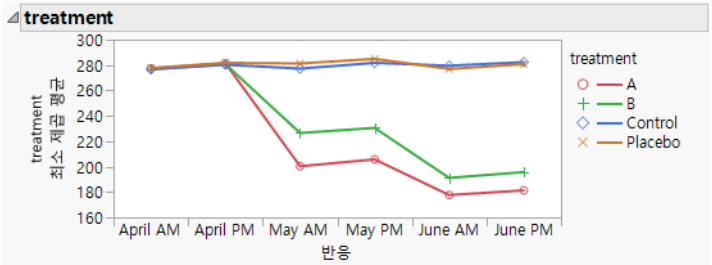
복합 다변량 모형의 예

JMP 는 반복 측정 계층으로 데이터를 처리할 수 있습니다 . Cholesterol.jmp 데이터 테이블을 예로 들어보겠습니다 . 5 명의 참가자로 구성된 그룹은 A, B, Control(대조), Placebo(플라시보) 라는 4 개의 treatment(치료) 그룹 중 하나에 속합니다 . 3 개월 동안 한 달에 한 번 아침에 콜레스테롤 수치를 측정하고 오후에 다시 측정했습니다 (가상 데이터). 이 예에서 반응 열은 월별 시간에 따라 순서대로 배열되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cholesterol.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .

3. April AM, April PM, May AM, May PM, June AM 및 June PM 을 선택하고 **Y** 를 클릭합니다 .
4. treatment 를 선택하고 **추가**를 클릭합니다 .
5. " 분석법 " 옆에서 **MANOVA** 를 선택합니다 .
6. **실행**을 클릭합니다 .

그림 9.10 treatment 그래프

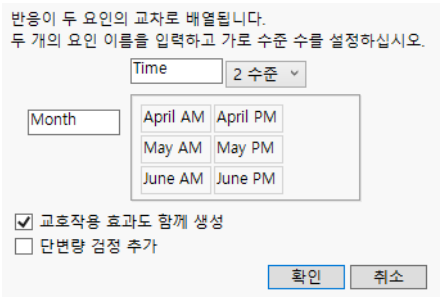


treatment 그래프에서 4 개의 치료 그룹이 매우 유사한 평균 콜레스테롤 값으로 연구를 시작했음을 알 수 있습니다 . A 및 B 치료 그룹은 실험 기간이 끝날 때 콜레스테롤 수치가 더 낮은 것으로 보입니다 . 대조 및 플라시보 그룹은 변경되지 않았습니다 .

7. **반응 선택** 메뉴를 클릭하고 **복합**을 선택합니다 .

   데이터 테이블에 반응이 배열되는 방식과 각 반응의 수준 수를 지정하여 이 창을 완료합니다 . 콜레스테롤 예에서 시간 열은 월 내에 배열됩니다. 따라서 시간을 하나의 요인으로 지정하고 월 효과를 다른 요인으로 지정합니다 . 교호작용 효과 검정은 선택 사항입니다 .
8. 그림 9.11 의 옵션을 사용하여 창을 완료합니다 .

그림 9.11 복합 창



9. **확인**을 클릭합니다 .

   각 효과에 대한 검정이 나타납니다 . 그림 9.12 에 보고서의 일부가 표시되어 있습니다 . 다음 사항에 유의하십시오 .

   – " 시간 " 보고서에는 시간과 treatment 간의 교호작용에 대한 p 값이 0.6038 로 표시되어 교호작용이 유의하지 않음을 나타냅니다 . 즉 , 오전 (AM) 과 오후 (PM) 사이에 treatment 차

이가 있다는 증거가 없음을 의미합니다 . 시간의 수준 (AM, PM) 이 두 개이므로 정확 F 검정이 나타납니다 .

– " 월 " 보고서에는 월과 treatment 간의 교호작용에 대한 p 값이 <.0001 로 표시되어 교호작용이 유의함을 나타냅니다 . 즉 , 치료 그룹 간의 차이가 월에 따라 달라진다는 것을 나타냅니다 . 그림 9.10 의 treatment 그래프를 보면 4 월 (April) 에는 그룹 간에 차이가 없지만 5 월 (May) 에는 치료 유형 (A, B, Control, Placebo) 간의 차이가 커지고 6 월 (June) 에는 훨씬 더 커집니다 .

– " 시간 * 월 " 보고서에는 treatment 에 대해 유의한 p 값이 표시되지 않습니다 . 즉 , 시간 , 월 , treatment 를 포함한 삼원 교호작용 효과가 통계적으로 유의하지 않음을 나타냅니다 .

그림 9.12 콜레스테롤 연구 결과

복합
Time
M 행렬
M-변환 모수 추정값
전체 모형
절편
treatment

검정	값	정확한 F	NumDF	DenDF	Prob>F
F-검정	0.1188721	0.6340	3	16	0.6038

복합
Month
M 행렬
M-변환 모수 추정값
전체 모형
절편
treatment

검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.013025	38.8109	6	30	<.0001*
Pillai 대각합	1.3128917	10.1907	6	32	<.0001*
Hotelling-Lawley	50.753209	123.4368	6	18.326	<.0001*
Roy 최대근	50.255302	268.0283	3	16	<.0001*

복합
Time*Month
M 행렬
M-변환 모수 추정값
전체 모형
절편
treatment

검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.6623742	1.1435	6	30	0.3619
Pillai 대각합	0.3582613	1.1638	6	32	0.3498
Hotelling-Lawley	0.4785668	1.1639	6	18.326	0.3671
Roy 최대근	0.4008469	2.1378	3	16	0.1355

판별 분석

판별 분석은 알려진 반응 값을 기반으로 일원 분류의 일부 수준을 예측하는 방법입니다. 이 기법은 측정 변수가 예측되는 수준의 다변량 평균에 얼마나 가까운지를 기반으로 합니다. 판별 분석은 판별 플랫폼을 사용하여 더 완전하게 구현됩니다. 자세한 내용은 Multivariate Methods의 에서 확인하십시오.

JMP에서 측정 변수를 *Y* 효과로 지정하고 분류 변수를 단일 *X* 효과로 지정합니다. 다변량 적합 플랫폼은 각 그룹에 대해 공분산이 동일하다는 가정하에 데이터의 평균 및 공분산 행렬 추정값을 제공합니다. MANOVA 플랫폼 이름 옆의 팝업 메뉴에서 **판별 저장** 옵션을 사용하여 판별 정보를 얻을 수 있습니다. 이 명령은 초기 **E** 및 **H** 행렬을 사용하여 거리와 확률을 현재 데이터 테이블의 열로 저장합니다.

수준이 *k* 개인 분류 변수의 경우 JMP는 *k* 개의 거리 열, *k* 개의 분류 확률 열, 예측 분류 열 및 두 개의 기타 계산 정보 열을 현재 데이터 테이블에 추가합니다.

판별 저장 옵션의 예

Mardia et al. 연구 자료 (1979) 에 나오는 Fisher 의 Iris 데이터를 검토합니다 . 각 표본에 대해 *k* = 3 개 수준의 종 (species) 과 4 개의 측정값이 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Sepal length, Sepal width, Petal length 및 Petal width 를 선택하고 **Y** 를 클릭합니다 .
4. Species 를 선택하고 **추가**를 클릭합니다 .
5. " 분석법 " 옆에서 **MANOVA** 를 선택합니다 .
6. **실행**을 클릭합니다 .
7. "MANOVA 적합 " 의 빨간색 삼각형을 클릭하고 **판별 저장**을 선택합니다 .

다음 열이 Iris.jmp 샘플 데이터 테이블에 추가됩니다 .

SqDist0 Mahalanobis 거리 계산에 필요한 2 차 형식입니다 .

SqDist[setosa] setosa 중심에서 관측값까지의 Mahalanobis 거리입니다 .

SqDist[versicolor] versicolor 중심에서 관측값까지의 Mahalanobis 거리입니다 .

SqDist[virginica] virginica 중심에서 관측값까지의 Mahalanobis 거리입니다 .

Prob0 아래에 사용된 Mahalanobis 거리의 음의 지수 합입니다 .

Prob[setosa] setosa 범주에 속할 확률입니다 .

Prob[versicolor] versicolor 범주에 속할 확률입니다 .

Prob[virginica] virginica 범주에 속할 확률입니다 .

Pred Species 확률에서 가장 가능성이 높은 종 (Species) 입니다 .

이제 다른 JMP 플랫폼에서 데이터 테이블의 새 열을 사용하여 보고서와 그래프로 판별 분석을 요약할 수 있습니다 . 예를 들면 다음과 같습니다 .

1. 새 열을 포함하는 업데이트된 Iris.jmp 데이터 테이블에서 **분석 > X 로 Y 적합**을 선택합니다 .
2. Species 를 선택하고 **Y, 반응**을 클릭합니다 .
3. Pred Species 를 선택하고 **X, 요인**을 클릭합니다 .
4. **확인**을 클릭합니다 .

"분할표"에 판별 분류가 요약되어 있습니다. 세 개의 오분류가 확인되었습니다.

그림 9.13 예측 및 실제 Species 의 분할표

분할표

Pred Species	개수 총 % 열 % 행 %	setosa	versicolor	virginica	합계
	setosa	50	0	0	50
		33.33	0.00	0.00	33.33
		100.00	0.00	0.00	
		100.00	0.00	0.00	
	versicolor	0	48	1	49
		0.00	32.00	0.67	32.67
		0.00	96.00	2.00	
		0.00	97.96	2.04	
	virginica	0	2	49	51
		0.00	1.33	32.67	34.00
		0.00	4.00	98.00	
		0.00	3.92	96.08	
	합계	50	50	50	150
		33.33	33.33	33.33	

MANOVA 분석법에 대한 통계 상세 정보

- " 다변량 검정 "
- " 근사 F- 검정 "
- " 정준 상세 정보 "

다변량 검정

아래에서 **E**는 잔차 교차곱 행렬이고, **H**는 모형 교차곱 행렬입니다. **E**의 주대각 요소는 각 변수의 잔차 제곱합입니다. **H**의 주대각 요소는 각 변수에 대한 모형의 제곱합입니다. 판별 분석 관련 문헌에서는 종종 **E**를 **W**라고 합니다(**W**는 within 을 나타냄).

다변량 결과 표의 검정 통계량은 $\mathbf{E}^{-1}\mathbf{H}$ 의 고유값 λ를 사용해서 구합니다. 다음 목록에서는 각 검정 통계량의 계산 방법을 설명합니다.

참고 : 반응 설계가 지정된 후 초기 **E** 및 **H** 행렬의 앞에 $\mathbf{M}'$ 를 곱하고 뒤에 **M** 을 곱합니다 .

- Wilks 람다

$$\Lambda = \frac{\det(\mathbf{E})}{\det(\mathbf{H}+\mathbf{E})} = \prod_{i=1}^{n}\left(\frac{1}{1+\lambda_i}\right)$$

- Pillai 대각합

$$V = \text{Trace}[\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}] = \sum_{i=1}^{n}\frac{\lambda_i}{1+\lambda_i}$$

- Hotelling-Lawley 대각합

$$U = \text{Trace}(\mathbf{E}^{-1}\mathbf{H}) = \sum_{i=1}^{n}\lambda_i$$

- Roy 최대근

$\Theta = \lambda_1$, $\mathbf{E}^{-1}\mathbf{H}$ 의 최대 고유값

E 와 **H** 는 다음과 같이 정의됩니다.

$$\mathbf{E} = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'(\mathbf{X}'\mathbf{X})\mathbf{b}$$

$$\mathbf{H} = (\mathbf{L}\mathbf{b})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-1}(\mathbf{L}\mathbf{b})$$

여기서 **b** 는 모형 계수의 추정 벡터이며, **A** 는 행렬 **A** 의 일반화 역행렬을 나타냅니다.

전체 모형 **L** 은 0(절편 값)으로 구성된 열로, 행 및 열 개수가 모형의 모수 개수와 같은 단위 행렬과 연결되어 있습니다. 효과에 대한 **L** 행렬은 전체 모형 **L** 행렬의 일부 행으로 구성한 부분집합입니다.

근사 F- 검정

F 값과 자유도를 계산하려면 p 의 순위가 $\mathbf{H}+\mathbf{E}$ 가 되도록 하고 q 의 순위는 $\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'$ 가 되도록 합니다. 여기서 **L** 행렬은 검정 중인 효과와 관련된 **X'X** 의 요소를 식별합니다. v 는 오차 자유도가 되도록 하고 s 는 p 및 q 의 최소값이 되도록 합니다. 또한 $m = 0.5(|p-q|-1)$ 및 $n = 0.5(v-p-1)$ 이 되도록 합니다.

표 9.2(423 페이지)에서는 해당 검정 통계량에서 구한 각 근사 F 의 계산을 보여 줍니다.

표 9.2 근사 *F* 통계량

검정	근사 F	분자 DF	분모 DF
Wilks 람다	$F = \left(\frac{1-\Lambda^{1/t}}{\Lambda^{1/t}}\right)\left(\frac{rt-2u}{pq}\right)$	pq	$rt-2u$
Pillai 대각합	$F = \left(\frac{V}{s-V}\right)\left(\frac{2n+s+1}{2m+s+1}\right)$	$s(2m+s+1)$	$s(2n+s+1)$
Hotelling-Lawley 대각합	$F = \frac{2(sn+1)U}{s^2(2m+s+1)}$	$s(2m+s+1)$	$2(sn+1)$
Roy 최대근	$F = \frac{\Theta(v-\max(p,q)+q)}{\max(p,q)}$	$\max(p,q)$	$v-\max(p,q)+q$

정준 상세 정보

검정 상세 정보 옵션에 대한 상세 정보

지정된 검정에 대해 "검정 상세 정보" 옵션을 선택하면 고유값, 정준 상관 및 고유 벡터가 보고서에 표시됩니다.

"검정 상세 정보" 옵션에 의해 생성된 정준 상관은 다음과 같이 계산됩니다.

$$\rho_i = \sqrt{\frac{\lambda_i}{1+\lambda_i}}$$

여기서 λ_i 는 다변량 검정 통계량을 계산하는 데 사용되는 $\mathbf{E}^{-1}\mathbf{H}$ 행렬의 i 번째 고유값입니다.

고유 벡터 라벨이 표시된 행렬은 지정된 검정에 대한 $\mathbf{E}^{-1}\mathbf{H}$의 고유 벡터 행렬인 **V** 행렬입니다.

참고 : 지정된 검정의 **E** 및 **H** 행렬은 원래 **E** 및 **H** 행렬에 대해 **M′EM** 과 **M′HM** 을 나타냅니다. **M** 행렬은 반응 설계에 의해 정의됩니다. 이 섹션에 사용된 **E** 및 **H** 는 "다변량 검정"(422 페이지) 에 정의되어 있습니다.

중심 그림 옵션에 대한 상세 정보

총 표본 중심 및 효과의 중심값은 다음과 같이 계산됩니다.

$$\text{총} = (c'_1\bar{y}, c'_2\bar{y}, \ldots, c'_g\bar{y})$$

$$\text{효과}_j = (c'_1\bar{x}_j, c'_2\bar{x}_j, \ldots, c'_g\bar{x}_j)$$

다음은 각 요소에 대한 설명입니다 .

$$c_i = \left(v'_i\left(\frac{E}{N-r}\right)v_i\right)^{-1/2} v_i$$

N - 관측값 수

v_i - 지정된 검정에 대한 $\mathbf{E}^{-1}\mathbf{H}$ 의 고유 벡터 행렬인 $\mathbf{V}$ 의 i 번째 열

$\bar{x}_j$ - j 번째 효과의 다변량 최소 제곱 평균

$\bar{y}$ - 반응의 전체 평균

g - 0 보다 큰 $\mathbf{E}^{-1}\mathbf{H}$ 의 고유값 수

r - $\mathbf{X}$ 행렬의 계수

참고 : 지정된 검정의 **E** 및 **H** 행렬은 원래 **E** 및 **H** 행렬에 대해 **M'EM** 과 **M'HM** 을 나타냅니다 . **M** 행렬은 반응 설계에 의해 정의됩니다 . 이 섹션에 사용된 **E** 및 **H** 는 " 다변량 검정 "(422 페이지) 에 정의되어 있습니다 .

효과에 대한 중심 반지름은 다음과 같이 계산됩니다 .

$$d = \sqrt{\frac{\chi^2_g(0.95)}{L(X'X)^{-1}L'}}$$

여기서 g 는 0 보다 큰 $\mathbf{E}^{-1}\mathbf{H}$ 의 고유값 수이고 , 분모의 **L** 행렬은 다변량 최소 제곱 평균 계산에서 가져옵니다 .

정준 점수 저장 옵션에 대한 상세 정보

정준 Y 값은 다음과 같이 계산됩니다 .

$$\tilde{Y} = \mathbf{YM'V}$$

다음은 각 요소에 대한 설명입니다 .

Y - 반응 변수의 행렬

M' - 반응 설계 행렬의 전치

V - 지정된 검정에 대한 $\mathbf{E}^{-1}\mathbf{H}$ 의 고유 벡터 행렬

참고 : 지정된 검정의 **E** 및 **H** 행렬은 원래 **E** 및 **H** 행렬에 대해 **M'EM** 과 **M'HM** 을 나타냅니다 . **M** 행렬은 반응 설계에 의해 정의됩니다 . 이 섹션에 사용된 **E** 및 **H** 는 " 다변량 검정 "(422 페이지) 에 정의되어 있습니다 .

0보다 큰 고유값에 해당하는 고유 벡터에 대해 정준 Y 값이 저장됩니다.

10 장

로그 선형 분산 모형
반응의 분산 및 평균 모델링

모형 적합 플랫폼의 로그 선형 분산 분석법을 사용하면 회귀 모형으로 반응의 기대값과 분산을 모두 모델링할 수 있습니다. 분산의 로그는 하나의 선형 모형에 적합되고 기대 반응은 여러 선형 모형에 동시에 적합됩니다.

참고 : 값을 추정하려면 잘 설계된 적합 데이터가 많이 있어야 합니다 . 평균보다 분산을 적합시키기 위한 데이터가 더 많이 필요합니다 .

많은 엔지니어의 실험 목표는 반응 자체를 최대화 또는 최소화하는 것이 아니라 목표 반응에 중점을 두고 변동성을 최소화하는 것입니다. 로그 선형 분산 모형은 분산을 모델링하는 매우 일반적이고 효과적인 방법을 제공하며 반복되지 않은 데이터와 반복이 있는 데이터에 모두 사용할 수 있습니다.

목차

로그 선형 분산 모형 개요 429
산포 효과 429
모형 규격 429
참고 430
로그 선형 분산 분석법 시작 430
예 : 로그 선형 분산 사용 431
로그 선형 보고서 433
로그 선형 플랫폼 옵션 434
열 저장 435
행 진단 435
잔차 검토 435
적합 모형 프로파일링 436
적합 모형 프로파일링의 예 436

로그 선형 분산 모형 개요

로그 선형 분산 모형은 선형 모형을 통해 분산을 모델링하는 방법을 제공합니다 . 자세한 내용은 Harvey 연구 자료 (1976), Cook and Weisberg 연구 자료 (1983), Aitken 연구 자료 (1987) 및 Carroll and Ruppert 연구 자료 (1988) 에서 확인하십시오 . 선형 모형에는 다음과 같이 평균 반응을 모델링하기 위한 회귀변수 항 외에도 분산의 로그를 모델링하기 위한 회귀변수 항이 있습니다 .

평균 모형 : $E(y) = \mathbf{X}\beta$

분산 모형 : $\log(\text{Variance}(y)) = \mathbf{Z}\,\lambda$,

또는 다음과 같이 지정될 수도 있습니다 .

$\text{Variance}(y) = \exp(\mathbf{Z}\,\lambda)$

여기서 X 의 열은 반응 평균에 대한 회귀변수이고 Z 의 열은 반응 분산에 대한 회귀변수입니다. 정규 선형 모형 모수는 β 로 표시되고 λ 는 분산 모형의 모수를 나타냅니다.

로그 선형 분산 모형은 REML 을 사용하여 추정됩니다.

산포 또는 **로그 분산** 효과는 반응 분산의 변화를 모델링할 수 있습니다. 이는 모형 적합 플랫폼에서 로그 선형 분산이라는 적합 분석법에 의해 구현됩니다.

산포 효과

산포 효과 모델링은 Taguchi 프레임워크를 제외하고는 교재에서 아주 자세하게 다루지 않습니다. Taguchi 스타일 실험에서는 외측 배열의 전체 설정에 대해 여러 측정을 수행하고, 이 외측 배열에서 목표를 벗어난 변동성을 측정하는 새 반응을 생성한 후 모형을 적합시켜 변동성을 최소화하는 요인을 찾는 방식으로 처리됩니다. 이 유형의 모델링에는 두 설계의 완전한 Cartesian 곱인 특수 설계가 필요합니다. 이 장에서 설명하는 방법은 더 유연한 모형 기반 방식으로 분산을 모델링합니다. Taguchi 가 변동성 모델링을 위해 권장하는 특정 성능 통계량은 $STD = -\log(s)$ 입니다. JMP 에 사용되는 방법론에서는 $\log(s^2)$ 이 모델링된 후 평균이 있는 모형과 결합됩니다. $\log(s^2)=2\log(s)$ 이므로 두 값은 기본적으로 동등합니다.

모형 규격

로그 선형 분산 효과는 "모형 적합" 대화상자에서 효과를 강조 표시한 후 **속성** 드롭다운 메뉴에서 **로그 분산 효과**를 선택하여 지정됩니다. 그러면 해당 효과의 끝에 **& 로그 분산**이 표시됩니다. 이 속성을 사용하면 맨 위의 적합 **분석법**도 **로그 선형 분산**으로 변경됩니다. 반응의 평균과 분산에 모두 효과를 사용하려면 효과를 두 번 지정해야 하며 한 번은 **로그 분산** 옵션을 사용합니다.

로그 분산 속성을 사용하여 지정하는 효과는 모형에서 *Zs* 를 생성하는 효과가 되고, 다른 효과는 모형에서 *Xs* 가 됩니다.

참고

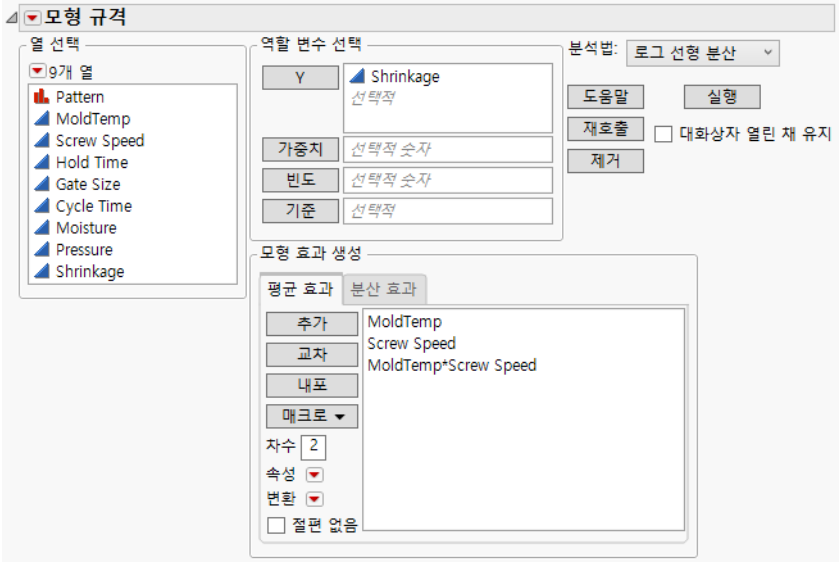
평균 모형에 대해 다른 모수가 추정될 때마다 하나 이상의 관측값이 필요하며 더 많으면 좋습니다. 그러나 분산 모수의 경우 적절한 추정값을 얻으려면 각 분산 모수에 대해 몇몇 관측값이 더 필요합니다. 평균보다 분산을 추정하는 데 데이터가 더 많이 필요합니다.

로그 선형 분산 모형은 산포 효과를 적합시키는 매우 유연한 방법이며, 지금까지 문헌에서 받은 것보다 훨씬 더 많은 주의를 기울여야 합니다.

로그 선형 분산 분석법 시작

분석 > 모형 적합을 선택하고 **Y** 에 대해 하나 이상의 열을 입력한 후 **분석법** 메뉴에서 **로그 선형 분산**을 선택하여 로그 선형 분산 분석법을 시작합니다.

그림 10.1 로그 선형 분산이 선택된 모형 적합 시작 창



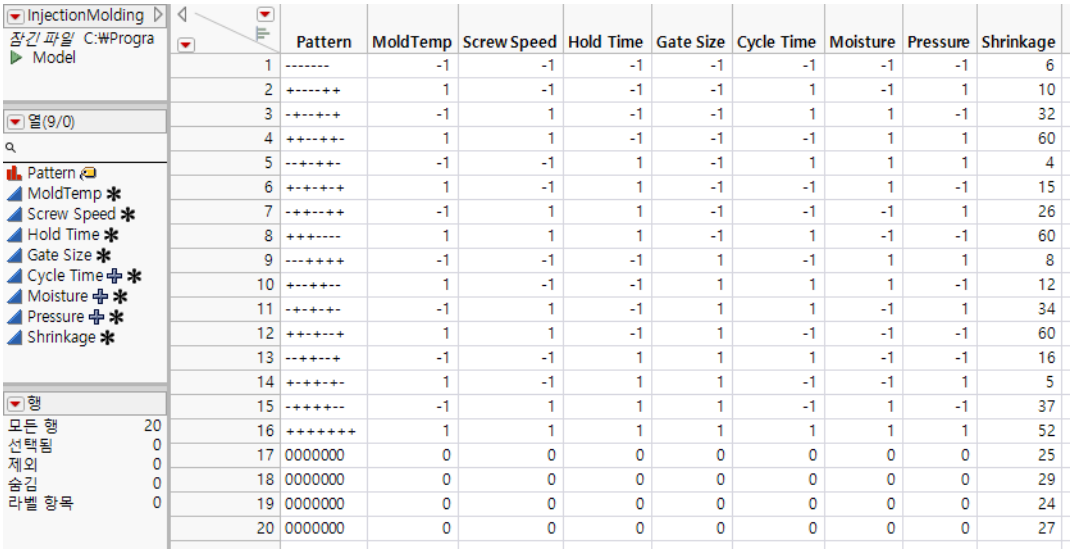
"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 "모형 규격" 장(29 페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 로그 선형 분산 분석법에 대한 정보를 제공합니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격"의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

예 : 로그 선형 분산 사용

InjectionMolding.jmp 데이터 테이블에는 네 개의 중앙점이 추가된 7 요인 2^{7-3} 부분 요인 설계의 실험 결과가 포함되어 있습니다(Montgomery 1991). 예비 조사에 따르면 평균 반응은 처음 두 요인, 즉 Mold Temperature 와 Screw Speed 에 따라서만 달라지고, 분산은 Holding Time 의 영향을 받는 것 같다고 확인되었습니다.

그림 10.2 사출 성형 데이터



	Pattern	MoldTemp	Screw Speed	Hold Time	Gate Size	Cycle Time	Moisture	Pressure	Shrinkage
1	-------	-1	-1	-1	-1	-1	-1	-1	6
2	+-----++	1	-1	-1	-1	1	-1	1	10
3	-+--+-+	-1	1	-1	-1	1	1	-1	32
4	++--++-	1	1	-1	-1	-1	1	1	60
5	--+-++-	-1	-1	1	-1	1	1	1	4
6	+-+-+-+	1	-1	1	-1	-1	1	-1	15
7	-++--++	-1	1	1	-1	-1	-1	1	26
8	+++----	1	1	1	-1	1	-1	-1	60
9	---++++	-1	-1	-1	1	-1	1	1	8
10	+--+--	1	-1	-1	1	1	1	-1	12
11	-+-+-+-	-1	1	-1	1	1	-1	1	34
12	++-+--+	1	1	-1	1	-1	-1	-1	60
13	--++--+	-1	-1	1	1	1	-1	-1	16
14	+-++-+-	1	-1	1	1	-1	-1	1	5
15	-++++--	-1	1	1	1	-1	1	-1	37
16	+++++++	1	1	1	1	1	1	1	52
17	0000000	0	0	0	0	0	0	0	25
18	0000000	0	0	0	0	0	0	0	29
19	0000000	0	0	0	0	0	0	0	24
20	0000000	0	0	0	0	0	0	0	27

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 InjectionMolding.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .

   데이터 테이블의 변수에 사전 선택된 역할이 할당되어 있으므로 분석이 자동으로 실행됩니다 .

그림 10.3 로그 선형 분산 보고서 창

로그선형 분산 적합

적합 통계량

-2*잔차 로그 가능도	90.712809
-2*로그 가능도	87.777035
AICc	106.23857
BIC	105.75143

Shrinkage에 대한 평균 모형

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	27.582516	0.380697	72.45	<.0001*
MoldTemp	7.683713	0.392907	19.56	<.0001*
Screw Speed	18.673515	0.392907	47.53	<.0001*
MoldTemp*Screw Speed	5.765297	0.392907	14.67	<.0001*

고정 효과 검정

소스	N모수	DF	DFDen	F 비	Prob > F
MoldTemp	1	1	5.758	382.4387	<.0001*
Screw Speed	1	1	5.758	2258.768	<.0001*
MoldTemp*Screw Speed	1	1	5.758	215.3094	<.0001*

Shrinkage에 대한 분산 모형

등분산에 대한 가능도비 검정

소스	-2*잔차 로그 가능도
초기에 등분산	101.76083063
적합 분산 이후	90.712808548
차이: 카이제곱	11.048022086
자유도	1
Prob > ChiSquare	0.0008878186

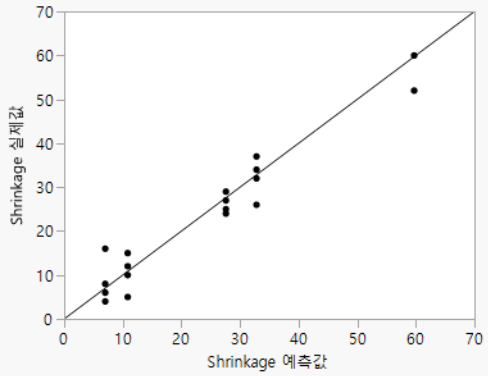
분산 모수 추정값

항	추정값	표준 오차	카이제곱	Prob>ChiSq	95% 하한	95% 상한
Hold Time	1.5788108	0.412603	11.0480	0.0009*	0.7057476	2.37325
잔차	6.6912718	2.445524	7.4864	0.0062*	3.2689762	13.696373

분산 효과 가능도비 검정

소스	검정 유형	DF	카이제곱	Prob>ChiSq
Hold Time	가능도	1	11.0480	0.0009*

실제값 대 예측값 그림



"Shrinkage에 대한 평균 모형" 보고서는 평균 모형의 모수를 제공하고 "Shrinkage에 대한 분산 모형" 보고서는 분산 모형의 모수를 제공합니다.

로그 선형 보고서

결과 보고서의 위쪽 부분에서는 기대 반응의 적합을 보여 줍니다. "적합 통계량 보고서 설명"(343 페이지)에 적합 통계량 계산 방법이 설명되어 있습니다. "모수 추정값" 및 "고정 효과 검정"은 REML(제한 최대 가능도)에서 도출되는 점을 제외하고 표준 최소 제곱 분석법의 보고서와 비슷합니다.

그림 10.4 평균 모형 출력

로그선형 분산 적합

적합 통계량

-2*잔차 로그 가능도	90.712809
-2*로그 가능도	87.777035
AICc	106.23857
BIC	105.75143

Shrinkage에 대한 평균 모형

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	27.582516	0.380697	72.45	<.0001*
MoldTemp	7.683713	0.392907	19.56	<.0001*
Screw Speed	18.673515	0.392907	47.53	<.0001*
MoldTemp*Screw Speed	5.765297	0.392907	14.67	<.0001*

고정 효과 검정

소스	N모수	DF	DFDen	F 비	Prob > F
MoldTemp	1	1	5.758	382.4387	<.0001*
Screw Speed	1	1	5.758	2258.768	<.0001*
MoldTemp*Screw Speed	1	1	5.758	215.3094	<.0001*

그림 10.5 분산 모형 출력

Shrinkage에 대한 분산 모형

등분산에 대한 가능도비 검정

소스	-2*잔차 로그 가능도
초기에 등분산	101.76083063
적합 분산 이후	90.712808548
차이: 카이제곱	11.048022086
자유도	1
Prob > ChiSquare	0.0008878186

분산 모수 추정값

항	추정값	표준 오차	카이제곱	Prob>ChiSq	95% 하한	95% 상한
Hold Time	1.5788108	0.412603	11.0480	0.0009*	0.7057476	2.37325
잔차	6.6912718	2.445524	7.4864	0.0062*	3.2689762	13.696373

분산 효과 가능도비 검정

소스	검정 유형	DF	카이제곱	Prob>ChiSq
Hold Time	가능도	1	11.0480	0.0009*

보고서의 두 번째 부분에서는 분산 모형의 적합을 보여 줍니다. **분산 모수 추정값** 보고서에 추정값 및 관련 통계량이 표시됩니다. 다음과 같은 두 개의 숨겨진 열이 제공됩니다.

- 숨겨진 열 **exp(추정값)** 은 추정값의 지수입니다. 따라서 요인이 +1 및 -1 값을 갖도록 코드화되는 경우 요인의 +1 수준은 분산에 **exp(추정값)** 값을 곱합니다. 마찬가지로 -1 수준은 분산

에 이 열의 역수를 곱합니다. 숨겨진 열을 표시하려면 보고서를 마우스 오른쪽 버튼으로 클릭한 후 표시된 **열** 메뉴에서 해당 열의 이름을 선택합니다.

- **exp(2l 추정값 l)** 라벨이 지정된 숨겨진 열은 회귀변수의 범위가 -1 ~ +1 일 때 하위 분산 대비 상위 분산의 비율입니다.

보고서에는 각 추정값의 표준 오차, 카이제곱, p 값 및 프로파일 가능도 신뢰 한계도 표시됩니다. 잔차 모수는 다른 모든 회귀변수가 0 일 때 분산의 전체 추정값입니다.

분산 모형의 적합도가 원래 모형보다 유의하게 나은지 확인해 봅니다. 이를 위한 가능도비 검정은 적합 모형을 절편을 제외한 모든 모수가 0 인 모형, 즉 등분산 모형과 비교합니다. 이 경우 p 값은 매우 유의합니다. Hold Time 변경에 따라 분산이 변경됩니다.

분산 효과 가능도비 검정은 각 항을 차례로 제외하고 모형을 다시 적합시켜 가능도비 검정을 생성합니다. 일반적으로 이러한 검정을 Wald 검정보다 더 신뢰할 수 있습니다.

로그 선형 플랫폼 옵션

" 로그선형 분산 적합 " 메뉴에는 다음 옵션이 포함되어 있습니다.

열 저장 데이터 테이블에 하나 이상의 열을 생성합니다. 자세한 내용은 " 열 저장 "(435 페이지) 에서 확인하십시오.

행 진단 행 진단 그림을 표시합니다. 자세한 내용은 " 행 진단 "(435 페이지) 에서 확인하십시오.

프로파일러 프로파일러, 등고선 프로파일러 또는 표면 프로파일러를 엽니다. 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 요인 프로파일링 "(150 페이지) 에서 확인하십시오.

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

열 저장

" 열 저장 " 에서 다음 옵션을 사용할 수 있습니다 .

예측 계산식 평균이라는 새 열을 생성합니다 . 새 열에는 지정된 모형에 의해 계산된 평균의 예측값이 포함됩니다 .

분산 계산식 분산이라는 새 열을 생성합니다 . 새 열에는 지정된 모형에 의해 계산된 분산의 예측값이 포함됩니다 .

표준편차 계산식 표준편차라는 새 열을 생성합니다 . 새 열에는 지정된 모형에 의해 계산된 표준편차의 예측값이 포함됩니다 .

잔차 잔차라는 새 열을 생성합니다 . 이 열에는 반응 관측값에서 예측값을 뺀 잔차가 포함됩니다 . 자세한 내용은 " 잔차 검토 "(435 페이지) 에서 확인하십시오 .

스튜던트화 잔차 스튜던트화 잔차라는 새 열을 생성합니다 . 새 열 값은 잔차를 표준 오차로 나눈 값입니다 .

예측값의 표준 오차 예측값의 표준 오차라는 새 열을 생성합니다 . 새 열에는 예측값의 표준 오차가 포함됩니다 .

개별값의 표준 오차 개별값의 표준 오차라는 새 열을 생성합니다 . 새 열에는 개별 예측값의 표준 오차가 포함됩니다 .

평균 신뢰 구간 평균의 95% 하한과 평균의 95% 상한이라는 두 개의 새 열을 생성합니다 . 새 열에는 예측 평균에 대한 신뢰 구간의 한계가 포함됩니다 .

개별값 신뢰 구간 개별값의 95% 하한과 개별값의 95% 상한이라는 두 개의 새 열을 생성합니다 . 새 열에는 개별 반응 값에 대한 신뢰 한계가 포함됩니다 .

행 진단

" 행 진단 " 에서 다음 옵션을 사용할 수 있습니다 .

실제값 대 예측값 그림 *Y* 에 대한 관측값 대 예측값 그림을 표시합니다 . 이 그림은 전체 모형에 대한 레버리지 그림입니다 .

스튜던트화 잔차 대 예측값 그림 *Y* 에 대한 스튜던트화 잔차 대 예측값 그림을 표시합니다 .

스튜던트화 잔차 대 행 그림 스튜던트화 잔차 대 행 그림을 표시합니다 .

잔차 검토

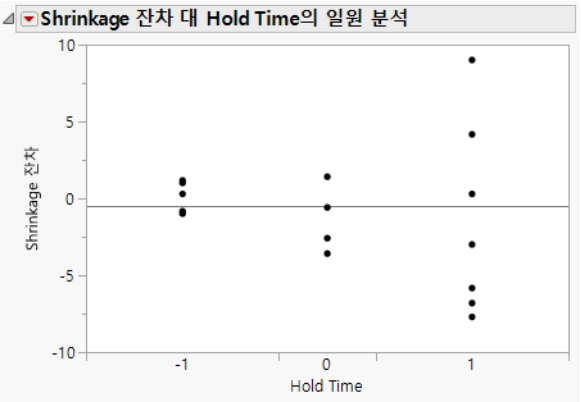
산포 효과를 확인하려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 InjectionMolding.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .

   데이터 테이블의 변수에 사전 선택된 역할이 할당되어 있으므로 분석이 자동으로 실행됩니다 .

3. " 로그선형 분산 적합 " 의 빨간색 삼각형을 클릭하고 **열 저장 > 잔차**를 선택합니다 .
4. InjectionMolding.jmp 샘플 데이터 테이블의 " 열 " 패널에서 Hold Time 옆에 있는 연속형 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **명목형**을 선택합니다 .
5. **분석 > X 로 Y 적합**을 선택합니다 .
6. Shrinkage 잔차를 선택하고 **Y, 반응**을 클릭합니다 .
7. Hold Time 을 선택하고 **X, 요인**을 클릭합니다 .
8. **확인**을 클릭합니다 .

그림 10.6 잔차 대 산포 효과



이 그림에서 Hold Time이 증가할수록 분산이 위로 올라가는 것을 쉽게 확인할 수 있습니다. 이는 Hold Time을 명목형 요인으로 처리하여 수행됩니다.

적합 모형 프로파일링

프로파일러, **등고선 프로파일러**, 또는 **표면 프로파일러**를 사용하면 적합 모형에 대한 더 많은 정보를 얻을 수 있습니다. 프로파일러 옵션을 선택하려면 "로그선형 분산 적합"의 빨간색 삼각형을 클릭한 후 **프로파일러** 메뉴에 있는 옵션 중 하나를 선택합니다.

적합 모형 프로파일링의 예

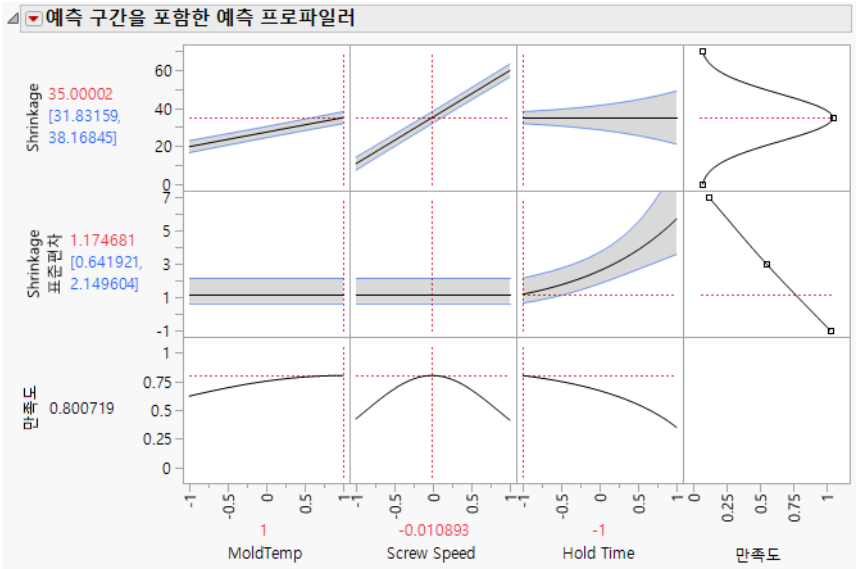
예를 들어 가장 작은 분산에서 반응에 대한 목표 (35) 를 달성하는 요인 설정을 찾는 것이 목적이라고 가정해 보겠습니다 . 모형을 적합시키고 보고서 메뉴에서 " 프로파일러 " 를 선택합니다 . 예를 들어 그림 10.7 에서는 평균의 목표값을 일치시키고 분산을 최소화하도록 설정된 프로파일러를 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 InjectionMolding.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .

데이터 테이블의 변수에 사전 선택된 역할이 할당되어 있으므로 분석이 자동으로 실행됩니다.

3. "로그선형 분산 적합"의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다.
4. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 설정**을 선택합니다.
5. "반응 목표" 창이 나타나면 "최대화"를 "목표값 일치"로 변경합니다.
6. **확인**을 클릭합니다.
7. 두 번째 "반응 목표" 창에서 **확인**을 클릭합니다.
8. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다.
9. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **예측 구간**을 선택합니다.

그림 10.7 예측 구간을 포함한 목표값 일치 및 분산 최소화 프로파일러



평균과 분산(둘 다 로그 분산 분석법으로 모델링됨)의 관계를 확인하는 가장 좋은 방법 중 하나는 평균에 대한 개별 예측 신뢰 구간을 검토하는 것입니다. 정규 신뢰 구간(프로파일러에 기본적으로 표시된 구간)은 개별 예측 신뢰 구간만큼 분산 모형에 대한 정보를 충분히 나타내지 않습니다. 예측 구간은 평균 및 분산 모형을 모두 한 그래프에 표시합니다.

Y 가 모델링된 반응이고 $x_{n,}$ 에서 새 관측값에 대한 예측 구간이 필요한 경우:

$$s^2\left|x_n = s^2_Y\right|x_n + s^2_{\hat{Y}}\left|x_n\right.$$

다음은 각 요소에 대한 설명입니다 .

$s^2\big|_{x_n}$ - x_n 에서 개별 예측에 대한 분산

$s^2_Y\big|_{x_n}$ - x_n 에서 Y 분포의 분산

$s^2_{\hat{Y}}\big|_{x_n}$ - $\hat{Y}$ 의 표본 분포 분산이며 평균의 분산

개별 예측의 분산은 Y 분포의 분산을 포함하므로 Y에 대한 분산이 변화할 때의 효과를 확인할 수 있습니다. 개별 예측 구간은 더 넓을 뿐만 아니라 분산 효과의 변화에 따라 형태를 변경할 수도 있습니다.

11 장

로지스틱 회귀 모형

명목형 또는 순서형 반응에 대한 회귀 모형 적합

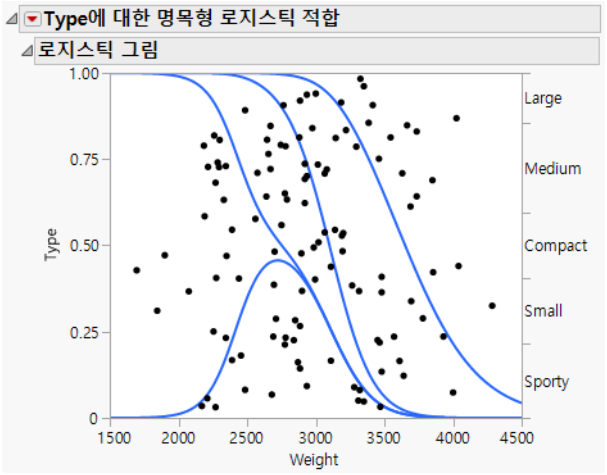
반응 변수에 이산형 값이 있는 경우 모형 적합 플랫폼을 사용하여 로지스틱 회귀 모형을 적합시킬 수 있습니다. 모형 적합 플랫폼은 로지스틱 회귀 모형 적합을 위한 두 가지 분석법을 제공합니다. 사용하는 분석법은 반응 열의 모델링 유형(명목형 또는 순서형)에 따라 달라집니다.

명목형 반응 변수의 경우 명목형 로지스틱 분석법으로 선형 모형을 다중 수준 로지스틱 반응 함수에 적합시킵니다.

순서형 반응 변수의 경우 순서형 로지스틱 분석법으로 누적 반응 확률을 선형 모형의 로지스틱 분포 함수에 적합시킵니다.

두 분석법 모두 모형의 가능도비 검정, 혼동 행렬, ROC 곡선 및 Lift 곡선을 제공합니다. 이항 반응의 경우 명목형 로지스틱 분석법은 승산비(해당 신뢰 구간 포함)도 제공합니다.

그림 11.1 명목형 로지스틱 회귀 모형에 대한 로지스틱 그림



목차

명목형 및 순서형 로지스틱 분석법 개요 441
명목형 로지스틱 회귀 441
순서형 로지스틱 회귀 441
로지스틱 회귀 모형을 적합시키는 기타 JMP 플랫폼 442
로지스틱 회귀의 예 442
명목형 로지스틱 회귀의 예 442
순서형 로지스틱 회귀의 예 444
명목형 및 순서형 로지스틱 분석법 시작 446
로지스틱 회귀 모형의 검증 447
로지스틱 적합 보고서 448
전체 모형 검정 449
적합 상세 정보 450
적합 결여 검정 450
로지스틱 적합 플랫폼 옵션 451
명목형 및 순서형 적합 옵션 451
명목형 적합 옵션 453
순서형 적합 옵션 454
로지스틱 회귀의 추가 예 455
역추정 예측의 예 455
명목형 로지스틱 모형에 대한 효과 요약의 사용 예 457
2차 순서형 로지스틱 모형의 예 459
예: 여러 열의 개수 쌓기 462
명목형 및 순서형 로지스틱 분석법에 대한 통계 상세 정보 463
로지스틱 회귀 모형 463
승산비 464
통계적 검정의 관계 465

명목형 및 순서형 로지스틱 분석법 개요

로지스틱 회귀는 범주형 *Y* 반응 변수 수준의 확률을 하나 이상의 *X* 효과에 대한 함수로 모델링합니다. 모형 적합 플랫폼은 로지스틱 회귀 모형 적합을 위한 두 가지 분석법을 제공합니다. 사용하는 분석법은 반응 열의 모델링 유형(명목형 또는 순서형)에 따라 달라집니다.

로지스틱 회귀 모형 적합에 대한 자세한 내용은 Walker and Duncan 연구 자료(1967), Nelson 연구 자료(1976), Harrell 연구 자료(1986) 및 McCullagh and Nelder 연구 자료(1989)에서 확인하십시오.

로지스틱 회귀 모형의 파라미터화에 대한 자세한 내용은 "로지스틱 회귀 모형"(463 페이지)에서 확인하십시오.

명목형 로지스틱 회귀

반응 변수의 모델링 유형이 명목형이면 플랫폼에서 최대 가능도를 사용하여 선형 모형을 다중 수준 로지스틱 반응 함수에 적합시킵니다. 따라서 하나를 제외한 모든 반응 수준은 *X* 효과의 값이 주어졌을 때 반응 수준의 확률을 나타내는 로지스틱 곡선으로 모델링됩니다. 최종 반응 수준의 확률은 1에서 다른 적합 확률의 합을 뺀 값입니다. 결과적으로 *X* 효과의 모든 값에서 반응 수준에 대한 적합 확률의 합은 1입니다.

반응 변수가 이항인 경우 모형 적합 창에서 목표 수준을 설정하여 확률을 모델링할 수준을 지정할 수 있습니다. 기본적으로 모형은 반응 변수의 상위 수준 확률을 추정합니다.

명목형 반응 변수의 모형 적합에 대한 자세한 내용은 "통계 상세 정보" 장의 "명목형 반응"(489 페이지)에서 확인하십시오.

순서형 로지스틱 회귀

반응 변수의 모델링 유형이 순서형이면 플랫폼에서 최대 가능도를 사용하여 누적 반응 확률을 선형 모형의 로지스틱 함수에 적합시킵니다. 각 반응 수준과 같은 수준 또는 그 아래 수준의 반응이 나타날 누적 확률은 곡선으로 모델링됩니다. 이 곡선들은 오른쪽 또는 왼쪽으로 이동된다는 점만 제외하고는 각 수준마다 동일합니다.

팁 : 반응 수준이 많은 경우 순서형 모형이 명목형 모형보다 적합 속도가 훨씬 빠르고 메모리를 적게 사용합니다 .

순서형 반응 변수가 있는 모형의 적합에 대한 자세한 내용은 "통계 상세 정보" 장의 "순서형 반응"(491 페이지)에서 확인하십시오.

로지스틱 회귀 모형을 적합시키는 기타 JMP 플랫폼

JMP 에서는 다음과 같은 다양한 방법을 사용하여 로지스틱 회귀 모형을 적합시킬 수 있습니다 .

- 연속형 주효과 하나만 있는 로지스틱 회귀 모형을 적합시키려면 X 로 Y 적합 플랫폼을 사용하여 각 효과에 대한 누적 로지스틱 확률도를 표시할 수 있습니다 . 자세한 내용은 **기본 분석** 의 에서 확인하십시오 .
- 로지스틱 회귀 모형에서 변수 선택을 수행하려면 모형 적합 플랫폼의 단계별 분석법을 사용할 수 있습니다 . 자세한 내용은 " 단계별 회귀 모형 " 장 (229 페이지) 에서 확인하십시오 .
- 로짓 연결 이외의 연결 함수를 사용하는 로지스틱 회귀 모형을 적합시키려면 모형 적합 플랫폼의 일반화 선형 모형 분석법을 사용할 수 있습니다 . 자세한 내용은 " 일반화 선형 모형 " 장 (467 페이지) 에서 확인하십시오 .
- JMP PRO 로지스틱 회귀 모형에서 변수 선택을 수행하고 벌점 로지스틱 회귀 모형을 적합시키려면 모형 적합 플랫폼의 일반화 회귀 분석법을 사용할 수 있습니다 . 자세한 내용은 " 일반화 회귀 모형 " 장 (263 페이지) 에서 확인하십시오 .

로지스틱 회귀의 예

이 섹션에는 모형 적합 플랫폼의 각 로지스틱 회귀 분석법 (명목형 및 순서형) 에 대한 두 가지 예가 포함되어 있습니다 .

- " 명목형 로지스틱 회귀의 예 "(442 페이지)
- " 순서형 로지스틱 회귀의 예 "(444 페이지)

명목형 로지스틱 회귀의 예

다양한 가열 시간 및 유지 시간으로 준비된 금속 주괴에 대한 실험을 수행한 후 시작 준비 상태를 테스트했습니다 . 자세한 내용은 Cox and Snell 연구 자료 (1989) 에서 확인하십시오 . 데이터는 Ingots.jmp 샘플 데이터 테이블에 포함되어 있습니다 . 이 예에서 모형 적합 플랫폼은 heat 및 soak 회귀변수가 있는 로지스틱 회귀 모형을 사용하여 Ready 반응의 확률을 적합시킵니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ingots.jmp 를 엽니다 .

   범주형 변수 ready 의 값 Ready 와 Not Ready 는 주괴를 사용할 준비가 되었는지 여부를 나타냅니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. ready 를 선택하고 **Y** 를 클릭합니다 .

   명목형 모델링 유형의 열을 선택했으므로 모형 적합 분석법이 " 명목형 로지스틱 " 으로 업데이트됩니다 .

   ready 는 두 개의 수준만 있는 명목형 열이므로 " 목표 수준 " 옵션이 나타납니다 . 이 옵션을 사용하면 확률을 모델링할 반응 수준을 지정할 수 있습니다 .

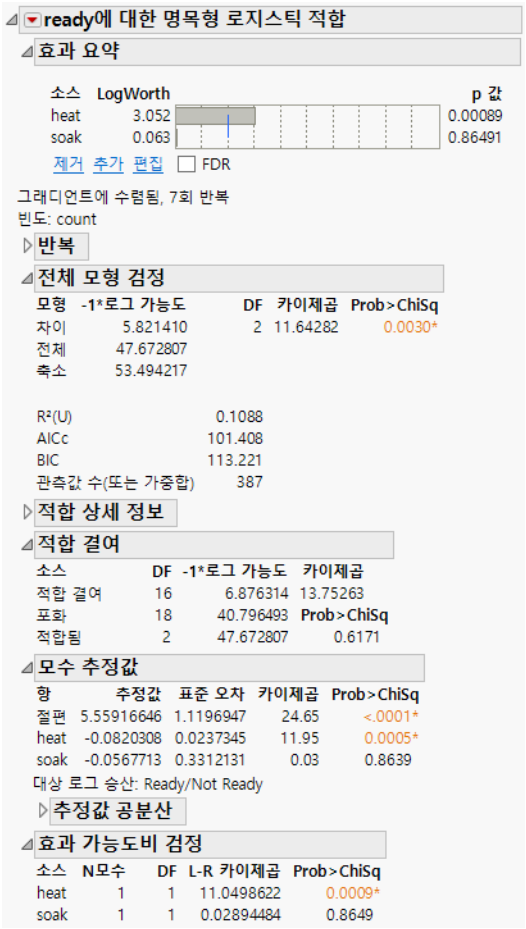
4. " 목표 수준 " 목록에서 "Ready" 를 선택합니다 .

   이 모형에서는 목표 수준이 Ready 이므로 Ready 반응의 확률을 모델링합니다 .

5. heat 와 soak 를 선택하고 **추가**를 클릭합니다 .
6. Count 를 선택하고 **빈도**를 클릭합니다 .
7. **실행**을 클릭합니다 .

   적합 프로세스가 수렴되면 명목형 회귀 보고서가 나타납니다 .

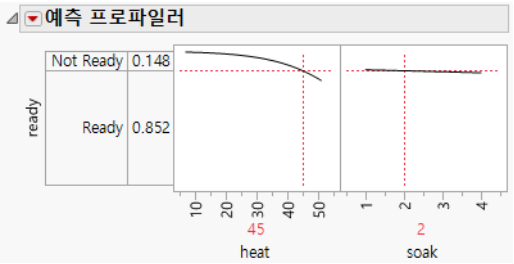
그림 11.2 명목형 로지스틱 적합 보고서



"전체 모형 검정" 보고서에서 카이제곱 통계량(11.64)의 p 값(0.0030)이 작으며 이는 전체 모형이 유의함을 나타냅니다. 그러나 soak에 대한 모수 추정값의 p 값은 0.8639이며 이는 유지 시간이 통계적으로 유의하지 않음을 나타냅니다.

8. "ready 에 대한 명목형 로지스틱 적합 " 옆의 빨간색 삼각형을 클릭하고 **프로파일러**를 선택합니다 .

그림 11.3 예측 프로파일러



heat가 45로 설정되고 soak가 2로 설정되면 ready의 확률이 0.85입니다.

이 시점에서 시작 준비 확률이 주어지고 특정 유지 시간에 대한 가열 시간을 찾기 위해 역추정 예측을 사용하는 데 관심이 있을 수 있습니다. 이에 대한 예는 "역추정 예측의 예"(455페이지)에서 계속됩니다.

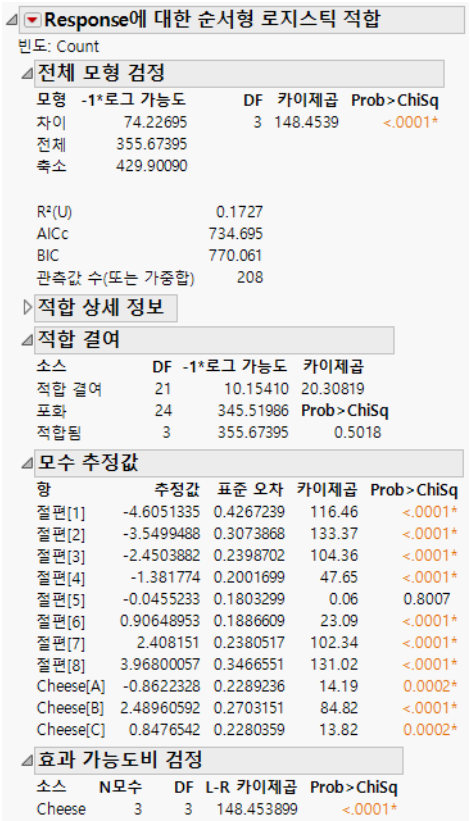
순서형 로지스틱 회귀의 예

다양한 치즈 첨가물 (A ~ D) 이 치즈 맛에 영향을 주는지 여부를 테스트하기 위한 실험이 수행되었습니다. 시식 패널이 맛을 측정하여 1(매우 싫음) ~ 9(최고의 맛) 의 순서형 척도를 기록했습니다 . 자세한 내용은 McCullagh and Nelder 연구 자료 (1989) 에서 확인하십시오 . 데이터는 Cheese.jmp 샘플 데이터 테이블에 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cheese.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. 응답을 선택하고 **Y** 를 클릭합니다 .

   순서형 모델링 유형의 열을 선택했으므로 모형 적합 분석법이 " 순서형 로지스틱 " 으로 업데이트됩니다 .
4. Cheese 를 선택하고 **추가**를 클릭합니다 .
5. Count 를 선택하고 **빈도**를 클릭합니다 .
6. **실행**을 클릭합니다 .

그림 11.4 순서형 로지스틱 적합 보고서



Response에 대한 순서형 로지스틱 적합

빈도: Count

전체 모형 검정

모형	-1*로그 가능도	DF	카이제곱	Prob>ChiSq
차이	74.22695	3	148.4539	<.0001*
전체	355.67395			
축소	429.90090			

R^2(U)	0.1727
AICc	734.695
BIC	770.061
관측값 수(또는 가중합)	208

적합 상세 정보

적합 결여

소스	DF	-1*로그 가능도	카이제곱
적합 결여	21	10.15410	20.30819
포화	24	345.51986	Prob>ChiSq
적합됨	3	355.67395	0.5018

모수 추정값

항	추정값	표준 오차	카이제곱	Prob>ChiSq
절편[1]	-4.6051335	0.4267239	116.46	<.0001*
절편[2]	-3.5499488	0.3073868	133.37	<.0001*
절편[3]	-2.4503882	0.2398702	104.36	<.0001*
절편[4]	-1.381774	0.2001699	47.65	<.0001*
절편[5]	-0.0455233	0.1803299	0.06	0.8007
절편[6]	0.90648953	0.1886609	23.09	<.0001*
절편[7]	2.408151	0.2380517	102.34	<.0001*
절편[8]	3.96800057	0.3466551	131.02	<.0001*
Cheese[A]	-0.8622328	0.2289236	14.19	0.0002*
Cheese[B]	2.48960592	0.2703151	84.82	<.0001*
Cheese[C]	0.8476542	0.2280359	13.82	0.0002*

효과 가능도비 검정

소스	N모수	DF	L-R 카이제곱	Prob>ChiSq
Cheese	3	3	148.453899	<.0001*

이 예의 모형 적합을 통해 절편만 있는 모형에 대한 -1*로그 가능도 429.9가 전체 모형에 대해 355.67로 감소합니다. 이 감소로 인해 자유도가 3인 전체 모형의 가능도비 카이제곱 통계량이 148.45로 산출됩니다. 따라서 인지된 치즈 맛의 차이는 매우 유의합니다.

가장 선호하는 치즈 첨가물은 음의 모수 추정값이 가장 큰 첨가물입니다. Cheese[D]는 설계 행렬의 자체 열이 없으므로 모수 추정값 보고서에 표시되지 않습니다. 그러나 Cheese D의 효과는 다른 효과의 음의 합계로 계산할 수 있으며 표 11.1에 표시됩니다.

표 11.1 Cheese.jmp 의 치즈 첨가물 선호도

Cheese	추정값	환경 설정
A	–1	두 번째 호감도
B	2.4896	최저 호감도
C	0.8477	세 번째 호감도
D	–1	최고 호감도

명목형 로지스틱 모형과의 비교

"적합 결여" 보고서에서는 모형이 데이터를 잘 적합시키는지에 대한 검정을 보여 줍니다.

순서형 문제로 볼 때, 처음 8개의 반응 수준에는 각각 절편이 있지만 네 개의 Cheese 수준에는 세 개의 모수만 있습니다. 따라서 순서형 모형의 자유도는 3입니다. 순서형 모형은 적합 결여 검정에서 적합 모형입니다.

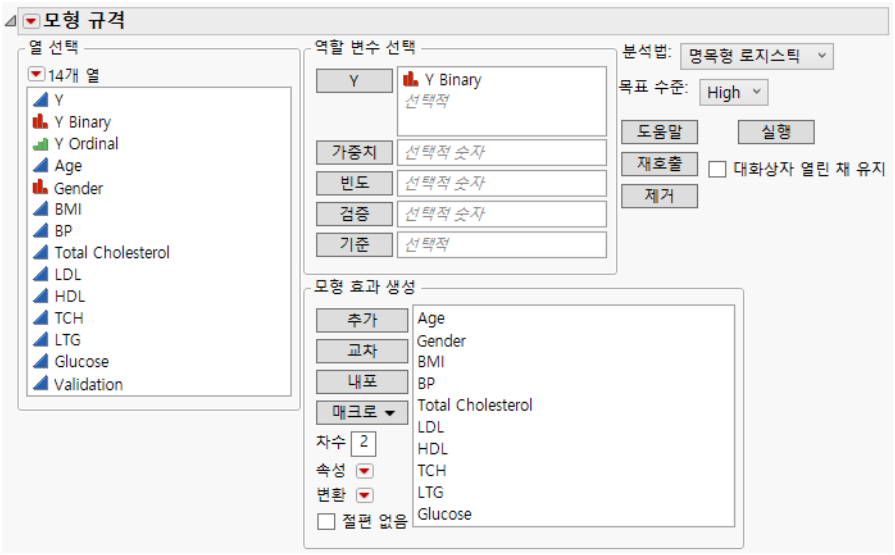
명목형 문제로 볼 때, 처음 8개의 반응 수준에는 각각 절편 및 네 개의 Cheese 수준에 대한 세 개의 모수가 있습니다. 따라서 명목형 모형의 자유도는 8 x 3 = 24입니다. 결과적으로 명목형 모형은 적합 결여 검정에서 포화 모형입니다.

이 예에서 순서형 모형의 적합 결여 검정은 명목형 모형을 비교 대상으로 순서형 반응 모형을 검정합니다. 적합 결여가 유의하지 않으면 순서형 모형이 적절하다고 간주합니다.

명목형 및 순서형 로지스틱 분석법 시작

분석 > 모형 적합을 선택하고 **Y** 에 대해 하나 이상의 비연속형 열을 입력하여 명목형 로지스틱 및 순서형 로지스틱 분석법을 시작합니다 . **Y** 에 여러 열이 입력되면 각 반응에 대한 모형이 적합됩니다 .

그림 11.5 명목형 로지스틱이 선택된 모형 적합 시작 창



"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 "모형 규격" 장(29페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 명목형 로지스틱 및 순서형 로지스틱 분석법에 대한 정보를 제공합니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격"의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

절편 항이 없는 모형을 지정하려면 "모형 적합" 창의 "모형 효과 생성" 패널에서 "절편 없음" 옵션을 선택합니다. 순서형 로지스틱 분석법에서는 "절편 없음" 옵션을 사용할 수 없습니다.

로지스틱 회귀 모형의 관심 사건은 "모형 적합" 창의 "목표 수준" 옵션에 의해 정의됩니다. 이 옵션은 명목형 로지스틱 분석법에서 이항 반응 열을 지정하는 경우에만 사용할 수 있습니다.

참고 : 모형 적합 플랫폼의 로지스틱 분석법을 사용하려면 모든 반응이 하나의 열에 있도록 데이터가 쌓인 형식이어야 합니다 . 데이터가 여러 열에 포함되도록 형식이 지정되는 경우도 있습니다 . 여러 열의 반응을 단일 열 반응으로 변환하는 예는 " 예 : 여러 열의 개수 쌓기 "(462 페이지)에서 확인하십시오 .

JMP PRO 로지스틱 회귀 모형의 검증

검증은 데이터 집합의 일부를 사용하여 모형 모수를 추정하고 다른 부분을 사용하여 모형의 예측 능력을 평가하는 프로세스입니다 .

- **훈련** 데이터 집합은 모형 모수를 추정하는 데 사용됩니다 .
- **검증** 데이터 집합은 모형 적합에서 모형의 예측 능력을 평가하거나 검증하는 데 사용됩니다.
- **테스트** 데이터 집합은 모형의 예측 능력에 대한 독립된 최종 평가입니다 . 테스트 데이터 집합은 검증 열을 사용하는 경우에만 사용할 수 있습니다 .

훈련, 검증 및 테스트 데이터 집합은 원래 데이터의 부분집합으로 생성됩니다. 이는 모형 적합 시작 창에서 검증 열을 사용하여 수행됩니다.

검증 열의 값은 데이터가 분할되는 방법과 검증에 사용되는 방법을 결정합니다 .

- 열에 두 개의 구분되는 값이 있으면 훈련 및 검증 데이터 집합이 생성됩니다 .
- 열에 세 개의 구분되는 값이 있으면 훈련 , 검증 및 테스트 데이터 집합이 생성됩니다 .
- 열에 네 개 이상의 구분되는 값이 있거나 값이 하나뿐이면 검증이 수행되지 않습니다 .

검증 열을 사용하면 "적합 상세 정보" 보고서에 훈련, 검증 및 테스트 데이터 집합에 대한 모형 적합 통계량이 제공됩니다. 또한 각 훈련, 검증 및 테스트 데이터 집합에 대해 별도의 ROC 곡선, Lift 곡선 및 혼동 행렬이 있습니다.

JMP 모델링 플랫폼에서 검증 열이 사용되는 방법에 대한 자세한 내용은 Predictive and Specialized Modeling의 에서 확인하십시오.

로지스틱 적합 보고서

명목형 로지스틱 또는 순서형 로지스틱 분석법을 사용하여 모형을 적합시키면 각각 명목형 로지스틱 적합 보고서 또는 순서형 로지스틱 적합 보고서가 생성됩니다 . 기본적으로 이러한 보고서에는 다음 보고서가 포함됩니다 .

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 보고서입니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 효과 요약 보고서 "(168 페이지) 에서 확인하십시오 .

로지스틱 그림 (모형이 단일 연속형 효과로 구성된 경우에만 사용 가능) 로지스틱 확률도에서는 로지스틱 모형의 적합 상태를 보여 줍니다 . 가로 축의 각 값에서 세로 방향의 확률 척도가 각 반응 범주의 확률로 분할됩니다 . 이 확률은 곡선 사이의 세로 거리로 측정되며 모든 반응 범주 확률의 합계는 1 이 됩니다 .

로지스틱 그림에서 점은 데이터 테이블의 관측값을 나타냅니다 . 각 점의 가로 위치는 연속형 요인의 값에 따라 결정됩니다 . 각 점의 세로 위치는 반응 범주의 값에 해당하는 곡선 사이에 오도록 무작위로 선택됩니다 . 고정된 난수 시드값이 사용되므로 세로 위치는 동일한 모형을 여러 번 적합시켜도 달라지지 않습니다 .

반복 (명목형 로지스틱 분석법에서만 사용 가능) 모형 적합을 시작한 후 반복 추정 프로세스가 시작되고 반복별로 보고됩니다 . 적합 프로세스가 완료되면 반복 기록 보고서를 열어 반복 단계를 확인할 수 있습니다 .

전체 모형 검정 절편 모수를 제외한 모든 회귀변수 효과가 생략된 모형과 전체 모형 적합을 비교하는 검정을 표시합니다 . 이 검정은 연속형 반응에 대한 분산 분석 테이블과 유사합니다 . 전체 모형 검정 보고서에 대한 자세한 내용은 " 전체 모형 검정 "(449 페이지) 에서 확인하십시오 .

적합 상세 정보 모형에 대한 다양한 적합 측도를 표시합니다 . 자세한 내용은 " 적합 상세 정보 "(450 페이지) 에서 확인하십시오 .

적합 결여 (*X* 효과와 관련한 반복된 점이 있고 모형이 포화되지 않은 경우에만 사용 가능) 모형에 항이 더 필요한지 여부를 확인하는 적합 결여 검정 (적합도 검정이라고도 함) 을 표시합니다 . 자세한 내용은 " 적합 결여 검정 "(450 페이지) 에서 확인하십시오 .

모수 추정값 모수 추정값 , 표준 오차 및 관련된 가설 검정을 표시합니다 . " 추정값 공분산 " 보고서에는 모수 추정값의 분산 및 공분산이 포함됩니다 .

참고 : " 추정값 공분산 " 보고서는 순서형 반응 변수가 아니라 명목형 반응 변수에 대해서만 표시됩니다 .

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 표시됨) 모형 항이 충족하는 선형 함수가 포함된 보고서를 표시합니다 .

효과 가능도비 검정 가능도비 카이제곱 검정은 전체 모형과 검정할 가설의 제약을 받는 모형 간의 로그 가능도 차이의 두 배로 계산됩니다 . 제약이 있는 모형은 효과를 포함하지 않는 모형입니다 . 이러한 검정은 각 검정마다 별도의 반복 집합이 필요하므로 시간이 걸릴 수 있습니다 .

참고 : 가능도비 검정은 완료하는 데 20 초 미만이 걸릴 것으로 예상되는 경우 플랫폼 기본값입니다 . 그렇지 않은 경우 기본 효과 검정은 Wald 검정입니다 .

전체 모형 검정

" 전체 모형 검정 " 테이블에는 절편 모수를 제외한 모든 회귀 모수가 생략된 모형과 전체 모형 적합을 비교하는 검정이 표시됩니다. 이 검정은 연속형 반응에 대한 분산 분석 테이블과 유사합니다. 음의 로그 가능도는 제곱합에 해당하고 카이제곱 검정은 *F* 검정에 해당합니다.

" 전체 모형 검정 " 테이블에는 다음과 같은 통계량이 표시됩니다 .

모형 모형 라벨을 나열합니다 .

차이 **전체** 모형과 **축소** 모형 간의 차이입니다 . 이 모형은 적합에 대한 전체 회귀변수의 유의성을 측정하는 데 사용됩니다 .

전체 절편 및 모든 효과를 포함하는 완전 모형입니다 .

축소 절편 모수만 포함하는 모형입니다 .

-1* 로그 가능도 각 모형에 대한 음의 로그 가능도입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

DF 전체 모형과 축소 모형 간의 차이에 대한 DF(자유도) 입니다 .

카이제곱 모든 회귀 모수가 0 이라는 가설에 대한 가능도비 카이제곱 검정 통계량입니다 . 이 검정 통계량은 적합 모형과 절편만 있는 축소 모형 간의 음의 로그 가능도 차이의 두 배로 계산됩니다 .

Prob>ChiSq 지정된 모형이 절편만 포함된 모형보다 더 잘 적합되지 않을 경우 카이제곱 값이 더 클 확률입니다 .

R^2(U) 모형 적합에 기인한 전체 불확실성의 비율로 , 축소된 음의 로그 가능도 값에 대한 차이의 비율로 정의됩니다 . R^2 값은 0(적합이 개선되지 않는 경우) 에서 1(완벽한 적합) 까지입니다 . R^2(U) 값이 1 이면 발생하는 사건의 예측 확률이 1 과 같음을 나타냅니다 . 즉 , 예측 확률에 불확실성이 없는 것입니다 . 로짓 모형의 경우 예측 확률은 거의 확실하지 않으므로 R^2(U) 가 작은 경향이 있습니다 .

R^2(U) 는 불확도 계수인 *U* 또는 McFadden 유사 R^2 이라고도 합니다 .

AICc 수정된 Akaike 정보 기준입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

BIC Bayes 정보 기준 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

관측값 수 (또는 가중합) 표본의 총 관측값 수입니다 . " 모형 적합 " 창에서 " 빈도 " 또는 " 가중치 " 열이 지정되는 경우 이 값은 빈도 또는 가중치 역할에 할당된 열 값의 합입니다 .

적합 상세 정보

" 적합 상세 정보 " 보고서에는 다음과 같은 통계량이 포함됩니다 .

엔트로피 R² R²(U) 와 동등합니다 . 자세한 내용은 " 전체 모형 검정 "(449 페이지) 에서 확인하십시오 .

일반화 R² 일반 회귀 모형에 적용할 수 있는 측도입니다 . 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다 . 일반화 R² 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R² 로 단순화됩니다 . 일반화 R² 를 Nagelkerke 또는 Craig and Uhler R^2 라고도 하는데 , 이는 Cox-Snell 유사 R^2 를 정규화한 것입니다 . 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오 .

평균 -Log p -log(*p*) 의 평균입니다 . 여기서 *p* 는 발생한 사건과 연관된 적합 확률입니다 .

RASE 제곱근 평균 제곱 오차입니다 . 여기서 차이는 반응과 *p*(실제로 발생한 사건의 적합 확률) 사이의 차이입니다 .

평균 절대 편차 반응과 p(실제로 발생한 사건의 적합 확률) 사이의 차이에 대한 절대값을 평균한 것입니다 .

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다 .

N 관측값 수입니다 .

엔트로피 R² 및 일반화 R² 의 경우 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 . 평균 -Log p, RASE, 평균 절대 편차 및 오분류 비율의 경우 값이 작을수록 더 나은 적합을 나타냅니다 .

효과가 전체적으로 유의한지 검정하기 위해 (전체 모형 검정) 카이제곱 통계량은 적합 모형과 절편만 있는 축소 모형 간의 음의 로그 가능도 차이의 두 배로 계산됩니다 .

검증 열을 지정한 경우 각 훈련 , 검증 및 테스트 데이터 집합에 대한 열이 " 적합 상세 정보 " 보고서에 포함됩니다 .

적합 결여 검정

적합 결여 검정은 현재 모형에 충분한 정보가 있는지 아니면 복합 항이 더 필요한지를 확인합니다 . 이 검정을 적합도 검정이라고도 합니다 . 적합 결여 검정은 데이터에서 모형 효과 값의 모든 조합에 대한 범주를 구성하여 순수 오차 음의 로그 가능도를 계산합니다 . " 적합 결여 " 테이블의 **포화** 행은 이 로그 가능도를 포함합니다 . " 적합 결여 " 보고서에는 **포화** 로그 가능도가 **적합됨** 모형보다 유의하게 나은지에 대한 검정도 포함되어 있습니다 .

포화 자유도는 *m*–1 이며 , 여기서 *m* 은 고유 모집단의 수입니다 . **적합됨** 자유도는 절편을 제외한 모수의 수입니다 .

" 적합 결여 " 테이블에는 **적합 결여** 로 인한 오차 , **포화** 모형의 오차 (순수 오차) 및 **적합** 모형의 총 오차에 대한 음의 로그 가능도가 포함됩니다 . 카이제곱 통계량은 적합 결여를 검정합니다 .

로지스틱 적합 플랫폼 옵션

- " 명목형 및 순서형 적합 옵션 "
- " 명목형 적합 옵션 "
- " 순서형 적합 옵션 "

명목형 및 순서형 적합 옵션

" 명목형 로지스틱 적합 " 과 " 순서형 로지스틱 적합 " 의 빨간색 삼각형 메뉴에서 모두 다음 옵션을 사용할 수 있습니다 .

로지스틱 그림 (모형이 단일 연속형 효과로 구성된 경우에만 사용 가능) " 로지스틱 그림 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 로지스틱 적합 보고서 "(448 페이지) 에서 확인하십시오 .

가능도비 검정 " 효과 가능도비 검정 " 보고서를 표시하거나 숨깁니다 . 가능도비 카이제곱 검정은 전체 모형과 검정할 가설의 제약을 받는 모형 간의 로그 가능도 차이의 두 배로 계산됩니다 . 제약이 있는 모형은 효과를 포함하지 않는 모형입니다 . 이러한 검정은 각 검정마다 별도의 반복 집합이 필요하므로 시간이 걸릴 수 있습니다 . 따라서 큰 문제의 경우 시간이 오래 걸릴 수 있습니다 .

참고 : 가능도비 검정은 완료하는 데 20 초 미만이 걸릴 것으로 예상되는 경우 플랫폼 기본값입니다 . 이 기본 옵션을 사용하는 것이 좋습니다 .

Wald 검정 " 효과 Wald 검정 " 보고서를 표시하거나 숨깁니다 . Wald 카이제곱은 가능도비 검정에 대한 2 차 근사이며 계산의 부산물입니다 . Wald 검정은 신뢰도가 낮다고 간주되지만 효과 선별을 위해 적절한 유의성 지표를 제공합니다. 각 모수 추정값 및 효과가 Wald 검정과 함께 표시됩니다 . 이 검정은 가능도비 검정을 완료하는 데 20 초를 초과할 것으로 예상되는 경우 기본 검정입니다 .

신뢰 구간 모형 모수에 대한 프로파일 - 가능도 신뢰 구간을 표시하거나 숨깁니다 . " 모형 적합 " 창의 " 모형 규격 " 빨간색 삼각형 메뉴에서 " 유의 수준 설정 " 을 선택하여 신뢰 수준을 변경할 수 있습니다 . 각 신뢰 한계에는 모형 적합에 일련의 반복이 필요하며 계산하는 데 시간이 오래 걸릴 수 있습니다 . 그뿐만 아니라 한계를 찾는 데 실패할 수도 있습니다 .

ROC 곡선 모형에 대한 ROC 곡선을 표시하거나 숨깁니다 . ROC(Receiver Operating Characteristic) 곡선은 반응 수준을 정렬하기 위해 모형의 적합 확률의 정렬 효율성을 측정합니다 . ROC 곡선은 진단 검정에서 기준점을 설정하는 데도 도움이 될 수 있습니다 . 주대각에서 곡선과의 거리가 멀어질수록 적합도가 높습니다. ROC 곡선에 대한 소개는 **기본 분석**의 에서 확인하십시오 .

로지스틱 적합에 세 개 이상의 반응 수준이 있는 경우 일반화 ROC 곡선 (파티션 플랫폼의 ROC 곡선과 동일) 이 생성됩니다 . 이 그림에는 각 반응 수준에 대한 곡선 , 즉 다른 모든 수

준에 대한 해당 수준의 ROC 곡선이 있습니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

검증 열을 지정한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 ROC 곡선이 표시됩니다 .

Lift 곡선 모형에 대한 Lift 곡선을 표시하거나 숨깁니다. Lift 곡선은 ROC 곡선과 동일한 정보를 보여 주되, 시작점에서 정렬의 풍부도를 극화하는 방법을 사용합니다 . 세로 축은 모집단 중 해당 부분이 해당 반응 수준 전체의 비율에 비해 선택된 반응 수준에 얼마나 풍부하게 포함되었는지를 보여 줍니다 . 검증 열을 지정한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 Lift 곡선이 표시됩니다 . Lift 곡선에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

혼동 행렬 " 혼동 행렬 " 보고서 및 " 혼동 비율 " 보고서가 포함된 혼동 통계량 보고서를 표시하거나 숨깁니다 . 두 보고서 모두 실제 반응 수준과 예측 반응 수준의 이원 분류입니다 . 예측 반응 수준은 시작 창에 지정된 목표 수준입니다 . 혼동 비율 보고서는 혼동 행렬 보고서와 동일하며 개수를 행 합계로 나눕니다 .

적합한 모형은 예측 반응 수준이 실제 반응 수준과 동일해야 합니다 . 혼동 행렬 보고서를 통해 예측 반응이 실제 반응에 잘 맞는지 평가할 수 있습니다 . 검증 열을 지정한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 혼동 행렬이 표시됩니다 .

반응이 명목형이고 " 수익 행렬 " 열 특성이 있는 경우 이 옵션을 선택하면 결정 행렬 보고서도 나타납니다 . 결정 행렬 보고서에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

프로파일러 모형의 요인 값이 변경될 때 지정된 반응 확률에 대한 적합된 값을 보여 주는 예측 프로파일러를 표시하거나 숨깁니다 . 이 기능은 명목형 반응과 순서형 반응에 모두 사용할 수 있습니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

등고선 프로파일러 (모형에 두 개 이상의 연속형 요인이 포함된 경우에만 사용 가능) 등고선 프로파일러를 표시하거나 숨깁니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다 .

효과 요약 모형의 효과를 대화식으로 업데이트할 수 있는 " 효과 요약 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 로지스틱 적합 보고서 "(448 페이지) 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

명목형 적합 옵션

"명목형 로지스틱 적합"의 빨간색 삼각형 메뉴에서 다음 옵션을 사용할 수 있습니다.

그림 옵션 (모형이 단일 연속형 효과로 구성된 경우에만 사용 가능) "그림 옵션" 메뉴에는 다음 옵션이 포함되어 있습니다.

점 표시 로지스틱 그림의 점을 표시하거나 숨깁니다.

비율 곡선 표시 효과 변수의 다양한 값에서 추정된 비율을 적합된 로지스틱 곡선과 비교할 수 있습니다. 이 옵션은 효과의 각 값에 대해 여러 점이 있는 경우에만 유용합니다. 이 경우 각 값의 적절한 비율 추정값을 구하고 이 비율을 적합된 로지스틱 곡선과 비교할 수 있습니다. 퇴화 점(일반적으로 0 또는 1 일 때)이 너무 많아지지 않도록 JMP 에서는 x 값에 점이 세 개 이상 있는 경우에 비율 값만 표시합니다.

선 색상 그림의 곡선 색상을 지정합니다.

승산비 단위 승산비와 범위 승산비가 포함된 승산비 보고서를 표시하거나 숨깁니다.

그림 11.6 승산비

승산비

ready의 Ready 대 Not Ready 승산

단위 승산 비율

회귀변수의 단위 변화당

항	승산 비율	하위 95%	상위 95%	역수
heat	0.921244	0.87937	0.965111	1.0854892
soak	0.94481	0.493646	1.808313	1.0584137

범위 승산 비율

전체 범위에 대한 회귀변수의 변화당

항	승산 비율	하위 95%	상위 95%	역수
heat	0.027069	0.003496	0.209605	36.942229
soak	0.8434	0.120295	5.913179	1.185677

승산비에 대한 검정 및 신뢰 구간이 Wald 방법에 기반합니다.

역추정 예측 (2 수준 명목형 반응에만 사용 가능) 지정된 확률이 되는 x 값을 찾습니다. 역추정 예측에 대한 자세한 내용은 부록 "통계 상세 정보"(487 페이지)에서 확인하십시오.

확률 계산식 저장 현재 데이터 테이블에 반응 수준의 선형 결합 계산식, 각 반응 수준에 대한 예측 계산식, 가장 확률이 높은 반응을 제공하는 예측 계산식을 포함하는 열을 생성합니다.

r 개 수준이 있는 명목형 반응 모형의 경우 JMP 에서 다음 열이 생성됩니다.

- 반응 수준 $j = 1, 2, \ldots r - 1$ 에 대한 회귀변수의 선형 결합을 포함하는 **선형 [j]** 열
- 마지막 수준 r 에 대한 적합 계산식을 포함하는 **확률 [r]** 열
- $j < r$ 일 때 j 수준에 대한 적합 계산식을 포함하는 **확률 [j]** 열

– 계산된 확률을 기반으로 각 행의 가장 확률이 높은 수준을 선택하는 **가장 확률이 높은 ' 반응 이름 ' 열**

JMP PRO 확률 계산식 게시 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

표시자 파라미터화 추정값 (모형 효과 중 명목형 열이 있는 경우에만 사용 가능) 명목형 열이 표시자 (SAS GLM) 파라미터화를 사용하여 코드화되고 연속형으로 처리되는 모형에 대한 모수 추정값을 제공하는 " 표시 함수 파라미터화 " 보고서를 표시하거나 숨깁니다 . 순서형 열은 일반적인 JMP 코딩 체계를 사용하여 코드화됩니다 . SAS GLM 및 JMP 코딩 체계에 대한 설명은 " 통계 상세 정보 " 부록의 " 요인 모형 "(492 페이지) 에서 확인하십시오 .

주의 : " 표시 함수 파라미터화 " 보고서에 제공된 표준 오차 및 카이제곱 값은 " 모수 추정값 " 보고서의 값과 다릅니다 . 이는 추정하는 모형 모수가 서로 다르기 때문입니다 .

순서형 적합 옵션

" 순서형 로지스틱 적합 " 의 빨간색 삼각형 메뉴에서 다음 옵션을 사용할 수 있습니다 .

저장 다음 저장 옵션을 포함합니다 .

확률 계산식 저장 현재 데이터 테이블에 반응 수준의 선형 결합 계산식 , 각 반응 수준에 대한 예측 계산식 , 가장 확률이 높은 반응을 제공하는 예측 계산식을 포함하는 열을 생성합니다 .

r 개 수준이 있는 순서형 반응 모형의 경우 JMP 에서 다음 열이 생성됩니다 .

– 절편 항이 없는 회귀변수의 선형 결합에 대한 계산식을 포함하는 **선형** 열

– *j* = 1, 2, ... *r* - 1 수준에 대해 각각 반응이 *j* 수준보다 작거나 같을 누적 확률에 대한 계산식을 포함하는 **누적 [j]** 열 . 모든 행에 대해 1 인 **누적 [j = 1, 2, ... r - 1]** 은 없습니다 .

– 1 < *j* < *r* 일 때 각각 반응이 *j* 수준일 확률에 대한 계산식을 포함하는 **확률 [j = 1, 2, ... r - 1]** 열 . **확률 [j]** 는 **누적 [j]** 와 **누적 [j – 1]** 의 차이입니다 . **확률 [1]** 은 **누적 [1]** 이고 **확률 [r]** 은 1 – **누적 [r – 1]** 입니다 .

– 계산된 확률을 기반으로 각 행의 가장 확률이 높은 수준을 선택하는 **가장 확률이 높은 ' 반응 이름 ' 열**

JMP PRO 확률 계산식 게시 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오 .

분위수 저장 (반응이 수치형이고 순서형 모델링 유형인 경우에만 사용 가능) 현재 데이터 테이블에 세 가지 확률에 대한 분위수를 적합시키는 OrdQ.05, OrdQ.50 및 OrdQ.95 열을 생성합니다 .

기대값 저장 (반응이 수치형이고 순서형 모델링 유형인 경우에만 사용 가능) 현재 데이터 테이블에 **기대 순서** 열을 생성합니다 . 이 열에는 각 행에 대한 적합 반응 확률과 반응 값의 선형 결합이 포함되며 기대값을 제공합니다 .

로지스틱 회귀의 추가 예

이 섹션에는 모형 적합 플랫폼에서 로지스틱 회귀를 사용하는 추가 예가 포함되어 있습니다 .

- " 역추정 예측의 예 "
- " 명목형 로지스틱 모형에 대한 효과 요약의 사용 예 "
- "2 차 순서형 로지스틱 모형의 예 "
- " 예 : 여러 열의 개수 쌓기 "

역추정 예측의 예

역추정 예측을 사용하면 지정된 반응 확률에서 하나의 독립 변수에 대한 값을 예측할 수 있습니다 . 독립 변수가 두 개 이상인 경우 예측 대상 변수를 제외한 모든 변수의 값을 지정해야 합니다 .

금속 주괴에 대한 실험이 수행되었습니다. 다양한 가열 시간 및 유지 시간으로 주괴를 준비한 후 시작 준비 상태를 테스트했습니다. 자세한 내용은 Cox and Snell 연구 자료 (1989) 에서 확인하십시오. 데이터는 JMP 의 Ingots.jmp 샘플 데이터 테이블에 포함되어 있습니다.

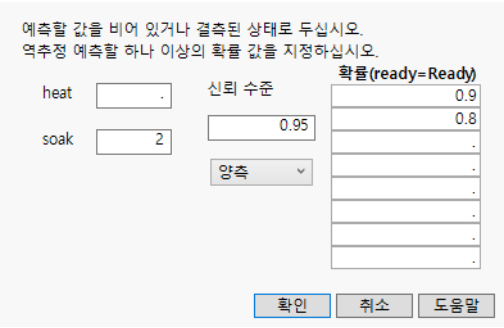
시작 준비 확률 0.8 과 0.9 에 대해 유지 시간 2 에서 가열 시간을 예측하는 데 관심이 있습니다 .

1. 분석을 위해 " 명목형 로지스틱 회귀의 예 "(442 페이지) 의 단계를 수행합니다 .
2. "ready 에 대한 명목형 로지스틱 적합 " 옆의 빨간색 삼각형을 클릭하고 **역추정 예측**을 선택합니다 .
3. heat 의 값을 삭제합니다 .

   heat 의 예측값을 찾아야 하므로 heat 값을 비워 둡니다 .
4. soak 에 2 를 입력합니다 .

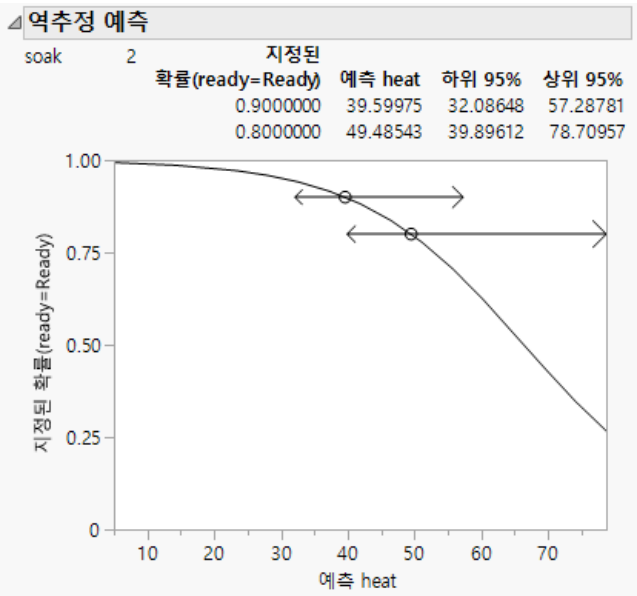
   soak 가 2 일 때 가열 시간을 예측하려고 합니다 .
5. " 확률 " 아래의 처음 두 행에 0.9 와 0.8 을 입력합니다 .

그림 11.7 역추정 예측 규격 창



6. **확인**을 클릭합니다.

그림 11.8 역추정 예측 보고서



soak가 2일 때 주괴 시작 준비가 완료될 확률이 90%인 heat의 예측값은 39.60이며, 32.09 ~ 57.29의 95% 신뢰 구간을 같이 가집니다. soak가 2일 때 주괴 시작 준비가 완료될 확률이 80%인 heat의 예측값은 49.49이며, 39.90 ~ 78.71의 95% 신뢰 구간을 같이 가집니다.

명목형 로지스틱 모형에 대한 효과 요약의 사용 예

세제 브랜드 선호도를 평가하기 위한 시장 조사 연구가 수행되었습니다 . 자세한 내용은 Ries and Smith 연구 자료 (1963) 에서 확인하십시오 . 연구 결과는 Detergent.jmp 샘플 데이터 테이블에 있습니다 . 모형은 다음과 같이 정의됩니다 .

- brand 반응 변수 - m, x 값 포함
- softness(수질 연성) 효과 - soft, medium, hard 값 포함
- previous use 효과 - yes, no 값 포함
- temperature 효과 - high, low 값 포함
- count 개수 변수 - 효과 범주의 각 조합에 대한 빈도 수 제공

3 요인 완전 요인 모형을 지정하여 연구를 시작합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Detergent.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 brand 를 선택하고 **Y** 를 클릭합니다 .

   명목형 반응 변수를 지정했으므로 " 분석법 " 이 " 명목형 로지스틱 " 으로 변경됩니다 .

   brand 는 두 개의 수준만 있는 명목형 열이므로 " 목표 수준 " 옵션이 나타납니다 . 이 옵션을 사용하면 확률을 모델링할 반응 수준을 지정할 수 있습니다 .
4. " 목표 수준 " 목록에서 "m" 을 선택합니다 .
5. Count 를 선택하고 **빈도**를 클릭합니다 .
6. softness ~ temperature 를 선택하고 **매크로 > 완전 요인**을 클릭합니다 .
7. **실행**을 클릭합니다 .

그림 11.9 3 요인 요인 모형에 대한 명목형 로지스틱 적합

brand에 대한 명목형 로지스틱 적합

효과 요약

그래디언트에 수렴됨, 3회 반복

빈도: count

반복

전체 모형 검정

모형	-1*로그 가능도	DF	카이제곱	Prob>ChiSq
차이	16.41281	11	32.82562	0.0006*
전체	682.24780			
축소	698.66061			

R^2(U)	0.0235
AICc	1388.81
BIC	1447.48
관측값 수(또는 가중합)	1008

적합 상세 정보

모수 추정값

효과 가능도비 검정

소스	N모수	DF	L-R 카이제곱	Prob>ChiSq
softness	2	2	0.09804239	0.9522
previous use	1	1	22.1316677	<.0001*
softness*previous use	2	2	3.78609668	0.1506
temperature	1	1	3.63914017	0.0564
softness*temperature	2	2	0.19617686	0.9066
previous use*temperature	1	1	2.26089203	0.1327
softness*previous use*temperature	2	2	0.73731691	0.6917

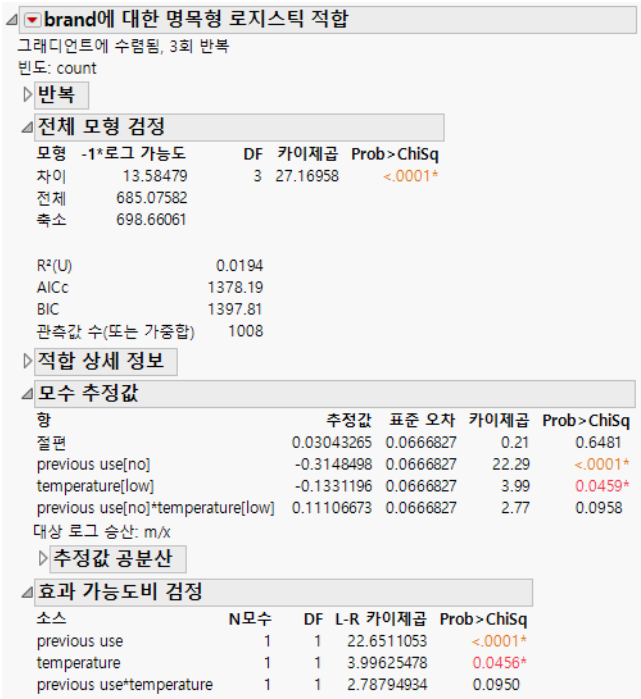
" 전체 모형 검정 " 보고서에서는 3 요인 완전 요인 모형이 전체적으로 유의하다는 것을 보여 줍니다 (Prob>ChiSq = 0.0006).

" 효과 가능도비 검정 " 보고서에서는 softness 를 포함하는 효과가 모형 적합에 유의하게 기여하지 않음을 보여 줍니다 . 따라서 softness 를 모형에서 제거할 수도 있습니다 . " 효과 요약 " 보고서에서 이를 수행할 수 있습니다 .

8. " 효과 요약 " 보고서의 " 소스 " 아래에서 softness*previous use ~ softness 를 선택하고 **제거** 를 클릭합니다 .

   보고서가 업데이트되어 2 요인 요인 모형이 표시됩니다 (그림 11.10). " 전체 모형 검정 " 보고서를 보면 2 요인 모형도 전체적으로 유의함을 알 수 있습니다 .

그림 11.10 2 요인 요인 모형에 대한 명목형 로지스틱 적합



brand에 대한 명목형 로지스틱 적합

그래디언트에 수렴됨, 3회 반복

빈도: count

▷ 반복

▲ 전체 모형 검정

모형	-1*로그 가능도	DF	카이제곱	Prob>ChiSq
차이	13.58479	3	27.16958	<.0001*
전체	685.07582			
축소	698.66061			

R^2(U)	0.0194
AICc	1378.19
BIC	1397.81
관측값 수(또는 가중합)	1008

▷ 적합 상세 정보

▲ 모수 추정값

항	추정값	표준 오차	카이제곱	Prob>ChiSq
절편	0.03043265	0.0666827	0.21	0.6481
previous use[no]	-0.3148498	0.0666827	22.29	<.0001*
temperature[low]	-0.1331196	0.0666827	3.99	0.0459*
previous use[no]*temperature[low]	0.11106673	0.0666827	2.77	0.0958

대상 로그 승산: m/x

▷ 추정값 공분산

▲ 효과 가능도비 검정

소스	N모수	DF	L-R 카이제곱	Prob>ChiSq
previous use	1	1	22.6511053	<.0001*
temperature	1	1	3.99625478	0.0456*
previous use*temperature	1	1	2.78794934	0.0950

세제 브랜드의 이전 사용 여부와 수온이 세제 선호도에 영향을 미친다는 결론을 내릴 수 있습니다. 또한 온도와 이전 사용 여부의 교호작용은 통계적으로 유의하지 않으므로 온도가 이전 사용에 따라 달라진다는 증거가 없습니다.

2 차 순서형 로지스틱 모형의 예

순서형 로지스틱 분석법은 순서형 반응의 높은 수준 또는 낮은 수준의 확률을 최적화하기 위해 2차 표면을 적합시킬 수 있습니다.

이 예에서 전자레인지 팝콘 제조업체는 소비자가 좋아하는 팝콘의 소금 양이 어느 정도인지 확인하려고 합니다. 이를 위해 제조업체는 팝콘 패키지에 추가되는 소금의 양에 따라 양호한 반응의 최대 확률을 찾습니다. 이 실험에서 소금의 양은 0, 1, 2 및 3 티스푼으로 조절됩니다. 응답자는 1(낮음) ~ 5(높음)의 척도로 맛을 평가합니다. 더 양호한 반응의 확률을 최대화하는 양이 최적의 소금 양입니다. 표 11.2에서는 각 소금 수준에 대한 10 개의 관측값을 보여 줍니다.

표 11.2 팝콘의 소금

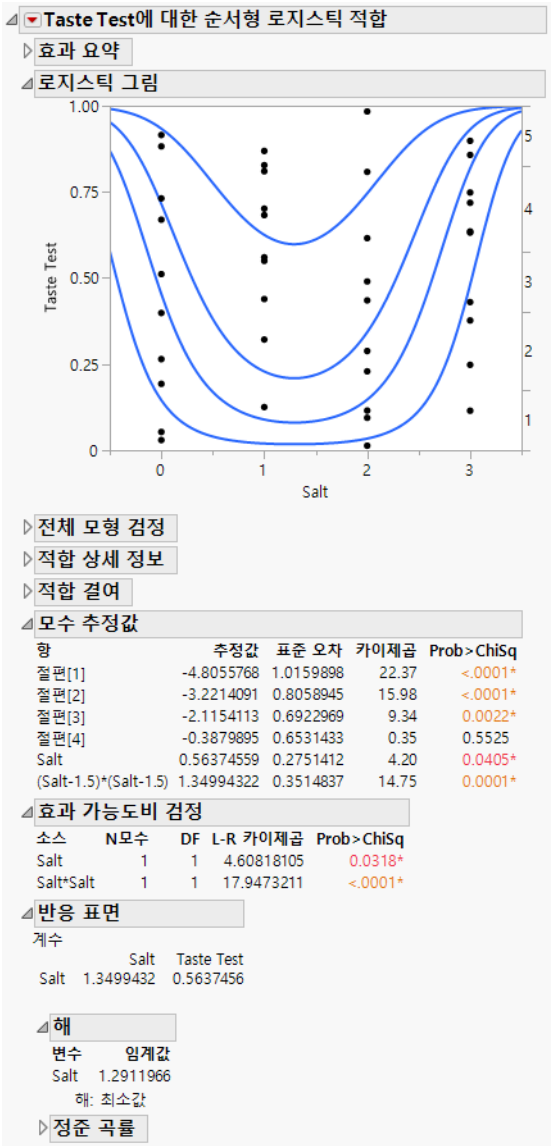
소금 양	소금 평가 반응
소금 없음	1, 3, 2, 4, 2, 2, 1, 4, 3, 4

표 11.2 팝콘의 소금 (계속)

소금 양	소금 평가 반응
1 티스푼	4, 5, 3, 4, 5, 4, 5, 5, 4, 5
2 티스푼	4, 3, 5, 1, 4, 2, 5, 4, 3, 2
3 티스푼	3, 1, 2, 3, 1, 2, 1, 2, 1, 2

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Salt in Popcorn.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 Taste Test 를 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 Salt 를 선택하고 **매크로 > 반응 표면**을 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 표면 " 옆의 표시 아이콘을 클릭하여 보고서를 엽니다 .

그림 11.11 Salt in Popcorn.jmp 에 대한 순서형 로지스틱 적합



이 보고서에서는 2차 모형이 반응 확률을 어떻게 적합시키는지 보여 줍니다. 곡선은 변하는 로지스틱 곡선이 아니라 동일한 점에서 최소값을 이루는 각 곡선이 쌓여 U자형이 됩니다. 임계값은 Mean(*X*) – 0.5 * (b1/b2)입니다. 여기서 b1은 선형 계수이고 b2는 2차 계수입니다. 이 계산식은 중심화된 *X*를 대상으로 합니다. "모수 추정값" 테이블에서 최적의 소금 양은 1.5 - 0.5 * (0.5637/1.3499) = 1.29 티스푼임을 계산할 수 있습니다.

특정 소금 양에서 각 곡선 사이의 세로 거리는 해당 소금 양에 대한 다섯 가지 반응 수준의 각 확률을 측정합니다. 가장 높은 반응 수준에 대한 확률은 상단 곡선에서 그림 직사각형 상단까지의

거리입니다. 이 거리는 소금 양이 약 1.3 티스푼일 때 최대값에 도달합니다. 모든 곡선은 동일한 임계점을 공유합니다.

Salt 및 Salt*Salt에 대한 모수 추정값은 임계값을 찾는 데 사용되는 계수입니다. 임계값은 로지스틱 그림에 최소값으로 표시되지만 가장 높은 반응의 확률을 최대화한다는 의미에서 최대값이 됩니다. 보고서의 "해" 부분은 그림 11.11의 "반응 표면" 아래에 표시되어 있으며, 여기서 **임계값** 아래에 1.29가 표시됩니다.

예 : 여러 열의 개수 쌓기

빈도(개수) 데이터가 데이터 테이블의 여러 열에 나열된 경우 데이터를 로지스틱 회귀에 필요한 형식으로 변환해야 합니다. 예를 들어 Ingots2.jmp 데이터 테이블(그림 11.12)에는 Nready 및 Nnotready 열이 있습니다. 이러한 열은 Heat 및 Soak 값의 각 조합에 대해 시작 준비가 된 주괴와 준비가 되지 않은 주괴의 수를 제공합니다.

그림 11.12 Ingots2.jmp 샘플 데이터 테이블

	Heat	Soak	Nready	Nnotready	Ntotal	P	Loss
1	7	1	10	0	10	0.5000	6.9315
2	7	1.7	17	0	17	0.5000	11.7835
3	7	2.2	7	0	7	0.5000	4.8520
4	7	2.8	12	0	12	0.5000	8.3178
5	7	4	9	0	9	0.5000	6.2383
6	14	1	31	0	31	0.5000	21.4876
7	14	1.7	43	0	43	0.5000	29.8053
8	14	2.2	31	2	33	0.5000	22.8739
9	14	2.8	31	0	31	0.5000	21.4876
10	14	4	19	0	19	0.5000	13.1698
11	27	1	55	1	56	0.5000	38.8162
12	27	1.7	40	4	44	0.5000	30.4985
13	27	2.2	21	0	21	0.5000	14.5561
14	27	2.8	21	1	22	0.5000	15.2492
15	27	4	15	1	16	0.5000	11.0904

로지스틱 회귀 모형을 적합시키기 전에 다음 단계에 따라 Nready 및 Nnotready 열을 단일 열에 쌓으십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ingots2.jmp 를 엽니다 .
2. **테이블 > 쌓기**를 선택합니다 .
3. " 열 선택 " 목록에서 Nready 와 NNotReady 를 선택하고 **열 쌓기**를 클릭합니다 .
4. **확인**을 클릭합니다 .

   이렇게 하면 그림 11.13 의 새 테이블이 생성됩니다 . **라벨**은 반응 (Y) 열이고 **데이터**는 빈도 열입니다 .

이 쌓인 데이터 테이블은 "명목형 로지스틱 회귀의 예"(442 페이지)에 사용된 Ingots.jmp 샘플 데이터 테이블과 동등합니다.

그림 11.13 쌓인 데이터 테이블

	Heat	Soak	Ntotal	P	Loss	Label	Data
1	7	1	10	0.5000	6.9315	Nready	10
2	7	1	10	0.5000	6.9315	Nnotready	0
3	7	1.7	17	0.5000	11.7835	Nready	17
4	7	1.7	17	0.5000	11.7835	Nnotready	0
5	7	2.2	7	0.5000	4.8520	Nready	7
6	7	2.2	7	0.5000	4.8520	Nnotready	0
7	7	2.8	12	0.5000	8.3178	Nready	12
8	7	2.8	12	0.5000	8.3178	Nnotready	0
9	7	4	9	0.5000	6.2383	Nready	9
10	7	4	9	0.5000	6.2383	Nnotready	0
11	14	1	31	0.5000	21.4876	Nready	31
12	14	1	31	0.5000	21.4876	Nnotready	0
13	14	1.7	43	0.5000	29.8053	Nready	43
14	14	1.7	43	0.5000	29.8053	Nnotready	0
15	14	2.2	33	0.5000	22.8739	Nready	31
16	14	2.2	33	0.5000	22.8739	Nnotready	2
17	14	2.8	31	0.5000	21.4876	Nready	31
18	14	2.8	31	0.5000	21.4876	Nnotready	0

명목형 및 순서형 로지스틱 분석법에 대한 통계 상세 정보

- "로지스틱 회귀 모형"
- "승산비"
- "통계적 검정의 관계"

로지스틱 회귀 모형

로지스틱 회귀는 명목형 Y 반응을 X 항의 선형 모형에 적합시킵니다. 더 정확히 말하면 로지스틱 함수를 사용하여 반응 수준에 대한 확률을 적합시킵니다. 두 개의 반응 수준에 대한 함수는 다음과 같습니다.

$P(Y = r_1) = (1 + e^{-Xb})^{-1}$ 여기서, r_1 은 첫 번째 반응 수준입니다.

또는 다음과 같이 지정될 수도 있습니다.

$\log\left(\frac{P(Y = r_1)}{P(Y = r_2)}\right) = Xb$ 여기서, r_1 과 r_2 는 각각 두 반응 수준입니다.

참고 : Y 가 이항 변수이고 명목형 모델링 유형인 경우 모형 적합 창에서 목표 수준을 설정하여 확률을 모델링할 수준을 지정할 수 있습니다 . 이 섹션에서는 목표 수준이 r_1 로 지정됩니다 .

$r(r > 2)$ 개의 명목형 반응 수준에 대해 모형은 다음과 같은 형식의 $r - 1$ 개 선형 모형 모수로 정의됩니다 .

$$\log\left(\frac{P(Y = j)}{P(Y = r)}\right) = Xb_j$$

최대 가능도 적합은 모형에서 발생한 반응에 기인한 결합 확률을 최대화하기 위해 β 가 선택됨을 의미합니다 . 이 적합은 음의 로그 가능도 (–LogLikelihood) 를 최소화하는 것과 동등합니다 .

$$\text{Loss} = -\text{logLikelihood} = \sum_{i=1}^{n} -\log(\text{Prob}(i\text{번째 행에 } y_j\text{번째 반응이 있음}))$$

승산비

두 개의 반응 수준에 대한 로지스틱 회귀 모형은 다음과 같이 지정됩니다 .

$$\log\left(\frac{\text{Prob}(Y = r_1)}{\text{Prob}(Y = r_2)}\right) = Xb$$ 여기서 , r_1 과 r_2 는 두 반응 수준입니다 .

따라서 **승산**은 다음과 같이 정의됩니다 .

$$\frac{\text{Prob}(Y = r_1)}{\text{Prob}(Y = r_2)} = \exp(X\beta) = \exp(\beta_0) \cdot \exp(\beta_1 X_1) \cdots \exp(\beta_i X_i)$$

$\exp(\beta_i(X_i + 1)) = \exp(\beta_i X_i)\exp(\beta_i)$ 입니다 . 즉 , X_i 가 단위 양만큼 변경되면 승산에 $\exp(\beta_i)$ 를 곱하고 이를 **단위 승산비**라고 합니다 . X_i 가 전체 범위에 대해 변경되면 승산에 $\exp((X_{\text{high}} - X_{\text{low}})\beta_i)$ 를 곱하고 이를 **범위 승산비**라고 합니다 . 이항 반응의 경우 반전된 반응 수준의 로그 승산비는 모수의 부호만 변경됩니다 . 따라서 보고된 값의 역수를 첫 번째 반응 수준 대신 마지막 수준에 초점을 맞출 수 있습니다 .

2 수준 명목형 효과는 첫 번째 수준과 두 번째 수준에 대해 각각 1 과 -1 로 코드화되므로 범위 승산비 또는 역수가 중요합니다 .

승산비 옵션을 선택할 때 모수 추정값의 신뢰 구간이 계산되지 않은 경우 승산비 옵션을 선택하면 해당 승산비에 대해 Wald 기반 신뢰 구간이 생성됩니다 .

다음 조건이 모두 충족될 경우 승산비 옵션을 선택하면 해당 승산비에 대해 프로파일 가능도 기반 신뢰 구간이 생성됩니다 .

- 가능도비 검정 옵션이 선택되었습니다 .
- 신뢰 구간 옵션이 선택되었습니다 .
- 각 반응의 모수 수가 8 개 미만입니다 .

- 행 수가 1000 개 미만입니다 .

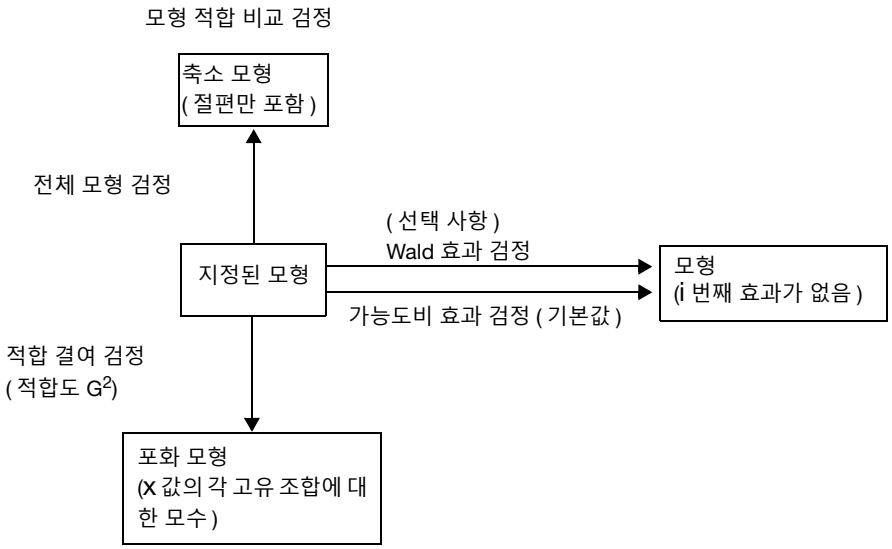
승산비에 대한 신뢰 구간을 계산하는 데 사용된 방법이 승산비 보고서 하단에 표시됩니다.

통계적 검정의 관계

로지스틱 회귀 보고서의 모든 통계적 검정은 그림 11.14 에 설명된 대로 지정된 모형의 적합을 부분집합 또는 상위집합 모형과 비교합니다 . 검정이 유의성을 나타내면 고차 모형이 타당성을 갖습니다 .

- 전체 모형 검정 : 지정된 모형이 절편을 제외한 다른 효과가 없는 축소 모형보다 유의하게 나은 경우
- 적합 결여 검정 : 포화 모형이 지정된 모형보다 유의하게 나은 경우
- 효과 검정 : 지정된 모형이 효과가 제공되지 않은 모형보다 유의하게 나은 경우

그림 11.14 통계적 검정의 관계



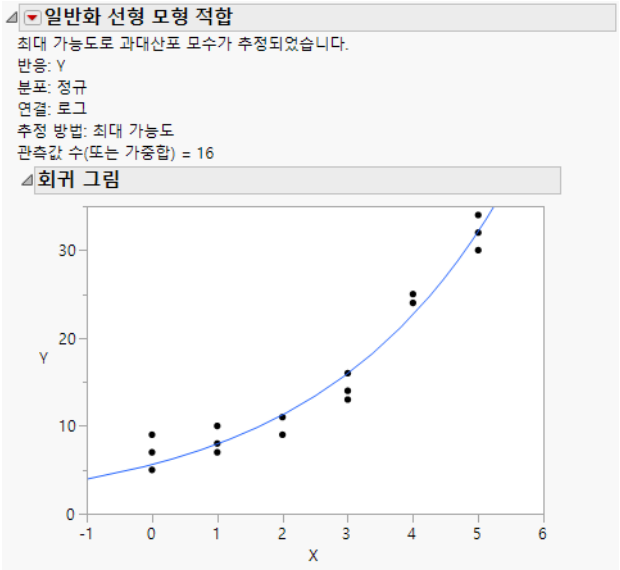
12 장

일반화 선형 모형
비정규 반응 분포에 대한 모형 적합

일반화 선형 모형은 기존 선형 모형의 일반적인 요구 사항에 맞지 않는 반응을 적합시키는 통합된 방법을 제공합니다. 예를 들어 빈도 수는 일반화 선형 모형을 사용하여 Poisson 분포 및 적합을 수행하는 경우가 많습니다.

모형 적합 플랫폼의 일반화 선형 모형 분석법을 사용하면 이항, 정규, Poisson 또는 지수 분포를 사용하여 반응에 대한 일반화 선형 모형을 적합시킬 수 있습니다. 이 플랫폼은 기존 선형 모형에 제공되는 것과 유사한 보고서를 제공합니다. 또한 Firth 수정을 사용하여 로지스틱 회귀 모형의 분리도 수용합니다.

그림 12.1 일반화 선형 모형 적합의 예



목차

일반화 선형 모형 분석법 개요 469
일반화 선형 모형의 예 469
일반화 선형 모형 분석법 시작 472
일반화 선형 모형 적합 보고서 473
    전체 모형 검정 474
일반화 선형 모형 적합 보고서 옵션 475
일반화 선형 모형 분석법의 추가 예 477
    대비를 사용하여 변수 수준의 차이 비교 477
    오프셋을 사용하는 Poisson 회귀 478
    로그 연결을 사용하는 정규 회귀 480
일반화 선형 모형 분석법에 대한 통계 상세 정보 482
    모형 선택 및 이탈도 484

일반화 선형 모형 분석법 개요

기존 선형 모형은 통계 데이터 분석에서 광범위하게 사용됩니다. 그러나 기존 선형 모형의 가정을 위반하는 상황이 있습니다. 이러한 경우 기존 선형 모형은 적절하지 않습니다. 기존 선형 모형은 반응이 연속형이고 모든 관측값에 대해 일정한 분산으로 정규 분포를 따른다고 가정합니다. 이러한 가정이 적절하지 않을 수 있습니다. 예를 들어 개수를 모델링하려는 경우 또는 반응이 증가할수록 관측된 반응의 분산이 증가하는 경우 이 가정은 적절하지 않습니다. 반응 평균이 0과 1 사이의 비율과 같은 특정 범위로 제한되는 경우도 기존 선형 모형의 가정을 위반한다고 볼 수 있습니다.

이와 같이 더 넓은 범위의 데이터 분석 문제에 해당하는 경우 일반화 선형 모형을 적용할 수 있습니다. 일반화 선형 모형은 기존 선형 모형의 확장입니다. 일반화 선형 모형은 선형 성분, 연결 함수 및 분산 함수로 구성됩니다. 연결 함수 $g(\mu_i) = x'_i\beta$는 Y_i의 기대값이 선형 예측 변수와 어떤 관계가 있는지 설명하는 단조 미분 가능 함수입니다. 일반화 선형 회귀의 한 예로 Poisson 회귀가 있으며 여기서 $\log(\mu_i)$는 연결 함수입니다. 모형 적합 플랫폼의 일반화 선형 모형 분석법을 사용하여 가능한 일반화 선형 회귀 모형의 전체 목록은 "일반화 선형 모형 분석법에 대한 통계 상세 정보"(482 페이지)에서 확인하십시오.

적합된 일반화 선형 모형은 기존 선형 모형과 동일한 통계량을 사용하여 요약하고 평가할 수 있습니다. 모형 적합 플랫폼은 일반화 선형 모형에 대한 모수 추정값, 표준 오차, 적합도 통계량, 신뢰 구간 및 가설 검정을 제공합니다. 정확 분포 이론이 일반화 선형 모형에 대해 제공되지 않거나 실현 가능하지 않은 경우도 있다는 점에 유의하십시오. 따라서 일부 추론 절차는 점근적 결과를 기반으로 합니다.

일반화 선형 모형 적합의 중요한 특징은 모형의 설명 변수를 선택하는 것입니다. 적합도 통계량의 변화는 대개 설명 변수의 부분집합이 특정 모형에 미치는 기여도를 평가하는 데 사용됩니다. **이탈도**는 도달 가능한 최대 로그 가능도 함수 값과 회귀 모수의 최대 가능도 추정값에서의 로그 가능도 함수 값 차이의 두 배로 정의됩니다. 이탈도는 주로 적합도의 측도로 사용됩니다. 도달 가능한 최대 로그 가능도는 모든 관측값에 대한 모수가 포함된 모형을 사용하여 얻을 수 있습니다.

JMP PRO 일반화 선형 모델링의 변수 선택 및 벌점 방법의 경우 JMP Pro에서 모형 적합 플랫폼의 일반화 회귀 분석법을 사용할 수 있습니다. 자세한 내용은 "일반화 회귀 모형" 장(263 페이지)에서 확인하십시오.

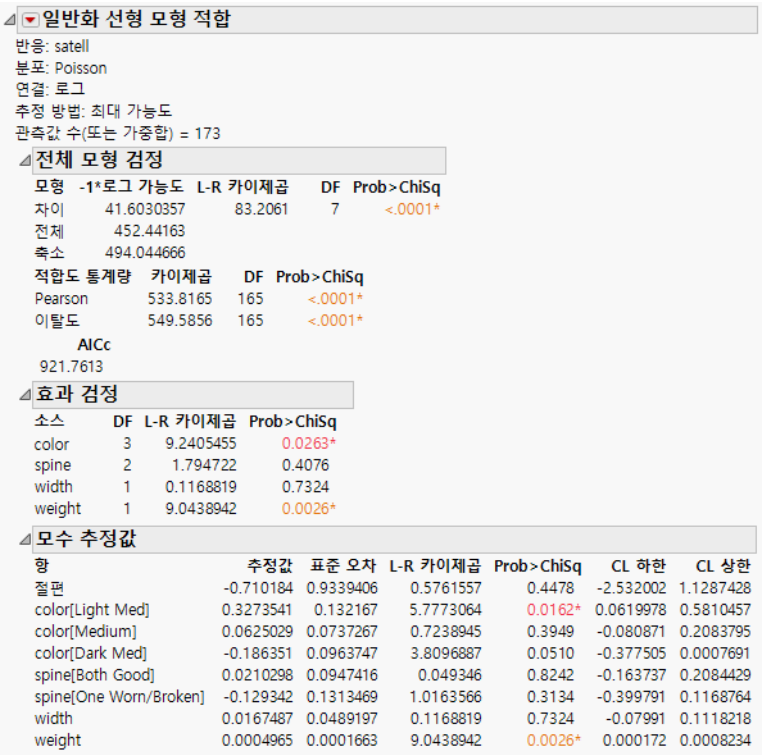
일반화 선형 모형의 예

이 예에서는 Poisson 회귀를 사용하여 투구게 둥지 연구의 개수 데이터를 모델링합니다. 각 암게의 둥지에 수게가 있습니다. 이 연구에서는 위성 수게라고 하는 다른 수게가 근처에 있는지 여부를 조사했습니다. CrabSatellites.jmp 데이터 테이블에는 위성 수게의 수를 나열하는 반응 변수와 암게의 색상, 가시 상태, 무게 및 등딱지 폭 길이를 설명하는 변수가 포함되어 있습니다. 위성 수게의 수와 암게를 설명하는 변수 사이의 관계에 관심이 있습니다.

Poisson 회귀 모형을 적합시키려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 CrabSatellites.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. satell 을 선택하고 **Y** 를 클릭합니다 .
4. color, spine, width 및 weight 를 선택하고 **추가**를 클릭합니다 .
5. " 분석법 " 목록에서 **일반화 선형 모형**을 선택합니다 .
6. " 분포 " 목록에서 **Poisson** 을 선택합니다 .
   " 연결 함수 " 목록에서 **로그**가 자동으로 선택되어야 합니다 .
7. **실행**을 클릭합니다 .

그림 12.2 위성 게 결과



일반화 선형 모형 적합

반응: satell
분포: Poisson
연결: 로그
추정 방법: 최대 가능도
관측값 수(또는 가중합) = 173

전체 모형 검정

모형	-1*로그 가능도	L-R 카이제곱	DF	Prob>ChiSq
차이	41.6030357	83.2061	7	<.0001*
전체	452.44163			
축소	494.044666			

적합도 통계량	카이제곱	DF	Prob>ChiSq
Pearson	533.8165	165	<.0001*
이탈도	549.5856	165	<.0001*

AICc
921.7613

효과 검정

소스	DF	L-R 카이제곱	Prob>ChiSq
color	3	9.2405455	0.0263*
spine	2	1.794722	0.4076
width	1	0.1168819	0.7324
weight	1	9.0438942	0.0026*

모수 추정값

항	추정값	표준 오차	L-R 카이제곱	Prob>ChiSq	CL 하한	CL 상한
절편	-0.710184	0.9339406	0.5761557	0.4478	-2.532002	1.1287428
color[Light Med]	0.3273541	0.132167	5.7773064	0.0162*	0.0619978	0.5810457
color[Medium]	0.0625029	0.0737267	0.7238945	0.3949	-0.080871	0.2083795
color[Dark Med]	-0.186351	0.0963747	3.8096887	0.0510	-0.377505	0.0007691
spine[Both Good]	0.0210298	0.0947416	0.049346	0.8242	-0.163737	0.2084429
spine[One Worn/Broken]	-0.129342	0.1313469	1.0163566	0.3134	-0.399791	0.1168764
width	0.0167487	0.0489197	0.1168819	0.7324	-0.07991	0.1118218
weight	0.0004965	0.0001663	9.0438942	0.0026*	0.000172	0.0008234

" 전체 모형 검정 " 에서는 전체 모형과 축소 모형에 대한 로그 가능도의 차이가 41.6 임을 보여 줍니다 . Prob>ChiSq 값은 전체 모형의 F 검정에 대한 p 값과 동등합니다 . p 값이 0.0001 보다 작으면 모형이 전체적으로 유의하다는 것을 나타냅니다 . 이 보고서에는 수정된 AICc(Akaike 정보 기준) 값 921.7613 도 포함되어 있습니다 . 이 값을 다른 모형과 비교하여 데이터에 가장 적합한 모형을 결정할 수 있습니다 . AICc 값이 작을수록 모형이 더 적합하다는 것을 나타냅니다 .

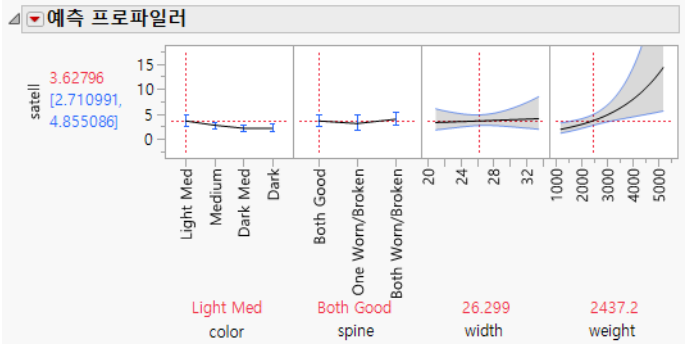
" 효과 검정 " 보고서에서는 weight 가 모형에서 유의한 요인임을 보여 줍니다 . weight 는 연속형 변수이므로 weight 의 p 값 0.0026 이 " 모수 추정값 " 테이블에서도 동일합니다 .

" 효과 검정 " 보고서를 통해 범주형 변수 color 도 유의하다는 것을 알 수 있습니다 . color 에는 Light Med, Medium, Dark Med 및 Dark 의 네 가지 수준이 있습니다 . 처음 세 수준의 모수 추정값이 " 모수 추정값 " 테이블에 나옵니다 . Dark 의 모수 추정값은 처음 세 수준에 대한 모수 추정값 합계의 음수입니다 .

예측 프로파일러를 사용하여 모형을 자세히 탐색할 수 있습니다 .

8. " 일반화 선형 모형 적합 " 옆의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다 .

그림 12.3 satell 에 대한 예측 프로파일러

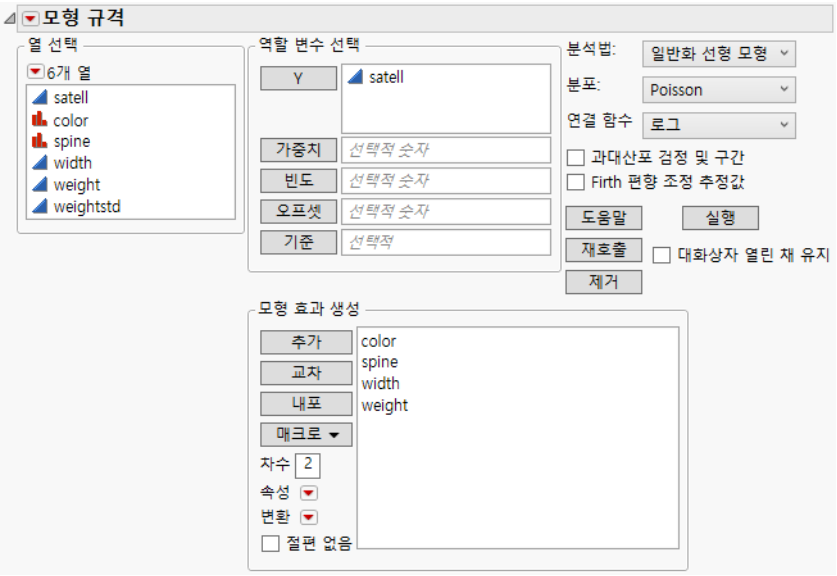


weight 의 신뢰 대역은 weight 값이 클 때보다 weight 값이 작을 때 모형의 변동성이 적음을 나타냅니다 . 프로파일러를 사용하면 범주형 변수의 수준을 쉽게 탐색할 수 있습니다 . 프로파일러를 통해 게의 색상이 연한 색에서 진한 색으로 변함에 따라 예측된 위성 게의 수가 줄어드는 것을 확인할 수 있습니다 .

일반화 선형 모형 분석법 시작

분석 > 모형 적합을 선택하고 **Y** 에 대해 하나 이상의 열을 입력한 후 **분석법** 메뉴에서 **일반화 선형 모형**을 선택하여 일반화 선형 모형 분석법을 시작합니다 .

그림 12.4 일반화 선형 모형이 선택된 모형 적합 시작 창



"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 "모형 규격" 장(29 페이지) 에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 일반화 선형 모형 분석법에 대한 정보를 제공합니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격" 의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

팁: 모형 적합 플랫폼의 일반화 선형 모형 분석법에서는 "절편 없음" 옵션을 사용할 수 없습니다.

" 분석법 " 에서 " 일반화 선형 모형 " 을 선택하면 모형 적합 시작 창이 변경되어 추가 옵션을 포함합니다 . 일반화 선형 모형 분석법에서는 다음 추가 옵션을 사용할 수 있습니다 .

분포 반응 변수에 대한 확률 분포를 지정합니다 .

연결 함수 선형 모형을 반응 변수와 관련짓는 연결 함수를 지정합니다 .

과대산포 검정 및 구간 과대산포 모수가 모형에 포함되도록 지정합니다 . 과대산포는 반응 분산이 반응 분포의 이론적 분산에 의해 예상되는 것보다 클 때 발생합니다 . Poisson 및 이항 반응 모형에서 과대산포가 발생할 수 있습니다 . McCullagh and Nelder 연구 자료 (1989) 에 따르면 과대산포는 실제로 흔히 일어나는 일입니다 .

참고 : 이 옵션은 " 전체 모형 검정 " 보고서의 " 적합도 통계량 " 테이블에 " 과대산포 " 열을 추가합니다 .

Firth 편향 조정 추정값 Firth 편향 조정 방법을 사용하여 모형을 적합시키도록 지정합니다. 이 최대 가능도 기반 방법은 편향 수정을 사용하지 않는 최대 가능도 기반 모형보다 더 나은 추정값과 검정을 생성하는 것으로 나타났습니다 . 또한 편향 수정된 MLE 는 로지스틱 유형의 모형에서 발생하기 쉬운 분리 문제도 개선합니다 . 로지스틱 회귀의 분리 문제에 대한 자세한 내용은 Firth 연구 자료 (1993) 및 Heinze and Schemper 연구 자료 (2002) 에서 확인하십시오 .

오프셋 (역할 변수 선택 버튼으로 표시됨) 오프셋 변수를 지정합니다 . 오프셋 변수는 모수가 1.0 으로 고정된 회귀 공변량처럼 처리되는 변수입니다 . 오프셋 변수는 로그 연결을 사용하는 Poisson 회귀에서 평균의 모델링을 척도화하는 데 가장 자주 사용됩니다 .

이항 반응 분포의 반응 지정

" 분포 " 에서 " 이항 " 을 선택하면 다음 방법 중 하나를 사용하여 반응 변수를 지정해야 합니다 .

- 데이터가 사건 빈도로 요약되지 않은 경우 단일 이진 열을 반응으로 지정합니다 . 반응 열은 명목형이어야 합니다 .
- 데이터가 사건 빈도로 요약된 경우 단일 이진 열을 반응으로 지정하고 " 빈도 " 역할에 빈도 변수를 지정합니다 . 반응 열은 명목형이어야 하고 빈도 변수는 각 반응 수준의 수를 포함합니다 .
- 데이터가 사건 빈도 및 시행 횟수로 요약된 경우 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 지정합니다 . 성공 횟수 대신 실패 횟수를 지정할 수도 있습니다 .

일반화 선형 모형 적합 보고서

기본적으로 " 일반화 선형 모형 적합 " 보고서에는 모형 규격에 대한 상세 정보와 다음 보고서가 포함됩니다 .

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 표시됨) 모형 항이 충족하는 선형 함수가 포함된 보고서를 표시합니다 .

회귀 그림 (연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개인 경우에만 표시됨) 세로 축에 반응을 나타내고 가로 축에 연속형 예측 변수를 나타내는 그림을 표시합니다 . 점 위에 회귀선이 표시됩니다 . 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 회귀선이 있고 그림 옆에 범례가 나타납니다 .

전체 모형 검정 절편 모수를 제외한 모든 효과가 생략된 모형과 전체 모형 적합을 비교하는 검정을 표시합니다 . 이 보고서에는 적합도 통계량 및 수정된 AICc(Akaike 정보 기준) 값도 포함됩니다 . 자세한 내용은 " 전체 모형 검정 "(474 페이지) 에서 확인하십시오 .

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 보고서입니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 효과 요약 보고서 "(168 페이지) 에서 확인하십시오 .

효과 검정 효과 검정은 개별 효과에 대한 모든 모수가 0 이라는 결합 검정입니다 . 연속형 효과와 마찬가지로 효과에 모수가 하나만 있는 경우 검정은 모수 추정값 테이블의 검정과 동일합니다.

참고 : Firth 조정이 사용되는 경우에도 효과 검정은 벌점 없는 가능도 함수를 기반으로 합니다 .

모수 추정값 모수 추정값 , 표준 오차 , 관련된 가설 검정 및 신뢰 한계를 표시합니다 . 단순 연속형 효과에는 모수가 하나만 있습니다 . 복합 분류 효과가 포함된 모형에는 각 예상 자유도에 대한 모수가 있습니다 .

참고 : 관측값 수가 1,000 을 초과하면 Wald 기반 신뢰 구간이 표시됩니다 . 그렇지 않으면 프로파일 - 가능도 신뢰 구간이 표시됩니다 .

스튜던트화 이탈도 잔차 대 예측값 세로 축에 스튜던트화 이탈도 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시합니다 .

전체 모형 검정

" 전체 모형 검정 " 테이블에는 절편 모수를 제외한 모든 회귀 모수가 생략된 모형과 전체 모형 적합을 비교하는 검정이 표시됩니다. 또한 모형 적합성을 평가하기 위한 두 가지 적합도 통계량과 AICc 값도 포함되어 있습니다.

" 전체 모형 검정 " 테이블에는 다음과 같은 통계량이 표시됩니다 .

모형 모형 라벨을 나열합니다 .

차이 **전체** 모형과 **축소** 모형 간의 차이입니다 . 이 모형은 적합에 대한 전체 회귀변수의 유의성을 측정하는 데 사용됩니다 .

전체 절편 및 모든 효과를 포함하는 완전 모형입니다 .

축소 절편 모수만 포함하는 모형입니다 .

-1* 로그 가능도 각 모형에 대한 음의 로그 가능도입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

L-R 카이제곱 모든 회귀 모수가 0 이라는 가설에 대한 가능도비 카이제곱 검정 통계량입니다 . 이 검정 통계량은 적합 모형과 절편만 있는 축소 모형 간의 음의 로그 가능도 차이의 두 배로 계산됩니다 .

DF 전체 모형과 축소 모형 간의 차이에 대한 DF(자유도) 입니다 .

Prob>ChiSq 지정된 모형이 절편만 포함된 모형보다 더 잘 적합되지 않을 경우 카이제곱 값이 더 클 확률입니다 .

적합도 통계량 두 가지 적합도 통계량인 Pearson 과 이탈도를 나열합니다 .

카이제곱 각 적합도 통계량에 대한 카이제곱 검정 통계량입니다 .

DF 각 적합도 통계량에 대한 자유도입니다 .

Prob>ChiSq 각 적합도 통계량에 대한 *p* 값입니다 .

과대산포 (시작 창에서 " 과대산포 검정 및 구간 " 옵션을 선택한 경우에만 표시됨) 과대산포 모수의 추정값입니다 . 자세한 내용은 " 일반화 선형 모형 분석법에 대한 통계 상세 정보 "(482 페이지) 에서 확인하십시오 .

AICc 수정된 Akaike 정보 기준입니다 . 자세한 내용은 " 통계 상세 정보 " 부록의 " 가능도 , AICc 및 BIC"(514 페이지) 에서 확인하십시오 .

일반화 선형 모형 적합 보고서 옵션

" 일반화 선형 모형 적합 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

사용자 검정 사용자 가설을 검정할 수 있습니다 . 사용자 검정에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 사용자 검정 "(115 페이지) 에서 확인하십시오 .

대비 변수 내의 수준에 대한 차이를 검정할 수 있습니다 . 대비에 공변량이 포함된 경우 대비를 검정할 공변량 값을 지정할 수 있습니다 . " 대비 " 옵션의 예는 " 대비를 사용하여 변수 수준의 차이 비교 "(477 페이지) 에서 확인하십시오 .

역추정 예측 (연속형 *X* 변수에만 사용 가능) *Y* 및 다른 *X* 변수에 대한 특정 값이 주어졌을 때 *X* 값을 예측할 수 있습니다 . " 역추정 예측 " 옵션에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 역추정 예측 "(131 페이지) 에서 확인하십시오 .

추정값 공분산 모형의 모든 효과에 대한 공분산 행렬을 표시하거나 숨깁니다 . 모수 추정량의 추정된 공분산 행렬은 다음과 같이 정의됩니다 .

$$\Sigma = -\mathbf{H}^{-1}$$

여기서 **H** 는 마지막 반복의 모수 추정값을 사용하여 평가된 헤시안 (또는 2 차 도함수) 행렬입니다 . 산포 모수는 추정되었는지 아니면 지정되었는지에 관계없이 **H** 에 통합되고 , 별칭이 지정된 모수에 해당하는 행과 열은 Σ 에 포함되지 않습니다 .

추정값 상관계수 모형의 모든 효과에 대한 상관 행렬을 표시하거나 숨깁니다 . 상관 행렬은 정규화된 공분산 행렬입니다 . Σ 의 각 σ_{ij} 요소에 대해 상관 행렬의 해당 요소는 $\sigma_{ij}/\sigma_i\sigma_j$ 이며 , 여기서 $\sigma_i = \sqrt{\sigma_{ii}}$ 입니다 .

프로파일러 다음과 같은 프로파일러 하위 메뉴를 표시합니다 .

프로파일러 각 X 변수에 대한 예측 트레이스를 검토할 수 있는 예측 프로파일러를 표시하거나 숨깁니다 . 예측 프로파일러에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 프로파일러 "(151 페이지) 에서 확인하십시오 .

등고선 프로파일러 대화식 등고선 프로파일러를 표시하거나 숨깁니다 . 등고선 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

표면 프로파일러 3D 표면 프로파일러를 표시하거나 숨깁니다 . 표면 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

진단 그림 잔차 , 예측값 및 실제값 그림을 위한 하위 메뉴를 표시합니다 . 이러한 그림을 통해 이상치를 검색하고 모형의 적합성을 판단할 수 있습니다 . 이탈도에 대한 자세한 내용은 " 모형 선택 및 이탈도 "(484 페이지) 에서 확인하십시오 . 다음과 같은 그림을 사용할 수 있습니다 .

스튜던트화 이탈도 잔차 대 예측값 세로 축에 스튜던트화 이탈도 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

스튜던트화 Pearson 잔차 대 예측값 세로 축에 스튜던트화 Pearson 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

이탈도 잔차 대 예측값 세로 축에 이탈도 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

Pearson 잔차 대 예측값 세로 축에 Pearson 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

회귀 그림 (연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개인 경우에만 사용 가능) 세로 축에 반응을 나타내고 가로 축에 연속형 예측 변수를 나타내는 그림을 표시하거나 숨깁니다. 점 위에 회귀선이 표시됩니다. 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 회귀선이 있고 그림 옆에 범례가 나타납니다.

선형 예측 변수 그림 (연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개인 경우에만 사용 가능) 세로 축에 역 연결 함수에 의해 변환된 반응을 나타내고 가로 축에 연속형 예측 변수를 나타내는 그림을 표시하거나 숨깁니다. 점 위에 변환된 회귀선이 표시됩니다. 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 변환된 회귀선이 있고 그림 옆에 범례가 나타납니다.

열 저장 특정 통계량을 현재 데이터 테이블의 새 열로 저장할 수 있는 옵션이 포함된 하위 메뉴를 표시합니다. 잔차 계산식에 대한 자세한 내용은 " 잔차 계산식 "(485 페이지) 에서 확인하십시오.

예측 계산식 현재 데이터 테이블에 모형을 예측하는 계산식 열을 생성합니다.

예측값 현재 데이터 테이블에 모형에서 예측한 값을 포함하는 열을 생성합니다.

평균 신뢰 구간 현재 데이터 테이블에 모형에 대한 예측 방정식의 95% 신뢰 한계를 포함하는 열을 생성합니다. 이러한 신뢰 한계는 모수 추정값의 변동을 반영합니다.

참고 :" 모형 적합 " 창의 " 모형 규격 " 빨간색 삼각형 메뉴에서 " 유의 수준 설정 " 을 선택하여 신뢰 한계의 α 수준을 변경할 수 있습니다.

개별 신뢰 한계 저장 현재 데이터 테이블에 모형에 대한 지정된 개별 값의 95% 신뢰 한계를 포함하는 열을 생성합니다. 이러한 신뢰 한계는 오차의 변동과 모수 추정값의 변동을 반영합니다.

참고 :" 모형 적합 " 창의 " 모형 규격 " 빨간색 삼각형 메뉴에서 " 유의 수준 설정 " 을 선택하여 신뢰 한계의 α 수준을 변경할 수 있습니다.

이탈도 잔차 현재 데이터 테이블에 이탈도 잔차를 포함하는 열을 생성합니다.

Pearson 잔차 현재 데이터 테이블에 Pearson 잔차를 포함하는 열을 생성합니다.

스튜던트화 이탈도 잔차 현재 데이터 테이블에 스튜던트화 이탈도 잔차를 포함하는 열을 생성합니다.

스튜던트화 Pearson 잔차 현재 데이터 테이블에 스튜던트화 Pearson 잔차를 포함하는 열을 생성합니다.

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다.

효과 요약 모형의 효과를 대화식으로 업데이트할 수 있는 " 효과 요약 " 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 효과 요약 보고서 "(168 페이지) 에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

일반화 선형 모형 분석법의 추가 예

- " 대비를 사용하여 변수 수준의 차이 비교 "
- " 오프셋을 사용하는 Poisson 회귀 "
- " 로그 연결을 사용하는 정규 회귀 "

대비를 사용하여 변수 수준의 차이 비교

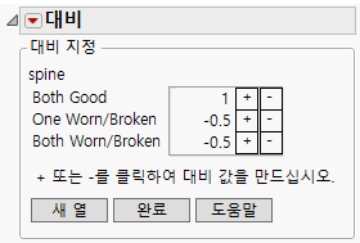
이 예에서는 " 일반화 선형 모형의 예 "(469 페이지) 의 위성 게 예를 계속합니다. 가시 상태가 좋은 암게가 유인하는 수게 (위성) 의 수가 가시가 닳거나 부러진 암게와 다른지 여부를 검정하려고 한다고 가정해 보겠습니다.

1. " 일반화 선형 모형의 예 "(469 페이지) 에 나오는 1 단계 ~ 7 단계를 완료합니다.
2. " 일반화 선형 모형 적합 " 옆의 빨간색 삼각형을 클릭하고 **대비**를 선택합니다. 그러면 " 대비 효과를 선택하십시오 " 창이 나타납니다.
3. spine(관심 변수) 을 선택하고 **확인**을 클릭합니다.

4. 가시 상태가 좋은 게를 가시가 닳거나 부러진 게와 비교하려면 Both Good 옆의 **+** 버튼을 클릭하고 One Worn/Broken 과 Both Worn/Broken 옆의 **-** 버튼을 각각 클릭합니다 .

   이렇게 하면 가시 상태가 좋은 암게를 가시가 닳거나 부러진 암게와 비교하는 대비 지정이 생성됩니다 .

그림 12.5 대비 지정 창



5. **완료**를 클릭합니다 .

그림 12.6 대비 보고서

Prob>Chisq 값 0.8242 는 0.05 보다 훨씬 크기 때문에 가시 상태에 따라 위성 게 유인에 차이가 있다는 결론을 내릴 수 없습니다.

오프셋을 사용하는 Poisson 회귀

Ship Damage.JMP 샘플 데이터 테이블은 McCullagh and Nelder 연구 자료(1989)의 데이터에서 채택되었습니다. 이 데이터 테이블에는 파도로 인해 선체 앞쪽 부분에 발생한 특정 유형의 손상 정보가 있습니다. 선체 건조 엔지니어는 선체 유형(Type), 선체 건조 연도(Yr Made) 및 선체 이용 연도(Yr Used)의 세 가지 변수와 관련된 손상 위험에 관심이 있습니다.

이 분석에서는 총 이용 개월 수의 로그인 Service 변수를 **오프셋 변수**로 사용합니다. 오프셋 변수는 모수가 1.0 으로 고정된 회귀 공변량처럼 처리되는 변수입니다. 오프셋 변수는 로그 연결을 사

용하는 Poisson 회귀에서 평균의 모델링을 척도화하는 데 가장 자주 사용됩니다. 이 예에서는 수리 횟수가 이용 개월 수에 비례할 것으로 예상되므로 log(이용 개월 수)를 사용합니다.

오프셋 변수가 어떻게 사용되는지 확인하기 위해 GLM 의 선형 성분을 η 라고 가정합니다 . 그런 다음 로그 연결 함수를 사용하여 오프셋이 포함된 평균 모형을 다음과 같이 지정합니다 .

exp[Log(이용 개월 수) + η] = [(이용 개월 수) * exp(η)].

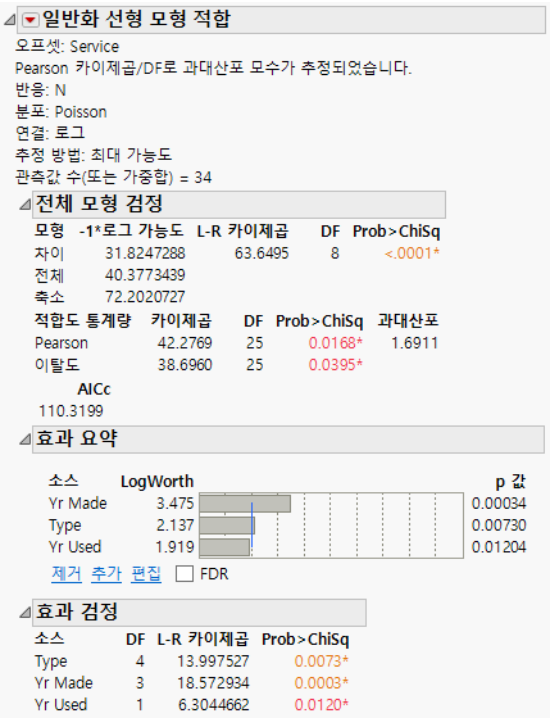
이 예를 실행하려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ship Damage.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 분석법 " 목록에서 **일반화 선형 모형**을 선택합니다 .
4. " 분포 " 목록에서 **Poisson** 을 선택합니다 .

   " 연결 함수 " 목록에서 **로그**가 자동으로 선택되어야 합니다 .
5. N 을 선택하고 **Y** 를 클릭합니다 .
6. Service 를 선택하고 **오프셋**을 클릭합니다 .
7. Type, Yr Made, Yr Used 를 선택하고 **추가**를 클릭합니다 .
8. **과대산포 검정 및 구간** 체크박스를 클릭합니다 .
9. **실행**을 클릭합니다 .

보고서에서 세 가지 효과 (Type, Yr Made, Yr Used) 가 모두 유의하다는 것을 알 수 있습니다 .

그림 12.7 오프셋을 사용하는 Poisson 모형에 대한 보고서 (일부분)

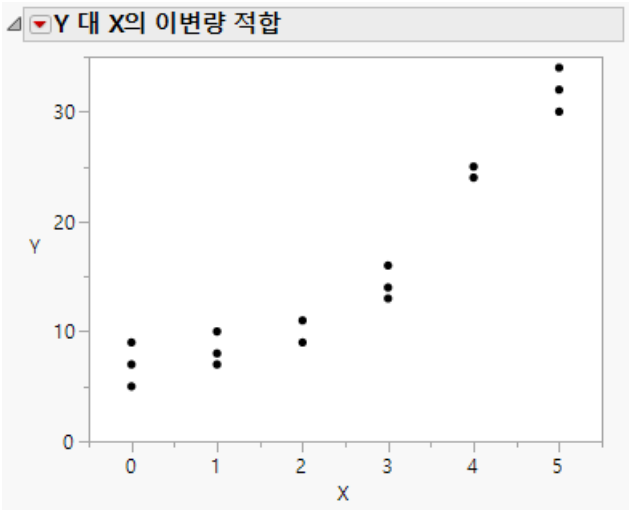

일반화 선형 모형 적합
오프셋: Service
Pearson 카이제곱/DF로 과대산포 모수가 추정되었습니다.
반응: N
분포: Poisson
연결: 로그
추정 방법: 최대 가능도
관측값 수(또는 가중합) = 34
전체 모형 검정
모형 -1*로그 가능도 L-R 카이제곱 DF Prob>ChiSq
차이 31.8247288 63.6495 8 <.0001*
전체 40.3773439
축소 72.2020727
적합도 통계량 카이제곱 DF Prob>ChiSq 과대산포
Pearson 42.2769 25 0.0168* 1.6911
이탈도 38.6960 25 0.0395*
AICc
110.3199
효과 요약
소스 LogWorth p 값
Yr Made 3.475 0.00034
Type 2.137 0.00730
Yr Used 1.919 0.01204
제거 추가 편집 FDR
효과 검정
소스 DF L-R 카이제곱 Prob>ChiSq
Type 4 13.997527 0.0073*
Yr Made 3 18.572934 0.0003*
Yr Used 1 6.3044662 0.0120*

로그 연결을 사용하는 정규 회귀

이 예에서는 Nor.jmp 데이터 테이블을 사용합니다 . 여기서 X 는 설명 변수이고 Y 는 반응 변수입니다 . 먼저 X 와 Y 사이의 관계를 탐색하여 모형 적합 플랫폼의 일반화 선형 모형 분석법에 사용할 적절한 연결 함수를 결정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nor.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Y 를 선택하고 **Y, 반응**을 클릭합니다 .
4. X 를 선택하고 **X, 요인**을 클릭한 후 **확인**을 클릭합니다 .

그림 12.8 Nor.jmp 에 대한 X 로 Y 적합 결과



Y 가 X 에 따라 비선형 방식으로 변하고 분산이 거의 일정하다는 것을 확인할 수 있습니다 . 따라서 이러한 데이터를 모델링하려면 로그 연결 함수를 사용하는 정규 분포가 적합합니다 . 즉 , $\log(\mu_i) = \mathbf{x}_i'\beta$ 이고 따라서 $\mu_i = \exp(\mathbf{x}_i'\beta)$ 입니다 .

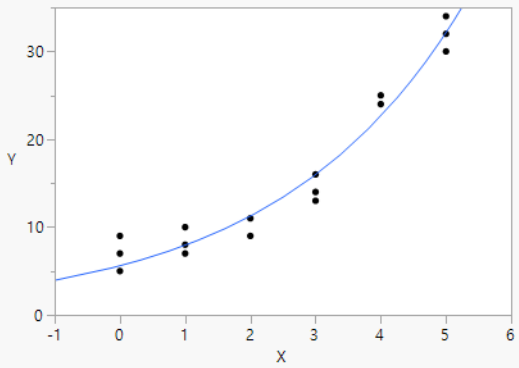
5. **분석 > 모형 적합**을 선택합니다 .
6. " 분석법 " 목록에서 **일반화 선형 모형**을 선택합니다 .
7. " 분포 " 목록에서 **정규**를 선택합니다 .
8. " 연결 함수 " 목록에서 **로그**를 선택합니다 .
9. Y 를 선택하고 **Y** 를 클릭합니다 .
10. x 를 선택하고 **추가**를 클릭합니다 .
11. **실행**을 클릭합니다 .

그림 12.9 Nor 결과

일반화 선형 모형 적합

최대 가능도로 과대산포 모수가 추정되었습니다.
반응: Y
분포: 정규
연결: 로그
추정 방법: 최대 가능도
관측값 수(또는 가중합) = 16

회귀 그림



전체 모형 검정

모형	-1*로그 가능도	L-R 카이제곱	DF	Prob>ChiSq
차이	26.5218281	53.0437	1	<.0001*
전체	32.1782768			
축소	58.7001049			

적합도 통계량	카이제곱	DF	Prob>ChiSq	과대산포
Pearson	52.3000	14	<.0001*	3.2687
이탈도	52.3000	14	<.0001*	

AICc
72.3566

효과 검정

소스	DF	L-R 카이제곱	Prob>ChiSq
X	1	53.043656	<.0001*

모수 추정값

항	추정값	표준 오차	L-R 카이제곱	Prob>ChiSq	CL 하한	CL 상한
절편	1.7213577	0.0893972	37.973777	<.0001*	1.5239809	1.8978651
X	0.3496015	0.0206493	53.043656	<.0001*	0.3081539	0.3944836

일반화 선형 모형 분석법에 대한 통계 상세 정보

일반화 선형 모형을 생성하려면 데이터에 대한 반응 변수와 설명 변수를 선택해야 합니다. 그런 다음 적절한 연결 함수와 반응 확률 분포를 선택해야 합니다. 설명 변수는 연속형 변수, 분류 변

수 및 교호작용의 조합으로 지정할 수 있습니다 . 표 12.1 에는 일반화 선형 모형의 몇 가지 일반적인 예가 나열되어 있습니다 .

표 12.1 일반화 선형 모형의 예

모형	반응 변수	분포	기본 연결 함수
기존 선형 모형	연속형	정규	항등, $g(\mu) = \mu$
로지스틱 회귀	개수 또는 이진 확률 변수	이항	로짓 , $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
로그 선형 모형의 Poisson 회귀	개수	Poisson	로그, $g(\mu) = \log(\mu)$
지수 회귀	연속형 양수	지수	$\frac{1}{\mu}$

플랫폼에서 모수 벡터의 최대 가능도 추정을 통해 일반화 선형 모형을 데이터에 적합시킵니다. 일반적으로 모수의 최대 가능도 추정에 대한 닫힌 형식의 해는 없습니다. 따라서 이 플랫폼에서는 Nelder and Wedderburn 연구 자료(1972)에서 최초로 공개된 기법을 사용하여 반복 적합 프로세스를 통해 모형의 모수를 수치적으로 추정합니다. 과대산포 모수 ϕ는 Pearson 적합도 통계량을 자유도로 나누어 추정합니다. 공분산, 표준 오차 및 신뢰 한계는 최대 가능도 추정량의 점근적 정규성을 기반으로 추정된 모수에 대해 계산됩니다.

모형 적합 플랫폼의 일반화 선형 모형 분석법에서 다양한 연결 함수 및 확률 분포를 사용할 수 있습니다. 표 12.2에는 기본 제공 연결 함수가 나열되어 있습니다.

표 12.2 기본 제공 연결 함수

연결 함수 이름	연결 함수 계산식
항등	$g(\mu) = \mu$
로짓	$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
프로빗	$g(\mu) = \Phi^{-1}(\mu)$. Φ = 표준 정규 누적 분포 함수
로그	$g(\mu) = \log(\mu)$
역수	$g(\mu) = \frac{1}{\mu}$

표 12.2 기본 제공 연결 함수 (계속)

연결 함수 이름	연결 함수 계산식
검정력	$g(\mu) = \begin{cases} \mu^{\lambda} & 단, (\lambda \neq 0) \\ \log(\mu) & 단, \lambda = 0 \end{cases}$
보 로그-로그	$g(m) = \log(-\log(1 - \mu))$

검정력 연결 함수를 선택하면 원하는 검정력을 입력할 수 있는 숫자 상자가 나타납니다.

표 12.3에는 반응 변수에 사용할 수 있는 분포와 관련된 분산 함수가 나열되어 있습니다.

표 12.3 반응 분포에 대한 분산 함수

분포	분산 함수
정규	$V(\mu) = 1$
이항	$V(\mu) = \mu(1 - \mu)$
Poisson	$V(\mu) = \mu$
지수	$V(\mu) = \mu^2$

모형 선택 및 이탈도

일반화 선형 모델링의 중요한 특징은 모형의 설명 변수를 선택하는 것입니다. 적합도 통계량의 변화는 대개 설명 변수의 부분집합이 특정 모형에 미치는 기여도를 평가하는 데 사용됩니다. **이탈도**는 도달 가능한 최대 로그 가능도와 회귀 모수의 최대 가능도 추정값에서의 로그 가능도 차이의 두 배로 정의됩니다. 이탈도는 주로 적합도의 측도로 사용됩니다. 도달 가능한 최대 로그 가능도는 모든 관측값에 대한 모수가 포함된 모형을 사용하여 얻을 수 있습니다. 표 12.4에는 반응 변수에 사용할 수 있는 각 분포에 대한 이탈도 계산식이 나열되어 있습니다.

표 12.4 반응 분포에 대한 이탈도 계산식

분포	이탈도 계산식
정규	$\sum_i w_i (y_i - \mu_i)^2$
이항	$2\sum_i w_i m_i \left[y_i \log\left(\frac{y_i}{\mu_i}\right) + (1 - y_i)\log\left(\frac{1 - y_i}{1 - \mu_i}\right) \right]$

표 12.4 반응 분포에 대한 이탈도 계산식 (계속)

분포	이탈도 계산식
Poisson	$2\sum_i w_i\left[y_i\log\left(\frac{y_i}{\mu_i}\right)-(y_i-\mu_i)\right]$
지수	$2\sum_i w_i\left[-\log\left(\frac{y_i}{\mu_i}\right)+\left(\frac{y_i-\mu_i}{\mu_i}\right)\right]$

Pearson 카이제곱 통계량은 다음과 같이 정의됩니다 .

$$X^2=\sum_i\frac{w_i(y_i-\mu_i)^2}{V(\mu_i)}$$

다음은 각 요소에 대한 설명입니다.

y_i = i 번째 반응

μ_i = 해당하는 예측 평균

$V(\mu_i)$ = 분산 함수

w_i = i 번째 관측값에 대한 알려진 가중치

참고 : 가중치를 지정하지 않으면 모든 관측값에 대해 $w_i=1$ 입니다 .

변수를 선택하는 한 가지 전략은 모형 시퀀스를 적합시키는 것입니다. 절편 항만 포함된 단순 모형으로 시작하여 각 연속 모형에 추가 설명 변수를 하나씩 포함합니다. 연속 모형 간의 이탈도 값이나 적합 로그 가능도 값의 차이로 추가 설명 변수의 중요도를 측정할 수 있습니다. 점근적 검정을 사용하면 추가 항의 통계적 유의성을 평가할 수 있습니다.

비정규 분포의 경우 역추정 예측에서 t 분포 임계값 대신 정규 임계값이 사용됩니다.

잔차 계산식

이탈도

$$r_{Di}=\sqrt{d_i}(sign(y_i-\mu_i))$$

스튜던트화 이탈도

$$r_{Di}=\frac{sign(y_i-\mu_i)\sqrt{d_i}}{\sqrt{\phi(1-h_i)}}$$

Pearson

$$r_{Pi} = \frac{(y_i - \mu_i)}{\sqrt{V(\mu_i)}}$$

스튜던트화 Pearson

$$r_{Pi} = \frac{y_i - \mu_i}{\sqrt{V(\mu_i)(1 - h_i)}}$$

다음은 각 요소에 대한 설명입니다.

$(y_i - \mu_i)$ = 원시 잔차

$\text{sign}(y_i - \mu_i) = (y_i - \mu_i)$ 가 양수인 경우 1, $(y_i - \mu_i)$ 가 음수인 경우 -1

d_i = 관측값 i 의 총 이탈도에 대한 기여도

ϕ = 산포 모수

$V(\mu_i)$ = 분산 함수

$h_i = W_e^{(1/2)}X(X'W_eX)^{-1}X'W_e^{(1/2)}$ 행렬의 i번째 주대각 요소. 여기서 W_e는 기대 정보 행렬을 계산하는 데 사용되는 가중치 행렬입니다 .

잔차 및 일반화 선형 모형에 대한 자세한 내용은 SAS Institute Inc. (2020a).

부록 A

통계 상세 정보
선형 모형 적합

이 부록에서는 다양한 유형의 반응 모형, 요인, 설계 코딩 및 파라미터화에 대해 설명합니다. 또한 본문에 설명된 방법에 대한 기타 여러 상세 정보가 포함되어 있습니다.

JMP 시스템은 선형 모형을 연속형, 순서형 및 명목형이라는 라벨이 지정된 세 가지 유형의 반응 모형에 적합시킵니다. 요인에 대한 많은 상세 정보는 다른 반응 모형에서도 동일하지만 JMP는 연속형 반응에 대해서만 그래픽 및 주변 프로파일을 지원합니다. 순서형 및 명목형 반응에는 이러한 기능이 지원되지 않습니다.

효과 적합 및 가설 검정 생성을 위해 컴퓨터 프로그램마다 서로 다른 설계 행렬 코딩을 사용하고 결과적으로 파라미터화에 영향을 줍니다. JMP는 SAS 시스템의 GLM 프로시저와 다른 코딩을 사용하지만 대부분의 경우 JMP와 SAS GLM 프로시저를 통해 생성되는 결과는 동일합니다. 다음 섹션에서는 JMP 코딩 상세 정보를 설명하고, SAS GLM 프로시저의 코딩과 다른 경우 해당 사례를 강조합니다.

목차

반응 모형 489
연속형 반응 489
명목형 반응 489
순서형 반응 491
요인 모형 492
연속형 요인 492
명목형 요인 493
순서형 요인 503
빈도 508
일반적인 가정 509
가정된 모형 509
상대적 유의성 509
다중 추론 509
타당도 평가 509
대체 방법 510
핵심 통계적 개념 510
불확도, 통합 개념 511
두 개의 기본 적합 기계 511
가능도 , AICc 및 BIC 514
검정력 계산 515
LSN 계산 516
LSV 계산 516
검정력 계산 517
조정 검정력 계산 518
신뢰 한계를 사용한 역추정 예측 519

반응 모형

JMP 는 선형 모형을 연속형, 명목형 및 순서형이라는 세 가지 유형의 반응에 적합시킵니다. JMP 에서 사용할 수 있는 모형과 방법은 실용적이고 널리 사용되며 통계 소프트웨어 도구에서 일반적인 접근 방식의 요구 사항에 잘 맞습니다. 모든 통계 소프트웨어에서 그렇듯이 사용하기로 선택한 모형의 가정 및 가정이 충족되지 않을 경우 발생하는 결과에 대해 학습해야 합니다. 자세한 내용은 이 장의 " 일반적인 가정 "(509 페이지) 에서 확인하십시오.

- " 연속형 반응 "
- " 명목형 반응 "
- " 순서형 반응 "

연속형 반응

반응 열 (Y 역할이 할당된 열) 이 연속형인 경우 JMP 에서 반응 값을 직접 적합시킵니다. 각 관측값에 대한 기본 모형은 다음과 같습니다.

Y = (X 및 모수의 함수) + 오차

통계 검정은 모형의 오차 항이 정규 분포를 따른다는 가정을 기반으로 합니다.

연속형 반응의 적합 원칙

이 적합 원칙을 **최소 제곱**이라고 합니다. 최소 제곱 방법은 오차 제곱합을 최소화하도록 모형의 모수를 추정합니다. 적합 모형의 오차(**잔차**)는 각 관측값의 실제값과 적합 모형에서 예측한 값 사이의 차이입니다.

오차가 정규 분포를 따르는 경우 최소 제곱 방법은 최대 가능도 추정 방법과 동등합니다. 이는 분석을 통해 가장 가능성이 높은 잔차를 제공하는 모형을 추정한다는 의미입니다. 로그 가능도는 정규 분포에 대한 오차 제곱합의 척도 배수입니다.

연속형 반응의 기준 모형

연속형 측정에 대한 가장 단순한 모형은 하나의 값만 적합시켜 모든 반응 값을 예측합니다. 이 값은 **평균**의 추정값입니다. 평균은 반응 값의 산술평균입니다. 다른 모든 모형이 이 기준 모형과 비교됩니다.

명목형 반응

명목형 반응은 로짓 모형의 간단한 확장을 통해 분석됩니다. 이항 (2 수준) 반응의 경우 로짓 반응 모형은 다음과 같이 지정됩니다.

$$\log\left(\frac{P(y=1)}{P(y=2)}\right) = X\beta$$

위의 모형을 다음과 같이 작성할 수도 있습니다 .

$$P(y = 1) = F(X\beta)$$

여기서 $F(x)$ 는 표준 로지스틱 분포의 누적 분포 함수입니다 .

$$F(x) = \frac{1}{1+e^{-x}} = \frac{e^{x}}{1+e^{x}}$$

반응 수준이 r 개인 경우 JMP 는 반응이 데이터 값에 의해 주어진 r 개의 여러 반응 수준 중 하나일 확률을 적합시킵니다 . 확률 추정값은 모두 양수여야 합니다 . 주어진 X 구성에 대해 확률 추정값의 합은 전체 반응 수준에서 1 이어야 합니다 . JMP 에서 확률을 예측하는 데 사용되는 함수는 선형 모형과 다중 반응 로지스틱 함수의 합성입니다 . 확률 비율의 로그가 선형 모형이므로 **로그 선형** 모형이라고도 합니다 . JMP 는 각 반응 확률을 r 번째 확률과 관련짓고 , 별도의 설계 모수 집합을 이러한 $r - 1$ 개 모형에 적합시킵니다 .

$$\log\left(\frac{P(y = j)}{P(y = r)}\right) = X\beta_{(j)} \quad j = 1, \ldots, r-1 \text{ 의 경우}$$

명목형 반응의 적합 원칙

이 적합 원칙을 **최대 가능도**라고 합니다. 이 방법은 데이터에 의해 제공된 모든 반응의 결합 확률이 모형에서 얻을 수 있는 최대값이 되도록 모수를 추정합니다. 결합 확률(가능도)을 직접 보고하는 것보다 가능도의 음의 로그 합계를 보고하는 것이 관리하기가 더 쉽습니다.

불확도 (음의 로그 가능도) 는 모형이 표본 데이터에서 실제로 발생한 반응에 기인한 확률의 음의 로그 합입니다 . 이 값은 크기가 n 인 표본에 대해 H 로 표시되고 다음과 같이 작성됩니다 .

$$H = \sum_{i=1}^{n} -\log(P(y = y_i))$$

확률 1을 발생한 각 사건에 기인한 것으로 볼 경우 완벽한 적합을 위한 음의 로그 합은 0입니다.

명목형 모형, 특히 반응 수준이 많은 모형을 적합시키려면 시간 및 메모리가 많이 소요될 수 있습니다. JMP는 음의 로그 가능도 값이 추정값에 수렴할수록 작아지는 것을 보여 주는 **반복 기록**을 사용하여 계산 진행 정도를 추적합니다.

명목형 반응의 기준 모형

명목형 반응에 대한 가장 단순한 모형은 전체 데이터 테이블에서 각 반응 수준에 대한 발생률로 적합된 일련의 상수 반응 확률입니다 . 즉 , y 가 반응 수준 j 일 확률은 총 표본 수 n 을 각 반응 수준 n_j 의 총계로 나누어 추정됩니다 . 이 확률은 다음과 같이 지정됩니다 .

$$p_j = \frac{n_j}{n}$$

다른 모든 모형이 이 기준 모형과 비교됩니다. 명목형 반응에 대한 기준 모형은 연속형 모형에 대한 표본 평균과 동일하게 사용됩니다.

R^2 통계량은 다음과 같이 모형에 의해 설명되는 불확도 비율을 측정합니다.

$$1 - \frac{H(\text{전체 모형})}{H(\text{기준 모형})}$$

그러나 실제로 범주형 모형에 대해 1에 가까운 R^2을 얻기는 어렵습니다.

순서형 반응

순서형 반응 (Y)을 사용할 경우 JMP는 명목형 반응과 마찬가지로 반응이 데이터에 의해 주어진 r개의 여러 반응 수준 중 하나일 확률을 적합시킵니다.

순서형 데이터에는 연속형 데이터와 같이 순서가 있습니다. 순서는 분석에 사용되지만 정렬된 순서 사이의 간격이나 거리는 사용되지 않습니다. 수치형 반응이 있지만 모형에서 값 간격을 무시하려면 해당 반응 열에 순서형 수준을 할당하면 됩니다. 분류 변수가 있고 수준이 낮음, 중간, 높음과 같이 자연 순서인 경우 순서형 모델링 유형을 사용할 수 있습니다.

순서형 반응은 일련의 병렬 로지스틱 곡선을 누적 확률에 적합시켜 모델링됩니다. 각 곡선은 설계 모수가 동일하지만 절편이 다르며 다음과 같이 지정됩니다.

$$\mathrm{P}(y \le j) = F(\alpha_j + \mathrm{X}\beta) \quad j = 1, \ldots, r-1 \text{ 의 경우}$$

여기서 r 개의 반응 수준이 있고 $F(x)$ 는 표준 로지스틱 누적 분포 함수입니다.

$$F(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$$

다음과 같이 여러 임계값을 통과할 때 순서형 반응이 변경되는 관측되지 않은 연속형 변수 z 의 관점으로 작성할 수도 있습니다.

$$y = \begin{cases} r & \alpha_{r-1} \le z \\ j & \alpha_{j-1} \le z < \alpha_j \\ 1 & z \le \alpha_1 \end{cases}$$

여기서 z 는 관측 불가능한 선형 모형과 오차의 함수입니다.

$$z = \mathrm{X}\beta + \varepsilon$$

ε는 로지스틱 분포를 따릅니다.

이러한 모형은 반응의 순서 특성을 인식하고 명목형 모형에 비해 필요한 모수 수가 훨씬 적으며 계산이 빠르기 때문에 유용합니다.

순서형 모형을 구성하는 방법은 승산을 모델링하는 대신 누적 분포를 모델링하는 명목형 모형의 경우와 다르지만 수학적으로 동등합니다. 마지막 수준을 제외한 모든 수준에 대해 함수를 적합시키는 대신 하나의 함수만 적합시킨 후 각 누적 반응 확률을 적합시킵니다.

순서형 반응의 적합 원칙

순서형 반응 모형에 대한 최대 가능도 적합 원칙은 명목형 반응 모형의 경우와 동일합니다. 이 방법은 발생하는 모든 반응의 결합 확률이 모형에서 얻을 수 있는 최대값이 되도록 모수를 추정합니다. 명목형 적합에 비해 더 빠르고 메모리를 절약하는 반복 방법을 사용합니다.

기준 모형

순서형 반응에 대한 가장 단순한 모형은 명목형 반응과 마찬가지로 전체 데이터 테이블에서 반응에 대한 발생률로 적합된 일련의 반응 확률입니다.

요인 모형

x 변수 (요인) 를 모델링하여 기대값 또는 확률을 예측하는 방법은 모형 요인의 핵심입니다 .

요인은 x 값과 추정할 모수의 선형 결합으로 예측 방정식에 참여합니다 . i 가 관측값을 인덱싱하고 j 가 모수를 인덱싱하는 연속형 반응 모형의 경우 일반적인 관측값의 가정된 모형 y_i 는 다음과 같이 작성됩니다 .

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

다음은 각 요소에 대한 설명입니다 .

y_i - 반응

x_{ij} - 데이터의 함수

ε_i - 관측 불가능한 랜덤 오차 구현

b_j - 추정할 미지의 모수

요인 항에서 선형 모형의 x가 구성되는 방식은 각 모델링 유형에 따라 다릅니다. 선형 모형 x는 교호작용 또는 내포 효과와 같은 복합 효과일 수도 있습니다. 복합 효과는 이후에 자세히 논의됩니다.

연속형 요인

연속형 요인은 설계 행렬에 직접 회귀변수로 배치됩니다. 열이 다른 열의 선형 함수인 경우 이 열의 모수는 **영점화됨** 또는 **추정할 수 없음**으로 표시됩니다. 연속형 요인은 다른 요인과 교차할 때 (교호작용 또는 다항식 항) 평균으로 중심화됩니다. 요인의 열 특성이 **혼합물** 또는 **코딩**이거나, 모형을 지정할 때 중심화 다항식 옵션이 해제된 경우에는 중심화가 억제됩니다. 코딩 열 특성이 있으면 적합 전에 요인이 코드화됩니다.

명목형 요인

명목형 요인은 설계 행렬에 대한 표시 변수로 변환됩니다. SAS GLM은 각 명목형 수준에 대한 표시자 열을 생성합니다. JMP는 마지막 수준을 제외한 각 명목형 수준에 대해 동일한 표시자 열을 생성합니다. 마지막 명목형 수준이 나오면 요인의 다른 모든 열에서 하나가 빠집니다. 아래와 같이 GLM 및 JMP에 대해 각각 세 가지 수준이 코드화된 명목형 요인 A를 예로 들어보겠습니다.

표 A.1 명목형 요인 A

	GLM			JMP	
A	A1	A2	A3	A13	A23
A1	1	0	0	1	0
A2	0	1	0	0	1
A3	0	0	1	–1	–1

GLM의 선형 모형 설계 행렬에는 열 사이에 선형 종속성이 있으며 최소 제곱 해는 일반화 역행렬을 사용합니다. 선택된 해는 A3 모수가 0으로 설정되게 합니다.

JMP의 선형 모형 설계 행렬은 결측 셀 또는 기타 부수적 공선성이 있는 경우를 제외하고 완전 계수(full rank)가 되도록 코드화됩니다. 마지막 수준에 대한 A 효과의 모수는 다른 수준의 음의 합이므로 모든 효과 수준에서 모수의 합이 0이 됩니다.

모수 해석

참고 : 명목형 수준의 모수는 해당 수준에 대한 예측 반응과 모든 수준에 대한 평균 예측 반응 간의 차이로 해석됩니다.

요인 수준에 대한 설계 열은 해당 요인 수준의 0-1 표시자에서 마지막 수준의 표시자를 뺀 값으로 생성됩니다. 이 코딩으로 인해 위의 모수 해석이 가능합니다.

표 A.2 모수 해석

JMP 모수 보고서	해석 방법	설계 열 코딩
절편	모든 수준에 대한 평균	**1**′
A[1]	$\alpha_1 - 1/3(\alpha_1 + \alpha_2 + \alpha_3)$	(A==1) – (A==3)
A[2]	$\alpha_2 - 1/3(\alpha_1 + \alpha_2 + \alpha_3)$	(A==2) – (A==3)

교호작용 및 교차효과

GLM 과 JMP 에서는 교차되는 요인의 설계 열 행에 직접곱을 적용하여 교호작용 효과를 생성합니다 . 예를 들어 다음 GLM 코드는

```
PROC GLM;
  CLASS A B;
  MODEL A B A*B;
```

아래와 같은 설계 행렬을 생성합니다.

표 A.3 설계 행렬

		A			B			AB								
A	B	1	2	3	1	2	3	11	12	13	21	22	23	31	32	33
A1	B1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
A1	B2	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
A1	B3	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
A2	B1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
A2	B2	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
A2	B3	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0
A3	B1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
A3	B2	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
A3	B3	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1

JMP **모형 적합** 명령을 사용하고 A 및 B 열에 대한 요인 모형을 요청하면 다음과 같은 설계 행렬이 생성됩니다. 이 행렬의 A13은 이전 행렬의 A1–A3입니다. 그러나 A13B13은 현재 행렬에서 A13*B13입니다.

표 A.4 현재 행렬

		A		B					
A	B	13	23	13	23	A1 B1	A1 B1	A1 B1	A1 B1
A1	B1	1	0	1	0	1	0	0	0
A1	B2	1	0	0	1	0	1	0	0
A1	B3	1	0	–1	–1	–1	–1	0	0

표 A.4 현재 행렬 (계속)

A2	B1	0	1	1	0	0	0	1	0
A2	B2	0	1	0	1	0	0	0	1
A2	B3	0	1	–1	–1	0	0	–1	–1
A3	B1	–1	–1	1	0	–1	0	–1	0
A3	B2	–1	–1	0	1	0	–1	0	–1
A3	B3	–1	–1	–1	–1	1	1	1	1

JMP 코딩을 사용하면 수준 수가 적은 요인의 교호작용 문제에 대해 메모리 및 계산 시간을 절약할 수 있습니다.

3x3 교차 모형에 대해 모수에 관한 셀의 기대값은 다음과 같습니다 .

표 A.5 3x3 교차 모형

	B1	B2	B3
A1	$\mu + \alpha_1 + \beta_1 + \alpha\beta_{11}$	$\mu + \alpha_1 + \beta_2 + \alpha\beta_{12}$	$\mu + \alpha_1 - \beta_1 - \beta_2 - \alpha\beta_{11} - \alpha\beta_{12}$
A2	$\mu + \alpha_2 + \beta_1 + \alpha\beta_{21}$	$\mu + \alpha_2 + \beta_2 + \alpha\beta_{22}$	$\mu + \alpha_2 - \beta_1 - \beta_2 - \alpha\beta_{21} - \alpha\beta_{22}$
A3	$\mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21}$	$\mu - \alpha_1 - \alpha_2 + \beta_2 - \alpha\beta_{12} - \alpha\beta_{22}$	$\mu - \alpha_1 - \alpha_2 - \beta_1 - \beta_2 + \alpha\beta_{11} + \alpha\beta_{12} + \alpha\beta_{21} + \alpha\beta_{22}$

내포 효과

GLM은 모형에 없는 것을 기준으로 우측 검정을 판별하므로 GLM의 내포 효과는 교호작용 효과와 동일하게 코드화됩니다. 모형에 포함되지 않은 효과는 포함하는 교호작용(내포와 동등) 효과에 의해 흡수될 수 있습니다.

JMP의 내포 효과는 다르게 코드화됩니다. JMP는 괄호로 묶은 항을 각 그룹에 대한 그룹화 항으로 사용합니다. 내포하는 항의 각 수준 조합에 대해 JMP는 괄호 밖의 효과를 생성합니다. 바깥쪽 항의 수준은 내포하는 항의 전체 수준에서 일치시킬 필요가 없습니다. 각 내포 수준은 설계 열 및 모수 생성과 관련하여 별도로 간주됩니다.

표 A.6 내포 효과

				B(A)					
				A1	A1	A2	A2	A3	A3
A	B	A13	A23	B13	B23	B13	B23	B13	B23
A1	B1	1	0	1	0	0	0	0	0
A1	B2	1	0	0	1	0	0	0	0
A1	B3	1	0	–1	–1	0	0	0	0
A2	B1	0	1	0	0	1	0	0	0
A2	B2	0	1	0	0	0	1	0	0
A2	B3	0	1	0	0	–1	–1	0	0
A3	B1	–1	–1	0	0	0	0	1	0
A3	B2	–1	–1	0	0	0	0	0	1
A3	B3	–1	–1	0	0	0	0	–1	–1

명목형 요인에 대한 최소 제곱 평균

최소 제곱 평균은 다른 모든 요인을 중립 값으로 설정한 후 일부 수준 조합에 해당하는 예측값입니다. 직접 배치된 연속형 회귀변수에 대한 중립 값은 표본 평균으로 정의됩니다. 연관이 없는 명목형 요인이 포함된 효과의 중립 값은 전체 수준에 적용된 평균 효과로 정의됩니다(코딩에서 모두 0이 됨). 순서형 요인은 다른 중립 값을 사용합니다("순서형 최소 제곱 평균"(506페이지) 참조). 최소 제곱 평균을 추정하지 못할 수 있으며, 이에 해당하는 경우 "추정할 수 없음"으로 표시됩니다. JMP의 최소 제곱 평균은 가중치가 사용되는 경우를 제외한 모든 경우에서 SAS PROC GLM의 최소 제곱 평균과 일치합니다(Goodnight and Harvey 1978). 가중치 변수가 사용되면 중립 값에 대해 JMP는 가중 평균을 사용하고 SAS PROC GLM은 비가중 평균을 사용합니다.

유효 가설 검정

일반적으로 JMP에서 생성하는 가설 검정은 SAS PROC GLM(Type III 및 IV 가설) 같은 대부분의 다른 신뢰할 수 있는 프로그램의 가설 검정과 일치합니다. 다음 두 섹션에서는 차이가 있는 경우에 대해 설명합니다.

SAS PROC GLM에서 Type III 및 IV에 대한 가설 검정은 일반 형식의 추정 가능한 함수를 사용하고, 관심 효과 및 관심 효과에 있는 효과만 포함하는 함수를 검색하여 생성됩니다(Goodnight 1978).

동일한 검정이 JMP에서 생성됩니다. 그러나 파라미터화가 다르므로 해당 효과의 모든 모수에 대한 결합 검정을 수행하여 효과를 검정할 수 있습니다(일단 완전 계수 가정). 포함하는 교호작

용 모수는 검정과 관련이 없습니다. 코딩으로 인해 이러한 모수는 포함된 효과에 대한 검정과 관련이 없기 때문입니다.

결측 셀 또는 기타 특이성이 있는 경우 JMP 검정은 GLM 검정과 다릅니다 . 아래에서 두 검정의 차이에 대해 설명합니다 .

- JMP 검정은 최소한 주효과에 대해 최소 제곱 평균이 다른지를 검정하는 것과 동등합니다 . 최소 제곱 평균을 추정할 수 없는 경우에는 검정에 일부 비교를 포함할 수 없으므로 자유도가 손실됩니다 . 교호작용의 경우 JMP 는 최소 제곱 평균의 차이가 모형의 포함하는 효과로 설명되는 주변 패턴보다 큰지를 검정합니다 .
- JMP 는 해당 효과가 있는 모형의 SSE 를 해당 효과가 없는 모형의 SSE 와 비교하여 효과를 검정합니다 . JMP 는 이 방법이 타당하도록 모형을 파라미터화합니다 .
- JMP 는 셀 평균 파라미터화가 아니라 구조적 파라미터화를 사용하지만 Hocking(1985, pp. 80–89, 163–166) 에서 설명하는 **유효 가설 검정**을 구현합니다 . 유효 가설 검정은 효과에 대해 원하는 가설로 시작하여 " 가능한 한 많은 " 가설을 포함합니다 . 물론 포함하는 효과에 결측 셀이 있으면 완전 가설을 추정할 수 없으므로 이 검정에서 가설의 일부를 제거해야 합니다 . 유효 가설은 완전 가설을 최대한 적게 제거합니다 .
- JMP 와 GLM 및 기타 프로그램에서 결측 셀의 존재와 관련된 가설 검정 간의 차이는 흥미로운 검정으로 간주되지 않습니다. 교호작용이 유의하면 포함된 주효과에 대한 검정은 관심 대상이 아닙니다. 교호작용이 유의하지 않으면 모형에서 해당 교호작용을 항상 제거할 수 있습니다. 일부 검정은 고유하지도 않습니다. 결측 셀 설계의 수준에 라벨을 다시 지정하면 GLM Type IV 검정이 변경될 수 있습니다 .

다음 섹션에서는 이 항목에 대한 내용을 더 자세히 설명합니다.

명목형 효과의 특이성 및 결측 셀

설계 열 사이의 선형 종속성을 가정해 보겠습니다. JMP 코딩을 사용하면 데이터가 부족하여 추정이 필요한 조합을 작성할 수 없거나 효과의 교락 또는 공선성 유형이 있는 경우 외에는 선형 종속성 문제가 발생하지 않습니다.

선형 종속성이 있으면 모수에 대한 최소 제곱 해가 고유하지 않을 수 있으며 일부 가설 검정을 수행할 수 없습니다. 설계 열이 모형의 이전 효과에 선형 종속되므로 JMP는 모수 추정값을 순서대로 0으로 설정하는 전략을 선택합니다. 영점화되는 모수 추정값과 추정 불가능한 모수 추정값을 보여주는 특수 열이 보고서에 포함됩니다. 별도의 **특이성** 보고서에는 선형 종속성이 표시됩니다.

특이성이 있는 경우 JMP에서 검정하는 가설이 GLM에서 선택하는 가설과 다를 수 있습니다. 일반적으로 JMP는 검정을 더 높은 주변화 수준으로 유지하므로 GLM보다 검정 자유도가 적습니다. 즉, JMP 검정은 해당 효과에 대한 전체 최소 제곱 평균의 검정에 해당하지만 GLM 검정은 이 특성이 항상 적용되지는 않습니다.

예를 들어 교호작용이 있는 이원 모형과 하나의 결측 셀을 가정해 봅니다. 여기서 A의 수준은 세 개, B의 수준은 두 개, A3B2는 결측 셀입니다.

표 A.7 교호작용이 있는 이원 모형

A B	A1	A2	B1	A1B1	A2B1	
A1 B1	1	0	1	1	0	
A2 B1	0	1	1	0	1	
A3 B1	–1	–1	1	–1	–1	
A1 B2	1	0	–1	–1	0	
A2 B2	0	1	–1	0	–1	
A3 B2	–1	–1	–1	1	1	이 교호작용이 없는 경우를 가정해 봅니다.

각 셀의 기대값은 다음과 같습니다 .

표 A.8 기대값

	B1	B2
A1	$\mu + \alpha_1 + \beta_1 + \alpha\beta_{11}$	$\mu + \alpha_1 - \beta_1 - \alpha\beta_{11}$
A2	$\mu + \alpha_2 + \beta_1 + \alpha\beta_{21}$	$\mu + \alpha_2 - \beta_1 - \alpha\beta_{21}$
A3	$\mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21}$	$\mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}$

데이터를 포함하는 셀에 추정 가능한 기대값이 있는 것은 분명합니다 . 결측 셀에는 추정할 수 없는 기대값이 있습니다 . 사실상 기대값은 특이성 보고서에 있는 설계 열의 선형 결합입니다 .

$$\mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}$$

B1 과 B2 의 최소 제곱 평균을 비교하는 검정을 생성하려는 경우를 가정해 보겠습니다 . 이 예에서 위의 표에 있는 행 평균은 이러한 최소 제곱 평균을 제공합니다 .

$$\begin{aligned} \text{LSM(B1)} = & (1/3)(\mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \\ & \mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \\ & \mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21}) \\ = & \ \mu + \beta_1 \end{aligned}$$

$$\begin{aligned} \text{LSM(B2)} = & (1/3)(\mu + \alpha_1 + -\beta_1 - \alpha\beta_{11} + \\ & \mu + \alpha_2 + -\beta_1 - \alpha\beta_{21} + \\ & \mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}) \\ = & \ \mu - \beta_1 \end{aligned}$$

$$LSM(B1) - LSM(B2) = 2\beta_1$$

이것은 β_1 모수에 대한 검정이 최소 제곱 평균이 동일한지 검정하는 것과 동등하다는 것을 보여 줍니다. 그러나 β_1은 추정할 수 없으므로 검정이 불가능하며, 이는 검정에 대한 자유도가 없다는 의미입니다.

이제 A 수준에 대한 최소 제곱 평균 검정을 생성합니다.

$$LSM(A1) = (1/2)(\mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \mu + \alpha_1 - \beta_1 - \alpha\beta_{11}) = \mu + \alpha_1$$

$$LSM(A2) = (1/2)(\mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \mu + \alpha_2 - \beta_1 - \alpha\beta_{21}) = \mu + \alpha_2$$

$$LSM(A3) = (1/2)(\mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21} + \mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}) = \mu - \alpha_1 - \alpha_2$$

$$LSM(A1) - LSM(A3) = 2\alpha_1 + \alpha_2$$

$$LSM(A2) - LSM(A3) = 2\alpha_2 + \alpha_1$$

이 중에서 어느 것도 추정할 수 없지만 추정 가능한 다른 비교가 있습니다. 결측 셀이 없는 A 열 두 개를 비교할 수 있습니다.

$$LSM(A1) - LSM(A2) = \alpha_1 - \alpha_2$$

이 조합은 효과에 두 개의 모수가 있지만 자유도가 1인 검정을 사용하여 JMP에서 검정됩니다.

내곱과 특이성 조합을 사용하고 이 값이 0 인지 검토하여 추정 가능성을 확인할 수 있습니다.

표 A.9 확인

	특이성 조합	
모수	조합	검정
m	1	0
a_1	–1	1
a_2	–1	–1
b_1	–1	0
ab_{11}	1	0
ab_{21}	1	0

교호작용에 대한 결측 셀의 설계 열은 항상 주효과(명목형 요인의 경우)의 자유도를 제거하는 것으로 나타났습니다. 따라서 최소 제곱 평균의 추정 불가능과 이러한 최소 제곱 평균에 해당하는 효과를 검정하기 위한 자유도의 손실 사이에는 직접적인 관계가 있습니다.

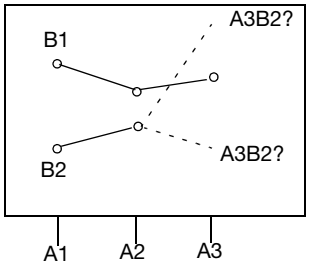
이 내용을 GLM의 방식과 비교해 볼 수 있습니다. 결측 셀이 없을 때는 GLM과 JMP에서 수행하는 검정이 동일합니다. 즉, 두 방법 모두 최소 제곱 평균이 동일한지 효과적으로 검정합니다. 그러나 GLM은 특이성이 발견되면 Type III인지 아니면 Type IV인지에 따라 이러한 셀을 다른 방식으로 처리합니다. Type IV의 경우 찾을 수 있는 추정 가능한 조합을 검색합니다. 이는 고유하지 않을 수 있으며 수준을 재정렬하면 다른 결과를 얻을 수 있습니다. Type III의 경우 고유한 검정을 얻기 위해 추정 가능한 함수의 직교화를 수행합니다. 그러나 셀 평균에 관해 검정을 해석하기가 매우 어려울 수 있습니다.

JMP 방식은 처음에는 GLM을 사용할 때보다 자유도 손실이 더 많아 곤란할 수 있지만 다음과 같은 몇 가지 장점이 있습니다 .

1. 검정이 이론적으로 LSM 에 연결됩니다 .
2. 재파라미터화된 모형의 제곱합 감소를 사용하여 검정을 쉽게 계산할 수 있습니다 .
3. 검정이 Hocking 의 " 유효 가설 검정 (Effective Hypothesis Tests)" 과 일치합니다 .
4. 검정은 **전체 주변 검정**입니다 . 즉 , 검정은 교호작용의 다른 효과를 완벽하게 가로질러 수행하게 됩니다 .

마지막으로 몇 가지 세부 사항을 설명합니다. A3B2가 결측 셀인 이전 예에서 셀 평균의 기대값 그래프를 생각해 봅니다.

그림 A.1 셀 평균의 기대값



이 그래프에서는 결측 셀이 있는 셀 평균의 기대값을 보여 줍니다. A1 및 A2 셀의 평균은 B 수준에 대해 프로파일링됩니다. JMP 방식을 사용할 경우 결측 셀의 평균은 교호작용 항에서 허용하는 무엇이든 될 수 있으므로 A3B2 결측 셀이 있으면 B 주효과를 검정할 수 없습니다. 결측 셀의 평균이 표시된 값보다 높으면 B 효과 검정이 유의할 가능성이 큽니다. 값이 더 낮으면 검정이 유의하지 않을 가능성이 큽니다. 요점은 사용자가 모른다는 것입니다. 최소 제곱 평균이 추정 불가능한 것으로 선언되는 것이 이 경우에 해당됩니다. 효과에 대한 가설도 마찬가지로 사용자가 알지 못한다는 것을 가정해야 합니다.

셀 부분집합에 대한 차이와 관련된 가설을 검정하려는 경우 GLM Type IV가 유용합니다. JMP에서는 많은 계산을 통해 효과를 분할하거나 대비를 사용하여 이러한 검정을 직접 생성해야 합니다.

JMP 및 GLM 가설

GLM은 JMP와 다르게 작동하며 결측 셀이 있을 때는 다른 가설 검정을 생성합니다. 특히 GLM은 JMP와 달리 효과에서 내포와 교차의 차이를 인식하지 못합니다. A, B(A), C(A B)라는 3계층의 내포 효과가 있고 지분 설계의 단계에서 수준 수가 다른 경우를 가정해 보겠습니다.

그림 A.10 에서는 GLM 모수에 관한 주효과 A 의 검정을 보여 줍니다 . 첫 번째 열 집합은 JMP 에서 수행한 검정입니다 . 두 번째 열 집합은 GLM Type IV 에서 수행한 검정입니다 . 세 번째 열 집합은 JMP 의 검정과 동등한 검정이며 처음 두 열에 다음 행렬을 곱한 결과입니다 .

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

이는 GLM 검정과 유사합니다. 마지막 열 집합은 GLM Type III 검정입니다. 검정에서 포함하는 효과를 분배하는 방식에 차이가 있습니다. JMP에서는 하향식 계층 구조입니다. GLM Type IV에서는 검정이 상향식입니다. 실제로는 검정 통계량이 유사한 경우가 많습니다.

표 A.10 GLM 및 JMP 가설 비교

모수	A 에 대한 JMP 검정		A 에 대한 GLM-IV 검정		JMP 회전 검정		A 에 대한 GLM-III 검정	
u	0	0	0	0	0	0	0	0
a1	0.6667	-0.3333	1	0	1	0	1	0
a2	–1	0.6667	0	1	0	1	0	1
a3	–1	-0.3333	-1	-1	-1	-1	-1	-1
a1b1	0.1667	-0.0833	0.2222	0	0.25	0	0.2424	0
a1b2	0.1667	-0.0833	0.3333	0	0.25	0	0.2727	0
a1b3	0.1667	-0.0833	0.2222	0	0.25	0	0.2424	0
a1b4	0.1667	-0.0833	0.2222	0	0.25	0	0.2424	0
a2b1	-0.1667	0.3333	0	0.5	0	0.5	0	0.5
a2b2	-0.1667	0.3333	0	0.5	0	0.5	0	0.5
a3b1	-0.1111	-0.1111	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333
a3b2	-0.1111	-0.1111	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333

표 A.10 GLM 및 JMP 가설 비교 (계속)

a3b3	-0.1111	-0.1111	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333
a1b1c1	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b1c2	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b2c1	0.0556	-0.0278	0.1111	0	0.0833	0	0.0909	0
a1b2c2	0.0556	-0.0278	0.1111	0	0.0833	0	0.0909	0
a1b2c3	0.0556	-0.0278	0.1111	0	0.0833	0	0.0909	0
a1b3c1	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b3c2	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b4c1	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b4c2	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a2b1c1	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a2b1c2	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a2b2c1	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a2b2c2	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a3b1c1	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b1c2	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b2c1	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b2c2	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b3c1	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b3c2	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667

JMP 파라미터화 관점에서 A 에 대한 검정은 다음과 같습니다 .

표 A.11 A 에 대한 검정

모수	GLM–IV 검정		JMP 검정	
m	0	0	0	0
a13	2	1	1	0
a23	1	2	0	1
a1:b14	0	0	0	0
a1:b24	0.11111	0	0	0
a1:b34	0	0	0	0
a2:b12	0	0	0	0
a3:b13	0	0	0	0
a3:b23	0	0	0	0
a1b1:c12	0	0	0	0
a1b2:c13	0	0	0	0
a1b2:c23	0	0	0	0
a1b3:c12	0	0	0	0
a1b4:c12	0	0	0	0
a2b1:c13	0	0	0	0
a2b2:c12	0	0	0	0
a3b1:c12	0	0	0	0
a3b2:c12	0	0	0	0
a3b3:c12	0	0	0	0

JMP 관점에서 보면 GLM 검정은 a1b24 모수에 계수를 적용하여 약간 다릅니다.

순서형 요인

순서형 모델링 유형으로 표시된 요인은 명목형 요인과 다르게 코드화됩니다. 모수 추정값이 다르게 해석되고 검정 및 최소 제곱 평균이 다릅니다.

순서형 요인의 첫 번째 수준은 대조 또는 기준 수준이며 , 모수는 순서형 요인이 각 후속 수준으로 설정될 때 반응에 미치는 영향을 측정합니다 . 순서형 요인 코딩은 첫 번째 용량이 0 일 때 다양한 용량을 나타내는 수준이 포함된 요인에 적합합니다 . 다음 표에서는 3 수준 순서형 요인의 예를 보여 줍니다 .

표 A.12 순서형 요인

항	코드화된 열		
A	a2	a3	
A1	0	0	대조 수준, 용량 0
A2	1	0	낮은 용량
A3	1	1	높은 용량

설계 패턴은 하위 삼각형 요소는 1이고 다른 요소는 0입니다.

단순 주효과 모형의 경우 다음과 같이 지정할 수 있습니다 .

$$y = \mu + \alpha_2 X_{(a \le 2)} + \alpha_3 X_{(a \le 3)} + \varepsilon$$

여기서 μ는 $A = 1$에서의 기대 반응, $\mu + \alpha_2$는 $A = 2$에서의 기대 반응, $\mu + \alpha_2 + \alpha_3$은 $A = 3$에서의 기대 반응을 나타냅니다. 따라서 α_2는 $A = 1$에서 $A = 2$로 이동 효과를 추정하고, α_3은 $A = 2$에서 $A = 3$으로 이동 효과를 추정합니다.

순서형 주효과에 대한 모든 모수의 부호가 동일하면 순서형 수준에서 반응 효과는 단조 특성을 보입니다.

순서형 교호작용

순서형 교호작용은 명목형 효과와 마찬가지로 요인 열의 수평 직접곱으로 생성됩니다. 두 개의 순서형 요인 A와 B가 있고 각 요인의 수준이 세 개인 경우를 예로 들겠습니다. JMP의 순서형 코딩은 다음에 표시된 설계 행렬을 생성합니다. 교호작용의 패턴은 1로 구성된 하삼각 행렬의 블록 하삼각 행렬입니다.

표 A.13 순서형 교호작용

						A*B			
						A2		A3	
A	B	A2	A3	B2	B3	B2	B3	B2	B3
A1	B1	0	0	0	0	0	0	0	0
A1	B2	0	0	1	0	0	0	0	0
A1	B3	0	0	1	1	0	0	0	0
A2	B1	1	0	0	0	0	0	0	0
A2	B2	1	0	1	0	1	0	0	0
A2	B3	1	0	1	1	1	1	0	0
A3	B1	1	1	0	0	0	0	0	0
A3	B2	1	1	1	0	1	0	1	0
A3	B3	1	1	1	1	1	1	1	1

참고 : 효과가 없는지 여부를 검정할 때 단순 모형에 대한 명목형 요인과 순서형 요인 사이에는 큰 차이가 없습니다. 그러나 교호작용이 지정되면 큰 차이를 보입니다. 대부분의 모형에 순서형 요인보다는 명목형 요인을 사용하는 것이 좋습니다.

순서형 교차 모형에 대한 가설 검정

모수의 의미를 확인하려면 모수에 관한 셀 평균 기대값을 보여 주는 아래 표를 검토합니다. 여기서 μ는 절편, α_2는 A2 수준의 모수 등입니다.

표 A.14 셀 평균 기대값

	B1	B2	B3
A1	μ	$\mu + \beta_2$	$\mu + \beta_2 + \beta_3$
A2	$\mu + \alpha_2$	$\mu + \alpha_2 + \beta_2 + \alpha\beta_{22}$	$\mu + \alpha_2 + \beta_2 + \beta_3 + \alpha\beta_{22} + \alpha\beta_{23}$
A3	$\mu + \alpha_2 + \alpha_3$	$\mu + \alpha_2 + \alpha_3 + \beta_2 + \alpha\beta_{22} + \alpha\beta_{32}$	$\mu + \alpha_2 + \alpha_3 + \beta_2 + \beta_3 + \alpha\beta_{22} + \alpha\beta_{23} + \alpha\beta_{32} + \alpha\beta_{33}$

실제로 A에 대한 주효과 검정은 첫 번째 수준에서 B를 포함하는 A 수준을 검정하는 것입니다. 마찬가지로 B에 대한 주효과 검정은 첫 번째 수준에서 A를 포함하는 여러 B 수준의 상단 행을 검정하는 것입니다. 이 검정은 두 요인이 모두 처치 용량이 다른 실험에 적합합니다. 주 관심은 각 처치의 단독 효능이며, 두 약물의 용량을 모두 투여했을 때 **약물 교호작용**을 찾는 것은 부수적입니다. 경우에 따라 각 약물이 다량 투여되면 위험할 수도 있습니다.

각 셀의 기대값은 현재 행과 열을 포함하여 왼쪽 및 위에 있는 각 셀과 관련된 모든 모수를 더하여 얻을 수 있습니다. 마지막 셀의 기대값은 모든 모수의 합입니다.

다른 효과에 포함된 효과에 대한 가설 검정은 순서형 코딩과 명목형 코딩에서 다르지만 다른 효과에 포함되지 않은 효과를 검정하는 것은 동일합니다. 위의 교차 설계에서 교호작용에 대한 검정은 A와 B의 적합이 명목형인지 아니면 순서형인지에 관계없이 동일합니다.

순서형 최소 제곱 평균

앞에서 설명했듯이 최소 제곱 평균은 다른 모든 요인을 중립 값으로 설정한 후 일부 수준 조합에 해당하는 예측값입니다. JMP는 연관이 없는 순서형 요인이 포함된 효과의 중립 값을 첫 번째 수준(대조 **기준** 수준)으로 정의합니다.

이와 같은 순서형 요인에 대한 최소 제곱 평균 정의는 포함된 효과의 가설 검정이 최소 제곱 평균의 동일 여부 검정과 동등하다는 개념을 유지합니다.

순서형 효과의 특이성 및 결측 셀

순서형 코딩을 사용하면 순서형 효과의 첫 번째 수준이 기준이라는 것을 의미합니다. 따라서 교호작용에 결측 셀이 있거나 교호작용의 데이터가 없는 경우에도 주효과에 대한 적합한 검정을 얻을 수 있습니다.

결측 셀이 있는 예

이 예는 A3B2 셀에 데이터가 없다는 점을 제외하고 셀당 두 개의 관측값이 있는 위의 예와 동일합니다. 이제 요인의 명목형 코드화 결과를 순서형 코드화 결과와 비교할 수 있습니다. 그림 A.2에서 알 수 있듯이 모형 적합은 동일합니다.

표 A.15 관측값

Y	A	B
12	1	1
14	1	1
15	1	2
16	1	2
17	2	1

표 A.15 관측값 (계속)

17	2	1
18	2	2
19	2	2
20	3	1
24	3	1

그림 A.2 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 요약 정보

적합 요약

R^2	0.891732
Adj-R^2	0.805118
제곱근 평균 제곱 오차	1.48324
반응 평균	17.2
관측값 수(또는 가중합)	10

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	4	90.60000	22.6500	10.2955
오차	5	11.00000	2.2000	Prob > F
수정 합계	9	101.60000		0.0125*

적합 요약

R^2	0.891732
Adj-R^2	0.805118
제곱근 평균 제곱 오차	1.48324
반응 평균	17.2
관측값 수(또는 가중합)	10

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	4	90.60000	22.6500	10.2955
오차	5	11.00000	2.2000	Prob > F
수정 합계	9	101.60000		0.0125*

코딩이 다르기 때문에 모수 추정값은 매우 차이가 있습니다. 결측 셀은 일부 명목형 모수의 추정 가능성에 영향을 주지만 순서형 모수에는 영향을 주지 않습니다.

그림 A.3 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 모수 추정값

모수 추정값

항		추정값	표준 오차	t 비	Prob>|t|
절편		17.333333	0.60553	28.63	<.0001*
A[1]		-4.333333	0.856349	-5.06	0.0039*
A[2]		-0.333333	0.856349	-0.39	0.7131
B[2-1]	편향	1.5	1.48324	1.01	0.3583
A[1]*B[2-1]	편향	1	2.097618	0.48	0.6537
A[2]*B[2-1]	영점화됨	0	0	.	.

모수 추정값

항		추정값	표준 오차	t 비	Prob>|t|
절편		13	1.048809	12.40	<.0001*
A[2-1]		4	1.48324	2.70	0.0429*
A[3-2]		5	1.48324	3.37	0.0199*
B[2-1]		2.5	1.48324	1.69	0.1527
A[2-1]*B[2-1]		-1	2.097618	-0.48	0.6537
A[3-2]*B[2-1]	영점화됨	0	0	.	.

특이성 상세 정보는 선형 종속성을 보여 주며 값을 검토하여 결측 셀을 식별합니다.

그림 A.4 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 특이성 상세 정보

특이성 상세 정보

항	상세 정보
B[2-1]	=A[1]*B[2-1] + A[2]*B[2-1]

특이성 상세 정보

항	상세 정보
A[3-2]*B[2-1]	= 0

명목형에 대한 효과 검정으로 인해 자유도가 손실됩니다. B의 경우 검정이 없습니다. 순서형의 경우 첫 번째 수준(기준)에 대한 결측 셀이 없으므로 손실이 없습니다.

그림 A.5 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 효과 검정

효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F	
A	2	2	81.333333	18.4848	0.0049*	
B	1	0	0.000000	.	.	LostDFs
A*B	2	1	0.500000	0.2273	0.6537	LostDFs

효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F	
A	2	2	81.333333	18.4848	0.0049*	
B	1	1	6.250000	2.8409	0.1527	
A*B	2	1	0.500000	0.2273	0.6537	LostDFs

최소 제곱 평균도 다릅니다. 명목형 LSM은 모두 추정 가능한 것은 아니지만 순서형 LSM은 가능합니다. 셀 평균을 검토하여 값을 확인할 수 있습니다. 두 경우에 대한 A*B LSM은 동일합니다. 그림 A.6에서는 명목형 적합 및 순서형 적합의 최소 제곱 평균을 보여 줍니다.

그림 A.6 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 최소 제곱 평균

효과 상세 정보

A

최소 제곱 평균 테이블

수준	최소 제곱 평균	표준 오차	평균
1	13.000000	1.0488088	14.2500
2	17.000000	1.0488088	17.7500
3	22.000000	1.0488088	22.0000

B

최소 제곱 평균 테이블

수준	최소 제곱 평균		표준 오차	평균
1	17.333333		0.60553007	17.3333
2	0.000000	추정할 수 없음	.	17.0000

A*B

최소 제곱 평균 테이블

수준	최소 제곱 평균		표준 오차
1,1	13.000000		1.0488088
1,2	15.500000		1.0488088
2,1	17.000000		1.0488088
2,2	18.500000		1.0488088
3,1	22.000000		1.0488088
3,2	0.000000	추정할 수 없음	.

효과 상세 정보

A

최소 제곱 평균 테이블

수준	최소 제곱 평균	표준 오차	평균
1	13.000000	1.0488088	14.2500
2	17.000000	1.0488088	17.7500
3	22.000000	1.0488088	22.0000

B

최소 제곱 평균 테이블

수준	최소 제곱 평균	표준 오차	평균
1	13.000000	1.0488088	17.3333
2	15.500000	1.0488088	17.0000

A*B

최소 제곱 평균 테이블

수준	최소 제곱 평균		표준 오차
1,1	13.000000		1.0488088
1,2	15.500000		1.0488088
2,1	17.000000		1.0488088
2,2	18.500000		1.0488088
3,1	22.000000		1.0488088
3,2	0.000000	추정할 수 없음	.

빈도

빈도 (정수가 아닌 값 포함) 가 분석에 미치는 영향은 손실 함수에 미치는 영향으로 설명됩니다. 반응 값 y_i 와 예측 변수 $x_{i1}, x_{i2},..., x_{in}$ 을 사용하여 θ 모수를 추정하려는 경우를 가정해 보겠습니다. 빈도 변수가 없다고 가정할 때 손실 함수는 다음과 같이 제공됩니다.

$$L(\theta|\underset{\sim}{y}) = \sum_{i=1}^{n} L(\theta|y_i, x_{i1}, x_{i2},\ldots,x_{in})$$

빈도 f_i 가 정의된 경우 손실 함수는 다음과 같습니다.

$$L(\theta|\underset{\sim}{y},\underset{\sim}{f}) = \sum_{i=1}^{n} f_i L(\theta|y_i, x_{i1}, x_{i2},\ldots,x_{in})$$

모수 추정값, 표준 오차, 가설 검정, 신뢰 구간 등 추론의 기준이 되는 모든 통계량은 이러한 형식의 손실 함수를 기반으로 계산됩니다.

일반적인 가정

통계량을 신뢰하려면 사용 기법에 대한 제한 사항과 값에 대해 잘 알고 있어야 합니다. 통계 방법은 도구일 뿐이므로 정확하지 않은 과학(근거 없는 통계적 가정) 또는 잘못된 데이터에 대비하거나 책임지지 않습니다.

가정된 모형

대부분의 통계량은 모형이 올바르다는 가정을 기반으로 합니다. 모형이 올바르지 않게 될 정도로 모형에서 생성되는 통계 보고서의 신뢰도를 줄여야 합니다.

상대적 유의성

대부분의 통계 검정에서는 모형을 절대적 의미로 평가하지 않습니다. 검정 통계량이 유의하면 모형이 평균과 같은 일부 축소 모형보다 더 잘 적합된다는 정도로만 해석할 수 있습니다. 모형이 데이터를 적합시킬 수 있지만 기본 물리적 모형을 완전히 잘 설명하지 못할 수 있습니다.

다중 추론

보통 통계 결과의 값은 그것을 있는 그대로 믿기보다는 어떤 발견의 실마리를 제공합니다. 그 발견을 사실로 확인하기 위해 추가 연구를 수행해야 할 수도 있습니다. 그렇지 않으면 그냥 데이터를 선별하는 것입니다.

예를 들어 충분한 분석을 수행하면 요인에 예측값이 없더라도 연구의 5%에서 5%의 유의한 효과를 발견할 수 있습니다. 마찬가지로 데이터에 대한 올바른 모형을 검정하는 대신 데이터를 사용하여 모형의 형태만 만들면 보고서의 유의 수준이 변질됩니다. 그러면 랜덤 오차가 모형 선택에 영향을 주고 모형이 실제보다 우수하다고 믿게 됩니다.

타당도 평가

다음과 같은 다양한 기법과 패턴을 사용하여 모형의 타당도를 평가할 수 있습니다 .

- 적합 결여 검정을 사용하면 포화 상태의 요인과 비교하여 모형 타당도를 확인할 수 있습니다 . 비포화 모형에 반복된 x 값이 포함된 경우 모형 적합 플랫폼에서 이러한 검정을 자동으로 제공합니다 .
- 모형 적합 플랫폼에서 잔차 그림과 스튜던트화 잔차 그림을 검토하여 연속형 반응에 대한 분포 가정을 확인할 수 있습니다 . 아니면 플랫폼 팝업 메뉴에서 **저장** 명령을 사용하여 데이터

테이블 열에 잔차를 저장합니다. 그런 다음 해당 열에서 **분석 > 분포**를 사용하여 정규 곡선이 있는 히스토그램 및 정규 분위수 그림을 검토합니다. 잔차가 완전히 독립적이지는 않지만 심각한 비정규 분포를 식별할 수 있습니다.

- 연속형 반응을 위한 최적의 종합 진단 도구는 각 점이 각 가설 검정에 미치는 영향을 보여 주는 레버리지 그림입니다. 데이터에 잘못된 값이 있다고 의심되는 경우 이 그림은 통계 검정이 단일 점의 영향을 심하게 받는지 여부를 판별하는 데 도움이 됩니다.
- 데이터에서 이상치를 스캔하여 올바른 관측값인지 여부를 검토하는 것이 좋습니다. 분포 플랫폼 보고서 및 그림에서 단변량 이상치를 발견할 수 있습니다. 이변량 이상치는 X 로 Y 적합 산점도 및 다변량 산점도 행렬에 나타납니다. 삼변량 이상치는 **그래프 > 3D 산점도**에서 생성된 3 차원 그림에서 확인할 수 있습니다. 더 높은 차원의 이상치는 주성분 또는 3D 산점도를 사용하거나, 다변량 플랫폼에서 계산 및 표시된 Mahalanobis 거리와 잭나이프 거리를 사용하여 찾을 수 있습니다.

대체 방법

통계 문헌에 특수한 비모수 방법 및 로버스트 방법이 설명되어 있지만 현재 JMP에서는 일부만 구현됩니다. 이러한 방법은 분포 가정이 더 적어야 하고(비모수) 오염의 영향을 덜 받습니다(로버스트). 그러나 일반적인 방법론적 접근법에는 도움이 되지 않고 검정 통계량에 대한 표본 확률이 작아 계산에 시간이 많이 걸릴 수 있습니다.

선형 순위 검정에 관심이 있고 정규 대표본 유의성 근사만 필요한 경우 데이터 순위를 분석하여 Wilcoxon 순위합 검정 또는 Kruskal-Wallis 일원 검정과 동등한 작업을 수행할 수 있습니다.

연속형 반응이 정규 가정을 적절하게 충족하는지 확실치 않은 경우, 모델링 유형을 연속형에서 순서형으로 변경하면 풍부한 표현 방식과 통계 검정력을 일부 포기하더라도 안전하게 분석할 수 있습니다.

핵심 통계적 개념

전통적 통계를 통합하고 통계 특성 및 적합 원칙을 시각화 가능한 형식으로 캡슐화하는 두 가지 핵심 개념이 있습니다.

- 불확도 통합 개념
- 두 개의 기본 적합 기계

이 두 가지 아이디어는 수리 통계학의 기초에 근거한 직관적 개념으로 통계를 이해하는 데 도움이 됩니다.

통계학과 과학의 관계는 회계와 비즈니스의 관계를 생각하면 됩니다. 관측한 증거의 중요도를 따져 보고 균형을 맞추기 위한 기술입니다. 통계 검정은 신용 감사와 같습니다. 그러나 통계 도구는 그 이상의 것을 가능하게 합니다. 데이터에 대해 예상치 못한 상황을 보여 주고 흥미로운 새로운 아이디어를 이끌어 낼 수 있는 발견 도구입니다. 이러한 강력한 도구를 사용하려면 먼저 도구의 작동 방식을 이해해야 합니다.

불확도 , 통합 개념

회계 작업 시 요약을 위해 자금을 합칩니다. 불확도 또는 잡음이 있는 상태에서 과학적 관측값을 검토하는 경우 데이터 요약을 위해 통계적 측정이 필요합니다. 돈이 가산되는 것처럼 적절한 측도를 선택하면 불확도는 가법성을 나타냅니다.

결합 확률을 얻으려면 관측값이 독립적이라고 가정한 후 확률을 더하는 것이 아니라 곱해야 하므로 직접 확률이 최적의 측도가 아닙니다. 각 확률의 로그를 사용하면 이를 합산할 수 있으며 합계는 결합 확률의 로그이므로 이 방법이 더 쉽습니다.

그러나 확률 로그는 0과 1 사이에 있는 숫자의 로그이므로 음수입니다. 양수를 유지하기 위해 JMP는 확률의 음의 로그를 사용합니다. 확률이 작아질수록 음의 로그가 커집니다. 이 측도를 불확도라고 하고 확률과 반대로 측정됩니다.

비즈니스에서는 수익을 최대화하고 비용을 최소화하려고 합니다. 과학에서는 불확도를 최소화하려고 합니다. 과학의 불확도는 비즈니스의 비용과 같은 역할입니다. 모든 통계 방법은 불확도가 최소화되도록 모형을 적합시킵니다.

불확도를 시각화하는 것은 어렵지 않습니다 . 동전 던지기가 각각 독립적일 때 일련의 동전 던지기를 생각해 보십시오 . 앞면을 던질 확률은 0.5 이고 밑이 2 인 로그의 -log(0.5) 는 1 입니다 . 연속해서 *h* 번 앞면을 던질 확률은 다음과 같이 정의됩니다 .

$$p = \left(\frac{1}{2}\right)^h$$

h 를 계산하면 다음과 같이 됩니다 .

$$h = -\log_2 p$$

어떤 사건의 불확도를 연속으로 "앞면"을 던져야 하는 횟수로 생각하여 동일한 희귀 사건을 가져올 수 있습니다.

통계적으로 수행하는 거의 모든 일에는 핵심에 불확도(-log*p*)가 있습니다. 통계 문헌에서는 불확도를 **음의 로그 가능도**라고 합니다.

두 개의 기본 적합 기계

통계적 적합에 대한 놀라운 사실은 대부분의 전통적 설계 방법이 스프링과 압력 실린더라는 두 가지 간단한 기계를 사용하는 것으로 축소된다는 것입니다.

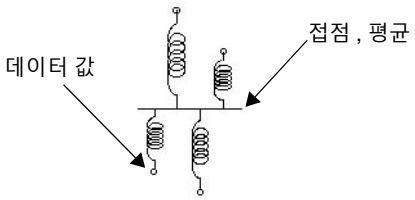
스프링

먼저, 스프링은 연속형 반응 모형에 대한 적합 기계입니다(Farebrother 1987). *n* 개의 점이 있고 점의 기대값(평균)을 알고 싶다고 가정해 보겠습니다. 점을 배치한 후 스프링으로 공통 접합부에 점을 연결하면 어떻게 될지 예상해 봅니다(그림 A.7). 손을 놓으면 스프링이 접점을 위아래로 움직이다가 중간 지점에서 멈춥니다. 이것은 물리학에 의해 발생하는 현상입니다.

스프링이 부착된 접점을 평균으로 데이터가 정규 분포를 따르는 경우 각 점의 스프링에서 발생하는 물리적 에너지는 데이터 점의 불확도에 비례합니다. 스프링의 에너지, 즉 불확도를 계산하려면 각 점에서 평균까지의 거리에 대한 제곱합을 계산해야 합니다.

관측된 데이터에 대한 불확도를 최소화하는 추정값을 선택하기 위해 스프링 안정점이 평균 추정값으로 선택됩니다. 이 점은 스프링을 늘리기 위해 최소 에너지가 필요한 지점이며 최소 제곱 적합과 동등합니다.

그림 A.7 데이터 점에 스프링 연결

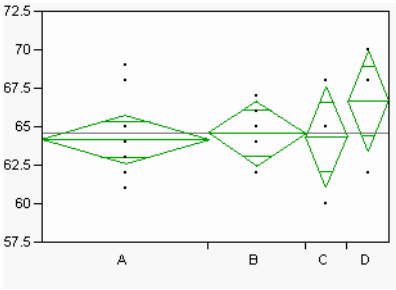


이 방법으로 하나 이상의 평균을 적합시킵니다. 또한 선, 평면 또는 초평면을 적합시킵니다. 이 방법은 연속형 데이터에 대한 거의 모든 모형을 적합시킬 수 있습니다. 스프링을 늘려야 하는 거리의 제곱합으로 에너지 또는 불확도를 측정합니다.

통계학자들은 최소한의 신뢰를 요구하는 정규 분포를 믿습니다. 어떤 의미에서는 가장 랜덤하다고 할 수 있습니다. 정규 분포의 모양은 분포에 대해 가장 무정보적입니다. 즉, 주어진 분산에 대해 예상되는 불확도가 가장 높은 분포입니다. 이 분포의 불확도는 제곱 거리로 측정됩니다. 대부분의 경우 분포가 혼합되어 있거나 독립 통계량의 합이 있을 때 이 분포가 극한 분포입니다. 이 분포를 사용하면 검정 통계량을 상당히 쉽게 측정할 수 있습니다.

가설에 의해 적합이 제약되는 경우 동일한 스프링 에너지를 측정하여 가설을 검정합니다. 실험에서 네 가지 서로 다른 처리의 반응이 있고 평균이 유의하게 다른지 검정하려는 경우를 가정해 보겠습니다. 먼저 그림 A.8에 나오는 것처럼 데이터가 그룹으로 표시되었지만 각 처리에 대한 개별 평균에 스프링이 연결되어 있다고 생각해 봅니다. 그런 다음 스프링력에 압력을 가하여 개별 평균을 공통 평균으로 이동합니다. 결과는 어떻게 될까요? 평균이 같아지도록 누르는 에너지의 양이 바로 필요한 검정 통계량입니다. 해당 에너지는 평균이 같은지 여부를 검정하는 가설에 대한 *F*-검정의 주요 성분입니다.

그림 A.8 연속형 반응 변수에 대한 일원 그림



압력 실린더

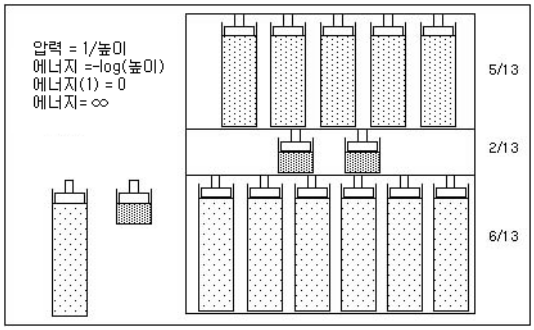
반응이 연속형 대신 범주형이면 어떻게 될까요? 예를 들어 반응이 자동차 표본의 생산 국가인 경우를 가정해 보겠습니다. 표본에 대해 American, European, Japanese 라는 세 가지 반응 수준의 확률이 있습니다. 생산 국가에 대한 이러한 확률을 몇 가지 추정값으로 설정한 후 데이터의 불확도를 평가할 수 있습니다. 이 불확도는 데이터에 의해 제공된 반응 확률의 음의 로그를 합산하여 구합니다. 이 행렬은 다음과 같이 정의됩니다.

$$H = \sum h_{y(i)} = -\sum \log p_{y(i)}$$

스프링 개념은 평균이 연속형 데이터에 어떻게 적합되는지 보여 줍니다. 반응이 범주형인 경우 통계 방법은 반응 확률을 직접 추정하고 데이터의 전체 불확도를 최소화하는 추정값을 선택합니다. 확률 추정값은 음수가 아니고 합이 1이어야 합니다. 반응 확률을 총 길이가 1인 척도의 구성 요소로 생각할 수 있습니다. 각 반응 관측값의 경우, 타이어 펌프와 같은 가스 압력 실린더를 반응 영역에 넣습니다. 최저 위치 에너지의 평형에 도달할 때까지 반응 수준 간의 파티션을 다르게 변경합니다. 그러면 결과 파티션의 크기를 사용하여 반응 확률이 추정됩니다.

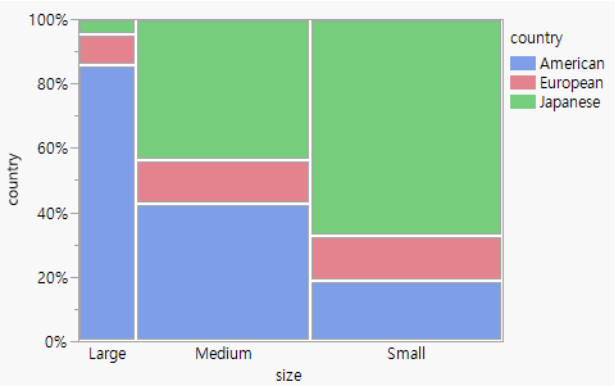
그림 A.9에서는 중형(medium) 자동차와 같은 단일 범주의 상황을 보여 줍니다(그림 A.10에 medium 라벨이 지정된 Carpoll.jmp의 모자이크 열 참조). 13개의 반응(자동차)이 있다고 가정합니다. 첫 번째 수준(American)의 반응은 6개이고 다음 수준과 마지막 수준의 반응은 각각 2개, 5개입니다. 총 에너지를 최소화하도록 반응 파티션에 대한 압력이 균형을 이루기 때문에 반응 확률은 각각 6/13, 2/13, 5/13가 됩니다.

그림 A.9 파티션에서 압력 실린더의 효과



연속형 데이터에 대한 스프링과 마찬가지로 일부 요인에 따라 표본을 나누고 개별 파티션 집합을 적합시킬 수 있습니다. 그런 다음 파티션을 동일하게 만드는 데 필요한 추가 에너지 양을 측정하여 반응률이 그룹 전체에서 같은지 검정합니다. 자동차 크기별로 그룹화된 자동차 생산 국가 확률에 대한 압력 실린더를 상상해 봅니다. 각 그룹의 파티션을 수평으로 정렬하는 데 필요한 에너지로 변수의 확률이 같은지 여부를 검정합니다. 그림 A.10에서는 이러한 파티션을 보여 줍니다.

그림 A.10 범주형 데이터에 대한 모자이크 그림



가능도 , AICc 및 BIC

JMP 의 많은 통계 모형은 **최대 가능도**라는 기법을 사용하여 적합됩니다 . 이 기법은 가능도 함수를 최대화하여 모형의 모수 (일반적으로 β 로 표시) 를 추정합니다 . L(β) 로 표시되는 가능도 함수는 관측된 데이터 값에서 평가된 확률 밀도 함수 또는 확률 질량 함수 (이산형 분포의 경우) 의 곱입니다 . 최대 가능도 추정은 관측된 데이터가 주어지면 모수 β 에 대해 L(β) 을 최대화하는 값을 찾습니다 .

가능도 함수 L(β) 을 최대화하는 것보다 가능도 함수의 음의 자연 로그 *-Log* L(β) 을 사용하는 것이 더 편리합니다. L(β) 최대화 문제는 음의 로그 가능도를 최소화하려는 최소화 문제로 다시 작성됩니다(*-1**로그 가능도 = *-Log* L(β)). 따라서 음의 로그 가능도 값이 더 작거나 음의 로그 가능도의 두 배(*-2**로그 가능도)인 경우 모형이 더 적합하다는 것을 나타냅니다.

음의 로그 가능도 값을 사용하여 모형을 선택하고 JMP 의 여러 플랫폼에서 생성된 모형 적합을 비교하는 사용자 가설 검정을 수행할 수 있습니다. 이 작업은 가능도비 검정을 사용하여 수행됩니다. 여러 JMP 플랫폼에서 *-2**로그 가능도*가 보고되는 한 가지 이유는 전체 모형과 축소 모형 *-2**로그 가능도* 값의 차이가 점근적으로 카이제곱 분포를 따르기 때문입니다. 이 가능도비 검정과 관련된 자유도는 두 모형의 모수 수 차이와 같습니다(Wilks 1938).

수정된 AICc(Akaike 정보 기준) 및 BIC(Bayes 정보 기준)는 정보를 기반으로 모형 적합을 평가하는 기준입니다. 둘 다 *-2**로그 가능도*를 기반으로 합니다.

AICc 는 다음과 같이 정의됩니다 .

$$\text{AICc} = -2LogLikelihood + 2k + 2k(k+1)/(n-k-1)$$

여기서 k는 모형의 추정된 모수 수를 나타내고 n은 모형에 사용된 관측값 수를 나타냅니다. 이 값은 동일한 데이터 집합에 대한 여러 모형을 비교하여 가장 적합한 모형을 결정하는 데 사용할 수 있습니다. Akaike(1974)에 설명된 대로 값이 가장 작은 모형이 일반적으로 선호됩니다.

BIC 는 다음과 같이 정의됩니다 .

$$\text{BIC} = -2LogLikelihood + k\ln(n)$$

여기서 k는 모형의 추정된 모수 수를 나타내고 n은 모형에 사용된 관측값 수를 나타냅니다. 두 모형의 BIC 값을 비교할 때 BIC 값이 작은 모형이 더 나은 것으로 간주됩니다.

일반적으로 BIC는 모수 수가 많은 모형에 AICc보다 벌점을 더 많이 부과합니다. 이러한 이유로 AICc와 비교하여 더 간결한 모형, 즉 모수 수가 더 적은 모형을 선택하게 됩니다. AICc와 BIC의 자세한 비교는 Burnham and Anderson 연구 자료(2004)에서 확인하십시오.

최소 제곱 회귀의 단순화된 AICc 및 BIC 계산식

최소 제곱 회귀의 경우 AICc 및 BIC 를 SSE(오차 제곱합) 에 따라 계산할 수도 있습니다 . SSE 를 기준으로 AICc 및 BIC 는 다음과 같이 정의됩니다 .

$$\text{AICc} = n\ln\left(\frac{SSE}{n}\right) + 2k + 2k(k+1)/(n-k-1) + n\ln(2\pi) + n$$

$$\text{BIC} = n\ln\left(\frac{SSE}{n}\right) + k\ln(n) + n\ln(2\pi) + n$$

여기서 k는 모형의 추정된 모수 수, n은 모형에 사용된 관측값 수, SSE는 모형의 오차 제곱합을 나타냅니다.

검정력 계산

다음 섹션에서는 LSN(최소 유의 개수), LSV(최소 유의값), 검정력 및 조정 검정력을 계산하는 데 사용되는 계산식을 제공합니다 . LSV 를 제외하고 각 효과 및 사용자 지정 대비 모음 (" 사용자 검정 " 및 " 최소 제곱 평균 대비 ") 에 대해 이러한 계산이 제공됩니다 . LSV 는 단일 선형 대비에 대해서만 계산됩니다 . 아래 상세 정보에서 **가설**은 원하는 대비의 모음을 나타냅니다 .

- "LSN 계산 "
- "LSV 계산 "
- " 검정력 계산 "
- " 조정 검정력 계산 "

LSN 계산

LSN(최소 유의 개수) 은 다음 방정식에서 N 값을 구합니다 .

$$\alpha = 1 - FDist\left[\frac{N\delta^2}{df_{Hyp}\sigma^2}, df_{Hyp}, N - df_{Hyp} - 1\right]$$

다음은 각 요소에 대한 설명입니다 .

FDist - 중심 *F* 분포의 누적 분포 함수

df_{Hyp} - 가설의 자유도

σ^2 - 오차 분산

δ^2 - 효과 크기 제곱

후향적 분석의 경우 δ^2은 가설의 제곱합을 *n*, 즉 현재 표본 크기로 나눈 값으로 추정됩니다. 효과 검정인 경우 δ^2은 해당 효과의 제곱합을 현재 연구의 관측값 수로 나눈 값으로 추정됩니다. 후향적 연구의 경우 오차 분산 σ^2은 평균 제곱 오차로 추정됩니다. 이러한 추정값은 α 값 0.05 와 함께 "검정력 상세 정보 창"에 기본값으로 입력됩니다.

향후 연구를 계획하기 위해 전향적 분석을 수행할 때 지정된 검정력을 얻을 수 있는 표본 크기를 결정해야 합니다("검정력 계산"(517 페이지) 참조).

LSV 계산

LSV(최소 유의값) 는 단일 선형 대비에 대해서만 계산됩니다.

단일 선형 대비 검정

자유도가 1 인 $L\beta = 0$ 검정을 가정해 봅니다 . 여기서 L 은 상수로 구성된 행 벡터입니다 . 이 가설에 대한 *t*- 검정의 검정 통계량은 다음과 같습니다 .

$$\frac{Lb}{s\sqrt{L(X'X)^{-1}L'}}$$

여기서 s는 제곱근 평균 제곱 오차입니다. 검정 통계량의 절대값이 자유도가 오차의 자유도와 같은 t 분포의 1 - $\alpha/2$ 분위수($t_{1-\alpha/2}$)를 초과하는 경우 유의 수준 α에서 가설을 기각합니다.

$(Lb)^{LSV}$ 로 표시된 최소 유의값을 찾으려면 Lb 값을 구합니다 .

$$(Lb)^{LSV} = t_{1-\alpha/2}s\sqrt{L(X'X)^{-1}L'}$$

단일 모수 검정

선형 대비가 단일 β_i 를 0 으로 설정하는 가설을 검정하는 특수한 경우에는 계산식이 다음과 같이 축소됩니다 .

$$b_i^{LSV} = t_{1-\alpha/2} s\sqrt{(X'X)_{ii}^{-1}} = t_{1-\alpha/2} StdError(b_i)$$

평균의 차이 검정

두 그룹의 평균을 비교하는 검정에 관심이 있는 경우 문헌에서 LSD(최소 유의차) 에 대한 설명을 참조하십시오. 모형에 명목형 변수 하나만 포함된 특수한 경우 단일 선형 대비 검정을 위한 계산식이 LSD 계산식으로 축소됩니다 .

$$\text{LSD} = t_{1-\alpha/2} s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

그러나 JMP에서 명목형 효과의 수준과 관련된 모수는 해당 수준의 평균과 모든 수준의 평균 간의 차이를 측정합니다. 따라서 이러한 비교에 대한 LSV는 평균 차이에 대한 LSD의 절반입니다.

참고 : 명목형 효과 수준에서 대비를 검정하는 경우 JMP 가 명목형 효과를 코드화하는 방식에 유의하십시오 . 즉 , 주어진 수준과 관련된 모수는 모든 수준의 평균에 대한 차이를 측정합니다 .

검정력 계산

유의 수준 α 및 표본 크기 N을 기반으로 선형 가설 검정의 검정력을 계산하는 데 관심이 있다고 가정해 보겠습니다. δ 크기의 효과를 감지하려고 합니다.

검정력을 계산하려면 먼저 선형 가설의 α 수준 *F*- 검정에 대한 임계값을 찾습니다 . 이 값은 방정식에서 F_C 값을 구하여 제공됩니다 .

$$\alpha = 1 - FDist[F_C, df_{Hyp}, N - df_{Model} - 1]$$

여기서 df_{Hyp}는 가설의 자유도, df_{Model}은 모형의 자유도, N은 제안된 표본 크기 또는 실제 표본 크기를 나타냅니다.

그런 다음 원하는 효과 크기와 관련된 비중심성 모수를 계산합니다 . 비중심성 모수는 다음과 같이 정의됩니다 .

$$\lambda = (N\delta^2)/\sigma^2$$

여기서 σ^2은 제안된 또는 추정된 오차 분산 값입니다.

δ 크기의 효과가 주어지면 검정 통계량은 비중심성 모수 λ 가 있는 비중심 *F* 분포를 따릅니다. 분포 함수는 아래에 *FDist* 로 표시됩니다. 검정력을 구하려면 검정 통계량이 임계값을 초과할 확률을 계산합니다.

$$\text{검정력} = 1\text{–}FDist\left[\mathrm{F_C}, df_{Hyp}, N - df_{Model} - 1, \frac{N\delta^2}{\sigma^2}\right]$$

관측값이 n 개인 연구의 후향적 검정력을 구할 때 JMP 는 비중심성 모수 $\lambda = (n\delta^2)/\sigma^2$ 을 $\lambda = SS_{Hyp}/\hat{\sigma}^2$ 으로 추정합니다. 여기서 SS_{Hyp} 는 가설로 인한 제곱합을 나타냅니다.

조정 검정력 계산

조정 검정력 계산(Wright and O'Brien 1988)은 후향적 검정력 분석에만 관련됩니다. 조정 검정력은 모수가 표본 추정값으로 대체될 때 발생하는 양의 편향을 제거하도록 조정된 비중심성 모수 추정값을 사용하여 검정력을 계산합니다.

표본 추정값으로 δ 와 σ 를 추정하여 구한 비중심성 모수 λ 의 추정값은 다음과 같이 계산됩니다.

$$\hat{\lambda} = SS_{Hyp}/MSE$$

Wright and O'Brien(1988)에서는 다음 식을 사용하여 비중심성 모수의 비편향 추정값을 구할 수 있다고 설명합니다.

$$[\hat{\lambda}(df_{Error} - 2)/df_{Error}] - df_{Hyp} = \frac{\hat{\lambda}(N - df_{Model} - 1 - 2)}{N - df_{Model} - 1} - df_{Hyp}$$

오른쪽 표현식에서는 후향적 검정력 분석을 위해 제안된 표본 크기 N이 연구 크기 n과 다를 때 비편향 비중심성 모수의 계산을 보여 줍니다. 여기서 df_{Hyp} 는 가설의 자유도를 나타내고 df_{Model} 은 전체 모형의 자유도를 나타냅니다.

이렇게 비중심성 추정값을 조정하면 음수 값이 생성될 수 있습니다. 음수 값은 0 으로 설정되고 약간의 편향이 다시 나타납니다. 조정된 비중심성 추정값은 다음과 같습니다.

$$\hat{\lambda}_{\text{adj}} = \text{최대값}\left[0, \frac{\hat{\lambda}(N - df_{Model} - 1 - 2)}{N - df_{Model} - 1} - df_{Hyp}\right]$$

조정 검정력은 다음과 같습니다.

$$\text{검정력}_{\text{adj}} = 1\text{–}FDist[\mathrm{F_C}, df_{Hyp}, N - df_{Model} - 1, \hat{\lambda}_{\text{adj}}]$$

비중심성 모수의 신뢰 한계는 Dwass(1955) 에 설명된 대로 구성됩니다.

$$\lambda \text{ 의 신뢰 하한} = \text{최대값}\left[0, [(\sqrt{SS_{\text{Hyp}}/MSE}) - \sqrt{df_{Hyp}\mathrm{F_C}}]^2\right]$$

$$\lambda \text{의 신뢰 상한} = [(\sqrt{SS_{\text{Hyp}}/MSE}) - \sqrt{df_{Hyp}F_C}]^2$$

검정력의 신뢰 한계는 λ 에 대한 이 신뢰 한계를 다음 방정식에 대입하여 구합니다 .

$$\text{검정력} = 1 - FDist[F_C, df_{Hyp}, N - df_{Model} - 1, \lambda]$$

신뢰 한계를 사용한 역추정 예측

역추정 예측은 반응 값에서 독립 변수의 값을 추정합니다. 생물학적 분석 문제의 경우 신뢰 한계를 사용한 역추정 예측이 특히 유용합니다. JMP 에서는 연속형 및 이항 반응 모형에 대한 역추정 예측 추정값을 요청할 수 있습니다. 반응이 연속형인 경우 개별 반응 또는 기대 반응에 대한 신뢰 한계를 요청할 수 있습니다.

신뢰 한계는 다음 논리를 기반으로 한 Fieller 의 정리 (1954) 를 사용하여 계산됩니다 . 다른 회귀변수 및 반응 값이 주어지면 단일 회귀변수의 값과 신뢰 한계를 예측하는 것이 목표입니다 .

- **b** 가 모수 β 를 추정하여 **b** 가 N(β,V) 분포를 따르도록 합니다 .
- **x** 를 관심 있는 회귀변수 값으로 합니다 (*i* 번째 값 추정).
- **y** 를 반응 값으로 합니다 .

다른 모든 **x** 값이 주어질 때 $\beta' x = y$ 가 되도록 **x**[i] 값에 대한 신뢰 영역을 원합니다.

역추정 예측은 다음과 같습니다 .

$$x[i] = \frac{y - \beta'_{(i)}x_{(i)}}{\beta[i]}$$

여기서 괄호 안의 서브스크립트 (*i*) 는 *i* 번째 성분이 생략되었음을 나타냅니다 . 다음과 같은 관계에서 신뢰 구간을 구성할 수 있습니다 .

$$(y - b'x)^2 < t^2 x'Vx$$

여기서 *t* 는 지정된 신뢰 수준에 대한 *t* 값입니다.

다음 방정식은

$$(y - b'x)^2 - t^2 x'Vx = 0$$

$z = x[i]$ 에 따라 2 차로 작성할 수 있습니다 .

$$gz^2 + hz + f = 0$$

다음은 각 요소에 대한 설명입니다 .

$$g = b[i]^2 - t^2 V[i,i]$$

$$h = -2yb[i] + 2b[i]b'_{(i)}x_{(i)} - 2t^2V[i,(i)]'x_{(i)}$$

$$f = y^2 - 2yb'_{(i)}x_{(i)} + (b'_{(i)}x_{(i)})^2 - t^2x_{(i)}'V_{(i)}x_{(i)}$$

g, h 및 f 값에 따라 부등식을 충족하는 값 집합 및 역추정 예측의 신뢰 구간은 다음과 같은 여러 형식을 가질 수 있습니다.

- $\phi_1 < \phi_2$ 조건에서 (ϕ_1, ϕ_2) 형식의 구간
- $\phi_1 < \phi_2$ 조건에서 $(-\infty, \phi_1) \cup (\phi_2, \infty)$ 형식의 겹치지 않는 두 구간
- 확장된 실직선 $(-\infty, \infty)$
- $-(\infty, \phi)$ 또는 (ϕ, ∞) 중 하나만

Fieller 신뢰 구간이 확장된 실직선인 경우 Wald 구간이 제공됩니다.

참고 : X 로 Y 적합 로지스틱 플랫폼 및 모형 적합 명목형 로지스틱 분석법에서는 역추정 예측을 위한 신뢰 구간을 계산할 때 *t* 값을 사용합니다. 모형 적합 일반화 선형 모형 분석법 및 SAS/STAT 의 PROC PROBIT 에서는 서로 다른 결과를 제공하는 *z* 값을 사용합니다.

부록 B

참조 자료

Aitken, M. (1987). "Modelling Variance Heterogeneity in Normal Regression Using GLIM." *Journal of the Royal Statistical Society, Series C* 36:332–339.

Akaike, H. (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* AC-19:716–723.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.

Benjamini, Y., and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57:289–300.

Box, G. E. P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, Part 2: Effects of Inequality of Variance and of Correlation between Errors in the Two-Way Classification." *Annals of Mathematical Statistics* 25:484–498.

Box, G. E. P., and Cox, D. R. (1964). "An Analysis of Transformations." *Journal of the Royal Statistical Society, Series B* 26:211–243.

Box, G. E. P., and Meyer, R. D. (1986). "An Analysis for Unreplicated Fractional Factorials." *Technometrics* 28:11–18.

Box, G. E. P., and Meyer, R. D. (1993). "Finding the Active Factors in Fractionated Screening Experiments." *Journal of Quality Technology* 25:94–105.

Burnham, K. P., and Anderson, D. R. (2004). "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods and Research* 33:261–304.

Burnham, K. P., Andersen, D. R., and Huyvaert, K. P. (2011). "AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons." *Behavioral Ecology and Sociobiology* 65:23–35.

Candes, E., and Tao, T. (2007). "The Dantzig Selector: Statistical Estimation when p is Much Larger than n." *The Annals of Statistics* 35:2313–2351.

Carroll, R. J., and Ruppert, D. (1988). *Transformation and Weighting in Regression*. London: Chapman & Hall.

Chilès, J.-P., and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. 2nd ed. New York: John Wiley & Sons.

Cobb, G. W. (1998). *Introduction to Design and Analysis of Experiments*. New York: Springer-Verlag.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Cole, J. W. L., and Grizzle, J. E. (1966). "Applications of Multivariate Analysis of Variance to Repeated Measures Experiments." *Biometrics* 22:810–828.

Conover, W. J. (1999). *Practical Nonparametric Statistics*. 3rd ed. New York: John Wiley & Sons.

Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman & Hall.

Cook, R. D., and Weisberg, S. (1983). "Diagnostics for Heteroscedasticity in Regression." *Biometrika* 70:1–10.

Cornell, J. A. (1990). *Experiments with Mixtures*. 2nd ed. New York: John Wiley & Sons.

Cox, D. R. (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society, Series B* 34:187–220.

Cox, D. R., and Snell, E. J. (1989). *The Analysis of Binary Data*. 2nd ed. London: Chapman & Hall.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Rev. ed. New York: John Wiley & Sons.

Daniel, C. (1959). "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments." *Technometrics* 1:311–341.

Dunnett, C. W. (1955). "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50:1096–1121.

Dwass, M. (1955). "A Note on Simultaneous Confidence Intervals." *Annals of Mathematical Statistics* 26:146–147.

Efron, B. (1977). "The Efficiency of Cox's Likelihood Function for Censored Data." *Journal of the American Statistical Association* 72:557–565.

Farebrother, R. W. (1987). "Mechanical Representations of the L1 and L2 Estimation Problems." In *Statistical Data Analysis Based on L_1 Norm and Related Methods*, edited by Y. Dodge, 455–464. Amsterdam: North-Holland.

Fieller, E. C. (1954). "Some Problems in Interval Estimation." *Journal of the Royal Statistical Society, Series B* 16:175–185.

Firth, D. (1993). "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80:27–38.

Fleming, T. R., and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: John & Sons.

Goodnight, J. H. (1978). *Tests of Hypotheses in Fixed Effects Linear Models*. Technical Report R–101, SAS Institute Inc., Cary, NC.

Goodnight, J. H., and Harvey, W. R. (1978). *Least-Squares Means in the Fixed-Effects General Linear Models*. Technical Report R-103, SAS Institute Inc., Cary, NC.

Goos, P., and Jones, B. (2011). *Optimal Design of Experiments: A Case Study Approach*. Chichester, UK: John Wiley & Sons.

Greenhouse, S. W., and Geisser, S. (1959). "On Methods in the Analysis of Profile Data." *Psychometrika* 32:95–112.

Harrell, F. E. (1986). "The LOGIST Procedure." In *SUGI Supplemental Library Guide, Version 5 Edition*. Cary, NC: SAS Institute Inc.

Harvey, A. C. (1976). "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica* 44:461–465.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer-Verlag.

Hayter, A. J. (1984). "A Proof of the Conjecture That the Tukey-Kramer Method Is Conservative." *Annals of Statistics* 12:61–75.

Heinze, G., and Schemper, M. (2002). "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21:2409–2419.

Hocking, R. R. (1985). *The Analysis of Linear Models*. Monterey, CA: Brooks/Cole.

Hoenig, J. M., and Heisey, D. M. (2001). "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis." *American Statistician* 55:19–24.

Hoerl, A. (1962). "Application of Ridge Analysis to Regression Problems." *Chemical Engineering Progress* 58:54–59.

Hoerl, A., and Kennard, R. (1970). "Ridge Regression: Biased Estimation for Non-orthogonal Problems." *Technometrics* 12:55–67.

Hsu, J. C. (1992). "The Factor Analytic Approach to Simultaneous Inference in the General Linear Model." *Journal of Computational and Graphical Statistics* 1:151–168.

Hsu, J. C. (1996). Multiple Comparisons: Theory and Methods. London: Chapman & Hall.

Hu, W., Laber, E. B., Barker, C., and Stefanski, L. A. (2019. "Assessing Tuning Parameter Selection Variability in Penalized Regression." *Technometrics* 61:154–164.

Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics*. 2nd ed. John Wiley & Sons.

Hui, F., Warton, D., and Foster, S. (2015). "Tuning Parameter Selection for the Adaptive Lasso Using ERIC." *Journal of the American Statistical Association* 110:262–269.

Huynh, H., and Feldt, L. S. (1970). "Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions." *Journal of the American Statistical Association* 65:1582–1589.

Huynh, H., and Feldt, L. S. (1976). "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split Plot Designs." *Journal of Educational Statistics* 1:69–82.

Kackar, R. N., and Harville, D. A. (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models." *Journal of the American Statistical Association* 79:853–862.

Kalbfleisch, J. D., and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Kenward, M. G., and Roger, J. H. (1997). "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood." *Biometrics* 53:983–997.

Koenker, R., and Hallock, K. (2001). "Quantile Regression: An Introduction." *Journal of Economic Perspectives* 15:143–156.

Kramer, C. Y. (1956). "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." *Biometrics* 12:307–310.

Lenth, R. V. (1989). "Quick and Easy Analysis of Unreplicated Factorials." *Technometrics* 31:469–473.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for Mixed Models*. 2nd ed. Cary, NC: SAS Institute Inc.

Mallows, C. L. (1973). "Some Comments on C_p." *Technometrics* 15:661–675.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

McClave, J. T., and Dietrich, F. H. (1988). *Statistics*. San Francisco: Dellen.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd ed. London: Chapman & Hall.

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons.

Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.

Miller, A. J. (1990). *Subset Selection in Regression*. New York: Chapman & Hall.

Montgomery, D. C. (1991). "Using Fractional Factorial Designs for Robust Process Development." *Quality Engineering* 3:193–205.

Muller, K. E., and Barton, C. N. (1989). "Approximate Power for Repeated-Measures ANOVA Lacking Sphericity." *Journal of the American Statistical Association* 84:549–555. Also see "Correction to 'Approximate Power for Repeated-Measures ANOVA Lacking Sphericity'," *Journal of the American Statistical Association* 86 (1991): 255–256.

Nagelkerke, N. J. D. (1991). "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78:691–692.

Nelder, J. A., and Wedderburn, R. W. M. (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society, Series A* 135:370–384.

Nelson, F. D. (1976). "On a General Computer Algorithm for the Analysis of Models with Limited Dependent Variables." *Annals of Economic and Social Measurement* 5:493–509.

Nelson, P. R., Wludyka, P. S., and Copeland, K. A. F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Philadelphia: SIAM.

Patterson, H. D., and Thompson, R. (1974). "Maximum Likelihood Estimation of Components of Variance." In *Proceedings of the Eighth International Biometric Conference,* 197–207. Washington, DC: International Biometric Society.

Portnoy, S., and Koenker, R. (1997). "The Gaussian Hare and the Laplacian Tortoise: Computation of Squared-Error vs. Absolute-Error Estimators." *Statistical Science* 12:279–300.

Rawlings, J. O. (1988). *Applied Regression Analysis: A Research Tool*. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Ries, P. N., and Smith, H. (1963). "The Use of Chi-Square for Preference Testing in Multidimensional Problems." *Chemical Engineering Progress* 59:39–43.

Sall, J. P. (1990). "Leverage Plots for General Linear Hypotheses." *American Statistician* 44:308–315.

SAS Institute Inc. (2020a). "The GENMOD Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/genmod.pdf.

SAS Institute Inc. (2020b). "The GLM Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/glm.pdf.

SAS Institute Inc. (2020c). "Introduction to Statistical Modeling with SAS/STAT Software." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/intromod.pdf.

SAS Institute Inc. (2020d). "The MIXED Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/mixed.pdf.

Satterthwaite, F. E. (1946). "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2:110–114.

Scheffé, H. (1958). "Experiments with Mixtures." *Journal of the Royal Statistical Society, Series B* 20:344–360.

Schuirmann, D. J. (1987). "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability." *Journal of Pharmacokinetics and Biopharmaceutics* 15:657–680.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. New York: John Wiley & Sons.

Seber, G. A. F. (1984). *Multivariate Observations*. New York: John Wiley & Sons.

Singer, J. D. (1998). "Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models." *Journal of Educational and Behavioral Statistics* 23:323–355.

Snedecor, G. W., and Cochran, W. G. (1967). Statistical Methods. 6th ed. Ames: Iowa State University Press.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., Jr., and McClelland, G. (2013). "Spotlights, Floodlights, and the Magic Number Zero: Simple Effects Tests in Moderated Regression." *Journal of Marketing Research* 50:277–288.

Stone, C., and Koo, C. Y. (1985). "Additive Splines in Statistics." In *Proceedings of the Statistical Computing Section*, 45–48. Alexandria, VA: American Statistical Association.

Sullivan, L. M., Dukes, K. A., and Losina, E. (1999). "An Introduction to Hierarchical Linear Modelling." *Statistics in Medicine* 18:855–888.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society, Series B* 58:267–288.

Tukey, J. W. (1953). "The Problem of Multiple Comparisons." In *Multiple Comparisons*, 1948–1983, edited by H. I. Braun, vol. 8 of *The Collected Works of John W. Tukey* (published 1994), 1–300. London: Chapman & Hall. Unpublished manuscript.

Walker, S. H., and Duncan, D. B. (1967). "Estimation of the Probability of an Event as a Function of Several Independent Variables." *Biometrika* 54:167–179.

Westfall, P. H., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple Comparisons and Multiple Tests Using SAS.* 2nd ed. Cary, NC: SAS Institute Inc.

Wilks, S. S. (1938). "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." *Annals of Mathematical Statistics*. 9:60–62.

Wolfinger, R. D., Tobias, R. D., and Sall, J. (1994). "Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models." *SIAM Journal on Scientific Computing* 15:1294–1310.

Wright, S. P., and O'Brien, R. G. (1988). "Power Analysis in an Enhanced GLM Procedure: What it Might Look Like." In *Proceedings of the Thirteenth Annual SAS Users Group International Conference*, 1097–1102. Cary, NC: SAS Institute Inc. http://www.sascommunity.org/sugi/SUGI88/Sugi-13-220%20Wright%20OBrien.pdf.

Zou, H. (2006). "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101:1418–1429.

Zou, H., and Hastie, T. (2005). "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society, Series B* 67:301–320.

부록 C

기술 라이선스 표시

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

버전 16

예측 및 전문 모델링

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는
것이 아니라 새로운 눈을 가지는 데 있다 ."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 예측 및 전문 모델링 16. Cary, NC: SAS Institute Inc.

JMP® 16 예측 및 전문 모델링

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

2021 년 3 월

2021 년 8 월

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP 를 처음 사용하든 오랫동안 사용해왔든 , JMP 와 관련해서 배워야 하는 것은 언제나 있습니다 .

JMP.com 을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다 .

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP 를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티 , 추가 기능 및 스크립트 예를 포함한 사용자용 리소스 , 포럼 , 블로그 , 컨퍼런스 정보 등

https://www.jmp.com/getstarted

목차

예측 및 전문 모델링

1 JMP 알아보기 17

설명서 및 추가 리소스

JMP 설명서에 사용되는 서식 규칙 19
JMP 도움말 20
JMP 설명서 라이브러리 20
JMP 학습을 위한 추가 리소스 26
JMP 자습서 26
샘플 데이터 테이블 26
통계 및 JSL 용어에 대해 알아보기 27
JMP 팁 및 힌트 알아보기 27
JMP 툴팁 27
JMP 사용자 커뮤니티 27
무료 온라인 Statistical Thinking 교육 과정 28
JMP 새로운 사용자를 위한 입문 키트 28
Statistics Knowledge 포털 28
JMP 교육 과정 28
사용자가 작성한 JMP 설명서 28
JMP 시작하기 창 28
JMP 기술 지원 29

2 예측 및 전문 모델링 소개 31

모델링 기법 개요

3 신경망 33

노드와 층을 사용하여 비선형 모형 적합

신경망 플랫폼 개요 35
신경망 플랫폼 시작 35
신경망 시작 창 36
모형 시작 제어판 37
신경망의 검증 방법 38
은닉층 구조 39
부스트 40

적합 옵션 40
모형 보고서 42
훈련 및 검증 적합 측도 42
혼동 통계량 43
신경망 플랫폼 옵션 43
신경망 모형 옵션 43
신경망의 예 45

4 파티션 모형 47
결정 트리를 사용하여 데이터 탐색 및 모델링

파티션 플랫폼 개요 49
파티션 플랫폼의 예 49
파티션 플랫폼 시작 52
분할 보고서 53
컨트롤 버튼 53
범주형 반응에 대한 보고서 54
연속형 반응에 대한 보고서 56
파티션 플랫폼 옵션 59
적합 상세 정보 표시 62
수익 행렬 지정 63
결정 행렬 보고서 64
결측값 정보화 65
실제값 대 예측값 그림 66
ROC 곡선 66
향상도 곡선 68
노드 옵션 69
파티션 검증 70
분할의 추가 예 71
연속형 반응의 예 71
결측값 정보화의 예 73
수익 행렬 및 결정 행렬 보고서의 예 76
파티션 플랫폼에 대한 통계 상세 정보 79
반응 및 요인 79
분할 기준 79
결정 트리 및 붓스트랩 포레스트의 예측 확률 80

5 붓스트랩 포레스트 81
여러 트리를 평균화하여 모형 적합

붓스트랩 포레스트 플랫폼 개요 83
범주형 반응이 있는 붓스트랩 포레스트의 예 83

붓스트랩 포레스트 모형 ... 83
결측값 ... 85
연속형 반응이 있는 붓스트랩 포레스트의 예 ... 86
붓스트랩 포레스트 플랫폼 시작 ... 88
시작 창 ... 88
규격 창 ... 89
붓스트랩 포레스트 보고서 ... 91
모형 검증 집합 요약 ... 93
규격 ... 93
전체 통계량 ... 93
누적 검증 ... 94
트리별 요약 ... 95
붓스트랩 포레스트 플랫폼 옵션 ... 95

6 부스티드 트리 ... 99
각각 이전 계층에 기반한 여러 트리 계층 적합

부스티드 트리 플랫폼 개요 ... 101
범주형 반응이 있는 부스티드 트리의 예 ... 101
연속형 반응이 있는 부스티드 트리의 예 ... 103
부스티드 트리 플랫폼 시작 ... 105
규격 창 ... 106
부스티드 트리 보고서 ... 109
모형 검증 집합 요약 ... 110
규격 ... 111
전체 통계량 ... 111
누적 검증 ... 112
부스티드 트리 플랫폼 옵션 ... 112
부스티드 트리 플랫폼에 대한 통계 상세 정보 ... 114
과대적합 벌점 ... 114

7 K 최근접 이웃 ... 115
인접한 관측값을 사용하여 반응 값 예측

K 최근접 이웃 플랫폼 개요 ... 117
범주형 반응이 있는 K 최근접 이웃의 예 ... 117
연속형 반응이 있는 K 최근접 이웃의 예 ... 119
K 최근접 이웃 플랫폼 시작 ... 121
K 최근접 이웃 보고서 ... 122
연속형 반응 ... 122
범주형 반응 ... 123
K 최근접 이웃 플랫폼 옵션 ... 124

모형 적합 옵션 124

8 Naive Bayes 127

확률적 가정을 사용하여 관측값 분류

Naive Bayes 플랫폼 개요 129
Naive Bayes 의 예 129
Naive Bayes 플랫폼 시작 132
Naive Bayes 보고서 133
적합 상세 정보 보고서 133
반응 열 보고서 134
혼동 행렬 보고서 134
ROC 곡선 134
Naive Bayes 플랫폼 옵션 134
Naive Bayes 의 추가 예 136
Naive Bayes 플랫폼에 대한 통계 상세 정보 137
알고리즘 137
저장된 확률 계산식 138

9 서포트 벡터 머신 141

서포트 벡터를 사용하여 반응 범주 예측

서포트 벡터 머신 개요 143
서포트 벡터 머신의 예 143
서포트 벡터 머신 플랫폼 시작 145
서포트 벡터 머신 시작 창 146
모형 시작 제어판 147
서포트 벡터 머신 보고서 149
모형 비교 보고서 149
서포트 벡터 머신 모형 보고서 150
서포트 벡터 머신 플랫폼 옵션 152
서포트 벡터 머신 모형 보고서 옵션 152
SVM 플랫폼의 추가 예 154
연속형 반응에 대한 서포트 벡터 회귀의 예 154

10 모형 선별 157

다중 예측 모형 적합 및 비교

연속형 반응에 대한 모형 선별의 예 159
모형 선별 플랫폼 시작 160
모형 선별 보고서 163
폴드 간 요약 보고서 164
훈련, 검증 및 테스트 적합 측도 165
모형 선별 플랫폼 옵션 166

결정 임계값 보고서 168

11 모형 비교 171

적합 모형의 예측 능력 비교

모형 비교의 예 173
모형 비교 플랫폼 시작 176
모형 비교 보고서 177
모형 비교 플랫폼 옵션 179
연속형 및 범주형 반응 179
연속형 반응 179
범주형 반응 179
모형 비교의 추가 예 180

12 검증 열 생성 183

데이터를 훈련 , 검증 및 테스트 데이터 집합으로 나누기

검증 열 생성 플랫폼 개요 185
검증 열 생성의 예 185
검증 열 생성 플랫폼 시작 187
검증 열 생성 보고서 189
비율 또는 상대 비율 지정 189
절단점 설정 190
옵션 190
시작 190
검증 열 생성 플랫폼 옵션 191
검증 열 생성 플랫폼의 추가 예 191

13 계산식 저장소 195

모형 관리 및 스코어링 코드 생성

계산식 저장소 플랫폼 개요 197
계산식 저장소의 예 197
계산식 저장소 플랫폼 시작 198
예측 계산식을 계산식 저장소에 게시하는 플랫폼 198
계산식 저장소 플랫폼 옵션 199
계산식 저장소 모형 옵션 200
계산식 저장소 플랫폼에서 스코어링 코드 생성 201

14 곡선 적합 203

기본 제공 비선형 모형을 데이터에 적합

곡선 적합 플랫폼 개요 205
곡선 적합 플랫폼을 사용한 예 205
곡선 적합 플랫폼 시작 208

곡선 적합 보고서 209
곡선 적합 옵션 210
모형 비교 보고서 211
모형 적합 보고서 212
모형 적합 옵션 212
병렬 검정 214
모수 추정값 비교 216
동등성 검정 217
곡선 적합 플랫폼에 대한 통계 상세 정보 219
모형 계산식 219

15 비선형 회귀 227
데이터에 사용자 비선형 모형 적합

사용자 모형 적합의 예 229
비선형 플랫폼 시작 232
비선형 적합 보고서 233
비선형 플랫폼 옵션 237
모형 라이브러리를 사용하여 계산식 생성 240
비선형 모형 라이브러리 사용자 정의 242
추가 예 243
최대 가능도의 예 : 로지스틱 회귀 243
이항 오차가 있는 프로빗 모형의 예 : 수치 미분 245
Poisson 손실 함수의 예 246
모수 한계 설정의 예 248
비선형 플랫폼에 대한 통계 상세 정보 250
프로파일 가능도 신뢰 한계 250
사용자 손실 함수의 작동 방식 251
도함수 관련 참고 사항 253
효과적인 비선형 모델링에 대한 참고 사항 254

16 함수 데이터 탐색기 257
함수 데이터 분석 수행

함수 데이터 탐색기 플랫폼 개요 259
함수 데이터 탐색기의 예 259
함수 데이터 탐색기 플랫폼 시작 263
함수 데이터 탐색기 보고서 265
모형 보고서 266
함수 데이터 탐색기 플랫폼 옵션 270
함수 데이터 탐색기 그룹 옵션 271
데이터 처리 보고서 옵션 272

모형 보고서 옵션 ... 275
함수 데이터 탐색기 플랫폼의 추가 예 ... 276
여러 함수 공정의 예 ... 276
함수 DOE 의 예 ... 279
함수 데이터 탐색기 플랫폼에 대한 통계 상세 정보 ... 282
함수 모형 적합 ... 282
함수 요약 상세 정보 ... 283

17 가우스 과정 ... 285
평활 모형을 사용하여 데이터 적합

가우스 과정의 예 ... 287
가우스 과정 플랫폼 시작 ... 289
가우스 과정 보고서 ... 290
실제값 대 예측값 그림 ... 290
모형 보고서 ... 290
주변 모형 그림 ... 291
가우스 과정 플랫폼 옵션 ... 291
가우스 과정 플랫폼의 추가 예 ... 292
가우스 과정 모형의 예 ... 292
범주형 예측 변수가 있는 가우스 과정 모형의 예 ... 294
가우스 과정 플랫폼에 대한 통계 상세 정보 ... 296
연속형 예측 변수가 있는 모형 ... 296
범주형 예측 변수가 있는 모형 ... 297
분산 계산식 파라미터화 ... 298
모형 적합 상세 정보 ... 298

18 시계열 분석 ... 299
시계열 모형 및 전이 함수 적합

시계열 플랫폼 개요 ... 301
시계열 플랫폼의 예 ... 301
시계열 플랫폼 시작 ... 304
시계열 분석 보고서 ... 305
시계열 그래프 ... 306
시계열 기본 진단 차트 ... 306
시계열 플랫폼 옵션 ... 308
시계열 진단 ... 308
차분 및 분해 ... 309
ARIMA 및 계절 ARIMA 모형 ... 311
평활 모형 ... 313
상태 공간 평활 모형 ... 314

전이 함수 모형 314
평활 모형 규격 창 317
보고서 319
차분 보고서 319
분해 보고서 320
모형 비교 보고서 322
ARIMA 및 계절 ARIMA 모형 보고서 324
상태 공간 평활 보고서 328
전이 함수 보고서 330
스펙트럼 밀도 보고서 331
시계열 플랫폼의 추가 예 332
시간 ID 열 생성의 예 332
Box-Cox 변환을 사용한 예 337
홀드백 집합을 사용한 예 339
시계열 플랫폼에 대한 통계 상세 정보 341
스펙트럼 밀도에 대한 통계 상세 정보 341
X-11 분해에 대한 통계 상세 정보 342
지수 평활 모형에 대한 통계 상세 정보 342
ARIMA 모형에 대한 통계 상세 정보 347
전이 함수에 대한 통계 상세 정보 348

19 시계열 예측 349
다중 시계열 예측
시계열 예측의 예 351
시계열 예측 플랫폼 시작 353
데이터 형식 353
시계열 예측 보고서 354
시간 패턴 분석 354
시간에 대한 그룹 요약 355
Y 에 대한 그룹 요약 355
모델링 규격 356
시계열 예측 플랫폼 옵션 358
모형 보고서 359

20 매칭 쌍 분석 361
동일한 개체에 대한 측정값 비교
매칭 쌍 플랫폼 개요 363
매칭 쌍 비교의 예 363
매칭 쌍 플랫폼 시작 364
여러 Y 열 365

매칭 쌍 보고서 366
차이 그림 및 보고서 366
그룹 간 367
매칭 쌍 플랫폼 옵션 367
그룹 간 매칭 쌍 비교의 예 368
매칭 쌍 플랫폼에 대한 통계 상세 정보 369
매칭 쌍 그래픽 369
반응 상관 370

21 모델링 유틸리티 373
데이터에서 이상치, 결측값 및 패턴 탐색

이상치 탐색 유틸리티 375
이상치 탐색 유틸리티의 예 375
이상치 탐색 유틸리티 시작 378
분위수 범위 이상치 379
로버스트 적합 이상치 382
로버스트 PCA 이상치 384
K 최근접 이웃 이상치 387
이상치 탐색 유틸리티 옵션 389
이상치 탐색 유틸리티의 추가 예 389
결측값 탐색 유틸리티 391
결측값 탐색 유틸리의 예 391
결측값 보고서 394
결측값 탐색 유틸리티 옵션 399
패턴 탐색 유틸리티 399
패턴 탐색 유틸리티의 예 399
패턴 탐색 유틸리티 시작 402
패턴 탐색 보고서 403
패턴 탐색 유틸리티 옵션 404
패턴 탐색 유틸리티의 추가 예 407
패턴 탐색 유틸리티에 대한 통계 상세 정보 409

22 반응 변수 선별 411
대규모 데이터의 여러 반응 변수 검정

반응 변수 선별 플랫폼 개요 413
반응 변수 선별의 예 414
반응 변수 선별 플랫폼 시작 416
반응 변수 선별 보고서 418
결과 테이블 418
FDR p 값 그림 420

FDR LogWorth 대 효과 크기 421
FDR LogWorth 대 R^2 422
p 값 데이터 테이블 422
p 값 데이터 테이블 스크립트 423
반응 변수 선별 플랫폼 옵션 423
평균 데이터 테이블 424
평균 비교 데이터 테이블 425
모형 적합의 반응 변수 선별 분석법 427
모형 적합의 반응 변수 선별 시작 428
반응 변수 선별 적합 보고서 429
반응 변수 선별 적합 옵션 429
PValues 데이터 테이블 430
Y Fits 데이터 테이블 431
반응 변수 선별의 추가 예 432
실제적 유의성 및 동등성 검정의 예 432
최대 LogWorth 옵션의 예 433
로버스트 적합의 예 435
반응 변수 선별 분석법 438
반응 변수 선별 플랫폼에 대한 통계 상세 정보 439
False Discovery Rate 439

23 공정 변수 선별 441
안정성 및 공정 능력을 위한 여러 공정 변수 선별
공정 변수 선별 플랫폼 개요 443
공정 변수 선별의 예 443
공정 변수 선별 플랫폼 시작 446
시작 창 역할 446
시작 창 옵션 447
한계 테이블 448
공정 변수 선별 보고서 450
공정 변수 선별 플랫폼 옵션 455
공정 성능 그래프 460
공정 변수 선별의 추가 예 461
공정 변수 선별 플랫폼에 대한 통계 상세 정보 464
중앙값을 사용하여 시그마를 추정하기 위한 척도 인수 464

24 예측 변수 선별 467
유의한 효과를 위한 여러 예측 변수 선별
예측 변수 선별 플랫폼 개요 469
예측 변수 선별의 예 469

예측 변수 선별 플랫폼 시작 471
예측 변수 선별 보고서 472
예측 변수 선별 플랫폼 옵션 472

25 연관성 분석 473
장바구니 분석 수행
연관성 분석 플랫폼 개요 475
연관성 분석 플랫폼의 예 476
연관성 분석 플랫폼 시작 478
데이터 형식 479
연관성 분석 보고서 479
빈발 항목 집합 480
규칙 480
연관성 분석 플랫폼 옵션 481
특이값 분해 481
SVD 보고서 482
SVD 보고서 옵션 483
주제 분석 484
주제 분석 보고서 484
주제 분석 보고서 옵션 485
추가 예 : SVD 분석 485
연관성 분석 플랫폼에 대한 통계 상세 정보 488
빈발 항목 집합 생성 488
연관성 분석 성능 측도 488

26 공정 기록 탐색기 491
복잡한 공정 기록에서 문제 성분 식별
공정 기록 탐색기의 예 493
공정 기록 탐색기 플랫폼 시작 495
데이터 형식 496
공정 기록 탐색기 보고서 496
공정 기록 탐색기 플랫폼 옵션 496

A 통계 상세 정보 501
예측 및 전문 모델링
적합 측도 503
엔트로피 R^2 503
JMP 모델링의 검증 503
검증 열 역할 504
K 폴드 및 홀드백 검증 506

B 참조 자료 509
C 기술 라이선스 고지 사항 513

1 장

JMP 알아보기
설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 19
JMP 도움말 20
JMP 설명서 라이브러리 20
JMP 학습을 위한 추가 리소스 26
JMP 자습서 26
샘플 데이터 테이블 26
통계 및 JSL 용어에 대해 알아보기 27
JMP 팁 및 힌트 알아보기 27
JMP 툴팁 27
JMP 사용자 커뮤니티 27
무료 온라인 Statistical Thinking 교육 과정 28
JMP 새로운 사용자를 위한 입문 키트 28
Statistics Knowledge 포털 28
JMP 교육 과정 28
사용자가 작성한 JMP 설명서 28
JMP 시작하기 창 28
JMP 기술 지원 29

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 Lucida Sans Typewriter 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 "JMP 설명서 라이브러리 "(20 페이지) 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
예측 및 전문 모델링	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
품질 및 공정 방법	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기 " 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기 "

창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

2 장

예측 및 전문 모델링 소개

모델링 기법 개요

예측 및 전문 모델링에서는 반응 변수 선별, 분할, 신경망 등의 좀 더 전문적인 모델링 기법에 대해 자세히 설명합니다.

- 신경망 플랫폼에서는 하나 또는 두 개의 층으로 완전히 연결된 다중 계층 퍼셉트론을 구현합니다. 신경망을 사용하면 입력 변수의 유연한 함수를 사용하여 하나 이상의 반응 변수를 예측할 수 있습니다. 자세한 내용은 "신경망"에서 확인하십시오.
- 파티션 플랫폼에서는 *X* 값과 *Y* 값 사이의 관계에 따라 데이터를 재귀적으로 분할하여 파티션의 결정 트리를 생성합니다. 자세한 내용은 "파티션 모형"에서 확인하십시오.
- JMP PRO 붓스트랩 포레스트 플랫폼을 사용하면 각각 훈련 데이터의 랜덤 부분집합에 적합되는 여러 결정 트리를 평균화하여 앙상블 모형을 적합시킬 수 있습니다. 자세한 내용은 "붓스트랩 포레스트"에서 확인하십시오.
- JMP PRO 부스티드 트리 플랫폼에서는 계층에 생성된 여러 개의 더 작은 결정 트리로 구성되는 가법 결정 트리 모형을 생성합니다. 각 계층의 트리는 소수의 분할로 구성됩니다(보통 5 개 이하). 각 계층은 재귀적 적합 방법론을 사용하여 적합됩니다. 자세한 내용은 "부스티드 트리"에서 확인하십시오.
- JMP PRO K 최근접 이웃 플랫폼에서는 지정된 관측값의 국소적 근방에 있는 관측값의 반응을 사용하여 해당 관측값에 대한 반응 값을 예측합니다. 분류를 위한 범주형 반응 및 예측을 위한 연속형 반응과 함께 사용할 수 있습니다. 자세한 내용은 "K 최근접 이웃"에서 확인하십시오.
- JMP PRO Naive Bayes 플랫폼에서는 관측값을 범주형 반응 변수의 수준에 의해 정의된 그룹으로 분류합니다. 분류에 사용되는 변수 또는 요인을 데이터 마이닝 문헌에서는 흔히 피처라고 합니다. 자세한 내용은 "Naive Bayes"에서 확인하십시오.
- JMP PRO 서포트 벡터 머신 플랫폼에서는 관측값을 범주형 반응 변수의 수준에 의해 정의된 그룹으로 분류합니다. 모형은 클래스를 구분하는 초평면을 최적화하여 데이터를 분류합니다. 자세한 내용은 "서포트 벡터 머신"에서 확인하십시오.
- JMP PRO 모형 선별 플랫폼을 사용하면 여러 예측 모형을 빠르게 실행하고 결과를 비교할 수 있습니다. 중첩된 진단 그림과 함께 각 모형에 대한 적합 측도가 제공됩니다. 자세한 내용은 "모형 선별"에서 확인하십시오.
- JMP PRO 모형 비교 플랫폼을 사용하면 여러 모형의 예측 능력을 비교할 수 있습니다. 중첩된 진단 그림과 함께 각 모형에 대한 적합 측도가 제공됩니다. 자세한 내용은 "모형 비교"에서 확인하십시오.
- JMP PRO 검증 열 생성 플랫폼을 사용하면 파티션을 생성하는 다섯 가지 방법 중 하나를 사용하여 데이터를 두 개 또는 세 개의 집합으로 분할할 수 있습니다. 자세한 내용은 "검증 열 생성"에서 확인하십시오.

- JMP PRO 계산식 저장소 플랫폼을 사용하면 배포할 모형을 구성, 비교 및 프로파일링하고 스코어링할 수 있습니다. 모형 탐색 작업의 경우 계산식 저장소를 사용하여 후보 모형을 JMP 데이터 테이블 외부에 저장할 수 있습니다. 자세한 내용은 "계산식 저장소"에서 확인하십시오.
- 곡선 적합 플랫폼에서는 다항식, 로지스틱, Gompertz, 지수, 정상점 및 약동학 모형과 같은 사전 정의된 모형을 제공합니다. 다양한 분석과 그래픽 기법을 사용하여 여러 그룹 또는 개체를 비교합니다. 자세한 내용은 "곡선 적합"에서 확인하십시오.
- 비선형 플랫폼을 사용하면 모형 계산식 및 추정할 모수가 포함된 사용자 비선형 모형을 적합시킬 수 있습니다. 자세한 내용은 "비선형 회귀"에서 확인하십시오.
- 가우스 과정 플랫폼에서는 연속형 반응과 하나 이상의 연속형 예측 변수 간의 관계를 모델링합니다. 이러한 모형은 유한 요소 코드의 출력과 같은 컴퓨터 시뮬레이션 실험 분야에서 흔히 볼 수 있으며 데이터를 완벽하게 보간하는 경우가 많습니다. 자세한 내용은 "가우스 과정"에서 확인하십시오.
- JMP PRO 함수 데이터 탐색기 플랫폼을 사용하면 함수 데이터를 다른 JMP 플랫폼에서 분석할 수 있는 형식으로 변환할 수 있습니다. 자세한 내용은 "함수 데이터 탐색기"에서 확인하십시오.
- 시계열 플랫폼을 사용하면 단변량 시계열을 탐색, 분석 및 예측할 수 있습니다. 자세한 내용은 "시계열 분석"에서 확인하십시오.
- 시계열 예측 플랫폼을 사용하면 다중 시계열을 모델링하고 예측할 수 있습니다. 최대 30 개의 지수 평활 모형 집합에서 최량 적합 모형이 자동으로 선택됩니다. 자세한 내용은 "시계열 예측"에서 확인하십시오.
- 매칭 쌍 플랫폼에서는 상관관계가 있는 두 개 이상의 변수 간 평균을 비교하고 차이를 평가합니다. 자세한 내용은 "매칭 쌍 분석"에서 확인하십시오.
- 모델링 유틸리티는 데이터 분석의 데이터 정리 및 전처리 단계를 지원합니다. 각 유틸리티에는 데이터를 더 정확하게 이해할 수 있는 탐색 도구가 있습니다. 자세한 내용은 "모델링 유틸리티"에서 확인하십시오.
- 반응 변수 선별 플랫폼에서는 다수의 반응 변수에 대한 검정 수행 과정을 자동화합니다. 검정 결과와 요약 통계량이 보고서 대신 데이터 테이블에 제공되어 데이터 탐색이 가능합니다. 자세한 내용은 "반응 변수 선별"에서 확인하십시오.
- 공정 변수 선별 플랫폼을 사용하면 장기간에 걸쳐 다수의 공정을 탐색할 수 있습니다. 이 플랫폼에서는 관리도, 공정 안정성 및 공정 능력 측정 기준을 계산하고 큰 공정 변화를 감지합니다. 자세한 내용은 "공정 변수 선별"에서 확인하십시오.
- 예측 변수 선별 플랫폼을 사용하면 유의한 예측 변수에 대한 데이터 집합을 선별할 수 있습니다. 자세한 내용은 "예측 변수 선별"에서 확인하십시오.
- JMP PRO 연관성 분석 플랫폼을 사용하면 서로 선호도가 있는 항목을 식별할 수 있습니다. 트랜잭션에서 자주 함께 나타나는 항목을 식별하기 위해 트랜잭션 데이터 (장바구니라고도 함) 를 분석하는 데 흔히 사용됩니다. 자세한 내용은 "연관성 분석"에서 확인하십시오.
- 공정 기록 탐색기 플랫폼을 사용하면 복잡한 공정 기록에서 문제를 일으키는 성분을 식별할 수 있습니다. 자세한 내용은 "공정 기록 탐색기"에서 확인하십시오.

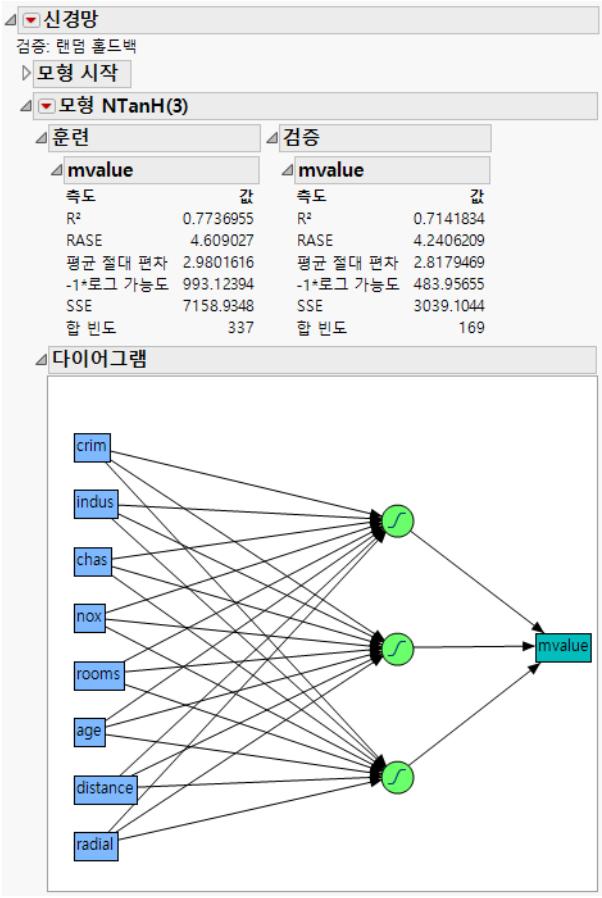
3 장

신경망
노드와 층을 사용하여 비선형 모형 적합

JMP PRO 이 플랫폼에 있는 대부분의 기능은 JMP Pro 에서만 사용할 수 있으며 그러한 경우 이 아이콘이 표시됩니다 .

신경망 플랫폼에서는 하나 또는 두 개의 층으로 완전히 연결된 다중 계층 퍼셉트론을 구현합니다. 신경망을 사용하면 입력 변수의 유연한 함수를 사용하여 하나 이상의 반응 변수를 예측할 수 있습니다. 신경망은 반응 표면의 함수 형태를 설명할 필요가 없거나 입력과 반응 사이의 관계를 설명할 필요가 없을 때 매우 좋은 예측 변수가 될 수 있습니다.

그림 3.1 신경망의 예



목차

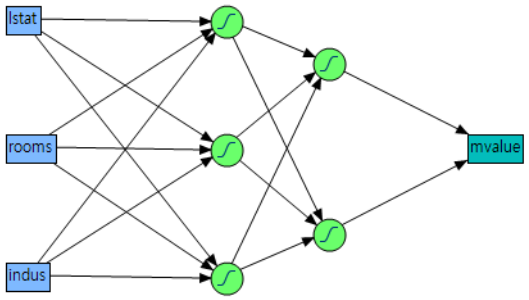
신경망 플랫폼 개요 35
신경망 플랫폼 시작 35
신경망 시작 창 36
모형 시작 제어판 37
신경망의 검증 방법 38
은닉층 구조 39
부스트 40
적합 옵션 40
모형 보고서 42
훈련 및 검증 적합 측도 42
혼동 통계량 43
신경망 플랫폼 옵션 43
신경망 모형 옵션 43
신경망의 예 45

신경망 플랫폼 개요

신경망을 은닉 노드라고 하는 파생 입력 집합의 함수로 생각할 수 있습니다. 은닉 노드는 원래 입력의 비선형 함수입니다. 최대 두 개의 은닉 노드 층을 지정할 수 있으며 각 층은 은닉 노드를 원하는 만큼 포함할 수 있습니다.

그림 3.2에서는 세 개의 X 변수와 하나의 Y 변수로 구성된 2층 신경망을 보여 줍니다. 이 예에서 첫 번째 층에는 두 개의 노드가 있고 각 노드는 두 번째 층에 있는 세 가지 모든 노드의 함수입니다. 두 번째 층에는 세 개의 노드가 있고 모든 노드는 세 가지 X 변수의 함수입니다. 예측된 Y 변수는 첫 번째 층에 있는 두 노드의 함수입니다.

그림 3.2 신경망 다이어그램



은닉층 노드에 적용되는 함수를 활성 함수라고 합니다. 활성 함수는 X 변수의 선형 결합을 변환한 것입니다. 활성 함수에 대한 자세한 내용은 "은닉층 구조"에서 확인하십시오. 반응에 적용되는 함수는 선형 결합(연속형 반응의 경우) 또는 로지스틱 변환(명목형 또는 순서형 반응의 경우)입니다.

신경망 모형의 주요 장점은 다양한 반응 표면을 효율적으로 모델링할 수 있다는 것입니다. 은닉 노드와 층이 충분하면 모든 표면을 원하는 정확도에 맞게 근사화할 수 있습니다. 신경망 모형의 주요 단점은 결과를 쉽게 해석할 수 없다는 것입니다. 정규 회귀 분석의 경우처럼 X 변수에서 Y 변수로의 직접 경로가 아니라 중간 층이 있기 때문입니다.

신경망 플랫폼 시작

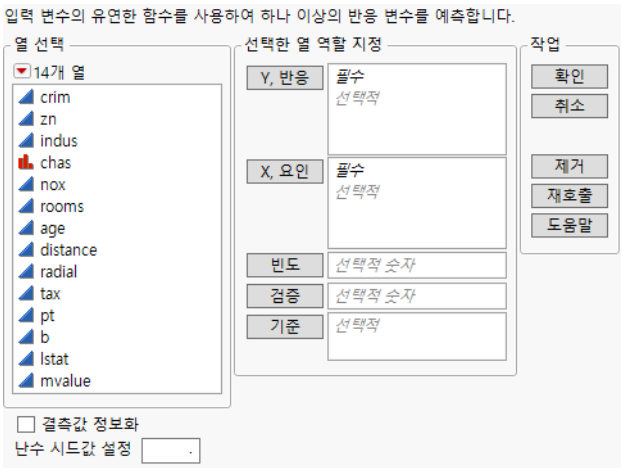
신경망 플랫폼을 시작하려면 **분석** > **예측 모델링** > **신경망**을 선택합니다.

신경망 플랫폼 시작은 2단계 프로세스입니다. 먼저 신경망 시작 창에 변수를 입력합니다. 그런 다음 "모형 시작" 제어판에서 옵션을 지정합니다.

JMP PRO 신경망 시작 창

신경망 시작 창을 사용하여 X 변수, Y 변수 및 검증 열을 지정하고 결측값 정보화 값 코딩을 활성화할 수 있습니다.

그림 3.3 신경망 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 반응 분석할 하나 이상의 반응 변수입니다. 반응을 여러 개 지정하면 반응에 대한 모형이 은닉층의 모든 모수(반응에 연결되지 않은 모수)를 공유합니다. 입력 변수가 주어지면 반응 변수는 조건부로 독립적이지만 이러한 변수를 통해 약간의 상관관계를 유지하여 하나의 전체 신경망 모형을 생성합니다.

X, 요인 예측 변수입니다.

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다.

JMP PRO **검증** 검증 데이터 집합을 정의하는 숫자 열입니다. 자세한 내용은 "신경망의 검증 방법"에서 확인하십시오. "열 선택" 목록에서 아무 열도 선택하지 않은 상태로 "검증" 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다. "검증 열 생성" 유틸리티에 대한 자세한 내용은 "검증 열 생성"에서 확인하십시오.

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다. 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다. 결과는 개별 보고서에 제공됩니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

JMP PRO **결측값 정보화** 결측값 정보화 코딩을 활성화하려면 이 상자를 선택합니다 . 이 코딩을 사용하면 결측값이 있어도 예측 모형을 추정할 수 있습니다 . 이 옵션은 결측 데이터에 정보가 포함된 경우 유용합니다 . 이 옵션을 선택하지 않으면 결측값이 있는 행이 무시됩니다 .

연속형 변수의 경우 결측값은 변수의 평균으로 대체됩니다 . 또한 < 열 이름 > 결측됨이라는 결측값 표시자가 생성되어 모형에 포함됩니다 . " 모형 시작 " 제어판의 " 공변량 변환 " 적합 옵션을 사용하여 변수를 변환하는 경우 결측값은 변환된 변수의 평균으로 대체됩니다 .

범주형 변수의 경우 결측값은 해당 변수에 대한 별도의 수준으로 처리됩니다 .

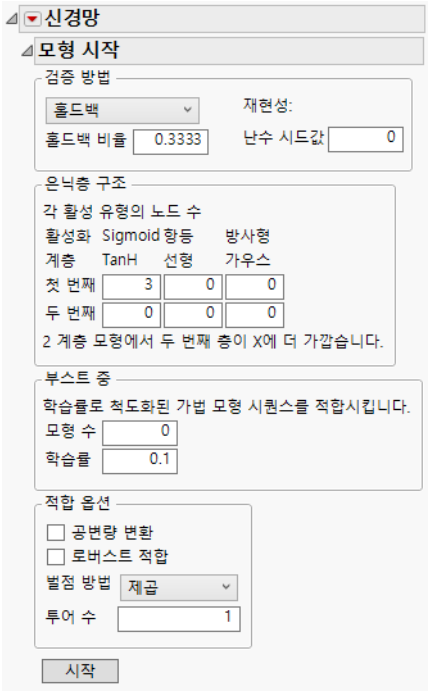
난수 시드값 설정 적합 절차에 사용되는 시작 값의 시드를 설정합니다 . " 난수 시드값 설정 " 은 분석을 재현하려는 경우에 유용합니다 . 난수 시드값을 설정하고 스크립트를 저장하면 시드값이 스크립트에 자동으로 저장됩니다 .

모형 시작 제어판

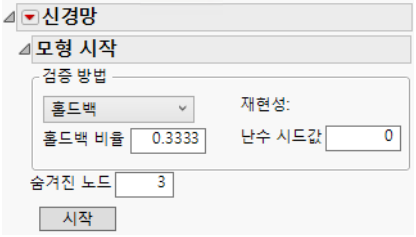
" 모형 시작 " 제어판을 사용하여 검증 방법 , 은닉층 구조 , 그래디언트 부스트 사용 여부 및 기타 적합 옵션을 지정할 수 있습니다 .

그림 3.4 모형 시작 제어판

JMP Pro



JMP



검증 방법 모형 검증에 사용할 방법을 선택합니다 . 자세한 내용은 " 신경망의 검증 방법 " 에서 확인하십시오 .

난수 시드값 이후의 신경망 플랫폼 실행을 위해 검증 할당을 재현하려면 0 이 아닌 숫자 난수 시드값을 지정합니다 . 기본적으로 난수 시드값은 0 으로 설정되어 재현 가능한 결과를 생성하지 않습니다 . 분석을 스크립트에 저장하면 사용자가 입력한 난수 시드값이 스크립트에 저장됩니다 .

은닉층 구조 또는 은닉 노드 각 층에서 각 유형의 은닉 노드 수를 지정합니다 . 자세한 내용은 " 은닉층 구조 " 에서 확인하십시오 .

참고 : JMP Standard Edition 에서는 TanH 활성 함수만 사용되며 하나의 은닉층이 있는 신경망만 적합시킬 수 있습니다 .

JMP PRO **부스트 중** 그래디언트 부스트 옵션을 지정합니다 . 자세한 내용은 " 부스트 " 에서 확인하십시오 .

JMP PRO **적합 옵션** 변수 변환 및 모형 적합 옵션을 지정합니다 . 자세한 내용은 " 적합 옵션 " 에서 확인하십시오 .

시작 신경망 모형을 적합시키고 모형 보고서를 표시합니다 .

시작을 클릭하여 모형을 적합시킨 후 "모형 시작" 제어판을 다시 열고 다른 모형 적합을 위해 설정을 변경할 수 있습니다 .

신경망의 검증 방법

신경망은 적합 루틴에 검증 방법을 활용합니다 . 검증 방법 선택에는 홀드백 , K 폴드 또는 검증 열 사용 (JMP Pro) 이 포함됩니다 . 신경망 플랫폼에서는 다음을 수행하여 모형을 적합시킵니다 .

- 모형 모수에 벌점 적용
- 검증 데이터 집합을 사용하여 모수에 대한 벌점 조정

다음 검증 방법 중 하나를 선택합니다 .

제외된 행 홀드백 행 상태를 사용하여 데이터의 부분집합을 생성합니다 . 제외되지 않은 행은 훈련 데이터 집합으로 사용되고 제외된 행은 검증 데이터 집합으로 사용됩니다 .

행 상태 사용 및 행 제외 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

홀드백 원래 데이터를 무작위로 훈련 데이터 집합과 검증 데이터 집합으로 나눕니다 . 원래 데이터에서 검증 데이터 집합으로 사용할 비율을 지정합니다 (홀드백). 랜덤 선택은 단순 랜덤 표집을 기반으로 할 때보다 더 균형 있는 훈련 데이터 집합 및 검증 데이터 집합을 생성하기 위해 모형 요인에 대한 층화 표집을 기반으로 합니다 .

K 폴드 데이터를 K 개의 부분집합 또는 폴드로 나눕니다 . 그러면 K 개의 각 폴드가 나머지 데이터에 대한 모형 적합을 검증하는 데 사용되어 총 K 개 모형을 적합시킵니다 . 최종 모형은 최량 검증 통계량을 제공하는 모형을 생성한 폴드에 대한 모형입니다 .

이 방법은 제한된 양의 데이터를 효율적으로 사용하기 때문에 소규모 데이터 집합에 유용합니다 .

JMP PRO **검증 열** 검증 데이터 집합을 정의하는 숫자 열을 사용합니다 . 다음과 같이 열의 값에 따라 데이터 분할 방법이 결정됩니다 .

- 검증 열에 두 개의 수준이 있는 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 검증 열에 세 개의 수준이 있는 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 검증 열에 4 개 이상의 고유 값이 있는 경우 K 폴드 검증이 수행됩니다 . 폴드 수는 검증 열의 값 수에 따라 결정됩니다 .

신경망 플랫폼에서는 검증 열을 사용하여 모형을 훈련 및 조정하거나 모형을 훈련 , 조정 및 평가합니다 . 검증에 대한 자세한 내용은 "JMP 모델링의 검증 " 에서 확인하십시오 .

은닉층 구조

참고 : JMP Standard Edition 에서는 TanH 활성 함수만 사용되며 하나의 은닉층이 있는 신경망만 적합시킬 수 있습니다 .

신경망 플랫폼은 층이 하나 또는 두 개인 신경망을 적합시킬 수 있습니다. 첫 번째 층의 노드 수를 늘리거나 두 번째 층을 추가하면 신경망이 더 유연해집니다. 각 층에 노드를 무제한으로 추가할 수 있습니다. 두 번째 층의 노드는 X 변수의 함수입니다. 첫 번째 층의 노드는 두 번째 층 노드의 함수입니다. Y 변수는 첫 번째 층 노드의 함수입니다.

주의 : 층이 두 개인 신경망에서는 부스트를 사용할 수 없습니다 . 두 번째 층에 0 이 아닌 값을 지정하고 부스트도 지정하면 두 번째 층이 무시됩니다 .

은닉층 노드에 적용되는 함수를 활성 함수라고 합니다 . 활성 함수는 X 변수의 선형 결합을 변환한 것입니다 . 다음 활성 함수를 사용할 수 있습니다 .

TanH 쌍곡선 탄젠트 함수는 시그모이드 함수입니다 . TanH 는 값을 -1 에서 1 사이로 변환하며 로지스틱 함수의 중심화 및 척도화된 버전입니다 . 쌍곡선 탄젠트 함수는 다음과 같습니다 .

$$\frac{e^{2x}-1}{e^{2x}+1}$$

여기서 x 는 X 변수의 선형 결합입니다 .

선형 항등 함수입니다 . X 변수의 선형 결합이 변환되지 않습니다 .

선형 활성 함수는 비선형 활성 함수 중 하나와 함께 사용되는 경우가 가장 흔합니다 . 이 경우 선형 활성 함수가 두 번째 층에 배치되고 비선형 활성 함수가 첫 번째 층에 배치됩니다. 이 방법은 먼저 X 변수의 차원을 축소한 후 Y 변수에 대한 비선형 모형을 사용하려는 경우에 유용합니다 .

연속형 Y 변수의 경우 선형 활성 함수만 사용하면 Y 변수의 모형이 X 변수의 선형 결합으로 축소됩니다 . 명목형 또는 순서형 Y 변수의 경우에는 모형이 로지스틱 회귀로 축소됩니다 .

가우스 가우스 함수입니다 . 반응 표면이 가우스 (정규) 형상일 때 또는 방사 기저 함수 동작에 대해 이 옵션을 사용합니다 . 가우스 함수는 다음과 같습니다 .

$$e^{-x^2}$$

여기서 x 는 X 변수의 선형 결합입니다 .

부스트

JMP PRO 부스트는 일련의 작은 모형을 적합시킨 후 이를 누적하여 더 큰 신경망 모형을 생성하는 프로세스입니다. 작은 모형은 각각 이전 모형의 척도화된 잔차에 대해 적합됩니다. 모형이 결합되어 더 큰 최종 모형이 생성됩니다. 이 프로세스는 검증을 사용하여 지정된 모형 수를 초과하지 않고 적합시킬 성분 모형 수를 평가합니다.

부스트는 큰 모형 하나를 적합시키는 것보다 빠른 경우가 많습니다. 그러나 기준 모형은 1 ~ 2 개 노드가 있는 단일층 모형이어야 합니다. 많은 수의 모형을 지정하면 적합 속도가 빠르다는 이점이 없어질 수 있습니다.

"모형 시작" 제어판의 "부스트" 패널을 사용하여 성분 모형 수와 학습률을 지정할 수 있습니다. "모형 시작" 제어판의 "은닉층 구조" 패널을 사용하여 기준 모형의 구조를 지정할 수 있습니다.

학습률은 $0 < r \leq 1$ 이어야 합니다. 학습률이 1 에 가까우면 최종 모형에 대한 수렴이 빨라지지만 데이터가 과대적합되는 경향도 높습니다. 적은 수의 모형이 지정된 경우 1 에 가까운 학습률을 사용하십시오 .

부스트 작동 방식을 보여 주기 위해 하나의 층과 두 개의 노드로 구성된 기준 모형을 지정하고 모형 수는 8 개라고 가정해 보겠습니다. 첫 번째 단계에서는 하나의 층과 두 개의 노드가 있는 모형을 적합시킵니다. 해당 모형의 예측값은 학습률로 척도화된 후 실제 값에서 빼서 척도화된 잔차를 구성합니다. 다음 단계에서는 하나의 층과 두 개의 노드가 있는 다른 모형을 적합시킵니다. 이때 반응 값은 이전 모형의 척도화된 잔차입니다. 이 프로세스는 8 개의 모형이 적합될 때까지 또는 다른 모형을 추가해도 검증 통계량이 개선되지 않을 때까지 계속됩니다. 성분 모형이 결합되어 더 큰 최종 모형이 생성됩니다. 이 예에서 6 개 모형이 적합된 후 중지되면 최종 모형은 하나의 층과 2 x 6 = 12 개의 노드로 구성됩니다.

주의 : 층이 두 개인 신경망에서는 부스트를 사용할 수 없습니다 . 두 번째 층에 0 이 아닌 값을 지정하고 부스트도 지정하면 두 번째 층이 무시됩니다 .

적합 옵션

JMP PRO 다음과 같은 모형 적합 옵션을 사용할 수 있습니다 .

공변량 변환 Johnson Su 또는 Johnson Sb 분포를 사용하여 모든 연속형 변수를 정규성에 가까운 변수로 변환합니다. 연속형 변수를 변환하면 이상치 또는 심하게 기울어진 분포가 미치는 부정적 영향을 줄이는 데 도움이 됩니다 . 자세한 내용은 " 신경망 모형 옵션 " 의 " 변환된 공변량 저장 " 옵션을 참조하십시오 .

로버스트 적합 최소 제곱 대신 최소 절대 편차를 사용하여 모형을 훈련합니다 . 이 옵션은 반응 이상치의 영향을 최소화하려는 경우에 유용합니다. 이 옵션은 연속형 반응에만 사용할 수 있습니다 .

벌점 방법 벌점 방법을 선택합니다 . 신경망에서 데이터가 과대적합되는 경향을 줄이기 위해 적합 프로세스에서 가능도에 대한 벌점을 통합합니다 . 자세한 내용은 " 벌점 방법 " 에서 확인하십시오 .

투어 수 반복할 때마다 모수 추정값에 대한 랜덤 시작점을 다르게 사용하여 적합 프로세스를 다시 시작할 횟수를 지정합니다 . 최량 검증 통계량을 제공하는 반복이 최종 모형으로 선택됩니다 .

벌점 방법

벌점은 $\lambda p(\beta_i)$ 이며 , 여기서 λ 는 벌점 모수이고 $p(\,)$ 는 벌점 함수라고 하는 모수 추정값의 함수입니다 . 벌점 모수의 최적값을 찾기 위해 검증이 사용됩니다 .

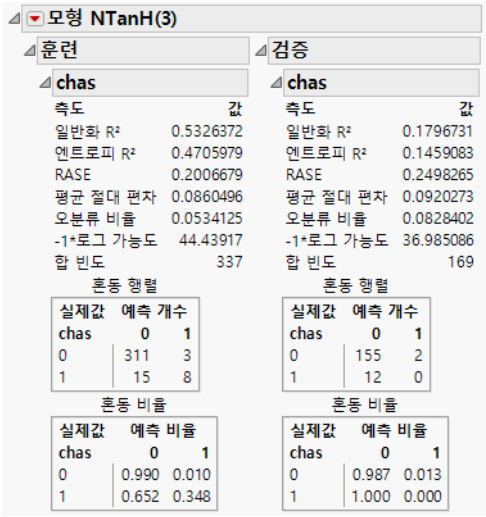
표 3.1 벌점 방법 설명

방법	벌점 함수	설명
제곱	$\sum \beta_i^2$	대부분의 X 변수가 모형의 예측 능력에 기여한다고 생각되면 이 방법을 사용합니다.
절대	$\sum \lvert\beta_i\rvert$	X 변수의 개수가 많고 이 중 몇 개가 다른 변수보다 모형의 예측 능력에 더 많이 기여한다고 생각되면 두 방법 중 하나를 사용합니다.
가중치 감소	$\sum \frac{\beta_i^2}{1+\beta_i^2}$	
벌점 없음	없음	벌점을 사용하지 않습니다. 많은 양의 데이터가 있고 적합 프로세스를 빠르게 진행하려는 경우 이 옵션을 사용할 수 있습니다. 그러나 이 옵션을 사용하면 벌점을 사용하는 모형보다 예측 성능이 낮은 모형이 생성될 수 있습니다.

모형 보고서

모든 신경망 모형에 대한 모형 보고서가 생성됩니다 . 훈련 데이터 집합과 검증 데이터 집합에 대한 적합 측도가 나타납니다 . 명목형 또는 순서형 반응의 경우 혼동 통계량도 나타납니다 .

그림 3.5 신경망 모형 보고서의 예



모형 NTanH(3)

훈련

chas

측도	값
일반화 R²	0.5326372
엔트로피 R²	0.4705979
RASE	0.2006679
평균 절대 편차	0.0860496
오분류 비율	0.0534125
-1*로그 가능도	44.43917
합 빈도	337

혼동 행렬

실제값	예측 개수	
chas	0	1
0	311	3
1	15	8

혼동 비율

실제값	예측 비율	
chas	0	1
0	0.990	0.010
1	0.652	0.348

검증

chas

측도	값
일반화 R²	0.1796731
엔트로피 R²	0.1459083
RASE	0.2498265
평균 절대 편차	0.0920273
오분류 비율	0.0828402
-1*로그 가능도	36.985086
합 빈도	169

혼동 행렬

실제값	예측 개수	
chas	0	1
0	155	2
1	12	0

혼동 비율

실제값	예측 비율	
chas	0	1
0	0.987	0.013
1	1.000	0.000

훈련 및 검증 적합 측도

훈련 데이터 집합과 검증 데이터 집합에 대한 적합 측도가 나타납니다 (그림 3.5).

일반화 R² 일반 회귀 모형에 적용할 수 있는 측도입니다 . 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다 . 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다 . 일반화 R² 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R² 로 단순화됩니다 . 일반화 R² 을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데 , 이는 Cox-Snell 유사 R^2 을 정규화한 것입니다 . 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오 .

엔트로피 R² (반응이 명목형 또는 순서형인 경우에만 표시됨) 적합 모형과 상수 확률 모형의 로그 가능도를 비교하는 적합 측도입니다 . 엔트로피 R² 범위는 0 ~ 1 이며 , 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 . 자세한 내용은 " 엔트로피 R²" 에서 확인하십시오 .

R² 모형에 대한 R² 을 제공합니다 .

RASE 제곱근 평균 제곱 오차를 제공합니다 . 반응이 명목형 또는 순서형인 경우 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 .

평균 절대 편차 반응과 예측 반응 사이의 차이에 대한 절대값의 평균입니다 . 반응이 명목형 또는 순서형인 경우 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 .

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다. 반응이 명목형 또는 순서형인 경우에만 나타납니다.

-1* 로그 가능도 음의 로그 가능도를 제공합니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

SSE 오차 제곱합을 제공합니다. 반응이 연속형인 경우에만 사용할 수 있습니다.

빈도 합 사용된 관측값 수를 제공합니다. 신경망 시작 창에서 빈도 변수를 지정한 경우 빈도 합은 빈도 열의 합을 제공합니다.

반응이 여러 개인 경우 각 반응에 대한 적합 통계량과 전체 일반화 R^2 및 음의 로그 가능도가 제공됩니다.

혼동 통계량

명목형 또는 순서형 반응의 경우 "혼동 행렬" 보고서와 "혼동 비율" 보고서가 제공됩니다(그림 3.5). "혼동 행렬" 보고서에는 실제 반응 수준과 예측 반응 수준의 이원 분류가 표시됩니다. 범주형 반응의 경우 예측 수준은 예측 확률이 가장 높은 수준입니다. 혼동 비율 보고서는 값을 행 합계로 나눈 혼동 행렬 보고서와 동일합니다.

신경망 플랫폼 옵션

"신경망"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

신경망 모형 옵션

각 모형 보고서에는 추가 출력을 생성하거나 결과를 저장하기 위한 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다. 모형 보고서의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

다이어그램 은닉층 구조를 나타내는 다이어그램을 표시합니다.

추정값 표시 보고서에 모수 추정값을 표시합니다.

프로파일러 예측 프로파일러를 시작합니다 . 명목형 또는 순서형 반응의 경우 각 반응 수준은 예측 프로파일러에서 별도의 행으로 표시됩니다. 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

범주형 프로파일러 예측 프로파일러를 시작합니다. 모든 범주형 확률이 단일 프로파일러 행으로 결합된다는 점을 제외하면 " 프로파일러 " 옵션과 유사합니다 . 명목형 또는 순서형 반응에만 사용할 수 있습니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

등고선 프로파일러 등고선 프로파일러를 시작합니다 . 이 옵션은 모형에 두 개 이상의 연속형 요인이 포함된 경우에만 사용할 수 있습니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

표면 프로파일러 표면 프로파일러를 시작합니다 . 이 옵션은 모형에 두 개 이상의 연속형 요인이 포함된 경우에만 사용할 수 있습니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

ROC 곡선 ROC 곡선을 생성합니다 . 명목형 또는 순서형 반응에만 사용할 수 있습니다 . ROC 곡선에 대한 자세한 내용은 "ROC 곡선 " 에서 확인하십시오 .

향상도 곡선 향상도 곡선을 생성합니다 . 명목형 또는 순서형 반응에만 사용할 수 있습니다 . 향상도 곡선에 대한 자세한 내용은 " 향상도 곡선 " 에서 확인하십시오 .

실제값 대 예측값 그림 실제 반응 대 예측 반응 그림을 표시합니다 . 연속형 반응에만 사용할 수 있습니다 .

잔차 대 예측값 그림 잔차와 예측 반응을 비교하는 그림을 표시합니다 . 연속형 반응에만 사용할 수 있습니다 .

계산식 저장 예측 반응 및 은닉층 노드에 대한 계산식이 포함된 새 열을 데이터 테이블에 생성합니다 .

프로파일 계산식 저장 예측 반응에 대한 계산식이 포함된 새 열을 데이터 테이블에 생성합니다 . 은닉층 노드에 대한 계산식이 이 계산식에 포함되어 있습니다 . 이 옵션은 대화식 HTML 버전의 프로파일러에서 사용할 수 있는 계산식을 생성합니다 .

빠른 계산식 저장 예측 반응에 대한 계산식이 포함된 새 열을 데이터 테이블에 생성합니다 . 은닉층 노드에 대한 계산식이 이 계산식에 포함되어 있습니다 . 이 옵션은 다른 옵션보다 빠르게 실행되는 계산식을 생성하지만 대화식 버전의 프로파일러에서 이 계산식을 사용할 수 없습니다 .

JMP PRO **예측 계산식 게시** 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

SAS Data 스텝 만들기 새 데이터 집합을 스코어링하는 데 사용할 수 있는 SAS 코드를 생성합니다 .

검증 저장 훈련 데이터 집합과 검증 데이터 집합에 사용된 행을 식별하는 새 열을 데이터 테이블에 생성합니다 . 신경망 시작 창에서 검증 열을 지정한 경우에는 이 옵션을 사용할 수 없습니다 . 자세한 내용은 " 신경망 시작 창 " 에서 확인하십시오 .

JMP PRO **변환된 공변량 저장** 변환된 공변량을 보여 주는 새 열을 데이터 테이블에 생성합니다 . 이 열에는 변환을 보여 주는 계산식이 포함됩니다 . 이 옵션은 " 모형 시작 " 제어판에서 " 공변량 변환 " 옵션을 선택한 경우에만 사용할 수 있습니다 . 자세한 내용은 " 적합 옵션 " 에서 확인하십시오 .

적합 제거 전체 모형 보고서를 제거합니다 .

신경망의 예

이 예에서는 Boston Housing.jmp 데이터 테이블을 사용합니다. 주택 가격 중앙값을 여러 인구 통계적 특성의 함수로 예측하기 위한 모형을 생성하려는 경우를 가정해 보겠습니다 . 아래 단계에 따라 신경망 모형을 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Boston Housing.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > 신경망**을 선택하여 신경망 플랫폼을 시작합니다 .
3. mvalue 를 **Y, 반응** 역할에 할당합니다 .
4. 다른 열 (crim ~ lstat) 을 **X, 요인** 역할에 할당합니다 .
5. **확인**을 클릭합니다 .
6. **홀드백 비율**에 "0.2" 를 입력합니다 .
7. **난수 시드값**에 "1234" 를 입력합니다 .

참고 : 일반적으로 검증 데이터 집합을 선택하는 랜덤 특성 때문에 결과가 달라집니다 . 위의 시드값을 입력하면 이 예에 표시된 결과를 재현할 수 있습니다 .

8. 첫 번째 층의 TanH 노드 수에 "3" 을 입력합니다 .
9. **공변량 변환** 옵션을 선택합니다 .
10. **시작**을 클릭합니다 .

그림 3.6 신경망 보고서

모형 NTanH(3)

훈련

mvalue	
측도	값
R^2	0.9074859
RASE	2.8561453
평균 절대 편차	2.1278654
-1*로그 가능도	997.23823
SSE	3295.6567
합 빈도	404

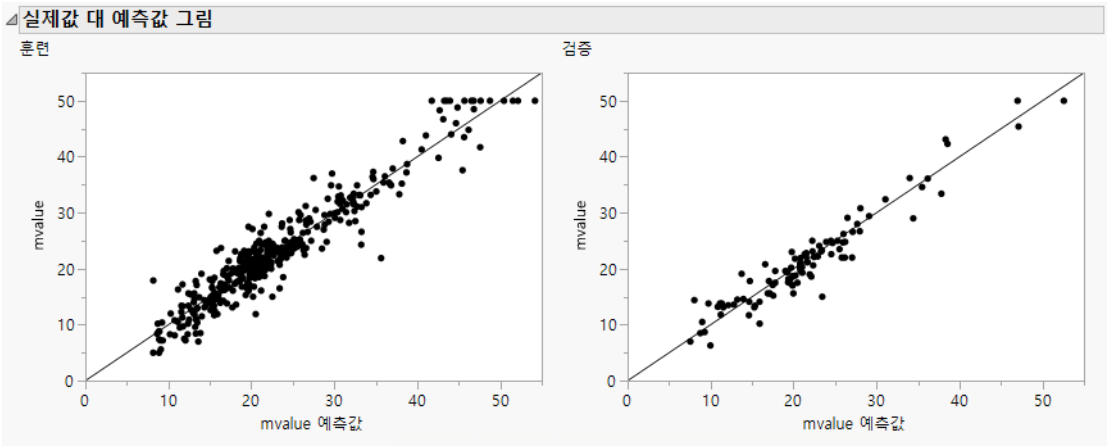
검증

mvalue	
측도	값
R^2	0.9125961
RASE	2.4304519
평균 절대 편차	1.8284984
-1*로그 가능도	235.31561
SSE	602.52384
합 빈도	102

훈련 데이터 집합과 검증 데이터 집합에 대한 보고서가 둘 다 제공됩니다. 검증 데이터 집합의 결과를 미래 관측값에 대한 모형의 예측 검정력을 나타내는 데 사용합니다.

검증 데이터 집합의 R^2 통계량은 0.913이며, 이는 모형이 모형 훈련에 사용되지 않는 데이터에 대해 잘 예측하고 있음을 나타냅니다. 다른 방법으로 모형 적합을 평가하기 위해 "모형 NTanH(3)" 옆의 빨간색 삼각형을 클릭하고 **실제값 대 예측값 그림**을 선택합니다.

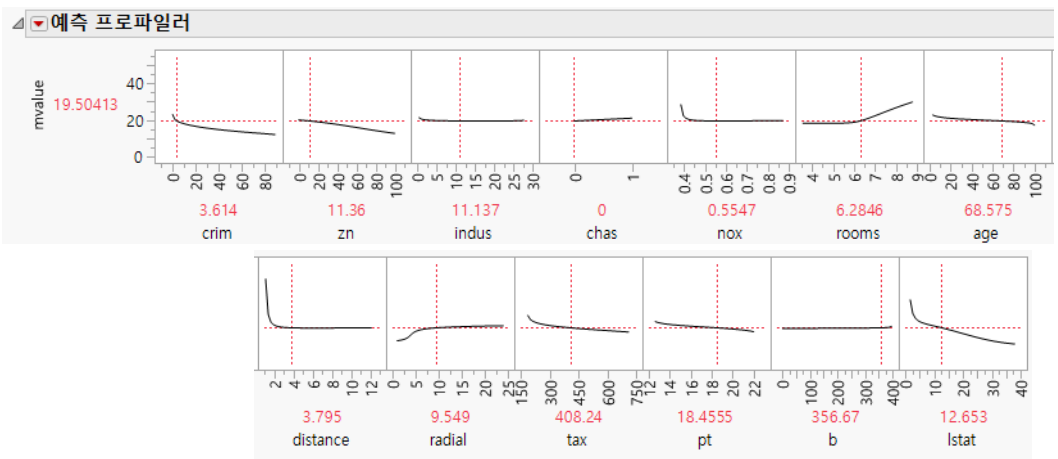
그림 3.7 실제값 대 예측값 그림



점이 선을 따라 분포되어 있으므로 예측값이 실제값과 유사하다는 것을 알 수 있습니다.

X 변수가 예측값에 미치는 영향에 대한 전반적인 이해를 위해 "모형 NTanH(3)" 옆의 빨간색 삼각형을 클릭하고 **프로파일러**를 선택합니다.

그림 3.8 프로파일러



변수 중 일부는 프로파일러의 기울기가 양수이고, 다른 변수는 프로파일러의 기울기가 음수입니다. 예를 들어 rooms 변수의 기울기는 양수입니다. 이는 집에 방이 많을수록 예측 중앙값이 더 높다는 것을 나타냅니다. pt 변수는 도시별 학생 교사 비율입니다. 이 변수의 기울기가 음수이므로 학생 대 교사 비율이 높을수록 중앙값이 낮다는 것을 나타냅니다.

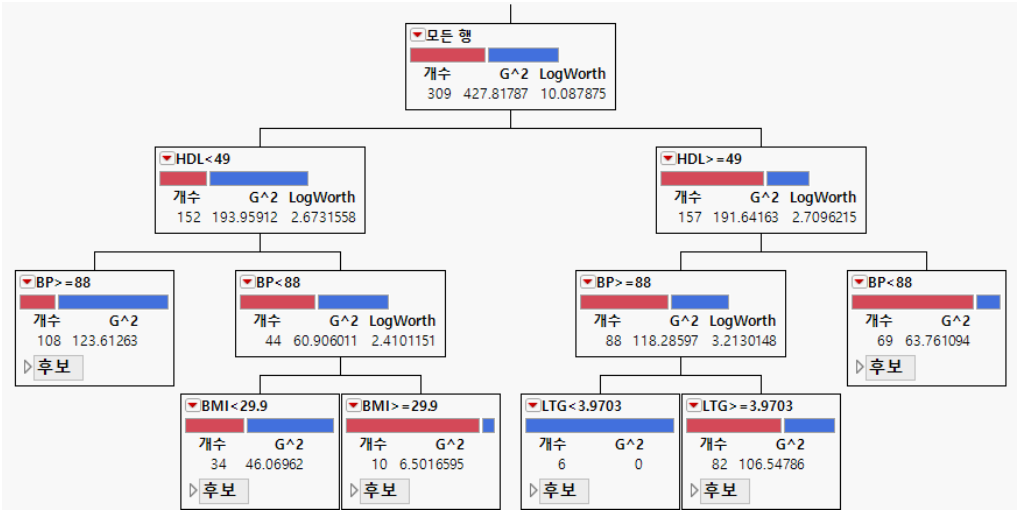
4 장

파티션 모형

결정 트리를 사용하여 데이터 탐색 및 모델링

파티션 플랫폼에서는 예측 변수와 반응 값 사이의 관계에 따라 데이터를 재귀적으로 분할하여 결정 트리를 생성합니다. 파티션 알고리즘은 반응을 가장 잘 예측하기 위해 예측 변수의 가능한 모든 분할을 검색합니다. 이러한 데이터 분할(파티션)은 재귀적으로 수행되어 결정 규칙 트리를 생성합니다. 원하는 적합을 얻을 때까지 분할이 계속됩니다. 파티션 알고리즘은 많은 수의 가능한 분할에서 최적 분할을 선택하므로 강력한 모델링 및 데이터 검색 도구라고 할 수 있습니다.

그림 4.1 결정 트리의 예



목차

파티션 플랫폼 개요 49
파티션 플랫폼의 예 49
파티션 플랫폼 시작 52
분할 보고서 53
컨트롤 버튼 53
범주형 반응에 대한 보고서 54
연속형 반응에 대한 보고서 56
파티션 플랫폼 옵션 59
적합 상세 정보 표시 62
수익 행렬 지정 63
결정 행렬 보고서 64
결측값 정보화 65
실제값 대 예측값 그림 66
ROC 곡선 66
향상도 곡선 68
노드 옵션 69
파티션 검증 70
분할의 추가 예 71
연속형 반응의 예 71
결측값 정보화의 예 73
수익 행렬 및 결정 행렬 보고서의 예 76
파티션 플랫폼에 대한 통계 상세 정보 79
반응 및 요인 79
분할 기준 79
결정 트리 및 붓스트랩 포레스트의 예측 확률 80

파티션 플랫폼 개요

파티션 플랫폼에서는 예측 변수와 반응 값 사이의 관계에 따라 데이터를 재귀적으로 분할하여 결정 트리를 생성합니다. 분할은 결정 트리, CART™, CHAID™, C4.5, C5 등 다양한 이름 및 브랜드 이름에 따라 몇 가지 변형이 있습니다 . 이 기법은 다음과 같은 이유 때문에 데이터 마이닝 기법으로 간주되는 경우가 많습니다 .

- 적합한 사전 모형 없이 관계를 탐색하는 데 유용합니다 .
- 많은 문제를 쉽게 처리합니다 .
- 결과를 해석할 수 있습니다 .

분할이 적용되는 대표적인 예는 질병에 대한 진단 휴리스틱을 만드는 것입니다. 여러 피실험자의 증상과 결과가 주어졌을 때 분할을 사용하여 새 환자를 진단하는 데 도움이 되는 질문 계층을 생성할 수 있습니다.

예측 변수는 연속형 또는 범주형(명목형 또는 순서형)일 수 있습니다. 예측 변수가 연속형이면 **절단 값**에 의해 분할이 생성됩니다. 표본은 이 절단 값보다 작은 값과 큰 값으로 구분됩니다. 예측 변수가 범주형이면 표본이 두 그룹의 수준으로 구분됩니다.

반응도 연속형 또는 범주형(명목형 또는 순서형)일 수 있습니다. 반응이 연속형이면 플랫폼에서 반응 값의 평균을 적합시키고, 오차 제곱합을 최소화하는 분할이 선택됩니다. 반응이 범주형이면 적합된 값은 반응 수준에 대한 확률이며, 잔차 로그 가능도 카이제곱을 최소화하는 분할이 선택됩니다.

분할 기준에 대한 자세한 내용은 "파티션 플랫폼에 대한 통계 상세 정보"에서 확인하십시오.

재귀 분할에 대한 자세한 내용은 Hawkins and Kass(1982) 및 Kass(1980) 연구 자료에서 확인하십시오.

파티션 플랫폼의 예

이 예에서는 파티션 플랫폼을 사용하여 당뇨병 환자의 1 년 질병 진행률 (낮음 또는 높음) 을 예측하는 결정 트리를 생성합니다 .

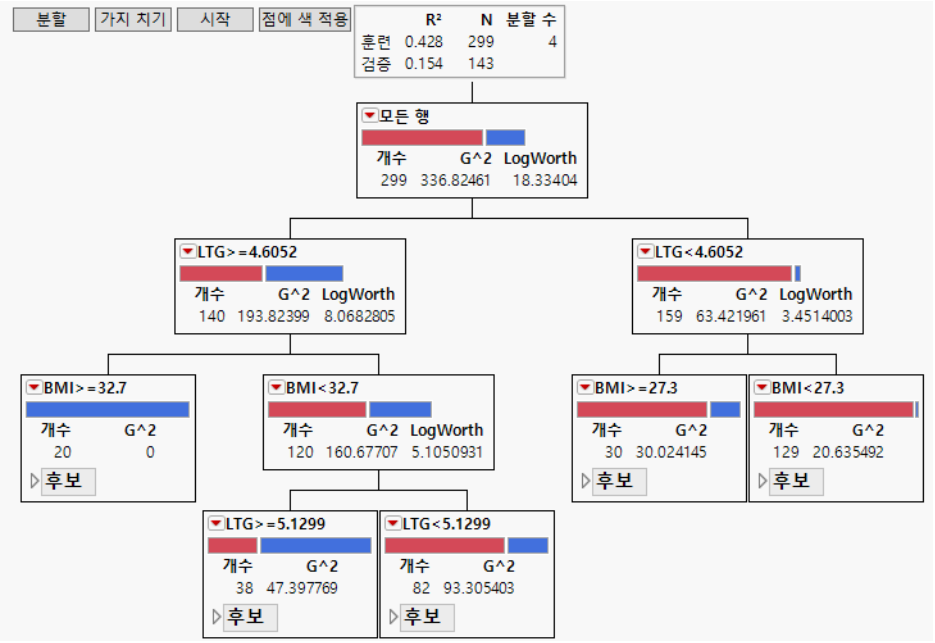
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
2. **분석** > **예측 모델링** > **파티션**을 선택합니다 .
3. Y Binary 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Age ~ Glucose 를 선택하고 **X, 요인**을 클릭합니다 .
5. **검증 데이터 비율**에 "0.33" 을 입력합니다 .

참고 : JMP Pro 에서는 검증을 위해 검증 열을 사용할 수 있습니다 . Validation 을 선택하고 **검증**을 클릭합니다 . 그런 다음 **검증 데이터 비율**을 0 으로 설정합니다 .

6. **확인**을 클릭합니다 .
7. 플랫폼 보고서 창에서 **시작**을 클릭하여 자동 분할을 수행합니다 .

> **참고 :** 랜덤 검증 데이터 비율이 사용되므로 결과가 그림 4.2 에 표시된 것과 다릅니다 .

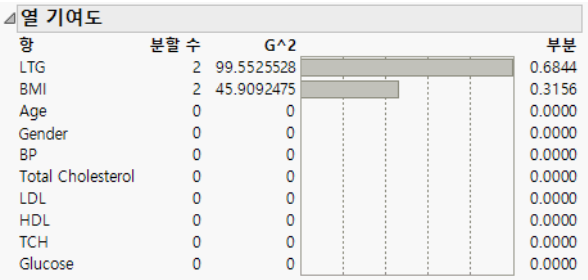
그림 4.2 Diabetes 에 대한 분할 보고서



자동 분할을 통해 네 개의 분할이 생성되었습니다 . 검증 데이터 집합의 최종 R^2 은 0.154 입니다 . 결정 트리는 네 개의 분할과 각 분할의 관측값 수를 보여 줍니다 .

8. "Y Binary 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **열 기여도**를 선택합니다 .

그림 4.3 열 기여도 보고서



" 열 기여도 " 보고서에서는 결정 트리 모형의 예측 변수가 LTG 와 BMI 뿐임을 보여 줍니다 . 각 열은 두 개의 분할에서 사용됩니다 . 사용자가 실행한 결과는 다를 수 있습니다 . 검증 데이터

비율을 사용하는 경우 데이터 테이블에서 검증 데이터 집합이 랜덤 선택됩니다. 분석을 다시 실행하면 새 검증 데이터 집합이 랜덤 선택되고 첫 번째 실행과 결과가 다를 수 있습니다.

9. "Y Binary 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **프로파일러**를 선택합니다.

그림 4.4 파티션 모형에 대한 프로파일러

프로파일러를 사용하면 BMI 및 LTG 값을 변경하여 Y Binary 결과를 예측할 수 있습니다. 다른 요인은 파티션 모형에 분할이 없었습니다. 해당 요인의 프로파일은 편평한 일직선입니다.

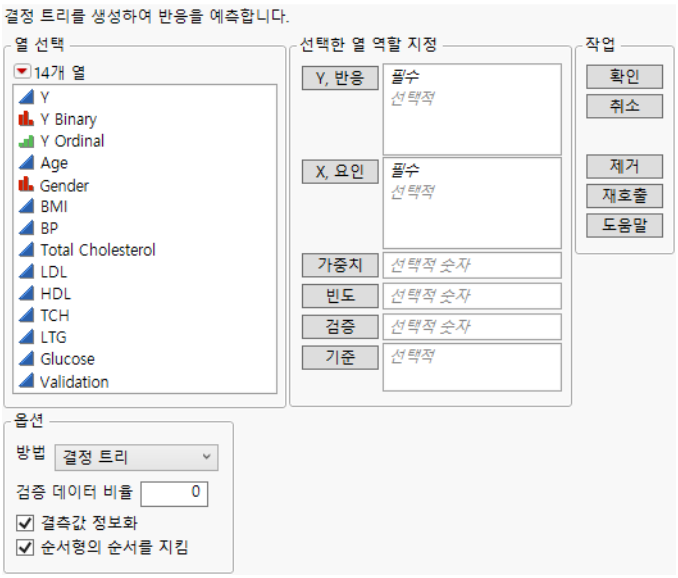
10. "Y Binary 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다.

Diabetes.jmp 데이터 테이블에 **확률 (Y Binary==Low)**, **확률 (Y Binary==High)** 및 **최대 확률 분류 Y Binary** 열이 추가됩니다. 이러한 반응 확률이 어떻게 계산되는지 보려면 " 열 " 패널에서 각 열 옆의 계산식 아이콘 을 두 번 클릭합니다.

파티션 플랫폼 시작

분석 > **예측 모델링** > **파티션**을 선택하여 파티션 플랫폼을 시작합니다 .

그림 4.5 파티션 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 .

X, 요인 예측 변수입니다 .

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다 .

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다 .

JMP PRO **검증** 검증 데이터 집합을 정의하는 숫자 열입니다 . 자세한 내용은 " 파티션 검증 " 에서 확인하십시오 . " 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다 . " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 " 검증 열 생성 " 에서 확인하십시오 .

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다 . 열의 각 수준에 대해 해당 행은 사용자가 지정하는 다른 변수를 사용하여 분석됩니다 . 결과는 별도의 보고서에 나타납니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

JMP PRO **방법** 분할 방법 (붓스트랩 포레스트 , 부스티드 트리 , K 최근접 이웃 또는 Naive Bayes) 을 선택할 수 있습니다 .

이러한 방법에 대한 자세한 내용은 " 붓스트랩 포레스트 ", " 부스티드 트리 ", "K 최근접 이웃 " 및 "Naive Bayes" 에서 확인하십시오 .

검증 데이터 비율 검증 데이터 집합으로 사용할 데이터 비율입니다 . 자세한 내용은 " 파티션 검증 " 에서 확인하십시오 .

결측값 정보화 이 옵션을 선택하면 범주형 예측 변수에 대한 결측값 범주화 및 연속형 예측 변수에 대한 결측값 정보화 처리가 활성화됩니다 . 자세한 내용은 " 결측값 정보화 " 에서 확인하십시오 .

순서형의 순서를 지킴 이 옵션을 선택하면 순서를 유지하는 분할만 고려하도록 제한됩니다 .

분할 보고서

초기 분할 보고서에는 분할 그림 , 컨트롤 버튼 , 요약 패널 및 결정 트리가 표시됩니다 . 분할 그림과 결정 트리는 분할 없이 초기화됩니다 . 보고서 상세 정보는 범주형 반응과 연속형 반응에 대해 서로 다릅니다 .

- " 컨트롤 버튼 "
- " 범주형 반응에 대한 보고서 "
- " 연속형 반응에 대한 보고서 "

컨트롤 버튼

컨트롤 버튼을 사용하여 결정 트리와 상호 작용할 수 있습니다 .

분할 최적 분할을 사용하여 데이터 파티션을 생성합니다 . 여러 분할을 지정하려면 Shift 키를 누른 채 **분할**을 클릭합니다 .

가지 치기 최악 분할을 제거합니다 .

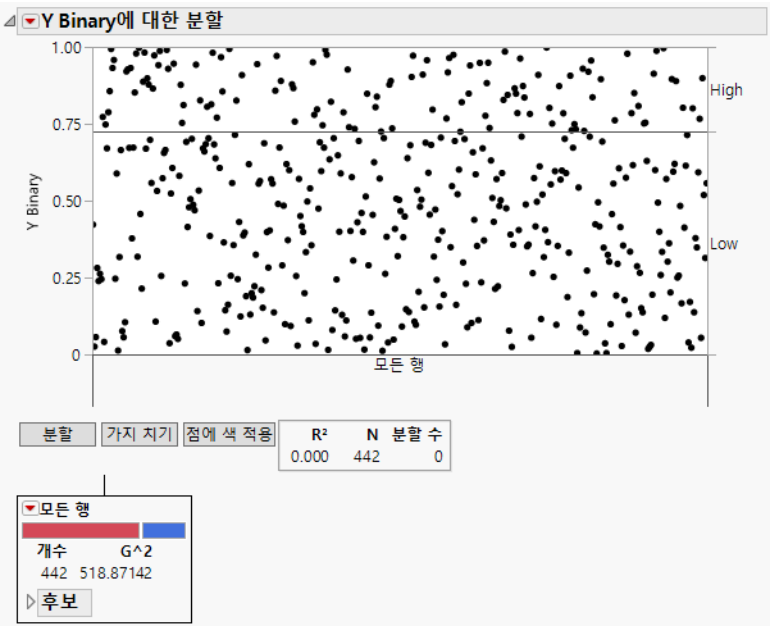
시작 (검증을 사용하는 경우 사용 가능) 검증 통계량이 최적화될 때까지 결정 트리에 분할을 자동으로 추가합니다 . 자세한 내용은 " 파티션 검증 " 에서 확인하십시오 . 검증을 사용하지 않으면 파티션 모형에 사용할 분할 수만 결정하면 됩니다 .

점에 색 적용 범주형 반응의 경우 반응 수준에 따라 관측값에 색상을 지정합니다 . 이 색상은 데이터 테이블에 추가됩니다 .

범주형 반응에 대한 보고서

Diabetes.jmp 샘플 데이터 테이블을 사용하여 범주형 반응 Y Binary 에 대한 보고서가 생성되었습니다 .

그림 4.6 범주형 반응에 대한 분할 보고서



분할 그림

분할 그림의 각 점은 데이터 테이블의 관측값을 나타냅니다. 검증을 사용하는 경우 훈련 데이터에 대한 그림만 표시됩니다. 초기 분할 그림에는 분할이 표시되지 않습니다.

다음을 확인할 수 있습니다 .

- 왼쪽 세로 축은 각 반응 결과의 비율입니다 .
- 오른쪽 세로 축은 반응 수준이 그림에 표시되는 순서를 보여 줍니다 .
- 가로선은 각 분할을 반응 변수별로 나눕니다 . 초기 가로선은 데이터 집합에서 그림에 첫 번째로 표시된 반응의 전체 비율을 보여 줍니다 .
- 분할은 텍스트 설명 및 그림의 관측값을 분할하는 세로선과 함께 X 축 아래에 표시됩니다 . 세로선은 그림까지 이어지고 각 노드의 경계를 나타냅니다 . 가장 최근 분할은 가로 축 바로 아래에서 기존 분할 위에 나타납니다 . 결정 트리의 분할 또는 가지 치기를 수행할 때마다 그림이 업데이트됩니다 .

요약 보고서

그림 4.7 범주형 반응에 대한 요약 보고서

	R^2	N	분할 수
훈련	0.428	299	4
검증	0.154	143	

요약 보고서에는 훈련 데이터와 검증 및 테스트 데이터 (사용된 경우) 에 대한 적합 통계량이 제공됩니다 . 결정 트리의 분할을 추가하거나 가지 치기를 수행하면 요약 패널의 적합 통계량이 업데이트됩니다 .

R^2 현재 R^2 값입니다 .

N 관측값 수입니다 .

분할 수 결정 트리의 현재 분할 수입니다 .

노드 보고서

트리의 각 노드에는 보고서 및 추가 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다 . 최종 노드에는 " 후보 " 보고서도 있습니다 .

그림 4.8 범주형 반응에 대한 최종 노드 보고서

모든 행

개수	G^2
442	518.87142

후보

항	후보 G^2	LogWorth	점 잘라내기
Age	10.5000264	1.71376465	51
Gender	1.8302510	0.75424581	2
BMI	92.8760803	31.25572705	27.3
BP	64.8300680	18.98689929	100
Total Cholesterol	20.3048623	4.01316712	194
LDL	12.5858490	0.82750128	122.2
HDL	44.2535587	11.48909721	46
TCH	64.3516993	17.86426783	4
LTG	102.8078418 *	35.97159929	4.8203
Glucose	43.2683018	11.05809993	99

개수 노드로 특성화된 훈련 관측값 수입니다 .

G^2 연속형 반응에 사용되는 제곱합 대신 범주형 반응에 사용되는 적합 통계량입니다 . 값이 작을수록 더 나은 적합을 나타냅니다 . 자세한 내용은 " 파티션 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오 .

후보 각 열에 대해 " 후보 " 보고서는 해당 열의 최적 분할에 대한 상세 정보를 제공합니다 . 모든 항에 대한 최적 분할은 별표로 표시됩니다 .

항 후보 열을 표시합니다 .

후보 G^2 최선 분할에 대한 가능도비 카이제곱입니다 . G^2 값이 가장 큰 예측 변수에서 분할하면 모형 G^2 의 축소가 최대화됩니다 .

LogWorth $-\log_{10}(p$ 값 $)$ 으로 정의된 LogWorth 통계량입니다 . 최적 분할은 LogWorth 를 최대화하는 분할입니다 . 자세한 내용은 " 파티션 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오 .

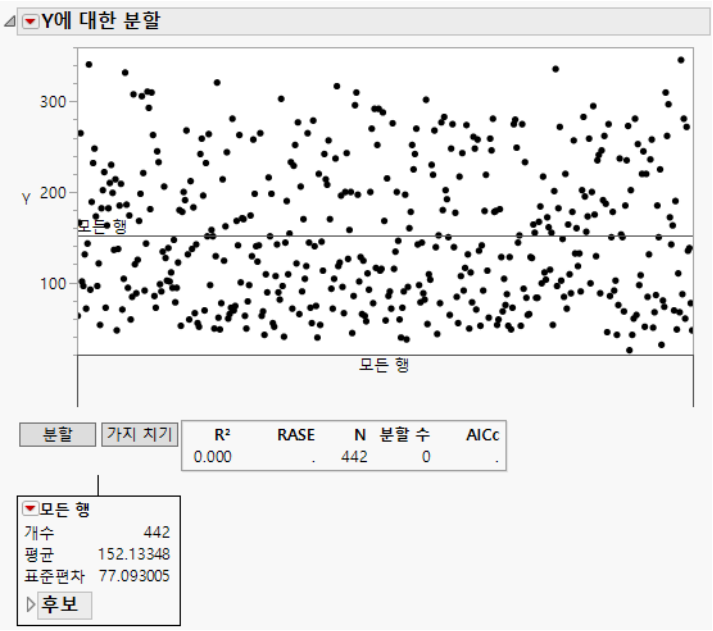
절단점 분할을 결정하는 예측 변수의 값입니다 . 범주형 항의 경우 맨 왼쪽 분할의 수준이 나열됩니다 .

최적 분할은 별표로 표시됩니다. 그러나 한 변수에 대해 후보 G^2이 더 높지만 다른 변수에 대해서는 Logworth가 더 높은 경우가 있습니다. 이 경우 > 및 < 기호를 사용하여 각 변수의 최적 방향을 가리킵니다. 별표는 합치 상태에 해당합니다. 자세한 내용은 "파티션 플랫폼에 대한 통계 상세 정보"에서 확인하십시오.

연속형 반응에 대한 보고서

Diabetes.jmp 샘플 데이터 테이블을 사용하여 연속형 반응 Y 에 대한 보고서가 생성되었습니다 .

그림 4.9 연속형 반응에 대한 분할 보고서



분할 그림

분할 그림은 분할 없이 초기화됩니다. 각 점은 데이터 테이블의 관측값을 나타냅니다. 검증을 사용하는 경우 훈련 데이터에 대한 그림만 표시됩니다.

다음을 확인할 수 있습니다 .

- 세로 축은 관측값의 반응 값을 나타냅니다 .

- 가로선은 결정 트리의 각 노드에 대한 평균 반응 값을 표시합니다 . 초기 가로선은 반응의 전체 평균 위치에 있습니다 .
- 세로 축 눈금 부분은 결정 트리의 분할을 나타냅니다 . 가장 최근 분할에 대한 텍스트 설명이 가로 축 아래에 나타납니다 . 분할이 생성되거나 제거되면 관측값이 각 노드로 재구성됩니다 .

팁 : 좁은 분할에 대한 툴팁을 보려면 분할 그림의 가로 축에 있는 라벨을 커서로 가리킵니다 .

요약 보고서

그림 4.10 연속형 반응에 대한 요약 보고서

	R²	RASE	N	분할 수	AICc
훈련	0.490	54.690663	299	4	3253.83
검증	0.366	61.16064	143		

요약 보고서에는 훈련 데이터와 검증 및 테스트 데이터 (사용된 경우) 에 대한 적합 통계량이 제공됩니다 . 결정 트리의 분할을 추가하거나 가지 치기를 수행하면 요약 패널의 적합 통계량이 업데이트됩니다 .

R² 현재 R^2 값입니다 .

RASE 제곱근 평균 제곱 오차입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

N 관측값 수입니다 .

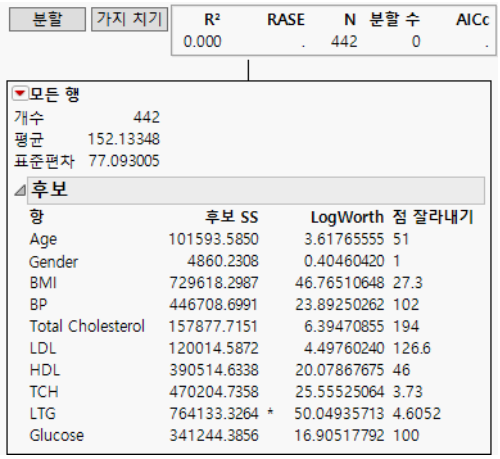
분할 수 결정 트리의 현재 분할 수입니다 .

AICc 수정된 Akaike 정보 기준입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

노드 보고서

트리의 각 노드에는 보고서 및 추가 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다. 최종 노드에는 " 후보 " 보고서도 있습니다.

그림 4.11 연속형 반응에 대한 최종 노드 보고서


분할 | 가지 치기

R^2	RASE	N	분할 수	AICc
0.000	.	442	0	.

모든 행

개수	442
평균	152.13348
표준편차	77.093005

후보

항	후보 SS		LogWorth	점 잘라내기
Age	101593.5850		3.61765555	51
Gender	4860.2308		0.40460420	1
BMI	729618.2987		46.76510648	27.3
BP	446708.6991		23.89250262	102
Total Cholesterol	157877.7151		6.39470855	194
LDL	120014.5872		4.49760240	126.6
HDL	390514.6338		20.07867675	46
TCH	470204.7358		25.55525064	3.73
LTG	764133.3264	*	50.04935713	4.6052
Glucose	341244.3856		16.90517792	100

개수 분지의 관측값 (행) 수입니다.

평균 해당 분지의 모든 관측값에 대한 평균 반응입니다.

표준편차 해당 분지의 모든 관측값에 대한 반응 표준편차입니다.

후보 각 열에 대해 " 후보 " 보고서는 해당 열의 최적 분할에 대한 상세 정보를 제공합니다. 모든 열에 대한 최적 분할은 별표로 표시됩니다.

항 후보 열을 표시합니다.

후보 SS 최선 분할에 대한 제곱합입니다.

LogWorth $-\log_{10}(p$ 값 $)$ 으로 정의된 LogWorth 통계량입니다. 최적 분할은 LogWorth 를 최대화하는 분할입니다. 자세한 내용은 " 파티션 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오.

절단점 분할을 결정하는 예측 변수의 값입니다. 범주형 항의 경우 맨 왼쪽 분할의 수준이 나열됩니다.

최적 분할은 별표로 표시됩니다. 그러나 한 변수에 대해 후보 SS가 더 높지만 다른 변수에 대해서는 Logworth가 더 높은 경우가 있습니다. 이 경우 > 및 < 기호를 사용하여 각 변수의 최적 방향을 가리킵니다. 별표는 합치 상태에 해당합니다. 자세한 내용은 "파티션 플랫폼에 대한 통계 상세 정보"에서 확인하십시오.

파티션 플랫폼 옵션

" 분할 " 의 빨간색 삼각형 메뉴 옵션을 사용하면 필요에 따라 보고서를 사용자 정의할 수 있습니다 . 사용 가능한 옵션은 분석에 사용되는 데이터 유형에 따라 결정됩니다 .

표시 옵션 보고서 요소를 표시하거나 숨기는 옵션이 포함되어 있습니다 .

점 표시 점을 표시합니다 . 범주형 반응의 경우 이 옵션을 선택하면 점 또는 색상이 지정된 패널이 표시됩니다 .

트리 표시 큰 분할 트리를 표시합니다 .

그래프 표시 분할 그래프를 표시합니다 .

분할 막대 표시 (범주형 반응에만 사용 가능) 각 잎의 분할 비율을 나타내는 색상 막대를 표시합니다 .

분할 통계량 표시 분할 통계량을 표시합니다 . 범주형 분할 통계량 G^2 에 대한 자세한 내용은 " 파티션 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오 .

분할 확률 표시 (범주형 반응에만 사용 가능) 노드 보고서에 비율 및 확률 통계량을 표시합니다 .

분할 개수 표시를 선택하면 비율 및 확률 통계량이 자동으로 표시됩니다 . 비율 및 확률에 대한 자세한 내용은 " 파티션 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오 .

분할 개수 표시 (범주형 반응에만 사용 가능) 노드 보고서에 빈도 수를 표시합니다 . 이 옵션을 선택하면 **분할 확률 표시**가 자동으로 선택됩니다 . **분할 확률 표시**를 선택 취소하면 개수가 표시되지 않습니다 .

분할 후보 표시 후보 보고서를 표시합니다 .

분할 후보 정렬 후보 보고서를 통계량 또는 log(worth) 를 기준으로 정렬합니다 .

최선 분할 트리를 최적 분할 지점에서 분할합니다 . 이 옵션은 **분할** 버튼을 클릭하는 것과 같습니다 .

가지 치기 - 최악 항목 최소 판별 능력이 있는 최종 분할을 제거합니다 . 이 옵션은 **가지 치기** 버튼을 클릭하는 것과 같습니다 .

최소 크기 분할 숫자 또는 총 표본 크기의 소수 부분을 입력하여 허용되는 최소 크기 분할을 정의합니다 . 숫자를 지정하려면 1 보다 크거나 같은 값을 입력합니다 . 표본 크기의 비율을 지정하려면 1 보다 작은 값을 입력합니다 . 기본값은 최대값 5 로 설정되거나 , 행 수를 10,000 으로 나눈 값 (내림) 으로 설정됩니다 .

열 잠그기 열이 분할 대상에 포함되지 않도록 해당 열을 대화식으로 잠급니다 . 개별 잠금에 영향을 주지 않고 표시를 해제하거나 다시 설정할 수 있습니다 .

실제값 대 예측값 그림 (연속형 반응에만 사용 가능) 실제값 대 예측값 그림을 표시합니다 . 자세한 내용은 " 실제값 대 예측값 그림 " 에서 확인하십시오 .

작은 트리 보기 분할 그래프 오른쪽에 소형 버전의 분할 트리를 표시합니다 .

3D 트리 트리 구조의 3D 그림을 표시합니다 . 이 옵션에 액세스하려면 Shift 키를 누른 채 빨간색 삼각형 메뉴를 클릭합니다 .

잎 보고서 보고서의 맨 아래 수준 잎에 대한 평균과 개수 또는 비율을 표시합니다 .

열 기여도 적합에 대한 각 입력 열의 기여도를 나타내는 보고서를 표시합니다 . 이 보고서에는 분할을 정의한 횟수와 해당 열에 기인한 총 G^2 또는 제곱합도 표시됩니다 .

분할 기록 R^2 대 분할 수 그림을 표시합니다 . 제외된 행 검증 , 홀드백 검증 또는 검증 열을 사용하는 경우 훈련 및 검증 R^2 값에 대한 별도의 곡선이 표시됩니다 . 훈련 데이터 집합의 R^2 곡선은 파란색이고 검증 데이터 집합의 R^2 곡선은 빨간색입니다 . K 폴드 교차 검증을 선택하는 경우 모든 데이터에 대한 R^2 곡선은 파란색이고 교차 검증 R^2 곡선은 녹색입니다 .

ROC 곡선 (범주형 반응에만 사용 가능) ROC(Receiver Operating Characteristic) 곡선은 반응 수준을 정렬하기 위한 모형 적합 확률의 효율성을 표시합니다 . 자세한 내용은 "ROC 곡선 " 에서 확인하십시오 .

향상도 곡선 (범주형 반응에만 사용 가능) 향상도 곡선은 파티션 모형의 예측 능력을 표시합니다 . 자세한 내용은 " 향상도 곡선 " 에서 확인하십시오 .

적합 상세 정보 표시 (반응이 범주형인 경우에만 표시됨) 적합 상세 정보 보고서에는 여러 적합 측도가 표시되고 혼동 행렬 보고서가 제공됩니다 . 자세한 내용은 " 적합 상세 정보 표시 " 에서 확인하십시오 .

열 저장 모형 및 트리 결과를 저장하고 SAS 코드를 생성하기 위한 옵션이 포함되어 있습니다 .

잔차 저장 모형의 잔차 값을 데이터 테이블에 저장합니다 .

예측값 저장 모형의 예측값을 데이터 테이블에 저장합니다 .

잎 번호 저장 트리의 잎 번호를 데이터 테이블의 열에 저장합니다 .

잎 라벨 저장 트리의 잎 라벨을 데이터 테이블에 저장합니다 . 라벨은 트리를 따라 행이 추적하는 각 분지에 대한 내용을 기록합니다 . 각 분지는 "&" 로 구분됩니다 . 예를 들면 "size(Small,Medium)&size(Small)" 로 라벨을 표시할 수 있습니다 . 그러나 JMP 에서는 반복되는 범주 라벨 형식으로 중복 정보를 포함하지 않습니다 . 잎의 범주 라벨은 상위 트리 노드의 포괄적 범주 목록을 나타낼 수도 있습니다 . 중복 라벨이 있는 트리 노드 위치에 캐럿 ('^') 이 나옵니다 . 따라서 "size(Small,Medium)&size(Small)" 은 ^&size(Small) 로 표시됩니다 .

예측 계산식 저장 예측 계산식을 데이터 테이블의 하나 또는 여러 열에 저장합니다 . 계산식은 트리 구조를 설명하는 내포된 조건부 절로 구성됩니다 . 반응이 연속형이면 " 예측 " 특성이 포함된 열 하나가 추가됩니다 . 반응이 범주형이면 반응의 각 수준에 대해 " 반응 확률 " 특성이 포함된 여러 열이 추가됩니다 . 또한 각 관측값에 대해 발생 확률이 가장 높은 반응 수준이 포함된 **최대 확률 분류** 열이 추가됩니다 .

허용 예측 계산식 저장 결측값이 있고 " 결측값 정보화 " 를 선택하지 않은 경우에도 예측하는 계산식을 저장합니다 . 예측 계산식은 결측 예측 변수에 대한 반응 값을 분할에 랜덤 할당하여 결측값을 허용합니다 . 반응이 연속형이면 열에 " 예측 " 특성이 포함됩니다 . 반응이 범주형이면 열에 " 반응 확률 " 특성이 포함됩니다 . " 결측값 정보화 " 를 선택한 경우 Shift 키를 누른 채 보고서의 빨간색 삼각형을 클릭하면 허용 예측 계산식을 저장할 수 있습니다 .

잎 번호 계산식 저장 잎 번호 계산식이 포함된 열을 데이터 테이블에 저장합니다 .

잎 라벨 계산식 저장 잎 라벨 계산식이 포함된 열을 데이터 테이블에 저장합니다 .

SAS DATA 스텝 만들기 새 데이터 집합을 스코어링하기 위한 SAS 코드를 생성합니다 .

JMP PRO **예측 계산식 게시** 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

JMP PRO **허용 예측 계산식 게시** 허용 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 . " 결측값 정보화 " 를 선택한 경우 Shift 키를 누른 채 보고서의 빨간색 삼각형을 클릭하면 이 옵션을 사용할 수 있습니다 .

수익 행렬 지정 (범주형 반응에만 사용 가능) 올바르거나 올바르지 않은 분류 결정과 관련된 수익 또는 비용을 지정할 수 있습니다 . 명목형 반응의 경우 확률 임계를 사용하여 수익 행렬 항목을 지정할 수 있습니다 . 자세한 내용은 " 적합 상세 정보 표시 " 에서 확인하십시오 .

프로파일러 대화식 프로파일러 보고서를 표시합니다 . 요인 값을 변경하면 추정된 분류 확률에 반영됩니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

점에 색 적용 (범주형 반응에만 사용 가능) 반응 수준에 따라 점에 색상을 적용합니다 . 이 옵션은 " 점에 색 적용 " 버튼을 클릭하는 것과 같습니다 . 자세한 내용은 " 범주형 반응에 대한 보고서 " 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

적합 상세 정보 표시

그림 4.12 범주형 반응에 대한 적합 상세 정보 (Diabetes.jmp 의 Y Binary)

적합 상세 정보

측도	훈련	검증	정의
엔트로피 R²	0.4276	0.1543	1-Loglike(model)/Loglike(0)
일반화 R²	0.5656	0.2464	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
평균 -Log p	0.3224	0.5312	∑ -Log(ρ[j])/n
RASE	0.3201	0.4080	√ ∑(y[j]-ρ[j])²/n
평균 절대 편차	0.2083	0.2917	∑ |y[j]-ρ[j]|/n
오분류 비율	0.1371	0.2238	∑ (ρ[j]≠ρMax)/n
N	299	143	n

혼동 행렬

훈련

실제값 Y Binary	예측 개수 Low	High
Low	212	12
High	29	46

실제값 Y Binary	예측 비율 Low	High
Low	0.946	0.054
High	0.387	0.613

검증

실제값 Y Binary	예측 개수 Low	High
Low	84	13
High	19	27

실제값 Y Binary	예측 비율 Low	High
Low	0.866	0.134
High	0.413	0.587

엔트로피 R² 적합 모형과 상수 확률 모형의 로그 가능도를 비교하는 적합 측도입니다. 엔트로피 R² 범위는 0 ~ 1 이며, 값이 1 에 가까울수록 더 나은 적합을 나타냅니다. 자세한 내용은 "엔트로피 R²" 에서 확인하십시오.

일반화 R² 일반 회귀 모형에 적용할 수 있는 측도입니다. 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다. 일반화 R² 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R² 로 단순화됩니다. 일반화 R² 을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데, 이는 Cox-Snell 유사 R^2 을 정규화한 것입니다. 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오. 값이 1 에 가까울수록 더 나은 적합을 나타냅니다.

평균 -Log p -log(p) 의 평균입니다. 여기서 p 는 발생한 사건과 연관된 적합 확률입니다. 값이 작을수록 더 나은 적합을 나타냅니다.

RASE 제곱근 평균 제곱 오차입니다. 여기서 차이는 반응과 p(실제로 발생한 사건의 적합 확률) 사이의 차이입니다. 값이 작을수록 더 나은 적합을 나타냅니다.

평균 절대 편차 반응과 p(실제로 발생한 사건의 적합 확률) 사이의 차이에 대한 절대값을 평균한 것입니다. 값이 작을수록 더 나은 적합을 나타냅니다.

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다. 값이 작을수록 더 나은 적합을 나타냅니다.

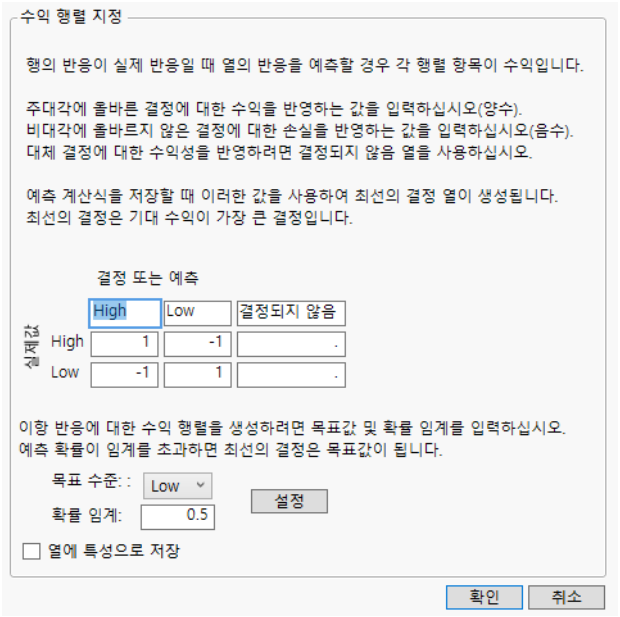
"혼동 행렬" 보고서에는 훈련 데이터 집합과 검증 및 테스트 데이터 집합(정의된 경우)에 대한 행렬이 표시됩니다. "혼동 행렬" 보고서에는 혼동 행렬과 혼동 비율 행렬이 포함됩니다. 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다. 혼동 비율 행렬은 값을 행 합계로 나눈 혼동 행렬과 동일합니다.

반응에 "수익 행렬" 열 특성이 있거나 "수익 행렬 지정" 옵션을 사용하여 비용을 지정하는 경우 "결정 행렬" 보고서가 나타납니다. 자세한 내용은 "결정 행렬 보고서"에서 확인하십시오.

수익 행렬 지정

수익 행렬은 범주형 반응과 함께 사용할 수 있습니다. 수익 행렬은 원하지 않는 결과에 비용을 할당하고 원하는 결과에 수익을 지정하는 데 사용됩니다.

그림 4.13 수익 행렬 지정 창



실제 반응 범주와 예측 반응 범주의 각 조합에 수익 및 비용 값을 할당할 수 있습니다. 대체 범주로 분류하는 비용을 지정하려면 "결정되지 않음" 열에 값을 입력합니다. 할당을 반응 열에 특성으로 저장하려면 **열에 특성으로 저장**을 선택합니다. 이 옵션을 선택하지 않으면 수익 행렬이 현재 분할 보고서에만 적용됩니다.

수익 행렬의 확률 임계 지정

이항 반응의 경우 수익 행렬에 가중치를 입력하는 대신 " 수익 행렬 " 창에서 확률 임계를 지정할 수 있습니다 . 수익 행렬에 대한 값을 계산하는 방법은 JMP 사용의 에서 확인하십시오 .

목표값 확률을 모델링할 수준입니다 .

확률 임계 목표 수준의 확률에 대한 임계입니다 . 관측값이 목표 수준 이내가 될 확률이 확률 임계를 초과할 경우 해당 관측값은 해당 수준으로 분류됩니다 .

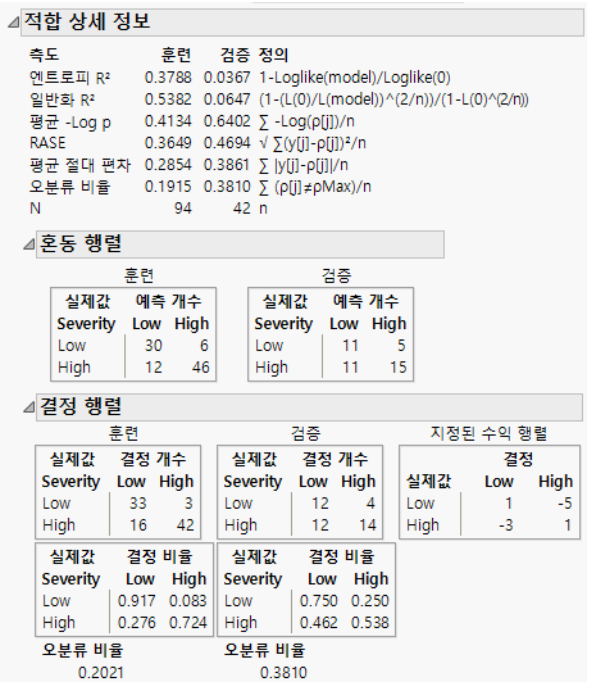
"수익 행렬 지정" 옵션을 사용하여 비용을 정의한 후 "적합 상세 정보 표시"를 선택하면 "결정 행렬" 보고서가 나타납니다. 자세한 내용은 "결정 행렬 보고서"에서 확인하십시오.

수익 행렬을 지정하고 모형 예측 계산식을 저장하면 다음을 포함하는 계산식 열이 데이터 테이블에 저장됩니다 .

- <level> 에 대한 수익 : 열의 각 반응 수준에 대해 각 관측값을 해당 수준으로 분류하기 위한 기대 수익을 지정합니다 .
- <column name> 에 대한 수익성이 가장 높은 예측 : 각 관측값에 대해 기대 수익이 가장 높은 반응의 수준을 지정합니다 .
- <column name> 에 대한 기대 수익 : 각 관측값에 대해 수익성이 가장 높은 예측 열에 의해 정의된 분류의 기대 수익을 지정합니다 .
- <column name> 에 대한 실제 수익 : 각 관측값에 대해 해당 관측값을 수익성이 가장 높은 예측 열에 의해 지정된 수준으로 분류하기 위한 실제 수익을 지정합니다 .

결정 행렬 보고서

그림 4.14 결정 행렬 보고서가 포함된 적합 상세 정보 보고서



적합 상세 정보

측도	훈련	검증	정의
엔트로피 R²	0.3788	0.0367	1-Loglike(model)/Loglike(0)
일반화 R²	0.5382	0.0647	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
평균 -Log p	0.4134	0.6402	∑ -Log(ρ[j])/n
RASE	0.3649	0.4694	√ ∑(y[j]-ρ[j])²/n
평균 절대 편차	0.2854	0.3861	∑ |y[j]-ρ[j]|/n
오분류 비율	0.1915	0.3810	∑ (ρ[j]≠ρMax)/n
N	94	42	n

혼동 행렬

훈련

실제값 Severity	예측 개수 Low	High
Low	30	6
High	12	46

검증

실제값 Severity	예측 개수 Low	High
Low	11	5
High	11	15

결정 행렬

훈련

실제값 Severity	결정 개수 Low	High
Low	33	3
High	16	42

실제값 Severity	결정 비율 Low	High
Low	0.917	0.083
High	0.276	0.724

오분류 비율 0.2021

검증

실제값 Severity	결정 개수 Low	High
Low	12	4
High	12	14

실제값 Severity	결정 비율 Low	High
Low	0.750	0.250
High	0.462	0.538

오분류 비율 0.3810

지정된 수익 행렬

실제값	결정 Low	High
Low	1	-5
High	-3	1

참고 : 이 보고서는 반응에 " 수익 행렬 " 열 특성이 있거나 " 수익 행렬 지정 " 옵션을 사용하여 비용을 지정하는 경우에만 사용할 수 있습니다 . 이 . 보고서는 " 적합 상세 정보 " 보고서에 포함됩니다 .

수익 행렬이 정의되면 파티션 알고리즘에서 행렬의 값을 사용하여 각 결정에 대한 수익을 계산합니다 . " 적합 상세 정보 표시 " 를 선택하면 " 결정 행렬 " 보고서가 나타납니다 .

"결정 행렬" 보고서에서 결정 개수는 수익 행렬의 가중치에 따라 수익성이 가장 높은 예측 결정을 반영합니다. 이 보고서는 훈련 데이터 집합과 검증 및 테스트 데이터 집합(정의된 경우)에 대한 결정 개수 및 결정 비율 행렬을 제공합니다. 참조를 위해 수익 행렬도 표시됩니다.

참고 : " 수익 행렬 지정 " 옵션을 사용하여 수익 행렬의 가중치를 변경하면 " 결정 행렬 " 보고서가 자동으로 업데이트되어 변경 내용을 반영합니다 .

결정 개수 행렬 행에 실제 반응을 사용하고 열에 분류 개수를 사용하여 이원 분류를 표시합니다.

지정된 수익 행렬 수익 행렬을 정의하는 가중치를 제공합니다 .

결정 비율 행렬 주어진 행에서 각 범주로 분류되는 관측값의 비율에 해당하는 비율 값을 표시합니다 . 모든 관측값이 올바르게 분류되면 대각선에 나타나는 비율이 모두 1 이 됩니다 .

팁 : 비용이 1 과 -1 인 기본 수익 행렬을 사용하여 반응에 대한 결정 비율 행렬을 얻을 수 있습니다 . 빨간색 삼각형 메뉴에서 **수익 행렬 지정**을 선택한 후 기본값을 변경하지 않고 **확인**을 클릭합니다 .

행렬은 다음과 같이 두 행으로 배열됩니다 .

- 결정 개수 행렬은 첫 번째 행에 있습니다 .
- 지정된 수익 행렬은 첫 번째 행의 오른쪽에 있습니다 .
- 결정 비율 행렬은 두 번째 행에 있습니다 .

결측값 정보화

" 결측값 정보화 " 옵션을 사용하면 예측 변수의 결측값을 정보로 처리할 수 있습니다 . 적합 모형은 결정적 모형입니다 . " 결측값 정보화 " 옵션은 시작 창에 있으며 기본적으로 선택되어 있습니다 . " 결측값 정보화 " 옵션을 선택하면 범주형 예측 변수와 연속형 예측 변수가 다르게 처리됩니다 .

- 범주형 예측 변수의 결측값을 포함하는 행은 해당 변수에 대한 별도의 수준으로 분석에 포함됩니다 .
- 연속형 예측 변수의 결측값을 포함하는 행은 다음과 같이 분할에 할당됩니다 . 연속형 예측 변수의 값이 정렬됩니다 . 먼저 결측 행이 정렬된 값의 아래쪽 끝에 있다고 간주됩니다 . 모든 분할이 생성됩니다 . 그런 다음 결측 행이 정렬된 값의 위쪽 끝에 있다고 간주됩니다 . 다시 모든 분할이 생성됩니다 . 최적 분할은 LogWorth 기준을 사용하여 결정됩니다 . 지정된 예측 변수에 대해 추가 분할을 수행하는 경우 알고리즘은 해당 예측 변수에 의해 유도된 첫 번째 분할에 따라 결측 행을 높은 값 또는 낮은 값으로 커밋합니다 .

" 결측값 정보화 " 옵션을 선택하지 않으면 결측값이 다음과 같이 처리됩니다 .

- 결측값이 있는 예측 변수가 분할 변수로 사용되면 해당 예측 변수에 대한 결측값이 있는 각 행이 분할의 양쪽 중 하나에 랜덤 할당됩니다 .

- 결측값이 있는 예측 변수가 분할 변수로 처음 사용되면 **결측값 대치** 열이 요약 보고서에 추가되어 결측값 대치 수를 표시합니다. 결측값 대치가 추가로 발생하면 결측값 대치 열이 업데이트됩니다 (그림 4.15). 여기서는 결측값 대치가 5 번 수행되었습니다.

참고 : 결측값 대치 수는 결측값이 포함된 행 수보다 클 수 있습니다. 결측값 대치는 각 분할에서 발생합니다. 결측값이 있는 행은 여러 번 랜덤 할당될 수 있습니다. 행이 랜덤 할당될 때마다 결측값 대치 수가 증가합니다.

그림 4.15 요약 보고서의 결측값 대치 메시지

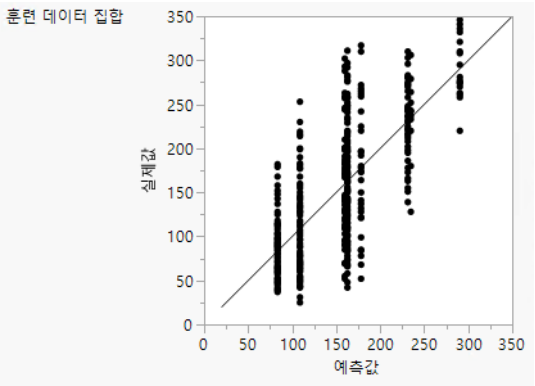
R^2	RASE	N	분할 수	AICc
0.448	6.8246278	501	1	3352.2

실제값 대 예측값 그림

연속형 반응의 경우 실제값 대 예측값 그림은 실제 반응 대 예측 반응을 나타내는 대표적인 그림입니다. 결정 트리를 적합시키면 잎의 모든 관측값이 동일한 예측값을 가집니다. 잎이 n 개인 경우 실제값 대 예측값 그림에는 최대 n 개의 개별 예측값이 표시됩니다. 실제 값은 n 개의 세로선에서 각 잎 평균 주위에 점이 분산된 산점도를 구성합니다.

대각선은 Y = X 선입니다. 모든 점이 이 대각선 위에 있을 때 완벽한 적합이 됩니다. 검증이 사용되면 훈련 데이터 집합과 검증 데이터 집합 둘 다에 대해 그림이 표시됩니다(그림 4.16).

그림 4.16 연속형 반응에 대한 실제값 대 예측값 그림



ROC 곡선

"ROC 곡선" 옵션은 범주형 반응에만 사용할 수 있습니다. ROC(Receiver Operating Characteristic) 곡선은 반응 수준을 정렬하기 위한 모형 적합 확률의 효율성을 표시합니다. ROC 곡선에 대한 소개는 **기본 분석**의 에서 확인하십시오.

파티션 모형의 각 관측값에 대한 예측 반응은 0과 1 사이의 값입니다. 예측 반응을 사용하여 관측값을 양수 또는 음수로 분류하려면 **절단점**이 사용됩니다. 예를 들어 절단점이 0.5인 경우 예측

반응이 0.5 이상인 관측값은 Positive로 분류되고 0.5 미만인 관측값은 Negative로 분류됩니다. 절단점이 달라지면 분류가 절충됩니다.

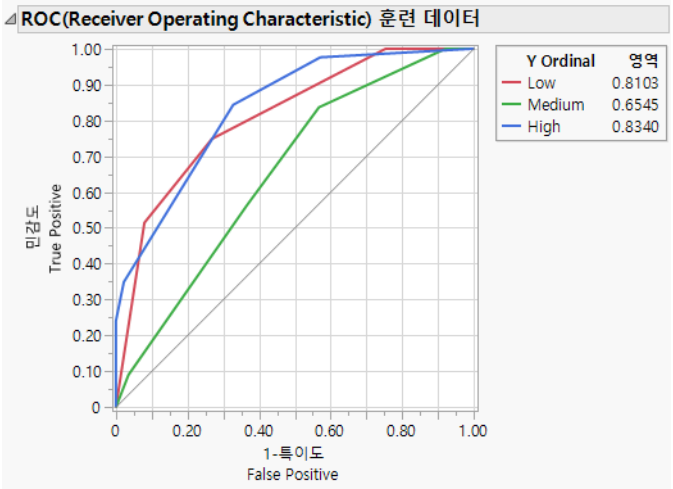
ROC 곡선을 생성하기 위해 각 예측 반응 수준을 가능한 절단점으로 간주하고 가능한 각 절단점에 대해 다음 값이 계산됩니다.

- **민감도**는 True Positive 비율 또는 예측 반응이 절단점보다 큰 Positive 관측값의 비율입니다.
- **특이도**는 True Negative 비율 또는 예측 반응이 절단점보다 작은 Negative 관측값의 비율입니다.

ROC 곡선은 민감도 대 1-특이도를 표시합니다. *n*개의 분할이 있는 파티션 모형에는 *n*+1개의 예측값이 있습니다. 파티션 모형의 ROC 곡선에는 *n*+1개의 선분이 있습니다.

반응 수준이 세 개 이상인 경우 분할 보고서에는 각 반응 수준 대 다른 수준에 대한 개별 ROC 곡선이 포함됩니다. 각 곡선은 수준을 Positive 반응 수준으로 표현한 것입니다. 수준이 두 개뿐인 경우 한 곡선은 다른 곡선을 반영합니다.

그림 4.17 수준이 세 개인 반응의 ROC 곡선



모형이 반응 값의 순위를 완벽하게 지정하면 정렬된 데이터에 먼저 모든 양의 값이 포함된 후 다른 모든 값이 포함됩니다. 이 경우 곡선은 오른쪽으로 이동하기 전에 맨 위로 이동합니다. 모형이 잘 예측하지 못하면 곡선은 그림의 왼쪽 아래에서 오른쪽 위로 대각선을 따릅니다.

실제로 ROC 곡선은 대각선 위에 있습니다. 곡선 아래 면적은 모형의 적합도를 나타내는 표시자입니다. 값이 1이면 완벽한 적합을 나타내고 0.5에 가까우면 모형이 그룹을 구별할 수 없음을 나타냅니다.

반응 수준이 세 개 이상인 경우 ROC 곡선 그림을 통해 곡선 아래 면적이 가장 큰 반응 범주를 확인할 수 있습니다.

향상도 곡선

"향상도 곡선" 옵션은 파티션 모형의 예측 능력을 표시하는 다른 그림을 제공합니다. 향상도 곡선은 향상도 대 관측값 비율을 표시합니다. 고유한 각 예측 확률 값에 대한 점이 있습니다. 반응 수준의 각 예측 확률은 예측 확률이 고유 예측 확률 값보다 크거나 같은 관측값의 비율을 정의합니다. 특정 반응 수준에 대한 **향상도** 값은 관측 반응의 전체 비율 대비 해당 부분의 관측 반응 비율을 나타내는 값입니다.

참고 : 작은 모형의 경우 대부분의 점이 동일한 예측 확률 값을 가질 수 있습니다 . 이 확률이 반응 수준에 대한 가장 높은 예측 확률인 경우에는 향상도 곡선이 부분 = 0 에서 시작하지 않습니다 . 그림 4.18 의 Low 향상도 곡선을 예로 들 수 있습니다 .

그림 4.18 향상도 곡선

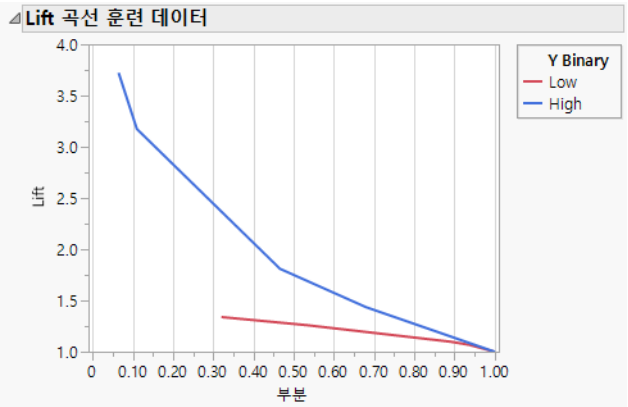


그림 4.19 향상도 곡선에 대한 향상도 테이블

Prob High	N > Prob High	Portion	N High in Portion	Portion High	Lift = portion high/ overall high of .27
0.97	20	0.06	20	1.00	3.72
0.77	44	0.14	39	0.89	3.30
0.33	68	0.22	47	0.69	2.57
0.31	168	0.54	78	0.46	1.73
0.04	309	1.00	83	0.27	1.00

그림 4.19 에서는 그림 4.18 에 나오는 High 향상도 곡선에 사용된 " 향상도 " 및 " 부분 " 의 계산을 보여 주기 위한 값 테이블을 제공합니다 . 반응 Y Binary 를 예측하기 위해 5 개의 분할이 있는 파티션 모형이 생성되었습니다 . Y Binary 에는 Low 와 High 라는 두 개의 수준이 있습니다 . 향상도 곡선은 309 개의 관측값을 기반으로 합니다 . 전체 비율 0.27 에 대해 관측된 High 반응은 83 개입니다 .

- Prob High: 파티션 모형에서 High 반응 수준에 대한 5 개의 고유 예측 확률 값
- N > Prob High: 예측 확률 값이 Prob High 값보다 크거나 같은 관측값 수
- Portion: N > Prob High 를 309, 즉 총 관측값 수로 나눈 값

- N High in Portion: 각 부분에서 High 반응이 관측된 관측값 수
- Portion High: N High in Portion 을 N > Prob High 로 나눈 값
- Lift: Portion High 를 0.27, 즉 관측된 High 반응의 전체 비율로 나눈 값

향상도는 각 부분에 대한 기대 High 반응 수와 비교하여 해당 부분에 속하는 High 반응 수를 측정합니다. 데이터 집합의 처음 6%에 대한 향상도는 3.72입니다. 모형을 사용하여 예측값이 가장 높은 관측값의 6%를 선택한 결과, 해당 6%를 랜덤 선택한 경우보다 High 반응이 3.72개 더 많습니다.

노드 옵션

이 섹션에서는 각 노드의 빨간색 삼각형 메뉴에 있는 옵션에 대해 설명합니다.

최선 분할 이 노드 또는 그 아래에서 최선 분할을 찾아 실행합니다.

여기에서 분할 선택한 노드에서 가장 적절한 열을 기준으로 분할합니다.

세부 분할 분할이 실행되는 위치를 지정할 수 있습니다. 이 옵션은 기준을 절단점의 함수로 표시하고 사용자 절단점을 결정할 때 유용합니다. 분할 열을 지정할 때 분할 방법에 대해 다음 옵션을 선택할 수 있습니다.

최적값 선택한 변수의 최적값에서 분할합니다.

지정한 값 분할이 실행되는 수준을 지정할 수 있습니다.

출력 분할 테이블 가능한 모든 분할 및 연관된 분할 값을 표시하는 데이터 테이블을 생성합니다.

가지 치기 - 아래 선택한 노드 아래의 분할을 제거합니다.

가지 치기 - 최악 항목 선택한 노드 아래에서 최악 분할을 찾아 제거합니다.

행 선택하기 이 잎에 해당하는 데이터 테이블 행을 선택합니다. Shift 키를 누른 채 다른 노드에서 이 명령을 선택하여 선택 영역을 확장할 수 있습니다.

상세 정보 표시 선택한 변수에 대한 분할 기준을 표시하는 데이터 테이블을 생성합니다. 분할 구간 및 관련 기준 값으로 구성된 데이터 테이블에는 해당 기준에 대한 그래프를 생성하는 스크립트가 연결되어 있습니다.

잠금 해당 노드 또는 하위 노드를 분할 대상으로 선택할 수 없게 합니다. 이 옵션을 선택하면 노드 제목에 잠금 아이콘이 나타납니다.

파티션 검증

파티션 모형이 쉽게 과대적합된다는 점에서 파티션 모형과 함께 검증을 사용하는 것은 중요합니다. 이 경우 모형은 모형을 생성하는 데 사용된 데이터를 매우 잘 예측하지만 미래 관측값은 제대로 예측하지 못합니다 . 검증은 데이터 집합의 일부를 사용하여 모형 모수를 추정하고 다른 부분을 사용하여 모형의 예측 능력을 평가하는 프로세스입니다 . 검증에 대한 자세한 내용은 "JMP 모델링의 검증 " 에서 확인하십시오 .

파티션에서 검증 방법을 사용할 때 **시작** 버튼이 나타납니다. **시작** 버튼을 사용하면 **분할** 버튼을 반복해서 클릭하지 않아도 분할을 반복할 수 있습니다. **시작** 버튼을 클릭하면 검증 R^2이 다음 10 개 분할을 통해 얻을 수 있는 값보다 나을 때까지 계속 분할합니다. 이 규칙은 해석하기 어렵지만 예측 검정력이 좋은 복잡한 트리를 생성할 수 있습니다.

시작 버튼을 사용하면 **분할 기록** 명령이 설정됩니다. **시작** 버튼을 사용하여 생성된 트리의 노드가 40 개를 초과하는 경우 **트리 표시** 명령이 해제됩니다.

다음 검증 방법 중 하나를 선택합니다 .

제외된 행 행 상태를 사용하여 데이터의 부분집합을 생성합니다 . 제외되지 않은 행은 훈련 데이터 집합으로 사용되고 제외된 행은 검증 데이터 집합으로 사용됩니다 .

행 상태 사용 및 행 제외 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

홀드백 원래 데이터를 무작위로 훈련 데이터 집합과 검증 데이터 집합으로 나눕니다 . 플랫폼 시작 창의 검증 데이터 비율은 검증 데이터 집합으로 사용할 원래 데이터의 비율을 지정하는 데 사용됩니다 (홀드백). 검증 데이터 비율에 대한 자세한 내용은 " 파티션 플랫폼 시작 " 에서 확인하십시오 .

JMP PRO **검증 열** 검증 데이터 집합을 정의하는 숫자 열을 사용합니다 . 이 열에는 최대 세 개의 구분되는 값이 포함되어야 합니다 .

- 검증 열에 두 개의 수준이 있는 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 검증 열에 세 개의 수준이 있는 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 검증 열에 네 개 이상의 수준이 있는 경우 가장 작은 세 개의 값을 포함하는 행이 검증 데이터 집합을 정의합니다 . 다른 모든 행은 분석에서 제외됩니다 .

" 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다 . " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 " 검증 열 생성 " 에서 확인하십시오 .

팁 : K 폴드 또는 내포 K 폴드 교차 검증을 사용하려면 모형 선별 플랫폼을 통해 파티션 모형을 적합시킵니다 . 자세한 내용은 " 모형 선별 " 에서 확인하십시오 .

분할의 추가 예

- " 연속형 반응의 예 "
- " 결측값 정보화의 예 "
- " 수익 행렬 및 결정 행렬 보고서의 예 "

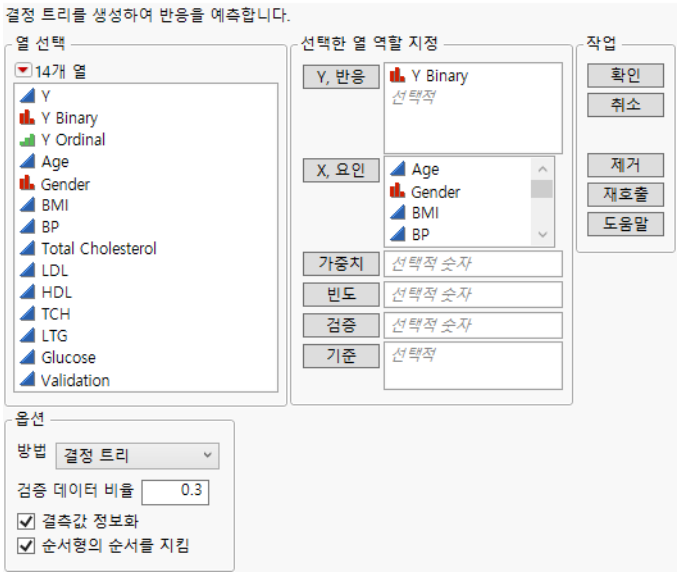
연속형 반응의 예

이 예에서는 파티션 플랫폼을 사용하여 당뇨병 환자의 1 년 질병 진행률 (양적 척도로 측정) 을 예측하는 결정 트리를 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
2. **분석** > **예측 모델링** > **파티션**을 선택합니다 .
3. Y 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Age ~ Glucose 를 선택하고 **X, 요인**을 클릭합니다 .
5. 설치된 JMP 유형에 따라 검증 절차를 선택합니다 .
   - JMP Pro 의 경우 Validation 을 선택하고 **검증**을 클릭합니다 .
   - JMP 의 경우 **검증 데이터 비율**에 "0.3" 을 입력합니다 .

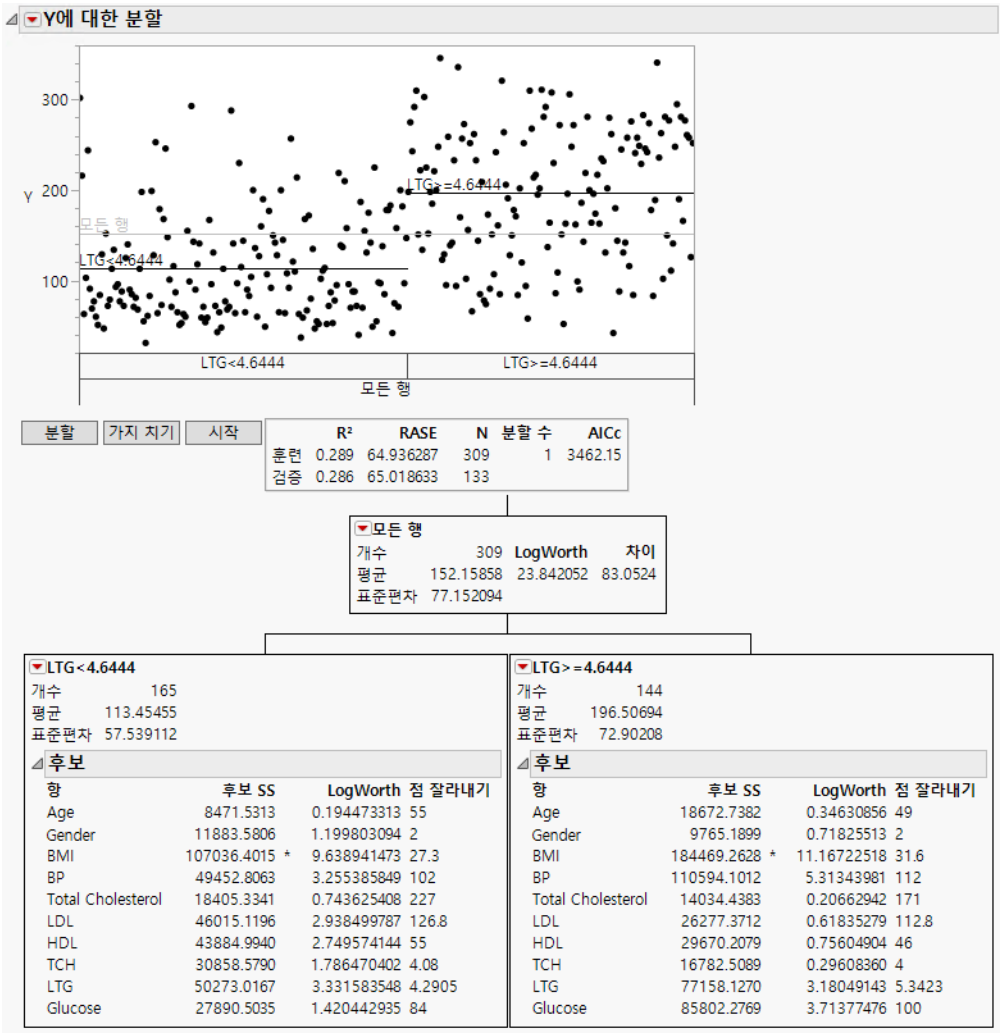
참고 : 검증 행의 랜덤 선택으로 인해 검증 비율을 사용한 결과는 여기에 표시된 것과 다를 수 있습니다 .

그림 4.20 검증 데이터 비율 = 0.3 인 완료된 시작 창



6. **확인**을 클릭합니다 .
7. 플랫폼 보고서 창에서 **분할**을 한 번 클릭하여 분할을 수행합니다 .

그림 4.21 결정 트리가 숨겨진 첫 번째 분할 후 보고서



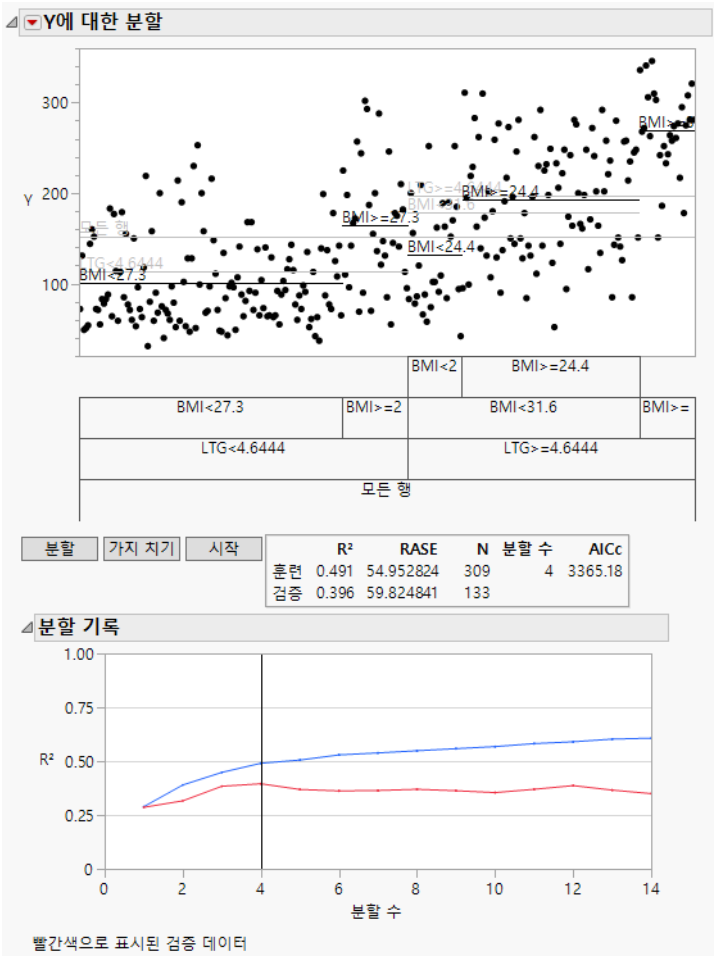
훈련 데이터 집합의 원래 값 309 개가 다음과 같이 두 부분으로 분할되었습니다 .

- 왼쪽 잎 (LTG < 4.6444) 에 165 개의 관측값이 있습니다 .
- 오른쪽 잎 (LTG >= 4.6444) 에 144 개의 관측값이 있습니다 .

다음에는 오른쪽 잎과 왼쪽 잎 모두 BMI 에서 분할됩니다 . BMI 에 대한 후보 SS 는 오른쪽 잎의 값이 왼쪽 잎의 값보다 높습니다 . 따라서 다음 분할은 오른쪽 잎에서 수행됩니다 .

8. **시작**을 클릭하여 자동 분할을 사용합니다 .

그림 4.22 검증을 사용한 자동 분할 후 보고서



네 개의 분할이 포함된 해를 찾았습니다. "분할 기록" 그림에서는 4 번의 분할 후 검증 데이터 집합에서 더 이상 개선이 없음을 보여 줍니다. 검증 데이터의 R^2 값 0.39 는 이 모형이 질병 진행의 강력한 예측 변수로 도움 되지 않음을 나타냅니다. 분할 그림에서 각 파티션 간 산포는 이 모형이 Y 수준을 잘 분리하지 않음을 나타냅니다.

결측값 정보화의 예

이 예에서는 고객이 신용 위험 상태인지 예측하기 위해 결정 트리를 생성합니다. 데이터 집합에 결측값이 포함되어 있으므로 결측값 정보화 옵션의 유효성도 살펴봅니다.

파티션 플랫폼 시작

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Equity.jmp 를 엽니다.

2. **분석** > **예측 모델링** > **파티션**을 선택합니다 .
3. BAD 를 선택하고 **Y, 반응**을 클릭합니다 .
4. LOAN ~ DEBTINC 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

결측값 정보화를 사용하여 결정 트리 및 ROC 곡선 생성

1. Shift 키를 누른 채 **분할**을 클릭합니다 .
2. 분할 수에 "5" 를 입력하고 **확인**을 클릭합니다 .
3. "BAD 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **ROC 곡선**을 선택합니다 .
4. "BAD 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다 .

   확률 (BAD==Good Risk) 및 확률 (BAD==Bad Risk) 열에 결측값 정보화 유틸리티에서 미래 대출 신청인의 신용 위험을 분류하기 위해 사용하는 계산식이 포함됩니다. 결측값 정보화를 사용하지 않는 모형과 비교하여 이 모형이 어떻게 수행되는지에 관심이 있습니다 .

결측값 정보화를 사용하지 않고 결정 트리 및 ROC 곡선 생성

1. "BAD 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **다시 실행 > 분석 다시 시작**을 선택합니다 .
2. **결측값 정보화**를 선택 취소합니다 .
3. " 확인 " 을 클릭하고 " 결측값 정보화를 사용하여 결정 트리 및 ROC 곡선 생성 " 의 단계를 반복합니다 .

   확률 (BAD==Good Risk) 2 및 확률 (BAD==Bad Risk) 2 에 결측값 정보화 유틸리티를 사용하지 않는 계산식이 포함됩니다 .

ROC 곡선 비교

두 모형의 ROC 곡선을 시각적으로 비교해 봅니다. 왼쪽 모형은 결측값 정보화를 사용한 모형이고 오른쪽 모형은 결측값 정보화를 사용하지 않은 모형입니다.

그림 4.23 결측값 정보화를 사용한 모형의 ROC 곡선 (왼쪽) 과 결측값 정보화를 사용하지 않은 모형의 ROC 곡선 (오른쪽)

결측값 정보화를 사용한 모형의 AUC(곡선 아래 면적) 값 0.8695 가 결측값 정보화를 사용하지 않은 모형의 AUC 값 0.7283 보다 높습니다 . 반응 수준이 두 개뿐이므로 각 모형의 ROC 곡선은 서로 다른 모형을 반영하고 AUC 가 같습니다 .

참고 : AUC 는 결측값 정보화를 사용하지 않은 모형에 대해 표시되는 것과 다를 수 있습니다 . 결측값 정보화를 사용하지 않는 경우 결측 행이 분할 측면에 랜덤 할당됩니다 . 분석을 다시 실행하면 결과에 약간 차이가 있을 수 있습니다 .

모형 비교 플랫폼 사용

다음으로 , 모형 비교 플랫폼에서 모형을 비교하여 4 단계 및 3 단계에서 생성한 두 계산식 집합을 비교합니다 .

1. **분석 > 예측 모델링 > 모형 비교**를 선택합니다 .
2. 확률 (BAD==Good Risk), 확률 (BAD==Bad Risk), 확률 (BAD==Good Risk) 2 및 확률 (BAD==Bad Risk) 2 를 선택하고 **Y, 예측 변수**를 클릭합니다 .

   첫 번째 계산식 열 쌍에는 결측값 정보화를 사용하는 모형의 계산식이 포함됩니다 . 두 번째 계산식 열 쌍에는 결측값 정보화를 사용하지 않는 모형의 계산식이 포함됩니다 .
3. **확인**을 클릭합니다 .

그림 4.24 모형 비교의 적합 측도

BAD에 대한 적합 측도

생성자	.2 .4 .6 .8	엔트로피 R^2	일반화 R^2	평균 -Log p	RASE	평균 절대 편차	오분류 비율	N
파티션		0.3813	0.5015	0.3092	0.3013	0.1817	0.1158	5960
파티션		0.1533	0.2248	0.4231	0.3637	0.2647	0.1829	5960

적합 측도 보고서에서는 결측값 정보화를 사용하여 적합된 첫 번째 모형이 결측값 정보화를 사용하지 않고 적합된 두 번째 모형보다 우수하다는 것을 보여 줍니다 . 첫 번째 모형은 R^2 값

이 더 높고 RMSE 값과 오분류 비율이 더 낮습니다 . 이러한 비교는 ROC 곡선 비교를 정렬합니다 .

참고 : 마찬가지로 결측값 정보화가 사용되지 않을 때는 랜덤 차이로 인해 결과가 다를 수 있습니다 .

수익 행렬 및 결정 행렬 보고서의 예

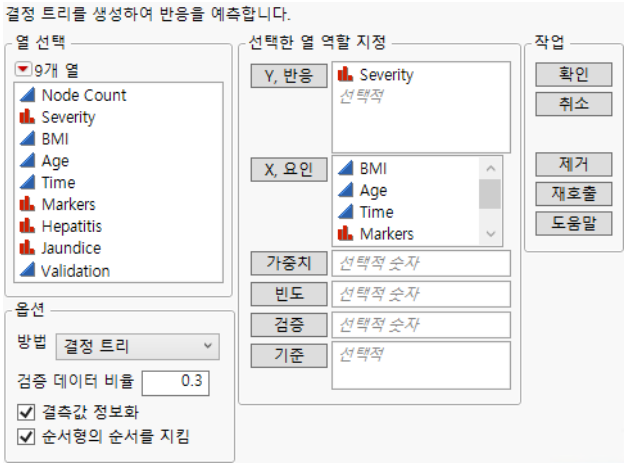
이 예에서는 간암 환자에 대한 연구를 가정합니다. 다양한 측정값 및 표식을 기반으로 각 환자의 질병 심각도(높음 또는 낮음)에 따라 환자를 분류하려고 합니다. 환자를 분류할 때 심각도가 높은 환자를 낮은 그룹으로 분류하거나 심각도가 낮은 환자를 높은 그룹으로 분류하는 두 가지 오류가 발생할 수 있습니다. 임상적으로 심각도가 높은 환자를 낮은 환자로 오분류하면 해당 환자가 자신에게 필요한 공격적 치료를 받지 못할 수 있으므로 비용이 많이 발생합니다. 심각도가 낮은 환자를 심각도가 높은 그룹으로 분류하는 오류는 비용이 더 적게 듭니다. 해당 환자는 필요 이상으로 더 공격적인 치료를 받을 수 있지만 이것은 큰 문제가 아닙니다.

이 예에서는 간암 연구 상황에서 수익 행렬을 정의하고 결정 행렬 보고서를 얻습니다 . " 결정 행렬 " 보고서는 수익 행렬의 비용과 관련하여 분류 비율을 평가하는 데 도움이 됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Liver Cancer.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > 파티션**을 선택합니다 .
3. Severity 를 선택하고 **Y, 반응**을 클릭합니다 .
4. BMI ~ Jaundice 를 선택하고 **X, 요인**을 클릭합니다 .
5. 설치된 JMP 유형에 따라 검증 절차를 선택합니다 .
   - JMP Pro 의 경우 Validation 을 선택하고 **검증**을 클릭합니다 .
   - JMP 의 경우 **검증 비율**에 "0.3" 을 입력합니다 .

참고 : 검증 행의 랜덤 선택으로 인해 검증 비율을 사용한 결과는 여기에 표시된 것과 다를 수 있습니다 .

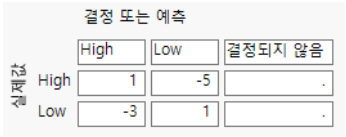
그림 4.25 검증 데이터 비율 = 0.3 인 완료된 시작 창



6. **확인**을 클릭합니다 .
7. Shift 키를 누른 채 **분할**을 클릭합니다 .
8. 분할 수에 "10" 을 입력하고 **확인**을 클릭합니다 .

   그림 아래의 패널에서 " 분할 수 " 가 10 인지 확인합니다 .
9. "Severity 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **수익 행렬 지정**을 선택합니다 .
10. 항목을 다음 값으로 변경합니다 .
    - High, High 상자에 1 입력
    - High, Low 상자에 -5 입력
    - Low, High 상자에 -3 입력
    - Low, Low 상자에 1 입력

그림 4.26 완료된 수익 행렬



팁 : 이후 분석에 사용할 수 있도록 이 수익 행렬을 열 특성으로 저장할 수 있습니다 . 수익 행렬 창 맨 아래의 " 열에 특성으로 저장 " 체크박스를 선택합니다 .

다음 사항에 유의하십시오 .

- 1 값은 각각 올바른 결정을 내릴 때 수익을 반영합니다 .
- -3 값은 심각도가 낮은 환자 (Low) 를 심각도가 높은 환자 (High) 로 분류할 경우 손실이 올바른 결정의 수익보다 3 배 많음을 나타냅니다 .

– -5 값은 심각도가 높은 환자 (High) 를 심각도가 낮은 환자 (Low) 로 분류할 경우 손실이 올바른 결정의 수익보다 5 배 많음을 나타냅니다 .

11. **확인**을 클릭합니다 .
12. "Severity 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **적합 상세 정보 표시**를 선택합니다 .

그림 4.27 혼동 행렬 및 결정 행렬 보고서

적합 상세 정보

측도	훈련	검증	정의
엔트로피 R^2	0.3788	0.0367	1-Loglike(model)/Loglike(0)
일반화 R^2	0.5382	0.0647	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
평균 -Log p	0.4134	0.6402	∑ -Log(ρ[j])/n
RASE	0.3649	0.4694	√ ∑(y[j]-ρ[j])²/n
평균 절대 편차	0.2854	0.3861	∑ |y[j]-ρ[j]|/n
오분류 비율	0.1915	0.3810	∑ (ρ[j]≠ρMax)/n
N	94	42	n

혼동 행렬

훈련

실제값 Severity	예측 개수 Low	High
Low	30	6
High	12	46

검증

실제값 Severity	예측 개수 Low	High
Low	11	5
High	11	15

결정 행렬

훈련

실제값 Severity	결정 개수 Low	High
Low	33	3
High	16	42

실제값 Severity	결정 비율 Low	High
Low	0.917	0.083
High	0.276	0.724

오분류 비율 0.2021

검증

실제값 Severity	결정 개수 Low	High
Low	12	4
High	12	14

실제값 Severity	결정 비율 Low	High
Low	0.750	0.250
High	0.462	0.538

오분류 비율 0.3810

지정된 수익 행렬

실제값	결정 Low	High
Low	1	-5
High	-3	1

" 혼동 행렬 " 보고서와 " 결정 행렬 " 보고서는 " 적합 상세 정보 " 보고서의 " 측도 " 목록을 따릅니다 . " 혼동 행렬 " 보고서와 " 결정 행렬 " 보고서의 혼동 행렬은 개수가 다르게 나타납니다. 이것은 수익 행렬의 가중치가 가중치 없이 예측된 확률과 다른 결정을 내리기 때문입니다.

검증 데이터 집합에 대한 혼동 행렬은 예측 확률에 따른 분류를 보여 줍니다. 이에 따라 High 심각도 환자 11 명이 Low 심각도로 분류되고 Low 심각도 환자 5 명이 High 심각도로 분류됩니다 .

" 결정 행렬 " 보고서에는 수익 행렬 가중치가 통합됩니다 . 이러한 가중치를 사용하면 High 심각도 환자 6 명만 Low 심각도로 분류됩니다 . 그러나 이 경우에는 Low 심각도 환자 6 명이 High 심각도 그룹 (추가 환자 1 명) 으로 잘못 분류되는 문제가 있습니다 .

13. "Severity 에 대한 분할 " 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다 .

8 개의 열이 데이터 테이블에 추가됩니다 .

팁 : 데이터 테이블로 빨리 돌아가려면 보고서 창 오른쪽 아래의 " 연결된 데이터 보기 " 아이콘 (Windows) 또는 도구 모음 메뉴의 " 데이터 테이블 표시 " 아이콘 (macOS) 을 클릭하십시오 .

- 처음 세 열은 예측 확률만 포함합니다 . 혼동 행렬 수는 예측 확률이 가장 높은 수준으로 환자를 분류하는 **최대 확률 분류 Severity** 열을 기반으로 합니다 . 이러한 확률은 **확률 (Severity == High)** 및 **확률 (Severity == Low)** 열에 제공됩니다 .
- 마지막 5 개 열은 수익 행렬 가중치를 포함합니다 . **Severity 에 대한 수익성이 가장 높은 예측** 열에는 수익 행렬을 기반으로 한 결정이 포함됩니다 . 환자의 결정은 수익성이 가장 높은 수준입니다 . 이익은 **High 에 대한 수익** 및 **Low 에 대한 수익** 열에 제공됩니다 .

파티션 플랫폼에 대한 통계 상세 정보

- " 반응 및 요인 "
- " 분할 기준 "
- " 결정 트리 및 붓스트랩 포레스트의 예측 확률 "

반응 및 요인

반응도 연속형 또는 범주형 (명목형 또는 순서형) 일 수 있습니다 .

- 반응이 범주형이면 반응 수준에 대해 추정된 확률을 적합시켜 잔차 로그 가능도 카이제곱 [2* 엔트로피] 을 최소화합니다 .
- 반응이 연속형이면 플랫폼에서 평균을 적합시켜 오차 제곱합을 최소화합니다 .

요인은 연속형 또는 범주형 (명목형 또는 순서형) 일 수 있습니다 .

- 요인이 연속형이면 요인에 대한 분할 " 절단값 " 에 따라 분할이 수행됩니다 .
- 요인이 범주형이면 *X* 범주를 두 개의 수준 그룹으로 나누고 가능한 모든 그룹을 두 수준으로 고려합니다 .

분할 기준

노드 분할은 노드에 대한 " 후보 " 보고서에 나오는 LogWorth 통계량을 기반으로 합니다 . LogWorth 는 다음과 같이 계산됩니다 .

$$-\log_{10}(p \text{ 값})$$

여기서 조정 *p* 값은 분할이 발생할 수 있는 다양한 방법을 고려하는 복합적 형식으로 계산됩니다 . 이 계산은 수준이 많은 *X* 를 우선으로 하는 비조정 *p* 값 및 수준이 적은 *X* 를 우선으로 하는 Bonferroni *p* 값에 비해 매우 공정합니다 . 이 방법에 대한 자세한 내용은 Sall(2002) 연구 자료에서 확인하십시오 .

연속형 반응의 경우 SS(제곱합)가 노드 보고서에 보고됩니다. 이 값은 분할로 인한 오차 제곱합의 변화입니다.

선택된 후보 SS 는 다음과 같습니다 .

$SS_{test} = SS_{parent} - (SS_{right} + SS_{left})$ 여기서 노드의 SS 는 $s^2(n - 1)$ 입니다 .

연속형 반응의 경우 차이 통계량도 보고됩니다. 이 값은 상위 노드의 두 하위 노드에 대한 예측값의 차이입니다.

범주형 반응의 경우 $\mathbf{G}^2$(가능도비 카이제곱)이 보고서에 나타납니다. 실제로 이 값은 [자연 로그] 엔트로피의 두 배 또는 엔트로피 변화의 두 배입니다. 엔트로피는 각 관측값에 대한 Σ -log(p)이며, 여기서 p는 발생한 반응에 기인한 확률입니다.

선택된 후보 G^2 은 다음과 같습니다 .

$G^2_{\text{test}} = G^2_{\text{parent}} - (G^2_{\text{left}} + G^2_{\text{right}})$

파티션에는 실제로 두 가지 비율이 있습니다. 하나는 훈련에 사용되는 일반적인 개수 대 합계 비율이며, 다른 하나는 0에서 약간 벗어난 비율입니다. 귀속 확률이 0이 아니므로 엔트로피 R^2에 사용되는 검증 또는 제외된 데이터 집합에 대해 확률 로그를 계산할 수 있습니다.

결정 트리 및 붓스트랩 포레스트의 예측 확률

결정 트리 및 붓스트랩 포레스트 방법에 대한 예측 확률은 아래 설명과 같이 확률 통계량에 의해 계산됩니다.

결정 트리의 범주형 반응에 대해 " 분할 확률 표시 " 명령은 다음 통계량을 표시합니다 .

비율 각 반응 수준에 대한 노드의 관측값 비율입니다 .

확률 트리의 해당 노드에 대한 예측 확률입니다 . 주어진 노드에서 i 번째 반응 수준에 대한 확률을 계산하는 방법은 다음과 같이 정의됩니다 .

$$\text{Prob}_i = \frac{n_i + \text{prior}_i}{\sum(n_i + \text{prior}_i)}$$

여기서 총합은 모든 반응 수준을 대상으로 합니다 . n_i 는 i 번째 반응 수준에 대한 노드의 관측값 수이고 , prior_i 는 i 번째 반응 수준에 대한 사전 확률이며 다음과 같이 계산됩니다 .

$\text{prior}_i = \lambda p_i + (1-\lambda)P_i$

여기서 p_i 는 상위 노드의 prior_i, P_i 는 상위 노드의 Prob_i, λ 는 현재 0.9 로 설정된 가중치 요인입니다 .

확률 (Prob) 계산 방법은 예측 확률이 항상 0 이 아님을 보장합니다 .

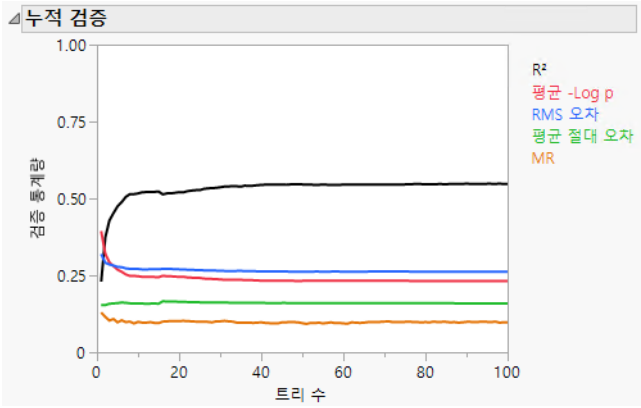
5 장

JMP PRO 붓스트랩 포레스트
여러 트리를 평균화하여 모형 적합

붓스트랩 포레스트 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

붓스트랩 포레스트 플랫폼은 각각 훈련 데이터의 붓스트랩 표본에 적합되는 여러 결정 트리를 평균화하여 앙상블 모형을 적합시킵니다. 각 트리의 각 분할은 예측 변수의 랜덤 부분집합을 고려합니다. 이런 방식으로 여러 약한 모형을 결합하여 더 강력한 모형을 생성합니다. 관측값에 대한 최종 예측은 모든 결정 트리에서 해당 관측값의 예측값을 평균한 것입니다.

그림 5.1 누적 검증 보고서의 예



목차

붓스트랩 포레스트 플랫폼 개요 83
범주형 반응이 있는 붓스트랩 포레스트의 예 83
붓스트랩 포레스트 모형 83
결측값 85
연속형 반응이 있는 붓스트랩 포레스트의 예 86
붓스트랩 포레스트 플랫폼 시작 88
시작 창 88
규격 창 89
붓스트랩 포레스트 보고서 91
모형 검증 집합 요약 93
규격 93
전체 통계량 93
누적 검증 94
트리별 요약 95
붓스트랩 포레스트 플랫폼 옵션 95

JMP PRO 붓스트랩 포레스트 플랫폼 개요

붓스트랩 포레스트 플랫폼은 많은 결정 트리에서 예측된 반응 값의 평균을 구하여 반응 값을 예측합니다. 각 트리는 훈련 데이터의 **붓스트랩 표본**에서 성장합니다. 붓스트랩 표본은 복원 추출 법을 사용한 관측값의 랜덤 표본입니다. 또한 예측 변수는 결정 트리의 각 분할에서 표집됩니다. 결정 트리는 "파티션 모형"에 설명된 재귀 분할 방법을 사용하여 적합됩니다.

다음은 훈련 데이터 집합에 대한 적합 프로세스입니다 .

1. 각 트리에 대해 관측값의 붓스트랩 표본을 선택합니다 .
2. 재귀 분할을 사용하여 각 결정 트리를 적합시킵니다 .
   - 각 분할에 대한 랜덤 예측 변수 집합을 선택합니다 .
   - " 붓스트랩 포레스트 규격 " 창에 지정된 중지 규칙이 충족될 때까지 분할을 계속합니다 .
3. " 붓스트랩 포레스트 규격 " 창에 지정된 트리 수에 도달하거나 조기 중지가 발생할 때까지 1 단계와 2 단계를 반복합니다 .

개별 트리의 경우 트리를 적합시키는 데 사용되는 관측값의 붓스트랩 표본은 복원 추출됩니다. 표집할 관측값 비율을 지정할 수 있습니다. 관측값이 복원 추출되므로 관측값의 100% 를 표집하도록 지정하는 경우 사용되지 않은 관측값의 기대 비율은 1/e 또는 약 36.8% 입니다. 각 개별 트리에 대해 사용되지 않은 이러한 관측값을 Out-Of-Bag 관측값이라고 합니다. 트리를 적합시키는 데 사용된 관측값은 In-Bag 관측값이라고 합니다. 연속형 반응의 경우 붓스트랩 포레스트 플랫폼은 Out-Of-Bag 관측값의 오차율에 대한 측도를 제공합니다 (Out-Of-Bag 오차).

연속형 반응의 경우 관측값의 예측값은 개별 트리 모음에 대한 예측값의 평균입니다. 범주형 반응의 경우 관측값의 예측 확률은 개별 트리 모음에 대한 예측 확률의 평균입니다. 관측값은 예측 확률이 가장 높은 수준으로 분류됩니다.

붓스트랩 포레스트에 대한 자세한 내용은 Hastie et al. (2009) 연구 자료에서 확인하십시오.

JMP PRO 범주형 반응이 있는 붓스트랩 포레스트의 예

이 예에서는 고객의 신용 위험이 낮은지 여부를 예측하기 위해 붓스트랩 포레스트 모형을 생성합니다 . 그러나 데이터 집합에 결측값이 포함되어 있으므로 결측값의 정도도 살펴봅니다 .

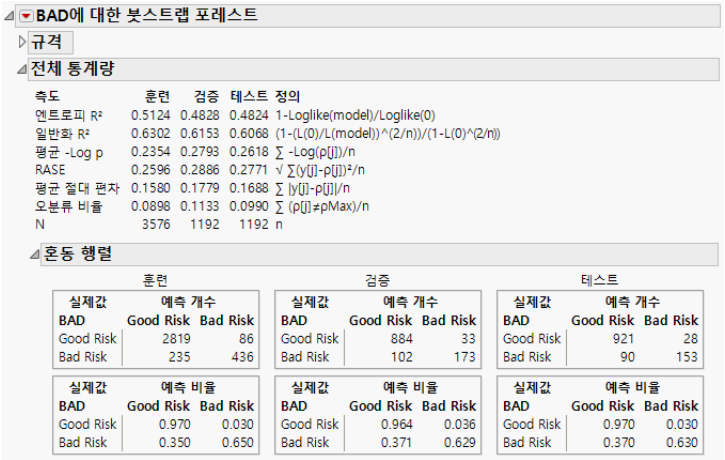
JMP PRO 붓스트랩 포레스트 모형

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Equity.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > 붓스트랩 포레스트**를 선택합니다 .
3. BAD 를 선택하고 **Y, 반응**을 클릭합니다 .
4. LOAN ~ DEBTINC 를 선택하고 **X, 요인**을 클릭합니다 .
5. Validation 을 선택하고 **검증**을 클릭합니다 .

6. **확인**을 클릭합니다 .
7. " 분할당 표집되는 항 수 " 옆에 "4" 를 입력합니다 .
8. " 트리당 최대 분할 수 " 옆에 "30" 을 입력합니다 .
9. **여러 항에 대한 다중 적합**을 선택하고 " 최대 항 수 " 옆에 "10" 을 입력합니다 .
10. (선택 사항) **멀티스레딩 제한**을 선택하고 " 난수 시드값 " 옆에 "123" 을 입력합니다 .

    붓스트랩 포레스트 방법에는 랜덤 표집이 포함되므로 이렇게 하면 아래 표시된 결과와 정확히 동일한 결과를 얻을 수 있습니다 .
11. **확인**을 클릭합니다 .

그림 5.2 전체 통계량 보고서



BAD에 대한 붓스트랩 포레스트

규격

전체 통계량

측도	훈련	검증	테스트	정의
엔트로피 R^2	0.5124	0.4828	0.4824	1-Loglike(model)/Loglike(0)
일반화 R^2	0.6302	0.6153	0.6068	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
평균 -Log p	0.2354	0.2793	0.2618	∑ -Log(ρ[j])/n
RASE	0.2596	0.2886	0.2771	√ ∑(y[j]-ρ[j])²/n
평균 절대 편차	0.1580	0.1779	0.1688	∑ |y[j]-ρ[j]|/n
오분류 비율	0.0898	0.1133	0.0990	∑ (ρ[j]≠ρMax)/n
N	3576	1192	1192	n

혼동 행렬

훈련

실제값	예측 개수	
BAD	Good Risk	Bad Risk
Good Risk	2819	86
Bad Risk	235	436

실제값	예측 비율	
BAD	Good Risk	Bad Risk
Good Risk	0.970	0.030
Bad Risk	0.350	0.650

검증

실제값	예측 개수	
BAD	Good Risk	Bad Risk
Good Risk	884	33
Bad Risk	102	173

실제값	예측 비율	
BAD	Good Risk	Bad Risk
Good Risk	0.964	0.036
Bad Risk	0.371	0.629

테스트

실제값	예측 개수	
BAD	Good Risk	Bad Risk
Good Risk	921	28
Bad Risk	90	153

실제값	예측 비율	
BAD	Good Risk	Bad Risk
Good Risk	0.970	0.030
Bad Risk	0.370	0.630

" 여러 항에 대한 다중 적합 " 옵션이 지정되었으므로 4, 5, 6, 8, 10 을 각 분할의 예측 변수 수로 사용하여 모형이 생성되었습니다 . " 모형 검증 집합 요약 " 보고서에서는 검증 데이터 집합의 엔트로피 R^2 이 가장 높은 모형이 5 항 모형임을 보여 줍니다 . 이 모형은 오분류 비율이 가장 작은 모형이기도 합니다 . 이 모형이 최적 모형으로 결정되고 이 모형에 대한 결과가 전체 보고서에 표시됩니다 .

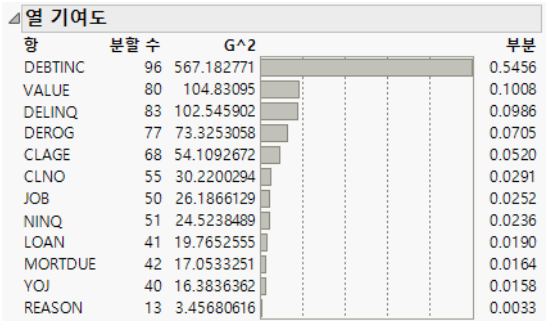
전체 보고서에서는 검증 데이터 집합과 테스트 데이터 집합의 오분류 비율이 각각 11.3% 와 9.9% 임을 보여 줍니다 . 혼동 행렬은 Bad Risk 고객을 Good Risk 고객으로 분류한 것이 오분류의 가장 큰 원인임을 나타냅니다 .

테스트 데이터 집합의 결과는 모형이 독립적 관측값으로 얼마나 잘 확장되는지를 나타냅니다 . 검증 데이터 집합은 붓스트랩 포레스트 모형을 선택하는 데 사용되었습니다 . 따라서 검증 데이터 집합의 결과는 모형이 독립 데이터로 일반화되는 방식을 나타내는 데 더 관심을 둡니다 .

모형에 가장 많이 기여한 예측 변수를 확인하려고 합니다 .

12. "BAD 에 대한 붓스트랩 포레스트 " 옆의 빨간색 삼각형을 클릭하고 **열 기여도**를 선택합니다 .

그림 5.3 열 기여도 보고서



열 기여도

항	분할 수	G^2	부분
DEBTINC	96	567.182771	0.5456
VALUE	80	104.83095	0.1008
DELINQ	83	102.545902	0.0986
DEROG	77	73.3253058	0.0705
CLAGE	68	54.1092672	0.0520
CLNO	55	30.2200294	0.0291
JOB	50	26.1866129	0.0252
NINQ	51	24.5238489	0.0236
LOAN	41	19.7652555	0.0190
MORTDUE	42	17.0533251	0.0164
YOJ	40	16.3836362	0.0158
REASON	13	3.45680616	0.0033

" 열 기여도 " 보고서에 따르면 신용 위험에 대해 가장 강력한 예측 변수는 총 부채 상환 비율을 나타내는 DEBTINC 입니다 . 다음으로 모형에 가장 많이 기여하는 변수는 연체 신용 한도 수를 나타내는 DELINQ 와 고객의 사정 평가액을 나타내는 VALUE 입니다 .

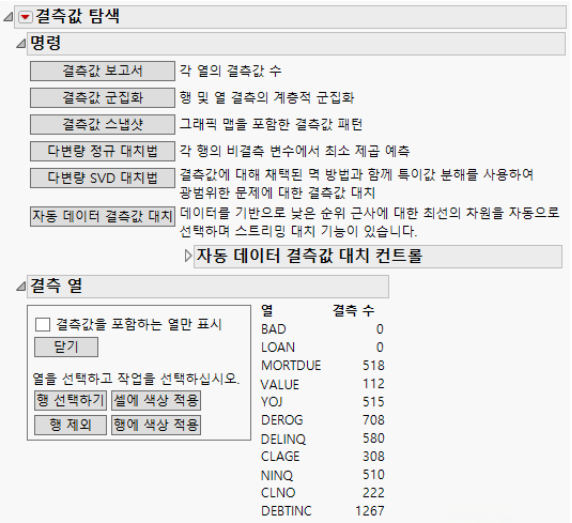
JMP PRO 결측값

이제 예측 변수 값을 결측값으로 판단하는 정도를 살펴봅니다 .

1. **분석 > 선별 > 결측값 탐색**을 선택합니다 .
2. Bad ~ DEBTINC 를 선택하고 **Y, 열**을 클릭합니다 .
3. 경고가 나타나면 **확인**을 클릭합니다 .

   REASON 및 JOB 열은 문자 데이터 유형이 있으므로 "Y, 열 " 목록에 추가되지 않습니다 . " 분포 "(이 예에서 설명하지 않음)를 사용하여 이 두 열에 대해 결측값 수를 확인할 수 있습니다.
4. **확인**을 클릭합니다 .

그림 5.4 결측값 보고서



DEBTINC 열에는 관측값의 약 21% 에 해당하는 1267 개의 결측값이 있습니다 . 붓스트랩 포레스트 분석에 포함된 대부분의 다른 열에도 결측값이 포함되어 있습니다 . 시작 창의 " 결측값 정보화 " 옵션을 사용할 경우 결측값은 해당 값이 전달하는 정보가 인식되도록 처리됩니다 . 자세한 내용은 " 결측값 정보화 " 에서 확인하십시오 .

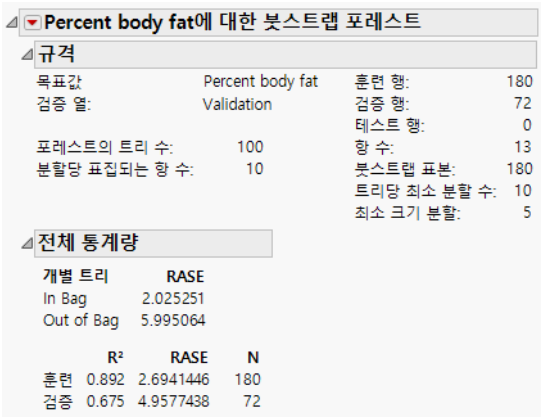
JMP PRO 연속형 반응이 있는 붓스트랩 포레스트의 예

이 예에서는 남성 피실험자의 체지방률을 예측하기 위해 붓스트랩 포레스트 모형을 생성합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Body Fat.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > 붓스트랩 포레스트**를 선택합니다 .
3. Percent body fat 을 선택하고 **Y, 반응**을 클릭합니다 .
4. Age (years) ~ Wrist circumference (cm) 를 선택하고 **X, 요인**을 클릭합니다 .
5. Validation 을 선택하고 **검증**을 클릭합니다 .
6. **확인**을 클릭합니다 .
7. (선택 사항) **멀티스레딩 제한**을 선택하고 " 난수 시드값 " 옆에 "123" 을 입력합니다 .

   붓스트랩 포레스트 방법에는 랜덤 표집이 포함되므로 이렇게 하면 아래 표시된 결과와 정확히 동일한 결과를 얻을 수 있습니다 .
8. **확인**을 클릭합니다 .

그림 5.5 전체 통계량



Percent body fat에 대한 붓스트랩 포레스트

규격

목표값	Percent body fat	훈련 행:	180
검증 열:	Validation	검증 행:	72
		테스트 행:	0
포레스트의 트리 수:	100	항 수:	13
분할당 표집되는 항 수:	10	붓스트랩 표본:	180
		트리당 최소 분할 수:	10
		최소 크기 분할:	5

전체 통계량

개별 트리	RASE
In Bag	2.025251
Out of Bag	5.995064

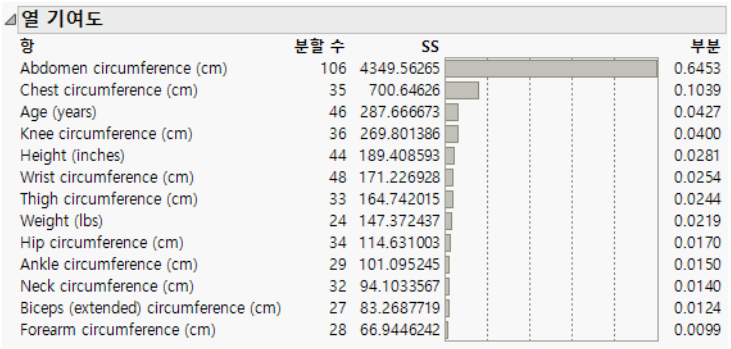
	R^2	RASE	N
훈련	0.892	2.6941446	180
검증	0.675	4.9577438	72

" 전체 통계량 " 보고서에는 " 검증 R^2" 이 0.675 로 표시됩니다 .

모형과 별개로 가장 중요한 예측 변수를 나타내려고 합니다 .

9. "Percent body fat 에 대한 붓스트랩 포레스트 " 옆의 빨간색 삼각형을 클릭하고 **열 기여도**를 선택합니다 .

그림 5.6 열 기여도



열 기여도

항	분할 수	SS	부분
Abdomen circumference (cm)	106	4349.56265	0.6453
Chest circumference (cm)	35	700.64626	0.1039
Age (years)	46	287.666673	0.0427
Knee circumference (cm)	36	269.801386	0.0400
Height (inches)	44	189.408593	0.0281
Wrist circumference (cm)	48	171.226928	0.0254
Thigh circumference (cm)	33	164.742015	0.0244
Weight (lbs)	24	147.372437	0.0219
Hip circumference (cm)	34	114.631003	0.0170
Ankle circumference (cm)	29	101.095245	0.0150
Neck circumference (cm)	32	94.1033567	0.0140
Biceps (extended) circumference (cm)	27	83.2687719	0.0124
Forearm circumference (cm)	28	66.9446242	0.0099

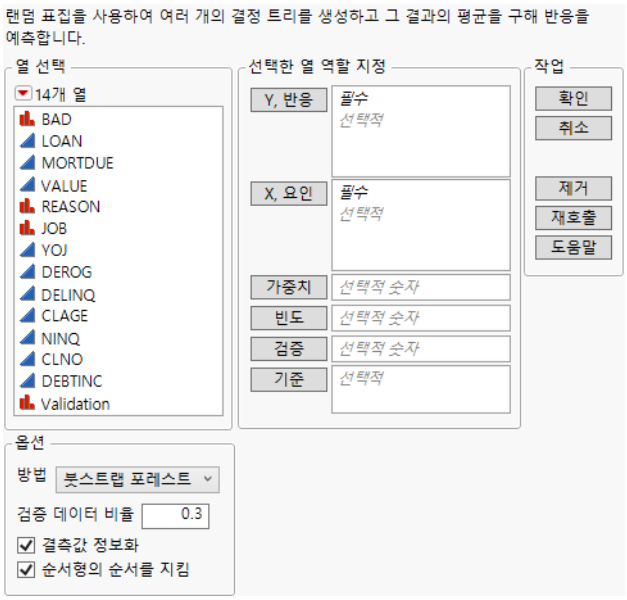
" 열 기여도 " 보고서에 따르면 Abdomen circumference (cm), Chest circumference (cm) 및 Age (years) 가 Percent body fat 에 대한 가장 강력한 예측 변수입니다 .

JMP PRO 붓스트랩 포레스트 플랫폼 시작

분석 > 예측 모델링 > 붓스트랩 포레스트를 선택하여 붓스트랩 포레스트 플랫폼을 시작합니다 .

JMP PRO 시작 창

그림 5.7 붓스트랩 포레스트 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

붓스트랩 포레스트 플랫폼 시작 창에는 다음과 같은 옵션이 있습니다 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 .

X, 요인 예측 변수입니다 .

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다 .

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다 .

검증 검증 데이터 집합을 정의하는 숫자 열입니다 . 이 열에는 최대 세 개의 구분되는 값이 포함되어야 합니다 .

- 검증 열에 두 개의 수준이 있는 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 검증 열에 세 개의 수준이 있는 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .

– 검증 열에 네 개 이상의 수준이 있는 경우 가장 작은 세 개의 값을 포함하는 행이 검증 데이터 집합을 정의합니다. 다른 모든 행은 분석에서 제외됩니다.

붓스트랩 포레스트 플랫폼에서는 검증 열을 사용하여 모형을 훈련 및 조정하거나 모형을 훈련, 조정 및 평가합니다. 검증에 대한 자세한 내용은 "JMP 모델링의 검증"에서 확인하십시오.

"열 선택" 목록에서 아무 열도 선택하지 않은 상태로 "검증" 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다. "검증 열 생성" 유틸리티에 대한 자세한 내용은 "검증 열 생성"에서 확인하십시오.

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다. 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다. 결과는 개별 보고서에 제공됩니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

방법 분할 방법 (결정 트리, 붓스트랩 포레스트, 부스티드 트리, K 최근접 이웃 또는 Naive Bayes)을 선택할 수 있습니다. 결정 트리를 제외한 이 대체 방법은 JMP Pro 에서 사용할 수 있습니다.

이러한 방법에 대한 자세한 내용은 "파티션 모형", "부스티드 트리", "K 최근접 이웃" 및 "Naive Bayes"에서 확인하십시오.

검증 데이터 비율 검증 데이터 집합으로 사용할 데이터 비율입니다.

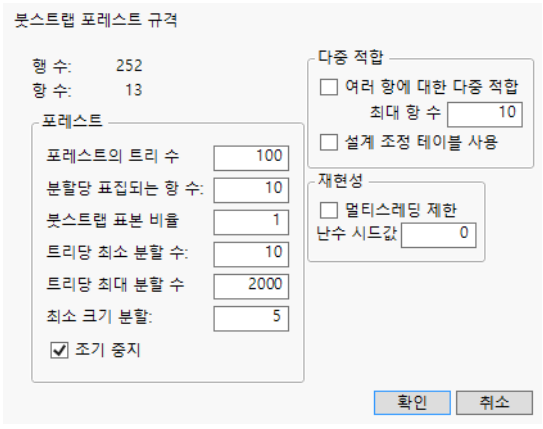
결측값 정보화 이 옵션을 선택하면 범주형 예측 변수에 대한 결측값 범주화 및 연속형 예측 변수에 대한 결측값 정보화 처리가 활성화됩니다. 자세한 내용은 "결측값 정보화"에서 확인하십시오.

순서형의 순서를 지킴 이 옵션을 선택하면 순서를 유지하는 분할만 고려하도록 제한됩니다.

JMP PRO 규격 창

시작 창에서 **확인**을 선택하면 "붓스트랩 포레스트 규격" 창이 나타납니다.

그림 5.8 붓스트랩 포레스트 규격 창



JMP PRO 규격 패널

행 수 데이터 테이블의 행 수입니다 .

항 수 예측 변수로 지정된 열 수입니다 .

JMP PRO 포레스트 패널

포레스트의 트리 수 증가 후 평균을 구할 트리 수입니다 .

분할당 표집되는 항 수 각 분할에서 분할 후보로 고려할 예측 변수의 수입니다 . 각 분할에 대해 예측 변수의 새 랜덤 표본이 후보 집합으로 사용됩니다 .

붓스트랩 표본 비율 각 트리를 증가시키기 위해 표집할 관측값의 비율입니다 (복원 추출법 사용). 각 트리에 대해 새 랜덤 표본이 생성됩니다 .

트리당 최소 분할 수 각 트리의 최소 분할 수입니다 .

트리당 최대 분할 수 각 트리의 최대 분할 수입니다 .

최소 크기 분할 후보 분할에 필요한 최소 관측값 수입니다 .

조기 중지 (검증을 사용하는 경우에만 사용 가능) 이 옵션을 선택하면 트리를 추가해도 검증 통계량이 개선되지 않을 경우 추가 트리 증가가 중지됩니다 . 검증 통계량은 검증 데이터 집합의 엔트로피 R^2 값 (범주형 반응의 경우) 및 R^2 값 (연속형 반응의 경우) 입니다 . 이 옵션을 선택하지 않으면 지정된 트리 수에 도달할 때까지 프로세스가 계속됩니다 .

JMP PRO 다중 적합 패널

여러 항에 대한 다중 적합 이 옵션을 선택하면 분할당 표집된 항 수의 여러 값에 대한 붓스트랩 포레스트가 생성됩니다 . 결과가 표시되는 모형은 검증 데이터 집합의 엔트로피 R^2 값 (범주형 반응의 경우) 또는 R^2 값 (연속형 반응의 경우) 이 가장 큰 모형입니다 .

하한은 " 분할당 표집되는 항 수 " 규격 값입니다 . 상한은 다음 옵션으로 지정됩니다 .

최대 항 수 분할에 대해 고려할 최대 항 수입니다 .

설계 조정 테이블 사용 " 포레스트 " 패널 조정 모수의 값이 포함된 데이터 테이블 (**설계 조정 테이블**) 을 선택할 수 있는 창을 엽니다 . 설계 조정 테이블에는 지정할 각 옵션에 대한 열과 각각 단일 붓스트랩 포레스트 모형 설계를 나타내는 하나 이상의 행이 있습니다 . 설계 조정 테이블에 옵션이 지정되지 않으면 기본값이 사용됩니다 .

JMP 에서는 지정된 조정 모수를 사용하여 테이블의 각 행에 대해 붓스트랩 포레스트 모형을 생성합니다 . 설계 조정 테이블에 둘 이상의 모형이 지정된 경우 " 모형 검증 집합 요약 " 보고서에 각 모형의 R^2 값이 나열됩니다 . " 붓스트랩 포레스트 " 보고서에는 R^2 값이 가장 큰 모형의 적합 통계량이 표시됩니다 .

실험 설계 기능을 사용하여 설계 조정 테이블을 생성할 수 있습니다 . 붓스트랩 포레스트 설계 조정 테이블에는 다음 열이 임의의 순서로 포함될 수 있습니다 .

– 트리 수

- 항 수
- 부분 붓스트랩
- 트리당 최소 분할 수
- 트리당 최대 분할 수
- 최소 크기 분할

JMP PRO 재현성 패널

멀티스레딩 제한 이 옵션을 선택하면 모든 계산이 단일 스레드에서 수행됩니다.

난수 시드값 이후 플랫폼 실행을 위해 결과를 재현하는 데 사용할 0 이 아닌 숫자 난수 시드값을 지정합니다. 기본적으로 난수 시드값은 0 으로 설정되어 재현 가능한 결과를 생성하지 않습니다. 분석을 스크립트에 저장하면 사용자가 입력한 난수 시드값이 스크립트에 저장됩니다.

JMP PRO 붓스트랩 포레스트 보고서

"붓스트랩 포레스트 규격" 창에서 **확인**을 클릭하면 JMP에서 모형을 적합시키는 데 걸리는 대략적인 시간을 알려 주는 창이 나타납니다. 예정 시간이 너무 길면 **현재 모형 채택**을 선택하여 현재 모형을 볼 수도 있습니다. 전체 모형이 적합되거나 현재 모형이 채택된 후 "붓스트랩 포레스트" 보고서가 나타납니다.

그림 5.9 현재 모형 채택 대화상자

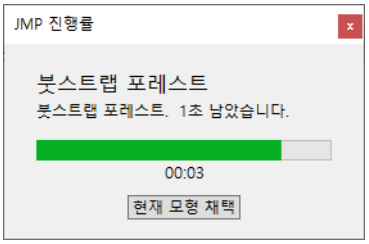
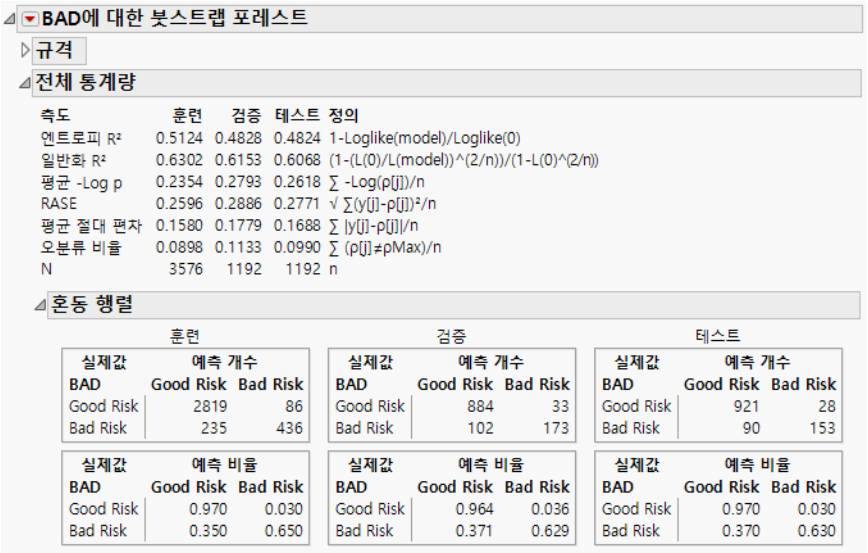


그림 5.10 범주형 반응에 대한 붓스트랩 포레스트 보고서



BAD에 대한 붓스트랩 포레스트

규격

전체 통계량

측도	훈련	검증	테스트	정의
엔트로피 R^2	0.5124	0.4828	0.4824	1-Loglike(model)/Loglike(0)
일반화 R^2	0.6302	0.6153	0.6068	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
평균 -Log p	0.2354	0.2793	0.2618	∑ -Log(ρ[j])/n
RASE	0.2596	0.2886	0.2771	√ ∑(y[j]-ρ[j])²/n
평균 절대 편차	0.1580	0.1779	0.1688	∑ |y[j]-ρ[j]|/n
오분류 비율	0.0898	0.1133	0.0990	∑ (ρ[j]≠ρMax)/n
N	3576	1192	1192	n

혼동 행렬

훈련

실제값	예측 개수	
BAD	Good Risk	Bad Risk
Good Risk	2819	86
Bad Risk	235	436

실제값	예측 비율	
BAD	Good Risk	Bad Risk
Good Risk	0.970	0.030
Bad Risk	0.350	0.650

검증

실제값	예측 개수	
BAD	Good Risk	Bad Risk
Good Risk	884	33
Bad Risk	102	173

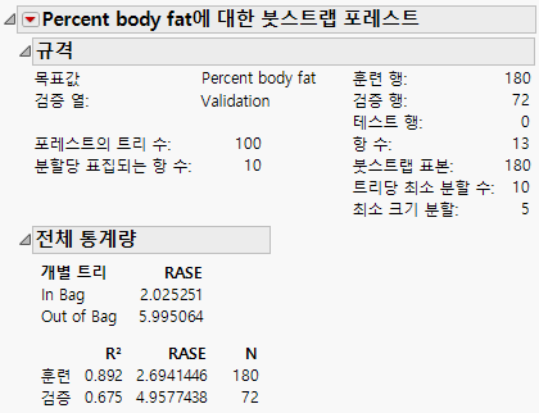
실제값	예측 비율	
BAD	Good Risk	Bad Risk
Good Risk	0.964	0.036
Bad Risk	0.371	0.629

테스트

실제값	예측 개수	
BAD	Good Risk	Bad Risk
Good Risk	921	28
Bad Risk	90	153

실제값	예측 비율	
BAD	Good Risk	Bad Risk
Good Risk	0.970	0.030
Bad Risk	0.370	0.630

그림 5.11 연속형 반응에 대한 붓스트랩 포레스트 보고서



Percent body fat에 대한 붓스트랩 포레스트

규격

목표값	Percent body fat	훈련 행:	180
검증 열:	Validation	검증 행:	72
		테스트 행:	0
포레스트의 트리 수:	100	항 수:	13
분할당 표집되는 항 수:	10	붓스트랩 표본:	180
		트리당 최소 분할 수:	10
		최소 크기 분할:	5

전체 통계량

개별 트리	RASE
In Bag	2.025251
Out of Bag	5.995064

	R^2	RASE	N
훈련	0.892	2.6941446	180
검증	0.675	4.9577438	72

반응이 범주형인지 아니면 연속형인지에 따라 다음과 같은 보고서가 제공됩니다 .

- " 모형 검증 집합 요약 "
- " 규격 "
- " 전체 통계량 "
- " 누적 검증 "
- " 트리별 요약 "

JMP PRO 모형 검증 집합 요약

("붓스트랩 포레스트 규격" 창에서 "여러 항에 대한 다중 적합" 옵션을 선택한 경우 사용 가능) 모든 모형 적합에 대한 적합 통계량을 제공합니다. 자세한 내용은 그림 5.10 및 "다중 적합 패널"에서 확인하십시오.

JMP PRO 규격

모형 적합에 사용된 설정을 표시합니다.

JMP PRO 전체 통계량

훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 적합 통계량을 제공합니다. 보고서의 특정 형식은 반응의 모델링 유형에 따라 다릅니다.

"붓스트랩 포레스트 규격" 창의 "여러 항에 대한 다중 적합" 옵션을 사용하여 여러 모형이 적합된다고 가정해 보겠습니다. 그러면 "전체 통계량" 및 "누적 검증" 보고서에 결과가 표시되는 모형은 검증 데이터 집합의 엔트로피 R^2 값(범주형 반응의 경우) 또는 R^2 값(연속형 반응의 경우)이 가장 큰 모형입니다.

JMP PRO 범주형 반응

측도 보고서

훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 다음 통계량을 제공합니다.

참고 : 엔트로피 R^2 및 일반화 R^2 의 경우 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 . 평균 -Log p, RASE, 평균 절대 편차 및 오분류 비율의 경우 값이 작을수록 더 나은 적합을 나타냅니다.

엔트로피 R^2 적합 모형과 상수 확률 모형의 로그 가능도를 비교하는 적합 측도입니다 . 값 범위는 0 에서 1 사이입니다 . 자세한 내용은 " 엔트로피 R^2" 에서 확인하십시오 .

일반화 R^2 일반 회귀 모형에 적용할 수 있는 측도입니다 . 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다 . 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다 . 일반화 R^2 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2 로 단순화됩니다 . 일반화 R^2 을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데 , 이는 Cox-Snell 유사 R^2 을 정규화한 것입니다 .

평균 -Log P 음의 $\log(p)$ 의 평균입니다 . 여기서 p 는 발생한 사건과 연관된 적합 확률입니다 .

RASE 제곱근 평균 제곱 예측 오차입니다 . 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 .

평균 절대 편차 반응과 예측 반응 사이의 차이에 대한 절대값의 평균입니다. 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다.

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다.

N 관측값 수입니다.

혼동 행렬 보고서

(범주형 반응에만 사용 가능) 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 분류 통계량을 표시합니다. "혼동 행렬" 보고서에는 혼동 행렬과 혼동 비율 행렬이 포함됩니다. 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다. 혼동 비율 행렬은 값을 행 합계로 나눈 혼동 행렬과 동일합니다.

결정 행렬

(범주형 반응을 사용하고 반응에 "수익 행렬" 열 특성이 있거나 "수익 행렬 지정" 옵션을 사용하여 비용을 지정하는 경우에만 사용 가능) 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 결정 개수 및 결정 비율 행렬을 제공합니다. 자세한 내용은 "분할의 추가 예"에서 확인하십시오.

JMP PRO 연속형 반응

개별 트리 보고서

모든 트리에 대해 In Bag 관측값과 Out of Bag 관측값의 평균인 RASE 값을 제공합니다. 트리를 생성하는 데 사용된 훈련 데이터 집합 관측값을 In-Bag 관측값이라고 합니다. 트리를 생성하는 데 사용되지 않는 훈련 관측값을 OOB(Out-Of-Bag) 관측값이라고 합니다.

각 트리에 대해 Out of Bag RASE는 오차 제곱합의 제곱근을 OOB 관측값 수로 나눈 값입니다. 각 트리에 대한 Out of Bag RASE 제곱은 "트리별 요약" 보고서에 "OOB SSE/N"으로 제공됩니다.

R^2 및 RASE 보고서

훈련 데이터 집합과 검증 및 테스트 데이터 집합(정의된 경우)에 대한 R^2, 제곱근 평균 제곱 예측 오차 및 관측값 수를 제공합니다.

JMP PRO 누적 검증

(검증을 사용하는 경우에만 사용 가능) 검증 데이터 집합 대 트리 수에 대한 적합 통계량 그림을 표시합니다.

연속형 반응의 단일 적합 통계량은 R^2 입니다. 범주형 반응의 적합 통계량은 아래에 나열되어 있으며 " 측도 보고서 " 에서 설명합니다.

- R^2(엔트로피 R^2)

- 평균 -Log p(평균 -Log p)
- 제곱근 평균 제곱 예측 오차 (RASE)
- 평균 절대 오차 (평균 절대 편차)
- MR(오분류 비율)

"누적 검증" 그림 아래의 "누적 상세 정보" 보고서에는 그림에 사용된 값이 제공됩니다.

JMP PRO 트리별 요약

" 트리별 요약 " 보고서에는 In-Bag 및 Out-Of-Bag 관측값 개념이 포함됩니다 . 개별 트리의 경우 트리를 적합시키는 데 사용되는 관측값의 붓스트랩 표본은 복원 추출됩니다 . 관측값이 복원 추출되므로 관측값의 100% 를 표집하도록 지정해도 사용되지 않은 관측값의 기대 비율은 $1/e$ 입니다 . 각 개별 트리에 대해 사용되지 않은 관측값을 Out-Of-Bag 관측값이라고 합니다 . 트리를 적합시키는 데 사용된 관측값은 In-Bag 관측값이라고 합니다 .

" 트리별 요약 " 보고서에는 각 트리에 대해 다음과 같은 요약 통계량이 표시됩니다 .

분할 결정 트리의 분할 수입니다 .

순위 트리의 OOB 손실 순위입니다(오름차순). OOB 손실이 가장 작은 트리의 순위가 1입니다.

OOB 손실 트리가 Out Of Bag 행에 적용될 경우 총 예측 부정확성을 나타내는 측도입니다 . 값이 작을수록 예측 정확도가 높다는 것을 나타냅니다 .

OOB 손실 /N OOB 손실을 OOB 행 수 , 즉 OOB N 으로 나눈 값입니다 .

R^2 (연속형 반응에만 사용 가능) 트리의 R^2 값입니다 .

IB SSE (연속형 반응에만 사용 가능) In Bag 행에 대한 오차 제곱합입니다 .

IB SSE/N (연속형 반응에만 사용 가능) In Bag 행에 대한 오차 제곱합을 In Bag 관측값 수로 나눈 값입니다 . In Bag 관측값 수는 훈련 데이터 집합의 관측값 수에 " 붓스트랩 포레스트 규격 " 창에 지정한 붓스트랩 표집 비율을 곱한 값과 같습니다 .

OOB N (연속형 반응에만 사용 가능) Out Of Bag 행 수입니다 .

OOB SSE (연속형 반응에만 사용 가능) 트리가 Out Of Bag 행에 적용될 경우 오차 제곱합입니다 .

OOB SSE/N (연속형 반응에만 사용 가능) OOB SSE 를 OOB 행 수 , 즉 OOB N 으로 나눈 값입니다 .

JMP PRO 붓스트랩 포레스트 플랫폼 옵션

" 붓스트랩 포레스트 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

실제값 대 예측값 그림 (연속형 반응에만 사용 가능) 실제값 대 예측값 그림을 제공합니다 .

열 기여도 적합에 대한 각 입력 열의 기여도를 보여 주는 보고서를 표시합니다 . 다음 항목도 보고서에 표시됩니다 .

- 지정된 열을 사용하여 데이터를 분할할 때 모든 트리에 대한 총 인스턴스 수
- 열에 기인한 총 G^2(범주형 반응의 경우) 또는 SS, 즉 제곱합 (연속형 반응의 경우)
- G^2 또는 SS 의 막대 차트
- 열에 기인한 G^2 또는 SS 의 비율

트리 표시 " 트리 보기 " 보고서에 트리를 표시하기 위한 다양한 옵션을 제공합니다 . 이 보고서에는 부스트 프로세스의 각 계층에서 적합된 트리 그림이 제공됩니다 . " 이름 범주 추정값 표시 " 옵션에 의해 표시되는 확률 열에 대한 설명은 " 결정 트리 및 붓스트랩 포레스트의 예측 확률 " 에서 확인하십시오 .

ROC 곡선 (범주형 반응에만 사용 가능) 자세한 내용은 "ROC 곡선 " 에서 확인하십시오 .

향상도 곡선 (범주형 반응에만 사용 가능) 자세한 내용은 " 향상도 곡선 " 에서 확인하십시오 .

열 저장 모형 및 트리 결과를 저장하고 SAS 코드를 생성하기 위한 옵션이 포함되어 있습니다 .

예측값 저장 모형의 예측값을 데이터 테이블에 저장합니다 .

예측 계산식 저장 예측 계산식을 데이터 테이블의 열에 저장합니다 . 계산식은 트리 구조를 설명하는 내포된 조건부 절로 구성됩니다 . 반응이 연속형이면 열에 " 예측 " 특성이 포함됩니다 . 반응이 범주형이면 열에 " 반응 확률 " 특성이 포함됩니다 .

허용 예측 계산식 저장 (이 옵션 대신 " 예측 계산식 저장 " 옵션을 사용해야 합니다 . " 예측 계산식 저장 " 을 사용할 수 없을 때만 이 옵션을 사용하십시오 .) 결측값이 있고 " 결측값 정보화 " 를 선택하지 않은 경우에도 예측하는 계산식을 저장합니다 . 예측 계산식은 결측 예측 변수에 대한 반응 값을 분할에 랜덤 할당하여 결측값을 허용합니다 . 반응이 연속형이면 열에 " 예측 " 특성이 포함됩니다 . 반응이 범주형이면 열에 " 반응 확률 " 특성이 포함됩니다 . " 결측값 정보화 " 를 선택한 경우 Shift 키를 누른 채 보고서의 빨간색 삼각형을 클릭하면 허용 예측 계산식을 저장할 수 있습니다 .

잔차 저장 (연속형 반응에만 사용 가능) 잔차를 데이터 테이블에 저장합니다 .

누적 상세 정보 저장 (검증을 사용하는 경우에만 사용 가능) 각 트리에 대한 적합 통계량이 포함된 데이터 테이블을 생성합니다 .

예측 계산식 게시 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

허용 예측 계산식 게시 (이 옵션 대신 " 예측 계산식 게시 " 옵션을 사용해야 합니다 . " 예측 계산식 게시 " 를 사용할 수 없을 때만 이 옵션을 사용하십시오 .) 허용 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 . " 결측값 정보화 " 를 선택한 경우 Shift 키를 누른 채 보고서의 빨간색 삼각형을 클릭하면 이 옵션을 사용할 수 있습니다 .

SAS DATA 스텝 만들기 새 데이터 집합을 스코어링하기 위한 SAS 코드를 생성합니다 .

수익 행렬 지정 (범주형 반응에만 사용 가능) 올바르거나 올바르지 않은 분류 결정과 관련된 수익 또는 비용을 지정할 수 있습니다. 자세한 내용은 "적합 상세 정보 표시"에서 확인하십시오.

프로파일러 예측 프로파일러를 표시합니다. 자세한 내용은 Profilers 의 에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

6 장

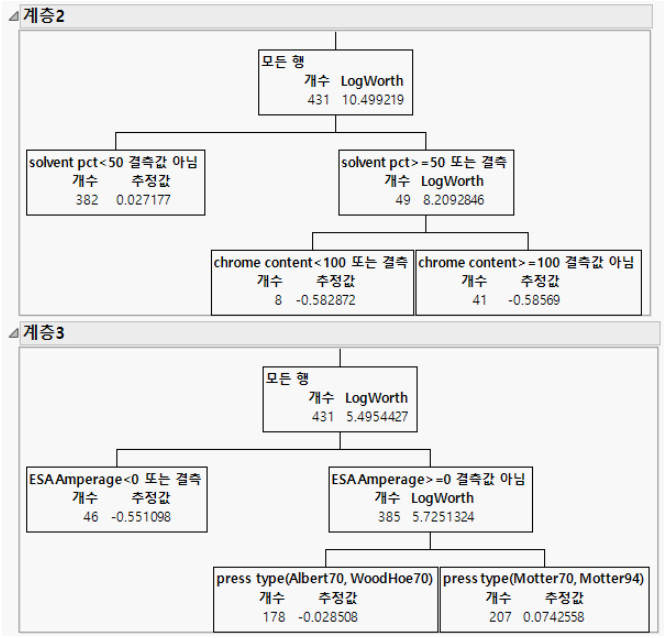
JMP PRO 부스티드 트리

각각 이전 계층에 기반한 여러 트리 계층 적합

부스티드 트리 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

부스트는 **계층**이라고 하는 일련의 작은 결정 트리를 적합시킨 후 이를 누적하여 더 큰 결정 트리를 생성하는 프로세스입니다. 각 계층의 트리는 적은 수의 분할로 구성됩니다. 트리는 이전 계층의 잔차를 기반으로 적합되므로 각 계층에서 이전 계층의 잘못된 적합 데이터에 대한 적합을 수정할 수 있습니다. 관측값에 대한 최종 예측은 모든 계층에서 해당 관측값의 예측을 합한 것입니다.

그림 6.1 부스티드 트리 계층의 예



목차

부스티드 트리 플랫폼 개요 101
범주형 반응이 있는 부스티드 트리의 예 101
연속형 반응이 있는 부스티드 트리의 예 103
부스티드 트리 플랫폼 시작 105
    규격 창 106
부스티드 트리 보고서 109
    모형 검증 집합 요약 110
    규격 111
    전체 통계량 111
    누적 검증 112
부스티드 트리 플랫폼 옵션 112
부스티드 트리 플랫폼에 대한 통계 상세 정보 114
    과대적합 벌점 114

JMP PRO 부스티드 트리 플랫폼 개요

부스티드 트리 플랫폼에서는 **계층**에 생성된 여러 개의 더 작은 결정 트리를 기반으로 하는 가법 결정 트리 모형을 생성합니다. 각 계층의 트리는 소수의 분할로 구성됩니다(보통 5개 이하). 각 계층은 "파티션 모형"에 설명된 재귀적 적합 방법론을 사용하여 적합됩니다. 차이는 지정된 수의 분할이 수행되면 적합이 중지된다는 것뿐입니다. 트리가 주어지면 잎에 있는 관측값의 예측값은 해당 잎에 있는 모든 관측값의 평균입니다.

적합 프로세스는 다음과 같습니다.

1. 초기 계층을 적합시킵니다.
2. 잔차를 계산합니다. 이 값은 잎 내의 관측값에 대한 실제 값에서 예측 평균을 뺀 값입니다.
3. 계층을 잔차에 적합시킵니다.
4. 가법 트리를 생성합니다. 주어진 관측값에 대해 전체 계층에서 예측값의 합계를 계산합니다.
5. 지정된 계층 수에 도달할 때까지 또는 검증을 사용하는 경우 추가 계층을 적합시켜도 검증 통계량이 더 이상 개선되지 않을 때까지 2 단계 ~ 4 단계를 반복합니다.

최종 예측은 모든 계층에서 관측값에 대한 예측을 합한 것입니다.

이전 계층의 잔차에 대해 연속적으로 계층을 적합시키면 각 계층에서 적합성을 향상시킬 수 있습니다.

범주형 반응의 경우 반응 수준이 두 개인 반응만 지원됩니다. 범주형 반응을 사용할 때 각 계층에서 적합된 잔차는 선형 로짓의 오프셋입니다. 최종 예측은 모든 계층에 대한 선형 로짓 합계의 로지스틱 변환입니다.

부스티드 트리에 대한 자세한 내용은 Hastie et al. (2009) 연구 자료에서 확인하십시오.

JMP PRO 범주형 반응이 있는 부스티드 트리의 예

이 예에서는 **밴딩** 결함의 영향을 받는 인쇄 작업을 예측하기 위해 부스티드 트리 모형을 생성합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Bands Data.jmp 를 엽니다.
2. **분석 > 예측 모델링 > 부스티드 트리**를 선택합니다.
3. Banding? 을 선택하고 **Y, 반응**을 클릭합니다.
4. Predictors 열 그룹을 선택하고 **X, 요인**을 클릭합니다.
5. **검증 데이터 비율**에 "0.2" 를 입력합니다.
6. **확인**을 클릭합니다.

   부스티드 트리 규격 창이 나타납니다.

7. (선택 사항) " 재현성 " 패널에서 **멀티스레딩 제한**을 선택하고 " 난수 시드값 " 에 "123" 을 입력합니다 .

   부스티드 트리 적합에는 랜덤 성분이 포함되므로 이렇게 하면 아래 표시된 것과 정확히 동일한 결과를 얻을 수 있습니다 .

8. **확인**을 클릭합니다 .

그림 6.2 명목형 반응에 대한 전체 통계량

전체 통계량

측도	훈련	검증	정의
엔트로피 R^2	0.7103	0.3299	1-Loglike(model)/Loglike(0)
일반화 R^2	0.8340	0.4803	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
평균 -Log p	0.1985	0.4383	∑ -Log(ρ[j])/n
RASE	0.2196	0.3801	√ ∑(y[j]-ρ[j])²/n
평균 절대 편차	0.1650	0.2828	∑ |y[j]-ρ[j]|/n
오분류 비율	0.0232	0.1852	∑ (ρ[j]≠ρMax)/n
N	431	108	n

혼동 행렬

훈련

실제값 Banding?	예측 개수 band	noband
band	179	9
noband	1	242

실제값 Banding?	예측 비율 band	noband
band	0.952	0.048
noband	0.004	0.996

검증

실제값 Banding?	예측 개수 band	noband
band	22	17
noband	3	66

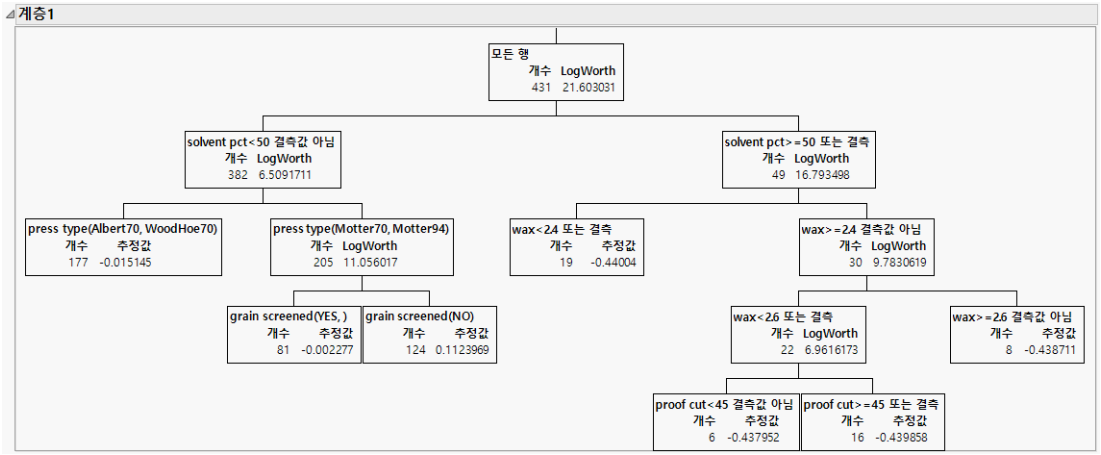
실제값 Banding?	예측 비율 band	noband
band	0.564	0.436
noband	0.043	0.957

반응 Banding? 이 범주형이므로 부스티드 트리 분석을 수행한 결과 " 측도 " 아래의 " 오분류 비율 " 과 " 혼동 행렬 " 보고서가 제공됩니다 . 검증 데이터 집합의 오분류 비율은 0.1852(약 19%) 입니다 .

9. "Banding? 에 대한 부스티드 트리 " 옆의 빨간색 삼각형을 클릭하고 **트리 표시 > 이름 범주 추정값 표시**를 선택합니다 .

   계층의 개요가 포함된 " 트리 보기 " 보고서가 나타납니다 . 계층을 검토하여 적합된 트리와 예측값을 확인할 수 있습니다 .

그림 6.3 부스티드 트리의 계층 1



10. "Banding? 에 대한 부스티드 트리 " 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다 .

    확률(Banding?==noband), 확률(Banding?==band) 및 최대 확률 분류 Banding? 열이 데이터 테이블에 추가됩니다 . **확률** (Banding?==noband) 열을 검토하여 계층에서 모형 예측이 계산되는 방식을 확인할 수 있습니다 .

JMP PRO 연속형 반응이 있는 부스티드 트리의 예

이 예에서는 명목형 요인과 연속형 요인의 조합이 주어졌을 때 체지방률을 예측하기 위해 부스티드 트리 모형을 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Body Fat.jmp 샘플 데이터 테이블을 엽니다 .
2. **분석 > 예측 모델링 > 부스티드 트리**를 선택합니다 .
3. Percent body fat 을 선택하고 **Y, 반응**을 클릭합니다 .
4. Age (years) ~ Wrist circumference (cm) 를 선택하고 **X, 요인**을 클릭합니다 .
5. Validation 을 선택하고 **검증**을 클릭합니다 .
6. **확인**을 클릭합니다 .
7. **확인**을 클릭합니다 .

그림 6.4 연속형 반응에 대한 전체 통계량

전체 통계량

	R^2	RASE	N
훈련	0.847	3.2116281	180
검증	0.611	5.4301412	72

" 전체 통계량 " 보고서에는 부스티드 트리 모형에 대한 R^2 및 RASE 가 제공됩니다 . 검증 데이터 집합의 R^2 은 0.611 입니다 . 검증 데이터 집합의 RASE 는 약 5.43 입니다 .

모형과 별개로 Percent body fat 에 대해 중요한 예측 변수를 나타내려고 합니다 .

8. "Percent body fat 에 대한 부스티드 트리 " 옆의 빨간색 삼각형을 클릭하고 **프로파일러**를 선택합니다 .
9. " 예측 프로파일러 " 옆의 빨간색 삼각형을 클릭하고 **변수 중요도 평가 > 독립 균등 입력**을 선택합니다 .

참고 : 변수 중요도 평가에는 랜덤화가 사용되므로 그림 6.5 의 결과와 정확히 일치하지 않을 수 있습니다 .

그림 6.5 변수 중요도의 요약 보고서

요약 보고서

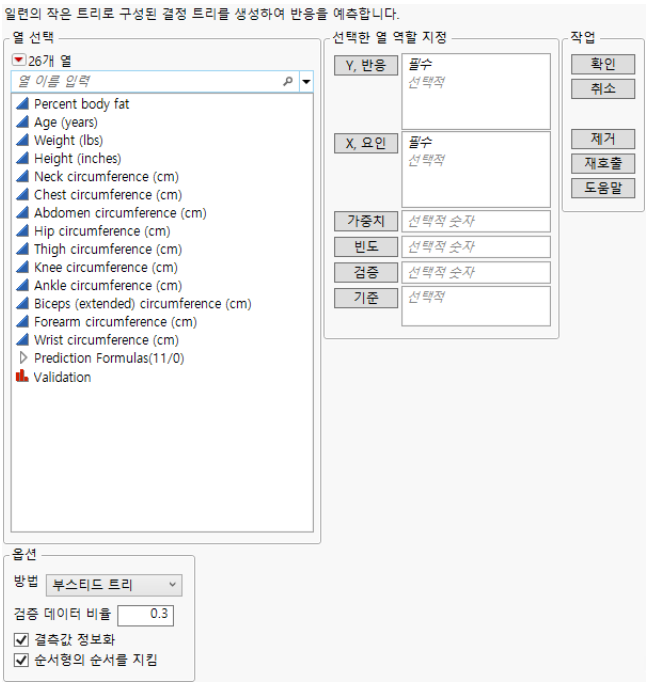
열	주효과	총 효과
Abdomen circumference (cm)	0.865	0.881
Age (years)	0.06	0.076
Wrist circumference (cm)	0.007	0.012
Biceps (extended) circumference (cm)	0.005	0.008
Hip circumference (cm)	0.004	0.007
Chest circumference (cm)	0.003	0.004
Ankle circumference (cm)	0.002	0.003
Height (inches)	0.002	0.003
Weight (lbs)	0.001	0.002
Thigh circumference (cm)	0.001	0.002
Knee circumference (cm)	0.001	0.002
Neck circumference (cm)	1e-17	2e-17
Forearm circumference (cm)	1e-17	2e-17

" 요약 보고서 " 에서는 Abdomen circumference (cm) 가 Percent body fat 의 가장 중요한 예측 변수임을 보여 줍니다 .

JMP PRO 부스티드 트리 플랫폼 시작

분석 > 예측 모델링 > 부스티드 트리를 선택하여 부스티드 트리 플랫폼을 시작합니다.

그림 6.6 Body Fat.jmp 를 사용한 부스티드 트리 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

부스티드 트리 플랫폼 시작 창에는 다음과 같은 옵션이 있습니다.

Y, 반응 분석할 하나 이상의 반응 변수입니다.

X, 요인 예측 변수입니다.

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다.

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다.

검증 검증 데이터 집합을 정의하는 숫자 열입니다. 이 열에는 최대 세 개의 구분되는 값이 포함되어야 합니다.

- 검증 열에 두 개의 수준이 있는 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다.
- 검증 열에 세 개의 수준이 있는 경우 작은 값부터 순서대로 훈련 데이터 집합, 검증 데이터 집합, 테스트 데이터 집합을 각각 정의합니다.

– 검증 열에 네 개 이상의 수준이 있는 경우 가장 작은 세 개의 값을 포함하는 행이 검증 데이터 집합을 정의합니다. 다른 모든 행은 분석에서 제외됩니다.

부스티드 트리 플랫폼에서는 검증 열을 사용하여 모형을 훈련 및 조정하거나 모형을 훈련, 조정 및 평가합니다. 검증에 대한 자세한 내용은 "JMP 모델링의 검증" 에서 확인하십시오.

"열 선택" 목록에서 아무 열도 선택하지 않은 상태로 "검증" 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다. "검증 열 생성" 유틸리티에 대한 자세한 내용은 "검증 열 생성" 에서 확인하십시오.

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다. 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다. 결과는 개별 보고서에 제공됩니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

방법 분할 방법 (결정 트리, 붓스트랩 포레스트, 부스티드 트리, K 최근접 이웃 또는 Naive Bayes) 을 선택할 수 있습니다. 결정 트리를 제외한 이 대체 방법은 JMP Pro 에서 사용할 수 있습니다.

이러한 방법에 대한 자세한 내용은 "파티션 모형", "붓스트랩 포레스트", "K 최근접 이웃" 및 "Naive Bayes" 에서 확인하십시오.

검증 데이터 비율 검증 데이터 집합으로 사용할 데이터 비율입니다.

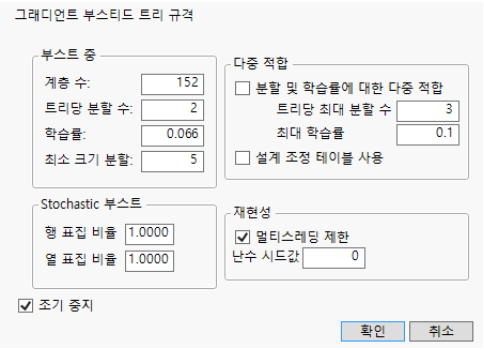
결측값 정보화 이 옵션을 선택하면 범주형 예측 변수에 대한 결측값 범주화 및 연속형 예측 변수에 대한 결측값 정보화 처리가 활성화됩니다. 자세한 내용은 "결측값 정보화" 에서 확인하십시오.

순서형의 순서를 지킴 이 옵션을 선택하면 순서를 유지하는 분할만 고려하도록 제한됩니다.

JMP PRO 규격 창

시작 창에서 **확인**을 선택하면 "그래디언트 부스티드 트리 규격" 창이 나타납니다.

그림 6.7 부스티드 트리 규격 창



JMP PRO 부스트 패널

계층 수 최종 트리에 포함할 최대 계층 수입니다.

트리당 분할 수 각 계층의 분할 수입니다.

학습률 $0 < r \leq 1$ 의 숫자입니다. 학습률이 1 에 가까우면 최종 트리에 대한 수렴이 빨라지지만 데이터가 과대적합되는 경향도 높습니다. 적은 수의 계층이 지정된 경우 1 에 가까운 학습률을 사용하십시오. 학습률은 모형의 수렴을 늦추는 작은 소수이며 보통 0.01 에서 0.1 사이입니다. 이렇게 하면 이후 계층이 이전 계층과 다른 분할을 사용할 수 있도록 합니다.

과대적합 벌점 (범주형 반응에만 사용 가능) 적합 확률이 0 이 되지 않도록 하는 데 도움이 되는 편향 모수입니다. 자세한 내용은 " 과대적합 벌점 " 에서 확인하십시오.

최소 크기 분할 후보 분할에 필요한 최소 관측값 수입니다.

JMP PRO 다중 적합 패널

분할 및 학습률에 대한 다중 적합 이 옵션을 선택하면 트리당 분할 수 (정수 증분) 와 학습률 (0.1 증분) 의 모든 조합에 대한 부스티드 트리가 생성됩니다.

조합의 하한은 " 트리당 분할 수 " 및 " 학습률 " 옵션으로 지정됩니다. 조합의 상한은 다음 옵션으로 지정됩니다.

트리당 최대 분할 수 트리당 분할 수의 상한입니다.

최대 학습률 학습률의 하한입니다.

설계 조정 테이블 사용 일부 조정 모수의 값이 포함된 데이터 테이블 (**설계 조정 테이블**) 을 선택할 수 있는 창을 엽니다. 설계 조정 테이블에는 지정할 각 옵션에 대한 열과 각각 단일 부스티드 트리 모형 설계를 나타내는 하나 이상의 행이 있습니다. 설계 조정 테이블에 옵션이 지정되지 않으면 기본값이 사용됩니다.

JMP 에서는 지정된 조정 모수를 사용하여 테이블의 각 행에 대해 부스티드 트리 모형을 생성합니다. 설계 조정 테이블에 둘 이상의 모형이 지정된 경우 " 모형 검증 집합 요약 " 보고서에 각 모형의 R^2 값이 나열됩니다. " 부스티드 트리 " 보고서에는 R^2 값이 가장 큰 모형의 적합 통계량이 표시됩니다.

실험 설계 기능을 사용하여 설계 조정 테이블을 생성할 수 있습니다. 부스티드 트리 설계 조정 테이블에는 다음 열이 임의의 순서로 포함될 수 있습니다.

- 계층 수
- 트리당 분할 수
- 학습률
- 최소 크기 분할
- 행 표집 비율
- 열 표집 비율

JMP PRO Stochastic 부스트 패널

행 표집 비율 각 계층에 대해 표집할 훈련 행의 비율입니다.

참고 : 반응이 범주형이면 층별 랜덤 표집을 사용하여 훈련 행이 표집됩니다.

열 표집 비율 각 계층에 대해 표집할 예측 변수 열의 비율입니다.

JMP PRO 재현성 패널

멀티스레딩 제한 이 옵션을 선택하면 모든 계산이 단일 스레드에서 수행됩니다.

난수 시드값 이후 플랫폼 실행을 위해 결과를 재현하는 데 사용할 0 이 아닌 숫자 난수 시드값을 지정합니다. 기본적으로 난수 시드값은 0 으로 설정되어 재현 가능한 결과를 생성하지 않습니다. 분석을 스크립트에 저장하면 사용자가 입력한 난수 시드값이 스크립트에 저장됩니다.

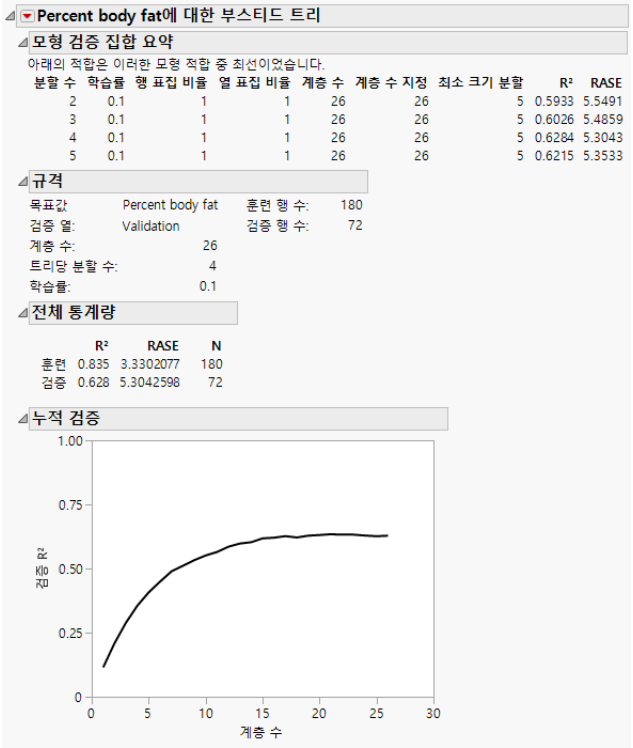
JMP PRO 조기 중지

조기 중지 이 옵션을 선택하면 계층을 추가해도 검증 통계량이 개선되지 않을 경우 부스트 프로세스에서 추가 계층 적합을 중지합니다. 이 옵션을 선택하지 않으면 지정된 계층 수에 도달할 때까지 부스트 프로세스가 계속됩니다. 이 옵션은 검증을 사용하는 경우에만 나타납니다.

JMP PRO 부스티드 트리 보고서

"그래디언트 부스티드 트리 규격" 창에서 "확인"을 클릭하면 부스티드 트리 보고서가 열립니다.

그림 6.8 연속형 반응에 대한 부스티드 트리 보고서



Percent body fat에 대한 부스티드 트리

모형 검증 집합 요약

아래의 적합은 이러한 모형 적합 중 최선이었습니다.

분할 수	학습률	행 표집 비율	열 표집 비율	계층 수	계층 수 지정	최소 크기 분할	R^2	RASE
2	0.1	1	1	26	26	5	0.5933	5.5491
3	0.1	1	1	26	26	5	0.6026	5.4859
4	0.1	1	1	26	26	5	0.6284	5.3043
5	0.1	1	1	26	26	5	0.6215	5.3533

규격

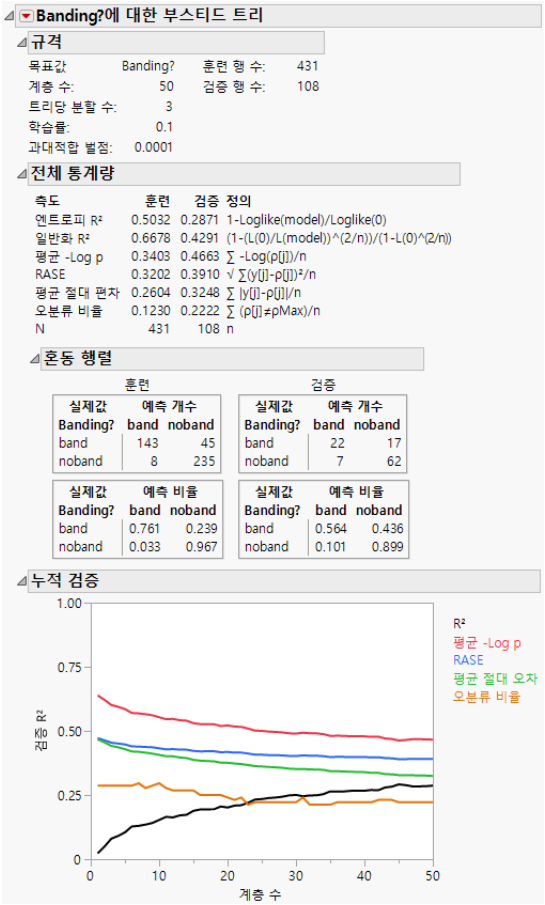
목표값	Percent body fat	훈련 행 수:	180
검증 열:	Validation	검증 행 수:	72
계층 수:	26		
트리당 분할 수:	4		
학습률:	0.1		

전체 통계량

	R^2	RASE	N
훈련	0.835	3.3302077	180
검증	0.628	5.3042598	72

누적 검증

그림 6.9 범주형 반응에 대한 부스티드 트리 보고서



반응이 범주형인지 아니면 연속형인지에 따라 다음과 같은 보고서가 제공됩니다.

- "모형 검증 집합 요약"
- "규격"
- "전체 통계량"
- "누적 검증"

JMP PRO 모형 검증 집합 요약

규격 창에서 "분할 및 학습률에 대한 다중 적합" 옵션을 선택한 경우 모든 모형 적합에 대한 적합 통계량을 표시합니다. 자세한 내용은 그림 6.8 및 "다중 적합 패널"에서 확인하십시오.

JMP PRO 규격

모형 적합에 사용된 설정을 표시합니다.

JMP PRO 전체 통계량

훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 적합 통계량을 표시합니다.

부스티드 트리 규격 창의 "분할 및 학습률에 대한 다중 적합" 옵션을 사용하여 여러 모형을 적합시키는 경우를 가정해 보겠습니다. 그러면 "전체 통계량" 및 "누적 검증" 보고서에 결과가 표시되는 모형은 검증 데이터 집합의 엔트로피 R^2 값(범주형 반응의 경우) 또는 R^2 값(연속형 반응의 경우)이 가장 큰 모형입니다.

JMP PRO 측도 보고서

(범주형 반응에만 사용 가능) 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 다음 통계량을 제공합니다.

참고 : 엔트로피 R^2 및 일반화 R^2 의 경우 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 . 평균 -Log p, RASE, 평균 절대 편차 및 오분류 비율의 경우 값이 작을수록 더 나은 적합을 나타냅니다.

엔트로피 R^2 적합 모형과 상수 확률 모형의 로그 가능도를 비교하는 적합 측도입니다 . 값 범위는 0 에서 1 사이입니다 . 자세한 내용은 " 엔트로피 R^2" 에서 확인하십시오 .

일반화 R^2 일반 회귀 모형에 적용할 수 있는 측도입니다 . 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다 . 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다 . 일반화 R^2 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2 으로 단순화됩니다 . 일반화 R^2 을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데 , 이는 Cox-Snell 유사 R^2 을 정규화한 것입니다 .

평균 -Log P 음의 $\log(p)$ 의 평균입니다 . 여기서 p 는 발생한 사건과 연관된 적합 확률입니다 .

RASE 제곱근 평균 제곱 예측 오차입니다 . 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 .

평균 절대 편차 반응과 예측 반응 사이의 차이에 대한 절대값의 평균입니다 . 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 .

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다 .

N 관측값 수입니다 .

JMP PRO 혼동 행렬 보고서

(범주형 반응에만 사용 가능) 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 분류 통계량을 표시합니다. "혼동 행렬" 보고서에는 혼동 행렬과 혼동 비율 행렬이 포함됩

니다. 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다. 혼동 비율 행렬은 값을 행 합계로 나눈 혼동 행렬과 동일합니다.

JMP PRO 결정 행렬

(범주형 반응을 사용하고 반응에 "수익 행렬" 열 특성이 있거나 "수익 행렬 지정" 옵션을 사용하여 비용을 지정하는 경우에만 사용 가능) 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 결정 개수 및 결정 비율 행렬을 제공합니다. 자세한 내용은 "분할의 추가 예"에서 확인하십시오.

JMP PRO 누적 검증

(검증을 사용하는 경우에만 사용 가능) 검증 데이터 집합 대 계층 수에 대한 적합 통계량 그림을 표시합니다.

연속형 반응의 단일 적합 통계량은 R^2 입니다. 범주형 반응의 적합 통계량은 아래에 나열되어 있으며 " 측도 보고서 " 에서 설명합니다.

- R^2(엔트로피 R^2)
- 평균 -Log p(평균 -Log p)
- 제곱근 평균 제곱 예측 오차 (RASE)
- 평균 절대 오차 (평균 절대 편차)
- MR(오분류 비율)

"누적 검증" 그림 아래의 "누적 상세 정보" 보고서에는 그림에 사용된 값이 제공됩니다.

JMP PRO 부스티드 트리 플랫폼 옵션

" 부스티드 트리 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

트리 표시 " 트리 보기 " 보고서에 트리를 표시하기 위한 옵션을 제공합니다. 이 보고서에는 부스트 프로세스의 각 계층에서 적합된 트리 그림이 제공됩니다.

실제값 대 예측값 그림 (연속형 반응에만 사용 가능) 실제값 대 예측값 그림을 제공합니다.

열 기여도 적합에 대한 각 입력 열의 기여도를 보여 주는 보고서를 표시합니다. 다음 항목도 보고서에 표시됩니다.

- 지정된 열을 사용하여 데이터를 분할할 때 모든 트리에 대한 총 인스턴스 수
- 열에 기인한 총 G^2(범주형 반응의 경우) 또는 SS, 즉 제곱합 (연속형 반응의 경우)
- G^2 또는 SS 의 막대 차트
- 열에 기인한 G^2 또는 SS 의 비율

ROC 곡선 (범주형 반응에만 사용 가능) 자세한 내용은 "ROC 곡선 " 에서 확인하십시오.

향상도 곡선 (범주형 반응에만 사용 가능) 자세한 내용은 " 향상도 곡선 " 에서 확인하십시오 .

열 저장 모형 및 트리 결과를 저장하고 SAS 코드를 생성하기 위한 옵션이 포함되어 있습니다 .

예측값 저장 모형의 예측값을 데이터 테이블에 저장합니다 .

예측 계산식 저장 예측 계산식을 데이터 테이블의 열에 저장합니다 . 계산식은 트리 구조를 설명하는 내포된 조건부 절로 구성됩니다 . 반응이 연속형이면 열에 " 예측 " 열 특성이 포함됩니다 . 반응이 범주형이면 열에 " 반응 확률 " 열 특성이 포함됩니다 .

허용 예측 계산식 저장 (이 옵션 대신 " 예측 계산식 저장 " 옵션을 사용해야 합니다 . " 예측 계산식 저장 " 을 사용할 수 없을 때만 이 옵션을 사용하십시오 .) 결측값이 있고 " 결측값 정보화 " 를 선택하지 않은 경우에도 예측하는 계산식을 저장합니다 . 예측 계산식은 결측 예측 변수에 대한 반응 값을 분할에 랜덤 할당하여 결측값을 허용합니다 . 반응이 연속형이면 열에 " 예측 " 열 특성이 포함됩니다 . 반응이 범주형이면 열에 " 반응 확률 " 열 특성이 포함됩니다 . " 결측값 정보화 " 를 선택한 경우 Shift 키를 누른 채 보고서의 빨간색 삼각형을 클릭하면 허용 예측 계산식을 저장할 수 있습니다 .

잔차 저장 (연속형 반응에만 사용 가능) 잔차를 데이터 테이블에 저장합니다 .

오프셋 추정값 저장 (범주형 반응에만 사용 가능) 선형 성분의 합을 저장합니다 . 이는 적합 확률의 로짓입니다 .

트리 상세 정보 저장 각 계층에 대한 분할 상세 정보 및 추정값이 포함된 데이터 테이블을 생성합니다 .

누적 상세 정보 저장 (검증을 사용하는 경우에만 사용 가능) 각 계층에 대한 적합 통계량이 포함된 데이터 테이블을 생성합니다 .

예측 계산식 게시 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

허용 예측 계산식 게시 (이 옵션 대신 " 예측 계산식 게시 " 옵션을 사용해야 합니다 . " 예측 계산식 게시 " 를 사용할 수 없을 때만 이 옵션을 사용하십시오 .) 허용 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 . " 결측값 정보화 " 를 선택한 경우 Shift 키를 누른 채 보고서의 빨간색 삼각형을 클릭하면 이 옵션을 사용할 수 있습니다 .

SAS DATA 스텝 만들기 새 데이터 집합을 스코어링하기 위한 SAS 코드를 생성합니다 .

수익 행렬 지정 (범주형 반응에만 사용 가능) 올바르거나 올바르지 않은 분류 결정과 관련된 수익 또는 비용을 지정할 수 있습니다 . 자세한 내용은 " 적합 상세 정보 표시 " 에서 확인하십시오 .

프로파일러 예측 프로파일러를 표시합니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

JMP PRO 부스티드 트리 플랫폼에 대한 통계 상세 정보

이 섹션에서는 부스티드 트리 플랫폼과 관련된 상세 정보를 설명합니다. 재귀 결정 트리에 대한 통계 상세 정보는 "파티션 플랫폼에 대한 통계 상세 정보"에서 확인하십시오.

JMP PRO 과대적합 벌점

반응이 범주형이면 모수 벌점이 부과됩니다. 각 계층에서 추정값은 음의 로그 가능도 + 벌점 값과 각 관측값에 대한 추정값의 제곱합을 곱한 값을 최소화합니다. 이 벌점은 각 새 계층이 훈련 데이터를 과대적합하지 않도록 합니다.

7 장

JMP PRO K 최근접 이웃

인접한 관측값을 사용하여 반응 값 예측

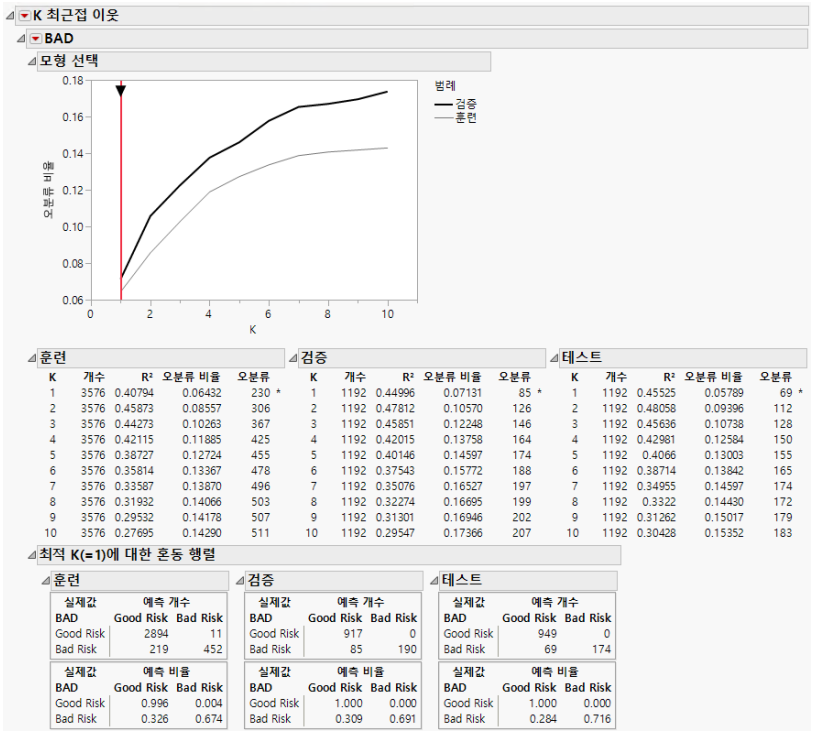
K 최근접 이웃 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

K 최근접 이웃 플랫폼에서는 지정된 관측값의 국소적 근방에 있는 관측값의 반응을 사용하여 해당 관측값에 대한 반응 값을 예측합니다. 이 플랫폼은 범주형 반응을 분류하고 연속형 반응을 예측하는 데 사용할 수 있습니다.

K 최근접 이웃은 인접한 관측값까지의 거리를 기반으로 하는 비모수 방법입니다. 이 때문에 K 최근접 이웃 방법은 불규칙한 예측 변수 값 경계를 사용하여 관측값을 분류할 수 있습니다. 그러나 이 알고리즘은 관련 없는 예측 변수의 영향을 받으므로 예측 변수를 선택하면 결과에 영향을 줄 수 있습니다.

K 최근접 이웃은 위성 이미지와 EKG(심전도) 패턴 등 다양한 응용 분야에서 사용되고 있습니다.

그림 7.1 K 최근접 이웃 플랫폼의 예



목차

K 최근접 이웃 플랫폼 개요 117
범주형 반응이 있는 K 최근접 이웃의 예 117
연속형 반응이 있는 K 최근접 이웃의 예 119
K 최근접 이웃 플랫폼 시작 121
K 최근접 이웃 보고서 122
    연속형 반응 122
    범주형 반응 123
K 최근접 이웃 플랫폼 옵션 124
    모형 적합 옵션 124

JMP PRO K 최근접 이웃 플랫폼 개요

K 최근접 이웃 플랫폼은 *k*개의 최근접 이웃의 반응을 기반으로 반응 값을 예측합니다. 주어진 관측값에 대한 *k*개의 최근접 이웃은 해당 관측값에 대한 예측 변수 값과 다른 각 관측값에 대한 예측 변수 값 사이의 최소 유클리드 거리 *k*개를 식별하여 결정됩니다. K 최근접 이웃 플랫폼에서는 연속형 반응과 범주형 반응을 모두 모델링합니다.

k 최근접 이웃 방법은 대규모 문제의 경우 예측 계산식이 복잡하고 해석하기 어려워 유용성이 제한될 수 있다는 단점을 가지고 있습니다. 또한 K 최근접 이웃 방법은 범주형 반응에 대한 확률을 계산하지 않습니다. *k* 최근접 이웃 방법에 대한 자세한 내용은 Hastie et al. (2009), Hand et al. (2001) 및 Shmueli et al. (2017) 연구 자료에서 확인하십시오.

연속형 반응

연속형 반응의 경우 예측값은 *k*개의 최근접 이웃에 대한 반응의 평균입니다. 각 연속형 예측 변수는 표준편차로 척도화됩니다. 이 척도를 사용하면 범위가 큰 단일 예측 변수가 거리 계산에 과도하게 영향을 미치지 않습니다. 연속형 예측 변수의 결측값은 해당 예측 변수의 평균으로 대체됩니다. 자세한 내용은 "연속형 반응이 있는 K 최근접 이웃의 예"에서 확인하십시오.

범주형 반응

범주형 반응의 경우 예측값은 *k*개의 최근접 이웃에 대해 가장 빈도가 높은 반응 수준입니다. 둘 이상의 수준이 가장 빈도가 높은 수준으로 묶여 있으면 이러한 수준 중 하나를 랜덤 선택하여 예측 반응이 할당됩니다.

참고: 범주형 반응의 경우 가장 빈도가 높은 수준에 대한 동점 연결이 무작위로 끊어지기 때문에 각 플랫폼 실행 결과가 다를 수 있습니다. 재현 가능한 결과를 얻으려면 시작 창에서 " 난수 시드 값 설정 " 옵션을 사용하거나 JSL 스크립트에 `Set Random Seed()` 함수를 포함하십시오.

범주형 예측 모형에서 각 범주형 예측 변수는 표시 변수로 표현되며, 이때 하나의 표시 변수가 각 수준을 나타냅니다. 범주형 예측 변수의 결측값이 있는 행은 해당 예측 변수에 대한 모든 표시 변수에서 값이 0으로 표시됩니다.

JMP PRO 범주형 반응이 있는 K 최근접 이웃의 예

주택 담보 대출을 신청한 5,960 명의 고객에 대한 과거 재무 데이터가 있습니다. 각 고객은 Good Risk(안전한 위험) 또는 Bad Risk(불안한 위험) 로 분류되었습니다. 많은 예측 변수에 대해 결측값이 있습니다. 미래 고객의 신용 위험을 분류하는 데 사용할 모형을 생성하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Equity.jmp 를 엽니다.
2. **분석 > 예측 모델링 > K 최근접 이웃**을 선택합니다.
3. BAD 를 선택하고 **Y, 반응**을 클릭합니다.

4. LOAN ~ CLNO 를 선택하고 **X, 요인**을 클릭합니다 .

   잠재적 예측 변수 중 하나인 DEBTINC 에는 결측값이 많이 있으므로 모형에 포함하지 않습니다 . 연속형 예측 변수의 결측값은 해당 예측 변수의 평균으로 대체됩니다 . 이 절차는 주로 결측값이 랜덤인 경우에 잘 맞습니다 . DEBINC 의 높은 결측 비율은 결측값이 정보를 제공할 수 있음을 나타내지만 이 예에서는 조사하지 않습니다 .

5. Validation 을 선택하고 **검증**을 클릭합니다 .

6. **확인**을 클릭합니다 .

그림 7.2 K 최근접 이웃 보고서

JMP 에서는 각 *K* 값에 대해 훈련 데이터 집합 관측값만 사용하여 모형을 생성합니다 . 이러한 각 모형은 검증 데이터 집합 관측값을 분류하는 데 사용됩니다 . 검증 데이터 집합 결과는 최적 모형을 선택하는 데 사용됩니다 . 이 예에서는 단일 최근접 이웃 ($K = 1$) 에 기반한 모형의 오분류 비율이 가장 작습니다 . 테스트 데이터 집합은 단일 최근접 이웃 모형이 독립 데이터에 대해 최상의 성능을 발휘하는지 확인합니다 .

7. "BAD" 의 빨간색 삼각형을 클릭하고 **예측 계산식 게시**를 선택합니다 .

8. **이웃 수 , K** 옆의 기본값 1 을 그대로 사용합니다 .
9. **확인**을 클릭합니다 .

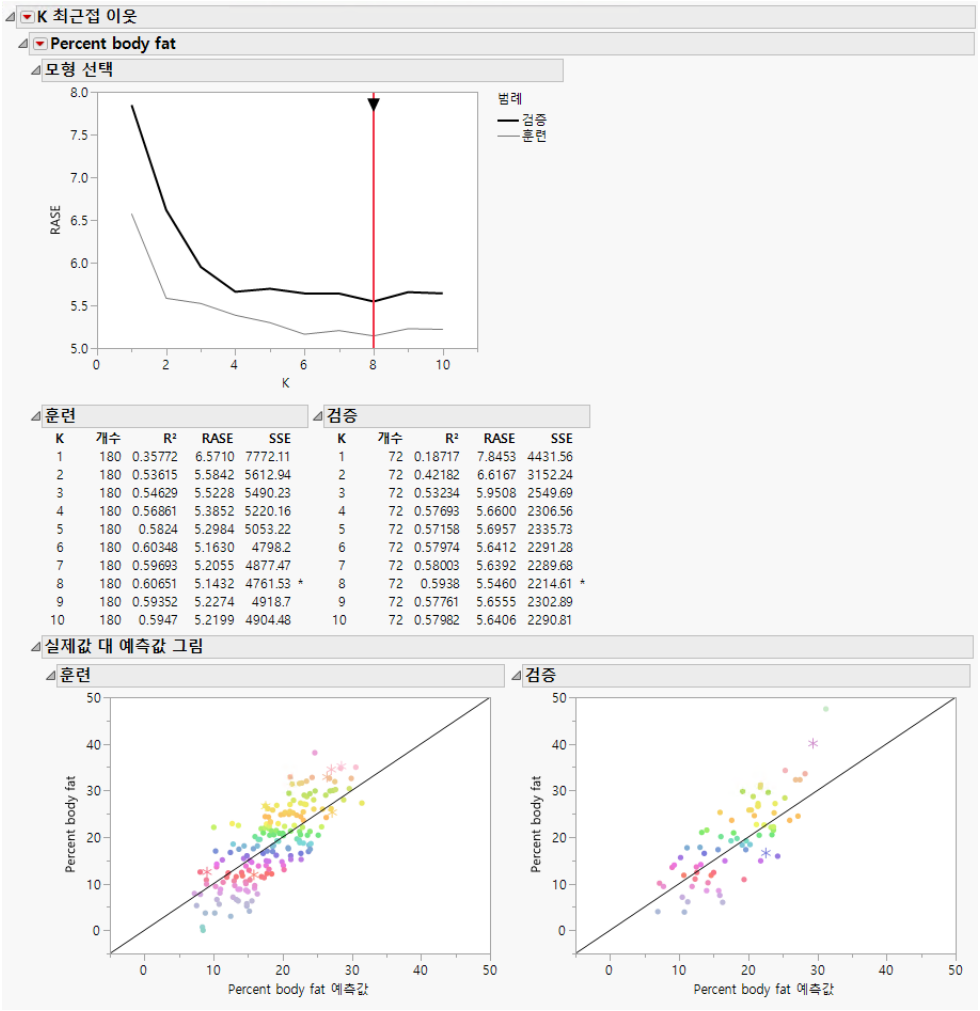
   예측 방정식이 계산식 저장소에 저장됩니다 . " 계산식 저장소 " 의 " 모형 비교 " 옵션을 사용하여 $K = 1$ 최근접 이웃 모형의 성능과 계산식 저장소에 게시된 대체 모형의 성능을 비교할 수 있습니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

JMP PRO 연속형 반응이 있는 K 최근접 이웃의 예

이 예에서는 13 개의 예측 변수를 사용하여 남성의 체지방률을 예측하려고 합니다 . Body Fat.jmp 샘플 데이터 테이블에는 수중체중법 및 다양한 신체 둘레 측정을 기반으로 하는 체지방률 추정값이 포함되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Body Fat.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > K 최근접 이웃**을 선택합니다 .
3. Percent body fat 을 선택하고 **Y, 반응**을 클릭합니다 .
4. Age (years) ~ Wrist circumference (cm) 를 선택하고 **X, 요인**을 클릭합니다 .
5. Validation 을 선택하고 **검증**을 클릭합니다 .
6. **확인**을 클릭합니다 .
7. "Percent body fat" 의 빨간색 삼각형을 클릭하고 **실제값 대 예측값 그림**을 선택합니다 .

그림 7.3 K 최근접 이웃 보고서



K	개수	R^2	RASE	SSE
1	180	0.35772	6.5710	7772.11
2	180	0.53615	5.5842	5612.94
3	180	0.54629	5.5228	5490.23
4	180	0.56861	5.3852	5220.16
5	180	0.5824	5.2984	5053.22
6	180	0.60348	5.1630	4798.2
7	180	0.59693	5.2055	4877.47
8	180	0.60651	5.1432	4761.53 *
9	180	0.59352	5.2274	4918.7
10	180	0.5947	5.2199	4904.48

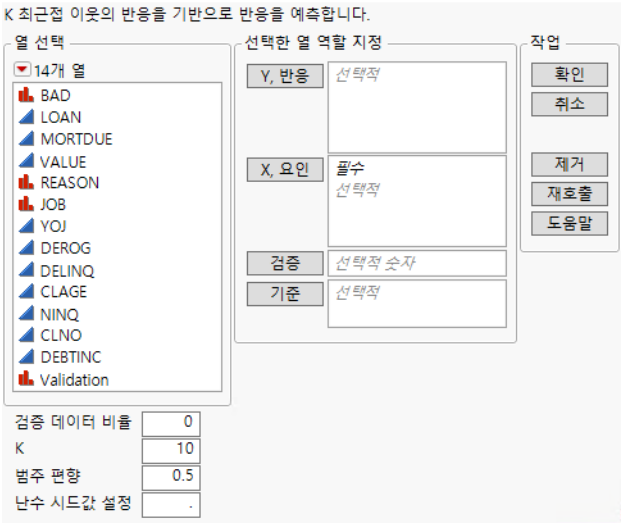
K	개수	R^2	RASE	SSE
1	72	0.18717	7.8453	4431.56
2	72	0.42182	6.6167	3152.24
3	72	0.53234	5.9508	2549.69
4	72	0.57693	5.6600	2306.56
5	72	0.57158	5.6957	2335.73
6	72	0.57974	5.6412	2291.28
7	72	0.58003	5.6392	2289.68
8	72	0.5938	5.5460	2214.61 *
9	72	0.57761	5.6555	2302.89
10	72	0.57982	5.6406	2290.81

$K = 8$ 모형이 검증 데이터 집합에 대한 RASE 가 가장 낮았습니다 . k 최근접 이웃 모형 중에서 8 개의 최근접 이웃에 기반한 모형의 성능이 가장 좋은 것으로 보입니다 . 훈련 데이터 집합의 실제값 대 예측값 그림을 보면 점이 선을 따라 분포되어 있으므로 예측값이 실제값과 유사하다는 것을 알 수 있습니다 . 검증 데이터 집합에 대한 그림에서는 대부분의 점이 선을 따라 분포되며 약간 멀리 떨어진 상단에 몇몇 관측값이 있습니다 .

JMP PRO K 최근접 이웃 플랫폼 시작

분석 > 예측 모델링 > K 최근접 이웃을 선택하여 K 최근접 이웃 플랫폼을 시작합니다 .

그림 7.4 K 최근접 이웃 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

K 최근접 이웃 시작 창에는 다음과 같은 옵션이 있습니다 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 .

> **참고 :** K 최근접 이웃 플랫폼은 반응 변수가 없는 경우에도 인접 관측값 사이의 거리를 결정하는 유틸리티로 사용할 수 있습니다 . 반응 변수를 지정하지 않으면 빈 보고서가 나타납니다 . 그러나 빨간색 삼각형 메뉴의 " 근접 이웃 행 저장 " 및 " 근접 이웃 거리 저장 " 옵션을 사용할 수 있습니다 .

X, 요인 예측 변수입니다 .

검증 검증 데이터 집합을 정의하는 숫자 열입니다 . 이 열에는 최대 세 개의 구분되는 값이 포함되어야 합니다 .

- 검증 열에 두 개의 수준이 있는 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 검증 열에 세 개의 수준이 있는 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 검증 열에 네 개 이상의 수준이 있는 경우 가장 작은 세 개의 값을 포함하는 행이 검증 데이터 집합을 정의합니다 . 다른 모든 행은 분석에서 제외됩니다 .

K 최근접 이웃 플랫폼에서는 검증 열을 사용하여 모형을 훈련 및 조정하거나 모형을 훈련 , 조정 및 평가합니다 . 검증에 대한 자세한 내용은 "JMP 모델링의 검증 " 에서 확인하십시오 .

" 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다 . " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 " 검증 열 생성 " 에서 확인하십시오 .

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다 . 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다. 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

검증 데이터 비율 검증 데이터 집합으로 사용할 데이터 비율입니다 .

이웃 수 , K 분석할 최근접 이웃의 최대 수입니다 . 모형은 *K* 에 지정한 값까지 하나의 최근접 이웃에 대해 적합됩니다 .

참고 : 최대 이웃 수 *K* 는 훈련 데이터 테이블의 행 수에서 1 을 뺀 값 이하여야 합니다 . 허용 가능한 최대 *K* 보다 큰 *K* 를 지정하면 경고가 나타납니다 .

범주 편향 범주형 반응의 적합 확률이 항상 양수가 되도록 보증하는 조정 모수를 지정합니다 . 범주 편향의 기본값은 0.5 입니다 .

난수 시드값 설정 명목형 및 순서형 반응에 대한 우선 순위 결정에 사용되는 랜덤화 프로세스의 시드값을 설정합니다 . " 검증 데이터 비율 " 을 지정하는 경우 이 옵션은 검증에 사용되는 행의 시드값도 설정합니다 . " 난수 시드값 설정 " 은 분석을 재현하려는 경우에 유용합니다 . 난수 시드값을 설정하고 스크립트를 저장하면 시드값이 스크립트에 자동으로 저장됩니다 .

JMP PRO K 최근접 이웃 보고서

"K 최근접 이웃 " 보고서에는 각 반응 변수에 대한 개별 보고서가 포함되어 있습니다 . 각 반응 변수 보고서에는 해당 반응에 대한 적합 모형 정보가 포함되어 있습니다 . 이 정보는 " 모형 선택 " 보고서 및 적합된 각 *k* 개 모형에 대한 요약 정보를 포함합니다 . 보고서에는 훈련 데이터 집합에 대한 테이블과 검증 및 테스트 데이터 집합 (검증을 사용하여 정의한 경우) 에 대한 테이블이 표시됩니다 .

"모형 선택" 보고서에는 범주형 반응의 오분류 비율 또는 연속형 반응의 RASE 를 기반으로 *K* 에 대한 해 경로 그림이 표시됩니다. 기본적으로 최적 성능 모형에 해당하는 *K* 값에 슬라이더가 배치됩니다. 슬라이더를 드래그하여 보고서에서 *K* 값을 변경할 수 있습니다.

반응의 모델링 유형에 따라 보고되는 통계량이 달라집니다. 요약 테이블의 각 행은 *k* 개의 최근접 이웃으로 정의된 모형에 해당하며, *K* 의 범위는 1 ~ 시작 창의 "이웃 수, K" 에 지정한 값입니다.

JMP PRO 연속형 반응

기본적으로 연속형 반응의 보고서에는 "모형 선택" 그래프 외에 "요약 테이블" 보고서가 포함됩니다.

요약 테이블

별표는 RASE 가 가장 작은 *K* 값에 대한 모형을 표시합니다 . 연속형 반응의 보고서에는 다음 열이 포함됩니다 .

K 모형에 사용된 최근접 이웃 수입니다 . *K* 의 범위는 1 ~ 시작 창의 " 이웃 수 , K" 에 지정한 값입니다 .

개수 관측값 수입니다 .

R^2 모형의 R^2 값입니다 .

RASE 모형에 대한 제곱근 평균 제곱 예측 오차입니다 . RASE 가 가장 작은 모형에 별표가 표시됩니다 . 동점 RASE 값이 있으면 *K* 가 가장 작은 모형에 별표가 표시됩니다 .

SSE 모형에 대한 오차 제곱합입니다 .

JMP PRO 범주형 반응

기본적으로 범주형 반응의 보고서에는 " 모형 선택 " 그래프 외에 " 요약 테이블 ", " 혼동 행렬 " 및 " 모자이크 그림 " 보고서가 포함됩니다 .

요약 테이블

별표는 오분류 비율이 가장 작은 *K* 값에 대한 모형을 표시합니다 . 범주형 반응의 보고서에는 다음 열이 포함됩니다 .

K 모형에 사용된 최근접 이웃 수입니다 . *K* 의 범위는 1 ~ 시작 창의 " 이웃 수 , K" 에 지정한 값입니다 .

개수 관측값 수입니다 .

오분류 비율 모형에 의해 잘못 분류된 관측값의 비율입니다 . 이 값은 " 오분류 " 를 " 개수 " 로 나누어 계산됩니다 . 오분류 비율이 가장 작은 모형에 별표가 표시됩니다 . 동점 오분류 비율 값이 있으면 *K* 가 가장 작은 모형에 별표가 표시됩니다 .

오분류 모형에서 잘못 예측한 관측값 수입니다 .

혼동 행렬

기본적으로 오분류 비율이 가장 작은 모형에 대한 "혼동 행렬" 보고서가 표시됩니다. 가장 작은 오분류 비율에 동점이 있으면 *K* 가 가장 작은 모형에 대한 보고서가 표시됩니다. "혼동 행렬" 보고서에는 혼동 행렬과 혼동 비율 행렬이 포함되어 있습니다. 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다. 혼동 비율 행렬은 값을 행 합계로 나눈 혼동 행렬과 동일합니다. 검증을 사용하는 경우 검증 데이터 집합과 테스트 데이터 집합에 대한 혼동 행렬 및 혼동 비율 행렬이 나타납니다. "혼동 행렬" 보고서와 오분류 비율을 사용하여 모형을 평가할 수 있습니다.

팁 : 해 경로 그림에서 슬라이더 위치를 변경하면 선택한 *K* 값에 대한 추가 " 혼동 행렬 " 보고서가 표시됩니다 . 이 보고서를 사용하여 대체 모형을 기본 최적 모형과 비교할 수 있습니다 .

모자이크 그림

기본적으로 오분류 비율이 가장 작은 모형에 대한 모자이크 그림이 표시됩니다. 가장 작은 오분류 비율에 동점이 있으면 *K*가 가장 작은 모형에 대한 모자이크 그림이 표시됩니다. 모자이크 그림은 각 세그먼트가 해당 그룹의 빈도 수에 비례하는 누적 막대 차트입니다. 모자이크 그림에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오. 검증을 사용하는 경우 검증 데이터 집합과 테스트 데이터 집합에 대한 모자이크 그림이 표시됩니다.

팁 : 해 경로 그림에서 슬라이더 위치를 변경하면 모자이크 그림이 업데이트되어 선택한 *K* 값의 결과가 표시됩니다 .

JMP PRO K 최근접 이웃 플랫폼 옵션

"K 최근접 이웃 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

근접 이웃 행 저장 *K*개 열을 데이터 테이블에 저장합니다. 열 이름은 **이웃 행 k** 입니다. 주어진 행의 *k* 번째 열에는 *k* 번째 최근접 이웃의 행 번호가 포함됩니다 .

주의 : 데이터 테이블에서 행을 재정렬할 때 **이웃 행 k** 열의 행 번호는 업데이트되지 않습니다 . 행을 재정렬하면 해당 열의 값이 올바르지 않습니다 .

근접 이웃 거리 저장 *K* 개 열을 데이터 테이블에 저장합니다 . 열 이름은 **거리 k** 입니다 . 주어진 행의 *k* 번째 열에는 해당 행의 *k* 번째 최근접 이웃까지의 거리가 포함됩니다 .

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

모형 적합 옵션

모자이크 그림 (명목형 또는 순서형 반응에만 사용 가능) 모자이크 그림을 표시하거나 숨깁니다 . 자세한 내용은 " 모자이크 그림 " 에서 확인하십시오 .

실제값 대 예측값 그림 (연속형 반응에만 사용 가능) RASE 가 가장 작은 모형에 대한 실제 반응 대 예측 반응 그림을 표시합니다 . 가장 작은 RASE 에 동점이 있는 경우 그림은 *K* 가 가장 작은 모형을 기반으로 합니다 .

팁 : 해 경로 그림의 슬라이더 위치를 다른 *K* 로 변경하면 " 실제값 대 예측값 그림 " 이 업데이트되어 선택한 *K* 값에 대한 모형이 반영됩니다 .

잔차 대 예측값 그림 (연속형 반응에만 사용 가능) RASE 가 가장 작은 모형에 대한 잔차 대 예측 반응 그림을 표시합니다 . 가장 작은 RASE 에 동점이 있는 경우 그림은 *K* 가 가장 작은 모형을 기반으로 합니다 .

팁 : 해 경로 그림의 슬라이더 위치를 다른 *K* 로 변경하면 " 잔차 대 예측값 그림 " 이 업데이트되어 선택한 *K* 값에 대한 모형이 반영됩니다 .

예측값 저장 *K*개의 예측값 열을 데이터 테이블에 저장합니다. 열 이름은 **예측 <반응> k**입니다. *k* 번째 열에는 *k* 번째 최근접 이웃을 기반으로 하는 모형에 대한 예측이 포함됩니다 . 여기서 **반응**은 반응 열 이름입니다 .

예측 계산식 저장 *k* 개의 특정 최근접 이웃 모형에 대한 예측 계산식이 포함된 열을 저장합니다 . 메시지가 표시되면 *K* 값을 입력합니다 . 예측 계산식에는 모든 훈련 데이터가 포함되므로 큰 데이터 테이블의 경우 이 옵션을 사용하면 실용성이 떨어질 수 있습니다 .

주의 : " 예측 계산식 저장 " 과 " 예측값 저장 " 을 사용한 결과 값이 반드시 일치하는 것은 아닙니다 . " 예측 계산식 저장 " 옵션을 사용하여 얻은 값은 예측값 행을 포함하여 훈련 데이터 집합의 모든 행을 사용합니다 . " 예측값 저장 " 옵션의 값은 훈련 데이터 집합에서 예측값 행을 제외한 다른 모든 행만 사용합니다 .

예측 계산식 게시 지정된 *k* 최근접 이웃 모형에 대한 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

8 장

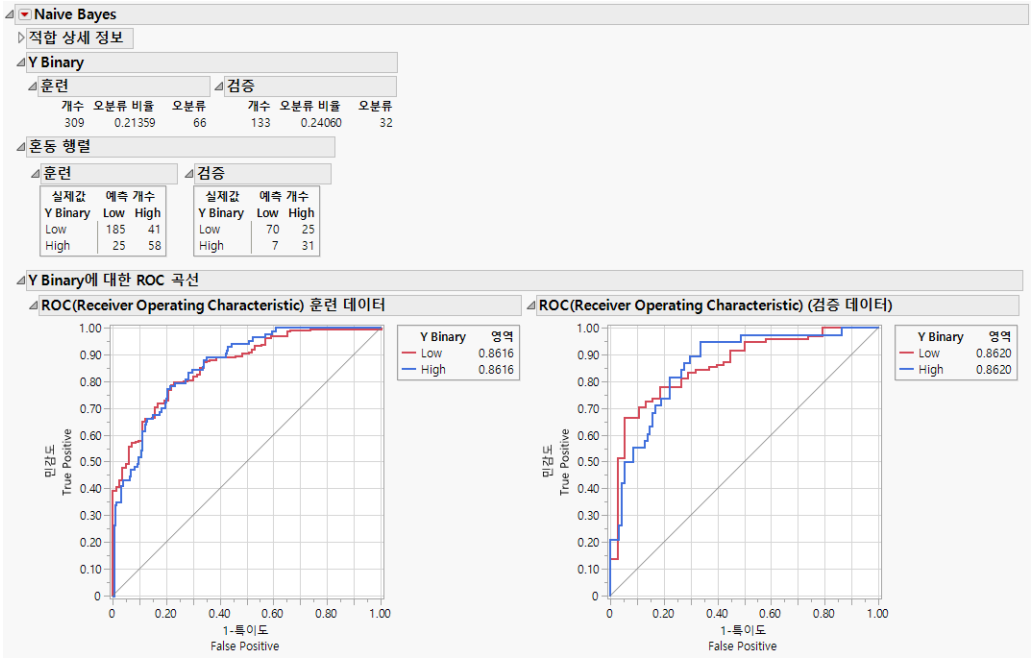
JMP PRO Naive Bayes

확률적 가정을 사용하여 관측값 분류

Naive Bayes 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

Naive Bayes 플랫폼은 범주형 변수의 값을 예측하기 위해 모형을 적합시킵니다. Naive Bayes는 빠르고 계산이 단순한 방법입니다. 특히 예측 변수의 수가 많은 경우에 알맞습니다.

그림 8.1 Naive Bayes 분석의 예



목차

Naive Bayes 플랫폼 개요 129
Naive Bayes 의 예 129
Naive Bayes 플랫폼 시작 132
Naive Bayes 보고서 133
적합 상세 정보 보고서 133
반응 열 보고서 134
혼동 행렬 보고서 134
ROC 곡선 134
Naive Bayes 플랫폼 옵션 134
Naive Bayes 의 추가 예 136
Naive Bayes 플랫폼에 대한 통계 상세 정보 137
알고리즘 137
저장된 확률 계산식 138

JMP PRO Naive Bayes 플랫폼 개요

Naive Bayes 플랫폼에서는 관측값을 범주형 반응 변수의 수준에 의해 정의된 **클래스**로 분류합니다. 분류에 사용되는 변수 또는 요인을 데이터 마이닝 문헌에서는 흔히 **피처**라고 합니다.

Naive Bayes 알고리즘은 각 클래스에 대해 각 피처 값이 발생하는 조건부 확률을 계산합니다. 피처가 연속형이면 조건부 주변 밀도가 추정됩니다. Naive Bayes 기법은 클래스 내에서 피처가 서로 독립적이라고 가정합니다. 이런 이유 때문에 이 기법을 "Naive"라고 합니다. 분류는 특정 클래스 내에서 피처 값의 조건부 확률이 높은 관측값이 해당 클래스에 속할 확률이 높다는 개념을 기반으로 합니다. 자세한 내용은 Hastie et al. (2009) 연구 자료에서 확인하십시오.

이 알고리즘은 1차원 밀도 또는 분포만 추정하므로 속도가 매우 빠릅니다. 따라서 큰 데이터 집합, 특히 피처 수가 많은 데이터 집합에 알맞습니다. 관측값에 대한 모든 비결측 피처 값은 조건부 확률을 계산하는 데 사용됩니다.

각 관측값에는 각 클래스에 대한 **Naive 스코어**가 할당됩니다. 지정된 클래스에 대한 관측값의 Naive 스코어는 해당 클래스에 속한 훈련 관측값의 비율에 관측값의 조건부 확률 곱을 곱한 값입니다. 관측값이 클래스에 속할 **Naive 확률**은 해당 클래스의 Naive 스코어를 모든 클래스의 Naive 스코어 합으로 나눈 값입니다. 관측값은 Naive 확률이 가장 높은 클래스에 할당됩니다.

주의 : 클래스 소속의 조건부 확률은 독립적이라고 가정되므로 Naive Bayes 추정 확률은 비효율적입니다 .

Naive Bayes를 사용하려면 모든 예측 변수 값과 클래스를 표현하기 위해 많은 수의 훈련 관측값이 필요합니다. 검증 데이터 집합의 관측값을 분류하는 중이고 해당 관측값에 훈련 데이터 집합에서 결측인 범주형 예측 변수 값이 있는 경우 플랫폼은 비결측 피처를 사용하여 예측합니다. 모든 예측 변수 값에 대해 관측값이 결측인 경우 예측 반응은 가장 빈도가 높은 반응입니다. 예측 계산식은 결측값을 관측값 스코어에 전혀 기여하지 않는 것으로 설정하여 결측값을 처리합니다.

Naive Bayes 기법에 대한 자세한 내용은 Hand et al. (2001) 및 Shmueli et al. (2010) 연구 자료에서 확인하십시오.

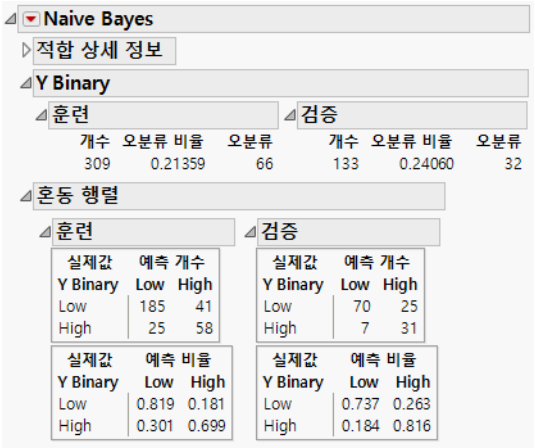
JMP PRO Naive Bayes 의 예

미래 환자의 질병 경과를 High 또는 Low 로 예측하는 데 사용할 분류 모형을 생성하려고 합니다 . 442 명의 당뇨병 환자에 대한 기준 의료 데이터가 있습니다 . 또한 각 환자가 병원에 처음 내원한 후 1년 동안 얻은 당뇨병 진행 경과의 이진 측정값도 있습니다 . 이 측정값은 질병 경과를 Low 또는 High 로 정량화합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > Naive Bayes** 를 선택합니다 .
3. Y Binary 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Age ~ Glucose 를 선택하고 **X, 요인**을 클릭합니다 .

5. Validation 을 선택하고 **검증**을 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 8.2 Naive Bayes 보고서

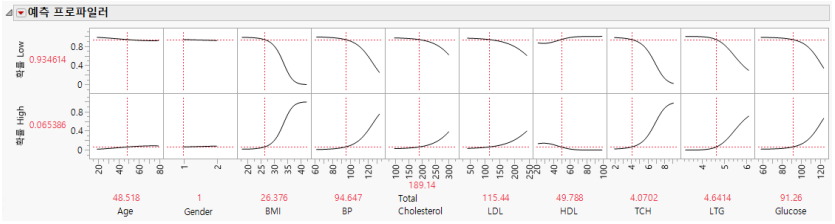


훈련 데이터 집합의 오분류 비율은 약 21% 이고 검증 데이터 집합의 오분류 비율은 약 24% 입니다 . " 혼동 행렬 " 을 보면 훈련 데이터 집합과 검증 데이터 집합 둘 다 질병 경과가 Low 인 환자를 High 로 분류한 것이 오분류의 큰 원인임을 알 수 있습니다 . 검증 데이터 집합 결과는 모형이 독립적 관측값으로 확장하는 방식을 나타냅니다 .

Naive Bayes 분류에 가장 큰 영향을 미치는 개별 예측 변수에 대해 알아보려고 합니다 .

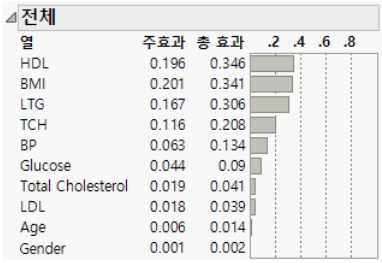
7. "Naive Bayes" 의 빨간색 삼각형을 클릭하고 **프로파일러**를 선택합니다 .

그림 8.3 질병 경과에 대한 예측 프로파일러



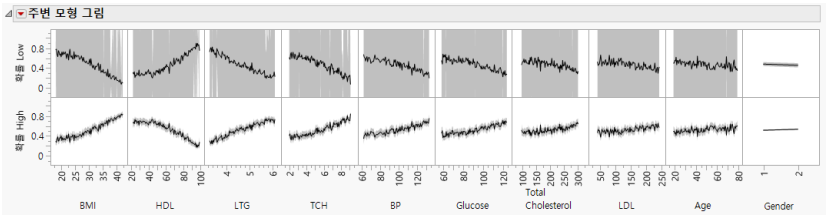
8. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **변수 중요도 평가 > 독립 균등 입력**을 선택합니다 .

그림 8.4 변수 중요도



" 요약 보고서 " 에서는 HDL, BMI 및 LTG 가 추정 확률에 가장 큰 영향을 미친다는 것을 나타냅니다 .

그림 8.5 주변 모형 그림 보고서

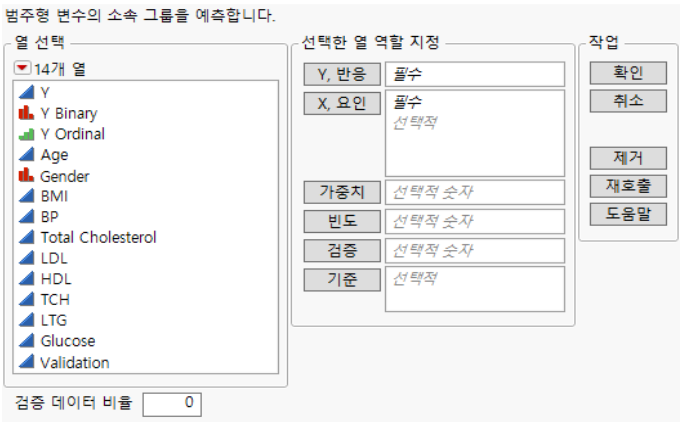


" 주변 모형 그림 " 보고서의 두 번째 행 그림에서는 HDL 값이 높을수록 환자를 High 로 분류할 확률이 낮음을 보여 줍니다 . 또한 BMI 와 LTG 값이 높을수록 환자를 High 로 분류할 확률이 높습니다 .

JMP PRO Naive Bayes 플랫폼 시작

분석 > 예측 모델링 > Naive Bayes 를 선택하여 Naive Bayes 플랫폼을 시작합니다 .

그림 8.6 Naive Bayes 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Naive Bayes 시작 창에는 다음과 같은 옵션이 있습니다 .

Y, 반응 관심 클래스가 값인 범주형 반응 열입니다 .

X, 요인 범주형 또는 연속형 예측 변수 열입니다 .

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다 .

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다 .

검증 검증 데이터 집합을 정의하는 숫자 열입니다 . 이 열에는 최대 세 개의 구분되는 값이 포함되어야 합니다 .

- 열의 고유 값이 세 개인 경우 다음과 같습니다 .

  가장 작은 값이 있는 행이 훈련 데이터 집합에 사용됩니다 .

  중간 값이 있는 행이 검증 데이터 집합에 사용됩니다 .

  가장 큰 값이 있는 행이 테스트 데이터 집합에 사용됩니다 .
- 열의 고유 값이 두 개인 경우 훈련 데이터 집합과 검증 데이터 집합만 사용됩니다 .
- 검증 열에 네 개 이상의 수준이 있는 경우 가장 작은 세 개의 값을 포함하는 행이 검증 데이터 집합을 정의합니다 . 다른 모든 행은 분석에서 제외됩니다 .

Naive Bayes 플랫폼에서는 검증 열을 사용하여 모형을 훈련하고 평가합니다 . 검증에 대한 자세한 내용은 "JMP 모델링의 검증 " 에서 확인하십시오 .

" 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다 . " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 " 검증 열 생성 " 에서 확인하십시오 .

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다 . 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다 . 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

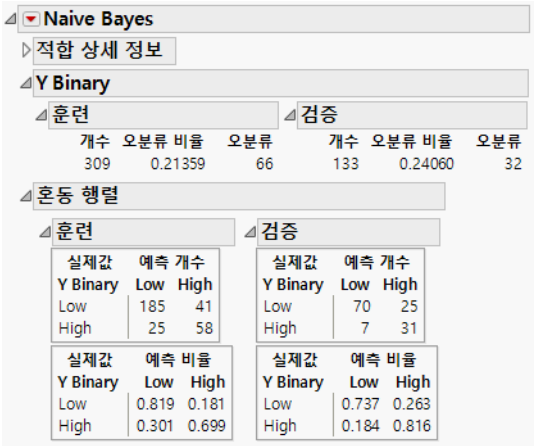
검증 데이터 비율 검증 데이터 집합으로 사용할 데이터 비율입니다 .

참고 : 시작 창에서 검증 열 또는 검증 데이터 비율을 지정하지 않은 경우 제외된 행이 있으면 이러한 행은 검증 데이터 집합으로 처리됩니다 . 검증에 대한 자세한 내용은 "JMP 모델링의 검증 " 에서 확인하십시오 .

JMP PRO Naive Bayes 보고서

시작 창에서 **확인**을 클릭하면 "Naive Bayes" 보고서가 나타납니다 . 기본적으로 "Naive Bayes" 보고서에는 적합 상세 정보 , 반응 열 , 혼동 행렬 , ROC 곡선 및 향상도 곡선에 대한 보고서가 포함됩니다 .

그림 8.7 Naive Bayes 보고서



JMP PRO 적합 상세 정보 보고서

"적합 상세 정보" 보고서에는 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 모형의 다양한 적합 측도가 표시됩니다 . "측도" 열에는 여러 적합 통계량이 나열되고 "정의"

열에는 해당 적합 통계량의 계산식이 표시됩니다. 자세한 내용은 "범주형 반응에 대한 적합 측도" 에서 확인하십시오. 기본적으로 "Naive Bayes" 보고서 창의 " 적합 상세 정보 " 보고서는 닫혀 있습니다.

JMP PRO 반응 열 보고서

반응 열 보고서에는 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 Naive Bayes 분류의 성능 통계량을 보여 주는 요약 테이블이 있습니다 . 요약 테이블에는 다음 열이 포함됩니다 .

개수 테이블에 해당하는 데이터 집합(훈련, 검증 또는 테스트 데이터 집합)의 관측값 수입니다.

오분류 비율 해당 데이터 집합에서 모형에 의해 잘못 분류된 관측값의 비율입니다 . 이 값은 " 오분류 " 를 " 개수 " 로 나누어 계산됩니다 .

오분류 해당 데이터 집합에서 잘못 분류된 관측값 수입니다 .

JMP PRO 혼동 행렬 보고서

"혼동 행렬" 보고서에는 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 혼동 행렬과 혼동 비율 행렬이 표시됩니다. 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다. 혼동 비율 행렬은 값을 행 합계로 나눈 혼동 행렬과 동일합니다.

JMP PRO ROC 곡선

이 보고서에는 훈련 데이터 집합과 검증 및 테스트 데이터 집합(지정된 경우)에 대한 ROC(Receiver Operating Characteristic) 곡선이 표시됩니다. ROC 곡선은 반응 수준을 올바르게 분류하기 위한 적합 확률 능력을 측정합니다. 주대각과 곡선의 거리가 멀수록 적합도가 높습니다. ROC 곡선에 대한 소개는 **기본 분석**의 에서 확인하십시오.

ROC 곡선에 대한 자세한 내용은 "ROC 곡선" 에서 확인하십시오.

JMP PRO Naive Bayes 플랫폼 옵션

"Naive Bayes" 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

ROC 곡선 ROC 그림을 표시하거나 숨깁니다 . ROC 그림은 기본적으로 표시됩니다 . 자세한 내용은 "ROC 곡선 " 에서 확인하십시오 .

향상도 곡선 모형에 대한 향상도 곡선을 표시하거나 숨깁니다 . 검증을 사용한 경우 훈련 데이터 집합과 검증 및 테스트 데이터 집합 (지정된 경우) 에 대해 각각 향상도 곡선이 표시됩니다 .

향상도 곡선은 적합 확률이 감소할 때 반응 수준이 얼마나 효과적으로 분류되는지 보여 줍니다 . 적합 확률은 가로 축을 따라 내림차순으로 표시됩니다 . 적합 확률의 수직 좌표는 해당 확률 이상에 대한 올바른 분류 비율을 전체 올바른 분류 비율로 나눈 값입니다 . 적합 확률이 임

계값을 초과하는 관측값만 선택한 경우 향상도 곡선을 사용하여 대부분의 관측값을 올바르게 분류할 수 있는지 확인할 수 있습니다 .

반응 수준이 두 개이면 향상도 곡선 그림에 반응의 첫 번째 수준에 대한 향상도 곡선만 표시됩니다 . 반응 수준이 세 개 이상이면 향상도 곡선 그림에 각 반응 수준에 대한 곡선의 하위 개요가 표시됩니다 . 향상도 곡선에 대한 자세한 내용은 " 향상도 곡선 " 에서 확인하십시오 .

예측값 저장 예측 분류를 데이터 테이블의 Naive 예측 <Y, 반응 > 열에 저장합니다 .

예측 계산식 저장 Naive 예측 계산식 <Y, 반응 > 열을 데이터 테이블에 저장합니다 . 이 열에는 분류에 대한 예측 계산식이 포함됩니다 .

확률 계산식 저장 각 관측값을 분류하는 데 사용되는 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 다음과 같은 세 개의 열 그룹이 저장됩니다 .

Naive 스코어 < 클래스 >, 합 클래스를 나타내는 각 열에 대해 이 열은 지정된 클래스의 소속 강도를 측정하는 스코어 계산식을 제공합니다 . "Naive 스코어 합 " 열에서 이러한 스코어는 전체 클래스에 대해 합산됩니다 . 자세한 내용은 " 저장된 확률 계산식 " 에서 확인하십시오 .

Naive 확률 < 클래스 > 각 클래스에 대해 이 열은 관측값이 해당 클래스에 속할 조건부 확률 계산식을 제공합니다 . 자세한 내용은 " 저장된 확률 계산식 " 에서 확인하십시오 .

Naive 예측 계산식 <Y, 반응 > 예측 클래스에 대한 계산식을 제공합니다 .

확률 계산식 게시 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않은 경우 이 옵션을 선택하면 " 계산식 저장소 " 창이 열립니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

프로파일러 대화식 프로파일러 보고서를 표시하거나 숨깁니다 . 요인 값을 변경하면 추정된 분류 확률에 반영됩니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO Naive Bayes 의 추가 예

주택 담보 대출을 신청한 5,960 명의 고객에 대한 과거 재무 데이터가 있습니다 . 각 고객은 Good Risk(안전한 위험) 또는 Bad Risk(불안한 위험) 로 분류되었습니다 . 대부분의 예측 변수에는 결측 데이터가 있습니다 . 미래 고객의 신용 위험을 분류하는 데 사용할 모형을 생성하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Equity.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > Naive Bayes** 를 선택합니다 .
3. BAD 를 선택하고 **Y, 반응**을 클릭합니다 .

   잠재적 예측 변수 중 하나인 DEBTINC 에는 정보가 될 수 있는 결측값이 많이 있습니다 . 그러나 Naive Bayes 는 많은 수의 결측값을 잘 처리할 준비가 되어 있지 않으므로 모형에 DEBTINC 를 포함하지 않습니다 .
4. LOAN ~ CLNO 를 선택하고 **X, 요인**을 클릭합니다 .
5. Validation 을 선택하고 **검증**을 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 8.8 BAD 에 대한 Naive Bayes 보고서

BAD

훈련

개수	오분류 비율	오분류
3576	0.18009	644

검증

개수	오분류 비율	오분류
1192	0.19128	228

테스트

개수	오분류 비율	오분류
1192	0.18960	226

혼동 행렬

훈련

실제값 BAD	예측 개수 Good Risk	Bad Risk
Good Risk	2691	214
Bad Risk	430	241

실제값 BAD	예측 비율 Good Risk	Bad Risk
Good Risk	0.926	0.074
Bad Risk	0.641	0.359

검증

실제값 BAD	예측 개수 Good Risk	Bad Risk
Good Risk	859	58
Bad Risk	170	105

실제값 BAD	예측 비율 Good Risk	Bad Risk
Good Risk	0.937	0.063
Bad Risk	0.618	0.382

테스트

실제값 BAD	예측 개수 Good Risk	Bad Risk
Good Risk	885	64
Bad Risk	162	81

실제값 BAD	예측 비율 Good Risk	Bad Risk
Good Risk	0.933	0.067
Bad Risk	0.667	0.333

훈련 , 검증 및 테스트 데이터 집합의 오분류 비율은 18% 에서 19% 사이입니다 . 모든 데이터 집합의 혼동 행렬은 Bad Risk 고객을 Good Risk 고객으로 분류한 것이 오분류의 가장 큰 원인임을 나타냅니다 .

특정 재무 배경 값을 가진 고객이 High Risk 로 분류될 확률에 관심이 있습니다 .

7. "Naive Bayes" 의 빨간색 삼각형을 클릭하고 **확률 계산식 저장**을 선택합니다 .

   3 개의 열 집합이 데이터 테이블에 추가됩니다 .
   - 세 개의 Naive 스코어 열에는 Good Risk, Bad Risk 및 이 둘의 합에 대한 Naive 스코어 계산식이 포함됩니다 .
   - 두 개의 Naive 확률 열에는 Good Risk 와 Bad Risk 에 대한 확률 계산식이 포함됩니다 .
   - Naive 예측 계산식 Bad 열에는 관측값의 Naive 확률이 가장 높은 클래스에 관측값을 할당하는 계산식이 포함됩니다 .

이 계산식을 사용하여 새 고객을 스코어링할 수 있습니다 . 계산식 열에 대한 자세한 내용은 " 저장된 확률 계산식 " 에서 확인하십시오 .

JMP PRO Naive Bayes 플랫폼에 대한 통계 상세 정보

- " 알고리즘 "
- " 저장된 확률 계산식 "

JMP PRO 알고리즘

Naive Bayes 방법은 피처 값이 주어졌을 때 관측값을 소속 확률이 가장 높은 클래스로 분류합니다. 이 방법에서는 피처가 각 클래스 내에서 조건부로 독립적이라고 가정합니다.

가능한 분류를 $C_1, \ldots, C_K$로 나타냅니다. 피처 또는 예측 변수를 $X_1, X_2, \ldots, X_p$로 나타냅니다.

예측 변수 값이 $x_1, x_2, \ldots, x_p$ 인 관측값이 C_k 클래스에 속할 조건부 확률은 다음과 같이 계산됩니다 .

$$P(C_k|(x_1, \ldots, x_p)) = \frac{\left(P(C_k)\prod_{j=1}^{p}[P(x_j|C_k)]\right)/(\mathbf{R})}{\sum_{k=1}^{K}\left(P(C_k)\left(\prod_{j=1}^{p}[P(x_j|C_k)]\right)/\mathbf{R}\right)},$$

여기서 **R** 은 정규화 상수입니다 . 위의 계산식에서 $X_j = x_j$ 인 관측값이 C_k, $P(x_j|C_k)$ 클래스에 속할 조건부 확률은 다음과 같이 결정됩니다 .

- X_j 가 범주형인 경우

$$P(x_j|C_k) = \frac{\text{다음 요소의 관측값 수 - } x_j|C_k}{\text{다음 요소의 관측값 수 - } C_k}$$

- X_j 가 연속형인 경우

$$P(x_j|C_k) = \frac{1}{s_{jk}}\phi((x_j - m_{jk})/s_{jk})$$

여기서 ϕ 는 표준 정규 밀도 함수이고 , m 과 s 는 각각 C_k 클래스 내 예측 변수 값의 평균과 표준편차입니다 .

관측값이 C_k, $P(C_k)$ 클래스에 속할 무조건부 확률은 다음과 같이 계산됩니다 .

$$P(C_k) = \frac{\text{다음 요소의 관측값 수 - } C_k + (0.5/K)}{\text{총 관측값 수} + 0.5}$$

관측값은 조건부 확률이 가장 큰 클래스로 분류됩니다.

참고 : $P(C_k)$ 계산식에서 0.5 는 사전 편향 요인입니다 . 이 값이 기본값입니다 . 사전 기본 요인을 변경하려면 파일 > 환경 설정 > 플랫폼 > Naive Bayes 로 이동한 후 " 사전 편향 " 체크박스를 선택하고 값을 변경하십시오 .

JMP PRO 저장된 확률 계산식

이 섹션에서는 " 확률 계산식 저장" 옵션을 사용하여 저장된 계산식에 대해 설명합니다. 예측 변수 값이 $x_1, x_2, \ldots, x_p$ 인 관측값이 C_k 클래스에 속할 조건부 확률은 "알고리즘" 섹션에 표시된 $P(C_k|(x_1,\ldots, x_p))$ 에서 제공하는 확률과 약간 다릅니다. 이는 계산 효율성을 위해서입니다.

Naive 스코어 계산식

지정된 클래스 C_k 에 대한 Naive 스코어 계산식 $S(C_k)$ 는 $P(C_k|(x_1,\ldots, x_p))$ 의 표현식에서 분자가 변형된 것으로 , 다음과 같이 계산됩니다 .

$$S(C_k) = \exp[\ln\{P(C_k)\} + \text{연속형} + \text{범주형} + \ln(\mathbf{R})]$$

Naive 스코어 계산식은 연속형 예측 변수와 범주형 예측 변수의 스코어 조합입니다. **R**은 정규화 상수입니다. 계산식의 연속형 부분은 다음과 같이 계산됩니다.

$$\text{연속형} = \sum_{j=1}^{p_1} \left\{ \text{Normal Log Density}\left(\frac{x_j - m_{jk}}{s_{jk}}\right) + \ln(1/s_{jk}) \right\}$$

다음은 각 요소에 대한 설명입니다.

$j = 1,\ldots, p_1$ 연속형 예측 변수

계산식의 범주형 부분은 다음과 같이 계산됩니다.

$$\text{범주형} = \sum_{r=1}^{p_2} \left\{ \sum_{l=1}^{L_r} 1_{rl} \ln(P(x_{rl}|C_k)) \right\}$$

다음은 각 요소에 대한 설명입니다.

$r = 1,...,p_2$ 범주형 변수

$l = r$ 번째 범주형 변수의 $1,..., L_r$ 수준

$1_{rl} = x_{rl}$ 이 r 번째 범주형 예측 변수의 l 번째 수준인 경우 1 이고 , 그렇지 않은 경우 0 인 표시 변수

Naive 스코어 합 계산식

Naive 스코어 합 계산식 S 는 모든 클래스에 대해 Naive 스코어 계산식을 합한 것입니다 . 이 계산식은 $P(C_k|(x_1,..., x_p))$ 의 표현식에서 분모가 변형되었습니다 .

$$S = \sum_{k=1}^{K} S(C_k)$$

Naive 확률 계산식

지정된 클래스 C_k 의 Naive 확률 계산식은 $P(C_k|(x_1,..., x_p))$ 입니다 . JMP 계산식에서는 다음과 같습니다 .

$$P(C_k|(x_1, \ldots, x_p)) = \frac{S(C_k)}{S}$$

Naive 예측 계산식

관측값의 Naive 예측 계산식은 해당 관측값을 $P(C_k|(x_1,..., x_p))$ 가 가장 큰 클래스로 분류합니다. 이것은 관측값을 Naive 스코어 계산식이 가장 큰 클래스 분류하는 것과 동일합니다.

9 장

JMP PRO 서포트 벡터 머신

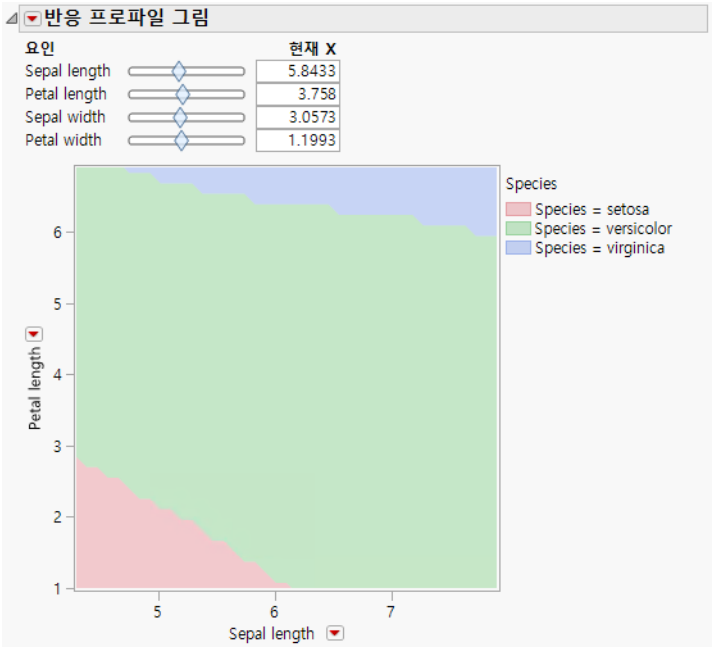
서포트 벡터를 사용하여 반응 범주 예측

서포트 벡터 머신 플랫폼은 JMP Pro 에서만 사용할 수 있습니다.

JMP Pro 의 SVM(서포트 벡터 머신) 플랫폼에서는 비선형 회귀와 분류를 위한 머신 학습 도구를 제공합니다. SVM 알고리즘은 훈련 데이터를 사용하여 새 관측값을 예측하거나 분류하는 모형을 생성하는 지도 학습 알고리즘입니다.

범주형 반응을 사용하는 경우 반응 프로파일 그림에서 모형 분류를 시각화할 수 있습니다. 또한 SVM 플랫폼에서는 모형이 데이터를 얼마나 잘 적합시키는지 확인할 수 있도록 오분류 비율, 혼동 행렬, ROC 곡선 및 향상도 곡선을 제공합니다. 연속형 반응을 사용하는 경우 모형이 데이터를 얼마나 잘 적합시키는지 확인하기 위해 적합 통계량 및 실제값 대 예측값 그림을 사용합니다.

그림 9.1 반응 프로파일 그림



목차

서포트 벡터 머신 개요 143
서포트 벡터 머신의 예 143
서포트 벡터 머신 플랫폼 시작 145
서포트 벡터 머신 시작 창 146
모형 시작 제어판 147
서포트 벡터 머신 보고서 149
모형 비교 보고서 149
서포트 벡터 머신 모형 보고서 150
서포트 벡터 머신 플랫폼 옵션 152
서포트 벡터 머신 모형 보고서 옵션 152
SVM 플랫폼의 추가 예 154
연속형 반응에 대한 서포트 벡터 회귀의 예 154

JMP PRO 서포트 벡터 머신 개요

SVM(서포트 벡터 머신) 모형은 새 관측값을 예측하거나 분류하는 데 사용되는 지도 학습 알고리즘입니다. 모형은 반응이 알려진 훈련 데이터 집합에 대해 적합됩니다. 그런 다음 모형을 사용하여 새 관측값의 반응을 예측합니다.

반응이 범주형인 경우 SVM 모형은 클래스를 구분하는 초평면을 최적화하여 데이터를 분류합니다. 이것은 클래스 간의 마진을 최대화하는 초평면을 찾는 것으로 볼 수도 있습니다. 단순한 문제에서 이 초평면은 선형입니다. 그러나 더 복잡한 데이터는 선형으로 구분할 수 없는 경우가 많습니다. 이러한 경우 SVM 플랫폼에서는 방사 기저 함수 커널을 사용하는 옵션을 제공하여 클래스를 더 쉽게 분리할 수 있는 비선형 차원에 점을 매핑합니다.

반응이 연속형인 경우 적합된 모형을 SVR(서포트 벡터 회귀) 모형이라고 합니다. 일반적인 회귀 분석 문제의 목표는 예측 반응과 실제 반응 사이의 오차를 최소화하는 모형을 적합시키는 것입니다. SVR 문제의 목표는 예측 반응과 실제 반응 사이의 오차가 $-\varepsilon \sim \varepsilon$ 범위 내에 속하도록 모형을 적합시키는 것입니다. 이렇게 하면 더 유연한 적합을 얻게 됩니다. JMP Pro에서 ε는 0.1과 같습니다. SVR 알고리즘은 $Y+\varepsilon$와 $Y-\varepsilon$라는 두 클래스를 생성하여 데이터를 두 배로 늘립니다. 그런 다음 분류 문제에 사용된 동일한 알고리즘이 예측(SVR) 문제에도 사용됩니다.

SVM 알고리즘의 최대화는 2차 계획법 문제의 해를 구하여 수행됩니다. JMP Pro에서 SVM 플랫폼에 사용되는 알고리즘은 1998년 John Platt이 소개한 SMO(Sequential Minimal Optimization) 알고리즘을 기반으로 합니다. 일반적으로 SVM 2차 계획법 문제는 상당히 큰 규모입니다. SMO 알고리즘은 전체 2차 계획법 문제를 일련의 더 작은 2차 계획법 문제로 분할합니다. 더 작은 2차 계획법 문제는 수치적 계산 대신 분석적으로 해를 구합니다. 즉, 닫힌 형식의 해가 생성됩니다. 따라서 SMO 알고리즘은 전체 2차 계획법 문제의 해를 구하는 것보다 더 효율적입니다(Platt 1998).

JMP PRO 서포트 벡터 머신의 예

442 명의 당뇨병 환자에 대한 기준 의료 데이터가 있습니다. 또한 각 환자가 병원에 처음 내원한 후 1 년 동안 얻은 당뇨병 진행 경과의 이진 측정값도 있습니다. 이 측정값은 질병 경과를 Low 또는 High 로 정량화합니다. 미래 환자의 질병 경과를 High 또는 Low 로 예측하기 위해 분류 모형을 생성하려고 합니다. 두 커널 함수 옵션을 모두 살펴봅니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다.
2. **분석 > 예측 모델링 > 서포트 벡터 머신**을 선택합니다.
3. Y Binary 를 선택하고 **Y, 반응**을 클릭합니다.
4. Age ~ Glucose 를 선택하고 **X, 요인**을 클릭합니다.
5. Validation 을 선택하고 **검증**을 클릭합니다.
6. **확인**을 클릭합니다.

7. " 모형 시작 " 제어판에서 커널 함수가 " 방사 기저 함수 " 인지 확인하고 " 설계 조정 " 을 선택합니다 .
8. " 런 수 " 옆에 "10" 을 입력합니다 .
9. **시작**을 클릭합니다 .
10. " 모형 시작 " 옆의 회색 삼각형을 클릭하여 " 모형 시작 " 제어판을 엽니다 .
11. 커널 함수를 " 선형 " 함수로 변경하고 " 설계 조정 " 을 선택합니다 .
12. " 런 수 " 옆에 "10" 을 입력합니다 .
13. **시작**을 클릭합니다 .

그림 9.2 모형 비교 보고서

모형 비교

표시	방법	커널 함수	비용	감마	SV 수	훈련 오분류 비율	검증 오분류 비율	검증 일반화 R^2	확률 임계	검증 오분류 비율 임계	최선
☐	모형 1	방사 기저 함수	2.57061	0.43877	251	0.00324	0.25564	0.00838	0.39079	0.25564	
☐	모형 2	방사 기저 함수	3.04271	0.31899	218	0.01294	0.28571	-0.4206	0.23223	0.28571	
☐	모형 3	방사 기저 함수	3.93906	0.46899	255	0.00324	0.28571	-0.0081	0.53372	0.28571	
☐	모형 4	방사 기저 함수	1.073	0.27968	208	0.0712	0.21805	0.22158	0.17372	0.21805	
☐	모형 5	방사 기저 함수	0.41677	0.0281	156	0.15534	0.18797	0.50485	0.36015	0.18797	
☐	모형 6	방사 기저 함수	0.23369	0.39898	246	0.26861	0.28571	0.22323	6e-8	0.28571	
☐	모형 7	방사 기저 함수	4.8013	0.15144	167	0.05178	0.2782	-0.4245	0.26219	0.2782	
☐	모형 8	방사 기저 함수	3.51202	0.20295	181	0.04207	0.28571	-0.454	0.22753	0.28571	
☐	모형 9	방사 기저 함수	4.2693	0.06093	140	0.11974	0.19549	0.42841	0.32812	0.19549	
☐	모형 10	방사 기저 함수	1.87641	0.0919	150	0.13269	0.21053	0.43499	0.30598	0.21053	
☐	모형 11	선형	4.42735	.	122	0.17476	0.17293	0.51944	0.47722	0.17293	
☐	모형 12	선형	4.99046	.	123	0.16828	0.17293	0.52037	0.47551	0.17293	
☐	모형 13	선형	3.8272	.	123	0.17476	0.17293	0.51925	0.47763	0.17293	
☐	모형 14	선형	2.74788	.	124	0.17152	0.17293	0.51723	0.48019	0.17293	
☐	모형 15	선형	3.22207	.	123	0.17476	0.17293	0.5185	0.47842	0.17293	
☐	모형 16	선형	1.29304	.	124	0.17152	0.17293	0.51523	0.4819	0.17293	
☐	모형 17	선형	1.74671	.	124	0.16828	0.17293	0.51556	0.48143	0.17293	
☐	모형 18	선형	0.83575	.	123	0.16505	0.15789	0.51498	0.48207	0.15789	
☐	모형 19	선형	2.22	.	125	0.17152	0.17293	0.5158	0.48255	0.17293	
☑	모형 20	선형	0.04975	.	138	0.15534	0.14286	0.52302	0.41946	0.14286	최소 오분류...

" 모형 비교 " 보고서에서는 오분류 비율과 R^2 을 기준으로 한 결과 모형 20 이 최적 모형임을 보여 줍니다 . 이 모형에는 비용 모수가 0.04975 인 선형 커널 함수가 있습니다 . 이 모형으로 추가 분석을 수행합니다 .

그림 9.3 최량 적합 모형에 대한 모형 보고서

"모형 요약" 보고서에서는 훈련 데이터 집합과 검증 데이터 집합의 오분류 비율이 매우 비슷하다는 것을 보여 줍니다. 이것은 모형이 데이터를 과대적합하지 않았음을 나타내는 좋은 표시입니다 . 혼동 행렬은 모형에 의해 잘못 분류된 관측값 유형에 대한 자세한 정보를 제공합니다 . 혼동 행렬의 왼쪽 위를 보면 대부분의 경우 모형이 Low 반응을 올바르게 분류한다는 것을 알 수 있습니다 (훈련 데이터 집합의 96%, 검증 데이터 집합의 92.6%). 그러나 High 반응 중 올바르게 분류된 것은 더 적습니다 (훈련 데이터 집합의 53%, 검증 데이터 집합의 68.4%). 따라서 오분류의 대부분은 High 반응이 Low 로 잘못 분류된 것입니다 .

JMP PRO 서포트 벡터 머신 플랫폼 시작

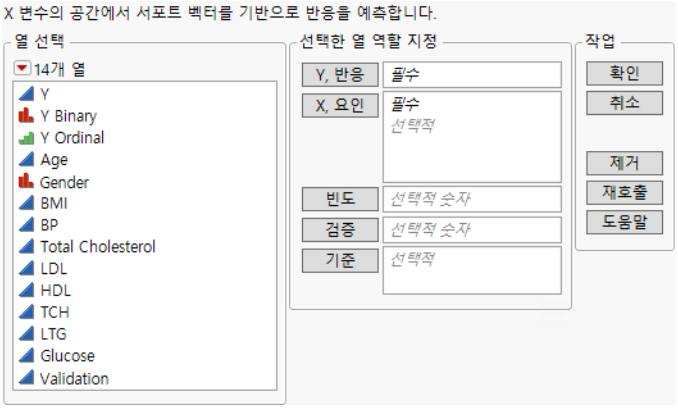
분석 > 예측 모델링 > 서포트 벡터 머신을 선택하여 SVM(서포트 벡터 머신) 플랫폼을 시작합니다 .

SVM 플랫폼 시작은 2 단계 프로세스입니다. 먼저 SVM 시작 창에 변수를 입력합니다. 그런 다음 "모형 시작" 제어판에서 옵션을 지정합니다.

JMP PRO 서포트 벡터 머신 시작 창

SVM 시작 창을 사용하여 반응 및 모형 요인을 지정합니다. 검증 및 기준 열이 사용되는 경우 해당 열도 시작 창에 지정됩니다.

그림 9.4 서포트 벡터 머신 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

서포트 벡터 머신 시작 창에는 다음과 같은 옵션이 있습니다.

Y, 반응 분석할 반응 변수입니다.

X, 요인 예측 변수입니다.

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다.

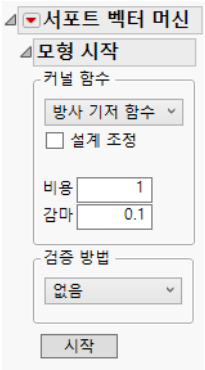
검증 검증 데이터 집합을 정의하는 숫자 열입니다. 자세한 내용은 "검증 방법"에서 확인하십시오. "열 선택" 목록에서 아무 열도 선택하지 않은 상태로 "검증" 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다. "검증 열 생성" 유틸리티에 대한 자세한 내용은 "검증 열 생성"에서 확인하십시오.

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다. 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다. 결과는 개별 보고서에 제공됩니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

JMP PRO 모형 시작 제어판

시작 창에서 " 확인 " 을 클릭하면 모형 적합을 위한 " 모형 시작 " 제어판을 보여 주는 " 서포트 벡터 머신 " 보고서 창이 나타납니다 . " 모형 시작 " 제어판을 사용하여 커널 함수 및 관련 모수 값과 검증 방법을 지정할 수 있습니다 .

그림 9.5 모형 시작 제어판



" 모형 시작 " 제어판에는 다음 옵션이 포함되어 있습니다 .

커널 함수 모형에 사용되는 커널 함수를 지정합니다 . 다음 커널 함수 중 하나를 선택할 수 있습니다 .

방사 기저 함수 클래스를 구분하는 비선형 초평면을 생성하려면 방사 기저 함수 커널을 선택합니다 .

– " 비용 " 모수는 훈련 데이터 집합의 관측값을 잘못 분류하는 것과 관련된 벌점입니다 . 비용 모수가 높을수록 훈련 데이터 집합의 점을 잘못 분류할 가능성이 낮은 알고리즘이 구현되는 반면 , 비용 모수가 낮을수록 마진이 더 넓어집니다 . " 비용 " 모수는 0 보다 커야 하며 기본값은 1 입니다 .

– " 감마 " 모수는 커널 함수의 모수입니다 . 이 모수는 결정선에 대한 곡률의 정도를 결정합니다 . 감마 값이 높을수록 곡률이 크다는 것을 나타냅니다 . 비선형 결정선은 더 유연한 적합을 제공하지만 곡률이 너무 크면 과대적합이 발생할 수 있습니다 . " 감마 " 모수는 0 보다 커야 하며 기본값은 1/(예측 변수의 수) 입니다 .

선형 클래스를 구분하는 선형 초평면을 생성하려면 선형 커널 함수를 선택합니다 .

– " 비용 " 모수는 훈련 데이터 집합의 관측값을 잘못 분류하는 것과 관련된 벌점입니다 . 비용 모수가 높을수록 훈련 데이터 집합의 점을 잘못 분류할 가능성이 낮은 알고리즘이 구현되는 반면 , 비용 모수가 낮을수록 마진이 더 넓어집니다 . " 비용 " 모수는 0 보다 커야 하며 기본값은 1 입니다 .

참고 : 범위를 벗어난 모수 값을 지정하면 기본값이 사용됩니다 .

팁 : 최량 적합 모형을 찾으려면 다양한 커널 함수 및 모수 값을 적합시키고 " 모형 비교 " 보고서를 사용하십시오 .

설계 조정 지정된 커널에 대한 모수 값 범위를 적합시킬 수 있습니다 . R^2 이 가장 크고 오분류 비율 또는 RASE 가 가장 작은 모형이 " 모형 비교 " 보고서에서 식별됩니다 . " 설계 조정 " 을 선택한 후에는 모수의 최소값과 최대값을 지정해야 합니다 . 데이터를 기반으로 한 기본값이 제공되며 최소값은 0 보다 커야 합니다 . " 런 수 " 의 값도 지정해야 합니다 . SVM 플랫폼은 최소값과 최대값에 의해 결정된 모수 값 격자에서 많은 모형을 적합시킵니다 .

검증 방법 모형 검증 방법을 지정합니다 . " 시작 " 버튼을 처음 클릭하면 지정된 검증 방법을 사용하여 첫 번째 SVM 모형이 적합됩니다 . 그런 다음 SVM 창 내에서 적합된 모든 SVM 모형에 이 검증 방법이 사용됩니다 . 이렇게 하면 보고서 창의 모든 모형이 동일한 검증 방법 및 검증 데이터 집합을 사용하여 적합됩니다 .

홀드백 원래 데이터를 무작위로 훈련 데이터 집합과 검증 데이터 집합으로 나눕니다 . 원래 데이터에서 검증 데이터 집합으로 사용할 비율을 지정할 수 있습니다 (홀드백).

K 폴드 (*Y* 가 연속형 또는 명목형인 경우에만 사용 가능) 원래 데이터를 무작위로 *K* 개의 부분집합으로 나눕니다 . 그러면 *K* 개의 각 집합이 나머지 데이터에 대한 모형 적합을 검증하는 데 사용되며 총 *K* 개 모형을 적합시킵니다 . *Y* 가 연속형이면 최량 검증 RASE 통계량을 제공하는 모형이 최종 모형으로 선택됩니다 . *Y* 가 명목형이면 최량 검증 오분류 비율을 제공하는 모형이 최종 모형으로 선택됩니다 .

검증 열 (시작 창에서 " 검증 " 열을 지정한 경우에만 사용 가능) 지정된 " 검증 " 열의 값을 사용하여 데이터를 여러 부분으로 나눕니다 . 열의 값은 데이터가 분할되는 방법과 검증에 사용되는 방법을 결정합니다 .

- 두 개의 값이 있는 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 세 개의 값이 있는 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 검증 열에 네 개 이상의 수준이 있는 경우 " 검증 열 K 폴드 " 가 사용됩니다 .

SVM 플랫폼에서는 " 설계 조정 " 이 사용되는 경우 외에는 검증 열을 사용하여 모형을 훈련하고 평가합니다 . " 설계 조정 " 옵션을 선택하는 경우 SVM 플랫폼에서는 검증 열을 사용하여 모형을 훈련 및 조정하거나 모형을 훈련 , 조정 및 평가합니다 . 검증에 대한 자세한 내용은 "JMP 모델링의 검증 " 에서 확인하십시오 .

참고 : 검증 열을 사용하여 데이터가 올바르게 분할되지 않으면 홀드백 검증 방법이 대신 사용됩니다 .

검증 열 K 폴드 ("Y, 반응 " 열에 정확히 두 개의 수준이 있고 시작 창에서 " 검증 " 열을 지정한 경우에만 사용 가능) 지정된 " 검증 " 열의 값을 사용하여 데이터를 *K* 개의 집합으로 나눕니다 . 여기서 *K* 는 열의 고유 값 수입니다 . 그런 다음 K 폴드 검증이 수행됩니다 .

없음 검증이 사용되지 않습니다 .

시작 지정된 SVM 모형을 적합시키고 모형 보고서를 표시합니다.

참고 : 데이터 테이블이 큰 경우 데이터에 적합되는 각 모형에 대해 진행률 표시줄이 나타납니다. 총 적합 모형 수는 $k!/2(k\text{-}2)!$ 이며, 여기서 k 는 반응 변수의 수준 수입니다. 각 진행률 표시줄에는 **현재 추정값 채택** 버튼이 있습니다. 적합 알고리즘을 조기에 중지하고 현재 추정값을 채택하려면 이 버튼을 클릭합니다. 이 버튼을 클릭한 후 예측 계산이 수행되므로 보고서가 나타나는 데 시간이 걸릴 수 있습니다.

결측값

예측 변수에 대해 결측값이 있는 행은 SVM 모델링 절차에 포함되지 않습니다. 따라서 데이터 테이블에 저장된 모든 열은 해당 행에 결측값을 포함합니다. 결측값이 있는 데이터를 SVM 모형에 포함하려면 전처리 작업이 필요합니다. 자세한 내용은 "결측값 탐색 유틸리티"에서 확인하십시오.

JMP PRO 서포트 벡터 머신 보고서

"모형 시작" 제어판에서 **시작**을 클릭하면 제어판 위에 "모형 비교" 보고서가 나타나고 제어판 아래에 모형 보고서가 나타납니다. 모형 보고서에는 순차적으로 번호가 지정됩니다. 새 모형이 적합될 때마다 "모형 비교" 보고서에 추가되고 새 모형 보고서가 창에 추가됩니다. 자세한 내용은 "모형 비교 보고서" 및 "서포트 벡터 머신 모형 보고서"에서 확인하십시오. "모형 비교" 보고서의 열을 기준으로 모형을 정렬하려면 원하는 열 머리글을 클릭합니다. 처음 클릭하면 오름차순으로 정렬되고 해당 열 머리글을 두 번째 클릭하면 내림차순으로 정렬됩니다.

JMP PRO 모형 비교 보고서

"모형 비교" 보고서에는 다음 정보가 포함되어 있습니다.

표시 해당 적합에 대한 "서포트 벡터 머신 모형" 보고서를 보고서 창에 표시할지 여부를 나타내는 체크박스입니다.

방법 모형 번호입니다.

커널 함수 해당 모형의 커널 함수입니다.

비용 해당 모형의 비용 모수 값입니다.

감마 해당 모형의 감마 모수 값입니다. 모형에 선형 커널이 사용되는 경우 이 값은 결측입니다.

SV 수 해당 모형에 사용된 서포트 벡터 수입니다.

훈련 오분류 비율 (반응이 범주형인 경우에만 표시됨) 훈련 데이터 집합의 관측값에 대한 오분류 비율입니다. 이 비율은 플랫폼에서 계산되는 분류 결정 규칙을 기반으로 합니다.

검증 오분류 비율 (반응이 범주형이고 검증을 사용하는 경우에만 표시됨) 검증 데이터 집합의 관측값에 대한 오분류 비율입니다. 이 비율은 플랫폼에서 계산되는 분류 결정 규칙을 기반으로 합니다.

테스트 오분류 비율 (반응이 범주형이고 검증 열을 사용하여 테스트 데이터 집합을 지정한 경우에만 표시됨) 테스트 데이터 집합의 관측값에 대한 오분류 비율입니다. 이 비율은 플랫폼에서 계산되는 분류 결정 규칙을 기반으로 합니다.

훈련 RASE (반응이 연속형인 경우에만 표시됨) 훈련 데이터 집합의 관측값에 대한 평균 제곱 예측 오차의 제곱근 (제곱근 평균 제곱 오차) 입니다. 자세한 내용은 "RASE" 에서 확인하십시오.

검증 RASE (반응이 연속형이고 검증을 사용하는 경우에만 표시됨) 검증 데이터 집합의 관측값에 대한 평균 제곱 예측 오차의 제곱근 (제곱근 평균 제곱 오차) 입니다. 자세한 내용은 "RASE" 에서 확인하십시오.

테스트 RASE (반응이 연속형이고 검증 열을 사용하여 테스트 데이터 집합을 지정한 경우에만 표시됨) 테스트 데이터 집합의 관측값에 대한 평균 제곱 예측 오차의 제곱근 (제곱근 평균 제곱 오차) 입니다. 자세한 내용은 "RASE" 에서 확인하십시오.

검증 일반화 R^2 (검증을 사용하는 경우에만 표시됨) 검증 데이터 집합의 관측값에 대한 일반화 R^2 값입니다. 자세한 내용은 " 적합 상세 정보 " 에서 확인하십시오.

확률 임계 (반응이 이항이고 검증을 사용하는 경우에만 표시됨) 플랫폼에서 계산되는 분류 결정 규칙에 의해 결정된 경계 확률입니다. 관측값은 예측 확률이 이 값을 초과할 경우 목표 수준으로 분류됩니다.

팁 : " 혼동 행렬 " 보고서의 " 확률 임계 설정 " 옵션을 사용하여 이 값을 변경할 수 있습니다. 자세한 내용은 " 혼동 행렬 " 에서 확인하십시오.

조건부 검증 오분류 비율 (반응이 이항이고 검증을 사용하는 경우에만 표시됨) 확률 임계값을 조건으로 하는 검증 데이터 집합의 관측값에 대한 오분류 비율입니다.

최적 오분류 비율이 가장 작은 모형 적합을 나타냅니다. 테스트 데이터 집합 없이 검증을 사용하는 경우 검증 오분류 비율이 가장 작은 모형입니다. 테스트 데이터 집합과 함께 검증을 사용할 경우에는 테스트 오분류 비율이 가장 작은 모형입니다.

모형 비교 그림

전체 모수 값 격자에서 검증 데이터 집합의 모형 성능을 나타내는 등고선 그림을 표시합니다. 이 보고서는 "모형 시작" 제어판에서 "설계 조정"을 지정한 경우에만 사용할 수 있습니다.

JMP PRO 서포트 벡터 머신 모형 보고서

반응 프로파일 그림 (반응이 범주형인 경우에만 사용 가능) 분류 모형의 시각적 표현을 제공합니다. 그림의 점은 실제 데이터 관측값이며 정확히 두 개의 연속형 변수가 있는 경우에만 기본적으로 표시됩니다. 표시된 두 변수의 경우 음영 등고선은 나머지 모형 요인의 고정 값에 의해 결정된 예측 공간의 평면을 나타냅니다. 예측은 플랫폼에서 계산되는 분류 결정 규칙을 기반으로 합니다. 고정 값에 대한 컨트롤은 그림 위에 있습니다. 슬라이더 또는 숫자 상자를 사용하여 요인의 고정 값을 변경하면 표시된 변수의 예측 공간이 자동으로 업데이트

됩니다 . 각 축의 빨간색 삼각형을 사용하면 그림의 축을 변경하여 연속형 요인을 표시할 수도 있습니다 .

" 반응 프로파일 그림 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

격자 밀도 음영 등고선의 기반이 되는 예측 격자의 정밀도를 결정합니다 . 격자 밀도가 높을수록 더 매끄러운 결정선이 표시됩니다 .

점 표시 그림에 점을 표시하거나 숨깁니다 . 변수가 두 개뿐인 경우 기본적으로 선택되어 있습니다 .

모형 요약 반응 열 이름 , 검증 방법 및 모형 적합에 사용된 커널 함수 유형을 제공합니다 . " 모형 요약 " 테이블에는 훈련 , 검증 및 테스트 데이터 집합에 대한 모형 적합 정보도 포함되어 있습니다 . 각 집합에 대해 관측값 수와 서포트 벡터 수가 보고됩니다 . 반응이 범주형이면 각 집합에 대해 오분류 비율이 보고됩니다 . 오분류 비율은 모형에 의해 잘못 분류된 관측값의 비율입니다 . 이 값은 오분류 수를 총 관측값 수로 나누어 계산됩니다 . 반응이 연속형이면 각 집합에 대해 RASE 및 R^2 값이 보고됩니다 .

추정 상세 정보 모형에 사용된 모수의 값을 제공합니다 .

적합 상세 정보 (반응이 범주형인 경우에만 사용 가능) 훈련 데이터 집합과 검증 및 테스트 데이터 집합 (지정된 경우) 에 대한 다음 통계량을 제공합니다 .

엔트로피 R^2 적합 모형과 상수 확률 모형의 로그 가능도를 비교하는 적합 측도입니다 . 엔트로피 R^2 범위는 0 ~ 1 이며 , 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 . 자세한 내용은 " 엔트로피 R^2" 에서 확인하십시오 .

일반화 R^2 일반 회귀 모형에 적용할 수 있는 측도입니다 . 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다 . 일반화 R^2 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2 로 단순화됩니다 . 일반화 R^2 을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데 , 이는 Cox-Snell 유사 R^2 을 정규화한 것입니다 . 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오 . 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 .

평균 -Log p -log(p) 의 평균입니다 . 여기서 p 는 발생한 사건과 연관된 적합 확률입니다 . 값이 작을수록 더 나은 적합을 나타냅니다 .

RASE 평균 제곱 예측 오차의 제곱근입니다 (제곱근 평균 제곱 오차). RASE 는 다음과 같이 계산되며 , 여기서 Source 는 훈련 , 검증 또는 테스트 데이터 집합을 나타냅니다 .

$$\text{RASE}_{Source} = \sqrt{\frac{SSE_{Source}}{n}}$$

평균 절대 편차 반응과 예측 반응 사이의 차이에 대한 절대값의 평균입니다 . 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 . 값이 작을수록 더 나은 적합을 나타냅니다 .

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다 .

참고 :" 적합 상세 정보 " 보고서의 오분류 비율이 " 혼동 행렬 " 보고서의 오분류 비율과 일치하지 않을 수 있습니다 . 이항 반응의 경우 " 적합 상세 정보 " 보고서의 비율은 확률 경계 0.5 를 사용하지만 " 혼동 행렬 " 보고서의 비율은 확률 임계값을 경계로 사용합니다 .

N 관측값 수입니다 .

혼동 행렬 (반응이 범주형인 경우에만 사용 가능) 모형에 지정된 각 데이터 집합 (훈련 , 검증 및 테스트) 에 대한 혼동 행렬이 표시됩니다 . 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다 . 혼동 행렬과 오분류 비율을 사용하여 모형을 평가할 수 있습니다 .

혼동 행렬과 오분류 비율은 확률 임계 상자의 값을 경계 값으로 사용합니다 . 기본적으로 이 값은 플랫폼에서 계산되는 분류 결정 규칙을 기반으로 합니다 . 슬라이더를 드래그하거나 " 확률 임계 " 옆의 상자에 새 값을 입력하여 경계 값을 변경할 수 있습니다 . 확률 임계를 변경하면 혼동 행렬과 오분류 비율이 자동으로 업데이트됩니다 . " 모형 비교 " 보고서의 확률 임계 및 조건부 검증 오분류 비율 열도 업데이트됩니다 .

실제값 대 예측값 그림 (반응이 연속형인 경우에만 사용 가능) 실제 반응 대 예측 반응 그림을 표시합니다 . 대각선은 Y = X 선입니다 . 점이 이 선에 가까울수록 모형이 데이터를 더 잘 적합시킵니다 . 검증이 사용되면 훈련 및 검증 데이터 집합과 테스트 데이터 집합 (지정된 경우) 에 대해 그림이 표시됩니다 .

JMP PRO 서포트 벡터 머신 플랫폼 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO 서포트 벡터 머신 모형 보고서 옵션

각 " 서포트 벡터 머신 모형 " 보고서에는 다음 옵션이 포함되어 있습니다 .

반응 프로파일 그림 (반응이 범주형인 경우에만 사용 가능) 반응 프로파일 그림을 표시하거나 숨깁니다 .

서포트 벡터 계수 서포트 벡터 계수 테이블을 표시하거나 숨깁니다 . 이 테이블에서 행을 선택하면 원래 데이터 테이블과 반응 프로파일 그림에서 서포트 벡터가 강조 표시됩니다 .

혼동 행렬 (반응이 범주형인 경우에만 사용 가능) 혼동 행렬을 표시하거나 숨깁니다 . 자세한 내용은 " 혼동 행렬 " 에서 확인하십시오 .

ROC 곡선 (반응이 범주형인 경우에만 사용 가능) ROC 곡선을 표시하거나 숨깁니다 . 검증을 사용하는 경우 훈련 데이터 집합에 대한 곡선 외에 검증 데이터 집합 및 테스트 데이터 집합 (지정된 경우) 에 대한 ROC 곡선이 나타납니다 . 자세한 내용은 "ROC 곡선 " 에서 확인하십시오 .

향상도 곡선 (반응이 범주형인 경우에만 사용 가능) 향상도 곡선을 표시하거나 숨깁니다 . 검증을 사용하는 경우 훈련 데이터 집합에 대한 곡선 외에 검증 데이터 집합 및 테스트 데이터 집합 (지정된 경우) 에 대한 향상도 곡선이 나타납니다 . 자세한 내용은 " 향상도 곡선 " 에서 확인하십시오 .

실제값 대 예측값 그림 (반응이 연속형인 경우에만 사용 가능) 실제값 대 예측값 그림을 표시하거나 숨깁니다 .

잔차 대 예측값 그림 (반응이 연속형인 경우에만 사용 가능) 잔차 대 예측값 그림을 표시하거나 숨깁니다 .

프로파일러 예측 프로파일러를 표시하거나 숨깁니다 . 각 반응 수준은 예측 프로파일러에서 별도의 행으로 표시됩니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

등고선 프로파일러 (모형에 두 개 이상의 연속형 요인이 포함된 경우에만 사용 가능) 등고선 프로파일러를 표시하거나 숨깁니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

표면 프로파일러 (모형에 하나 이상의 연속형 요인이 포함된 경우에만 사용 가능) 표면 프로파일러를 표시하거나 숨깁니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

예측값 저장 임계 확률 경계 값에 기반한 예측 분류를 데이터 테이블의 <Y, 반응 > 예측값 열에 저장합니다 .

예측 계산식 저장 임계 확률 경계 값에 기반한 예측 계산식을 데이터 테이블에 저장합니다 . 반응이 이항이면 단일 예측 계산식 열이 데이터 테이블에 추가됩니다 . 반응 수준이 세 개 이상이면 최종 예측에 대한 계산식 열 외에 각 쌍별 모형에 대한 예측 계산식 열이 데이터 테이블에 추가됩니다 .

예측 계산식 게시 임계 확률 경계 값에 기반한 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않은 경우 이 옵션을 선택하면 " 계산식 저장소 " 창이 열립니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

확률 저장 최대 확률 분류 <Y, 반응 > 열 및 각 반응 수준의 확률을 데이터 테이블에 개별 열로 저장합니다 .

확률 계산식 저장 (반응이 이항인 경우에만 사용 가능) 다음 계산식 열을 데이터 테이블에 저장합니다 .

결정 값 결정 값에 대한 계산식을 포함하는 숨겨진 열입니다 .

확률 이항 반응의 각 수준에 대해 하나씩 총 두 개의 확률 계산식 열입니다 .

최대 확률 분류 확률 값이 더 높은 예측 분류입니다 .

임계 예측 분류 혼동 행렬에서 얻은 임계 확률 경계 값에 기반한 예측 분류입니다 . 기본적으로 이 값은 플랫폼에서 계산되는 분류 결정 규칙을 기반으로 합니다 . 임계값을 0.5 로 변경할 경우 **임계 예측 분류** 열은 **최대 확률 분류** 열과 동일하므로 데이터 테이블에 저장되지 않습니다 .

확률 계산식 게시 (반응이 이항인 경우에만 사용 가능) 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않은 경우 이 옵션을 선택하면 " 계산식 저장소 " 창이 열립니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

검증 저장 (" 홀드백 " 또는 "K 폴드 " 검증 방법을 사용하는 경우에만 사용 가능) 훈련 데이터 집합과 검증 데이터 집합에 사용된 행을 식별하는 새 열을 데이터 테이블에 생성합니다 .

적합 제거 전체 모형 보고서를 제거합니다 .

SVM 플랫폼의 추가 예

연속형 반응에 대한 서포트 벡터 회귀의 예

SVR(서포트 벡터 회귀) 모형은 SVM 알고리즘을 사용하여 연속형 반응 데이터를 예측합니다 . 이 예에서는 " 서포트 벡터 머신의 예 " 에서 사용된 것과 동일한 당뇨병 환자 의료 데이터를 사용합니다 . 단 , 당뇨병 진행 경과의 이진 측정값 대신 연속형 측정값이 제공됩니다 . 병이 많이 진행될수록 값이 높습니다 . 방사 기저 함수의 비용 모수를 변경하여 최량 적합 모형을 찾을 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > 서포트 벡터 머신**을 선택합니다 .
3. Y 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Age ~ Glucose 를 선택하고 **X, 요인**을 클릭합니다 .
5. Validation 을 선택하고 **검증**을 클릭합니다 .
6. **확인**을 클릭합니다 .
7. " 모형 시작 " 제어판에서 커널 함수가 " 방사 기저 함수 " 이고 " 비용 " 모수가 1, " 감마 " 모수가 0.1 인지 확인합니다 .
8. **시작**을 클릭합니다 .
9. " 모형 시작 " 옆의 회색 삼각형을 클릭하여 " 모형 시작 " 제어판을 엽니다 .

10. 비용 모수를 0.1 로 변경합니다 .
11. **시작**을 클릭합니다 .
12. " 모형 시작 " 옆의 회색 삼각형을 클릭하여 " 모형 시작 " 제어판을 엽니다 .
13. 비용 모수를 2 로 변경합니다 .
14. **시작**을 클릭합니다 .

그림 9.6 모형 비교 보고서

모형 비교

표시	방법	커널 함수	비용	감마	SV 수	훈련 RASE	검증 RASE	검증 일반화 R^2	최선
☑	모형 1	방사 기저 함수	1	0.1	266	43.8147	58.3285	0.42552	
☑	모형 2	방사 기저 함수	0.1	0.1	280	52.5804	57.5025	0.44167	최소 RASE 최대 R^2
☑	모형 3	방사 기저 함수	2	0.1	265	40.553	60.7801	0.37621	

모형 비교 그림

검증 열: Validation

" 모형 비교 " 보고서에는 서로 다른 비용 모수를 사용한 모형의 결과가 포함되어 있습니다 . 비용 모수가 높을수록 점을 잘못 분류할 가능성이 낮은 알고리즘이 구현되는 반면 , 비용 모수가 낮을수록 더 유연한 알고리즘이 구현됩니다 . 이 시나리오에서는 더 유연한 모형 , 즉 비용 모수가 0.1 인 모형이 최적 모형 적합을 생성합니다 .

15. " 모형 비교 " 보고서의 " 표시 " 열에서 " 모형 1" 과 " 모형 3" 을 선택 취소합니다 .

그림 9.7 연속형 반응에 대한 모형 보고서

서포트 벡터 머신 모형 2

모형 요약

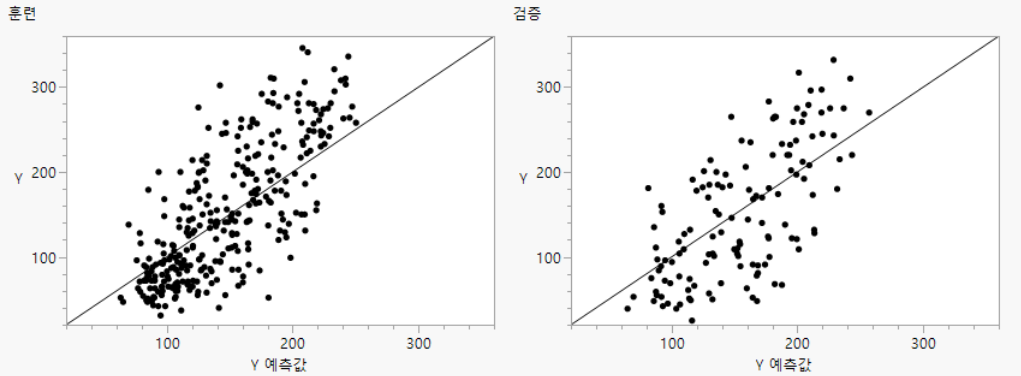
응답 Y
검증 방법 검증 열
커널 함수 방사 기저 함수

측도	훈련	검증
행 수	309	133
빈도 합	309	133
RASE	52.58037	57.502491
R^2	0.5340282	0.4416728
서포트 벡터 수	280	280

추정 상세 정보

비용 0.1
감마 0.1

실제값 대 예측값 그림



연속형 반응에 대한 모형 보고서에는 " 모형 요약 ", " 추정 상세 정보 " 및 " 실제값 대 예측값 그림 " 이 포함됩니다 . " 모형 요약 " 보고서에서 훈련 데이터 집합의 RASE 는 52.58, 검증 데이터 집합의 RASE 는 57.50 입니다 .

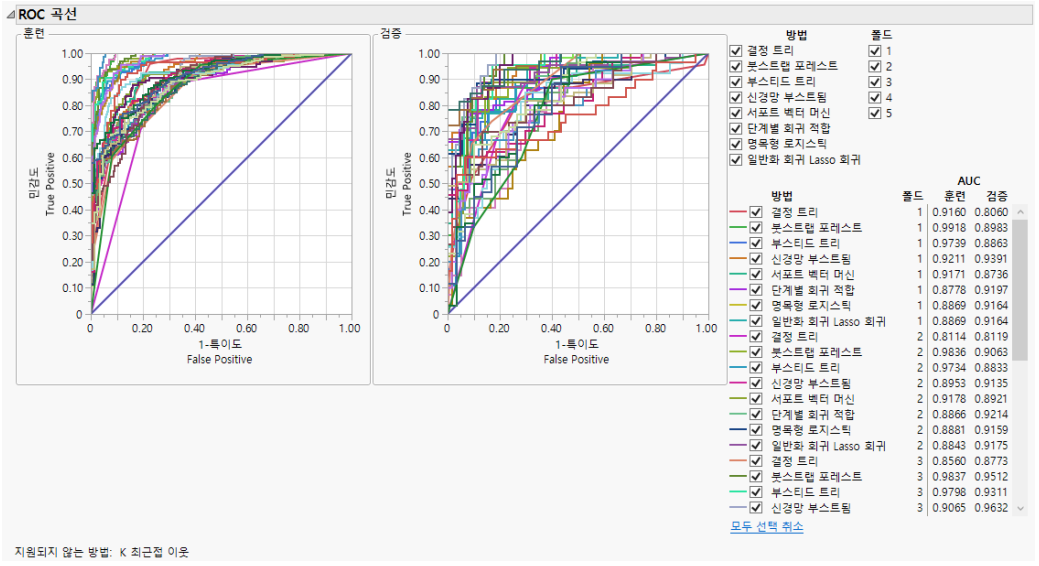
10 장

JMP PRO 모형 선별
다중 예측 모형 적합 및 비교

모형 선별 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

JMP Pro의 모형 선별 플랫폼을 사용하면 하나의 시작 창에서 여러 예측 모델링 플랫폼을 실행하고 다양한 방법의 요약을 조합할 수 있습니다. 이렇게 하면 개별 모형 플랫폼을 시작하지 않고도 최적 성능 모형을 선택할 수 있습니다. "모형 선별" 보고서 창에서 원하는 모형의 개별 플랫폼을 손쉽게 실행하여 보다 세분화할 수도 있습니다.

그림 10.1 모형 선별 보고서



목차

연속형 반응에 대한 모형 선별의 예 159
모형 선별 플랫폼 시작 160
모형 선별 보고서 163
폴드 간 요약 보고서 164
훈련, 검증 및 테스트 적합 측도 165
모형 선별 플랫폼 옵션 166
결정 임계값 보고서 168

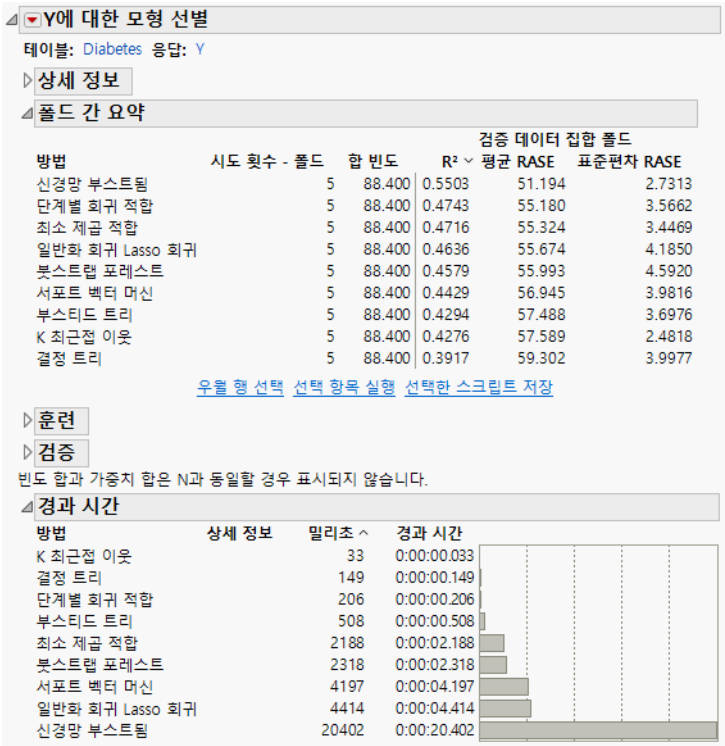
JMP PRO 연속형 반응에 대한 모형 선별의 예

442 명의 당뇨병 환자에 대한 기준 의료 데이터가 있습니다 . 또한 각 환자가 병원에 처음 내원한 후 1 년 동안 얻은 당뇨병 진행 경과의 연속형 측정값도 있습니다 . 병이 많이 진행될수록 값이 높습니다 . 당뇨병 진행 경과를 예측하기 위해 다양한 모형의 성능을 평가하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > 모형 선별**을 선택합니다 .
3. Y 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Age ~ Glucose 를 선택하고 **X, 요인**을 클릭합니다 .
5. " 폴드 교차 검증 " 섹션에서 **K 폴드 교차 검증** 옆의 상자를 선택합니다 .
6. (선택 사항) " 옵션 " 섹션에서 **난수 시드값 설정** 옆에 "42920" 을 입력합니다 .

   예제 결과와 일치시키려면 난수 시드값을 사용하십시오 .
7. **확인**을 클릭합니다 .
8. "Y 에 대한 모형 선별 " 옆의 빨간색 삼각형을 클릭하고 **선택적 보고서 > 경과 시간**을 선택합니다 .

그림 10.2 연속형 반응에 대한 모형 보고서



Y에 대한 모형 선별

데이블: Diabetes **응답:** Y

상세 정보

폴드 간 요약

방법	시도 횟수 - 폴드	합 빈도	R^2	검증 데이터 집합 폴드 평균 RASE	검증 데이터 집합 폴드 표준편차 RASE
신경망 부스트됨	5	88.400	0.5503	51.194	2.7313
단계별 회귀 적합	5	88.400	0.4743	55.180	3.5662
최소 제곱 적합	5	88.400	0.4716	55.324	3.4469
일반화 회귀 Lasso 회귀	5	88.400	0.4636	55.674	4.1850
붓스트랩 포레스트	5	88.400	0.4579	55.993	4.5920
서포트 벡터 머신	5	88.400	0.4429	56.945	3.9816
부스티드 트리	5	88.400	0.4294	57.488	3.6976
K 최근접 이웃	5	88.400	0.4276	57.589	2.4818
결정 트리	5	88.400	0.3917	59.302	3.9977

우열 행 선택 선택 항목 실행 선택한 스크립트 저장

훈련

검증

빈도 합과 가중치 합은 N과 동일할 경우 표시되지 않습니다.

경과 시간

방법	상세 정보	밀리초	경과 시간
K 최근접 이웃		33	0:00:00.033
결정 트리		149	0:00:00.149
단계별 회귀 적합		206	0:00:00.206
부스티드 트리		508	0:00:00.508
최소 제곱 적합		2188	0:00:02.188
붓스트랩 포레스트		2318	0:00:02.318
서포트 벡터 머신		4197	0:00:04.197
일반화 회귀 Lasso 회귀		4414	0:00:04.414
신경망 부스트됨		20402	0:00:20.402

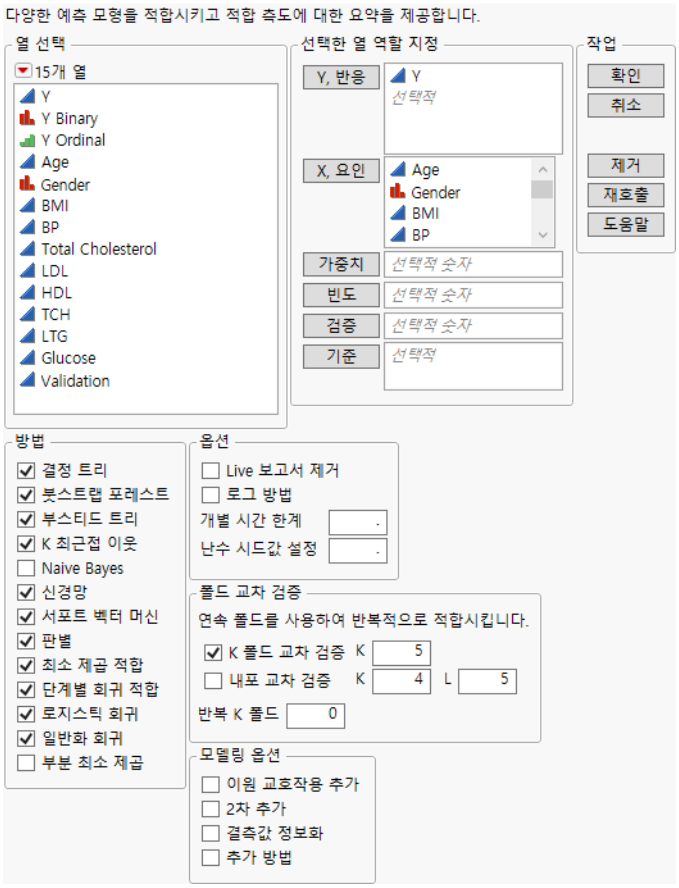
전체 폴드에서 평균화된 검증 R^2을 기반으로 하는 최적 성능 모형은 "부스트된 신경망"입니다. "부스트된 신경망"의 평균 검증 R^2은 0.5503입니다. "경과 시간" 보고서에 따르면 "부스트된 신경망" 모형이 적합 시간도 가장 오래 걸렸습니다. 이 경우 경과 시간은 5초에 불과하지만 데이터 집합이 크거나 모형이 복잡한 경우에는 이 정보가 중요할 수 있습니다.

팁 : 개별 모형의 성능 측정 기준을 보려면 " 검증 " 옆의 회색 표시 아이콘을 클릭하십시오 .

JMP PRO 모형 선별 플랫폼 시작

모형 선별 플랫폼을 시작하려면 **분석 > 예측 모델링 > 모형 선별**을 선택합니다 .

그림 10.3 모형 선별 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 .

X, 요인 예측 변수입니다 .

가중치 (K 최근접 이웃 , 서포트 벡터 머신 또는 신경망 모델링 플랫폼에는 적용할 수 없음) 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다 .

빈도 (K 최근접 이웃 모델링 플랫폼에는 적용할 수 없음) 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다 .

검증 (시작 창에서 " 교차 검증 " 옵션을 선택한 경우에는 적용할 수 없음) 검증 데이터 집합을 정의하는 숫자 열입니다 . " 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다 . " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 " 검증 열 생성 " 에서 확인하십시오 .

참고 : 네 개 이상의 수준이 있는 검증 열을 지정하는 경우 이 열은 K 폴드 교차 검증을 수행하는 데 사용됩니다 .

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다 . 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다 . 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

방법 원하는 모델링 플랫폼을 선택할 수 있습니다 . 기본적으로 적합되는 모델링 플랫폼은 " 결정 트리 (파티션)", " 붓스트랩 포레스트 ", " 부스티드 트리 ", "K 최근접 이웃 ", " 신경망 ", " 서포트 벡터 머신 ", " 판별 ", " 최소 제곱 적합 ", " 단계별 회귀 적합 ", " 로지스틱 회귀 " 및 "일반화 회귀"입니다 . "Naive Bayes", "부분 최소 제곱" 및 "XGBoost"도 사용할 수 있습니다 .

참고 :

- XGBoost는 JMP에서 지원되지 않으며 XGBoost 추가기능이 설치된 경우에만 사용할 수 있습니다 . XGBoost 에 대한 자세한 내용은 https://community.jmp.com 에서 확인하십시오 .
- 결정 트리 (파티션), 판별 및 부분 최소 제곱 방법 모두 모형을 적합시키려면 특정 유형의 검증 데이터 집합이 필요합니다 .
- 검증 데이터 집합의 관측값이 20 개 미만인 경우 결정 트리 (파티션) 모형을 적합시킬 수 없습니다 .
- 모델링 플랫폼에서는 기본 옵션 및 조정 모수를 사용하여 모형을 적합시킵니다 . 플랫폼을 직접 호출하고 다른 옵션을 선택하여 기본 설정을 사용할 때보다 적합성을 향상시킬 수 있습니다 .

옵션

추가 옵션을 제공합니다 .

Live 보고서 제거 " 모형 선별 " 보고서 창에 개별 모형 플랫폼 보고서를 포함하지 않습니다 .

팁 : 많은 방법과 적합이 포함된 큰 문제의 경우 이 옵션을 선택하여 메모리를 확보할 수 있습니다 .

로그 방법 적합 플랫폼이 호출될 때마다 진행 메시지를 로그에 기록합니다 .

개별 시간 한계 각 적합에 대한 시간 한계 (초) 를 지정합니다 . 조기 중지를 지원하는 플랫폼의 경우 해당 시점까지의 최적 추정값이 제공됩니다. 조기 중지를 지원하지 않는 플랫폼의 경우 결과가 제공되지 않습니다 .

난수 시드값 설정 모형 적합 루틴의 랜덤 성분에 사용되는 난수 시드값을 설정합니다 . 이렇게 하면 플랫폼을 다시 실행하고 동일한 모형 적합을 얻을 수 있습니다 .

폴드 교차 검증

다양한 유형의 교차 검증에 대한 옵션을 제공합니다.

K 폴드 교차 검증 데이터를 무작위로 *K* 개의 부분 또는 폴드로 나눕니다 . 각 모형이 데이터에 *K* 번 적합되고 , 매번 교차 검증 집합으로 홀드아웃된 서로 다른 폴드를 사용합니다 . 총 *K* 개 모형이 적합됩니다 . *K* 의 기본값은 5 입니다 .

- *K* 는 K 폴드 교차 검증을 위한 폴드 수를 지정합니다 . 기본값은 5 이고 *K* 는 1 보다 커야 합니다 .

내포 교차 검증 교차 검증을 위해 데이터를 내포된 폴드로 나눕니다 . 먼저 데이터를 $k = 1, ..., K$ 개의 동일한 부분 또는 **폴드**로 나눕니다 . 각 폴드에 대해 *k* 번째 폴드는 테스트 데이터 집합으로 사용되고 , 나머지 데이터는 *L* 개의 동일한 부분으로 추가 분할됩니다 . 이러한 *L* 개의 구분을 **내측 폴드**라고 합니다 . 그런 다음 모형이 데이터에 *L* 번 적합되고 , 매번 교차 검증 집합으로 홀드아웃된 다른 내측 폴드가 사용됩니다 . 그런 다음 *L* 개의 모형이 *k* 번째 폴드를 공통 테스트 데이터 집합으로 사용합니다 . 총 *K***L* 개 모형이 적합됩니다 . *K* 의 기본값은 4 이고 *L* 의 기본값은 5 입니다 .

예를 들어 K = 2, L = 3 으로 설정합니다 . 처음에는 데이터가 두 개의 폴드로 구분됩니다 . 첫 번째 폴드는 테스트 데이터 집합으로 홀드아웃되고 두 번째 폴드는 3 개의 내측 폴드로 구분됩니다 . 세 개의 모형이 데이터에 적합되고 , 매번 교차 검증 집합으로 홀드아웃된 서로 다른 내측 폴드가 사용됩니다 . 그러면 세 개 모형이 모두 첫 번째 폴드에서 테스트됩니다 .

그런 다음 두 번째 폴드가 테스트 데이터 집합으로 홀드아웃되고 첫 번째 폴드는 3 개의 내측 폴드로 구분됩니다 . 세 개의 모형이 데이터에 적합되고 , 매번 교차 검증 집합으로 홀드아웃된 서로 다른 내측 폴드가 사용됩니다 . 그러면 세 개 모형이 모두 두 번째 폴드에서 테스트됩니다 .

- *K* 는 내포 교차 검증을 위한 폴드 수를 지정합니다 . 기본값은 4 이고 *K* 는 1 보다 커야 합니다 .
- *L* 은 내포 교차 검증을 위한 내측 폴드 수를 지정합니다 . 기본값은 5 이고 *L* 은 1 보다 커야 합니다 .

참고 : "K 폴드 교차 검증 " 과 " 내포 교차 검증 " 을 모두 선택하면 " 내포 교차 검증 " 이 수행됩니다 .

반복 K 폴드 K 폴드 교차 검증 또는 내포 교차 검증 프로세스의 반복 횟수를 지정합니다 .

모델링 옵션

모델링 플랫폼에 대한 추가 옵션을 제공합니다.

이원 교호작용 추가 모든 이원 교호작용 효과를 선형 모형에 추가합니다 .

2 차 추가 연속형 변수의 제곱에 대한 효과를 선형 모형에 추가합니다 .

결측값 정보화 모든 플랫폼에 대해 결측값 정보화를 활성화합니다 .

추가 방법 일반화 회귀 플랫폼에서 능형 회귀 , Elastic Net 및 Lasso 회귀와 같은 몇 가지 추가 방법을 호출합니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

주의 : 이로 인해 추가 모형 적합이 생성됩니다 .

"확인"을 클릭하면 지정된 모형이 적합되고 진행률 표시줄 집합이 표시됩니다. 위쪽 진행률 표시줄은 모든 적합에 대한 진행률을 보고합니다. 아래쪽 진행률 표시줄은 현재 개별 모형 적합에 대한 진행률을 보고합니다. 아래쪽 진행률 표시줄은 조기 중지를 사용하여 중지할 수 있으며 위쪽 진행률 표시줄은 계속 실행됩니다.

JMP PRO 모형 선별 보고서

"모형 선별" 보고서의 내용은 구현된 교차 검증 유형과 반응 변수의 모델링 유형에 따라 달라집니다. "모형 선별" 보고서 맨 위에 데이터 테이블 이름과 반응 변수의 열 이름이 표시됩니다. 검증 열을 지정한 경우 해당 열 이름도 표시됩니다.

기본적으로 "상세 정보" 보고서가 다음에 나타납니다. 이 보고서에는 각 폴드 및 시행에 대한 모든 모형의 개별 모형 보고서가 포함되어 있습니다. 이러한 보고서는 각 플랫폼의 실제 출력이며 모든 보고서 및 명령 기능을 포함합니다. 예를 들어 최적 모형을 결정한 경우 "상세 정보" 보고서의 해당 플랫폼 보고서로 이동하여 원하는 예측 계산식 열을 저장할 수 있습니다. 시작 창에서 "Live 보고서 제거" 옵션을 선택하여 "상세 정보" 보고서를 숨길 수 있습니다.

시작 창에서 "K 폴드 교차 검증" 옵션 또는 "내포 교차 검증" 옵션을 지정한 경우 "폴드 간 요약" 보고서가 나타납니다. 이 보고서는 테스트 폴드에서 평균화된 적합 측도를 요약한 것입니다.

각 개별 모형 적합, 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합(해당하는 경우)에 대한 적합 측도가 표시됩니다. 시작 창에서 검증 열을 지정하여 교차 검증이 사용된 경우 개별 모형 적합 측도만 보고서에 표시됩니다.

폴드 간 요약 보고서

" 폴드 간 요약 " 보고서에는 전체 폴드에 대한 적합 측도와 " 반복 K 폴드 " 가 지정된 경우 전체 시행에 대한 적합 측도가 요약되어 있습니다 . "K 폴드 교차 검증 " 옵션을 지정한 경우 검증 데이터 집합 전체에 대해 적합 측도가 요약됩니다 . " 내포 교차 검증 " 옵션을 지정한 경우에는 테스트 데이터 집합 전체에 대해 적합 측도가 요약됩니다 . 이 보고서에는 다음 열을 포함하는 테이블이 있습니다 .

방법 모형을 적합시키는 데 사용된 방법의 이름입니다 .

시행 횟수 - 폴드 모든 폴드 및 시행 (해당하는 경우) 에 대해 적합된 총 모형 수입니다 .

빈도 합 검증 데이터 집합 또는 테스트 데이터 집합의 평균 관측값 수입니다 . 테스트 데이터 집합은 모형 성능을 추정하는 데 사용됩니다 .

R^2 모든 검증 데이터 집합 폴드 또는 테스트 데이터 집합 폴드에 대한 평균 R^2 입니다 . 반응이 범주형인 경우 엔트로피 R^2 이 이 열에 포함됩니다 .

평균 RASE 모든 검증 데이터 집합 폴드 또는 테스트 데이터 집합 폴드에 대한 평균 RASE(제곱근 평균 제곱 오차) 입니다 .

표준편차 RASE 모든 검증 데이터 집합 또는 테스트 데이터 집합에 대한 RASE의 표준편차입니다.

평균 AUC (범주형 반응에만 사용 가능) 모든 검증 데이터 집합 또는 테스트 데이터 집합에 대한 평균 AUC(ROC 곡선 아래 면적) 입니다 .

평균 MR (범주형 반응에만 사용 가능) 모든 검증 데이터 집합 또는 테스트 데이터 집합에 대한 평균 MR(오분류 비율) 입니다 .

테이블 아래에 다음과 같은 옵션이 있습니다 .

우월 행 선택 모형 적합 기준 조합에 따라 다른 모든 모형보다 우수하거나 동일한 각 모형을 선택합니다 . 이를 파레토 프론티어 선택이라고도 합니다 . 연속형 반응의 경우 우월 모형을 결정할 때 R^2 과 빈도 합을 고려합니다 . 범주형 반응의 경우 우월 모형을 결정할 때 엔트로피 R^2, 오분류 비율 , AUC 및 빈도 합을 고려합니다 .

선택 항목 실행 선택한 각 행에 대해 개별 모형을 실행합니다 . 시작 창에서 폴드 교차 검증 유형을 지정한 경우 최종 모형에 해당하는 폴드 , 내측 폴드 및 시행 조합을 사용하여 모형이 실행됩니다 . 최종 모형은 가중 평균 R^2 이 가장 높은 모형입니다 . 가중 평균 R^2 은 훈련 R^2, 검증 R^2 및 테스트 R^2(테스트 데이터 집합이 사용된 경우) 의 관측값 수에 따라 가중치가 적용된 평균입니다 . " 선택 항목 실행 " 을 클릭하면 필요한 폴드 , 내측 폴드 및 시행 조합에 대한 검증 열이 데이터 테이블에 생성됩니다 .

선택한 스크립트 저장 선택한 각 행의 스크립트 창에 모형 스크립트를 저장합니다 . 시작 창에서 폴드 교차 검증 유형을 지정한 경우 스크립트에 지정된 검증 데이터 집합은 최종 모형에 해당하는 폴드 , 내측 폴드 및 시행 조합을 사용합니다 . 최종 모형은 가중 평균 R^2 이 가장 높은 모형입니다 . 가중 평균 R^2 은 훈련 R^2, 검증 R^2 및 테스트 R^2(테스트 데이터 집합이 사용된 경우) 의 관측값 수에 따라 가중치가 적용된 평균입니다 . " 선택한 스크립트 저장 " 을 클릭하면 요청된 폴드 , 내측 폴드 및 시행 조합에 대한 검증 열이 데이터 테이블에 생성됩니다 .

훈련 , 검증 및 테스트 적합 측도

지정된 각 모형 데이터 집합에 대한 적합 측도 보고서가 있습니다 . 훈련 데이터 집합 , 훈련 및 검증 데이터 집합 또는 훈련 , 검증 및 테스트 데이터 집합이 될 수 있습니다 . 각 보고서에는 다음 열을 포함하는 테이블이 있습니다 .

방법 모형을 적합시키는 데 사용된 방법의 이름입니다 .

N 데이터 집합의 관측값 수입니다 .

가중치 합 가중치의 합입니다 .

R^2 (연속형 반응에만 사용 가능) 적합 모형의 R^2 값입니다 .

엔트로피 R^2 (범주형 반응에만 사용 가능) 적합 모형과 상수 확률 모형의 로그 가능도를 비교하는 적합 측도입니다 . 엔트로피 R^2 범위는 0 ~ 1 이며 , 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 . 자세한 내용은 " 엔트로피 R^2" 에서 확인하십시오 .

오분류 비율 (범주형 반응에만 사용 가능) 모형에 의해 잘못 분류된 관측값의 비율입니다 . 값이 작을수록 더 나은 적합을 나타냅니다 .

참고 : 이러한 테이블에서 오분류 비율은 항상 확률 임계 0.5 를 사용하여 계산됩니다 .

AUC (범주형 반응에만 사용 가능) ROC 곡선 아래 면적입니다 . 값이 1 에 가까울수록 더 나은 적합을 나타냅니다 .

RASE 평균 제곱 예측 오차의 제곱근입니다 (제곱근 평균 제곱 오차). RASE 는 다음과 같이 계산되며 , 여기서 Source 는 훈련 , 검증 또는 테스트 데이터 집합을 나타냅니다 .

$$\text{RASE}_{Source} = \sqrt{\frac{SSE_{Source}}{n}}$$

일반화 R^2 (범주형 반응에만 사용 가능) 일반 회귀 모형에 적용할 수 있는 측도입니다 . 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다 . 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다 . 일반화 R^2 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2로 단순화됩니다. 일반화 R^2을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데 , 이는 Cox-Snell 유사 R^2 을 정규화한 것입니다 . 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오 .

폴드 (시작 창에서 "K 폴드 교차 검증 " 옵션 또는 " 내포 교차 검증 " 옵션을 지정한 경우에만 사용 가능) 해당 행에서 모형 적합을 위해 홀드아웃된 폴드를 식별합니다 .

내측 폴드 (시작 창에서 " 내포 교차 검증 " 옵션을 지정한 경우에만 사용 가능) 해당 행에서 모형 적합을 위해 홀드아웃된 내측 폴드를 식별합니다 .

시행 (시작 창에서 " 반복 K 폴드 " 옵션을 지정한 경우에만 사용 가능) 해당 행에서 모형 적합에 대한 시행 번호를 식별합니다 .

각 테이블 아래에 다음과 같은 옵션이 있습니다 .

우월 행 선택 모형 적합 기준 조합에 따라 다른 모든 모형보다 우수하거나 동일한 각 모형을 선택합니다 . 연속형 반응의 경우 우월 모형을 결정할 때 R^2 과 빈도 합을 고려합니다 . 범주형 반응의 경우 우월 모형을 결정할 때 엔트로피 R^2, 오분류 비율 , AUC 및 빈도 합을 고려합니다 .

선택 항목 실행 선택한 각 행에 지정된 개별 모형을 실행합니다 .

선택한 스크립트 저장 선택한 각 행의 스크립트 창에 모형 스크립트를 저장합니다 .

JMP PRO 모형 선별 플랫폼 옵션

실제값 대 예측값 그림 (연속형 반응에만 사용 가능) 모든 모형 적합의 실제값 대 예측값 점을 중첩하는 그림 보고서를 표시하거나 숨깁니다 . 훈련 , 검증 및 테스트 데이터 집합에 대한 개별 그림이 있습니다 . 이 보고서에는 방법 , 폴드 , 내측 폴드 및 시행 조합을 기반으로 그림에 나타나는 적합을 지정할 수 있는 체크박스가 포함된 " 방법 " 범례가 있습니다 . 범례 맨 아래에 있는 " 모두 선택 취소 " 를 클릭하면 그림에서 모든 모형 적합이 제거됩니다 .

참고 :

- 훈련 , 검증 또는 테스트 보고서에서 모형의 부분집합을 선택한 후 " 실제값 대 예측값 그림 " 을 선택하면 선택한 모형에 대한 그림만 표시됩니다 .
- 일부 모델링 플랫폼에서는 실제값 대 예측값 그림이 지원되지 않습니다 . 지원되지 않는 플랫폼 목록이 그림 아래에 표시됩니다 .

프로파일러 (연속형 반응에만 사용 가능) 각 모형 적합 유형에 대한 예측 프로파일러를 표시하거나 숨깁니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 . 시작 창에서 폴드 교차 검증 유형을 지정한 경우 예측 프로파일러의 선택된 모형은 최종 모형에 해당하는 폴드 , 내측 폴드 및 시행 조합을 사용합니다 . 최종 모형은 가중 평균 R^2 이 가장 높은 모형입니다 . 가중 평균 R^2 은 훈련 R^2, 검증 R^2 및 테스트 R^2(테스트 데이터 집합이 사용된 경우) 의 관측값 수에 따라 가중치가 적용된 평균입니다 .

참고 :

- 표준 최소 제곱 모형 및 일반화 회귀 모형의 예측 프로파일러에 신뢰 구간이 표시됩니다 .
- K 최근접 이웃 모형에는 예측 프로파일러를 사용할 수 없습니다 .

ROC 곡선 (범주형 반응에만 사용 가능) 모든 모형 적합에 대한 ROC 곡선을 중첩하는 그림 보고서를 표시하거나 숨깁니다 . 훈련 , 검증 및 테스트 데이터 집합에 대한 개별 그림이 있습니다 . 이 보고서에는 그림에 나타나는 적합을 지정할 수 있는 체크박스가 포함된 두 개의 범례가 있습니다 . 범례 맨 아래에 있는 " 모두 선택 취소 " 를 클릭하면 그림에서 모든 모형 적합이 제거됩니다 . ROC 곡선에 대한 자세한 내용은 "ROC 곡선 " 에서 확인하십시오 .

참고 :

- 훈련 , 검증 또는 테스트 보고서에서 모형의 부분집합을 선택한 후 "ROC 곡선 " 을 선택하면 선택한 모형에 대한 그림만 표시됩니다 .

– 일부 모델링 플랫폼에서는 ROC 그림이 지원되지 않습니다 . 지원되지 않는 플랫폼 목록이 그림 아래에 표시됩니다 .

수익 표시 (" 수익 행렬 " 열 특성이 있는 이항 범주형 반응에만 사용 가능) 각 모형에 대한 기대 수익을 표시하거나 숨깁니다 .

결정 임계값 (이항 범주형 반응에만 사용 가능) 훈련 , 검증 및 테스트 (지정된 경우) 데이터 집합에 대한 " 결정 임계값 " 보고서를 표시하거나 숨깁니다 . 각 보고서에는 임계값을 조정하고 모형 적합을 비교할 수 있도록 각 모형에 대한 적합 확률 분포 그래프 , 각 모형에 대한 혼동 행렬 및 분류 그래프가 포함되어 있습니다 . 자세한 내용은 " 결정 임계값 보고서 " 에서 확인하십시오 .

참고 : 훈련 , 검증 또는 테스트 보고서에서 모형의 부분집합을 선택한 후 " 결정 임계값 " 을 선택하면 선택한 모형에 대한 정보만 표시됩니다 .

결과 테이블 저장 검증 보고서의 정보를 " 모형 선별 통계량 검증 데이터 집합 " 이라는 새 데이터 테이블에 저장합니다 . " 내포 교차 검증 " 옵션이 지정된 경우와 같이 테스트 데이터 집합이 있는 경우 테스트 보고서의 정보도 " 모형 선별 통계량 테스트 데이터 집합 " 이라는 새 데이터 테이블에 저장됩니다 .

K 폴드 결과 테이블 저장 (시작 창에서 교차 검증 옵션을 지정한 경우에만 사용 가능) " 폴드 간 요약 " 보고서의 정보를 "K 폴드 결과 " 라는 새 데이터 테이블에 저장합니다 .

예측 계산식 저장 (Live 보고서가 모형 보고서에 포함된 경우에만 사용 가능) 선택한 모형에 대한 예측 계산식을 데이터 테이블에 저장합니다 . " 폴드 간 요약 " 보고서에서 모형을 선택하고 시작 창에 폴드 교차 검증 유형이 지정된 경우 최종 모형에 해당하는 폴드 , 내측 폴드 및 시행 조합에 대한 예측 계산식이 저장됩니다 . 최종 모형은 가중 평균 R^2 이 가장 높은 모형입니다 . 가중 평균 R^2 은 훈련 R^2, 검증 R^2 및 테스트 R^2(테스트 데이터 집합이 사용된 경우) 의 관측값 수에 따라 가중치가 적용된 평균입니다 .

참고 :

– 단계별 모형의 경우 예측 계산식이 저장되지 않습니다 .

– 순서형 반응을 사용하는 SVM 모형의 경우 예측 계산식이 저장되지 않습니다 .

선택적 보고서 추가 보고서 옵션 하위 메뉴를 표시합니다 .

경과 시간 각 모형 유형을 적합시키는 데 소요된 총 경과 시간을 포함하는 보고서를 표시하거나 숨깁니다 .

예측 변수 카디널리티 각 범주형 예측 변수의 선형 모형 적합에 사용되는 수준 수와 모수 수에 대한 보고서를 표시하거나 숨깁니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

결정 임계값 보고서

"결정 임계값" 보고서를 사용하면 이항 분류 모형에 대한 임계값을 탐색할 수 있습니다. 검증 방법으로 지정된 각 모형 데이터 집합에 대한 "결정 임계값" 보고서가 있습니다. 훈련 데이터 집합, 훈련 및 검증 데이터 집합 또는 훈련, 검증 및 테스트 데이터 집합이 될 수 있습니다. 각 "결정 임계값" 보고서에는 적합 확률 분포 그래프, 분류 막대 차트 및 혼동 행렬이 포함되어 있습니다. 모두 모형 적합, 폴드, 시행 및 클래스 수준별로 구성됩니다. 보고서에는 분류 정확성 측도에 대한 탭 형식 섹션과 수익 행렬 설정 옵션도 포함되어 있습니다. 확률 임계를 조정하면 보고서가 업데이트됩니다.

적합 확률 분포

적합 확률 분포 또는 모형 스코어를 통해 각 개별 모형 적합이 두 클래스를 어떻게 구분하는지 확인할 수 있습니다. 그래프의 세로선은 각 관측값의 분류를 결정하는 확률 임계를 나타냅니다. 기본적으로 확률 임계는 0.5입니다. 세로선을 드래그하거나, "확률 임계" 값을 클릭하고 새 값을 입력하여 확률 임계를 변경할 수 있습니다. 그러면 전체 "결정 임계값" 보고서에서 확률 임계값이 변경됩니다. 확률 임계값은 0에서 1 사이여야 합니다.

분류 개수

분류 막대 차트에는 현재 임계에서 반응 변수의 각 수준에 대한 분류 개수가 표시됩니다. 녹색 막대는 올바르게 분류된 관측값을 나타내고 빨간색 막대는 잘못 분류된 관측값을 나타냅니다.

혼동 행렬

분할표라고도 하는 혼동 행렬은 각 개별 모형 적합에 대한 실제 반응과 예측 반응의 이원 분류를 보여 줍니다. 혼동 비율 행렬도 표시됩니다. 비율은 혼동 행렬의 값을 행 합계로 나눈 것과 같습니다.

분류 정확성 측도

임계별 분류 (False) 오분류 수 대 확률 임계 그림과 오분류 비율 대 확률 임계 그림을 표시합니다 . 각 그림에는 각 개별 모형 적합에 대한 두 개의 곡선이 포함됩니다 . 낮은 반응 범주의 곡선은 실선이고 높은 반응 범주의 곡선은 파선입니다 . 곡선은 각 반응 수준에 대해 동일한 오분류

수 또는 비율을 생성하는 임계값에서 교차합니다 . 또한 각 그래프에는 현재 확률 임계값을 나타내는 세로선이 있습니다 . 이 세로선을 드래그하여 확률 임계값을 변경할 수 있습니다 . 그러면 전체 보고서에서 확률 임계값이 변경됩니다 .

비율별 분류 (False) 오분류 수 대 순위별 스코어 비율 그림 또는 오분류 비율 대 순위별 스코어 비율 그림을 표시합니다 . 각 그림에는 각 개별 모형 적합에 대한 두 개의 곡선이 포함됩니다 . 낮은 반응 범주의 곡선은 실선이고 높은 반응 범주의 곡선은 파선입니다 .

임계별 분류 (True) True 수 대 확률 임계 그림과 True 비율 대 확률 임계 그림을 표시합니다 . 각 그림에는 각 개별 모형 적합에 대한 두 개의 곡선이 포함됩니다 . 낮은 반응 범주의 곡선은 파선이고 높은 반응 수준의 곡선은 실선입니다 . 곡선은 각 반응에 대해 동일한 올바른 분류 수 또는 비율을 생성하는 임계값에서 교차합니다. 또한 각 그래프에는 현재 확률 임계값을 나타내는 세로선이 있습니다 . 이 세로선을 드래그하여 확률 임계값을 변경할 수 있습니다 . 그러면 전체 보고서에서 확률 임계값이 변경됩니다 .

비율별 분류 (True) True 수 대 순위별 스코어 비율 그림 또는 True 비율 대 순위별 스코어 비율 그림을 표시합니다 . 각 그림에는 각 개별 모형 적합에 대한 두 개의 곡선이 포함됩니다 . 낮은 반응 범주의 곡선은 파선이고 높은 반응 범주의 곡선은 실선입니다 .

임계별 수익 (수익 행렬을 지정한 경우에만 사용 가능) 평균 수익 대 확률 임계 그림을 표시합니다 . 각 개별 모형 적합에 대한 곡선과 현재 확률 임계값을 나타내는 세로선이 있습니다 . 지정된 수익 행렬도 그림 옆에 표시됩니다 .

측정 기준 각 모형에 대한 분류 정확성 측정 기준 테이블을 표시합니다 . 각 열의 측정 기준이 계산되는 방법을 설명하는 범례가 제공됩니다 .

수익 행렬 설정

원하지 않는 결과에 비용을 할당하고 원하는 결과에 수익을 할당할 수 있습니다. 자세한 내용은 "수익 행렬 지정"에서 확인하십시오. 수익 행렬 창에서 확률 임계를 변경하고 "확인"을 클릭하면 해당 값을 확률 임계로 사용하여 "결정 임계값" 보고서가 업데이트됩니다.

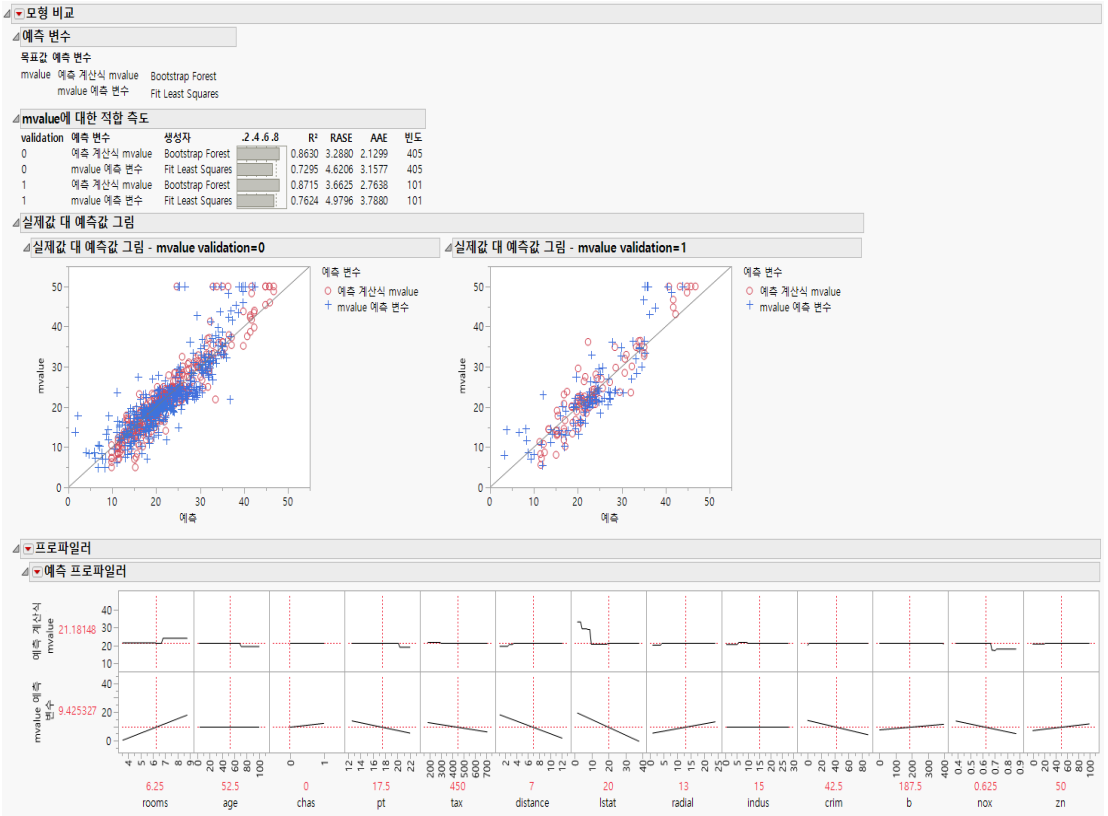
11 장

JMP PRO 모형 비교
적합 모형의 예측 능력 비교

모형 비교 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

JMP Pro의 모형 비교 플랫폼을 사용하면 여러 모형의 예측 능력을 비교할 수 있습니다. 중첩된 진단 그림과 함께 각 모형에 대한 적합 측도가 제공됩니다.

그림 11.1 모형 비교의 예



목차

모형 비교의 예 173
모형 비교 플랫폼 시작 176
모형 비교 보고서 177
모형 비교 플랫폼 옵션 179
연속형 및 범주형 반응 179
연속형 반응 179
범주형 반응 179
모형 비교의 추가 예 180

JMP PRO 모형 비교의 예

이 섹션에서는 모형 비교 플랫폼을 사용하는 예를 제공합니다. 이 예에서는 인구 통계학 데이터를 사용하여 중위수 주택 가격에 대한 모형을 생성합니다. 회귀 모형과 붓스트랩 포레스트 모형을 비교합니다.

도움말 > 샘플 데이터 라이브러리를 선택하고 Boston Housing.jmp 를 열어 시작합니다.

검증 열 생성

1. **분석** > **예측 모델링** > **검증 열 생성**을 선택합니다 .
2. 시작 창에서 열을 선택하지 않습니다 .

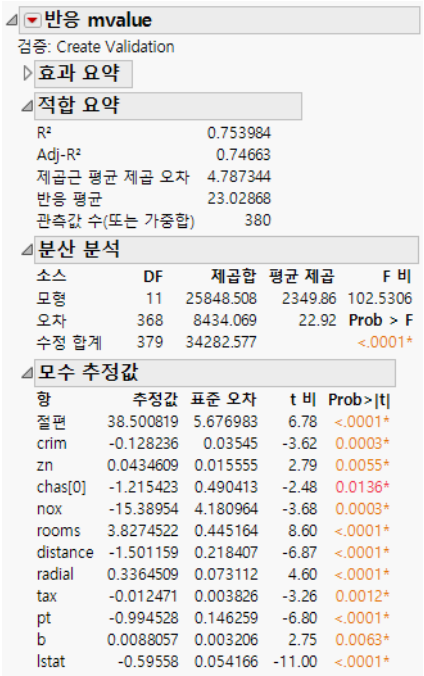
   이는 플랫폼에서 단순 랜덤 검증 열을 생성한다는 것을 나타냅니다 .
3. **확인**을 클릭합니다 .
4. " 새 열 이름 " 옆의 상자에 "Create Validation" 을 입력합니다 .
5. " 난수 시드값 " 옆의 상자에 "1234" 를 입력합니다 .
6. **시작**을 클릭합니다 .

   새 검증 열이 생성됩니다 . 0 이 할당된 행은 훈련 데이터 집합입니다 . 1 이 할당된 행은 검증 데이터 집합입니다 .

회귀 모형 생성 및 열에 예측 계산식 저장

1. **분석 > 모형 적합**을 선택합니다 .
2. mvalue 를 선택하고 **Y** 를 클릭합니다 .
3. crim ~ lstat 를 선택하고 **추가**를 클릭합니다 .
4. Create Validation 을 선택하고 **검증**을 클릭합니다 .
5. " 분석법 " 목록에서 **단계별**을 선택합니다 .
6. **실행** 버튼을 클릭합니다 .
7. " 중지 규칙 " 목록에서 **p 값 임계**를 선택합니다 .
8. **시작** 버튼을 클릭합니다 .
9. **모형 실행** 버튼을 클릭합니다 .

그림 11.2 모형 적합 보고서



반응 mvalue

검증: Create Validation

효과 요약

적합 요약

R^2	0.753984
Adj-R^2	0.74663
제곱근 평균 제곱 오차	4.787344
반응 평균	23.02868
관측값 수(또는 가중합)	380

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	11	25848.508	2349.86	102.5306
오차	368	8434.069	22.92	Prob > F
수정 합계	379	34282.577		<.0001*

모수 추정값

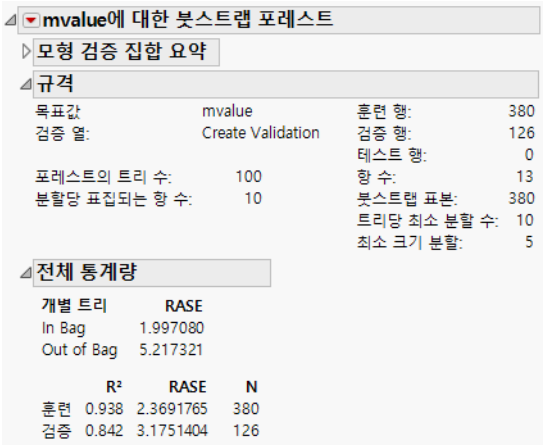
항	추정값	표준 오차	t 비	Prob>|t|
절편	38.500819	5.676983	6.78	<.0001*
crim	-0.128236	0.03545	-3.62	0.0003*
zn	0.0434609	0.015555	2.79	0.0055*
chas[0]	-1.215423	0.490413	-2.48	0.0136*
nox	-15.38954	4.180964	-3.68	0.0003*
rooms	3.8274522	0.445164	8.60	<.0001*
distance	-1.501159	0.218407	-6.87	<.0001*
radial	0.3364509	0.073112	4.60	<.0001*
tax	-0.012471	0.003826	-3.26	0.0012*
pt	-0.994528	0.146259	-6.80	<.0001*
b	0.0088057	0.003206	2.75	0.0063*
lstat	-0.59558	0.054166	-11.00	<.0001*

10. 예측 계산식을 열에 저장하려면 " 반응 " 의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식**을 선택합니다 .

붓스트랩 포레스트 모형 생성 및 열에 예측 계산식 저장

1. **분석 > 예측 모델링 > 붓스트랩 포레스트**를 선택합니다 .
2. mvalue 를 선택하고 **Y, 반응**을 클릭합니다 .
3. crim ~ lstat 를 선택하고 **X, 요인**을 클릭합니다 .
4. Create Validation 을 선택하고 **검증**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. **조기 중지** 체크박스를 선택합니다 .
7. **여러 항에 대한 다중 적합** 체크박스를 선택합니다 .
8. **난수 시드값** 옆의 상자에 "617" 을 입력합니다 .
9. **확인**을 클릭합니다 .

그림 11.3 붓스트랩 포레스트 모형



10. 예측 계산식을 열에 저장하려면 " 붓스트랩 포레스트 " 의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다 .

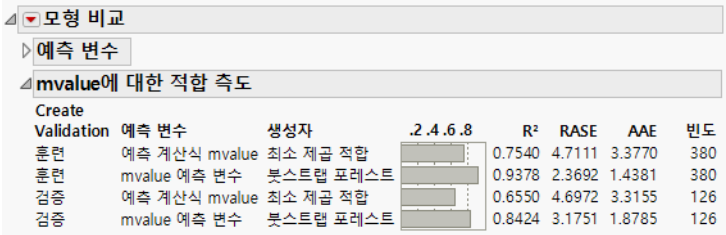
모형 비교

1. **분석 > 예측 모델링 > 모형 비교**를 선택합니다 .
2. 예측 계산식 열 두 개를 선택하고 **Y, 예측 변수**를 클릭합니다 .
3. Create Validation 을 선택하고 **그룹**을 클릭합니다 .

> **팁 : 그룹** 열을 지정하지 않으면 JMP 에서 동일한 검증 열이 모든 예측 변수에 사용될 때 이를 자동으로 인식하여 그룹화 변수로 추가하라는 메시지가 표시됩니다 .

4. **확인**을 클릭합니다 .

그림 11.4 모형 비교 보고서

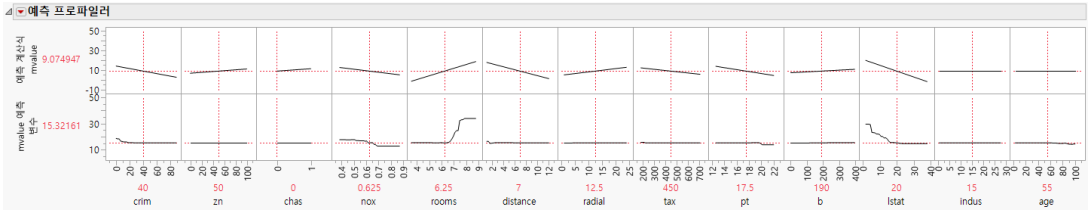


훈련 데이터 집합의 행은 모형을 생성하는 데 사용되었으므로 Create Validation = 훈련에 대한 R^2 통계량이 인위적으로 부풀려질 수 있습니다 . 이 경우 통계량은 모형의 미래 예측 능력을 대표하지 못합니다 . 붓스트랩 포레스트 모형의 경우 특히 그렇습니다 .

Create Validation = 검증에 대한 통계량을 사용하여 모형을 비교합니다 . 이 경우 붓스트랩 포레스트 모형이 회귀 모형보다 더 잘 예측합니다 .

5. "모형 비교"의 빨간색 삼각형을 클릭하고 "프로파일러"를 선택합니다.

그림 11.5 모든 모형에 대한 예측 프로파일러



예측 프로파일러를 사용하면 서로 다른 모형에서 각 요인의 영향을 비교할 수 있습니다. 프로파일러는 회귀 모형과 파티션 모형을 사용하는 경우와 같이 여러 유형의 모형을 비교할 때 특히 유용합니다.

관련 정보

- 선형 모형 적합의
- "파티션 모형"

JMP PRO 모형 비교 플랫폼 시작

모형 비교 플랫폼을 시작하려면 **분석 > 예측 모델링 > 모형 비교**를 선택합니다.

그림 11.6 모형 비교 시작 창

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 예측 변수 비교할 모형에 대한 예측이 포함된 열입니다. 계산식 열 또는 데이터 열일 수 있습니다. JMP 플랫폼에서 생성된 예측 계산식 열에는 " 예측 " 또는 " 반응 확률 " 열 특성이 있습니다. 이러한 특성 중 하나가 포함되지 않은 열을 지정하면 플랫폼에서 지정된 Y 열에 의해 예측되는 열을 지정하라는 메시지가 표시됩니다.

수준이 *k* 개 있는 범주형 반응의 경우 대부분의 모형 적합 플랫폼은 각각 수준의 확률을 예측하는 *k* 개 열을 데이터 테이블에 저장합니다. *k* 개 열은 모두 **Y, 예측 변수**로 지정해야 합니다. *k* 개의 확률 열을 저장하지 않는 플랫폼의 경우 예측 반응 수준이 포함된 열을 **Y, 예측 변수** 열로 지정할 수 있습니다.

Y, 예측 변수 열을 지정하지 않으면 데이터 테이블에서 " 예측 " 또는 " 반응 확률 " 열 특성이 있는 예측 계산식 열이 사용됩니다.

그룹 데이터를 별도로 평가되는 그룹으로 구분하는 열입니다. **그룹** 열을 지정하지 않으면 JMP에서 동일한 검증 열이 모든 예측 변수에 사용될 때 이를 자동으로 인식하여 그룹화 변수로 추가하라는 메시지가 표시됩니다.

다른 역할 버튼은 JMP 플랫폼 간에 공통적입니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

JMP PRO 모형 비교 보고서

그림 11.7 에서는 연속형 반응에 대한 초기 모형 비교 보고서의 예를 보여 줍니다.

그림 11.7 초기 모형 비교 보고서

모형 비교

예측 변수

mvalue에 대한 적합 측도

Create Validation	예측 변수	생성자	.2 .4 .6 .8	R^2	RASE	AAE	빈도
훈련	예측 계산식 mvalue	최소 제곱 적합		0.7540	4.7111	3.3770	380
훈련	mvalue 예측 변수	붓스트랩 포레스트		0.9378	2.3692	1.4381	380
검증	예측 계산식 mvalue	최소 제곱 적합		0.6550	4.6972	3.3155	126
검증	mvalue 예측 변수	붓스트랩 포레스트		0.8424	3.1751	1.8785	126

"예측 변수" 보고서에는 모든 반응 및 각 반응에 대해 비교되는 모든 모형이 표시됩니다. 예측 변수 열을 생성한 적합 플랫폼도 나열됩니다.

"적합 측도" 보고서에는 각 모형에 대한 적합 측도가 표시됩니다. 연속형 반응과 범주형 반응의 열은 서로 다릅니다.

연속형 반응에 대한 적합 측도

R^2 R^2 통계량입니다. 결측값이 없는 데이터 테이블에서는 "모형 비교" 보고서의 R^2 통계량이 원래 모형과 일치합니다. 그러나 결측값이 있으면 R^2 통계량이 다릅니다.

RASE 평균 제곱 예측 오차의 제곱근입니다. 이 값은 다음과 같이 계산됩니다.

– 예측 오차 (실제 반응과 예측 반응의 차이) 를 제곱하고 합하여 *SSE* 를 구합니다.

- 관측값 수를 n 으로 나타냅니다 .
- RASE 는 다음과 같습니다 .

$$\text{RASE} = \sqrt{\frac{SSE}{n}}$$

AAE 평균 절대 오차입니다 .

빈도 각 행의 빈도 수를 포함하는 열입니다 .

범주형 반응에 대한 적합 측도

엔트로피 R² 1 에서 적합 모형과 상수 확률 모형의 음의 로그 가능도 비율을 뺀 값입니다 . 값 범위는 0 에서 1 사이입니다 .

일반화 R² 일반 회귀 모형에 적용할 수 있는 측도입니다 . 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다 . 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다 . 일반화 R² 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R² 로 단순화됩니다 . 일반화 R² 을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데 , 이는 Cox-Snell 유사 R^2 을 정규화한 것입니다 . 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오 .

평균 -Log p -log(p) 의 평균입니다 . 여기서 p 는 발생한 사건과 연관된 적합 확률입니다 .

RASE 제곱근 평균 제곱 예측 오차입니다 . 범주형 반응의 경우 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 .

평균 절대 편차 반응과 예측 반응 사이의 차이에 대한 절대값의 평균입니다 . 범주형 반응의 경우 차이는 1 에서 p(실제로 발생한 반응 수준의 적합 확률) 사이입니다 .

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다 .

N 관측값 수입니다 .

관련 정보

"훈련 및 검증 적합 측도"에서는 범주형 반응의 적합 측도에 대한 자세한 정보를 제공합니다.

JMP PRO 모형 비교 플랫폼 옵션

JMP PRO 연속형 및 범주형 반응

모형 평균화 예측값 (연속형 반응의 경우) 또는 예측 확률 (범주형 반응의 경우) 의 산술 평균이 포함된 새 열을 생성합니다 .

JMP PRO 연속형 반응

실제값 대 예측값 그림 실제값 대 예측값의 산점도를 표시합니다 . 여러 모형에 대한 그림이 중첩됩니다 .

잔차 대 행 번호 그림 잔차 대 행 번호 그림을 표시합니다 . 여러 모형에 대한 그림이 중첩됩니다 .

프로파일러 데이터의 예측 계산식 열을 기반으로 각 반응에 대한 프로파일러를 표시합니다 . 프로파일러에는 비교되는 각 모형에 대한 행이 있습니다 .

JMP PRO 범주형 반응

ROC 곡선 반응 변수의 각 수준에 대한 ROC 곡선을 표시합니다 . 여러 모형에 대한 곡선이 중첩됩니다 .

AUC 비교 각 모형의 AUC(ROC 곡선 아래 면적) 비교를 제공합니다 . 곡선 아래 면적은 적합도를 나타내는 표시자이며 값이 1 이면 완벽한 적합을 나타냅니다 .

이 보고서에는 다음과 같은 정보가 포함됩니다 .

- 각 AUC 에 대한 표준 오차 및 신뢰 구간
- 각 AUC 쌍 간의 차이에 대한 표준 오차 , 신뢰 구간 및 가설 검정
- 모든 AUC 가 동일한지 여부를 검정하기 위한 전체 가설 검정

향상도 곡선 반응 변수의 각 수준에 대한 향상도 곡선을 표시합니다 . 여러 모형에 대한 곡선이 중첩됩니다 .

누적 이득 곡선 반응 변수의 각 수준에 대한 누적 이득 곡선을 표시합니다 . 누적 이득 곡선은 모든 반응의 비율 대비 모형에서 식별된 반응 수준의 비율을 보여 주는 그림입니다 . 완벽한 모형의 누적 이득 곡선은 반응 수준의 전체 비율에서 1.0 에 도달합니다 . 여러 모형에 대한 곡선이 중첩됩니다 .

혼동 행렬 각 모형에 대한 혼동 행렬을 표시합니다 . 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다 . 개수 및 비율 혼동 행렬이 표시됩니다 . 그룹 변수의 각 수준에 대해 별도의 혼동 행렬이 생성됩니다 .

반응에 " 수익 행렬 " 열 특성이 있으면 " 실제값 대 결정 개수 " 및 " 실제값 대 결정 비율 " 행렬이 혼동 행렬 오른쪽에 표시됩니다 . 이러한 행렬에 대한 자세한 내용은 " 분할의 추가 예 " 에서 확인하십시오 .

프로파일러 데이터의 예측 계산식 열을 기반으로 각 반응에 대한 프로파일러를 표시합니다. 프로파일러에는 비교되는 각 모형에 대한 행이 있습니다.

결정 임계값 (이항 범주형 반응에만 사용 가능) 훈련, 검증 및 테스트 (지정된 경우) 데이터 집합에 대한 " 결정 임계값 " 보고서를 표시하거나 숨깁니다. 각 보고서에는 모형 적합을 비교할 수 있도록 각 모형에 대한 적합 확률 분포 그래프, 각 모형에 대한 혼동 행렬 및 분류 그래프가 포함되어 있습니다. 자세한 내용은 " 결정 임계값 보고서 " 에서 확인하십시오.

관련 정보

- "ROC 곡선 "
- " 향상도 곡선 "

JMP PRO 모형 비교의 추가 예

이 예에서는 자동차 데이터를 사용하여 구매한 차량의 크기를 예측하는 모형을 생성합니다. 로지스틱 회귀 모형과 결정 트리 모형을 비교합니다.

도움말 > 샘플 데이터 라이브러리를 선택하고 Car Physical Data.jmp 를 열어 시작합니다.

로지스틱 회귀 모형 생성

1. **분석 > 모형 적합**을 선택합니다.
2. Type 을 선택하고 **Y** 를 클릭합니다.
3. Country, Weight, Turning Cycle, Displacement 및 Horsepower 열을 선택하고 **추가**를 클릭합니다.
4. **실행**을 클릭합니다.

   명목형 로지스틱 적합 보고서가 나타납니다.
5. 예측 계산식을 열에 저장하려면 " 명목형 로지스틱 " 의 빨간색 삼각형을 클릭하고 **확률 계산식 저장**을 선택합니다.

결정 트리 모형 생성 및 열에 예측 계산식 저장

1. **분석 > 예측 모델링 > 파티션**을 선택합니다.
2. Type 을 선택하고 **Y, 반응**을 클릭합니다.
3. Country, Weight, Turning Cycle, Displacement 및 Horsepower 열을 선택하고 **X, 요인**을 클릭합니다.
4. " 방법 " 목록에서 **결정 트리**가 선택되어 있는지 확인합니다.
5. **확인**을 클릭합니다.

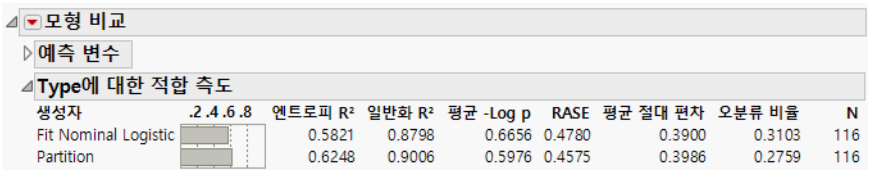
   분할 보고서가 나타납니다.
6. **분할**을 10 번 클릭합니다.

7. 예측 계산식을 열에 저장하려면 " 분할 " 의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다 .

모형 비교

1. **분석 > 예측 모델링 > 모형 비교**를 선택합니다 .
2. 확률로 시작하는 모든 열을 선택하고 **Y, 예측 변수**를 클릭합니다 .
3. **확인**을 클릭합니다 .

그림 11.8 초기 모형 비교 보고서

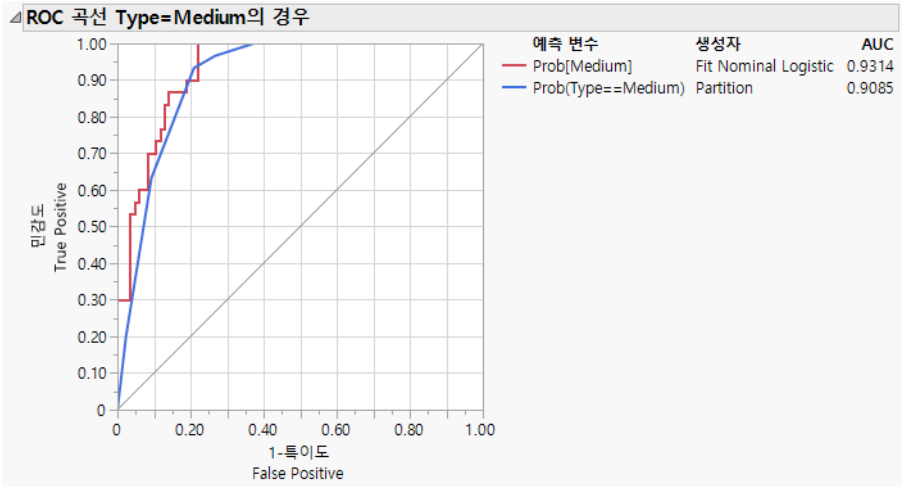


보고서에 따르면 파티션 모형의 엔트로피 R^2 및 일반화 R^2 값이 약간 더 높고 오분류 비율 값이 약간 더 낮습니다 .

4. " 모형 비교 " 의 빨간색 삼각형을 클릭하고 **ROC 곡선**을 선택합니다 .

   각 Type 에 대한 ROC 곡선이 나타나며 , 그 중 하나가 그림 11.9 에 표시되어 있습니다 .

그림 11.9 Medium 에 대한 ROC 곡선

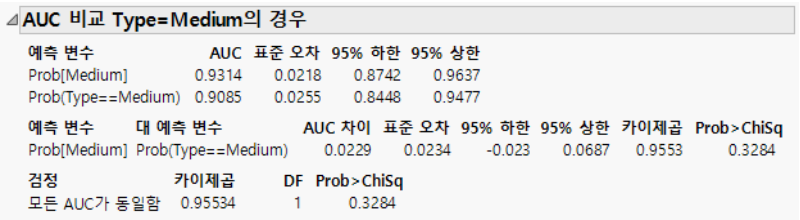


모든 ROC 곡선을 검토하면 두 모형의 예측 능력이 비슷하다는 것을 알 수 있습니다 .

5. " 모형 비교 " 의 빨간색 삼각형을 클릭하고 **AUC 비교**를 선택합니다 .

   각 Type 에 대한 AUC 비교 보고서가 나타나며 , 그 중 하나가 그림 11.10 에 표시되어 있습니다 .

그림 11.10 Medium 에 대한 AUC 비교



AUC 비교 Type=Medium의 경우

예측 변수	AUC	표준 오차	95% 하한	95% 상한
Prob[Medium]	0.9314	0.0218	0.8742	0.9637
Prob(Type==Medium)	0.9085	0.0255	0.8448	0.9477

예측 변수	대 예측 변수	AUC 차이	표준 오차	95% 하한	95% 상한	카이제곱	Prob>ChiSq
Prob[Medium]	Prob(Type==Medium)	0.0229	0.0234	-0.023	0.0687	0.9553	0.3284

검정	카이제곱	DF	Prob>ChiSq
모든 AUC가 동일함	0.95534	1	0.3284

이 보고서에는 AUC 값(ROC 곡선 아래 면적)의 차이에 대한 가설 검정 결과가 표시됩니다. 결과를 검토하면 모든 Type 수준의 값 사이에 통계적 차이가 없음을 알 수 있습니다.

다음과 같은 이유로 두 모형의 예측 능력에 큰 차이가 없다는 결론을 내릴 수 있습니다.

- R^2 값 및 ROC 곡선이 비슷합니다.
- AUC 값 사이에 통계적으로 유의한 차이가 없습니다.

12 장

JMP PRO 검증 열 생성

데이터를 훈련 , 검증 및 테스트 데이터 집합으로 나누기

검증 열 생성 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

검증은 데이터 집합의 일부를 사용하여 모형 모수를 추정하고 다른 부분을 사용하여 모형의 예측 능력을 평가하는 프로세스입니다 . 데이터를 두 개 또는 세 개의 집합으로 분할할 수 있습니다 . 검증 열 생성 플랫폼에서는 이러한 분할을 생성하는 다섯 가지 방법을 제공합니다 .

목차

검증 열 생성 플랫폼 개요 185
검증 열 생성의 예 185
검증 열 생성 플랫폼 시작 187
검증 열 생성 보고서 189
비율 또는 상대 비율 지정 189
절단점 설정 190
옵션 190
시작 190
검증 열 생성 플랫폼 옵션 191
검증 열 생성 플랫폼의 추가 예 191

JMP PRO 검증 열 생성 플랫폼 개요

검증은 데이터 집합의 일부를 사용하여 모형 모수를 추정하고 다른 부분을 사용하여 모형의 예측 능력을 평가하는 프로세스입니다. 복잡한 데이터의 경우 이 방법을 사용하면 모형 과대적합 위험을 줄일 수 있습니다.

검증 열은 데이터를 두 개 또는 세 개의 부분으로 분할합니다.

- 훈련 데이터 집합은 모형 모수를 추정하는 데 사용됩니다.
- 검증 데이터 집합은 예측 능력이 좋은 모형을 선택하는 데 사용됩니다.
- 테스트 데이터 집합은 모형이 선택된 후 해당 모형의 예측 능력을 확인합니다.

검증 열은 많은 JMP 플랫폼에서 검증 방법으로 사용할 수 있습니다. 자세한 내용은 "JMP 모델링의 검증"에서 확인하십시오.

검증 열 생성 플랫폼에서는 다양한 방법을 사용하여 훈련, 검증 및 테스트 데이터 집합을 생성할 수 있습니다. 층화, 그룹화 또는 절단점 열을 지정하여 검증 열 생성에 사용되는 방법을 결정할 수 있습니다.

JMP PRO 검증 열 생성의 예

Lipid Data.jmp 데이터 테이블에는 캘리포니아 병원에 있는 피실험자 95 명의 혈액 측정, 신체 측정 및 설문지 데이터가 포함되어 있습니다. 향후 분석 중에 사용할 검증 열을 생성하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Lipid Data.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. Gender 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.

그림 12.1 Lipid Data.jmp 의 Gender 분포

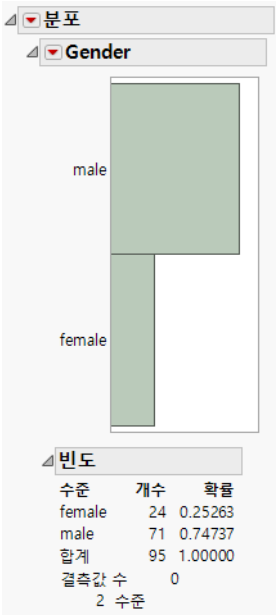


그림 12.1 에서는 데이터 집합의 Gender 분포를 보여 줍니다 . 남성과 여성의 비율이 같지 않다는 점에 유의하십시오 . 데이터 내에 여성 수가 더 적으므로 검증 데이터 집합과 훈련 데이터 집합에서 성별 균형을 유지하려고 합니다 .

5. **분석 > 예측 모델링 > 검증 열 생성**을 선택합니다 .
6. Gender 를 선택하고 **층화 열**을 클릭합니다 .
7. **확인**을 클릭합니다 .

   선택한 검증 방법에 대한 설명이 포함된 " 검증 열 생성 " 보고서가 나타납니다 . 비율 , 열 유형을 변경하거나 시드값을 설정하는 옵션도 있습니다 .
8. (선택 사항) 보고서의 " 옵션 " 섹션에서 **난수 시드값** 옆에 "1234" 를 입력합니다 .
9. **시작**을 클릭합니다 .

   " 검증 " 열이 데이터 테이블에 추가됩니다 . 모자이크 그림을 생성하여 검증 데이터 집합과 훈련 데이터 집합의 분포를 탐색할 수 있습니다 .
10. **분석 > X 로 Y 적합**을 선택합니다 .
11. 검증을 **Y, 반응**에 할당하고 Gender 를 **X, 요인**에 할당합니다 .
12. **확인**을 클릭합니다 .

그림 12.2 검증 데이터 집합과 훈련 데이터 집합의 Gender 분포

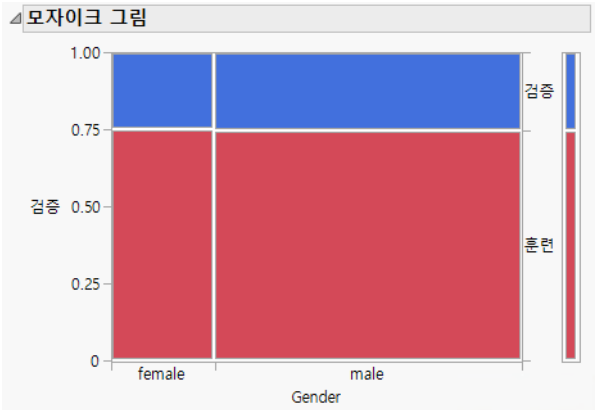
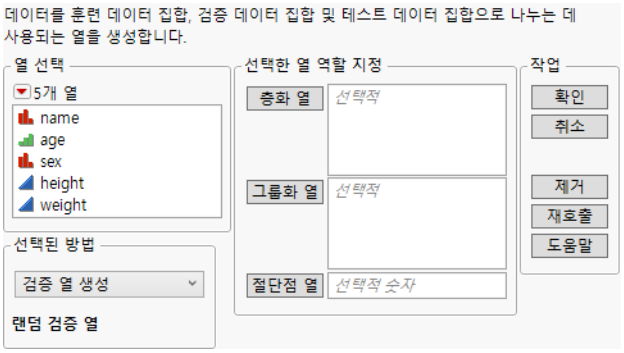


그림 12.2에서는 각 검증 데이터 집합과 훈련 데이터 집합에 대한 Gender 분포를 보여 줍니다. 여성과 남성의 약 75%가 훈련 데이터 집합에 있고 약 25%가 검증 데이터 집합에 있습니다.

JMP PRO 검증 열 생성 플랫폼 시작

분석 > 예측 모델링 > 검증 열 생성을 선택하여 검증 열 생성 플랫폼을 시작합니다.

그림 12.3 검증 열 생성 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

검증 열 생성 시작 창에는 다음과 같은 옵션이 있습니다.

층화 열 하나 이상의 층화 열을 할당합니다.

그룹화 열 하나 이상의 그룹화 열을 할당합니다.

절단점 열 숫자 절단점 열을 할당합니다.

절단점 배치 ID 절단점 열이 할당되면 절단점 배치 ID 열도 할당할 수 있습니다. 이렇게 하면 " 절단점 배치 ID" 열의 각 수준 내에서 절단점 값을 결정할 수 있습니다.

선택된 방법

두 가지 검증 방법을 제공합니다.

검증 열 생성 지정된 층화, 그룹화 및 절단점 열을 기반으로 검증 열을 생성합니다. 지정된 층화, 그룹화 및 절단점 열에 의해 결정된 검증 열 방법이 상자 아래에 설명됩니다. 방법을 선택하고 " 확인 " 을 클릭한 후 " 검증 열 생성 " 보고서에서 각 집합에 대한 할당을 지정합니다. 자세한 내용은 " 비율 또는 상대 비율 지정 " 및 " 절단점 설정 " 에서 확인하십시오. 홀드백 집합을 구성하는 방법에는 5 가지가 있습니다.

랜덤 검증 열 시작 창에서 열을 할당하지 않은 경우 기본 방법입니다. 이 방법은 " 검증 열 생성 " 보고서에 입력한 할당 방식에 따라 데이터 집합을 분할합니다.

층화 검증 열 하나 이상의 층화 열을 할당한 경우 선택된 방법입니다. 이 방법은 지정된 층화 열의 수준에 따라 균일한 비율로 데이터 집합을 분할합니다. " 랜덤 검증 열 " 방법과 마찬가지로 " 검증 열 생성 " 보고서에 입력한 할당 방식에 따라 행이 홀드백 집합에 랜덤 할당됩니다. 그러나 이 방법은 층화 열의 각 수준 또는 수준 조합에서 수행됩니다. 각 훈련, 검증 및 테스트 데이터 집합에서 열 수준을 균일한 비율로 나타내려면 이 방법을 사용합니다.

그룹화 검증 열 하나 이상의 그룹화 열을 지정한 경우 선택된 방법입니다. 이 방법은 지정된 열의 전체 수준 또는 둘 이상 열의 수준 조합이 동일한 집합에 배치되도록 데이터 집합을 분할합니다. 이 때문에 결과 집합의 크기는 사용자가 지정한 크기와 약간 다릅니다. 홀드백 집합 간에 수준을 분할하는 것이 적절하지 않은 경우 이 옵션을 사용합니다.

그룹별 층화 검증 열 층화 열과 그룹화 열을 둘 다 지정한 경우 선택된 방법입니다. 이 방법은 층화 열 간에 수준 균형을 맞추면서 지정된 그룹이 동일한 홀드백 집합에 함께 유지되는 방식으로 데이터를 분할합니다. " 그룹화 검증 열 " 과 마찬가지로 그룹을 지정된 열의 수준 또는 둘 이상 열의 수준 조합으로 생성할 수 있습니다 . 결과 집합의 크기는 사용자가 지정한 크기와 약간 다릅니다.

절단점 검증 열 절단점 열을 지정한 경우 선택된 방법입니다. 이 방법은 시계열 절단점에 따라 데이터 집합을 분할합니다. 기간에 따라 홀드백 집합에 데이터를 할당하려는 경우 이 옵션을 사용합니다. 훈련 데이터 집합은 첫 번째 절단점과 두 번째 절단점 사이의 행으로 구성됩니다. 검증 데이터 집합은 두 번째 절단점과 세 번째 절단점 사이의 행으로 구성됩니다. 테스트 데이터 집합은 나머지 행으로 구성됩니다. 이러한 집합은 " 절단점 설정 " 보고서의 옵션을 기반으로 선택됩니다.

자동 검증 테이블 생성 원래 데이터 테이블의 행에 연결된 원래 데이터 테이블의 복제 행을 포함하는 새 데이터 테이블을 생성합니다. 교차 검증에 사용할 수 있는 새 데이터 테이블에는 다음과 같은 네 개의 추가 열이 있습니다.

Valid Set 원래 데이터에 0 을 할당하고 복제 데이터에 1 을 할당합니다. 이 열의 값은 훈련 데이터 집합과 검증 데이터 집합을 지정합니다. 분석 시작 창의 " 검증 " 역할에 이 열을 사용합니다.

Valid ID 원래 관측값의 행 번호를 할당합니다 . 이렇게 하면 각 원래 관측값에 대한 훈련 데이터 집합 행과 검증 데이터 집합 행을 일치시킬 수 있습니다 .

Valid Weight 분석 시작 창의 " 빈도 " 역할에 사용할 자동 검증 가중치를 할당합니다 . Valid ID 의 각 값에 대해 훈련 관측값과 검증 관측값에 대한 동일한 균등 난수가 생성됩니다 . 훈련 데이터 집합의 경우 Valid Weight 가 다음과 같이 계산됩니다 .

Valid Weight = -log(1 - Valid Uniform)

검증 데이터 집합의 경우 Valid Weight 가 다음과 같이 계산됩니다 .

Valid Weight = -log(Valid Uniform)

Valid Weight 열은 훈련 데이터 가중치가 검증 데이터 가중치와 음의 상관관계를 갖도록 구성됩니다 . 이렇게 하면 검증 데이터의 적합 차이로 인해 적합 방법에 대한 효과적인 교차 검증이 가능합니다 .

Null Factor Valid ID 의 각 값에 대해 동일한 정규 난수를 할당합니다 .

팁 : 부분집합을 훈련 데이터로 사용하면 추정 문제가 발생할 수 있는 작은 데이터 테이블의 경우 " 자동 검증 테이블 생성 " 을 사용하십시오 .

JMP PRO 검증 열 생성 보고서

시작 창에서 " 확인 " 을 클릭하면 " 검증 열 생성 " 보고서 창이 나타납니다 . 보고서에는 선택한 검증 방법에 대한 설명과 비율 , 절단점 , 열 유형 및 난수 시드값 설정을 위한 관련 옵션이 포함되어 있습니다 .

JMP PRO 비율 또는 상대 비율 지정

이 보고서에서는 훈련, 검증 및 테스트 데이터 집합에 대한 할당을 지정할 수 있습니다. "훈련 데이터 집합", "검증 데이터 집합" 및 "테스트 데이터 집합" 옆의 상자에 각 집합에 포함할 행의 수 또는 비율을 나타내는 값을 입력합니다. 기본값은 행의 약 75%를 포함하는 훈련 데이터 집합과 약 25%를 포함하는 검증 데이터 집합을 구성합니다.

데이터 테이블의 행 수와 선택한 검증 방법에 따라 결과 집합의 실제 크기가 사용자가 지정한 크기와 약간 다를 수 있습니다. "조정 비율", "행 수" 및 "그룹 수" 열은 집합의 실제 크기를 제공합니다.

JMP PRO 절단점 설정

이 보고서는 시작 창에서 절단점 열을 지정한 경우에만 표시됩니다. 절단점은 다음 네 가지 방법 중 하나에 의해 결정됩니다.

비율 각 집합에 대해 지정된 행의 비율에 따라 절단점을 결정합니다. "훈련 데이터 집합", "검증 데이터 집합" 및 "테스트 데이터 집합" 옆의 상자에 각 집합에 포함할 비율을 나타내는 값을 입력합니다. 행 수와 비율에 따라 결과 집합의 실제 크기가 사용자가 지정한 크기와 약간 다를 수 있습니다.

행 수 각 집합에 대해 지정된 행 수에 따라 절단점을 결정합니다. "훈련 데이터 집합", "검증 데이터 집합" 및 "테스트 데이터 집합" 옆의 상자에 각 집합에 포함할 행 수를 나타내는 값을 입력합니다. 이 옵션을 사용하면 집합을 정확하게 지정할 수 있습니다.

고정 시간 또는 날짜 지정된 절단점 열의 고정 데이터 점에 따라 절단점을 결정합니다. 이 옵션을 선택하면 절단점 열의 최소값 및 최대값이 참조용으로 표시됩니다. "훈련 데이터 집합", "검증 데이터 집합" 및 "테스트 데이터 집합" 옆의 상자에 각 집합에 포함할 최소값을 나타내는 값을 입력합니다.

경과 시간 절단점 열의 첫 번째 타임스탬프 이후 경과한 시간을 기준으로 절단점을 결정합니다. 이 옵션을 선택하면 첫 번째 값에서 마지막 값까지 총 경과 시간이 참조용으로 표시됩니다. "훈련 데이터 집합", "검증 데이터 집합" 및 "테스트 데이터 집합" 옆의 상자에 각 집합에 포함할 경과 시간을 나타내는 값을 입력합니다.

JMP PRO 옵션

다음과 같은 추가 옵션을 지정할 수도 있습니다.

새 열 이름 검증 열 이름을 지정할 수 있습니다.

검증 열 유형 ("절단점 검증 열" 방법에는 사용 불가능) 검증 열 유형을 "고정" 또는 "계산식"으로 지정할 수 있습니다.

- 고정 열 유형은 고정 값을 포함하는 열을 생성합니다.
- 계산식 열 유형은 계산식 열을 검증 열로 생성합니다.

난수 시드값 ("절단점 검증 열" 방법을 사용하거나 계산식 열 유형을 지정한 경우에는 사용 불가능) 이후에 동일한 검증 열을 재현하기 위해 난수 시드값을 설정할 수 있습니다.

JMP PRO 시작

"검증 열 생성" 보고서에서 **시작**을 클릭하면 검증 열이 데이터 테이블에 추가됩니다. 검증 열에는 층화, 그룹화 또는 절단점 변수를 제공하는 "노트" 특성이 있습니다.

JMP PRO 검증 열 생성 플랫폼 옵션

" 검증 열 생성 " 유틸리티에는 다음과 같은 빨간색 삼각형 메뉴 옵션이 있습니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

JMP PRO 검증 열 생성 플랫폼의 추가 예

이 예에서는 미국 전역의 16 개 기상 관측소에서 1 년 동안 수집한 주간 기상 데이터를 사용합니다 . 데이터 테이블의 "Weather Station Locations" 스크립트를 실행하여 위치 지도를 표시합니다 . 모든 기상 관측소가 매주 (년 중) 에 대한 주간 온도 측정값을 가지고 있는 것은 아닙니다 . 데이터 수집 날짜를 기반으로 이 데이터 테이블에 대한 검증 열을 생성하는 것이 목적입니다 . 각 기상 관측소에 대해 처음 60% 의 관측값을 훈련에 사용하고 , 다음 25% 관측값을 검증에 사용하고 , 마지막 15% 관측값을 테스트에 사용하려고 합니다 . 이 예에서는 이 유형의 시나리오에서 배치 ID 열 사용의 중요성을 보여 줍니다 .

절단점을 사용하여 검증 열 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Functional Data/Weekly Weather Data.jmp 를 엽니다 .
2. **분석 > 예측 모델링 > 검증 열 생성**을 선택합니다 .
3. Week of Year 를 선택하고 **절단점 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 절단점 결정 기준 " 옆의 목록에서 **비율**을 선택합니다 .
6. " 훈련 데이터 집합 ", " 검증 데이터 집합 " 및 " 테스트 데이터 집합 " 옆의 상자에 각각 "0.60", "0.25", "0.15" 를 입력합니다 .
7. " 새 열 이름 " 옆의 상자에 "Cutpoint Validation" 을 입력합니다 .
8. **시작**을 클릭합니다 .

   "Cutpoint Validation" 이라는 검증 열이 데이터 테이블에 추가됩니다 .
9. **분석 > 테이블 생성**을 선택합니다 .
10. ID 를 클릭하고 " 행에 대한 놓기 영역 " 으로 드래그합니다 .
11. Cutpoint Validation 을 클릭하고 N 위로 드래그합니다 .

12. 행 % 를 클릭하고 셀 위로 드래그합니다 .

그림 12.4 절단점 검증 열 비율

	Cutpoint Validation		
	훈련	검증	테스트
ID	행 %	행 %	행 %
Bristol, TN	62.79%	20.93%	16.28%
Corpus Christi, TX	70.27%	18.92%	10.81%
Denver, CO	58.14%	23.26%	18.60%
Goodwell, OK	64.71%	17.65%	17.65%
Grand Canyon National Park, AZ	53.66%	26.83%	19.51%
Greenville, ME	60.47%	20.93%	18.60%
Harrisburg, PA	60.53%	26.32%	13.16%
Helena, MT	60.00%	25.00%	15.00%
Lynchburg, VA	71.79%	17.95%	10.26%
Madera, CA	67.50%	22.50%	10.00%
Miami Beach, FL	63.16%	26.32%	10.53%
Minneapolis/St. Paul, MN	57.50%	27.50%	15.00%
N. Myrtle Beach, SC	69.05%	16.67%	14.29%
Peason Ridge Range, LA	57.89%	28.95%	13.16%
Philip, SD	59.52%	23.81%	16.67%
Terra Haute, IN	47.22%	33.33%	19.44%

그림 12.4 에서는 일부 기상 관측소에서 제공하는 훈련 , 검증 및 테스트 데이터 집합 비율이 정확하지 않음을 보여 줍니다 . 정확한 백분율을 얻으려면 배치 ID 열을 사용합니다 .

절단점과 배치 ID 를 사용하여 검증 열 생성

1. **분석 > 예측 모델링 > 검증 열 생성**을 선택합니다 .
2. Week of Year 를 선택하고 **절단점 열**을 클릭합니다 .
3. ID 를 선택하고 **절단점 배치 ID** 를 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 절단점 결정 기준 " 옆의 목록에서 **비율**을 선택합니다 .
6. " 훈련 데이터 집합 ", " 검증 데이터 집합 " 및 " 테스트 데이터 집합 " 옆의 상자에 각각 "0.60", "0.25", "0.15" 를 입력합니다 .
7. " 새 열 이름 " 옆의 상자에 "Cutpoint Batch Validation" 을 입력합니다 .
8. **시작**을 클릭합니다 .

   "Cutpoint Batch Validation" 이라는 검증 열이 데이터 테이블에 추가됩니다 .
9. **분석 > 테이블 생성**을 선택합니다 .
10. ID 를 클릭하고 " 행에 대한 놓기 영역 " 으로 드래그합니다 .
11. Cutpoint Batch Validation 을 클릭하고 N 위로 드래그합니다 .
12. 행 % 를 클릭하고 셀 위로 드래그합니다 .

그림 12.5 배치 ID 비율을 사용한 절단점 검증 열

	Cutpoint Batch Validation		
	훈련	검증	테스트
ID	행 %	행 %	행 %
Bristol, TN	60.47%	25.58%	13.95%
Corpus Christi, TX	59.46%	24.32%	16.22%
Denver, CO	60.47%	25.58%	13.95%
Goodwell, OK	58.82%	26.47%	14.71%
Grand Canyon National Park, AZ	60.98%	24.39%	14.63%
Greenville, ME	60.47%	25.58%	13.95%
Harrisburg, PA	60.53%	26.32%	13.16%
Helena, MT	60.00%	25.00%	15.00%
Lynchburg, VA	58.97%	25.64%	15.38%
Madera, CA	60.00%	25.00%	15.00%
Miami Beach, FL	60.53%	26.32%	13.16%
Minneapolis/St. Paul, MN	60.00%	25.00%	15.00%
N. Myrtle Beach, SC	59.52%	26.19%	14.29%
Peason Ridge Range, LA	60.53%	26.32%	13.16%
Philip, SD	59.52%	26.19%	14.29%
Terra Haute, IN	61.11%	25.00%	13.89%

그림 12.5 에서는 절단점 배치 ID 열을 사용하면 각 기상 관측소의 훈련 , 검증 및 테스트 데이터 집합에 대한 비율이 지정된 값에 훨씬 가깝다는 것을 보여 줍니다 .

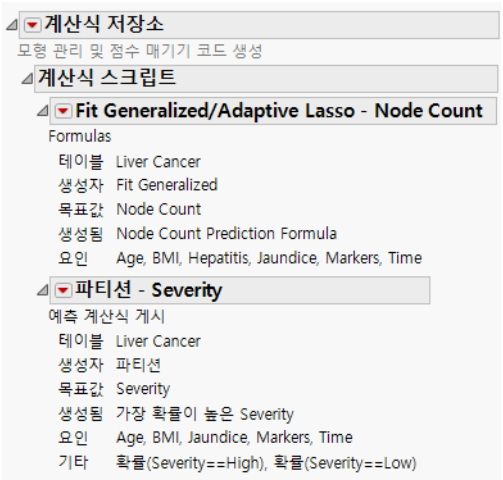
13 장

JMP PRO 계산식 저장소
모형 관리 및 스코어링 코드 생성

계산식 저장소 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

계산식 저장소는 모형을 구성, 비교 및 프로파일링하기 위한 저장소입니다. 계산식 저장소에 게시된 모형에 대해 JMP 내부 또는 외부 배포를 위한 스코어링 코드를 생성할 수 있습니다. 모형 탐색 작업의 경우 계산식 저장소를 사용하여 후보 모형을 JMP 데이터 테이블 외부에 저장할 수 있습니다. 계산식 저장소에서 모형 프로파일러 및 모형 비교 플랫폼에 액세스할 수 있습니다. 나중에 사용하기 위해 선택한 모형을 현재 JMP 테이블에 저장하거나 스코어링할 새 데이터와 함께 JMP 테이블에 저장할 수 있습니다. C, Python, JavaScript, SAS 또는 SQL 로 스코어링 코드를 생성하면 JMP 외부 환경에서 사용할 수 있습니다

그림 13.1 계산식 저장소의 예



목차

계산식 저장소 플랫폼 개요 197
계산식 저장소의 예 197
계산식 저장소 플랫폼 시작 198
예측 계산식을 계산식 저장소에 게시하는 플랫폼 198
계산식 저장소 플랫폼 옵션 199
계산식 저장소 모형 옵션 200
계산식 저장소 플랫폼에서 스코어링 코드 생성 201

JMP PRO 계산식 저장소 플랫폼 개요

계산식 저장소는 배포할 모형을 구성, 비교, 프로파일링 및 스코어링하기 위한 저장소입니다. 모형은 예측 계산식입니다. 예측 계산식은 계산식 저장소에 열 스크립트로 저장됩니다. 예측 계산식을 JMP 테이블에 추가하여 데이터를 스코어링할 수 있습니다. 또한 계산식 저장소를 사용하여 예측 계산식에 대한 스코어링 코드를 생성하면 JMP 외부 환경에서 모형을 쉽게 배포할 수 있습니다.

계산식 저장소를 사용하면 다음 작업을 수행할 수 있습니다 .

- 예측 계산식을 데이터 테이블이 아닌 다른 곳에 저장
- 여러 데이터 테이블의 예측 계산식을 공통 위치에 저장
- 예측 계산식 내에 중간 정리 계산식 저장
- 모형 비교
- 모형 프로파일링
- JMP 에서 새 데이터를 스코어링하기 위해 데이터 테이블에 예측 계산식 추가
- JMP 외부에 모형을 배포하기 위한 스코어링 코드 (C, Python, JavaScript, SAS 또는 SQL) 생성
- 계산식 저장소는 생성된 스코어링 코드에 중간 계산식을 포함합니다. 열 계산식이 있는 입력을 통해 생성된 모형은 해당 계산식을 스코어링 코드에 포함합니다 . 예를 들어 모형에 재코딩된 변수가 있는 경우 재코딩 계산식이 스코어링 코드에 포함됩니다. 이렇게 하면 처리되지 않은 원시 데이터에 예측 계산식을 적용할 수 있습니다 .

참고 : 결측값 정보화는 지원되는 중간 계산식이 아닙니다 .

JMP PRO 계산식 저장소의 예

Liver Cancer.jmp 샘플 데이터 테이블에는 연구에 참여한 환자의 간암 중증도에 대한 데이터가 포함되어 있습니다 . 모형에 대한 여러 테이블 스크립트도 파일에 포함되어 있습니다 . 이 예에서는 이러한 스크립트를 사용하여 계산식 저장소를 설명하기 위한 모형을 생성합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Liver Cancer.jmp 를 엽니다 .
2. "Lasso Poisson, Validation Column" 스크립트 옆의 녹색 삼각형을 클릭합니다 .
3. " 검증 열을 사용한 Poisson 적응형 Lasso 회귀 " 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 게시**를 선택합니다 .

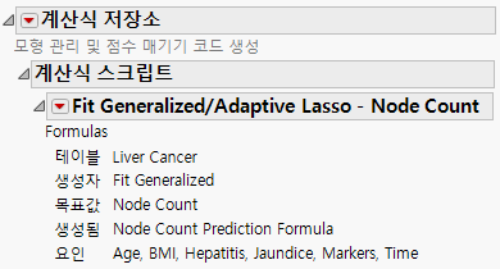
   이 옵션을 선택하면 일반화 적합 모형에 대한 예측 계산식이 포함된 계산식 저장소가 열립니다 .
4. 예측 계산식을 데이터 테이블에 추가하려면 계산식 저장소의 " 일반화 적합 - Node Count" 빨간색 삼각형에서 **스크립트 실행**을 선택합니다 .

참고 : 3 단계에서 **열 저장 > 예측 계산식 저장**을 사용하면 3 단계와 4 단계의 결과를 얻을 수 있습니다 . 그러나 이 경우에는 예측 계산식이 계산식 저장소에 포함되지 않습니다 .

5. JMP 외부에서 사용할 스코어링 코드를 생성하려면 " 일반화 적합 - Node Count" 의 빨간색 삼각형에서 코드 유형을 선택합니다 . 코드가 포함된 스크립트 창이 나타납니다 . 자세한 내용은 " 계산식 저장소 플랫폼에서 스코어링 코드 생성 " 에서 확인하십시오 .
6. 계산식 저장소를 저장하려면 **파일 > 저장**을 선택합니다 .

참고 : 계산식 저장소는 JMP 보고서 파일 (*.jrp) 로 저장됩니다 .

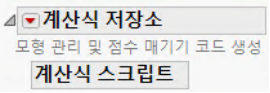
그림 13.2 일반화 모형이 있는 계산식 저장소



JMP PRO 계산식 저장소 플랫폼 시작

분석 > 예측 모델링 > 계산식 저장소를 선택하여 계산식 저장소를 시작합니다 .

그림 13.3 시작 창의 빈 계산식 저장소



또는 열려 있는 계산식 저장소가 없는 경우 " 게시 " 명령을 선택하면 계산식 저장소가 열립니다 .

JMP PRO 예측 계산식을 계산식 저장소에 게시하는 플랫폼

예측 계산식을 게시하고 스코어링 코드를 생성하는 플랫폼은 다음과 같습니다 .

- 판별
- 최소 제곱 회귀
- 로지스틱 회귀
- 파티션
- Uplift

- K 최근접 이웃
- Naive Bayes
- 신경망
- 잠재 계층 분석
- 주성분
- 일반화 회귀
- PLS
- 가우스 과정

예측 계산식을 계산식 저장소에 게시하지 않는 플랫폼에서는 예측 계산식을 데이터 테이블에 저장할 수 있습니다. 데이터 테이블에서 **열에서 계산식 추가**를 선택하여 예측 계산식을 계산식 저장소에 추가합니다. 그러나 스코어링 코드는 이러한 모형에 대해 올바르게 작동하지 않을 수 있습니다. 스코어링 코드에 대한 자세한 내용은 "계산식 저장소 플랫폼에서 스코어링 코드 생성"에서 확인하십시오.

또한 계산식 편집기에서 직접 계산식을 게시할 수 있습니다. 계산식을 선택하고 마우스 오른쪽 버튼을 클릭한 후 **계산식 저장소에 게시**를 선택합니다.

JMP PRO 계산식 저장소 플랫폼 옵션

열에서 계산식 추가 데이터 테이블 열의 기존 예측 계산식을 현재 계산식 저장소에 추가할 수 있습니다. 테이블에 계산식이 하나뿐인 경우 해당 계산식이 저장소에 추가됩니다. 열 계산식이 여러 개인 경우 대화상자에서 원하는 계산식을 선택할 수 있습니다 . 데이터 테이블에서 계산식 열이 선택되면 대화상자에서 해당 열이 미리 선택됩니다.

스크립트 표시 현재 계산식 저장소에 있는 모든 계산식에 대한 스크립트가 포함된 새 " 계산식 창 " 을 열거나, 열려 있는 " 계산식 창 " 에 추가합니다.

스크립트 복사 현재 계산식 저장소의 모든 스크립트를 클립보드에 복사합니다.

계산식을 함수로 복사 현재 계산식 저장소에서 모형을 선택하여 JSL 에서 사용할 함수로 복사할 수 있습니다. 선택한 모형은 클립보드에 복사됩니다. 함수는 열 대신 스칼라 변수에 대해 작동합니다. 선택한 각 모형에 대해 하나의 네임스페이스가 생성됩니다. 각 네임스페이스에는 모형 입력을 모수로 사용하고 모형 예측 변수를 반환하는 predict 함수 하나가 있습니다.

계산식을 변환으로 복사 현재 계산식 저장소에서 복사할 모형을 선택할 수 있습니다. 선택한 모형은 클립보드의 `Transform Columns()` 문 내에 복사됩니다.

스크립트 실행 현재 계산식 저장소의 모형을 JMP 데이터 테이블의 새 열에 저장할 수 있습니다. 중간 열 계산식이 있는 계산식은 중간 열을 JMP 데이터 테이블에 추가합니다.

참고 : 데이터 테이블이 두 개 이상 열려 있으면 " 스크립트 실행 " 에 사용할 데이터 테이블을 지정할 수 있는 " 테이블 선택 " 창이 나타납니다.

C 코드 생성 , Python 코드 생성 , JavaScript 코드 생성 , SAS 코드 생성 및 SQL 코드 생성 현재 계산식 저장소에서 코드 생성을 위한 모형을 선택할 수 있습니다 . 선택한 모형에 대한 각 스코어링 코드 (C, Python, JavaScript, SAS DS2 또는 SQL 형식) 가 포함된 새 스크립트 창이 나타납니다 . 이 코드를 사용하면 선택한 환경 또는 프레임워크에서 모형을 쉽게 배포할 수 있습니다 . 자세한 내용은 " 계산식 저장소 플랫폼에서 스코어링 코드 생성 " 에서 확인하십시오 .

모형 비교 현재 계산식 저장소에서 모형 비교 유틸리티를 사용하여 비교할 모형을 선택할 수 있습니다 . 여러 테이블의 모형이 저장소에 저장된 경우 먼저 관심 테이블을 선택한 후 관심 모형을 선택합니다 . Shift 키를 누른 채 " 모형 비교 " 옵션을 선택하면 새 창에서 모형 비교가 시작됩니다 . 자세한 내용은 " 모형 비교 " 에서 확인하십시오 .

참고 : 데이터 테이블이 두 개 이상 열려 있으면 " 모형 비교 " 에 사용할 데이터 테이블을 지정할 수 있는 " 테이블 선택 " 창이 나타납니다 .

모형 비교 제거 모든 모형 비교 보고서를 현재 계산식 저장소에서 제거합니다 .

프로파일러 현재 계산식 저장소에서 프로파일러를 사용하여 프로파일링할 모형을 선택할 수 있습니다 . 여러 테이블의 모형이 저장소에 저장된 경우 먼저 관심 테이블을 선택한 후 관심 모형을 선택합니다 . Shift 키를 누른 채 " 프로파일러 " 옵션을 선택하면 새 창에서 프로파일러가 시작됩니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

참고 : 데이터 테이블이 두 개 이상 열려 있으면 " 프로파일러 " 에 사용할 데이터 테이블을 지정할 수 있는 " 테이블 선택 " 창이 나타납니다 .

프로파일러 제거 모든 프로파일러를 현재 계산식 저장소에서 제거합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

JMP PRO 계산식 저장소 모형 옵션

계산식 저장소에 저장된 각 예측 계산식에는 개별 옵션 메뉴가 있습니다 . 메인 메뉴 옵션에 해당하는 표시, 복사, 실행 및 코드 생성 옵션 외에 다음과 같은 개별 메뉴 옵션이 포함되어 있습니다.

새 열 이름 바꾸기 모형 이름을 바꿀 수 있습니다. 스크립트를 실행하여 예측 계산식을 JMP 데이터 테이블에 추가하면 이 새 이름이 열 이름으로 적용됩니다 . 또한 생성된 코드에도 이 이름이 포함됩니다 .

제거 계산식 저장소에서 모형을 제거합니다 . 이 명령은 실행 취소할 수 없습니다 .

JMP PRO 계산식 저장소 플랫폼에서 스코어링 코드 생성

스코어링 코드를 생성하는 목적은 JMP 에서 생성된 모형을 프로덕션 환경 또는 선택한 기타 프레임워크에서 쉽게 사용하기 위해서입니다 . 많은 플랫폼에서 예측 계산식을 계산식 저장소에 게시하지만 모든 예측 계산식이 전체 코드를 생성하는 것은 아닙니다 . 코드는 전체 코드이거나 코드 조각일 수 있으며 , 구현을 위해 추가 프로그래밍이 필요한 지원되지 않는 함수를 포함할 수도 있습니다 . 대상 언어로 추가 프로그래밍이 필요한 지원되지 않는 함수 호출을 나타내는 오류 메시지가 표시되는 경우가 종종 있습니다 . 좀 더 복잡한 함수를 구현하기 위해 대상 언어별 라이브러리가 필요할 수도 있습니다 . 이러한 라이브러리는 아래의 언어별 섹션에서 설명합니다 . 예를 들어 다음과 같은 작업을 수행할 수 있습니다 .

- 생성된 SAS 코드를 사용하여 모형을 SAS 모형 관리자에 배포합니다 .
- 생성된 SQL 코드를 사용하여 데이터베이스 내 스코어링으로 ETL 프로세스를 확대합니다 .
- 생성된 C 코드로 빌드된 응용 프로그램을 사용하여 데이터 변환 파이프라인에 대한 노드를 생성합니다 .
- 생성된 Python 코드를 사용하여 실시간 스코어링 결과를 표시할 Jupyter 노트를 생성합니다 .
- 생성된 JavaScript 를 포함하는 웹 응용 프로그램으로 고객이 자신의 데이터를 직접 스코어링할 수 있습니다 .

C, Python 및 JavaScript 언어의 경우 생성된 코드를 배포하거나 컴파일할 때 .h 파일 및 유틸리티 라이브러리와 같은 지원 코드를 포함해야 합니다. 이러한 파일은 JMP 설치 폴더 아래의 Scoring 폴더에 있습니다.

팁: 사용자 컴퓨터에서 설치 폴더 위치를 찾으려면 JSL 명령 `Get Path Variable("$ALL_HOME/Scoring");` 을 사용하십시오 .

C 코드 생성된 C 스코어링 코드는 라이브러리로 컴파일된 후 응용 프로그램에 연결되어야 합니다 . 정적 연결 또는 동적 연결 방식을 사용할 수 있습니다 . 컴파일 및 연결에 필요한 jmp_lib.h, jmp_parms.h, jmp_score.h 파일은 JMP 설치 폴더 아래의 Scoring/C 폴더에서 찾을 수 있습니다 .

Python 코드 Python 스코어링 응용 프로그램을 실행하는 데 필요한 jmp_score.py 파일은 JMP 설치 폴더 아래의 Scoring/Python 폴더에서 찾을 수 있습니다 . 모형을 배포할 때 jmp_score.py 를 실행 가능한 스코어링 코드와 동일한 위치에 복사합니다 .

생성된 코드를 호출하여 데이터를 스코어링하는 경우 입력 데이터가 JMP 에서 생성된 코드와 호환되는 데이터 구조로 전달되어야 합니다 . Python 스코어링 코드는 입력 및 출력 인수가 사전과 유사한 유형이어야 합니다 . Spark, Dask 및 Pandas 프레임워크의 DataFrame 행 개체 또는 표준 Python "dict" 클래스와 같은 키를 사용하여 값에 액세스하고 쓸 수 있습니다 .

선형 대수 연산자 또는 NumPy Python 라이브러리가 필요한 고급 연산자 (예 : Vec Quadratic, Design Norm) 를 사용합니다 .

JavaScript 코드 JavaScript 스코어링 응용 프로그램을 실행하는 데 필요한 jmp_score.js 파일은 Scoring/JavaScript 폴더에서 찾을 수 있습니다 .

SAS 코드 PROC DS2 문으로 래핑한 생성된 코드 조각은 SAS In-Database Code Accelerator 를 포함한 SAS 응용 프로그램에서 사용할 수 있습니다 . 코드는 SAS 창에 추가되며 코드 앞의 주석 블록에 변수 이름 매핑 섹션을 포함합니다 . 중간 열과 관련된 임시 변수 이름은 drop 명령을 통해 DATA 스텝 끝에서 삭제됩니다 .

팁 : 범주형 반응에 대한 로지스틱 또는 신경망 모형과 같이 ifmax 호출이 있는 모형의 경우 임시 변수 선언을 **method run()** 문 앞으로 이동하십시오 .

SQL 코드 select 문으로 래핑된 SQL 코드 조각은 대부분의 주 데이터베이스 서버에 대한 SQL 쿼리에 사용할 수 있습니다 .

참고 : " 자리 표시자 " 또는 "ERROR" 가 포함된 코드는 지원되지 않는 함수 호출을 나타냅니다 .

14 장

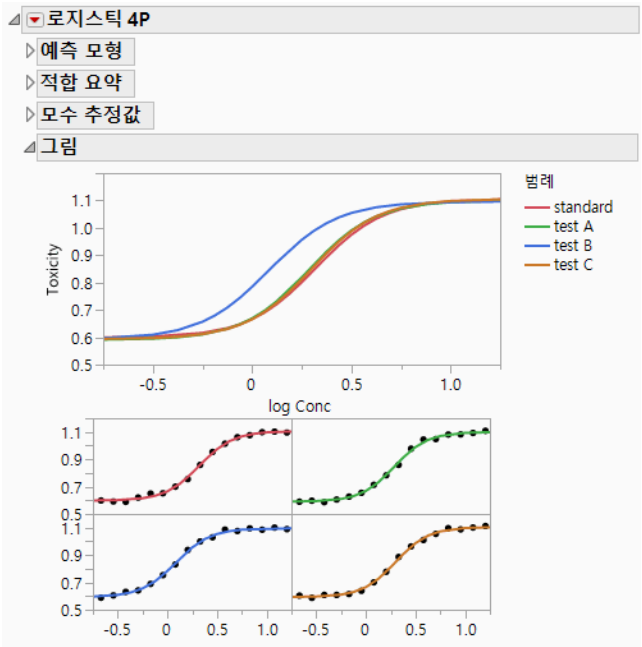
곡선 적합

기본 제공 비선형 모형을 데이터에 적합

많은 경우, 특히 물리학과 생물학에서 잘 알려진 비선형 방정식은 변수 간의 관계를 설명합니다. 예를 들어 약물 생리학 실험에서는 약물에 대한 반응 강도가 어떻게 변하는지를 약물 농도 함수로 나타낼 수 있습니다. 흔히 시그모이드 곡선은 반응 강도를 약물 농도 함수로 정확하게 모델링합니다. 또한 시간 경과에 따른 인구 규모를 모델링할 수 있는 지수 성장 곡선도 예로 들 수 있습니다.

곡선 적합 플랫폼에서는 모수 추정값의 시작 값을 지정하거나 모형 계산식을 생성할 필요가 없습니다. 시작 값을 지정하고 모형 계산식을 생성하려면 모든 비선형 모형을 적합시킬 수 있는 더 강력한 사용자 비선형 플랫폼을 사용합니다. 자세한 내용은 "비선형 회귀"에서 확인하십시오.

그림 14.1 곡선 적합 플랫폼에서 비선형 적합의 예



목차

곡선 적합 플랫폼 개요 205
곡선 적합 플랫폼을 사용한 예 205
곡선 적합 플랫폼 시작 208
곡선 적합 보고서 209
곡선 적합 옵션 210
    모형 비교 보고서 211
모형 적합 보고서 212
모형 적합 옵션 212
    병렬 검정 214
    모수 추정값 비교 216
    동등성 검정 217
곡선 적합 플랫폼에 대한 통계 상세 정보 219
    모형 계산식 219

곡선 적합 플랫폼 개요

일부 모형은 모수에 대해 **선형**(예: 2차 또는 기타 다항식)이고, 다른 모형은 선형이 되도록 변환(예: x의 로그 변환을 사용하는 경우)할 수 있습니다. 이러한 상황에서는 모형 적합 또는 X로 Y 적합 플랫폼이 더 적절합니다. **선형 모형 적합**의 에 나오는 예에서는 산소 섭취량과 달리기 시간 사이의 유의한 선형 관계를 보여 줍니다. 모형 적합에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오. X로 Y 적합에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오.

곡선 적합 플랫폼을 사용하면 모수에 대해 **비선형**인 모형을 적합시킬 수 있습니다. 이 장의 처음에 나오는 예에서는 비선형 관계 분석(농도 함수로 나타낸 약물 독성)을 보여 줍니다. 독성에 대한 농도 효과는 낮은 투여량에서 높은 투여량으로 변하므로 이 관계는 비선형입니다.

다음은 선형 및 비선형 함수에 대한 방정식 예입니다.

선형 함수: $Y = \beta_0 + \beta_1 e^{x}$

비선형 함수: $Y = \beta_0 + \beta_1 e^{\beta_2 x}$

곡선 적합 플랫폼에서는 다항식, 로지스틱, 프로빗, Gompertz, 지수, 정상점 및 약동학 모형과 같은 사전 정의된 모형을 제공합니다. 그룹화 변수를 지정하면 그룹화 변수의 각 수준에 대해 개별 모형 모수를 추정할 수 있습니다. 적합 모형과 추정 모수는 그룹화 변수의 전체 수준에서 비교할 수 있습니다.

곡선 적합을 사용하면 예측 계산식을 생성하기 위한 모형을 만들 수도 있습니다. 그런 다음 비선형의 모수 상한과 하한을 설정합니다. 자세한 내용은 "모수 한계 설정의 예"에서 확인하십시오. 또한 JMP Pro에서는 곡선 적합 플랫폼 내에서 보조 변수 집합을 지정하고 일반화 회귀 모형을 적합시켜 이러한 변수가 반응에 어떤 영향을 미치는지 확인할 수 있습니다.

곡선 적합 플랫폼을 사용한 예

이 예에서는 약물 농도의 함수로 나타낸 독성에 대한 모형을 생성하는 방법을 보여 줍니다. standard 약물 제제가 있고 이를 세 개의 새로운 제제와 비교하려고 합니다.

각 약물의 특정 농도에서 생존 세포에 대한 비생존 세포의 독성 비율에 관심이 있습니다. 사전 조사를 통해 각 약물 제제의 16개 농도에 대한 독성 비율을 알고 있습니다. 비율이 높으면 독성이 더 많음을 나타내며 이는 약물 개발에 해로울 수 있습니다. 농도 값 범위를 줄이고 곡선의 차이를 더 쉽게 감지할 수 있도록 로그 농도가 계산되었습니다.

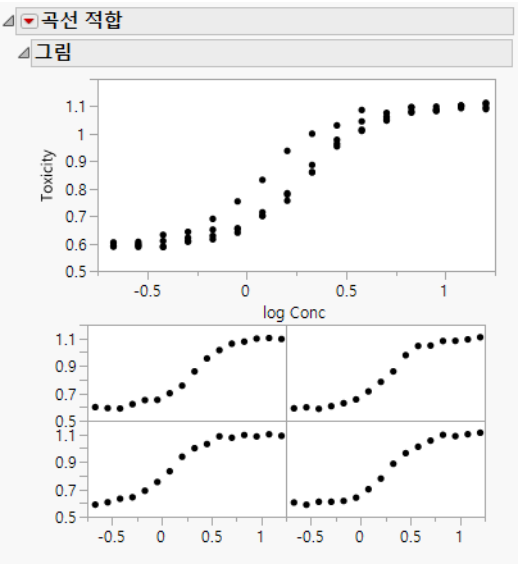
다음 단계에 따라 모형을 생성합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nonlinear Examples/Bioassay.jmp 를 엽니다.
2. **분석 > 전문 모델링 > 곡선 적합**. 을 선택합니다.

3. Toxicity 를 **Y, 반응** 역할에 할당합니다 .
4. log Conc 를 **X, 회귀변수** 역할에 할당합니다 .
5. Formulation 을 **그룹** 역할에 할당합니다 .
6. **확인**을 클릭합니다 .

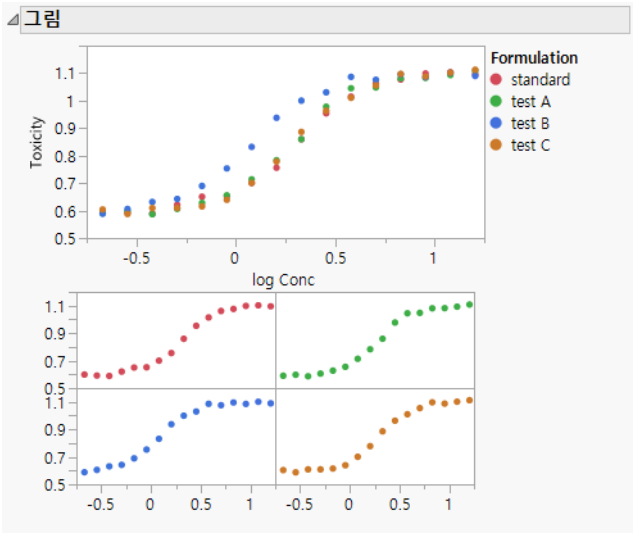
   " 곡선 적합 " 보고서가 나타납니다 . " 그림 " 보고서에는 각 제제의 적합 모형에 대한 중첩 그림이 포함되어 있습니다 .

그림 14.2 초기 곡선 적합 보고서



7. 각 약물 제제를 식별하는 범례를 표시하려면 그래프 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **행 범례**를 선택합니다 . **Formulation** 열을 선택하고 **확인**을 클릭합니다 .

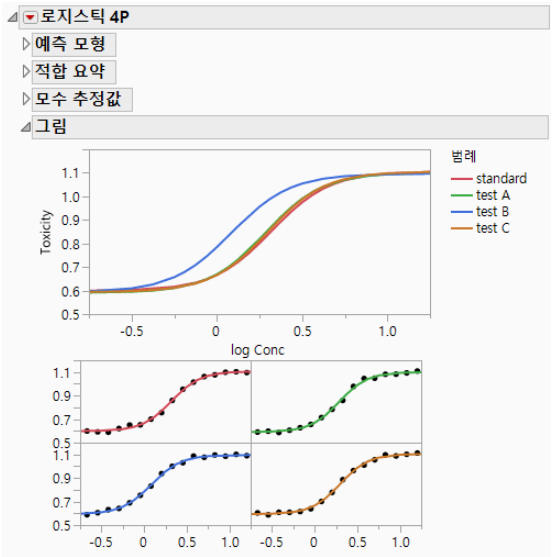
그림 14.3 그림 범례가 포함된 곡선 적합 보고서



곡선이 S 형태로 나타나므로 시그모이드 곡선이 적합할 수 있습니다. 표 14.1 에서는 곡선 적합 플랫폼에서 제공하는 여러 유형의 모형에 대한 계산식과 그래픽 표현을 보여 줍니다.

8. "곡선 적합"의 빨간색 삼각형을 클릭하고 **시그모이드 곡선 > 로지스틱 곡선 > 로지스틱 4P 적합**을 선택합니다.

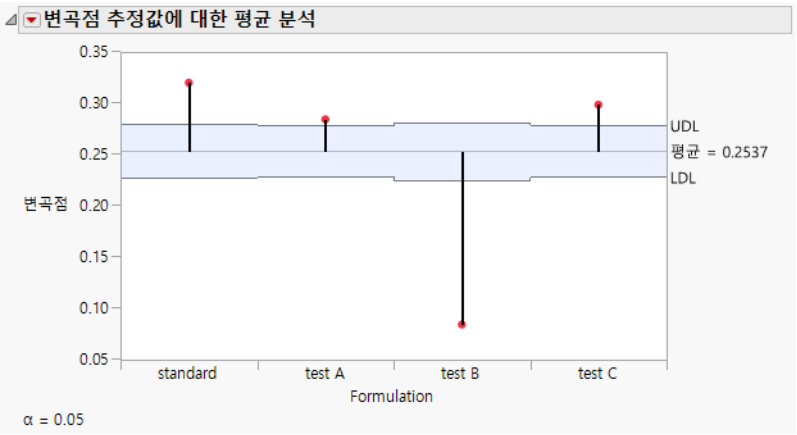
그림 14.4 로지스틱 4P 보고서



" 로지스틱 4P" 보고서가 나타납니다 . 각 약물 제제에 대한 별도의 그림도 있습니다 . 적합 곡선 그림을 보면 test B 곡선이 다른 곡선에 비해 빠르게 상승하므로 B 제제가 다를 수 있음을 알 수 있습니다 . 변곡점 모수로 인해 상승이 발생합니다 .

9. " 로지스틱 4P" 의 빨간색 삼각형을 클릭하고 **모수 추정값 비교**를 선택합니다 .

그림 14.5 모수 비교 보고서

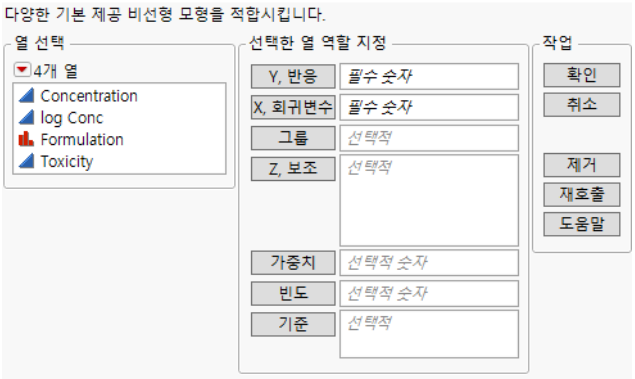


test B 제제에 대한 변곡점 모수가 평균 변곡점보다 유의하게 낮습니다. 이는 그림 14.4에 표시된 그림과 일치합니다. 약물 제제 B는 다른 제제보다 독성 비율이 더 높습니다.

곡선 적합 플랫폼 시작

곡선 적합 플랫폼을 시작하려면 **분석 > 전문 모델링 > 곡선 적합**을 선택합니다 . 시작 창이 그림 14.6 에 표시되어 있습니다 .

그림 14.6 곡선 적합 플랫폼 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

곡선 적합 플랫폼 시작 창에는 다음과 같은 기능이 있습니다.

Y, 반응 *Y* 변수를 지정합니다.

X, 회귀변수 *X* 변수를 지정합니다.

그룹 그룹화 변수를 지정합니다. 적합 모형에는 그룹화 변수의 각 수준에 대한 개별 모수가 있습니다. 이를 통해 그룹화 변수의 전체 수준에서 적합 모형과 추정 모수를 비교할 수 있습니다.

Z, 보조 하나 이상의 보조 변수를 지정합니다. 보조 변수는 곡선 적합 플랫폼의 계산에 사용되지 않으며 이 변수를 포함해도 결과에 영향을 주지 않습니다. 보조 변수는 향후 곡선 적합 결과를 분석하는 데 사용할 수 있는 변수입니다. 보조 변수를 지정하면 " 모수 테이블 생성 " 옵션으로 생성된 테이블에 포함됩니다. JMP Pro 에서 하나 이상의 보조 변수와 그룹화 변수를 지정하면 적합 모형에 대해 " 곡선 DOE 분석 " 옵션을 사용할 수 있습니다. 자세한 내용은 " 곡선 DOE 분석 " 에서 확인하십시오.

가중치 관측값의 가중치를 포함하는 변수를 지정합니다.

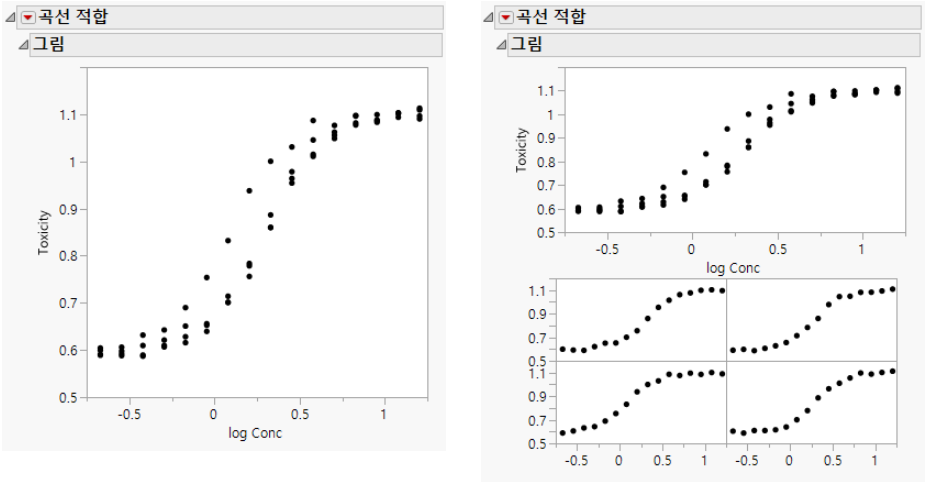
빈도 관측값의 빈도를 포함하는 변수를 지정합니다.

기준 변수의 모든 수준에 대해 개별 분석을 수행할 변수를 지정합니다.

곡선 적합 보고서

처음에는 " 곡선 적합 " 보고서에 Y 대 X 그림만 포함됩니다. " 그룹 " 변수를 지정하면 적합 모형의 각 그룹에 대한 중첩 및 개별 그림이 보고서에 포함됩니다.

그림 14.7 곡선 적합 보고서 : 그룹화 변수 없음 (왼쪽), 그룹 변수 있음 (오른쪽)



모형을 적합시키면 적합 모형이 그림에 나타납니다(플랫폼 시작 창에서 그룹화 변수를 지정하지 않은 경우). 그림 위에 "모형 비교" 보고서가 나타나고 그림 아래에 모형 보고서가 나타납니다. 새 모형이 적합될 때마다 해당 모형이 "모형 비교" 보고서에 나타나고 새 모형 보고서가 보고서 창에 나타납니다. 자세한 내용은 "모형 비교 보고서" 및 "모형 적합 보고서"에서 확인하십시오.

참고 : 데이터 테이블에서 제외된 모든 행은 곡선 적합 그림에서도 숨겨집니다 .

곡선 적합 옵션

" 곡선 적합 " 의 빨간색 삼각형 메뉴에서 다음과 같은 기본 제공 모형 중 하나를 선택합니다 . 자세한 내용은 " 모형 계산식 " 에서 확인하십시오 .

다항 1 차 ~ 5 차 다항식 모형을 적합시킵니다 .

시그모이드 곡선 로지스틱 , 프로빗 , Gompertz 및 Weibull 모형을 적합시킵니다 . 이러한 모형은 S 형태이며 상위 점근선과 하위 점근선이 모두 있습니다 . 로지스틱 2P, 3P 및 4P 모형과 프로빗 2P 및 4P 모형은 대칭입니다 . 로지스틱 5P 와 두 개의 Gompertz 모형은 대칭이 아닙니다 . 로지스틱 2P 는 반응이 0 에서 1 사이인 경우에만 사용할 수 있습니다 . Weibull 성장은 반응값과 회귀변수 값이 모두 음수가 아닌 경우에만 사용할 수 있습니다 . 시그모이드 곡선의 예로는 학습 곡선과 종양 성장 모델링이 있으며 둘 다 처음에는 증가하다가 차츰 감소합니다 .

지수적으로 증가 및 감소 지수 , 이중 지수 , 기계적 증가 및 셀 성장 모형을 적합시킵니다 . 지수 2P 와 3P 는 비슷하지만 3P 모형에는 점근선이 있습니다 . 이중 지수 모형은 두 개의 개별 증가 및 감소 프로세스가 있다고 가정합니다 . 기계적 증가 및 지수 3P 모형은 항상 증가 (또는 감소) 하지만 증가율 (또는 감소율) 이 느려서 모형에 점근선이 생깁니다 . 지수적으로 증가 및 감소의 예로는 각각 바이러스 확산과 약물 반감기가 있습니다 .

정상점 모형 가우스 정상점 및 로렌츠 정상점 모형을 적합시킵니다 . 이러한 모형은 정상점까지 증가한 후 감소합니다 . 가우스 정상점 모형은 가우스 PDF(확률 밀도 함수) 의 척도화된 결과입니다 . 로렌츠 정상점 모형은 Cauchy 분포 (연속형 확률 분포) 의 척도화된 결과입니다 . 이러한 모형은 일부 화학 농도 시험 및 인공 신경망에 사용할 수 있습니다 .

약동학 모형 1 구획 경구 투여 모형 , 2 구획 정맥 투여 모형 및 이중 지수 4P 모형을 적합시킵니다 . 이 옵션은 체내 약물 농도를 모델링하는 데 사용됩니다 .

Michaelis Menten 적합 기질 농도에 대한 효소 반응 속도를 연관시키는 Michaelis Menten 생화학 운동 모형을 적합시킵니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

모형 비교 보고서

그림 14.8에 표시된 보고서를 생성하려면 "곡선 적합"의 빨간색 삼각형 메뉴에서 **시그모이드 곡선 > 로지스틱 곡선 > 로지스틱 4P 적합** 및 **시그모이드 곡선 > Gompertz 4P 적합**을 선택합니다.

그림 14.8 모형 비교 보고서

곡선 적합

모형 비교

모형	AICc	AICc 가중치	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R^2
로지스틱 4P	-372.3021	1		-348.9054	0.005325	0.0001109	0.0105327	0.9980584
Gompertz 4P	-327.5372	1.903e-10		-304.1405	0.0107173	0.0002233	0.0149425	0.9960922

" 모형 비교 " 보고서에는 여러 모형을 비교하는 데 사용되는 적합 통계량이 표시됩니다 . 통계량은 AICc, AICc 가중치 , BIC, SSE, MSE, RMSE 및 R^2 이며 아래에 정의되어 있습니다 .

AICc 두 개 이상의 모형을 비교하는 데 사용할 수 있는 추정된 통계 모형의 적합도를 나타내는 측도를 제공합니다 . AICc 는 작은 표본에 맞게 AIC 를 조정한 것입니다 . AICc 는 데이터 점수가 모수 수보다 두 개 이상 많은 경우에만 계산할 수 있습니다 . AICc 값이 가장 낮은 모형이 최적 모형이며 이 예에서는 로지스틱 4P 입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

AICc 가중치 합계가 1 이 되는 정규화된 AICc 값을 제공합니다 . AICc 가중치는 적합 모형 중 하나가 true 일 때 특정 모형이 true 모형일 확률로 해석할 수 있습니다 . 따라서 AICc 가중치가 1 에 가장 가까운 모형이 최량 적합 모형입니다 . 이 예에서는 로지스틱 4P 모형이 분명히 더 나은 적합입니다 . AICc 가중치는 비결측 AICc 값만 사용해서 계산됩니다 .

AICcWeight = exp[-0.5(AICc-min(AICc))] / sum(exp[-0.5(AICc-min(AICc))])

여기서 min(AICc) 는 적합 모형 중 가장 작은 AICc 값입니다 . AICc 가중치 열은 내림차순으로 정렬됩니다 .

BIC 여러 모형을 비교할 때 유용한 모형 적합 측도를 제공합니다 (가능도 함수를 기반으로 함). BIC 값이 낮은 모형이 더 나은 적합입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

SSE 각 관측값과 예측값 간의 차이에 대한 제곱합입니다 .

MSE 각 값의 오차 제곱 평균을 제공합니다 .

RMSE 랜덤 오차의 표준편차를 추정하는 MSE 의 제곱근입니다 .

R^2 랜덤 오차가 아니라 모형에 기인할 수 있는 반응의 변동 비율을 추정합니다 . R^2 값이 1 에 가장 가까운 모형이 더 나은 적합입니다 .

모형 비교 플랫폼에서는 잔차와 실제값 그림과 같은 추가 옵션을 제공합니다. 자세한 내용은 "모형 비교"에서 확인하십시오.

모형 적합 보고서

각 적합 모형에 대해 보고서가 생성됩니다. 각 모형 보고서의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

예측 모형 예측 계산식과 모수의 대수 형식을 제공합니다.

적합 요약 "모형 비교" 보고서와 동일한 적합 통계량을 제공합니다.

모수 추정값 모수 추정값, 표준 오차, 카이제곱 통계량, *p* 값 및 신뢰 구간을 제공합니다. 추정값의 상관 및 공분산도 제공됩니다.

팁 : 정렬 기준으로 사용할 열을 클릭하여 "모수 추정값" 보고서 테이블을 정렬할 수 있습니다.

그림 적합 모형을 사용한 데이터 그림을 제공합니다 (그림 14.7). 그림은 플랫폼 시작 창에서 그룹화 변수를 선택한 경우에만 표시됩니다.

모형 적합 옵션

각 모형 보고서에는 다음 옵션 중 일부 또는 전체가 포함된 빨간색 삼각형 메뉴가 있습니다.

병렬 검정 곡선이 X 축을 따라 이동할 때 모형이 비슷한지 여부를 판별하는 데 도움이 됩니다. 특정 상황에서는 그룹 간 추가 비교를 수행하기 전에 병렬 처리를 설정하는 것이 중요합니다. 플랫폼 시작 창에서 그룹 변수를 지정한 경우에만 이 옵션을 사용할 수 있습니다. 이 옵션은 고차 다항식을 제외한 선형 회귀 모형과 시그모이드 모형 (로지스틱 및 Gompertz) 에 사용할 수 있습니다. 자세한 내용은 "병렬 검정"에서 확인하십시오.

곡선 아래 면적 적합 곡선 아래 면적을 제공합니다. 이 옵션은 1 구획, 2 구획, 가우스 정상점 및 로렌츠 정상점 모형에 사용할 수 있습니다. 이중 지수 4P 모형에도 이 옵션을 사용할 수 있지만 모든 모수가 양수일 때만 가능합니다. 통합 범위는 모형의 유형에 따라 다르며 보고서에 지정됩니다.

플랫폼 시작 창에서 그룹화 변수를 지정한 경우 그룹 간 추정값 비교를 위해 평균 분석이 수행됩니다. 그룹에 대한 결과가 결정 한계를 초과하면 결과가 AUC 의 전체 평균과 다르다고 간주됩니다.

반응 정상점까지의 시간 적합 곡선의 정상점에서 회귀변수 *X* 의 추정값을 표시합니다. 추정값의 표준 오차도 표시됩니다. 이 옵션은 셀 성장 4P 및 1 구획 모형에 사용할 수 있습니다.

반응 정상점 적합 곡선의 정상점에서 반응 *Y* 의 추정값을 표시합니다. 추정값의 표준 오차도 표시됩니다. 이 옵션은 셀 성장 4P 및 1 구획 모형에 사용할 수 있습니다.

모수 추정값 비교 그룹화 변수의 전체 수준에서 모수의 동일성 검정을 위한 분석을 제공합니다. 플랫폼 시작 창에서 그룹 변수를 지정한 경우에만 이 옵션을 사용할 수 있습니다. 자세한 내용은 "모수 추정값 비교"에서 확인하십시오.

동등성 검정 그룹화 변수의 전체 수준에서 모형의 동등성 검정을 위한 분석을 제공합니다. 플랫폼 시작 창에서 그룹 변수를 지정한 경우에만 이 옵션을 사용할 수 있습니다. 자세한 내용은 "동등성 검정"에서 확인하십시오.

JMP PRO **곡선 DOE 분석** (시작 창에서 그룹화 변수와 하나 이상의 보조 변수를 지정한 경우에만 사용 가능) 곡선 적합 플랫폼 내에서 "일반화 회귀" 보고서를 시작합니다. 일반화 회귀 모형은 보조 변수를 모형 효과로 사용하여 비선형 모형의 각 모수에 적합됩니다. 기본적으로 2 차 요인 모형이 적합되고 "추정 방법"은 "전진 선택"입니다. 그러면 이러한 모형이 결합되어 반응 프로파일러를 회귀변수 및 보조 변수의 함수로 생성합니다. 그런 다음 CDOE 프로파일러를 사용하여 보조 변수가 반응에 어떤 영향을 주는지 탐색할 수 있습니다.

"곡선 DOE 분석" 보고서에는 다음과 같은 빨간색 삼각형 메뉴 옵션이 있습니다.

모형 모수에 대한 일반화 회귀 각 모형 모수에 대한 "일반화 회귀" 보고서를 표시하거나 숨깁니다. "일반화 회귀" 모형 보고서에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

진단 그림 반응 변수에 대한 실제값 대 예측값 그림과 잔차 그림을 표시하거나 숨깁니다.

CDOE 프로파일러 보조 변수에 따라 반응이 어떻게 변하는지 탐색할 수 있는 CDOE 프로파일러를 표시하거나 숨깁니다. "CDOE 프로파일러" 의 빨간색 삼각형 메뉴 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오.

예측 계산식 저장 반응에 대한 예측 계산식을 데이터 테이블의 새 열에 저장합니다.

모수 테이블 생성 모수 추정값, 표준 오차 및 t 비를 데이터 테이블에 저장합니다. 플랫폼 시작 창에서 그룹 변수를 지정한 경우에만 이 옵션을 사용할 수 있습니다.

실제값 대 예측값 그림 세로 축에 실제 Y 값을 표시하고 가로 축에 예측 Y 값을 표시합니다.

잔차 대 예측값 그림 세로 축에 잔차를 표시하고 가로 축에 예측 Y 값을 표시합니다.

프로파일러 적합 예측 함수의 프로파일러를 표시하거나 숨깁니다. 도함수는 *X* 변수에 대한 예측 함수의 도함수입니다. 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오.

계산식 저장 다양한 계산식 열을 데이터 테이블에 저장하기 위한 옵션이 포함되어 있습니다.

예측 계산식 저장 예측 방정식을 저장합니다.

예측값의 표준 오차 저장 예측값의 표준 오차를 저장합니다.

모수 예측 계산식 저장 예측 방정식을 모수 형식으로 저장합니다. 이렇게 하면 사용자 비선형 플랫폼에서 적합 모형을 사용하려는 경우에 유용합니다.

잔차 계산식 저장 잔차를 저장합니다.

스튜던트화 잔차 계산식 저장 표준화 잔차를 추정된 표준편차로 나누는 스튜던트화 잔차 계산식을 저장합니다.

1 차 도함수 저장 *X* 변수에 대한 예측 함수의 도함수를 저장합니다 .

1 차 도함수의 표준 오차 저장 1 차 도함수의 표준 오차 방정식을 저장합니다 .

역추정 예측 계산식 저장 *Y* 에서 *X* 를 예측하기 위한 방정식을 저장합니다 .

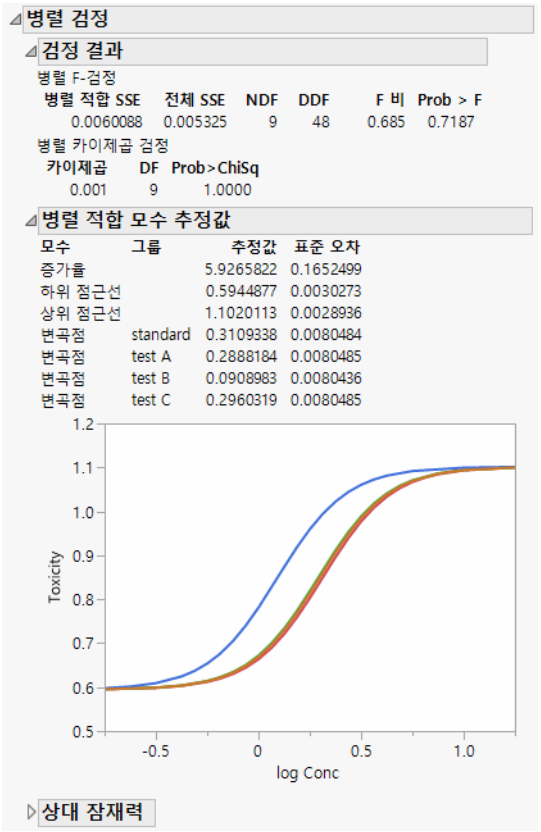
사용자 역추정 예측 특정 *Y* 값에 대한 *X* 값을 예측합니다 . 역추정 예측에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

적합 제거 모형 보고서 , " 모형 비교 " 보고서의 항목 및 그림의 적합선을 제거합니다 .

병렬 검정

"병렬 검정" 옵션은 그룹 간의 적합 모형이 형상은 동일하지만 X 축을 따라 이동하는지 검정하기 위한 분석을 제공합니다(그림 14.9). Bioassay 예에서 약물 제제 B에 대한 곡선은 다른 세 곡선의 왼쪽으로 이동합니다. 그러나 곡선의 형상이 여전히 같은지(병렬) 또는 B 제제가 다른지는 알 수 없습니다. 병렬 검정을 통해 서로 다른 약물 제제의 형상이 비슷하고 가로 축을 따라 이동하는지 여부를 알 수 있습니다. 적합 모형의 빨간색 삼각형 메뉴에서 **병렬 검정**을 선택하여 보고서를 추가합니다.

그림 14.9 병렬 검정



보고서는 다음과 같은 결과를 제공합니다 .

검정 결과 병렬 *F*- 검정 및 카이제곱 검정 결과를 제공합니다 . 이 검정에서는 전체 모형의 오차 제곱합 (전체 SSE) 과 축소 모형의 오차 제곱합 (적합 SSE) 을 비교합니다 . 전체 모형은 각 그룹에 서로 다른 모수를 제공합니다 . 축소 모형은 변곡점을 제외한 모든 모수를 그룹이 공유하도록 강제 적용합니다 . *p* 값이 작으면 그룹 모형이 서로 유의하게 다르다는 것을 나타냅니다 . 따라서 모수를 공유한 축소 모형은 충분하지 않으며 전체 모형을 사용해야 합니다 . 그림 14.9 에서 *p* 값이 0.05 보다 크므로 곡선 사이에 차이가 있다는 결론을 내릴 증거가 충분하지 않습니다 . 이 경우에는 모수를 공유한 축소 모형이 적절합니다 .

적합 통계량은 다음과 같이 계산됩니다 .

F 비 [(적합 SSE - 전체 SSE)/NDF]/[전체 SSE/DDF]

카이제곱 적합 SSE - 전체 SSE

병렬 적합 모수 추정값 축소 모형의 모수 추정값을 제공합니다 (변곡점을 제외한 동일한 모수). 축소 모형의 적합 곡선 그림이 제공됩니다 . 약물 제제 B 의 변곡점이 다른 세 약물 제제의 변곡점보다 훨씬 낮습니다 .

상대 잠재력 그룹화 변수의 각 수준에 대한 잠재력 및 상대 잠재력을 제공합니다 . 잠재력은 10^(EC_{50}) 입니다 . 여기서 EC_{50} 은 기준선과 최대값 사이의 중간 반응이 생성되는 농도입니다 . 로지스틱 2P, 3P 및 4P 의 경우 잠재력은 10^(변곡점 모수) 입니다 . 상대 잠재력은 그룹화 변수의 한 수준에 대한 잠재력을 다른 수준의 잠재력으로 나눈 값입니다 . 예를 들어 그림 14.10 에서 standard 에 대한 test A 의 상대 잠재력은 $\text{Potency}_{\text{test A}}/\text{Potency}_{\text{standard}}$ 입니다 .

그림 14.10 그룹별 상대 잠재력

상대 잠재력

상대 잠재력 대 standard

그룹	잠재력	상대 잠재력	표준 오차
standard	2.0461329	1	0
test A	1.9445466	1.0522416	0.0246535
test B	1.2328161	1.6597227	0.0388878
test C	1.977115	1.0349084	0.0242474

상대 잠재력 대 test A

그룹	잠재력	상대 잠재력	표준 오차
standard	2.0461329	0.950352	0.0222663
test A	1.9445466	1	0
test B	1.2328161	1.5773209	0.036957
test C	1.977115	0.9835273	0.0230435

상대 잠재력 대 test B

그룹	잠재력	상대 잠재력	표준 오차
standard	2.0461329	0.6025103	0.014117
test A	1.9445466	0.6339864	0.0148545
test B	1.2328161	1	0
test C	1.977115	0.6235429	0.0146098

상대 잠재력 대 test C

그룹	잠재력	상대 잠재력	표준 오차
standard	2.0461329	0.9662691	0.0226392
test A	1.9445466	1.0167486	0.0238219
test B	1.2328161	1.6037388	0.037576
test C	1.977115	1	0

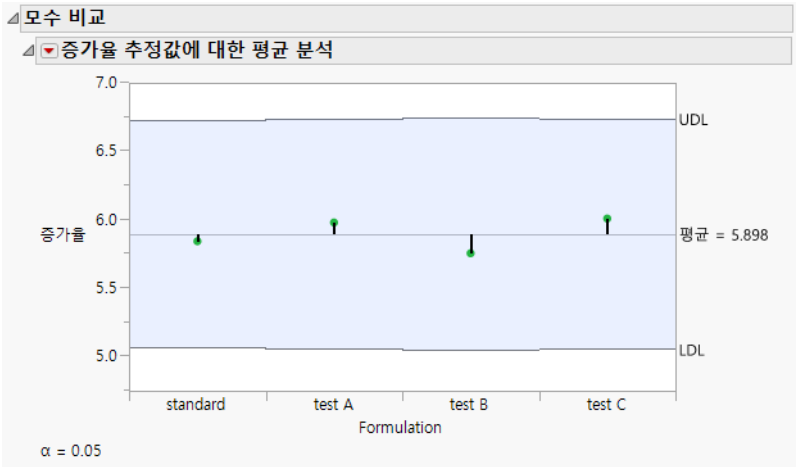
그림 14.10 의 **상대 잠재력 대 standard** 패널에서 약물 제제 A 와 C 의 상대 잠재력은 거의 1 입니다 . 이는 해당 제제의 잠재력이 standard 제제의 잠재력과 유사하다는 것을 나타냅니다 . 약물 제제 B 의 잠재력은 standard 의 잠재력보다 낮습니다 . 이는 농도 함수로 나타내는 약물 제제 B 의 독성이 standard 보다 빠르게 증가한다는 의미입니다 .

병렬 검정에서 곡선이 평행하므로 상대 잠재력을 계산할 수 있습니다. 상대 잠재력을 기반으로 B 제제가 다른 약물 제제보다 더 강력하다는 결론을 내릴 수 있습니다. 이전 조사 결과와 함께 볼 때 약물 제제 B의 독성이 더 강한 것 같습니다.

모수 추정값 비교

"모수 추정값 비교" 보고서에는 그룹화 변수의 전체 수준에서 모수의 동일성 검정 결과가 제공됩니다. 각 모수에 대해 해당 모수가 전체 평균과 같은지 여부를 검정하는 ANOM(평균 분석) 보고서가 있습니다. 모수에 대한 결과가 결정 한계를 초과하면 해당 모수가 전체 평균과 다릅니다. 그림 14.11에서는 증가율 추정값에 대한 ANOM 보고서를 보여 줍니다. 적합 모형의 빨간색 삼각형 메뉴에서 **모수 추정값 비교**를 선택하여 보고서를 추가합니다.

그림 14.11 증가율 추정값에 대한 모수 비교



" 평균 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

유의 수준 설정 검정의 유의 수준을 설정합니다 .

요약 보고서 표시 모수 추정값 , 결정 한계 및 모수의 한계 초과 여부를 포함하는 보고서를 표시하거나 숨깁니다 .

표시 옵션 결정 한계 , 음영 및 중심선을 표시하거나 숨기기 위한 옵션이 포함되어 있습니다 . 점 모양을 변경하는 옵션도 포함되어 있습니다 .

평균 분석 보고서에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오.

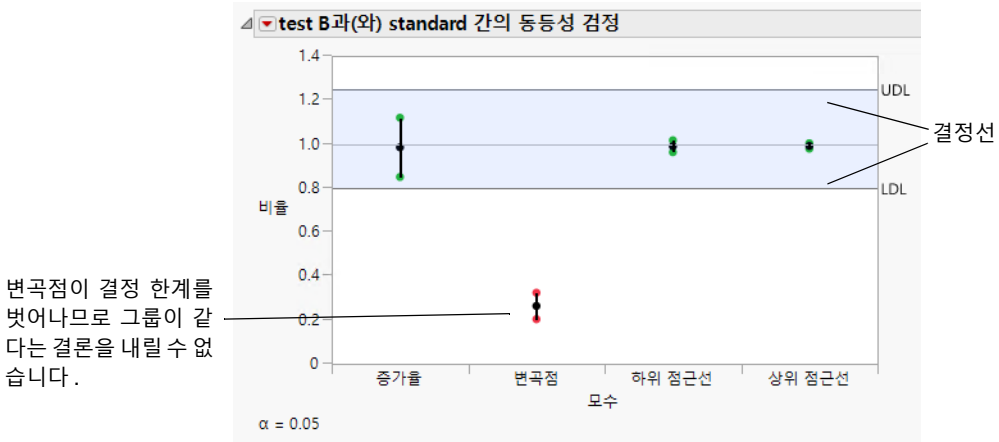
동등성 검정

"동등성 검정" 보고서에는 그룹화 변수의 전체 수준에서 모형의 동등성 검정을 위한 분석이 제공됩니다(그림 14.12). 옵션을 선택한 후에는 다른 모든 수준과 대조하여 검정할 그룹화 변수 수준을 지정합니다. 모든 수준과 선택한 수준을 비교하는 보고서가 있습니다. 적합 모형의 빨간색 삼각형 메뉴에서 **동등성 검정**을 선택하여 보고서를 추가합니다.

모수의 동일성 검정은 모수의 비율을 분석하여 수행됩니다. 기본 결정선은 0.8 및 1.25 비율 값에 배치되어 25% 차이를 나타냅니다.

모든 신뢰 구간이 결정선 내에 있으면 두 그룹이 실제적으로 동일합니다. 단일 구간이 선을 벗어나면(그림 14.12) 그룹이 같다는 결론을 내릴 수 없습니다. 약물 제제 B의 변곡점이 standard 보다 낮으며, 이는 이전 결과와 일치합니다.

그림 14.12 동등성 검정



" 동등성 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

유의 수준 설정 검정의 유의 수준을 설정합니다 . 기본값은 0.05 입니다 .

결정선 설정 비율에 대한 결정 한계를 변경합니다 . 기본값은 0.8 과 1.25 로 설정되어 25% 차이를 나타냅니다 .

요약 보고서 표시 모수 추정값 , 결정 한계 및 모수의 한계 초과 여부를 포함하는 보고서를 표시하거나 숨깁니다 .

표시 옵션 결정 한계 , 음영 및 중심선을 표시하거나 숨기기 위한 옵션이 포함되어 있습니다 . 점 모양을 변경하는 옵션도 포함되어 있습니다. 추가 형식 지정 옵션을 보려면 그래프를 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다 .

곡선 적합 플랫폼에 대한 통계 상세 정보

모형 계산식

표 14.1 에서는 " 곡선 적합 " 빨간색 삼각형 메뉴의 모형에 대한 계산식을 제공합니다 .

표 14.1 곡선 적합 모형 계산식

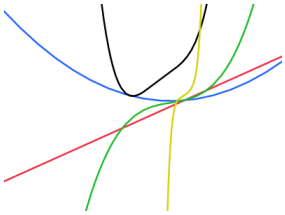
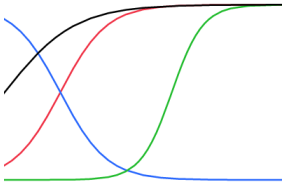
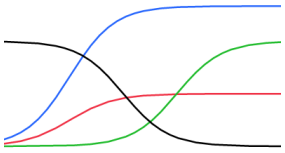
모형	계산식
다항 	$\beta_0 + \sum_{i=1}^{k} \beta_i x^i$ 여기서 k는 다항식의 차수입니다. 이러한 모형은 모형 적합 플랫폼과 X로 Y 적합 플랫폼을 사용하여 적합시킬 수도 있습니다.
로지스틱 2P 	$\frac{1}{1 + \text{Exp}(-a(x-b))}$ a = 증가율 b = 변곡점 모든 반응 값이 0에서 1 사이인 경우에만 사용할 수 있습니다.
로지스틱 3P 	$\frac{c}{1 + \text{Exp}(-a(x-b))}$ a = 증가율 b = 변곡점 c = 점근선

표 14.1 곡선 적합 모형 계산식 (계속)

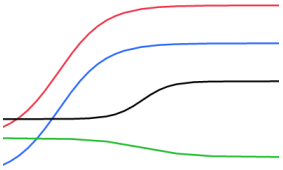
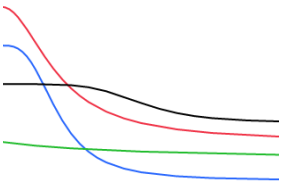
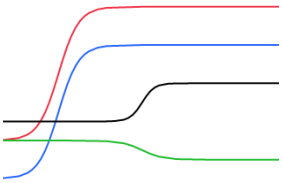
모형	계산식
로지스틱 4P 	$c + \frac{d-c}{1+\text{Exp}(-a(x-b))}$ a = 증가율 b = 변곡점 c = 하위 점근선 d = 상위 점근선
로지스틱 4P Rodbard 	$c + \frac{d-c}{1+(x/b)^a}$ a = 증가율 b = 변곡점 c = 하위 점근선 d = 상위 점근선 회귀변수 값이 양수인 경우에만 사용할 수 있습니다.
로지스틱 4P Hill 	$c + \frac{d-c}{1+10^{(-a(x-b))}}$ a = 증가율 b = 변곡점 c = 하위 점근선 d = 상위 점근선

표 14.1 곡선 적합 모형 계산식 (계속)

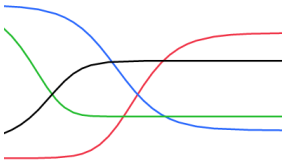
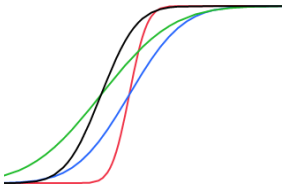
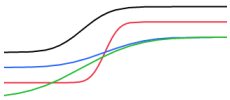
모형	계산식
로지스틱 5P 	$c + \frac{d-c}{(1+\text{Exp}(-a(x-b)))^f}$ a = 증가율 b = 변곡점 c = 점근선 1 d = 점근선 2 f = 검정력
프로빗 2P 	$\Phi\left(\frac{x-b}{a}\right)$ a = 증가율 b = 변곡점 Φ = 정규 분포 CDF 모든 반응 값이 0에서 1 사이인 경우에만 사용할 수 있습니다.
프로빗 4P 	$c + (d-c) \cdot \Phi\left(\frac{x-b}{a}\right)$ a = 증가율 b = 변곡점 c = 점근선 1 d = 점근선 2 Φ = 정규 분포 CDF

표 14.1 곡선 적합 모형 계산식 (계속)

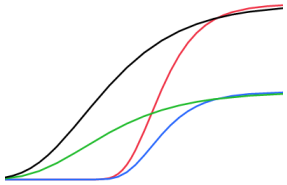
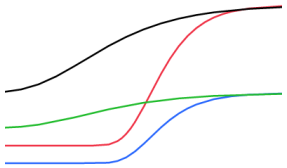
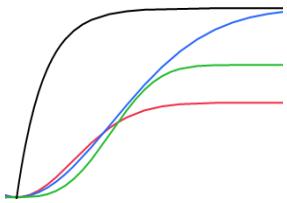
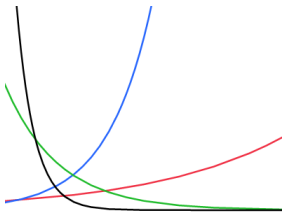
모형	계산식
Gompertz 3P 	$a\text{Exp}(-\text{Exp}(-b(x-c)))$ a = 점근선 b = 증가율 c = 변곡점
Gompertz 4P 	$a+(b-a)\text{Exp}(-\text{Exp}(-c(x-d)))$ a = 하위 점근선 b = 상위 점근선 c = 증가율 d = 변곡점
Weibull 성장 	$a(1-\text{Exp}(-(x/c)^{b}))$ a = 상위 점근선 b = 증가율 c = 변곡점 반응 값과 회귀변수 값이 모두 음수가 아닌 경우에만 사용할 수 있습니다.
지수 2P 	$a\text{Exp}(bx)$ a = 척도 b = 증가율

표 14.1 곡선 적합 모형 계산식 (계속)

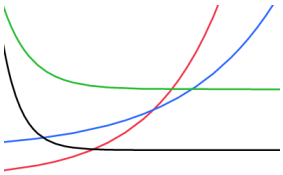
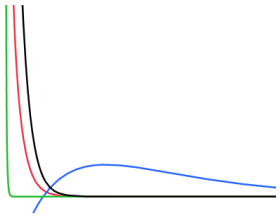
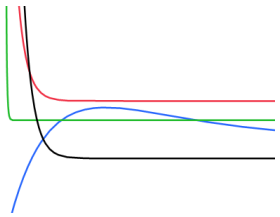
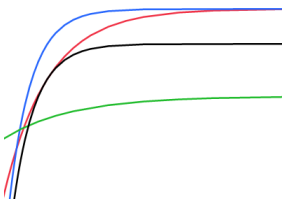
모형	계산식
지수 3P 	$a + b\text{Exp}(cx)$ a = 점근선 b = 척도 c = 증가율
이중 지수 4P 	$a\text{Exp}(-bx) + c\text{Exp}(-dx)$ a = 척도 1 b = 감소율 1 c = 척도 2 d = 감소율 2 반응 값이 양수인 경우에만 사용할 수 있습니다.
이중 지수 5P 	$a + b\text{Exp}(-cx) + d\text{Exp}(-fx)$ a = 점근선 b = 척도 1 c = 감소율 1 d = 척도 2 f = 감소율 2
기계적 증가 	$a(1 - b\text{Exp}(-cx))$ a = 점근선 b = 척도 c = 증가율

표 14.1 곡선 적합 모형 계산식 (계속)

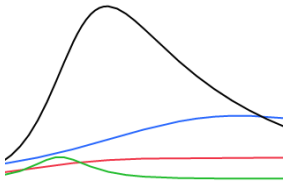
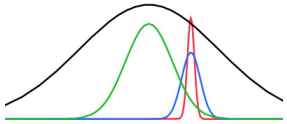
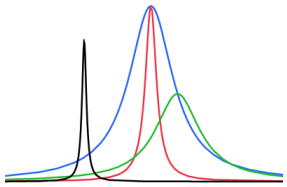
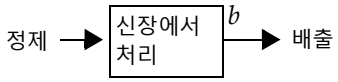
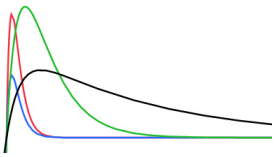
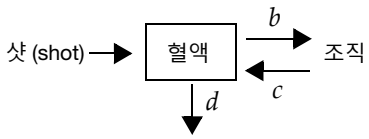
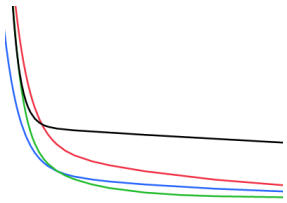
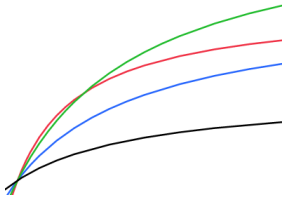
모형	계산식
셀 성장 4P 	$\frac{ab}{(a-b)\text{Exp}(-cx)+b\text{Exp}(dx)}$ a = 사멸 비율 d가 0인 경우 정상점 값 b = 시간 0에서 반응 c = 셀 분화 비율 d = 셀 사멸 비율 반응 값이 양수이고 회귀변수 값이 음수가 아닌 경우에만 사용할 수 있습니다.
가우스 정상점 	$a\text{Exp}\left(-\frac{1}{2}\left(\frac{x-b}{c}\right)^2\right)$ a = 정상점 값 b = 임계점 c = 증가율
로렌츠 정상점 	$\frac{ab^2}{(x-c)^2+b^2}$ a = 정상점 값 b = 증가율 c = 임계점
1구획 경구 투여  	$\frac{abc}{c-b}(\text{Exp}(-bx)-\text{Exp}(-cx))$ a = 곡선 아래 면적 b = 제거율 c = 흡수율 반응 값과 회귀변수 값이 모두 양수인 경우에만 사용할 수 있습니다.

표 14.1 곡선 적합 모형 계산식 (계속)

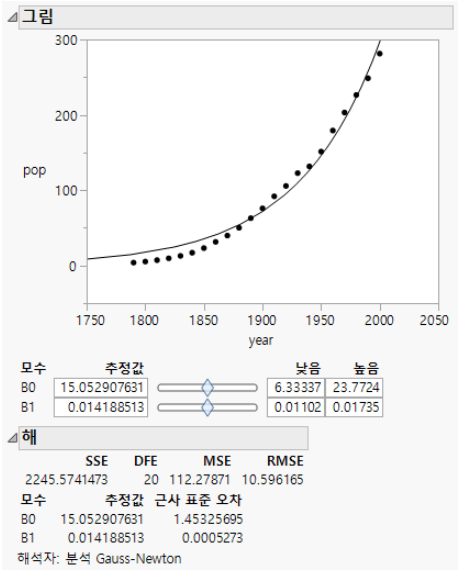
모형	계산식
2구획 정맥 투여  	$\frac{a}{\alpha-\beta}((\alpha-b)\mathrm{Exp}(-\alpha x)-(\beta-b)\mathrm{Exp}(-\beta x))$ $\alpha = \frac{1}{2}(b+c+d+\sqrt{(b+c+d)^2-4bd})$ $\beta = \frac{1}{2}(b+c+d-\sqrt{(b+c+d)^2-4bd})$ a = 초기 농도 b = 유입 속도 c = 유출 속도 d = 제거율 반응 값과 회귀변수 값이 모두 양수인 경우에만 사용할 수 있습니다.
Michaelis Menten 	$\frac{ax}{b+x}$ a = 최대 반응 속도 b = 역 선호도 반응 값과 회귀변수 값이 모두 양수인 경우에만 사용할 수 있습니다.

15 장

비선형 회귀
데이터에 사용자 비선형 모형 적합

비선형 플랫폼은 모수에 대해 **비선형**인 모형에 사용됩니다. 이 장에서는 모형 계산식 및 추정할 모수가 포함된 사용자 비선형 모형에 대해 중점적으로 다룹니다. 기본 최소 제곱 손실 함수 또는 사용자 손실 함수를 사용하여 모형을 적합시킵니다. 이 플랫폼은 전체 관측값에서 손실 함수의 합을 최소화합니다.

그림 15.1 사용자 비선형 적합의 예



비선형 플랫폼에서는 다항식, 로지스틱, Gompertz, 지수, 정상점 및 약동학 모형과 같은 사전 정의된 모형도 제공합니다. 자세한 내용은 "곡선 적합"에서 확인하십시오.

참고 : 일부 모형은 모수에 대해 **선형** (예 : 2 차 또는 기타 다항식) 이거나 , 선형이 되도록 변환 (예 : x 의 로그 변환을 사용하는 경우) 할 수 있습니다 . 이러한 상황에서는 모형 적합 또는 X 로 Y 적합 플랫폼이 더 적절합니다 . 이 플랫폼에 대한 자세한 내용은 **선형 모형 적합**의 및 **기본 분석**의 에서 확인하십시오 .

목차

사용자 모형 적합의 예 229
비선형 플랫폼 시작 232
비선형 적합 보고서 233
비선형 플랫폼 옵션 237
모형 라이브러리를 사용하여 계산식 생성 240
    비선형 모형 라이브러리 사용자 정의 242
추가 예 243
    최대 가능도의 예: 로지스틱 회귀 243
    이항 오차가 있는 프로빗 모형의 예: 수치 미분 245
    Poisson 손실 함수의 예 246
    모수 한계 설정의 예 248
비선형 플랫폼에 대한 통계 상세 정보 250
    프로파일 가능도 신뢰 한계 250
    사용자 손실 함수의 작동 방식 251
    도함수 관련 참고 사항 253
    효과적인 비선형 모델링에 대한 참고 사항 254

사용자 모형 적합의 예

이 예에서는 모형에 대한 계산식 열을 생성한 후 비선형 플랫폼으로 모형을 적합시켜 사용자 모형을 적합시킵니다. 이 방법을 사용하면 기본 제공 모형을 적합시키는 것보다 몇 단계가 더 필요하지만 비선형 모형을 적합시킬 수 있습니다. 또한 사용자 손실 함수를 제공하고 적합 프로세스에 대한 몇 가지 다른 옵션을 지정할 수도 있습니다.

팁 : 사용자 모형이 여러 데이터 테이블에 자주 사용되는 모형인 경우 모형 라이브러리에 모형을 추가할 수 있습니다. 자세한 내용은 " 비선형 모형 라이브러리 사용자 정의 " 에서 확인하십시오.

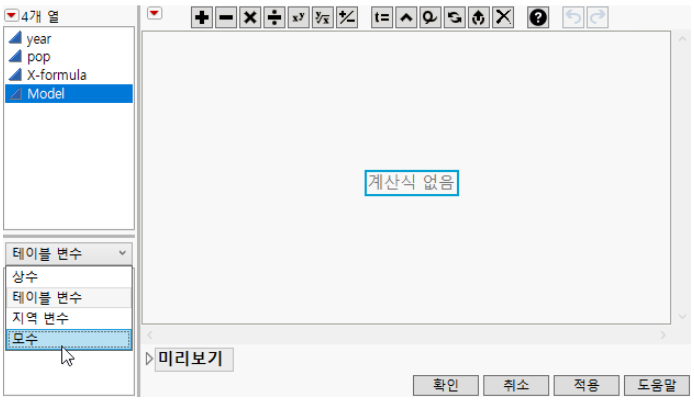
데이터는 US Population.jmp 데이터 테이블에 있습니다. 반응 변수는 미국 인구 (백만 단위) 이고 예측 변수는 연도입니다.

계산식 열 생성

사용자 모형을 적합시키려면 먼저 초기 모수 추정값으로 모형 열을 생성해야 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nonlinear Examples/US Population.jmp 를 엽니다.
2. 모형이라는 새 열을 생성합니다.
3. 모형 열을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 계산식**을 선택합니다.
   계산식 편집기가 나타납니다.
4. 열 목록 아래의 목록에서 **모수**를 선택합니다.

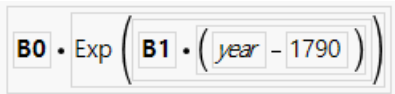
그림 15.2 모수 선택



5. **새 모수**를 선택합니다.
6. " 이름 " 에 "B0" 을 입력합니다.
7. " 값 " 에 "3.9" 를 입력합니다. 이 값이 모수의 초기 추정값입니다.
8. **확인**을 클릭합니다.

9. **새 모수**를 선택합니다 .
10. " 이름 " 에 "B1" 을 입력하고 " 값 " 에 "0.022" 를 입력합니다 .
11. **확인**을 클릭합니다 .
12. 계산식 편집기 함수 , year 열 및 모수를 사용하여 모형 계산식을 입력합니다 .

그림 15.3 완료된 모형 계산식



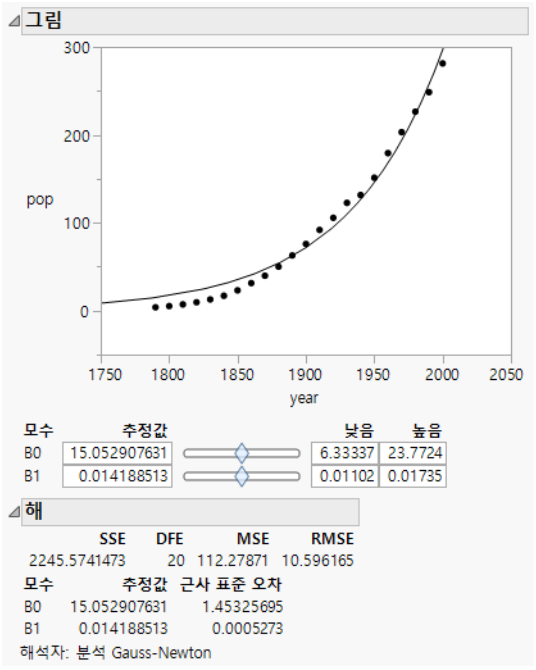
팁 : Exp 명령을 찾으려면 " 초월함수 " 옆의 회색 삼각형을 클릭합니다 .

13. **확인**을 클릭합니다 .

비선형 모형 적합

1. **분석 > 전문 모델링 > 비선형**을 선택합니다 .
2. 모형을 **X, 예측변수 계산식** 역할에 할당합니다 .
3. pop 를 **Y, 반응** 역할에 할당합니다 .
4. **확인**을 클릭합니다 .
5. " 제어판 " 에서 **시작**을 클릭하여 모형을 적합시킵니다 .

그림 15.4 그림 및 해 보고서



해 보고서에 최종 모수 추정값이 다른 적합 통계량과 함께 표시됩니다 . 적합 모형은 그림에 표시됩니다 .

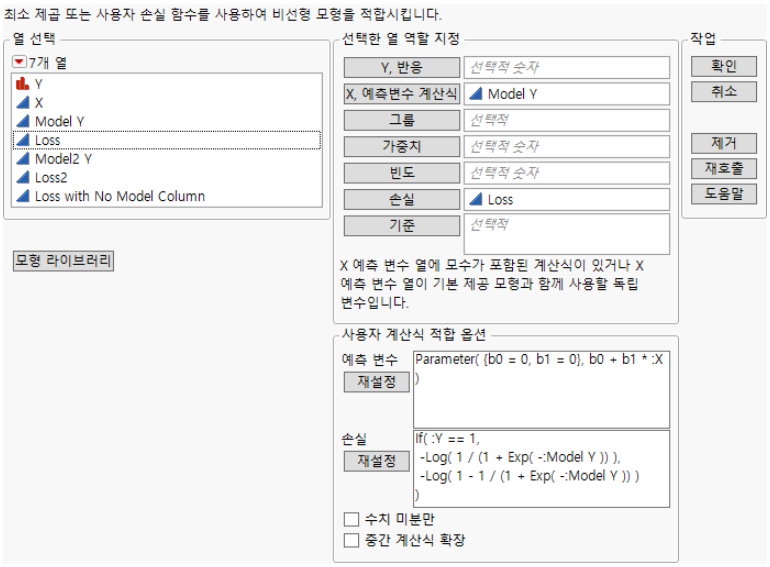
그룹화 변수가 있는 모형의 모수

계산식 편집기에서 모수를 추가할 때 **범주로 확장, 열 선택** 체크박스에 주목하십시오. 이 옵션은 한 번에 여러 모수(예: 범주형 변수의 각 수준에 대해 하나씩)를 추가하는 데 사용됩니다. 이 옵션을 선택하면 열을 선택할 수 있는 대화상자가 나타납니다. 선택 후 "모수" 목록에 이름이 D_열 이름인 새 모수가 나타납니다. 여기서 D는 모수에 지정한 이름입니다. 이 모수를 계산식에 사용하면 그룹화 변수의 각 수준에 대한 개별 모수를 포함하는 Match 표현식이 삽입됩니다.

비선형 플랫폼 시작

비선형 플랫폼을 시작하려면 **분석 > 전문 모델링 > 비선형**을 선택합니다 .

그림 15.5 비선형 플랫폼 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

비선형 플랫폼 시작 창에는 다음과 같은 기능이 있습니다 .

Y, 반응 *Y* 변수를 선택합니다 .

X, 예측변수 계산식 *X* 변수를 선택하거나, 모수를 사용한 모형 계산식이 포함된 열을 선택합니다.

참고 : 모수를 사용한 모형 계산식이 포함되지 않은 열을 선택하고 사용자 계산식도 지정하지 않으면 곡선 적합 플랫폼이 시작됩니다 .

참고 : 할당이 없는 모수 목록 항목이 포함된 계산식 열을 선택하면 JMP 에서 자동으로 해당 모수를 0 으로 설정하고 로그에 메시지를 표시합니다 .

그룹 그룹화 변수를 지정합니다 . 적합 모형에는 그룹화 변수의 각 수준에 대한 개별 모수가 있습니다 . 이를 통해 그룹화 변수의 전체 수준에서 적합 모형과 추정 모수를 비교할 수 있습니다 .

참고 : 계산식에 각 그룹 수준에 대한 개별 모수가 없으면 JMP 에서 그룹 변수가 무시됩니다 .

가중치 관측값의 가중치를 포함하는 변수를 지정합니다 .

빈도 관측값의 빈도를 나타내는 변수를 지정합니다 .

손실 손실 함수를 제공하는 계산식 열을 지정합니다 .

기준 변수의 모든 수준에 대해 개별 분석을 수행할 변수를 지정합니다 .

모형 라이브러리 계산식 열 생성을 위해 초기값을 선택하는 데 도움이 되는 " 모형 라이브러리 " 도구를 실행합니다 . 자세한 내용은 " 모형 라이브러리를 사용하여 계산식 생성 " 에서 확인하십시오 .

사용자 계산식 적합 옵션 모수를 사용한 모형 계산식이 포함되지 않은 열을 "X, 예측변수 계산식 " 역할에 지정하는 경우 이 옵션을 사용하여 사용자 계산식을 생성할 수 있습니다 . " 예측 변수 " 필드에서 모수 , 모수 값 , 선택한 *X* 변수를 사용하는 사용자 계산식을 지정합니다 . 그런 다음 **재설정** 버튼을 클릭하여 플랫폼을 시작합니다 .

수치 미분만 수치 미분만 사용합니다 . 이 옵션은 모형이 정리되지 않아 해석 도함수를 사용할 수 없는 경우에 유용합니다 . 또한 어려운 상황에서도 수렴에 도달하는 데 도움이 될 수 있습니다 . 이 옵션은 "X, 예측변수 계산식 " 역할에 계산식 열이 제공되는 경우에만 사용됩니다 .

중간 계산식 확장 모형에 대한 성분 열이 자체 계산식이 있는 열인 경우 해당 열이 다른 열을 참조하면 JMP 에서 내부 계산식을 대체하도록 지정합니다 . 성분 열이 확장되지 않도록 하려면 **기타** 열 특성을 사용하여 이름이 " 계산식 확장 " 이고 값이 0 인 특성을 할당합니다 . 이 옵션은 "X, 예측변수 계산식 " 역할에 계산식 열이 제공되는 경우에만 사용됩니다 .

비선형 적합 보고서

초기 " 비선형 적합 " 보고서에는 다음 항목이 포함됩니다 (그림 15.6 참조).

제어판 적합 프로세스를 제어하기 위한 옵션을 제공합니다 .

시작 적합 프로세스를 시작합니다 .

중지 적합 프로세스를 중지합니다 .

단계 적합 프로세스를 한 번에 하나씩 순서대로 반복합니다 .

재설정 계산식에 대해 편집 가능한 값을 재설정하고 , 반복 값을 재설정한 후 이러한 새 값으로 SSE 를 계산합니다 .

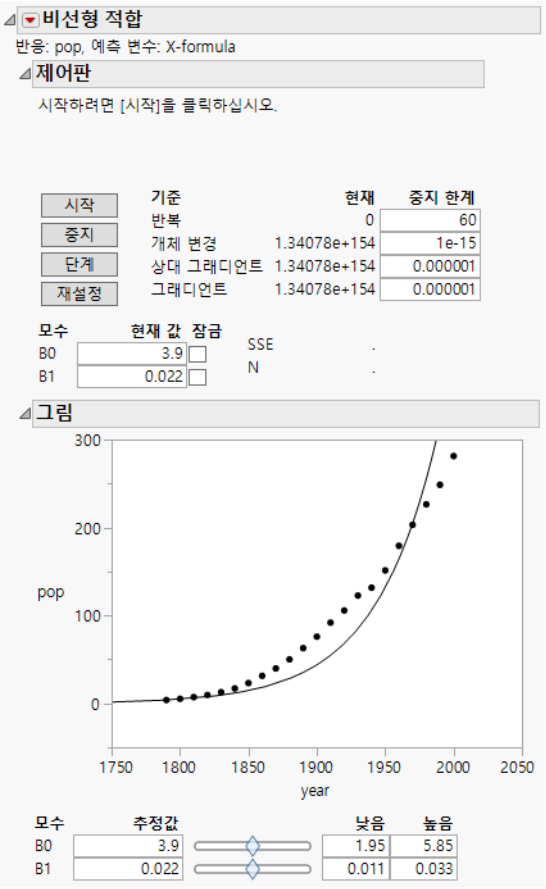
기준 적합 프로세스의 반복 측도를 표시합니다 .

현재 각 기준의 현재 값을 표시합니다 .

중지 한계 " 기준 " 아래에 나열된 측도에 대한 한계를 설정합니다 .

그림 *X* 변수가 하나뿐인 모형의 *X* 및 *Y* 변수 그림을 표시합니다 . 현재 값에 기반한 모형이 그림에 표시됩니다 . 모수의 현재 값을 변경하려면 그림 아래의 슬라이더 또는 편집 상자를 사용합니다 .

그림 15.6 초기 비선형 적합 보고서



시작을 클릭하여 모형을 적합시키면 다음과 같은 추가 항목이 보고서에 포함됩니다 (그림 15.7 참조).

추정값 저장 현재 모수 값을 계산식 열의 모수에 저장합니다 .

신뢰 한계 모든 모수에 대한 신뢰 구간을 계산합니다 . 이 구간은 프로파일 가능도 신뢰 구간이며 " 해 " 보고서에 표시됩니다 . 신뢰 한계 계산에는 각 모수의 각 한계에 대한 새 반복 집합이 포함되며 반복을 통해 한계를 찾지 못하는 경우가 종종 있습니다 . **알파 편집** 및 **수렴 기준** 옵션은 신뢰 구간 계산에 사용됩니다 . **CL 에 대한 목표 SSE** 에 대한 자세한 내용은 " 프로파일 가능도 신뢰 한계 " 에서 확인하십시오 .

해 모수 추정값 및 기타 통계량을 표시합니다 .

SSE 잔차 오차 제곱합을 표시합니다 . SSE 는 최소화하는 것이 목표입니다 . 사용자 손실 함수가 지정된 경우 이 값은 손실 함수의 합입니다 .

DFE 오차의 자유도 , 즉 사용된 관측값 수에서 적합된 모수의 수를 뺀 값을 표시합니다 .

MSE 평균 제곱 오차를 표시합니다. 이 값은 잔차 오차 분산의 추정값, 즉 SSE 를 DFE 로 나눈 값입니다.

RMSE 잔차 오차의 표준편차, 즉 MSE 의 제곱근을 추정합니다.

R^2 (모형에 절편이 있고 잔차 합이 0 에 가까운 경우에만 사용 가능) R^2 값을 표시합니다.

모수 적합 계산식에서 모수에 지정한 이름을 나열합니다.

추정값 생성된 모수 추정값을 나열합니다. 비선형 회귀의 경우 모든 것이 작동하는 것처럼 보여도 이 추정값에 문제가 있을 수 있습니다.

근사 표준 오차 선형 회귀와 유사하게 계산되는 근사 표준 오차를 나열합니다. 이 값은 RMSE 의 곱과 도함수 교차곱 역행렬의 주대각 제곱근에 의해 생성됩니다.

참고 : 관측값 수가 적합 모형의 모수 수보다 작거나 같으면 근사 표준 오차가 보고되지 않습니다.

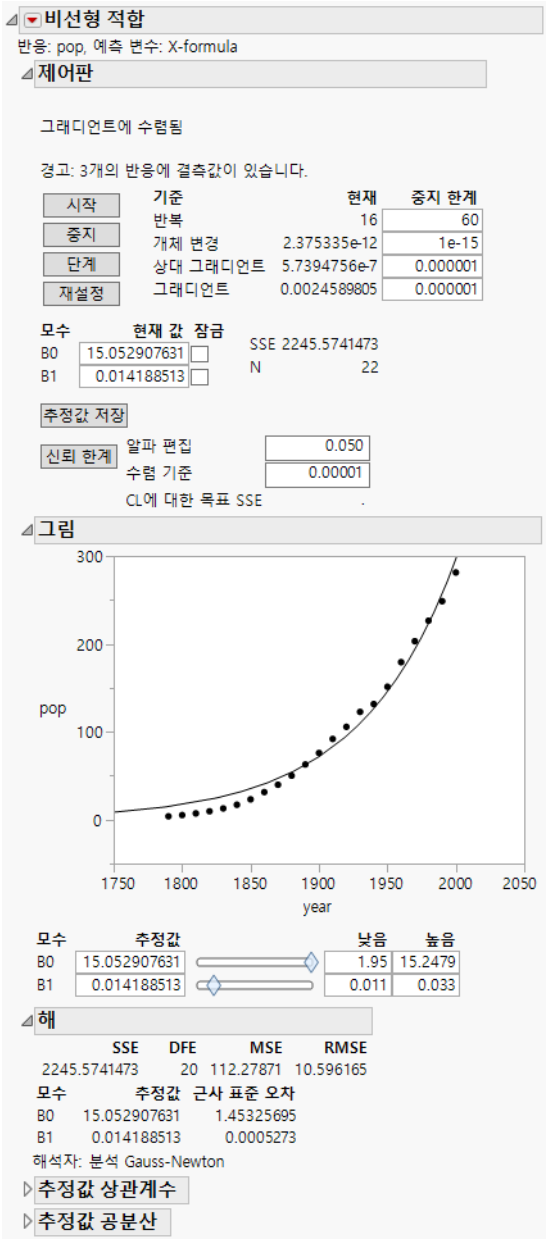
CL 하한 및 CL 상한 모수에 대한 신뢰 한계를 표시합니다. " 제어판 " 에서 **신뢰 한계**를 클릭하지 않으면 이 값은 결측입니다. 신뢰 구간에 대한 자세한 내용은 " 프로파일 가능도 신뢰 한계 " 에서 확인하십시오.

제외된 데이터 제외된 행에 대한 적합 통계량을 보여 주는 보고서를 표시합니다. 이 옵션은 모형 적합에 사용되지 않은 관측값에 대해 모형을 검증하는 경우에 유용합니다. 이 기능을 " 해 저장 " 옵션과 함께 사용하여 제외 항목을 변경하고 다른 제외 항목을 반영하는 새 보고서를 얻을 수 있습니다.

추정값 상관계수 모수 추정값 간의 상관을 표시합니다.

추정값 공분산 모수 추정값 간의 공분산을 표시합니다.

그림 15.7 적합 모형 보고서



비선형 플랫폼 옵션

" 비선형 적합 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

모수 경계 모수의 경계를 설정합니다 . 이 옵션을 선택하면 " 제어판 " 에 편집 가능한 상자가 나타납니다 . 필드를 비워 두면 경계가 없는 모수임을 나타냅니다 .

그림 *X* 변수가 하나뿐인 모형의 *X* 및 *Y* 변수 그림을 표시하거나 숨깁니다 . 그림에 표시된 모형은 모수의 현재 값을 기반으로 합니다 . 모수의 현재 값을 변경하려면 그림 아래의 슬라이더 또는 편집 상자를 사용합니다 . 시작할 때 그룹 변수를 지정하면 각 그룹에 대한 곡선이 표시됩니다 .

반복 옵션 적합 알고리즘에 대한 옵션을 지정합니다 .

반복 로그 적합 프로세스의 각 단계를 새 창에 기록합니다 .

수치 미분만 모형이 정리되지 않아 해석 도함수를 사용할 수 없는 경우에 유용합니다 . 또한 어려운 상황에서도 수렴에 도달하는 데 도움이 될 수 있습니다 .

중간 계산식 확장 수식에 대한 성분 열이 자체 계산식이 있는 열인 경우 해당 열이 다른 열을 참조하면 JMP 에서 내부 계산식을 대체하도록 지정합니다 . 성분 열이 확장되지 않도록 하려면 **기타** 열 특성을 사용하여 이름이 " 계산식 확장 " 이고 값이 0 인 특성을 할당합니다 .

Newton Gauss-Newton(정규 최소 제곱의 경우) 또는 Newton-Raphson(손실 함수가 있는 모형의 경우) 이 최적화 방법인지 여부를 지정합니다 .

QuasiNewton SR1 QuasiNewton SR1 을 최적화 방법으로 지정합니다 .

QuasiNewton BFGS QuasiNewton BFGS 를 최적화 방법으로 지정합니다 .

현재 추정값 채택 추정값이 수렴되지 않은 경우에도 JMP 에서 현재 추정값을 사용하여 해 보고서를 생성하도록 지정합니다 .

도함수 표시 비선형 계산식의 도함수를 JMP 로그에 표시합니다 . 도함수에 대한 기술 정보는 " 도함수 관련 참고 사항 " 에서 확인하십시오 .

언스레드 주 계산 스레드에서 반복을 실행합니다 . 대부분의 경우 JMP 는 별도의 계산 스레드에서 계산을 수행합니다 . 이렇게 하면 비선형 계산 중에 다른 작업을 수행하는 동안 JMP 의 응답성이 향상됩니다 . 그러나 주 스레드에서 실행해야 하는 일부 고립 상황 (예 : 디스플레이 루틴을 호출하는 파생 효과가 있는 모형) 이 있으므로 이 옵션을 설정해야 합니다 .

프로파일러 반응 표면을 보기 위한 다양한 프로파일러를 제공합니다 .

프로파일러 예측 프로파일러를 표시합니다 . 프로파일러를 사용하면 각 *x* 변수에서 차례대로 표면의 수직 단면을 볼 수 있으며 요인의 최적 값을 찾을 수 있습니다 .

등고선 프로파일러 등고선 프로파일러를 표시합니다 . 등고선 프로파일러를 사용하면 2 차원 등고선과 3 차원 그물 그림을 볼 수 있습니다 .

표면 프로파일러 3 차원 표면 그림을 생성합니다 . 이 옵션은 X 변수가 두 개 이상 있는 모형에만 사용할 수 있습니다 .

모수 프로파일러 예측 프로파일러를 표시하고 모수의 함수로 나타낸 SSE 또는 손실을 프로파일링합니다 .

모수 등고선 프로파일러 등고선 프로파일러를 표시하고 모수의 함수로 나타낸 SSE 또는 손실 등고선을 그립니다 .

모수 표면 프로파일러 3 차원 표면 그림을 생성하고 모수의 함수로 나타낸 SSE 또는 손실을 프로파일링합니다 . 이 옵션은 모수가 두 개 이상 있는 모형에만 사용할 수 있습니다 .

SSE 격자 해 추정값 주변의 값 격자를 생성하고 각 값의 오차 제곱합을 계산합니다 . 해 추정값의 SSE 가 최소값이어야 합니다 . 이 옵션을 선택하면 다음과 같은 기능이 포함된 **출력에 대한 격자 지정** 보고서가 표시됩니다 .

모수 모형의 모수를 나열합니다 .

최소값 격자 계산에 사용된 최소 모수 값을 표시합니다 . 기본적으로 **최소값**은 해 추정값에서 **근사 표준 오차**의 2.5 배를 뺀 값입니다 .

최대값 격자 계산에 사용된 최대 모수 값을 표시합니다 . 기본적으로 **최대값**은 해 추정값에 **근사 표준 오차**의 2.5 배를 더한 값입니다 .

점의 수 각 모수에 대해 생성할 점의 수를 제공합니다 . 새 격자 테이블의 총 점 수를 계산하려면 모든 " 점의 수 " 값을 곱합니다 . 처음에 " 점의 수 " 값은 처음 두 모수에 대해 11 이고 나머지 모수에 대해 3 입니다 . 새 값을 지정하는 경우 홀수 값을 사용하여 격자 테이블에 해 추정값이 포함되도록 해야 합니다 . 모수에 대해 " 점의 수 " 를 0 으로 설정하면 격자 테이블에 해 추정값만 기록됩니다 .

시작을 클릭하면 JMP 에서 새 테이블에 점 격자를 생성합니다 . 강조 표시된 행은 테이블에 해가 있는 경우 해 추정값 행을 나타냅니다 .

원래 모수로 되돌리기 플랫폼을 원래 모수 값 (계산식 열 모수에 지정된 값) 으로 재설정합니다 .

해 저장 현재 모수 추정값과 요약 통계량을 포함하는 " 저장된 모형 " 보고서를 생성합니다 . 여러 모형의 결과를 저장하고 비교할 수 있습니다 . 이 기능은 서로 다른 모수 제한에 기반한 모형 또는 다른 옵션을 사용하여 적합된 모형을 비교하려는 경우에 유용합니다 . 특정 모형에 대한 라디오 버튼을 클릭하면 " 그림 " 에 해당 모형이 표시되고 " 제어판 " 에 모수 추정값이 표시됩니다 .

사용자 추정값 모수 함수의 추정값을 제공합니다 . 사용자는 모수만 포함된 표현식을 제공합니다 . 그러면 JMP 에서 현재 모수 추정값을 사용하여 표현식을 계산하고 1 차 Taylor 급수 근사를 사용하여 표현식의 표준 오차도 계산합니다 .

사용자 역추정 예측 주어진 *Y* 값에 대한 *X* 값을 추정합니다 . 또한 추정된 *X* 의 표준 오차를 계산합니다 . JMP 에서 모형을 반전할 수 있어야 합니다 . 표준 오차는 반전 표현식을 사용하는 1 차 Taylor 급수 근사를 기반으로 합니다 . 신뢰 구간은 *t*- 분위수를 표준 오차와 함께 사용하며 Wald 구간입니다 .

예측 표현식 표시 보고서 맨 위에 예측 모형 또는 손실 함수를 표시합니다 .

예측 신뢰 한계 저장 모형 예측에 대한 점근적 신뢰 한계를 저장합니다. 이 값은 주어진 X 값에서 평균 Y 에 대한 신뢰 구간입니다.

개별 신뢰 한계 저장 개별 예측에 대한 점근적 신뢰 한계를 저장합니다. 이 값은 주어진 X 값에서 개별 Y 값에 대한 신뢰 구간입니다.

계산식 저장 모형 결과를 데이터 테이블 열에 저장하기 위한 옵션을 제공합니다.

예측 계산식 저장 현재 모수 추정값을 사용하여 예측 계산식을 저장합니다.

예측값의 표준 오차 저장 모형 예측에 대한 표준 오차를 저장합니다. 이 값은 주어진 X에 대해 평균 Y를 예측하기 위한 표준 오차입니다. 계산식은 `Sqrt(VecQuadratic(matrix1,vector1))` 형식입니다. 여기서 `matrix1` 은 모수 추정값과 관련된 공분산 행렬이고 `vector1` 은 각 모수에 대한 모형의 편도함수 합성입니다.

개별값의 표준 오차 저장 개별 예측에 대한 표준 오차를 저장합니다. 이 값은 주어진 X 값에 대해 개별 Y 값을 예측하기 위한 표준 오차입니다. 계산식은 `Sqrt(VecQuadratic(matrix1, vector1)+mse)` 형식입니다. 여기서 `matrix1` 은 모수 추정값과 관련된 공분산 행렬, `vector1` 은 각 모수에 대한 모형의 편도함수 합성, `mse` 는 오차 분산 추정값입니다.

잔차 계산식 저장 잔차를 계산하기 위한 계산식을 저장합니다.

예측 신뢰 한계 계산식 저장 모형 예측에 대한 신뢰 구간을 계산하는 계산식을 저장합니다. 이 값은 주어진 X 에 대한 평균 Y 의 신뢰 구간입니다.

개별 신뢰 한계 계산식 저장 개별 예측에 대한 신뢰 구간을 계산하는 계산식을 저장합니다. 이 값은 주어진 X 에 대한 개별 Y 의 신뢰 구간입니다.

역추정 예측 계산식 저장 모형의 역추정, 역추정 예측의 표준 오차 및 개별 역추정 예측의 표준 오차에 대한 계산식을 저장합니다.

지정된 풀이 계산식 저장 간단한 사례의 경우 " 역추정 예측 계산식 저장 " 과 동일합니다. 그러나 이 명령을 사용하면 계산식이 여러 변수의 함수가 될 수 있으며 표현식을 대체할 수 있습니다. 이 기능은 계산식에서 한 번만 나오는 가역 연산자 및 함수를 쉽게 풀이하는 데만 사용할 수 있습니다.

이 명령을 선택하면 해석할 변수를 선택할 수 있는 대화상자가 나타납니다. 결과 테이블의 열 이름을 편집할 수도 있습니다. 또한 이 대화상자에서 이름을 값으로 대체할 수도 있습니다. 이러한 경우 해당 값에 대한 계산식이 풀이됩니다.

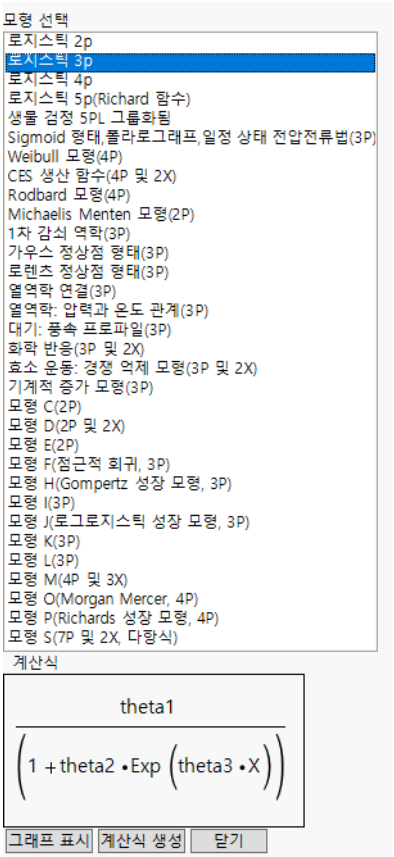
참고 : 표준 오차, 신뢰 구간 및 가설 검정은 최소 제곱 추정이 수행되거나, 최대 가능도 추정이 적절한 음의 로그 가능도와 함께 사용되는 경우에만 올바릅니다.

추정값을 테이블에 저장 모수 추정값을 새 데이터 테이블에 저장합니다.

모형 라이브러리를 사용하여 계산식 생성

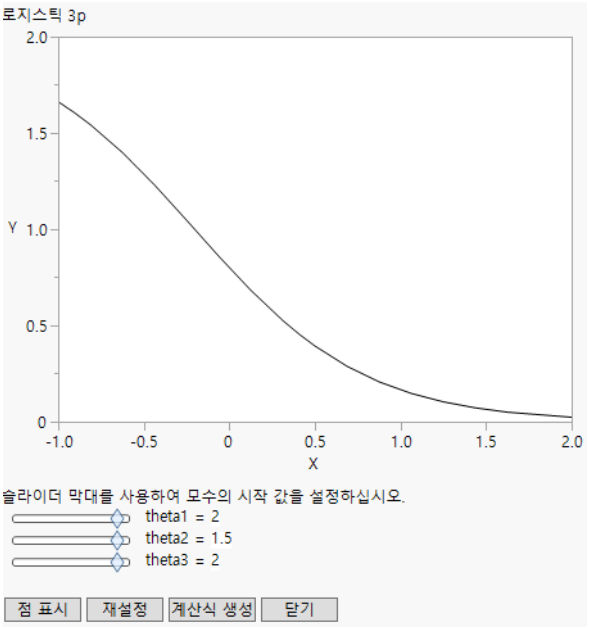
모형 라이브러리를 사용하면 모수와 초기값으로 계산식 열을 생성하는 데 도움이 됩니다. 비선형 시작 창에서 **모형 라이브러리** 버튼을 클릭하여 라이브러리를 엽니다 . 목록에서 모형을 선택하면 **계산식** 상자에 해당 모형의 계산식이 표시됩니다 .

그림 15.8 비선형 모형 라이브러리 대화상자



그래프 표시를 클릭하면 1 모수 모형의 2D 이론적 곡선과 2 모수 모형의 3D 표면 그림이 표시됩니다. 설명 (X) 변수가 세 개 이상 있는 모형에는 그래프를 사용할 수 없습니다. 그래프 창에서 슬라이더를 사용하거나, 값을 직접 클릭한 후 입력하여 모수의 기본 초기값을 변경합니다.

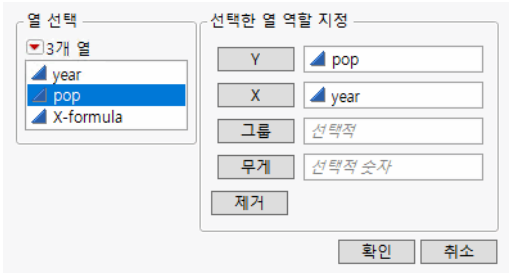
그림 15.9 모형 라이브러리의 그래프 예



재설정 버튼은 모수의 초기값을 기본값으로 다시 설정합니다.

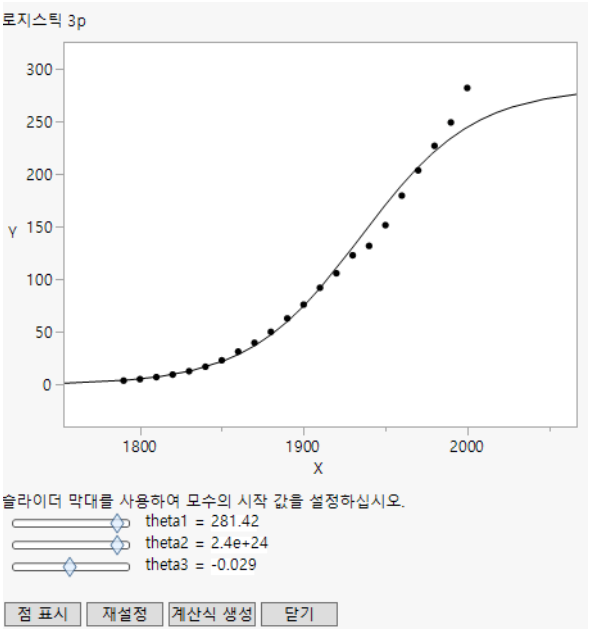
실제 데이터 점을 그림에 중첩하려면 **점 표시**를 클릭합니다. *X* 및 *Y* 역할과 선택적 그룹 역할에 열을 할당하라는 대화상자가 열립니다(그림 15.10 참조). 그룹 역할을 사용하면 범주형 변수의 모든 수준에 모형을 적합시킬 수 있습니다. 여기에 그룹 역할을 지정하는 경우 플랫폼 시작 창의 그룹 열도 지정합니다.

그림 15.10 역할 선택



대부분의 모형에서 시작 값은 상수입니다. 점을 표시하면 모수 값을 조정하여 모형이 모수의 여러 값에 따라 얼마나 잘 적합되는지 확인할 수 있습니다. 미국 인구 예의 경우 그림 15.11에 점이 표시되어 있습니다.

그림 15.11 점 표시



점 표시를 사용한 후 **계산식 생성**을 클릭하면 모형 라이브러리에서 선택한 모형을 따라 명명된 새 데이터 테이블 열이 생성됩니다. 이 열에는 최신 모수 시작 값의 함수로 나타낸 계산식이 있습니다.

참고 : 그래프 표시 또는 **점 표시** 버튼을 사용하기 전에 **계산식 생성**을 클릭하면 *X* 및 *Y* 역할과 선택적 그룹 역할을 제공해야 합니다 (그림 15.10). 그런 다음 원할 경우 모수 시작 값을 조정할 수 있도록 그림으로 돌아갑니다 . 이때 **계산식 생성**을 다시 클릭하여 새 열을 생성합니다 .

데이터 테이블에 계산식이 생성되면 비선형 시작 대화상자에서 새 열을 **X, 예측변수 계산식**으로 할당하여 분석을 계속합니다.

비선형 모형 라이브러리 사용자 정의

모형 라이브러리는 JMP(Windows) 가 포함된 폴더 또는 Application Package(macOS) 의 Resources/Builtins 폴더에 기본 제공된 NonlinLib.jsl 스크립트에 의해 생성됩니다. 이 스크립트를 수정하여 비선형 라이브러리 스크립트를 사용자 정의할 수 있습니다.

참고 : 파일을 직접 편집하지 못할 수도 있습니다 . 이 경우 스크립트를 열고 원하는 대로 편집한 후 컴퓨터의 다른 위치에 저장합니다 . 이때 동일한 파일 이름을 사용하십시오 . 그런 다음 Resources/Builtins 폴더의 기존 NonlinLib.jsl 스크립트를 삭제하고 새 NonlinLib.jsl 파일을 Resources/Builtins 폴더로 이동합니다 .

모형을 추가하려면 Listofmodellist# 목록에 세 줄을 추가해야 합니다 . 이 세 줄은 사실 다음과 같은 세 부분으로 구성된 목록입니다 .

- 모형 (따옴표로 묶은 문자열)
- 모형 계산식 (표현식)
- 모형 척도

예를 들어 다음과 같은 형식의 "Simple Exponential Growth" 라는 모형을 추가하려는 경우를 가정해 보겠습니다 .

$$y = b_1 e^{kx}$$

다음 줄을 NonlinLib.jsl 스크립트에 추가합니다 .

```
{//Simple Exponential Growth
      "Simple Exponential Growth",
      Expr(Parameter({b1=2, k=0.5}, b1*exp(k * :X))),
      lowx = -1; highx = 2; lowy = 0;  highy = 2},
```

다음 사항에 유의하십시오 .

- 첫 번째 줄은 여는 대괄호 (목록 시작) 와 선택적 주석입니다 . 두 번째 줄은 모형 라이브러리 창에 표시되는 문자열입니다 .
- lowx, highx, logy 및 highy 값은 이론적 그래프의 초기 창을 지정합니다 .
- 위의 예에서는 쉼표가 마지막 문자로 사용됩니다 . Listofmodellist# 목록의 마지막 항목에는 쉼표를 생략할 수 있습니다 .
- 모형에 세 개 이상의 모수가 사용되는 경우 마지막 줄 (그래프 한계 포함) 을 따옴표로 묶은 "String Not Available" 문자열로 바꿉니다 .

모형을 삭제하려면 Listofmodellist# 목록에서 해당하는 세 줄 목록을 삭제합니다.

추가 예

- " 최대 가능도의 예 : 로지스틱 회귀 "
- " 이항 오차가 있는 프로빗 모형의 예 : 수치 미분 "
- "Poisson 손실 함수의 예 "
- " 모수 한계 설정의 예 "

최대 가능도의 예 : 로지스틱 회귀

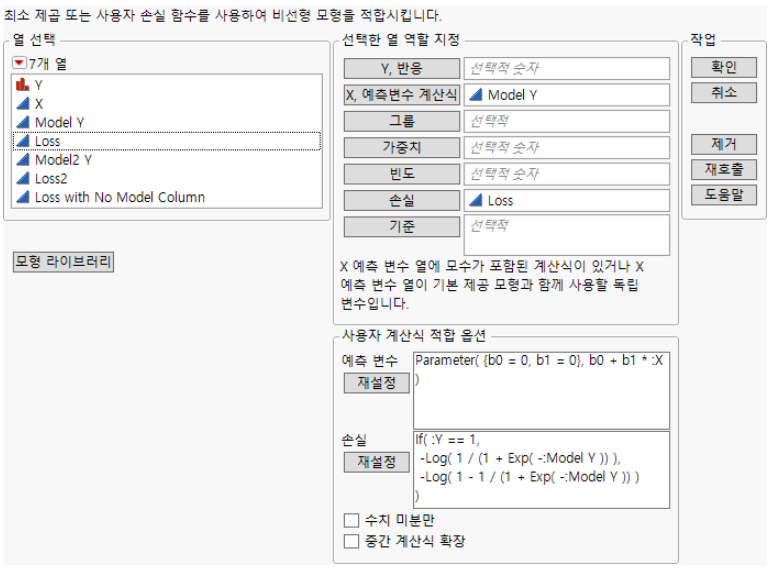
이 예에서는 비선형 플랫폼을 사용하여 손실 함수를 최소화하는 방법을 보여 줍니다. 손실 함수는 음의 로그 가능도 함수이므로 최대 가능도 추정값을 생성합니다.

Nonlinear Examples 샘플 데이터 폴더의 Logistic w Loss.jmp 데이터 테이블에는 손실 함수를 사용하여 로지스틱 회귀를 적합시키는 예제가 있습니다. Y 열에는 사건의 경우 1, 비사건(non-event)의 경우 0 값이 포함되어 있습니다. Model Y 열에는 선형 모형이 있고 Loss 열에는 손실 함수가 있습니다. 이 예에서 손실 함수는 각 관측값에 대한 음의 로그 가능도 또는 관측된 반응을 얻을 확률의 음의 로그입니다.

아래 단계에 따라 모형을 실행합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nonlinear Examples/Logistic w Loss.jmp 를 엽니다.
2. **분석 > 전문 모델링 > 비선형**을 선택합니다.
3. Model Y 를 **X, 예측변수 계산식** 역할에 할당합니다.
4. Loss 를 **손실** 역할에 할당합니다.

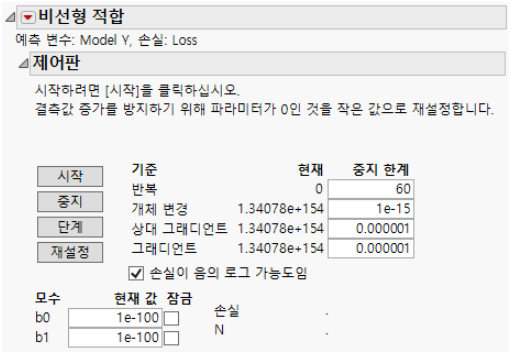
그림 15.12 비선형 시작 창



5. **확인**을 클릭합니다.

   비선형 적합 제어판이 나타납니다.

그림 15.13 비선형 적합 제어판



6. **시작**을 클릭합니다.

   모수 추정값이 " 해 " 보고서에 표시됩니다.

그림 15.14 해 보고서

"해" 보고서의 "손실" 값은 모수 추정값에서 계산된 음의 로그 가능도입니다.

이항 오차가 있는 프로빗 모형의 예 : 수치 미분

Ingots2.jmp 샘플 데이터 테이블에는 가열 및 균열 처리 시간을 다르게 한 후 준비 상태를 테스트한 주괴 수가 포함되어 있습니다. 반응 변수 NReady 는 테스트한 주괴 수 (Ntotal) 와 가열 및 균열 시간에 따라 이항입니다. 이항 오차가 있는 프로빗 모형의 모수에 대한 최대 가능도 추정값은 다음을 사용하여 계산됩니다.

- 수치 미분
- 손실 함수로 나타낸 음의 로그 가능도
- Newton-Raphson 방법

준비된 평균 주괴 수는 테스트한 수와 주어진 가열 및 균열 처리 시간에서 주괴 사용 준비가 완료될 확률의 곱입니다. 프로빗 모형을 사용하여 P 열에는 다음과 같은 모형 계산식이 포함됩니다.

```
Normal Distribution(b0+b1*Heat+b2*Soak)
```

`Normal Distribution` 함수에 대한 인수는 처리의 선형 모형입니다.

이항 오차를 지정하기 위해 손실 함수 Loss 에는 다음 계산식이 포함됩니다 .

```
-(Nready*Log(p) + (Ntotal - Nready)*Log(1 - p))
```

다음 단계에 따라 모형을 적합시킵니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ingots2.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 비선형**을 선택합니다 .
3. P 를 **X, 예측변수 계산식** 역할에 할당합니다 .
4. Loss 를 **손실** 역할에 할당합니다 .
5. **수치 미분만** 옵션을 선택합니다 .
6. **확인**을 클릭합니다 .
7. **시작**을 클릭합니다 .

이 플랫폼에서는 그림 15.15 에 표시된 모수 추정값을 얻기 위해 수치 SR1 방법을 사용했습니다 .

그림 15.15 Ingots2 데이터의 해

해

손실	DFE	평균 손실	제곱근 평균 손실
47.479945327	16	2.9674966	1.7226423

모수	추정값	근사 표준 오차
b0	-2.8934153	0.51256572
b1	0.0399554554	0.01202329
b2	0.0362537934	0.15017139

해석자: 수치 SR1

Poisson 손실 함수의 예

Poisson 분포는 개수 데이터를 모델링하는 데 주로 사용됩니다 .

$$P(Y = n) = \frac{e^{-\mu}\mu^{n}}{n!}, n = 0, 1, 2, \ldots$$

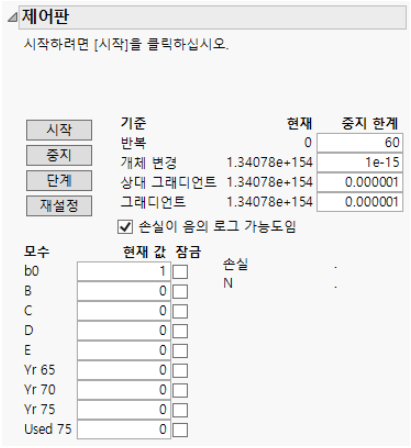
여기서 μ는 단일 모수이거나, 여러 모수가 있는 선형 모형일 수 있습니다. 많은 교재와 논문에서 모형을 변환하고 반복 재가중 최소 제곱을 사용하여 적합시킬 수 있는 방법을 보여 줍니다 (Nelder and Wedderburn 1972 연구 자료). 그러나 JMP 에서는 모형을 직접 적합시키는 것이 더 간단합니다. 예를 들어 McCullagh and Nelder 연구 자료(1989)에서는 화물 운송 선박에 대해 파도로 인해 보고된 피해 사고의 수를 분석하는 방법을 보여 줍니다.

데이터는 Ship Damage.jmp 샘플 데이터 테이블에 있습니다 . 모형 계산식은 model 열에 있고 손실 함수 (음의 로그 가능도) 는 Poisson 열에 있습니다 . 모형을 적합시키려면 아래 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ship Damage.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 비선형**을 선택합니다 .

3. model 을 **X, 예측변수 계산식** 역할에 할당합니다 .
4. Poisson 을 **손실** 역할에 할당합니다 .
5. **확인**을 클릭합니다 .
6. b0 의 **현재 값** (초기값) 을 1 로 설정하고 다른 모수는 0 으로 설정합니다 .

그림 15.16 새 모수 입력



7. **시작**을 클릭합니다 .
8. **신뢰 한계** 버튼을 클릭합니다 .

"해" 보고서가 나타납니다. 결과에는 모수 추정값, 신뢰 구간 및 기타 요약 통계량이 포함됩니다.

그림 15.17 Poisson 손실 예의 해 테이블

해

손실	DFE	평균 손실	제곱근 평균 손실
68.281234087	25	2.7312494	1.6526492

모수	추정값	근사 표준 오차	CL 하한	CL 상한
b0	-6.405914771	0.21744445	-6.8430512	-5.9896833
B	-0.543353982	0.17758996	-0.881379	-0.183552
C	-0.687418388	0.32904722	-1.3764541	-0.0745299
D	-0.075979835	0.2905789	-0.6715296	0.47523936
E	0.325581683	0.23587934	-0.143451	0.78520431
Yr 65	0.6971496668	0.1496412	0.40752821	0.99512758
Yr 70	0.8184263427	0.16977314	0.48728046	1.15369087
Yr 75	0.4534435316	0.23317069	-0.0123098	0.90388895
Used 75	0.3844880718	0.11827198	0.15341509	0.61742312

해석자: 분석 NR

참고 : 표준 오차 , 신뢰 구간 및 가설 검정은 최소 제곱 추정이 수행되거나 , 최대 가능도 추정이 적절한 음의 로그 가능도와 함께 사용되는 경우에만 올바릅니다 .

모수 한계 설정의 예

곡선 적합 분석법을 사용하면 모형을 적합시킨 후 비선형 플랫폼의 전체 분석법에서 예측 방정식을 사용할 수 있습니다. 이 방법은 더 많은 단계와 사용자 입력이 필요하지만 비선형 모형을 적합시킬 수 있습니다.

"곡선 적합 플랫폼을 사용한 예"의 단계를 완료하여 모형을 적합시킵니다. 이 예에서는 곡선 적합에서 예측 계산식을 저장한 후 비선형에서 모수 한계를 설정하는 방법을 보여 줍니다.

1. " 로지스틱 4P" 의 빨간색 삼각형을 클릭하고 **계산식 저장 > 모수 예측 계산식 저장**을 선택합니다.

   Toxicity 예측 변수라는 새 열이 데이터 테이블에 나타납니다.
2. **분석 > 전문 모델링 > 비선형**을 선택합니다.
3. Toxicity 를 **Y, 반응** 역할에 할당합니다.
4. Toxicity 예측 변수를 **X, 예측변수 계산식** 역할에 할당합니다.
5. Formulation 을 **그룹** 역할에 할당합니다.
6. **확인**을 클릭합니다.

   " 비선형 적합 " 창이 나타납니다. " 제어판 " 에 모수 값과 잠금 옵션이 표시됩니다. 각 모수 앞에 나열된 문자는 곡선 적합 기능에서 예측 모형의 변수에 해당합니다.

그림 15.18 비선형 적합 제어판

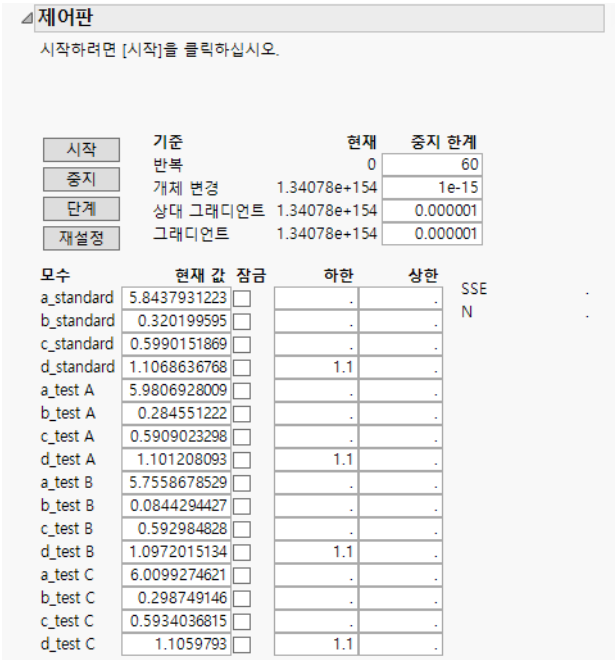
팁 : 사전 정보에서 값을 알고 있는 경우 모수를 잠글 수 있습니다 .

7. " 비선형 적합 " 의 빨간색 삼각형을 클릭하고 **모수 경계**를 선택합니다 .

   모수 상한 및 하한을 설정하기 위한 옵션이 모수 옆에 나타납니다 .

8. 그림 15.19 에 표시된 대로 모수의 하한을 설정합니다 . 사전 경험을 통해 약물의 최대 독성이 1.1 이상이라는 것을 알고 있습니다 .

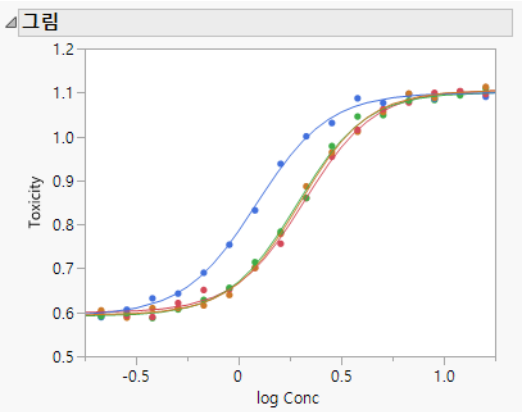
그림 15.19 모수 경계 설정



9. **시작**을 클릭합니다 .

   해 보고서에 최종 모수 추정값이 다른 적합 통계량과 함께 표시됩니다 . 적합 모형은 그림에 표시됩니다 .

그림 15.20 비선형 적합 그림 및 모수 추정값



모수	추정값		낮음	높음
a_standard	5.843793154		2.9219	8.76569
b_standard	0.3201995954		0.1601	0.4803
c_standard	0.5990151873		0.29951	0.89852
d_standard	1.1068636765		0.55343	1.6603
a_test A	5.9806928191		2.99035	8.97104
b_test A	0.2845512223		0.14228	0.42683
c_test A	0.5909023301		0.29545	0.88635
d_test A	1.1012080929		0.5506	1.65181
a_test B	5.6527018259		2.87793	8.6338
b_test B	0.0860027315		0.04221	0.12664
c_test B	0.5917244683		0.29649	0.88948
d_test B	1.1		0.5486	1.6458
a_test C	6.0099274821		3.00496	9.01489
b_test C	0.2987491449		0.14937	0.44812
c_test C	0.5934036814		0.2967	0.89011
d_test C	1.1059792995		0.55299	1.65897

그림 아래의 옵션을 사용하여 모수 한계 및 추정값을 조정할 수 있습니다.

비선형 플랫폼에 대한 통계 상세 정보

- " 프로파일 가능도 신뢰 한계 "
- " 사용자 손실 함수의 작동 방식 "
- " 도함수 관련 참고 사항 "
- " 효과적인 비선형 모델링에 대한 참고 사항 "

프로파일 가능도 신뢰 한계

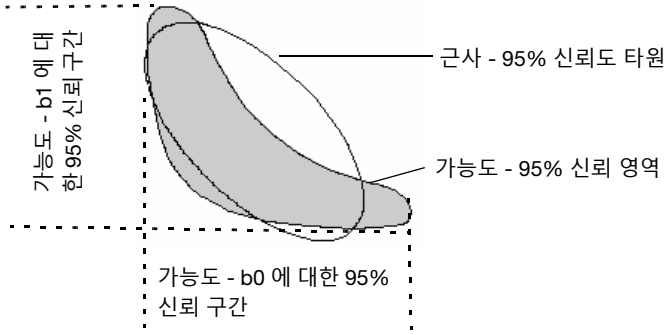
모수의 신뢰 상한 및 하한은 각 모수를 다른 모수에 대해 최소화한 후 해당 모수의 값 검색을 기반으로 합니다. 검색을 통해 해의 최소 SSE보다 특정 값만큼 큰 SSE를 생성하는 값을 찾습니다. 이 차이의 목표는 *F* 분포를 기반으로 합니다. 이러한 구간을 **가능도 신뢰 구간** 또는 **프로파일 가능도 신뢰 구간**이라고도 합니다(Bates and Watts 1988, Ratkowsky 1990).

프로파일 신뢰 한계는 모두 **목표 SSE**로 시작합니다. 이 값은 *F*-검정이 주어진 유의 수준에서 해 SSE와 유의하게 다르다고 간주하는 오차 제곱합(또는 손실 함수의 합)입니다. 손실 함수가 음의 로그 가능도로 지정되면 *F* 분위수 대신 카이제곱 분위수가 사용됩니다. 각 모수의 신뢰 상한은 SSE가 목표 SSE에 도달할 때까지 모수 값이 증가합니다. 모수 값이 상승하므로 다른 모든 모수는 프로파일링된 모수의 변경에 따라 최소 제곱 추정값으로 조정됩니다. 개념상 이것은 내포된 반복의 복합 집합입니다. 내부적으로는 Johnston과 DeLong이 개발한 1회 반복 작업을 통해 이를 수행하는 방법이 있습니다. 자세한 내용은 SAS/ETS User's Guide(SAS Institute Inc. 2020)에서 확인하십시오.

그림 15.21 에서는 음영 영역 내부에 최소 제곱 (또는 최소 손실) 해가 있는 목표 SSE 또는 음의 가능도의 등고선을 보여 줍니다 .

- 점근적 표준 오차는 타원체가 있는 영역에 가깝고 극단 (수평 및 수직 접선) 의 모수 값을 사용하는 신뢰 구간을 생성합니다 .
- 프로파일 신뢰 한계는 근사 타원체 대신 실제 영역의 극단에서 모수 값을 찾습니다 .

그림 15.21 모수에 대한 신뢰 한계 다이어그램



가능도 신뢰 구간은 근사 표준 오차에서 계산된 신뢰 구간보다 신뢰도가 높습니다 . 특정 한계를 찾을 수 없으면 다음 한계에 대한 계산이 시작됩니다 . 수렴이 어려운 경우 다음을 시도해 보십시오 .

- 더 큰 알파 값을 사용하면 구간이 짧아지고 동작이 개선될 수 있습니다 .
- 신뢰 한계 기준을 완화합니다 .

사용자 손실 함수의 작동 방식

비선형 기능은 기본 잔차 제곱합 이외의 함수를 최소화하거나 최대화할 수 있습니다. 이 섹션에서는 이를 수행하는 방식을 수리적으로 보여 줍니다.

$f(\beta)$ 를 모형으로 가정해 보겠습니다 . 그런 다음 비선형 플랫폼에서 다음과 같이 정의된 손실 함수의 합을 최소화하려고 시도합니다 .

$$L = \sum_{i=1}^{n} \rho(f(\beta))$$

각 행에 대한 손실 함수 $\rho(\bullet)$ 은 데이터 테이블에 있는 다른 변수의 함수일 수 있습니다 . 여기에는 0이 아닌 1차 및 2차 도함수가 있어야 합니다. 기본 $\rho(\bullet)$ 함수인 잔차 제곱은 다음과 같습니다.

$$\rho(f(\beta)) = (y - f(\beta))^2$$

사용자 손실 함수가 있는 모형을 지정하려면 데이터 테이블에 변수를 생성하고 손실 함수를 구성합니다. 비선형 플랫폼을 시작한 후 손실 함수가 포함된 열을 손실 변수로 선택합니다.

비선형 최소화 계산식은 모형에 대해 $\rho(\bullet)$ 의 처음 두 도함수를 가져오고 , 다음과 같이 그래디언트 및 근사 헤시안 행렬을 구성하여 작동합니다 .

$$L = \sum_{i=1}^{n} \rho(f(\beta))$$

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \rho(f(\beta))}{\partial f} \frac{\partial f}{\partial \beta_j}$$

$$\frac{\partial^2 L}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n} \left[\frac{\partial^2 \rho(f(\beta))}{(\partial f)^2} \frac{\partial f}{\partial \beta_j} \frac{\partial f}{\partial \beta_k} + \frac{\partial \rho(f(\beta))}{\partial f} \frac{\partial^2 f}{\partial \beta_k \partial \beta_j} \right]$$

$f(\bullet)$ 가 모수에 대해 선형인 경우 마지막 방정식의 두 번째 항은 0 입니다 . 그렇지 않은 경우에도 첫 번째 항에 비해 합계가 작을 수 있으며 다음을 사용합니다 .

$$\frac{\partial^2 L}{\partial \beta_j \partial \beta_k} \cong \sum_{i=1}^{n} \frac{\partial^2 \rho(f(\beta))}{(\partial f)^2} \frac{\partial f}{\partial \beta_j} \frac{\partial f}{\partial \beta_k}$$

잔차 합이 작으므로 ρ 이 잔차 제곱이면 두 번째 항이 작을 수 있습니다. 절편 항이 있는 경우 항은 0입니다. 최소 제곱의 경우 이 항은 Gauss-Newton 과 Newton-Raphson 을 구별하는 항입니다.

참고 : 표준 오차 , 신뢰 구간 및 가설 검정은 최소 제곱 추정이 수행되거나 , 최대 가능도 추정이 적절한 음의 로그 가능도와 함께 사용되는 경우에만 올바릅니다 .

도함수 관련 참고 사항

비선형 플랫폼에서는 가장 일반적인 연산을 사용하는 계산식에 대해 기호 도함수를 사용합니다. 이 섹션에서는 도함수 표현식의 결과 유형을 보여 줍니다.

Negative Exponential.jmp 비선형 샘플 데이터 예제를 열면 Nonlinear 열의 실제 계산식이 다음과 같습니다 .

```
Parameter({b0=0.5, b1=0.5}, b0*(1-Exp(-b1*X)))
```

계산식 편집기를 사용하는 경우 계산식의 Parameter 블록이 숨겨집니다. 계산식은 이러한 방식으로 열에 저장되고 비선형 시작 대화상자에 나타납니다. 두 모수 **b0**과 **b1**은 초기값이 지정되고 적합 계산식에 사용됩니다.

비선형 플랫폼에서는 계산식의 별도 복사본을 만들고 이를 편집하여 표현식에서 모수를 추출합니다. 그런 다음 모수에 대한 참조를 해당 모수가 추정된 위치에 매핑합니다. 비선형에서는 모수에 대해 예측 계산식의 해석 도함수를 사용합니다. **도함수 표시** 명령을 사용하면 다음과 같은 결과 계산식이 로그에 나열됩니다.

예측 모형

```
b0 * First(T#1=1-(T#2=Exp(-b1*X)), T#3=-(-1*T#2*X))
```

모수에 대한 **모형**의 도함수

```
{T#1, T#3*b0}
```

도함수 기능은 다음과 같이 작동합니다 .

- 하위 표현식을 반복해서 계산하지 않도록 하기 위해 예측 모형에는 도함수 계산에 필요한 하위 표현식 값을 저장하는 할당이 스레드됩니다 . T#1, T#2 등과 같은 이름에 할당이 수행됩니다 .
- 예측 모형은 평가된 추가 하위 표현식이 필요한 경우 First 함수를 사용합니다. 이 함수는 첫 번째 인수 표현식의 값을 반환하고 다른 인수도 평가합니다 . 이 경우 도함수에 대한 추가 할당이 필요합니다 .
- 도함수 테이블은 그 자체가 표현식 목록이며 , 각 모수에 대해 하나의 표현식을 적합시킵니다 . 예를 들어 **b0** 에 대한 모형의 도함수는 T#1 이며 , 예측 모형의 스레드는 1-(Exp(-b1*X)) 입니다 . **b1** 에 대한 도함수는 T#3*b0 이며 , 위의 할당을 대체할 경우 -(-1*Exp(-b1*X)*X)*b0 입니다 . 최적화가 많이 수행되지만 연산이 항상 최적으로 결합되는 것은 아닙니다 . 이중 부정을 제거하지 않는 T#3 의 표현식으로 이를 확인할 수 있습니다 .

손실 함수를 지정하면 계산식 편집기에서 모수(있는 경우)에 대한 도함수를 사용합니다. 그리고 모형에 대해 1차 도함수와 2차 도함수를 사용합니다(있는 경우).

도함수 메커니즘이 함수의 해석 도함수를 사용하는 방법을 모르면 NumDeriv 함수를 사용하여 수치 미분을 수행합니다. 이 경우 플랫폼에서는 인수의 델타 변경과 관련하여 함수의 변경을 평가하는 데 사용되는 델타를 표시합니다. 적절한 수치 미분을 얻으려면 다른 델타 설정으로 실험해야 할 수도 있습니다.

팁

주어진 모형을 나타내는 방법에는 항상 여러 가지가 있으며, 어떤 방법은 다른 형식보다 훨씬 우수한 성능을 보입니다. Ratkowsky 연구 자료(1990)에서 이러한 대체 형식을 다룹니다.

계산식의 여러 위치에서 반복되는 하위 표현식이 있는 경우 임시 변수에 할당하는 것이 좋습니다. 그런 다음 나중에 계산식에서 이 변수를 참조합니다. 예를 들어 위의 모형 계산식 중 하나는 다음과 같습니다.

```
If(Y==0, Log(1/(1+Exp(model))), Log(1 - 1/(1 + Exp(model))));
```

표현식을 인수 분해한 후 지역 변수에 할당하여 계산식을 단순화할 수 있습니다.

```
temp=1/(1+Exp(model));
If(Y==0, Log(temp), Log(1-temp));
```

도함수 기능은 할당과 조건 전체에서 도함수를 추적할 수 있습니다.

효과적인 비선형 모델링에 대한 참고 사항

중심화 다항식을 수행하는 것이 좋습니다.

모수에 대해 선형인 완전 다항식 항이 있는 곳이면 항상 다항식을 중심화하는 것이 좋습니다. 이렇게 하면 최적화를 위한 수치 표면의 조건이 개선됩니다. 다음과 같은 표현식을 예로 들어보겠습니다.

$$a_1 + b_1 x + c_1 x^2$$

다음과 같이 변환해야 합니다.

$$a_2 + b_2(x - \bar{x}) + c_2(x - \bar{x})^2$$

두 모형은 모수 변환을 제외하고는 동일하지만 모형이 비선형이면 두 번째 모형을 훨씬 쉽게 적합시킬 수 있습니다.

모수 변환은 쉽게 풀이할 수 있습니다.

$$\begin{aligned} a_1 &= a_2 - b_2\bar{x} + c_2\bar{x}^2 \\ b_1 &= b_2 - 2c_2\bar{x} \\ c_1 &= c_2 \end{aligned}$$

반복 수가 계속 최대값까지 진행되면 최대 반복 수를 늘리거나 수렴 기준 중 하나를 완화합니다.

모든 문제에 대해 잘 맞는 총괄 최적화 방법은 실제로 없습니다. JMP에는 비선형 플랫폼에서 해를 구할 수 있는 문제의 범위를 확대할 수 있도록 **Newton**, **QuasiNewton BFGS**, **QuasiNewton SR1** 및 **수치 미분만** 등의 옵션이 있습니다.

기본 설정이 특정 문제에 대한 해로 수렴할 수 없는 경우 이 설정의 다양한 조합을 사용하여 수렴 승산을 높입니다.

일부 모형은 모수의 시작 값에 매우 민감합니다. 새 시작 값으로 작업하는 것이 효과적일 때가 많습니다. 시작 값을 편집하고 **재설정**을 클릭하여 결과를 확인합니다. 그림도 유용합니다. 슬라이더를 사용하여 더 잘 적합되도록 곡선을 시각적으로 수정합니다. 모수 프로파일러가 도움을 줄 수 있지만 데이터 집합이 작은 경우 외에는 너무 느릴 수 있습니다.

16 장

JMP PRO 함수 데이터 탐색기

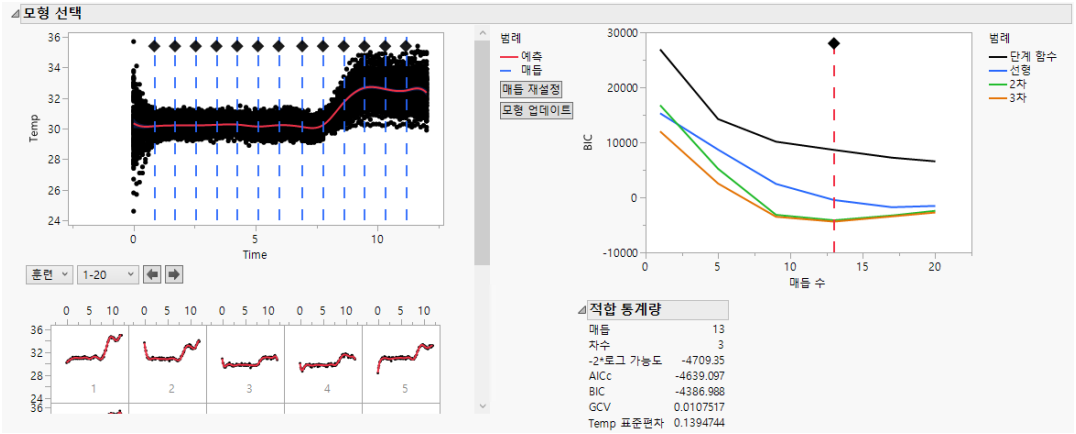
함수 데이터 분석 수행

함수 데이터 탐색기 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

FDE(함수 데이터 탐색기)는 함수, 신호 또는 계열 데이터를 위한 플랫폼입니다. 이 플랫폼은 탐색적 데이터 분석 도구 또는 차원 축소 기법으로 사용할 수 있습니다. 차원 축소의 경우 FDE 플랫폼은 함수 데이터를 다른 JMP 플랫폼에서 분석할 수 있는 형식으로 변환합니다.

출력 데이터에 대한 몇 가지 유형의 변환 및 입력 데이터에 대한 정렬을 포함하여 데이터 전처리 도구를 FDE 플랫폼에서 사용할 수 있습니다. B-스플라인, P-스플라인 또는 Fourier 기저 모형을 데이터에 적합시켜 함수 모형이 생성됩니다. 모형이 적합되면 함수 모형에 대해 함수 PCA(함수 주성분 분석)가 자동으로 수행됩니다. 함수 PCA 의 결과(예: 함수 주성분 스코어)는 특징 추출 또는 차원 축소를 위해 별도의 데이터 테이블에 저장됩니다.

그림 16.1 FDE 의 B- 스플라인 모형 선택



목차

함수 데이터 탐색기 플랫폼 개요 259
함수 데이터 탐색기의 예 259
함수 데이터 탐색기 플랫폼 시작 263
함수 데이터 탐색기 보고서 265
    모형 보고서 266
함수 데이터 탐색기 플랫폼 옵션 270
    함수 데이터 탐색기 그룹 옵션 271
    데이터 처리 보고서 옵션 272
    모형 보고서 옵션 275
함수 데이터 탐색기 플랫폼의 추가 예 276
    여러 함수 공정의 예 276
    함수 DOE의 예 279
함수 데이터 탐색기 플랫폼에 대한 통계 상세 정보 282
    함수 모형 적합 282
    함수 요약 상세 정보 283

JMP PRO 함수 데이터 탐색기 플랫폼 개요

함수 데이터는 측정값 집합이 곡선 또는 이미지를 구성하는 연속 도메인에 대해 기록된 데이터로 정의할 수 있습니다. 대개 도메인은 시간이고 측정값 집합은 ID 변수로 정의됩니다. 도메인의 특정 지점 t에서 ID 수준 i에 대한 함수 데이터 관측값을 $f_i(t)$로 표기합니다. 함수 데이터 탐색기 플랫폼을 사용하면 함수 데이터를 탐색하고 분석할 수 있습니다.

함수 데이터의 형식은 조밀하거나 희박할 수 있습니다. 조밀한 함수 데이터는 관측값이 ID 변수의 모든 수준에 대해 균등한 간격의 동일한 점 격자 위에 있을 때 발생합니다. 희박한 함수 데이터는 관측값이 도메인 전체에서 균등하지 않은 간격으로 배치되어 ID 수준의 관측값 수가 다를 때 발생합니다. 함수 데이터 탐색기 플랫폼에서는 두 가지 형태의 함수 데이터를 모두 처리할 수 있습니다.

함수 데이터를 여러 가지 방법으로 표현할 수 있지만 일반적으로 다음과 같은 두 경우로 분류할 수 있습니다 .

- 관심 반응 $f(t)$ 는 함수 형태를 가집니다 .
- 함수 형태를 갖는 하나 이상의 공변량 $f(t)'$가 있습니다. 이를 함수 또는 신호 공정이라고도 합니다 .

함수 데이터 탐색기 플랫폼은 모든 유형의 함수 데이터에 대한 탐색 도구로 유용합니다. 그러나 이 플랫폼의 강점은 많은 함수 공정(스칼라 반응과 연관될 수 있음)을 사용하고 이후 모델링에 사용할 주요 특징을 추출하는 것입니다. 이 작업을 수행하려면 먼저 B-스플라인, P-스플라인 또는 Fourier 기저 모형을 사용하여 함수 모형을 데이터에 적합시킵니다. 그런 다음 이 함수 모형에 대해 함수 PCA(함수 주성분 분석)가 수행됩니다. FPC(함수 주성분) 스코어와 같은 함수 PCA의 결과는 저장된 후 다른 모델링 플랫폼(예: 모형 적합 플랫폼의 일반화 회귀 분석법)에서 특징 추출 및 분석에 사용됩니다. 또는 FDE 플랫폼 내에서 보조 변수 집합을 지정하고 일반화 회귀 모형을 적합시켜 이러한 변수가 반응에 어떤 영향을 미치는지 확인할 수 있습니다.

함수 데이터 분석에 대한 자세한 내용은 Ramsay and Silverman(2005) 연구 자료에서 확인하십시오.

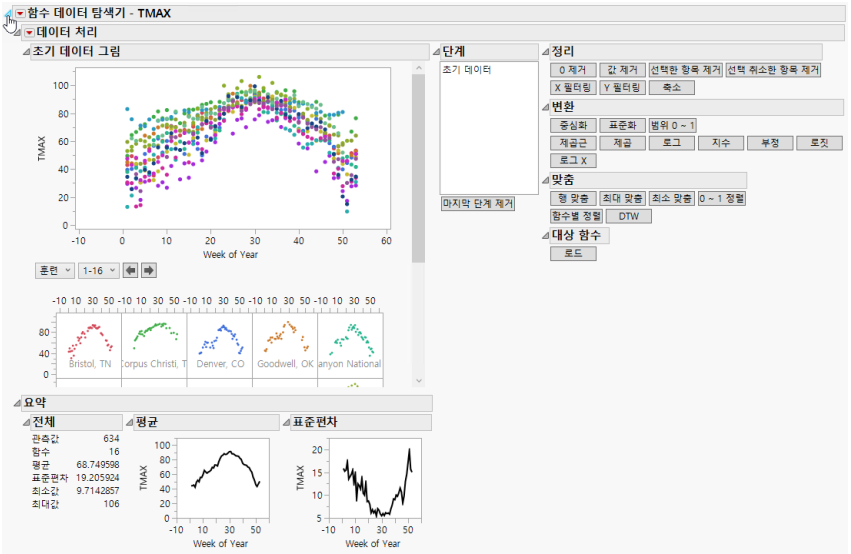
JMP PRO 함수 데이터 탐색기의 예

이 예에서는 미국 전역의 16 개 기상 관측소에서 수집한 주간 기상 데이터를 분석합니다 . 데이터 테이블의 "Weather Station Locations" 스크립트를 실행하여 위치 지도를 표시합니다. 일별 온도는 주간 평균으로 요약됩니다 . 모든 기상 관측소가 매주 (년 중) 에 대한 주간 온도 측정값을 가지고 있는 것은 아닙니다 . 이것은 희박 함수 데이터의 예입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Functional Data/Weekly Weather Data.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 함수 데이터 탐색기**를 선택합니다 .
3. TMAX 를 선택하고 **Y, 출력**을 클릭합니다 .

4. Week of Year 를 선택하고 **X, 입력**을 클릭합니다 .
5. ID 를 선택하고 **ID, 함수**를 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 16.2 초기 함수 데이터 탐색기 보고서



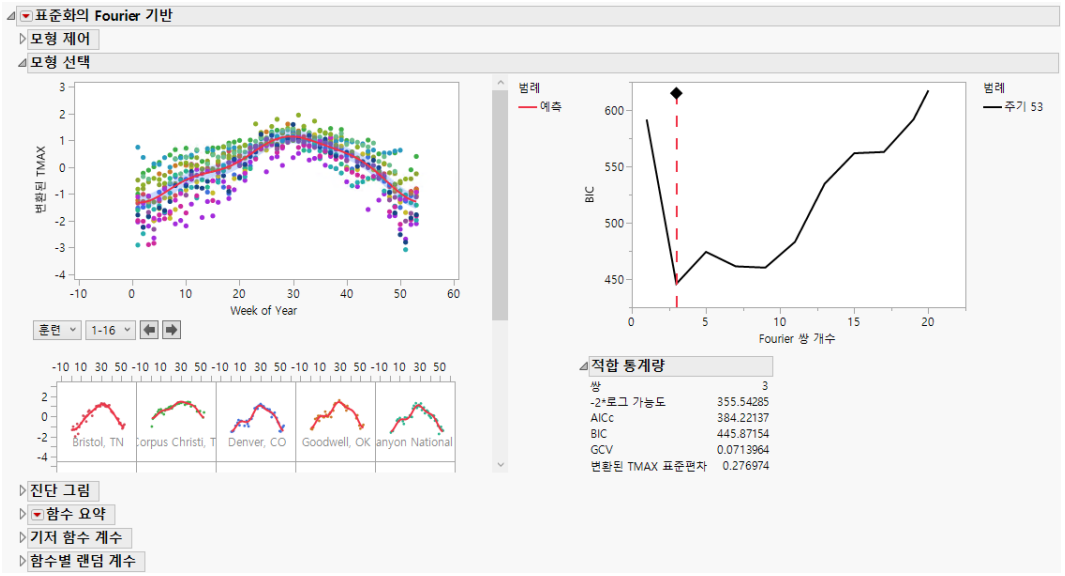
초기 " 함수 데이터 탐색기 " 보고서에는 원시 데이터 그림 , 요약 통계량 , 데이터의 함수 평균 및 함수 표준편차 요약 그림이 포함되어 있습니다 . 데이터 처리 옵션을 위한 버튼도 있습니다 . 데이터 처리 옵션은 " 데이터 처리 " 의 빨간색 삼각형 메뉴에서도 액세스할 수 있습니다 . 대개 모델링하기 전에 출력 데이터를 표준화하는 것이 좋습니다 .

7. " 변환 " 메뉴에서 **표준화** 버튼을 클릭합니다 .

   지정된 변환을 기반으로 데이터 그림과 요약 통계량이 업데이트됩니다 . " 단계 " 목록에 " 표준화 " 가 추가됩니다 .

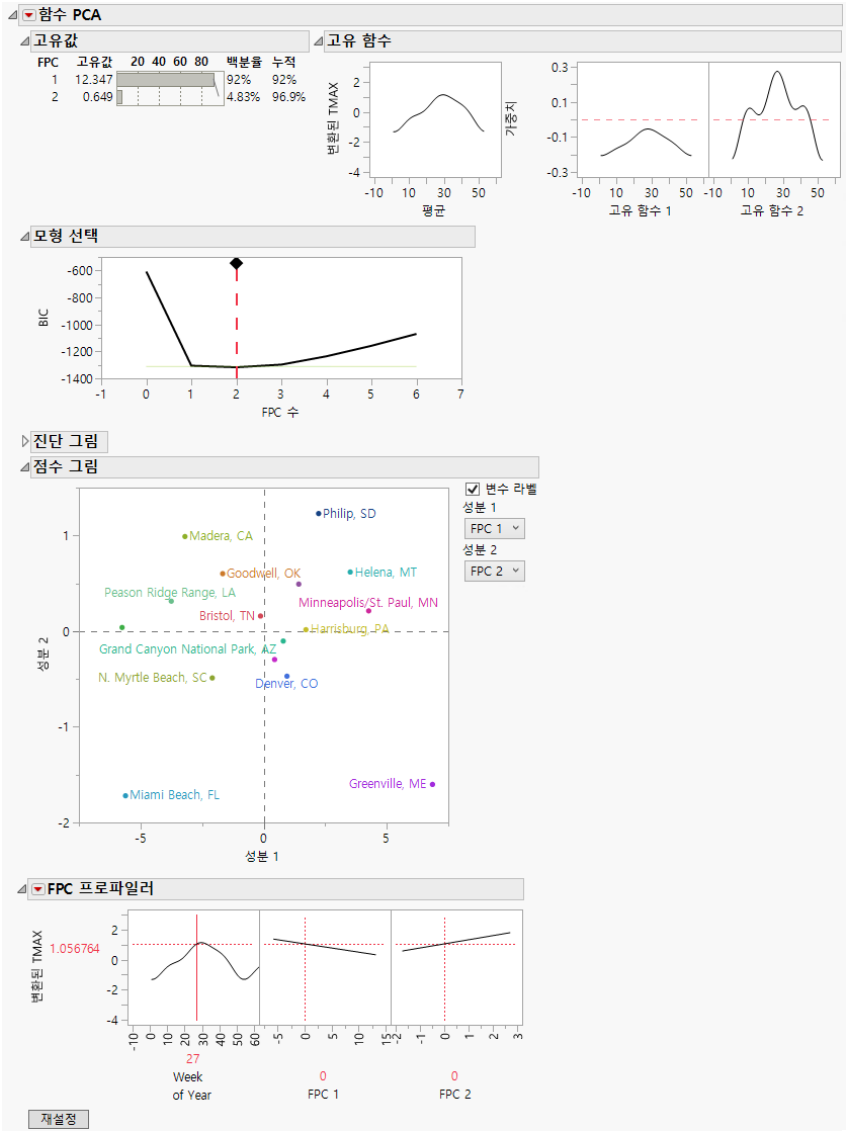
8. " 함수 데이터 탐색기 " 의 빨간색 삼각형을 클릭하고 **모형 > Fourier 기저**를 선택합니다 .

그림 16.3 Fourier 기저 모형 보고서



"Fourier 기저 " 보고서에는 선택한 모형에 대한 정보를 제공하는 여러 보고서가 포함되어 있습니다 . " 모형 선택 " 보고서에는 BIC 적합 기준에 따라 최량 적합 모형이 표시됩니다 . 기상 데이터의 경우 선택된 Fourier 기저 모형의 주기는 53 이고 기저 함수 쌍은 3 개입니다 . 모형에 대한 적합 통계량과 계수도 제공됩니다 . 아래로 스크롤하여 " 함수 PCA" 보고서를 봅니다 .

그림 16.4 함수 PCA 보고서



" 함수 PCA" 보고서에서는 처음 두 고유값이 데이터 변동의 거의 97% 를 설명한다는 것을 보여 줍니다 . " 모형 선택 " 그래프에서 빨간색 파선을 클릭한 후 3 FPC 로 드래그하면 처음 세 고유값이 데이터 변동의 99% 를 설명한다는 것을 확인할 수 있습니다 . 그러나 첫 번째 고유값만으로 92% 를 설명합니다 . " 스코어 그림 " 을 사용하여 다른 함수의 이상치인 개별 함수를 감지할 수 있습니다 . " 스코어 그림 " 에서 "Miami Beach, FL" 과 "Greenville, ME" 위치를 제외한 대부분의 위치가 함께 모여 있습니다 . 개별 함수 그림이 나올 때까지 위로 스크롤합니다 . Miami Beach 위치의 함수는 더 편평하므로 다른 위치에 비해 온도 변동이 적다는 것을

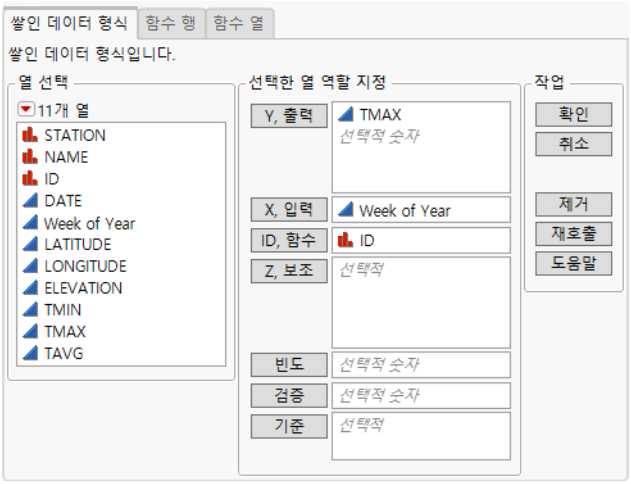
나타냅니다 . Greenville 위치의 함수는 최대값이 더 낮으므로 다른 위치에 비해 온도가 일관되게 낮음을 나타냅니다 .

팁 : 이상치를 더 잘 식별하려면 " 스코어 그림 " 보고서에서 " 변수 라벨 " 옵션을 선택 취소하십시오 .

JMP PRO 함수 데이터 탐색기 플랫폼 시작

분석 > 전문 모델링 > 함수 데이터 탐색기를 선택하여 함수 데이터 탐색기 플랫폼을 시작합니다 .

그림 16.5 함수 데이터 탐색기 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

함수 데이터 탐색기 시작 창에는 다양한 유형의 데이터 형식 탭이 있습니다 . 사용하는 데이터 형식에 따라 탭을 선택합니다 .

쌓인 데이터 형식 데이터 테이블의 각 행이 단일 관측값에 해당하는 경우 선택합니다 . 출력 , 입력 및 ID 변수에 대한 개별 열이 있습니다 .

참고 : 데이터 형식 중 " 쌓인 데이터 형식 " 만 함수 공정을 여러 개 지정할 수 있습니다 . " 쌓인 데이터 형식 " 탭의 **Y, 출력**에 둘 이상의 열을 할당하면 각 Y 변수가 개별적으로 분석됩니다 . " 적합 그룹 " 보고서에는 Y 변수에 대한 개별 보고서가 포함됩니다 .

함수 행 데이터 테이블의 각 행이 ID 변수의 한 수준에 대한 전체 출력 함수에 해당하는 경우 선택합니다 . 각 열은 입력 변수의 수준입니다 .

주의 : " 함수 행 " 형식은 "FDE X" 열 특성이 사용된 경우를 제외하고 입력 도메인에서 관측값의 간격이 균등하다고 가정합니다 . "FDE X" 열 특성을 사용하면 열 이름에 지정된 입력 변수를 이 데이터 형식에 사용할 수 있습니다 .

함수 열 데이터 테이블의 각 열이 ID 변수의 한 수준에 대한 전체 출력 함수에 해당하는 경우 선택합니다 . 각 행은 입력 변수의 수준에 해당합니다 .

JMP PRO 시작 창 옵션

Y, 출력 함수 공정 $f(t)$ 를 할당합니다 . ID 변수의 각 수준에 대해 두 개 이상의 관측된 출력 값이 있어야 합니다 .

참고 : 관측된 출력 값이 두 개 미만인 함수는 분석에서 제거됩니다 .

X, 입력 (" 쌓인 데이터 형식 " 및 " 함수 열 " 에 사용 가능) 입력 변수 t 를 할당합니다 . "X, 입력 " 에 변수를 지정하지 않으면 행 번호가 사용됩니다 . 행 번호를 사용하면 입력 도메인에서 관측값의 간격이 균등하다고 가정합니다 .

ID, 함수 (" 쌓인 데이터 형식 " 및 " 함수 행 " 에 사용 가능) 각 함수에 ID 변수를 할당합니다 . 쌓인 데이터의 경우 ID 변수가 할당되지 않으면 모든 관측값을 하나의 함수에서 가져온다고 가정합니다 .

Z, 보조 (" 쌓인 데이터 형식 " 및 " 함수 행 " 에 사용 가능) 하나 이상의 보조 변수를 할당합니다 . 보조 변수는 함수 데이터 탐색기 플랫폼의 계산에 사용되지 않으며 이 변수를 포함해도 결과에 영향을 주지 않습니다 . 보조 변수는 향후 함수 데이터 탐색기의 결과를 분석하는 데 사용할 수 있는 변수입니다 . 보조 변수를 지정하면 " 데이터 저장 " 및 " 요약 저장 " 옵션으로 생성되는 테이블의 Supplementary 열 그룹에 포함됩니다 . 이러한 열에는 원래 데이터 테이블에 지정된 열 특성이 그대로 유지됩니다 . 적합 모형에 " 함수 DOE 분석 " 옵션도 사용할 수 있습니다 . 자세한 내용은 " 함수 DOE 분석 " 에서 확인하십시오 .

빈도 (" 쌓인 데이터 형식 " 에만 사용 가능) 분석의 각 행에 대한 빈도를 나타내는 숫자 값이 포함된 열을 할당합니다 . 빈도 열의 효과는 데이터 테이블을 확장하여 정수 빈도가 k 인 행을 k 개의 동일한 행으로 확장하는 것입니다 .

검증 검증 데이터 집합을 정의하는 선택적 숫자 열을 할당합니다 . 이 열에는 구분되는 값이 두 개만 포함되어야 합니다 . 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 . 값이 세 개 이상인 경우 가장 작은 값이 훈련 데이터 집합을 정의하고 다른 모든 값은 검증 데이터 집합을 정의합니다 . FDE 플랫폼에서는 검증 열을 사용하여 모형을 훈련하고 평가합니다 . 검증에 대한 자세한 내용은 "JMP 모델링의 검증 " 에서 확인하십시오 .

참고 : " 검증 " 옵션을 사용하면 각 함수의 관측값 표본이 아니라 전체 함수를 홀드아웃할 수 있습니다 . 따라서 ID 값이 동일한 모든 관측값이 테스트 또는 검증으로 분류되어야 합니다 . ID 값이 동일한 관측값이 두 데이터 집합에 모두 포함될 수 없습니다 . 이 유형의 검증 열에 대한 자세한 내용은 " 그룹화 검증 열 " 에서 확인하십시오 .

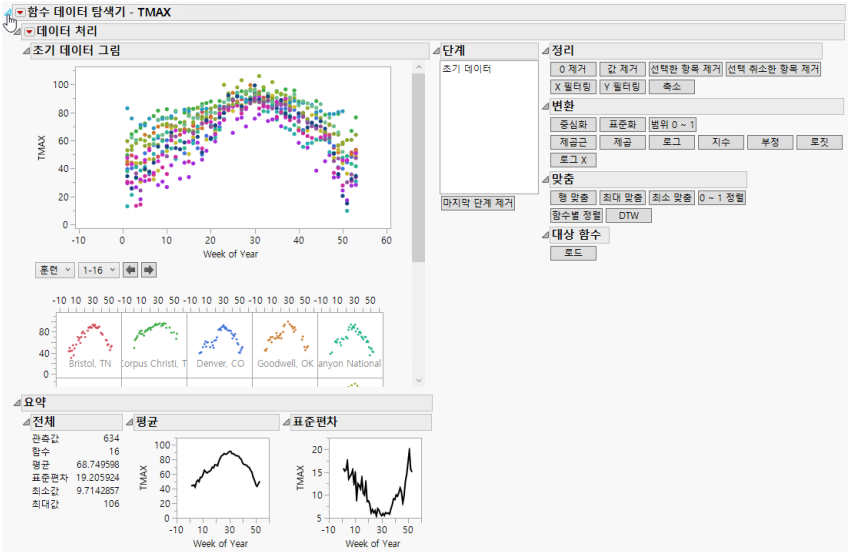
주의 : " 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하고 데이터 테이블에 검증 열을 추가하면 그룹화 검증 열이 되지 않습니다 . 대신 검증 열 생성 플랫폼을 사용합니다 . 자세한 내용은 " 그룹화 검증 열 " 에서 확인하십시오 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 할당합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

JMP PRO 함수 데이터 탐색기 보고서

초기 " 함수 데이터 탐색기 " 보고서에는 " 데이터 처리 " 보고서와 " 요약 " 보고서가 포함됩니다 .

그림 16.6 초기 함수 데이터 탐색기 보고서



" 데이터 처리 " 보고서에는 모든 관측값의 초기 데이터 그림과 ID 변수 수준에 해당하는 개별 데이터 그림 격자가 포함되어 있습니다 . 격자에 표시되는 최대 개별 그림 수는 20 개입니다 . 격자 위에는 표시할 개별 그림을 대화식으로 선택할 수 있는 드롭다운 메뉴와 화살표가 있습니다 . 검증 데이터 집합을 지정한 경우 첫 번째 드롭다운 메뉴를 사용하여 훈련 데이터 집합과 검증 데이터 집합 중에서 표시되는 개별 그림의 출처를 선택할 수 있습니다 . 함수가 20 개를 초과하는 경우 두 번째 드롭다운 메뉴와 화살표를 사용하여 다양한 그림 그룹을 표시할 수 있습니다 . 모든 그래프에는 입력 변수의 값에 대한 함수 반응 데이터가 표시됩니다 . " 데이터 처리 " 보고서에는 데이터 처리 버튼도 포함되어 있습니다 .

" 요약 " 보고서에는 관측값 수 , 함수 수 , 전체 평균 , 표준편차 , 최소값 및 최대값을 포함한 전체 요약 통계량 테이블이 포함되어 있습니다 . 함수 평균 및 함수 표준편차 그림도 있습니다 . 그림에 표시된 함수 요약 통계량은 입력 변수의 각 고유 값에서 계산됩니다 .

"데이터 처리" 및 "요약" 보고서는 모형을 적합시키거나 직접 함수 주성분 분석을 수행하기 전에 데이터를 전처리하는 데 유용합니다. "데이터 처리" 보고서에 있는 데이터 정리, 변환 및 맞춤 버튼은 "데이터 처리"의 빨간색 삼각형 메뉴에 있는 옵션에 해당합니다. 자세한 내용은 "데이터 처리 보고서 옵션"에서 확인하십시오. 전처리 단계를 수행할 때마다 그래프와 요약 결과가 자동으로 업데이트됩니다. 여러 전처리 단계를 순서대로 수행할 수 있습니다. 각 단계가 "데이터 처리" 보고서의 "단계" 목록에 추가됩니다. 가장 최근 단계를 제거하려면 "단계" 개요 아래의 **마지막 단계 제거** 버튼을 클릭합니다.

JMP PRO 모형 보고서

"함수 데이터 탐색기"의 빨간색 삼각형 메뉴에 있는 "모형" 하위 메뉴 옵션을 사용하여 모형을 데이터에 적합시킬 수 있습니다. 사용 가능한 모형은 "모형"에서 확인하십시오. 다른 유형의 모형을 데이터에 적합시킬 때마다 모형 보고서가 나타납니다. 각 모형 보고서에는 다음 보고서가 포함됩니다.

- "모형 제어"
- "모형 선택"
- "진단 그림"
- "함수 요약"
- "기저 함수 계수"
- "함수별 랜덤 계수"
- "함수 PCA"

JMP PRO 모형 제어

"모형 제어" 보고서를 사용하면 "모형 선택" 보고서에서 비교할 모형의 모수를 정의할 수 있습니다. "모형 제어" 보고서의 모양은 적합되는 모형 유형에 따라 달라집니다.

B- 스플라인 및 P- 스플라인 모형 제어

B- 스플라인 또는 P- 스플라인 모형이 적합되면 다음 모수를 지정할 수 있습니다.

매듭 수 각 스플라인의 매듭 수를 추가 또는 제거하거나 범위를 지정합니다. 매듭 수는 0 이 아닌 정수여야 합니다.

> **참고:** B- 스플라인 모형에 허용되는 최대 매듭 수는 함수당 최대 관측값 수 또는 고유 입력 수보다 한 개 적습니다. P- 스플라인 모형에 허용되는 최대 매듭 수는 고유 입력 수보다 두 개 적습니다. 최대값보다 큰 값을 지정하면 경고 메시지가 나타납니다.

스플라인 차수 "모형 선택" 보고서에서 스플라인 차수 적합을 추가하거나 제거합니다.

Fourier 기저 모형 제어

Fourier 기저 모형이 적합되면 다음 모수를 지정할 수 있습니다.

Fourier 쌍 개수 비교할 Fourier 쌍 개수를 추가 또는 제거하거나 범위를 지정합니다.

주기 함수의 주기를 변경합니다.

모형 제어를 지정한 후 **시작**을 클릭하면 "모형 선택" 보고서에 업데이트된 모형이 표시됩니다.

팁: 모형을 적합시키기 전에 모형 제어를 지정하려면 Shift 키를 누른 채 "함수 데이터 탐색기"의 빨간색 삼각형을 클릭한 후 원하는 모형을 선택합니다. 자세한 내용은 "모형"에서 확인하십시오.

JMP PRO 모형 선택

"모형 선택" 보고서에는 전체 예측 그림, 개별 예측 그림 격자, 해 경로 그림 및 적합 통계량 테이블이 포함되어 있습니다. 개별 예측 그림 격자는 "데이터 처리" 보고서의 개별 그림 격자와 레이아웃 및 컨트롤이 동일합니다. 한 번에 최대 20개의 그림이 표시됩니다. 개별 예측 그림의 다양한 그룹을 볼 수 있는 드롭다운 메뉴와 화살표가 있습니다. 해 경로 그림에서는 모형 모수의 값에 대해 표시된 모형 선택 기준을 보여 줍니다. BIC(Bayes 정보 기준)가 기본 적합 기준입니다. 자세한 내용은 "모형 보고서 옵션"에서 확인하십시오. B-스플라인 및 P-스플라인 모형의 경우 정의된 매듭 수에 대해 각 스플라인 차수의 개별 해 경로가 표시됩니다. Fourier 기저 모형의 경우 정의된 주기 동안 Fourier 쌍 개수에 대해 해 경로가 표시됩니다. "모형 선택" 옵션을 사용하여 해 경로 그림 모수를 변경할 수 있습니다.

현재 해는 해 경로 그림에서 수직 점선으로 지정됩니다. 기본적으로 모형 선택 기준 값이 가장 작은 모형에 해당하는 매듭 수 또는 Fourier 쌍 개수 위치에 슬라이더가 있습니다. 수직 점선 맨 위의 슬라이더를 드래그하여 현재 모형의 매듭 수 또는 Fourier 쌍 개수를 변경할 수 있습니다. 슬라이더를 드래그하면 "모형 선택" 보고서의 예측 그림과 다른 모든 보고서의 정보가 자동으로 업데이트됩니다.

"적합 통계량" 테이블에는 현재 해 모형에 대한 설명이 포함되어 있습니다. -2*로그 가능도, AICc, BIC 및 GCV 모형 적합 기준 값, "<Y, **출력**> 표준편차"로 표시되는 반응 표준편차 값도 표시됩니다. 반응 표준편차는 적합 모형의 잔차 시그마로 정의됩니다. P-스플라인 모형을 선택한 경우 벌점 모수 λ(람다)도 표시됩니다.

예측 그림에는 원시 데이터 및 현재 모형에 해당하는 예측 곡선이 표시됩니다. 검증 데이터 집합이 있는 경우 검증 데이터 집합에 포함된 함수에 대해서는 예측 곡선이 표시되지 않습니다. 스플라인 모형의 경우 선택된 기본 모형은 최량 적합의 스플라인 차수입니다. 해 경로 그림 또는 범례에서 특정 스플라인을 클릭하여 현재 모형 선택을 변경할 수 있습니다. 전체 예측 그림의 곡선은 평균 예측 곡선입니다. 개별 예측 그림의 곡선은 각 특정 함수에 대한 예측 곡선입니다. B-스플라인 모형의 경우 전체 예측 그림에 매듭 위치도 표시됩니다. 파란색 슬라이더 막대를 다른 위치로 드래그하여 매듭 위치를 변경할 수 있습니다. 새 매듭 위치에 따라 모형 보고서를 업데이트하려면 **모형 업데이트** 버튼을 클릭합니다. 매듭을 기본 위치로 재설정하려면 **매듭 재설정** 버튼을 클릭합니다.

JMP PRO 진단 그림

"진단 그림" 보고서에는 "실제값 대 예측값 그림"과 "잔차 대 예측값 그림"이 포함되어 있습니다. 이러한 그림은 현재 모형이 데이터에 얼마나 적합한지 평가하는 데 도움이 됩니다. "진단 그림" 보고서는 기본적으로 닫혀 있습니다.

JMP PRO 함수 요약

ID 변수의 각 수준에 대한 함수 PCA 의 요약을 표시합니다 . 데이터 변동을 1% 넘게 설명하는 고유값과 관련된 함수 주성분이 기본적으로 표시됩니다 . 평균 , 표준편차 , 중앙값 , 최소값 , 최대값 , 통합 차이 , RISE(제곱근 적분 제곱 오차) 및 RIFS(제곱근 적분 제곱 함수) 도 표시됩니다 . 통합 차이 및 RISE 요약 값은 ID 특정 함수가 전체 평균 함수와 얼마나 다른지 판단하는 데 사용됩니다 . RIFS 요약 값은 최적 곡선 적합을 위해 사용됩니다 . 자세한 내용은 " 함수 요약 상세 정보"에서 확인하십시오. "함수 요약"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

함수 요약 사용자 정의 " 함수 요약 " 보고서에 표시되는 FPC 수와 요약 통계량을 선택할 수 있는 창을 표시합니다 . 표시할 FPC 수가 지정되면 " 함수 PCA" 보고서도 업데이트됩니다 . 또한 " 요약 저장 " 옵션으로 생성되는 데이터 테이블에 각 함수에 대한 그래프를 포함할지 여부를 결정하는 " 그래프 저장 " 체크박스도 있습니다 .

팁 : 함수 공정이 여러 개인 경우 Ctrl 키를 클릭한 채 " 함수 요약 사용자 정의 " 를 선택하면 동일한 요약 값을 표시하도록 모든 " 함수 요약 " 보고서를 사용자 정의할 수 있습니다 .

요약 저장 " 함수 요약 " 보고서에 지정된 요약 통계량을 새 데이터 테이블에 저장합니다 . 새 데이터 테이블의 이름에 모형 적합이 설명되어 있습니다 . 이 데이터 테이블에는 고유 함수 , 평균 함수 , 예측 함수 및 조건부 예측 함수에 대한 계산식 열이 포함됩니다 . 또한 각 함수에 대해 지정된 모형 적합 및 원시 데이터의 그래프 이미지를 포함하는 열도 있습니다 . 데이터 테이블에는 예측 계산식 및 조건부 예측 계산식에 대한 예측 프로파일러를 시작하는 프로파일러 스크립트가 있습니다 . 이러한 계산식은 입력 변수 , ID 변수 및 고유 함수의 함수입니다 .

JMP PRO 기저 함수 계수

추정된 기저 함수 계수와 해당 표준편차를 표시합니다. 이 값은 ID 변수의 모든 수준에서 공통이며 혼합 모형 프레임워크의 고정 추정값입니다. 계수에 대한 표준 오차 및 신뢰 구간을 보려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열**을 선택합니다.

JMP PRO 함수별 랜덤 계수

각 기저 함수 및 함수 공정 조합에 대해 추정된 랜덤 계수를 표시합니다. 이 값은 ID 변수의 각 수준에 대해 고유하며 혼합 모형 프레임워크의 임의 효과 추정값입니다.

JMP PRO 함수 PCA

적합된 함수 모형에 대해 함수 PCA(함수 주성분 분석)가 수행됩니다. "함수 PCA" 보고서에는 각 FPC(함수 주성분)에 해당하는 고유값이 가장 큰 것부터 순서대로 나열됩니다. 각 FPC에 의해 설명되는 변동 백분율과 누적 백분율이 나열되고 막대 차트에 표시됩니다. 평균 함수의 그래프 및 각 성분에 대한 그래프가 있습니다. 성분 그래프에는 고유 함수의 값이 표시됩니다.

"함수 PCA" 보고서에서 모형 선택을 수행하여 선택한 함수 주성분 수를 다시 정의할 수 있습니다. BIC(Bayes 정보 기준) 대 FPC 수를 보여 주는 해 경로 그림이 있습니다. 현재 FPC 수는 해 경로 그림에서 수직 점선으로 지정됩니다. FPC 수가 다른 모형의 적합이 비슷할 수 있습니다. 따라서 해 경로 그림에서는 BIC 통계량 값의 구간을 나타내는 영역을 제공합니다. 녹색 영역과 노란색 영역이 있습니다. 녹색 영역에는 최소 BIC에서 최소 BIC + 4까지 구간에 해당하는 값이 포함되고, 노란색 영역에는 최소 BIC + 4에서 최소 BIC + 10까지 구간에 해당하는 값이 포함됩니다. 기본적으로 녹색 영역 내에서 FPC 수가 가장 적은 모형이 선택됩니다. 수직선 맨 위의 슬라이더를 드래그하여 FPC 수를 변경할 수 있습니다. 슬라이더를 드래그하면 "함수 PCA" 보고서의 다른 정보가 자동으로 업데이트됩니다.

참고 : 영역이 그림에서 작게 나타날 수 있습니다 . y 축을 확대하면 영역이 더 보기 좋게 표시됩니다 .

직접 함수 PCA가 수행되면 전체 예측 그림과 개별 예측 그림 격자도 표시됩니다. 개별 예측 그림 격자는 "데이터 처리" 보고서의 개별 그림 격자와 레이아웃 및 컨트롤이 동일합니다. 한 번에 최대 20개의 그림이 표시되며, 개별 예측 그림의 다양한 그룹을 볼 수 있는 드롭다운 메뉴와 화살표가 있습니다. FPC 수를 업데이트하면 예측 그림도 자동으로 업데이트됩니다.

예측 그림에는 원시 데이터 및 현재 모형에 해당하는 예측 곡선이 표시됩니다. 검증 데이터 집합이 있는 경우 검증 데이터 집합에 포함된 함수에 대해서는 예측 곡선이 표시되지 않습니다. 전체 예측 그림의 곡선은 지정된 FPC 수에서 평균 예측 곡선입니다. 개별 예측 그림의 곡선은 지정된 FPC 수에서 각 특정 함수에 대한 예측 곡선입니다.

참고 : 단일 함수만 모델링된 경우 " 함수 PCA" 보고서가 표시되지 않습니다 . 그렇지 않은 경우 JMP 에서 함수 PCA 를 수행할 수 없으면 " 함수 PCA" 보고서에 오류 메시지가 나타납니다 .

" 함수 PCA" 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

진단 그림 " 실제값 대 예측값 그림 " 과 " 잔차 대 예측값 그림 " 을 표시하거나 숨깁니다 . 이러한 그림은 선택된 함수 주성분 수를 고려하여 모형이 데이터에 얼마나 적합한지 평가하는 데 도움이 됩니다 .

스코어 그림 FPC 스코어의 스코어 그림을 표시하거나 숨깁니다 . " 성분 선택 " 아래의 목록을 사용하여 스코어 그림의 각 축에 표시할 FPC 를 지정할 수 있습니다 . FPC 가 하나뿐인 경우 $y = x$ 선에 FPC 스코어가 표시되고 성분 변경 목록은 표시되지 않습니다 . 스코어 그림은 이상치를 감지하는 데 유용합니다 . FPC 스코어의 경우 스코어 그림을 사용하여 ID 변수에서 이상치 함수가 있는 수준을 감지할 수 있습니다 . 스코어 그림에서 점을 선택하면 FPC 프로파일러가 해당 함수의 스코어로 설정됩니다 .

팁 : 스코어 그림의 점을 커서로 가리키면 ID 변수의 해당 수준에 대한 적합 곡선 그림이 표시됩니다 .

FPC 프로파일러 FPC 스코어의 프로파일러를 표시하거나 숨깁니다 . "FPC 프로파일러 " 에는 입력 변수에 대한 열과 각 FPC 스코어에 대한 열이 포함됩니다 . 지정된 각 대상 함수에 대해 두 개의 추가 프로파일러가 있습니다 . 하나는 대상 함수와의 차이를 측정하고 다른 하나는 대상 함수와의 적분 오차를 측정합니다 . "FPC 프로파일러 " 의 빨간색 삼각형 메뉴 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

팁: 프로파일러에서 모든 FPC 스코어를 0으로 재설정하려면 "재설정" 버튼을 사용하십시오.

FPC 수 사용자 정의 " 함수 PCA" 에 표시할 FPC 스코어 수를 지정합니다 . 이 옵션에서 FPC 스코어 수를 지정하면 " 함수 요약 " 보고서도 업데이트됩니다 . 평균 모형을 보려면 FPC 수를 0 으로 설정합니다 .

JMP PRO 함수 데이터 탐색기 플랫폼 옵션

" 함수 데이터 탐색기 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

요약 다음과 같은 함수 요약 통계량 옵션을 제공하는 하위 메뉴입니다 .

평균 그림 " 요약 " 보고서에 함수 평균 그림을 표시하거나 숨깁니다 . 기본적으로 설정되어 있습니다 .

표준편차 그림 "요약" 보고서에 함수 표준편차 그림을 표시하거나 숨깁니다 . 기본적으로 설정되어 있습니다 .

중앙값 그림 " 요약 " 보고서에 함수 중앙값 그림을 표시하거나 숨깁니다 .

모형 다음과 같은 모형 옵션을 제공하는 하위 메뉴입니다 .

B- 스플라인 기저 스플라인 (B- 스플라인) 모형을 데이터에 적합시킵니다 . 비주기 데이터의 경우 B- 스플라인 모형을 사용합니다 .

P- 스플라인 벌점 기저 스플라인 (P- 스플라인) 모형을 데이터에 적합시킵니다 .

Fourier 기저 Fourier 기저 모형을 데이터에 적합시킵니다 . 주기 데이터의 경우 Fourier 기저 모형을 사용합니다 . 주기 모형은 함수의 시작 위치와 완료 위치가 같다고 가정합니다 . 자세한 내용은 "Fourier 기저 모형 " 에서 확인하십시오 .

참고 : 고유 입력 값이 세 개 미만인 경우 Fourier 기저 모형이나 P- 스플라인 모형을 데이터에 적합시킬 수 없으며 경고 메시지가 나타납니다 .

모형 제어 모형을 적합시키기 전에 " 모형 제어 " 패널을 열 수 있는 하위 메뉴를 표시합니다 . 자세한 내용은 " 모형 제어 " 에서 확인하십시오 .

직접 함수 PCA (단일 함수만 있는 경우에는 사용 불가능) 기저 함수 모형을 먼저 적합시키지 않고 데이터에 대해 함수 주성분 분석을 직접 수행합니다 . 이렇게 하면 특히 대규모 데이터 집합의 경우 계산 시간이 감소합니다 . 직접 함수 PCA 는 다음과 같이 구현됩니다 .

1. 입력 데이터를 0 에서 1 사이로 맞추고 관측값을 입력 값의 공통 격자에 보간합니다 .
2. 데이터에 대해 함수 주성분 분석을 수행합니다 .
3. 각 격자점에 매듭이 있는 P-스플라인 모형을 사용하여 첫 번째 고유 함수를 평활화합니다 .
4. 데이터에서 첫 번째 평활 고유 함수를 제거하고 , 많은 양의 데이터 변동이 설명될 때까지 2 단계 ~ 4 단계를 반복합니다 .

직접 함수 PCA 를 수행하면 " 함수 PCA" 보고서가 표시됩니다 . 자세한 내용은 " 함수 PCA" 에서 확인하십시오 .

데이터 저장 처리된 데이터를 새 데이터 테이블에 저장합니다 . 처리된 데이터는 쌓인 데이터 형식으로 저장됩니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO 함수 데이터 탐색기 그룹 옵션

함수 데이터 탐색기 그룹 옵션은 시작 창의 " 쌓인 데이터 형식 " 탭에서 함수 공정을 여러 개 지정한 경우에만 사용할 수 있습니다 .

요약 모든 함수 공정에 대한 함수 요약 통계량 그림을 표시하거나 숨기는 옵션 하위 메뉴입니다 . 자세한 내용은 " 요약 " 에서 확인하십시오 .

데이터 처리 모든 함수 공정에 적용할 수 있는 데이터 처리 옵션 하위 메뉴입니다 . 자세한 내용은 " 데이터 처리 보고서 옵션 " 에서 확인하십시오 .

모형 모든 함수 공정에 적합시킬 수 있는 모형 옵션 하위 메뉴입니다 . 자세한 내용은 " 모형 " 에서 확인하십시오 .

데이터 저장 모든 함수 공정에 대해 처리된 데이터를 새 데이터 테이블에 저장합니다. 처리된 데이터는 쌓인 데이터 형식으로 저장됩니다 .

요약 저장 각 " 함수 요약 " 보고서에 지정된 요약 통계량을 새 데이터 테이블에 저장합니다 . 새 데이터 테이블의 이름에 모형 적합이 설명되어 있습니다 . 이 데이터 테이블에는 각 Y 변수에 대한 고유 함수 , 평균 함수 , 예측 함수 및 조건부 예측 함수의 계산식 열이 포함됩니다 . 데이터 테이블에는 각 Y 변수에 대해 예측 계산식 및 조건부 예측 계산식에 대한 예측 프로파일러를 시작하는 프로파일러 스크립트가 있습니다 . 이러한 계산식은 입력 변수 , ID 변수 및 고유 함수의 함수입니다 .

행으로 배열 창에서 가로로 표시되는 보고서 수를 지정합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO 데이터 처리 보고서 옵션

" 데이터 처리 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

정리 다음과 같은 데이터 정리 옵션이 포함된 하위 메뉴입니다 .

0 제거 값이 0 인 관측값을 제거합니다 . 데이터에 0 이 없으면 0 값이 없음을 나타내는 경고가 표시됩니다 .

값 제거 데이터에서 제거할 값을 지정할 수 있는 규격 창을 표시합니다 .

선택한 항목 제거 데이터 테이블에서 선택한 행에 해당하는 관측값을 제거합니다 .

선택 취소한 항목 제거 데이터 테이블에서 선택하지 않은 행에 해당하는 관측값을 제거합니다 .

주의 : " 선택한 항목 제거 " 및 " 선택 취소한 항목 제거 " 는 행 번호를 제거합니다 . " 자동 재계산 " 이 활성화된 경우 이러한 옵션을 사용하기 전에 행을 추가하거나 삭제해야 합니다 .

X 필터링 지정된 구간을 벗어나는 X 값을 제거합니다 . "X 필터링 " 옵션을 선택하면 " 다음 값 미만 " 및 " 다음 값 초과 " 값을 지정해야 합니다 . 지정된 구간을 벗어나는 X 값은 분석에 사용되지 않습니다 .

Y 필터링 지정된 구간을 벗어나는 Y 값을 제거합니다 . "Y 필터링 " 옵션을 선택하면 " 다음 값 미만 " 및 " 다음 값 초과 " 값을 지정해야 합니다 . 지정된 구간을 벗어나는 Y 값은 분석에 사용되지 않습니다 .

축소 다음 기법 중 하나를 사용하여 X 값에 대한 데이터를 줄입니다 .

– " 격자 " 탭을 사용하여 관측값을 값의 공통 격자에 보간합니다 . 격자 크기를 지정할 수 있습니다 . 기본적으로 격자 크기는 고유 입력 값 수의 절반이므로 총 관측값 수가 감소합니다 . 총 관측값 수를 줄이는 것에는 관심이 없지만 관측값이 동일한 격자에 있게 하려면 격자 크기를 고유 입력 값 수로 지정합니다 .

– " 계급 " 탭을 사용하여 고유 X 값에서 균등한 간격으로 지정된 수의 계급을 생성합니다 . 각 함수 (또는 "ID, 함수 " 변수의 수준) 에 대해 계급 내의 관측값을 평균화하여 해당 계급 수준에 대한 Y 값을 생성합니다 .

– " 줄임 " 탭을 사용하여 X 값에 대한 모든 N 개 관측값을 제거합니다 . 여기서 N 은 지정된 줄임 비율에 의해 결정됩니다 . 이 작업은 각 함수 (또는 "ID, 함수 " 변수의 수준) 에 대해 수행됩니다 . 줄임 비율의 기본값은 2 이며 , 이 값은 각 함수에서 관측값의 절반을 제거합니다 .

참고 : " 제거 " 옵션을 사용하면 지정된 관측값이 분석 및 모델링 보고서에서 제외되지만 데이터 테이블에서 해당 관측값이 변경되지는 않습니다 .

변환 다음과 같은 데이터 변환 옵션이 포함된 하위 메뉴입니다 .

중심화 결과를 가운데에 맞춥니다 .

표준화 평균 0, 분산 1 이 되도록 데이터를 중심에 맞추고 척도를 조정하여 결과를 표준화합니다 .

범위 0 ~ 1 0 ~ 1 범위에 들어가도록 결과의 척도를 조정합니다 .

제곱근 결과의 제곱근을 계산하여 데이터를 변환합니다. 출력 값은 음수가 아니어야 합니다.

제곱 결과의 제곱을 계산하여 데이터를 변환합니다 .

로그 결과의 자연 로그를 계산하여 데이터를 변환합니다 .

지수 결과의 지수 함수를 계산하여 데이터를 변환합니다 .

부정 결과에 음수를 취해 데이터를 변환합니다 .

로짓 결과의 로짓 함수를 계산하여 데이터를 변환합니다. 출력 값은 0에서 1 사이여야 합니다.

로그 X 입력의 자연 로그를 계산하여 데이터를 변환합니다 .

맞춤 다음과 같은 입력 데이터 맞춤 옵션이 포함된 하위 메뉴입니다 .

행 맞춤 입력 값을 행 번호로 바꿉니다 .

최대 맞춤 각 ID 수준에 대해 관측된 최대 출력 값을 사용하여 함수를 정렬합니다 . 관측된 최대 출력 값과 관련된 입력 값은 각 ID 수준에 대해 0 으로 설정되고 , 다른 입력 값은 관측된 최대값과 0 사이의 차이에 따라 위 또는 아래로 이동합니다 .

최소 맞춤 각 ID 수준에 대해 관측된 최소 출력 값을 사용하여 함수를 정렬합니다 . 관측된 최소 출력 값과 관련된 입력 값은 각 ID 수준에 대해 0 으로 설정되고 , 다른 입력 값은 관측된 최소값과 0 사이의 차이에 따라 위 또는 아래로 이동합니다 .

0 ~ 1 정렬 입력 값의 범위가 0 ~ 1 이 되도록 출력 함수를 정렬합니다 .

팁 : "0 ~ 1 정렬 " 은 P- 스플라인 모형을 적합시킬 때 특히 유용합니다 .

함수별 정렬 각 함수가 입력 값의 전체 최소값에서 시작하여 입력 값의 전체 최대값에서 끝나도록 출력 함수를 정렬합니다 .

DTW(동적 시간 뒤틀림) (함수가 두 개 이상인 경우에만 사용 가능) DTW(동적 시간 뒤틀림) 를 사용하여 출력 함수를 정렬합니다 . DTW 는 둘 이상의 함수를 서로 맞추기 위해 최적의 뒤틀림을 찾는 함수 맞춤 기법입니다 . DTW 옵션을 선택하면 " 참조 함수 선택 " 창이 나타납니다 . 이를 사용하여 참조 함수를 선택합니다 . 참조 함수는 나머지 함수를 맞출 때 사용할 기준 함수입니다 .

참조 함수를 선택하고 " 확인 " 을 클릭하면 나머지 쿼리 함수 목록과 함께 뒤틀림 함수 그림이 표시됩니다 . 뒤틀림 함수 그림에서 참조 함수는 y 축에 있고 선택한 쿼리 함수는 x 축에 있습니다 . 빨간색 대각선 ($y = x$) 과의 편차는 맞춤을 개선하기 위해 쿼리 함수의 입력에 뒤틀림이 적용되었음을 나타냅니다 .

대상 함수 (함수가 두 개 이상인 경우에만 사용 가능) 대상 함수를 로드할 수 있는 하위 메뉴입니다 .

대상 로드 대상 함수를 지정할 수 있는 창을 표시합니다 . 대상 함수는 모든 함수가 대상 함수처럼 보이도록 곡선 매칭에 사용됩니다 . 두 개의 대상 함수를 지정하여 나머지 곡선을 " 최선 " 및 " 더 나쁜 " 함수와 비교할 수도 있습니다 .

하나 이상의 대상 함수를 지정하면 함수의 데이터가 모형 적합에 사용되지 않습니다 . 지정된 각 대상 함수에 대해 두 개의 행이 FPC 프로파일러에 추가됩니다 . 자세한 내용은 "FPC 프로파일러 " 에서 확인하십시오 .

참고 : 다른 전처리 단계를 수행하기 전에 대상 함수를 로드해야 합니다 .

DTW(동적 시간 뒤틀림) 옵션

뒤틀림 함수 그림 뒤틀림 함수 그림을 표시하거나 숨깁니다 . 기본적으로 설정되어 있습니다 .

거리 행렬 저장 거리 행렬을 별도의 데이터 테이블에 저장합니다 . 거리 행렬은 함수를 군집화하는 데 유용할 수 있습니다 . 거리 행렬 데이터 테이블에는 " 계층적 군집화 " 스크립트가 포함되어 있습니다 .

뒤틀림 함수 저장 뒤틀림 함수를 별도의 데이터 테이블에 저장합니다. 데이터 테이블의 각 행에는 DTW 조정 입력 변수, 원래 입력 변수 및 ID 변수가 포함됩니다.

JMP PRO 모형 보고서 옵션

모형 선택 모형 선택 기준으로 사용할 옵션이 포함된 하위 메뉴를 표시합니다. AICc, BIC 및 GCV 중에서 선택할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

그림 기반 모든 기저 함수 그림을 하나의 그래프에 표시하거나 숨깁니다.

진단 그림 " 진단 그림 " 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 진단 그림 " 에서 확인하십시오.

함수 요약 " 함수 요약 " 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 함수 요약 " 에서 확인하십시오.

기저 함수 계수 " 기저 함수 계수 " 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 기저 함수 계수 " 에서 확인하십시오.

함수별 랜덤 계수 " 함수별 랜덤 계수 " 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 함수별 랜덤 계수 " 에서 확인하십시오.

함수 PCA " 함수 PCA" 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 함수 PCA" 에서 확인하십시오.

함수 DOE 분석 (시작 창에서 하나 이상의 보조 변수를 지정한 경우에만 사용 가능) FDE 플랫폼 내에서 " 일반화 회귀 " 보고서를 시작합니다. 일반화 회귀 모형은 보조 변수를 모형 효과로 사용하여 각 FPC 스코어 함수에 적합됩니다. 기본적으로 2 차 요인 모형이 적합되고 " 추정 방법 " 은 " 최량 부분집합 " 입니다. 범주형 2 차 효과는 포함되지 않습니다. 모형의 항 수가 21 개를 초과하거나 함수 수가 1000 개를 초과하면 " 추정 방법 " 이 자동으로 " 가지 치기 전진 선택 " 으로 전환됩니다. 아니면 원래 데이터 테이블에서 원하는 모형 적합을 정의하는 모형 스크립트를 지정할 수도 있습니다. 보조 변수를 사용하여 FPC 스코어를 모델링하면 모형 적합을 사용하여 보조 변수에 따라 반응이 어떻게 변하는지 확인할 수 있습니다. 보조 변수가 반응에 어떤 영향을 주는지 탐색하려면 FDOE 프로파일러를 사용합니다.

" 함수 DOE 분석 " 보고서에는 다음과 같은 빨간색 삼각형 메뉴 옵션이 있습니다.

FPC 스코어에 대한 일반화 회귀 각 FPC 스코어에 대한 " 일반화 회귀 " 보고서를 표시하거나 숨깁니다. " 일반화 회귀 " 모형 보고서에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

진단 그림 반응 변수에 대한 실제값 대 예측값 그림과 잔차 그림을 표시하거나 숨깁니다.

FDOE 프로파일러 보조 변수에 따라 반응이 어떻게 변하는지 탐색할 수 있는 FDOE 프로파일러를 표시하거나 숨깁니다. "FPC 프로파일러 " 의 빨간색 삼각형 메뉴 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오.

열 저장 " 예측 계산식 저장 " 및 " 잔차 계산식 저장 " 옵션이 있는 목록을 표시합니다. 이러한 옵션은 해당 계산식을 데이터 테이블의 새 열에 저장합니다.

데이터 저장 모델링된 데이터를 새 데이터 테이블에 저장합니다 . 모델링된 데이터는 쌓인 데이터 형식으로 저장됩니다 .

적합 제거 지정된 적합에 대한 모형 보고서를 제거합니다 .

JMP PRO 함수 데이터 탐색기 플랫폼의 추가 예

- " 여러 함수 공정의 예 "
- " 함수 DOE 의 예 "

JMP PRO 여러 함수 공정의 예

이 예에서는 Fermentation Process.jmp 및 Fermentation Process Batch Yield Results.jmp 샘플 데이터 테이블을 사용하여 효소 생산을 분석합니다 . Yield 는 유전자 변형 효모에 의해 생산된 효소의 양입니다 . 12 시간 동안 동일한 시간 간격으로 수집된 배치당 100 개의 공정 측정값이 있습니다 .

함수 데이터 탐색기 플랫폼을 사용하여 모형을 데이터에 적합시키고 함수 주성분을 새 데이터 테이블에 저장합니다 . 그런 다음 모형 적합 플랫폼의 일반화 회귀 분석법으로 함수 주성분을 분석합니다 .

함수 모형 적합

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Functional Data/Fermentation Process.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 함수 데이터 탐색기**를 선택합니다 .
3. " 쌓인 데이터 형식 " 탭에서 Ethanol ~ pH 를 선택하고 **Y, 출력**을 클릭합니다 .
4. Time 을 선택하고 **X, 입력**을 클릭합니다 .
5. BatchID 를 선택하고 **ID, 함수**를 클릭합니다 .
6. **확인**을 클릭합니다 .
7. " 함수 데이터 탐색기 그룹 " 의 빨간색 삼각형을 클릭하고 **데이터 처리 > 맞춤 > 0 ~ 1 정렬**을 선택합니다 . 이렇게 하면 각 " 함수 데이터 탐색기 " 보고서에서 입력 변수의 범위가 0 에서 1 사이로 맞춰집니다 .
8. " 함수 데이터 탐색기 그룹 " 의 빨간색 삼각형을 클릭하고 **모형 > B- 스플라인**을 선택합니다 . 이렇게 하면 각 함수 공정에 B- 스플라인 모형이 적합됩니다 .

그림 16.7 Ethanol 에 대한 함수 데이터 탐색기 보고서

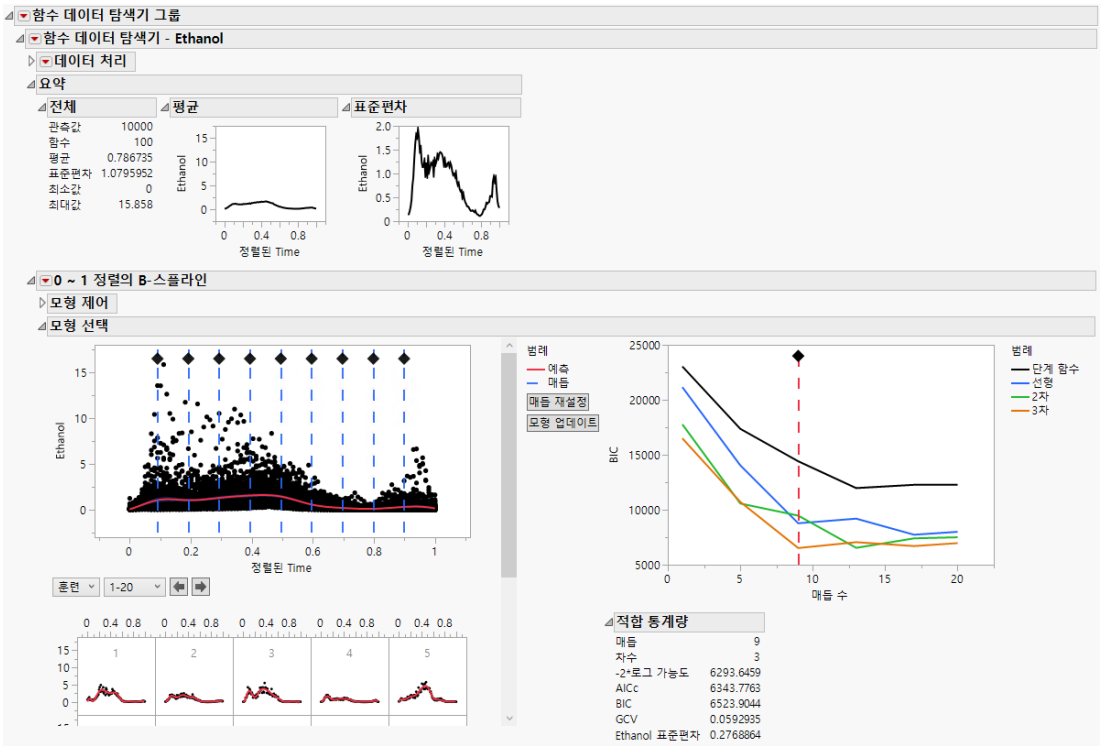


그림 16.8 Ethanol 에 대한 모형 요약 보고서

함수 요약

BatchID	FPC 1	FPC 2	FPC 3	FPC 4	FPC 5	FPC 6	평균	표준편차	중앙값	최소값	최대값	통합 차이	제곱근 적분 제곱 오차	제곱근 적분 함수 제곱
1	0.6845185	0.4272826	0.3090187	-0.174881	-0.261418	0.2534972	1.2856117	1.2574705	0.7225838	0.0275282	3.4532315	-0.297809	0.9637933	1.7832112
2	0.2003363	-0.162552	0.1458985	-0.006107	0.0059254	0.0042297	0.848664	0.71297	0.7468215	0.0333247	2.0280261	-0.736315	0.317081	1.1007059
3	1.0570202	0.1670821	-0.025524	0.1554196	0.1599784	-0.213975	1.5221477	1.3465031	1.4533318	-0.018659	3.8349448	-0.060525	1.1267728	2.0232175
4	-0.225892	-0.342518	0.0456513	0.0469127	0.095793	-0.009734	0.5248517	0.4519468	0.4568174	0.0261908	1.6643826	-1.061051	0.4292202	0.6897988
5	0.7066471	0.8878144	-0.416058	0.1123721	0.0512395	0.1779	1.4912892	1.4352359	0.828356	0.068398	4.6865036	-0.091374	1.2326783	2.0610547
6	-0.164355	0.0402334	0.0698129	0.0188387	-0.004275	-0.031643	0.648615	0.4651278	0.6306233	0.0425897	1.5070566	-0.936894	0.1937517	0.7960802
7	0.4840975	0.0604706	-0.02545	0.138502	-0.044549	-0.079275	1.188724	0.8264775	1.3331523	0.0450854	2.337651	-0.395091	0.5320304	1.4384966
8	0.0598675	0.3232751	-0.194786	0.1468215	0.0999753	-0.021161	0.8913351	0.7786373	0.7196766	0.0029737	2.5753376	-0.69338	0.4349227	1.1820858
9	-0.35775	-0.244627	0.0950901	0.0453786	0.0302305	0.027442	0.4330103	0.3329415	0.4001359	0.0494026	1.0654932	-1.153327	0.4507855	0.5436156
10	0.1030905	0.3791323	-0.32985	-0.309402	0.1014623	0.1254908	1.0564703	0.84232	0.8304595	0.068398	3.0158387	-0.528012	0.622775	1.343811

기저 함수 계수

모수	평균 추정값	표준편차 추정값
정렬된 Time[0]	0.068398	0
정렬된 Time[0.04]	0.3628335	0.6072559
정렬된 Time[0.1]	1.446524	2.3982214
정렬된 Time[0.19]	0.8752812	0.9861143
정렬된 Time[0.29]	1.3486495	1.4351681
정렬된 Time[0.39]	1.5859873	1.4953555
정렬된 Time[0.49]	1.69221	1.2326668
정렬된 Time[0.6]	0.396788	0.7335251
정렬된 Time[0.7]	0.2061623	0.2751322
정렬된 Time[0.8]	0.0235344	0
정렬된 Time[0.89]	0.3734323	0.6151888
정렬된 Time[0.95]	0.3903473	1.3136877
정렬된 Time[1]	0.1561339	0.1818355

그림 16.7 및 그림 16.8 에서는 함수 공정 변수 중 하나인 Ethanol 에 대한 모형 보고서를 보여 줍니다 . 전체 보고서를 스크롤하여 각 공정 변수에 대한 모형 적합을 확인합니다 . 그런 다음 " 함수 요약 " 보고서의 FPC 를 분석에 사용합니다 .

FPC 저장 및 Yield 결과 연결

1. Ctrl 키를 누른 채 " 함수 요약 " 의 빨간색 삼각형을 클릭하고 **함수 요약 사용자 정의**를 선택합니다 .
2. **표시할 FPC 수 입력** 옆의 상자에 "3" 을 입력합니다 .
3. **모든 요약 선택 취소** 상자를 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 함수 데이터 탐색기 그룹 " 의 빨간색 삼각형을 클릭하고 **요약 저장**을 선택합니다 .
6. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Functional Data/Fermentation Process Batch Yield Results.jmp 를 엽니다 .
7. 함수 데이터 탐색기 모형 요약 .jmp 데이터 테이블에서 BatchID 를 마우스 오른쪽 버튼으로 클릭하고 **연결 ID** 를 선택 취소합니다 .
8. 함수 데이터 탐색기 모형 요약 .jmp 데이터 테이블에서 BatchID 를 마우스 오른쪽 버튼으로 클릭하고 **연결 참조** > Fermentation Process Batch Yield Results.jmp 를 선택합니다 .

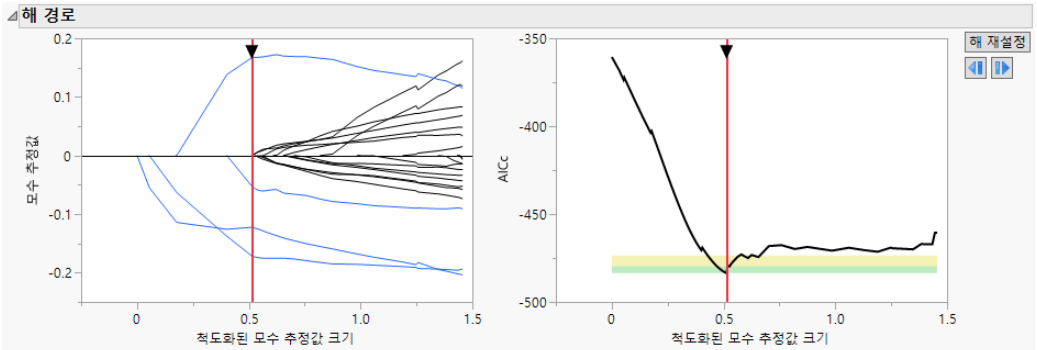
   이렇게 하면 yield 데이터 테이블과 요약 데이터 테이블이 결합됩니다 .

일반화 회귀 모형 적합

모형 적합 플랫폼의 일반화 회귀 분석법을 사용하여 함수 공정 변수가 Yield 에 미치는 영향을 확인합니다 .

1. 함수 데이터 탐색기 모형 요약 .jmp 데이터 테이블에서 **분석 > 모형 적합**을 선택합니다 .
2. "BatchID 에 의해 Fermentation Process Batch Yield Results 로 참조됨 " 옆의 삼각형을 클릭합니다 .
3. Yield[BatchID] 를 선택하고 **Y** 를 클릭합니다 .
4. Time 및 BatchID 를 제외한 나머지 열을 선택하고 **추가**를 클릭합니다 .
5. " 분석법 " 을 " 일반화 회귀 " 로 변경합니다 .
6. **실행**을 클릭합니다 .
7. **적응형** 상자를 선택합니다 .
8. **시작**을 클릭합니다 .

그림 16.9 Batch Yield 에 대한 일반화 회귀 보고서



원래 예측 변수에 대한 모수 추정값

항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	0.53436	0.0020214	69880.551	<.0001*	0.5303981	0.5383219
Ethanol FPC 1	0	0	0	1.0000	0	0
Ethanol FPC 2	-0.042765	0.0060948	49.23396	<.0001*	-0.054711	-0.03082
Ethanol FPC 3	0	0	0	1.0000	0	0
Temp FPC 1	0	0	0	1.0000	0	0
Temp FPC 2	0	0	0	1.0000	0	0
Temp FPC 3	0	0	0	1.0000	0	0
Molasses Feed FPC 1	0	0	0	1.0000	0	0
Molasses Feed FPC 2	8.4212e-5	1.4668e-5	32.960213	<.0001*	5.5463e-5	0.000113
Molasses Feed FPC 3	0	0	0	1.0000	0	0
NH3 Feed FPC 1	0	0	0	1.0000	0	0
NH3 Feed FPC 2	-0.000598	8.3927e-5	50.851939	<.0001*	-0.000763	-0.000434
NH3 Feed FPC 3	0	0	0	1.0000	0	0
Air FPC 1	-9.712e-6	3.8557e-6	6.3454451	0.0118*	-1.727e-5	-2.156e-6
Air FPC 2	0	0	0	1.0000	0	0
Air FPC 3	0	0	0	1.0000	0	0
Tank Level FPC 1	0	0	0	1.0000	0	0
Tank Level FPC 2	0	0	0	1.0000	0	0
Tank Level FPC 3	0	0	0	1.0000	0	0
pH FPC 1	0	0	0	1.0000	0	0
pH FPC 2	0	0	0	1.0000	0	0
pH FPC 3	0	0	0	1.0000	0	0

"일반화 회귀" 보고서에 따르면 Ethanol, Molasses Feed, NH3 Feed 및 Air 의 특정 성분이 Yield 에 유의한 영향을 줍니다. 모형의 R^2 은 0.73225 입니다. 먼저 FDE 를 사용하여 함수 공정에 대해 차원 축소를 수행하면 변수의 수를 크게 줄이는 동시에 적절한 예측 모형을 생성할 수 있습니다.

함수 DOE 의 예

이 예에서는 Formulation For Homogeneity DOE.jmp 데이터 테이블을 사용하여 서로 다른 온도에서 성분 조합이 화합물의 균질성 등급에 어떤 영향을 주는지 살펴봅니다. 화합물을 구성하는 세 가지 성분은 Active, Water 및 Solvent 라는 라벨이 지정된 기본 성분입니다. 32 개의 성분 조합이 있으며 이러한 조합은 Formulation 열의 값으로 표시됩니다. 각 배합 (Formulation) 에 대해 섭씨 20 도에서 70 도까지 5 도 간격으로 Homogeneity Grade 가 측정됩니다. 서로 다른 온도에 대한 이러한 측정값을 그림에 표시하여 각 배합에 대한 곡선을 만들 수 있습니다. 함수 데이터 탐색기 플랫폼을 사용하여 Active, Water 및 Solvent 값에 따라 곡선이 어떻게 변하는지 확인합니다.

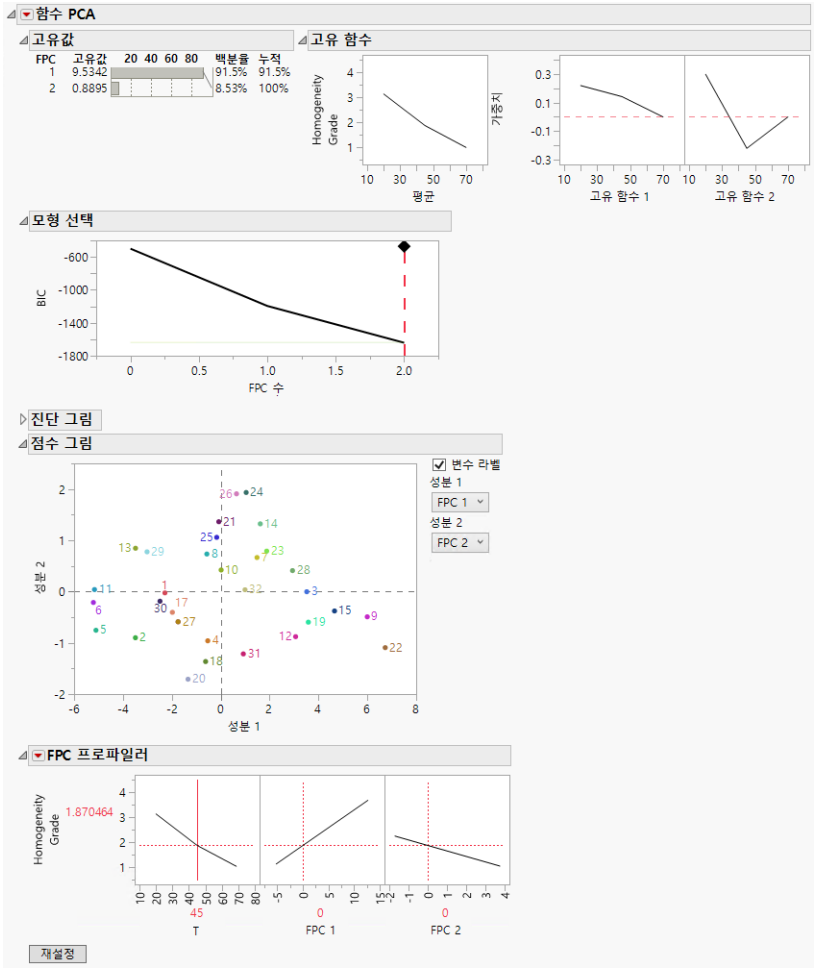
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Functional Data/Formulation For Homogeneity DOE.jmp 를 엽니다.

2. **분석 > 전문 모델링 > 함수 데이터 탐색기**를 선택합니다 .
3. " 쌓인 데이터 형식 " 탭에서 Homogeneity Grade 를 선택하고 **Y, 출력**을 클릭합니다 .
4. T 를 선택하고 **X, 입력**을 클릭합니다 .
5. Formulation 을 선택하고 **ID, 함수**를 클릭합니다 .
6. Solvent, Active, Water 를 선택하고 **Z, 보조**를 클릭합니다 .

   이러한 변수는 초기 함수 데이터 분석에 사용되지 않습니다 . 그러나 여기서 보조 변수로 지정하면 나중에 모형을 적합시킬 때 사용할 수 있습니다 .
7. **확인**을 클릭합니다 .
8. " 함수 데이터 탐색기 " 의 빨간색 삼각형을 클릭하고 **모형 > B- 스플라인**을 선택합니다 .

   최량 적합 모형은 매듭이 하나인 선형 스플라인 모형입니다 .

그림 16.10 Homogeneity Grade 에 대한 함수 PCA



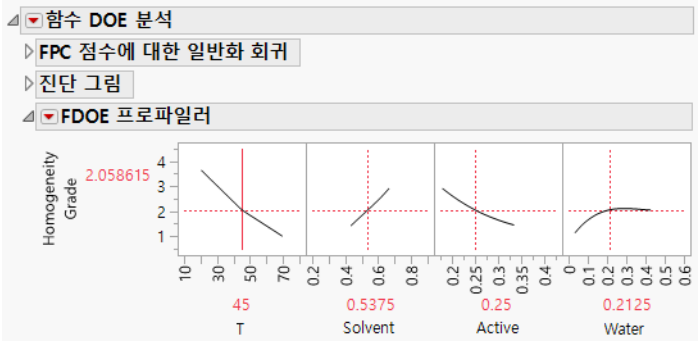
" 함수 PCA" 보고서에서는 데이터의 변동을 설명하려면 두 개의 함수 주성분이 필요하다는 것을 보여 줍니다. "FPC 프로파일러 " 에서는 서로 다른 온도에서 FPC 가 Homogeneity Grade 에 미치는 영향을 보여 줍니다 . 그러나 실제 성분이 Homogeneity Grade 에 어떤 영향을 미치는지 알 수 있는 방법은 없습니다 .

9. " 초기 데이터의 B- 스플라인 " 옆에 있는 빨간색 삼각형을 클릭하고 **함수 DOE 분석**을 선택합니다 .

   " 함수 DOE 분석 " 옵션은 세 개의 보조 변수를 예측 변수로 사용하여 일반화 회귀 모형을 각 주성분 스코어에 적합시킵니다 . 데이터 테이블에 "Model" 스크립트가 포함되어 있으므로 적합된 모형은 스크립트에 정의된 모형을 기반으로 합니다 .

참고 : 데이터 테이블에 "Model" 스크립트가 포함되어 있지 않으면 2 차 요인 모형이 적합됩니다 .

그림 16.11 Homogeneity Grade 에 대한 FDOE 프로파일러



"FDOE 프로파일러 " 를 사용하면 실제로 관심 있는 예측 변수에 따라 Homogeneity Grade 가 어떻게 변하는지 탐색할 수 있습니다 . 전반적으로 Active 양이 증가하면 Homogeneity Grade 가 감소하는 것으로 보입니다 .

참고 : 이 예에서는 Active, Solvent 및 Water 값의 합이 1 이 되도록 제한됩니다 . 이 제약 조건은 프로파일러에 반영됩니다 .

JMP PRO 함수 데이터 탐색기 플랫폼에 대한 통계 상세 정보

- "함수 모형 적합"
- "함수 요약 상세 정보"

함수 모형 적합

함수 데이터 탐색기 플랫폼의 모든 모형 적합은 기저 함수 확장에 의존합니다. 기저 함수는 독립 함수의 집합입니다. 함수 $f(t)$ 는 ϕ_k 로 표시되는 K 개 기저 함수의 선형 조합을 사용하여 근사화할 수 있습니다. 일반적으로 함수는 다음과 같이 근사화됩니다.

$$\hat{f}(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

여기서 c_k 모수는 기저 계수입니다. 평활 정도는 기저 함수의 수 K에 의해 결정됩니다. 기저 함수 확장에 대한 자세한 내용은 Ramsay and Silverman(2005) 연구 자료에서 확인하십시오.

JMP PRO Fourier 기저 모형

Fourier 기저를 정의하는 $k = 1,..., K$ 개 함수는 $\phi_0 = 1$, $\phi_{2k-1} = \sin(r\omega t)$ 및 $\phi_{2k} = \cos(r\omega t)$ 가 되도록 정의됩니다. 그러면 근사 함수가 다음과 같이 정의됩니다.

$$\hat{f}_i(t) = c_{i0} + c_{i1}\sin\omega t + c_{i2}\cos\omega t + c_{i3}\sin 2\omega t + c_{i4}\cos 2\omega t + \ldots$$

이는 $A = 2\pi/\omega$ 로 정의되는 주기 A 를 갖는 주기 함수에 근사한 값을 산출합니다. 계수는 고정 기저 함수 계수와 함수별 랜덤 계수의 조합입니다. 각 c 는 다음과 같이 정의됩니다.

$$c_{ik} = \beta_k + \alpha_{ik}$$

여기서 β_k는 기저 함수 k의 고정 계수이고, α_{ik}는 특정 함수 공정 i에 대한 기저 함수 k의 랜덤 계수입니다. Fourier 기저 모형의 경우 β_k 및 α_{ik} 모수의 추정값은 각각 "기저 함수 계수" 테이블과 "함수별 랜덤 계수" 테이블에서 확인할 수 있습니다.

Fourier 기저 모형에는 하나의 절편 항과 Fourier 쌍이라고 하는 동일한 수의 사인 항 및 코사인 항이 있습니다. 따라서 K는 항상 홀수입니다. 예를 들어 $K = 7$이면 절편 항 하나와 세 개의 Fourier 쌍이 있습니다.

함수 요약 상세 정보

"함수 요약" 보고서에는 통합 차이, RISE(제곱근 적분 제곱 오차) 및 RIFS(제곱근 적분 제곱 함수) 요약 통계량이 포함됩니다. 이 섹션에서는 이러한 요약 통계량을 정의합니다. 이 섹션에 나오는 모든 방정식은 입력 값이 0 ~ 1 범위에 맞춰져 있다고 가정합니다. 모든 함수에 대해 다음 표기가 사용됩니다.

$\hat{f}_i(x)$ 입력 값 x 에서 곡선 i 에 대한 추정 함수

$\hat{m}(x)$ 입력 값 x 에서의 추정 평균 함수

통합 차이

통합 차이는 ID 특정 곡선의 점과 전체 평균 곡선의 점에 대한 평균 차이입니다. 이 값을 사용하면 ID 특정 곡선이 대체로 평균 곡선 위에 있는지 아니면 아래에 있는지 판별할 수 있습니다. 곡선 i 에 대한 통합 차이는 다음과 같이 정의됩니다.

$$\mathrm{ID}_i = \int_0^1 (\hat{f}_i(x) - \hat{m}(x))dx$$

RISE

RISE(제곱근 적분 제곱 오차) 는 ID 특정 곡선과 전체 평균 곡선 사이의 평균 거리입니다. 이 값을 사용하면 전체 평균 곡선과 가장 유사하거나 가장 다른 ID 특정 곡선을 판별할 수 있습니다. 곡선 i 에 대한 RISE 요약 값은 다음과 같이 정의됩니다.

$$\mathrm{RISE}_i = \sqrt{\int_0^1 (\hat{f}_i(x) - \hat{m}(x))^2 dx}$$

RIFS

RIFS(제곱근 적분 제곱 함수) 는 최적 곡선 시나리오를 분석하기 위한 요약 값입니다. 최적 곡선을 미리 빼는 경우 가장 작은 RIFS 값을 사용하여 최적 곡선에 가장 가까운 함수를 판별할 수 있습니다. 곡선 i 에 대한 RIFS 요약 값은 다음과 같이 정의됩니다.

$$\mathrm{RIFS}_i = \sqrt{\int_0^1 \hat{f}_i(x)^2 dx}$$

17 장

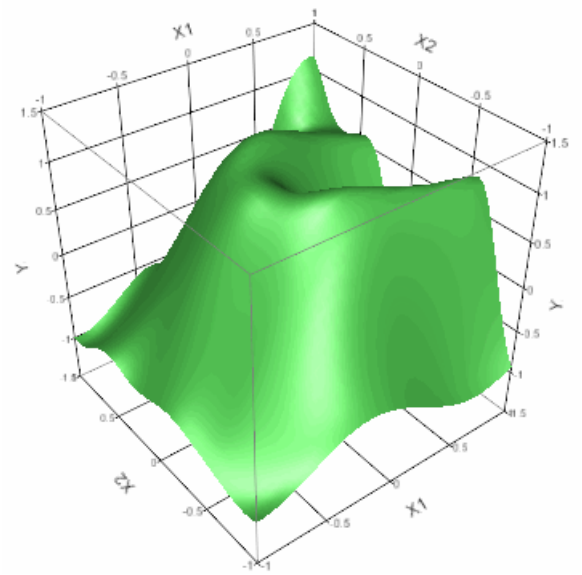
가우스 과정
평활 모형을 사용하여 데이터 적합

가우스 과정 플랫폼을 사용하여 연속형 반응과 하나 이상의 예측 변수 간의 관계를 모델링할 수 있습니다. 이 유형의 모형은 유한 요소 코드의 출력과 같은 컴퓨터 시뮬레이션 실험에서 흔히 볼 수 있으며 데이터를 완벽하게 보간하는 경우가 많습니다. 가우스 과정은 오차 항이 없는 모형을 처리할 수 있습니다. 이러한 모형은 입력 값이 동일하면 항상 동일한 출력 값이 생성됩니다.

가우스 과정 플랫폼에서는 공간 상관 모형을 데이터에 적합시킵니다. 독립 변수의 값이 멀어질수록 두 관측값 사이의 반응 상관이 감소합니다.

이 플랫폼을 사용하는 한 가지 목적은 추가 분석 및 최적화에 사용할 수 있는 예측 계산식을 얻는 것입니다.

그림 17.1 가우스 과정 예측 표면 예제



목차

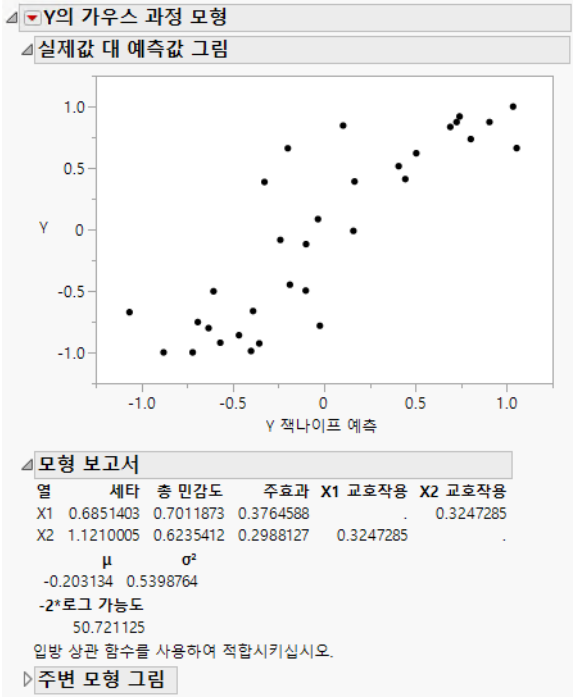
가우스 과정의 예 287
가우스 과정 플랫폼 시작 289
가우스 과정 보고서 290
실제값 대 예측값 그림 290
모형 보고서 290
주변 모형 그림 291
가우스 과정 플랫폼 옵션 291
가우스 과정 플랫폼의 추가 예 292
가우스 과정 모형의 예 292
범주형 예측 변수가 있는 가우스 과정 모형의 예 294
가우스 과정 플랫폼에 대한 통계 상세 정보 296
연속형 예측 변수가 있는 모형 296
범주형 예측 변수가 있는 모형 297
분산 계산식 파라미터화 298
모형 적합 상세 정보 298

가우스 과정의 예

이 예에서는 Y(반응) 에 대한 결정적 방정식과 함께 두 변수의 공간 채움 설계 데이터를 사용합니다 . 가우스 과정 플랫폼을 사용하여 Y 에 대한 X1 및 X2 의 설명력을 찾을 수 있습니다 . 열 계산식에서 Y 에 대한 방정식을 볼 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 2D Gaussian Process Example.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 가우스 과정**을 선택합니다 .
3. X1 과 X2 를 선택하고 **X** 를 클릭합니다 .
4. Y 를 선택하고 **Y** 를 클릭합니다 .
5. **상관 유형 > 입방**을 선택합니다 .
6. JMP PRO **빠른 GASP** 를 선택 취소합니다 .
7. **확인**을 클릭합니다 .

그림 17.2 가우스 과정 보고서

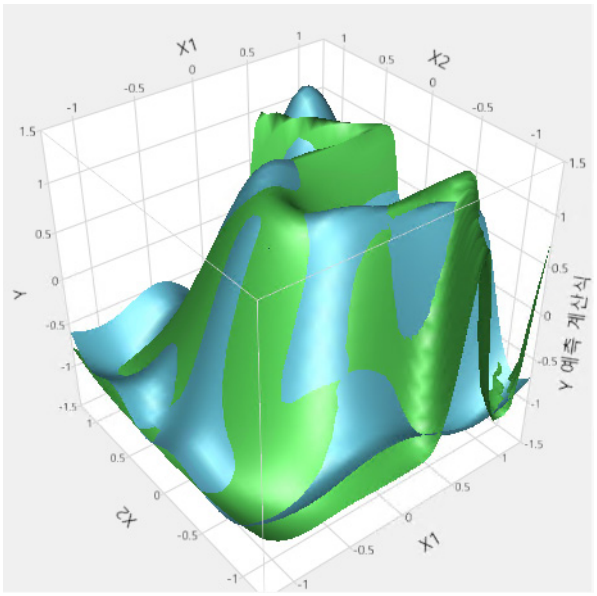


참고 : 최소화 루틴의 다른 시작점 , 상관 유형 선택 및 너겟 모수 포함으로 인해 추정된 모수가 다를 수 있습니다 .

이제 적합된 표면을 원래 표면과 비교하여 시각화합니다 .

8. "Y 의 가우스 과정 모형 " 옆의 빨간색 삼각형을 클릭하고 **예측 계산식 저장**을 선택합니다 .
9. **그래프 > 표면 그림**을 선택합니다 .
10. X1 ~ Y 예측 계산식을 선택하고 **열**을 클릭합니다 .
11. **확인**을 클릭합니다 .
12. " 표면 " 열에서 "Y 예측 계산식 " 에 대해 **양쪽**을 선택합니다 .

그림 17.3 실제 Y 와 예측 Y 에 대한 3D 표면 그림

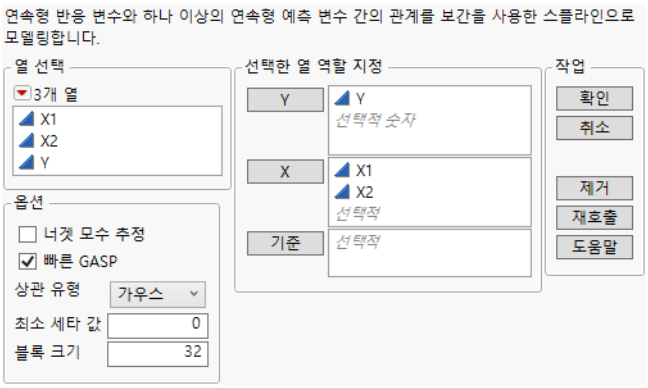


두 표면이 비슷합니다 . 반응 Y 에 대한 X1 및 X2 의 영향을 시각화할 수 있습니다 . 그림을 회전하여 다른 각도에서 볼 수 있습니다 . 반응에 대한 요인의 영향을 이해하는 데 사용할 수 있는 또 다른 도구에는 주변 그림이 있습니다 .

가우스 과정 플랫폼 시작

분석 > 전문 모델링 > 가우스 과정을 선택하여 가우스 과정 플랫폼을 시작합니다 .

그림 17.4 가우스 과정 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y 분석할 연속형 열을 할당합니다 .

X 설명 변수로 사용할 열을 할당합니다 ." 빠른 GASP" 옵션이 지정된 경우 JMP Pro 에서 범주형 변수를 사용할 수 있습니다 .

너겟 모수 추정 추정 절차에 능형 회귀 모수를 도입합니다 . 능형 회귀 모수는 반응에 잡음 또는 랜덤성이 있고 예측 모형으로 완벽한 보간 대신 잡음을 평활화하려는 경우에 유용합니다 .

JMP PRO **빠른 GASP** 빠른 GASP 알고리즘을 사용하기 위한 옵션입니다 . 빠른 GASP 는 계산 속도를 높이기 위해 가우스 과정 모형을 작은 조각 (블록) 으로 분할합니다 . 블록을 사용하면 여러 CPU 와 병렬 처리를 사용할 수 있습니다 .

참고 : 관측값이 2,500 개가 넘으면 빠른 GASP 알고리즘이 필요합니다 .

빠른 GASP 에 대한 자세한 내용은 Parker(2015) 연구 자료에서 확인하십시오 .

상관 유형 모형의 상관 구조를 선택합니다 . 이 플랫폼에서는 공간 상관 모형을 데이터에 적합시킵니다 . 여기서는 독립 변수의 값이 멀어질수록 두 관측값 사이의 반응 상관이 감소합니다 .

가우스 두 점 사이의 상관을 점 사이 거리에 관계없이 항상 0 이 아닌 것으로 제한합니다 .

입방 충분히 떨어져 있는 두 점 사이의 상관이 0 이 되도록 허용합니다 . 이 방법은 3 차 스플라인을 일반화한 것입니다 .

JMP PRO 빠른 GASP 알고리즘은 3 차 상관 함수를 지원하지 않습니다 .

최소 세타 값 적합 모형에 사용할 최소 세타 값을 설정합니다. 기본값은 0 입니다. 세타 값은 정규 회귀 모형의 기울기 모수와 유사합니다. 세타 값이 작으면 변수가 예측값에 거의 영향을 미치지 않음을 나타냅니다.

JMP PRO **블록 크기** 빠른 GASP 알고리즘에 사용되는 각 계산 블록의 관측값 수입니다. 블록당 25 개 이상의 관측값이 있어야 하고 데이터 집합의 최대 행 수는 2,500 개입니다.

가우스 과정 보고서

- "실제값 대 예측값 그림"
- "모형 보고서"
- "주변 모형 그림"

실제값 대 예측값 그림

"실제값 대 예측값 그림"에서는 Y 축의 실제 Y 값과 X 축의 잭나이프 예측값을 보여 줍니다. 적합도의 한 가지 측도는 점 배치가 그림의 대각선 (Y = X) 을 얼마나 잘 따르는지 살펴보는 것입니다.

잭나이프 값은 제외된 각 Y에 대해 모형이 관련 행으로 재적합되지 않는다는 점에서 정확한 잭나이프 값이 아닙니다. 오히려 관련된 각 Y에 대한 예측 모형에서 행이 제외되지만 상관 모수는 행의 기여도를 유지합니다. 데이터를 완벽하게 보간하는 가우스 과정의 경우 이 잭나이프 절차는 입력과 동일하지 않은 예측을 제공합니다.

모형 보고서

"모형 보고서" 에서는 모형 모수 추정값에 대한 함수 ANOVA 테이블을 보여 줍니다. 구체적으로 말하면 함수 기반 방법을 사용하여 변동을 계산하는 분산 분석 테이블입니다.

세타 가우스 과정 모형 모수 추정값입니다. 자세한 내용은 "가우스 과정 플랫폼에 대한 통계 상세 정보" 에서 확인하십시오.

총 민감도 각 요인에 대한 주효과 및 모든 교호작용 항의 합입니다. 이 값은 한 요인과 해당 요인의 모든 이원 교호작용이 반응 변수에 미치는 영향을 나타내는 측도입니다.

총 변동은 전체 실험 공간에 대한 통합 변동입니다.

주효과 각 요인의 함수 주효과는 해당 요인 단독에 의한 통합 총 변동입니다. 주효과는 모형의 각 요인에 대한 함수 효과와 총 변동의 비율입니다.

교호작용 함수 교호작용 효과는 주효과와 비슷한 방식으로 계산됩니다.

JMP PRO **범주형 입력** 모형에 범주형 요인이 포함된 경우 각 범주형 요인에 대한 상관 행렬이 제공됩니다. 비대각 항목은 가우스 과정 모형 모수 추정값에 해당합니다. 자세한 내용은 "범주형 예측 변수가 있는 모형" 에서 확인하십시오.

Mu 와 시그마 ^{2} 평균 및 분산 모형 모수입니다 .

너겟 추정된 너겟 값입니다 . 가우스 과정 시작 창에서 **너겟 모수 추정**을 선택한 경우 너겟 값이 보고됩니다 . JMP 에서 특이 공분산 행렬을 방지하기 위해 너겟 모수를 추가한 경우에도 너겟 값이 보고됩니다 .

-2* 로그 가능도 음의 로그 가능도 함수의 두 배에 대한 추정값입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

주변 모형 그림

모형의 각 요인에 대한 주변 그림이 나타납니다. 각 그림마다 하나를 제외한 모든 요인은 데이터에서 다른 요인의 범위에 대해 균등한 분포를 사용하여 통합됩니다. 나머지 관심 요인에 대한 주변 반응 예측은 파란색으로 표시됩니다. 주변 모형이 데이터를 얼마나 잘 적합시키는지 나타내기 위해 관측된 반응 값에 대한 데이터 점이 포함됩니다. 요인 수가 적은 모형의 경우 이러한 그림을 진단 그림으로 사용할 수 있습니다.

가우스 과정 플랫폼 옵션

" 가우스 과정 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

프로파일러 표준 프로파일러를 엽니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

등고선 프로파일러 등고선 프로파일러를 엽니다. 자세한 내용은 Profilers 의 에서 확인하십시오.

표면 프로파일러 표면 프로파일러를 엽니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

예측 계산식 저장 활성 데이터 테이블에 새 예측 계산식 열을 생성합니다 .

분산 계산식 저장 활성 데이터 테이블에 새 분산 계산식 열을 생성합니다 .

JMP PRO **예측 계산식 게시** 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

JMP PRO **분산 계산식 게시** 분산 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 " 계산식 저장소 " 에서 확인하십시오 .

잭나이프 예측값 저장 잭나이프 예측값을 활성 데이터 테이블에 저장합니다 . 이 값은 실제값 대 예측값 그림의 X 축 값입니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

가우스 과정 플랫폼의 추가 예

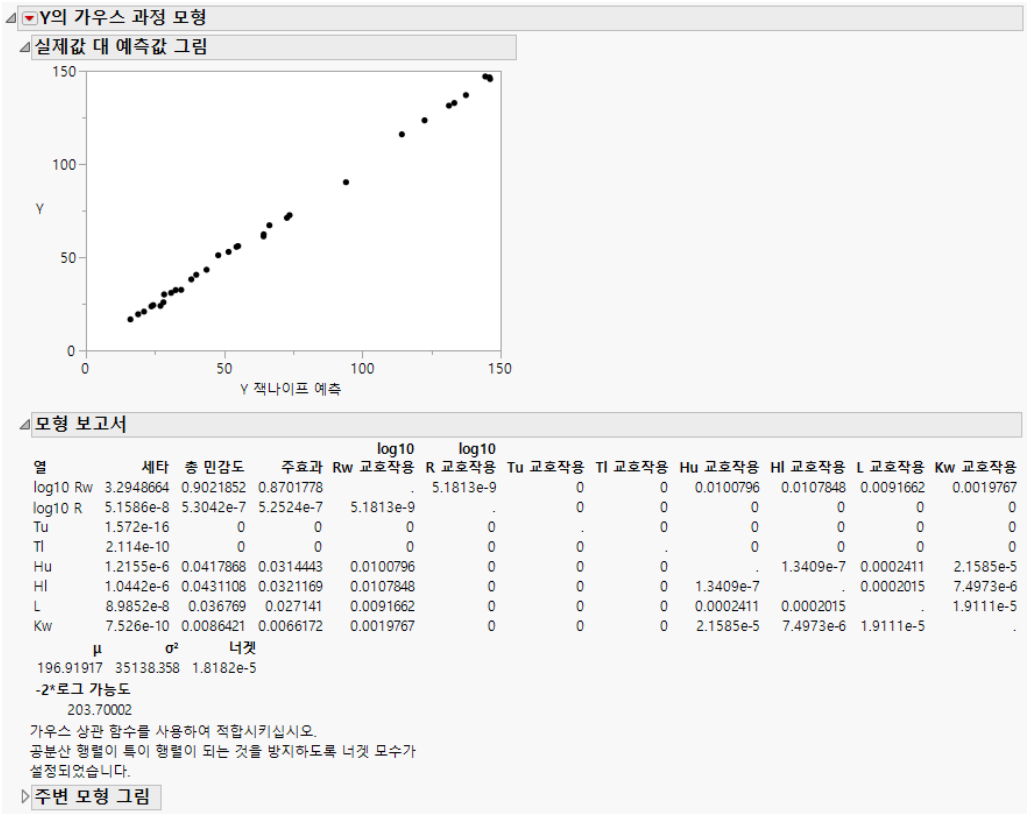
- " 가우스 과정 모형의 예 "
- " 범주형 예측 변수가 있는 가우스 과정 모형의 예 "

가우스 과정 모형의 예

이 예에서는 지표면에서 두 개의 대수층을 통과하여 뚫은 보어홀을 통한 물의 흐름을 보여 주는 데이터를 사용합니다 . 지정된 공학 모형이 주어지면 가우스 과정을 통해 모형에 포함된 요인이 반응 Y 에 미치는 영향을 이해할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Borehole Latin Hypercube.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 가우스 과정**을 선택합니다 .
3. log10 Rw ~ Kw 를 선택하고 **X** 를 클릭합니다 .
4. Y 를 선택하고 **Y** 를 클릭합니다 .
5. JMP PRO JMP Pro 에서 분석을 더 빠르게 실행하려면 " 빠른 GASP" 를 선택된 상태로 둡니다 .
6. **확인**을 클릭합니다 .

그림 17.5 Borehole Latin Hypercube 보고서



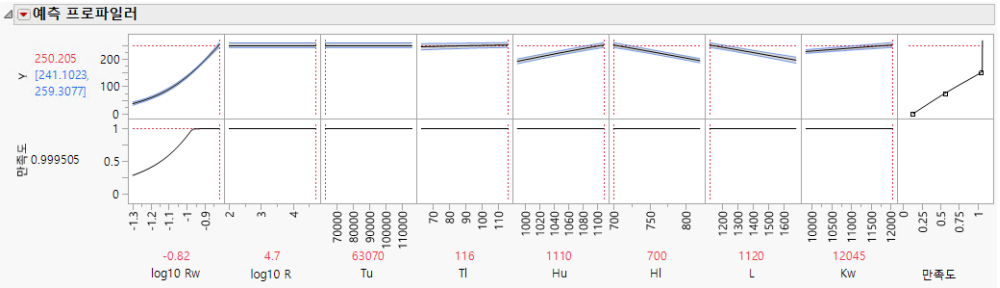
실제값 대 예측값 그림의 데이터가 Y = X 선을 따라 분포되어 있으므로 가우스 과정 예측 모형이 실제 함수와 유사한 근사값을 산출한다는 것을 나타냅니다 . " 모형 보고서 " 에서 첫 번째 요인 log10 Rw 의 총 민감도가 가장 높음을 확인할 수 있습니다 . Log10 Rw 의 추정된 총 민감도는 반응의 변동을 90% 넘게 설명합니다 . 세타 값이 작은 요인은 예측 계산식에 거의 또는 전혀 영향을 미치지 않습니다. 프로파일러를 사용하여 민감도를 시각화할 수 있습니다.

7. "Y 의 가우스 과정 모형 " 의 빨간색 삼각형을 클릭하고 **프로파일러**를 선택합니다 .

   반응 Y 에 " 반응 한계 " 열 특성이 있으므로 만족도 함수가 자동으로 나타납니다 .

8. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다 .

그림 17.6 가우스 과정 모형 프로파일러



Y 의 " 반응 한계 " 열 특성에 목표가 " 최대화 " 로 설정되어 있으므로 만족도 함수는 반응 Y 를 최대화하도록 설정됩니다 . 만족도 함수를 최대화하면 반응을 최대화하는 요인 값이 식별됩니다 . 음영이 적용된 대역은 95% 신뢰 구간을 나타냅니다 .

참고 : 표시된 추정값은 빠른 GASP 알고리즘을 사용하여 계산된 그림 17.5 의 추정값과 다를 수 있습니다 .

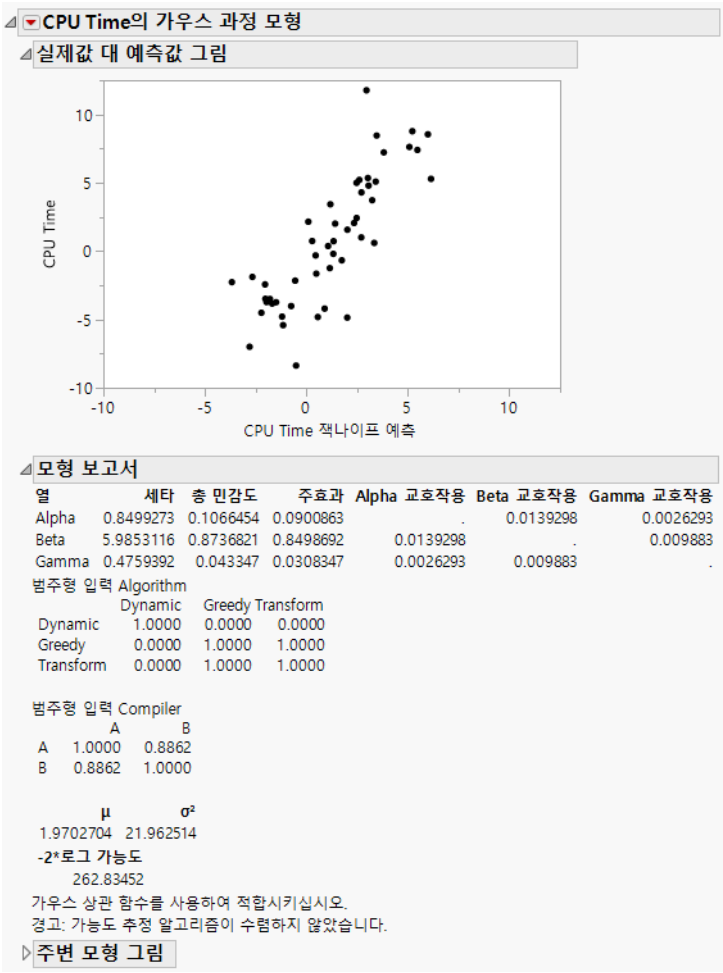
JMP PRO 범주형 예측 변수가 있는 가우스 과정 모형의 예

이 예에서는 Algorithm Data.jmp 샘플 데이터 테이블을 사용합니다 . 이 데이터는 공간 채움 설계 실험을 50 회 실행하여 얻은 시뮬레이션된 CPU 시간입니다 . Algorithm Factors.jmp 샘플 데이터 테이블에는 설계에 대한 요인 및 설정이 제공됩니다. 설계에는 세 개의 연속형 요인과 두 개의 범주형 요인이 있습니다 . 연속형 요인과 범주형 요인이 모두 포함된 가우스 과정 모형을 사용하여 CPU Time 을 예측하는 것이 목표입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Algorithm Data.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 가우스 과정**을 선택합니다 .
3. Alpha ~ Compiler 를 선택하고 **X** 를 클릭합니다 .
4. CPU Time 을 선택하고 **Y** 를 클릭합니다 .
5. JMP PRO 분석을 실행하려면 " 빠른 GASP" 를 선택된 상태로 둡니다 . **확인**을 클릭합니다 .

참고 : 범주형 요인이 포함된 모형에는 " 빠른 GASP" 옵션을 사용해야 합니다 . 자세한 내용은 " 범주형 예측 변수가 있는 모형 " 에서 확인하십시오 .

그림 17.7 Algorithm Data 보고서



" 실제값 대 예측값 그림 " 에서는 실제 CPU 시간과 예측 CPU 시간 사이의 강한 상관관계를 보여 줍니다 . 이는 가우스 과정 예측 모형이 실제 함수와 유사한 근사값을 산출한다는 것을 나타냅니다 . " 모형 보고서 " 에서 Beta 예측 변수의 총 민감도가 가장 높습니다 . 이는 연속형 예측 변수 중 Beta 가 반응의 변동을 가장 많이 설명한다는 것을 나타냅니다 . 각 범주형 예측 변수 , 즉 Algorithm 과 Compiler 에 대한 별도의 " 범주형 입력 " 행렬이 있습니다 . 이러한 행렬은 상관 행렬이며 각 범주형 예측 변수에 대한 수준 사이의 상관관계를 나타냅니다 . 행렬의 비대각은 τ 모수입니다 .

가우스 과정 플랫폼에 대한 통계 상세 정보

- "연속형 예측 변수가 있는 모형"
- "범주형 예측 변수가 있는 모형"
- "분산 계산식 파라미터화"
- "모형 적합 상세 정보"

연속형 예측 변수가 있는 모형

가우스 과정 모형에 연속형 예측 변수만 포함된 경우 가우스 과정 플랫폼에서는 "가우스"와 "입방"이라는 두 개의 상관 구조를 구현합니다.

가우스 상관 구조는 멱이 2 인 곱 지수 상관 함수를 추정 모형으로 사용합니다. 이 모형은 Y 가 평균 μ 와 공분산 행렬 $\sigma^2\mathbf{R}$ 의 정규 분포를 따른다고 가정합니다. $\mathbf{R}$ 행렬의 요소는 다음과 같이 정의됩니다.

$$r_{ij} = \exp\left(-\sum_{k=1}^{K} \theta_k (x_{ik} - x_{jk})^2\right)$$

다음은 각 요소에 대한 설명입니다.

K = 연속형 예측 변수의 수

θ_k = k 번째 예측 변수의 세타 모수

x_{ik} = 개체 i 에 대한 k 번째 예측 변수의 값

x_{jk} = 개체 j 에 대한 k 번째 예측 변수의 값

입방 상관 구조도 Y 가 평균 μ 와 공분산 행렬 $\sigma^2\mathbf{R}$ 의 정규 분포를 따른다고 가정합니다. $\mathbf{R}$ 행렬은 다음 요소로 구성됩니다.

$$r_{ij} = \prod_k \rho(d;\theta_k)$$

여기서,

$$d = x_{ik} - x_{jk}$$

$$\rho(d;\theta) = \begin{cases} 1 - 6(d\theta)^2 + 6(|d|\theta)^3, & |d| \le \frac{1}{2\theta} \\ 2(1 - |d|\theta)^3, & \frac{1}{2\theta} < |d| \le \frac{1}{\theta} \\ 0, & \frac{1}{\theta} < |d| \end{cases}$$

자세한 내용은 Santer(2003) 연구 자료에서 확인하십시오. 입방 상관 구조에 사용되는 세타 모수는 문헌에 자주 사용되는 모수의 역수입니다. 세타가 모형에 영향을 미치지 않을 때 역수가 사용되어 rho의 값이 무한대가 아니라 0이 됩니다.

JMP PRO 범주형 예측 변수가 있는 모형

가우스 과정 모형에 범주형 예측 변수가 포함된 경우 가우스 상관 구조가 사용됩니다. **R** 행렬의 요소는 다음과 같이 정의됩니다.

$$r_{ij} = \left(\prod_{p=1}^{P} \tau_{p_{ij}} \right) \exp\left(- \sum_{k=1}^{K} \theta_k (x_{ik} - x_{jk})^2 \right)$$

다음은 각 요소에 대한 설명입니다.

K = 연속형 예측 변수의 수

P = 범주형 예측 변수의 수

θ_k = k 번째 연속형 예측 변수의 세타 모수

x_{ik} = 개체 i 에 대한 k 번째 연속형 예측 변수의 값

x_{jk} = 개체 j 에 대한 k 번째 연속형 예측 변수의 값

$\tau_{p_{ij}}$ = 개체 i 에 대한 예측 변수 p 의 관측된 수준과 개체 j 에 대한 예측 변수 p 의 관측된 수준 사이의 상관

범주형 변수의 각 수준 조합에 대한 τ 모수가 있습니다. 여기서 τ_{ij}는 개체 i와 개체 j의 관측된 수준에 의해 구성된 고유 조합에 해당합니다. 따라서 공분산 요소 r_{ij}는 i번째 관측값과 j번째 관측값에서 얻은 범주형 예측 변수의 수준 조합에 따라 달라집니다. 자세한 내용은 Qian et al. (2012) 연구 자료에서 확인하십시오.

분산 계산식 파라미터화

저장된 분산 계산식은 모형에 범주형 예측 변수가 포함된 경우를 제외하고 이전에 정의된 **R** 파라미터화를 사용합니다. 가우스 과정 모형에 범주형 예측 변수가 포함된 경우에는 저장된 분산 계산식에 다음과 같은 **R** 파라미터화가 사용됩니다.

$$r_{ij} = \exp\left(-\sum_{k=1}^{K} \theta_k (x_{ik} - x_{jk})^2 - \sum_{p=1}^{P} \phi_{p_{ij}}\right)$$

여기서 $\phi_{p_{ij}} = -\ln(\tau_{p_{ij}})$ 및 다른 모든 변수는 이전에 정의된 대로입니다.

모형 적합 상세 정보

모형 모수는 최대 가능도를 통해 적합됩니다. 적합된 모수는 플랫폼 보고서에 제공됩니다. 모수는 다음과 같습니다.

- μ = 가우스 과정 평균
- σ^2 = 가우스 과정 분산
- 세타는 **R** 정의에서 θ_k 값에 해당합니다.
- 범주형 입력 상관 행렬의 비대각은 **R** 정의에서 τ_{pij} 값에 해당합니다.

참고 : JMP 에서 분산 행렬을 반전할 수 있도록 능형 회귀 모수를 추가한 경우 보고서에 **공분산 행렬이 특이 행렬이 되는 것을 방지하도록 너겟 모수가 설정되었습니다.** 라는 노트가 포함됩니다.

18 장

시계열 분석

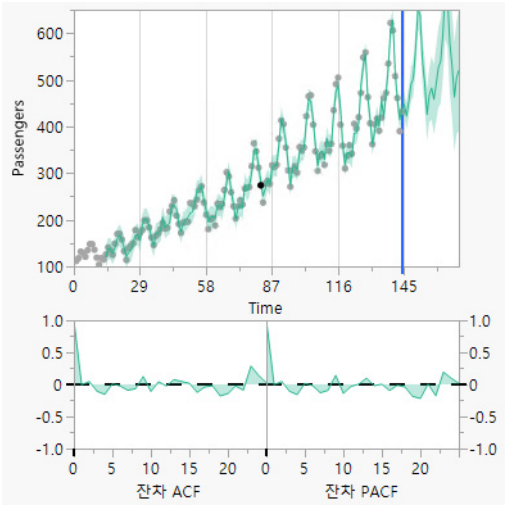
시계열 모형 및 전이 함수 적합

시계열 플랫폼을 사용하면 단변량 시계열을 탐색, 분석 및 예측할 수 있습니다. 시계열은 일정 간격의 일련의 기간 동안 얻은 관측값 집합입니다. 일반적으로 시간상 서로 가까운 관측값은 상관되어 있습니다. 시계열 방법론은 관측값 간의 이러한 종속성을 활용하여 미래에 시계열이 어떤 형태가 될지 더 잘 예측할 수 있습니다.

시계열 데이터의 공통 특성에는 계절성, 추세 및 자기상관이 포함됩니다. 시계열 플랫폼에서는 이러한 특성을 처리할 수 있는 옵션을 제공합니다. 변동도, 자기상관 그림, 편자기상관 그림 및 스펙트럼 밀도 그림과 같은 그래프를 사용하여 시계열을 설명하고 예측하는 데 적절한 모형 유형을 식별할 수 있습니다. 또한 이 플랫폼에는 데이터의 계절 또는 일반적 추세를 제거하여 분석을 단순화할 수 있는 몇 가지 분해 방법이 있습니다. 아니면 계절성과 장기 추세를 하나의 모형에 모두 포함할 수 있는 더 정교한 ARIMA 모형 및 상태 공간 평활 모형을 적합시킬 수 있습니다. Box-Cox 변환을 수행하고 변환된 계열을 분석 및 모델링할 수도 있습니다.

모형의 예측 성능을 평가하는 몇 가지 방법이 있습니다. "홀드백에 대한 예측" 기능은 시계열을 모형 생성을 위한 훈련 부분과 예측 성능 평가를 위한 홀드백 부분으로 분할합니다.

그림 18.1 예측 그림



목차

시계열 플랫폼 개요 301
시계열 플랫폼의 예 301
시계열 플랫폼 시작 304
시계열 분석 보고서 305
  시계열 그래프 306
  시계열 기본 진단 차트 306
시계열 플랫폼 옵션 308
  시계열 진단 308
  차분 및 분해 309
  ARIMA 및 계절 ARIMA 모형 311
  평활 모형 313
  상태 공간 평활 모형 314
  전이 함수 모형 314
  평활 모형 규격 창 317
보고서 319
  차분 보고서 319
  분해 보고서 320
  모형 비교 보고서 322
  ARIMA 및 계절 ARIMA 모형 보고서 324
  상태 공간 평활 보고서 328
  전이 함수 보고서 330
  스펙트럼 밀도 보고서 331
시계열 플랫폼의 추가 예 332
  시간 ID 열 생성의 예 332
  Box-Cox 변환을 사용한 예 337
  홀드백 집합을 사용한 예 339
시계열 플랫폼에 대한 통계 상세 정보 341
  스펙트럼 밀도에 대한 통계 상세 정보 341
  X-11 분해에 대한 통계 상세 정보 342
  지수 평활 모형에 대한 통계 상세 정보 342
  ARIMA 모형에 대한 통계 상세 정보 347
  전이 함수에 대한 통계 상세 정보 348

시계열 플랫폼 개요

시계열은 일정 간격의 일련의 기간 동안 얻은 관측값 집합입니다($y_1, y_2, \ldots, y_N$). 예를 들어 분기별 판매 보고서, 월별 평균 온도, 흑점 개수 등이 시계열 데이터에 포함됩니다. 시계열 플랫폼을 사용하면 이러한 유형의 데이터에서 발견되는 패턴과 추세를 탐색할 수 있습니다. 그런 다음 이러한 패턴과 추세를 사용하여 미래의 결과를 예측할 수 있습니다.

시계열 데이터의 공통 특성에는 계절성, 추세 및 자기상관이 포함됩니다. **계절성**은 알려진 기간 동안 나타나는 패턴을 말합니다. 예를 들어 월별로 수집되는 데이터는 모든 데이터 수집 연도의 여름철에 유사한 형태를 보일 수 있습니다. **추세**는 시간에 따른 값의 점진적 증가 또는 감소와 같이 계열의 장기적인 움직임을 말합니다. **자기상관**은 계열의 각 점이 계열의 이전 값과 상관되는 정도를 나타냅니다.

시계열 플랫폼에는 다양한 모형과 예측 방법이 있습니다. 그러나 일부 방법은 추세 또는 계절성을 처리할 수 없습니다. 적절한 모형을 선택하려면 계열에 어떤 특성이 있는지 확인하는 것이 중요합니다. 시계열 플랫폼에서는 시계열의 진화를 설명하고 예측하는 데 적절한 모형 유형을 식별할 수 있는 변동도, 자기상관 그림, 편자기상관 그림 및 스펙트럼 밀도 그림과 같은 그래프를 제공합니다. 또한 이 플랫폼에는 데이터의 계절 또는 일반적 추세를 제거하여 분석을 탐색하고 단순화할 수 있는 몇 가지 차분 및 분해 방법이 있습니다. Box-Cox 변환을 보고 데이터에 적용할 수도 있습니다.

아니면 계절성과 장기 추세를 포함할 수 있는 더 정교한 모형을 적합시킬 수 있습니다. 플랫폼에서 이 기능을 가진 모형 중 하나는 고급 지수 평활 모형인 Winter 의 가법 방법입니다. 또한 ARIMA(AutoRegressive Integrated Moving Average) 모형과 상태 공간 평활 모형을 적합시킬 수 있습니다. 이 두 가지 유형의 모형은 통계적으로 가장 복잡하지만 유연성도 가장 높습니다. 고급 지수 평활, ARIMA 및 상태 공간 평활 모형은 해석하기 어렵지만 예측에 탁월한 도구입니다.

시계열 플랫폼에서는 입력 계열과 함께 제공될 경우 전이 함수 모형도 적합시킬 수 있습니다.

시계열 플랫폼의 예

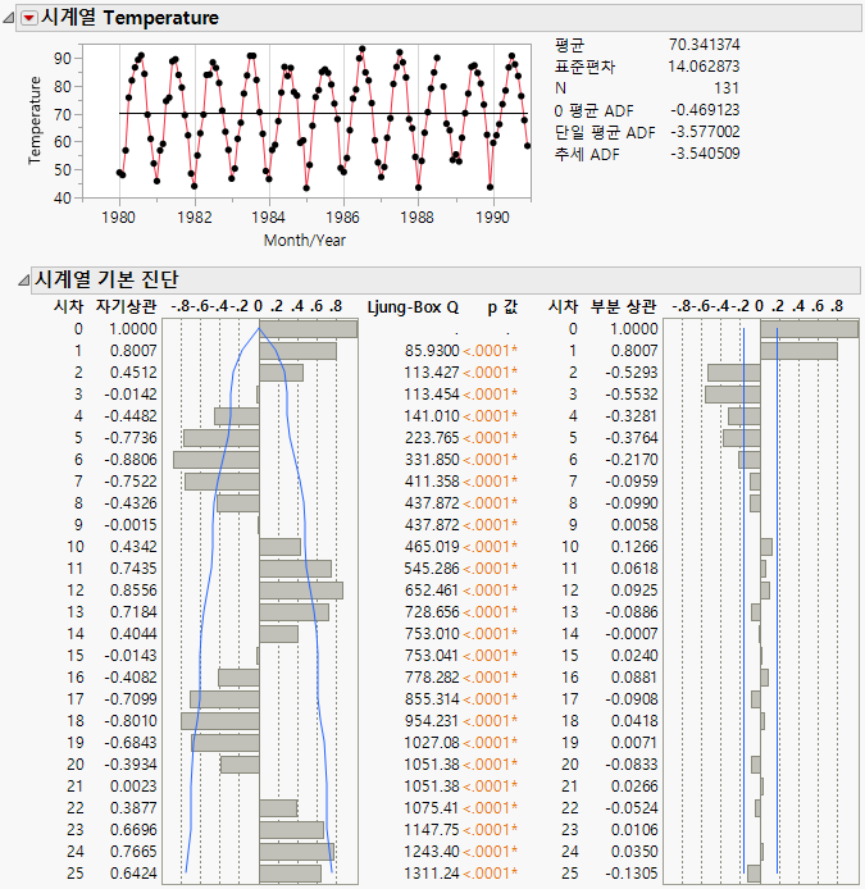
이 예에서는 1980 년에서 1990 년까지 측정한 최대 월별 온도 (화씨) 가 포함된 Raleigh Temps.jmp 샘플 데이터 테이블을 사용합니다. 시계열 플랫폼을 사용하여 계열을 검토하고 향후 2 년간의 최대 월별 온도를 예측할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Time Series/Raleigh Temps.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 시계열**을 선택합니다 .
3. Temperature 를 선택하고 **Y, 시계열**을 클릭합니다 .
4. Month/Year 를 선택하고 **X, 시간 ID** 를 클릭합니다 .
5. **예측 주기** 옆의 상자에 "24" 를 입력합니다 .

   이 값은 데이터에 적합된 모형을 사용하여 예측되는 미래 기간의 수입니다 . 다음 2 년 , 즉 24 개월 동안의 월별 온도를 예측하려고 합니다 .

6. **확인**을 클릭합니다 .

그림 18.2 Raleigh Temps.jmp 에 대한 시계열 분석 보고서



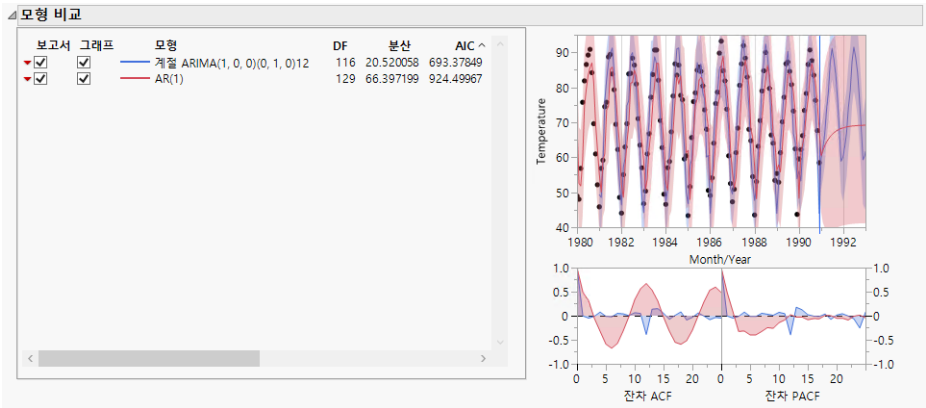
시차	자기상관	Ljung-Box Q	p 값	시차	부분 상관
0	1.0000	.	.	0	1.0000
1	0.8007	85.9300	<.0001*	1	0.8007
2	0.4512	113.427	<.0001*	2	-0.5293
3	-0.0142	113.454	<.0001*	3	-0.5532
4	-0.4482	141.010	<.0001*	4	-0.3281
5	-0.7736	223.765	<.0001*	5	-0.3764
6	-0.8806	331.850	<.0001*	6	-0.2170
7	-0.7522	411.358	<.0001*	7	-0.0959
8	-0.4326	437.872	<.0001*	8	-0.0990
9	-0.0015	437.872	<.0001*	9	0.0058
10	0.4342	465.019	<.0001*	10	0.1266
11	0.7435	545.286	<.0001*	11	0.0618
12	0.8556	652.461	<.0001*	12	0.0925
13	0.7184	728.656	<.0001*	13	-0.0886
14	0.4044	753.010	<.0001*	14	-0.0007
15	-0.0143	753.041	<.0001*	15	0.0240
16	-0.4082	778.282	<.0001*	16	0.0881
17	-0.7099	855.314	<.0001*	17	-0.0908
18	-0.8010	954.231	<.0001*	18	0.0418
19	-0.6843	1027.08	<.0001*	19	0.0071
20	-0.3934	1051.38	<.0001*	20	-0.0833
21	0.0023	1051.38	<.0001*	21	0.0266
22	0.3877	1075.41	<.0001*	22	-0.0524
23	0.6696	1147.75	<.0001*	23	0.0106
24	0.7665	1243.40	<.0001*	24	0.0350
25	0.6424	1311.24	<.0001*	25	-0.1305

시계열 그래프에서는 계열이 순환적임을 보여 줍니다 . 이 순환 성분은 자기상관 차트에서도 분명하게 확인할 수 있습니다 . 1 시차의 점은 양의 상관관계가 있습니다 (" 자기상관 " 값 = 0.8007). 점이 멀어질수록 음의 상관관계가 된 후 다시 양의 상관관계를 보이고 패턴이 반복됩니다 . 시계열 그래프와 자기상관 차트는 시계열에 계절성 특성이 있다는 증거를 제공합니다 .

7. " 시계열 " 의 빨간색 삼각형을 클릭하고 **ARIMA** 를 선택합니다 .
8. 계열에 자기상관 증거가 보였으므로 "*p*, 자기회귀 차수 " 를 "1" 로 설정합니다 .
9. **추정**을 클릭합니다 .
10. " 시계열 " 의 빨간색 삼각형을 클릭하고 **계절 ARIMA** 를 선택합니다 .
11. 계열에 자기상관 증거가 보였으므로 "ARIMA" 상자에서 "*p*, 자기회귀 차수 " 를 "1" 로 설정합니다 .

12. 계열에 계절성 증거가 보였으므로 " 계절 ARIMA" 상자에서 "*D*, 차분 차수 " 를 "1" 로 설정합니다 .
13. **추정**을 클릭합니다 .
14. 모형 비교 테이블에서 두 모형에 대한 " 그래프 " 아래의 상자를 선택합니다 .

그림 18.3 Raleigh Temps.jmp 에 대한 모형 비교 테이블

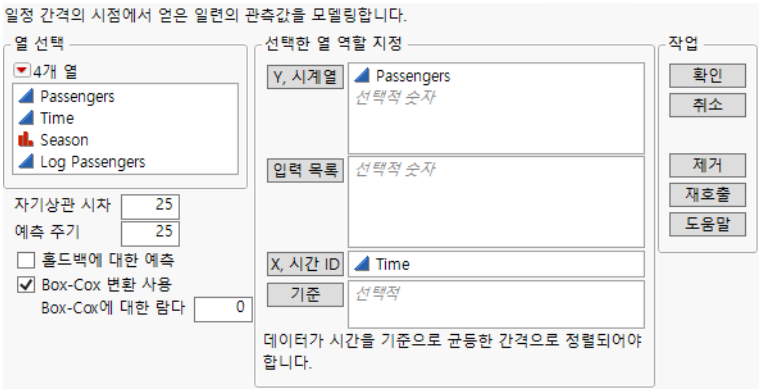


모형 비교 테이블은 AIC 통계량에 따라 내림차순으로 정렬됩니다 . 즉 , 최량 적합 모형이 보고서 맨 위에 나타납니다 . 계절 ARIMA 모형의 AIC 값 (693.4) 이 정규 ARIMA 모형의 AIC 값 (924.5) 보다 훨씬 작습니다 . 그래프에서는 ARIMA 모형이 관측된 점을 비교적 잘 예측하지만 잔차가 계절 ARIMA 모형의 잔차보다 크다는 것을 보여 줍니다 . 또한 계절 ARIMA 모형의 예측 구간이 더 좁으므로 미래 관측값에 대해 보다 현실적인 예측을 제공합니다 . 계열이 계절 성분에 대한 증거를 보여 주었으므로 이러한 결과가 타당합니다 .

시계열 플랫폼 시작

분석 > 전문 모델링 > 시계열을 선택하여 시계열 플랫폼을 시작합니다 . 그림 18.4 에서는 Seriesg.jmp 샘플 데이터 테이블에 대한 시계열 시작 창을 보여 줍니다 .

그림 18.4 시계열 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

시계열 플랫폼 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y, 시계열 하나 이상의 열을 시계열 변수로 할당합니다 . Y 축에 표시됩니다 .

입력 목록 하나 이상의 열을 입력 계열 변수로 할당합니다 . " 입력 시계열 패널 " 에 표시되고 전이 함수 모형에 사용됩니다 . 입력 계열 변수는 시계열 또는 표시자로 숫자여야 합니다 .

X, 시간 ID 시간 축 (X 축) 에 라벨을 지정하기 위한 하나의 변수를 할당합니다 . " 시간 ID" 에 변수를 지정하지 않으면 행 번호가 대신 사용됩니다 .

참고 : "X, 시간 ID" 변수를 사용하는 경우 " 시간 주기 " 열 특성을 사용하여 시간 주기를 지정할 수 있습니다 . " 매년 ", " 매분기 ", " 매월 ", " 매주 ", " 매일 ", " 매시간 ", " 매분 " 및 " 매초 " 중에서 선택할 수 있습니다 . 이렇게 하면 JMP 에서 예측 값을 그림에 표시할 때 데이터 간격을 결정하는 데 도움이 됩니다 . 주기를 지정하지 않으면 데이터가 일정한 간격의 수치형 데이터로 간주됩니다 .

주의 : X, 시간 ID 에 할당된 변수의 관측값은 간격이 일정하다고 가정합니다 . 그러나 시계열 플랫폼에서는 타임스탬프가 증가하는지 여부만 확인합니다 . 관측값의 간격이 일정한지 여부는 확인하지 않습니다 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 할당합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

참고 : " 기준 " 변수를 사용하는 경우 " 기준 " 변수의 각 수준에 대한 관측값 수에 따라 자기상관 시차 수를 변경해야 할 수도 있습니다 . 시차 수는 1 보다 크고 수준당 관측값 수보다 작아야 합니다 .

자기상관 시차 자기상관 및 부분 상관을 계산하는 데 사용할 시차 수를 지정합니다 . 이 값은 이러한 상관을 계산하는 데 사용된 점 사이의 최대 주기 수입니다 . 1 보다 크고 행 수보다 작아야 합니다 . 기본 시차 수는 25 입니다 .

팁 : 최대 시차 수에 일반적으로 사용되는 규칙은 $n/4$ 이며 , 여기서 n 은 관측값 수입니다 .

예측 주기 데이터에 적합된 각 모형을 사용하여 예측되는 관측값 수를 지정합니다 . 기본 예측 수는 25 입니다 .

- " 홀드백에 대한 예측 " 옵션을 선택하는 경우 " 예측 주기 " 값은 모형 적합에 포함되지 않은 계열 끝의 관측값 수를 지정합니다 . 그러면 적합 모형을 사용하여 이러한 값을 예측합니다 .
- " 홀드백에 대한 예측 " 옵션을 선택하지 않는 경우 " 예측 주기 " 값은 계열이 끝난 후 예측되는 미래 관측값 수를 지정합니다 .

홀드백에 대한 예측 예측이 미래 관측값에 대해 수행되는지 아니면 홀드백 관측값에 대해 수행되는지 결정합니다 . " 홀드백에 대한 예측 " 옵션을 선택하면 " 예측 주기 " 옵션에 지정된 수에 따라 결정된 홀드백 집합에서 예측이 수행됩니다 .

Box-Cox 변환 사용 "Box-Cox 에 대한 람다 " 옵션에 지정된 람다로 Box-Cox 변환을 사용하여 원래 데이터를 변환합니다 . 이 옵션을 선택하면 변환된 데이터에서 시계열 보고서의 모든 분석이 수행됩니다 . 그러나 모든 예측은 다시 변환되어 원래 척도로 보고됩니다 .

Box-Cox 에 대한 람다 ("Box-Cox 변환 사용 " 을 선택한 경우에만 사용 가능) 원래 데이터의 Box-Cox 변환에 사용되는 람다 모수를 지정합니다 .

참고 : 시계열 플랫폼은 시계열의 순서와 간격을 유지하기 위해 데이터 테이블의 제외된 행을 결측값으로 처리합니다 .

시계열 분석 보고서

초기 시계열 보고서에는 시계열 그래프 , 시계열 변수에 대한 요약 통계량 및 검정 , 기본 진단 보고서가 표시됩니다 . 시작 창에서 " 입력 목록 " 에 열을 지정한 경우 " 전이 함수 분석 " 및 " 입력 시계열 패널 " 보고서가 표시됩니다 . " 전이 함수 분석 " 보고서와 " 입력 시계열 패널 " 보고서에는 시계열 보고서와 동일한 초기 정보가 포함됩니다 .

- " 시계열 그래프 "
- " 시계열 기본 진단 차트 "

시계열 그래프

시계열 그래프에는 각 시계열 대 시간 ID 가 표시됩니다 . 시간 ID 를 지정하지 않으면 행 번호가 대신 사용됩니다 . 홀드백 집합이 지정된 경우 훈련 데이터와 홀드백 데이터를 구분하는 세로 참조선이 그래프에 있습니다 .

이 플랫폼에서는 ADF(Augmented Dickey-Fuller) 검정을 사용하여 몇 가지 안정성 검정도 수행합니다. 다음 검정 및 요약 통계량이 시계열 그래프 옆에 표시됩니다.

Box-Cox 에 대한 람다 (Box-Cox 변환을 지정한 경우에만 표시됨) Box-Cox 변환에 사용되는 람다 값입니다 .

평균 표본 평균입니다 .

표준편차 표본 표준편차입니다 .

N 계열 길이입니다 .

0 평균 ADF 0 평균을 사용하여 확률 보행과 대조하는 검정으로 , 다음과 같이 정의됩니다 .

$$x_t = \phi x_{t-1} + e_t$$

단일 평균 ADF 0 이 아닌 평균을 사용하여 확률 보행과 대조하는 검정으로 , 다음과 같이 정의됩니다 .

$$x_t - \mu = \phi(x_{t-1} - \mu) + e_t$$

추세 ADF 0 이 아닌 평균과 선형 추세를 사용하여 확률 보행과 대조하는 검정으로 , 다음과 같이 정의됩니다 .

$$x_t - \mu - \beta t = \phi[x_{t-1} - \mu - \beta(t-1)] + e_t$$

시계열 기본 진단 차트

"시계열 기본 진단" 차트에 표시되는 정보는 시계열 보고서의 빨간색 삼각형 메뉴 옵션에 따라 다릅니다. 진단 차트에서 정보를 표시하거나 숨기는 빨간색 삼각형 메뉴 옵션은 "자기상관", "편자기상관", "변동도" 및 "AR 계수"입니다. 기본적으로 "자기상관" 및 "편자기상관"이 표시됩니다.

자기상관 차트

" 자기상관 " 옵션은 " 시계열 기본 진단 " 차트에서 다음 열을 표시하거나 숨깁니다 .

시차 점 사이의 주기 수입니다 .

> **참고 :** 시차 수는 기본적으로 0 에서 시작합니다 . 시차 1 로 시작하는 상관을 계산하려면 그래프를 생성하기 전에 JMP 환경 설정을 수정합니다. **파일 > 환경 설정 > 플랫폼 > 시계열**을 선택한 후 **ACF 및 PACF 에 시차 0 표시 안 함**을 선택하십시오 .

자기상관 k 번째 시차에 대한 자기상관이며 다음과 같이 계산됩니다 .

$$r_k = \frac{c_k}{c_0} \quad \text{여기서 ,} \quad c_k = \frac{1}{N} \sum_{t=k+1}^{N} (y_t - \bar{y})(y_{t-k} - \bar{y})$$

및 $\bar{y}$ 는 시계열에서 비결측 점 N 개의 평균입니다 . 정의에 따라 1 차 자기상관 (시차 0) 의 길이는 항상 1 입니다 .

막대는 자기상관을 그래픽으로 나타냅니다. 파란색 곡선은 다음과 같이 계산되는 큰 시차 표준 오차의 두 배 (±2 표준 오차) 를 나타냅니다 .

$$\mathrm{SE}_k = \sqrt{\frac{1}{N}\left(1 + 2\sum_{i=1}^{k-1} r_i^2\right)}$$

Ljung-Box Q 자기상관 그룹이 0 과 유의하게 다른지 여부를 검정하거나 모형의 잔차를 백색잡음과 구별할 수 있는지 검정하는 데 사용됩니다 . Q 는 검정 통계량입니다 .

p 값 Ljung-Box 검정의 p 값입니다 .

편자기상관 차트

" 편자기상관 " 옵션은 " 시계열 기본 진단 " 차트에서 다음 열을 표시하거나 숨깁니다 .

시차 점 사이의 주기 수입니다 .

편자기상관 k 번째 시차에 대한 편자기상관입니다 .

막대는 편자기상관을 그래픽으로 나타냅니다 . 파란색 선은 근사 95% 예측 한계에 대한 ±2 표준 오차를 나타내며 , 여기서 표준 오차는 다음과 같이 계산됩니다 .

$$\mathrm{SE}_k = \frac{1}{\sqrt{n}} \quad \text{, 모든 } k \text{ 에 대해}$$

변동도 차트

" 변동도 " 옵션은 " 시계열 기본 진단 " 차트에서 다음 열을 표시하거나 숨깁니다 .

시차 점 사이의 주기 수입니다 .

변동도 변동도는 k 시차가 있는 점의 차이 분산을 측정하여 1 시차가 있는 점의 차이 분산과 비교합니다 . 변동도는 다음과 같이 자기상관에서 계산됩니다 .

$$V_k = \frac{1 - r_{k+1}}{1 - r_1}$$

여기서 r_k 는 시차 k 에서의 자기상관입니다 .

AR 계수 차트

"AR 계수 " 옵션은 " 시계열 기본 진단 " 차트에서 다음 열을 표시하거나 숨깁니다 .

시차 점 사이의 주기 수입니다 .

AR 계수 계수는 고차의 순수 자기회귀 모형을 적합시켜 얻을 수 있는 것과 비슷합니다 .

시계열 플랫폼 옵션

- " 시계열 진단 "
- " 차분 및 분해 "
- "ARIMA 및 계절 ARIMA 모형 "
- " 평활 모형 "
- " 상태 공간 평활 모형 "
- " 전이 함수 모형 "
- " 평활 모형 규격 창 "

시계열 진단

그래프 시계열 그림 모양을 제어하는 옵션이 포함된 하위 메뉴를 표시합니다 .

시계열 그래프 시계열 그래프를 표시하거나 숨깁니다 .

점 표시 시계열 그래프에 점을 표시하거나 숨깁니다 .

연결선 시계열 그래프의 점을 연결하는 선을 표시하거나 숨깁니다 .

평균 선 시계열 그래프에서 시계열의 평균을 나타내는 가로선을 표시하거나 숨깁니다 .

자기상관 " 시계열 기본 진단 " 차트에서 자기상관 그림을 표시하거나 숨깁니다 . 자기상관 그래프는 주어진 시간 간격 (시차) 에 대해 시계열에 있는 모든 점 쌍 사이의 상관을 설명합니다 . 자세한 내용은 " 자기상관 차트 " 에서 확인하십시오 .

팁 : 표본의 자기상관 그래프를 **표본 자기상관 함수**라고도 합니다 .

편자기상관 " 시계열 기본 진단 " 차트에서 편자기상관 그림을 표시하거나 숨깁니다 . 편자기상관 그래프는 주어진 시간 간격 (시차) 에 대해 시계열에 있는 모든 점 쌍 사이의 부분 상관을 설명합니다 . 자세한 내용은 " 편자기상관 차트 " 에서 확인하십시오 .

팁 : 자기상관 및 편자기상관 그래프는 시계열이 정상 (시간 경과에 따라 고정 평균 및 표준 편차를 갖는다는 의미) 인지 여부를 확인하고 시계열 적합에 적절한 모형을 결정하는 데 도움이 될 수 있습니다 .

변동도 "시계열 기본 진단" 차트에서 변동도 그래프를 표시하거나 숨깁니다. 자세한 내용은 "변동도 차트"에서 확인하십시오.

AR 계수 "시계열 기본 진단" 차트에서 AR(자기회귀) 계수의 최소 제곱 추정값 그래프를 표시하거나 숨깁니다. 자세한 내용은 "AR 계수 차트"에서 확인하십시오.

스펙트럼 밀도 스펙트럼 밀도 그래프를 주기와 빈도의 함수로 표시하거나 숨깁니다. 스펙트럼 밀도 옵션은 데이터에 대한 두 가지 검정 결과를 제공하는 "백색잡음 검정" 보고서도 표시합니다. 자세한 내용은 "스펙트럼 밀도 보고서" 및 "스펙트럼 밀도에 대한 통계 상세 정보"에서 확인하십시오.

차분 및 분해

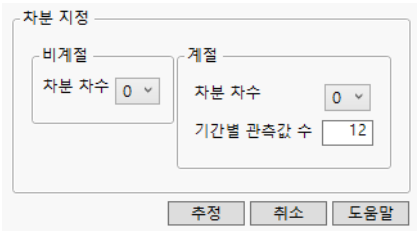
차분 "차분 규격" 창을 표시합니다 (그림 18.5). 이 창에서 시계열에 적용할 차분 연산을 지정할 수 있습니다. 시계열의 값을 차분하면 비정상 계열을 정상 계열로 변환할 수 있습니다. 차분 계열은 다음 방정식에 의해 제공됩니다.

$$w_t = (1 - \mathbf{B})^d (1 - \mathbf{B}^s)^D y_t$$

여기서 t 는 시간 인덱스이고, $\mathbf{B}$ 는 $\mathbf{B}y_t = y_{t-1}$ 로 정의된 후방 연산자입니다.

참고 : 많은 시계열은 추세 또는 계절성이 있는 시계열과 같은 고정 평균을 보이지 않습니다. 이러한 비정상 계열은 ARMA 모형과 같이 정상 시계열을 가정하는 시계열 모형에 의한 설명에 적합하지 않습니다. 추세 및 / 또는 계절성을 제거하면 정상 차분 계열이 생성되고 안정성을 가정한 모형을 사용하여 계열을 설명할 수 있습니다.

그림 18.5 차분 규격 창



"차분 규격" 창을 사용하여 비계절 차분 차수 (d), 계절 차분 차수 (D) 및 기간별 관측값 수 (s) 를 지정할 수 있습니다. 차분 차수의 값을 0 으로 선택하면 해당 유형의 차분이 없다는 것을 나타냅니다. 차분 연산을 지정하고 **추정**을 클릭할 때마다 보고서 창에 새 "차분" 보고서가 표시됩니다. 자세한 내용은 "시계열 플랫폼의 추가 예"에서 확인하십시오.

분해 (시작 창에서 "홀드백에 대한 예측" 또는 "Box-Cox 변환 사용" 을 선택한 경우에는 사용 불가능) 분해 방법 하위 메뉴를 표시합니다. 시계열 데이터 분해는 시계열에서 선형 추세와 계절 주기를 분리하고 제거합니다. 이렇게 하면 모형 추정에 도움이 될 수 있습니다. 세 가지 분해 옵션이 제공됩니다.

선형 추세 제거 선형 회귀 모형을 사용하여 시계열의 선형 추세를 추정하고 데이터에서 선형 추세를 제거합니다. 추세제거 계열에 대한 시계열 보고서가 선형 추세 정보와 함께 보고서 창에 추가됩니다. 자세한 내용은 " 시계열 분석 보고서 " 및 " 선형 추세 보고서 " 에서 확인하십시오.

순환 제거 단일 코사인파를 사용하여 시계열의 순환 성분을 추정한 후 데이터에서 순환 성분을 제거합니다. " 순환 제거 " 옵션을 선택하면 " 순환 정의 " 대화상자가 나타납니다. 이 대화상자 창에서 순환별 단위 수를 지정하고 데이터에서 상수를 뺄지 여부를 나타낼 수 있습니다. 순환제거 계열에 대한 시계열 보고서가 순환 정보와 함께 보고서 창에 추가됩니다. 자세한 내용은 " 시계열 분석 보고서 " 및 " 순환 보고서 " 에서 확인하십시오.

X11 미국 인구조사국(US Bureau of the Census)에서 개발한 X-11 방법을 사용하여 추세 및 계절 효과를 제거합니다 (Shiskin et. al. 1967 연구 자료 참조). X-11 방법에 대한 자세한 내용은 "X-11 분해에 대한 통계 상세 정보 " 에서 확인하십시오. 이 옵션을 선택하면 " 분해 유형 선택 " 대화상자가 나타납니다. 이 대화상자 창에서 승법 또는 가법 X-11 조정을 지정할 수 있습니다. **확인**을 클릭하면 X11 보고서가 보고서 창에 추가됩니다. 자세한 내용은 "X11 보고서 " 에서 확인하십시오.

X11 옵션은 월별 또는 분기별 데이터에만 사용할 수 있습니다. X, 시간 ID 열에 공백이나 결측값 없이 월 또는 분기별로 일정한 간격의 숫자 값이 포함되어야 합니다. 이러한 요구사항을 충족하지 않는 시간 열에 대해 X11 을 요청하면 JMP 에서 오류가 반환됩니다. X11 분석에 적절한 시간 열의 예는 " 적절한 시간 ID 열 생성 " 에서 확인하십시오.

참고 : 선형 추세 제거 또는 **순환 제거** 옵션을 선택하면 추세제거 또는 순환제거 데이터가 포함된 열이 데이터 테이블에 추가됩니다. 이 열이 데이터 테이블에 이미 있는 경우 옵션을 선택하면 자동으로 기존 열을 덮어씁니다.

팁 : 일반적으로 선형 추세를 제거한 후 12 개월 주기와 같은 긴 주기를 제거하여 분해를 시작합니다. 그런 다음 6 개월 주기와 같은 짧은 주기를 제거할 수 있습니다.

시차 그림 표시 시간 *t* 의 관측값이 Y 축에 있고 시간 *t* +/- *p* 의 관측값이 X 축에 있는 그림을 표시하거나 숨깁니다. +/- *p* 는 시차라고 합니다. 이 그림은 시간 *t* 의 관측값이 시간 *t* +/- *p* 의 다른 관측값과 어떤 관계가 있는지 확인하는 데 유용합니다. 그러나 그림에 구조가 있으면 시간에 따라 관측값 사이에 어떤 관계가 있음을 나타냅니다 . 구조를 식별하면 시계열 모형을 생성할 때 도움이 됩니다.

Box-Cox 변환 그림 표시 (시작 창에서 "Box-Cox 변환 사용 " 을 지정한 경우에는 사용 불가능) "Box-Cox 변환 그림 " 을 표시하거나 숨깁니다 . Box-Cox 변환은 람다 모수 (λ) 를 기반으로 하는 멱 변환이며 다음 계산식을 사용합니다 .

$$Y_\lambda = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(y) & \text{단, } \lambda = 0 \end{cases}$$

"Box-Cox 변환 그림 " 에서는 변환된 시계열 그래프와 람다 모수 (λ) 설정 옵션을 보여 줍니다 . " 람다 " 상자에는 모수의 현재 값이 표시됩니다 . " 람다 " 슬라이더를 사용하여 모수 값을 변경할 수도 있으며 , 그러면 그림이 자동으로 업데이트됩니다 . " 낮음 " 및 " 높음 " 상자는 슬라이더 범위를 설정합니다 .

팁 : "Box-Cox 변환 그림 " 을 사용하여 다양한 람다 값에 대해 변환된 시계열을 확인할 수 있습니다 . 값을 결정한 후에는 분석을 다시 시작하고 **Box-Cox 변환 사용**을 선택한 후 "Box-Cox 에 대한 람다 " 옆에 원하는 람다 값을 입력합니다 . 이렇게 하면 변환된 계열에 대해 시계열 플랫폼이 실행됩니다 .

교차 상관 (" 전이 함수 분석 " 의 빨간색 삼각형 메뉴에서만 사용 가능) 보고서에 교차 상관 그림을 표시하거나 숨깁니다 . 그림의 길이는 자기상관 그림 길이의 두 배 또는 2 × ACF 길이 + 1 입니다 . 이 그림에는 출력 계열 대 모든 입력 계열 그림이 숫자 형식과 그래픽 형식으로 포함됩니다 . 파란색 선은 2 표준 오차를 나타냅니다 .

참고 : 교차 상관 그림의 경우 귀무가설 하의 표준 오차가 1/sqrt(n - k) 로 계산됩니다 . 여기서 n 은 비결측 데이터 값의 수이고 k 는 자기상관 시차 수입니다 .

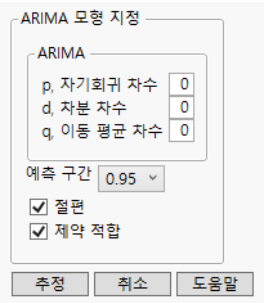
사전백색화 (" 입력 계열 " 의 빨간색 삼각형 메뉴에서만 사용 가능) 사전백색화 차수를 설정할 수 있는 " 사전백색화 규격 " 창을 표시합니다 . 사전백색화는 전이 함수 모형을 식별하는 데 사용되는 기법입니다 . 이 기법은 잔차가 백색잡음과 동일하도록 ARIMA 모형을 입력 계열에 적합시키는 방법을 사용합니다 . 그런 다음 동일한 모형을 사용하여 출력 계열을 추정합니다 . 필터링된 입력 계열과 필터링된 출력 계열 사이의 교차 상관을 사용하여 전이 함수 모형에 적절한 시차를 결정할 수 있습니다 . 사전백색화에 대한 자세한 내용은 Box et al. (1994) 연구 자료에서 확인하십시오 .

ARIMA 및 계절 ARIMA 모형

ARIMA 적합시킬 ARIMA 모형을 지정할 수 있는 "ARIMA 규격 " 창을 표시합니다 . ARIMA 모형은 과거 값과 일련의 오차 (**확률 충격** 또는 **개변**이라고도 함) 의 선형 조합을 통해 시계열의 미래 값을 예측합니다 . ARIMA 모형은 시계열에 대해 지정된 ARIMA 모형의 최대 가능도 적합을 수행합니다 . 자세한 내용은 "ARIMA 모형 " 에서 확인하십시오 .

참고 : 일반적으로 ARIMA 모형은 ARIMA(*p*,*d*,*q*) 로 표시됩니다 . *p*, *d* 또는 *q* 가 0 이면 해당 문자가 삭제됩니다 . 예를 들어 *p* 와 *d* 가 0 이면 모형은 MA(*q*) 로 표시되는 단순 이동 평균 모형이 됩니다 .

그림 18.6 ARIMA 규격 창



***p*, 자기회귀 차수** 다항식 φ(**B**) 연산자의 차수 *p* 입니다 .

***d*, 차분 차수** 차분 연산자의 차수 *d* 입니다 .

***q*, 이동 평균 차수** 차분 연산자 θ(**B**) 의 차수 *q* 입니다 .

예측 구간 예측 구간에 대해 0 에서 1 사이의 예측 수준을 설정할 수 있습니다 .

절편 절편 항 μ 를 모형에 포함할지 여부를 결정합니다 .

제약 적합 이 옵션을 선택하면 적합 절차에서 자기회귀 모수가 항상 안정 영역 내에 있고 이동 평균 모수가 가역 영역 내에 있도록 제한합니다 .

팁 : 적합기가 최적 값을 찾기 어렵거나 적합 속도를 높이려면 " 제약 적합 " 옵션을 선택 취소하십시오 . " 모형 요약 " 테이블을 사용하여 결과 적합 모형이 안정적이고 가역적인지 확인할 수 있습니다 .

모형을 지정하고 **추정**을 클릭하면 모형 보고서가 보고서 창에 추가됩니다 . 자세한 내용은 " 보고서 " 에서 확인하십시오 .

계절 ARIMA 적합시킬 계절 ARIMA 모형을 지정할 수 있는 " 계절 ARIMA 규격 " 창을 표시합니다 . 이 창에는 "ARIMA 규격 " 창과 동일한 요소가 있지만 계절 요소 규격도 포함됩니다 . " 기간별 관측값 수 " 옵션을 사용하면 *s* 로 표시되는 기간별 관측값 수를 지정할 수 있습니다 . 계절 ARIMA 모형에 대한 자세한 내용은 " 계절 ARIMA 모형 " 에서 확인하십시오 .

참고 : 계절 ARIMA 모형은 " 계절 ARIMA(*p*,*d*,*q*)(*P*,*D*,*Q*)*s*" 로 표시됩니다 .

모형을 지정하고 **추정**을 클릭하면 모형 보고서가 보고서 창에 추가됩니다 . 자세한 내용은 " 보고서 " 에서 확인하십시오 .

평활 모형

평활 모형 하위 메뉴를 표시합니다 . 평활 모형을 선택하면 규격 창이 나타납니다 . 자세한 내용은 " 평활 모형 창 " 에서 확인하십시오 . 지정된 각 모형에 대한 평활 모형 보고서가 보고서 창에 나타납니다 . 자세한 내용은 "ARIMA 및 계절 ARIMA 모형 보고서 " 에서 확인하십시오 . 평활 모형은 모형에 의한 시계열의 진화를 나타냅니다 .

$$y_t = \mu_t + \beta_t t + s(t) + a_t$$

다음은 각 요소에 대한 설명입니다 .

μ_t = 시간에 따라 변하는 평균 항

β_t = 시간에 따라 변하는 기울기 항

$s(t)$ = 시간에 따라 변하는 계절 항 중 하나

a_t = 확률 충격

일반 평활 모형 방정식에 대한 자세한 내용은 " 지수 평활 모형에 대한 통계 상세 정보 " 에서 확인하십시오 . 다음과 같은 평활 모형을 사용할 수 있습니다 .

단순 이동 평균 (시작 창에서 " 홀드백에 대한 예측 " 또는 "Box-Cox 변환 사용 " 을 선택한 경우에는 사용 불가능) 평활 창에서 정의된 인접한 여러 점의 평균을 사용하여 값을 추정하는 모형입니다 . " 단순 평활 평균 지정 " 창에서 평활 창의 특성을 지정할 수 있습니다 . 평활 창 옵션을 지정하면 " 단순 이동 평균 " 보고서가 표시됩니다 . 기본적으로 이 보고서는 시간 창에서 연속 관측값의 평균과 동일한 값이 표시된 그림을 생성합니다 . 여러 단순 이동 평균 모형을 추가하고 동일한 그림에 표시할 수 있습니다 . 자세한 내용은 " 단순 평활 평균 지정 창 " 에서 확인하십시오 .

단순 지수 평활 수준 성분이 있는 모형입니다 . 자세한 내용은 " 단순 지수 평활 " 에서 확인하십시오 .

이중 지수 평활 수준 성분과 추세 성분이 있는 모형입니다 . 이것은 선형 지수 평활의 특수한 경우입니다 . 자세한 내용은 " 이중 (Brown) 지수 평활 " 에서 확인하십시오 .

선형 지수 평활 수준 성분과 추세 성분이 있는 모형입니다 . 자세한 내용은 " 선형 (Holt) 지수 평활 " 에서 확인하십시오 .

진폭감소 추세 선형 지수 평활 수준 성분과 진폭감소 추세 성분이 있는 모형입니다 . 이 모형은 선형 추세보다 더 복잡한 추세를 나타내는 계열에 적절합니다 . 자세한 내용은 " 진폭감소 추세 선형 지수 평활 " 에서 확인하십시오 .

계절 지수 평활 수준 성분과 계절 성분이 있는 모형입니다 . 자세한 내용은 " 계절 지수 평활 " 에서 확인하십시오 .

Winters 방법 수준 성분 , 추세 성분 및 계절 성분이 있는 모형입니다 . 자세한 내용은 "Winters 방법 (가법 모형)" 에서 확인하십시오 .

참고: 단순 이동 평균 모형을 제외한 각 평활 모형은 기존 지수 평활 모형이며 ARIMA 모형과 동일합니다. 일부 평활 모형은 ARIMA 옵션에서 ARIMA 모형 모수를 허용하지 않도록 제한하는 방식이 내재되어 있으므로 ARIMA 옵션을 사용하여 동등한 ARIMA 모형을 지정하지 못할 수도 있습니다.

상태 공간 평활 모형

다양한 상태 공간 평활 모형을 적합시킬 수 있는 "상태 공간 평활 모형 지정" 창을 표시합니다. 이 내용은 Hyndman et al. (2008) 연구 자료에 정의되어 있습니다. 상태 공간 평활 모형은 오차, 추세 성분 및 계절 성분을 기반으로 정의됩니다.

- 오차는 가법 (A) 또는 승법 (M) 일 수 있습니다.
- 추세 성분은 없음 (N), 가법 (A), 진폭감소 가법 (A_d), 승법 (M) 또는 진폭감소 승법 (M_d) 일 수 있습니다.
- 계절 성분은 없음 (N), 가법 (A) 또는 승법 (M) 일 수 있습니다.

특정 모형은 ETS (오차, 추세, 계절)로 나타낼 수 있습니다. "상태 공간 평활 모형 지정" 창의 체크박스를 사용하여 원하는 모형에 대한 오차, 추세 및 계절성을 선택합니다. 플랫폼에서 권장하는 모형에 해당하는 체크박스를 선택하려면 **권장 선택**을 클릭합니다. 권장 모형이 선택된 상태로 창이 열립니다. **모두 선택**을 클릭하여 모든 체크박스를 선택하거나 **모두 선택 취소**를 클릭하여 모든 체크박스를 선택 취소할 수 있습니다. 이 창에는 다음 옵션도 포함되어 있습니다.

주기 모형 적합 프로세스에서 고려할 계절성 값을 지정합니다.

모수 제한 관측값이 현재에서 멀어질수록 모형의 현재 상태에 미치는 영향이 적어지도록 모수를 제한합니다. 상태 공간 평활 모형에서 모든 이전 관측값이 주어진 경우 시간 t 에서의 예측은 시간 t 까지의 모든 관측값의 가중합과 동일합니다. 가중치는 모수의 함수입니다. 따라서 모수를 제한하면 과거 관측값에 대한 가중치가 0 이 되고 관측값이 현재에서 멀어질수록 가중치가 더 빨리 0 이 됩니다.

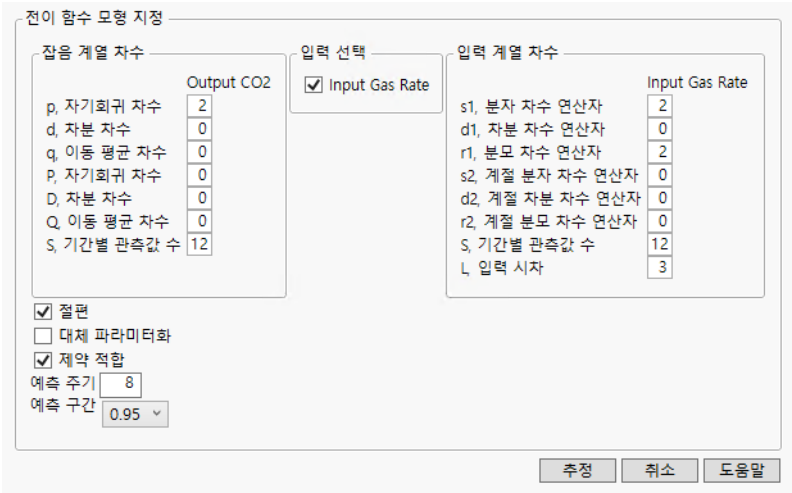
확인을 클릭하면 지정된 모형 집합이 적합됩니다. 각 상태 공간 평활 모형의 요약 값이 모형 비교 테이블에 추가됩니다. 개별 적합 보고서는 보고서 창에 추가됩니다.

전이 함수 모형

전이 함수 ("전이 함수 분석"의 빨간색 삼각형 메뉴에서만 사용 가능) "전이 함수 모형 규격" 창을 표시합니다. 전이 함수 모형을 생성하는 것은 ARIMA 모형을 생성하는 것과 비슷하며 모형 탐색, 적합 및 비교를 반복하는 프로세스입니다. 모형을 생성하기 전 데이터 탐색 프로세스 중에 데이터를 사전백색화하는 것이 유용할 때가 있습니다. 자세한 내용은 "사전백색화"에서 확인하십시오. 전이 함수에 대한 자세한 내용은 "전이 함수에 대한 통계 상세 정보"에서 확인하십시오.

참고 : 현재 " 전이 함수 " 옵션은 결측값을 제한적으로 지원합니다 . 시계열 플랫폼에서는 제외된 행도 결측값으로 처리하므로 제외된 행이 포함된 데이터 테이블에 대해서는 전이 함수 모형을 적합시킬 수 없습니다 .

그림 18.7 전이 함수 모형 규격 창



" 전이 함수 모형 규격 " 창에는 다음 섹션이 포함되어 있습니다 .

잡음 계열 차수 잡음 계열에 대한 규격을 포함합니다 . 소문자는 비계절 다항식의 계수이고 대문자는 계절 다항식의 계수입니다 .

입력 선택 모형에 대한 입력 계열을 선택할 수 있습니다 .

입력 계열 차수 입력 계열에 대한 규격을 포함합니다 . 처음 세 차수는 비계절 다항식과 관련됩니다 . 다음 네 차수는 계절 다항식과 관련됩니다 . 마지막 옵션은 입력 시차에 대한 것입니다 .

다음과 같이 모형 적합을 제어하는 세 가지 추가 옵션이 있습니다 .

절편 모형에 절편이 있는지 여부를 지정합니다 . 이 옵션을 선택하지 않으면 μ 가 0 으로 가정됩니다 .

대체 파라미터화 분자 다항식에서 일반 회귀 계수를 인수 분해할지 여부를 지정합니다 .

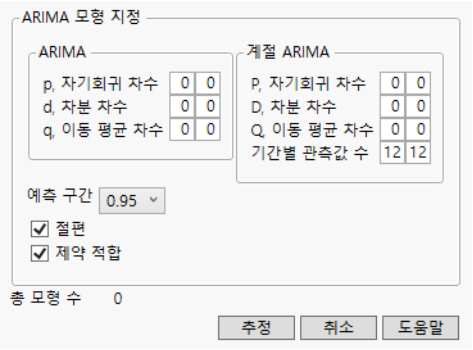
제약 적합 AR 및 MA 계수의 제약을 전환합니다 .

예측 주기 예측에 사용되는 예측 주기 수를 지정합니다 . 데이터 테이블의 끝에 Y 변수에 대한 결측값과 입력 변수에 대한 비결측 값이 포함된 행이 있으면 이러한 행이 초기 예측 설정에 사용됩니다 . 입력 변수의 값은 입력 변수의 미래 값으로 간주됩니다 .

예측 구간 예측 구간에 대한 신뢰 수준을 지정합니다 .

ARIMA 모형 그룹 차수 범위를 지정하여 다양한 ARIMA 모형 또는 계절 ARIMA 모형을 적합시킬 수 있는 "ARIMA 모형 그룹 " 창을 표시합니다 . 창에 범위를 입력하면 그에 따라 " 총 모형 수 " 가 업데이트됩니다 .

그림 18.8 ARIMA 모형 그룹 규격 창



모형을 지정하고 **추정**을 클릭하면 지정된 각 모형에 대한 모형 보고서가 보고서 창에 추가됩니다 . 자세한 내용은 " 보고서 " 에서 확인하십시오 .

모형의 예측 결합 및 저장 보고서에 있는 모든 모형 적합의 결과를 결합하여 새 데이터 테이블을 생성합니다 .

스펙트럼 밀도 저장 *(i+1)*번째 행이 빈도 $f_i = i/N$(즉, $1/N$의 i번째 조화)에 해당하는 주기도와 스펙트럼 밀도가 포함된 새 데이터 테이블을 생성합니다 . 새 데이터 테이블에는 다음과 같은 열이 있습니다 .

주기 i 번째 조화의 주기입니다 ($1/f_i$).

빈도 조화의 빈도입니다 (f_i).

각 빈도수 조화의 각 빈도수입니다 ($2\pi f_i$).

사인 Fourier 사인 계수입니다 (a_i).

코사인 Fourier 코사인 계수입니다 (b_i).

주기도 주기도입니다 ($I(f_i)$).

스펙트럼 밀도 주기도의 평활 버전인 스펙트럼 밀도입니다 .

예측 주기 수 적합 모형에 대해 예측되는 미래 기간의 수를 설정할 수 있는 창을 표시합니다 . 초기값은 시계열 시작 창에서 설정됩니다 . 이 값이 변경되면 모든 기존 및 미래 예측 결과에 새 값이 사용됩니다 .

최대 반복 수 ARIMA 모형 적합에 사용되는 미래 최적화에 대한 최대 반복 수를 재설정할 수 있는 창을 표시합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

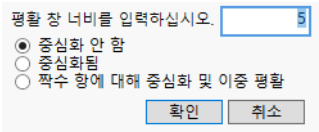
그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

평활 모형 규격 창

단순 평활 평균 지정 창

" 단순 이동 평균 " 을 평활 모형으로 선택하면 " 단순 평활 평균 지정 " 창이 나타납니다 . w 를 SMA(단순 이동 평균) 모형의 평활 창 너비라고 하겠습니다 . $f_t=(y_t+y_{t-1}+y_{t-2}+...+y_{t-(w-2)}+y_{t-(w-1)})/w$ 를 특정 시점 t 에 대한 연속된 w 개 관측값의 평균이라고 하겠습니다 .

그림 18.9 단순 평활 평균 지정 창



평활 창 너비를 입력하십시오 . 평균을 구할 연속 점 수를 정의하는 평활 창 너비 w 입니다 . 창 너비가 클수록 계열이 더 많이 평활됩니다 .

중심화 안 함 평활 창은 계열이 추정되는 점인 시점 t 까지 이어지는 연속 점으로 구성됩니다 . 즉 , f_t 는 시간 t 에 대해 그림에 표시되는 값입니다 .

중심화됨 평활 창은 계열이 추정되는 시점을 중심으로 합니다 .

- 홀수 w 의 경우 f_t 는 시간 t-(w-1)/2 에 대해 그림에 표시되는 값입니다 .
- 짝수 w 의 경우 f_t 는 시간 t-(w-1)/2 에 대해 그림에 표시되는 값입니다 . 데이터 테이블에 저장할 때 f_t 는 t-(w-2)/2 입니다 .

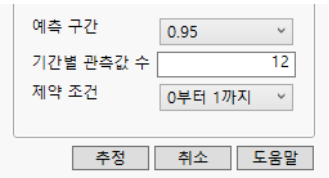
짝수 항에 대해 중심화 및 이중 평활 짝수 w 의 경우 평활 창은 계열이 추정되는 시점을 중심으로 할 수 없습니다 . 이 옵션은 거의 중심에 있는 두 개의 평활 창을 생성하여 평균을 구합니다 . 평활 추정값은 다음과 같이 계산됩니다 .

$$f_{t-\frac{w}{2}} = \frac{y_t + 2\sum_{i=1}^{w-1} y_{t-i} + y_{t-w}}{2w}$$

평활 모형 창

"단순 이동 평균" 이외의 평활 모형 옵션 중 하나를 선택하면 평활 모형 규격 창이 나타납니다. 창 제목과 사용 가능한 옵션은 선택한 평활 모형 옵션에 따라 다릅니다.

그림 18.10 평활 모형 규격 창



예측 구간 예측 구간에 대해 예측 수준을 설정할 수 있습니다.

기간별 관측값 수 (계절 평활 모형에만 사용 가능) 계절 평활 모형에서 기간별 관측값 수를 설정할 수 있습니다.

제약 조건 적합 중 평활 가중치에 적용할 제약 조건 유형을 지정할 수 있습니다. 다음과 같은 제약 조건 옵션을 사용할 수 있습니다.

0 부터 1 까지 평활 가중치 값을 0 부터 1 까지의 범위로 제한합니다.

무제약 모수의 범위를 자유롭게 지정할 수 있습니다.

안정적 가역적 동등한 ARIMA 모형이 안정적이고 가역적이 되도록 모수를 제한합니다.

사용자 정의 개별 평활 가중치에 대한 제약 조건을 설정할 수 있도록 대화상자를 확장합니다. 각 평활 가중치는 가중치 이름 옆의 팝업 메뉴 설정에 따라 **유계 (bounded)**, **고정** 또는 **무제약**일 수 있습니다. 고정 또는 유계 가중치에 대한 값을 입력할 때 해당 값은 양의 실수 또는 음의 실수일 수 있습니다.

그림 18.11 사용자 평활 가중치

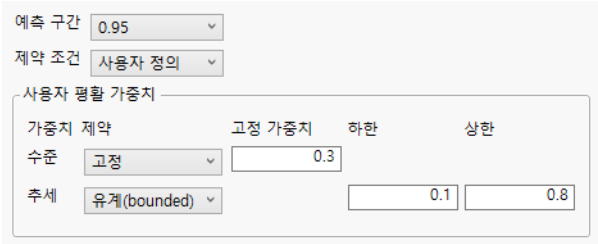


그림 18.11 에 표시된 예제에는 0.3 값으로 고정된 " 수준 " 가중치 (α) 와 0.1 ~ 0.8 범위로 제한된 " 추세 " 가중치 (γ) 가 있습니다 . 이 경우 추세 가중치 값은 0.1 ~ 0.8 범위 내에서 달라질 수 있지만 수준 가중치는 0.3 으로 유지됩니다 . 이러한 사용자 제약 조건을 사용하여 모든 평활 가중치를 미리 지정할 수 있습니다 . 이 경우 예측과 잔차는 계산되지만 가중치는 데이터에서 추정되지 않습니다 .

보고서

- " 차분 보고서 "
- " 분해 보고서 "
- " 모형 비교 보고서 "
- "ARIMA 및 계절 ARIMA 모형 보고서 "
- " 상태 공간 평활 보고서 "
- " 전이 함수 보고서 "
- " 스펙트럼 밀도 보고서 "

차분 보고서

"차분" 보고서에는 차분 계열의 자기상관 및 편자기상관 그래프가 포함됩니다. 이 그래프를 사용하여 차분 계열이 정상인지 여부를 확인할 수 있습니다.

시계열 플랫폼에서 사용할 수 있는 ARIMA 모형과 계절 ARIMA 모형은 차분 연산을 지원합니다. 이러한 모형은 먼저 차분 연산에 따라 시계열을 차분한 후 차분 계열을 적합시키는 2 단계 프로세스를 수행합니다. "차분" 옵션은 ARIMA 모형에 지정할 차분 차수를 결정하는 데 유용한 전처리 도구입니다.

" 차분 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

그래프 차분 계열 그림의 모양을 제어하는 옵션이 포함된 하위 메뉴를 표시합니다. 자세한 내용은 " 시계열 플랫폼 옵션 " 에서 확인하십시오 .

자기상관 차분 계열의 자기상관을 표시하거나 숨깁니다 .

편자기상관 차분 계열의 편자기상관을 표시하거나 숨깁니다 .

변동도 차분 계열의 변동도를 표시하거나 숨깁니다 .

저장 차분 계열의 값이 포함된 새 열을 원래 데이터 테이블에 저장합니다 . 차분 과정에서 일부 선행 요소가 손실됩니다 . 이러한 요소는 저장된 " 차분 " 열에 결측값으로 나타납니다 .

적합 제거 보고서 창에서 " 차분 " 보고서를 제거합니다 .

분해 보고서

이 섹션에서는 세 가지 분해 옵션을 통해 생성되는 보고서에 대한 상세 정보를 제공합니다 .

- " 선형 추세 보고서 "
- " 순환 보고서 "
- "X11 보고서 "

선형 추세 보고서

데이터에 적합되는 선형 회귀 모형의 β_0 및 β_1 값을 포함합니다 .

$$\text{Trend}_t = \beta_0 + \beta_1 time$$

추세제거 계열은 $D_t = O_t$ - Trend_t와 같습니다. 여기서 O_t는 원래 시계열입니다.

순환 보고서

데이터에 적합되는 순환 성분의 값을 포함합니다 .

$$\text{Cycle}_t = C + A \cdot \cos\left(2 \cdot \pi \cdot \left(\left(\frac{1}{U}\right) \cdot t + P\right)\right)$$

모수 값은 다음과 같이 정의됩니다 .

- C = 상수 (선택적)
- A = 코사인파의 진폭
- U = 순환별 단위 수
- P = 코사인파의 시기입니다 .
- t = 주어진 관측값의 행 번호에서 1 을 뺀 값

순환제거 계열은 $D_t = O_t$ - Cycle_t와 같습니다. 여기서 O_t는 원래 시계열입니다.

X11 보고서

분해 유형 선택에 따라 X11 옵션은 "X11 - 승법 " 보고서 또는 "X11 - 가법 " 보고서를 추가합니다 . 보고서에는 다음과 같은 네 개의 동일한 그림이 포함됩니다 .

원래 및 조정 원래 시계열 O_t 위에 X11 조정 시계열을 중첩합니다 . X11 조정 값은 승법 조정의 경우 O_t/S_t 이고 가법 조정의 경우 $O_t - S_t$ 입니다 .

D10 - 최종 계절 요인 시간에 따른 계절 요인 성분 S_t 를 표시합니다 .

D12 - 최종 추세 순환 시간에 따른 추세 순환 성분 C_t 를 표시합니다 .

D13 - 최종 불규칙 계열 시간에 따른 불규칙 성분 I_t 를 표시합니다 .

X11 보고서 옵션

X11 보고서에는 다음과 같은 빨간색 삼각형 옵션이 있습니다 .

테이블 표시 X11 요약 테이블을 표시하거나 숨깁니다 . 자세한 내용은 Shiskin et. al. (1967) 연구 자료에서 확인하십시오 . 테이블은 B ~ F 라벨이 지정된 5 개의 범주로 그룹화됩니다 (표 18.1 의 설명 참조).

열 저장 계절 조정 시계열 , 계절 성분 (S_t), 추세 순환 성분 (C_t) 및 불규칙 계열 성분 (I_t) 에 대한 네 개의 열을 데이터 테이블에 저장합니다 .

모든 열 저장 보고서에서 **테이블 표시** 옵션으로 생성된 모든 테이블에 대한 열을 데이터 테이블에 저장합니다 .

적합 제거 보고서 창에서 X11 보고서를 제거합니다 .

표 18.1 X11 출력 테이블의 범주 설명

맨 앞의 문자	범주 설명
B	계절, 추세 순환 및 불규칙 성분의 예비 추정값
C	계절, 추세 순환 및 불규칙 성분의 중간 추정값
D	계절, 추세 순환 및 불규칙 성분의 최종 추정값
E	분석 테이블
F	요약 지표

X11 출력 테이블의 내용에 대한 자세한 정보는 Shisken et. al. (1967) 연구 자료 또는 SAS/ETS 15.2 User's Guide("The output from PROC X11" 검색) 에서 확인하십시오 .

모형 비교 보고서

모형이 적합되면 보고서 창에 " 모형 비교 " 보고서가 표시됩니다 . 이 보고서에는 모형 비교 테이블과 모형 그림이 포함되어 있습니다 . 새 모형이 적합될 때마다 고유한 색상 코딩을 사용하여 새 행이 모형 비교 테이블에 추가됩니다. 모형 비교 테이블은 각 모형에 대한 적합 통계량의 요약 정보를 제공하며 , 동일한 시계열에 적합된 여러 모형을 비교하는 데 사용됩니다 . 기본적으로 모형은 AIC 통계량에 따라 오름차순으로 정렬됩니다 . 다른 통계량을 기준으로 모형을 정렬하려면 모형 비교 테이블에서 아무 열이나 마우스 오른쪽 버튼으로 클릭하고 " 열별 정렬 " 을 선택합니다 . 아니면 열 머리글을 클릭하여 정렬할 수도 있습니다 . 적합 통계량에 대한 설명은 " 모형 요약 테이블 " 에서 확인하십시오 . 모형 비교 테이블에만 있는 적합 통계량은 " 가중치 " 입니다 . 이 적합 통계량은 정규화된 AIC 가중치입니다 . 모형의 AIC 가중치는 다음과 같이 계산됩니다 .

$$\text{AICWeight} = \exp[-0.5(\text{AIC} - \text{BestAIC})] / \sum_{k=1}^{K} (\exp[-0.5(\text{AIC}_k - \text{BestAIC})])$$

K는 총 모형 수, AIC_k는 모형 k에 대한 AIC 값, BestAIC는 AIC 값이 가장 작은 모형에 대한 AIC 값입니다.

홀드백 집합을 지정하면 훈련 데이터에 대해 계산된 적합 통계량이 모형 비교 테이블에 표시되지 않습니다. 대신 홀드백 집합에 대한 개별 모형의 예측 성능을 판별하기 위해 RMSE(제곱근 평균 제곱 오차), MSE(평균 제곱 오차), MAPE(평균 절대 백분율 오차) 및 MAE(평균 절대 오차)와 같은 측정 기준이 계산됩니다. 기본적으로 모형은 RMSE 통계량에 따라 오름차순으로 정렬됩니다. 다른 통계량을 기준으로 모형을 정렬하려면 모형 비교 테이블에서 아무 열이나 마우스 오른쪽 버튼으로 클릭하고 "열별 정렬"을 선택합니다. 아니면 열 머리글을 클릭하여 정렬할 수도 있습니다. 적합 통계량에 대한 설명은 "모형 요약 테이블"에서 확인하십시오.

주의 : 홀드백 집합을 지정하지 않고 상태 공간 평활 모형과 ARIMA 모형의 조합을 적합시키면 모형 비교 테이블 맨 위에 주의 메시지가 표시됩니다 . 서로 다른 두 모형 계층 간에는 가능도 값 및 가능도 기반 정보 기준 (예 : AIC 통계량) 을 비교할 수 없으므로 이 경우 상태 공간 평활 모형을 ARIMA 모형과 비교할 수 없습니다 . 그러나 홀드백 집합을 지정하면 RMSE 통계량이 사용되므로 모형을 비교할 수 있습니다 .

그림 18.12 모형 비교 테이블

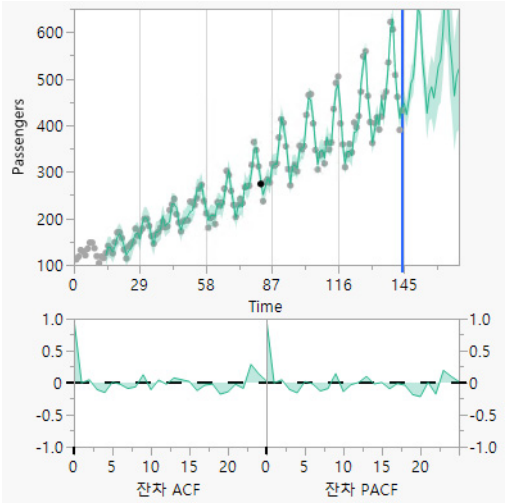
모형 비교

보고서	그래프	모형	DF	분산	AIC ^	SBC	R^2	-2*로그 가능도	가중치	.2 .4 .6 .8	MAPE	MAE
☑	☐	계절 ARIMA(0, 1, 1)(0, 1, 1)12	128	138.49122	1020.9047	1029.5302	0.990	1014.9047	1.000000		3.197010	8.990978
☑	☐	ARIMA(1, 1, 1)	140	978.9437	1394.1215	1403.0101	0.932	1388.1215	0.000000		8.687058	24.364166
☑	☐	ARMA(1, 1)	141	989.13835	1407.7483	1416.6577	0.919	1401.7483	0.000000		9.704486	25.348507
☑	☐	AR(1)	142	1134.3871	1426.1794	1432.1190	0.909	1422.1794	0.000000		9.873726	26.556219
☑	☐	MA(1)	142	4264.6121	1616.8626	1622.8022	0.693	1612.8626	0.000000		23.865796	53.716974

보고서 체크박스를 사용하여 보고서 창에 표시할 전체 모형 보고서를 선택할 수 있습니다. 모형 비교 테이블 오른쪽에 두 개의 모형 그림이 나타납니다. 위의 그림은 데이터, 예측 및 예측 한계

에 대한 시계열 그림입니다. 아래에는 자기상관 및 편자기상관 함수에 대한 그림이 있습니다. **그래프** 체크박스를 사용하여 모형 그림에 표시할 모형을 선택할 수 있습니다.

그림 18.13 모형 그림



모형 비교 보고서 옵션

" 모형 비교 " 보고서의 각 모형에는 다음과 같은 빨간색 삼각형 메뉴 옵션이 있습니다 .

새로 적합 모형에 대한 규격 창을 엽니다 . 설정을 변경하여 다른 모형을 적합시킬 수 있습니다 .

한 번 시뮬레이션 (홀드백 집합을 지정한 경우에는 사용 불가능) 시간 축의 위쪽 끝에 모형 시뮬레이션을 한 번 제공합니다 . 모형 비교 시계열 그림에 시뮬레이션이 표시됩니다 .

추가 시뮬레이션 (홀드백 집합을 지정한 경우에는 사용 불가능) 시간 축의 위쪽 끝에 지정된 수의 모형 시뮬레이션을 제공합니다 . 모형 비교 시계열 그림에 시뮬레이션이 표시됩니다 .

모형 시뮬레이션 제거 (홀드백 집합을 지정한 경우에는 사용 불가능) 모형에 대한 시뮬레이션을 제거합니다 .

모든 시뮬레이션 제거 (홀드백 집합을 지정한 경우에는 사용 불가능) 모든 모형에 대한 시뮬레이션을 제거합니다 .

시뮬레이션 생성 (홀드백 집합을 지정한 경우에는 사용 불가능) 모형에 대한 시뮬레이션을 생성하고 결과를 데이터 테이블에 저장합니다 . 난수 시드값 , 시뮬레이션 수 및 예측 주기 수를 지정할 수 있습니다 .

시드값 설정 (홀드백 집합을 지정한 경우에는 사용 불가능) 시뮬레이션된 궤적을 생성하기 위한 시드값을 지정합니다 .

적합 제거 보고서 창에서 해당 모형을 제거합니다 .

모든 모형 제거 보고서 창에서 모형 비교 테이블에 나열된 모든 모형을 제거합니다 .

선택 취소한 항목 제거 보고서 창에서 " 보고서 " 체크박스가 선택되지 않은 모든 모형을 제거합니다 .

선택한 항목 제거 보고서 창에서 " 보고서 " 체크박스가 선택된 모든 모형을 제거합니다 .

모든 보고서 숨기기 보고서 창에서 모형 비교 테이블에 나열된 모든 모형을 숨깁니다 . 이것은 " 보고서 " 체크박스를 모두 선택 취소하는 것과 같습니다 .

모든 그래프 숨기기 모형 그림에서 모형 비교 테이블에 나열된 모든 모형을 숨깁니다 . 이것은 " 그래프 " 체크박스를 모두 선택 취소하는 것과 같습니다 .

참고 : 시계열 플랫폼에서는 제외된 행을 결측값으로 처리합니다 . 데이터 테이블에서 마지막 몇몇 관측값이 제외되거나 결측인 경우 " 모형 비교 " 보고서의 시뮬레이션 옵션이 제대로 작동하지 않습니다 . 시뮬레이션 절차에는 시계열의 마지막 몇몇 관측값이 사용됩니다 . 마지막 몇몇 관측값이 결측인 경우 시뮬레이션을 수행하면 모든 미래 결측값이 생성됩니다 .

ARIMA 및 계절 ARIMA 모형 보고서

시계열 모델링 옵션은 이론적 모형을 계열에 적합시키고 적합 모형을 사용하여 계열의 미래 값을 예측하는 데 사용됩니다 . 또한 이러한 옵션을 사용하면 선택한 모형의 적합성을 확인할 수 있는 통계량 및 잔차가 생성됩니다 . 모델링 옵션을 여러 번 선택할 수 있습니다 . 모형을 선택할 때마다 해당 모형이 모형 비교 테이블에 추가됩니다 . 모형 비교 테이블에서 모형 옆의 "보고서" 체크박스를 선택하면 해당 모형에 대한 보고서가 생성됩니다 . 보고서 제목에 모형이 지정됩니다 .

기본적으로 다음 보고서가 표시됩니다 .

- 모형 요약 테이블
- 모수 추정값 테이블
- 예측 그림
- 잔차
- 반복 기록

모형 요약 테이블 모형에 대한 적합 통계량을 포함합니다 . 아래 계산식에서 n 은 계열의 길이이고 k 는 모형에 있는 적합 모수의 수입니다 .

DF 적합의 자유도입니다 $(n - k)$.

개변 제곱합 개변 제곱의 합입니다 .

잔차 제곱합 잔차 제곱의 합입니다 .

분산 추정값 개변 제곱합을 자유도 $(n - k)$ 로 나눈 값입니다 . 이 값은 "ARIMA 모형 " 섹션에 설명된 확률 충격 a_t 의 분산에 대한 표본 추정값입니다 .

표준편차 분산 추정값의 제곱근입니다 . 이 값은 확률 충격 a_t 의 표준편차에 대한 표본 추정값입니다 .

Akaike 정보 기준 –2* 로그 가능도 + 2*k* 로 계산되는 AIC 값입니다 . AIC 값이 작을수록 더 나은 적합을 나타냅니다 .

Schwarz Bayes 기준 –2* 로그 가능도 + *k**ln(*n*) 으로 계산되는 SBC 값입니다 . SBC 값이 작을수록 더 나은 적합을 나타냅니다 . Schwarz Bayes 기준은 BIC(Bayes 정보 기준) 와 같습니다 .

R² 다음과 같이 계산되는 R² 값입니다 .

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

다음은 각 요소에 대한 설명입니다 .

$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$\hat{y}_i$ 1 시차 후 예측

$\bar{y}$ 평균

모형이 계열에 잘 적합되지 않으면 모형 SSE(오차 제곱합) 가 SST(총 제곱합) 보다 클 수 있습니다 . 따라서 R^2 이 음수일 수 있습니다 .

Adj-R² 다음과 같이 계산되는 조정 R^2 값입니다 .

$$1 - \left[\frac{(n-1)}{(n-k)}(1 - R^2)\right]$$

MAPE 다음과 같이 계산되는 평균 절대 백분율 오차 값입니다 .

$$\frac{100}{n} \sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

MAE 다음과 같이 계산되는 평균 절대 오차 값입니다 .

$$\frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

-2* 로그 가능도 최량 적합 모수 추정값에서 계산된 음의 로그 가능도 함수의 두 배입니다 . 값이 작을수록 더 나은 적합을 나타냅니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

안정적 자기회귀 연산자가 안정적인지 여부를 나타냅니다 . 즉 , $\phi(z) = 0$ 의 모든 제곱근이 단위 원을 벗어나는지 여부입니다 .

가역적 이동 평균 연산자가 가역적인지 여부를 나타냅니다 . 즉 , $\theta(z) = 0$ 의 모든 제곱근이 단위 원을 벗어나는지 여부입니다 .

참고 : ϕ 및 θ 연산자는 "ARIMA 모형 " 섹션에 정의되어 있습니다 .

모수 추정값 테이블 시계열 모형 모수의 추정값을 표시합니다 . 각 모형 유형마다 고유한 모수 집합이 있으며 , 이러한 모수는 특정 시계열 모형에 대한 섹션에 설명되어 있습니다 . 각 모수 추정값 테이블에는 다음과 같은 열이 포함됩니다 .

항 각 모형 유형에 대한 섹션에 설명된 모수의 이름입니다 . 일부 모형에는 **절편** 또는 평균 항이 포함됩니다 . 이러한 모형에서는 관련 **상수 추정값**도 표시됩니다 . 상수 추정값에 대한 정의는 ARIMA 모형 설명 아래에 제공됩니다 .

요인 (승법 계절 ARIMA 모형에만 표시됨) 모수를 포함하는 모형의 요인입니다 . 승법 계절 모형에서 요인 1 은 비계절 요인이고 요인 2 는 계절 요인입니다 .

시차 (ARIMA 및 계절 ARIMA 모형에만 표시됨) 모수를 곱한 항에 적용되는 시차 또는 후방 연산자의 차수입니다 .

추정값 시계열 모형의 모수 추정값입니다 .

표준 오차 모수 추정값의 표준 오차 추정값입니다 . 이 추정값은 검정 및 예측 구간을 계산하는 데 사용됩니다 .

t 비 각 모수가 0 이라는 가설에 대한 검정 통계량입니다 . 모수에 대한 검정 통계량은 표준 오차에 대한 모수 추정값의 비율입니다 . 가설이 참이면 이 통계량은 근사 스튜던트 *t* 분포를 따릅니다 . 절대값이 2 보다 큰 *t* 비는 0.05 유의 수준에 가까우므로 이러한 값을 찾는 것이 유의성을 판단하는 일반적인 규칙입니다 .

Prob>|t| 각 모수에 대해 계산되는 관측된 *p* 값입니다 . *p* 값은 가설이 참일 때 *t* 비 (절대값 기준) 가 계산된 값보다 클 확률입니다 .

상수 추정값 (절편 또는 평균 항이 포함된 모형에만 표시됨) 상수 추정값에 대한 정의는 ARIMA 모형 아래에 제공됩니다 .

Mu (ARIMA 및 계절 ARIMA 모형에만 표시됨) ARIMA 또는 계절 ARIMA 모형의 절편 값에 대한 추정값입니다 .

예측 시계열에 대한 관측값과 예측값을 모두 보여 주는 그림입니다 . 이 그림은 세로선에 의해 두 영역으로 구분됩니다 .

홀드백 집합을 지정한 경우 세로선은 훈련 데이터와 홀드백 데이터를 구분합니다 . 세로선 왼쪽에는 1 시차 후 예측이 훈련 데이터 점과 중첩됩니다 . 세로선 오른쪽에는 모형에서 예측한

값이 홀드백 데이터 점과 중첩됩니다 . 예측 구간도 표시됩니다 . 플랫폼 시작 창의 " 예측 주기 " 옵션을 사용하여 예측 값 수를 제어할 수 있습니다 .

홀드백 집합을 지정하지 않은 경우 세로선은 관측된 데이터와 미래 예측을 구분합니다. 세로선 왼쪽에는 1 시차 후 예측이 관측된 데이터 점과 중첩됩니다 . 세로선 오른쪽에는 모형에서 예측한 미래 값과 예측 구간이 표시됩니다 . 플랫폼 시작 창의 " 예측 주기 " 옵션을 사용하거나 " 시계열 " 의 빨간색 삼각형 메뉴에서 " 예측 주기 수 " 를 선택하여 미래 예측 값 수를 제어할 수 있습니다 .

잔차 적합 모형을 기반으로 한 잔차 값을 보여 주는 그래프입니다 . 이러한 값은 시계열의 관측값에서 1 시차 후 예측값을 뺀 값입니다 . 잔차에 대한 자기상관 및 편자기상관 보고서도 표시됩니다. 이 보고서를 사용하면 적합 모형이 데이터를 설명하기에 적절한지 여부를 결정할 수 있습니다 . 잔차 그림의 점이 0 주위에서 정규 분포를 따르고 , 0 보다 큰 시차에 대해 잔차의 자기상관 및 편자기상관에 유의한 성분이 없으면 적합 모형이 적절하다고 할 수 있습니다 .

홀드백 계열에 대한 예측 오차 (홀드백 집합을 지정한 경우에만 사용 가능) 홀드백 집합의 관측값에 대한 예측 오차 그림입니다 . 오차에 대한 자기상관 및 편자기상관 보고서도 표시됩니다 .

반복 기록 각 반복에서 목적 함수의 값을 포함합니다 . 이 옵션은 적합 절차의 문제를 진단하는 데 유용할 수 있습니다 . 데이터에 잘 맞지 않는 모형을 적합시키려고 하면 가능도에 대한 최적값에서 수렴하지 못할 수 있습니다 . " 반복 기록 " 테이블에는 다음과 같은 통계량이 포함됩니다 .

반복 반복 번호입니다 .

반복 기록 각 단계의 목적 함수 값입니다 .

단계 반복 단계의 유형입니다 .

목적 함수 - 기준 목적 함수의 그래디언트 norm 입니다 .

ARIMA 및 계절 ARIMA 모형 보고서 옵션

각 ARIMA 및 계절 ARIMA 모형 보고서에는 다음 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다 .

점 표시 예측 그래프에 데이터 점을 표시하거나 숨깁니다 .

예측 구간 표시 예측 그래프에 예측 구간을 표시하거나 숨깁니다 .

열 저장 모형 결과를 나타내는 열이 포함된 새 데이터 테이블을 생성합니다 .

예측 계산식 저장 (홀드백 집합을 지정한 경우에는 사용 불가능) 데이터 및 예측 계산식을 새 데이터 테이블에 저장합니다 .

SAS 작업 생성 SAS 에서 모형 분석을 복제하는 SAS 코드를 생성합니다 .

SAS 로 전송 모형 분석을 복제하는 SAS 코드를 SAS 로 전송합니다 . SAS 서버에 연결되어 있지 않은 경우 이 옵션을 선택하면 연결 과정을 안내합니다 .

잔차 통계량 모형에 대해 나타낼 잔차 통계량의 표시 유형을 제어합니다 . 이러한 표시 유형은 "시계열 플랫폼 옵션" 섹션에서 설명합니다. 그러나 이 설정은 일련의 잔차에 적용됩니다.

적합 제거 보고서에서 개별 모형 적합을 제거합니다 .

상태 공간 평활 보고서

각 상태 공간 평활 모형이 모형 비교 테이블에 추가됩니다 . 모형 비교 테이블에서 상태 공간 평활 모형에 대한 " 보고서 " 체크박스를 선택하면 " 상태 공간 평활 " 보고서가 나타납니다 . 각 " 상태 공간 평활 " 보고서에는 모형을 식별하는 고유한 개요 이름이 있습니다 . 다음 보고서가 포함됩니다 .

모형 유형 계열의 최량 적합 모형에 대한 오차 , 추세 및 계절성을 설명합니다 .

모형 요약 모형에 대한 적합 통계량을 포함합니다 . 아래 계산식에서 n 은 계열의 길이이고 k 는 모형에 있는 적합 모수의 수입니다 .

-2* 로그 가능도 최량 적합 모수 추정값에서 계산된 음의 로그 가능도 함수의 두 배입니다 . 값이 작을수록 더 나은 적합을 나타냅니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

AIC –2*로그 가능도 + 2k로 계산되는 AIC(Akaike 정보 기준) 값입니다. AIC 값이 작을수록 더 나은 적합을 나타냅니다 .

BIC –2*로그 가능도 + k*ln(n)으로 계산되는 SBC(Schwarz Bayes 기준) 값입니다. SBC 값이 작을수록 더 나은 적합을 나타냅니다 . Schwarz Bayes 기준은 BIC(Bayes 정보 기준) 와 같습니다 .

모수 수 정보 기준 계산의 모수 수입니다 . 이 값은 가능도 함수의 자유 모수 수보다 하나 더 큽니다 .

> **참고 :** " 모수 수 " 는 모수 추정값 테이블에 있는 모수의 수가 아닙니다 .

시그마 1 시차 후 예측 표준편차의 추정값입니다 .

R² 다음과 같이 계산되는 R² 값입니다 .

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

다음은 각 요소에 대한 설명입니다 .

$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$\hat{y}_i$ 1 시차 후 예측

$\bar{y}$ 평균

모형이 계열에 잘 적합되지 않으면 모형 SSE(오차 제곱합) 가 SST(총 제곱합) 보다 클 수 있습니다 . 따라서 R^2 이 음수일 수 있습니다 .

Adj-R² 다음과 같이 계산되는 조정 R^2 값입니다 .

$$1-\left[\frac{(n-1)}{(n-k)}(1-R^2)\right]$$

MAPE 다음과 같이 계산되는 평균 절대 백분율 오차 값입니다 .

$$\frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|$$

여기서 ,

$\hat{y}_i$ 1 시차 후 예측

MAE 다음과 같이 계산되는 평균 절대 오차 값입니다 .

$$\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|$$

여기서 ,

$\hat{y}_i$ 1 시차 후 예측

모수 추정값 모형의 각 모수에 대한 추정값 , 표준 오차 및 95% 신뢰 구간을 보여 주는 테이블입니다 .

예측 시계열에 대한 관측값과 예측값을 모두 보여 주는 그림입니다 . 자세한 내용은 " 예측 " 에서 확인하십시오 .

1 시차 후 예측 오차 시간에 따른 1 시차 후 예측 오차에 대한 그림입니다 . 오차는 각 시점 t 에 대해 t 에서의 시계열 관측값에서 t 에서의 1 시차 후 예측값을 뺀 값으로 계산됩니다 . 홀드백 집합을 지정하면 훈련 데이터에 대해서만 오차가 계산됩니다. 오차에 대한 자기상관 및 편자기상관 보고서도 표시됩니다 .

홀드백 계열에 대한 예측 오차 (홀드백 집합을 지정한 경우에만 사용 가능) 홀드백 집합의 관측값에 대한 예측 오차 그림입니다 . 오차에 대한 자기상관 및 편자기상관 보고서도 표시됩니다 .

성분 상태 모형의 각 상태에 대한 그림과 모형의 해당 계산식을 포함하는 보고서입니다 . 각 그림에는 계열의 상태 추정값과 상태 예측이 포함됩니다 . 홀드백 집합을 지정한 경우 상태 추정값은 훈련 부분을 대상으로 하고 상태 예측은 홀드백 부분을 대상으로 합니다 . 홀드백 집합을 지정하지 않은 경우 상태 추정값은 시계열의 마지막 관측값까지 포함하는 부분을 대상으로 하고 상태 예측은 시계열의 마지막 관측값 이후를 대상으로 합니다 . 모형은 수준 , 추세 및 계절이라는 최대 세 개의 상태를 가질 수 있습니다 .

" 성분 상태 " 의 빨간색 삼각형 메뉴에는 " 상태 저장 " 옵션이 있습니다 . 각 시점의 성분 상태를 별도의 데이터 테이블에 저장하려면 이 옵션을 선택합니다 .

승법 오차 모형의 1 시차 후 상대 예측 오차 (승법 오차 모형에만 사용 가능) 시간에 따른 1 시차 후 상대 예측 오차에 대한 그림입니다 . 오차는 각 시점 *t* 에 대해 *t* 에서의 1 시차 후 예측 오차를 *t* 에서의 1 시차 후 예측값으로 나누어 계산됩니다 .

상태 공간 평활 보고서 옵션

각 상태 공간 평활 보고서에는 다음 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다 .

점 표시 예측 그래프에 데이터 점을 표시하거나 숨깁니다 .

예측 구간 표시 예측 그래프에 예측 구간을 표시하거나 숨깁니다 .

열 저장 모형 결과를 나타내는 열이 포함된 새 데이터 테이블을 생성합니다 .

적합 제거 보고서에서 개별 모형 적합을 제거합니다 .

전이 함수 보고서

각 전이 함수 모형이 모형 비교 테이블에 추가됩니다. 모형 비교 테이블에서 전이 함수 모형에 대한 " 보고서 " 체크박스를 선택하면 " 전이 함수 모형 " 보고서가 생성됩니다 . 홀드백 집합이 지정되면 " 모형 비교 " 보고서에 모형 그림도 제공됩니다 .

각 " 전이 함수 모형 " 보고서에는 다음 보고서가 포함됩니다 .

- 모형 요약
- 모수 추정값
- 잔차
- 홀드백 계열에 대한 예측 오차
- 대화식 예측
- 반복 기록

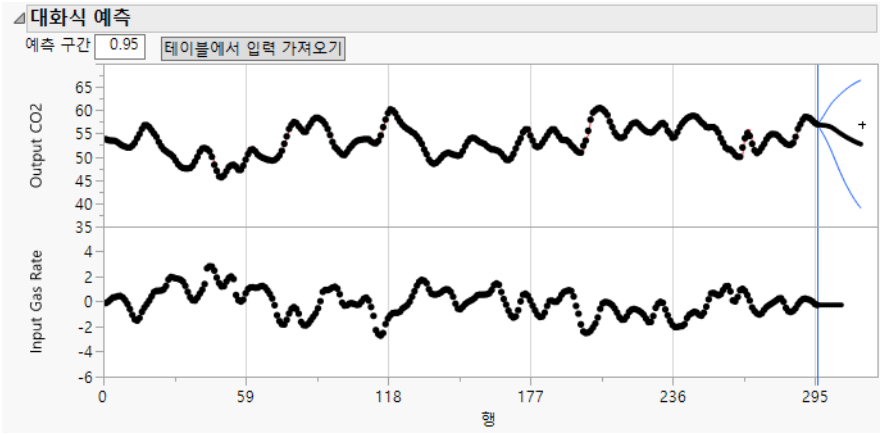
"모형 요약", "모수 추정값", "잔차", "홀드백 계열에 대한 예측 오차" 및 "반복 기록" 보고서의 정보는 ARIMA 및 계절 ARIMA 모형 보고서와 동일합니다. 이러한 보고서에 대한 자세한 내용은 "ARIMA 및 계절 ARIMA 모형 보고서"에서 확인하십시오. "모수 추정값" 테이블 다음에는 모형의 계산식이 나옵니다. 여기서 **B**는 후방 연산자입니다.

대화식 예측

"대화식 예측" 보고서에는 "전이 함수 모형 규격" 창에 지정된 예측 구간을 기반으로 예측 그래프가 제공됩니다. 예측 구간은 파란색으로 표시됩니다. 그래프 위의 "예측 구간" 상자에 숫자를 입력하여 이 예측 구간에 대한 신뢰 수준을 변경합니다.

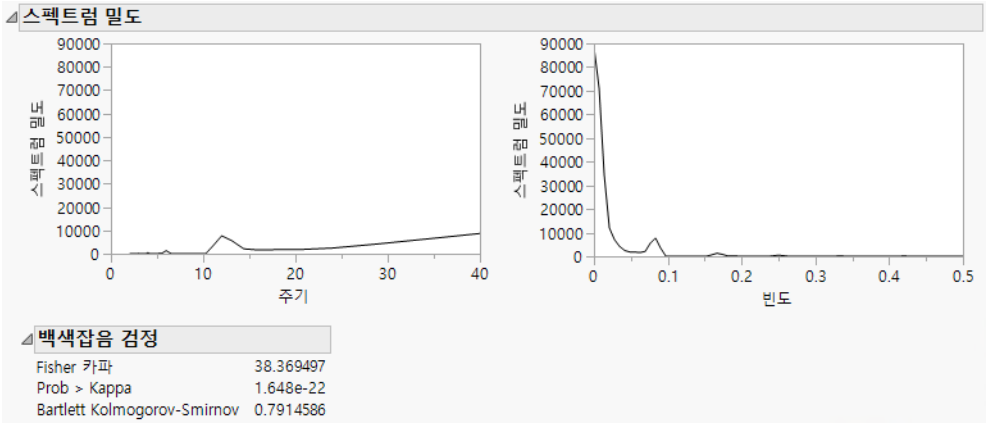
그래프에서 + 기호를 드래그하여 그래프의 예측 주기 수를 변경할 수 있습니다. 예측 주기에서는 "테이블에서 입력 가져오기" 버튼을 사용하거나, 입력 그래프의 점을 다른 값으로 드래그하여 입력 값을 변경할 수 있습니다. 변경 결과는 출력 그래프의 예측 주기에 반영됩니다.

그림 18.14 대화식 예측 그래프



스펙트럼 밀도 보고서

그림 18.15 스펙트럼 밀도 그림 및 백색잡음 검정 보고서



" 백색잡음 검정 " 보고서에는 다음 통계량이 포함됩니다 .

Fisher 카파 계열의 값을 분산이 1인 정규 분포에서 가져온다는 귀무가설을 계열에 주기 성분이 있다는 대립가설과 대조하여 검정합니다 . 카파는 주기도의 최대값 $I(f_i)$ 와 평균 값의 비율입니다 .

Prob > Kappa 귀무가설이 참일 때 카파보다 큰 값을 관측할 확률이며 다음 방정식으로 지정됩니다 .

$$Pr(k > \kappa) = 1 - \sum_{j=0}^{q} (-1)^j \binom{q}{j} \left[\max\left(1 - \frac{jk}{q}, 0\right)\right]^{q-1}$$

다음은 각 요소에 대한 설명입니다 .

$q = N/2$(N 이 짝수인 경우), $q = (N - 1)/2$(N 이 홀수인 경우)

κ = 관측된 카파 값

이 확률이 유의 수준 α 보다 작으면 귀무가설이 기각됩니다 .

Bartlett Kolmogorov-Smirnov 정규화된 누적 주기도를 (0, 1) 구간에서 균등 분포의 누적 분포 함수와 비교합니다 . 검정 통계량은 누적 주기도와 균등 CDF 의 최대 절대 차이와 같습니다 . 일반적으로 값이 $a/(\sqrt{q})$ 를 초과하면 계열이 정규 분포를 따른다는 가설을 기각합니다 . $a = 1.36$ 및 $a = 1.63$ 값은 각각 5% 와 1% 의 유의 수준에 해당합니다 .

시계열 플랫폼의 추가 예

- " 시간 ID 열 생성의 예 "
- "Box-Cox 변환을 사용한 예 "
- " 홀드백 집합을 사용한 예 "

시간 ID 열 생성의 예

이 예에서는 SeriesP.jmp 샘플 데이터 테이블을 사용하여 시계열 분석을 수행하는 방법을 보여줍니다. 먼저 시간 ID에 적절한 새 열을 생성합니다.

적절한 시간 ID 열 생성

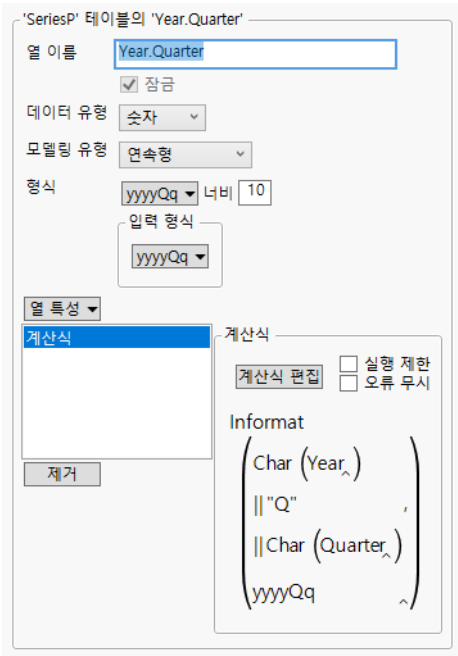
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Time Series/SeriesP.jmp 를 엽니다 .

   SeriesP.jmp 데이터 테이블에는 반응이 관측된 기간을 식별하는 Year 열과 Quarter 열이 포함되어 있습니다 . 그러나 시계열 플랫폼에서 X 축에 라벨을 지정하려면 일정 간격의 고유한 시점이 포함된 하나의 열이 필요합니다 . 시간 ID 를 지정하지 않으면 행 번호를 사용하여 기간을 식별합니다 . 이를 방지하고 보고서를 더 쉽게 해석하기 위해 Year 및 Quarter 에서 시간 ID 열을 생성합니다 .
2. **열 > 새 열**을 선택합니다 . " 열 이름 " 상자에 Year.Quarter 를 입력합니다 .

3. " 형식 " 옆에서 **최적 > 날짜 > yyyyQq** 를 선택합니다 .
4. **열 특성 > 계산식**을 선택합니다 .
5. " 날짜 / 시간 " 옆의 회색 삼각형을 클릭하고 **Informat** 을 클릭합니다 .
6. " 문자 " 옆의 회색 삼각형을 클릭합니다 .
7. Year 를 선택하고 **Char, Concat** 를 차례로 클릭합니다 .
8. 상자에 "Q"(인용 부호 포함) 를 입력하고 Enter 키를 누릅니다 .
9. **Concat** 를 클릭합니다 .
10. Quarter 를 선택하고 **Char** 을 클릭합니다 .
11. <formatName> 상자를 클릭하고 yyyyQq 를 입력한 후 Enter 키를 누릅니다 .
12. **확인**을 클릭합니다 .

    그림 18.16 에 나오는 것과 같은 완료된 " 새 열 " 창이 나타납니다 .

그림 18.16 새 열



13. **확인**을 클릭합니다 .

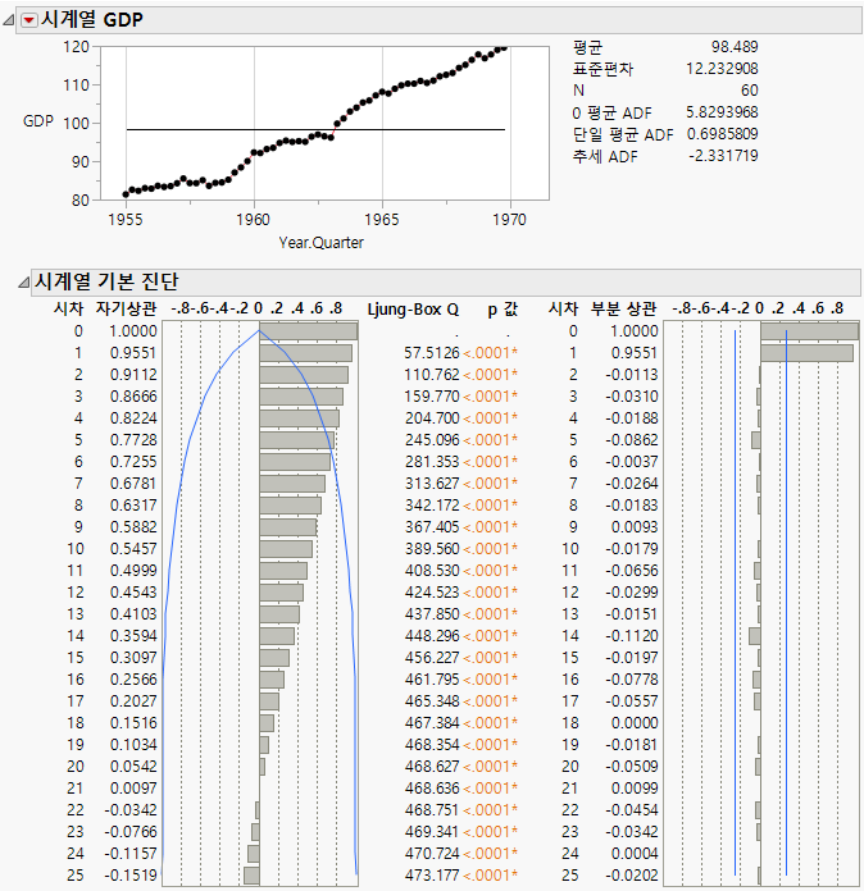
참고 : 이 시간 열은 X11 분석에도 사용됩니다 .

시계열 분석

데이터 테이블에 적절한 시간 ID 열이 포함되었으므로 분석을 진행합니다.

1. **분석 > 전문 모델링 > 시계열**을 선택합니다 .
2. GDP 를 선택하고 **Y, 시계열**을 클릭합니다 .
3. Year.Quarter 를 선택하고 **X, 시간 ID** 를 클릭합니다 .
4. **확인**을 클릭합니다 .

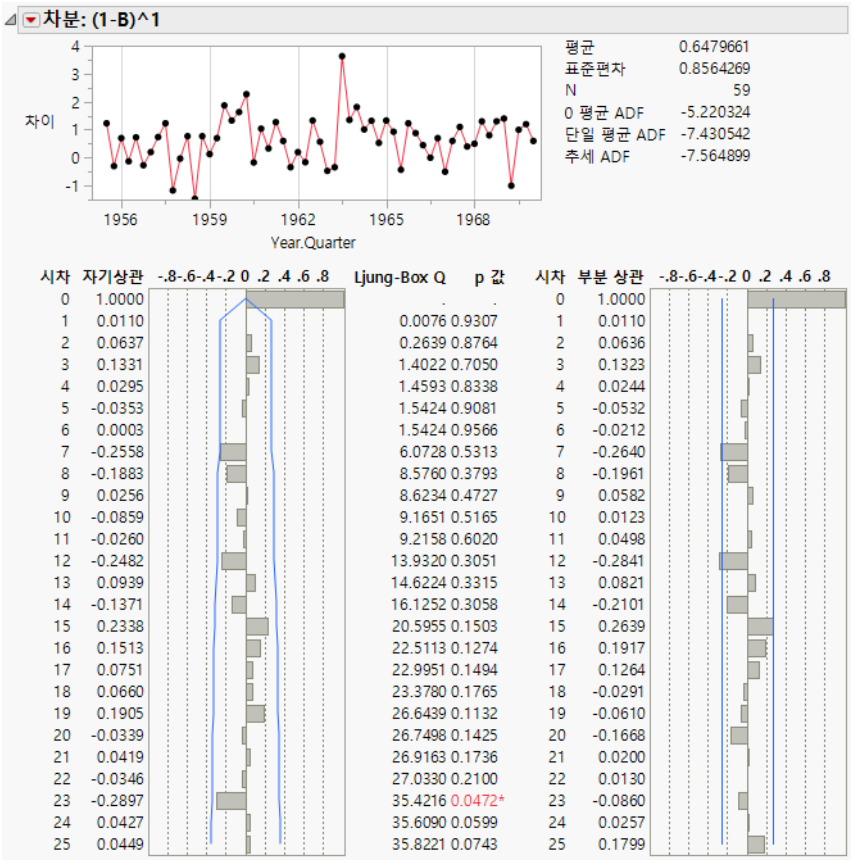
그림 18.17 SeriesP.jmp 에 대한 시계열 보고서



계열은 시간이 지남에 따라 선형에 가까운 증가 추세를 보입니다. 또한 자기상관 차트에서는 서로 가까운 점 사이에 강한 상관관계가 있음을 보여 줍니다 . 시차가 1, 2, 3 인 점의 자기상관 값은 각각 0.9551, 0.9112, 0.8666 입니다 .

5. " 시계열 GDP" 의 빨간색 삼각형을 클릭하고 **차분**을 선택합니다 .
6. " 비계절 차분 차수 " 에 대해 "1" 을 선택하고 **추정**을 클릭합니다 .

그림 18.18 SeriesP.jmp 에 대한 차분 보고서



" 차분 " 보고서는 원래 시계열에 적합시킬 적절한 모형을 결정하는 데 도움이 됩니다 . 차분 그림에서는 차분 계열에 원래 데이터에서 관측된 추세가 더 이상 없음을 보여 줍니다 . 이는 시차 -1 차분이 적절한 선택임을 나타냅니다 . 또한 추세를 제거한 후에도 계열에 계절성 징후가 보이지 않습니다 . 이러한 이유로 원래 계열을 적합시키는 모형은 선형 추세를 처리할 수 있어야 하지만 계절성을 반드시 처리할 필요는 없습니다. 선형 지수 평활 및 ARIMA 모형이 적절합니다 .

7. " 시계열 GDP" 의 빨간색 삼각형을 클릭하고 **평활 모형 > 선형 지수 평활**을 선택합니다 .
8. **추정**을 클릭합니다 .
9. " 시계열 GDP" 의 빨간색 삼각형을 클릭하고 **ARIMA 모형 그룹**을 선택합니다 . 이렇게 하면 (*p*,*d*,*q*)(*P*,*D*,*Q*) 의 값 범위에 대해 여러 ARIMA 모형을 적합시킬 수 있습니다 .
10. "ARIMA" 상자에서 다음 범위를 설정합니다 .
    - 차분 보고서에서 시차 -1 차분이 적절하다는 것을 확인했으므로 1 ~ 1 범위를 설정하여 "*d*, 차분 차수 " 를 1 로 수정합니다 .
    - 원래 계열에 자기상관 증거가 보였으므로 "*p*, 자기회귀 차수 " 를 0 ~ 1 범위로 설정합니다 .

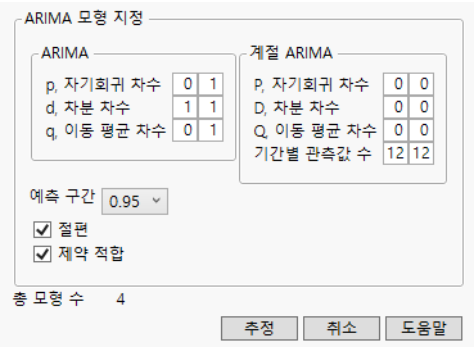
– "*q*, 이동 평균 차수 " 를 0 ~ 1 범위로 설정합니다 .

참고 : 대부분의 경우 *p* 와 *q* 를 작게 유지하면 충분합니다 .

– 계열에 계절성 증거가 없었으므로 *P, D, Q* 는 설정된 값 (0) 그대로 둡니다 .

이러한 설정을 통해 총 네 개의 모형이 적합됩니다 .

그림 18.19 ARIMA 모형 그룹 규격



11. **추정**을 클릭합니다 .

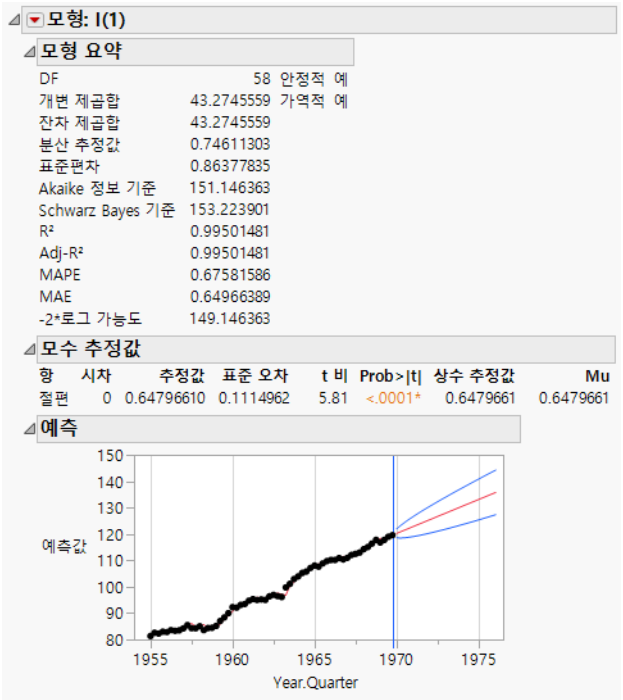
그림 18.20 모형 비교 테이블

모형 비교

보고서	그래프	모형	DF	분산	AIC	SBC	R^2	-2*로그 가능도	가중치	.2 .4 .6 .8	MAPE	MAE
☑	☐	I(1)	58	0.746113	151.14636	153.22390	0.995	149.14636	0.477865		0.675816	0.649664
☑	☐	ARI(1, 1)	57	0.7591096	153.13924	157.29432	0.995	149.13924	0.176424		0.676216	0.649960
☑	☐	IMA(1, 1)	57	0.7591199	153.14002	157.29510	0.995	149.14002	0.176355		0.676179	0.649933
☑	☐	ARIMA(1, 1, 1)	56	0.748633	153.91882	160.15143	0.995	147.91882	0.119474		0.668189	0.641981
☑	☐	선형 (Holt) 지수 평활	56	0.7878954	155.66571	159.78660	0.994	151.66571	0.049882		0.716452	0.686737

모형 비교 테이블은 AIC 기준에 따라 최량 적합 모형이 목록의 맨 위에 오도록 정렬됩니다. 이 경우 보고서에 "I(1)"로 표시된 ARIMA(0,1,0) 모형이 원래 시계열에 가장 적합합니다. 또한 I(1) 모형이 "최량" 모형이지만 모든 모형의 적합 통계량 값이 매우 비슷하다는 점도 유의해야 합니다. 모두 적절하다고 간주할 수 있습니다.

그림 18.21 ARIMA(0,1,0) 에 대한 모형 보고서



I(1) 모형 보고서에는 예측 그래프가 표시됩니다. 파란색 선은 예측 구간을 나타냅니다. GDP는 선형 비율로 계속 증가한다고 예측됩니다.

Box-Cox 변환을 사용한 예

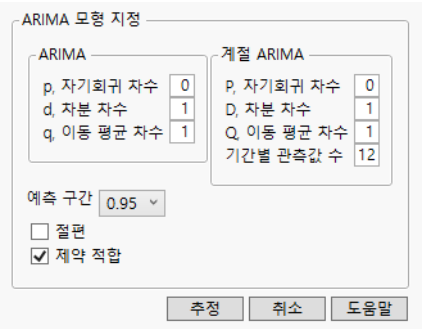
이 예에서는 Box-Cox 변환을 사용하여 Seriesg.jmp 데이터 테이블의 Passengers 열을 분석합니다. 람다 값을 0으로 설정하여 Box-Cox 변환을 근사 로그 변환으로 지정합니다. Seriesg.jmp 데이터 테이블의 스크립트는 Log Passengers 열에 대한 시계열 분석을 실행합니다. 분석에 Box-Cox 변환을 사용하면 모든 예측이 다시 변환되어 원래 척도로 보고되므로 결과를 더 쉽게 해석할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Time Series/Seriesg.jmp 를 엽니다.
2. **분석 > 전문 모델링 > 시계열**을 선택합니다.
3. Passengers 를 선택하고 **Y, 시계열**을 클릭합니다.
4. Time 을 선택하고 **X, 시간 ID** 를 클릭합니다.
5. **Box-Cox 변환 사용**을 선택합니다.
   람다에 기본값 0 을 사용합니다.
6. **확인**을 클릭합니다.

7. " 시계열 Box-Cox 변환 Passengers" 옆의 빨간색 삼각형을 클릭하고 **계절 ARIMA** 를 선택합니다 .
8. " 계절 ARIMA 규격 " 창의 "ARIMA" 섹션에서 다음 모수를 설정합니다 .
   - "*p*, 자기회귀 차수 " 를 "0" 으로 설정합니다 .
   - "*d*, 차분 차수 " 를 "1" 로 설정합니다 .
   - "*q*, 이동 평균 차수 " 를 "1" 로 설정합니다 .
9. " 계절 ARIMA 규격 " 창의 " 계절 ARIMA" 섹션에서 다음 모수를 설정합니다 .
   - "*P*, 자기회귀 차수 " 를 "0" 으로 설정합니다 .
   - "*D*, 차분 차수 " 를 "1" 로 설정합니다 .
   - "*Q*, 이동 평균 차수 " 를 "1" 로 설정합니다 .
10. " 절편 " 을 선택 취소하여 절편이 없는 모형을 실행합니다 .

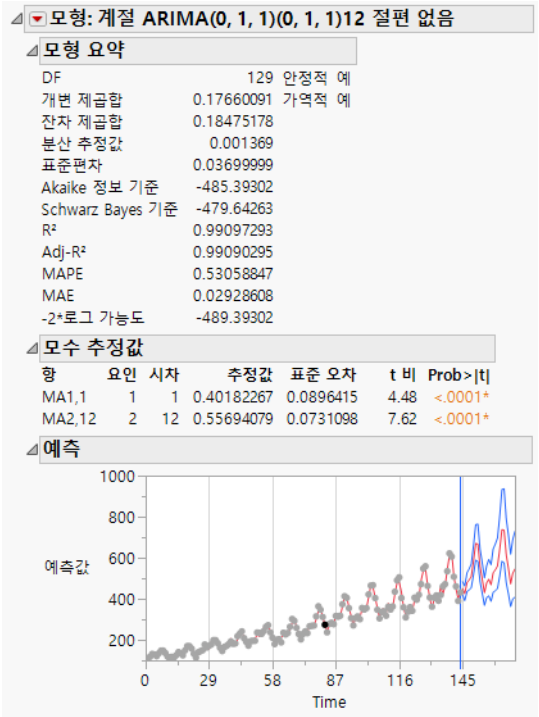
    현재 창이 그림 18.22 에 표시된 창과 일치해야 합니다 .

그림 18.22 계절 ARIMA 규격



11. **추정**을 클릭합니다 .

그림 18.23 Box-Cox 변환 Passengers 데이터에 대한 계절 ARIMA 모형



모수 추정값은 Log Passengers 열에 대해 동일한 모형을 실행하여 얻은 값과 동일합니다. 그러나 그림 18.23 의 예측 및 예측 구간은 원래 척도로 표시됩니다.

홀드백 집합을 사용한 예

이 예에서는 홀드백 집합을 사용하여 월별 납 생산 데이터에 대한 예측을 평가합니다. 홀드백 집합이 사용되므로 상태 공간 평활 모형과 ARIMA 모형을 모두 적합시키고 비교하여 데이터에 대한 최량 적합 모형을 찾을 수 있습니다.

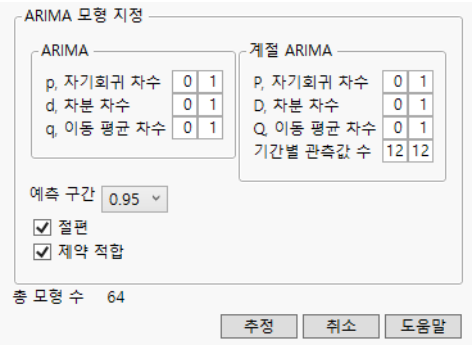
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Time Series/Lead Production.jmp 를 엽니다.
2. **분석 > 전문 모델링 > 시계열**을 선택합니다.
3. LEADPROD 를 선택하고 **Y, 시계열**을 클릭합니다.
4. DATE 를 선택하고 **X, 시간 ID** 를 클릭합니다.
5. **홀드백에 대한 예측**을 선택합니다.

   이 옵션을 선택하면 미래 관측값 대신 홀드백 집합에 대한 예측이 실행됩니다.
6. " 예측 주기 " 옆에 "12" 를 입력합니다.

   이렇게 하면 데이터의 마지막 12 개 관측값이 홀드백 집합에 할당됩니다. 월별 데이터이므로 이 값은 작년에 해당합니다.

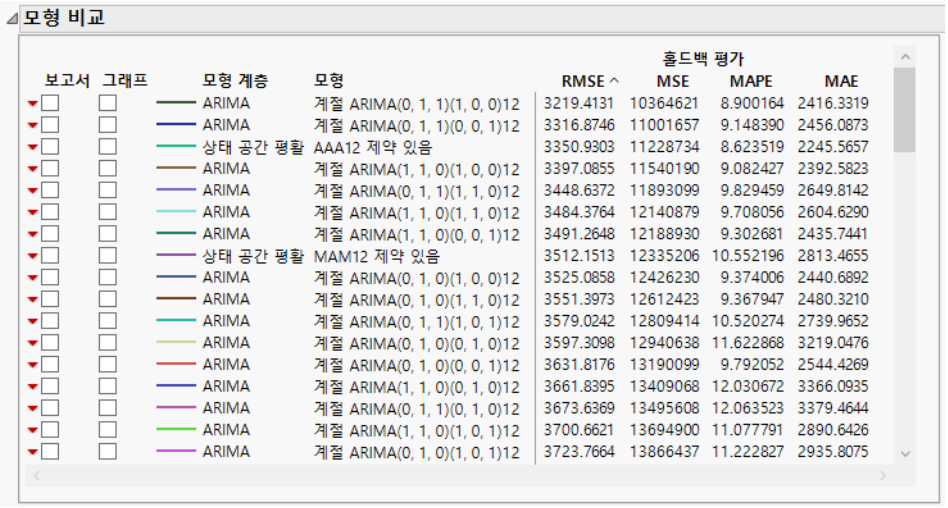
7. **확인**을 클릭합니다 .
8. " 시계열 LEADPROD" 의 빨간색 삼각형을 클릭하고 **상태 공간 평활**을 선택합니다 .
9. " 주기 " 옆에 "12" 를 입력합니다 .
10. " 모수 제한 " 을 선택합니다 .
11. **확인**을 클릭합니다 .
12. " 시계열 LEADPROD" 의 빨간색 삼각형을 클릭하고 **ARIMA 모형 그룹**을 선택합니다 .
13. 모든 모수의 범위를 0 ~ 1 로 설정합니다 . "ARIMA 모형 지정 " 창은 그림 18.24 에 나오는 것과 같습니다 .

그림 18.24 ARIMA 모형 그룹 규격



14. **추정**을 클릭합니다 .

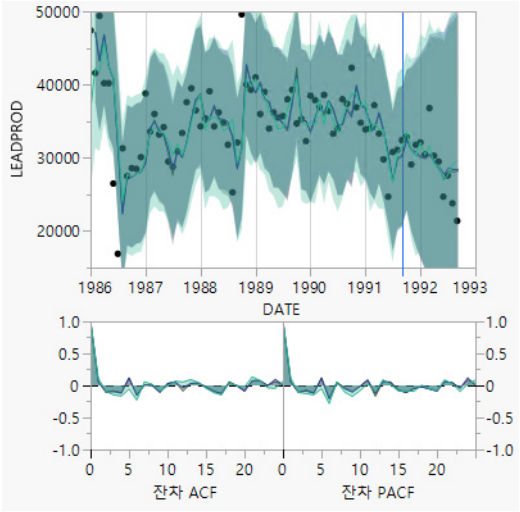
그림 18.25 모형 비교



상태 공간 평활 모형과 ARIMA 모형이 홀드백 집합에 대해 평가되므로 두 모형의 적합 통계량을 비교할 수 있습니다 .

15. " 모형 비교 " 보고서에서 처음 세 모형 옆의 " 그래프 " 체크박스를 선택합니다 .

그림 18.26 모형 비교 그래프



평가 통계량과 예측 그림을 보면 상위 모형 간의 예측 성능에 큰 차이가 없음을 알 수 있습니다 . 보고서에 따르면 ARIMA 모형 또는 상태 공간 평활 모형이 이 데이터에 적절합니다 .

시계열 플랫폼에 대한 통계 상세 정보

- " 스펙트럼 밀도에 대한 통계 상세 정보 "
- "X-11 분해에 대한 통계 상세 정보 "
- " 지수 평활 모형에 대한 통계 상세 정보 "
- "ARIMA 모형에 대한 통계 상세 정보 "
- " 전이 함수에 대한 통계 상세 정보 "

스펙트럼 밀도에 대한 통계 상세 정보

Fourier 계열의 계수에 대한 최소 제곱 추정값은 다음과 같이 계산됩니다 .

$$a_t = \frac{2}{N}\sum_{i=1}^{N} y_t \cos(2\pi f_i t)$$

$$b_t = \frac{2}{N}\sum_{i=1}^{N} y_t \sin(2\pi f_i t)$$

그런 다음 $f_i = i/N$가 결합되어 빈도 f_i에서의 강도를 나타내는 주기도 $I(f_i) = (N/2)(a_i^2 + b_i^2)$를 구성합니다.

그런 다음 주기도가 평활되고 $1/(4\pi)$로 척도가 조정되어 스펙트럼 밀도를 구성합니다.

X-11 분해에 대한 통계 상세 정보

이 방법은 승법 또는 가법 분해를 사용하여 원래 시계열을 조정합니다. 세 가지 X-11 성분(추세 순환, 계절, 불규칙)을 추정하기 위해 반복 프로세스를 사용하여 모형이 적합됩니다. 추세 순환 성분은 장기 추세 및 장기 순환 효과를 모두 포함합니다. 불규칙 성분은 추세 성분과 계절 성분으로 설명되지 않는 변동 효과를 포함합니다. 과거 X-11 방법 개발에 대한 개요를 보려면 SAS/ETS 15.2 User's Guide("Historical Development of X-11" 검색)를 참조하십시오.

승법 조정은 다음 모형을 적합시킵니다.

$$O_t = C_t \cdot S_t \cdot I_t$$

다음은 각 요소에 대한 설명입니다.

O_t = 원래 시계열

C_t = 추세 순환 성분

S_t = 계절 성분

I_t = 불규칙 성분

조정 승법 추세는 O_t/S_t입니다.

가법 조정은 다음 모형을 적합시킵니다.

$$O_t = C_t + S_t + I_t$$

조정 가법 추세는 $O_t - S_t$입니다.

지수 평활 모형에 대한 통계 상세 정보

평활 모형은 다음과 같이 정의됩니다.

$$y_t = \mu_t + \beta_t t + s(t) + a_t$$

다음은 각 요소에 대한 설명입니다 .

μ_t = 시간에 따라 변하는 평균 항

β_t = 시간에 따라 변하는 기울기 항

$s(t)$ = 시간에 따라 변하는 계절 항 중 하나

a_t = 확률 충격

추세가 없는 모형은 $\beta_t = 0$ 이고 비계절 모형은 $s(t) = 0$ 입니다 . 시간에 따라 변하는 이러한 항의 추정량은 다음과 같이 정의됩니다 .

L_t = μ_t 를 추정하는 평활 수준

T_t = β_t 를 추정하는 평활 추세

$S_{t-j}(j = 0, 1,..., s - 1) = s(t)$ 의 추정값

각 평활 모형은 이러한 추정량의 진화를 설명하는 재귀 평활 방정식 집합을 정의합니다 . 평활 방정식은 **평활 가중치**라고 하는 다음 모형 모수로 작성됩니다 .

α = 수준 평활 가중치

γ = 추세 평활 가중치

φ = 진폭감소 추세 가중치

δ = 계절 평활 가중치

이러한 모수는 각 모형에 서로 다른 방식으로 입력(또는 전혀 입력되지 않음)되지만 가중치가 클수록 최근 데이터에 영향을 많이 주고 가중치가 작을수록 최근 데이터에 영향을 적게 준다는 공통된 특성이 있습니다. 평활 가중치에 대한 자세한 내용은 SAS/ETS 15.2 User's Guide("Forecasting Process Details" 장)에서 확인하십시오.

단순 지수 평활

단순 지수 평활 모형은 $y_t = \mu_t + \alpha_t$ 입니다.

평활 방정식 $L_t = \alpha y_t + (1 - \alpha)L_{t-1}$ 은 단일 평활 가중치 α 로 정의됩니다 . 이 모형은 다음 조건이 참일 때 ARIMA(0, 1, 1) 모형과 동일합니다 .

$$(1 - B)y_t = (1 - \theta B)\alpha_t \quad \text{여기서 , } \theta = 1 - \alpha$$

모형의 이동 평균 형식은 다음과 같이 정의됩니다 .

$$y_t = a_t + \sum_{j-1}^{\infty} \alpha a_{t-j}$$

이중 (Brown) 지수 평활

이중 지수 평활 모형은 $y_t = \mu_t + \beta_1 t + a_t$입니다.

평활 방정식은 단일 평활 가중치 α 를 사용하여 다음과 같이 정의됩니다 .

$$L_t = \alpha y_t + (1-\alpha)L_{t-1}$$

$$T_t = \alpha(L_t - L_{t-1}) + (1-\alpha)T_{t-1}$$

이 모형은 다음 조건이 참일 때 ARIMA(0, 1, 1)(0, 1, 1)$_1$ 모형과 동일합니다 .

$$(1-\mathrm{B})^2 y_t = (1-\theta\mathrm{B})^2 a_t \quad \text{여기서} \quad \theta_{1,1} = \theta_{2,1} \ , \ \theta = 1-\alpha$$

모형의 이동 평균 형식은 다음과 같이 정의됩니다 .

$$y_t = a_t + \sum_{j=1}^{\infty} (2\alpha + (j-1)\alpha^2) a_{t-j}$$

선형 (Holt) 지수 평활

선형 지수 평활 모형은 $y_t = \mu_t + \beta_t t + a_t$입니다.

평활 방정식은 α 및 γ 평활 가중치를 사용하여 다음과 같이 정의됩니다 .

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

이 모형은 다음 조건이 참일 때 ARIMA(0, 2, 2) 모형과 동일합니다 .

$$(1-\mathrm{B})^2 y_t = (1-\theta\mathrm{B}-\theta_2\mathrm{B}^2) a_t \quad \text{여기서} \quad \theta = 2-\alpha-\alpha\gamma \ , \ \theta_2 = \alpha - 1$$

모형의 이동 평균 형식은 다음과 같이 정의됩니다 .

$$y_t = a_t + \sum_{j=1}^{\infty} (\alpha + j\alpha\gamma) a_{t-j}$$

진폭감소 추세 선형 지수 평활

진폭감소 추세 선형 지수 평활 모형은 $y_t = \mu_t + \beta_t t + a_t$입니다.

평활 방정식은 α, γ 및 φ 평활 가중치를 사용하여 다음과 같이 정의됩니다 .

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + \varphi T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)\varphi T_{t-1}$$

이 모형은 다음 조건이 참일 때 ARIMA(1, 1, 2) 모형과 동일합니다 .

$$(1-\varphi B)(1-B)y_t = (1-\theta_1 B - \theta_2 B^2)a_t$$

다음은 각 요소에 대한 설명입니다 .

$$\theta_1 = 1 + \varphi - \alpha - \alpha\gamma\varphi$$

$$\theta_2 = (\alpha - 1)\varphi$$

모형의 이동 평균 형식은 다음과 같이 정의됩니다 .

$$y_t = \alpha_t + \sum_{j=1}^{\infty}\left(\frac{\alpha + \alpha\gamma\varphi(\varphi^j - 1)}{\varphi - 1}\right)\alpha_{t-j}$$

계절 지수 평활

계절 지수 평활 모형은 $y_t = \mu_t + s(t) + a_t$입니다.

평활 방정식은 α 및 δ 평활 가중치를 사용하여 다음과 같이 정의됩니다 .

$$L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)L_{t-1}$$

$$S_t = \delta(y_t - L_{t-s}) + (1-\delta)S_{t-s}$$

이 모형은 계절 ARIMA$(0, 1, s+1)(0, 1, 0)_s$ 모형과 동일합니다 .

$$(1-B)(1-B^s)y_t = (1-\theta_1 B - \theta_2 B^2 - \theta_3 B^{s+1})a_t$$

다음은 각 요소에 대한 설명입니다 .

$$\theta_1 = 1 - \alpha$$

$$\theta_2 = (1-\delta)(1-\alpha)$$

$$\theta_3 = (1-\alpha)(\delta-1)$$

모형의 이동 평균 형식은 다음과 같이 정의됩니다 .

$$y_t = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \quad \text{여기서 , } \psi = \begin{cases} \alpha \text{ 단, } j \bmod s \neq 0 \\ \alpha + \delta(1-\alpha) \text{ 단, } j \bmod s = 0 \end{cases}$$

Winters 방법 (가법 모형)

Winters 방법의 가법 버전 모형은 $y_t = \mu_t + \beta_t t + s(t) + a_t$입니다 .

평활 방정식은 α, γ 및 δ 가중치를 사용하여 다음과 같이 정의됩니다 .

$$L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

$$S_t = \delta(y_t - L_t) + (1-\delta)S_{t-s}$$

이 모형은 다음과 같이 정의된 계절 ARIMA(0, 1, *s*+1)(0, 1, 0)s 모형과 동일합니다 .

$$(1-B)(1-B^2)y_t = \left(1 - \sum_{i=1}^{s+1} \theta_i B^i\right) a_t$$

모형의 이동 평균 형식은 다음과 같이 정의됩니다 .

$$y_t = a_t + \sum_{j=1}^{\infty} \Psi_j a_{t-j}$$

여기서 ,

$$\psi = \begin{cases} \alpha + j\alpha\gamma , & j \bmod s \neq 0 \\ \alpha + j\alpha\gamma + \delta(1-\alpha) , & j \bmod s = 0 \end{cases}$$

ARIMA 모형에 대한 통계 상세 정보

ARIMA 모형

반응 계열 $\{y_t\}$ 에 대한 ARIMA 모형의 일반 형식은 다음과 같이 정의됩니다 .

$$\phi(B)(w_t - \mu) = \theta(B)a_t$$

다음은 각 요소에 대한 설명입니다 .

t = 시간 인덱스

B = $By_t = y_{t-1}$ 로 정의된 후방 연산자

$w_t = (1 - B)^d y_t$ = 차분 후의 반응 계열

μ = 절편 또는 평균 항

$\phi(B)$ 및 $\theta(B)$ 는 각각 자기회귀 연산자와 이동 평균 연산자이며 다음과 같이 정의됩니다 .

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$$

다음은 각 요소에 대한 설명입니다 .

a_t = 확률 충격 시퀀스

a_t 는 평균이 0 이고 분산이 일정한 독립 정규 분포를 따른다고 가정합니다 .

모형을 다음과 같이 다시 작성할 수 있습니다 .

$$\phi(B)w_t = \delta + \theta(B)a_t$$

상수 추정값 δ 는 다음과 같은 관계식으로 제공됩니다 .

$$\delta = \phi(B)\mu = \mu - \phi_1\mu - \phi_2\mu - \ldots - \phi_p\mu$$

계절 ARIMA 모형

계절 ARIMA 모델링의 경우 차분 , 자기회귀 및 이동 평균 연산자는 계절 다항식과 비계절 다항식의 곱입니다 .

$$w_t = (1 - B)^d (1 - B^s)^D y_t$$

$$\varphi(B) = (1 - \varphi_{1,1}B - \varphi_{1,2}B^2 - \ldots - \varphi_{1,p}B^p)(1 - \varphi_{2,s}B^s - \varphi_{2,2s}B^{2s} - \ldots - \varphi_{2,Ps}B^{Ps})$$

$$\theta(B) = (1 - \theta_{1,1}B - \theta_{1,2}B^2 - \ldots - \theta_{1,q}B^q)(1 - \theta_{2,s}B^s - \theta_{2,2s}B^{2s} - \ldots - \theta_{2,Qs}B^{Qs})$$

여기서 s는 기간별 관측값 수입니다. 계수의 첫 번째 인덱스는 요인 번호(1 - 비계절, 2 - 계절)이고 두 번째 인덱스는 항의 시차입니다.

전이 함수에 대한 통계 상세 정보

m 개의 입력이 있는 일반적인 전이 함수 모형은 다음과 같이 나타낼 수 있습니다.

$$Y_t - \mu = \frac{\omega_1(B)}{\delta_1(B)}X_{1,t-d1} + \ldots + \frac{\omega_m(B)}{\delta_m(B)}X_{m,m-dm} + \frac{\theta(B)}{\varphi(B)}e_t$$

다음은 각 요소에 대한 설명입니다.

Y_t = 출력 계열

$X_1 \sim X_m = m$ 개의 입력 계열

e_t = 잡음 계열

$X_{1,t-d1} = X_1$ 계열이 $d1$ 단계 시차의 t 에 의해 인덱싱됨

μ = 모형의 평균 수준

$\varphi(B)$ 및 $\theta(B)$ = ARIMA 모형의 자기회귀 다항식 및 이동 평균 다항식

$\omega_k(B)$ 및 $\delta_k(B)$ = 개별 전이 함수의 분자 및 분모 요인(다항식). 여기서, k는 $1 \sim m$ 개의 개별 입력에 대한 인덱스를 나타냅니다.

위 모형의 각 다항식은 계절 ARIMA에서와 같이 두 부분(비계절과 계절 또는 두 성분의 곱)을 포함할 수 있습니다. 모형을 지정할 때 모형에 포함하지 않을 부분에는 기본값 0을 그대로 둡니다.

19 장

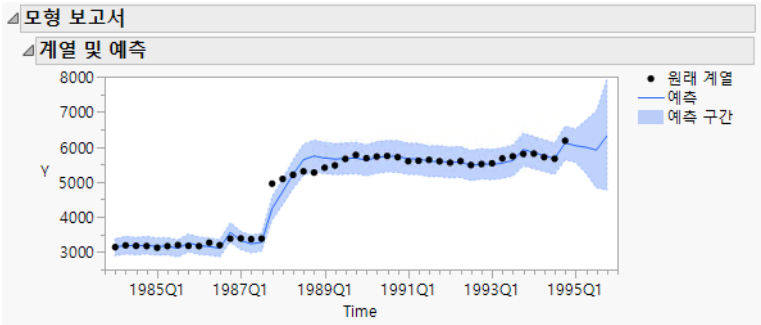
시계열 예측

다중 시계열 예측

시계열 예측 플랫폼을 사용하면 계열 수가 매우 많은 다중 시계열을 예측할 수 있습니다. 이 플랫폼에서는 다중 시계열에 대한 간단한 요약을 표시하고, 지수 평활족에서 최대 30 개의 서로 다른 예측 모형을 효율적으로 적합시킬 수 있는 옵션을 제공합니다. 최량 적합 모형이 자동으로 선택됩니다.

시계열 예측 플랫폼은 지수 평활 모형 집합에서 최량 적합 모형을 자동으로 선택하기 위해 만들어졌습니다. 이는 최량 적합 모형을 수동으로 선택해야 하고 ARIMA 모형을 사용하는 시계열 플랫폼과 대조적입니다.

그림 19.1 원래 계열 및 예측



목차

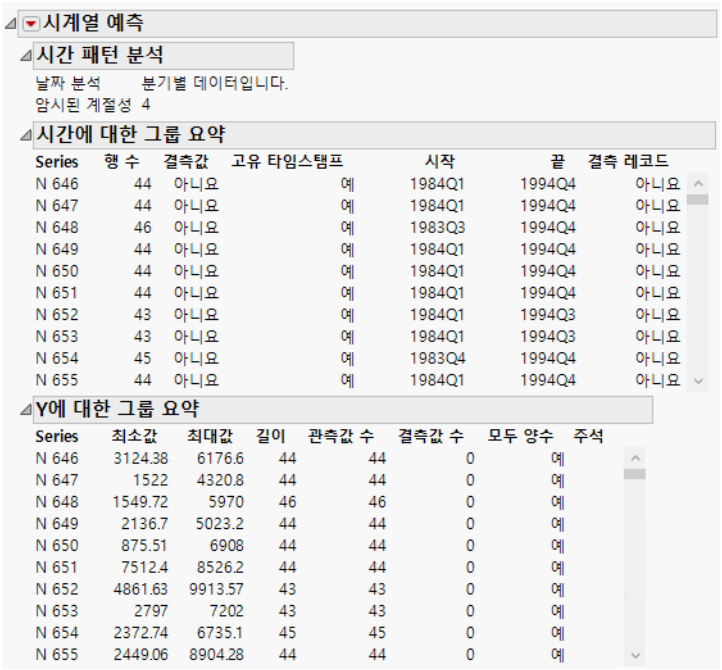
시계열 예측의 예 351
시계열 예측 플랫폼 시작 353
데이터 형식 353
시계열 예측 보고서 354
시간 패턴 분석 354
시간에 대한 그룹 요약 355
Y에 대한 그룹 요약 355
모델링 규격 356
시계열 예측 플랫폼 옵션 358
모형 보고서 359

시계열 예측의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Time Series/M3C Quarterly.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 시계열 예측**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Series 를 선택하고 **그룹화**를 클릭합니다 .
5. Time 을 선택하고 **시간**을 클릭합니다 .
6. **확인**을 클릭합니다 .

   초기 " 시계열 예측 " 보고서가 나타납니다 . 여기에는 시간 및 Y 변수에 대한 요약이 포함되어 있습니다 . 예측 모형 적합 옵션을 제공하는 " 모델링 규격 " 보고서도 있습니다 .

그림 19.2 시계열 예측의 초기 보고서



시계열 예측

시간 패턴 분석

날짜 분석 분기별 데이터입니다.
암시된 계절성 4

시간에 대한 그룹 요약

Series	행 수	결측값	고유 타임스탬프	시작	끝	결측 레코드
N 646	44	아니요	예	1984Q1	1994Q4	아니요
N 647	44	아니요	예	1984Q1	1994Q4	아니요
N 648	46	아니요	예	1983Q3	1994Q4	아니요
N 649	44	아니요	예	1984Q1	1994Q4	아니요
N 650	44	아니요	예	1984Q1	1994Q4	아니요
N 651	44	아니요	예	1984Q1	1994Q4	아니요
N 652	43	아니요	예	1984Q1	1994Q3	아니요
N 653	43	아니요	예	1984Q1	1994Q3	아니요
N 654	45	아니요	예	1983Q4	1994Q4	아니요
N 655	44	아니요	예	1984Q1	1994Q4	아니요

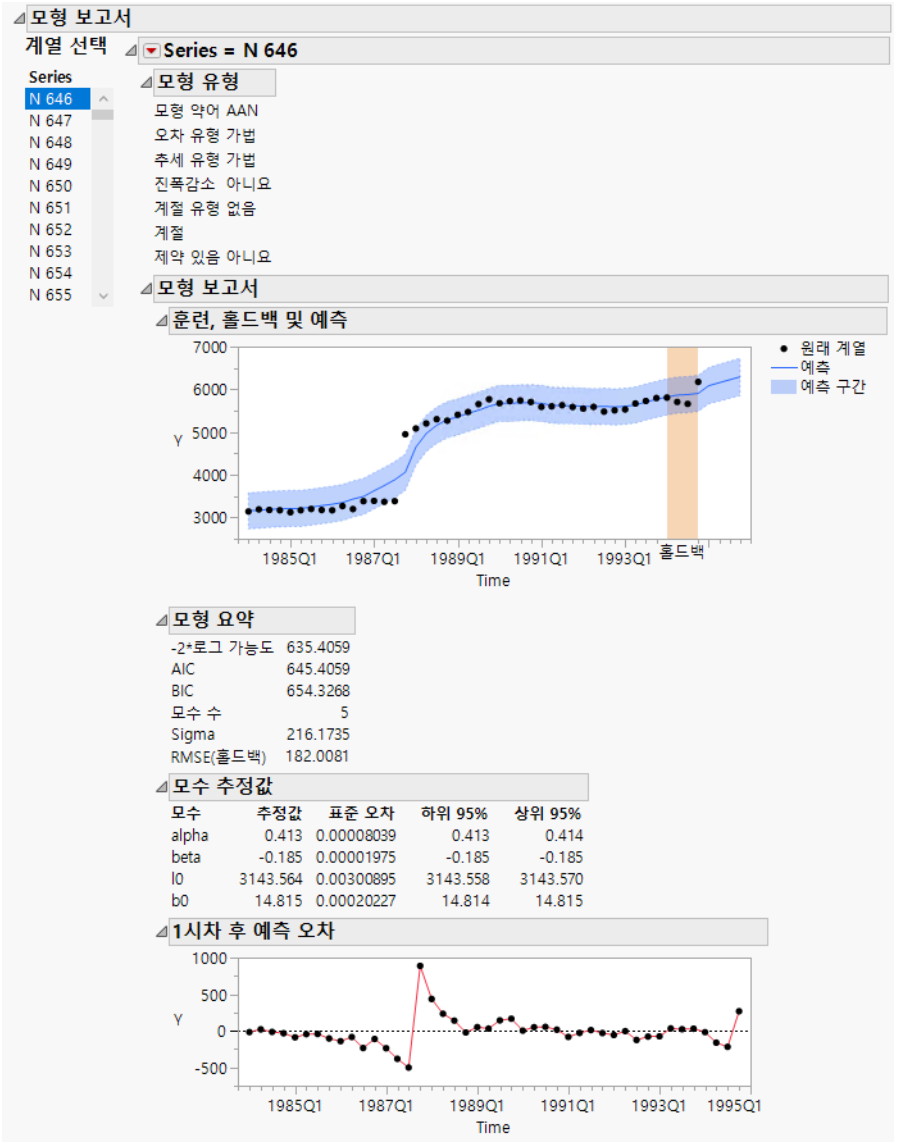
Y에 대한 그룹 요약

Series	최소값	최대값	길이	관측값 수	결측값 수	모두 양수	주석
N 646	3124.38	6176.6	44	44	0	예	
N 647	1522	4320.8	44	44	0	예	
N 648	1549.72	5970	46	46	0	예	
N 649	2136.7	5023.2	44	44	0	예	
N 650	875.51	6908	44	44	0	예	
N 651	7512.4	8526.2	44	44	0	예	
N 652	4861.63	9913.57	43	43	0	예	
N 653	2797	7202	43	43	0	예	
N 654	2372.74	6735.1	45	45	0	예	
N 655	2449.06	8904.28	44	44	0	예	

7. " 모델링 규격 " 보고서에서 " 완전한 규격 " 탭을 클릭합니다 .
8. " 모형 선택 전략 " 아래의 메뉴를 클릭하고 " 예측 성능 " 을 선택합니다 .
9. **실행**을 클릭합니다 .

   이렇게 하면 권장 모형 집합을 데이터에 적합시킵니다 . 그룹화 변수의 각 수준에 대한 최적 모형이 선택되고 " 모형 보고서 " 에 보고됩니다 . 자세한 내용은 " 모형 보고서 " 에서 확인하십시오 .

그림 19.3 모형 보고서

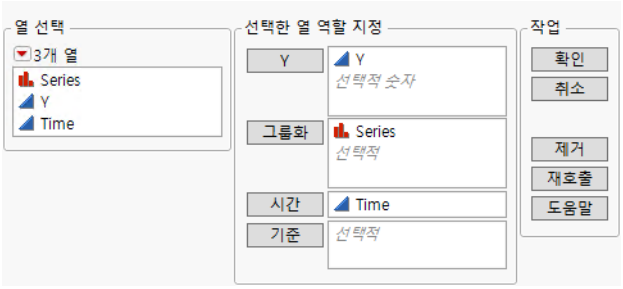


N646 계열에 대한 모형 보고서가 표시됩니다 . 이것은 " 계열 선택 " 목록에서 선택한 계열 이름에 해당합니다 . 아래쪽 화살표 키를 사용하여 각 계열에 대한 모형 보고서를 스캔할 수 있습니다 . 또는 목록에서 다른 이름을 클릭하여 선택한 계열에 대한 모형 보고서를 표시할 수도 있습니다 .

시계열 예측 플랫폼 시작

분석 > 전문 모델링 > 시계열 예측을 선택하여 시계열 예측 플랫폼을 시작합니다 .

그림 19.4 시계열 예측 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

시계열 예측 플랫폼 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y 하나 이상의 열을 시계열 변수로 할당합니다 . 다중 시계열을 테이블의 개별 열로 구성하거나 , 하나 이상의 그룹화 변수를 사용하여 단일 열에 쌓을 수 있습니다 .

그룹화 하나 이상의 열을 그룹화 변수로 할당합니다 . 그룹화 변수는 쌓인 형식의 개별 시계열을 식별하는 데 사용됩니다 .

시간 시간 변수에 대한 열을 할당합니다 . 개별 시계열이 개별 열로 구성된 경우 시간 열의 타임스탬프에 중복 항목이 포함되지 않아야 합니다 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 할당합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

데이터 형식

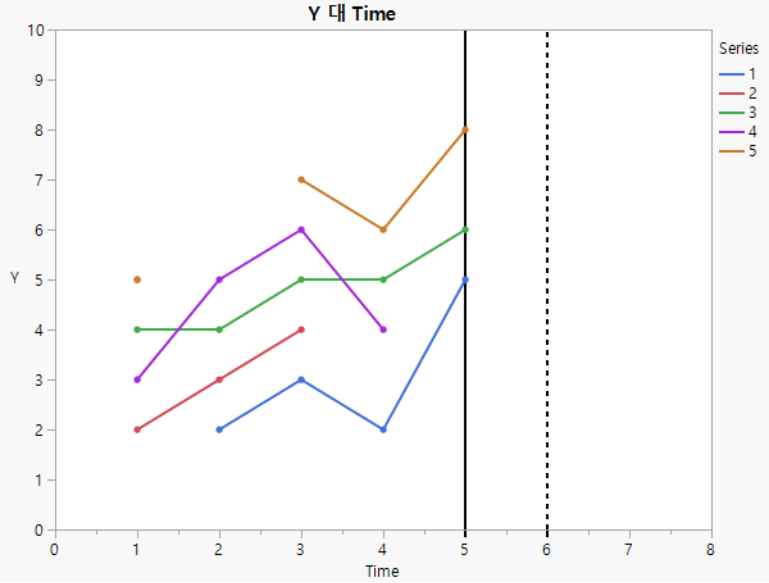
다중 시계열을 테이블의 개별 열로 구성하거나 , 개별 계열을 식별하기 위한 하나 이상의 그룹화 변수를 사용하여 단일 열에 쌓을 수 있습니다 . 여러 Y 열과 여러 그룹화 열을 조합할 수도 있습니다 .

이 플랫폼에서는 데이터에 대해 다음과 같이 가정합니다 .

- 모든 시계열은 동일한 시간 해상도로 기록됩니다 .
- 모든 시계열은 동일한 최종 관측 시점까지 수집됩니다 . 이 최종 관측 시점은 개별 시계열에서 발생하는 맨 마지막 시점으로 정의됩니다 . 최종 관측 시점까지 데이터가 없는 계열은 결측값이 있는 것으로 가정합니다 . 예를 들어 그림 19.5 에서 최종 관측 시점은 검은색 실선으로 표시된 5 입니다 . 계열 2 는 시점 4 와 5 에 결측 관측값이 있고 계열 4 는 시점 5 에 결측 관측값이 있다고 간주됩니다 .

- 모든 예측은 동일한 시점에서 시작합니다. 이 시작 시점은 최종 관측 시점 이후의 다음 시점입니다. 예를 들어 그림 19.5 에서 최종 관측 시점은 검은색 실선으로 표시된 5 입니다. 따라서 모든 예측은 검은색 파선으로 표시된 시점 6 에서 시작합니다.

그림 19.5 가정 데모



시계열 예측 보고서

"시계열 예측" 보고서에는 시작 창에 지정된 변수를 기반으로 한 요약 테이블이 포함됩니다. 그룹별 시계열 요약 테이블이 항상 제공됩니다. 시간 열이 지정된 경우에는 시간 패턴 요약 및 그룹별 시간 요약 테이블도 포함됩니다. 이 보고서에는 각 시계열에 모형을 적합시킬 수 있는 "모델링 규격" 보고서도 포함되어 있습니다.

- "시간 패턴 분석"
- "시간에 대한 그룹 요약"
- "Y 에 대한 그룹 요약"
- "모델링 규격"

시간 패턴 분석

시간 열이 지정되면 시계열 예측 플랫폼에서 열 항목을 분석하여 패턴을 결정합니다. 이 플랫폼은 일별, 영업일별, 월별, 격주별, 분기별 등 다양한 시간 패턴을 인식하도록 되어 있습니다. 패턴은 "날짜 분석"으로 보고됩니다. 시간 패턴에 따라 "암시된 계절성"도 제공됩니다.

시간에 대한 그룹 요약

시간 열이 지정되면 보고서에 개별 시계열에 대한 요약이 포함됩니다 . 요약에는 각 시계열에 대한 다음 정보가 포함됩니다 .

< 그룹화 열 이름 > (그룹화 변수를 지정한 경우에만 사용 가능) 데이터가 쌓인 형식으로 제공될 때 개별 계열을 정의하는 그룹화 변수의 수준입니다 .

행 수 계열의 관측값 수입니다 .

결측값 시간 값에 결측값이 있는지 여부를 나타냅니다 .

고유 타임스탬프 타임스탬프가 고유한지 여부를 나타냅니다 .

시작 시작 타임스탬프입니다 .

끝 종료 타임스탬프입니다 .

결측 레코드 시작 및 종료 타임스탬프와 데이터 해상도 / 패턴을 기반으로 데이터 테이블에 결측 관측값이 있는지 여부를 나타냅니다 .

Y 에 대한 그룹 요약

보고서에 각 Y 변수에 대한 요약 테이블이 포함됩니다 . 요약에는 각 시계열에 대한 다음 정보가 포함됩니다 .

변수 (여러 개의 Y 열과 하나 이상의 그룹화 열이 있는 경우에만 사용 가능) 시계열 변수의 열 이름입니다 .

시계열 (계열이 여러 개 있는 경우에만 사용 가능) 개별 시계열의 이름입니다 . 데이터가 쌓인 형식으로 제공된 경우 이 값은 그룹화 변수의 수준입니다 . 데이터가 여러 개의 열과 함께 가로형 형식으로 제공된 경우에는 개별 계열의 열 이름입니다 .

참고 : 데이터가 쌓인 형식으로 제공되는 경우 테이블의 열 머리글은 그룹화 열의 이름입니다 .

최소값 개별 계열에 대한 Y 의 최소값입니다 .

최대값 개별 계열에 대한 Y 의 최대값입니다 .

길이 계열 기간입니다 . 이 값은 시작 및 종료 타임스탬프와 시간 패턴 분석 결과를 기반으로 결정됩니다 .

관측값 수 관측값 수입니다 . 결측 레코드가 없으면 " 길이 " 값과 같습니다 .

결측값 수 결측 Y 값 수입니다 .

모두 양수 모든 Y 값이 양수인지 여부를 나타냅니다 .

주석 계열에 문제 (예 : 고유하지 않은 타임스탬프) 가 있는지 여부를 나타냅니다 .

모델링 규격

"모델링 규격" 보고서를 사용하면 모형을 개별 시계열에 적합시킬 수 있습니다. 다양한 모형이 고려됩니다. Hyndman et al. (2008) 연구 자료에서는 오차, 추세 성분 및 계절 성분을 기반으로 상태 공간 평활 모형을 정의합니다.

- 오차는 가법 (A) 또는 승법 (M) 일 수 있습니다.
- 추세 성분은 없음 (N), 가법 (A), 진폭감소 가법 (A_d), 승법 (M) 또는 진폭감소 승법 (M_d) 일 수 있습니다.
- 계절 성분은 없음 (N), 가법 (A) 또는 승법 (M) 일 수 있습니다.

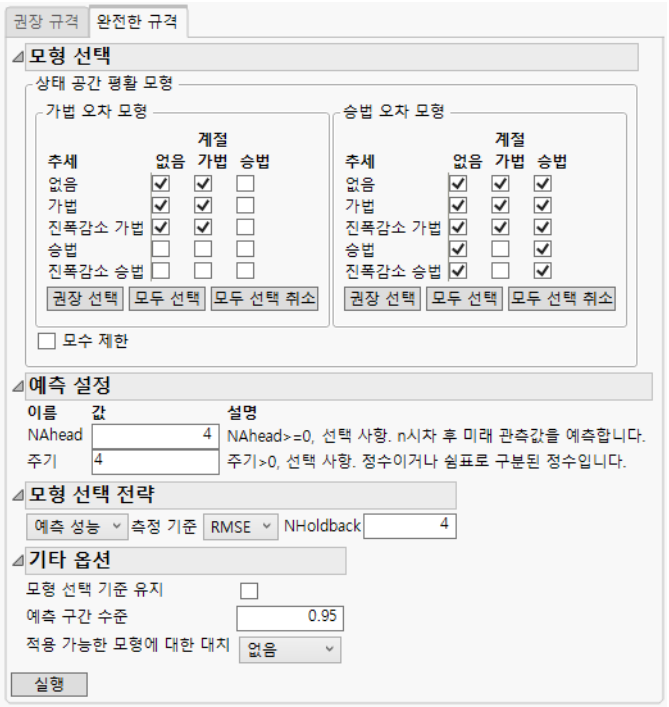
특정 모형은 ETS (오차, 추세, 계절)로 나타낼 수 있습니다. 이러한 모형이 시계열 예측에 사용됩니다.

"모델링 규격" 보고서에는 두 개의 탭이 있습니다. 플랫폼에서 선택한 상태 공간 평활 모형 집합을 적합시키려면 "권장 규격" 탭을 사용합니다. 적합시킬 특정 상태 공간 평활 모형을 선택하려면 "완전한 규격" 탭을 사용합니다. 그런 다음 각 개별 시계열에 대해 주어진 집합의 최량 적합 모형이 예측에 사용됩니다.

모형 선택

"완전한 규격 탭"에 표시된 "완전한 규격" 탭에서만 사용할 수 있습니다. "모형 선택" 보고서를 사용하면 개별 시계열에 적합시킬 상태 공간 평활 모형을 지정할 수 있습니다. 체크박스를 사용하여 모형에 대한 오차, 추세 및 계절성을 선택합니다. 플랫폼에서 권장하는 모형에 해당하는 체크박스를 선택하려면 **권장 선택**을 클릭합니다. **모두 선택**을 클릭하여 모든 체크박스를 선택하거나 **모두 선택 취소**를 클릭하여 모든 체크박스를 선택 취소할 수 있습니다. 관측값이 현재에서 멀어질수록 모형의 현재 상태에 미치는 영향이 적어지도록 모수를 제한하려면 **모수 제한**을 클릭합니다. 상태 공간 평활 모형에서 모든 이전 관측값이 주어진 경우 시간 t에서의 예측은 시간 t까지의 모든 관측값의 가중합과 동일합니다. 가중치는 모수의 함수입니다. 따라서 모수를 제한하면 과거 관측값에 대한 가중치가 0이 되고 관측값이 현재에서 멀어질수록 가중치가 더 빨리 0이 됩니다.

그림 19.6 완전한 규격 탭



예측 설정

다음과 같은 선택적 설정을 지정할 수 있습니다 .

NAhead 예측 전 단계 수를 지정합니다 . 단계 수는 음수가 아니어야 합니다 .

주기 모형 적합 프로세스에서 고려할 계절성 값을 지정합니다 . 기본적으로 주기는 " 시간 패턴 분석 " 보고서의 암시된 계절성으로 설정됩니다 . 계절성이 다른 모형을 고려하려면 여기에 추가 주기를 지정하고 쉼표로 구분합니다 . 주기는 0 보다 커야 합니다 .

모형 선택 전략

최적 모형을 선택하는 방법을 지정합니다 .

정보 기준 지정된 정보 기준에 따라 최적 모형을 결정합니다 . 사용 가능한 정보 기준은 AIC 와 BIC 입니다 .

> **팁 :** 시계열이 긴 경우 BIC 를 사용하십시오 .

예측 성능 홀드백 집합을 포함하는 성능 선택 알고리즘에 따라 최적 모형을 결정합니다 . 먼저 시계열이 훈련 데이터 집합과 홀드백 집합으로 분할됩니다 . "NHoldback" 값은 홀드백 집합의 관측값 수를 지정합니다 . 그런 다음 알고리즘이 훈련 데이터 집합의 데이터에 대해 모든 권장 모형 또는 사용자 지정 모형을 적합시킵니다 . 예측은 개별 적합 모형을 사용하여 홀드

백 집합에 대해 수행됩니다 . 이러한 예측은 실제 홀드백 관측값과 비교되며 지정된 측정 기준을 사용하여 모형이 평가됩니다 . 측정 기준이 가장 적합한 모형이 최종 모형으로 선택됩니다 . 마지막으로 , 선택된 최종 모형이 전체 시계열 (훈련 데이터 집합과 홀드백 집합) 에 적합되고 재적합 모형은 시계열의 마지막 관측값 이후를 예측하는 데 사용됩니다 .

측정 기준 개별 모형 적합에 의한 예측을 평가할 측정 기준을 지정합니다 . 사용 가능한 측정 기준은 RMSE(제곱근 평균 제곱 오차), MSE(평균 제곱 오차) 및 MAE(평균 절대 오차) 입니다 .

NHoldback 홀드백 집합에 사용되는 관측값 수를 지정합니다 .

기타 옵션

다음 옵션을 지정할 수 있습니다 .

모형 선택 기준 유지 데이터에 적합된 모형 집합에 대한 모형 선택 기준을 저장합니다 . 모형 적합 후 이 정보는 " 모든 계열에 대한 모든 모형의 정보 기준 " 보고서에 표시됩니다 .

예측 구간 수준 예측에 대한 예측 구간 수준입니다 . 이 값은 예측 그림에서 음영 영역의 너비를 변경합니다 .

적용 가능한 모형에 대한 대치 모형 적합 프로세스 중에 사용되는 결측값 대치 방법을 지정합니다 .

참고 : 이러한 옵션은 모형 적합에 사용되는 데이터만 변경합니다 . 원시 데이터는 변경되지 않습니다 .

없음 결측값을 대치하지 않습니다 .

마지막 값 결측 관측값 시퀀스 전에 나오는 마지막 값을 사용하여 결측값을 대치합니다 .

중간 값 결측 관측값 시퀀스 전에 나오는 마지막 값과 결측 관측값 시퀀스 후에 나오는 첫 번째 값의 평균을 구하여 결측값을 대치합니다 .

선형 보간 결측 관측값 시퀀스 전에 나오는 마지막 값과 결측 관측값 시퀀스 후에 나오는 첫 번째 값 간의 선형 보간을 생성하여 결측값을 대치합니다 .

실행을 클릭하면 지정된 모형 집합이 적합됩니다 . 활성 스레드 수 , 완료된 작업 수 , 총 작업 수 및 완료된 작업의 백분율을 보고하는 진행률 표시줄이 나타납니다 . 작업은 하나의 계열과 반응 변수 조합에 대해 지정된 모든 모형을 적합시키는 것으로 정의됩니다 . 적합 프로세스가 완료되면 "모형 보고서"가 표시됩니다 . 자세한 내용은 "모형 보고서"에서 확인하십시오. "모델링 규격" 보고서의 옵션을 변경하고 **실행**을 다시 클릭하면 "모형 보고서"가 새 보고서로 대체됩니다 .

시계열 예측 플랫폼 옵션

" 시계열 예측 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

보고서 옵션 (모형을 적합시킨 후에만 적용) 다음과 같은 보고서 옵션을 포함합니다 .

예측 구간 수준 설정 예측 구간 수준을 변경할 수 있습니다 . 개별 모형 보고서가 자동으로 업데이트됩니다 .

NAhead 설정 예측 전 단계 수를 변경할 수 있습니다 . 개별 모형 보고서가 자동으로 업데이트됩니다 .

NHoldout 설정 홀드백 집합의 기간을 변경할 수 있습니다 . 개별 모형 보고서가 자동으로 업데이트됩니다 .

결과 저장 (모형을 적합시킨 후에만 적용) " 예측 " 이라는 별도의 데이터 테이블에 예측을 저장합니다 ." 예측 " 데이터 테이블은 원래 데이터 테이블과 레이아웃 및 형식이 동일합니다 . NHoldout 으로 홀드아웃 집합을 지정한 경우 홀드아웃 데이터의 실제값과 예측을 저장하는 추가 데이터 테이블이 생성됩니다 . 이 테이블을 사용하여 여러 유형의 오차 추정값을 계산할 수 있습니다 .

다음과 같은 추가 저장 옵션이 " 출력 요청 " 창에 표시됩니다 .

예측 구간 저장 하한 및 상한 예측 구간을 " 예측 " 데이터 테이블의 열로 저장합니다 .

1 시차 후 예측 저장 1 시차 후 예측과 예측을 쌓고 이를 별도의 데이터 테이블에 저장합니다 .

원래 계열 저장 원래 계열과 예측을 쌓고 이를 별도의 데이터 테이블에 저장합니다 .

예측 결과를 원래 테이블에 저장 결과를 원래 데이터 테이블의 새 열에 저장합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

모형 보고서

" 모형 보고서 " 에는 " 계열 선택 " 목록과 하나 이상의 개별 계열 모델링 보고서가 포함되어 있습니다 . 표시되는 개별 보고서는 " 계열 선택 " 목록에 의해 결정됩니다 . 목록에서 계열 이름을 클릭하면 선택한 계열에 대한 보고서가 표시됩니다 . Ctrl 키 또는 Shift 키를 사용하여 계열 이름을

두 개 이상 선택하면 선택한 각 계열에 대한 개별 모형 적합 보고서가 표시됩니다 . " 계열 선택 " 목록에서 하나 이상의 계열을 선택 취소하면 해당 계열에 대한 보고서가 제거됩니다 .

"계열 선택" 목록은 데이터 테이블에 연결되어 있습니다. "계열 선택" 목록에서 계열을 선택하면 해당 계열의 관측값과 반응 변수 조합도 데이터 테이블에서 선택됩니다. "계열 선택" 목록은 "Y에 대한 그룹 요약" 테이블에도 연결됩니다. 시간 변수를 지정한 경우에는 "계열 선택" 목록이 "시간에 대한 그룹 요약" 테이블에도 연결됩니다. "계열 선택" 목록에서 계열을 선택하면 해당 항목이 테이블에서도 선택됩니다.

각 개별 모형 보고서에는 다음 정보가 포함되어 있습니다 .

모형 유형 계열의 최량 적합 모형에 대한 오차 , 추세 및 계절성을 설명합니다 .

모형 보고서 계열에 적합된 모형을 기반으로 한 보고서 그룹입니다 .

계열 및 예측 (" 정보 기준 " 을 모형 선택 전략으로 지정한 경우에만 사용 가능) 원래 계열과 예측에 대한 그림입니다 . 원래 계열의 기간 동안 1 시차 후 예측이 표시됩니다 . 원래 계열의 기간이 지나면 예측이 표시됩니다 . 음영 영역은 예측 구간을 나타냅니다 .

참고 : 모든 예측은 모든 개별 계열에서 마지막 시점의 바로 다음 시점에서 시작됩니다 . 자세한 내용은 " 데이터 형식 " 에서 확인하십시오 .

훈련 , 홀드백 및 예측 (" 예측 성능 " 을 모형 선택 전략으로 지정한 경우에만 사용 가능) 훈련 계열 , 홀드백 계열 및 예측 계열에 대한 그림입니다 . 세로 음영 영역은 홀드백 계열을 나타냅니다 . 전체 시계열에 대한 모형 재적합이므로 전체 원래 시계열의 기간 동안 1 시차 후 예측이 표시됩니다 . 마지막 관측값 이후의 기간 동안 예측이 표시됩니다 . 음영 영역은 예측 구간을 나타냅니다 .

모형 요약 -2* 로그 가능도 , AIC, 모수 수 및 시그마 값을 제공합니다 . 시그마 값은 1 시차 후 예측 표준편차의 추정값입니다 . " 예측 성능 " 이 모형 선택 전략으로 지정된 경우 홀드백 집합의 RMSE(제곱근 평균 제곱 오차) 도 표시됩니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

모수 추정값 모형에 대한 모수 추정값을 제공합니다 . 모수는 " 모형 유형 " 보고서에 설명된 최량 적합 모형의 모수에 해당합니다 .

1 시차 후 예측 오차 시간에 따른 1 시차 후 예측 오차에 대한 그림입니다 . 각 시점 t 에 대해 오차는 $e(t) = y(t) - \mu(t)$ 로 계산됩니다 .

승법 오차 모형의 1 시차 후 상대 예측 오차 (승법 오차 모형에만 사용 가능) 시간에 따른 1 시차 후 상대 예측 오차에 대한 그림입니다 . 각 시점 t 에 대해 오차는 $e(t) = [y(t) - \mu(t)]/\mu(t)$ 로 계산됩니다 .

20 장

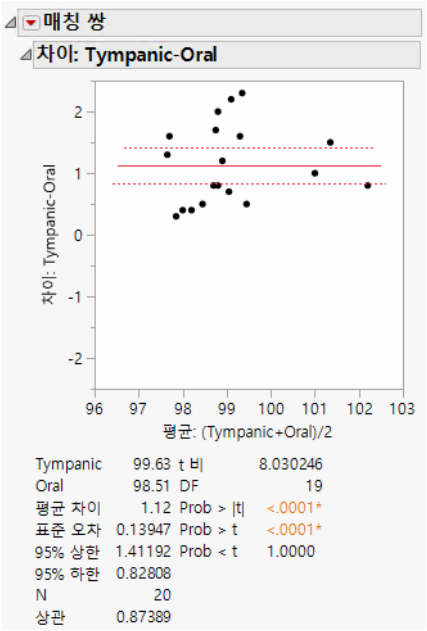
매칭 쌍 분석

동일한 개체에 대한 측정값 비교

매칭 쌍 플랫폼에서는 상관관계가 있는 두 개 이상의 변수 간 평균을 비교하고 차이를 평가합니다. 예를 들어 동일한 피실험자에 대해 치료 전과 치료 후의 혈압 측정값을 비교할 수 있습니다. **쌍체 t-검정**이라는 통계적 방법은 상관된 반응을 고려합니다.

이 플랫폼에서는 세 가지 모든 대립가설에 대한 쌍체 *t*-검정 결과와 쌍체 평균별 쌍체 차이 그래프를 제공합니다. 세 개 이상의 매칭 반응과 그룹화 열에 대해 단순 반복 측정 분석으로 표본 간 검정을 수행할 수 있는 추가 기능도 제공합니다.

그림 20.1 매칭 쌍 분석의 예



목차

매칭 쌍 플랫폼 개요 363
매칭 쌍 비교의 예 363
매칭 쌍 플랫폼 시작 364
여러 Y 열 365
매칭 쌍 보고서 366
차이 그림 및 보고서 366
그룹 간 367
매칭 쌍 플랫폼 옵션 367
그룹 간 매칭 쌍 비교의 예 368
매칭 쌍 플랫폼에 대한 통계 상세 정보 369
매칭 쌍 그래픽 369
반응 상관 370

매칭 쌍 플랫폼 개요

매칭 쌍 플랫폼에서는 쌍체 *t*-검정을 사용하여 두 반응 열 간의 행별 차이를 비교합니다. 두 열은 동일한 피실험자에 대한 처리 전과 처리 후의 측정값을 나타내는 경우가 많습니다. 또는 측정값이 두 개의 다른 측정기로 동일한 개체에서 수집한 데이터를 나타낼 수 있습니다.

데이터 테이블의 두 열에 쌍체 데이터가 배열된 경우 매칭 쌍 플랫폼을 사용할 수 있습니다 . 그러나 모든 측정값이 단일 열에 있는 경우 다음 작업 중 하나를 수행하십시오 .

- **테이블** 메뉴의 **분할** 옵션을 사용하여 측정값 열을 두 개의 열로 분할합니다 . 그런 다음 매칭 쌍 플랫폼을 사용할 수 있습니다 .
- 반응 열이 두 개인 경우 두 반응 간의 차이를 계산하는 세 번째 열을 생성합니다 . 그런 다음 분포 플랫폼을 사용하여 차이 열의 평균이 0 인지 검정합니다 .
- 두 반응이 단일 열에 저장된 경우 이원 분산 분석을 수행할 수 있습니다 . 한 요인 (ID 변수) 은 두 반응을 식별하고 다른 요인은 개체를 식별합니다 . 블록 변수 (개체 열) 와 함께 X 로 Y 적합 일원 분석 플랫폼을 사용하거나 , 모형 적합 플랫폼을 사용하여 이원 ANOVA 를 수행합니다 . ID 요인에 대한 검정은 쌍체 *t*- 검정과 동일합니다 .

참고 : 쌍체 데이터의 경우 일반적인 독립 *t*- 검정을 수행하지 마십시오 . 데이터를 한 열에 쌓지 말고 , 블록 변수 없이 ID 에 대해 X 로 Y 적합 일원 ANOVA 를 사용합니다 . 이렇게 하면 반응 간의 상관을 무시하는 효과가 있습니다 . 이 경우 검정을 수행할 때 반응에 음의 상관관계가 있으면 효과를 과대추정하고 반응에 양의 상관관계가 있으면 효과를 과소추정합니다 .

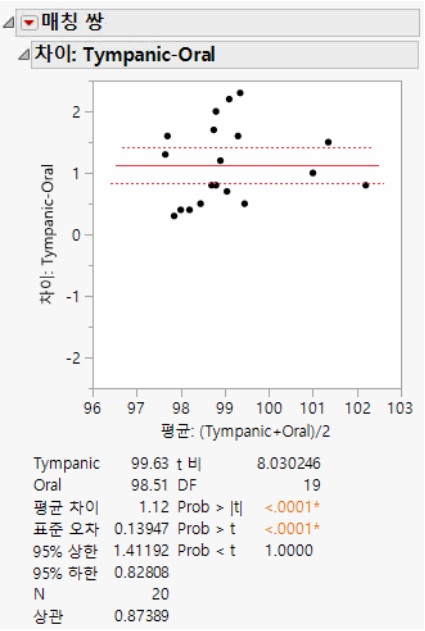
매칭 쌍 비교의 예

이 예에서는 Therm.jmp 샘플 데이터 테이블을 사용합니다 . 데이터에는 20 명의 체온 측정값이 포함되어 있습니다 . 체온은 구강 체온계와 고막 (귀) 체온계 두 종류를 사용하여 측정되었습니다 . 두 종류의 체온계로 측정한 체온이 동일한지 여부를 확인하려고 합니다 . 다른 사람과의 체온 차이는 중요하지 않습니다 . 매칭 쌍 분석은 체온계 간의 차이를 검정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Therm.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 매칭 쌍**을 선택합니다 .
3. Oral 과 Tympanic 을 선택하고 **Y, 쌍체 반응**을 클릭합니다 .
4. **확인**을 클릭합니다 .

   보고서 창이 나타납니다 .

그림 20.2 매칭 쌍 보고서 창



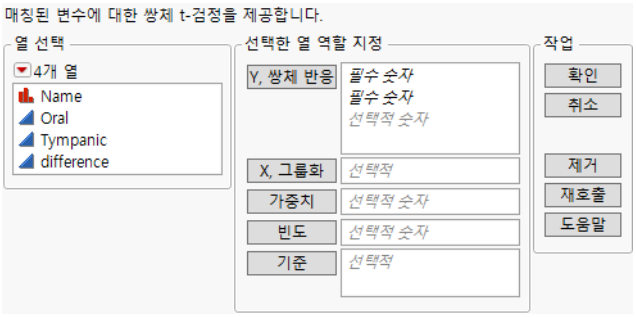
여기에서는 대체로 고막 체온계가 구강 체온계보다 1.12도 높게 측정된다는 결과를 보여 줍니다. *p* 값(Prob > |t|)이 작으므로 이 차이는 우연히 발생한 것이 아니라 통계적으로 유의하다는 것을 나타냅니다.

이 매칭 쌍 분석에서는 어떤 체온계가 정확한지(둘 중 하나인 경우)는 나타내지 않고 체온계 간에 차이가 있다는 것만 나타냅니다.

매칭 쌍 플랫폼 시작

분석 > 전문 모델링 > 매칭 쌍을 선택하여 매칭 쌍 플랫폼을 시작합니다.

그림 20.3 매칭 쌍 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 쌍체 반응 두 개의 반응 열을 제공합니다. 세 개 이상의 반응을 분석하는 방법에 대한 자세한 내용은 "여러 Y 열"에서 확인하십시오.

X, 그룹화 그룹 간 차이를 비교하기 위한 그룹화 변수를 제공합니다. 자세한 내용은 "그룹 간"에서 확인하십시오.

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 하나의 열을 식별합니다.

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 하나의 열을 식별합니다.

기준 기준 변수의 각 수준에 대해 개별 매칭 쌍 분석을 수행합니다.

확인을 클릭하면 "매칭 쌍" 보고서 창이 나타납니다. 자세한 내용은 "매칭 쌍 보고서"에서 확인하십시오.

여러 Y 열

반응이 세 개 이상일 수 있습니다. 반응 수가 홀수이면 가능한 모든 쌍이 분석됩니다. 다음 표에서는 세 가지 반응의 예를 보여 줍니다.

Y1 대 Y2	Y1 대 Y3
	Y2 대 Y3

반응 수가 짝수이면 매칭 쌍 플랫폼에서 가능한 모든 쌍을 수행할지 묻습니다. 가능한 모든 쌍을 수행하지 않으면 인접한 반응이 쌍으로 분석됩니다. 다음 표에서는 네 개 반응에 대한 분석 배열을 보여 줍니다.

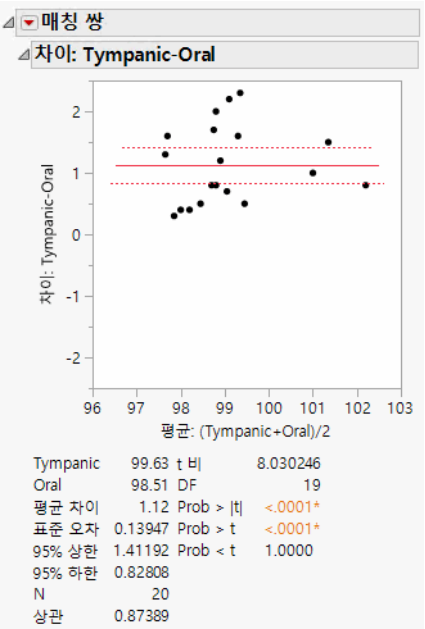
Y1 대 Y2	Y3 대 Y4

매칭 쌍 보고서

그림 20.4에 표시된 것과 같은 보고서 창을 생성하려면 "매칭 쌍 비교의 예"에 설명된 방법을 따르십시오.

" 매칭 쌍 " 보고서에는 Tukey 평균 - 차이 그림 , 요약 통계량 및 쌍체 *t*- 검정 결과가 표시됩니다 . 자세한 내용은 " 차이 그림 및 보고서 " 에서 확인하십시오 . **X, 그룹화** 변수를 지정한 경우 보고서에 " 그룹 간 " 보고서도 포함됩니다 . 자세한 내용은 " 그룹 간 " 에서 확인하십시오 .

그림 20.4 매칭 쌍 보고서의 예



참고 : 빨간색 삼각형 메뉴에는 초기 보고서 창에 보고서를 추가할 수 있는 다른 옵션이 더 있습니다 . 자세한 내용은 " 매칭 쌍 플랫폼 옵션 " 에서 확인하십시오 .

차이 그림 및 보고서

" 차이 " 그림에는 평균별 차이가 표시됩니다 . " 차이 " 그림에서 다음 사항에 유의하십시오 .

- 평균 차이는 가로선으로 표시되고 위 / 아래에는 95% 신뢰 구간이 파선으로 표시됩니다 . 신뢰 영역에 0 이 포함되면 평균이 0.05 수준에서 유의한 차이가 없음을 나타냅니다 . 이 예에서는 차이가 유의합니다 .
- 참조 프레임을 추가하면 쌍의 평균이 세로선으로 표시됩니다. 참조 프레임에 대한 자세한 내용은 " 매칭 쌍 플랫폼 옵션 " 에서 확인하십시오 .

"차이" 보고서에는 각 반응의 평균, 평균 차이 및 차이에 대한 신뢰 구간이 표시됩니다. 쌍체 *t*-검정에 대한 결과도 "차이" 보고서에 표시됩니다.

그룹 간

참고 : " 그룹 간 " 보고서는 **X, 그룹화** 변수를 지정한 경우에만 나타납니다 .

그룹 간 분석은 단순 반복 측정 분석에 해당합니다 . 모형 적합 플랫폼의 **MANOVA** 분석법을 사용하여 동일한 검정 결과를 얻을 수 있습니다 .

평균 차이 두 쌍체 열 사이에 각 그룹의 전체 행 차이의 평균을 표시합니다 . 다시 말하면 , 개체 내와 개체 간 교호작용 또는 분할구와 주구 교호작용입니다 .

평균 평균 두 쌍체 열에 대해 각 그룹의 전체 행 평균의 평균을 표시합니다 . 다시 말하면 , 개체 간 또는 주구 효과입니다 .

그룹 간 검정 다음 두 *F*- 검정은 그룹 간 값이 다른지 여부를 판별합니다 .

평균 차이 반응 쌍 간의 변화가 그룹마다 다른지 검정합니다 .

평균 평균 개체에 대한 평균 반응이 그룹마다 다른지 검정합니다 .

관련 정보

- " 그룹 간 매칭 쌍 비교의 예 "

매칭 쌍 플랫폼 옵션

" 매칭 쌍 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

평균별 차이 그림 쌍체 평균별 쌍체 차이 그림을 표시하거나 숨깁니다 . 이 그림에 대한 자세한 설명은 " 차이 그림 및 보고서 " 에서 확인하십시오 .

행별 차이 그림 행 번호별 쌍체 차이 그림을 표시하거나 숨깁니다 .

참조 프레임 평균별 차이 그림에 참조 프레임을 표시하거나 숨깁니다 . 이 그림 안에서 Y2 대 Y1 그림의 위치를 보여 주는 직사각형으로 기울어져 있고 , 눌린 모양일 수 있습니다 . 평균의 평균을 나타내는 빨간색 세로선이 표시됩니다 . 차이 범위가 데이터 범위의 절반보다 클 때 참조 프레임이 처음부터 표시됩니다 .

Wilcoxon 부호 순위 Wilcoxon 부호 순위 검정을 표시하거나 숨깁니다 . Wilcoxon 부호 순위 검정은 쌍체 차이에 적용됩니다 . 이 검정은 양수 차이의 크기를 음수 차이의 크기와 비교하는 비모수 검정입니다 . 이 검정에서는 Pratt 방법을 사용하여 0 차이 값을 처리합니다 . 또한 차이의 분포가 대칭적이라고 가정합니다 . 자세한 내용은 **기본 분석**의 에서 확인하십시오 . Lehman(2006), Conover(1999, p. 350) 및 Cureton(1967) 연구 자료도 확인하십시오 .

부호 검정 부호 검정을 표시하거나 숨깁니다 . 이 검정은 차이 값의 부호 (양수 또는 음수) 만 검정에 사용하는 쌍체 *t*- 검정의 비모수 버전입니다 .

α **수준 설정** 분석에 사용되는 유의 수준을 변경합니다 . 보고서와 그림의 신뢰 구간에 영향을 줍니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

그룹 간 매칭 쌍 비교의 예

이 예에서는 Dogs.jmp 샘플 데이터 테이블을 사용합니다 . 여기에서는 그룹 간 매칭 쌍 보고서 및 모형 적합을 사용한 해당 MANOVA 보고서를 모두 생성하는 방법을 보여 줍니다 .

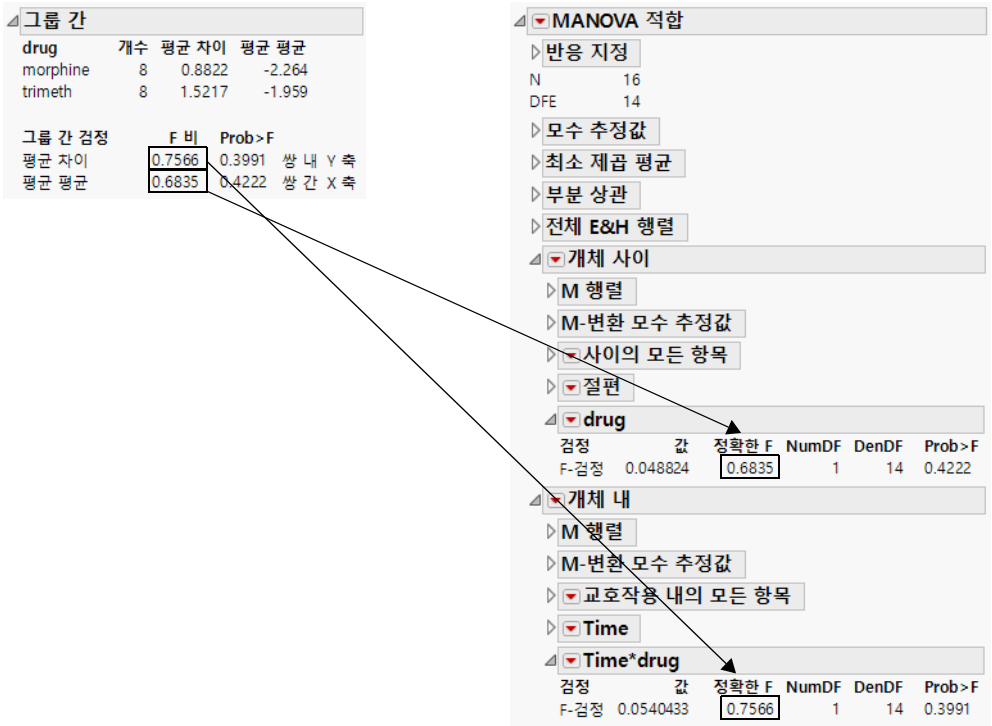
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Dogs.jmp 를 엽니다 .
2. **분석 > 전문 모델링 > 매칭 쌍**을 선택합니다 .
3. LogHist0 및 LogHist1 을 선택하고 **Y, 쌍체 반응**을 클릭합니다 .
4. drug 를 선택하고 **X, 그룹화**를 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 20.5 의 왼쪽 보고서가 나타납니다.

이제 동일한 데이터 테이블을 사용하여 모형 적합 보고서를 생성합니다.

1. **분석 > 모형 적합**을 선택합니다 .
2. LogHist0 및 LogHist1 을 선택하고 **Y** 를 클릭합니다 .
3. drug 를 선택하고 **추가**를 클릭합니다 .
4. **Manova** 분석법을 선택합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 지정 " 보고서의 **반응 선택** 메뉴에서 **반복 측정**을 선택합니다 .
7. **확인**을 클릭합니다 .

그림 20.5 그룹 간 매칭 쌍 및 반복 측정을 사용한 모형 적합 MANOVA 의 예



"그룹 간" 보고서에서 "평균 차이"의 "F 비"는 "개체 내" 보고서에서 "시간*drug"의 F 비에 해당합니다. "그룹 간" 보고서에서 "평균 평균"의 "F 비"는 "MANOVA 적합" 보고서의 "개체 사이"에서 "drug"의 F 비에 해당합니다.

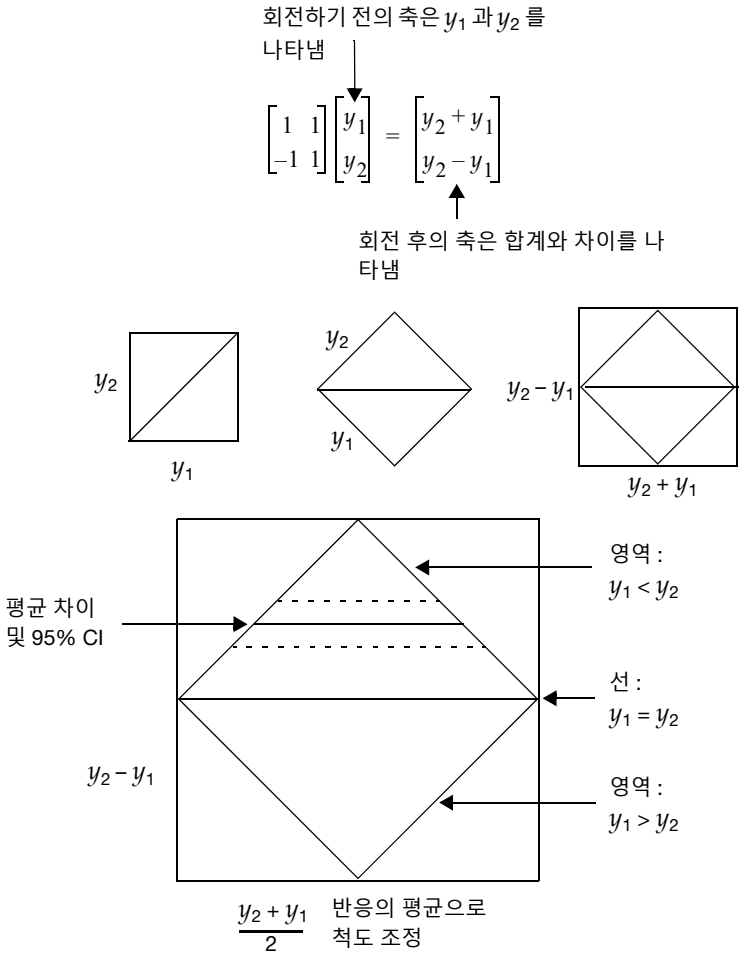
매칭 쌍 플랫폼에 대한 통계 상세 정보

- "매칭 쌍 그래픽"
- "반응 상관"

매칭 쌍 그래픽

플랫폼의 기본 그래프는 Tukey 평균-차이입니다(Cleveland 1994, p. 130). 이 그래프는 두 반응의 차이가 세로 축에 표시되고 두 반응의 평균이 가로 축에 표시됩니다. 이 그래프는 두 원래 변수의 산점도를 45도 회전한 것과 같습니다. 45도 회전하고 척도를 조정하면 원래 좌표가 차이와 평균이 됩니다.

그림 20.6 평균별 차이로 변환 (45 도 회전) 의 예

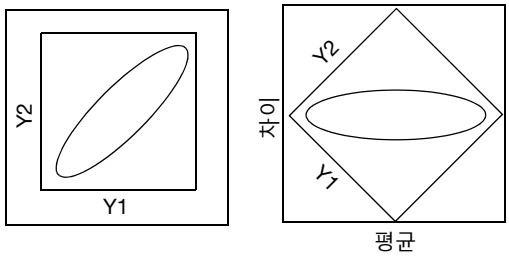


반응 상관

동일한 개체에서 서로 다른 시간에 한 쌍의 측정값을 가져오는 대부분의 경우 두 값은 양의 상관관계가 있습니다. 그러나 경쟁 반응을 나타내는 값일 경우 음의 상관관계가 있을 수 있습니다.

그림 20.7에서는 두 반응의 양의 상관관계가 차이(세로 축)에 대한 작은 분산이 되는 방식을 보여 줍니다. 음의 상관관계가 있으면 타원이 다른 방향을 향하고 회전한 그래프의 분산이 세로 축에서 큽니다.

그림 20.7 회전 전후의 양의 상관관계 예제



21 장

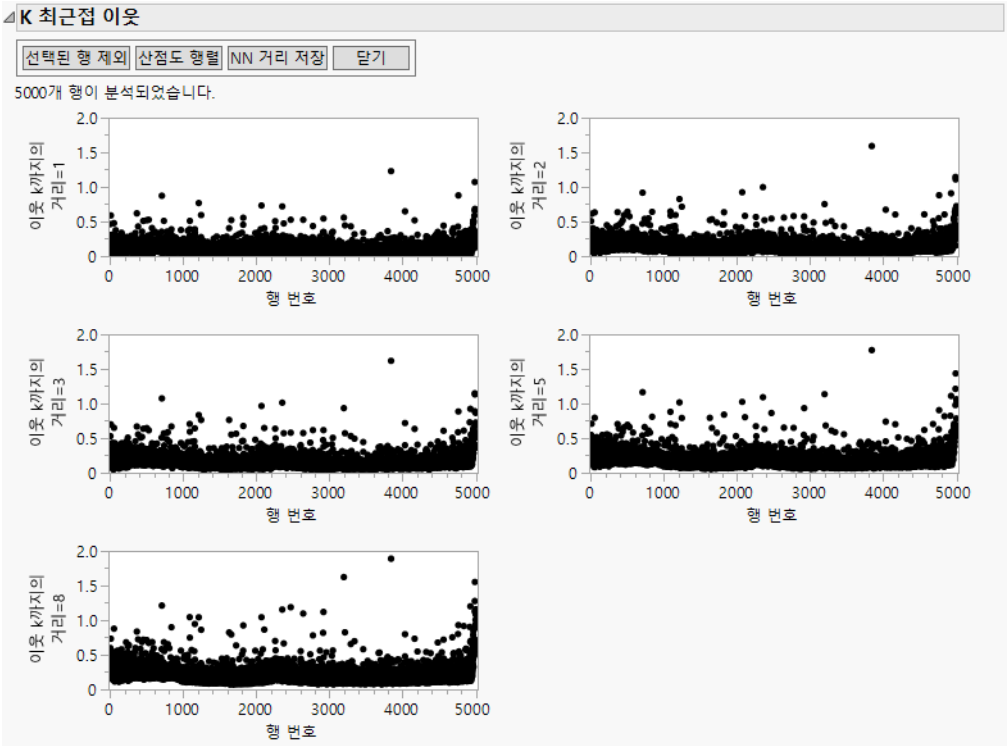
모델링 유틸리티

데이터에서 이상치 , 결측값 및 패턴 탐색

모델링 유틸리티는 데이터 분석의 데이터 정리 및 전처리 단계를 지원하는 용도의 유틸리티 모음입니다. 각 유틸리티에는 데이터를 더 정확하게 이해할 수 있는 탐색 도구가 있습니다. 모델링 유틸리티를 사용하여 다음을 수행할 수 있습니다.

- 단변량 및 다변량 사례 둘 다에서 이상치 탐색
- 데이터에서 결측값 탐색 및 대치
- 데이터에서 일반적이지 않은 특징이나 패턴 탐색 및 찾기

그림 21.1 다변량 k- 최근접 이웃 이상치 예



목차

이상치 탐색 유틸리티 375
이상치 탐색 유틸리티의 예 375
이상치 탐색 유틸리티 시작 378
분위수 범위 이상치 379
로버스트 적합 이상치 382
로버스트 PCA 이상치 384
K 최근접 이웃 이상치 387
이상치 탐색 유틸리티 옵션 389
이상치 탐색 유틸리티의 추가 예 389
결측값 탐색 유틸리티 391
결측값 탐색 유틸리의 예 391
결측값 보고서 394
결측값 탐색 유틸리티 옵션 399
패턴 탐색 유틸리티 399
패턴 탐색 유틸리티의 예 399
패턴 탐색 유틸리티 시작 402
패턴 탐색 보고서 403
패턴 탐색 유틸리티 옵션 404
패턴 탐색 유틸리티의 추가 예 407
패턴 탐색 유틸리티에 대한 통계 상세 정보 409

이상치 탐색 유틸리티

이상치 탐색 도구는 이상치를 식별, 탐색 및 관리하기 위한 네 가지 옵션을 제공합니다. 데이터에서 이상치를 탐색하고 이해하는 것은 분석의 중요한 부분입니다. 데이터의 이상치는 데이터 수집 또는 보고 단계에서의 실수, 측정 시스템 오류, 데이터 집합에 오류 또는 결측값 포함, 단순 비정상 값 등이 원인이 될 수 있습니다. 이상치가 있으면 추정값이 왜곡되고 결과가 해당 이상치로 치우칠 수 있습니다.

이상치로 인해 표본 분산이 부풀려지기도 합니다. 경우에 따라 데이터의 이상치를 유지해야 하는데 이를 제거하면 표본 분산이 과소추정되고 데이터가 반대 방향으로 편향될 수 있습니다.

이상치는 제거 여부에 상관없이 찾는 것이 좋습니다. 데이터를 시각적으로 검사할 수 있는 여러 가지가 있습니다. 예를 들어 상자 그림, 히스토그램 및 산점도에서는 이러한 극단값을 쉽게 표시할 수 있습니다. 자세한 내용은 JMP **살펴보기**의 에서 확인하십시오.

이상치 탐색 도구에는 다음과 같은 옵션이 있습니다.

단변량 단변량 데이터에서 이상치를 탐색하기 위한 두 가지 옵션이 있습니다.

분위수 범위 이상치 각 열의 분위수 분포를 사용하여 이상치를 극단값으로 식별합니다. 이 도구는 데이터 내에서 결측값 또는 오류 코드를 검색하는 데 유용합니다. 이 방법은 데이터의 이상치 탐색을 시작하는 데 권장됩니다. 자세한 내용은 " 분위수 범위 이상치 " 에서 확인하십시오.

로버스트 적합 이상치 각 열의 중심 및 산포에 대한 로버스트 추정값을 찾고 이상치를 해당 값에서 멀리 떨어진 데이터 점으로 식별합니다. 자세한 내용은 " 로버스트 적합 이상치 " 에서 확인하십시오.

다변량 다변량 데이터에서 이상치를 탐색하기 위한 두 가지 옵션이 있습니다.

로버스트 PCA 이상치 데이터를 낮은 계수 행렬과 잔차로 분해하고 잔차를 사용하여 이상치를 감지합니다. 자세한 내용은 " 로버스트 PCA 이상치 " 에서 확인하십시오.

K 최근접 이웃 이상치 이상치를 k 개 최근접 이웃에서 멀리 떨어진 값으로 식별합니다. 자세한 내용은 "K 최근접 이웃 이상치 " 에서 확인하십시오.

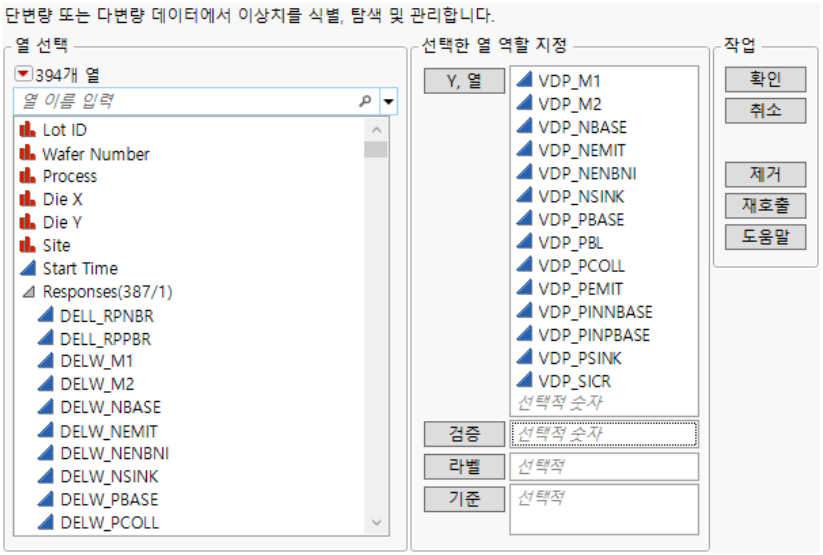
이상치 탐색 유틸리티의 예

이상치 탐색 유틸리티를 사용하면 분포 플랫폼에서 검토할 수 있는 이상치를 식별할 수 있습니다. Probe.jmp 샘플 데이터 테이블에는 5800 개의 반도체 웨이퍼에서 측정한 387 개의 특성 (Responses 열 그룹) 이 포함되어 있습니다. Lot ID 및 Wafer Number 열은 웨이퍼를 고유하게 식별합니다. 데이터 집합의 선택된 열 그룹 내에서 이상치를 식별하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Probe.jmp 샘플 데이터 테이블을 엽니다.
2. **분석 > 선별 > 이상치 탐색**을 선택합니다.
3. Responses(387/0) 옆의 삼각형을 클릭하여 그룹의 모든 열을 표시합니다.

4. VDP_M1 ~ VDP_SICR 열을 선택하고 **Y, 열**을 클릭합니다 . 14 개 열이 선택되어야 합니다 .

그림 21.2 이상치 탐색 시작 창



5. **확인**을 클릭합니다 .
6. **분위수 범위 이상치**를 클릭합니다 .

   " 분위수 범위 이상치 " 보고서에 각 열이 표시되고 발견된 이상치의 수와 ID 가 나열됩니다 .
7. " 분위수 범위 이상치 " 보고서에서 **이상치를 포함하는 열만 표시** 를 선택합니다 . 이렇게 하면 이상치를 포함하는 열만 표시되도록 열 목록이 제한됩니다 .

   몇몇 열에 이상치 값 9999 가 포함되어 있습니다 . 많은 업계에서 9 모음값을 결측값 코드로 사용합니다 .
8. "9 모음값 " 보고서에서 각 열을 선택합니다 .
9. **결측값 코드에 가장 큰 9 모음값 추가**를 클릭합니다 .

   원래 데이터를 유지하려면 **다른 이름으로 저장** 명령을 사용해야 함을 나타내는 "JMP 경고 " 가 표시됩니다 .
10. **확인**을 클릭합니다 .
11. " 분위수 범위 이상치 " 보고서에서 **다시 스캔**을 클릭합니다 .
12. **검색을 정수로 제한**을 선택합니다 .

    연속형 데이터에서 정수 값은 종종 오류 코드 또는 코드화된 다른 데이터 값을 나타냅니다 . 이 열 집합에는 추가 오류 코드가 포함되어 있지 않습니다 .
13. **검색을 정수로 제한**을 선택 취소합니다 .

데이터 검토

1. " 분위수 범위 이상치 " 보고서에서 나머지 열을 모두 선택합니다 .
2. **행 선택하기**를 클릭합니다 .
3. **분석 > 분포**를 선택합니다 .
4. 선택된 열을 **Y, 열** 역할에 할당합니다 . " 분위수 범위 이상치 " 보고서에서 이러한 열 이름을 선택했으므로 " 분포 " 시작 창에 이미 선택되어 있습니다 .
5. **확인**을 클릭합니다 .

그림 21.3 이상치가 선택된 열의 분포

VDP_M1 및 VDP_PEMIT 열에서는 선택된 이상치 중 일부가 대부분의 데이터에 다소 가깝습니다 . 나머지 열의 경우 선택된 이상치가 대부분의 데이터에서 멀리 떨어져 있습니다 . 데이터 점을 조사하고 분석에서 제외하십시오 .

제외된 이상치 구체화

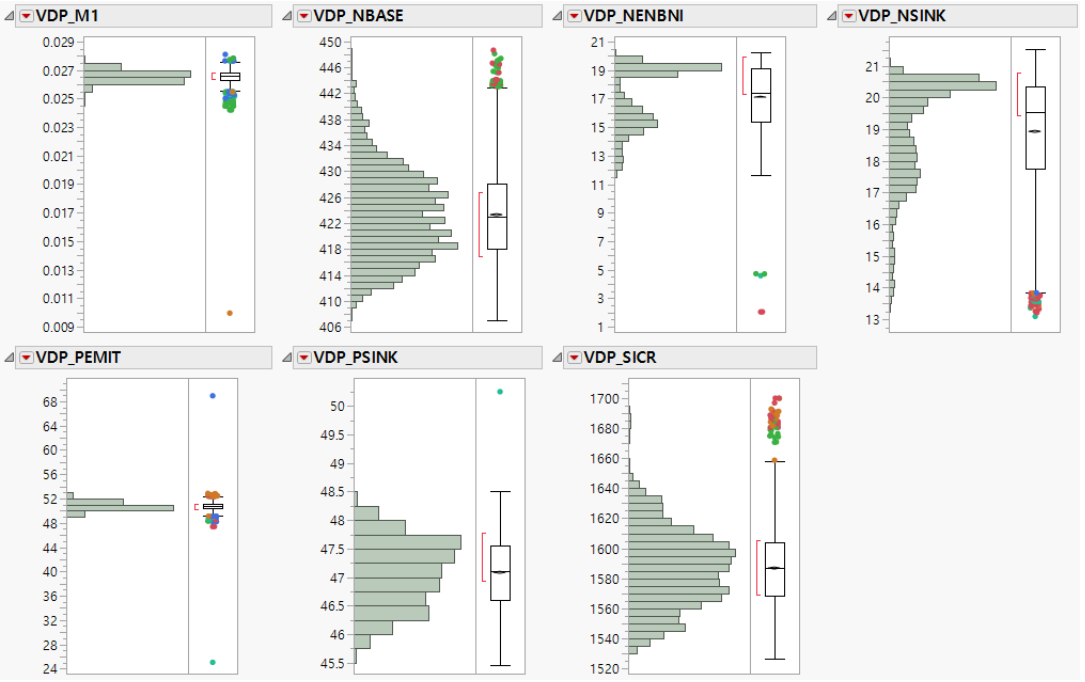
1. " 분위수 범위 이상치 " 보고서에서 Ctrl 키를 누른 채 VDP_M1 및 VDP_PEMIT 열을 선택 취소합니다 .
2. 나머지 열은 보고서에서 선택된 상태로 **행 제외**를 클릭합니다 .
3. "Q" 를 "20" 으로 변경합니다 .
4. **다시 스캔**을 클릭합니다 .

5. 보고서에서 VDP_M1 및 VDP_PEMIT 열을 선택합니다 .
6. **행 선택하기**를 클릭합니다 .

데이터 재검토

1. " 분포 " 보고서를 다시 검토합니다 . 선택된 이상치가 대부분의 데이터에서 충분히 떨어져 있어 분석에서 선택 및 제외할 수 있습니다 .
2. " 분위수 범위 이상치 " 보고서에서 **행 제외**를 클릭합니다 .
3. " 분포 " 보고서에서 " 분포 " 의 빨간색 삼각형을 클릭하고 **다시 실행 > 분석 다시 실행**을 선택합니다 .

그림 21.4 이상치가 제외된 열의 분포



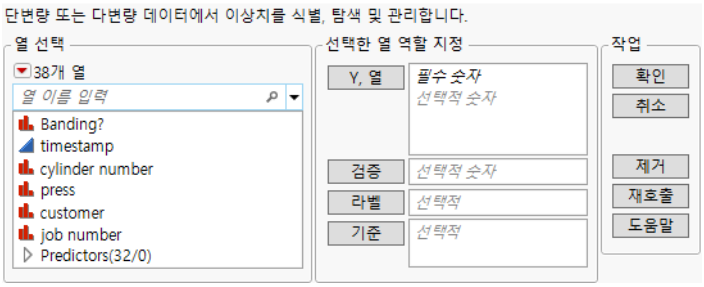
이제 이상치 없이 더 유용한 정보가 데이터 분포에 표시됩니다 .

이상치 탐색 유틸리티 시작

이상치 탐색을 시작하려면 **분석 > 선별 > 이상치 탐색**을 선택합니다 .

참고 : " 이상치 탐색 " 명령은 연속형 모델링 유형의 열만 분석합니다 . 다른 열을 시작 창에 입력할 수 있지만 무시됩니다 .

그림 21.5 이상치 탐색 유틸리티 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 열 분석할 열을 지정합니다 .

검증 로버스트 PCA 이상치에 사용되는 검증 열을 지정합니다 .

라벨 다변량 분석 보고서의 행 번호를 라벨로 바꾸는 열을 지정합니다 .

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다 . 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다 . 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

확인 을 클릭하면 " 이상치 탐색 " 보고서가 나타납니다 . 다음과 같은 네 가지 이상치 분석 명령을 사용할 수 있습니다 .

- " 분위수 범위 이상치 "
- " 로버스트 적합 이상치 "
- " 로버스트 PCA 이상치 "
- "K 최근접 이웃 이상치 "

팁: 기준 변수의 모든 수준에 대해 이상치 분석을 실행하려면 Ctrl 키를 누른 채 원하는 이상치 분석 명령 버튼을 클릭합니다 .

분위수 범위 이상치

이상치 감지를 위한 "분위수 범위 이상치" 방법은 열 값의 분위수 분포를 사용하여 극단값을 찾습니다. 분위수는 관련된 분포 가정이 없으므로 이상치를 감지하는 데 유용합니다. 데이터는 가장 작은 것부터 오름차순으로 정렬됩니다. 예를 들어 20번째 분위수는 값의 20%가 더 작은 위치의 기준 값입니다. 극단값은 분위수 간 범위, 즉 지정된 두 분위수 간 거리의 승수를 사용하여 구합니다. 분위수 계산 방법에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오.

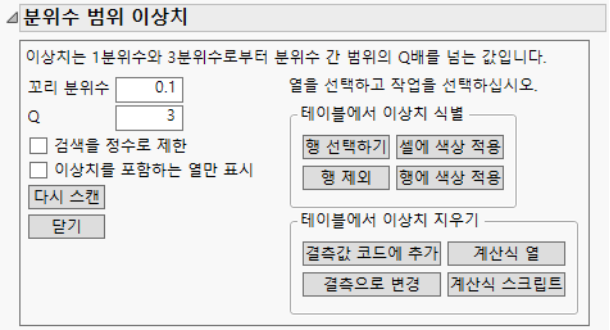
분위수 범위 이상치 유틸리티는 데이터 내에 저장된 결측값 코드를 식별하는 데도 유용합니다. 앞에서 언급했듯이 일부 업계에서는 결측값이 9모음값(예: 999 또는 9999)으로 입력됩니다. 이

유틸리티는 3사분위수보다 큰 9모음값을 의심되는 결측값 코드로 찾습니다. 사용자는 이 유틸리티를 사용하여 해당 결측값 코드를 데이터 테이블의 열 특성으로 추가할 수 있습니다.

분위수 범위 이상치 옵션

"분위수 범위 이상치" 패널에서는 이상치를 식별하는 방법과 관리하는 방법을 지정할 수 있습니다.

그림 21.6 분위수 범위 이상치 창



이상치는 하위 분위수와 상위 분위수의 분위수 간 범위의 *Q* 배를 넘는 값으로 간주됩니다. *Q* 값과 분위수 간 범위 크기를 조정할 수 있습니다.

꼬리 분위수 분위수 간 범위를 계산하는 데 사용되는 하위 분위수에 대한 확률입니다. 상위 분위수 확률은 1 - 꼬리 분위수로 간주됩니다. 예를 들어 꼬리 분위수 값이 0.1 이면 분위수 간 범위는 데이터의 0.1 분위수에서 0.9 분위수 사이임을 의미합니다. 기본값은 0.1 입니다.

Q 이상치 임계값을 결정하는 승수입니다. 꼬리 분위수 또는 1 - 꼬리 분위수 값을 넘어 분위수 간 범위의 *Q* 배를 초과하는 값이 이상치로 식별됩니다. *Q* 값이 크면 작은 값보다 더 보수적인 이상치 집합이 제공됩니다. 기본값은 3 입니다.

검색을 정수로 제한 이상치 값을 정수 값으로만 제한합니다. 이 설정은 업종별 결측값 코드 및 오류 코드를 찾기 위해 이상치 검색을 제한합니다.

이상치를 포함하는 열만 표시 보고서의 열 목록을 이상치가 포함된 열로 제한합니다.

사용자가 지정한 내용에 따라 보고서가 표시된 후 다양한 방법을 사용하여 이러한 극단값을 처리할 수 있습니다. "분위수 범위 이상치" 보고서에서 지정된 열을 선택하여 열의 이상치를 선택할 수 있습니다.

팁 : 보고서에서 열을 선택하지 않고 다음 버튼 중 하나를 클릭하면 모든 열을 선택할 수 있는 "JMP 경고" 가 나타납니다.

행 선택하기 데이터 테이블에서 선택한 열의 이상치 행을 선택합니다.

행 제외 "분위수 범위 이상치" 보고서에서 선택한 열의 이상치에 대해 행 제외 상태를 설정합니다. "분위수 범위 이상치" 보고서를 업데이트하려면 **다시 스캔**을 클릭합니다.

셀에 색상 적용 데이터 테이블에서 선택한 이상치의 셀에 색상을 적용합니다 .

행에 색상 적용 데이터 테이블에서 선택한 열의 이상치가 포함된 행에 색상을 적용합니다 .

결측값 코드에 추가 선택한 이상치를 결측값 코드 열 특성에 추가합니다 . 이 옵션을 사용하면 데이터 내에서 알려진 결측값 또는 오류 코드를 식별할 수 있습니다 . 결측값 및 오류 코드는 대개 정수이고 9 모음값일 때도 있습니다 . " 분위수 범위 이상치 " 보고서를 업데이트하려면 **다시 스캔**을 클릭합니다 .

참고 : 시작 창에서 기준 변수를 지정한 경우 " 분위수 범위 이상치 " 에서 " 결측값 코드에 추가 " 를 사용할 수 없습니다 .

결측으로 변경 데이터 테이블에서 이상치 값을 결측값으로 변경합니다 . 값을 결측값으로 변경할 때 주의해야 합니다 . 데이터가 올바르지 않거나 정확하지 않을 때만 결측값으로 변경합니다 . " 분위수 범위 이상치 " 보고서를 업데이트하려면 **다시 스캔**을 클릭합니다 .

참고 : 선택한 이상치가 결측값 코드에 추가된 경우 이상치가 결측값으로 변경되지 않습니다 .

계산식 열 시작 창에 지정된 각 열에 대한 새 계산식 열을 생성합니다 . 각 새 열은 원래 열의 값이 이상치 제한 범위 내에 있는 경우 해당 값을 포함하고 그렇지 않은 경우 결측으로 설정됩니다 . 새 열은 원래 열과 구분하기 위해 사용자가 지정한 이름을 접두사 또는 접미사로 사용합니다 . 기본적으로 접미사는 " 선별 " 로 설정됩니다 .

계산식 스크립트 데이터 테이블에 추가되는 스크립트를 생성합니다 . 스크립트가 실행되면 시작 창에 지정된 각 열에 대한 새 계산식 열을 생성합니다 . 각 새 열은 원래 열의 값이 이상치 제한 범위 내에 있는 경우 해당 값을 포함하고 그렇지 않은 경우 결측으로 설정됩니다 . 새 열은 원래 열과 구분하기 위해 사용자가 지정한 이름을 접두사 또는 접미사로 사용합니다 . 기본적으로 접미사는 " 선별 " 로 설정됩니다 .

다시 스캔 이상치 처리 동작을 수행한 후 데이터를 다시 스캔합니다 .

참고 : 모든 명령 그룹에서 다시 스캔하려면 Ctrl 키를 누른 채 **다시 스캔**을 클릭합니다 .

닫기 " 분위수 범위 이상치 " 패널을 닫습니다 .

참고 : 모든 명령 창을 닫으려면 Ctrl 키를 누른 채 **닫기**를 클릭합니다 .

분위수 범위 이상치 보고서

"분위수 범위 이상치" 보고서에는 지정된 옵션을 사용하여 찾은 이상치가 있는 모든 열이 나열됩니다 . 이 보고서에는 상위 분위수 / 하위 분위수 및 낮은 임계 / 높은 임계에 대한 값이 표시됩니다 . 이러한 임계값 한계를 벗어나는 값을 이상치로 간주합니다 . 각 열의 이상치 수가 표시됩니다 . 각 이상치 값이 보고서의 마지막 열에 나열됩니다 . 한 열에 두 번 이상 나오는 이상치는 괄호 안에 개수가 표시됩니다 . 이상치가 없는 열을 보고서에서 제거하려면 **이상치를 포함하는 열만 표시**를 선택합니다 .

이 보고서를 읽을 때 몇 가지 주의할 항목이 있습니다 .

- 오류 코드 . 일부 연속형 데이터의 경우 비정상적으로 큰 정수 값은 오류 코드일 가능성이 높습니다 . 예를 들어 상위 및 하위 분위수 값이 모두 0.5 미만인 경우 1049 또는 -777 과 같은 이상치는 오류 코드일 수 있습니다 .
- 0. 경우에 따라 0 이 결측값을 나타낼 수 있습니다 . 데이터의 대부분이 상당히 크고 0 이 이상치로 나타나면 결측 데이터가 원인일 가능성이 높습니다 .

9 모음값 보고서

"분위수 범위 이상치" 창의 "9모음값" 보고서에는 확률이 높은 결측값 코드가 포함된 열 목록이 표시됩니다. 이러한 결측값 코드는 연속된 9(보통 9999)로 표시되며, 모두 9이고 상위 분위수보다 높은 가장 큰 숫자입니다. 그 수가 많으면 이러한 이상치가 실제로 결측값 코드일 가능성이 높습니다. 그 수가 매우 적으면 추가 분석을 통해 해당 값이 이상치인지 아니면 결측값 코드인지 판별해야 합니다. "9모음값" 보고서에는 상위 분위수 값이 포함됩니다.

이 보고서는 확률이 높은 결측값 코드가 식별된 경우에만 표시됩니다 .

결측값 코드에 가장 큰 9 모음값 추가 선택한 이상치 값을 결측값 코드 열 특성에 추가합니다 . " 분위수 범위 이상치 " 보고서를 업데이트하려면 **다시 스캔**을 클릭해야 합니다 .

> **참고 :** 시작 창에서 기준 변수를 지정한 경우 " 분위수 범위 이상치 " 에서 " 결측값 코드에 가장 큰 9 모음값 추가 " 를 사용할 수 없습니다 .

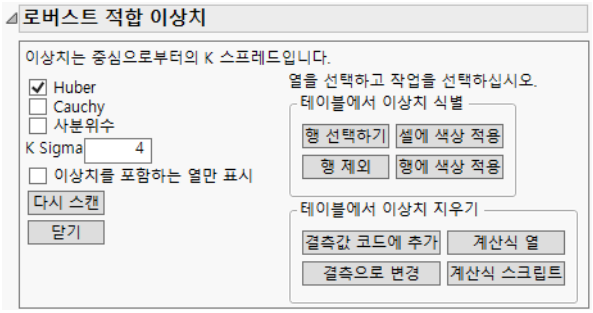
가장 높은 9 모음값을 결측값으로 변경 선택한 이상치 값을 데이터 테이블의 결측값으로 바꿉니다 .

참고 : 데이터 변경을 위해 처음으로 작업 (예 : **결측으로 변경** 또는 **행 제외**) 을 선택하면 원래 데이터의 복사본을 유지하기 위해 **다른 이름으로 저장** 명령을 사용하여 데이터 테이블을 새 파일로 저장하라는 경고 창이 표시됩니다 . 이 창이 나타나면 **확인**을 클릭합니다 . 새 데이터 파일을 저장하기로 결정하면 파일을 새 이름으로 저장하라는 메시지가 자동으로 표시됩니다 .

로버스트 적합 이상치

모수의 로버스트 추정값은 비로버스트 추정값보다 이상치에 덜 민감합니다. "로버스트 적합 이상치"에서는 데이터의 중심 및 산포에 대한 여러 유형의 로버스트 추정값을 제공하여 이상치 식별을 위한 임계값을 결정합니다.

그림 21.7 로버스트 적합 이상치 창



로버스트 적합 이상치 옵션

중심 및 산포에 대한 로버스트 추정값이 주어지면 이상치는 로버스트 중심에서 로버스트 산포의 *K* 배 이상 떨어진 값으로 정의됩니다 . " 로버스트 적합 이상치 " 창에서는 로버스트 추정값 및 승수 *K* 를 계산하기 위한 몇 가지 옵션과 발견된 이상치를 관리하는 도구를 제공합니다 .

Huber Huber M- 추정을 사용하여 중심과 산포를 추정합니다 . 이 옵션이 기본값입니다 . 자세한 내용은 Huber 와 Ronchetti 연구 자료 (2009) 에서 확인하십시오 .

Cauchy Cauchy 분포를 가정하여 중심 및 산포 추정값을 계산합니다 . Cauchy 추정값은 높은 붕괴점을 가지며 일반적으로 Huber 추정값보다 더 로버스트합니다 . 그러나 데이터가 군집으로 분리된 경우 Cauchy 분포는 군집을 더 가깝게 만드는 데이터의 절반만 고려하고 나머지는 무시하는 경향이 있습니다 .

사분위수 IQR(사분위수 범위) 을 사용하여 산포를 추정합니다 . 중심 추정값은 중앙값입니다 . 산포 추정값은 IQR 을 1.34898 로 나눈 값입니다 . 데이터가 정규 분포를 따르는 경우 IQR 을 이 인수로 나누면 산포가 1 표준편차에 해당합니다 .

K 중심에서 산포의 *K* 배 이상 떨어진 값을 이상치로 결정하는 승수입니다 . *K* 값이 크면 작은 값보다 더 보수적인 이상치 집합이 제공됩니다 . 기본값은 4 입니다 .

이상치를 포함하는 열만 표시 보고서의 열 목록을 이상치가 포함된 열로 제한합니다 .

사용자가 지정한 내용에 따라 보고서가 표시되면 다양한 방법을 사용하여 이러한 극단값을 탐색할 수 있습니다 . " 로버스트 추정값 및 이상치 " 보고서에서 지정된 행을 선택하여 행의 이상치를 선택할 수 있습니다 .

팁 : 보고서에서 열을 선택하지 않고 다음 버튼 중 하나를 클릭하면 모든 열을 선택할 수 있는 "JMP 경고 " 가 나타납니다 .

행 선택하기 데이터 테이블에서 선택한 열의 이상치가 포함된 행을 선택합니다 .

행 제외 데이터 테이블에서 선택한 열의 이상치에 대해 행 제외 상태를 설정합니다 . " 로버스트 추정값 및 이상치 " 보고서를 업데이트하려면 **다시 스캔**을 클릭합니다 .

셀에 색상 적용 데이터 테이블에서 선택한 이상치의 셀에 색상을 적용합니다 .

행에 색상 적용 데이터 테이블에서 선택한 열의 이상치가 포함된 행에 색상을 적용합니다 .

결측값 코드에 추가 선택한 이상치를 선택한 열의 결측값 코드 열 특성에 추가합니다 . 이 옵션을 사용하면 데이터 내에서 알려진 결측값 또는 오류 코드를 식별할 수 있습니다 . " 로버스트 추정값 및 이상치 " 보고서를 업데이트하려면 **다시 스캔**을 클릭합니다 .

참고 : 시작 창에서 기준 변수를 지정한 경우 " 로버스트 적합 이상치 " 에서 " 결측값 코드에 추가 " 를 사용할 수 없습니다 .

결측으로 변경 데이터 테이블에서 이상치 값을 결측값으로 변경합니다 . " 로버스트 추정값 및 이상치 " 보고서를 업데이트하려면 **다시 스캔**을 클릭합니다 .

계산식 열 시작 창에 지정된 각 열에 대한 새 계산식 열을 생성합니다 . 각 새 열은 원래 열의 값이 이상치 제한 범위 내에 있는 경우 해당 값을 포함하고 그렇지 않은 경우 결측으로 설정됩니다 . 새 열은 원래 열과 구분하기 위해 사용자가 지정한 이름을 접두사 또는 접미사로 사용합니다 . 기본적으로 접미사는 " 선별 " 로 설정됩니다 .

계산식 스크립트 데이터 테이블에 추가되는 스크립트를 생성합니다 . 스크립트가 실행되면 시작 창에 지정된 각 열에 대한 새 계산식 열을 생성합니다 . 각 새 열은 원래 열의 값이 이상치 제한 범위 내에 있는 경우 해당 값을 포함하고 그렇지 않은 경우 결측으로 설정됩니다 . 새 열은 원래 열과 구분하기 위해 사용자가 지정한 이름을 접두사 또는 접미사로 사용합니다 . 기본적으로 접미사는 " 선별 " 로 설정됩니다 .

다시 스캔 이상치 처리 동작을 수행한 후 데이터를 다시 스캔합니다 .

참고 : 모든 명령 그룹에서 다시 스캔하려면 Ctrl 키를 누른 채 **다시 스캔**을 클릭합니다 .

닫기 " 로버스트 적합 이상치 " 패널을 닫습니다 .

참고 : 모든 명령 창을 닫으려면 Ctrl 키를 누른 채 **닫기**를 클릭합니다 .

로버스트 PCA 이상치

로버스트 PCA 이상치 유틸리티를 사용하면 상관된 다변량 데이터에서 이상치 셀을 빠르게 식별할 수 있습니다 . 다른 많은 다변량 방법은 이상치 행만 식별하므로 이 방법이 유용합니다 . 이 방법을 데이터에 적용하기 전에 먼저 열을 중심화 (선택적) 하고 척도화합니다 . 척도 인수는 다음과 같이 정의됩니다 .

max [Q(.75) - Q(.50), Q(.50) - Q(.25)]/[normalQuantile(0.75)]

여기서 ,

Q(*p*) 는 *p* 번째 분위수입니다 .

참고 : Q(75) 또는 Q(25) 가 중앙값과 같으면 0 이 아닌 범위가 될 때까지 더 극단적인 분위수가 사용됩니다 .

데이터 중심화 및 척도화가 완료되면 로버스트 PCA 이상치 유틸리티에서 특이값 분해 및 임계 지정 단계를 순차적으로 수행하여 데이터 행렬을 분해합니다 . 데이터는 낮은 계수 행렬과 잔차

희소 행렬로 분해됩니다. 임계값은 잔차가 매우 크거나 이상치가 되도록 또는 이상치가 아닌 경우 0에 매우 가깝도록 지정됩니다. 이 알고리즘은 이상치나 작은 잡음 없이 체계적인 변동을 포착하는 데 적절한 행렬 순위를 결정합니다. 낮은 계수 공간에 없는 이상치는 잔차를 기반으로 감지됩니다. 자세한 내용은 Candes et al(2009) 및 Lin et al(2013) 연구 자료에서 확인하십시오. 결측값이 있는 경우 처음에는 중심화 및 척도화 단계 후에 0으로 대체됩니다. 그런 다음 각 SVD(특이값 분해) 반복 후 SVD의 예측값으로 결측값이 업데이트됩니다.

로버스트 PCA 이상치 보고서

명령 목록에서 " 로버스트 PCA 이상치 " 를 선택하는 경우 람다 값을 지정하고 데이터를 중심화할지 여부를 선택해야 합니다. Shift 키를 누른 채 " 로버스트 PCA 이상치 " 버튼을 클릭하면 다음 옵션도 사용할 수 있습니다.

람다 잔차 행렬의 희소성을 결정하는 값을 지정합니다. 람다 값이 클수록 잔차 행렬의 희소성도 커집니다. n 개의 훈련 행과 p 개의 열이 있는 데이터 테이블의 경우 기본 람다 값은 다음과 같이 정의됩니다.

$$\lambda = \frac{2}{\sqrt{\max(n, p)}}$$

최대 반복 최대 SVD 반복 수를 지정합니다.

수렴 기준 알고리즘의 중지 시기를 결정합니다.

이상치 임계값 " 큰 잔차 셀 " 테이블에 표시되는 이상치를 결정하는 이상치 임계값을 지정합니다. 척도화된 잔차가 다음 값보다 크면 관측값이 표시됩니다.

min[0.99 × max{abs(residuals)}, 이상치 임계값]

이상치 임계값의 기본값은 2 입니다.

중심화 로버스트 PCA 이상치 알고리즘을 수행하기 전에 데이터를 중심화할지 여부를 결정합니다.

참고 : 행 수가 10 보다 작거나 같으면 데이터가 중심화되지 않습니다.

척도화 로버스트 PCA 이상치 알고리즘을 수행하기 전에 데이터를 척도화할지 여부를 결정합니다.

참고 : 행 수가 10 보다 작거나 같으면 데이터가 척도화되지 않습니다.

" 로버스트 PCA 이상치 " 보고서에는 이 방법에 대한 정보를 제공하는 테이블이 있습니다. 이 테이블에는 낮은 계수 행렬의 순위 , SVD 반복 수 , 수렴 기준 , 람다 값 및 대치된 결측값 수가 포함됩니다. 다음 보고서와 테이블도 보고서에 포함됩니다.

큰 잔차 셀 가장 큰 이상치 셀을 보여 주는 테이블입니다. 표시되는 셀 수는 이상치 임계값에 의해 결정됩니다. 이 테이블에는 셀의 열 이름과 행 번호 , 잔차 값 및 척도화된 잔차 값이 포함됩니다.

팁 : 데이터 테이블의 특정 이상치 셀에 색상을 적용하려면 " 큰 잔차 셀 " 테이블에서 행을 선택하고 **색상 지정**을 클릭합니다 .

행 제곱근 평균 제곱 데이터 테이블의 각 행에 대한 제곱근 평균 제곱 값을 보여 주는 테이블입니다 . 제곱근 평균 제곱은 척도화된 잔차를 사용하여 계산됩니다 .

팁 : " 행 제곱근 평균 제곱 " 테이블에서 행을 선택하면 데이터 테이블에서 해당 행이 선택됩니다 .

열 제곱근 평균 제곱 시작 창에 지정된 각 열에 대한 제곱근 평균 제곱 값을 보여 주는 테이블입니다 . 제곱근 평균 제곱은 척도화된 잔차를 사용하여 계산됩니다 .

팁 : " 열 제곱근 평균 제곱 " 테이블에서 행을 선택하고 **열 선택**을 클릭하면 데이터 테이블에서 해당 열이 선택됩니다 .

스냅샷 데이터 테이블에 있는 이상치 셀의 그래픽 표현입니다 . 이상치 셀은 빨간색으로 표시됩니다 .

잔차 행렬 분해의 잔차 행렬입니다 . 척도화된 잔차의 절대값이 다음 값보다 크면 셀에 색상이 적용됩니다 .

$$\min[0.99 \times \max\{\text{abs}(\text{residuals})\}, \text{이상치 임계값}]$$

낮은 계수 근사 행렬 분해의 척도화된 잔차 행렬입니다 .

특이값 SVD 의 특이값 벡터입니다 .

로버스트 PCA 이상치 옵션

"로버스트 PCA 이상치" 보고서 아래쪽에 보고서의 각 항목을 저장하는 옵션을 제공하는 버튼이 있습니다.

닫기 " 로버스트 PCA 이상치 " 보고서를 닫습니다 .

큰 이상치 저장 " 큰 잔차 셀 " 테이블의 정보를 새 데이터 테이블에 저장합니다 .

정리된 항목 저장 임계값을 기반으로 이상치를 정리하고 새 열을 데이터 테이블에 저장하는 몇 가지 기법을 제공하는 창을 엽니다 .

절사 해당하는 척도화된 절대 잔차가 지정된 임계값보다 크면 이상치 셀을 제거합니다 . 기본적으로 임계값은 10 입니다 . 이상치 셀을 빨간색으로 표시하려면 " 색상 " 을 선택합니다 . 절사 셀은 비척도화 임계값으로 설정됩니다 .

결측값 대치 해당하는 척도화된 절대 잔차가 지정된 임계값보다 크면 이상치 셀을 낮은 계수 근사의 값으로 설정합니다 . 기본적으로 임계값은 100 입니다 . 해당 셀을 녹색으로 표시하려면 " 색상 " 을 선택합니다 .

결측값 생성 해당하는 척도화된 절대 잔차가 지정된 임계값보다 크면 이상치 셀을 결측으로 설정합니다 . 기본적으로 임계값은 1000 입니다 . 해당 셀을 파란색으로 표시하려면 " 색상 " 을 선택합니다 .

결측값의 대치된 값 색상 적용 이 옵션을 선택하면 원래 있던 결측값이 대치된 셀에 색상을 적용합니다.

잔차 저장 잔차를 원래 데이터 테이블의 새 열에 저장합니다.

척도화된 잔차 저장 척도화된 잔차를 원래 데이터 테이블의 새 열에 저장합니다.

낮은 계수 근사 저장 낮은 계수 근사를 원래 데이터 테이블의 새 열에 저장합니다.

K 최근접 이웃 이상치

K 최근접 이웃 이상치를 사용하면 최근접 이웃과의 거리를 기반으로 이상치를 식별할 수 있습니다. 각 *k* 값에 대해 K 최근접 이웃 이상치 유틸리티는 각 점에서 *k* 번째 최근접 이웃까지의 유클리드 거리 그림을 표시합니다. *K* 로 표시되는 가장 큰 *k* 값을 지정합니다. 그림이 너무 많이 표시되지 않도록 피보나치 수열을 사용하여 *k* = 1,2,3,5,...,*K* 에 대한 그림이 제공됩니다.

최근접 이웃을 계산하기 전에 열을 중심화하고 척도화합니다. 척도 인수는 다음과 같습니다.

max [Q(.75) - Q(.50), Q(.50) - Q(.25)]/[normalQuantile(0.75)]

여기서,

Q(*p*) 는 *p* 번째 분위수입니다.

참고 : Q(75) 또는 Q(25) 가 중앙값과 같으면 0 이 아닌 범위가 될 때까지 더 극단적인 분위수가 사용됩니다.

이 방법은 지정된 *k* 값에 민감합니다. *k* 값이 작으면 점을 이상치로 식별하지 못할 수 있고 *k* 값이 크면 점을 이상치로 잘못 분류할 수 있습니다.

- 소수의 이웃만 연구하기 위해 *K* 값을 작게 지정하는 경우를 가정해 보겠습니다. *K* 개를 초과하는 점을 포함하고 나머지 점에서 멀리 떨어져 있는 군집의 경우 군집 내의 점은 최근접 이웃과의 거리가 짧습니다. 이상치 군집을 감지하지 못할 수 있습니다.
- 많은 수의 이웃을 연구하기 위해 *K* 값을 크게 지정하는 경우를 가정해 보겠습니다. *K* 개 미만의 데이터 점을 포함하는 군집의 경우 이러한 군집 내의 점은 이상치일 수 있습니다. 군집이 점으로 구성된다는 점을 간과하고 대신 개별 군집 멤버를 이상치로 해석할 수 있습니다.

K 최근접 이웃 이상치 보고서

명령 목록에서 "K 최근접 이웃 이상치"를 선택하는 경우 고려할 가장 먼 이웃의 상한으로 사용하기 위한 *K* 값을 지정해야 합니다. 결측값 대치 여부도 지정해야 합니다. 기본적으로 *K* 는 8 로 설정되고 "결측값 대치"가 선택되어 있습니다.

이 보고서에는 선택한 *k* 값에 대한 그림이 표시됩니다(최대 *K* 값까지). 각 그림의 *k* 값은 세로 축 라벨에 표시됩니다. 라벨은 "이웃 k까지의 거리=<a>" 형식으로 표시되며, 여기서 "a"는 가장 가까운 *a* 번째 이웃을 나타내는 정수입니다. 각 그림에는 *i* 번째 행의 점에서 *a* 번째 최근접 이웃까지의 거리가 표시됩니다. 여러 *k* 값에 대해 이웃과의 거리가 먼 점은 이상치일 가능성이 높습니다.

그림 위의 버튼으로 다음을 수행합니다 .

선택된 행 제외 선택한 점에 해당하는 행을 추가 분석에서 제외합니다 . 데이터 테이블에서 해당 행에 " 제외 " 행 상태가 할당됩니다 . "K 최근접 이웃 " 보고서를 다시 실행할지 아니면 닫을지 묻는 메시지가 표시됩니다 . 분석을 다시 실행하면 새로운 최근접 이웃이 식별됩니다 . 그림이 업데이트되고 제외된 점이 표시되지 않습니다 .

산점도 행렬 분석의 모든 열에 대한 산점도 행렬이 포함된 별도의 창을 엽니다. 잠재적 이상치를 "K 최근접 이웃 " 그림에서 선택하고 산점도 행렬에서 확인하여 탐색할 수 있습니다 .

NN 거리 저장 각 행에서 *n*번째 최근접 이웃까지의 거리를 데이터 테이블의 새 열로 저장합니다.

닫기 "K 최근접 이웃 " 보고서를 닫습니다 .

가장 큰 이상치

보고서에 가장 큰 이상치 테이블도 포함되어 있습니다 . 이 테이블에는 *K* 번째 최근접 이웃과의 거리가 가장 큰 20 개의 관측값이 포함됩니다 . 테이블에 다음 열이 있습니다 .

행 관측값의 행 번호입니다 .

거리 지정된 행의 관측값과 *K* 번째 최근접 이웃의 거리입니다 . 테이블은 이 열을 기준으로 내림차순으로 정렬됩니다 .

근접 이웃 *k*개의 최근접 이웃에 대한 행 번호를 나열합니다. 첫 번째 행 번호가 가장 가까운 이웃입니다 . 마지막 행 번호는 *K* 번째 최근접 이웃이며 이 관측값과 지정된 행 사이의 거리가 " 거리 " 열에 표시됩니다 .

열 <n> 해당 RMS 값의 열 이름을 지정합니다 .

RMS<n> 각 열에 대한 *k* 개의 최근접 이웃에서 제곱근 평균 제곱 차이를 계산합니다 . 가장 큰 RMS 값 5 개가 순서대로 표시되며 , 여기서 RMS1 이 최대 RMS 값입니다 . *p* 번째 RMS 값은 다음과 같이 계산됩니다 .

$$\mathrm{RMS}_p = \sqrt{\frac{\sum_{k=1}^{K} (D_{p,i} - D_{p,i_k})^2}{K}}$$

다음은 각 요소에 대한 설명입니다 .

D_p = *p* 번째 열

$D_{p,i}$ = *i* 행에 대한 *p* 번째 열의 값

$D_{p,ik}$ = *i* 행의 *k* 번째 최근접 이웃에 대한 *p* 번째 열의 값

참고 : 가장 큰 이상치 테이블에 표시된 " 열 " 및 "RMS" 열의 수는 시작 창에 지정된 열 수와 5 중 더 작은 값입니다 .

이상치 탐색 유틸리티 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

이상치 탐색 유틸리티의 추가 예

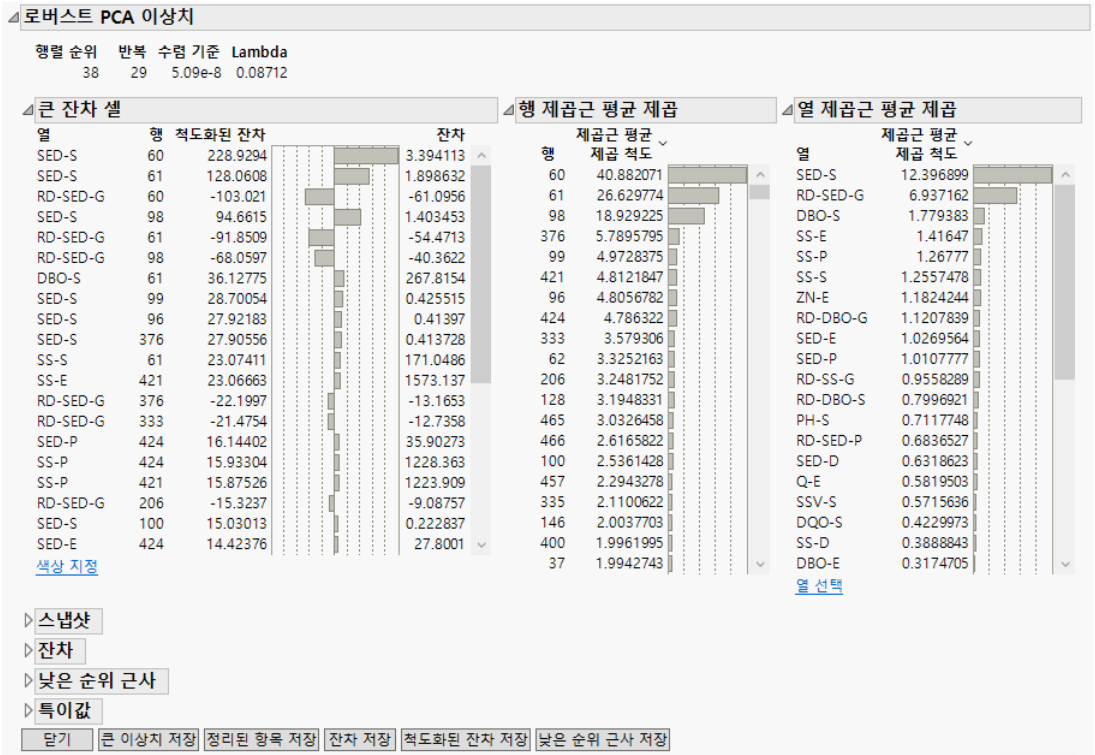
로버스트 PCA 이상치 예

Water Treatment.jmp 데이터 집합에는 도시 폐수 처리장에 있는 38 개 센서의 일별 측정값이 포함되어 있습니다. 이러한 데이터에 잠재적 이상치가 있는지 탐색하려고 합니다. 잠재적 이상치에는 센서 고장, 악천후 및 기타 상황이 포함될 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Water Treatment.jmp 를 엽니다.
2. **분석 > 선별 > 이상치 탐색**을 선택합니다.
3. Sensor Measurements 열 그룹을 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. Shift 키를 누른 채 **로버스트 PCA 이상치**를 클릭합니다.
6. " 이상치 임계값 " 옆에 "10" 을 입력합니다.

   다른 옵션에는 기본값을 사용합니다.
7. **확인**을 클릭합니다.

그림 21.8 로버스트 PCA 이상치 보고서



" 큰 잔차 셀 " 테이블에는 척도화된 잔차가 10 보다 크거나 -10 보다 작은 셀이 표시됩니다 . SED-S 열에 척도화된 잔차가 큰 셀이 여러 개 있습니다 . " 열 제곱근 평균 제곱 " 테이블을 살펴보면 실제로 SED-S 열의 제곱근 평균 제곱 척도 값이 가장 크다는 것을 알 수 있습니다 . 제곱근 평균 제곱 척도 값이 가장 큰 행은 60 행입니다 .

8. 보고서 맨 아래에서 **정리된 항목 저장**을 클릭합니다 .

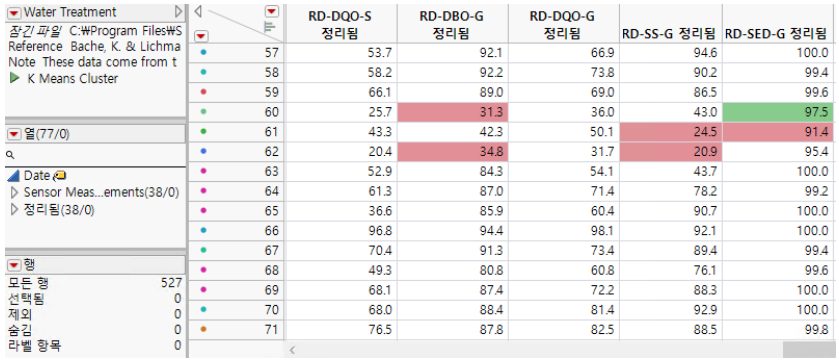
   이 옵션을 사용하면 값을 절사하고 , 새 값을 대치하거나 , 셀을 결측으로 설정하여 이상치가 큰 셀을 조정할 수 있습니다 . " 정리된 열 저장 " 보고서에 이러한 옵션이 제공됩니다 .

9. " 결측값 대치 " 를 선택합니다 . " 절사 " 는 기본적으로 선택되어 있습니다 .

10. **확인**을 클릭합니다 .

    Water Treatment.jmp 데이터 테이블에 38 개의 " 정리됨 " 열이 새로 추가되었습니다 . 60 행의 제곱근 평균 제곱 척도 값이 가장 크다는 것을 알고 있으므로 이 행으로 스크롤하여 절사 및 결측값 대치된 일부 셀을 확인합니다 . 이제 모든 유형의 분석에서 정리된 열을 사용할 수 있습니다 .

그림 21.9 로버스트 PCA 이상치의 정리된 열



	RD-DQO-S 정리됨	RD-DBO-G 정리됨	RD-DQO-G 정리됨	RD-SS-G 정리됨	RD-SED-G 정리됨
57	53.7	92.1	66.9	94.6	100.0
58	58.2	92.2	73.8	90.2	99.4
59	66.1	89.0	69.0	86.5	99.6
60	25.7	31.3	36.0	43.0	97.5
61	43.3	42.3	50.1	24.5	91.4
62	20.4	34.8	31.7	20.9	95.4
63	52.9	84.3	54.1	43.7	100.0
64	61.3	87.0	71.4	78.2	99.2
65	36.6	85.9	60.4	90.7	100.0
66	96.8	94.4	98.1	92.1	100.0
67	70.4	91.3	73.4	89.4	99.4
68	49.3	80.8	60.8	76.1	99.6
69	68.1	87.4	72.2	88.3	100.0
70	68.0	88.4	81.4	92.9	100.0
71	76.5	87.8	82.5	88.5	99.8

결측값 탐색 유틸리티

데이터 집합에 결측값이 있으면 해당 데이터를 사용하여 결론을 내릴 때 영향을 줄 수 있습니다. 예를 들어 건강한 몇몇 참가자가 종단 연구에서 빠지고 그들의 데이터가 계속 누락된 경우 연구 결과는 남아 있는 건강하지 않은 개인에게 편향될 수 있습니다. 결측 데이터 값을 식별해야 하며 추가 분석을 수행하기 전에 해당 값을 파악해야 합니다.

결측값 탐색 유틸리티는 데이터의 결측값을 식별하고 파악할 수 있는 여러 가지 방법을 제공합니다. 또한 결측값에 대해 다변량 정규 대치를 수행하는 방법도 제공합니다. 이러한 결측값 대치법은 데이터의 **결측값이 랜덤**이라고 가정합니다. 즉, 결측 데이터와 비결측 데이터 간의 차이를 연구에 포함된 다른 변수의 값으로 설명할 수 없다는 의미입니다. 결측값이 랜덤하지 **않다고** 의심되면 여러 플랫폼에서 사용할 수 있는 "결측값 정보화" 절차를 사용해 보십시오. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

주의: 결측값 대치 후 데이터를 분석할 때는 결과가 편향될 가능성이 있으므로 주의해야 합니다.

결측값 탐색 유틸리의 예

Arrhythmia.jmp 샘플 데이터 테이블에는 환자 ECG(심전도) 정보 452개가 포함되어 있습니다. 데이터는 원래 여러 ECG 패턴을 심장 부정맥으로 분류하기 위해 수집되었습니다. 그러나 이 데이터 테이블에 결측값이 있습니다. 주로 이러한 결측값을 탐색하고 필요한 경우 결측값을 대치하는 데 관심이 있습니다. 연속형 모델링 유형의 열에만 결측값 대치를 수행할 수 있으므로 두 단계로 분석을 수행합니다.

결측값 검토

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Arrhythmia.jmp 를 엽니다.
2. **분석 > 선별 > 결측값 탐색**을 선택합니다.
3. 모든 열 (총 280 개) 을 선택하고 **Y, 열**을 클릭합니다.

4. **확인**을 클릭합니다 . **결측값을 포함하는 열만 표시** 체크박스를 선택합니다 .

그림 21.10 결측값 보고서

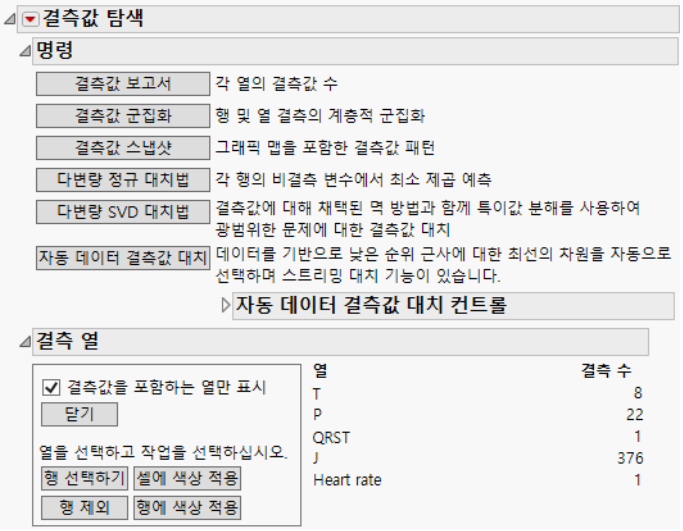


그림 21.10 에 표시된 " 결측 열 " 보고서에서는 5 개 열에만 결측 데이터가 있음을 나타냅니다 . 총 452 개 행 중에서 J 열에 376 개의 결측값이 있습니다 . 대부분의 값이 결측이므로 대치된 값을 사용한 분석 결과가 의미 없을 수 있습니다 . 이러한 데이터의 경우 " 결측값 정보화 " 옵션을 지원하는 플랫폼에서 J 열에 대해 " 결측값 정보화 " 옵션을 사용하여 모형을 탐색할 수 있습니다 .

두 가지 결측값 대치 옵션인 " 다변량 정규 대치법 " 과 " 다변량 SVD 대치법 " 이 표시되지 않습니다 . 분석에 포함된 일부 열이 범주형이므로 결측값 대치를 사용할 수 없다는 메시지가 표시됩니다 . 수치형이지만 모델링 유형이 명목형인 열이 데이터 테이블에 일부 포함되어 있습니다 . 이러한 열은 결측값 대치에 사용할 수 없습니다 .

결측값 대치

결측값이 있는 5 개 열이 연속형입니다 . 데이터 테이블의 연속형 열에 대해 다변량 대치법을 사용하여 J 열을 제외한 4 개 열의 결측값을 대치합니다 . 이렇게 하면 값이 결측일 확률은 제외된 명목형 변수의 값이 아니라 연속형 변수의 값에만 의존한다고 암묵적으로 가정합니다 . 이 새 분석을 수행하려면 결측값 탐색 유틸리티를 다시 시작해야 합니다 .

1. **분석 > 선별 > 결측값 탐색**을 선택합니다 .
2. 시작 창에서 **280 개 열** 옆의 빨간색 삼각형을 클릭합니다 .

   열 필터 메뉴를 사용하여 " 열 선택 " 목록에서 연속형 모델링 유형의 열만 표시합니다 .
3. **모델링 유형 > 모두 선택 해제**를 선택합니다 .

   이렇게 하면 " 열 선택 " 목록에서 모든 열이 제거됩니다 .
4. **모델링 유형 > 연속형**을 선택합니다 .

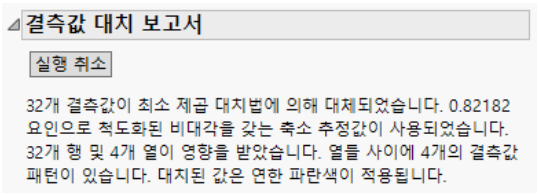
이제 " 열 선택 " 목록에 207 개의 연속형 열만 포함됩니다 .

5. 207 개 열을 모두 선택합니다 . 그런 다음 Ctrl 키를 누른 채 J 열을 클릭하여 선택 취소하고 **Y, 열**을 클릭합니다 .
6. **확인**을 클릭합니다 .
7. **다변량 정규 대치법**을 클릭합니다 .

   공분산에 대해 축소 추정량을 사용할지 묻는 창이 나타납니다 .
8. **예 (축소)** 를 클릭합니다 .

   원래 데이터를 유지하려면 **다른 이름으로 저장** 명령을 사용해야 함을 알리는 "JMP 경고 " 가 나타납니다 .
9. **확인**을 클릭합니다 .

그림 21.11 결측값 대치 보고서



" 결측값 대치 보고서 " 에는 대치된 결측값 수와 결측값 대치 관련 상세 정보가 표시됩니다 . 결측값이 있었던 4 개 열에 결측 데이터가 남아 있지 않습니다 .

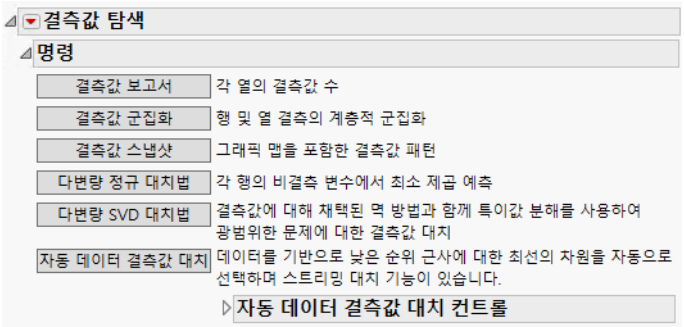
결측값 탐색 유틸리티 시작

분석 > 선별 > 결측값 탐색을 선택하여 결측값 탐색 모델링 유틸리티를 시작합니다. 관심 열을 "Y, 열" 목록에 입력합니다. **기준** 변수도 지정할 수 있습니다.

참고 : 결측값 탐색 유틸리티에서는 모델링 유형이 수치형인 열만 입력할 수 있습니다 .

결측값 보고서

그림 21.12 Arrhythmia.jmp 의 연속형 변수에 대한 결측값 보고서



시작 창에서 " 확인 " 을 클릭하면 보고서가 열리고 " 명령 " 개요와 " 결측 열 " 보고서가 표시됩니다 . 다음과 같은 명령이 포함되어 있습니다 .

- " 결측값 보고서 "
- " 결측값 군집화 "
- " 결측값 스냅샷 "
- " 다변량 정규 대치법 " (시작 창에서 명목형 또는 순서형 모델링 유형의 숫자 열을 입력한 경우에는 사용 불가능)
- " 다변량 SVD 대치법 " (시작 창에서 명목형 또는 순서형 모델링 유형의 숫자 열을 입력한 경우에는 사용 불가능)
- JMP PRO " 자동 데이터 결측값 대치 " (시작 창에서 명목형 또는 순서형 모델링 유형의 숫자 열을 입력한 경우에는 사용 불가능)

팁 : 기준 변수의 모든 수준에서 결측값 명령을 실행하려면 Ctrl 키를 누른 채 원하는 명령 버튼을 클릭합니다 .

결측값 보고서

" 결측값 보고서 " 는 각 열의 이름과 해당 열의 결측값 수를 나열하는 " 결측 열 " 보고서를 엽니다.

결측값을 포함하는 열만 표시 결측값이 없는 열을 목록에서 제거합니다 .

닫기 " 결측 열 " 보고서를 닫습니다 .

행 선택하기 데이터 테이블에서 " 결측 열 " 보고서에서 선택한 열의 결측값이 포함된 행을 선택합니다 .

행 제외 데이터 테이블에서 " 결측 열 " 보고서에서 선택한 열의 결측값이 포함된 행에 " 제외 " 행 상태를 적용합니다 .

셀에 색상 적용 데이터 테이블에서 " 결측 열 " 보고서에서 선택한 열의 결측값이 포함된 셀에 색상을 적용합니다 .

행에 색상 적용 데이터 테이블에서 " 결측 열 " 보고서에서 선택한 열의 결측값이 포함된 행에 색상을 적용합니다 .

결측값 군집화

" 결측값 군집화 " 는 결측 데이터에 대한 계층적 군집화 분석을 제공합니다 .

- 그림 오른쪽의 덴드로그램에는 결측 데이터 패턴 행의 군집이 표시됩니다. 이러한 행은 테이블 > 결측 데이터 패턴을 사용하여 얻을 수 있습니다 .
- 그림 아래의 덴드로그램에는 변수 군집이 표시됩니다 .

이 보고서를 사용하여 특정 열 그룹의 결측값 패턴이 비슷한지 여부를 확인할 수 있습니다.

그림의 행은 결측 데이터 패턴으로 정의되며 각 패턴에 대한 행이 있습니다. 열은 변수에 해당합니다. 빨간색 각 셀은 그림 아래에 나열된 열의 결측값 그룹을 나타냅니다. 셀을 커서로 가리키면 표시된 값 목록을 볼 수 있습니다. 그림을 클릭하면 결측 데이터 패턴 행이 선택됩니다. 선택된 패턴을 나타내는 세로 막대가 표시됩니다.

결측값 스냅샷

"결측값 스냅샷"은 결측값에 대한 셀 그림을 표시합니다. 열은 변수를 나타냅니다. 검은색 셀은 결측값을 나타냅니다. 이 그림은 데이터 수집 기간이 끝나기 전에 피실험자가 연구에서 빠질 수 있는 종단 데이터의 결측을 이해하는 데 특히 유용합니다.

다변량 정규 대치법

다변량 정규 대치법 유틸리티는 다변량 정규 분포를 기반으로 결측값을 대치합니다. 절차를 수행하려면 모든 변수의 모델링 유형이 연속형이어야 합니다. 이 알고리즘은 최소 제곱 대치법을 사용합니다. 공분산 행렬은 쌍별 공분산을 사용하여 구성됩니다. 대각 항목(분산)은 각 변수에 대한 모든 비결측 값을 사용하여 계산됩니다. 두 변수의 비대각 항목은 두 변수에 대해 비결측인 모든 관측값을 사용하여 계산됩니다. 공분산 행렬이 특이 행렬인 경우 이 알고리즘은 Moore-Penrose 유사역행렬을 기반으로 하는 최소-노름 최소 제곱 대치법을 사용합니다.

다변량 정규 대치법 옵션을 사용하면 공분산에 대한 축소 추정량을 사용할 수 있습니다. 축소 추정량 사용은 공분산 행렬의 추정을 개선할 수 있는 방법입니다. 축소 추정량에 대한 자세한 내용은 Schäfer and Strimmer(2005) 연구 자료에서 확인하십시오.

참고 : 검증 열이 지정되면 훈련 데이터 집합의 관측값을 사용하여 공분산 행렬이 계산됩니다 .

다변량 정규 대치법 보고서

결측값 대치 보고서에 다변량 대치법 프로세스의 결과가 설명됩니다 . 다음 내용이 결과에 포함됩니다 .

- 결측값 대치법 (최소 제곱 또는 최소 - 노름 최소 제곱)
- 대체된 값의 수
- 축소 추정량 사용 여부
- 비대각 항목의 척도화 인수
- 영향을 받은 행과 열의 수
- 서로 다른 결측값 패턴의 수

결측값 대치가 완료되면 데이터 테이블에서 대치된 값에 해당하는 셀이 연한 파란색으로 표시됩니다 . " 결측 열 " 보고서가 열려 있는 경우 결측값을 표시하지 않도록 업데이트됩니다 .

결측값 대치를 취소하고 대치된 데이터를 결측값으로 바꾸려면 **실행 취소**를 클릭합니다 .

다변량 SVD 대치법

다변량 SVD 대치법 유틸리티는 SVD(특이값 분해)를 사용하여 결측값을 대치합니다 . 이 유틸리티는 변수의 수가 수백 또는 수천 개인 데이터에 유용합니다 . SVD 계산에는 공분산 행렬 계산이 필요하지 않으므로 변수가 많이 포함된 광범위한 문제의 경우 SVD 방법이 권장됩니다 . 절차를 수행하려면 모든 변수의 모델링 유형이 연속형이어야 합니다 .

특이값 분해는 관측값 행렬 **X** 를 **X** = **UDV**′ 로 나타냅니다 . 여기서 **U** 와 **V** 는 직교 행렬이고 **D** 는 대각 행렬입니다 .

다변량 SVD 대치법 유틸리티에서 기본적으로 사용되는 SVD 알고리즘은 IRLBA(Implicitly Restarted Lanczos Bidiagonalization) 방법이라고도 하는 희소 Lanczos 방법입니다 . 자세한 내용은 Baglama 와 Reichel(2005) 연구 자료에서 확인하십시오 . 이 알고리즘은 다음을 수행합니다 .

1. 각 결측값이 해당 열의 평균으로 대체됩니다 .
2. 관측값 행렬 **X** 에 대해 SVD 분해가 수행됩니다 .
3. 결측값이 있었던 각 셀이 SVD 분해를 통해 얻은 **UDV**′ 행렬의 해당 요소로 대체됩니다 .
4. SVD 가 행렬 **X** 로 수렴할 때까지 2 단계와 3 단계를 반복합니다 .

결측값 대치법 창

" 다변량 SVD 대치법 " 을 클릭하면 " 결측값 대치법 " 창에 권장 설정이 표시됩니다 .

특이 벡터 수 계산된 후 결측값 대치에 사용되는 특이 벡터의 수입니다 .

> **참고 :** 특이 벡터를 너무 많이 지정하지 않아야 합니다 . 그렇지 않으면 반복되는 동안 SVD 와 결측값 대치가 변하지 않습니다 .

최대 반복 수 결측값 대치에 사용되는 반복 수입니다 .

반복 로그 표시 반복 횟수를 표시하고 기준에 대한 정보를 제공하는 " 상세 정보 " 보고서를 엽니다 .

규모가 큰 문제의 경우 SVD가 완료된 차원 수를 보여 주는 진행률 표시줄이 나타납니다. 언제든지 결측값 대치를 중지하고 해당 차원 수를 사용할 수 있습니다.

다변량 SVD 대치법 보고서

결측값 대치 보고서에 다변량 대치법 프로세스의 결과가 설명됩니다 .

- 결측값 대치법
- 대체된 값의 수
- 영향을 받은 행과 열의 수

결측값 대치가 완료되면 대치된 열에 결측값이 없음을 나타내는 "결측 열" 보고서가 자동으로 표시됩니다. 대치된 값은 연한 파란색으로 표시됩니다.

결측값 대치를 취소하고 대치된 데이터를 결측값으로 바꾸려면 **실행 취소**를 클릭합니다.

JMP PRO 자동 데이터 결측값 대치

ADI(자동 데이터 결측값 대치) 유틸리티는 행렬 완성이라고도 하는 낮은 계수 행렬 근사 방법을 사용하여 결측값을 대치합니다. 훈련된 ADI 모형은 스코어링 계산식을 통해 스트리밍 데이터에 대한 결측 데이터 대치를 수행할 수 있습니다. 스트리밍 데이터는 시간 경과에 따라 이용 가능해지고, 대치된 모형의 조정 또는 검증에 사용되지 않은 관측값 행이 추가된 것입니다. 이 유틸리티는 낮은 계수 근사에 대한 최선의 차원을 선택할 수 있도록 유연하고, 로버스트하며, 자동화되었습니다. 이러한 기능을 통해 ADI가 다양한 유형의 데이터 집합에 대해 잘 작동할 수 있습니다.

행렬의 낮은 계수 근사는 $\mathbf{X} = \mathbf{UDV}'$ 형식이며 SVD(특이값 분해) 의 확장으로 간주할 수 있습니다 . ADI 는 Soft-Impute 방법을 결측값 대치 모형으로 사용하며 , 데이터가 낮은 계수 근사의 순위를 결정하도록 되어 있습니다 .

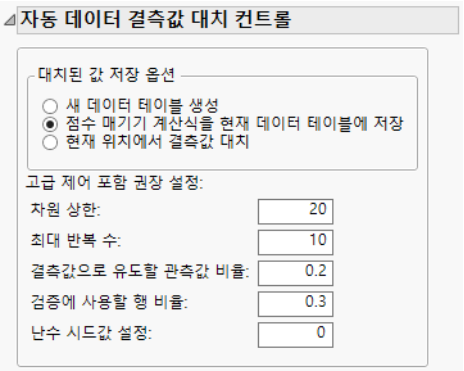
ADI 알고리즘은 다음 단계를 수행합니다 .

1. 데이터가 훈련 데이터 집합과 검증 데이터 집합으로 분할됩니다 .
2. 훈련 데이터 집합의 관측값을 사용하여 각 집합이 중심화 및 척도화됩니다 .
3. 분할된 각 데이터 집합에 대해 IM(유도된 결측값) 이라고 하는 추가 결측값이 각 열 내에 더해집니다 .
4. 조정 모수의 해 경로를 따라 결측값 대치 모형이 훈련 데이터 집합에 대해 적합됩니다. IM 값은 조정 모수의 최량 값을 결정하는 데 사용됩니다 .
5. 4 단계에서 선택한 결측값 대치 모형 결과의 편향을 제거하여 훈련 데이터 집합을 사용한 순위 축소가 추가로 수행됩니다 .
6. 스트리밍 데이터에 대한 모형을 보정하고 과대적합을 방지하기 위해 최종 순위 축소가 수행됩니다 . 5 단계에서 결정된 순위를 상한으로 사용하고 검증 데이터 집합에 대해 결측값 대치 모형을 적합시켜 이 작업을 완료합니다 .

자동 데이터 결측값 대치 컨트롤

ADI 유틸리티에는 대치된 값을 저장하는 옵션과 고급 컨트롤이 포함되어 있습니다 .

그림 21.13 ADI 컨트롤



대치된 값 저장 옵션 다음 세 가지 옵션을 사용하여 ADI 방법의 대치된 값을 저장할 수 있습니다 .

새 데이터 테이블 생성 원래 데이터 테이블과 동일한 차원의 새 데이터 테이블을 생성합니다 . 새 데이터 테이블에는 시작 창에서 선택한 열에 대치된 값이 포함됩니다 .

스코어링 계산식을 현재 데이터 테이블에 저장 시작 창에 지정한 대치된 열을 포함하는 **결측값 대치됨** 열 그룹을 현재 데이터 테이블에 저장합니다 . **ADI 결측값 대치 열**이라는 숨겨진 열도 현재 데이터 테이블에 추가됩니다 . 이 열에는 데이터 결측값 대치에 사용된 스코어링 계산식과 대치된 벡터가 포함됩니다 . 데이터 테이블에 행이 추가되면 열 계산식이 자동으로 업데이트되므로 스트리밍 데이터에 대한 결측 데이터 대치가 가능합니다 . 이 옵션은 기본 옵션입니다 .

현재 위치에서 결측값 대치 현재 데이터 테이블에서 결측값을 대치합니다 . 대치된 값은 연한 파란색으로 표시됩니다 .

고급 컨트롤 데이터에 따라 권장 설정이 기본값으로 사용되는 다음 고급 컨트롤을 포함합니다 .

차원 상한 낮은 계수 근사에 허용되는 최대 순위를 결정합니다 . 이 값은 선택한 열로 구성되는 행렬의 차원에 의해 결정됩니다 .

최대 반복 수 결측값 대치 모형의 조정 모수를 결정하기 위해 반복되는 값의 수를 결정합니다 . 기본값은 10 입니다 .

결측값으로 유도할 관측값 비율 훈련 및 검증 데이터 집합에 추가되는 IM 값의 비율을 결정합니다 . 각 집합에 대한 기본 비율은 0.2 입니다 .

검증에 사용할 행 비율 훈련 및 검증 데이터 집합에 사용할 행 비율을 결정합니다 . 검증 데이터 집합에 대한 기본 비율은 0.3 입니다 .

난수 시드값 설정 ADI 에 사용할 난수 시드값을 결정합니다 . 재현 가능한 결과를 얻으려면 이 옵션을 사용합니다 .

결측값 탐색 유틸리티 옵션

ADI 적재 행렬 (ADI 유틸리티에만 사용 가능) 각 성분에 대한 요인 적재에 해당하는 열이 포함된 보고서를 표시하거나 숨깁니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

패턴 탐색 유틸리티

데이터가 정리되지 않아 육안 검사만으로 데이터의 타당성을 판단하기 어려울 때가 있습니다 . 패턴 탐색 유틸리티는 데이터에서 비정상적이거나 예기치 않은 패턴을 감지하는 도구입니다 . 우연히 또는 정상 메커니즘에 의해 발생된 패턴이지만 위조되거나 무단 변경된 데이터를 나타낼 수도 있습니다 . 이 유틸리티는 값이 정확하여 매칭과 패턴이 우연히 발생할 가능성이 매우 낮을 때 가장 유용합니다 . 또한 패턴 탐색 유틸리티를 사용하면 큰 데이터 테이블을 효율적으로 처리할 수 있습니다 .

패턴 탐색 유틸리티는 중복된 값과 중복된 값 시퀀스를 찾습니다. 이러한 유형의 패턴은 동일한 값을 여러 번 붙여넣기, 동일한 값을 전체 셀 범위에 붙여넣기 또는 데이터 범위를 복사하여 다른 곳에 붙여넣기와 같은 데이터 변조 전략을 나타낼 수 있습니다. 또한 이 유틸리티는 형식이 지정된 값의 특성을 분석하여 데이터가 실제 측정값인지 아니면 난수 생성기로 생성되었을 수 있는지 판단합니다. 분포 특성을 검토하고 행 그룹에서 열 간의 선형 관계를 감지하는 옵션도 있습니다.

참고: 패턴 탐색 유틸리티는 제외된 행을 무시하지만 결과가 행을 삭제한 경우와 다를 수 있습니다 . " 가장 긴 런 " 및 " 가장 긴 중복 시퀀스 " 옵션과 같은 행 런의 경우 제외된 행이 있으면 런이 중지됩니다 .

패턴 탐색 유틸리티의 예

Nicardipine Lab Patterns.jmp 샘플 데이터 테이블은 니카르디핀 (Nicardipine) 약물 연구에서 채택되었습니다 . 데이터 테이블에는 27 개의 수치형 실험 결과 열이 포함되어 있습니다 . 패턴 탐색 유틸리티를 사용하여 실험 결과에 예기치 않은 패턴이 있는지 여부를 확인합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nicardipine Lab Patterns.jmp 샘플 데이터 테이블을 엽니다 .

2. **분석 > 선별 > 패턴 탐색**을 선택합니다 .
3. Laboratory Results 열 그룹을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 단변량 요약 " 옆의 회색 표시 아이콘을 클릭하여 " 단변량 요약 " 보고서를 엽니다 .

그림 21.14 패턴 탐색 단변량 요약 보고서

단변량 요약

가장 긴 런

열	시작 행	개수	값	희귀도
Creatinine	2265	7	0.03536	17.6
Hematocrit	3048	4	28.999999	12.3
BUN	2267	5	2.142	7.8
Prothrombin Time	2931	6	12	6.7
LDH	3456	3	602	5.8
Activated PTT	1460	3	30.299999	4.8
Platelet	521	3	150	4.8
Platelet	1092	3	336	4.1
Activated PTT	3435	3	29.5	3.3
CO2	1390	4	82.612	3.1
Creatine Kinase	3456	3	63	3.1
Erythrocytes	815	3	4.3000002	3.0

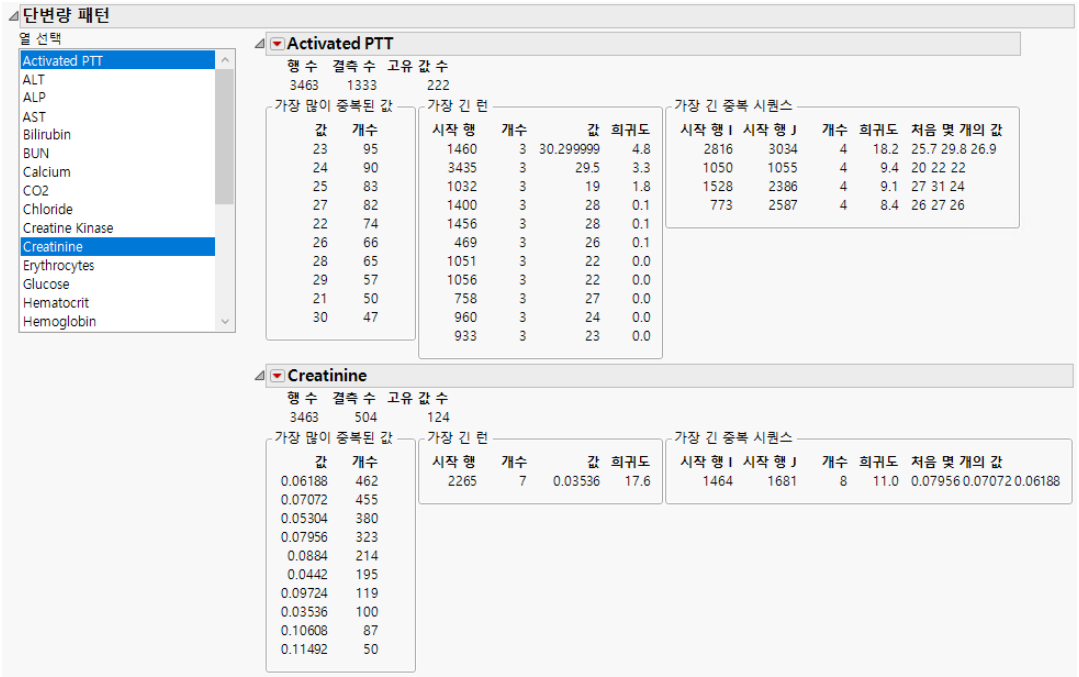
가장 긴 중복 시퀀스

열	시작 행 I	시작 행 J	개수	희귀도
Activated PTT	2816	3034	4	18.2
Prothrombin Time	2942	2966	6	14.9
Hematocrit	3051	3057	3	14.6
Leukocytes	547	1275	3	11.1
Creatinine	1464	1681	8	11.0
Leukocytes	975	1485	3	10.8
ALT	840	2425	3	10.5
Sodium	2339	2976	6	9.8
Leukocytes	456	2805	3	9.8
Activated PTT	1050	1055	4	9.4
Potassium	1046	1067	6	9.4
Hemoglobin	1247	3273	3	9.2

" 단변량 요약 " 보고서에는 모든 열에서 발견된 가장 긴 런과 가장 긴 중복 시퀀스가 요약됩니다 . 각 테이블은 런 또는 시퀀스가 우연히 발생할 가능성을 계산한 희귀도에 따라 정렬됩니다 . 가장 긴 런에서 희귀도 스코어가 가장 높은 열은 Creatinine 인 반면 , 가장 긴 중복 시퀀스에서 희귀도 스코어가 가장 높은 열은 Activated PTT 입니다 . " 단변량 패턴 " 보고서에서 이러한 열에 대해 자세히 살펴보겠습니다 .

6. 처음에는 첫 번째 열인 Activated PTT 에 대한 보고서가 " 단변량 패턴 " 보고서에 표시됩니다 . Ctrl 키를 누른 채 " 열 선택 " 목록에서 Creatinine 을 선택합니다 .

그림 21.15 패턴 탐색 단변량 패턴 보고서

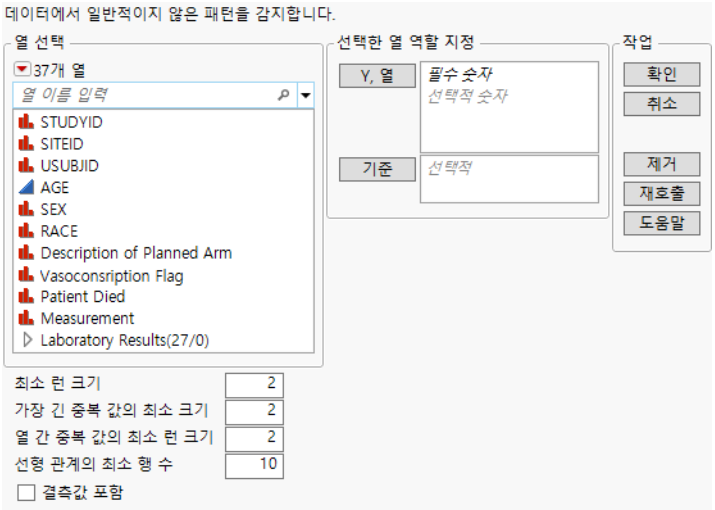


" 열 선택 " 목록에서 열 이름을 선택하면 해당 열 보고서가 표시됩니다 . Activated PTT 보고서에서는 이 열에 개수가 3 인 런과 4 인 시퀀스가 여러 개 포함되어 있음을 보여 줍니다 . Creatinine 보고서에서는 이 열이 " 가장 긴 런 " 에 대해 희귀도 스코어가 가장 높지만 이 열에 런이 하나뿐임을 보여 줍니다 . 이 열에는 개수가 8 인 시퀀스도 하나뿐입니다 .

패턴 탐색 유틸리티 시작

분석 > 선별 > 패턴 탐색을 선택하여 패턴 탐색 유틸리티를 시작합니다.

그림 21.16 패턴 탐색 유틸리티 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 열 분석할 열을 할당합니다. 변수는 수치형이고 연속형이어야 합니다.

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열을 할당합니다. 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다. 결과는 개별 보고서에 제공됩니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

최소 런 크기 행에서 시퀀스를 런으로 보고할 최소 값 수를 지정합니다.

가장 긴 중복 값의 최소 크기 열 내에서 중복 시퀀스로 보고할 시퀀스의 최소 길이를 지정합니다.

열 간 중복 값의 최소 런 크기 열 간에 중복 시퀀스로 보고할 시퀀스의 최소 길이를 지정합니다.

선형 관계의 최소 행 수 행 그룹에서 열 간의 선형 관계를 감지하는 데 사용되는 최소 행 수를 지정합니다.

참고 : 최소 행 수는 3 보다 크거나 같아야 합니다.

결측값 포함 결측값을 패턴 감지에 포함하도록 지정합니다.

패턴 탐색 보고서

초기 " 패턴 탐색 " 보고서에는 " 제어판 ", " 단변량 요약 " 보고서 및 " 단변량 패턴 " 보고서가 포함됩니다. 이러한 보고서의 여러 테이블에는 선택 가능한 행이 포함되어 있습니다. 행을 선택하고 첫 번째 열을 마우스 오른쪽 버튼으로 클릭하면 데이터 테이블에서 관련 행과 열을 선택하거나 색상을 지정할 수 있는 메뉴가 나타납니다.

제어판

제어판에는 보고서 창에 표시할 보고서를 선택할 수 있는 옵션이 있습니다. 이러한 옵션은 "패턴 탐색"의 빨간색 삼각형 메뉴에 있는 옵션과 동일합니다. 자세한 내용은 "패턴 탐색 유틸리티 옵션"에서 확인하십시오. 보고서 설정을 지정하는 제어판 섹션도 있습니다. 이러한 설정은 시작 창의 설정과 동일합니다. 자세한 내용은 "패턴 탐색 유틸리티 시작"에서 확인하십시오. 제어판에서 설정을 변경하면 보고서가 자동으로 업데이트됩니다.

단변량 요약

"단변량 요약" 보고서는 전체 열에 대한 여러 패턴 보고서의 결과를 결합합니다. 기본적으로 "가장 긴 런"과 "가장 긴 중복 시퀀스"에 대한 요약 테이블이 있습니다. 희귀도 ≥ 2인 가장 긴 런 및 가장 긴 중복 시퀀스가 테이블에 표시됩니다.

"분수 길이" 옵션을 선택하면 "분수 길이" 요약 테이블도 표시됩니다. 이 테이블에는 각 열에서 분수 길이가 15보다 크거나 같은 값의 비율이 표시됩니다. 분수 길이가 15보다 크거나 같은 값을 포함하지 않는 열은 "분수 길이" 요약 테이블에 표시되지 않습니다.

하나 이상의 Y 열에 "규격 한계" 열 특성이 있고 "규격 한계에 대한 분포" 옵션을 선택한 경우 "규격 한계 분포" 요약 테이블도 표시됩니다. 이 테이블에는 통계적으로 유의한 수의 관측값이 규격 한계를 벗어나는 열이 표시됩니다.

"단변량 요약" 보고서 내용에 대한 자세한 설명은 "패턴 탐색 유틸리티 옵션"에서 확인하십시오.

단변량 패턴

"단변량 패턴" 보고서에는 "열 선택" 목록과 하나 이상의 열별 패턴 보고서가 포함되어 있습니다. 표시되는 열별 보고서는 "열 선택" 목록에서 결정됩니다. 각 열의 보고서를 검토하려면 첫 번째 열을 선택한 후 아래쪽 화살표 키를 연속 사용하여 다음 열로 이동합니다. 아니면 이 목록에서 열 이름을 클릭하여 선택한 열의 패턴 보고서를 표시할 수도 있습니다. Ctrl 키 또는 Shift 키를 사용하여 둘 이상의 열 이름을 선택하면 선택한 각 열에 대한 패턴 보고서가 표시됩니다. "열 선택" 목록에서 하나 이상의 열을 선택 취소하면 해당 열에 대한 보고서가 제거됩니다.

각 개별 패턴 보고서에는 지정된 열에 대한 행 수, 결측 관측값 수 및 고유 값 수를 포함하는 요약 테이블이 표시됩니다. "가장 많이 중복된 값", "가장 긴 런" 및 "가장 긴 중복 시퀀스" 테이블도 기본적으로 제공됩니다. 제어판이나 "패턴 탐색"의 빨간색 삼각형 메뉴에서 "형식 지정 값 너비", "분수 길이" 또는 "선행 및 후행 숫자" 옵션을 선택하면 해당 테이블이 각 개별 패턴 보고서에 추가됩니다.

"단변량 패턴" 보고서의 테이블에서 행을 마우스 오른쪽 버튼으로 클릭하면 마우스 오른쪽 버튼 클릭 옵션 메뉴가 나타납니다.

행 및 열 선택 "단변량 패턴" 보고서의 특정 테이블에서 선택한 행에 해당하는 열과 행을 원래 데이터 테이블에서 선택합니다. 예를 들어 특정 열에 대한 "선행 숫자" 테이블의 두 번째 행을 마우스 오른쪽 버튼으로 클릭하면 해당 열에서 선행 숫자가 2 인 행이 원래 데이터 테이블에서 선택됩니다.

셀 색상 지정 "단변량 패턴" 보고서의 특정 테이블에서 선택한 행에 해당하는 원래 데이터 테이블의 셀에 색상을 지정합니다. 또한 색상이 지정된 셀의 행이 선택됩니다.

열별 패턴 보고서에는 빨간색 삼각형 옵션이 있습니다. 자세한 내용은 "단변량 패턴 보고서 옵션"에서 확인하십시오.

패턴 탐색 유틸리티 옵션

제어판과 "패턴 탐색" 의 빨간색 삼각형 메뉴에는 요약 및 패턴 테이블에 대한 옵션이 포함되어 있습니다.

가장 많이 중복된 값 데이터에서 가장 많이 나오는 값을 보여 주는 테이블을 표시하거나 숨깁니다. 각 열에 대해 최대 10 개의 값을 표시합니다.

가장 긴 런 동일한 값의 가장 긴 런을 보여 주는 테이블을 표시하거나 숨깁니다. 이 테이블에는 다음 정보가 포함됩니다.

시작 행 런 시작에 해당하는 행 번호입니다.

개수 런 길이입니다.

값 런의 항목 값입니다.

희귀도 런이 우연히 발생할 가능성을 측정한 값입니다. 희귀도가 높으면 해당 런이 우연히 발생했을 가능성이 낮다는 의미입니다. 자세한 내용은 "가장 긴 런의 희귀도"에서 확인하십시오.

가장 긴 중복 시퀀스 같은 열에 두 번 이상 나오는 값의 가장 긴 시퀀스를 보여 주는 테이블을 표시하거나 숨깁니다. 반복 값 시퀀스도 고려됩니다. 시퀀스가 두 번 이상 중복되면 각 인스턴스가 테이블에 나타납니다. 이 테이블에는 다음 정보가 포함됩니다.

시작 행 I 첫 번째 시퀀스의 시작에 해당하는 행 번호입니다.

시작 행 J 중복 시퀀스의 시작에 해당하는 행 번호입니다.

개수 시퀀스 길이입니다.

희귀도 중복 시퀀스가 우연히 발생할 가능성을 측정한 값입니다. 희귀도가 높으면 해당 중복 시퀀스가 우연히 발생했을 가능성이 낮다는 의미입니다. 자세한 내용은 "가장 긴 시퀀스의 희귀도"에서 확인하십시오.

처음 몇 개의 값 중복 시퀀스의 처음 몇 개 값입니다. 각 시퀀스에 대해 최대 3 개의 값을 표시합니다.

형식 지정 값 너비 형식이 지정된 값의 전체 너비 및 소수점 이하 너비를 보여 주는 테이블을 표시하거나 숨깁니다 . 이 테이블에는 다음 정보가 포함됩니다 .

너비 너비 크기입니다 .

전체 기준 개수 해당 열에서 전체 너비가 지정된 너비와 같은 관측값의 수입니다 .

소수점 이하 기준 개수 해당 열에서 소수점 이하 너비가 지정된 너비와 같은 관측값의 수입니다 .

분수 길이 값의 연속된 소수 부분 길이를 보여 주는 테이블을 표시하거나 숨깁니다 . 연속된 소수 부분은 수를 연속해서 나눈 항의 시퀀스로 표현한 것입니다 . 분수 길이는 어떤 수가 유리수인지 아닌지를 나타낼 수 있습니다 . 길이가 15 이상이면 값이 무리수 (예 : 루트 함수의 경우), 무리수를 포함하는 함수 또는 난수 생성기 (Benford, 1938) 에 의한 값임을 나타냅니다 .

참고 : 패턴 탐색에서 최대 분수 길이는 15 입니다 . 분수 길이가 15 이상인 숫자는 모두 15 로 보고됩니다 .

길이 연속된 소수 부분의 길이입니다 .

개수 해당 열에서 연속된 소수 부분 길이가 지정된 길이와 같은 관측값의 수입니다 .

선행 및 후행 숫자 1 ~ 9 의 각 선행 및 후행 숫자에 대한 개수 테이블을 표시하거나 숨깁니다 . 선행 숫자의 경우 해당 선행 숫자를 포함하는 최소값 및 최대값과 기대 비율을 나타내는 열도 있습니다 . 기대 비율은 값의 첫 번째 숫자가 a 와 동일한 빈도가 $F_a = \log_{10}(a+1/a)$ 라는 Benford 법칙을 기반으로 합니다 . 이 법칙은 자릿수가 모두 네 개 이상인 큰 범위의 숫자를 포함하는 데이터에 적용됩니다 .

규격 한계 매칭 (하나 이상의 열에 " 규격 한계 " 열 특성이 있는 경우 사용 가능) 정확히 규격 하한 , 규격 상한 또는 목표값에 해당하는 값 테이블을 표시하거나 숨깁니다 . 정확히 매칭되는 항목이 없으면 보고서가 표시되지 않습니다 .

규격 한계에 대한 분포 (하나 이상의 열에 " 규격 한계 " 열 특성이 있는 경우에만 사용 가능) 규격 한계를 벗어나는 관측값 대 기대값 테이블을 표시하거나 숨깁니다 . 카이제곱 및 p 값 열은 실제 관측값 수가 기대 관측값 수보다 훨씬 많은 경우에만 표시됩니다 .

열 간 중복 값 " 열 간 중복 값 " 보고서를 표시하거나 숨깁니다 . 이 보고서에는 둘 이상의 열에서 동일한 행에 나타나는 값 시퀀스 테이블이 포함되어 있습니다 . 보고서 맨 위에는 중복으로 간주할 최소 시퀀스 길이를 나타내는 노트가 있습니다 . 이 값은 시작 창 또는 제어판에서 " 열 간 중복 값의 최소 런 크기 " 로 지정됩니다 . " 색상 지정 " 옵션도 보고서 맨 위에 있습니다 . 자세한 내용은 " 열 간 중복 값 색상 지정 " 에서 확인하십시오 .

열 간 중복 값 색상 지정 (" 열 간 중복 값 " 을 선택한 후에만 사용 가능) 데이터 테이블에서 " 열 간 중복 값 " 보고서의 중복 매칭 항목에 해당하는 셀에 색상을 지정합니다 . 또한 중복 항목을 포함하는 행이 선택됩니다 .

선형 관계 " 변수 간 선형 관계 " 보고서를 표시하거나 숨깁니다 . 이 보고서에는 지정된 수의 순차적 행에 대해 열 간의 정확 선형 관계를 보여 주는 테이블이 포함됩니다 . 정확 선형 관계는 R^2 값이 0.999999 보다 크거나 같은 선형 관계로 정의됩니다 . 보고서 맨 위에는 고려된 최소 행 수를 나타내는 노트가 있습니다 . 이 값은 시작 창 또는 제어판에서 " 선형 관계의 최소 행

수 " 로 지정됩니다 . " 색상 지정 " 옵션도 보고서 맨 위에 있습니다 . 자세한 내용은 " 선형 관계 색상 지정 " 에서 확인하십시오 .

선형 관계 색상 지정 (" 선형 관계 " 를 선택한 후에만 사용 가능) 데이터 테이블에서 " 변수 간 선형 관계 " 테이블의 선형 관계에 해당하는 셀에 색상을 지정합니다 . 또한 선형 관계를 포함하는 행이 선택됩니다 .

열 저장 다음 저장 옵션을 포함합니다 .

런 테이블 저장 모든 열의 가장 긴 런을 새 데이터 테이블에 저장합니다 . 이 데이터 테이블은 " 단변량 패턴 " 보고서의 " 가장 긴 런 " 테이블과 동일한 열을 포함하고 " 시작 행 " 을 기준으로 정렬됩니다 .

중복 시퀀스 테이블 저장 모든 열의 가장 긴 중복 시퀀스를 새 데이터 테이블에 저장합니다 . 이 데이터 테이블은 " 단변량 패턴 " 보고서의 " 가장 긴 중복 시퀀스 " 테이블과 동일한 열을 포함하고 " 시작 행 I" 를 기준으로 정렬됩니다 .

열 간 중복 값 저장 " 열 간 중복 값 " 테이블을 새 데이터 테이블에 저장합니다 .

선형 관계 저장 " 변수 간 선형 관계 " 테이블을 새 데이터 테이블에 저장합니다 .

열 순서화 정렬 옵션이 포함되어 있습니다 . 정렬을 변경하면 " 단변량 패턴 " 보고서의 열 목록이 재정렬됩니다 . 맨 위 열이 자동으로 선택되고 해당 열에 대한 보고서만 표시됩니다 .

런 희귀도별로 정렬 가장 긴 런의 희귀도 값을 기준으로 열 목록을 정렬합니다 . 더 높은 희귀도 값이 맨 위에 나옵니다 .

시퀀스 희귀도별로 정렬 긴 중복 시퀀스의 희귀도 값을 기준으로 열 목록을 정렬합니다 . 더 높은 희귀도 값이 맨 위에 나옵니다 .

열 이름별로 정렬 열 이름을 기준으로 열 목록을 정렬합니다 (사전순).

원래 순서 시작 창에서 열이 지정된 원래 순서대로 열 목록을 정렬합니다 .

셀 색상 지우기 데이터 테이블에서 선택한 열 또는 모든 열 (데이터 테이블에서 열이 선택되지 않은 경우) 에 대해 셀 색상을 지웁니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

단변량 패턴 보고서 옵션

단변량 패턴의 각 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

가장 많이 중복된 값 선택 데이터 테이블에서 " 가장 많이 중복된 값 " 테이블의 값에 해당하는 관측값을 선택합니다 .

가장 긴 런 선택 데이터 테이블에서 "가장 긴 런" 테이블의 값에 해당하는 관측값을 선택합니다.

가장 긴 중복 시퀀스 선택 데이터 테이블에서 " 가장 긴 중복 시퀀스 " 테이블의 값에 해당하는 관측값을 선택합니다 .

형식 지정 값 너비별로 선택 데이터 테이블에서 " 형식 지정 값 너비 " 테이블의 값 중 사용자가 지정한 전체 너비의 값에 해당하는 관측값을 선택합니다 .

소수점 이하 너비별로 선택 데이터 테이블에서 " 형식 지정 값 너비 " 테이블의 값 중 사용자가 지정한 소수점 이하 너비의 값에 해당하는 관측값을 선택합니다 .

분수 길이별로 선택 데이터 테이블에서 " 분수 길이 " 테이블의 값 중 사용자가 지정한 분수 길이의 값에 해당하는 관측값을 선택합니다 .

선행 숫자별로 선택 데이터 테이블에서 " 선행 및 후행 숫자 " 테이블의 값 중 사용자가 지정한 선행 숫자의 값에 해당하는 관측값을 선택합니다 .

패턴 탐색 유틸리티의 추가 예

Water Treatment.jmp 데이터 테이블에는 도시 폐수 처리장의 데이터가 포함되어 있습니다 . 열은 폐수 처리장에 있는 여러 센서의 일별 측정값입니다 . 패턴 탐색 유틸리티를 사용하여 센서 측정값에 예기치 않은 패턴이 있는지 여부를 확인합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Water Treatment.jmp 샘플 데이터 테이블을 엽니다 .
2. **분석 > 선별 > 패턴 탐색**을 선택합니다 .
3. Sensor Measurements 열 그룹을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 제어판 " 의 " 열 간 " 에서 " 열 간 중복 값 " 옆에 있는 상자를 클릭합니다 .

그림 21.17 열 간 중복 값 보고서

열 간 중복 값

2회 이상의 런 사용 색상 지정

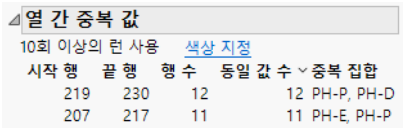
시작 행	끝 행	행 수	동일 값 수	중복 집합
219	230	12	12	PH-P, PH-D
207	217	11	11	PH-E, PH-P
214	217	4	8	PH-E, PH-P, PH-D
330	333	4	8	PH-E, PH-P, PH-D
198	204	7	7	PH-P, PH-D
91	96	6	6	PH-P, PH-D
94	96	3	6	PH-E, PH-P, PH-D
208	210	3	6	PH-E, PH-P, PH-D
216	217	2	6	PH-E, PH-P, PH-D, PH-S
228	229	2	6	PH-E, PH-P, PH-D, PH-S
247	252	6	6	PH-E, PH-P
290	295	6	6	PH-E, PH-P
293	295	3	6	PH-E, PH-P, PH-D
427	429	3	6	PH-E, PH-P, PH-D
472	477	6	6	PH-E, PH-P
26	30	5	5	PH-E, PH-P
63	67	5	5	PH-P, PH-D
141	145	5	5	PH-E, PH-D
299	303	5	5	PH-E, PH-P
352	356	5	5	PH-P, PH-D

" 열 간 중복 값 " 보고서에 중복 열이 포함된 긴 목록이 표시됩니다 . 그러나 최소 런 크기가 2 이고 작은 런이 목록에 많이 있습니다 . 10 이상의 더 큰 런에 초점을 맞추려고 합니다 .

6. "제어판"의 "설정"에서 "열 간 중복 값의 최소 런 크기" 옆에 있는 상자에 "10"을 입력합니다.

팁 : 상자 바깥쪽의 아무 곳이나 클릭하면 보고서가 자동으로 업데이트됩니다 .

그림 21.18 런 크기가 10 이상인 열 간 중복 값 보고서



열 간 중복 값

10회 이상의 런 사용 색상 지정

시작 행	끝 행	행 수	동일 값 수	중복 집합
219	230	12	12	PH-P, PH-D
207	217	11	11	PH-E, PH-P

업데이트된 " 열 간 중복 값 " 보고서에서는 런 횟수가 10 이상인 중복 집합이 두 개 있음을 보여 줍니다 . 두 집합 모두 PH-P 열을 포함합니다 .

7. " 열 간 중복 값 " 보고서에서 **색상 지정**을 클릭합니다 .

그림 21.19 색상이 지정된 Water Treatments.jmp 데이터 테이블

	PH-E	DBO-E	DQO-E	SS-E	SSV-E	SED-E	COND-E	PH-P	DBO-P	SS-P	SSV-P	SED-P	COND-P	PH-D
206	7.7	124.0	234.0	120.0	61.7	1.8	1550.0	7.9	123.0	220.0	59.0	4.0	1651.0	7.8
207	7.6	172.0	394.0	226.0	51.3	5.5	1324.0	7.6	197.0	302.0	47.0	4.0	1306.0	7.7
208	7.6	268.0	324.0	226.0	53.1	4.0	1390.0	7.6	265.0	236.0	53.4	4.5	1389.0	7.6
209	7.6	176.0	200.0	176.0	50.0	3.0	859.0	7.6	149.0	172.0	51.2	2.0	860.0	7.6
210	7.6	146.0	243.0	164.0	57.3	3.2	1095.0	7.6	193.0	208.0	55.8	3.0	1089.0	7.6
211	7.8	139.0	380.0	256.0	53.1	5.5	2040.0	7.8	201.0	340.0	47.6	8.5	2000.0	7.9
212	7.9	205.0	529.0	264.0	61.4	6.5	2750.0	7.9	220.0	324.0	60.5	6.0	2850.0	7.8
213	7.8	117.0	366.0	500.0	27.2	4.2	3230.0	7.8	135.0	620.0	32.3	7.0	3170.0	7.7
214	7.6	148.0	400.0	564.0	31.6	7.0	1554.0	7.6	235.0	798.0	28.6	11.0	1515.0	7.6
215	7.8	158.0	256.0	194.0	47.4	3.0	1831.0	7.8	218.0	240.0	46.7	4.0	1860.0	7.8
216	7.7	123.0	246.0	186.0	39.8	2.5	1288.0	7.7	241.0	300.0	25.3	4.5	1328.0	7.7
217	7.7	133.0	410.0	204.0	62.7	4.5	1490.0	7.7	161.0	256.0	57.8	4.5	1421.0	7.7
218	7.5	237.0	388.0	178.0	57.3	4.0	1620.0	7.6	300.0	240.0	53.3	4.5	1571.0	7.7
219	7.7	203.0	357.0	222.0	58.6	3.0	1828.0	7.6	215.0	208.0	73.1	5.0	1817.0	7.6
220	7.8	199.0	380.0	244.0	55.7	5.0	1453.0	7.8	110.0	196.0	60.2	3.5	1430.0	7.8
221	7.8	214.0	941.0	304.0	53.3	6.5	1454.0	7.8	396.0	320.0	57.5	7.0	1528.0	7.8
222	7.7	208.0	298.0	260.0	37.7	4.0	1236.0	7.8	245.0	284.0	46.5	5.0	1285.0	7.8
223	7.9	206.0	305.0	188.0	56.4	4.0	1200.0	7.9	220.9	272.0	50.0	6.0	1228.0	7.9
224	7.7	247.0	361.0	286.0	51.0	4.5	1180.0	7.7	146.0	212.0	56.4	2.5	1251.0	7.7
225	7.5	438.0	681.0	370.0	56.8	6.0	1396.0	7.6	212.0	348.0	64.4	6.0	1446.0	7.6
226	7.8	255.0	282.0	166.0	67.5	2.0	950.0	7.8	284.0	172.0	65.1	2.0	965.0	7.8
227	7.7	283.0	310.0	170.0	61.2	2.5	1065.0	7.8	235.0	270.0	51.1	5.5	1100.0	7.8
228	7.9	313.0	341.0	512.0	33.6	6.0	915.0	7.9	205.0	520.0	33.8	7.5	976.0	7.9
229	7.7	300.0	610.0	452.0	45.1	8.0	1313.0	7.7	375.0	556.0	43.9	10.0	1312.0	7.7
230	7.8	177.0	412.0	196.0	69.4	4.5	2190.0	7.8	330.0	476.0	52.1	7.5	2330.0	7.8

"색상 지정" 기능은 원래 데이터 테이블에서 보고서의 중복 집합에 해당하는 셀에 색상을 지정합니다.

패턴 탐색 유틸리티에 대한 통계 상세 정보

가장 긴 런의 희귀도

가장 긴 런의 희귀도를 계산하려면 먼저 다음 변수를 정의합니다.

n = 열의 행 수

k = 열에서 특정 값이 나타나는 횟수

$p = k/n$ = 열에서 특정 값을 관측할 확률

m = 런 길이

N = 고유 런의 수

가장 긴 런의 희귀도는 다음과 같이 계산됩니다.

$$\text{희귀도} = -\log_2(1 - (1 - p^{m-1})^N)$$

가장 긴 시퀀스의 희귀도

가장 긴 시퀀스의 희귀도를 계산하려면 먼저 다음 변수를 정의합니다 .

p = 열에서 특정 시퀀스를 한 번 관측할 확률

k = 열에서 시퀀스의 시작 값이 나타나는 횟수

가장 긴 시퀀스의 희귀도는 다음과 같이 계산됩니다 .

$$\text{희귀도} = -\log_2(1 - (1 - p)^k)$$

22 장

반응 변수 선별
대규모 데이터의 여러 반응 변수 검정

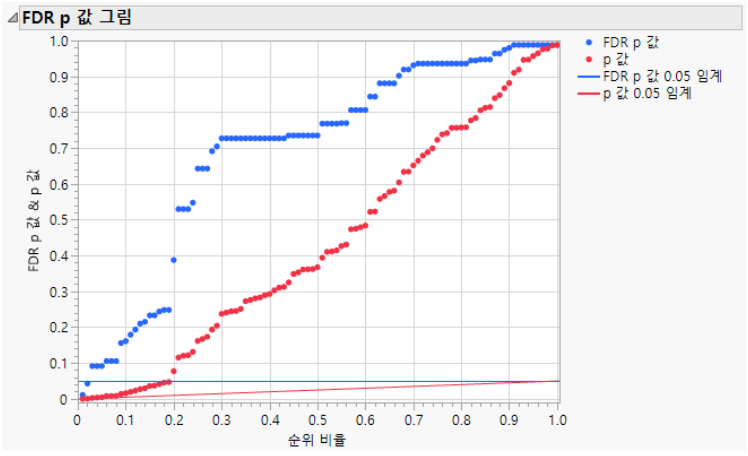
부품이나 유기체에 대해 수백 또는 수천 번의 측정을 수행하는 대규모 데이터 집합을 분석하려면 혁신적인 접근 방식이 필요합니다. 오해의 소지가 없더라도 요인 효과에 대해 많은 반응 변수를 검정하는 것은 적절한 방법론 없이는 어려울 수 있습니다.

반응 변수 선별 플랫폼에서는 다수의 반응 변수에 대한 검정 수행 과정을 자동화합니다. 검정 결과와 요약 통계량이 보고서 대신 데이터 테이블에 제공되어 데이터 탐색이 가능합니다. False Discovery Rate 방법은 유의성이 잘못 선언되지 않도록 보호합니다. *p* 값 그림은 LogWorth를 사용하여 척도화되므로 쉽게 해석할 수 있습니다.

대규모 데이터 집합은 정리되지 않은 경우가 많으므로 반응 변수 선별에서는 불규칙하게 분포된 데이터 및 결측 데이터를 처리하는 방법을 제공합니다. 로버스트 추정값 방법을 사용하면 이상치가 데이터에 남아 있을 수 있지만 이러한 이상치에 대한 검정 민감도가 줄어듭니다. 결측 데이터 옵션을 사용하면 결측값을 분석에 포함할 수 있습니다. 이러한 기능을 통해 먼저 광범위한 데이터 품질 분석을 수행하지 않고도 데이터를 분석할 수 있습니다.

관측값이 많으면 실제적 관심 영역이 아닌 차이도 통계적으로 유의할 수 있습니다. 반응 변수 선별에서는 감지하려는 차이를 지정할 수 있는 실제적 차이 검정을 제공합니다. 다른 한편으로는 차이가 주어진 크기를 초과하지 않는지, 즉 평균이 동일한지 여부를 알고 싶을 수도 있습니다. 이를 위해 반응 변수 선별에서 동등성 검정을 제공합니다.

그림 22.1 반응 변수 선별 그림의 예



목차

반응 변수 선별 플랫폼 개요 . . . 413
반응 변수 선별의 예 . . . 414
반응 변수 선별 플랫폼 시작 . . . 416
반응 변수 선별 보고서 . . . 418
결과 테이블 . . . 418
FDR p 값 그림 . . . 420
FDR LogWorth 대 효과 크기 . . . 421
FDR LogWorth 대 R^2 . . . 422
p 값 데이터 테이블 . . . 422
p 값 데이터 테이블 스크립트 . . . 423
반응 변수 선별 플랫폼 옵션 . . . 423
평균 데이터 테이블 . . . 424
평균 비교 데이터 테이블 . . . 425
모형 적합의 반응 변수 선별 분석법 . . . 427
모형 적합의 반응 변수 선별 시작 . . . 428
반응 변수 선별 적합 보고서 . . . 429
반응 변수 선별 적합 옵션 . . . 429
PValues 데이터 테이블 . . . 430
Y Fits 데이터 테이블 . . . 431
반응 변수 선별의 추가 예 . . . 432
실제적 유의성 및 동등성 검정의 예 . . . 432
최대 LogWorth 옵션의 예 . . . 433
로버스트 적합의 예 . . . 435
반응 변수 선별 분석법 . . . 438
반응 변수 선별 플랫폼에 대한 통계 상세 정보 . . . 439
False Discovery Rate . . . 439

반응 변수 선별 플랫폼 개요

반응 변수 선별 플랫폼에서는 다수의 반응 변수에 대한 검정 수행 과정을 자동화합니다. 지정한 각 반응 변수를 지정한 각 요인에 대해 검정합니다. 반응 변수 선별은 대규모 데이터와 관련된 두 가지 주요 문제를 해결합니다. 검정을 많이 수행해야 한다는 것과 이상치 및 결측값을 효과적으로 처리하기 위한 요구 사항이 이러한 문제입니다.

반응 변수 선별은 플랫폼 및 모형 적합 분석법으로 사용할 수 있습니다. 표 22.1에 표시된 대로 두 경우 모두 X로 Y 적합 플랫폼의 검정과 유사한 검정을 수행합니다. 분석법으로 사용될 경우에는 개별 모형 효과에 대한 반응을 검정합니다.

반응 변수 선별에서는 필요한 여러 추론을 지원하기 위해 다음과 같은 기능을 제공합니다 .

데이터 테이블 결과가 보고서뿐만 아니라 데이터 테이블에도 표시되므로 결과를 탐색, 정렬, 검색 및 그림에 표시할 수 있습니다 . *p* 값의 LogWorth 와 같이 그림을 쉽게 해석할 수 있는 통계량이 제공됩니다 .

False Discovery Rate 많은 수의 검정을 수행하므로 검정이 유의하다고 선언하는 전체 비율을 제어해야 합니다 . 반응 변수 선별은 False Discovery Rate 를 제어합니다 . FDR(False Discovery Rate) 은 유의한 것으로 잘못 선언할 유의성 검정의 기대 비율입니다 (Benjamini and Hochberg 1995 및 Westfall et al. 2011 참조).

실제적 유의성 검정 관측값이 많으면 실제적으로 중요하지 않은 작은 효과도 통계적으로 유의하다는 결과가 나올 수 있습니다 . 이 문제를 해결하기 위해 **실제로 유의하다고** 간주할 효과 크기를 정의할 수 있습니다 . 그런 다음 실제적 유의성 검정을 수행하여 현실적으로 관심을 끌 수 있는 큰 효과만 감지합니다 .

동등성 검정 많은 요인을 연구할 때 반응에 근본적으로 동일한 영향을 미치는 요인에 관심을 가질 수 있습니다 . 이 경우 실제적 동등성을 정의하는 효과 크기를 지정한 후 동등성 검정을 수행할 수 있습니다 .

반응 변수 선별에서는 정리되지 않은 데이터를 처리할 때 발생하는 문제를 해결하기 위해 이상치 및 결측 데이터를 처리하는 기능을 제공합니다. 이러한 기능을 사용하면 데이터 품질 문제를 해결하기 위해 노력할 필요 없이 데이터를 직접 분석할 수 있습니다.

로버스트 추정 데이터에 이상치가 있으면 표준 오차 추정값이 커지고 이로 인해 검정이 실제 효과에 민감하지 않게 됩니다 . 로버스트 옵션을 선택하여 Huber M- 추정을 수행합니다 . 이상치가 데이터에 남아 있지만 이러한 이상치에 대한 검정 민감도가 감소합니다 .

결측값 옵션 이 플랫폼에는 범주형 예측 변수의 결측값을 정보가 포함된 것으로 간주하는 옵션이 포함되어 있습니다 .

표 22.1 반응 변수 선별에서 수행되는 분석

반응	요인	X 로 Y 적합 분석	설명
연속형	범주형	일원 분석	분산 분석
연속형	연속형	이변량	단순 선형 회귀

표 22.1 반응 변수 선별에서 수행되는 분석 (계속)

반응	요인	X 로 Y 적합 분석	설명
범주형	범주형	분할 분석	카이제곱
범주형	연속형	로지스틱	단순 로지스틱 회귀

반응 변수 선별 플랫폼에서는 "반응 변수 선별" 보고서와 "p 값" 테이블을 생성합니다. "반응 변수 선별" 분석법을 사용하면 "반응 변수 선별 적합" 보고서와 "PValues" 및 "Y Fits"라는 두 개의 데이터 테이블이 생성됩니다.

JSL 명령 `Summarize Y by X`는 반응 변수 선별 플랫폼과 동일한 기능을 수행하지만 플랫폼 창을 생성하지 않습니다. 자세한 내용은 JSL Syntax Reference의 에서 확인하십시오.

반응 변수 선별의 예

Probe.jmp 샘플 데이터 테이블에는 5800 개의 웨이퍼에서 측정한 387 개의 특성 (Responses 열 그룹) 이 포함되어 있습니다 . Lot ID 및 Wafer Number 열은 웨이퍼를 고유하게 식별합니다 . 공정 변경 (Process) 에 대해 서로 다른 값을 표시하는 특성을 알아보려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Probe.jmp 를 엽니다 .
2. **분석 > 선별 > 반응 변수 선별**을 선택합니다 .
3. Responses 열 그룹을 선택하고 **Y, 반응**을 클릭합니다 .
4. Process 를 선택하고 **X** 를 클릭합니다 .
5. **최대 LogWorth** 상자에 "100" 을 입력합니다 .

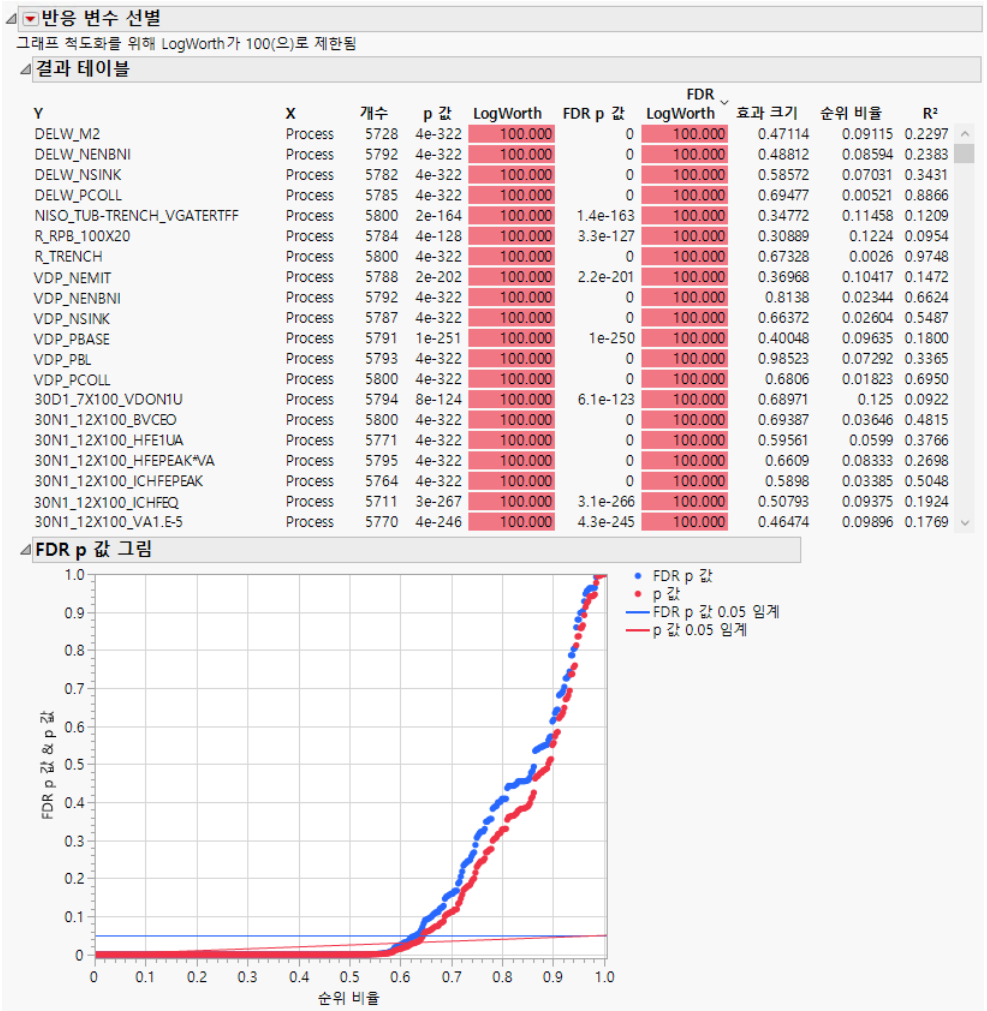
   100 이상의 logworth(-$\log_{10}$(*p* 값)) 는 극도로 작은 *p* 값에 해당합니다 . " 최대 LogWorth" 값을 설정하면 보고되는 logworth 값이 제한되므로 그림의 척도를 제어하는 데 도움이 됩니다 .
6. **확인**을 클릭합니다 .

   보고서에 *p* 값 결과 테이블과 "FDR p 값 그림 " 이 표시되고 두 개의 다른 그림 보고서도 포함됩니다 .

"FDR p 값 그림"에는 387개의 각 검정에 대해 "FDR p 값"과 "p 값"이라는 두 가지 유형의 *p* 값이 표시됩니다. "순위 비율"에 대한 p 값이 그림에 표시됩니다. "p 값"은 Process에 대한 Y 검정의 일반적인 *p* 값입니다. "FDR p 값"은 주어진 FDR(False Discover Rate)을 보장하기 위해(이 예의 경우 0.05) 조정된 *p* 값입니다. 그림에 FDR p 값은 파란색으로 표시되고 p 값은 빨간색으로 표시됩니다. "순위 비율"은 FDR *p* 값을 가장 작은 것부터 순서대로, 즉 유의성이 큰 것부터 내림차순으로 정렬합니다.

그림의 파란색 가로선과 경사진 빨간색 선은 모두 FDR 유의성에 대한 임계입니다. 파란색 선 아래의 FDR *p* 값을 사용한 검정은 False Discover Rate에 대해 조정된 0.05 수준에서 유의합니다. 빨간색 선 아래의 일반 *p* 값을 사용한 검정은 0.05 수준에서 유의합니다. 이런 방식으로 그림을 사용하여 두 *p* 값 집합 중 하나에서 FDR 유의성을 이해할 수 있습니다.

그림 22.2 Process 에 대한 387 개 검정의 반응 변수 선별 보고서



Y	X	개수	p 값	LogWorth	FDR p 값	FDR LogWorth	효과 크기	순위 비율	R^2
DELW_M2	Process	5728	4e-322	100.000	0	100.000	0.47114	0.09115	0.2297
DELW_NENBNI	Process	5792	4e-322	100.000	0	100.000	0.48812	0.08594	0.2383
DELW_NSINK	Process	5782	4e-322	100.000	0	100.000	0.58572	0.07031	0.3431
DELW_PCOLL	Process	5785	4e-322	100.000	0	100.000	0.69477	0.00521	0.8866
NISO_TUB-TRENCH_VGATERTFF	Process	5800	2e-164	100.000	1.4e-163	100.000	0.34772	0.11458	0.1209
R_RPB_100X20	Process	5784	4e-128	100.000	3.3e-127	100.000	0.30889	0.1224	0.0954
R_TRENCH	Process	5800	4e-322	100.000	0	100.000	0.67328	0.0026	0.9748
VDP_NEMIT	Process	5788	2e-202	100.000	2.2e-201	100.000	0.36968	0.10417	0.1472
VDP_NENBNI	Process	5792	4e-322	100.000	0	100.000	0.8138	0.02344	0.6624
VDP_NSINK	Process	5787	4e-322	100.000	0	100.000	0.66372	0.02604	0.5487
VDP_PBASE	Process	5791	1e-251	100.000	1e-250	100.000	0.40048	0.09635	0.1800
VDP_PBL	Process	5793	4e-322	100.000	0	100.000	0.98523	0.07292	0.3365
VDP_PCOLL	Process	5800	4e-322	100.000	0	100.000	0.6806	0.01823	0.6950
30D1_7X100_VDON1U	Process	5794	8e-124	100.000	6.1e-123	100.000	0.68971	0.125	0.0922
30N1_12X100_BVCEO	Process	5800	4e-322	100.000	0	100.000	0.69387	0.03646	0.4815
30N1_12X100_HFE1UA	Process	5771	4e-322	100.000	0	100.000	0.59561	0.0599	0.3766
30N1_12X100_HFEPEAK*VA	Process	5795	4e-322	100.000	0	100.000	0.6609	0.08333	0.2698
30N1_12X100_ICHFEPEAK	Process	5764	4e-322	100.000	0	100.000	0.5898	0.03385	0.5048
30N1_12X100_ICHFEQ	Process	5711	3e-267	100.000	3.1e-266	100.000	0.50793	0.09375	0.1924
30N1_12X100_VA1.E-5	Process	5770	4e-246	100.000	4.3e-245	100.000	0.46474	0.09896	0.1769

"FDR p 값 그림"에서는 유의한 검정이 60%가 넘는다는 것을 보여 줍니다. 조정되지 않은 p 값을 사용할 경우 유의하지만 FDR p 값을 사용할 경우에는 유의하지 않은 소수의 검정이 있습니다. 이러한 검정은 빨간색 선 위에 있지만 파란색 선 아래에 있는 빨간색 점에 해당합니다.

Process 간에 유의하게 다른 특성을 식별하기 위해 그림에서 적절한 점 주위를 직사각형 형태로 드래그할 수 있습니다. 그러면 첫 번째 열에 특성 이름이 제공된 "결과 테이블"에서 이러한 점에 해당하는 행이 선택됩니다.

"결과 테이블"에는 Responses 그룹의 각 반응 측도당 하나씩 총 387개 행이 포함되어 있습니다. 반응은 첫 번째 열, Y에 표시되고 각 반응은 X 열, 즉 Process의 효과에 대해 검정됩니다.

나머지 열은 X에 대한 Y 검정의 정보를 제공합니다. 여기서 검정은 일원 분산 분석입니다. 이 테이블에는 검정의 p 값, LogWorth, FDR(False Discovery Rate) p 값 및 FDR LogWorth 등이 제공

됩니다. 이 테이블을 사용하여 다양한 통계량을 기준으로 정렬하고, 행을 선택하거나 관심 통계량을 그림에 표시할 수 있습니다.

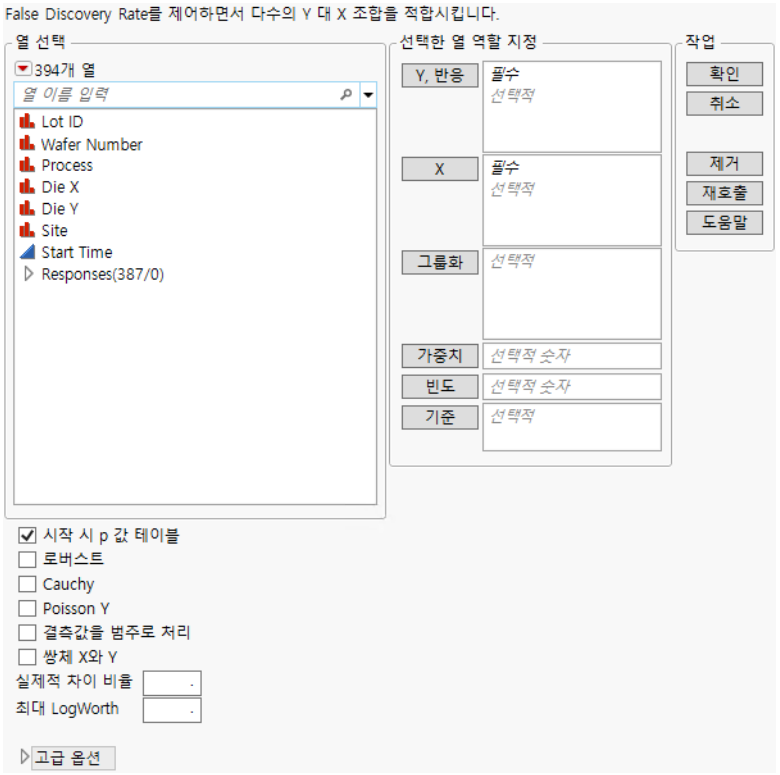
시작 창에서 "최대 LogWorth"를 "100"으로 설정했으므로 1e-100 이하의 p 값에 해당하는 LogWorth 및 FDR LogWorth 값은 100으로 보고됩니다. 또한 2보다 큰 FDR LogWorth 값에 해당하는 셀은 강도 그래디언트가 적용된 색상이 지정됩니다.

보고서 및 p 값 테이블에 대한 자세한 내용은 "반응 변수 선별 보고서"에서 확인하십시오.

반응 변수 선별 플랫폼 시작

분석 > 선별 > 반응 변수 선별을 선택하여 반응 변수 선별 플랫폼을 시작합니다.

그림 22.3 반응 변수 선별 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

시작 창 역할

Y, 반응 분석할 측정값이 포함된 반응 열을 식별합니다.

X 반응을 검정할 열을 식별합니다.

그룹화 지정된 열의 각 수준에 할당된 행을 개별적으로 분석합니다. 모든 결과는 단일 테이블과 보고서에 제공됩니다.

가중치 각 행에 가중치를 할당하는 데 사용할 값이 있는 열을 식별합니다. 이러한 값은 분석에서 가중치로 사용됩니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 열을 식별합니다. 이러한 값을 사용하면 사전 요약된 데이터를 고려할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

기준 지정된 열의 각 수준에 대해 해당하는 Y 및 X 를 분석하고 결과를 별도의 테이블과 보고서에 제공합니다.

시작 창 옵션

시작 시 p 값 테이블 *p* 값 및 개별 모형 적합 통계량에 대한 데이터 테이블을 생성합니다. 이 데이터 테이블은 반응 변수 선별 보고서의 " 결과 테이블 " 에 연결됩니다.

로버스트 (연속형 반응에만 사용 가능) Huber M- 추정 방법 (Huber 와 Ronchetti 2009) 을 사용하여 회귀 및 ANOVA 모형을 적합시킵니다. 이 방법은 Huber 손실 함수를 최소화하여 모수를 추정합니다. Huber 손실 함수는 이상치에 벌점을 부과하여 이상치가 검정에 미치는 영향을 줄입니다. 이상치가 없으면 추정값이 최소 제곱 추정값에 가깝습니다. " 로버스트 " 옵션을 선택한 경우 여러 로버스트 측도 및 모형 적합 통계량 열이 " 결과 테이블 " 에 추가됩니다. Huber M- 추정에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오. 예는 " 로버스트 적합의 예 " 에서 확인하십시오.

주의 : " 로버스트 " 옵션을 사용하면 처리 시간이 길어질 수 있습니다.

Cauchy (연속형 반응에만 사용 가능) 최대 가능도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다. 이 추정 방법은 오차가 Cauchy 분포를 따른다고 가정합니다. Cauchy 분포는 정규 분포보다 꼬리가 더 두꺼우므로 이상치의 영향이 줄어듭니다. 이 옵션은 데이터에 포함된 이상치의 비율이 높은 경우에 유용할 수 있습니다. 하지만 데이터가 정규 분포에 가깝고 이상치가 일부만 있는 경우에 이 옵션을 사용하면 추론이 잘못될 수 있습니다. "Cauchy" 옵션을 선택하면 여러 Cauchy 측도 및 모형 적합 통계량 열이 " 결과 테이블 " 에 추가됩니다.

Poisson Y 각 Y 반응을 Poisson 분포를 따르는 개수로 적합시킵니다. 검정이 범주형 X 에 대해서만 수행됩니다. 이 옵션은 반응이 개수일 때 적절합니다.

결측값을 범주로 처리 범주형 X 변수에 대해 X 의 결측값을 범주로 처리합니다.

쌍체 X 와 Y **Y, 반응** 및 **X** 목록의 순서에 따라 X 열과 쌍을 이룬 Y 열에 대해서만 검정을 수행합니다. 첫 번째 Y 는 첫 번째 X 와 쌍이 되고 두 번째 Y 는 두 번째 X 와 쌍이 되는 식입니다.

실제적 차이 비율 실용적으로 의미가 있다고 간주하는 차이를 나타내는 규격 범위 또는 추정된 6 표준편차 범위의 비율입니다. " 규격 한계 " 가 열 특성으로 설정되지 않은 경우에는 반응에

대한 6 표준편차 범위가 추정됩니다 . 표준편차 추정값은 IQR(사분위수 범위) 을 사용하여 $\hat{\sigma} = (IQR)/(1.3489795)$ 로 계산됩니다 .

" 실제적 차이 비율 " 을 지정하지 않은 경우 기본값은 0.10 입니다 . 실제적 유의성 검정과 동등성 검정의 경우 이 차이를 사용하여 실제적 차이를 확인합니다 . 자세한 내용은 " 평균 비교 데이터 테이블 " 에서 확인하십시오 .

최대 LogWorth LogWorth 값 (p 값의 $-\log_{10}$) 과 관련된 그림의 척도를 제어합니다 . 최대 LogWorth를 초과하는 LogWorth 값은 LogWorth 그림에서 극단 척도를 방지하기 위해 최대 LogWorth 로 표시됩니다 . 예는 " 최대 LogWorth 옵션의 예 " 에서 확인하십시오 .

고급 옵션 다음 옵션이 포함되어 있습니다 .

카파 " 카파 " 라는 새 열을 데이터 테이블에 추가합니다 . Y 와 X 가 모두 범주형이고 수준이 같으면 카파가 제공됩니다 . 이 값은 Y 와 X 간의 합치도 측도입니다 .

상관 " 상관 " 옵션은 값 순서화로 정의된 인덱스에 따라 Pearson 곱적률 상관계수를 계산합니다 .

X 와 Y 가 모두 이항인 경우 Pearson 곱적률을 계산하면 Spearman Rho 및 Kendall Tau-B 가 제공됩니다 . 그렇지 않은 경우 " 상관 " 값이 크면 연관성을 나타내고 " 상관 " 값이 작으면 연관성을 배제하지 않습니다 .

X 범주형 강제 적용 모델링 유형을 무시하고 모든 X 열을 범주형 열로 처리합니다 .

X 연속형 강제 적용 모델링 유형을 무시하고 모든 X 열을 연속형 열로 처리합니다 .

Y 범주형 강제 적용 모델링 유형을 무시하고 모든 Y 열을 범주형 열로 처리합니다 .

Y 연속형 강제 적용 모델링 유형을 무시하고 모든 Y 열을 연속형 열로 처리합니다 .

언스레드 계산 속도를 위해 멀티스레딩 사용을 제한합니다 .

반응 변수 선별 보고서

"반응 변수 선별" 보고서는 결과 테이블과 여러 개의 그림으로 구성됩니다. 이러한 그림은 FDR(False Discovery Rate) 통계량에 초점을 맞춥니다. 자세한 내용은 "False Discovery Rate"에서 확인하십시오.

기본 그림은 "FDR p 값 그림", "FDR LogWorth 대 효과 크기" 및 "FDR LogWorth 대 R^2"입니다. 시작 창에서 "로버스트" 옵션을 선택한 경우 이러한 각 보고서의 로버스트 버전도 제공됩니다. 또한 로버스트 LogWorth 대 FDR LogWorth 그림도 제공되므로 로버스트 적합을 사용할 때의 영향을 평가하는 데 도움이 됩니다.

결과 테이블

반응 변수 선별 결과 테이블에는 Y 및 X 변수의 각 쌍에 대한 행이 포함되어 있습니다. 테이블의 열에는 선택한 적합과 Y 및 X 모델링 유형에 고유한 측도와 모형 적합 통계량이 포함됩니다.

그룹 (그룹화 변수가 있는 경우에만 표시됨) 그룹화 열의 수준입니다 .

Y 지정된 반응 열입니다 .

X 지정된 요인 열입니다 .

개수 검정에 사용된 행 수 또는 빈도 변수나 가중치 변수의 해당 합입니다 .

p 값 Y 와 X 변수 쌍에 해당하는 유의성 검정의 *p* 값입니다 . X 로 Y 적합 통계량에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오 .

LogWorth $-\log_{10}(p$ 값) 의 계산 값입니다 . 이 변환은 그래프 작성에 적절한 척도를 제공하도록 *p* 값을 조정합니다 . 2 를 초과하는 값은 0.01 수준에서 유의합니다 ($(-\log_{10}(0.01) = 2)$).

FDR p 값 Benjamini-Hochberg 기법을 사용하여 계산된 False Discovery Rate *p* 값입니다 . 이 기법은 여러 검정에 대한 FDR(False Discovery Rate) 을 제어하기 위해 *p* 값을 조정합니다 . 그룹 변수가 없으면 테이블에 표시된 모든 검정이 여러 검정 집합에 포함됩니다 . 그룹 변수가 있으면 그룹 변수의 각 수준에 대해 수행된 모든 검정이 여러 검정 집합에 포함됩니다 . FDR 수정에 대한 자세한 내용은 Benjamini and Hochberg 연구 자료 (1995) 에서 확인하십시오 . False Discovery Rate 에 대한 자세한 내용은 "False Discovery Rate" 에서 확인하십시오 .

FDR LogWorth $-\log_{10}$(FDR p 값) 의 계산 값입니다 . 이 통계량은 유의성을 그림에 표시하고 평가하는 데 사용됩니다 . *p* 값이 작으면 FDR LogWorth 값이 커집니다 . 2 보다 큰 FDR LogWorth 값 (0.01 보다 작은 *p* 값) 에 해당하는 셀은 강도 그래디언트가 적용된 색상이 지정됩니다 .

효과 크기 X 의 수준 또는 값에 따라 반응 값이 달라지는 정도를 나타냅니다 . 효과 크기는 척도 불변입니다 .

- Y 가 연속형인 경우 효과 크기는 가설 검정에서 구한 평균 제곱합의 제곱근을 반응 표준편차의 로버스트 추정값으로 나눈 값입니다 . IQR(사분위수 범위) 이 0 이 아니고 IQR > 범위 /20 이면 표준편차 추정값은 IQR/1.3489795 입니다 . 그렇지 않으면 표본 표준편차가 사용됩니다 .
- Y 가 범주형이고 X 가 연속형인 경우 효과 크기는 전체 모형 검정에 대한 평균 카이제곱 값의 제곱근입니다 .
- Y 와 X 가 모두 범주형인 경우 효과 크기는 평균 Pearson 카이제곱의 제곱근입니다 .

순위 비율 FDR LogWorth 순위를 검정 수의 비율로 나타낸 것입니다 . 검정 수가 *m* 이면 가장 큰 FDR LogWorth 값의 순위 비율은 $1/m$ 이고 가장 작은 값의 순위 비율은 1 입니다 . 순위 비율은 p 값과 FDR p 값을 순위별로 유의성이 큰 것부터 그림에 표시하는 데 사용됩니다 .

R^2 (Y 가 연속형인 경우에만 표시됨) 모형에 의해 설명되는 총 변동의 비율을 측정하는 결정 계수입니다 .

카파 (Y 와 X 가 모두 범주형이고 수준 수가 같은 경우에만 표시됨) Y 와 X 간의 합치도 측도입니다 .

상관 (Y 와 X 가 모두 범주형인 경우에만 표시됨) 값 순서화로 정의된 인덱스를 기준으로 하는 Pearson 곱적률 상관계수입니다 .

시작 창에서 "로버스트" 옵션을 선택한 경우 다음 열이 "결과 테이블"에 추가됩니다. "로버스트" 옵션은 Y가 연속형일 때만 적용되므로 Y가 범주형이면 로버스트 열 셀이 비어 있습니다.

로버스트 p 값 로버스트 적합을 사용한 Y 와 X 변수 쌍에 해당하는 유의성 검정의 *p* 값입니다 .

로버스트 LogWorth $-\log_{10}$(로버스트 p 값) 의 계산 값입니다 .

로버스트 FDR p 값 Benjamini-Hochberg 기법을 사용하여 로버스트 p 값에 대해 계산된 False Discovery Rate 입니다 . 그룹 변수가 없으면 테이블에 표시된 모든 검정에 다중 검정 조정이 적용됩니다 . 그룹 변수가 있으면 그룹 변수의 각 수준에 대해 수행된 모든 검정에 다중 검정 조정이 적용됩니다 .

로버스트 FDR LogWorth $-\log_{10}$(로버스트 FDR p 값) 의 계산 값입니다 .

로버스트 순위 비율 로버스트 FDR LogWorth 순위를 검정 수의 비율로 나타낸 것입니다 .

로버스트 카이제곱 로버스트 검정과 관련된 카이제곱 값입니다 .

로버스트 시그마 오차 표준편차의 로버스트 추정값입니다 .

로버스트 이상치 비율 로버스트 평균에서의 거리가 로버스트 시그마의 3 배를 초과하는 값의 비율입니다 .

시작 창에서 "Cauchy" 옵션을 선택하면 다음 열이 " 결과 테이블 " 에 추가됩니다 . "Cauchy" 옵션은 Y 가 연속형일 때만 적용되므로 Y 가 범주형이면 Cauchy 열 셀이 비어 있습니다 .

Cauchy p 값 Cauchy 적합을 사용한 Y 와 X 변수 쌍에 해당하는 유의성 검정의 *p* 값입니다 .

Cauchy LogWorth $-\log_{10}$(Cauchy p 값) 의 계산 값입니다 .

Cauchy FDR p 값 Benjamini-Hochberg 기법을 사용하여 Cauchy p 값에 대해 계산된 False Discovery Rate 입니다 . 그룹 변수가 없으면 테이블에 표시된 모든 검정에 다중 검정 조정이 적용됩니다 . 그룹 변수가 있으면 그룹 변수의 각 수준에 대해 수행된 모든 검정에 다중 검정 조정이 적용됩니다 .

Cauchy FDR LogWorth $-\log_{10}$(Cauchy FDR p 값) 의 계산 값입니다 .

Cauchy 순위 비율 Cauchy FDR LogWorth 순위를 검정 수의 비율로 나타낸 것입니다 .

Cauchy 카이제곱 Cauchy 적합의 검정과 관련된 카이제곱 값입니다 .

Cauchy 시그마 Cauchy 적합으로부터의 오차 표준편차의 추정값입니다 .

Cauchy 이상치 비율 Cauchy 적합의 평균에서의 거리가 Cauchy 시그마의 3 배를 초과하는 값의 비율입니다 .

FDR p 값 그림

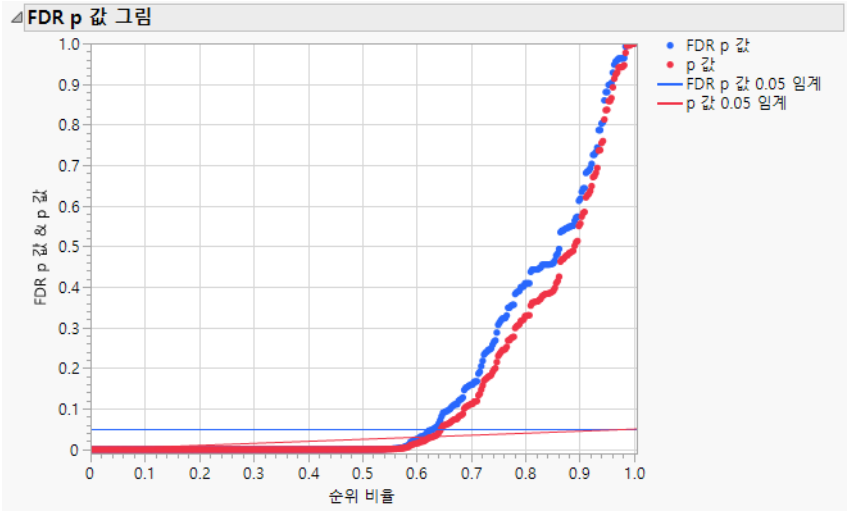
"FDR p 값 그림 " 보고서에는 순위 비율에 대한 FDR p 값 및 p 값 그림이 표시됩니다. 순위 비율은 p 값을 유의성이 큰 것부터 순서대로 정렬합니다. FDR p 값은 파란색으로 표시되고 p 값은 빨간색으로 표시됩니다.

파란색 가로선은 0.05 유의 수준을 나타냅니다. 세로 축을 두 번 클릭하고 현재 참조선을 제거한 후 새 참조선을 추가하여 이 수준을 변경할 수 있습니다.

증가하는 빨간색 선은 조정되지 않은 *p* 값에 대한 FDR 임계를 제공합니다. FDR 조정 *p* 값이 파란색 선 아래에 있으면 *p* 값이 빨간색 선 아래에 있습니다. 이를 통해 조정되거나 조정되지 않은 *p* 값 중 하나를 사용하여 FDR 을 기준으로 유의성을 이해할 수 있습니다.

그림 22.4에서는 Probe.jmp 샘플 데이터 테이블에 대한 FDR p 값 그림을 보여 줍니다. 일부 검정은 일반적인 *p* 값에 따라 유의하지만 FDR *p* 값에 따르면 유의하지 않습니다.

그림 22.4 FDR p 값 그림



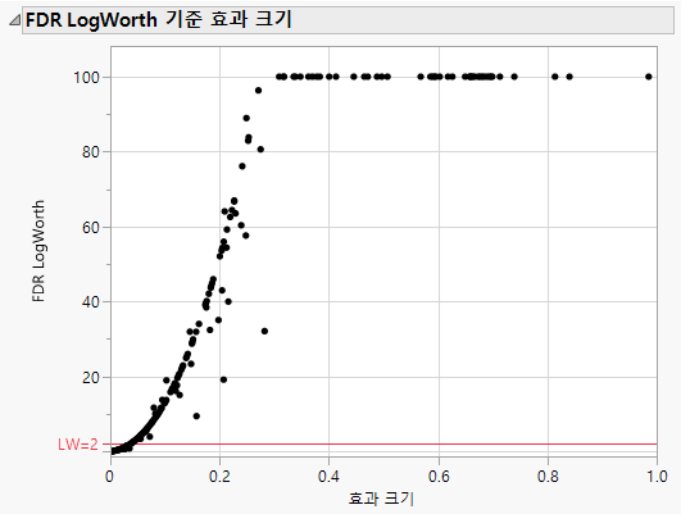
FDR LogWorth 대 효과 크기

효과가 크면 연관된 *p* 값이 매우 작은 경우가 많습니다. 이러한 작은 값을 그래픽으로 시각화하려면 어려울 수 있습니다. LogWorth($-\log_{10}(p$ 값)) 척도로 변환하면 매우 유의한 *p* 값은 LogWorth가 크고 유의하지 않은 *p* 값은 LogWorth가 낮습니다. LogWorth가 0이면 유의하지 않은 *p* 값 1에 해당합니다. LogWorth가 2를 초과하면 *p* 값이 0.01 미만입니다.

"FDR LogWorth 대 효과 크기" 그림에서 세로 축은 FDR LogWorth이고 가로 축은 효과 크기입니다. 일반적으로 효과가 클수록 *p* 값이 더 유의하고 LogWorth가 더 큽니다. 그러나 유의성은 오차 분산에도 영향을 받으므로 이 관계가 항상 성립하는 것은 아닙니다. 사실 오차 분산의 크기 때문에 큰 LogWorth가 작은 효과와 연관될 수 있고 작은 LogWorth가 큰 효과와 연관될 수 있습니다. "FDR LogWorth 대 효과 크기" 그림을 사용하여 이 관계를 살펴볼 수 있습니다.

그림 22.5에서는 최대 LogWorth가 100으로 설정된 Probe.jmp 샘플 데이터 테이블의 FDR LogWorth 대 효과 크기 그림을 보여 줍니다. 대부분의 FDR LogWorth 값이 2를 초과하므로 대부분의 효과가 0.01 수준에서 유의하다는 것을 알 수 있습니다. FDR LogWorth 값 100은 극도로 작은 *p* 값에 해당합니다.

그림 22.5 FDR LogWorth 대 효과 크기



FDR LogWorth 대 R^2

"FDR LogWorth 대 R^2" 그림의 세로 축에는 FDR LogWorth가 표시되고 가로 축에는 R^2 값이 표시됩니다. LogWorth 값이 클수록 더 큰 R^2 값과 연관되는 경향이 있지만 이 관계는 관측값 수에도 영향을 받습니다.

p 값 데이터 테이블

시작 창에서 "시작 시 p 값 테이블" 옵션을 선택한 경우 반응 변수 선별 플랫폼을 시작할 때 p 값 데이터 테이블이 생성됩니다. 이 데이터 테이블에는 "반응 변수 선별" 보고서의 "결과 테이블"과 동일한 정보가 포함됩니다. p 값 테이블에는 다음과 같은 모형 관련 열이 더 포함됩니다.

Y 평균 Y 의 평균입니다.

SSE Y 가 연속형일 때 표시됩니다. 오차의 제곱합입니다.

DFE Y 가 연속형일 때 표시됩니다. 오차의 자유도입니다.

MSE Y 가 연속형일 때 표시됩니다. 평균 제곱 오차입니다.

F 비 Y 가 연속형일 때 표시됩니다. 분산 분석 또는 회귀 검정의 F 비입니다.

절편 Y 와 X 가 모두 연속형일 때 표시됩니다. 해당하는 Y 및 X 변수 쌍과 관련된 회귀 모형의 절편입니다.

기울기 Y 와 X 가 모두 연속형일 때 표시됩니다. 해당하는 Y 및 X 변수 쌍과 관련된 회귀 모형의 기울기입니다.

DF Y 와 X 가 모두 범주형일 때 표시됩니다. 카이제곱 검정의 자유도입니다.

LR 카이제곱 Y 와 X 가 모두 범주형일 때 표시됩니다. 가능도비 카이제곱 통계량의 값입니다.

로버스트 경과 시간 (시작 창에서 " 로버스트 " 를 지정한 경우에만 표시됨) 로버스트 보고서를 생성하는 데 필요한 계산 시간 (초) 입니다 .

p 값 데이터 테이블 스크립트

관련 스크립트가 p 값 데이터 테이블에 저장됩니다. 이 중 하나를 제외한 모든 스크립트는 보고서에 제공된 그림을 재현합니다. p 값 테이블에서 행을 선택하는 경우 "선택한 항목 적합" 스크립트가 적절한 X로 Y 적합 분석을 생성합니다.

반응 변수 선별 플랫폼 옵션

" 반응 변수 선별 " 의 빨간색 삼각형 메뉴에는 화면 표시를 사용자 정의하고 계산된 데이터를 저장하는 옵션이 포함되어 있습니다 .

결과 테이블 결과 테이블을 표시하거나 숨깁니다 .

선택한 항목 적합 선택한 관계에 대해 적절한 X 로 Y 적합 보고서를 " 반응 변수 선별 " 보고서에 추가합니다 . 그림의 점 , 결과 테이블의 행 또는 p 값 데이터 테이블 (열린 경우) 의 행을 선택하여 관계를 선택할 수 있습니다 .

열 선택 결과 테이블에서 선택한 행 , 반응 변수 선별 보고서 창의 그림에서 선택한 점 또는 p 값 테이블 (열린 경우) 의 행에 해당하는 열을 원래 데이터 테이블에서 선택합니다 . 먼저 행 또는 점을 선택한 후 " 열 선택 " 을 선택합니다 . 데이터 테이블에서 해당 열이 선택됩니다 . 먼저 p 값 테이블의 추가 행 또는 그림의 점을 선택한 후 다시 " 열 선택 " 을 선택하면 이에 해당하는 열을 선택할 수 있습니다 . 다른 행이나 점에 해당하는 열을 선택하려면 먼저 원래 데이터 테이블에서 현재 열 선택을 지웁니다 .

선택 조건 " 선택 조건 " 창을 엽니다 . " 비교 " 메뉴와 " 값 " 텍스트 상자를 사용하여 결과 테이블에서 특정 조건에 해당하는 특정 반응을 선택할 수 있습니다 . 예를 들어 " 효과 크기 > 0.80" 에 해당하는 모든 반응을 선택할 수 있습니다 . " 확인 " 을 클릭하면 " 결과 테이블 " 에서 반응이 선택됩니다 .

팁 : " 결과 테이블 " 의 아무 곳이나 마우스 오른쪽 버튼으로 클릭하여 " 선택 조건 " 창에 액세스할 수도 있습니다 .

p 값 저장 p 값 데이터 테이블을 생성합니다 . 자세한 내용은 "p 값 데이터 테이블 " 에서 확인하십시오 .

평균 저장 연속형 Y 와 범주형 X 의 경우 범주형 변수의 각 수준에 대한 개수 , 평균 및 표준편차가 포함된 데이터 테이블을 생성합니다 . " 로버스트 " 옵션을 선택한 경우 로버스트 평균이 포함됩니다 .

평균 비교 저장 연속형 Y 와 범주형 X 의 경우 범주형 변수의 모든 수준에서 전체 쌍별 비교를 검정합니다 . 각 비교에 대해 데이터 테이블에는 일반적인 *t*- 검정 , 실제적 유의성 검정 , 동등성 검정 및 결과 요약을 위해 색상 코딩을 사용하는 열이 제공됩니다. 실제적 LogWorth 대 상대 실제적 차이 그림을 표시하는 "Practical LogWorth by Relative Practical Difference" 스크립

트도 데이터 테이블에 포함되어 있습니다. 자세한 내용은 "평균 비교 데이터 테이블" 에서 확인하십시오. 예는 "실제적 유의성 및 동등성 검정의 예" 에서 확인하십시오.

표준화 잔차 저장 "잔차 그룹" 이라는 새 열 그룹을 원래 데이터 테이블에 저장하고 이 열 그룹에 열을 배치합니다. 각 연속형 Y 와 범주형 X 에 대해 잔차를 추정된 표준편차로 나눈 값을 포함하는 열이 생성됩니다. 즉, 표준화 잔차가 열에 포함됩니다. 열은 계산식으로 정의됩니다.

"로버스트" 옵션을 선택한 경우 로버스트 적합 및 로버스트 추정값을 사용하여 표준화 잔차 열이 생성됩니다.

이상치 표시자 저장 "이상치 그룹" 이라는 새 열 그룹을 원래 데이터 테이블에 저장하고 이 열 그룹에 열을 배치합니다. "이상치 표시자 저장" 은 "로버스트" 옵션을 선택한 경우에 가장 효과적입니다.

각 연속형 Y 와 범주형 X 에 대해 이상치를 나타내는 열이 생성됩니다. 이상치는 예측값까지의 거리가 시그마 추정값의 3 배를 초과하는 점입니다. 즉, 이상치는 표준화 잔차가 3 을 초과하는 점입니다. 열은 계산식으로 정의됩니다.

"로버스트" 옵션을 선택한 경우 로버스트 적합 및 로버스트 추정값이 사용됩니다. 이상치는 예측값까지의 거리가 로버스트 시그마 추정값의 3 배를 초과하는 점입니다.

"이상치 군집화" 스크립트가 원래 데이터 테이블에 추가됩니다. 이 스크립트는 데이터의 계층적 군집화 그림에 이상치를 표시합니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

평균 데이터 테이블

평균 데이터 테이블에는 반응과 X 수준의 각 조합에 대한 행이 포함되어 있습니다. Probe.jmp 샘플 데이터 테이블의 경우 387 개의 반응 변수가 있고 각 변수가 두 수준의 Process 에 대해 검정되었습니다. 평균 테이블에는 387x2 = 774 개 행이 포함됩니다.

그림 22.6 평균 데이터 테이블

Save Means

열(6/0)

Y
X
수준
개수
평균
표준편차

행

모든 행	774
선택됨	0
제외	0
숨김	0
라벨 항목	0

	Y	X	수준	개수	평균	표준편차
1	DELL_RPNBR	Process	New	3044	0.2035840106	0.1916031302
2	DELL_RPNBR	Process	Old	2750	0.2346329436	0.1278299414
3	DELL_RPPBR	Process	New	3044	-0.068072506	0.1784161107
4	DELL_RPPBR	Process	Old	2750	-0.043321135	1.9165048167
5	DELW_M1	Process	New	3039	-0.04387197	1.0774346893
6	DELW_M1	Process	Old	2750	-0.071400861	0.0389764703
7	DELW_M2	Process	New	3039	0.8014275806	0.0698264259
8	DELW_M2	Process	Old	2689	0.7274958056	0.0649383327
9	DELW_NBASE	Process	New	3028	0.1957655669	77.801881218
10	DELW_NBASE	Process	Old	2750	1.4765185729	0.1326787074
11	DELW_NEMIT	Process	New	3038	0.3411755212	0.0940762694
12	DELW_NEMIT	Process	Old	2750	0.3813085954	0.0976561441
13	DELW_NENBNI	Process	New	3043	3.7172025841	1.1419433505
14	DELW_NENBNI	Process	Old	2749	4.6556238745	0.1893626325
15	DELW_NSINK	Process	New	3032	9.2876152816	1.2840573691
16	DELW_NSINK	Process	Old	2750	7.8902297629	0.3781489659
17	DELW_PBASE	Process	New	3036	1.1651478779	0.1351193171

평균 데이터 테이블에는 다음 열이 포함됩니다.

Y 연속형 반응 변수입니다.

X 범주형 변수입니다.

수준 범주형 X 변수의 수준입니다.

개수 해당 수준의 값 개수입니다.

평균 지정된 수준에 대한 Y 변수의 평균입니다.

표준편차 지정된 수준에 대한 Y 변수의 표준편차입니다.

로버스트 평균 평균의 로버스트 M-추정값입니다. 시작 창에서 "로버스트" 옵션을 선택한 경우 표시됩니다.

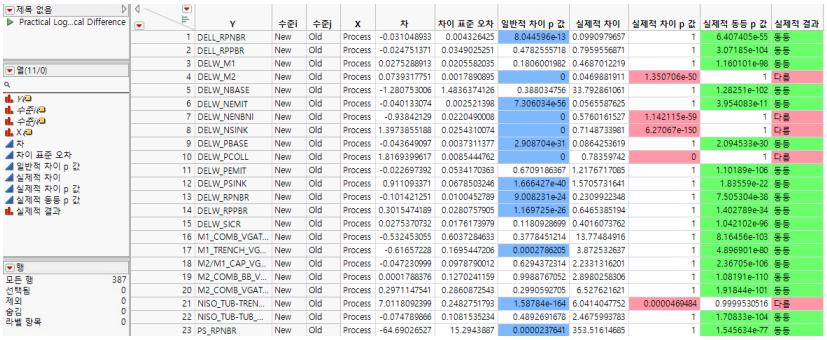
평균 비교 데이터 테이블

데이터 테이블이 많은 수의 행으로 구성된 경우, 즉 n이 큰 경우 검정에 사용되는 표준 오차는 매우 작을 수 있습니다. 이로 인해 사실은 관측된 차이가 너무 작아서 실제로 중요하지 않지만 검정 결과는 통계적으로 유의할 수 있습니다. 실제적 유의성 검정을 수행하면 감지하도록 고려할 차이의 크기를 지정할 수 있습니다. 이 차이를 **실제적 차이**라고 합니다. 차이가 0인지 검정하는 대신 차이가 실제적 차이를 초과하는지 여부를 검정합니다. 결과적으로 검정이 더 의미 있고 자세히 조사할 검정 수가 적습니다.

동등성 검정을 사용하면 반응에 대한 두 수준의 효과가 본질적으로 동일한지 여부를 실제적 관점에서 판단할 수 있습니다. 즉, 동등성 검정은 차이가 실제적 차이보다 작은지 여부를 검정합니다.

평균 비교 데이터 테이블에는 실제적 차이 검정과 실제적 동등성 검정 둘 다에 대한 결과가 제공됩니다. 각 행은 범주형 요인의 두 수준에 대한 반응을 비교합니다. 쌍별 비교 결과는 쉽게 해석할 수 있도록 색상으로 구분됩니다. 실제적 차이를 지정하는 방법에 대한 설명은 "실제적 차이"에서 확인하십시오. 예는 "실제적 유의성 및 동등성 검정의 예"에서 확인하십시오.

그림 22.7 평균 비교 데이터 테이블



제목 없음
Practical Log...cal Difference

열(11/0)
Y
수준i
수준j
X
차
차이 표준 오차
일반적 차이 p 값
실제적 차이
실제적 차이 p 값
실제적 동등 p 값
실제적 결과

행
모든 행 387
선택됨 0
제외 0
숨김 0
라벨 함목 0

	Y	수준i	수준j	X	차	차이 표준 오차	일반적 차이 p 값	실제적 차이	실제적 차이 p 값	실제적 동등 p 값	실제적 결과
1	DELL_RPNBR	New	Old	Process	-0.031048933	0.004326425	8.044596e-13	0.0990979657	1	6.407405e-55	동등
2	DELL_RPPBR	New	Old	Process	-0.024751371	0.0349025251	0.4782555718	0.7959556871	1	3.07185e-104	동등
3	DELW_M1	New	Old	Process	0.0275288913	0.0205582035	0.1806001982	0.4687012219	1	1.160101e-98	동등
4	DELW_M2	New	Old	Process	0.0739317751	0.0017890895	0	0.0469881911	1.350706e-50	1	다름
5	DELW_NBASE	New	Old	Process	-1.280753006	1.4836374126	0.388034756	33.792861061	1	1.28251e-102	동등
6	DELW_NEMIT	New	Old	Process	-0.040133074	0.002521398	7.306034e-56	0.0565587625	1	3.954083e-11	동등
7	DELW_NENBNI	New	Old	Process	-0.93842129	0.0220490008	0	0.5760161527	1.142115e-59	1	다름
8	DELW_NSINK	New	Old	Process	1.3973855188	0.0254310074	0	0.7148733981	6.27067e-150	1	다름
9	DELW_PBASE	New	Old	Process	-0.043649097	0.0037311377	2.908704e-31	0.0864253619	1	2.094533e-30	동등
10	DELW_PCOLL	New	Old	Process	1.8169399617	0.0085444762	0	0.78359742	0	1	다름
11	DELW_PEMIT	New	Old	Process	-0.022697392	0.0534170363	0.6709186367	1.2176717085	1	1.10189e-106	동등
12	DELW_PSINK	New	Old	Process	0.911093371	0.0678503246	1.666427e-40	1.5705731641	1	1.83559e-22	동등
13	DELW_RPNBR	New	Old	Process	-0.101421251	0.0100452789	9.008231e-24	0.2309922348	1	7.505304e-38	동등
14	DELW_RPPBR	New	Old	Process	0.3015474189	0.0280757905	1.169725e-26	0.6465385194	1	1.402789e-34	동등
15	DELW_SICR	New	Old	Process	0.0275370732	0.0176173979	0.1180928699	0.4016073762	1	1.042102e-96	동등
16	M1_COMB_VGAT...	New	Old	Process	-0.532453055	0.6037284633	0.3778451214	13.77484916	1	8.16456e-103	동등
17	M1_TRENCH_VG...	New	Old	Process	-0.61657228	0.1695447206	0.0002786205	3.872532637	1	4.896901e-80	동등
18	M2/M1_CAP_VG...	New	Old	Process	-0.047230999	0.0978790012	0.6294372314	2.2331316201	1	2.36705e-106	동등
19	M2_COMB_88_V...	New	Old	Process	0.0001788376	0.1270241159	0.9988767052	2.8980258306	1	1.08191e-110	동등
20	M2_COMB_VGAT...	New	Old	Process	0.2971147541	0.2860872543	0.2990592705	6.5276721621	1	1.81844e-101	동등
21	NISO_TUB-TREN...	New	Old	Process	7.0118092399	0.2482751793	1.58784e-164	6.0414047752	0.0000469484	0.9999530516	다름
22	NISO_TUB-TUB_...	New	Old	Process	-0.074789866	0.1081535234	0.4892691678	2.4675993783	1	1.70833e-104	동등
23	PS_RPNBR	New	Old	Process	-64.69026527	15.2943887	0.0000237641	353.51614685	1	1.545634e-77	동등

평균 비교 데이터 테이블에는 실제적 LogWorth 대 상대 실제적 차이 그림을 표시하는 "Practical LogWorth by Relative Practical Difference" 스크립트가 포함되어 있습니다 . 상대 실제적 차이는 실제 차이를 실제적 차이로 나눈 값으로 정의됩니다 .

Y 연속형 반응 변수입니다 .

X 범주형 변수입니다 .

수준 i 범주형 X 변수의 수준입니다 .

수준 j 수준 i 와 비교되는 범주형 X 변수의 수준입니다 .

차 두 수준에서 추정된 평균 차이입니다 . " 로버스트 " 옵션을 선택한 경우 평균의 로버스트 추정값이 사용됩니다 .

차이 표준 오차 평균 차이의 표준 오차입니다 . " 로버스트 " 옵션을 선택한 경우 이 값은 로버스트 추정값입니다 .

일반적 차이 p 값 쌍별 비교를 위한 일반적인 스튜던트 *t*- 검정의 *p* 값입니다 . " 로버스트 " 옵션을 선택한 경우 이 값은 *t*- 검정의 로버스트 버전입니다 . 0.05 수준에서 유의한 검정이 강조 표시됩니다 .

실제적 차이 실제적 관심 영역으로 간주할 평균 차이입니다 . Y 변수에 " 규격 한계 " 특성을 할당하는 경우 실제적 차이는 규격 한계 간의 차이에 실제적 차이 비율을 곱한 값으로 계산됩니다 . " 실제적 차이 비율 " 을 지정하지 않은 경우 실제적 차이는 규격 한계 간의 차이에 0.10 을 곱한 값입니다 .

Y 변수에 " 규격 한계 " 특성을 할당하지 않은 경우 표준편차 추정값은 IQR(사분위수 범위) 에서 계산됩니다 . 이 추정값은 $\hat{\sigma} = (IQR)/(1.3489795)$ 입니다 . 실제적 차이는 $6\hat{\sigma}$ 에 실제적 차이 비율을 곱한 값으로 계산됩니다 . " 실제적 차이 비율 " 을 지정하지 않은 경우 실제적 차이는 $6\hat{\sigma}$ 에 0.10 을 곱한 값으로 계산됩니다 .

실제적 차이 p 값 수준 i 와 수준 j 간의 Y 평균 차이의 절대값이 실제적 차이보다 작거나 같은지 여부를 검정하기 위한 *p* 값입니다 . *p* 값이 작으면 절대 차이가 실제적 차이를 초과한다는 것을 나타냅니다 . 이는 수준 i 와 수준 j 가 실제적으로 중요한 차이를 설명한다는 의미입니다 .

실제적 동등 p 값 TOST(Two One-Sided Tests: 두 개의 단측 검정) 방법을 사용하여 평균 간의 실제적 차이를 검정합니다 (Schuirmann 1987). 실제적 차이는 더 작은 차이가 실제적으로 동등한 것으로 간주되는 임계 차이를 지정합니다. 두 개의 귀무가설, 즉 실제 차이가 실제적 차이를 초과한다는 가설과 실제 차이가 실제적 차이의 음의 값보다 작다는 가설에 대해 단측 *t*- 검정이 생성됩니다. 두 검정이 모두 기각되면 평균의 절대 차이가 실제적 차이 내에 있음을 나타냅니다. 따라서 그룹은 실제적으로 동등하다고 간주됩니다.

실제적 동등 p 값은 단측 *t*- 검정에서 구한 가장 큰 *p* 값입니다. 실제적 동등 p 값이 작으면 실제적으로 수준 i 의 평균 반응이 수준 j 의 평균과 같다는 것을 나타냅니다.

실제적 결과 실제적 차이 및 동등성 검정 결과에 대한 설명입니다. 값은 유의한 결과를 식별하는 데 도움이 되도록 색상으로 구분됩니다.

- 다름 (분홍색): 절대 차이가 실제적 차이보다 유의하게 크다는 것을 나타냅니다.
- 동등 (녹색): 절대 차이가 실제적 차이 내에 유의하게 포함된다는 것을 나타냅니다.
- 판단할 수 없음 (회색): 실제적 차이 검정과 실제적 동등성 검정이 모두 유의하지 않다는 것을 나타냅니다.

모형 적합의 반응 변수 선별 분석법

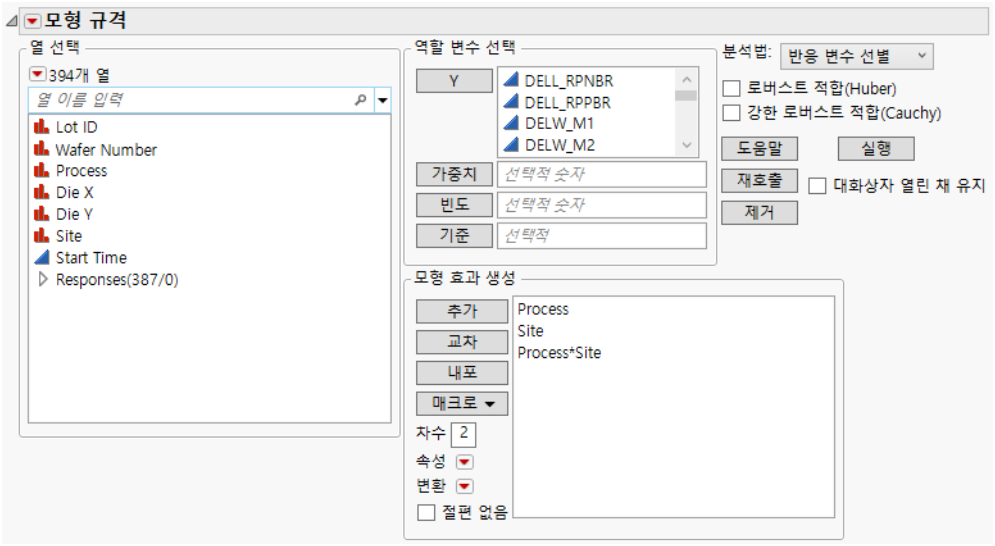
선형 모형 효과에 대한 단변량 검정에 관심이 있는 경우 모형 적합의 반응 변수 선별 분석법을 적합시킬 수 있습니다. 생성된 보고서 및 테이블은 모든 모형 효과에 대해 모든 반응을 검정합니다.

- " 모형 적합의 반응 변수 선별 시작 "
- " 반응 변수 선별 적합 보고서 "
- " 반응 변수 선별 적합 옵션 "
- "PValues 데이터 테이블 "
- "Y Fits 데이터 테이블 "

모형 적합의 반응 변수 선별 시작

분석 > 모형 적합을 선택합니다 . Y 및 모형 효과를 입력합니다 . " 분석법 " 목록에서 **반응 변수 선별**을 선택합니다 .

그림 22.8 모형 적합 창의 반응 변수 선별



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로버스트 추정을 위한 다음 두 가지 옵션이 있습니다 .

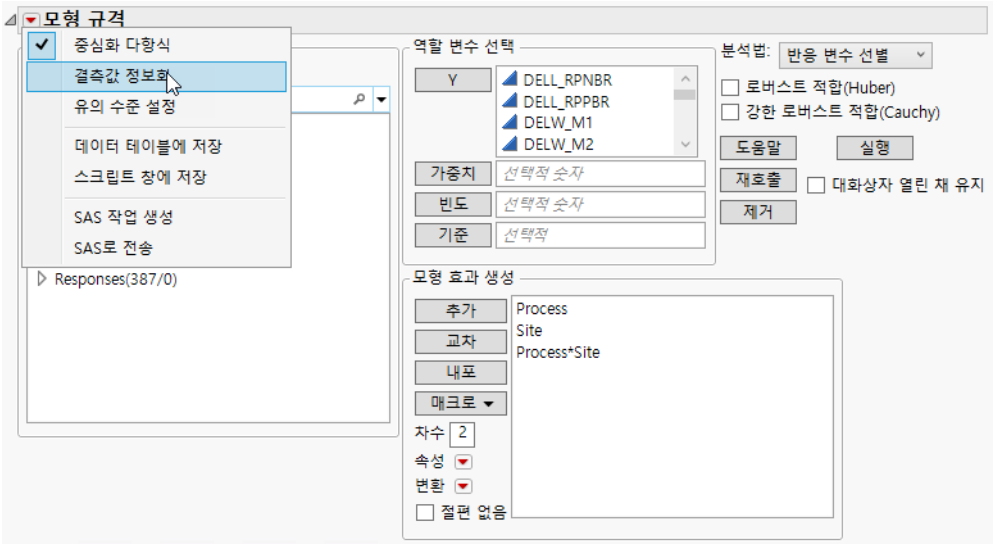
로버스트 적합 (Huber) 연속형 반응의 이상치에 대한 비중 강하 (down weight) 를 위해 로버스트 (Huber) 추정을 지정합니다 . 이상치가 없으면 이러한 추정값이 최소 제곱 추정값에 가깝습니다 . 이 옵션을 사용하면 처리 시간이 늘어납니다 .

강한 로버스트 적합 (Cauchy) 연속형 반응의 이상치에 대한 비중 강하 (down weight) 를 위해 강한 로버스트 (Cauchy) 추정을 지정합니다 . 오차가 Cauchy 분포를 따른다고 가정합니다 . Cauchy 분포는 정규 분포보다 꼬리가 더 두꺼우므로 이상치의 영향이 줄어듭니다 . 이 옵션은 데이터에 포함된 이상치의 비율이 높은 경우에 유용할 수 있습니다 . 하지만 데이터가 정규 분포에 가깝고 이상치가 일부만 있는 경우에 이 옵션을 사용하면 추론이 잘못될 수 있습니다 . Cauchy 옵션은 최대 가능도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다 .

팁 : 두 로버스트 옵션을 모두 선택하면 플랫폼에서 Cauchy 추정만 사용합니다 .

결측값 정보화 옵션은 결측값에 대한 코딩 시스템을 제공합니다. 결측값 정보화 코딩을 사용하면 결측값이 있어도 예측 모형을 추정할 수 있습니다. 이 옵션은 결측 데이터에 정보가 포함된 경우 유용합니다. "모형 규격"의 빨간색 삼각형 메뉴에서 이 옵션을 선택합니다.

그림 22.9 결측값 정보화 옵션



모형 적합 창에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

반응 변수 선별 적합 보고서

" 반응 변수 선별 적합 " 보고서에는 다음과 같은 두 개의 그림이 표시됩니다 .

- FDR PValue 그림
- FDR LogWorth 대 순위 비율 그림

"FDR PValue 그림"은 플랫폼 자체와 동일한 방식으로 해석됩니다. 자세한 내용은 "반응 변수 선별 보고서"에서 확인하십시오.

"FDR LogWorth 대 순위 비율 " 그림에는 p 값 순위에 대한 FDR LogWorth 값이 표시됩니다 . 표시된 점은 순위 비율이 증가하면 감소하거나 일정하게 유지됩니다 . 이 그림은 검정의 유의한 비율을 나타냅니다 . 반응 변수 선별 분석법을 사용한 예는 " 반응 변수 선별 분석법 " 에서 확인하십시오 .

반응 변수 선별 적합 옵션

" 반응 변수 선별 적합 " 의 빨간색 삼각형 메뉴에는 계산된 데이터를 저장하는 옵션이 포함되어 있습니다 .

모형 대화상자 제공된 보고서를 얻기 위해 실행한 모형 대화상자가 포함된 창을 엽니다 .

추정값 저장 각 행이 반응에 해당하고 열이 모형 항에 해당하는 데이터 테이블을 엽니다. 항목은 지정된 모형을 적합시켜 구한 모수 추정값입니다. 이 데이터 테이블에는 분석에 사용된 데이터 테이블 이름을 제공하는 " 원래 데이터 " 라는 테이블 변수도 포함되어 있습니다 . 기준 변

수를 지정한 경우 JMP 에서 기준 변수의 각 수준에 대한 추정값 테이블이 생성되고 " 원래 데이터 " 변수는 기준 변수와 해당 수준을 제공합니다 .

예측 계산식 저장 모든 반응에 대한 예측 방정식이 포함된 열을 원래 데이터 테이블에 추가합니다 .

최소 제곱 평균 저장 각 행이 반응 및 효과 설정 조합에 해당하는 데이터 테이블을 엽니다 . 이 행은 해당 설정 조합의 표준 오차 및 최소 제곱 평균을 포함합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

PValues 데이터 테이블

"PValues" 데이터 테이블에는 Y 변수와 모형 효과로 구성된 각 쌍에 대한 행이 포함되어 있습니다 . 테이블의 열은 다음 항목을 포함합니다 . 시작 창에서 " 로버스트 적합 " 옵션을 선택하면 Huber M- 추정을 사용하여 모형이 적합됩니다 .

Y 지정된 반응 열입니다 .

Effect 지정된 모형 효과입니다 .

FRatio 효과 (Effect) 검정에 대한 검정 통계량입니다 . 이 값은 최소 제곱 적합의 " 효과 검정 " 보고서에 나오는 값과 같습니다 .

PValue F 비 (FRatio) 에 해당하는 유의성 검정의 p 값입니다 . 효과 검정에 대한 자세한 내용은 선형 모형 적합의 에서 확인하십시오 .

LogWorth $-\log_{10}(p$ 값) 의 계산 값입니다 . 이 변환은 그래프 작성에 적절한 척도를 제공하도록 p 값을 조정합니다 . $-\log_{10}(0.01) = 2$ 이므로 2 를 초과하는 값은 0.01 수준에서 유의합니다 .

FDR PValue Benjamini-Hochberg 기법을 사용하여 계산된 False Discovery Rate p 값입니다 . 이 기법은 여러 검정에 대한 FDR(False Discovery Rate) 을 제어하기 위해 p 값을 조정합니다 . FDR 수정에 대한 자세한 내용은 Benjamini and Hochberg 연구 자료 (1995) 에서 확인하십시오 . False Discovery Rate 에 대한 자세한 내용은 "False Discovery Rate" 또는 Westfall et al. 연구 자료 (2011) 에서 확인하십시오 .

FDR LogWorth $-\log_{10}$(FDR p 값) 의 계산 값입니다 . 이 통계량은 유의성을 그림에 표시하고 평가하는 데 가장 적합합니다 . p 값이 작으면 FDR LogWorth 값이 커집니다 .

Rank Fraction FDR LogWorth 순위를 검정 수의 비율로 나타낸 것입니다. 검정 수가 m 이면 가장 큰 FDR LogWorth 값의 순위 비율은 $1/m$ 이고 가장 작은 값의 순위 비율은 1 입니다. 마찬가지로 순위 비율 (Rank Fraction) 은 p 값을 검정 수의 비율로 작은 것부터 순서대로 정렬합니다. 순위 비율은 p 값과 FDR p 값을 순위별로 유의성이 큰 것부터 그림에 표시하는 데 사용됩니다.

Test DF 효과 검정에 대한 자유도입니다.

"PValues" 데이터 테이블에는 분석에 사용된 데이터 테이블 이름을 제공하는 "원래 데이터"라는 테이블 변수도 포함되어 있습니다. 기준 변수를 지정한 경우 JMP 에서 기준 변수의 각 수준에 대한 PValues 테이블이 생성되고 "원래 데이터" 변수는 기준 변수와 해당 수준을 제공합니다.

Y Fits 데이터 테이블

"Y Fits" 데이터 테이블에는 Y 변수에 대한 행이 포함되어 있습니다. 테이블의 열에는 각 Y 에 대해 모형 적합에 대한 정보가 요약되어 있습니다. 시작 창에서 " 로버스트 적합 " 옵션을 선택하면 Huber M- 추정을 사용하여 모형이 적합됩니다.

Y 지정된 반응 열입니다.

RSquare 다중 상관계수입니다.

RMSE 제곱근 평균 제곱 오차입니다.

Count 관측값 수 또는 가중치 변수의 합입니다.

Overall FRatio 최소 제곱 적합의 " 분산 분석 " 보고서에 나오는 모형 적합에 대한 검정 통계량입니다.

Overall PValue 전체 모형 유의성 검정에 대한 p 값입니다.

Overall LogWorth 전체 모형 유의성 검정에 대한 p 값의 LogWorth 입니다.

Overall FDR PValue False Discover Rate 에 대해 조정된 전체 p 값입니다. 자세한 내용은 " 반응 변수 선별 보고서 " 에서 확인하십시오.

Overall FDR LogWorth Overall FDR PValue 의 LogWorth 입니다.

Overall Rank Fraction Overall FDR LogWorth 순위를 검정 수의 비율로 나타낸 것입니다. 검정 수가 m 이면 가장 큰 Overall FDR LogWorth 값의 순위 비율은 $1/m$ 이고 가장 작은 값의 순위 비율은 1 입니다.

< 효과 > PValue 이러한 열에는 각 모형 효과 검정의 p 값이 포함됩니다. 이러한 열은 열 패널에서 "PValue" 그룹으로 정렬됩니다.

< 효과 > LogWorth 이러한 열에는 각 모형 효과 검정의 p 값에 대한 LogWorth 가 포함됩니다. 이러한 열은 열 패널에서 "LogWorth" 그룹으로 정렬됩니다.

< 효과 > FDR LogWorth 이러한 열에는 각 모형 효과 검정의 FDR LogWorth 가 포함됩니다. 이러한 열은 열 패널에서 "FDR LogWorth" 그룹으로 정렬됩니다.

"Y Fits" 데이터 테이블에는 분석에 사용된 데이터 테이블 이름을 제공하는 " 원래 데이터 " 라는 테이블 변수도 포함되어 있습니다. 기준 변수를 지정한 경우 JMP 에서 기준 변수의 각 수

준에 대한 Y Fits 테이블이 생성되고 " 원래 데이터 " 변수는 기준 변수와 해당 수준을 제공합니다 .

반응 변수 선별의 추가 예

- " 실제적 유의성 및 동등성 검정의 예 "
- " 최대 LogWorth 옵션의 예 "
- " 로버스트 적합의 예 "
- " 반응 변수 선별 분석법 "

실제적 유의성 및 동등성 검정의 예

이 예에서는 Probe.jmp 샘플 데이터 테이블을 사용하여 실제적 차이를 검정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Probe.jmp 를 엽니다 .
2. **분석 > 선별 > 반응 변수 선별**을 선택합니다 .

   " 반응 변수 선별 " 시작 창이 나타납니다 .
3. Responses 열 그룹을 선택하고 **Y, 반응**을 클릭합니다 .
4. Process 를 선택하고 **X** 를 클릭합니다 .
5. " 실제적 차이 비율 " 상자에 "0.15" 를 입력합니다 .
6. **확인**을 클릭합니다 .
7. " 반응 변수 선별 " 의 빨간색 삼각형을 클릭하고 **평균 비교 저장**을 선택합니다 .

   그림 22.10 에는 데이터 테이블의 일부가 표시됩니다 . Y 의 각 반응에 대해 해당 행은 Process 의 New 수준과 Old 수준에 대한 검정 정보를 제공합니다 .

그림 22.10 평균 비교 테이블 (일부분)

	Y	수준i	수준j	X	차	차이 표준 오차	일반적 차이 p 값	실제적 차이	실제적 차이 p 값	실제적 동등 p 값	실제적 결과
1	DELL_RPNBR	New	Old	Process	-0.031048933	0.004326425	8.044596e-13	0.1486469486	1	1.57079e-153	동등
2	DELL_RPPBR	New	Old	Process	-0.024751371	0.0349025251	0.4782555718	1.1939335307	1	2.40983e-225	동등
3	DELW_M1	New	Old	Process	0.0275288913	0.0205582035	0.1806001982	0.7030518329	1	1.40119e-217	동등
4	DELW_M2	New	Old	Process	0.0739317751	0.0017890895	0	0.0704822866	0.0269479498	0.9730520502	다름
5	DELW_NBASE	New	Old	Process	-1.280753006	1.4836374126	0.388034756	50.689291592	1	6.62236e-223	동등
6	DELW_NEMIT	New	Old	Process	-0.040133074	0.002521398	7.306034e-56	0.0848381437	1	7.74847e-69	동등
7	DELW_NENBNI	New	Old	Process	-0.93842129	0.0220490008	0	0.864024229	0.0003726256	0.9996273744	다름
8	DELW_NSINK	New	Old	Process	1.3973855188	0.0254310074	0	1.0723100972	3.228438e-37	1	다름
9	DELW_PBASE	New	Old	Process	-0.043649097	0.0037311377	2.908704e-31	0.1296380428	1	8.21675e-113	동등
10	DELW_PCOLL	New	Old	Process	1.8169399617	0.0085444762	0	1.17539613	0	1	다름
11	DELW_PEMIT	New	Old	Process	-0.022697392	0.0534170363	0.6709186367	1.8265075627	1	1.2499e-228	동등
12	DELW_PSINK	New	Old	Process	0.911093371	0.0678503246	1.666427e-40	2.3558597461	1	3.147655e-97	동등
13	DELW_RPNBR	New	Old	Process	-0.101421251	0.0100452789	9.008231e-24	0.3464883523	1	1.6441e-125	동등
14	DELW_RPPBR	New	Old	Process	0.3015474189	0.0280757905	1.169725e-26	0.9698077792	1	7.48088e-120	동등
15	DELW_SICR	New	Old	Process	0.0275370732	0.0176173979	0.1180928699	0.6024110643	1	7.77189e-215	동등
16	M1_COMB_VGAT...	New	Old	Process	-0.532453055	0.6037284633	0.3778451214	20.662273741	1	1.83138e-223	동등
17	M1_TRENCH_VG...	New	Old	Process	-0.61657228	0.1695447206	0.0002786205	5.8087989555	1	2.39165e-191	동등
18	M2/M1_CAP_VG...	New	Old	Process	-0.047230999	0.0978790012	0.6294372314	3.3496974301	1	2.54456e-228	동등
19	M2_COMB_BB_V...	New	Old	Process	0.0001788376	0.1270241159	0.9988767052	4.3470387459	1	3.07472e-234	동등
20	M2_COMB_VGAT...	New	Old	Process	0.2971147541	0.2860872543	0.2990592705	9.7914324315	1	1.42412e-221	동등
21	NISO_TUB-TREN...	New	Old	Process	7.0118092399	0.2482751793	1.58784e-164	9.0621071628	1	9.077234e-17	동등
22	NISO_TUB-TUB_...	New	Old	Process	-0.074789866	0.1081535234	0.4892691678	3.7013990674	1	8.97914e-226	동등
23	PS_RPNBR	New	Old	Process	-64.69026527	15.2943887	0.0000237641	530.27422027	1	3.12381e-189	동등

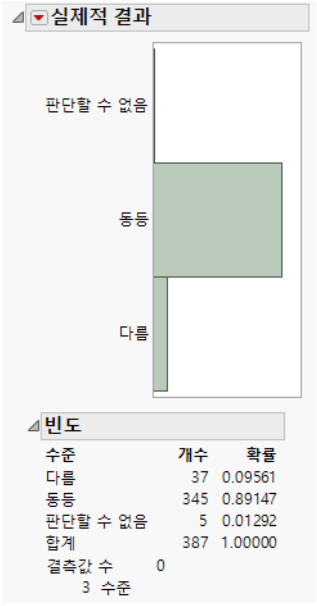
Probe.jmp 에 규격 한계가 열 특성으로 저장되지 않았으므로 JMP 에서 각 반응의 실제적 차이에 대한 값이 계산됩니다 . 사용자가 지정한 실제적 차이 0.15 에 반응의 6σ 범위 추정값을 곱합니다 . 이 값을 사용하여 실제적 차이 및 동등성을 검정합니다 . 이 값은 **실제적 차이** 열에 표시됩니다 .

일반적 차이 열에는 *p* 값이 유의성을 나타내는 반응이 표시됩니다 . " 일반적 차이 p 값 " 및 " 실제적 동등 p 값 " 열은 실제적 차이와 실제적 동등성 검정에 대한 *p* 값을 제공합니다 . 많은 열이 **통계적으로** 유의한 차이를 보이지만 **실제적으로** 유의한 차이는 보이지 않습니다 .

8. 평균 비교 데이터 테이블을 표시하고 **분석 > 분포**를 선택합니다 .
9. **실제적 결과**를 선택하고 **Y, 열**을 클릭합니다 .
10. **확인**을 클릭합니다 .

그림 22.11 에서는 실제적 유의성 결과의 분포를 보여 줍니다 . 지정된 실제적 차이에 대한 검정에 따라 37 개의 검정만 다릅니다 . 5 개의 반응은 검정으로 판단할 수 없었습니다 . 즉 , 반응이 Process 간에 실제적 차이를 나타내는지 여부를 구별할 수 없습니다 .

그림 22.11 실제적 유의성 결과의 분포



그림에서 해당 막대를 클릭하여 추가 연구에 사용할 37 개 반응을 선택할 수 있습니다 .

최대 LogWorth 옵션의 예

반응 변수 선별의 " 최대 LogWorth" 옵션을 사용하면 LogWorth 척도가 매우 큰 값으로 왜곡되지 않도록 제어할 수 있습니다 . 데이터 집합에 관측값이 많은 경우 *p* 값이 매우 작을 수 있습니다 .

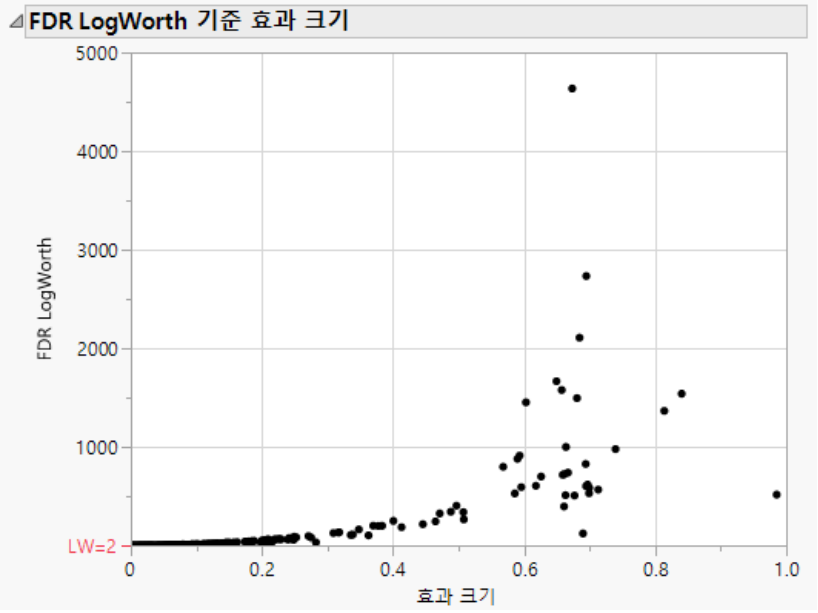
LogWorth 값은 이러한 경우 *p* 값을 그래픽으로 연구하는 유용한 방법을 제공합니다 . 일부 데이터 집합에서는 *p* 값이 너무 작아 LogWorth 척도가 매우 큰 값으로 왜곡됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Probe.jmp 를 엽니다 .
2. **분석 > 선별 > 반응 변수 선별**을 선택합니다 .
3. " 반응 변수 선별 " 시작 창에서 Responses 열 그룹을 선택하고 **Y, 반응**을 클릭합니다 .
4. Process 를 선택하고 **X** 를 클릭합니다 .
5. **로버스트** 체크박스를 선택합니다 .
6. **확인**을 클릭합니다 .

   수치가 많아서 분석을 완료하는 데 시간이 걸릴 수 있습니다 .
7. " 반응 변수 선별 " 보고서에서 " 로버스트 FDR LogWorth 대 효과 크기 " 보고서를 엽니다 .

   " 로버스트 FDR LogWorth" 값 (약 58,000) 이 너무 커서 그림의 상세 정보를 보기가 어렵습니다 . 그래프에 상세 정보가 충분히 표시되게 하려면 LogWorth 의 최대값을 설정합니다 .

그림 22.12 로버스트 FDR LogWorth 대 효과 크기 (최대 LogWorth 설정 안 함)

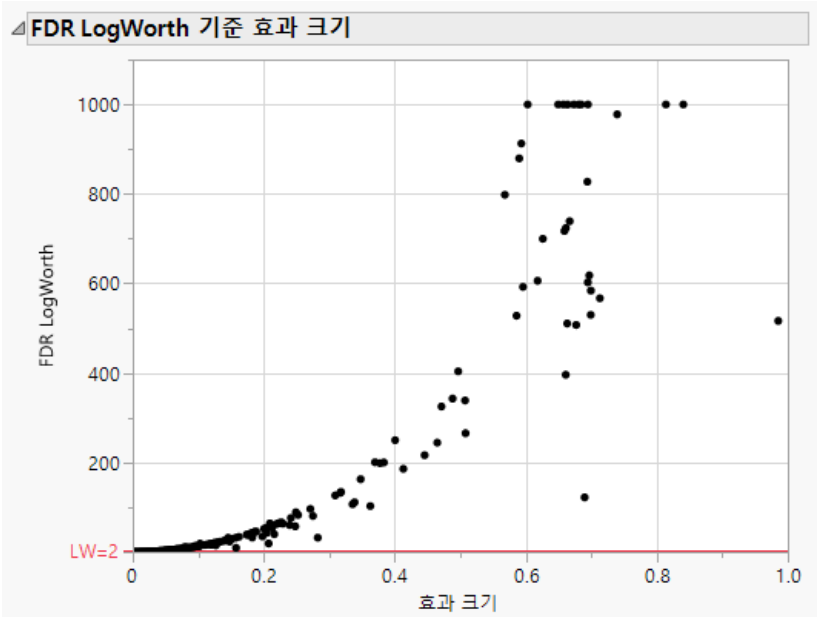


8. 1 단계 ~ 5 단계를 반복합니다 .
9. 시작 창 맨 아래의 " 최대 LogWorth" 상자에 "1000" 을 입력합니다 .
10. **확인**을 클릭합니다 .

    분석을 완료하는 데 시간이 걸릴 수 있습니다 .
11. " 반응 변수 선별 " 보고서에서 " 로버스트 FDR LogWorth 대 효과 크기 " 보고서를 엽니다 .

    이제 그림의 상세 정보가 분명하게 보입니다 .

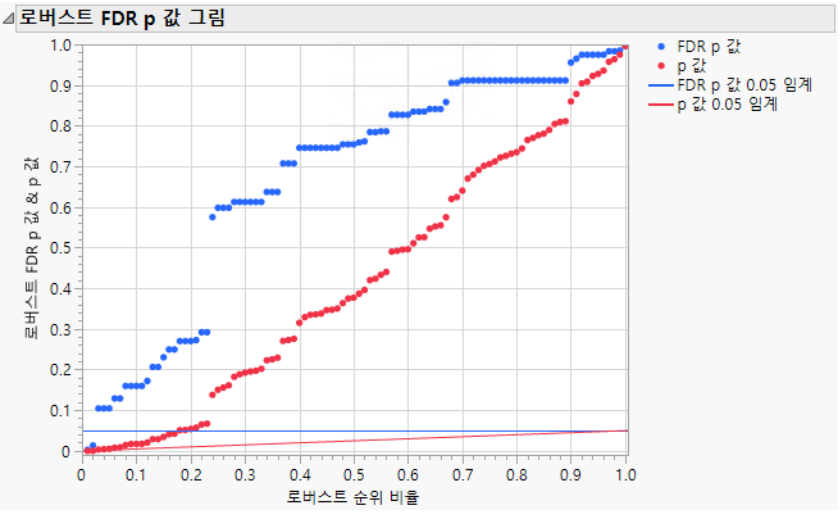
그림 22.13 로버스트 FDR LogWorth 대 효과 크기 (최대 LogWorth = 1000)



로버스트 적합의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drosophila Aging.jmp 를 엽니다 .
2. **분석 > 선별 > 반응 변수 선별**을 선택합니다 .
3. 연속형 열을 모두 선택하고 **Y, 반응**을 클릭합니다 .
4. line 을 선택하고 **X** 를 클릭합니다 .
5. **로버스트**를 선택합니다 .
6. **확인**을 클릭합니다 .

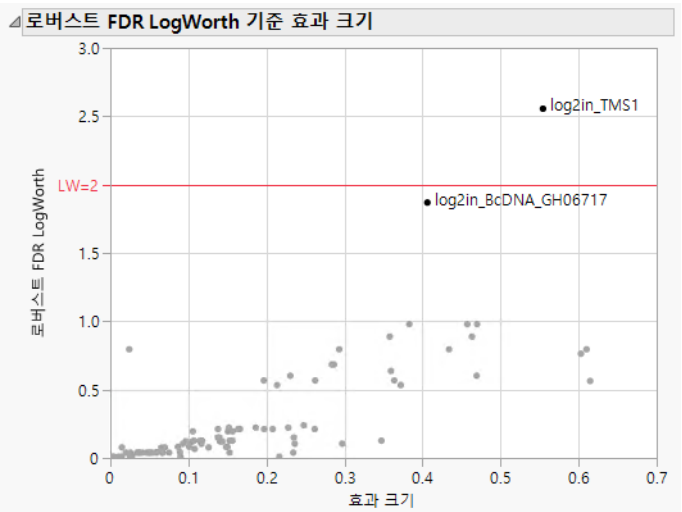
그림 22.14 Drosophila 데이터에 대한 로버스트 FDR p 값 그림



0.05 미만의 빨간색 점으로 표시된 것처럼 조정되지 않은 로버스트 *p* 값을 사용하여 많은 검정이 유의합니다 . 그러나 로버스트 FDR *p* 값에 따르면 두 가지 검정만 유의합니다 . 이 두 점은 FDR LogWorth 를 보여 주는 그림에서 더 쉽게 식별됩니다 .

7. " 로버스트 FDR LogWorth 대 효과 크기 " 표시 아이콘을 클릭합니다 .
8. " 로버스트 FDR LogWorth" 값이 1.5 를 초과하는 점 주위를 직사각형 형태로 드래그합니다 .
9. 그래프에서 마우스 오른쪽 버튼을 클릭하고 **행 라벨**을 선택합니다 .

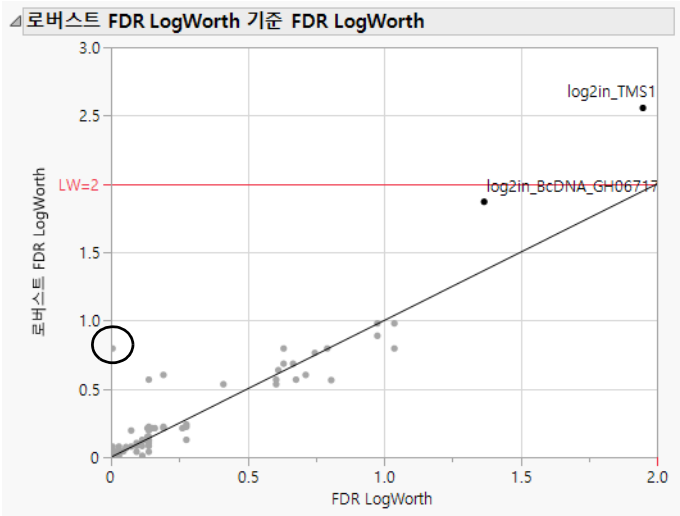
그림 22.15 Drosophila 데이터에 대한 로버스트 LogWorth 대 효과 크기



빨간색 선 (2) 위의 점은 유의 수준이 0.01($-\log_{10}(0.01) = 2$) 미만입니다. 가로선 (약 1.3) 은 0.05 유의 수준에 해당합니다.

10. " 로버스트 FDR LogWorth 대 FDR LogWorth" 표시 아이콘을 클릭합니다.

그림 22.16 Drosophila 데이터에 대한 로버스트 FDR LogWorth 대 FDR LogWorth



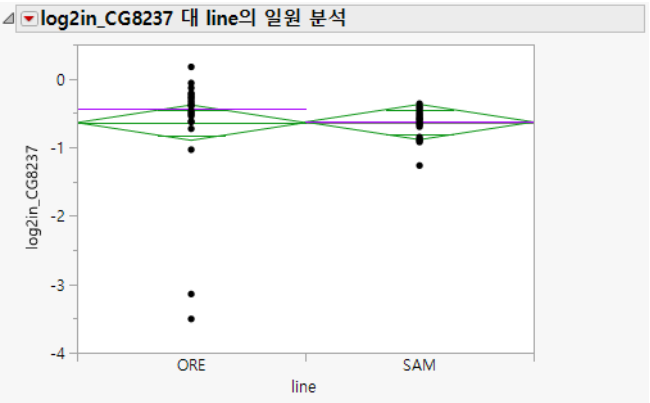
반응에 대한 로버스트 검정이 일반 검정과 동일할 경우 해당 점은 그림 22.16 에서와 같이 대각선 위에 분포합니다. 그림에서 원 안의 점은 로버스트 LogWorth 값이 LogWorth 값을 초과하므로 선에서 떨어져 있습니다.

11. 그림에서 이 점 주위를 직사각형 형태로 드래그합니다.
12. " 결과 테이블 " 에서 이 점에 해당하는 행을 찾습니다.

    반응 log2in_CG8237 의 p 값은 0.9568 이고 로버스트 p 값은 0.0176 입니다.
13. " 반응 변수 선별 " 의 빨간색 삼각형을 클릭하고 **선택한 항목 적합**을 선택합니다.

    반응 log2in_CG8237 에 대한 일원 분석이 포함된 " 선택한 항목 적합 " 보고서가 표시됩니다. 그림에 ORE line 에 대한 두 개의 이상치가 표시됩니다 (그림 22.17). 이러한 이상치는 로버스트 검정과 일반 검정의 결과가 서로 다른 이유를 나타냅니다 . 이상치는 비로버스트 검정의 오차 분산을 부풀리므로 유의한 효과를 알아보기가 더 어렵습니다. 반면 , 로버스트 적합은 이러한 이상치의 비중을 낮추어 오차 분산에 대한 기여도를 줄입니다.

그림 22.17 log2in_CG8237 에 대한 일원 분석



반응 변수 선별 분석법

모형 적합의 반응 변수 선별 분석법을 사용하면 선형 모형 효과에 대한 다중 반응 검정을 연구할 수 있습니다 .

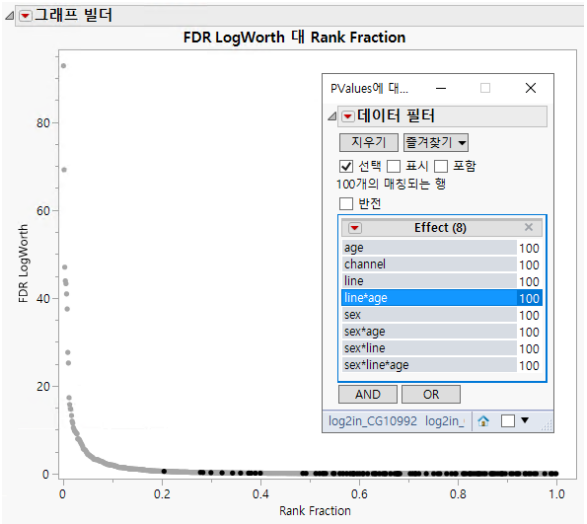
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drosophila Aging.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. 연속형 열을 모두 선택하고 **Y** 를 클릭합니다 .
4. channel 을 선택하고 **추가**를 클릭합니다 .
5. sex, line, age 를 선택하고 **매크로 > 완전 요인**을 선택합니다 .
6. " 분석법 " 목록에서 **반응 변수 선별**을 선택합니다 .
7. **실행**을 클릭합니다 .

   " 반응 변수 선별 적합 " 보고서가 나타납니다 . 두 개의 데이터 테이블도 제공됩니다 . "Y Fits" 테이블에는 전체 모형 검정이 요약되어 있고 "PValues" 테이블에는 각 Y에 대한 모형의 개별 효과 검정이 포함되어 있습니다 .

   어떤 효과가 중요한지 일반적인 개념을 얻으려면 다음을 수행합니다 .
8. "PValues" 데이터 테이블에서 FDR LogWorth by Rank Fraction 스크립트를 실행합니다 .
9. **행 > 데이터 필터**를 선택합니다 .
10. " 데이터 필터 " 창에서 Effect 를 선택하고 **추가**를 클릭합니다 .
11. " 데이터 필터 " 의 모형 효과 목록을 클릭하면서 "FDR LogWorth 대 순위 비율 " 그림에서 선택한 점을 관찰합니다 .

    2 를 초과하는 LogWorth 값은 0.01 수준에서 유의합니다 . 데이터 필터를 사용하면 sex 와 channel 을 제외하고 모형 효과가 0.01 수준에서 거의 유의하지 않음을 알 수 있습니다 . 그림 22.18 에서는 2 위치의 참조선을 보여 줍니다 . line*age 교호작용 효과 검정을 위한 점이 선택되어 있습니다 . 이러한 점은 0.01 수준에서 모두 유의하지 않습니다 .

그림 22.18 line*age 검정이 선택된 FDR LogWorth 대 Rank Fraction 그림



반응 변수 선별 플랫폼에 대한 통계 상세 정보

False Discovery Rate

모든 반응 변수 선별 그림에는 Benjamini and Hochberg(1995)에 설명된 FDR 기법을 사용하여 수행되는 검정의 p 값이 포함됩니다. Westfall et al. (2011) 연구 자료도 확인하십시오. 이 방법에서는 p 값이 독립적이고 균등 분포를 따른다고 가정합니다.

JMP 는 다음 절차를 사용하여 α 수준에서 False Discovery Rate 를 제어합니다 .

1. 관심 있는 가설 검정 m 개를 수행하여 p 값 ($p_1, p_2, \ldots, p_m$) 을 구합니다 .
2. p 값의 순위를 가장 작은 것부터 순서대로 정렬합니다 . 이러한 순위화를 $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$ 으로 나타냅니다 .
3. $p_{(i)} \leq (i/m)\alpha$ 에 대해 가장 큰 p 값을 찾습니다 . 이 첫 번째 p 값을 k 번째로 큰 $p_{(k)}$ 라고 가정합니다 .
4. $p_{(k)}$ 보다 작거나 같은 p 값과 관련된 k 개의 가설을 기각합니다 .

이 절차를 사용하면 False Discovery Rate가 α를 초과하지 않습니다.

False Discovery Rate 에 대해 조정된 p 값은 $p_{(i),FDR}$ 로 표시하고 다음과 같이 계산됩니다 .

$$p_{(i),FDR} = \begin{cases} p_{(m)} & \text{for } i = m \\ min\left[p_{(i+1),FDR}, \left(\frac{m}{i}\right)p_{(i)}\right] & \text{for } i = m-1, \ldots, 1 \end{cases}$$

가설의 FDR 조정 p 값이 α 미만이면 절차에 의해 가설이 기각됩니다.

23 장

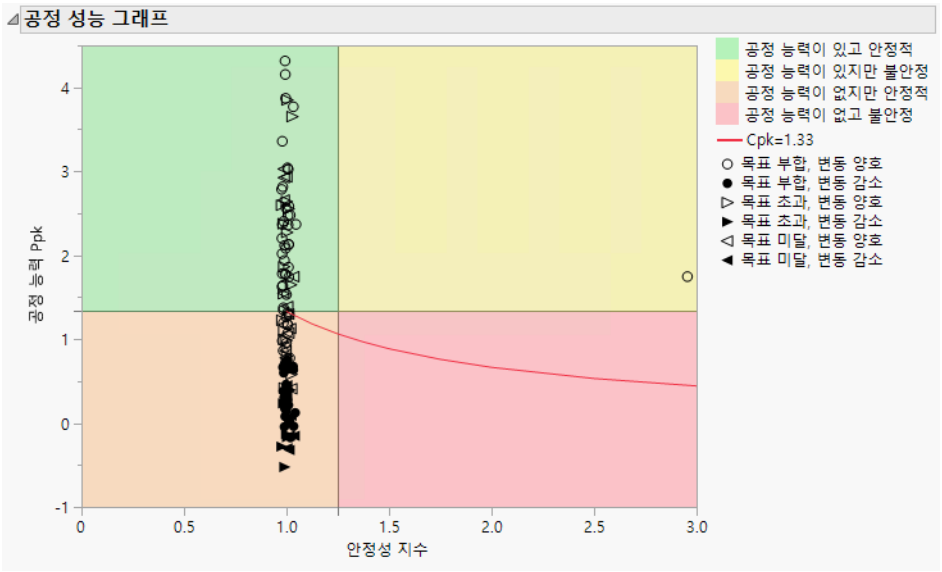
공정 변수 선별

안정성 및 공정 능력을 위한 여러 공정 변수 선별

장기간에 걸쳐 다수의 공정을 탐색하려면 공정 변수 선별 플랫폼을 사용합니다. 이 플랫폼에서는 공정 안정성 및 공정 능력 측정 기준을 계산합니다. 여기에서는 관리도를 생성하고 큰 공정 변화를 감지합니다. 이 플랫폼은 불안정하거나, 규격 한계를 충족할 수 없거나, 평균 변화에 영향을 받는 공정에 빠르게 집중할 수 있도록 지원하여 매우 많은 수의 공정을 신속히 평가하는 것이 목적입니다.

초기 결과에 따라 특정 공정을 그래픽으로 탐색하거나 심층 분석을 수행하도록 선택할 수 있습니다. 관리도 빌더 및 공정 능력 플랫폼에 쉽게 액세스할 수 있습니다. 모든 공정 또는 특정 공정에 대한 상세 결과를 저장할 수 있습니다.

그림 23.1 공정 성능 그래프의 예



목차

공정 변수 선별 플랫폼 개요 443
공정 변수 선별의 예 443
공정 변수 선별 플랫폼 시작 446
시작 창 역할 446
시작 창 옵션 447
한계 테이블 448
공정 변수 선별 보고서 450
공정 변수 선별 플랫폼 옵션 455
공정 성능 그래프 460
공정 변수 선별의 추가 예 461
공정 변수 선별 플랫폼에 대한 통계 상세 정보 464
중앙값을 사용하여 시그마를 추정하기 위한 척도 인수 464

공정 변수 선별 플랫폼 개요

공정 변수 선별 플랫폼을 사용하면 다수의 공정에서 데이터의 안정성 및 공정 능력 평가 작업을 쉽게 할 수 있습니다 . 결과는 주로 관리도 계산을 기반으로 하여 공정이 관리이탈 상태인 경우를 확인합니다 . 공정 변수 선별 플랫폼을 사용하면 다음을 수행할 수 있습니다 .

- 일정한 부분군 크기를 지정하거나, 부분군을 사용한 관리도 계산을 위해 부분군 식별자가 포함된 변수를 사용합니다 .
- 그룹화 변수를 사용합니다 . 그룹 변수의 각 값 조합에 대해 각 공정 변수 분석이 제공됩니다 .
- 중앙값을 사용하여 이상치에 대해 로버스트한 중심선 및 시그마 계산을 수행합니다 .
- 공정 평균에서 큰 변화가 있는 위치에 대한 정보를 얻습니다 .

공정 평균 및 산포 변화에 대한 검정을 포함하여 특정 관리도 검정을 표시하도록 요약 보고서를 사용자 정의할 수 있습니다 . 규격 한계를 지정하면 보고서에 공정 능력 정보도 제공됩니다 . 공정 성능 그래프에서는 안정성과 공정 능력에 따른 공정 성능을 시각적으로 보여 줍니다 . 변화 그래프에는 상향 변화 및 하향 변화의 위치가 표시됩니다 .

팁 : 규격 한계를 추가하는 방법에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오 .

공정 변수 선별을 사용하면 추가 분석을 위한 특정 공정을 쉽게 선택할 수 있습니다 . 이 플랫폼에서는 이러한 공정에 대한 작은 런 차트를 제공하며 , 그림 크기 때문에 상당수의 차트를 한 번에 쉽게 볼 수 있습니다 . 선택한 공정의 분석을 위해 관리도 빌더 및 공정 능력 플랫폼에 연결할 수도 있습니다 .

전체 공정 집합 또는 선택한 공정만 대상으로 하는 다양한 형식의 결과가 포함된 데이터 테이블을 저장할 수 있습니다 .

공정 변수 선별의 예

Semiconductor Capability.jmp 샘플 데이터 테이블에는 128 개의 공정 측정값 열이 포함되어 있습니다 . 각 열에는 1,455 개의 측정값이 있습니다 . 여기서 불안정한 공정을 식별하려고 합니다 . 또한 각 열에는 " 규격 한계 " 열 특성이 포함되어 있습니다 . 공정이 안정적이면 공정 능력을 계산하는 것이 적절합니다 . 이 데이터 테이블에 대해 안정성과 공정 능력을 모두 평가합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Semiconductor Capability.jmp 를 엽니다 .
2. **분석 > 선별 > 공정 변수 선별**을 선택합니다 .
3. Processes 열 그룹을 선택하고 **공정 변수**를 클릭합니다 .

   " 관리도 유형 " 이 "I-MR" 로 설정되어 있습니다 .
4. **확인**을 클릭합니다 .

그림 23.2 초기 보고서 (일부분)

공정 변수 선별

I-MR

		변동		요약		관리도 경보					공정 능력		
열	안정성 지수	군내 표준편차	전체 표준편차	평균	개수	경보 비율	검정1	최근 경보	Ppk	Cpk	규격 이탈 개수	규격 이탈 비율	최근 규격 이탈
PMS1	2.96	6.4e-15	1.9e-14	-50	1455	0.03643	53	3	1.746	5.159	0	0	.
IVP8	1.05	0.45226	0.47293	-21.69	1455	0.00550	8	49	2.369	2.477	0	0	.
B1	1.04	5.70202	5.94318	125.972	1455	0.00275	4	43	0.124	0.130	604	0.4151	1
IVP7	1.04	5.92093	6.14904	-79.865	1455	0.00275	4	114	-0.151	-0.157	1032	0.7093	1
SIT2	1.03	3.59488	3.71963	41.1722	1455	0.00550	8	21	1.745	1.806	0	0	.
CAP	1.03	0.92619	0.95805	38.1075	1455	0.00412	6	12	-0.037	-0.039	1149	0.7897	1
VPM7	1.03	3.20356	3.3107	85.321	1455	0.00550	8	107	0.634	0.655	54	0.0371	13
NPN10	1.03	1.76238	1.82043	73.5168	1455	0.00481	7	38	3.771	3.895	0	0	.
INV2	1.03	0.6307	0.65112	17.3705	1455	0.00481	7	11	0.676	0.697	33	0.0227	25
NPN9	1.03	2.40507	2.47045	47.8068	1455	0.00344	5	9	3.653	3.752	0	0	.
IVP9	1.03	1.61702	1.65762	-30.721	1455	0.00344	5	137	0.410	0.420	145	0.0997	1
VTN210	1.02	2.30557	2.35977	0.09116	1455	0.00344	5	9	0.582	0.595	55	0.0378	6
SIT1	1.02	15.3977	15.7506	149.659	1455	0.00481	7	16	0.088	0.090	581	0.3993	2
INM1	1.02	3.28224	3.34868	82.4373	1455	0.00412	6	3	1.649	1.682	0	0	.
IVP3	1.02	3.08312	3.14263	-49.493	1455	0.00206	3	3	0.395	0.403	168	0.1155	4
VTP210	1.02	3.31312	3.3769	1.29622	1455	0.00344	5	25	1.128	1.150	1	0.0007	1052
PBL1	1.02	0.26689	0.27188	2.71456	1455	0.00481	7	65	0.669	0.681	31	0.0213	53
NPN6	1.02	1.12479	1.14543	43.2968	1455	0.00206	3	69	-0.173	-0.176	1111	0.7636	2
E2A1	1.02	0.00132	0.00134	0.62966	1455	0.00344	5	21	2.478	2.522	0	0	.
VDP1	1.02	0.23752	0.24127	28.8111	1455	0.00137	2	17	0.775	0.787	15	0.0103	27
M1_M1	1.02	0.20527	0.20846	0.23703	1455	0.00412	6	26	-0.149	-0.151	1007	0.6921	1
A2N	1.01	8.53741	8.66111	56.0799	1455	0.00344	5	25	0.019	0.019	925	0.6357	1

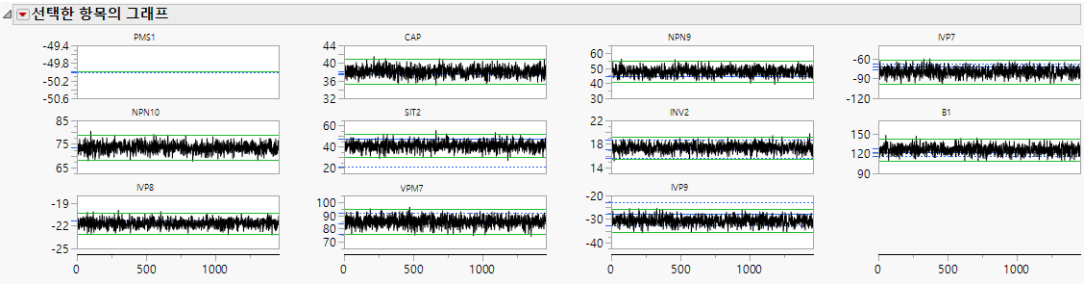
각 공정에 대한 결과 테이블을 보여 주는 " 공정 변수 선별 " 창이 나타납니다 . 테이블은 " 안정성 지수 " 를 기준으로 정렬됩니다 . 이것은 공정 안정성을 나타내는 측도이며 , 안정적인 공정의 안정성 지수는 1 에 가깝습니다 . 안정성 지수의 값이 높을수록 공정이 덜 안정적임을 나타냅니다 . 보고서의 " 안정성 지수 " 옆에 캐럿이 표시되어 정렬을 나타냅니다 . 안정성 지수 값이 1.03 이상인 공정을 자세히 살펴보려고 합니다 .

5. 보고서 창에서 PMS1 ~ IVP9 공정을 선택합니다 .

   이러한 처음 11 개 공정은 각각 " 안정성 지수 " 열의 값이 1.03 이상입니다 .

6. 선택한 공정을 마우스 오른쪽 버튼으로 클릭하고 **선택한 항목에 대한 빠른 그래프**를 선택합니다 .

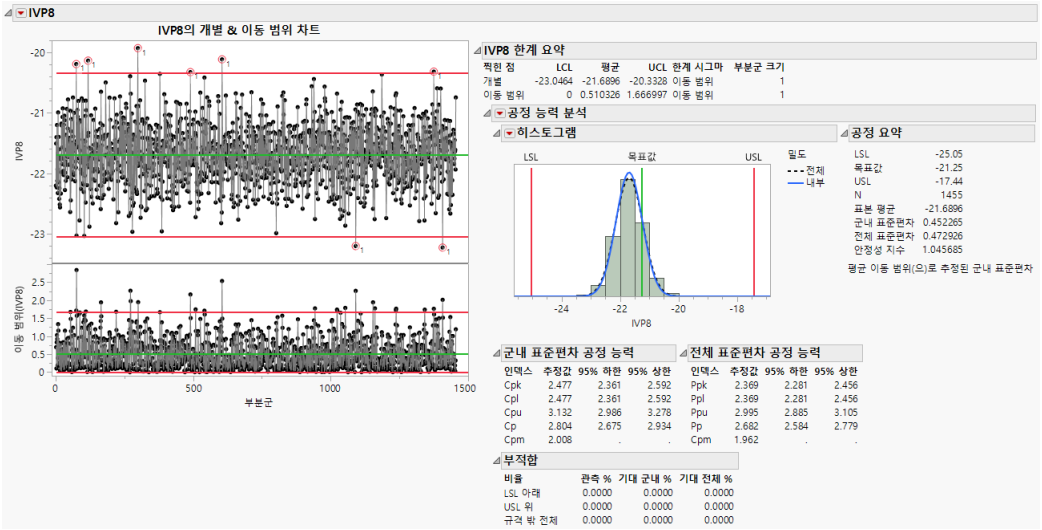
그림 23.3 경보 비율이 가장 높은 공정에 대한 빠른 그래프



IVP8(" 선택한 항목의 그래프 " 에서 3 행 , 1 열) 을 자세히 살펴보기로 했습니다 .

7. 요약 테이블에서 IVP8 에 해당하는 두 번째 공정을 선택합니다 .
8. 선택한 공정을 마우스 오른쪽 버튼으로 클릭하고 **선택한 항목에 대한 관리도**를 선택합니다 .

그림 23.4 IVP8 에 대한 관리도 빌더 보고서

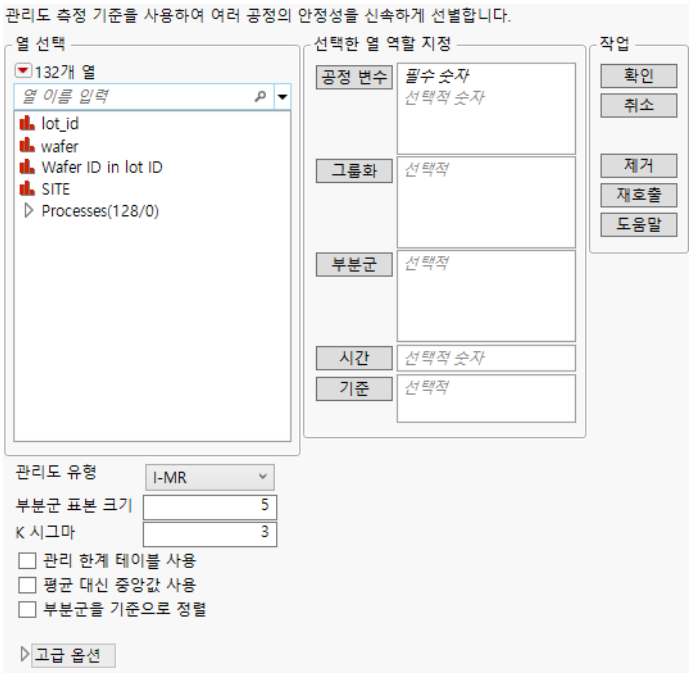


" 관리도 빌더 " 보고서가 나타납니다 . IVP8 에 " 규격 한계 " 열 특성이 있으므로 보고서에 공정 능력 분석도 포함됩니다 .

공정 변수 선별 플랫폼 시작

분석 > 선별 > 공정 변수 선별을 선택하여 공정 변수 선별 플랫폼을 시작합니다.

그림 23.5 공정 변수 선별 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

시작 창 역할

공정 변수 분석할 측정값이 포함된 공정 데이터 열입니다. 열의 데이터 유형은 숫자여야 합니다. 공정 변수에 " 관리 한계 " 열 특성이 있으면 공정 변수 선별 플랫폼에서 해당 한계를 사용하여 지정된 시그마를 계산합니다.

> **참고 :** 이 플랫폼에서는 " 관리 한계 " 열 특성에 지정된 관리 한계를 산포 (R) 차트에 지원하지 않습니다.

그룹화 그룹화 변수로 할당된 열입니다. 그룹화 변수의 각 수준 조합에서 각 공정 변수가 분석됩니다. 결과는 단일 보고서에 제공됩니다.

부분군 하나 이상의 부분군 변수를 할당합니다. 부분군 변수가 두 개 이상 할당된 경우 부분군 수준은 여러 열의 조합으로 정의됩니다.

참고 : I-MR 차트에서는 부분군 역할이 무시됩니다 .

시간 데이터의 시간 순서에 사용되는 값이 포함된 숫자 열입니다 . 타임스탬프가 있는 데이터에 대해 " 시간 " 역할을 사용합니다 . 타임스탬프는 빠른 그래프 및 변화 그래프의 시간 축에 사용됩니다 . 공정 데이터는 계산하기 전에 시간 변수를 기준으로 정렬됩니다 .

기준 각 수준으로 개별 분석을 정의하는 열입니다 . 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다 . 결과는 별도의 테이블과 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

시작 창 옵션

관리도 유형 "I-MR"(개별 측정값 및 이동 범위), "Xbar-R", "Xbar-S", "Xbar MR 및 R", "Xbar MR 및 S" 의 다섯 가지 관리도 유형 중 하나를 선택합니다 . "Xbar MR 및 R" 관리도와 "Xbar MR 및 S" 관리도는 삼원 관리도라고도 합니다 . 통계 상세 정보에 대한 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

참고 : 공정의 부분군 크기가 1 이면 차트가 자동으로 "I-MR" 차트로 전환됩니다 .

부분군 표본 크기 부분군의 상수 표본 크기를 지정합니다 . 최소 부분군 크기는 2 입니다 . 기본 부분군 크기는 5 입니다 . 부분군 변수가 지정된 경우 또는 "I-MR" 차트에 대해서는 " 부분군 표본 크기 " 규격이 무시됩니다 .

K 시그마 시그마 승수를 지정합니다 . "K 시그마 " 는 관리 한계 계산에서 시그마에 곱하는 값입니다 . 기본적으로 "K 시그마 " 는 3 입니다 .

관리 한계 테이블 사용 데이터 테이블에서 과거 관리 한계 및 규격 한계를 가져올 수 있습니다 . 시작 창에서 이 옵션을 선택하고 " 확인 " 을 클릭하면 " 한계 테이블 선택 " 창이 나타납니다 . 한계 테이블을 선택하고 " 확인 " 을 클릭하면 " 한계 규격 " 창이 나타납니다 . 한계 테이블의 열을 적절한 역할에 할당하고 " 확인 " 을 클릭합니다 . 자세한 내용은 " 한계 테이블 " 에서 확인하십시오 .

참고 : 이 옵션을 선택하지 않으면 공정 변수에 대한 " 관리 한계 " 및 " 규격 한계 " 열 특성에서 한계를 가져옵니다 . 이 옵션을 선택하지 않고 공정 변수에 관리 한계 열 특성이 없으면 데이터에서 관리 한계가 계산됩니다 .

평균 대신 중앙값 사용 관측값의 중앙값을 사용하여 중심선을 추정합니다 . 시그마는 몬테카를로 시뮬레이션을 통해 구한 척도 인수를 사용하여 추정됩니다 . 인수 테이블은 " 공정 변수 선별 플랫폼에 대한 통계 상세 정보"에 나와 있습니다 . 계산은 선택한 차트 유형에 따라 다릅니다 .

- "Xbar-R" 차트 또는 "I-MR" 차트 계산의 경우 범위의 척도화된 중앙값을 사용하여 시그마를 추정합니다 .
- "Xbar-S" 차트 계산의 경우 표준편차의 척도화된 중앙값을 사용하여 시그마를 추정합니다 .

– 부분군 크기가 동일하지 않은 경우 척도 인수는 가장 가까운 정수로 반올림된 평균 부분군 크기에 해당합니다 .

하나 이상의 이상치가 중심선 위치에 영향을 주는 경우 많은 부분군이 관리이탈 상태일 수 있습니다 . 중앙값을 사용하면 이 문제가 완화됩니다 .

참고 : " 평균 대신 중앙값 사용 " 을 선택하면 " 선택한 항목에 대한 관리도 " 또는 " 선택한 항목에 대한 공정 능력 분석 " 의 빨간색 삼각형 메뉴 옵션에서 얻은 결과가 " 공정 변수 선별 " 결과와 일치하지 않습니다 .

부분군을 기준으로 정렬 계산하기 전에 부분군 변수 또는 내포된 부분군 변수의 조합을 기준으로 공정 데이터를 정렬합니다 .

고급 옵션 다음 옵션이 포함되어 있습니다 .

변화 임계 변화 그래프의 민감도를 제어하는 값을 지정합니다 . " 변화 임계 " 는 기본적으로 3 으로 설정됩니다 . 이상치 제거 후 변화 그래프에는 " 변화 임계 " 에 지정된 군내 표준편차 단위 수를 초과하는 모든 공정 변화의 시간 발생 그림이 표시됩니다 . 자세한 내용은 " 변화 그래프 " 에서 확인하십시오 .

이상치 임계값 최근의 큰 변화 감지 및 변화 그래프에 대한 이상치 제거의 민감도를 제어하는 값을 지정합니다 . " 이상치 임계값 " 은 기본적으로 5 로 설정됩니다 . 특정 관측값에서 인접한 두 관측값까지의 군내 표준편차 단위 수가 지정된 이상치 임계값을 초과할 경우 해당 관측값은 가장 가까운 인접 관측값에서 1 * 군내 표준편차 단위만큼 떨어진 값으로 대체됩니다 . 자세한 내용은 " 변화 크기 및 위치 " 에서 확인하십시오 .

변화 람다 변화 그래프에 사용되는 EWMA(지수 가중 이동 평균) 가중치를 변경할 수 있습니다 . 자세한 내용은 " 변화 크기 및 위치 " 에서 확인하십시오 .

흐름 베타 흐름 감지를 위해 EWMA(지수 가중 이동 평균) 에 사용되는 가중치입니다 . " 흐름 베타 " 는 기본적으로 0.05 로 설정됩니다 .

최소 공정 길이 공정이 분석에 포함되기 위해 필요한 최소 데이터 값 수입니다 . 기본적으로 이 값은 3 으로 설정됩니다 .

한계 테이블

한계 테이블에는 공정 데이터 테이블에서 공정 변수 및 그룹화 변수에 의해 정의된 각 공정에 대한 행이 포함됩니다. 한계 테이블을 사용하는 경우 "한계 규격" 창에서 아래 나열된 역할로 변수를 지정할 수 있습니다. 모든 역할에 대해 변수를 지정할 필요는 없습니다. 이러한 역할은 모두 선택적입니다.

그림 23.6 한계 규격 창

한계 테이블에서 역할 버튼과 일치하는 이름을 가진 열은 자동으로 입력됩니다. 예를 들어 "공정", "열" 또는 "모수" 등의 열은 "공정 변수" 목록에 자동으로 입력됩니다.

관리 한계가 있지만 " 중심 " 또는 " 시그마 " 에 대한 열이 없는 경우 " 유도된 시그마 " 옵션을 사용할 수 있습니다 .

공정 변수 공정 데이터 테이블의 열 이름에 해당하는 값이 포함된 열입니다 .

그룹화 공정 데이터 테이블에 대한 그룹화 변수 값이 포함된 하나 이상의 열입니다 .

중심 각 공정에 대한 중심선 값이 포함된 열입니다 . 일반적으로 과거 공정 평균입니다 .

시그마 각 공정에 대한 군내 표준편차 값이 포함된 열입니다 . 일반적으로 과거 표준편차입니다 . 이 값은 " 공정 변수 선별 " 보고서에서 " 지정된 시그마 " 로 라벨이 지정됩니다 .

유도된 시그마 주어진 관리 한계와 부분군 크기를 기반으로 시그마를 계산합니다 . 이 값은 " 공정 변수 선별 " 보고서에서 " 지정된 시그마 " 로 라벨이 지정됩니다 .

참고 : 지정된 시그마가 있으면 공정 변수 선별 플랫폼과 관리도 빌더 플랫폼 간에 군내 표준편차 , 안정성 지수 및 안정성 비율 값이 일치하지 않습니다 . 공정 변수 선별 플랫폼의 경우 지정된 시그마를 사용하여 관리 한계를 계산하지만 군내 표준편차는 데이터를 사용하여 계산합니다 . 관리도 빌더 플랫폼의 경우 지정된 시그마를 과거 시그마로 사용하며 관리도에서 과거 시그마가 군내 표준편차로 사용됩니다 .

유도된 시그마는 다음과 같이 계산됩니다 .

$$\text{시그마} = d^*(UCL - LCL)/6$$

여기서 d 는 부분군 크기의 제곱근입니다 .

LCL 각 공정에 대한 관리 하한이 포함된 열입니다 .

UCL 각 공정에 대한 관리 상한이 포함된 열입니다 .

부분군 크기 각 공정에 대한 부분군 크기가 포함된 열입니다 .

LSL 각 공정에 대한 규격 하한이 포함된 열입니다 .

USL 각 공정에 대한 규격 상한이 포함된 열입니다 .

목표값 각 공정에 대한 목표값이 포함된 열입니다 .

중요도 각 공정에 대한 중요도 값이 포함된 열입니다 . " 중요도 " 열은 원하는 순서로 공정을 정렬하는 메커니즘을 제공합니다 .

공정 변수 선별 보고서

"공정 변수 선별" 보고서는 공정 안정성 결과가 포함된 요약 테이블과 함께 열립니다. 요약 테이블 위에는 결과의 기반이 되는 관리도 유형을 보여 주는 라벨이 있습니다. I-MR을 제외한 모든 차트 유형의 경우 표본 크기나 부분군 변수도 요약 테이블 위에 표시됩니다. 규격 한계를 제공한 경우에는 공정 능력 결과도 요약 테이블에 포함됩니다. 공정과 그룹은 처음에는 안정성 지수에 따라 내림차순으로 정렬됩니다. 사용자가 중요도 값을 지정한 경우 공정과 그룹은 중요도 내의 안정성 지수에 따라 내림차순으로 정렬됩니다. "안정성 지수", "Ppk", "Cpk", "Cp" 및 "목표 지수" 열은 녹색, 노란색, 빨간색으로 표시되어 안정성 또는 공정 능력이 각각 적합, 경계, 나쁨 상태임을 나타냅니다. 이 색상 코딩 체계는 공정 성능 그래프 색상 코딩 체계와 일치합니다.

팁 : 보고서를 열별로 정렬하려면 열 이름을 클릭합니다 . 해당 열 이름 오른쪽에 캐럿이 나타납니다 . 캐럿 방향은 내림차순 정렬인지 아니면 오름차순 정렬인지 나타냅니다 . 정렬 순서를 변경하려면 열 이름을 다시 클릭합니다 .

요약 테이블의 관리도 계산에는 Nelson 검정과 범위 한계 초과 검정이 포함됩니다 . 이러한 검정에서는 관리도 한계에 대해 다음과 같이 가정합니다 .

- XBar 또는 X 관리도의 중심선은 모든 측정값의 평균에 의해 지정됩니다 . " 평균 대신 중앙값 사용 " 옵션을 선택하면 관측값의 중앙값에 의해 중심선이 지정됩니다 .

- 관리 한계는 중심선으로부터 *K* 시그마 단위만큼 떨어진 위치에 배치됩니다. 시작 창의 "K 시그마 " 옵션을 사용하여 *K* 를 지정합니다 . K 시그마 기본값은 3 입니다 .
- 시그마는 사용자가 지정한 관리도 유형에 해당하는 규칙을 사용하여 추정되거나 , 평균 대신 중앙값을 사용하는 경우 " 평균 대신 중앙값 사용 " 의 설명대로 추정됩니다 .

팁 : 공정 변수 선별 플랫폼의 8 가지 Nelson 검정은 관리도 빌더 플랫폼 환경 설정의 검정 설정을 따릅니다 . 파일 > 환경 설정 > 플랫폼 > 관리도 빌더에서 검정을 사용자 정의할 수 있습니다 .

요약 테이블에는 다음과 같은 정보가 포함될 수 있습니다 .

열 공정 역할로 입력한 열입니다 . 공정 및 그룹화 열의 고유한 각 조합에 대한 행이 있습니다 . 공정 열이 하나뿐이면 이 열이 표시되지 않습니다 .

팁 : 선택한 항목에서 작동하는 옵션에 액세스하려면 **열** 열을 마우스 오른쪽 버튼으로 클릭합니다 .

그룹화 열 데이터 테이블에서 그룹화 역할로 입력한 각 열에 대한 보고서 열이 있습니다 . 그룹화 열의 수준이 나열되므로 공정 이름과 그룹화 열 값의 각 조합에 대한 고유 행이 보고서 테이블에 있습니다 .

중요도 (사용자가 한계 테이블의 입력으로 지정한 중요도 값이 있는 경우에만 표시됨) 공정에 대해 사용자가 지정한 중요도 값입니다 .

변동 다음 열이 포함됩니다 .

안정성 지수 공정의 안정성 측도입니다 . 안정적인 공정의 안정성 지수는 1 에 가깝습니다 . 값이 클수록 안정성이 낮습니다 . 안정성 지수는 다음과 같이 정의됩니다 .

(전체 표준편차 / 군내 표준편차)

시작할 때 삼원 관리도를 선택한 경우 안정성 지수는 다음과 같이 정의됩니다 .

(전체 표준편차 / 군간 - 군내 표준편차)

안정성 비율 공정의 안정성 측도입니다 . 안정적인 공정의 안정성 비율은 1 에 가깝습니다 . 값이 클수록 안정성이 낮습니다 . 안정성 비율은 다음과 같이 정의됩니다 .

$$(\text{전체 표준편차} / \text{군내 표준편차})^2$$

시작할 때 삼원 관리도를 선택한 경우 안정성 비율은 다음과 같이 정의됩니다 .

$$(\text{전체 표준편차} / \text{군간 - 군내 표준편차})^2$$

군내 표준편차 부분군 내 변동에 기반한 표준편차 추정값입니다 . 추정값은 사용자가 지정한 관리도 유형을 기반으로 하며 단기 변동 측도입니다 . 통계 상세 정보는 **품질 및 공정 방법** 의 에서 확인하십시오 . 평균 대신 중앙값을 선택하면 " 평균 대신 중앙값 사용 " 에 설명된 대로 군내 표준편차가 계산됩니다 .

전체 표준편차 모든 관측값에 기반한 일반적인 표준편차 추정값입니다 .

군간 표준편차 (시작 창에서 삼원 관리도를 선택한 경우에만 표시됨) 부분군 간 변동에 기반한 표준편차 추정값입니다. 통계 상세 정보는 **품질 및 공정 방법**의 에서 확인하십시오.

군간 - 군내 표준편차 (시작 창에서 삼원 관리도를 선택한 경우에만 표시됨) 부분군 간 변동과 부분군 내 변동에 기반한 표준편차 추정값입니다. 군간 - 군내 표준편차 추정값은 다음과 같이 정의됩니다.

$$\text{군간-군내 표준편차} = \sqrt{\text{군내 표준편차}^2 + \text{군간 표준편차}^2}$$

지정된 시그마 " 한계 규격 " 대화상자의 시그마 관리 한계로 지정된 표준편차 또는 관리 한계와 부분군 크기에서 도출된 표준편차 추정값입니다. 자세한 내용은 " 유도된 시그마 " 에서 확인하십시오. 관리 한계와 부분군 크기는 한계 테이블 또는 " 관리 한계 " 열 특성을 사용하여 지정할 수 있습니다.

요약 다음 열이 포함됩니다.

중심선 (시작 창에서 " 평균 대신 중앙값 사용 " 을 선택하거나, 한계 테이블을 사용하여 중심 값을 가져오거나, 시그마 유도를 위해 관리 한계를 가져오는 경우에 표시됨) " 중심선 " 아래에 나열된 값은 관리도 계산에서 중심선으로 사용됩니다.

- 시작 창에서 " 평균 대신 중앙값 사용 " 을 선택하는 경우 관측값의 전체 중앙값이 표시됩니다.
- 한계 테이블에서 중심 값을 가져오는 경우 해당 값이 표시됩니다.
- 관리 한계를 가져오는 경우 (UCL + LCL)/2 로 계산된 값이 표시됩니다.

평균 모든 관측값의 평균입니다.

개수 관측값 수입니다.

부분군 수 부분군 수입니다.

관리도 경보 8 개의 각 Western Electric 규칙을 비롯하여 다양한 검정에 대해 경보를 발생시키는 부분군의 정보가 포함되어 있습니다. 표준편차 추정값은 군내 표준편차 값입니다. 기본적으로 " 경보 비율 ", " 검정 1" 및 " 최근 경보 " 열만 요약 테이블에 표시됩니다.

경보 비율 " 검정 선택 " 옵션에서 선택한 검정에 대해 경보를 발생시킨 부분군의 수 (모든 경보) 를 비결측 부분군의 수 (부분군) 로 나눈 값입니다.

모든 경보 (검정 열이 두 개 이상 있는 경우에만 표시됨) " 검정 선택 " 옵션에서 선택한 검정에 대해 경보를 발동하는 부분군의 수입니다. 여기에는 8 개의 Nelson 검정과 " 범위 한계 초과 " 에 해당하는 검정이 포함됩니다.

팁 : 공정 변수 선별 플랫폼의 8 가지 Nelson 검정은 관리도 빌더 플랫폼 환경 설정의 검정 설정을 따릅니다. 파일 > 환경 설정 > 플랫폼 > 관리도 빌더에서 검정을 사용자 정의할 수 있습니다.

검정 1 하나의 점이 중심선으로부터 3 표준편차를 벗어납니다 . 해당 점과 연관된 부분군에서 경보를 발동합니다 .

검정 2 연속된 9 개 이상의 점이 중심선의 같은 쪽에 있습니다 . 9 번째 점과 연관된 부분군에서 경보를 발동합니다 .

검정 3 연속된 6 개 이상의 점이 계속 증가하거나 감소합니다 . 6 번째 점과 연관된 부분군에서 경보를 발동합니다 .

검정 4 연속된 14 개의 점이 증가하다가 감소하거나 감소하다가 증가하는 식으로 방향을 바꿉니다 . 14 번째 점과 연관된 부분군에서 경보를 발동합니다 .

검정 5 중심선의 같은 쪽에 있는 연속된 3 개 점 중 2 개가 중심선으로부터 2 표준편차를 벗어납니다 . 2 표준편차를 벗어나는 두 번째 점과 연관된 부분군에서 경보를 발동합니다 .

검정 6 중심선의 같은 쪽에 있는 연속된 5 개 점 중 4 개가 중심선으로부터 2 표준편차 이상 벗어납니다 . 1 표준편차를 벗어나는 네 번째 점과 연관된 부분군에서 경보를 발동합니다 .

검정 7 중심선의 양쪽에 있는 연속된 15 개의 점이 모두 중심선의 1 표준편차 내에 있습니다 . 15 번째 점과 연관된 부분군에서 경보를 발동합니다 .

검정 8 중심선의 양쪽에 있는 연속된 8 개의 점이 모두 중심선의 1 표준편차를 벗어납니다 . 8 번째 점과 연관된 부분군에서 경보를 발동합니다 .

범위 한계 초과 R, S 또는 MR 차트 계산에서 관리 상한을 초과하는 부분군의 수입니다 .

이동 범위 한계 초과 삼원 관리도 계산에서 이동 범위 한계를 초과하는 부분군의 수입니다 .

최근 경보 마지막 부분군부터 세어 Nelson 또는 범위 한계 초과 검정에 대해 가장 최근에 경보 신호를 보낸 부분군의 위치입니다 .

공정 능력 (일부 공정에 대해 " 규격 한계 " 를 지정한 경우에만 표시됨) 다음 옵션이 포함되어 있습니다 .

Ppk 정규 분포를 가정하고 전체 표준편차에 기반한 공정 능력 지수입니다 . 통계 상세 정보는 **품질 및 공정 방법**의 에서 확인하십시오 . 기본적으로 설정되어 있습니다 .

Cpk 정규 분포를 가정하고 군내 표준편차 또는 군간 - 군내 표준편차에 기반한 공정 능력 지수입니다 . 통계 상세 정보는 **품질 및 공정 방법**의 에서 확인하십시오 . 기본적으로 설정되어 있습니다 .

Cp 목표 및 흐름 문제가 해결된 경우 잠재적 공정 능력입니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

목표 지수 공정 평균이 목표값과 다른 단기 표준편차의 수입니다 . 이 값은 목표값에 도달할 공정 능력을 측정합니다 . 목표 지수는 3(Cp - Cpk) 으로 계산됩니다 . 목표 지수는 1 을 초과하면 나쁨 , 0.5 에서 1 사이이면 경계 , 0.5 미만이면 적합으로 간주됩니다 . 자세한 내용은 White et al. (2018) 연구 자료에서 확인하십시오 .

규격 이탈 개수 규격 한계를 벗어나는 관측값의 수입니다 . 기본적으로 설정되어 있습니다 .

규격 이탈 비율 규격 한계를 벗어나는 관측값의 비율입니다 . 기본적으로 설정되어 있습니다 .

기대 규격 이탈 비율 규격 한계를 벗어나는 관측값의 기대 비율입니다 . 기대 규격 이탈 비율은 안정적이고 정규 분포를 따르는 공정을 가정하고 전체 표준편차를 사용합니다 .

최근 규격 이탈 마지막 관측값부터 가장 최근 관측값까지 세어 규격 한계를 벗어나는 관측값의 수입니다 . 기본적으로 설정되어 있습니다 .

(평균 - 목표)/ 규격 범위 규격으로 중심화된 평균입니다 . 이 값은 " 공정 능력 " 의 " 규격으로 표준화된 평균 변화 " 와 동일합니다 . 통계 상세 정보는 **품질 및 공정 방법**의 에서 확인하십시오 .

표준편차 / 규격 범위 규격으로 척도화된 표준편차입니다 . 이 값은 " 공정 능력 " 의 " 규격으로 표준화된 표준편차 " 와 동일합니다 . 통계 상세 정보는 **품질 및 공정 방법**의 에서 확인하십시오 .

LSL 규격 하한입니다 .

USL 규격 상한입니다 .

목표값 목표값입니다 .

변화 크기 및 위치 (" 공정 변수 선별 " 의 빨간색 삼각형 메뉴에서 " 변화 감지 " 옵션을 선택한 경우에만 표시됨) 변화 감지는 1 * 군내 표준편차 단위를 초과하는 변화를 식별하기 위해 수행됩니다 . 이 알고리즘은 개별 관측값에 대해 이상치 수정 및 EWMA 평활 방식을 사용합니다 . 알고리즘은 다음과 같습니다 .

- 이상치 단독으로 변화를 나타내지 않도록 이상치가 제거됩니다 . 시작 창에서 이상치 임계값 (기본값 5) 으로 지정된 값은 이상치 제거 민감도를 제어합니다 . 특정 관측값에서 인접한 두 관측값까지의 군내 표준편차 단위 수가 지정된 이상치 임계값을 초과할 경우 해당 관측값은 가장 가까운 인접 관측값에서 1 * 군내 표준편차 단위만큼 떨어진 값으로 대체됩니다 .
- 정방향 시간 순서로 부분군 평균에 대한 EWMA 적합이 생성되고, 역방향 시간 순서로 부분군 평균에 대한 다른 EWMA 적합이 생성됩니다 . EWMA 적합의 람다는 0.3 입니다 .
- 연속된 EWMA 값 사이의 최대 양수 및 음수 차이 중 1 * 군내 표준편차 단위를 초과하는 차이가 식별됩니다 .
- 이러한 차이의 절대값을 군내 표준편차 추정값으로 나눈 값이 최대 상향 변화 및 최대 하향 변화 값으로 보고됩니다 .
- 이러한 최대 변화와 관련된 첫 번째 부분군의 위치가 상향 변화 위치 및 하향 변화 위치를 정의합니다 .

최대 상향 변화 1 * 군내 표준편차 단위를 초과하는 최대 상향 변화의 크기입니다 (군내 표준편차 단위로 보고됨)

상향 변화 위치 또는 상향 변화 < 시간 변수 > 최대 상향 변화를 갖는 부분군의 위치입니다 . 시간 변수를 지정하면 요약 테이블의 열 이름이 " 상향 변화 < 시간 변수 >" 로 지정되고 변화 위치가 시간 변수로 제공됩니다 .

최대 하향 변화 1 * 군내 표준편차 단위를 초과하는 최대 하향 변화의 크기입니다 (군내 표준편차 단위로 보고됨)

하향 변화 위치 또는 하향 변화 < 시간 변수 > 최대 하향 변화를 갖는 부분군의 위치입니다 . 시간 변수를 지정하면 요약 테이블의 열 이름이 " 하향 변화 < 시간 변수 >" 로 지정되고 변화 위치가 시간 변수로 제공됩니다 .

흐름 크기 및 위치 (" 공정 변수 선별 " 의 빨간색 삼각형 메뉴에서 " 흐름 요약 " 옵션을 선택한 경우에만 표시됨) 흐름 감지는 공정에서 더 작고 점진적인 변화를 감지하기 위해 수행됩니다 . 이 알고리즘은 흐름 감지가 EWMA 대신 Holt 이중 지수 평활기를 사용한다는 점을 제외하면 변화 감지에 사용되는 알고리즘과 동일합니다 . 알고리즘은 다음과 같습니다 .

- 변화 감지에 사용되는 동일한 과정을 통해 이상치가 제거됩니다 . 자세한 내용은 " 변화 크기 및 위치 " 에서 확인하십시오 .
- 흐름 감지 알고리즘은 정방향 시간 순서로 부분군 평균에 대한 Holt 이중 지수 평활 모형을 적합시키고 , 역방향 시간 순서로 부분군 평균에 대한 다른 Holt 이중 지수 평활 모형을 적합시킵니다 . 두 평활 모형 적합은 각각 α(수준) 와 β(기울기) 라는 두 개의 평활 상수를 가집니다 . β 평활 상수는 0.05 로 설정되고 α 평활 상수는 오차를 최소화하도록 추정됩니다 .
- 양수 흐름 값 , 음수 흐름 값 및 절대 흐름 값의 평균이 각각 " 평균 상향 흐름 ", " 평균 하향 흐름 ", " 평균 절대 흐름 " 으로 보고됩니다 .

평균 상향 흐름 양수 흐름 값의 합을 개수로 나눈 값입니다 .

평균 하향 흐름 음수 흐름 값의 합을 개수로 나눈 값입니다 .

평균 절대 흐름 모든 흐름 값의 절대값 합을 개수로 나눈 값입니다 .

참고 : 공정 변수 선별은 " 분포 " 또는 " 공정 능력 분포 " 열 특성을 사용하지 않으며 항상 정규 분포를 가정하여 공정 능력 분석을 수행합니다 . 이러한 열 특성 중 하나에 대한 분포 유형이 비정규이면 요약 테이블 아래에 " 경고 " 개요가 표시됩니다 .

공정 변수 선별 플랫폼 옵션

" 공정 변수 선별 " 의 빨간색 삼각형 메뉴에는 화면 표시를 사용자 정의하고 계산된 통계량을 저장하는 옵션이 포함되어 있습니다 . 요약 테이블에서 **열** 열을 마우스 오른쪽 버튼으로 클릭하여 선택한 항목에서 작동하는 옵션에 액세스할 수도 있습니다 .

요약 요약 테이블을 표시하거나 숨깁니다 . 자세한 내용은 " 공정 변수 선별 보고서 " 에서 확인하십시오 .

찾기 및 선택 시작 창에서 공정 변수 또는 그룹화로 입력한 열에 대한 검색 문자열을 입력합니다 . 각 열에 대한 패널이 나타납니다 . 해당하는 공정이 요약 테이블에서 선택됩니다 .

선택 조건 " 선택 조건 " 창을 엽니다 . " 비교 " 메뉴와 " 값 " 텍스트 상자를 사용하여 요약 테이블에서 특정 조건에 해당하는 특정 공정을 선택할 수 있습니다 . 예를 들어 "Ppk < 1.33" 에 해당하는 모든 공정을 선택할 수 있습니다 . " 확인 " 을 클릭하면 요약 테이블에서 공정이 선택됩니다 .

팁 : 요약 테이블의 아무 곳이나 마우스 오른쪽 버튼으로 클릭하여 " 선택 조건 " 창에 액세스할 수도 있습니다 .

요약 열 표시 요약 테이블에 표시되는 요약 열을 선택할 수 있습니다 .

선택한 항목에 대한 빠른 그래프 요약 테이블에서 선택한 공정에 대한 작은 그래프를 " 선택한 항목의 그래프 " 보고서에 표시합니다 . 이 보고서를 통해 한 번에 여러 공정을 보고 비교할 수 있습니다 . 그림은 시작 창에 입력한 순서에 따라 정렬됩니다 . 각 빠른 그래프에는 규격 한계에 대한 선도 있습니다 . 목표값을 나타내는 파란색 실선과 LSL 및 USL 값을 나타내는 파란색 점선이 있습니다 . 빠른 그래프를 제거하려면 " 선택한 항목의 그래프 " 의 빨간색 삼각형을 클릭하고 " 제거 " 를 선택합니다 .

팁 : 빠른 그래프의 레이아웃을 변경하려면 " 선택한 항목의 그래프 " 의 빨간색 삼각형을 클릭하고 " 가로 그림 수 " 를 선택합니다 .

선택한 항목에 대한 관리도 요약 테이블에서 선택한 공정에 대한 " 관리도 빌더 " 보고서 창을 엽니다 . 관리도는 공정 변수 선별 시작 창에서 선택한 유형에 해당합니다 .

참고 :

- 공정 변수 선별에서 선택한 검정만 관리도 빌더에 전달됩니다 . 해제된 검정은 전달되지 않습니다 .
- 여러 부분군 변수가 있는 경우 , 내부 부분군 변수가 외부 부분군 변수 내에 정렬되어 있지 않았다면 선택한 항목에 대한 관리도를 시작할 때 새 부분군 변수 열이 생성되어 사용됩니다 . 공정 변수 선별 시작 창에서 " 부분군을 기준으로 정렬 " 을 선택한 경우에는 적용되지 않습니다 .

선택한 항목에 대한 공정 능력 분석 요약 테이블에서 선택한 공정에 대한 " 개별 상세 정보 보고서 " 를 보여 주는 " 공정 능력 " 보고서 창을 엽니다 . 규격 한계가 지정되지 않은 공정을 선택하면 " 규격 한계 " 창이 나타납니다 . 이 창에서 데이터 테이블을 선택하거나 값을 직접 입력하여 규격 한계를 지정할 수 있습니다 .

공정 능력 분석은 정규 분포를 가정하고 , 다음과 같이 공정 변수 선별 시작 창에서 선택한 관리도 유형에 해당하는 군내 표준편차 값을 사용합니다 .

- I-MR 의 경우 이동 범위
- Xbar-R 의 경우 범위 평균
- Xbar-S 의 경우 비편향 표준편차의 평균

참고 :

- " 한계 규격 " 창에서 시그마 관리 한계 또는 " 유도된 시그마 " 옵션을 사용하여 시그마를 지정하는 경우 이 값은 공정 능력에 사용되지 않습니다 . 공정 변수 선별 플랫폼과 공정 능력 플랫폼에서 이러한 관리 한계를 다르게 사용하기 때문입니다 .
- 여러 부분군 변수가 있는 경우 , 내부 부분군 변수가 외부 부분군 변수 내에 정렬되어 있지 않았다면 선택한 항목에 대한 공정 능력 분석을 시작할 때 새 부분군 변수 열이 생성되어

사용됩니다 . 공정 변수 선별 시작 창에서 " 부분군을 기준으로 정렬 " 을 선택한 경우에는 적용되지 않습니다 .

선택한 항목에 색상 적용 선택한 색상을 요약 테이블에서 선택한 행의 값에 적용합니다 .

선택한 항목 제거 요약 테이블에서 선택한 행을 제거하고 해당 공정 없이 분석을 다시 실행합니다 .

검정 표시 " 공정 변수 선별 " 보고서의 요약 테이블에 " 검정 선택 " 옵션에서 선택한 Nelson 검정에 대한 검정 결과를 표시하거나 숨깁니다 .

검정 선택 " 경보 비율 " 및 " 모든 경보 " 계산에 포함할 검정을 선택할 수 있습니다 .

팁 : 여러 검정을 선택하려면 Alt 키를 누른 채 " 공정 변수 선별 " 의 빨간색 삼각형을 클릭하여 모든 플랫폼 옵션이 포함된 메뉴를 엽니다 .

변화 감지 이상치 제거 후 변화를 감지하기 위한 옵션을 제공합니다 . 자세한 내용은 " 변화 크기 및 위치 " 에서 확인하십시오 .

최대 상향 변화 요약 테이블에 " 최대 상향 변화 " 및 " 상향 변화 위치 " 열을 추가합니다 . 계열에서 1 * 군내 표준편차 단위를 초과하는 최대 상향 변화가 식별됩니다 . 자세한 내용은 " 변화 크기 및 위치 " 에서 확인하십시오 .

최대 하향 변화 요약 테이블에 " 최대 하향 변화 " 및 " 하향 변화 위치 " 열을 추가합니다 . 계열에서 1 * 군내 표준편차 단위를 초과하는 최대 하향 변화가 식별됩니다 . 자세한 내용은 " 변화 크기 및 위치 " 에서 확인하십시오 .

변화 그래프 (일부 공정에 변화 임계를 초과하는 변화가 있는 경우에만 사용 가능) " 변화 임계 " 에 지정된 군내 표준편차 단위 수 (기본값 3) 를 초과하는 모든 공정 변화의 시간 발생 그림을 표시합니다 . " 변화 그래프 " 의 슬라이더 막대를 사용하여 변화 임계를 변경할 수도 있습니다 . 다이아몬드 모양을 커서로 가리키면 값이 표시됩니다 . 녹색 표식은 상향 변화를 나타내고 빨간색 표식은 하향 변화를 나타냅니다 . 표식은 변화의 국소 정상점에 있습니다 .

한 번 이상의 변화 발생에 해당하는 공정을 식별하려면 점을 선택하고 " 공정 선택 " 을 클릭합니다 . 해당하는 공정이 요약 테이블에서 선택됩니다 . 변화 임계 (군내 표준편차 단위 수) 를 초과하는 변화가 없는 공정은 표시되지 않습니다 .

참고 : 변화 크기가 지정된 변화 임계 (군내 표준편차 단위 수) 보다 작으면 요약 테이블에 나타나는 최대 상향 변화 및 최대 하향 변화 값의 위치가 " 변화 그래프 " 에 표시되지 않습니다 . 자세한 내용은 " 공정 변수 선별의 추가 예 " 에서 확인하십시오 .

빠른 그래프에 변화 표시 (보고서 창에 빠른 그래프가 추가된 경우에만 사용 가능) 녹색 및 빨간색 세로선을 사용하여 빠른 그래프에 변화 위치를 표시합니다 .

흐름 요약 요약 테이블에 " 평균 상향 흐름 ", " 평균 하향 흐름 " 및 " 평균 절대 흐름 " 열을 추가합니다 . 자세한 내용은 " 흐름 크기 및 위치 " 에서 확인하십시오 .

선택 항목 흐름 그래프 요약 테이블에서 선택한 각 공정에 대한 흐름 그래프를 표시합니다 . 흐름 그래프를 사용하면 선택한 공정에서 더 작고 점진적인 변화를 감지할 수 있습니다 . 표시된 값은 Holt 이중 지수 평활 모형에서 구한 기울기 추정값입니다 . 자세한 내용은 " 흐름 크기 및 위치 " 에서 확인하십시오 . 각 흐름 그래프에는 Holt 이중 지수 평활 모형의 두 평활 모수에 대한 추정값 및 군내 표준편차 값이 표시됩니다 .

공정 성능 그래프 (하나 이상의 공정 변수에 대해 규격 한계가 정의된 경우에만 사용 가능) 안정성과 공정 능력 측면에서 공정 성능을 평가하는 4 개의 사분면 그래프를 표시합니다 . 자세한 내용은 " 공정 성능 그래프 " 에서 확인하십시오 .

공정 성능 그래프 경계 (공정 성능 그래프가 보고서 창에 있는 경우에만 사용 가능) 공정 성능 그래프의 안정성 지수 경계 및 Ppk 공정 능력 경계에 대한 값을 설정할 수 있는 창을 엽니다 .

팁 : 파일 > 환경 설정 > 플랫폼 > 공정 변수 선별에서 안정성 지수 경계 및 Ppk 공정 능력 경계에 대해 원하는 환경 설정을 지정할 수 있습니다 .

목표 그림 (하나 이상의 공정 변수에 대해 규격 한계가 정의된 경우에만 사용 가능) " 목표 그림 " 에는 각 변수에 대해 규격으로 정규화된 평균 변화가 X 축에 표시되고 규격으로 정규화된 표준편차가 Y 축에 표시됩니다 . 공정의 중요도 값을 정의하면 중요도에 따라 목표 그림의 점 크기가 조정됩니다 . 이 그림을 사용하면 변수가 규격 한계에 적합한 정도를 빠르게 요약하여 볼 수 있습니다 . 목표 그림 및 해당 기능에 대한 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

팁 : 목표 그림의 점을 커서로 가리키면 해당 공정에 대한 빠른 그래프가 표시됩니다 . 빠른 그래프를 클릭하여 보고서 창에 추가할 수 있습니다 .

" 목표 그림 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

군내 표준편차 점 표시 군내 표준편차 추정값을 사용하여 계산된 점을 표시하거나 숨깁니다 .

군내 또는 군간 - 군내 표준편차 점 표시 (시작 창에서 삼원 관리도를 선택한 경우에만 사용 가능) 군간 - 군내 표준편차 추정값을 사용하여 계산된 점을 표시하거나 숨깁니다 .

전체 표준편차 점 표시 전체 표준편차 추정값을 사용하여 계산된 점을 표시하거나 숨깁니다 .

음영 수준 Ppk 수준 음영을 표시하거나 숨깁니다 . 이 옵션은 기본적으로 설정되어 있습니다 . 음영 영역은 p 와 Ppk 사이의 관계에 따라 달라지며 , 이때 p 는 Ppk 아래의 상자에 표시된 값을 나타냅니다 . 기본적으로 $p = 1.33$ 입니다 .

- 빨간색 영역의 점은 $\text{Ppk} < p$ 입니다 .
- 노란색 영역의 점은 $p < \text{Ppk} < 2p$ 입니다 .
- 녹색 영역의 점은 $2p < \text{Ppk}$ 입니다 .

군내 표준편차 점에 라벨 지정 군내 표준편차 추정값을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

군내 또는 군간 - 군내 표준편차 점에 라벨 지정 (시작 창에서 삼원 관리도를 선택한 경우에만 사용 가능) 군간 - 군내 표준편차 추정값을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

전체 표준편차 점에 라벨 지정 전체 표준편차 추정값을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

결함 비율 등고선 지정된 결함 비율을 나타내는 등고선을 표시하거나 숨깁니다 .

공정 능력 표시 (하나 이상의 공정 변수에 대해 규격 한계가 정의된 경우에만 사용 가능) 요약 테이블에 공정 능력을 표시하거나 숨길 수 있는 하위 메뉴를 표시합니다 .

참고 : " 규격으로 중심화된 평균 " 및 " 규격으로 척도화된 표준편차 " 를 선택하면 요약 테이블에 "(평균 - 목표)/ 규격 범위 " 와 " 표준편차 / 규격 범위 " 가 표시됩니다 . " 규격 한계 " 옵션은 요약 테이블에 "LSL", "USL" 및 " 목표값 " 열을 표시합니다 .

하위 메뉴에서 다음 옵션도 사용할 수 있습니다 .

규격 이탈 값에 색상 적용 데이터 테이블에서 규격 이탈 값에 해당하는 셀에 색상을 적용합니다 . 값이 USL 을 초과하는 셀은 파란색으로 표시되고 LSL 미만인 셀은 빨간색으로 표시됩니다 .

팁 : 특정 셀의 색상을 제거하려면 관심 있는 모든 셀을 선택합니다 . 셀 중 하나를 마우스 오른쪽 버튼으로 클릭하고 " 색상 지우기 " 를 선택합니다 .

요약 테이블 저장 요약 테이블에 나타낼 수 있는 모든 정보를 " 공정 요약 " 데이터 테이블에 저장합니다 . 하나 이상의 공정에 대해 규격 한계가 지정된 경우 " 공정 요약 " 데이터 테이블에 규격 한계 상세 정보도 포함됩니다 . 요약 테이블과 마찬가지로 " 안정성 지수 ", "Ppk", "Cpk", "Cp" 및 " 목표 지수 " 열은 녹색 , 노란색 , 빨간색으로 표시되어 안정성 또는 공정 능력이 각각 적합 , 경계 , 나쁨 상태임을 나타냅니다 .

그래프와 함께 요약 테이블 저장 그래프 열이 추가된 " 공정 요약 " 데이터 테이블을 생성합니다 . 그래프 열에는 " 공정 요약 " 테이블의 각 공정에 대한 빠른 그래프가 포함되어 있습니다 .

상세 정보 테이블 저장 관리도 계산에 대한 상세 정보를 " 공정 상세 정보 " 데이터 테이블에 저장합니다 . 이 테이블에는 공정과 그룹화 변수의 각 조합에 대해 다음을 보여 주는 각 부분군의 행이 포함됩니다 .

- 부분군 변수의 값
- 부분군 표본 통계량의 값
- 관리 한계
- 부분군 크기
- 경보가 발동한 경우 이에 대한 표시자 목록 . Nelson 검정에 대한 경보는 검정 번호로 표시됩니다 . 범위 한계 초과 검정에 대한 경보는 "R" 로 표시됩니다 .
- 흐름

선택한 상세 정보 저장 요약 테이블에서 선택한 행에 대해 " 상세 정보 테이블 저장 " 을 선택할 때 저장되는 정보가 포함된 " 공정 상세 정보 " 테이블을 저장합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

공정 성능 그래프

"공정 성능 그래프"는 공정 능력과 안정성을 비교하는 4개의 사분면 그림입니다. 규격 한계가 제공된 각 공정을 표식으로 나타냅니다. 표식 유형은 공정이 목표 지수 및 Cp 값 지침을 위반하는지 여부에 따라 다릅니다. 왼쪽을 가리키는 삼각형은 목표 미달인 공정, 원은 목표에 부합하는 공정, 오른쪽을 가리키는 삼각형은 목표 초과인 공정을 각각 나타냅니다. 비어 있는 표식은 공정 변동이 적절함을 나타내고 채워진 표식은 공정 변동을 줄일 수 있음을 나타냅니다. 표 23.1에 표식 조합에 대한 전체 설명이 나와 있습니다. 또한 공정의 중요도 값을 정의하면 중요도에 따라 공정 성능 그래프의 표식 크기가 조정됩니다.

표 23.1 공정 성능 그래프 표식

그래프 표식	목표 지수	Cp
○	목표 부합	변동 양호
●	목표 부합	변동 감소
▷	목표 초과	변동 양호
▶	목표 초과	변동 감소
◁	목표 미달	변동 양호
◀	목표 미달	변동 감소

그래프에서 가로 좌표는 공정의 안정성 지수를 나타내고 세로 좌표는 공정 능력 (Ppk) 을 나타냅니다 . 그래프는 다음과 같은 기본 경계에 따라 4 개의 사분면으로 구분됩니다 .

– 안정성 지수가 1.25 를 초과하면 공정이 불안정하다는 것을 나타냅니다 .

– Ppk 가 1.33 보다 작으면 공정 능력이 없음을 나타냅니다 .

또한 Cpk 값이 1.33 인 위치를 나타내는 빨간색 선이 그래프에 있으며 1.33 에서 안정성 지수에 의해 구분됩니다 . 이 선은 공정을 수정 방법에 따라 불안정한 사분면과 공정 능력이 없는 사분면으로 분류합니다 . 빨간색 선 아래의 공정은 특수 원인에 의해 수정 가능하며 빨간색 선 위의 공정은 변동을 줄여 수정 가능합니다 . 그래프에서 점을 선택하면 요약 테이블에서 해당 공정이 선택됩니다 .

팁 : 공정 성능 그래프의 점을 커서로 가리키면 해당 공정에 대한 빠른 그래프가 표시됩니다 . 빠른 그래프를 클릭하여 보고서 창에 추가할 수 있습니다 .

공정 변수 선별의 추가 예

이 예에서는 Consumer Prices.jmp 를 사용하여 그룹화 열 사용 및 변화 그래프 생성을 보여 줍니다 .

소비자 물가 지수 데이터 테이블에는 17 개 상품에 대한 월별 데이터가 포함되어 있습니다 . 기간은 제품마다 다릅니다 . 정렬된 데이터는 17 개 종류의 모든 상품이 Series 라는 단일 열에 나열되어 있습니다 . 제품을 구분하려면 Series 를 그룹화 열로 처리해야 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Consumer Prices.jmp 를 엽니다 .
2. **분석 > 선별 > 공정 변수 선별**을 선택합니다 .

3. Price 를 선택하고 **공정 변수**를 클릭합니다 .
4. Series 를 선택하고 **그룹화**를 클릭합니다 .

   이렇게 하면 Series 의 각 수준이 개별 공정으로 처리됩니다 .
5. Date 를 선택하고 **시간**을 클릭합니다 .
6. **관리도 유형**을 "Xbar-R" 로 설정합니다 .
7. **부분군 표본 크기**를 "3" 으로 설정합니다 .

   데이터가 월별로 제공되므로 부분군 크기 3 은 분기를 나타냅니다 .
8. **확인**을 클릭합니다 .
9. Alt 키를 누른 채 " 공정 변수 선별 " 옆의 빨간색 삼각형을 클릭합니다 .

   그러면 사용 가능한 빨간색 삼각형 옵션을 보여 주는 창이 열립니다 . 이 창에서 한 번에 여러 옵션을 선택할 수 있습니다 .
10. **최대 상향 변화** , **최대 하향 변화** 및 **변화 그래프**를 선택합니다 .
11. **확인**을 클릭합니다 .

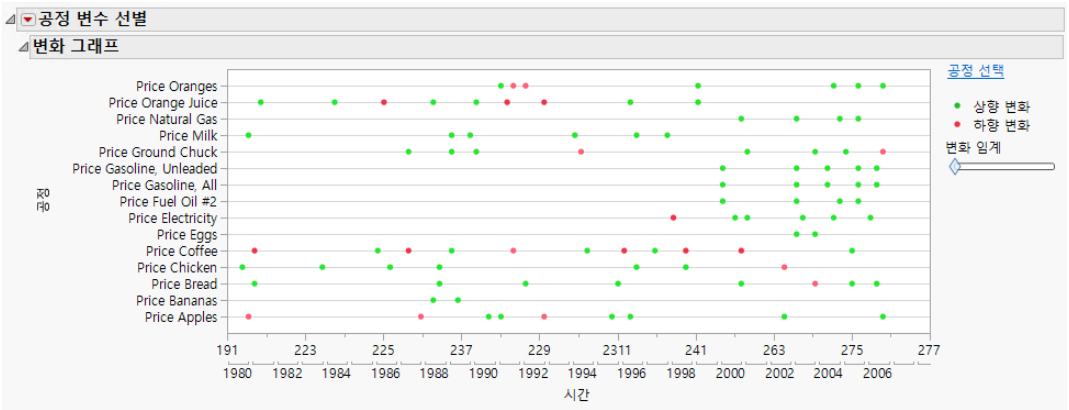
    " 최대 상향 변화 ", " 상향 변화 Date", " 최대 하향 변화 " 및 " 하향 변화 Date" 열이 요약 테이블에 추가됩니다 . 변화는 1 * 군내 표준편차 단위를 초과하는 최대 변화입니다 . 각 변화 위치는 시간 변수 Date 를 기준으로 제공됩니다 . 자세한 내용은 " 변화 크기 및 위치 " 에서 확인하십시오 .

    " 변화 그래프 " 도 나타납니다 . " 변화 그래프 " 에는 기본적으로 3 으로 설정된 **변화 임계** 군내 표준편차 단위 수를 초과하는 모든 변화가 표시됩니다 . 자세한 내용은 " 변화 그래프 " 에서 확인하십시오 . 녹색 점은 상향 변화에 해당하고 빨간색 점은 하향 변화에 해당합니다 .

    요약 테이블에 Gasoline, All 의 최대 상향 변화 값과 최대 하향 변화 값이 모두 있습니다 . 최대 하향 변화 값 1.8296 은 3 보다 작습니다 . " 변화 그래프 " 에는 군내 표준편차 단위 수가 3 이상인 변화만 표시되므로 Gasoline, All 의 최대 하향 변화 값은 " 변화 그래프 " 에 표시되지 않습니다 .

    Tomatoes 도 " 변화 그래프 " 에 포함되지 않습니다 . Tomatoes 의 경우 군내 표준편차 단위 수가 3 이상인 변화가 없었습니다 .
12. " 변화 그래프 " 의 가로 축을 두 번 클릭하여 "X 축 설정 " 창을 엽니다 .
13. " 눈금 / 계급 증분 " 패널에서 **보조 눈금 수**를 "1" 로 설정합니다 .
14. **라벨 행 내포**를 "2" 로 설정합니다 .
15. **확인**을 클릭합니다 .

그림 23.7 변화 그래프



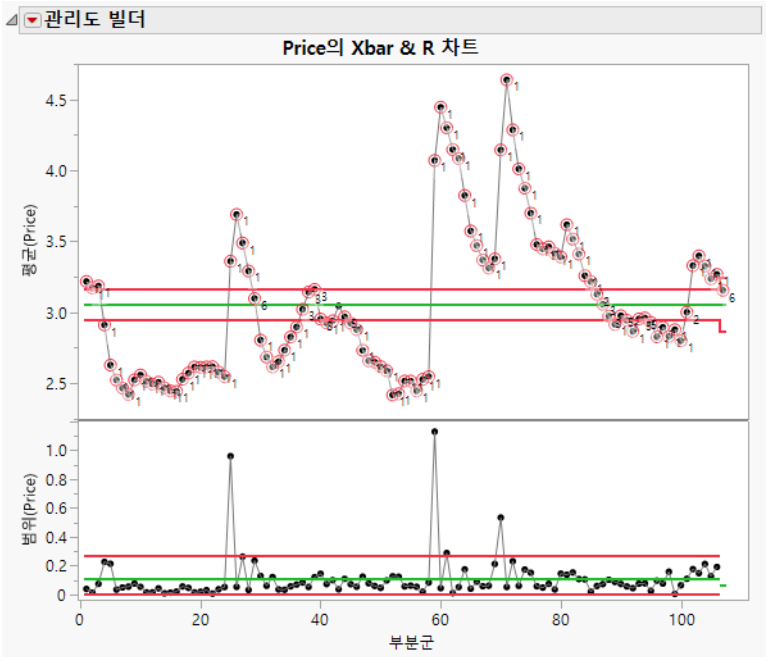
대부분의 계열에 주로 상향 변화가 표시됩니다 . 그러나 "Price Coffee" 의 경우 하향 변화와 상향 변화가 몇 번 번갈아 나타납니다 . 이 계열을 더 잘 이해하기 위해 관리도를 생성합니다 .

16. " 변화 그래프 " 에서 Price Coffee 오른쪽의 한 점을 선택하고 **공정 선택**을 클릭합니다 .

    이렇게 하면 요약 테이블에서 Coffee 에 해당하는 행이 선택됩니다 .

17. 요약 테이블에서 선택한 공정을 마우스 오른쪽 버튼으로 클릭하고 **선택한 항목에 대한 관리도**를 선택합니다 .

그림 23.8 Coffee 에 대한 관리도



관리도에는 " 변화 그래프 " 에서 식별된 상향 변화 및 하향 변화가 표시됩니다 .

요약 테이블을 보면 1994 년 9 월이 포함된 부분군에 대해 최대 상향 변화 (25.399 군내 표준편차 단위) 가 발생한다는 것을 알 수 있습니다 . 그림 23.8 의 관리도에서 이 부분군의 위치는 59 입니다 . 또한 요약 테이블에서는 1981 년 3 월을 포함하는 부분군에 대해 최대 하향 변화 (9.1674 군내 표준편차 단위) 가 발생한다는 것도 알 수 있습니다 . 관리도에서 이 부분군의 위치는 5 입니다 .

변화는 EWMA 평활 계열 및 이상치 수정 알고리즘을 사용하여 계산되므로 Shewhart 관리도에서 변화가 시작되는 부분군과 변화 위치가 정확하게 일치하지 않을 수 있습니다 .

공정 변수 선별 플랫폼에 대한 통계 상세 정보

중앙값을 사용하여 시그마를 추정하기 위한 척도 인수

"평균 대신 중앙값 사용"을 선택하면 척도화된 중앙값 범위 또는 중앙값 표준편차를 사용하여 시그마를 추정합니다. 아래 표에는 몬테카를로 시뮬레이션을 사용하여 구한 척도 인수가 제공됩니다.

다음 내용은 정규 분포에서 추출한 크기 *n* 인 부분군에 대해 참입니다 .

- 범위의 이론적 중앙값은 약 *d2_Median**σ 이며 , 여기서 *d2_Median* 은 *n* 에 해당하는 값입니다.
- 표준편차의 이론적 중앙값은 약 *c4_Median**σ 이며 , 여기서 *c4_Median* 은 *n* 에 해당하는 값입니다 .

표 23.2 중앙값 범위 및 중앙값 표준편차에 대한 척도 상수

n	d2_Median	c4_Median
2	0.953	0.675
3	1.588	0.833
4	1.978	0.888
5	2.257	0.917
6	2.471	0.933
7	2.646	0.944
8	2.792	0.952
9	2.915	0.959
10	3.024	0.963
11	3.118	0.967
12	3.208	0.969
13	3.286	0.972
14	3.357	0.975
15	3.422	0.976
16	3.483	0.978
17	3.539	0.979
18	3.590	0.980
19	3.640	0.981
20	3.685	0.982
21	3.731	0.983
22	3.770	0.984
23	3.811	0.984
24	3.846	0.985
25	3.883	0.986

24 장

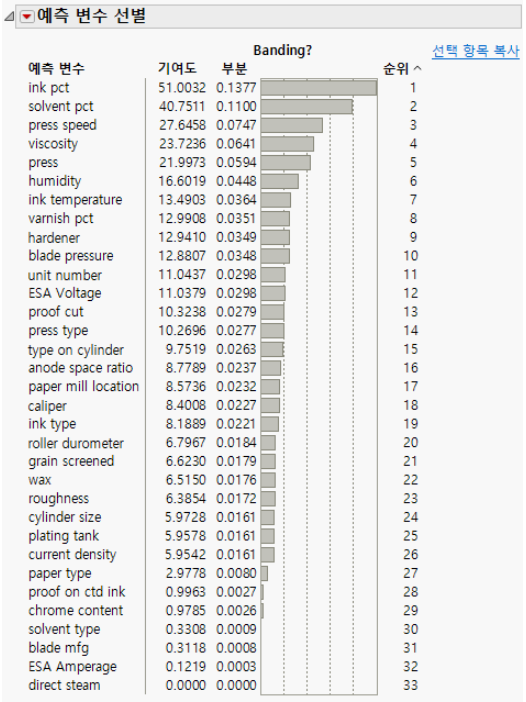
예측 변수 선별

유의한 효과를 위한 여러 예측 변수 선별

부품, 공정 또는 표본에 대해 수백 또는 수천 번의 측정을 수행하는 대규모 데이터 집합을 분석하려면 혁신적인 접근 방식이 필요합니다. 예측 변수 선별 플랫폼에서는 결과 예측 능력을 위해 많은 예측 변수를 선별하는 방법을 제공합니다. 예를 들어 예측 변수 선별을 사용하면 상태 예측을 위한 조건이 있는 환자와 없는 환자의 표본에서 테스트된 수천 개의 생체 지표를 식별할 수 있습니다.

예측 변수 선별은 반응 변수 선별과 다릅니다. 반응 변수 선별은 요인을 한 번에 하나씩 반응의 예측 변수로 검정합니다. 예측 변수 선별은 붓스트랩 포레스트 분할을 사용하여 반응에 대한 예측 변수의 기여도를 평가합니다. 파티션 모형은 여러 예측 변수를 기반으로 합니다. 예측 변수 선별은 단독으로는 약하지만 다른 예측 변수와 함께 사용하면 강력한 예측 변수를 식별할 수 있습니다. 자세한 내용은 "반응 변수 선별"에서 확인하십시오.

그림 24.1 예측 변수 선별 보고서의 예



예측 변수 선별

Banding?

선택 항목 복사

예측 변수	기여도	부분	순위
ink pct	51.0032	0.1377	1
solvent pct	40.7511	0.1100	2
press speed	27.6458	0.0747	3
viscosity	23.7236	0.0641	4
press	21.9973	0.0594	5
humidity	16.6019	0.0448	6
ink temperature	13.4903	0.0364	7
varnish pct	12.9908	0.0351	8
hardener	12.9410	0.0349	9
blade pressure	12.8807	0.0348	10
unit number	11.0437	0.0298	11
ESA Voltage	11.0379	0.0298	12
proof cut	10.3238	0.0279	13
press type	10.2696	0.0277	14
type on cylinder	9.7519	0.0263	15
anode space ratio	8.7789	0.0237	16
paper mill location	8.5736	0.0232	17
caliper	8.4008	0.0227	18
ink type	8.1889	0.0221	19
roller durometer	6.7967	0.0184	20
grain screened	6.6230	0.0179	21
wax	6.5150	0.0176	22
roughness	6.3854	0.0172	23
cylinder size	5.9728	0.0161	24
plating tank	5.9578	0.0161	25
current density	5.9542	0.0161	26
paper type	2.9778	0.0080	27
proof on ctd ink	0.9963	0.0027	28
chrome content	0.9785	0.0026	29
solvent type	0.3308	0.0009	30
blade mfg	0.3118	0.0008	31
ESA Amperage	0.1219	0.0003	32
direct steam	0.0000	0.0000	33

목차

예측 변수 선별 플랫폼 개요 469
예측 변수 선별의 예 469
예측 변수 선별 플랫폼 시작 471
예측 변수 선별 보고서 472
예측 변수 선별 플랫폼 옵션 472

예측 변수 선별 플랫폼 개요

예측 변수 선별은 다수의 후보에서 유의한 예측 변수를 식별하는 데 유용합니다. 수백 개의 X가 있고 이 중 결과의 예측 변수로 가장 유의한 X를 결정해야 한다고 가정해 보겠습니다.

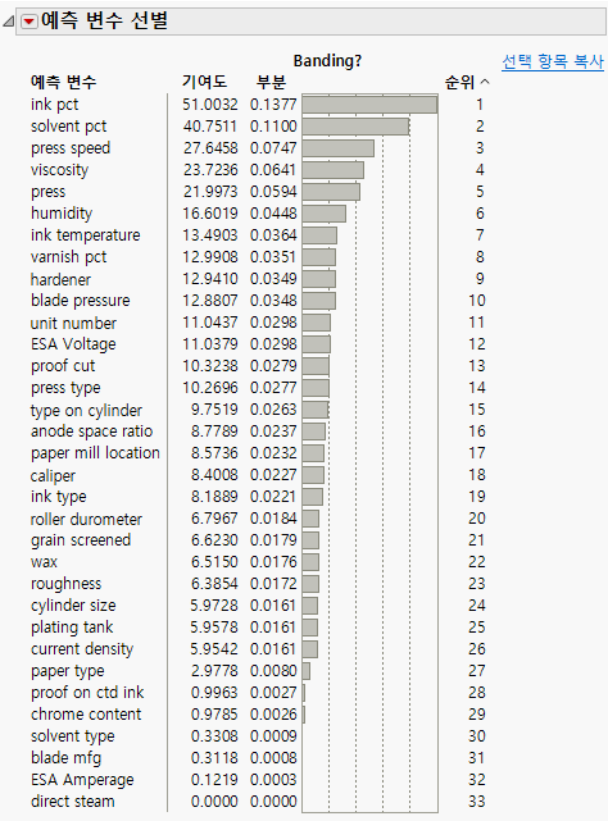
예측 변수 선별 플랫폼에서는 붓스트랩 포레스트를 사용하여 반응의 잠재적 예측 변수를 식별합니다. 각 반응에 대해 100개의 결정 트리를 사용하는 붓스트랩 포레스트 모형이 생성됩니다. 각 예측 변수의 붓스트랩 포레스트 모형에 대한 열 기여도가 순위화됩니다. 붓스트랩 포레스트 방법은 랜덤 성분을 포함하므로 보고서를 다시 실행하면 열 기여도가 다를 수 있습니다. 결정 트리에 대한 자세한 내용은 "파티션 모형"에서 확인하십시오.

예측 변수 선별의 예

Bands Data.jmp 데이터 테이블에는 그라비어 인쇄 회사의 기계에서 측정한 값이 포함되어 있습니다 . 데이터 집합에는 539 개의 레코드와 38 개의 변수가 있습니다 . 반응 Y 는 Banding? 열이고 값은 "BAND" 와 "NOBAND" 입니다 . 반응에 기여할 가능성이 가장 높은 특성이 무엇인지 알아보려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Bands Data.jmp 를 엽니다 .
2. **분석 > 선별 > 예측 변수 선별**을 선택합니다 .
3. Banding? 을 **Y, 반응**으로 선택합니다 .
4. 그룹화된 grain screened ~ chrome content 열을 선택하고 **X** 를 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 24.2 순위화된 열 기여도



예측 변수	기여도	부분	순위
ink pct	51.0032	0.1377	1
solvent pct	40.7511	0.1100	2
press speed	27.6458	0.0747	3
viscosity	23.7236	0.0641	4
press	21.9973	0.0594	5
humidity	16.6019	0.0448	6
ink temperature	13.4903	0.0364	7
varnish pct	12.9908	0.0351	8
hardener	12.9410	0.0349	9
blade pressure	12.8807	0.0348	10
unit number	11.0437	0.0298	11
ESA Voltage	11.0379	0.0298	12
proof cut	10.3238	0.0279	13
press type	10.2696	0.0277	14
type on cylinder	9.7519	0.0263	15
anode space ratio	8.7789	0.0237	16
paper mill location	8.5736	0.0232	17
caliper	8.4008	0.0227	18
ink type	8.1889	0.0221	19
roller durometer	6.7967	0.0184	20
grain screened	6.6230	0.0179	21
wax	6.5150	0.0176	22
roughness	6.3854	0.0172	23
cylinder size	5.9728	0.0161	24
plating tank	5.9578	0.0161	25
current density	5.9542	0.0161	26
paper type	2.9778	0.0080	27
proof on ctd ink	0.9963	0.0027	28
chrome content	0.9785	0.0026	29
solvent type	0.3308	0.0009	30
blade mfg	0.3118	0.0008	31
ESA Amperage	0.1219	0.0003	32
direct steam	0.0000	0.0000	33

참고: 이 분석은 랜덤 선택 성분이 있는 붓스트랩 포레스트 방법을 기반으로 하므로 결과가 그림 24.2 에 표시된 것과 약간 다를 수 있습니다 . 자세한 내용은 " 붓스트랩 포레스트 " 에서 확인하십시오 .

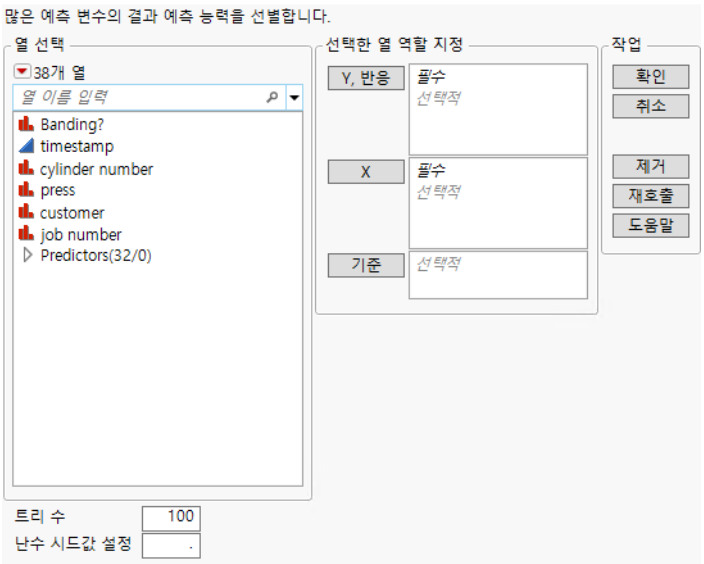
열은 붓스트랩 포레스트 모형에서 기여도에 따라 정렬되고 순위가 매겨집니다. 기여도가 가장 높은 예측 변수가 관심 반응에 대한 강력한 후보 예측 변수입니다.

팁 : 테이블에서 열 정렬 방식을 변경하려면 열 머리글을 클릭합니다 .

예측 변수 선별 플랫폼 시작

분석 > 선별 > 예측 변수 선별을 선택하여 예측 변수 선별 플랫폼을 시작합니다 .

그림 24.3 예측 변수 선별 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 반응 반응 열입니다 .

X 예측 변수 열입니다 .

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다 . 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다 . 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

트리 수 붓스트랩 포레스트 모형의 결정 트리 수입니다 .

난수 시드값 설정 적합 절차에 사용되는 시작 값의 시드를 설정합니다 . " 난수 시드값 설정 " 은 분석을 재현하려는 경우에 유용합니다 . 난수 시드값을 설정하고 스크립트를 저장하면 시드값이 스크립트에 자동으로 저장됩니다 .

예측 변수 선별 보고서

보고서(그림 24.2)에는 각 기여도 및 순위가 포함된 예측 변수 목록이 표시됩니다. 기여도가 가장 높은 예측 변수가 Y 예측에 중요할 수 있습니다.

"기여도" 열에는 붓스트랩 포레스트 모형에 대한 각 예측 변수의 기여도가 표시됩니다. 보고서의 "부분" 열에는 각 변수의 기여도가 백분율로 표시됩니다.

"예측 변수 선별" 보고서에서 중요한 예측 변수를 선택할 수 있습니다. 중요한 예측 변수를 선택하면 데이터 테이블에서 해당 열이 선택되므로 모델링 플랫폼의 시작 창에 열을 쉽게 입력할 수 있습니다. "선택 항목 복사" 버튼을 사용하면 선택한 열을 클립보드에 복사하여 모형에도 붙여넣을 수 있습니다. 이런 방식으로 예측 변수 선별은 모델링 프로세스를 향상시킵니다

예측 변수 선별 플랫폼 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

25 장

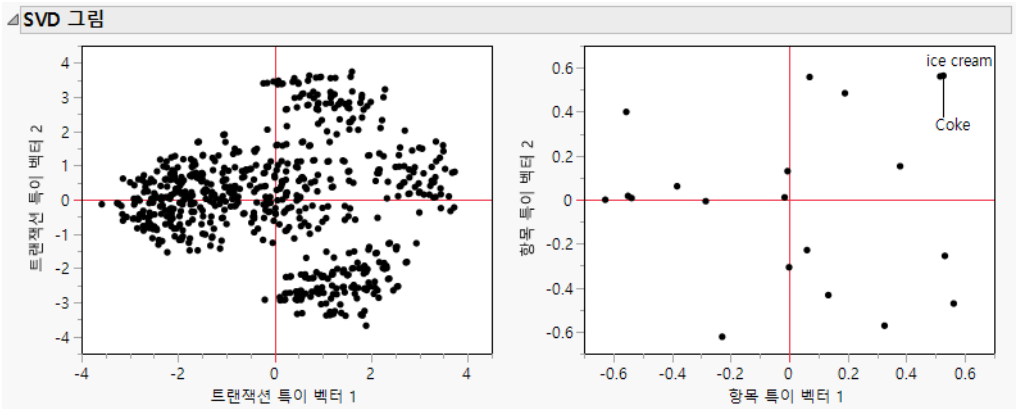
JMP PRO 연관성 분석
장바구니 분석 수행

JMP PRO 연관성 분석 플랫폼은 JMP Pro 에서만 사용할 수 있습니다 .

연관성 분석 플랫폼을 사용하면 서로 선호도가 있는 항목을 식별할 수 있습니다. 트랜잭션에서 자주 함께 나타나는 항목을 식별하기 위해 트랜잭션 데이터(장바구니라고도 함)를 분석하는 데 흔히 사용됩니다. 예를 들어 식료품점과 온라인 판매자는 연관성 분석을 사용하여 함께 구매하는 경향이 있는 상품을 전략적으로 구성하고 추천합니다.

연관성 분석은 종속 사건 또는 연관된 사건을 식별할 때도 사용됩니다. 예를 들어 비슷한 시기에 고장 난 것 같은 자동차 부품을 식별할 수 있습니다. 이 경우 차량 검사를 장바구니로 간주하고 각 검사에서 발견된 결함 부품 그룹 간의 연관성을 분석합니다.

그림 25.1 특이값 분해 그림의 예



목차

연관성 분석 플랫폼 개요 475
연관성 분석 플랫폼의 예 476
연관성 분석 플랫폼 시작 478
데이터 형식 479
연관성 분석 보고서 479
빈발 항목 집합 480
규칙 480
연관성 분석 플랫폼 옵션 481
특이값 분해 481
SVD 보고서 482
SVD 보고서 옵션 483
주제 분석 484
주제 분석 보고서 484
주제 분석 보고서 옵션 485
추가 예 : SVD 분석 485
연관성 분석 플랫폼에 대한 통계 상세 정보 488
빈발 항목 집합 생성 488
연관성 분석 성능 측도 488

JMP PRO 연관성 분석 플랫폼 개요

연관성 분석 플랫폼에서는 독립 사건 또는 **트랜잭션**에서 항목 그룹 간의 연관성을 식별합니다. 연관성 분석에서 **항목**은 기본 관심 대상입니다. 예를 들어 항목은 제품, 웹 페이지 또는 서비스일 수 있습니다. **항목 집합**은 하나 이상의 항목 목록입니다.

두 항목 집합 간의 관계는 **연관 규칙**으로 정의됩니다. 연관 규칙은 **조건** 항목 집합과 **결과** 항목 집합으로 구성됩니다. **선행 항목**은 조건 항목 집합의 개별 항목입니다. 연관성 분석은 조건 항목 집합이 이미 트랜잭션에 있는 경우 결과 항목 집합이 트랜잭션에 있을 것으로 예측하는 연관 규칙을 식별합니다. 일부 연관 규칙은 다른 규칙보다 더 강력하므로 더 유용합니다. 다음과 같은 세 가지 성능 측도로 연관 규칙 강도를 설명합니다.

- **지지도**는 항목 집합이 나타나는 트랜잭션의 비율입니다. 지지도 값이 높으면 해당 항목 집합이 자주 발생한다는 것을 나타냅니다.
- **신뢰도**는 트랜잭션에 조건 항목 집합이 있을 때 결과 항목 집합을 포함하는 트랜잭션의 비율입니다. 신뢰도는 연관 규칙의 영향력 또는 예측 검정력을 측정합니다. 연관성 분석의 신뢰도는 신뢰 구간의 개념과 관련이 없습니다.
- **향상도**는 조건 항목 집합과 결과 항목 집합이 트랜잭션에서 독립적으로 발생한다고 가정할 때 기대 신뢰도에 대한 연관 규칙 신뢰도의 비율입니다. 향상도는 조건 항목 집합이 있는지 여부에 따라 결과 항목 집합이 얼마나 달라지는지 측정합니다. 향상도 최소값은 0 입니다.
  - 향상도 비율이 1 보다 작으면 조건과 결과가 함께 발생하는 빈도가 우연히 발생할 경우보다 낮으므로 서로 밀어낸다는 것을 나타냅니다.
  - 향상도 비율이 1 에 가까우면 결과가 조건이 포함된 트랜잭션에서 발생하는 비율이 우연히 발생할 비율과 같음을 나타냅니다.
  - 향상도 비율이 1 보다 크면 결과 항목 집합에 조건 항목 집합에 대한 선호도가 있음을 나타냅니다. 결과 항목 집합은 우연히 발생하는 것보다 조건 항목 집합과 함께 발생할 빈도가 더 높습니다.

이러한 성능 측도에 대한 자세한 내용은 "연관성 분석 성능 측도"에서 확인하십시오.

연관성 분석 플랫폼을 사용하면 특이값 분해도 수행할 수 있습니다. SVD(특이값 분해)는 연관성 분석에 사용되는 것과 다른 행렬 축소 방법론을 사용하여 유사한 트랜잭션 및 유사한 항목을 그룹화합니다. 연관성 분석에서 얻은 결과를 보완하는 정보를 제공하려면 SVD 방법론을 사용합니다.

연관성 분석에 대한 자세한 내용은 Hastie et al. (2009) 및 Shmueli et al. (2010) 연구 자료에서 확인하십시오. 특이값 분해에 대한 자세한 내용은 Jolliffe(2002) 연구 자료에서 확인하십시오.

JMP PRO 연관성 분석 플랫폼의 예

이 예에서는 식료품점에서 보고한 트랜잭션 데이터가 포함된 Grocery Purchases.jmp 샘플 데이터 테이블을 사용합니다 . 데이터 테이블에는 각각 고유한 고객 ID 가 할당된 1001 명의 고객이 구입한 항목이 나열되어 있습니다 . 소비자 행동 패턴을 식별하기 위해 항목 간의 연관성을 탐색하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Grocery Purchases.jmp 를 엽니다 .
2. **분석 > 선별 > 연관성 분석**을 선택합니다 .
3. Product 를 선택하고 **항목**을 클릭합니다 .
4. Customer ID 를 선택하고 **ID** 를 클릭합니다 .
5. **확인**을 클릭합니다 .

   기본적으로 " 규칙 " 보고서는 **신뢰도**에 따라 내림차순으로 정렬됩니다 . 그러나 신뢰도가 매우 높은 연관 규칙은 조건 집합의 항목 수가 많은 더 경향도 있습니다 . 더 작은 조건 집합의 연관 규칙을 검토할 예정이므로 보고서를 **신뢰도**에 따라 오름차순으로 정렬합니다 .
6. " 규칙 " 보고서에서 마우스 오른쪽 버튼을 클릭하고 **열별 정렬**을 선택합니다 .

   " 열 선택 " 창이 나타납니다 .
7. 신뢰도를 선택한 후 " 오름차순 " 옵션을 선택합니다 .
8. **확인**을 클릭합니다 .

그림 25.2 연관성 분석 보고서

규칙

조건	규칙 결과	신뢰도	Lift
apples	sardines	40%	1.357
Coke	turkey	40%	1.422
sardines	chicken, Heineken	40%	2.002
herring, olives	ham, turkey	40%	3.006
ice cream	turkey	40%	1.424
artichoke	baguette, Heineken	40%	1.547
ham	corned beef, herring	40%	1.648
ham	corned beef, olives	40%	1.703
peppers	sardines	41%	1.371
sardines	peppers	41%	1.371
Coke	Heineken, sardines	41%	2.23
sardines	Coke, Heineken	41%	2.332
avocado, Heineken	crackers, ham	41%	2.707
soda	baguette, herring	41%	1.631
corned beef	ham	41%	1.343
Heineken, soda	bourbon, crackers, olives	41%	3.152
artichoke, Heineken	avocado, baguette	41%	1.921
Heineken, herring	avocado, baguette	41%	1.924

9. 신뢰도 값 "58%" 가 나올 때까지 보고서를 아래로 스크롤합니다 .

   " 규칙 " 보고서 테이블에 avocado(아보카도) 를 구입한 고객의 58% 가 artichoke(아티초크) 도 구입했음을 나타내는 항목이 있습니다 . " 향상도 " 값은 1.908 로 , 종속성 가능성이 있음을 나타냅니다 . 트랜잭션에서 아보카도와 아티초크가 발생하는 비율이 유의한지 확인하려고 합니다 .

10. " 빈발 항목 집합 " 옆의 표시 아이콘을 클릭합니다 .

그림 25.3 빈발 항목 집합 보고서

빈발 항목 집합

항목 집합	지지도	항목 수
{Heineken}	60%	1
{crackers}	49%	1
{herring}	49%	1
{olives}	47%	1
{bourbon}	40%	1
{baguette}	39%	1
{corned beef}	39%	1
{crackers, Heineken}	37%	2
{avocado}	36%	1
{soda}	32%	1
{chicken}	31%	1
{apples}	31%	1
{ice cream}	31%	1
{artichoke}	30%	1
{ham}	30%	1
{Coke}	30%	1
{peppers}	30%	1
{sardines}	30%	1
{Heineken, herring}	29%	2
{turkey}	28%	1
{baguette, Heineken}	26%	2
{Heineken, soda}	26%	2
{herring, olives}	26%	2
{artichoke, Heineken}	25%	2

" 빈발 항목 집합 " 보고서에 따르면 고객의 36% 가 아보카도를 구입했습니다 . 그림 25.2 의 " 규칙 " 보고서에서는 이러한 고객의 58% 가 아티초크도 구입했음을 보여 줍니다 . 이 행동을 보이는 고객의 비율이 높기 때문에 식료품점 관리자는 이 정보를 사용하여 아보카도와 아티초크를 전략적으로 배치할 수 있습니다 .

향상도가 가장 높은 연관 규칙도 살펴보려고 합니다 .

11. " 규칙 " 보고서 테이블에서 마우스 오른쪽 버튼을 클릭하고 **열별 정렬**을 선택합니다 .

" 열 선택 " 창이 나타납니다 .

12. **향상도**를 선택하고 **확인**을 클릭합니다 .

" 규칙 " 테이블이 향상도 값에 따라 내림차순으로 정렬됩니다 . 두 번째 연관 규칙은 향상도가 5.642 이고 신뢰도가 83% 입니다 . 조건 집합 {chicken, ice cream} 과 결과 항목 집합 {Coke, sardines} 모두 지지도가 적절한지 확인하려고 합니다 .

13. " 빈발 항목 집합 " 보고서에서 마우스 오른쪽 버튼을 클릭하고 **열별 정렬**을 선택합니다 .

" 열 선택 " 창이 나타납니다 .

14. **항목 집합**을 선택하고 " 오름차순 " 옵션을 선택합니다 .

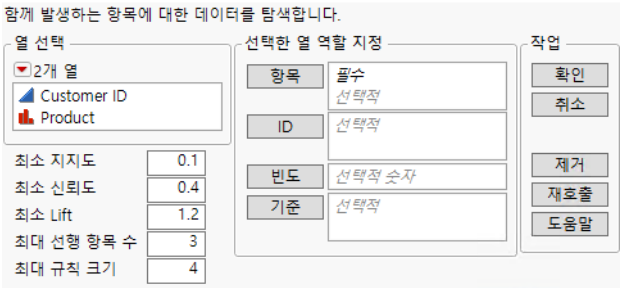
15. **확인**을 클릭합니다 .

" 빈발 항목 집합 " 테이블이 항목 집합에 따라 사전순으로 정렬됩니다 . 목록을 스크롤하여 조건 항목 집합 {chicken, ice cream} 의 지지도가 14% 인지 확인합니다 . 결과 항목 집합 {Coke, sardines} 의 지지도는 15% 입니다 . 이 연관 규칙은 향상도 값이 높지만 첫 번째로 검토한 연관 규칙보다 트랜잭션 수가 적음을 나타냅니다 .

JMP PRO 연관성 분석 플랫폼 시작

분석 > 선별 > 연관성 분석을 선택하여 연관성 분석 플랫폼을 시작합니다.

그림 25.4 연관성 분석 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

항목 분석할 항목 데이터가 포함된 하나 이상의 범주형 열입니다. 열의 모델링 유형이 "다중 반응"인 경우 각 행의 각 다중 반응이 하나의 항목으로 처리됩니다. 여러 열을 지정한 경우 각 행의 각 열이 하나의 항목으로 처리됩니다.

ID 항목이 속한 트랜잭션을 식별하는 열입니다. "항목" 역할에 할당된 열의 모델링 유형이 "다중 반응"이거나 "항목" 역할에 여러 열이 지정된 경우를 제외하고 "ID" 역할은 필수입니다.

참고 : "ID" 역할이 지정되면 ID 가 동일한 모든 행이 하나의 트랜잭션으로 결합됩니다.

빈도 "항목" 열에서 트랜잭션 빈도를 식별하는 열입니다. 예를 들어 빈도가 3 인 트랜잭션은 "항목" 열에서 항목이 포함된 세 개의 동일한 트랜잭션으로 간주됩니다.

참고 : "항목" 역할에 할당된 열의 모델링 유형이 "다중 반응"이 아니면 "빈도" 역할이 무시됩니다.

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

최소 지지도 항목 집합 발생 비율에 대한 최소값을 지정합니다. 이 값은 0 에서 1 사이여야 합니다. 지지도가 이 값보다 크거나 같은 항목 집합만 분석에 포함됩니다.

최소 신뢰도 조건 항목 집합이 포함된 트랜잭션 내에서 결과 항목 집합이 발생할 비율에 대한 최소값을 지정합니다. 이 값은 0 에서 1 사이여야 합니다. 신뢰도가 이 값보다 크거나 같은 연관 규칙만 보고서에 나타납니다.

최소 향상도 최소 종속성 비율을 지정합니다. 향상도 값은 0 이상이어야 합니다. 향상도가 이 값보다 크거나 같은 연관 규칙만 보고서에 나타납니다.

최대 선행 항목 수 조건 항목 집합의 최대 항목 수를 지정합니다. 조건 집합의 항목 수가 이 값보다 큰 연관 규칙은 분석에 포함되지 않습니다.

최대 규칙 크기 조건 항목 집합과 결과 항목 집합의 결합에 나타나는 최대 항목 수를 지정합니다. 결합 항목 수가 이 값보다 큰 연관 규칙은 분석에 포함되지 않습니다. 기본값은 4 입니다.

참고 : 시작 창에서 최소 지지도, 최대 선행 항목 수 및 최대 규칙 크기 옵션을 사용하여 대규모 데이터 집합의 계산 시간을 줄일 수 있습니다. 이러한 측도에 대한 자세한 내용은 " 연관성 분석 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오.

데이터 형식

연관성 분석 플랫폼에는 다음과 같은 세 가지 데이터 형식이 허용됩니다.

- 각 행의 단일 항목 반응을 " 항목 " 역할에 지정하고 "ID" 역할을 사용하여 각 트랜잭션에 포함되는 항목을 식별할 수 있습니다. 이 형식의 경우 " 항목 " 및 "ID" 역할은 필수이고 " 빈도 " 역할은 무시됩니다.
- 모델링 유형이 " 다중 반응 " 인 열을 " 항목 " 역할에 지정할 수 있습니다. 이 형식의 경우 " 항목 " 역할은 필수이고 "ID" 및 " 빈도 " 역할은 선택적입니다. "ID" 역할이 지정되면 ID 가 동일한 모든 행이 하나의 트랜잭션으로 결합됩니다. " 빈도 " 역할이 지정되면 트랜잭션의 특정 항목이 아니라 트랜잭션에 빈도가 적용됩니다.
- 다중 항목 반응 열을 " 항목 역할에 지정할 수 있습니다. 이 형식의 경우 "ID" 및 " 빈도 " 역할은 선택적입니다. "ID" 역할이 지정되면 ID 가 동일한 모든 행이 하나의 트랜잭션으로 결합됩니다. " 빈도 " 역할이 지정되면 트랜잭션의 특정 항목이 아니라 트랜잭션에 빈도가 적용됩니다.

JMP PRO 연관성 분석 보고서

기본적으로 " 연관성 분석 " 보고서에는 다음 보고서가 포함됩니다.

- " 빈발 항목 집합 "
- " 규칙 "

팁 : 보고서의 테이블 내용을 특정 열을 기준으로 정렬하려면 테이블에서 마우스 오른쪽 버튼을 클릭하고 **열별 정렬**을 선택합니다.

JMP PRO 빈발 항목 집합

"빈발 항목 집합" 보고서에는 항목 집합이 지지도에 따라 내림차순으로 나열됩니다. 나열된 항목 집합은 시작 창에 지정한 최소 지지도 값을 충족합니다. 각 항목 집합은 조건 항목 집합 및 연관 규칙을 구성하기 위한 결과 항목 집합으로 간주됩니다. 이 테이블에는 다음 열이 포함됩니다.

항목 집합 조건 집합 또는 연관 규칙을 위한 결과 집합으로 간주되는 항목 집합입니다.

지지도 항목 집합의 모든 항목이 발생하는 트랜잭션의 비율입니다.

항목 수 항목 집합의 항목 수입니다.

JMP PRO 규칙

"규칙" 보고서에는 신뢰도에 따라 내림차순으로 정렬된 연관 규칙 테이블이 표시됩니다. 시작 창에서 지정한 최소 지지도, 최소 신뢰도, 최소 향상도, 최대 선행 항목 수 및 최대 규칙 크기 요구 사항을 충족하는 연관 규칙만 이 보고서에 나타납니다.

"규칙" 보고서 테이블에는 다음 열이 포함됩니다.

규칙 조건 항목 집합과 결과 항목 집합을 결합하여 구성된 연관 규칙입니다.

조건 트랜잭션 내에서 결과 항목 집합의 존재 여부에 영향을 미치는 것으로 간주되는 항목 집합입니다.

결과 존재 여부가 조건 항목 집합의 존재 여부에 영향을 받는 것으로 간주되는 항목 집합입니다.

신뢰도 조건 항목 집합이 트랜잭션에 있을 때 결과 항목 집합을 포함하는 트랜잭션의 비율입니다. 신뢰도는 연관 규칙의 영향력 또는 예측 검정력을 측정합니다.

향상도 조건 항목 집합과 결과 항목 집합이 트랜잭션에서 독립적으로 발생한다고 가정할 때 기대 신뢰도에 대한 연관 규칙 신뢰도의 비율입니다. 향상도는 조건 항목 집합이 있는지 여부에 따라 결과 항목 집합이 얼마나 달라지는지 측정합니다. 향상도 최소값은 0 입니다.

- 향상도 비율이 1 보다 작으면 조건 항목 집합과 결과 항목 집합이 함께 발생하는 빈도가 우연히 발생할 경우보다 낮으므로 서로 밀어낸다는 것을 나타냅니다.
- 향상도 비율이 1에 가까우면 결과 항목 집합이 조건 항목 집합이 포함된 트랜잭션에서 발생하는 비율이 우연히 발생할 비율과 같음을 나타냅니다.
- 향상도 비율이 1 보다 크면 결과 항목 집합에 조건 항목 집합에 대한 선호도가 있음을 나타냅니다. 결과 항목 집합은 우연히 발생하는 것보다 조건 항목 집합과 함께 발생할 빈도가 더 높습니다.

JMP PRO 연관성 분석 플랫폼 옵션

" 연관성 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

트랜잭션 목록 각 트랜잭션 ID 값 및 해당 트랜잭션에 포함된 항목을 나열하는 테이블을 표시하거나 숨깁니다 . 테이블은 " 트랜잭션 ID" 열을 기준으로 정렬됩니다 .

빈발 항목 집합 지지도가 시작 창에 지정된 최소 지지도 값을 초과하는 항목 집합 목록을 표시하거나 숨깁니다 . 자세한 내용은 " 빈발 항목 집합 " 에서 확인하십시오 .

규칙 시작 창에 지정된 최소 지지도 , 최소 신뢰도 , 최소 향상도 , 최대 선행 항목 수 및 최대 규칙 크기 요구 사항을 충족하는 연관 규칙 테이블을 표시하거나 숨깁니다 . 자세한 내용은 " 규칙 " 에서 확인하십시오 .

SVD 항목에 대한 따름 행렬의 부분 SVD(특이값 분해) 보고서를 표시하거나 숨깁니다 . 이 분해는 따름 행렬을 분석을 위해 사용자가 지정한 차원 수로 축소합니다 . SVD 보고서에는 특이 벡터 , 군집 트랜잭션 및 군집 항목을 시각화하고 , 특이 벡터를 저장하거나 , 회전된 SVD 분석을 수행하는 옵션도 포함되어 있습니다 . 자세한 내용은 " 특이값 분해 " 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO 특이값 분해

SVD(특이값 분해)는 서로 선호도가 있는 항목을 식별할 수 있는 다른 방법을 제공하여 연관성 분석을 보완합니다. 트랜잭션 항목 행렬의 특이값 분해는 행렬을 관리 가능한 차원 수로 축소하여 유사한 트랜잭션 및 유사한 항목을 그룹화할 수 있도록 합니다. SVD 분석은 상관 행렬에 대해 PCA(주성분 분석)를 수행하는 것과 동일합니다.

트랜잭션 항목 행렬은 각 행이 트랜잭션에 해당하고 각 열이 항목에 해당하는 행렬입니다. 행렬의 항목은 0과 1입니다. 트랜잭션에서 항목이 발생하면 해당 행과 열 항목이 1입니다. 그렇지 않으면 행과 열 항목이 0입니다. 일반적으로 트랜잭션 항목 행렬에는 1보다 0 값이 더 많으므로 **희소 행렬**이라고 합니다.

부분 특이값 분해는 세 개의 행렬 **U**, **S** 및 **V**′ 를 사용하여 열 표준화 트랜잭션 항목 행렬에 근사한 값을 산출합니다. 이러한 행렬 간의 관계는 다음과 같이 정의됩니다.

트랜잭션 항목 행렬 ≈ **U** * **S** * **V**′

*nTran*을 트랜잭션 항목 행렬의 트랜잭션(행) 수, *nItem*을 트랜잭션 항목 행렬의 항목(열) 수, *nVec*를 지정된 특이 벡터 수로 각각 정의합니다. *nVec*는 min(*nTran*, *nItem*)보다 작거나 같아야 합니다. 따라서 **U**는 트랜잭션 항목 행렬의 왼쪽 특이 벡터를 포함하는 *nTran* x *nVec* 행렬입니다. **S**는 *nVec* 차 대각 행렬입니다. **S**의 대각 항목은 트랜잭션 항목 행렬의 특이값입니다. **V**′는 *nVec* x *nItem* 행렬입니다. **V**′의 행(또는 **V**의 열)은 오른쪽 특이 벡터입니다.

오른쪽 특이 벡터는 함수 또는 주제 영역이 유사한 서로 다른 항목 간의 연관성을 포착합니다. 세 개의 항목이 동일한 트랜잭션에 나타나는 경향이 있는 경우 SVD에서는 **V**′에 이 세 개 항목에 대해 큰 값을 갖는 특이 벡터를 생성할 가능성이 높습니다. **U** 특이 벡터는 이 새로운 항목 공간에 투영된 트랜잭션을 나타냅니다.

SVD는 간접 관계도 포착합니다. 두 개의 항목이 동일한 트랜잭션에 함께 나타나지는 않지만 일반적으로 다른 세 번째 항목이 있는 트랜잭션에 나타나는 경우 SVD는 이러한 연관성을 포착할 수 있습니다. 두 개의 트랜잭션에 공통된 항목은 없지만 차원 축소 공간에서 연결되는 항목이 포함된 경우 이 두 트랜잭션은 SVD 그림에서 유사 벡터에 매핑됩니다.

SVD는 트랜잭션 데이터를 고정 차원 벡터 공간으로 변환하여 군집화, 분류 및 회귀 기법에 적용할 수 있도록 합니다. 저장 옵션을 사용하면 이 벡터 공간을 다른 JMP 플랫폼에서 분석할 수 있는 형태로 내보낼 수 있습니다.

특이값 분해가 수행되기 전에 트랜잭션 항목 행렬을 중심화하고 척도화한 후 *nTran* - 1로 나눕니다. 이 분석은 트랜잭션 항목 행렬의 상관 행렬에 대한 PCA와 동일합니다. SVD 구현은 트랜잭션 항목 행렬의 희소성을 활용합니다.

JMP PRO SVD 보고서

SVD 옵션은 특이값 분해를 통해 두 개의 SVD 그림과 한 개의 특이값 테이블을 생성합니다.

JMP PRO SVD 그림

첫 번째 그림에는 각 트랜잭션에 대한 점이 포함되어 있습니다. 특정 트랜잭션에 대해 표시되는 점은 처음 두 개의 특이 벡터(**U** 행렬의 처음 두 열)에 있는 트랜잭션 값에 대각 특이값 행렬(**S**)을 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 스코어 그림과 동일합니다. 이 그림의 각 점은 트랜잭션을 나타냅니다. 트랜잭션 SVD 그림에서 눈에 띄게 그룹화된 점은 구성이 유사한 트랜잭션을 나타냅니다.

두 번째 그림에는 각 항목에 대한 점이 포함되어 있습니다. 특정 항목에 대해 표시되는 점은 처음 두 개의 특이 벡터(**V**′ 행렬의 처음 두 행)에 있는 항목 값에 대각 특이값 행렬(**S**)을 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 적재 그림과 동일합니다. 이 그림의 각 점은 항목을 나타냅니다. 항목 SVD 그림에서 눈에 띄게 그룹화된 항목은 함수 또는 주제 영역이 유사한 항목을 나타냅니다.

JMP PRO 특이값

트랜잭션 및 항목 SVD 그림 아래에 특이값 테이블이 나타납니다. 이 특이값은 트랜잭션 항목 행렬의 특이값 분해에서 S 행렬의 대각 항목입니다. 특이값 테이블에는 동등한 주성분 분석의 해당 고유값이 들어 있는 열이 포함되어 있습니다. 주성분 플랫폼에서와 마찬가지로 각 고유값(또는 특이값)이 설명하는 변동의 백분율 및 누적 백분율을 나타내는 열이 있습니다. "누적 백분율" 열을 사용하여 트랜잭션 항목 행렬에서 유지할 분산의 백분율을 결정한 후 해당하는 수의 특이 벡터를 사용할 수 있습니다.

JMP PRO SVD 보고서 옵션

SVD 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

SVD 산점도 행렬 항목 및 트랜잭션 특이값 분해 벡터에 대한 산점도 행렬을 표시하거나 숨깁니다. 이 옵션을 선택할 경우 산점도 행렬의 크기를 선택해야 합니다. 이 산점도 행렬을 사용하면 특이값 분해의 처음 두 차원보다 더 많은 차원을 시각화할 수 있습니다 .

주제 분석 , 회전된 SVD 항목 트랜잭션 행렬에 대한 Varimax 회전 부분 특이값 분해를 수행하여 주제라는 항목 그룹을 생성합니다 . 이 옵션을 여러 번 선택하여 서로 다른 개수의 주제를 찾을 수 있습니다 . 자세한 내용은 " 주제 분석 " 에서 확인하십시오 .

군집 항목 데이터에 있는 항목의 계층적 군집화 분석을 표시하거나 숨깁니다 . 덴드로그램 오른쪽에 다음과 같은 옵션이 있습니다 .

군집 설정 덴드로그램에서 색상이 적용되고 " 군집 저장 " 옵션에 사용되는 군집 수를 지정합니다 . 항목 SVD 그림의 점도 군집별로 색상이 적용됩니다 .

군집 저장 항목 및 해당 항목의 군집 할당을 포함하는 데이터 테이블을 저장합니다 . " 군집 저장 " 을 클릭하기 전에 군집 수를 설정해야 합니다 .

군집화 및 덴드로그램에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

군집 트랜잭션 데이터에 있는 트랜잭션의 계층적 군집화 분석을 표시하거나 숨깁니다 . 덴드로그램 오른쪽에 다음과 같은 옵션이 있습니다 .

군집 설정 덴드로그램에서 색상이 적용되고 " 군집 저장 " 옵션에 사용되는 군집 수를 지정합니다 . 트랜잭션 SVD 그림의 점도 군집별로 색상이 적용됩니다 .

군집 저장 트랜잭션 및 해당 트랜잭션의 군집 할당을 포함하는 데이터 테이블을 저장합니다 . " 군집 저장 " 을 클릭하기 전에 군집 수를 설정해야 합니다 .

군집화 및 덴드로그램에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

트랜잭션 특이 벡터 저장 트랜잭션 특이값 분해에서 사용자가 지정한 수의 특이 벡터를 새 데이터 테이블의 열로 저장합니다 . 이 데이터 테이블에서 각 행은 하나의 트랜잭션에 해당합니다 . 저장된 처음 두 열은 트랜잭션 SVD 그림에 표시된 점을 나타냅니다 . 자세한 내용은 " 특이값 분해 " 에서 확인하십시오 .

항목 특이 벡터 저장 항목 특이값 분해에서 사용자가 지정한 수의 특이 벡터를 새 데이터 테이블의 열로 저장합니다. 이 데이터 테이블에서 각 행은 하나의 항목에 해당합니다. 저장된 처음 두 열은 항목 SVD 그림에 표시된 점을 나타냅니다. 자세한 내용은 "특이값 분해"에서 확인하십시오.

JMP PRO 주제 분석

"주제 분석, 회전된 SVD" 옵션은 트랜잭션 항목 행렬의 부분 SVD(특이값 분해)에 대해 Varimax 회전을 수행합니다. 자세한 내용은 "특이값 분해"에서 확인하십시오. 트랜잭션 항목 행렬에서 유지할 주제 수에 해당하는 회전된 특이 벡터의 수를 지정해야 합니다. 주제 수를 지정하면 주제 분석 보고서가 나타납니다.

주제 분석은 회전 PCA(주성분 분석)와 동일합니다. Varimax 회전은 특이 벡터 집합을 선택하고 이를 회전시켜 더 직접적으로 좌표 방향(항목 방향)을 향하게 합니다. 이렇게 회전하면 각 회전 벡터가 항목 집합을 향하므로 벡터가 트랜잭션의 구성을 설명하는 데 도움이 됩니다. 음수 값은 반발력을 나타냅니다. 음수 값을 갖는 항목은 양수 값을 갖는 항목에 비해 주제에 나타나는 빈도가 낮습니다.

JMP PRO 주제 분석 보고서

주제 분석 보고서에서는 회전 후 각 주제에서 적재량이 가장 큰 항목을 보여 줍니다. 추가 보고서에서는 회전된 특이값 분해의 성분을 보여 줍니다.

"주제별 상위 적재" 보고서에서는 각 주제에 대한 테이블을 보여 줍니다. 각 테이블의 항목은 각 주제에 대한 적재량 절대값이 가장 큰 항목입니다. 각 테이블은 적재량 절대값을 기준으로 내림차순 정렬되어 있습니다. 이러한 테이블을 사용하여 각 주제에 해당하는 개념 테마를 판별할 수 있습니다.

주제 분석 보고서에는 다음 보고서도 포함됩니다.

주제 적재 각 항목의 주제별 적재량 행렬이 포함됩니다. 이 행렬은 회전 PCA 의 요인 적재 행렬과 동일합니다.

주제별 단어 클라우드 각 주제당 하나씩의 단어 클라우드가 포함됩니다.

주제 스코어 각 주제의 트랜잭션 스코어 행렬이 포함됩니다. 주제에서 스코어가 높은 트랜잭션은 해당 주제와 연관성이 있을 가능성이 높습니다.

주제 스코어 그림 모든 트랜잭션에 대한 주제 스코어를 1 차원 산점도에 표시합니다. 음수 값은 주제와 음의 관계를 갖는 트랜잭션을 나타냅니다. 이 그림을 사용하여 각 주제 내에서 트랜잭션의 분포를 살펴볼 수 있습니다.

각 주제에 의해 설명된 분산 각 주제에 의해 설명되는 분산의 테이블이 포함됩니다. 이 테이블에는 각 주제에 의해 설명되는 변동이 백분율 또는 누적 백분율에 대한 열도 포함됩니다.

회전 행렬 Varimax 회전에 대한 회전 행렬이 포함됩니다.

JMP PRO 주제 분석 보고서 옵션

" 주제 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

주제 산점도 행렬 회전된 특이값 분해 벡터에 대한 산점도 행렬을 표시하거나 숨깁니다 .

표시 옵션 주제 분석 보고서에 나타나는 내용을 표시하거나 숨기기 위한 옵션이 포함되어 있습니다 . 자세한 내용은 " 주제 분석 보고서 " 에서 확인하십시오 .

주제 이름 바꾸기 하나 이상의 주제에 대해 설명이 포함된 이름을 추가할 수 있는 대화상자를 표시합니다 .

트랜잭션 주제 벡터 저장 회전 특이값 분해에서 구한 특이 벡터 중 사용자가 지정한 수만큼을 데이터 테이블에 열로 저장합니다 .

항목 주제 벡터 저장 주제 벡터를 " 항목 주제 스코어 " 라는 새 데이터 테이블의 열로 저장합니다 .

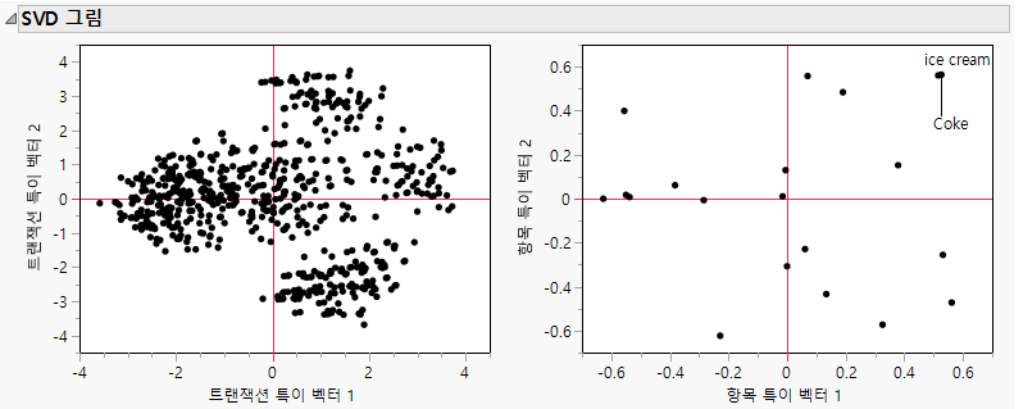
제거 SVD 보고서에서 주제 분석 보고서를 제거합니다 .

JMP PRO 추가 예 : SVD 분석

이 예에서는 Grocery Purchases.jmp 샘플 데이터에 대해 더 자세히 파악하기 위해 트랜잭션 항목 행렬의 특이값 분해를 사용합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Grocery Purchases.jmp 를 엽니다 .
2. **분석 > 선별 > 연관성 분석**을 선택합니다 .
3. Product 를 선택하고 **항목**을 클릭합니다 .
4. Customer ID 를 선택하고 **ID** 를 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 연관성 분석 " 의 빨간색 삼각형을 클릭하고 **SVD** 를 선택합니다 .
7. **확인**을 클릭합니다 .

그림 25.5 SVD 그림



트랜잭션 SVD 그림은 두 개 또는 세 개의 트랜잭션 그룹이 있을 수 있음을 나타냅니다 . 항목 SVD 그림의 오른쪽 위에 Coke(콜라) 와 ice cream(아이스크림) 을 나타내는 점이 중첩되어 있습니다 . 이 두 항목의 근접성은 두 항목 간에 강한 선호도가 있음을 나타냅니다 .

8. "SVD" 의 빨간색 삼각형을 클릭하고 **주제 분석 , 회전된 SVD** 를 선택합니다 .
9. " 주제 수 (회전된 특이 벡터)" 옆에 "3" 을 입력하고 **확인**을 클릭합니다 .

   주제 항목 및 주제 스코어 보고서가 나타납니다 .

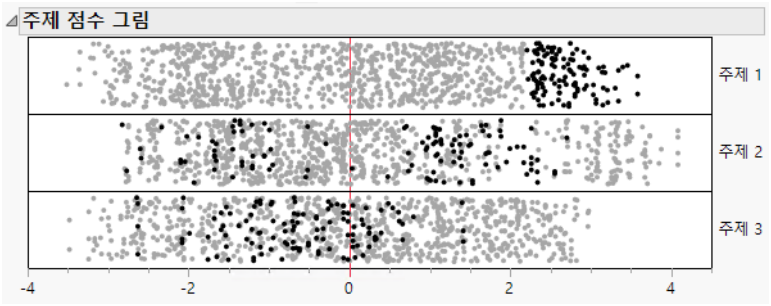
그림 25.6 주제 항목 보고서

주제별 상위 적재

주제 1		주제 2		주제 3	
항목	적재	항목	적재	항목	적재
avocado	-0.7191	Coke	0.7017	crackers	0.7050
olives	0.7142	ice cream	0.6945	apples	-0.6437
baguette	-0.6089	herring	-0.6550	soda	0.6411
turkey	0.5697	sardines	0.6122	Heineken	0.6409
artichoke	-0.4889	chicken	0.5158	sardines	-0.3618
bourbon	0.4860	corned beef	-0.4257	steak	-0.3468
Heineken	-0.4329	steak	-0.3395	corned beef	-0.3409
corned beef	0.3836	olives	-0.3035	bourbon	0.3340
Coke	0.3096	ham	-0.3022		

세 개의 그룹 또는 주제가 생성되어 주제 항목 보고서에 표시됩니다 . 주제 항목 테이블에 나열된 첫 번째 항목은 해당 그룹의 기본 항목을 나타냅니다 . 예를 들어 " 주제 1" 은 avocado(아보카도) 를 포함하지 않지만 olives(올리브) 를 포함하는 트랜잭션으로 주로 식별되는 그룹입니다 .

그림 25.7 주제 스코어



1001 개의 각 트랜잭션에 할당된 주제 스코어가 " 주세 스코어 " 보고서에 표시됩니다 . 해당 트랜잭션이 다른 주제와 어떤 관련이 있는지 보려면 주제에 대한 점 그룹을 선택합니다 . 예를 들어 주제 1 에서 값이 매우 높은 트랜잭션이 주제 2 와 3 에서는 값이 낮은 경향이 있습니다 .

10. " 특이값 " 보고서를 엽니다 .

그림 25.8 특이값 테이블

특이값

숫자	특이값	고유값	백분율	누적 백분율
1	1.7399	3.0272	15.1362	15.1362
2	1.6494	2.7204	13.6021	28.7383
3	1.5093	2.2780	11.3901	40.1284
4	1.4077	1.9816	9.9079	50.0363
5	1.2565	1.5788	7.8938	57.9301
6	1.1331	1.2839	6.4196	64.3497
7	1.0333	1.0676	5.3381	69.6877
8	0.9927	0.9854	4.9270	74.6148
9	0.9265	0.8583	4.2916	78.9063
10	0.7888	0.6223	3.1114	82.0177
11	0.7247	0.5253	2.6263	84.6440
12	0.6647	0.4419	2.2094	86.8534
13	0.6465	0.4179	2.0897	88.9431
14	0.6379	0.4069	2.0346	90.9777
15	0.6266	0.3927	1.9634	92.9411
16	0.6140	0.3770	1.8848	94.8259
17	0.6077	0.3693	1.8465	96.6724
18	0.5817	0.3383	1.6917	98.3641
19	0.5590	0.3125	1.5623	99.9264
20	0.1213	0.0147	0.0736	100.000

그림 25.8 에서 알 수 있듯이 처음 두 특이값은 식료품점 데이터의 변동을 약 30% 만 설명합니다 . 변동을 충분히 설명하려면 다른 차원이 더 필요할 수 있습니다 .

JMP PRO 연관성 분석 플랫폼에 대한 통계 상세 정보

- "빈발 항목 집합 생성"
- "연관성 분석 성능 측도"

JMP PRO 빈발 항목 집합 생성

연관성 분석 플랫폼에서는 빈발 항목 집합을 생성할 때 Apriori 알고리즘을 사용하여 계산 시간을 줄입니다. Apriori 알고리즘은 항목 집합의 지지도가 부분집합의 지지도보다 크지 않다는 사실을 활용합니다. 플랫폼에서는 최소 지지도 수준을 충족하는 더 작은 항목 집합의 조합에서 더 큰 항목 집합을 생성합니다. 또한 지정된 최대 선행 항목 수 또는 최대 규칙 크기를 초과하는 항목 집합을 생성하지 않습니다. 가능한 총 규칙 수는 항목 수에 따라 기하급수적으로 증가하므로 대규모 데이터 집합으로 작업할 때 이러한 옵션이 유용합니다. Apriori 알고리즘에 대한 자세한 내용은 Agrawal and Srikant(1994) 연구 자료에서 확인하십시오.

JMP PRO 연관성 분석 성능 측도

이 섹션에서는 연관성 분석 플랫폼에 사용되는 성능 측도를 정의합니다. 조건 항목 집합 X와 결과 항목 집합 Y를 사용하는 연관 규칙을 $X \Rightarrow Y$로 나타냅니다. Hahsler(2015)에는 지지도, 신뢰도 및 향상도를 비롯하여 연관성 분석에 사용되는 성능 측도 모음이 포함되어 있습니다.

JMP PRO 지지도

지지도는 항목 집합이 발생하는 트랜잭션의 비율입니다. 지지도를 트랜잭션에 항목 집합이 포함될 확률로 볼 수도 있습니다.

조건 항목 집합 X의 지지도 S는 다음과 같이 정의됩니다.

$$supp(X) = \frac{N_X}{N} = Pr(X)$$

다음은 각 요소에 대한 설명입니다.

N_X = 항목 집합 X를 포함하는 트랜잭션의 수

N = 총 트랜잭션 수

연관 규칙의 지지도는 다음과 같이 정의됩니다.

$$supp(X \Rightarrow Y) = supp(X \cup Y) = Pr(X \cap Y)$$

이 경우 지지도는 트랜잭션에 항목 집합 X와 Y가 모두 포함될 확률과 동일합니다.

항목 집합과 연관 규칙의 경우 모두 지지도 범위는 0% ~ 100% 사이입니다.

JMP PRO 신뢰도

신뢰도는 트랜잭션에 조건 항목 집합이 있을 때 결과 항목 집합을 포함하는 트랜잭션의 비율입니다. 신뢰도를 트랜잭션에 조건 항목 집합이 있을 때 해당 트랜잭션에 결과 항목 집합이 포함될 조건부 확률로 볼 수도 있습니다.

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} = \frac{Pr(X \cap Y)}{Pr(X)} = Pr(Y|X)$$

신뢰도 범위는 0% ~ 100% 사이입니다. 신뢰도가 0%인 연관 규칙의 결과 항목 집합은 조건 항목 집합이 있는 모든 트랜잭션에 나타나지 않습니다. 신뢰도가 100%이면 조건 항목 집합을 포함하는 모든 트랜잭션에 결과 항목 집합도 포함된다는 것을 나타냅니다.

참고 : 연관성 분석의 신뢰도는 신뢰 구간의 개념과 관련이 없습니다.

JMP PRO 향상도

향상도는 X 와 Y 간의 종속성을 측정합니다.

$$lift(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) \times supp(Y)} = \frac{Pr(X \cap Y)}{Pr(X)Pr(Y)}$$

향상도의 분자는 X와 Y가 공동으로 발생하는 트랜잭션의 비율입니다. 분모는 X와 Y가 독립적으로 발생한다고 가정할 때 예상되는 결합 발생의 추정값입니다.

향상도 범위는 0 ~ ∞ 사이입니다. 향상도 값이 1이면 X와 Y가 트랜잭션에서 공동으로 발생하는 빈도가 단지 우연히 발생할 가능성과 같음을 나타냅니다. 향상도 값이 증가하면 X가 있을 때 Y가 예상보다 자주 발생한다는 것을 나타냅니다.

참고 : $X \Rightarrow Y$ 연관 규칙의 향상도는 $Y \Rightarrow X$ 연관 규칙의 향상도와 같습니다.

26 장

공정 기록 탐색기
복잡한 공정 기록에서 문제 성분 식별

참고 : 공정 기록 탐색기는 새로운 시험용 플랫폼입니다. 향후 릴리스에서 변경될 수 있습니다.

반도체 공장과 같은 복잡한 제조 환경에서 도구 또는 경로 조합에 결함이 있으면 수율과 품질이 저하될 수 있습니다. 그러나 하나의 유닛에 서로 다른 여러 도구를 포함하는 수백 개의 단계를 거쳐야 할 수 있습니다. 낮은 수율의 원인이 되는 정확한 요인 조합을 판별하기 어려울 수 있습니다.

공정 기록 탐색기 플랫폼에서는 모든 생산 추적 데이터를 정리하고 낮은 수율과 관련된 것으로 보이는 요인을 식별할 수 있습니다. 그런 다음 단계별 회귀를 수행할 수 있습니다. 공정에서 단계 또는 도구 사이의 유닛 대기 시간을 분석하고 이러한 대기 시간이 수율과 어떤 관련이 있는지 확인할 수 있습니다. 또한 다른 도구로의 전환이 수율에 미치는 영향을 분석할 수도 있습니다.

그림 26.1 공정 기록 탐색기 보고서

공정 기록 탐색기

1250개의 유닛에서 다음 ID 열 사용: Lot, Wafer
125개의 단계에서 다음 열 사용: Layer, Operation
156250번의 작업에서 다음 시간 열 사용: TimeIn, TimeOut
195개의 수준이 다음 X 열에 있음: Tool, Route
1 반응: Yield

반응 Yield

최저 수율 수준

항	수준	개수	평균	표준 오차	표준편차
Tool	PlanarTool008	1000	0.75299	0.00188	0.05952
Tool	DepTool003	1375	0.75577	0.00168	0.06245
Tool	EtchTool019	750	0.76192	0.0021	0.05749
Tool	EtchTool012	500	0.76203	0.0026	0.05807
Tool	EtchTool034	850	0.76279	0.00199	0.0581
Tool	PlanarTool021	1050	0.7637	0.00186	0.06028
Tool	PlanarTool030	1025	0.76539	0.00184	0.05894
Tool	EtchTool016	900	0.76566	0.00191	0.05734
Tool	EtchTool009	925	0.76598	0.00197	0.05994
Tool	Dep2Tool018	1525	0.76766	0.00156	0.06088
Tool	Dep2Tool014	1825	0.76813	0.00138	0.05893
Tool	DepTool019	1650	0.76839	0.00143	0.05794
Tool	EtchTool020	1000	0.76909	0.00177	0.05586
Tool	Dep2Tool015	1475	0.76948	0.0015	0.05775
Tool	PhotoTool010	3525	0.76996	0.00098	0.05822
Tool	Dep2Tool013	1825	0.7705	0.00141	0.06025
Tool	DepTool012	1750	0.77063	0.00145	0.06051
Tool	Dep2Tool006	1400	0.77063	0.00158	0.05905
Tool	PlanarTool028	900	0.77103	0.00202	0.06071
Tool	Dep2Tool011	1600	0.77124	0.00145	0.05797

목차

공정 기록 탐색기의 예 493
공정 기록 탐색기 플랫폼 시작 495
데이터 형식 496
공정 기록 탐색기 보고서 496
공정 기록 탐색기 플랫폼 옵션 496

공정 기록 탐색기의 예

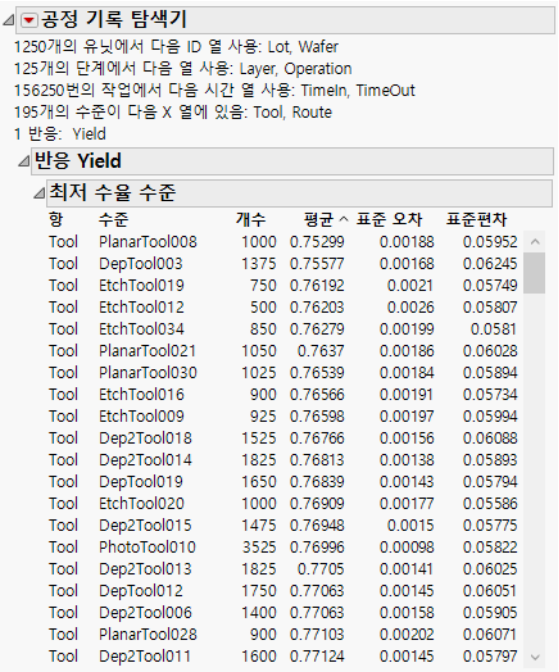
이 예에서는 두 개의 데이터 테이블을 사용합니다 . 첫 번째 데이터 테이블에는 복잡한 제조 공정을 거친 50 개 로트의 25 개 웨이퍼에 대한 단계별 기록이 포함되어 있습니다 . 두 번째 데이터 테이블에는 각 로트의 각 웨이퍼에 대한 수율이 포함됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control > Lot Wafer History.jmp 및 Quality Control > Lot Wafer Yield.jmp 를 엽니다 .
2. Lot Wafer History.jmp 데이터 테이블에서 **분석 > 선별 > 공정 기록 탐색기**를 선택합니다 .
3. Lot 와 Wafer 를 선택하고 **ID** 를 클릭합니다 .
4. Tool 과 Route 를 선택하고 **X, 공정**을 클릭합니다 .
5. Layer 와 Operation 을 선택하고 **단계**를 클릭합니다 .
6. TimeIn 과 TimeOut 을 선택하고 **타임스탬프**를 클릭합니다 .
7. **확인**을 클릭합니다 .

   해당하는 수율 테이블을 선택할 수 있는 창이 나타납니다 .
8. Lot Wafer Yield 를 선택하고 **확인**을 클릭합니다 .

   수율 열을 선택할 수 있는 창이 나타납니다 .
9. Yield 를 선택하고 **확인**을 클릭합니다 .
10. " 공정 기록 탐색기 " 의 빨간색 삼각형을 클릭하고 **최저 수율 수준**을 선택합니다 .

그림 26.2 공정 기록 탐색기 보고서



공정 기록 탐색기

1250개의 유닛에서 다음 ID 열 사용: Lot, Wafer
125개의 단계에서 다음 열 사용: Layer, Operation
156250번의 작업에서 다음 시간 열 사용: TimeIn, TimeOut
195개의 수준이 다음 X 열에 있음: Tool, Route
1 반응: Yield

반응 Yield

최저 수율 수준

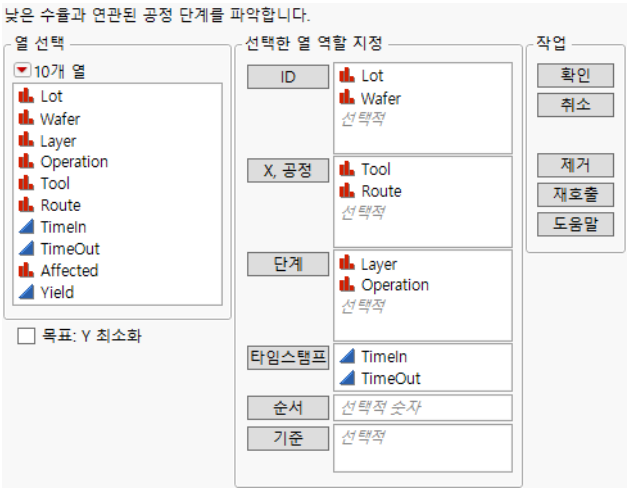
항	수준	개수	평균	표준 오차	표준편차
Tool	PlanarTool008	1000	0.75299	0.00188	0.05952
Tool	DepTool003	1375	0.75577	0.00168	0.06245
Tool	EtchTool019	750	0.76192	0.0021	0.05749
Tool	EtchTool012	500	0.76203	0.0026	0.05807
Tool	EtchTool034	850	0.76279	0.00199	0.0581
Tool	PlanarTool021	1050	0.7637	0.00186	0.06028
Tool	PlanarTool030	1025	0.76539	0.00184	0.05894
Tool	EtchTool016	900	0.76566	0.00191	0.05734
Tool	EtchTool009	925	0.76598	0.00197	0.05994
Tool	Dep2Tool018	1525	0.76766	0.00156	0.06088
Tool	Dep2Tool014	1825	0.76813	0.00138	0.05893
Tool	DepTool019	1650	0.76839	0.00143	0.05794
Tool	EtchTool020	1000	0.76909	0.00177	0.05586
Tool	Dep2Tool015	1475	0.76948	0.0015	0.05775
Tool	PhotoTool010	3525	0.76996	0.00098	0.05822
Tool	Dep2Tool013	1825	0.7705	0.00141	0.06025
Tool	DepTool012	1750	0.77063	0.00145	0.06051
Tool	Dep2Tool006	1400	0.77063	0.00158	0.05905
Tool	PlanarTool028	900	0.77103	0.00202	0.06071
Tool	Dep2Tool011	1600	0.77124	0.00145	0.05797

이 보고서에는 유닛, 단계 및 작업 수와 X 열의 수준 수를 비롯하여 공정에 대한 정보가 포함됩니다. "최저 수율 수준" 테이블에는 X 열의 195개 수준에 대한 개수 및 수율 정보가 포함되어 있습니다. PlanarTool008 도구의 수율이 가장 낮고 1,000개의 유닛이 해당 도구를 거쳤습니다.

공정 기록 탐색기 플랫폼 시작

분석 > 선별 > 공정 기록 탐색기를 선택하여 공정 기록 탐색기 플랫폼을 시작합니다 .

그림 26.3 공정 기록 탐색기 시작



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

시작 창 역할

ID 공정의 유닛을 식별하는 하나 이상의 열을 지정합니다 .

> **참고 :** ID 열은 수율 테이블에 해당 열이 있어야 합니다 .

X, 공정 공정 단계를 구성하는 성분 또는 특정 상세 정보를 식별하는 하나 이상의 열을 지정합니다 .

단계 유닛이 거쳐 가는 단계 및 하위 단계를 지정하는 선택적 열입니다 . 예를 들어 Layer 와 해당 Layer 내의 Operation 유형입니다 .

타임스탬프 작업 시간을 지정합니다 . 시간은 단일 시점으로 지정하거나 , 작업 시작 및 종료 시간을 나타내는 한 쌍의 시점으로 지정할 수 있습니다 .

순서 유닛 배치가 "X, 공정" 역할에 지정된 성분을 거쳐 가는 순서를 지정하는 선택적 열입니다 .

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

시작 창 옵션

목표 : Y 최소화 분석 목표가 반응을 최소화하는 것임을 지정합니다 . 기본적으로 분석은 반응을 최대화하려고 합니다 .

참고 : 반응은 플랫폼을 시작할 때 사용된 데이터 테이블이 아니라 두 번째 데이터 테이블에 지정됩니다 .

데이터 형식

공정 기록 탐색기에 사용할 데이터는 두 개의 데이터 테이블에 지정되어야 합니다. 첫 번째 테이블은 플랫폼을 시작하는 데 사용되며 공정의 단계 및 상세 정보를 포함합니다. 두 번째 테이블은 ID 열의 각 수준에 대한 반응 또는 수율을 포함합니다.

Quality Control > Lot Wafer History 및 Quality Control > Lot Wafer Yield 샘플 데이터 테이블은 데이터가 두 데이터 테이블에 지정되는 방식을 보여 주는 예입니다.

공정 기록 탐색기 보고서

초기 " 공정 기록 탐색기 " 보고서에서는 데이터 테이블의 정보를 요약하여 보여 줍니다 .

- ID 열에 의해 결정되는 유닛 수 . ID 열이 두 개 이상인 경우 고유 조합의 수입니다 .
- 단계 열에 의해 결정되는 단계 수 . 단계 열이 두 개 이상인 경우 고유 조합의 수입니다 .
- 타임스탬프 열에 의해 결정되는 작업 수 . 이 값은 데이터 테이블의 행 수와 같습니다 .
- X, 공정 열에 의해 결정되는 수준 수 . X, 공정 열이 두 개 이상인 경우 이 값은 각 열에 있는 수준 수의 합입니다 .
- Yield 열에 의해 결정되는 반응 수

"공정 기록 탐색기"의 빨간색 삼각형 메뉴에서 옵션을 선택하여 분석을 계속합니다. 자세한 내용은 "공정 기록 탐색기 플랫폼 옵션"에서 확인하십시오.

공정 기록 탐색기 플랫폼 옵션

" 공정 기록 탐색기 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

최저 수율 수준 각 X, 공정 변수의 각 수준에 대한 수율 요약 통계량 테이블을 표시합니다 . 각 수준에 대해 제공되는 요약 통계량은 작업 수와 수율의 평균 , 표준 오차 및 표준편차입니다 . 기본적으로 이 테이블은 평균 수율에 따라 오름차순으로 정렬됩니다 .

시간 필터를 사용한 최저 수율 수준 수율이 가장 낮은 상위 50 개 수준의 시간 구간에 대한 요약 통계량 테이블을 표시합니다 . 이 루틴은 구간 내 평균 수율을 최소화하는 " 최저 수율 수준 " 테이블에서 상위 50 개의 각 수준에 대한 시간 구간을 검색합니다 . 각 구간은 해당 수준의 작업

을 25% 이상 포함해야 합니다 . 각 수준에 대해 제공되는 요약 통계량은 식별된 구간의 작업 수 , 식별된 구간의 평균 수율 , 수준의 첫 번째 작업과 마지막 작업에 대한 타임스탬프 (첫 번째 시간 , 마지막 시간), 식별된 구간의 첫 번째 작업과 마지막 작업에 대한 타임스탬프 (런시작 시간 , 런종료 시간) 입니다 .

단계별 회귀 단계별 회귀 분석을 위한 옵션을 지정할 수 있는 " 단계별 회귀 " 창을 표시합니다 . 단계별 회귀를 사용하여 공정에서 수율 감소의 원인이 되는 문제를 찾을 수 있습니다 . 각 회귀 단계에서 음수 수율이 가장 큰 항에 따라 항이 입력됩니다 . 이렇게 하면 수율에 가장 큰 영향을 미치는 공정 단계를 식별하는 데 도움이 됩니다 . 단계별 회귀는 각 회귀 단계에서 항이 모형에 입력된 순서를 보여 주는 테이블을 제공합니다 .

참고 : 시작 창에서 " 목표 : Y 최소화 " 옵션을 선택한 경우에는 단계별 회귀가 수율 증가의 원인이 되는 항을 감지하도록 조정됩니다 .

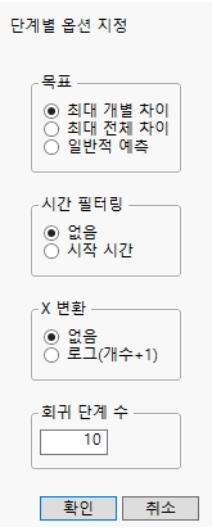
목표 단계별 회귀 모형에 사용되는 기준을 지정합니다 . " 최대 개별 차이 " 옵션은 개별적으로 수율을 가장 많이 변경시키는 항을 식별합니다 . " 최대 전체 차이 " 옵션은 전체 차이 (차이에 영향을 받는 유닛 수를 곱한 값) 를 가장 크게 만드는 항을 식별합니다 . " 일반적 예측 " 옵션은 기존의 단계별 회귀 모형을 생성합니다 .

시간 필터링 단계별 회귀 모형에서 성분의 시작 시간을 사용할 수 있습니다 . " 없음 " 옵션을 선택하면 전체 기간에 대한 평균이 시간으로 사용됩니다 . " 시작 시간 " 옵션을 선택하면 특정 시작 시간에 성분이 실패할 수 있습니다 . 시작 시간은 음수 수율 효과를 최대화하는 시간입니다 . " 시작 시간 " 옵션을 선택하면 " 단계별 회귀 " 보고서 테이블에 " 시작 시간 " 열이 포함됩니다 .

X 변환 단계별 회귀에서 X, 공정 변수에 대해 일반 개수 데이터 또는 로그 (개수 +1) 데이터를 사용할 수 있습니다 . " 로그 (개수 +1)" 옵션은 빈도 수가 높으면 회귀 분석에서 선형 영향력이 적도록 회귀 분석의 변수를 변환합니다 . " 없음 " 옵션을 선택하면 높은 빈도 수가 회귀 분석에서 선형 영향력을 가집니다 .

회귀 단계 수 단계별 회귀 분석의 단계 수입니다 . 기본값은 10 입니다 .

그림 26.4 단계별 회귀 창



대기 시간 분석 (타임스탬프 변수를 두 개 지정한 경우에만 사용 가능) 각 고유 ID 는 한 번에 하나의 단계 수준만 진행할 수 있습니다. 또한 각 단계는 한 번에 하나의 ID 만 처리할 수 있습니다. 대기 시간 분석은 각 ID 에 대한 대기 시간 요약 통계량 테이블을 제공합니다. 이 테이블에서 대기 시간은 ID 가 한 단계에서 완료된 후 다른 단계를 시작할 때까지의 시간입니다. 다른 대기 시간 분석 테이블도 제공됩니다. 이러한 테이블에는 각 단계 변수의 수준 내에서 각 ID 에 대한 대기 시간 요약 통계량이 포함됩니다. 각 테이블에서 대기 시간은 ID 가 단계 변수의 특정 수준에서 완료된 후 동일한 수준을 다시 진행할 때까지의 시간입니다. 이 테이블에는 다음 통계량이 제공됩니다.

대기 시간 합계 주어진 ID 및 단계 변수 수준 조합에 대한 대기 시간의 합입니다.

대기 시간 개수 주어진 ID 및 단계 변수 수준 조합에 대한 대기 시간의 수입니다.

평균 대기 시간 주어진 ID 및 단계 변수 수준 조합에 대한 대기 시간의 평균입니다.

최대 대기 시간 주어진 ID 및 단계 변수 수준 조합에 대한 최대 대기 시간입니다.

전환 분석 결과 수율에 문제가 되는 성분 간 전환을 식별하는 데 유용한 보고서를 표시합니다. 이 보고서는 공정에서 정렬 및 등록 문제를 식별하는 데 도움이 될 수 있습니다. 예를 들어 공정 열의 일부 수준이 해당 공정 열의 다른 수준에 대해 정렬되지 않았을 수 있습니다. 전환 문제가 있으면 특정 수준을 반복하는 항목이 공정 열의 여러 수준을 거치는 항목보다 수율이 더 높을 것으로 기대할 수 있습니다.

이 옵션을 선택하면 분석에 포함할 하나 이상의 X, 공정 열과 각 열에서 하나 이상의 수준을 선택해야 합니다. 분석을 위해 지정한 각 열에 대해 네 개의 테이블이 포함된 보고서가 나타납니다. 첫 번째 테이블에는 열의 특정 수준에서 열의 다른 특정 수준으로 전환하는 경우의 수율과 개수가 표시됩니다. 두 번째 테이블에는 열의 특정 수준에서 외부로 전환하는 경우의 수율과 개수가 표시됩니다. 세 번째 테이블에는 열의 특정 수준으로 전환하는 경우의 수율과

개수가 표시됩니다. 네 번째 테이블에는 유닛이 열의 특정 수준을 반복하는 전환에 대한 수율과 개수가 표시됩니다.

순서 교호작용 (시작 창에서 순서 변수를 지정한 경우에만 사용 가능) 순서 열을 기준으로 수율에 대한 일원 분석 보고서를 표시합니다. 각 X, 공정 열에 대한 테이블이 보고서에 포함됩니다. 각 테이블에서 공정 열의 수준이 해당 분산 분석 모형의 LogWorth 값에 따라 내림차순으로 나열되어 있습니다. 이 테이블에는 공정 열의 각 수준에 대한 p 값과 히트맵도 포함됩니다.

개수 테이블 저장 X, 공정 변수의 각 수준을 거친 유닛 수가 포함된 새 데이터 테이블을 생성합니다. 개수 테이블에는 ID 값에 의해 결정되는 각 고유 유닛에 대한 행이 있습니다.

로그 개수 테이블 저장 로그 (개수 +1) 항목을 제외하고 개수 테이블과 동일한 형식의 새 데이터 테이블을 생성합니다.

대기 시간 저장 (타임스탬프 변수를 두 개 지정한 경우에만 사용 가능) 대기 시간 분석 보고서의 각 테이블을 별도의 데이터 테이블에 저장합니다.

부록 A

통계 상세 정보
예측 및 전문 모델링

이 부록에서는 각 플랫폼에서 사용할 수 있는 적합 측도 및 교차 검증 유형에 대해 설명합니다.

목차

적합 측도 503
엔트로피 R^2 503
JMP 모델링의 검증 503
검증 열 역할 504
K 폴드 및 홀드백 검증 506

적합 측도

엔트로피 R²

엔트로피 R² 은 훈련 데이터 집합과 검증 및 테스트 데이터 집합 (검증이 사용되는 경우) 에 대해 계산됩니다 . 훈련 데이터 집합에 대한 엔트로피 R² 은 다음과 같이 계산됩니다 .

- 훈련 데이터 집합을 사용하여 지정된 모형이 적합됩니다 .
- 모형을 기반으로 반응의 각 수준에 대한 예측 확률이 계산됩니다 .
- 이러한 예측 확률을 사용하여 훈련 데이터 집합의 관측값에 대한 가능도가 계산됩니다 . 이를 $Likelihood_Full_{Training}$ 이라고 합니다 .
- 훈련 데이터 집합을 사용하여 축소 모형 (예측 변수 없음) 이 적합됩니다 .
- 축소 모형의 반응 수준에 대한 예측 확률을 사용하여 훈련 데이터 집합의 관측값에 대한 가능도가 계산됩니다 . 이를 $Likelihood_Reduced_{Training}$ 이라고 합니다 .
- 훈련 데이터 집합에 대한 엔트로피 R² 은 다음과 같습니다 .

$$\text{엔트로피 } R^2_{Training} = 1 - \frac{\log(Likelihood_Full_{Training})}{\log(Likelihood_Reduced_{Training})}$$

검증 데이터 집합 및 테스트 데이터 집합의 엔트로피 R² 은 훈련 데이터 집합의 엔트로피 R² 과 유사한 방식으로 계산됩니다 .

JMP 모델링의 검증

과대적합을 방지하고 양호한 예측 모형을 선택하기 위해 모델링 전에 데이터를 여러 데이터 집합으로 분할할 수 있습니다 . 이 프로세스는 원래 데이터의 일부를 사용하여 모수를 추정하고 나머지 데이터를 사용하여 모수를 조정 또는 평가하거나 둘 다 수행합니다 . JMP Pro 에서는 다음과 같은 방법을 사용하여 데이터를 두 개 또는 세 개의 집합으로 분할할 수 있습니다 .

훈련 및 평가 데이터를 훈련 데이터 집합과 검증 데이터 집합이라는 두 집합으로 분할합니다 . 훈련 데이터 집합은 모형 모수를 추정하는 데 사용됩니다 . 검증 데이터 집합은 적합 모형의 성능을 독립적으로 평가하는 데 사용됩니다 .

훈련 및 조정 데이터를 훈련 데이터 집합과 검증 데이터 집합이라는 두 집합으로 분할합니다 . 훈련 데이터 집합은 모형 모수를 추정하는 데 사용됩니다 . 검증 데이터 집합은 모형 적합 알고리즘에서 모형 모수를 조정하고 궁극적으로 예측 능력이 양호한 모형을 선택하는 데 사용됩니다 . 이 경우에는 독립적인 모형 평가가 수행되지 않습니다 .

훈련 , 조정 및 평가 데이터를 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합이라는 세 집합으로 분할합니다 . 훈련 데이터 집합은 모형 모수를 추정하는 데 사용됩니다 . 검증 데이터 집합은 모형 적합 알고리즘에서 모형 모수를 조정하고 궁극적으로 예측 능력이 양호한 모

형을 선택하는 데 사용됩니다. 그런 다음 테스트 데이터 집합을 사용하여 적합 모형의 성능을 독립적으로 평가합니다.

JMP PRO 검증 열 역할

검증 열 역할은 JMP Pro에서만 사용할 수 있습니다. JMP의 경우 "제외된 행을 검증 홀드백으로 사용"의 내용을 참조하십시오.

데이터 파티션을 생성하는 한 가지 방법은 검증 열 역할을 사용하는 것입니다. 검증 열 역할은 해당 열의 값을 사용하여 데이터를 여러 부분으로 나눕니다. 이 열은 플랫폼 시작 창에서 "검증" 역할을 사용하여 할당합니다. 검증 열을 생성하는 방법에 대한 자세한 내용은 "검증 열 생성"에서 확인하십시오.

주의 : 검증 열 사용은 플랫폼에 따라 다릅니다. 플랫폼마다 검증 열의 수준을 다르게 사용합니다. 표 A.1 의 " 참고 " 내용을 참조하십시오.

표 A.1 플랫폼별 검증 열

플랫폼	훈련 및 평가	훈련 및 조정	훈련, 조정 및 평가	참고
모형 적합				
최소 제곱 적합	예	아니요	아니요	수준 수가 3개를 초과하면 검증 열이 무시됩니다.
단계별 회귀	아니요	예	예	수준 수가 3개를 초과하면 K 폴드 교차 검증이 사용됩니다.
로지스틱 회귀	예	아니요	아니요	수준 수가 3개를 초과하면 검증 열이 무시됩니다.
일반화 회귀	아니요	예	예	수준 수가 3개를 초과하면 K 폴드 교차 검증이 사용됩니다.
부분 최소 제곱	아니요	예	예	수준 수가 3개를 초과하면 K 폴드 교차 검증이 사용됩니다.
예측 모형				
신경망	아니요	예	예	수준 수가 3개를 초과하면 K 폴드 교차 검증이 사용됩니다.
파티션	아니요	예	예	수준 수가 3개를 초과하면 가장 작은 3개의 값을 포함하는 행만 사용됩니다.

플랫폼	훈련 및 평가	훈련 및 조정	훈련 , 조정 및 평가	참고
붓스트랩 포레스트	아니요	예	예	수준 수가 3 개를 초과하면 가장 작은 3 개의 값을 포함하는 행만 사용됩니다.
부스티드 트리	아니요	예	예	수준 수가 3 개를 초과하면 가장 작은 3 개의 값을 포함하는 행만 사용됩니다.
K 최근접 이웃	아니요	예	예	수준 수가 3 개를 초과하면 가장 작은 3 개의 값을 포함하는 행만 사용됩니다.
Naive Bayes	아니요	예	예	수준 수가 3 개를 초과하면 가장 작은 3 개의 값을 포함하는 행만 사용됩니다.
서포트 벡터 머신	예	예	예	수준 수가 3 개를 초과하면 K 폴드 교차 검증이 사용됩니다.
전문 모형				
함수 데이터 탐색기	예	아니요	아니요	그룹별 랜덤 검증 열로 생성해야 합니다. 수준 수가 2 개를 초과하면 가장 작은 값으로 훈련 데이터 집합이 정의되고 다른 모든 값으로 검증 데이터 집합이 정의됩니다.
다변량 모형				
판별	예	예	예	수준 수가 3 개를 초과하면 가장 작은 3 개의 값을 포함하는 행만 사용됩니다.
부분 최소 제곱	아니요	예	예	수준 수가 3 개를 초과하면 K 폴드 교차 검증이 사용됩니다.
Uplift	아니요	예	예	수준 수가 3 개를 초과하면 가장 작은 3 개의 값을 포함하는 행만 사용됩니다.

K 폴드 및 홀드백 검증

검증 열 역할은 데이터를 교차 검증 데이터 집합으로 분할하기 위한 프레임워크를 제공합니다. 또한 일부 JMP 플랫폼에서는 K 폴드 및 다양한 유형의 홀드백 검증을 지원합니다.

K 폴드 교차 검증 원래 데이터를 *K* 개의 부분집합으로 나눕니다. 그러면 *K* 개의 각 집합이 나머지 데이터에 대한 모형 적합을 검증하는 데 사용되며 총 *K* 개 모형을 적합시킵니다. 최량 검증 통계량을 생성하는 모형이 최종 모형으로 선택되고, 해당 모형을 생성하는 데 사용되지 않은 폴드는 테스트 데이터 집합 성능 통계량을 제공합니다.

참고 : 일부 플랫폼의 경우 모형 제어판에서 K 폴드 교차 검증을 지정해야 합니다. 다른 플랫폼의 경우 플랫폼 시작 창에서 K 폴드 교차 검증을 지정해야 합니다. 수준이 네 개 이상 포함된 검증 열을 통해 K 폴드 교차 검증을 지정해야 하는 플랫폼도 있습니다.

랜덤 검증 홀드백 (특정 플랫폼의 시작 옵션으로 사용 가능) 원래 데이터를 무작위로 훈련 데이터 집합과 검증 데이터 집합으로 나눕니다. 테스트 데이터 집합을 포함할 수도 있습니다. 원래 데이터 중 각 데이터 집합에 사용할 데이터의 비율을 지정할 수 있습니다.

Leave-One-Out 검증 홀드백 (특정 플랫폼의 옵션으로 사용 가능) 관측값을 한 번에 하나씩 제외하면서 모형을 반복적으로 적합시킵니다. Leave-One-Out 검증은 잭나이프 절차라고도 합니다.

제외된 행을 검증 홀드백으로 사용 데이터 테이블에서 제외된 행을 검증 홀드백 집합으로 사용합니다. JMP Pro 의 경우 플랫폼 환경 설정에서 이 옵션을 선택하여 사용할 수 있습니다.

참고 : 제외된 행을 검증 홀드백 집합으로 사용할 수 있는 플랫폼에서는 시작 창에서 검증 열 또는 검증 비율을 지정하지 않은 경우에만 제외된 행이 사용됩니다.

표 A.2 플랫폼별 K 폴드 및 홀드백 검증

플랫폼	제외된 행을 검증 홀드백으로 사용	랜덤 검증 홀드백	Leave-One-Out 홀드백	K 폴드 교차 검증
모형 적합				
최소 제곱 적합	아니요	아니요	아니요	아니요
단계별 회귀	아니요	아니요	아니요	예(연속형 반응 모형의 경우만 해당)
로지스틱 회귀	아니요	아니요	아니요	아니요
JMP PRO 일반화 회귀	아니요	예	예	예(모형 제어판을 통해)

플랫폼	제외된 행을 검증 홀드백으로 사용	랜덤 검증 홀드백	Leave-One-Out 홀드백	K 폴드 교차 검증
JMP PRO 부분 최소 제곱	아니요	예	예	예(모형 제어판을 통해)
예측 모형				
신경망	예	예	아니요	예(모형 시작 창 또는 검증 열을 통해)
파티션	예	예	아니요	예(플랫폼 환경 설정에서 옵션 선택)
JMP PRO 붓스트랩 포레스트	예	예	아니요	아니요
JMP PRO 부스티드 트리	예	예	아니요	아니요
JMP PRO K 최근접 이웃	예	예	아니요	아니요
JMP PRO Naive Bayes	예	예	아니요	아니요
JMP PRO 서포트 벡터 머신	아니요	예	아니요	예(모형 시작 창을 통해)
전문 모형				
JMP PRO 함수 데이터 탐색기	아니요	아니요	아니요	아니요
다변량 모형				
판별	선택적	아니요	아니요	아니요
부분 최소 제곱	아니요	예	예	예(모형 시작 창 또는 검증 열을 통해)
JMP PRO Uplift	아니요	예		아니요

부록 B

참조 자료

Agrawal, R., and Srikant, R. (1994). "Fast Algorithms for Mining Association Rules." In *Proceedings of the 20th VLDB Conference*. Santiago, Chile: IBM Almaden Research Center. Accessed July 5, 2016. https://web.stanford.edu/class/cs345d-01/rl/ar-mining.pdf.

Baglama, J., and Reichel, L. (2005). "Augmented implicitly restarted Lanczos bidiagonalization methods." *SIAM Journal on Scientific Computing*, 27:19–42.

Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: John Wiley & Sons.

Benford, F. (1938). "The law of anomalous numbers." *Proceedings of the American philosophical society*, 551–572.

Benjamini, Y., and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society*, *Series B* 57:289–300.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

Candes, E. J., Li, X., Ma, Y., and Wright, J. (2009). "Robust Principal Component Analysis?" *Journal of the ACM*, 58:1–37.

Cleveland, W. S. (1994). *Visualizing Data*, Summit, NJ: Hobart Press.

Conover, W. J. (1999). *Practical Nonparametric Statistics*. 3rd ed. New York: John Wiley & Sons.

Cureton, E. E. (1967). "The Normal Approximation to the Signed-Rank Sampling Distribution when Zero Differences are Present." *Journal of the American Statistical Association* 62: 319, 1068–1069.

Hahsler, M. (2015). "A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules." Accessed September 17, 2018. http://michael.hahsler.net/research/association_rules/measures.html.

Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*. Cambridge, MA: MIT Press.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer-Verlag.

Hawkins D. M., and Kass G. V. (1982). "Automatic Interaction Detection." In *Topics in Applied Multivariate Analysis*, edited by D. M. Hawkins, 267–300. Cambridge: Cambridge University Press.

Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics*. 2nd ed. New York: John Wiley & Sons.

Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer-Verlag.

Jolliffe, I. T. (2002). *Principal Component Analysis*. New York: Springer-Verlag.

Kass, G. V. (1980). "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Journal of the Royal Statistical Society, Series C* 29:119–127

Lehman, E. L. (2006). *Nonparametrics: Statistical Methods Based on Ranks.* 2nd ed. New York: Springer.

Lin, Z., Chen, M., and Ma, Y. (2013). The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055.

Mason, R. L., and Young, J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications*. Philadelphia: SIAM.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models.* 2nd ed. London: Chapman & Hall.

Nagelkerke, N. J. D. (1991). "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78:691–692.

Nelder, J. A., and Wedderburn, R. W. M. (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society, Series A* 135:370–384.

Parker, R. J. (2015). *Efficient Computational Methods for Large Spatial Data*. Ph.D. diss., Department of Statistics, North Carolina State University. Accessed June 30, 2016. https://repository.lib.ncsu.edu/ir/bitstream/1840.16/10572/1/etd.pdf.

Platt, J. (1998). *Sequential minimal optimization: A fast algorithm for training support vector machines*. Technical Report MST-TR-98-14, Microsoft Research. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-98-14.pdf.

Qian, P. Z., Huaiquing, W., and Wu, C. F. (2012). "Gaussian process models for computer experiments with qualitative and quantitative factors." *Technometrics* 50:383–396.

Ramsay, J. O., and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd ed. New York: Springer.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models*. New York: Marcel Dekker.

Sall, J. (2002). "Monte Carlo Calibration of Distributions of Partition Statistics." SAS Institute Inc.,Cary, NC. Accessed July 29, 2015. https://www.jmp.com/content/dam/jmp/documents/en/white-papers/montecarlocal.pdf.

Santer, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.

SAS Institute Inc. (2020). *SAS/ETS 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/etsug/15.2/content/etsug.pdf.

Schäfer, J., and Strimmer, K. (2005). "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics." *Statistical Applications in Genetics and Molecular Biology* 4, Article 32.

Schuirmann, D. J. (1987). "A Comparison of the Two One-sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability." *Journal of Pharmacokinetics and Biopharmaceutics* 15:657–680.

Shiskin, J., Young, A. H., and Musgrave, J. C. (1967). *The X-11 Variant of the Census Method II Seasonal Adjustment Program.* Technical Report 15, US Department of Commerce, Bureau of the Census.

Shmueli, G., Patel, N. R., and Bruce, P. C. (2010). *Data Mining For Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Shmueli, G., Bruce, P. C., Stephens M. L., and Patel, N. R. (2017). *Data Mining For Business Intelligence: Concepts, Techniques, and Applications with JMP Pro*. Hoboken, NJ: John Wiley & Sons.

Westfall, P. H., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple Comparisons and Multiple Tests Using SAS*. 2nd ed. Cary, NC: SAS Institute Inc.

White, K., Szarka, J., and Jensen, W. (2018). "A Recommended Set of Indices for Evaluating Process Health." Presentation at the 2018 Fall Technical Conference, sponsored by the American Statistical Association and the American Society for Quality.

부록 C

기술 라이선스 고지 사항

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

• LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

• clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

버전 16

다변량 방법

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는 것이 아니라 새로운 눈을 가지는 데 있다."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 16 다변량 방법. Cary, NC: SAS Institute Inc.

JMP® 16 다변량 방법

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

August 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP를 처음 사용하든 오랫동안 사용해왔든, JMP와 관련해서 배워야 하는 것은 언제나 있습니다.

JMP.com을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다.

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티, 추가 기능 및 스크립트 예를 포함한 사용자용 리소스, 포럼, 블로그, 컨퍼런스 정보 등

https://www.jmp.com/getstarted

목차
다변량 방법

1 JMP 알아보기 13
설명서 및 추가 리소스
JMP 설명서에 사용되는 서식 규칙 15
JMP 도움말 16
JMP 설명서 라이브러리 16
JMP 학습을 위한 추가 리소스 22
JMP 자습서 22
샘플 데이터 테이블 22
통계 및 JSL 용어에 대해 알아보기 23
JMP 팁 및 힌트 알아보기 23
JMP 툴팁 23
JMP 사용자 커뮤니티 23
무료 온라인 Statistical Thinking 교육 과정 24
JMP 새로운 사용자를 위한 입문 키트 24
Statistics Knowledge 포털 24
JMP 교육 과정 24
사용자가 작성한 JMP 설명서 24
JMP 시작하기 창 24
JMP 기술 지원 25

2 다변량 분석 소개 27
다변량 기법 개요

3 상관 및 다변량 기법 29
변수의 다차원 동작 탐색
다변량 플랫폼의 예 31
다변량 플랫폼 시작 33
다변량 보고서 34
다변량 플랫폼 옵션 35
비모수 상관 38
산점도 행렬 39
이상치 분석 40

항목 신뢰도 43
결측 데이터 대치 43
항목 신뢰도의 예 43
계산 및 통계 상세 정보 44
추정 방법 45
Pearson 곱적률 상관계수 45
비모수 연관성 측도 46
역상관 행렬 47
거리 측도 47
Cronbach α 49

4 주성분 51

데이터 차원 축소

주성분 분석 플랫폼 개요 53
주성분 분석의 예 53
주성분 플랫폼 시작 54
결측 데이터 56
주성분 보고서 56
주성분 보고서 옵션 58
이상치 분석 66
주성분 분석 플랫폼에 대한 통계 상세 정보 68
추정 방법 68
DModX 계산 69
이상치 분석을 위한 계산 70

5 판별 분석 71

연속형 변수를 기반으로 분류 예측

판별 분석 플랫폼 개요 73
판별 분석의 예 73
판별 분석 플랫폼 시작 75
단계별 변수 선택 76
판별 방법 80
공분산 축소 82
판별 분석 보고서 82
주성분 83
정준 그림 및 정준 구조 84
판별 스코어 87
스코어 요약 88
판별 분석 옵션 91
스코어 옵션 92

정준 옵션 94
3D 정준 그림의 예 97
사전 확률 지정 98
새 수준 고려 98
판별 행렬 저장 99
산점도 행렬 99
JMP 및 JMP Pro 에서 검증 100
판별 분석 플랫폼에 대한 통계 상세 정보 101
와이드 선형 알고리즘 설명 101
저장된 계산식 101
다변량 검정 108
근사 F- 검정 109
그룹 간 공분산 행렬 110

6 부분 최소 제곱 모형 111

Y 와 X 간의 상관을 사용하여 모형 개발

부분 최소 제곱 플랫폼 개요 113
부분 최소 제곱의 예 114
부분 최소 제곱 플랫폼 시작 117
중심화 및 척도화 120
X 표준화 120
모형 시작 제어판 121
부분 최소 제곱 옵션 122
부분 최소 제곱 보고서 123
모형 비교 요약 123
교차 검증 보고서 124
모형 적합 보고서 128
모형 적합 옵션 129
변수 중요도 그림 131
VIP 대 계수 그림 132
열 저장 133
부분 최소 제곱의 추가 예 134
부분 최소 제곱 플랫폼에 대한 통계 상세 정보 136
부분 최소 제곱 136
van der Voet T^2 137
T^2 그림 138
X 스코어 산점도 행렬의 신뢰도 타원 138
예측 표준 오차 및 신뢰 한계 139
표준화된 스코어 및 적재량 140
PLS-DA(PLS 판별 분석) 141

7 다중 대응 분석 143

범주형 변수 수준 간의 연관성 식별

다중 대응 분석의 예 145
다중 대응 분석 플랫폼 시작 148
다중 대응 분석 보고서 149
다중 대응 분석 플랫폼 옵션 150
대응 분석 옵션 151
그림 표시 152
상세 정보 표시 153
조정 관성 표시 153
좌표 표시 153
요약 통계량 표시 154
관성에 대한 부분 기여도 표시 154
제곱코사인 표시 155
Cochran Q 검정 155
교차표 155
다중 대응 분석 플랫폼의 추가 예 156
보조 변수를 사용한 예 156
보조 ID 를 사용한 예 157
Cochran Q 검정의 예 158
다중 대응 분석 플랫폼에 대한 통계 상세 정보 159
상세 정보 보고서 160
조정 관성 160
요약 통계량 161
관성에 대한 부분 기여도 161
Cochran Q 통계량 162

8 구조 방정식 모형 163

구조 방정식 모형 적합

구조 방정식 모형 개요 165
구조 방정식 모형의 예 167
구조 방정식 모형 플랫폼 시작 171
구조 방정식 모형 보고서 172
모형 규격 보고서 173
모형 비교 보고서 178
카이제곱 차이 검정 보고서 179
구조 방정식 모형 적합 보고서 179
구조 방정식 모형 플랫폼 옵션 181
모형 옵션 182
SEM 경로 다이어그램 사용자 정의 186

경로 다이어그램 팝업 메뉴 옵션 186
다이어그램 모양 사용자 정의 옵션 188
구조 방정식 모형의 추가 예 189
잠재 경로 변수 모형의 예 189
잠재 성장 곡선 모형의 예 196
측정 모형 평가 보고서 사용의 예 198
구조 방정식 모형 플랫폼에 대한 통계 상세 정보 201
적합 요약 측도에 대한 통계 상세 정보 201

9 요인 분석 205
데이터의 잠재 변수 식별
요인 분석 플랫폼 개요 207
요인 분석 플랫폼의 예 207
요인 분석 플랫폼 시작 210
요인 분석 보고서 211
모형 시작 212
회전 방법 213
요인 분석 플랫폼 옵션 214
요인 분석 모형 적합 옵션 215

10 다차원 척도법 219
개체 집합 간의 근접성 시각화
다차원 척도법 플랫폼 개요 221
다차원 척도법의 예 221
다차원 척도법 플랫폼 시작 224
다차원 척도 보고서 225
다차원 척도 그림 225
Shepard 다이어그램 225
적합 상세 정보 226
다차원 척도법 플랫폼 옵션 226
Waern 링크 227
다차원 척도법 플랫폼의 추가 예 228
다차원 척도법 플랫폼에 대한 통계 상세 정보 230
스트레스 230
변환 230
속성 목록 형식 231

11 항목 분석 233
항목 및 대상별 검정 결과 분석
항목 분석의 예 235
항목 분석 플랫폼 시작 237

로지스틱 3PL 모형 상세 정보 ... 238
데이터 형식 ... 238
항목 분석 보고서 ... 238
특성 곡선 ... 238
정보 그림 ... 239
쌍대 그림 ... 240
모수 추정값 ... 240
항목 분석 플랫폼 옵션 ... 241
항목 분석 플랫폼에 대한 통계 상세 정보 ... 241
항목 반응 곡선 ... 241
항목 반응 곡선 모형 ... 242
IRT 모형 가정 ... 244
IRT 모형 적합 ... 245
능력 계산식 ... 246

12 계층적 군집화 ... 247
군집 트리를 사용하여 관측값 그룹화

계층적 군집화 플랫폼 개요 ... 249
관측값 군집화 플랫폼 개요 ... 249
군집화의 예 ... 250
계층적 군집화 플랫폼 시작 ... 253
군집화 방법 ... 254
거리 계산 방법 ... 254
데이터 구조 ... 255
Y, 열 변수로 변환 ... 256
계층적 군집화 보고서 ... 257
덴드로그램 보고서 ... 258
덴드로그램 및 거리 그래프 예시 ... 258
군집화 기록 보고서 ... 260
계층적 군집화 옵션 ... 260
계층적 군집화 플랫폼의 추가 예 ... 263
거리 행렬의 예 ... 263
공간 측도를 사용한 웨이퍼 결함 분류의 예 ... 265
계층적 군집화 플랫폼에 대한 통계 상세 정보 ... 267
공간 측도 ... 267
거리 방법 계산식 ... 269

13 K 평균 군집화 ... 271
거리를 사용하여 관측값 그룹화

K 평균 군집화 플랫폼 개요 ... 273

관측값 군집화 플랫폼 개요 ... 273
K 평균 군집화의 예 ... 274
K 평균 군집화 플랫폼 시작 ... 277
반복적 군집화 보고서 ... 278
반복적 군집화 옵션 ... 278
반복적 군집화 제어판 ... 279
K 평균 보고서 ... 280
군집 비교 보고서 ... 280
K 평균 보고서 ... 280
K 평균 보고서 옵션 ... 280
자기 조직화 지도 ... 282
자기 조직화 지도 제어판 ... 282
자기 조직화 지도 보고서 ... 283
SOM 알고리즘 설명 ... 284
K 평균 군집화 플랫폼의 추가 예 ... 284
자기 조직화 지도의 예 ... 284

14 정규 혼합 ... 287
확률을 사용하여 관측값 그룹화

정규 혼합 군집화 플랫폼 개요 ... 289
관측값 군집화 플랫폼 개요 ... 289
정규 혼합 군집화의 예 ... 290
정규 혼합 군집화 플랫폼 시작 ... 292
모형 기반 군집화 보고서 ... 293
모형 기반 군집화 옵션 ... 293
모형 기반 군집화 제어판 ... 294
정규 혼합 보고서 ... 295
군집 비교 보고서 ... 295
정규 혼합 보고서 ... 295
정규 혼합 보고서 옵션 ... 295
정규 혼합 군집화 플랫폼에 대한 통계 상세 정보 ... 297

15 잠재 계층 분석 ... 299
범주형 변수의 관측값 그룹화

잠재 계층 분석 플랫폼 개요 ... 301
잠재 계층 분석의 예 ... 301
잠재 계층 분석 플랫폼 시작 ... 305
잠재 계층 분석 보고서 ... 306
군집 비교 보고서 ... 306
잠재 계층 모형 보고서 ... 306

잠재 계층 분석 플랫폼 옵션 308
잠재 계층 분석 308
잠재 계층 모형 옵션 309
잠재 계층 분석 플랫폼의 추가 예 309
소속 군집 확률 그림 309
잠재 계층 분석 플랫폼에 대한 통계 상세 정보 311
잠재 계층 모형 적합 311
최대 군집 수 312

16 변수 군집화 313
유사한 변수를 대표 그룹으로 그룹화

변수 군집화 플랫폼 개요 315
변수 군집화 플랫폼의 예 315
변수 군집화 플랫폼 시작 317
변수 군집화 보고서 317
상관 색상 맵 318
군집 요약 318
군집 멤버 318
표준화 성분 319
변수 군집화 플랫폼 옵션 319
변수 군집화 플랫폼의 추가 예 320
상관 색상 맵의 예 320
차원 축소를 위한 변수 군집화 플랫폼의 예 321
변수 군집화 플랫폼에 대한 통계 상세 정보 324
변수 군집화 알고리즘 324

A 통계 상세 정보 327
다변량 방법

와이드 선형 방법 및 특이값 분해 329
특이값 분해 329
SVD 및 공분산 행렬 330
SVD 및 공분산 역행렬 330
SVD 계산 331

B 참조 자료 333

C 기술 라이선스 고지 사항 339

1 장

JMP 알아보기
설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 15
JMP 도움말 16
JMP 설명서 라이브러리 16
JMP 학습을 위한 추가 리소스 22
JMP 자습서 22
샘플 데이터 테이블 22
통계 및 JSL 용어에 대해 알아보기 23
JMP 팁 및 힌트 알아보기 23
JMP 툴팁 23
JMP 사용자 커뮤니티 23
무료 온라인 Statistical Thinking 교육 과정 24
JMP 새로운 사용자를 위한 입문 키트 24
Statistics Knowledge 포털 24
JMP 교육 과정 24
사용자가 작성한 JMP 설명서 24
JMP 시작하기 창 24
JMP 기술 지원 25

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 *`Lucida Sans Typewriter`* 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 "JMP 설명서 라이브러리 "(16 페이지) 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
예측 및 전문 모델링	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
다변량 방법	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
품질 및 공정 방법	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기 " 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기 "

창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

2 장

다변량 분석 소개

다변량 기법 개요

다변량 방법에서는 여러 변수를 동시에 분석하는 다음 기법에 대해 설명합니다 .

- 다변량 플랫폼에서는 여러 변수가 서로 어떤 관계가 있는지 검토합니다 . 자세한 내용은 " 상관 및 다변량 기법 " 에서 확인하십시오 .
- 주성분 플랫폼에서는 측정된 변수 집합 중 원래 변수의 변동성을 가능한 한 많이 포착하는 소수의 독립적 선형 결합 (주성분) 을 도출합니다 . 이 방법은 유용한 탐색 기법이며 예측 모형을 생성하는 데 도움이 될 수 있습니다 . 자세한 내용은 " 주성분 " 에서 확인하십시오 .
- 판별 플랫폼에서는 알려진 연속형 반응 (Y) 을 기반으로 명목형 또는 순서형 분류 (X) 변수를 예측하는 방법을 찾습니다 . 이 방법은 MANOVA(다변량 분산 분석) 의 역추정 예측으로 간주할 수 있습니다 . 자세한 내용은 " 판별 분석 " 에서 확인하십시오 .
- 부분 최소 제곱 플랫폼에서는 요인 , 즉 설명 변수 (X) 의 선형 결합을 기반으로 선형 모형을 적합시킵니다 . PLS 는 X 와 Y 사이의 상관을 이용하여 기본 잠재 구조를 나타냅니다 . 자세한 내용은 " 부분 최소 제곱 모형 " 에서 확인하십시오 .
- MCA(다중 대응 분석) 플랫폼에서는 여러 범주형 변수를 사용하여 해당 변수 수준 간의 연관성을 식별합니다 . MCA 는 특히 프랑스와 일본에서 사회 과학 분야에 주로 사용됩니다 . 설문 조사 분석에서 질문 합치도를 식별하는 데 사용할 수 있습니다 . 자세한 내용은 " 다중 대응 분석 " 에서 확인하십시오 .
- JMP PRO 구조 방정식 모형 플랫폼에서는 확증적 요인 분석 , 잠재 변수를 사용하거나 사용하지 않는 경로 모형 , 측정 오차 모형 등의 다양한 모형을 적합시킬 수 있습니다 . 자세한 내용은 " 구조 방정식 모형 " 에서 확인하십시오 .
- 요인 분석 플랫폼을 사용하면 더 큰 관측 변수 집합에서 요인을 구성할 수 있습니다 . 이러한 요인은 관측 변수 부분집합의 선형 결합으로 표현됩니다 . 요인 분석을 사용하면 측정 및 관측된 변수 집합으로 설명되는 요인의 수 , 요인과 변수 간 관계의 강도 등을 탐색할 수 있습니다 . 자세한 내용은 " 요인 분석 " 에서 확인하십시오 .
- MDS(다차원 척도법) 플랫폼을 사용하면 개체 집합 간의 근접성 패턴 (유사성 , 비유사성 또는 거리) 을 시각적으로 표현할 수 있습니다 . 자세한 내용은 " 다차원 척도법 " 에서 확인하십시오 .
- 항목 분석 플랫폼을 사용하면 항목 반응 이론 모형을 적합시킬 수 있습니다 . IRT(항목 반응 이론) 방법은 검사 및 설문지와 같이 측정기의 분석 및 스코어링에 사용됩니다 . IRT 는 개인이 어떤 항목을 지지하거나 올바르게 응답할 확률을 특성 또는 능력과 연관시키기 위해 모형 시스템을 사용합니다 . 표준화 검사 , 인지 발달 및 소비자 선호도를 연구할 때 IRT 를 사용할 수 있습니다 . 자세한 내용은 " 항목 분석 " 에서 확인하십시오 .

- 계층적 군집화 플랫폼에서는 여러 변수에 대해 유사한 값을 공유하는 행을 함께 그룹화합니다. 이 방법은 데이터의 군집 구조를 이해하는 데 도움이 되는 유용한 탐색 기법입니다. 자세한 내용은 " 계층적 군집화 " 에서 확인하십시오.
- K 평균 군집화 플랫폼에서는 여러 변수에 대해 유사한 값을 공유하는 관측값을 그룹화합니다. 자세한 내용은 "K 평균 군집화 " 에서 확인하십시오.
- 정규 혼합 플랫폼을 사용하면 중첩 정규 분포의 데이터를 가져올 때 관측값을 군집화할 수 있습니다. 자세한 내용은 " 정규 혼합 " 에서 확인하십시오.
- 잠재 계층 분석 플랫폼에서는 범주형 반응 변수에 대한 관측값 군집을 찾습니다. 모형은 다항 혼합 모형의 형태를 취합니다. 자세한 내용은 " 잠재 계층 분석 " 에서 확인하십시오.
- 변수 군집화 플랫폼에서는 유사한 변수를 대표 그룹으로 그룹화합니다. 변수 군집화를 차원 축소 방법으로 사용할 수 있습니다. 큰 변수 집합을 사용하여 모델링하는 대신 군집에서 가장 대표적인 변수의 군집 성분을 사용하여 데이터 변동의 대부분을 설명할 수 있습니다. 자세한 내용은 " 변수 군집화 " 에서 확인하십시오.

3 장

상관 및 다변량 기법
변수의 다차원 동작 탐색

다변량 데이터에는 하나(단변량) 또는 두 개(이변량)가 아니라 많은 변수가 포함됩니다. 다변량 플랫폼을 사용하여 여러 변수가 서로 어떤 관련이 있는지 탐색할 수 있습니다. 다변량 플랫폼에서는 각 반응 변수 쌍 간의 선형 관계 강도를 요약하고 검정하는 다양한 기법을 제공합니다. 모수 상관 검정과 비모수 상관 검정 모두 플랫폼에서 사용할 수 있습니다. 또한 산점도 행렬과 색상 맵 같은 그래픽 기능을 사용하여 변수 사이의 종속성, 이상치 및 군집을 식별할 수도 있습니다.

주성분 분석, 이상치 분석 및 항목 신뢰도를 포함하여 변수 간의 관계를 더 자세히 조사할 수 있는 추가 다변량 분석 기법이 있습니다. 이러한 기법은 "다변량" 보고서를 통해 사용할 수 있습니다. JMP의 주성분 분석 플랫폼과 이상치 분석 플랫폼을 사용하여 이러한 기법을 더 심층적으로 구현할 수 있습니다.

그림 3.1 다변량 보고서의 예

목차

다변량 플랫폼의 예 31
다변량 플랫폼 시작 33
다변량 보고서 34
다변량 플랫폼 옵션 35
비모수 상관 38
산점도 행렬 39
이상치 분석 40
항목 신뢰도 43
결측 데이터 대치 43
항목 신뢰도의 예 43
계산 및 통계 상세 정보 44
추정 방법 45
Pearson 곱적률 상관계수 45
비모수 연관성 측도 46
역상관 행렬 47
거리 측도 47
Cronbach α 49

다변량 플랫폼의 예

이 예에서는 Body Measurements.jmp 샘플 데이터 테이블을 사용하여 다양한 신체 측정값 간의 관계를 검토합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Body Measurements.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 다변량**을 선택합니다 .
3. Mass 를 제외한 모든 열을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

   초기 다변량 보고서에는 상관 행렬과 산점도 행렬이 포함됩니다. 행별 추정 방법이 사용되었음을 알려 주는 노트도 있습니다 . 모든 변수는 양의 상관관계가 있지만 강도가 다양합니다 .
5. " 다변량 " 의 빨간색 삼각형을 클릭하고 **산점도 행렬**을 선택 취소합니다 .
6. " 다변량 " 의 빨간색 삼각형을 클릭하고 **상관 확률**을 선택합니다 .
7. " 다변량 " 의 빨간색 삼각형을 클릭하고 **색상 맵 > 상관 색상 맵**을 선택합니다 .

그림 3.2 Body Measurements 데이터에 대한 다변량 보고서

다변량

상관

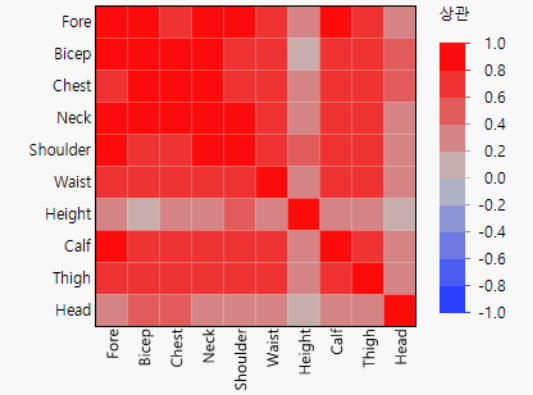
	Fore	Bicep	Chest	Neck	Shoulder	Waist	Height	Calf	Thigh	Head
Fore	1.0000	0.8134	0.7671	0.8134	0.8670	0.7577	0.3537	0.8130	0.7877	0.3029
Bicep	0.8134	1.0000	0.8552	0.8240	0.7202	0.6652	0.0657	0.6421	0.6974	0.4199
Chest	0.7671	0.8552	1.0000	0.8279	0.7999	0.7154	0.2017	0.6436	0.7802	0.4379
Neck	0.8134	0.8240	0.8279	1.0000	0.8312	0.7530	0.3912	0.6095	0.7055	0.3223
Shoulder	0.8670	0.7202	0.7999	0.8312	1.0000	0.7462	0.4839	0.6847	0.6825	0.2258
Waist	0.7577	0.6652	0.7154	0.7530	0.7462	1.0000	0.3513	0.6893	0.7403	0.2997
Height	0.3537	0.0657	0.2017	0.3912	0.4839	0.3513	1.0000	0.3977	0.2860	0.1576
Calf	0.8130	0.6421	0.6436	0.6095	0.6847	0.6893	0.3977	1.0000	0.7497	0.2742
Thigh	0.7877	0.6974	0.7802	0.7055	0.6825	0.7403	0.2860	0.7497	1.0000	0.2363
Head	0.3029	0.4199	0.4379	0.3223	0.2258	0.2997	0.1576	0.2742	0.2363	1.0000

행별 방법으로 상관계수를 추정합니다.

상관 확률

	Fore	Bicep	Chest	Neck	Shoulder	Waist	Height	Calf	Thigh	Head
Fore	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1064	<.0001	<.0001	0.1706
Bicep	<.0001	<.0001	<.0001	<.0001	0.0002	0.0007	0.7715	0.0013	0.0003	0.0517
Chest	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.3681	0.0012	<.0001	0.0415
Neck	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0718	0.0026	0.0002	0.1436
Shoulder	<.0001	0.0002	<.0001	<.0001	<.0001	<.0001	0.0225	0.0004	0.0005	0.3124
Waist	<.0001	0.0007	0.0002	<.0001	<.0001	<.0001	0.1089	0.0004	<.0001	0.1753
Height	0.1064	0.7715	0.3681	0.0718	0.0225	0.1089	<.0001	0.0668	0.1969	0.4836
Calf	<.0001	0.0013	0.0012	0.0026	0.0004	0.0004	0.0668	<.0001	<.0001	0.2168
Thigh	<.0001	0.0003	<.0001	0.0002	0.0005	<.0001	0.1969	<.0001	<.0001	0.2898
Head	0.1706	0.0517	0.0415	0.1436	0.3124	0.1753	0.4836	0.2168	0.2898	<.0001

상관 색상 맵

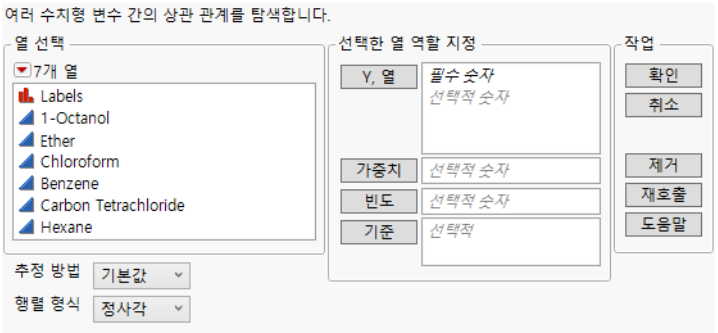


" 상관 색상 맵 " 보고서는 상관 행렬의 정보를 더 간결하게 나타냅니다 . 그림의 대부분은 진한 빨간색으로 표시되어 대부분의 변수가 높은 상관관계가 있음을 나타냅니다. 더 연하게 표시된 두 행과 열은 Height(키) 와 Head(머리 둘레) 측정값이 다른 변수와 큰 상관관계가 없음을 나타냅니다 . 이 사실은 " 상관 확률 " 테이블에서 Height 와 Head 의 p 값이 대부분 유의하지 않음을 통해 뒷받침됩니다 .

다변량 플랫폼 시작

분석 > 다변량 방법 > 다변량을 선택하여 다변량 플랫폼을 시작합니다 .

그림 3.3 다변량 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 열 하나 이상의 반응 열을 식별합니다 . 반응 열의 데이터 유형은 숫자여야 하지만 모델링 유형은 연속형 또는 순서형일 수 있습니다 .

참고 : 순서형 반응 변수를 지정하는 경우 시작 창에서 " 확인 " 을 클릭하면 JMP 경고가 나타납니다 . 이 경고는 어떤 변수가 순서형인지 나타내고 사용자가 분석에 순서형 변수를 포함하려고 했음을 확인합니다 .

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 하나의 열을 식별합니다 .

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 하나의 열을 식별합니다 .

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

추정 방법 상관 계산 방법을 지정합니다 . "REML" 과 " 쌍별 " 방법이 가장 자주 사용됩니다 . 이러한 방법 중 일부는 결측 데이터를 처리합니다 . " 행별 " 을 제외한 다른 방법을 사용하고 " 결측 데이터 대치 " 명령을 선택하여 결측값을 추정할 수도 있습니다 . 자세한 내용은 " 결측 데이터 대치 " 에서 확인하십시오 .

기본값 " 기본값 " 옵션은 " 행별 ", " 쌍별 " 또는 "REML" 방법을 사용합니다 .

- 행별 추정은 결측값이 없는 데이터 테이블에 사용됩니다 .
- 쌍별 추정은 결측값이 있고 열이 10 개 , 행이 5,000 개를 각각 초과하거나 행보다 열이 더 많은 데이터 테이블에 사용됩니다 .
- REML 추정은 그 외의 경우에 사용됩니다 .

참고 :" 기본값 " 옵션이 "REML" 추정을 사용하지만 적합이 올바르게 수렴되지 않으면 플랫폼이 " 쌍별 " 방법으로 돌아갑니다 . 이는 결측값이 있는 데이터 테이블에서 열이 10 개 미만 , 행이 5,000 개 미만 또는 열이 행보다 적은 경우 중 하나 이상에 해당할 때 발생할 수 있습니다. 표시된 추정 방법이 "쌍별"이면 REML 적합이 수렴되지 않았음을 의미합니다.

REML REML(제한 최대 가능도) 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다 . 데이터 집합이 크고 결측값이 많은 경우 이 방법을 사용하면 편향 수정 요인 때문에 처리 속도가 느립니다 . 따라서 REML 은 작은 데이터 집합에 가장 유용합니다 . 데이터에 결측 셀이 없으면 REML 및 ML 추정값이 동일하고 표본 공분산 행렬과 같습니다 . 결측 셀이 있으면 REML 의 분산 및 공분산 추정값이 ML 추정의 추정값보다 덜 편향됩니다 . 통계 상세 정보는 "REML" 에서 확인하십시오 .

ML ML(최대 가능도) 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다 . ML 추정값은 빠르게 생성되므로 이 방법은 결측 데이터가 있는 큰 데이터 테이블에 가장 유용합니다 .

로버스트 로버스트 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다 . 이 방법은 극단값의 비중을 낮추므로 이상치가 있을 수 있는 데이터 테이블에 유용합니다 . 통계 상세 정보는 " 로버스트 " 에서 확인하십시오 .

행별 행별 추정은 각 열 쌍에 대한 Pearson 상관을 계산합니다. 통계 상세 정보는 "Pearson 곱적률 상관계수 " 에서 확인하십시오 . 행별 추정에서는 결측값이 있는 행을 사용하지 않습니다 . 이 방법은 결측 데이터가 있는 관측값을 제외하는 데 유용합니다 .

쌍별 쌍별 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다 . 이 추정 방법은 두 열에 대해 결측값이 없는 모든 관측값을 사용하여 각 열 쌍에 대한 Pearson 상관을 계산합니다 . 통계 상세 정보는 "Pearson 곱적률 상관계수 " 에서 확인하십시오 . 쌍별 추정은 결측값이 있는 데이터 테이블에서 행보다 열이 더 많거나 , 열이 10 개를 초과하거나 , 행이 5,000 개를 초과하는 경우에 가장 유용합니다 .

참고 : "REML", "ML" 또는 " 로버스트 " 를 선택한 경우 데이터 테이블에 행보다 열이 더 많고 결측값이 있으면 JMP 에서 " 추정 방법 " 이 " 쌍별 " 로 전환됩니다 .

행렬 형식 산점도 행렬의 형식 옵션을 선택합니다 . " 정사각 " 옵션은 정렬된 모든 열 조합에 대한 그림을 표시합니다 . " 하삼각 " 은 대각선 아래에 그림을 표시하며 처음 n - 1 개 열이 가로 축에 나타납니다 . " 상삼각 " 은 대각선 위에 그림을 표시하며 처음 n - 1 개 열이 세로 축에 나타납니다 .

다변량 보고서

기본 다변량 보고서에는 표준 상관 행렬, 산점도 행렬, 상관 추정에 사용된 방법을 나타내는 노트가 표시됩니다. 경우에 따라 지정된 방법이 사용된 이유를 설명하는 정보도 표시됩니다. 플랫폼 메뉴에는 여러 변수를 검토하기 위한 추가 상관 옵션 및 기타 기법이 나열되어 있습니다. 자세한 내용은 "다변량 플랫폼 옵션"에서 확인하십시오.

다변량 플랫폼 옵션

참고 : " 다변량 " 보고서에 표시된 상관 값 및 *p* 값 형식은 " 환경 설정 " 의 조건부 형식 설정을 사용하여 지정됩니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

" 다변량 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

다변량 상관 " 상관 " 테이블을 표시하거나 숨깁니다 . 이 테이블은 각 반응 (*Y*) 변수 쌍 간의 선형 관계 강도가 요약된 상관계수 행렬입니다 . 이 옵션은 기본적으로 설정되어 있습니다 . 자세한 내용은 "Pearson 곱적률 상관계수 " 에서 확인하십시오 .

참고 : 이 상관 행렬은 시작 창에서 선택한 방법으로 계산됩니다 .

상관 확률 " 상관 확률 " 테이블을 표시하거나 숨깁니다 . 이 테이블은 *p* 값 행렬입니다 . 각 *p* 값은 변수 사이의 실제 상관이 0 이라는 귀무가설 검정에 해당합니다 . 즉 , 두 반응 변수 사이에 선형 관계가 없는지에 대한 검정입니다 .

상관 CI 상관의 양쪽 꼬리 신뢰 구간을 표시하거나 숨깁니다 .

팁 : 기본 신뢰 계수는 95% 입니다 . 신뢰 계수를 변경하려면 "α 수준 설정 " 옵션을 사용하십시오 .

역상관 역상관 행렬을 나타내는 " 역상관 " 보고서를 표시하거나 숨깁니다 . 행렬의 대각 요소는 변수가 다른 변수의 선형 함수에 얼마나 가까운지에 대한 함수입니다. 역상관 테이블에서 대각선은 $1/(1-R^2)$ 의 값입니다 . R^2 은 다른 모든 변수를 기반으로 해당 변수를 모델링하는 단순 선형 모형에서 계산됩니다 . 다중상관이 0 이면 대각선 역요소가 1 입니다 . 다중상관이 1 이면 역요소가 무한대가 되고 결측으로 보고됩니다 . 역상관에 대한 통계 상세 정보는 " 역상관 행렬 " 에서 확인하십시오 .

부분 상관 부분 상관 행렬을 나타내는 " 부분 상관 " 보고서를 표시하거나 숨깁니다 . 부분 상관 행렬에는 다른 모든 변수의 효과를 조정한 후 변수 쌍 사이의 관계에 대한 측도가 표시됩니다 . 이 테이블은 역상관 행렬에 음수를 취하고 , 단위 대각값으로 척도화된 것입니다 . 즉 , 대각 요소가 1 이 되도록 행렬이 척도화됩니다 .

부분 상관 확률 *p* 값 행렬을 나타내는 " 부분 상관 확률 " 보고서를 표시하거나 숨깁니다 . 각 *p* 값은 변수 사이의 실제 부분 상관이 0 이라는 귀무가설 검정에 해당합니다 . 즉 , 다른 변수의 효과를 조정한 후 두 반응 변수 사이에 선형 관계가 없는지에 대한 검정입니다 .

참고 : 자유도가 충분하지 않으면 " 부분 상관 확률 " 옵션을 사용할 수 없습니다 . 이는 관측값보다 변수가 더 많을 때 발생할 수 있습니다 .

공분산 행렬 변수 쌍이 함께 변화하는 정도를 측정하는 공분산 행렬을 표시하거나 숨깁니다 .

쌍별 상관 각 Y 변수 쌍에 대한 Pearson 곱적률 상관계수가 나열된 " 쌍별 상관 " 보고서를 표시하거나 숨깁니다 . 상관은 쌍별 제거 방법으로 계산됩니다 . 변수 쌍 중 어느 한 변수에 대해

결측값이 있으면 개수 값이 달라집니다 . 또한 " 쌍별 상관 " 보고서는 유의 확률을 보여 주고 막대 차트로 상관관계를 비교합니다 . 모든 결과는 쌍별 방법을 기반으로 합니다 .

참고 : 이 옵션은 고려 중인 변수 중 하나에 대해 결측인 행을 제외합니다 .

Hotelling T^2 검정 Y 로 입력한 변수의 다변량 분포 평균에 대한 1 표본 검정을 수행할 수 있습니다 . "Hotelling T^2 검정 " 옵션을 선택하면 귀무가설 하에서 평균 벡터를 지정할 수 있는 창이 나타납니다 . 각 변수에 대해 가설 평균을 입력합니다 . 이 검정에서는 Y 변수의 다변량 정규성을 가정합니다 . "Hotelling T^2 검정 " 보고서에는 다음 정보가 제공됩니다 .

변수 Y 로 입력된 변수입니다 .

평균 각 변수의 표본 평균입니다 .

가설 평균 지정한 귀무가설 평균입니다 .

검정 통계량 Hotelling T^2 통계량 값입니다 .

F 비 검정 통계량 값입니다 . n 개의 행과 k 개의 변수가 있는 경우 F 비는 다음과 같이 정의됩니다 .

$$\frac{n-k}{k(n-1)}T^2$$

Prob > F 검정의 p 값입니다 . 귀무가설 하에서 F 비는 k 및 n - k 자유도의 F 분포를 따릅니다 .

참고 : 보고서를 제거하려면 "Hotelling T^2 검정 " 옆의 빨간색 삼각형을 클릭하고 **검정 제거** 를 선택합니다 .

단순 통계량 이 메뉴에는 각 열에 대한 단순 통계량 (평균 , 표준편차 , 합 , 최소값 및 최대값) 을 표시하거나 숨기는 두 가지 옵션이 포함되어 있습니다 . 결측값이 있거나 " 로버스트 " 방법을 사용하는 경우 단변량 및 다변량 단순 통계량이 다를 수 있습니다 .

단변량 단순 통계량 다른 열의 값이나 결측값에 관계없이 각 열에 대해 계산되는 통계량을 표시합니다 . 이러한 값은 분포 플랫폼에서 생성되는 값과 일치합니다 .

다변량 단순 통계량 시작 창에서 선택한 추정 방법과 결측 데이터가 있는지 여부에 부합하는 통계량을 표시합니다 . 결측 관측값이 없으면 " 로버스트 " 방법에만 이 옵션을 사용할 수 있습니다 . 결측 관측값이 있으면 " 쌍별 " 을 제외한 모든 추정 방법에 이 옵션을 사용할 수 있습니다 . "REML", "ML" 또는 " 로버스트 " 방법의 경우 선택한 방법으로 평균 벡터와 공분산 행렬이 추정됩니다 . " 행별 " 방법의 경우 하나 이상의 결측값이 있는 모든 행이 평균 및 분산 계산에서 제외됩니다 .

비모수 상관 이 메뉴에는 Spearman Rho, Kendall Tau, Hoeffding D 라는 세 가지 비모수 측도가 포함되어 있습니다 . 각 옵션은 해당 측도에 대한 비모수 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 비모수 상관 " 에서 확인하십시오 .

α 수준 설정 상관 신뢰 구간에 대한 알파 값을 지정할 수 있습니다 . 0.01, 0.05, 0.10 및 0.50 의 네 가지 알파 값이 나열됩니다 . 다른 값을 입력하려면 **기타** 를 선택합니다 .

산점도 행렬 각 반응 변수 쌍의 산점도 행렬을 표시하거나 숨깁니다. 이 옵션은 기본적으로 설정되어 있습니다. 자세한 내용은 "산점도 행렬"에서 확인하십시오.

색상 맵 "색상 맵" 메뉴에는 7 개 유형의 색상 맵이 포함되어 있습니다. 메뉴의 각 옵션은 해당 유형의 색상 맵을 표시하거나 숨깁니다. 다음 유형의 색상 맵을 사용할 수 있습니다.

상관 색상 맵 빨간색 (+1) ~ 파란색 (-1) 의 척도로 변수 간 상관을 보여 주는 셀 그림을 생성합니다.

p 값 색상 맵 $p = 0$(빨간색) ~ $p = 1$(파란색)의 척도로 상관 유의성을 보여 주는 셀 그림을 생성합니다.

상관 군집화 유사한 변수를 함께 군집화하는 셀 그림을 생성합니다. 상관은 "상관 색상 맵"의 상관과 같지만 변수 위치가 다를 수 있습니다.

쌍별 상관 색상 맵 빨간색 (+1) ~ 파란색 (-1) 의 척도로 변수 간 쌍별 상관을 보여 주는 셀 그림을 생성합니다.

Spearman ρ 색상 맵 빨간색 (+1) ~ 파란색 (-1) 의 척도로 변수 간 Spearman ρ 비모수 상관을 보여 주는 셀 그림을 생성합니다.

Kendall τ 색상 맵 빨간색 (+1) ~ 파란색 (-1) 의 척도로 변수 간 Kendall τ 비모수 상관을 보여 주는 셀 그림을 생성합니다.

Hoeffding D 색상 맵 빨간색 (+1) ~ 파란색 (-1) 의 척도로 변수 간 Hoeffding D 비모수 상관을 보여 주는 셀 그림을 생성합니다.

평행 좌표 그림 변수의 평행 좌표 그림을 표시하거나 숨깁니다.

타원 3D 그림 95% 신뢰 타원체가 있는 3 차원 산점도를 표시하거나 숨깁니다. 이 옵션을 선택하면 세 개의 변수와 해당 축을 지정해야 합니다.

주성분 이 메뉴에는 주성분 보고서를 표시하거나 숨기기 위한 옵션이 포함되어 있습니다. 상관, 공분산 또는 비척도화 값을 기반으로 하는 주성분 보고서를 선택할 수 있습니다. 주성분 보고서가 이미 표시된 상태에서 이러한 옵션 중 하나를 선택하면 보고서가 새 옵션으로 바뀝니다. 보고서를 제거하려면 **없음**을 선택합니다.

주성분은 원래 변수의 선형 결합을 계산하는 기법입니다. 첫 번째 주성분이 최대 변동을 가지며, 두 번째 주성분이 그 다음으로 큰 변동을 가지는 방식이며 첫 번째 주성분에 대해 직교합니다. 자세한 내용은 "주성분"에서 확인하십시오.

이상치 분석 이 메뉴에는 Mahalanobis 거리, 잭나이프 거리 및 T^2 통계량 방법 중 하나를 사용하여 다변량 관점의 거리를 측정하는 그림을 각각 표시하거나 숨기기 위한 옵션이 포함되어 있습니다. 자세한 내용은 "이상치 분석"에서 확인하십시오.

항목 신뢰도 이 메뉴에는 각각 항목 신뢰도 보고서를 표시하거나 숨기기 위한 옵션이 포함되어 있습니다. 보고서는 Cronbach α 또는 표준화 α 를 사용하여 일련의 측정기에서 전체 반응을 얼마나 일관되게 측정하는지 나타냅니다. 자세한 내용은 "항목 신뢰도"에서 확인하십시오.

결측 데이터 대치 (데이터 테이블에 결측값이 포함된 경우에만 사용 가능) 데이터 테이블을 복제하고 모든 결측값을 추정값으로 대체하는 새 데이터 테이블을 생성합니다. 자세한 내용은 "결측 데이터 대치"에서 확인하십시오.

결측값 대치 계산식 저장 (데이터 테이블에 결측값이 포함된 경우에만 사용 가능) 결측값이 포함된 열에 대해 결측값 추정에 사용된 계산식을 포함하는 새 열을 데이터 테이블에 저장합니다. 새 열 이름은 **결측값 대치됨** < 열 이름 > 입니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

비모수 상관

" 비모수 상관 " 메뉴는 쌍별 상관에 대한 세 가지 비모수 측도를 제공합니다. 각 비모수 상관 보고서에서는 선택한 연관성 측도에 대한 유의 확률을 제공하고 연관성 값을 막대 차트에 표시합니다. 세 가지 비모수 상관 측도가 있습니다.

Spearman Rho 데이터 값 대신 데이터 값의 순위에 대해 계산된 상관계수입니다.

Kendall Tau 관측값의 부합 / 비부합 쌍의 개수를 기반으로 합니다. 관측값의 X 값이 더 클 때 Y 값도 더 큰 쌍이 **부합** 쌍이고, 관측값의 X 값이 더 클 때 Y 값이 더 작은 쌍은 **비부합** 쌍입니다. 동일 쌍, 즉 *X* 값이 같거나 *Y* 값이 같은 관측값 쌍에 대해서는 수정이 적용됩니다.

Hoeffding D –0.5 ~ 1 범위의 통계 척도입니다. 큰 양수 값은 종속성을 나타냅니다. 이 통계량은 2 x 2 분류 테이블에 대한 카이제곱 통계량의 관측값에 대해 가중 합에 근사한 값을 산출합니다. 2 x 2 테이블은 각 데이터 값을 임계값으로 설정하여 만들어집니다. 이 통계량은 일반적인 독립성 이탈을 더 많이 감지합니다.

참고: 비모수 상관은 시작 창에서 다른 추정 방법을 선택한 경우에도 쌍별 방법을 사용하여 계산됩니다.

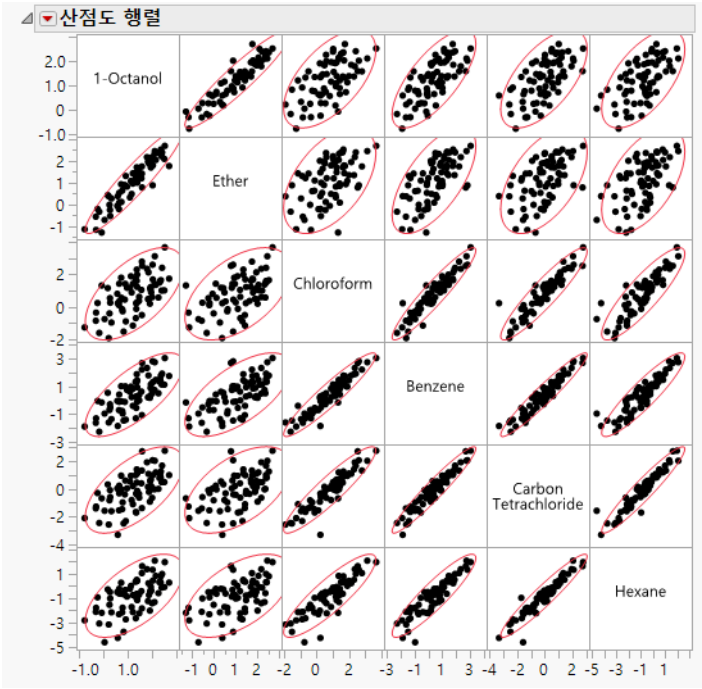
참고 : 가중치 변수가 지정된 경우 결측 및 0 값 가중치는 비모수 상관 계산에서 제외됩니다. 다른 모든 가중치 값은 1 로 처리됩니다.

이 세 가지 방법에 대한 통계 상세 정보는 "비모수 연관성 측도"에서 확인하십시오.

산점도 행렬

산점도 행렬은 각 반응 변수 쌍 사이의 상관을 시각화하는 데 도움이 됩니다.

그림 3.4 산점도 행렬



"밀도 타원" 옵션을 설정하면 각 산점도에 95% 이변량 정규 밀도 타원이 표시됩니다. 각 변수 쌍이 이변량 정규분포를 따른다고 가정하면 이 타원은 전체 점의 약 95%를 둘러쌉니다. 타원의 너비는 변수의 상관관계 정도를 반영합니다. 타원이 상당히 둥근 형태이고 대각선 방향이 아니면 해당 변수 간에 상관관계가 없는 것입니다. 타원이 좁고 대각선 방향이면 해당 변수 간에 상관관계가 높은 것입니다.

팁 :

- 한 셀의 크기를 조정하면 모든 셀의 크기가 조정됩니다 .
- 라벨 셀을 다른 라벨 셀로 드래그하면 행렬이 재정렬됩니다 .
- 산점도 행렬에서 패턴을 찾을 때 변수가 상관관계에 따라 그룹으로 군집화되는 것을 확인할 수 있습니다 . 그림 3.4 에서는 두 개의 상관 군집을 보여 줍니다 (왼쪽 위의 두 변수와 오른쪽 아래의 네 변수).

산점도 행렬 옵션

" 산점도 행렬 " 의 빨간색 삼각형 메뉴를 사용하면 색상과 밀도 타원을 사용하고 α 수준을 설정하여 행렬을 사용자 정의할 수 있습니다 .

점 표시 산점도에 점을 표시하거나 숨깁니다 .

선형 적합 회귀선 및 적합 회귀선에 대한 95% 수준 신뢰 곡선을 표시하거나 숨깁니다 .

밀도 타원 산점도에 95% 밀도 타원을 표시하거나 숨깁니다 . " 타원 α" 메뉴를 사용하여 α 수준을 변경할 수 있습니다 .

음영 타원 각 타원에 색상을 적용합니다 . " 타원 투명도 " 및 " 타원 색상 " 메뉴를 사용하여 투명도와 색상을 변경할 수 있습니다 .

상관 표시 각 산점도의 왼쪽 위에 각 변수 쌍의 상관을 표시하거나 숨깁니다 .

행렬 옵션 (시작 창에서 " 정사각 " 행렬 형식을 선택한 경우에만 사용 가능) 산점도 행렬의 오른쪽 상단 삼각형 영역 모양을 변경하는 옵션이 포함된 하위 메뉴를 표시합니다 . 다음 옵션을 한 번에 하나만 선택할 수 있습니다 .

유의성 원 산점도 행렬의 오른쪽 상단 삼각형 영역에 상관 원을 표시하거나 숨깁니다 . 각 원의 색상은 각 변수 쌍 간의 상관관계를 빨간색 (+1) ~ 파란색 (-1) 의 척도로 나타냅니다 . 각 원의 크기는 변수 간의 유의성 검정을 나타냅니다 . 원이 클수록 더 유의한 관계를 나타냅니다 .

히트맵 산점도 행렬의 오른쪽 상단 삼각형 영역에 상관 히트맵을 표시하거나 숨깁니다 . 히트맵의 각 셀 색상은 각 변수 쌍 간의 상관관계를 빨간색 (+1) ~ 파란색 (-1) 의 척도로 나타냅니다 .

히스토그램 표시 라벨 셀에 가로 또는 세로 히스토그램을 표시하거나 숨깁니다 . 히스토그램을 추가한 후 **개수 표시**를 선택하여 히스토그램의 각 막대에 개수로 라벨을 지정할 수 있습니다 . **수평** 또는 **수직**을 선택하여 히스토그램 방향을 변경하거나 히스토그램을 제거합니다 .

타원 α 타원에 사용되는 α 수준을 설정합니다 . 메뉴에서 표준 α 수준 중 하나를 선택하거나 , **기타**를 선택하여 다른 값을 입력합니다 .

타원 투명도 색상이 적용되는 타원의 투명도를 설정합니다 . 기본 수준 중 하나를 선택하거나 , **기타**를 선택하여 다른 값을 입력합니다 . 기본값은 0.2 입니다 .

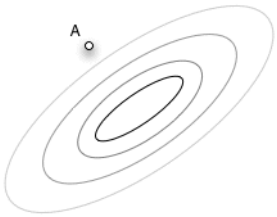
타원 색상 색상이 적용되는 타원의 색상을 설정합니다 . 색상 팔레트에서 하나를 선택하거나 , **기타**를 선택하여 다른 색상을 사용합니다 . 기본값은 빨간색입니다 .

비모수 밀도 데이터 점의 밀도를 묘사하는 평활 비모수 이변량 표면을 기반으로 음영 밀도 등고선을 표시하거나 숨깁니다 . 비모수 표면의 10% 및 50% 분위수에 대한 등고선이 표시됩니다 .

이상치 분석

"이상치 분석" 메뉴에는 상관 구조와 관련하여 다변량 관점의 거리를 측정하는 그림을 각각 표시하거나 숨기기 위한 옵션이 포함되어 있습니다 . 예를 들어 그림 3.5에서 점 A는 좌표 방향에서 이상치가 아니지만 상관 구조를 벗어나므로 이상치입니다 .

그림 3.5 이상치의 예



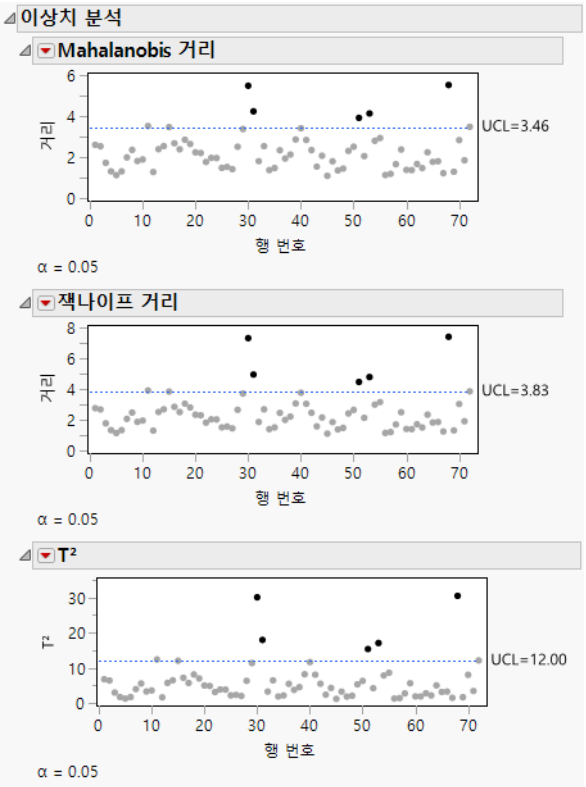
" 이상치 분석 " 메뉴의 각 옵션은 해당 거리의 그림을 표시하거나 숨깁니다 . 그림 아래쪽에 표시된 유의 수준으로 검정이 수행됩니다 . 다음 옵션을 사용할 수 있습니다 .

Mahalanobis 거리 다변량 평균 (중심) 과 각 점 사이의 Mahalanobis 거리를 표시하거나 숨깁니다 . 표준 Mahalanobis 거리는 데이터의 평균 , 표준편차 및 상관관계 추정값에 따라 달라집니다 . 각 관측값 번호에 대해 거리가 표시됩니다 . 거리 값이 가장 큰 점을 강조 표시하여 극단 다변량 이상치를 식별할 수 있습니다 . 자세한 내용은 "Mahalanobis 거리 측도 " 에서 확인하십시오 .

잭나이프 거리 잭나이프 기법을 사용하여 계산된 거리를 표시하거나 숨깁니다 . 각 관측값의 거리는 관측값 자체를 포함하지 않는 평균 , 표준편차 및 상관 행렬의 추정값을 사용하여 계산됩니다 . 잭나이프 거리는 이상치가 있을 때 유용합니다 . 이 경우 Mahalanobis 거리는 왜곡되어 이상치를 감추거나 다른 점이 이상치보다 더 멀리 보이게 하는 경향이 있습니다 . 자세한 내용은 " 잭나이프 거리 측도 " 에서 확인하십시오 .

T^2 Mahalanobis 거리의 제곱인 거리를 표시하거나 숨깁니다 . 이 그림은 다변량 관리도에서 선호됩니다 . 계산된 T^2 통계량 값과 관리 상한이 그림에 포함됩니다 . 이 한계를 벗어나는 값은 이상치일 수 있습니다 . 자세한 내용은 "T^2 거리 측도 " 에서 확인하십시오 .

그림 3.6 이상치 분석 그림



거리 및 값 저장

그림의 빨간색 삼각형 메뉴에서 **저장** 옵션을 선택하여 거리를 데이터 테이블에 저장할 수 있습니다.

참고 : 잭나이프 거리 열에는 계산식이 저장되지 않습니다. 즉, 데이터 테이블을 수정해도 거리가 다시 계산되지 **않습니다**. 열을 추가 또는 삭제하거나 데이터 테이블의 값을 변경하는 경우 **분석 > 다변량 방법 > 다변량**을 다시 선택하여 새 잭나이프 거리를 계산해야 합니다.

각 행의 거리 값을 저장하는 것 외에도 지정된 이상치 분석 유형에 대한 UCL(관리 상한) 값을 포함하는 열 특성이 생성됩니다.

항목 신뢰도

항목 신뢰도는 일련의 측정기에서 전체 반응을 얼마나 일관되게 측정하는지 나타냅니다 . Cronbach α(Cronbach 1951) 는 신뢰도 측도 중 하나입니다 . Cronbach α 가 적용되는 두 가지 주요 분야는 공업용 측정기 신뢰도와 설문지 분석입니다 .

Cronbach α는 측정 척도에서 항목의 평균 상관을 기반으로 합니다. 이는 데이터 테이블에서 모든 반분 상관의 평균을 계산하는 것과 같습니다. 항목의 분산이 광범위한 경우 표준화 α를 요청할 수 있습니다.

참고 : Cronbach α 는 유의 수준 α 와 관련이 없습니다 . 또한 항목 신뢰도는 생존 시간 신뢰성 분석과 관련이 없습니다 .

JMP에서 개별 항목의 영향을 검토하기 위해 해당 항목을 계산에서 제외하고 Cronbach α 값의 효과를 표시합니다. 변수(항목)를 제외할 때 α가 증가하면 해당 변수는 다른 변수와 높은 상관관계가 없습니다. α가 감소하면 해당 변수가 척도의 다른 항목과 상관되어 있다는 결론을 내릴 수 있습니다.

계산에 대한 자세한 내용은 "Cronbach α" 에서 확인하십시오.

결측 데이터 대치

결측 데이터를 대치하려면 "다변량" 의 빨간색 삼각형을 클릭하고 **결측 데이터 대치**를 선택합니다. 새 데이터 테이블이 생성됩니다. 이 데이터 테이블은 사용 중인 데이터 테이블을 복제하여 결측값을 추정값으로 대체한 것과 같습니다.

대치된 값은 각 행의 비결측값에 따른 조건부 기대값입니다. 시작 창에서 선택한 방법으로 추정된 평균 및 공분산 행렬이 결측값 대치 계산에 사용됩니다. 그런 다음 모든 다변량 검정 및 옵션을 대치된 데이터 집합에 사용할 수 있습니다.

항목 신뢰도의 예

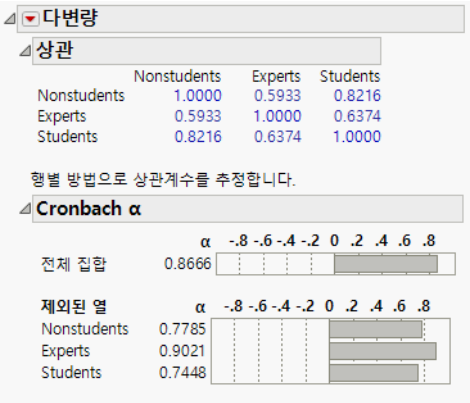
이 예에서는 샘플 데이터 폴더의 Danger.jmp 데이터를 사용합니다 . 이 테이블에는 일정 수준의 위험이 내재된 30 개 항목이 나열되어 있습니다 . 세 그룹의 사람들 (학생 , 학생 아님 , 전문가) 이 인식한 위험 수준에 따라 항목에 순위를 매겼습니다 . Nuclear power(원자력) 의 경우 학생 그룹과 학생이 아닌 그룹은 모두 매우 위험 (1) 하다고 평가했지만 전문가 그룹은 위험 수준을 낮게 (20) 평가했습니다 . 반면 , 오토바이의 위험 수준은 세 심사 그룹 모두 5 위 또는 6 위로 평가했습니다 .

Cronbach α 를 사용하여 그룹이 항목 순위를 매기는 인식 방식에서 합치도를 평가할 수 있습니다 . 값이 각 그룹에 대해 동일한 순위 집합을 나타내는 이 유형의 예에서는 데이터를 표준화하는 것이 아무 효과가 없습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Danger.jmp 를 엽니다 .

2. **분석 > 다변량 방법 > 다변량**을 선택합니다 .
3. Activity 를 제외한 모든 열을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 다변량 " 의 빨간색 삼각형을 클릭하고 **항목 신뢰도 > Cronbach** α 를 선택합니다 .
6. (선택 사항) " 다변량 " 의 빨간색 삼각형을 클릭하고 **산점도 행렬**을 선택하여 해당 그림을 숨깁니다 .

그림 3.7 Cronbach α 보고서



Cronbach α 결과에는 전체 α 0.8666 이 표시되어 세 그룹 간의 순위 값이 높은 상관관계가 있음을 나타냅니다. 또한 분석에서 Experts(전문가 그룹) 를 제거하면 Nonstudents 와 Students 는 Cronbach α 스코어가 각각 0.7785 와 0.7448 로, 위험 순위를 거의 동일하게 평가했습니다.

계산 및 통계 상세 정보

- " 추정 방법 "
- "Pearson 곱적률 상관계수 "
- " 비모수 연관성 측도 "
- " 역상관 행렬 "
- " 거리 측도 "
- "Cronbach α"

추정 방법

REML

데이터에 결측값이 포함된 경우 REML(제한 최대 가능도) 추정값은 ML(최대 가능도) 추정 방법보다 덜 편향됩니다. REML 방법은 오차 대비를 기반으로 주변 가능도를 최대화합니다. REML 방법은 주로 분산 및 공분산을 추정하는 데 사용됩니다. 주성분 플랫폼의 REML 방법은 비정형 공분산 행렬을 사용하는 반복 측정 데이터에 대한 혼합 모형의 REML 추정과 동일합니다. 혼합 모형의 REML 추정에 대한 자세한 내용은 SAS Institute Inc. (2020e) 의 "MIXED Procedure" 장에서 확인하십시오.

로버스트

기본적으로 이 방법은 이상치의 비중을 크게 줄여 해당 값을 무시합니다 . 데이터의 반복 재가중 적합 시퀀스는 다음 가중치를 사용하여 수행됩니다 .

$Q < K$ 인 경우 $w_i = 1.0$, 그렇지 않은 경우 $w_i = K/Q$

여기서 K 는 자유도가 데이터 테이블의 열 수와 같은 카이제곱 분포의 0.75 분위수와 동일한 상수입니다 . Q 는 다음과 같이 정의됩니다 .

$$Q = (y_i - \mu)^T (S^2)^{-1} (y_i - \mu)$$

이 방정식에서 $y_i = i$번째 관측값에 대한 반응, μ = 평균 벡터의 현재 추정값, S^2 = 공분산 행렬의 현재 추정값, T = 전치 행렬 연산입니다. 마지막 단계는 분산 행렬의 편향 축소입니다.

이 방법을 사용하면 데이터에 이상치가 많지 않을 때는 분산 추정값이 더 높지만 이상치가 있을 때 훨씬 더 정확한 분산 추정값을 얻을 수 있습니다.

Pearson 곱적률 상관계수

Pearson 곱적률 상관계수는 두 변수 간의 선형 관계 강도를 측정합니다 . 반응 변수 X 와 Y 에 대해 r 로 표시되고 다음과 같이 계산됩니다 .

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

두 변수가 정확히 선형 관계에 있는 경우 상관계수는 변수가 양의 관계인지 아니면 음의 관계인지에 따라 1 또는 –1 입니다. 선형 관계가 없으면 상관계수가 0에 가까워집니다.

비모수 연관성 측도

Spearman, Kendall 또는 Hoeffding 상관의 경우 데이터의 순위가 먼저 매겨집니다. 그런 다음 데이터 값 순위에 대해 계산이 수행됩니다. 동점일 경우 평균 순위가 사용됩니다.

참고 : 가중치 변수가 지정된 경우 결측 및 0 값 가중치는 비모수 상관 계산에서 제외됩니다 . 다른 모든 가중치 값은 1 로 처리됩니다 .

Spearman ρ(Rho) 계수

Spearman ρ 상관계수는 앞에서 설명한 Pearson 상관 계산식을 사용하여 데이터 순위에 대해 계산됩니다.

Kendall τ_b 계수

Kendall τ_b 계수는 부합/비부합 쌍의 개수를 기반으로 합니다. 두 변수의 행 쌍은 더 큰 변수가 일치할 때 **부합** 쌍입니다. 그렇지 않은 경우 비부합 쌍 또는 동점입니다.

아래 계산식은

$$\tau_b = \frac{\sum_{i<j} \operatorname{sgn}(x_i - x_j)\operatorname{sgn}(y_i - y_j)}{\sqrt{(T_0 - T_1)(T_0 - T_2)}}$$

Kendall τ_b 를 계산합니다 . 다음은 각 요소에 대한 설명입니다 .

$$T_0 = (n(n-1))/2$$

$$T_1 = \sum((t_i)(t_i-1))/2$$

$$T_2 = \sum((u_i)(u_i-1))/2$$

다음 사항에 유의하십시오 .

- sgn(z) 은 z > 0 인 경우 1, z = 0 인 경우 0, z < 0 인 경우 –1 입니다 .
- $t_i(u_i)$ 는 동점 x(각 y) 값의 i 번째 그룹에서 동점 x(각 y) 값의 수입니다 .
- n 은 관측값 수입니다 .
- Kendall τ_b 의 범위는 –1 ~ 1 입니다 . 가중치 변수가 지정되면 무시됩니다 .

계산은 다음과 같은 방식으로 진행됩니다 .

- 첫 번째 변수의 값에 따라 순서대로 관측값 순위가 매겨집니다 .
- 그런 다음 두 번째 변수의 값에 따라 관측값 순위가 다시 매겨집니다 .
- 첫 번째 변수의 교환 횟수를 사용하여 Kendall τ_b 를 계산합니다 .

Hoeffding D 통계량

Hoeffding *D*(1948) 계산식은 다음과 같습니다 .

$$D = 30\left(\frac{(n-2)(n-3)D_1 + D_2 - 2(n-2)D_3}{n(n-1)(n-2)(n-3)(n-4)}\right)$$

다음은 각 요소에 대한 설명입니다 .

$$D_1 = \Sigma_i(Q_i - 1)(Q_i - 2)$$
$$D_2 = \Sigma_i(R_i - 1)(R_i - 2)(S_i - 1)(S_i - 2)$$
$$D_3 = \Sigma_i(R_i - 2)(S_i - 2)(Q_i - 1)$$

다음 사항에 유의하십시오 .

- R_i 와 S_i 는 x 및 y 값의 순위입니다 .
- Q_i(이변량 순위라고도 함) 는 x 값과 y 값이 모두 i 번째 점보다 작은 점의 수에 1 을 더한 값입니다 .
- x 값 또는 y 값 중 하나만 같은 점은 다른 값이 i 번째 점의 해당 값보다 작으면 Q_i 에 1/2 을 적용합니다 . x 와 y 둘 다 같은 점은 Q_i 에 1/4 을 적용합니다 .

관측값 중에서 동점이 없으면 *D* 통계량의 값은 –0.5에서 1 사이이며, 여기서 1은 완전 종속을 나타냅니다. 가중치 변수가 지정되면 무시됩니다.

역상관 행렬

역상관 행렬은 유용한 다변량 정보를 제공합니다 . 역상관 행렬의 대각 요소는 변수가 다른 변수의 선형 함수에 얼마나 가까운지에 대한 함수이며 , VIF(분산 팽창 계수) 라고도 합니다 . 구체적으로 말하면 상관 행렬을 **R** 로 나타내고 역상관 행렬을 $\mathbf{R}^{-1}$ 로 나타낼 경우 대각 요소는 r_{ii} 로 나타내고 다음과 같이 계산됩니다 .

$$r_{ii} = \mathrm{VIF}_i = \frac{1}{1 - \mathrm{R}_i^2}$$

여기서 R_i^2은 다른 설명 변수를 기반으로 i번째 설명 변수를 모델링하는 모형의 변동 계수입니다. 따라서 r_{ii}가 크면 i번째 변수가 다른 변수와 높은 상관관계가 있음을 나타냅니다.

거리 측도

"이상치 분석" 그림에는 데이터 테이블의 각 점에 대해 지정된 거리 측도가 표시됩니다.

Mahalanobis 거리 측도

Mahalanobis 거리는 데이터의 상관 구조와 개별 척도를 고려합니다 . 각 값에 대해 Mahalanobis 거리는 M_i 로 나타내고 다음과 같이 계산됩니다 .

$$M_i = \sqrt{(\mathrm{Y}_i - \overline{\mathrm{Y}})'\mathrm{S}^{-1}(\mathrm{Y}_i - \overline{\mathrm{Y}})}$$

다음은 각 요소에 대한 설명입니다 .

$\mathrm{Y}_i = i$ 번째 행의 데이터

$\overline{\mathrm{Y}}$ = 평균 행

S = 데이터에 대한 추정된 공분산 행렬

"Mahalanobis 거리 " 그림에 표시된 UCL 참조선 (Mason 과 Young 2002) 은 다음과 같이 계산됩니다 .

$$UCL_{Mahalanobis} = \sqrt{\frac{(n-1)^2}{n}\beta_{\left[1-\alpha;\frac{p}{2};\frac{n-p-1}{2}\right]}}$$

다음은 각 요소에 대한 설명입니다 .

n = 관측값 수

p = 변수 (열) 수

$\beta_{\left[1-\alpha;\frac{p}{2};\frac{n-p-1}{2}\right]}$ = 베타 $\left(\frac{p}{2}, \frac{n-p-1}{2}\right)$ 분포의 (1–α) 번째 분위수

변수가 다른 변수의 정확한 선형 결합인 경우 상관 행렬은 특이 행렬이고 해당 변수의 행과 열은 0으로 설정됩니다. 결과로 생성되는 일반화 역행렬은 거리를 구성하는 데 유효합니다.

잭나이프 거리 측도

잭나이프 거리는 관측값 자체를 포함하지 않는 평균 , 표준편차 및 상관 행렬의 추정값을 사용하여 계산됩니다 . 각 값에 대해 잭나이프 거리는 다음과 같이 계산됩니다 .

$$J_i = \sqrt{\frac{(n-2)n^2}{(n-1)^3} \times \frac{M_i^2}{1 - \frac{nM_i^2}{(n-1)^2}}}$$

다음은 각 요소에 대한 설명입니다 .

n = 관측값 수

p = 변수 (열) 수

M_i = i 번째 관측값의 Mahalanobis 거리

" 잭나이프 거리 " 그림에 표시된 UCL 참조선 (Penny 1996) 은 다음과 같이 계산됩니다 .

$$UCL_{Jackknife} = \sqrt{\frac{(n-2)n^2}{(n-1)^3} \times \frac{UCL_{Mahalanobis}^{\ 2}}{1 - \frac{n \bullet UCL_{Mahalanobis}^{\ 2}}{(n-1)^2}}}$$

T^2 거리 측도

T^2 거리는 Mahalanobis 거리의 제곱이므로 $T_i^2 = M_i^2$입니다.

T^2 거리의 UCL 은 다음과 같습니다 .

$$UCL_{T^2} = \frac{(n-1)^2}{n} \beta_{\left[1-\alpha;\frac{p}{2};\frac{n-p-1}{2}\right]} = (UCL_{Mahalanobis})^2$$

다음은 각 요소에 대한 설명입니다 .

n = 관측값 수

p = 변수 (열) 수

$\beta_{\left[1-\alpha;\frac{p}{2};\frac{n-p-1}{2}\right]}$ = 베타 $\left(\frac{p}{2}, \frac{n-p-1}{2}\right)$ 분포의 (1–α) 번째 분위수

다변량 거리는 여러 차원에서 이상치를 찾는 데 유용합니다. 그러나 변수가 다변량 관점에서 높은 상관관계가 있는 경우에는 점을 차원의 부분집합에서 비정상으로 보지 않고 다변량 공간의 이상치로 간주할 수 있습니다. 즉, 값이 상관되어 있는 경우 하나 또는 두 개의 축을 따라 보면 점이 특별하지 않을 수 있지만 상관관계를 위반하여 이상치가 될 수 있습니다.

Cronbach α

Cronbach α 는 다음과 같이 정의됩니다 .

$$\alpha = \frac{kc}{v + (k-1)c}$$

다음은 각 요소에 대한 설명입니다 .

k = 척도의 항목 수

c = 항목 간의 평균 공분산

v = 항목 간의 평균 분산

항목이 일정한 분산을 가지도록 표준화되는 경우 계산식은 다음과 같습니다 .

$$\alpha = \frac{k(r)}{1 + (k-1)r}$$

다음은 각 요소에 대한 설명입니다 .

r = 항목 간의 평균 상관

전체 α 계수가 클수록 항목이 신뢰할 수 있는 척도 또는 검정에 기여한다는 것을 더 확신할 수 있습니다. 상관관계가 높은 항목이 많은 경우 계수가 1.0에 근접할 수 있습니다.

4 장

주성분
데이터 차원 축소

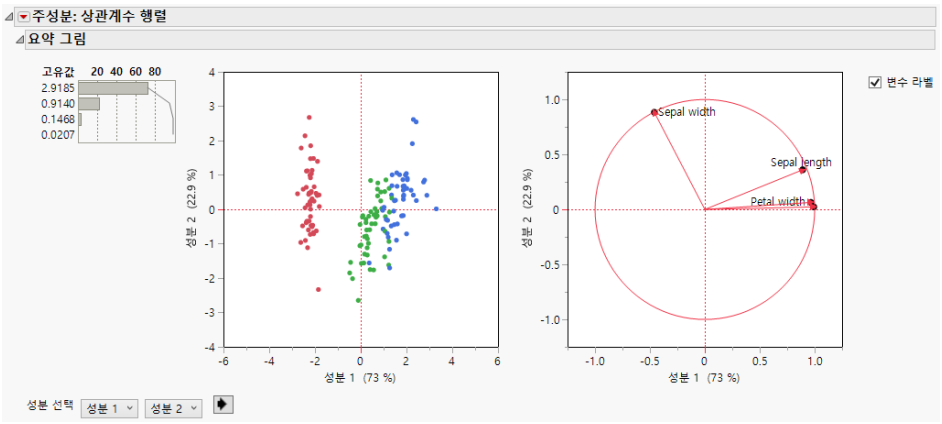
주성분 분석의 목적은 측정된 변수 집합 중 원래 변수의 변동성을 가능한 한 많이 포착하는 소수의 독립적 선형 결합(**주성분**)을 도출하는 것입니다. 주성분 분석은 탐색적 데이터 분석 도구이자 차원 축소 기법입니다. 주성분 분석은 **주성분 분석 회귀**(PCA 회귀 또는 PCR이라고도 함)에서와 같이 예측 모형을 생성하는 데도 유용합니다.

변수가 매우 많은 데이터를 위해 주성분 플랫폼에서는 "와이드" 추정 방법을 제공합니다. "와이드" 방법을 사용하면 짧은 시간에 주성분을 계산할 수 있습니다. 그런 다음 이러한 주성분을 PCA 회귀에 사용할 수 있습니다.

대부분 0을 포함하는 데이터(**희소 데이터**라고도 함)를 위해 주성분 플랫폼에서는 "희소" 추정 방법을 제공합니다. "와이드" 방법과 같이 "희소" 방법도 짧은 시간에 주성분을 계산합니다. 그러나 "와이드" 방법과 달리 "희소" 방법은 전체 집합이 아니라 사용자가 정의한 고정된 수의 주성분을 계산합니다.

주성분 플랫폼에서는 요인 분석도 지원합니다. JMP에서는 추출된 성분을 해석하는 데 도움이 되는 여러 유형의 직교 및 사각 요인 분석 스타일 회전을 제공합니다. 요인 분석에 대한 자세한 내용은 "요인 분석"에서 확인하십시오.

그림 4.1 주성분의 예



목차

주성분 분석 플랫폼 개요 53
주성분 분석의 예 53
주성분 플랫폼 시작 54
결측 데이터 56
주성분 보고서 56
주성분 보고서 옵션 58
이상치 분석 66
주성분 분석 플랫폼에 대한 통계 상세 정보 68
추정 방법 68
DModX 계산 69
이상치 분석을 위한 계산 70

주성분 분석 플랫폼 개요

주성분 분석은 변수 집합의 변동을 원래 변수보다 더 적은 수의 **주성분**(원래 변수의 독립적 선형 결합)으로 모델링합니다.

상관된 많은 변수에 대한 점 배열을 보려면 주성분 분석을 사용하여 가장 눈에 띄는 고차원 데이터의 방향을 표시할 수 있습니다. 주성분 분석을 사용하면 데이터 집합의 차원이 축소됩니다. 주성분은 가능한 한 적은 수의 변수를 사용하여 데이터 구조를 최대한 완벽하게 나타낼 수 있는 방법입니다.

변수가 p 개인 경우 다음과 같은 방식으로 p 개의 주성분이 구성됩니다.

- 첫 번째 주성분은 표준화된 원래 변수 중에서 가능한 가장 큰 분산을 갖는 선형 결합입니다.
- 이후의 각 주성분은 가능한 분산이 가장 큰 변수 선형 결합이며 이전에 정의된 모든 성분과 상관관계가 없습니다.

각 주성분은 상관 행렬(또는 공분산 행렬이나 제곱합 및 교차곱 행렬)의 고유 벡터를 변수와 선형 결합하여 계산됩니다. 고유값은 각 성분의 분산을 나타냅니다.

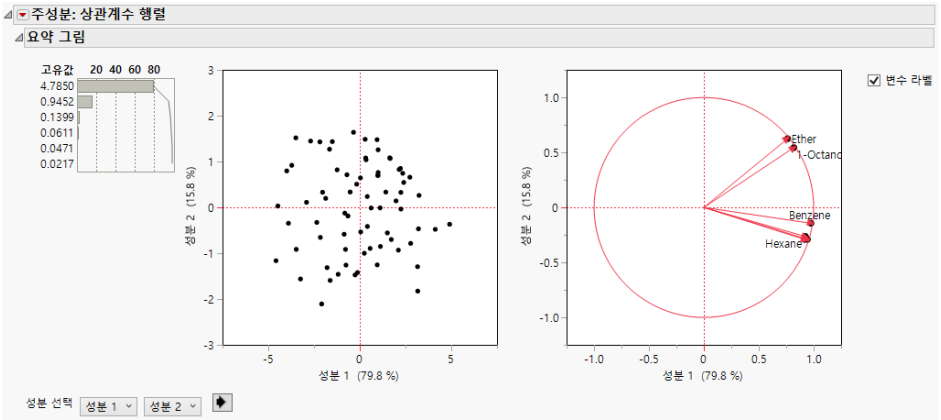
주성분 플랫폼을 사용하면 상관 행렬, 공분산 행렬 또는 비척도화 데이터에 대한 분석을 수행할 수 있습니다. 주성분 플랫폼 내에서 요인 분석을 수행할 수도 있습니다. 자세한 내용은 "요인 분석"에서 확인하십시오.

주성분 분석의 예

두 요인에 대한 데이터 테이블의 주성분 분석 보고서 예를 보려면 다음을 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Solubility.jmp 를 엽니다.
2. **분석 > 다변량 방법 > 주성분**을 선택합니다.
3. 연속형 열을 모두 선택하고 **Y, 열**을 클릭합니다.
4. "기본값"으로 설정된 "추정 방법"을 그대로 유지하고 **확인**을 클릭합니다.

그림 4.2 상관계수 행렬을 이용한 주성분 보고서



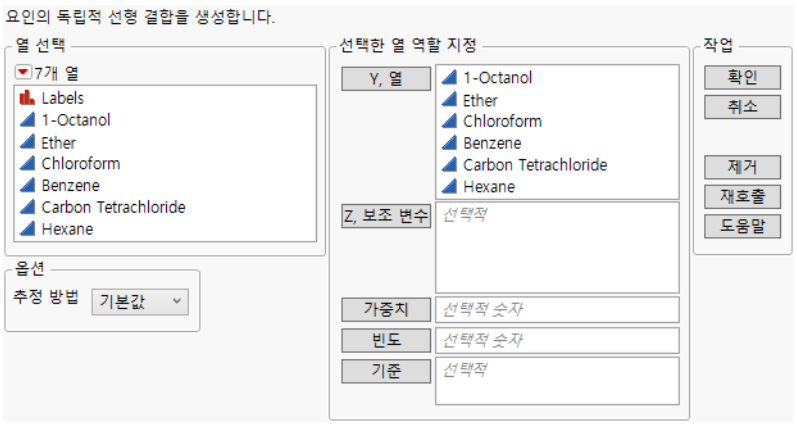
이 보고서에는 각 주성분이 설명하는 변동의 백분율을 나타내는 막대 차트와 고유값이 제공됩니다. 이 예에서는 첫 번째 주성분이 데이터 변동의 거의 80%를 설명합니다. 처음 두 주성분이 함께 데이터 변동을 거의 모두(95.5%) 설명합니다. 스코어 그림과 적재 그림도 있습니다. 자세한 내용은 "주성분 보고서"에서 확인하십시오.

주성분 플랫폼 시작

분석 > 다변량 방법 > 주성분을 선택하여 주성분 플랫폼을 시작합니다 . 다변량 및 3D 산점도 플랫폼을 사용하여 주성분 분석을 사용할 수도 있습니다 .

"주성분 분석의 예"에 설명된 예에서는 Solubility.jmp 샘플 데이터 테이블의 모든 연속형 변수를 사용합니다 .

그림 4.3 주성분 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 열 성분에 대해 분석할 변수입니다.

Z, 보조 변수 표시할 보조 변수입니다. 보조 변수는 주성분 계산에 포함되지 않으며 포함해도 결과에 영향을 주지 않습니다 . 연속형 보조 변수는 적재 그림에 투영되어 해석을 향상시키는 데 사용할 수 있습니다.

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 하나의 열을 식별합니다.

참고 : " 와이드 " 및 " 희소 " 추정 방법에서는 가중치 역할이 무시됩니다.

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 하나의 열을 식별합니다.

참고 : " 와이드 " 및 " 희소 " 추정 방법에서는 빈도 역할이 무시됩니다.

기준 각 그룹에 대한 개별 분석을 수행할 수 있도록 " 기준 " 열에 지정된 각 값의 주성분 보고서를 생성합니다.

추정 방법 상관 계산 방법을 지정합니다. 이러한 방법 중 일부는 결측 데이터를 처리합니다.

기본값 **기본값** 옵션은 " 행별 ", " 쌍별 " 또는 "REML" 방법을 사용합니다. 해당하는 경우 " 와이드 " 방법으로 전환할 것을 권장하는 JMP 경고가 표시됩니다.

- **행별** 추정은 결측값이 없는 데이터 테이블에 사용됩니다.
- **쌍별** 추정은 결측값이 있고 열이 10 개 , 행이 5,000 개를 각각 초과하거나 행보다 열이 더 많은 데이터 테이블에 사용됩니다.
- **REML** 추정은 그 외의 경우에 사용됩니다.
- **와이드** 추정은 열이 500 개를 초과하고 행보다 열이 더 많은 데이터 테이블의 경우 JMP 경고 창이 표시되어 사용을 권장합니다. 열 수가 많은 경우 다른 방법을 사용하면 계산 시간이 길어질 수 있기 때문입니다 . **와이드**를 클릭하여 와이드 방법으로 전환하거나 **계속**을 클릭하여 원래 선택한 방법을 사용합니다.

REML REML(제한 최대 가능도) 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다 . 데이터 집합이 크고 결측값이 많은 경우 이 방법을 사용하면 편향 수정 요인 때문에 처리 속도가 느립니다. 따라서 REML 은 작은 데이터 집합에 가장 유용합니다. 데이터에 결측 셀이 없으면 REML 및 ML 추정값이 동일하고 표본 공분산 행렬과 같습니다 . 결측 셀이 있으면 REML 의 분산 및 공분산 추정값이 ML 추정의 추정값보다 덜 편향됩니다 . 통계 상세 정보는 "REML" 에서 확인하십시오.

ML ML(최대 가능도) 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다. ML 추정값은 빠르게 생성되므로 이 방법은 결측 데이터가 있는 큰 데이터 테이블에 가장 유용합니다.

로버스트 로버스트 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다 . 이 방법은 극단값의 비중을 낮추므로 이상치가 있을 수 있는 데이터 테이블에 유용합니다. 통계 상세 정보는 " 로버스트 " 에서 확인하십시오.

행별 행별 추정은 각 열 쌍에 대한 Pearson 상관을 계산합니다. 통계 상세 정보는 "Pearson 곱적률 상관계수 " 에서 확인하십시오. 행별 추정에서는 결측값이 있는 관측값을 사용하지 않습니다. 이 방법은 결측 데이터가 있는 관측값을 제외하는 데 유용합니다.

쌍별 쌍별 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다. 이 추정 방법은 두 열에 대해 결측값이 없는 모든 관측값을 사용하여 각 열 쌍에 대한 Pearson 상관을 계산합니다. 통계 상세 정보는 "Pearson 곱적률 상관계수 " 에서 확인하십시오. 쌍별 추정은 결측값이 있는 데이터 테이블에서 행보다 열이 더 많거나, 열이 10 개를 초과하거나, 행이 5,000 개를 초과하는 경우에 가장 유용합니다.

와이드 와이드 추정은 결측값이 있는 관측값을 사용하지 않으므로 결측 셀이 포함된 행은 방법을 적용하기 전에 삭제됩니다. 이 추정 방법은 전체 특이값 분해를 기반으로 하는 알고리즘을 사용합니다. 이 알고리즘은 공분산 행렬을 계산하지 않으므로 계산이 효율적입니다. 데이터에 열이 매우 많은 경우에 유용합니다. 통계 상세 정보는 " 와이드 " 에서 확인하십시오.

JMP PRO **희소** 희소 추정은 결측값이 있는 경우에도 모든 데이터를 사용합니다. 이 추정 방법은 부분 특이값 분해를 기반으로 하는 알고리즘을 사용하여 첫 번째 지정한 수의 특이값과 특이값 벡터만 계산합니다. 이 알고리즘은 공분산 행렬 및 불필요한 주성분을 계산하지 않으므로 계산이 효율적입니다. 데이터가 희소할 때, 즉 0 이 많이 포함되거나 데이터에 열이 많은 경우에 유용합니다. 통계 상세 정보는 " 희소 " 에서 확인하십시오.

참고 : "REML", "ML" 또는 " 로버스트 " 를 선택한 경우 데이터 테이블에 행보다 열이 더 많고 결측값이 있으면 JMP 에서 " 추정 방법 " 이 " 쌍별 " 로 전환됩니다.

성분 수 (" 희소 " 를 추정 방법으로 지정한 경우에만 사용 가능) 추정할 성분 수를 지정합니다. 일반적으로 성분 수는 데이터의 차원보다 훨씬 적습니다.

결측 데이터

다양한 방식으로 결측 데이터를 처리할 수 있도록 여러 가지 추정 방법이 제공됩니다. 다음과 같은 방법으로 결측값을 추정할 수도 있습니다.

- **다변량 방법 > 다변량** 아래에 있는 " 결측 데이터 대치 " 옵션을 사용합니다. 자세한 내용은 " 결측 데이터 대치 " 에서 확인하십시오.
- **분석 > 선별 > 결측값 탐색**에 있는 " 다변량 정규 대치법 " 또는 " 다변량 SVD 대치법 " 유틸리티를 사용합니다. 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

주성분 보고서

"와이드" 또는 "희소" 이외의 추정 방법을 선택한 경우 "주성분: 상관계수 행렬 이용" 보고서가 처음에 나타납니다. "주성분"의 빨간색 삼각형 메뉴에서 "주성분" 옵션에 대해 "공분산 행렬 이용" 또는 "비척도화 행렬 이용"을 선택하면 이 보고서의 제목이 바뀝니다.

"와이드" 방법을 선택하면 "와이드 주성분" 보고서가 나타납니다. "희소" 방법을 선택하면 "희소 주성분" 보고서가 나타납니다.

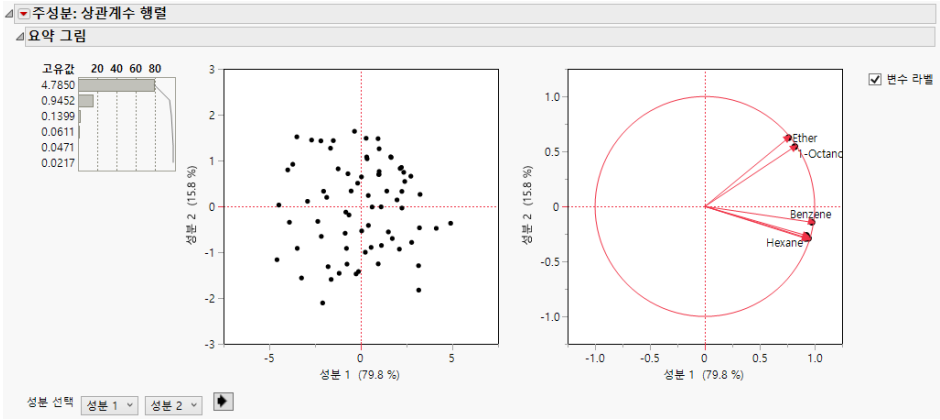
초기 주성분 보고서는 상관계수 행렬을 이용한 분석을 보여 줍니다. 여기에는 주성분을 사용하여 지정된 *Y* 변수의 변동이 요약되어 있습니다(그림 4.4 참조). 빨간색 삼각형 메뉴에서 "주성분" 옵션을 선택하여 공분산 행렬 또는 비척도화 데이터에 기반한 분석으로 전환할 수 있습니다.

선택한 옵션에 따라 다음 고유값 분해 중 하나에서 주성분이 도출됩니다 .

- 상관 행렬
- 공분산 행렬
- 비척도화되고 비중심화된 데이터에 대한 제곱합 및 교차곱 행렬

보고서에는 주성분이 데이터의 변동을 흡수하는 방식이 상세하게 표시됩니다. 변수의 고유 벡터 선형 결합에서 주성분 점이 파생됩니다.

그림 4.4 상관계수 행렬을 이용한 주성분 보고서



이 보고서에는 각 주성분이 설명하는 변동의 백분율을 나타내는 막대 차트와 고유값이 제공됩니다. 스코어 그림과 적재 그림도 있습니다. 고유값은 각 성분이 기여하는 분산의 양을 기반으로 추출된 총 성분 수를 나타냅니다.

스코어 그림에는 다른 성분과의 관계에 따라 평균 및 표준편차 값을 조정하여 계산된 각 성분의 값이 그래프로 표시됩니다.

적재 그림에는 변수와 성분 사이의 회전되지 않은 적재 행렬이 그래프로 표시됩니다. 값이 1에 가까울수록 성분이 변수에 미치는 효과가 커집니다.

기본적으로 이 보고서에는 처음 두 주성분에 대한 스코어 그림과 적재 그림이 표시됩니다. "성분 선택" 옆의 목록을 사용하여 스코어 그림 및 적재 그림에 그래프로 표시되는 주성분을 지정할 수 있습니다.

주성분 보고서 옵션

"주성분"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

참고 : " 와이드 " 또는 " 희소 " 추정 방법에서는 일부 옵션을 사용할 수 없습니다 .

주성분 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) **상관계수** , **공분산** 또는 **비척도화** 데이터를 기반으로 주성분을 생성할 수 있습니다 .

상관 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 변수 간의 상관 행렬입니다 .

참고 : 대각선 값은 1.0 입니다 .

공분산 행렬 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 변수의 공분산을 표시하거나 숨깁니다 .

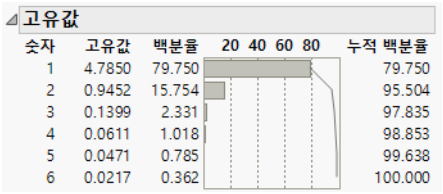
고유값 각 주성분에 해당하는 고유값을 가장 큰 것부터 순서대로 나열합니다 . 고유값은 다변량 표본에서 총 변동의 분할을 나타냅니다 .

고유값의 척도는 주성분을 추출하기 위해 선택한 행렬에 따라 다릅니다 .

- " 상관계수 행렬 이용 " 옵션을 선택하는 경우 고유값이 변수 수의 합계로 척도화됩니다 .
- " 공분산 행렬 이용 " 옵션을 선택하는 경우 고유값이 척도화되지 않습니다 .
- " 비척도화 행렬 이용 " 옵션을 선택하는 경우 고유값을 총 관측값 수로 나눕니다 .

빨간색 삼각형 메뉴에서 **Bartlett 검정** 옵션을 선택하면 각 고유값에 대한 가설 검정 (그림 4.6) 이 제공됩니다 (Jackson 2003 참조).

그림 4.5 고유값



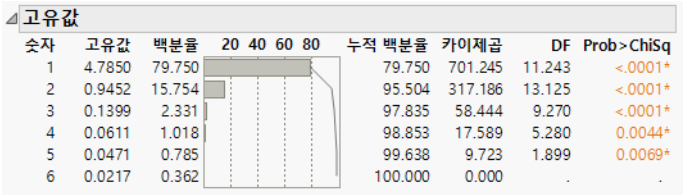
고유 벡터 각 주성분에 대한 고유 벡터 테이블을 왼쪽에서 오른쪽 순서로 표시하거나 숨깁니다 . 이러한 계수를 사용하여 원래 변수의 선형 결합을 구성하면 주성분 변수가 생성됩니다 . 표준 규칙에 따라 고유 벡터의 norm 은 1 입니다 .

참고 : 표시된 고유 벡터 수는 상관 행렬의 계수와 같거나 , " 희소 " 방법을 선택한 경우에는 시작 창에 지정된 성분 수와 같습니다 .

Bartlett 검정 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 동질성 검정 결과를 표시하거나 숨깁니다 (" 고유값 " 테이블에 추가됨). 이 검정은 검정에 대한 카이제곱 , DF(자유도)

및 *p* 값 (Prob > ChiSq) 을 계산하여 고유값의 분산이 동일한지 여부를 판별합니다 . 자세한 내용은 Bartlett(1937, 1954) 연구 자료에서 확인하십시오 .

그림 4.6 Bartlett 검정



고유값

숫자	고유값	백분율	20 40 60 80	누적 백분율	카이제곱	DF	Prob>ChiSq
1	4.7850	79.750		79.750	701.245	11.243	<.0001*
2	0.9452	15.754		95.504	317.186	13.125	<.0001*
3	0.1399	2.331		97.835	58.444	9.270	<.0001*
4	0.0611	1.018		98.853	17.589	5.280	0.0044*
5	0.0471	0.785		99.638	9.723	1.899	0.0069*
6	0.0217	0.362		100.000	0.000	.	.

적재 행렬 각 성분에 대한 적재 테이블을 표시하거나 숨깁니다 . 이러한 값은 적재 그림에 그래프로 표시됩니다 . 테이블 값의 투명도는 0 과 절대 적재 값 사이의 거리를 나타냅니다 . 0 에 더 가까운 절대 적재 값은 0 에서 더 멀리 있는 절대 적재 값보다 투명합니다 .

보조 변수를 지정하면 각 연속형 보조 변수와 범주형 보조 변수의 각 수준에 대한 추가 좌표 테이블이 표시됩니다. 이러한 값은 적재 그림에서 연속형 보조 변수에 대한 그래프로 표시됩니다 .

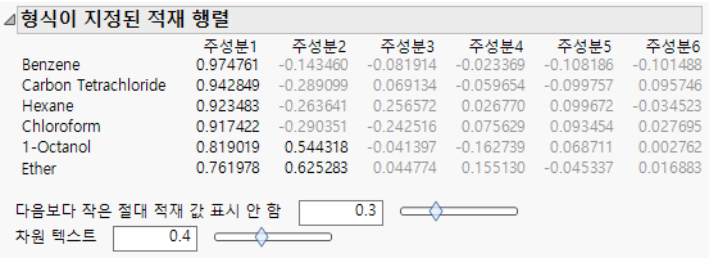
적재 및 좌표의 척도는 주성분을 추출하기 위해 선택한 행렬에 따라 다릅니다 .

- " 상관계수 행렬 이용 " 옵션을 선택하는 경우 적재의 *i* 번째 열은 *i* 번째 고유 벡터에 *i* 번째 고유값의 제곱근을 곱한 값입니다 . *i,j* 번째 적재는 *i* 번째 변수와 *j* 번째 주성분 간의 상관입니다 .
- " 공분산 행렬 이용 " 옵션을 선택하는 경우 적재의 *i* 번째 열에서 *j* 번째 항목은 *i* 번째 고유 벡터에 *i* 번째 고유값의 제곱근을 곱하고 *j* 번째 변수의 표준편차로 나눈 값입니다 . *i,j* 번째 적재는 *i* 번째 변수와 *j* 번째 주성분 간의 상관입니다 .
- " 비척도화 행렬 이용 " 옵션을 선택하는 경우 적재의 *i* 번째 열에서 *j* 번째 항목은 *i* 번째 고유 벡터에 *i* 번째 고유값의 제곱근을 곱하고 *j* 번째 변수의 표준 오차로 나눈 값입니다 . *j* 번째 변수의 표준 오차는 제곱합 및 교차곱 행렬의 *j* 번째 대각 항목을 행 수로 나눈 값입니다 (**X**′**X**/n).

참고 : 비척도화 데이터를 분석하는 경우 *i,j* 번째 적재는 *i* 번째 변수와 *j* 번째 주성분 간의 상관이 아닙니다 .

형식이 지정된 적재 행렬 각 성분에 대한 적재 테이블을 표시하거나 숨깁니다 . 이 테이블은 첫 번째 주성분에 대한 적재가 큰 것부터 내림차순으로 정렬됩니다 . 따라서 변수는 첫 번째 성분에 대한 적재가 큰 것부터 내림차순으로 나열됩니다 .

그림 4.7 형식이 지정된 적재 행렬


형식이 지정된 적재 행렬

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6
Benzene	0.974761	-0.143460	-0.081914	-0.023369	-0.108186	-0.101488
Carbon Tetrachloride	0.942849	-0.289099	0.069134	-0.059654	-0.099757	0.095746
Hexane	0.923483	-0.263641	0.256572	0.026770	0.099672	-0.034523
Chloroform	0.917422	-0.290351	-0.242516	0.075629	0.093454	0.027695
1-Octanol	0.819019	0.544318	-0.041397	-0.162739	0.068711	0.002762
Ether	0.761978	0.625283	0.044774	0.155130	-0.045337	0.016883

다음보다 작은 절대 적재 값 표시 안 함 0.3
차원 텍스트 0.4

다음보다 작은 절대 적재 값 표시 안 함 "형식이 지정된 적재 행렬" 보고서에 사용할 수 없는 적재를 결정하는 값입니다. 텍스트 상자 또는 슬라이더를 사용하여 절대값이 선택한 값 미만인 적재를 흐리게 표시할 수 있습니다.

텍스트 흐리게 " 형식이 지정된 적재 행렬 " 보고서에서 흐리게 표시된 값의 투명도입니다. 텍스트 상자 또는 슬라이더를 사용하여 흐리게 표시되는 적재의 투명도를 설정할 수 있습니다. 투명도 범위는 0 ~ 1 이며, 낮은 값이 높은 값보다 더 투명합니다. 예를 들어 투명도를 0 으로 설정하면 사용할 수 없는 적재가 행렬에서 완전히 제거되지만 투명도를 1 로 설정하면 적재를 계속 사용할 수 있습니다.

변수의 제곱코사인 변수의 제곱코사인을 포함하는 테이블을 표시하거나 숨깁니다. 보조 변수를 지정하면 보조 변수의 제곱코사인 테이블이 추가로 표시됩니다. 전체 주성분에 대한 제곱코사인 값의 합은 각 변수에 대해 1 과 같습니다. 제곱코사인을 사용하면 주성분이 변수를 얼마나 잘 표현하는지 확인할 수 있습니다. 또한 특정 변수를 나타내는 데 필요한 주성분 수를 결정할 수 있습니다. 이 옵션은 처음 세 주성분에 대한 제곱코사인 그림도 표시합니다.

참고 : " 희소 " 추정 방법을 사용하고 선택한 성분 수가 3 개 미만이면 지정된 수의 성분만 그림에 표시됩니다.

변수의 부분 기여도 변수의 부분 기여도를 포함하는 테이블을 표시하거나 숨깁니다. 부분 기여도를 사용하면 각 변수가 각 주성분에 기여하는 비율을 확인할 수 있습니다. 이 옵션은 처음 세 주성분에 대한 부분 기여도 그림도 표시합니다.

참고 : " 희소 " 추정 방법을 사용하고 선택한 성분 수가 3 개 미만이면 지정된 수의 성분만 그림에 표시됩니다.

요약 그림 생성된 요약 정보를 기본 보고서에 표시하거나 숨깁니다. 이 요약 정보에는 고유값 그림, 스코어 그림 및 적재 그림이 포함됩니다. 기본적으로 이 보고서에는 처음 두 주성분에 대한 스코어 그림과 적재 그림이 표시됩니다. 표시할 주성분을 지정하는 옵션이 보고서에 있습니다. 자세한 내용은 " 주성분 보고서 " 에서 확인하십시오.

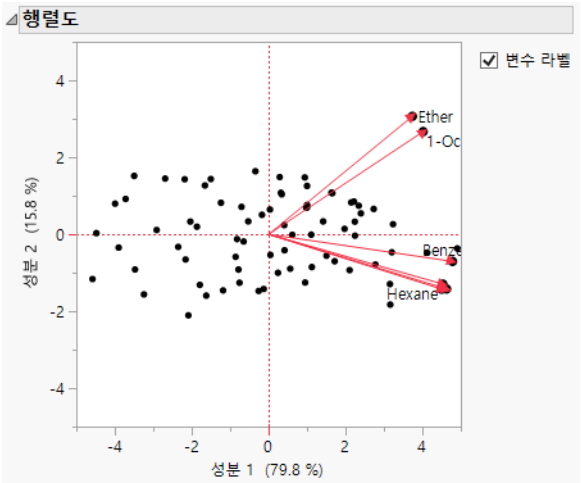
참고 : 데이터에 결측값이 포함되어 있으면 결측값 대치된 스코어가 스코어 그림에 표시됩니다.

팁 : 적재 그림에서 화살표 끝을 선택하면 데이터 테이블에서 해당 열이 선택됩니다 . Ctrl 키를 누른 채 화살표 끝을 클릭하면 열이 선택 취소됩니다 .

행렬도 지정된 수의 성분에 대한 스코어 그림과 적재 그림을 중첩하는 그림을 표시하거나 숨깁니다 .

참고 : 데이터에 결측값이 포함되어 있으면 결측값 대치된 스코어가 행렬도에 표시됩니다 .

그림 4.8 행렬도

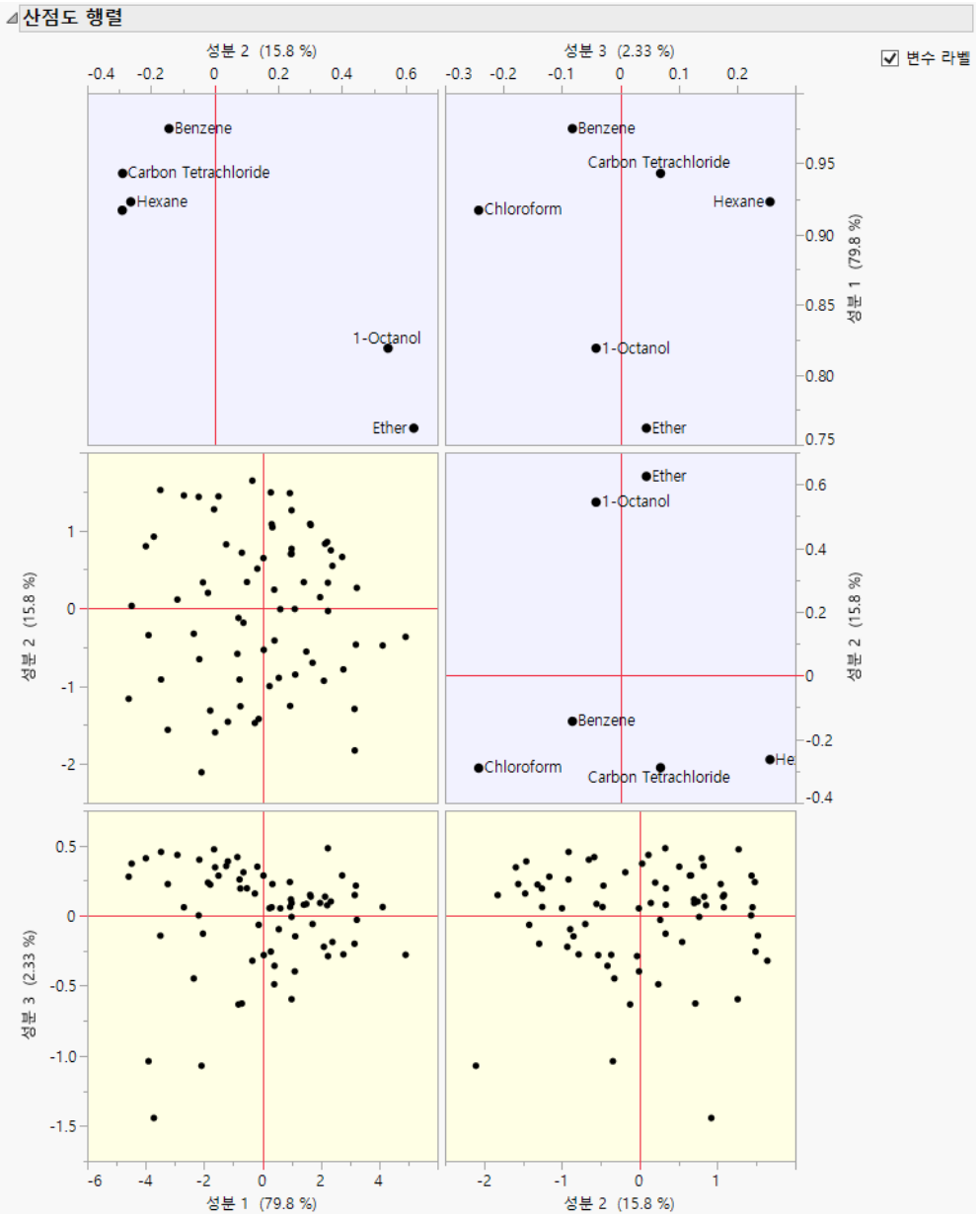


참고 : 스코어 그림 표식은 점이고 적재 그림 표식은 다이아몬드입니다 .

산점도 행렬 지정된 수의 주성분에 대한 스코어 그림과 적재 그림의 행렬을 표시하거나 숨깁니다 . 산점도 행렬은 스코어 그림과 적재 그림을 모두 한 공간에 배열합니다 . 스코어 그림에는 노란색 음영 배경이 있습니다 . 적재 그림에는 파란색 음영 배경이 있습니다 .

참고 : 데이터에 결측값이 포함되어 있으면 결측값 대치된 스코어가 산점도 행렬에 표시됩니다 .

그림 4.9 산점도 행렬



참고 : 산점도 행렬에 표시되는 적재 그림 행렬은 " 적재 그림 " 옵션을 선택하여 얻는 적재 그림 행렬의 전치입니다 .

스크리 그림 각 성분에 대한 고유값 그래프를 표시하거나 숨깁니다. 이 스크리 그림은 데이터 공간의 차원을 시각화하는 데 도움이 됩니다 .

스코어 그림 지정된 수의 성분에 대한 주성분 쌍의 스코어 산점도 행렬을 표시하거나 숨깁니다. 이 그림은 그림 4.4 에 나와 있습니다 (맨 왼쪽 그림).

적재 그림 지정된 수의 성분에 대한 요인 적재의 2 차원 표현 행렬을 표시하거나 숨깁니다. 변수가 30 개 이하이면 적재 그림의 변수에 라벨이 지정됩니다. 변수가 30 개를 초과하면 라벨이 기본적으로 해제되어 있습니다. 이 정보는 그림 4.4 에 나와 있습니다 (맨 오른쪽 그림).

팁 : 적재 그림에서 화살표 끝을 선택하면 데이터 테이블에서 해당 열이 선택됩니다. Ctrl 키를 누른 채 화살표 끝을 클릭하면 열이 선택 취소됩니다.

결측값 대치를 사용한 스코어 그림 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 결측값을 대치하고 스코어 그림을 생성합니다. 이 옵션은 결측값이 있는 경우에만 사용할 수 있습니다.

3D 스코어 그림 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 세 가지 주성분 스코어의 3D 산점도를 표시하거나 숨깁니다. 명령을 처음 호출하면 처음 세 개의 주성분이 표시됩니다.

참고 : 데이터에 결측값이 포함되어 있으면 결측값 대치된 스코어가 3D 스코어 그림에 표시됩니다.

그림 4.10 3D 산점도 스코어 그림

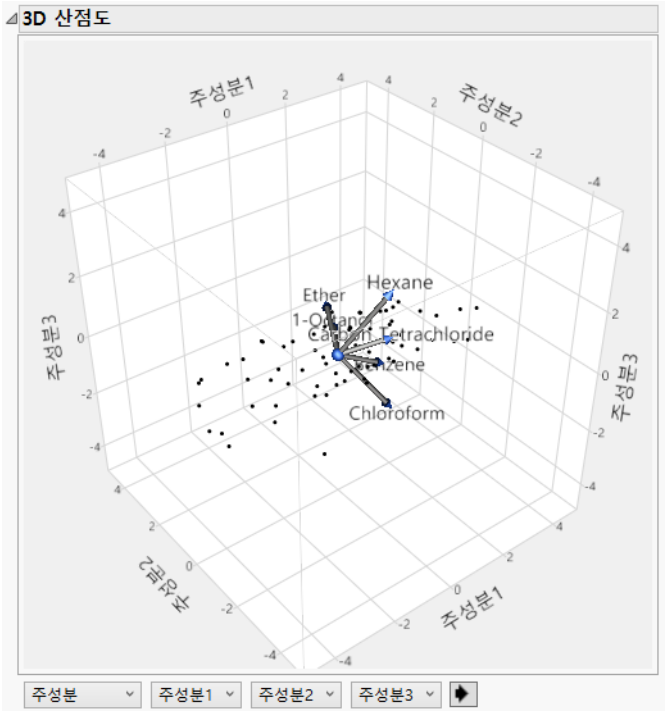


그림 소스 그림에 있는 데이터 점의 소스입니다. 사용 가능한 옵션은 " 주성분 ", " 회전된 주성분 " 및 " 데이터 열 " 입니다.

축 컨트롤 각 축의 콘텐츠입니다 . " 주성분 " 옵션 또는 " 회전된 성분 " 옵션을 선택한 경우 " 축 컨트롤 " 옵션은 주성분입니다 . " 데이터 열 " 옵션을 선택한 경우 옵션은 분석의 변수입니다 .

순환 버튼 가능한 모든 축 콘텐츠 항목을 순환합니다 .

변수는 그림에 선으로 표시됩니다. **행렬도 선**이라고 하는 이러한 선은 축에서 주성분의 함수로 변수를 근사화합니다 . 변수가 두 개 또는 세 개만 있는 경우 선은 변수를 정확하게 나타냅니다 . 이 선은 주성분 적재에 해당합니다 .

스코어 타원 각 주성분 쌍에 대한 요약 스코어 그림에서 타원을 표시하거나 숨깁니다. 타원은 유의 수준에 기반한 신뢰도 타원 또는 관측값이 중심에서 얼마나 떨어져 있는지에 기반한 관리 한계 타원으로 구성됩니다 . 기본적으로 타원은 95% 신뢰도 타원입니다 .

스코어 타원 범위 스코어 타원의 구성 방법을 변경할 수 있는 하위 메뉴를 표시합니다 . 신뢰 수준을 지정하거나 , 중심과의 거리를 *k*- 시그마로 지정하여 스코어 타원을 지정합니다 . 신뢰 수준 p 와 *k*- 시그마 간의 관계는 $p = 1 - \exp(-k^2/2)$ 입니다 .

표시 옵션

화살표 선 화살표를 표시할 수 있는 모든 그림에서 화살표를 표시하거나 숨깁니다 . 변수가 1000 개 이하이면 화살표가 표시됩니다 . 변수가 1000 개를 초과하면 화살표가 기본적으로 표시되지 않습니다 .

보조 변수 표시 (보조 변수를 지정한 경우에만 사용 가능) 행렬도 , 스코어 그림 및 적재 그림에서 화살표 선 (연속형 보조 변수의 경우) 또는 라벨 표식 (범주형 보조 변수의 경우) 을 표시하거나 숨깁니다 .

이상치 분석 T^2 및 기여도 통계량을 통해 데이터에서 이상치를 감지할 수 있는 " 이상치 분석 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 이상치 분석 " 에서 확인하십시오 .

요인 분석 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 주성분의 요인 분석 스타일 회전 또는 요인 분석을 수행합니다 . 자세한 내용은 " 요인 분석 " 에서 확인하십시오 .

JMP PRO **변수 군집화** (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 변수를 서로 겹치지 않는 군집으로 나누어 변수에 대한 군집 분석을 수행합니다 . 변수 군집화는 유사한 변수를 대표 그룹으로 그룹화하는 방법을 제공합니다 . 그러면 각 군집을 단일 성분 또는 변수로 나타낼 수 있습니다 . 성분은 군집에 있는 모든 변수의 선형 결합입니다 . 아니면 군집을 해당 군집에서 가장 대표적인 멤버로 식별되는 변수로 나타낼 수도 있습니다 . 자세한 내용은 " 변수 군집화 " 에서 확인하십시오 .

참고 : 변수 군집화에서는 " 공분산 행렬 이용 " 또는 " 비척도화 행렬 이용 " 옵션을 선택한 경우에도 모든 계산에 상관 행렬을 사용합니다 .

모형 기반 다변량 관리도 지정된 수의 주성분에 대한 계산식을 저장하고 MDMCC(모형 기반 다변량 관리도) 시작 창을 실행합니다 . 주성분 계산식은 MDMCC 시작 창에서 공정 열로 할당됩니다 . 공정을 추가 또는 제거하고 , 시간 ID 를 추가하거나 , 과거 데이터 끝 위치를 설정한 후 " 확인 " 을 클릭할 수 있습니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

주성분 저장 지정한 수의 주성분을 각 성분에 대한 계산식과 함께 데이터 테이블에 저장합니다. 계산식은 결측값이 있는 행을 계산할 수 없습니다.

주성분 계산은 주성분을 추출하기 위해 선택한 행렬에 따라 다릅니다.

- "상관계수 행렬 이용" 옵션을 선택하는 경우 *i* 번째 주성분은 *i* 번째 고유 벡터의 항목을 계수로 사용하여 중심화 및 척도화된 관측값의 선형 결합입니다.
- "공분산 행렬 이용" 옵션을 선택하는 경우 *i* 번째 주성분은 *i* 번째 고유 벡터의 항목을 계수로 사용하여 중심화된 관측값의 선형 결합입니다.
- "비척도화 행렬 이용" 옵션을 선택하는 경우 *i* 번째 주성분은 *i* 번째 고유 벡터의 항목을 계수로 사용하는 원시 관측값의 선형 결합입니다.

참고 : 지정된 성분 수가 상관 행렬의 계수를 초과하는 경우 저장되는 성분 수는 상관 행렬의 계수로 설정됩니다.

예측값 저장 주성분 수가 지정되면 예측된 변수를 데이터 테이블의 새 열에 저장합니다.

예측값을 성분 계산식으로 저장 지정된 수의 성분을 선형 결합한 예측된 변수를 계산식으로 저장합니다. 계산식에 사용된 주성분도 데이터 테이블에 저장됩니다.

DModX 저장 주성분 수가 지정되면 DModX(주성분 모형에 대한 관측값 거리) 를 데이터 테이블의 새 열에 저장합니다. DModX 값이 클수록 데이터의 이상치가 사소하거나 중간 정도임을 나타냅니다. 자세한 내용은 "DModX 계산 " 에서 확인하십시오.

개별 제곱코사인 저장 주성분 수가 지정되면 개별 제곱코사인을 데이터 테이블의 새 열에 저장합니다.

개별 부분 기여도 저장 주성분 수가 지정되면 개별 부분 기여도를 데이터 테이블의 새 열에 저장합니다.

회전된 성분 저장 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 회전된 성분을 성분에 대한 계산식과 함께 데이터 테이블에 저장합니다. 이 옵션은 " 요인 분석 " 옵션을 사용한 후에만 사용할 수 있습니다. 계산식은 결측값이 있는 행을 계산할 수 없습니다.

결측값 대치를 사용한 주성분 저장 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 결측값을 대치하고 주성분을 데이터 테이블에 저장합니다. 결측값 대치 및 주성분 계산을 수행하기 위한 계산식이 열에 포함됩니다. 이 옵션은 결측값이 있는 경우에만 사용할 수 있습니다.

결측값 대치를 사용한 회전된 성분 저장 (" 와이드 " 또는 " 희소 " 추정 방법에는 사용 불가능) 결측값을 대치하고 회전된 성분을 데이터 테이블에 저장합니다. 결측값 대치 및 회전된 성분 계산을 수행하기 위한 계산식이 열에 포함됩니다. 이 옵션은 " 요인 분석 " 옵션을 사용한 후 결측값이 있는 경우에만 사용할 수 있습니다.

결측값 대치 계산식 저장 (데이터 테이블에 결측값이 포함된 경우에만 사용 가능) 결측값이 포함된 열에 대해 결측값 추정에 사용된 계산식을 포함하는 새 열을 데이터 테이블에 저장합니다. 새 열 이름은 **결측값 대치됨 < 열 이름 >** 입니다.

JMP PRO **성분 계산식 게시** 지정된 수의 주성분 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다. 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

JMP PRO **DModX 계산식 게시** 지정된 수의 주성분에 기반한 DModX 계산식을 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

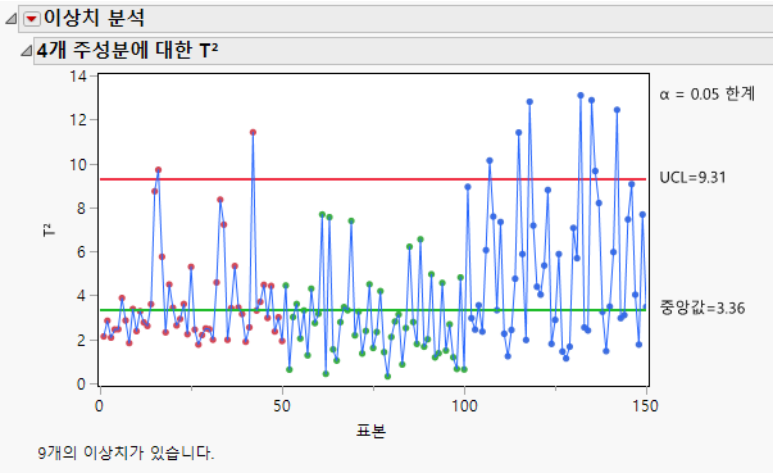
다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

이상치 분석

그림 4.11 이상치 분석 보고서



기본적으로 "이상치 분석" 보고서에는 "<*A*>개 주성분에 대한 T^2" 그림이 표시됩니다. 이 그림에는 각 관측값에 대한 T^2 값이 표시되고 중앙값 및 UCL(관리 상한)을 나타내는 가로선이 있습니다. T^2 값, 중앙값 및 UCL을 계산하는 방법에 대한 자세한 내용은 "이상치 분석을 위한 계산"에서 확인하십시오.

UCL을 계산하는 데 사용된 α 수준이 그림 옆에 표시됩니다. 감지된 이상치 수는 그림 아래에 표시됩니다. 이 수는 T^2 값이 UCL보다 큰 관측값 수입니다.

팁 : T^2 그림의 점을 커서로 가리키면 해당 관측값에 대한 T^2 기여 비율 그림을 볼 수 있습니다 . T^2 기여 비율 그림을 클릭하면 보고서 창에 그림이 추가됩니다 .

이상치 분석 보고서 옵션

T² 그림 T^2 그림을 표시하거나 숨깁니다 . 기본적으로 설정되어 있습니다 .

기여도 히트맵 모든 관측값에 대한 T^2 기여도 값의 히트맵을 표시하거나 숨깁니다 .

기여 비율 히트맵 모든 관측값에 대한 T^2 기여도 값을 개별 행의 T^2 비율로 표현한 히트맵을 표시하거나 숨깁니다 . 개별 행의 비율은 기여도 제곱을 계산한 후 해당 개별 행에 대한 모든 기여도의 제곱합으로 나누어 구합니다 .

선택된 표본에 대한 기여도 그림 (T^2 그림에서 점을 하나 이상 선택한 경우에만 사용 가능) 선택된 각 표본에 대한 T^2 기여도 그림이 포함된 보고서를 표시합니다 . T^2 기여도 그림은 표본의 T^2 통계량에 대한 각 변수의 기여도를 보여 줍니다 . 기여도를 계산하는 방법에 대한 자세한 내용은 " 이상치 분석을 위한 계산 " 에서 확인하십시오 . 기여도 그림을 사용하여 이상치를 조사할 수 있습니다 . 양의 기여도 또는 음의 기여도가 가장 큰 변수가 T^2 값이 큰 표본에 가장 많이 기여하는 변수입니다 . 빨간색 삼각형 메뉴 옵션에 대한 자세한 내용은 " 기여도 그림 보고서 옵션 " 에서 확인하십시오 .

선택된 표본에 대한 기여 비율 그림 (T^2 그림에서 점을 하나 이상 선택한 경우에만 사용 가능) 선택된 각 표본에 대한 T^2 기여 비율 그림이 포함된 보고서를 표시합니다 . T^2 기여 비율 그림에는 선택된 관측값에 대한 기여도 값이 개별 행의 T^2 비율로 표시됩니다 . 이것은 기여 비율 히트맵의 정보를 또 다르게 표현한 것입니다 . 빨간색 삼각형 메뉴 옵션에 대한 자세한 내용은 " 기여도 그림 보고서 옵션 " 에서 확인하십시오 .

정규화된 DModX 그림 (성분 수가 변수 수보다 작은 경우에만 사용 가능) 정규화된 DModX 값의 그림을 표시하거나 숨깁니다 . DModX 값은 데이터에서 중간 이상치를 감지하는 데 유용합니다 .

성분 수 T^2 및 T^2 기여도 통계량에 사용되는 주성분 수를 지정할 수 있습니다 . 성분 수를 변경하면 T^2 그림 , 히트맵 및 정규화된 DModX 그림이 자동으로 업데이트됩니다 .

α 수준 설정 α 수준을 지정할 수 있습니다 .

T² 저장 T^2 값을 데이터 테이블의 새 열에 저장합니다 .

기여도 저장 T^2 기여도를 데이터 테이블의 새 열에 저장합니다 . 각 Y 변수에 대해 열이 하나씩 있습니다 .

정규화된 DModX 저장 (성분 수가 변수 수보다 작은 경우에만 사용 가능) 정규화된 DModX 값을 데이터 테이블의 새 열에 저장합니다 .

기여도 그림 보고서 옵션

"선택된 표본에 대한 T^2 기여도 그림", "선택된 표본에 대한 T^2 기여 비율 그림" 및 "선택된 표본에 대한 T^2 평균 기여 비율 그림"에는 다음과 같은 빨간색 삼각형 메뉴 옵션이 포함되어 있습니다.

막대 라벨 지정 기여도 그림의 막대에 라벨을 지정하는 옵션이 포함된 하위 메뉴를 표시합니다. 라벨 옵션은 "라벨 없음", "값별 라벨" 및 "열별 라벨"입니다.

그림 제거 보고서에서 기여도 그림을 제거합니다.

선택한 항목에 대한 관리도 (기여도 그림에서 막대 세그먼트를 하나 이상 선택한 경우에만 사용 가능) 선택한 각 공정 및 그룹에 대한 관리도 결과가 포함된 관리도 빌더 창을 엽니다.

팁 : 기여도 그림에서 막대 세그먼트를 커서로 가리켜 관리도를 볼 수도 있습니다. 관리도를 클릭하면 관리도 빌더 창이 열립니다.

주성분 분석 플랫폼에 대한 통계 상세 정보

- "추정 방법"
- "DModX 계산"
- "이상치 분석을 위한 계산"

추정 방법

REML

데이터에 결측값이 포함된 경우 REML(제한 최대 가능도) 추정값은 ML(최대 가능도) 추정 방법보다 덜 편향됩니다. REML 방법은 오차 대비를 기반으로 주변 가능도를 최대화합니다. REML 방법은 주로 분산 및 공분산을 추정하는 데 사용됩니다. 주성분 플랫폼의 REML 방법은 비정형 공분산 행렬을 사용하는 반복 측정 데이터에 대한 혼합 모형의 REML 추정과 동일합니다. 혼합 모형의 REML 추정에 대한 자세한 내용은 SAS PROC MIXED 관련 문서에서 확인하십시오.

와이드

와이드 방법은 공분산 행렬을 계산하지 않는 계산 효율적인 알고리즘을 사용합니다. 이 알고리즘은 특이값 분해를 기반으로 합니다. 다음 표기를 고려하십시오.

- n = 행 수
- p = 변수 수
- $\mathbf{X}$ = 데이터 값의 $n \times p$ 행렬

0이 아닌 고유값 수와 이에 따른 주성분 수는 $\mathbf{X}$의 상관 행렬 계수와 같습니다. 0이 아닌 고유값 수는 n과 p 중 더 작은 값을 초과할 수 없습니다.

와이드 방법을 선택하면 데이터가 표준화됩니다 . 값을 표준화하려면 평균을 뺀 후 표준편차로 나눕니다 . 표준화된 데이터 값의 $n \times p$ 행렬을 $\mathbf{X}_s$ 로 나타냅니다 . 그러면 표준화된 데이터의 공분산 행렬은 $\mathbf{X}$ 의 상관 행렬이며 다음과 같이 정의됩니다 .

$$Cov = \mathbf{X}_s'\mathbf{X}_s/(n-1)$$

특이값 분해를 사용하면 $\mathbf{X}_s$가 $\mathbf{U}Diag(\Lambda)\mathbf{V}'$로 표기됩니다. 이 표현은 $\mathbf{X}_s'\mathbf{X}_s$의 고유 벡터 및 고유값을 얻는 데 사용됩니다. 주성분 또는 스코어는 $\mathbf{X}_s\mathbf{V}$로 제공됩니다. 추가 배경 정보는 "와이드 선형 방법 및 특이값 분해"에서 확인하십시오.

JMP PRO 희소

희소 방법은 와이드 방법과 마찬가지로 특이값 분해를 기반으로 합니다. 따라서 희소 방법 알고리즘은 공분산 행렬을 계산하지 않으므로 계산이 효율적입니다.

" 와이드 " 에 설명된 것과 동일한 $\mathbf{X}$ 표기 및 표준화를 고려하십시오 . $\mathbf{X}$ 의 상관 행렬은 $\mathbf{X}_s$ 의 공분산 행렬로 나타냅니다 .

$$Cov(\mathbf{X}_s) = \mathbf{X}_s'\mathbf{X}_s/(n-1)$$

희소 방법은 특이값 분해 계산에서 와이드 방법과 다릅니다. 와이드 방법은 전체 특이값 분해를 수행합니다. 그러나 희소 방법은 특이값 분해에서 첫 번째 지정한 수의 특이값과 특이 벡터만 계산하는 알고리즘을 사용합니다. 따라서 첫 번째 지정한 수의 고유값과 주성분만 반환됩니다. 이 알고리즘에 대한 자세한 내용은 Baglama 와 Reichel(2005) 연구 자료에서 확인하십시오.

DModX 계산

DModX 는 주성분 모형까지의 관측 거리이며 다음과 같이 정의됩니다 .

$$\text{DModX} = \sqrt{\frac{\sum_k e_{ik}^2}{K-A}}$$

다음은 각 요소에 대한 설명입니다 .

e_{ik} = 모형의 잔차

K = 변수 수

A = 주성분 수

DModX 값이 클수록 데이터의 이상치가 사소하거나 중간 정도임을 나타냅니다.

이상치 분석을 위한 계산

이상치 분석 보고서의 계산에는 다음 표기가 사용됩니다 .

n = 관측값 수

A = 주성분 수

X_{ci} = i 번째 관측값에 대한 표준화된 데이터

T^2 통계량

i 번째 관측값에 대한 T^2 통계량은 다음과 같이 계산됩니다 .

$$T_i^2 = X_{ci} P_A L^{-1} P_A^{\mathrm{T}} X_{ci}^{\mathrm{T}}$$

여기서 P_A는 처음 A개 고유 벡터를 포함하는 행렬이고, L은 처음 A개 고유값을 포함하는 대각 행렬입니다.

T^2 그림의 중앙값과 UCL 은 다음과 같이 계산됩니다 .

$$CL_{T^2, q} = \frac{(n-1)^2}{n} \beta\left[q, \frac{A}{2}, \frac{n-A-1}{2}\right]$$

다음은 각 요소에 대한 설명입니다 .

$\beta\left[q, \frac{A}{2}, \frac{n-A-1}{2}\right]$ = 베타 $\left(\frac{A}{2}, \frac{n-A-1}{2}\right)$ 분포의 q 번째 분위수

중앙값을 계산하려면 $q = 0.5$를 사용합니다. UCL을 계산하려면 $q = (1 - \alpha)$를 사용합니다.

기여도 통계량

i 번째 관측값에 대한 T^2 기여도 통계량 벡터는 다음과 같이 계산됩니다 .

$$con_i = X_{ci} P_A L^{-1/2} P_A^{\mathrm{T}}$$

참고 : 개별값의 기여도 제곱합은 해당 개별값에 대한 T_i^2 과 같습니다 .

5 장

판별 분석

연속형 변수를 기반으로 분류 예측

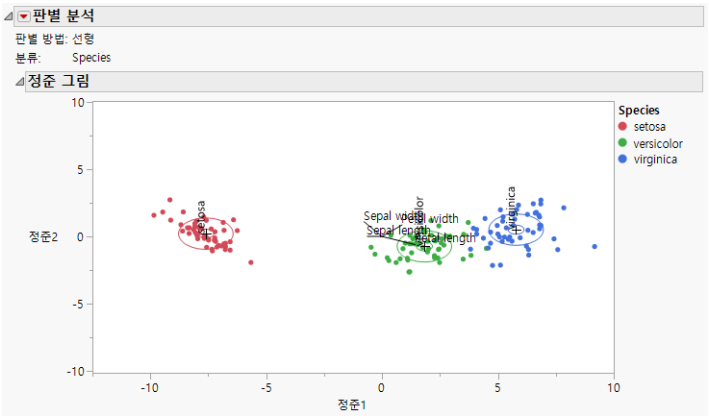
판별 분석은 여러 연속형 변수의 관측값을 기반으로 그룹 또는 범주의 소속을 예측합니다. 특히 판별 분석은 알려진 연속형 반응(Y)을 기반으로 범주형 분류(X) 변수를 예측합니다. 판별 분석을 위한 데이터는 소속 그룹이 알려진 관측값 표본과 연속형 변수에 대한 값으로 구성됩니다.

예를 들어 기대 수익성에 따라 대출 신청인을 저금리 대출, 장기 대출 또는 대출 없음이라는 세 가지 대출 범주(X)로 분류하려고 할 수도 있습니다. 현재 급여, 현재 직장에서의 근무 기간, 나이 및 부채 부담과 같은 연속형 변수(Y)를 사용하여 개인에게 가장 유리한 대출 범주를 예측할 수 있습니다. 판별 분석을 사용하여 개인을 대출 범주로 분류하는 예측 모형을 생성할 수 있습니다.

판별 플랫폼의 기능은 다음과 같습니다 .

- 잘 구별하는 변수를 선택하는 데 도움이 되는 단계별 선택 옵션
- 적합 방법 : 선형 , 2 차 , 정규화 및 와이드 선형
- 정준 그림 및 오분류 요약
- 각 그룹에 대한 제곱 거리 및 판별 스코어
- 예측 거리와 확률을 데이터 테이블에 저장하는 옵션

그림 5.1 정준 그림



목차

판별 분석 플랫폼 개요 73
판별 분석의 예 73
판별 분석 플랫폼 시작 75
단계별 변수 선택 76
판별 방법 80
공분산 축소 82
판별 분석 보고서 82
주성분 83
정준 그림 및 정준 구조 84
판별 스코어 87
스코어 요약 88
판별 분석 옵션 91
스코어 옵션 92
정준 옵션 94
3D 정준 그림의 예 97
사전 확률 지정 98
새 수준 고려 98
판별 행렬 저장 99
산점도 행렬 99
JMP 및 JMP Pro 에서 검증 100
판별 분석 플랫폼에 대한 통계 상세 정보 101
와이드 선형 알고리즘 설명 101
저장된 계산식 101
다변량 검정 108
근사 F-검정 109
그룹 간 공분산 행렬 110

판별 분석 플랫폼 개요

판별 분석은 연속형 변수에 대한 값으로 설명되는 관측값을 그룹으로 분류하려고 시도합니다. 범주형 변수 *X*로 정의되는 소속 그룹은 연속형 변수에 의해 예측됩니다. 이러한 변수를 **공변량**이라고 하며 *Y*로 나타냅니다.

판별 분석은 로지스틱 회귀와 다릅니다. 로지스틱 회귀에서 분류 변수는 확률 변수이며 연속형 변수에 의해 예측됩니다. 판별 분석에서 분류는 고정되어 있으며 공변량(*Y*)이 확률 변수의 구현입니다. 그러나 두 기법 모두 연속형 변수에 의해 범주형 값이 예측됩니다.

판별 플랫폼에서는 모형 적합을 위한 네 가지 방법을 제공합니다. 모든 방법은 Mahalanobis 거리를 사용하여 각 관측값에서 각 그룹의 다변량 평균(**중심**)까지의 거리를 추정합니다. 소속 그룹의 사전 확률을 지정할 수 있으며 이러한 확률은 거리 계산에서 고려됩니다. 관측값은 가장 근접한 그룹으로 분류됩니다.

다음과 같은 적합 방법을 사용할 수 있습니다 .

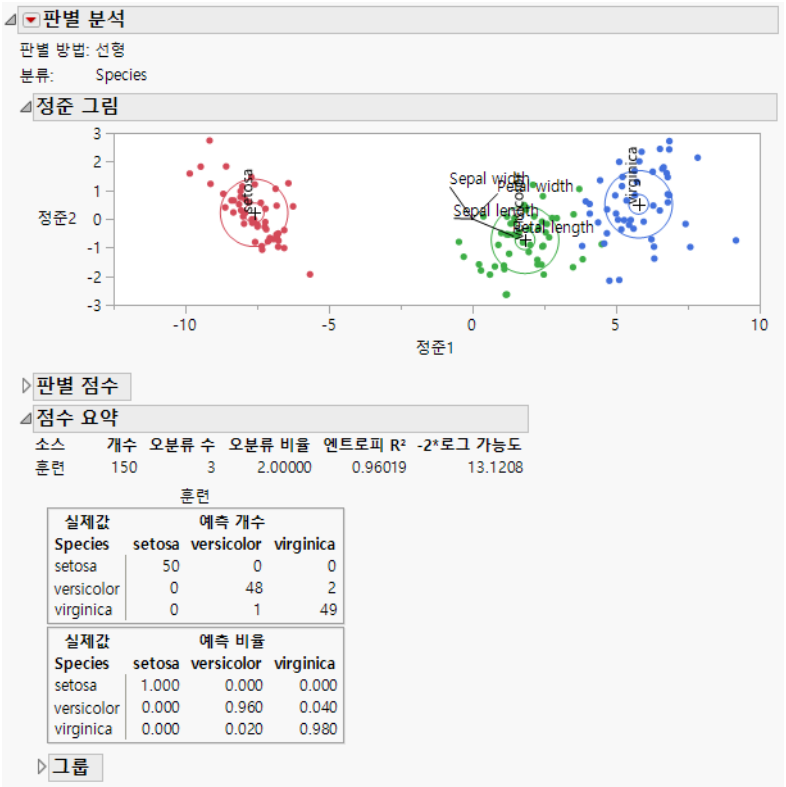
- 선형 — 그룹 내 공분산 행렬이 같다고 가정합니다 . *X* 에 의해 정의되는 그룹의 공변량 평균은 서로 다르다고 간주됩니다 .
- 2 차 — 그룹 내 공분산 행렬이 다르다고 가정합니다 . 이 방법은 선형 방법보다 더 많은 모수를 추정해야 합니다 . 그룹 표본 크기가 작으면 추정값이 불안정할 수 있습니다 .
- 정규화 — 그룹 내 공분산 행렬이 다를 때 추정값에 안정성을 부여하는 두 가지 방법을 제공합니다 . 이 옵션은 그룹 표본 크기가 작을 때 유용합니다 .
- 와이드 선형 — 다른 방법을 사용하면 계산상 문제가 발생할 수 있는 다수의 공변량을 기반으로 모형을 적합시키는 데 유용합니다 . 이 경우 모든 공분산 행렬이 같다고 가정합니다 .

판별 분석의 예

Fisher 의 Iris 데이터 집합을 사용하여 세 가지 다른 종으로 구성된 Iris 화초 표본에서 네 개의 측정값을 추출했습니다 . 목표는 네 개의 측정값을 사용하여 종을 정확하게 식별하는 것입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 판별**을 선택합니다 .
3. Sepal length, Sepal width, Petal length 및 Petal width 를 선택하고 **Y, 공변량**을 클릭합니다 .
4. Species 를 선택하고 **X, 범주**를 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 5.2 판별 분석 보고서 창

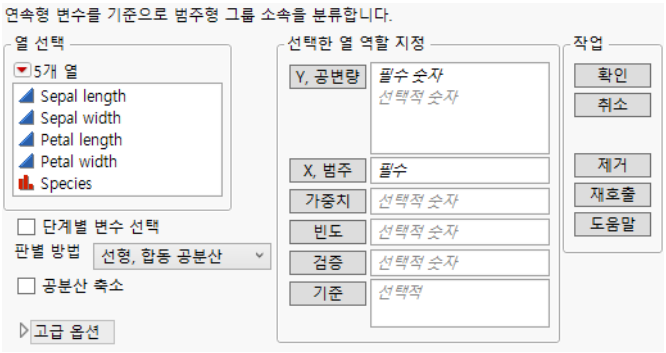


Species에 대한 클래스가 세 개이므로 두 개의 정준 변수가 있습니다. "정준 그림"에서 두 정준 좌표에 대해 각 관측값이 표시됩니다. 이 그림에서는 두 좌표에 세 가지 종이 구분되어 있음을 보여 줍니다. 검증 데이터 집합이 없으므로 "스코어 요약" 보고서에 훈련 데이터 집합에 대한 패널만 표시됩니다. 검증 데이터 집합이 없으면 전체 데이터 집합이 훈련 데이터 집합으로 간주됩니다. 150 개 관측값 중 세 개만 잘못 분류되었습니다.

판별 분석 플랫폼 시작

분석 > 다변량 방법 > 판별을 선택하여 판별 플랫폼을 시작합니다 .

그림 5.3 Iris.jmp 에 대한 판별 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

참고 : " 검증 " 버튼은 JMP Pro 에만 나타납니다 . JMP 에서는 제외된 행을 사용하여 검증 데이터 집합을 정의할 수 있습니다 . 자세한 내용은 "JMP 및 JMP Pro 에서 검증 " 에서 확인하십시오 .

Y, 공변량 관측값을 범주로 분류하는 데 사용할 연속형 변수를 포함하는 열입니다 .

X, 범주 관측값을 분류할 범주 또는 그룹을 포함하는 열입니다 .

가중치 분석을 위해 각 행에 가중치를 할당하는 값이 들어 있는 열입니다 .

빈도 분석을 위해 각 행에 빈도를 할당하는 값이 들어 있는 열입니다 . 일반적으로 빈도 열의 효과는 데이터 테이블을 확장하여 정수 빈도가 *k* 인 행을 *k* 개의 행으로 확장하는 것입니다 . 빈도 비율을 지정할 수 있습니다 .

JMP PRO **검증** 검증 데이터 집합을 정의하는 숫자 열입니다 . 이 열에는 최대 세 개의 구분되는 값이 포함되어야 합니다 .

- 두 개의 값이 있는 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 세 개의 값이 있는 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 검증 열에 네 개 이상의 수준이 있는 경우 가장 작은 세 개의 값을 포함하는 행이 검증 데이터 집합을 정의합니다 . 다른 모든 행은 분석에서 제외됩니다 .

판별 플랫폼에서는 " 단계별 변수 선택 " 이 사용되는 경우 외에는 검증 열을 사용하여 모형을 훈련하고 평가합니다 . 시작 창에서 " 단계별 변수 선택 " 옵션을 선택하는 경우 판별 플랫폼

에서는 검증 열을 사용하여 모형을 훈련 및 조정하거나 모형을 훈련, 조정 및 평가합니다. 검증에 대한 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

팁 : " 단계별 변수 선택 " 을 사용하지 않는 경우 검증 열에는 구분되는 값이 두 개만 포함되어야 합니다.

" 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다. " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

기준 지정된 열의 각 수준에 대해 별도의 분석을 수행합니다.

단계별 변수 선택 공분산 분석 및 p 값을 사용하여 단계별 변수 선택을 수행합니다. 자세한 내용은 " 단계별 변수 선택 " 에서 확인하십시오.

검증 데이터 집합을 지정한 경우 검증 데이터 집합에 대한 통계량도 나타납니다. 검증 데이터 집합 통계량은 " 시작 " 버튼을 사용하는 경우 수행할 단계 수를 결정하는 데 사용됩니다.

참고 : 이 옵션은 " 와이드 선형 " 판별 방법에는 제공되지 않습니다.

판별 방법 판별 분석을 수행하기 위한 네 가지 방법을 제공합니다. 자세한 내용은 " 판별 방법 " 에서 확인하십시오.

공분산 축소 합동 그룹 내 공분산 행렬과 그룹 내 공분산 행렬의 비대각 요소를 축소합니다. 이로 인해 안정성이 향상되고 예측 분산을 줄일 수 있습니다. 자세한 내용은 " 공분산 축소 " 에서 확인하십시오.

고급 옵션 다음 옵션이 포함되어 있습니다.

비중심화된 정준 이전 버전 JMP 와의 호환성을 위해 정준 스코어 중심화를 제한합니다.

유사역행렬 사용 공분산 행렬이 특이 행렬일 때 Moore-Penrose 유사역행렬을 분석에 사용합니다. 결과 스코어에 모든 공변량이 포함됩니다. 이 옵션을 선택하지 않으면 **Y, 공변량** 목록에서 앞에 나오는 공변량의 선형 결합인 공변량이 분석에서 삭제됩니다.

제외된 행으로 교차 검증 제외된 행이 적합 통계량이 계산되는 검증 데이터 집합을 구성하도록 지정합니다.

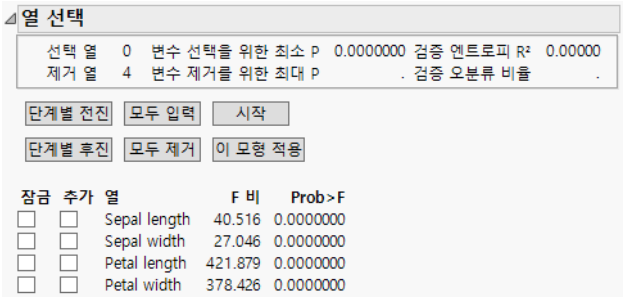
단계별 변수 선택

참고 : " 와이드 선형 " 방법에는 " 단계별 변수 선택 " 을 사용할 수 없습니다.

시작 창에서 "단계별 변수 선택" 옵션을 선택하면 "열 선택" 패널이 표시된 "판별 분석" 보고서가 열립니다. 버튼을 사용하여 변수를 선택하거나, "잠금" 및 "추가" 체크박스로 변수를 수동 선택하여 단계별 분석을 수행할 수 있습니다. 선택한 항목에 따라 F 비와 p 값이 업데이트됩니다. 업데이트 방식에 대한 자세한 내용은 "F 비와 Prob>F 업데이트"에서 확인하십시오.

검증 데이터 집합 유형을 지정하면 "시작" 버튼이 나타납니다. "시작"을 클릭하면 JMP에서 검증 데이터 집합 통계량을 사용하여 수행할 단계 수를 결정합니다.

그림 5.4 검증 데이터 집합을 사용한 Iris.jmp 의 열 선택 패널



F 비와 Prob>F 업데이트

모형에서 변수를 입력하거나 제거하면 다음과 같은 구조의 공분산 모형 분석을 기반으로 F 비와 Prob>F 값이 업데이트됩니다.

- 고려 중인 공변량이 반응입니다.
- 모형에 이미 입력된 공변량이 예측 변수입니다.
- 그룹 변수가 예측 변수입니다.

단계별 보고서에 제공된 "F 비"와 "Prob>F" 값은 그룹 변수에 대한 공분산 분석 검정의 *F* 비와 *p* 값입니다. 그룹 변수에 대한 공분산 분석 검정은 고려 중인 공변량에 상대적인 판별력을 나타내는 지표입니다.

통계량

선택 열 판별 모형에 입력하기 위해 현재 선택된 열의 수입니다.

제외 열 판별 모형에 현재 입력할 수 있는 열의 수입니다.

변수 선택을 위한 최소 P 모형에 입력할 수 있는 모든 공변량의 *p* 값 중 가장 작은 *p* 값입니다.

변수 제거를 위한 최대 P 모형에 입력하기 위해 현재 선택된 모든 공변량의 *p* 값 중 가장 큰 *p* 값입니다.

검증 엔트로피 R² 검증 데이터 집합에 대한 엔트로피 R² 입니다. 값이 클수록 더 나은 적합을 나타냅니다. "엔트로피 R²" 값이 1 이면 분류가 완벽하게 예측되었음을 나타냅니다. 예측 확률의 불확도는 판별 모형에서 일반적으로 사용되므로 엔트로피 R² 값은 작은 경향이 있습니다.

자세한 내용은 " 엔트로피 R²" 에서 확인하십시오. 검증 데이터 집합이 사용되는 경우에만 사용할 수 있습니다.

참고 : 검증 엔트로피 R² 이 음수일 수 있습니다.

검증 오분류 비율 검증 데이터 집합에 대한 오분류 비율입니다 . 값이 작을수록 더 나은 분류를 나타냅니다. 검증 데이터 집합이 사용되는 경우에만 사용할 수 있습니다.

버튼

단계별 전진 아직 입력되지 않은 공변량 중에서 가장 유의한 공변량을 입력합니다. 검증 데이터 집합을 사용하는 경우 "Prob>F" 값은 훈련 데이터 집합을 기반으로 합니다.

단계별 후진 입력되었지만 잠기지 않은 공변량 중에서 가장 유의하지 않은 공변량을 제거합니다. 검증 데이터 집합을 사용하는 경우 "Prob>F" 값은 훈련 데이터 집합을 기반으로 합니다.

모두 입력 "추가" 열에 잠겨 있지 않은 모든 공변량을 선택하여 모든 공변량을 입력합니다.

모두 제거 "추가" 열에서 잠겨 있지 않은 모든 공변량을 선택 취소하여 제거합니다.

이 모형 적용 "추가" 열에 선택된 공변량을 기반으로 판별 분석 보고서를 생성합니다. "열 선택" 개요가 닫히고 "판별 분석" 창이 업데이트되어 선택한 판별 방법을 기반으로 분석 결과가 표시됩니다.

팁 : 이 모형 적용을 클릭하면 선택한 열이 "스코어 요약" 보고서 맨 위에 나타납니다.

시작 검증 엔트로피 R^2 이 감소하기 시작할 때까지 전진 선택 단계에서 공변량을 입력합니다. 검증 엔트로피 R^2 이 개선되지 않은 상태로 전진 선택 단계가 두 번 수행되면 입력이 종료됩니다. JMP 의 제외된 행 또는 JMP Pro 의 검증 열에 대해서만 사용할 수 있습니다.

열

잠금 버튼을 사용한 단계별 변수 선별에 관계없이 공변량을 강제로 현재 상태로 유지합니다.

다음 사항에 유의하십시오.

- 공변량을 입력한 후 해당 공변량에 대해 **잠금**을 선택하면 컨트롤 버튼을 사용하여 선택한 내용에 관계없이 공변량이 모형에 유지됩니다. 잠긴 공변량에 대한 **추가** 상자에는 체크 표시가 흐리게 표시되어 해당 공변량이 모형에 있음을 나타냅니다.
- 입력되지 않은 공변량에 대해 **잠금**을 선택하면 컨트롤 버튼을 사용하여 선택한 내용에 관계없이 공변량이 모형에 추가되지 않습니다.

추가 현재 모형에 있는 열을 나타냅니다. 모형에 있는 열 또는 없는 열을 수동으로 선택할 수 있습니다. 흐린 체크 표시는 모형에 입력된 잠긴 공변량을 나타냅니다.

열 관심 공변량입니다.

F 비 공분산 분석 모형을 사용하여 얻은 그룹 변수의 검정에 대한 *F* 비입니다. 자세한 내용은 "F 비와 Prob>F 업데이트"에서 확인하십시오.

Prob > F 공분산 분석 모형을 사용하여 얻은 그룹 변수의 검정에 대한 *p* 값입니다. 자세한 내용은 "F 비와 Prob>F 업데이트"에서 확인하십시오.

단계별의 예

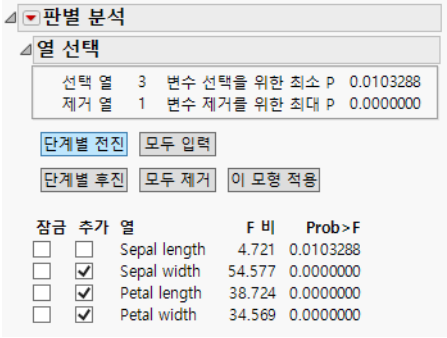
"단계별" 사용 방법을 보여 주기 위해 Iris.jmp 샘플 데이터 테이블을 예로 들어보겠습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다.
2. **분석 > 다변량 방법 > 판별**을 선택합니다.

3. Sepal length, Sepal width, Petal length 및 Petal width 를 선택하고 **Y, 공변량**을 클릭합니다 .
4. Species 를 선택하고 **X, 범주**를 클릭합니다 .
5. **단계별 변수 선택**을 선택합니다 .
6. **확인**을 클릭합니다 .
7. **단계별 전진**을 세 번 클릭합니다 .

   세 개의 공변량이 모형에 입력됩니다 . " 변수 선택을 위한 최소 P" 가 상단 패널에 나타납니다 . 이 값이 0.0103288 이므로 남은 공변량 Sepal length 도 Species 에 대한 판별 분석에서 중요할 수 있음을 나타냅니다 .

그림 5.5 Iris.jmp 의 변수 선별에 의한 모형

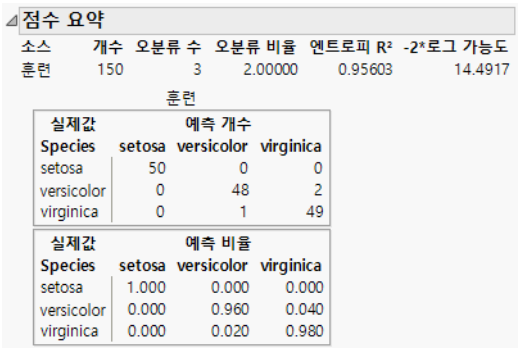


8. **이 모형 적용**을 클릭합니다 .

   " 열 선택 " 개요가 닫힙니다 . 창이 업데이트되어 입력된 공변량과 선택한 판별 방법에 따라 적합 보고서가 표시됩니다 .

   모형에 대해 선택한 공변량이 " 스코어 요약 " 보고서 맨 위에 나열됩니다 .

그림 5.6 선택한 공변량을 보여 주는 스코어 요약 보고서



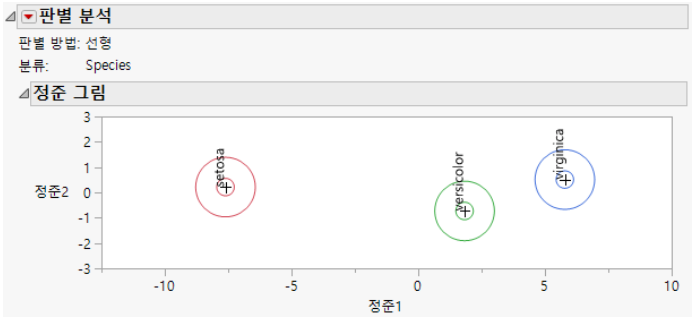
판별 방법

JMP에서는 판별 분석을 수행하기 위해 선형, 2차, 정규화 및 와이드 선형이라는 방법을 제공합니다. 처음 세 가지 방법은 기본 모형이 다릅니다. 와이드 선형 방법은 공변량 수가 많을 때 선형 모형을 적합시키는 효율적인 방법입니다.

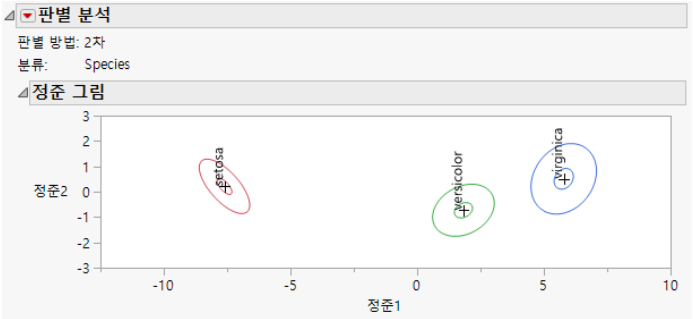
참고 : 공변량을 500 개 넘게 입력하면 와이드 선형 방법으로 전환할 것을 제안하는 JMP 경고가 나타납니다 . 열 수가 많은 경우 다른 방법을 사용하면 계산 시간이 길어질 수 있기 때문입니다 . 와이드 선형 방법으로 전환하려면 **와이드 선형 , 복수 열**을 클릭합니다 . 원래 선택한 방법을 사용하려면 **계속**을 클릭합니다 .

그림 5.7 선형 , 2 차 및 정규화 판별 분석

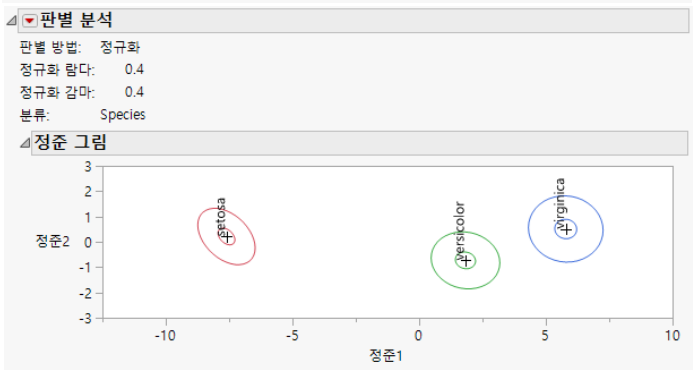
선형



2 차



정규화
(λ=0.4, γ=0.4)



선형 , 2 차 및 정규화 방법이 그림 5.7 에 나와 있습니다 . 여기에서는 각 방법을 간략하게 설명합니다 . 자세한 내용은 " 저장된 계산식 " 에서 확인하십시오 .

선형 , 합동 공분산 선형 판별 분석을 수행합니다 . 이 방법에서는 그룹 내 공분산 행렬이 같다고 가정합니다 . 자세한 내용은 " 선형 판별 방법 " 에서 확인하십시오 .

2 차 , 각 공분산 2 차 판별 분석을 수행합니다 . 이 방법에서는 그룹 내 공분산 행렬이 다르다고 가정합니다 . 이 방법은 선형 방법보다 더 많은 모수를 추정해야 합니다 . 그룹 표본 크기가 작으면 추정값이 불안정할 수 있습니다 . 자세한 내용은 "2 차 판별 방법 " 에서 확인하십시오 .

공변량이 X 변수의 수준에 대해 일정하면 그룹 내 공분산 행렬의 관련 항목은 공분산이 0 입니다 . 역행렬을 계산하기 위해 공분산 0 은 해당하는 합동 그룹 내 공분산으로 대체됩니다 . 이 작업이 완료되면 문제가 있는 공변량과 X 수준을 식별하는 노트가 보고서 창에 나타납니다 .

팁 : 데이터 집합이 작은 경우 2 차 방법의 단점이 있습니다 . 가역적이고 안정적인 공분산 행렬을 생성하기 어려울 수 있습니다 . 정규화 방법은 그룹 간 차이를 허용하면서 이러한 문제를 개선합니다 .

정규화 , 절충 방법 그룹 내 공분산 행렬이 다를 때 추정값에 안정성을 부여하는 두 가지 방법을 제공합니다 . 이 옵션은 그룹 표본 크기가 작을 때 유용합니다 . 자세한 내용은 " 정규화 , 절충 방법 " 및 " 정규화 판별 방법 " 에서 확인하십시오 .

와이드 선형 , 복수 열 다른 방법을 사용하면 계산상 문제가 발생할 수 있는 다수의 공변량을 기반으로 모형을 적합시키는 데 유용합니다 . 이 방법에서는 그룹 내 공분산 행렬이 모두 같다고 가정합니다 . 이 방법은 특이값 분해 방식을 사용하여 합동 그룹 내 공분산 행렬의 역행렬을 계산합니다 . 자세한 내용은 " 와이드 선형 알고리즘 설명 " 에서 확인하십시오 .

참고 : " 와이드 선형 " 옵션을 사용하면 다른 판별 방법에 일반적으로 나타나는 몇 가지 기능을 사용할 수 없습니다 . 이 알고리즘은 매우 큰 합동 그룹 내 공분산 행렬을 명시적으로 계산하지 않기 때문입니다 .

정규화 , 절충 방법

정규화 판별 분석은 음수가 아닌 두 모수에 의해 관리됩니다 .

- 첫 번째 모수 (**람다 , 합동 공분산에 가깝게 축소**) 는 개별 공분산 행렬과 그룹 공분산 행렬을 혼합하는 방법을 지정합니다 . 이 모수의 경우 1 은 선형 판별 분석에 해당하고 0 은 2 차 판별 분석에 해당합니다 .
- 두 번째 모수 (**감마 , 대각 행렬에 가깝게 축소**) 는 비대각 요소 (변수 간 공분산) 에 적용할 수축 크기를 지정하는 승수입니다 . 1 을 선택하면 공분산 행렬이 강제로 대각화됩니다 .

이 두 모수에 각각 0 을 할당하면 2 차 판별 분석을 요청하는 것과 동일합니다 . 마찬가지로 람다에 1 을 할당하고 감마에 0 을 할당하면 선형 판별 분석이 요청됩니다 . 표 5.1 의 내용을 참조하면 정규화를 결정하는 데 도움이 됩니다 . 선형 , 2 차 및 정규화 판별 분석의 예는 그림 5.7 에서 확인하십시오 .

표 5.1 정규화 판별 분석

더 작은 람다 사용	더 큰 람다 사용	더 작은 감마 사용	더 큰 감마 사용
공분산 행렬이 다름	공분산 행렬이 같음	변수에 상관관계가 있음	변수에 상관관계가 없음
행 수가 많음	행 수가 적음		
변수 수가 적음	변수 수가 많음		

공분산 축소

판별 시작 창에서 공분산 축소 옵션을 선택할 수 있습니다. 이 옵션은 일부 그룹의 관측값 수가 적을 때 권장됩니다. 판별 분석을 수행하려면 공분산 행렬의 역행렬을 계산해야 합니다. 비대각 항목을 축소하면 안정성이 향상되고 예측 분산이 줄어듭니다. 공분산 축소 옵션은 Schafer 와 Strimmer(2005) 연구 자료에 설명된 방법을 사용하여 결정된 계수로 비대각 항목을 축소합니다.

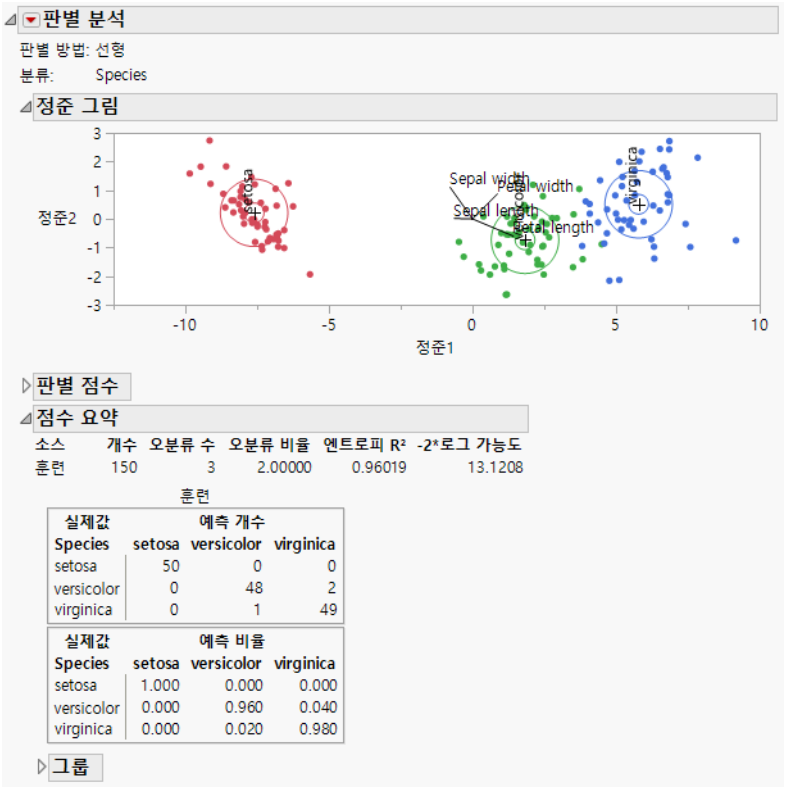
시작 창에서 선형 판별 방법과 함께 "공분산 축소" 옵션을 선택하면 람다 및 감마 값이 적절하게 지정된 정규화 판별 방법에서 제공하는 축소와 동일하게 공분산 행렬이 축소됩니다. "공분산 축소" 옵션을 선택하고 분석을 실행하면 "축소" 보고서에 "전체 축소" 값과 "전체 람다" 값이 제공됩니다. 정규화 방법을 사용하여 동일한 분석을 얻으려면 "정규화 모수" 창에서 "람다"에 1을 입력하고 "감마"에 "축소" 보고서의 "전체 람다" 값을 입력합니다.

판별 분석 보고서

"판별 분석" 보고서에는 선택한 판별 방법에 기반한 판별 결과가 제공됩니다. "판별 방법" 및 "분류" 변수가 보고서 맨 위에 표시됩니다. "정규화" 방법을 선택한 경우 관련 모수도 표시됩니다.

"판별 분석"의 빨간색 삼각형 메뉴에서 옵션을 선택하여 판별 방법을 변경할 수 있습니다. 그러면 보고서 결과가 업데이트되어 선택한 방법이 반영됩니다.

그림 5.8 판별 분석 보고서의 예



기본 " 판별 분석 " 보고서에는 다음 섹션이 포함되어 있습니다 .

- " 와이드 선형 " 판별 방법을 선택하면 " 주성분 " 보고서가 나타납니다 . 자세한 내용은 " 주성분 " 에서 확인하십시오 .
- " 정준 그림 " 에는 그룹을 가장 잘 분리하는 두 개의 차원에 점과 다변량 평균이 표시됩니다 . 자세한 내용은 " 정준 그림 및 정준 구조 " 에서 확인하십시오 .
- " 판별 스코어 " 보고서에서는 각 관측값이 어떻게 분류되는지 자세히 보여 줍니다 . 자세한 내용은 " 판별 스코어 " 에서 확인하십시오 .
- " 스코어 요약 " 보고서에서는 관측값이 얼마나 잘 분류되었는지 간략하게 보여 줍니다 . 자세한 내용은 " 판별 스코어 " 에서 확인하십시오 .

주성분

이 보고서는 " 와이드 선형 " 방법을 선택한 경우에만 나타납니다 . 다음 표기를 고려하십시오 .

- 공변량의 $n \times p$ 행렬을 $\mathbf{Y}$ 로 나타냅니다 . 여기서 n 은 관측값 수이고 p 는 공변량 수입니다 .
- $\mathbf{Y}$ 의 각 관측값에 대해 공변량 평균을 빼고 그 차이를 해당 공변량의 합동 표준편차로 나눕니다 . 결과 행렬을 $\mathbf{Y}_S$ 로 나타냅니다 .

보고서에 다음과 같은 항목이 제공됩니다 .

개수 추출된 고유값 수입니다 . 고유값은 누적 백분율이 99.99% 이상이 될 때까지 추출됩니다 . 이는 변동의 99.99% 가 설명되었음을 나타냅니다 .

고유값 $\mathbf{Y}_s$ 에 대한 공분산 행렬의 고유값 , 즉 $(\mathbf{Y}_s'\mathbf{Y}_s)/(n - p)$ 이며 내림차순으로 정렬됩니다 .

누적 백분율 고유값의 누적합을 모든 고유값 합계의 백분율로 나타낸 값입니다 . 고유값 합계는 $\mathbf{Y}_s'\mathbf{Y}_s$ 의 계수입니다 .

특이값 $\mathbf{Y}_s$ 의 특이값이며 내림차순으로 정렬됩니다 .

정준 그림 및 정준 구조

정준 그림은 변수의 정준 상관 구조를 설명하는 행렬도입니다.

정준 구조

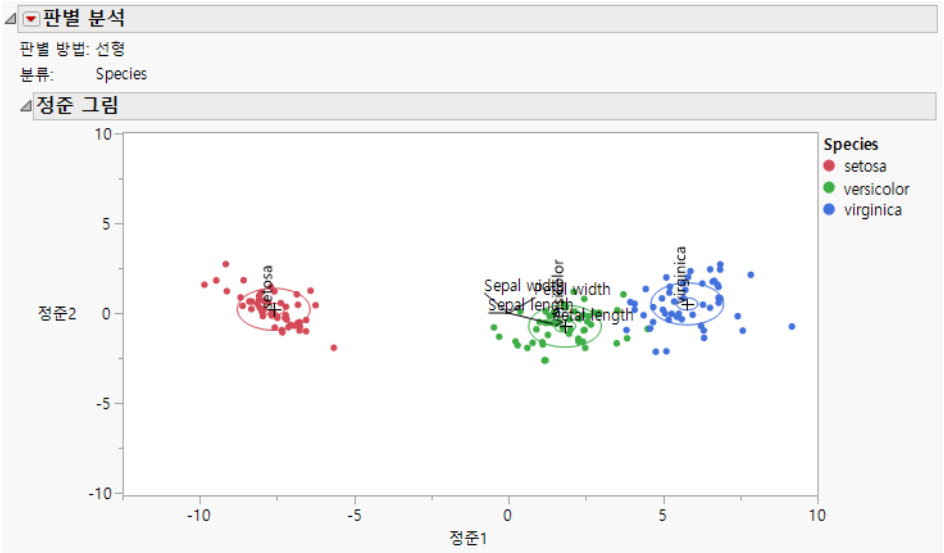
"X, 범주" 열의 각 수준은 표시 변수를 정의합니다. 범주를 나타내는 표시 변수 집합과 공변량 간에 정준 상관이 수행됩니다. 여기서 공변량의 선형 결합, 즉 **정준 변수**가 도출됩니다. 이러한 정준 변수는 범주 간 변동을 요약하려고 합니다.

첫 번째 정준 변수는 범주 표시 변수와 공변량 간의 다중상관을 최대화하는 공변량 선형 결합입니다. 두 번째 정준 변수는 첫 번째 정준 변수와 상관관계가 없으며 범주와의 다중상관을 최대화하는 선형 결합입니다. "X, 범주" 열에 k 개의 수준이 있으면 정준 변수의 수가 $k - 1$ 개입니다.

정준 그림

그림 5.9 에서는 Iris.jmp 데이터 테이블의 선형 판별 분석에 대한 정준 그림을 보여 줍니다. Species 별로 점에 색상이 적용되어 있습니다.

그림 5.9 Iris.jmp 에 대한 정준 그림



행렬도 축은 처음 두 정준 변수입니다 . 이는 그룹 간에 최대 분리를 제공하는 두 차원을 정의합니다 . 각 정준 변수는 공변량의 선형 결합입니다 . 자세한 내용은 " 정준 구조 " 에서 확인하십시오 . 행렬도에서는 각 관측값이 정준 변수에 따라 어떻게 표시되는지 그리고 각 공변량이 정준 변수에 어떻게 기여하는지 보여 줍니다 .

- 각 그룹의 관측값 및 다변량 평균은 행렬도에 점으로 표시됩니다 . 이는 처음 두 정준 변수로 표현됩니다 .
  - 각 다변량 평균에 해당하는 점은 더하기 ("+") 표식으로 나타냅니다 .
  - 각 평균에 대해 95% 신뢰 수준 타원이 그림에 표시됩니다 . 두 그룹이 유의하게 다르면 신뢰도 타원이 교차하지 않는 경향이 있습니다 .
  - 각 그룹에 대해 50% 등고선을 나타내는 타원이 그림에 표시됩니다 . 이는 정규성을 가정하고 처음 두 정준 변수의 공간에서 관측값의 약 50% 를 포함하는 영역을 나타냅니다 .
- 그림에 표시되는 선 집합은 공변량을 나타냅니다 .
  - 각 정준 변수에 대해 선형 결합된 공변량의 계수를 **가중치**로 해석할 수 있습니다 .
  - 가중치 간에 쉽게 비교하기 위해 각 공변량이 평균이 0, 표준편차가 1 이 되도록 표준화됩니다 . 표준화된 공변량의 계수를 **정준 가중치**라고 합니다 . 공변량의 정준 가중치가 클수록 정준 변수와의 연관성이 커집니다 .
  - 행렬도에서 각 선의 길이와 방향은 해당 공변량과 처음 두 정준 변수의 연관성 정도를 나타냅니다 . 선 길이는 정준 가중치의 배수입니다 .
  - 선은 데이터의 총 평균을 정준 변수 측면에서 나타내는 점 (0,0) 에서 시작됩니다 .

– " 판별 분석 " 의 빨간색 삼각형 메뉴에서 **정준 옵션 > 정준 상세 정보 표시**를 선택하여 가중치 계수 값을 얻을 수 있습니다 . " 정준 상세 정보 " 보고서 맨 아래에서 " 표준화 스코어링 계수 " 를 클릭합니다 . 자세한 내용은 " 표준화 스코어링 계수 " 에서 확인하십시오 .

정준 그림 수정

다음과 같이 추가 옵션을 사용하여 행렬도를 수정할 수 있습니다 .

- " 판별 분석 " 의 빨간색 삼각형 메뉴에서 **정준 옵션 > 평균 CL 타원 표시**를 선택하여 95% 신뢰도 타원을 표시하거나 숨깁니다 .
- " 판별 분석 " 의 빨간색 삼각형 메뉴에서 **정준 옵션 > 행렬도 선 표시**를 선택하여 선을 표시하거나 숨깁니다 .
- 행렬도 선의 가운데를 그래프의 다른 위치로 드래그합니다 . " 판별 분석 " 의 빨간색 삼각형 메뉴에서 **정준 옵션 > 행렬도 선 위치**를 선택하여 위치와 척도를 지정합니다 . 선을 표시하기 위해 조정이 필요한 경우를 제외하고 정준 그림의 기본 반지름 척도는 1.5 입니다 .
- " 판별 분석 " 의 빨간색 삼각형 메뉴에서 **정준 옵션 > 정규 50% 등고선 표시**를 선택하여 50% 등고선을 표시하거나 숨깁니다 .
- " 판별 분석 " 의 빨간색 삼각형 메뉴에서 **정준 옵션 > 점에 색 적용**을 선택하여 타원과 일치하도록 점에 색상 코딩을 적용합니다 .

세 개 이상의 범주로 분류

Iris.jmp 데이터의 경우 세 개의 Species 가 있으므로 정준 변수는 두 개뿐입니다 . 그림 5.9 의 그림에서는 두 정준 변수를 사용하여 세 그룹을 양호하게 분리한 상태를 보여 줍니다 .

그림의 선은 다음을 나타냅니다 .

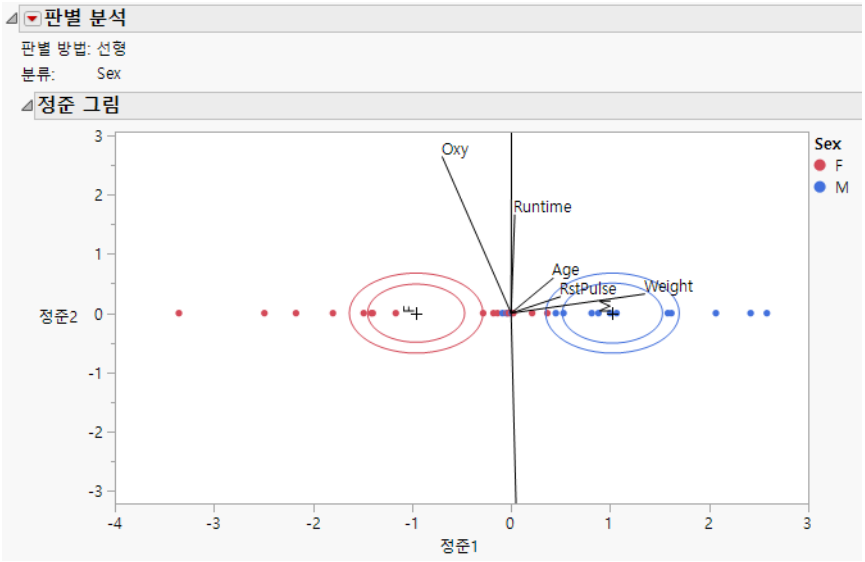
- Petal length 는 정준 1 과 양의 관계를 갖고 정준 2 와 음의 관계를 갖습니다 . 따라서 정준 2 보다 정준 1 을 정의하는 데 가중치를 더 부여합니다 .
- Petal width 는 정준 1 및 정준 2 와 모두 양의 관계를 갖습니다 . 따라서 두 정준 변량을 정의하는 데 동일한 가중치를 부여합니다 .
- Sepal width 는 정준 1 과 음의 관계를 갖고 정준 2 와 양의 관계를 갖습니다 . 따라서 정준 1 보다 정준 2 를 정의하는 데 가중치를 더 부여합니다 .
- Sepal length 는 정준 1 을 정의할 때 음의 가중치를 가지며 정준 2 를 정의할 때 매우 약하게 연관됩니다 .

두 개의 범주로 분류

분류 변수의 수준이 두 개뿐인 경우 그림에서 "정준1"로 표시되는 단일 정규 변수에 대해 점이 표시됩니다. 각 공변량의 정준 가중치는 정준1에만 관련됩니다. 선은 공변량을 분리하기 위해 연직 성분으로만 표시됩니다. 선을 정준1 축에 투영하여 단일 정준 변수와의 상대적 연관성을 비교할 수 있습니다.

그림 5.10 에서는 Fitness.jmp 샘플 데이터 테이블에 대한 정준 그림을 보여 줍니다. 7 개의 연속형 변량이 개인을 범주 M(남성) 또는 F(여성)로 분류하는 데 사용됩니다. 분류 변수의 범주가 두 개뿐이므로 정준 변수가 하나만 있습니다.

그림 5.10 Fitness.jmp 에 대한 정준 그림



정준 그림의 점에 Sex별로 색상이 적용되어 있습니다. 두 그룹은 정준1의 값으로 잘 분리되어 있습니다.

7 개의 공변량에 해당하는 선에 연직 성분이 있지만 이 경우 선을 정준 1 축에 대한 투영의 관점으로만 해석해야 합니다 . 다음 사항에 유의하십시오 .

- MaxPulse, Runtime 및 RunPulse 는 정준 1 과 연관성이 거의 없습니다 .
- Weight, RstPulse 및 Age 는 정준 1 과 양의 관계를 갖습니다 . Weight 의 연관성이 가장 높습니다 . RstPulse 및 Age 공변량은 연관성이 비슷하지만 더 작습니다 .
- Oxy 는 정준 1 과 음의 관계를 갖습니다 .

판별 스코어

" 판별 스코어 " 보고서에는 각 관측값의 예측 분류 및 지원 정보가 제공됩니다 .

행 데이터 테이블의 관측값 행입니다 .

실제값 데이터 테이블에 제공된 관측값 분류입니다 .

SqDist(실제값) 데이터 테이블에 제공된 관측값 분류에 대해 저장된 계산식 SqDist[<level>] 의 값입니다 . 자세한 내용은 " 스코어 옵션 " 에서 확인하십시오 .

참고 : 계산식의 오프셋 항 때문에 SqDist(실제값) 이 음수일 수 있습니다 .

Prob(실제값) 관측값의 실제 분류에 대한 추정 확률입니다 .

-Log(확률) Prob(실제값) 의 음의 로그입니다 . 이 음의 로그 가능도 값이 크면 실제 범주에 소속될 확률이 잘 예측되지 않는 관측값을 식별할 수 있습니다 .

"-Log(확률)" 값 오른쪽에 -Log(확률) 그림이 나타납니다 . 막대가 크면 잘못된 예측을 나타냅니다 . 별표 (*) 는 잘못 분류된 관측값을 나타냅니다 .

검증 또는 테스트 데이터 집합을 사용하는 경우 검증 데이터 집합의 관측값은 "v" 로 표시되고 테스트 데이터 집합의 관측값은 "t" 로 표시됩니다 .

예측값 관측값의 예측 분류입니다 . 예측 분류는 예측 소속 확률이 가장 높은 범주입니다 .

확률 (예측) 관측값의 예측 분류에 대한 추정 확률입니다 .

기타 예측 확률이 0.1 을 초과하는 다른 범주를 나열합니다 (있는 경우).

그림 5.11 에서는 Iris.jmp 샘플 데이터 테이블에 대해 선형 판별 방법을 사용한 "판별 스코어" 보고서를 보여 줍니다. **스코어 옵션 > 관심 있는 행만 표시** 옵션이 선택되어 예측 확률이 0.05 에서 0.95 사이인 행 또는 잘못 분류된 행만 표시됩니다.

그림 5.11 관심 있는 행만 표시

판별 점수

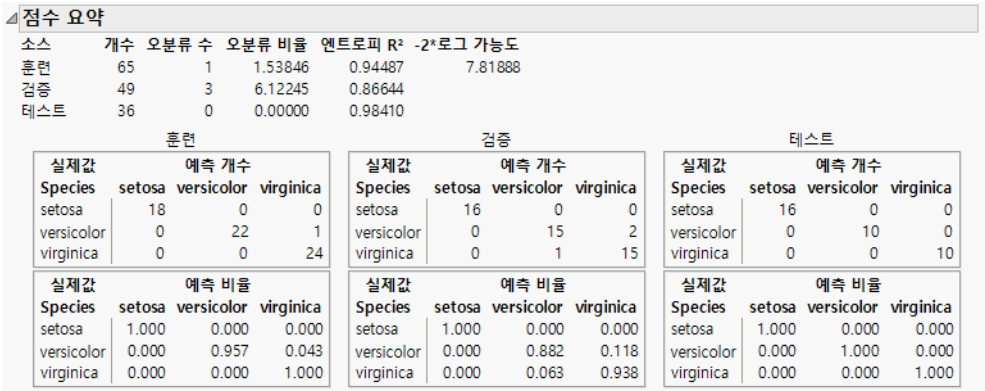
행	실제값	SqDist(실제값)	Prob(실제값)	-Log(확률)		예측값	확률(예측)	기타
71	versicolor	8.66970	0.2532	1.373	*	virginica	0.7468	
73	versicolor	4.87619	0.8155	0.204		versicolor	0.8155	virginica 0.18
78	versicolor	4.66698	0.6892	0.372		versicolor	0.6892	virginica 0.31
84	versicolor	8.43926	0.1434	1.942	*	virginica	0.8566	
120	virginica	8.19641	0.7792	0.249		virginica	0.7792	versicolor 0.22
124	virginica	3.57858	0.9029	0.102		virginica	0.9029	
127	virginica	3.90184	0.8116	0.209		virginica	0.8116	versicolor 0.19
128	virginica	3.31470	0.8658	0.144		virginica	0.8658	versicolor 0.13
130	virginica	9.08495	0.8963	0.109		virginica	0.8963	versicolor 0.10
134	virginica	7.23593	0.2706	1.307	*	versicolor	0.7294	
135	virginica	15.83301	0.9340	0.068		virginica	0.9340	
139	virginica	4.09385	0.8075	0.214		virginica	0.8075	versicolor 0.19

'*'는 잘못 분류되었음을 나타냅니다.

스코어 요약

"스코어 요약" 보고서에서는 판별 스코어에 대한 개요를 제공합니다. 그림 5.12의 테이블에는 실제값 및 예측 분류가 표시됩니다. 모든 관측값이 올바르게 분류되면 비대각 수가 0입니다.

그림 5.12 Iris.jmp 에 대한 스코어 요약



점수 요약

소스	개수	오분류 수	오분류 비율	엔트로피 R²	-2*로그 가능도
훈련	65	1	1.53846	0.94487	7.81888
검증	49	3	6.12245	0.86644	
테스트	36	0	0.00000	0.98410	

훈련

실제값	예측 개수		
Species	setosa	versicolor	virginica
setosa	18	0	0
versicolor	0	22	1
virginica	0	0	24

실제값	예측 비율		
Species	setosa	versicolor	virginica
setosa	1.000	0.000	0.000
versicolor	0.000	0.957	0.043
virginica	0.000	0.000	1.000

검증

실제값	예측 개수		
Species	setosa	versicolor	virginica
setosa	16	0	0
versicolor	0	15	2
virginica	0	1	15

실제값	예측 비율		
Species	setosa	versicolor	virginica
setosa	1.000	0.000	0.000
versicolor	0.000	0.882	0.118
virginica	0.000	0.063	0.938

테스트

실제값	예측 개수		
Species	setosa	versicolor	virginica
setosa	16	0	0
versicolor	0	10	0
virginica	0	0	10

실제값	예측 비율		
Species	setosa	versicolor	virginica
setosa	1.000	0.000	0.000
versicolor	0.000	1.000	0.000
virginica	0.000	0.000	1.000

" 스코어 요약 " 보고서에는 다음 정보가 제공됩니다 .

열 " 단계별 변수 선택 " 을 사용하여 모형을 생성한 경우 모형에 추가된 열이 나열됩니다 (그림 5.6 참조).

소스 검증을 사용하지 않으면 모든 관측값이 훈련 데이터 집합을 구성합니다 . 검증을 사용하면 훈련 및 검증 데이터 집합 또는 훈련 , 검증 및 테스트 데이터 집합에 대한 행이 표시됩니다 .

오분류 수 지정된 집합에서 잘못 분류된 관측값의 수를 제공합니다 .

오분류 비율 지정된 집합에서 잘못 분류된 관측값의 백분율을 제공합니다 .

엔트로피 R² 적합 측도입니다 . 값이 클수록 더 나은 적합을 나타냅니다 . " 엔트로피 R²" 값이 1 이면 분류가 완벽하게 예측되었음을 나타냅니다 . 예측 확률의 불확도는 판별 모형에서 일반적으로 사용되므로 엔트로피 R² 값은 작은 경향이 있습니다 .

자세한 내용은 " 엔트로피 R²" 에서 확인하십시오 .

참고 : 엔트로피 R² 이 음수일 수 있습니다 .

-2* 로그 가능도 모형을 기반으로 훈련 데이터 집합에 있는 관측값의 음의 로그 가능도를 두 배 곱한 값입니다 . 값이 작을수록 더 나은 적합을 나타냅니다 . 훈련 데이터 집합에만 제공됩니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

혼동 행렬 범주형 X 의 각 수준에 대한 실제값 대 예측값 개수 행렬을 표시합니다 . JMP Pro 에서 검증을 사용하는 경우 각 관측값 집합에 대해 행렬이 제공됩니다 . JMP 에서 제외된 행을 사용하는 경우 제외된 행이 검증 데이터 집합으로 간주되고 별도의 검증 행렬이 제공됩니다 . 자세한 내용은 "JMP 및 JMP Pro 에서 검증 " 에서 확인하십시오 .

엔트로피 R²

엔트로피 R²은 적합 측도입니다. 이 값은 훈련 데이터 집합에 대해 계산되고, 검증을 사용하는 경우 검증 데이터 집합 및 테스트 데이터 집합에 대해 계산됩니다.

훈련 데이터 집합에 대한 엔트로피 R^2

훈련 데이터 집합에 대한 엔트로피 R^2 은 다음과 같이 계산됩니다 .

- 훈련 데이터 집합을 사용하여 판별 모형이 적합됩니다 .
- 모형을 기반으로 한 예측 확률을 구합니다 .
- 이러한 예측 확률을 사용하여 훈련 데이터 집합의 관측값에 대한 가능도가 계산됩니다. 이를 $Likelihood_Full_{Training}$ 이라고 합니다 .
- 훈련 데이터 집합을 사용하여 축소 모형 (예측 변수 없음) 이 적합됩니다 .
- 축소 모형의 X 수준에 대한 예측 확률을 사용하여 훈련 데이터 집합의 관측값에 대한 가능도가 계산됩니다 . 이를 $Likelihood_Reduced_{Training}$ 이라고 합니다 .
- 훈련 데이터 집합에 대한 엔트로피 R^2 은 다음과 같습니다 .

$$\text{엔트로피 } R^2_{Training} = 1 - \frac{\log(Likelihood_Full_{Training})}{\log(Likelihood_Reduced_{Training})}$$

검증 및 테스트 데이터 집합에 대한 엔트로피 R^2

검증 데이터 집합에 대한 엔트로피 R^2 은 다음과 같이 계산됩니다 .

- 훈련 데이터 집합만 사용하여 판별 모형이 적합됩니다 .
- 모든 관측값에 대해 훈련 데이터 집합 모형을 기반으로 한 예측 확률을 구합니다 .
- 이러한 예측 확률을 사용하여 검증 데이터 집합의 관측값에 대한 가능도가 계산됩니다. 이를 $Likelihood_Full_{Validation}$ 이라고 합니다 .
- 훈련 데이터 집합만 사용하여 축소 모형 (예측 변수 없음) 이 적합됩니다 .
- 축소 모형의 X 수준에 대한 예측 확률을 사용하여 검증 데이터 집합의 관측값에 대한 가능도가 계산됩니다 . 이를 $Likelihood_Reduced_{Validation}$ 이라고 합니다 .
- 검증 엔트로피 R^2 은 다음과 같습니다 .

$$\text{검증 엔트로피 } R^2 = 1 - \frac{\log(Likelihood_Full_{Validation})}{\log(Likelihood_Reduced_{Validation})}$$

테스트 데이터 집합의 엔트로피 R^2 은 검증 데이터 집합의 엔트로피 R^2 과 유사한 방식으로 계산됩니다.

판별 분석 옵션

" 판별 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

단계별 변수 선택 (" 와이드 선형 " 방법에는 사용 불가능) " 열 선택 " 제어판을 표시하거나 숨깁니다 . 이 제어판에는 공분산 분석 및 *p* 값을 사용하여 단계별 변수 선택을 수행할 수 있는 옵션이 포함되어 있습니다 . 자세한 내용은 " 단계별 변수 선택 " 에서 확인하십시오 .

판별 방법 판별 방법을 지정합니다 . " 선형 ", "2 차 ", " 정규화 " 및 " 와이드 선형 " 중에서 선택합니다 . 자세한 내용은 " 판별 방법 " 에서 확인하십시오 .

판별 스코어 각 행의 판별 스코어 테이블을 표시하거나 숨깁니다 .

스코어 옵션 관측값 스코어링을 위한 옵션을 제공합니다 . 특히 스코어링 계산식을을 저장할 수 있습니다 . 자세한 내용은 " 스코어 옵션 " 에서 확인하십시오 .

정준 그림 정준 그림을 표시하거나 숨깁니다 . 자세한 내용은 " 정준 그림 및 정준 구조 " 에서 확인하십시오 .

정준 옵션 정준 그림에 영향을 주는 옵션을 제공합니다 . 자세한 내용은 " 정준 옵션 " 에서 확인하십시오 .

3D 정준 그림 3 차원 정준 그림을 표시합니다 . 이 옵션은 범주형 X 의 수준이 네 개 이상인 경우에만 사용할 수 있습니다 . 자세한 내용은 "3D 정준 그림의 예 " 에서 확인하십시오 .

사전 확률 지정 *X* 변수의 각 수준에 대한 사전 확률을 지정할 수 있습니다 . 자세한 내용은 " 사전 확률 지정 " 에서 확인하십시오 .

새 수준 고려 일부 점은 알려진 그룹에 적합하지 않을 수 있으며 스코어가 없는 새 그룹에 속한 것으로 간주되어야 함을 지정합니다 . 자세한 내용은 " 새 수준 고려 " 에서 확인하십시오 .

그룹 내 공분산 표시 다음 보고서를 표시하거나 숨깁니다 .

- 합동 그룹 내 공분산 및 상관 행렬을 제공하는 공분산 행렬 보고서
- "2 차 " 및 " 정규화 " 방법을 선택한 경우 그룹 내 상관 행렬을 표시하는 그룹별 상관 보고서

  각 그룹에 대해 그룹 내 공분산 행렬의 행렬식 로그도 표시됩니다 .
- "2 차 " 판별 방법을 선택한 경우 그룹 내 공분산 행렬을 표시하는 " 공분산 행렬 " 보고서에 " 그룹 공분산 " 개요가 추가됩니다 .

" 와이드 선형 " 판별 방법에는 " 그룹 내 공분산 표시 " 를 사용할 수 없습니다 .

그룹 평균 표시 각 공변량의 평균을 제공하는 " 그룹 평균 " 보고서를 표시하거나 숨깁니다 . *X* 변수의 각 수준에 대한 평균과 전체 평균이 나타납니다 .

판별 행렬 저장 판별 분석 결과라는 스크립트를 데이터 테이블에 저장합니다 . 이 스크립트는 JSL 에서 사용하기 위한 다음 개체의 목록입니다 .

- 공변량 (Y) 목록
- 범주형 변수 X
- X 수준 목록

- 공변량 평균 대 X 수준의 행렬
- 합동 그룹 내 공분산 행렬

" 와이드 선형 " 판별 방법에는 " 판별 행렬 저장 " 을 사용할 수 없습니다 . 자세한 내용은 " 판별 행렬 저장 " 에서 확인하십시오 .

산점도 행렬 각 공변량 쌍에 대한 산점도가 있는 행렬을 보여 주는 " 산점도 행렬 " 보고서를 엽니다 . 이 옵션은 각 그룹에 대해 음영 밀도 타원이 있는 산점도 행렬 플랫폼을 호출합니다 . 산점도에는 검증이 사용된 경우에도 데이터 테이블의 모든 관측값이 포함됩니다 . 자세한 내용은 " 산점도 행렬 " 에서 확인하십시오 .

" 와이드 선형 " 판별 방법에는 사용할 수 없습니다 .

프로파일러 범주형 확률이 단일 프로파일러 행으로 결합된 대화식 프로파일러 보고서를 표시하거나 숨깁니다 . 요인 값을 변경하면 추정된 분류 확률에 반영됩니다 . 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

스코어 옵션

" 스코어 옵션 " 은 스코어를 처리하는 다음과 같은 선택 항목을 제공합니다 .

관심 있는 행만 표시 " 판별 스코어 " 보고서에서 잘못 분류된 행과 예측 확률이 0.05 에서 0.95 사이인 행만 표시합니다 .

분류 개수 표시 " 스코어 요약 " 보고서에 혼동 행렬과 혼동 비율 행렬을 표시하거나 숨깁니다 . 혼동 행렬은 실제 반응과 예측 반응의 이원 분류입니다 . 혼동 비율 행렬은 값을 행 합계로 나눈 혼동 행렬과 동일합니다 . 기본적으로 " 스코어 요약 " 보고서에는 범주형 X 의 각 수준에 대한 혼동 행렬과 혼동 비율 행렬이 표시됩니다 . JMP Pro 에서 검증을 사용하는 경우 각 관측값 집합에 대해 행렬이 제공됩니다 . JMP 에서 제외된 행을 사용하는 경우 이러한 행이 검증 데이터 집합으로 간주되고 별도의 검증 행렬이 제공됩니다 . 자세한 내용은 "JMP 및 JMP Pro 에서 검증 " 에서 확인하십시오 .

각 그룹까지의 거리 표시 각 그룹 평균에 대한 각 관측값의 Mahalanobis 거리 제곱을 포함하는 보고서를 표시하거나 숨깁니다 .

각 그룹에 대한 확률 표시 관측값이 범주형 X 로 정의된 각 그룹에 속할 확률을 포함하는 보고서를 표시하거나 숨깁니다 .

ROC 곡선 " 스코어 요약 " 보고서에 ROC(Receiver Operating Characteristic) 곡선을 표시하거나 숨깁니다 . ROC 곡선에 대한 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오 .

오분류 행 선택 행별 목록을 표시하는 보고서 창과 데이터 테이블에서 잘못 분류된 행을 선택합니다 .

불확실한 행 선택 행별 목록을 표시하는 보고서 창과 데이터 테이블에서 분류가 불확실한 행을 선택합니다 . 불확실한 행은 **어떤** 그룹에 대한 소속 그룹일 확률이 0 에 가깝지도 않고 1 에 가깝지도 않은 행입니다 .

이 옵션을 선택하면 불확도를 반영하는 예측 확률 범위를 지정할 수 있는 창이 열립니다 . 기본적으로 확률이 0 또는 1 과 0.1 넘게 차이 나는 행은 불확실한 것으로 정의됩니다 . 따라서 0.1 에서 0.9 사이의 확률을 갖는 행이 기본적으로 선택됩니다 .

계산식 저장 거리 , 확률 및 소속 확률 예측 계산식을 데이터 테이블에 저장합니다 . 자세한 내용은 " 저장된 계산식 " 에서 확인하십시오 .

- 거리 계산식은 SqDist0 및 SqDist[<level>] 이며 , 여기서 <level> 은 X 의 수준을 나타냅니다 . 거리 계산식은 Mahalanobis 거리 계산과 관계가 있는 중간 값을 생성합니다 .
- 확률 계산식은 Prob[<level>] 이며 , 여기서 <level> 은 X 의 수준을 나타냅니다 . 각 확률 열은 해당 X 수준에서 관측값의 소속에 대한 사후 확률을 제공합니다 . " 반응 확률 " 열 특성이 각 확률 열에 저장됩니다 . " 반응 확률 " 열 특성에 대한 자세한 내용은 **JMP 사용**에서 확인하십시오 .
- 소속 확률 예측 계산식은 Pred <X> 이며 " 최대 확률 분류 수준 " 이라는 분류 규칙을 포함합니다 .
- 와이드 선형 방법은 공변량 벡터와 **판별 주성분** 계산식을 포함하는 **판별 데이터 행렬** 열도 저장합니다 . 자세한 내용은 " 와이드 선형 판별 방법 " 에서 확인하십시오 .

참고 : 와이드 선형을 제외한 다른 방법의 경우 계산식을 저장하면 "RowEdit Prob" 스크립트가 데이터 테이블에 저장됩니다 . 이 스크립트는 데이터 테이블에서 불확실한 행을 선택합니다 . 스크립트는 확률이 0 또는 1 과 0.1 넘게 차이 나는 행을 불확실한 것으로 정의합니다 . 또한 불확실한 행을 검토할 수 있도록 " 행 편집기 " 창도 엽니다 . 와이드 선형을 제외한 새 모형을 적합시키고 " 계산식 저장 " 을 선택하면 기존 "RowEdit Prob" 스크립트가 새 적합에 적용되는 스크립트로 대체됩니다 .

스코어링 스크립트 만들기 (JMP Standard 에서만 사용 가능) " 계산식 저장 " 옵션으로 저장된 계산식 열을 구성하는 스크립트를 생성합니다 . 이 스크립트를 저장한 후 다른 데이터 테이블과 함께 사용하여 소속 확률을 계산하고 소속 그룹을 예측하는 계산식 열을 생성할 수 있습니다 .

JMP PRO **확률 계산식 게시** (JMP Pro 에서만 사용할 수 있음) 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다 . " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다 . 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오 .

정준 옵션

아래 나열된 첫 번째 옵션은 정준 그림 또는 3D 정준 그림의 모양과 관련이 있습니다. 나머지 옵션은 그림과 관련된 계산에 대해 상세 정보를 제공합니다.

참고 : 3D 정준 그림은 공변량이 세 개 이상 있고 그룹화 변수의 범주가 네 개 이상인 경우에만 사용할 수 있습니다 .

그림 모양과 관련된 옵션

점 표시 정준 그림 및 3D 정준 그림에서 점을 표시하거나 숨깁니다 .

평균 CL 타원 표시 정규성을 가정하고 정준 그림 및 3D 정준 그림에서 각 그룹의 평균에 대한 95% 신뢰도 타원을 표시하거나 숨깁니다 .

정규 50% 등고선 표시 각 그룹에 대해 50% 예측 타원 또는 타원체를 표시하거나 숨깁니다 . 정준 그림에서 각 타원은 다변량 정규성을 가정하고 처음 두 정준 변수의 공간에서 각 그룹에 대한 관측값의 약 50% 가 속할 위치를 추정하는 영역을 나타냅니다 . 3D 정준 그림에서 각 타원체는 다변량 정규성을 가정하고 처음 세 정준 변수의 공간에서 관측값의 약 50% 가 속할 위치를 추정하는 영역을 나타냅니다 .

행렬도 선 표시 정준 그림 및 3D 정준 그림에서 행렬도 선을 표시하거나 숨깁니다 . 라벨이 지정된 선은 정준 공간에서 공변량의 방향을 표시합니다 . 이러한 선은 각 공변량과 각 정준 변수의 연관성 정도를 나타냅니다 .

행렬도 선 위치 정준 그림 및 3D 정준 그림에서 행렬도 선의 위치와 반지름 척도를 지정할 수 있습니다 .

- 기본적으로 선은 데이터의 총 평균을 정준 변수 측면에서 나타내는 점 (0,0) 에서 시작됩니다 . 정준 그림에서 선을 드래그하거나 이 옵션을 사용하여 좌표를 지정할 수 있습니다 .
- 선을 표시하기 위해 조정이 필요한 경우를 제외하고 정준 그림의 기본 반지름 척도는 1.5 입니다 . 반지름 척도는 표준화 스코어링 계수와 관련하여 완료됩니다 .

점에 색 적용 *X* 변수의 수준을 기반으로 정준 그림 및 3D 정준 그림의 점에 색상을 지정합니다 . 데이터 테이블의 행에 색상 표식이 추가됩니다 . 이 옵션은 **행 > 열 값에 따른 색상 또는 표식** 을 선택하고 *X* 변수를 선택하는 것과 같습니다 . 또한 그래프를 마우스 오른쪽 버튼으로 클릭하고 **행 범례** 를 선택한 후 분류 열에 따라 색상을 적용하는 것과도 같습니다 .

계산과 관련된 옵션

정준 상세 정보 표시 " 정준 상세 정보 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 정준 상세 정보 표시 " 에서 확인하십시오 .

정준 구조 표시 " 정준 구조 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 정준 구조 표시 " 에서 확인하십시오 . " 와이드 선형 " 판별 방법에는 사용할 수 없습니다 .

정준 스코어 저장 데이터 테이블에 각 관측값에 대한 정준 스코어 계산식이 포함된 열을 생성합니다 . *k* 번째 정준 스코어의 열에는 **정준** [<k>] 라는 이름이 지정됩니다 .

팁 : 스크립트에서 **새 데이터 테이블에 저장** 스크립트 명령을 판별 개체에 보내면 정준 변수에 대한 그룹 평균 , 표준화 스코어링 계수의 반지름 척도가 1.5 인 행렬도 선 및 정준 스코어가 새 데이터 테이블에 저장됩니다 . " 와이드 선형 " 판별 방법에는 사용할 수 없습니다 .

정준 상세 정보 표시

"정준 상세 정보" 보고서에서는 공변량과 그룹화 변수 *X* 사이의 관계를 확인하는 검정을 보여 줍니다. 보고서 아래쪽에 관련 행렬이 제공됩니다.

그림 5.13 Iris.jmp 에 대한 정준 상세 정보

정준 상세 정보

전체 합동 그룹 내 공분산 행렬에서 계산된 정준 상세 정보입니다.

고유값	백분율	누적 백분율	정준 상관	가능도비	근사 F	NumDF	DenDF	Prob>F
32.1919292	99.1213	99.1213	0.98482089	0.02343863	199.1453	8	288	<.0001*
0.28539104	0.8787	100.0000	0.47119702	0.77797337	13.7939	3	145	<.0001*

검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.0234386	199.1453	8	288	<.0001*
Pillai 대각합	1.1918988	53.4665	8	290	<.0001*
Hotelling-Lawley	32.47732	582.1970	8	203.4	<.0001*
Roy 최대근	32.191929	1166.9574	4	145	<.0001*

▷ 행렬 내
▷ 행렬 간
▷ 스코어링 계수
▷ 표준화 스코어링 계수

참고 : 보고서의 결과를 계산하는 데 사용되는 행렬은 합동 그룹 내 공분산 행렬입니다 (" 그룹 내 행렬 " 로 표시됨). 이 행렬은 모든 판별 방법에 대해 " 정준 상세 정보 " 보고서의 기초로 사용됩니다 . " 정준 상세 정보 " 보고서의 통계량과 검정은 모든 판별 방법에 대해 동일합니다 .

통계량 및 검정

" 정준 상세 정보 " 보고서에는 고유값이 나열되고 0 고유값에 대한 가능도비 검정이 제공됩니다 . 정준 상관이 0 이라는 귀무가설에 대해 네 가지 검정이 제공됩니다 .

고유값 그룹 내 행렬의 역행렬과 그룹 간 행렬의 곱을 나타내는 고유값입니다 . 이 값은 가장 큰 것부터 순서대로 나열됩니다. 고유값 크기는 연관된 판별 함수로 설명되는 분산의 양을 반영합니다 .

백분율 주어진 고유값이 나타내는 고유값 합계의 비율입니다 .

누적 백분율 비율의 누적합입니다 .

정준 상관 공변량과 범주형 *X* 로 정의된 그룹 간의 정준 상관입니다 . *X* 로 정의된 그룹을 나타내기 위해 숫자 표시 변수를 정의하는 경우를 가정해 보겠습니다 . 그런 다음 공변량을 한 변수 집합으로 사용하고 *X* 의 그룹을 나타내는 표시 변수를 다른 집합으로 사용하여 정준 상관 분석을 수행합니다 . " 정준 상관 " 값은 이 분석의 결과로 생성된 정준 상관 값입니다 .

가능도비 해당하는 정규 상관과 더 작은 모든 상관의 모집단 값이 0 인지 여부를 검정하기 위한 가능도비 통계량입니다. 이 비율은 주어진 정준 상관 및 더 작은 모든 정준 상관 값의 곱과 같습니다 (1 - 정준 상관 ^{2}).

검정 모든 그룹에서 공변량 평균이 같다는 귀무가설에 대한 네 가지 표준 검정이 나열됩니다 (Wilk 람다 , Pillai 대각합 , Hotelling-Lawley 및 Roy 최대근). 자세한 내용은 " 다변량 검정 " 및 " 근사 F- 검정 " 에서 확인하십시오 .

근사 F 해당 검정과 관련된 F 값입니다. 특정 검정의 경우 F 값이 근사값 또는 상한입니다. 자세한 내용은 " 근사 F- 검정 " 에서 확인하십시오 .

NumDF 해당 검정에 대한 분자 자유도입니다 .

DenDF 해당 검정에 대한 분모 자유도입니다 .

Prob>F 해당 검정에 대한 p 값입니다 .

행렬

정준 구조와 관련된 네 개의 행렬이 보고서 아래쪽에 제공됩니다. 행렬을 보려면 행렬 이름 옆의 표시 아이콘을 클릭합니다. 숨기려면 행렬 이름을 클릭합니다 .

그룹 내 행렬 합동 그룹 내 공분산 행렬입니다 .

그룹 간 행렬 그룹 간 공분산 행렬 $\mathbf{S}_B$ 입니다 . 자세한 내용은 " 그룹 간 공분산 행렬 " 에서 확인하십시오 .

스코어링 계수 원시 데이터를 기준으로 정준 스코어를 계산하는 데 사용되는 계수입니다 . 이러한 계수는 **정준 옵션 > 정준 스코어 저장** 옵션에 사용됩니다 . 계산 방법에 대한 자세한 내용은 SAS Institute Inc. (2020b) 의 "CANDISC Procedure" 장에서 확인하십시오 .

표준화 스코어링 계수 표준화 데이터를 기준으로 정준 스코어를 계산하는 데 사용되는 계수입니다 . 정준 가중치라고도 합니다 . 계산 방법에 대한 자세한 내용은 SAS Institute Inc. (2020b) 의 "CANDISC Procedure" 장에서 확인하십시오 .

정준 구조 표시

"정준 구조" 보고서에는 정준 변수와 공변량 간의 상관을 제공하는 세 개의 행렬이 있습니다. 다른 행렬에는 그룹 변수 수준에 대한 평균이 표시됩니다. 행렬을 보려면 행렬 이름 옆의 표시 아이콘을 클릭합니다. 숨기려면 행렬 이름을 클릭합니다.

그림 5.14 그룹 간 정준 구조를 보여 주는 Iris.jmp 의 정준 구조

총 정준 구조 정준 변수와 공변량 간의 상관입니다 . 적재라고도 합니다 .

그룹 간 정준 구조 정준 변수에 대한 그룹 평균과 공변량에 대한 그룹 평균 간의 상관입니다 .

합동 그룹 내 정준 구조 정준 변수와 공변량 간의 부분 상관이며 그룹 변수에 대해 조정됩니다 .

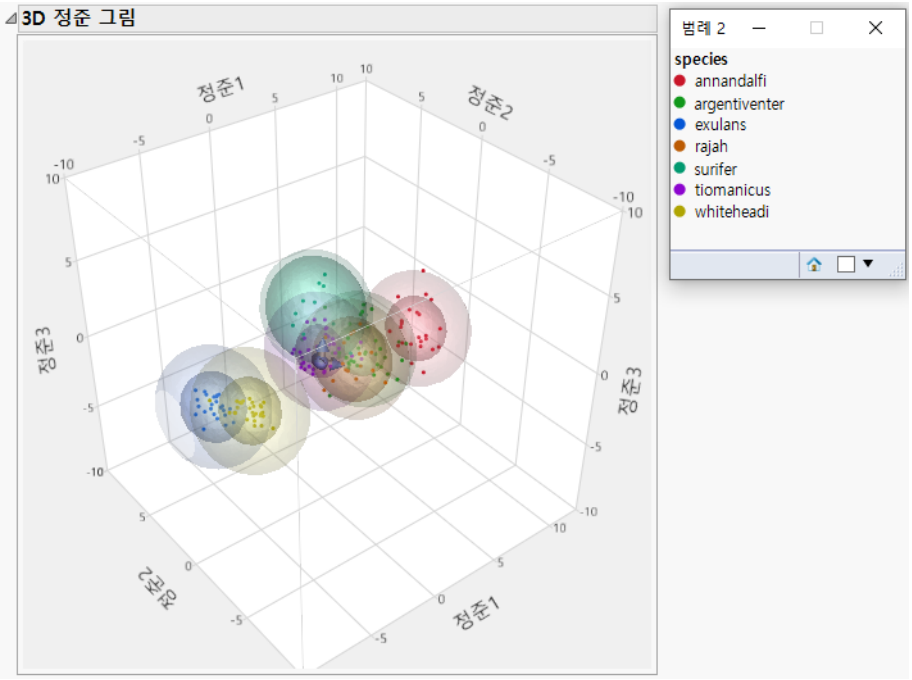
정준 변수에 대한 클래스 평균 각 정준 변수에 대한 그룹 변수 전체 수준의 평균을 제공합니다 .

3D 정준 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Owl Diet.jmp 를 엽니다 .
2. 180 행 ~ 294 행을 선택합니다 .
   이 행은 species 값이 결측입니다 . 이러한 행을 숨기고 제외합니다 .
3. **행 > 숨기고 제외하기**를 선택합니다 .
4. **행 > 열 값에 따른 색상 또는 표식**을 선택합니다 .
5. species 를 선택합니다 .
6. " 색상 " 메뉴에서 **JMP 진한 설정**을 선택합니다 .
7. **범례가 포함된 창 만들기**를 선택합니다 .
8. **확인**을 클릭합니다 .
   작은 "범례" 창이 나타납니다. 데이터 테이블의 행에 species별로 색상이 지정되어 있습니다.
9. **분석 > 다변량 방법 > 판별**을 선택합니다 .
10. skull length, teeth row, palatine foramen 및 jaw length 를 **Y, 공변량**으로 지정합니다 .
11. species 를 **X, 범주**로 지정합니다 .
12. **확인**을 클릭합니다 .
13. " 판별 분석 " 의 빨간색 삼각형을 클릭하고 **3D 정준 그림**을 클릭합니다 .

팁 : " 범례 " 에서 범주를 클릭하면 "3D 정준 그림 " 에서 해당 점이 강조 표시됩니다 . 3D 그림 내부를 클릭한 후 드래그하면 그림이 회전합니다 .

그림 5.15 범례 창이 있는 3D 정준 그림



사전 확률 지정

다음 옵션을 사용하여 사전 확률을 지정할 수 있습니다.

동등 확률 모든 그룹에 동일한 사전 확률을 할당합니다. 이 옵션이 기본값입니다.

발생 횟수에 비례 관측된 데이터에 나타나는 빈도에 비례하여 해당 그룹에 사전 확률을 할당합니다.

기타 사용자 사전 확률을 지정할 수 있습니다.

새 수준 고려

일부 관측값이 범주형 변수의 지정된 수준에 대한 이상치라고 의심되는 경우 "새 수준 고려" 옵션을 사용합니다. 이 옵션을 선택하면 새 수준의 사전 확률을 지정하라는 메뉴가 나타납니다.

새 그룹을 사용하는 것이 더 적합한 관측값은 "기타"라는 새 수준에 할당됩니다. "기타" 그룹에 소속될 확률은 이러한 관측값이 **그룹 구조가 가정되지 않은** 전체 관측값 집합의 분포를 따른다고 가정합니다. 따라서 공분산 구조와 관련된 정규 등고선이 상대적으로 넓어집니다. 거리 계산은 지정된 사전 확률에 따라 조정됩니다.

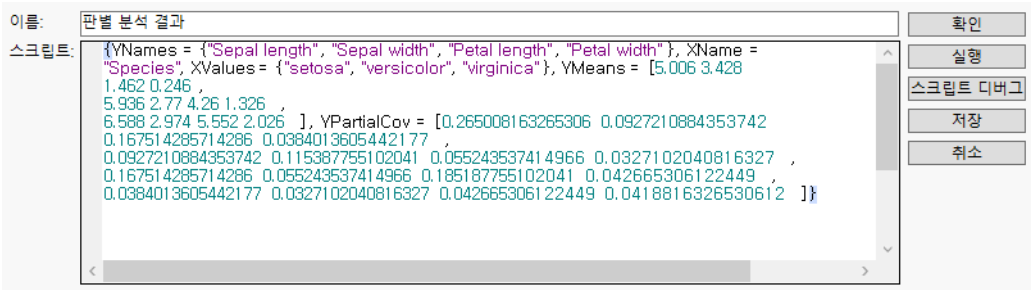
판별 행렬 저장

판별 행렬 저장은 JMP 스크립트 언어에서 사용할 전역 목록(DiscrimResults)을 생성합니다. 목록에는 훈련 데이터 집합에 대해 계산된 다음 항목이 포함됩니다.

- YNames - 공변량 (Y) 목록
- XName - 범주형 변수
- XValues - X 수준 목록
- YMeans - 공변량 평균 대 X 수준의 행렬
- YPartialCov - 그룹 내 공분산 행렬

Iris.jmp 샘플 데이터 테이블의 **Discriminant** 스크립트를 사용하여 얻은 분석을 예로 들어보겠습니다. "판별 분석"의 빨간색 삼각형 메뉴에서 **판별 행렬 저장**을 선택하면 **판별 분석 결과** 스크립트가 데이터 테이블에 저장됩니다.

그림 5.16 Iris.jmp 에 대한 판별 분석 결과 테이블 스크립트



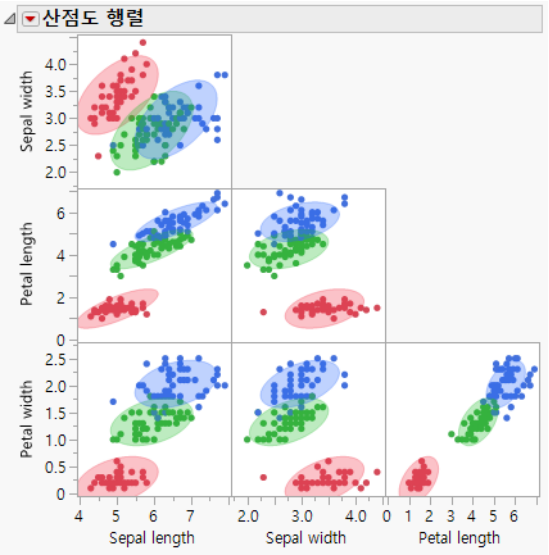
참고 : 스크립트에서 " 판별 행렬 가져오기 " 스크립트 명령을 판별 플랫폼 개체에 보낼 수 있습니다. 이렇게 하면 " 판별 행렬 저장 " 과 동일한 값을 얻을 수 있지만 데이터 테이블에 저장되지는 않습니다.

산점도 행렬

"산점도 행렬" 명령을 사용하면 공변량에 대한 하삼각 산점도 행렬이 포함된 산점도 행렬 플랫폼이 별도의 창에 호출됩니다. 데이터 테이블의 모든 관측값에 대한 점이 표시됩니다.

범주형 변수 X의 각 수준에 대해 90% 범위의 타원이 표시됩니다. 선형 판별 방법의 경우 합동 그룹 내 공분산 행렬을 기반으로 합니다. 그림 5.17에서는 Iris.jmp 샘플 데이터 테이블에 대한 "산점도 행렬" 창을 보여 줍니다.

그림 5.17 Iris.jmp 에 대한 산점도 행렬



"산점도 행렬"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 Essential Graphing 에서 확인하십시오.

JMP 및 JMP Pro 에서 검증

JMP에서는 검증 데이터 집합을 구성하는 행을 제외하여 검증 데이터 집합을 지정할 수 있습니다. 검증 데이터 집합으로 사용할 행을 선택한 후 **행 > 제외/제외 해제**를 선택합니다. 제외되지 않은 행은 훈련 데이터 집합으로 처리됩니다.

참고 : JMP Pro 에서는 판별 시작 창에서 검증 열을 지정할 수 있습니다 . 검증 열은 숫자 데이터 유형이어야 하고 두 개 이상의 구분되는 값을 포함해야 합니다 .

다음 사항에 유의하십시오 .

- 열에 두 개의 값이 포함된 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다 .
- 열에 세 개의 값이 포함된 경우 작은 값부터 순서대로 훈련 데이터 집합 , 검증 데이터 집합 , 테스트 데이터 집합을 각각 정의합니다 .
- 열에 네 개 이상의 구분되는 값이 포함된 경우 가장 작은 세 개의 값과 관련 관측값만 사용하여 순서대로 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합을 각각 정의합니다 .

검증 데이터 집합이 지정되면 판별 플랫폼에서 다음을 수행합니다 .

- 훈련 데이터를 사용하여 모형이 적합됩니다 .

- " 단계별 변수 선택 " 옵션을 선택하면 모형에 대한 " 검증 엔트로피 R^2" 및 " 검증 오분류 비율 " 통계량이 제공됩니다 . 자세한 내용은 " 통계량 " 및 " 검증 및 테스트 데이터 집합에 대한 엔트로피 R^2" 에서 확인하십시오 .
- " 판별 스코어 " 보고서에 검증 데이터 집합과 테스트 데이터 집합의 행을 식별하는 표시자가 나타납니다 .
- " 스코어 요약 " 보고서에 훈련 데이터 집합 , 검증 데이터 집합 및 테스트 데이터 집합에 대한 실제 분류 대 예측 분류가 표시됩니다 .

판별 분석 플랫폼에 대한 통계 상세 정보

- " 와이드 선형 알고리즘 설명 "
- " 저장된 계산식 "
- " 다변량 검정 "
- " 근사 F- 검정 "
- " 그룹 간 공분산 행렬 "

와이드 선형 알고리즘 설명

와이드 선형 판별 분석은 다음과 같이 수행됩니다 .

- 데이터는 그룹 평균을 빼고 합동 표준편차로 나누어 표준화됩니다 .
- 특이값 분해를 사용하여 특이 벡터 집합에서 주성분 변환 행렬을 얻습니다 .
- 유지된 성분 수는 특이값 제곱합의 최소 0.9999 를 나타냅니다 .
- 변환된 데이터에 대해 선형 판별 분석이 수행되며 , 이때 데이터가 그룹 평균에 의해 변화되지 않습니다 . 합동 그룹 내 공분산 행렬이 대각 행렬이므로 빠르게 계산할 수 있습니다 .

저장된 계산식

이 섹션에서는 **스코어 옵션 > 계산식 저장**을 통해 저장된 계산식의 파생을 보여 줍니다. 계산식은 판별 방법에 따라 달라집니다.

범주형 변수 X에 의해 정의된 각 그룹에 대해 공변량의 관측값은 p 차원 다변량 정규 분포를 따른다고 가정됩니다. 여기서 p는 공변량 수입니다. 계산식에 사용되는 표기는 표 5.2에 나와 있습니다.

표 5.2 계산식 저장 옵션을 통해 제공된 계산식의 표기

p	공변량 수
T	총 그룹 수(X 수준)

표 5.2 계산식 저장 옵션을 통해 제공된 계산식의 표기 (계속)

$t = 1,..., T$	X에 의해 정의된 그룹을 구별하기 위한 첨자
n_t	그룹 t의 관측값 수
$n = n_1 + n_2 + ... + n_T$	총 관측값 수
$\mathbf{y}$	관측값에 대한 p x 1 공변량 벡터
$\mathbf{y}_{it} = (y_{i1t}, y_{i2t}, \ldots, y_{ipt})$	p개 공변량의 벡터로 구성된 그룹 t의 i번째 관측값
$\bar{\mathbf{y}}_t$	그룹 t의 관측값에 대한 p x 1 공변량 벡터 $\mathbf{y}$의 평균
$\mathbf{y}_{bar}$	모든 관측값에 대한 p x 1 공변량 벡터의 평균
$\mathbf{S}_t = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (\mathbf{y}_{it} - \bar{\mathbf{y}}_t)(\mathbf{y}_{it} - \bar{\mathbf{y}}_t)'$	그룹 t에 대해 추정된 $(p \times p)$ 그룹 내 공분산 행렬
$\mathbf{S}_p = \frac{1}{n - T} \sum_{t=1}^{T} (n_t - 1)\mathbf{S}_t$	추정된 $(p \times p)$ 합동 그룹 내 공분산 행렬
q_t	그룹 t에 속할 사전 확률
$p(t \mid \mathbf{y})$	$\mathbf{y}$가 그룹 t에 속할 사후 확률
$\lvert\mathbf{A}\rvert$	행렬 A의 행렬식

선형 판별 방법

선형 판별 분석에서는 그룹 내 공분산 행렬이 모두 같다고 가정합니다. 공통 공분산 행렬은 $\mathbf{S}_p$로 추정됩니다. 표기는 표 5.2에서 확인하십시오.

관측값 $\mathbf{y}$ 에서 그룹 t 까지의 Mahalanobis 거리는 다음과 같이 정의됩니다 .

$$d_t^2 = (\mathbf{y} - \bar{\mathbf{y}}_t)' \mathbf{S}_p^{-1} (\mathbf{y} - \bar{\mathbf{y}}_t)$$

그룹 t 에서 관측값 $\mathbf{y}$ 의 가능도는 다음과 같이 추정됩니다 .

$$l_t(\mathbf{y}) = (2\pi)^{-T/2}\left|\mathbf{S}_p\right|^{-1/2}\exp(-(\mathbf{y}-\bar{\mathbf{y}}_t)'\mathbf{S}_p^{-1}(\mathbf{y}-\bar{\mathbf{y}}_t)/2)$$
$$= (2\pi)^{-T/2}\left|\mathbf{S}_p\right|^{-1/2}\exp(-d_t^2/2)$$

합동 공분산 행렬에 대해 추정해야 할 모수의 수는 $p(p+1)/2$ 이고 평균에 대해서는 Tp 입니다. 추정해야 할 모수의 총 수는 $p(p+1)/2 + Tp$ 입니다.

그룹 t 에 소속될 사후 확률은 다음과 같이 정의됩니다 .

$$p(t|\mathbf{y}) = \frac{q_t l_t(\mathbf{y})}{\sum_{u=1}^{T} q_u l_u(\mathbf{y})} = \frac{1}{1+\sum_{u\neq t}\exp(-[(d_u^2-2\log(q_u))-(d_t^2-2\log(q_t))]/2)}$$

관측값 $\mathbf{y}$ 는 사후 확률이 가장 큰 그룹에 할당됩니다.

선형 판별 방법에 의해 저장되는 계산식은 다음과 같이 정의됩니다 .

SqDist[0]	$\mathbf{y}'\mathbf{S}_p^{-1}\mathbf{y}$
SqDist[<group t>]	$d_t^2 - 2\log(q_t)$
Prob[<group t>]	$p(t\|\mathbf{y})$
Pred <X>	$p(t\|\mathbf{y})$ 가 최대값이 되는 $t(t = 1, ..., T)$

2 차 판별 방법

2차 판별 분석에서는 그룹 내 공분산 행렬이 같다고 가정되지 않습니다. 그룹 t에 대한 그룹 내 공분산 행렬은 $\mathbf{S}_t$로 추정됩니다. 따라서 그룹 내 공분산 행렬에 대해 추정해야 할 모수의 수는 $Tp(p+1)/2$ 이고 평균에 대해서는 Tp 입니다. 추정해야 할 모수의 총 수는 $Tp(p+3)/2$ 입니다.

그룹 표본 크기가 p에 비해 작으면 그룹 내 공분산 행렬의 추정값이 크게 변동되는 경향이 있습니다. 판별 스코어는 그룹 내 공분산 행렬에 대한 역행렬의 가장 작은 고유값에 크게 영향을 받습니다. 자세한 내용은 Friedman(1989) 연구 자료에서 확인하십시오. 따라서 그룹 표본 크기가 p 에 비해 작은 경우 "정규화 판별 방법"에 설명된 정규화 방법을 고려할 수도 있습니다.

표기는 표 5.2 에서 확인하십시오 . 관측값 **y** 에서 그룹 *t* 까지의 Mahalanobis 거리는 다음과 같이 정의됩니다 .

$$d_t^2 = (\mathbf{y} - \bar{\mathbf{y}}_t)' \mathbf{S}_t^{-1} (\mathbf{y} - \bar{\mathbf{y}}_t)$$

그룹 *t* 에서 관측값 **y** 의 가능도는 다음과 같이 추정됩니다 .

$$\begin{aligned} l_t(\mathbf{y}) &= (2\pi)^{-T/2} \left|\mathbf{S}_t\right|^{-1/2} \exp(-(\mathbf{y} - \bar{\mathbf{y}}_t)' \mathbf{S}_t^{-1} (\mathbf{y} - \bar{\mathbf{y}}_t)/2) \\ &= (2\pi)^{-T/2} \left|\mathbf{S}_t\right|^{-1/2} \exp(-d_t^2/2) \end{aligned}$$

그룹 *t* 에 소속될 사후 확률은 다음과 같습니다 .

$$\begin{aligned} p(t|\mathbf{y}) &= (q_t l_t(\mathbf{y})) / \left(\sum_{u=1}^{T} q_u l_u(\mathbf{x}) \right) \\ &= \frac{1}{1 + \sum_{u \neq t} \exp(-[(d_u^2 + \log\left|\mathbf{S}_u\right| - 2\log(q_u)) - (d_t^2 + \log\left|\mathbf{S}_t\right| - 2\log(q_t))]/2)} \end{aligned}$$

관측값 **y** 는 사후 확률이 가장 큰 그룹에 할당됩니다 .

2 차 판별 방법에 의해 저장되는 계산식은 다음과 같이 정의됩니다 .

SqDist[<group *t*>]	$d_t^2 + \log\left\|\mathbf{S}_t\right\| - 2\log(q_t)$
Prob[<group *t*>]	$p(t\|\mathbf{y})$
Pred <*X*>	$p(t\|\mathbf{y})$ 가 최대값이 되는 $t(t = 1, ..., T)$

참고 : SqDist[<group *t*>] 는 음수일 수 있습니다 .

정규화 판별 방법

정규화 판별 분석에서는 λ 와 γ 라는 두 모수를 사용할 수 있습니다 .

- λ 모수는 동일하다고 가정되지 않는 합동 공분산 행렬과 그룹 내 공분산 행렬에 할당된 가중치의 균형을 맞춥니다 .

- γ 모수는 대각 행렬에 가깝게 축소할 크기를 결정합니다 .

이 방법을 사용하면 정규화의 두 가지 측면을 활용하여 2차 판별 분석의 추정값에 안정성을 부여할 수 있습니다. 자세한 내용은 Friedman(1989) 연구 자료에서 확인하십시오. 표기는 표 5.2에서 확인하십시오.

정규화 방법의 경우 그룹 t 에 대한 공분산 행렬은 다음과 같습니다 .

$$\Sigma_t = (1-\gamma)(\lambda \mathbf{S}_p + (1-\lambda)\mathbf{S}_t) + \gamma Diag((\lambda \mathbf{S}_p + (1-\lambda)\mathbf{S}_t))$$

관측값 $\mathbf{y}$ 에서 그룹 t 까지의 Mahalanobis 거리는 다음과 같이 정의됩니다 .

$$d_t^2 = (\mathbf{y} - \bar{\mathbf{y}}_t)' \Sigma_t^{-1} (\mathbf{y} - \bar{\mathbf{y}}_t)$$

그룹 t 에서 관측값 $\mathbf{y}$ 의 가능도는 다음과 같이 추정됩니다 .

$$\begin{aligned} l_t(\mathbf{y}) &= (2\pi)^{-T/2} \left|\Sigma_t\right|^{-1/2} \exp(-(\mathbf{y} - \bar{\mathbf{y}}_t)' \Sigma_t^{-1} (\mathbf{y} - \bar{\mathbf{y}}_t)/2) \\ &= (2\pi)^{-T/2} \left|\Sigma_t\right|^{-1/2} \exp(-d_t^2/2) \end{aligned}$$

그룹 t 에 소속될 사후 확률은 다음과 같이 지정됩니다 .

$$\begin{aligned} p(t|\mathbf{y}) &= (q_t l_t(\mathbf{y}))/\left(\sum_{u=1}^{T} q_u l_u(\mathbf{x}) \right) \\ &= \frac{1}{1 + \sum_{u \neq t} \exp(-[(d_u^2 + \log\left|\Sigma_u\right| - 2\log(q_u)) - (d_t^2 + \log\left|\Sigma_t\right| - 2\log(q_t))]/2)} \end{aligned}$$

관측값 $\mathbf{y}$는 사후 확률이 가장 큰 그룹에 할당됩니다.

정규화 판별 방법에 의해 저장되는 계산식은 아래와 같이 정의됩니다 .

SqDist[<group *t*>]	$d_t^2 + \log\left\|\Sigma_t\right\| - 2\log(q_t)$
Prob[<group *t*>]	$p(t\|\mathbf{y})$
Pred <*X*>	$p(t\|\mathbf{y})$ 가 최대값이 되는 $t(t = 1, ..., T)$

참고 : SqDist[<group *t*>] 는 음수일 수 있습니다 .

와이드 선형 판별 방법

와이드 선형 방법은 공변량 수가 많은 경우, 특히 공변량 수가 관측값 수를 초과($p > n$)하는 경우에 유용합니다. 이 방법은 합동 그룹 내 공분산 행렬 $\mathbf{S}_p$의 역행렬 또는 전치를 효율적으로 계산하는 것에 중점을 둡니다($p > n$인 경우). 여기에서는 큰 공분산 행렬에 대한 공간 할당 및 역행렬 계산이 발생하지 않도록 특이값 분해를 사용합니다.

와이드 선형 방법은 그룹 내 공분산 행렬이 같다고 가정하며, 관측값 수가 공변량 수와 같거나 이를 초과하는 경우 선형 방법과 동일합니다.

와이드 선형 계산

표기는 표 5.2 에서 확인하십시오 . 다음 단계를 사용하여 와이드 선형 계산을 수행합니다 .

1. 그룹 내 표본 평균의 T x p 행렬 **M** 을 계산합니다 . **M** 의 (t,j) 번째 항목인 m_{tj} 는 j 번째 공변량에 대한 그룹 t 멤버의 표본 평균입니다 .
2. 각 공변량 j 에 대해 그룹 전체의 합동 표준편차를 계산합니다 . 이를 s_{jj} 라고 합니다 .
3. 대각 항목 s_{jj} 가 있는 대각 행렬을 $\mathbf{S}_{diag}$ 로 나타냅니다 .
4. 각 공변량에 대한 값을 다음과 같이 중심화 및 척도화합니다 .
   - 관측값이 속한 그룹의 평균을 뺍니다 .
   - 차이를 합동 표준편차로 나눕니다 .

   그룹 t 의 관측값 i 에 대해 j 번째 공변량의 그룹 중심화 및 척도화된 값은 다음과 같이 표기합니다 .

$$y^*_{ij} = \frac{y_{ij} - m_{t(i)j}}{s_{jj}}$$

   $t(i)$ 표기는 관측값 i 가 속한 그룹 t 를 나타냅니다 .
5. y^*_{ij} 값 행렬을 $\mathbf{Y}_s$ 로 나타냅니다 .
6. 그룹 중심화 및 척도화된 공변량의 합동 그룹 내 공분산 행렬을 **R** 로 나타냅니다 . **R** 행렬은 다음과 같이 지정됩니다 .

$$\mathbf{R} = (\mathbf{Y}_s'\mathbf{Y}_s)/(n - T)$$

7. $\mathbf{Y}_s$ 에 특이값 분해를 적용합니다 .

$$\mathbf{Y}_s = \mathbf{UDV}'$$

   여기서 **U** 와 **V** 는 직교정규이며 **D** 는 대각선에 양수 항목 (특이값) 이 있는 대각 행렬입니다 . 자세한 내용은 " 특이값 분해 " 에서 확인하십시오 .

그러면 **R** 을 다음과 같이 쓸 수 있습니다 .

$$\mathbf{R} = (\mathbf{Y}_s'\mathbf{Y}_s)/(n-T) = (\mathbf{V}\mathbf{D}^2\mathbf{V}')/(n-T)$$

8. **R** 이 완전 계수인 경우 $\mathbf{R}^{-1/2}$ 은 다음과 같습니다 .

$$\mathbf{R}^{-1/2} = (\mathbf{V}\mathbf{D}^{-1}\mathbf{V}')/\sqrt{n-T}$$

여기서 $\mathbf{D}^{-1}$ 은 대각 항목이 **D** 의 대각 항목에 대한 역행렬인 대각 행렬입니다 .
R 이 완전 계수가 아닌 경우 **R** 에 대한 유사역행렬을 다음과 같이 정의합니다 .

$$\mathbf{R}^- = (\mathbf{V}\mathbf{D}^{-2}\mathbf{V}')/(n-T)$$

그러면 **R**의 역제곱근을 다음과 같이 정의합니다.

$$(\mathbf{R}^-)^{1/2} = (\mathbf{V}\mathbf{D}^{-1}\mathbf{V}')/\sqrt{n-T}$$

9. **R** 이 완전 계수이면 $\mathbf{R}^- = \mathbf{R}^{-1}$ 이 됩니다 . 따라서 완비성을 위해 유사역행렬을 사용하여 계속합니다 .
$p \times p$ 행렬 $\mathbf{T}_s$ 를 다음과 같이 정의합니다 .

$$\mathbf{T}_s = (\mathbf{S}_{diag}^{-1}\mathbf{V}\mathbf{D}^-)/(\sqrt{n-T})$$

그러면 다음과 같이 구성할 수 있습니다 .

$$(\mathbf{T}_s\mathbf{T}_s') = \left(\mathbf{S}_{diag}^{-1}\mathbf{V}(\mathbf{D}^-)^2\mathbf{V}'\mathbf{S}_{diag}^{-1}\right)/(n-T) = \mathbf{S}_{diag}^{-1}\mathbf{R}^-\mathbf{S}_{diag}^{-1} = \mathbf{S}_p^-$$

여기서 $\mathbf{S}_p^-$ 는 원래 데이터에 대해 SVD 를 사용하여 계산된 합동 그룹 내 공분산 행렬의 일반화 역행렬입니다 .

Mahalanobis 거리

Mahalanobis 거리, 가능도 및 사후 확률에 대한 계산식은 "선형 판별 방법"의 계산식과 동일합니다. 그러나 $\mathbf{S}_p$의 역행렬은 특이값 분해를 사용하여 계산된 일반화 역행렬로 대체됩니다.

계산식을 저장하면 분해 형식으로 Mahalanobis 거리가 제공됩니다 . 관측값 **y** 에 대해 그룹 *t* 까지의 제곱 거리는 다음과 같습니다 . 여기서 마지막 등식의 *SqDist*[0] 및 **판별 주성분**은 " 저장된 계산식 " 에 정의되어 있습니다 .

$$\begin{aligned} d_t^2 &= (\mathbf{y}-\bar{\mathbf{y}}_t)'\bar{\mathbf{S}}_p(\mathbf{y}-\bar{\mathbf{y}}_t) \\ &= (\mathbf{y}-\bar{\mathbf{y}}_t)'\mathbf{T}_s\mathbf{T}_s'(\mathbf{y}-\bar{\mathbf{y}}_t) \\ &= ((\mathbf{y}-\bar{\mathbf{y}})-(\bar{\mathbf{y}}_t-\bar{\mathbf{y}}))'\mathbf{T}_s\mathbf{T}_s'((\mathbf{y}-\bar{\mathbf{y}})-(\bar{\mathbf{y}}_t-\bar{\mathbf{y}})) \\ &= (\mathbf{T}_s'(\mathbf{y}-\bar{\mathbf{y}}))'(\mathbf{T}_s'(\mathbf{y}-\bar{\mathbf{y}}))-2(\mathbf{T}_s'(\bar{\mathbf{y}}_t-\bar{\mathbf{y}}))'(\mathbf{T}_s'(\mathbf{y}-\bar{\mathbf{y}}))+(\mathbf{T}_s'(\bar{\mathbf{y}}_t-\bar{\mathbf{y}}))'(\mathbf{T}_s'(\bar{\mathbf{y}}_t-\bar{\mathbf{y}})) \\ &= SqDist[0]-2(\mathbf{T}_s'(\bar{\mathbf{y}}_t-\bar{\mathbf{y}}))'Discrim\ Prin\ Comp+(\mathbf{T}_s'(\bar{\mathbf{y}}_t-\bar{\mathbf{y}}))'(\mathbf{T}_s'(\bar{\mathbf{y}}_t-\bar{\mathbf{y}})) \end{aligned}$$

저장된 계산식

다음은 와이드 선형 판별 방법에 의해 저장되는 계산식입니다 .

판별 데이터 행렬	공변량에 대한 관측값 벡터
판별 주성분	데이터를 그룹 내에서 상관관계가 없도록 렌더링하는 주성분 스코어링 행렬에 의해 변환된 데이터입니다. $\mathbf{T}_s'(\mathbf{y}-\bar{\mathbf{y}})$ 로 지정되며, 여기서 $\bar{\mathbf{y}}$ 는 전체 평균을 포함하는 p x 1 벡터입니다.
SqDist[0]	$(\mathbf{y}-\bar{\mathbf{y}})'\mathbf{T}'_s\mathbf{T}_s'(\mathbf{y}-\bar{\mathbf{y}})$
SqDist[<group *t*>]	관측값에서 그룹 중심까지의 Mahalanobis 거리입니다. 자세한 내용은 "Mahalanobis 거리"에서 확인하십시오.
Prob[<group *t*>]	$p(t\|\mathbf{y})$ ("선형 판별 방법"에 제공됨)
Pred <*X*>	$p(t\|\mathbf{y})$ 가 최대값이 되는 $t(t = 1, ..., T)$

다변량 검정

아래에서 **E**는 잔차 교차곱 행렬이고, **H**는 모형 교차곱 행렬입니다. **E**의 주대각 요소는 각 변수의 잔차 제곱합입니다. **H**의 주대각 요소는 각 변수에 대한 모형의 제곱합입니다. 판별 분석 관련 문헌에서는 종종 **E**를 **W**라고 합니다(**W**는 within을 나타냄).

다변량 결과 표의 검정 통계량은 $\mathbf{E}^{-1}\mathbf{H}$ 의 고유값 λ를 사용해서 구합니다. 다음 목록에서는 각 검정 통계량의 계산 방법을 설명합니다.

참고 : 반응 설계가 지정된 후 초기 **E** 및 **H** 행렬의 앞에 **M**′ 를 곱하고 뒤에 **M** 을 곱합니다 .

- Wilks 람다

$$\Lambda = \frac{\det(\mathbf{E})}{\det(\mathbf{H}+\mathbf{E})} = \prod_{i=1}^{n}\left(\frac{1}{1+\lambda_i}\right)$$

- Pillai 대각합

$$V = \text{Trace}[\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}] = \sum_{i=1}^{n}\frac{\lambda_i}{1+\lambda_i}$$

- Hotelling-Lawley 대각합

$$U = \text{Trace}(\mathbf{E}^{-1}\mathbf{H}) = \sum_{i=1}^{n}\lambda_i$$

- Roy 최대근

$\Theta = \lambda_1$, $\mathbf{E}^{-1}\mathbf{H}$ 의 최대 고유값

E 와 **H** 는 다음과 같이 정의됩니다.

$$\mathbf{E} = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'(\mathbf{X}'\mathbf{X})\mathbf{b}$$

$$\mathbf{H} = (\mathbf{L}\mathbf{b})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-1}(\mathbf{L}\mathbf{b})$$

여기서 **b** 는 모형 계수의 추정 벡터이며, **A** 는 행렬 **A** 의 일반화 역행렬을 나타냅니다.

전체 모형 **L** 은 0(절편 값)으로 구성된 열로, 행 및 열 개수가 모형의 모수 개수와 같은 단위 행렬과 연결되어 있습니다. 효과에 대한 **L** 행렬은 전체 모형 **L** 행렬의 일부 행으로 구성한 부분집합입니다.

근사 F- 검정

F 값과 자유도를 계산하려면 p 의 순위가 $\mathbf{H}+\mathbf{E}$ 가 되도록 하고 q 의 순위는 $\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'$ 가 되도록 합니다. 여기서 **L** 행렬은 검정 중인 효과와 관련된 $\mathbf{X}'\mathbf{X}$ 의 요소를 식별합니다. v 는 오차 자유도가 되도록 하고 s 는 p 및 q 의 최소값이 되도록 합니다. 또한 $m = 0.5(|p-q|-1)$ 및 $n = 0.5(v-p-1)$ 이 되도록 합니다.

표 5.3(110 페이지)에서는 해당 검정 통계량에서 구한 각 근사 F 의 계산을 보여 줍니다.

표 5.3 근사 F 통계량

검정	근사 F	분자 DF	분모 DF
Wilks 람다	$F = \left(\frac{1-\Lambda^{1/t}}{\Lambda^{1/t}}\right)\left(\frac{rt-2u}{pq}\right)$	pq	$rt-2u$
Pillai 대각합	$F = \left(\frac{V}{s-V}\right)\left(\frac{2n+s+1}{2m+s+1}\right)$	$s(2m+s+1)$	$s(2n+s+1)$
Hotelling-Lawley 대각합	$F = \frac{2(sn+1)U}{s^2(2m+s+1)}$	$s(2m+s+1)$	$2(sn+1)$
Roy 최대근	$F = \frac{\Theta(v-\max(p,q)+q)}{\max(p,q)}$	$\max(p,q)$	$v-\max(p,q)+q$

그룹 간 공분산 행렬

이 행렬은 표 5.2 의 표기를 사용하여 다음과 같이 정의됩니다 .

$$\mathbf{S}_B = \frac{1}{T-1}\sum_{t=1}^{T} T\left(\frac{n_t}{n}\right)(\bar{\mathbf{y}}_t - \mathbf{y}_{bar})(\bar{\mathbf{y}}_t - \mathbf{y}_{bar})'$$

6 장

부분 최소 제곱 모형

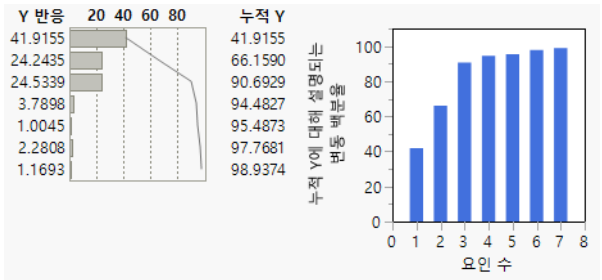
Y 와 X 간의 상관을 사용하여 모형 개발

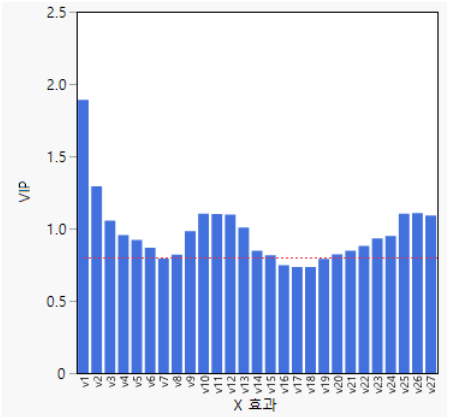
PLS(부분 최소 제곱) 플랫폼에서는 요인, 즉 설명 변수(*X*)의 선형 결합을 기반으로 선형 모형을 적합시킵니다. 이러한 요인은 *X*와 하나 이상의 반응(*Y*) 간의 공분산을 최대화하는 방식으로 얻어집니다. PLS는 *X*와 *Y* 사이의 상관을 이용하여 기본 잠재 구조를 나타냅니다.

JMP PRO JMP Pro에서는 PLS-DA(PLS 판별 분석) 수행, 다양한 모형 효과 포함, 여러 검증 방법 사용, 결측 데이터 대치, 다양한 통계량 분포의 붓스트랩 추정값 계산 등의 작업이 가능한 추가 기능을 제공합니다.

부분 최소 제곱은 X 변수가 관측값보다 많은 경우, X 변수 간의 상관관계가 높은 경우, X 변수가 많은 경우, Y 변수가 몇 개 있고 X 변수가 많은 경우와 같이 보통 최소 제곱을 사용하여 만족스러운 결과를 얻을 수 없는 상황에서 효과적입니다.

그림 6.1 부분 최소 제곱 보고서 (일부)





목차

부분 최소 제곱 플랫폼 개요 113
부분 최소 제곱의 예 114
부분 최소 제곱 플랫폼 시작 117
중심화 및 척도화 120
X 표준화 120
모형 시작 제어판 121
부분 최소 제곱 옵션 122
부분 최소 제곱 보고서 123
모형 비교 요약 123
교차 검증 보고서 124
모형 적합 보고서 128
모형 적합 옵션 129
변수 중요도 그림 131
VIP 대 계수 그림 132
열 저장 133
부분 최소 제곱의 추가 예 134
부분 최소 제곱 플랫폼에 대한 통계 상세 정보 136
부분 최소 제곱 136
van der Voet T^2 137
T^2 그림 138
X 스코어 산점도 행렬의 신뢰도 타원 138
예측 표준 오차 및 신뢰 한계 139
표준화된 스코어 및 적재량 140
PLS-DA(PLS 판별 분석) 141

부분 최소 제곱 플랫폼 개요

보통 최소 제곱과 달리 예측 변수가 관측값보다 많을 때 PLS를 사용할 수 있습니다. PLS는 분광학, 계량화학, 유전체학, 심리학, 교육, 경제학, 정치학 및 환경 과학 등의 분야에서 고차원 데이터를 모델링하는 데 광범위하게 사용됩니다.

PLS 방법을 사용한 모형 적합은 설명 변수가 관측값보다 많거나 설명 변수의 상관관계가 높을 때 특히 유용합니다. PLS를 사용하여 단일 모형을 여러 반응에 동시에 적합시킬 수 있습니다. 자세한 내용은 Garthwaite(1994), Wold(1994), Wold et al. (2001), Eriksson et al. (2006) 및 Cox 와 Gaudard(2013) 연구 자료에서 확인하십시오.

NIPALS(Nonlinear Iterative Partial Least Squares) 와 SIMPLS("Statistically Inspired Modification of PLS") 라는 두 가지 모형 적합 알고리즘을 사용할 수 있습니다. NIPALS에 대한 자세한 내용은 Wold(1980) 연구 자료에서 확인하십시오. SIMPLS에 대한 자세한 내용은 De Jong(1993) 연구 자료에서 확인하십시오. 두 방법에 대한 설명은 Boulesteix와 Strimmer(2007) 연구 자료에서 확인하십시오. SIMPLS 알고리즘은 특정 최적성 문제의 해를 구하는 것을 목표로 개발되었습니다. 단일 반응의 경우 두 방법 모두 동일한 모형을 제공합니다. 다중 반응의 경우 약간 차이가 있습니다.

JMP에서는 분석 > 다변량 방법 > 부분 최소 제곱을 통해서만 PLS 플랫폼에 액세스할 수 있습니다. JMP Pro에서는 분석 > 모형 적합을 통해 부분 최소 제곱 분석법에 액세스할 수도 있습니다.

JMP PRO JMP Pro 에서는 다음을 수행할 수 있습니다 .

- 모형 적합에서 부분 최소 제곱 분석법을 사용하여 명목형 모델링 유형의 반응을 적합시키는 방식으로 PLS-DA(PLS 판별 분석) 를 수행합니다 .
- 모형 적합에서 부분 최소 제곱 분석법을 사용하여 다항식 , 교호작용 및 범주형 효과를 적합시킵니다 .
- 검증 및 교차 검증을 위한 여러 방법 중에서 선택합니다 .
- 결측 데이터를 대치합니다 .
- 다양한 통계량 분포의 붓스트랩 추정값을 구합니다 . 관심 있는 보고서를 마우스 오른쪽 버튼으로 클릭할 수 있습니다 . 붓스트랩 추정값에 대한 자세한 내용은 **기본 분석**에서 확인하십시오 .

부분 최소 제곱은 추출할 최적 요인 수를 선택하는 데 도움이 되도록 van der Voet T^2 검정 및 교차 검증을 사용합니다 .

- JMP 플랫폼에서는 leave-one-out 교차 검증 방법을 사용합니다 . 검증을 사용하지 않도록 선택할 수도 있습니다 .
- JMP PRO JMP Pro 에서는 K 폴드 , Leave-One-Out 또는 랜덤 홀드백 교차 검증을 선택하거나 검증 열을 지정할 수 있습니다 . 검증을 사용하지 않도록 선택할 수도 있습니다 .

부분 최소 제곱의 예

이 예는 부분 최소 제곱이 매우 효과적인 분광계 보정 분야에서 가져왔습니다. 발트해 오염에 대한 연구를 진행 중이라고 가정해 보겠습니다. 해수 표본 스펙트럼을 사용하여 표본에 존재하는 세 가지 화합물의 양을 확인하려고 합니다.

세 가지 관심 화합물은 다음과 같습니다 .

- 리그닌술포산염 (ls) - 펄프 산업 공해
- 부식산 (ha) - 천연 임산물
- 세제 (dt) 의 형광 표백제

반응은 각 표본에서 해당 화합물의 양입니다. 예측 변수는 v1 ~ v27 의 파장 범위로 측정된 스펙트럼 방출 강도입니다.

모형 보정을 위해 화학적 조성이 알려진 표본을 사용합니다 . 보정 데이터는 농도가 알려진 리그닌술포산염 , 부식산 및 세제의 16 개 표본으로 구성됩니다 . 27 개의 등거리 파장에서 방출 강도가 기록됩니다 . 부분 최소 제곱 플랫폼을 사용하여 스펙트럼 방출 강도에서 화합물 양을 예측하는 모형을 생성할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Baltic.jmp 를 엽니다 .

   참고 : Baltic.jmp 데이터 테이블의 데이터는 Umetrics(1995) 에 보고되어 있습니다 . 원본 출처는 Lindberg, Persson 및 Wold(1983) 연구 자료입니다 .

2. **분석 > 다변량 방법 > 부분 최소 제곱**을 선택합니다 .
3. ls, ha 및 dt 를 **Y, 반응** 역할에 할당합니다 .
4. 27 개의 강도 변수 (v1 ~ v27) 가 포함된 Intensities 를 **X, 요인** 역할에 할당합니다 .
5. **확인**을 클릭합니다 .

   부분 최소 제곱 모형 시작 제어판이 나타납니다 .
6. " 검증 방법 " 에서 **Leave-One-Out** 을 선택합니다 .
7. **시작**을 클릭합니다 .

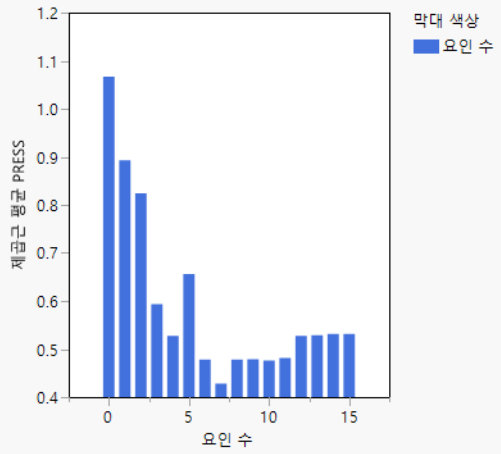
   Van der Voet 검정은 랜덤화 검정이므로 "Prob > van der Voet T^2" 값이 약간 다를 수 있습니다 .

그림 6.2 부분 최소 제곱 보고서

방법=NIPALS을(를) 사용한 Leave-One-Out 교차 검증(빠름 SVD 사용)

요인 수	제곱근 평균 PRESS	van der Voet T²	Prob > van der Voet T²	Q²	누적 Q²	R²X	누적 R²X	R²Y	누적 R²Y
0	1.066667	10.070473	0.0010*	-0.137778	-0.137778	0.000000	0.000000	0.000000	0.000000
1	0.892851	11.141942	<.0001*	-0.098077	-0.098077	0.974480	0.974480	0.421421	0.421421
2	0.824551	8.605908	<.0001*	0.773385	0.751159	0.022034	0.996514	0.247724	0.669146
3	0.593922	8.034522	0.0080*	0.774570	0.943904	0.001772	0.998286	0.241683	0.910829
4	0.527742	7.879423	0.0080*	0.656229	0.980716	0.001153	0.999438	0.037043	0.947872
5	0.656013	6.883024	0.0180*	0.356257	0.987586	0.000386	0.999825	0.012502	0.960374
6	0.478479	3.446269	0.2890	0.640049	0.995532	0.000113	0.999938	0.019350	0.979724
7	0.428141	0.000000	1.0000	0.809320	0.999148	0.000017	0.999955	0.010643	0.990367
8	0.478562	7.031041	0.0100*	0.818516	0.999845	0.000011	0.999967	0.004666	0.995032
9	0.478739	3.066627	0.3990	0.810850	0.999971	0.000012	0.999979	0.001452	0.996485
10	0.476024	3.847943	0.2420	0.845631	0.999995	0.000007	0.999986	0.001417	0.997902
11	0.481440	4.822874	0.1100	0.820847	0.999999	0.000004	0.999990	0.001248	0.999150
12	0.527436	3.968884	0.2040	0.810490	1.000000	0.000004	0.999994	0.000637	0.999787
13	0.529249	4.028960	0.2070	0.812180	1.000000	0.000003	0.999997	0.000158	0.999945
14	0.531553	4.408278	0.1670	0.816103	1.000000	0.000003	1.000000	0.000055	1.000000
15	0.531553	4.408278	0.1800	0.816103	1.000000	0.000000	1.000000	0.000000	1.000000

제곱근 평균 PRESS 그림

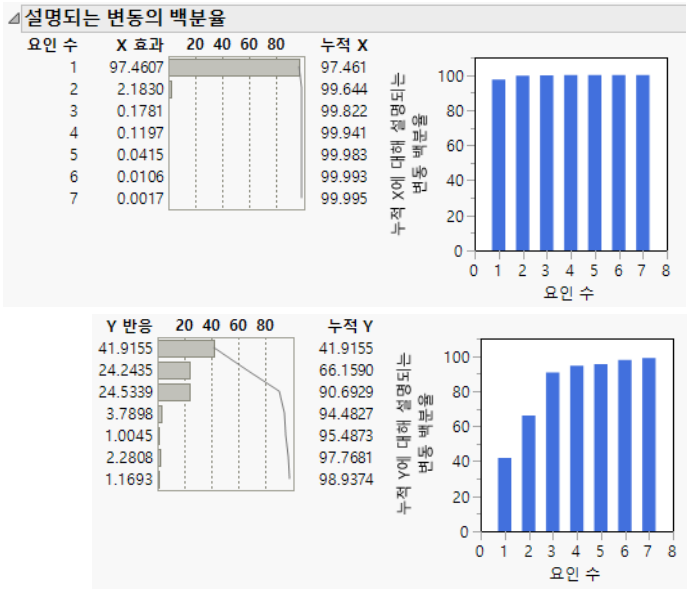


참고: 최소 제곱근 평균 PRESS는 0.42814이고 최소 요인 수는 7입니다.

" 제곱근 평균 PRESS(예측 잔차 제곱합) 그림 " 에서는 요인 수가 7 일 때 제곱근 평균 PRESS 가 최소화된다는 것을 보여 줍니다 . 이 내용은 " 제곱근 평균 PRESS 그림 " 아래의 " 참고 " 에 명시되어 있습니다 . **7 개의 요인을 사용한 NIPALS 적합 (빠른 SVD 사용)** 보고서가 생성되었습니다 . 그림 6.3 에 보고서의 일부가 표시되어 있습니다 .

van der Voet T^2 검정 통계량은 요인 수가 다른 모형이 최소 PRESS 값을 가진 모형과 유의하게 다른지 여부를 확인하는 데 사용됩니다 . 일반적인 방법은 van der Voet 유의 수준이 0.10 을 초과하는 가장 작은 수의 요인을 추출하는 것입니다 (SAS Institute Inc 2020f, Tobias 1995 참조). 이 방법을 여기에 적용하려면 **모형 시작** 패널의 **요인 수**에 6 을 입력하여 새 모형을 적합시킵니다 .

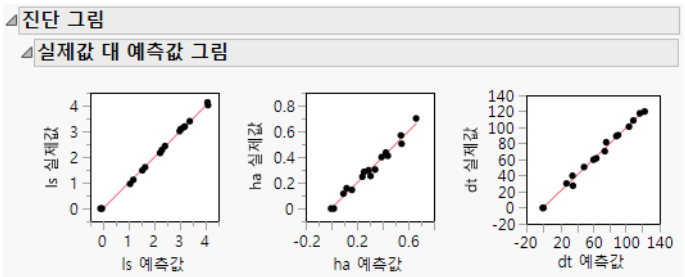
그림 6.3 추출된 7 개 요인



8. "7 개의 요인을 사용한 NIPALS 적합 (빠른 SVD 사용)" 의 빨간색 삼각형을 클릭하고 **진단 그림**을 선택합니다 .

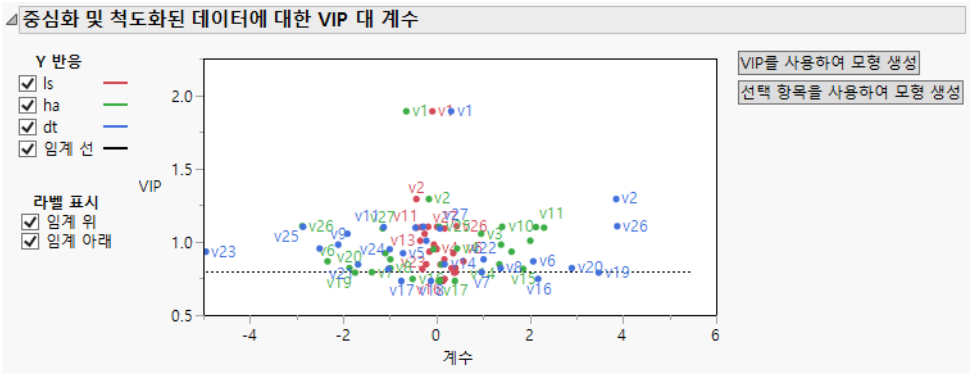
   그러면 실제값 대 예측값 그림을 보여 주는 보고서와 다양한 잔차 그림을 보여 주는 세 개의 보고서가 제공됩니다 . " 실제값 대 예측값 그림 " 에서는 예측 화합물 양이 실제 양과 일치하는 정도를 보여 줍니다 .

그림 6.4 진단 그림



9. "7 개의 요인을 사용한 NIPALS 적합 (빠른 SVD 사용)" 의 빨간색 삼각형을 클릭하고 **VIP 대 계수 그림**을 선택합니다 .

그림 6.5 VIP 대 계수 그림



VIP 대 계수 그림은 다양한 반응의 적합도와 관련하여 영향력 있는 변수를 식별하는 데 도움이 됩니다. 예를 들어 v23, v2 및 v26 은 VIP 값이 0.8 을 초과하고 계수도 상대적으로 큽니다.

부분 최소 제곱 플랫폼 시작

다음과 같은 두 가지 방법으로 부분 최소 제곱 플랫폼을 시작할 수 있습니다.

- **분석 > 다변량 방법 > 부분 최소 제곱**을 선택합니다.
- JMP PRO **분석 > 모형 적합**을 선택하고 " 분석법 " 메뉴에서 **부분 최소 제곱**을 선택합니다. 이 방법을 사용하면 다음을 수행할 수 있습니다.
  - 범주형 변수를 *Y* 또는 *X* 로 입력합니다. 범주형 *Y* 를 입력하여 PLS-DA 를 수행합니다.
  - 교호작용 및 다항식 항을 모형에 추가합니다.
  - "X 표준화 " 옵션을 사용하여 중심화 및 척도화된 열로 고차 항을 구성합니다.
  - 모형 규격 스크립트를 저장합니다.

모형 적합 시작 창의 일부 기능은 부분 최소 제곱 분석법에 사용할 수 없습니다.

- 가중치, 내포, 속성, 변환, 절편 없음

  팁 : " 열 선택 " 상자에서 변수를 마우스 오른쪽 버튼으로 클릭하고 " 변환 " 옵션을 선택하여 변수를 변환할 수 있습니다.

- 매크로 : 혼합물 반응 표면, Scheffe 입방

그림 6.6 JMP Pro 부분 최소 제곱 시작 창 (결측값 대치법에 EM 선택됨)

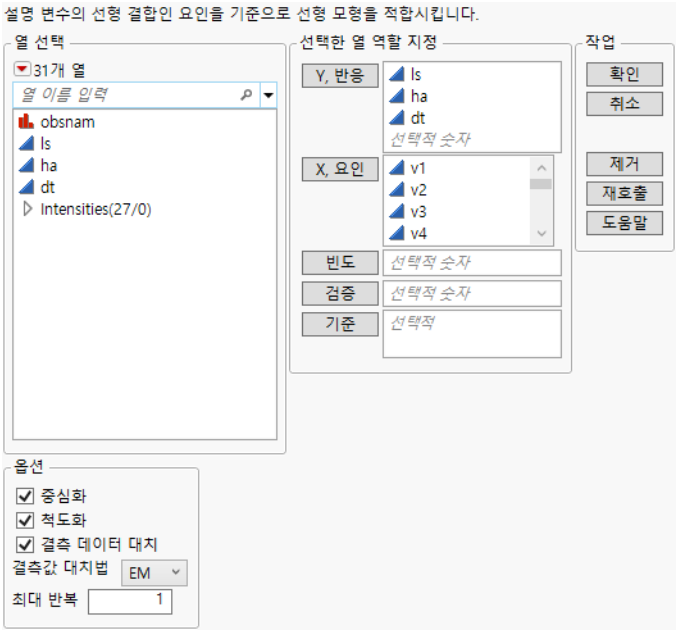
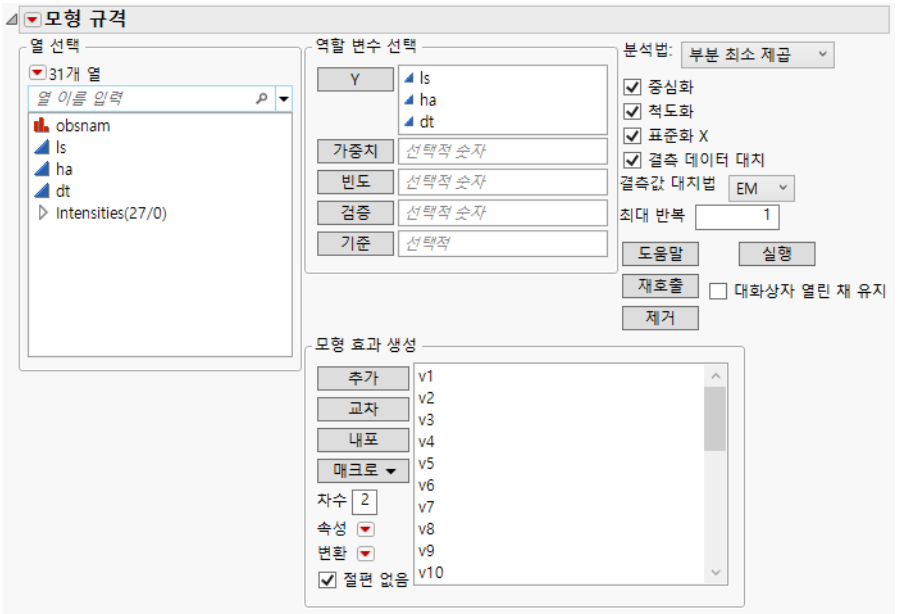


그림 6.7 JMP Pro 모형 적합 부분 최소 제곱 시작 창



모형 적합 시작 옵션에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

부분 최소 제곱 시작 창에는 다음 옵션이 포함되어 있습니다.

Y, 반응 수치형 반응 열을 입력합니다. 열을 여러 개 입력하면 공동으로 모델링됩니다.

JMP PRO JMP Pro 에서는 모형 적합 시작 창에 명목형 반응 열을 입력하여 PLS-DA 를 수행할 수 있습니다. 자세한 내용은 "PLS-DA(PLS 판별 분석)" 에서 확인하십시오.

X, 요인 예측 변수 열을 입력합니다. 부분 최소 제곱 시작 창에는 수치형 예측 변수만 허용됩니다.

JMP PRO JMP Pro 에서는 모형 적합 시작 창에 명목형 및 순서형 모형 효과를 입력할 수 있습니다. 순서형 효과는 명목형으로 처리됩니다.

빈도 데이터가 요약된 경우 값에 각 행에 대한 개수가 포함된 열을 입력합니다.

JMP PRO **검증** 검증 데이터 집합을 정의하는 숫자 열입니다 . 검증 열에는 다음과 같이 연속된 정수 값만 포함되어야 합니다.

- 검증 열에 두 개의 수준이 포함된 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다.
- 검증 열에 세 개의 수준이 포함된 경우 작은 값부터 순서대로 훈련 데이터 집합, 검증 데이터 집합, 테스트 데이터 집합을 각각 정의합니다.
- 검증 열에 네 개 이상의 수준이 포함된 경우 K 폴드 교차 검증이 사용됩니다 . 기타 검증 옵션에 대한 자세한 내용은 " 검증 방법 " 에서 확인하십시오.

PLS 플랫폼에서는 검증 열을 사용하여 모형을 훈련 및 조정하거나 모형을 훈련, 조정 및 평가합니다. 검증에 대한 자세한 내용은 예측 및 전문 모델링의 에서 확인하십시오.

참고 : " 열 선택 " 목록에서 아무 열도 선택하지 않은 상태로 " 검증 " 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다 . " 검증 열 생성 " 유틸리티에 대한 자세한 내용은 예측 및 전문 모델링의 에서 확인하십시오.

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 입력합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다.

중심화 각 열에서 평균을 빼서 모든 Y 변수와 모형 효과를 중심화합니다. 자세한 내용은 " 중심화 및 척도화 " 에서 확인하십시오.

척도화 각 열을 표준편차로 나누어 모든 Y 변수와 모형 효과의 척도를 조정합니다. 자세한 내용은 " 중심화 및 척도화 " 에서 확인하십시오.

JMP PRO **X 표준화** (모형 적합 시작 창에서만 사용 가능) 모형 효과를 구성하는 데 사용되는 모든 열을 중앙에 맞추고 척도를 조정합니다. 이 옵션을 선택하지 않으면 원래 데이터 테이블 열을 사용하여 고차 효과가 구성됩니다. 그런 다음 선택한 " 중심화 " 및 " 척도화 " 옵션에 따라 각 고차 효과가 중심화되거나 척도화됩니다. "X 표준화 " 는 Y 변수를 중앙에 맞추거나 척도를 조정하지 않습니다. 자세한 내용은 "X 표준화 " 에서 확인하십시오.

JMP PRO **결측 데이터 대치** Y 또는 X 의 결측 데이터 값을 비결측 값으로 대체합니다 . **결측값 대치법** 목록에서 적절한 방법을 선택합니다 .

결측 데이터 대치를 선택하지 않으면 X 변수의 관측값이 결측인 행이 분석에서 제외되고 해당 행에 대한 예측이 계산되지 않습니다. X 변수에는 결측 관측값이 없지만 Y 변수에 결측 관측값이 있는 행도 분석에서 제외되지만 예측이 계산됩니다 .

JMP PRO **결측값 대치법** (**결측 데이터 대치**를 선택한 경우에만 표시됨) 다음 결측값 대치 방법 중에서 선택합니다 .

평균 각 모형 효과 또는 반응 열에 대해 결측값을 비결측 값의 평균으로 대체합니다 .

EM 반복 EM(Expectation-Maximization) 방식을 사용하여 결측값을 대치합니다 . 첫 번째 반복에서는 지정된 모형이 효과 또는 반응의 결측값을 평균으로 대체한 데이터에 적합됩니다 . Y 모형의 예측값과 X 모형의 예측값을 사용하여 결측값을 대치합니다 . 이후 반복의 경우 현재 추정값을 사용한 조건부 분포가 주어지면 결측값이 해당 예측값으로 대체됩니다 .

결측값 대치를 위해 다항식 항은 별도의 예측 변수로 처리됩니다 . 다항식 항이 지정되면 해당 항은 원래 데이터에서 계산되거나 , "X 표준화 " 가 선택된 경우 표준화된 열 값에서 계산됩니다 . 다항식 항 정의와 관련된 열의 값이 결측인 행이 있으면 해당 항목은 다항식 항에 대해 결측입니다 . 이 방식으로 정의된 다항식 항을 사용하여 결측값 대치가 수행됩니다 .

EM 방법에 대한 자세한 내용은 Nelson, Taylor 및 MacGregor(1996) 연구 자료에서 확인하십시오 .

JMP PRO **최대 반복** (" 결측값 대치법 " 에서 "EM" 을 선택한 경우에만 표시됨) 알고리즘에 사용되는 최대 반복 수를 설정할 수 있습니다 . 결측값의 현재 추정값과 이전 추정값 사이의 최대 차이가 10^{-8} 로 제한되면 알고리즘이 종료됩니다 .

시작 창을 완료하고 **확인**을 클릭하면 "모형 시작" 제어판이 나타납니다. 자세한 내용은 "모형 시작 제어판"에서 확인하십시오.

중심화 및 척도화

"중심화" 및 "척도화" 옵션은 기본적으로 선택되어 있습니다. 즉, 예측 변수와 반응이 평균 0, 표준편차 1이 되도록 중심화되고 척도가 조정됩니다. 예측 변수와 반응을 중심화하면 변동에 비례하여 해당 변수와 반응이 동등하게 됩니다. 중심화하지 않으면 변수의 평균과 해당 평균에 대한 변동이 모두 연속 요인을 생성하는 데 포함됩니다. 설명을 위해 Time과 Temp가 두 개의 예측 변수라고 가정해 보겠습니다. 이 변수를 척도화하면 Time의 1 표준편차 변화가 Temp의 1 표준편차 변화와 거의 동일하다는 것을 나타냅니다.

X 표준화

JMP PRO "모형 적합" 창에서 "부분 최소 제곱" 분석법을 선택하면 "X 표준화" 옵션이 기본적으로 선택됩니다. 이렇게 하면 모형 효과로 입력한 모든 열과 교호작용 또는 다항식 항에 포함된 모든 열이 표준화됩니다.

X1 과 X2 라는 두 개의 열이 있고 " 모형 적합 " 창에서 교호작용 항 X1*X2 를 모형 효과로 입력한다고 가정해 보겠습니다 . "X 표준화 " 옵션을 선택하면 교호작용 항을 구성하기 전에 X1 과 X2 가 모두 중심화 및 척도화됩니다 . 구성되는 교호작용 항은 다음과 같이 계산됩니다 .

$$\left(\frac{X1 - mean(X1)}{std(X1)}\right) \times \left(\frac{X2 - mean(X2)}{std(X2)}\right)$$

그런 다음 모든 모형 효과가 모형에 포함되기 전에 **중심화** 및 **척도화** 옵션 선택에 따라 중심화되거나 척도화됩니다 .

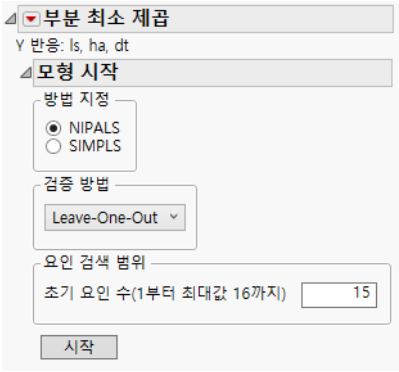
"X 표준화 " 옵션을 선택하지 않고 **중심화** 및 **척도화** 옵션을 모두 선택하면 모형에 입력되는 항이 다음과 같이 계산됩니다 .

$$\frac{X1 \times X2 - mean(X1 \times X2)}{std(X1 \times X2)}$$

모형 시작 제어판

플랫폼 시작 창에서 **확인**을 클릭하거나 " 모형 적합 " 창에서 **실행**을 클릭하면 " 모형 시작 " 제어판이 나타납니다 .

그림 6.8 부분 최소 제곱 모형 시작 제어판



참고 : JMP Pro 에서는 " 모형 시작 " 제어판의 " 검증 방법 " 부분이 다르게 나타납니다 .

" 모형 시작 " 제어판에는 다음과 같은 선택 옵션이 포함되어 있습니다 .

방법 지정 모형 적합 알고리즘 유형을 선택합니다 . **NIPALS** 와 **SIMPLS** 라는 두 가지 알고리즘이 있습니다 . 반응 변수가 하나뿐인 경우 두 방법은 동일한 계수 추정값을 생성합니다 . 두 알고리즘의 차이에 대한 자세한 내용은 " 부분 최소 제곱 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오 .

검증 방법 검증 방법을 선택합니다 . 검증은 추출할 최적 요인 수를 결정하는 데 사용됩니다 . JMP Pro 의 경우 플랫폼 시작 창에서 검증 열을 지정하면 이러한 옵션이 나타나지 않습니다 .

JMP PRO **홀드백** 검증 데이터 집합에 대해 지정된 비율의 데이터를 무작위로 선택하고 데이터의 다른 부분을 사용하여 모형을 적합시킵니다 . 랜덤 선택은 단순 랜덤 표집을 기반으로 할 때보다 더 균형 있는 훈련 데이터 집합 및 검증 데이터 집합을 생성하기 위해 모형 요인에 대한 층화 표집을 기반으로 합니다 .

JMP PRO **K 폴드** 데이터를 *K* 개의 부분집합 또는 **폴드**로 분할합니다 . 그런 다음 각 폴드는 나머지 데이터에 적합된 모형을 검증하는 데 사용되어 총 *K* 개의 모형을 적합시킵니다 . 이 방법은 제한된 양의 데이터를 효율적으로 사용하기 때문에 소규모 데이터 집합에 가장 적합합니다 .

Leave-One-Out leave-one-out 교차 검증을 수행합니다 .

없음 추출할 요인 수를 선택하기 위해 검증을 사용하지 않습니다 . 요인 수는 " 요인 검색 범위 " 에 지정됩니다 .

요인 검색 범위 검증을 사용하지 않는 경우 추출할 잠재 요인 수를 지정합니다 . 검증을 사용 중인 경우에는 최적 요인 수를 선택하기 전에 플랫폼에서 적합시키려고 하는 최대 요인 수입니다 . 최대 요인 수는 비결측 행 수와 요인 수 중 더 작은 값입니다 . 이 경우 초기 요인 수는 15 와 최대 요인 수 중 더 작은 값입니다 .

요인 지정 **시작**을 클릭하여 초기 모형을 적합시키면 나타납니다 . 새 모형을 적합시키는 데 사용할 요인 수를 지정합니다 .

시작 지정된 규격으로 모형 적합을 시작합니다 .

부분 최소 제곱 옵션

" 부분 최소 제곱 " 의 빨간색 삼각형 메뉴에 포함된 일부 옵션은 " 모형 시작 " 제어판에서 " 시작 " 을 클릭하기 전에 사용하면 유용합니다 .

JMP PRO **난수 시드값 설정** K 폴드 및 홀드백 검증에 사용되는 난수화 과정을 위한 시드값을 설정합니다 . 분석을 재현하려는 경우에 유용합니다 . 시드를 양수 값으로 설정하고 스크립트를 저장하면 시드값이 스크립트에 자동으로 저장됩니다 . 스크립트를 실행하면 항상 동일한 교차검증 분석이 생성됩니다 . " 검증 방법 " 을 " 없음 " 으로 설정하거나 검증 열을 사용하는 경우에는 이 옵션이 나타나지 않습니다 .

JMP PRO **SVD** 모형 적합 알고리즘에서 SVD 알고리즘의 빠른 또는 전통적 방법 적용을 선택할 수 있는 하위 메뉴를 표시합니다 . 기본은 빠른 SVD 입니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

부분 최소 제곱 보고서

그림 6.8 의 "모형 시작" 제어판에서 **시작**을 처음 클릭하면 "검증 방법" 패널이 "모형 시작" 창에서 제거됩니다. 검증 열을 지정했거나 "검증 방법" 패널에서 "홀드백"을 선택한 경우 보고서의 모든 모형 적합은 훈련 데이터를 기반으로 합니다. 그렇지 않은 경우 모든 모형 적합은 전체 데이터 집합을 기반으로 합니다.

검증을 사용한 경우 다음 세 가지 보고서가 나타납니다 .

- 모형 비교 요약
- 교차 검증 보고서
- <N> 개의 요인을 사용한 NIPALS(또는 SIMPLS) 적합

CV 방법을 **없음**으로 선택한 경우 다음 두 보고서가 나타납니다 .

- 모형 비교 요약
- <N> 개의 요인을 사용한 NIPALS(또는 SIMPLS) 적합

추가 모형을 적합시키려면 "모형 시작" 패널에서 원하는 요인 수를 지정합니다.

모형 비교 요약

"모형 비교 요약"에는 각 적합 모형에 대한 요약 결과가 표시됩니다.

그림 6.9 모형 비교 요약

모형 비교 요약

방법	SVD	행 수	요인 수	누적 X에 대해 설명되는 변동 백분율	누적 Y에 대해 설명되는 변동 백분율	VIP 수 > 0.8
NIPALS	빠름	16	7	99.995152	98.937438	22
NIPALS	빠름	16	6	99.993471	97.768092	22

이 보고서에는 다음과 같은 요약 정보가 포함됩니다 .

방법 " 모형 시작 " 제어판에서 지정한 분석 방법을 표시합니다 .

SVD 지정한 SVD 알고리즘 구현 방법을 표시합니다 .

행 수 훈련 데이터 집합에 사용된 관측값 수를 표시합니다 .

요인 수 추출된 요인 수를 표시합니다 .

누적 X 에 대해 설명되는 변동 백분율 모형에 의해 설명되는 X 의 변동 백분율을 표시합니다 .

누적 Y 에 대해 설명되는 변동 백분율 모형에 의해 설명되는 Y 의 변동 백분율을 표시합니다 .

VIP 수 > 0.8 VIP(Variable Importance for Projection) 값이 0.8 보다 큰 모형 효과의 수를 표시합니다 . VIP 스코어는 X 및 Y 모델링에 대한 변수의 상대적 중요도를 나타내는 측도입니다 (Wold 1994, Eriksson et al. 2006 참조).

교차 검증 보고서

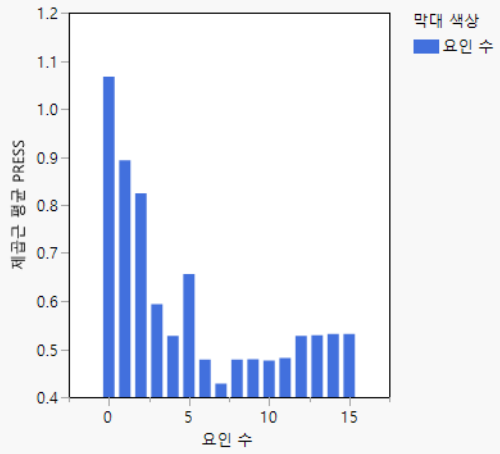
이 보고서는 "모형 시작" 제어판에서 교차 검증 형식을 검증 방법으로 선택한 경우에만 나타납니다. 보고서 제목은 제어판에서 선택한 교차 검증 및 방법 옵션에 따라 동적으로 "<방법 = 방법 지정>"을 사용한 <교차 검증 방법>"으로 지정됩니다. 여기에는 "모형 시작" 제어판에 지정한 대로 0 ~ 추출된 최대 요인 수를 사용하여 모형 적합에 대한 요약 통계량이 표시됩니다. 또한 제곱근 평균 PRESS 값을 보여 주는 그림도 제공됩니다. 자세한 내용은 "제곱근 평균 PRESS 그림"에서 확인하십시오. 최적 요인 수는 최소 제곱근 평균 PRESS 통계량을 사용하여 식별됩니다.

그림 6.10 교차 검증 보고서

방법=NIPALS을(를) 사용한 Leave-One-Out 교차 검증(빠름 SVD 사용)

요인 수	제곱근 평균 PRESS	van der Voet T²	Prob > van der Voet T²	Q²	누적 Q²	R²X	누적 R²X	R²Y	누적 R²Y
0	1.066667	10.070473	0.0010*	-0.137778	-0.137778	0.000000	0.000000	0.000000	0.000000
1	0.892851	11.141942	<.0001*	-0.098077	-0.098077	0.974480	0.974480	0.421421	0.421421
2	0.824551	8.605908	<.0001*	0.773385	0.751159	0.022034	0.996514	0.247724	0.669146
3	0.593922	8.034522	0.0080*	0.774570	0.943904	0.001772	0.998286	0.241683	0.910829
4	0.527742	7.879423	0.0080*	0.656229	0.980716	0.001153	0.999438	0.037043	0.947872
5	0.656013	6.883024	0.0180*	0.356257	0.987586	0.000386	0.999825	0.012502	0.960374
6	0.478479	3.446269	0.2890	0.640049	0.995532	0.000113	0.999938	0.019350	0.979724
7	0.428141	0.000000	1.0000	0.809320	0.999148	0.000017	0.999955	0.010643	0.990367
8	0.478562	7.031041	0.0100*	0.818516	0.999845	0.000011	0.999967	0.004666	0.995032
9	0.478739	3.066627	0.3990	0.810850	0.999971	0.000012	0.999979	0.001452	0.996485
10	0.476024	3.847943	0.2420	0.845631	0.999995	0.000007	0.999986	0.001417	0.997902
11	0.481440	4.822874	0.1100	0.820847	0.999999	0.000004	0.999990	0.001248	0.999150
12	0.527436	3.968884	0.2040	0.810490	1.000000	0.000004	0.999994	0.000637	0.999787
13	0.529249	4.028960	0.2070	0.812180	1.000000	0.000003	0.999997	0.000158	0.999945
14	0.531553	4.408278	0.1670	0.816103	1.000000	0.000003	1.000000	0.000055	1.000000
15	0.531553	4.408278	0.1800	0.816103	1.000000	0.000000	1.000000	0.000000	1.000000

제곱근 평균 PRESS 그림



참고: 최소 제곱근 평균 PRESS는 0.42814이고 최소 요인 수는 7입니다.

JMP PRO **X 표준화** 옵션을 선택하면 표준화가 전체 데이터 테이블에 한 번 적용됩니다. 개별 훈련 데이터 집합에 다시 적용되지는 않습니다. 그러나 **중심화** 또는 **척도화** 옵션 조합을 선택하면 이 선택 조합이 각 교차 검증 훈련 데이터 집합에 적용됩니다. 이러한 옵션을 선택한 경우 개별적으로 중심화 및 척도화되는 훈련 데이터 집합을 사용하여 교차 검증이 진행됩니다.

다음 통계량이 보고서에 표시됩니다. 검증 또는 교차 검증이 사용되는 경우 보고된 결과는 훈련 데이터 집합 통계량의 요약입니다.

요인 수 모형 적합에 사용된 요인 수입니다.

제곱근 평균 PRESS 모든 반응에 대한 PRESS 값 평균의 제곱근입니다. 자세한 내용은 " 제곱근 평균 PRESS" 에서 확인하십시오.

van der Voet T^2 추출된 요인 수가 다른 모형이 최적 모형과 유의하게 다른지 여부를 검정하는 van der Voet 검정의 검정 통계량입니다 . 각 van der Voet T^2 검정에 대한 귀무가설은 해당

요인 수를 기반으로 하는 모형이 최적 모형과 다르지 않다는 것입니다 . 대립가설은 모형이 최적 모형과 다르다는 것입니다 . 자세한 내용은 "van der Voet T^2" 에서 확인하십시오 .

Prob > van der Voet T^2 van der Voet T^2 검정의 p 값입니다 . 자세한 내용은 "van der Voet T^2" 에서 확인하십시오 .

Q^2 다음과 같이 Y 의 총 제곱합으로 나눈 PRESS 값 비율을 1 에서 뺀 값으로 정의되는 무차원 예측 능력 측도입니다 .

1 – PRESS/SSY

자세한 내용은 "Q^2 계산 " 에서 확인하십시오 .

누적 Q^2 주어진 수 이하의 요인을 사용한 모형의 예측 능력을 나타내는 지표입니다 . 주어진 요인 수 f 에 대해 누적 Q^2 은 다음과 같이 정의됩니다 .

$$1 - \prod_{i=1}^{f} (\text{Press}_i/\text{SYY}_i)$$

여기서 $PRESS_i$ 와 SSY_i 는 i 개 요인에 대한 값에 해당합니다 .

R^2X 지정된 요인에 의해 설명되는 X 변동의 백분율입니다 . R^2X 가 큰 성분이 X 변수의 변동을 많이 설명합니다. 자세한 내용은 "검증이 사용되는 경우 R^2X 및 R^2Y 계산"에서 확인하십시오.

누적 R^2X 주어진 요인 수를 사용한 모형에 의해 설명되는 X 변동의 백분율입니다 . 이 값은 R^2X 값의 합계입니다 (i = 1 ~ 주어진 요인 수).

R^2Y 지정된 요인에 의해 설명되는 Y 변동의 백분율입니다 . R^2Y 가 큰 성분이 Y 변수의 변동을 많이 설명합니다. 자세한 내용은 "검증이 사용되는 경우 R^2X 및 R^2Y 계산"에서 확인하십시오.

누적 R^2Y 주어진 요인 수를 사용한 모형에 의해 설명되는 Y 변동의 백분율입니다 . 이 값은 R^2Y 값의 합계입니다 (i = 1 ~ 주어진 요인 수).

Q^2 및 누적 R^2Y 해석

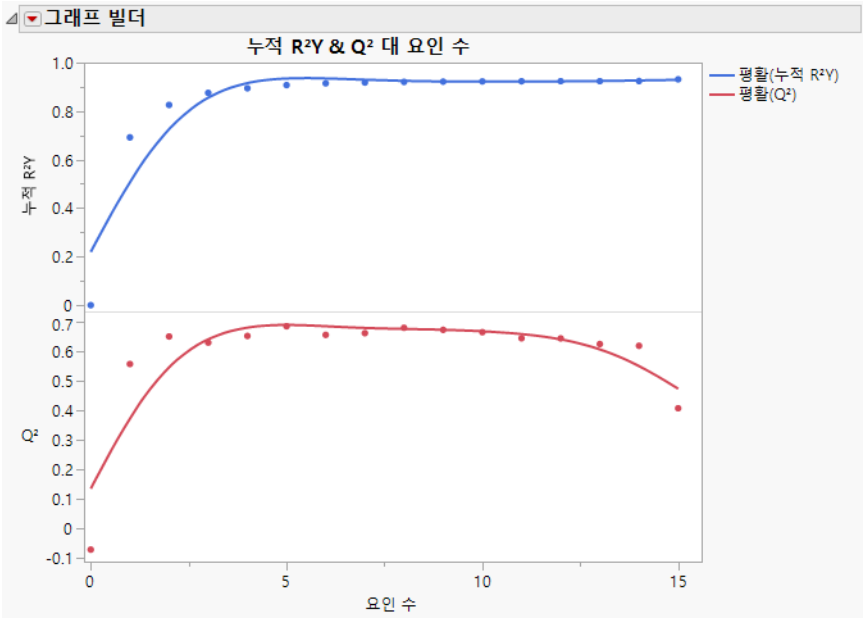
Q^2 및 누적 R^2Y 통계량은 둘 다 예측 능력을 측정하지만 방식이 다릅니다 .

- 누적 R^2Y 는 요인 수가 증가함에 따라 함께 증가합니다 . 모형에 요인이 추가될수록 더 많은 변동이 설명되기 때문입니다 .
- Q^2 은 요인 수가 증가함에 따라 증가했다가 감소하거나 , 적어도 증가를 중단하는 경향이 있습니다 . 요인이 추가될수록 모형이 훈련 데이터 집합에 맞게 조정되고 새 데이터에 대해 잘 일반화되지 않아 PRESS 통계량이 감소하기 때문입니다 .

Q^2 및 누적 R^2Y 분석은 모형에 포함할 요인 수를 결정하기 위해 van der Voet 검정 대신 사용할 수 있습니다. Q^2이 크고 감소가 시작되지 않은 요인 수를 선택합니다. 누적 R^2Y도 커야 합니다.

그림 6.11에서는 Leave-One-Out 검증 방법을 사용하는 Penta.jmp 데이터 테이블의 요인 수에 대한 누적 R^2Y 및 Q^2 그림을 보여 줍니다. 누적 R^2Y는 약 네 개 요인에 대해 증가 후 안정됩니다. Q^2 통계량은 두 요인에 대해 가장 크고 그 이후에 안정되기 시작합니다. 이 그림은 요인이 두 개인 모형이 데이터를 과대적합하지 않고 Y 변동의 많은 부분을 설명한다는 것을 나타냅니다.

그림 6.11 Penta.jmp 에 대한 누적 R^2Y 및 Q^2



제곱근 평균 PRESS 그림

이 막대 차트에는 요인 수가 가로 축에 표시되고 제곱근 평균 PRESS 값이 세로 축에 표시됩니다. 이는 교차 검증 보고서의 "제곱근 평균 PRESS" 열 오른쪽에 나타나는 가로 막대 차트와 동일합니다(그림 6.10 참조).

제곱근 평균 PRESS

지정된 요인 수 *a* 에 대한 제곱근 평균 PRESS 는 다음 단계를 사용하여 계산됩니다 .

1. 요인이 *a* 개인 모형을 각 훈련 데이터 집합에 적합시킵니다 .
2. 결과 예측 계산식을 검증 데이터 집합의 관측값에 적용합니다 .
3. 각 Y 에 대해 다음을 수행합니다 .
   - 각 검증 데이터 집합에 대해 관측된 각 검증 데이터 집합 값과 예측값 간의 차이 제곱을 계산합니다 (제곱 예측 오차).
   - 각 검증 데이터 집합에 대해 이러한 차이 제곱의 평균을 구하고 결과를 반응에 대한 분산 추정값으로 나눕니다 . K 폴드 및 Leave-One-Out 검증 방법의 경우 전체 반응 열의 분산으로 나눕니다 . 홀드백 검증의 경우 훈련 데이터 집합에 있는 반응 값의 분산으로 나눕니다 .
   - 이러한 평균을 합산하고 검증 데이터 집합이 두 개 이상인 경우 합계를 검증 데이터 집합 수 - 1 로 나눕니다 . 이 값이 주어진 Y 에 대한 PRESS 통계량입니다 .
4. *a* 개 요인의 제곱근 평균 PRESS 는 모든 반응에 대한 PRESS 값 평균의 제곱근입니다 .

5. 여러 Y 의 PRESS 통계량은 모든 반응에 대해 3 단계에서 구한 PRESS 통계량의 평균을 계산한 값입니다 .

Q^2 계산

Q^2 통계량은 1 - *PRESS*/*SSY* 로 정의됩니다. *PRESS* 통계량은 훈련 데이터 집합을 기반으로 개발되었지만 검증 데이터 집합에서 평가된 모형의 모든 반응에 대해 평균을 구한 예측 오차 제곱합입니다. *SSY* 값은 모든 반응에 대해 평균을 구하고 검증 데이터 집합의 관측값을 기반으로 하는 Y 의 제곱합입니다.

교차 검증 보고서의 Q^2 통계량은 선택한 검증 방법에 따라 다음과 같은 방식으로 계산됩니다 .

Leave-One-Out Q^2 은 1 - 평균 (*PRESS*)/ 평균 (*SSY*) 값입니다 . 각 *Y* 에 대한 *PRESS* 는 관측값을 한 번에 하나씩 제외하여 생성된 모형에서 계산됩니다 . *SSY* 는 각 *Y* 열의 모든 값에서 계산됩니다 .

K 폴드 Q^2 은 각 *K* 폴드를 제외하여 생성된 *K* 개 모형을 기반으로 한 검증 데이터 집합에 대해 계산되는 1 - *PRESS*/*SSY* 값의 평균입니다 .

홀드백 또는 검증 데이터 집합 Q^2 은 단일 훈련 데이터 집합을 사용하여 생성된 모형을 기반으로 한 검증 데이터 집합에 대해 계산되는 1 - *PRESS*/*SSY* 값입니다 .

검증이 사용되는 경우 R^2X 및 R^2Y 계산

교차 검증 보고서의 R^2X 및 R^2Y 통계량은 선택한 검증 방법에 따라 다음과 같은 방식으로 계산됩니다.

참고 : 이러한 모든 계산에서 R^2Y 는 유사하게 계산됩니다 .

Leave-One-Out R^2X 는 관측값을 한 번에 하나씩 제외하여 생성된 모형에 대한 "X 효과에 대해 설명되는 변동 백분율 " 값의 평균입니다 .

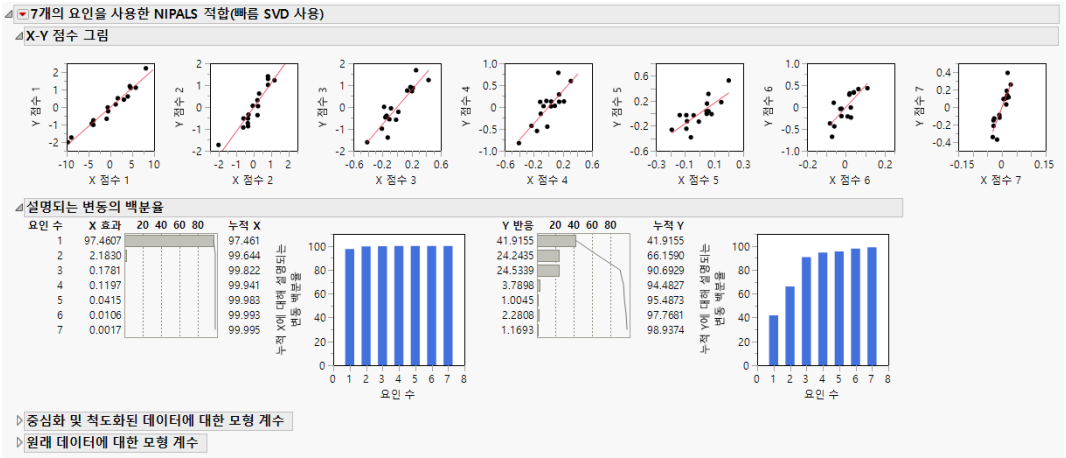
K 폴드 R^2X 는 각 폴드를 제외하여 생성된 K 개 모형에 대한 "X 효과에 대해 설명되는 변동 백분율 " 값의 평균입니다 .

홀드백 또는 검증 데이터 집합 R^2X 는 훈련 데이터를 사용하여 생성된 모형에 대한 "X 효과에 대해 설명되는 변동 백분율 " 값입니다 .

모형 적합 보고서

모형 적합 보고서에는 각 적합 모형에 대한 상세 결과가 표시됩니다. 적합에는 교차 검증에 기반한 최적 요인 수가 사용되거나, 교차 검증 방법이 지정되지 않은 경우 지정된 요인 수가 사용됩니다. 보고서 제목은 NIPALS 또는 SIMPLS 중 사용된 방법과 빠른 SVD 또는 전통적 SVD 중 사용된 방법을 나타내고 추출된 요인 수를 제공합니다.

그림 6.12 모형 적합 보고서



모형 적합 보고서에는 다음과 같은 요약 정보가 포함됩니다 .

X-Y 스코어 그림 추출된 각 요인에 대한 X 및 Y 스코어의 산점도입니다 .

설명되는 변동의 백분율 X 와 Y 둘 다에 대해 설명되는 변동 백분율 및 누적 변동 백분율을 표시합니다 . 추출된 각 요인에 대한 결과가 제공됩니다 .

중심화 및 척도화된 데이터에 대한 모형 계수 각 Y 를 대상으로 중심화 및 척도화된 데이터를 기반으로 하는 모형에 대한 X 의 계수를 표시합니다 .

모형 적합 옵션

모형 적합의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

변동 백분율 그림 "X 효과에 대해 설명되는 변동 백분율 " 과 "Y 효과에 대해 설명되는 변동 백분율 " 이라는 두 개의 그림을 추가합니다 . 여기에는 *X* 및 *Y* 에 대해 추출된 각 요인에 의해 설명되는 변동 백분율을 나타내는 누적 막대 차트가 표시됩니다 .

변수 중요도 그림 각 X 변수의 VIP 값을 그림에 표시합니다 . VIP 스코어가 " 변수 중요도 테이블 " 에 나타납니다 . 자세한 내용은 " 변수 중요도 그림 " 에서 확인하십시오 .

VIP 대 계수 그림 모형 계수에 대한 VIP 통계량을 그림에 표시합니다 . 선택한 *Y* 에 해당하는 점만 표시할 수 있습니다 . 추가 라벨 지정 옵션이 제공됩니다 . 중심화 및 척도화된 데이터에 대한 그림과 원래 데이터에 대한 그림이 둘 다 있습니다 . 자세한 내용은 "VIP 대 계수 그림 " 에서 확인하십시오 .

VIP 임계 설정 변수 중요도 그림 , 변수 중요도 테이블 및 VIP 대 계수 그림에 사용할 임계 수준을 설정합니다 .

계수 그림 전체 X 변수에서 각 반응에 대한 모형 계수를 그림에 표시합니다. 선택한 *Y* 에 해당하는 점만 표시할 수 있습니다. 중심화 및 척도화된 데이터에 대한 그림과 원래 데이터에 대한 그림이 둘 다 있습니다.

적재 그림 추출된 각 요인에 대한 X 및 Y 적재를 그림에 표시합니다. *X* 와 *Y* 에 대해 각각 별도의 그림이 있습니다.

적재 산점도 행렬 X 적재 및 Y 적재의 산점도 행렬을 표시합니다.

상관 적재 그림 동일한 그림에 중첩된 X 및 Y 적재의 단일 산점도 또는 산점도 행렬을 표시합니다. 이 옵션을 선택할 때 그림에 표시할 요인 수를 지정합니다.

- 두 개의 요인을 지정하면 단일 상관 적재 산점도가 나타납니다. 그림 아래에서 축을 정의하는 두 요인을 선택합니다. 오른쪽 화살표 버튼을 클릭하면 각 요인 조합이 연속해서 그림에 표시됩니다.
- 세 개 이상의 요인을 지정하면 선택한 개수까지의 요인 쌍에 대한 셀이 포함된 산점도 행렬이 나타납니다.

두 경우 모두 체크박스를 사용하여 라벨 지정을 제어합니다.

X-Y 스코어 그림 다음 옵션을 포함합니다.

선형 적합 "X-Y 스코어 그림 " 에 점을 통과하는 적합선을 표시하거나 숨깁니다.

신뢰 대역 표시 "X-Y 스코어 그림 " 에 적합선에 대한 95% 신뢰 대역을 표시하거나 숨깁니다.

스코어 산점도 행렬 X 스코어의 산점도 행렬과 Y 스코어의 산점도 행렬을 표시합니다. 각 X 스코어 산점도에는 이상치 감지에 사용할 수 있는 95% 신뢰도 타원이 표시됩니다 . 신뢰도 타원에 대한 통계 상세 정보는 "X 스코어 산점도 행렬의 신뢰도 타원 " 에서 확인하십시오.

거리 그림 다음을 나타내는 그림을 표시합니다.

- 각 관측값에서 X 모형까지의 거리
- 각 관측값에서 Y 모형까지의 거리
- X 모형까지의 거리와 Y 모형까지의 거리를 둘 다 보여 주는 산점도

적합한 모형에서는 X 거리와 Y 거리가 모두 작으므로 점이 원점 (0,0) 에 가깝습니다. 그림을 사용하여 X 또는 Y 에 대한 이상치를 찾을 수 있습니다. 점 그룹이 함께 군집화되면 공통 특징을 가질 수 있으며 개별적으로 분석할 수 있습니다. 검증 데이터 집합 또는 검증 및 테스트 데이터 집합이 사용 중인 경우 이러한 데이터 집합과 훈련 데이터 집합에 대해 별도의 보고서가 제공됩니다.

T 제곱 그림 각 관측값에 대한 T^2 통계량 그림과 관리 한계를 함께 표시합니다. 관측값의 T^2 통계량은 추출된 요인에 대한 해당 관측값의 스코어를 기반으로 계산됩니다. 관리 한계 및 T^2 계산에 대한 자세한 내용은 "T^2 그림 " 에서 확인하십시오.

진단 그림 모형 적합을 평가하기 위한 진단 그림을 표시합니다. " 실제값 대 예측값 그림 ," 잔차 대 예측값 그림 ," 잔차 대 행 번호 그림 " 및 " 잔차 정규 분위수 그림 " 이라는 네 가지 유형의 그림이 있습니다. 각 반응에 대한 그림이 제공됩니다. 검증 데이터 집합 또는 검증 및 테스트 데이터 집합이 사용 중인 경우 이러한 데이터 집합과 훈련 데이터 집합에 대해 별도의 보고서가 제공됩니다.

프로파일러 각 Y 변수에 대한 프로파일러를 표시합니다.

스펙트럼 프로파일러 모든 반응 변수가 그림의 첫 번째 셀에 나타나는 단일 프로파일러를 표시합니다. 이 프로파일러는 X 변수의 변화가 Y 변수에 미치는 영향을 동시에 시각화하는 데 유용합니다.

열 저장 다양한 계산식 및 결과를 저장하기 위한 옵션이 포함되어 있습니다. 자세한 내용은 " 열 저장 " 에서 확인하십시오.

적합 제거 주 플랫폼 보고서에서 모형 보고서를 제거합니다.

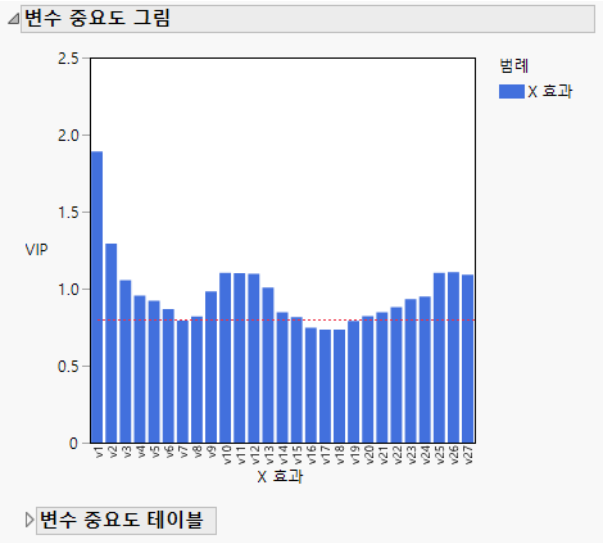
VIP 를 사용하여 모형 생성 시작 창을 연 후, 적절한 반응을 *Y* 로 입력하고 VIP 값이 특정 임계를 초과하는 변수를 *X* 로 입력하여 창을 채웁니다. " 중심화 및 척도화된 데이터에 대한 VIP 대 계수 " 보고서의 버튼과 동일한 기능을 수행합니다. 자세한 내용은 "VIP 대 계수 그림 " 에서 확인하십시오.

저장된 X 스코어에 대한 모형 기반 다변량 관리도 각 X 스코어에 대한 계산식을 저장하고 MDMCC(모형 기반 다변량 관리도) 시작 창을 실행합니다. MDMCC 시작 창에서 스코어 계산식이 공정 열로 할당됩니다. 공정을 추가 또는 제거하고, 시간 ID 를 추가하거나, 과거 데이터 끝 위치를 설정한 후 " 확인 " 을 클릭할 수 있습니다. 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오.

변수 중요도 그림

"변수 중요도 그림"에는 각 X 변수에 대한 VIP 값이 그래프로 표시됩니다. "변수 중요도 테이블"에는 VIP 스코어가 표시됩니다. VIP 스코어는 X와 Y를 모델링할 때 변수의 중요도를 나타내는 측도입니다. 변수의 계수 및 VIP가 작으면 해당 변수는 모형에서 삭제할 후보가 됩니다(Wold 1994 참조). 값이 0.8이면 일반적으로 VIP가 작은 것으로 간주되며(Eriksson et al. 2006 참조) 그림의 0.8 위치에 빨간색 파선이 표시됩니다.

그림 6.13 변수 중요도 그림



VIP 대 계수 그림

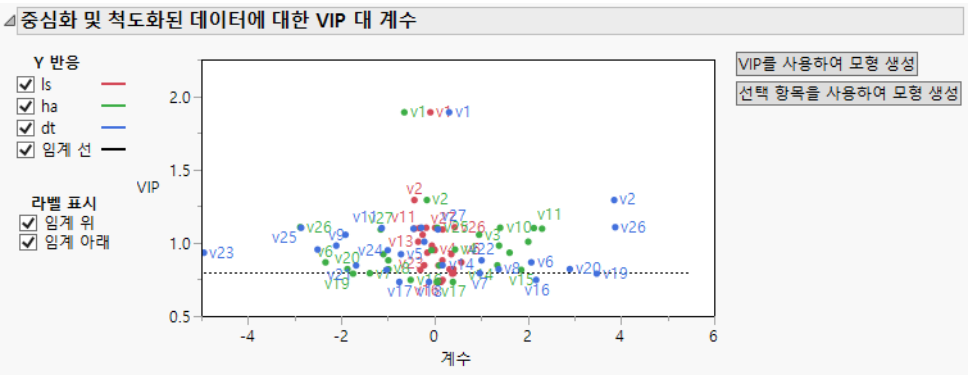
그림 오른쪽에 있는 다음 두 옵션을 사용하면 변수 축소 및 모형 생성을 쉽게 할 수 있습니다.

VIP 를 사용하여 모형 생성 시작 창을 연 후, 적절한 반응을 *Y* 로 입력하고 VIP 값이 특정 임계를 초과하는 변수를 *X* 로 입력하여 창을 채웁니다.

선택 항목을 사용하여 모형 생성 그림에서 직접 *X* 를 선택할 수 있으며 그러면 *Y* 와 선택한 *X* 만 시작 창에 입력됩니다.

현재 열 선택을 기반으로 다른 플랫폼을 사용하려면 원하는 플랫폼을 엽니다. 선택 항목이 시작 창에 그대로 유지됩니다. 역할 버튼을 클릭하면 선택한 열이 채워집니다.

그림 6.14 중심화 및 척도화된 데이터에 대한 VIP 대 계수 그림



열 저장

예측 계산식 저장 각 Y 변수에 대해 예측 계산식이 포함된 **예측 계산식 < 반응 >** 열을 데이터 테이블에 저장합니다 .

예측을 X 스코어 계산식으로 저장 각 Y 변수에 대해 X 스코어에 대한 예측 계산식이 포함된 **예측 계산식 < 반응 >** 열을 데이터 테이블에 저장합니다 .

예측 계산식 표준 오차 저장 각 Y 변수에 대해 예측 평균의 표준 오차가 포함된 **예측값 표준 오차 < 반응 >** 열을 데이터 테이블에 저장합니다 . 자세한 내용은 " 예측 표준 오차 및 신뢰 한계 " 에서 확인하십시오 .

평균 신뢰 한계 계산식 저장 각 Y 변수에 대해 **평균 < 반응 > 의 95% 하한** 및 **평균 < 반응 > 의 95% 상한**이라는 두 열을 데이터 테이블에 저장합니다 . 이 열에는 반응 평균에 대한 95% 신뢰 한계가 포함됩니다 . 자세한 내용은 " 예측 표준 오차 및 신뢰 한계 " 에서 확인하십시오 .

개별값 신뢰 한계 계산식 저장 각 Y 변수에 대해 **개별 < 반응 > 의 95% 하한** 및 **개별 < 반응 > 의 95% 상한**이라는 두 열을 데이터 테이블에 저장합니다 . 이 열에는 개별값에 대한 95% 예측 한계가 포함됩니다 . 자세한 내용은 " 예측 표준 오차 및 신뢰 한계 " 에서 확인하십시오 .

스코어 계산식 저장 다음과 같은 두 개의 열 집합을 데이터 테이블에 저장합니다 .

- 각 X 스코어에 대한 계산식을 포함하는 **X 스코어 <N> 계산식** 열
- 각 Y 스코어에 대한 계산식을 포함하는 **Y 스코어 <N> 계산식** 열

자세한 내용은 " 부분 최소 제곱 " 에서 확인하십시오 .

Y 예측값 저장 Y 변수에 대한 예측값을 데이터 테이블의 열에 저장합니다 .

Y 잔차 저장 Y 변수에 대한 잔차 값을 데이터 테이블의 열에 저장합니다 .

X 예측값 저장 X 변수에 대한 예측값을 데이터 테이블의 열에 저장합니다 .

X 잔차 저장 X 변수에 대한 잔차 값을 데이터 테이블의 열에 저장합니다 .

X 효과에 대해 설명되는 변동 백분율 저장 추출된 모든 요인을 대상으로 각 X 변수에 대해 설명되는 변동 백분율을 새 데이터 테이블에 저장합니다 .

Y 효과에 대해 설명되는 변동 백분율 저장 추출된 모든 요인을 대상으로 각 Y 변수에 대해 설명되는 변동 백분율을 새 데이터 테이블에 저장합니다 .

스코어 저장 추출된 각 요인에 대한 X 스코어와 Y 스코어를 데이터 테이블에 저장합니다 .

적재 저장 X 적재와 Y 적재를 새 데이터 테이블에 저장합니다 .

표준화된 스코어 저장 상관 적재 그림을 생성하는 데 사용된 표준화된 X 스코어 및 Y 스코어를 데이터 테이블에 저장합니다 . 계산식은 " 표준화된 스코어 및 적재량 " 에서 확인하십시오 .

표준화된 적재량 저장 상관 적재 그림을 생성하는 데 사용된 표준화된 X 적재량 및 Y 적재량을 새 데이터 테이블에 저장합니다 . 계산식은 "표준화된 스코어 및 적재량"에서 확인하십시오 .

T 제곱 저장 T^2 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 이 열의 값은 T 제곱 그림에도 사용됩니다 .

거리 저장 X 모형까지의 거리 (DModX) 및 Y 모형까지의 거리 (DModY) 값을 데이터 테이블에 저장합니다 . 이러한 값은 거리 그림에 사용됩니다 .

X 가중치 저장 추출된 모든 요인에 대한 각 X 변수의 가중치를 새 데이터 테이블에 저장합니다 .

JMP PRO **검증 저장** 각 관측값이 검증에 사용된 방법을 기술하는 새 열을 데이터 테이블에 저장합니다. 홀드백 검증의 경우 행이 훈련 또는 검증에 사용되었는지 여부를 식별합니다. K 폴드 검증의 경우 행이 할당된 부분군의 번호를 식별합니다.

JMP PRO **결측값 대치 저장** "결측 데이터 대치"를 선택한 경우 X 및 Y 로 지정된 데이터 테이블 열이 포함된 새 데이터 테이블이 열리고 결측값은 대치된 값으로 바뀝니다. 다항식 항에 대한 열은 표시되지 않습니다. 검증 열을 지정한 경우 검증 열도 포함됩니다.

JMP PRO **예측 계산식 게시** 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

JMP PRO **스코어 계산식 게시** X 및 Y 스코어 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

JMP PRO 부분 최소 제곱의 추가 예

이 예에서는 모형 적합의 부분 최소 제곱 분석법을 사용하여 범주형 반응과 다수의 변수에 대한 분석을 수행합니다. 모형 적합을 통해 부분 최소 제곱을 사용하면 K 폴드 교차 검증도 사용할 수 있습니다. 예제 데이터 집합은 165 명의 남성에게서 수집한 혈청 표본으로 구성되어 있습니다. 남성 중 84 명은 전립선암을 앓고 있고 81 명은 그렇지 않습니다. 분석 목표는 혈청 표본을 기반으로 전립선암에 걸린 남성을 판별할 수 있도록 하는 것입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Prostate Cancer.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Status 를 선택하고 **Y** 를 클릭합니다.
4. Proteins 열 그룹을 선택하고 **추가**를 클릭합니다.
5. "분석법" 목록에서 **부분 최소 제곱**을 선택합니다.
6. **실행**을 클릭합니다.
7. (선택 사항) "부분 최소 제곱"의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다.
8. (선택 사항) "난수 시드값 지정" 옆에 "1234" 를 입력합니다.

   난수 시드값을 지정하면 이 예에 표시된 결과를 재현할 수 있습니다.
9. (선택 사항) **확인**을 클릭합니다.
10. "모형 시작" 에서 **시작**을 클릭합니다.

그림 6.15 부분 최소 제곱 보고서

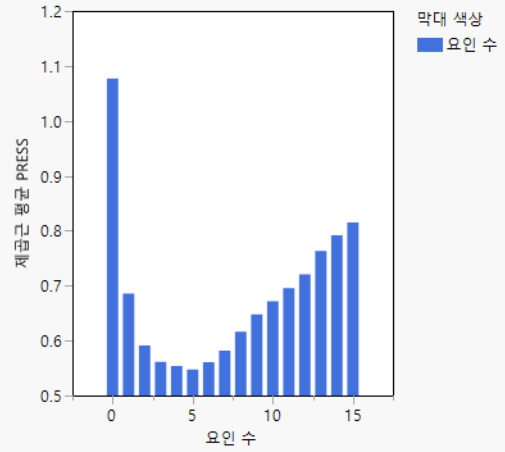
모형 비교 요약

방법	SVD	행 수	요인 수	누적 X에 대해 설명되는 변동 백분율	누적 Y에 대해 설명되는 변동 백분율	VIP 수 > 0.8
NIPALS	빠름	165	5	52.886468	85.671251	626

K=7 및 방법=NIPALS을(를) 사용한 K 폴드 교차 검증(빠름 SVD 사용)

요인 수	제곱근 평균 PRESS	van der Voet T²	Prob > van der Voet T²	Q²	누적 Q²	R²X	누적 R²X	R²Y	누적 R²Y
0	1.076926	130.46053	<.0001*	-0.000625	-0.000625	0.000000	0.000000	0.000000	0.000000
1	0.685794	16.150335	0.0010*	0.593265	0.593265	0.342651	0.342651	0.608907	0.608907
2	0.591463	2.927629	0.0790	0.696945	0.876737	0.077029	0.419680	0.132190	0.741097
3	0.561730	0.480087	0.4740	0.726683	0.966310	0.039013	0.458693	0.066929	0.808026
4	0.553949	0.278455	0.5910	0.734125	0.991043	0.043112	0.501805	0.034490	0.842515
5	0.547744	0.000000	1.0000	0.740188	0.997673	0.055223	0.557027	0.017535	0.860051
6	0.560870	1.357190	0.2350	0.728045	0.999367	0.048092	0.605120	0.015012	0.875063
7	0.581726	3.490817	0.0600	0.707162	0.999815	0.034371	0.639491	0.012864	0.887927
8	0.616081	7.673681	0.0010*	0.671501	0.999939	0.023573	0.663064	0.013162	0.901089
9	0.647871	12.869719	<.0001*	0.636867	0.999978	0.020427	0.683491	0.011323	0.912412
10	0.671625	14.823639	<.0001*	0.610463	0.999991	0.015562	0.699053	0.010699	0.923112
11	0.695867	17.385858	<.0001*	0.581816	0.999996	0.019040	0.718092	0.007945	0.931057
12	0.720865	19.399541	<.0001*	0.551351	0.999998	0.021476	0.739568	0.007050	0.938107
13	0.763351	22.554517	<.0001*	0.496903	0.999999	0.011706	0.751274	0.008857	0.946963
14	0.791689	26.405173	<.0001*	0.458986	1.000000	0.011510	0.762784	0.006236	0.953199
15	0.815424	28.291824	<.0001*	0.426187	1.000000	0.011048	0.773832	0.005424	0.958623

제곱근 평균 PRESS 그림



참고: 최소 제곱근 평균 PRESS는 0.54774이고 최소 요인 수는 5입니다.

" 제곱근 평균 PRESS 그림 " 에서는 요인 수가 5 일 때 제곱근 평균 PRESS 가 최소화된다는 것을 보여 줍니다 . 이 내용은 " 제곱근 평균 PRESS 그림 " 아래의 " 참고 " 에 명시되어 있습니다 . **5 개의 요인을 사용한 NIPALS 적합 (빠른 SVD 사용)** 보고서가 생성되었습니다 .

부분 최소 제곱 플랫폼에 대한 통계 상세 정보

이 섹션에서는 부분 최소 제곱 플랫폼에 사용되는 몇 가지 방법에 대한 상세 정보를 제공합니다. 자세한 내용은 Hoskuldsson(1988), Garthwaite(1994) 또는 Cox 와 Gaudard(2013) 연구 자료에서 확인하십시오.

- " 부분 최소 제곱 "
- "van der Voet T^2"
- "T^2 그림 "
- "X 스코어 산점도 행렬의 신뢰도 타원 "
- " 예측 표준 오차 및 신뢰 한계 "
- " 표준화된 스코어 및 적재량 "
- "PLS-DA(PLS 판별 분석)"

부분 최소 제곱

부분 최소 제곱은 요인, 즉 설명 변수(*X*)의 선형 결합을 기반으로 선형 모형을 적합시킵니다. 이러한 요인은 *X*와 하나 이상의 반응(*Y*) 간의 공분산을 최대화하는 방식으로 얻어집니다. 이런 식으로 PLS는 *X*와 *Y* 사이의 상관을 이용하여 기본 잠재 구조를 나타냅니다. 요인은 반응 변동과 예측 변수 변동을 설명해야 하는 합동 목표를 해결합니다. 부분 최소 제곱은 X 변수가 관측값보다 많거나 X 변수의 상관관계가 높을 때 특히 유용합니다.

NIPALS

NIPALS 방법은 한 번에 하나의 요인을 추출하는 방식으로 작동합니다. $\mathbf{X} = \mathbf{X}_0$ 은 중심화 및 척도화된 예측 변수 행렬이고, $\mathbf{Y} = \mathbf{Y}_0$ 은 중심화 및 척도화된 반응 값 행렬이라고 하겠습니다. PLS 방법은 예측 변수의 선형 결합인 $\mathbf{t} = \mathbf{X}_0\mathbf{w}$ 로 시작합니다. 여기서 **t** 는 **스코어 벡터**이고 **w** 는 연관된 **가중치 벡터**입니다. PLS 방법은 **t** 에 대한 회귀 분석을 통해 $\mathbf{X}_0$ 과 $\mathbf{Y}_0$ 을 모두 예측합니다.

$\hat{\mathbf{X}}_0 = \mathbf{tp}'$, 여기서 $\mathbf{p}' = (\mathbf{t}'\mathbf{t})^{-1}\mathbf{t}'\mathbf{X}_0$

$\hat{\mathbf{Y}}_0 = \mathbf{tc}'$, 여기서 $\mathbf{c}' = (\mathbf{t}'\mathbf{t})^{-1}\mathbf{t}'\mathbf{Y}_0$

벡터 **p** 와 **c** 를 각각 *X* 및 *Y* **적재**라고 합니다.

특정 선형 결합 $\mathbf{t} = \mathbf{X}_0\mathbf{w}$ 는 일부 반응 선형 결합 $\mathbf{u} = \mathbf{Y}_0\mathbf{q}$ 에서 최대 공분산 $\mathbf{t}'\mathbf{u}$ 를 갖는 결합입니다. 또 다른 특성은 *X* 및 *Y* 가중치를 나타내는 **w** 와 **q** 가 공분산 행렬 $\mathbf{X}_0'\mathbf{Y}_0$ 의 첫 번째 왼쪽 및 오른쪽 특이 벡터에 비례한다는 것입니다. 이는 $\mathbf{X}_0'\mathbf{Y}_0\mathbf{Y}_0'\mathbf{X}_0$ 과 $\mathbf{Y}_0'\mathbf{X}_0\mathbf{X}_0'\mathbf{Y}_0$ 의 첫 번째 고유 벡터에 각각 해당합니다.

이것은 첫 번째 PLS 요인이 추출되는 방법을 설명합니다. 두 번째 요인은 $\mathbf{X}_0$ 과 $\mathbf{Y}_0$ 을 첫 번째 요인의 X 및 Y 잔차로 대체하여 동일한 방식으로 추출됩니다.

$$\mathbf{X}_1 = \mathbf{X}_0 - \hat{\mathbf{X}}_0$$

$$\mathbf{Y}_1 = \mathbf{Y}_0 - \hat{\mathbf{Y}}_0$$

이러한 잔차를 **수축된** X 및 Y 블록이라고도 합니다. 스코어 벡터 추출 및 데이터 행렬 수축 과정이 추출된 요인 수만큼 반복됩니다.

SIMPLS

SIMPLS 알고리즘은 통계 기준을 최적화하기 위해 개발되었습니다. 이 방법은 X 스코어가 직교해야 한다는 요건에 따라 X와 Y의 선형 결합 간에 공분산을 최대화하는 스코어 벡터를 찾습니다. $\mathbf{X}_0$ 및 $\mathbf{Y}_0$ 행렬이 수축되는 NIPALS와 달리 SIMPLS는 교차곱 행렬 $\mathbf{X}_0'\mathbf{Y}_0$을 수축시킵니다.

단일 Y 변수의 경우 이 두 알고리즘이 동일합니다. 그러나 다변량 Y의 경우 모형이 다릅니다. SIMPLS는 De Jong(1993)에 의해 제안되었습니다.

van der Voet T^2

van der Voet T^2 검정은 지정된 수만큼 추출된 요인이 있는 모형이 제안된 최적 모형과 유의하게 다른지 여부를 판별하는 데 도움이 됩니다. 이 검정은 두 모형의 잔차 제곱이 동일한 분포를 따른다는 귀무가설에 기반한 랜덤화 검정입니다. 귀무가설은 두 모형의 예측 능력이 동일하다는 것을 나타낸다고 직관적으로 생각할 수 있습니다.

교차 검증 보고서에 제공된 van der Voet T^2 통계량을 얻기 위해 각 검증 데이터 집합에 대해 아래 계산이 수행됩니다. 단일 검증 데이터 집합의 경우 결과는 보고된 값입니다. Leave-One-Out 및 K 폴드 검증의 경우 각 검증 데이터 집합의 결과를 구하여 평균을 계산합니다.

추출된 요인이 i 개 있는 모형의 반응 k 에 대한 j 번째 예측 잔차를 $R_{i,jk}$ 로 나타냅니다. 제안된 최적 요인 수 opt 에 기반한 모형의 해당 값을 $R_{opt,jk}$ 로 나타냅니다. 검정 통계량은 다음 차이를 기반으로 합니다.

$$D_{i,jk} = R^2_{i,jk} - R^2_{opt,jk}$$

K 개의 반응이 있다고 가정해 보겠습니다. 다음 표기를 고려하십시오.

$$\mathbf{d}_{i,j} = (D_{i,j1}, D_{i,j2}, \ldots, D_{i,jK})'$$

$$\mathbf{d}_{i,.} = \sum_j \mathbf{d}_{i,j}$$

$$\mathbf{S_i} = \sum_j \mathbf{d}_{i,j}\mathbf{d}_{i,j}'$$

추출된 i 개의 요인에 대한 van der Voet 통계량은 다음과 같이 정의됩니다 .

$$C_i = \mathbf{d}_{i,.}'\mathbf{S_i}^{-1}\mathbf{d}_{i,.}$$

$R^2_{i,jk}$ 및 $R^2_{opt,jk}$를 무작위로 교환한 결과 값의 분포와 C_i를 비교하여 유의 수준을 구합니다. 이러한 값의 몬테카를로 표본이 시뮬레이션되고, 유의 수준은 C_i보다 크거나 같은 시뮬레이션된 임계값의 비율로 근사값이 산출됩니다.

T^2 그림

i 번째 관측값의 T^2 값은 다음과 같이 계산됩니다 .

$$T_i^2 = (n-1)\sum_{j=1}^{p}\left(t_{ij}^2 / \sum_{k=1}^{n} t_{kj}^2\right)$$

여기서 t_{ij} = i번째 행과 j번째 추출된 요인의 X 스코어, p = 추출된 요인 수, n = 모형을 훈련하는 데 사용된 관측값 수입니다. 검증을 사용하지 않는 경우 n = 총 관측값 수입니다.

T^2 그림의 관리 한계는 다음과 같이 계산됩니다 .

$((n\text{-}1)^2/n)$*BetaQuantile(0.95, p/2, $(n\text{-}p\text{-}1)/2$)

여기서 p = 추출된 요인 수, n = 모형을 훈련하는 데 사용된 관측값 수입니다. 검증을 사용하지 않는 경우 n = 총 관측값 수입니다. 자세한 내용은 Tracy et al. (1992) 연구 자료에서 확인하십시오.

X 스코어 산점도 행렬의 신뢰도 타원

"스코어 산점도 행렬" 옵션은 X 스코어 산점도에 95% 신뢰도 타원을 추가합니다. NIPALS 알고리즘과 SIMPLE 알고리즘 모두 직교 스코어 벡터를 생성하므로 X 스코어에 상관관계가 없습니다. 타원은 X 스코어의 각 쌍이 상관계수가 0인 이변량 정규 분포를 따른다고 가정합니다.

세로 축의 스코어 i 와 가로 축의 스코어 j 에 대한 산점도를 예로 들어보겠습니다 . 타원의 위쪽 , 아래쪽 , 왼쪽 및 오른쪽 극단에 대한 좌표는 다음과 같이 정의됩니다 .

- 위쪽 / 아래쪽 극단 = +/-sqrt(var(score i)*z)
- 왼쪽 / 오른쪽 극단 = +/-sqrt(var(score j)*z)

여기서 z = $((n\text{-}1)*(n\text{-}1)/n)$*BetaQuantile(0.95, 1, $(n\text{-}3)/2$)입니다. Z 값에 대한 배경 정보는 Tracy et al. (1992) 연구 자료에서 확인하십시오.

예측 표준 오차 및 신뢰 한계

X를 예측 변수 행렬로 나타내고 **Y**를 반응 값 행렬로 나타냅니다. 이러한 값은 시작 창에서 선택한 항목에 따라 중심화 및 척도화될 수 있습니다. **Y**의 성분은 공통 분산이 σ^2인 독립 정규 분포를 따른다고 가정합니다.

Hoskuldsson(1988)은 **Y**에 대한 PLS 모형을 스코어 관점으로 보면 다중 선형 회귀 모형과 형식적으로 유사하다는 것을 발견했습니다. 또한 이 유사성을 사용하여 예측값의 분산에 대한 근사 계산식을 도출했습니다. Umetrics(1995) 연구 자료도 확인하십시오. 그러나 Denham(1997)은 PLS에 의해 예측된 값은 Y의 비선형 함수임을 지적했습니다. 또한 예측 구간을 얻기 위해 붓스트랩 및 교차 검증 기법을 제안했습니다. PLS 플랫폼에서는 Umetrics(1995)에 설명된 정규성 기반 방식을 사용합니다.

열에 스코어가 포함된 행렬을 **T**로 나타내고 **X**의 새 관측값 $\mathbf{x}_0$을 예로 들어보겠습니다. **Y**에 대한 예측 모형은 **T**를 기반으로 **Y**를 모델링하는 회귀를 통해 얻습니다. $\mathbf{x}_0$과 관련된 스코어 벡터를 $\mathbf{t}_0$으로 나타냅니다.

a가 요인 수를 나타낸다고 하겠습니다. s^2을 $df = n - a - 1$(데이터가 중심화된 경우) 및 $df = n - a$(데이터가 중심화되지 않은 경우)로 나눈 잔차 제곱합으로 정의합니다. s^2 값은 σ^2의 추정값입니다.

예측 계산식의 표준 오차

$\mathbf{x}_0$에서 예측 평균의 표준 오차는 다음과 같이 추정됩니다.

$$SE(\bar{Y}_{x_0}) = s\sqrt{\left(\frac{1}{n} + \mathbf{t}_0(\mathbf{T}'\mathbf{T})^{-1}\mathbf{t}_0'\right)}$$

평균 신뢰 한계 계산식

$t_{0.975, df}$를 자유도가 $df = n - a - 1$(데이터가 중심화된 경우) 및 $df = n - a$(데이터가 중심화되지 않은 경우)인 t 분포의 0.975 분위수를 나타낸다고 하겠습니다.

평균에 대한 95% 신뢰 구간은 다음과 같이 계산됩니다.

$$\bar{Y}_{x_0} \pm t_{0.975, df} SE(\bar{Y}_{x_0})$$

개별값 신뢰 한계 계산식

$\mathbf{x}_0$에서 예측된 개별 반응의 표준 오차는 다음과 같이 추정됩니다.

$$SE(\hat{Y}_{x_0}) = s\sqrt{\left(\frac{1}{n} + 1 + \mathbf{t}_0(\mathbf{T}'\mathbf{T})^{-1}\mathbf{t}_0'\right)}$$

$t_{0.975, df}$를 자유도가 $df = n - a - 1$(데이터가 중심화된 경우) 및 $df = n - a$(데이터가 중심화되지 않은 경우)인 t 분포의 0.975 분위수를 나타낸다고 하겠습니다.

개별 반응에 대한 95% 신뢰 구간은 다음과 같이 계산됩니다.

$$\overline{Y}_{x_0} \pm t_{0.975,\, df} SE(\hat{Y}_{x_0})$$

표준화된 스코어 및 적재량

다음 표기를 고려하십시오.

- n_{tr} = 훈련 데이터 집합의 관측값 수
- m = X 의 효과 수
- k = Y 의 반응 수
- $VarX_i$ = i 번째 요인에 의해 설명되는 X 의 변동 백분율
- $VarY_i$ = i 번째 요인에 의해 설명되는 Y 의 변동 백분율
- $\mathbf{XScore}_i$ = i 번째 요인에 대한 X 스코어 벡터
- $\mathbf{YScore}_i$ = i 번째 요인에 대한 Y 스코어 벡터
- $\mathbf{XLoad}_i$ = i 번째 요인에 대한 X 적재량 벡터
- $\mathbf{YLoad}_i$ = i 번째 요인에 대한 Y 적재량 벡터

표준화된 스코어

i 번째 표준화된 X 스코어의 벡터는 다음과 같이 정의됩니다.

$$\frac{\mathbf{XScore}_i}{(n_{tr} - 1)\sqrt{mVarX_i / n_{tr}}}$$

i 번째 표준화된 Y 스코어의 벡터는 다음과 같이 정의됩니다.

$$\frac{\mathbf{YScore}_i}{(n_{tr} - 1)\sqrt{kVarY_i / n_{tr}}}$$

표준화된 적재량

i 번째 표준화된 X 적재량의 벡터는 다음과 같이 정의됩니다.

$$\mathbf{XLoad}_i \sqrt{mVarX_i}$$

i 번째 표준화된 Y 적재량의 벡터는 다음과 같이 정의됩니다 .

$$\mathbf{YLoad}_i\sqrt{kVarY_i}$$

PLS-DA(PLS 판별 분석)

시작 창에서 범주형 변수를 Y로 입력하면 표시자 코딩을 사용하여 변수가 코드화됩니다. 수준이 k개인 경우 각 수준은 해당 수준의 행이면 값이 1이고 그렇지 않으면 0인 표시 변수로 나타냅니다. 결과로 생성된 k개의 표시 변수가 연속형으로 처리되고, 연속형 Y를 사용하는 경우와 동일하게 PLS 분석이 진행됩니다.

7 장

다중 대응 분석

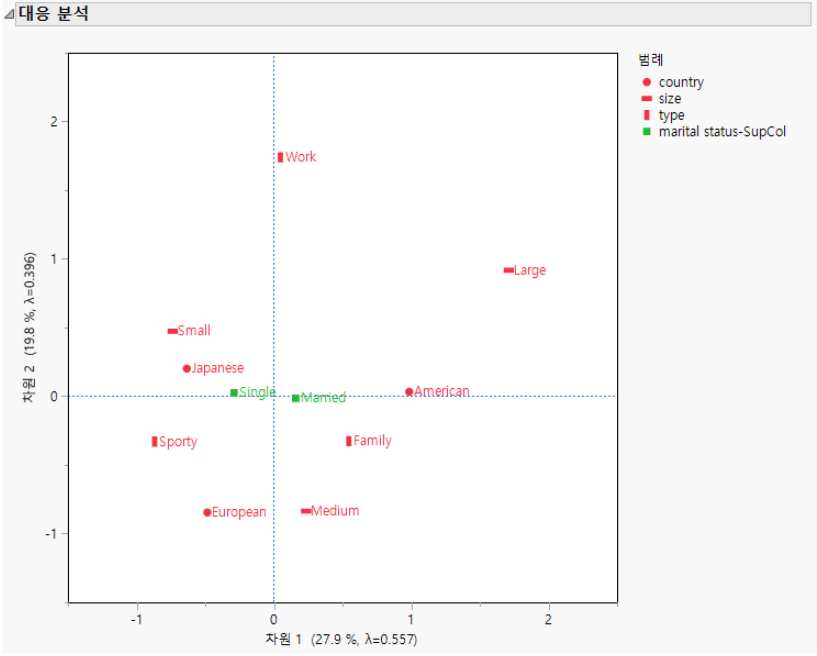
범주형 변수 수준 간의 연관성 식별

MCA(다중 대응 분석)에서는 여러 범주형 변수를 사용하여 해당 변수 수준 간의 연관성을 식별합니다. MCA는 두 개의 변수에서 다수로 대응 분석을 확장합니다. 이는 양적 변수에 대한 주성분 분석과 유사한 것으로 간주될 수 있습니다. 또한 다른 다변량 방법과 마찬가지로 차원 축소 방법이며, 데이터를 2차원 또는 3차원 공간의 점으로 나타냅니다.

다중 대응 분석은 사회 과학 분야에 주로 사용됩니다. 설문 조사 분석에서 질문 합치도를 식별하는 데 사용할 수 있습니다. 또한 상품의 잠재 시장을 파악하기 위해 소비자 조사에도 사용됩니다.

다중 대응 분석에 대한 자세한 내용은 LeRoux와 Rouanet(2010) 연구 자료에서 확인하십시오.

그림 7.1 다중 대응 분석



목차

다중 대응 분석의 예 145
다중 대응 분석 플랫폼 시작 148
다중 대응 분석 보고서 149
다중 대응 분석 플랫폼 옵션 150
대응 분석 옵션 151
그림 표시 152
상세 정보 표시 153
조정 관성 표시 153
좌표 표시 153
요약 통계량 표시 154
관성에 대한 부분 기여도 표시 154
제곱코사인 표시 155
Cochran Q 검정 155
교차표 155
다중 대응 분석 플랫폼의 추가 예 156
보조 변수를 사용한 예 156
보조 ID를 사용한 예 157
Cochran Q 검정의 예 158
다중 대응 분석 플랫폼에 대한 통계 상세 정보 159
상세 정보 보고서 160
조정 관성 160
요약 통계량 161
관성에 대한 부분 기여도 161
Cochran Q 통계량 162

다중 대응 분석의 예

이 예에서는 55 명의 JMP 직원을 대상으로 다음 영역에서 개인의 선호도 또는 취향에 대해 수집한 데이터를 사용합니다.

- 선호하는 TV 프로그램 (8 개 범주): 뉴스 , 코미디 , 경찰 , 자연 , 스포츠 , 영화 , 드라마 또는 연속극
- 선호하는 영화 (8 개 범주): 액션 , 코미디 , 시대극 , 다큐멘터리 , 공포 , 뮤지컬 , 로맨스 또는 SF
- 선호하는 미술 유형 (7 개 범주): 행위 미술 , 풍경화 , 르네상스 , 정물화 , 초상화 , 현대 미술 또는 인상파
- 선호하는 외식 장소 (6 개 범주): 버거 & 감자튀김 , 펍 , 인도 식당 , 이탈리아 식당 , 프랑스 식당 또는 스테이크 전문 식당

설문 조사 대상자에 대한 인구 통계도 데이터에 포함되어 있습니다. 네 가지 질문에 대한 55개 반응 간의 연관성을 탐색하려고 합니다. 이를 통해 다양한 그룹의 직원을 이해할 수 있습니다.

참고 : 선호도 측도는 LeRoux 와 Rouanet(2010) 의 질문을 기반으로 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Employee Taste.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 다중 대응 분석**을 선택합니다 .
3. TV, Film, Art 및 Restaurant 를 선택하고 **Y, 반응**을 클릭합니다 .

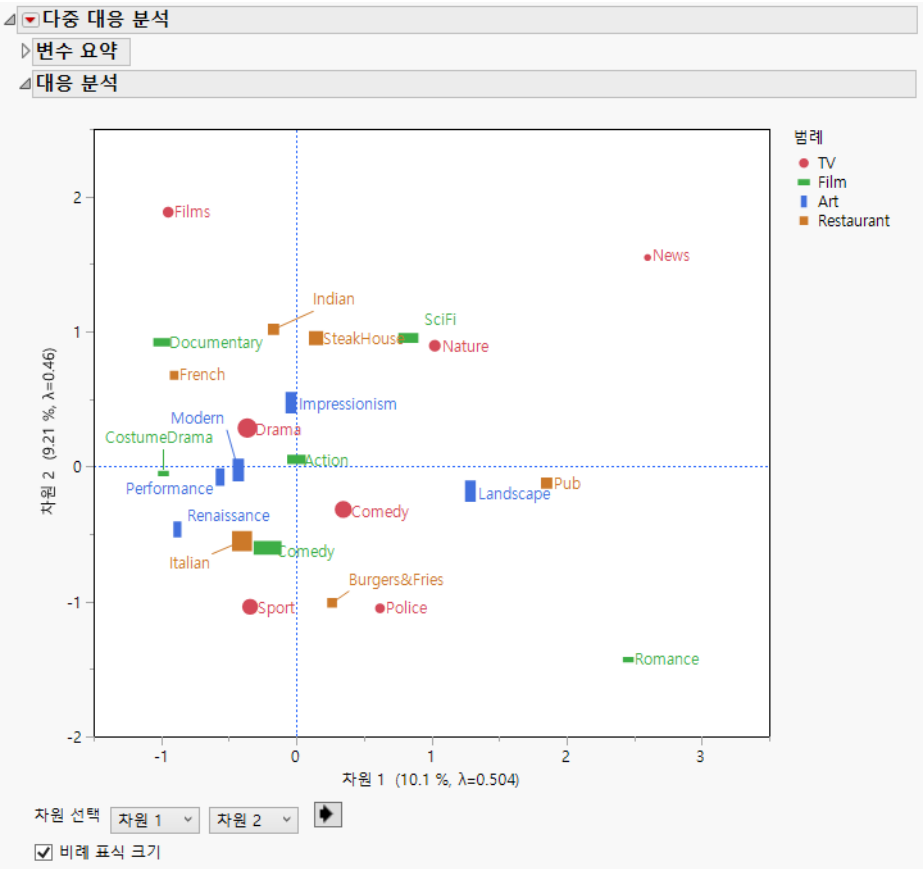
   주성분 분석에서처럼 MCA 에서 요인은 일반적으로 반응과 설명 변수가 아니라 반응으로 간주됩니다 .
4. **확인**을 클릭합니다 .

팁 : 중첩 라벨을 클릭한 후 드래그하여 재배열할 수 있습니다 .

5. 그림 아래에서 **비례 표식 크기**를 선택합니다 .

   표식 크기는 해당 범주에서 반응의 상대적 비율을 나타냅니다 .

그림 7.2 초기 다중 대응 분석 보고서 (일부)



" 대응 분석 " 보고서에는 처음 두 개의 주 축 (차원) 에 투영된 네 가지 변수의 범주가 표시됩니다 . 컨트롤을 사용하여 그림에 표시된 차원을 변경할 수 있습니다 . 점 사이의 거리는 직원의 반응 패턴 간 비유사성을 나타냅니다 .

이 그래프에서 주제에 대한 지식을 활용하여 결과를 해석할 수 있습니다 .

- "Sport" 와 "Police" TV 프로그램 선호 군집이 "Burgers&Fries" 식당 선호 항목 근처에 그룹화되어 있습니다 . 이 군집을 " 대중문화 " 그룹이라고 할 수 있습니다 .
- 차원 2 를 보면 개인의 취향이 대중문화 그룹에서 다큐멘터리 영화를 좋아하고 , TV 드라마를 선호하며 , 프랑스와 인도 요리 같은 전문 식당을 선호하는 사람처럼 좀 더 " 정교한 " 취향으로 분류될 수 있는 것으로 옮겨갑니다 .

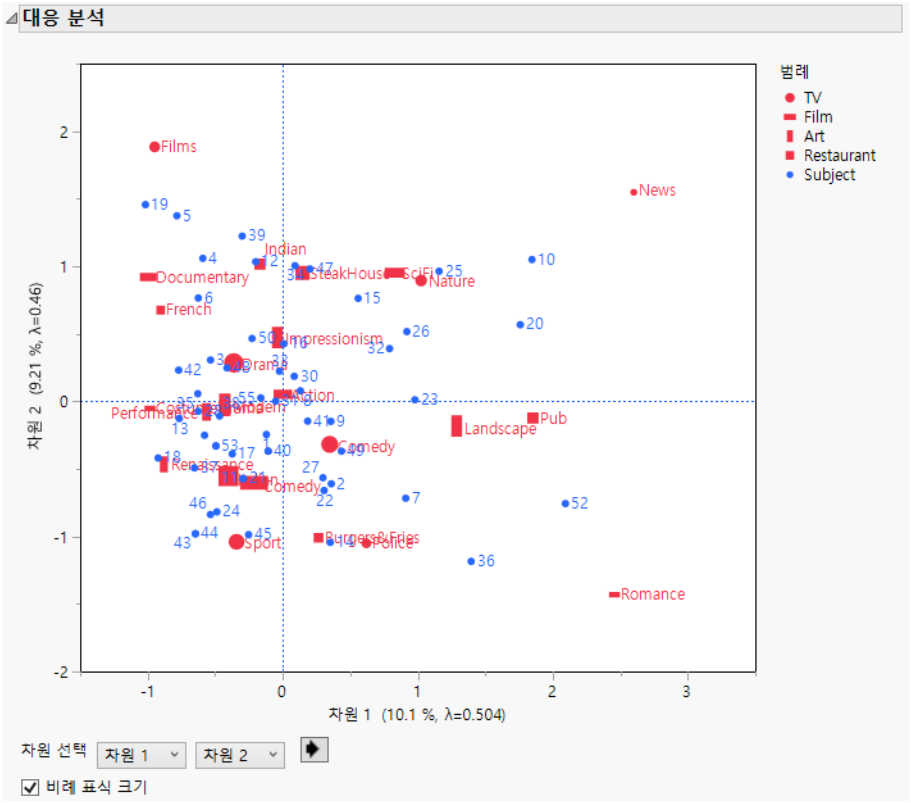
개별값에 대한 스코어를 얻으려면 다음과 같이 Subject 를 **X, 요인**으로 추가합니다 .

6. " 변수 요약 " 개요를 엽니다 .

   " 변수 요약 " 패널을 사용하면 플랫폼을 다시 시작하지 않고 분석을 수정할 수 있습니다 . 또한 완료된 분석에 대한 간략한 보기를 제공합니다 .

7. Subject 를 선택하고 **X 추가**를 클릭합니다 . 그러면 분석이 자동으로 업데이트됩니다 .

그림 7.3 Subject 가 포함된 다중 대응 분석 보고서

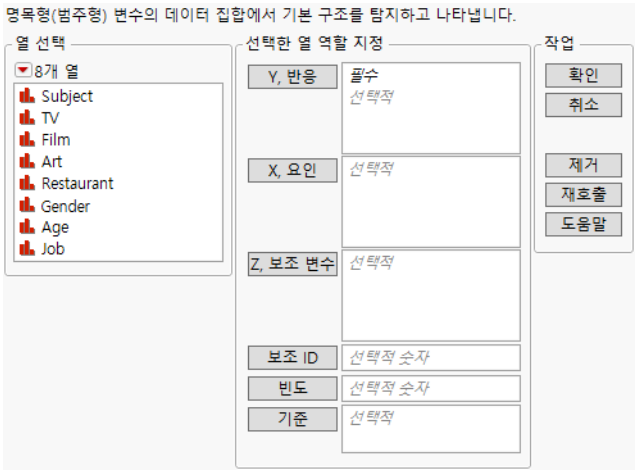


Subject 를 포함하면 그림의 왼쪽 아래 사분면에 취향이 비슷한 직원 군집이 강조 표시됩니다. 해당 직원들은 대중문화 그룹으로 식별한 영역에 속합니다.

다중 대응 분석 플랫폼 시작

분석 > 다변량 방법 > 다중 대응 분석을 선택하여 다중 대응 분석 플랫폼을 시작합니다 .

그림 7.4 다중 대응 분석 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 반응 분석할 범주형 열을 할당합니다 . MCA 에서는 일반적으로 변수 간의 연관성에 관심이 있지만 명시적 " 설명 " 변수와 " 반응 " 변수가 없습니다 .

X, 요인 요인 또는 설명 변수로 사용할 하나 이상의 범주형 열을 할당합니다 .

참고 : 개별값에 대한 스코어를 얻으려면 개체 ID 열을 단일 X 로 사용합니다 .

Z, 보조 변수 보조 변수로 사용할 열을 할당합니다. 이러한 변수는 연관성을 식별하는 데 관심이 있지만 계산에 포함되지 않습니다 . 보조 변수는 데이터 해석을 향상시키는 데 사용됩니다 .

보조 ID 보조로 사용할 행을 식별하는 열을 할당합니다 . 일반적으로 보조 ID 열의 값은 1 과 0 입니다 . ID 0 과 관련된 행은 보조 행으로 처리됩니다 . 비보조 행에 없는 X 또는 Y 변수의 수준이 보조 행에 있으면 보조 ID 열이 무시됩니다 .

참고 : " 보조 ID" 역할과 "Z, 보조 변수 " 역할은 한 번에 하나만 지정할 수 있습니다 .

빈도 이 역할에 빈도 변수를 할당합니다 . 이 기능은 데이터를 요약할 경우에 유용합니다 .

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

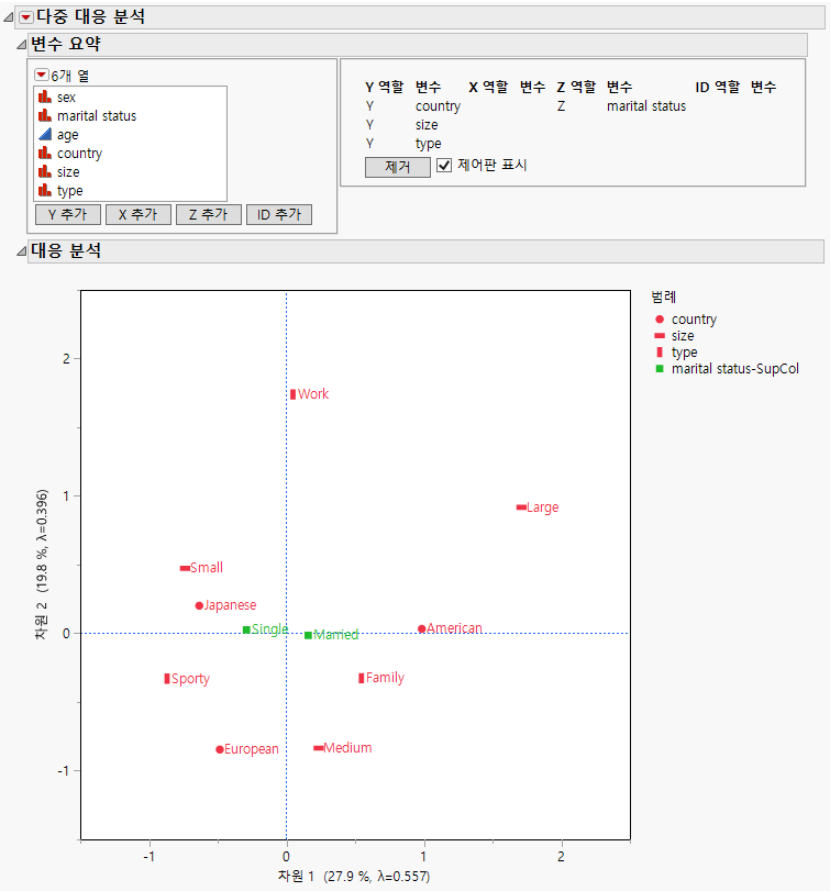
참고 : 다중 대응 분석 플랫폼에서는 결측값을 많은 다른 JMP 플랫폼과 다르게 처리합니다 . 분석할 때 모든 비결측 셀 쌍을 하나의 행에 사용합니다 . 계산에서 전체 행이 제거되지 않습니다 .

다중 대응 분석 보고서

초기 다중 대응 분석 보고서에는 변수 요약, 대응 분석 그림 및 중요도 순으로 나타낸 데이터 차원 상세 정보가 표시됩니다. 범주 또는 개별값 클라우드 그림을 통해 데이터 내에 존재하는 연관성을 식별할 수 있습니다. 상세 정보에는 그림에 표시된 두 차원이 데이터 내의 관계를 해석하기에 충분한지 여부에 대한 정보가 제공됩니다.

대화식 "변수 요약" 패널을 사용하면 플랫폼을 다시 시작하지 않고 분석을 수정할 수 있습니다. 이 패널에는 분석에 사용된 열과 역할이 표시됩니다. "제어판 표시" 체크박스를 선택하면 데이터 테이블의 열 목록이 왼쪽에 나타납니다. 열을 선택하고 "Y 추가", "X 추가", "Z 추가" 또는 "ID 추가"를 클릭하여 분석의 열을 변경할 수 있습니다. 또는 열을 변수 요약 테이블의 머리글로 드래그할 수 있습니다. 이렇게 하면 시작 창으로 돌아가지 않고 분석을 수정할 수 있습니다.

그림 7.5 제어판 표시가 선택된 다중 대응 분석 보고서



다중 대응 분석 플랫폼 옵션

" 다중 대응 분석 " 의 빨간색 삼각형 메뉴 옵션을 사용하면 필요에 따라 보고서를 사용자 정의할 수 있습니다 . 사용할 수 있는 보고서는 사용자가 수행하는 분석 유형에 따라 결정됩니다 .

교차표 선택한 변수 역할에 맞게 Burt 표 또는 분할표를 표시하거나 숨깁니다 . 자세한 내용은 " 교차표 " 에서 확인하십시오 .

대응 분석 대응 분석 보고서를 표시하거나 숨깁니다 . 이러한 보고서에는 그림 , 상세 정보 , 좌표 및 요약 통계량이 제공됩니다 . 자세한 내용은 " 다중 대응 분석 플랫폼의 추가 예 " 에서 확인하십시오 .

보조 행 교차표 (하나 이상의 보조 Z 및 X 변수가 지정되거나 ID 가 지정된 경우 사용 가능) 다음 규칙에 따라 Burt 표 또는 분할표를 표시하거나 숨깁니다 .

- 하나 이상의 X 및 Z 변수가 지정된 경우 X 변수 대 반응 변수의 분할표를 표시하거나 숨깁니다 .
- ID 가 지정되고 X 변수가 지정되지 않은 경우 추가 관측값에 대한 Burt 표를 표시하거나 숨깁니다 . 추가 관측값은 ID 변수가 0 인 관측값입니다 .
- ID 와 하나 이상의 X 변수가 지정된 경우 추가 관측값에 대한 분할표를 표시하거나 숨깁니다 . 추가 관측값은 ID 변수가 0 인 관측값입니다 .

보조 열 교차표 (하나 이상의 보조 Z 변수가 지정된 경우 사용 가능) 다음 규칙에 따라 분할표를 표시하거나 숨깁니다 .

- X 변수가 지정되지 않은 경우 반응 변수 대 보조 변수의 분할표를 표시하거나 숨깁니다 .
- 하나 이상의 X 변수가 지정된 경우 X 변수 대 보조 변수의 분할표를 표시하거나 숨깁니다 .

모자이크 그림 (X 변수와 Y 변수가 각각 하나씩 있는 경우에만 사용 가능) Y 대 X 모자이크 그림을 표시하거나 숨깁니다 . 모자이크 그림은 각 세그먼트가 해당 그룹의 빈도 수에 비례하는 누적 막대 차트입니다 .

독립성에 대한 검정 (X 변수와 Y 변수가 각각 하나씩 있는 경우에만 사용 가능) X 변수와 Y 변수 간의 독립성 검정을 표시하거나 숨깁니다 . 이 검정에는 Pearson 형식과 가능도비 형식의 두 가지 버전이 있으며 둘 다 카이제곱 통계량을 사용합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

대응 분석 옵션

" 대응 분석 " 에서 사용할 수 있는 보고서는 사용자가 수행하는 다중 대응 분석 유형에 따라 결정됩니다 . 이러한 보고서 중 일부는 기본적으로 표시됩니다 .

그림 표시 처음 두 개의 주 축에 의해 설명되는 평면에 2 차원 범주 클라우드를 표시합니다 . 이 그림은 기본적으로 나타납니다 . 이 그림에는 등간격 척도가 사용됩니다 .

상세 정보 표시 특이값 , 관성 , 카이제곱 통계량 , 백분율 및 누적 백분율을 포함하는 분석 상세 정보를 제공합니다 . 이 보고서는 기본적으로 나타납니다 . 자세한 내용은 " 상세 정보 표시 " 에서 확인하십시오 .

조정 관성 표시 Benzecri 조정 관성 보고서와 Greenacre 조정 관성 보고서를 제공합니다. 자세한 내용은 Benzecri(1979) 및 Greenacre(1984) 연구 자료에서 확인하십시오 . 하나 이상의 X 변수가 있는 경우 이 옵션을 사용할 수 없습니다 . 자세한 내용은 " 조정 관성 표시 " 에서 확인하십시오 .

좌표 표시 필요한 경우 분석의 범주에 대해 최대 처음 세 개의 주 좌표 보고서를 제공합니다 . 자세한 내용은 " 좌표 표시 " 에서 확인하십시오 .

요약 통계량 표시 분석의 각 범주에 대한 품질 , 질량 및 관성 요약 통계량 보고서를 제공합니다 . 자세한 내용은 " 요약 통계량 표시 " 에서 확인하십시오 .

관성에 대한 부분 기여도 표시 처음 세 개의 각 차원에서 관성에 대한 각 범주의 기여도를 보여 주는 보고서를 제공합니다. 자세한 내용은 "관성에 대한 부분 기여도 표시"에서 확인하십시오.

제곱코사인 표시 처음 세 개의 각 차원에서 각 범주의 제곱코사인을 보여 주는 보고서를 제공합니다 . 이 보고서에는 각 Y 변수의 각 수준에서 처음 세 개의 각 차원에 대한 제곱코사인 값을 표시하는 막대 차트가 포함되어 있습니다 . 자세한 내용은 " 제곱코사인 표시 " 에서 확인하십시오 .

Cochran Q 검정 (모든 Y 변수에 두 수준으로 구성된 동일한 집합이 있고 X 변수에 각 행에 대한 고유 값이 있는 경우에만 사용 가능) 특정 반응의 주변 확률이 Y 변수에 따라 변하지 않음을 검정하는 Cochran *Q* 통계량을 제공합니다 . Cochran *Q* 통계량은 세 개 이상의 반응 변수에 대한 McNemar 통계량을 일반화한 것입니다 . Agresti(2013) 연구 자료를 참조할 수 있습니다 . 자세한 내용은 "Cochran Q 검정 " 에서 확인하십시오 .

3D 대응 분석 처음 세 개의 주 축에 의해 설명되는 공간에 Y, X 및 Z 변수의 3 차원 범주 클라우드를 표시합니다 . 차원이 세 개 미만인 경우 이 옵션을 사용할 수 없습니다 .

좌표 저장 주 좌표를 하나 이상의 JMP 데이터 테이블에 저장합니다 . 열 좌표 , 행 좌표 , 보조 열 좌표 및 보조 행 좌표가 별도의 JMP 데이터 테이블에 저장됩니다 . 저장할 열 수를 선택할 수 있습니다 .

좌표 계산식 저장 여러 차원의 주 좌표에 대한 계산식 열을 데이터 테이블에 저장합니다 . 각 관측값에 대한 값은 각 차원에 대한 특이값으로 척도화된 Y 변수의 좌표 평균입니다 . 저장할 열 수를 선택할 수 있습니다 .

그림 표시

그림에는 처음 두 개의 주 축에 의해 설명되는 평면에 투영된 범주 또는 개별값이 표시됩니다. 이 그림에는 등간격 척도가 사용됩니다. 그림 아래의 "차원 선택" 컨트롤을 사용하여 그림에 표시되는 차원을 전환할 수 있습니다. 첫 번째 컨트롤은 그림의 가로 축을 정의하고 두 번째 컨트롤은 그림의 세로 축을 정의합니다. 화살표 버튼을 클릭하여 그림에 표시되는 차원을 순환할 수 있습니다. 그림의 점 크기를 각 점에 해당하는 관측값 수에 비례하여 표시할지 여부를 지정하려면 **비례 표식 크기** 체크박스를 사용합니다.

참고 : 대응 분석 그림에서 점을 선택하면 보고서 창의 다른 테이블에 있는 해당 행도 선택됩니다 . 그러나 데이터 테이블의 행은 선택되지 않습니다 . 그림에서 특정 변수와 관련된 모든 점을 선택하려면 그림 범례에서 변수 이름을 선택합니다 .

상세 정보 표시

특이값 테이블을 표시합니다 .

특이값 분할표 또는 Burt 표의 특이값 분해에서 특이값을 표시합니다 . 계산식은 " 상세 정보 보고서 " 에서 확인하십시오 .

관성 정준 차원의 측면을 고려한 상대적인 변동을 반영하는 특이값 제곱을 나열합니다 .

카이제곱 Burt 표 또는 분할표에 대한 전체 카이제곱에서 차원이 나타내는 부분을 나열합니다 .

백분율 총 관성에 대한 관성 비율입니다 .

누적 백분율 관성의 누적 비율을 보여 줍니다 . 처음 두 개의 특이값에서 큰 관성이 발견되는 경우에는 2D 대응 분석 그림으로 테이블에서의 관계를 충분히 나타낼 수 있습니다 .

조정 관성 표시

MCA 에서 Burt 표의 주 관성은 고유값입니다 . 이러한 관성을 사용하면 비관적인 적합 표시가 제공된다는 문제가 있습니다 . Benzecri 는 관성 조정을 제안했습니다 . Greenacre 는 Benzecri 의 조정이 적합 품질을 과대추정한다고 주장하며 대체 조정 방법을 제안했습니다 . 참조를 위해 두 조정 값이 모두 계산됩니다 . 자세한 내용은 " 조정 관성 " 에서 확인하십시오 .

관성 정준 차원의 측면을 고려한 상대적인 변동을 반영하는 특이값 제곱을 나열합니다 .

조정 관성 Benzecri 또는 Greenacre 조정에 따라 조정된 관성을 나열합니다 .

백분율 총 관성에 대한 조정 관성의 비율입니다 .

누적 백분율 조정 관성의 누적 비율을 보여 줍니다 . 처음 두 개의 특이값에서 큰 관성이 발견되는 경우에는 2D 대응 분석 그림으로 테이블에서의 관계를 충분히 나타낼 수 있습니다 .

좌표 표시

" 열 좌표 " 테이블 또는 " 행 및 열 좌표 " 테이블을 표시합니다 .

X X, 요인 변수로 지정된 열을 나열합니다 .

Y Y, 반응 변수로 지정된 열을 나열합니다 .

Z Z, 보조 변수로 지정된 열을 나열합니다 .

범주 X, Y 또는 Z 변수의 수준을 나열합니다 .

차원 1, 차원 2, 차원 3 표시된 주 축에서 각 수준 또는 각 반응의 좌표를 나열합니다 . 기본적으로 처음 세 개 차원에 대한 좌표가 테이블에 표시됩니다 . 추가 차원에 대한 좌표 열은 숨겨집니다 . 이러한 선택적 열을 표시하려면 테이블에서 마우스 오른쪽 버튼을 클릭하고 **열** 하위 메뉴에서 차원 열을 선택합니다 .

참고 : X, 요인 변수로 지정된 열이 있는 경우 동일한 머리글을 사용하는 좌표 보고서에 X 테이블과 Y 테이블이 모두 표시됩니다 . Z, 보조 변수가 지정된 경우에는 X 및 Y 좌표 아래에 해당 좌표가 나열됩니다 .

요약 통계량 표시

" 열 점에 대한 요약 통계량 " 테이블 또는 " 행 및 열 점에 대한 요약 통계량 " 테이블을 표시합니다 . Y 테이블에는 **열 점**이라고 하는 각 반응의 각 수준에 대한 품질 , 질량 및 관성이 제공됩니다 . X 테이블에는 X, 요인 변수의 각 수준에 대한 품질 , 질량 및 관성이 제공됩니다 . 자세한 내용은 " 요약 통계량 " 에서 확인하십시오 .

X X, 요인 변수로 지정된 열을 나열합니다 .

Y Y, 반응 변수로 지정된 열을 나열합니다 .

범주 X 또는 Y 변수의 수준을 나열합니다 .

품질 (차원 =2) 해별로 수준 표현의 품질을 나열합니다 .

질량 반응 수준에 대한 행 빈도를 총 빈도로 나눈 값을 나열합니다 . Burt 표에서 이 값은 각 행의 총 % 입니다 .

관성 반응 수준에 의해 설명되는 총 관성의 비율을 나열합니다 . 전체 수준 및 해당 반응에 대해 관성 값의 합은 1 입니다 .

참고 : X, 요인 변수로 지정된 열이 있는 경우 동일한 머리글을 사용하는 요약 통계량 보고서에 X 테이블과 Y 테이블이 모두 표시됩니다 .

관성에 대한 부분 기여도 표시

" 열 점의 관성에 대한 부분 기여도 " 테이블 또는 " 행 및 열 점의 관성에 대한 부분 기여도 " 테이블을 표시합니다 . " 열 점의 관성에 대한 부분 기여도 그림 " 도 표시됩니다 . 이 그림은 각 Y 변수의 각 수준을 대상으로 테이블에 표시된 각 차원에 대한 부분 기여도를 보여 주는 막대 차트입니다 .

X X, 요인 변수로 지정된 열을 나열합니다 .

Y Y, 반응 변수로 지정된 열을 나열합니다 .

범주 X 또는 Y 변수의 수준을 나열합니다 .

차원 1, 차원 2, 차원 3 표시된 차원의 관성에 대한 반응 또는 요인 수준의 기여도를 나열합니다 . 기본적으로 처음 세 개 차원에 대한 열이 테이블에 표시됩니다 . 추가 열은 숨겨집니다 . 이러한 선택적 열을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열** 하위 메뉴에서 차원 열을 선택합니다 .

각 반응의 각 수준은 각 차원의 관성에 기여합니다 . 각 차원 내에서 부분 기여도의 합은 1 입니다 .

참고 : X, 요인 변수로 지정된 열이 있는 경우 동일한 머리글을 사용하는 관성에 대한 부분 기여도 보고서에 X 테이블과 Y 테이블이 모두 표시됩니다 . 자세한 내용은 " 관성에 대한 부분 기여도 " 에서 확인하십시오 .

제곱코사인 표시

" 열 점에 대한 제곱코사인 " 테이블 또는 " 행 및 열 점에 대한 제곱코사인 " 테이블을 표시합니다 . " 열 점의 제곱코사인 그림 " 도 표시됩니다 . 이 그림은 각 Y 변수의 각 수준을 대상으로 표시된 처음 세 개의 각 차원에 대한 제곱코사인 값을 보여 주는 막대 차트입니다 .

X X, 요인 변수로 지정된 열을 나열합니다 .

Y Y, 반응 변수로 지정된 열을 나열합니다 .

범주 X 또는 Y 변수의 수준을 나열합니다 .

차원 1, 차원 2, 차원 3 표시된 차원별로 수준 표현의 **품질**을 나열합니다 . 기본적으로 처음 세 개 차원에 대한 결과가 테이블에 표시됩니다 . 추가 열은 숨겨집니다 . 이러한 선택적 열을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열** 하위 메뉴에서 차원 열을 선택합니다 .

값은 표시된 차원에 대한 각 점의 품질을 나타냅니다 . 제곱코사인을 점과 차원의 상관 제곱으로 해석할 수 있습니다 . 처음 두 차원의 제곱코사인 합은 요약 통계량 보고서의 " 품질 (차원 =2)" 값과 같습니다 . 자세한 내용은 " 요약 통계량 " 에서 확인하십시오 .

참고 : X, 요인 변수로 지정된 열이 있는 경우 동일한 머리글을 사용하는 제곱코사인 보고서에 X 테이블과 Y 테이블이 모두 표시됩니다 .

Cochran Q 검정

Cochran *Q* 검정은 세 개 이상의 이분 결과에 대한 전체 매칭 표본의 비모수 동질성 검정입니다. 이 검정을 사용하여 매칭 쌍의 비율 차이를 판별할 수 있습니다. Cochran *Q* 검정은 결과가 두 개인 경우에 사용되는 McNemar 검정을 확장한 것입니다.

교차표

교차표 옵션은 Burt 표 또는 분할표를 표시하거나 숨깁니다. X, 요인 열 없이 Y, 반응 열을 여러 개 선택하면 Burt 표가 생성됩니다. X, 요인 열을 선택하면 Burt 표 대신 기존의 분할표가 생성됩니다. 개요 노드 제목은 교차표의 구조를 반영합니다.

Burt 표는 다중 대응 분석의 기반입니다. 이 표는 모든 범주형 변수 쌍에 대한 분할된 대칭 테이블입니다. 대각 분할은 대각 행렬입니다(변수 자체의 교차표). 비대각 분할은 일반적인 분할표입니다.

"Burt 표 " 또는 " 분할표 " 의 빨간색 삼각형 메뉴에는 표에 표시할 다음 통계량 옵션이 포함되어 있습니다 .

개수 셀 빈도 , 빈도 소계 및 총 합계 (총 표본 크기) 입니다 . 이 값은 기본적으로 표시됩니다 .

총 % 총 합계 대비 셀 개수 및 소계의 백분율입니다 . 이 값은 기본적으로 표시됩니다 .

셀 카이제곱 각 셀에 대해 $(O - E)^2 / E$ 로 계산된 카이제곱 값입니다 .

열 % 열 합계 대비 각 셀 개수의 백분율입니다 .

행 % 행 합계 대비 각 셀 개수의 백분율입니다 .

기대값 독립성을 가정한 상태에서 각 셀의 기대 빈도 (*E*) 입니다 . 해당 행 합계와 열 합계의 곱을 총 합계로 나눠서 계산됩니다 .

편차 관측된 셀 빈도 (*O*) 에서 기대 셀 빈도 (*E*) 를 뺀 값입니다 .

열 누적 누적 열 합계입니다 .

열 누적 % 누적 열 백분율입니다 .

행 누적 누적 행 합계입니다 .

행 누적 % 누적 행 백분율입니다 .

데이터 테이블로 만들기 표에 표시된 각 통계량에 대해 하나씩 데이터 테이블을 생성합니다 .

다중 대응 분석 플랫폼의 추가 예

- " 보조 변수를 사용한 예 "
- " 보조 ID 를 사용한 예 "
- "Cochran Q 검정의 예 "

보조 변수를 사용한 예

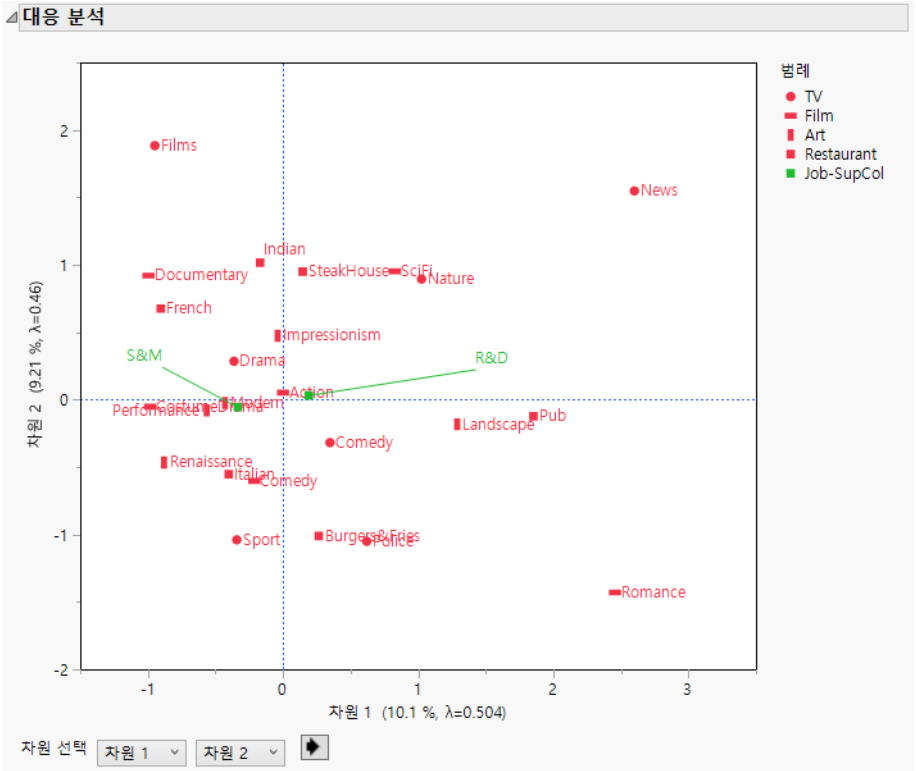
이 예에서는 " 다중 대응 분석의 예 " 에 설명된 JMP 직원의 선호도에 대한 샘플 데이터를 사용합니다 .

직원의 업무와 선호도 간의 관계를 탐색하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Employee Taste.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 다중 대응 분석**을 선택합니다 .
3. TV, Film, Art 및 Restaurant 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Job 을 선택하고 **Z, 보조 변수**를 클릭합니다 .
5. **확인**을 클릭합니다 .

참고 : Job 은 보조 변수이므로 계산에 사용되지 않습니다 . 계산이 완료된 후 Job 유형이 그림에 표시됩니다 .

그림 7.6 보조 변수를 사용한 MCA 보고서



그림에서 R&D(연구 개발) 직원과 S&M(영업 및 마케팅) 직원에 대한 데이터 점이 서로 가까이 있다는 것을 알 수 있습니다. 이는 직원의 업무 기능으로 인한 선호도에 큰 차이가 없다는 의미로 해석됩니다. 즉, 직원의 업무 역할은 선호도에 유의한 요인이 아닙니다.

보조 ID 를 사용한 예

미국 인구조사를 통해 지난 세기 동안의 인구 증가를 검토할 수 있습니다 . US Regional Population.jmp 샘플 데이터 테이블에는 1920 년부터 2010 년까지의 각 인구조사 연도에 대해 지역별로 그룹화된 미국 50 개 주의 인구가 포함되어 있습니다 . Alaska 와 Hawaii 는 전체 기간 중에 주가 아니었고 미국 본토에 속하지 않으므로 보조 지역으로 처리됩니다 . 이 두 주의 인구 증가가 미국의 나머지 주와 다른지 여부에 관심이 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 US Regional Population.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 다중 대응 분석**을 선택합니다 .

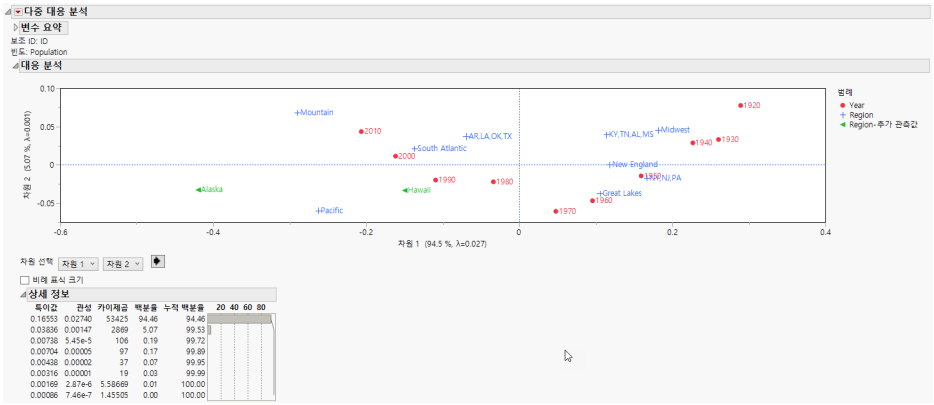
3. Year 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Region 을 선택하고 **X, 요인**을 클릭합니다 .
5. ID 를 선택하고 **보조 ID** 를 클릭합니다 .
6. Population 을 선택하고 **빈도**를 클릭합니다 .
7. **확인**을 클릭합니다 .

"상세 정보" 보고서에서는 연도와 지역 간의 연관성이 첫 번째 차원에 의해 거의 전부 설명된다는 것을 보여 줍니다. 그림을 보면 첫 번째 차원에서 연도가 정확한 순서로 되어 있음을 알 수 있습니다. 이 순서는 대응 분석을 통해 자연스럽게 일어났으며 순서 정보가 분석에 제공되지 않았습니다. 데이터를 표시하는 데 사용된 등간격 척도가 그림에서 강조 표시됩니다.

지역 순서는 Midwest 에서 시작하여 Northeast, South, Mountain West 로 이어지는 인구 이동을 반영합니다.

Alaska 와 Hawaii 는 분석 계산에 사용되지 않았지만 결과를 기반으로 그림에 표시되었습니다. 증가 패턴은 Pacific 주와 가장 유사합니다. Alaska 의 증가는 Pacific 지역보다 훨씬 더 극단적입니다.

그림 7.7 보조 ID 를 사용한 MCA 보고서



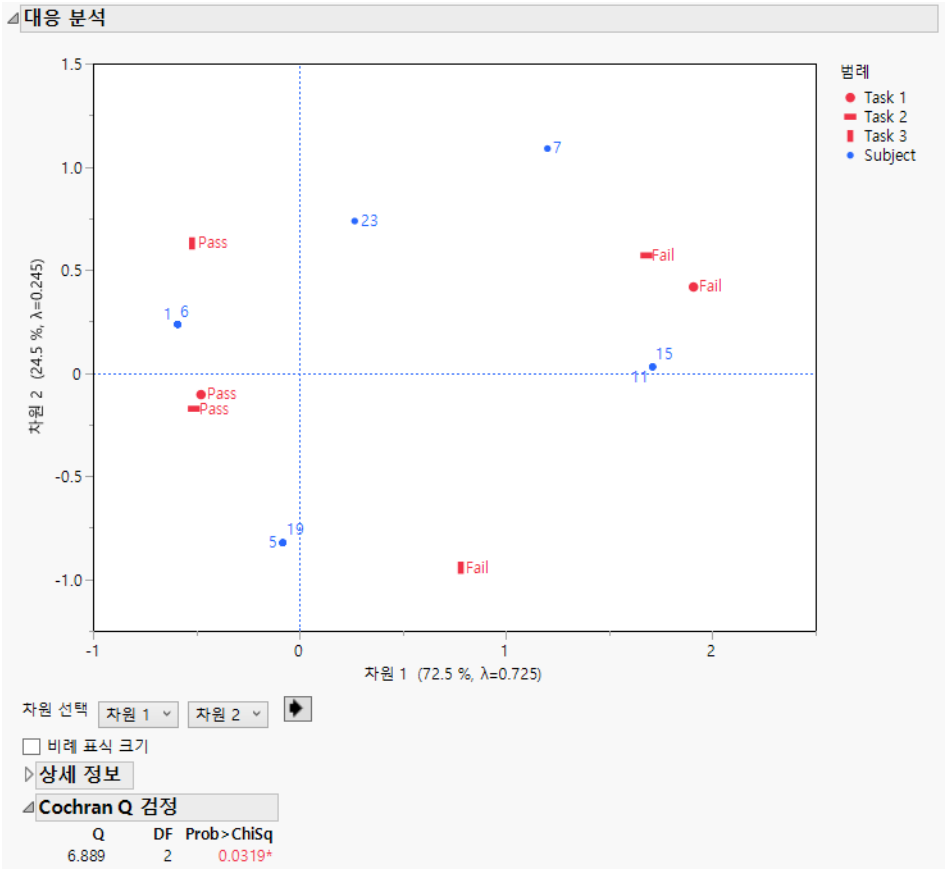
Cochran Q 검정의 예

이 예에서는 시뮬레이션된 데이터 테이블을 사용하여 30 명의 피실험자를 대상으로 세 가지 작업의 완료 용이성을 평가합니다. Task 1 은 완료율 80% 로 시뮬레이션되었습니다. Tasks 2 와 3 은 각각 피실험자의 95% 와 80% 에 대해 Task 1 과 동일한 결과를 갖도록 시뮬레이션되었습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cohcrans Q.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 다중 대응 분석**을 선택합니다 .
3. Task 1, Task 2, Task 3 을 선택하고 **Y, 반응**을 클릭합니다 .
4. Subject 를 선택하고 **X, 요인**을 클릭합니다 .

5. **확인**을 클릭합니다 .
6. " 다중 대응 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **대응 분석 > Cochran Q 검정**을 선택합니다 .

그림 7.8 Cochran *Q* 검정



대응 분석 그림에서 Pass 및 Fail 결과가 함께 군집화되어 있으므로 Task 1 과 Task 2 간의 유사성을 보여 줍니다. 그러나 Task 3 은 Tasks 1 과 2 에서 멀리 떨어져 있습니다. Cochran *Q* 검정 통계량 6.889 의 관련 *p* 값은 0.0319 입니다. 이 *p* 값은 0.05 유의 수준에서 모든 작업에 대한 통과율이 동일하다는 가설이 기각됨을 뒷받침합니다.

다중 대응 분석 플랫폼에 대한 통계 상세 정보

- " 상세 정보 보고서 "
- " 조정 관성 "

- "요약 통계량"
- "관성에 대한 부분 기여도"
- "Cochran Q 통계량"

상세 정보 보고서

단순 대응 분석이 수행되면 다음과 같은 특이값 분해의 특이값이 보고서에 나열됩니다.

$$\mathbf{D}_r^{-0.5}(\mathbf{P} - rc')\mathbf{D}_c^{-0.5} = \mathbf{U}Diag(\Lambda)\mathbf{V}'$$

다음은 각 요소에 대한 설명입니다.

- $\mathbf{P}$ = 개수를 총 빈도로 나눈 값의 행렬
- r, c = $\mathbf{P}$ 의 행 및 열 합계
- $\mathbf{D}$ 행렬 = r 및 c 값의 대각 행렬
- Λ = 특이값의 열 벡터

다중 대응 분석이 수행되면 특이값 분해가 다음 방정식으로 확장됩니다.

$$\mathbf{D}^{-0.5}(\mathbf{C} - dd')\mathbf{D}^{-0.5} = \mathbf{U}Diag(\Lambda)\mathbf{V}'$$

다음은 각 요소에 대한 설명입니다.

- $\mathbf{C}$ = Burt 표
- d = $\mathbf{C}$ 의 열 합계에 대한 열 벡터 ($\mathbf{C}$ 가 대칭이므로 d = 행 합계)
- $\mathbf{D}$ = d 값의 대각 행렬

"상세 정보" 보고서에서 관성은 열 벡터 Λ입니다. 특이값은 관성 벡터의 제곱근입니다. 열 좌표는 다음과 같이 계산됩니다.

$$\mathbf{D}^{-0.5}\mathbf{V}Diag\left(\Lambda^{0.5}\right)$$

조정 관성

MCA 에서 m 개의 범주형 변수로 생성된 Burt 표의 일반적인 주 관성은 Λ^2 의 고유값 λ_k 입니다. 이러한 관성은 비관적인 적합 표시를 제공합니다. Benzecri(1979)는 다음과 같은 관성 조정을 제안했습니다. 이 내용은 Greenacre(1984, p. 145) 연구 자료에도 설명되어 있습니다.

$$\left(\frac{m}{m-1}\right)^2 \times \left(\lambda_k - \frac{1}{m}\right)^2 \quad \text{단,} \quad \lambda_k > \frac{1}{m}$$

이 조정은 $1/m$ 보다 큰 모든 관성에 대해 조정 관성의 합을 기준으로 조정 관성의 백분율을 계산합니다.

Greenacre(1984, p. 156) 연구 자료에는 Benzecri 조정이 적합 품질을 과대추정한다는 내용이 나옵니다 . 대신 Greenacre 는 다음을 기준으로 조정 관성의 백분율을 계산할 것을 제안합니다 .

$$\frac{m}{m-1}\left(\text{tr}\left(Diag\left(\Lambda^4\right)\right)-\frac{n_c-m}{m^2}\right)$$

이 계산은 $1/m$ 보다 큰 모든 관성을 대상으로 합니다. 여기서 $\text{tr}(Diag(\Lambda^4))$ 은 관성 제곱합이고, n_c 는 m 개의 변수에 대한 총 범주 수입니다.

요약 통계량

품질은 지정된 수의 차원으로 정의되는 공간에서 원점과 점 사이의 거리 제곱과 최대 차원 수를 갖는 공간에서 원점과의 거리에 대한 비율입니다. 카이제곱 측정 기준의 경우 주어진 차원의 점 품질은 차원을 정의하는 벡터를 사용하여 계산된 코사인을 통해 얻을 수 있습니다. 또한 품질은 지정된 차원의 관성 합 대 모든 차원의 관성 합 비율과 같습니다. 품질은 저차원 공간에서 점이 얼마나 잘 표현되는지를 나타냅니다.

질량은 총 빈도에 대한 행 또는 열 빈도의 비율입니다.

관성은 주성분 분석의 분산과 유사합니다. 전체 관성은 테이블에 있는 모든 관측값의 합으로 나눈 이원 빈도 테이블의 총 Pearson 카이제곱입니다.

상대 관성은 전체 관성에 대한 점의 기여 비율입니다. 요약 통계량 테이블의 "관성" 열에 상대 관성이 나열됩니다.

관성에 대한 부분 기여도

차원의 관성에 대한 행 또는 열의 기여도는 다음과 같이 계산됩니다 .

기여도 = (질량)(좌표)2/(차원 관성)

Cochran Q 통계량

Cochran *Q* 검정 통계량은 다음과 같이 계산됩니다 .

$$Q = k(k-1)\frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{b} X_{ij} - \frac{N}{k}\right)^2}{\sum_{i=1}^{b}\left(\sum_{j=1}^{k} X_{ij}\left(k - \sum_{j=1}^{k} X_{ij}\right)\right)}$$

다음은 각 요소에 대한 설명입니다 .

k = 처리 수

b = 블록 수

X_{ij} = k 번째 처리에 대한 i 번째 개체의 반응 (0 또는 1)

N = 모든 개체와 처리에 대한 양의 반응의 총 합계

8 장

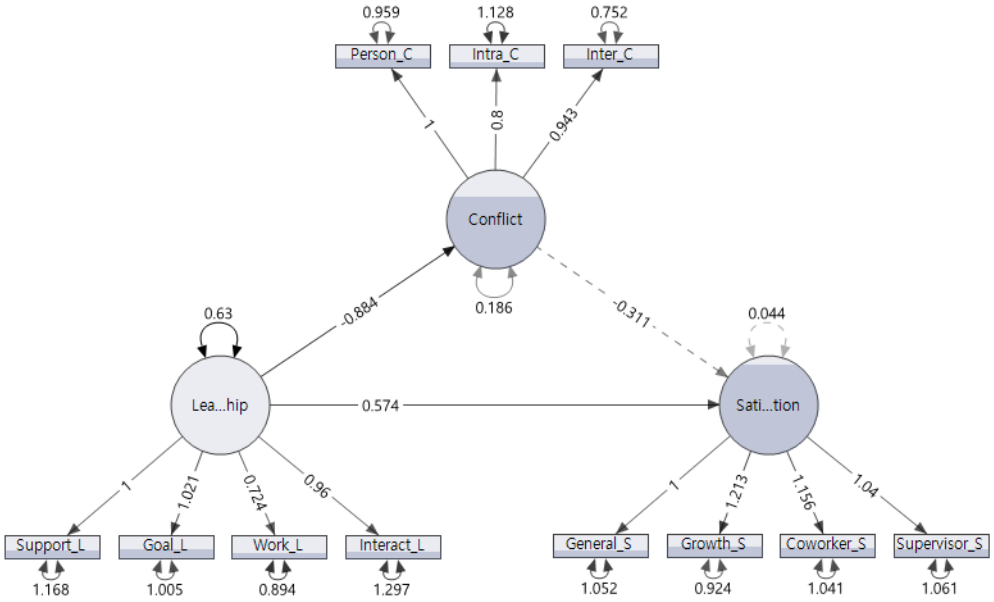
JMP PRO 구조 방정식 모형

구조 방정식 모형 적합

구조 방정식 모형 플랫폼은 JMP Pro 에서만 사용할 수 있습니다.

구조 방정식 모형 플랫폼을 사용하면 다양한 모형을 적합시킬 수 있습니다. 이러한 모형에는 확증적 요인 분석, 잠재 변수를 사용하거나 사용하지 않는 경로 모형, 측정 오차 모형 및 잠재 성장 곡선 모형이 포함됩니다. **시작점** 및 **끝점** 변수를 선택하고 두 변수의 연결 방법을 결정하여 모형을 지정할 수 있습니다. 연결은 다이어그램에서 단방향 또는 양방향 화살표로 표시됩니다. 이러한 선택 항목으로 모형의 두 가지 대체 보기(동적 경로 다이어그램, 모형 엣지 목록)를 채웁니다. 또한 이 플랫폼에서는 모형을 지정할 때 모형 식별을 확인합니다.

그림 8.1 구조 방정식 모형 경로 다이어그램



목차

구조 방정식 모형 개요 165
구조 방정식 모형의 예 167
구조 방정식 모형 플랫폼 시작 171
구조 방정식 모형 보고서 172
모형 규격 보고서 173
모형 비교 보고서 178
카이제곱 차이 검정 보고서 179
구조 방정식 모형 적합 보고서 179
구조 방정식 모형 플랫폼 옵션 181
모형 옵션 182
SEM 경로 다이어그램 사용자 정의 186
경로 다이어그램 팝업 메뉴 옵션 186
다이어그램 모양 사용자 정의 옵션 188
구조 방정식 모형의 추가 예 189
잠재 경로 변수 모형의 예 189
잠재 성장 곡선 모형의 예 196
측정 모형 평가 보고서 사용의 예 198
구조 방정식 모형 플랫폼에 대한 통계 상세 정보 201
적합 요약 측도에 대한 통계 상세 정보 201

JMP PRO 구조 방정식 모형 개요

SEM(구조 방정식 모형) 플랫폼을 사용하면 변수 간의 관계 이론을 검정하는 데 사용할 수 있는 다양한 모형을 적합시킬 수 있습니다. 모형의 변수는 관측된 **명시** 변수이거나 관측되지 않은 **잠재** 변수일 수 있습니다. 구조 방정식 모델링은 사회 과학 및 행동 과학 분야에서 많이 사용됩니다.

기본적으로 이 플랫폼에서는 모든 변수에 대한 평균과 분산이 있는 모형을 지정합니다. 그런 다음 모형이 생성되는 동안 여러 보기를 확인할 수 있는 모형 생성 인터페이스를 제공합니다. 또한 모형 생성 과정 중에 몇 가지 모형 상세 정보를 제공하여 모형을 실행하기 전에 유지할 수 없는 모형에 대해 알려 줍니다.

하나 이상의 모형을 적합시킨 후 "모형 비교" 보고서에서 적합 모형과 두 개의 기준 모형을 비교할 수 있습니다. 기준 모형은 제약 없는 모형과 독립 모형입니다. 제약 없는 모형은 데이터에 구조를 적용하지 않고 지정된 모형 변수의 모든 평균, 분산 및 공분산을 적합시키는 완전 포화 모형입니다. 독립 모형은 지정된 모형 변수의 모든 평균과 분산을 적합시킵니다. 지정된 모형 변수에서 모든 공분산이 0으로 고정되므로 매우 제한적인 모형이 됩니다.

SEM 플랫폼에서는 전체 정보 최대 가능도(Finkbeiner 1979) 방법을 사용합니다. 이렇게 하면 랜덤 결측값이 있는 관측값 비율이 높은 경우에도 데이터에서 사용 가능한 모든 정보를 완전히 사용할 수 있습니다.

구조 방정식 모델링에 대한 자세한 내용은 SAS Institute Inc. (2020a)의 "CALIS Procedure" 장, Bollen(1989) 및 Kline(2016) 연구 자료에서 확인하십시오.

참고 : 구조 방정식 모형 플랫폼의 모든 모형은 평균 구조로 추정되므로 Constant 항이 포함됩니다 . 관측 변수의 평균에 구조를 적용하지 않으려면 기본 모형 규격에서와 같이 평균이 자유롭게 추정되어야 합니다 .

모형 유형

이 섹션에서는 구조 방정식 모형 플랫폼에서 적합시킬 수 있는 다양한 모형 유형 중 일부에 대해 설명합니다 .

- **경로 분석**을 사용하면 관측 변수 간의 연관성에 대한 대체 설명 모형을 검정할 수 있습니다 . 이 기법은 연구에서 관심 구성당 하나의 변수만 사용할 수 있을 때 주로 사용됩니다 . 대개 가장 단순한 경로 분석 모형은 X 가 Y 를 예측하는 표준 회귀 모형입니다 . SEM 플랫폼을 사용하면 이 단순 회귀 모형을 적합시킬 수 있지만 더 흥미로운 모형을 지정할 수도 있습니다 . 예를 들어 이론이나 이전 조사에 기초하여 $X \Rightarrow Y$ 관계의 중재 변수로 간주되는 Z 변수가 있을 수 있습니다 . 즉 , X 가 Z 를 예측한 후 Y 를 예측합니다 . 따라서 원래 $X \Rightarrow Y$ 관계는 Z 가 원래 모형에서 제외되어야만 존재할 수 있습니다 . SEM 플랫폼에서 다음 단계에 따라 경로 분석을 수행할 수 있습니다 .
  1. 시작 창에서 모든 관측 변수를 선택하고 "모형 변수"를 클릭한 후 "확인"을 클릭합니다 .
  2. "모형 규격" 보고서의 "시작점 목록" 에서 예측 변수를 선택하고 "끝점 목록" 에서 해당하는 결과를 선택한 후 단방향 화살표 버튼을 클릭합니다 .

참고 : 상관계수가 0 이라는 가설을 검정하는 경우 외에는 모든 외생 변수 (단방향 화살표가 가리키지 않는 변수) 가 모형에서 자유롭게 상관되어야 합니다 . 이러한 공분산은 양방향 화살표 버튼으로 지정됩니다 .

- CFA(확증적 요인 분석) 를 사용하면 대체 측정 모형을 검정할 수 있습니다 . CFA 는 조사 개발에서 주로 사용되며 구조 회귀 모형을 적합시키기 전의 초기 단계로 사용됩니다 . SEM 플랫폼을 사용하면 다음 단계를 수행하여 확증적 요인 분석 모형을 적합시킬 수 있습니다 .

  1. 시작 창에서 모든 관측 변수를 선택하고 " 모형 변수 " 를 클릭한 후 " 확인 " 을 클릭합니다 .
  2. " 모형 규격 " 의 " 끝점 목록 " 을 사용하여 잠재 변수에 적재될 것으로 간주되는 변수를 선택합니다 .
  3. " 끝점 목록 " 아래의 상자에 잠재 변수 이름을 입력하고 잠재 변수 추가 버튼 ⊞을 클릭합니다 .
  4. 모형에 대한 모든 잠재 변수가 지정될 때까지 이 과정을 반복합니다 .

  SEM 플랫폼에는 항상 평균 구조가 포함되므로 모든 관측 변수가 " 평균 / 절편 " 목록에 Constant 항 결과로 나열됩니다 . 또한 모든 잠재 변수는 첫 번째 표시자의 적재를 1 로 설정하거나 (기본값), 시작 창에서 " 잠재 변수 표준화 " 옵션을 선택한 경우 분산을 1 로 설정하여 자동으로 식별됩니다 . 마지막으로 기존의 CFA 모형에서는 모든 잠재 변수가 공변할 수 있습니다 . " 시작점 목록 " 과 " 끝점 목록 " 에서 모든 잠재 변수를 선택한 후 양방향 화살표 버튼을 클릭하여 이러한 공분산을 지정할 수 있습니다 .
- SR(구조 회귀) 모형은 잠재 변수가 있는 경로 분석이라고도 합니다 . 이러한 모형은 주로 CFA(확증적 요인 분석) 를 통해 적절한 측정 모형을 식별한 후에 사용됩니다 . SR 모형을 사용하면 잠재 변수 간의 특정 관계 패턴을 검정할 수 있습니다 . 즉 , CFA 는 잠재 변수 사이의 효과에 방향성을 적용하지 않지만 (모든 잠재 변수가 자유롭게 공변할 수 있음) SR 모형은 방향성을 적용합니다 . 경영진의 리더십 (Leadership) 이 직장 내 팀 갈등 (Conflict) 을 줄이고 직원 만족도 (Satisfaction) 를 높일 수 있다고 가정한 예에서 Leadership 잠재 변수는 Conflict 및 Satisfaction 잠재 변수를 예측할 수 있습니다 . 다음 단계에 따라 CFA 를 수행한 후 이러한 방향 효과 (회귀) 를 지정할 수 있습니다 .

  1. 시작 창에서 모든 관측 변수를 선택하고 " 모형 변수 " 를 클릭한 후 " 확인 " 을 클릭합니다 .
  2. " 모형 규격 " 의 " 끝점 목록 " 을 사용하여 잠재 변수에 적재될 것으로 간주되는 변수를 선택합니다 .
  3. " 끝점 목록 " 아래의 상자 옆에 잠재 변수 이름을 입력하고 잠재 변수 추가 버튼 ⊞을 클릭합니다 .
  4. 모형에 대한 모든 잠재 변수가 지정될 때까지 이 과정을 반복합니다 .
  5. " 모형 규격 " 보고서의 " 시작점 목록 " 에서 예측 변수를 선택하고 " 끝점 목록 " 에서 해당하는 결과를 선택한 후 단방향 화살표 버튼을 클릭합니다 .
- LGC(잠재 성장 곡선) 모형을 사용하면 반복 측정 데이터에 대체 잠재 궤적을 적합시키고 검정할 수 있습니다 . 이러한 모형은 혼합 모형 프레임워크의 임의 효과 모형과 매우 유사합니다 . 무성장 모형을 선형 모형과 비교하려는 경우가 종종 있습니다 . 무성장 모형에서는 개체의 시작점이 다를 수 있지만 궤적이 평평합니다 . 선형 모형에서는 개체의 시작점과 시간에

따른 선형 기울기가 모두 다를 수 있습니다 . 사용 가능한 데이터가 충분한 경우에는 이러한 모형을 개체의 시작점과 시간에 따라 선형 및 2 차 비율이 달라질 수 있는 2 차 모형과 비교할 수도 있습니다 . SEM 플랫폼 모형 규격을 사용하여 LGC 모형을 적합시킬 수 있지만 플랫폼에서 다음 단계를 수행하여 LGC 모형 적합을 간소화합니다 .

1. 시작 창에서 모든 관측 변수 (반복 측정) 를 선택하고 " 모형 변수 " 를 클릭한 후 " 확인 " 을 클릭합니다 .

   참고 : 잠재 성장 모형 바로 가기에서 모형을 올바르게 지정하려면 관측 변수가 시간 오름차순으로 나열되고 시간 간격이 동일해야 합니다 .

2. " 모형 바로 가기 " 옵션을 사용하여 종단 분석 > 절편만 있는 잠재 성장 곡선 모형을 선택한 후 " 실행 " 을 클릭합니다 .
3. " 모형 바로 가기 " 옵션을 사용하여 종단 분석 > 선형 잠재 성장 곡선 모형을 선택한 후 " 실행 " 을 클릭합니다 .
4. " 모형 바로 가기 " 옵션을 사용하여 종단 분석 > 2 차 잠재 성장 곡선 모형을 선택한 후 " 실행 " 을 클릭합니다 .

" 모형 비교 " 테이블에 대체 적합 지수가 표시되고 최적 모형을 선택할 수 있습니다 .

- **조건부 잠재 성장 곡선** 모형은 위의 단계에 따라 이상적인 성장 궤적을 식별한 후 사용할 수 있습니다 . 이때 변경 요인 및 절편의 예측 변수를 모형에 추가할 수 있습니다 . 이러한 예측 변수는 성장 과정의 초기 스코어와 이후의 변화를 결정하는 데 중요한 요인일 수 있습니다 . 조건부 LGC 를 적합시키려면 시작 창에서 잠재 변수의 가설 예측 변수를 포함하여 모든 관측 변수 (반복 측정) 를 선택합니다 . 다음 단계를 쉽게 수행하려면 예측 변수가 " 모형 변수 " 목록의 마지막 변수여야 합니다 .

  1. " 모형 바로 가기 " 옵션을 사용하여 적절한 성장 궤적을 선택합니다 . 이 옵션은 예측 변수를 포함하여 LGC 모형의 모든 변수를 지정합니다 . 따라서 예측 변수를 성장 과정에서 제외하고 예측 변수로 올바르게 지정해야 합니다 .
  2. " 적재 " 목록에서 예측 변수를 찾고 해당 변수를 포함하는 모든 효과를 선택한 후 " 제거 " 를 클릭합니다 .
  3. " 시작점 목록 " 에서 예측 변수를 선택하고 " 끝점 목록 " 에서 " 절편 " 또는 " 기울기 " 를 선택합니다 .
  4. 단방향 화살표를 클릭하여 조건부 LGC 를 지정합니다 .

  참고 : 예측 변수가 두 개 이상인 경우 " 시작점 목록 " 과 " 끝점 목록 " 에서 예측 변수를 선택하고 양방향 화살표 버튼을 클릭하여 공분산을 지정해야 합니다 .

JMP PRO 구조 방정식 모형의 예

예를 들어 인사과 직원이 주요 직장 구성 요인을 측정하기 위한 설문 조사를 개발하려고 합니다 . 이 예에서는 구조 방정식 모형 플랫폼을 사용하여 200 명을 대상으로 실시한 직장 관련 설문 조

사의 응답을 분석합니다 . 설문 조사는 직무 만족도와 관련된 11 개 질문에 대한 응답을 포함합니다 . 설문 조사 질문에 대한 답변을 리더십 특성 , 역할 갈등 및 전반적인 직무 만족도의 잠재 변수와 연관시키는 확증적 요인 분석 모형을 생성하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Job Satisfaction.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 구조 방정식 모형**을 선택합니다 .
3. Support_L ~ Supervisor_S 를 선택하고 **모형 변수**를 클릭합니다 .
4. **확인**을 클릭합니다 .

   " 구조 방정식 모형 " 보고서의 " 모형 규격 " 개요가 나타납니다 .

잠재 변수 생성

5. "끝점 목록"에서 Support_L ~ Interact_L을 선택하고 "끝점 목록" 아래의 상자에 "Leadership"을 입력한 후 잠재 변수 추가 버튼 ⊞을 클릭합니다 .
6. " 끝점 목록 " 에서 Person_C ~ Inter_C 를 선택하고 " 끝점 목록 " 아래의 상자에 "Conflict" 를 입력한 후 잠재 변수 추가 버튼 ⊞을 클릭합니다 .
7. "끝점 목록"에서 General_S ~ Supervisor_S를 선택하고 "끝점 목록" 아래의 상자에 "Satisfaction"을 입력한 후 잠재 변수 추가 버튼 ⊞을 클릭합니다 .

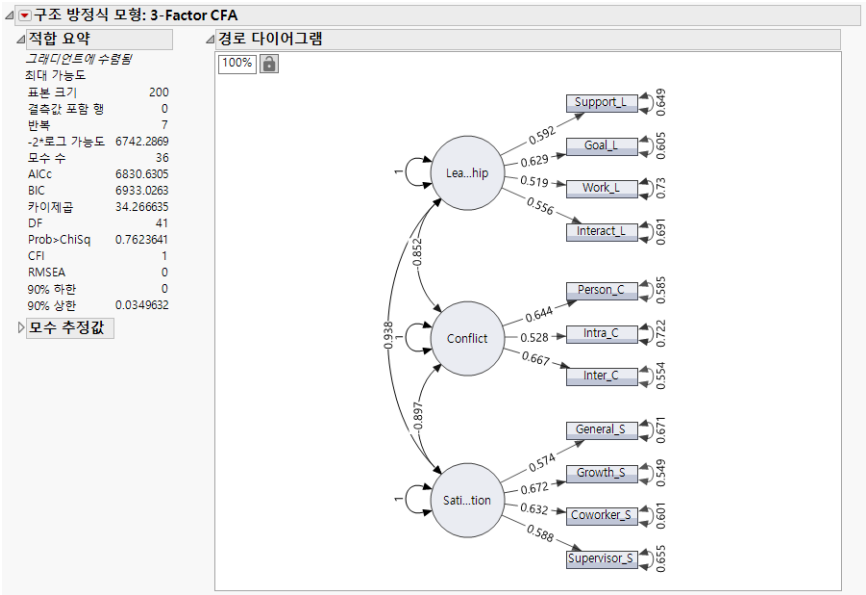
공분산 추가 및 모형 적합

8. "시작점 목록"에서 Leadership, Conflict, Satisfaction을 선택하고 "끝점 목록"에서 Leadership, Conflict, Satisfaction 을 선택한 후 양방향 화살표 버튼 ↔을 클릭합니다 .
9. " 모형 규격 " 보고서의 왼쪽 위에 있는 " 모형 이름 " 아래 텍스트 상자에 "3-Factor CFA" 를 입력합니다 .
10. **실행**을 클릭합니다 .
11. (선택 사항) " 구조 방정식 모형 : 잠재 변수를 사용한 경로 분석 " 옆의 빨간색 삼각형을 클릭하고 **경로 다이어그램 설정 > 레이아웃 > 위쪽에서 아래쪽으로**를 선택합니다 .
12. (선택 사항) " 모수 추정값 " 옆의 회색 표시 아이콘을 클릭합니다 .

    " 모수 추정값 " 보고서를 닫으면 전체 경로 다이어그램을 볼 수 있습니다 .
13. 경로 다이어그램을 마우스 오른쪽 버튼으로 클릭하고 **표시 > 추정값 표시 > 표준화**를 선택합니다 .

    이제 " 경로 다이어그램 " 의 숫자는 모형에 대한 표준화 모수 추정값을 나타냅니다 .

그림 8.2 구조 방정식 모형 보고서

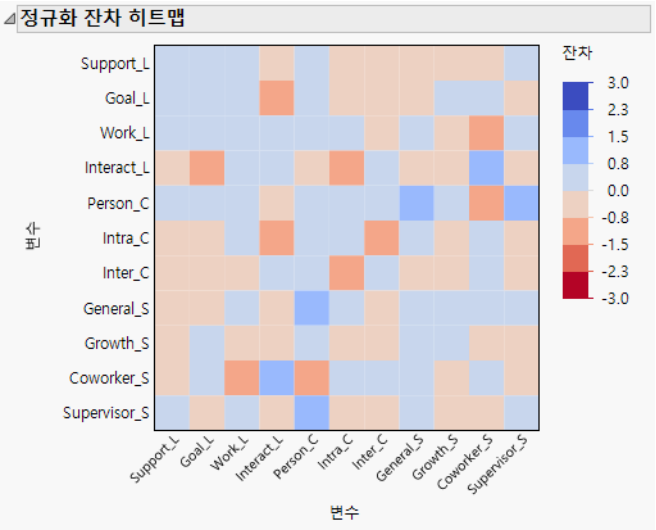


" 적합 요약 " 보고서에 나열된 이 모형의 카이제곱 통계량은 34.27 이고 자유도는 41 입니다 . 해당 p 값은 0.7624 로 , 유의하지 않습니다 . 이는 모형이 잘 적합된다는 귀무가설을 기각할 증거가 없음을 나타냅니다 . 따라서 이 모형이 데이터를 적절하게 적합시킨다는 결론을 내릴 수 있습니다 .

카이제곱 값은 표본 크기에 따라 달라지므로 일부 적합 모형에서 유의한 카이제곱 값을 생성할 수 있습니다 . CFI(비교 적합 지수) 와 RMSEA(근사의 제곱근 평균 제곱 오차) 는 모형 적합을 결정하는 추가 지표를 제공합니다 . 이러한 지수는 0 과 1 사이로 제한됩니다 . CFI 값이 0.90 보다 크고 RMSEA 값이 0.10 보다 작으면 좋습니다 (Browne 과 Cudeck 1993, Hu 와 Bentler 1999 참조). 여기서 CFI 1 과 RMSEA 0 은 적합도가 탁월하다는 것을 나타냅니다 . 이 설문 조사는 리더십 , 갈등 및 만족도에 대한 잠재 변수를 측정하는 데 좋은 도구라는 결론을 내릴 수 있습니다 .

14. " 구조 방정식 모형 : 3-Factor CFA" 옆의 빨간색 삼각형을 클릭하고 **정규화 잔차 히트맵**을 선택합니다 .

그림 8.3 정규화 잔차 히트맵

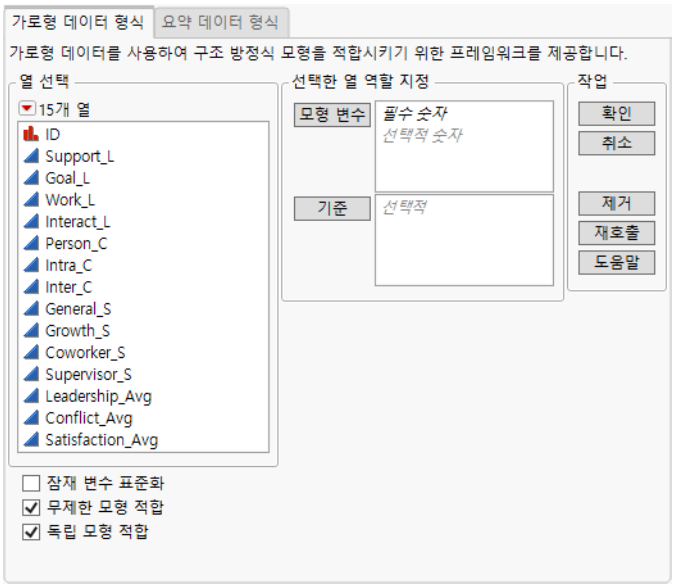


" 정규화 잔차 히트맵 " 에 양의 방향 또는 음의 방향으로 2.0 단위를 초과하는 값이 없습니다 . 이는 모형이 데이터를 잘 적합시킨다는 추가적인 증거입니다 . 잔차는 더 세분화된 수준에서 부적합 모형을 진단할 수도 있습니다 . 이 모형의 정규화 잔차는 국소적 부적합 증거를 나타내지 않습니다 .

JMP PRO 구조 방정식 모형 플랫폼 시작

분석 > 다변량 방법 > 구조 방정식 모형을 선택하여 구조 방정식 모형 플랫폼을 시작합니다.

그림 8.4 구조 방정식 모형 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

시작 창에는 두 가지 데이터 형식에 대한 탭이 있습니다.

가로형 데이터 형식 데이터 테이블의 각 행이 단일 관측값에 해당하고 열에 모형에 사용할 변수가 포함된 경우에 선택합니다. 결측값만 포함된 행은 분석에서 제외됩니다.

요약 데이터 형식 데이터 테이블의 데이터가 상관 행렬 또는 공분산 행렬로 요약된 경우에 선택합니다. 평균과 표준편차도 열로 지정할 수 있습니다. 이러한 값을 지정하지 않으면 평균은 0 으로 간주되고 표준편차는 입력 행렬 대각선의 제곱근으로 간주됩니다.

JMP PRO 시작 창 옵션

모형 변수 모형에 포함할 열입니다. 하나 이상의 열을 지정해야 합니다. 모든 열에는 숫자 데이터 유형과 연속형 모델링 유형이 있어야 합니다.

요약 데이터 형식의 경우 모형 변수 열은 요약 데이터에 대한 상관 행렬 또는 공분산 행렬을 포함하는 열입니다.

기준 (" 가로형 데이터 형식 " 에만 사용 가능) 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열입니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

주의 : 기준 변수를 사용하면 다중 그룹 분석이 수행되지 않습니다 .

평균 (" 요약 데이터 형식 " 에만 사용 가능) 상관 행렬 또는 공분산 행렬의 각 행에 있는 변수에 해당하는 평균 열입니다 . 지정하지 않으면 평균이 0 으로 간주됩니다 .

표준편차 (" 요약 데이터 형식 " 에만 사용 가능) 상관 행렬 또는 공분산 행렬의 각 행에 있는 변수에 해당하는 표준편차 열입니다 . 지정하지 않으면 표준편차가 입력 행렬 대각선의 제곱근으로 간주됩니다 .

라벨 (" 요약 데이터 형식 " 에만 사용 가능) 상관 행렬 또는 공분산 행렬의 각 행에 있는 변수에 해당하는 라벨 열입니다 . 지정하지 않으면 입력 행렬이 포함된 열의 이름이 " 모형 규격 " 의 변수에 사용됩니다 .

잠재 변수 표준화 이 옵션을 선택하면 분산을 1 로 고정하고 모든 적재를 자유롭게 추정할 수 있도록 하여 잠재 변수의 척도를 설정합니다 .

제약 없는 모형 적합 이 옵션을 선택하면 첫 번째 모형을 적합시킬 때 제약 없는 모형이 자동으로 적합됩니다 . 그런 다음 " 모형 비교 " 보고서에서 적합 모형을 제약 없는 모형과 비교할 수 있습니다 . 제약 없는 모형은 데이터에 구조를 적용하지 않고 지정된 모형 변수의 모든 평균 , 분산 및 공분산을 적합시키는 완전 포화 모형입니다 .

독립 모형 적합 이 옵션을 선택하면 첫 번째 모형을 적합시킬 때 독립 모형이 자동으로 적합됩니다 . 그런 다음 " 모형 비교 " 보고서에서 적합 모형을 독립 모형과 비교할 수 있습니다 . 독립 모형은 지정된 모형 변수의 모든 평균과 분산을 적합시킵니다 . 지정된 모형 변수에서 모든 공분산이 0 으로 고정되므로 매우 제한적인 모형이 됩니다 .

표본 크기 (" 요약 데이터 형식 " 에만 사용 가능) 요약 데이터로 나타내는 관측값 수를 지정합니다 .

JMP PRO 구조 방정식 모형 보고서

- " 모형 규격 보고서 "
- " 모형 비교 보고서 "
- " 카이제곱 차이 검정 보고서 "
- " 구조 방정식 모형 적합 보고서 "

JMP PRO 모형 규격 보고서

" 모형 규격 " 보고서에는 모형을 지정하기 위한 컨트롤이 포함되어 있습니다 . 시작 창에서 " 확인 " 을 클릭하면 " 모형 규격 " 보고서의 " 다이어그램 " 및 " 목록 " 탭에 기본 독립 모형이 나타납니다 . 독립 모형은 지정된 모형 변수의 모든 평균과 분산을 포함합니다 .

JMP PRO 규격 패널

" 규격 " 패널에는 모형을 생성하기 위한 컨트롤이 있습니다 .

모형 이름 모형 이름을 지정할 수 있습니다 . " 작업 " 패널에서 " 실행 " 을 클릭하면 " 모형 이름 " 의 텍스트를 사용하는 모형 보고서가 생성됩니다 .

시작점 목록 시작 창에서 지정한 모형 변수와 Constant 항을 나열합니다 . Constant 항은 상수에서 각 변수를 회귀하여 관측 변수 또는 잠재 변수의 평균을 추정합니다 . 이 목록에서 하나 이상의 변수를 선택하고 " 끝점 목록 " 에서 하나 이상의 변수를 선택한 후 화살표 버튼 중 하나를 클릭하면 모형에 항이 추가됩니다 . 잠재 변수를 생성하면 " 시작점 목록 " 및 " 끝점 목록 " 에 추가됩니다 . 목록 맨 위에 있는 검색 필터 컨트롤을 사용하여 이 목록의 항목을 필터링할 수 있습니다 . 자세한 내용은 " 검색 필터 옵션 " 에서 확인하십시오 .

끝점 목록 시작 창에서 지정한 모형 변수와 Constant 항을 나열합니다 . Constant 항은 상수에서 각 변수를 회귀하여 관측 변수 또는 잠재 변수의 평균을 추정합니다 . " 시작점 목록 " 에서 하나 이상의 변수를 선택하고 이 목록에서 하나 이상의 변수를 선택한 후 화살표 버튼 중 하나를 클릭하면 모형에 항이 추가됩니다 . 잠재 변수를 생성하면 " 시작점 목록 " 및 " 끝점 목록 " 에 추가됩니다 . 모형에 잠재 변수를 추가하려면 " 끝점 목록 " 의 변수를 선택해야 합니다 . 목록 맨 위에 있는 검색 필터 컨트롤을 사용하여 이 목록의 항목을 필터링할 수 있습니다 . 자세한 내용은 " 검색 필터 옵션 " 에서 확인하십시오 .

단방향 화살표 " 시작점 목록 " 과 " 끝점 목록 " 에서 선택한 변수 간의 관계 유형을 지정합니다 . 단방향 화살표는 회귀 효과와 동일합니다 .

양방향 화살표 " 시작점 목록 " 과 " 끝점 목록 " 에서 선택한 변수 간의 관계 유형을 지정합니다 . 양방향 화살표는 공분산 효과와 동일합니다 .

잠재 변수 추가 " 시작점 목록 " 과 " 끝점 목록 " 에 잠재 변수를 추가합니다 . 변수 이름은 버튼 왼쪽에 있는 상자의 텍스트를 기반으로 합니다 .

참고 : 버튼을 클릭하기 전에 " 끝점 목록 " 에서 잠재 변수에 대한 표시자를 선택해야 합니다 .

잠재 변수 제거 " 시작점 목록 " 또는 " 끝점 목록 " 에서 현재 선택된 잠재 변수를 제거합니다 . "시작점 목록"과 "끝점 목록"에서 잠재 변수를 제거하면 모형에서도 잠재 변수가 제거됩니다.

모형 바로 가기 공통 모형을 빠르게 생성하기 위한 세 가지 범주의 옵션이 포함되어 있습니다 . 이러한 옵션 중 하나를 선택하면 현재 모형 규격이 모두 지워집니다 .

규격 지원 외생 변수 또는 내생 변수에 대한 공분산 추가 과정을 단축하는 옵션이 포함되어 있습니다 .

횡단면 (전통적) 확증적 요인 분석 및 중재 분석을 위한 옵션이 포함되어 있습니다 .

종단 분석 절편만 있는 잠재 성장 곡선 , 선형 잠재 성장 곡선 , 2 차 잠재 성장 곡선 및 잠재 변수 기반 성장 곡선과 같은 공통 잠재 성장 곡선 모형에 대한 옵션이 포함되어 있습니다 . 또한 " 성장 모형 적합 및 비교 " 옵션을 사용하면 한 번 클릭으로 무성장 모형 , 선형 성장 모형 및 2 차 성장 모형을 적합시킬 수 있습니다 .

참고 : 잠재 성장 곡선 옵션 중 하나를 선택할 때 " 끝점 목록 " 의 변수를 선택하면 선택한 변수에만 모형이 적용됩니다 . 그렇지 않으면 " 끝점 목록 " 의 모든 변수에 모형이 적용됩니다 .

JMP PRO 작업 패널

" 작업 " 패널의 버튼은 " 목록 " 탭의 지정된 모형 변수 목록에서 작동합니다 . 다음과 같은 작업을 수행할 수 있습니다 .

모수 고정 현재 선택된 효과의 모수 값을 상수 값으로 고정할 수 있습니다 . 변수의 모수 값을 고정하면 항 이름 뒤의 괄호 안에 고정 값이 표시됩니다 .

참고 : 기본값 0 을 유지하고 " 확인 " 을 클릭하면 선택한 효과가 모형에서 제거됩니다 .

등가 설정 두 개 이상 선택한 효과의 모수 값이 같도록 제한할 수 있습니다 . 둘 이상의 효과에 대한 모수 값이 같도록 설정하면 항 이름 뒤의 괄호 안에 임의의 영숫자 라벨 ("c1") 이 나타납니다 . 모수 값이 서로 같도록 설정된 효과 집합이 여러 개인 경우 번호가 순차적으로 지정됩니다 ("c1", "c2" 등).

참고 : 동등 제약 조건은 유형이 동일한 모수 내에서만 허용됩니다 .

제약 해제 모형에서 선택한 효과에 적용된 제한을 제거할 수 있습니다 . 제한에는 고정 값으로 설정된 효과 또는 다른 효과와 동일하게 설정된 효과가 포함됩니다 .

제거 선택한 효과를 모형에서 제거합니다 .

실행 취소 모형에 대한 마지막 수정 작업을 실행 취소합니다 .

다시 실행 모형에 대해 마지막으로 실행 취소한 수정 작업을 다시 실행합니다 .

실행 현재 지정된 모형을 적합시키고 보고서를 생성합니다 .

재설정 모형 규격을 독립 모형 (기본값) 으로 재설정합니다 .

JMP PRO 다이어그램 패널

"다이어그램" 패널에는 모형 다이어그램 레이아웃을 조정할 수 있는 버튼이 포함되어 있습니다. 첫 번째 버튼을 사용하면 다이어그램에서 직사각형으로 표시된 명시 변수를 회전할 수 있습니다. 두 번째 버튼을 사용하면 다이어그램의 두 가지 다른 배열을 순환할 수 있습니다. 세 번째 버

튼을 사용하여 다이어그램을 사용자 정의할 수 있습니다. 자세한 내용은 "SEM 경로 다이어그램 사용자 정의"에서 확인하십시오.

JMP PRO 상세 정보 패널

"상세 정보" 패널에는 현재 "모형 규격" 보고서에 지정된 모형에 대한 정보가 포함되어 있습니다. 모형에 대한 명시 변수의 수, 잠재 변수의 수, 자유 추정 모수의 수 및 자유도 정보를 제공합니다. 최대 반복 수를 조정할 수도 있습니다.

JMP PRO 다이어그램 탭

"다이어그램" 탭에는 현재 지정된 모형을 시각화할 수 있는 모형 다이어그램이 포함됩니다. 잠재 변수는 원으로 표시되고 명시 변수는 직사각형으로 표시됩니다. 단방향 화살표는 적재 및 회귀를 나타냅니다. 양방향 화살표는 분산 및 공분산을 나타냅니다. 분산은 해당 변수 자신에게 돌아오는 양방향 화살표로 지정됩니다. 공분산은 두 변수 사이의 양방향 화살표로 지정됩니다.

경로 다이어그램은 대화식이므로 항목을 드래그하여 배열할 수 있습니다. 화살표 키를 사용하여 경로 다이어그램에서 선택한 항목을 이동할 수도 있습니다. "다이어그램" 패널의 버튼을 사용하여 경로 다이어그램의 항목을 재배열할 수 있습니다. 경로 다이어그램 팝업 메뉴의 옵션을 사용하여 경로 다이어그램에서 항목을 추가하거나 제거할 수 있습니다. 경로 다이어그램 자체를 마우스 오른쪽 버튼으로 클릭하면 경로 다이어그램 팝업 메뉴가 나타납니다. 팝업 메뉴 내용은 빈 캔버스 부분을 클릭했는지 아니면 경로 다이어그램 요소를 클릭했는지에 따라 다릅니다. 경로 다이어그램 왼쪽 위에 있는 확대/축소 설정을 조정하여 경로 다이어그램을 확대하거나 축소할 수도 있습니다. 또한 경로 다이어그램에는 경로 다이어그램에서 항목 위치를 잠글 수 있는 "잠금" 버튼이 왼쪽 위에 있습니다. 경로 다이어그램이 잠기면 "잠금" 버튼이 파란색입니다. 경로 다이어그램에서 선택한 후 팝업 메뉴에서 **선택 > 선택 항목 숨기기** 옵션을 선택하여 해당 항목을 숨길 수 있습니다. 경로 다이어그램의 항목이 숨겨지면 숨겨진 모든 항목을 표시할 수 있는 "모두 숨기기 해제" 버튼이 경로 다이어그램 맨 위에 나타납니다.

경로 다이어그램 사용자 정의 옵션에 대한 자세한 내용은 "SEM 경로 다이어그램 사용자 정의"에서 확인하십시오. 자세한 내용은 "다이어그램 패널", "상세 정보 패널" 및 "경로 다이어그램 설정"에서 확인하십시오.

팁 : 경로 다이어그램을 이미지로 복사하려면 다이어그램을 마우스 오른쪽 버튼으로 클릭하고 **다이어그램 복사**를 선택합니다. 최상의 품질을 유지하려면 클립보드 이미지를 벡터 그래픽으로 붙여 넣습니다.

JMP PRO 목록 탭

"목록" 탭에는 모형의 각 변수 유형에 대한 목록이 포함됩니다. 이러한 목록은 모형 다이어그램에 나타내는 데 사용되는 화살표로 분류됩니다. 단방향 화살표는 평균이나 절편, 적재 및 회귀를 지정하는 데 사용됩니다. 양방향 화살표는 분산 및 공분산을 지정하는 데 사용됩니다. 모형에 대한 평균과 절편은 "목록" 탭에만 나타나고 "다이어그램" 탭에는 나타나지 않습니다.

각 목록 맨 위에 있는 검색 필터 컨트롤을 사용하여 각 목록의 항목을 필터링할 수 있습니다. 자세한 내용은 "검색 필터 옵션"에서 확인하십시오.

JMP PRO 상태 탭

"상태" 탭에는 모형 식별성 확인이 포함됩니다. 신뢰할 수 있는 추정값을 얻으려면 구조 방정식 모형을 식별해야 합니다. SEM 컨텍스트에서 식별은 입력 데이터의 모집단 공분산 행렬을 기반으로 모형의 모든 모수에 대한 고유 해를 도출할 수 있음을 의미합니다. SEM 프레임워크에서 다양한 모형을 지정할 수 있으므로 식별 규칙도 다양합니다. "상태" 탭은 "식별 규칙", "모형 상세 정보" 및 "데이터 상세 정보"라는 세 가지 정보 패널로 구성됩니다.

- " 식별 규칙 " 패널에는 지정된 모형에 적합한 규칙 (최대 8 개) 목록이 포함됩니다 . 모형 식별을 보장하기 위한 필요 충분 규칙이 있습니다 . 필요 규칙이 실패할 경우 모형을 적합시키기 전에 규칙을 수정하는 단계를 수행해야 합니다 . 충분 규칙이 실패할 경우 문제를 반드시 해결할 필요는 없습니다 . 경우에 따라 충분 규칙이 모형에 문제가 있다는 증거 없이 실패할 수 있습니다 . 모든 규칙은 양정치 공분산 행렬을 가정합니다 . 공분산 행렬이 양정치가 아니면 " 모형 상세 정보 " 패널 아래에 경고가 나타납니다 .

  팁 : 특정 식별 규칙에 대한 자세한 내용을 보려면 규칙 테이블의 행을 클릭하십시오 . 그러면 테이블 오른쪽에 규칙 설명이 나타납니다 .

- " 모형 상세 정보 " 패널에는 현재 지정된 모형을 기술하는 값 목록이 포함됩니다 . 이러한 값을 사용하여 모형의 잠재적 문제를 감지할 수 있습니다 .
- " 데이터 상세 정보 " 패널에는 입력 데이터를 기술하는 값 목록이 포함됩니다 . 이러한 값을 사용하여 데이터의 잠재적 문제를 감지할 수 있습니다 .

지정된 데이터 열에 특이성이 있으면 보고서에 "특이성 상세 정보" 테이블이 포함됩니다.

"상태" 탭 자체에는 지정된 모형의 현재 상태를 보여 주는 동적 아이콘이 있습니다. 아이콘은 다음과 같이 상태를 지정합니다 .

- 적용 가능한 모든 식별 규칙을 통과합니다 .
- 하나 이상의 필수 식별 규칙이 충족되지 않으며 모형을 적합시키기 전에 문제 해결 단계를 수행해야 합니다 .
- 필수가 아닌 식별 규칙 중 하나 이상이 충족되지 않으며 모형이 올바르게 지정되었는지 여부를 판별하기 위해 추가 조사가 필요합니다 . SEM 고급 응용 프로그램에서 이러한 상황이 종종 발생하지만 모형에 문제가 있다는 의미는 아닙니다 . 이는 식별 규칙이 모형 식별을 보장할 수 없음을 나타냅니다 .

그림 8.5 상태 탭의 예

JMP PRO 검색 필터 옵션

"모형 규격" 보고서의 "시작점 목록", "끝점 목록" 및 ""목록" 탭의 목록 상자에는 특정 목록의 항목을 필터링할 수 있는 검색 필터가 포함되어 있습니다.

검색 상자 옆의 아래쪽 화살표 버튼을 클릭하여 검색을 구체화할 수 있습니다.

용어 포함 검색 기준의 일부가 포함된 항목을 반환합니다. "ease oom" 을 검색하면 "Release Zoom" 과 같은 메시지가 반환됩니다.

구 포함 검색 기준이 정확히 포함된 항목을 반환합니다. "text box" 를 검색하면 "text" 와 "box" 가 바로 연이어 포함된 항목 (예 : "Context Box" 및 "Text Box") 이 반환됩니다.

구로 시작 검색 기준으로 시작하는 항목을 반환합니다.

구로 끝남 검색 기준으로 끝나는 항목을 반환합니다.

전체 구 전체 문자열로 구성된 항목을 반환됩니다 . "text box" 를 검색하면 "text box" 만 포함된 항목이 반환됩니다 .

정규 표현식 검색 상자에서 와일드카드 (*) 와 마침표 (.) 를 사용할 수 있습니다 . "get.*name" 을 검색하면 "get" 다음에 하나 이상의 단어가 포함된 항목을 찾을 수 있습니다 . 즉 , "Get Color Theme Names", "Get Name Info" 및 "Get Effect Names" 등이 반환됩니다 .

결과 반전 검색 기준과 매칭되지 않는 항목을 반환합니다 .

모든 용어 일치 문자열이 모두 포함된 항목을 반환합니다 . "t test" 를 검색하면 검색 문자열 중 하나 또는 둘 모두가 포함된 요소 (예 : "Pat Test", "Shortest Edit Script" 및 "Paired t test") 가 반환됩니다 .

대 / 소문자 무시 검색 기준의 대 / 소문자를 무시합니다 .

전체 단어 일치 " 모든 용어 일치 " 설정에 따라 문자열의 각 단어가 포함된 항목을 반환합니다 . " 모든 용어 일치 " 옵션이 선택되어 있는 경우 "data filter" 를 검색하면 "data" 와 "filter" 가 반환됩니다 .

JMP PRO 모형 비교 보고서

"모형 비교" 보고서에는 적합된 모든 모형을 포함하는 테이블이 포함됩니다. 두 번째 열의 아이콘을 사용하여 "모형 비교" 보고서 아래에 나타나는 모형 보고서를 제어할 수 있습니다. "모형 이름" 오른쪽의 나머지 열을 사용하면 다양한 기준에 따라 모형을 비교할 수 있습니다.

모형의 "AICc 가중치" 값은 적합 모형 중 하나가 사실일 때 특정 모형이 참모형일 확률로 해석할 수 있습니다. 따라서 AICc 가중치가 1 에 가장 가까운 모형이 최량 적합 모형입니다. AICc 가중치는 비결측 AICc 값만 사용하여 계산되고 다음과 같이 정의됩니다.

AICcWeight = exp[-0.5(AICc-min(AICc))] / sum(exp[-0.5(AICc-min(AICc))])

여기서 min(AICc) 는 적합 모형 중 가장 작은 AICc 값입니다 .

"모형 비교" 보고서의 다른 기준에 대한 자세한 내용은 "구조 방정식 모형 적합 보고서"에서 확인하십시오.

참고: 모형이 수렴되지 않으면 "모형 비교" 보고서의 모형 이름 시작 부분에 별표가 표시됩니다.

적합 모형 성능에 대한 컨텍스트를 제공하기 위해 다음 두 모형이 " 모형 비교 " 보고서에 기본적으로 표시됩니다 .

제약 없음 데이터에 구조를 적용하지 않고 지정된 모형 변수의 모든 평균 , 분산 및 공분산을 적합시킵니다 .

독립 지정된 모형 변수의 모든 평균과 분산을 적합시키고 모든 공분산을 0 으로 고정합니다 .

모형 비교 보고서 옵션

" 모형 비교 " 테이블 아래에 두 가지 옵션이 나타납니다 . 이러한 옵션은 " 모형 비교 " 테이블에서 행을 클릭하여 강조 표시한 후 사용할 수 있습니다 .

선택한 모형 비교 (테이블에서 두 개 이상의 행이 강조 표시된 경우에만 사용 가능) 선택한 행의 모든 내포 모형 조합 사이의 내포 카이제곱 차이 검정을 계산합니다 .

팁 : 선택한 모형이 내포되지 않은 경우 경고가 나타나고 내포되지 않은 모형 조합은 " 카이제곱 차이 검정 " 보고서에 표시되지 않습니다 .

선택 해제 (테이블에서 하나 이상의 행이 강조 표시된 경우에만 사용 가능) 테이블 행 선택을 모두 지웁니다 .

JMP PRO 카이제곱 차이 검정 보고서

"카이제곱 차이 검정 " 보고서에는 내포 카이제곱 검정 테이블이 포함됩니다 . 이 테이블에는 모형을 정의하는 두 열이 있습니다 . 첫 번째 열은 두 모형 중 더 제한된 모형을 포함하고 두 번째 열은 두 모형 중 덜 제한된 모형을 포함합니다 . 더 작은 모형이 더 큰 모형 안에 내포됩니다 . 나머지 열에는 카이제곱 값 , 자유도 , CFI 및 RMSEA 의 차이와 내포 카이제곱 검정의 p 값이 표시됩니다 . 열 이름의 Δ 표기는 차이를 나타냅니다 . 유의한 Δ 카이제곱 값은 내포 모형의 추가 제약 조건으로 인해 통계적으로 유의한 부적합 증가가 발생하므로 덜 제한된 모형을 유지해야 함을 나타냅니다 . 카이제곱 검정은 표본 크기의 영향을 받습니다 . 즉 , 표본 크기가 증가할수록 더 유의할 가능성이 높기 때문에 "ΔCFI" 와 "ΔRMSEA" 도 고려해야 합니다 . 이상적으로는 ΔCFI 가 –0.01 을 초과하지 않고 ΔRMSEA 가 0.015 를 초과하지 않아야 합니다 (Chen 2007 참조).

주의 : 이 보고서의 차이 검정은 내포 모형에서만 의미가 있습니다 .

빨간색 X 버튼을 클릭하여 보고서의 행을 제거할 수 있습니다 . 보고서를 완전히 제거하려면 테이블의 모든 행을 제거해야 합니다 . 주 보고서에서 " 구조 방정식 모형 " 노드를 제거하면 해당 모형이 포함된 모든 차이 검정이 테이블에서 제거됩니다 .

JMP PRO 구조 방정식 모형 적합 보고서

" 실행 " 을 클릭할 때마다 지정된 모형에 대한 구조 방정식 모형 보고서가 나타납니다 . 기본적으로 이 보고서에는 " 적합 요약 " 보고서 , " 모수 추정값 " 보고서 및 " 경로 다이어그램 " 이 포함됩니다 .

적합 요약 수렴 상태 및 추정 방법을 포함한 모형 적합 정보 테이블입니다 . 다음 통계량이 이 테이블에 보고됩니다 .

표본 크기 모형을 적합시키는 데 사용된 관측값 (행) 수입니다 .

결측값 포함 행 결측값이 하나 이상 포함된 관측값 (행) 수입니다 . 모든 결측값은 전체 정보 최대 가능도를 사용하여 처리됩니다 (Finkbeiner 1979 참조).

반복 모형을 적합시키는 데 사용된 반복 수입니다 .

-2* 로그 가능도 적합 모형의 로그 가능도에 -2 를 곱한 값입니다 . 이 값은 내포 모형을 비교하는 데 사용할 수 있습니다 . 두 모형의 -2* 로그 가능도 값 차이는 자유도가 모형 간의 자유도 차이와 동일한 카이제곱 분포입니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

모수 수 모형에서 자유 추정 모수의 수입니다 .

AICc 수정된 Akaike 정보 기준입니다 . 이 값은 모형을 비교하는 데 사용할 수 있으며 값이 작을수록 더 나은 모형 적합을 나타냅니다 . 자세한 내용은 "AICc 및 BIC" 에서 확인하십시오 .

BIC Bayes 정보 기준입니다 . 이 값은 모형을 비교하는 데 사용할 수 있으며 값이 작을수록 더 나은 모형 적합을 나타냅니다 . 자세한 내용은 "AICc 및 BIC" 에서 확인하십시오 .

카이제곱 모형에 대한 카이제곱 통계량입니다 .

DF 모형 적합에 대한 카이제곱 검정의 자유도입니다 .

Prob>ChiSq 모형에 대한 카이제곱 통계량의 *p* 값입니다 .

CFI Bentler CFI(비교 적합 지수) 는 모형 적합을 결정하는 추가 지표를 제공합니다 . CFI 는 0 과 1 사이로 제한됩니다 . 0.90 보다 큰 값이 좋습니다 (Browne 과 Cudeck 1993, Hu 와 Bentler 1999 참조). 자세한 내용은 "CFI" 에서 확인하십시오 .

RMSEA RMSEA(근사의 제곱근 평균 제곱 오차) 는 모형 적합을 결정하는 추가 지표를 제공합니다 . RMSEA 는 0 과 1 사이로 제한됩니다 . 0.10 보다 작은 값이 좋습니다 (Browne 과 Cudeck 1993, Hu 와 Bentler 1999 참조). 자세한 내용은 "RMSEA" 에서 확인하십시오 .

90% 하한 RMSEA 에 대한 90% 신뢰 하한입니다 . 자세한 내용은 "RMSEA" 에서 확인하십시오 .

90% 상한 RMSEA 에 대한 90% 신뢰 상한입니다 . 자세한 내용은 "RMSEA" 에서 확인하십시오 .

모수 추정값 모형 모수에 대한 추정값 테이블입니다 . 이 테이블은 " 평균 / 절편 ", " 적재 ", " 회귀 " 및 " 분산 " 섹션으로 구성됩니다 . 각 추정값에 대해 표준 오차 (" 표준 오차 "), Wald 검정 통계량 ("Wald Z") 및 해당 *p* 값 ("Prob>|Z|") 이 제공됩니다 .

팁 : " 모수 추정값 " 테이블에는 각 모형 모수의 " 시작 " 및 " 끝 " 성분을 식별하는 숨겨진 열이 포함되어 있습니다 . 이러한 열을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열 > 시작** 및 **열 > 끝**을 선택합니다 .

경로 다이어그램 적합 모형의 경로 다이어그램 표현을 표시합니다 . 자세한 내용은 " 다이어그램 탭 " 에서 확인하십시오 .

JMP PRO 구조 방정식 모형 플랫폼 옵션

" 구조 방정식 모형 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

경로 다이어그램 설정 경로 다이어그램 표시를 수정하기 위한 다음 옵션을 포함합니다 .

다이어그램 사용자 정의 경로 다이어그램의 여러 측면을 사용자 정의할 수 있습니다 . 자세한 내용은 " 다이어그램 모양 사용자 정의 옵션 " 에서 확인하십시오 .

레이아웃 경로 다이어그램의 전체 모양을 변경하는 두 가지 옵션을 포함합니다 . " 왼쪽에서 오른쪽으로 " 레이아웃 또는 " 위쪽에서 아래쪽으로 " 레이아웃 중에서 선택할 수 있습니다 .

팁 : 경로 다이어그램에서 항목을 드래그하여 특정 항목의 배열을 변경할 수도 있습니다 .

다이어그램 복사 경로 다이어그램 이미지를 클립보드에 저장합니다 . 최상의 품질을 유지하려면 클립보드 이미지를 벡터 그래픽으로 붙여 넣습니다 .

다이어그램 특성 복사 현재 경로 다이어그램 특성을 클립보드에 복사합니다 . 그런 다음 특성을 다른 SEM 경로 다이어그램에 붙여 넣을 수 있습니다 .

다이어그램 특성 붙여넣기 클립보드의 경로 다이어그램 특성을 현재 SEM 경로 다이어그램에 붙여 넣습니다 .

기술 통계량 기술 통계량을 생성하기 위한 다음 옵션을 포함합니다 .

단변량 단순 통계량 단변량 단순 통계량 보고서를 표시하거나 숨깁니다 . 이 보고서의 통계량은 결측 데이터가 있을 수 있는 다른 열과 독립적으로 각 열에 대해 추정됩니다 .

전체 정보 다변량 통계량 다변량 단순 통계량 보고서를 표시하거나 숨깁니다 . 이 보고서의 통계량은 결측 데이터를 설명하기 위해 전체 정보 최대 가능도를 사용하여 추정됩니다 .

이상치 탐색 시작 이상치 탐색 플랫폼을 시작합니다 . 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오 .

결측값 탐색 시작 (결측값이 없는 경우에는 사용 불가능) 결측값 탐색 플랫폼을 시작합니다 . 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오 .

명시 변수 추가 현재 모형에 새 명시 변수를 추가할 수 있습니다 . 추가할 새 명시 변수를 선택하면 새 " 구조 방정식 모형 " 보고서가 시작됩니다 . 새 보고서에는 명시 변수가 새로 추가된 기존 모형 규격이 사용됩니다 .

명시 변수 제거 지정된 명시 변수가 제거된 새 " 구조 방정식 모형 " 보고서를 시작할 수 있습니다 . 제거할 명시 변수를 선택하면 새 보고서가 시작됩니다 . 새 보고서에는 지정된 명시 변수가 제거된 기존 모형 규격이 사용됩니다 .

모형 규격 복사 현재 구조 방정식 모형 규격을 클립보드에 복사합니다 . 그런 다음 모형 규격을 다른 SEM 플랫폼 보고서에 붙여 넣을 수 있습니다 .

모형 규격 붙여넣기 클립보드의 모형 규격을 현재 모형 규격에 붙여 넣습니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

JMP PRO 모형 옵션

" 실행 " 을 클릭하면 지정된 모형에 대한 구조 방정식 모형 보고서가 나타납니다. 이 보고서에는 다음 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다.

경로 다이어그램 표시 모형 보고서에 경로 다이어그램을 표시하거나 숨깁니다.

경로 다이어그램 설정 모형의 경로 다이어그램을 수정하기 위한 다음 옵션을 포함합니다.

다이어그램 사용자 정의 경로 다이어그램의 여러 측면을 사용자 정의할 수 있습니다. 자세한 내용은 " 다이어그램 모양 사용자 정의 옵션 " 에서 확인하십시오.

레이아웃 경로 다이어그램의 전체 모양을 변경하는 두 가지 옵션을 포함합니다. " 왼쪽에서 오른쪽으로 " 레이아웃 또는 " 위쪽에서 아래쪽으로 " 레이아웃 중에서 선택할 수 있습니다.

팁 : 경로 다이어그램에서 항목을 드래그하여 특정 항목의 배열을 변경할 수도 있습니다.

다이어그램 복사 경로 다이어그램 이미지를 클립보드에 저장합니다. 최상의 품질을 유지하려면 클립보드 이미지를 벡터 그래픽으로 붙여 넣습니다.

다이어그램 특성 복사 현재 경로 다이어그램 특성을 클립보드에 복사합니다. 그런 다음 특성을 다른 SEM 경로 다이어그램에 붙여 넣을 수 있습니다.

다이어그램 특성 붙여넣기 클립보드의 경로 다이어그램 특성을 현재 SEM 경로 다이어그램에 붙여 넣습니다.

지수 적합 적합 모형을 평가할 수 있는 다양한 지수 값이 포함된 보고서를 표시하거나 숨깁니다. " 지수 적합 " 보고서에는 " 적합 요약 " 보고서에 나오는 값 외에 다음과 같은 지수 값도 포함됩니다.

TLI TLI(Tucker-Lewis 지수) 는 모형 적합을 결정하는 추가 지표를 제공합니다. 이 지수는 NNFI(비표준 적합 지수) 라고도 합니다. TLI 는 0 과 1 사이로 제한됩니다. 0.95 보다 큰 값이 좋습니다 (West et al. 2012 참조). 자세한 내용은 "TLI" 에서 확인하십시오.

NFI NFI(표준 적합 지수) 는 모형 적합을 결정하는 추가 지표를 제공합니다 . NFI 는 0 과 1 사이로 제한됩니다 . 0.95 보다 큰 값이 좋습니다 (West et al. 2012 참조). 자세한 내용은 "NFI" 에서 확인하십시오 .

수정 GFI 수정 적합도 지수는 모형 적합을 결정하는 추가 지표를 제공합니다 . 수정 GFI 는 0 과 1 사이로 제한됩니다 . 0.95 보다 큰 값이 좋습니다 (West et al. 2012 참조). 자세한 내용은 " 수정 GFI 및 수정 AGFI" 에서 확인하십시오 .

수정 AGFI 수정 조정 적합도 지수는 모형 적합을 결정하는 추가 지표를 제공합니다 . 수정 AGFI 는 0 과 1 사이로 제한됩니다 (West et al. 2012 참조). 자세한 내용은 " 수정 GFI 및 수정 AGFI" 에서 확인하십시오 .

RMR RMR(제곱근 평균 제곱 잔차)은 모형 적합을 결정하는 추가 지표를 제공합니다. RMR 의 잔차는 관측 공분산과 모형 내재 공분산 간의 차이에서 가져옵니다 . RMR 은 양수이고 값이 작을수록 좋습니다 (West et al. 2012 참조). 자세한 내용은 "RMR 및 SRMR" 에서 확인하십시오 .

SRMR SRMR(표준화된 제곱근 평균 제곱 잔차) 은 모형 적합을 결정하는 추가 지표를 제공합니다 . SRMR 의 잔차는 관측 공분산과 모형 내재 공분산 간의 표준화된 차이에서 가져옵니다 . SRMR 은 양수이고 값이 작을수록 좋습니다 (West et al. 2012 참조). 자세한 내용은 "RMR 및 SRMR" 에서 확인하십시오 .

참고 : " 지수 적합 " 보고서의 다른 값에 대한 설명은 " 구조 방정식 모형 적합 보고서 " 에서 확인하십시오 .

적합 요약 모형 적합에 대한 상세 정보가 포함된 보고서를 표시하거나 숨깁니다 .

모수 추정값 모형에 대한 표준화되지 않은 모수 추정값이 포함된 보고서를 표시하거나 숨깁니다 .

표준화 모수 추정값 모형에 대한 표준화된 모수 추정값이 포함된 보고서를 표시하거나 숨깁니다 .

신뢰 구간 "모수 추정값" 및 "표준화 모수 추정값" 보고서에 신뢰 구간을 표시하거나 숨깁니다 .

총 효과 (모형에 하나 이상의 회귀 또는 적재 변수가 포함되고 효과가 수렴하는 경우에만 사용 가능) 모형에 있는 총 효과의 표준화되지 않은 계수 및 표준화된 계수 테이블을 표시하거나 숨깁니다 . 표준 오차도 포함됩니다 . 효과 수렴에 대한 검정은 Bentler 와 Freeman(1983) 연구 자료에 설명되어 있습니다 .

간접 효과 (모형에 중재 변수가 포함되고 효과가 수렴하는 경우에만 사용 가능) 모형에 있는 간접 효과의 표준화되지 않은 계수 및 표준화된 계수 테이블을 표시하거나 숨깁니다 . 표준 오차도 포함됩니다 . 효과 수렴에 대한 검정은 Bentler 와 Freeman(1983) 연구 자료에 설명되어 있습니다 .

예측 프로파일러 결과 변수 집합의 조건부 기대값에 대한 예측 변수 집합의 효과를 볼 수 있습니다 . 이 옵션을 선택하면 하나 이상의 예측 변수와 하나 이상의 결과를 선택해야 하는 창이 나타납니다 . 예측 및 95% 신뢰 구간은 모형 내재 공분산 행렬을 기반으로 합니다 . 예측 프로파일러에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

참고 : 설정 창의 초기 변수 목록은 모형과 일치하는 변수로 제한됩니다 . 예를 들어 " 예측 변수 선택 " 목록에는 모형에서 특정 항목을 예측하는 변수만 포함되고 " 결과 선택 " 목록에는 모형의 다른 변수에 의해 예측되는 변수만 포함됩니다 . 두 목록의 모든 모형 변수를 보려면 **모든 변수 표시** 상자를 선택합니다 .

모형 내재 공분산 모형에 내재된 공분산 행렬이 포함된 보고서를 표시하거나 숨깁니다 .

모형 내재 평균 모형에 내재된 각 변수에 대한 평균이 포함된 보고서를 표시하거나 숨깁니다 .

잔차 모형의 잔차 행렬이 포함된 보고서를 표시하거나 숨깁니다 . 이 행렬은 모형 내재 공분산 행렬과 표본 공분산 행렬 간의 차이입니다 .

정규화 잔차 모형의 정규화 잔차 행렬이 포함된 보고서를 표시하거나 숨깁니다 .

정규화 잔차 히트맵 모형의 정규화 잔차 히트맵이 포함된 보고서를 표시하거나 숨깁니다 .

RAM 행렬 RAM(Reticular Action Model) 표기법에 사용되는 모형 행렬이 포함된 보고서를 표시하거나 숨깁니다 .

추정값 공분산 모형의 모수 추정값 공분산 행렬이 포함된 보고서를 표시하거나 숨깁니다 .

내생 변수의 R^2 (모형이 재귀적이고 내생 변수가 포함된 경우에만 사용 가능) 모형의 각 내생 변수에 대한 R^2 값이 포함된 보고서를 표시하거나 숨깁니다 . 이 값은 1 에서 각 내생 변수에 대한 잔차 분산과 모형 내재 분산의 비율을 뺀 값으로 계산됩니다 . R^2 값은 내생 변수에서 모형에 의해 설명되는 분산 크기를 나타냅니다 . 내생 변수는 경로 다이어그램에서 해당 변수를 가리키는 경로가 있는 변수입니다 .

수정 지수 모형 수정 지수의 모든 추정값 또는 일부를 표시할 수 있습니다 . 이러한 값은 모형 적합을 개선하기 위해 모형에 추가할 수 있는 모수를 결정하는 데 사용될 수 있습니다 . 각 테이블은 카이제곱 열을 기준으로 내림차순 정렬됩니다 .

모든 수정 지수 모든 모형 수정 지수의 추정값을 포함하는 테이블을 표시하거나 숨깁니다 . 이 테이블에는 각 추정값의 모수 유형을 나타내는 열이 있습니다 .

평균에 대한 수정 지수 평균 및 절편에 대한 모형 수정 지수 추정값을 포함하는 테이블을 표시하거나 숨깁니다 .

적재에 대한 수정 지수 적재 모수에 대한 모형 수정 지수 추정값을 포함하는 테이블을 표시하거나 숨깁니다 .

회귀에 대한 수정 지수 회귀 모수에 대한 모형 수정 지수 추정값을 포함하는 테이블을 표시하거나 숨깁니다 .

공분산에 대한 수정 지수 공분산 모수에 대한 모형 수정 지수 추정값을 포함하는 테이블을 표시하거나 숨깁니다 .

측정 모형 평가 (고유 요인 간의 공분산이 없는 확증적 요인 모형에만 사용 가능) 표시자 신뢰도 , 오메가 및 H 계수 , 구성 타당도 행렬 등 검정과 측도의 신뢰도 및 타당도를 정량화하기 위한 다양한 통계량과 그래프를 표시하거나 숨깁니다 .

"표시자 신뢰도" 그림에는 잠재 변수의 표준화된 적재량 제곱과 제안된 최소 허용 가능 신뢰도 임계값 (0.25) 이 표시됩니다. 변수의 값이 낮으면 변수가 해당하는 잠재 변수의 변동을 포착하는 데 그다지 유용하지 않음을 나타냅니다.

"종합 신뢰도" 및 "구성 최대 신뢰도" 보고서에는 각 잠재 변수에 대한 오메가 (McDonald 1999) 계수와 H(Hancock 과 Mueller 2001) 계수가 각각 표시됩니다. 이러한 값의 범위는 0 ~ 1 이며 약 0.70 이상을 권장합니다. 오메가는 관측된 종합 스코어에서 잠재 변수의 분산 비율을 나타냅니다. H 는 표시자가 나타내는 잠재 변수의 분산 비율을 나타냅니다. 이러한 추정값은 모형에 따라 다릅니다. 한 요인 모형이 적합되면 결과 오메가는 일반 오메가입니다. 둘 이상의 잠재 변수가 있는 요인 모형이 적합되면 결과 오메가 추정값을 하위 척도 오메가라고 합니다. 그러나 이중 요인 모형이 적합되는 경우 일반 요인 추정값의 오메가는 계층적 오메가이고 그룹 요인은 계층적 하위 척도 오메가라고 합니다 (Rodriguez et al. 2015 참조). 제안된 임계값은 설문 조사 목표의 컨텍스트에서 사용되어야 합니다. 종합 스코어를 사용하여 개별값에 대한 결정을 내리려면 제안된 임계값보다 신뢰도가 높아야 하지만 (약 0.90 이상) 조사 목적으로 종합 스코어를 사용하려면 임계값의 아래쪽 끝이 허용됩니다 (Nunnally 1978 참조).

"구성 타당도 행렬" 보고서는 사용자가 생각하는 대상을 잠재 변수가 측정하고 있는지 여부를 판별하는 데 도움이 됩니다.

- 하삼각 항목은 잠재 변수 상관을 포함합니다. 이러한 항목을 사용하면 잠재 변수가 서로 어느 정도 상관되어 있는지 확인하고 가설 상관 강도와 비교할 수 있습니다.
- 상삼각 항목은 잠재 변수의 상관 제곱입니다. 이러한 항목을 사용하면 잠재 변수 간의 분산 중첩에 초점을 맞출 수 있습니다. 이 통계량은 행렬의 대각 항목과 비교할 때 특히 유용합니다.
- 대각 항목은 각 잠재 변수에 의해 추출된 평균 분산 크기를 포함합니다. 이 값은 각 잠재 변수에 대한 표시자 신뢰도의 평균과 동일합니다. 대각 항목은 표시자가 잠재 변수를 올바르게 정의하는 데 충분한 계통적 분산을 가지므로 대각 항목의 값이 높은 잠재 변수가 좋습니다. 이상적으로는 각 잠재 변수의 대각 항목이 위쪽 및 오른쪽 항목보다 높아야 합니다.

구성 타당도 행렬을 시각화하면 대각 항목을 상삼각 항목과 비교할 수 있습니다.

자세한 내용은 "측정 모형 평가 보고서 사용의 예" 에서 확인하십시오.

열 저장 적합 구조 방정식 모형을 기반으로 한 열을 데이터 테이블에 저장할 수 있습니다.

요인 스코어 저장 (모형에 잠재 변수가 있는 경우에만 사용 가능) 각 잠재 변수에 대해 회귀 방법을 사용하여 계산된 요인 스코어를 포함하는 열을 데이터 테이블에 저장합니다. 요인 스코어는 데이터 테이블에도 추가되는 숨겨진 열에서 계산됩니다 . 이 숨겨진 열은 `Estimate Factor Score()` JSL 함수를 사용합니다. 이 함수에 대한 자세한 내용은 도움말 > 스크립트 인덱스에서 확인하십시오.

Bartlett 요인 스코어 저장 (모형에 잠재 변수가 있는 경우에만 사용 가능) 각 잠재 변수에 대해 Bartlett 방법을 사용하여 계산된 요인 스코어를 포함하는 열을 데이터 테이블에 저장합니다 . 요인 스코어는 데이터 테이블에도 추가되는 숨겨진 열에서 계산됩니다 . 이 숨겨진 열은 `Estimate Bartlett Factor Score()` JSL 함수를 사용합니다 . 이 함수에 대한 자세한 내용은 도움말 > 스크립트 인덱스에서 확인하십시오 .

예측 계산식 저장 (모형에 하나 이상의 내생 변수 또는 종속 변수가 있는 경우에만 사용 가능) 각 변수에 대한 관측 결과의 예측값 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 모형에 잠재 변수가 있으면 Bartlett 방법을 사용하여 계산된 요인 스코어도 데이터 테이블에 저장됩니다 .

관측 잔차 저장 (모형에 하나 이상의 내생 변수 또는 종속 변수가 있는 경우에만 사용 가능) 각 변수에 대한 관측 결과의 잔차 값이 포함된 열을 데이터 테이블에 저장합니다 . 모형에 잠재 변수가 있으면 Bartlett 방법을 사용하여 계산된 요인 스코어도 데이터 테이블에 저장됩니다 .

모형 규격 복사 현재 구조 방정식 모형 규격을 클립보드에 복사합니다 . 그런 다음 모형 규격을 다른 SEM 플랫폼 보고서에 붙여 넣을 수 있습니다 .

모형 규격으로 재호출 모형 규격 보고서의 모형을 지정된 모형으로 설정합니다 .

적합 제거 보고서 창에서 지정된 모형 보고서를 제거합니다 .

JMP PRO SEM 경로 다이어그램 사용자 정의

다양한 방법으로 경로 다이어그램을 사용자 정의할 수 있습니다 . 대부분의 사용자 정의 옵션은 경로 다이어그램의 팝업 메뉴와 " 다이어그램 모양 사용자 정의 " 창 중 하나 또는 둘 다에 있습니다 .

- " 경로 다이어그램 팝업 메뉴 옵션 "
- " 다이어그램 모양 사용자 정의 옵션 "

JMP PRO 경로 다이어그램 팝업 메뉴 옵션

경로 다이어그램 자체를 마우스 오른쪽 버튼으로 클릭하면 경로 다이어그램 팝업 메뉴가 나타납니다 . 메뉴에서 사용할 수 있는 특정 옵션 집합은 다이어그램에서 마우스 오른쪽 버튼을 클릭한 위치와 마우스 오른쪽 버튼을 클릭할 때 경로 다이어그램 요소가 선택되었는지 여부에 따라 달라집니다 .

선택 " 시작점 목록 " 및 " 끝점 목록 " 에서 선택 , 다이어그램에 단방향 및 양방향 화살표 추가 , 새 잠재 변수 생성 , 선택한 항목 숨기기 및 숨겨진 모든 항목 표시를 위한 옵션을 포함합니다 . 경로 다이어그램에서 항목을 선택하지 않으면 이 하위 메뉴에 있는 대부분의 옵션을 사용할 수 없습니다 .

팁 : 경로 다이어그램의 항목이 숨겨지면 숨겨진 모든 항목을 표시할 수 있는 " 모두 숨기기 해제 " 버튼이 경로 다이어그램 맨 위에 나타납니다 .

표시 경로 다이어그램의 다양한 요소를 표시하거나 숨기는 옵션을 포함합니다 . 표시된 요소의 원래 집합으로 돌아가려면 **기본값 표시**를 선택합니다 .

참고 : " 기본값 표시 " 옵션은 " 평균 / 절편 표시 " 옵션 설정에 영향을 주지 않습니다 .

적합 모형 보고서에 있는 경로 다이어그램의 경우 이 하위 메뉴에는 경로 다이어그램의 화살표에 표시되는 추정값을 변경하는 옵션도 포함되어 있습니다 . 표준화되지 않은 모수 추정값 또는 표준화된 모수 추정값을 표시하거나 추정값을 표시하지 않도록 선택할 수 있습니다 .

이 유형 모두 선택 (경로 다이어그램에서 항목을 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 경로 다이어그램에서 마우스 오른쪽 버튼으로 클릭한 항목과 유형이 같은 다른 모든 항목을 선택합니다 .

잠재 변수 그룹 선택 (경로 다이어그램에서 잠재 변수 항목을 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 선택한 잠재 변수와 연결된 전체 항목 그룹을 선택합니다 . 잠재 변수를 선택하지 않은 경우 이 옵션은 마우스 오른쪽 버튼으로 클릭한 잠재 변수와 연결된 항목 그룹만 선택합니다 .

변수 이름 바꾸기 (경로 다이어그램에서 잠재 변수 항목을 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 경로 다이어그램에서 잠재 변수 이름을 변경할 수 있습니다 .

참고 : 경로 다이어그램에서 잠재 변수 이름을 바꾸면 " 시작점 목록 " 및 " 끝점 목록 " 에서 이름이 업데이트됩니다 .

회귀 추가 (경로 다이어그램에서 변수를 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 마우스 오른쪽 버튼으로 클릭한 변수와 경로 다이어그램에 있는 다른 변수 사이의 회귀를 나타내는 단방향 화살표를 추가할 수 있습니다 . 이 옵션을 선택하는 경우 화살표로 가리킬 변수를 클릭해야 합니다 .

공분산 추가 (경로 다이어그램에서 변수를 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 마우스 오른쪽 버튼으로 클릭한 변수와 경로 다이어그램에 있는 다른 변수 사이의 공분산을 나타내는 양방향 화살표를 추가할 수 있습니다 . 이 옵션을 선택하는 경우 화살표로 가리킬 변수를 클릭해야 합니다 .

다이어그램 사용자 정의 " 다이어그램 모양 사용자 정의 " 창을 시작합니다 . 자세한 내용은 " 다이어그램 모양 사용자 정의 옵션 " 에서 확인하십시오 .

잠재 변수 그룹 회전 경로 다이어그램에서 선택한 잠재 변수에 대해 시계 방향으로 90 씩 증분하여 잠재 변수 표시자를 회전합니다 . 잠재 변수를 선택하지 않은 경우 이 옵션은 경로 다이어그램의 모든 잠재 변수 표시자를 회전합니다 .

레이아웃 경로 다이어그램의 항목에 대한 두 가지 배열 옵션 중 하나를 선택할 수 있습니다 .

다이어그램 복사 경로 다이어그램 이미지를 클립보드에 저장합니다 . 최상의 품질을 유지하려면 클립보드 이미지를 벡터 그래픽으로 붙여 넣습니다 .

다이어그램 특성 복사 현재 경로 다이어그램 특성을 클립보드에 복사합니다 . 그런 다음 특성을 다른 SEM 경로 다이어그램에 붙여 넣을 수 있습니다 .

다이어그램 특성 붙여넣기 클립보드의 경로 다이어그램 특성을 현재 SEM 경로 다이어그램에 붙여 넣습니다 .

실행 취소 경로 다이어그램에 대한 마지막 변경 내용을 실행 취소합니다 .

다시 실행 경로 다이어그램에 대해 마지막으로 실행 취소한 변경 내용을 다시 실행합니다 .

편집 JMP 의 그래프에 대한 표준 옵션을 포함합니다 . 이러한 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

레이아웃 재설정 경로 다이어그램을 원래대로 재설정합니다 .

JMP PRO 다이어그램 모양 사용자 정의 옵션

" 다이어그램 " 패널에서 사용자 정의 버튼을 클릭하거나 , 빨간색 삼각형 메뉴에서 " 다이어그램 사용자 정의 " 를 선택하거나 , 경로 다이어그램 팝업 메뉴에서 " 다이어그램 사용자 정의 " 를 선택하면 " 다이어그램 모양 사용자 정의 " 창이 나타납니다 . " 다이어그램 모양 사용자 정의 " 창에는 사전 설정된 두 개의 색상 테마 중에서 선택할 수 있는 드롭다운 메뉴와 네 개의 패널이 있습니다 .

테마 사전 설정 사전 설정된 색상 테마 옵션을 포함합니다 . " 흑백 " 테마 또는 " 파란색 " 테마 중에서 선택할 수 있습니다 . 두 테마 중 하나를 선택하면 해당 테마에 맞게 설정이 빠르게 업데이트됩니다 .

변수 모양 경로 다이어그램에 있는 항목의 채우기 색상 , 테두리 색상 , 텍스트 색상 및 크기를 설정하는 옵션을 포함합니다 . 명시 변수 , 잠재 변수 및 상수 변수의 모양과 이러한 각 항목의 글꼴을 설정할 수 있습니다 .

경로 설정 화살표 모양에 대한 옵션을 포함합니다 . 각 경로 화살표의 두께 및 투명도를 고정 값으로 설정하거나 표준화 추정값을 기반으로 할 수 있습니다 . 유의하지 않은 p 값을 파선으로 나타내고 , 비유의성에 대한 α 수준을 설정하고 , 화살표와 관련된 색상과 글꼴을 변경할 수도 있습니다 . 기본적으로 화살표 두께는 고정 값으로 설정되고 화살표 투명도는 표준화 모수 추정값에 비례합니다 .

기타 설정 다음 옵션이 포함되어 있습니다 .

격자 활성화 다이어그램의 항목을 정렬하는 데 사용할 수 있는 격자를 경로 다이어그램에 표시하거나 숨깁니다 .

다이어그램 잠금 경로 다이어그램에서 항목 배치를 잠그거나 잠금 해제합니다 . 경로 다이어그램이 잠겨 있으면 항목을 드래그할 수 없으며 경로 다이어그램 맨 위의 잠금 표시자가 강조 표시됩니다 .

평균 / 절편 표시 경로 다이어그램에 평균과 절편을 표시하거나 숨깁니다 . 기본적으로 평균 구조는 경로 다이어그램에 표시되지 않습니다 .

잠재 변수 표시자 표시 경로 다이어그램에 잠재 변수 표시자를 표시하거나 숨깁니다 .

회귀 표시 경로 다이어그램에 회귀 화살표를 표시하거나 숨깁니다.

분산 표시 경로 다이어그램에 분산 화살표를 표시하거나 숨깁니다.

공분산 표시 경로 다이어그램에 공분산 화살표를 표시하거나 숨깁니다.

등식 제약 조건 표시 경로 다이어그램에 등식 제약 조건을 표시하거나 숨깁니다.

R 제곱 값 표시 경로 다이어그램의 각 노드에 대한 R^2 값을 표시하거나 숨깁니다.

R 제곱으로 노드 채우기 경로 다이어그램에서 각 노드에 대한 R^2 값에 해당하는 각 노드의 채우기 색상을 표시하거나 숨깁니다.

미리보기 창의 현재 설정을 기반으로 하는 경로 다이어그램 모양 미리보기를 포함합니다.

팁 : 이후 경로 다이어그램에 동일한 설정을 사용할 수 있도록 창의 현재 설정을 플랫폼 환경 설정에 저장하려면 " 다이어그램 모양 사용자 정의 " 창 아래쪽에 있는 **환경 설정에 저장** 버튼을 사용합니다.

JMP PRO 구조 방정식 모형의 추가 예

- " 잠재 경로 변수 모형의 예 "
- " 잠재 성장 곡선 모형의 예 "
- " 측정 모형 평가 보고서 사용의 예 "

잠재 경로 변수 모형의 예

이 예에서는 75 개 개발도상국의 데이터를 사용하는 Bollen(1989) 연구 자료에 설명된 산업화 및 정치적 민주주의에 대한 구조 방정식 모형을 생성합니다. 데이터 테이블의 변수에는 1960 년과 1965 년의 네 가지 민주주의 측도와 1960 년의 세 가지 산업화 측도가 포함됩니다. 이러한 변수는 데이터 테이블의 각 열에서 " 노트 " 열 특성에 설명되어 있습니다. " 노트 " 열 특성을 보려면 열 이름을 마우스 오른쪽 버튼으로 클릭하고 " 열 정보 " 를 선택한 후 " 열 특성 " 아래의 " 노트 " 를 선택합니다. 생성하는 구조 방정식 모형의 유형은 구조 회귀 모형입니다.

모형 규격 과정은 잠재 변수 생성, 적재 및 회귀 변수 추가, 공분산 항 추가 및 적재 변수에 제약 조건 적용이라는 네 가지 주 단계로 구성됩니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Political Democracy.jmp 를 엽니다.
2. **분석 > 다변량 방법 > 구조 방정식 모형**을 선택합니다.
3. Prod60 ~ Legis65 를 선택하고 **모형 변수**를 클릭합니다.
4. **확인**을 클릭합니다.

   " 구조 방정식 모형 " 보고서의 " 모형 규격 " 개요가 나타납니다.
5. 보기 패널 상자에서 **목록** 탭을 클릭합니다.

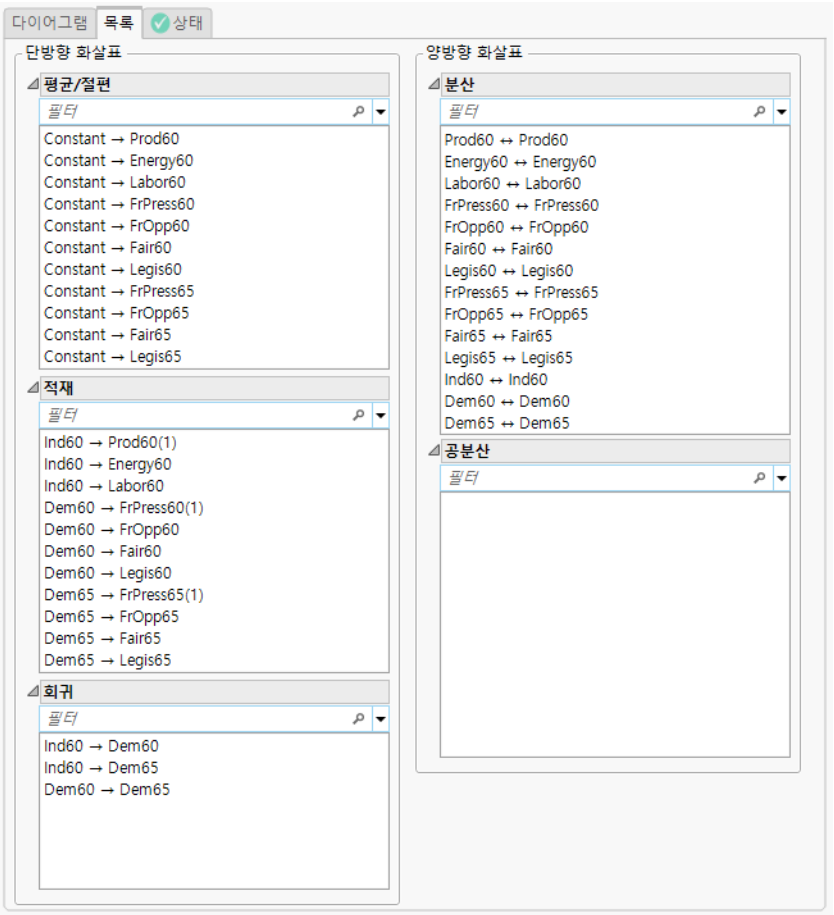
잠재 변수 생성

6. " 끝점 목록 " 에서 Prod60 ~ Labor60 을 선택하고 " 끝점 목록 " 아래의 상자에 "Ind60" 을 입력한 후 잠재 변수 추가 버튼 ⊕을 클릭합니다 .
7. " 끝점 목록 " 에서 FrPress60 ~ Legis60 을 선택하고 " 끝점 목록 " 아래의 상자에 "Dem60" 을 입력한 후 잠재 변수 추가 버튼 ⊕을 클릭합니다 .
8. " 끝점 목록 " 에서 FrPress65 ~ Legis65 를 선택하고 " 끝점 목록 " 아래의 상자에 "Dem65" 를 입력한 후 잠재 변수 추가 버튼 ⊕을 클릭합니다 .

적재 및 회귀 변수 추가

9. " 시작점 목록 " 에서 Ind60 을 선택하고 " 끝점 목록 " 에서 Dem60 을 선택한 후 단방향 화살표 버튼 →을 클릭합니다 .
10. " 시작점 목록 " 에서 Ind60 을 선택하고 " 끝점 목록 " 에서 Dem65 를 선택한 후 단방향 화살표 버튼 →을 클릭합니다 .
11. " 시작점 목록 " 에서 Dem60 을 선택하고 " 끝점 목록 " 에서 Dem65 를 선택한 후 단방향 화살표 버튼 →을 클릭합니다 .

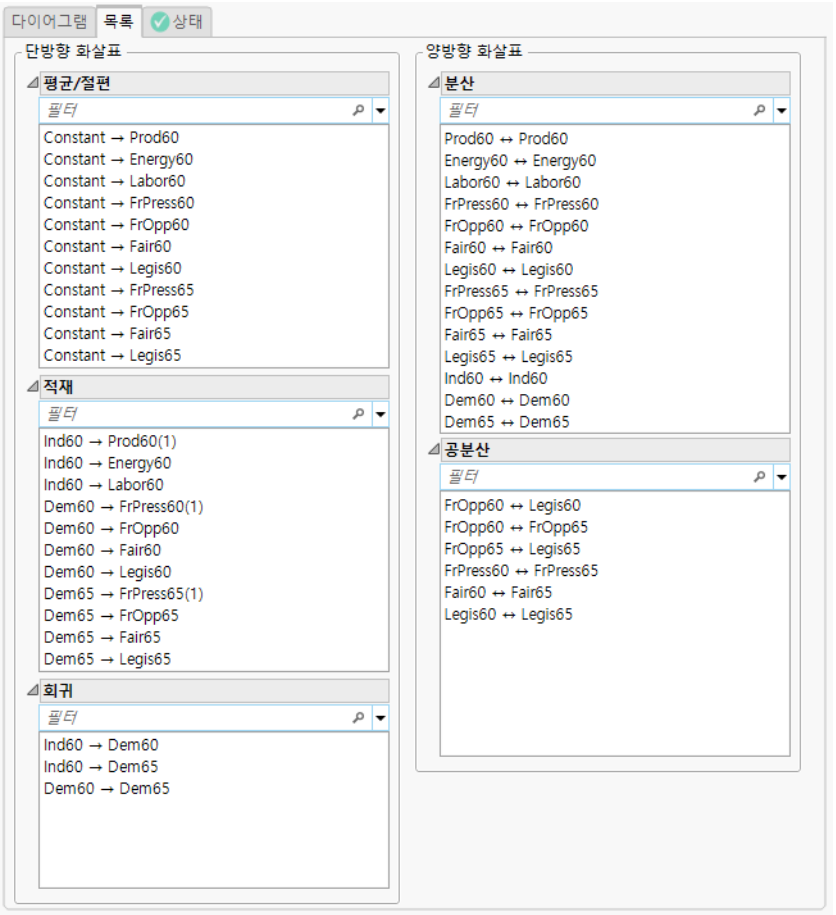
그림 8.6 적재 및 회귀



공분산 추가

12. " 시작점 목록 " 에서 FrOpp60 을 선택하고 " 끝점 목록 " 에서 Legis60 과 FrOpp65 를 선택한 후 양방향 화살표 버튼 ↔ 을 클릭합니다 .
13. " 시작점 목록 " 에서 FrOpp65 를 선택하고 " 끝점 목록 " 에서 Legis65 를 선택한 후 양방향 화살표 버튼 ↔ 을 클릭합니다 .
14. " 시작점 목록 " 에서 FrPress60 을 선택하고 " 끝점 목록 " 에서 FrPress65 를 선택한 후 양방향 화살표 버튼 ↔ 을 클릭합니다 .
15. " 시작점 목록 " 에서 Fair60 을 선택하고 " 끝점 목록 " 에서 Fair65 를 선택한 후 양방향 화살표 버튼 ↔ 을 클릭합니다 .
16. " 시작점 목록 " 에서 Legis60 을 선택하고 " 끝점 목록 " 에서 Legis65 를 선택한 후 양방향 화살표 버튼 ↔ 을 클릭합니다 .

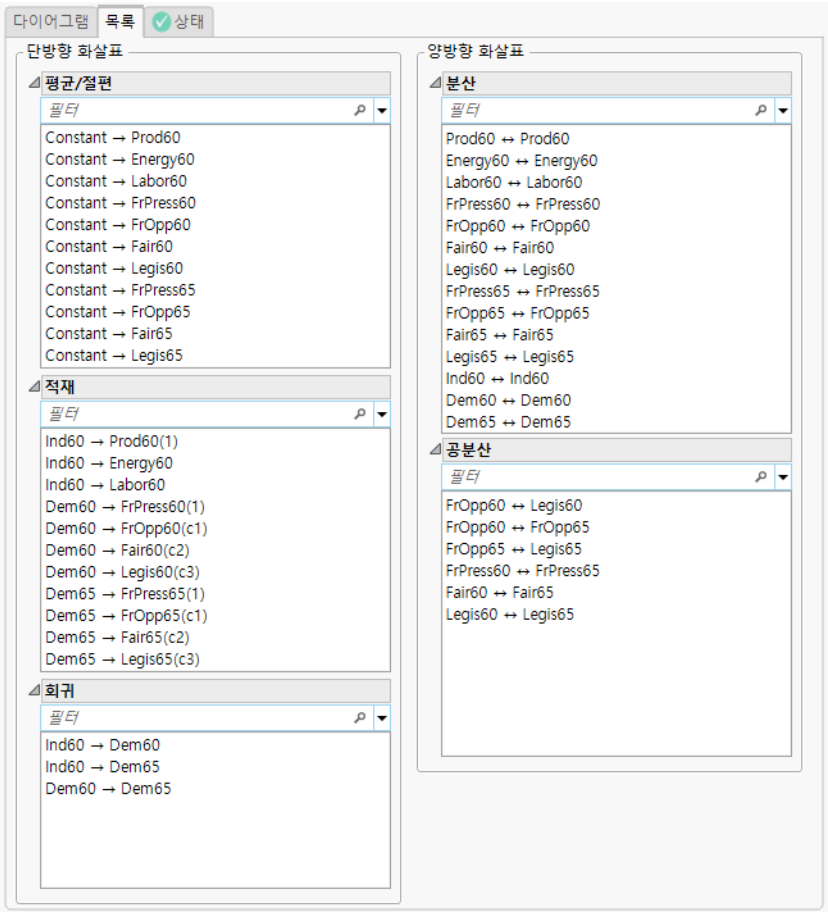
그림 8.7 공분산



적재에 제약 조건 추가

17. " 적재 " 목록에서 Dem60->FrOpp60 과 Dem65->FrOpp65 를 선택하고 **등가 설정**을 클릭합니다 .
18. " 적재 " 목록에서 Dem60->Fair60 과 Dem65->Fair65 를 선택하고 **등가 설정**을 클릭합니다 .
19. " 적재 " 목록에서 Dem60->Legis60 과 Dem65->Legis65 를 선택하고 **등가 설정**을 클릭합니다 .

그림 8.8 완료된 모형 규격



적재에 대한 제약 조건은 영숫자 라벨로 지정됩니다 . 예를 들어 Dem60->FrOpp60 과 Dem65->FrOpp65 는 둘 다 "c1" 이라는 라벨이 지정되어 있으므로 등가 설정된 것을 알 수 있습니다 .

20. " 모형 이름 " 아래의 텍스트 상자에 "Industrialization and Political Democracy" 를 입력합니다 .
21. **실행**을 클릭합니다 .

그림 8.9 구조 방정식 모형의 적합 요약 보고서

적합 요약

그래디언트에 수렴됨	
최대 가능도	
표본 크기	75
결측값 포함 행	0
반복	7
-2*로그 가능도	3097.6362
모수 수	39
AICc	3264.779
BIC	3266.0182
카이제곱	40.179489
DF	38
Prob>ChiSq	0.3738824
CFI	0.9967743
RMSEA	0.0276538
90% 하한	0
90% 상한	0.0870678

" 적합 요약 " 보고서에 나열된 이 모형의 카이제곱 통계량은 40.18 이고 자유도는 38 입니다 . 해당 p 값은 0.3739 로 , 유의하지 않습니다 . 이는 모형이 잘 적합된다는 귀무가설을 기각할 증거가 없음을 나타냅니다 . 따라서 이 모형이 데이터를 적절하게 적합시킨다는 결론을 내릴 수 있습니다 .

카이제곱 값은 표본 크기에 따라 달라지므로 일부 적합 모형에서 유의한 카이제곱 값을 생성할 수 있습니다 . CFI(비교 적합 지수) 와 RMSEA(근사의 제곱근 평균 제곱 오차) 는 모형 적합을 결정하는 추가 지표를 제공합니다 . 이러한 지수는 0 과 1 사이로 제한됩니다 . CFI 값이 0.90 보다 크고 RMSEA 값이 0.10 보다 작으면 좋습니다 (Browne 과 Cudeck 1993, Hu 와 Bentler 1999 참조). 여기서 CFI 0.9968 과 RMSEA 0.0277 은 적합도가 매우 좋음을 나타냅니다 .

그림 8.10 구조 방정식 모형의 모수 추정값 보고서

모수 추정값

평균/절편	추정값	표준 오차	Wald Z	Prob>|Z|
Constant → Energy60	4.7921946	0.1732697	27.657432	<.0001*
Constant → Labor60	3.5576898	0.1612318	22.065682	<.0001*
Constant → FrPress60	5.4646667	0.2989114	18.281892	<.0001*
Constant → FrOpp60	4.2564429	0.4389904	9.6959807	<.0001*
Constant → Fair60	6.5631103	0.3939809	16.658446	<.0001*
Constant → Legis60	4.452533	0.3796307	11.728591	<.0001*
Constant → FrPress65	5.1362519	0.3044457	16.870832	<.0001*
Constant → FrOpp65	2.9780741	0.3923504	7.5903426	<.0001*
Constant → Fair65	6.1962639	0.3643999	17.004021	<.0001*
Constant → Legis65	4.0433897	0.3753526	10.772243	<.0001*
적재	**추정값**	**표준 오차**	**Wald Z**	**Prob>|Z|**
Ind60 → Prod60	1	.	.	.
Ind60 → Energy60	2.1796569	0.1389146	15.690626	<.0001*
Ind60 → Labor60	1.8182095	0.152128	11.951843	<.0001*
Dem60 → FrPress60	1	.	.	.
Dem60 → FrOpp60	1.1907835	0.1416451	8.4068115	<.0001*
Dem60 → Fair60	1.1745402	0.1197987	9.8042797	<.0001*
Dem60 → Legis60	1.2509806	0.1229262	10.176679	<.0001*
Dem65 → FrPress65	1	.	.	.
Dem65 → FrOpp65	1.1907835	0.1416451	8.4068115	<.0001*
Dem65 → Fair65	1.1745402	0.1197987	9.8042797	<.0001*
Dem65 → Legis65	1.2509806	0.1229262	10.176679	<.0001*
회귀	**추정값**	**표준 오차**	**Wald Z**	**Prob>|Z|**
Ind60 → Dem60	1.471331	0.3914721	3.7584565	0.0002*
Ind60 → Dem65	0.6004747	0.2382837	2.5199993	0.0117*
Dem60 → Dem65	0.8650414	0.0756835	11.429724	<.0001*
분산	**추정값**	**표준 오차**	**Wald Z**	**Prob>|Z|**
Prod60 ↔ Prod60	0.0813877	0.0196996	4.13144	<.0001*
Energy60 ↔ Energy60	0.1204277	0.0699024	1.7227975	0.0849
Labor60 ↔ Labor60	0.4666599	0.0891233	5.2361183	<.0001*
FrPress60 ↔ FrPress60	1.8546493	0.4569855	4.0584422	<.0001*
FrOpp60 ↔ FrOpp60	7.5813412	1.3449672	5.6368223	<.0001*
Fair60 ↔ Fair60	4.9556729	0.96122	5.1556073	<.0001*
Legis60 ↔ Legis60	3.2244878	0.7417156	4.3473373	<.0001*
FrPress65 ↔ FrPress65	2.3130434	0.4834138	4.7848104	<.0001*
FrOpp65 ↔ FrOpp65	4.9681894	0.8945151	5.554059	<.0001*
Fair65 ↔ Fair65	3.5600372	0.7379206	4.8244178	<.0001*
Legis65 ↔ Legis65	3.3076951	0.7128724	4.6399537	<.0001*
Ind60 ↔ Ind60	0.448599	0.0867473	5.1713335	<.0001*
Dem60 ↔ Dem60	3.8753207	0.8886194	4.3610581	<.0001*
Dem65 ↔ Dem65	0.1644634	0.2333453	0.7048069	0.4809
공분산	**추정값**	**표준 오차**	**Wald Z**	**Prob>|Z|**
FrOpp60 ↔ Legis60	1.4401257	0.6909886	2.0841528	0.0371*
FrOpp60 ↔ FrOpp65	2.1830097	0.7311078	2.9858931	0.0028*
FrOpp65 ↔ Legis65	1.3717876	0.5781788	2.3726008	0.0177*
FrPress60 ↔ FrPress65	0.5825448	0.3644286	1.5985156	0.1099
Fair60 ↔ Fair65	0.7115828	0.6194143	1.1487994	0.2506
Legis60 ↔ Legis65	0.362804	0.4607898	0.7873525	0.4311

다음으로 , " 회귀 " 아래의 모수 추정값은 Dem60 과 Dem65 에 대한 Ind60 의 양의 효과 및 Dem65 에 대한 Dem60 의 양의 효과를 보여 줍니다 . 따라서 Ind60 의 스코어가 높을수록 Dem60 과 Dem65 의 스코어가 높고 Dem60 의 스코어가 높을수록 Dem65 의 스코어가 높습니다 . 모수 추정값에 대한 해당 p 값은 " 회귀 " 아래에 표시됩니다 . 3 개의 회귀 모수는 $\alpha = 0.05$ 수준에서 모두 유의합니다 . 따라서 잠재 변수 사이에 0 이 아닌 관계가 있다는 결론을 내릴 수 있습니다 .

잠재 성장 곡선 모형의 예

LGC(잠재 성장 곡선) 모형을 사용하면 분석자가 종단 데이터를 모델링하고 시간에 따른 궤적을 특성화할 수 있습니다. 가장 일반적인 잠재 성장 곡선 모형은 과정의 전체 궤적과 전체 궤적에서 개별 관측값의 편차를 포착하기 위해 절편 및 기울기 잠재 변수를 지정하는 선형 성장 모형입니다. 이 모형은 랜덤 계수 모형과 매우 유사합니다. LGC 모형의 주요 특징은 데이터의 평균 구조에 제약 조건을 적용하는 것입니다. 평균 구조를 모델링하면 성장과 관련된 가설을 검정할 수 있습니다.

이 예에서는 4 년 동안 학업 프로그램을 이수한 학생들의 성취도를 모델링합니다 . 선형 궤적이 이 과정의 특징이라는 가설을 검정하려고 합니다 .

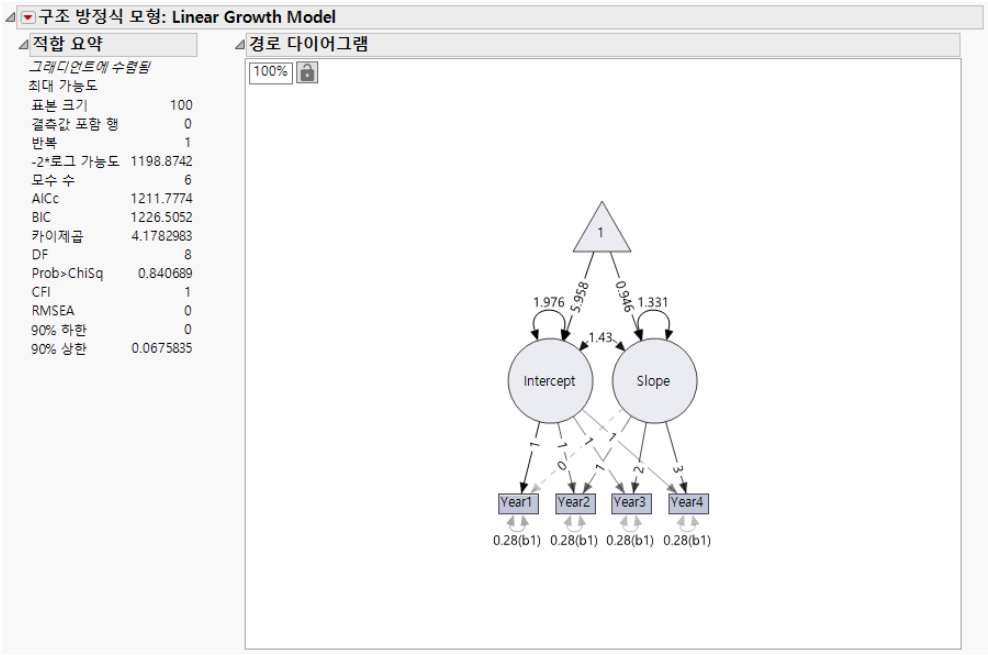
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Academic Achievement.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 구조 방정식 모형**을 선택합니다 .
3. Year1 ~ Year4 를 선택하고 **모형 변수**를 클릭합니다 .
4. **확인**을 클릭합니다 .

   " 구조 방정식 모형 " 보고서의 " 모형 규격 " 개요가 나타납니다 .
5. **모형 바로 가기 > 종단 분석 > 선형 잠재 성장 곡선**을 선택합니다 .

   이제 모형 규격 다이어그램에 선형 잠재 성장 곡선 모형이 표시됩니다 .
6. 보기 패널 상자에서 **목록** 탭을 클릭합니다 .
7. " 분산 " 목록에서 처음 4 개 항목을 선택하고 **등가 설정**을 클릭합니다 .

   이렇게 하면 각 변수의 잔차 분산이 동일한 값으로 제한되며 이는 분산 가정의 동질성과 유사합니다 . SEM 에서는 이 가정을 검정할 수 있으므로 등식 제약 조건 없이 다른 모형을 적합시키고 카이제곱 차이 검정과 비교할 수 있습니다 .
8. " 모형 이름 " 아래의 텍스트 상자에 "Linear LGC with Equal Variances" 를 입력합니다 .
9. **실행**을 클릭합니다 .

그림 8.11 선형 LGC 모형에 대한 적합 요약 및 경로 다이어그램



" 적합 요약 " 보고서에 나열된 이 모형의 카이제곱 통계량은 4.18 이고 자유도는 8 입니다 . 해당 p 값은 0.8407 로 , 유의하지 않습니다 . 이는 모형이 잘 적합된다는 귀무가설을 기각할 증거가 없음을 나타냅니다. 따라서 이 모형이 데이터를 적절하게 적합시킨다는 결론을 내릴 수 있습니다 . 또한 CFI 및 RMSEA 적합 지수가 각각 0.9 보다 크고 0.1 보다 작으므로 적합도가 탁월하다는 것을 나타냅니다 .

그림 8.12 선형 LGC 모형에 대한 모수 추정값

구조 방정식 모형: Linear Growth Model

모수 추정값

평균/절편	추정값	표준 오차	Wald Z	Prob>|Z|
Constant → Intercept	5.9579957	0.1473906	40.423176	<.0001*
Constant → Slope	0.9461125	0.1177755	8.0331884	<.0001*
적재	**추정값**	**표준 오차**	**Wald Z**	**Prob>|Z|**
Intercept → Year1	1	.	.	.
Intercept → Year2	1	.	.	.
Intercept → Year3	1	.	.	.
Intercept → Year4	1	.	.	.
Slope → Year1	0	.	.	.
Slope → Year2	1	.	.	.
Slope → Year3	2	.	.	.
Slope → Year4	3	.	.	.
분산	**추정값**	**표준 오차**	**Wald Z**	**Prob>|Z|**
Year1 ↔ Year1	0.2803392	0.0280339	10	<.0001*
Year2 ↔ Year2	0.2803392	0.0280339	10	<.0001*
Year3 ↔ Year3	0.2803392	0.0280339	10	<.0001*
Year4 ↔ Year4	0.2803392	0.0280339	10	<.0001*
Intercept ↔ Intercept	1.9761611	0.3078496	6.4192413	<.0001*
Slope ↔ Slope	1.3310383	0.1962465	6.7824803	<.0001*
공분산	**추정값**	**표준 오차**	**Wald Z**	**Prob>|Z|**
Intercept ↔ Slope	1.429507	0.2197848	6.5041213	<.0001*

절편의 평균 추정값은 5.96 이며 , 이는 첫 해의 전체 성취도를 나타냅니다 . 즉 , 첫 번째 측정에서 학생들의 학업 성취도가 평균 5.96 입니다 . 첫 번째 측정에 해당하는 변수에 대한 기울기 요인의 적재가 0 이므로 절편의 평균은 이 변수에서 중심화됩니다 .

기울기의 평균 추정값은 0.95 이며 , 이는 전체 성취도가 매년 0.95 씩 증가한다는 것을 나타냅니다 . 절편 및 기울기 요인의 분산 추정값이 유의하므로 평균 궤적의 유의한 변동을 나타냅니다 . 학생들마다 시작 지점도 다르고 증가 비율도 다릅니다 . 외부 변수를 절편 및 기울기의 예측 변수로 사용하여 개인의 궤적 차이를 발생시키는 요인을 파악할 수 있습니다 .

마지막으로 , 절편과 기울기 사이의 양의 공분산은 첫 해에 더 높은 수준의 학업 성취도로 시작하는 학생이 시간이 지남에 따라 더 빠르게 증가하는 경향이 있음을 나타냅니다 .

LGC 모형은 시간에 따른 과정 변화를 조사하기 위한 유연한 접근 방식을 제공합니다 . "모형 바로 가기" 메뉴는 무성장 궤적 , 2 차 궤적 및 잠재 변수 기반 비선형 궤적을 포함하여 최량 모형 적합을 확인하기 위해 서로 적합시키고 검정할 수 있는 대체 궤적을 제공합니다 .

측정 모형 평가 보고서 사용의 예

CFA(확증적 요인 분석) 모형을 사용하면 대체 측정 모형을 검정할 수 있습니다 . "측정 모형 평가" 보고서는 검정과 측도의 신뢰도 및 타당도를 정량화하기 위한 도구를 제공합니다 . 결과에는 표시자 신뢰도 , 오메가 및 H 계수 , 구성 타당도 행렬이 포함됩니다 .

이 예에서는 소비자 데이터 조사의 타당도와 신뢰도를 평가합니다 . Privacy, Security, Reputation, Trust 및 Purchase Intent 라는 다섯 가지 잠재 변수에 대한 확증적 요인 분석 모형을 적합시킵니다 . 그런 다음 " 측정 모형 평가 " 옵션을 사용하여 설문 조사 신뢰도를 검정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Online Consumer Data.jmp 를 엽니다 .
2. **SEM: CFA** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

이 스크립트는 설문 조사 데이터에 대한 확증적 요인 분석 모형을 실행합니다 .

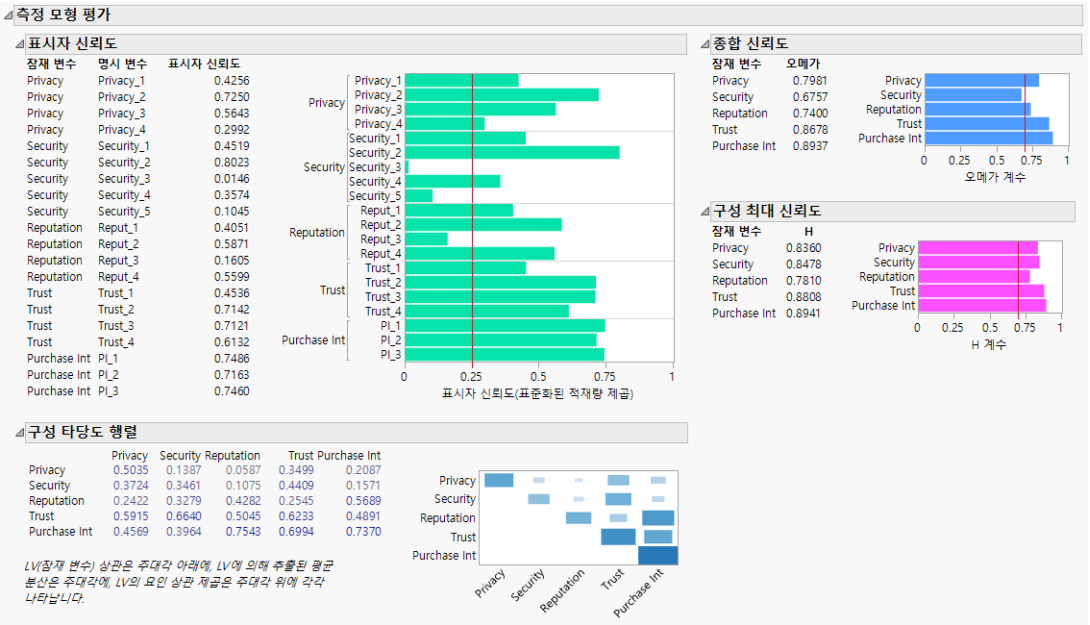
3. "모형 규격" 및 "모형 비교" 옆의 회색 표시 아이콘을 클릭하여 해당 보고서 부분을 숨깁니다.

그림 8.13 CFA 모형에 대한 적합 요약

" 적합 요약 " 보고서에 나열된 이 모형의 카이제곱 통계량은 808.11 이고 자유도는 160 입니다 . 해당 p 값은 유의합니다 . 이는 모형이 잘 적합된다는 귀무가설을 기각할 몇 가지 증거가 있음을 나타냅니다 . 그러나 카이제곱 통계량은 표본 크기에 큰 영향을 받으며 , 관측값 수가 200 개 ~ 300 개를 초과하는 표본의 데이터에 모형이 잘 적합되는 경우에도 유의한 것으로 나타났습니다 . 이 예에서 표본 크기는 843 입니다 . 따라서 CFI 및 RMSEA 적합 지수로도 모형 적합을 평가해야 합니다 . 이 예에서는 두 지수가 각각 0.9 보다 크고 0.1 보다 작으므로 모두 양호한 적합을 나타냅니다 .

4. " 구조 방정식 모형 : CFA" 옆의 빨간색 삼각형을 클릭하고 **측정 모형 평가**를 선택합니다 .

그림 8.14 CFA 모형에 대한 측정 모형 평가 보고서



"표시자 신뢰도" 그림에는 잠재 변수의 표준화된 적재량 제곱과 제안된 최소 허용 가능 신뢰도 임계값 (0.25) 이 표시됩니다 . Security 관련 질문 두 개와 Reputation 관련 질문 하나가 해당 잠재 변수의 변동을 포착하는 데 그다지 유용하지 않다는 것을 알 수 있습니다 . Security_3 질문은 값이 매우 낮으므로 구성의 신뢰도에 거의 영향을 주지 않고 제거할 수 있는 반면 , Security_5 및 Reput_3 질문은 신뢰도 향상을 위해 수정할 수 있습니다 .

"종합 신뢰도" 및 "구성 최대 신뢰도" 보고서에는 각 잠재 변수에 대한 오메가 계수와 H 계수가 각각 표시됩니다 . 이러한 값의 범위는 0 ~ 1 이며 약 0.70 이상을 권장합니다 . 오메가는 관측된 종합 스코어에서 잠재 변수의 분산 비율을 나타냅니다 . H 는 표시자가 나타내는 잠재 변수의 분산 비율을 나타냅니다 . 이러한 측도에 따르면 Security 의 오메가 계수는 제안된 임계값 0.70 보다 아주 약간 낮지만 조사 목적으로는 이 종합 신뢰도가 충분하다고 할 수 있습니다 . 제안된 임계값은 설문 조사 목표의 컨텍스트에서 사용되어야 합니다 . 종합 스코어를 사용하여 개별값에 대한 결정을 내리려면 제안된 임계값보다 신뢰도가 높아야 하지만 (약 0.90 이상) 조사 목적으로 종합 스코어를 사용하려면 임계값의 아래쪽 끝이 허용됩니다 (Nunnally 1978 참조). Security_3, Security_5 및 Reput_4 질문에 대한 표시자 신뢰도 향상에 초점을 맞추면 Security 및 Reputation 의 종합 신뢰도를 향상시킬 수 있습니다 .

"구성 타당도 행렬" 보고서는 사용자가 생각하는 대상을 잠재 변수가 측정하고 있는지 여부를 판별할 수 있는 방법을 제공합니다 . 행렬 시각화 그림에서 Privacy, Trust 및 Purchase Intent 의 대각 값이 위쪽 및 오른쪽 값보다 큽니다 . 그러나 Security 와 Reputation 의 경우에는 그렇지 않습니다 . 이는 Security 및 Reputation 관련 질문을 개선하면 Security 및 Reputation 잠재 변수를 측정할 수 있다는 추가적인 증거입니다 .

Security 및 Reputation 관련 질문 중 일부를 삭제하거나 수정하여 설문 조사를 개선할 수 있다는 결론을 내릴 수 있습니다.

보고서 구성 요소에 대한 자세한 내용은 "측정 모형 평가"에서 확인하십시오.

JMP PRO 구조 방정식 모형 플랫폼에 대한 통계 상세 정보

JMP PRO 적합 요약 측도에 대한 통계 상세 정보

JMP PRO AICc 및 BIC

AICc 와 BIC 의 계산식은 다음과 같이 정의됩니다 .

$$\text{AICc} = \text{-2logL} + 2k + \frac{2k(k+1)}{n-(k+1)}$$

$$\text{BIC} = \text{-2logL} + k\ln(n)$$

다음은 각 요소에 대한 설명입니다 .

- -2logL = 음의 로그 가능도의 두 배
- n = 표본 크기
- k = 모수 수

"모형 비교" 보고서의 가능도 기반 측도에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

JMP PRO CFI

CFI(비교 적합 지수) 는 다음과 같이 계산됩니다 .

$$\text{CFI} = \frac{max(\chi_0^2 - df_0, 0) - max(\chi_{min}^2 - df_{min}, 0)}{max(\chi_0^2 - df_0, 0)}$$

다음은 각 요소에 대한 설명입니다 .

- χ_0^2 = 독립 모형의 카이제곱 통계량
- df_0 = 독립 모형의 자유도
- χ_{min}^2 = 적합 모형의 카이제곱 통계량
- $df_{\min}$ = 적합 모형의 자유도

CFI에 대한 자세한 내용은 Bentler(1990) 연구 자료에서 확인하십시오 .

JMP PRO TLI

TLI(Tucker-Lewis 지수) 는 다음과 같이 정의됩니다 .

$$\text{TLI} = \frac{\frac{\chi_0^2}{df_0} - \frac{\chi_{min}^2}{df_{min}}}{\frac{\chi_0^2}{df_0} - 1}$$

다음은 각 요소에 대한 설명입니다.

- χ_0^2 = 독립 모형의 카이제곱 통계량
- df_0 = 독립 모형의 자유도
- χ_{min}^2 = 적합 모형의 카이제곱 통계량
- df_{min} = 적합 모형의 자유도

자세한 내용은 West et al. (2012) 연구 자료에서 확인하십시오.

JMP PRO NFI

Bentler-Bonett NFI(표준 적합 지수) 는 다음과 같이 정의됩니다 .

$$\text{NFI} = \frac{\chi_0^2 - \chi_{min}^2}{\chi_0^2}$$

다음은 각 요소에 대한 설명입니다.

- χ_0^2 = 독립 모형의 카이제곱 통계량
- χ_{min}^2 = 적합 모형의 카이제곱 통계량

자세한 내용은 West et al. (2012) 연구 자료에서 확인하십시오.

JMP PRO 수정 GFI 및 수정 AGFI

수정 GFI(수정 적합도 지수) 는 다음과 같이 정의됩니다 .

$$\text{수정 GFI} = \frac{p}{p + 2\left(\frac{\chi_{min}^2 - df_{min}}{n-1}\right)}$$

다음은 각 요소에 대한 설명입니다 .

- χ^2_{min} = 적합 모형의 카이제곱 통계량
- $df_{\min}$ = 적합 모형의 자유도
- p = 적합 모형에서 관측 변수의 수
- n = 표본 크기

수정 AGFI(수정 조정 적합도 지수) 는 다음과 같이 정의됩니다 .

$$\text{수정 AGFI} = 1 - \frac{p^*}{df_{min}}(1 - RevisedGFI)$$

다음은 각 요소에 대한 설명입니다 .

- p^* = 공분산 행렬 및 관측 변수 평균 벡터의 고유 항목 수
- $df_{\min}$ = 적합 모형의 자유도

자세한 내용은 West et al. (2012) 연구 자료에서 확인하십시오 .

JMP PRO RMSEA

RMSEA(근사의 제곱근 평균 제곱 오차) 는 다음과 같이 계산됩니다 .

$$\text{RMSEA} = \sqrt{\frac{\max(\chi^2_{min} - df_{min}, 0)}{n \times df_{min}}}$$

다음은 각 요소에 대한 설명입니다 .

- n = 표본 크기
- $df_{\min}$ = 적합 모형의 자유도
- χ^2_{min} = 적합 모형의 카이제곱 통계량

RMSEA 에 대한 신뢰 한계는 비중심 카이제곱 분포 $\Phi(x|\lambda, d)$ 의 누적 분포 함수를 사용하여 계산됩니다 . 90% 신뢰 한계는 다음과 같이 계산됩니다 .

$$\text{하한} = \sqrt{\frac{\lambda_L}{n \times df_{min}}}$$

$$\text{상한} = \sqrt{\frac{\lambda_U}{n \times df_{min}}}$$

다음은 각 요소에 대한 설명입니다 .

- λ_L 은 $\Phi(\chi^2_{min} \mid \lambda_L, df_{min}) = 0.95$ 를 충족합니다 .
- λ_U 는 $\Phi(\chi^2_{min} \mid \lambda_U, df_{min}) = 0.05$ 를 충족합니다 .

자세한 내용은 Maydeu-Olivares et al. (2017) 연구 자료에서 확인하십시오 .

JMP PRO RMR 및 SRMR

RMR 및 SRMR 계산식은 다음과 같이 정의됩니다 .

$$\text{RMR} = \sqrt{\frac{1}{b}\left[\sum_{i}^{p}\sum_{j}^{i}(s_{ij} - \hat{\sigma}_{ij})^2 + \sum_{i}^{p}(\bar{x}_i - \hat{\mu}_i)^2\right]}$$

$$\text{SRMR} = \sqrt{\frac{1}{b}\left[\sum_{i}^{p}\sum_{j}^{i}\frac{(s_{ij} - \hat{\sigma}_{ij})^2}{s_{ii}s_{jj}} + \sum_{i}^{p}\frac{(\bar{x}_i - \hat{\mu}_i)^2}{s_{ii}}\right]}$$

다음은 각 요소에 대한 설명입니다 .

- p = 명시 변수의 수
- b = 공분산 행렬 및 관측 변수 평균 벡터의 고유 항목 수

$$b = \frac{p(p+1)}{2} + p$$

- s_{ij} = 입력 공분산 행렬의 (i, j) 번째 요소
- $\hat{\sigma}_{ij}$ = 예측 공분산 행렬의 (i, j) 번째 요소
- $\bar{x}_i$ = 표본 평균 벡터의 i 번째 요소
- $\hat{\mu}_i$ = 예측 평균 벡터의 i 번째 요소

자세한 내용은 SAS Institute Inc. (2020a) 의 "CALIS Procedure" 장에서 확인하십시오 .

9 장

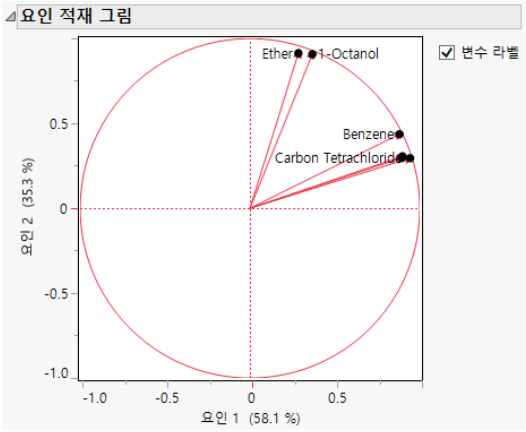
요인 분석
데이터의 잠재 변수 식별

요인 분석은 관측된 변수를 더 적은 수의 관측 불가능한 잠재 변수 또는 요인의 관점에서 설명하려고 합니다. 요인 분석은 공통 요인 분석 및 탐색 요인 분석이라고도 합니다. 요인은 관측 변수(오차 포함)의 선형 결합으로 정의됩니다. 이러한 요인은 관측 변수에 **공통된** 변동을 설명하기 위해 구성됩니다. 요인 분석의 목적은 관측되지 않은 요인의 관점에서 관측 변수에 대한 의미 있는 해석을 찾는 것입니다. 변수 수를 줄이는 목적도 있습니다.

요인 분석은 심리학, 사회학 및 교육에 기원을 둔 여러 분야에 사용됩니다. 이러한 분야에서 요인 분석은 관측된 행동이 기본 패턴 및 구조 측면에서 어떻게 해석될 수 있는지 이해하는 데 사용됩니다. 예를 들어 외부 활동 참여, 취미, 운동 및 여행 측도는 모두 "적극적인 성격 유형과 소극적인 성격 유형"으로 설명될 수 있는 요인과 관련이 있을 수 있습니다.

데이터의 기본 패턴과 구조를 탐색하거나 해석해야 하는 경우 요인 분석을 사용합니다. 또한 더 적은 수의 잠재 변수를 사용하여 변수의 정보를 요약할 때도 사용해 보십시오.

그림 9.1 회전된 요인 적재



목차

요인 분석 플랫폼 개요 207
요인 분석 플랫폼의 예 207
요인 분석 플랫폼 시작 210
요인 분석 보고서 211
    모형 시작 212
    회전 방법 213
요인 분석 플랫폼 옵션 214
요인 분석 모형 적합 옵션 215

요인 분석 플랫폼 개요

요인 분석은 더 적은 수의 관측 불가능한 요인의 관점에서 관측 가능 변수 집합을 모델링합니다. 요인은 관측 변수 간의 상관 또는 공분산을 설명하기 위해 구성됩니다. 요인의 해석력을 높이기 위해 요인 회전을 사용하여 요인의 기준 축을 변경할 수 있습니다.

10 개의 관측 변수 $X_1, X_2, \ldots, X_{10}$ 이 있는 경우를 가정해 보겠습니다 . 이러한 10 개의 변수를 F_1 과 F_2 라는 두 잠재 요인의 관점에서 모델링하려고 합니다 . 편의상 요인은 상관관계가 없으며 각 요인의 평균은 0, 분산은 1 이라고 가정합니다 . 도출하려는 모형의 형식은 다음과 같습니다 .

$$X_i = \beta_{i0} + \beta_{i1}F_1 + \beta_{i2}F_2 + \varepsilon_i$$

따라서 $Var(X_i) = \beta_{i1}^2 + \beta_{i2}^2 + Var(\varepsilon_i)$ 입니다 . X_i 의 분산 중 요인 , 즉 공통 분산 또는 **공통분**으로 인한 부분은 $\beta_{i1}^2 + \beta_{i2}^2$ 입니다 . 나머지 분산 $Var(\varepsilon_i)$ 는 고유 분산이며 X_i 에 고유한 특정 분산과 오차 분산의 조합으로 간주됩니다 .

이 플랫폼에서는 상관 행렬 또는 공분산 행렬의 고유값에 대한 스크리 그림을 제공합니다 . 이 그림을 참고로 사용하여 추출할 요인 수를 결정할 수 있습니다 . 플랫폼의 기본 요인 수는 1 을 초과하는 고유값 수입니다 .

이 플랫폼에서는 이 모형의 모수를 추정하기 위해 "주 축"과 "최대 가능도"라는 두 가지 요인화 방법을 제공합니다 . 각 변수의 공통 요인이 기여하는 분산 비율을 추정하기 위한 두 가지 "사전 공통분" 옵션이 있습니다 . 이러한 옵션은 상관 행렬 또는 공분산 행렬의 대각에 대한 가정을 설정합니다 . "주성분" 옵션은 대각의 값이 1 인 상관 행렬 또는 대각의 값이 분산인 공분산 행렬을 분석할 구조로 처리합니다 . "공통 요인 분석" 옵션은 대각 항목을 SMC(다중상관제곱) 로 설정합니다 . 이러한 값은 다른 변수와 공유하는 변동 비율을 반영합니다 .

요인 회전은 추출된 요인의 해석력을 뒷받침하는 데 사용됩니다 . 요인 분석 플랫폼에서는 직교 회전과 사각 회전을 모두 포함하는 다양한 회전 방법을 제공합니다 .

공통 분산을 고려하는 요인 분석과 달리 주성분 분석은 관측 변수의 총 분산을 설명합니다 . 자세한 내용은 "주성분"에서 확인하십시오 .

요인 분석에 대한 자세한 내용은 Jöreskog(1977) 또는 Cudeck 과 MacCallum(2007) 연구 자료에서 확인하십시오 .

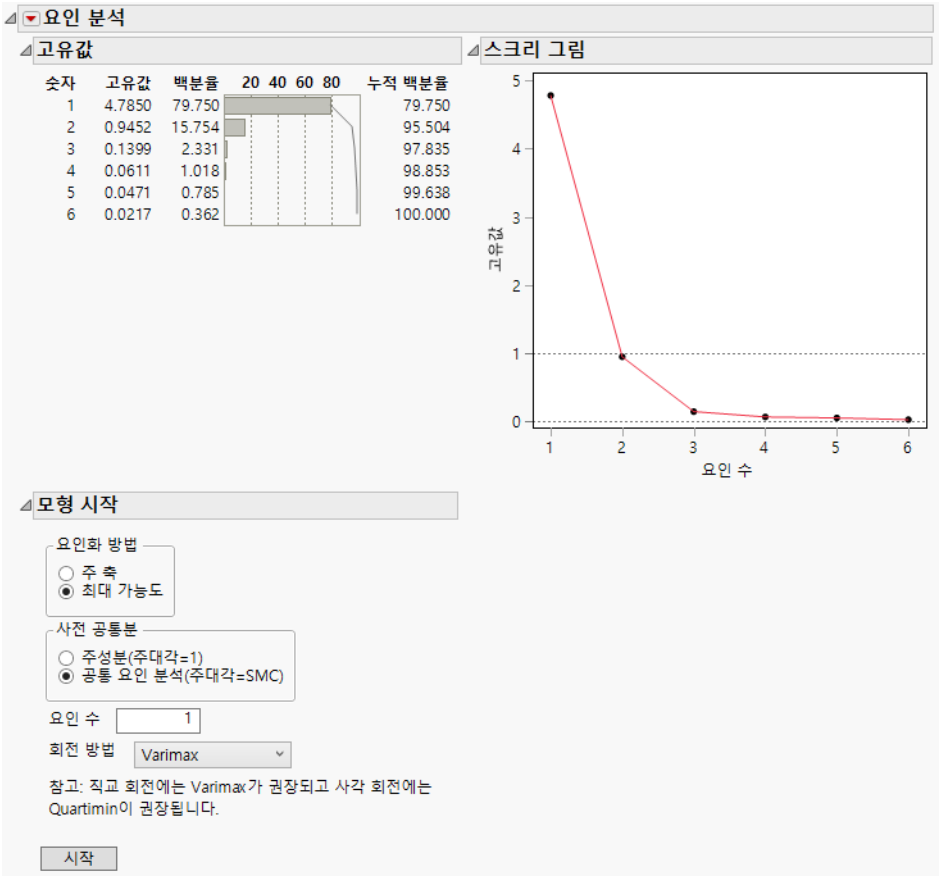
요인 분석 플랫폼의 예

이 예에서는 Solubility.jmp 샘플 데이터 테이블을 사용하여 6 개의 용매로 설명되는 두 가지 요인을 추출합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Solubility.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 요인 분석**을 선택합니다 .
3. 1-Octanol ~ Hexane 을 선택하고 **Y, 열**을 클릭합니다 .

4. **확인**을 클릭합니다 .

그림 9.2 초기 요인 분석 보고서



스크리 그림의 팔꿈치 부분을 보면 요인이 두 개인 모형이 적절하다는 것을 알 수 있습니다.

5. " 모형 시작 " 개요에서 다음과 같이 선택합니다 .
   - 요인화 방법 = **최대 가능도**
   - 사전 공통분 = **공통 요인 분석**
   - 요인 수 = 2
   - 회전 방법 = **Varimax**
6. **시작**을 클릭합니다 .

그림 9.3 요인 분석 보고서

요인 분석

고유값 스크리 그림

모형 시작

요인 분석: 상관계수 행렬 2개 요인 포함: 최대 가능도 / Varimax

최종 공통분 추정값

각 요인에 의해 설명된 분산

유의성 검정

검정	DF	카이제곱	Prob>ChiSq
H0: 공통 요인이 없습니다.	15	691.106	<.0001*
HA: 공통 요인이 하나 이상 있습니다.			

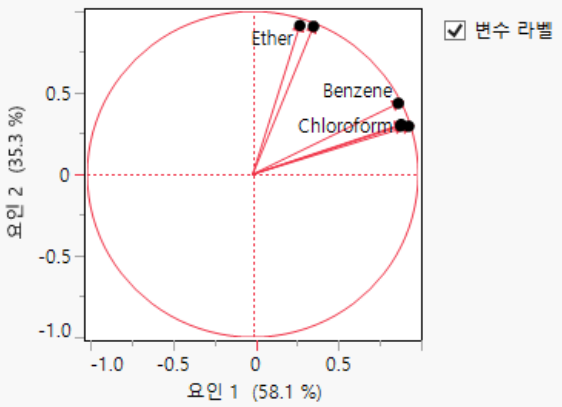
검정	DF	기준	카이제곱	Prob>ChiSq
H0: 2개 요인은 충분합니다.	4	0.389	26.031	<.0001*
HA: 더 많은 요인이 필요합니다.				

적합 측도

적합 측도	적합 지수
Bartlett 수정이 없는 카이제곱	27.654
AIC	19.654
BIC	10.547
Tucker-Lewis 지수	0.878
근사의 제곱근 평균 제곱 오차	0.289

회전된 요인 적재

요인 적재 그림



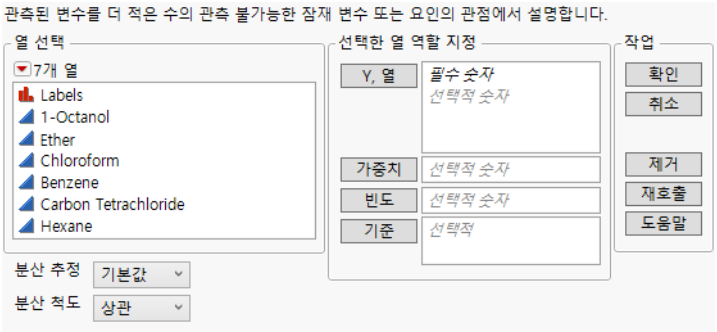
보고서에는 공통분 추정값, 분산 추정값, 유의성 검정, 적합 측도, 회전된 요인 적재 및 요인 적재 그림이 나열됩니다. "회전된 요인 적재" 및 "요인 적재 그림"에서는 요인 1이 Carbon Tetrachloride-Chloroform-Benzene-Hexane 변수 군집과 관련되고, 요인 2가 Ether–1-Octanol 변수 군집과 관련된다는 것을 나타냅니다. 보고서에 표시된 정보에 대한 자세한 내용은 "요인 분석 모형 적합 옵션"에서 확인하십시오.

팁 : " 요인 적재 그림 " 의 점을 클릭하여 라벨을 선택하고 이동할 수 있습니다 . 오른쪽 아래 모서리를 클릭하면 라벨을 더 쉽게 볼 수 있도록 그림 크기가 커집니다 .

요인 분석 플랫폼 시작

분석 > 다변량 방법 > 요인 분석을 선택하여 요인 분석 플랫폼을 시작합니다 .

그림 9.4 요인 분석 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 열 분석할 열입니다 . 데이터 유형이 숫자여야 합니다 .

가중치 데이터 테이블의 각 관측값에 대한 가중치를 포함하는 열입니다 . 값이 0 보다 큰 행만 분석에 포함됩니다 .

빈도 분석의 각 행에 빈도를 할당합니다 . 데이터를 요약할 경우에 유용합니다 .

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

분산 추정 분석을 위한 분산 - 공분산 행렬을 추정하는 방법을 나열합니다 . 방법에 대한 자세한 내용은 " 다변량 보고서 " 에서 확인하십시오 .

분산 척도 요인 분석을 수행하기 위한 척도화 방법을 나열합니다 .

상관 상관계수 행렬을 이용한 분석을 수행할 수 있는 기본 방법입니다 .

공분산 가중치가 변수의 분산인 가중 상관 행렬을 이용한 분석을 수행할 수 있습니다 .

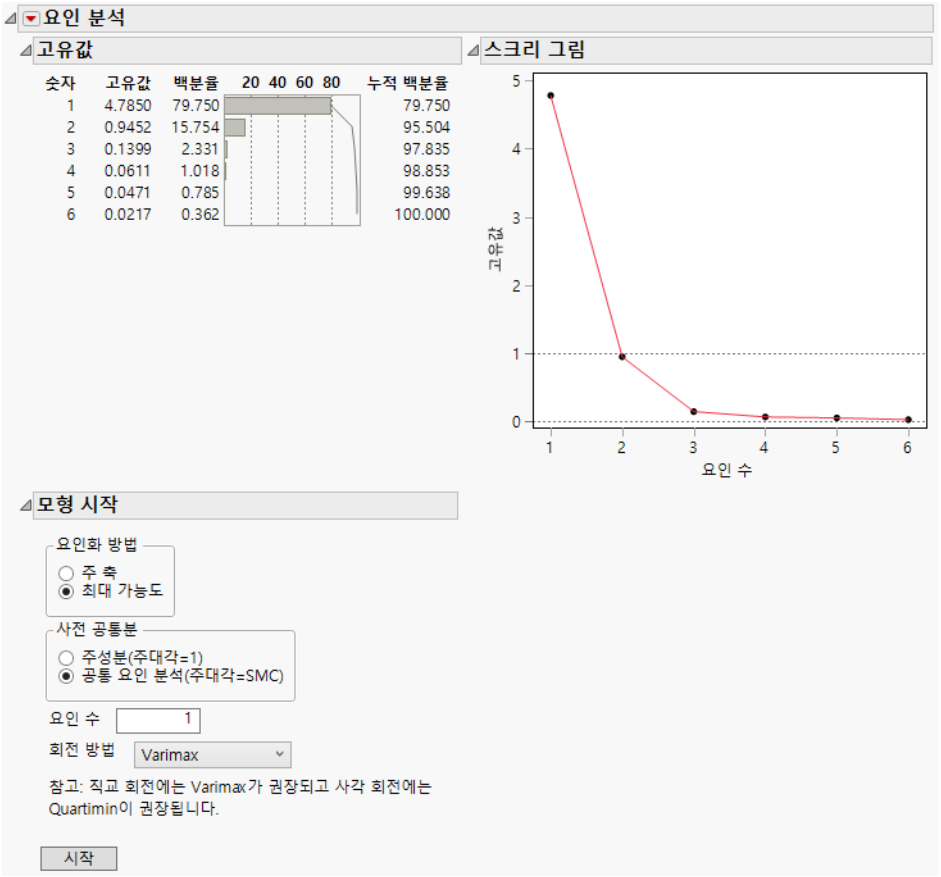
비척도화 이미 중심화 및 척도화된 변수를 분석할 수 있습니다 .

요인 분석 보고서

초기 요인 분석 보고서에는 고유값과 스크리 그림이 표시됩니다. 고유값은 주성분 분석을 통해 계산됩니다. 스크리 그림에는 이러한 고유값이 그래프로 표시됩니다. "모형 시작"의 초기 요인 수는 1.0을 초과하는 고유값 수와 같습니다. 추출할 요인 수를 변경할 수 있습니다.

요인 수를 결정할 때 스크리 그림을 참고로 사용할 수 있습니다. 스크리 그림이 평평하게 안정되기 전의 고유값 수는 요인 수에 대한 상한을 제공합니다. 이 예에서는 스크리 그림에 의해 두 개의 성분(요인)이 제안됩니다.

그림 9.5 요인 분석 보고서

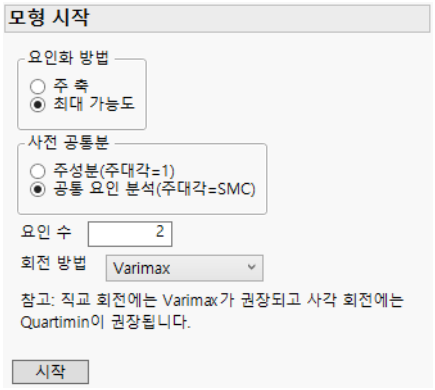


"고유값" 테이블을 보면 첫 번째 고유값이 변동의 79.75%를 설명하고 두 번째 고유값이 15.75%를 설명한다는 것을 알 수 있습니다. 따라서 처음 두 고유값이 총 변동의 95.50%를 설명합니다. 세 번째 고유값은 변동의 2.33%만 설명하고 나머지 고유값의 기여도는 무시해도 될 정도의 수준입니다. **요인 수** 상자는 처음에 1로 설정되었지만 이 분석을 통해 2개 요인을 추출하는 것이 적절하다는 것을 알 수 있습니다.

모형 시작

"모형 시작" 제어판에서 요인 분석을 구성합니다. **시작**을 클릭하면 요인 분석 결과를 얻을 수 있습니다.

그림 9.6 모형 시작



요인화 방법 요인 추출 방법을 정의합니다.

주 축 행렬의 대각 항목이 변수의 공통분 추정값으로 대체된 축소 상관 행렬 또는 축소 공분산 행렬에 대해 고유값 분해를 수행합니다. 이 방법은 계산이 효율적이지만 가설 검정을 허용하지 않습니다.

최대 가능도 공통 요인 수에 대한 가설을 검정하고 모형 적합 통계량을 구할 수 있습니다.

참고 : 최대 가능도 방법을 사용하려면 양정치 상관 행렬이 필요합니다. 상관 행렬이 양정치가 아닌 경우 주 축 방법을 선택하십시오.

사전 공통분 각 변수의 공통 요인이 기여하는 분산 비율을 추정하는 방법을 정의합니다.

주성분 (대각 = 1) 모든 공통분을 1 로 설정하여 모든 요인에 의해 각 변수의 분산이 100% 설명된다는 것을 나타냅니다.

팁 : "요인화 방법"을 "주 축"으로 설정하고 이 옵션을 사용하면 주성분 분석이 수행됩니다.

공통 요인 분석 (대각 = SMC) 공통분을 SMC(다중상관제곱) 계수로 동일하게 설정합니다. 변수가 주어지면 SMC 는 다른 모든 변수에 대한 해당 변수의 회귀 분석 R^2 입니다.

요인 수 분석에서 추출할 요인 수를 지정합니다. 기본값은 1.0 보다 크거나 같은 고유값 수입니다. 요인 수를 1 이상으로 설정할 수 있으며 변수의 수를 초과할 수 없습니다.

회전 방법 회전 방법을 정의합니다. 기본값은 "Varimax" 입니다. 사용 가능한 회전 방법에 대한 설명은 " 회전 방법 " 에서 확인하십시오.

회전 방법

회전은 요인을 더 쉽게 해석할 수 있도록 하기 위해 요인의 기준 축을 변경하는 데 사용됩니다. 데이터에서 추출된 요인에 회전이 적용됩니다. 회전 방법은 다양한 복잡성 또는 단순성 함수를 기반으로 합니다. 회전에 대한 자세한 내용은 SAS Institute Inc. (2020c)의 "FACTOR Procedure" 장, Browne(2001) 또는 Frank와 Todeschini(1994) 연구 자료에서 확인하십시오.

초기 추출 후에는 요인 간에 상관관계가 없습니다. 직교 변환에 의해 요인이 회전하는 경우 회전된 요인도 상관관계가 없습니다. 사각 변환에 의해 요인이 회전하는 경우 회전된 요인은 상관관계를 가집니다. 사각 회전이 직교 회전보다 해석 가능한 요인을 더 많이 생성합니다. 그러나 요인이 상관되면 변수를 설명할 때 요인 중요도를 결정하는 명확한 측도가 없어집니다.

직교 회전 방법

Varimax 모든 변수에 대해 요인 적재 제곱의 분산 합계를 최대화합니다. 가장 일반적인 이 방법을 사용하면 각 요인에 대한 각 변수의 적재 값이 작거나 커집니다. $\gamma = 1$ 인 Orthomax 와 같습니다.

Biquartimax Varimax 및 Quartimax 회전의 균등 가중 해입니다. $\gamma = 0.5$ 인 Orthomax 와 같습니다.

Equamax Varimax 회전과 Quartimax 회전 사이의 가중 해입니다. $\gamma = N/2$ 인 Orthomax 와 같습니다 (N = 요인 수).

Factorparsimax 요인 복잡성을 최소화하는 것이 목표인 해입니다. 변수 복잡성은 알고리즘에서 고려되지 않으므로 이 방법을 사용하면 교차 적재가 발생할 수 있습니다. $\gamma = N$ 인 Orthomax 와 같습니다 (N = 요인 수).

Orthomax 가중치가 γ 로 표시되는 일반적인 가중 회전 방법입니다. 대부분의 특정 직교 회전 방법은 특정 γ 를 사용하는 Orthomax 회전입니다.

Parsimax 변수와 요인 복잡성 간에 균형을 맞춥니다. $\gamma = (I(N-1))/(I+N-2)$ 인 Orthomax 와 같습니다 (I = 항목 수, N = 요인 수).

Quartimax 각 변수를 설명하는 데 필요한 요인 수를 최소화합니다. $\gamma = 1$ 인 Orthomax 와 같습니다.

사각 회전 방법

Biquartimin 공분산 비율을 최소화하기 위한 회전입니다. $\tau = 0.5$ 인 Oblimin 과 같습니다.

Covarimin 사각 Varimax 회전입니다. $\tau = 1$ 인 Oblimin 과 같습니다.

Obbiquartimax 사각 Biquartimax 회전입니다.

Obequamax 사각 Equamax 회전입니다.

Obfactorparsimax 사각 요인 Parsimax 회전입니다.

Oblimin 가중치가 τ 로 표시되는 일반적인 가중 사각 회전 방법입니다. 대부분의 특정 사각 회전 방법은 특정 τ 를 사용하는 Oblimin 회전입니다.

Obparsimax 사각 Parsimax 회전입니다.

Obquartimax 사각 Quartimax 회전입니다 (Quartimin 방법과 동등).

Obvarimax 사각 Varimax 회전입니다 .

Quartimin 사각 Quartimin 회전입니다(사각 Quartimax와 동등). $\tau = 0$인 Oblimin과 같습니다.

Promax Varimax 를 먼저 수행한 후 Procrustes 회전을 사용하여 간단한 구조를 얻는 2 단계 회전입니다 . 이 방법은 Oblimin 대신 사용할 수 있으며 계산이 효율적입니다 .

요인 분석 플랫폼 옵션

" 요인 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

고유값 원래 상관 행렬, 공분산 행렬 또는 비척도화 행렬의 고유값 테이블을 표시하거나 숨깁니다 . 이 테이블에는 각 고유값이 나타내는 총 분산의 백분율, 기여도 백분율을 보여 주는 막대 차트 및 연속된 각 고유값이 기여하는 누적 백분율이 포함됩니다 . 1.0 보다 크거나 같은 고유값 수를 분석할 요인 수의 지표로 사용할 수 있습니다 .

스크리 그림 고유값 대 성분 (요인) 수를 나타내는 그림을 표시하거나 숨깁니다 . 이 그림을 추가 지표로 사용하여 최대 분산 크기에 기여하는 요인 수를 결정할 수 있습니다 . 그림에 표시된 선이 평평하게 안정되는 점을 분석에 필요한 충분 요인 수로 사용할 수 있습니다 .

Kaiser-Meyer-Olkin 검정 KMO(Kaiser-Meyer-Olkin) 검정 결과를 표시하거나 숨깁니다 . 이 검정은 기본 요인에 의한 공통 분산일 수 있는 분산 비율의 지표입니다 . 검정 통계량은 각 개별 변수와 모든 변수 집합에 대해 계산되며 , JMP 에서는 MSA(표집 적합성 측도) 라고 합니다 . 다음 값은 요인 분석에 대한 변수의 적합성을 나타냅니다 .

- 0.00 ~ 0.49 허용 불가
- 0.50 ~ 0.59 매우 나쁨
- 0.60 ~ 0.69 조금 나쁨
- 0.70 ~ 0.79 보통
- 0.80 ~ 0.89 조금 좋음
- 0.90 ~ 1.00 매우 좋음

참고 : 상관 행렬이 특이 행렬인 경우 KMO 검정을 사용할 수 없습니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

요인 분석 모형 적합 옵션

요인 분석 모형 적합의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

사전 공통분 (공통 요인 분석에만 사용 가능) 각 변수에 대한 공통분의 초기 추정값을 표시하거나 숨깁니다. 변수가 주어지면 이 추정값은 SMC(다중상관제곱) 계수이거나, 다른 모든 변수에 대한 해당 변수의 회귀 분석 R^2 입니다.

고유값 (공통 요인 분석에만 사용 가능) 축소 상관 행렬의 고유값과 해당 값이 설명하는 공통 분산의 백분율을 표시하거나 숨깁니다. 축소 상관 행렬은 대각 항목이 공통분 추정값으로 대체된 상관 행렬입니다. 고유값은 요인에 의해 설명되는 공통 분산을 나타냅니다. 축소 상관 행렬은 양정치가 아닐 수도 있고 음수 고유값을 가질 수 있으므로 "누적 백분율"이 100%를 초과할 수 있습니다.

이 테이블에는 분석을 위해 유지된 요인 수가 나타납니다.

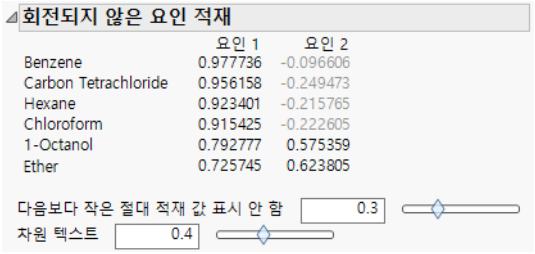
정렬되지 않고 회전되지 않은 요인 적재 정렬과 회전을 수행하기 전의 요인 적재 행렬을 표시하거나 숨깁니다.

회전되지 않은 요인 적재 회전하기 전의 요인 적재 행렬을 표시하거나 숨깁니다. 요인 적재는 공통 요인이 변수에 미치는 영향을 측정합니다. 회전되지 않은 요인은 직교하므로 요인 적재 행렬은 변수와 요인 간의 상관 행렬입니다. 적재 절대값이 1에 가까울수록 변수에 대한 요인의 효과가 커집니다.

테이블에서 슬라이더와 **다음보다 작은 절대 적재 값 표시 안 함**의 값을 사용하여 지정된 값 미만인 값을 표시하지 않을 수 있습니다. 표시되지 않는 값은 **텍스트 흐리게**에 지정한 설정에 따라 흐리게 나타납니다.

텍스트 흐리게 슬라이더와 값을 사용하여 **다음보다 작은 절대 적재 값 표시 안 함** 값보다 절대값이 작은 요인 값에 대한 테이블의 글꼴 투명도 그래디언트를 제어할 수 있습니다.

그림 9.7 텍스트 흐리게 컨트롤이 있는 회전되지 않은 요인 적재



참고 : " 회전되지 않은 요인 적재 " 행렬은 동일한 요인과 관련된 변수가 서로 옆에 나타나도록 정렬됩니다 .

회전 행렬 요인 적재 그림과 요인 적재 행렬을 회전하는 데 사용된 값을 표시하거나 숨깁니다 .

요인 간 상관 (사각 회전에만 사용 가능) 요인 간 상관 행렬을 표시하거나 숨깁니다 .

목표 행렬 (Promax 회전에만 사용 가능) Varimax 요인 패턴이 회전되는 행렬을 표시하거나 숨깁니다 .

요인 구조 (사각 회전에만 사용 가능) 변수와 공통 요인 간의 상관 행렬을 표시하거나 숨깁니다.

최종 공통분 추정값 요인 모형이 적합된 후 공통분 추정값을 표시하거나 숨깁니다 . 요인이 직교할 경우 변수의 최종 공통분 추정값은 해당 변수의 적재 제곱합과 같습니다 .

표준 스코어 계수 회전된 요인을 소스 데이터 테이블에 저장할 때 요인 스코어를 추정하는 데 사용되는 승수 테이블을 표시하거나 숨깁니다 .

각 요인에 의해 설명된 분산 (직교 회전에만 사용 가능) 회전된 각 요인에 의해 설명되는 공통 분산의 분산 , 백분율 및 누적 백분율을 표시하거나 숨깁니다 .

다른 요인을 무시하고 각 요인에 의해 설명된 분산 (사각 회전에만 사용 가능) 다른 요인에 관계없이 회전된 각 요인에 의해 설명되는 공통 분산의 분산 및 백분율을 표시하거나 숨깁니다 .

유의성 검정 (최대 가능도 요인화 방법에만 사용 가능) 두 개의 카이제곱 검정 결과를 제공합니다 .

첫 번째 검정은 "H_0: 공통 요인이 없습니다 ." 에 대한 것입니다 . 이 귀무가설은 변수 간 상호 상관관계를 설명하는 공통 요인이 없음을 나타냅니다 . 이 검정은 요인의 상관 행렬이 단위 행렬이라는 귀무가설을 사용하는 구형성 Bartlett 검정입니다 (Bartlett, 1954 참조).

두 번째 검정은 "H_0: *N* 개 요인은 충분합니다 ." 에 대한 것입니다 . 여기서 *N* 은 지정된 요인 수입니다 . 이 귀무가설의 기각은 변수 간 상호 상관관계를 설명하기 위해 요인이 더 필요할 수 있음을 나타냅니다 (Bartlett, 1954 참조). " 기준 " 은 로그 가능도 목적 함수 값입니다 .

적합 측도 (최대 가능도 요인화 방법에만 사용 가능) "Bartlett 수정이 없는 카이제곱 ", "AIC", "BIC", "Tucker-Lewis 지수 " 및 " 근사의 제곱근 평균 제곱 오차 " 와 같은 적합 측도를 표시하거나 숨깁니다 .

요인 스코어 측도 " 다중 R", " 다중 R^2" 및 " 최소 상관 " 과 같은 요인 스코어 결정 측도를 표시하거나 숨깁니다 . 이러한 측도는 요인 스코어가 2 차 분석에 유용한지 여부를 평가하는 데 사용됩니다 .

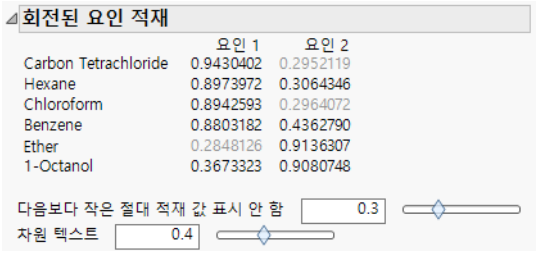
정렬되지 않고 회전된 요인 적재 회전 후 정렬되지 않은 요인 적재 행렬을 표시하거나 숨깁니다 .

회전된 요인 적재 회전 후 요인 적재 행렬을 표시하거나 숨깁니다 . 직교 회전을 사용한 경우 이러한 값은 변수와 회전된 요인 간의 상관입니다 .

테이블에서 슬라이더와 **다음보다 작은 절대 적재 값 표시 안 함**의 값을 사용하여 지정된 값 미만인 값을 표시하지 않을 수 있습니다 . 표시되지 않는 값은 **텍스트 흐리게**에 지정한 설정에 따라 흐리게 나타납니다 .

텍스트 흐리게 슬라이더와 값을 사용하여 테이블의 글꼴 투명도 그래디언트를 제어할 수 있습니다 . 값이 낮을수록 **다음보다 작은 절대 적재 값 표시 안 함**에 지정된 값보다 절대값이 작은 요인 값의 글꼴이 더 투명해집니다 .

그림 9.8 텍스트 흐리게 컨트롤이 있는 회전된 요인 적재



참고 : " 회전된 요인 적재 " 행렬은 동일한 요인과 관련된 변수가 서로 옆에 나타나도록 정렬됩니다 .

요인 적재 그림 회전된 요인 적재 그림을 표시하거나 숨깁니다 . 요인을 세 개 이상 모델링하는 경우 적재 그림은 그림의 행렬입니다 .

스코어 그림 추정된 요인 스코어의 산점도를 표시하거나 숨깁니다 . 요인을 세 개 이상 모델링하는 경우 스코어 그림은 그림의 행렬입니다 .

결측값 대치를 사용한 스코어 그림 (결측값이 있는 경우에만 사용 가능) 결측값을 대치 값으로 채우고 추정된 요인 스코어의 산점도를 표시하거나 숨깁니다 .

표시 옵션 적재 그림에 화살표를 표시하거나 숨길 수 있습니다 .

회전된 요인 저장 회전된 요인 스코어와 계산식을 데이터 테이블에 저장합니다 .

참고 : 계산식은 결측값이 있는 행을 계산할 수 없습니다 .

적합 제거 " 요인 분석 " 보고서에서 모형 적합 결과를 제거합니다 . 이 옵션을 사용하면 새 보고서에 대한 " 모형 시작 " 구성을 변경할 수 있습니다 .

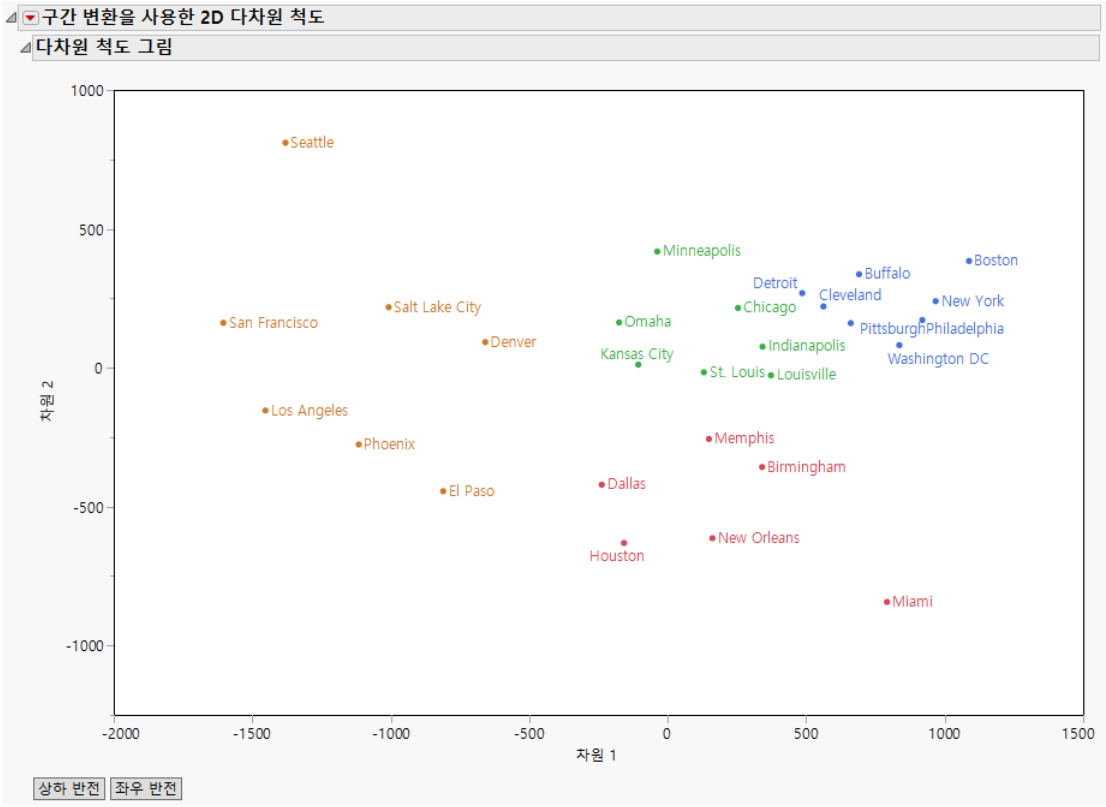
10 장

다차원 척도법
개체 집합 간의 근접성 시각화

MDS(다차원 척도법)는 개체 집합 간의 근접성 패턴(유사성, 비유사성 또는 거리)을 시각적으로 표현하는 데 사용되는 기법입니다. 예를 들어 도시 간의 거리 행렬이 주어진 경우 MDS를 사용하여 2차원 도시 맵을 생성할 수 있습니다.

다차원 척도법은 연구원이 브랜드, 취향 또는 기타 상품 속성과 관련된 인식 측도를 가지고 있는 소비자 조사에 주로 사용됩니다. MDS는 근접성 또는 속성 집합을 기반으로 개체의 근접성을 시각화하는 데 관심이 있는 다른 많은 분야에 적용할 수 있습니다.

그림 10.1 다차원 척도법의 예



목차

다차원 척도법 플랫폼 개요 221
다차원 척도법의 예 221
다차원 척도법 플랫폼 시작 224
다차원 척도 보고서 225
다차원 척도 그림 225
Shepard 다이어그램 225
적합 상세 정보 226
다차원 척도법 플랫폼 옵션 226
Waern 링크 227
다차원 척도법 플랫폼의 추가 예 228
다차원 척도법 플랫폼에 대한 통계 상세 정보 230
스트레스 230
변환 230
속성 목록 형식 231

다차원 척도법 플랫폼 개요

다차원 척도법 플랫폼에서는 개체 집합 간의 근접성 그림을 생성합니다. 이 그림을 사용하여 데이터 집합의 구조를 시각적으로 탐색할 수 있습니다. MDS는 적은 수의 차원에서 개체 집합 간의 근접성 패턴(거리, 유사성)을 시각화하는 데 사용되는 다변량 기법입니다. MDS는 거리 행렬에 적용됩니다. 스트레스 함수(실제 근접성과 예측된 근접성의 차이)를 최소화하여 MDS 그림의 좌표를 얻습니다.

거리라는 용어가 물리적 거리의 측도를 나타낼 수 있습니다(예: 도시 간). 거리는 정확한 측정보다는 주관적 평가인 경우가 더 많습니다. 근접성은 제품 브랜드 간의 인지된 유사성, 범죄율의 상관관계 또는 표본 국가에 대한 경제적 유사성을 측정할 수 있습니다. 거리를 근접성 또는 유사성(비유사성)이라고 할 수도 있습니다. 속성 목록 데이터가 주어진 경우 먼저 속성 목록에서 거리 행렬이 생성됩니다.

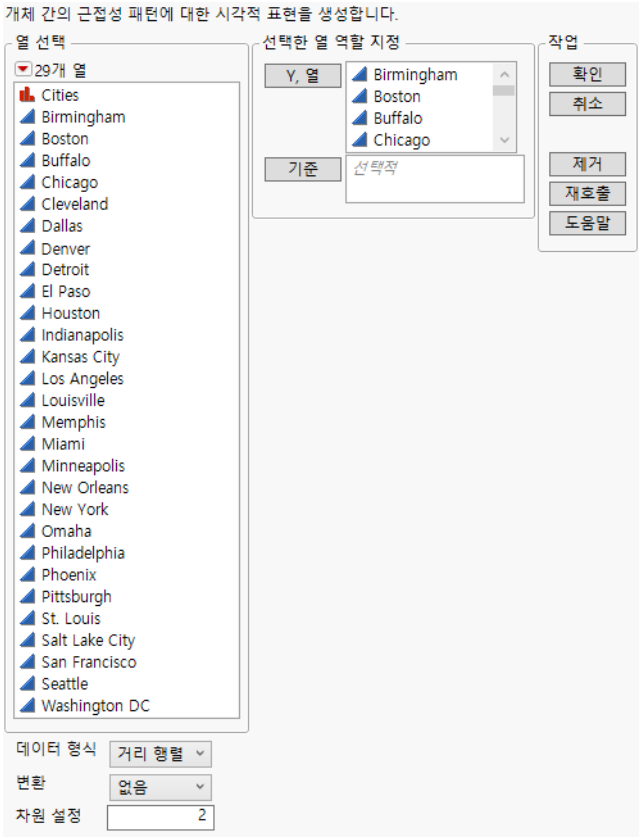
다차원 척도법에 대한 자세한 내용은 Borg와 Groenen(2005) 또는 Jackson(2003) 연구 자료에서 확인하십시오.

다차원 척도법의 예

이 예에서는 미국 28 개 도시 간의 비행 거리 행렬을 나타내는 Flight Distances.jmp 샘플 데이터 테이블을 사용합니다 . MDS 를 사용하여 데이터 테이블의 쌍별 거리를 기반으로 하는 2 차원 도시 맵을 생성할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Flight Distances.jmp 를 엽니다 .
2. **분석 > 다변량 방법 > 다차원 척도법**을 선택합니다 .
3. Birmingham ~ Washington DC 를 선택하고 **Y, 열**을 클릭합니다 .

그림 10.2 완료된 다차원 척도법 시작 창



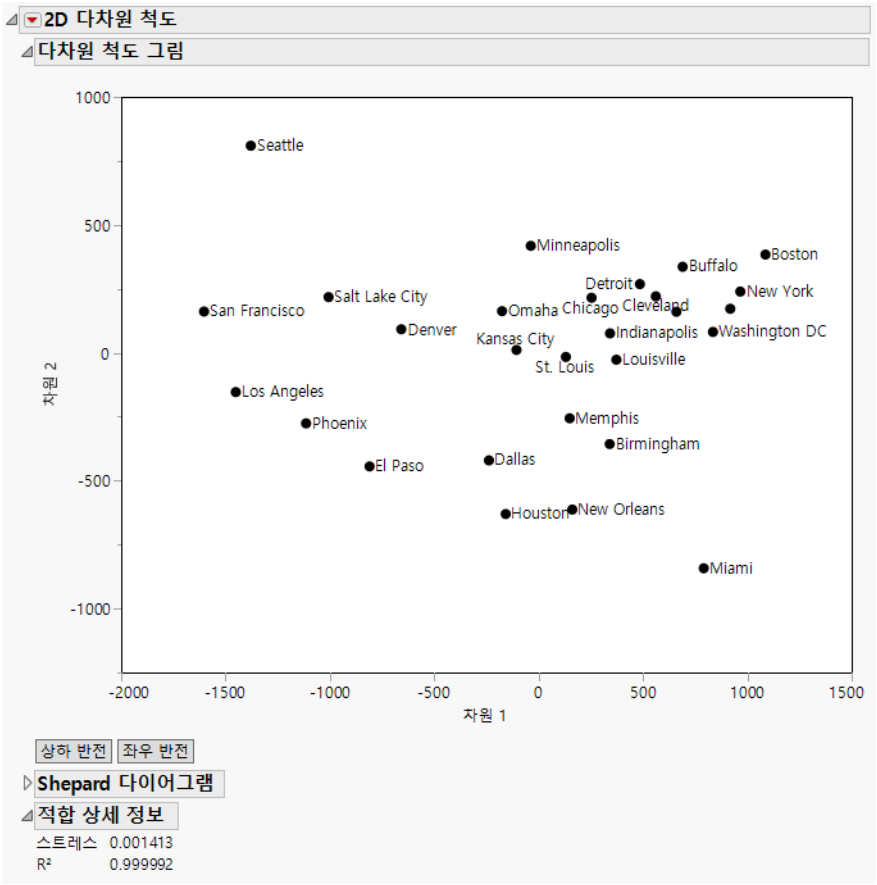
4. **확인**을 클릭합니다.

   " 다차원 척도 그림 " 에서 데이터 점을 커서로 가리키면 행 번호 또는 행 라벨이 표시됩니다. 다음 7 개 단계를 수행하여 MDS 그림에 라벨을 지정하고 그림을 회전합니다.

5. Flight Distances 데이터 테이블을 선택합니다.
6. Cities 열을 마우스 오른쪽 버튼으로 클릭하고 **라벨 / 라벨 해제**를 선택합니다.
7. **행 > 행 선택 > 모든 행 선택**을 선택합니다.
8. **행 > 라벨 / 라벨 해제**를 선택합니다.
9. " 다차원 척도 그림 " 을 선택합니다.
10. **상하 반전** 버튼을 클릭합니다.
11. **좌우 반전** 버튼을 클릭합니다.

    " 상하 반전 " 및 " 좌우 반전 " 버튼을 사용하여 MDS 그림의 방향을 변경할 수 있습니다. MDS 결과는 방향에 영향을 받지 않습니다. 물리적 위치와 같은 알려진 방향이 결과에 있는 경우 그림을 회전하거나 대칭 이동할 수 있습니다.

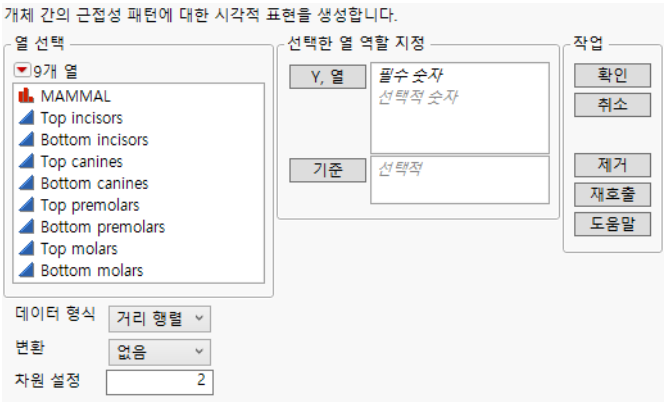
그림 10.3 다차원 척도 그림



다차원 척도법 플랫폼 시작

분석 > 다변량 방법 > 다차원 척도법을 선택하여 다차원 척도법 플랫폼을 시작합니다 .

그림 10.4 다차원 척도법 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 열 분석할 열입니다 . 데이터 유형이 숫자여야 합니다 .

기준 개별 분석을 정의하는 수준이 포함된 하나 이상의 열입니다 . 지정된 열의 각 수준에 대해 해당 행은 사용자가 지정한 다른 변수를 사용하여 분석됩니다 . 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

참고 : 거리 행렬을 사용하는 경우 기준 변수에 해당 기준 변수의 각 수준에 대한 완전 행렬이 필요합니다 .

데이터 형식 MDS 는 다음과 같은 두 가지 데이터 형식을 지원합니다 .

거리 행렬 행 수가 열 수와 동일한 완전 대칭 , 하삼각 또는 상삼각 행렬입니다 . 대각 항목은 0 또는 결측일 수 있습니다 .

속성 목록 개체의 품질 또는 특성에 대한 측도가 포함된 열 집합입니다 . 일반적으로 열에 개체 이름이 지정됩니다 . 개체 열은 분석에 사용되지 않고 MDS 그림에서 데이터 점의 라벨로 사용됩니다 .

변환 지원되는 변환은 " 비 ", " 구간 " 및 " 순서형 " 입니다 .

없음 변환이 사용되지 않습니다 .

비 데이터에 가장 작은 것부터 큰 것까지 순서가 있고 , 값의 차이에 의미가 있으며 , 척도 0 은 절대 0 입니다 . MDS 그림을 척도화하는 데 사용됩니다 .

구간 데이터에 가장 작은 것부터 큰 것까지 순서가 있고 , 값의 차이에 의미가 있습니다 . MDS 그림을 척도화하고 이동하는 데 사용됩니다 .

순서형 데이터에 가장 작은 것부터 큰 것까지 순서가 있습니다. 순서형 데이터에 사용됩니다.

차원 설정 개체 간의 근접성을 시각적으로 표현하기 위한 차원 수입니다 . 일반적으로 2 차원 또는 3 차원이 사용됩니다 . 3 차원보다 크면 시각화가 복잡해집니다 .

참고 : 1 에서 n - 1 사이의 차원 (n = 개체 수) 을 선택할 수 있습니다 . 그렇지 않은 경우 차원이 2 로 설정됩니다 .

다차원 척도 보고서

초기 다차원 척도 보고서에는 "다차원 척도 그림", "Shepard 다이어그램" 및 "적합 상세 정보"가 표시됩니다. 시작 창에서 적합 차원을 3 이상으로 지정하면 표시할 차원을 선택할 수 있는 컨트롤이 "다차원 척도 그림"에 제공됩니다.

MDS 그림에서 서로 가까이 있는 개체는 유사한 속성을 공유합니다. 그림에 라벨과 색상을 추가하면 유사한 그룹을 식별하는 데 도움이 될 수 있습니다. Shepard 다이어그램과 적합 통계량 요약은 MDS 그림이 개체의 근접성을 얼마나 잘 나타내는지를 보여 줍니다.

다차원 척도 그림

MDS 그림에는 다차원 척도가 2차원으로 표시됩니다. 그림 아래에는 축을 상하 또는 좌우로 반전할 수 있는 두 개의 버튼이 있습니다. 점 간 근접성을 변경하지 않고 MDS 해를 반사, 회전 또는 이동할 수 있습니다. 축을 반사 또는 회전하는 것은 맵 방향이 알려진 지리적 개체로 작업할 때 가장 일반적입니다.

분석에 세 개 이상의 차원이 사용된 경우 그림 아래의 **차원 선택** 컨트롤을 사용하여 그림에 표시되는 차원을 전환할 수 있습니다. 첫 번째 컨트롤은 그림의 가로 축을 정의하고 두 번째 컨트롤은 그림의 세로 축을 정의합니다.

Shepard 다이어그램

Shepard 다이어그램은 실제 또는 변환된 근접성 대 예측된 근접성을 보여 주는 그림입니다. 이 그림에는 다차원 척도 그림에 실제 근접성이 얼마나 잘 반영되는지 나타납니다. Shepard 는 실제값 대 예측값 그림과 유사합니다. 빨간색으로 표시된 Y = X 선 위에 점이 있는 그림이 이상적입니다.

적합 상세 정보

" 적합 상세 정보 " 에는 MDS 근접성이 실제 근접성과 얼마나 일치하는지 요약하는 통계량과 사용된 변환 정보가 제공됩니다 .

스트레스 적합 절차에서 최소화된 스트레스 함수 (스트레스 1) 의 값입니다 . 스트레스 값은 0 에서 1 사이일 수 있으며 값이 작을수록 더 나은 적합을 나타냅니다 .

R^2 실제 또는 변환된 근접성 대 예측된 근접성의 선형 적합에 대한 R^2 값입니다 .

기울기 비율 또는 구간 변환이 사용된 경우 변환에 대한 기울기가 제공됩니다 . 변환된 근접성에 대한 실제 근접성의 선형 회귀 기울기입니다 .

절편 구간 변환이 사용된 경우 변환에 대한 절편이 제공됩니다 . 변환된 근접성에 대한 실제 근접성의 선형 회귀 절편입니다 .

다차원 척도법 플랫폼 옵션

" 다차원 척도 " 의 빨간색 삼각형 메뉴 옵션을 사용하면 필요에 따라 보고서를 사용자 정의할 수 있습니다 . 사용할 수 있는 옵션은 분석에 사용하는 데이터 유형 및 차원 수에 따라 결정됩니다 .

MDS 그림 MDS 그림을 표시하거나 숨깁니다 .

진단 MDS 에 대한 진단을 제공합니다 .

Shepard 다이어그램 실제 근접성 (변환이 사용된 경우 변환된 근접성) 대 예측된 근접성 그림을 표시합니다 . 이 보고서는 기본적으로 나타납니다 . 자세한 내용은 "Shepard 다이어그램 " 에서 확인하십시오 .

Waern 링크 MDS 그림에 Waern 링크를 표시합니다 . 이 옵션을 선택하면 링크를 표시할 부분 (가장 작은 부분 또는 가장 큰 부분) 을 결정하는 컨트롤을 사용할 수 있습니다 . 자세한 내용은 "Waern 링크 " 에서 확인하십시오 .

좌표 표시 해 좌표 보고서를 제공합니다 . 이는 " 다차원 척도 그림 " 에 표시된 점의 좌표입니다 . 보고서에는 최대 3 개 차원의 좌표가 표시됩니다 . 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열**을 선택하여 보고서에 차원을 추가할 수 있습니다 . 최대 차원 수는 시작 창에 설정한 차원 수입니다 .

근접성 표시 근접성 보고서를 제공합니다 . 각 개체 쌍 사이의 원래 근접성과 도출된 근접성 (거리) 이 제공됩니다 . 쌍은 " 시작 개체 " 열과 " 끝 개체 " 열에서 식별됩니다 . 변환이 사용된 경우 변환된 근접성도 테이블에 포함됩니다 .

근접성 저장 (데이터 형식이 " 속성 목록 " 인 경우에만 사용 가능) 거리 행렬을 데이터 테이블에 저장합니다 .

3D 그림 (시작 창에서 " 차원 설정 " 에 세 개 이상의 차원을 지정한 경우에만 사용 가능) 처음 세 개 차원의 3D 그림을 표시합니다 .

좌표 저장 해 좌표를 데이터 테이블에 별도의 열로 저장합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

Waern 링크

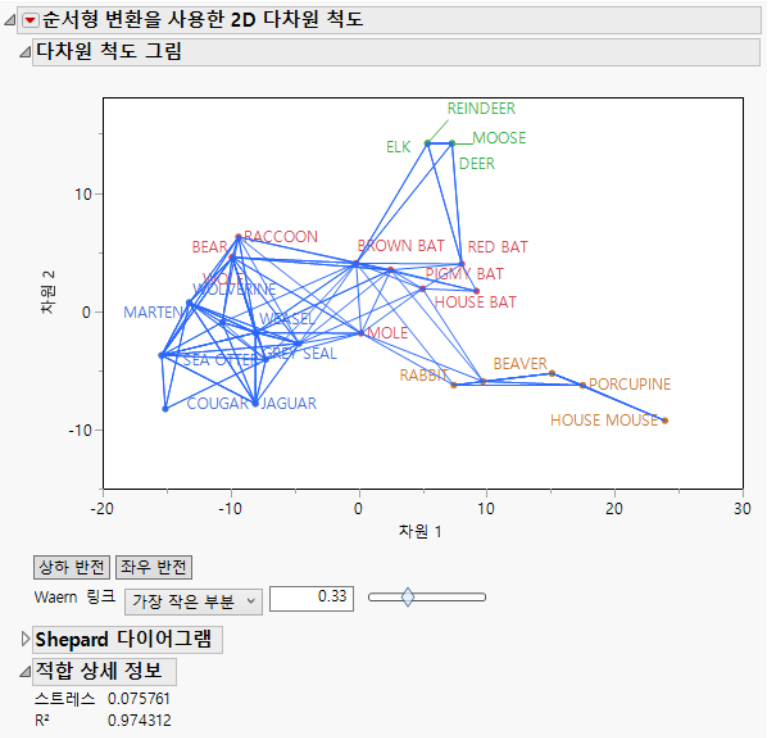
Waern 링크를 표시하면 실제 근접성과 예측된 근접성을 비교하여 MDS 결과를 시각적으로 확인할 수 있습니다. 이 링크는 실제 근접성을 기반으로 다차원 척도 그림의 점을 연결합니다. 근접성이 가장 작거나 가장 큰 개체가 연결됩니다. 일반적으로 개체 간 근접성의 가장 작은 33% 를 고려해야 합니다. MDS 그림이 근접성을 잘 표현하면 가장 작은 실제 근접성의 링크가 그림에서 가장 가까운 개체를 연결해야 합니다. 가장 작은 근접성의 링크가 그림에서 늘어나 멀리 있는 개체를 연결하면 MDS 적합을 의심해 볼 수 있습니다.

Waern 링크 컨트롤

그림에 링크의 가장 작은 부분을 표시할지 아니면 가장 큰 부분을 표시할지 선택할 수 있는 목록이 있습니다. 표시된 링크 부분은 상자에 값을 입력하거나 슬라이더를 사용하여 제어합니다. 그림 10.5에서는 Teeth.jmp 데이터 테이블에 대한 Waern 링크를 보여 줍니다. 이 링크는 가장 작은 33% 부분을 표시합니다.

Waern 링크에 대한 자세한 내용은 Waern(1972) 연구 자료에서 확인하십시오.

그림 10.5 Waern 링크가 있는 MDS 그림



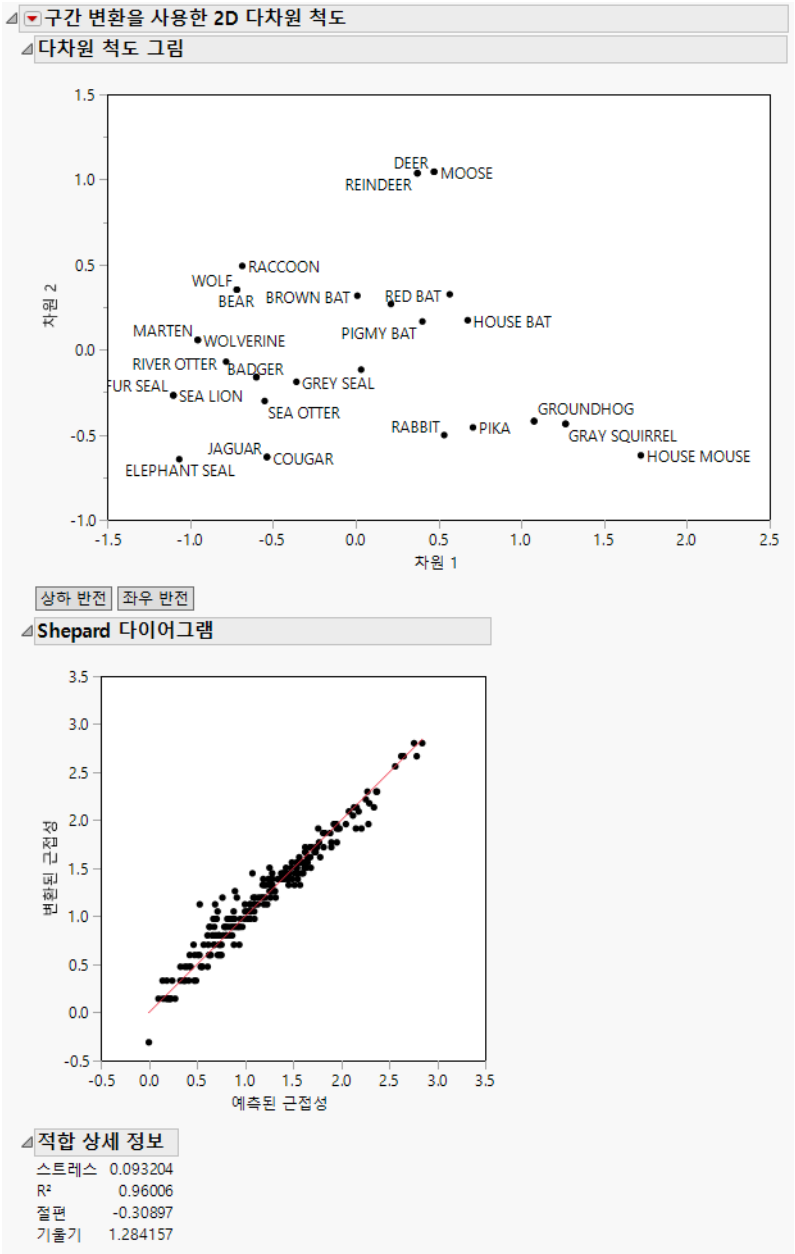
다차원 척도법 플랫폼의 추가 예

이 예에서는 Teeth.jmp 샘플 데이터 테이블을 사용합니다. 이 테이블은 32 마리 포유 동물에 대한 8 가지 치아 종류의 개수가 포함된 속성 목록입니다. MDS 를 사용하여 치아를 기반으로 포유 동물의 유사성을 탐색할 수 있습니다. 구간 변환을 사용하여 해당 변환의 결과를 보여 줍니다. 데이터에는 의미를 갖는 순서가 있습니다 (치아 네 개는 치아 두 개보다 2 배 많음). 순서형 변환과 같은 다른 변환을 탐색할 수도 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Teeth.jmp 를 엽니다.
2. MAMMAL 열을 마우스 오른쪽 버튼으로 클릭하고 **라벨 / 라벨 해제**를 선택합니다.
3. **행 > 행 선택 > 모든 행 선택**을 선택합니다.
4. **행 > 라벨 / 라벨 해제**를 선택합니다.
5. **분석 > 다변량 방법 > 다차원 척도법**을 선택합니다.
6. Top incisors ~ Bottom molars 를 선택하고 **Y, 열**을 클릭합니다.
7. **데이터 형식 > 속성 목록**을 선택합니다.
8. **변환 > 구간**을 선택합니다.

9. **확인**을 클릭합니다 .

그림 10.6 다차원 척도 보고서



"Shepard 다이어그램 " 과 " 적합 상세 정보 " 에서는 MDS 그림에 치아 유사성으로 인한 동물의 유사성이 잘 표현되어 있음을 나타냅니다 . " 스트레스 " 통계량 (0.093) 은 낮고 , 변환된 근

접성 대 예측된 근접성의 R^2 적합 (0.96) 은 높습니다 . 또한 " 적합 상세 정보 " 에서는 실제 근접성의 변환을 위한 절편과 기울기를 제공합니다 .

다차원 척도법 플랫폼에 대한 통계 상세 정보

JMP 에서는 MDS 좌표를 결정하기 위해 Quasi Newton 최적화 방법을 사용하여 스트레스 함수를 최소화합니다 . 이 최소화로 인해 차원의 각 쌍별 집합에 대해 도출된 근접성 측도를 최소화하는 좌표 집합이 미리 결정된 수의 차원에 생성됩니다 . 데이터가 순서형이면 단조 회귀가 사용됩니다 . 그렇지 않으면 표준 최소 제곱 회귀가 사용됩니다 .

- " 스트레스 "
- " 변환 "
- " 속성 목록 형식 "

스트레스

다음 표기를 사용하여 스트레스를 정의합니다 .

- i, j - 개체 수에 대한 인덱스
- d_{ij} - 개체 i 와 j 사이의 유도 거리
  δ_{ij} - 개체 i 와 j 사이의 관측된 상대 거리
- $f(\delta_{rs})$ - 거리에 대한 변환 함수

스트레스 함수는 다음과 같이 정의됩니다 .

$$\text{스트레스} = \left[\frac{\sum_{i<j} [f(\delta_{ij}) - d_{ij}]^2}{\sum_{i<j} d_{ij}^2} \right]^{\frac{1}{2}}$$

이 스트레스 측도는 Kruskal 스트레스, 제 1 종 또는 간단히 스트레스 1 이라고도 합니다 .

변환

이 섹션에서는 "스트레스"에 설명된 표기를 사용합니다. 변환은 실제 근접성을 척도화하는 데 사용됩니다. 변환은 데이터의 특정 구조를 고려하여 실제 근접성의 MDS 표현을 향상시키는 것으로 간주됩니다. 변환 함수의 모수는 최소화 알고리즘의 추가 모수가 됩니다.

비 변환

비율 데이터의 경우 :

$$f(\delta_{rs}) = b\delta_{rs}$$

구간 변환

구간 데이터의 경우 :

$$f(\delta_{rs}) = a + b\delta_{rs}$$

순서형 변환

순서형 데이터의 경우 데이터가 변환되지 않으며 최소 제곱 회귀 대신 단조 회귀가 알고리즘에 사용됩니다.

속성 목록 형식

데이터가 속성 목록인 경우 거리 행렬로 변환된 후 MDS가 적용됩니다. 거리 행렬은 유클리드 거리를 사용하여 계산됩니다. 각 항목 쌍에 대해 다음과 같이 항목 간의 거리를 정의합니다.

$$\delta_{ij} = \sqrt{\sum_{k} \frac{(x_{ki} - x_{kj})^2}{k}}$$

여기서 k는 속성 수입니다.

참고 : MDS 플랫폼에 대한 고급 예는 San Francisco Crime Distances.jmp 샘플 데이터 테이블과 해당 테이블의 소스 스크립트를 참조하십시오 . 이 스크립트는 쌍별 상관을 사용하여 거리 행렬을 생성합니다 . 그런 다음 결과 거리 행렬을 사용하여 범죄 범주 간의 관계를 탐색합니다 .

11 장

항목 분석
항목 및 대상별 검정 결과 분석

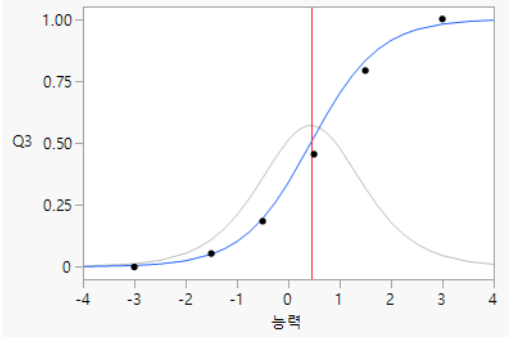
항목 분석 플랫폼을 사용하면 **항목 반응 이론** 모형을 적합시킬 수 있습니다. IRT(항목 반응 이론) 방법은 검사 및 설문지와 같이 측정기의 분석 및 스코어링에 사용됩니다. 항목 반응 이론은 개인이 어떤 항목을 지지하거나 올바르게 응답할 확률을 특성 또는 능력과 연관시키기 위해 모형 시스템을 사용합니다. 관심 특성 또는 능력은 직접 측정할 수 없는 경우가 많으므로 **잠재적**이라고 합니다. 표준화 검사, 인지 발달 및 소비자 선호도를 연구할 때 IRT를 사용할 수 있습니다. IRT는 항목 스코어보다 관측된 총 스코어에 초점을 맞추는 CTT(고전 검사 이론) 대신 사용할 수 있는 방법입니다.

항목 분석 플랫폼에서는 다음과 같은 결과로 IRT 방법을 구현합니다.

- 측정기가 항목 수준에서 스코어링되므로 잠재 반응에 대한 각 항목의 기여도를 파악할 수 있습니다.
- 응답자와 항목에 대한 스코어가 모두 동일한 척도로 산출됩니다.
- 응답자 및 항목 스코어가 단일 그림에 표시됩니다.
- 항목 특성 곡선이 표시됩니다. 이 곡선은 항목과 응답자의 기본 특성 또는 능력 간의 관계를 탐색하는 데 사용할 수 있습니다.

항목 반응 이론에 대한 자세한 내용은 de Ayala(2009) 연구 자료에서 확인하십시오.

그림 11.1 항목 분석 특성 그림



목차

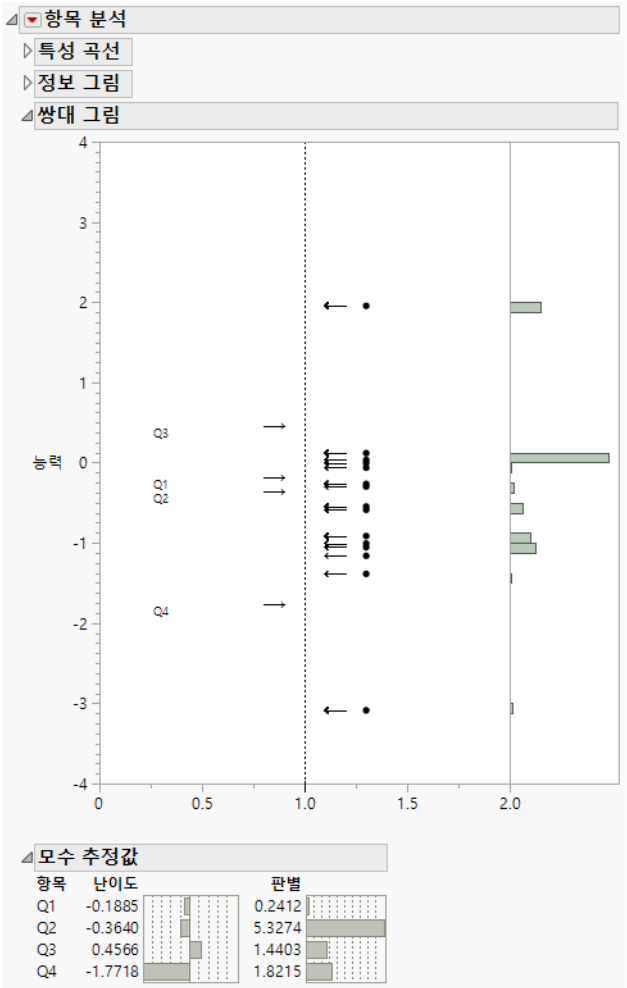
항목 분석의 예 235
항목 분석 플랫폼 시작 237
로지스틱 3PL 모형 상세 정보 238
데이터 형식 238
항목 분석 보고서 238
특성 곡선 238
정보 그림 239
쌍대 그림 240
모수 추정값 240
항목 분석 플랫폼 옵션 241
항목 분석 플랫폼에 대한 통계 상세 정보 241
항목 반응 곡선 241
항목 반응 곡선 모형 242
IRT 모형 가정 244
IRT 모형 적합 245
능력 계산식 246

항목 분석의 예

이 예에서는 1996년에 실시된 TIMSS(Third International Mathematics and Science Study) 관련 데이터에서 가져온 MathScienceTest.jmp 샘플 데이터 테이블을 사용합니다. 데이터 테이블에는 14 개 질문에 대한 1263 명 학생의 스코어 (1 = 정답, 0 = 오답) 가 포함되어 있습니다. 질문과 응답자의 수학적 능력 간의 관계를 이해하기 위해 처음 네 가지 질문을 검토합니다. 시험 문제는 잠재적 수학 능력을 측정하는 데 사용되는 항목입니다. 2PL 모형을 이 데이터에 적합시킵니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 MathScienceTest.jmp 를 엽니다.
2. **분석 > 다변량 방법 > 항목 분석**을 선택합니다.
3. Q1 ~ Q4 를 선택하고 **Y, 테스트 항목**을 클릭한 후 **확인**을 클릭합니다.

그림 11.2 항목 반응 보고서

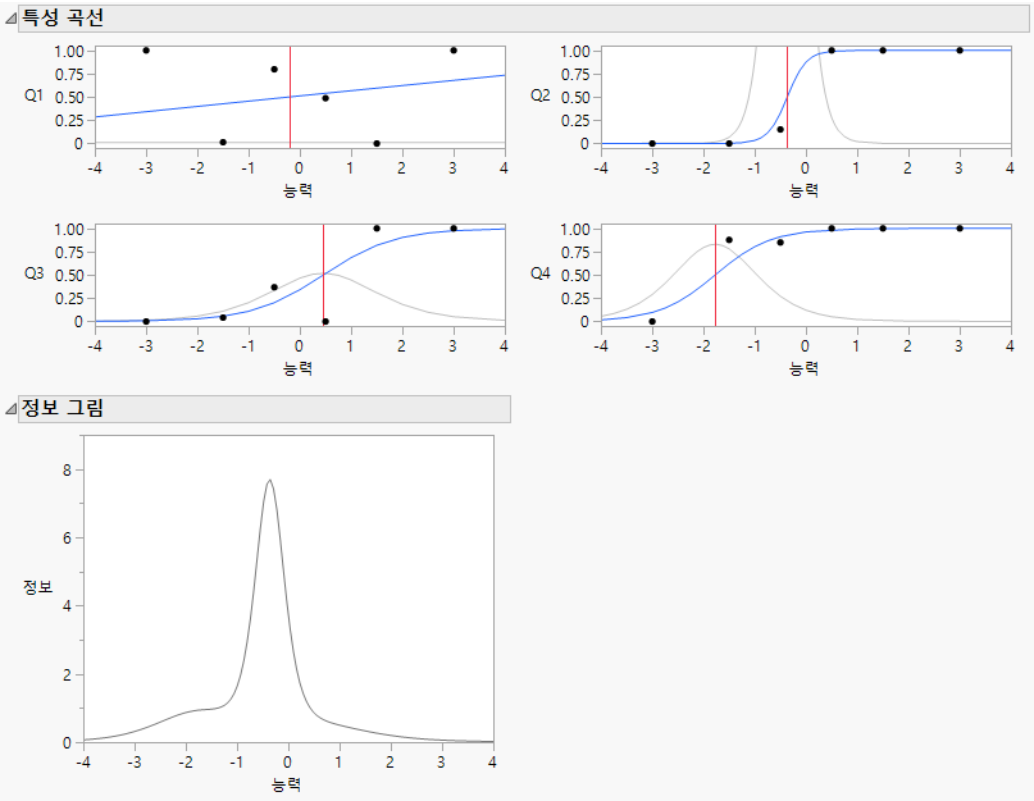


쌍대 그림에서 Q4 의 난이도 스코어가 -1.78 로 가장 낮으므로 Q4 가 가장 쉬운 문제라는 것을 알 수 있습니다 . Q3 이 난이도 스코어 0.46 으로 가장 어렵습니다 . 대부분의 응답자는 그래프 가운데 부분의 데이터 점에서 볼 수 있듯이 능력 척도의 중간에서 아래쪽 끝에 속합니다 . 히스토그램을 보면 응답자의 약 40% 가 능력 척도에서 0 보다 약간 위에 있는 것을 알 수 있습니다 .

참고 : 모두 오답이거나 모두 정답인 개인에 대해서는 능력 스코어가 계산되지 않습니다 . 자세한 내용은 "IRT 모형 적합 " 에서 확인하십시오 .

4. " 특성 곡선 " 보고서의 회색 표시 아이콘을 클릭하여 엽니다 .
5. " 항목 분석 " 의 빨간색 삼각형을 클릭하고 **가로 그림 수**를 선택합니다 .
6. "2" 를 입력하고 **확인**을 클릭합니다 .
7. " 정보 그림 " 보고서의 회색 표시 아이콘을 클릭하여 엽니다 .

그림 11.3 항목 반응의 예



Q1 은 특성 곡선과 정보 곡선이 모두 편평합니다 . 이는 Q1 이 응답자의 수학적 능력을 판별할 수 있는 많은 정보를 제공하지 않는다는 것을 나타냅니다 . Q2 의 특성 곡선은 가파르므로 Q2 가 응답

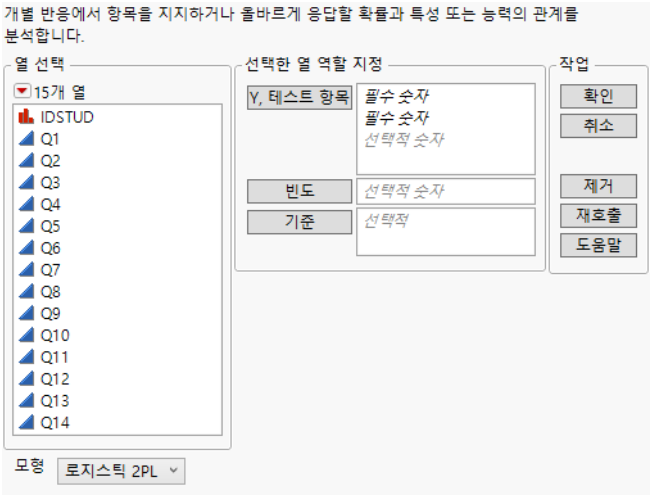
자 능력을 판별하는 데 유용하다는 것을 알 수 있습니다. 각 그림의 세로선은 특성 곡선의 변곡점에 있습니다. 이 세로선은 응답자가 지정된 문제에 정답을 맞출 확률이 50%인 능력 수준입니다.

정보 그림에서는 네 가지 질문을 함께 분석한 결과 -1과 0 사이의 능력 수준에 대한 정보를 가장 많이 제공한다는 것을 나타냅니다. 난이도가 높은 질문을 모형에 더 많이 포함하면 정보 곡선이 넓어질 수 있습니다.

항목 분석 플랫폼 시작

분석 > 다변량 방법 > 항목 분석을 선택하여 항목 분석 플랫폼을 시작합니다.

그림 11.4 항목 분석 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 테스트 항목 분석할 두 개 이상의 열을 할당합니다. 열은 연속형 숫자여야 하고 0과 1만 포함할 수 있습니다.

> **팁 :** 데이터를 0과 1로 재코딩해야 하는 경우 **열 > 재코딩**을 사용합니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

빈도 이 역할에 빈도 변수를 할당합니다. 이 기능은 데이터를 요약할 경우에 유용합니다.

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

모형 다음 옵션에서 원하는 모형을 지정합니다.

로지스틱 2PL 2모수 로지스틱 모형입니다.

로지스틱 3PL 3 모수 로지스틱 모형입니다 .

로지스틱 1PL Rasch 파라미터화를 사용한 1 모수 로지스틱 모형입니다 .

로지스틱 3PL 모형 상세 정보

로지스틱 3PL 모형을 선택하고 " 확인 " 을 클릭하면 추측 모수에 대한 벌점을 입력하라는 메시지가 나타납니다 . 3PL 모형의 경우 벌점 기본값은 0 입니다 . 그러나 *c* 모수 (각 항목에 대한 추측) 에 0 이 아닌 벌점을 입력할 수 있습니다 . 이 벌점은 능형 회귀에서 사용하는 벌점 모수 유형과 유사합니다. 벌점은 추정된 추측 모수의 분산에 적용됩니다. 벌점을 사용하면 다음과 같은 이점이 있습니다 .

- 모형 모수 추정을 안정화합니다 .
- 계산 속도가 빠릅니다 .
- 약간의 편향은 있지만 항목 전체에서 추측 모수의 변동을 줄입니다 .

벌점 값이 크면 추측 모수가 0 이 되며 값이 작을수록 항목 전체에서 추측 모수의 변동을 줄이는 데 도움이 됩니다. 0 값을 사용하면 벌점이 없습니다.

데이터 형식

항목 분석 플랫폼에는 각 개인에 대한 행과 각 항목에 대한 열이 포함된 데이터 테이블이 필요합니다. 항목 열은 수치형이어야 하고 올바르지 않은 반응을 나타내는 0 과 올바른 반응을 나타내는 1 만 포함할 수 있습니다. MathScienceTest.jmp 샘플 데이터 테이블에서는 14 개의 시험 문제에 대한 1,263 명의 반응으로 항목 반응 분석을 수행하기 위해 필요한 데이터 형식을 보여 줍니다.

항목 분석 보고서

- " 특성 곡선 "
- " 정보 그림 "
- " 쌍대 그림 "
- " 모수 추정값 "

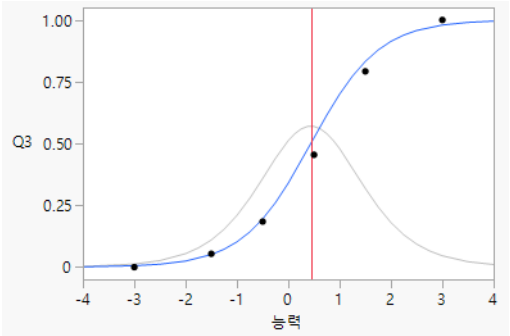
특성 곡선

"특성 곡선"에는 시작 창에서 지정한 각 항목에 대한 ICC(항목 특성 곡선)가 포함됩니다. "특성 곡선"은 처음에 닫혀 있습니다.

항목 특성 곡선은 항목에 올바르게 답할 확률 대 능력을 표시합니다. 능력은 표준화된 척도로 측정되므로 능력이 0 인 응답자는 평균 능력의 응답자입니다. 고정된 능력 수준에 대해 관측된 정답 확률의 데이터 점이 그림에 표시됩니다. 적합 특성 곡선을 데이터 점과 비교하면 각 개별 항목에 대한 모형 적합도를 시각적으로 평가할 수 있습니다. 또한 특성 그림에는 배경 정보 곡선이 있

으며 특성 곡선 변곡점 위치에 세로선이 있습니다. 배경 정보 곡선은 변곡점에서 최대화되는 항목 특성 곡선의 기울기 그림입니다.

그림 11.5 항목 특성 곡선

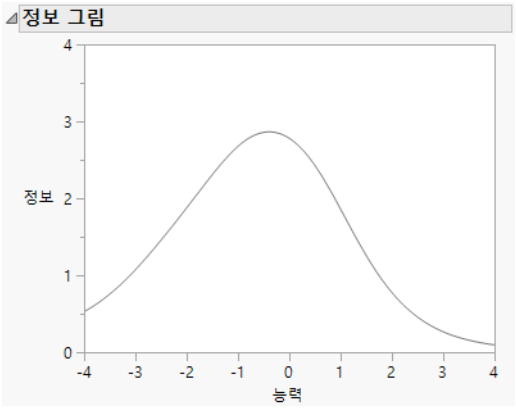


팁 : " 항목 분석 " 의 빨간색 삼각형 메뉴에 있는 " 가로 그림 수 " 옵션을 사용하여 보고서의 각 행에 나타나는 특성 곡선의 수를 조정할 수 있습니다 .

정보 그림

"정보 그림" 보고서에는 개별 항목 정보 곡선을 합하여 생성되는 전체 정보 곡선 그림이 포함됩니다. 이 정보 그림을 통해 테스트로 측정할 수 있는 적절한 능력 수준을 파악할 수 있습니다. 그림 11.6에서는 능력 수준이 높은 개인보다 능력이 평균 수준에서 낮은 수준인 개인을 평가하는 데 더 적절한 항목을 사용한 테스트를 보여 줍니다. 이 그림은 처음에 닫혀 있습니다.

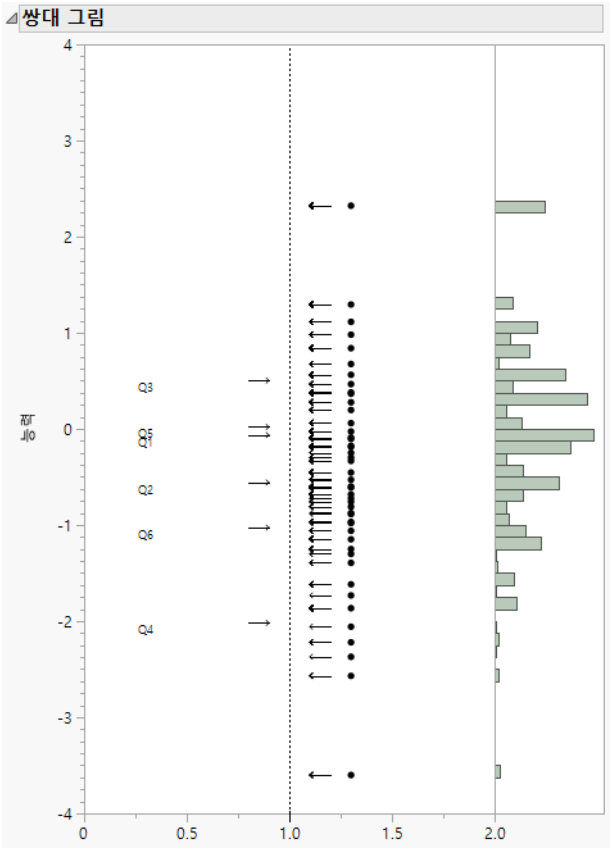
그림 11.6 정보 그림



쌍대 그림

"쌍대 그림" 보고서에는 항목 난이도와 대상 능력을 하나의 그림에 보여 주는 그림이 포함됩니다. 난이도와 능력은 세로 축에 표시된 공통 표준화 척도를 사용합니다. 항목은 그림 왼쪽에 난이도별로 표시됩니다. 대상은 그림 오른쪽에 데이터 점과 히스토그램으로 표시됩니다. 쌍대 그림을 사용하면 각 항목의 난이도를 각 응답자의 능력과 연관시킬 수 있습니다.

그림 11.7 쌍대 그림



모수 추정값

"모수 추정값" 보고서에는 각 항목에 대한 추정 모수 테이블이 포함됩니다. 제공되는 모수는 분석에 사용된 모형 (1PL, 2PL 또는 3PL) 에 따라 다릅니다.

항목 테스트 항목입니다.

난이도 *b* 모수 또는 항목 난이도 측도입니다. 난이도 추정값 옆에 난이도 모수 히스토그램이 표시됩니다.

판별 (2PL 및 3PL 모형에만 사용 가능) *a* 모수 또는 항목 판별 측도입니다. 판별 추정값 옆에 판별 모수 히스토그램이 표시됩니다.

하위 점근선 (3PL 모형에만 사용 가능) *c* 모수 또는 추측 측도입니다.

항목 분석 플랫폼 옵션

가로 그림 수 "특성 곡선" 보고서의 각 그림 행에 표시할 ICC 그림 수를 지정할 수 있습니다. 기본값은 행당 하나의 ICC 그림입니다.

능력 계산식 저장 능력 계산식을 데이터 테이블의 새 열에 저장합니다. 이 옵션은 `IRT Ability()` JSL 함수를 사용합니다. 이 함수에 대한 자세한 내용은 도움말 > 스크립트 인덱스에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

항목 분석 플랫폼에 대한 통계 상세 정보

IRT(항목 반응 이론)는 일련의 방정식을 사용하여 항목을 관측되지 않은(잠재적) 특성 또는 능력과 연관시킵니다. 항목 또는 질문은 직접 관측할 수 없는 기본 잠재 구성의 지표입니다. 데이터를 수집할 때는 대상 능력과 항목 특성 둘 다 알 수 없습니다.

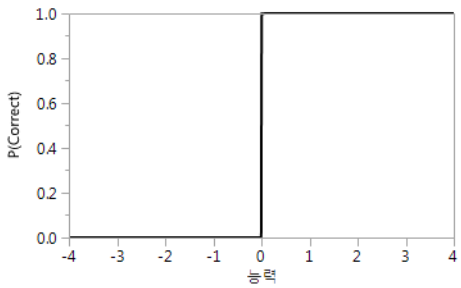
- "항목 반응 곡선"
- "항목 반응 곡선 모형"
- "IRT 모형 가정"
- "IRT 모형 적합"
- "능력 계산식"

항목 반응 곡선

항목 반응 곡선(항목 특성 곡선)은 **능력** 척도로 정의된 능력과 각 항목 간의 관계를 기술하는 데 사용됩니다. 항목 반응 곡선은 다양한 수준의 능력에 대해 항목에 올바르게 답할 확률을 표시합

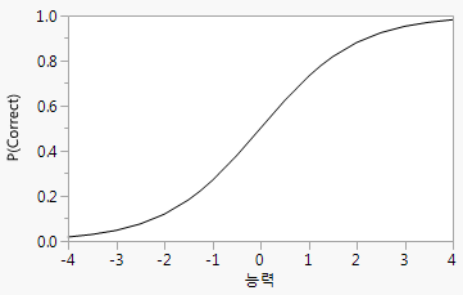
니다. 완벽한 판별력을 가진 항목의 경우 능력이 임계값 미만인 응답자가 올바르게 답할 확률은 0%이고 능력이 임계값을 초과하는 대상이 올바르게 답할 확률은 100%입니다.

그림 11.8 완벽한 판별력을 가진 항목에 대한 특성 곡선



항목에 올바르게 답할 확률과 능력 간의 일반적인 관계는 하위 점근선과 상위 점근선이 있는 S 자 형태의 함수입니다. 응답자의 능력이 향상될수록 항목에 올바르게 답할 확률은 100%까지 증가합니다. 특정 항목에 대한 곡선 모양은 항목의 난이도 및 판별 특성과 관련됩니다.

그림 11.9 일반적 항목 반응 곡선



항목 반응 곡선 모형

1 모수 , 2 모수 및 3 모수 로지스틱 모형을 사용하여 항목 반응 곡선을 모델링할 수 있습니다 . 3PL(3 모수 로지스틱) 모형은 다음과 같이 정의됩니다 .

$$P(\theta) = c + \frac{1-c}{1+e^{-(a)(\theta-b)}}$$

- $P(\theta)$ 는 능력 수준 θ에 대해 항목에 올바르게 답할 확률입니다 . 항목 반응 이론 모형 적합에 대한 자세한 내용은 "IRT 모형 적합 " 에서 확인하십시오 .
- a 모수는 변곡점에서 곡선의 경사도를 정의합니다 . 이 모수는 항목의 판별력 추정값을 제공합니다 .

- *b* 모수는 능력 축에서 변곡점 위치를 정의합니다 . 이 모수는 항목의 난이도 추정값을 제공합니다 .
- *c* 모수는 하위 점근선입니다 . 이 모수는 추측을 통해 항목에 올바르게 답할 확률의 추정값을 제공합니다 .
- 2PL 모형의 경우 *c* 모수는 0 으로 설정됩니다 .

$$P(\theta) = \frac{1}{1 + e^{-(a)(\theta - b)}}$$

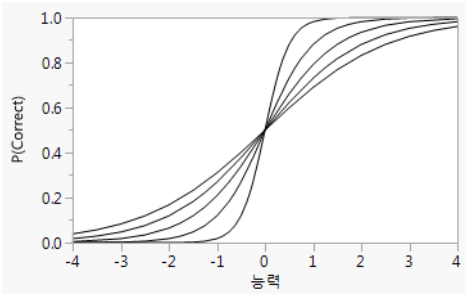
- 1PL 모형의 경우 *c* 모수는 0 으로 설정되고 *a* 모수는 1 로 설정됩니다 . 이 파라미터화를 Rasch 모형이라고도 합니다 (Rasch 1980 참조).

$$P(\theta) = \frac{1}{1 + e^{-(\theta - b)}}$$

a 모수 : 항목 판별

2PL 및 3PL 모형의 *a* 모수, 즉 변곡점에서 곡선의 경사도는 항목의 판별력 측도를 제공합니다. 항목의 판별력 또는 판별은 능력 수준이 낮은 응답자와 능력 수준이 높은 응답자를 항목이 얼마나 잘 구별할 수 있는지를 나타냅니다. 가파른 항목 반응 곡선은 항목의 판별력이 크다는 것을 나타냅니다. 능력 수준이 낮은 응답자는 항목에 올바르게 반응할 확률이 낮은 반면 능력이 높은 응답자는 올바르게 반응할 확률이 높습니다. 비교적 곡선이 편평한 항목은 판별력이 낮습니다. 판별력이 낮은 항목은 측정기에서 탈락할 후보입니다.

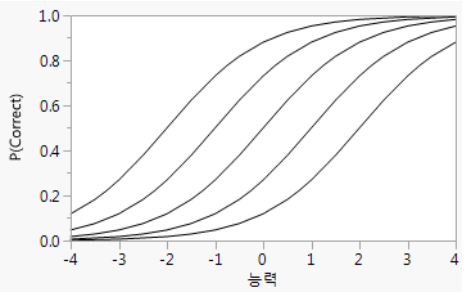
그림 11.10 여러 *a* 값에 대한 로지스틱 모형



b 모수 : 항목 난이도

b 모수, 즉 능력에 대한 변곡점 위치는 항목 난이도 측도를 제공합니다. 능력 척도에서 변곡점 위치가 오른쪽으로 더 멀리 있는 항목 반응 곡선은 변곡점이 왼쪽에 있는 항목보다 답하기 어려운 항목을 나타냅니다. 1PL 및 2PL 모형에서 *b* 모수는 항목에 올바르게 답할 확률이 50% 가 되기 위해 필요한 능력 수준의 추정값을 제공합니다.

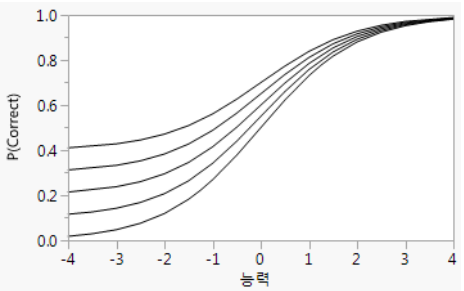
그림 11.11 여러 *b* 값에 대한 로지스틱 곡선



c 모수 : 추측

3PL 모형에서 *c* 모수, 즉 항목 반응 곡선의 하위 점근선은 추측 모수 측도를 제공합니다. 0이 아닌 하위 점근선은 능력 수준이 매우 낮은 사람이 항목에 올바르게 답할 확률이 0이 아님을 나타냅니다.

그림 11.12 여러 *c* 값에 대한 로지스틱 모형



IRT 모형 가정

2PL 모형은 항목 분석 플랫폼의 기본 모형입니다. 1PL 모형은 모든 항목의 판별력이 동일하다고 가정할 수 있을 때 적절합니다. 이 가정이 적절하지 않을 때는 2PL 또는 3PL 모형을 사용해야 합니다. 2PL 모형은 특히 작은 데이터 집합의 경우 3PL 모형보다 수치 안정성이 큽니다. 또한 2PL 모형에서 *b*는 응답자가 항목에 올바르게 답할 확률이 50%가 되기 위해 필요한 능력 수준으로 해석할 수 있습니다.

IRT 모형은 기본 특성이 1차원이라고 가정합니다. 즉, 기본 잠재 구성이 1개입니다. 서로 복잡하게 상호 작용하는 몇 가지 특성을 측정하는 경우 1차원 모형이 적절하지 않습니다. IRT 모형은 연속형 잠재 변수에 적절합니다. 범주형 잠재 변수의 경우 잠재 계층 모형을 고려해야 합니다. 자세한 내용은 "잠재 계층 분석"에서 확인하십시오. IRT 모형은 항목 불변으로 간주됩니다. 항목 불변성은 $P(\theta)$가 능력 수준이 θ인 개인 집합에 대한 올바른 반응 확률로 해석된다는 의미입니다. 능력 수준이 동일한 많은 개인이 항목에 답한 경우 $P(\theta)$는 항목에 올바르게 답할 사람의 비율을 예측합니다. 이는 IRT 모형이 테스트 대상 그룹에 관계없이 동일한 모수를 갖는다는 의미입

니다. 또한 IRT 모형은 국소적 독립을 가정합니다. 즉, 잠재적 구성이 고려된 후에는 항목이 서로 독립적입니다.

IRT 모형 적합

IRT 모형은 MMLE(주변 최대 가능도 추정) 를 사용하여 적합됩니다 . MMLE 는 JLE(결합 최대 가능도 추정) 대신 사용할 수 있는 방법입니다 . MMLE 는 대상을 임의 효과로 처리합니다 . 항목과 능력은 조건부 확률로 관련됩니다 . 계산식은 다음과 같이 정의됩니다 .

$$p(\mathbf{x}|\theta, \vartheta) = \prod_{j=1}^{L} p_j(\theta)^{x_j}(1 - p_j(\theta))^{1-x_j}$$

여기서 $p(\mathbf{x}|\theta, \vartheta)$는 대상 능력 θ와 항목 모수 벡터 ϑ가 주어졌을 때 반응 벡터 $\mathbf{x}$의 확률입니다. 항목 모수의 수는 사용된 모형(1PL, 2PL 또는 3PL)에 따라 다릅니다.

MMLE 는 항목 모수 추정값을 얻기 위해 가우스 구적을 사용하여 대상 효과를 통합합니다 . 반응 벡터 $\mathbf{x}$ 의 확률은 다음과 같이 계산됩니다 .

$$p(\mathbf{x}) = \int_{-\infty}^{\infty} p(\mathbf{x}|\theta, \vartheta) g(\theta|\upsilon) d\theta$$

여기서 $g(\theta|\nu)$는 대상의 분포이고 ν는 모집단 위치와 척도 모수의 벡터입니다. JMP에서는 평균이 0이고 표준편차가 1인 정규 분포가 $g(\theta|\nu)$에 사용됩니다.

참고 : 테스트 질문에 대한 결측값은 올바르지 않은 반응으로 처리됩니다 . 모두 올바르지 않게 답한 개인 또는 모두 올바르게 답한 개인에 대해서는 능력 스코어가 계산되지 않습니다 . 이러한 대상에 대한 반응 패턴이 모형 추정에 포함됩니다 .

IRT 모형 적합을 위한 MMLE 절차는 두 단계로 임의 효과 모형을 적합시키는 것과 비교할 수 있습니다. 능력 모수는 분산이 1인 임의 효과로 처리됩니다. 첫 번째 단계에서는 가우스 구적을 사용하여 이러한 임의 효과가 통합됩니다. 항목 모수는 능력 모수가 통합된 상태에서 주변 가능도의 ML을 사용하여 추정되는 고정 효과로 처리됩니다. 능력 모수는 본질적으로 완전 비통합 (결합) 가능도를 사용하여 추정되는 최량 선형 비편향 예측입니다. 이때 항목 모수를 알려진 것으로 처리하고 첫 번째 단계에서 얻은 값으로 고정합니다.

L 개 항목에 대한 2^L 반응 패턴이 있습니다. 각 패턴의 능력 수준은 θ 가 수렴할 때까지 다음을 적용하여 반응 패턴에 대한 확률이 가장 높은 능력 수준을 찾는 방식으로 계산할 수 있습니다.

$$\theta_i^{t+1} = \theta_i^{t} - \frac{X_i - \sum_{j=1}^{L} p_{ij}(t)}{-\sum_{j=1}^{L} p_{ij}(t)(1-p_{ij}(t))}$$

다음은 각 요소에 대한 설명입니다.

θ = 반응 패턴을 얻을 가능도를 최대화합니다.

t = 반복 수

L = 항목 수

X_i = 관측된 스코어

p_{ij} = 항목 모수를 기반으로 i 번째 사람이 j 번째 항목에 대해 올바르게 반응할 확률

능력 계산식

"항목 분석"의 빨간색 삼각형 메뉴에 있는 "능력 계산식 저장" 옵션은 능력 계산식을 데이터 테이블의 새 열에 저장합니다. 이 계산식을 사용하여 데이터 테이블에 추가된 다른 대상을 스코어링하거나, 새 테이블에 복사하여 새 대상 집합을 스코어링할 수 있습니다.

테이블에 저장된 함수를 `IRT Ability` 함수라고 합니다. 항목 모수 추정값은 이 함수의 행렬에 저장됩니다. 이 함수에 대한 자세한 내용은 도움말 > 스크립트 인덱스에서 확인하십시오.

12 장

계층적 군집화
군집 트리를 사용하여 관측값 그룹화

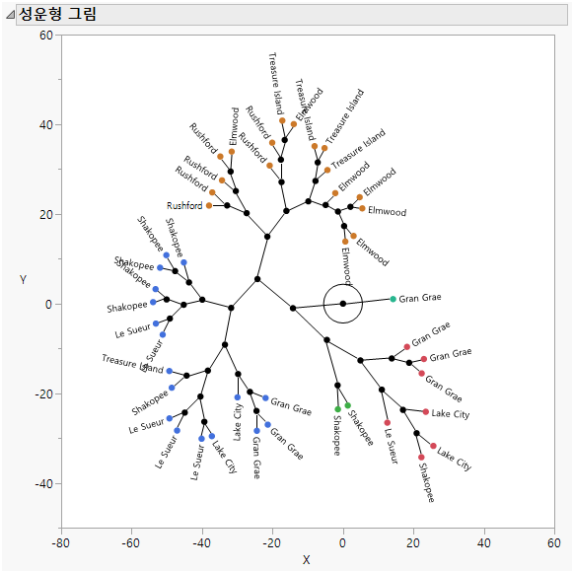
군집화는 여러 변수에 대해 유사한 값을 공유하는 관측값을 함께 그룹화하는 다변량 기법입니다. 이 방법을 사용하여 데이터의 군집 구조를 이해할 수 있습니다.

계층적 군집화는 군집을 연속해서 결합합니다. 이 방법은 각 관측값을 자체 군집화하는 것으로 시작합니다. 그런 다음 각 단계에서 거리가 가장 가까운 두 군집을 단일 군집으로 결합합니다. 결과는 **덴드로그램**이라고 하는 트리로 표시됩니다.

행 수가 수만 개 이하인 작은 데이터 테이블의 경우 계층적 군집화를 사용합니다. 이 알고리즘은 시간이 많이 소요되며 큰 데이터 테이블의 경우 느리게 실행될 수 있습니다. 더 큰 데이터 테이블의 경우 K 평균 군집화 또는 정규 혼합 방법을 사용합니다.

참고: 계층적 군집화는 문자 열을 지원하고 K 평균 군집화 또는 정규 혼합에는 숫자 열이 필요합니다.

그림 12.1 성운형 그림의 예



목차

계층적 군집화 플랫폼 개요 249
관측값 군집화 플랫폼 개요 249
군집화의 예 250
계층적 군집화 플랫폼 시작 253
군집화 방법 254
거리 계산 방법 254
데이터 구조 255
Y, 열 변수로 변환 256
계층적 군집화 보고서 257
덴드로그램 보고서 258
덴드로그램 및 거리 그래프 예시 258
군집화 기록 보고서 260
계층적 군집화 옵션 260
계층적 군집화 플랫폼의 추가 예 263
거리 행렬의 예 263
공간 측도를 사용한 웨이퍼 결함 분류의 예 265
계층적 군집화 플랫폼에 대한 통계 상세 정보 267
공간 측도 267
거리 방법 계산식 269

계층적 군집화 플랫폼 개요

계층적 군집화는 JMP에서 관측값을 군집화하기 위해 제공하는 네 가지 플랫폼 중 하나입니다. 네 가지 방법을 모두 비교하려면 "관측값 군집화 플랫폼 개요"의 내용을 참조하십시오.

계층적 군집화 방법은 각 관측값이 자체 군집을 형성하는 것으로 시작합니다. 군집화 과정의 각 단계에서는 모든 군집 쌍 간의 거리를 계산하고 가장 가까운 두 군집을 결합합니다. 모든 점이 하나의 군집에 포함될 때까지 이 과정이 계속됩니다. 계층적 군집화는 사용하는 결합 방식 때문에 **병합 군집화**라고도 합니다.

병합 과정은 덴드로그램이라는 트리로 표현됩니다. JMP에서는 군집 수를 결정하는 데 도움이 되도록 거리 그래프를 제공합니다. 군집 간의 거리가 실제적으로 더 이상 중요하지 않은 시점을 확인하여 군집 수를 선택할 수 있습니다.

계층적 군집화는 문자 열도 지원합니다. 두 가지 방법을 사용하여 거리를 정의할 수 있습니다.

- 열이 순서형인 경우 군집화에 사용되는 값은 순서가 지정된 범주의 인덱스로, 연속형 데이터처럼 처리됩니다. 이러한 값은 연속형 데이터처럼 표준화됩니다.
- 열이 명목형인 경우 범주가 일치하는 두 관측값 간의 거리는 0 입니다. 범주가 다르면 거리가 1 입니다.

계층적 군집화를 사용하면 "평균", "중심", "Ward", "단일" 및 "완전"이라는 다섯 가지 규칙 중에서 선택하여 군집 간 거리를 정의할 수 있습니다. 각 규칙은 서로 다른 군집 시퀀스를 생성할 수 있습니다.

팁 : 계층적 군집화 과정은 " 빠른 Ward" 방법을 사용하는 경우를 제외하고 n 개의 관측값에 대해 $n(n+1)/2$ 개의 거리로 시작합니다. 따라서 이 방법은 n 이 큰 경우 실행 시간이 오래 걸릴 수 있습니다. 숫자 관측값 수가 많은 경우에는 K 평균 군집화 또는 정규 혼합 방법을 고려해 보십시오.

관측값 군집화 플랫폼 개요

군집화는 여러 변수에 대해 유사한 값을 공유하는 관측값을 함께 그룹화하는 다변량 기법입니다. 일반적으로 관측값은 p 차원 공간에 고르게 분산되지 않습니다. 여기서 p 는 변수 수입니다. 대신 관측값은 군집을 형성합니다. 이러한 군집을 식별하면 데이터를 더 세밀하게 파악할 수 있습니다.

참고 : JMP 에서는 변수를 군집화할 수 있는 플랫폼도 제공합니다. 자세한 내용은 " 변수 군집화 " 에서 확인하십시오.

JMP 에서 제공하는 다음 네 가지 플랫폼을 사용하여 관측값을 군집화할 수 있습니다.

- 계층적 군집화는 최대 수만 개의 행이 있는 작은 테이블에 유용하며 문자 데이터를 허용합니다. 계층적 군집화는 행을 계층적 순서로 결합하여 트리로 표현합니다. 트리가 생성되면 데이터에 가장 적절한 군집 수를 선택할 수 있습니다.

- K 평균 군집화는 최대 수백만 개의 행이 있는 큰 테이블에 적절하며 숫자 데이터만 허용합니다. 군집 수 *k* 를 미리 지정해야 합니다. 이 알고리즘은 군집 시드 점을 추측합니다. 그런 다음 군집에 점을 번갈아 할당하고 군집 중심을 다시 계산하는 반복 과정을 수행합니다.
- 정규 혼합은 중첩될 수 있는 다변량 정규 분포의 혼합에서 데이터를 수집한 경우에 적절하며 숫자 데이터만 허용합니다. 다변량 이상치가 있는 경우 균등 분포를 가정하고 이상치 군집을 사용할 수 있습니다.

  군집 수를 미리 지정해야 합니다. 최대 가능도는 혼합 비율과 평균, 표준편차 및 상관을 공동으로 추정하는 데 사용됩니다. 각 그룹에 속할 확률이 각 점에 할당됩니다. EM 알고리즘은 추정값을 구하는 데 사용됩니다.
- 잠재 계층 분석은 대부분의 변수가 범주형일 때 적절합니다. 군집 수를 미리 지정해야 합니다. 이 알고리즘은 다항 혼합 분포를 가정하는 모형을 적합시킵니다. 각 관측값에 대해 소속 군집의 최대 가능도 추정값이 계산됩니다. 관측값은 소속 확률이 가장 큰 군집으로 분류됩니다.

표 12.1 군집화 방법 요약

방법	데이터 유형 또는 모델링 유형	데이터 테이블 크기	군집 수 지정
계층적 군집화	임의	빠른 Ward 를 사용하는 경우 최대 200,000 개 행 다른 방법을 사용하는 경우 최대 5,000 개 행	아니요
K 평균 군집화	수치형	최대 수백만 개 행	예
정규 혼합	수치형	모든 크기	예
잠재 계층 분석	명목형 또는 순서형	모든 크기	예

일부 군집화 플랫폼에는 데이터의 이상치를 처리하는 옵션이 있습니다. 그러나 데이터에 이상치가 있으면 분석하기 전에 먼저 이상치를 탐색하는 것이 가장 좋습니다. 이 작업은 이상치 탐색 유틸리티를 사용하여 수행할 수 있습니다. 자세한 내용은 의"모델링 유틸리티" 장에서 확인하십시오.

군집화의 예

이 예에서는 2009 년 인구 1,000 명당 조출생률 및 사망률을 기준으로 여러 국가를 그룹화합니다.

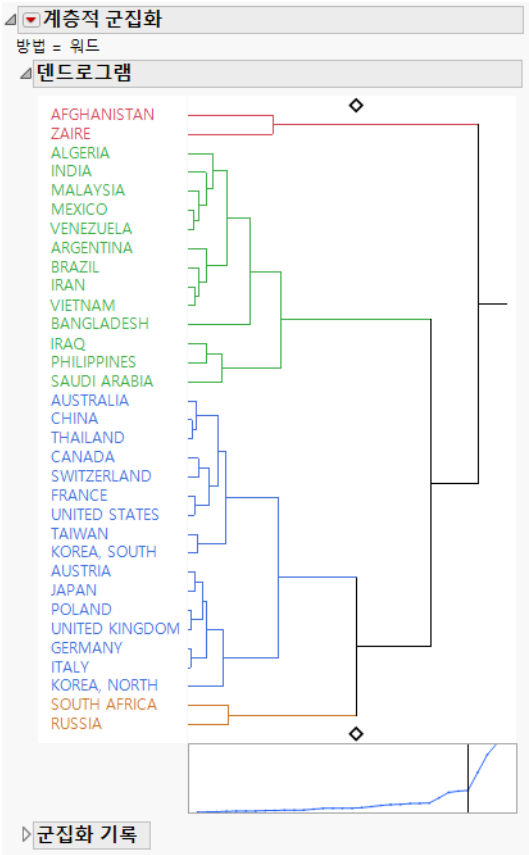
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Birth Death Subset.jmp 를 엽니다.
2. **분석 > 군집화 > 계층적 군집화**를 선택합니다.
3. birth 와 death 를 선택하고 **Y, 열**을 클릭합니다.

4. country 를 선택하고 **라벨**을 클릭합니다 .

   이렇게 선택하고 " 확인 " 을 클릭하면 표시된 덴드로그램에 행 번호 대신 country 열을 사용하여 라벨을 지정할 수 있습니다 .

5. **확인**을 클릭합니다 .
6. " 계층적 군집화 " 의 빨간색 삼각형을 클릭하고 **군집 색 표시**를 선택합니다 .

그림 12.2 계층적 군집화 보고서



덴드로그램에는 군집화 수행 방법이 표시됩니다 . 덴드로그램을 왼쪽에서 오른쪽으로 읽으면 군집화 과정을 볼 수 있습니다 . 각 단계는 **가장 가까운** 두 군집을 단일 군집으로 결합하는 것으로 구성됩니다 .

덴드로그램에서 군집 간의 상대 거리는 군집을 연결하는 세로선 사이의 가로 거리로 지정됩니다 . 예를 들어 예를 들어 Mexico 와 Venezuela 로 구성된 군집과의 차이는 Malaysia 보다 Afghanistan 과 Zaire 가 더 큽니다 .

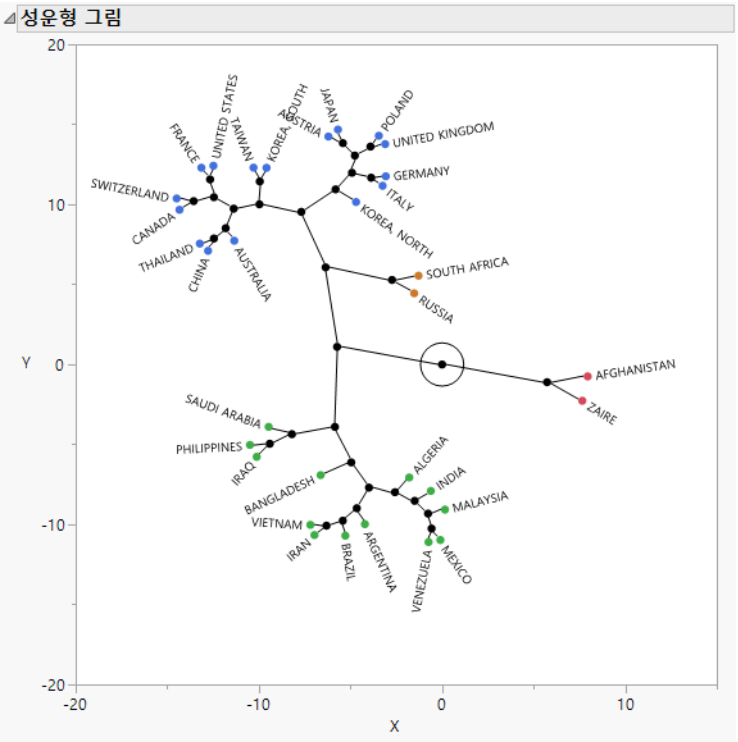
덴드로그램 아래에 표시된 그림에는 두 군집이 단일 군집으로 결합되는 각 단계에 대한 점이 있습니다 . 가로 좌표는 군집 수를 나타내며 왼쪽에서 오른쪽으로 감소합니다 . 점의 세로 좌표는 지정된 수의 군집을 형성하기 위해 결합된 두 군집 간의 거리입니다 . 덴드로그램에서

다이아몬드 중 하나를 클릭하고 선을 드래그하여 데이터를 가장 잘 나타내는 군집 수를 선택할 수 있습니다 . " 계층적 군집화 " 의 빨간색 삼각형 메뉴에서 " 군집 수 " 옵션을 사용하여 군집 수를 선택할 수도 있습니다 .

거리 그래프를 보면 네 개의 군집에서 기울기가 눈에 띄게 변했습니다 . 기울기 변화는 네 개의 군집이 남아 있는 점까지 결합된 군집의 차이가 비교적 작다는 것을 나타냅니다 . 따라서 군집 수를 4 로 선택하는 것이 좋습니다 . 이 값은 기본적으로 표시된 군집 수입니다 .

7. " 계층적 군집화 " 의 빨간색 삼각형을 클릭하고 **성운형 그림**을 선택합니다 .

그림 12.3 성운형 그림

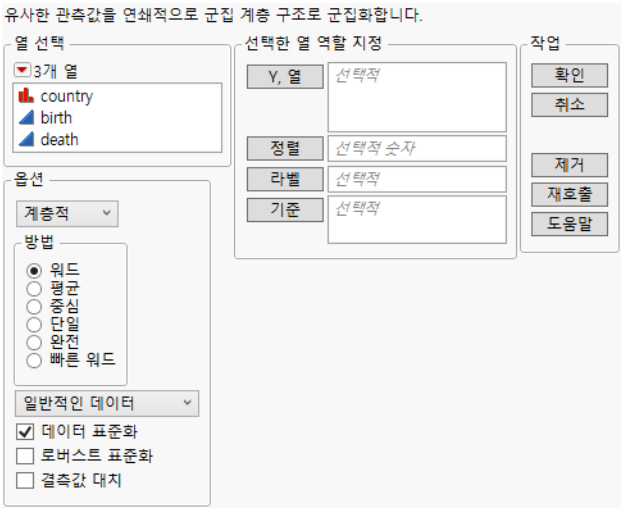


이 성운형 그림은 국가를 끝점으로 배열하고 각 군집 결합을 새 점으로 배열합니다 . 선은 소속 군집을 나타냅니다 . 군집 결합 사이의 선 길이는 결합된 군집 사이의 거리와 거의 같습니다 . 성운형 그림을 보면 Zaire 와 Afghanistan 을 포함하는 군집과 나머지 국가를 포함하는 군집 간의 거리가 그림의 위쪽 절반과 아래쪽 절반에 각각 포함되는 나머지 국가로 구성된 두 군집 간의 거리와 거의 비슷합니다 .

계층적 군집화 플랫폼 시작

분석 > 군집화 > 계층적 군집화를 선택하여 계층적 군집화 플랫폼을 시작합니다.

그림 12.4 계층적 군집화 시작 대화상자



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 열 관측값을 군집화하는 데 사용되는 변수입니다.

정렬 지정된 열에 기반한 평균 값을 기준으로 군집을 정렬합니다.

> **팁:** 주성분 분석을 수행하여 얻은 첫 번째 주성분을 정렬 열로 사용합니다. 군집은 이러한 값을 기준으로 정렬됩니다.

속성 ID (**데이터 쌓여짐**을 데이터 구조로 선택한 경우에만 사용 가능) 쌓이는 변수를 지정합니다.

개체 ID (**요약된 데이터** 또는 **데이터 쌓여짐**을 데이터 구조로 선택한 경우에만 사용 가능) 측정값이 쌓이는 각 단위에 대한 고유 식별자를 제공하는 하나 이상의 열입니다.

라벨 보고서의 덴드로그램에 라벨을 지정하는 데 사용되는 열입니다.

> **참고:** 선택한 데이터 구조가 **데이터가 거리 행렬임**인 경우 라벨 열의 데이터 유형은 문자여야 합니다.

기준 각 수준으로 개별 분석을 정의하는 열입니다. 지정된 열의 각 수준에 대해 해당하는 행이 분석됩니다. 결과는 개별 보고서에 제공됩니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다.

시작 창에는 다음과 같은 메뉴 및 옵션이 있습니다.

- " 군집화 방법 "
- " 거리 계산 방법 "
- " 데이터 구조 "
- "Y, 열 변수로 변환 "

군집화 방법

" 계층적 " 이 기본 군집화 방법이지만 대화상자에서 "K 평균 " 또는 " 정규 혼합 " 으로 전환할 수 있습니다 . "K 평균 " 또는 " 정규 혼합 " 을 선택하고 " 확인 " 을 클릭하면 다음과 같은 방법을 선택할 수 있는 " 제어판 " 이 나타납니다 .

K 평균 군집화 자세한 내용은 "K 평균 군집화 " 에서 확인하십시오 .

정규 혼합 자세한 내용은 " 정규 혼합 " 에서 확인하십시오 .

로버스트 정규 혼합 자세한 내용은 " 정규 혼합 보고서 " 에서 확인하십시오 .

자기 조직화 지도 자세한 내용은 " 자기 조직화 지도 " 에서 확인하십시오 .

거리 계산 방법

거리 계산에 사용되는 방법을 선택합니다 . 거리 계산식은 " 거리 방법 계산식 " 에서 확인하십시오 .

Ward Ward 의 최소 분산 방법에서 두 군집 간의 거리는 모든 변수에 대해 합산된 두 군집 간의 ANOVA 제곱합입니다 . 각 생성 단계마다 이전 생성 단계에서 얻은 두 군집을 병합하여 얻을 수 있는 모든 파티션에 대해 군집 내 제곱합이 최소화됩니다 . 제곱합을 총 제곱합으로 나누어 분산 비율 (준부분 상관 제곱) 을 제공하면 제곱합을 더 쉽게 해석할 수 있습니다 .

Ward 방법은 다변량 정규 혼합 , 구형성 공분산 행렬 및 등확률 표집을 가정하고 계층의 각 수준에서 가능도를 최대화하도록 군집을 결합합니다 .

Ward 방법은 관측값 수가 적은 군집을 결합하는 경향이 있으며 관측값 수가 거의 같은 군집을 생성하는 쪽으로 크게 편향되어 있습니다 . 또한 이상치에도 매우 민감합니다 . 자세한 내용은 Milligan(1980) 연구 자료에서 확인하십시오 .

평균 두 군집 간의 거리는 관측값 쌍 간의 평균 거리입니다 . 평균 연결법은 분산이 작은 군집을 결합하는 경향이 있으며 분산이 같은 군집을 생성하는 쪽으로 약간 편향되어 있습니다. 자세한 내용은 Sokal 과 Michener(1958) 연구 자료에서 확인하십시오 .

중심 두 군집 간의 거리는 평균 사이의 유클리드 거리 제곱으로 정의됩니다 . 중심 방법은 대부분의 다른 계층적 방법보다 이상치에 더 로버스트하지만 다른 측면에서는 Ward 방법 또는 평균 연결법만큼 잘 수행되지 않을 수 있습니다 . 자세한 내용은 Milligan(1980) 연구 자료에서 확인하십시오 .

단일 두 군집 간의 거리는 한 군집의 관측값과 다른 군집의 관측값 사이의 최소 거리입니다 . 단일 연결법은 바람직한 이론적 특성을 많이 가지고 있지만 몬테카를로 연구에서는 잘 수행되지 않았습니다 . 자세한 내용은 Jardine 과 Sibson(1971), Fisher 와 Van Ness(1971),

Hartigan(1981) 및 Milligan(1980) 연구 자료에서 확인하십시오 . 단일 연결법은 Florek et al. (1951a, 1951b) 에 의해 시작되었으며 이후에 McQuitty(1957) 와 Sneath(1957) 에 의해 재창조되었습니다 .

단일 연결법은 군집 모양에 제약 조건을 두지 않으므로 길고 불규칙한 군집을 감지하는 대신 조밀한 군집 복구 성능이 저하됩니다 . 단일 연결법은 주 군집을 분리하기 전에 분포의 꼬리를 잘라내는 경향이 있습니다 . 자세한 내용은 Hartigan(1981) 연구 자료에서 확인하십시오 .

완전 두 군집 간의 거리는 한 군집의 관측값과 다른 군집의 관측값 사이의 최대 거리입니다. 완전 연결법은 지름이 거의 같은 군집을 생성하는 쪽으로 크게 편향되어 있으며 중간 이상치에 의해 심하게 왜곡될 수 있습니다. 자세한 내용은 Milligan(1980) 연구 자료에서 확인하십시오 .

빠른 Ward 많은 수의 행에 대해 Ward 방법을 더 빠르게 계산하는 알고리즘을 적용합니다 . 이 알고리즘은 거리 행렬을 계산할 필요가 없으므로 계산 시간이 더 짧습니다 . 행 수가 2,000 개를 초과할 때마다 자동으로 사용됩니다 .

데이터 구조

다음 옵션은 다변량 거리를 계산하는 데 사용되는 데이터의 형식을 설명합니다 .

일반적인 데이터 각 관측값마다 하나의 행과 각 변수마다 하나의 열이 있는 직사각형 데이터입니다 .

요약된 데이터 하나 이상의 식별 열의 수준으로 요약된 데이터입니다 . 이 옵션을 선택하면 시작 창에 " 개체 ID" 텍스트 상자가 나타납니다 . 식별 열을 개체 ID 로 지정합니다 . **요약된 데이터** 옵션은 수준 평균을 계산하고 이러한 평균을 입력 데이터로 처리합니다 .

데이터가 거리 행렬임 관측값 간의 거리로 구성된 데이터입니다 . 관측값이 n 개인 경우 거리 테이블에는 n 개의 행과 $n+1$ 개의 열이 있어야 합니다 . 하나의 열 (대개 첫 번째 열) 에 n 개의 각 관측값에 대한 고유 식별자가 포함되어야 합니다 . 나머지 열에는 해당 관측값과 n 개 관측값 사이의 거리가 포함됩니다 . 다음 사항에 유의하십시오 .

- 점은 자신과의 거리가 0 이므로 테이블의 대각 요소는 0 이거나 결측이어야 합니다 . 0 이 아닌 값 또는 비결측 값은 0 으로 처리되며 보고서에 이에 대한 설명이 나타납니다 .
- 거리 열은 대칭 정방 행렬이거나 , 아래쪽 또는 위쪽 부분에 결측 항목이 있는 상삼각 또는 하삼각 행렬일 수 있습니다 . 거리가 정방 행렬로 제공된 경우 테이블이 대칭이 아니면 보고서에 경고가 나타납니다 .
- 다른 데이터 구조로 시작한 후 거리 행렬을 저장할 수 있습니다 . 자세한 내용은 " 거리 행렬 저장 " 에서 확인하십시오 .

  데이터가 거리 행렬임 옵션을 선택하는 경우 거리 열을 "Y, 열 " 로 입력하고 식별자 열을 " 라벨 " 로 입력합니다 . 라벨 열의 데이터 유형은 문자여야 합니다 . 예는 " 거리 행렬의 예 " 에서 확인하십시오 .

데이터 쌓여짐 관심 반응이 하나이고 각 개체에 대한 행이 여러 개 있는 데이터입니다 .

데이터 쌓여짐 옵션을 선택하면 시작 창에 "속성 ID" 및 "개체 ID" 텍스트 상자가 나타납니다.

– 단일 열을 "Y, 열 " 로 입력합니다 .
– "Y, 열 " 변수의 그룹화를 설명하는 열을 " 속성 ID" 로 입력합니다 . 두 개의 열만 입력하고 " 공간 측도 추가 " 를 선택하면 군집 분석에 사용할 공간 성분을 추가할 수 있습니다 . 자세한 내용은 " 공간 측도 추가 " 에서 확인하십시오 .
– 개체에 대한 식별 열을 " 개체 ID" 로 입력합니다 .

수행되는 분석은 " 속성 ID" 열을 기준으로 "Y, 열 " 변수를 분할한 후 반응 열을 표준화하지 않고 계층적 군집화를 수행하는 것과 같습니다 .

팁 : 이 옵션을 " 공간 측도 추가 " 옵션과 함께 사용하여 2 차원 공간 군집화를 수행할 수 있습니다 . 예를 들어 웨이퍼 데이터는 주로 각 다이당 하나의 행을 사용하여 기록됩니다 . 주요 관심은 웨이퍼를 군집화하는 것입니다 . 자세한 내용은 " 공간 측도를 사용한 웨이퍼 결함 분류의 예 " 에서 확인하십시오 .

주의 : 측정값 열이 하나이므로 쌓인 데이터는 " 데이터 표준화 " 옵션이 적절하지 않습니다 .

경고 - 비결측 데이터가 충분하지 않습니다 .

요약된 데이터 또는 **데이터 쌓여짐** 데이터 구조를 사용하는 경우 **비결측 데이터가 충분하지 않습니다 .** 라는 JMP 경고를 이해하기 어려울 수 있습니다 . 다음과 같은 경우에 경고가 발생합니다 .

- **일반적인 데이터**를 선택했을 때 모든 행 또는 하나를 제외한 모든 행에서 "Y, 열 " 변수의 값이 하나 이상 결측인 경우
- **요약된 데이터**를 선택하고 데이터가 " 개체 ID" 열에 대해 요약되었을 때 모든 행 또는 하나를 제외한 모든 행에서 요약된 "Y, 열 " 변수의 값이 하나 이상 결측인 경우 . 군집화 플랫폼에서 분석하는 데이터 구조를 확인하려면 테이블 > 요약을 선택한 후 " 개체 ID" 열을 " 그룹 " 으로 입력하고 "Y, 열 " 변수를 통계량 > 평균으로 입력합니다 .
- **데이터 쌓여짐**을 선택하고 " 속성 ID" 열에 대해 데이터가 분할될 때 모든 행 또는 하나를 제외한 모든 행에서 분할 "Y, 열 " 의 값이 하나 이상 결측인 경우 . 군집화 플랫폼에서 분석하는 데이터 구조를 확인하려면 **테이블 > 분할**을 선택한 후 " 속성 ID" 열을 " 분할 기준 " 으로 , "Y, 열 " 변수를 " 분할 열 " 로 , " 개체 ID" 열을 " 그룹 " 으로 각각 입력합니다 .

Y, 열 변수로 변환

다음 옵션은 군집 분석에 사용할 Y, 열 변수의 형식을 지정합니다 .

데이터 표준화 연속형 열과 순서형 열에 대해 서로 다른 측정 척도 문제를 해결합니다 . **데이터 쌓여짐** 옵션을 선택한 경우를 제외하고 각 열의 값은 열 평균을 뺀 후 열 표준편차로 나누어 표준화됩니다 . 표준화된 값으로 군집 거리를 계산하지 않으려면 " 데이터 표준화 " 체크박스를 선택 취소합니다 .

로버스트 표준화 연속형 열과 순서형 열의 평균 및 표준편차 추정값에 대한 이상치의 영향을 줄입니다 . 이 옵션은 평균 및 표준편차의 Huber M- 추정값을 사용합니다 (Huber 1964, Huber

1973, Huber 와 Ronchetti 2009 참조). 이상치가 있는 열의 경우 이 옵션을 사용하면 다변량 거리를 결정할 때 표준화된 값을 더 크게 표현할 수 있습니다 .

참고 : " 데이터 표준화 " 와 " 로버스트 표준화 " 를 모두 선택하는 경우 각 열은 로버스트 열 평균을 뺀 후 로버스트 표준편차로 나누어 표준화됩니다 . 이 옵션은 열이 서로 다른 측정 척도를 나타내는 경우 **또는** 관측값이 특정 차원에서만 이상치인 경향이 있는 경우에 유용합니다 .

참고 : " 데이터 표준화 " 를 선택하지 않고 " 로버스트 표준화 " 를 선택하는 경우 결합된 모든 열의 값에 대한 로버스트 평균과 로버스트 표준편차를 사용하여 각 열을 표준화합니다 . 이 옵션은 열이 모두 동일한 측정 척도를 나타내는 경우 **및** 관측값이 모든 차원에서 이상치인 경향이 있는 경우에 유용할 수 있습니다 .

결측값 대치 결측값을 대치합니다 . 변수 수가 50 개 이하이거나 행 수의 절반 미만이면 다변량 정규 대치법이 사용됩니다 . 그렇지 않으면 다변량 SVD 대치법이 사용됩니다 .

다변량 정규 대치법은 쌍별 공분산을 계산하여 반응 열에 대한 공분산 행렬을 생성합니다 . 그런 다음 주어진 관측값에 대해 결측값이 없는 모든 예측 변수를 사용한 회귀 예측과 동일한 방법으로 각 결측값이 대치됩니다 . 생성된 공분산 행렬이 양정치가 아니면 열 평균을 사용하여 결측값이 대치됩니다 .

다변량 SVD 대치법은 특이값 분해를 사용하여 공분산 행렬 생성을 방지합니다 . 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오 .

주의 : 결측값 대치에서는 군집이 없고 , 다변량 정규 분포의 데이터를 가져왔으며 , 값이 랜덤으로 완전히 결측되었다고 가정합니다 . 대개 이러한 가정은 실제로는 적절하지 않으므로 이 기능을 사용할 때 주의해야 합니다 . 그러나 이 기능을 사용하면 데이터의 대부분을 삭제하는 것보다 더 유익한 결과를 얻을 수 있습니다 .

공간 측도 추가 (**데이터 쌓여짐**을 데이터 구조로 선택한 경우에만 사용 가능) 데이터가 쌓여 있고 공간 좌표 (예 : 가로 및 세로 좌표) 에 해당하는 두 속성 열을 포함하는 경우 **공간 측도 추가** 옵션을 선택합니다 . 이 옵션을 선택하면 결함 패턴을 군집화하는 데 도움이 되도록 공간 성분을 선택하고 가중치를 지정할 수 있는 창이 열립니다 . 이 방법은 특별하며 매우 특정한 설정에만 적용할 수 있습니다 . 자세한 내용은 " 공간 측도 " 및 " 공간 측도를 사용한 웨이퍼 결함 분류의 예 " 에서 확인하십시오 .

계층적 군집화 보고서

" 계층적 군집화 " 보고서에는 사용된 방법 , 덴드로그램 및 " 군집화 기록 " 테이블이 표시됩니다 . 시작 창에서 라벨 열을 할당한 경우 해당 열의 값으로 덴드로그램의 각 관측값을 식별합니다 .

- " 덴드로그램 보고서 "
- " 덴드로그램 및 거리 그래프 예시 "
- " 군집화 기록 보고서 "

덴드로그램 보고서

덴드로그램은 관측값이 군집으로 병합되는 것을 나타내는 트리 다이어그램입니다. 덴드로그램은 군집의 비유사성 정도에 대한 정보도 제공합니다.

덴드로그램을 왼쪽에서 오른쪽으로 읽으면 군집화 과정을 볼 수 있습니다 . 각 단계는 **가장 가까운** 두 군집을 단일 군집으로 결합하는 것으로 구성됩니다 .

- 군집 결합은 세로선에 의해 연결되는 가로선으로 표시됩니다 .
- 세로선의 가로 위치는 지정된 수의 군집을 형성하기 위해 가장 최근에 결합된 두 군집 사이의 거리를 나타냅니다 .

참고 : 관측값 수가 256 개 미만인 경우 거리는 거리 그래프에 표시된 거리에 비례합니다 . 그렇지 않은 경우 기하 간격이 사용됩니다 . 자세한 내용은 " 덴드로그램 척도 " 에서 확인하십시오 .

다음과 같은 작업을 수행할 수 있습니다 .

- 덴드로그램의 위쪽 또는 아래쪽에 있는 다이아몬드 모양 핸들을 클릭한 채로 드래그하여 주어진 군집 수를 식별합니다 .
- 군집 줄기를 클릭하여 덴드로그램과 데이터 테이블에서 군집의 모든 멤버를 선택합니다 .

거리 그래프

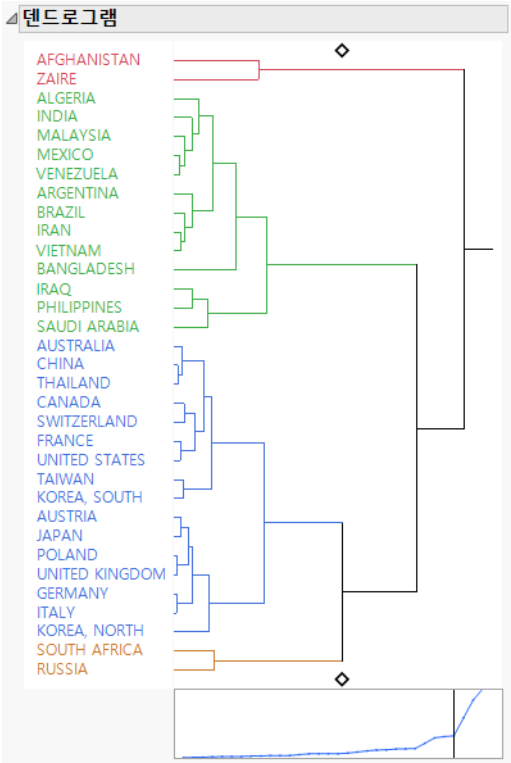
거리 그래프는 덴드로그램 아래에 나타나는 그림입니다. 이 그래프에는 두 군집이 단일 군집으로 결합되는 각 단계에 대한 점이 있습니다. 가로 좌표는 군집 수를 나타내며 왼쪽에서 오른쪽으로 감소합니다. 점의 세로 좌표는 주어진 단계에서 결합된 군집 간의 거리입니다.

덴드로그램에서 다이아몬드 모양 핸들 중 하나를 클릭한 채로 드래그하면 선택된 군집 수를 제어할 수 있습니다. 다이아몬드를 클릭한 채로 드래그하면 군집 수에 따라 움직이는 세로선이 그림에 나타납니다. 거리 그래프의 기울기가 평평해지는 점이 종종 있습니다. 이러한 점은 자연적인 단락점을 나타내며 군집 수를 결정하는 데 도움이 됩니다.

덴드로그램 및 거리 그래프 예시

"군집화의 예"의 Birth Death Subset.jmp에 대한 덴드로그램 보고서를 예로 들어보겠습니다.

그림 12.5 Birth Death Subset.jmp 에 대한 덴드로그램 보고서



다이아몬드는 네 개의 군집 위치에서 설정됩니다. 네 개의 군집 모형을 형성하기 위해 가장 최근에 결합된 두 군집은 Algeria ~ Bangladesh로 구성된 군집과 Iraq ~ Saudi Arabia로 구성된 군집입니다. 이러한 두 군집 간의 거리는 다이아몬드가 4로 설정될 때 거리 그림에서 세로선으로 표시된 점입니다. 거리는 "군집화 기록" 보고서에서 "군집 수" = 4 옆에 제공됩니다. 여기서 거리는 1.618708760이고, Algeria와 Iraq로 시작하는 군집이 결합되어 네 개의 군집을 생성한다는 것을 알 수 있습니다.

다섯 개의 군집을 생성하기 위해 결합되는 두 군집은 Australia ~ Korea, South로 구성된 군집과 Austria ~ Korea, North로 구성된 군집입니다. 덴드로그램에서 이러한 군집의 세로 결합선은 네 개의 군집 모형을 형성하기 위해 결합된 군집, 즉 Algeria ~ Bangladesh와 Iraq ~ Saudi Arabia 군집에 대한 세로 결합선의 왼쪽에서 거의 동일한 가로 거리 위치에 있습니다. 따라서 네 개의 군집 모형을 형성하기 위해 결합된 군집과 다섯 개의 군집 모형을 형성하기 위해 결합된 군집 사이의 거리에 큰 차이가 없습니다.

거리 그림이 네 개의 군집 모형부터 안정화된다는 것은 해당 점까지의 군집 그룹화가 군집 간의 거리를 크게 설명하지 못한다는 것을 나타냅니다. 그러나 네 개 군집 모형에서는 군집 간의 분리가 양호하다는 것을 보여 줍니다.

군집화 기록 보고서

" 군집화 기록 " 테이블에는 군집화 기록이 포함됩니다 .

군집 수 피 군집 개체와 군집 개체로 표시된 결합이 수행된 **후** 생성되는 군집 수를 나열합니다 . 군집 수는 첫 번째 결합으로 시작합니다 (*n - 1* 개의 군집이 있는 경우). 여기서 *n* 은 개체 수입니다 . 보고서에는 모든 개체가 하나의 군집에 포함될 때까지 군집 수가 내림차순으로 나열됩니다 . 군집화 기록은 이러한 방식으로 덴드로그램의 순서 (왼쪽에서 오른쪽으로) 를 따릅니다 .

거리 시작 창에서 선택한 거리 방법에 따라 계산된 군집 간의 거리입니다 . 자세한 내용은 " 거리 계산 방법 " 에서 확인하십시오 .

피 군집 개체 덴드로그램에서 결합되는 첫 번째 군집의 대표 개체입니다 . " 피 군집 개체 " 열에 표시된 군집 순서와 대표 개체는 데이터 정렬 방식에 따른 것이며 내재된 의미는 없습니다 .

군집 개체 덴드로그램에서 결합되는 두 번째 군집의 대표 개체입니다 . " 군집 개체 " 열에 표시된 군집 순서와 대표 개체는 데이터 정렬 방식에 따른 것이며 내재된 의미는 없습니다 .

계층적 군집화 옵션

" 계층적 군집화 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

군집 색 표시 소속 군집에 따라 덴드로그램 및 연결된 결합 막대의 라벨 색상을 지정합니다 . 데이터 테이블의 행에도 해당 색상을 할당합니다 . 군집 수를 변경하면 색상이 업데이트됩니다 . 이 옵션을 선택 취소하면 군집 수에 따라 색상이 더 이상 업데이트되지 않습니다 .

군집 표시 행이 속한 군집에 해당하는 데이터 테이블의 행에 표식을 할당합니다 . 군집 수를 변경하면 표식이 업데이트됩니다 . 이 옵션을 선택 취소하면 군집 수에 따라 표식이 더 이상 업데이트되지 않습니다 .

군집 수 군집 수 입력 상자가 나타나고 덴드로그램 슬라이더를 해당 숫자에 배치합니다 .

군집 기준 전체 군집 수 범위에 대한 CCC(Cubic Clustering Criterion) 를 제공합니다 . CCC 는 군집 수를 추정하는 데 사용됩니다 . 이 통계량은 거리 기반 군집화 알고리즘과 함께 사용할 수 있습니다 . CCC 값이 클수록 군집 수 측면에서 더 적합하다는 것을 나타냅니다 . 자세한 내용은 SAS Institute Inc. (1983) 에서 확인하십시오 . **데이터가 거리 행렬임**을 선택한 경우에는 사용할 수 없습니다 .

덴드로그램 표시 " 덴드로그램 " 보고서를 표시하거나 숨깁니다 .

덴드로그램 척도 덴드로그램 척도를 지정하기 위한 다음 옵션을 포함합니다 .

거리 간격 시작 창에 지정된 거리 방법을 기반으로 두 결합점 사이의 가로 거리를 해당 점에서 결합된 두 군집 간의 거리로 표시합니다 . 거리 척도는 거리 그래프에 사용되는 척도와 동일하며 덴드로그램의 기본 척도입니다 .

균등 간격 두 결합점 사이의 가로 거리를 동일하게 표시합니다 .

기하 간격 군집 수가 증가함에 따라 결합점 사이의 가로 거리를 증가합니다 . 이 옵션은 개체가 많고 , 큰 군집보다 작은 군집을 더 잘 보이게 표시하려는 경우에 유용합니다 .

거리 그래프 덴드로그램 아래에 거리 그림을 표시하거나 숨깁니다 .

군집 수 핸들 표시 덴드로그램에서 군집 수를 수동으로 변경하는 데 사용되는 핸들을 표시하거나 숨깁니다 .

선택 행 확대 / 축소 덴드로그램에서 군집을 선택한 후 특정 군집을 선택하고 확대합니다 . 또는 군집을 두 번 클릭하여 확대할 수 있습니다 . 원래 보기로 돌아가려면 " 확대 / 축소 해제 " 를 사용합니다 .

확대 / 축소 해제 확대 / 축소 후 덴드로그램을 원래 보기로 되돌립니다 .

선택한 군집에서 피벗 현재 선택된 군집의 두 하위 군집 순서를 역순으로 바꿉니다 .

색상 맵 값별로 색상이 지정된 각 "Y, 열 " 변수를 표시하는 색상 맵 또는 히트맵을 추가하기 위한 옵션을 제공합니다 . 하위 메뉴에 몇 가지 색상 테마 선택 옵션이 제공됩니다 .

이원 군집화 행과 "Y, 열 " 에 지정된 변수를 기준으로 군집화합니다 . 맨 아래의 "Y, 열 " 변수에 대한 색상 맵이 덴드로그램에 추가됩니다 . 일반적으로 이원 군집화의 경우 변수가 동일한 척도로 측정되므로 " 데이터 표준화 " 를 선택하지 않습니다 . **데이터 쌓여짐**을 선택한 경우에는 사용할 수 없습니다 .

위치 지정 라벨 및 덴드로그램의 다른 부분에 대한 위치를 변경하는 옵션을 제공합니다 .

범례 색상 맵에 사용된 색상의 범례를 표시하거나 숨깁니다 . 이 옵션은 색상 맵이 활성화된 경우에만 사용할 수 있습니다 .

추가 색상 맵 열 지정된 열에 대한 색상 맵을 추가합니다 . **요약된 데이터** , **데이터가 거리 행렬임** 또는 **데이터 쌓여짐**을 선택한 경우에는 사용할 수 없습니다 .

성운형 그림 계층적 군집화 덴드로그램의 정보를 표시하는 대체 방법을 표시하거나 숨깁니다 . 각 관측값 (행) 을 끝점으로 나타내고 각 군집 결합을 새 점으로 나타냅니다 . 표시된 선은 소속 군집을 나타냅니다 . 선 길이는 군집 간 거리를 나타냅니다 . 선이 길수록 군집 간의 거리가 더 멀다는 것을 나타냅니다 .

성운형 그림의 선을 커서로 가리키면 길이를 확인할 수 있습니다 . 그러나 길이 값은 서로에 대해서만 의미가 있습니다 . 축 척도 , 점 방향 및 선 각도는 임의적입니다 . 단지 , 노드 끝 사이에 간격을 두고 그림이 복잡해 보이지 않도록 값이 결정되므로 데이터 집합이 더 큰 경우 중요합니다 .

끝점의 라벨을 해제하려면 성운형 그림 안쪽을 마우스 오른쪽 버튼으로 클릭하고 **라벨 표시**를 선택 취소합니다 .

성운형 좌표 저장 성운형 그림의 좌표를 데이터 테이블에 저장합니다 . **요약된 데이터** , **데이터가 거리 행렬임** 또는 **데이터 쌓여짐**을 선택한 경우에는 사용할 수 없습니다 .

군집 저장 군집 번호를 포함하는 데이터 테이블 열을 생성합니다 . 시작 창에서 " 공간 측도 추가 " 를 선택한 경우 군집 번호가 Hough 데이터 테이블에도 저장됩니다 .

가장 근접한 군집에 대한 계산식 저장 가장 가까운 군집에 대한 계산식을 포함하는 데이터 테이블 열을 생성합니다 . 이 옵션은 각 군집의 중심에 대한 유클리드 거리 제곱을 계산하고 가장 가까운 군집을 선택합니다 . 군집이 다르게 결정되므로 이 계산식은 계층적 군집화에서 제공된

군집 할당을 재현하지 못할 수도 있습니다. 그러나 군집 할당은 매우 유사합니다. **요약된 데이터**, **데이터가 거리 행렬임** 또는 **데이터 쌓여짐**을 선택한 경우에는 사용할 수 없습니다.

표시 순서 저장 행이 덴드로그램에 나타나는 순서를 포함하는 데이터 테이블 열을 생성합니다.

군집 계층 저장 사용자 덴드로그램에 대한 스크립트를 작성하는 데 필요한 정보를 포함하는 데이터 테이블을 생성합니다. 각 군집 결합에 대해 세 개의 행이 있습니다. 첫 번째 행은 군집 개체, 두 번째 행은 피 군집 개체, 세 번째 행은 군집 중심, 크기 및 기타 정보를 제공하는 결과에 대한 것입니다.

군집 트리 저장 JMP 와 SAS 간의 군집 트리를 비교하는 데 필요한 정보를 포함하는 새 데이터 테이블을 생성합니다. 각 군집 결합에 대해 새 군집마다 군집 크기와 기타 정보가 포함된 행이 각각 하나씩 있습니다.

거리 행렬 저장 관측값 사이의 거리를 포함하는 새 데이터 테이블을 생성합니다.

군집 평균 저장 각 군집에서 각 열의 평균과 행 수를 포함하는 새 데이터 테이블을 생성합니다.

군집 요약 (**데이터가 거리 행렬임**을 선택한 경우에는 사용 불가능) 다음 정보를 표시합니다.

군집 평균 각 군집에 대해 관측값 수 (쌓여진 데이터의 경우 개체 ID) 및 각 변수의 평균을 제공하는 테이블입니다.

군집 표준편차 각 군집에 대해 관측값 수 (쌓여진 데이터의 경우 개체 ID) 및 각 변수의 표준편차를 제공하는 테이블입니다.

군집 평균 그림 군집 평균의 평행 그림 또는 2 차원 히트맵입니다.

데이터 쌓여짐을 선택하고 두 개의 " 속성 ID" 변수가 있는 경우 외에는 평행 그림이 표시됩니다. 평행 그림의 경우 각 변수의 축이 척도화됩니다.

– " 데이터 표준화 " 를 선택한 경우 축 범위는 평균의 위와 아래로 2 표준편차이며, 여기서 표준편차와 평균은 원시 데이터에 대해 계산됩니다. 군집 평균이 이 범위를 벗어나면 이 값을 포함하도록 축이 확장됩니다.

– " 데이터 표준화 " 를 선택하지 않은 경우 척도가 표시되는 공통 세로 축이 있습니다. 척도는 " 그래프 빌더 " 의 " 균등 척도 " 옵션과 동일합니다.

데이터 쌓여짐을 선택하고 두 개의 " 속성 ID" 변수가 있는 경우 각 군집에 대해 각 위치에서 Y 변수의 평균을 보여 주는 2 차원 그림이 표시됩니다. 이러한 그림은 " 파랑 - 회색 - 빨강 " 색상 그래디언트를 사용하여 색상이 적용됩니다.

열 요약 각 변수에 대해 군집에 의해 설명되는 변동 비율을 나타내는 R^2 값을 제공합니다. 이 값은 군집의 변수 회귀에 대한 R^2 값입니다. 이 옵션은 R^2 값의 막대 그래프도 제공합니다.

산점도 행렬 모든 변수를 사용하여 산점도 행렬을 생성합니다. **요약된 데이터**, **데이터가 거리 행렬임** 또는 **데이터 쌓여짐**을 선택한 경우에는 사용할 수 없습니다.

평행 좌표 그림 각 군집에 대한 평행 좌표 그림을 생성합니다. **요약된 데이터**, **데이터가 거리 행렬임** 또는 **데이터 쌓여짐**을 선택한 경우에는 사용할 수 없습니다. 축은 " 군집 평균 그림 " 에 설명된 대로 척도화됩니다. 자세한 내용은 " 군집 평균 그림 " 에서 확인하십시오.

군집 처리 비교 (Shift 키를 누른 채 " 계층적 군집화 " 의 빨간색 삼각형을 클릭한 경우에만 사용 가능) 반응 열과 2 수준 처리 열을 선택합니다 . " 계층적으로 군집화된 차이 " 보고서를 생성합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

계층적 군집화 플랫폼의 추가 예

- " 거리 행렬의 예 "
- " 공간 측도를 사용한 웨이퍼 결함 분류의 예 "

거리 행렬의 예

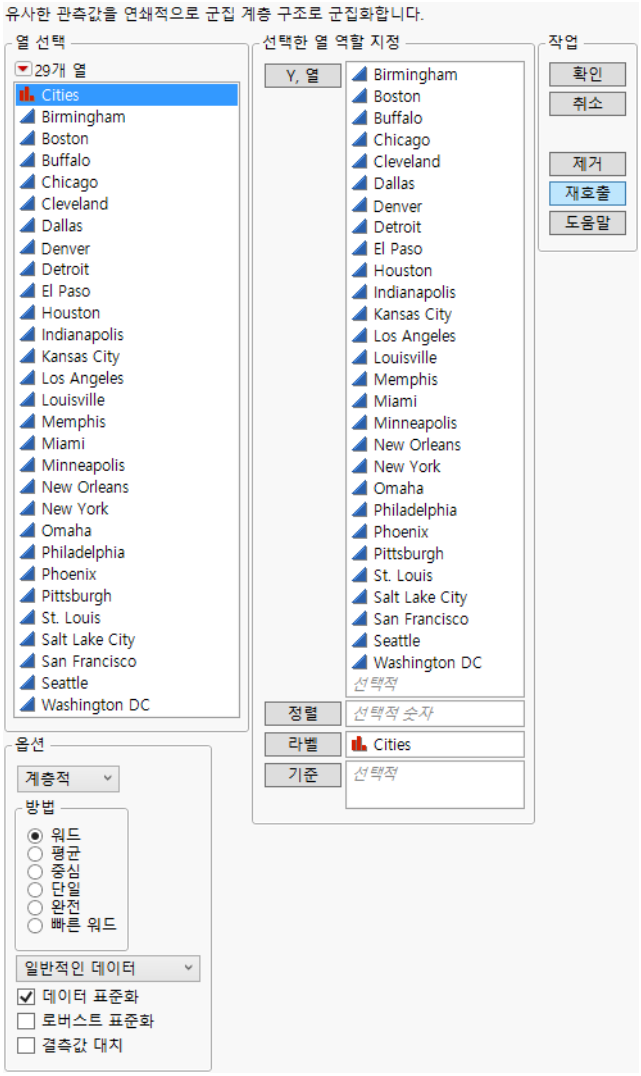
거리 행렬에 대한 적절한 데이터 테이블 구조는 다음으로 구성됩니다 .

- 문자 데이터 유형을 사용하는 식별자 열 (대개 첫 번째 열)
- *n* 개 열의 집합 . 여기서 *n* 은 행 수이기도 합니다 . 이러한 *n* 개의 열은 대각선에 0 또는 결측값이 있는 대칭 행렬을 정의합니다 .

Flight Distances.jmp 의 거리 행렬은 위의 형식을 따릅니다 .

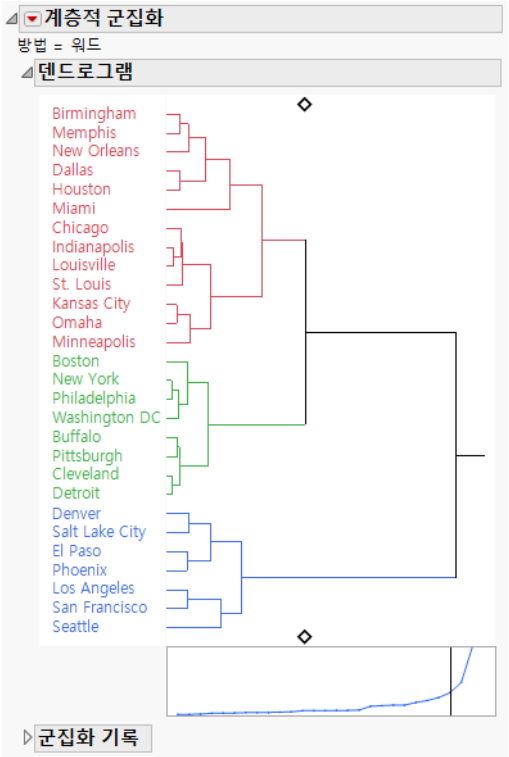
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Flight Distances.jmp 를 엽니다 .
2. **분석 > 군집화 > 계층적 군집화**를 선택합니다 .
3. 시작 창 왼쪽 아래에 있는 목록에서 **일반적인 데이터**를 **데이터가 거리 행렬임**으로 변경합니다 .
4. Cities 를 선택하고 **라벨**을 클릭합니다 .
5. 나머지 열을 모두 선택하고 **Y, 열**을 클릭합니다 .

그림 12.6 완료된 거리 행렬 시작 창



6. **확인**을 클릭합니다.
7. "계층적 군집화"의 빨간색 삼각형을 클릭하고 **군집 색 표시**를 선택합니다.

그림 12.7 Flight Distances 에 대한 덴드로그램 보고서



비행 거리 "덴드로그램" 보고서의 다이아몬드 위치는 모형이 도시를 세 개의 군집으로 그룹화했음을 나타냅니다. 군집은 덴드로그램에서 색상으로 구분됩니다. 보고서 해석 방법에 대한 자세한 내용은 "덴드로그램 보고서"에서 확인하십시오.

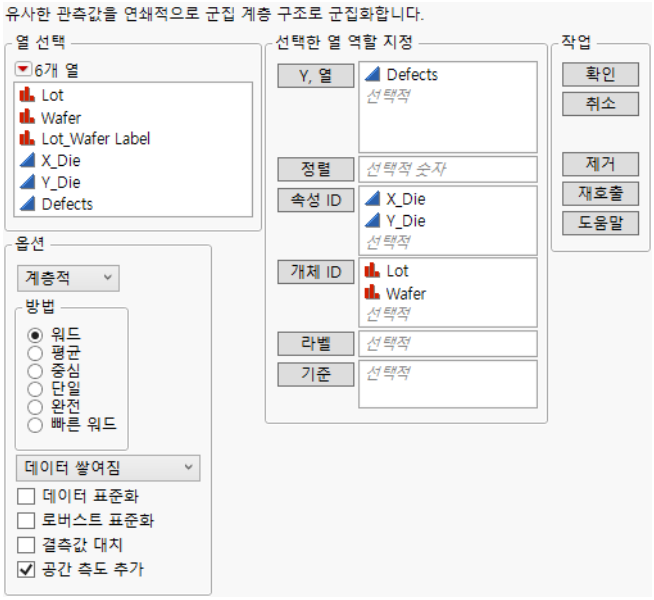
공간 측도를 사용한 웨이퍼 결함 분류의 예

계층적 군집화 플랫폼에서는 "공간 측도"라는 특별한 군집화 옵션을 사용할 수 있습니다. 이 예에서는 이 옵션을 사용하여 웨이퍼를 군집화합니다. 옵션에 대한 자세한 내용은 "공간 측도"에서 확인하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Wafer Stacked.jmp 를 엽니다.
2. **분석 > 군집화 > 계층적 군집화**를 선택합니다.
3. 왼쪽 아래에 있는 목록에서 **일반적인 데이터**를 **데이터 쌓여짐**으로 변경합니다.
   쌓인 데이터에 대한 추가 옵션이 시작 창에 나타납니다.
4. Defects 를 선택하고 **Y, 열**을 클릭합니다.
5. X_Die 와 Y_Die 를 선택하고 **속성 ID** 를 클릭합니다.
6. Lot 와 Wafer 를 선택하고 **개체 ID** 를 클릭합니다.

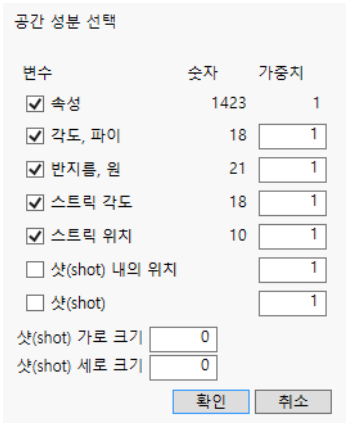
7. 왼쪽 아래의 옵션 목록에서 **공간 측도 추가**를 선택합니다 .

그림 12.8 완료된 군집화 시작 창



8. **확인**을 클릭합니다 .

그림 12.9 공간 성분 창



1423 개 위치에서 Defects 가 측정되었으므로 1423 개의 " 속성 " 변수가 있습니다 .

9. **확인**을 클릭하여 " 공간 " 창의 선택 내용을 적용합니다 .

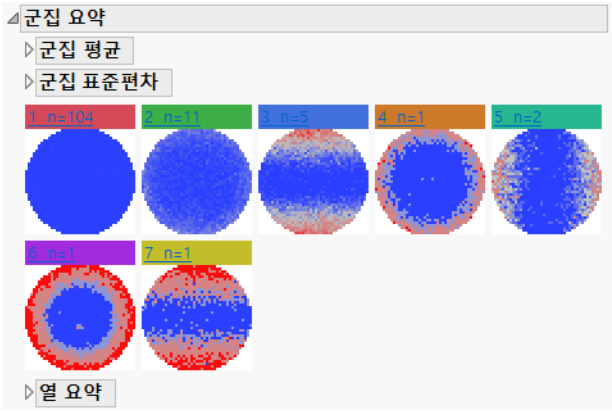
" 계층적 군집화 " 보고서와 "Wafer Stacked Defects 공간 " 데이터 테이블이라는 두 개의 창이 열립니다 .

10. " 덴드로그램 " 그림에서 맨 위에 있는 다이아몬드 모양 핸들을 클릭한 채로 드래그하여 다양한 군집 수를 탐색합니다 .

    핸들을 드래그하면 덴드로그램 아래의 거리 그래프에서 세로선이 해당 군집 수로 움직입니다 . 세로 좌표는 주어진 단계에서 결합된 군집 간의 거리를 제공합니다 . 군집 수가 7 일 때 그래프가 수평이 되는 것 같습니다 .

11. " 계층적 군집화 " 의 빨간색 삼각형을 클릭하고 **군집 수**를 선택합니다 .
12. "7" 을 입력하고 **확인**을 클릭합니다 .
13. " 계층적 군집화 " 의 빨간색 삼각형을 클릭하고 **군집 요약**을 선택합니다 .

그림 12.10 군집 요약 보고서



웨이퍼 맵은 각 군집에 대한 결함의 공간적 특성을 나타냅니다 . 군집 1 에는 104 개의 웨이퍼가 포함되어 있으며 비교적 적은 수의 결함이 웨이퍼 전체에 퍼져 있습니다 . 군집 3 에는 5 개의 웨이퍼가 있으며 상단 및 하단 반구체의 극단에 결함이 집중되어 있습니다 . 군집 분석으로 생성된 데이터 테이블에서 개별 웨이퍼에 대한 맵과 Hough 공간 맵을 볼 수 있습니다 . 자세한 내용은 " 공간 측도 " 에서 확인하십시오 .

계층적 군집화 플랫폼에 대한 통계 상세 정보

- " 공간 측도 "
- " 거리 방법 계산식 "

공간 측도

"공간 측도 추가" 옵션을 사용하려면 데이터가 쌓여 있고, 공간 좌표에 해당하는 두 개의 속성 열이 포함되어 있어야 합니다. 일부 공간 측도는 Hough 변환을 사용하여 생성됩니다. 자세한 내용은 White et al. (2008) 및 Ballard(1981) 연구 자료에서 확인하십시오. "공간 측도를 사용한 웨이퍼 결함 분류의 예"의 내용도 참조하십시오.

공간 성분 선택 창

시작 창에서 다음을 수행하면 " 공간 성분 선택 " 창이 나타납니다 .

- " 데이터 쌓여짐 " 데이터 구조 선택
- 공간 좌표에 해당하는 두 개의 열을 " 속성 ID" 로 지정
- " 개체 ID" 지정
- " 공간 측도 추가 " 선택

" 공간 성분 선택 " 창에서 군집 분석에 사용할 공간 성분을 선택하고 가중치를 지정합니다 . 이러한 성분은 군집 분석에 사용되는 변수를 생성하는 데 사용됩니다 . 각 개체에 대한 행이 있는 새 테이블이 열립니다 . 이 테이블에는 각 개체에 대해 계산된 공간 성분이 포함되어 있습니다 .

변수 군집 분석에서 생성 및 사용되는 변수 유형입니다 . 변수는 공간 성분과 반응 Y 를 사용하여 생성됩니다 .

속성 각 개체에 대해 두 개의 " 속성 ID" 변수로 정의된 각 위치에서 계산된 Y 변수의 값입니다 .

각도 , 파이 파이 모양 또는 반구체 모양을 반영하는 변수입니다 .

반지름 , 원 원형 모양을 반영하는 변수입니다 .

스트릭 각도 동일한 각도의 스트릭을 반영하는 변수입니다 .

스트릭 위치 동일한 공간 위치의 스트릭을 반영하는 변수입니다 .

샷 (shot) 내의 위치 샷 (shot) 에서 다이 위치를 기반으로 하는 변수입니다 . " 샷 (shot) 내의 위치 " 변수는 "ShotPos[vShotSize, hShotSize]" 로 표시되며 , 여기서 vShotSize 와 hShotSize 는 정의된 세로 및 가로 샷 (shot) 크기입니다 .

샷 (shot) 개체가 있는 직사각형을 식별하는 변수이며 , 직사각형에 있는 개체의 가로 및 세로 위치 수를 지정합니다 . 샷 (shot) 이라는 용어는 반도체 웨이퍼 데이터에서 웨이퍼 전체에 대해 함께 이미징되는 다이를 식별하기 위해 사용됩니다 .

" 샷 (shot) 가로 크기 " 및 " 샷 (shot) 세로 크기 " 에 값을 입력합니다 . 가로 샷 (shot) 크기를 4 로 지정하고 세로 샷 (shot) 크기를 5 로 지정하면 한 샷 (shot) 에 최대 20 개의 다이가 있음을 나타냅니다 . 생성되는 총 식별자 수는 다음과 같이 계산됩니다 .

floor[(hSize+hShotSize-1)/hShotSize]* floor[(vSize+vShotSize-1)/vShotSize]

여기서 hSize 와 vSize 는 각각 최대 가로 위치 수와 최대 세로 위치 수이며 , hShotSize = 샷 (shot) 가로 크기이고 vShotSize = 샷 (shot) 세로 크기입니다 .

참고 : 샷 (shot) 변수는 "Shot[vert, horiz]" 로 표시되며 , 여기서 vert 와 horiz 는 각각 세로 및 가로 다이 위치를 나타냅니다 .

개수 주어진 유형으로 생성되는 총 변수 수입니다 .

가중치 군집을 결정하는 데 사용되는 주어진 유형의 변수에 대한 중요도 측도입니다 .

공간 측도 보고서

"공간 성분 선택" 창에서 "확인"을 클릭하면 두 개의 창이 나타납니다.

계층적 군집화 보고서

쌓인 데이터와 두 개의 속성 ID를 사용하여 분석을 수행하면 "군집 요약" 보고서에 Y 변수의 공간 맵이 표시됩니다. 각 그림은 "속성 ID" 변수로 정의된 각 위치에 대한 군집 평균을 표시하는 2차원 그림입니다. 이 그림에는 파랑-회색-빨강 색상 그래디언트와 분위수 척도가 사용됩니다. 분위수 척도를 사용하면 이상치로 인한 영향이 줄어듭니다.

공간 데이터 테이블

공간 측도 데이터 테이블에는 고유한 각 개체 ID 에 대한 행이 있습니다 . 열은 Y 변수를 나타내기 위해 파랑 - 회색 - 빨강 기본 색상 그래디언트를 사용하여 표시됩니다 . 이 테이블에는 다음 열이 포함됩니다 .

Object 두 개의 " 속성 ID" 변수로 정의된 각 공간 위치에서 Y 변수의 히트맵을 표시하는 표현식 열입니다 .

Hough 각 개체에 대한 Hough 공간의 히트맵을 표시하는 표현식 열입니다 . 자세한 내용은 White et al. (2008) 연구 자료에서 확인하십시오 .

공간 측도 각 개체에 대해 계산된 값을 표시하는 각 공간 측도의 열입니다 . 셀은 값에 따라 색상이 지정됩니다 .

거리 방법 계산식

이 섹션에서는 시작 창에서 선택한 방법을 기반으로 거리를 계산하는 데 사용되는 계산식을 제공합니다. 방법에 대한 설명은 "거리 계산 방법"에서 확인하십시오.

계산식에는 다음 표기가 사용됩니다 . 여기서 소문자 기호는 일반적으로 관측값과 관련되고 대문자 기호는 군집과 관련됩니다 .

n = 관측값 수

v = 변수 수

x_i = i 번째 관측값

C_K = K 번째 군집 ({1, 2,..., n} 의 부분집합)

N_K = C_K 의 관측값 수

$\bar{x}$ = 표본 평균 벡터

$\bar{x}_K$ = C_K 군집의 평균 벡터

$\|x\|$ = **x** 요소의 제곱합 제곱근 (**x** 벡터의 유클리드 길이)

$$d(x_i, x_j) = \left\| x_i - x_j \right\|^2$$

평균 연결법 평균 연결 군집 방법의 거리는 다음과 같습니다 .

$$D_{KL} = \sum_{i \in C_K} \sum_{j \in C_L} \frac{d(x_i, x_j)}{N_K N_L}$$

중심 방법 중심 군집화 방법의 거리는 다음과 같습니다 .

$$D_{KL} = \left\| \overline{x}_K - \overline{x}_L \right\|^2$$

Ward Ward 방법의 거리는 다음과 같습니다 .

$$D_{KL} = \frac{\left\| \overline{x}_K - \overline{x}_L \right\|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$

단일 연결법 단일 연결 군집 방법의 거리는 다음과 같습니다 .

$$D_{KL} = \min_{i \in C_K} \min_{j \in C_L} d(x_i, x_j)$$

완전 연결법 완전 연결 군집 방법의 거리는 다음과 같습니다 .

$$D_{KL} = \max_{i \in C_K} \max_{j \in C_L} d(x_i, x_j)$$

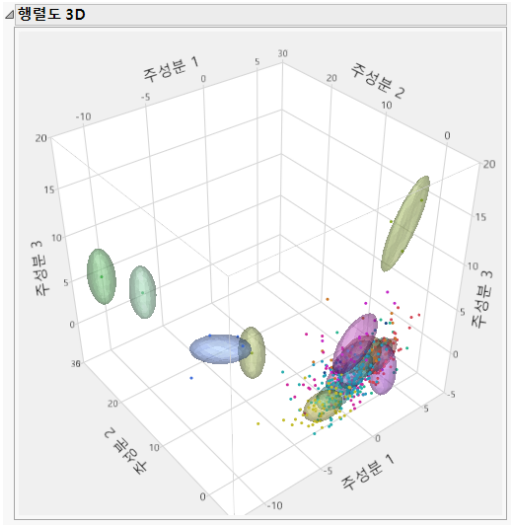
13 장

K 평균 군집화
거리를 사용하여 관측값 그룹화

K 평균 군집화 플랫폼을 사용하여 여러 변수에 대해 유사한 값을 공유하는 관측값을 그룹화할 수 있습니다. 약 200 개 ~ 100,000 개의 관측값을 포함하는 큰 데이터 테이블에 *k* 평균 방법을 사용합니다.

K 평균 군집화 플랫폼에서는 관측값을 분할하는 반복 알고리즘을 사용하여 지정된 수의 군집을 생성합니다. k **평균**이라고 하는 이 방법은 군집 중심까지의 거리를 최소화하기 위해 관측값을 군집으로 분할합니다. 군집 수 *k*를 미리 지정해야 합니다. 그러나 여러 *k* 값의 결과를 비교하여 데이터에 대한 최적 군집 수를 선택할 수 있습니다.

그림 13.1 3D 행렬도



목차

K 평균 군집화 플랫폼 개요 273
관측값 군집화 플랫폼 개요 273
K 평균 군집화의 예 274
K 평균 군집화 플랫폼 시작 277
반복적 군집화 보고서 278
반복적 군집화 옵션 278
반복적 군집화 제어판 279
K 평균 보고서 280
군집 비교 보고서 280
K 평균 보고서 280
K 평균 보고서 옵션 280
자기 조직화 지도 282
자기 조직화 지도 제어판 282
자기 조직화 지도 보고서 283
SOM 알고리즘 설명 284
K 평균 군집화 플랫폼의 추가 예 284
자기 조직화 지도의 예 284

K 평균 군집화 플랫폼 개요

K 평균 군집화는 JMP에서 관측값을 군집화하기 위해 제공하는 네 가지 플랫폼 중 하나입니다. 네 가지 방법을 모두 비교하려면 "관측값 군집화 플랫폼 개요"의 내용을 참조하십시오.

K 평균 군집화 플랫폼에서는 반복 적합 프로세스를 사용하여 지정된 수의 군집을 형성합니다. *k* 평균 알고리즘은 먼저 **군집 시드**라고 하는 *k*개의 점 집합을 군집 평균의 초기 추측 값으로 선택합니다. 각 관측값은 가장 가까운 군집 시드에 할당되어 임시 군집 집합을 형성합니다. 그런 다음 시드가 군집 평균으로 대체되고 점이 재할당되며 군집에서 더 이상 변화가 없을 때까지 프로세스가 계속됩니다.

k 평균 알고리즘은 **EM 알고리즘**의 특수한 경우로, 여기서 *E*는 기대(Expectation), *M*은 최대화(Maximization)를 나타냅니다. *k* 평균 알고리즘의 경우 임시 군집 평균의 계산은 기대 단계를 나타내고 가장 가까운 군집에 점을 할당하는 것은 최대화 단계를 나타냅니다.

K 평균 군집화는 숫자 열만 지원합니다. *K* 평균 군집화는 모델링 유형(명목형 및 순서형)을 무시하고 모든 숫자 열을 연속형으로 처리합니다.

군집 수 *k* 또는 *k* 값 범위를 미리 지정해야 합니다. 그러나 여러 *k* 값의 결과를 비교하여 데이터에 대한 최적 군집 수를 선택할 수 있습니다.

K 평균 군집화에 대한 배경 정보는 SAS Institute Inc. (2020d)의 "FASTCLUS Procedure" 장 및 Hastie et al. (2009) 연구 자료에서 확인하십시오.

관측값 군집화 플랫폼 개요

군집화는 여러 변수에 대해 유사한 값을 공유하는 관측값을 함께 그룹화하는 다변량 기법입니다. 일반적으로 관측값은 *p* 차원 공간에 고르게 분산되지 않습니다. 여기서 *p*는 변수 수입니다. 대신 관측값은 군집을 형성합니다. 이러한 군집을 식별하면 데이터를 더 세밀하게 파악할 수 있습니다.

참고 : JMP 에서는 변수를 군집화할 수 있는 플랫폼도 제공합니다 . 자세한 내용은 " 변수 군집화 " 에서 확인하십시오 .

JMP 에서 제공하는 다음 네 가지 플랫폼을 사용하여 관측값을 군집화할 수 있습니다 .

- 계층적 군집화는 최대 수만 개의 행이 있는 작은 테이블에 유용하며 문자 데이터를 허용합니다 . 계층적 군집화는 행을 계층적 순서로 결합하여 트리로 표현합니다 . 트리가 생성되면 데이터에 가장 적절한 군집 수를 선택할 수 있습니다 .
- K 평균 군집화는 최대 수백만 개의 행이 있는 큰 테이블에 적절하며 숫자 데이터만 허용합니다 . 군집 수 *k* 를 미리 지정해야 합니다 . 이 알고리즘은 군집 시드 점을 추측합니다 . 그런 다음 군집에 점을 번갈아 할당하고 군집 중심을 다시 계산하는 반복 과정을 수행합니다 .
- 정규 혼합은 중첩될 수 있는 다변량 정규 분포의 혼합에서 데이터를 수집한 경우에 적절하며 숫자 데이터만 허용합니다. 다변량 이상치가 있는 경우 균등 분포를 가정하고 이상치 군집을 사용할 수 있습니다 .

군집 수를 미리 지정해야 합니다. 최대 가능도는 혼합 비율과 평균, 표준편차 및 상관을 공동으로 추정하는 데 사용됩니다. 각 그룹에 속할 확률이 각 점에 할당됩니다. EM 알고리즘은 추정값을 구하는 데 사용됩니다.

- 잠재 계층 분석은 대부분의 변수가 범주형일 때 적절합니다. 군집 수를 미리 지정해야 합니다. 이 알고리즘은 다항 혼합 분포를 가정하는 모형을 적합시킵니다. 각 관측값에 대해 소속 군집의 최대 가능도 추정값이 계산됩니다. 관측값은 소속 확률이 가장 큰 군집으로 분류됩니다.

표 13.1 군집화 방법 요약

방법	데이터 유형 또는 모델링 유형	데이터 테이블 크기	군집 수 지정
계층적 군집화	임의	빠른 Ward 를 사용하는 경우 최대 200,000 개 행 다른 방법을 사용하는 경우 최대 5,000 개 행	아니요
K 평균 군집화	수치형	최대 수백만 개 행	예
정규 혼합	수치형	모든 크기	예
잠재 계층 분석	명목형 또는 순서형	모든 크기	예

일부 군집화 플랫폼에는 데이터의 이상치를 처리하는 옵션이 있습니다. 그러나 데이터에 이상치가 있으면 분석하기 전에 먼저 이상치를 탐색하는 것이 가장 좋습니다. 이 작업은 이상치 탐색 유틸리티를 사용하여 수행할 수 있습니다. 자세한 내용은 의"모델링 유틸리티" 장에서 확인하십시오.

K 평균 군집화의 예

이 예에서는 Cytometry.jmp 샘플 데이터 테이블을 사용하여 K 평균 군집화로 관측값을 군집화합니다. 세포분석법은 세포 표면의 표지를 감지하는 데 사용되며 이러한 표지의 판독값은 특정 질병을 진단하는 데 도움이 됩니다. 이 예에서 관측값은 세포분석에서 감지된 네 개의 표지 판독값을 기반으로 그룹화됩니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cytometry.jmp 를 엽니다.
2. **분석 > 군집화 > K 평균 군집화**를 선택합니다.
3. CD3, CD8, CD4 및 MCB 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. **군집 수** 옆에 "3" 을 입력합니다.
6. **군집 범위 (선택 사항)** 옆에 "15" 를 입력합니다.

군집 범위가 15 로 설정되었으므로 플랫폼에서 3 개 ~ 15 개 군집에 대한 적합을 제공합니다 . 그런 다음 원하는 군집 수를 결정할 수 있습니다 .

7. **시작**을 클릭합니다 .

그림 13.2 군집 비교 보고서

군집 비교

방법	군집 수	CCC	최선
K 평균 군집화	3	23.1784	
K 평균 군집화	4	8.80709	
K 평균 군집화	5	29.5123	
K 평균 군집화	6	52.5517	
K 평균 군집화	7	49.5876	
K 평균 군집화	8	56.5308	
K 평균 군집화	9	54.053	
K 평균 군집화	10	69.8707	
K 평균 군집화	11	70.5239	최적 CCC
K 평균 군집화	12	61.5326	
K 평균 군집화	13	68.1277	
K 평균 군집화	14	66.4044	
K 평균 군집화	15	69.9928	

" 군집 비교 " 보고서가 보고서 창 맨 위에 나타납니다 . 최량 적합은 가장 큰 CCC 값에 의해 결정됩니다 . 이 경우에는 11 개 군집을 적합시킬 때 최량 적합이 발생합니다 .

8. "K 평균 군집 수 =11" 보고서가 나올 때까지 스크롤합니다 .

그림 13.3 K 평균 군집 수 =11 보고서

K 평균 군집 수=11

개별적으로 척도화된 열

군집 요약

군집	개수	단계	기준
1	816	24	0
2	482		
3	157		
4	577		
5	856		
6	498		
7	447		
8	549		
9	377		
10	123		
11	118		

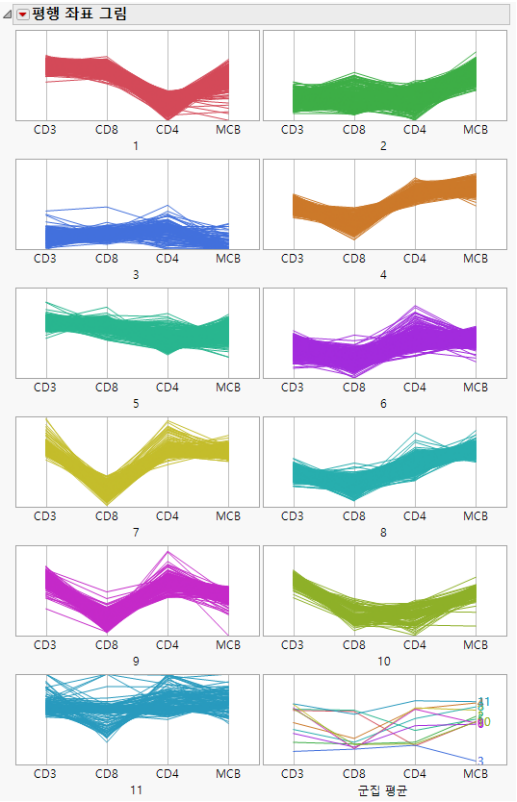
군집 평균

군집	CD3	CD8	CD4	MCB
1	314.073529	300.270833	106.615196	183.4375
2	140.091286	116.365145	127.024896	205.014523
3	90.7898089	87.5477707	109.356688	15.0191083
4	247.426343	148.064125	314.074523	261.831889
5	320.287383	307.372664	193.117991	193.174065
6	188.686747	99.0863454	220.062249	171.682731
7	347.496644	89.9574944	320.782998	231.557047
8	210.258652	126.125683	260.327869	245.588342
9	328.206897	89.7824934	312.909814	175.965517
10	322.105691	113.715447	116.666667	181.731707
11	349.864407	284.813559	360.932203	267.474576

" 군집 요약 " 보고서에는 11 개 군집 각각의 관측값 수가 표시됩니다 . " 군집 평균 " 보고서에는 각 군집에 대한 표지 판독값의 평균이 표시됩니다 .

9. "K 평균 군집 수 =11" 의 빨간색 삼각형을 클릭하고 **평행 좌표 그림**을 선택합니다 .

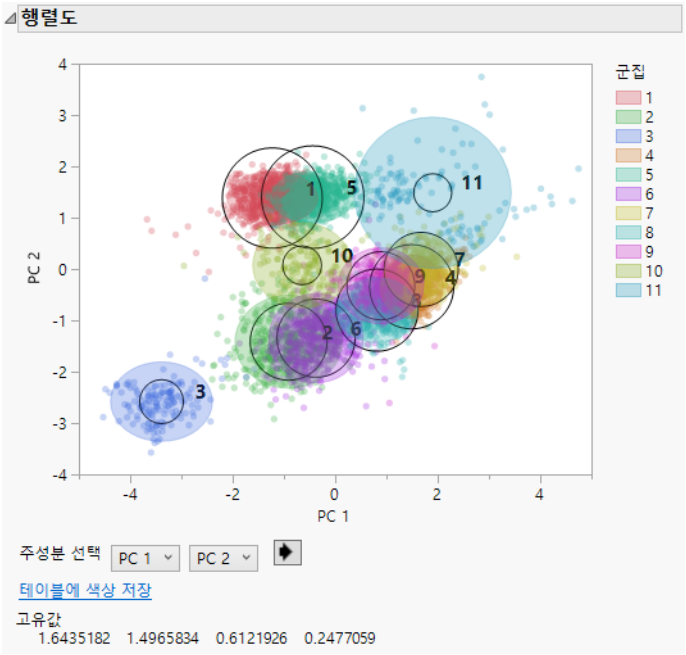
그림 13.4 Cytometry 데이터에 대한 평행 좌표 그림



" 평행 좌표 그림 " 에는 각 군집의 관측값 구조가 표시됩니다 . 이러한 그림을 사용하여 군집이 어떻게 다른지 확인할 수 있습니다 . 군집 4, 6, 7, 8, 9 는 비교적 CD8 값이 낮고 CD4 값이 높은 경향이 있습니다 . 반면에 군집 1 은 CD8 값이 더 높고 CD4 값이 더 낮습니다 .

10. "K 평균 군집 수 =11" 의 빨간색 삼각형을 클릭하고 **행렬도**를 선택합니다 .

그림 13.5 Cytometry 데이터에 대한 행렬도

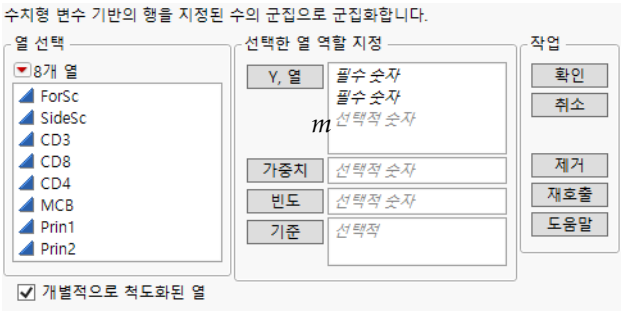


군집 색상을 식별하는 범례가 그림 오른쪽에 표시됩니다 . 처음 두 주성분을 기반으로 다른 군집과 가장 많이 떨어져 있는 군집은 3, 10 및 11 군집입니다 . 이는 다른 군집에 대한 그림과 다른 그림 13.4 의 평행 좌표 그림에서도 입증됩니다 . 그림 아래의 목록을 사용하여 다른 주성분 조합에 대한 행렬도를 확인할 수 있습니다 .

K 평균 군집화 플랫폼 시작

분석 > 군집화 > K 평균 군집화를 선택하여 K 평균 군집화 플랫폼을 시작합니다 .

그림 13.6 K 평균 군집화 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 열 관측값을 군집화하는 데 사용되는 변수입니다.

참고 : *K* 평균 군집화는 숫자 열만 지원합니다.

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다.

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다.

기준 각 수준으로 개별 분석을 정의하는 열입니다. 지정된 열의 각 수준에 대해 해당하는 행이 분석됩니다. 결과는 개별 보고서에 제공됩니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다.

시작 창 옵션

개별적으로 척도화된 열 각 열을 다른 열과 독립적으로 척도화합니다. 변수가 공통 측정 척도를 공유하지 않고 한 변수가 군집화 과정을 지배하지 않게 하려는 경우 이 옵션을 사용합니다. 예를 들어 한 변수의 값이 0 에서 1000 사이이고 다른 변수의 값이 0 에서 10 사이일 수 있습니다. 이 경우 군집화 과정이 첫 번째 변수에 의해 지배되지 않도록 이 옵션을 사용할 수 있습니다.

"확인"을 클릭하면 "제어판"이 나타납니다. 자세한 내용은 "반복적 군집화 제어판"에서 확인하십시오.

반복적 군집화 보고서

시작 창에서 "확인"을 클릭하면 "반복적 군집화" 보고서 창이 나타나고 모형 적합을 위한 "제어판"이 표시됩니다. 자세한 내용은 "반복적 군집화 제어판"에서 확인하십시오. 모형을 적합시키면 해당 보고서가 창에 추가됩니다. 자세한 내용은 "K 평균 보고서"에서 확인하십시오.

반복적 군집화 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

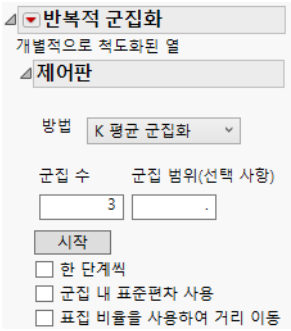
스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

반복적 군집화 제어판

Cytometry.jmp 데이터 테이블에 대한 " 제어판 " 이 그림 13.7 에 나와 있습니다 . 다른 수의 군집을 반복적으로 적합시키거나 , " 군집 범위 " 옵션을 사용하여 범위를 지정할 수 있습니다 .

그림 13.7 반복적 군집화 제어판



" 제어판 " 에는 다음 옵션이 있습니다 .

방법 다음과 같은 군집화 방법을 사용할 수 있습니다 .

K 평균 군집화 이 장에서 설명합니다 .

자기 조직화 지도 " 자기 조직화 지도 제어판 " 에서 설명합니다 .

군집 수 형성할 군집 수를 지정합니다 .

군집 범위 (선택 사항) 형성할 군집 수에 대한 상한을 제공합니다 . 여기에 숫자를 입력하면 플랫폼에서 " 군집 수 " 와 " 군집 범위 (선택 사항)" 로 입력한 값 사이의 모든 정수에 대해 개별 분석을 생성합니다 .

시작 군집을 자동으로 적합시킵니다 (" 한 단계씩 " 을 선택한 경우 제외).

한 단계씩 군집화 과정을 한 번에 한 단계씩 반복할 수 있습니다 . " 한 단계씩 " 을 선택하고 " 시작 " 을 클릭하면 군집 할당이 없지만 " 시작 " 및 " 단계 " 버튼이 포함된 "K 평균 군집화 " 보고서가 나타납니다 .

- 한 번에 한 단계씩 반복하려면 " 단계 " 버튼을 클릭합니다 .
- 군집을 자동으로 적합시키려면 " 시작 " 버튼을 클릭합니다 .

군집 내 표준편차 사용 각 군집 내 관측값에 대해 각 변수의 표준편차 추정값을 사용하여 거리를 척도화합니다 . 이 옵션을 선택하지 않으면 각 변수의 전체 표준편차 추정값을 사용하여 거리를 척도화합니다 .

표집 비율을 사용하여 거리 이동 군집 크기에 따라 거리를 조정합니다 . 군집 크기가 같지 않은 경우 관측값은 더 큰 군집에서 나올 사전 확률이 더 높으므로 더 큰 군집에 관측값이 할당될 확률이 더 높아야 합니다 .

K 평균 보고서

" 제어판 " 에서 " 시작 " 을 클릭하면 다음 보고서가 나타납니다 .

- 군집 비교 보고서 . 자세한 내용은 " 군집 비교 보고서 " 에서 확인하십시오 .
- 하나 이상의 K 평균 보고서 . 보고서 이름은 *k*(적합된 군집 수) 에 따라 동적으로 "K 평균 군집 수 =<k>" 로 지정됩니다 . 수행한 각 적합에 대한 "K 평균 군집 수 =<k>" 보고서가 나타납니다 .

CD3 ~ MCB 변수를 "Y, 열"로 사용한 Cytometry.jmp 데이터 테이블의 "군집 비교" 보고서와 "K 평균 군집 수=<11>" 보고서가 그림 13.2 및 "K 평균 군집 수=11 보고서"에 나와 있습니다.

군집 비교 보고서

"군집 비교" 보고서에서는 다양한 모형을 비교하기 위한 적합 통계량을 제공합니다. 적합 통계량은 CCC(Cubic Clustering Criterion)입니다. CCC 값이 클수록 더 나은 적합을 나타냅니다. 최량 적합은 "최선" 열에 "최적 CCC"를 지정하여 표시됩니다. 자세한 내용은 SAS Institute Inc. (1983)에서 확인하십시오. 상수 열은 CCC 계산에 포함되지 않습니다.

K 평균 보고서

각 K 평균 보고서에서는 각 군집에 대해 다음과 같은 요약 통계량을 제공합니다 .

- " 군집 요약 " 보고서에는 군집 수와 각 군집의 관측값 수 및 필요한 반복 수가 제공됩니다 .
- " 군집 평균 " 보고서에는 각 군집에서 각 변수에 대한 관측값의 평균이 제공됩니다 .
- " 군집 표준편차 " 보고서에는 각 군집에서 각 변수에 대한 관측값의 표준편차가 제공됩니다 .

K 평균 보고서 옵션

각 K 평균 보고서에는 다음 옵션이 포함됩니다 .

행렬도 데이터의 처음 두 주성분으로 표현된 점 및 군집의 그림을 군집 색상을 식별하는 범례와 함께 표시합니다 . 군집 중심을 기준으로 원이 그려지고 원 크기는 군집 내의 개수에 비례합니다 . 음영 영역은 평균 주위의 밀도 등고선입니다 . 기본적으로 이 영역은 해당 군집의 관측값 중 90% 가 속하는 위치를 나타냅니다 (Mardia et al. 1980 참조). 그림 아래의 목록을 사용하여 그림 축을 다른 주성분으로 변경할 수 있습니다 . 또는 화살표 버튼을 사용하여 가능한 모든 축 조합을 순환할 수 있습니다 . 군집 색상을 데이터 테이블에 저장하는 옵션도 그림 아래에 있습니다 . 자세한 내용은 " 테이블에 색상 저장 " 에서 확인하십시오 . 고유값은 내림차순으로 표시됩니다 .

참고 : 시작 창에서 " 개별적으로 척도화된 열 " 을 선택한 경우 행렬도에 상관 행렬이 사용됩니다 . " 개별적으로 척도화된 열 " 을 선택하지 않은 경우에는 행렬도에 공분산 행렬이 사용됩니다 .

행렬도 옵션 행렬도 모양을 제어하기 위한 다음 옵션을 포함합니다 .

행렬도 선 표시 행렬도 선을 표시합니다 . 라벨이 지정된 선은 주성분으로 정의된 부분 공간에서 공변량의 방향을 표시합니다 . 이러한 선은 각 변수와 각 주성분의 연관성 정도를 나타냅니다 .

행렬도 선 위치 행렬도 선의 위치와 반지름 척도를 지정할 수 있습니다 . 기본적으로 선은 점 (0,0) 에서 시작됩니다 . 그림에서 선을 드래그하거나 이 옵션을 사용하여 좌표를 지정할 수 있습니다 . 반지름 척도 옵션을 사용하여 선이 더 잘 보이게 선 척도를 조정할 수도 있습니다 .

행렬도 등고선 밀도 밀도 등고선에 대한 신뢰 수준을 지정할 수 있습니다 . 기본 신뢰 수준은 90% 입니다 .

군집 표시 데이터 테이블의 행에 군집을 식별하는 표식을 할당합니다 .

행렬도 3D 데이터의 3 차원 행렬도를 표시합니다 . 변수가 세 개 이상인 경우에만 사용할 수 있습니다 .

평행 좌표 그림 각 군집에 대한 평행 좌표 그림을 생성합니다 . 이 그림 보고서에는 데이터와 평균을 표시하거나 숨기는 옵션이 있습니다 . 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .

산점도 행렬 모든 Y 변수를 사용하여 산점도 행렬을 생성합니다 .

테이블에 색상 저장 데이터 테이블의 행에 군집을 식별하는 색상을 할당합니다 . 보고서 창에 " 행렬도 " 가 있는 경우 데이터 테이블에 저장된 색상은 행렬도의 군집 색상과 일치합니다 . " 행렬도 " 에서 색상을 변경하고 " 테이블에 색상 저장 " 옵션을 다시 선택하면 테이블의 색상이 행렬도의 색상과 일치하도록 업데이트됩니다 .

참고 : 저장 옵션을 선택하면 저장된 각 열에 해당 특정 열의 데이터에 대한 군집 수를 지정하는 " 노트 " 열 특성이 포함됩니다 . 이렇게 하면 둘 이상의 군집 적합에서 열을 저장하고 열 특성을 사용하여 저장된 열을 가져온 군집화 적합을 식별할 수 있습니다 .

군집 저장 다음 두 열을 데이터 테이블에 저장합니다 .

- 군집 열은 주어진 행이 할당된 군집의 번호를 포함합니다 .
- (자기 조직화 지도에는 사용 불가능) 거리 열은 주어진 관측값과 군집 평균 사이의 유클리드 거리 제곱을 포함합니다 . 각 변수에 대해 관측값과 해당 변수의 군집 평균 간의 차이를 변수의 전체 표준편차로 나눕니다 . 변수 전체에 대해 이러한 척도화된 차이의 제곱합이 계산됩니다 .

군집 거리 저장 (자기 조직화 지도에는 사용 불가능) 거리 열을 데이터 테이블에 저장합니다 . 이 열은 **군집 저장** 옵션에서 얻은 거리 열과 동일합니다 .

군집 계산식 저장 " 군집 계산식 " 이라는 계산식 열을 데이터 테이블에 저장합니다 . 이 계산식은 각 소속 군집을 식별합니다 .

거리 계산식 저장 (자기 조직화 지도에는 사용 불가능) " 거리 계산식 " 이라는 계산식 열을 데이터 테이블에 저장합니다 . 이 계산식은 할당된 군집까지의 거리를 계산합니다 .

K 군집 거리 저장 (자기 조직화 지도에는 사용 불가능) 각 군집 중심까지의 유클리드 거리 제곱을 포함하는 *k* 개 열을 저장합니다 .

K 거리 계산식 저장 (자기 조직화 지도에는 사용 불가능) 각 군집 중심까지의 유클리드 거리 제곱 계산식을 포함하는 *k* 개 열을 저장합니다 .

군집 계산식 게시 " 군집 계산식 저장 " 옵션에 사용된 것과 동일한 스코어링 코드를 " 계산식 저장소 " 에 게시합니다 .

군집 시뮬레이션 군집 평균 및 표준편차를 사용하여 Y 변수에 대한 시뮬레이션된 군집 관측값을 포함하는 새 데이터 테이블을 생성합니다 .

제거 군집화 보고서를 제거합니다 .

자기 조직화 지도

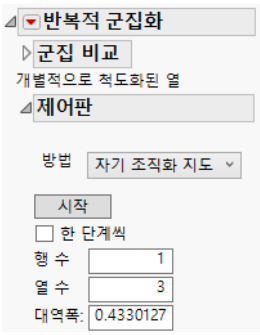
SOM(자기 조직화 지도) 기법은 Teuvo Kohonen(1989, 1990) 에 의해 개발되었으며 다른 신경망 지지자 및 통계학자에 의해 확장되었습니다. 원래 SOM 은 원래의 신경망 알고리즘처럼 학습 프로세스였지만 여기서 구현된 버전은 *k* 평균 군집화의 변형입니다. SOM 문헌에서는 이 변형을 국소 가중 선형 평활기를 사용한 배치 알고리즘이라고 합니다.

SOM 의 목표는 군집 격자의 특정 레이아웃에 군집을 형성하는 것뿐만 아니라 SOM 격자에서 서로 가까운 군집의 점이 다변량 공간에서도 서로 가까이 있도록 하는 것입니다. 전통적 *k* 평균 군집화에서는 군집 구조가 임의적이지만 SOM 의 군집은 격자 구조를 가집니다. 격자 구조는 군집을 2 차원으로 해석하는 데 도움이 됩니다. 즉, 서로 가까운 군집이 멀리 있는 군집보다 더 유사합니다. 자세한 내용은 "SOM 알고리즘 설명" 에서 확인하십시오.

자기 조직화 지도 제어판

반복적 군집화 제어판의 " 방법 " 목록에서 " 자기 조직화 지도 " 옵션을 선택합니다 .

그림 13.8 자기 조직화 지도 제어판



제어판의 일부 옵션은 " 반복적 군집화 제어판 " 에 설명되어 있습니다 . 나머지 옵션은 아래에서 설명합니다 .

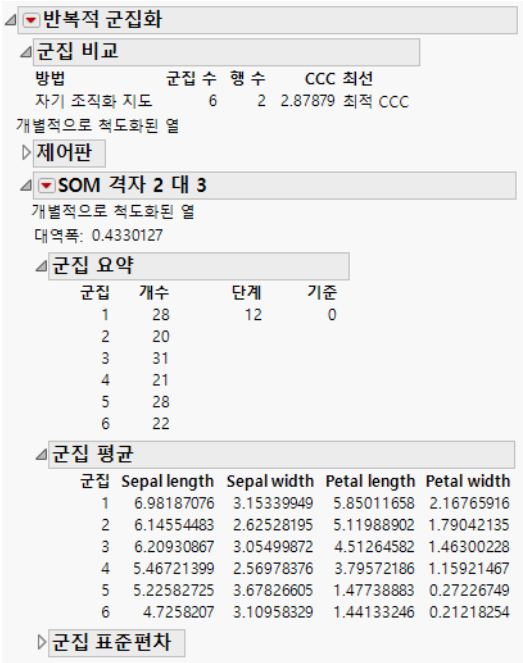
행 수 군집 격자의 행 수입니다 .

열 수 군집 격자의 열 수입니다 .

대역폭 중심 예측을 위한 인접 군집의 효과를 지정합니다 . 대역폭이 작을수록 더 가까운 군집에 가중치가 더 많이 지정됩니다 .

자기 조직화 지도 보고서

그림 13.9 자기 조직화 지도 보고서



반복적 군집화

군집 비교

방법	군집 수	행 수	CCC	최선
자기 조직화 지도	6	2	2.87879	최적 CCC

개별적으로 척도화된 열

제어판

SOM 격자 2 대 3

개별적으로 척도화된 열

대역폭: 0.4330127

군집 요약

군집	개수	단계	기준
1	28	12	0
2	20		
3	31		
4	21		
5	28		
6	22		

군집 평균

군집	Sepal length	Sepal width	Petal length	Petal width
1	6.98187076	3.15339949	5.85011658	2.16765916
2	6.14554483	2.62528195	5.11988902	1.79042135
3	6.20930867	3.05499872	4.51264582	1.46300228
4	5.46721399	2.56978376	3.79572186	1.15921467
5	5.22582725	3.67826605	1.47738883	0.27226749
6	4.7258207	3.10958329	1.44133246	0.21218254

군집 표준편차

" 군집 비교 " 보고서에는 총 군집 수와 요청된 행 수가 표시됩니다 . SOM 보고서는 요청된 격자 크기에 따라 이름이 지정됩니다 . " 대역폭 " 은 "SOM 격자 " 보고서 맨 위에 제공됩니다 . 보고서 자체는 "K 평균 군집 수 " 보고서와 유사합니다 . 자세한 내용은 "K 평균 보고서 " 에서 확인하십시오 .

"자기 조직화 지도"의 빨간색 삼각형 옵션에 대한 자세한 내용은 "K 평균 보고서 옵션"에서 확인하십시오.

SOM 알고리즘 설명

이 섹션에는 JMP 의 SOM 구현 단계가 포함되어 있습니다 .

- 다차원 공간에 대한 적절한 범위를 제공하는 방식으로 초기 군집 시드가 선택됩니다 . JMP 에서 주성분을 사용하여 데이터의 가장 큰 변동을 포착하는 두 방향을 결정합니다 .
- 그런 다음 JMP 가 각 방향의 중간에서 2.5 표준편차를 갖는 이 주성분 공간에 격자를 배치합니다 . 이 격자를 변수의 원래 공간으로 다시 변환하여 군집 시드가 결정됩니다 .
- *k* 평균과 마찬가지로 군집 할당이 진행됩니다 . 각 점이 가장 가까운 위치의 군집에 할당됩니다 .
- *k* 평균과 마찬가지로 각 군집에 대해 평균이 추정됩니다 . 그런 다음 JMP 에서 이러한 평균을 사용하여 가중 회귀를 설정합니다 . 이때 각 변수를 회귀의 반응으로 지정하고 SOM 격자 좌표를 회귀변수로 지정합니다 . 가중치 함수는 중심이 추정되는 군집에 큰 가중치를 제공하는 커널 함수를 사용합니다 . SOM 격자의 군집에서 더 멀리 있는 군집에 더 작은 가중치가 제공됩니다 . 새 군집 평균은 이 회귀 분석의 예측값입니다 .
- 과정이 수렴될 때까지 이러한 반복이 계속됩니다 .

K 평균 군집화 플랫폼의 추가 예

자기 조직화 지도의 예

이 예에서는 세 가지 종류의 붓꽃에 대한 꽃받침 길이 (Sepal length), 꽃받침 폭 (Sepal width), 꽃잎 길이 (Petal length) 및 꽃잎 폭 (Petal width) 측정값이 포함된 Iris.jmp 샘플 데이터 테이블을 사용합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다 .
2. **분석 > 군집화 > K 평균 군집화**를 선택합니다 .
3. Sepal length, Sepal width, Petal length 및 Petal width 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 제어판 " 의 " 방법 " 메뉴에서 **자기 조직화 지도**를 선택합니다 .
6. **행 수**를 "1" 로 설정하고 **열 수**를 "2" 로 설정합니다 .
7. **시작**을 클릭합니다 .
8. " 제어판 " 보고서를 엽니다 .
9. **행 수**를 "1" 로 설정하고 **열 수**를 "3" 으로 설정합니다 .
10. **시작**을 클릭합니다 .
11. " 제어판 " 보고서를 엽니다 .
12. **행 수**를 "2" 로 설정하고 **열 수**를 "2" 로 설정합니다 .

13. **시작**을 클릭합니다 .

그림 13.10 SOM 군집 비교

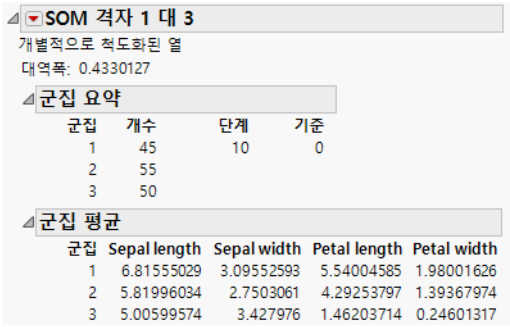
군집 비교

방법	군집 수	행 수	CCC 최선
자기 조직화 지도	2	1	3.35952
자기 조직화 지도	3	1	4.95837 최적 CCC
자기 조직화 지도	4	2	3.64938

" 군집 비교 " 보고서가 보고서 창 맨 위에 나타납니다 . 최량 적합은 가장 큰 CCC 값에 의해 결정됩니다 . 가장 큰 CCC 를 제공하는 군집 수는 종의 수와 같은 3 입니다 .

14. "SOM 격자 1 X 3" 보고서가 나올 때까지 스크롤합니다 . 분류가 완벽하지 않음을 알 수 있습니다 . 각 군집은 각각 50 개의 행이 있는 각 종을 나타내야 합니다 .

그림 13.11 Iris.jmp 에 대한 SOM 보고서



SOM 격자 1 대 3

개별적으로 척도화된 열

대역폭: 0.4330127

군집 요약

군집	개수	단계	기준
1	45	10	0
2	55		
3	50		

군집 평균

군집	Sepal length	Sepal width	Petal length	Petal width
1	6.81555029	3.09552593	5.54004585	1.98001626
2	5.81996034	2.7503061	4.29253797	1.39367974
3	5.00599574	3.427976	1.46203714	0.24601317

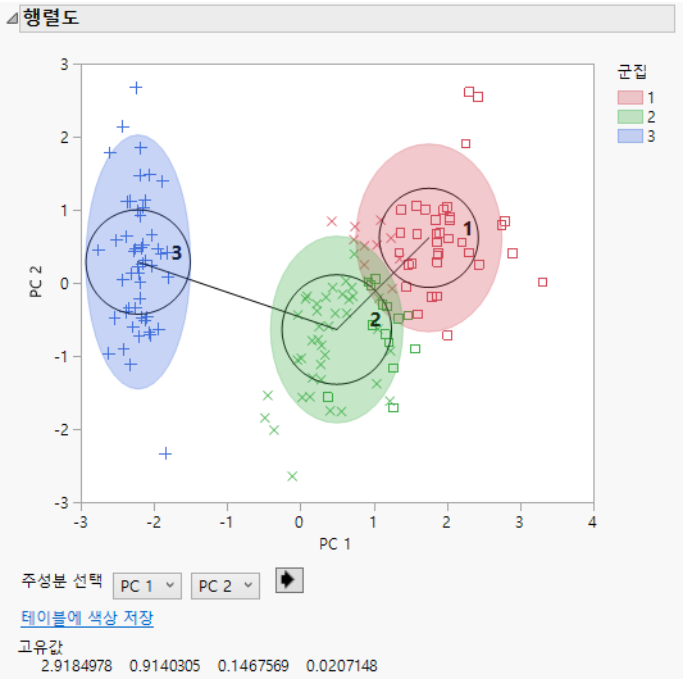
15. 데이터 테이블에서 Species 열을 선택하고 **행 > 열 값에 따른 색상 또는 표식**을 선택합니다 .

16. " 표식 " 에서 **클래식** 옵션을 선택합니다 .

17. **확인**을 클릭합니다 .

18. "SOM 격자 1 X 3" 옆의 빨간색 삼각형 메뉴를 클릭하고 **행렬도**를 선택합니다 .

그림 13.12 SOM 행렬도



군집 3 의 모든 행이 setosa 종으로 올바르게 식별되었음을 확인할 수 있습니다 . 다른 두 종 , virginica 와 versicolor 는 약간 중첩되어 서로 잘못 볼 수 있습니다 .

14 장

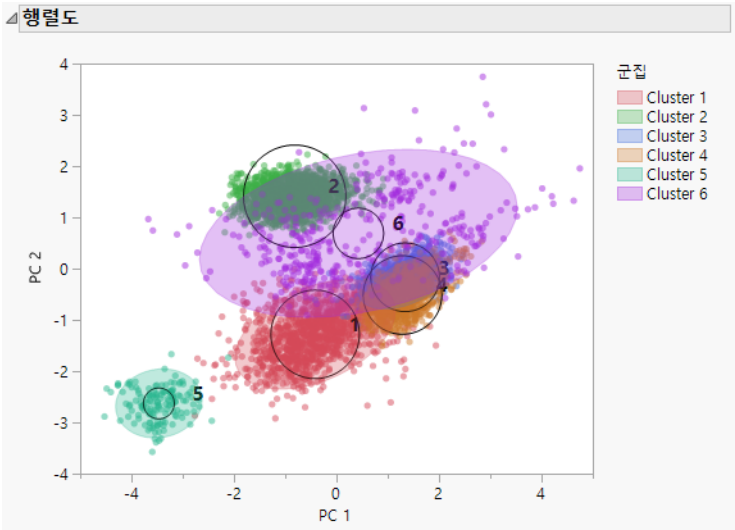
정규 혼합
확률을 사용하여 관측값 그룹화

중첩된 정규 분포의 데이터를 가져오는 경우 정규 혼합을 사용하여 군집화합니다. 군집 수를 미리 지정해야 합니다.

정규 혼합은 다변량 정규 분포의 혼합을 사용하여 관측값의 결합 확률 분포가 근사화된다는 가정을 기반으로 하는 반복 기법입니다. 이러한 혼합은 서로 다른 군집을 나타냅니다. 개별 군집은 다변량 정규 분포를 따릅니다.

군집이 잘 분리되면 계층적 군집화 및 k 평균 군집화가 제대로 작동합니다. 그러나 군집이 중첩된 경우에는 정규 혼합이 더 나은 방법입니다. 이 방법은 경계에 기반한 임의의 군집 할당 대신 소속 군집 확률을 기반으로 하기 때문입니다.

그림 14.1 정규 혼합 행렬도



목차

정규 혼합 군집화 플랫폼 개요 289
관측값 군집화 플랫폼 개요 289
정규 혼합 군집화의 예 290
정규 혼합 군집화 플랫폼 시작 292
모형 기반 군집화 보고서 293
모형 기반 군집화 옵션 293
모형 기반 군집화 제어판 294
정규 혼합 보고서 295
군집 비교 보고서 295
정규 혼합 보고서 295
정규 혼합 보고서 옵션 295
정규 혼합 군집화 플랫폼에 대한 통계 상세 정보 297

정규 혼합 군집화 플랫폼 개요

정규 혼합은 JMP에서 관측값을 군집화하기 위해 제공하는 네 가지 플랫폼 중 하나입니다. 네 가지 방법을 모두 비교하려면 "관측값 군집화 플랫폼 개요"의 내용을 참조하십시오.

정규 혼합은 수치 변수를 위한 반복 군집화 기법입니다. 그러나 각 군집 내에서 기대 반응의 비율도 예측합니다. 정규 혼합은 서로 다른 군집을 나타내는 다변량 정규 분포의 혼합을 사용하여 측정값 열의 결합 확률 분포를 근사화할 수 있다고 가정합니다. 각 군집에 대해 평균 벡터와 공분산 행렬이 추정됩니다. 자세한 내용은 McLachlan과 Krishnan(1997) 연구 자료 및 Hand et al. (2001)의 Section 9.6에서 확인하십시오.

참고 : 정규 혼합 알고리즘에는 군집 중심에 대한 랜덤 추측을 반복하는 작업이 포함됩니다. 이 때문에 분석을 실행할 때마다 결과가 약간 다를 수 있습니다.

다변량 이상치가 있다고 의심되는 경우 두 가지 옵션이 있습니다. 이상치 군집 또는 이상치 탐색 유틸리티를 사용할 수 있습니다. 이상치 군집 옵션은 균등 분포를 가정하고 표준 정규 혼합 방법보다 이상치에 덜 민감합니다. 이상치 탐색 유틸리티를 사용하면 분석 전에 이상치를 탐색하고 처리할 수 있습니다. 자세한 내용은 "이상치 군집" 및 **예측 및 전문 모델링**의 에서 확인하십시오.

관측값 군집화 플랫폼 개요

군집화는 여러 변수에 대해 유사한 값을 공유하는 관측값을 함께 그룹화하는 다변량 기법입니다. 일반적으로 관측값은 p 차원 공간에 고르게 분산되지 않습니다. 여기서 p는 변수 수입니다. 대신 관측값은 군집을 형성합니다. 이러한 군집을 식별하면 데이터를 더 세밀하게 파악할 수 있습니다.

참고 : JMP 에서는 변수를 군집화할 수 있는 플랫폼도 제공합니다. 자세한 내용은 " 변수 군집화 " 에서 확인하십시오 .

JMP 에서 제공하는 다음 네 가지 플랫폼을 사용하여 관측값을 군집화할 수 있습니다 .

- 계층적 군집화는 최대 수만 개의 행이 있는 작은 테이블에 유용하며 문자 데이터를 허용합니다 . 계층적 군집화는 행을 계층적 순서로 결합하여 트리로 표현합니다 . 트리가 생성되면 데이터에 가장 적절한 군집 수를 선택할 수 있습니다 .
- K 평균 군집화는 최대 수백만 개의 행이 있는 큰 테이블에 적절하며 숫자 데이터만 허용합니다 . 군집 수 k 를 미리 지정해야 합니다 . 이 알고리즘은 군집 시드 점을 추측합니다 . 그런 다음 군집에 점을 번갈아 할당하고 군집 중심을 다시 계산하는 반복 과정을 수행합니다 .
- 정규 혼합은 중첩될 수 있는 다변량 정규 분포의 혼합에서 데이터를 수집한 경우에 적절하며 숫자 데이터만 허용합니다. 다변량 이상치가 있는 경우 균등 분포를 가정하고 이상치 군집을 사용할 수 있습니다 .

군집 수를 미리 지정해야 합니다. 최대 가능도는 혼합 비율과 평균, 표준편차 및 상관을 공동으로 추정하는 데 사용됩니다. 각 그룹에 속할 확률이 각 점에 할당됩니다. EM 알고리즘은 추정값을 구하는 데 사용됩니다.

- 잠재 계층 분석은 대부분의 변수가 범주형일 때 적절합니다. 군집 수를 미리 지정해야 합니다. 이 알고리즘은 다항 혼합 분포를 가정하는 모형을 적합시킵니다. 각 관측값에 대해 소속 군집의 최대 가능도 추정값이 계산됩니다. 관측값은 소속 확률이 가장 큰 군집으로 분류됩니다.

표 14.1 군집화 방법 요약

방법	데이터 유형 또는 모델링 유형	데이터 테이블 크기	군집 수 지정
계층적 군집화	임의	빠른 Ward 를 사용하는 경우 최대 200,000 개 행 다른 방법을 사용하는 경우 최대 5,000 개 행	아니요
K 평균 군집화	수치형	최대 수백만 개 행	예
정규 혼합	수치형	모든 크기	예
잠재 계층 분석	명목형 또는 순서형	모든 크기	예

일부 군집화 플랫폼에는 데이터의 이상치를 처리하는 옵션이 있습니다. 그러나 데이터에 이상치가 있으면 분석하기 전에 먼저 이상치를 탐색하는 것이 가장 좋습니다. 이 작업은 이상치 탐색 유틸리티를 사용하여 수행할 수 있습니다. 자세한 내용은 **예측 및 전문 모델링**의"모델링 유틸리티" 장에서 확인하십시오.

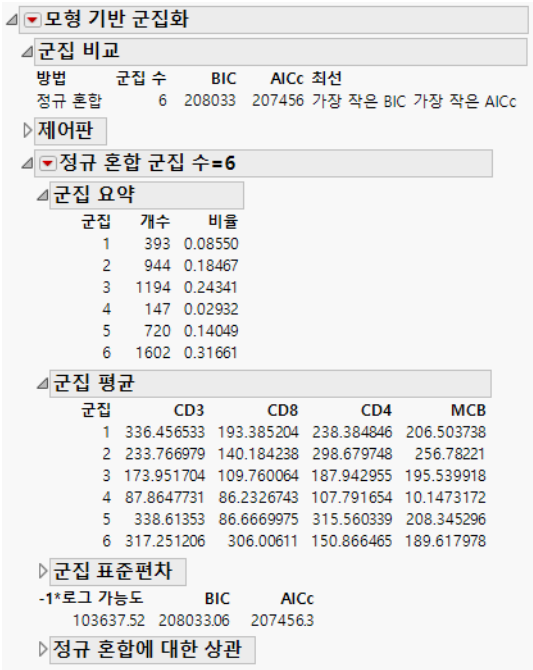
정규 혼합 군집화의 예

세포분석법은 세포의 다양한 특성을 측정하는 데 사용됩니다. 세포 표지를 측정하면 특정 질병을 진단하는 데 도움이 됩니다. 이 예에서는 세포분석에서 감지된 네 개의 표지 판독값을 기반으로 관측값을 군집화합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cytometry.jmp 를 엽니다.
2. **분석 > 군집화 > 정규 혼합**을 선택합니다.
3. CD3, CD8, CD4 및 MCB 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. **군집 수** 옆에 "6" 을 입력합니다.
6. **시작**을 클릭합니다.

참고 : 알고리즘에 랜덤 시작 값이 있으므로 결과가 다를 수 있습니다 .

그림 14.2 정규 혼합 군집 수 =6 보고서



모형 기반 군집화

군집 비교

방법	군집 수	BIC	AICc	최선
정규 혼합	6	208033	207456	가장 작은 BIC 가장 작은 AICc

제어판

정규 혼합 군집 수=6

군집 요약

군집	개수	비율
1	393	0.08550
2	944	0.18467
3	1194	0.24341
4	147	0.02932
5	720	0.14049
6	1602	0.31661

군집 평균

군집	CD3	CD8	CD4	MCB
1	336.456533	193.385204	238.384846	206.503738
2	233.766979	140.184238	298.679748	256.78221
3	173.951704	109.760064	187.942955	195.539918
4	87.8647731	86.2326743	107.791654	10.1473172
5	338.61353	86.6669975	315.560339	208.345296
6	317.251206	306.00611	150.866465	189.617978

군집 표준편차

-1*로그 가능도	BIC	AICc
103637.52	208033.06	207456.3

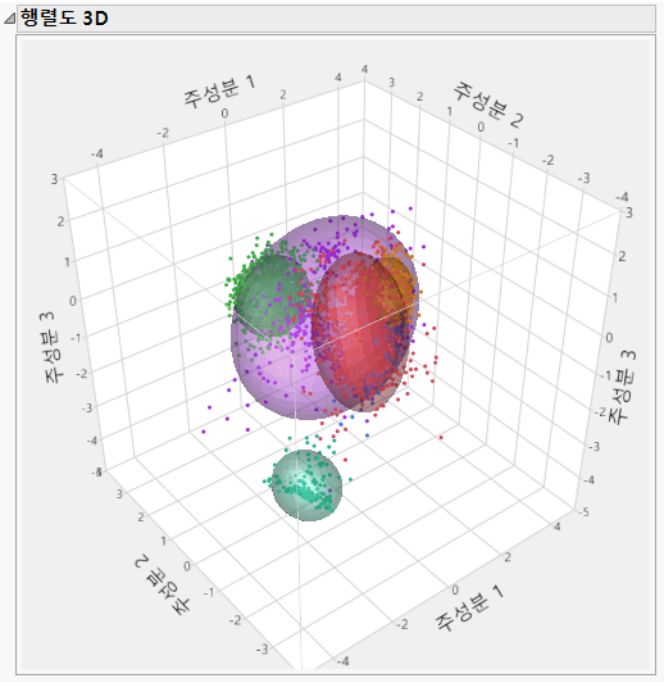
정규 혼합에 대한 상관

" 군집 요약 " 보고서에는 6 개 군집 각각의 관측값 수가 표시됩니다 . " 군집 평균 " 보고서에는 각 군집에 대한 표지 판독값의 평균이 표시됩니다 .

7. " 정규 혼합 군집 수 =6" 옆의 빨간색 삼각형을 클릭하고 **행렬도 3D** 를 선택합니다 .

참고 : 알고리즘에 랜덤 시작 값이 있으므로 3D 행렬도가 다르게 나타날 수 있습니다 .

그림 14.3 Cytometry 데이터에 대한 3D 행렬도

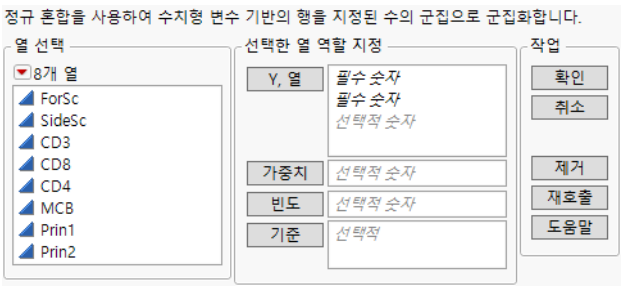


이 그림에는 군집에 적합된 정규 밀도의 등고선이 표시됩니다. 한 군집이 처음 세 개의 주성분을 기반으로 다른 군집과 뚜렷하게 분리된 것 같습니다.

정규 혼합 군집화 플랫폼 시작

분석 > 군집화 > 정규 혼합을 선택하여 정규 혼합 군집화 플랫폼을 시작합니다.

그림 14.4 정규 혼합 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y, 열 관측값을 군집화하는 데 사용되는 변수입니다 .

참고 : 정규 혼합 군집화는 숫자 열만 지원합니다 .

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다 .

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다 .

기준 각 수준으로 개별 분석을 정의하는 열입니다 . 지정된 열의 각 수준에 대해 해당하는 행이 분석됩니다 . 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

"확인"을 클릭하면 "제어판"이 나타납니다. 자세한 내용은 "모형 기반 군집화 제어판"에서 확인하십시오.

모형 기반 군집화 보고서

시작 창에서 " 확인 " 을 클릭하면 " 모형 기반 군집화 " 보고서 창이 열리고 모형 적합을 위한 " 제어판 " 이 표시됩니다 . 자세한 내용은 " 모형 기반 군집화 제어판 " 에서 확인하십시오 . 모형을 적합시키면 해당 보고서가 창에 추가됩니다 . 자세한 내용은 " 정규 혼합 보고서 " 에서 확인하십시오 .

모형 기반 군집화 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

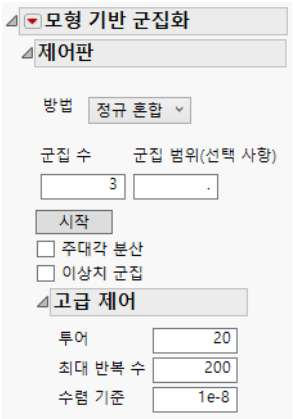
다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

모형 기반 군집화 제어판

CD3 ~ MCB 변수를 **Y, 열**로 사용한 Cytometry.jmp 데이터 테이블의 " 제어판 " 이 그림 14.5 에 나와 있습니다 . " 제어판 " 을 반복적으로 사용하여 다양한 수의 군집을 적합시키거나 , " 군집 범위 " 옵션을 사용하여 범위를 지정할 수 있습니다 .

그림 14.5 정규 혼합 방법에 대한 제어판



모형 기반 군집화 제어판에는 다음과 같은 옵션이 있습니다 .

군집 수 형성할 군집 수를 지정합니다 .

군집 범위 (선택 사항) 형성할 군집 수에 대한 상한을 제공합니다 . 여기에 숫자를 입력하면 플랫폼에서 **군집 수**와 **군집 범위 (선택 사항)** 로 입력한 값 사이의 모든 정수에 대해 개별 분석을 생성합니다 .

시작 군집을 적합시킵니다 .

대각 분산 공분산 행렬의 비대각 요소를 0 으로 제한합니다 . 플랫폼에서는 변수 간에 상관관계가 없는 다변량 정규 분포를 적합시킵니다 .

참고 : 변수보다 관측값 수가 적은 경우 특이 공분산 행렬이 생성되는 것을 방지하기 위해 " 대각 분산 " 옵션이 필요할 때가 있습니다 . 또한 많은 수의 변수에 대해 매우 큰 공분산 행렬을 추정하는 것을 방지하는 데 사용할 수도 있습니다 .

이상치 군집 정규 군집에 속하지 않는 이상치를 포착하기 위해 군집을 적합시킵니다 . 이 군집이 생성되면 군집 0 으로 지정되고 관측값 수가 " 군집 요약 " 보고서에 나타납니다 . 이상치 군집에 속하는 관측값의 분포는 관측값을 포함하는 초입방체에 대해 균등하다고 가정합니다 .

고급 컨트롤 다음과 같은 고급 컨트롤을 사용할 수 있습니다 .

투어 추정 프로세스의 독립적 재시작 횟수입니다 . 다시 시작할 때마다 시작 값이 다릅니다 . 독립적으로 시작하면 국소 해를 찾지 않도록 할 수 있습니다 .

최대 반복 수 EM 알고리즘의 수렴 단계에 대한 최대 반복 수입니다 .

수렴 기준 EM 반복이 종료되는 가능도 차이입니다 .

정규 혼합 보고서

" 제어판 " 에서 " 시작 " 을 클릭하면 다음 보고서가 나타납니다 .

- 군집 비교 보고서 . 자세한 내용은 " 군집 비교 보고서 " 에서 확인하십시오 .
- 하나 이상의 정규 혼합 보고서 . 보고서 이름은 *k*(적합된 군집 수) 에 따라 동적으로 " 정규 혼합 군집 수 =<k>" 로 지정됩니다 . 수행한 각 적합에 대한 " 정규 혼합 군집 수 =<k>" 보고서가 나타납니다 .

CD3 ~ MCB 변수를 **Y, 열**로 사용한 Cytometry.jmp 데이터 테이블의 "군집 비교" 보고서와 "정규 혼합 군집 수=<6>" 보고서가 그림 14.2 및 "정규 혼합 군집 수=6 보고서"에 나와 있습니다.

군집 비교 보고서

"군집 비교" 보고서에서는 다양한 모형을 비교하기 위한 적합 통계량을 제공합니다. 적합 통계량은 BIC와 AICc입니다. 각 값이 작을수록 더 나은 적합을 나타냅니다. 최량 적합은 "최선" 열에 표시됩니다.

정규 혼합 보고서

정규 혼합 보고서에서는 각 군집에 대해 다음과 같은 요약 통계량을 제공합니다 .

- " 군집 요약 " 보고서에는 각 군집에 대한 관측값 수와 비율이 제공됩니다 .
- " 군집 평균 " 보고서에는 각 군집에서 각 변수에 대한 관측값의 평균이 제공됩니다 .
- " 군집 표준편차 " 보고서에는 각 군집에서 각 변수에 대한 관측값의 표준편차가 제공됩니다 .
- "-1* 로그 가능도 " 테이블에는 음의 로그 가능도 , BIC 및 AICc 가 제공됩니다 . 자세한 내용은 선형 모형 적합의 에서 확인하십시오 .
- " 정규 혼합에 대한 상관 " 보고서에는 각 군집에 대한 추정 상관 행렬이 제공됩니다 .

정규 혼합 보고서 옵션

행렬도 데이터의 처음 두 주성분으로 표현된 점 및 군집의 그림을 군집 색상을 식별하는 범례와 함께 표시합니다 . 군집 중심을 기준으로 원이 그려지고 원 크기는 군집 내의 개수에 비례합니다 . 음영 영역은 평균 주위의 밀도 등고선입니다 . 기본적으로 이 영역은 해당 군집의 관측값 중 90% 가 속하는 위치를 나타냅니다 (Mardia et al. 1980 참조). 그림 아래의 목록을 사용하여 그림 축을 다른 주성분으로 변경할 수 있습니다 . 또는 화살표 버튼을 사용하여 가능한 모든 축 조합을 순환할 수 있습니다 . 군집 색상을 데이터 테이블에 저장하는 옵션도 그림 아래에 있습니다 . 자세한 내용은 " 테이블에 색상 저장 " 에서 확인하십시오 . 고유값은 내림차순으로 표시됩니다 .

참고 : 행렬도에서는 항상 상관 행렬을 사용하여 주성분을 계산합니다 .

행렬도 옵션 행렬도 모양을 제어하기 위한 옵션을 포함합니다 .

행렬도 선 표시 행렬도 선을 표시합니다 . 라벨이 지정된 선은 주성분으로 정의된 부분 공간에서 공변량의 방향을 표시합니다 . 이러한 선은 각 변수와 각 주성분의 연관성 정도를 나타냅니다 .

행렬도 선 위치 행렬도 선의 위치와 반지름 척도를 지정할 수 있습니다 . 기본적으로 선은 점 (0,0) 에서 시작됩니다 . 그림에서 선을 드래그하거나 이 옵션을 사용하여 좌표를 지정할 수 있습니다 . 반지름 척도 옵션을 사용하여 선이 더 잘 보이게 선 척도를 조정할 수도 있습니다 .

행렬도 등고선 밀도 밀도 등고선에 대한 신뢰 수준을 지정할 수 있습니다 . 기본 신뢰 수준은 90% 입니다 .

군집 표시 데이터 테이블의 행에 군집을 식별하는 표식을 할당합니다 .

행렬도 3D 데이터의 3 차원 행렬도를 표시합니다 . 변수가 세 개 이상인 경우에만 사용할 수 있습니다 .

평행 좌표 그림 각 군집에 대한 평행 좌표 그림을 생성합니다 . 이 그림 보고서에는 데이터와 평균을 표시하거나 숨기는 옵션이 있습니다 . 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .

산점도 행렬 모든 Y 변수를 사용하여 산점도 행렬을 생성합니다 .

테이블에 색상 저장 데이터 테이블의 행에 군집을 식별하는 색상을 할당합니다 . 보고서 창에 " 행렬도 " 가 있는 경우 데이터 테이블에 저장된 색상은 행렬도의 군집 색상과 일치합니다 . " 행렬도 " 에서 색상을 변경하고 " 테이블에 색상 저장 " 옵션을 다시 선택하면 테이블의 색상이 행렬도의 색상과 일치하도록 업데이트됩니다 .

참고 : 저장 옵션을 선택하면 저장된 각 열에 해당 특정 열의 데이터에 대한 군집 수를 지정하는 " 노트 " 열 특성이 포함됩니다 . 이렇게 하면 둘 이상의 군집 적합에서 열을 저장하고 열 특성을 사용하여 저장된 열을 가져온 군집화 적합을 식별할 수 있습니다 .

군집 저장 주어진 행이 할당된 군집의 번호를 포함하는 **군집** 열을 데이터 테이블에 저장합니다 . 정규 혼합의 경우 이 군집은 최대 확률 분류 군집입니다 .

군집 계산식 저장 **군집 계산식**이라는 계산식 열을 데이터 테이블에 저장합니다 . 이 계산식은 행이 속한 군집을 식별합니다 .

군집 계산식 게시 " 군집 계산식 저장 " 옵션에 사용된 것과 동일한 스코어링 코드를 " 계산식 저장소 " 에 게시합니다 . " 군집 계산식 게시 " 를 선택하고 " 계산식 저장소 " 내의 모형에서 " 스크립트 실행 " 을 선택하는 경우 , 데이터 테이블에 저장된 열은 " 군집 계산식 저장 " 을 선택할 때 저장된 열과 일치해야 합니다 .

혼합 확률 저장 각 군집에 대해 관측값이 해당 군집에 속할 확률을 포함하는 **군집 확률** <k> 열을 추가합니다 .

혼합 계산식 저장 혼합 확률을 계산하는 데 사용된 계산식을 포함하는 열을 데이터 테이블에 추가합니다. 이러한 계산식 열은 제외된 데이터 또는 테이블에 추가한 데이터의 확률을 스코어링하는 데 사용됩니다.

거리 계산식 <k> 관측값에서 계산된 군집 <k>에 대한 추정된 다변량 정규 밀도 함수입니다.

총 거리 거리 계산식 열의 합계입니다. 이 열의 계산식은 " 밀도 계산식 저장 " 옵션으로 생성된 **혼합 밀도** 열의 계산식과 동일합니다.

확률 계산식 <k> 관측값이 군집 <k>에 속할 확률입니다. 이러한 열에는 " 혼합 확률 저장 " 옵션으로 생성된 **군집 확률 <k>** 열의 값을 제공하는 계산식이 포함됩니다. 혼합 확률을 계산하기 위한 열 계산식은 다음과 같습니다.

$$\text{확률 계산식 <k>} = \frac{\text{거리 계산식 <k>}}{\text{총 거리}}$$

밀도 계산식 저장 정규 혼합에 대한 추정 밀도 함수를 포함하는 **혼합 밀도** 열을 데이터 테이블에 추가합니다.

군집 시뮬레이션 혼합 밀도를 사용하여 예측 변수 값을 시뮬레이션합니다. 이러한 값과 해당 값이 분류되는 군집을 새 데이터 테이블에 저장합니다.

제거 군집화 보고서를 제거합니다.

정규 혼합 군집화 플랫폼에 대한 통계 상세 정보

정규 혼합에서는 Newton-Raphson 알고리즘보다 안정적인 EM 알고리즘을 사용하여 적합을 수행합니다. 또한 JMP에서는 공분산 행렬이 특이 행렬인 경우 이를 원활하게 처리할 수 있는 Bayes 정규화 버전의 EM 알고리즘을 사용합니다. 추정값은 초기 추측에 크게 의존하므로 플랫폼에서 여러 번 투어를 반복합니다. 각 투어마다 초기 중심에 대해 랜덤 선택된 점이 있습니다.

투어를 여러 번 반복하면 추정 프로세스에 시간이 많이 소요되므로 큰 문제의 경우 상당히 오래 기다려야 합니다. 컨트롤을 사용하여 투어 및 반복 제한을 지정할 수 있습니다.

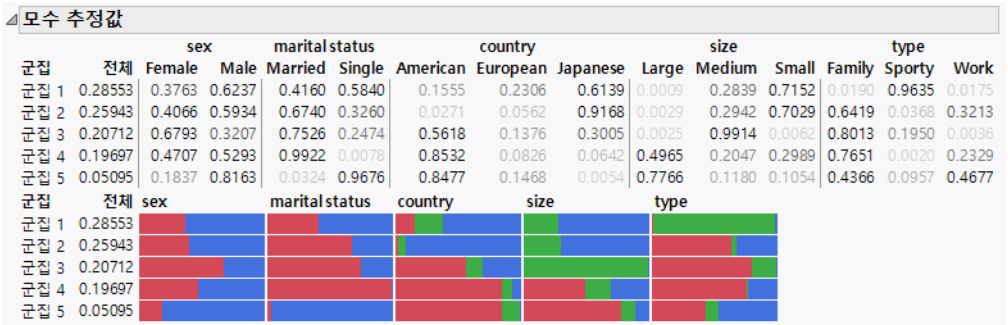
15 장

잠재 계층 분석

범주형 변수의 관측값 그룹화

잠재 계층 분석을 사용하면 범주형 반응 변수에 대한 관측값 군집을 찾을 수 있습니다. 잠재 변수는 관측 불가능한 그룹화 변수입니다. 잠재 변수의 각 수준을 잠재 계층이라고 합니다. 잠재 계층 분석 플랫폼에서는 잠재 계층 모형을 적합시키고 각 관측값에 대한 잠재 계층 또는 최대 확률 분류 군집을 결정합니다. 대부분의 상황에서 주제 관련 전문가는 잠재 계층 분석 결과를 사용하여 계층의 특성을 기반으로 각 잠재 계층에 대한 정의를 생성합니다.

그림 15.1 잠재 계층 분석의 예



모수 추정값

		sex		marital status		country			size			type		
군집	전체	Female	Male	Married	Single	American	European	Japanese	Large	Medium	Small	Family	Sporty	Work
군집 1	0.28553	0.3763	0.6237	0.4160	0.5840	0.1555	0.2306	0.6139	0.0009	0.2839	0.7152	0.0190	0.9635	0.0175
군집 2	0.25943	0.4066	0.5934	0.6740	0.3260	0.0271	0.0562	0.9168	0.0029	0.2942	0.7029	0.6419	0.0368	0.3213
군집 3	0.20712	0.6793	0.3207	0.7526	0.2474	0.5618	0.1376	0.3005	0.0025	0.9914	0.0062	0.8013	0.1950	0.0036
군집 4	0.19697	0.4707	0.5293	0.9922	0.0078	0.8532	0.0826	0.0642	0.4965	0.2047	0.2989	0.7651	0.0020	0.2329
군집 5	0.05095	0.1837	0.8163	0.0324	0.9676	0.8477	0.1468	0.0054	0.7766	0.1180	0.1054	0.4366	0.0957	0.4677

군집	전체	sex	marital status	country	size	type
군집 1	0.28553					
군집 2	0.25943					
군집 3	0.20712					
군집 4	0.19697					
군집 5	0.05095					

목차

잠재 계층 분석 플랫폼 개요 301
잠재 계층 분석의 예 301
잠재 계층 분석 플랫폼 시작 305
잠재 계층 분석 보고서 306
    군집 비교 보고서 306
    잠재 계층 모형 보고서 306
잠재 계층 분석 플랫폼 옵션 308
    잠재 계층 분석 308
    잠재 계층 모형 옵션 309
잠재 계층 분석 플랫폼의 추가 예 309
    소속 군집 확률 그림 309
잠재 계층 분석 플랫폼에 대한 통계 상세 정보 311
    잠재 계층 모형 적합 311
    최대 군집 수 312

잠재 계층 분석 플랫폼 개요

잠재 계층 분석 플랫폼에서는 범주형 반응 변수에 잠재 계층 모형을 적합시키고 각 관측값에 대한 잠재 계층 또는 최대 확률 분류 군집을 결정합니다. **잠재 변수**는 관측 불가능한 그룹화 변수입니다. 잠재 변수의 각 수준을 **잠재 계층**이라고 합니다. 예를 들어 위험 우선 순위에 따라 그룹화된 설문 조사 응답자 군집이 잠재 계층일 수 있습니다.

모형은 다항 혼합 모형의 형태를 취합니다. 모형에는 γ 모수와 ρ 모수의 두 가지 모수 집합이 있습니다. γ 모수는 소속 군집의 전체 확률을 나타냅니다. ρ 모수는 소속 군집을 조건으로 주어진 반응을 관측할 확률을 나타냅니다. 잠재 계층은 이러한 조건부 확률의 패턴을 특징으로 합니다.

분석 결과가 유의하려면 주제 관련 전문가가 플랫폼에서 생성되는 군집을 해석해야 합니다. 이 전문가는 잠재 계층의 특성을 조사하고 해당 특성을 기반으로 각 계층에 대한 정의를 생성합니다.

참고 : 반응 열에 결측값이 있는 행은 분석에서 제외됩니다 .

잠재 계층 모형에 대한 자세한 내용은 Collins 와 Lanza(2010) 및 Goodman(1974) 연구 자료에서 확인하십시오.

잠재 계층 분석의 예

이 예에서는 잠재 계층 분석 플랫폼을 사용하여 2005 년 미국 고등학생을 대상으로 한 설문 조사의 응답을 분석합니다. 설문 조사에서 학생들에게 건강 위험 행동과 관련된 다양한 선다형 질문을 했습니다.

이 예에서는 잠재 계층 모형을 적합시켜 12 개 질문에 대한 학생들의 반응을 기반으로 학생 군집을 식별합니다 . 분석 대상 열은 선다형 설문 조사 질문에서 응답을 두 개의 계층 (Yes/No) 으로 분류하여 얻었습니다 .

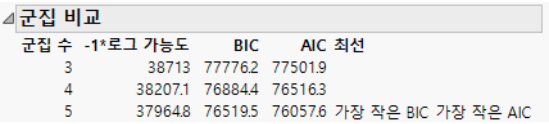
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Health Risk Survey.jmp 를 엽니다 .
2. Health Risk Survey 데이터 테이블에서 Launch LCA Platform 스크립트 옆의 녹색 삼각형을 클릭합니다 .

   이 스크립트는 12 개의 관심 대상 열을 선택하고 " 잠재 계층 분석 " 시작 창을 연 후 12 개 열을 Y 로 입력합니다 .

   참고 : LCA 플랫폼을 직접 시작하려면 분석 > 군집화 > 잠재 계층 분석을 선택합니다 .
3. **최대** 옆의 상자에 "5" 를 입력합니다 .

   이 옵션은 3 개의 군집 및 최대 5 개의 군집에 대한 잠재 계층 모형을 적합시킵니다 .
4. **확인**을 클릭합니다 .

그림 15.2 군집 비교 보고서



군집 비교

군집 수	-1*로그 가능도	BIC	AIC	최선
3	38713	77776.2	77501.9	
4	38207.1	76884.4	76516.3	
5	37964.8	76519.5	76057.6	가장 작은 BIC 가장 작은 AIC

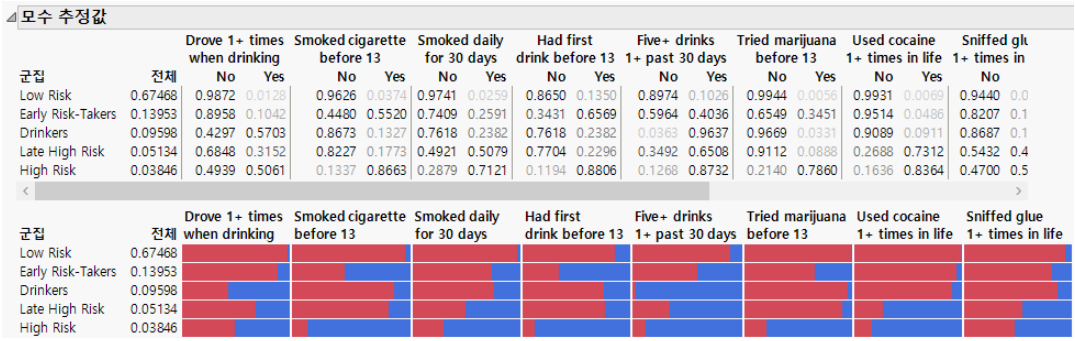
" 잠재 계층 분석 " 개요에는 " 군집 비교 " 보고서와 세 개의 개별 잠재 계층 모형 보고서가 포함됩니다 . 잠재 계층 모형 보고서에는 3 개 , 4 개 , 5 개 군집에 대한 모형이 각각 표시됩니다 . " 군집 비교 " 보고서를 보면 5 개 군집 모형의 BIC 및 AIC 가 가장 작으며 이는 세 모형 중 이 모형이 최량 적합 모형이라는 것을 나타냅니다 . 즉 , 이 모형이 분석 대상입니다 .

5. "5 개 군집의 잠재 계층 모형 " 보고서에서 " 모수 추정값 " 아래의 막대 차트를 검토합니다 . 다음 사항에 유의하십시오 .
   - 군집 1 은 모든 위험 행동에 대한 대답이 대부분 "No" 입니다 .
   - 군집 2 는 13 세 이전의 네 가지 위험 행동에 대해 "Yes" 대답이 많습니다 .
   - 군집 3 은 음주 운전 및 지난 30 일 동안 5 회 이상 음주에 대해 "Yes" 대답이 많습니다 .
   - 군집 4 는 13 세 이전의 행동을 제외한 대부분의 위험 행동에 대해 "Yes" 대답이 많습니다 .
   - 군집 5 는 대부분의 위험 행동에 대해 "Yes" 대답이 가장 많습니다 .

   이 정보를 사용하여 군집에 의미 있는 이름을 지정합니다 .
6. "5 개 군집의 잠재 계층 모형 " 옆의 빨간색 삼각형을 클릭하고 **군집 이름 바꾸기**를 선택합니다 .
   - 군집 1: Low Risk 입력
   - 군집 2: Early Risk-Takers 입력
   - 군집 3: Drinkers 입력
   - 군집 4: Late High Risk 입력
   - 군집 5: High Risk 입력
7. **확인**을 클릭합니다 .
8. JMP 경고가 나타나면 **확인**을 클릭합니다 .

참고 : 새 군집 이름은 스크립트에 저장되지 않습니다 .

그림 15.3 모수 추정값 보고서 (일부)



모수 추정값

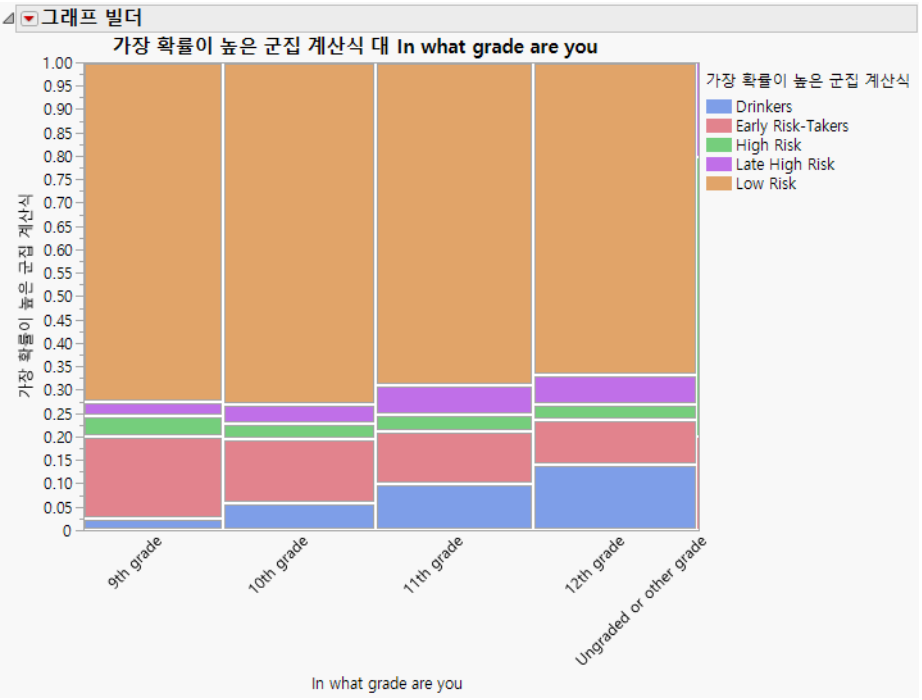
군집	전체	Drove 1+ times when drinking		Smoked cigarette before 13		Smoked daily for 30 days		Had first drink before 13		Five+ drinks 1+ past 30 days		Tried marijuana before 13		Used cocaine 1+ times in life		Sniffed gl 1+ times in
		No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Low Risk	0.67468	0.9872	0.0128	0.9626	0.0374	0.9741	0.0259	0.8650	0.1350	0.8974	0.1026	0.9944	0.0056	0.9931	0.0069	0.9440
Early Risk-Takers	0.13953	0.8958	0.1042	0.4480	0.5520	0.7409	0.2591	0.3431	0.6569	0.5964	0.4036	0.6549	0.3451	0.9514	0.0486	0.8207
Drinkers	0.09598	0.4297	0.5703	0.8673	0.1327	0.7618	0.2382	0.7618	0.2382	0.0363	0.9637	0.9669	0.0331	0.9089	0.0911	0.8687
Late High Risk	0.05134	0.6848	0.3152	0.8227	0.1773	0.4921	0.5079	0.7704	0.2296	0.3492	0.6508	0.9112	0.0888	0.2688	0.7312	0.5432
High Risk	0.03846	0.4939	0.5061	0.1337	0.8663	0.2879	0.7121	0.1194	0.8806	0.1268	0.8732	0.2140	0.7860	0.1636	0.8364	0.4700

그림 15.3 에서는 분석의 처음 8 개 변수에 대한 모수 추정값을 보여 줍니다 . 보고서 창에 새 군집 이름이 나타납니다 .

이제 "In what grade are you" 라는 인구 통계 질문에 대해 소속 군집을 비교합니다 .

9. "5 개 군집의 잠재 계층 모형 " 옆의 빨간색 삼각형을 클릭하고 **혼합 및 군집 계산식 저장**을 선택합니다 .
10. **그래프 > 그래프 빌더**를 선택합니다 .
11. In what grade are you 를 **X** 로 입력합니다 .
12. **최대 확률 분류 군집 계산식**을 **Y** 로 입력합니다 .
13. 모자이크 요소를 선택합니다 .
14. **완료**를 클릭합니다 .

그림 15.4 소속 군집 대 학년의 모자이크 그림

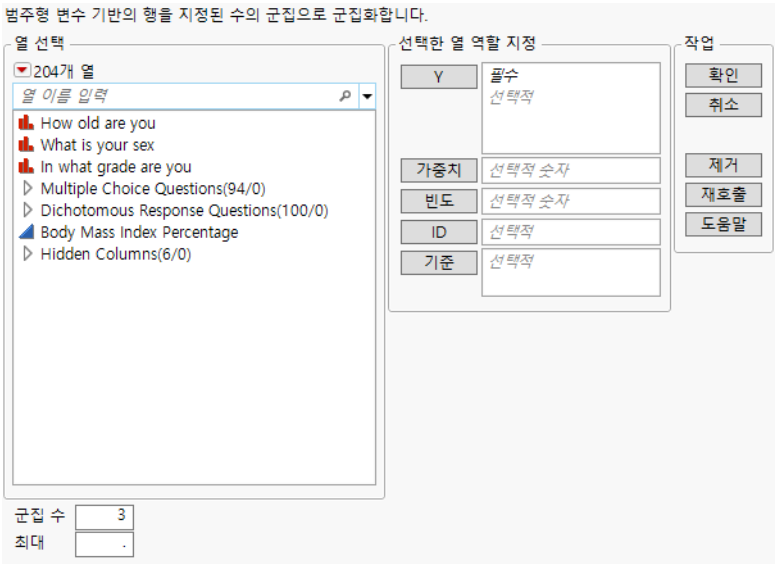


대부분의 응답자가 "Low Risk" 군집에 속한다는 것을 알 수 있습니다 . "Drinkers" 라벨이 지정된 계층은 학년이 올라갈수록 더 많은 응답자를 포함합니다 .

잠재 계층 분석 플랫폼 시작

분석 > 군집화 > 잠재 계층 분석을 선택하여 잠재 계층 분석 플랫폼을 시작합니다.

그림 15.5 잠재 계층 분석 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

잠재 계층 분석 플랫폼 시작 창에는 다음 옵션이 포함되어 있습니다.

Y 분석할 하나 이상의 열입니다. 명목형, 순서형 또는 다중 반응 모델링 유형의 열을 분석할 수 있습니다. 명목형 또는 순서형 반응을 분석하려면 두 개 이상의 열이 필요합니다. 반응이 여러 개 포함되고 모델링 유형이 다중 반응인 경우에는 열이 하나만 필요합니다.

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다.

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다.

ID 개별 응답자를 식별하는 데 사용되는 열입니다. 이 ID 는 일부 출력 테이블에 사용됩니다.

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열입니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다.

군집 수 분석에서 계산할 군집 수입니다.

최대 최대 군집 수를 지정합니다. 이 수가 "군집 수"에 지정된 값을 초과하면 "군집 수"와 "최대"사이의 범위에 있는 각 정수 값과 동일한 군집 수로 모형 보고서가 생성됩니다. 이러한 보고서는 "잠재 계층 분석" 보고서 개요의 일부로 나타납니다.

주의 : LCA 모형이 적절하게 적합시킬 수 있는 최대 군집 수가 있습니다 . 최대값보다 많은 군집을 요청하면 보고서 창에 경고 메시지가 나타납니다 . LCA 플랫폼에서는 열에 지원되는 최대 군집 수까지 적합시킵니다 . 최대 군집 수를 결정하는 방법에 대한 자세한 내용은 " 최대 군집 수 " 에서 확인하십시오 .

잠재 계층 분석 보고서

초기 "잠재 계층 분석" 보고서에는 "군집 비교" 보고서와 지정된 각 군집 수에 대한 잠재 계층 모형 보고서가 포함됩니다.

군집 비교 보고서

"군집 비교" 보고서에서는 다양한 모형을 비교하기 위한 적합 통계량을 제공합니다. 적합 통계량은 음의 로그 가능도(-1*로그 가능도), BIC 및 AIC입니다. 각 값이 작을수록 더 나은 적합을 나타냅니다. 최량 적합은 **최선** 열에 표시됩니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

잠재 계층 모형 보고서

각 잠재 계층 모형 보고서 이름은 *k*(적합된 군집 수) 에 따라 동적으로 "<k> 군집의 잠재 계층 모형 " 으로 지정됩니다 . 보고서에는 다음 결과와 개요가 포함됩니다 .

- " 모형 요약 "
- " 모수 추정값 "
- " 전치된 모수 추정값 "
- " 효과 크기 "
- "MDS 그림 "
- " 혼합 확률 "

모형 요약

기본적으로 지정된 수의 군집에 대한 모형 요약이 각 잠재 계층 모형 보고서 맨 위에 나타납니다. 모형 요약에는 -1*로그 가능도, 모수 수, BIC 및 AIC가 포함됩니다. 이러한 요약 값을 사용하여 모형이 데이터에 얼마나 잘 적합되는지 확인할 수 있습니다. -1*로그 가능도, BIC 및 AIC 값이 작을수록 더 나은 적합을 나타냅니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오. "모수 수" 값은 잠재 계층 모형의 고유 모수 수를 제공합니다. 자세한 내용은 "잠재 계층 분석 플랫폼에 대한 통계 상세 정보"에서 확인하십시오.

모수 추정값

"모수 추정값" 보고서는 기본적으로 표시되며 테이블 형식과 그래픽 형식의 모수 추정값 요약을 포함합니다. 각 요약에는 모형 군집에 해당하는 행이 포함됩니다.

"전체" 열에는 관측값이 각 군집에 속할 확률이 표시됩니다. 이는 γ 모수입니다. 자세한 내용은 "잠재 계층 분석 플랫폼에 대한 통계 상세 정보"에서 확인하십시오.

표시된 나머지 열은 " 잠재 계층 분석 " 시작 창에 지정된 Y 열에 따라 세로 구분자로 그룹화됩니다 .

- 범주형 반응 열의 각 그룹에는 각 반응 내의 각 수준에 대한 열이 있습니다. 각 그룹에서 주어진 행과 열의 값은 행으로 식별된 군집에 관측값이 속할 경우 열이 나타내는 반응의 조건부 확률입니다 . 이는 ρ 모수입니다 .
- 다중 반응 열의 각 그룹에는 다중 반응 내의 각 범주에 대한 열이 있습니다. 각 그룹에서 주어진 행과 열의 값은 행으로 식별된 군집에 관측값이 속할 경우 지정된 범주의 하위 수준에서 반응의 조건부 확률입니다 . 이는 ρ 모수입니다 .

그래픽 보기에는 조건부 확률 값이 **점유율 차트**로 표시됩니다. 각 군집 및 각 Y에 대해 주어진 소속 군집의 조건부 확률이 가로 누적 막대 차트로 표시됩니다. 이항 또는 명목형 반응 열의 경우 이러한 차트의 백분율 합계는 각 반응에 대해 1입니다. 다중 반응 열의 경우 각 범주의 하위 수준에 대한 백분율이며 합계가 1이 아닙니다. 막대 누적은 값 테이블에 표시되는 변수의 순서를 따릅니다. 막대를 커서로 가리켜 변수의 수준 또는 범주를 볼 수도 있습니다.

팁 : " 모수 추정값 " 보고서의 한 테이블에서 하나 이상의 행을 선택하여 해당 군집에 할당된 관측값을 선택할 수 있습니다 .

전치된 모수 추정값

"전치된 모수 추정값" 보고서에는 "모수 추정값" 보고서 테이블의 전치 테이블이 포함됩니다. 여기서는 군집이 열로 표시됩니다. 분석에 포함된 각 Y 열의 각 반응 범주에 대해 각 군집의 조건부 확률이 표시됩니다.

참고 : " 전체 " 열의 추정값은 전치된 테이블에 포함되지 않습니다 .

효과 크기

"효과 크기" 테이블은 기본적으로 표시되며 군집에 대해 Y 열을 비교합니다. 이 테이블의 각 행에 있는 통계량은 Y 열의 수준 또는 범주별로 소속 군집에 대한 기대 개수의 분할표 분석에서 얻습니다. 기대 개수는 각 군집의 관측값 수에 Y 열의 각 수준 또는 범주에 대한 조건부 확률을 곱한 값입니다.

각 반응에서 군집별 수준의 기대 개수 분할표에 대해 Pearson 카이제곱 통계량 χ^2 이 계산됩니다 . n 이 관측값 수를 나타낸다고 하겠습니다 . " 효과 크기 " 열의 값은 다음과 같이 정의됩니다 .

$$\text{효과 크기} = \sqrt{\frac{\chi^2}{n}}$$

"LR Logworth" 열의 각 값은 -log10(p_{LR})을 나타내며, 여기서 p_{LR}은 기대 개수의 분할표에 대한 가능도비 검정 p 값입니다. 2를 초과하는 Logworth 값은 0.01 유의 수준의 유의성에 해당합니다.

팁 : " 효과 크기 " 테이블에서 하나 이상의 행을 선택하여 데이터 테이블에서 해당 열을 선택할 수 있습니다 .

MDS 그림

"MDS 그림"은 기본적으로 표시되며 각 군집마다 하나의 점을 포함합니다. 이 그림은 군집 근접성을 2차원으로 표현한 것입니다. 서로 더 가까운 군집이 더 유사합니다. 이 그림은 ρ 모수의 비유사성 행렬에서 생성됩니다. MDS 그림에 대한 자세한 내용은 "다차원 척도법"에서 확인하십시오.

혼합 확률

"혼합 확률" 테이블에는 각 행의 소속 군집 확률이 표시됩니다. "최대 확률 분류 군집" 열은 각 행의 소속 확률이 가장 높은 군집을 나타냅니다.

참고 : 하나 이상의 Y 열에 대한 결측값이 포함된 행은 분석에서 제외되고 " 혼합 확률 " 테이블에 나타나지 않습니다 .

잠재 계층 분석 플랫폼 옵션

- " 잠재 계층 분석 "
- " 잠재 계층 모형 옵션 "

잠재 계층 분석

" 잠재 계층 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

새 군집 수 다른 개수의 군집을 사용하여 다른 분석을 실행할 수 있습니다 . 새 분석 보고서가 현재 보고서에 추가됩니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

잠재 계층 모형 옵션

"<k> 군집의 잠재 계층 모형 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

모형 보고서 사용 가능한 모형 보고서를 표시하거나 숨길 수 있습니다. 모형 보고서에 대한 자세한 내용은 " 잠재 계층 분석 보고서 " 에서 확인하십시오.

군집별 색상 데이터 테이블의 각 행을 최대 확률 분류 군집에 따라 색상을 적용합니다. 예는 " 잠재 계층 분석 플랫폼의 추가 예 " 에서 확인하십시오.

혼합 및 군집 계산식 저장 각 군집에 대한 계산식 열과 최대 확률 분류 군집에 대한 계산식 열을 데이터 테이블에 저장합니다.

군집 계산식만 저장 최대 확률 분류 군집을 결정하는 계산식이 포함된 열을 데이터 테이블에 저장합니다.

JMP PRO **확률 계산식 게시** 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. " 계산식 저장소 " 보고서가 열려 있지 않으면 " 계산식 저장소 " 보고서가 생성됩니다. 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

혼합 확률 저장 " 혼합 확률 " 테이블의 값을 데이터 테이블의 해당 행에 저장합니다.

군집만 저장 각 행에 대해 최대 확률 분류 군집이 포함된 새 열을 데이터 테이블에 저장합니다. 이 열에는 계산식이 포함되지 않습니다.

군집 이름 바꾸기 보고서의 군집에 의미 있는 이름을 지정할 수 있습니다.

참고 : 보고서에 난수 시드값을 지정한 경우 외에는 새 군집 이름이 스크립트에 저장되지 않습니다. 스크립트를 통해 보고서를 시작할 때만 난수 시드값을 설정할 수 있습니다.

적합 제거 보고서 창에서 지정된 군집화 보고서를 제거합니다.

잠재 계층 분석 플랫폼의 추가 예

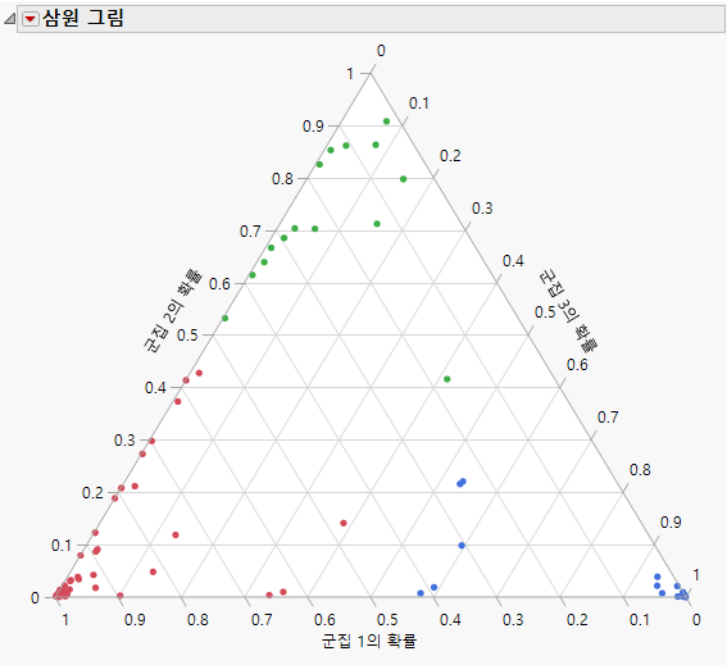
소속 군집 확률 그림

이 예에서는 자동차 소유자 및 제조회사에 대한 설문 조사 데이터가 포함된 Car Poll.jmp 샘플 데이터 테이블을 사용합니다. 자동차 소유자를 세 개의 군집으로 분류하고, 소속 군집 확률을 시각화하는 그림을 생성하려고 합니다. 군집이 세 개인 경우 삼원 그림을 사용하면 시각화하는 데 유용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다.
2. **분석 > 군집화 > 잠재 계층 분석**을 선택합니다.

3. age 를 제외한 모든 열을 선택하고 **Y** 를 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "3 개 군집의 잠재 계층 모형 " 옆의 빨간색 삼각형을 클릭하고 **군집별 색상**을 선택합니다 .
6. "3 개 군집의 잠재 계층 모형 " 옆의 빨간색 삼각형을 클릭하고 **혼합 확률 저장**을 선택합니다 .
7. Car Poll 데이터 테이블 창의 열 목록에서 LCA 군집 확률 열 그룹을 선택합니다 .
8. **그래프 > 삼원 그림**을 선택합니다 .
9. **X, 그림에 그리기**를 클릭합니다 .
10. **확인**을 클릭합니다 .

그림 15.6 소속 군집 확률의 삼원 그림



각 관측값에 대한 군집 확률 삼원 그림에서 대부분의 소속 군집 확률이 꼭지점 근처에 있습니다 . 이는 한 군집에 대해 높은 값이 있고 다른 두 군집에 대해 낮은 값이 있음을 나타냅니다 . 그러나 그림 중간에 몇 개의 점이 있으며 , 이는 해당 관측값이 어느 군집에도 소속될 확률이 높지 않음을 나타냅니다 . 이러한 관측값은 더 철저한 조사가 필요하거나 , 데이터를 더 잘 표현하기 위해 군집이 더 필요하다는 것을 나타낼 수 있습니다 .

참고 : 난수 시드값이 지정되지 않았으므로 결과가 다를 수 있습니다 .

잠재 계층 분석 플랫폼에 대한 통계 상세 정보

- "잠재 계층 모형 적합"
- "최대 군집 수"

잠재 계층 모형 적합

이 섹션에서는 잠재 계층 분석 플랫폼에서 적합되는 잠재 계층 모형에 대해 설명합니다. 잠재 계층 모형에 대한 자세한 내용은 Collins와 Lanza(2010) 및 Agresti(2013) 연구 자료에서 확인하십시오.

참고 : 텍스트 탐색기 플랫폼에서 사용되는 LCA 알고리즘은 문서 용어 행렬의 희소성을 활용합니다. 이러한 이유로 텍스트 탐색기 플랫폼의 LCA 결과는 잠재 계층 분석 플랫폼의 결과와 정확하게 매칭되지 않습니다.

$j = 1, ..., J$가 관측된 반응 열을 나타낸다고 하겠습니다. 이는 잠재 계층 분석 플랫폼 시작 창의 Y 열입니다. j 열의 수준 수를 R_j로 나타냅니다.

J 변수의 다차원 분할표에는 $W = R_1 * ... * R_J$ 셀이 포함됩니다. 이러한 각 셀은 J 변수에 대한 반응 패턴으로 정의됩니다. 따라서 각 반응 패턴은 $\mathbf{y} = y_1, ..., y_j$ 형식의 J 길이 벡터입니다. $\mathbf{Y}$를 행 벡터로 간주되는 모든 반응 패턴의 $W \times J$ 배열로 정의합니다. $\mathbf{Y}$에 있는 각 요소 $\mathbf{y}_w$의 확률은 $\Pr(\mathbf{y}_w)$입니다. 이러한 확률의 합은 1 입니다.

$$\sum_{w=1}^{W} Pr(y_w) = 1$$

다음 표기를 고려하십시오.

- C = 잠재 계층 모형의 군집 수
- γ_c = 군집 c에 소속될 확률. γ_c는 잠재 계층 출현율입니다. 이러한 모수의 합은 1 입니다.
- $r_{j,k}$ = j 번째 반응의 k 번째 수준
- $\rho_{j,k|c}$ = c 계층의 소속을 조건으로 j 열에서 $r_{j,k}$ 반응을 관측할 확률. $\rho_{j,k|c}$는 항목 - 반응 확률입니다. 지정된 군집 및 반응 변수 j에 대해 $\rho_{j,k|c}$의 합은 1 입니다.
- $I(y_j = r_{j,k}) = y_j$ 반응이 j 번째 반응의 k 번째 수준이면 1, 그렇지 않으면 0 인 표시 함수

반응 $\mathbf{y}_w = y_1, ..., y_j$의 특정 벡터를 관측할 확률은 C 개의 각 잠재 계층에 대해 해당 반응 벡터를 관측할 조건부 확률의 합계입니다.

$$\Pr(y) = \sum_{c=1}^{C} \gamma_c \prod_{j=1}^{J} \prod_{k=1}^{R_j} \rho_{j,k|c}^{I(y_j = r_{j,k})}$$

이 방정식은 "잠재 계층 모형"의 빨간색 삼각형 메뉴에서 "혼합 및 군집 계산식 저장" 옵션을 선택하여 데이터 테이블에 저장할 수 있는 **확률 계산식 군집** 계산식의 분모입니다. **확률 계산식 군집** 열의 계산식은 $\Pr(\text{군집} = c \mid \mathbf{y}_w)$을 제공하며, 이는 $\Pr(\mathbf{y}_w, \text{군집} = c)/\Pr(\mathbf{y}_w)$과 같습니다.

잠재 계층 모형의 γ 및 ρ 모수는 반복 EM(Expectation-Maximization) 알고리즘을 사용하여 추정됩니다. 잠재 계층 모형의 고유 모수 수는 다음과 같이 정의됩니다.

$$(C-1) + C\sum_{j=1}^{J}(R_j - 1)$$

최대 군집 수

LCA 모형에 적합시킬 수 있는 최대 군집 수는 모형 자유도에 따라 다릅니다. LCA 모형의 자유도는 열에 의해 생성된 분할표 크기를 기반으로 합니다. 분할표 크기는 테이블에서 하나 이상의 관측값이 포함된 셀 개수이며 K로 나타냅니다. 모든 셀에 하나 이상의 관측값이 포함된 경우 K는 반응 열의 수준 수를 곱한 값입니다. 자유도 계산식은 다음과 같이 정의됩니다.

DF = K - {nCluster - 1 + nCluster(nTotalLevels - nCols)} - 1

다음은 각 요소에 대한 설명입니다.

nCluster = 군집 수

nTotalLevels = 반응 열 수준의 합계

nCols = 반응 열 수

LCA 모형을 적절하게 적합시키려면 자유도가 양수여야 합니다. 따라서 DF > 0 을 위해 최대 군집 수는 다음과 같이 정의됩니다.

max(nCluster) < floor[K/(1 + nTotal Levels – nCols)]

16 장

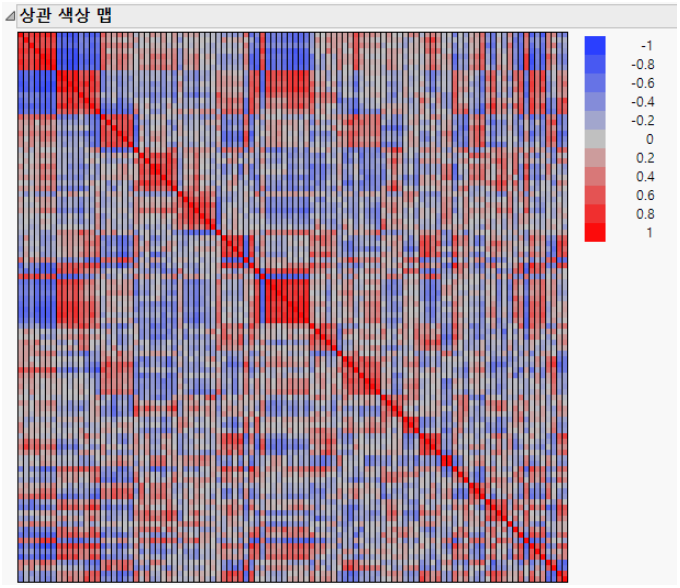
변수 군집화

유사한 변수를 대표 그룹으로 그룹화

변수 군집화는 유사한 변수를 대표 그룹으로 그룹화하는 방법을 제공합니다. 각 군집은 단일 성분 또는 변수로 나타낼 수 있습니다. 성분은 군집에 있는 모든 변수의 선형 결합입니다. 아니면 군집을 해당 군집에서 가장 대표적인 멤버로 식별되는 변수로 나타낼 수도 있습니다.

변수 군집화를 차원 축소 방법으로 사용할 수 있습니다. 큰 변수 집합을 사용하여 모델링하는 대신 군집 성분 또는 군집에서 가장 대표적인 변수를 사용하여 데이터 변동의 대부분을 설명할 수 있습니다. 또한 변수 군집화를 사용한 차원 축소는 주성분을 사용한 차원 축소보다 의미 있는 해석이 가능한 경우가 많습니다.

그림 16.1 변수 상관 맵의 예



목차

변수 군집화 플랫폼 개요 315
변수 군집화 플랫폼의 예 315
변수 군집화 플랫폼 시작 317
변수 군집화 보고서 317
상관 색상 맵 318
군집 요약 318
군집 멤버 318
표준화 성분 319
변수 군집화 플랫폼 옵션 319
변수 군집화 플랫폼의 추가 예 320
상관 색상 맵의 예 320
차원 축소를 위한 변수 군집화 플랫폼의 예 321
변수 군집화 플랫폼에 대한 통계 상세 정보 324
변수 군집화 알고리즘 324

변수 군집화 플랫폼 개요

주성분 분석은 분석에 포함된 모든 변수의 선형 결합인 성분을 생성합니다. 반면에 변수 군집화 옵션은 유사한 변수로 구성된 군집에 있는 변수의 선형 결합인 성분을 생성합니다. 전체 변수 집합이 군집으로 분할됩니다. 각 군집에 대해 해당 군집 변수의 첫 번째 주성분을 사용하여 **군집 성분**이 생성됩니다. 이 군집 성분은 해당 군집의 변수 중에서 가능한 한 많은 변동을 설명하는 선형 결합입니다.

변수 군집화 옵션을 차원 축소 방법으로 사용할 수 있습니다. 흔히 큰 변수 집합에서 변동의 상당 부분을 군집 성분 또는 군집에서 가장 대표적인 변수로 나타낼 수 있습니다. 그런 다음 이러한 새 변수를 예측 또는 기타 모델링 기법에 사용할 수 있습니다. 일반적으로 새 군집 기반 변수는 모든 변수에 기반한 주성분보다 해석하기가 쉽습니다.

공통 변수 집합에서 생성된 주성분은 직교합니다. 그러나 군집 성분은 구분되는 변수 집합에서 생성되므로 직교하지 않습니다.

큰 변수 집합을 사용하는 경우 변수 군집화 플랫폼에서는 특이값 분해를 기반으로 하는 알고리즘을 사용하여 계산 시간을 단축합니다. 추가 배경 정보는 "와이드 선형 방법 및 특이값 분해"에서 확인하십시오.

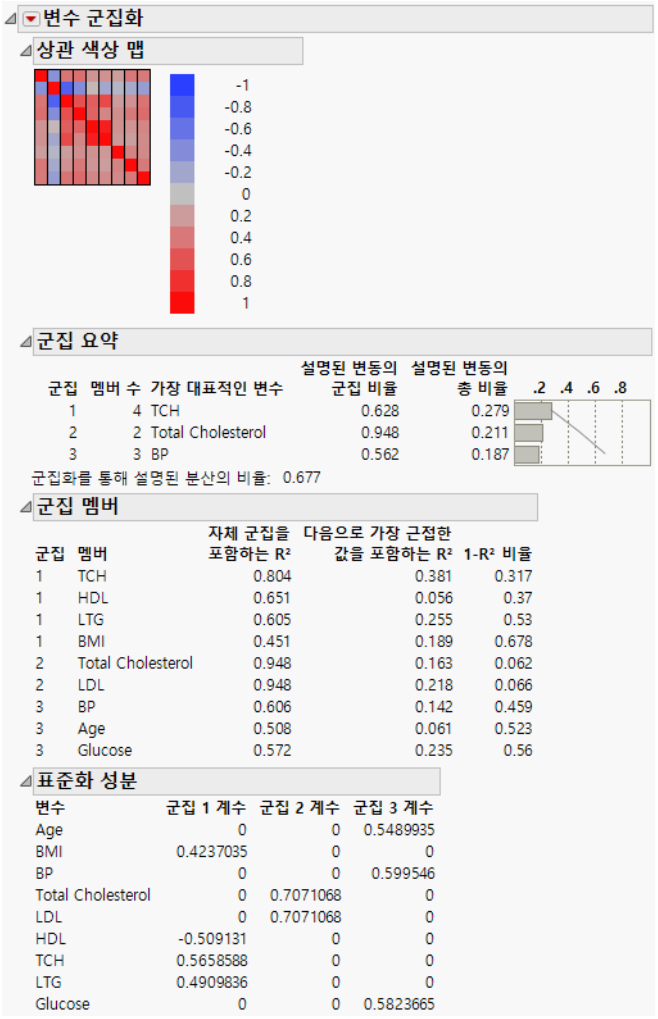
변수 군집화 플랫폼의 예

Diabetes.jmp 샘플 데이터 테이블에는 질병 경과를 모델링하는 데 사용되는 10 개의 기준 변수가 포함되어 있습니다 . 이 예에서는 연속형 기준 변수를 군집화합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다 .
2. **분석 > 군집화 > 변수 군집화**를 선택합니다 .
3. Gender 를 제외한 Age ~ Glucose 열 (Age, BMI, BP, Total Cholesterol, LDL, HDL, TCH, LTG 및 Glucose) 을 선택하고 **Y, 열**을 클릭합니다 .

   변수 군집화에는 연속형 수치 변수가 필요하므로 Gender 열을 포함할 수 없습니다 .
4. **확인**을 클릭합니다 .

그림 16.2 Diabetes 데이터에 대한 변수 군집화 보고서



변수 군집화

상관 색상 맵

-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

군집 요약

군집	멤버 수	가장 대표적인 변수	설명된 변동의 군집 비율	설명된 변동의 총 비율
1	4	TCH	0.628	0.279
2	2	Total Cholesterol	0.948	0.211
3	3	BP	0.562	0.187

.2 .4 .6 .8

군집화를 통해 설명된 분산의 비율: 0.677

군집 멤버

군집	멤버	자체 군집을 포함하는 R^2	다음으로 가장 근접한 값을 포함하는 R^2	$1-R^2$ 비율
1	TCH	0.804	0.381	0.317
1	HDL	0.651	0.056	0.37
1	LTG	0.605	0.255	0.53
1	BMI	0.451	0.189	0.678
2	Total Cholesterol	0.948	0.163	0.062
2	LDL	0.948	0.218	0.066
3	BP	0.606	0.142	0.459
3	Age	0.508	0.061	0.523
3	Glucose	0.572	0.235	0.56

표준화 성분

변수	군집 1 계수	군집 2 계수	군집 3 계수
Age	0	0	0.5489935
BMI	0.4237035	0	0
BP	0	0	0.599546
Total Cholesterol	0	0.7071068	0
LDL	0	0.7071068	0
HDL	-0.509131	0	0
TCH	0.5658588	0	0
LTG	0.4909836	0	0
Glucose	0	0	0.5823665

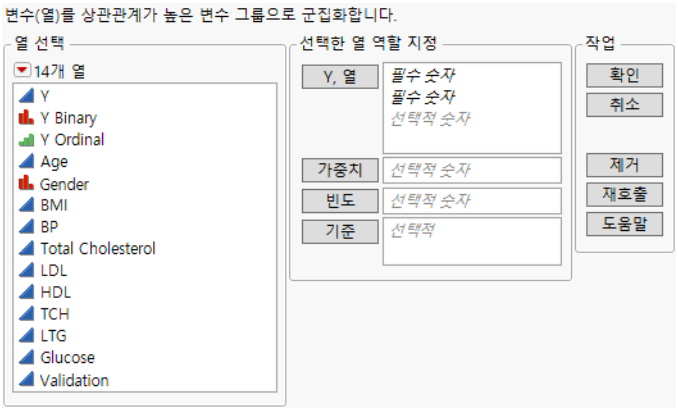
" 군집 요약 " 보고서에는 다음과 같이 변수가 세 개의 군집으로 그룹화되어 있습니다 .

- " 군집 멤버 " 보고서에 표시된 것처럼 군집 1 은 TCH, HDL, LTG 및 BMI 로 구성됩니다 . " 군집 요약 " 보고서에 따르면 TCH 가 군집 1 의 가장 대표적인 변수이고 , 군집 1 변수의 경우 변동의 62.8% 가 첫 번째 주성분에 의해 설명됩니다 .
- 군집 2 는 Total Cholesterol 과 LDL 로 구성됩니다 . " 군집 요약 " 보고서에 따르면 Total Cholesterol 이 군집 2 의 가장 대표적인 변수이고 , 군집 2 변수의 경우 변동의 94.8% 가 첫 번째 주성분에 의해 설명됩니다 .
- 군집 3 은 BP, Age 및 Glucose 로 구성됩니다 . " 군집 요약 " 보고서에 따르면 BP 가 가장 대표적인 변수이고 , 군집 3 변수의 경우 변동의 56.2% 가 첫 번째 주성분에 의해 설명됩니다 .

변수 군집화 플랫폼 시작

분석 > 군집화 > 변수 군집화를 선택하여 변수 군집화 플랫폼을 시작합니다 .

그림 16.3 변수 군집화 시작 대화상자



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y, 열 군집화할 변수입니다 . 변수는 수치형이고 연속형이어야 합니다 .

가중치 분석의 각 행에 가중치를 할당하는 숫자 값이 포함된 열입니다 .

빈도 분석의 각 행에 빈도를 할당하는 숫자 값이 포함된 열입니다 .

기준 각 수준으로 개별 분석을 정의하는 열입니다 . 지정된 열의 각 수준에 대해 해당하는 행이 분석됩니다 . 결과는 개별 보고서에 제공됩니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다 .

변수 군집화 보고서

기본적으로 " 변수 군집화 " 보고서에는 다음 항목이 표시됩니다 .

- " 상관 색상 맵 "
- " 군집 요약 "
- " 군집 멤버 "
- " 표준화 성분 "

팁 : " 변수 군집화 " 보고서에서 행을 선택하면 데이터 테이블에서 해당 열이 선택됩니다 . Ctrl 키를 누른 채 행을 클릭하면 데이터 테이블에서 열이 선택 취소됩니다 .

상관 색상 맵

"상관 색상 맵" 보고서에는 변수 간 상관관계에 대한 색상 맵이 표시됩니다. 변수는 "군집 멤버" 보고서에 나열된 순서대로 배열됩니다. 이렇게 배열하면 같은 군집의 멤버가 상관 그림에서 인접하게 배치됩니다. 자세한 내용은 "상관 색상 맵의 예"에서 확인하십시오.

팁 : 색상 맵에서 정사각형을 커서로 가리키면 해당 정사각형에 포함된 변수와 상관 값을 볼 수 있습니다 .

같은 군집의 변수는 다른 군집의 변수보다 절대 상관이 더 높은 경향이 있습니다(더 진한 빨간색 또는 파란색). 따라서 상관 맵에서 주어진 성분의 변수에 해당하는 셀로 형성된 정사각형이 대각선을 따라 두드러지는 경우가 많습니다.

상관은 행별 방법을 사용하여 계산됩니다. 이 방법은 변수에 대한 결측 데이터가 있는 관측값을 상관 계산에서 제외합니다. 행별 추정 방법에 대한 자세한 내용은 "추정 방법"에서 확인하십시오.

군집 요약

" 군집 요약 " 보고서에서는 다음 정보를 제공합니다 .

군집 군집 식별자입니다 .

멤버 수 군집의 변수 수입니다 .

가장 대표적인 변수 군집 성분과의 상관 제곱이 가장 큰 군집 변수입니다 .

설명된 변동의 군집 비율 군집의 변수 중 첫 번째 주성분에 의해 설명되는 군집 변동 비율입니다 . 군집의 변수가 하나뿐이면 이 값이 1 입니다 . 이 통계량은 모든 변수가 아니라 군집 내의 변수만 기반으로 합니다 .

설명된 변동의 총 비율 군집 성분에 의해 설명되는 전체 변동 비율입니다 . 이것은 각 군집 내의 변수만 사용하여 첫 번째 주성분을 계산하는 것과 동일합니다 .

테이블 아래에는 모든 군집 성분에 의해 설명되는 총 변동 비율에 대한 내용이 나와 있습니다 .

군집 멤버

" 군집 멤버 " 보고서에서는 다음 정보를 제공합니다 .

군집 군집 식별자입니다 .

멤버 군집에 포함된 변수입니다 .

자체 군집과의 R^2 변수와 해당 군집 성분의 상관 제곱입니다 .

다음으로 가장 근접한 군집과의 R^2 다음으로 가장 근접한 군집에 대한 변수와 군집 성분의 상관 제곱입니다 . 다음으로 가장 근접한 군집은 변수와 군집 성분의 상관 제곱이 두 번째로 높은 군집입니다 .

1-R² 비율 변수가 속한 군집과 다음으로 가장 근접한 군집 사이의 상대적 근접성에 대한 측도입니다. 이 값은 다음과 같이 정의됩니다.

(1 - 자체 군집과의 R^2)/(1 - 다음으로 가장 근접한 군집과의 R^2)

표준화 성분

"표준화 성분" 보고서에서는 군집 성분을 정의하는 계수를 제공합니다. 이러한 계수는 각 군집 내에서 첫 번째 주성분의 고유 벡터입니다.

변수 군집화 플랫폼 옵션

"변수 군집화"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

상관 색상 맵 "상관 색상 맵" 그림을 표시하거나 숨깁니다. 자세한 내용은 "상관 색상 맵"에서 확인하십시오.

군집 요약 "군집 요약" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "군집 요약"에서 확인하십시오.

군집 멤버 "군집 멤버" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "군집 멤버"에서 확인하십시오.

군집 성분 "표준화 성분" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "표준화 성분"에서 확인하십시오.

군집 성분 저장 군집 성분이라는 열 그룹을 데이터 테이블에 저장합니다. 각 열은 군집 <i> 성분으로 이름이 지정되고 군집 성분을 비중심화 및 비척도화 변수로 표현하는 계산식을 포함합니다.

모형 적합 시작 "모형 효과 생성" 목록에 각 군집의 가장 대표적인 변수가 입력된 "모형 규격" 창을 엽니다. 이 옵션을 사용하여 가장 대표적인 변수를 기반으로 모형을 생성할 수 있습니다.

팁 : 성분을 사용하여 모형을 적합시키려면 먼저 **군집 성분 저장** 옵션을 선택합니다. 그런 다음 "모형 효과 생성" 목록에서 각 군집에 대한 가장 대표적인 변수를 원하는 군집 성분 열로 바꿉니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

변수 군집화 플랫폼의 추가 예

- "상관 색상 맵의 예"
- "차원 축소를 위한 변수 군집화 플랫폼의 예"

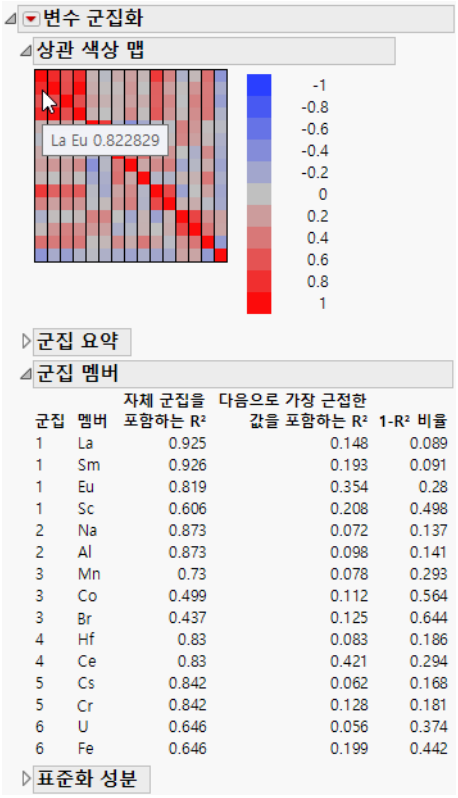
상관 색상 맵의 예

이 예에서는 상관 색상 맵을 생성하고 검토합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cherts.jmp 를 엽니다.
2. **분석 > 군집화 > 변수 군집화**를 선택합니다.
3. 모든 연속형 변수를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "군집 요약" 및 "표준화 성분" 보고서를 닫습니다.
6. 색상 맵에서 두 번째 행과 첫 번째 열의 셀을 커서로 가리킵니다.

   이 셀에 해당하는 변수가 La 와 Eu 이고 변수 상관이 0.822829 임을 보여 주는 툴팁이 나타납니다.

그림 16.4 Cherts.jmp 에 대한 상관 색상 맵



"군집 멤버" 보고서에서는 군집 1에 네 개의 변수가 있음을 보여 줍니다. "상관 색상 맵"에서 이 네 개 변수에 해당하는 왼쪽 위의 4x4 정사각형 셀은 뚜렷한 양의 상관관계 패턴을 보여 줍니다. 색상 맵에는 두 개 군집(군집 4 및 5)의 변수에 대한 양의 상관관계 패턴도 표시됩니다. 색상 맵에서 군집 6의 두 변수에 해당하는 오른쪽 아래의 2x2 정사각형 셀은 변수에 음의 상관관계가 있음을 나타냅니다. 자세한 내용은 "상관 색상 맵"에서 확인하십시오.

차원 축소를 위한 변수 군집화 플랫폼의 예

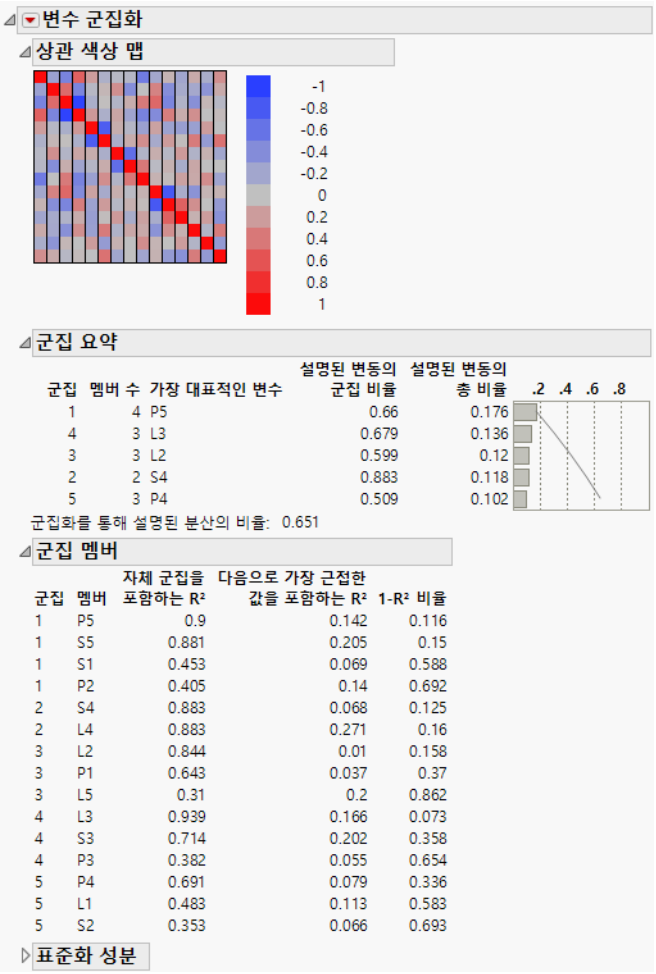
이 예에서는 변수 군집화 플랫폼을 모델링을 위한 차원 축소 도구로 사용합니다. Penta.jmp 샘플 데이터 테이블에는 반응 변수 log RAI를 예측하는 데 사용되는 15개의 변수가 포함되어 있습니다. 변수 군집화를 사용하여 이 수를 줄일 수 있습니다.

변수 군집화

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penta.jmp 를 엽니다 .
2. **분석 > 군집화 > 변수 군집화**를 선택합니다 .
3. logRAI 를 제외한 모든 연속형 변수를 선택하고 **Y, 열**을 클릭합니다 .

4. **확인**을 클릭합니다 .
5. " 변수 군집화 " 의 빨간색 삼각형을 클릭하고 **군집 성분 저장**을 선택합니다 .
   그룹화된 다섯 개의 계산식 열이 데이터 테이블에 추가됩니다 .

그림 16.5 Penta.jmp 에 대한 변수 군집화 보고서



"군집 요약" 및 "군집 멤버" 보고서에서는 변수가 다섯 개 그룹으로 군집화되어 다섯 개의 군집 성분 변수가 있음을 보여 줍니다.

모형 적합

이제 다음 두 모형을 적합시키고 비교하여 logRAI 를 예측합니다 .

- 모든 연속형 변수를 예측 변수로 사용하는 모형

- 군집 성분을 예측 변수로 사용하는 모형

1. " 변수 군집화 " 의 빨간색 삼각형을 클릭하고 **모형 적합 시작**을 선택합니다 .
2. logRAI 를 선택하고 **Y** 를 클릭합니다 .

   가장 대표적인 변수를 나타내는 다섯 개 군집이 " 모형 효과 생성 " 목록에 입력되어 있습니다 . 그러나 모든 예측 변수를 입력하려고 합니다 .
3. 모든 연속형 변수 (S1 ~ P5) 를 선택하고 **추가**를 클릭합니다 .

   Obs Name 을 포함하지 않도록 주의하십시오 .
4. **대화상자 열린 채 유지** 옆의 상자를 선택합니다 .
5. **실행**을 클릭합니다 .

그림 16.6 모든 연속형 예측 변수를 사용한 모형의 최소 제곱 적합 보고서

반응 log RAI

효과 요약

적합 요약

R^2	0.929316
Adj-R^2	0.853582
제곱근 평균 제곱 오차	0.331225
반응 평균	0.734333
관측값 수(또는 가중합)	30

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	15	20.193596	1.34624	12.2709
오차	14	1.535941	0.10971	Prob > F
수정 합계	29	21.729537		<.0001*

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	-0.802632	0.924946	-0.87	0.4002
P5	2.0563803	1.651272	1.25	0.2335
S4	-0.062354	0.134935	-0.46	0.6511
L2	0.0860287	0.061206	1.41	0.1817
L3	0.3185383	0.080091	3.98	0.0014*
P4	0.4136598	0.394449	1.05	0.3121
S1	-0.09783	0.038948	-2.51	0.0249*
L1	0.032362	0.049732	0.65	0.5258
P1	-0.107951	0.085209	-1.27	0.2259
S2	0.086703	0.044276	1.96	0.0704
P2	0.0847235	0.086297	0.98	0.3429
S3	-0.037728	0.055602	-0.68	0.5085
P3	-0.027313	0.233655	-0.12	0.9086
L4	-0.029756	0.152012	-0.20	0.8476
S5	2.7123146	2.222039	1.22	0.2424
L5	-0.209128	0.270401	-0.77	0.4521

효과 검정

효과 상세 정보

6. " 모형 적합 " 창의 " 모형 효과 생성 " 목록에서 모든 변수를 선택하고 **제거**를 클릭합니다 .
7. 군집 성분 그룹을 선택하고 **추가**를 클릭합니다 .
8. **실행**을 클릭합니다 .

그림 16.7 군집 성분을 예측 변수로 사용한 모형의 최소 제곱 적합 보고서

반응 log RAI

효과 요약

적합 요약

R^2	0.8214
Adj-R^2	0.784191
제곱근 평균 제곱 오차	0.402125
반응 평균	0.734333
관측값 수(또는 가중합)	30

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	5	17.848635	3.56973	22.0757
오차	24	3.880902	0.16170	Prob > F
수정 합계	29	21.729537		<.0001*

모수 추정값

항	추정값	표준 오차	t 비	Prob>|t|
절편	0.6651552	0.074349	8.95	<.0001*
군집 1 성분	-0.018483	0.056013	-0.33	0.7443
군집 2 성분	0.0035891	0.069032	0.05	0.9590
군집 3 성분	0.2043072	0.066714	3.06	0.0053*
군집 4 성분	0.5754553	0.065423	8.80	<.0001*
군집 5 성분	-0.046594	0.069786	-0.67	0.5107

효과 검정

효과 상세 정보

다섯 개의 군집 성분만 예측 변수로 포함하는 모형은 상당한 양의 반응 변동을 설명합니다(조정 $R^2 = 0.784$). 15개의 예측 변수를 모두 사용하는 모형의 조정 R^2은 약간 더 높은 0.853입니다(그림 16.6 참조).

변수 군집화 플랫폼에 대한 통계 상세 정보

변수 군집화 알고리즘

군집화 알고리즘은 더 이상 분할할 수 없을 때까지 반복적으로 변수 군집을 분할하고 변수를 군집에 재할당합니다. 초기 군집은 모든 변수로 구성됩니다. 이 알고리즘은 SAS에서 개발했으며 PROC VARCLUS(SAS Institute Inc. 2020g)에서 구현됩니다.

참고 : 이 알고리즘은 "Y, 열 " 목록의 변수에 대해 결측값이 없는 관측값만 사용합니다 .

알고리즘의 반복 단계는 다음과 같습니다 .

1. 모든 군집에 대해 다음을 수행합니다 .
   a. 각 군집의 변수에 대한 주성분을 계산합니다 .
   b. 모든 군집의 두 번째 고유값이 1 보다 작으면 알고리즘을 종료합니다 .

2. 다음 단계에 따라 두 번째 고유값이 가장 큰 (1 보다 큼) 군집을 두 개의 새 군집으로 분할합니다 .
   a. 직교 사각 회전을 사용하여 현재 군집의 변수에 대한 주성분을 회전합니다 .
   b. 현재 군집에서 첫 번째 회전된 주성분에 대한 상관 제곱이 두 번째 주성분에 대한 상관 제곱보다 높은 변수로 구성된 군집 하나를 정의합니다 .
   c. 원래 군집의 나머지 변수로 구성된 다른 군집을 정의합니다 . 이러한 변수는 두 번째 주성분과 더 높은 상관관계가 있습니다 .
   d. 두 개의 새 군집에 대한 주성분을 계산합니다 .
3. 데이터 집합의 변수를 다른 군집에 할당해야 하는지 여부를 검정합니다 . 각 변수에 대해 다음을 수행합니다 .
   a. 각 군집의 첫 번째 주성분과 변수의 상관 제곱을 계산합니다 .
   b. 상관 제곱이 가장 큰 변수를 군집에 포함합니다 .

참고 : 직교 사각 회전은 원시 Quartimax 회전이라고도 합니다 . 자세한 내용은 Harris 와 Kaiser(1964) 연구 자료에서 확인하십시오 .

부록 A

통계 상세 정보

다변량 방법

이 부록에서는 와이드 선형 방법에 대한 내용과 특이값 분해 사용에 대해 설명합니다.

목차

와이드 선형 방법 및 특이값 분해 329
특이값 분해 329
SVD 및 공분산 행렬 330
SVD 및 공분산 역행렬 330
SVD 계산 331

와이드 선형 방법 및 특이값 분해

군집화, 주성분 및 판별 플랫폼의 와이드 선형 방법을 사용하면 수천 개 또는 수백만 개의 변수가 있는 데이터 집합을 분석할 수 있습니다. 대부분의 다변량 기법에서는 공분산 행렬 또는 공분산 행렬의 역행렬을 계산해야 합니다. 다변량 분석에 많은 수의 변수가 포함될 경우 공분산 행렬이 엄청나게 커질 수 있으므로 공분산 행렬 또는 역행렬을 계산하는 데 문제가 있고 계산 시간이 많이 소요됩니다.

데이터가 n개의 행과 p개의 열로 구성되어 있다고 가정해 보겠습니다. 공분산 행렬의 계수는 n과 p 중 더 작은 값입니다. 가로형 데이터 집합에서는 보통 p가 n보다 훨씬 큽니다. 이러한 경우 공분산 행렬의 역행렬은 최대 n개의 0이 아닌 고유값을 가집니다. 와이드 선형 방법에서는 이 사실을 특이값 분해와 함께 사용하여 효율적인 계산을 제공합니다. 자세한 내용은 "SVD 계산"에서 확인하십시오.

특이값 분해

SVD(**특이값 분해**) 를 사용하면 선형 변환을 회전 , 척도화 , 다른 회전으로 표현할 수 있습니다 . SVD 에 따르면 $n \times p$ 행렬 **X** 를 다음과 같이 쓸 수 있습니다 .

$$\mathbf{X} = \mathbf{U}Diag(\Lambda)\mathbf{V}'$$

r을 **X**의 계수라고 하겠습니다. $r \times r$ 단위 행렬을 $\mathbf{I}_r$로 나타냅니다.

U, $Diag(\Lambda)$ 및 **V** 행렬은 다음과 같은 특성을 가집니다 .

U 는 $\mathbf{U}'\mathbf{U} = \mathbf{I}_r$ 인 $n \times r$ 반직교 행렬입니다 .

V 는 $\mathbf{V}'\mathbf{V} = \mathbf{I}_r$ 인 $p \times r$ 반직교 행렬입니다 .

$Diag(\Lambda)$는 열 벡터 $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_r)'$로 지정되는 양수 대각 요소가 포함된 $r \times r$ 대각 행렬입니다 . 여기서 $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > 0$ 입니다 .

λ_i는 **X**의 0이 아닌 **특이값**입니다.

다음은 SVD 를 정방 행렬의 스펙트럼 분해와 관련시킵니다 .

- λ_i 의 제곱은 $\mathbf{X}'\mathbf{X}$ 의 0 이 아닌 고유값입니다 .
- **V** 의 r 열은 $\mathbf{X}'\mathbf{X}$ 의 고유 벡터입니다 .

참고 : 문헌에는 **U**, **V** 행렬 및 특이값을 포함하는 행렬의 차원과 관련된 다양한 규칙이 있습니다 . 그러나 차이가 **X** 의 계수까지의 분해에 실제적인 영향을 미치지 않습니다 .

특이값 분해에 대한 자세한 내용은 Press et al. (1998, Section 2.6) 연구 자료에서 확인하십시오.

SVD 및 공분산 행렬

이 섹션에서는 SVD를 사용하여 공분산 행렬의 고유 벡터와 고유값을 구하는 방법을 설명합니다. 관심 행렬이 적어도 하나의 큰 차원을 갖는 경우 SVD를 계산하는 것이 공분산 행렬 및 고유값 분해를 계산하는 것보다 훨씬 더 효율적입니다.

n을 관측값 수, p를 관심 있는 다변량 분석에 포함된 변수의 수라고 하겠습니다. 데이터 값의 n x p 행렬을 $\mathbf{X}$로 나타냅니다.

SVD 는 대개 표준화된 데이터에 적용됩니다 . 값을 표준화하려면 평균을 뺀 후 표준편차로 나눕니다 . 표준화된 데이터 값의 n x p 행렬을 $\mathbf{X_s}$ 로 나타냅니다 . 그러면 표준화된 데이터의 공분산 행렬은 $\mathbf{X}$ 의 상관 행렬이며 다음과 같이 정의됩니다 .

$$Cov = \mathbf{X}_s'\mathbf{X}_s/(n-1)$$

SVD를 $\mathbf{X}_s$에 적용하여 $\mathbf{X}_s'\mathbf{X}_s$의 고유 벡터와 고유값을 얻을 수 있습니다. 이렇게 하면 $\mathbf{X}$ 행렬이 심한 가로형(열이 많음) 또는 세로형(행이 많음)일 때 고유 벡터와 고유값을 효율적으로 계산할 수 있습니다. 이 기법은 와이드 PCA의 기본입니다. 자세한 내용은 "주성분 보고서"에서 확인하십시오.

SVD 및 공분산 역행렬

일부 다변량 기법을 사용할 때 공분산 행렬의 역행렬을 계산해야 합니다. 이 섹션에서는 SVD를 사용하여 공분산 행렬의 역행렬을 계산하는 방법을 설명합니다.

표준화된 데이터 행렬을 $\mathbf{X_s}$ 로 나타내고 $\mathbf{S} = \mathbf{X_s}'\mathbf{X_s}$ 를 정의합니다 . 특이값 분해를 사용하면 $\mathbf{S}$ 를 다음과 같이 쓸 수 있습니다 .

$$\mathbf{S} = (\mathbf{U}Diag(\Lambda)\mathbf{V}')'(\mathbf{U}Diag(\Lambda)\mathbf{V}') = \mathbf{V}Diag(\Lambda)^2\mathbf{V}'$$

$\mathbf{S}$ 가 완전 계수이면 $\mathbf{V}$ 는 p x p 직교정규 행렬이고 $\mathbf{S}^{-1}$ 을 다음과 같이 쓸 수 있습니다 .

$$\mathbf{S}^{-1} = (\mathbf{V}Diag(\Lambda)^2\mathbf{V}')^{-1} = \mathbf{V}Diag(\Lambda)^{-2}\mathbf{V}'$$

$\mathbf{S}$ 가 완전 계수가 아니면 $\mathrm{Diag}(\Lambda)^{-1}$ 을 일반화 역행렬 $\mathrm{Diag}(\Lambda)^{+}$ 로 대체할 수 있습니다 . 여기서 $Diag(\Lambda)$ 의 대각 요소는 역수로 대체됩니다 . 즉 , $\mathbf{S}$ 의 일반화 역행렬이 다음과 같이 정의됩니다 .

$$\mathbf{S}^{-} = \mathbf{V}(Diag(\Lambda)^{+})^2\mathbf{V}'$$

이 일반화 역행렬은 SVD만 사용하여 계산할 수 있습니다.

와이드 선형 판별 분석에 SVD를 적용하는 데 대한 자세한 내용은 "와이드 선형 판별 방법"에서 확인하십시오.

SVD 계산

다변량 방법 플랫폼에서 JMP는 Golub와 Kahan(1965) 연구 자료에 제안된 방법에 따라 행렬의 SVD를 계산합니다. Golub와 Kahan 방법은 2단계 절차를 포함합니다. 첫 번째 단계에서는 **M** 행렬을 이중 대각 행렬 **J**로 축소합니다. 두 번째 단계에서는 **J**의 특이값을 계산합니다. 이 값은 원래 행렬 **M**의 특이값과 동일합니다. 일반적으로 **M** 행렬의 열은 변수가 계산에 미치는 효과를 균등화하기 위해 표준화됩니다. Golub와 Kahan 방법은 계산이 효율적입니다.

부록 B

참조 자료

Agresti, A. (2013). *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: John Wiley & Sons.

Baglama, J., and Reichel, L. (2005). "Augmented implicitly restarted Lanczos bidiagonalization methods." *SIAM Journal on Scientific Computing* 27:19–42.

Ballard, D. H. (1981). "Generalizing the Hough Transform to Detect Arbitrary Shapes." *Pattern Recognition* 13:111–122.

Bartlett, M. S. (1937). "Properties of sufficiency and statistical tests." *Proceedings of the Royal Society of London, Series A* 160:268–282.

Bartlett, M. S. (1954). "A Note on the Multiplying Factors for Various Chi Square Approximations." *Journal of the Royal Statistical Society, Series B* 16:296–298.

Benzécri, J. P. (1979). "Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire, addendum et erratum à [BIN. MULT.]." *Cahiers de l'Analyse des Données* 4:377–378.

Bentler, P. M. (1990). "Comparative Fit Indexes in Structural Models." *Psychological Bulletin* 107:238.

Bentler, P. M., and Freeman, E. H. (1983). "Tests for Stability in Linear Structural Equation Systems." *Psychometrika* 48:143–145.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Borg, I., and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: Theory and Applications.* 2nd ed. New York: Springer.

Boulesteix, A.-L., and Strimmer, K. (2007). "Partial Least Squares: A Versatile Tool for the Analysis of High-Dimensional Genomic Data." *Briefings in Bioinformatics* 8:32–44.

Browne, M. (2001). "An Overview of Analytic Rotation in Exploratory Factor Analysis." *Multivariate Behavioral Research* 36:111–150.

Browne, M. W., and Cudeck, R. (1993). "Alternative Ways of Assessing Model Fit." In *Testing Structural Equation Models*, edited by K. A. Bollen, and J. S. Long, 136–162. Newbury Park, CA: Sage Publications.

Chen, F. F. (2007). "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling: A Multidisciplinary Journal* 14:464–504.

Collins, L., and Lanza, S. (2010). *Latent Class and Latent Transition Analysis*. Hoboken NJ: John Wiley & Sons.

Cox, I., and Gaudard, M. (2013). *Discovering Partial Least Squares with JMP*. Cary, NC: SAS Institute Inc.

Cronbach, L. J. (1951). "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16:297–334.

Cudeck, R., and MacCallum, R. C., eds. (2007). Factor Analysis at 100, Historical Developments and Future Directions. Mahwah, NJ: Lawrence Erlbaum Associates.

de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: Guilford Press.

De Jong, S. (1993). "SIMPLS: An Alternative Approach to Partial Least Squares Regression." *Chemometrics and Intelligent Laboratory Systems* 18:251–263.

Denham, M. C. (1997). "Prediction Intervals in Partial Least Squares." *Journal of Chemometrics* 11:39–52.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikstrom, C., and Wold, S. (2006). *Multi- and Megavariate Data Analysis Basic Principles and Applications (Part I)*. Chapter 4. Umetrics.

Finkbeiner, C. (1979). "Estimation for the Multiple Factor Model when Data are Missing." *Psychometrika* 44:409–420.

Fisher, L., and Van Ness, J. W. (1971). "Admissible Clustering Procedures." *Biometrika* 58:91–104.

Florek, K., Lukaszewicz, J., Perkal, J., and Zubrzycki, S. (1951a). "Sur la liaison et la division des points d'un ensemble fini." *Colloquium Mathematicae* 2:282–285.

Florek, K., Lukaszewicz, J., Perkal, J., and Zubrzycki, S. (1951b). "Taksonomia Wroclawska." *Przeglad Antropologiczny* 17:193–211.

Frank, I. E., and Todeschini, T. (1994). *The Data Analysis Handbook*. New York: Elsevier.

Friedman, J. H. (1989). "Regularized Discriminant Analysis." *Journal of the American Statistical Association* 84:165–175.

Garthwaite, P. (1994). "An Interpretation of Partial Least Squares." *Journal of the American Statistical Association* 89:122–127.

Golub, G. H., and Kahan, W. (1965). "Calculating the singular values and pseudo-inverse of a matrix." *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis* 2:205–224.

Goodman, L. A. (1974). "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–231.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.

Hancock, G. R., and Mueller, R. O. (2001). "Rethinking Construct Reliability within Latent Variable Systems." In *Structural Equation Modeling: Present and Future – A Festschrift in Honor of Karl Jöreskog*, edited by R. Cudeck, S. du Toit, and D. Sörbom, 195–216. Lincolnwood, IL: Scientific Software International.

Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*. Cambridge, MA: MIT Press.

Harris, C. W., and Kaiser, H. F. (1964). "Oblique Factor Analytic Solutions by Orthogonal Transformation." *Psychometrika* 32:363–379.

Hartigan, J. A. (1981). "Consistency of Single Linkage for High–Density Clusters." *Journal of the American Statistical Association* 76:388–394.

Hastie, T., Tibshirani, R., and Friedman, J. H.(2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer Verlag.

Hoskuldsson, A. (1988). "PLS Regression Methods." *Journal of Chemometrics* 2:211–228.

Hoeffding, W. (1948). "A Non-Parametric Test of Independence." *Annals of Mathematical Statistics* 19:546–557.

Hu, L.-T., and Bentler, P. M. (1999). "Cutoff Criteria for Fit Indices in Covariance Structure Analysis: Conventional Criteria versus New Alternatives." *Structural Equation Modeling* 6:1–55.

Huber, P. J. (1964). "Robust Estimation of a Location Parameter." *Annals of Mathematical Statistics* 35:73–101.

Huber, P. J. (1973). "Robust Regression: Asymptotics, Conjecture, and Monte Carlo." *Annals of Statistics* 1:799–821.

Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics.* 2nd ed. Hoboken, NJ: John Wiley & Sons.

Jackson, J. E. (2003). *A User's Guide to Principal Components*. Hoboken, NJ: John Wiley & Sons.

Jardine, N., and Sibson, R. (1971). *Mathematical Taxonomy*. New York: John Wiley & Sons.

Jöreskog, K. G. (1977). "Factor Analysis by Least-Squares and Maximum Likelihood Methods." In *Statistical Methods for Digital Computers*, edited by K. Enslein, A. Ralston, and H. Wilf, 125 - 165. New York: John Wiley & Sons.

Kline, R. B. (2016). *Principles and Practice of Structural Equation Methodology*. 4th ed. New York: The Guilford Press.

Kohonen, T. (1989). *Self-Organization and Associative Memory*. 3rd ed. Vol. 8 of Springer Series in Information. Berlin: Springer-Verlag.

Kohonen, T. (1990). "The Self-Organizing Map." *Proceedings of the IEEE* 78:1464–1480.

Lindberg, W., Persson, J.-A., and Wold, S. (1983). "Partial Least-Squares Method for Spectrofluorimetric Analysis of Mixtures of Humic Acid and Ligninsulfonate." *Analytical Chemistry* 55:643–648.

LeRoux, B., and Rouanet, H. (2010). Multiple Correspondence Analysis. Vol.07–163 of Sage University Paper Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage Publications.

Mardia, K., Kent, J., and Bibby, J. (1980). *Multivariate Analysis.* New York: Academic Press.

Mason, R. L., and Young, J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications.* Philadelphia: SIAM.

Maydeu-Olivares, A., Shi, D., and Rosseel, Y. (2017). "Assessing Fit in Structural Equation Models: A Monte-Carlo Evaluation of RMSEA Versus SRMR Confidence Intervals and Tests of Close Fit." *Structural Equation Model: A Multidisciplinary Journal* 25:389–402.

McDonald, R. P. (1999). *Test Theory: A Unified Approach*. Mahwah, NJ: Erlbaum.

McLachlan, G. J., and Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: John Wiley & Sons.

McQuitty, L. L. (1957). "Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies." *Educational and Psychological Measurement* 17:207–229.

Milligan, G. W. (1980). "An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms." *Psychometrika* 45:325–342.

Nelson, P. R. C., Taylor, P. A., and MacGregor, J. F. (1996). "Missing Data Methods in PCA and PLS: Score calculations with incomplete observations." *Chemometrics and Intelligent Laboratory Systems* 35:45–65.

Nunnally, J. C. (1978). *Psychometric theory*. 2nd ed. New York: McGraw-Hill.

Penny, K. I. (1996). "Appropriate Critical Values When Testing for a Single Multivariate Outlier by Using the Mahalanobis Distance." *Journal of the Royal Statistical Society, Series C* 45:73–81.

Press, W. H, Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1998). *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Cambridge, England: Cambridge University Press.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). "Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices. *Psychological Methods* 21:137.

SAS Institute Inc. (1983). *SAS Technical Report A-108: Cubic Clustering Criterion*. Cary, NC: SAS Institute Inc. Accessed December 16, 2015. https://support.sas.com/kb/22/addl/fusion_22540_1_a108_5903.pdf.

SAS Institute Inc. (2020a). "The CALIS Procedure." *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/calis.pdf.

SAS Institute Inc. (2020b). "The CANDISC Procedure." *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/candisc.pdf.

SAS Institute Inc. (2020c). "The FACTOR Procedure." *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/factor.pdf.

SAS Institute Inc. (2020d). "The FASTCLUS Procedure." *SAS/STAT 15.2 User's Guide.* Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/fastclus.pdf.

SAS Institute Inc. (2020e). "The MIXED Procedure." *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/mixed.pdf.

SAS Institute Inc. (2020f). "The PLS Procedure." *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/pls.pdf.

SAS Institute Inc. (2020g). "The VARCLUS Procedure." *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/varclus.pdf.

Schafer, J., and Strimmer, K. (2005). "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics." *Statistical Applications in Genetics and Molecular Biology* 4 Article 32.

Sneath, P. H. A. (1957). "The Application of Computers to Taxonomy." *Journal of General Microbiology* 17:201–226.

Sokal, R. R., and Michener, C. D. (1958). "A Statistical Method for Evaluating Systematic Relationships." *University of Kansas Science Bulletin* 38:1409–1438.

Steiger, J. H. (1989). *EzPATH: A Supplementary Module for SYSTAT and SYGRAPH*. Evanston, IL: Systat.

Steiger, J. H. (1990). "Structural Model Evaluation and Modification: An Interval Estimation Approach." *Multivariate Behavioral Research* 25:173–180.

Tobias, R. D. (1995). "An Introduction to Partial Least Squares Regression." In *Proceedings of the Twentieth Annual SAS Users Group International Conference*, 1250–1257. Cary, NC: SAS Institute Inc. http://www.sascommunity.org/sugi/SUGI95/Sugi-95-210%20Tobias.pdf.

Tracy, N. D., Young, J. C., and Mason, R. R. (1992). "Multivariate Control Charts for Individual Observations." *Journal of Quality Technology* 24:88–95.

Umetrics. (1995). *Multivariate Analysis (3-day course)*. Winchester, MA.

Waern, Y. (1972). "Structure in Similarity Matrices: A Graphic Approach." *Scandinavian Journal of Psychology* 13:5–16.

West, S. G., Taylor, A. B., and Wu, W. (2012). "Model Fit and Model Selection in Structural Equation Modeling." In *Handbook of Structural Equation Modeling*, edited by R. H. Hoyle, 209–231. New York: The Guilford Press.

White, K. P., Jr., Kundu, B., and Mastrangelo, C. M. (2008). "Classification of Defect Clusters on Semiconductor Wafers Via the Hough Transform." *IEEE Transactions on Semiconductor Manufacturing* 21:272–278.

Wold, S. (1994). "PLS for Multivariate Linear Modeling." In *QSAR: Chemometric Methods in Molecular Design. Methods and Principles in Medicinal Chemistry*, edited by H. van de Waterbeemd, pp. 195–218. Weinheim, Germany: Verlag-Chemie.

Wold, S., Sjostrom, M., and Eriksson, L. (2001). "PLS-Regression: A Basic Tool of Chemometrics." *Chemometrics and Intelligent Laboratory Systems* 58:109–130.

부록 C

기술 라이선스 고지 사항

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

버전 16

품질 및 공정 방법

"진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는 것이 아니라 새로운 눈을 가지는 데 있다."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 16 품질 및 공정 방법. Cary, NC: SAS Institute Inc.

JMP® 16 품질 및 공정 방법

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

2021년 3월

2021년 8월

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

JMP 활용하기

JMP를 처음 사용하든 오랫동안 사용해왔든, JMP와 관련해서 배워야 하는 것은 언제나 있습니다.

JMP.com을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다.

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티, 추가 기능 및 스크립트 예를 포함한 사용자용 리소스, 포럼, 블로그, 컨퍼런스 정보 등

https://www.jmp.com/getstarted

목차
품질 및 공정 방법

1 JMP 알아보기 13
설명서 및 추가 리소스
JMP 설명서에 사용되는 서식 규칙 15
JMP 도움말 16
JMP 설명서 라이브러리 16
JMP 학습을 위한 추가 리소스 22
JMP 자습서 22
샘플 데이터 테이블 22
통계 및 JSL 용어에 대해 알아보기 23
JMP 팁 및 힌트 알아보기 23
JMP 툴팁 23
JMP 사용자 커뮤니티 23
무료 온라인 Statistical Thinking 교육 과정 24
JMP 새로운 사용자를 위한 입문 키트 24
Statistics Knowledge 포털 24
JMP 교육 과정 24
사용자가 작성한 JMP 설명서 24
JMP 시작하기 창 24
JMP 기술 지원 25

2 품질 및 공정 방법 소개 27
공정 및 제품 개선을 위한 도구

3 관리도 빌더 29
대화식으로 관리도 생성
관리도 빌더 개요 32
관리도 빌더의 예 32
관리도 유형 34
Shewhart 계량형 관리도 34
Shewhart 계수형 관리도 35
희귀 사건 관리도 36
관리도 유형 37
관리도 빌더 시작 39

관리도 빌더 대화식 작업 영역 40
특정 관리도의 시작 창 42
관리도 빌더 창 43
관리도 빌더 옵션 45
관리도 빌더의 빨간색 삼각형 메뉴 옵션 45
옵션 패널 및 마우스 오른쪽 버튼 클릭 차트 옵션 48
관리도 빌더의 마우스 오른쪽 버튼 클릭 축 옵션 57
관리 한계 작업 58
관리 한계의 예 58
관리도 빌더에서 제외 및 숨겨진 표본 66
관리도 빌더의 추가 예 67
개별 측정값 및 이동 범위 차트 예제 67
Xbar-R 차트 단계 예제 69
부분군 크기가 다른 Xbar-S 차트 예제 72
런 차트 예제 74
P 차트 예제 75
NP 차트 예제 76
C 차트 예제 78
U 차트 예제 79
G 차트 예제 80
T 차트 예제 81
삼원 관리도 예제 83
여러 관리도의 예 85
관리도 빌더에 대한 통계 상세 정보 85
Xbar-R 차트에 대한 관리 한계 86
Xbar-S 차트에 대한 관리 한계 87
개별 측정값 및 이동 범위 차트에 대한 관리 한계 87
P 및 NP 차트에 대한 관리 한계 89
U 차트 및 C 차트에 대한 관리 한계 89
Levey-Jennings 차트 90
G 차트에 대한 관리 한계 90
T 차트에 대한 관리 한계 91
삼원 관리도에 대한 시그마 계산 92

4 측정 시스템 분석 95
EMP 방법을 사용하여 연속형 측정 공정 평가

측정 시스템 분석 개요 97
측정 시스템 분석의 예 97
측정 시스템 분석 플랫폼 시작 100
데이터 형식 101

측정 시스템 분석 플랫폼 옵션 102
평균 차트 103
범위 차트 또는 S 관리도 104
EMP 결과 104
유효 해상도 106
변화 감지 프로파일러 106
편향 비교 111
검사 - 재검사 오차 비교 111
측정 시스템 분석의 추가 예 112
측정 시스템 분석에 대한 통계 상세 정보 118
급내 상관 및 확률 오차 계산 118

5 계량형 게이지 차트 121
게이지 R&R 을 사용하여 연속형 측정 공정 평가

계량형 차트 개요 123
계량형 차트의 예 124
계량형 / 계수형 게이지 차트 플랫폼 시작 125
데이터 형식 126
계량형 게이지 차트 126
계량형 게이지 플랫폼 옵션 128
이분산성 검정 130
분산 성분 130
게이지 R&R 방법 정보 132
게이지 R&R 옵션 133
판별비 135
오분류 확률 135
편향 보고서 136
선형성 연구 136
계량형 차트의 추가 예 137
이분산성 검정의 예 137
편향 보고서 옵션의 예 139
계량형 차트에 대한 통계 상세 정보 142
분산 성분에 대한 통계 상세 정보 142
판별비에 대한 통계 상세 정보 143
오분류 확률에 대한 통계 상세 정보 144

6 계수형 게이지 차트 145
합치도 측도를 사용하여 범주형 측정 공정 평가

계수형 게이지 차트 개요 147
계수형 게이지 차트의 예 147

계량형 / 계수형 게이지 차트 플랫폼 시작 149
계수형 게이지 차트 및 보고서 150
합치도 보고서 151
유효성 보고서 152
계수형 게이지 플랫폼 옵션 153
계수형 게이지 차트에 대한 통계 상세 정보 154
합치도 보고서에 대한 통계 상세 정보 155

7 공정 능력 159
시간 경과에 따른 공정의 변동성 측정

공정 능력 플랫폼 개요 161
정규 변수가 있는 공정 능력 플랫폼의 예 163
비정규 변수가 있는 공정 능력 플랫폼의 예 165
공정 능력 플랫폼 시작 171
공정 선택 172
공정 부분군 지정 172
이동 범위 옵션 173
과거 정보 173
분포 옵션 174
기타 규격 175
규격 한계 입력 175
규격 한계 창 175
한계 데이터 테이블 176
규격 한계 열 특성 177
공정 능력 보고서 178
목표 그림 179
공정 능력 상자 그림 181
공정 능력 지수 그림 183
공정 능력 플랫폼 옵션 185
개별 상세 정보 보고서 188
정규화된 상자 그림 194
공정 성능 그림 195
요약 보고서 196
목표 그림 요약 테이블 생성 198
공정 능력 플랫폼의 추가 예 199
안정적인 공정에 대한 공정 능력 199
불안정한 공정에 대한 공정 능력 202
비정규 공정 Ppk 에 대한 신뢰 한계 시뮬레이션 206
공정 능력 플랫폼에 대한 통계 상세 정보 213
변동 통계량 213

목표 그림 및 공정 능력 상자 그림에 대한 표기 216
목표 그림 217
목표값이 결측인 공정에 대한 공정 능력 상자 그림 218
정규 분포에 대한 공정 능력 지수 219
비정규 분포에 대한 공정 능력 지수 : 백분위수 및 Z 스코어 방법 224
분포의 파라미터화 225

8 CUSUM 관리도 229
결정 한계가 포함된 테이블 형식 CUSUM 관리도 생성
CUSUM 관리도 플랫폼 개요 231
CUSUM 관리도의 예 231
CUSUM 관리도 플랫폼 시작 233
CUSUM 관리도 플랫폼 보고서 234
CUSUM 관리도의 제어판 234
CUSUM 차트 235
CUSUM 관리도 플랫폼 옵션 236
평균 런 길이 (ARL) 보고서 238
CUSUM 관리도의 추가 예 239
CUSUM 관리도의 데이터 단위 옵션 예 239
부분군이 있는 CUSUM 차트의 예 240
CUSUM 관리도 플랫폼에 대한 통계 상세 정보 241
CUSUM 관리도 생성에 대한 통계 상세 정보 241
변화 감지에 대한 통계 상세 정보 242
평균 런 길이에 대한 통계 상세 정보 242

9 EWMA 관리도 245
지수 가중 관리도 생성
EWMA 관리도 플랫폼 개요 247
EWMA 관리도 플랫폼의 예 247
EWMA 관리도 플랫폼 시작 249
EWMA 관리도 플랫폼 보고서 250
EWMA 관리도의 제어판 250
EWMA 차트 보고서 251
EWMA 관리도 플랫폼 옵션 252
EWMA 관리도의 평균 런 길이 (ARL) 보고서 255
EWMA 관리도 플랫폼의 추가 예 256
EWMA 관리도 플랫폼에 대한 통계 상세 정보 259

10 다변량 관리도 263
여러 공정 특성 동시 모니터링
다변량 관리도 개요 265

다변량 관리도의 예 265
다변량 관리도 플랫폼 시작 268
다변량 관리도 269
다변량 관리도 플랫폼 옵션 270
분할된 T 제곱 271
변화 지점 감지 271
주성분 272
다변량 관리도의 추가 예 273
부분군으로 묶은 데이터를 사용한 공정 모니터링의 예 273
분할된 T 제곱의 예 276
변화 지점 감지의 예 278
다변량 관리도에 대한 통계 상세 정보 280
개별 관측값에 대한 통계 상세 정보 280
합리적 부분군의 관측값에 대한 통계 상세 정보 281
변화 지점 감지에 대한 통계 상세 정보 284

11 모형 기반 다변량 관리도 287
복합 공정 모니터링 및 진단

모형 기반 다변량 관리도 개요 289
모형 기반 다변량 관리도의 예 289
모형 기반 다변량 관리도 플랫폼 시작 292
데이터 형식 292
모형 기반 다변량 관리도 보고서 293
모형 기반 다변량 관리도 플랫폼 옵션 294
그림 옵션 295
스코어 그림 296
모형 기반 다변량 관리도 플랫폼의 추가 예 297
과거 데이터를 사용하는 MDMVCC 의 예 297
PLS 모형을 사용하는 MDMVCC 의 예 298
모형 기반 다변량 관리도 플랫폼에 대한 통계 상세 정보 299
모니터링 통계량 299
한계 301
기여도 303
스코어 그림 그룹 비교 306
스코어 그림 적재 306

12 레거시 관리도 307
계량형 및 계수형 관리도 생성

레거시 관리도의 예 309
레거시 관리도 유형 310

계량형 관리도 310
계수형 관리도 313
레거시 관리도 플랫폼 시작 314
공정 정보 315
한계 규격 317
지정된 통계량 317
레거시 관리도 보고서 318
V- 마스크 CUSUM 차트 보고서 320
양측 V-마스크 CUSUM 차트 해석 320
단측 CUSUM 차트 해석 322
레거시 관리도 플랫폼 옵션 322
레거시 관리도에 대한 창 옵션 322
레거시 관리도에 대한 차트 옵션 325
V- 마스크 CUSUM 관리도에 대한 차트 옵션 327
한계 저장 및 검색 328
제외 , 삭제 및 숨겨진 표본 331
관리도 플랫폼의 추가 예 333
사전 집계 차트 예제 333
V- 마스크 CUSUM 차트 예제 335
단측 CUSUM 차트 예제 337
UWMA 차트 예제 338
관리도 플랫폼에 대한 통계 상세 정보 339
중앙값 이동 범위 차트에 대한 관리 한계 339
공정 능력 분석에 대한 통계 상세 정보 339
V- 마스크 CUSUM 관리도에 대한 통계 상세 정보 345
가중 이동 평균 차트에 대한 통계 상세 정보 348

13 파레토도 349
근본 원인을 개선하기 위한 노력에 집중
파레토도 플랫폼 개요 351
파레토도 플랫폼의 예 351
파레토도 플랫폼 시작 354
파레토도 보고서 355
파레토도 플랫폼 옵션 356
원인 옵션 357
파레토도 플랫폼의 추가 예 358
결합 원인의 임계 예제 358
그룹 전체에서 상수 크기 사용 예제 360
그룹 전체에서 상수가 아닌 표본 크기 사용 예제 361
일원 비교 파레토도 예제 363

이원 비교 파레토도 예제 364
파레토도 플랫폼에 대한 통계 상세 정보 366
가능도비 카이제곱 검정 366

14 특성 요인도 367
근본 원인 식별
특성 요인도 개요 369
특성 요인도의 예 369
데이터 준비 370
다이어그램 플랫폼 시작 370
특성 요인도 371
마우스 오른쪽 버튼 클릭 메뉴 371
특성 요인도 메뉴 옵션 374
다이어그램 저장 374
다이어그램을 데이터 테이블로 저장 374
다이어그램을 저널로 저장 375
다이어그램을 스크립트로 저장 375

15 품질 유틸리티 377
규격 한계 관리 및 OC 곡선 생성
규격 한계 관리 유틸리티 379
규격 한계 관리 유틸리티의 예 379
규격 한계 관리 옵션 382
검사 특성 곡선 유틸리티 383
OC 곡선 유틸리티 시작 383
관리도에 대한 OC 곡선 옵션 383
합격 표집에 대한 OC 곡선 옵션 384

A 참조 자료 385

B 기술 라이선스 고지 사항 389

1 장

JMP 알아보기
설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙 15
JMP 도움말 16
JMP 설명서 라이브러리 16
JMP 학습을 위한 추가 리소스 22
JMP 자습서 22
샘플 데이터 테이블 22
통계 및 JSL 용어에 대해 알아보기 23
JMP 팁 및 힌트 알아보기 23
JMP 툴팁 23
JMP 사용자 커뮤니티 23
무료 온라인 Statistical Thinking 교육 과정 24
JMP 새로운 사용자를 위한 입문 키트 24
Statistics Knowledge 포털 24
JMP 교육 과정 24
사용자가 작성한 JMP 설명서 24
JMP 시작하기 창 24
JMP 기술 지원 25

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다 .

- 샘플 데이터 테이블 이름 , 열 이름 , 경로 이름 , 파일 이름 , 파일 확장자 및 폴더는 Helvetica(또는 sans-serif online) 글꼴로 표시됩니다 .
- 코드는 `Lucida Sans Typewriter`(또는 monospace online) 글꼴로 표시됩니다 .
- 코드 출력은 Lucida Sans Typewriter 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다 .
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다 .
  - 버튼
  - 체크박스
  - 명령
  - 선택 가능한 목록 이름
  - 메뉴
  - 옵션
  - 탭 이름
  - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다 .
  - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
  - 설명서 제목
  - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 https://www.jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다 .

팁 : 팁 섹션에는 유용한 정보가 표시됩니다 .

JMP 도움말

" 도움말 " 메뉴의 "JMP 도움말 " 에서는 JMP 기능 , 통계적 방법 및 JSL(JMP Scripting Language) 에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다 .
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다 .
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다 . **도구** 메뉴에서 도움말 도구 를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다 .
- JMP 창 내에서 **도움말** 버튼을 클릭합니다 .

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다 . 자세한 내용은 JMP 설명서 라이브러리에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 https://www.jmp.com/documentation 에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다 .

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
품질 및 공정 방법	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 https://jmp.com/tools에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 " 도움말 > 통계 분석 인덱스 " 를 선택합니다 . JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다 .

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다 .

스크립트 인덱스 JSL 함수 , 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다 . " 스크립트 인덱스 " 에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다 .

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨) 이 제공됩니다 .

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁 : Windows 의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다 . **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다 . 이 옵션은 macOS 에서는 사용할 수 없습니다 .

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 https://community.jmp.com을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 https://www.jmp.com/statisticalthinking에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 https://www.jmp.com/welcome에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간략한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 https://www.jmp.com/skp에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 https://www.jmp.com/training에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 https://www.jmp.com/books 에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기 " 창을 살펴보십시오 . 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다 . "JMP 시작하기 "

창에는 " 분석 ", " 그래프 ", " 테이블 " 및 " 파일 " 메뉴에 있는 다양한 옵션이 포함됩니다 . 또한 이 창에는 JMP Pro 의 기능 및 플랫폼도 나열됩니다 .

- "JMP 시작하기 " 창을 열려면 **보기** (macOS 의 경우 **창**) **> JMP 시작하기**를 선택합니다 .
- Windows 에서 JMP 를 열 때 자동으로 "JMP 시작하기 " 를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 " 초기 JMP 창 " 목록에서 **JMP 시작하기**를 선택합니다 . macOS 에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다 .

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP 의 교육을 받은 엔지니어가 제공하며 , 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다 .

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 https://www.jmp.com/support 에서 제공됩니다 .

2 장

품질 및 공정 방법 소개
공정 및 제품 개선을 위한 도구

품질 및 공정 방법에서는 다음과 같이 품질 및 공정 성능을 평가하고 개선하기 위해 JMP에서 사용할 수 있는 다양한 방법과 도구에 대해 설명합니다.

- 관리도는 주요 변수에 대한 피드백을 제공하고 공정이 통계적 관리 상태이거나 관리 이탈일 때 이를 표시합니다. "관리도 빌더"에서는 관리도 빌더라고 하는 대화식 관리도 플랫폼을 사용하여 관리도를 생성하는 JMP 접근 방식에 대해 설명합니다.
- 측정 시스템 분석 플랫폼은 시스템의 정밀도, 일관성 및 편향을 평가합니다. 공정을 연구하려면 먼저 공정을 정확하고 정밀하게 측정할 수 있는지 확인해야 합니다. 만일 변동의 대부분이 측정 공정 자체에 기인하는 경우 공정에 대해 신뢰할 수 있는 정보를 얻지 못합니다. 이 분석을 통해 시스템 성능을 확인할 수 있습니다. 자세한 내용은 "측정 시스템 분석"에서 확인하십시오.
- 계량형 / 계수형 게이지 차트 플랫폼은 계량형 또는 계수형 게이지 차트를 생성합니다. 계량형 차트는 연속형 측정값을 분석하여 시스템 성능을 표시합니다. 계수형 차트는 범주형 측정값을 분석하여 반응 간의 합치 측도를 보여 줍니다. 게이지 연구를 수행하여 데이터의 변동 측도를 확인할 수도 있습니다. 자세한 내용은 "계량형 게이지 차트" 및 "계수형 게이지 차트"에서 확인하십시오.
- 공정 능력 플랫폼은 규격 한계를 충족하는 공정 능력을 측정합니다. 공정 중심 및 변동성별로 요약된 공정 성능을 규격 한계와 비교할 수 있습니다. 이 플랫폼은 장기 변동과 단기 변동을 둘 다 기준으로 하여 공정 능력 지수를 계산합니다. 분석을 수행하여 규격을 기준으로 변동을 식별할 수 있으며, 이로 인해 점점 더 높은 적합 값을 얻을 수 있습니다. 자세한 내용은 "공정 능력"에서 확인하십시오.
- CUSUM 차트를 사용하면 누적합을 기반으로 의사 결정을 내릴 수 있습니다. 이 차트를 통해 공정의 작은 변화를 감지할 수 있습니다. 자세한 내용은 "CUSUM 관리도"에서 확인하십시오.
- EWMA(지수 가중 이동 평균) 차트를 사용하여 공정의 작은 변화를 감지할 수도 있습니다. 자세한 내용은 "EWMA 관리도"에서 확인하십시오.
- 여러 공정 특성을 동시에 모니터링해야 하는 경우 "다변량 관리도"의 내용을 참조하십시오.
- MDMVCC(모형 기반 다변량 관리도) 플랫폼에서는 주성분 또는 부분 최소 제곱 모형을 기반으로 관리도를 생성할 수 있습니다. 자세한 내용은 "모형 기반 다변량 관리도"에서 확인하십시오.
- "레거시 관리도"에서는 JMP 의 이전 관리도 플랫폼에 대해 설명합니다. 이러한 플랫폼을 사용하지 않고 새로 추가된 CUSUM 및 EWMA 관리도 플랫폼과 함께 관리도 빌더 플랫폼을 사용하는 것이 좋습니다.

- 파레토도 플랫폼은 품질 관련 공정 또는 작업에서 문제의 빈도를 보여 줍니다 . 파레토도는 문제의 빈도와 심각도를 강조 표시하여 어떤 문제를 먼저 해결해야 할지 결정하는 데 도움이 됩니다 . 자세한 내용은 " 파레토도 " 에서 확인하십시오 .
- 다이어그램 플랫폼은 브레인스토밍을 위해 또는 추가 실험의 변수를 식별하기 위한 예비 분석 용도로 문제의 원인을 구성하는 특성 요인도를 생성합니다. 완료되면 추가 분석을 수행하여 문제의 근본 원인을 식별할 수 있습니다. 자세한 내용은 "특성 요인도"에서 확인하십시오.
- 규격 한계 관리 유틸리티를 사용하면 여러 열에 대한 많은 규격 한계를 한 번에 빠르게 추가하거나 편집할 수 있습니다 . 자세한 내용은 " 규격 한계 관리 유틸리티 " 에서 확인하십시오 .
- OC(검사 특성) 곡선 유틸리티를 사용하면 관리도 및 계수형 합격 표집 계획에 대한 OC 곡선을 생성할 수 있습니다 . 자세한 내용은 " 검사 특성 곡선 유틸리티 " 에서 확인하십시오 .

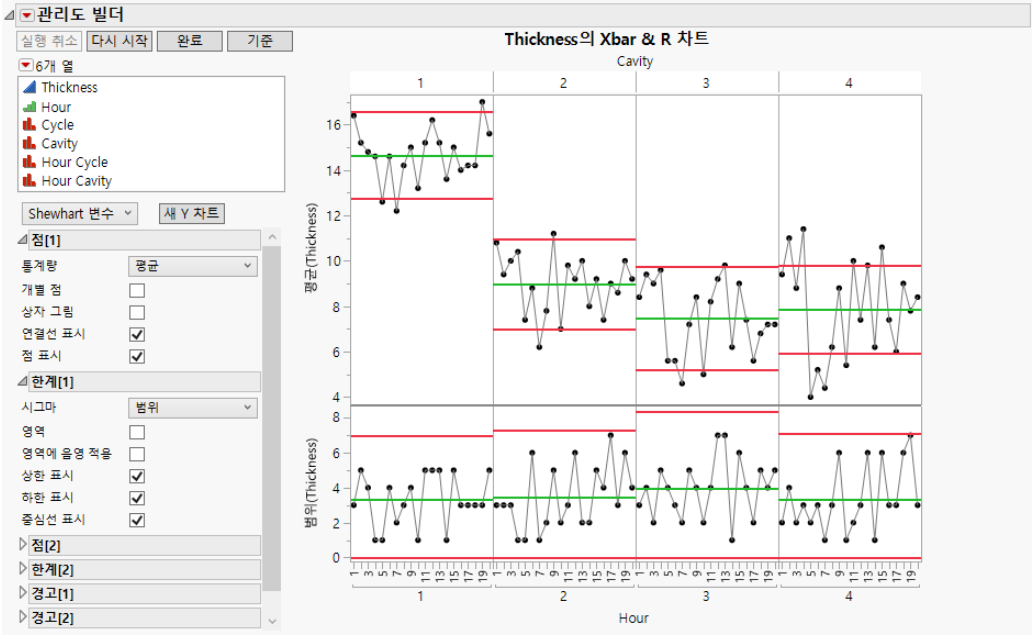
3 장

관리도 빌더
대화식으로 관리도 생성

관리도는 공정 변동을 모니터링하기 위한 그래픽 및 분석 도구입니다. 공정 자체에서 발생하는 고유 변동은 관리 한계 집합을 사용하여 정량화할 수 있습니다. 관리 한계는 우연 원인 변동과 특수 원인 변동을 구별하는 데 도움이 됩니다. 일반적으로 특수 원인 변동을 식별 및 제거하기 위한 작업이 수행됩니다. 공정의 우연 원인 변동을 정량화하는 것도 중요합니다. 이는 공정 능력을 결정하기 때문입니다.

관리도 빌더를 사용하여 공정 데이터의 관리도를 생성합니다. 관리도 빌더는 대화식 작업 영역으로 시작하거나 특정 관리도 메뉴 옵션에서 시작할 수 있습니다. 대화식 작업 영역에서 차트를 작성할 변수를 선택하여 영역으로 드래그합니다. JMP에서 해당 데이터를 기반으로 적절한 차트 유형이 자동으로 선택됩니다. 다른 유형의 차트를 빠르게 생성하거나 기존 차트의 현재 설정을 변경할 수 있습니다.

그림 3.1 관리도 빌더의 예



목차

관리도 빌더 개요 32
관리도 빌더의 예 32
관리도 유형 34
Shewhart 계량형 관리도 34
Shewhart 계수형 관리도 35
희귀 사건 관리도 36
관리도 유형 37
관리도 빌더 시작 39
관리도 빌더 대화식 작업 영역 40
특정 관리도의 시작 창 42
관리도 빌더 창 43
관리도 빌더 옵션 45
관리도 빌더의 빨간색 삼각형 메뉴 옵션 45
옵션 패널 및 마우스 오른쪽 버튼 클릭 차트 옵션 48
관리도 빌더의 마우스 오른쪽 버튼 클릭 축 옵션 57
관리 한계 작업 58
관리 한계의 예 58
관리도 빌더에서 제외 및 숨겨진 표본 66
관리도 빌더의 추가 예 67
개별 측정값 및 이동 범위 차트 예제 67
Xbar-R 차트 단계 예제 69
부분군 크기가 다른 Xbar-S 차트 예제 72
런 차트 예제 74
P 차트 예제 75
NP 차트 예제 76
C 차트 예제 78
U 차트 예제 79
G 차트 예제 80
T 차트 예제 81
삼원 관리도 예제 83
여러 관리도의 예 85
관리도 빌더에 대한 통계 상세 정보 85
Xbar-R 차트에 대한 관리 한계 86
Xbar-S 차트에 대한 관리 한계 87
개별 측정값 및 이동 범위 차트에 대한 관리 한계 87
P 및 NP 차트에 대한 관리 한계 89
U 차트 및 C 차트에 대한 관리 한계 89
Levey-Jennings 차트 90

G 차트에 대한 관리 한계 90
T 차트에 대한 관리 한계 91
삼원 관리도에 대한 시그마 계산 92

관리도 빌더 개요

관리도는 공정 변동을 모니터링하고 공정의 특수 원인 변동을 식별하기 위한 그래픽 및 분석 도구입니다. 관리 한계를 설정하여 일상적 변동을 필터링하면 공정이 안정적이고 예측 가능한지 여부를 판별하는 데 도움이 됩니다. 공정의 변동이 필요 이상인 경우 공정을 조정하여 잠재적인 비용 절감 효과와 함께 더 높은 품질의 결과를 생성할 수 있습니다.

공정은 시간에 따라 모니터링되므로 모든 공정에서 측정 변동을 보입니다 . 공정 측정에는 다음과 같은 두 가지 유형의 변동이 있습니다 .

- 일상적 또는 우연 원인 변동 . 안정적인 공정에서 얻은 측정값도 랜덤 변동을 보입니다 . 공정 측정 결과 우연 원인 변동만 나타나는 경우 측정값은 예상 한계 내에 유지됩니다 .
- 비정상 또는 특수 원인 변동 . 특수 원인 변동은 관리도에서 관측된 패턴으로 표시됩니다 . 공정 평균의 변화 , 관리 한계 위 / 아래의 점 또는 위 / 아래로 추세를 나타내는 측정값 등을 예로 들 수 있습니다 . 공정 측정의 이러한 변화는 도구 또는 기계 파손 , 장비 성능 저하 , 원재료 변경과 같은 요인이 원인일 수 있습니다 . 공정의 변화 또는 결함은 대부분 공정 측정의 비정상적 변동에 의해 식별할 수 있습니다 .

관리도 빌더를 사용하면 Shewhart 및 희귀 사건 관리도를 포함하여 여러 유형의 관리도를 생성할 수 있습니다 . Shewhart 관리도는 크게 계량형 관리도와 계수형 관리도로 분류됩니다 . 희귀 사건 차트는 드물게 발생하는 사건에 사용하기 위한 용도로 만들어졌습니다 . JMP 에서는 관리도 생성을 위해 유연한 사용자 정의 방식을 제공합니다 . 다음과 같은 방법으로 관리도를 생성할 수 있습니다 .

- 대화식 관리도 빌더 작업 영역을 사용합니다. 데이터 열을 작업 영역으로 드래그하면 관리도 빌더가 데이터 유형 및 표본 크기를 기반으로 적절한 차트를 생성합니다 .
- 관리도 메뉴 옵션을 사용하여 시작 창으로 특정 관리도를 생성합니다 .

위의 방법 중 하나를 통해 초기 차트가 생성되면 메뉴 및 기타 옵션을 사용하여 차트 유형 변경, 차트의 통계량 변경, 차트 형식 조정 또는 차트 추가 등을 수행할 수 있습니다.

관리도 빌더의 예

Socket Thickness.jmp 샘플 데이터 테이블에는 소켓 두께 측정값이 포함되어 있습니다 . 생산 중 결함 수가 증가했으며 이 문제의 발생 원인을 조사하려고 합니다 . 이 예에서는 대화식 방식 또는 시작 창 방식 중 하나를 사용하여 관리도 빌더에서 조사를 수행하는 방법을 보여 줍니다 . 생성할 관리도 유형을 알고 있으면 두 번째 방식이 편리합니다 .

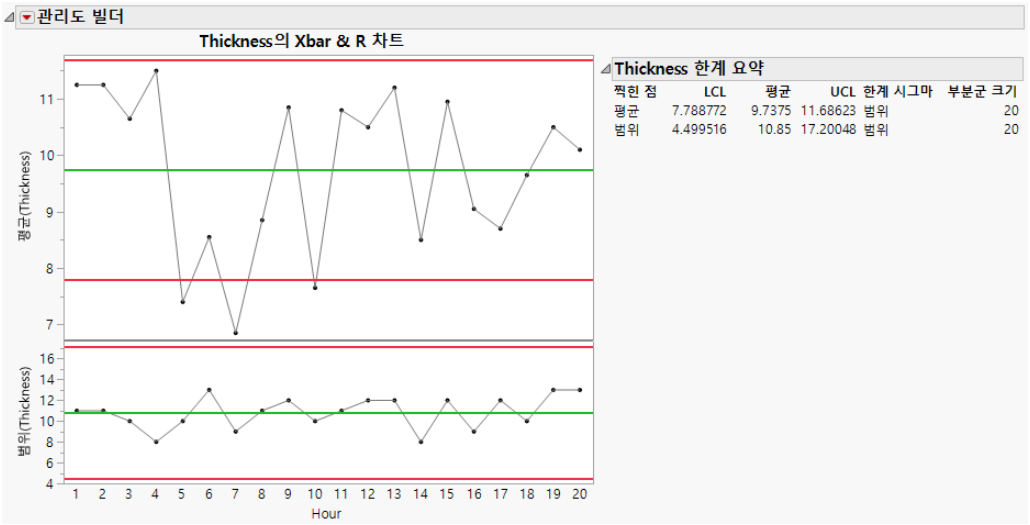
관리도 빌더 대화식 방법

대화식 관리도 빌더 작업 영역을 사용하여 공정 데이터의 변동을 조사할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Socket Thickness.jmp를 엽니다.
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .

3. Thickness 를 **Y** 영역으로 드래그합니다 .
4. Hour 를 **부분군** 영역 (아래쪽) 으로 드래그합니다 .

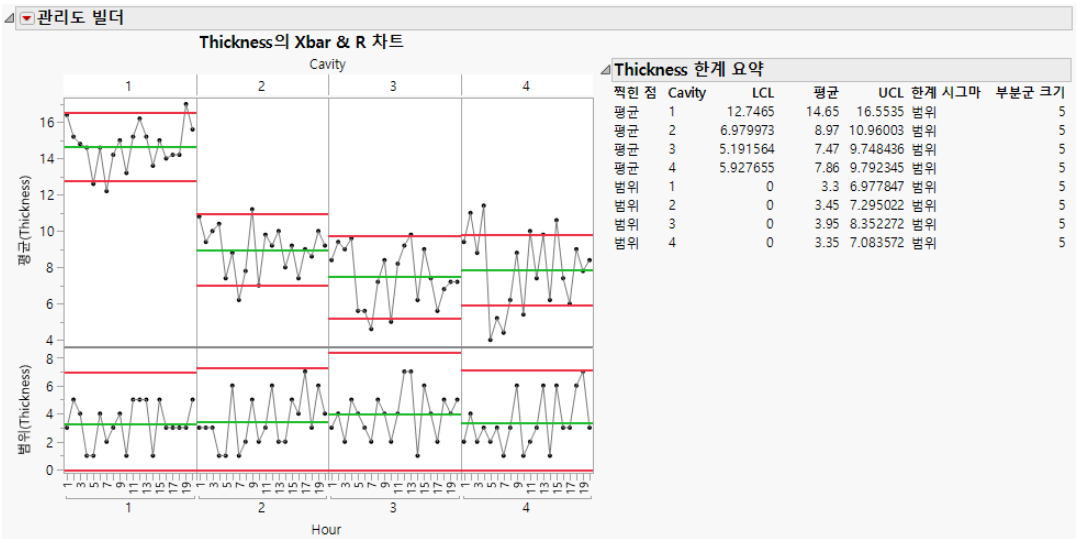
그림 3.2 Thickness(소켓 두께) 에 대한 관리도



" 평균 " 차트를 보면 관리 하한인 7.788772 아래에 점이 여러 개 있음을 알 수 있습니다 . 다른 변수가 문제의 원인이 될 수 있는지 확인하려고 합니다 .

5. Cavity 를 **단계** 영역으로 드래그합니다 .
6. **완료**를 클릭합니다 .

그림 3.3 각 Cavity(캐비티) 에 대한 관리도



" 평균 " 차트에서 다음과 같은 결론을 내릴 수 있습니다 .

- 캐비티 간에 차이가 있으므로 각 캐비티에 대해 별도의 관리 한계가 필요합니다 .
- Cavity 1 은 다른 캐비티보다 평균 두께 값이 큰 소켓을 생산합니다 . 이는 캐비티 간의 차이에 대한 추가 조사가 필요하다는 것을 의미합니다 .
- 모든 캐비티에 관리 한계를 벗어나는 점이 있습니다 . 따라서 각 캐비티에 대한 공정의 관리 결여를 조사해야 합니다 .

각 캐비티에 대한 " 범위 " 차트에서는 부분군 내 측정값이 관리 상태이며 캐비티 전체에서 비슷하다는 것을 보여 줍니다.

관리도 빌더 시작 창 방법

"XBar 관리도 " 시작 창을 사용하여 그림 3.3 과 동일한 차트를 얻을 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Socket Thickness.jmp를 엽니다.
2. **분석 > 품질 및 공정 > 관리도 > XBar 관리도**를 선택합니다 .
3. Thickness 를 선택하고 **Y** 를 클릭합니다 .
4. Hour 를 선택하고 **부분군**을 클릭합니다 .
5. Cavity 를 선택하고 **단계**를 클릭합니다 .
6. **확인**을 클릭합니다 .
   그림 3.3 과 동일한 관리도를 볼 수 있습니다 .

관리도 유형

관리도 빌더를 사용하면 Shewhart 계량형, Shewhart 계수형 및 희귀 사건 차트를 포함하여 여러 유형의 관리도를 생성할 수 있습니다.

- "Shewhart 계량형 관리도 "
- "Shewhart 계수형 관리도 "
- " 희귀 사건 관리도 "
- " 관리도 유형 "

Shewhart 계량형 관리도

계량형 관리도는 차트에 표시된 부분군 요약 통계량에 따라 분류됩니다 .

- XBar 차트는 부분군 평균을 표시하는 위치 차트의 한 유형입니다 .
- R 차트는 부분군 범위 (최대값 – 최소값) 를 표시하는 산포 차트의 한 유형입니다 .
- S 차트는 부분군 표준편차를 표시하는 산포 차트의 한 유형입니다 .

- 개별 측정값 차트는 개별 측정값을 표시하는 위치 차트의 한 유형입니다 .
- 런 차트는 개별 측정값을 연결된 일련의 점으로 표시하는 위치 차트의 한 유형입니다 .
- 사전 집계 차트에는 부분군 평균과 표준편차가 모두 표시됩니다 .
- 이동 범위 차트는 연속된 두 측정값의 이동 범위를 표시하는 산포 차트의 한 유형입니다 .

참고: 산포 차트를 제거하거나, 파일 > 환경 설정 > 플랫폼 > 관리도 빌더에서 "두 개의 Shewhart 차트 표시" 옵션을 해제하면 위치 차트만 표시됩니다. 연결된 스크립트에는 Off(0)로 설정된 JSL 옵션 "Show Two Shewhart Charts" 가 포함됩니다 .

XBar, R 및 S 차트

연속형 척도로 측정된 품질 특성에 대해 일반적인 분석을 수행하면 해당하는 R 또는 S 차트 위에 평균 차트를 정렬하여 공정 평균과 변동성이 모두 표시됩니다.

개별 측정값 차트

개별 측정값 차트에는 개별 측정값이 표시됩니다. 각 표집 시점에 대해 하나의 측정값만 사용 가능한 경우 개별 측정값 차트가 적합합니다. 개별 측정값을 차트에 표시하는 경우 개별 측정값 차트는 해당 이동 범위 차트 위에 표시됩니다. 이동 범위 차트에는 연속된 두 측정값의 이동 범위가 표시됩니다.

사전 집계 차트

데이터가 동일한 공정 단위의 반복 측정값으로 구성된 경우 이를 해당 단위에 대한 하나의 측정값으로 결합할 수 있습니다. 데이터에 각 공정 또는 측정 단위에 대한 반복 측정값이 포함된 경우 외에는 사전 집계하지 않는 것이 좋습니다.

사전 집계는 선택한 표본 크기 또는 표본 라벨을 기반으로 공정 열을 표본 평균 및/또는 표준편차로 요약합니다. 그런 다음 창에서 선택한 옵션에 따라 요약 데이터를 차트에 표시합니다.

Levey-Jennings 차트

Levey-Jennings 차트는 장기 시그마를 기반으로 한 관리 한계와 공정 평균을 보여 줍니다. 관리 한계는 중심선에서 3*s* 떨어진 위치에 배치됩니다. Levey-Jennings 차트에 대한 표준편차 *s* 는 분포 플랫폼의 표준편차와 동일한 방식으로 계산됩니다.

Shewhart 계수형 관리도

계수형 차트는 개수 데이터에 적용할 수 있습니다. 계수형 차트는 이항 및 Poisson 모형을 기반으로 합니다. 개수는 부분군별로 측정되므로 여러 차트를 비교할 때 차트 간에 부분군의 항목 수가 비슷한지 확인해야 합니다. 계수형 차트는 계량형 차트와 마찬가지로 차트에 표시된 부분군 표본 통계량에 따라 분류됩니다.

표 3.1 계수형 차트 결정

시그마 계산에 사용되는 분포	통계량 유형 : 비율	통계량 유형 : 개수
이항	P 차트	NP 차트
Poisson	U 차트	C 차트

관리도 빌더는 선택된 변수를 기반으로 몇 가지 결정을 내립니다 . 기본 차트가 생성되면 메뉴 및 기타 옵션을 사용하여 차트의 유형 , 통계량 , 형식 등을 변경할 수 있습니다 .

- P 차트에는 부분군 크기가 변동 가능한 표본에서 부적합품 (불량) 의 비율이 표시됩니다 . P 차트의 각 부분군은 N_i개 항목으로 구성되며 각 항목이 적합 또는 부적합으로 평가되므로 부분군의 최대 부적합품 수는 N_i 개입니다 .
- NP 차트에는 부분군 표본의 부적합품 (불량) 수가 표시됩니다 . NP 차트의 각 부분군은 N_i 개 항목으로 구성되며 각 항목이 적합 또는 부적합으로 평가되므로 부분군 i 의 최대 부적합품 수는 N_i 개입니다 .
- C 차트에는 대개 하나의 검사 단위로 구성된 (반드시 그런 것은 아님) 부분군 표본의 결점 (결함) 수가 표시됩니다 .
- U 차트에는 검사 단위 수가 변할 수 있는 부분군 표본의 단위별 결점(결함) 수가 표시됩니다.

희귀 사건 관리도

희귀 사건 차트는 드물게 발생하는 사건에서 데이터를 가져오는 공정에 대한 정보를 제공하는 관리도입니다. 기존 관리도에서 드물게 발생하는 추적 공정은 비효과적인 경향이 있습니다. 희귀 사건 차트는 희귀 사건 시나리오에서 관리도의 한계에 대응하여 개발되었습니다. 관리도 빌더는 두 가지 유형의 희귀 사건 차트를 제공합니다(G 차트, T 차트). G 차트와 T 차트의 차이는 희귀 사건 사이의 간격을 측정하는 데 사용되는 양(quantity)입니다. G 차트는 일정 시간 사이에 발생한 사건 수를 측정하는 반면 T 차트는 발생 사건 사이의 시간 간격을 측정합니다.

표 3.2 시그마를 기반으로 희귀 사건 차트 결정

시그마 계산에 사용되는 분포	차트 유형
음이항	G 차트
Weibull	T 차트

G 차트

G 차트는 드물게 발생하는 오류 또는 부적합 발생 사건 수를 측정하여 시간 경과에 따른 공정 차트를 생성합니다. 차트의 각 점은 상대적으로 드문 사건 발생 사이의 단위 수를 나타냅니다. 예

를 들어 물품이 매일 생산되는 환경에서 예기치 않게 생산 라인이 중단될 수 있습니다. G 차트를 사용하여 생산 라인 중단 사이에 생산된 단위 수를 검토할 수 있습니다.

G 차트를 읽을 때 관리 상한 위의 점은 오류 사이의 사건 수가 증가했음을 나타냅니다. 드물게 발생하는 오류 또는 부적합 발생 사건 수가 증가했다면 관리도 차원에서 좋은 의미로 해석할 수 있습니다. 따라서 G 차트로 작업할 때는 일반적으로 한계를 초과하여 관리이탈로 플래그 지정된 점이 바람직한 효과로 간주됩니다.

T 차트

T 차트는 마지막 사건 이후 경과된 시간 간격을 측정합니다. 차트의 각 점은 이전에 희귀 사건이 발생한 이후 경과된 시간 간격 수를 나타냅니다. T 차트는 음수가 아닌 숫자 데이터, 날짜/시간 데이터 및 시간 간격 데이터에 사용할 수 있습니다. 이러한 데이터의 기존 그림에는 0 위치에 많은 점이 포함되고 가끔 1 위치에 점이 포함될 수 있으므로 T 차트를 사용하면 많은 수의 점이 관리이탈 상태로 플래그 지정되는 것을 방지할 수 있습니다. 관리도 빌더에서 T 차트의 데이터 점은 정수 값으로 제한됩니다.

T 차트를 읽을 때 관리 상한 위의 점은 사건 사이의 시간이 증가했음을 나타냅니다. 이는 부작용 발생률이 감소했다는 의미입니다. 따라서 T 차트로 작업할 때는 일반적으로 한계를 초과하여 관리이탈로 플래그 지정된 점이 바람직한 효과로 간주됩니다.

관리도 유형

관리도 빌더와 분석 > 품질 및 공정 > 관리도 메뉴의 플랫폼에서 가장 일반적인 관리도를 사용할 수 있습니다. 차트를 쉽고 빠르게 생성하려면 관리도 빌더를 첫 번째 옵션으로 선택하십시오. JMP에서는 데이터를 기반으로 적절한 차트 유형이 자동으로 선택됩니다. 표 3.3 ~ 표 3.7에는 다양한 관리도 유형이 요약되어 있습니다.

표 3.3 그룹화 (X) 변수 또는 요약되지 않은 데이터가 없는 계량형 차트

차트 유형	관리도 빌더 옵션	
	점 > 통계량	한계 > 시그마
개별	개별	이동 범위
개별값에 대한 이동 범위	이동 범위	이동 범위
개별(중앙값 이동 범위에서 계산된 한계)	개별	중앙값 이동 범위
개별값에 대한 중앙값 이동 범위	이동 범위	중앙값 이동 범위
Levey Jennings	개별	Levey Jennings

표 3.4 그룹화 (X) 변수 또는 요약 데이터가 있는 계량형 차트

차트 유형	관리도 빌더 옵션	
	점 > 통계량	한계 > 시그마
XBar(범위에서 계산된 한계)	평균	범위
XBar(표준편차에서 계산된 한계)	평균	표준편차
R	범위	범위
S	표준편차	표준편차
Levey Jennings	평균	Levey Jennings 또는 전체 표준편차

표 3.5 사전 집계 차트

차트 유형	관리도 빌더 옵션	
	점 > 통계량	한계 > 시그마
그룹 평균에 대한 개별값	평균	이동 범위
그룹 평균에 대한 개별값(중앙값 이동 범위에서 계산된 한계)	평균	중앙값 이동 범위
그룹 표준편차에 대한 개별값	표준편차	이동 범위
그룹 표준편차에 대한 개별값(중앙값 이동 범위에서 계산된 한계)	표준편차	중앙값 이동 범위
그룹 평균에 대한 이동 범위	평균에 대한 이동 범위	이동 범위
그룹 평균에 대한 중앙값 이동 범위	평균에 대한 이동 범위	중앙값 이동 범위
그룹 표준편차에 대한 이동 범위	표준편차에 대한 이동 범위	이동 범위
그룹 표준편차에 대한 중앙값 이동 범위	표준편차에 대한 이동 범위	중앙값 이동 범위

표 3.6 계수형 차트

차트 유형	관리도 빌더 옵션	
	점 > 통계량	한계 > 시그마
P 차트	비율	이항
NP 차트	개수	이항
C 차트	개수	Poisson
U 차트	비율	Poisson

표 3.7 희귀 사건 차트

차트 유형	관리도 빌더 옵션	
	점 > 통계량	한계 > 시그마
G 차트	개수	음이항
T 차트	개수	Weibull

관리도 빌더 시작

다음 두 가지 방법으로 관리도 빌더를 시작할 수 있습니다 .

- 어떤 유형의 관리도가 데이터에 적합한지 확실하지 않은 경우 **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 . 이 방법을 사용하면 데이터 열을 작업 영역으로 드래그할 수 있으며 관리도 빌더가 데이터 유형 및 표본 크기를 기반으로 적절한 차트를 생성합니다 . 자세한 내용은 " 관리도 빌더 대화식 작업 영역 " 에서 확인하십시오 .
- 데이터에 적합한 관리도 유형을 알고 있는 경우 **분석 > 품질 및 공정 > 관리도** 하위 메뉴에서 적절한 차트를 선택합니다 . 그러면 관리도 시작 창이 표시됩니다 . 자세한 내용은 " 특정 관리도의 시작 창 " 에서 확인하십시오 . 다음 관리도에 대한 관리도 빌더 시작 창이 있습니다 .
  - I/MR 관리도
  - XBar 관리도
  - 런 차트
  - P 관리도
  - NP 관리도
  - C 관리도
  - U 관리도

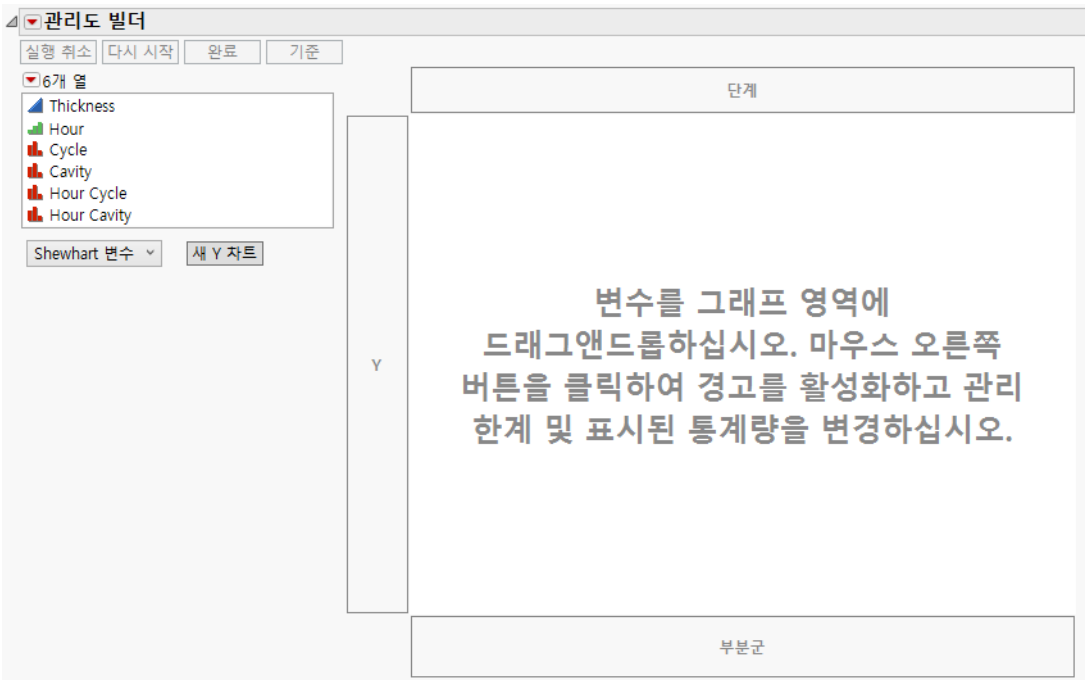
- Levey Jennings 관리도
- 평균 I/MR 관리도
- 삼원 관리도

참고 : CUSUM 관리도 , EWMA 관리도 및 다변량 관리도는 관리도 빌더에서 시작되는 대신 자체 플랫폼에서 시작되며 별도로 문서화됩니다 . 자세한 내용은 "CUSUM 관리도 ", " 다변량 관리도 " 및 "EWMA 관리도 " 에서 확인하십시오 .

시작 창에서 "확인"을 클릭하면 기본적으로 제어판이 숨겨져 있는 관리도 빌더 창이 나타납니다. 다른 모든 옵션 및 기능은 동일합니다.

관리도 빌더 대화식 작업 영역

그림 3.4 대화식 관리도 빌더 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

관리도 생성을 시작하려면 **열 선택** 상자의 변수를 영역으로 드래그합니다 . 변수를 가운데에 놓으면 해당 변수가 연속형인지 아니면 범주형인지에 따라 JMP 에서 변수 위치를 추측합니다 . 관리도 빌더 작업 영역에는 다음 영역이 포함되어 있습니다 .

Y 공정 변수를 할당합니다 .

부분군 부분군 변수를 할당합니다 . 부분군 수준을 여러 열의 조합으로 정의하려면 **부분군** 영역에 여러 변수를 추가합니다 . 부분군 변수가 할당되면 관리도의 각 점은 부분군의 모든 점에 대한 요약 통계량에 해당합니다 .

단계 단계 변수를 할당합니다 . 단계 변수가 할당되면 각 단계에 대해 별도의 관리 한계가 계산됩니다 . 자세한 내용은 "Xbar-R 차트 단계 예제 " 에서 확인하십시오 .

초기 " 관리도 빌더 " 창에는 다음 버튼이 포함되어 있습니다 .

실행 취소 창에서 마지막으로 변경한 사항을 되돌립니다 .

다시 시작 창을 기본 상태로 되돌려 모든 데이터를 제거하고 모든 영역을 지웁니다 .

완료 버튼과 " 열 선택 " 상자를 숨기고 모든 놓기 영역 외곽선을 제거합니다 . 프레젠테이션에 적합한 이 형식을 사용하여 그래프를 다른 프로그램에 복사할 수 있습니다 . 창을 대화식 모드로 복원하려면 " 관리도 빌더 " 의 빨간색 삼각형을 클릭하고 **제어판 표시**를 선택합니다 .

기준 변수를 식별하고 열에 나타나는 각 값에 대해 별도의 분석을 생성합니다 .

Shewhart 계량형 /Shewhart 계수형 / 희귀 사건 "Shewhart 계량형 ", "Shewhart 계수형 " 또는 " 희귀 사건 " 관리도 유형을 선택할 수 있습니다 . 계수형 차트 유형을 선택하면 차트에 " 시행 횟수 " 상자와 영역이 나타납니다 .

시행 횟수 (계수형 차트에 사용 가능) 계수형 관리도에 사용할 로트 크기를 할당합니다 .

새 Y 차트 " 열 선택 " 상자에서 선택한 모든 열에 대해 현재 차트의 복사본을 생성합니다 . 선택한 열은 새 차트의 Y 역할에 사용됩니다 .

변수를 차트로 드래그하면 차트에서 항목을 표시하거나 숨기거나 전환할 수 있는 다른 버튼과 옵션이 왼쪽에 나타납니다 (그림 3.7). 이러한 기능 (점 , 한계 , 경고 등) 중 대부분은 차트를 마우스 오른쪽 버튼으로 클릭하면 사용할 수 있는 기능과 동일합니다 . 자세한 내용은 " 옵션 패널 및 마우스 오른쪽 버튼 클릭 차트 옵션 " 에서 확인하십시오 . 경고와 규칙에 대한 자세한 내용은 " 검정 " 및 "Westgard 규칙 " 에서 확인하십시오 .

삼원 관리도 계량형 차트 유형에 대한 삼원 관리도를 생성할 수 있습니다 . 부분군 크기는 1 보다 커야 합니다 . 부분군 평균 , 부분군 내 변동 또는 부분군 간 변동에 따라 도표화되는 통계량이 달라집니다 . 세 개의 기본 집합에는 이동 범위 한계를 사용한 평균의 사전 집계 차트 , 이동 범위 차트 및 범위 차트가 포함됩니다 .

사건 선택기 선택이 변경될 때 해당 내용에 대한 반응이 실시간으로 차트에 적용될 수 있습니다 . 인식되어 사전 스코어링되는 몇 가지 표준 그룹이 있습니다 (예 : 통과 / 실패 , 예 / 아니요 , Likert 척도 , 적합 / 부적합 및 결함 / 무결함). 설문 조사 결과를 분석 중이고 하나 이상의 질문에 대해 결과의 특정 부문에만 초점을 맞추려는 경우 화면에서 선택할 수 있습니다 . 선택하면 차트가 스코어링되고 즉시 다시 표시됩니다 . 사건 선택기에서 선택한 수준은 사건으로 간주되고 다른 모든 수준은 비사건 (non-event) 으로 간주됩니다 .

사건 선택기는 반응 열의 모델링 유형이 명목형 또는 순서형인 계수형 차트에 사용할 수 있습니다 . 숫자 정수 값 명목형 또는 순서형 반응 열에서 사건 선택기를 사용하려면 " 관리도 빌더 " 의 빨간색 삼각형 메뉴에서 " 사건 선택기 사용 " 옵션을 선택해야 합니다 . 모델링 유형이 연속형인 반응 열에는 사건 선택기가 나타나지 않습니다 .

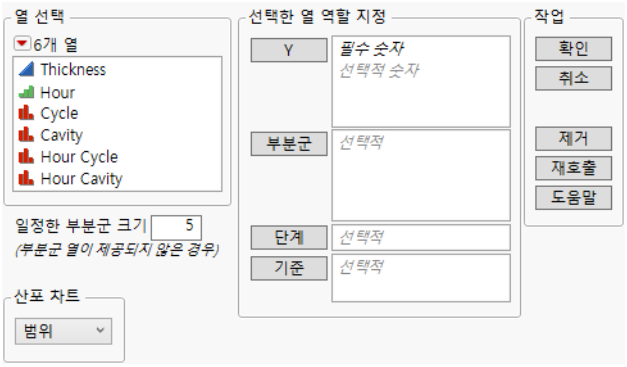
특정 관리도의 시작 창

시작 창에 표시되는 옵션은 시작하는 관리도가 계량형인지 아니면 계수형인지에 따라 달라집니다.

계량형 관리도의 시작 창

이 섹션에는 I/MR, XBar, 런 , Levey Jennings, 평균 I/MR 및 삼원 관리도의 시작 창에 대한 정보가 포함되어 있습니다 .

그림 3.5 계량형 관리도의 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

Y 공정 변수를 할당합니다 .

부분군 (XBar, Levey Jennings, 평균 I/MR 및 삼원 관리도에만 사용 가능) 부분군 변수를 할당합니다 . 부분군 변수가 할당되면 관리도의 각 점은 부분군의 모든 점에 대한 요약 통계량에 해당합니다 .

표본 라벨 (I/MR 차트에만 사용 가능) 개별 측정값에 고유 라벨을 지정하기 위해 표본 라벨 변수를 할당합니다 .

단계 (런 차트에는 사용 불가능) 단계 변수를 할당합니다 . 단계 변수가 할당되면 각 단계에 대해 별도의 관리 한계가 계산됩니다 .

기준 열에 나타나는 각 값에 대해 별도의 분석을 생성하기 위해 변수를 식별합니다 .

표본 라벨을 기준으로 정렬 표본 라벨을 기준으로 모든 변수를 정렬합니다 .

일정한 부분군 크기 (XBar 차트에만 사용 가능) 각 부분군을 구성하는 관측값 수를 지정합니다 . 부분군당 관측값 수가 일정하지 않으면 " 부분군 " 역할에 열을 지정해야 합니다 .

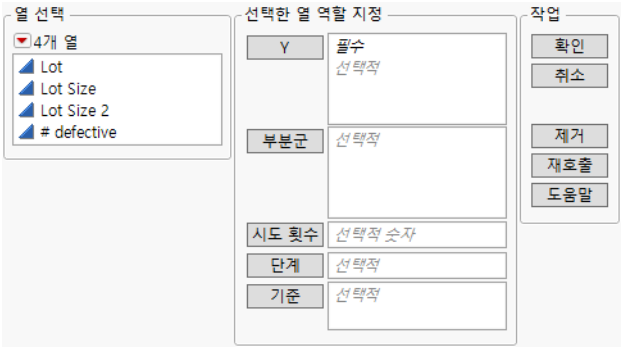
참고: 부분군 변수를 지정하면 플랫폼에서 "일정한 부분군 크기" 옵션의 설정이 무시됩니다.

산포 차트 (XBar 차트에만 사용 가능) XBar 차트 아래에 나타나는 산포 차트의 유형을 지정합니다. 산포 차트 없음, 범위 차트 (XBar & R) 또는 S 관리도 (XBar & S) 중에서 선택할 수 있습니다.

계수형 관리도의 시작 창

이 섹션에는 NP, P, C 및 U 관리도의 시작 창에 대한 정보가 포함되어 있습니다.

그림 3.6 계수형 관리도의 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Y 공정 변수를 할당합니다.

부분군 부분군 변수를 할당합니다. 부분군 변수가 할당되면 관리도의 각 점은 부분군의 모든 점에 대한 요약 통계량에 해당합니다.

시행 횟수 부분군 표본 크기를 할당합니다.

단계 단계 변수를 할당합니다. 단계 변수가 할당되면 각 단계에 대해 별도의 관리 한계가 계산됩니다.

기준 열에 나타나는 각 값에 대해 별도의 분석을 생성하기 위해 변수를 식별합니다.

관리도 빌더 창

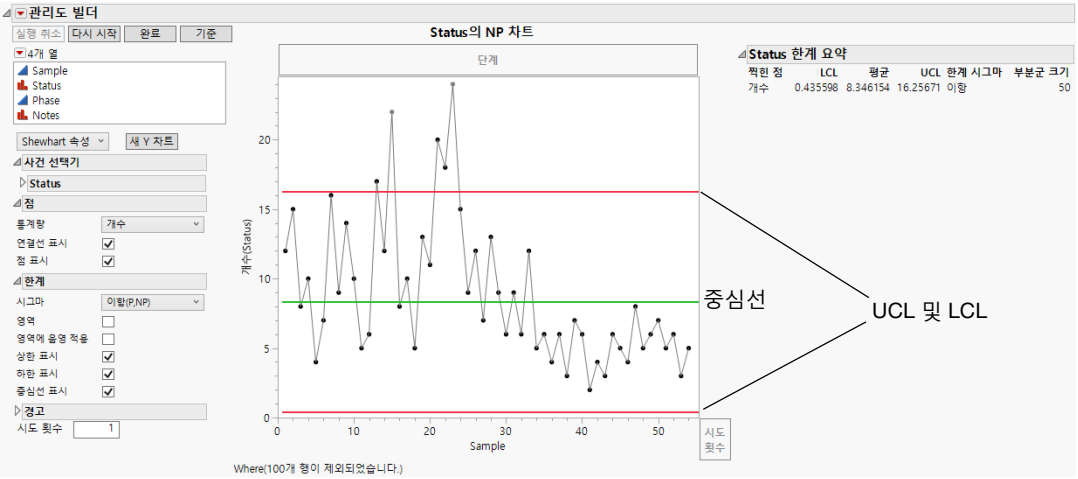
관리도 빌더를 사용하여 공정 데이터에 대한 관리도를 생성할 수 있습니다. 분석을 통해 공정이 통계적 관리 상태인지 여부를 평가하는 데 사용할 수 있는 차트가 생성됩니다. 보고서는 선택한 차트 유형에 따라 달라집니다. 데이터 테이블에서 데이터가 추가되거나 변경되면 관리도가 동적으로 업데이트됩니다. 그림 3.7 에는 Bottle Tops.jmp 샘플 데이터 테이블에 대한 관리도 빌더 창이 표시되어 있습니다.

차트를 생성하려면 다음을 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Bottle Tops.jmp 를 엽니다.

2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Status 를 **Y** 영역으로 드래그합니다 .
4. Sample 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .

그림 3.7 관리도 빌더 창



다른 변수를 다양한 영역으로 드래그하여 분석을 확대하고 "관리도 빌더 옵션"을 사용하여 데이터를 자세히 검토할 수 있습니다. 또한 마우스 오른쪽 버튼을 클릭하면 나타나는 차트 옵션 중 일부(예: 점, 한계, 경고 및 영역 표시 또는 숨기기, 통계량 및 시그마 옵션 선택)가 차트 왼쪽에 표시되어 쉽게 액세스할 수 있습니다.

관리도에는 다음과 같은 특성이 있습니다 .

- 차트에 표시된 각 점은 개별 공정 측정값 또는 요약 통계량을 나타냅니다 . 부분군은 추론에 근거하여 **합리적으로** 선택해야 합니다 . 즉 , 부분군 **간의** 공정 신호가 참일 확률을 최대화하도록 선택해야 합니다 . 대개 가장 효과적인 그룹화 전략을 결정하려면 공정에 대한 지식이 필요합니다 . 자세한 내용은 Wheeler(2004), Woodall and Adams(1998) 연구 자료에서 확인하십시오 .
- 관리도의 세로 축은 요약 통계량과 동일한 단위로 척도화됩니다 .
- 관리도의 가로 축은 부분군 표본을 식별하고 시간 순서로 지정됩니다. 시간별 공정을 관측하는 것은 공정의 변화 여부를 평가하는 데 중요합니다 .

녹색 선은 중심선 , 즉 데이터의 평균입니다 . 중심선은 공정이 통계적 관리 상태일 때 요약 통계량의 평균 값 (기대값) 을 나타냅니다 . 측정값은 중심선의 양쪽에 균등하게 나타나야 합니다 . 그렇지 않으면 공정 평균이 변하고 있다는 증거가 될 수 있습니다 .

- 빨간색 선 두 개는 각각 UCL 및 LCL 라벨이 지정된 관리 상한과 관리 하한입니다 . 관리 한계는 공정이 통계적 관리 상태일 때 요약 통계량에 예상되는 변동 범위를 제공합니다 . 공정에서 일상적 변동만 나타나는 경우 모든 점이 해당 범위 내에 무작위로 포함되어야 합니다 .

참고 : 산포 , 계수형 및 희귀 사건 차트에서 관리 하한을 숨기려면 옵션 패널의 " 하한 표시 " 옵션을 선택 취소합니다 . 관리 하한을 항상 숨기도록 기본값을 변경하려면 파일 > 환경 설정 > 플랫폼 > 관리도 빌더에서 " 관리 하한 표시 " 환경 설정을 선택 취소합니다 .

- 관리 한계를 벗어난 점은 특수한 변동 원인이 있음을 나타냅니다 .

관리도 빌더 창의 옵션은 표본을 받고 기록하거나 데이터 테이블에 추가할 때 동적으로 업데이트할 수 있는 관리도를 생성합니다. 관리도에 비정상적인 변동이 나타날 때 공정 성능이 저하된 경우에는 공정을 통계적 관리 상태로 되돌리는 조치를 취해야 합니다. 비정상적인 변동이 공정 개선을 나타낼 경우에는 변동 원인을 연구하고 구현해야 합니다.

축을 두 번 클릭하면 형식, 축 값, 눈금 수, 격자선, 참조선 및 표시할 기타 옵션을 지정할 수 있는 적절한 축 규격 창이 나타납니다.

관리도 빌더 옵션

관리도 빌더 옵션은 빨간색 삼각형 메뉴에 표시되거나 차트 또는 축을 마우스 오른쪽 버튼으로 클릭하면 나타납니다 . 마우스 오른쪽 버튼을 클릭하여 나타나는 차트 옵션 중 일부는 차트 왼쪽에도 표시되어 쉽게 액세스할 수 있습니다. 또한 파일 > 환경 설정 > 플랫폼 > 관리도 빌더에서 관리도 빌더의 다양한 옵션에 대한 환경 설정을 지정할 수 있습니다 .

- " 관리도 빌더의 빨간색 삼각형 메뉴 옵션 "
- " 옵션 패널 및 마우스 오른쪽 버튼 클릭 차트 옵션 "
- " 관리도 빌더의 마우스 오른쪽 버튼 클릭 축 옵션 "

관리도 빌더의 빨간색 삼각형 메뉴 옵션

" 관리도 빌더 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

제어판 표시 다음 요소를 표시하거나 숨깁니다 .

- 버튼
- 열 선택 상자
- 놓기 영역 테두리
- 체크박스 및 드롭다운 메뉴

한계 요약 표시 한계 요약 보고서를 표시하거나 숨깁니다 . 이 보고서에는 관리 한계 (LCL 및 UCL), 중심선 (평균), 찍힌 점과 한계 및 차트의 표본 크기가 표시됩니다 . 희귀 사건 차트의 경우 표본 크기가 표시되지 않습니다 .

공정 능력 표시 (규격 한계가 있는 Shewhart 계량형 차트에만 사용 가능) 공정 능력 분석 보고서를 표시하거나 숨깁니다. 공정 능력 보고서는 한계 요약 보고서의 일부이므로 " 한계 요약 표시 " 옵션을 선택한 경우에만 나타납니다. 자세한 내용은 " 공정 능력 보고서 " 에서 확인하십시오. 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 관리도 빌더의 공정 능력 보고서에 있는 다양한 옵션에 대한 환경 설정을 지정할 수 있습니다.

참고 : 반응 변수에 변동이 없으면 " 공정 능력 표시 " 를 사용할 수 없습니다.

경보 보고서 표시 관리이탈 표본에 대한 정보가 포함된 보고서를 표시하거나 숨깁니다. 이 보고서는 각 차트에서 현재 활성화된 검정의 오류를 반영하며 다른 검정이 활성화 / 비활성화되고 데이터 및 행 상태가 변경되면 자동으로 업데이트됩니다. 두 번째 테이블에는 각 차트에 대해 현재 활성화된 검정이 나열됩니다. 첫 번째 테이블에는 다음 열이 포함되어 있습니다.

위치 보고서의 맨 위에서 시작하여 차트 위치를 숫자로 나타냅니다.

총 관리이탈 표본 수 선택한 검정 중 하나 이상에 실패한 표본 수를 계산합니다.

경보 비율 총 관리이탈 표본 수를 총 비결측 표본 수로 나눈 값입니다. 관리이탈 비율이라고도 합니다.

참고 : 경보 비율을 계산할 때 제외된 표본 수도 포함하려면 " 제외된 부분군 검정 " 및 " 제외된 영역 표시 " 옵션이 모두 선택되어 있어야 합니다.

한계 라벨 표시 각 차트에서 한계의 라벨을 표시하거나 숨깁니다. 한계는 차트의 오른쪽 프레임 안에 표시됩니다.

시그마 보고서 표시 (Shewhart 계량형 차트에만 사용 가능) 시그마 값 테이블이 포함된 공정 시그마 보고서를 표시하거나 숨깁니다. " 공정 시그마 보고서 " 에는 전체 표본 크기, 부분군 수, 표본 평균, 전체 표준편차, 군내 표준편차 및 안정성 지수가 포함됩니다. 삼원 관리도의 경우 군간 표준편차 및 군간 - 군내 표준편차 값도 표시됩니다. 단계 변수가 지정되면 각 단계에 대한 값 집합이 제공됩니다.

참고 : " 공정 시그마 보고서 " 는 한계 요약 보고서가 실행된 경우에만 나타납니다.

한계 가져오기 열려 있거나 저장된 데이터 테이블에 포함되어 있는 관리 한계를 가져옵니다.

제외된 영역 표시 차트에서 표본이 제외된 영역을 표시 (설정) 하거나 제거 (해제) 합니다. 완전히 제외된 부분군은 위치 차트에 흐린 점으로 표시되어 해당 부분군이 제외되었음을 나타냅니다.

주의 : " 제외된 영역 표시 " 옵션은 차트에 영향을 줍니다. 제외된 표본은 이 옵션의 설정 여부에 관계없이 관리 한계 계산에서 제거됩니다. 제외된 표본은 " 제외된 부분군 검정 " 및 " 제외된 영역 표시 " 옵션이 모두 선택된 경우에만 경보 비율 계산에 포함됩니다.

부분군 크기 설정 (부분군 변수가 지정된 경우에는 사용 불가능) 부분군 크기를 설정합니다. 한계 및 시그마를 계산할 때 결측값이 고려됩니다.

참고 : " 부분군 크기 설정 " 옵션을 사용하는 경우 " 제외된 영역 표시 " 옵션이 자동으로 설정됩니다 .

한계 저장 다음 중 한 가지 방법으로 관리 한계를 저장합니다 .

열에 관리 한계를 반응 변수의 기존 데이터 테이블에 열 특성으로 저장합니다 . 한계가 일정하면 보고서의 각 차트 유형에 대한 LCL, 평균 및 UCL 값이 저장됩니다 . 단계 차트에는 이 옵션을 사용할 수 없습니다 . 또한 각 차트에 대해 표본 크기가 일정하지 않으면 이 옵션을 사용해도 아무 효과가 없습니다 .

새 테이블에 각 차트의 표준편차 및 평균을 새 데이터 테이블에 저장합니다 . 한계가 일정하면 각 차트의 LCL, 평균 , UCL 및 표본 크기도 저장됩니다 . 단계가 있는 경우 각 단계에 대한 새 값 집합이 저장됩니다 . 각 통계량에 대한 행과 각 Y 변수에 대한 열이 있습니다 .

새 세로형 테이블에 (희귀 사건 , 계수형 또는 단계 차트에는 사용 불가능) 각 차트의 표준편차 , 평균 및 시그마를 새 데이터 테이블에 저장합니다 . 한계가 일정하면 각 차트의 LCL, 평균 , UCL 및 표본 크기도 저장됩니다 . 각 Y 변수에 대한 행과 각 통계량에 대한 열이 있습니다 . 공정 변수 선별 플랫폼에서 사용할 수 있는 시그마 열도 저장됩니다 .

요약 저장 표본 라벨 , 표본 크기 , 표시되는 통계량 , 중심선 , 관리 한계 , 검정 , 경고 및 오류와 같은 정보를 포함하는 새 데이터 테이블을 생성합니다 . 테이블에 포함되는 특정 통계량은 차트 유형에 따라 다릅니다 .

그래프 간격 그래프 사이의 간격 크기를 설정합니다 .

결측 범주 포함 그래프에서 범주형 열에 결측값이 있는 행을 수집할 수 있도록 하고 그래프의 결측값을 별도의 범주로 표시합니다 . 이 옵션을 비활성화하면 결측 X 값이 있는 모든 행이 계산에서 제거되고 그래프에서도 숨겨집니다 .

수집된 결측값을 관련 축에 표시할 수 있는 확실한 방법이 없으므로 연속형 X 변수 또는 범주형 Y 변수에는 이 옵션을 사용할 수 없습니다 . 기본적으로 이 옵션은 활성화되어 있습니다 .

참고 : " 결측 범주 포함 " 이 활성화된 경우 범주형 X 변수에 결측값이 있으면 관리도 빌더의 공정 능력 분석 결과가 공정 능력 플랫폼의 결과와 일치하지 않습니다 .

사건 선택기 사용 (비연속형 수치 Y 변수가 있는 계수형 차트에만 사용 가능) 순서형 수치 데이터를 범주화하고 개별 수치 수준의 모델링 선택을 제공합니다 .

경보 스크립트 데이터가 특수 원인 검정에 실패했음을 나타내는 스크립트를 작성하고 실행할 수 있습니다 . 자세한 내용은 " 검정 " 에서 확인하십시오 . 결과를 파일 또는 로그에 쓰거나 이메일로 보낼 수 있습니다 . 검정이 실패한 이유에 대한 설명을 포함하는 옵션이 있습니다 .

경보 스크립트가 호출되면 실행된 스크립트와 후속 JSL 스크립트에서 모두 다음 변수를 사용할 수 있습니다 .

`qc_col` - 열 이름

`qc_test` - 실패한 검정

`qc_sample` - 표본 번호

qc_phase - 오류가 발생한 단계의 라벨

사용자 경보 스크립트 작성에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

참고 : 로컬 데이터 필터를 사용하는 보고서에서는 경보 스크립트를 사용할 수 없습니다 .

행 순서별 정렬 모든 부분군 및 단계 변수를 데이터 테이블에 수준이 나타나는 순서대로 정렬합니다 . 이 옵션은 내포된 부분군 및 단계 변수의 모든 조합에 적용됩니다 .

제외된 부분군 검정 (" 제외된 영역 표시 " 옵션이 선택된 경우에만 사용 가능) 완전히 제외된 부분군을 검정 계산에 포함 (설정) 하거나 제외 (해제) 합니다 . 제외된 부분군이 표시된 경우 " 제외된 부분군 검정 " 옵션을 선택하지 않으면 제외된 부분군이 결측값으로 처리됩니다 .

참고 : 연속된 점을 사용하는 검정의 경우 (런 검정) 완전히 제외된 부분군이 결측값으로 처리되고 연속된 점 개수가 다시 계산됩니다 .

관리도 대화상자 (관리도 시작 창을 통해 관리도를 시작한 경우에만 사용 가능) 관리도를 생성하는 데 사용된 원래 설정으로 관리도 시작 창을 엽니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

- 열 전환기는 연결된 차트가 두 개 이하인 단일 Y 변수에만 사용할 수 있습니다 . 선택한 차트 유형에 따라 Y 역할에 적합한 열만 열 전환기 열 목록에 포함됩니다 .
- 관리도 빌더에서 " 자동 재계산 " 옵션은 기본적으로 설정되어 있으며 해제할 수 없습니다 .
- 로컬 데이터 필터를 사용하는 경우 " 제외된 영역 표시 " 옵션의 선택을 취소하면 더 집중적으로 탐색할 수 있습니다 .

옵션 패널 및 마우스 오른쪽 버튼 클릭 차트 옵션

다음 옵션은 관리도 빌더 보고서의 왼쪽에 나타나거나 , 차트를 마우스 오른쪽 버튼으로 클릭하면 나타납니다 . 일부 옵션은 두 위치에 모두 표시됩니다 .

사건 선택기 (반응 열의 모델링 유형이 명목형 또는 순서형인 계수형 차트에만 사용 가능) 선택이 변경될 때 해당 내용에 대한 반응이 실시간으로 차트에 적용될 수 있습니다 . 자세한 내용은 " 사건 선택기 " 에서 확인하십시오 .

점 다음 옵션을 제공합니다 .

통계량 차트에 표시되는 통계량을 변경합니다. 자세한 내용은 "통계량"에서 확인하십시오.

개별 점 부분군의 개별 관측값을 표시하거나 숨깁니다. 부분군 변수 또는 " 표본 크기 설정 " 에서만 사용할 수 있습니다. 계수형 차트 유형 또는 희귀 사건 차트에는 이 옵션을 사용할 수 없습니다.

상자 그림 (Shewhart 계량형 차트에만 사용 가능) 차트에 상자 그림을 표시하거나 숨깁니다.

연결선 표시 점 사이의 연결선을 표시합니다.

점 표시 차트에 점을 표시하거나 숨깁니다.

한계 다음 옵션을 제공합니다.

시그마 시그마 계산 방법을 지정합니다. 자세한 내용은 " 시그마 " 에서 확인하십시오.

영역 (계량형 및 계수형 차트 유형에만 사용 가능) 차트에 영역을 표시하거나 숨깁니다. 평균의 양쪽에 너비가 동일한 영역이 세 개씩 있습니다. LCL 아래 또는 UCL 위에는 영역이 표시되지 않습니다. 계량형 차트의 한계가 평균을 기준으로 중심화되지 않으면 min(평균-LCL,UCL-평균)/3이 각 영역의 너비로 사용됩니다. 계수형 차트의 영역에서는 (UCL-평균)/3 을 너비로 사용합니다.

영역에 음영 적용 범위별로 음영 영역을 표시하거나 숨깁니다. C 영역은 녹색 , A 및 B 영역은 노란색으로 각각 음영이 적용되고 A 영역을 벗어난 부분은 빨간색으로 음영이 적용됩니다.

규격 한계 (데이터 테이블에 " 규격 한계 " 열 특성이 있거나 , " 규격 한계 추가 " 옵션을 사용하여 규격 한계를 지정하는 경우에만 사용 가능) 차트에 규격 한계를 표시하거나 숨깁니다. 기본적으로 " 규격 한계 " 열 특성에서 " 그래프 참조선으로 표시 " 옵션이 선택된 경우 규격 한계가 표시됩니다. 규격 한계 열 특성을 추가하는 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

규격 한계 추가 (마우스 오른쪽 버튼 클릭 메뉴에서 사용 가능) 규격 한계를 입력할 수 있습니다.

관리 한계 설정 (마우스 오른쪽 버튼 클릭 메뉴에서 사용 가능) 검정의 관리 한계를 입력할 수 있습니다. " 관리 한계 설정 " 창에서 " 확인 " 을 클릭하면 지정된 관리 한계가 전체 그룹에 균등하게 설정됩니다. 지정된 관리 한계를 제거하려면 이 옵션을 다시 선택합니다.

한계 추가 (마우스 오른쪽 버튼 클릭 메뉴에서 사용 가능) 차트에 표시할 추가 관리 한계를 지정합니다. 이러한 한계는 검정에 사용되지 않습니다.

상한 표시 차트에 관리 상한을 표시하거나 숨깁니다. 차트에서 관리 상한을 숨기면 " 한계를 벗어난 경우 검정 " 및 " 검정 1" 옵션은 숨겨진 관리 상한과 연결된 점에 플래그를 지정하지 않습니다.

하한 표시 차트에 관리 하한을 표시하거나 숨깁니다. 차트에서 관리 하한을 숨기면 " 한계를 벗어난 경우 검정 " 및 " 검정 1" 옵션은 숨겨진 관리 하한과 연결된 점에 플래그를 지정하지 않습니다.

중심선 표시 차트에 중심선을 표시하거나 숨깁니다.

산포 차트 추가 (마우스 오른쪽 버튼 클릭 메뉴에서 사용 가능) 차트 영역에 산포 차트를 추가합니다 . " 점 " 옵션으로 차트 유형을 변경합니다 . 산포 차트에서는 범위 , 표준편차 또는 이동 범위를 포함한 여러 형태의 산포 중 하나를 표시하여 데이터의 변동을 보여 줍니다 . 계량형 차트 유형에만 사용할 수 있습니다 .

참고 : 파일 > 환경 설정 > 플랫폼 > 관리도 빌더에서 " 산포 차트 " 및 " 요약된 산포 차트 " 환경 설정을 사용하여 기본 산포 차트 유형을 사용자 정의할 수 있습니다 .

부분군 크기 설정 부분군 크기를 설정합니다. 한계 및 시그마를 계산할 때 결측값이 고려됩니다.

경고 다음 옵션을 제공합니다 .

검정 사용자 정의 (마우스 오른쪽 버튼 클릭 메뉴에서 사용 가능) 사용자 검정을 설계하고 동시에 여러 검정을 선택하거나 선택 취소할 수 있습니다 . 이 옵션을 선택하면 검정 설계를 위한 " 검정 사용자 정의 " 창이 나타납니다 . 검정 설명을 선택하고 원하는 번호 (n) 및 라벨을 입력합니다 . 설정을 환경 설정에 저장할 수 있으며 기본 설정을 복원할 수도 있습니다 . 계량형 및 계수형 차트 유형에만 사용할 수 있습니다 .

검정 활성화할 통계적 관리 검정을 선택할 수 있습니다 . 검정에 대한 자세한 내용은 " 검정 " 에서 확인하십시오 . 계량형 및 계수형 차트 유형에만 사용할 수 있습니다 .

참고 : 차트에서 플래그 지정된 점을 커서로 가리키면 실패한 검정에 대한 설명을 확인할 수 있습니다 .

Westgard 규칙 활성화된 Westgard 통계적 관리 검정 집합을 지정합니다 . Westgard 규칙은 영역이 아니라 시그마를 기반으로 하므로 일정한 표본 크기를 고려하지 않고 계산할 수 있습니다 . 검정에 대한 자세한 내용은 "Westgard 규칙 " 에서 확인하십시오 . 계량형 및 계수형 차트 유형에만 사용할 수 있습니다 .

한계를 벗어난 경우 검정 (JMP 에서는 검정 15 라고 함) 관리 한계를 벗어나는 점에 대한 검정을 활성화합니다 . 이러한 점은 차트에서 식별됩니다 . 이 검정은 표본 크기가 동일한지에 관계없이 한계가 있는 모든 차트에서 작동합니다 .

참고 : 관리 상한 또는 하한을 숨기면 " 한계를 벗어난 경우 검정 " 옵션은 관리도에 표시되지 않는 한계를 벗어나는 점에 플래그를 지정하지 않습니다 .

그래프 제거 (마우스 오른쪽 버튼 클릭 메뉴에서 사용 가능) 관리도를 제거합니다 .

위치 차트 제거 (위치 차트를 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 보고서에서 위치 차트를 제거합니다 .

산포 차트 제거 (산포 차트를 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 보고서에서 산포 차트를 제거합니다 .

참고 : 행 , 그래프 , 사용자 정의 및 편집 메뉴에 대한 설명은 JMP 사용에서 확인하십시오 .

삼원 관리도 계량형 차트 유형에 대한 삼원 관리도를 생성할 수 있습니다 . 자세한 내용은 " 삼원 관리도 " 에서 확인하십시오 .

통계량

차트에 점으로 표시되는 통계량을 변경할 수 있습니다. 사용 가능한 옵션은 선택한 차트 유형에 따라 다릅니다.

계량형 차트 유형의 경우 다음 옵션을 사용하여 차트에 점으로 표시되는 통계량을 변경할 수 있습니다 .

개별 각 점이 데이터 테이블의 개별 값을 나타내는 차트를 생성합니다 .

평균 각 점이 부분군 값의 평균을 나타내는 차트를 생성합니다 .

범위 각 점이 부분군 값의 범위를 나타내는 차트를 생성합니다 .

표준편차 각 점이 부분군 값의 표준편차를 나타내는 차트를 생성합니다 .

평균에 대한 이동 범위 연속된 두 부분군 평균 간의 범위 차이를 계산합니다 .

표준편차에 대한 이동 범위 연속된 두 부분군 표준편차 간의 범위 차이를 계산합니다 .

이동 범위 각 점이 연속된 두 관측값 간의 차이를 나타내는 차트를 생성합니다 .

참고 : 평균 , 범위 , 표준편차 , 평균에 대한 이동 범위 및 표준편차에 대한 이동 범위 방법은 표본 크기가 1 보다 큰 부분군 변수가 지정되거나 표본 크기가 설정된 경우에만 나타납니다 .

계수형 차트 유형의 경우 다음 옵션을 사용하여 차트에 점으로 표시되는 통계량을 변경할 수 있습니다.

비율 각 점이 부분군 표본의 항목 비율을 나타내는 차트를 생성합니다 .

개수 각 점이 부분군 표본의 항목 수를 나타내는 차트를 생성합니다 .

희귀 사건 차트 유형의 경우 "개수" 옵션을 사용하여 통계량을 차트에 점으로 나타냅니다.

시그마

차트에 대한 시그마 계산 방법을 변경할 수 있습니다. 사용 가능한 옵션은 선택한 차트 유형에 따라 다릅니다.

계량형 차트 유형의 경우 다음 옵션을 사용할 수 있습니다 .

범위 부분군 데이터의 범위를 사용하여 시그마를 추정합니다 .

표준편차 부분군 데이터의 표준편차를 사용하여 시그마를 추정합니다 .

이동 범위 이동 범위를 사용하여 시그마를 추정합니다 . 이동 범위는 연속된 두 점 사이의 차이입니다 .

중앙값 이동 범위 평균 이동 범위가 아니라 중앙값 이동 범위를 사용하여 시그마를 추정합니다 .

Levey-Jennings 모든 관측값의 표준편차를 사용하여 시그마를 추정합니다 . 차트에 단계가 있는 경우 시그마는 각 단계에 대해 개별적으로 계산됩니다 .

계수형 차트 유형의 경우 다음 옵션을 사용할 수 있습니다.

이항 이항 분포 모형을 사용하여 시그마를 추정합니다 . 이 모형은 각 실험이 특정 확률로 성공하는 일련의 순차적 실험에서 성공 횟수를 나타냅니다 . " 이항 " 을 선택하면 P 또는 NP 차트가 생성됩니다 .

Poisson Poisson 분포 모형을 사용하여 시그마를 추정합니다 . 이 모형은 사건의 수와 지정된 시간 간격 동안 이러한 사건이 발생하는 시간을 나타냅니다 . "Poisson" 을 선택하면 C 또는 U 차트가 생성됩니다 .

희귀 사건 차트 유형의 경우 다음 옵션을 사용할 수 있습니다.

음이항 음이항 분포 모형을 사용하여 시그마를 추정합니다. 이 모형은 지정된 실패 횟수에 도달하기 전에 일련의 시행에서 성공 횟수를 나타냅니다 . " 음이항 " 을 선택하면 G 차트가 생성됩니다 .

Weibull Weibull 분포 모형을 사용하여 시그마를 추정합니다 . 이 모형은 고장 사이의 평균 시간을 나타냅니다 . "Weibull" 을 선택하면 T 차트가 생성됩니다 .

검정

팝업 메뉴 또는 창 왼쪽에 있는 "경고" 옵션은 검정 선택을 위한 하위 메뉴를 표시합니다. 메뉴에서 특수 원인에 대한 하나 이상의 검정(Western Electric 규칙)을 선택할 수 있습니다. Nelson(1984)은 관리도에서 특수 검정을 식별하는 데 사용되는 번호 매기기 표기법을 개발했습니다. 크기가 동일한 표본과 동일하지 않은 표본에서 모두 검정이 가능합니다.

선택한 검정이 특정 표본에 대해 양성인 경우 해당 점의 라벨에 검정 번호가 지정됩니다. 표시할 검정을 여러 개 선택한 경우 특정 점에서 둘 이상의 검정에 대한 신호가 나타나면 가장 낮은 숫자가 지정된 검정의 라벨이 해당 점 옆에 나타납니다. 차트에서 플래그 지정된 점을 커서로 가리키면 실패한 검정에 대한 설명을 확인할 수 있습니다.

팁 : 동시에 여러 검정을 추가하거나 제거하려면 제어판의 **경고 > 검정**에서 검정을 선택 또는 선택 취소하십시오 .

표 3.8 에는 8 개의 검정 목록 및 해석이 포함되어 있고 그림 3.9 에서는 검정을 보여 줍니다 . 각 검정에 다음 규칙이 적용됩니다 .

- 상한과 하한 사이의 영역은 각각 너비가 1 표준편차인 6 개 영역으로 구분됩니다 .
- 영역에 A, B, C, C, B, A 라벨이 지정되며 C 영역이 중심선에서 가장 가깝습니다 .
- C 영역과 B 영역을 구분하는 선을 벗어난 점은 B 영역 내에 있거나 영역을 벗어납니다 . 즉 , 중심선으로부터의 표준편차가 1 을 초과하는 경우입니다 .
- 두 영역을 구분하는 선 위에 있는 점은 가장 안쪽 영역에 속하는 것으로 간주됩니다 . 따라서 A 영역과 B 영역 사이의 선 위에 있는 점은 B 영역에 있는 것으로 간주됩니다 .
- 단계 변수가 지정되면 각 단계를 시작할 때 각 검정의 개수가 재설정됩니다 .

참고 :

- 검정 1 ~ 검정 8 은 모든 Shewhart 계량형 차트 유형에 적용됩니다 .

- 검정 1, 2, 5, 6 은 차트의 위쪽 및 아래쪽 절반에 각각 별도로 적용됩니다 .
- 검정 3, 4, 7, 8 은 전체 차트에 적용됩니다 .
- 연속 관측값을 기반으로 하는 런 검정이 발동하면 다음 표본으로 이동할 때 개수가 0 으로 재설정되지 않습니다 .
- 런 검정은 " 제외된 영역 표시 " 및 " 제외된 부분군 검정 " 옵션의 설정에 따라 제외된 행을 처리합니다 .
  - 기본적으로 두 옵션이 모두 선택되어 있으며 제외된 행이 런 검정 계산에 포함됩니다 .
  - " 제외된 영역 표시 " 옵션을 선택하고 " 제외된 부분군 검정 " 옵션을 선택하지 않으면 제외된 행이 결측값으로 처리되고 다음 표본으로 이동할 때 런 검정 개수가 0 으로 재설정됩니다 .
  - " 제외된 영역 표시 " 옵션을 선택하지 않으면 제외된 행이 삭제된 것처럼 처리됩니다 .
- 계수형 차트에는 검정 5 ~ 검정 8 을 사용할 수 없습니다 .

이러한 검정의 사용 방법에 대한 자세한 권장 사항은 Nelson 연구 자료 (1984, 1985) 에서 확인하십시오 .

그림 3.8 Western Electric 규칙에 대한 영역

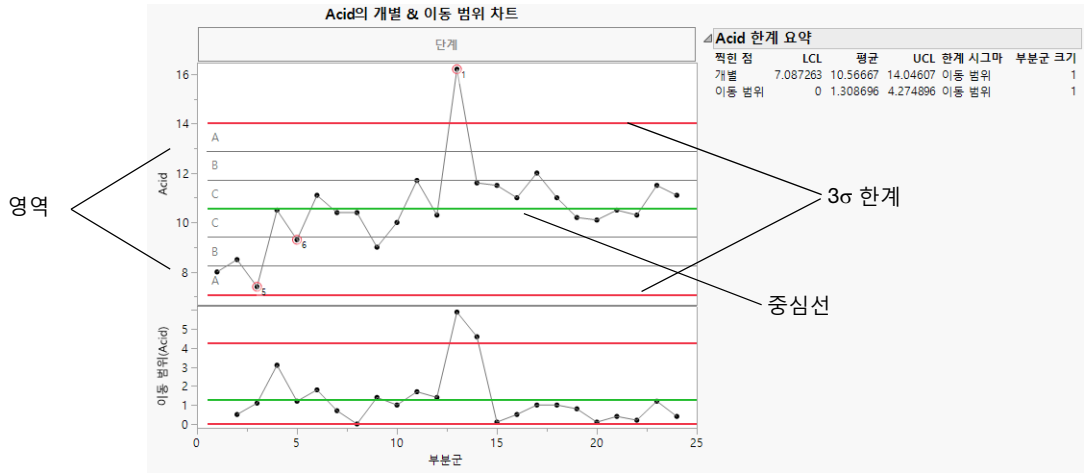


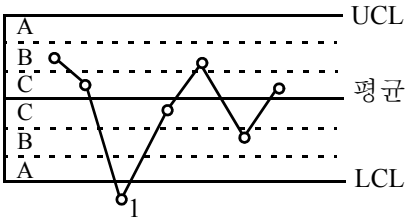
표 3.8 특수 원인 검정에 대한 설명 및 해석[a]

검정 1	1 개의 점이 영역 A 를 벗어나는 경우(위쪽 또는 아래쪽)	공정에서 평균의 변화, 표준편차의 증가 또는 단일 이탈을 감지합니다. 검정 1 해석을 위해 산포 차트(R, S 또는 MR)를 사용하여 변동 증가를 제외할 수 있습니다. 관리 상한 또는 하한을 숨기면 "검정 1" 옵션은 관리도에 표시되지 않는 한계와 연결된 점에 플래그를 지정하지 않습니다.
검정 2	연속된 9 개의 점이 영역 C 의 한쪽(위쪽 또는 아래쪽)에 있거나 벗어나는 경우	공정 평균의 변화를 감지합니다.
검정 3	연속된 6 개의 점이 점진적으로 증가하거나 감소하는 경우(차트의 모든 위치)	공정 평균의 추세 또는 흐름을 감지합니다.
검정 4	연속된 14 개의 점이 세로 방향으로 번갈아 나오는 경우(차트의 모든 위치)	교대로 사용되는 두 기계, 벤더 또는 작업자와 같은 체계적 효과를 감지합니다.
검정 5	연속된 3 개의 점 중 2 개의 점이 영역 A 에 있거나 벗어나고 점 자체가 영역 A 에 있거나 벗어나는 경우. 2 개의 점은 같은 쪽(위쪽 또는 아래쪽)에 있어야 합니다.	공정 평균의 변화 또는 표준편차의 증가를 감지합니다. 3 개 점 중 2 개의 점이 양성 검정을 제공합니다.
검정 6	연속된 5 개의 점 중 4 개의 점이 영역 B 에 있거나 벗어나고 점 자체가 영역 B 에 있거나 벗어나는 경우. 4 개의 점은 같은 쪽(위쪽 또는 아래쪽)에 있어야 합니다.	공정 평균의 변화를 감지합니다. 5 개 점 중 4 개의 점이 양성 검정을 제공합니다.
검정 7	연속된 15 개의 점이 영역 C, 즉 중심선 위/아래에 있는 경우	단일 부분군의 관측값을 평균이 다른 여러 소스에서 가져오는 경우 부분군의 층화를 감지합니다. 변동 감소도 감지합니다.
검정 8	연속된 8 개의 점이 중심선 양측에 있고 영역 C 에는 없는 경우	한 부분군의 관측값은 단일 소스에서 가져오지만 부분군을 평균이 다른 여러 소스에서 가져오는 경우 부분군의 층화를 감지합니다.

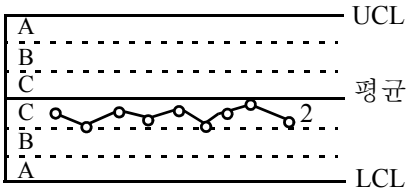
a. Nelson (1984, 1985)

그림 3.9 특수 원인 검정에 대한 그림[1]

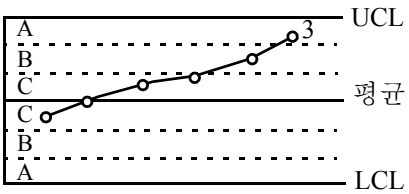
검정 1: 1 개의 점이 영역 A 를 벗어나는 경우



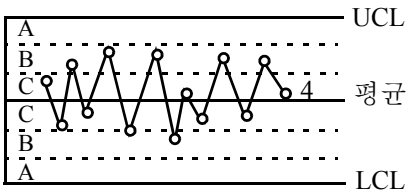
검정 2: 연속된 9 개의 점이 영역 C 의 한쪽 (위쪽 또는 아래쪽) 에 있거나 벗어나는 경우



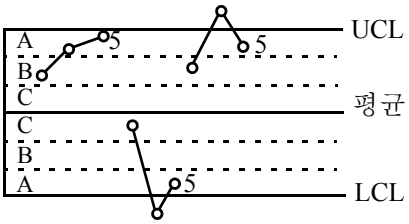
검정 3: 연속된 6 개의 점이 점진적으로 증가하거나 감소하는 경우



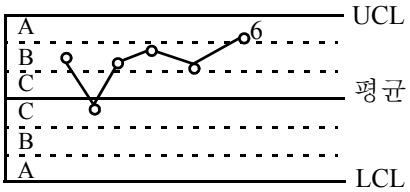
검정 4: 연속된 14 개의 점이 세로 방향으로 번갈아 나오는 경우



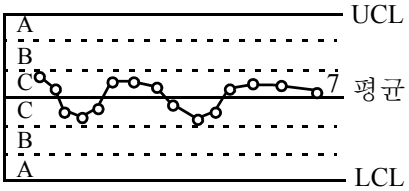
검정 5: 연속된 3 개의 점 중 2 개의 점이 영역 A 에 있거나 벗어나는 경우



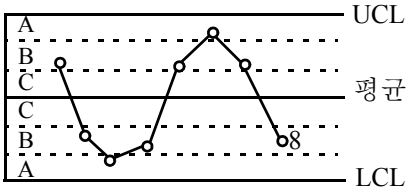
검정 6: 연속된 5 개의 점 중 4 개의 점이 영역 B 에 있거나 벗어나는 경우



검정 7: 연속된 15 개의 점이 영역 C, 즉 중심선 위 / 아래에 있는 경우



검정 8: 연속된 8 개의 점이 중심선 양측에 있고 영역 C 에는 없는 경우



Westgard 규칙

Westgard 규칙은 차트를 마우스 오른쪽 버튼으로 클릭할 때 나타나거나 창 왼쪽에 있는 "경고" 옵션의 "Westgard 규칙" 하위 메뉴에서 구현됩니다. 특정 검정에 대한 결정 규칙을 사용하여 약

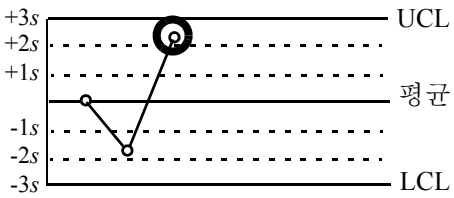
1. Nelson 1984, 1985)

어 형태로 표현된 여러 검정이 있습니다. 예를 들어 **1 2s**는 1 개의 점과 평균 사이의 거리가 2 표준편차인 검정을 나타냅니다.

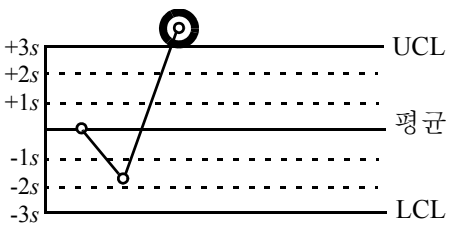
참고 :

- 연속 관측값을 기반으로 하는 런 검정이 발동하면 다음 표본으로 이동할 때 개수가 0 으로 재설정되지 않습니다 .
- 런 검정은 " 제외된 영역 표시 " 및 " 제외된 부분군 검정 " 옵션의 설정에 따라 제외된 행을 처리합니다 .
  - 기본적으로 두 옵션이 모두 선택되어 있으며 제외된 행이 런 검정 계산에 포함됩니다 .
  - " 제외된 영역 표시 " 옵션을 선택하고 " 제외된 부분군 검정 " 옵션을 선택하지 않으면 제외된 행이 결측값으로 처리되고 다음 표본으로 이동할 때 런 검정 개수가 0 으로 재설정됩니다 .
  - " 제외된 영역 표시 " 옵션을 선택하지 않으면 제외된 행이 삭제된 것처럼 처리됩니다 .

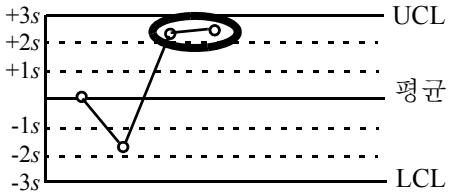
규칙 1 2S(JMP 에서는 검정 9 라고 함)는 관리 한계와 평균 사이의 거리가 2 표준편차로 설정된 Levey-Jennings 차트에서 주로 사용됩니다. 1 개의 점이 이러한 한계를 초과할 때 규칙이 발동합니다.



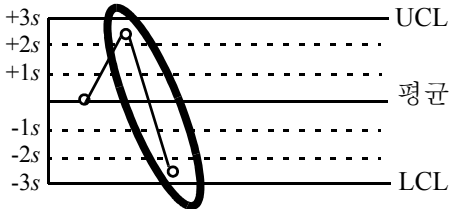
규칙 1 3S(JMP 에서는 검정 10 이라고 함)는 관리 한계와 평균 사이의 거리가 3 표준편차로 설정된 Levey-Jennings 차트에 공통적인 규칙을 나타냅니다. 1 개의 점이 이러한 한계를 초과할 때 규칙이 발동합니다.



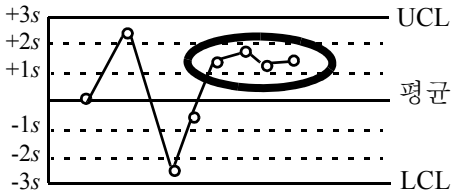
규칙 2 2S(JMP 에서는 검정 11 이라고 함)는 관리도에서 연속된 두 측정값이 평균으로부터 2 표준편차를 벗어날 때 발동합니다.



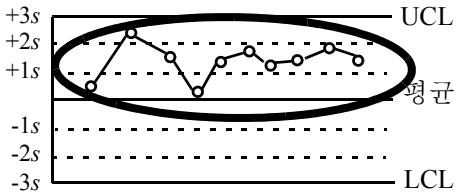
규칙 R 4S(JMP에서는 검정 12라고 함)는 한 측정값과 평균 사이의 거리가 2 표준편차보다 크고 이전 측정값과 평균 사이의 거리가 반대 방향으로 2 표준편차보다 큰 경우, 즉 차이가 4 표준편차보다 클 때 발동합니다.



규칙 4 1S(JMP에서는 검정 13이라고 함)는 연속된 4개의 측정값이 평균으로부터 1 표준편차를 벗어날 때 발동합니다.



규칙 10 X(JMP에서는 검정 14라고 함)는 연속된 10개의 점이 평균의 한쪽 부분에 있을 때 발동합니다.



관리도 빌더의 마우스 오른쪽 버튼 클릭 축 옵션

그래프 제거 전체 그래프를 제거합니다.

제거 변수를 제거합니다.

참고 : 그래프에 차트 유형이 두 개 이상 있으면 여러 차트를 나열하는 하위 메뉴가 표시됩니다. 제거할 차트를 선택할 수 있습니다.

축 설정, 축 되돌리기, 축 라벨 추가 또는 제거, 열 특성에 저장 및 편집 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

관리 한계 작업

관리 한계는 공정 성능을 기반으로 하며 공정의 변동성에 대한 정보를 제공합니다. 관리도가 생성될 때 데이터에서 UCL(관리 상한), 중심선 및 LCL(관리 하한)이 계산됩니다. 계산된 관리 한계를 사용하여 공정이 변경된 시기를 나타낼 수 있습니다.

관리 한계는 주로 공정 능력 분석에 사용되는 규격 한계와 다릅니다.

표 3.9 관리 한계와 규격 한계 비교

관리 한계	규격 한계
데이터에서 계산됨	고객 또는 설계에 의해 정의됨
변동성에 기반	시스템 요구 사항에 기반
VOP(공정의 소리)	VOC(고객의 소리)

관리 한계의 예

회사의 인쇄 공정을 예로 들겠습니다. 변동으로 인해 기울어짐, 두께 및 길이 문제를 포함하여 선이 왜곡될 수 있습니다. 이 예에서는 선 길이를 검토합니다. 인쇄된 선 길이가 16cm +/- 0.2cm 이면 양호한 것으로 간주됩니다. 더 길면 문장이 페이지 밖으로 나갈 수 있습니다. 더 짧으면 페이지에서 버려지는 공간이 많을 수 있습니다. 인쇄를 실행할 때마다 첫 번째 책과 마지막 책이 측정에 사용됩니다. 각 책의 중간에 지정된 페이지에서 선 길이가 측정됩니다.

이 공정이 관리 상태(안정적)인지 여부, 인쇄 품질이 일관적인지 여부, 인쇄 공정을 개선했을 때 나타나는 결과, 품질이 향상되는지 등에 대해 알고 싶습니다. 이러한 질문의 답을 얻기 위해 관리도를 생성하고 관리 한계를 사용해야 합니다.

이 예는 세 부분으로 나누어집니다. 대부분의 경우 JMP에서 관리 한계를 자동으로 계산할 수 있는 "기준 관리도 생성" 작업으로 시작합니다. 그런 다음 이러한 관리 한계를 새 데이터에 적용하기 위해 "관리 한계 지정" 또는 "여러 관리 한계 집합 지정"(단계 데이터의 경우) 작업을 수행합니다.

기준 관리도 생성

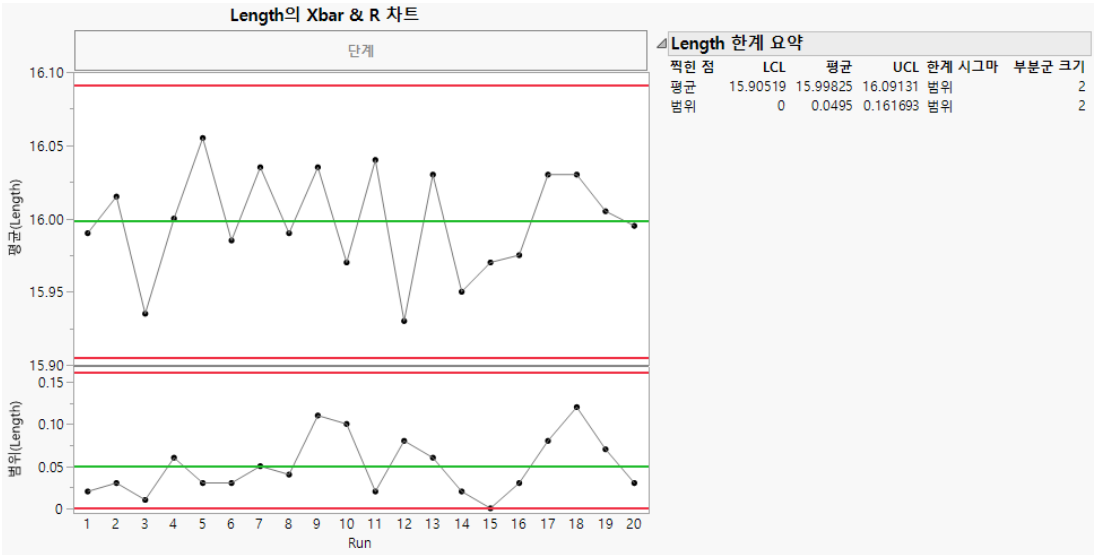
먼저 기존 공정이 관리 상태인지 여부를 검토합니다 . 관리 상태이면 JMP 에서 생성한 관리 한계를 기준 또는 과거 한계로 사용할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Line Length.jmp 를 엽니다 .

2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Length 를 **Y** 영역으로 드래그합니다 .

   "Length 의 개별값 및 이동 범위 차트 " 가 나타납니다 . 이 차트는 데이터에 자연적 부분군 지정이 없는 경우에 적합합니다 . 그러나 이 예에는 자연적 부분군 지정 , 즉 각 인쇄 실행이 있습니다 .
4. Run 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .

그림 3.10 인쇄 실행별 Length(선 길이) 의 Xbar-R 차트



Xbar-R 차트에 가로로 세 개의 선이 그려집니다 . 이 선은 계산된 LCL(관리 하한), 평균 및 UCL(관리 상한) 입니다 .

모든 점이 관리 한계 내에 있고 이러한 한계 내에 점이 무작위로 배치되는 것이 가장 이상적입니다 . 그래프를 보면 관리 한계를 벗어나는 점이 없고 점에 패턴이 있는 것 같지 않습니다 . 더 자세히 조사하려면 Western Electric 검정을 수행하여 이러한 검정의 실패 원인이 되는 패턴과 추세가 있는지 확인하십시오 . Western Electric 검정은 Nelson 검정이라고도 합니다 .

5. XBar 차트에서 마우스 오른쪽 버튼을 클릭하고 **경고 > 검정 > 모든 검정**을 선택합니다 .

   원 안에 포함되거나 플래그가 지정된 점이 없습니다. 이는 공정이 관리 상태이거나 안정적임을 의미합니다 .

   공정이 관리 상태가 아니라고 판단한 경우 관리이탈 점을 조사하거나 관리 상태가 되도록 공정을 변경하는 작업을 수행하게 됩니다. 이 예에서는 공정이 이미 관리 상태이거나 안정적이므로 해당 단계를 건너뛸 수 있습니다 . 이제 새 데이터에 이러한 관리 한계를 사용할 수 있습니다 . 단계 데이터가 있는 예는 " 관리 한계 지정 " 또는 " 여러 관리 한계 집합 지정 " 에서 확인하십시오 .

관리 한계 지정

공정이 관리 상태임을 확인했으므로 이러한 과거 한계를 새 데이터와 함께 사용하여 새 데이터가 기존 공정과 어떻게 비교되는지 확인할 수 있습니다. 과거 한계를 사용하려면 관리 한계를 JMP에서 자동으로 계산하지 않고 사용자가 지정해야 합니다.

다음과 같은 몇 가지 방법으로 JMP 에서 관리 한계를 지정할 수 있습니다 .

- " 관리 한계 설정 옵션 "
- " 열 특성 추가 "
- " 한계 가져오기 옵션 사용 "
- " 행 제외 "

관리 한계 설정 옵션

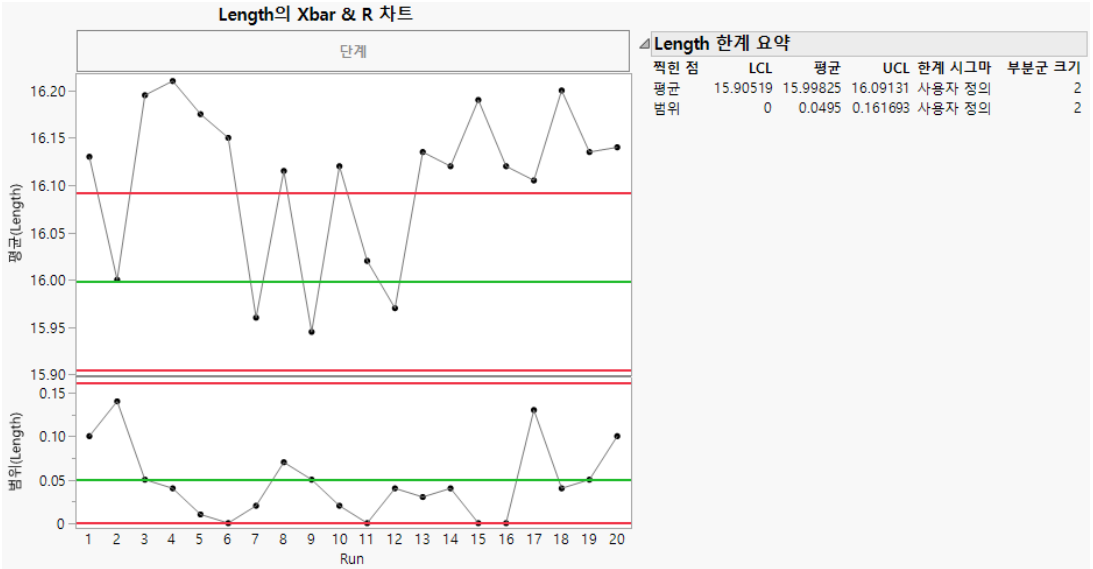
관리 한계를 지정하는 간단한 방법 중 하나는 관리도 빌더의 " 관리 한계 설정 " 옵션을 사용하는 것입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/New Length Data.jmp 를 엽니다 .
   이 테이블에는 새 데이터가 포함되어 있습니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Length 를 **Y** 영역으로 드래그합니다 .
4. Run 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .
5. 평균 (XBar) 차트에서 마우스 오른쪽 버튼을 클릭하고 **한계 > 관리 한계 설정**을 선택합니다 .
6. 다음 한계를 입력합니다 .
   - LCL - 15.90519
   - 평균 - 15.99825
   - UCL - 16.09131

   이 값은 그림 3.10 에 있는 평균 (XBar) 차트의 과거 한계입니다 .
7. **확인**을 클릭합니다 .
8. 범위 (R) 차트에서 마우스 오른쪽 버튼을 클릭하고 **한계 > 관리 한계 설정**을 선택합니다 .
9. 다음 한계를 입력합니다 .
   - LCL - 0
   - 평균 - 0.0495
   - UCL - 0.161693

   이 값은 그림 3.10 에 있는 범위 (R) 차트의 과거 한계입니다 .
10. **확인**을 클릭합니다 .

그림 3.11 과거 한계를 사용한 Length(선 길이) 의 Xbar-R 차트



JMP는 데이터에서 한계를 계산하는 대신 사용자가 정의한 과거 관리 한계를 사용했습니다. 이제 "Length 한계 요약" 테이블의 "한계 시그마"에 "사용자 정의"라고 표시됩니다. 지금은 많은 점이 한계를 벗어났습니다. 또한 평균이 기준 공정의 평균보다 높습니다. 이 공정은 기준 관리 한계를 계산하는 데 사용한 원래 공정과 다르게 나타납니다.

열 특성 추가

관리 한계를 지정하는 또 다른 방법은 " 관리 한계 " 열 특성을 새 데이터 테이블의 열에 추가하는 것입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/New Length Data.jmp 를 엽니다 .
   이 테이블에는 새 데이터가 포함되어 있습니다 .
2. Length 열을 선택하고 **열 > 열 정보**를 클릭합니다 .
3. **열 특성 > 관리 한계**를 클릭합니다 .
4. XBar 가 선택되었으므로 평균 (XBar) 차트에 대한 다음 고정 한계를 입력합니다 .
   - 평균 - 15.99825
   - LCL - 15.90519
   - UCL - 16.09131

   이 값은 그림 3.10 에 있는 평균 (XBar) 차트의 과거 한계입니다 .
   " 부분군 크기 " 값은 비워 둡니다 . 이 값은 관리도 빌더 플랫폼에서 사용되지 않습니다 .
5. **XBar > R** 을 클릭하고 범위 (R) 차트에 대한 다음 고정 한계를 입력합니다 .
   - 평균 - 0.0495

- LCL - 0
- UCL - 0.161693

이 값은 그림 3.10 에 있는 범위 (R) 차트의 과거 한계입니다 .

" 부분군 크기 " 값은 비워 둡니다 . 이 값은 관리도 빌더 플랫폼에서 사용되지 않습니다 .

6. **확인**을 클릭합니다 .

   Length 열에 대한 Xbar-R 차트의 관리 한계를 입력했습니다 . 이제 관리도를 생성할 수 있습니다 .

7. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
8. Length 를 **Y** 영역으로 드래그합니다 .
9. Run 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .

그림 3.11 의 관리도와 동일한 관리도가 표시됩니다.

한계 가져오기 옵션 사용

" 한계 가져오기 " 를 사용하여 관리 한계를 지정하는 것은 가장 유연한 방법입니다 . 다음과 같은 경우 이 방법을 사용해야 합니다 .

- 여러 공정에 대한 관리 한계가 있는 경우
- 각 단계마다 다른 관리 한계가 있는 경우 (" 여러 관리 한계 집합 지정 " 참조)

" 한계 가져오기 " 방법을 사용하려면 과거 한계를 정의하는 데이터 테이블이 필요합니다. 한계 테이블을 생성하는 방법에 대한 자세한 내용은 " 한계 저장 및 검색 "에서 확인하십시오.

참고 : 부분군 변수가 지정되지 않은 경우 " 한계 가져오기 " 옵션은 한계 테이블의 부분군 크기 (_Sample Size) 를 사용합니다 . 또한 파일에 한계가 없는 경우에는 시그마 (_Std Dev) 도 검색합니다 . 한계 파일에 LCL 또는 UCL 이 지정되지 않은 경우 (평균과 시그마가 모두 있고 부분군 크기가 일정한 경우) 이 옵션은 평균 , 부분군 크기 및 시그마를 기반으로 한계를 설정합니다 .

이 예에서는 한계 데이터 테이블이 이미 생성되었습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/New Length Data.jmp 를 엽니다 .

   이 테이블에는 새 데이터가 포함되어 있습니다 .

2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Length 를 **Y** 영역으로 드래그합니다 .
4. Run 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .
5. " 관리도 빌더 " 의 빨간색 삼각형을 클릭하고 **한계 가져오기**를 선택합니다 .
6. **기타**를 선택하고 **확인**을 클릭합니다 .
7. 이 예에 사용할 한계 데이터 테이블 (Length Limits.jmp) 을 찾아서 엽니다 . 기본적으로 이 파일은 다음 위치에 있습니다 .
   - Windows: C:\Program Files\SAS\JMP\16\Samples\Data\Quality Control

– macOS: \Library\Application Support\JMP\16\Samples\Data\Quality Control

그림 3.11 의 관리도와 동일한 관리도가 표시됩니다.

행 제외

관리 한계를 지정하는 또 다른 방법은 데이터 테이블에서 행을 제외하는 것입니다. 이 방법을 사용하면 동일한 그래프에서 과거 데이터와 새 데이터를 모두 볼 수 있다는 장점이 있습니다. 이렇게 하면 데이터 수집 기간 사이에 발생하는 차이를 시각화하고 조사할 수 있습니다.

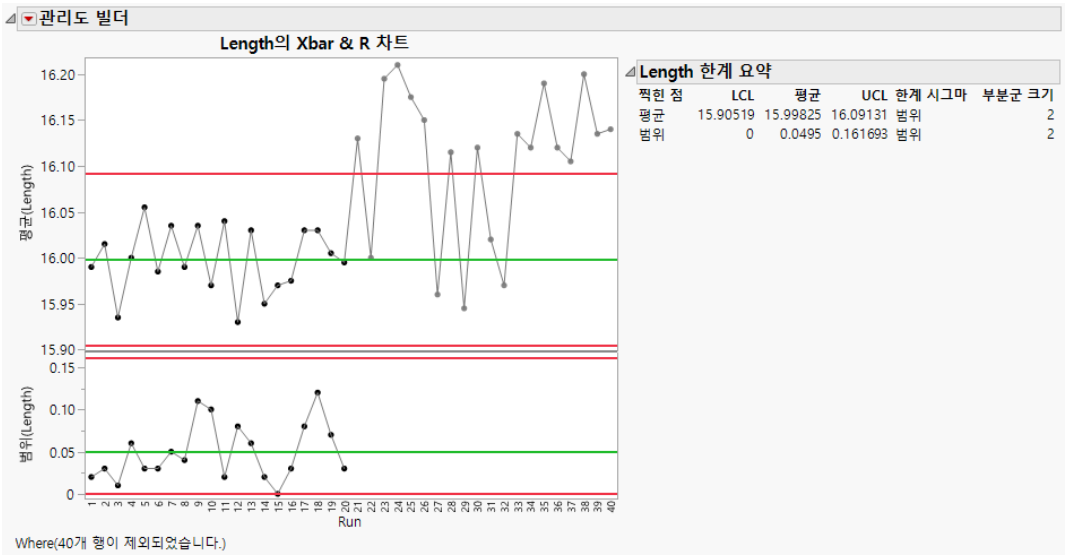
이 방법을 사용하려면 다음 기준을 충족해야 합니다 .

- 새 데이터와 이전 데이터가 동일한 데이터 테이블에 있어야 합니다 .
- 과거 데이터와 새 데이터의 부분군 크기가 모두 같아야 합니다 .
- 모든 새 데이터가 데이터 테이블에서 제외되어야 합니다 (행 > 제외 / 제외 해제 사용).

이 예에서는 새 데이터가 이미 제외되었습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Combined.jmp 를 엽니다 .
   이 테이블에는 이전 데이터와 새 데이터가 포함되어 있고 새 데이터에 해당하는 행이 제외되었습니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Length 를 **Y** 영역으로 드래그합니다 .
4. Run 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .

그림 3.12 제외된 데이터가 있는 Length(선 길이) 의 Xbar-R 차트



JMP는 제외되지 않은 행(과거 데이터)만 사용하여 관리 한계를 생성합니다. 새 데이터(제외된 데이터)는 그래프에 흐리게 표시되지만 이러한 데이터는 계산에 사용되지 않았습니다.

여러 관리 한계 집합 지정

이 예에서는 공정의 여러 단계에 대해 서로 다른 관리 한계를 설정하려고 합니다. 열 특성, 관리 한계 설정 및 제외된 행 상태 방법은 전체 차트에 대해 하나의 관리 한계 집합으로 제한되므로 이 경우에는 사용할 수 없습니다. 단계가 있는 관리도의 경우 한계 가져오기 방법을 사용해야 합니다.

인쇄 회사의 목표는 서로 다른 세 현장에서 종이와 책등 사이의 결합을 끊는 데 필요한 힘의 변동을 줄이는 것입니다 . 각 현장마다 사용하는 기계 및 작업자가 다르며 국가도 다르므로 고유한 과거 한계 집합을 가지고 있습니다 . 이 회사는 세 현장 모두에 대해 다음을 수행합니다 .

1. 기존 공정 데이터를 기반으로 기준 관리도를 생성합니다 .
2. 설계된 실험을 기반으로 공정을 변경합니다 .
3. 새 공정에서 데이터를 수집합니다 .
4. 새 공정 데이터를 기반으로 새 관리도를 생성합니다 .

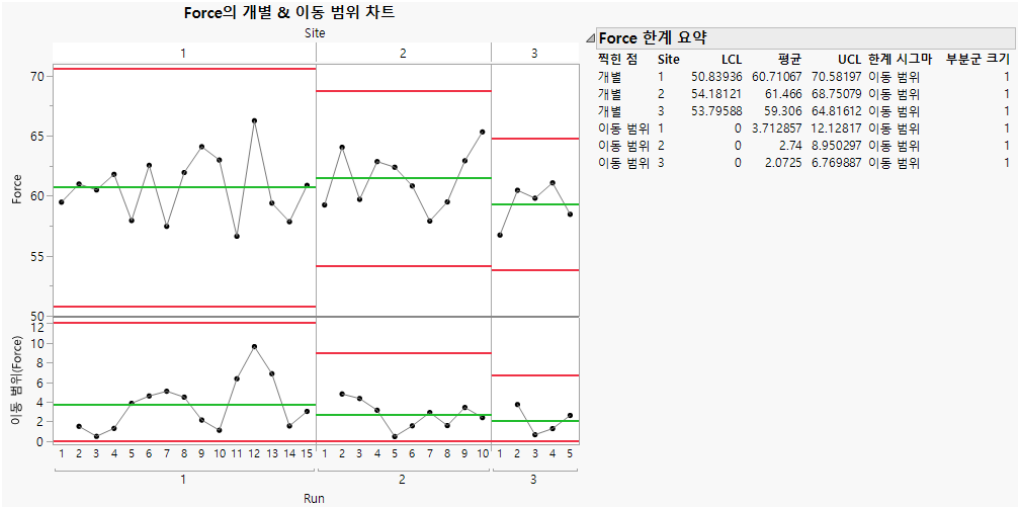
목표는 이전 공정의 과거 한계를 사용하여 관리도에 새 데이터를 표시하는 것입니다. 인쇄 회사는 이 방식으로 새 공정을 이전 공정 한계와 비교할 수 있습니다.

기존 공정을 기반으로 관리도 생성

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Phase Historical Data.jmp 를 엽니다 .

   이 테이블에는 세 현장 모두에 대한 기존 공정 데이터가 포함되어 있습니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Force 를 **Y** 영역으로 드래그합니다 .
4. Run 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .
5. Site 를 **단계** 영역으로 드래그합니다 .

그림 3.13 기존 데이터의 기준 관리도



업데이트된 공정을 기반으로 관리도 생성

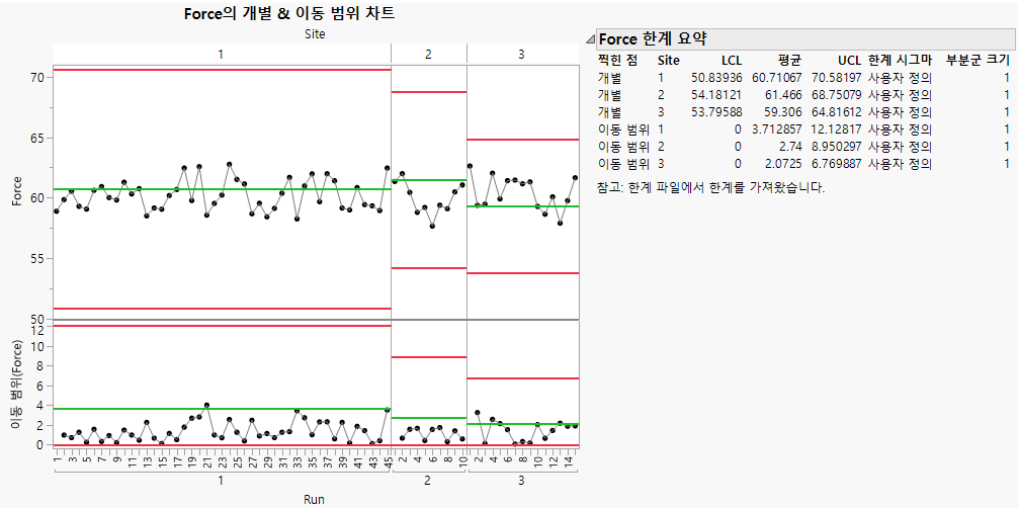
1. 그림 3.13 의 보고서에서 " 관리도 빌더 " 의 빨간색 삼각형을 클릭하고 **한계 저장 > 새 테이블에**를 선택합니다 .

   이렇게 하면 한계 테이블이 생성됩니다 .
2. 나중에 액세스할 수 있도록 이 새 한계 테이블을 원하는 위치에 저장합니다 .
3. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Phase New Data.jmp 를 엽니다 .

   이 데이터는 공정 변경 후 세 현장 모두에서 수집되었습니다 .
4. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
5. Force 를 **Y** 영역으로 드래그합니다 .
6. Run 을 **부분군** 영역 (아래쪽) 으로 드래그합니다 .
7. Site 를 **단계** 영역으로 드래그합니다 .
8. " 관리도 빌더 " 의 빨간색 삼각형을 클릭하고 **한계 가져오기**를 선택합니다 . 2 단계에서 저장한 한계 테이블을 엽니다 .

   이렇게 하면 관리도 빌더 보고서의 새 데이터에 과거 한계가 적용됩니다 .

그림 3.14 과거 한계를 기반으로 한 새 데이터의 관리도



이제 새 데이터(공정 변경 후)가 과거 공정 한계(공정 변경 전)와 어떻게 비교되는지 확인할 수 있습니다. 위치 차트 또는 산포 차트에 대해 관리 한계를 벗어나는 점이 없습니다. 변동을 줄이는 것이 목표였습니다. 이동 범위 차트를 보면 대부분의 점이 평균 선 아래에 있음을 확인할 수 있습니다. Site(현장) 1과 2의 경우 페이지와 책등 사이의 결합을 끊는 데 필요한 힘의 변동이 감소했음이 분명합니다. Site 3의 감소는 Site 1과 2만큼 크지 않습니다. 인쇄 공정을 개선한 결과 변동을 줄이는 데 성공한 것으로 보입니다.

관리도 빌더에서 제외 및 숨겨진 표본

다음 글머리 기호에는 관리도 빌더에서 제외되거나 숨겨진 표본의 사용에 대한 정보가 요약되어 있습니다.

- 제외된 부분군은 관리 한계 계산에 사용되지 않으며 기본적으로 차트에 흐린 점으로 표시됩니다. " 제외된 영역 표시 " 옵션을 선택하지 않으면 제외된 부분군의 점이 차트에 표시되지 않고 특수 원인 검정에서 결측값으로 처리되며 특수 원인 검정을 위한 점 개수에 포함되지 않습니다.
- 숨겨진 관측값은 관리 한계 계산에 사용되지만 차트에 표시되지 않습니다.
- " 제외된 부분군 검정 " 옵션을 선택한 경우 숨겨진 행과 제외된 행은 특수 원인 검정을 위한 점 개수에 모두 포함됩니다. 제외된 행은 특수 원인 플래그로 라벨을 지정할 수 있습니다. 숨겨진 점에는 라벨을 지정할 수 없습니다. 숨겨진 점에 특수 원인 검정에 대한 플래그가 있는 경우 차트에 표시되지 않습니다.
- 일부가 제외된 부분군의 경우 , 부분군 내에서 하나 이상의 관측값이 제외되고 하나 이상의 관측값이 포함되면 제외된 관측값은 점 통계량 또는 한계를 계산할 때 포함되지 않습니다.
- 음수 및 비정수 데이터에 대한 검사는 제외된 값을 포함한 전체 데이터를 대상으로 합니다.

- 검정은 " 제외된 부분군 검정 " 옵션이 선택된 경우에만 제외된 모든 부분군에 적용됩니다 .

관리도 빌더의 추가 예

참고 : 이 섹션에는 제어판을 표시하는 예와 표시하지 않는 예가 나옵니다 . 제어판을 표시하거나 숨기려면 " 관리도 빌더 " 의 빨간색 삼각형 메뉴에서 " 제어판 표시 " 를 선택하십시오 .

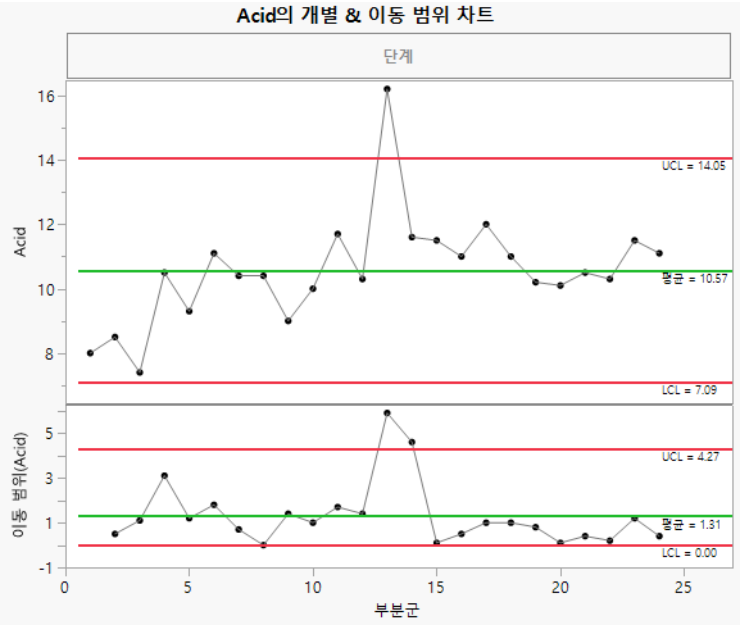
- " 개별 측정값 및 이동 범위 차트 예제 "
- "Xbar-R 차트 단계 예제 "
- " 부분군 크기가 다른 Xbar-S 차트 예제 "
- " 런 차트 예제 "
- "P 차트 예제 "
- "NP 차트 예제 "
- "C 차트 예제 "
- "U 차트 예제 "
- "G 차트 예제 "
- "T 차트 예제 "
- " 삼원 관리도 예제 "
- " 여러 관리도의 예 "

개별 측정값 및 이동 범위 차트 예제

Quality Control 샘플 데이터 폴더의 Pickles.jmp 데이터에는 피클 통의 산 함량이 포함되어 있습니다 . 피클은 산도에 민감하고 큰 통에서 생산되므로 산도가 높으면 피클 통 전체를 망치게 됩니다 . 매일 오후 1 시 , 2 시 , 3 시에 네 통의 산도를 측정합니다 . 날짜 , 시간 및 산도 측정값이 데이터 테이블에 기록됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Pickles.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Acid 를 **Y** 역할로 드래그합니다 .
4. **관리도 빌더**의 빨간색 삼각형을 클릭하고 **한계 라벨 표시**를 선택합니다 .
   이 옵션은 두 차트에서 모두 관리 한계와 평균에 라벨을 지정합니다 .

그림 3.15 Acid(산) 에 대한 개별 측정값 및 이동 범위 차트

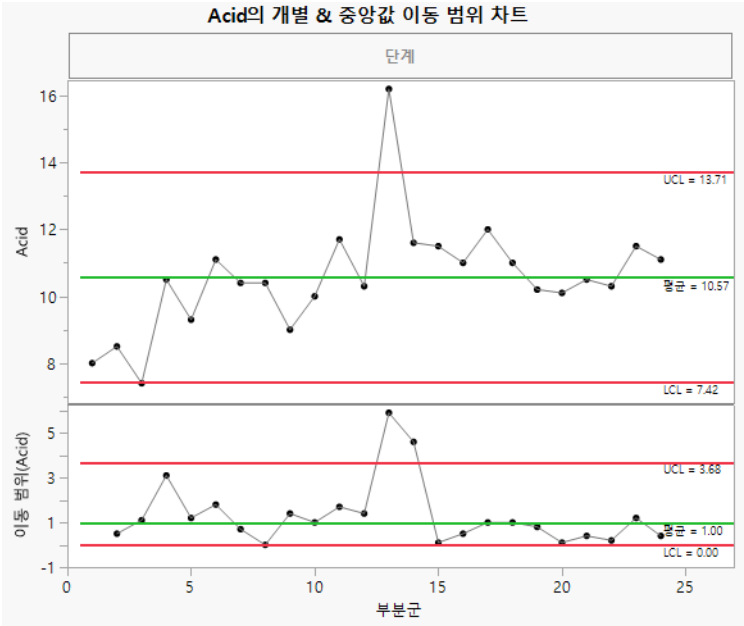


개별 측정값 및 이동 범위 차트는 생산된 각 통 (부분군 크기 1) 의 산도를 모니터링합니다 . Vat 13 의 산도는 관리 상한인 14.05 를 초과합니다 .

중앙값 이동 범위 차트도 볼 수 있습니다. 중앙값 이동 범위를 사용하도록 차트를 변경하려면 다음 단계를 계속하십시오 .

5. " 한계 [1]" 개요에서 **시그마** 설정을 **중앙값 이동 범위**로 변경합니다 .
6. " 한계 [2]" 개요에서 **시그마** 설정을 **중앙값 이동 범위**로 변경합니다 .

그림 3.16 Acid(산) 에 대한 개별 측정값 및 중앙값 이동 범위 차트



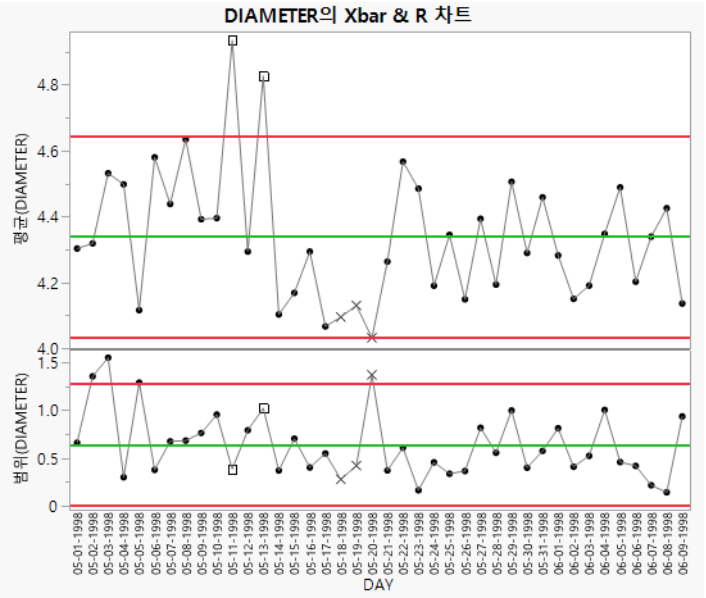
개별 측정값 및 중앙값 이동 범위 차트의 한계는 평균 이동 범위가 아니라 중앙값 이동 범위를 시그마로 사용합니다. 그러면 Acid에 대한 관리 한계가 약간 좁아집니다.

Xbar-R 차트 단계 예제

의료용 튜브 제조업체에서 새 프로토타입에 대한 튜브 지름 데이터를 수집했습니다 . 지난 40 일 간의 생산 공정에서 데이터가 수집되었습니다 . 처음 20 일 (단계 1) 후 제조 장비가 일부 조정되었습니다 . 데이터를 분석하여 최근 20 일 (단계 2) 간의 생산 공정이 관리 상태인지 여부를 판별합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Diameter.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. DIAMETER 를 **Y** 역할로 드래그합니다 .
4. DAY 를 **부분군** 역할 (아래쪽) 로 드래그합니다 .

그림 3.17 Diameter(지름) 에 대한 관리도

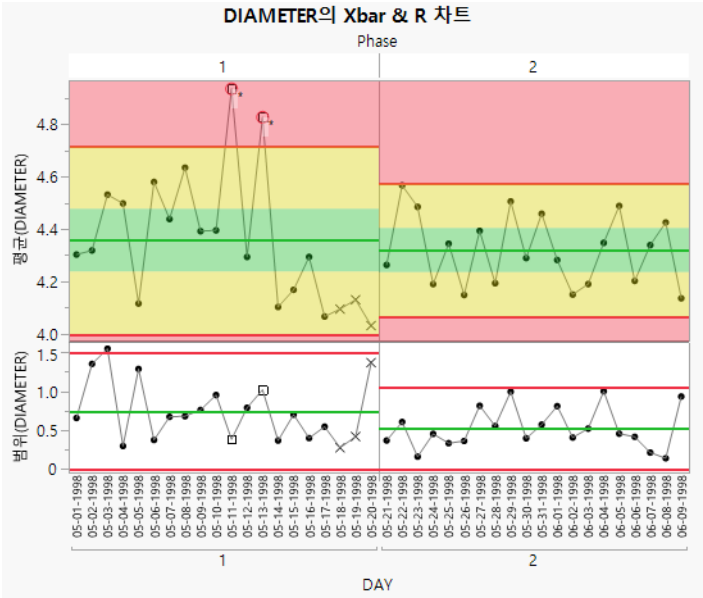


처음 20 일은 변동성이 큰 것으로 보이고 평균 차트에서 세 개의 관측값이 관리 한계를 벗어납니다 . 제조 장비가 조정되고 새 관리 한계가 통합되었습니다 .

각 단계에 대해 별도의 관리 한계를 계산하려면 다음을 수행하십시오 .

5. Phase 를 **단계** 역할로 드래그합니다 .
6. 평균 차트에서 마우스 오른쪽 버튼을 클릭하고 **경고 > 한계를 벗어난 경우 검정**을 선택합니다.
7. " 한계 [1]" 개요에서 **영역에 음영 적용**을 선택합니다 .

그림 3.18 각 Phase(단계) 에 대한 관리도



단계 변수를 포함하면 단계 2의 관리 한계가 단계 2의 데이터만 기반으로 한다는 것을 의미합니다. 단계 2의 모든 관측값이 관리 한계를 벗어나지 않습니다. 이는 차트에 영역 음영을 포함하여 강조 표시됩니다. 따라서 조정 후 공정이 관리 상태에 있다는 결론을 내릴 수 있습니다.

다른 변수를 기준으로 관리도 필터링

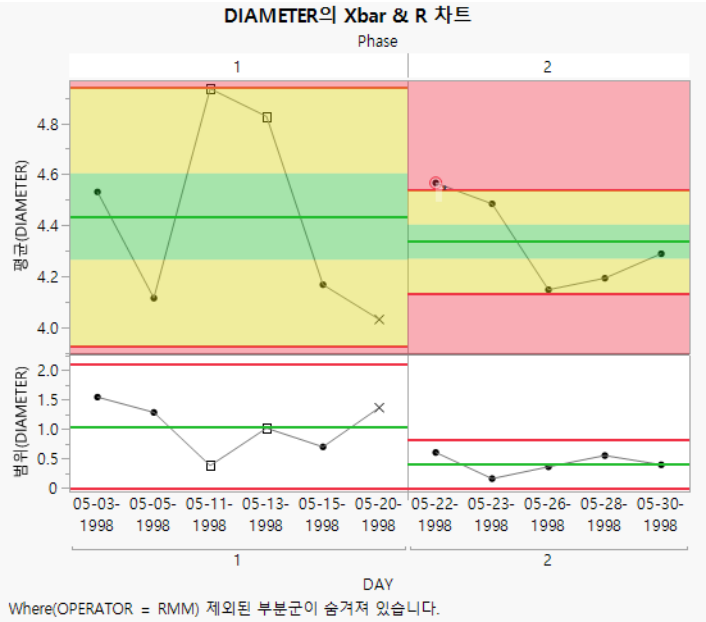
이 데이터 테이블 Diameter.jmp에는 각 표본의 기계 작업자에 대한 열이 포함되어 있습니다. 관리도 빌더에서 로컬 데이터 필터를 사용하여 작업자 부분집합에 대한 데이터를 표시할 수 있습니다.

8. **관리도 빌더**의 빨간색 삼각형을 클릭하고 **제외된 영역 표시**를 선택 취소합니다.

   " 제외된 영역 표시 " 옵션을 해제하면 로컬 데이터 필터에서 선택할 때 로컬 데이터 필터의 설정에 의해 제외된 부분군이 관리도의 가로 축에 더 이상 나타나지 않습니다. 즉, 데이터에서 관심 있는 부분만 볼 수 있습니다.

9. **관리도 빌더**의 빨간색 삼각형을 클릭하고 **로컬 데이터 필터**를 선택합니다.
10. " 로컬 데이터 필터 " 에서 OPERATOR 를 클릭한 후 + 버튼을 클릭합니다.
11. " 로컬 데이터 필터 " 에서 "RMM" 라벨이 표시된 막대를 선택합니다.

그림 3.19 Operator RMM(작업자 RMM) 에 대한 Diameter(지름) 의 Xbar-R 차트



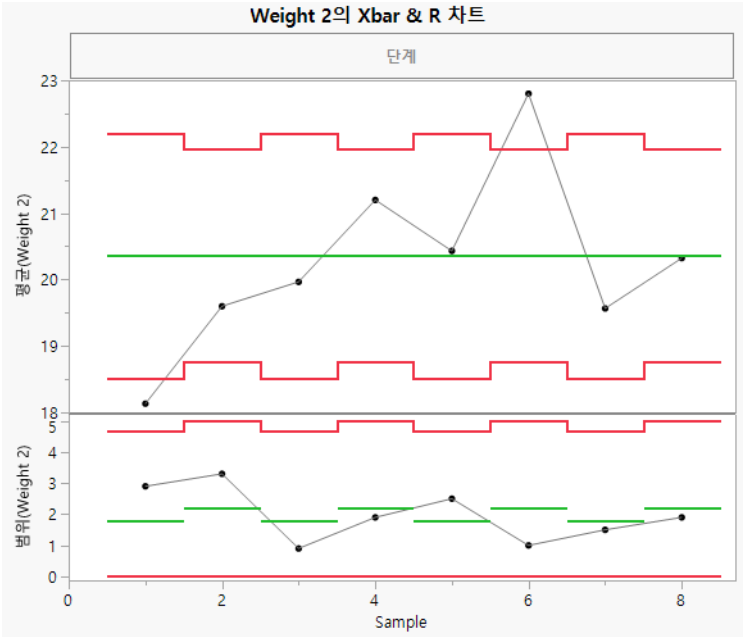
이제 차트 아래의 Where() 문에 표시된 RMM 작업자에 대한 점만 Xbar-R 차트에 표시됩니다. 두 단계에 대한 한계는 다른 세 작업자에 대한 관측값이 제외되었음을 반영하도록 조정되었습니다.

부분군 크기가 다른 Xbar-S 차트 예제

이 예에서는 Coating.jmp 데이터 테이블을 사용합니다. 관심 있는 품질 특성은 Weight 2 열입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Coating.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Weight 2 를 **Y** 역할로 드래그합니다 .
4. Sample 을 **부분군** 역할 (아래쪽) 로 드래그합니다 .

그림 3.20 변하는 부분군 크기에 대한 Xbar-S 차트

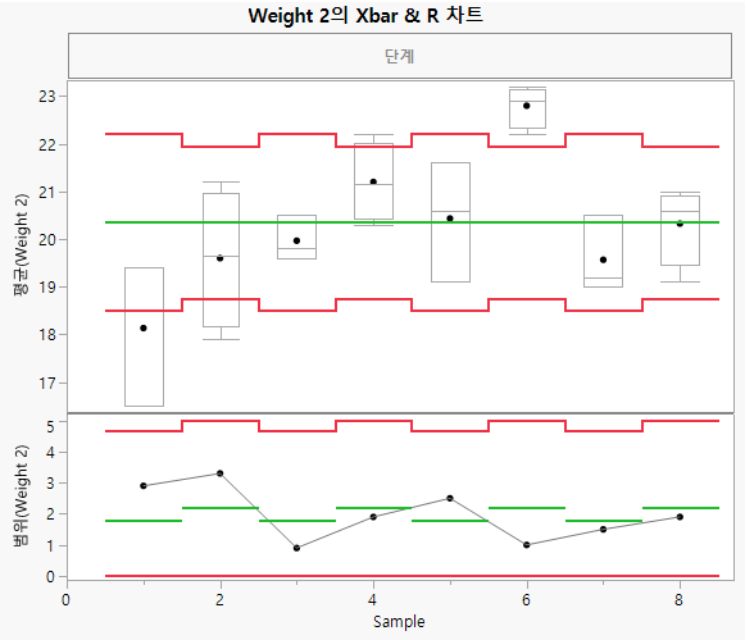


Weight 2 데이터에 결측값이 여러 개 있으므로 차트에 표시된 한계가 고르지 않습니다. 각 표본의 관측값 수는 동일하지만 표본 1, 3, 5, 7 에는 각각 결측값이 있습니다.

각 표본의 평균을 연결하는 선을 표시하는 대신 각 표본의 상자 그림 보기로 전환할 수 있습니다.

5. **점 [1]** 개요에서 **연결선 표시** 옵션을 선택 취소합니다.
6. **점 [1]** 개요에서 **상자 그림** 옵션을 선택합니다.

그림 3.21 상자 그림이 표시된 Xbar-S 차트

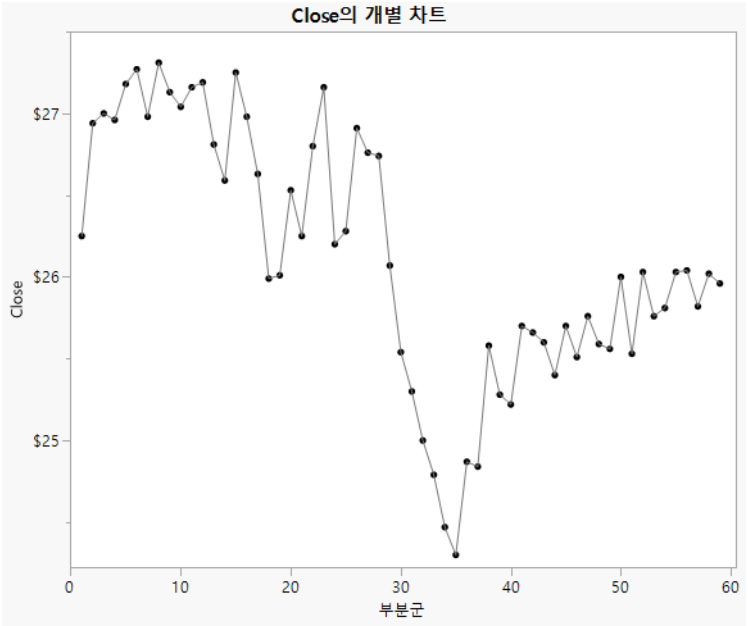


런 차트 예제

런 차트는 데이터 열을 연결된 일련의 점으로 표시합니다 . 이 예는 Stock Averages.jmp 의 Close 변수에 대한 런 차트입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Stock Averages.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > 런 차트**를 선택합니다 .
3. Close 를 선택하고 **Y** 를 클릭합니다 .
4. **확인**을 클릭합니다 .

그림 3.22 주식 평균 종가에 대한 런 차트



P 차트 예제

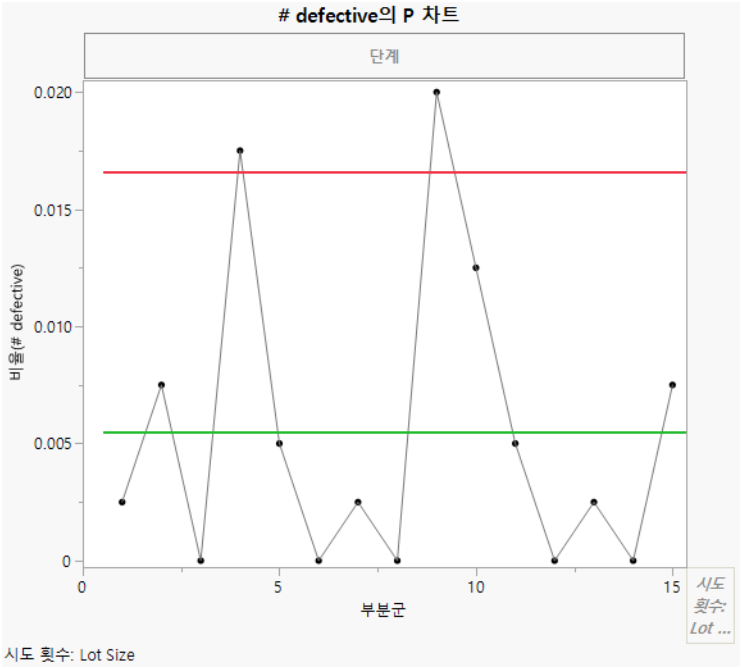
Washers.jmp 샘플 데이터에는 American Society for Testing and Materials(미국 재료 시험 협회 1976) 의 ASTM Manual on Presentation of Data and Control Chart Analysis 와 다른 두 로트 크기에 대한 결함 데이터가 포함되어 있습니다. 상수 표본 크기와 가변 표본 크기의 차이를 보기 위해 Lot Size 및 Lot Size 2 에 대한 차트를 비교할 수 있습니다.

Quality Control 샘플 데이터 폴더의 Washers.jmp 데이터에는 400 개의 아연 도금 와셔 중 15 개 로트의 결함 수가 포함되어 있습니다 . 와셔에 매끈하지 않은 아연 도금 및 철 노출과 같은 마감 결함이 있는지 검사했습니다 . 와셔에 마감 결함이 포함된 경우 부적합 또는 불량으로 간주되었습니다 . 따라서 결함 수는 크기 400 의 각 로트에 대한 불량 와셔 수를 나타냅니다 . Washers.jmp 데이터 테이블을 사용하여 변하는 표본 크기를 허용하는 표본 크기 변수를 지정합니다 . 이 데이터 테이블에는 모든 상수 표본 크기가 포함되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Washers.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > P 관리도**를 선택합니다 .
3. # defective 를 선택하고 **Y** 를 클릭합니다 .
4. Lot Size 를 선택하고 **시행 횟수**를 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 한계 " 개요에서 **하한 표시** 옵션을 선택 취소합니다 .

   이렇게 하면 현재 상황에서 관심 대상이 아닌 하한이 숨겨집니다 .

그림 3.23 표본 크기를 사용한 # defective(결함 수) 의 P 차트



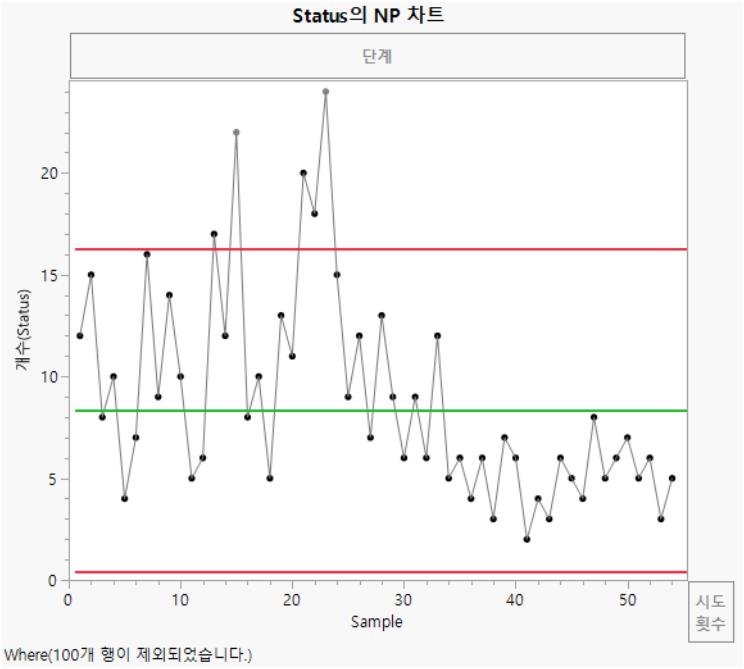
상수 표본 크기와 가변 표본 크기의 차이를 보려면 변수를 시행 횟수 영역으로 드래그하여 Lot Size 및 Lot Size 2에 대한 차트를 비교할 수 있습니다.

NP 차트 예제

Bottle Tops.jmp 샘플 데이터에는 병뚜껑 제조 공정에서 시뮬레이션된 데이터가 포함되어 있습니다 . Sample 은 각 병의 표본 ID 번호입니다 . Status 는 병뚜껑이 설계 기준을 준수하는지 여부를 나타냅니다 . Phase 열에서 첫 번째 단계는 공정 조정 전의 시간을 나타냅니다 . 두 번째 단계는 공정 조정 후의 시간을 나타냅니다 . 공정 변경에 대한 참고 사항 (Notes) 도 포함됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Bottle Tops.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Sample 을 **부분군** 역할로 드래그합니다 .
4. Status 를 **Y** 역할로 드래그합니다 .

그림 3.24 Status(부적합 상태) 의 NP 차트

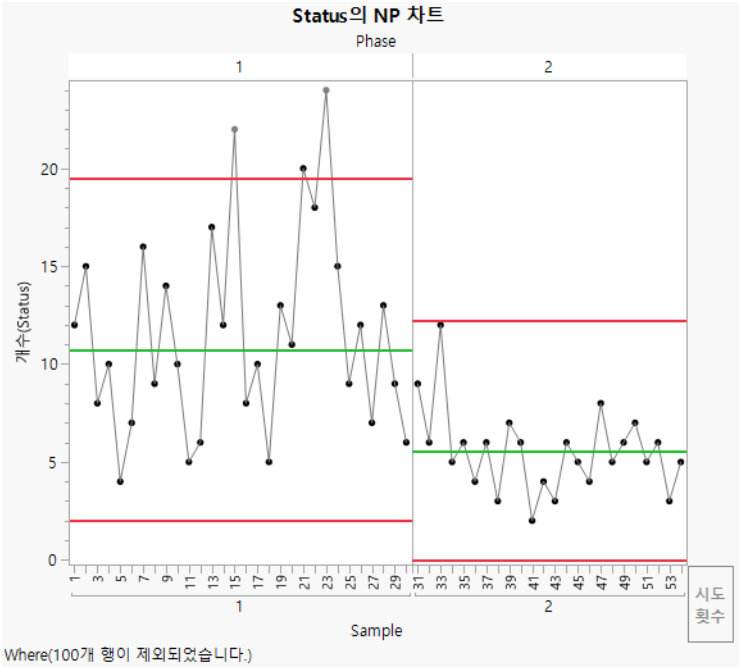


원래 관측값은 변동성이 큰 것으로 보이고 5 개의 관측값 (Sample 13, 15, 21, 22, 23) 이 관리 상한을 벗어납니다 . Sample 15 와 23 은 각각 새 재료 및 작업자가 공정에 투입되었음을 나타냅니다 . 단계가 끝날 때 제조 장비가 조정되었습니다 . 따라서 전체 계열에 대한 관리 한계를 단계 2 에서 관리 수준을 평가하는 데 사용하면 안 됩니다 .

각 단계에 대해 별도의 관리 한계를 계산하려면 다음을 수행하십시오 .

5. Phase 를 **단계** 영역으로 드래그합니다 .

그림 3.25 단계 (Phase) 별 NP 차트



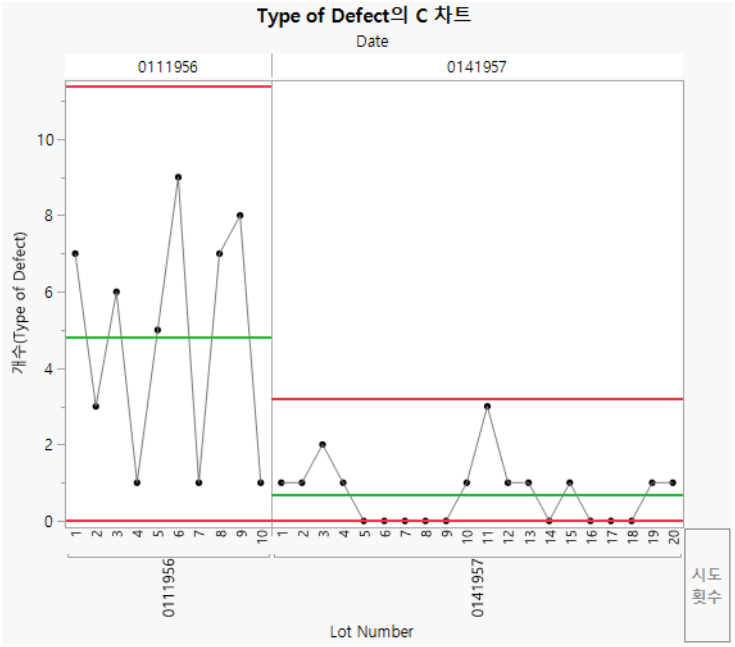
단계 변수를 포함하면 단계 2의 관리 한계가 단계 2의 데이터만 기반으로 한다는 것을 의미합니다. 단계 2의 모든 관측값이 관리 한계를 벗어나지 않습니다. 따라서 조정 후 공정이 관리 상태에 있다는 결론을 내릴 수 있습니다.

C 차트 예제

Cabinet Defects.jmp 샘플 데이터 테이블에는 두 기간 동안 캐비닛 제조 과정에서 발견된 다양한 결함 관련 데이터가 포함되어 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Cabinet Defects.jmp 를 엽니다.
2. **분석 > 품질 및 공정 > 관리도 > C 관리도**를 선택합니다.
3. Type of Defect 를 선택하고 **Y** 를 클릭합니다.
4. Lot Number 를 선택하고 **부분군**을 클릭합니다.
5. Date 를 선택하고 **단계**를 클릭합니다.
6. **확인**을 클릭합니다.

그림 3.26 단계가 있는 Type of Defect(결함 유형) 의 C 차트



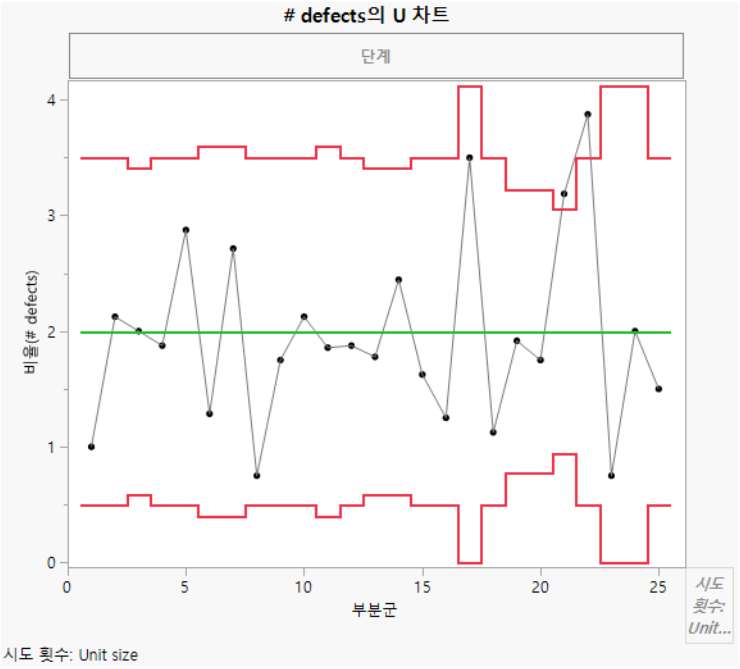
이제 서로 다른 두 날짜의 결과를 볼 수 있습니다. 둘 다 한계 내에 있는 것으로 보입니다. 다른 결함 유형 동작을 검토하려면 **사건 선택기 > Type of Defect**에서 다른 결함 유형을 선택한 후 한계가 업데이트되면 결과를 확인합니다.

U 차트 예제

Quality Control 샘플 데이터 폴더의 Braces.jmp 는 자동차 지지 가새 상자의 결함 수를 기록한 데이터입니다 . 가새 상자가 하나의 검사 단위입니다 . 검사한 상자 수 (일별) 는 부분군 표본 크기이며 달라질 수 있습니다 . 그림 3.27 의 U 차트는 부분군 표본 크기당 가새 결함 수를 모니터링합니다 . 상한 및 하한은 검사한 단위 수에 따라 달라집니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Braces.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > U 관리도**를 선택합니다 .
3. # defects 를 선택하고 **Y** 를 클릭합니다 .
4. Unit size 를 선택하고 **시행 횟수**를 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 3.27 # defects(결함 수) 의 U 차트



부분군 전체에서 표본 크기가 같지 않으므로 한계가 고르지 않습니다. 마지막 5개 표본 중 2개가 관리 한계를 벗어납니다.

G 차트 예제

G 차트는 희귀 사건이 예상보다 더 자주 발생하고 개입이 필요한지 여부를 파악할 수 있는 효과적인 방법입니다 . 자세한 내용은 " 희귀 사건 관리도 " 에서 확인하십시오 .

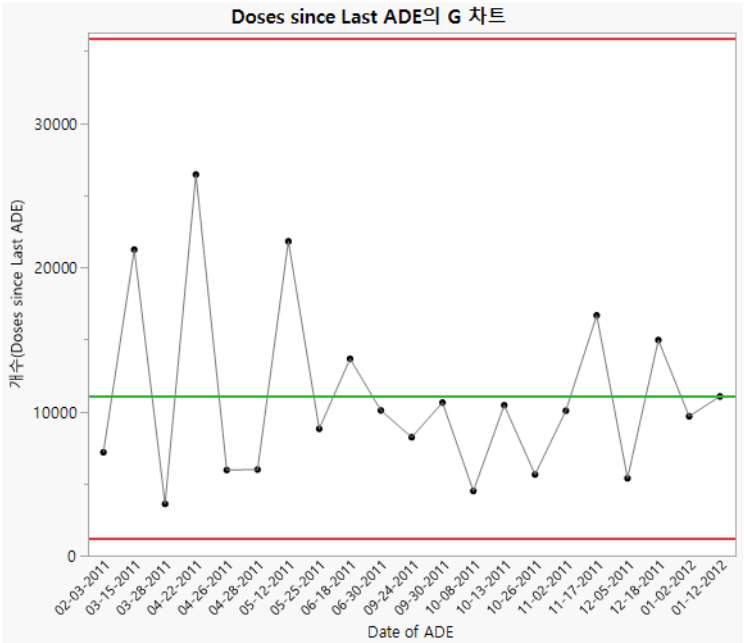
Adverse Reactions.jmp 샘플 데이터 테이블에는 병원 환자 그룹에 의해 보고된 ADE(약물 부작용 사례)에 대한 시뮬레이션 데이터가 포함되어 있습니다. ADE는 환자가 약물 복용 후 경험한 모든 유형의 부상 또는 반응입니다. 반응 날짜 및 마지막 반응 후 경과한 기간(일)이 기록되었습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Adverse Reactions.jmp를 엽니다.
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Doses since Last ADE 를 **Y** 역할로 드래그합니다 .
4. Date of ADE 를 **부분군** 역할로 드래그합니다 .

   "Doses since Last ADE 의 개별값 및 이동 범위 차트 " 가 나타납니다 .
5. 드롭다운 목록에서 **Shewhart 계량형** 대신 **희귀 사건**을 선택합니다 .

마지막 사건 이후 제공된 복용량을 보여 주는 "Doses since Last ADE 의 G 차트 " 가 나타납니다 .

그림 3.28 Doses since Last ADE(마지막 ADE 후 복용량) 의 G 차트



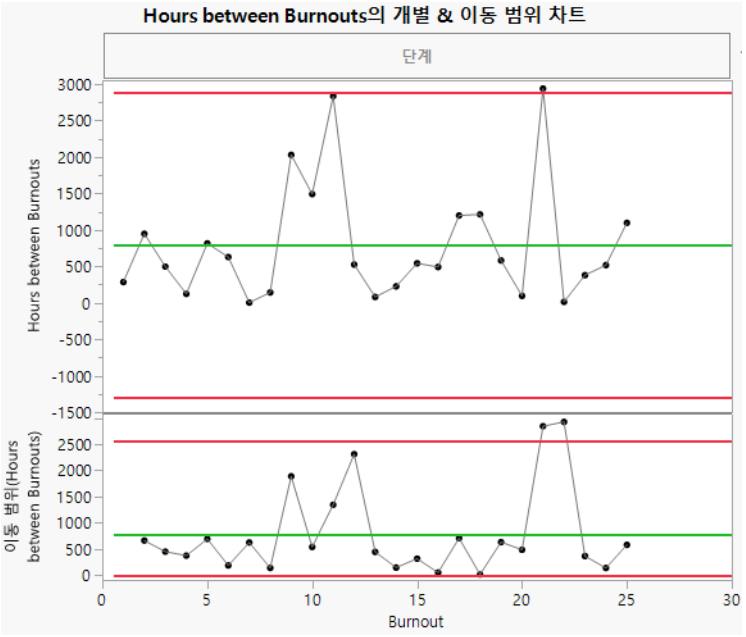
T 차트 예제

T 차트는 마지막 사건 이후 경과한 시간을 측정하는 데 사용됩니다 . 자세한 내용은 " 희귀 사건 관리도 " 에서 확인하십시오 .

Fan Burnout.jmp 샘플 데이터 테이블에는 팬 제조 공정에 대한 시뮬레이션 데이터가 포함되어 있습니다. 첫 번째 열은 연소된 각 팬을 식별합니다. 두 번째 열은 각 연소 사이의 시간을 식별합니다.

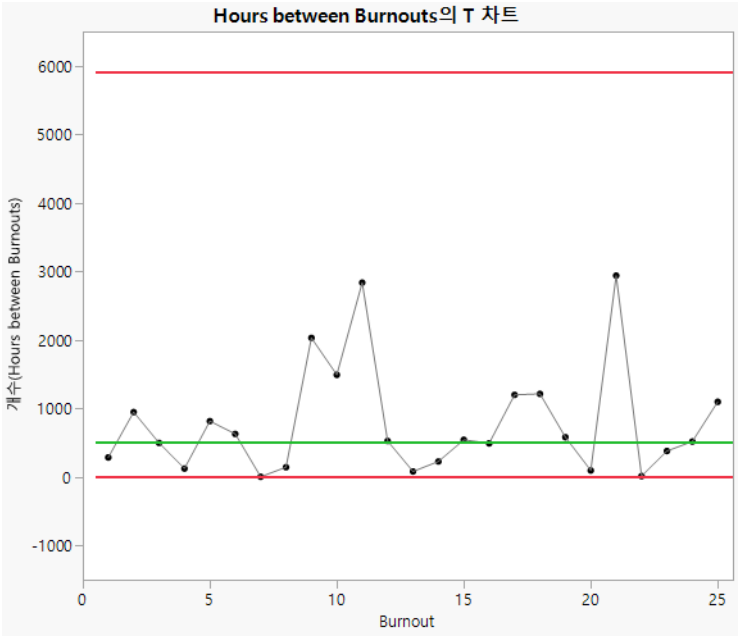
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Fan Burnout.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Hours between Burnouts 을 **Y** 역할로 드래그합니다 .
4. Burnout 을 **부분군** 역할로 드래그합니다 .

그림 3.29 Hours Between Burnouts(연소 사이 시간) 의 개별값 및 이동 범위 차트



5. 드롭다운 목록에서 **Shewhart 계량형** 대신 **희귀 사건**을 선택합니다 .
6. " 한계 " 아래의 **시그마**를 **음이항**에서 **Weibull** 로 변경합니다 .

그림 3.30 Hours Between Burnouts(연소 사이 시간) 의 T 차트



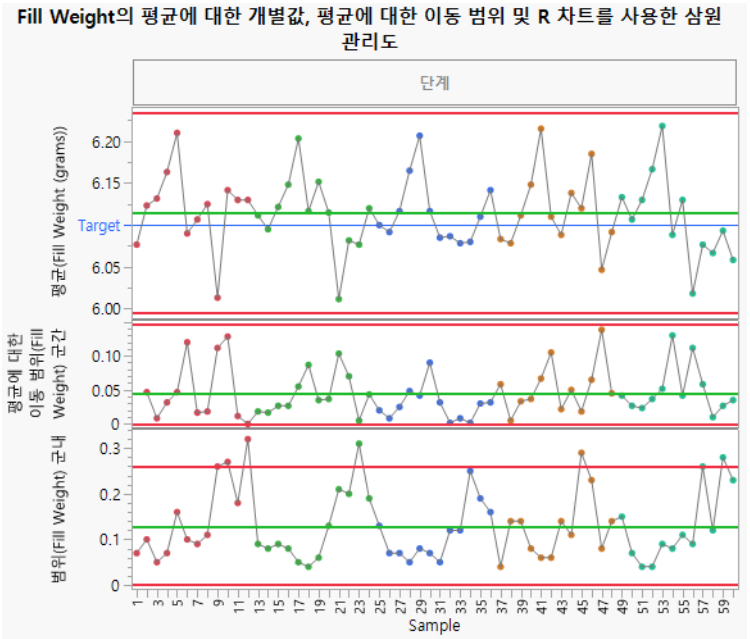
T 차트에서 모든 점이 관리 한계 내에 있는 것으로 나타납니다. 개별값 및 이동 범위 차트는 한계가 너무 좁아서 분석에 적합하지 않았습니다.

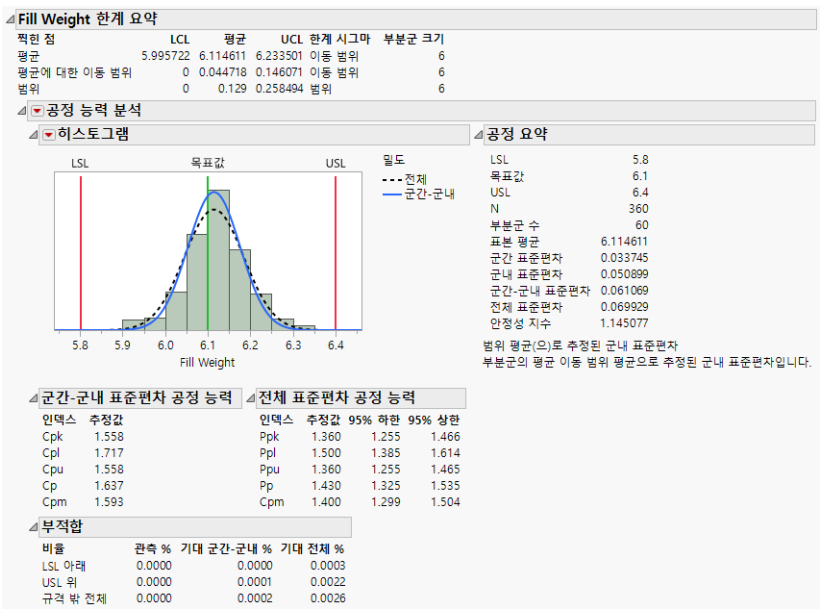
삼원 관리도 예제

삼원 관리도는 배치 간 변동과 배치 내 변동이 있을 때 유용합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Vial Fill Weights.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > 삼원 관리도**를 선택합니다 .
3. Fill Weight 를 선택하고 **Y** 를 클릭합니다 .
4. Sample 을 선택하고 **부분군**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 3.31 Fill Weight(충전 양) 에 대한 삼원 관리도





범위 차트와 평균 차트 사이에 이동 범위 차트가 나타납니다 . 이제 평균 (XBar) 차트의 한계는 각 표본 간의 이동 범위를 사용하여 계산됩니다 .

여러 관리도의 예

대화식 작업 영역의 "새 Y 차트" 옵션을 사용하거나 특정 관리도 시작 창에서 여러 공정을 지정하여 동일한 창에 여러 관리도를 생성할 수 있습니다. 이 예에서는 I/MR 관리도 시작 창을 사용한 후 관리 한계를 새 데이터 테이블에 저장합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Semiconductor Capability.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > IMR 관리도**를 선택합니다 .
3. "Processes" 옆의 삼각형을 클릭하여 모든 공정을 표시합니다 .
4. 처음 5 개 공정을 선택하고 **Y** 를 클릭합니다 .
5. **확인**을 클릭합니다 .

   보고서 창을 스크롤하면 선택한 5 개 공정 모두에 대한 개별값 및 이동 범위 관리도가 있습니다 . 공정에 " 규격 한계 " 열 특성이 있으므로 각 공정에 대한 " 공정 능력 분석 " 보고서도 있습니다 .
6. " 관리도 빌더 " 의 빨간색 삼각형을 클릭하고 **한계 저장 > 새 세로형 테이블에**를 선택합니다 .

그림 3.32 새 세로형 테이블

	공정	_Std Dev	_Sample Size	_Sigma	_Mean	_LCL	_UCL	_AvgR	_LCLR	_UCLR
1	NPN1	2.6350722541	1	2.6350722541	114.79276556	106.8875488	122.69798232	2.9733606353	0	9.7125774226
2	PNP1	59.430128559	1	59.430128559	313.06971647	134.77933079	491.36010214	67.059718964	0	219.05271249
3	PNP2	79.27035928	1	79.27035928	456.61569285	218.80461501	694.42677069	89.44702198	0	292.18155238
4	NPN2	2.1316521676	1	2.1316521676	115.74213802	109.34718152	122.13709452	2.4053118974	0	7.8570280886
5	PNP3	6.1609122855	1	6.1609122855	137.6145741	119.13183725	156.09731096	6.9518450733	0	22.708423829

각 Y 변수 행과 각 통계량 열이 포함된 새 데이터 테이블에 관리 한계가 저장됩니다 .

관리도 빌더에 대한 통계 상세 정보

참고 :

- 관리 한계는 기본적으로 3 으로 설정된 시그마 승수 K 를 사용하여 계산됩니다 . K 의 기본값을 변경하려면 파일 > 환경 설정 > 플랫폼 > 관리도 빌더에서 "K 시그마 " 환경 설정을 사용합니다 .
- 최대 $n = 50$ 의 표본 크기에 대해 JMP 는 Harter 연구 자료 (1960) 의 Table 2 에 정의된 관리도 상수 $d_2(n)$ 및 $d_3(n)$ 을 사용합니다 . 크기가 50 을 초과하는 표본의 경우 JMP 는 시그마 계산과 관리 한계 계산 둘 다에서 표본 크기 50 에 대한 관리도 상수 값을 사용합니다 .

- "Xbar-R 차트에 대한 관리 한계 "
- "Xbar-S 차트에 대한 관리 한계 "
- " 개별 측정값 및 이동 범위 차트에 대한 관리 한계 "
- "P 및 NP 차트에 대한 관리 한계 "
- "U 차트 및 C 차트에 대한 관리 한계 "

- "Levey-Jennings 차트 "
- "G 차트에 대한 관리 한계 "
- "T 차트에 대한 관리 한계 "
- " 삼원 관리도에 대한 시그마 계산 "

Xbar-R 차트에 대한 관리 한계

JMP 에서는 다음 계산식을 사용하여 Xbar-R 차트에 대한 관리 한계를 생성합니다 .

XBar 차트의 LCL = $\overline{X}_w - \frac{K\hat{\sigma}}{\sqrt{n_i}}$

XBar 차트의 UCL = $\overline{X}_w + \frac{K\hat{\sigma}}{\sqrt{n_i}}$

R 차트의 LCL = $\max\left(d_2(n_i)\hat{\sigma} - Kd_3(n_i)\hat{\sigma}, 0\right)$

R 차트의 UCL = $d_2(n_i)\hat{\sigma} + Kd_3(n_i)\hat{\sigma}$

R 차트의 중심선 : 기본적으로 i 번째 부분군의 중심선은 R_i 의 기대 추정값을 나타냅니다 . 이 값은 $d_2(n_i)\hat{\sigma}$ 로 계산됩니다 . 여기서 $\hat{\sigma}$ 는 σ 의 추정값입니다 .

Xbar-R 차트의 표준편차는 다음 계산식을 사용하여 추정됩니다 .

$$\hat{\sigma} = \frac{\frac{R_1}{d_2(n_1)} + \ldots + \frac{R_N}{d_2(n_N)}}{N}$$

다음은 각 요소에 대한 설명입니다 .

$\overline{X}_w$ = 부분군 평균의 가중 평균

K = 시그마 승수 (기본값 3)

σ = 공정 표준편차

n_i = i 번째 부분군의 표본 크기

$d_2(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 기대값

$d_3(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 표준편차

R_i = i 번째 부분군의 범위

N = 부분군 수 ($n_i \geq 2$)

Xbar-S 차트에 대한 관리 한계

JMP 에서는 다음 계산식을 사용하여 Xbar-S 차트에 대한 관리 한계를 생성합니다 .

XBar 차트의 LCL = $\overline{X}_w - \frac{K\hat{\sigma}}{\sqrt{n_i}}$

XBar 차트의 UCL = $\overline{X}_w + \frac{K\hat{\sigma}}{\sqrt{n_i}}$

S 차트의 LCL = $\max\left(c_4(n_i)\hat{\sigma} - Kc_5(n_i)\hat{\sigma}, 0\right)$

S 차트의 UCL = $c_4(n_i)\hat{\sigma} + Kc_5(n_i)\hat{\sigma}$

S 차트의 중심선 : 기본적으로 i 번째 부분군의 중심선은 s_i 의 기대 추정값을 나타냅니다 . 이 값은 $c_4(n_i)\hat{\sigma}$ 로 계산됩니다 . 여기서 $\hat{\sigma}$ 는 σ 의 추정값입니다 .

Xbar-S 차트의 표준편차 추정값은 다음과 같습니다 .

$$\hat{\sigma} = \frac{\frac{s_1}{c_4(n_1)} + \ldots + \frac{s_N}{c_4(n_N)}}{N}$$

다음은 각 요소에 대한 설명입니다 .

$\overline{X}_w$ = 부분군 평균의 가중 평균

K = 시그마 승수 (기본값 3)

σ = 공정 표준편차

n_i = i 번째 부분군의 표본 크기

$c_4(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 표준편차 기대값

$c_5(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 표준편차에 대한 표준 오차

N = 부분군 수 ($n_i \geq 2$)

s_i = i 번째 부분군의 표본 표준편차

개별 측정값 및 이동 범위 차트에 대한 관리 한계

개별 측정값 차트에 대한 관리 한계는 다음과 같이 계산됩니다 .

개별 측정값 차트의 LCL = $\overline{X} - K\hat{\sigma}$

개별 측정값 차트의 UCL = $\overline{X} + K\hat{\sigma}$

중앙값 이동 범위로 추정된 시그마를 갖는 개별 측정값 차트의 관리 한계는 다음과 같이 계산됩니다.

개별 측정값 차트의 LCL = $\overline{X} - K\hat{\sigma}_{MMR}$

개별 측정값 차트의 UCL = $\overline{X} + K\hat{\sigma}_{MMR}$

이동 범위 차트에 대한 관리 한계는 다음과 같이 계산됩니다.

이동 범위 차트의 LCL = $\max\{d_2(n)\hat{\sigma} - Kd_3(n)\hat{\sigma}, 0\}$

이동 범위 차트의 UCL = $d_2(n)\hat{\sigma} + Kd_3(n)\hat{\sigma}$

중앙값 이동 범위 차트에 대한 관리 한계는 다음과 같이 계산됩니다.

$LCL_{MMR} = \max(0, MMR - Kd_3(n)\hat{\sigma}_{MMR})$

$UCL_{MMR} = MMR + Kd_3(n)\hat{\sigma}_{MMR}$

개별 측정값 및 이동 범위 차트의 표준편차는 다음과 같이 추정됩니다.

$$\hat{\sigma} = \frac{\overline{MR}}{d_2(n)}$$

중앙값을 사용하는 경우 개별 측정값 및 이동 범위 차트의 표준편차는 다음과 같이 추정됩니다.

$\hat{\sigma}_{MMR} = MMR/0.954$

다음은 각 요소에 대한 설명입니다.

$\overline{X}$ = 개별 측정값의 평균

K = 시그마 승수 (기본값 3)

$\overline{MR}$ = $(MR_2+MR_3+...+MR_N)/(N-1)$ 로 계산된 비결측 이동 범위의 평균 ($MR_i = |x_i - x_{i-1}|$)

MMR = 비결측 이동 범위의 중앙값

σ = 공정 표준편차

$d_2(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 기대값

$d_3(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 표준편차

참고 : 관리도 빌더의 이동 범위 차트에는 $n = 2$ 범위 길이가 사용됩니다.

P 및 NP 차트에 대한 관리 한계

다음 계산식을 사용하여 LCL(관리 하한) 및 UCL(관리 상한) 이 계산됩니다 .

P 차트 LCL = $\max(\bar{p} - K\sqrt{\bar{p}(1-\bar{p})/n_i}, 0)$

P 차트 UCL = $\min(\bar{p} + K\sqrt{\bar{p}(1-\bar{p})/n_i}, 1)$

NP 차트 LCL = $\max(n_i\bar{p} - K\sqrt{n_i\bar{p}(1-\bar{p})}, 0)$

NP 차트 UCL = $\min(n_i\bar{p} + K\sqrt{n_i\bar{p}(1-\bar{p})}, n_i)$

다음은 각 요소에 대한 설명입니다 .

$\bar{p}$ = 전체 부분군에 대한 부적합품의 평균 비율

$$\bar{p} = \frac{n_1 p_1 + \ldots + n_N p_N}{n_1 + \ldots + n_n} = \frac{X_1 + \ldots + X_N}{n_1 + \ldots + n_N}$$

$n_i = i$ 번째 부분군의 항목 수

K = 시그마 승수 (기본값 3)

U 차트 및 C 차트에 대한 관리 한계

다음 계산식을 사용하여 LCL(관리 하한) 및 UCL(관리 상한) 이 계산됩니다 .

U 차트 LCL = $\max(\bar{u} - K\sqrt{\bar{u}/n_i}, 0)$

U 차트 UCL = $\bar{u} + K\sqrt{\bar{u}/n_i}$

C 차트 LCL = $\max(n_i\bar{u} - K\sqrt{n_i\bar{u}}, 0)$

C 차트 UCL = $n_i\bar{u} + K\sqrt{n_i\bar{u}}$

한계는 n_i 에 따라 달라집니다 .

$u_i = i$ 번째 부분군의 단위별 결함 수 . 일반적으로 $u_i = c_i/n_i$ 입니다 .

K = 시그마 승수 (기본값 3)

$c_i = i$ 번째 부분군의 총 결함 수

$n_i = i$ 번째 부분군의 검사 단위 수

$\bar{u}$ = 전체 부분군에 대한 단위별 평균 결함 수 . $\bar{u}$ 양 (quantity) 은 가중 평균으로 계산됩니다 .

$$\bar{u} = \frac{n_1 u_1 + \ldots + n_N u_N}{n_1 + \ldots + n_N} = \frac{c_1 + \ldots + c_N}{n_1 + \ldots + n_N}$$

N = 부분군 수

Levey-Jennings 차트

Levey-Jennings 차트는 장기 시그마를 기반으로 한 관리 한계와 공정 평균을 보여 줍니다. 관리 한계는 중심선에서 K*s 거리에 배치되며 기본적으로 K = 3 입니다.

Levey-Jennings 차트에 대한 표준편차 s 는 분포 플랫폼의 표준편차와 동일한 방식으로 계산됩니다.

$$s = \sqrt{\sum_{i=1}^{N} \frac{(y_i - \bar{y})^2}{N-1}}$$

자세한 내용은 Levey and Jennings(1950) 및 Westgard(2002) 연구 자료에서 확인하십시오.

G 차트에 대한 관리 한계

음이항 분포는 기하(Poisson) 분포의 확장이며 Poisson과 관련하여 과대산포를 허용합니다. 음이항 분포를 사용하면 개수 데이터에 대해 정확 관리 한계와 근사 관리 한계를 모두 구성할 수 있습니다. 근사 관리 한계는 음이항에 대한 카이제곱 근사를 기반으로 구할 수 있습니다. 모든 데이터는 부분군 크기에 관계없이 개별 관측값으로 사용됩니다.

모수가 (μ, k)인 음이항 분포의 X를 가정해 보겠습니다. 그러면 다음과 같이 구성할 수 있습니다.

$$P(X \le r) \sim P\left(\chi^2_v \le \frac{2r+1}{1+\mu k}\right)$$

다음은 각 요소에 대한 설명입니다.

χ^2_v = 자유도가 $v = 2\mu/(1+\mu k)$ 인 카이제곱 변량

이 근사를 기반으로 관리 상한 및 관리 하한의 근사값을 산출할 수 있습니다. 한 방향의 명목형 수준 α 제 1 종 오류 확률의 경우 근사 관리 상한은 다음 방정식이 참이 되는 한계 (UCL) 입니다.

$$P(X > \mathrm{UCL}) = 1 - P\left(\chi^2_v \le \frac{2\mathrm{UCL}+1}{1+\mu k}\right) = \alpha$$

마찬가지로 근사 LCL(관리 하한) 은 다음 방정식이 참이 되는 한계입니다.

$$P(X < \mathrm{LCL}) = 1 - P\left(\chi^2_v \ge \frac{2\mathrm{LCL}+1}{1+\mu k}\right) = \alpha$$

따라서 다음 계산식을 사용하여 근사 수준의 LCL(관리 하한) 및 UCL(관리 상한)이 계산됩니다.

$$\mathrm{UCL} = \frac{\chi^2_{v,\,1-\alpha}(1+\mu k) - 1}{2}$$

$$LCL = \max\left\{0, \frac{\chi^2_{v,\alpha}(1+\mu k)-1}{2}\right\}$$

다음은 각 요소에 대한 설명입니다 .

$\chi^2_{v,1-\alpha}(\chi^2_{v,\alpha})$ = 자유도가 $v = 2\mu/(1+\mu k)$ 인 카이제곱 분포의 상위 (하위) 백분위수 . 음수 관리 하한은 0 으로 설정할 수 있습니다 .

음이항 관리 한계에 대한 자세한 내용은 Hoffman 연구 자료(2003)에서 확인하십시오.

T 차트에 대한 관리 한계

형상 모수 및 척도 모수의 추정값은 데이터에서 계산되며 Weibull 분포의 백분위수를 구하는 데 사용됩니다.

참고 : Weibull 분포 모수를 추정할 때 반응 값이 0 인 부분군에는 가중치가 0 으로 지정됩니다 .

다음 통계량을 정의합니다.

p1 = Normal (0,1) 에 대한 normalDist(-*K*)

p2 = Normal (0,1) 에 대한 normalDist(0)

p3 = Normal (0,1) 에 대한 normalDist(*K*)

이제 다음 계산식을 사용하여 한계가 계산됩니다.

LCL = Weibull Quantile (p1, β, α)

CL = Weibull Quantile (p2, β, α)

UCL = Weibull Quantile (p3, β, α)

다음은 각 요소에 대한 설명입니다.

β = Weibull Quantile 함수의 형상 모수 , α = Weibull Quantile 함수의 척도 모수

K = 시그마 승수 (기본값 3)

Weibull Quantile 함수에 대한 자세한 내용은 **도움말 > 스크립트 인덱스**에서 확인하십시오.

삼원 관리도에 대한 시그마 계산

범위 평균에 기반한 군내 표준편차

범위 평균을 사용하여 추정되는 삼원 관리도의 군내 표준편차 추정값은 평균에 대한 개별값, 평균에 대한 이동 범위 및 R 차트에 사용할 수 있습니다.

$$\hat{\sigma}_{\text{within}} = \frac{\frac{R_1}{d_2(n_1)} + \ldots + \frac{R_N}{d_2(n_N)}}{N}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

$R_i = i$ 번째 부분군의 범위

$n_i = i$ 번째 부분군의 표본 크기

$d_2(n_i)$ = 단위 표준편차를 갖는 n_i 개 독립 정규 분포 변수의 범위 기대값

N = 부분군 수 ($n_i \geq 2$)

비편향 표준편차의 평균에 기반한 군내 표준편차

비편향 표준편차의 평균을 사용하여 추정되는 삼원 관리도의 군내 표준편차 추정값은 평균에 대한 개별값, 평균에 대한 이동 범위 및 S 차트에 사용할 수 있습니다.

$$\hat{\sigma}_{\text{within}} = \frac{\frac{s_1}{c_4(n_1)} + \ldots + \frac{s_N}{c_4(n_N)}}{N}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

$s_i = i$ 번째 부분군의 표본 표준편차

$n_i = i$ 번째 부분군의 표본 크기

$c_4(n_i)$ = 단위 표준편차를 갖는 n_i 개 독립 정규 분포 변수의 표준편차 기대값

N = 부분군 수 ($n_i \geq 2$)

군간 표준편차

삼원 관리도의 군간 표준편차 추정값은 부분군 평균의 이동 범위를 사용하여 추정됩니다.

$$\hat{\sigma}_{\text{between}} = \sqrt{\left(\frac{\overline{MR}}{d_2(2)}\right)^2 - \frac{\hat{\sigma}^2_{\text{within}}}{H}}$$

이 계산식에서는 다음과 같은 표기를 사용합니다 .

$\overline{MR}$ = (MR_2+MR_3+...+MR_N)/(*N-1*) 로 계산된 비결측 이동 범위의 평균 ($MR_i = |y_i - y_{i-1}|$)

$d_2(2)$ = 단위 표준편차를 갖는 두 독립 정규 분포 변수의 범위 기대값

$H = \dfrac{N}{\frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_N}}$, 부분군 표본 크기의 조화 평균

참고 : 음수 값으로 추정된 군간 표준편차는 0 으로 설정됩니다 .

군간 - 군내 표준편차

삼원 관리도의 군간 - 군내 표준편차 추정값은 군내 표준편차 및 군간 표준편차 추정값의 조합을 사용하여 추정됩니다 .

$$\hat{\sigma}_{\text{between-and-within}} = \sqrt{\hat{\sigma}^2_{\text{within}} + \hat{\sigma}^2_{\text{between}}}$$

4 장

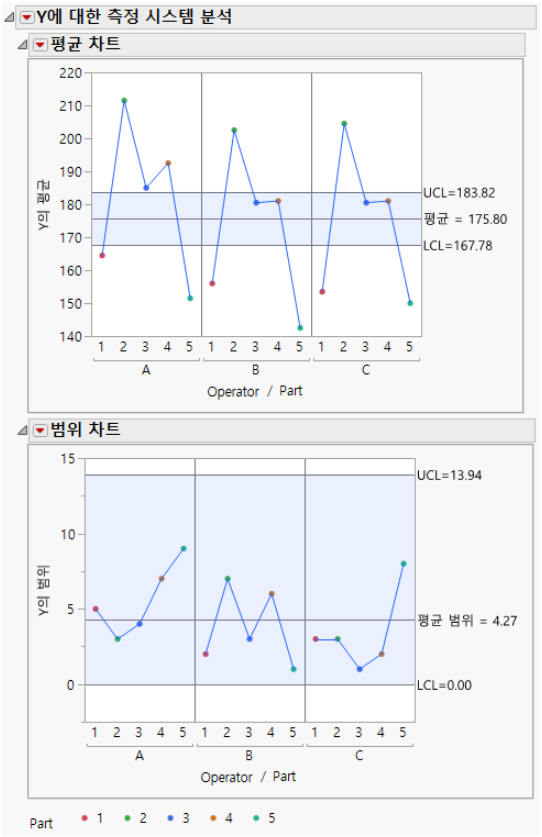
측정 시스템 분석

EMP 방법을 사용하여 연속형 측정 공정 평가

MSA(측정 시스템 분석) 플랫폼은 측정 시스템의 정밀도, 일관성 및 편향을 평가합니다. 공정 자체를 연구하려면 먼저 공정을 정확하고 정밀하게 측정할 수 있는지 확인해야 합니다. 만일 변동의 대부분이 측정 공정 자체에 기인하는 경우 공정에 대해 신뢰할 수 있는 정보를 얻지 못합니다. MSA 를 사용하여 측정 시스템의 성능을 확인할 수 있습니다.

이 장에서는 EMP 방법에 대해 설명합니다. 게이지 R&R 방법은 "계량형 게이지 차트"에서 설명됩니다.

그림 4.1 측정 시스템 분석의 예



목차

측정 시스템 분석 개요 97
측정 시스템 분석의 예 97
측정 시스템 분석 플랫폼 시작 100
데이터 형식 101
측정 시스템 분석 플랫폼 옵션 102
평균 차트 103
범위 차트 또는 S 관리도 104
EMP 결과 104
유효 해상도 106
변화 감지 프로파일러 106
편향 비교 111
검사-재검사 오차 비교 111
측정 시스템 분석의 추가 예 112
측정 시스템 분석에 대한 통계 상세 정보 118
급내 상관 및 확률 오차 계산 118

측정 시스템 분석 개요

측정 시스템 분석 플랫폼의 EMP(Evaluating the Measurement Process) 방법은 주로 Donald J. Wheeler 의 저서 EMP III Using Imperfect Data(2006)에 제시된 방법을 기반으로 합니다. EMP 방법은 해석하기 쉬운 시각적 정보와 결과를 제공하며, 측정 시스템의 성능을 최대한 발휘할 수 있도록 개선하는 데 도움이 됩니다.

게이지 R&R 방법은 작업자 변동(재현성) 및 측정 변동(반복성)으로 인한 변동이 어느 정도인지 분석합니다. 게이지 R&R은 모형의 균형 여부에 관계없이 교차 모형과 내포 모형의 여러 조합에 사용할 수 있습니다. 자세한 내용은 "계량형 게이지 차트"에서 확인하십시오.

Six Sigma DMAIC 방법론에서 MSA(측정 시스템 분석)는 측정 단계를 다루고 공정 동작 차트(관리도)는 관리 단계를 다룹니다. MSA는 미래 결과를 예측하고 특성화하는 데 도움이 됩니다. MSA에서 수집한 정보를 사용하여 공정 동작 차트를 해석하고 구성할 수 있습니다.

관리도에 대한 자세한 내용은 "관리도 빌더"에서 확인하십시오.

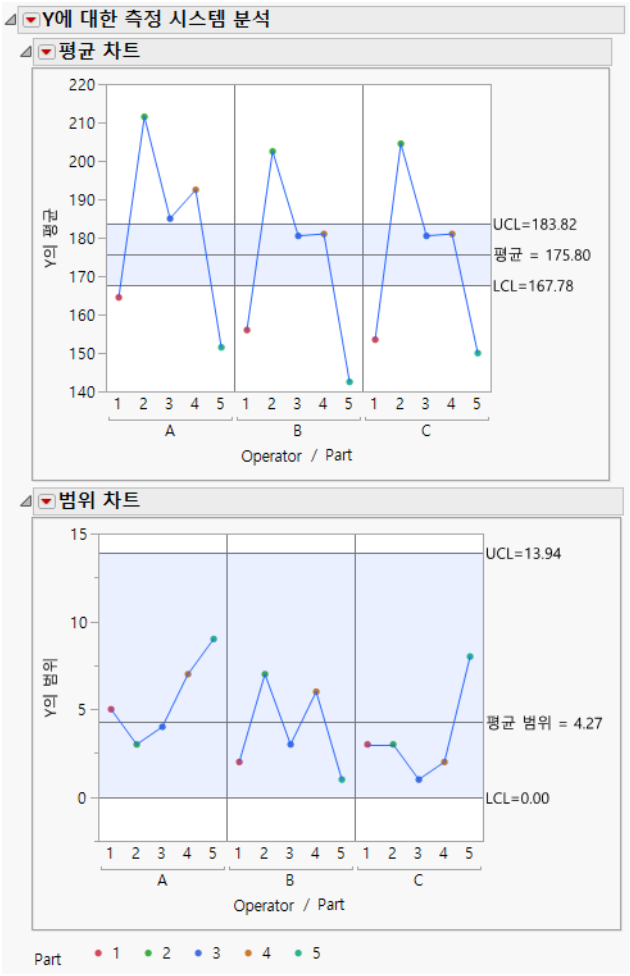
측정 시스템 분석의 예

이 예에서는 세 명의 작업자가 동일 부품 5 개를 측정했습니다 . 측정값에서 발견된 변동량을 기반으로 측정 시스템의 성능을 확인합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/Gasket.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 측정 시스템 분석**을 선택합니다 .
3. Y 를 **Y, 반응** 역할에 할당합니다 .
4. Part 를 **부품 , 표본 ID** 역할에 할당합니다 .
5. Operator 를 **X, 그룹화** 역할에 할당합니다 .

   MSA 방법은 **EMP**, **산포 차트 유형**은 **범위** , **모형 유형**은 **교차**로 각각 설정되어 있습니다 .
6. **확인**을 클릭합니다 .

그림 4.2 MSA 초기 보고서



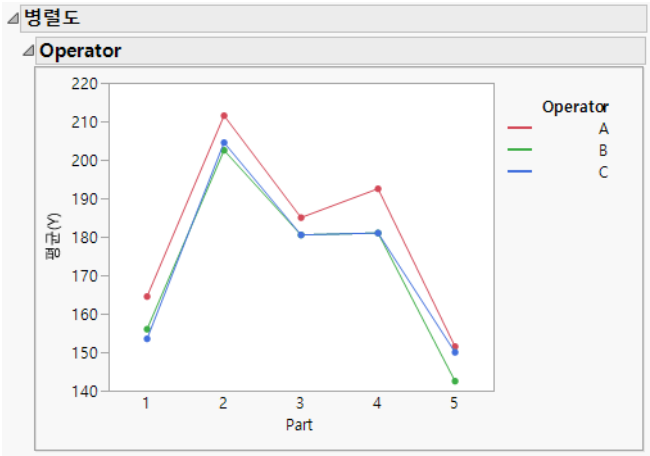
" 평균 차트 " 에는 각 작업자와 부품 조합에 대한 평균 측정값이 표시됩니다 . 이 예에서는 부품 측정값 평균이 대부분 관리 한계를 벗어납니다 . 이 결과는 부품 간 변동을 감지할 수 있음을 나타내므로 바람직합니다 .

" 범위 차트 " 에는 각 작업자와 부품 조합에 대한 변동성이 표시됩니다 . 이 예에서는 범위가 관리 한계 내에 있습니다 . 이 결과는 작업자가 동일한 방식 및 비슷한 변동으로 부품을 측정하고 있음을 나타내므로 바람직합니다 .

차트 아래 범례에 각 부품의 색상 코딩이 표시됩니다 .

7. "Y 에 대한 측정 시스템 분석 " 옆의 빨간색 삼각형을 클릭하고 **병렬도**를 선택합니다 .

그림 4.3 Operator(작업자) 와 Part(부품) 에 대한 병렬도

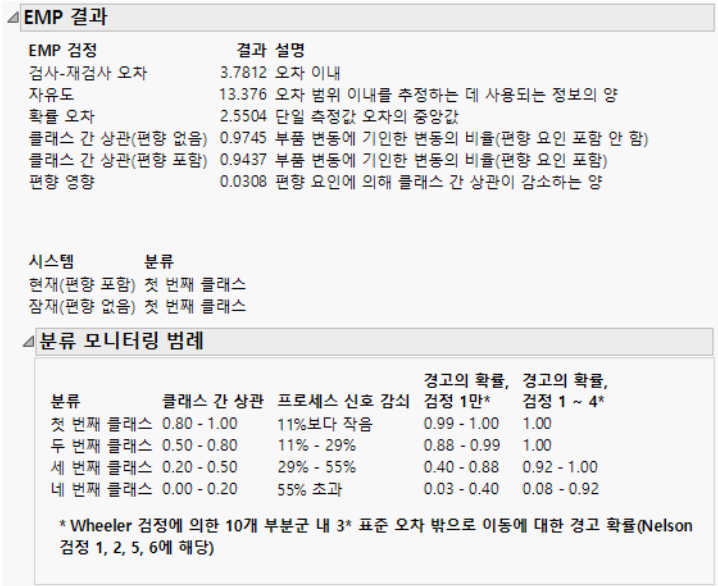


" 병렬도 " 차트에는 작업자별로 각 부품의 평균 측정값이 표시됩니다 . 전체적으로 선이 평행하고 큰 교차가 없으므로 작업자와 부품 간에 교호작용이 없다는 결론을 내릴 수 있습니다 .

팁 : 교호작용은 추가 조사가 필요한 심각한 문제를 나타냅니다 .

8. "Y 에 대한 측정 시스템 분석 " 옆의 빨간색 삼각형을 클릭하고 **EMP 결과**를 선택합니다 .

그림 4.4 EMP 결과 보고서



EMP 결과

EMP 검정	결과	설명
검사-재검사 오차	3.7812	오차 이내
자유도	13.376	오차 범위 이내를 추정하는 데 사용되는 정보의 양
확률 오차	2.5504	단일 측정값 오차의 중앙값
클래스 간 상관(편향 없음)	0.9745	부품 변동에 기인한 변동의 비율(편향 요인 포함 안 함)
클래스 간 상관(편향 포함)	0.9437	부품 변동에 기인한 변동의 비율(편향 요인 포함)
편향 영향	0.0308	편향 요인에 의해 클래스 간 상관이 감소하는 양

시스템	분류
현재(편향 포함)	첫 번째 클래스
잠재(편향 없음)	첫 번째 클래스

분류 모니터링 범례

분류	클래스 간 상관	프로세스 신호 감쇠	경고의 확률, 검정 1만*	경고의 확률, 검정 1 ~ 4*
첫 번째 클래스	0.80 - 1.00	11%보다 작음	0.99 - 1.00	1.00
두 번째 클래스	0.50 - 0.80	11% - 29%	0.88 - 0.99	1.00
세 번째 클래스	0.20 - 0.50	29% - 55%	0.40 - 0.88	0.92 - 1.00
네 번째 클래스	0.00 - 0.20	55% 초과	0.03 - 0.40	0.08 - 0.92

* Wheeler 검정에 의한 10개 부분군 내 3* 표준 오차 밖으로 이동에 대한 경고 확률(Nelson 검정 1, 2, 5, 6에 해당)

"EMP 결과" 보고서는 측정 시스템을 평가하고 분류하는 데 도움이 되는 여러 통계량을 계산합니다 . "급내 상관"은 총 변동에서 부품이 기여하는 비율을 나타냅니다 .

"EMP 결과 " 보고서를 통해 다음과 같은 결론을 내릴 수 있습니다 .

- " 급내 상관 " 값이 1 에 가까우며 이는 대부분의 변동이 측정 시스템이 아니라 부품에서 발생한다는 것을 나타냅니다 .
- 분류가 " 첫 번째 클래스 " 이며 이는 공정 신호 강도가 11% 미만으로 약화됨을 의미합니다 .
- 검정 1 만 사용하여 경고를 감지할 확률은 99% 이상입니다 .
- 검정 1 ~ 4 를 사용하여 경고를 감지할 확률은 100% 입니다 .

참고: 검정 및 공정 변화 감지에 대한 자세한 내용은 "변화 감지 프로파일러"에서 확인하십시오.

작업자와 부품 간에 교호작용이 없고 측정값 변동도 거의 없습니다(분류 = 첫 번째 클래스). 따라서 측정 시스템이 매우 잘 수행되고 있다는 결론을 내릴 수 있습니다.

측정 시스템 분석 플랫폼 시작

분석 > 품질 및 공정 > 측정 시스템 분석을 선택하여 측정 시스템 분석 플랫폼을 시작할 수 있습니다.

그림 4.5 측정 시스템 분석 창

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

측정 시스템 분석 창에는 다음 기능이 포함되어 있습니다 .

열 선택 현재 데이터 테이블의 모든 변수를 나열합니다 . 선택한 열을 역할로 이동합니다 .

MSA 방법 **EMP**(Evaluating the Measurement Process) 또는 **게이지 R&R** 중 사용할 방법을 선택합니다. 이 장에서는 **EMP** 방법에 대해 설명합니다. **게이지 R&R** 방법에 대한 자세한 내용은 " 계량형 게이지 차트 " 에서 확인하십시오.

산포 차트 유형 변동을 표시할 차트 유형을 지정합니다. **범위** 또는 **표준편차** 옵션을 선택합니다.

참고 : EMP 방법의 경우 산포 차트 유형에 따라 EMP 결과 보고서의 통계량을 계산하는 방법이 결정됩니다. **범위** 옵션을 선택하고 1 요인 또는 2 요인의 균형 잡힌 교차 모형을 사용하는 경우 이 보고서의 통계량은 범위를 기반으로 합니다. 그렇지 않은 경우 이 보고서의 통계량은 표준편차를 기반으로 합니다.

모형 유형 다음과 같은 모형 유형을 지정합니다.

주 명목형 또는 순서형 모델링 유형의 변수를 주효과로 처리합니다.

교차 모든 요인의 모든 수준이 다른 모든 요인의 모든 수준과 함께 발생하면 모형이 교차됩니다.

두 개의 요인 교호작용과 교차 두 요인의 각 수준이 다른 요인의 모든 수준과 함께 발생하면 모형이 교차됩니다.

내포 한 요인의 모든 수준이 다른 요인의 단일 수준 내에만 나타나면 모형이 내포됩니다.

교차 후 내포 (3 개 요인만) 3 개 요인에 대해 요인이 교차된 후 내포됩니다.

내포 후 교차 (3 개 요인만) 3 개 요인에 대해 요인이 내포된 후 교차됩니다.

옵션 다음 옵션이 포함되어 있습니다.

분석 설정 REML 최대 반복 수와 수렴을 설정합니다.

알파 지정 1-α 신뢰 수준을 지정합니다.

Y, 반응 측정값 열입니다.

부품 , 표본 ID 부품 또는 단위를 지정하는 열입니다.

X, 그룹화 그룹화 변수를 나타내는 열입니다.

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다.

데이터 형식

측정 시스템 분석 플랫폼을 사용하려면 모든 반응 측정값이 단일 반응 열에 있어야 합니다. 반응이 여러 열에 기록되는 경우도 있습니다. 이때 각 행은 한 설계 요인의 수준이고 각 열은 다른 설계 요인의 수준입니다. 측정 시스템 분석 플랫폼을 실행하기 전에 이 형식의 데이터를 쌓아야 합니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

측정 시스템 분석 플랫폼 옵션

플랫폼 옵션은 " 측정 시스템 분석 " 옆의 빨간색 삼각형 메뉴에 나타납니다 . 옵션을 선택하면 MSA 보고서 창에 해당 그래프 또는 보고서가 생성됩니다 . 옵션을 선택 취소하면 그래프 또는 보고서가 제거됩니다 . 다음 옵션 중에서 선택합니다 .

평균 차트 부품과 *X* 변수의 각 조합에 대한 평균 측정값을 보여 주는 그림입니다 . 평균 차트는 측정 변동이 있더라도 제품 변동을 감지하는 데 도움이 됩니다 . 평균 차트에서는 부품 간 변동을 감지하는 관리이탈 데이터가 바람직합니다 . 자세한 내용은 " 평균 차트 " 에서 확인하십시오 .

범위 차트 부품과 *X* 변수의 각 조합에 대한 변동 통계량을 보여 주는 그림입니다 . 시작 창에서 " 산포 차트 유형 " 을 **범위**로 선택한 경우에만 나타납니다 . 범위 차트는 부분군 내의 일관성을 확인하는 데 도움이 됩니다 . 범위 차트에서는 오차의 동질성을 나타내는 한계 내의 데이터가 바람직합니다 . 자세한 내용은 " 범위 차트 또는 S 관리도 " 에서 확인하십시오 .

S 관리도 부품과 *X* 변수의 각 조합에 대한 표준편차 통계량을 보여 주는 그림입니다 . 시작 창에서 " 산포 차트 유형 " 을 **표준편차**로 선택한 경우에만 나타납니다 . S 관리도는 부분군 내의 일관성을 확인하는 데 도움이 됩니다 . S 관리도에서는 오차의 동질성을 나타내는 한계 내의 데이터가 바람직합니다 . 자세한 내용은 " 범위 차트 또는 S 관리도 " 에서 확인하십시오 .

병렬도 각 부품에 대한 측정값 평균을 반영하는 중첩 그림입니다 . 선이 비교적 평행하지 않거나 교차하면 부품과 *X* 변수 간에 교호작용이 있을 수 있습니다 .

팁 : 교호작용은 추가 조사가 필요한 심각한 문제를 나타냅니다 . 예를 들어 부품과 작업자 간의 교호작용은 작업자가 대체로 각 부품마다 다르게 측정한다는 것을 의미합니다 . 따라서 측정 변동을 예측할 수 없습니다 . 이 문제는 작업자가 부품에 대해 동일한 패턴이나 프로파일을 갖지 않는 이유를 찾기 위해 추가 조사가 필요합니다 .

EMP 결과 측정 시스템을 평가하고 분류하는 데 도움이 되는 여러 통계량을 계산하는 보고서입니다 . 자세한 내용은 "EMP 결과 " 에서 확인하십시오 .

유효 해상도 측정 시스템의 해상도 결과를 포함하는 보고서입니다 . 자세한 내용은 " 유효 해상도 " 에서 확인하십시오 .

편향 비교 *X* 변수의 평균이 서로 다른지 여부를 검정하기 위한 평균 분석 차트입니다 . 자세한 내용은 " 편향 비교 " 에서 확인하십시오 .

검사 - 재검사 오차 비교 그룹의 검사 - 재검사 오차 수준이 서로 다른지 여부를 검정하기 위한 " 분산에 대한 평균 분석 " 또는 " 평균 범위 분석 " 차트입니다 . 자세한 내용은 " 검사 - 재검사 오차 비교 " 에서 확인하십시오 .

변화 감지 프로파일러 공정 동작 차트에서 경고를 받을 확률을 확인하기 위해 조정할 수 있는 대화식 차트 집합입니다 . 자세한 내용은 " 변화 감지 프로파일러 " 에서 확인하십시오 .

분산 성분 주어진 모형에 대한 분산 성분 추정값을 포함하는 보고서입니다 . 이 보고서의 계산은 범위가 아니라 분산을 기반으로 합니다 . 균형 데이터는 EMS 방법을 사용합니다 . 불균형 데이터는 REML 방법을 사용합니다 .

참고 : 이 보고서는 Bayes 분산 성분 추정값을 계산하지 않는다는 점을 제외하고 계량형 차트 플랫폼의 분산 성분 보고서와 비슷합니다 . 자세한 내용은 " 분산 성분 " 에서 확인하십시오 .

EMP 게이지 R&R 결과 측정 변동성을 부품 변동과 측정 시스템 변동으로 분할하는 보고서입니다 . 이 보고서의 계산은 범위가 아니라 분산을 기반으로 합니다 . 음의 분산 성분이 0 으로 설정되므로 0 값은 결과의 이상치를 나타낼 수 있습니다 .

참고 : 이 보고서는 계량형 차트 플랫폼의 게이지 R&R 보고서와 비슷합니다 . 그러나 기본적으로는 재현성 계산에 교호작용이 포함되지 않습니다 . 재현성 계산에 교호작용을 포함하도록 지정하려면 " 재현성에 교호작용 포함 " 플랫폼 환경 설정을 선택합니다 . 이 환경 설정은 파일 > 환경 설정 > 플랫폼 > EMP 측정 시스템 분석에 있습니다 . 게이지 R&R 연구에 대한 자세한 내용은 " 게이지 R&R 방법 정보 " 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

평균 차트

" 평균 차트 " 옆의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

총 평균 표시 차트에 *Y* 변수의 전체 평균을 그립니다 .

연결된 평균 표시 모든 평균 측정값을 연결하는 선을 그립니다 .

관리 한계 표시 UCL(관리 상한) 및 LCL(관리 하한) 을 나타내는 선을 그리고 해당 값에 라벨을 지정합니다 . 평균 차트의 관리 한계에는 XBar 관리도와 동일한 계산이 사용됩니다 . 자세한 내용은 "Xbar-R 차트에 대한 관리 한계 " 에서 확인하십시오 .

관리 한계 음영 표시 UCL 과 LCL 사이에 음영을 추가합니다 .

구분 기호 표시 *X* 변수 사이를 구분하는 세로선을 그립니다 .

데이터 표시 차트에 데이터 점을 추가합니다 .

참고 : 두 가지 방법 중 하나로 평균 차트의 변수를 바꿀 수 있습니다 . 한 가지 방법은 한 축의 변수를 다른 축으로 드래그한 후 놓아 기존 변수를 바꾸는 것이고 , 다른 방법은 연결된 데이터 테이블의 " 열 " 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다 .

범위 차트 또는 S 관리도

" 범위 차트 " 또는 "S 관리도 " 옆의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

평균 산포 표시 차트에 평균 범위 또는 표준편차를 그립니다 .

연결된 점 표시 모든 범위 또는 표준편차를 연결하는 선을 그립니다 .

관리 한계 표시 UCL(관리 상한) 및 LCL(관리 하한) 을 나타내는 선을 그리고 해당 값에 라벨을 지정합니다 . 범위 차트에 사용되는 한계 계산에 대한 자세한 내용은 "Xbar-R 차트에 대한 관리 한계 " 에서 확인하십시오 . S 관리도에 사용되는 한계 계산에 대한 자세한 내용은 "Xbar-S 차트에 대한 관리 한계 " 에서 확인하십시오 .

관리 한계 음영 표시 UCL 과 LCL 사이에 음영을 추가합니다 .

구분 기호 표시 *X* 변수 사이를 구분하는 세로선을 그립니다 .

참고 : 두 가지 방법 중 하나로 범위 차트 또는 S 관리도의 변수를 바꿀 수 있습니다 . 한 가지 방법은 한 축의 변수를 다른 축으로 드래그한 후 놓아 기존 변수를 바꾸는 것이고 , 다른 방법은 연결된 데이터 테이블의 " 열 " 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다 .

EMP 결과

참고 : MSA 방법 및 **산포 차트 유형**을 각각 **EMP** 와 **범위** 로 선택하고 , 1 요인 또는 2 요인의 균형 잡힌 교차 모형을 사용하는 경우 이 보고서의 통계량은 범위를 기반으로 합니다 . 그렇지 않은 경우 이 보고서의 통계량은 분산을 기반으로 합니다 .

"EMP 결과 " 보고서는 측정 시스템을 평가하고 분류하는 데 도움이 되는 여러 통계량을 계산합니다 . 이 보고서를 사용하여 다음을 확인할 수 있습니다 .

- 공정 차트에 미치는 영향
- 설정할 검정
- 공정 신호의 감쇄 비율
- 편향 요인에 의해 시스템에 미치는 영향 및 잠재적 급내 상관계수의 감소 정도

"EMP 결과 " 보고서에는 다음 계산이 포함됩니다 .

검사 - 재검사 오차 측정 변동 또는 반복성을 나타냅니다 (군내 오차 또는 순수 오차라고도 함).

자유도 군내 오차를 추정하는 데 사용되는 정보의 양을 나타냅니다 .

확률 오차 단일 측정값 오차의 중앙값입니다. 측정 해상도 품질을 나타내며 측정값을 기록할 때 사용할 자릿수를 결정하는 데 도움이 됩니다. 자세한 내용은 "유효 해상도"에서 확인하십시오.

급내 상관 총 변동에서 부품이 기여하는 비율을 나타냅니다. 측정 변동이 거의 없는 경우 이 값은 1 에 더 가깝습니다.

급내 상관 (편향 없음) 결과를 계산할 때 편향 또는 교호작용 요인을 고려하지 않습니다.

급내 상관 (편향 포함) 결과를 계산할 때 편향 요인 (예 : 작업자, 기기 등) 을 고려합니다.

급내 상관 (편향 및 교호작용 포함) 결과를 계산할 때 편향 및 교호작용 요인을 고려합니다. 이 계산은 모형이 교차 모형이고 범위 대신 표준편차를 사용하는 경우에만 나타납니다.

편향 영향 편향 요인에 의해 급내 상관이 감소하는 양입니다.

편향 및 교호작용 영향 편향 및 교호작용 요인에 의해 급내 상관이 감소하는 양입니다. 이 계산은 모형이 교차 모형이고 범위 대신 표준편차를 사용하는 경우에만 나타납니다.

공정 클래스 모니터

시스템 및 분류 파라미터를 이해하려면 먼저 모니터링 분류 범례를 이해해야 합니다.

그림 4.6 모니터링 분류 범례

분류 모니터링 범례

분류	클래스 간 상관	프로세스 신호 감쇠	경고의 확률, 검정 1만*	경고의 확률, 검정 1 ~ 4*
첫 번째 클래스	0.80 - 1.00	11%보다 작음	0.99 - 1.00	1.00
두 번째 클래스	0.50 - 0.80	11% - 29%	0.88 - 0.99	1.00
세 번째 클래스	0.20 - 0.50	29% - 55%	0.40 - 0.88	0.92 - 1.00
네 번째 클래스	0.00 - 0.20	55% 초과	0.03 - 0.40	0.08 - 0.92

* Wheeler 검정에 의한 10개 부분군 내 3* 표준 오차 밖으로 이동에 대한 경고 확률(Nelson 검정 1, 2, 5, 6에 해당)

이 범례는 첫 번째, 두 번째, 세 번째 및 네 번째 클래스 분류를 설명합니다. 각 분류는 다음을 나타냅니다.

- 해당하는 급내 상관 값
- 공정 신호 감쇄 (감소) 양
- Wheeler 검정 1 개 또는 4 개를 모두 사용하여 10 개 부분군 내에서 3* 표준 오차 변화를 감지할 확률

Wheeler(2006)는 Western Electric Zone Test라고 하는 네 가지 감지 검정을 식별합니다. "변화 감지 프로파일러"에는 선택할 수 있는 8개의 검정이 있습니다. Wheeler 검정에 해당하는 검정은 첫 번째, 두 번째, 다섯 번째 및 여섯 번째 검정입니다.

팁 : 범례가 나타나지 않게 하려면 "EMP 측정 시스템 분석" 플랫폼 환경 설정에서 **모니터링 분류 범례 표시**를 선택 취소하십시오.

유효 해상도

유효 해상도 보고서는 측정 증분이 제대로 작동하는지 확인하는 데 도움이 됩니다 . 측정값을 기록할 때 자릿수를 추가하거나 삭제해야 할 수도 있고 현재 증분이 그대로 적용될 수도 있습니다 . 다음 사항에 유의하십시오 .

- " 확률 오차 " 는 측정값 오차의 중앙값을 계산합니다 .
- " 현재 측정 증분 " 은 현재 반올림하는 자릿수를 반영하며 데이터에서 최근접 10 의 거듭제곱으로 가져옵니다 . 이 값을 " 가장 작은 유효 증분 ", " 하한 증분 " 및 " 가장 큰 유효 증분 " 과 비교합니다 . 이 비교 결과를 바탕으로 작업이 추천됩니다 .
- 측정 증분이 크면 마지막 자릿수에서 불확도는 적지만 중앙값 오차가 큽니다 . 측정 증분이 작으면 중앙값 오차는 작지만 마지막 자릿수에서 불확도가 더 큽니다 .

변화 감지 프로파일러

변화 감지 프로파일러를 사용하면 공정 모니터링에 사용하는 관리도의 민감도를 평가할 수 있습니다. 변화 감지 프로파일러는 제품 평균 또는 제품 표준편차의 변화를 감지할 확률을 추정합니다. 관리도 한계에는 측정 오차 변동의 원인이 포함됩니다. 변화 감지 프로파일러는 이러한 한계를 기반으로 경고 확률을 추정합니다. 이것은 공정 평균을 모니터링하는 관리도에서 다음 k 개 부분군에 대해 경고 신호를 보낼 확률입니다.

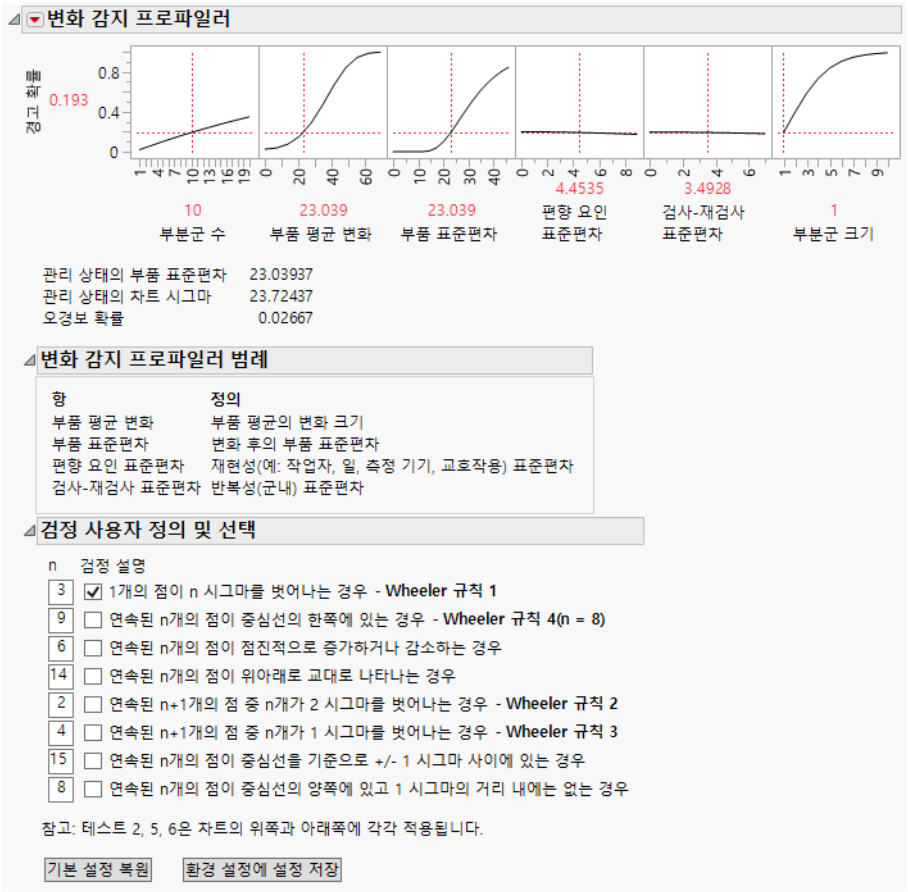
관리도에 사용할 부분군 크기를 설정할 수 있습니다 . 다음 사항에 유의하십시오 .

- 부분군 크기가 1 인 경우 관리도는 개별 측정값 차트입니다 .
- 부분군 크기가 1 을 초과할 경우 관리도는 *X*Bar 차트입니다 .

부분군 크기가 관리도의 민감도에 미치는 영향을 탐색할 수 있습니다. 또한 편향 및 검사-재검사 오차를 줄여 얻을 수 있는 이점도 확인할 수 있습니다.

그림 4.7에서는 Variability Data 폴더에 있는 Gasket.jmp 샘플 데이터 테이블에 대한 "변화 감지 프로파일러" 보고서를 보여 줍니다.

그림 4.7 Gasket.jmp 에 대한 변화 감지 프로파일러



경고 확률

경고 확률은 공정의 변화를 감지할 확률입니다. "변화 감지 프로파일러"의 "부품 평균 변화" 및 "부품 표준편차" 설정에 의해 변화가 정의됩니다. 확률을 계산할 때는 "검정 사용자 정의 및 선택" 개요에서 선택한 검정이 프로파일러에 지정된 "부분군 수"에 적용된다고 가정합니다.

개별 측정값 차트(부분군 크기 = 1) 및 XBar 차트(부분군 크기 > 1)의 관리 한계는 관리 상태의 차트 시그마를 기반으로 합니다. 관리 상태의 시그마는 편향 요인(재현성) 변동과 검사-재검사(반복성) 변동을 고려합니다. 처음에는 MSA 연구에서 얻은 값으로 설정됩니다. 관리 상태의 차트 시그마는 관리 상태의 부품 표준편차도 통합합니다. 이 두 값은 관리 상태의 차트 시그마를 기반으로 한 오경보 확률과 함께 프로파일러 아래에 나타납니다.

관리 상태의 부품 표준편차 측정 오차를 제외한 안정적인 공정에 대한 실제 부품 값의 표준편차입니다. 관리 상태의 부품 표준편차 기본값은 부품 성분의 표준편차입니다. 부품 성분의 표준편차는 MSA 분석으로 추정되고 분산 성분 보고서에 있습니다.

MSA 연구의 부품은 특정한 특성을 갖도록 선택되며 생산 시 볼 수 있는 부품 간 변동을 반영하지 못하는 경우가 많습니다. 이 때문에 " 변화 감지 프로파일러 " 의 빨간색 삼각형 메뉴에서 **관리 상태의 부품 표준편차 변경**을 선택하여 관리 상태의 부품 표준편차를 지정할 수 있습니다.

관리 상태의 차트 시그마 관리 한계를 계산하는 데 사용되는 시그마 값입니다. 이 값은 변화 감지 프로파일러에 지정된 관리 상태의 부품 표준편차, 편향 요인 표준편차 및 검사 - 재검사 표준편차와 부분군 크기를 사용하여 계산됩니다. 이때 재현성 요인은 부분군 내에서 일정하다고 가정합니다.

부분군 크기가 n 인 경우 관리 한계는 다음 값으로 설정됩니다.

$\pm 3(\text{관리 상태의 차트 시그마})/(\sqrt{n})$

관리 상태의 차트 시그마는 다음 항의 제곱합 제곱근입니다.

- 관리 상태의 부품 표준편차
- 변화 감지 프로파일러에 지정된 편향 요인 표준편차에 $\sqrt{n}$ 을 곱한 값
- 변화 감지 프로파일러에 지정된 검사 - 재검사 표준편차

재현성 요인이 부분군 내에서 일정하다는 가정을 설명하기 위해 편향 요인 표준편차에 $\sqrt{n}$ 을 곱합니다.

관리 상태의 부품 표준편차, 편향 요인 표준편차, 검사 - 재검사 표준편차 또는 부분군 크기를 변경하면 관리 상태의 차트 시그마가 업데이트됩니다.

오경보 확률 부품 평균 또는 표준편차가 변경되지 않았는데 관리도 검정에서 경고 신호가 발생할 확률입니다. " 검정 사용자 정의 및 선택 " 에서 부분군 수 또는 검정을 변경하면 오경보 확률이 업데이트됩니다.

분산 성분 보고서에 대한 자세한 내용은 "분산 성분"에서 확인하십시오.

변화 감지 프로파일러 설정

부분군 수 경고 확률이 계산되는 부분군의 수입니다. 부분군 수가 k 로 설정되면 프로파일러는 이 k 개 부분군을 기반으로 관리도에서 하나 이상의 경고 신호를 보낼 확률을 제공합니다. 부분군 수는 기본적으로 10 으로 설정됩니다. 부분군 수를 변경하려면 그림에서 세로선을 드래그합니다.

부품 평균 변화 부품 평균의 변화 크기입니다. 기본적으로 프로파일러는 1 시그마 변화를 감지하도록 설정됩니다. 초기값은 부품 성분의 표준편차입니다. 부품 성분의 표준편차는 MSA 분석으로 추정되고 분산 성분 보고서에 있습니다. 부품 평균 변화를 변경하려면 그림에서 세로선을 드래그하거나 그림 아래의 값을 클릭합니다.

부품 표준편차 측정 오차를 제외한 실제 부품 값의 표준편차입니다. 부품 표준편차 초기값은 부품 성분의 표준편차입니다. 부품 성분의 표준편차는 MSA 분석으로 추정되고 분산 성분 보

고서에 있습니다 . 부품 표준편차를 변경하려면 그림에서 세로선을 드래그하거나 그림 아래의 값을 클릭합니다 .

편향 요인 표준편차 재현성과 관련된 요인의 표준편차입니다 . 편향 요인에는 작업자와 기기가 포함됩니다 . 편향 요인 변동에는 부품 및 반복성 (군내) 변동이 포함되지 않습니다 . 초기값은 재현성 및 교호작용 분산 성분을 사용하여 파생됩니다 . 이 분산 성분은 MSA 분석으로 추정되고 분산 성분 보고서에 있습니다 . 편향 요인 표준편차를 변경하려면 그림에서 세로선을 드래그하거나 그림 아래의 값을 클릭합니다 .

검사 - 재검사 표준편차 모형의 검사 - 재검사 (반복성) 변동의 표준편차입니다 . 초기값은 군내 성분의 표준편차입니다. 군내 성분의 표준편차는 MSA 분석으로 추정되고 분산 성분 보고서에 있습니다 . 검사 - 재검사 표준편차를 변경하려면 그림에서 세로선을 드래그하거나 그림 아래의 값을 클릭합니다 .

부분군 크기 각 부분군에 사용되는 표본 크기입니다 . 이 값은 기본적으로 1 로 설정됩니다 . 표본 크기를 늘려 관리도 성능 개선을 조사할 수 있습니다 . 표본 크기를 1 에서 늘리면 개별 측정값 차트에서 *X*Bar 차트로 이동할 때 어떻게 되는지 결과를 보여 줍니다 . 부분군 크기를 변경하려면 그림에서 세로선을 드래그합니다 .

변화 감지 프로파일러 옵션

" 변화 감지 프로파일러 " 의 빨간색 삼각형 메뉴는 몇 가지 옵션을 제공합니다 . 여기에서는 하나의 옵션만 설명합니다 .

관리 상태의 부품 표준편차 변경 안정적 공정에 대한 부품 표준편차 값을 지정합니다 . 관리 상태의 부품 표준편차는 측정 오차를 제외한 실제 부품 값의 변동을 반영해야 합니다 . 새 값을 입력하고 " 확인 " 을 클릭합니다 .

관리 상태의 부품 표준편차는 처음에 부품 성분의 표준편차로 설정됩니다. 부품 성분의 표준편차는 MSA 분석으로 추정되고 분산 성분 보고서에 있습니다 .

이 옵션은 EMP 연구를 위해 선택한 부품이 공정에서 추출한 랜덤 표본이 아닌 경우에 유용합니다 .

요인 격자 재설정 요인의 현재 설정에 대한 특정 값을 입력하고 , 해당 설정을 잠그거나 , 격자의 여러 요소를 제어할 수 있도록 각 요인에 대한 창을 표시합니다 . 자세한 내용은 Profilers 의에서 확인하십시오 .

요인 설정 다음 옵션으로 구성된 하위 메뉴입니다 .

설정 기억 " 설정 기억 " 명령이 호출될 때마다 현재 설정 값을 누적하는 개요 노드를 보고서에 추가합니다 . 기억된 각 설정 앞에는 해당 설정으로 재설정하는 데 사용되는 라디오 버튼이 있습니다 . " 기억된 설정 " 의 빨간색 삼각형 메뉴에는 선택한 설정 또는 모든 설정을 제거하는 옵션이 있습니다 .

설정 스크립트 복사 현재 프로파일러의 설정을 클립보드에 복사합니다 .

설정 스크립트 붙여넣기 클립보드의 프로파일러 설정을 다른 보고서의 프로파일러에 붙여넣습니다 .

스크립트 설정 요인이 변경될 때마다 호출되는 스크립트를 설정합니다. 설정된 스크립트는 다음 형식의 인수 목록을 받습니다.

```
{factor1 = n1, factor2 = n2, ...}
```

예를 들어 이 목록을 로그에 기록하려면 먼저 다음 함수를 정의합니다.

```
ProfileCallbackLog = Function({arg},show(arg));
```

그런 다음 스크립트 설정 대화상자에 ProfileCallbackLog 를 입력합니다.

이와 비슷한 다음 함수는 요인 값을 전역 값으로 변환합니다.

```
ProfileCallbackAssign = Function({arg},evalList(arg));
```

또는 다음과 같이 한 번에 하나씩 값에 액세스합니다.

```
 ProfileCallbackAccess =
Function({arg},f1=arg["factor1"];f2=arg["factor2"]);
```

변화 감지 프로파일러 범례

이 패널에서는 변화 감지 프로파일러의 네 가지 설정에 대해 간략하게 설명합니다. 자세한 내용은 "변화 감지 프로파일러 설정"에서 확인하십시오.

팁 : 범례가 나타나지 않게 하려면 "EMP 측정 시스템 분석 " 플랫폼 환경 설정에서 **변화 감지 프로파일러 범례 표시**를 선택 취소하십시오.

검정 사용자 정의 및 선택

"검정 사용자 정의 및 선택" 패널에서는 관리도의 k 개 부분군에 적용할 검정을 선택하고 사용자 정의합니다. 여기에 나오는 8 개 검정은 Nelson 연구 자료 (1984) 를 기반으로 합니다. 검정에 대한 자세한 내용은 "검정"에서 확인하십시오.

변화 감지 프로파일러 계산을 수행할 때 이러한 검정이 고려됩니다. 검정을 더 추가하면 경고 확률 및 오경보 확률 값이 증가합니다. 준난수 시뮬레이션을 기반으로 계산되므로 프로파일러를 업데이트할 때 약간 지연될 수 있습니다.

" 검정 사용자 정의 및 선택 " 패널에는 다음 옵션이 있습니다.

기본 설정 복원 설정이 환경 설정에 저장되지 않은 경우 이 옵션은 선택한 검정을 첫 번째 검정으로만 재설정합니다. n 값도 " 검정 " 에 설명된 값으로 재설정됩니다. 설정이 환경 설정에 저장된 경우 이 옵션은 선택한 검정 및 n 값을 환경 설정에 지정된 값으로 재설정합니다.

참고 : 파일 > 환경 설정 > 플랫폼 > 관리도 빌더를 선택하여 관리도 검정의 환경 설정에 액세스할 수 있습니다. 사용자 검정 1 ~ 8 은 " 검정 사용자 정의 및 선택 " 에 표시된 8 개 검정에 해당합니다.

환경 설정에 설정 저장 선택한 검정 및 *n* 값을 향후 분석에 사용할 수 있도록 저장합니다 . 이러한 환경 설정은 " 관리도 빌더 " 플랫폼 환경 설정에 추가됩니다 .

편향 비교

편향 비교 옵션은 평균 분석 차트를 생성합니다 . 이 차트에서는 그룹화 변수의 각 수준에 대한 평균 값을 보여 주고 전체 평균과 비교합니다 . 이 차트를 사용하면 작업자가 대체로 부품을 너무 높게 측정하는지 아니면 너무 낮게 측정하는지 확인할 수 있습니다 .

" 평균 분석 " 옆의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

유의 수준 설정 가장 일반적인 유의 수준 옵션 중에서 하나를 선택하거나 , **기타**를 선택하고 원하는 수준을 직접 지정합니다 . 유의 수준을 변경하면 결정 상한 및 결정 하한이 수정됩니다 .

요약 보고서 표시 그룹 평균과 결정 한계가 포함된 보고서를 표시하고 , 그룹 평균이 결정 상한을 초과하거나 결정 하한 미만인 경우 보고합니다 .

표시 옵션 다음 옵션을 포함합니다 .

결정 한계 표시 UDL(결정 상한) 및 LDL(결정 하한) 을 나타내는 선을 그리고 해당 값을 정의합니다 .

결정 한계 음영 표시 UDL 과 LDL 사이에 음영을 추가합니다 .

중심선 표시 평균을 나타내는 중심선 통계량을 그립니다 .

점 옵션 차트 표시를 바늘 , 연결된 점 또는 점으로 변경합니다 .

검사 - 재검사 오차 비교

검사 - 재검사 오차 비교 옵션은 " 분산에 대한 평균 분석 " 또는 " 평균 범위 분석 " 차트 유형을 생성합니다 . 이 차트에는 작업자 간에 검사 - 재검사 오차의 차이가 있는지 여부가 표시됩니다 . 예를 들어 이 차트를 사용하면 각 작업자의 측정 방식에 불일치가 있는지 확인할 수 있습니다 . 분산 성분에 범위가 사용되는 경우 평균 범위 분석 차트가 표시됩니다 .

- "Operator 분산 검정 " 옆의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 " 편향 비교 " 에서 확인하십시오 .
- " 분산에 대한 평균 분석 " 차트에 대한 자세한 내용은 " 분산 성분 " 에서 확인하십시오 .

측정 시스템 분석의 추가 예

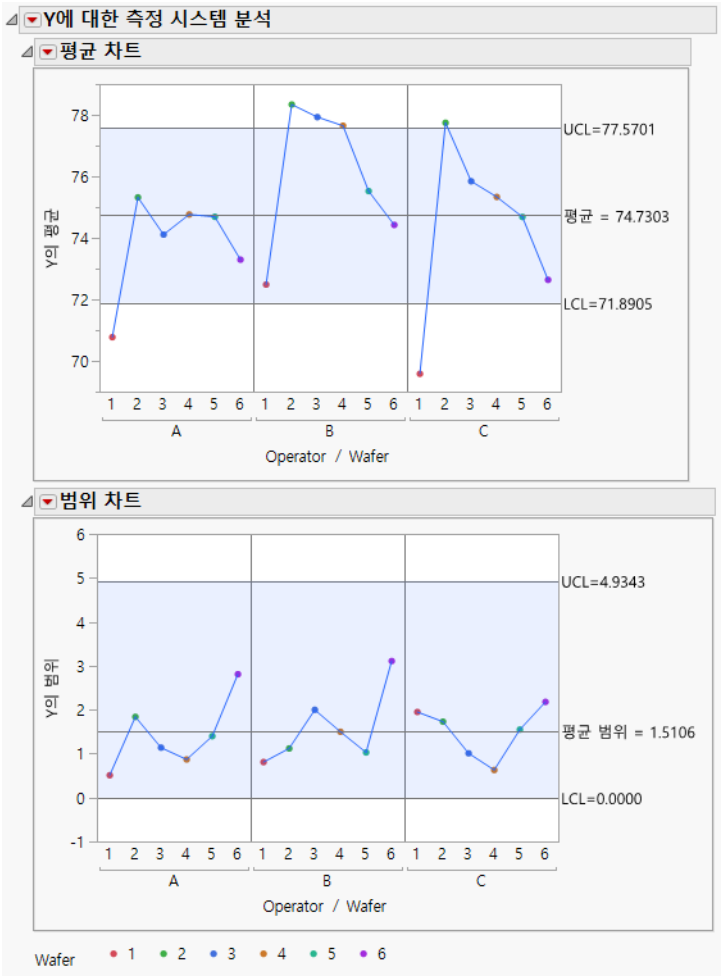
이 예에서는 세 명의 작업자가 6 개의 웨이퍼에 대해 각각 두 번씩 단일 특성을 측정했습니다 . 측정 시스템 성능을 확인하기 위해 상세한 분석을 수행합니다 .

초기 분석 수행

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/Wafer.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 측정 시스템 분석**을 선택합니다 .
3. Y 를 **Y, 반응** 역할에 할당합니다 .
4. Wafer 를 **부품 , 표본 ID** 역할에 할당합니다 .
5. Operator 를 **X, 그룹화** 역할에 할당합니다 .

   MSA 방법은 **EMP**, **산포 차트 유형**은 **범위** , **모형 유형**은 **교차**로 각각 설정되어 있습니다 .
6. **확인**을 클릭합니다 .

그림 4.8 평균 차트와 범위 차트



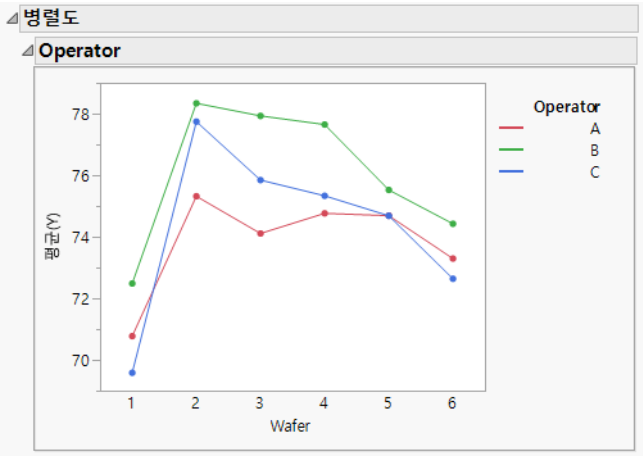
"평균 차트"에서는 평균 부품 측정값 중 일부가 관리 한계를 벗어난다는 것을 보여 줍니다. 이 결과는 측정 가능한 부품 간 변동을 나타내므로 바람직합니다.

"범위 차트"에서는 관리 한계를 벗어나는 점이 없습니다. 이 결과는 작업자 측정값이 부품 내에서 일관성이 있음을 나타내므로 바람직합니다.

교호작용 검토

작업자와 부품 간의 교호작용을 자세히 살펴봅니다. "Y에 대한 측정 시스템 분석" 옆의 빨간색 삼각형을 클릭하고 **병렬도**를 선택합니다.

그림 4.9 병렬도

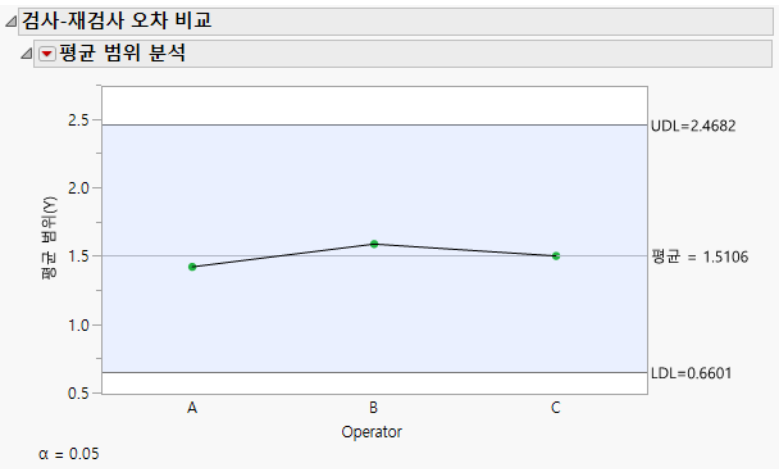


작업자별 병렬도를 살펴보면 선이 비교적 평행하고 약간의 교차만 있음을 확인할 수 있습니다.

작업자 일관성 검토

작업자 간의 분산을 자세히 살펴봅니다. "Y에 대한 측정 시스템 분석" 옆의 빨간색 삼각형을 클릭하고 **검사-재검사 오차 비교**를 선택합니다.

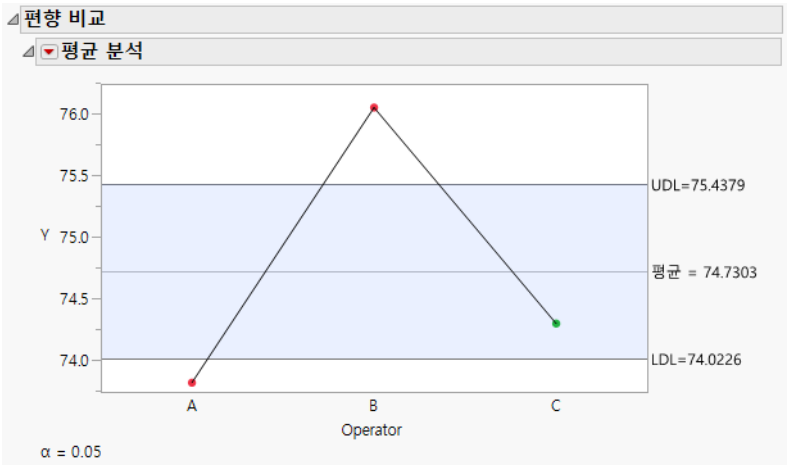
그림 4.10 검사 - 재검사 오차 비교



검사-재검사 오차 비교를 살펴보면 검사-재검사 오차가 전체 검사-재검사 오차와 유의하게 다른 작업자가 없음을 확인할 수 있습니다. 작업자가 일관되게 측정하는 것으로 보입니다.

확실히 하기 위해 작업자가 부품을 너무 높거나 낮게 측정하는지를 나타내는 편향 비교 차트를 살펴보기로 했습니다. "Y에 대한 측정 시스템 분석" 옆의 빨간색 삼각형을 클릭하고 **편향 비교**를 선택합니다.

그림 4.11 편향 비교



편향 비교 차트를 살펴보면 다음과 같은 내용을 관찰할 수 있습니다 .

- Operator A 와 Operator B 의 측정값은 전체 평균과 유의한 차이가 있으므로 측정 편향을 감지할 수 있습니다 .
- Operator A 는 유의하게 낮게 편향되어 있습니다 .
- Operator B 는 유의하게 높게 편향되어 있습니다 .
- Operator C 는 전체 평균과 유의한 차이를 보이지 않습니다 .

측정 시스템 분류

EMP 결과 보고서를 검토하여 측정 시스템을 분류하고 개선 가능성이 있는지 살펴봅니다. "Y에 대한 측정 시스템 분석" 옆의 빨간색 삼각형을 클릭하고 **EMP 결과**를 선택합니다.

그림 4.12 EMP 결과

EMP 결과

EMP 검정	결과	설명
검사-재검사 오차	1.3387	오차 이내
자유도	16.006	오차 범위 이내를 추정하는 데 사용되는 정보의 양
확률 오차	0.9029	단일 측정값 오차의 중앙값
클래스 간 상관(편향 없음)	0.7385	부품 변동에 기인한 변동의 비율(편향 요인 포함 안 함)
클래스 간 상관(편향 포함)	0.6272	부품 변동에 기인한 변동의 비율(편향 요인 포함)
편향 영향	0.1113	편향 요인에 의해 클래스 간 상관이 감소하는 양

시스템	분류
현재(편향 포함)	두 번째 클래스
잠재(편향 없음)	두 번째 클래스

분류 모니터링 범례

분류	클래스 간 상관	프로세스 신호 감쇠	경고의 확률, 검정 1만*	경고의 확률, 검정 1 ~ 4*
첫 번째 클래스	0.80 - 1.00	11%보다 작음	0.99 - 1.00	1.00
두 번째 클래스	0.50 - 0.80	11% - 29%	0.88 - 0.99	1.00
세 번째 클래스	0.20 - 0.50	29% - 55%	0.40 - 0.88	0.92 - 1.00
네 번째 클래스	0.00 - 0.20	55% 초과	0.03 - 0.40	0.08 - 0.92

* Wheeler 검정에 의한 10개 부분군 내 3* 표준 오차 밖으로 이동에 대한 경고 확률(Nelson 검정 1, 2, 5, 6에 해당)

"분류"가 "두 번째 클래스"이며, 이는 검정 1만 사용하여 10개의 부분군 내에서 3* 표준 오차 변화를 감지할 가능성이 88%보다 높음을 의미합니다. 편향 요인이 급내 상관에 미치는 영향은 11%입니다. 즉, 편향 요인을 제거하면 급내 상관계수가 11% 향상됩니다.

공정 변경을 감지하는 관리도 기능 살펴보기

변화 감지 프로파일러를 사용하면 관리도에서 공정 변경을 감지할 수 있는 확률을 탐색할 수 있습니다. "Y에 대한 측정 시스템 분석" 옆의 빨간색 삼각형을 클릭하고 **변화 감지 프로파일러**를 선택합니다.

그림 4.13 변화 감지 프로파일러

기본적으로 3 시그마 한계를 초과하는 점에 대한 검정만 선택됩니다. 또한 기본 부분군 크기가 1 이며, 이는 개별 측정값 차트를 사용 중임을 나타냅니다.

변화 후 10개 부분군에서 2 * 부품 표준편차 평균의 변화를 감지할 수 있는지 살펴봅니다. **부품 평균 변화** 값 2.1701 을 클릭하고 4.34(2.17 * 2)로 변경합니다. 부품 표준편차의 2 배 변화를 감지할 확률은 56.9% 입니다.

이제 편향을 제거할 경우 2 * 부품 표준편차의 변화를 감지할 확률에 어떤 영향을 미치는지 확인합니다. **편향 요인 표준편차** 값을 1.1256 에서 0 으로 변경합니다. 변화를 감지할 확률이 67.8% 로 증가합니다.

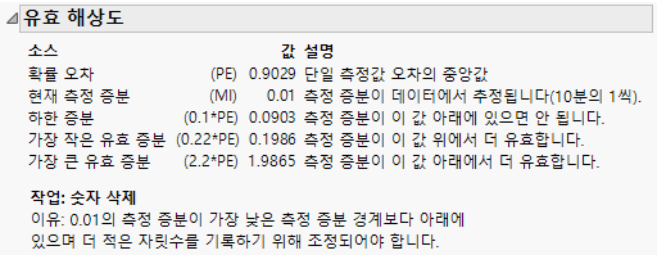
마지막으로, 검정을 더 추가할 경우 2 * 부품 표준편차 변화를 감지할 확률이 어떻게 달라지는지 확인합니다. 첫 번째 검정 외에 두 번째, 다섯 번째 및 여섯 번째 검정(Wheeler 규칙 4, 2, 3)을 선택합니다. 이 네 가지 검정(편향 변동 없음)을 사용할 경우 변화를 감지할 확률은 99.9% 입니다.

더 큰 부분군 크기를 기반으로 하는 관리도를 사용할 경우의 결과도 확인할 수 있습니다. 부분군 크기가 2 이상인 경우 관리도는 XBar 차트입니다. **편향 요인 표준편차** 값을 다시 1.1256 으로 변경하고 첫 번째 검정을 제외한 모든 검정을 선택 취소합니다. 프로파일러에서 **부분군 크기**를 4 로 설정합니다. 2 * 부품 표준편차 변화를 감지할 확률은 98.5% 입니다.

측정 증분 검토

마지막으로, 측정 증분이 제대로 작동하는지 확인합니다. "Y에 대한 측정 시스템 분석" 옆의 빨간색 삼각형을 클릭하고 **유효 해상도**를 선택합니다.

그림 4.14 유효 해상도



현재 측정 증분(0.01)이 하한 증분(0.09) 미만이며, 이는 하나 더 적은 자릿수를 기록하기 위해 이후 측정을 조정해야 함을 나타냅니다.

측정 시스템 분석에 대한 통계 상세 정보

범위 차트에 사용되는 한계 계산에 대한 자세한 내용은 "Xbar-R 차트에 대한 관리 한계"에서 확인하십시오. S 관리도에 사용되는 한계 계산에 대한 자세한 내용은 "Xbar-S 차트에 대한 관리 한계"에서 확인하십시오.

급내 상관 및 확률 오차 계산

급내 상관 (편향 없음) 은 다음과 같이 계산됩니다 .

$$r_{pe} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{pe}^2}$$

급내 상관 (편향 포함) 은 다음과 같이 계산됩니다 .

$$r_b = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_b^2 + \hat{\sigma}_{pe}^2}$$

급내 상관 (편향 및 교호작용 요인 포함) 은 다음과 같이 계산됩니다 .

$$r_{int} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_b^2 + \hat{\sigma}_{int}^2 + \hat{\sigma}_{pe}^2}$$

확률 오차는 다음과 같이 계산됩니다 .

$$Z_{0.75} \times \hat{\sigma}_{pe}$$

다음 사항에 유의하십시오 .

$\hat{\sigma}^2_{pe}$ = 순수 오차에 대한 분산 추정값

$\hat{\sigma}^2_{p}$ = 제품에 대한 분산 추정값

$\hat{\sigma}^2_{b}$ = 편향 요인에 대한 분산 추정값

$\hat{\sigma}^2_{int}$ = 교호작용 요인에 대한 분산 추정값

$Z_{0.75}$ = 표준 정규 분포의 75% 분위수

5 장

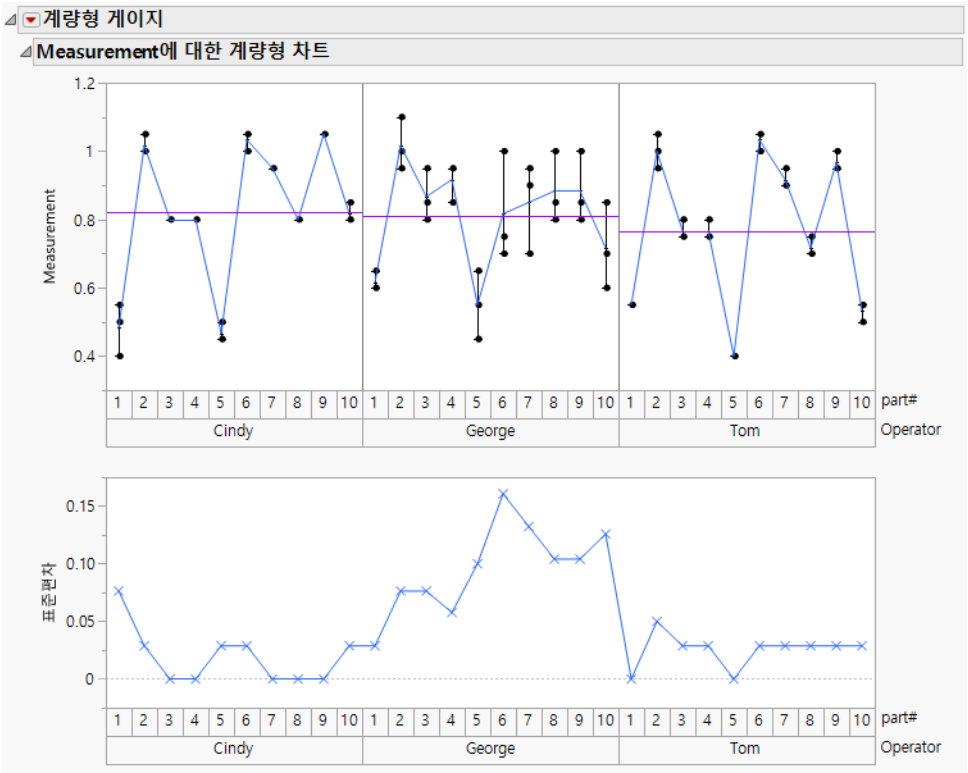
계량형 게이지 차트

게이지 R&R 을 사용하여 연속형 측정 공정 평가

계량형 게이지 차트는 연속형 측정값을 분석하고 측정 시스템의 성능을 나타낼 수 있습니다. 게이지 연구를 수행하여 데이터의 변동 측도를 확인할 수도 있습니다.

팁 : 이 장에서는 계량형 차트만 설명합니다 . 계수형 차트에 대한 자세한 내용은 " 계수형 게이지 차트 " 에서 확인하십시오 .

그림 5.1 계량형 차트의 예



목차

계량형 차트 개요 123
계량형 차트의 예 124
계량형 / 계수형 게이지 차트 플랫폼 시작 125
데이터 형식 126
계량형 게이지 차트 126
계량형 게이지 플랫폼 옵션 128
이분산성 검정 130
분산 성분 130
게이지 R&R 방법 정보 132
게이지 R&R 옵션 133
판별비 135
오분류 확률 135
편향 보고서 136
선형성 연구 136
계량형 차트의 추가 예 137
이분산성 검정의 예 137
편향 보고서 옵션의 예 139
계량형 차트에 대한 통계 상세 정보 142
분산 성분에 대한 통계 상세 정보 142
판별비에 대한 통계 상세 정보 143
오분류 확률에 대한 통계 상세 정보 144

계량형 차트 개요

팁 : 계량형 차트의 기존 이름은 **다변수** 차트이지만 이 이름이 잘 알려져 있지 않기 때문에 좀 더 일반적인 계량형 차트라는 용어가 대신 사용됩니다 .

시간에 따른 공정의 변동을 표시하는 관리도처럼 계량형 차트에는 부품, 작업자, 반복, 측정기 등의 범주에 대한 동일한 유형의 변동이 표시됩니다. 계량형 차트에는 그룹화 요인의 각 수준에 대한 데이터와 평균이 표시되며 모든 그림은 나란히 배치됩니다. 데이터와 함께 각 범주에 대한 데이터의 평균, 범위 및 표준편차를 보고 범주 간에 이 값이 어떻게 달라지는지 확인할 수 있습니다. 보고서 옵션은 범주 간에 평균 및 분산이 어떻게 달라지는지에 가장 관심이 있다는 가정을 기반으로 합니다.

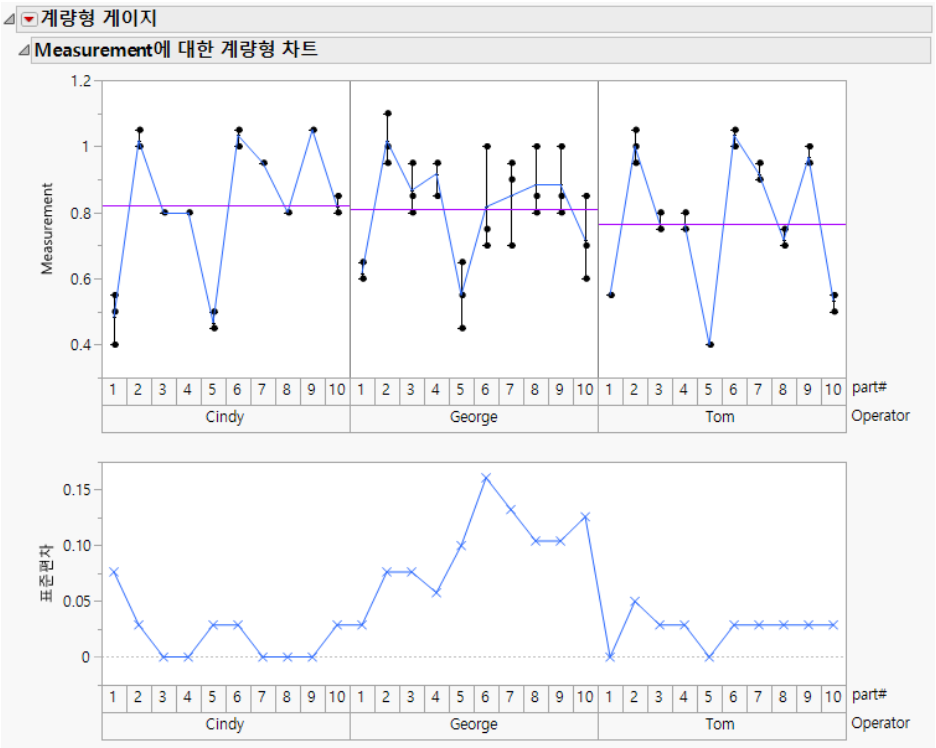
계량형 차트는 게이지 R&R 같은 측정 시스템 분석에 주로 사용됩니다. 이 분석은 작업자 변동(재현성) 및 측정 변동(반복성)으로 인한 변동이 어느 정도인지 검토합니다. 게이지 R&R은 모형의 균형 여부에 관계없이 교차 모형과 내포 모형의 여러 조합에 사용할 수 있습니다.

계량형 차트의 예

부품 측정값이 포함된 데이터가 있다고 가정해 보겠습니다. Cindy, George, Tom 이라는 세 명의 작업자가 각각 10 개의 부품을 측정했습니다. 각 부품을 세 번씩 측정하여 총 90 개의 관측값이 수집되었습니다. 작업자 간의 변동을 식별하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/2 Factors Crossed.jmp를 엽니다.
2. **분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트**를 선택합니다.
3. **차트 유형**에서 **계량형**을 선택합니다.
4. Measurement 를 선택하고 **Y, 반응**을 클릭합니다.
5. Operator 를 선택하고 **X, 그룹화**를 클릭합니다.
6. part# 을 선택하고 **부품 , 표본 ID** 를 클릭합니다.
7. **확인**을 클릭합니다.
8. "계량형 게이지"의 빨간색 삼각형을 클릭하고 **그룹 평균 표시** 및 **셀 평균 연결**을 선택합니다.

그림 5.2 계량형 차트의 예

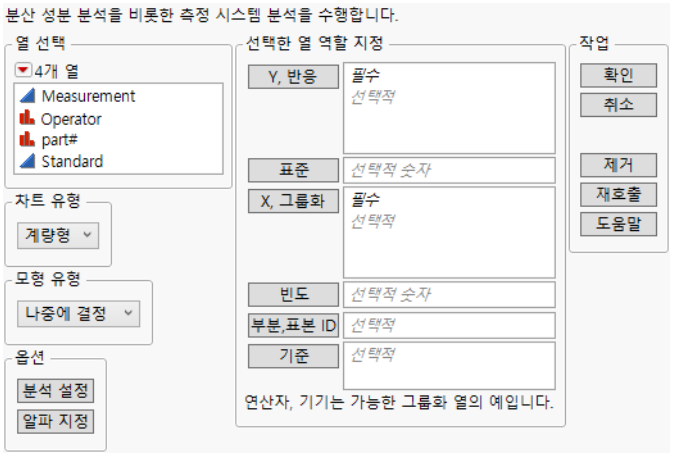


표준편차 차트를 살펴보면 부품을 가장 일관되게 측정한 것으로 보이는 Tom 보다 Cindy 와 George 의 측정값에 변동이 더 많은 것을 확인할 수 있습니다. George 의 측정값 변동이 가장 크기 때문에 부품을 가장 일관되지 않게 측정한다고 볼 수 있습니다.

계량형 / 계수형 게이지 차트 플랫폼 시작

분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트를 선택하여 계량형 / 계수형 게이지 차트 플랫폼을 시작할 수 있습니다. **차트 유형**을 **계량형**으로 설정합니다.

그림 5.3 계량형 / 계수형 게이지 차트 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

차트 유형 계량형 게이지 분석 (연속형 반응의 경우) 또는 계수형 게이지 분석 (대개 " 통과 " 또는 " 실패 " 로 구분되는 범주형 반응의 경우) 중에서 선택합니다.

> **참고 :** 이 장에서는 **계량형** 차트 유형에 대한 내용만 다룹니다. **계수형** 차트 유형에 대한 자세한 내용은 " 계수형 게이지 차트 " 에서 확인하십시오.

모형 유형 모형 유형 (**주 효과** , **교차** , **내포** 등) 을 선택합니다. 자세한 내용은 " 분산 성분에 대한 통계 상세 정보 " 에서 확인하십시오.

분석 설정 분산 성분을 계산하기 위한 방법을 지정합니다. 자세한 내용은 " 분석 설정 " 에서 확인하십시오.

알파 지정 플랫폼에서 사용되는 유의 수준을 지정합니다.

Y, 반응 측정값 열을 지정합니다. *Y* 열을 두 개 이상 지정하면 각 반응에 대해 별도의 계량형 차트가 생성됩니다.

표준 측정된 부품에 대해 " 실제 " 또는 알려진 값을 포함하는 표준 또는 참조 열을 지정합니다 . 이 열을 포함하면 **편향** 및 **선형성 연구** 옵션이 활성화됩니다 . 이러한 옵션은 관측된 측정값과 참조 또는 표준 값 간의 차이를 분석합니다 . 자세한 내용은 " 편향 보고서 " 및 " 선형성 연구 " 에서 확인하십시오 .

X, 그룹화 측정값을 그룹화하는 분류 열을 지정합니다 . 요인이 내포된 계층 구조를 형성하는 경우 고차 항을 먼저 지정합니다 . 게이지 연구를 수행하는 경우에는 작업자를 먼저 지정한 후 부품을 지정합니다 .

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 데이터 테이블 열을 식별합니다 . 요약 데이터가 있는 경우 유용하게 사용됩니다 .

부품 , 표본 ID 측정되는 부품 또는 표본을 식별합니다 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다 .

데이터 형식

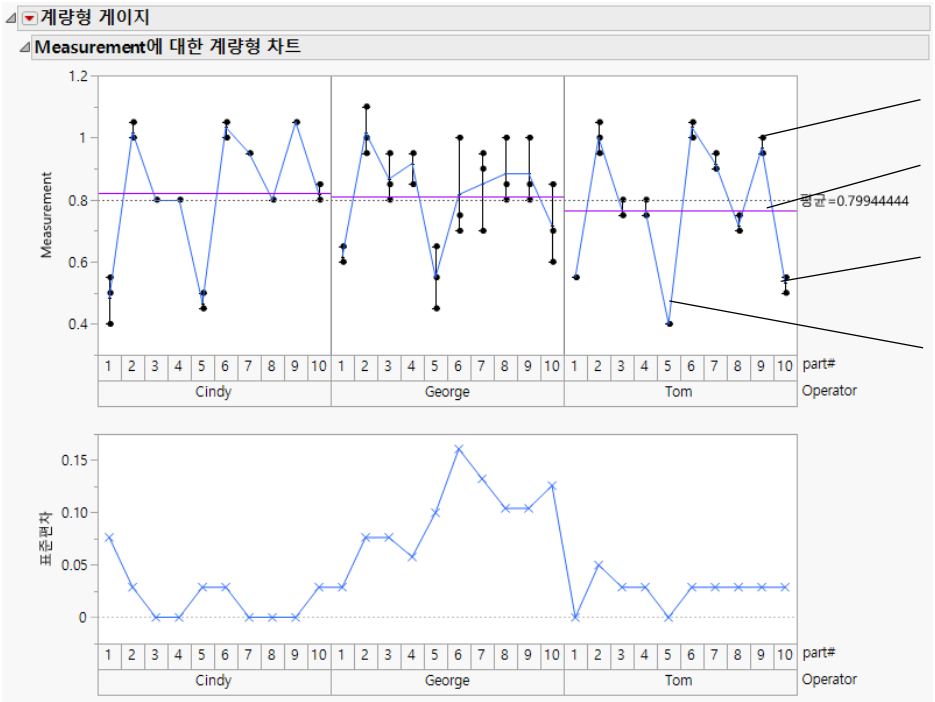
계량형 차트 플랫폼을 사용하려면 모든 반응 측정값이 단일 반응 열에 있어야 합니다. 반응이 여러 열에 기록되는 경우도 있습니다. 이때 각 행은 한 설계 요인의 수준이고 각 열은 다른 설계 요인의 수준입니다. 계량형 차트 플랫폼을 실행하기 전에 이 형식의 데이터를 쌓아야 합니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

계량형 게이지 차트

계량형 차트와 S 관리도에는 변동 패턴이 표시됩니다. 이 관리도를 사용하여 가능한 변동 그룹을 식별할 수 있습니다(부분군 내, 부분군 간, 시간 경과에 따름). 이러한 변동 원인이 크다고 판단되면 해당 원인에 대한 변동을 줄이는 작업을 수행할 수 있습니다.

그림5.4에 표시된 것과 같은 결과를 생성하려면 "계량형 차트의 예"에 설명된 방법을 따르십시오.

그림 5.4 계량형 게이지 차트



이 차트에는 y 축의 반응 및 분류된 다수준 x 축이 표시됩니다.

그림 5.4의 Measurement 차트에는 각 작업자의 부품별 측정값 범위가 표시됩니다. 각 측정값이 차트에 나타납니다. 최대값 및 최소값 막대는 각 셀의 값 범위를 나타내고 셀 평균 막대는 각 값 조합의 중앙값을 나타냅니다. 표준편차 차트에는 작업자가 각 부품에 대해 측정한 값의 표준편차가 표시됩니다.

그림 5.4에 나타난 대로 차트에 기능을 추가할 수 있습니다. 자세한 내용은 "계량형 게이지 플랫폼 옵션"에서 확인하십시오.

차트에서 변수를 바꾸려면 다음 중 하나를 수행하십시오 .

- 한 축 라벨에서 다른 축 라벨로 변수를 드래그하여 기존 변수를 바꿉니다 . 변수를 차트로 드래그하거나 축 라벨을 클릭하면 축 라벨이 강조 표시됩니다 . 이 위치에 변수를 놓을 수 있습니다 .
- 연결된 데이터 테이블의 " 열 " 패널에서 변수를 클릭한 후 축 라벨로 드래그합니다 .

다른 플랫폼에서는 연결된 데이터 테이블에서 제외된 행도 차트 또는 그림에 나타납니다. 그러나 계량형 차트에서는 제외된 행이 차트에 나타나지 않습니다.

계량형 게이지 플랫폼 옵션

빨간색 삼각형 옵션을 사용하여 차트 모양을 수정하고 게이지 R&R 분석을 수행하거나 분산 성분을 계산할 수 있습니다.

참고 : 그림 5.4 에서는 이러한 옵션 중 일부를 보여 줍니다 .

팁 : 이러한 옵션의 기본 동작을 설정하려면 **파일 > 환경 설정 > 플랫폼 > 계량형 차트**를 선택하십시오 .

수직 차트 레이아웃을 가로 또는 세로로 변경합니다 .

계량형 차트 계량형 차트를 표시하거나 숨깁니다 .

점 표시 개별 행에 대한 점을 표시하거나 숨깁니다 .

범위 막대 표시 각 셀의 최대값과 최소값을 나타내는 막대를 표시하거나 숨깁니다 .

셀 평균 표시 각 셀의 평균 표식을 표시하거나 숨깁니다 .

셀 평균 연결 셀 그룹 내에서 셀 평균을 연결하거나 연결을 끊습니다 .

구분 기호 표시 **X, 그룹화** 변수의 각 수준 사이에 구분선을 표시하거나 숨깁니다 .

그룹 평균 표시 (X, 그룹화 변수가 두 개 이상인 경우 또는 X, 그룹화 변수와 부품 , 표본 ID 변수가 각각 한 개인 경우에만 사용 가능) 가로 실선으로 나타나는 셀 그룹 평균을 표시하거나 숨깁니다 . 그룹화 변수 중 하나를 선택하라는 창이 나타납니다 .

총 평균 표시 전체 그래프에서 가로 방향의 회색 점선으로 나타나는 전체 평균을 표시하거나 숨깁니다 .

총 중앙값 표시 전체 그래프에서 가로 방향의 파란색 점선으로 나타나는 전체 중앙값을 표시하거나 숨깁니다 .

상자 그림 표시 상자 그림을 표시하거나 숨깁니다 .

평균 다이아몬드 평균 다이아몬드를 표시하거나 숨깁니다 . 각 셀의 그룹 내 표준편차가 신뢰 구간에 사용됩니다 .

Xbar 관리 한계 계량형 차트의 UCL 및 LCL 에서 선을 표시하거나 숨깁니다 . 이러한 한계를 계산하는 방법에 대한 자세한 내용은 "관리도 빌더에 대한 통계 상세 정보"에서 확인하십시오.

지터링된 점 일치하는 점이 서로 포개지지 않도록 그림에 표시된 점에 랜덤 잡음을 추가합니다 .

표준 평균 표시 (**표준** 변수를 지정한 경우에만 사용 가능) 표준 열의 평균을 표시하거나 숨깁니다.

변동 요약 보고서 평균 , 표준편차 , CV(변동 계수), 평균의 표준 오차 , 신뢰 하한 / 상한 , 최소값 , 최대값 및 관측값 수를 제공하는 보고서를 표시하거나 숨깁니다 .

S 관리도 전체 범주 셀의 셀 표준편차를 표시하는 별도의 그래프를 표시하거나 숨깁니다 .

표준편차의 평균 S 관리도에서 평균 표준편차에 선을 표시하거나 숨깁니다 .

S 관리 한계 S 관리도에서 LCL 및 UCL 을 보여 주는 선을 표시하거나 숨깁니다 . 이러한 한계를 계산하는 방법에 대한 자세한 내용은"관리도 빌더에 대한 통계 상세 정보"에서 확인하십시오.

표준편차의 그룹 평균 S 관리도에서 평균 선을 표시하거나 숨깁니다 .

이분산성 검정 그룹 간 분산을 비교하기 위한 검정을 수행합니다 . 자세한 내용은 " 이분산성 검정 " 에서 확인하십시오 .

분산 성분 특정 모형에 대한 분산 성분을 추정합니다 . 분산 성분은 주효과 , 교차 , 내포 , 교차 후 내포 (3 개 요인만) 및 내포 후 교차 (3 개 요인만) 모형에 대해 계산됩니다 . 자세한 내용은 " 분산 성분 " 에서 확인하십시오 .

게이지 연구 다음 옵션이 포함되어 있습니다 .

게이지 R&R 첫 번째 요인과 마지막 요인을 각각 그룹화 열 및 부품으로 해석하고 , 추정된 분산 성분을 사용하여 게이지 R&R 보고서를 생성합니다 . 부품 필드도 시작 창에 있습니다 . 자세한 내용은 " 게이지 R&R 옵션 " 에서 확인하십시오 .

판별비 특정 제품에 대해 주어진 측정의 상대적 유용성을 나타냅니다 . 측정의 총 분산을 측정 오차의 분산과 비교합니다 . 자세한 내용은 " 판별비 " 에서 확인하십시오 .

오분류 확률 적합한 부품을 기각하고 잘못된 부품을 채택할 확률을 표시합니다 . 자세한 내용은 " 오분류 확률 " 에서 확인하십시오 .

편향 보고서 관측된 값과 표준 값 사이의 평균 차이를 표시합니다 . 편향 평균 그래프 및 요약 테이블이 나타납니다 . 이 옵션은 시작 창에서 표준 변수를 지정하는 경우에만 사용할 수 있습니다 . 자세한 내용은 " 편향 보고서 " 에서 확인하십시오 .

선형성 연구 표준 값을 X 변수로 사용하고 편향을 Y 변수로 사용하여 회귀 분석을 수행합니다 . 이 분석은 편향과 부품 크기 간의 관계를 검토합니다 . 기울기가 0 이면 가장 이상적입니다 . 기울기가 0 이 아니면 크기가 다른 부품에서 게이지가 다르게 작동한다는 의미입니다 . 이 옵션은 시작 창에서 표준 변수를 지정하는 경우에만 사용할 수 있습니다 . 자세한 내용은 " 선형성 연구 " 에서 확인하십시오 .

게이지 R&R 그림 평균 그림 (모형의 각 주효과에 의한 평균 반응) 및 표준편차 그림을 표시하거나 숨깁니다 . 순수 내포 모형의 경우 내포 구조의 그래프가 나타납니다 . 순수 교차 모형의 경우 교호작용 그래프가 나타납니다 . 그 외의 경우에는 그래프가 각 효과에서 독립적으로 표시됩니다 . 표준편차 그림의 빨간색 선은 각 효과에 대한 $\sqrt{\text{평균 가중 분산}}$ 값을 연결합니다 .

AIAG 라벨 자동차 분석에 광범위하게 사용되는 AIAG 표준에 따라 품질 통계량에 라벨을 설정해야 함을 지정할 수 있습니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

이분산성 검정

참고 : 자세한 내용은 " 이분산성 검정의 예 " 에서 확인하십시오 .

이분산성 검정 옵션은 그룹 간 분산을 비교하기 위한 검정을 수행합니다. 이 검정은 ANOMV(분산에 대한 평균 분석) 에 기반한 방법입니다. 이 방법은 그룹 표준편차가 평균 그룹 분산의 제곱근과 다른지 여부를 나타냅니다.

이 방법은 비정규 데이터에 대한 로버스트를 위해 순열 시뮬레이션을 사용하여 결정 한계를 계산합니다. 이 방법에 대한 자세한 내용은 Wludyka and Sa 연구 자료 (2004) 에서 확인하십시오. 시뮬레이션을 사용하는 방법이므로 결정 한계가 매번 약간 다를 수 있습니다. 항상 동일한 결과를 얻으려면 Ctrl+Shift 를 누르고 옵션을 선택한 후 동일한 난수 시드값을 지정합니다.

검정 보고서의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

유의 수준 설정 검정에 사용할 유의 수준을 설정합니다 .

요약 보고서 표시 검정에 대한 요약 보고서를 표시하거나 숨깁니다 . 보고서의 값은 그림에 표시된 값과 동일합니다 .

표시 옵션 결정 한계 , 음영 , 중심선 및 바늘을 표시하거나 숨깁니다 .

분산 성분

분산 성분 옵션은 측정값 사이의 변동을 모델링합니다. 반응은 다양한 분류 수준과 연결된 임의 효과에 상수 평균을 더한 것으로 가정됩니다.

참고 : 시작 창에서 **모형 유형**을 선택하지 않은 경우 (**나중에 결정**을 선택한 경우) **분산 성분** 옵션을 선택하면 모형 유형을 선택하는 대화상자가 나타납니다 . 모형 유형에 대한 자세한 내용은 " 계량형 / 계수형 게이지 차트 플랫폼 시작 " 에서 확인하십시오 .

그림 5.5 분산 성분 보고서의 예

분산 분석

소스	DF	SS	평균 제곱	F 비	Prob > F
Operator	2	0.054889	0.02744	1.3150	0.2931
part#	9	2.633583	0.29262	14.0209	<.0001*
Operator*part#	18	0.375667	0.02087	5.0425	<.0001*
범위 내	60	0.248333	0.00414		
합계	89	3.312472	0.03722		

분산 성분

성분	분산 성분	% 총계	20 40 60 80	제곱근(분산 성분)
Operator	0.00021914	0.5461		0.01480
part#	0.03019444	75.2		0.17377
Operator*part#	0.00557716	13.9		0.07468
범위 내	0.00413889	10.3		0.06433
합계	0.04012963	100.0		0.20032

"분산 분석" 보고서는 EMS 방법을 사용하여 분산 성분을 추정하는 경우에만 나타납니다. 이 보고서에서는 모형에 포함된 각 효과의 유의성을 보여 줍니다.

"분산 성분" 보고서에는 추정값이 표시됩니다. 자세한 내용은 "분산 성분에 대한 통계 상세 정보"에서 확인하십시오.

분석 설정

시작 창에서 **분석 설정**을 클릭하여 분산 성분 계산 방법을 선택합니다.

그림 5.6 분석 설정 창

분석 설정

- ◉ 최선의 분석(EMS, REML 또는 Bayes) 선택
- ○ 최선의 분석(EMS 또는 REML) 선택
- ○ REML 분석 사용
- ○ Bayes 분석 사용

최대 반복 수	100	(REML 분석에만 영향을 줌)
수렴 한계	1e-8	(REML 분석에만 영향을 줌)
통합 가로 좌표 수	128	(Bayes 분석에만 영향을 줌)
함수 실행 최대 횟수	65536	(Bayes 분석에만 영향을 줌)

확인 취소 도움말

최선의 분석 (EMS, REML 또는 Bayes) 선택 다음 논리를 사용하여 EMS, REML 또는 Bayes 중 최선의 분석을 선택합니다.

- 균형 데이터를 사용하는 경우 음의 분산 성분이 없으면 EMS(기대 평균 제곱) 방법을 사용하여 분산 성분을 추정합니다 .
- 불균형 데이터를 사용하는 경우 분산 성분이 음수로 추정되지 않으면 REML(제한 최대 가능도) 방법이 사용되고 음수로 추정되면 Bayes 방법이 사용됩니다 .
- EMS 방법을 사용하여 분산 성분이 음수로 추정되면 Bayes 방법이 사용됩니다 .
- 분산 성분에 교락이 있으면 유계 REML 방법이 사용되며 음의 분산 성분 추정값이 0 으로 설정됩니다 .

최선의 분석 (EMS 또는 REML) 선택 **최선의 분석 (EMS, REML 또는 Bayes) 선택** 옵션과 동일한 논리를 사용하여 EMS 또는 REML 중 최선의 분석을 선택합니다 . 그러나 이 옵션은 분산 성분이 음수인 경우에도 Bayes 방법을 사용하지 않습니다 . 유계 REML 방법이 사용되며 음의 분산 성분은 0 으로 강제 설정됩니다 .

REML 분석 사용 균형 데이터인 경우에도 유계 REML 방법을 사용합니다 . 유계 REML 방법은 불균형 데이터를 처리할 수 있으며 음의 분산 성분을 0 으로 강제 설정합니다 .

Bayes 분석 사용 Bayes 방법을 사용합니다 . Bayes 방법은 불균형 데이터를 처리할 수 있으며 모든 분산 성분을 양수의 0 이 아닌 값으로 강제 설정합니다 . 분산 성분에 교락이 있으면 유계 REML 방법이 사용되며 음의 분산 성분 추정값이 0 으로 설정됩니다 . JMP 에서 구현되는 이 방법은 수정된 버전의 Jeffreys 사전 확률을 사용하여 사후 평균을 계산합니다 . 자세한 내용은 Portnoy(1971) 및 Sahai(1974) 연구 자료에서 확인하십시오 .

최대 반복 수 (REML 방법에만 적용 가능) 어려운 문제의 경우 반복 횟수를 늘릴 수 있습니다 . 이 값을 늘리면 JMP 가 최적화 단계에서 해를 찾기 위해 더 많이 시행합니다 .

수렴 한계 (REML 방법에만 적용 가능) 더 높은 정밀도가 필요한 문제의 경우 수렴 한계를 더 작게 변경할 수 있습니다 . 이 값을 줄이면 JMP 가 최적화 단계에서 더 높은 수준의 정확도로 해를 찾습니다 . 그러나 이렇게 하면 해를 찾는 데 걸리는 시간이 길어질 수 있습니다 . 수렴 값이 클수록 결과가 빨리 반환되지만 정밀도는 더 낮습니다 .

통합 가로 좌표 수 (Bayes 방법에만 적용 가능) 정확도를 높이기 위해 통합 가로 좌표 수를 늘릴 수 있습니다 . 그러나 이렇게 하면 해를 찾는 데 걸리는 시간이 길어질 수 있습니다 . 이 값이 작을수록 결과가 빨리 반환되지만 정밀도는 더 낮습니다 .

함수 실행 최대 횟수 (Bayes 방법에만 적용 가능) 정확도를 높이기 위해 함수 실행 최대 횟수를 늘릴 수 있습니다 . 그러나 이렇게 하면 해를 찾는 데 걸리는 시간이 길어질 수 있습니다 . 이 값이 작을수록 결과가 빨리 반환되지만 정밀도는 더 낮습니다 .

게이지 R&R 방법 정보

게이지 R&R 방법은 작업자 변동(재현성) 및 측정 변동(반복성)으로 인한 측정 시스템의 변동이 어느 정도인지 분석합니다. 게이지 R&R 연구는 모형의 균형 여부에 관계없이 교차 모형과 내포 모형의 여러 조합에 사용할 수 있습니다.

팁 : EMP 방법을 사용하여 측정 시스템을 평가할 수도 있습니다 . 자세한 내용은 " 측정 시스템 분석 " 에서 확인하십시오 .

게이지 R&R 연구를 수행하기 전에 공정에서 전체 부품 크기 범위에 대해 부품의 랜덤 표본을 수집합니다 . 여러 작업자를 무작위로 선택하여 각 부품을 여러 번 측정합니다 . 그러면 변동은 다음과 같은 원인에 기인합니다 .

- 부품 간 **공정 변동** . 이 변동은 측정을 신뢰할 수 있는 경우 연구해야 할 최종 목표입니다 .
- 여러 측정 , 즉 **반복성**에 내재된 변동 . 표 5.1 에서는 이를 **범위 내 변동**이라고 합니다 .
- 여러 작업자가 부품을 측정하기 때문에 발생하는 변동 , 즉 **재현성**

게이지 R&R 분석을 수행하면 반복성 및 재현성을 기반으로 변동이 보고됩니다.

표 5.1 게이지 R&R 분석의 용어 및 합계 정의

분산 합	용어 및 약어	대체 용어
V(군내)	반복성(EV)	장비 변동
V(작업자)+V(작업자*부품)	재현성(AV)	평가자 변동
V(작업자*부품)	교호작용(IV)	교호작용 변동
V(군내)+V(작업자)+V(작업자*부품)	게이지 RR(RR)	측정 변동
V(부품)	부품 변동(PV)	부품 변동
V(군내)+V(작업자)+ V(작업자*부품)+V(부품)	총 변동(TV)	총 변동

Shewhart 관리도는 시간이 경과하면서 관리이탈 상태가 되는 공정을 식별할 수 있습니다. 또한 계량형 차트는 평균 또는 분산이 체계적으로 다른 작업자, 측정기 또는 부품 원인을 식별하는 데 도움이 됩니다.

게이지 R&R 옵션

게이지 R&R 옵션은 작업자 및 부품의 게이지 연구를 위해 해석되는 변동 측도를 보여 줍니다.

모형 유형을 아직 선택하지 않은 경우 **게이지 R&R** 옵션을 선택하면 모형 유형을 선택하는 대화 상자가 나타납니다. 그런 다음 게이지 R&R 규격을 수정합니다.

참고 : 계량형 차트에 대한 플랫폼 환경 설정에는 게이지 R&R 규격 대화상자 옵션이 포함됩니다. 이 환경 설정은 기본적으로 선택되어 있습니다. 데이터 테이블에 정의된 규격 한계를 사용하려면 이 환경 설정을 선택 취소합니다.

게이지 R&R 규격 입력 / 확인

" 게이지 R&R 규격 입력 / 확인 " 창에는 다음 옵션이 포함되어 있습니다.

공차 입력 방법 선택 다음 중 공차 입력 방법을 선택합니다.

공차 (USL – LSL) 를 직접 입력하려면 **공차 구간**을 선택합니다.

규격 한계를 입력한 후 공차를 자동으로 계산하려면 **LSL 및 / 또는 USL** 을 입력합니다.

K, 시그마 승수 **K** 는 시그마와 곱하기 위해 선택하는 상수 값입니다. 예를 들어 6 을 입력하면 6* 시그마 또는 6 시그마 공정을 표시할 수 있습니다.

팁 : 파일 > 환경 설정 > 플랫폼 > 계량형 차트를 선택하여 **K** 기본값을 수정할 수 있습니다.

공차 구간 , USL-LSL 공정에 대한 규격 상한과 규격 하한 사이의 차이를 나타내는 공차를 입력합니다 .

규격 한계 규격 상한 및 하한을 입력합니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

과거 평균 단측 규격 한계에 대한 공차 범위를 계산합니다(USL-과거 평균 또는 과거 평균-LSL). 과거 평균을 입력하지 않으면 총 평균이 사용됩니다 .

과거 시그마 변동을 설명하는 값을 입력합니다 (기록 또는 이전 경험에서 얻을 수 있음).

게이지 R&R 보고서

그림 5.7 게이지 R&R 보고서의 예

게이지 R&R

측정 소스		변동(6*표준편차)		다음에 대한 6*sqrt
반복성	(EV)	0.3860052	장비 변동	V(범위 내)
재현성	(AV)	0.4568005	평가자 변동	V(Operator) + V(Operator*part#)
Operator		0.0888194		V(Operator)
Operator*part#		0.4480823		V(Operator*part#)
게이지 R&R	(RR)	0.5980524	측정 변동	V(범위 내) + V(Operator) + V(Operator*part#)
부품 변동	(PV)	1.0425929	부품 변동	V(part#)
총 변동	(TV)	1.2019429	총 변동	V(범위 내) + V(Operator) + V(Operator*part#) + V(part#)

6 k
49.7571 % 게이지 R&R = 100*(RR/TV)
0.57362 부품 변동에 대한 정밀도 = RR/PV
2 개별 범주 수 = 1.41 * (PV/RR)
부품에 대해 마지막 열 'part#'을(를) 사용합니다.

전체 게이지 R&R 보고서

게이지 R&R

측정		변동(6*표준편차)		다음에 대한 6*sqrt
Operator*part#	(IV)	0.4480823	교호작용 변동	V(Operator*part#)
반복성	(EV)	0.3860052	장비 변동	V(범위 내)
재현성	(AV)	0.4568005	평가자 변동	V(Operator)+V(Operator*part#)
게이지 R&R	(RR)	0.5980524	측정 변동	V(범위 내)+V(Operator)+V(Operator*part#)
부품 변동	(PV)	1.0425929	부품 변동	V(part#)
총 변동	(TV)	1.2019429	총 변동	V(범위 내)+V(Operator)+V(Operator*part#)+V(part#)

6 k
49.7571 % 게이지 R&R = 100*(RR/TV)
0.57362 부품 변동에 대한 정밀도 = RR/PV
2 개별 범주 수 = 1.41 * (PV/RR)

축소 게이지 R&R 보고서

참고 : 축소 게이지 R&R 보고서를 생성하려면 **파일 > 환경 설정 > 플랫폼 > 계량형 차트** > **축소 게이지 R&R 보고서**를 선택합니다 .

이 예에서 "변동" 열의 값은 *k* 값(이 예의 경우 6)으로 척도화된 분산 성분 합의 제곱근입니다.

표 5.2에서는 Barrentine(1991)이 제안한 측정 변동에 대한 지침을 보여 줍니다.

표 5.2 허용 가능한 측정 변동 백분율

< 10%	우수
11% ~ 20%	양호
21% ~ 30%	조건부 허용
> 30%	부족

다음 사항에 유의하십시오 .

- " 게이지 R&R 규격 입력 / 확인 " 창에서 **공차 구간**을 입력한 경우 " 게이지 R&R" 보고서에 **% 공차** 열이 나타납니다 . 이 열은 100*(변동 / 공차) 로 계산됩니다 . 또한 보고서 아래쪽에 **정밀도 / 공차** 비율이 표시됩니다 . 이 비율은 게이지 변동으로 인해 손실된 공차 또는 공정 능력 구간의 비율을 나타냅니다 .
- " 게이지 R&R 규격 입력 / 확인 " 창에서 **과거 시그마**를 입력한 경우 " 게이지 R&R" 보고서에 **% 공정** 열이 나타납니다 . 이 열은 100*(변동 /(K* 과거 시그마)) 로 정의됩니다 .
- NDC(개별 범주 수) 는 가장 가까운 정수로 내림된 (1.41 * (PV/RR)) 값으로 정의됩니다 .

판별비

판별비는 특정 제품에 대해 주어진 측정의 상대적 유용성을 나타냅니다. 일반적으로 판별비가 2 보다 작으면 측정에서 제품 변동을 감지할 수 없으므로 측정 공정을 개선해야 합니다. 판별비가 4 보다 크면 허용할 수 없는 제품 변동을 적절하게 감지하므로 생산 공정을 개선해야 합니다.

자세한 내용은 "판별비에 대한 통계 상세 정보"에서 확인하십시오.

오분류 확률

측정 변동으로 인해 적합한 부품이 기각되고 잘못된 부품이 채택될 수 있습니다. 이것을 오분류라고 합니다. 측정 변동이 감소하면 오분류 비율도 줄어듭니다. 모형 유형 선택 및 규격 한계 입력을 아직 수행하지 않은 경우 **오분류 확률** 옵션을 선택하면 이 작업을 위한 대화상자가 나타납니다.

그림 5.8 오분류 확률 보고서의 예

오분류 확률

설명	확률
P(적합한 부품이 잘못 기각됨)	0.0802
P(잘못된 부품이 잘못 채택됨)	0.2787
P(부품이 적합하며 기각됨)	0.0735
P(부품이 잘못되었으며 채택됨)	0.0235
P(부품이 적합함)	0.9157

오분류 확률은 *Y*(측정된 부품 값) 와 *X*(실제 부품 값) 의 결합 확률 함수를 기반으로 합니다 . 사용된 결합 확률 밀도 함수는 이변량 정규 분포입니다 . 설명을 이해하기 위해 다음 확률을 정의합니다 .

$\delta = P[(LSL \le X \le USL) \text{ 및 } (Y < LSL \text{ 또는 } Y > USL)]$

$\beta = P[(X < LSL \text{ 또는 } X > USL) \text{ 및 } (LSL \le Y \le USL)]$

$\pi = P(LSL \le X \le USL)$

설명

P(적합한 부품이 잘못 기각됨) 부품이 적합한 경우 기각될 조건부 확률 , 즉 δ/π 입니다 .

P(잘못된 부품이 잘못 채택됨) 부품이 잘못된 경우 채택될 조건부 확률 , 즉 β/(1-π) 입니다 .

P(부품이 적합하며 기각됨) 부품이 적합하며 기각될 결합 확률 , 즉 δ 입니다 .

P(부품이 잘못되었으며 채택됨) 부품이 잘못되었으며 채택될 결합 확률 , 즉 β 입니다 .

P(부품이 적합함) 부품이 적합할 확률 , 즉 π 입니다 .

자세한 내용은 "오분류 확률에 대한 통계 상세 정보" 및 Burdick et al. 연구 자료 (2005) 에서 확인하십시오 .

편향 보고서

편향 보고서에는 전체 측정 편향 그래프와 표준별 측정 편향 그래프가 각각 요약 테이블과 함께 표시됩니다 . X 변수의 각 수준에 대해 관측된 값과 표준 값 사이의 차이 또는 편향 평균이 나타납니다 . 편향에 대한 *t* 검정도 제공됩니다 .

편향 보고서 옵션은 시작 창에서 **표준** 변수를 제공한 경우에만 사용할 수 있습니다 .

" 측정 편향 보고서 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

신뢰 구간 각 부품의 편향 평균에 대한 신뢰 구간을 계산하고 " 표준별 측정 편향 보고서 " 그림에 표식을 추가합니다 .

측정 오차 그래프 측정 오차 대 모든 그룹화 열의 그래프를 함께 생성합니다 . 각 그룹화 열별 측정 오차 그래프도 별도로 표시됩니다 .

선형성 연구

선형성 연구는 표준 변수를 X 변수로 사용하고 편향을 Y 변수로 사용하여 회귀 분석을 수행합니다 . 이 분석은 편향과 부품 크기 간의 관계를 검토합니다 . 기울기가 0 이면 가장 이상적입니다 . 기울기가 0 과 유의한 차이가 있으면 표준으로 측정된 변수 또는 부품 크기와 측정 능력 사이의 관계에 유의성이 있다는 결론을 내릴 수 있습니다 .

선형성 연구 옵션은 시작 창에서 **표준** 변수를 제공한 경우에만 사용할 수 있습니다 .

이 보고서에는 다음과 같은 정보가 표시됩니다 .

- 각 표준에 대한 편향 요약 통계량
- 선의 기울기가 0 인지 여부를 검정하는 분산분석표
- 기울기 (선형성) 및 절편 (편향) 에 대한 검정을 포함하는 선 모수 . 절편 검정은 기울기에 대한 검정이 기울기 = 0 의 가설을 기각하지 못하는 경우에만 유용합니다 .
- 그래프 바로 아래에 나타나는 선 방정식

"선형성 연구"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

유의 수준 설정 편향 신뢰 구간에 사용되는 유의 수준을 변경합니다 .

그룹별 선형성 시작 창에서 지정한 **X, 그룹화** 변수의 각 수준에 대한 개별 선형성 그림을 생성합니다 .

계량형 차트의 추가 예

- " 이분산성 검정의 예 "
- " 편향 보고서 옵션의 예 "

이분산성 검정의 예

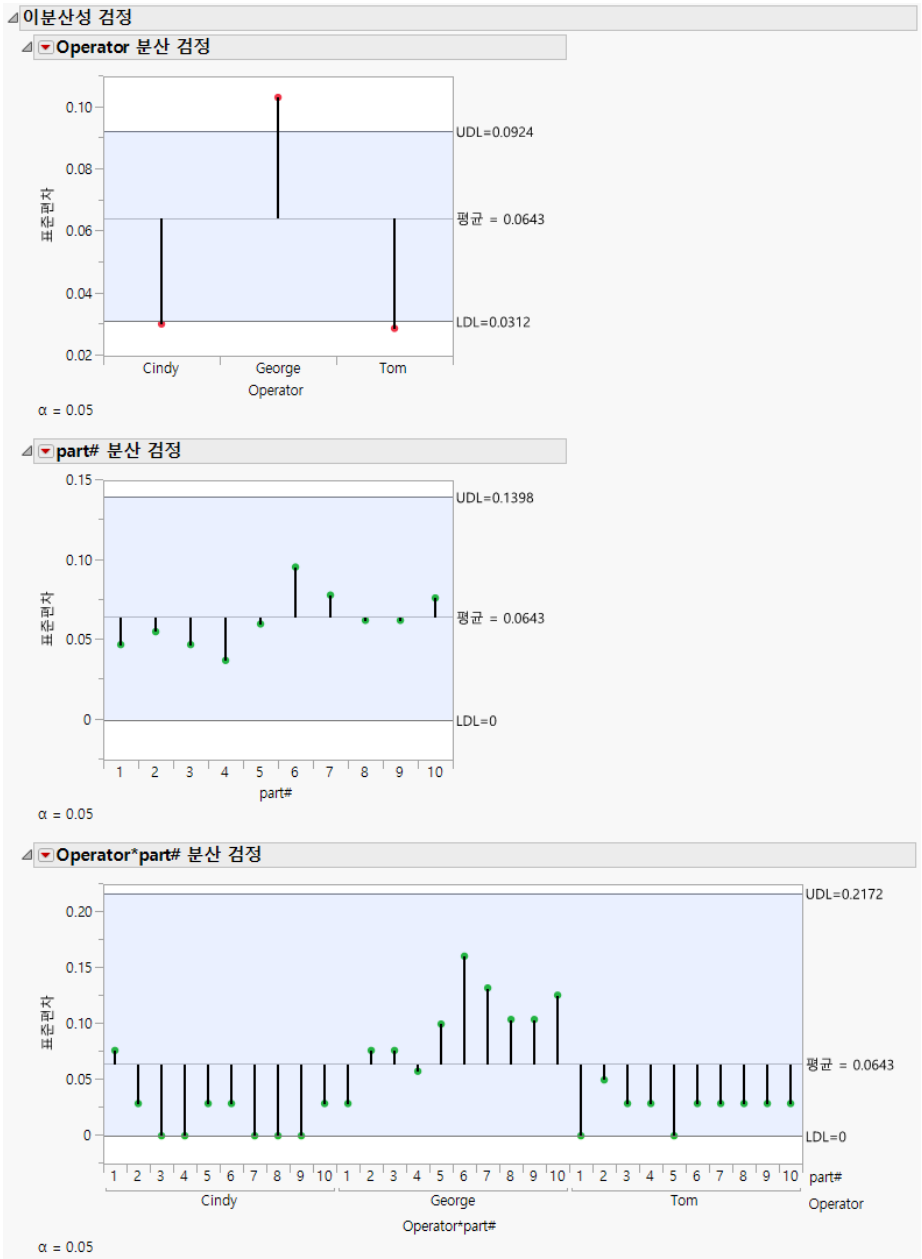
부품 측정값이 포함된 데이터가 있다고 가정해 보겠습니다. Cindy, George, Tom 이라는 세 명의 작업자가 각각 10 개의 부품을 측정했습니다 . 각 부품을 세 번씩 측정하여 총 90 개의 관측값이 수집되었습니다 . 다음 내용을 검토하려고 합니다 .

- 각 작업자에 대한 측정값의 분산이 동일한지 여부
- 각 부품에 대한 분산이 동일한지 여부
- 각 작업자 * 부품 조합에 대한 분산이 동일한지 여부

각 그룹에 대한 모든 분산이 통계적으로 동일한 것으로 간주되면 가장 이상적입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/2 Factors Crossed.jmp를 엽니다.
2. **분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트**를 선택합니다 .
3. Measurement 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Operator 를 선택하고 **X, 그룹화**를 클릭합니다 .
5. part# 을 선택하고 **부품 , 표본 ID** 를 클릭합니다 .
6. **확인**을 클릭합니다 .
7. " 계량형 게이지 " 의 빨간색 삼각형을 클릭하고 **이분산성 검정**을 선택합니다 .
8. **교차**를 선택합니다 .
9. **확인**을 클릭합니다 .

그림 5.9 이분산성 검정 보고서



참고 : 시뮬레이션을 사용하는 방법이므로 결정 한계가 매번 약간 다를 수 있습니다 .

"Operator 분산 검정"에서 세 수준이 모두 결정 상한 및 하한을 초과합니다. 이를 통해 각 작업자의 변동성이 평균 그룹 분산의 제곱근과 다르다는 결론을 내립니다. 각 작업자 간의 변동이 다른 이유를 검토할 수도 있습니다.

"part# 분산 검정" 및 "Operator*part# 분산 검정"(교호작용)의 경우 결정 한계를 초과하는 수준이 없습니다. 이를 통해 분산이 평균 그룹 분산의 제곱근과 통계적으로 다르지 않다는 결론을 내립니다. 즉, 각 부품(part)은 다른 부품과 비슷한 분산을 가지며 각 작업자*부품#(Operator*part#) 조합도 다른 조합과 비슷한 분산을 가집니다.

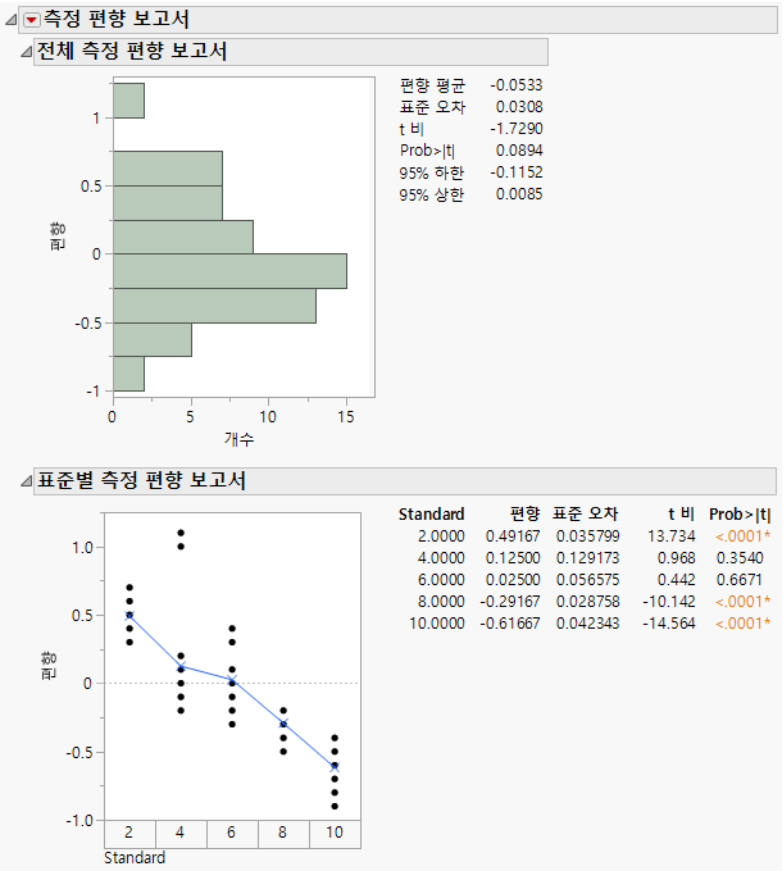
편향 보고서 옵션의 예

참고 : 이 데이터는 Automotive Industry Action Group(2002) 에서 가져왔습니다 .

공장 감독자가 공정에 새 측정 시스템을 도입하는 경우를 가정합니다 . PPAP(Production Part Approval Process) 의 일부로 측정 시스템의 편향 및 선형성을 평가해야 합니다 . 문서화된 공정 변동을 기반으로 측정 시스템의 전체 작동 범위에서 5개의 부품이 선택되었습니다. 레이아웃 검사로 각 부품을 측정하여 참조 값을 결정했습니다 . 그런 다음 책임 작업자가 각 부품을 12 번 측정했습니다 . 부품은 하루 중 무작위로 선택되었습니다 . 이 예에서는 전체 편향 및 개별 측정값 편향 (표준별) 을 검토하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/MSALinearity.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트**를 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Standard 를 선택하고 **표준**을 클릭합니다 .
5. Part 를 선택하고 **X, 그룹화**를 클릭합니다 .
6. **확인**을 클릭합니다 .
7. " 계량형 게이지 " 의 빨간색 삼각형을 클릭하고 **게이지 연구 > 편향 보고서**를 선택합니다 .

그림 5.10 측정 편향 보고서



모든 측정에 대한 편향(반응 - 표준)이 계산됩니다. "전체 측정 편향 보고서"에는 편향 히스토그램과 편향 평균이 0인지 확인하기 위한 *t* 검정이 표시됩니다. 편향 평균이 0이 아니라 -0.0533임을 알 수 있습니다. 그러나 신뢰 구간(-0.1152, 0.0085) 내에 0이 포함되어 있어 편향 평균이 0과 유의하게 다르지 않습니다. 유의 수준 0.05를 사용하면 *p* 값이 0.05보다 크다는 것을 알 수 있으며, 이는 편향 평균이 0과 유의하게 다르지 않음을 나타냅니다.

"표준별 측정 편향 보고서"에는 각 부품에 대한 편향 평균 값이 표시됩니다. 편향 평균이 모든 부품에 대한 실제 편향 값과 함께 그래프에 표시되므로 산포를 볼 수 있습니다. 이 예에서는 부품 번호 1(표준 값 2)이 높게 편향되어 있고 부품 4와 5(표준 값 8과 10)가 낮게 편향되어 있습니다.

팁 : 편향에 대한 신뢰 구간을 보려면 테이블에서 마우스 오른쪽 버튼을 클릭하고 **열 > 95% 하한 및 95% 상한**을 선택합니다 .

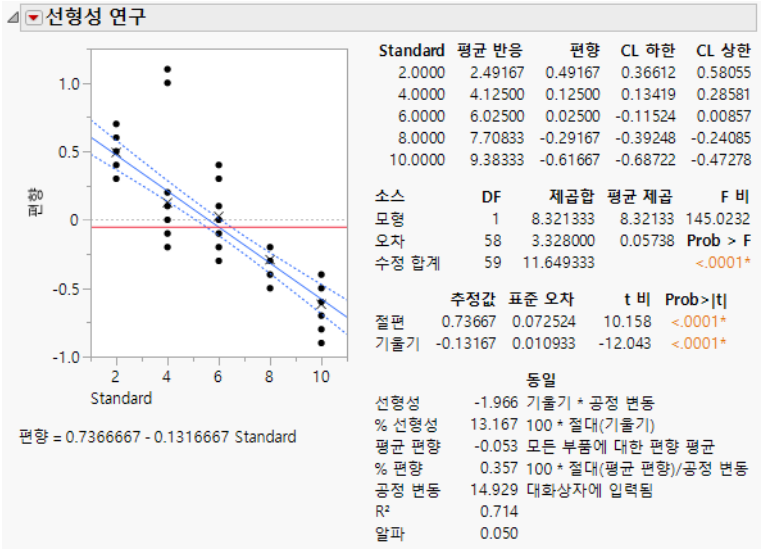
선형성 연구의 예

이제 **편향 보고서** 옵션과 동일한 데이터 및 시나리오를 사용하여 선형성을 검토할 수 있습니다. 이를 통해 부품 크기와 작업자의 측정 기량 사이에 유의한 관계가 있는지 여부를 판별합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/MSALinearity.jmp 를 엽니다.
2. **분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트**를 선택합니다.
3. Response 를 선택하고 **Y, 반응**을 클릭합니다.
4. Standard 를 선택하고 **표준**을 클릭합니다.
5. Part 를 선택하고 **X, 그룹화**를 클릭합니다.
6. **확인**을 클릭합니다.
7. " 계량형 게이지 " 의 빨간색 삼각형을 클릭하고 **게이지 연구 > 선형성 연구**를 선택합니다.
8. 공정 변동을 지정하라는 창이 나타나면 14.9286 을 입력합니다.

   14.9286 은 반응의 표준편차인 2.488105 의 6 배입니다.

그림 5.11 선형성 연구



다음 사항에 유의하십시오.

- 기울기는 -0.131667 입니다. 이 값은 그래프 아래의 방정식에 포함되며 세 번째 테이블에도 표시됩니다.
- 기울기 검정과 관련된 *p* 값 (<.0001) 은 매우 작습니다. 기울기에 대한 *t* 검정은 편향이 표준 값에 따라 변하는지 여부를 검정합니다.

p 값이 작으므로 부품 크기와 작업자의 측정 기량 사이에 유의한 선형 관계가 있다는 결론을 내릴 수 있습니다. 그래프에서도 이를 확인할 수 있습니다. 부품 또는 표준 값이 작으면 편향이 높고 값이 크면 편향이 낮습니다.

계량형 차트에 대한 통계 상세 정보

- "분산 성분에 대한 통계 상세 정보"
- "판별비에 대한 통계 상세 정보"
- "오분류 확률에 대한 통계 상세 정보"

분산 성분에 대한 통계 상세 정보

데이터 수집 방법에 따라 선택하는 정확 모형 유형이 달라집니다. 예를 들어 작업자가 동일한 부품을 측정 (교차 설계 사용) 할 수도 있고 다른 부품을 측정 (지분 설계 사용) 할 수도 있습니다. *B* 가 *A* 에 내포되고, 여러 측정값이 *B* 와 *A* 둘 다에 내포되며, *na*•*nb*•*nw* 측정값이 있는 모형을 가정하면 다음과 같은 명령문이 가능합니다.

- *na* 임의 효과는 *A* 에 기인합니다.
- *na*•*nb* 임의 효과는 *A* 내의 각 *nb B* 수준에 기인합니다.
- *na*•*nb*•*nw* 임의 효과는 *A* 에 내포된 *B* 내의 각 *nw* 수준에 기인합니다.

$$y_{ijk} = u + Za_i + Zb_{ij} + Zw_{ijk}.$$

Z 는 각 분류 수준에 대한 임의 효과입니다. *Z* 는 각각 평균이 0 이고 다른 모든 랜덤 항과 독립적이라고 가정합니다. 반응 *y* 의 분산은 각 *z* 성분으로 인한 분산의 합입니다.

$$\mathrm{Var}(y_{ijk}) = \mathrm{Var}(Za_i) + \mathrm{Var}(Zb_{ij}) + \mathrm{Var}(Zw_{ijk}).$$

표 5.3에서는 지원되는 모형 및 모형에 포함되는 효과를 보여 줍니다.

표 5.3 계량형 차트 플랫폼에서 지원되는 모형

모형	요인	모형의 효과
주효과	1	A
	2	A, B
	무제한	추가 요인에 대한 기타 효과

표 5.3 계량형 차트 플랫폼에서 지원되는 모형 (계속)

모형	요인	모형의 효과
교차	1	A
	2	A, B, A*B
	3	A, B, A*B, C, A*C, B*C, A*B*C
	4	A, B, A*B, C, A*C, B*C, A*B*C, D, A*D, B*D, A*B*D, C*D, A*C*D, B*C*D, A*B*C*D,
	무제한	추가 요인에 대한 기타 효과
내포	1	A
	2	A, B(A)
	3	A, B(A), C(A,B)
	4	A, B(A), C(A,B), D(A,B,C)
	무제한	추가 요인에 대한 기타 효과
교차 후 내포	3	A, B, A*B, C(A,B)
내포 후 교차	3	A, B(A), C, A*C, C*B(A)

판별비에 대한 통계 상세 정보

판별비는 측정값 M 의 총 분산을 측정 오차 E 의 분산과 비교합니다 . 내포 주효과를 포함하여 모든 주효과에 대해 판별비가 계산됩니다 . 판별비 D 는 다음과 같이 계산됩니다 .

$$D = \sqrt{2\left(\frac{P}{T-P}\right)+1}$$

다음은 각 요소에 대한 설명입니다 .

P = 요인에 대한 추정 분산
T = 추정된 총 분산

오분류 확률에 대한 통계 상세 정보

이 섹션에서는 오분류 확률 보고서의 확률 계산에 대해 설명합니다. 오분류 확률은 Y(측정된 부품 값) 와 X(실제 부품 값) 의 결합 확률 함수를 기반으로 합니다. 결합 확률 분포 함수 $F_{YX}(y, x)$ 는 평균 벡터 [μ, μ] 과 다음 공분산 행렬이 있는 이변량 정규 분포를 사용합니다.

$$\begin{bmatrix} \gamma_P + \gamma_M & \gamma_P \\ \gamma_P & \gamma_P \end{bmatrix}$$

여기서 γ_P는 부품 간 변동, γ_M은 측정 변동, μ는 총 평균을 나타냅니다. 이러한 통계량은 보고서 창의 통계량에서 찾거나 파생될 수 있습니다. 구체적으로 $\gamma_P + \gamma_M$은 TV(총 변동) 를 6 으로 나눈 값의 제곱, 즉 $(TV/6)^2$과 같고, γ_P는 PV(부품 변동) 를 6 으로 나눈 값의 제곱, 즉 $(PV/6)^2$과 같습니다. Y와 X 간의 상관 ρ_{YX}는 $\gamma_P/(\gamma_P + \gamma_M)$의 제곱근으로 정의됩니다.

이제 다음 확률을 정의합니다.

$\delta = P[(LSL \le X \le USL)$ 및 $(Y < LSL$ 또는 $Y > USL)]$

$\beta = P[(X < LSL$ 또는 $X > USL)$ 및 $(LSL \le Y \le USL)]$

$\pi = P(LSL \le X \le USL)$

이러한 확률은 결합 확률 분포 함수인 $F_{YX}(y, x)$와 Y 및 X에 대한 주변 확률 분포 함수인 $F_Y(y)$ 및 $F_X(x)$로 표현할 수 있습니다.

$\delta = F_{YX}(LSL, USL) - F_{YX}(LSL, LSL) - F_{YX}(USL, USL) + F_{YX}(USL, LSL) + F_X(USL) - F_X(LSL)$

$\beta = F_{YX}(USL, LSL) - F_{YX}(LSL, LSL) - F_{YX}(USL, USL) + F_{YX}(LSL, USL) + F_Y(USL) - F_Y(LSL)$

$\pi = F_X(USL) - F_X(LSL)$

P(적합한 부품이 잘못 기각됨) = δ/π

P(잘못된 부품이 잘못 채택됨) = $\beta/(1-\pi)$

P(부품이 적합하며 기각됨) = δ

P(부품이 잘못되었으며 채택됨) = β

P(부품이 적합함) = π

6 장

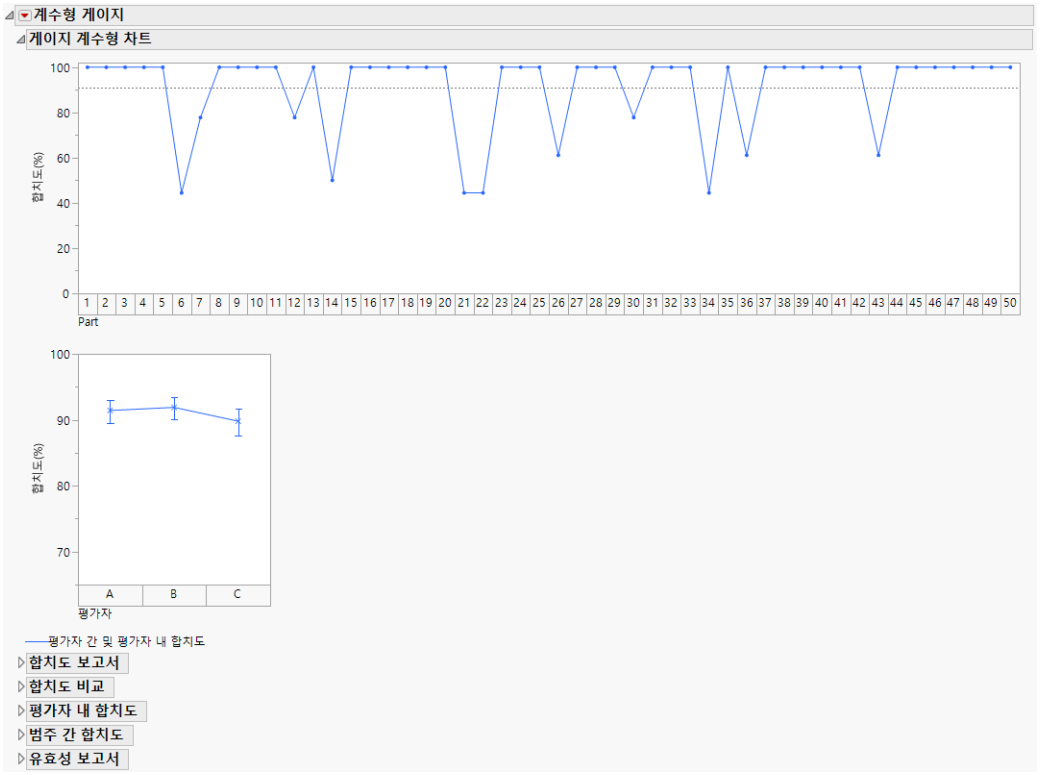
계수형 게이지 차트

합치도 측도를 사용하여 범주형 측정 공정 평가

계수형 차트는 범주형 측정값을 분석하여 반응(예: 평가자) 간의 합치 측도를 나타내는 데 도움이 될 수 있습니다. **계수형 데이터**에서 관심 변수는 한정된 수의 범주를 가집니다. 일반적으로 데이터에는 통과/실패와 같은 두 가지 결과만 가능합니다. 평가자가 얼마나 효과적으로 부품을 분류했는지, 평가자 간의 합치 정도 및 여러 평가 과정에서 각 평가자 자체의 합치 정도 등을 검토할 수 있습니다.

팁 : 이 장에서는 계수형 차트만 설명합니다 . 계량형 차트에 대한 자세한 내용은 " 계량형 게이지 차트 " 에서 확인하십시오 .

그림 6.1 계수형 차트의 예



목차

계수형 게이지 차트 개요 147
계수형 게이지 차트의 예 147
계량형 / 계수형 게이지 차트 플랫폼 시작 149
계수형 게이지 차트 및 보고서 150
합치도 보고서 151
유효성 보고서 152
계수형 게이지 플랫폼 옵션 153
계수형 게이지 차트에 대한 통계 상세 정보 154
합치도 보고서에 대한 통계 상세 정보 155

계수형 게이지 차트 개요

계수형 게이지 차트를 생성하기 전에 다음 지침을 사용하여 데이터 형식을 지정해야 합니다 .

- 평가자 간의 합치도를 비교하려면 데이터 테이블의 각 평가자가 개별 열에 있어야 합니다 . 이러한 열은 시작 창에서 **Y, 반응** 역할에 할당됩니다 . 그림 6.2 에서 각 평가자 (A, B, C) 는 개별 열에 있습니다 .
- 다른 열의 반응은 문자 (통과 또는 실패) 또는 숫자 (0 또는 1) 일 수 있습니다 . 그림 6.2 에서 평가자 반응은 숫자입니다(0 - 통과, 1 - 실패). 모든 반응 열의 데이터 유형이 동일해야 합니다.
- **X, 그룹화** 변수로 사용할 다른 모든 관심 변수는 각각 하나의 열에 쌓인 형식이어야 합니다 (그림 6.2 의 Part 열 참조). 표준 열도 정의할 수 있습니다 . 이 열은 평가자를 표준과 비교하는 보고서를 생성합니다 . 표준 열과 반응 열의 데이터 유형이 동일해야 합니다 .

그림 6.2 계수형 게이지 데이터

Attribute Gauge
잠긴 파일 C:₩Progra
Attribute Chart
열(7/0)
Part
Standard
Code
A
B
C
RefValue

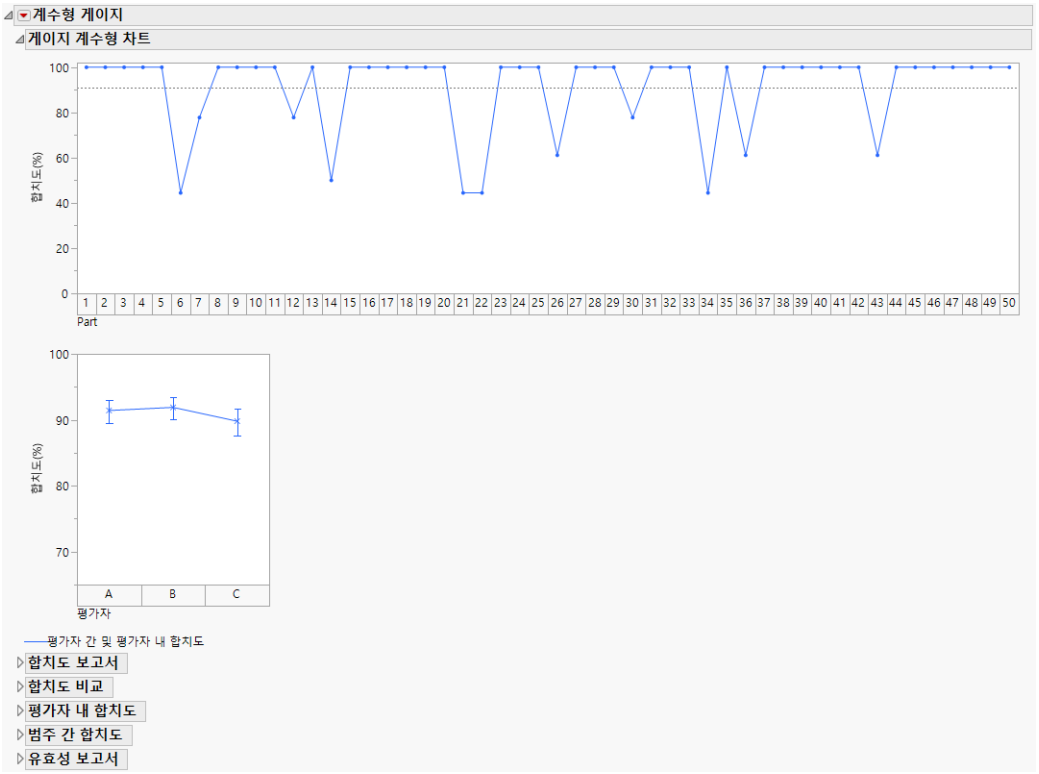
	Part	Standard	Code	A	B	C	RefValue
1	1	1	+	1	1	1	0.476901
2	1	1	+	1	1	1	0.476901
3	1	1	+	1	1	1	0.476901
4	2	1	+	1	1	1	0.509015
5	2	1	+	1	1	1	0.509015
6	2	1	+	1	1	1	0.509015
7	3	0	-	0	0	0	0.576459
8	3	0	-	0	0	0	0.576459
9	3	0	-	0	0	0	0.576459
10	4	0	-	0	0	0	0.566152
11	4	0	-	0	0	0	0.566152

계수형 게이지 차트의 예

부품 평가 (통과 또는 실패) 가 포함된 데이터가 있다고 가정해 보겠습니다 . A, B, C 로 식별되는 세 명의 평가자가 각각 50 개 부품에 대해 세 번씩 0(통과) 또는 1(실패) 을 기록했습니다 . 평가자가 부품을 올바르게 분류하는 데 얼마나 효과적이었는지 , 평가자 간의 합치 정도 및 평가 과정에서 평가자 자체의 합치 정도를 검토하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Attribute Gauge.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트**를 선택합니다 .
3. **차트 유형**에서 **계수형**을 선택합니다 .
4. A, B, C 를 선택하고 **Y, 반응**을 클릭합니다 .
5. Standard 를 선택하고 **표준**을 클릭합니다 .
6. Part 를 선택하고 **X, 그룹화**를 클릭합니다 .
7. **확인**을 클릭합니다 .

그림 6.3 계수형 차트의 예



첫 번째 차트(Part)에서는 각 부품에 대해 평가자 간의 합치도를 보여 줍니다. 예를 들어 6, 12, 14, 21, 22 등의 부품에 대해 합치도(%)가 떨어진 것을 확인할 수 있습니다. 해당 부품은 분류하기가 더 어려웠을 수 있습니다.

두 번째 차트(평가자)에서는 특정 부품에 대한 각 평가자의 자체 합치도 및 다른 평가자와의 합치도를 모든 부품에 대해 요약하여 보여 줍니다. 이 예에서는 평가자의 성과가 비교적 유사한 것으로 보입니다. 평가자 C의 합치도가 가장 낮지만 차이가 크지는 않습니다(91% 대신 약 89%).

8. " 유효성 보고서 " 를 열고 " 적합 보고서 " 가 나올 때까지 아래로 스크롤합니다 .

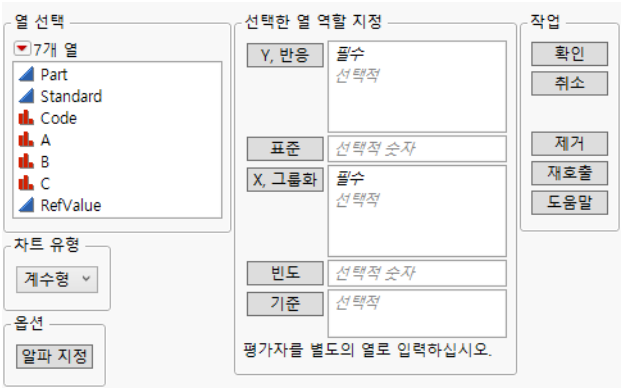
   0 = 부적합 (실패) 및 1 = 적합 (통과) 을 확인할 수 있습니다 . 그러나 이 데이터에서는 정반대로 0 이 통과이고 1 이 실패입니다 . 이 설정을 거꾸로 바꿔야 합니다 .

9. " 적합 보고서 " 의 빨간색 삼각형을 클릭하고 **적합 범주 변경**을 선택합니다 .

계량형 / 계수형 게이지 차트 플랫폼 시작

분석 > 품질 및 공정 > 계량형 / 계수형 게이지 차트를 선택하여 계량형 / 계수형 게이지 차트 플랫폼을 시작할 수 있습니다 . **차트 유형**을 **계수형**으로 설정합니다 .

그림 6.4 계량형 / 계수형 게이지 차트 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

차트 유형 계량형 게이지 분석 (연속형 반응의 경우) 또는 계수형 게이지 분석 (대개 " 통과 " 또는 " 실패 " 로 구분되는 범주형 반응의 경우) 중에서 선택합니다 .

> **참고 :** 이 장에서는 **계수형** 차트 유형에 대한 내용만 다룹니다 . **계량형** 차트 유형에 대한 자세한 내용은 " 계량형 게이지 차트 " 에서 확인하십시오 .

알파 지정 플랫폼에서 사용되는 유의 수준을 지정합니다 .

Y, 반응 각 평가자가 제공한 평가 열을 지정합니다 . 둘 이상의 평가 열을 지정해야 합니다 .

표준 부품에 대해 " 실제 " 또는 알려진 값을 포함하는 표준 또는 참조 열을 지정합니다 . 평가자를 표준과 비교하는 " 유효성 보고서 " 및 " 합치도 비교 " 보고서의 기타 섹션이 보고서 창에 나타납니다 .

X, 그룹화 측정값을 그룹화하는 분류 열을 지정합니다 . 요인이 내포된 계층 구조를 형성하는 경우 고차 항을 먼저 지정합니다 .

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 데이터 테이블 열을 식별합니다 . 요약 데이터가 있는 경우 유용하게 사용됩니다 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다 .

계수형 게이지 차트 및 보고서

계수형 게이지 차트에는 연구의 모든 부품에 대한 평가자 합치 측정값인 합치도(%)가 표시됩니다. 각 부품에 대한 합치도는 해당 부품의 모든 평가에 대해 모든 평가자 쌍의 평가를 비교하여 계산됩니다. 자세한 내용은 "계수형 게이지 차트에 대한 통계 상세 정보"에서 확인하십시오.

그림 6.5에 표시된 것과 같은 결과를 생성하려면 "계수형 게이지 차트의 예"에 설명된 방법을 따르십시오.

그림 6.5 계수형 게이지 차트

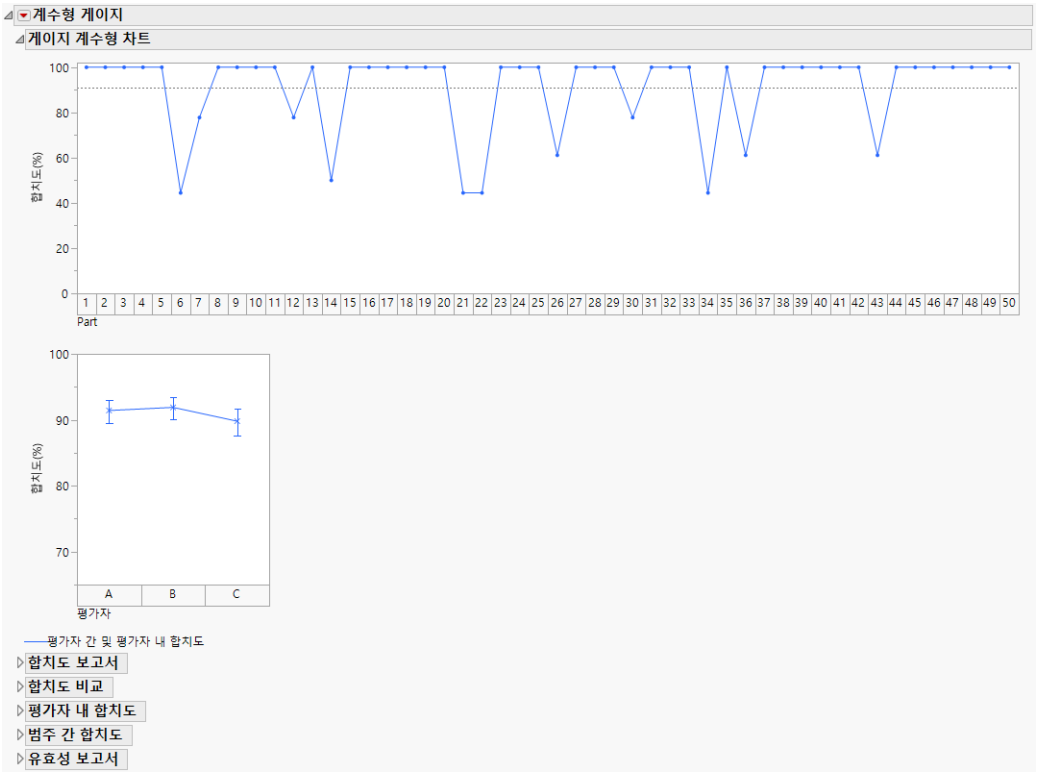


그림 6.5 의 첫 번째 차트에는 모든 X 그룹화 변수 (이 경우 'Part') 가 *x* 축에 사용됩니다 . 두 번째 차트에는 일반적으로 모든 Y 변수가 *x* 축에 사용됩니다 (이 경우 ' 평가자 ').

- 첫 번째 그래프에서는 합치도 (%) 값이 낮은 부품을 찾고 해당 부품의 측정값에 대해 평가자 의견이 일치하지 않는 이유를 조사할 수 있습니다 .
- 두 번째 그래프에서는 합치도 (%) 값이 낮은 평가자를 찾고 해당 평가자가 자체적으로 또는 다른 평가자와 의견이 일치하지 않는 이유를 조사할 수 있습니다 .

기타 옵션에 대한 자세한 내용은 "계수형 게이지 플랫폼 옵션"에서 확인하십시오.

합치도 보고서

참고 : 카파 값은 합치도를 나타내는 통계량입니다 . 카파 값이 1 에 가까울수록 합치도가 높습니다 . 카파 값이 0 에 가까울수록 합치도가 낮습니다 .

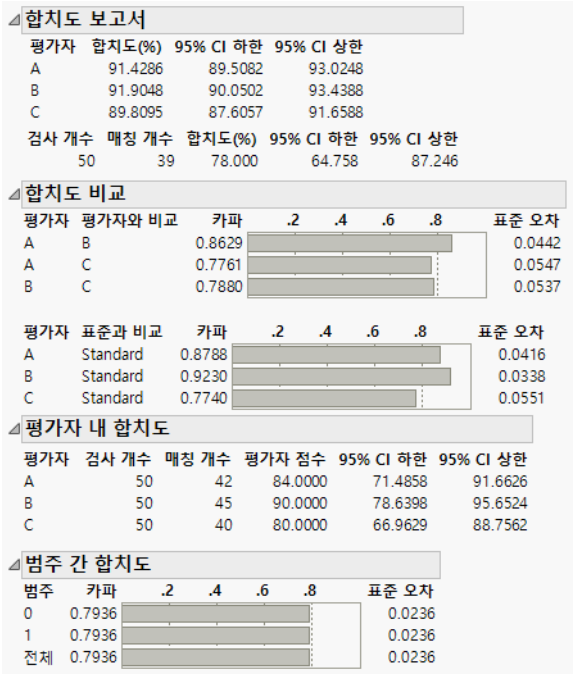
"합치도 보고서"에는 각 평가자에 대해 요약된 합치도와 전체 합치도가 표시됩니다. 이 보고서는 "계수형 게이지 차트" 보고서(그림 6.5)의 두 번째 차트에 표시된 데이터를 숫자 형식으로 나타낸 것입니다.

"합치도 비교" 보고서에서는 카파 통계량을 사용하여 각 평가자를 다른 모든 평가자와 비교하여 보여 줍니다. 평가자와 표준의 비교는 시작 창에서 표준 변수를 지정한 경우에만 수행됩니다.

"평가자 내 합치도" 보고서에는 검사한 항목 수가 표시됩니다. 신뢰 구간은 Agresti and Coull 연구 자료(1998)에서 제안된 스코어 신뢰 구간입니다. "매칭 개수"는 평가자가 개별 항목의 각 검사에 대해 동의한 검사 항목 수의 합입니다. "평가자 스코어"는 매칭 개수를 검사 개수로 나눈 값입니다.

"범주 간 합치도" 보고서에는 분류에서 우연에 의해 기대할 수 있는 합치도가 표시됩니다. 항목을 분류할 때 고정된 수의 평가자 간 합치도를 평가합니다.

그림 6.6 합치도 보고서



합치도 보고서

평가자	합치도(%)	95% CI 하한	95% CI 상한
A	91.4286	89.5082	93.0248
B	91.9048	90.0502	93.4388
C	89.8095	87.6057	91.6588

검사 개수	매칭 개수	합치도(%)	95% CI 하한	95% CI 상한
50	39	78.000	64.758	87.246

합치도 비교

평가자	평가자와 비교	카파	.2 .4 .6 .8	표준 오차
A	B	0.8629		0.0442
A	C	0.7761		0.0547
B	C	0.7880		0.0537

평가자	표준과 비교	카파	.2 .4 .6 .8	표준 오차
A	Standard	0.8788		0.0416
B	Standard	0.9230		0.0338
C	Standard	0.7740		0.0551

평가자 내 합치도

평가자	검사 개수	매칭 개수	평가자 점수	95% CI 하한	95% CI 상한
A	50	42	84.0000	71.4858	91.6626
B	50	45	90.0000	78.6398	95.6524
C	50	40	80.0000	66.9629	88.7562

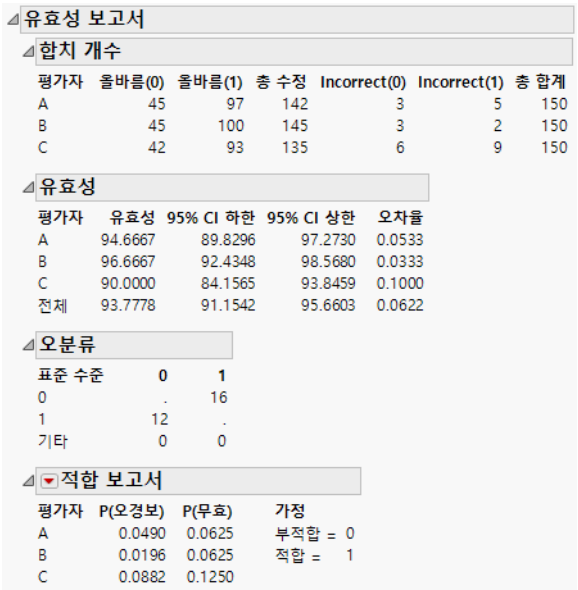
범주 간 합치도

범주	카파	.2 .4 .6 .8	표준 오차
0	0.7936		0.0236
1	0.7936		0.0236
전체	0.7936		0.0236

유효성 보고서

"유효성 보고서"는 시작 창에서 표준 변수를 지정한 경우에만 나타납니다. 표준 변수에 대한 설명은 "계량형/계수형 게이지 차트 플랫폼 시작"에서 확인하십시오. 이 보고서에서는 모든 평가자를 표준과 비교합니다.

그림 6.7 유효성 보고서



유효성 보고서

합치 개수

평가자	올바름(0)	올바름(1)	총 수정	Incorrect(0)	Incorrect(1)	총 합계
A	45	97	142	3	5	150
B	45	100	145	3	2	150
C	42	93	135	6	9	150

유효성

평가자	유효성	95% CI 하한	95% CI 상한	오차율
A	94.6667	89.8296	97.2730	0.0533
B	96.6667	92.4348	98.5680	0.0333
C	90.0000	84.1565	93.8459	0.1000
전체	93.7778	91.1542	95.6603	0.0622

오분류

표준 수준	0	1
0	.	16
1	12	.
기타	0	0

적합 보고서

평가자	P(오경보)	P(무효)	가정
A	0.0490	0.0625	부적합 = 0
B	0.0196	0.0625	적합 = 1
C	0.0882	0.1250	

"합치 개수" 테이블에는 표준의 모든 수준에 대해 일치 값과 불일치 값의 셀 개수가 표시됩니다. 그림 6.7에서 표준 변수는 두 개의 수준(0, 1)을 가집니다. 평가자 A는 수준 0에 대해 45개의 일치 반응과 3개의 불일치 반응을 기록하고 수준 1에 대해 97개의 일치 반응과 5개의 불일치 반응을 기록했습니다.

유효성은 일치 판단 수를 총 기회 수로 나눈 값으로 정의됩니다. 예를 들어 평가자 A가 모든 부품을 세 번 표집했다고 가정해 보겠습니다. 6번째 부품에 대한 판단 중 하나가 일치하지 않았습니다(예: 통과, 통과, 실패). 이때 나머지 두 판단은 일치 판단으로 간주됩니다. 이 유효성 정의는 MSA 3rd Edition과 다릅니다. MSA에 따르면 부품 6에 대한 평가자 A의 세 번의 기회가 모두 불일치 값으로 계산됩니다. 모든 검사를 개별적으로 포함하면 전체 검사 공정에 대해 더 많은 정보를 얻을 수 있습니다.

"유효성" 테이블에는 유효성에 대한 95% 신뢰 구간이 제공됩니다. 이 구간은 스코어 신뢰 구간입니다. 스코어 신뢰 구간은 특히 관측값이 경계에 가까이 있을 때 포함 확률을 증가시키는 것으로 입증되었습니다. 자세한 내용은 Agresti and Coull 연구 자료(1998)에서 확인하십시오.

"오분류" 테이블에서는 잘못된 라벨 지정을 보여 줍니다. 행은 표준의 수준 또는 허용되는 참조 값을 나타냅니다. 열은 평가자가 제공한 수준을 포함합니다.

적합 보고서

"적합 보고서"에는 오경보 확률과 무효 확률이 표시됩니다. "적합 보고서"는 평가 수준이 두 개(예: 통과/실패 또는 0/1)인 경우에만 나타납니다.

다음 설명이 적용됩니다.

오경보 부품이 실제로 적합한데 부적합한 것으로 결정됩니다.

무효 부품이 실제로 부적합한데 적합한 것으로 결정됩니다.

P(오경보) 부적합으로 잘못 판정된 부품의 수를 적합 판정된 총 부품 수로 나눈 값입니다.

P(무효) 적합으로 잘못 판정된 부품의 수를 실제로 부적합한 총 부품 수로 나눈 값입니다.

"적합 보고서"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

적합 범주 변경 적합한 것으로 간주되는 반응 범주를 반대로 바꿉니다.

탈출 속도 계산 부적합품이 생산되었지만 감지되지 않을 확률을 나타내는 탈출 속도를 계산합니다. 탈출 속도는 공정에서 부적합품이 생산될 확률에 무효 확률을 곱한 값으로 계산됩니다. 공정에서 부적합품을 생산할 확률, 즉 부적합 확률을 지정해야 합니다.

참고 : 이 플랫폼에서는 결측값이 별도의 범주로 간주됩니다. 이와 같은 별도의 범주를 만들지 않으려면 데이터 테이블에서 결측값 행을 제외하십시오.

계수형 게이지 플랫폼 옵션

"계수형 게이지"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

계수형 게이지 차트 게이지 계수형 차트 및 효율성 차트를 표시하거나 숨깁니다.

합치 점 표시 차트에 합치 점을 표시하거나 숨깁니다.

합치 점 연결 차트의 합치 점을 연결합니다.

평가자별 합치도 신뢰 구간 효율성 차트에 평가자별 합치도 신뢰 구간을 표시하거나 숨깁니다.

합치 그룹 평균 표시 게이지 계수형 차트에 합치 그룹 평균을 표시하거나 숨깁니다. 이 옵션은 X, 그룹화 변수를 두 개 이상 지정하는 경우에 사용할 수 있습니다.

합치 총 평균 표시 게이지 계수형 차트에 전체 합치 평균을 표시하거나 숨깁니다.

유효성 점 표시 차트에 유효성 점을 표시하거나 숨깁니다.

유효성 점 연결 차트의 유효성 점 사이에 선을 그립니다.

평가자별 유효성 신뢰 구간 계수형 게이지 차트 보고서의 두 번째 차트에 신뢰 구간을 표시하거나 숨깁니다 (그림 6.5).

유효성 보고서 유효성 보고서를 표시하거나 숨깁니다. 이 보고서에서는 카파 통계량을 사용하여 모든 평가자를 표준과 비교합니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

계수형 게이지 차트에 대한 통계 상세 정보

그림 6.5 에서 모든 **X, 그룹화** 변수가 x 축에 표시된 첫 번째 차트의 경우 합치도 (%) 는 다음과 같이 계산됩니다 .

$$\text{부품의 합치도(\%) } i = \frac{\sum_{l=1}^{k} \binom{\text{number of responses for level } l}{2}}{\binom{N_i}{2}}$$

그림 6.5 에서 모든 **Y, 반응** 변수가 x 축에 표시된 두 번째 차트의 경우 합치도 (%) 는 다음과 같이 계산됩니다 .

$$\text{평가자의 합치도(\%) } k = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{r_i} \text{반복 수가 j일 때 부품 i 내에서 이 평가자 k에 대한 number of uncounted matching levels}\right)}{\sum_{i=1}^{n}\left(\sum_{i=1}^{r_i} N_i - j\right)}$$

다음 사항에 유의하십시오 .

- n = 부품 수 (그룹화 변수)
- r_i = 부품 $i(i = 1,...,n)$ 의 반복 수
- m = 평가자 수
- k = 수준 수
- $N_i = m \times r_i$. 부품 $i(i = 1,...,n)$ 에 대한 평가 수 . 여기에는 모든 평가자의 반응 및 부품에 대한 반복 평가가 포함됩니다 . 예를 들어 3 명의 평가자가 부품 i 를 각각 3 번씩 측정하는 경우 N_i 는 $3 \times 3 = 9$ 입니다 .

한 부품을 각각 세 번씩 반복 실험하는 세 명의 평가자에 대한 데이터 테이블을 예로 들어보겠습니다.

표 6.1 평가자 A, B, C 에 대한 세 번의 반복 실험

	A	B	C
1	1	1	1
2	1	1	0
3	0	0	0

이 테이블을 사용하여 다음과 같은 계산을 수행할 수 있습니다 .

$$\text{합치도 (\%)} = \frac{\binom{4}{2} + \binom{5}{2}}{\binom{9}{2}} = \frac{16}{36} = 0.444$$

$$\text{합치도 (\%) [평가자 A] = 합치도 (\%) [평가자 B]} = \frac{4+3+3}{8+7+6} = \frac{10}{21} = 0.476,$$

$$\text{합치도 (\%) [평가자 C]} = \frac{4+3+2}{8+7+6} = \frac{9}{21} = 0.4286$$

합치도 보고서에 대한 통계 상세 정보

단순 카파 계수는 평가자 간 합치도의 측도입니다 .

$$\hat{\kappa} = \frac{P_0 - P_e}{1 - P_e}$$

다음은 각 요소에 대한 설명입니다 .

$$P_0 = \sum_i p_{ii}$$

및

$$P_e = \sum_i p_{i.} p_{.i}$$

두 개의 반응 변수를 n 개 부품에 대한 두 개의 독립된 평가로 볼 경우 평가자의 의견이 완전히 합치되면 카파 계수가 +1 이 됩니다. 관측된 합치도가 우연에 의한 합치 확률을 초과하면 카파 계수가 양수가 되며 그 크기는 합치 강도를 반영합니다. 실제로는 흔치 않지만 관측된 합치도가 우연에 의한 합치 확률보다 작으면 카파가 음수가 됩니다. 카파의 최소값은 주변 비율에 따라 -1 에서 0 사이입니다.

다음 방정식을 사용하여 단순 카파 계수의 점근적 분산을 추정합니다 .

$$\text{var} = \frac{A + B - C}{(1 - P_e)^2 n}$$

다음은 각 요소에 대한 설명입니다 .

$$A = \sum_i p_{ii}\left[1 - (p_{i.} + p_{.i})(1 - \hat{\kappa})\right]$$

$$B = (1 - \hat{\kappa})^2 \sum_{i \neq j} \sum p_{ij}(p_{.i} + p_{j.})^2$$

및

$$C = \left[\hat{\kappa} - P_e(1 - \hat{\kappa})\right]^2$$

카파가 그림에 표시되고 표준 오차도 제공됩니다.

참고 : 계수형 차트 플랫폼의 카파 통계량은 변수 수준이 불균형한 경우에도 표시됩니다 .

범주형 카파 통계량 (Fleiss 1981) 은 " 범주 간 합치도 " 보고서에 있습니다 .

다음과 같이 가정해 보겠습니다 .

- n = 부품 수 (그룹화 변수)
- m = 평가자 수
- k = 수준 수
- r_i = 부품 $i(i = 1,...,n)$ 의 반복 수
- $N_i = m \times r_i$. 부품 $i(i = 1, 2,...,n)$ 에 대한 평가 수 . 여기에는 모든 평가자의 반응 및 부품에 대한 반복 평가가 포함됩니다 . 예를 들어 2 명의 평가자가 부품 i 를 각각 3 번씩 측정하는 경우 N_i 는 3 x 2 = 6 입니다 .
- x_{ij} = 수준 $j(j = 1, 2,...,k)$ 에 대한 부품 $i(i = 1, 2,...,n)$ 의 평가 수

개별 범주 카파는 다음과 같이 정의됩니다 .

$$\hat{\kappa}_j = 1 - \frac{\sum_{i=1}^{n} x_{ij}(N_i - x_{ij})}{(\bar{p}_j \bar{q}_j) \sum_{i=1} N_i(N_i - 1)} \quad \text{여기서,} \quad \bar{p}_j = \frac{\sum_{i=1}^{n} x_{ij}}{\sum_{i=1} N_i} \quad \bar{q}_j = 1 - \bar{p}_j$$

전체 카파는 다음과 같이 정의됩니다 .

$$\hat{\bar{\kappa}} = \frac{\sum_{j=1}^{k} \bar{q}_j \bar{p}_j \hat{\kappa}_j}{\sum_{j=1}^{k} \bar{p}_j \bar{q}_j}$$

$\hat{\kappa}_j$ 및 $\hat{\bar{\kappa}}$ 의 분산은 다음과 같이 계산됩니다 .

$$\text{var}(\hat{\kappa}_j) = \frac{2}{nN(N-1)}$$

$$\text{var}(\hat{\bar{\kappa}}) = \frac{2}{\left(\sum_{j=1}^{k} \bar{p}_j \bar{q}_j\right)^2 nN(N-1)} \times \left[\left(\sum_{j=1}^{k} \bar{p}_j \bar{q}_j\right)^2 - \sum_{j=1}^{k} \bar{p}_j \bar{q}_j (\bar{q}_j - \bar{p}_j)\right]$$

$\hat{\kappa}_j$ 및 $\hat{\bar{\kappa}}$ 의 표준 오차는 부품당 평가 수가 같을 때만 표시됩니다(예 : 모든 $i = 1,\ldots,n$ 에 대해 $N_i = N$).

7 장

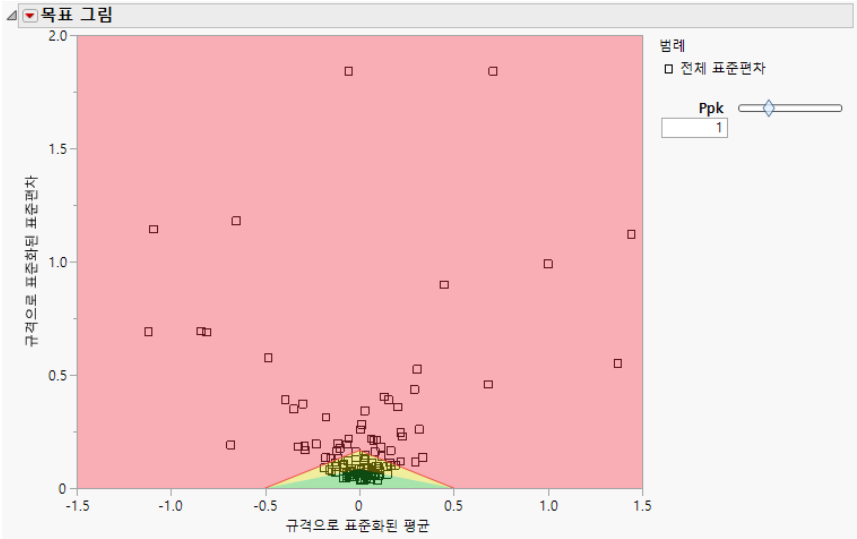
공정 능력
시간 경과에 따른 공정의 변동성 측정

공정 관리에 사용되는 공정 능력 분석은 지정된 규격 한계와 비교하여 공정이 얼마나 잘 수행되고 있는지를 측정합니다. 양호한 공정은 안정적이고 규격 한계 이내인 제품을 일관되게 생산하는 공정입니다. 공정 능력 지수는 공정 중심 및 변동성별로 요약된 공정 성능을 규격 한계와 관련짓는 측정값입니다.

목표 그림과 상자 그림 같은 그래픽 도구를 사용하면 규격 내에 있는 공정 또는 제품 특성을 시각적으로 빠르게 식별할 수 있습니다. 개별 상세 정보 보고서에는 분석의 각 변수에 대한 공정 능력 보고서가 표시됩니다. 분석을 수행하여 규격 또는 요구 사항을 기준으로 변동을 식별할 수 있으며, 이로 인해 점점 더 높은 적합 값을 얻을 수 있습니다.

부분군을 지정하여 공정의 전체 변동을 부분군 내 변동과 비교할 수 있습니다. 다양한 분포를 따르는 측정값을 생성하는 공정에 대한 공정 능력 지수를 계산할 수 있습니다. 지정된 분포를 따르지 않는 데이터의 경우 비모수 공정 능력 지수를 계산할 수 있습니다.

그림 7.1 공정 능력 플랫폼의 예



목차

공정 능력 플랫폼 개요 161
정규 변수가 있는 공정 능력 플랫폼의 예 163
비정규 변수가 있는 공정 능력 플랫폼의 예 165
공정 능력 플랫폼 시작 171
공정 선택 172
공정 부분군 지정 172
이동 범위 옵션 173
과거 정보 173
분포 옵션 174
기타 규격 175
규격 한계 입력 175
규격 한계 창 175
한계 데이터 테이블 176
규격 한계 열 특성 177
공정 능력 보고서 178
목표 그림 179
공정 능력 상자 그림 181
공정 능력 지수 그림 183
공정 능력 플랫폼 옵션 185
개별 상세 정보 보고서 188
정규화된 상자 그림 194
공정 성능 그림 195
요약 보고서 196
목표 그림 요약 테이블 생성 198
공정 능력 플랫폼의 추가 예 199
안정적인 공정에 대한 공정 능력 199
불안정한 공정에 대한 공정 능력 202
비정규 공정 Ppk에 대한 신뢰 한계 시뮬레이션 206
공정 능력 플랫폼에 대한 통계 상세 정보 213
변동 통계량 213
목표 그림 및 공정 능력 상자 그림에 대한 표기 216
목표 그림 217
목표값이 결측인 공정에 대한 공정 능력 상자 그림 218
정규 분포에 대한 공정 능력 지수 219
비정규 분포에 대한 공정 능력 지수: 백분위수 및 Z 스코어 방법 224
분포의 파라미터화 225

공정 능력 플랫폼 개요

공정 능력 플랫폼은 주어진 규격을 공정이 어느 정도 준수하는지를 측정하는 데 필요한 도구를 제공합니다. 기본적으로 정규 분포를 사용하여 적합시키는 변수에 대한 목표 그림, 공정 능력 상자 그림 및 공정 능력 지수 그림이 표시됩니다. 비정규 변수에 대한 공정 능력 지수는 공정 능력 지수 그림에 표시됩니다. 분석의 변수에 대해 정규화된 상자 그림, 요약 보고서 및 개별 상세 정보 보고서를 추가할 수 있습니다.

다음과 같은 여러 가지 방법으로 규격 한계를 제공할 수 있습니다 .

- 데이터 테이블에서 열 특성 사용
- 시작 창에서 규격 한계 대화상자 요청
- 규격 한계 데이터 테이블에서 한계 로드
- 규격 한계 관리 유틸리티 사용 (분석 > 품질 및 공정 > 규격 한계 관리)

양측, 단측 또는 비대칭 규격 한계를 지정할 수 있습니다.

참고 : 공정 능력 플랫폼은 분석 > 분포 및 분석 > 품질 및 공정 > 관리도를 통해 사용할 수 있는 공정 능력 분석의 확장입니다 .

공정 능력 지수

공정 능력 지수는 규격 한계를 충족하는 제품을 생산하기 위한 공정의 능력을 나타내는 비율입니다. 이 지수는 품질 특성의 평균 및 표준편차 추정값을 규격 한계와 관련시킵니다. 군내 공정 능력 추정값은 부분군 내 변동에서 생성된 표준편차의 추정값을 기반으로 합니다. 전체 공정 능력 추정값은 모든 공정 데이터에서 생성된 표준편차의 추정값을 사용합니다. 자세한 내용은 "정규 분포에 대한 공정 능력 지수" 및 "변동 통계량"에서 확인하십시오.

평균 또는 표준편차의 추정값은 중심 또는 산포와 관련된 공정이 **안정적**인 경우에만 잘 정의됩니다. 따라서 군내 공정 능력 지수를 해석하려면 공정 산포가 안정적이어야 합니다. 전체 공정 능력 지수를 해석하려면 공정 중심과 산포가 모두 안정적이어야 합니다.

작은 표본에서 생성된 공정 능력 지수는 변동성이 클 수 있습니다. 공정 능력 플랫폼은 대부분의 공정 능력 지수에 대한 신뢰 구간을 제공합니다. 이를 사용하여 품질 특성의 실제 공정 능력에 대해 가능한 값 범위를 결정할 수 있습니다.

참고 : JMP PRO 신뢰 구간이 제공되지 않는 경우 (예 : 비정규 분포의 경우) 시뮬레이션 기능을 사용하여 신뢰 구간을 생성할 수 있습니다 . 예는 " 비정규 공정 Ppk 에 대한 신뢰 한계 시뮬레이션 " 에서 확인하십시오 .

공정 능력 지수 값에 대한 지침은 Montgomery(2013)에서 확인할 수 있습니다. 최소 권장 값은 1.33 입니다. Six Sigma 이니셔티브는 DPPM(Defective Parts Per Million) 비율이 매우 낮은, 훨씬 높은 수준의 공정 능력 수준을 목표로 합니다.

비정규 공정에 대한 공정 능력 지수

공정 능력 플랫폼은 정규, 베타, 지수, 감마, Johnson, 로그 정규, 2 정규 혼합, 3 정규 혼합, SHASH 및 Weibull 분포를 사용하여 공정 측정에 대한 공정 능력 지수를 생성합니다. "최량 적합" 옵션은 이러한 분포 중에서 가장 적합한 분포를 결정하고 이 적합에 대한 공정 능력 지수를 제공합니다. 또한 이 플랫폼에는 공정 능력의 비모수 추정값을 제공하는 비모수 적합 옵션도 있습니다.

비정규 방법의 경우 ISO/분위수 방법(백분위수)과 Bothe/Z 스코어 방법(Z 스코어)이라는 두 가지 방식으로 추정값을 생성합니다. 이 방법에 대한 자세한 내용은 "비정규 분포에 대한 공정 능력 지수: 백분위수 및 Z 스코어 방법"에서 확인하십시오.

참고 : 개별 반응에 대한 공정 능력 분석은 분석 > 품질 및 공정 > 관리도 빌더를 통해 액세스할 수 있습니다 . 그러나 비정규 분포는 공정 능력 플랫폼에서만 사용할 수 있습니다 .

전체 및 군내 표준편차 추정값

공정 능력 플랫폼에서 대부분의 공정 능력 지수는 **전체**(장기) 변동 및 **부분군 내**(단기) 변동의 추정값을 기반으로 계산할 수 있습니다. 공정이 안정적이면 전체 변동과 부분군 내 변동이 비슷하므로 이러한 두 변동 측도에서 산출되는 결과도 비슷합니다. 정규화된 상자 그림과 요약 테이블은 전체 변동 또는 부분군 내 변동을 사용하여 계산할 수 있습니다. 안정적인 공정 및 불안정한 공정에 대한 공정 능력 지수를 계산하는 예는 "공정 능력 플랫폼의 추가 예"에서 확인하십시오.

시작 창에서 부분군 내 변동을 추정하기 위한 부분군을 지정할 수 있습니다. 부분군을 정의하는 열을 지정하거나 일정한 부분군 크기를 선택할 수 있습니다. 이러한 각 방법에 대해 비편향 표준편차의 평균을 사용하거나 범위 평균을 사용하여 공정 변동을 추정하도록 선택할 수 있습니다. 공정 능력 플랫폼에서 부분군을 지정하지 않으면 크기가 2인 부분군의 이동 범위를 사용하여 공정 변동의 부분군 내 추정값을 생성합니다. 마지막으로, 공정 표준편차의 추정값으로 사용할 과거 시그마를 지정할 수 있습니다.

공정 능력 지수 표기

공정 능력 플랫폼은 두 가지 공정 능력 지수 집합을 제공합니다 . 공정 능력 지수 계산에 대한 자세한 내용은 " 정규 분포에 대한 공정 능력 지수 " 에서 확인하십시오 .

- Cpk, Cpl, Cpu, Cp 및 Cpm. 이 지수는 공정 표준편차의 부분군 내 (단기) 추정값을 기반으로 합니다 .
- Ppk, Ppl, Ppu, Pp 및 Cpm. 이 지수는 공정 표준편차의 전체 (장기) 추정값을 기반으로 합니다 . 공정이 안정적이지 않으면 공정 표준편차가 존재하지 않습니다 . 자세한 내용은 Montgomery 연구 자료 (2013) 에서 확인하십시오 .

공정 능력 플랫폼은 공정 능력 지수에 적절한 AIAG 표기를 사용합니다. Ppk 라벨은 전체 변동 추정값에서 생성된 지수를 나타내고 Cpk는 부분군 내 변동 추정값에서 생성된 지수를 나타냅니다.

참고 : 기본적으로 "AIAG(Ppk) 라벨 " 플랫폼 환경 설정이 선택되어 있습니다 . " 공정 능력 " 에서 이 환경 설정을 선택 취소하면 Cp 표기만 사용하도록 보고를 변경할 수 있습니다 .

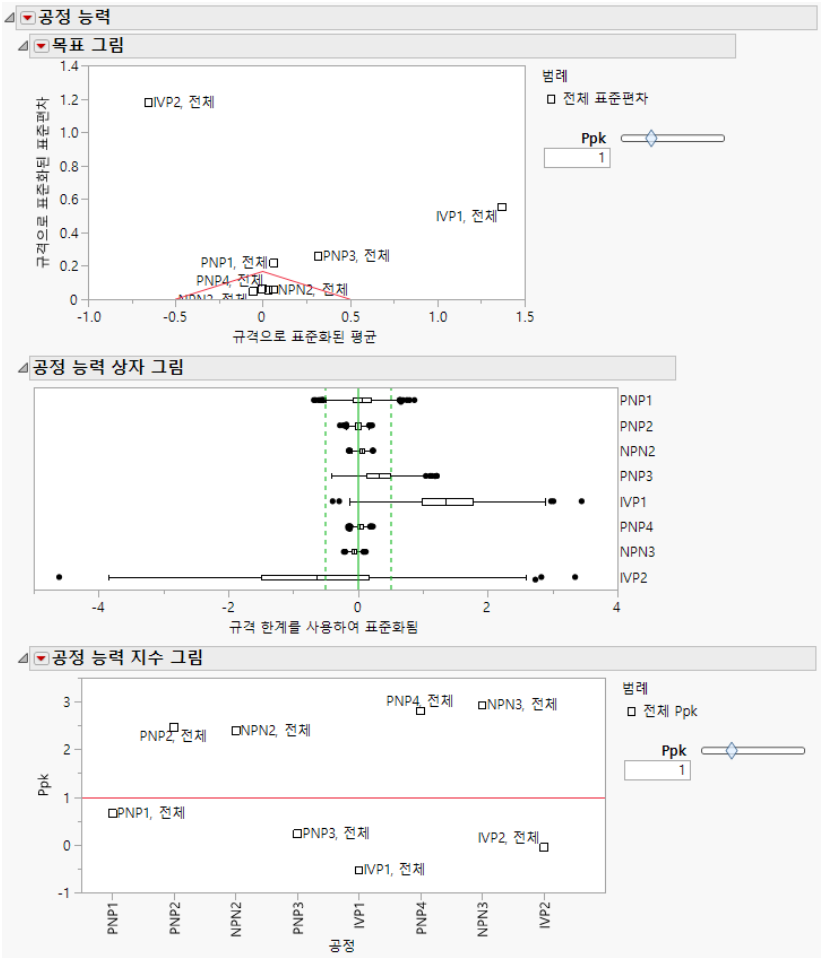
공정 능력 분석에 대한 자세한 내용은 Montgomery(2013) 및 Wheeler(2004) 연구 자료에서 확인할 수 있습니다.

정규 변수가 있는 공정 능력 플랫폼의 예

이 예에서는 Semiconductor Capability.jmp 샘플 데이터 테이블을 사용합니다 . 변수는 반도체 제조업체에서 가공 중인 웨이퍼에 대해 수행할 수 있는 표준 측정을 나타냅니다 . 열 특성 > 규격 한계 특성을 통해 변수의 규격 한계가 데이터 테이블에 입력되었습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Semiconductor Capability.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 공정 능력**을 선택합니다 .
3. Processes 옆의 흰색 삼각형을 클릭하여 모든 연속형 변수를 표시합니다 .
4. PNP1, PNP2, NPN2, PNP3, IVP1, PNP4, NPN3 및 IVP2 를 선택하고 **Y, 공정**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 목표 그림 " 의 빨간색 삼각형을 클릭하고 **전체 표준편차 점에 라벨 지정**을 선택합니다 .
7. " 공정 능력 지수 그림 " 의 빨간색 삼각형을 클릭하고 **전체 표준편차 점에 라벨 지정**을 선택합니다 .

그림 7.2 Semiconductor Capability.jmp 예의 결과



"목표 그림"에서는 각 변수에 대해 규격으로 정규화된 평균 변화를 x 축에 표시하고 규격으로 정규화된 표준편차를 y 축에 표시합니다. 그림의 아래쪽 가운데에 있는 빨간색 선으로 정의된 삼각형 영역이 목표 삼각형입니다. 이 삼각형은 공정 능력 지수 값 영역을 정의합니다. 그림 오른쪽의 Ppk 슬라이더를 사용하여 목표 삼각형을 조정할 수 있습니다. 슬라이더가 1로 설정된 경우 PNP1, PNP3, IVP1 및 IVP2가 목표 삼각형 외부에 있으므로 규격을 벗어난 것 같습니다.

"공정 능력 상자 그림" 보고서에는 분석의 각 변수에 대한 상자 그림이 표시됩니다. 각 열의 값은 목표값으로 중심화되고 규격 범위로 척도화됩니다. 이 예에서는 모든 공정 변수에 규격 상한과 규격 하한이 둘 다 있으며 이 한계는 목표값을 기준으로 대칭입니다. 따라서 녹색 실선은 목표값 위치를 나타내고 파선은 규격 한계를 나타냅니다.

IVP1의 경우 대부분의 점이 USL(규격 상한)을 초과하고 IVP2의 경우 대부분의 점이 목표값 미만인 것으로 나타납니다. PNP2는 모든 데이터 값이 규격 한계 내의 목표값에 있는 것 같습니다.

"공정 능력 지수 그림"에는 각 변수에 대한 Ppk 값이 표시됩니다. Ppk 값이 2 이상인 네 개 변수는 공정 능력이 충분한 공정의 변수입니다. 다른 네 개 변수의 Ppk 값은 1 미만입니다.

비정규 변수가 있는 공정 능력 플랫폼의 예

Process Measurements.jmp 데이터 테이블에는 제품을 생산하는 데 사용되는 7가지 공정에 대한 측정값이 포함되어 있습니다. 각 공정의 규격 한계는 열 특성으로 저장됩니다. 먼저 공정 데이터의 분포를 검토합니다. 데이터가 정규 분포가 아님을 알 수 있습니다. 그러면 공정 능력 플랫폼의 비정규 공정 능력 기능을 사용하여 공정 능력 지수를 계산합니다.

분포 보기

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Process Measurements.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. **열 선택** 목록에서 7 개 열을 모두 선택하고 **Y, 열**을 클릭합니다 .
4. **히스토그램만** 옆의 상자를 선택합니다 .
5. **확인**을 클릭합니다 .

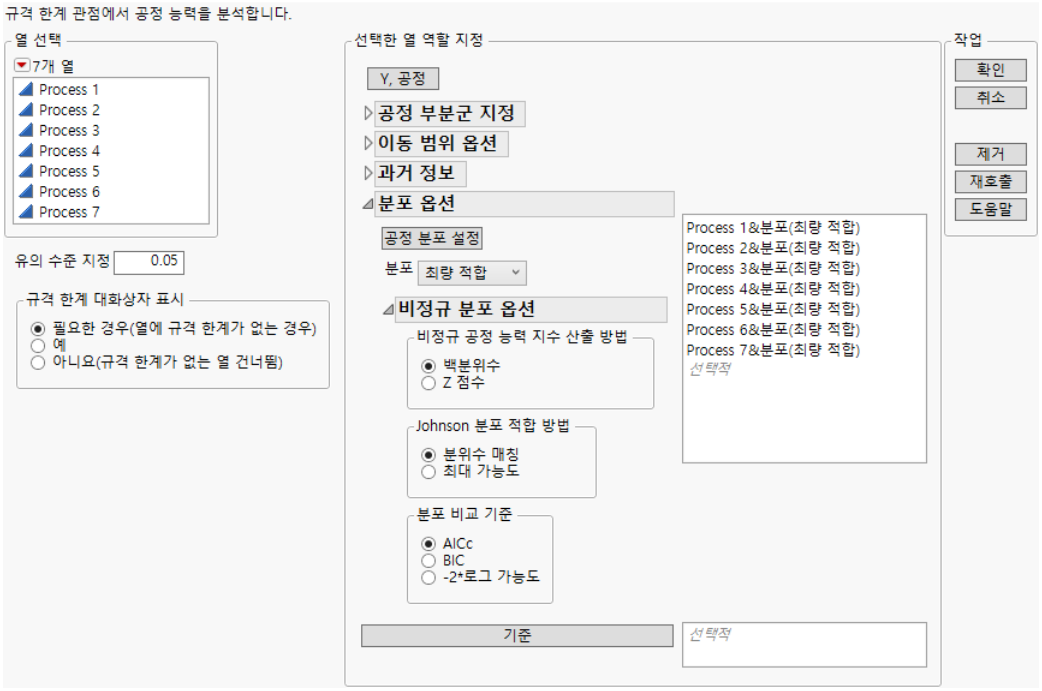
   대부분의 공정에서 히스토그램은 측정값의 이론적 분포가 편중되어 정규 분포를 따르지 않는다는 증거를 보여 줍니다 . 따라서 각 공정에 대해 가능한 모든 모수 분포 중에서 가장 적합한 분포를 찾습니다 .

공정 능력 분석 수행

1. **분석 > 품질 및 공정 > 공정 능력**을 선택합니다 .
2. **열** 목록에서 7 개 열을 모두 선택하고 **Y, 공정**을 클릭합니다 .
3. **Y, 공정** 목록에서 7 개 열을 모두 선택합니다 .
4. **분포 옵션** 패널을 열고 **분포** 목록에서 **최량 적합**을 선택합니다 .
5. **공정 분포 설정**을 클릭합니다 .

   "Y, 공정 " 목록의 각 변수 이름에 **& 분포 (최량 적합)** 접미사가 추가됩니다 . " 최량 적합 " 옵션은 가장 적합한 모수 분포를 각 변수에 적합시키도록 지정합니다 . 사용 가능한 모수 분포는 정규, 베타, 지수, 감마, Johnson, 로그 정규, 2 정규 혼합, 3 정규 혼합, SHASH 및 Weibull 입니다 (그림 7.3).
6. **비정규 분포 옵션** 개요를 엽니다 . " 비정규 공정 능력 지수 산출 방법 " 은 **백분위수** , "Johnson 분포 적합 방법 " 은 **분위수 매칭** , " 분포 비교 기준 " 은 **AICc** 로 각각 설정되어 있습니다 .

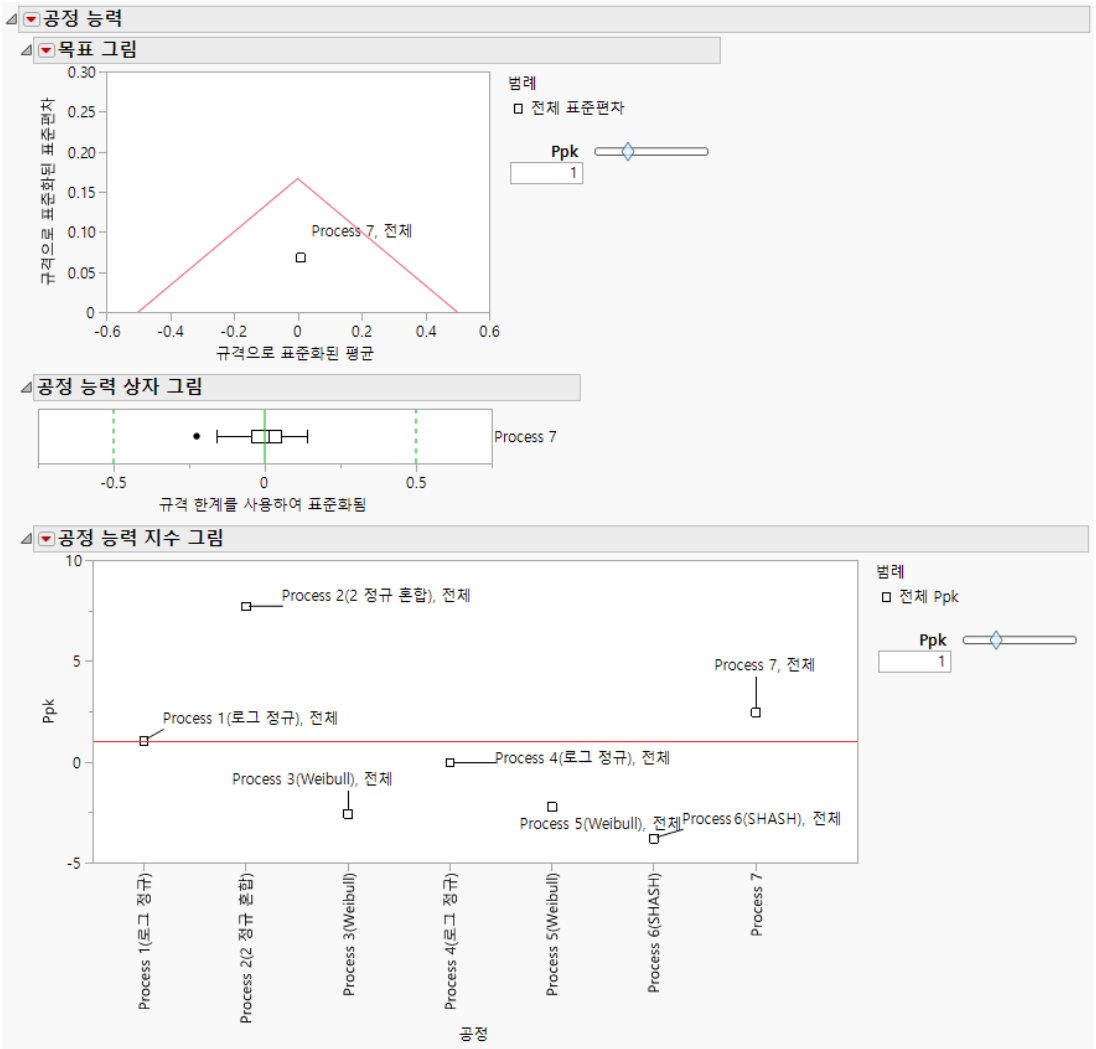
그림 7.3 완료된 시작 창



분위수 매칭 방법은 최대 가능도에 비해 안정성과 속도 때문에 Johnson 분포를 적합시킬 때 기본으로 사용됩니다. 최대 가능도는 분포 플랫폼에서 사용됩니다.

7. **확인**을 클릭합니다.
8. " 목표 그림 " 의 빨간색 삼각형을 클릭하고 **전체 표준편차 점에 라벨 지정**을 선택합니다.
9. " 공정 능력 지수 그림 " 의 빨간색 삼각형을 클릭하고 **전체 표준편차 점에 라벨 지정**을 선택합니다.

그림 7.4 변수에 라벨이 지정된 초기 보고서



참고 : 라벨을 클릭하여 드래그하면 그림을 더 의미 있게 해석할 수 있습니다 . " 공정 능력 지수 그림 " 의 오른쪽 프레임을 클릭하여 오른쪽으로 드래그하면 라벨을 더 쉽게 구분할 수 있습니다 .

" 목표 그림 " 에는 Process 7 에 해당하는 하나의 점만 표시됩니다 . " 공정 능력 상자 그림 " 보고서에는 Process 7 에 대한 단일 상자 그림이 표시됩니다 . 이는 Process 7 에 가장 적합한 분포가 정규 분포이기 때문입니다 .

10. " 공정 능력 지수 그림 " 의 오른쪽에서 Ppk 값을 2 로 설정합니다 .

    " 공정 능력 지수 그림 " 에는 7 개 공정 모두에 대한 Ppk 값이 표시됩니다 . 두 공정 (Process 2, Process 7) 만 공정 능력 값이 2 를 초과합니다 . " 공정 능력 지수 그림 " 에서 변수 이름 오

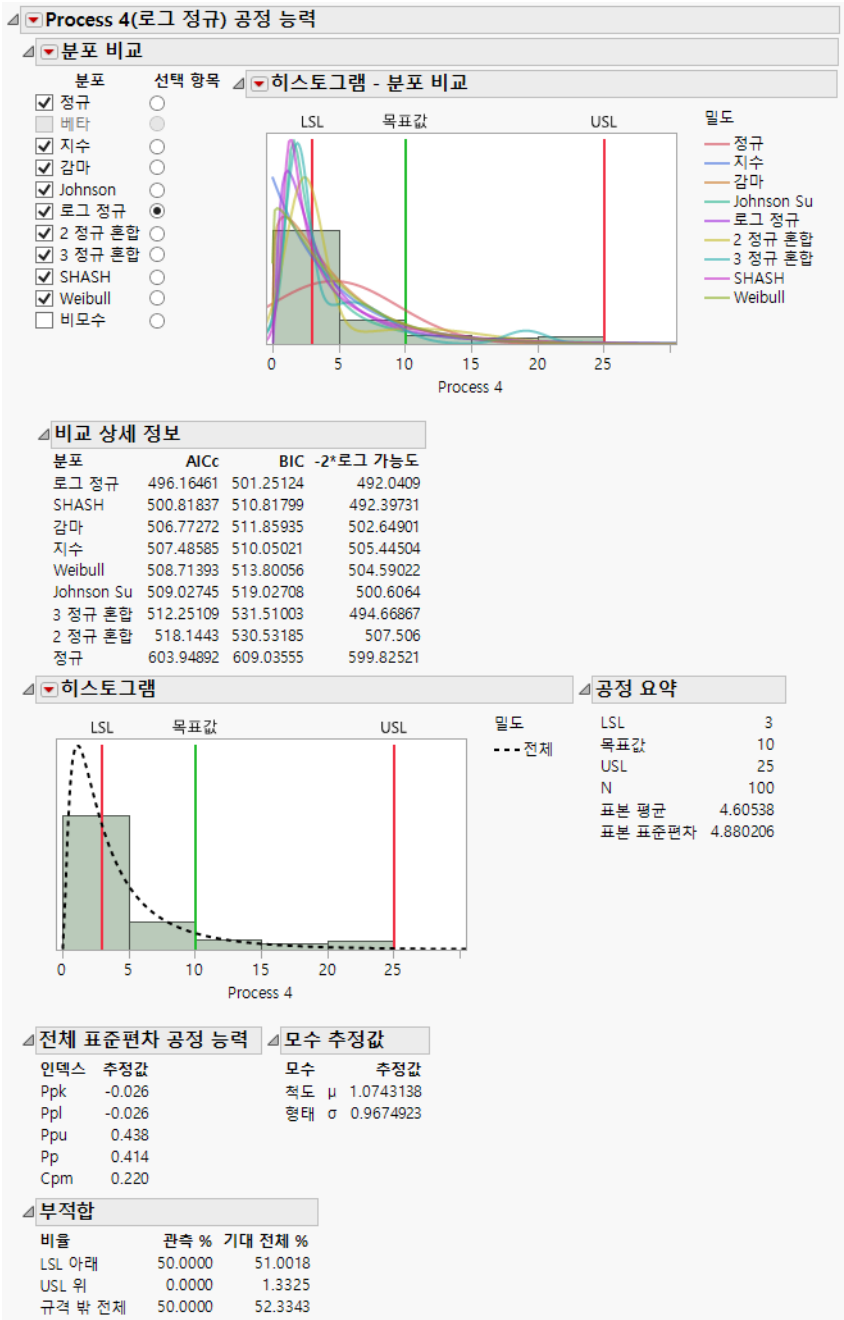
른쪽의 괄호 안에 가장 적합한 비정규 분포가 표시됩니다 . Process 7 의 가장 적합한 분포는 정규 분포이므로 표시되지 않습니다 .

11. " 공정 능력 " 의 빨간색 삼각형을 클릭하고 **개별 상세 정보 보고서**를 선택합니다 .

    시작 창에서 최량 적합을 요청했으므로 각 분포의 빨간색 삼각형 메뉴에서 " 분포 비교 " 옵션이 선택되었습니다 .

12. **Process 4(로그 정규) 공정 능력** 보고서가 나올 때까지 스크롤합니다 .

그림 7.5 Process 4 의 개별 상세 정보 보고서



Process 4 보고서의 제목은 로그 정규 적합을 기반으로 공정 능력이 계산된다는 것을 나타냅니다 . " 분포 비교 " 보고서에서 " 비모수 " 와 " 베타 " 를 제외한 모든 체크박스가 선택되어 있

으므로 이 9 개 분포가 적합하다는 것을 나타냅니다 . 이 결과는 시작 창에서 최량 적합을 요청했기 때문입니다 ." 선택 항목 " 열에서 선택된 버튼은 공정 능력 및 부적합 추정을 위해 "Process 4(로그 정규) 공정 능력 " 보고서의 나머지 부분에 사용된 분포가 로그 정규 분포임을 나타냅니다 .

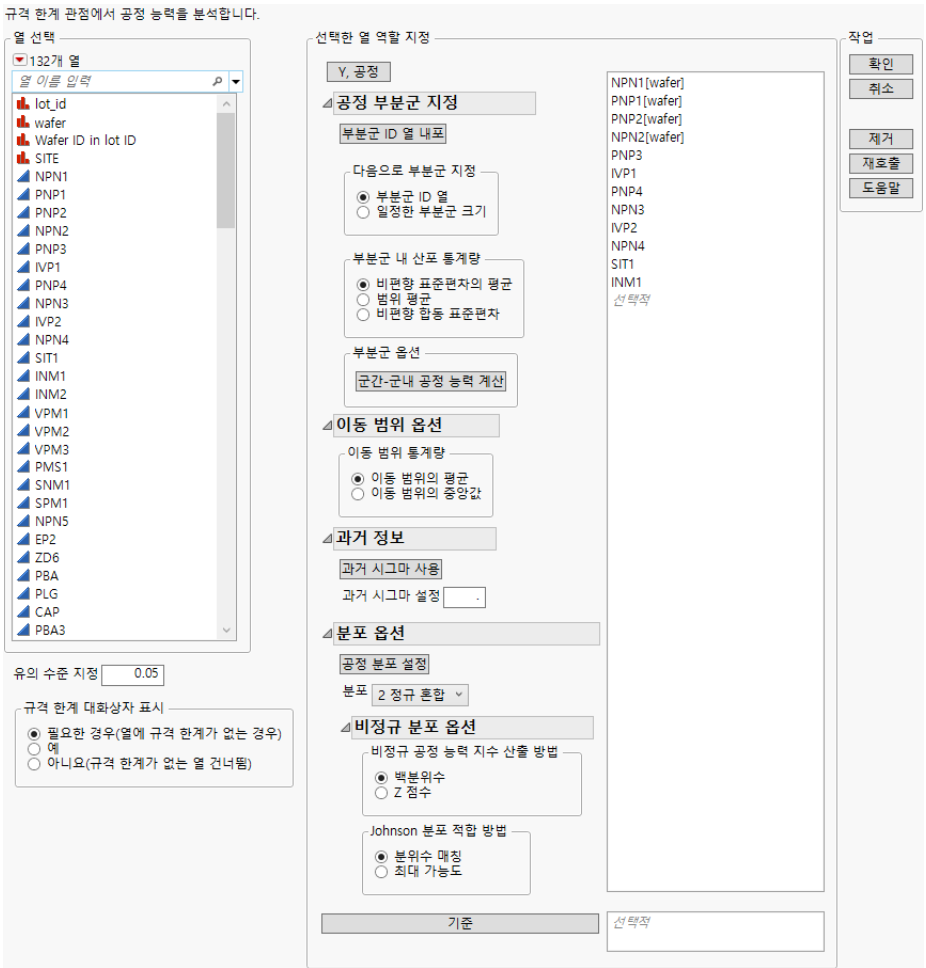
" 분포 비교 " 보고서를 사용하면 9 개의 분포 적합을 비교할 수 있습니다 ." 히스토그램 - 분포 비교 " 보고서에서는 적합을 시각적으로 평가할 수 있으며 " 비교 상세 정보 " 보고서에서는 선택한 분포의 적합 통계량을 보여 줍니다 . 그림과 적합 통계량 모두 로그 정규 분포가 선택한 분포 중에서 가장 적합하다는 것을 나타냅니다 .

" 개별 상세 정보 보고서 " 에 기본으로 표시되는 정보에는 추정된 최량 적합 분포를 보여 주는 히스토그램 , 공정 정보 요약 , 전체 표준편차 추정값에 기반한 공정 능력 지수 , 적합된 로그 정규 분포의 모수 추정값 , 관측 및 기대 부적합 수준이 포함됩니다 .

공정 능력 플랫폼 시작

분석 > 품질 및 공정 > 공정 능력을 선택하여 공정 능력 플랫폼을 시작할 수 있습니다 . Semiconductor Capability.jmp 데이터 테이블을 사용하는 그림 7.6 에는 모든 개요와 패널이 열려 있습니다 .

그림 7.6 공정 능력 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

공정 능력 시작 창에는 다음 개요 및 옵션이 포함되어 있습니다 .

- " 공정 선택 "
- " 공정 부분군 지정 "

- "이동 범위 옵션"
- "과거 정보"
- "분포 옵션"
- "기타 규격"

다음 중 하나에 해당하는 경우를 제외하고 시작 창에서 "확인"을 클릭하면 "규격 한계" 창이 나타납니다.

- 모든 열에 규격 한계가 포함된 경우
- 시작 창에서 **아니요 (규격 한계가 없는 열 건너뜀)** 를 선택한 경우

시작 창에서 "예"를 선택하는 경우에도 "규격 한계" 창이 나타납니다. 그렇지 않으면 "공정 능력" 보고서 창이 나타납니다.

공정 선택

공정 능력 분석에 포함할 공정 변수를 선택합니다.

Y, 공정 분석하려는 변수를 할당합니다.

참고 :

- 공정 능력 시작 창의 "열 선택" 목록에는 "변환" 메뉴를 사용할 수 없습니다. 공정 능력에서 사용할 변환 열을 생성하려면 데이터 테이블에서 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **새 계산식 열**을 선택합니다. 새 계산식 열 생성에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.
- 가상으로 결합된 테이블의 참조 열은 공정 능력 플랫폼에서 사용할 수 없습니다.

공정 부분군 지정

이 옵션 그룹을 사용하면 "Y, 공정" 목록의 각 변수에 부분군 ID 열 또는 일정한 부분군 크기를 할당할 수 있습니다.

ID 열을 사용하여 부분군 생성

1. "Y, 공정" 목록에서 하나 이상의 변수를 선택합니다.
2. **다음으로 부분군 지정** 옵션에서 **부분군 ID 열**을 선택합니다.
3. "열 선택" 목록에서 부분군 ID 열을 선택합니다.
4. **부분군 ID 열 내포**를 클릭합니다.

"Y, 공정" 목록에서 변수 이름 오른쪽의 대괄호 안에 부분군 ID 열이 나타납니다.

일정한 부분군 크기를 사용하여 부분군 생성

1. "Y, 공정" 목록에서 하나 이상의 변수를 선택합니다.

2. **다음으로 부분군 지정** 옵션에서 **일정한 부분군 크기**를 선택합니다 .
3. **일정한 부분군 크기 설정** 옆에 부분군 크기를 입력합니다 .
4. **크기별 부분군 지정**을 클릭합니다 .

"Y, 공정 " 목록에서 변수 이름 오른쪽의 대괄호 안에 부분군 크기가 나타납니다 .

부분군 ID 열 내포 (" 부분군 ID 열 " 을 선택한 경우 사용 가능) " 열 선택 " 목록에서 선택한 열을 할당하여 선택한 "Y, 공정 " 열의 부분군을 정의합니다 .

크기별 부분군 지정 (" 일정한 부분군 크기 " 를 선택한 경우 사용 가능) " 일정한 부분군 크기 설정 " 상자에 지정한 부분군 크기를 할당하여 선택한 "Y, 공정 " 열의 부분군을 정의합니다 .

일정한 부분군 크기 설정 (" 일정한 부분군 크기 " 를 선택한 경우 사용 가능) 선택한 "Y, 공정 " 열에 사용할 일정한 부분군 크기를 지정합니다 . " 크기별 부분군 지정 " 을 사용하여 이 값을 할당해야 합니다 .

부분군 내 변동 통계량 (" 공정 부분군 지정 " 을 사용하는 경우 사용 가능) 부분군 내 표준편차 추정값을 계산할 때 표준편차를 사용할지 아니면 범위를 사용할지 지정합니다 .

군간 - 군내 공정 능력 계산 (" 공정 부분군 지정 " 을 사용하는 경우 사용 가능) 부분군의 군간 - 군내 표준편차 추정값을 공정 능력 분석에 사용하도록 지정합니다 .

이동 범위 옵션

이 개요를 사용하면 부분군 지정이 사용되지 않을 때 군내 표준편차 추정값에 사용할 이동 범위 통계량을 지정할 수 있습니다 .

참고 : 부분군을 지정하고 **군간 - 군내 공정 능력 계산**을 클릭하는 경우 " 이동 범위 옵션 " 개요를 사용하여 군간 표준편차 추정값에 사용할 이동 범위 통계량을 지정할 수 있습니다 .

이동 범위 평균 이동 범위의 평균을 사용하여 표준편차를 추정합니다 . 이동 범위는 연속된 두 점 사이의 차이입니다 .

이동 범위 중앙값 이동 범위의 중앙값을 사용하여 표준편차를 추정합니다 .

과거 정보

이 개요를 사용하면 과거에 승인된 표준편차 값을 "Y, 공정 " 목록의 변수에 할당할 수 있습니다 .

1. "Y, 공정 " 목록에서 하나 이상의 변수를 선택합니다 .
2. " 과거 시그마 설정 " 옆에 값을 입력합니다 .
3. " 과거 시그마 사용 " 을 선택하여 해당 값을 선택한 변수에 할당합니다 .

지정한 값이 "Y, 공정 " 목록에서 변수 이름 오른쪽의 "& 시그마 ()" 표현식 괄호 안에 나타납니다 .

참고 : 과거 시그마를 설정하면 선택한 공정 변수에 대한 부분군 할당이 더 이상 관련이 없으며 제거됩니다 .

분포 옵션

별도로 지정한 경우를 제외하고 모든 "Y, 공정 " 변수는 정규 분포를 따른다는 가정을 사용하여 분석됩니다 . " 분포 옵션 " 개요를 사용하면 "Y, 공정 " 목록의 변수에 다른 분포 또는 계산 방법을 할당하고 비정규 계산과 관련된 옵션을 지정할 수 있습니다 .

- 사용 가능한 분포는 정규 , 베타 , 지수 , 감마 , Johnson, 로그 정규 , 2 정규 혼합 , 3 정규 혼합 , SHASH 및 Weibull 분포입니다 . Johnson 분포 외에는 최대 가능도 추정을 사용하여 분포를 적합시킵니다 . 자세한 내용은 "Johnson 분포 적합 방법 " 에서 확인하십시오 .
- " 최량 적합 " 옵션은 사용 가능한 분포 중에서 가장 적합한 분포를 결정하고 이 적합을 적용합니다 .
- " 비모수 " 옵션은 커널 밀도 추정을 사용하여 분포를 적합시킵니다 .

비정규 적합과 관련된 추가 옵션은 "비정규 분포 옵션"에서 확인하십시오.

분포 지정

1. "Y, 공정 " 목록에서 하나 이상의 변수를 선택합니다 .
2. " 분포 " 목록에서 분포를 선택합니다 .
3. " 공정 분포 설정 " 을 선택하여 해당 분포를 선택한 변수에 할당합니다 .

지정한 분포가 "Y, 공정" 목록에서 변수 이름 오른쪽의 "&분포()" 표현식 괄호 안에 나타납니다.

참고 : " 정규 " 를 제외한 다른 분포를 선택하면 부분군 ID 열 또는 과거 시그마를 할당할 수 없습니다 . 이러한 선택은 비정규 공정 능력 지수를 계산하는 데 사용되는 방법에서 지원되지 않습니다 . 자세한 내용은 " 비정규 분포에 대한 공정 능력 지수 : 백분위수 및 Z 스코어 방법 " 에서 확인하십시오 .

비정규 분포 옵션

비정규 공정 능력 지수 산출 방법 비정규 분포에 대한 공정 능력 지수를 계산하는 데 사용되는 방법을 지정합니다 . 자세한 내용은 " 비정규 분포에 대한 공정 능력 지수 : 백분위수 및 Z 스코어 방법 " 에서 확인하십시오 .

Johnson 분포 적합 방법 가장 적합한 Johnson 분포를 찾는 데 사용되는 방법을 지정합니다 . 모수를 추정하기 전에 Johnson Su, Sb 및 Sl 계열 중에서 가장 적합한 분포 계열이 결정됩니다 . Slifker and Shapiro(1980) 에 설명된 절차를 사용하여 가장 적합한 계열을 찾습니다 .

분위수 매칭 기본 방법입니다 . 이 방법은 최대 가능도보다 더 안정적이고 빠릅니다 . 분위수 매칭 모수 추정값은 가장 적합한 계열이라는 가정하에 분위수 매칭 방식을 사용하여 산출됩니다 . 자세한 내용은 Slifker and Shapiro 연구 자료 (1980) 에서 확인하십시오 .

최대 가능도 최대 가능도를 사용하여 가장 적합한 계열의 모수가 결정됩니다 .

분포 비교 기준 (" 최량 적합 " 분포를 선택한 경우 사용 가능) 최량 적합 분포를 결정하는 데 사용할 기준을 지정합니다 . 이 기준은 " 비교 상세 정보 " 보고서에 나오는 분포 순서에도 영향을 줍니다 . 자세한 내용은 " 비교 기준별 정렬 " 에서 확인하십시오 .

기타 규격

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

유의 수준 지정 신뢰 한계의 유의 수준을 지정합니다 .

규격 한계 대화상자 표시 규격 한계가 없는 열을 처리하는 방법을 지정합니다 .

참고 : 모든 공정 변수의 규격 한계를 " 규격 한계 " 열 특성으로 지정하거나 한계 데이터 테이블에서 규격 한계를 로드하는 것이 좋습니다 (" 한계 데이터 테이블 " 참조). 그렇지 않으면 시작 창에서 " 확인 " 을 클릭한 후 나타나는 " 규격 한계 " 창에서 대화식으로 한계를 지정할 수 있습니다 (시작 창에서 **아니요 (규격 한계가 없는 열 건너뜀)** 를 선택하지 않은 경우).

규격 한계 입력

LSL(규격 하한), USL(규격 상한) 및 목표값은 품질 공정에 대한 하한, 상한 및 목표값을 정의합니다.

다음과 같은 몇 가지 방법으로 규격 한계를 입력할 수 있습니다 .

- 시작 창에서 열을 선택한 후 " 규격 한계 " 창에 한계를 입력합니다 . 자세한 내용은 " 규격 한계 창 " 에서 확인하십시오 .
- 한계 테이블이라고 하는 JMP 데이터 테이블에서 한계를 가져옵니다 . 자세한 내용은 " 한계 데이터 테이블 " 에서 확인하십시오 .
- 데이터 테이블에서 한계를 " 규격 한계 " 열 특성으로 입력합니다 . 자세한 내용은 " 규격 한계 열 특성 " 에서 확인하십시오 .
- JSL 스크립트를 실행하여 공정 능력 보고서를 생성하는 경우 스크립트에 한계를 입력합니다 . 자세한 내용은 " 공정 능력 보고서 " 에서 확인하십시오 .

선택한 열에 대해 하나의 규격 한계만 필요합니다. USL만 지정하면 상자 그림 및 목표 그림 점이 파란색으로 표시됩니다. LSL만 지정하면 상자 그림 및 목표 그림 점이 빨간색으로 표시됩니다.

규격 한계 창

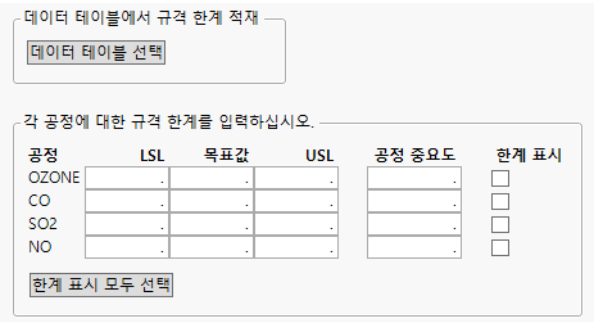
한계가 포함된 열이 없고, 시작 창에서 **아니요(규격 한계가 없는 열 건너뜀)**를 선택하지 않은 경우 "확인"을 클릭하면 "규격 한계" 창이 나타납니다. 시작 창에서 "예"를 선택하는 경우에도 "규격 한계" 창이 나타납니다. 그림7.7에서는 Cities.jmp 샘플 데이터 테이블에 대한 시작 창에서

OZONE, CO, SO2 및 NO를 공정 변수로 선택한 후의 "규격 한계" 창을 보여 줍니다. 알려진 규격 한계를 입력하고 "확인"을 클릭하면 공정 능력 보고서가 표시됩니다.

각 열에 대해 공정 중요도 값을 지정할 수 있습니다. 공정 중요도 값은 원하는 순서대로 공정을 정렬하는 메커니즘을 제공합니다. 공정 능력 보고서의 여러 그래프에서 공정 중요도 값을 사용하여 표식 크기를 지정할 수 있습니다.

공정에 대한 "한계 표시" 옵션을 선택한 후 규격 한계를 열 특성에 저장하면 저장된 "규격 한계" 열 특성에서 "그래프 참조선으로 표시" 옵션이 선택됩니다. 공정에 대한 "한계 표시" 옵션을 선택한 후 규격 한계를 새 테이블에 저장하면 새 테이블의 "한계 표시" 열에 해당 공정에 대한 1이 포함됩니다. **한계 표시 모두 선택** 버튼을 클릭하면 모든 공정에 대해 "한계 표시" 옵션이 선택됩니다.

그림 7.7 Cities.jmp 의 규격 한계 창



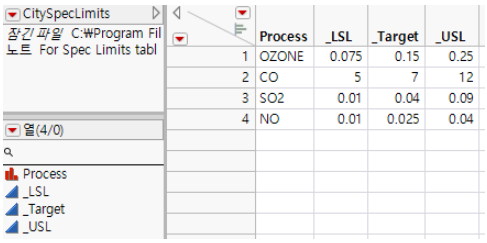
한계 데이터 테이블

"규격 한계" 창에서 **데이터 테이블에서 규격 한계 적재** 옵션을 사용하여 한계 데이터 테이블을 지정할 수도 있습니다. **데이터 테이블 선택** 버튼을 클릭한 후 분석에 대한 규격 한계가 포함된 데이터 테이블을 선택합니다. 적절한 한계 테이블을 선택하면 창에 값이 채워집니다. **확인**을 클릭하면 공정 능력 보고서가 표시됩니다.

한계 데이터 테이블은 **세로형** 또는 **가로형**이라는 두 가지 형식을 사용할 수 있습니다. 세로형 한계 데이터 테이블에는 반응 열 하나가 있고 한계 키워드는 다른 열입니다. 가로형 한계 데이터 테이블에는 한계 키 라벨을 지정하는 하나의 열과 각 반응에 대한 열이 있습니다. **데이터 테이블에서 규격 한계 적재** 옵션을 사용하여 이러한 형식 중 하나를 읽을 수 있습니다.

- 세로형 테이블에는 네 개 또는 다섯 개의 열이 포함되며 각 공정에 대한 행이 하나씩 있습니다. 첫 번째 열은 문자 데이터 유형을 사용하며 공정 능력 플랫폼에서 분석된 열의 이름을 포함합니다. 다음 세 열은 LSL, 목표값 및 USL 로 이름을 지정해야 합니다. 이러한 열 이름 앞에 밑줄 문자를 사용할 수도 있습니다. 마지막 한계 표시 열 (선택적) 은 선택한 분석 그림에 규격 한계를 참조선으로 표시할지 여부를 지정합니다.

그림 7.8 세로형 규격 한계 테이블의 예


CitySpecLimits
잠긴 파일 C:₩Program Fil
노트 For Spec Limits tabl
열(4/0)
Process
_LSL
_Target
_USL

	Process	_LSL	_Target	_USL
1	OZONE	0.075	0.15	0.25
2	CO	5	7	12
3	SO2	0.01	0.04	0.09
4	NO	0.01	0.025	0.04

- 가로형 테이블에는 공정 능력 플랫폼에서 분석된 각 열에 대한 하나의 열과 세 개의 행 및 _LimitsKey 열이 포함됩니다. _LimitsKey 열의 세 개 행은 각각 _LSL, _목표값 및 _USL 식별자를 포함해야 합니다.

그림 7.9 가로형 규격 한계 테이블의 예

	_LimitsKey	OZONE	CO	SO2	NO
1	_LSL	0	5	0	0
2	_Target	0.05	10	0.03	0.025
3	_USL	0.1	20	0.08	0.6

한계 데이터 테이블을 생성하는 가장 쉬운 방법은 공정 능력 플랫폼에서 계산된 결과를 저장하는 것입니다. "공정 능력"의 빨간색 삼각형 메뉴에 있는 "저장" 옵션을 사용하면 표본 값에서 한계를 저장할 수 있습니다. 규격 한계를 입력하거나 로드한 후 다음을 수행할 수 있습니다.

- **저장 > 규격 한계를 열 특성으로 저장**을 선택하여 한계를 데이터 테이블의 열에 "규격 한계" 열 특성으로 저장합니다.
- **저장 > 분포를 열 특성으로 저장**을 선택하여 공정 능력 계산에 사용된 분포를 데이터 테이블의 열에 "공정 능력 분포" 열 특성으로 저장합니다.
- **저장 > 새 테이블에 규격 한계 저장**을 선택하여 한계를 세로형의 새 규격 한계 데이터 테이블에 저장합니다. 비정규 분포를 하나 이상 선택한 경우에는 지정된 분포를 포함하는 **분포** 열도 한계 데이터 테이블에 추가됩니다.

자세한 내용은 "공정 능력 플랫폼 옵션"에서 확인하십시오.

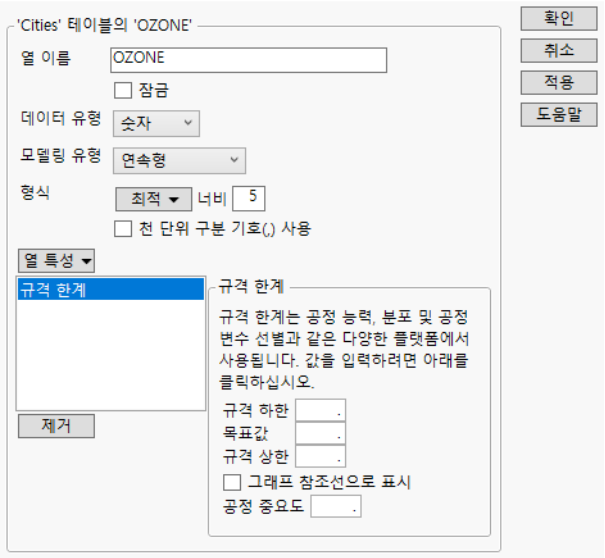
규격 한계 열 특성

공정 능력 분석을 수행하는 경우 열 특성 > 규격 한계를 사용하여 규격 한계를 열 특성으로 저장할 수 있습니다. "규격 한계" 특성은 숫자 열에만 적용됩니다.

일부 공정에는 단측 규격이 설정됩니다. 목표값이 없는 공정도 있습니다. 규격 하한, 규격 상한, 목표값 또는 공정 중요도 값 중 적용되는 항목을 입력할 수 있습니다.

그림 7.10에서는 Cities.jmp 샘플 데이터 테이블에 있는 OZONE 열에 대한 "열 특성" 창의 "규격 한계" 섹션을 표시합니다.

그림 7.10 열 특성 창의 규격 한계 섹션



팁 : 규격 한계를 열 특성으로 저장하면 분석을 반복할 때 일관성이 보장됩니다 .

공정 능력 보고서

기본적으로 공정 능력 플랫폼은 다음 보고서를 제공합니다 .

- " 목표 그림 "(하나 이상의 변수가 정규 분포로 적합된 경우에만 제공되며 정규 분포로 적합된 변수에 대한 점만 표시)
- " 공정 능력 상자 그림 "(하나 이상의 변수가 정규 분포로 적합된 경우에만 제공되며 정규 분포로 적합된 변수에 대한 상자 그림만 표시)
- " 공정 능력 지수 그림 "

그림 7.2 에서는 기본 공정 능력 보고서의 예를 보여 줍니다 .

"공정 능력"의 빨간색 삼각형 메뉴를 사용하여 개별 상세 정보 보고서, 정규화된 상자 그림 및 요약 보고서를 추가할 수 있습니다 . 빨간색 삼각형 메뉴에는 데이터 테이블에서 규격 이탈 값 식별, 요약 데이터 테이블 생성, 분석된 열의 표시 순서 변경 및 규격 한계 저장을 수행할 수 있는 옵션도 있습니다 . 이러한 옵션에 대한 자세한 내용은 "공정 능력 플랫폼 옵션"에서 확인하십시오 .

파일 > 환경 설정 > 플랫폼 > 공정 능력에서 기본 보고서를 변경할 수 있습니다 . 파일 > 환경 설정 > 플랫폼에서 관련 공정 능력 항목을 선택하여 해당 옵션으로 생성된 보고서의 모양을 변경할 수도 있습니다 .

목표 그림

"목표 그림"에서는 각 변수에 대해 규격으로 정규화된 평균 변화를 *x* 축에 표시하고 규격으로 정규화된 표준편차를 *y* 축에 표시합니다. 이 그림을 사용하면 변수가 규격 한계에 적합한 정도를 빠르게 요약하여 볼 수 있습니다. 기본적으로 "목표 그림"에는 각 열에 대해 전체 표준편차를 사용하여 계산된 점만 표시됩니다. 각 점을 커서로 가리키면 해당 점을 계산하는 데 사용된 표준편차 방법 및 변수 이름을 볼 수 있습니다. 목표 그림의 좌표 계산에 대한 자세한 내용은 "목표 그림"에서 확인하십시오.

참고 : 정규 분포를 제외한 다른 분포를 사용하는 공정 변수는 목표 그림에 표시되지 않습니다 .

목표 그림 점

목표 그림의 점은 행이 아니라 열에 해당합니다. 목표 그림에서 점을 선택하면 데이터 테이블에서 해당 열이 선택됩니다. 공정 중요도 값을 지정하면 중요도에 따라 목표 그림의 점 크기가 조정됩니다.

목표 그림의 점을 커서로 가리키면 해당 공정에 대한 관리도를 볼 수 있습니다. 관리도를 클릭하면 해당 관리도와 공정 능력 보고서가 포함된 관리도 빌더가 시작됩니다.

참고: 비편향 합동 표준편차를 공정의 그룹 내 변동 통계량으로 선택한 경우에는 해당 공정에 대한 관리도를 사용할 수 없습니다 .

목표 그림의 점은 목표 그림 요약 테이블의 행에도 연결되며, 이 테이블의 각 행은 열에 해당합니다. 목표 그림에서 점을 선택하고 마우스 오른쪽 버튼을 클릭한 후 행 상태를 적용할 수 있습니다. 이러한 행 상태는 목표 그림 요약 테이블의 행에 적용됩니다. 목표 그림 요약 테이블에서 적용하는 행 상태는 목표 그림에 반영됩니다. 이 테이블을 표시하려면 "공정 능력"의 빨간색 삼각형 메뉴에서 "목표 그림 요약 테이블 생성"을 선택합니다. 자세한 내용은 "목표 그림 요약 테이블 생성"에서 확인하십시오.

팁 : 목표 그림에서 점을 숨길 경우 목표 그림 요약 테이블에서 해당 행 상태를 변경하여 점을 다시 표시할 수 있습니다 .

목표 그림 삼각형

목표 그림의 아래쪽 가운데에 목표 그림 삼각형이 나타납니다. 그림 오른쪽의 슬라이더를 사용하여 그림에서 목표 삼각형 크기를 조정할 수 있습니다.

기본적으로 Ppk 슬라이더와 그 아래의 값은 Ppk = 1로 설정됩니다. 정규 분포를 사용하는 경우 이 값은 부적합 비율(0.0027)에 가깝습니다. 목표 삼각형은 상자에 표시된 Ppk를 나타냅니다. Ppk 값을 변경하려면 슬라이더를 움직이거나 상자에 숫자를 입력합니다.

JMP에서는 기본적으로 Ppk 값을 기반으로 목표 그림을 제공합니다. 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 이 환경 설정을 변경할 수 있습니다. "AIAG(Ppk) 라벨" 환경 설정을 선택하지 않으면 목표 그림 오른쪽의 슬라이더 라벨을 포함하여 모든 Ppk 라벨이 Cpk 라벨로 변경됩니다.

목표 그림 옵션

" 목표 그림 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

군내 표준편차 점 표시 군내 표준편차 추정값을 사용하여 계산된 점을 표시하거나 숨깁니다 .

군내 또는 군간 - 군내 표준편차 점 표시 (시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) 군내 표준편차 추정값 또는 군간 - 군내 표준편차 추정값 (지정된 경우) 을 사용하여 계산된 점을 표시하거나 숨깁니다 .

전체 표준편차 점 표시 전체 표준편차 추정값을 사용하여 계산된 점을 표시하거나 숨깁니다 .

음영 수준 Ppk 수준 음영을 표시하거나 숨깁니다 (그림 7.11). " 음영 수준 " 을 선택하면 음영 영역이 그림에 나타납니다 . 음영 영역은 p 와 Ppk 사이의 관계에 따라 달라지며 , 이때 p 는 Ppk 아래의 상자에 표시된 값을 나타냅니다 .

- 빨간색 영역의 점은 $\text{Ppk} < p$ 입니다 .
- 노란색 영역의 점은 $p < \text{Ppk} < 2p$ 입니다 .
- 녹색 영역의 점은 $2p < \text{Ppk}$ 입니다 .

군내 표준편차 점에 라벨 지정 군내 표준편차 추정값을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

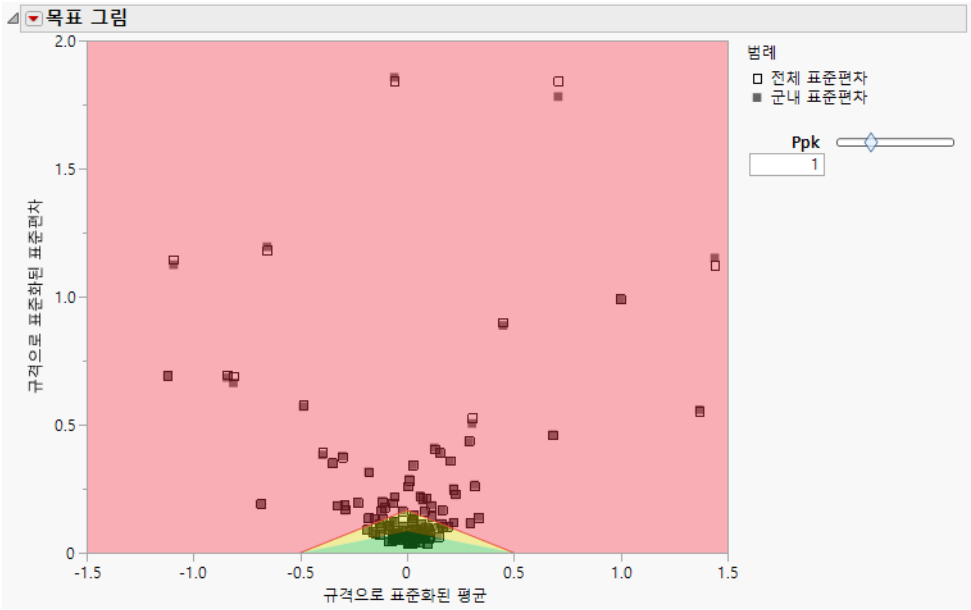
군내 또는 군간 - 군내 표준편차 점에 라벨 지정 (시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) 군내 표준편차 추정값 또는 군간 - 군내 표준편차 추정값 (지정된 경우) 을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

전체 표준편차 점에 라벨 지정 전체 표준편차 추정값을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

결함 비율 등고선 지정된 결함 비율을 나타내는 등고선을 표시하거나 숨깁니다 .

그림 7.11 에서는 Semiconductor Capability.jmp 샘플 데이터 테이블의 전체 데이터 집합에 대해 "목표 그림"의 빨간색 삼각형 메뉴에서 "음영 수준" 및 "군내 표준편차 점 표시"를 선택한 후 표시되는 목표 그림을 보여 줍니다.

그림 7.11 목표 그림



단측 또는 결측 규격 한계

열의 규격 한계가 하나뿐인 경우 표식과 색상은 다음과 같은 방식으로 사용됩니다 .

- USL(규격 상한) 만 지정된 경우 목표 그림의 점은 오른쪽을 가리키는 삼각형으로 나타나고 파란색으로 표시됩니다 .
- LSL(규격 하한) 만 지정된 경우 목표 그림의 점은 왼쪽을 가리키는 삼각형으로 나타나고 빨간색으로 표시됩니다 .
- 하나 이상의 공정에 규격 상한만 있는 경우 목표 삼각형의 오른쪽 절반이 파란색입니다 .
- 하나 이상의 공정에 규격 하한만 있는 경우 목표 삼각형의 왼쪽 절반이 빨간색입니다 .

규격 상한만 있는 공정은 파란색으로 나타나며 목표 삼각형의 파란색(오른쪽)과 비교해야 합니다. 규격 하한만 있는 공정은 빨간색으로 나타나며 목표 삼각형의 빨간색(왼쪽)과 비교해야 합니다. 점 좌표를 계산하는 방법에 대한 자세한 내용은 "목표 그림"에서 확인하십시오.

공정 능력 상자 그림

" 공정 능력 상자 그림 " 에는 분석의 각 변수에 대한 상자 그림이 표시됩니다 . 각 열의 값은 목표값으로 중심화되고 규격 한계 간의 차이로 척도화됩니다 . 목표값이 규격 한계 사이에서 중심화

되지 않으면 목표값과 규격 한계 간 최소 차이의 2 배로 척도화됩니다. 각 공정 열 Y_j 의 경우 다음과 같습니다(표기에 대한 설명은 "목표 그림 및 공정 능력 상자 그림에 대한 표기"에서 확인).

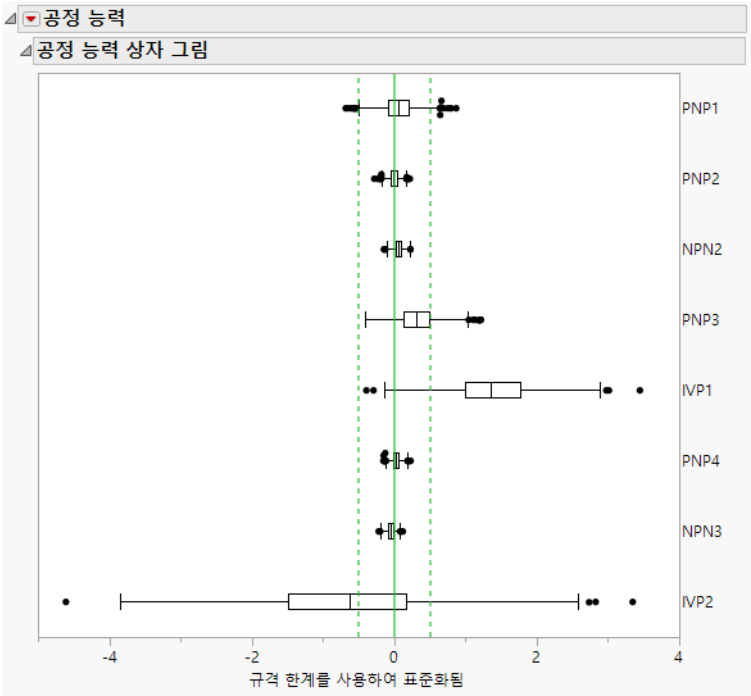
$$Z_{ij} = \frac{Y_{ij} - T_j}{2 \times \min(T_j - \mathrm{LSL}_j, \mathrm{USL}_j - T_j)}$$

단측 규격이 지정된 공정에 대한 내용은 "단측 또는 결측 규격 한계"에서 확인하십시오. 목표값이 지정되지 않은 경우에 대한 내용은 "목표값이 결측인 공정에 대한 공정 능력 상자 그림"에서 확인하십시오.

참고 : 정규 분포를 제외한 다른 분포를 사용하는 공정 변수는 공정 능력 상자 그림에 표시되지 않습니다.

그림 7.11 에서는 Semiconductor Capability.jmp 샘플 데이터 테이블의 8 개 변수에 대한 "공정 능력 상자 그림" 보고서를 보여 줍니다.

그림 7.12 공정 능력 상자 그림



이 그림에는 ±0.5 위치에 녹색 점선이 그려져 있습니다.

- 목표값이 규격 한계 사이에서 중심화된 공정의 경우 녹색 점선은 표준화된 규격 한계를 나타냅니다.

- 목표값이 규격 한계 사이에서 중심화되지 않은 공정의 경우 녹색 점선 중 하나는 목표값에 더 가까운 한계에 대한 표준화된 규격 한계를 나타냅니다 . 다른 녹색 점선은 반대 방향의 동일한 거리를 나타냅니다 .

이 그림은 규격 한계와 관련하여 변수를 비교하는 데 유용합니다 . 예를 들어 그림 7.12 에서 IVP1 의 점 대부분은 USL 을 초과하고 IVP2 의 점 대부분은 목표값 미만입니다 . PNP2 는 모든 데이터 점이 규격 한계 내의 목표값에 있는 것 같습니다 .

단측 또는 결측 규격 한계

열의 규격 한계가 하나뿐인 경우 색상은 다음과 같은 방식으로 사용됩니다 .

- USL(규격 상한) 만 지정된 경우 상자 그림은 파란색으로 표시됩니다 .
- LSL(규격 하한) 만 지정된 경우 상자 그림은 빨간색으로 표시됩니다 .
- 하나 이상의 공정에 규격 상한만 있는 경우 0.5 위치에 파란색 점선이 표시됩니다 .
- 하나 이상의 공정에 규격 하한만 있는 경우 -0.5 위치에 빨간색 점선이 표시됩니다 .

규격 하한만 지정되고 공정 목표값이 지정된 경우를 가정해 보겠습니다 . 공정 능력 상자 그림은 변환된 관측값에 대한 다음 값을 기반으로 합니다 . 표기에 대한 설명은 " 목표 그림 및 공정 능력 상자 그림에 대한 표기 " 에서 확인하십시오 .

$$Z_{ij} = \frac{Y_{ij} - T_j}{2(T_j - \mathrm{LSL}_j)}$$

규격 상한만 지정되고 공정 목표값이 지정된 경우를 가정해 보겠습니다 . 공정 능력 상자 그림은 변환된 관측값에 대한 다음 값을 기반으로 합니다 .

$$Z_{ij} = \frac{Y_{ij} - T_j}{2(\mathrm{USL}_j - T_j)}$$

단측 규격 한계가 있는 경우 결측 목표값을 처리하는 방법에 대한 자세한 내용은 "단일 규격 한계가 지정되고 목표값 없음"에서 확인하십시오 .

공정 능력 지수 그림

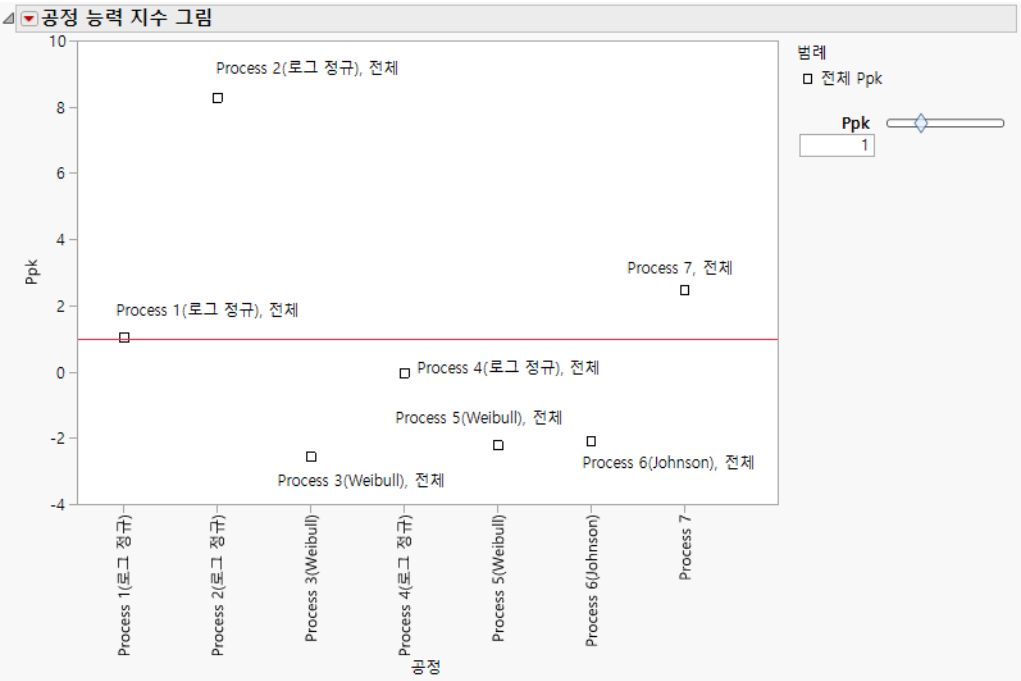
" 공정 능력 지수 그림 " 에는 "Y, 공정 " 으로 입력한 모든 변수에 대한 Ppk 값이 표시됩니다 . 각 변수 이름이 가로 축에 나타나고 Ppk 값이 세로 축에 나타납니다 . 비정규 분포를 적합시키면 적합 분포 이름이 변수 이름의 괄호 안 접미사로 그림에 표시됩니다 . 공정 중요도 값을 지정하면 중요도에 따라 공정 능력 지수 그림의 점 크기가 조정됩니다 . 그림 오른쪽의 슬라이더로 지정된 Ppk 값에 가로선이 추가됩니다 .

공정 능력 지수 그림의 점을 커서로 가리키면 해당 공정에 대한 관리도를 볼 수 있습니다 . 관리도를 클릭하면 해당 관리도와 공정 능력 보고서가 포함된 관리도 빌더가 시작됩니다 .

참고: 비편향 합동 표준편차를 공정의 그룹 내 변동 통계량으로 선택한 경우에는 해당 공정에 대한 관리도를 사용할 수 없습니다.

그림 7.13에서는 Process Measurements.jmp 샘플 데이터 테이블에 대한 "공정 능력 지수 그림" 보고서를 보여 줍니다. 변수 중 6개는 비정규 분포로 적합되었습니다. Process 7은 정규 분포로 적합되었습니다. "공정 능력 지수 그림"의 빨간색 삼각형 메뉴에 있는 "전체 표준편차 점에 라벨 지정" 옵션을 사용하여 점에 라벨이 지정되었습니다.

그림 7.13 비정규 분포가 포함된 공정 능력 지수 그림



공정 능력 지수 그림 옵션

"공정 능력 지수 그림"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

군내 표준편차 점 표시 군내 표준편차 추정값을 사용하여 계산된 점을 표시하거나 숨깁니다.

군내 또는 군간 - 군내 표준편차 점 표시 (시작 창에서 하나 이상의 공정에 대해 "군간 - 군내 공정 능력 계산"을 선택한 경우에만 사용 가능) 군내 표준편차 추정값 또는 군간 - 군내 표준편차 추정값 (지정된 경우)을 사용하여 계산된 점을 표시하거나 숨깁니다.

전체 표준편차 점 표시 전체 표준편차 추정값을 사용하여 계산된 점을 표시하거나 숨깁니다.

음영 수준 Ppk 수준 음영을 표시하거나 숨깁니다. "음영 수준"을 선택하면 음영 영역이 그림에 나타납니다. 음영 영역은 p 와 Ppk 사이의 관계에 따라 달라지며, 이때 p 는 Ppk 아래의 상자에 표시된 값을 나타냅니다.

- 빨간색 영역의 점은 Ppk < p 입니다 .
- 노란색 영역의 점은 p < Ppk < 2p 입니다 .
- 녹색 영역의 점은 2p < Ppk 입니다 .

군내 표준편차 점에 라벨 지정 군내 표준편차 추정값을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

군내 또는 군간 - 군내 표준편차 점에 라벨 지정 (시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) 군내 표준편차 추정값 또는 군간 - 군내 표준편차 추정값 (지정된 경우) 을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

전체 표준편차 점에 라벨 지정 전체 표준편차 추정값을 사용하여 계산된 점의 라벨을 표시하거나 숨깁니다 .

공정 능력 플랫폼 옵션

" 공정 능력 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

개별 상세 정보 보고서 분석의 각 변수에 대한 개별 상세 정보 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 개별 상세 정보 보고서 " 에서 확인하십시오 .

목표 그림 데이터에 대한 목표 그림을 표시하거나 숨깁니다 . " 목표 그림 " 에서는 각 변수에 대해 규격으로 정규화된 평균 변화를 x 축에 표시하고 규격으로 정규화된 표준편차를 y 축에 표시합니다 . 자세한 내용은 " 목표 그림 " 에서 확인하십시오 . 정규 분포를 지정하는 변수만 그림에 표시됩니다 .

공정 능력 상자 그림 분석의 각 변수에 대한 공정 능력 상자 그림을 표시하거나 숨깁니다 . 각 열의 값은 목표값으로 중심화되고 , 목표값과 규격 한계 간 최소 차이의 2 배로 척도화됩니다 . 자세한 내용은 " 공정 능력 상자 그림 " 에서 확인하십시오 . 상자 그림은 정규 분포를 지정하는 변수에 대해서만 표시됩니다 .

정규화된 상자 그림 각 공정 변수의 정규화된 상자 그림을 표시하는 그림에 대한 두 가지 옵션을 제공합니다 . 각 열은 평균을 빼고 열의 표준편차 추정값으로 나누어 표준화됩니다 . 이 상자 그림은 표준화된 값의 분위수를 사용하여 생성됩니다 . 자세한 내용은 " 정규화된 상자 그림 " 에서 확인하십시오 . 정규화된 상자 그림은 정규 분포를 지정하는 변수에 대해서만 표시됩니다 .

군내 표준편차 정규화된 상자 그림 " 군내 표준편차 정규화된 상자 그림 " 을 표시하거나 숨깁니다 . 이 상자 그림은 표준편차의 부분군 내 추정값을 사용하여 생성됩니다 .

군내 또는 군간 - 군내 표준편차 정규화된 상자 그림 (시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) " 군내 또는 군간 - 군내 표준편차 정규화된 상자 그림 " 을 표시하거나 숨깁니다 . 이 상자 그림은 표준편차의 그룹 내 추정값 또는 군간 - 군내 추정값 (지정된 경우) 을 사용하여 생성됩니다 .

전체 표준편차 정규화된 상자 그림 " 전체 표준편차 정규화된 상자 그림 " 을 표시하거나 숨깁니다 . 이 상자 그림은 표준편차의 전체 추정값을 사용하여 생성됩니다 .

공정 능력 지수 그림 "Y, 공정 " 으로 입력한 모든 변수에 대한 전체 Ppk 값을 표시합니다 . 자세한 내용은 " 공정 능력 지수 그림 " 에서 확인하십시오 .

공정 성능 그림 공정 능력과 안정성을 비교하는 4 개의 사분면 그림을 표시하거나 숨깁니다 . 규격 한계가 하나 이상 있는 각 공정이 점으로 표시됩니다 . 자세한 내용은 " 공정 성능 그림 " 에서 확인하십시오 .

요약 보고서 공정 능력 지수 요약 보고서에 대한 두 가지 옵션을 제공합니다 . 자세한 내용은 " 요약 보고서 " 에서 확인하십시오 .

군내 표준편차 요약 보고서 표준편차의 부분군 내 추정값을 사용하여 계산된 공정 능력 지수의 요약 보고서를 표시하거나 숨깁니다 . 정규 분포가 지정된 변수에 대한 결과만 사용할 수 있습니다 .

군내 또는 군간 - 군내 표준편차 요약 보고서 (시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) 표준편차의 그룹 내 추정값 또는 그룹 간 - 그룹 내 추정값 (지정된 경우) 을 사용하여 계산된 공정 능력 지수의 요약 보고서를 표시하거나 숨깁니다 .

전체 표준편차 요약 보고서 표준편차의 전체 추정값을 사용하여 계산된 공정 능력 지수의 요약 보고서를 표시하거나 숨깁니다 .

작업 옵션

다음과 같은 빨간색 삼각형 메뉴 옵션은 작업을 수행합니다 .

규격 이탈 값 데이터 테이블에서 규격 이탈 값을 포함하는 셀에 대한 옵션을 제공합니다 .

규격 이탈 값 선택 데이터 테이블에서 규격 한계를 벗어나는 값이 하나 이상 포함된 모든 행과 열을 선택합니다 .

규격 이탈 값에 색상 적용 데이터 테이블에서 규격 이탈 값에 해당하는 셀에 색상을 적용합니다 . 값이 USL 을 초과하는 셀은 파란색으로 표시되고 값이 LSL 미만인 셀은 빨간색으로 표시됩니다 .

팁 : 특정 셀의 색상을 제거하려면 관심 있는 모든 셀을 선택합니다 . 셀 중 하나를 마우스 오른쪽 버튼으로 클릭하고 " 색상 지우기 " 를 선택합니다 . 모든 셀의 색상을 제거하려면 " 규격 이탈 값에 색상 적용 " 을 선택 취소합니다 .

목표 그림 요약 테이블 생성 목표 그림에 표시된 점에 대한 요약 테이블을 생성합니다 . 이 테이블에는 변수 이름 , 규격으로 표준화된 평균 변화 및 규격으로 표준화된 표준편차가 포함됩니다 . 이 테이블에서 모든 변수는 각각 표준편차 유형 (군내 , 전체) 을 나타내는 두 개의 행을 가집니다 . 자세한 내용은 " 목표 그림 요약 테이블 생성 " 에서 확인하십시오 .

정렬 기준 상자 그림 , 요약 보고서 및 개별 상세 정보 보고서를 재정렬합니다 . 초기 순서 , 초기 순서 역순 , 군내 표준편차 Cpk 오름차순 , 군내 또는 군간 - 군내 표준편차 Cpk 오름차순 , 군내 표준편차 Cpk 내림차순 , 군내 또는 군간 - 군내 표준편차 Cpk 내림차순 , 전체 표준편차 Ppk 오름차순 또는 전체 표준편차 Ppk 내림차순을 기준으로 재정렬할 수 있습니다 . 군내 표

준편차를 기준으로 정렬하는 옵션은 정규 분포가 지정된 변수에 대해서만 그림 요소를 재정렬합니다.

참고 : 군내 또는 군간 - 군내 표준편차를 기준으로 정렬하는 옵션은 시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용할 수 있습니다.

저장 규격 한계 및 분포를 저장하기 위한 옵션을 제공합니다.

규격 한계를 열 특성으로 저장 규격 한계를 분석의 각 변수에 대한 열 특성에 저장합니다. " 규격 한계 " 열 특성이 없으면 열 특성이 새로 생성됩니다. " 규격 한계 " 열 특성이 있으면 열 특성의 값을 덮어씁니다. 자세한 내용은 " 규격 한계 열 특성 " 에서 확인하십시오.

분포를 열 특성으로 저장 공정 능력 계산에 사용된 분포를 " 공정 능력 분포 " 열 특성으로 저장합니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

열에 비정규 분포를 지정하는 " 분포 " 특성이 포함되어 있고 " 공정 능력 분포 " 특성이 없으면 공정 능력 플랫폼에서 비정규 적합을 적용합니다. 공정 능력 플랫폼은 " 분포 " 열 특성에 지정된 분포를 사용하거나, 해당 분포가 공정 능력에서 지원되지 않는 경우 Johnson 적합을 사용합니다. 열에 " 공정 능력 분포 " 특성이 포함된 경우 공정 능력 플랫폼은 " 공정 능력 분포 " 열 특성에 지정된 분포를 사용합니다.

참고 : 공정 능력 플랫폼에서 특정 분포를 사용하려면 해당 분포를 " 공정 능력 분포 " 열 특성으로 저장하십시오.

새 테이블에 규격 한계 저장 각 공정에 대한 규격 한계 및 한계 표시 설정을 세로형 한계 데이터 테이블에 저장합니다. 자세한 내용은 " 한계 데이터 테이블 " 에서 확인하십시오.

대화상자 다시 시작 플랫폼 시작 창을 열고 보고서를 생성하는 데 사용되었던 설정을 재호출합니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

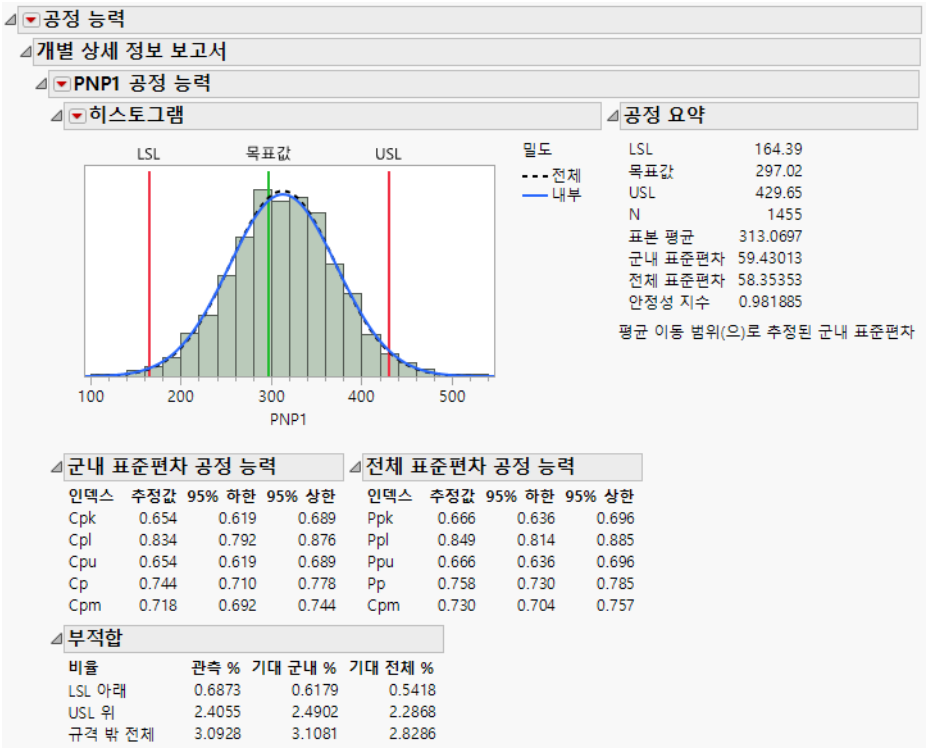
개별 상세 정보 보고서

"개별 상세 정보 보고서" 옵션은 분석의 각 변수에 대한 공정 능력 보고서를 표시합니다.

정규 분포

그림7.14에서는 "정규 변수가 있는 공정 능력 플랫폼의 예"에 설명된 Semiconductor Capability.jmp 샘플 데이터 테이블의 PNP1에 대한 개별 상세 정보 보고서를 보여 줍니다.

그림 7.14 개별 상세 정보 보고서



정규 분포를 사용하는 변수의 개별 상세 정보 보고서에는 히스토그램, 공정 요약 상세 정보, 공정 능력 통계량 및 부적합 통계량이 표시됩니다. 히스토그램에서는 값 분포, 규격 하한, 규격 상한, 공정 목표값(지정된 경우) 및 가정된 분포를 표시하는 하나 또는 두 개의 곡선을 보여 줍니다. 그림7.14의 히스토그램에는 두 개의 정규 곡선이 표시됩니다. 하나는 표준편차의 전체 추정값을 기반으로 하고 다른 하나는 부분군 내 추정값을 기반으로 합니다.

정규 분포를 사용하여 공정을 적합시키는 경우 공정의 안정성 측도인 **안정성 지수**가 공정 요약에 포함됩니다. 안정성 지수는 다음과 같이 정의됩니다.

(전체 표준편차 / 군내 표준편차)

시작 창에서 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 지정한 경우 해당 공정의 안정성 지수는 다음과 같이 계산됩니다 .

(전체 표준편차 / 군간 - 군내 표준편차)

안정적인 공정의 안정성 지수는 1 에 가깝습니다 . 값이 클수록 안정성이 낮음을 나타냅니다 .

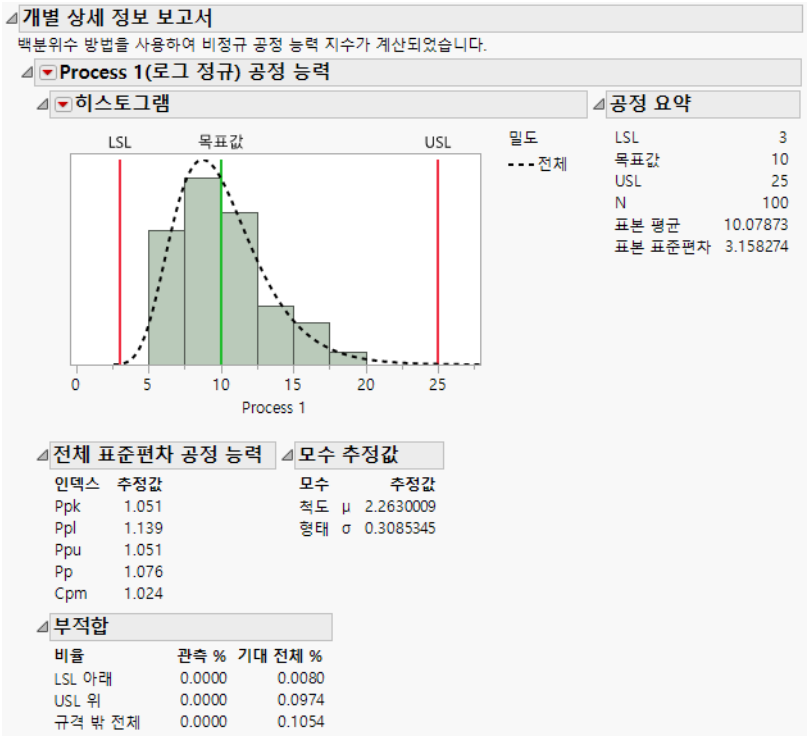
참고 : 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 안정성 평가 유형 환경 설정을 변경할 수 있습니다 . 이렇게 하면 공정 능력 플랫폼을 통해 사용되는 안정성 평가 유형이 변경됩니다 .

비정규 분포

참고 : 비정규 분포를 지정한 공정에는 부분군 내 변동 및 안정성 지수를 기반으로 하는 공정 능력 지수를 사용할 수 없습니다 .

그림7.15에서는 "비정규 변수가 있는 공정 능력 플랫폼의 예"에 설명된 Process Measurements.jmp 샘플 데이터 테이블의 Process 1 에 대한 개별 상세 정보 보고서를 보여 줍니다 .

그림 7.15 Process 1 의 개별 상세 정보 보고서



시작 창에서 선택한 비정규 분포 옵션을 요약하는 노트가 포함된 보고서가 열립니다 .

비정규 분포를 사용하는 변수의 개별 상세 정보 보고서에는 히스토그램, 공정 요약 상세 정보, 공정 능력 통계량 및 부적합 통계량이 표시됩니다. 히스토그램에서는 값 분포, 규격 하한, 규격 상한 및 공정 목표값(지정된 경우)을 보여 줍니다. 적합 분포를 나타내는 곡선이 히스토그램 위에 겹쳐집니다. 비모수 분포를 선택한 경우에는 비모수 밀도 곡선이 히스토그램에 표시됩니다.

이 보고서에는 "모수 추정값" 보고서(비정규 모수 분포를 선택한 경우) 또는 "비모수 밀도" 보고서(비모수 적합을 선택한 경우)도 표시됩니다. 자세한 내용은 "모수 추정값" 및 "비모수 밀도"에서 확인하십시오.

개별 상세 정보 보고서 옵션

"개별 상세 정보 보고서" 섹션에서 각 변수에 대한 개요 제목은 "<변수 이름> 공정 능력" 형식입니다. 그러나 비정규 공정 능력을 요청할 경우에는 관련 분포 이름이 개요 제목의 괄호 안에 표시됩니다.

각 공정 능력 보고서에는 다음 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다.

분포 비교 공정의 분포를 비교하는 제어판을 표시하거나 숨깁니다. 자세한 내용은 "분포 비교"에서 확인하십시오.

공정 요약 전체 표준편차 추정값을 포함하여 해당 변수에 대한 요약 통계량을 표시하거나 숨깁니다. 정규 분포를 지정한 경우에는 군내 표준편차 추정값과 안정성 지수도 추가됩니다. 시작 창에서 하나 이상의 공정에 대해 "군간 - 군내 공정 능력 계산"을 지정한 경우 군간 표준편차 및 군간 - 군내 표준편차에 대한 추정값도 포함됩니다.

히스토그램 변수 값의 히스토그램을 표시하거나 숨깁니다. 히스토그램 보고서에는 다음과 같은 히스토그램 기능을 제어하는 빨간색 삼각형 메뉴가 포함되어 있습니다.

규격 한계 표시 히스토그램에서 공정의 규격 한계 위치에 빨간색 세로선을 표시하거나 숨깁니다.

목표값 표시 히스토그램에서 공정 목표값 위치에 녹색 세로선을 표시하거나 숨깁니다.

군내 표준편차 정규곡선 표시 표본 평균으로 지정된 평균과 군내 표준편차 추정값으로 지정된 표준편차를 사용하여 히스토그램에서 근사 정규 밀도 함수를 표시하거나 숨깁니다.

군간 - 군내 표준편차 정규곡선 표시 (시작 창에서 하나 이상의 공정에 대해 "군간 - 군내 공정 능력 계산"을 선택한 경우에만 사용 가능) 표본 평균으로 지정된 평균과 군간 - 군내 표준편차 추정값으로 지정된 표준편차를 사용하여 히스토그램에서 근사 정규 밀도 함수를 표시하거나 숨깁니다.

전체 표준편차 정규곡선 표시 표본 평균으로 지정된 평균과 전체 표준편차 추정값으로 지정된 표준편차를 사용하여 히스토그램에서 근사 정규 밀도 함수를 표시하거나 숨깁니다.

개수 축 표시 히스토그램 그림의 오른쪽에 관측값 수를 보여 주는 추가 축을 표시하거나 숨깁니다.

밀도 축 표시 히스토그램 그림의 오른쪽에 밀도를 보여 주는 추가 축을 표시하거나 숨깁니다.

공정 능력 지수 다음과 같은 공정 능력 지수 보고서의 표시를 제어합니다.

군내 표준편차 공정 능력 (정규 분포인 경우 사용 가능) 군내 (단기) 표준편차를 기반으로 공정 능력 지수 및 신뢰 구간을 표시하거나 숨깁니다 .

군간 - 군내 표준편차 공정 능력 (정규 분포를 사용하고 시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) 군간 - 군내 표준편차를 기반으로 공정 능력 지수를 표시하거나 숨깁니다 .

군내 표준편차 벤치마크 Z (정규 분포인 경우 사용 가능) 군내 (단기) 표준편차를 기반으로 벤치마크 Z 지수를 표시하거나 숨깁니다 .

군간 - 군내 표준편차 Z 벤치마크 (정규 분포를 사용하고 시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) 군간 - 군내 표준편차를 기반으로 벤치마크 Z 지수를 표시하거나 숨깁니다 .

군내 표준편차 목표 지수 (정규 분포인 경우 사용 가능) 군내 (단기) 표준편차를 기반으로 하는 목표 지수 추정값을 표시하거나 숨깁니다 .

군간 - 군내 표준편차 목표 지수 (정규 분포를 사용하고 시작 창에서 하나 이상의 공정에 대해 " 군간 - 군내 공정 능력 계산 " 을 선택한 경우에만 사용 가능) 군간 - 군내 표준편차를 기반으로 하는 목표 지수 추정값을 표시하거나 숨깁니다 .

전체 표준편차 공정 능력 전체 (장기) 표준편차를 기반으로 공정 능력 지수 및 신뢰 구간을 표시하거나 숨깁니다 .

전체 표준편차 벤치마크 Z (정규 분포인 경우 사용 가능) 전체 (장기) 표준편차를 기반으로 벤치마크 Z 지수를 표시하거나 숨깁니다 .

참고 : 기본적으로 공정 능력 지수의 신뢰 구간은 $\alpha = 0.05$ 를 기반으로 구성됩니다 . 기본 신뢰 수준을 변경하려면 파일 > 환경 설정 > 플랫폼 > 공정 능력을 선택합니다 .

부적합 LSL 미만 , USL 초과 및 규격 한계를 벗어난 관측값의 관측 백분율과 기대 백분율을 표시하거나 숨깁니다 . " 부적합 " 테이블에는 관측 PPM, 기대 PPM 및 개수에 대한 숨겨진 열이 포함되어 있습니다 .

대화식 공정 능력 그림 대화식 공정 능력 그림을 표시하거나 숨깁니다 . 대화식 공정 능력 그림을 사용하면 하나 이상의 요약 통계량 값을 변경하고 이 변경으로 인해 공정 능력 분석에 어떤 영향을 미치는지 확인할 수 있습니다 . 요약 통계량의 원래 값과 새 값 , 공정 능력 지수 및 기대 PPM 을 표시하는 " 원래 값 " 보고서와 " 새 값 " 보고서가 있습니다 . 슬라이더 컨트롤 또는 텍스트 상자를 사용하여 규격 한계 , 평균 및 전체 표준편차를 원래 값에서 변경할 수 있습니다 . " 평균 변화 " 상자를 사용하여 원래 표준편차의 계수만큼 평균을 바꿀 수 있습니다 . " 대화식 공정 능력 그림 " 보고서에는 다음과 같은 빨간색 삼각형 메뉴 옵션이 있습니다 .

공정 능력 " 원래 값 " 보고서와 " 새 값 " 보고서에 공정 능력 지수를 표시하거나 숨깁니다 .

기대 PPM " 원래 값 " 보고서와 " 새 값 " 보고서에 기대 PPM 값을 표시하거나 숨깁니다 .

원래 값으로 되돌리기 대화식 공정 능력 그림 및 " 새 값 " 보고서의 요약 값을 원래 값으로 되돌립니다 .

새 규격 한계를 열 특성으로 저장 새 규격 한계를 원래 데이터 테이블의 열에 " 규격 한계 " 열 특성으로 저장합니다 .

참고 :" 자동 재계산 " 옵션을 설정하지 않으면 분석이 새 규격 한계로 다시 실행되지 않습니다 .

모수 추정값 (" 정규 " 또는 " 비모수 " 를 제외한 다른 분포를 선택한 경우 사용 가능) 선택한 분포의 모수에 대한 추정값을 제공하는 " 모수 추정값 " 보고서를 표시하거나 숨깁니다 .

Johnson 계열 분포를 제외한 모든 분포의 추정값은 최대 가능도를 사용하여 계산됩니다 . Johnson 계열 적합에 대한 자세한 내용은 "Johnson 분포 적합 방법 " 에서 확인하십시오 .

정규 , 베타 , 지수 , 감마 , Johnson, 로그 정규 및 Weibull 분포의 모수와 확률 밀도 함수에 대한 자세한 내용은 " 비정규 분포에 대한 공정 능력 지수 : 백분위수 및 Z 스코어 방법 " 에서 확인하십시오 . 이러한 파라미터화는 공정 능력에서 임계 모수를 지원하지 않는다는 점을 제외하고 분포 플랫폼에 사용되는 것과 동일합니다 . 자세한 내용은 **기본 분석**의 에서 확인하십시오 .

모수 고정 (" 정규 " 또는 " 비모수 " 를 제외한 다른 분포를 선택한 경우 사용 가능) 비정규 분포에서 하나 이상의 모수 값을 고정할 수 있는 창을 표시합니다 . 고정할 모수의 " 사용자 정의 값 " 열에 값을 입력합니다 . 고정 모수 값이 지정된 경우 " 확인 " 을 클릭하면 생략된 모수 값이 다시 추정됩니다 . 다시 추정된 모수 값은 고정된 모수를 나타내는 열과 함께 " 모수 추정값 " 보고서에 나타납니다 .

비모수 밀도 (" 비모수 " 분포를 선택한 경우 사용 가능) 비모수 분포를 적합시키는 데 사용된 **커널 대역폭**을 제공하는 " 비모수 밀도 " 보고서를 표시하거나 숨깁니다 . 커널 대역폭은 다음 식을 사용하여 계산됩니다 . 여기서 n 은 관측값 수이고 S 는 비수정 표본 표준편차입니다 .

$$\text{대역폭} = \frac{0.9S}{n^{1/5}}$$

분포 비교

" 분포 비교 " 보고서를 사용하면 다양한 분포 적합을 비교하고 적용할 수 있습니다 . 다음 사항에 유의하십시오 .

- 선택한 분포가 " 선택 항목 " 열에 표시되어 있습니다 .
- 처음에는 이 보고서의 " 비교 상세 정보 " 보고서에 선택한 분포의 적합 통계량 및 기타 적합 분포가 표시됩니다 ." 최량 적합 " 을 선택한 경우에는 " 비교 상세 정보 " 보고서에 모든 모수 적합에 대한 통계량이 처음에 표시됩니다 .
- " 분포 " 목록에서 비교할 분포를 선택합니다 .
  - 선택한 각 계열의 가장 적합한 분포에 대한 확률 밀도 함수가 " 히스토그램 - 분포 비교 " 보고서의 히스토그램 위에 겹쳐집니다 .
  - 모수 분포의 경우 해당 계열에 대한 적합 결과를 포함하는 행이 " 비교 상세 정보 " 보고서에 추가됩니다 .

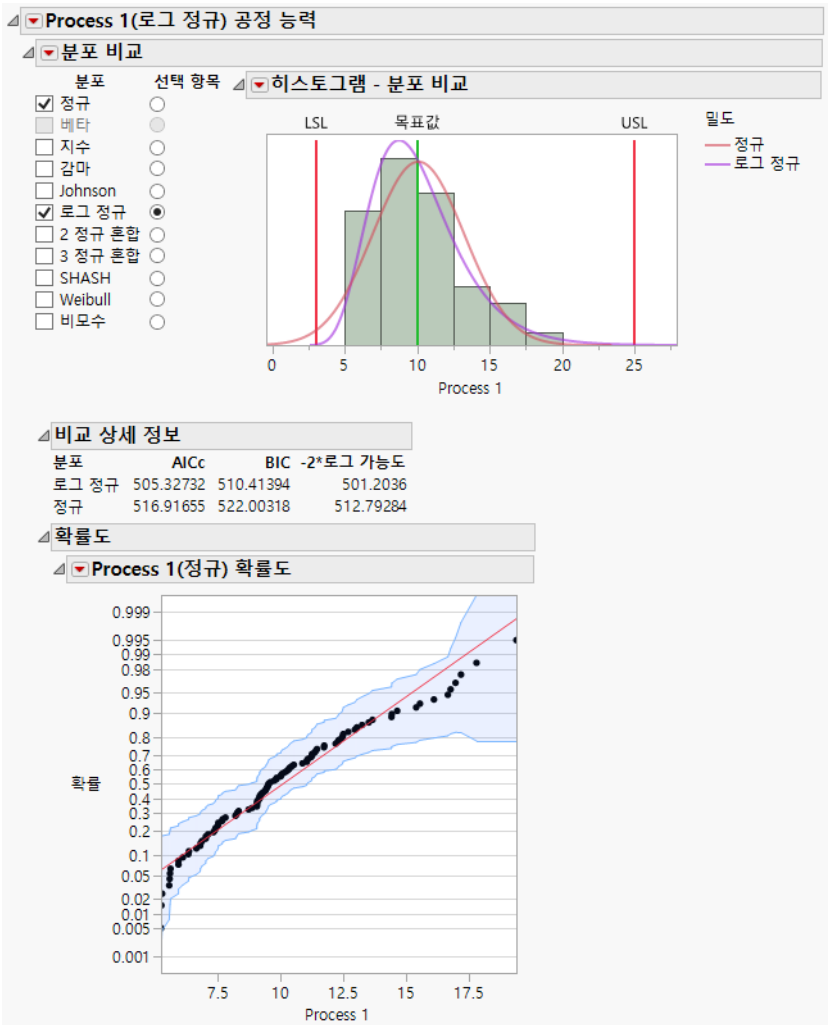
– "분포" 목록에서 "비모수" 를 선택한 경우 자동 선택된 커널 대역폭을 보여 주는 비모수 밀도 보고서가 "분포 비교" 보고서에 추가됩니다. 자세한 내용은 "비모수 밀도" 에서 확인하십시오.

– "선택 항목" 아래의 라디오 버튼을 선택하여 선택 분포를 변경할 수 있습니다. 그러면 공정 능력 보고서가 업데이트되어 선택한 분포에 대한 결과가 표시됩니다.

그림 7.16에서는 Process Measurements.jmp 샘플 데이터 테이블의 Process 1에 대한 분포 비교 보고서를 보여 줍니다. 선택한 로그 정규 분포를 정규 분포와 비교합니다. "비교 상세 정보" 보고서에서는 두 분포에 대한 적합 통계량을 보여 줍니다.

확률도를 생성하려면 "분포 비교"의 빨간색 삼각형을 클릭하고 "확률도"를 선택합니다. 그림 7.16에서 정규 분포 확률도의 점은 선형을 따르지 않습니다. 이는 불량 적합을 나타냅니다.

그림 7.16 정규 분포의 확률도를 사용한 분포 비교



분포 비교 옵션

" 분포 비교 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

비교 상세 정보 각 분포의 AICc, BIC 및 -2* 로그 가능도 값을 제공합니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 . 비모수 적합에는 사용할 수 없습니다 .

비교 히스토그램 히스토그램 보고서를 표시하거나 숨깁니다 .

확률도 적합시키는 각 모수 분포의 확률도를 나타내는 보고서를 표시하거나 숨깁니다 (그림 7.16). 관측값의 가로 좌표는 관측된 데이터 값입니다 . 관측값의 세로 좌표는 관측값 순위에 대한 적합 분포의 분위수 값입니다 . 정규 분포의 경우 전체 표준편차 추정값을 사용하여 적합 분포를 결정합니다 .

각 확률도의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

동시 경험적 신뢰 한계 선택한 모수 계열의 데이터를 사용한다는 가정하에 실제 확률 함수를 포함할 동시 신뢰 수준이 95% 인 신뢰 한계를 표시하거나 숨깁니다 . 이러한 한계는 모든 점에서 추정 정밀도가 동일합니다 . 선택한 모수 분포가 데이터를 잘 적합시키는지 여부를 판별할 때 이 값을 사용합니다 . 자세한 내용은 Nair(1984) 및 Meeker and Escobar(1998) 연구 자료에서 확인하십시오 .

동시 경험적 신뢰 한계 음영 동시 경험적 신뢰 한계 사이의 영역에 음영을 표시하거나 숨깁니다 .

모수 선형 적합 적합 분포를 기반으로 관측값에 대한 예측 확률을 나타내는 선을 표시하거나 숨깁니다 .

모수 적합 신뢰 한계 음영 모수 적합 신뢰 구간 사이의 영역에 음영을 표시하거나 숨깁니다 . 모수 적합 신뢰 한계의 신뢰 수준은 $(1-\alpha)$ 입니다 . 여기서 α 는 시작 창에서 지정한 값입니다 . 모수 적합 신뢰 한계가 의미가 있고 한계를 계산할 수 있는 경우에만 사용할 수 있습니다 .

가능한 경우 모수 분포 F 를 위치 - 척도 계열로 표현하여 $F(y) = G(z)$ 가 되도록 구간을 계산합니다 . 여기서 $z = (y - \mu)/\sigma$ 입니다 . 한 점에서 적합 위치 - 척도 성분의 근사 표준 오차는 델타 방법을 사용하여 계산됩니다 . 표준 오차 추정값을 사용하면 각 점에 대해 z 의 Wald 신뢰 구간이 계산됩니다 . 누적 분포 함수 F 의 신뢰 구간은 G 를 사용한 Wald 구간 변환을 통해 산출됩니다 . 경우에 따라 공정 측정 구간의 끝점 근처에 적절한 구간을 제공하기 위해 특별한 조정 작업이 필요할 수도 있습니다 .

비교 기준별 정렬 선택한 기준에 따라 " 비교 상세 정보 " 보고서의 분포를 정렬합니다 . 시작 창의 " 분포 비교 기준 " 패널에서 다른 기준을 선택한 경우 외에 기본 순서는 AICc 를 기준으로 합니다 .

정규화된 상자 그림

"정규화된 상자 그림" 옵션은 제목에 지정된 표준편차를 사용하여 정규화된 상자 그림을 표시하거나 숨깁니다. JMP에서는 정규화된 상자 그림을 그릴 때 평균을 빼고 표준편차로 나누어 각 열을 표준화합니다. 이러한 표준화된 값을 사용하여 각 열에 대한 상자 그림이 생성됩니다.

그림 7.17 군내 표준편차 정규화된 상자 그림

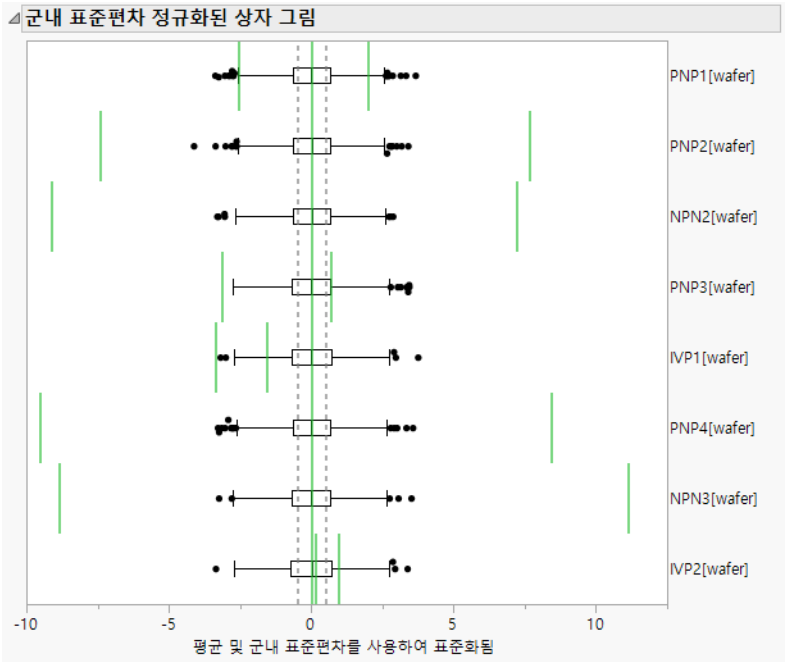


그림 7.17 에서는 Semiconductor Capability.jmp 샘플 데이터 테이블에서 wafer 를 부분군 변수로 사용하는 공정 변수 선택에 대한 "군내 표준편차 정규화된 상자 그림" 을 보여 줍니다.

녹색 세로선은 각 변수의 평균 및 표준편차로 정규화된 각 변수의 규격 한계를 나타냅니다. 데이터가 표준편차 1 로 표준화되므로 ±0.5 위치에 회색 세로 점선이 표시됩니다.

공정 성능 그림

" 공정 성능 그림 " 옵션은 공정 능력과 안정성을 비교하는 4 개의 사분면 그림을 표시하거나 숨깁니다 . 규격 한계가 있는 각 공정이 점으로 표시됩니다 . 공정 중요도 값을 지정하면 중요도에 따라 점 크기가 조정됩니다. 각 점의 가로 좌표는 공정의 안정성 지수이고 세로 좌표는 공정의 전체 Ppk 공정 능력입니다 . 그림은 다음과 같은 기본 경계에 따라 음영 처리된 4 개의 사분면으로 구분됩니다 .

- 1.25 를 초과하는 안정성 지수는 공정이 불안정하다는 것을 나타냅니다 .
- 1.0 보다 작은 Ppk 는 공정 능력이 없음을 나타냅니다 .

또한 그래프의 빨간색 선은 Cpk 값이 1 인 위치를 나타냅니다. 4 개의 사분면을 정의하는 경계는 그림 오른쪽에 있는 Ppk 및 안정성 지수 슬라이더 컨트롤을 사용하여 조정할 수 있습니다. 파일 > 환경 설정 > 플랫폼 > 공정 성능 그림에서 원하는 공정 능력 경계 및 안정성 경계 환경 설정을 지정하고 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 안정성 평가 유형을 지정할 수도 있습니다.

범례에는 음영 영역에 대한 설명이 포함되어 있습니다. 규격 하한 또는 규격 상한이 없는 공정이 있으면 해당 공정에 사용되는 표식도 범례에 표시됩니다. 그림의 모든 공정에 규격 하한 및 상한이 포함되어 있으면 범례에 표식이 나타나지 않습니다. 자세한 내용은 "단측 또는 결측 규격 한계"에서 확인하십시오.

공정 성능 그림의 점을 커서로 가리키면 해당 공정에 대한 관리도를 볼 수 있습니다. 관리도를 클릭하면 해당 관리도와 공정 능력 보고서가 포함된 관리도 빌더가 시작됩니다.

참고: 비편향 합동 표준편차를 공정의 그룹 내 변동 통계량으로 선택한 경우에는 해당 공정에 대한 관리도를 사용할 수 없습니다.

"공정 성능 그림"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

라벨 점 공정 성능 그림의 각 점에 대한 라벨을 표시하거나 숨깁니다.

군내 Cpk 곡선 표시 공정 성능 그림에서 군내 Cpk 곡선을 표시하거나 숨깁니다.

그림 7.18 공정 성능 그림

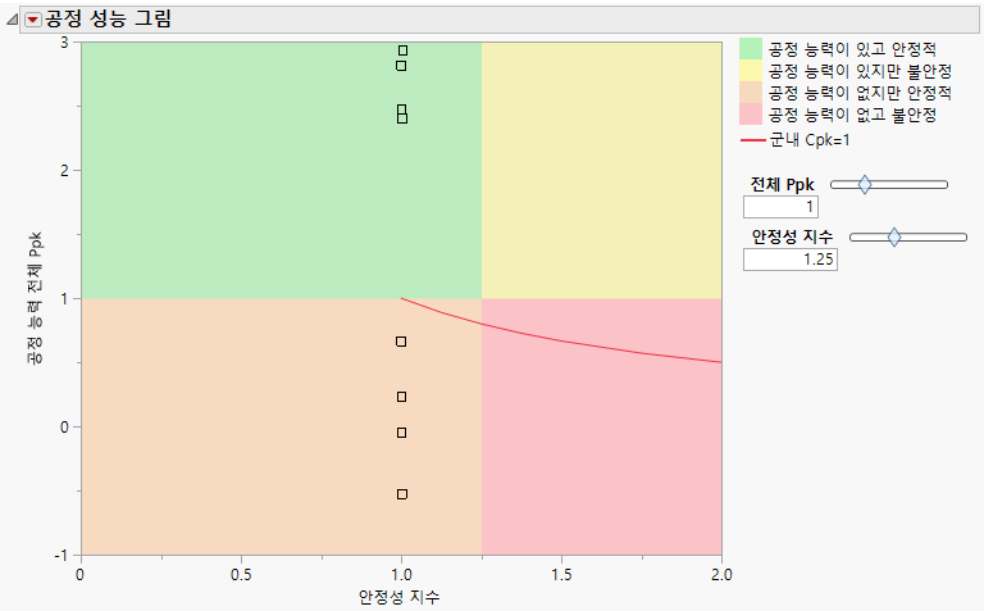


그림 7.18에서는 Semiconductor Capability.jmp 샘플 데이터 테이블에서 wafer를 부분군 변수로 사용하는 공정 변수 선택에 대한 "공정 성능 그림"을 보여 줍니다.

요약 보고서

"요약 보고서" 옵션은 각 변수에 대한 LSL, 목표값, USL, 표본 평균, 다양한 표준편차 추정값, 안정성 지수, Cpk, Cpl, Cpu, Cp, Cpm 및 부적합 통계량이 포함된 테이블을 표시하거나 숨깁니다. 비결측 공정 중요도 값이 하나 이상 있는 경우 "중요도" 열도 요약 보고서에 포함됩니다. 이러한

통계량은 보고서 제목에 지정된 표준편차 추정값을 사용하여 계산됩니다. 안정성 지수, Cpk, Cpl, Cpu 및 Cp 열은 녹색, 노란색, 빨간색으로 표시되어 안정성 또는 공정 능력이 각각 적합, 경계, 매우 나쁨 상태임을 나타냅니다. 이 색상 코딩 체계는 공정 성능 그래프에 표시되는 것과 일치합니다.

참고 : 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 안정성 평가 유형 환경 설정을 변경할 수 있습니다 . 이렇게 하면 공정 능력 플랫폼을 통해 사용되는 안정성 평가 유형이 변경됩니다 .

그림 7.19 에서는 " 정규 변수가 있는 공정 능력 플랫폼의 예 " 에 설명된 두 요약 보고서의 열 중 일부를 보여 줍니다 . 이 보고서에는 다음 열을 선택적으로 사용할 수 있습니다 .

- Cpk, Cpl, Cpu, CP 및 Cpm 에 대한 신뢰 구간
- 기대 PPM 및 관측 PPM 통계량 (규격 밖 , LSL 아래 , USL 위)

  참고 : 기대 PPM 통계량은 선택한 분포를 기반으로 예상되는 백분율입니다 . 기본적으로 정규 분포가 선택됩니다 . 관측 PPM 통계량은 실제 데이터에 기반한 백분율입니다 .

- 표본 표준편차
- 표본 크기 (N), 최소값 , 최대값
- 목표 지수

  참고 : 목표 지수는 군내 표준편차 공정 능력 요약 보고서에서만 사용할 수 있습니다 .

이러한 선택적 열을 표시하려면 보고서를 마우스 오른쪽 버튼으로 클릭하고 "열" 하위 메뉴에서 열 이름을 선택합니다.

전체 표준편차를 기반으로 하는 보고서에는 군내 공정 능력 지수 Cpk, Cpl, Cpu 및 Cp 대신 전체 공정 능력 지수 Ppk, Ppl, Ppu 및 Pp 가 표시됩니다. 전체 공정 능력 지수의 라벨은 "AIAG(Ppk) 라벨" 환경 설정 값에 따라 달라집니다.

그림 7.19 군내 표준편차 / 전체 표준편차 공정 능력 요약 보고서

공정 능력

군내 표준편차 공정 능력 요약 보고서

공정	LSL	목표값	USL	표본 평균	군내 표준편차	안정성 지수	Cpk	Cpl	Cpu	Cp	Cpm
PNP1	164.39	297.02	429.65	313.0697	59.43013	0.981885	0.654	0.834	0.654	0.744	0.718
PNP2	-136.12	465.44	1067.01	456.6157	79.27036	1.007008	2.492	2.492	2.567	2.530	2.514
NPN2	96.59	113.75	130.9	115.7421	2.131652	0.98552	2.370	2.995	2.370	2.683	1.959
PNP3	118.68	130.29	141.9	137.6146	6.160912	0.983744	0.232	1.024	0.232	0.628	0.404
IVP1	59.62	63.41	67.2	73.78072	4.238298	0.990097	-0.518	1.114	-0.518	0.298	0.113
PNP4	-54.43	238.74	531.91	256.3756	33.22573	0.981389	2.764	3.118	2.764	2.941	2.598
NPN3	97.32	120.8	144.29	118.1352	2.362847	1.000809	2.936	2.936	3.690	3.313	2.198
IVP2	139.2	142.31	145.41	138.2432	7.406516	0.989286	-0.043	-0.043	0.323	0.140	0.122

전체 표준편차 공정 능력 요약 보고서

공정	LSL	목표값	USL	표본 평균	전체 표준편차	안정성 지수	Ppk	Ppl	Ppu	Pp	Cpm
PNP1	164.39	297.02	429.65	313.0697	58.35353	0.981885	0.666	0.849	0.666	0.758	0.730
PNP2	-136.12	465.44	1067.01	456.6157	79.82589	1.007008	2.475	2.475	2.549	2.512	2.497
NPN2	96.59	113.75	130.9	115.7421	2.100786	0.98552	2.405	3.039	2.405	2.722	1.975
PNP3	118.68	130.29	141.9	137.6146	6.060762	0.983744	0.236	1.041	0.236	0.639	0.407
IVP1	59.62	63.41	67.2	73.78072	4.196326	0.990097	-0.523	1.125	-0.523	0.301	0.113
PNP4	-54.43	238.74	531.91	256.3756	32.60738	0.981389	2.817	3.177	2.817	2.997	2.636
NPN3	97.32	120.8	144.29	118.1352	2.364757	1.000809	2.934	2.934	3.687	3.310	2.197
IVP2	139.2	142.31	145.41	138.2432	7.327164	0.989286	-0.044	-0.044	0.326	0.141	0.123

목표 그림 요약 테이블 생성

"목표 그림 요약 테이블 생성" 옵션은 각 변수의 이름, 규격으로 정규화된 평균 변화(규격으로 표준화된 평균) 및 규격으로 표준화된 표준편차를 포함하는 요약 데이터 테이블을 생성합니다. 각 변수마다 표준편차 유형을 나타내는 행이 있습니다.

참고: 변수가 정규 분포를 제외한 다른 분포로 적합된 경우 적합 분포 이름이 변수 이름의 괄호 안에 추가됩니다 . 비정규 변수에는 " 규격으로 표준화된 평균 " 및 " 규격으로 표준화된 표준편차 " 값이 제공되지 않습니다 .

목표 그림의 점은 목표 그림 요약 테이블의 행에 연결됩니다. 목표 그림의 점에 행 상태를 적용하면 목표 그림 요약 테이블에서 해당 행 상태를 변경할 수 있습니다. 반대로 목표 그림 요약 테이블에서 행 상태를 적용하면 목표 그림에도 반영됩니다.

그림7.20에서는 "정규 변수가 있는 공정 능력 플랫폼의 예"에 설명된 Semiconductor Capability.jmp 샘플 데이터 테이블에 대한 목표 그림 요약 테이블을 보여 줍니다.

그림 7.20 요약 테이블

	공정	표준편차 유형	규격으로 표준화된 평균	규격으로 표준화된 표준편차
1	PNP1	군내	0.0605056038	0.2240448185
2	PNP2	군내	-0.00733452	0.0658873257
3	NPN2	군내	0.0580798256	0.0621472935
4	PNP3	군내	0.3154424678	0.2653278331
5	IVP1	군내	1.3681693325	0.5591421926
6	PNP4	군내	0.0300774244	0.0566663241
7	NPN3	군내	-0.05674621	0.0503161575
8	IVP2	군내	-0.65593547	1.1945993146
9	PNP1	전체	0.0605056038	0.2199861461
10	PNP2	전체	-0.00733452	0.0663490674
11	NPN2	전체	0.0580798256	0.0612473972
12	PNP3	전체	0.3154424678	0.2610147096
13	IVP1	전체	1.3681693325	0.5536050447
14	PNP4	전체	0.0300774244	0.0556117297

공정 능력 플랫폼의 추가 예

- " 안정적인 공정에 대한 공정 능력 "
- " 불안정한 공정에 대한 공정 능력 "
- " 비정규 공정 Ppk 에 대한 신뢰 한계 시뮬레이션 "

안정적인 공정에 대한 공정 능력

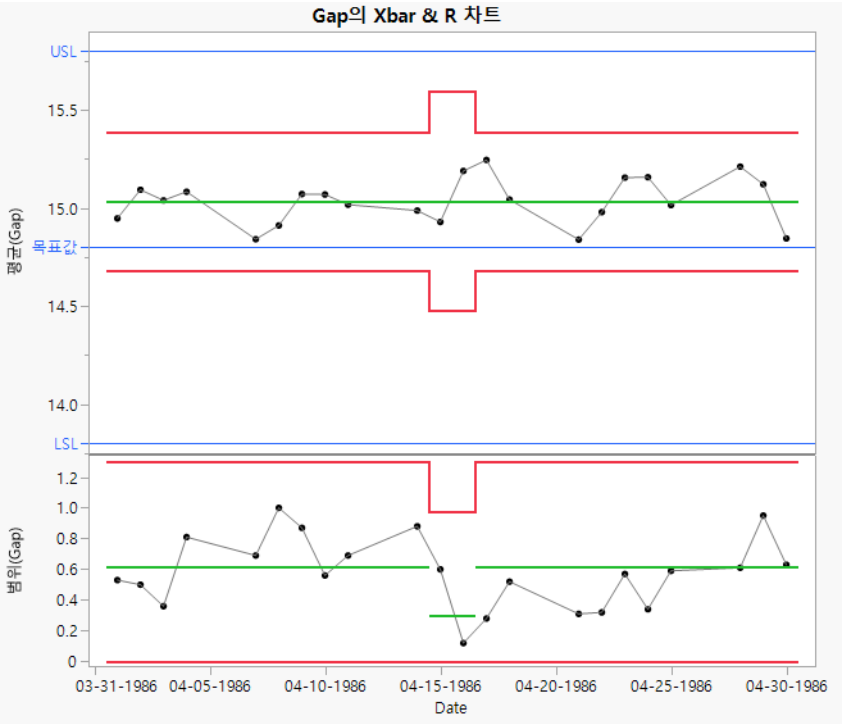
이 예에서는 공정 능력 분석을 기반으로 PPM 불량률을 추정할 수 있는 가정을 확인합니다. 관리도 빌더를 통해 공정 능력에 직접 액세스합니다. 데이터는 크기가 5인 22개의 부분군으로 구성되어 있습니다. 결측 측정값은 총 6개이며, 연속된 두 부분군에 각각 3개씩 있습니다.

관리도 빌더를 통한 공정 능력

관리도 빌더를 사용하여 공정 특성에 대한 정규성 가정 및 공정 안정성을 확인할 수 있습니다. 관리도 빌더 내에서 직접 공정 능력 정보를 얻을 수도 있습니다 .

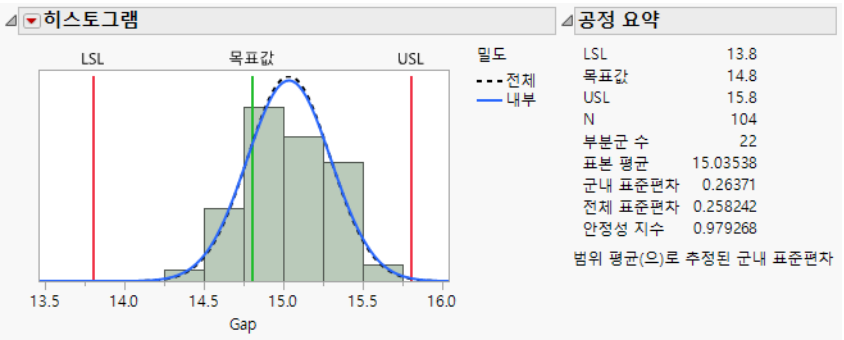
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Clips2.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 빌더**를 선택합니다 .
3. Date 를 **부분군** 영역으로 드래그합니다 .
4. Gap 을 **Y** 영역으로 드래그합니다 .

그림 7.21 Gap 에 대한 Xbar-R 관리도



관리도를 보면 Gap 이 시간이 지남에 따라 안정적임을 알 수 있습니다 . Gap 에 " 규격 한계 " 열 특성이 있으므로 관리도 오른쪽에 " 공정 능력 분석 " 보고서가 나타납니다 .

그림 7.22 Gap 에 대한 공정 능력 분석의 히스토그램



히스토그램과 파란색 적합 정규 곡선은 Gap 의 데이터가 정규 분포에 근사하다는 것을 나타냅니다 . 공정이 안정적이지만 Gap 분포가 규격 범위의 오른쪽으로 이동되었습니다 .

" 공정 요약 " 보고서에서는 " 규격 한계 " 열 특성에 저장된 규격 한계를 보여 줍니다 . 또한 부분군 내 변동에서 계산된 표준편차 추정값 (군내 표준편차) 이 표본 표준편차로 지정된 전체

추정값 (전체 표준편차) 과 크게 다르지 않음을 보여 줍니다 . 결과적으로 안정적 지수는 1 에 가깝습니다 (0.979268). 이것은 공정이 안정적이므로 예상된 결과입니다 .

5. " 부적합 " 보고서의 본문을 마우스 오른쪽 버튼으로 클릭하고 " 열 " 하위 메뉴에서 **기대 군내 PPM** 을 선택합니다 .

그림 7.23 공정 능력 지수 및 부적합 보고서

군내 표준편차 공정 능력

인덱스	추정값	95% 하한	95% 상한
Cpk	0.966	0.805	1.128
Cpl	1.562	1.314	1.808
Cpu	0.966	0.805	1.127
Cp	1.264	1.071	1.457
Cpm	0.943	0.828	1.058

전체 표준편차 공정 능력

인덱스	추정값	95% 하한	95% 상한
Ppk	0.987	0.838	1.136
Ppl	1.595	1.367	1.821
Ppu	0.987	0.837	1.135
Pp	1.291	1.115	1.467
Cpm	0.954	0.841	1.072

부적합

비율	관측 %	기대 군내 %	기대 전체 %	기대 군내 PPM
LSL 아래	0.0000	0.0001	0.0001	1.402263
USL 위	0.0000	0.1869	0.1534	1869.0329
규격 밖 전체	0.0000	0.1870	0.1535	1870.4352

부분군 변동을 사용하여 계산된 Cpk 값은 0.966 이며 , 이는 공정 능력이 좋지 않음을 나타냅니다 . Cpl 값은 양호한 성능을 나타내지만 이는 공정이 규격 하한에서 멀리 이동했기 때문입니다 . 일반적으로 불량품은 Gap 값이 클 때 발생합니다 .

Cpk 에 대한 신뢰 구간 범위는 0.805 ~ 1.128 입니다 . 104 개의 관측값이 있는데도 범위가 넓습니다 . 공정 능력 지수는 비율이기 때문에 변동성이 매우 큽니다 . 따라서 공정 능력 지수의 점 추정값을 기반으로 잘못된 결론을 내리기 쉽습니다 .

" 부적합 " 보고서에 제공된 규격 이탈 제품의 추정값은 공정 성능을 나타내는 직접적인 측도입니다 . " 부적합 " 보고서의 PPM 값은 Gap 이 규격 하한 미만인 경우가 거의 없음을 나타냅니다 (1.4PPM). 그러나 Gap 이 규격 상한을 초과하는 부품의 수는 1869.0PPM 입니다 .

비중심화 공정에 대한 Cp 값은 중심화되도록 공정을 조정한 경우의 잠재적 공정 능력을 나타냅니다 . 이 공정이 목표값 14.8 로 중심화되도록 조정된 경우 공정 능력은 1.264 이며 , 1.071 ~ 1.457 범위의 신뢰 구간을 갖습니다 .

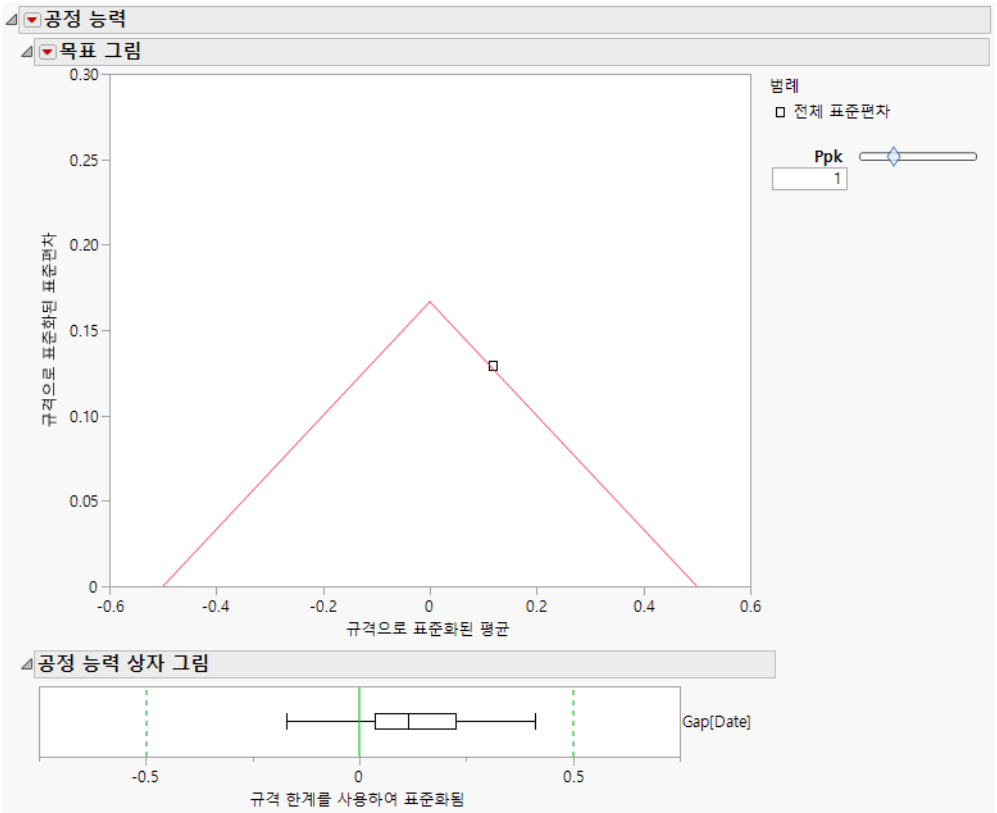
공정 능력 플랫폼

Gap 에 대한 안정성 및 정규성을 확인했으므로 공정 능력 플랫폼에서 추가 정보를 얻을 수 있습니다 .

1. **분석 > 품질 및 공정 > 공정 능력**을 선택합니다 .
2. Gap 을 선택하고 **Y, 공정**을 클릭합니다 .
3. **공정 부분군 지정** 개요를 엽니다 .
4. " 열 선택 " 목록에서 Date 를 선택하고 역할 목록에서 Gap 을 선택합니다 .
5. **부분군 ID 열 내포**를 클릭합니다 .

   기본적으로 " 부분군 내 변동 통계량 " 은 " 비편향 표준편차의 평균 " 으로 선택되어 있습니다 . 관리도 빌더 예 (" 관리도 빌더를 통한 공정 능력 ") 에서는 부분군 범위가 사용되었습니다 .
6. **확인**을 클릭합니다 .

그림 7.24 Gap 에 대한 목표 그림 및 상자 그림



목표 그림에서는 Gap 의 Ppk 지수가 기본적으로 1 임을 보여 줍니다 . 상자 그림에서는 대부분의 값이 규격 내에 있지만 데이터 값의 우세가 규격 범위 내에서 오른쪽으로 이동되었음을 보여 줍니다 .

7. " 공정 능력 " 의 빨간색 삼각형을 클릭하고 " 개별 상세 정보 보고서 " 를 선택합니다 .

   이 보고서는 군내 표준편차가 평균 범위 대신 표준편차 평균에 기반한다는 점을 제외하고 관리도 빌더를 사용하여 생성됩니다 . 자세한 내용은 "Gap 에 대한 공정 능력 분석의 히스토그램 " 및 " 공정 능력 지수 및 부적합 보고서 " 에서 확인하십시오 .

불안정한 공정에 대한 공정 능력

이 예에서는 공정이 불안정하여 전체 변동이 범위 내 변동과 다른 경우를 보여 줍니다. Quality Control 샘플 데이터 폴더의 Coating.jmp 데이터 테이블이 사용됩니다(ASTM Manual on Presentation of Data and Control Chart Analysis 에서 가져옴). 관심 공정 변수는 Sample 열에 의해 부분군으로 그룹화된 Weight 열입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Coating.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 공정 능력**을 선택합니다 .

3. Weight 를 선택하고 **Y, 공정**을 클릭합니다 .
4. **공정 부분군 지정** 개요를 엽니다 .
5. 왼쪽의 **열 선택** 목록에서 Sample 을 선택합니다 .
6. 오른쪽의 **선택한 열 역할 지정** 목록에서 Weight 를 선택합니다 .
7. **부분군 ID 열 내포**를 클릭합니다 .
8. **확인**을 클릭합니다 .
9. **규격 한계** 창에서 **LSL** 에 16, **목표값**에 20, **USL** 에 24 를 각각 입력합니다 .
10. **확인**을 클릭합니다 .
11. " 목표 그림 " 의 빨간색 삼각형을 클릭하고 **군내 표준편차 점 표시**를 선택합니다 .
12. " 공정 능력 " 의 빨간색 삼각형을 클릭하고 **개별 상세 정보 보고서**를 선택합니다 .

그림 7.25 Coating.jmp 데이터에 대한 공정 능력 보고서

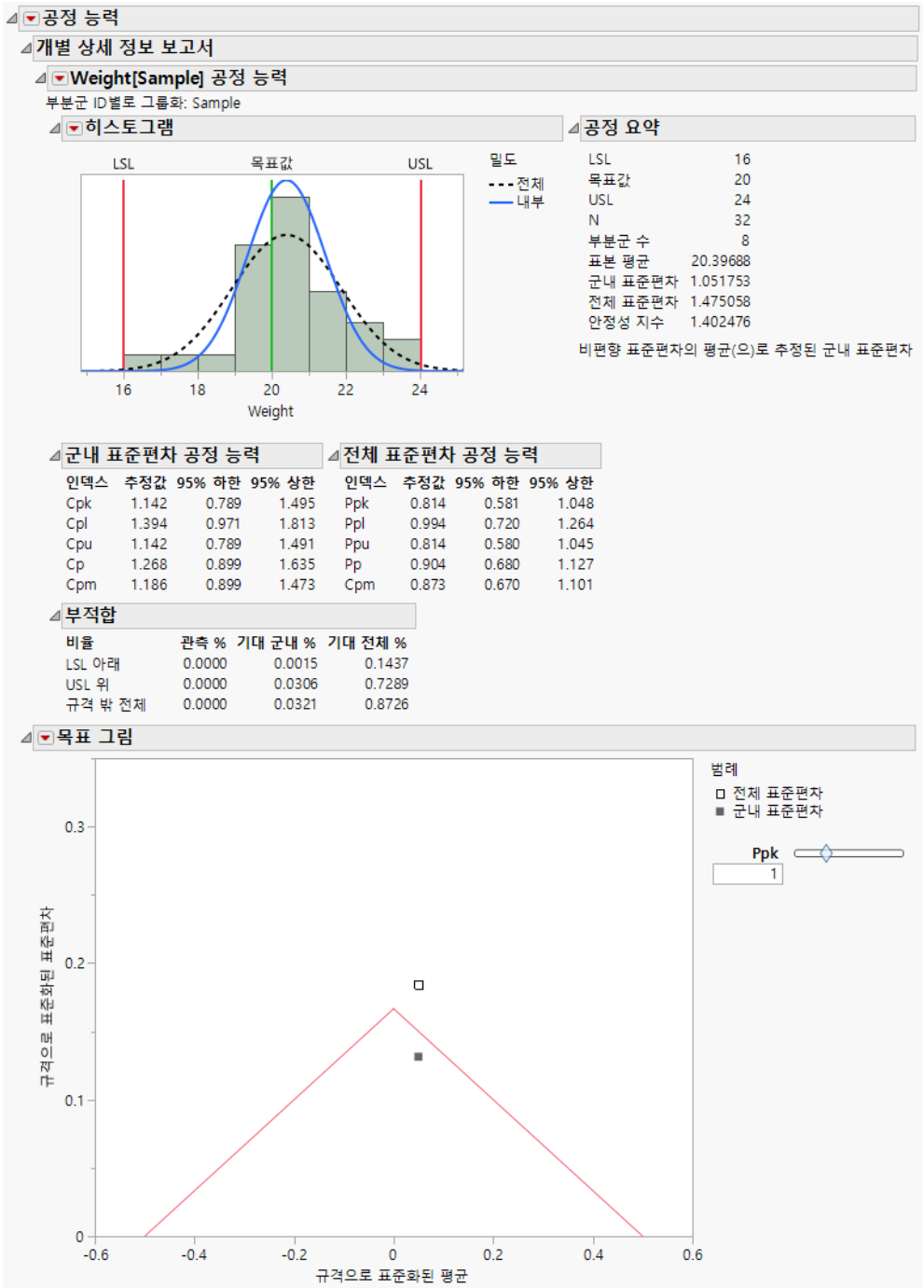


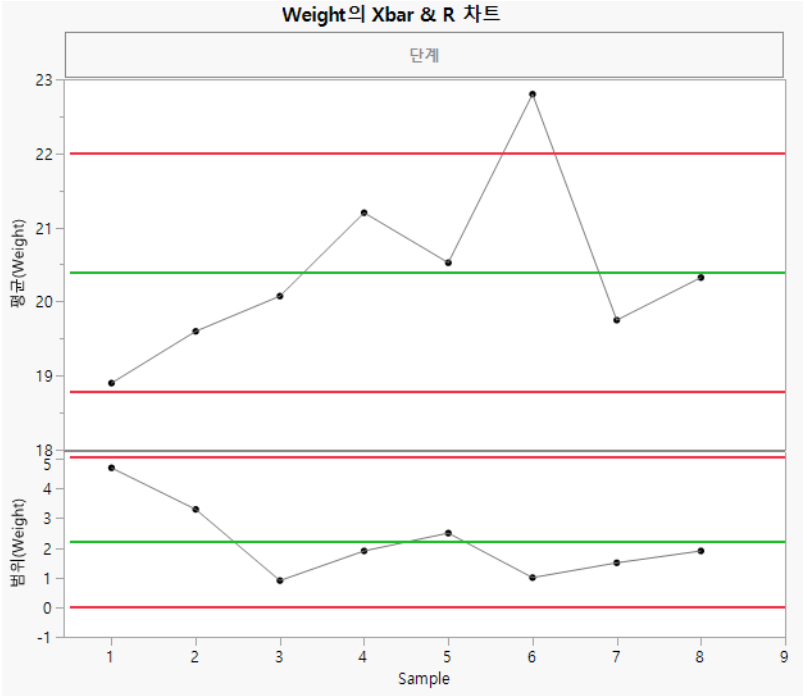
그림 7.25 에서는 결과로 생성된 공정 능력 보고서를 보여 줍니다. 목표 그림에는 규격 한계로 표준화된 평균 변화 및 표준편차를 나타내는 두 개의 점이 표시됩니다. 목표 그림 옆에 이 두 점을 식별하는 범례가 있습니다. 전체 표준편차 점은 전체 표본 표준편차를 사용하여 계산됩니다. 군내 표준편차 점은 표준편차의 부분군 내 추정값을 사용하여 계산됩니다.

전체 표준편차를 사용하여 계산된 점은 Ppk = 1 에 해당하는 목표 삼각형을 벗어납니다. 이는 Weight 변수로 인해 부적합 제품이 생산된다는 의미입니다.

그러나 군내 표준편차를 사용하여 계산된 점은 목표 삼각형 안에 있습니다. 이는 공정이 안정적이라면 Weight 값이 규격 한계 내에 포함될 확률이 높다는 의미입니다.

13. 목표 그림의 점 중 하나를 커서로 가리킵니다.
14. 관리도를 클릭합니다.

그림 7.26 Weight 에 대한 Xbar-R 차트



관리도를 보면 Weight 측정값이 불안정하다는 것을 알 수 있습니다. 공정은 특수 원인의 영향을 받으며 예측할 수 없습니다. 따라서 공정 능력 지수 해석 및 부적합 추정값이 매우 불확실합니다. 공정을 예측할 수 없으므로 전체 표준편차를 기반으로 한 추정값도 불확실합니다.

그림 7.25 의 히스토그램에는 두 표준편차 추정값을 사용하여 히스토그램 위에 겹쳐진 정규 밀도 곡선과 함께 Weight 값의 분포가 표시됩니다. 전체 표준편차 추정값을 사용하는 정규 곡선은 군내 표준편차 추정값을 사용하는 정규 곡선에 비해 더 평평하고 넓습니다. 공정을 불안정하게 만드는 특수 원인에 의해 전체 표준편차 추정값이 부풀려지므로 이 정규 곡선이 더 퍼집니다. 공정이 안정적이라면 더 좁은 정규 곡선에 공정 동작이 반영됩니다.

Cpk 추정값(1.142)을 Ppk 추정값(0.814)과 비교할 수도 있습니다. Ppk가 Cpk보다 훨씬 작으므로 역시 공정을 예측할 수 없다는 결론을 내릴 수 있습니다. Cpk 추정값은 공정을 안정적인 상태로 전환하여 얻을 수 있는 공정 능력에 대한 예측값입니다.

참고 : " 개별 상세 정보 보고서 구분 " 환경 설정은 개별 보고서가 기본적으로 표시되는지 여부를 결정합니다 . 이 환경 설정이 활성화된 경우 공정 변수의 수가 환경 설정에 지정된 수보다 작거나 같으면 개별 보고서가 기본적으로 나타납니다 . 환경 설정 > 플랫폼 > 공정 능력에서 이 환경 설정을 변경할 수 있습니다 .

비정규 공정 Ppk 에 대한 신뢰 한계 시뮬레이션

이 예에서는 먼저 Tablet Measurements.jmp에 있는 세 개의 비정규 변수에 대한 공정 능력 분석을 수행합니다. 그런 다음 시뮬레이션을 사용하여 Purity 변수의 부적합 비율에 대한 신뢰 한계를 찾습니다.

비정규 공정 능력 분석

아래 단계를 수행하지 않으려면 Tablet Measurements.jmp 의 **Process Capability** 테이블 스크립트를 실행하여 이 섹션의 결과를 얻을 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Tablet Measurements.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 공정 능력**을 선택합니다 .
3. Weight, Thickness, Purity 를 선택하고 **Y, 공정**을 클릭합니다 .
4. 오른쪽의 **선택한 열 역할 지정** 목록에서 Weight, Thickness, Purity 를 선택합니다 .
5. **분포 옵션** 개요를 엽니다 .
6. 분포 목록에서 **최량 적합**을 선택합니다 .
7. **공정 분포 설정**을 클릭합니다 .

   오른쪽 목록의 각 열 이름에 **& 분포 (최량 적합)** 접미사가 추가됩니다 .
8. **확인**을 클릭합니다 .

   Ppk 값을 보여 주는 공정 능력 지수 그림이 나타납니다 . Thickness 변수만 Ppk = 1 을 나타내는 선 위에 있습니다 . Purity 는 거의 선 위에 있습니다 . 측정값 수 250 이 많은 것 같지만 추정된 Ppk 값은 여전히 변동성이 큽니다 . 이 때문에 실제 Purity Ppk 값에 대한 신뢰 구간을 생성합니다 .

   참고 : 목표 그림이 표시되지 않았으므로 정규 분포 적합이 세 변수에 대한 최량 적합이 아니라는 결론을 내릴 수 있습니다 .
9. " 공정 능력 " 의 빨간색 삼각형을 클릭하고 **개별 상세 정보 보고서**를 선택합니다 .

   각 공정마다 최량 적합이 다릅니다 .

   – Weight: 로그 정규

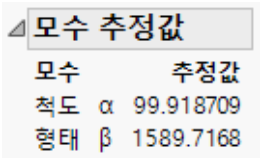
- Thickness: SHASH
- Purity: Weibull

시뮬레이션 열 생성

시뮬레이션 유틸리티를 사용하여 Ppk 신뢰 한계를 추정하려면 적합된 Weibull 분포를 반영하는 시뮬레이션 계산식을 생성해야 합니다 . 아래 단계를 수행하지 않으려면 **Add Simulation Column** 테이블 스크립트를 실행하여 이 섹션의 결과를 얻을 수 있습니다 .

1. "Purity(Weibull) 공정 능력 " 보고서가 나올 때까지 스크롤하여 " 모수 추정값 " 보고서를 찾습니다 .

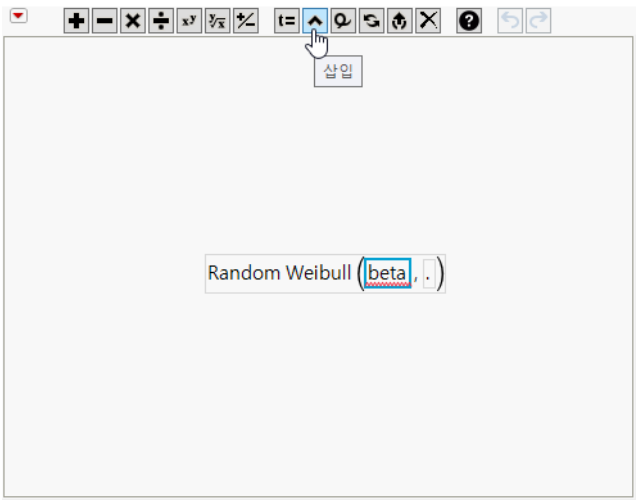
그림 7.27 Purity 에 대한 Weibull 모수 추정값



이 값은 최량 적합 분포인 Weibull 에 대한 모수 추정값입니다 .

2. Tablet Measurements.jmp 샘플 데이터 테이블에서 **열 > 새 열**을 선택합니다 .
3. **열 이름** 옆에 Simulated Purity 를 입력합니다 .
4. **열 특성** 목록에서 **계산식**을 선택합니다 .
5. 계산식 편집기에서 **난수 > Random Weibull** 을 선택합니다 .
6. **beta** 자리 표시자가 선택되어 있습니다 . 삽입 요소 (**^**) 를 클릭합니다 .

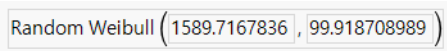
그림 7.28 Simulated Purity 열에 대한 계산식 편집기



이렇게 하면 **alpha** 파라미터의 자리 표시자가 추가됩니다 .

7. " 공정 능력 " 보고서의 "Purity(Weibull) 공정 능력 " 에서 " 모수 추정값 " 보고서 테이블을 마우스 오른쪽 버튼으로 클릭하고 **데이터 테이블로 만들기**를 선택합니다 .
8. **추정값** 열의 2 행에 입력된 값 (1589.7167836) 을 복사합니다 .
9. 계산식 편집기 창에서 **Random Weibull** 계산식의 **beta** 자리 표시자를 선택하고 **beta** 자리 표시자에 1589.7167836 을 붙여 넣습니다 .
10. " 모수 추정값 " 보고서에서 생성한 데이터 테이블에서 **추정값** 열의 1 행에 입력된 값 (99.918708989) 을 복사합니다 .
11. 계산식 편집기 창에서 **Random Weibull** 계산식의 **alpha** 자리 표시자를 선택하고 **alpha** 자리 표시자에 99.918708989 를 붙여 넣습니다 .

그림 7.29 완료된 계산식 창



12. 계산식 편집기 창에서 **확인**을 클릭합니다 .
13. Tablet Measurements.jmp 데이터 테이블에서 Simulated Purity 열을 마우스 오른쪽 버튼으로 클릭하고 **열 특성 > 규격 한계**를 선택합니다 .
14. **규격 하한** 옆에 99.5 를 입력합니다 .
15. " 새 열 " 창에서 **확인**을 클릭합니다 .

Simulated Purity 열에 최량 적합 분포의 값을 시뮬레이션하는 계산식이 포함됩니다 .

Purity Ppk 및 기대 부적합 비율에 대한 신뢰 구간 시뮬레이션

시뮬레이션을 사용하면 지정한 횟수만큼 전체 분석이 실행됩니다. 필요한 분석만 실행하여 계산량을 최소화하면 계산 시간을 줄일 수 있습니다. 이 예에서는 적합된 Weibull 분포를 사용하는 Purity에만 관심이 있으므로 시뮬레이션을 실행하기 전에 이 분석만 수행합니다.

참고 : 계산 시간을 무시하려면 이전 섹션에서 생성한 동일한 보고서를 사용하여 7 단계로 시작할 수 있습니다 .

---

1. " 공정 능력 " 보고서에서 " 공정 능력 " 의 빨간색 삼각형을 클릭하고 **대화상자 다시 시작**을 선택합니다 .
2. (선택 사항) " 공정 능력 " 보고서를 닫습니다 .
3. 시작 창의 **선택한 열 역할 지정** 목록에서 Weight& 분포 (로그 정규) 및 Thickness& 분포 (SHASH) 를 선택합니다 .
4. **제거**를 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 공정 능력 " 의 빨간색 삼각형을 클릭하고 **개별 상세 정보 보고서**를 선택합니다 .

Ppk 값과 Ppl 값이 모두 제공되지만 Purity 에는 규격 하한만 있으므로 두 값이 동일합니다 .

7. " 전체 표준편차 공정 능력 " 보고서에서 **추정값** 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .

   스위치 아웃할 열 목록에서 Purity 가 선택되어 있는지 확인합니다 . **스위치 인할 열** 목록에서 Simulated Purity 가 선택되어 있는지 확인합니다 .

8. **표본 수** 옆에 **500** 을 입력합니다 .

   참고 : 다음 단계를 반드시 수행할 필요는 없습니다 . 그러나 이 예에 표시된 시뮬레이션 값을 정확하게 얻을 수 있습니다 .

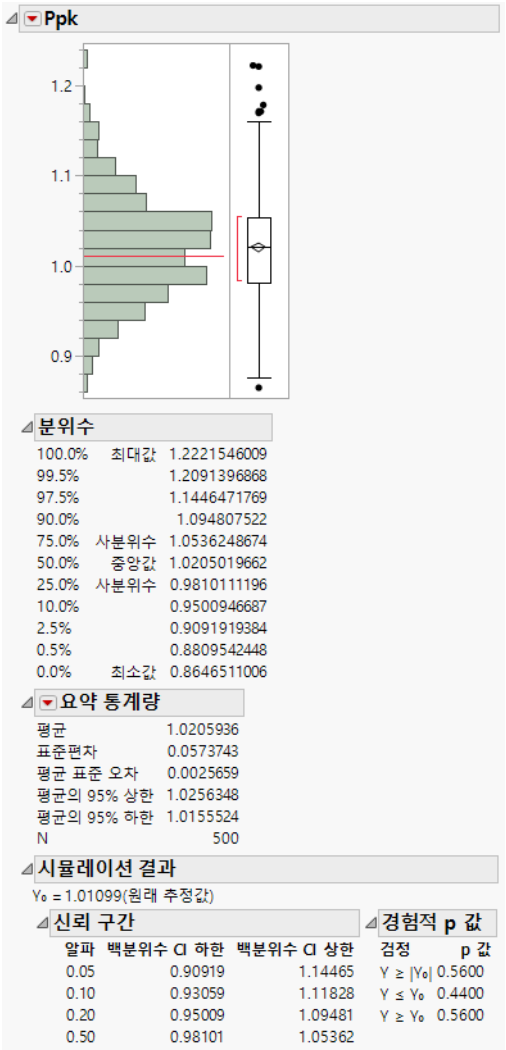
9. (선택 사항) **난수 시드값** 옆에 **12345** 를 입력합니다 .

10. **확인**을 클릭합니다 .

    계산하는 데 몇 초 정도 걸릴 수 있습니다 . " 공정 능력 시뮬레이션 결과 (추정값)" 데이터 테이블이 나타납니다 . 이 테이블의 Ppk 및 Ppl 열은 각각 Simulated Purity 계산식을 기반으로 계산된 500 개의 값을 포함합니다 . 첫 번째 행 (제외됨) 은 원래 분석에서 얻은 Purity 값을 포함합니다 . Purity 에는 규격 하한만 있으므로 Ppk 값이 Ppl 값과 동일합니다 .

11. " 공정 능력 시뮬레이션 결과 (추정값)" 데이터 테이블에서 **분포** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

그림 7.30 Purity 에 대해 시뮬레이션된 Ppk 값 분포



참고 : Simulated Purity 열 계산식에서 파라미터의 소수점 정밀도에 따라 표시되는 값이 약간 다를 수 있습니다 .

Ppk 와 Ppl 에 대해 각각 하나씩 총 두 개의 " 분포 " 보고서가 표시됩니다 . 그러나 Purity 에는 규격 하한만 있으므로 Ppk 와 Ppl 의 값이 동일합니다 . 따라서 " 분포 " 보고서도 동일합니다 .

" 시뮬레이션 결과 " 보고서에서는 Ppk 에 대한 95% 신뢰 구간이 0.909 ~ 1.145 임을 보여 줍니다 . 이는 실제 Ppk 값이 1.0 을 초과할 수 있음을 의미합니다 . 즉 , " 비정규 공정 능력 분석 " 에서 생성한 " 공정 능력 지수 그림 " 에서 Ppk = 1 선 위에 Purity 가 있습니다 .

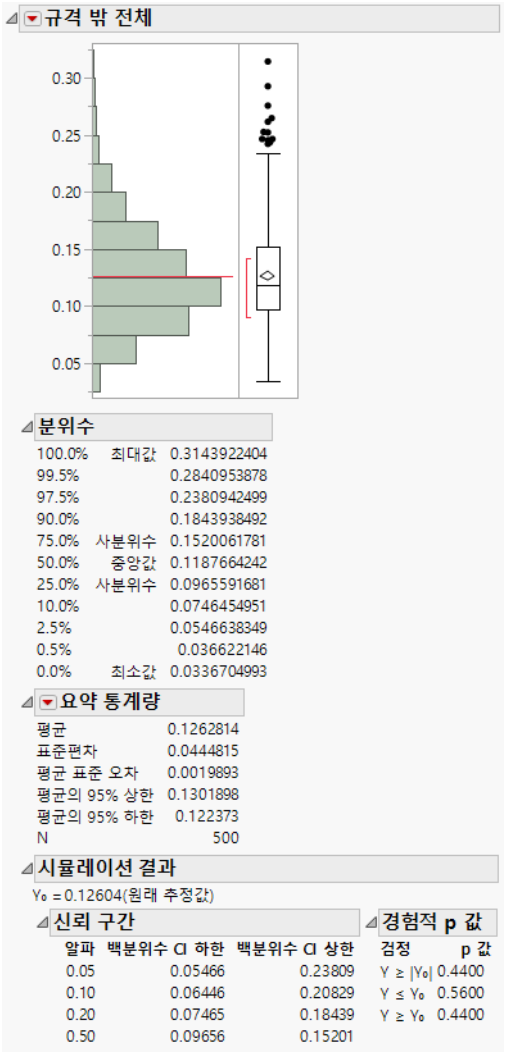
12. " 공정 능력 " 보고서에서 " 부적합 " 보고서의 **기대 전체 %** 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .

스위치 아웃할 열 목록에서 Purity 가 선택되어 있는지 확인합니다 . **스위치 인할 열** 목록에서 Simulated Purity 가 선택되어 있는지 확인합니다 .

13. **표본 수** 옆에 **500** 을 입력합니다 .
14. (선택 사항) **난수 시드값** 옆에 **12345** 를 입력합니다 .
15. **확인**을 클릭합니다 .

    계산하는 데 몇 초 정도 걸릴 수 있습니다 . " 공정 능력 시뮬레이션 결과 (기대 전체 %)" 데이터 테이블이 나타납니다 . Purity 에는 규격 하한만 있으므로 LSL 아래 값이 규격 밖 전체 값과 동일합니다 .
16. " 공정 능력 시뮬레이션 결과 (기대 전체 %)" 데이터 테이블에서 **분포** 스크립트 옆의 녹색 삼각형을 클릭합니다 .

그림 7.31 Purity 에 대해 시뮬레이션된 규격 밖 전체 값 분포



참고 : Simulated Purity 열 계산식에서 파라미터의 소수점 정밀도에 따라 표시되는 값이 약간 다를 수 있습니다 .

동일한 두 개의 " 분포 " 보고서가 나타납니다 . " 시뮬레이션 결과 " 보고서에서는 기대 전체 %(부적합) 에 대한 95% 신뢰 구간이 0.055 ~ 0.238 임을 보여 줍니다 .

공정 능력 플랫폼에 대한 통계 상세 정보

- " 변동 통계량 "
- " 목표 그림 및 공정 능력 상자 그림에 대한 표기 "
- " 목표 그림 "
- " 목표값이 결측인 공정에 대한 공정 능력 상자 그림 "
- " 정규 분포에 대한 공정 능력 지수 "
- " 비정규 분포에 대한 공정 능력 지수 : 백분위수 및 Z 스코어 방법 "
- " 분포의 파라미터화 "

변동 통계량

공정 표준편차를 σ로 나타냅니다. 공정 능력 플랫폼은 두 가지 유형의 공정 능력 지수를 제공합니다. Ppk 지수는 부분군에 의존하지 않는 방식으로 모든 데이터를 사용하는 σ 추정값을 기반으로 합니다. 이 전체 추정값은 우연 원인 변동뿐만 아니라 특수 원인 변동을 반영할 수 있습니다. Cpk 지수는 우연 원인 변동만 포착하는 추정값을 기반으로 합니다. Cpk 지수는 σ에 대한 부분군의 군내 추정값 또는 군간-군내 추정값을 사용하여 생성됩니다. 이러한 방식으로 실제 공정 표준편차를 반영합니다. 공정이 안정적이지 않으면 σ의 여러 추정값이 크게 다를 수 있습니다.

전체 표준편차

전체 표준편차는 부분군에 의존하지 않습니다 . σ 의 전체 추정값은 다음과 같이 계산됩니다 .

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

이 계산식에서는 다음과 같은 표기를 사용합니다 .

N = 전체 데이터 집합의 비결측값 수

y_i = i 번째 관측값

$\bar{y}$ = 전체 데이터 집합의 비결측값 평균

주의 : 공정이 안정적인 경우 전체 표준편차는 공정 표준편차를 추정합니다. 공정이 불안정한 경우에는 공정 표준편차를 알 수 없으므로 σ 의 전체 추정값이 불확실합니다 .

부분군 내 변동에 기반한 표준편차 추정값

부분군 내 변동을 기반으로 하는 σ 추정값은 다음 방법 중 하나로 생성할 수 있습니다 .

- 범위 평균으로 추정된 군내 표준편차

- 비편향 표준편차의 평균으로 추정된 군내 표준편차
- 이동 범위로 추정된 군내 표준편차
- 비편향 합동 표준편차로 추정된 군내 표준편차

시작 창에서 부분군 ID 열 또는 일정한 부분군 크기를 지정하는 경우 원하는 부분군 내 변동 통계량을 지정할 수 있습니다. 자세한 내용은 "공정 능력 플랫폼 시작"에서 확인하십시오. 부분군 ID 열, 일정한 부분군 크기 또는 과거 시그마를 지정하지 않으면 JMP에서 세 번째 방법(크기가 2인 부분군의 이동 범위)을 사용하여 군내 표준편차를 추정합니다.

범위 평균에 기반한 군내 표준편차

범위 평균으로 추정된 군내 표준편차는 XBar-*R* 차트의 표준편차 추정값과 동일합니다.

$$\hat{\sigma} = \frac{\frac{R_1}{d_2(n_1)} + \ldots + \frac{R_N}{d_2(n_N)}}{N}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

$R_i = i$ 번째 부분군의 범위

$n_i = i$ 번째 부분군의 표본 크기

$d_2(n_i)$ = 단위 표준편차를 갖는 n_i 개 독립 정규 분포 변수의 범위 기대값

N = 부분군 수 ($n_i \geq 2$)

비편향 표준편차의 평균에 기반한 군내 표준편차

비편향 표준편차의 평균으로 추정된 군내 표준편차는 XBar-*S* 차트의 표준편차 추정값과 동일합니다.

$$\hat{\sigma} = \frac{\frac{s_1}{c_4(n_1)} + \ldots + \frac{s_N}{c_4(n_N)}}{N}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

$n_i = i$ 번째 부분군의 표본 크기

$c_4(n_i)$ = 단위 표준편차를 갖는 n_i 개 독립 정규 분포 변수의 표준편차 기대값

N = 부분군 수 ($n_i \geq 2$)

$s_i = i$ 번째 부분군의 표본 표준편차

평균 이동 범위에 기반한 군내 표준편차

평균 이동 범위로 추정된 군내 표준편차는 개별 측정값 및 이동 범위 차트의 표준편차 추정값과 동일합니다.

$$\hat{\sigma} = \frac{\overline{MR}}{d_2(2)}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

$\overline{MR}$ = (MR_2+MR_3+...+MR_N)/(N-1) 로 계산된 비결측 이동 범위의 평균 ($MR_i = |y_i - y_{i-1}|$)

$d_2(2)$ = 단위 표준편차를 갖는 두 독립 정규 분포 변수의 범위 기대값

중앙값 이동 범위에 기반한 군내 표준편차

중앙값 이동 범위로 추정된 군내 표준편차는 다음과 같습니다.

$$\hat{\sigma} = \frac{MMR}{0.954}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

MMR = Median(MR_2, MR_3,..., MR_N)으로 계산된 비결측 이동 범위의 중앙값($MR_i = |y_i - y_{i-1}|$)

비편향 합동 표준편차에 기반한 군내 표준편차

비편향 합동 표준편차로 추정된 군내 표준편차는 다음과 같습니다.

$$\hat{\sigma} = \frac{\sqrt{(n_1 - 1)s_1^2 + \ldots + (n_N - 1)s_N^2}}{c_4(n)\sqrt{n_1 + \ldots + n_2 - N}}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

$n_i = i$ 번째 부분군의 표본 크기

$n = n_1 + \ldots + n_N$, 총 표본 크기

$c_4(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 표준편차 기대값

N = 부분군 수 ($n_i \geq 2$)

$s_i = i$ 번째 부분군의 표본 표준편차

그룹 간 변동에 기반한 표준편차 추정값

이동 범위에 기반한 군간 표준편차

부분군 간 변동을 기반으로 하는 σ 추정값은 부분군 평균의 이동 범위로 추정됩니다 .

$$\hat{\sigma} = \sqrt{\left(\frac{\overline{MR}}{d_2(2)}\right)^2 - \frac{\hat{\sigma}^2_{\text{Within}}}{H}}$$

이 계산식에서는 다음과 같은 표기를 사용합니다 .

$\overline{MR}$ = $(MR_2+MR_3+...+MR_N)/(N-1)$ 로 계산된 비결측 이동 범위의 평균 ($MR_i = |y_i - y_{i-1}|$)

$d_2(2)$ = 단위 표준편차를 갖는 두 독립 정규 분포 변수의 범위 기대값

σ^2_{within} = 지정된 군내 표준편차 추정값

$H = \dfrac{N}{\frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_N}}$, 부분군 표본 크기의 조화 평균

그룹의 군간 - 군내 변동에 기반한 표준편차 추정값

군간 - 군내 표준편차

그룹의 결합된 군간 - 군내 변동을 기반으로 하는 표준편차 추정값은 다음과 같이 정의됩니다 .

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2_{\text{within}} + \hat{\sigma}^2_{\text{between}}}$$

목표 그림 및 공정 능력 상자 그림에 대한 표기

다음은 목표 그림 및 공정 능력 상자 그림의 계산식에 사용되는 표기입니다 .

Y_{ij} = 공정 j 의 i 번째 관측값

$\bar{Y}_j$ = 공정 j 에 대한 관측값의 평균

SD(Y_j) = 공정 j 에 대한 관측값의 표준편차

T_j = 공정 j 의 목표값

LSL_j = 공정 j 의 규격 하한

USL_j = 공정 j 의 규격 상한

목표 그림

이 섹션에서는 목표 그림에 표시된 규격 값으로 표준화된 평균 변화 및 표준편차의 계산에 대해 자세히 설명합니다. 여기에서는 "목표 그림 및 공정 능력 상자 그림에 대한 표기"에 정의된 표기를 사용합니다.

j 번째 열에 대해 규격 한계로 표준화된 평균 변화 및 표준편차는 다음과 같이 정의됩니다 .

$$\text{규격으로 표준화된 평균} = \frac{\overline{Y}_j - T_j}{2 \times \min(T_j - \text{LSL}_j, \text{USL}_j - T_j)}$$

$$\text{규격으로 표준화된 표준편차} = \frac{SD(Y_j)}{2 \times \min(T_j - \text{LSL}_j, \text{USL}_j - T_j)}$$

참고 : LSL_j 또는 USL_j 중 하나가 결측이면 목표값과 비결측 규격 한계 사이 간격의 2 배가 목표 그림 좌표의 분모에 사용됩니다 .

목표값이 결측인 공정에 대한 목표 그림 점

공정에 규격 하한과 상한이 모두 있지만 목표값이 없는 경우를 가정해 보겠습니다. 그러면 "목표 그림"에 제공된 계산식에서 T_j를 두 규격 한계의 중간점으로 대체하여 사용됩니다.

공정에 규격 한계가 하나만 있고 목표값이 없는 경우를 가정해 보겠습니다 . 목표 그림의 한 점에 대한 (x,y) 좌표를 구하려면 해당 공정의 공정 능력 지수가 사용됩니다 . 이론적 평균 및 표준편차에 대한 정의는 " 정규 분포에 대한 공정 능력 지수 " 에서 확인하십시오 . 표본 관측값의 경우 다음 관계가 유지됩니다 .

$$C_{pu} = \frac{\text{USL}_j - \overline{Y}_j}{3SD(Y_j)}$$

$$C_{pl} = \frac{\overline{Y}_j - \text{LSL}_j}{3SD(Y_j)}$$

공정에 두 개의 규격 한계가 있고 한계의 중간점이 목표값인 경우 목표 그림의 한 점에 대한 (x,y) 좌표는 다음 관계를 충족합니다 .

$$C_{pu} = (0.5 - x)/3y$$

$$C_{pl} = (0.5 + x)/3y$$

규격 한계가 하나뿐이고 목표값이 없는 경우 좌표를 구하려면 이러한 관계가 사용됩니다. 고유점을 식별하려면 원점에서 점이 속한 선의 기울기에 대한 가정이 필요합니다. 규격 상한과 하한의 경우 기울기는 각각 0.5 와 -0.5 로 가정됩니다. 공정 능력 값이 1 이고 목표 그림 삼각형의 Ppk 슬라이더가 1 로 설정된 경우 이러한 기울기는 목표 그림 삼각형 선 위에 점을 배치합니다.

규격 상한만 있고 목표값이 없는 경우를 예로 들어보겠습니다 . 원점에서 0.5 기울기의 선 위에 (x,y) 좌표가 배치된다는 가정을 사용하여 x 와 y 를 계산하면 다음 좌표가 제공됩니다 .

$$x = 1/(3C_{pu} + 2)$$
$$y = 1/(6C_{pu} + 4)$$

규격 하한만 있고 목표값이 없는 경우를 예로 들어보겠습니다 . 원점에서 -0.5 기울기의 선 위에 (x,y) 좌표가 배치된다는 가정을 사용하여 x 와 y 를 계산하면 다음 좌표가 제공됩니다 .

$$x = -1/(3C_{pl} + 2)$$
$$y = 1/(6C_{pl} + 4)$$

참고 : C_{pu} 또는 C_{pl} 이 -0.6 미만이면 위의 계산식에서 값이 -0.6 으로 설정됩니다 . 값 -2/3 에서 x 의 분모는 값 0 을 가정합니다 . 공정 능력 값을 -0.6 으로 제한하면 분모 값을 0 으로 가정하거나 부호가 바뀌는 것을 방지할 수 있습니다 .

목표값이 결측인 공정에 대한 공정 능력 상자 그림

목표값이 없는 열에 규격 상한과 하한이 모두 있거나 규격 한계가 하나만 있을 수 있습니다. 이 섹션에서는 "목표 그림 및 공정 능력 상자 그림에 대한 표기"에 정의된 표기를 사용합니다.

두 개의 규격 한계가 지정되고 목표값 없음

j 번째 열에 목표값이 지정되지 않은 경우 공정 능력 상자 그림은 변환된 관측값에 대한 다음 값을 기반으로 합니다 .

$$Z_{ij} = \frac{Y_{ij} - (\mathrm{LSL}_j + \mathrm{USL}_j)/2}{\mathrm{USL}_j - \mathrm{LSL}_j}$$

단일 규격 한계가 지정되고 목표값 없음

규격 하한만 지정된 경우를 가정해 보겠습니다. 비슷한 방식으로 규격 상한만 지정된 경우도 가능합니다.

j 번째 열에 목표값이 지정되지 않은 경우 공정 능력 상자 그림은 변환된 관측값에 대한 다음 값을 기반으로 합니다 .

$$Z_{ij} = \frac{Y_{ij} - \bar{Y}_j}{2(\bar{Y}_j - \mathrm{LSL}_j)}$$

참고: 열에 규격 한계가 하나뿐이고 목표값이 없는 경우 표본 평균이 규격 구간을 벗어나면 해당 열에 대한 공정 능력 상자 그림이 표시되지 않습니다 .

정규 분포에 대한 공정 능력 지수

이 섹션에서는 정규 데이터에 대한 공정 능력 지수의 계산에 대해 자세히 설명합니다.

평균이 μ 이고 표준편차가 σ 인 공정 특성에 대한 모집단 기반 공정 능력 지수는 다음과 같이 정의됩니다 . 표본 관측값의 경우 모수가 다음 추정값으로 대체됩니다 .

$$Cp = \frac{USL - LSL}{6\sigma}$$

$$Cpl = \frac{\mu - LSL}{3\sigma}$$

$$Cpu = \frac{USL - \mu}{3\sigma}$$

$$Cpk = \min(Cpl, Cpu)$$

$$Cpm = \frac{\min(T - LSL, USL - T)}{3\sigma\sqrt{1 + \left(\frac{T - \mu}{\sigma}\right)^2}}$$

$$\text{목표 지수} = 3(Cp - Cpk)$$

이 계산식에서는 다음과 같은 표기를 사용합니다 .

LSL = 규격 하한
USL = 규격 상한
T = 목표값

군내 표준편차 공정 능력 추정값의 경우 사용자가 선택한 부분군 지정 방법을 사용하여 σ를 추정합니다. 전체 표준편차 공정 능력 추정값의 경우 표본 표준편차를 사용하여 σ를 추정합니다. 규격 한계가 하나뿐인 경우 결측 규격 한계를 포함하는 공정 능력 지수가 결측값으로 보고됩니다.

참고: 기본 AIAG(Ppk) 라벨을 사용하면 전체 표준편차를 기반으로 하는 지수가 Pp, Ppl, Ppu 및 Ppk 로 표시됩니다 . 전체 표준편차를 사용할 때 Cpm 지수의 라벨은 변경되지 않습니다 . 이 섹션의 계산식은 Cp 라벨을 사용하여 정의됩니다 .

공정 능력 지수에 대한 신뢰 구간

공정 능력 지수에 대한 신뢰 구간은 정규 분포를 따르는 공정에만 사용할 수 있습니다. 신뢰 구간은 군내 표준편차 및 전체 표준편차 공정 능력에 대해 모두 계산되며 개별 상세 정보 보고서에 표시됩니다.

Cp

Cp 에 대한 100(1 - α)% 신뢰 구간은 다음과 같이 계산됩니다 .

$$\left(\hat{\mathrm{Cp}} \sqrt{\frac{\chi^2_{\alpha/2,\,df}}{df}},\ \hat{\mathrm{Cp}} \sqrt{\frac{\chi^2_{1-\alpha/2,\,df}}{df}} \right)$$

다음은 각 요소에 대한 설명입니다 .

$\hat{\mathrm{Cp}}$ = Cp 에 대한 추정값

$\chi^2_{\alpha/2,\,df}$ = 자유도가 *df* 인 카이제곱 분포의 (α/2) 번째 분위수

df = 자유도

N = 관측값 수

m = 부분군 수

전체 표준편차 공정 능력의 경우 자유도는 *N* - 1 입니다.

군내 표준편차 공정 능력의 경우 자유도는 부분군 지정 및 군내 표준편차 추정 방법에 따라 달라집니다.

- 불균형 부분군을 사용하는 군내 표준편차 공정 능력의 경우 자유도 계산은 군내 표준편차 추정 방법에 관계없이 동일합니다 . 자유도는 *N* - *m* 입니다 .
- 균형 부분군 (크기 *n* = *N*/*m*) 을 사용하는 군내 표준편차 공정 능력의 경우 자유도 계산은 군내 표준편차 추정 방법에 따라 달라집니다 .
  - 비편향 표준편차의 평균을 사용하여 군내 표준편차를 추정하는 경우 자유도는 *f* * (*N* - *m*) 입니다 . 배율 *f* 는 0.875 와 1 사이에서 변하며 다음과 같이 정의됩니다 .

$$f = \frac{1}{2(n-1)\left(\frac{(n-1)}{2} \left(\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \right)^2 - 1 \right)}$$

    여기서 Γ(*n*) 는 *n* 에서 실행된 감마 함수입니다 .

    자세한 내용은 Bissell 연구 자료 (1990) 에서 확인하십시오 .

  - 범위 평균을 사용하여 군내 표준편차를 추정하는 경우 자유도는 *df* = 1/*A* - (3/16) * *A* + (3/64) * A^2 + 0.25 로 계산됩니다 . *A* 는 다음과 같이 정의됩니다 .

$$A = \frac{2d_3(n)^2}{m \cdot d_2(n)^2}$$

$d_2(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 기대값

$d_3(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 표준편차

자세한 내용은 David 연구 자료 (1951) 에서 확인하십시오 .

- 비편향 합동 표준편차를 사용하여 군내 표준편차를 추정하는 경우 자유도는 $N - m$ 입니다.

- 부분군이 없는 군내 표준편차 공정 능력의 경우 자유도 계산은 군내 표준편차 추정 방법에 따라 달라집니다 .
  - 평균 이동 범위를 사용하여 군내 표준편차를 추정하는 경우 자유도는 0.62 * (N - 1) 로 계산됩니다 .
  - 중앙값 이동 범위를 사용하여 군내 표준편차를 추정하는 경우 자유도는 0.32 * (N - 1) 로 계산됩니다 .

  자세한 내용은 Wheeler 연구 자료 (2004, p. 82) 에서 확인하십시오 .

Cpk

Cpk 에 대한 100(1 - α)% 신뢰 구간은 다음과 같이 계산됩니다 .

$$\left(\hat{C}_{pk}\left[1-\Phi^{-1}{}_{1-\alpha/2}\sqrt{\frac{1}{9N\hat{C}_{pk}^2}+\frac{1}{2df}}\right],\ \hat{C}_{pk}\left[1+\Phi^{-1}{}_{1-\alpha/2}\sqrt{\frac{1}{9N\hat{C}_{pk}^2}+\frac{1}{2df}}\right]\right)$$

다음은 각 요소에 대한 설명입니다 .

$\hat{C}_{pk}$ = Cpk 에 대한 추정값

$\Phi^{-1}{}_{1-\alpha/2}$ = 표준 정규 분포의 (1 - α/2) 번째 분위수

df = 자유도

N = 관측값 수

m = 부분군 수

전체 표준편차 공정 능력의 경우 자유도는 N - 1 입니다.

군내 표준편차 공정 능력의 경우 자유도는 부분군 지정 및 군내 표준편차 추정 방법에 따라 달라집니다.

- 불균형 부분군을 사용하는 군내 표준편차 공정 능력의 경우 자유도 계산은 군내 표준편차 추정 방법에 관계없이 동일합니다 . 자유도는 $N - m$ 입니다 .
- 균형 부분군 (크기 $n = N/m$) 을 사용하는 군내 표준편차 공정 능력의 경우 자유도 계산은 군내 표준편차 추정 방법에 따라 달라집니다 .
  - 비편향 표준편차의 평균을 사용하여 군내 표준편차를 추정하는 경우 자유도는 f * ($N - m$) 입니다 . 배율 f 는 0.875 와 1 사이에서 변하며 다음과 같이 정의됩니다 .

$$f = \frac{1}{2(n-1)\left(\frac{(n-1)}{2}\left(\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right)^2 - 1\right)}$$

여기서 $\Gamma(n)$ 는 n 에서 실행된 감마 함수입니다 .

자세한 내용은 Bissell 연구 자료 (1990) 에서 확인하십시오 .

– 범위 평균을 사용하여 군내 표준편차를 추정하는 경우 자유도는 $df = 1/A - (3/16) * A + (3/64) * A^2 + 0.25$ 로 계산됩니다 . A 는 다음과 같이 정의됩니다 .

$$A = \frac{2d_3(n)^2}{m \cdot d_2(n)^2}$$

$d_2(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 기대값

$d_3(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 표준편차

자세한 내용은 David 연구 자료 (1951) 에서 확인하십시오 .

– 비편향 합동 표준편차를 사용하여 군내 표준편차를 추정하는 경우 자유도는 $N - m$ 입니다 .

- 부분군이 없는 군내 표준편차 공정 능력의 경우 자유도 계산은 군내 표준편차 추정 방법에 따라 달라집니다 .
  – 평균 이동 범위를 사용하여 군내 표준편차를 추정하는 경우 자유도는 $0.62 * (N - 1)$ 로 계산됩니다 .
  – 중앙값 이동 범위를 사용하여 군내 표준편차를 추정하는 경우 자유도는 $0.32 * (N - 1)$ 로 계산됩니다 .

  자세한 내용은 Wheeler 연구 자료 (2004, p. 82) 에서 확인하십시오 .

Cpm

참고 : Cpm 에 대한 신뢰 구간은 목표값이 규격 하한과 규격 상한 사이에서 중심화되는 경우에만 계산됩니다 .

Cpm 에 대한 $100(1 - \alpha)\%$ 신뢰 구간은 다음과 같이 계산됩니다 .

$$\left(\hat{Cpm}\sqrt{\frac{\chi^2_{\alpha/2,\gamma}}{\gamma}}, \hat{Cpm}\sqrt{\frac{\chi^2_{1-\alpha/2,\gamma}}{\gamma}}\right)$$

다음은 각 요소에 대한 설명입니다 .

$\hat{Cpm}$ = Cpm 에 대한 추정값

$\chi^2_{\alpha/2,\gamma}$ = 자유도가 γ 인 카이제곱 분포의 (α/2) 번째 분위수

$$\gamma = \frac{N\left(1+\left(\frac{\bar{x}-T}{s}\right)^2\right)^2}{1+2\left(\frac{\bar{x}-T}{s}\right)^2}$$

N = 관측값 수

$\bar{x}$ = 관측값의 평균

T = 목표값

s = 표준편차 추정값

전체 표준편차 공정 능력의 경우 s는 전체 표준편차 추정값입니다. 군내 표준편차 공정 능력의 경우 s는 군내 표준편차 추정값으로 대체됩니다.

팁 : Cp, Cpk 및 Cpm 의 신뢰 구간에 대한 자세한 내용은 Pearn and Kotz 연구 자료 (2006) 에서 확인하십시오 .

Cpl, Cpu

Cpl 과 Cpu 의 신뢰 하한 및 상한은 Chou et al. (1990) 의 방법을 사용하여 계산됩니다 .

Cpl 에 대한 100(1 - α)% 신뢰 한계 (CPL_L, CPL_U 로 표시) 는 다음 방정식을 충족합니다 .

$\Pr[t_{n-1}(\delta_L) \geq 3\hat{C}_{pl}\sqrt{n}] = \alpha/2$ 여기서 , $\delta_L = 3CPL_L\sqrt{n}$

$\Pr[t_{n-1}(\delta_U) \leq 3\hat{C}_{pl}\sqrt{n}] = \alpha/2$ 여기서 , $\delta_U = 3CPL_U\sqrt{n}$

다음은 각 요소에 대한 설명입니다 .

$t_{n\text{-}1}(\delta)$ 은 n - 1 자유도와 비중심성 모수 δ 를 사용하는 비중심 t 분포를 따릅니다 .

$\hat{C}_{pl}$ = Cpl 에 대한 추정값

Cpu 에 대한 100(1 - α)% 신뢰 한계 (CPU_L, CPU_U 로 표시) 는 다음 방정식을 충족합니다 .

$\Pr[t_{n-1}(\delta_L) \geq 3\hat{C}_{pu}\sqrt{n}] = \alpha/2$ 여기서 , $\delta_L = 3CPU_L\sqrt{n}$

$\Pr[t_{n-1}(\delta_U) \leq 3\hat{C}_{pu}\sqrt{n}] = \alpha/2$ 여기서 , $\delta_U = 3CPU_U\sqrt{n}$

다음은 각 요소에 대한 설명입니다.

$t_{n\text{-}1}(\delta)$ 은 n - 1 자유도와 비중심성 모수 δ 를 사용하는 비중심 t 분포를 따릅니다 .

$\hat{C}_{pu}$ = Cpu 에 대한 추정값

비정규 분포에 대한 공정 능력 지수 : 백분위수 및 Z 스코어 방법

이 섹션에서는 비정규 분포에 대한 공정 능력 지수를 계산하는 방법을 설명합니다. 백분위수(ISO/분위수라고도 함)와 Z 스코어(Bothe/Z 스코어라고도 함)라는 두 가지 방법이 설명됩니다. 비정규 공정 변수에 대한 분포를 선택할 때 모수 분포 또는 비모수 분포를 적합시킬 수 있습니다. 백분위수 또는 Z 스코어 방법을 사용하여 관심 공정 변수에 대한 공정 능력 지수를 계산할 수 있습니다. 그러나 데이터 양이 매우 많은 경우가 아니면 비모수 적합은 분포의 꼬리 부분에서 동작을 정확히 반영하지 못할 수 있습니다.

참고 : 데이터가 정규 분포를 따르는 경우 백분위수 방법과 Z 스코어 방법 둘 다 공정 능력 계산식이 정규성 기반 공정 능력 지수의 계산식으로 축소됩니다 .

두 방법을 설명할 때 다음 표기가 사용됩니다 .

LSL = 규격 하한
USL = 규격 상한
T = 목표값

백분위수 (ISO/ 분위수) 방법

백분위수 방법은 표준 공정 능력 계산식의 평균을 적합 분포의 중앙값으로 대체하고 6σ 범위의 값을 해당 백분위수 범위로 대체합니다. 이 방법에 대한 설명은 AIAG 연구 자료(2005)에서 확인하십시오.

적합 분포의 α*100 번째 백분위수를 P_α 로 나타냅니다 . 백분위수 방법 공정 능력 지수는 다음과 같이 정의됩니다 .

$$P_{pk} = \min\left(\frac{P_{0.5} - LSL}{P_{0.5} - P_{0.00135}}, \frac{USL - P_{0.5}}{P_{0.99865} - P_{0.5}}\right)$$

$$P_{pl} = \frac{P_{0.5} - LSL}{P_{0.5} - P_{0.00135}}$$

$$P_{pu} = \frac{USL - P_{0.5}}{P_{0.99865} - P_{0.5}}$$

$$P_p = \frac{USL - LSL}{P_{0.99865} - P_{0.00135}}$$

$$C_{pm} = \frac{\min\left(\frac{T - LSL}{P_{0.5} - P_{0.00135}}, \frac{USL - T}{P_{0.99865} - P_{0.5}}\right)}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

Z 스코어 (Bothe/Z 스코어) 방법

Z 스코어 방법은 규격 한계를 표준 정규 척도에서 동일한 확률을 갖는 값으로 변환합니다. 이 방법을 사용하면 적합된 비정규 분포와 위험 수준이 동일한 정규 분포에 해당하는 공정 능력 측도가 계산됩니다.

*F*가 규격 하한 및 상한(LSL, USL)이 지정된 공정 변수의 적합 분포를 나타낸다고 하겠습니다. 동등한 표준 정규 규격 한계는 다음과 같이 정의됩니다 .

$$\begin{bmatrix} LSL_F = \Phi^{-1}(F(LSL)) \\ USL_F = \Phi^{-1}(F(USL)) \end{bmatrix}$$

Z 스코어 방법 공정 능력 지수는 다음과 같이 정의됩니다 .

$$P_{pk} = \min(-LSL_F/3, USL_F/3)$$

$$P_{pl} = -LSL_F/3$$

$$P_{pu} = USL_F/3$$

$$P_p = (USL_F - LSL_F)/6$$

참고 : Cpm 은 목표값 기반 측도이므로 Z 스코어 방법을 사용하여 계산할 수 없습니다 .

참고 : 공정 능력이 충분한 데이터의 경우 *F(LSL)* 과 *F(USL)* 이 각각 0 과 1 에 너무 가까워서 LSL_F 또는 USL_F 를 계산할 수 없습니다 . 이러한 경우 기본적으로 Z 스코어 방법이 백분위수 방법으로 자동 전환됩니다 . 결과적으로 더 유의한 공정 능력 지수가 제공됩니다 . 이 기본 설정을 해제하려면 파일 > 환경 설정 > 플랫폼 > 공정 능력을 선택하십시오 .

분포의 파라미터화

이 섹션에서는 공정 능력 플랫폼에 사용되는 분포에 대한 밀도 함수 *f*를 제공합니다. Johnson 및 SHASH 분포를 제외한 모든 분포에 대한 기대값과 분산도 제공합니다.

정규

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

$$E(x) = \mu$$

$$Var(x) = \sigma^2$$

베타

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 \le x \le 1,\ \alpha > 0,\ \beta > 0$$

$$E(x) = \frac{\alpha}{\alpha + \beta}$$

$$Var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

여기서 $B(\cdot)$는 베타 함수입니다.

지수

$$f(x|\sigma) = \frac{1}{\sigma}\exp(-x/\sigma),\ x > 0,\ \sigma > 0$$

$$E(x) = \sigma$$

$$Var(x) = \sigma^2$$

감마

$$f(x|\alpha, \sigma) = \frac{1}{\Gamma(\alpha)\sigma^{\alpha}} x^{\alpha-1}\exp(-x/\sigma),\ x > 0,\ \alpha > 0,\ \sigma > 0$$

$$E(x) = \alpha\sigma$$

$$Var(x) = \alpha\sigma^2$$

여기서 $\Gamma(\cdot)$는 감마 함수입니다.

Johnson

Johnson Su

$$f(x|\gamma, \delta, \sigma, \theta) = \frac{\delta}{\sigma}\left[1 + \left(\frac{x-\theta}{\sigma}\right)^2\right]^{-1/2} \phi\left[\gamma + \delta \sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)\right],\ -\infty < x, \theta, \gamma < \infty,\ \theta > 0,\ \delta > 0$$

Johnson Sb

$$f(x|\gamma, \delta, \sigma, \theta) = \phi\left[\gamma + \delta \ln\left(\frac{x-\theta}{\sigma-(x-\theta)}\right)\right]\left(\frac{\delta\sigma}{(x-\theta)(\sigma-(x-\theta))}\right),\ \theta < x < \theta + \sigma,\ \sigma > 0$$

Johnson Sl

$$f(x|\gamma, \delta, \sigma, \theta) = \frac{\delta}{|x-\theta|}\phi\left[\gamma + \delta \ln\left(\frac{x-\theta}{\sigma}\right)\right],\ x > 0 \text{ 단, } \sigma = 1,\ x < 0 \text{ 단, } \sigma = -1$$

여기서 $\phi(\cdot)$는 표준 정규 확률 밀도 함수입니다.

로그 정규

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \frac{\exp\left[\frac{-(\log(x)-\mu)^2}{2\sigma^2}\right]}{x}, \; x > 0, \; -\infty < \mu < \infty, \; \sigma > 0$$

$$E(x) = \exp(\mu + \sigma^2/2)$$

$$Var(x) = \exp(2(\mu + \sigma^2)) - \exp(2\mu + \sigma^2)$$

정규 혼합

2 정규 혼합 및 3 정규 혼합 분포 옵션은 다음 파라미터화를 공유합니다 .

$$f(x|\mu_i, \sigma_i, \pi_i) = \sum_{i=1}^{k} \frac{\pi_i}{\sigma_i} \phi\left(\frac{x-\mu_i}{\sigma_i}\right)$$

$$E(x) = \sum_{i=1}^{k} \pi_i \mu_i$$

$$Var(x) = \sum_{i=1}^{k} \pi_i(\mu_i^2 + \sigma_i^2) - \left(\sum_{i=1}^{k} \pi_i \mu_i\right)^2$$

여기서 μ_i, σ_i 및 π_i는 각각 i번째 그룹의 평균, 표준편차 및 비율이며, $\phi(\cdot)$는 표준 정규 확률 밀도 함수입니다. 2 정규 혼합의 경우 k는 2입니다. 3 정규 분포 혼합의 경우 k는 3입니다. 혼합물 내 각 그룹에 대해 평균, 표준편차 및 전체 대비 비율이 개별적으로 추정됩니다.

SHASH

$$f(x|\gamma, \delta, \sigma, \theta) = \frac{\delta\cosh(w)}{\sqrt{\sigma^2 + (x-\theta)^2}} \phi[\sinh(w)], \; -\infty < \gamma, x, \theta < \infty, \; 0 < \delta, \; 0 < \sigma$$

다음은 각 요소에 대한 설명입니다 .

$\phi(\cdot)$ = 표준 정규 pdf

$$w = \gamma + \delta\sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)$$

참고: $\gamma = 0$이고 $\delta = 1$인 경우 SHASH 분포는 위치가 θ이고 척도가 σ인 정규 분포와 동등합니다.

Weibull

$$f(x|\alpha,\beta) = \frac{\beta}{\alpha^{\beta}}x^{\beta-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right],\ \alpha>0,\ \beta>0$$

$$\mathrm{E}(x)=\alpha\Gamma\left(1+\frac{1}{\beta}\right)$$

$$\mathrm{Var}(x)=\alpha^{2}\left\{\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^{2}\left(1+\frac{1}{\beta}\right)\right\}$$

여기서 $\Gamma(\cdot)$는 감마 함수입니다.

8 장

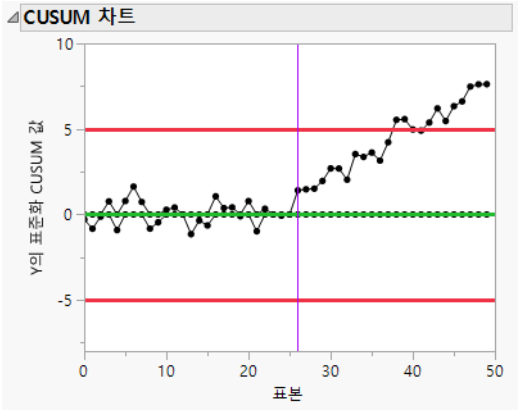
CUSUM 관리도

결정 한계가 포함된 테이블 형식 CUSUM 관리도 생성

CUSUM(누적합) 관리도를 사용하면 공정의 작은 변화를 감지할 수 있습니다. 점진적 흐름과 같이 시간이 지남에 따라 발생하고 반드시 갑작스러운 변동이 동반되지는 않는 변화를 감지하는 데 유용합니다. CUSUM 관리도는 결정 한계가 포함된 CUSUM 차트를 생성하며 Shewhart 차트와 유사합니다. 이 차트를 **테이블 형식 CUSUM 차트**라고도 합니다. V-마스크 누적합 관리도를 생성하려면 "V-마스크 CUSUM 관리도"의 내용을 참조하십시오.

CUSUM 관리도 플랫폼에서는 ARL(평균 런 길이)에 대한 정보도 제공합니다. **평균 런 길이**는 관리이탈 신호가 발생하기 전에 이를 예상할 수 있는 평균 표본 또는 관측값 수입니다. 특정 모수가 제공되고 분산이 일정하다고 가정할 때 평균 런 길이를 사용하여 CUSUM 차트의 성능을 평가할 수 있습니다.

그림 8.1 CUSUM 관리도



목차

CUSUM 관리도 플랫폼 개요 231
CUSUM 관리도의 예 231
CUSUM 관리도 플랫폼 시작 233
CUSUM 관리도 플랫폼 보고서 234
CUSUM 관리도의 제어판 234
CUSUM 차트 235
CUSUM 관리도 플랫폼 옵션 236
평균 런 길이(ARL) 보고서 238
CUSUM 관리도의 추가 예 239
CUSUM 관리도의 데이터 단위 옵션 예 239
부분군이 있는 CUSUM 차트의 예 240
CUSUM 관리도 플랫폼에 대한 통계 상세 정보 241
CUSUM 관리도 생성에 대한 통계 상세 정보 241
변화 감지에 대한 통계 상세 정보 242
평균 런 길이에 대한 통계 상세 정보 242

CUSUM 관리도 플랫폼 개요

테이블 형식 CUSUM 차트는 두 개의 단측 결정 한계 차트가 하나의 차트에 겹쳐져 있습니다. 이 차트는 공정이 관리이탈 상태일 때 신호를 보내는 결정 한계를 포함하고 차트에서 변화 발생 지점으로 의심되는 위치에 변화 선을 표시합니다. CUSUM 관리도 플랫폼을 사용하려면 중요한 것으로 간주되는 최소 평균 변화를 결정해야 합니다. 표준편차 단위 또는 데이터 단위로 CUSUM 관리도를 볼 수 있습니다. 테이블 형식 CUSUM 차트에 대한 자세한 내용은 Woodall and Adams(1998) 및 Montgomery(2013) 연구 자료에서 확인하십시오.

누적합 관리도의 또 다른 유형에는 V-마스크 차트가 있습니다. V-마스크 CUSUM 차트를 생성하려면 "V-마스크 CUSUM 관리도"의 내용을 참조하십시오.

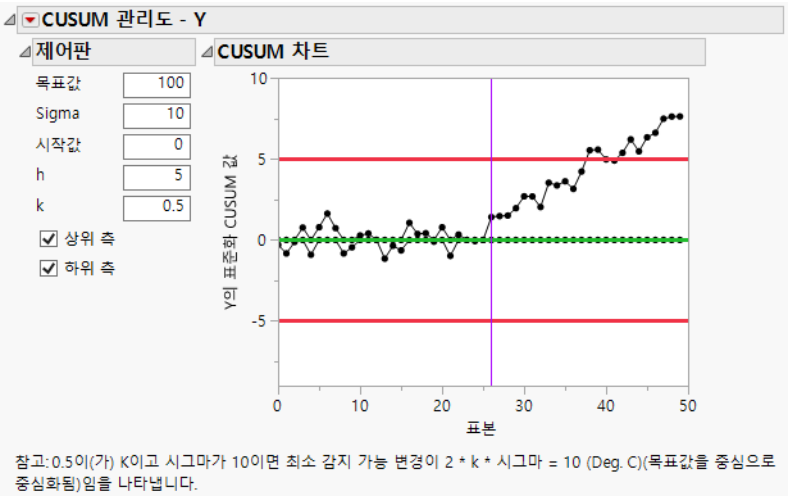
참고 : CUSUM 관리도 플랫폼의 요약 결과가 V- 마스크 CUSUM 플랫폼의 요약 결과와 항상 일치하지는 않습니다 . 특히 양측 V- 마스크 CUSUM 차트의 요약 결과는 " 상위 측 " 옵션과 " 하위 측 " 옵션이 모두 선택된 CUSUM 관리도의 요약 결과와 일치하지 않습니다 . 그러나 CUSUM 관리도 플랫폼과 V- 마스크 CUSUM 플랫폼의 단측 요약 보고서는 일치합니다 .

CUSUM 관리도의 예

엔진 온도의 작은 변화를 감지하려고 합니다 . 데이터 테이블에는 엔진 온도 조절 장치에서 수집한 온도 측정값이 포함되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Engine Temperature Sensor.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > CUSUM 관리도**를 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 목표값 " 상자에 100 을 입력합니다 .
6. " 시그마 " 상자에 10 을 입력합니다 .

그림 8.2 CUSUM 관리도 보고서



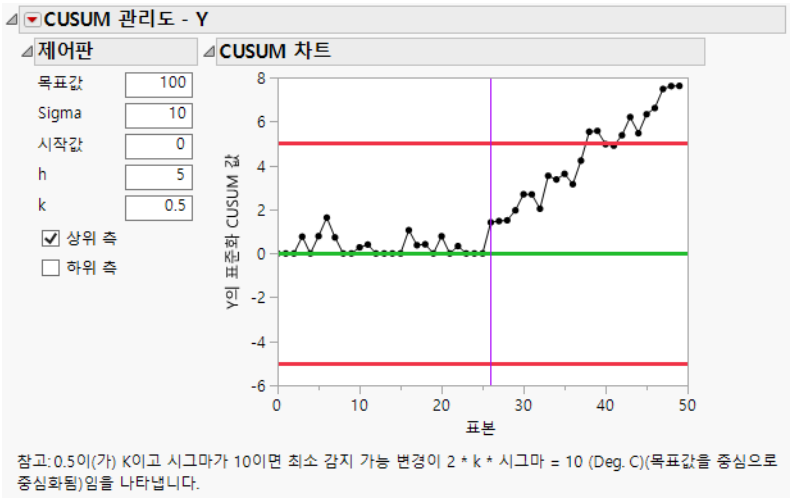
CUSUM 차트의 세로선은 온도 측정값 변화가 대략 표본 26에서 시작되었음을 나타냅니다.

참고 : Engine Temperature Sensor.jmp 의 "IMR Chart" 테이블 스크립트를 실행하여 이 결과를 개별값 및 이동 범위 관리도와 비교할 수 있습니다 . IMR 차트에서는 Nelson 검정에 걸리는 것이 없습니다 .

단측 CUSUM 관리도의 예

이전 예에서 온도 변화 증가에만 관심이 있다고 가정하여 계속합니다. 그림 8.2 의 CUSUM 관리도를 단측 차트로 변경하려면 "하위 측" 체크박스의 선택을 취소합니다. 그러면 음수 누적합에 대한 점이 차트에서 제거됩니다. 즉, CUSUM 관리도에 양수 누적합 점만 포함되어 있습니다.

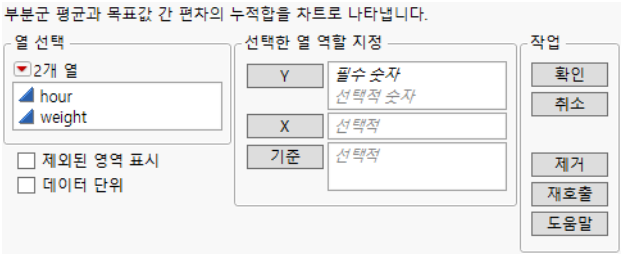
그림 8.3 단측 CUSUM 관리도 보고서



CUSUM 관리도 플랫폼 시작

분석 > 품질 및 공정 > 관리도 > CUSUM 관리도를 선택하여 CUSUM 관리도 플랫폼을 시작할 수 있습니다 .

그림 8.4 CUSUM 관리도 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

CUSUM 관리도 플랫폼 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y 차트를 생성할 변수를 식별합니다 .

참고 : 데이터 테이블의 행은 관측값을 수집한 순서대로 정렬되어야 합니다 .

X 부분군 변수를 식별합니다 . CUSUM 차트의 가로 축은 부분군 변수로 라벨이 지정됩니다 . 이 열의 값이 두 번 이상 나오면 각 X 값의 평균 반응이 CUSUM 차트에 표시됩니다 .

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

제외된 영역 표시 (X 변수가 지정된 경우에만 적용 가능) 완전히 제외된 부분군이 CUSUM 관리도의 가로 축에 표시되도록 지정합니다.

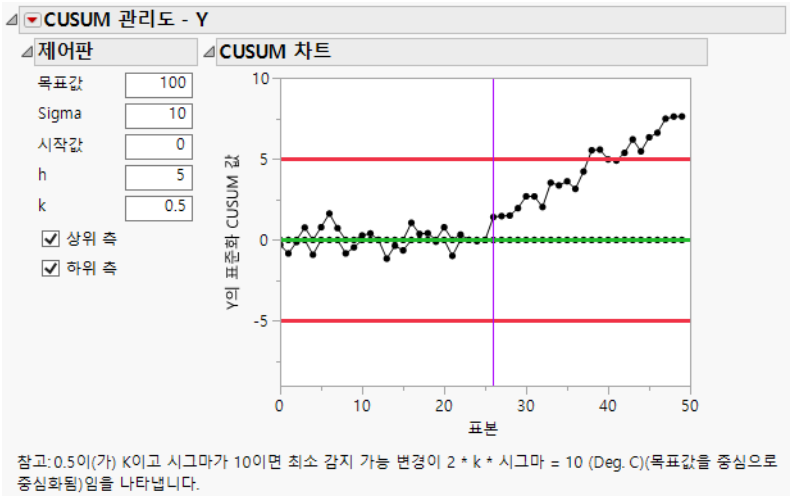
데이터 단위 표준편차 단위 대신 데이터 단위를 보고서에 사용하도록 지정합니다. 기본적으로 차트와 모수는 표준편차 단위로 표시됩니다. 그러나 시작 창에서 "데이터 단위" 옵션을 선택하면 차트와 모수가 분석 대상 데이터 열의 단위로 표시됩니다.

표준편차 단위를 사용하면 h 및 k 모수의 값이 공정 표준편차에 따라 달라지지 않습니다. 이것은 장점이 될 수 있습니다.

CUSUM 관리도 플랫폼 보고서

기본적으로 CUSUM 관리도 플랫폼에서는 모수 제어판과 CUSUM 차트가 포함된 보고서를 생성합니다.

그림 8.5 CUSUM 관리도 보고서



CUSUM 관리도의 제어판

"제어판" 보고서에는 차트 모수의 현재 값이 포함됩니다. 현재 값은 모수 값을 업데이트할 수 있는 상자에 있습니다. "상위 측" 및 "하위 측" 상자도 있습니다. 시작 창에서 "데이터 단위" 옵션을 지정한 경우 "제어판"의 체크박스 아래에 이 설정이 표시됩니다.

제어판 보고서에는 다음 옵션이 표시됩니다 .

목표값 알려진 평균 값입니다 . 이 값은 차트의 중심선 값입니다 . 기본적으로 이 모수는 Y 열의 " 규격 한계 " 열 특성에 지정된 " 목표값 " 으로 설정됩니다 . Y 열의 " 규격 한계 " 열 특성에 목표값이 지정되지 않은 경우 이 모수는 Y 열의 전체 평균으로 설정됩니다 .

참고 : Y 열의 " 규격 한계 " 열 특성에 목표값이 지정된 경우에도 Y 열의 전체 평균을 중심선 값으로 사용하려면 " 중심선에 공정 평균 사용 " 플랫폼 환경 설정을 선택하십시오 . 이 환경 설정은 파일 > 환경 설정 > 플랫폼 > CUSUM 관리도에 있습니다 .

시그마 알려진 표준편차 값입니다 . 기본적으로 이 모수는 Y 열의 평균 이동 범위로 설정됩니다 . X 변수가 있는 경우 시그마 모수는 요약 데이터의 평균 이동 범위로 설정됩니다 .

시작값 첫 번째 표본 이전의 누적합 값입니다 . 0 이 아닌 값에서 누적합을 시작하면 표본 시작 근처에서 CUSUM 차트의 민감도가 증가합니다 . 이 모수는 FIR(Fast Initial Response) 값이라고도 합니다 . 기본적으로 이 모수는 0 으로 설정됩니다 .

h 또는 H 한계를 정의하는 모수의 값입니다 . 시작 창에서 " 데이터 단위 " 옵션을 선택하지 않은 경우 h 모수입니다 . 시작 창에서 " 데이터 단위 " 옵션을 선택한 경우에는 H 모수입니다 . H 는 h 에 시그마를 곱한 값입니다 . 기본적으로 h 는 5 로 설정되고 H 는 시그마 값의 5 배입니다 .

k 또는 K 감지할 최소 평균 변화를 정의하는 모수의 값입니다 . 시작 창에서 " 데이터 단위 " 옵션을 선택하지 않은 경우 k 모수입니다 . 시작 창에서 " 데이터 단위 " 옵션을 선택한 경우에는 K 모수입니다 . K 는 k 에 시그마를 곱한 값입니다 . 기본적으로 k 는 0.5 로 설정되고 K 는 시그마 값의 1/2 입니다 .

상위 측 차트에 누적합에 대한 양수 값을 표시하거나 숨깁니다 . 이러한 값은 C^+ 값입니다 .

하위 측 차트에 누적합에 대한 음수 값을 표시하거나 숨깁니다 . 이러한 값은 C^- 값입니다 .

데이터 단위 사용 이 텍스트가 있으면 시작 창에서 " 데이터 단위 " 옵션이 선택되었고 CUSUM 차트의 값이 중심화되지만 표준화되지 않았음을 나타냅니다 .

CUSUM 차트

"CUSUM 차트" 보고서에는 차트 모수의 현재 값에 의해 지정된 결정 한계가 있는 누적합 관리도가 포함됩니다. 표본 또는 부분군(X 변수를 지정한 경우)이 가로 축에 표시됩니다. 세로 축에는 누적합에 대한 양수 값과 음수 값의 중심화된 값이 표시됩니다. 시작 창에서 "데이터 단위" 옵션을 선택하지 않은 경우 표준화된 반응 값의 누적합이 세로 축에 표시됩니다. 시작 창에서 "데이터 단위" 옵션을 선택한 경우에는 표준화되지 않은 반응 값의 누적합이 세로 축에 표시됩니다.

CUSUM 관리도 플랫폼 옵션

"CUSUM 관리도 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

한계 표시 CUSUM 차트에 결정 상한 및 하한을 표시하거나 숨깁니다 .

중심선 표시 CUSUM 차트에 중심선을 표시하거나 숨깁니다 .

변화 선 표시 (데이터에서 변화를 감지한 경우에만 사용 가능) CUSUM 차트에 변화를 지정하는 세로선을 표시하거나 숨깁니다 . 변화 선은 변화가 시작되는 지점에 그립니다 .

- 차트에서 C^+ 값이 상한선 위에 있으면 양의 변화가 발생합니다 . 변화 시작은 C^+ 에 대한 가장 최근 0 값 이후 첫 번째 점으로 정의됩니다 .
- 차트에서 C^- 값이 하한선 아래에 있으면 음의 변화가 발생합니다 . 변화 시작은 C^- 에 대한 가장 최근 0 값 이후 첫 번째 점으로 정의됩니다 .

ARL 표시 평균 런 길이 (ARL) 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 평균 런 길이 (ARL) 보고서 " 에서 확인하십시오 .

ARL 프로파일러 평균 런 길이 대 h 및 k 모수의 프로파일러를 표시하거나 숨깁니다 . 시작 창에서 " 데이터 단위 " 옵션을 지정한 경우 평균 런 길이 대 H 및 K 모수의 프로파일러가 표시됩니다 .

지정된 변화에 대한 ARL(평균 런 길이) 은 관리이탈 신호가 발생하기 전에 예상되는 평균 런 수입니다 . 예를 들어 ARL(0) 은 공정이 관리 상태일 때 오경보 신호가 발생하기 전에 예상되는 평균 런 수를 나타냅니다 . 공정이 관리 상태일 때 변화 크기는 0 입니다 .

ARL 프로파일러를 사용하면 모수의 다양한 설정이 해당 CUSUM 차트의 성능에 어떤 영향을 주는지 살펴볼 수 있습니다 . " 제어판 " 보고서의 모수가 업데이트되면 ARL 프로파일러도 업데이트됩니다 . 이상적인 CUSUM 차트는 ARL(0) 값이 높고 ARL(Δ) 값이 낮습니다 . 여기서 Δ 는 관심 있는 크기 변화입니다 .

ARL 프로파일러는 "제어판" 보고서의 "상위 측" 및 "하위 측" 옵션 설정에 따라 달라집니다 .

- " 상위 측 " 옵션과 " 하위 측 " 옵션을 모두 선택한 경우 프로파일러는 CUSUM 차트에서 결정 상한 또는 하한을 벗어나는 평균 런 길이를 나타냅니다 .
- " 상위 측 " 옵션만 선택한 경우 프로파일러는 CUSUM 차트에서 결정 상한에 대한 평균 런 길이를 나타냅니다 .
- " 하위 측 " 옵션만 선택한 경우 프로파일러는 CUSUM 차트에서 결정 하한에 대한 평균 런 길이를 나타냅니다 .

"ARL 프로파일러 " 옆의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 Profilers 의 에서 확인하십시오 .

제어판 모수의 현재 값을 보여 주는 보고서를 표시하거나 숨깁니다 . 이 보고서를 사용하면 모수 값 및 CUSUM 차트의 측면성을 변경할 수 있습니다 .

모수 보고서 모수의 현재 값을 보여 주는 보고서를 표시하거나 숨깁니다 .

한계를 벗어난 경우 검정 CUSUM 차트에서 상한선 위에 있거나 하한선 아래에 있는 모든 점 주위에 빨간색 원을 표시하거나 숨깁니다.

요약 저장 CUSUM 차트의 각 부분군에 대한 통계량을 포함하는 새 데이터 테이블을 생성합니다. 부분군 수와 크기, 부분군 표본 평균, 변화 시작 표시자, 변화 시작 사이의 각 간격을 나타내는 값, 상위 / 하위 누적합과 해당하는 연속 런 수 및 LCL/UCL 값 등의 통계량이 새 데이터 테이블에 저장됩니다.

차트 조정 Y 변수에 대한 허용 범위를 설정할 수 있는 컨트롤을 표시하거나 숨깁니다. 허용 범위의 최소값과 최대값을 조정하고 **완료**를 클릭합니다. 이때 k 모수의 새 값을 기반으로 CUSUM 차트가 업데이트됩니다.

- 시작 창에서 " 데이터 단위 " 옵션을 지정한 경우 결측값 대치된 k 모수는 최소값과 최대값의 평균입니다.
- " 데이터 단위 " 옵션을 지정하지 않은 경우 결측값 대치된 K 모수는 최소값과 최대값의 평균을 시그마 모수로 나눈 값입니다.

Y 변수에 대한 허용 범위를 설정하면 CUSUM 차트의 실제적 유의성을 설정할 수 있습니다. 특히 검정 간격이 빈번할 때 유용하며 이로 인해 실제적 평균 런 길이가 훨씬 줄어들 수 있습니다.

기본값으로 재설정 모든 모수를 기본값으로 재설정합니다.

경보 스크립트 데이터가 특수 원인 검정에 실패했음을 나타내는 스크립트를 작성하고 실행할 수 있습니다. 결과를 파일 또는 로그에 쓰거나 이메일로 보낼 수 있습니다. 검정이 실패한 이유에 대한 설명을 포함하는 옵션이 있습니다.

경보 스크립트가 호출되면 실행된 스크립트와 후속 JSL 스크립트에서 모두 다음 변수를 사용할 수 있습니다.

`qc_col` - 열 이름

`qc_test` - 실패한 검정

`qc_sample` - 표본 번호

팁 : 경보 스크립트를 지정한 후 " 한계를 벗어난 경우 검정 " 옵션을 설정하면 경보 스크립트가 호출됩니다.

사용자 경보 스크립트 작성에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

평균 런 길이 (ARL) 보고서

"평균 런 길이(ARL)" 보고서에는 ARL 값의 테이블과 그래프가 포함됩니다. 지정된 변화에 대한 ARL(평균 런 길이)은 관리이탈 신호가 발생하기 전까지 예상되는 평균 런 수입니다. 예를 들어 ARL(0)은 공정이 관리 상태일 때 오경보 신호가 발생하기 전까지 예상되는 평균 런 수를 나타냅니다. 공정이 관리 상태일 때 변화 크기는 0입니다.

ARL 보고서의 테이블과 그래프를 사용하면 모수의 다양한 설정이 해당 CUSUM 차트의 성능에 어떤 영향을 주는지 살펴볼 수 있습니다. "제어판" 보고서의 h 및 k 모수가 업데이트되면 ARL 보고서도 업데이트됩니다. 이상적인 CUSUM 차트는 ARL(0) 값이 높고 ARL(Δ) 값이 낮습니다. 여기서 Δ는 관심 있는 크기 변화입니다.

" 평균 런 길이 (ARL)" 보고서는 " 제어판 " 보고서의 " 상위 측 " 및 " 하위 측 " 옵션 설정에 따라 달라집니다 . 옵션을 하나만 선택한 경우 ARL 보고서에는 해당하는 단측 CUSUM 차트에 대한 계산이 사용됩니다 . 두 옵션을 모두 선택한 경우 ARL 보고서에는 양측 CUSUM 차트에 대한 계산이 사용됩니다 . 양측 ARL 값은 다음 방정식에 의해 단측 양의 ARL 및 음의 ARL 값과 관련됩니다 .

$$\frac{1}{\text{ARL}} = \frac{1}{\text{양의 ARL}} + \frac{1}{\text{음의 ARL}}$$

ARL 테이블

ARL 테이블에는 0.25씩 증가하는 0 ~ 3 사이의 변화(Δ)에 대한 평균 런 길이가 표시됩니다. "데이터 단위" 옵션이 지정된 경우 변화는 2*K/시그마^{2}으로 나타냅니다. "데이터 단위" 옵션이 지정되지 않은 경우 변화는 2*k/시그마로 나타냅니다.

ARL 그래프

"ARL 그래프"에는 0 ~ 3 사이의 변화(Δ)에 대한 평균 런 길이가 표시됩니다. 이 그래프에는 ARL 그래프 왼쪽에 있는 ARL 테이블과 동일한 데이터 점이 포함됩니다.

CUSUM 관리도의 추가 예

- "CUSUM 관리도의 데이터 단위 옵션 예 "
- " 부분군이 있는 CUSUM 차트의 예 "

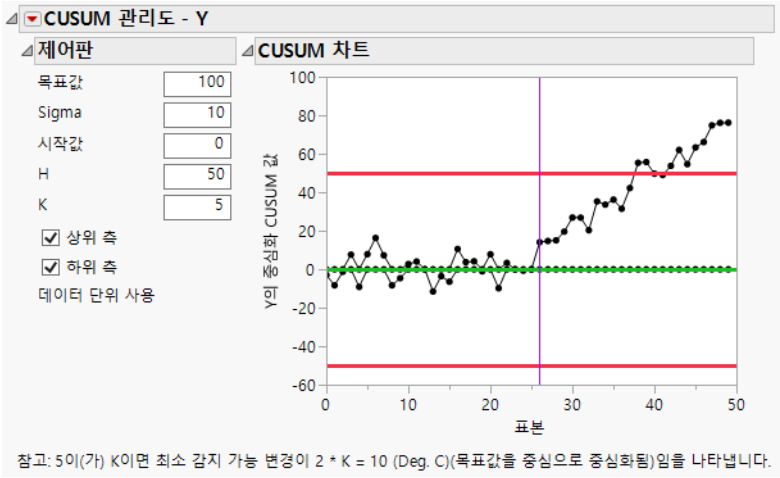
CUSUM 관리도의 데이터 단위 옵션 예

이 예에서는 데이터 단위 옵션을 사용하여 "CUSUM 관리도의 예 " 에 나오는 분석을 재현합니다 . 엔진 온도의 작은 변화를 감지하려고 합니다 . 데이터 테이블에는 엔진 온도 조절 장치에서 수집한 온도 측정값이 포함되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Engine Temperature Sensor.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > CUSUM 관리도**를 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. " 데이터 단위 " 옆의 상자를 선택합니다 .
5. **확인**을 클릭합니다 .
6. " 목표값 " 상자에 100 을 입력합니다 .
7. " 시그마 " 상자에 10 을 입력합니다 .

   " 시작값 " 아래의 옵션이 h 와 k 대신 H 와 K 입니다 . 이제 이 모수는 표준편차 단위가 아니라 데이터 열의 단위로 지정됩니다 .

그림 8.6 CUSUM 관리도 보고서



시그마 단위를 사용한 예와 마찬가지로 CUSUM 차트의 세로선은 온도 측정값 변화가 대략 표본 26에서 시작되었음을 나타냅니다.

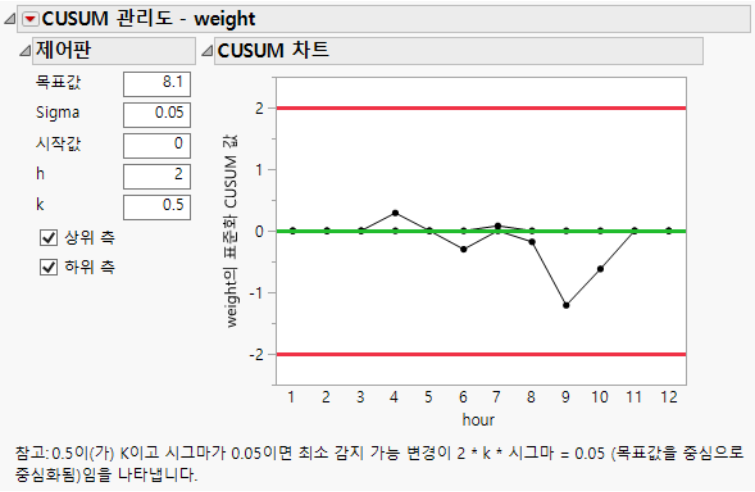
부분군이 있는 CUSUM 차트의 예

기계 하나가 8온스 캔에 2사이클 엔진 오일 첨가제를 주입합니다. 주입 공정은 통계적 관리 상태를 유지한다고 간주됩니다. 가득 주입된 캔의 평균 무게(μ_0)가 8.10 온스가 되도록 공정을 설정합니다. 이전 분석에 따르면 주입 무게의 표준편차(σ_0)는 0.05온스입니다.

4 개 캔이 포함된 부분군 표본을 선택하고 12 시간 동안 1 시간에 한 번씩 무게를 측정했습니다. Oil1 Cusum.jmp 데이터 테이블의 각 관측값은 무게 (weight) 값 하나와 관련 시간 (hour) 값을 포함합니다. 관측값은 시간 값이 작은 것부터 순서대로 정렬됩니다. 공정에서 2σ 변화를 감지하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Oil1 Cusum.jmp 를 엽니다.
2. **분석 > 품질 및 공정 > 관리도 > CUSUM 관리도**를 선택합니다.
3. weight 를 선택하고 **Y** 를 클릭합니다.
4. hour 를 선택하고 **X** 를 클릭합니다.
5. **확인**을 클릭합니다.
6. " 목표값 " 상자에 8.1 을 입력합니다.
   이 값은 공정의 목표 평균입니다.
7. " 시그마 " 상자에 0.05 를 입력합니다.
   이 값은 공정에 대해 알려진 표준편차입니다.
8. h 상자에 2 를 입력합니다.
   이 값은 결정 한계를 각 방향에서 2 표준편차로 정의합니다.

그림 8.7 부분군이 있는 CUSUM 관리도



CUSUM 차트에서 결정 상한 또는 하한을 벗어나는 점이 없습니다. 즉, 공정의 변화가 발생했다는 증거가 없습니다.

참고 : Montgomery(2013) 에서는 " 규모 경제가 유의하거나 , 1 보다 큰 크기의 표본을 추출해야 하는 다른 타당한 이유가 있는 경우에만 1 보다 큰 크기의 부분군을 CUSUM 과 함께 사용해야 한다 " 고 설명합니다 . 테이블 형식 CUSUM 차트에서 합리적인 부분군을 사용한다고 해서 항상 차트 성능이 개선되는 것은 아닙니다 .

CUSUM 관리도 플랫폼에 대한 통계 상세 정보

- "CUSUM 관리도 생성에 대한 통계 상세 정보 "
- " 변화 감지에 대한 통계 상세 정보 "
- " 평균 런 길이에 대한 통계 상세 정보 "

CUSUM 관리도 생성에 대한 통계 상세 정보

이 섹션에서는 CUSUM 차트 생성에 사용되는 통계량을 정의합니다 . 이 중 일부 통계량은 " 요약 저장 " 명령을 사용하여 생성되는 데이터 테이블에도 저장됩니다 .

단측 상위 및 하위 누적합

C^+ 및 C^- 의 정의는 " 데이터 단위 " 옵션 설정에 따라 달라집니다 .

참고 : 요약 저장 데이터 테이블에서 C^+ 와 C^- 에는 각각 상위 누적합과 하위 누적합이라는 라벨이 지정됩니다 .

표준화된 단위의 누적합

" 데이터 단위 " 옵션을 선택하지 않은 경우 각 단계의 C^+ 및 C^- 는 다음과 같이 정의됩니다 .

$$C_i^+ = \max\left(0, \frac{x_i - T}{\sigma} - k + C_{i-1}^+\right)$$

$$C_i^- = \min\left(0, \frac{x_i - T}{\sigma} + k + C_{i-1}^-\right)$$

다음은 각 요소에 대한 설명입니다 .

$x_i = i$ 번째 단계의 반응 값

T = 공정의 목표값

σ = 공정의 표준편차

k = 표준편차 단위로 표현된 참조 값

"시작값"에 값을 지정하는 경우 해당 값이 초기 C^+ 값으로 사용되고, 이 값의 음수 값이 초기 C^- 값으로 사용됩니다. 그렇지 않은 경우 C^+와 C^-의 초기값은 0입니다.

데이터 단위의 누적합

" 데이터 단위 " 옵션을 선택한 경우 각 단계의 C^+ 및 C^- 는 다음과 같이 정의됩니다 .

$$C_i^+ = \max(0, (x_i - T) - K + C_{i-1}^+)$$

$$C_i^- = \min(0, (x_i - T) + K + C_{i-1}^-)$$

다음은 각 요소에 대한 설명입니다 .

$x_i = i$ 번째 단계의 반응 값

T = 공정의 목표값

σ = 공정의 표준편차

K = 데이터 단위로 표현된 참조 값

"시작값"에 값을 지정하는 경우 해당 값이 초기 C^+ 값으로 사용되고, 이 값의 음수 값이 초기 C^- 값으로 사용됩니다. 그렇지 않은 경우 C^+와 C^-의 초기값은 0입니다.

양의 런 및 음의 런 카운터

각 단계에서 N^+는 C^+에 대한 가장 최근 0 값 이후의 단계 수입니다. 각 단계에서 N^-는 C^-에 대한 가장 최근 0 값 이후의 단계 수입니다.

참고 : 요약 저장 데이터 테이블에서 N^+ 와 N^- 에는 각각 양의 런과 음의 런이라는 라벨이 지정됩니다 .

변화 감지에 대한 통계 상세 정보

차트에서 C^+ 값이 상한선 위에 있으면 양의 변화가 발생합니다. 변화 시작은 C^+에 대한 가장 최근 0 값 이후 첫 번째 점으로 정의됩니다.

차트에서 C^- 값이 하한선 아래에 있으면 음의 변화가 발생합니다. 변화 시작은 C^-에 대한 가장 최근 0 값 이후 첫 번째 점으로 정의됩니다.

평균 런 길이에 대한 통계 상세 정보

단측 ARL(평균 런 길이) 값은 Goel and Wu(1971)에 설명된 적분 방정식 방법(24개의 가우스 점)을 사용하여 계산됩니다. "시작값"에 지정된 값이 0보다 큰 경우에는 Lucas and Crosier(1982)의 Appendix A.1에 설명된 방법으로 값이 계산됩니다.

양측 ARL 값은 다음 방정식에 의해 단측 양의 ARL 및 음의 ARL 값과 관련됩니다 .

$$\frac{1}{\text{ARL}} = \frac{1}{\text{양의 ARL}} + \frac{1}{\text{음의 ARL}}$$

Lucas and Crosier(1982) 에서는 초기 CUSUM S_0이 0이 아닌 값으로 설정된 CUSUM 차트에 대한 "시작값" 모수의 특성을 설명합니다. 이 모수는 FIR(Fast Initial Response) 기능이라고도 합니다. 이 차트에서 제공하는 평균 런 길이 계산을 통해 공정이 관리 상태일 때는 FIR 기능이 미치는 효과가 거의 없으며, 결과적으로 초기 관리이탈 상태에 대한 반응이 표준 CUSUM 차트보다 빠르다는 것을 알 수 있습니다.

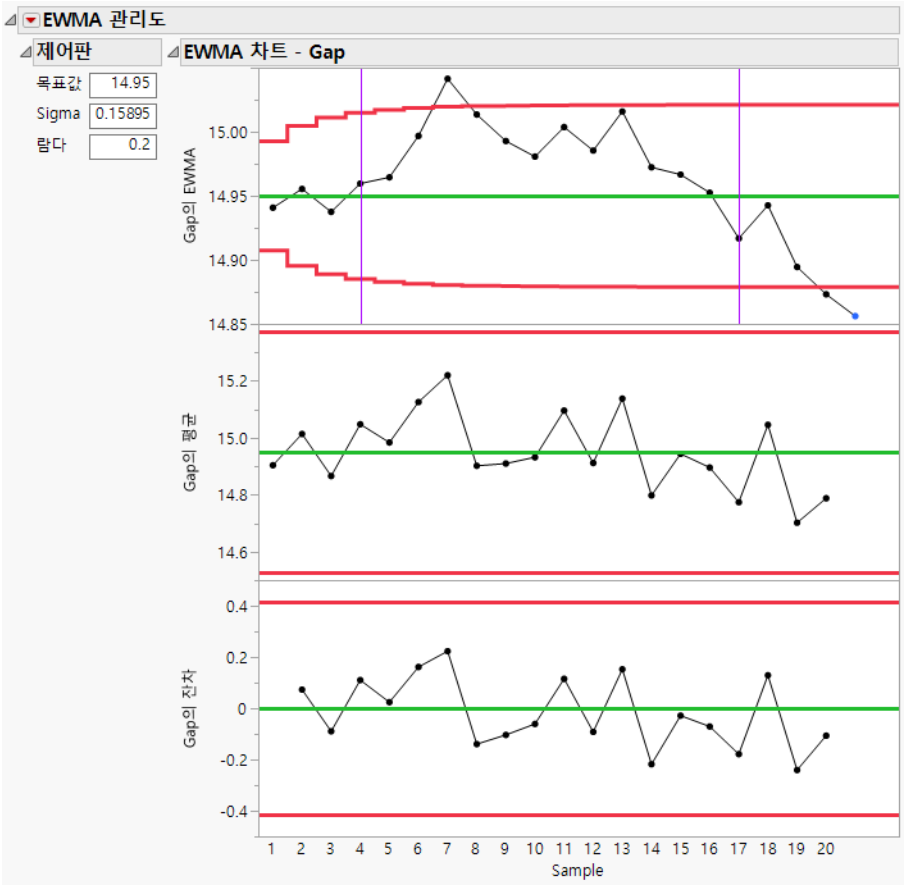
9 장

EWMA 관리도

지수 가중 관리도 생성

EWMA(지수 가중 이동 평균) 차트를 사용하여 공정의 작은 변화를 감지할 수 있습니다. EWMA 차트의 각 점은 현재 부분군 표본의 평균을 포함하여 이전의 모든 부분군 평균에 대한 가중 평균입니다. 가중치는 시간에 따라 지수적으로 감소합니다.

그림 9.1 EWMA 관리도 보고서



목차

EWMA 관리도 플랫폼 개요 247
EWMA 관리도 플랫폼의 예 247
EWMA 관리도 플랫폼 시작 249
EWMA 관리도 플랫폼 보고서 250
EWMA 관리도의 제어판 250
EWMA 차트 보고서 251
EWMA 관리도 플랫폼 옵션 252
EWMA 관리도의 평균 런 길이(ARL) 보고서 255
EWMA 관리도 플랫폼의 추가 예 256
EWMA 관리도 플랫폼에 대한 통계 상세 정보 259

EWMA 관리도 플랫폼 개요

EWMA(지수 가중 이동 평균) 차트를 사용하여 공정의 작은 변화를 감지할 수 있습니다. EWMA 차트의 각 점은 현재 부분군 표본의 평균을 포함하여 이전의 모든 부분군 평균에 대한 가중 평균입니다. 가중치는 시간에 따라 지수적으로 감소합니다. 지수 가중 이동 평균 차트에 대한 자세한 내용은 Box et al. (2009) 및 Montgomery(2013) 연구 자료에서 확인하십시오.

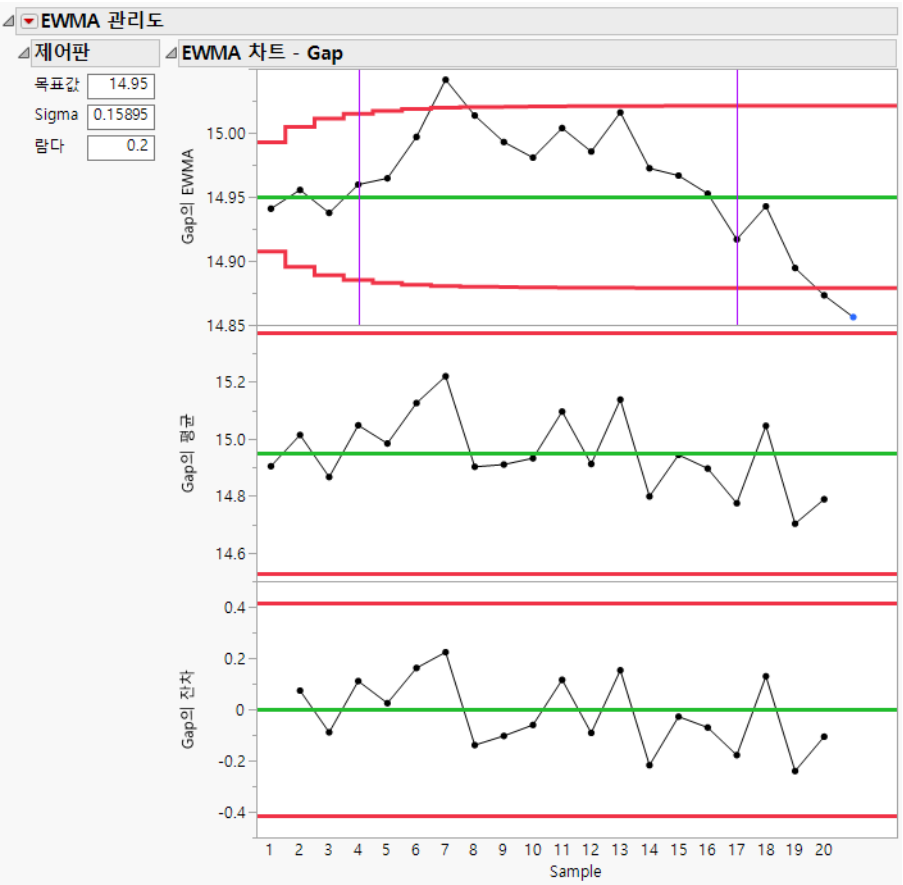
EWMA 관리도 플랫폼은 EWMA 차트와 X 차트 및 잔차 차트를 쌍으로 구성합니다. 부분군 변수를 지정하지 않은 경우 X 차트는 개별 측정값을 보여 줍니다. 부분군 변수를 지정하고 하나 이상의 부분군 크기가 1보다 큰 경우 X 차트는 XBar 차트입니다.

EWMA 관리도 플랫폼의 예

Clips1.jmp 샘플 데이터 테이블에서 관심 측도는 제조된 금속 클립의 끝부분 사이 간격입니다 . 공정에서 평균 간격의 변화를 모니터링하기 위해 클립 5 개로 구성된 부분군 표본을 매일 측정했습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Clips1.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > EWMA 관리도**를 선택합니다 .
3. Gap 을 선택하고 **Y** 를 클릭합니다 .
4. Sample 을 선택하고 **부분군**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 9.2 EWMA 관리도

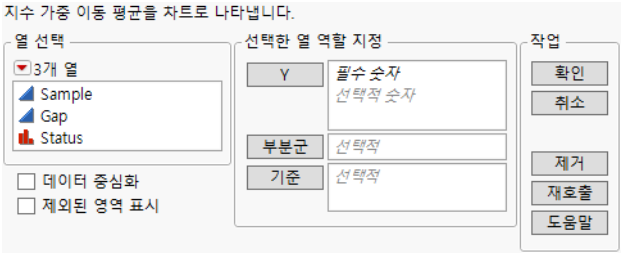


EWMA 차트의 보라색 세로선은 변화를 나타냅니다. 표본 4와 17에서 변화 시작이 감지되었습니다.

EWMA 관리도 플랫폼 시작

분석 > 품질 및 공정 > 관리도 > EWMA 관리도를 선택하여 EWMA 관리도 플랫폼을 시작할 수 있습니다 .

그림 9.3 EWMA 관리도 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

EWMA 관리도 플랫폼 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y 차트를 생성할 변수를 식별합니다 .

> **참고 :** 부분군 변수를 지정하지 않은 경우 데이터 테이블의 행은 관측값을 수집한 순서대로 정렬되어야 합니다 .

부분군 부분군 변수를 식별합니다. EWMA 차트의 가로 축은 부분군 변수로 라벨이 지정됩니다.

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

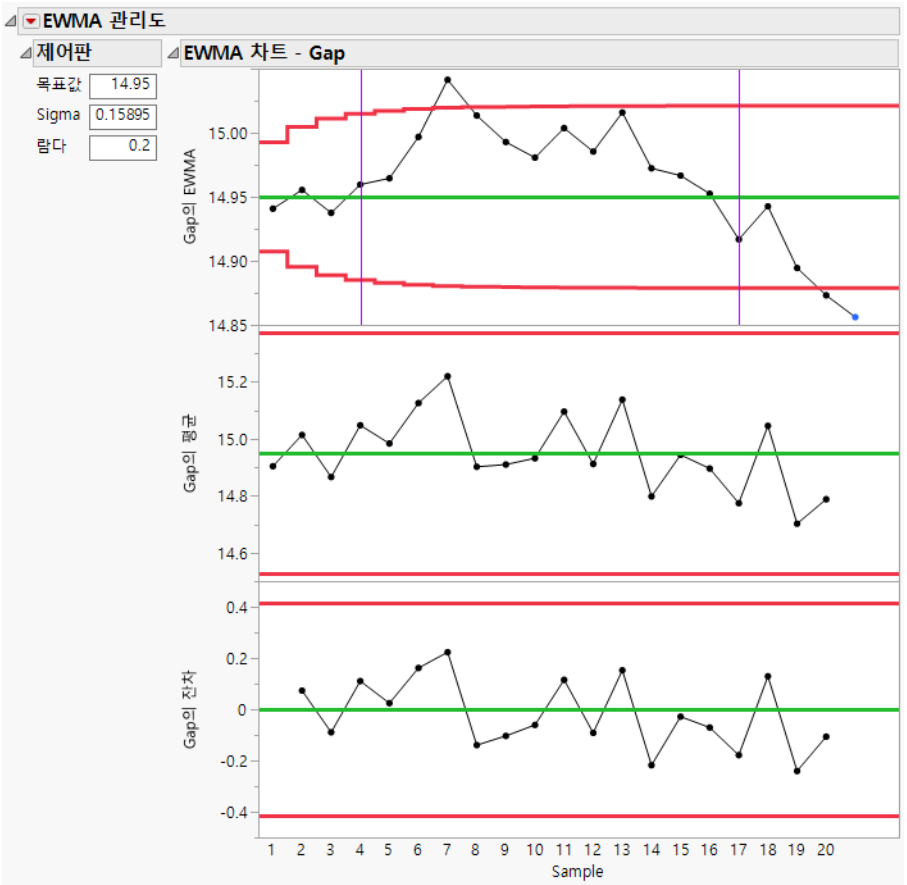
데이터 중심화 각 관측값에서 목표값을 빼서 데이터를 중심화하도록 지정합니다 .

제외된 영역 표시 (부분군 변수가 지정된 경우에만 적용 가능) 데이터 테이블에서 완전히 제외된 부분군이 EWMA 관리도에 표시되도록 지정합니다 .

EWMA 관리도 플랫폼 보고서

기본적으로 EWMA 관리도 플랫폼에서는 모수 제어판 , EWMA 차트 , X 차트 및 잔차 차트가 포함된 보고서를 생성합니다 .

그림 9.4 EWMA 관리도 보고서



EWMA 관리도의 제어판

"제어판"에는 차트 모수의 현재 값이 포함됩니다. 현재 값은 모수 값을 업데이트할 수 있는 상자에 있습니다.

제어판에는 다음 옵션이 표시됩니다 .

목표값 알려진 평균 값입니다 . 이 값은 차트의 중심선 값입니다 . 기본적으로 이 모수는 Y 열의 " 규격 한계 " 열 특성에 지정된 " 목표값 " 으로 설정됩니다 . Y 열의 " 규격 한계 " 열 특성에 목표값이 지정되지 않은 경우 이 모수는 Y 열의 전체 평균으로 설정됩니다 .

다음과 같은 경우 목표값을 편집할 수 없습니다 .

- 시작 창에서 " 데이터 중심화 " 옵션을 지정한 경우
- *Y* 변수의 " 규격 한계 " 열 특성에 목표값이 포함된 경우 . 그러나 빨간색 삼각형 메뉴에서 " 목표값에 전체 평균 사용 " 옵션을 선택하여 " 규격 한계 " 열 특성의 목표값과 *Y* 변수의 전체 평균 간에 전환할 수 있습니다 .

시그마 알려진 표준편차 값입니다 . 기본적으로 이 모수는 Y 열의 평균 이동 범위로 설정됩니다 . 부분군 변수가 있는 경우 시그마 모수는 부분군 평균의 이동 범위 평균으로 설정됩니다 .

참고 : 기본 시그마 값 계산에는 제외된 행이 포함되지 않습니다 .

람다 사전 표본에 가중치를 지정하기 위한 평활 상수 값입니다 . 기본적으로 이 모수는 0.2 로 설정됩니다 .

*** 중심화된 데이터 사용** 이 텍스트가 있으면 시작 창에서 " 데이터 중심화 " 옵션이 선택되었고 EWMA 차트와 X 차트의 값이 목표값을 기준으로 중심화되었음을 나타냅니다 .

EWMA 차트 보고서

"EWMA 차트 " 보고서에는 EWMA 차트 , X 차트 및 잔차 차트라는 세 가지 차트가 포함됩니다 . 이 차트의 해석에 대한 자세한 내용은 Box et al. 연구 자료 (2009) 에서 확인하십시오 .

EWMA 차트

EWMA 차트는 차트 모수의 현재 값에 의해 지정된 결정 한계와 함께 EWMA(지수 가중 이동 평균) 를 보여 줍니다 .

- 부분군 변수를 지정하고 하나 이상의 부분군 크기가 1 보다 큰 경우 가로 축은 부분군을 나타냅니다 .
- 그렇지 않은 경우 가로 축은 표본을 나타냅니다 .

두 경우 모두 세로 축은 지수 가중 이동 평균을 나타냅니다. 시작 창에서 "데이터 중심화" 옵션을 지정한 경우 세로 축은 목표값을 뺀 지수 가중 이동 평균을 나타냅니다. 각 표본 또는 부분군의 단일 점이 차트에 표시됩니다. 예측 점을 나타내는 파란색 점도 하나 더 있습니다.

참고 : 마지막 표본 또는 부분군이 숨겨지고 제외된 경우 마지막 표본 또는 부분군을 예측 점에 연결하는 선이 표시되지 않습니다 .

X 차트

X 차트는 관측값의 Shewhart 관리도입니다 .

- 부분군 변수를 지정하고 하나 이상의 부분군 크기가 1 보다 큰 경우 X 차트는 평균 값의 XBar 차트입니다 . 가로 축은 부분군을 나타냅니다 . 세로 축은 부분군 평균을 나타냅니다 . 시작 창에서 " 데이터 중심화 " 옵션을 선택한 경우 세로 축은 목표값을 뺀 부분군 평균을 나타냅니다 . 각 부분군의 단일 점이 차트에 표시됩니다 .
- 그렇지 않은 경우 X 차트는 개별 측정값을 보여 줍니다 . 가로 축은 표본을 나타냅니다 . 세로 축은 측정값을 나타냅니다 . 시작 창에서 " 데이터 중심화 " 옵션을 선택한 경우 세로 축은 목표값을 뺀 측정값을 나타냅니다 . 각 표본의 단일 점이 차트에 표시됩니다 .

X 차트의 한계에 대한 자세한 내용은 "관리도 빌더에 대한 통계 상세 정보"에서 확인하십시오.

잔차 차트

잔차 차트를 사용하면 자기상관을 시각적으로 확인할 수 있습니다 .

- 부분군 변수를 지정하고 하나 이상의 부분군 크기가 1 보다 큰 경우 잔차 차트는 각 부분군 평균과 이전 부분군의 EWMA 값 간의 차이를 보여 줍니다 . *i* 번째 잔차는 $r_i = X_i - EWMA_{i-1}$ 로 계산되며 , 여기서 X_i 는 *i* 번째 부분군 평균 , $EWMA_{i-1}$ 은 (*i*-1) 번째 EWMA 값을 각각 나타냅니다 .
- 그렇지 않은 경우 잔차 차트는 각 표본 값과 이전 표본의 EWMA 값 간의 차이를 보여 줍니다 . *i* 번째 잔차는 $r_i = X_i - EWMA_{i-1}$ 로 계산되며 , 여기서 X_i 는 *i* 번째 표본 값 , $EWMA_{i-1}$ 은 (*i*-1) 번째 EWMA 값을 각각 나타냅니다 .

잔차 차트의 한계는 ±3**ResidSigma* 이며 , 여기서 *ResidSigma* 는 잔차의 표준편차입니다 .

EWMA 관리도 플랫폼 옵션

"EWMA 관리도 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

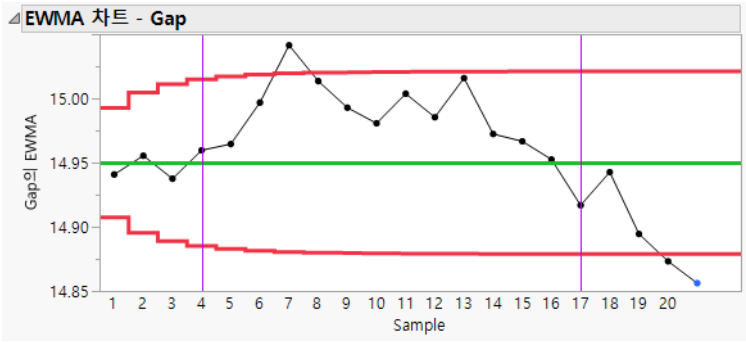
한계 표시 EWMA 차트 , X 차트 및 잔차 차트에 결정 상한 및 하한을 표시하거나 숨깁니다 .

중심선 표시 EWMA 차트 , X 차트 및 잔차 차트에 중심선을 표시하거나 숨깁니다 .

변화 선 표시 EWMA 차트에 변화를 지정하는 세로선을 표시하거나 숨깁니다 . 변화 선은 변화가 시작되는 지점에 그립니다 . 변화 시작은 EWMA 값이 특정 방향으로 중심선과 교차한 이후 첫 번째 점으로 정의됩니다 .

- 차트에서 EWMA 값이 상한선 위에 있으면 양의 변화가 발생합니다 . 변화 시작은 목표값 선 아래에 있는 가장 최근 EWMA 값 이후 첫 번째 점으로 정의됩니다 .
- 차트에서 EWMA 값이 하한선 아래에 있으면 음의 변화가 발생합니다 . 변화 시작은 목표값 선 위에 있는 가장 최근 EWMA 값 이후 첫 번째 점으로 정의됩니다 .

그림 9.5 변화 선의 예



이 예에서는 표본 7 의 점이 상한선 위에 있습니다 . 표본 7 이전을 살펴보면 목표값 선 아래에 있는 가장 최근 점이 표본 3 입니다 . 따라서 표본 4 가 목표값 선 아래에 있는 가장 최근 점 이후 목표값 위의 첫 번째 점이며 양의 변화가 시작되는 지점으로 표시됩니다 .

마찬가지로 표본 20 의 점이 하한선 아래에 있고 목표값 선 위에 있는 가장 최근 점은 표본 16 입니다 . 따라서 표본 17 이 음의 변화가 시작되는 지점으로 표시됩니다 .

한계를 벗어난 경우 검정 EWMA 차트와 X 차트에서 상한선 위에 있거나 하한선 아래에 있는 모든 점 주위에 빨간색 원을 표시하거나 숨깁니다 .

ARL 표시 평균 런 길이 (ARL) 보고서를 표시하거나 숨깁니다 . 자세한 내용은 "EWMA 관리도의 평균 런 길이 (ARL) 보고서 " 에서 확인하십시오 .

제어판 모수의 현재 값을 보여 주는 보고서를 표시하거나 숨깁니다 . 이 보고서를 사용하면 EWMA 차트의 모수 값을 변경할 수 있습니다 .

모수 보고서 모수의 현재 값을 보여 주는 보고서를 표시하거나 숨깁니다 .

상수 한계 EWMA 차트의 한계가 일정하도록 점근적 표현식을 사용하여 EWMA 차트 한계를 계산하도록 지정합니다 .

주의 : 부분군의 표본 크기가 동일하지 않으면 " 상수 한계 " 옵션을 사용해도 아무 효과가 없습니다 .

요약 저장 EWMA 차트의 각 부분군에 대한 통계량을 포함하는 새 데이터 테이블을 생성합니다 . 부분군 수 , 부분군 라벨 , 부분군 크기 , 부분군 평균 , 변화 시작 표시자 , 변화 시작 사이의 각 간격을 나타내는 값 , 각 부분군의 지수 가중 이동 평균 , 연속된 양의 런 / 음의 런 수 , LCL/UCL 값 및 검정 오류 표시자 등의 통계량이 새 데이터 테이블에 저장됩니다 . 예측 값은 요약 테이블의 마지막 행에 저장됩니다 .

기본값으로 재설정 모든 모수를 기본값으로 재설정합니다 . " 람다 " 플랫폼 환경 설정에 람다 값이 지정된 경우 플랫폼 환경 설정에 지정된 값으로 재설정됩니다 .

참고 : *Y* 변수의 " 규격 한계 " 열 특성에 목표값이 포함된 경우 " 기본값으로 재설정 " 옵션은 목표값을 " 규격 한계 " 열 특성에 지정된 값으로 설정합니다 .

빈 부분군 다음에 EWMA 다시 시작 빈 부분군이 있을 때 이동 평균 및 한계 계산을 처리하는 방법을 지정합니다. 부분군의 모든 관측값이 결측이거나 제외된 행에 포함된 경우 부분군이 비어 있을 수 있습니다 . 이 옵션을 선택하면 빈 부분군 다음에 나오는 첫 번째 비결측 부분군에서 계산이 다시 시작됩니다 . 계산을 다시 시작하면 이동 평균이 전체 평균으로 재설정됩니다 . 이 옵션을 선택하지 않으면 가장 최근의 비결측 부분군 이동 평균으로 EWMA 계산이 계속됩니다 .

목표값에 전체 평균 사용 (*Y* 변수의 " 규격 한계 " 열 특성에 목표값이 포함된 경우에만 사용 가능) EWMA 차트의 목표값을 *Y* 변수의 전체 평균으로 설정합니다. 이 옵션을 선택하지 않으면 " 규격 한계 " 열 특성에 지정된 목표값이 EWMA 차트의 목표값으로 설정됩니다 .

중첩 차트 개별 위치 값을 EWMA 차트에 중첩할지 여부를 지정합니다 . 이 옵션을 선택하면 위치 차트가 더 이상 표시되지 않습니다 . 대신 위치 차트의 점은 연결되지 않은 회색 X 로 EWMA 차트에 나타납니다 . " 중첩 차트에 X 한계 표시 " 옵션을 선택하지 않으면 위치 차트의 한계가 EWMA 차트에 나타나지 않습니다 .

중첩 차트에 X 한계 표시 (" 중첩 차트 " 옵션을 선택한 경우에만 사용 가능) " 중첩 차트 " 옵션을 선택한 경우 EWMA 차트에 위치 차트의 한계를 표시하거나 숨깁니다 . 위치 차트 한계는 EWMA 차트에 파선으로 표시됩니다 .

참고 : " 한계를 벗어난 경우 검정 " 옵션은 " 중첩 차트에 X 한계 표시 " 옵션을 선택한 경우에만 중첩 차트의 위치 차트 값에 적용됩니다 .

람다 슬라이더 람다 모수의 값을 대화식으로 변경할 수 있는 슬라이더 컨트롤을 표시하거나 숨깁니다 .

X 차트 표시 EWMA 차트 아래에 위치 차트를 표시하거나 숨깁니다 .

참고 : " 중첩 차트 " 옵션이 선택된 경우 "X 차트 표시 " 옵션은 EWMA 차트에 중첩된 개별 위치 값과 한계를 표시하거나 숨깁니다 .

잔차 차트 표시 잔차 차트를 표시하거나 숨깁니다 .

경보 스크립트 데이터가 특수 원인 검정에 실패했음을 나타내는 스크립트를 작성하고 실행할 수 있습니다 . 결과를 파일 또는 로그에 쓰거나 이메일로 보낼 수 있습니다 . 검정이 실패한 이유에 대한 설명을 포함하는 옵션이 있습니다 .

경보 스크립트가 호출되면 실행된 스크립트와 후속 JSL 스크립트에서 모두 다음 변수를 사용할 수 있습니다 .

`qc_col` - 열 이름

`qc_test` - 실패한 검정

`qc_sample` - 표본 번호

팁 : 경보 스크립트를 지정한 후 " 한계를 벗어난 경우 검정 " 옵션을 설정하면 경보 스크립트가 호출됩니다 .

사용자 경보 스크립트 작성에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

EWMA 관리도의 평균 런 길이 (ARL) 보고서

"평균 런 길이 (ARL)" 보고서에는 ARL 값의 테이블과 그래프가 포함됩니다. 지정된 변화에 대한 ARL(평균 런 길이) 은 관리이탈 신호가 발생하기 전에 예상되는 평균 런 수입니다. 예를 들어 ARL(0) 은 공정이 관리 상태일 때 오경보 신호가 발생하기 전까지 예상되는 평균 런 수를 나타냅니다. 공정이 관리 상태일 때 변화 크기는 0 입니다.

ARL 보고서의 테이블과 그래프를 사용하면 람다 모수의 다양한 설정이 해당 EWMA 차트의 성능에 어떤 영향을 주는지 살펴볼 수 있습니다. 또한 테이블과 그래프를 통해 EWMA 차트 성능을 Shewhart 차트 (예 : EWMA 차트 보고서의 X 차트) 와 비교할 수도 있습니다. Shewhart ARL 열은 람다가 1 로 설정된 경우 EWMA ARL 열과 동일합니다.

EWMA 의 ARL(0) 값은 " 상수 한계 " 옵션 설정에 따라 달라집니다 .

- " 상수 한계 " 옵션을 선택하면 공정이 시작 값의 영향을 무시할 수 있을 만큼 오랫동안 관리 상태를 유지하고 있다고 가정합니다 . 이 경우를 정상 상태의 EWMA 차트라고도 하며 Crowder(1987) 에 설명된 방법을 사용하여 EWMA ARL(0) 값을 계산합니다 .
- " 상수 한계 " 옵션을 선택하지 않으면 Knoth(2004) 에 설명된 방법을 사용하여 EWMA ARL(0) 값을 계산합니다 . 이 경우를 0 상태의 EWMA 차트라고도 합니다 .

ARL 보고서

"ARL 보고서 " 에는 0.25 씩 증가하는 0 ~ 3 사이의 표준화된 변화 (Δ) 에 대한 평균 런 길이가 표시됩니다. 변화는 sqrt(*n*)*RawShift/Sigma 로 나타내며 , 여기서 *n* 은 각 부분군의 상수 표본 크기입니다. 관리 한계를 계산하는 데 사용되는 승수의 경우 K = 3 이 가정됩니다. 이 테이블에는 EWMA 차트 및 Shewhart 차트에 대한 ARL 값이 포함됩니다.

ARL 그래프

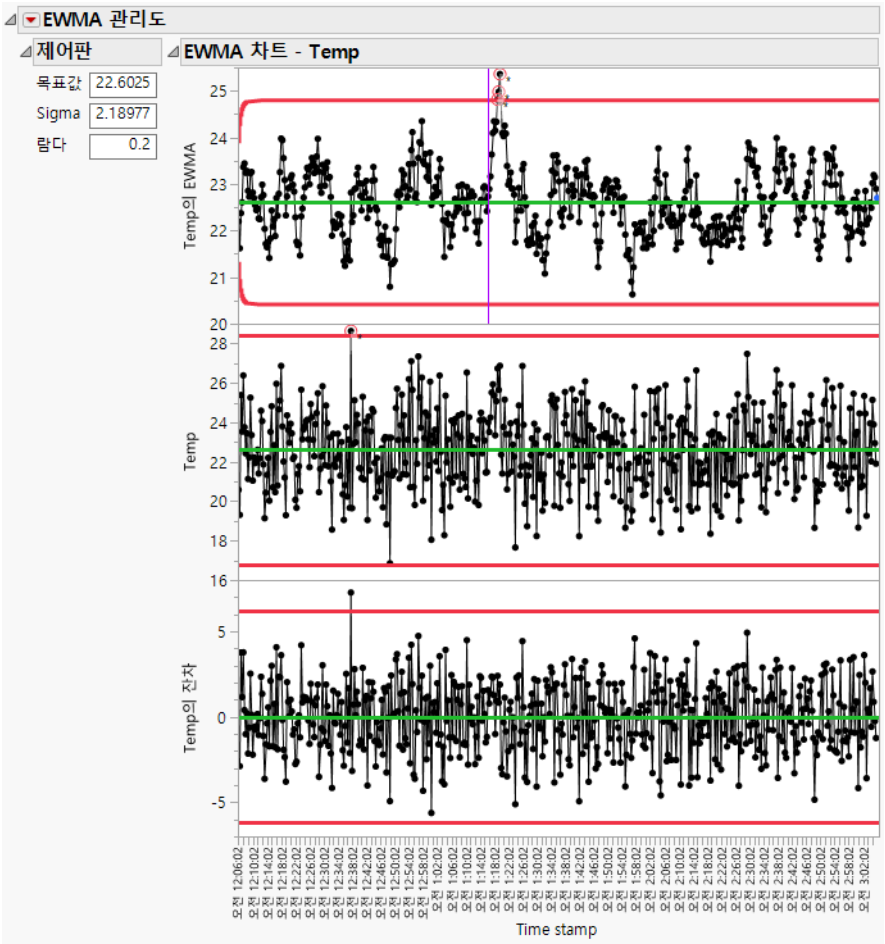
"ARL 그래프 " 에는 0 ~ 3 사이의 표준화된 변화 (Δ) 에 대한 평균 런 길이가 표시됩니다. 이 그래프에는 ARL 그래프 왼쪽에 있는 ARL 테이블과 동일한 데이터 점이 포함됩니다. 실선은 EWMA ARL 값에 해당하고 파선은 Shewhart ARL 값에 해당합니다.

EWMA 관리도 플랫폼의 추가 예

이 예에서는 EWMA 관리도에 부분군 점과 한계를 중첩하는 차트를 생성합니다. 이렇게 하면 부분군 점의 컨텍스트에서 EWMA 점을 볼 수 있습니다 . Quench.jmp 샘플 데이터 테이블에는 자기 냉각 시 온도에 대해 시뮬레이션된 관측값이 포함되어 있습니다 . 초전도 코일의 일부가 저항 상태가 되고 온도가 상승하면 자기 냉각이 발생합니다 . 이 경우 핵융합로 격납 용기가 위험에 노출될 수 있습니다 . 누출을 방지하려면 즉시 자석 전원 공급을 차단해야 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Quench.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > EWMA 관리도**를 선택합니다 .
3. Temp 를 선택하고 **Y** 를 클릭합니다 .
4. Time stamp 를 선택하고 **부분군**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "EWMA 관리도 " 의 빨간색 삼각형을 클릭하고 **한계를 벗어난 경우 검정**을 선택합니다 .

그림 9.6 한계를 벗어난 점이 강조 표시된 EWMA 차트



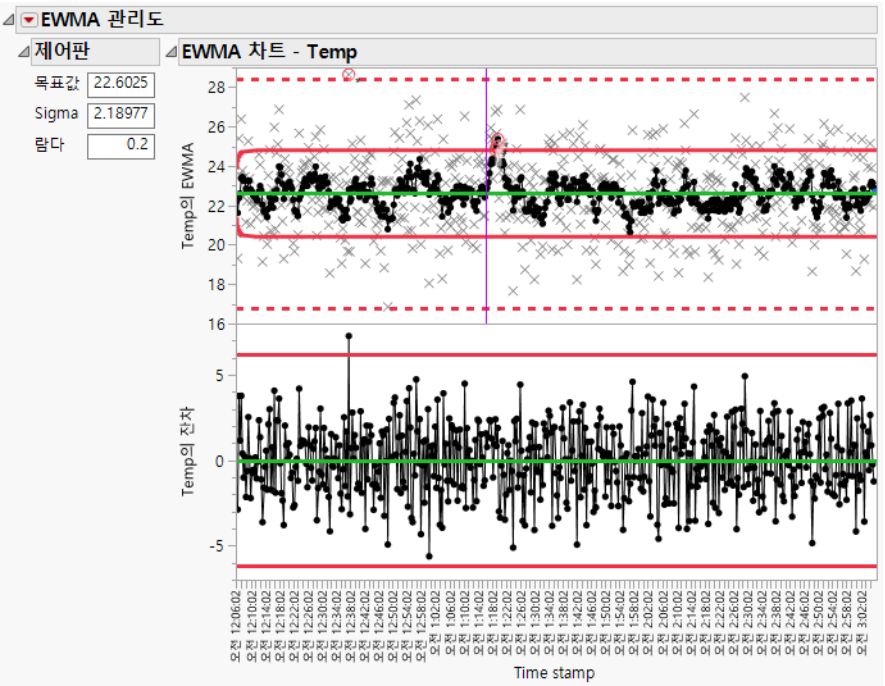
EWMA 차트에서 세 개의 관측값이 한계를 벗어나고 X 차트에서는 한 개의 관측값이 한계를 벗어납니다. 한계를 벗어난 관측값은 빨간색 원으로 표시되어 있습니다. EWMA 차트의 보라색 세로선은 대략 오전 1 시 16 분 정도에 발생한 상향 변화를 나타냅니다.

7. "EWMA 관리도 " 의 빨간색 삼각형을 클릭하고 **중첩 차트**를 선택합니다.

   기본적으로 X 차트 한계는 중첩 차트에 표시되지 않으므로 X 한계를 벗어나는 관측값을 원으로 표시하려면 중첩 차트에 X 한계를 표시하도록 설정해야 합니다.

8. "EWMA 관리도 " 의 빨간색 삼각형을 클릭하고 **중첩 차트에 X 한계 표시**를 선택합니다.

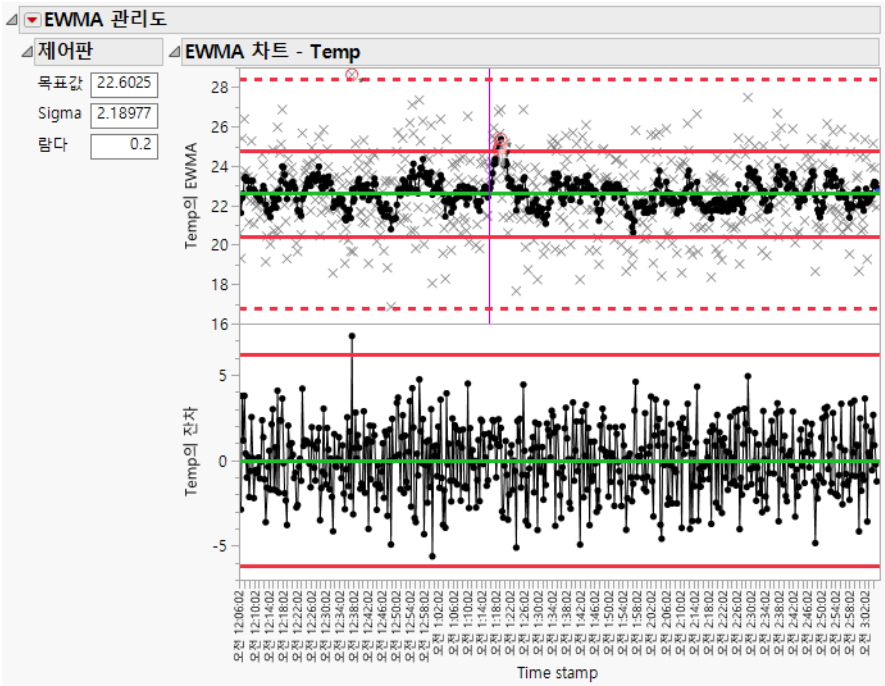
그림 9.7 X 차트가 중첩된 EWMA 차트



위쪽 차트의 실선과 원은 EWMA 차트의 한계와 점을 나타냅니다. 파선과 x 표식은 X 차트의 한계와 점을 나타냅니다 . EWMA 한계 및 X 한계를 벗어나는 관측값은 원으로 표시되어 있습니다 .

9. "EWMA 관리도 " 의 빨간색 삼각형을 클릭하고 **상수 한계**를 선택합니다 .

그림 9.8 상수 한계를 사용한 EWMA 차트



데이터의 시간 범위가 너무 길어서 점근적 한계가 처음 몇 개 부분군에 대해서만 원래 EWMA 한계와 다릅니다 . 이 경우 EWMA 한계 대신 상수 (점근적) 한계를 사용할 수 있습니다 .

EWMA 관리도 플랫폼에 대한 통계 상세 정보

이 섹션에서는 EWMA 차트 생성에 사용되는 통계량을 정의합니다. 이 중 일부 통계량은 "요약 저장" 명령을 사용하여 생성되는 데이터 테이블에도 저장됩니다.

EWMA 차트의 i 번째 점은 다음과 같이 계산됩니다 .

$$EWMA_i = \lambda X_i + (1 - \lambda) EWMA_{i-1}$$

다음은 각 요소에 대한 설명입니다 .

λ = 람다 모수

X_i = i 번째 표본 값 또는 부분군 평균

$EWMA_{i-1}$ = (i-1) 번째 EWMA 값

$i = 1$ 이면 $EWMA_0$을 목표값으로 정의합니다.

참고 : " 빈 부분군 다음에 EWMA 다시 시작 " 옵션을 선택한 경우 빈 부분군 다음에 나오는 $EWMA_{i-1}$ 값은 목표값입니다. "빈 부분군 다음에 EWMA 다시 시작" 옵션을 선택하지 않은 경우 빈 부분군 다음에 나오는 $EWMA_{i-1}$ 값은 비어 있지 않은 가장 최근 부분군에 대한 *EWMA* 값입니다 .

EWMA 차트의 관리 한계 계산은 "상수 한계" 옵션 설정에 의해 결정됩니다.

EWMA 한계

" 상수 한계 " 옵션을 선택하지 않은 경우 부분군 크기가 같지 않으면 EWMA 관리 한계가 다음과 같이 계산됩니다 .

$$\text{LCL} = T - K\sigma\lambda\sqrt{\sum_{j=1}^{i}\frac{(1-\lambda)^{2(i-j)}}{n_j}}$$

$$\text{UCL} = T + K\sigma\lambda\sqrt{\sum_{j=1}^{i}\frac{(1-\lambda)^{2(i-j)}}{n_j}}$$

다음은 각 요소에 대한 설명입니다 .

T = 목표값

K = 시그마 승수 (기본값 3)

σ = 시그마 값

i = 표본 또는 부분군 수

n_i = 부분군 i 의 크기

" 상수 한계 " 옵션을 선택하지 않은 경우 부분군 크기가 같으면 EWMA 관리 한계 계산식이 다음과 같이 단순화됩니다 .

$$\text{LCL} = T - K\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}[1-(1-\lambda)^{2i}]}$$

$$\text{UCL} = T + K\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}[1-(1-\lambda)^{2i}]}$$

다음은 각 요소에 대한 설명입니다 .

T = 목표값

K = 시그마 승수 (기본값 3)

σ = 시그마 값

i = 표본 또는 부분군 수

n = 각 부분군의 크기 (부분군이 지정되지 않은 경우 1)

상수 한계

" 상수 한계 " 옵션을 선택한 경우 EWMA 관리 한계가 다음과 같이 계산됩니다 .

$$\text{LCL} = T - K\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}}$$

$$\text{UCL} = T + K\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}}$$

다음은 각 요소에 대한 설명입니다 .

T = 목표값

K = 시그마 승수 (기본값 3)

σ = 시그마 값

n = 부분군 크기 (부분군이 지정되지 않은 경우 1)

지수 가중 이동 평균 차트의 생성에 대한 자세한 내용은 Montgomery 연구 자료 (2013) 에서 확인하십시오 .

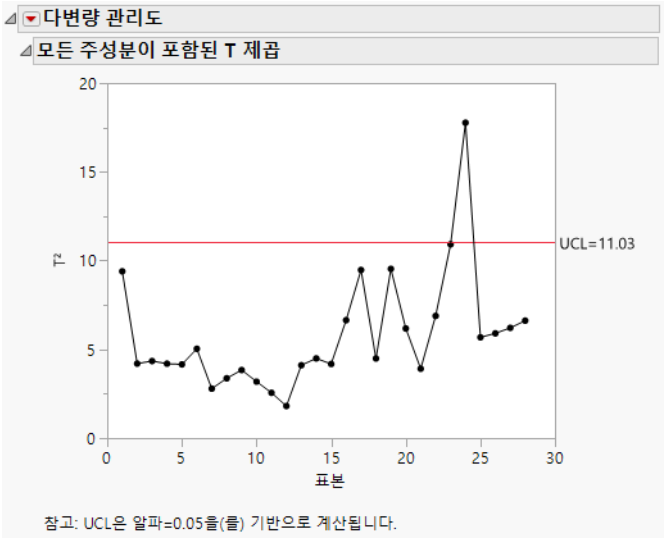
10 장

다변량 관리도
여러 공정 특성 동시 모니터링

다변량 관리도는 서로 관련된 두 개 이상의 공정 변수를 모니터링하는 데 사용됩니다. 단변량 관리도를 사용하여 단일 독립 공정 특성을 모니터링하는 경우 공정 변수 간에 상관관계가 있으면 다변량 관리도가 필요합니다. 다변량 관리도 플랫폼을 사용하면 Hotelling T^2 차트를 생성할 수 있습니다. 공정의 안정성 여부를 판별하는 것은 물론 새 데이터를 수집할 때 공정을 모니터링하려면 이 플랫폼을 사용할 수 있습니다.

복합 공정의 모니터링 및 진단에 대한 자세한 내용은 "모형 기반 다변량 관리도"에서 확인하십시오.

그림 10.1 다변량 관리도의 예



목차

다변량 관리도 개요 265
다변량 관리도의 예 265
다변량 관리도 플랫폼 시작 268
다변량 관리도 269
다변량 관리도 플랫폼 옵션 270
분할된 T 제곱 271
변화 지점 감지 271
주성분 272
다변량 관리도의 추가 예 273
부분군으로 묶은 데이터를 사용한 공정 모니터링의 예 273
분할된 T 제곱의 예 276
변화 지점 감지의 예 278
다변량 관리도에 대한 통계 상세 정보 280
개별 관측값에 대한 통계 상세 정보 280
합리적 부분군의 관측값에 대한 통계 상세 정보 281
변화 지점 감지에 대한 통계 상세 정보 284

다변량 관리도 개요

다변량 관리도는 서로 관련된 두 개 이상의 공정 변수를 모니터링하는 데 사용됩니다. 단변량 관리도를 사용하여 단일 독립 공정 특성을 모니터링하는 경우 공정 변수 간에 상관관계가 있으면 다변량 관리도가 필요합니다. Hotelling T^2 차트(줄여서 T^2 차트)는 다변량 관리도의 한 유형입니다. T^2 차트는 서로 관련된 여러 변수 간의 관계 또는 평균 변화를 감지할 수 있습니다. 관측값은 공정 변수의 개별 관측값이거나 합리적 부분군으로 그룹화할 수 있습니다.

현재 또는 과거 데이터를 사용하여 다변량 관리도를 생성할 수 있습니다. 현재 데이터를 사용하여 생성되는 경우 단계 I 차트라고 하고, 과거 데이터 집합의 목표 통계량을 사용하여 생성되는 경우 단계 II 차트라고 합니다. 단계 I에서는 공정이 안정적인지 확인하고 공정에 대한 목표 통계량을 계산할 과거 데이터 집합을 구성합니다. 단계 II에서는 다변량 관리도에 단계 I의 목표 통계량을 사용하여 새 공정 관측값을 모니터링합니다.

단계 II 다변량 관리도를 생성하려면 먼저 공정이 안정적이고 공정 능력이 있는 기간을 식별합니다.

1. 이 기간 동안 공정이 안정적인지 확인하기 위해 단계 I 관리도를 만듭니다.

   단계 I 에 사용된 데이터는 과거 데이터 집합을 제공합니다.
2. 이 과거 데이터 집합에 대한 목표 통계량을 저장합니다.
3. 2 단계에서 저장한 목표 통계량을 기반으로 하는 단계 II 관리도를 사용하여 진행 중인 공정을 모니터링합니다.

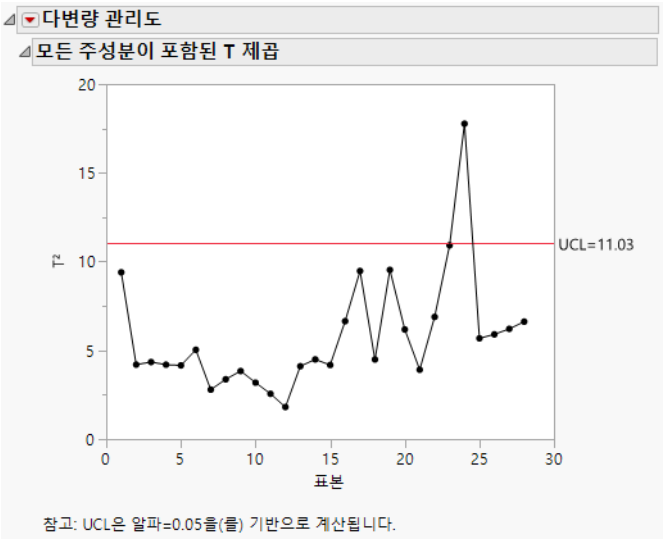
다변량 관리도의 예

이 예에서는 부분군으로 묶지 않은 데이터에 대한 관리도를 생성하는 것을 보여 줍니다. 데이터는 증기 터빈 엔진에서 측정한 값입니다. 부분군으로 묶은 데이터를 사용하는 예는 "부분군으로 묶은 데이터를 사용한 공정 모니터링의 예 " 에서 확인하십시오.

1 단계 : 공정의 안정성 여부 확인

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Steam Turbine Historical.jmp 를 엽니다.
2. **분석 > 품질 및 공정 > 관리도 > 다변량 관리도**를 선택합니다.
3. 열을 모두 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.

그림 10.2 초기 다변량 관리도



관리이탈 점이 하나뿐이므로 공정이 합리적인 통계적 관리 상태에 있는 것으로 보입니다. 따라서 이 데이터를 기반으로 목표값을 생성하는 것이 좋습니다.

2 단계 : 목표 통계량 저장

1. " 다변량 관리도 " 옆의 빨간색 삼각형을 클릭하고 **목표 통계량 저장**을 선택합니다 .
   이렇게 하면 공정에 대한 목표 통계량이 포함된 새 데이터 테이블이 생성됩니다 .

그림 10.3 증기 터빈 데이터에 대한 목표 통계량

	Ref_Stats	Fuel	Steam Flow	Steam Temp	MW	Cool Temp	Pressure
1	_SampleSize	28	28	28	28	28	28
2	_NumSample	1	1	1	1	1	1
3	_Mean	237595.78571	179015.78571	846.39285714	20.647142857	53.871428571	29.139285714
4	_Std	7247.6859825	4374.3063819	2.9481857034	0.5341650261	0.2088010623	0.0497347461
5	_Corr_Fuel	1	0.8714382899	-0.549875041	0.8558570808	-0.270049819	-0.469928462
6	_Corr_Steam Flow	0.8714382899	1	-0.629023927	0.9852529223	-0.223127002	-0.533056185
7	_Corr_Steam Te...	-0.549875041	-0.629023927	1	-0.595214609	0.2475387217	0.2192147319
8	_Corr_MW	0.8558570808	0.9852529223	-0.595214609	1	-0.207305813	-0.50447312
9	_Corr_Cool Temp	-0.270049819	-0.223127002	0.2475387217	-0.207305813	1	0.3617461646
10	_Corr_Pressure	-0.469928462	-0.533056185	0.2192147319	-0.50447312	0.3617461646	1

2. 새 데이터 테이블을 Steam Turbine Targets.jmp 로 저장합니다 .

목표 통계량이 설정되었으므로 공정을 모니터링하는 다변량 관리도를 생성합니다.

3 단계 : 공정 모니터링

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Steam Turbine Current.jmp 를 엽니다 .

이 샘플 데이터 테이블에는 공정의 최근 관측값이 포함되어 있습니다 .

2. **분석 > 품질 및 공정 > 관리도 > 다변량 관리도**를 선택합니다 .
3. 열을 모두 선택하고 **Y, 열**을 클릭합니다 .
4. **목표값 가져오기**를 클릭합니다 .
5. 저장한 Steam Turbine Targets.jmp 테이블을 엽니다 .
6. **확인**을 클릭합니다 .

기본 유의 수준이 0.05로 설정되어 있습니다. 이 값을 0.001로 변경합니다.

7. " 다변량 관리도 " 옆의 빨간색 삼각형을 클릭하고 **유의 수준 설정 > 기타**를 선택합니다 .
8. 0.001 을 입력하고 **확인**을 클릭합니다 .

그림 10.4 Steam Turbine(증기 터빈) 관리도

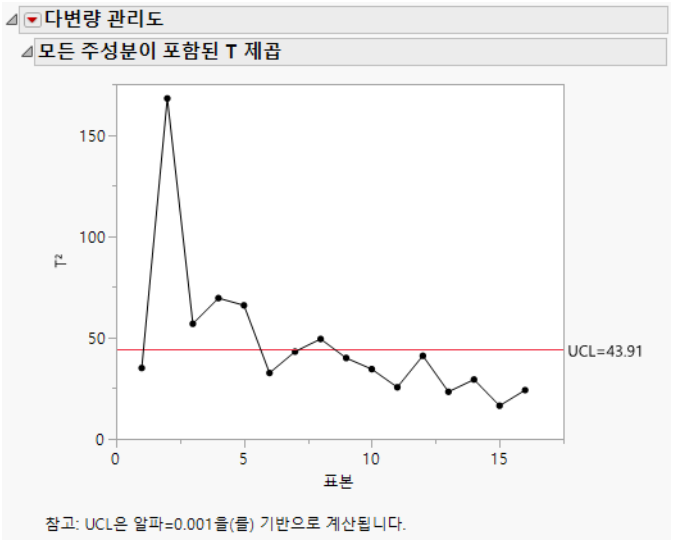
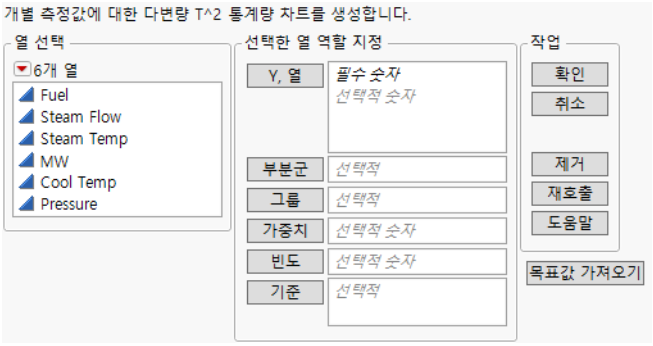


그림 10.4에서는 관측값 2, 3, 4, 5, 8에서 발생하는 관리이탈 상태를 보여 줍니다. 이 결과는 해당 관측값이 Steam Turbine Historical.jmp의 과거 데이터에 부합하지 않으며 공정에 대한 추가 조사가 필요하다는 의미입니다. 지정 가능한 원인을 찾기 위해 개별 단변량 관리도를 검토하거나 다른 단변량 절차를 수행할 수 있습니다.

다변량 관리도 플랫폼 시작

분석 > 품질 및 공정 > 관리도 > 다변량 관리도를 선택하여 다변량 관리도 플랫폼을 시작할 수 있습니다 .

그림 10.5 다변량 관리도 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

다변량 관리도 플랫폼 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y, 열 분석할 열을 지정합니다 .

부분군 부분군으로 묶은 데이터가 포함된 열을 입력합니다 . 이 그룹은 계층에서 **그룹**에 내포됩니다 .

그룹 최상위 계층 수준에서 소속 그룹을 지정하는 열을 입력합니다 .

가중치 데이터에 가중치 (중요도 또는 영향력) 를 할당하는 데 사용할 변수가 있는 데이터 테이블 열을 식별합니다 .

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 데이터 테이블 열을 식별합니다 . 데이터 테이블에 요약 데이터가 포함된 경우 유용할 수 있습니다 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다 .

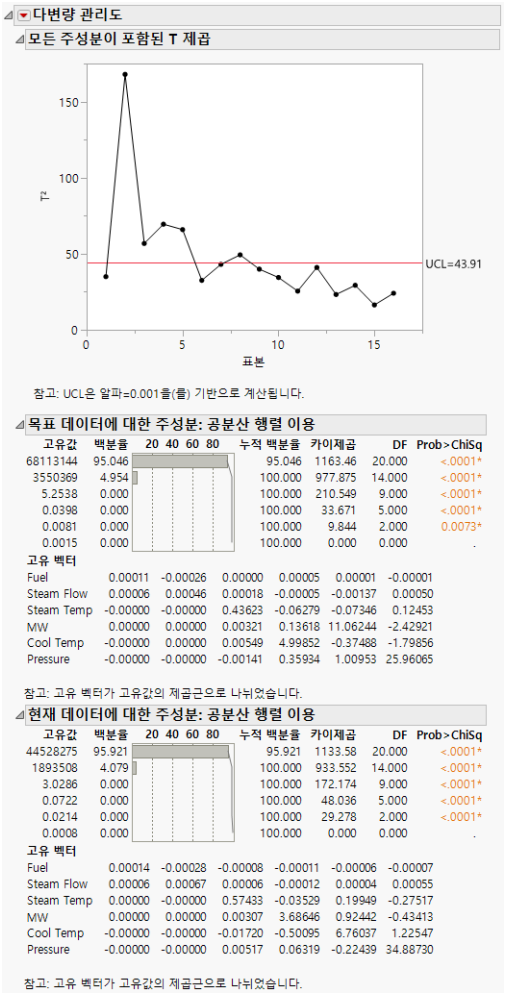
목표값 가져오기 클릭하여 공정의 과거 목표값이 포함된 JMP 테이블을 선택합니다 .

다변량 관리도

다변량 관리도를 사용하여 공정의 변화를 빠르게 식별하고 , 공정을 모니터링하여 특수 원인을 표시할 수 있습니다 .

그림 10.6에 표시된 것과 같은 결과를 생성하려면 "다변량 관리도의 예"에 설명된 방법을 따르십시오 .

그림 10.6 다변량 관리도



팁 : 기타 옵션에 대한 자세한 내용은 " 다변량 관리도 플랫폼 옵션 " 에서 확인하십시오 .

다변량 관리도에는 Hotelling T^2 통계량이 표시됩니다. 목표값이 지정되었는지 여부에 따라 관리 한계 계산이 달라집니다. T^2 통계량 및 UCL(관리 상한)의 계산 방식을 이해하려면 "다변량 관리도에 대한 통계 상세 정보"의 내용을 참조하십시오. 관리 한계에 대한 자세한 내용은 Tracy et al. 연구 자료 (1992)에서 확인하십시오.

이 예에서 두 데이터 집합에 대한 주성분 보고서는 첫 번째 주성분에 해당하는 첫 번째 고유값이 변수에서 발생하는 총 변동의 95%를 설명한다는 것을 나타냅니다. 두 고유 벡터 테이블의 값은 첫 번째 주성분이 주로 Fuel 및 Steam Flow 변수를 기반으로 한다는 것을 나타냅니다. 이 정보를 사용하여 첫 번째 성분만 기반으로 하는 민감도가 더 높은 관리도를 생성할 수 있습니다. 주성분 보고서에 대한 자세한 내용은 "주성분"에서 확인하십시오.

다변량 관리도 플랫폼 옵션

" 다변량 관리도 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

T 제곱 차트 T^2 차트를 표시합니다 . Hotelling T^2 차트는 상관을 고려한 *X*Bar 차트의 다변량 확장입니다 .

분할된 T 제곱 Y 의 주성분을 기반으로 다변량 관리도를 생성합니다 . T^2 에 대한 주성분 수를 지정합니다 . 자세한 내용은 " 분할된 T 제곱 " 에서 확인하십시오 .

유의 수준 설정 관리 한계를 계산하는 데 사용되는 α 수준을 설정합니다. 기본값은 α=0.05입니다.

공분산 표시 공분산 보고서를 표시합니다 . 공분산은 두 변수 간의 선형 관계를 나타내는 측도입니다 . 부분군 변수를 지정한 경우 합동 공분산 보고서가 표시됩니다 .

상관 표시 상관 보고서를 표시합니다. 부분군 변수를 지정한 경우 합동 상관 보고서가 표시됩니다.

공분산 역행렬 표시 공분산 역행렬 보고서를 표시합니다 . 공분산 역행렬이 특이 행렬인 경우 공분산 행렬의 일반화 역행렬이 보고됩니다 . 부분군 변수를 지정한 경우 합동 공분산 역행렬 보고서가 표시됩니다 .

역상관 표시 역상관 보고서를 표시합니다 . 역상관이 특이 행렬인 경우 상관 행렬의 일반화 역행렬이 보고됩니다 . 부분군 변수를 지정한 경우 합동 역상관 보고서가 표시됩니다 .

평균 표시 각 그룹의 평균을 포함하는 그룹 평균 보고서를 표시합니다 .

T 제곱 저장 T^2 값이 포함된 새 열을 데이터 테이블에 생성합니다 .

T 제곱 계산식 저장 데이터 테이블에 새 열을 생성합니다. T^2 값의 계산식을 이 열에 저장합니다.

목표 통계량 저장 공정에 대한 목표 통계량이 포함된 데이터 테이블을 새로 생성합니다 . 목표 통계량에는 표본 크기 , 표본 수 , 평균 , 표준편차 및 상관이 포함됩니다 .

변화 지점 감지 (부분군으로 묶은 데이터에는 사용 불가능) 행 번호별로 검정 통계량의 변화 지점 감지 그림을 표시하고 변화 지점이 나타나는 행 번호를 알려 줍니다 . 자세한 내용은 " 변화 지점 감지 " 에서 확인하십시오 .

주성분 고유값과 해당 고유 벡터를 보여 주는 보고서를 표시합니다. 주성분은 모니터링 중인 많은 변수 중 공정 변동의 주요 원인이 되는 변수를 파악하는 데 도움이 됩니다 . 자세한 내용은 " 주성분 " 에서 확인하십시오 .

주성분 저장 척도화된 주성분이 포함된 새 열을 데이터 테이블에 생성합니다.

분할된 T 제곱

상관관계가 있는 많은 공정 특성을 모니터링하는 경우 "분할된 T 제곱" 옵션을 사용하여 주성분에 기반한 관리도를 생성할 수 있습니다. 소수의 주성분이 측정값 변동의 많은 부분을 설명하는 경우 이러한 큰 주성분에 기반한 다변량 관리도가 차원이 높은 원래 데이터에 기반한 차트보다 더 민감할 수 있습니다.

"분할된 T 제곱" 옵션은 공분산 행렬의 상태가 잘못된 경우에도 유용합니다. 이 경우 고유값이 작은 성분, 즉 변동을 거의 설명하지 않는 성분은 T^2에 잘못된 영향을 크게 미칠 수 있습니다. 공정 동작을 연구할 때 중요도가 낮은 이러한 성분을 분리하면 유용합니다.

"분할된 T 제곱" 옵션을 선택한 후에는 사용할 주성분 수를 결정해야 합니다.

이 옵션은 "큰 주성분이 포함된 T 제곱"과 "작은 주성분이 포함된 T 제곱"이라는 두 개의 다변량 관리도를 생성합니다. 옵션을 처음 선택할 때 *r*을 주성분 수로 입력한다고 가정해 보겠습니다. 큰 주성분이 포함된 차트는 *r*개의 최대 고유값에 해당하는 *r*개의 주성분을 기반으로 합니다. 이는 "주성분: 공분산 행렬 이용" 보고서의 "백분율" 및 "누적 백분율" 열에 표시된 것과 같이 변동을 가장 많이 설명하는 *r*개의 성분입니다. 작은 주성분이 포함된 차트는 나머지 주성분을 기반으로 합니다.

부분군이 주어진 경우 큰 주성분 차트의 T^2 값과 작은 주성분 차트의 T^2 값을 합산하면 "모든 주성분이 포함된 T^2" 보고서에 표시된 전체 T^2 통계량이 됩니다. 분할된 T^2 값의 계산 방식에 대한 자세한 내용은 Kourti and MacGregor 연구 자료 (1996)에서 확인하십시오.

팁 : T^2 통계량을 분해하는 경우 모형 기반 다변량 관리도 플랫폼을 사용하는 것이 좋습니다. 자세한 내용은 " 모형 기반 다변량 관리도 " 에서 확인하십시오.

변화 지점 감지

데이터 집합이 다변량 개별 관측값으로 구성된 경우 평균 벡터, 공분산 행렬 또는 둘 다에서 변화를 감지하는 관리도를 만들 수 있습니다. 이 방법은 데이터를 분할하고 변화에 대한 가능도비 검정 통계량을 계산합니다. 관리도에 표시되는 통계량은 관측값의 가능도비 검정 통계량을 다음 값의 곱으로 나눈 값입니다.

- 근사 기대값 (변화가 없다고 가정)
- 관리 상한에 대한 근사값

근사 관리 상한으로 나누면 유효 관리 상한인 1에 대해 점을 표시할 수 있습니다. 변화 지점 감지 그림은 관리도 통계량의 최대화된 값에서 발생하는 변화에 대한 변화 지점을 쉽게 보여 줍니다. JMP의 변화 지점 감지 구현은 Sullivan and Woodall(2000)을 기반으로 하며 "변화 지점 감지에 대한 통계 상세 정보"에서 설명합니다.

참고 : 변화 지점 감지 방법은 데이터의 단일 변화를 표시하는 것이 목적입니다 . 이 방법을 반복 적용하여 여러 변화를 감지할 수 있습니다 .

변화 지점 감지 그림의 경우 다음 사항에 유의하십시오 .

- 1.0 보다 큰 값은 데이터의 변화 가능성을 나타냅니다 .
- 변화 지점 감지 그림의 관리도 통계량은 관심 있는 가능도비 통계량 (평균 벡터 또는 공분산 행렬) 을 정규화 계수로 나누어 구할 수 있습니다 .
- 변화 지점 감지 그림에 대한 검정 통계량 값이 최대인 관측값에서 데이터 변화 지점이 나타납니다 .

산점도 행렬의 경우 다음 사항에 유의하십시오 .

- 이 그림에서는 표본 평균 벡터의 변화를 보여 줍니다 .
- " 변화 지점 감지의 예 " 에서는 데이터가 두 그룹으로 구분됩니다 . 처음 24 개 관측값은 첫 번째 그룹으로 분류됩니다 . 나머지 관측값은 두 번째 그룹으로 분류됩니다 .

주성분

주성분 보고서에는 다음 정보가 포함됩니다 .

고유값 공분산 행렬에 대한 고유값입니다 .

백분율 해당하는 고유 벡터에 의해 설명된 변동의 백분율입니다 . 포함된 막대 차트도 표시됩니다 .

누적 백분율 고유값에 해당하는 고유 벡터에 의해 설명된 변동의 누적 백분율입니다 .

카이제곱 데이터에 남아 있는 상관이 랜덤 특성인지 여부에 대한 검정을 제공합니다 . 이 검정은 구형성 Bartlett 검정입니다 . 이 검정이 귀무가설을 기각하면 데이터에 이 고유값과 관련된 구조가 남아 있음을 의미합니다 .

DF 카이제곱 검정과 관련된 자유도입니다 .

Prob > ChiSq 검정의 p 값입니다 .

고유 벡터 고유값에 해당하는 고유 벡터 테이블입니다 . 각 고유 벡터는 해당 고유값의 제곱근으로 나눕니다 .

주성분에 대한 자세한 내용은 Multivariate Methods 의 에서 확인하십시오 .

다변량 관리도의 추가 예

- "부분군으로 묶은 데이터를 사용한 공정 모니터링의 예"
- "분할된 T 제곱의 예"
- "변화 지점 감지의 예"

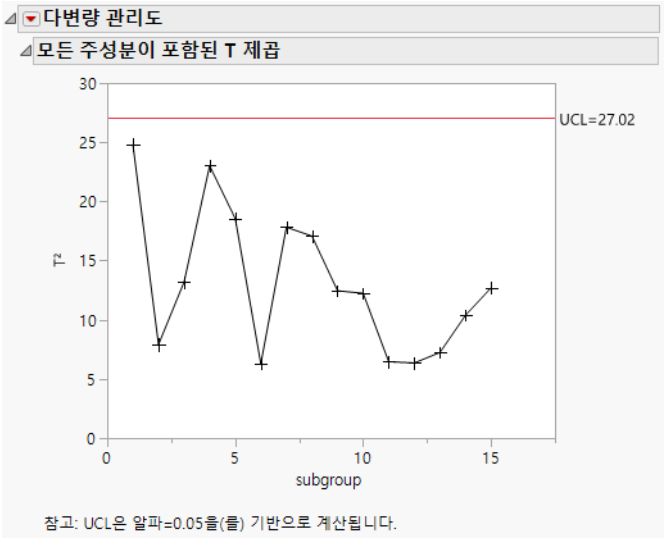
부분군으로 묶은 데이터를 사용한 공정 모니터링의 예

부분군으로 묶은 데이터를 사용하는 다변량 공정 모니터링의 워크플로우는 개별 데이터에 대한 워크플로우와 비슷합니다. 자세한 내용은 "다변량 관리도의 예"에서 확인하십시오. 초기 관리도를 생성하여 목표 통계량을 저장한 후 이러한 목표값을 사용하여 공정을 모니터링합니다.

1 단계 : 공정의 안정성 여부 확인

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Aluminum Pins Historical.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > 다변량 관리도**를 선택합니다 .
3. Diameter 및 Length 열을 모두 선택하고 **Y, 열**을 클릭합니다 .
4. subgroup 을 선택하고 **부분군**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 10.7 부분군으로 묶은 데이터의 다변량 관리도 - 1 단계



공정이 통계적 관리 상태에 있는 것으로 보이며 이 데이터를 사용하여 목표값을 생성할 수 있습니다.

2 단계 : 목표 통계량 저장

1. " 다변량 관리도 " 옆의 빨간색 삼각형을 클릭하고 **목표 통계량 저장**을 선택합니다.
   이렇게 하면 공정에 대한 목표 통계량이 포함된 새 데이터 테이블이 생성됩니다.
2. 새 데이터 테이블을 Aluminum Pins Targets.jmp 로 저장합니다.
   목표 통계량이 설정되었으므로 공정 모니터링을 위한 다변량 관리도를 생성합니다.

3 단계 : 공정 모니터링

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Aluminum Pins Current.jmp를 엽니다.
   이 샘플 데이터 테이블에는 공정의 최근 관측값이 포함되어 있습니다.
2. **분석 > 품질 및 공정 > 관리도 > 다변량 관리도**를 선택합니다.
3. Diameter 및 Length 열을 모두 선택하고 **Y, 열**을 클릭합니다.
4. subgroup 을 선택하고 **부분군**을 클릭합니다.
5. **목표값 가져오기**를 클릭합니다.
6. 저장한 Aluminum Pins Targets.jmp 테이블을 엽니다.
7. **확인**을 클릭합니다.
8. " 다변량 관리도 " 옆의 빨간색 삼각형을 클릭하고 **평균 표시**를 선택합니다.
   " 평균 표시 " 옵션은 각 부분군에 대한 평균을 제공합니다. 그러면 그룹 간에 비유사성이 가장 높은 그룹을 관측할 수 있습니다.

그림 10.8 부분군으로 묶은 데이터의 다변량 관리도 - 3 단계

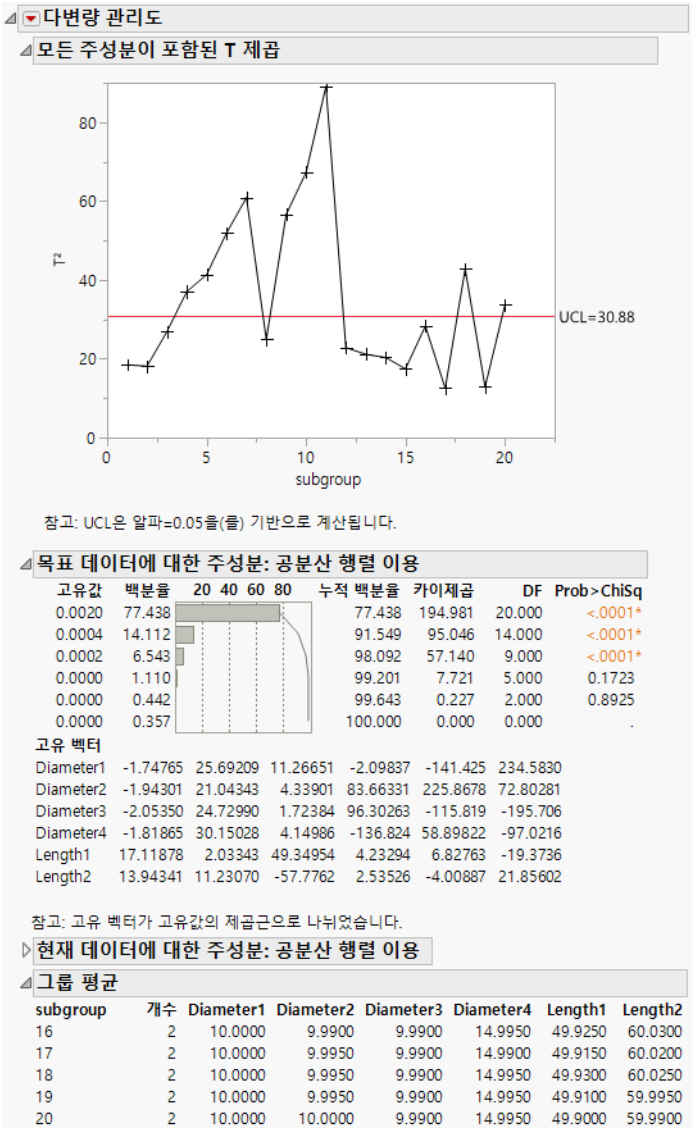


그림 10.8에서는 4-7, 9-11, 18 및 20 부분군에서 불안정성이 나타나는 것을 보여 줍니다. 이 결과는 해당 관측값이 Aluminum Pins Historical.jmp의 과거 데이터에 부합하지 않으며 공정에 대한 추가 조사가 필요하다는 의미입니다. 해당 지점에서 공정이 관리이탈 상태가 된 이유를 판별하기 위해 개별 단변량 관리도를 검토하거나 다른 단변량 절차를 수행할 수 있습니다.

이 공정을 모니터링하는 다른 방법은 큰 주성분을 기반으로 합니다. 이 예에서 과거 데이터의 경우 처음 세 개의 주성분이 변동의 약 98%를 차지합니다. 이를 기반으로 처음 세 개의 주성분에 대한 차트를 생성할 수 있습니다. 그런 다음 이 세 가지 성분을 사용하여 현재 데이터를 모니터링

합니다. 공정 모니터링에 사용되는 차트의 관리 한계는 과거 데이터에 대한 해당 차트를 기반으로 해야 합니다.

분할된 T 제곱의 예

공정 동작을 연구할 때 " 분할된 T 제곱 " 을 사용하여 중요도가 높은 성분과 낮은 성분을 분리할 수 있습니다 . 이 예에서는 막대 사이 간격이 균등한 12 개 위치에서 50 개 막대의 코팅을 각각 측정했습니다 . 측정값의 변동을 검토하고 변동 원인에 대한 추가 조사가 필요한지 여부를 판별하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Thickness.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 관리도 > 다변량 관리도**를 선택합니다 .
3. Thickness 열을 모두 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

   현재 유의 수준은 5% 오경보 비율에 해당하는 0.05 로 설정되어 있습니다 . 오경보 비율을 1% 로 설정하려고 합니다 .
5. 유의 수준을 변경하려면 " 다변량 관리도 " 옆의 빨간색 삼각형을 클릭하고 **유의 수준 설정** , **0.01** 을 차례로 선택합니다 .

그림 10.9 Thickness.jmp 에 대한 초기 다변량 관리도

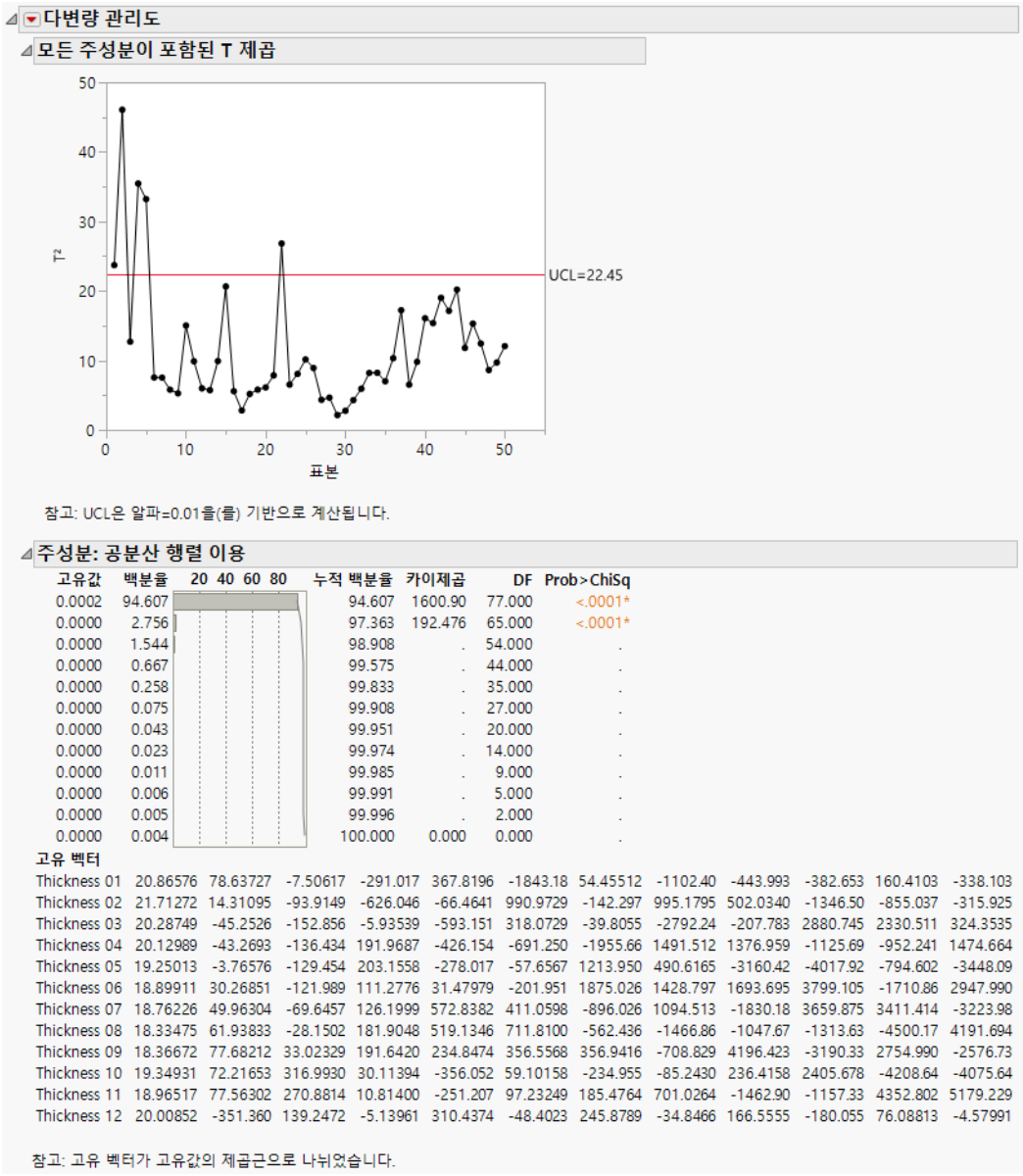
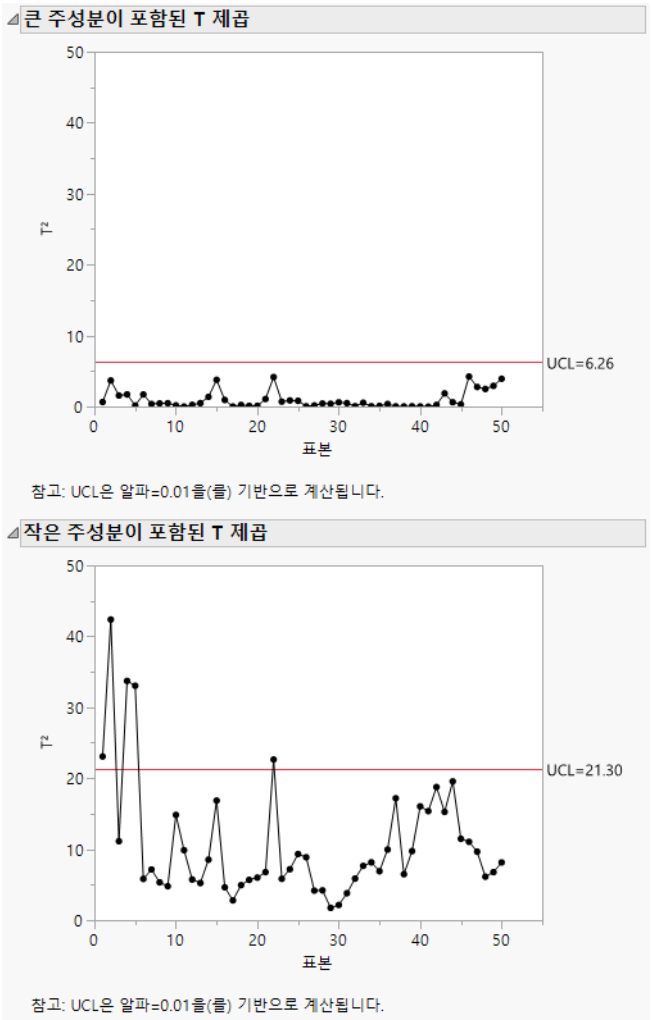


그림 10.9 의 전체 관리도에서는 특수 원인이 막대 1, 2, 4, 5, 22 에 영향을 미쳤음을 나타냅니다 . 주성분 보고서를 살펴보면 12 개 두께 측정값의 변동 중 약 95% 가 첫 번째 주성분에 의해 설명된다는 것을 알 수 있습니다 . 이 주성분과 관련된 변동을 더 자세히 연구하려고 합니다 .

6. " 다변량 관리도 " 의 빨간색 삼각형을 클릭하고 **분할된 T 제곱**을 선택합니다 .
7. **확인**을 클릭하여 주성분의 기본값 1 을 적용합니다 .

그림 10.10 분할된 T 제곱 관리도



주성분 보고서와 달리 첫 번째 성분에 대한 변동만 반영하는 "큰 주성분이 포함된 T 제곱" 차트에는 특수 원인에 대한 증거가 보이지 않습니다. "작은 주성분이 포함된 T 제곱" 차트에서는 나머지 더 작은 성분에 있는 특수 원인 표시를 볼 수 있습니다. 이러한 작은 성분은 큰 변동을 설명하지 않으며 랜덤 잡음을 나타낼 가능성이 있습니다. 따라서 두께 측정값의 변동은 중요한 문제가 아니라는 결론을 내릴 수 있습니다.

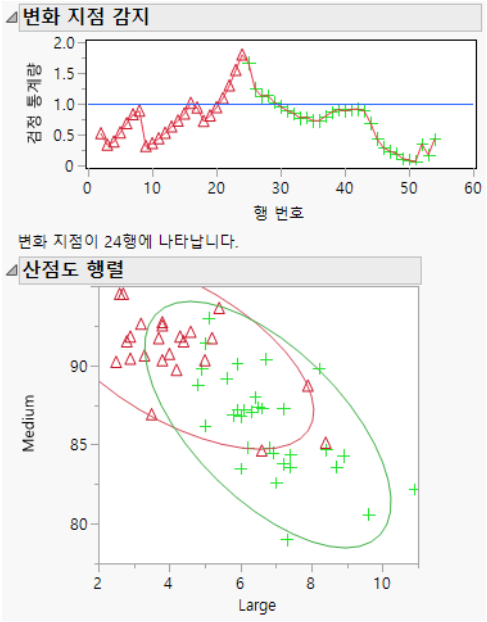
변화 지점 감지의 예

변화 지점 감지를 사용하여 데이터에서 변화가 발생하는 지점을 찾을 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Gravel.jmp 를 엽니다 .

2. **분석 > 품질 및 공정 > 관리도 > 다변량 관리도**를 선택합니다 .
3. Large 와 Medium 을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 다변량 관리도 " 옆의 빨간색 삼각형을 클릭하고 **변화 지점 감지**를 선택합니다 .

그림 10.11 Gravel.jmp 에 대한 변화 지점 감지



팁 : 데이터에 따라 두 그룹의 밀도 타원을 보려면 축을 드래그해야 할 수도 있습니다 .

변화 지점 감지 그림에서 1.0 보다 큰 값은 데이터의 변화 가능성을 나타냅니다. 적어도 하나의 변화는 명백하게 나타납니다. 관측값 24 에서 변화 지점이 보이고 24 직후에 변화가 발생합니다. 두 그룹의 95% 예측 영역은 크기, 형상 및 방향이 거의 같으므로 표본 공분산 행렬이 유사하다는 것을 시각적으로 나타냅니다.

다변량 관리도에 대한 통계 상세 정보

- " 개별 관측값에 대한 통계 상세 정보 "
- " 합리적 부분군의 관측값에 대한 통계 상세 정보 "
- " 변화 지점 감지에 대한 통계 상세 정보 "

개별 관측값에 대한 통계 상세 정보

부분군으로 묶지 않은 측정값, 즉 자연 부분군 크기가 $n = 1$인 측정값을 예로 들어보겠습니다. 관측값 수를 m으로 나타내고 측정된 변수의 수를 p로 나타냅니다. 각 관측값에 대한 T^2 통계량이 계산되어 그림에 표시됩니다. T^2 통계량 및 UCL(관리 상한) 계산은 목표 통계량의 소스에 따라 달라집니다. 단계 I 차트의 한계는 관리도에 표시되는 것과 동일한 데이터를 기반으로 합니다. 단계 II 차트의 한계는 과거 데이터 집합에서 계산된 목표 통계량을 기반으로 합니다. Hotelling T^2 관리도의 T^2 통계량 및 관리 한계 계산에 대한 자세한 내용은 Montgomery 연구 자료(2013)에서 확인하십시오.

단계 I 관리도에 대한 계산

단계 I 관리도에서 i 번째 관측값의 T^2 통계량은 다음과 같이 정의됩니다.

$$T_i^2 = (Y_i - \overline{Y})' S^{-1} (Y_i - \overline{Y})$$

다음은 각 요소에 대한 설명입니다.

$Y_i = i$ 번째 관측값에 대한 p 개 측정값의 열 벡터
$\overline{Y} = p$ 개 변수에 대한 표본 평균의 열 벡터
S^{-1} = 표본 공분산 행렬의 역

i개의 각 관측값에 대한 T_i^2 값은 다변량 관리도에 표시된 점입니다.

단계 I 관리 한계를 계산할 때 UCL 은 베타 분포를 기반으로 합니다. UCL(관리 상한) 은 다음과 같이 정의됩니다.

$$UCL = \frac{(m-1)^2}{m} \beta_{\left[1-\alpha; \frac{p}{2}; \frac{m-p-1}{2}\right]}$$

다음은 각 요소에 대한 설명입니다.

p = 변수 수
m = 관측값 수
$\beta_{\left[1-\alpha; \frac{p}{2}; \frac{m-p-1}{2}\right]}$ = 베타 $\left(\frac{p}{2}, \frac{m-p-1}{2}\right)$ 분포의 (1–α) 번째 분위수

단계 II 관리도에 대한 계산

단계 II 관리도에서 과거 데이터 집합을 X 로 정의합니다. 그러면 i 번째 관측값의 T^2 통계량이 다음과 같이 정의됩니다.

$$T_i^2 = (Y_i - \overline{X})' S_X^{-1} (Y_i - \overline{X})$$

다음은 각 요소에 대한 설명입니다.

$Y_i = i$ 번째 관측값에 대한 p 개 측정값의 열 벡터

$\overline{X}$ = 과거 데이터 집합에서 계산된 p 개 변수에 대한 표본 평균의 열 벡터

S_X^{-1} = 과거 데이터 집합에서 계산된 표본 공분산 행렬의 역

i 개의 각 관측값에 대한 T_i^2 값은 다변량 관리도에 표시된 점입니다.

단계 II 관리 한계를 계산할 때 새 관측값은 과거 데이터 집합과 독립적입니다. 이 경우 UCL(관리 상한) 은 F 분포의 함수이며, 목표값을 계산하는 과거 데이터 집합의 관측값 수에 따라 부분적으로 달라집니다. UCL 은 다음과 같이 정의됩니다.

$$UCL = \begin{cases} \dfrac{p(m+1)(m-1)}{m(m-p)} F_{[1-\alpha, p, m-p]} & \text{단, } m \le 100 \\ \dfrac{p(m-1)}{m-p} F_{[1-\alpha, p, m-p]} & \text{단, } m > 100 \end{cases}$$

다음은 각 요소에 대한 설명입니다.

p = 변수 수

m = 과거 데이터 집합의 관측값 수

$F_{[1-\alpha, p, m-p]} = F(p, m-p)$ 분포의 $(1-\alpha)$ 번째 분위수

합리적 부분군의 관측값에 대한 통계 상세 정보

p 개 변수를 모니터링하고 $n > 1$ 크기의 m 개 부분군을 구하는 경우를 예로 들어보겠습니다. 각 부분군에 대한 T^2 통계량이 계산되어 그림에 표시됩니다. T^2 통계량 및 UCL(관리 상한) 계산은 목표 통계량의 소스에 따라 달라집니다. 단계 I 차트의 한계는 관리도에 표시되는 것과 동일한 데이터를 기반으로 합니다. 단계 II 차트의 한계는 과거 데이터 집합에서 계산된 목표 통계량을 기반으로 합니다. Hotelling T^2 관리도의 T^2 통계량 및 관리 한계 계산에 대한 자세한 내용은 Montgomery 연구 자료(2013)에서 확인하십시오.

단계 I 관리도에 대한 계산

단계 I 관리도에서 j 번째 부분군의 T^2 통계량은 다음과 같이 정의됩니다 .

$$T_j^2 = (\overline{Y}_j - \overline{Y})' S_p^{-1} (\overline{Y}_j - \overline{Y})$$

다음은 각 요소에 대한 설명입니다 .

$\overline{Y}_j$ = j 번째 부분군에 대한 p 개 측정값의 n 개 열 벡터의 평균

$\overline{Y} = \frac{1}{m} \sum_{j=1}^{m} \overline{Y}_j$ = 부분군 평균의 평균

S_j = j 번째 부분군의 n 개 관측값에 대한 표본 공분산 행렬

$S_p = \frac{1}{m} \sum_{j=1}^{m} S_j$ = 부분군 내 공분산 행렬의 평균으로 계산되는 합동 공분산 행렬

단계 I UCL(관리 상한) 은 다음과 같이 정의됩니다 .

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{[1-\alpha, p, mn-m-p+1]}$$

다음은 각 요소에 대한 설명입니다 .

p = 변수 수

n = 각 부분군의 표본 크기

m = 부분군 수

$F_{[1-\alpha, p, mn-m-p+1]} = F(p, mn-m-p+1)$ 분포의 (1–α) 번째 분위수

단계 II 관리도에 대한 계산

단계 II 관리도에서 , 목표 통계량을 계산하는 과거 데이터 집합을 X. 로 정의합니다 . 그러면 j 번째 부분군의 T^2 통계량이 다음과 같이 정의됩니다 .

$$T_j^2 = (\overline{Y}_j - \overline{X})' S_p^{-1} (\overline{Y}_j - \overline{X})$$

다음은 각 요소에 대한 설명입니다 .

$\overline{Y}_j$ = j 번째 부분군에 대한 p 개 측정값의 n 개 열 벡터의 평균

$\overline{X}_k$ = 과거 데이터 집합에서 k 번째 부분군에 대한 p 개 측정값의 n 개 열 벡터의 평균

$\overline{X} = \frac{1}{m}\sum_{k=1}^{m} \overline{X}_k$ = 관측값의 전체 평균

S_k = 과거 데이터 집합에서 k 번째 부분군의 n 개 관측값에 대한 표본 공분산 행렬

$S_p = \frac{1}{m}\sum_{k=1}^{m} S_k$ = 부분군 내 공분산 행렬의 평균으로 계산되는 합동 공분산 행렬

단계 II UCL(관리 상한) 은 다음과 같이 정의됩니다 .

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{[1-\alpha,\, p,\, mn-m-p+1]}$$

다음은 각 요소에 대한 설명입니다 .

p = 변수 수

n = 부분군 표본 크기

m = 과거 데이터 집합의 부분군 수

$F_{[1-\alpha,\, p,\, mn-m-p+1]}$ = $F(p, mn-m-p+1)$ 분포의 (1–α) 번째 분위수

합리적 부분군의 관측값에 대한 검정 통계량의 가법성

독립이며 정규 분포를 따르는 mn 개의 관측값 표본이 각각 n 크기의 합리적 부분군 m 개로 그룹화되는 경우 , T^2_M 을 j 번째 부분군의 평균 ($\overline{Y}_j$) 과 목표값 사이의 거리로 정의합니다 . T^2_M 은 이전 섹션에 나오는 합리적 부분군의 관측값에 대한 T^2 과 동등합니다 . 각 부분군의 내부 변동 및 목표값 기준의 전체 변동과 관련된 T^2 통계량도 계산할 수 있습니다 . T^2 통계량의 성분은 제곱합과 상당히 비슷한 가법성을 나타냅니다 . 구체적으로 말하면 m 개의 각 부분군에 대해 다음 관계가 성립됩니다 .

$$T^2_{A_j} = T^2_{M_j} + T^2_{D_j}$$

다음 모든 정의에서 S_p 는 관리도 유형 (단계 I 또는 단계 II 관리도) 에 따라 이전 섹션에서와 같이 정의됩니다 . 또한 μ 를 $\overline{Y}$ (단계 I 관리도의 경우) 및 $\overline{X}$ (단계 II 관리도의 경우) 로 정의합니다 .

j 번째 부분군의 목표값에서 떨어진 거리는 다음과 같이 정의됩니다 .

$$T^2_{M_j} = n(\overline{Y}_j - \mu)' S_P^{-1} (\overline{Y}_j - \mu)$$

j 번째 부분군의 내부 변동은 다음과 같이 정의됩니다 .

$$T^2_{D_j} = \sum_{i=1}^{n} (Y_{ji} - \overline{Y}_j)' S_P^{-1} (Y_{ji} - \overline{Y}_j)$$

여기서 Y_{ji} 는 j 번째 부분군에 대한 p 개 측정값의 i 번째 열 벡터입니다 .

j 번째 부분군의 전체 변동은 다음과 같이 정의됩니다 .

$$T^2_{A_j} = \sum_{i=1}^{n} (Y_{ji} - \mu)' S_P^{-1} (Y_{ji} - \mu)$$

여기서 Y_{ji}는 j번째 부분군에 대한 p개 측정값의 i번째 열 벡터입니다.

참고 : " 다변량 관리도 " 의 빨간색 삼각형 메뉴에서 **T 제곱 저장** 또는 **T 제곱 계산식 저장** 옵션을 선택하는 경우 각 행에 저장된 세 값은 위에 나오는 세 가지 정의의 i 값 중 하나에 해당합니다 .

변화 지점 감지에 대한 통계 상세 정보

이 섹션의 설명은 Sullivan and Woodall(2000) 에서 발전된 내용을 기반으로 합니다.

가정

평균 벡터가 μ_i 이고 공분산 행렬이 Σ_i 인 p 차원의 다변량 분포를 $N_p(\mu_i,\Sigma_i)$ 로 나타냅니다 . 다음과 같이 x_i 를 이러한 분포와 독립적인 $m(m > p)$ 개의 관측값이라고 가정해 보겠습니다 .

$$x_i \sim N_p(\mu_i,\Sigma_i), \quad i = 1, \ldots, m$$

공정이 안정적이면 평균 μ_i 및 공분산 행렬 Σ_i가 공통 값과 동일하여 x_i가 $N_p(\mu, \Sigma)$ 분포를 따릅니다.

m_1 개 관측값과 m_1+1 개 관측값 사이에 평균 벡터와 공분산 행렬 중 하나 또는 둘 다에서 단일 변화가 발생한다고 가정해 보겠습니다 . 그러면 다음 조건이 충족됩니다 .

- $1 \sim m_1$ 관측값은 평균 벡터 및 공분산 행렬이 동일합니다 (μ_a,Σ_a).
- $m_1 + 1 \sim m$ 관측값은 평균 벡터 및 공분산 행렬이 동일합니다 (μ_b,Σ_b).
- 다음 중 하나가 발생합니다 .
  - 변화가 평균에 영향을 미치는 경우 $\mu_a \neq \mu_b$ 입니다 .
  - 변화가 공분산 행렬에 영향을 미치는 경우 $\Sigma_a \neq \Sigma_b$ 입니다 .
  - 변화가 평균과 공분산 행렬에 모두 영향을 미치는 경우 $\mu_a \neq \mu_b$, $\Sigma_a \neq \Sigma_b$ 입니다 .

개요

가능도비 검정 방법은 평균 벡터와 공분산 행렬 중 하나 또는 둘 다에서 발생하는 변화를 식별하는 데 사용됩니다. 가능도비 검정 통계량은 근사 관리 상한이 1인 관리도 통계량을 계산하는 데 사용됩니다. 가능한 모든 m_1 값에 대해 관리도 통계량이 표시됩니다. 관측값의 관리도 통계량이 관리 상한 1을 초과하면 변화가 발생했음을 나타냅니다. 변화가 정확히 한 번 발생한다고 가정하면 관리도 통계량 값이 최대인 관측값 직후에 변화가 시작된다고 간주됩니다.

가능도비 검정 통계량

처음 m_1 개 관측값에 대한 로그 가능도 함수의 두 배인 최대값은 다음과 같이 정의됩니다.

$$l_1 = -m_1 k_1 \log[2\pi] - m_1 \log\left[\left|S_1\right|_{k_1}\right] - m_1 k_1$$

l_1 에 대한 방정식에서는 다음과 같은 표기를 사용합니다.

- S_1 은 처음 m_1 개 관측값에 대한 공분산 행렬의 최대 가능도 추정값입니다.
- $k_1 = \text{Min}[p, m_1\text{-}1]$ 은 $p \times p$ 행렬 S_1 의 계수입니다.
- $\left|S_1\right|_{k_1}$ 표기는 S_1 행렬의 일반화된 행렬식을 나타내며, 다음과 같이 k_1 개의 양의 고유값 λ_j 에 대한 곱으로 정의됩니다.

  $$\left|S_1\right|_{k_1} = \prod_{j=1}^{k_1} \lambda_j$$

  일반화된 행렬식은 S_1 이 완전 계수를 가질 때 보통 행렬식과 같습니다.

후속 $m_2 = m - m_1$ 개 관측값에 대한 로그 가능도 함수의 두 배인 최대값을 l_2로 나타내고, 모든 m 개 관측값에 대한 로그 가능도 함수의 두 배인 최대값을 l_0으로 나타냅니다. l_2와 l_0 모두 l_1에 제공된 것과 비슷한 표현식이 제공됩니다.

가능도비 검정 통계량은 $l_1 + l_2$ 합을 l_0과 비교합니다. $l_1 + l_2$ 합은 m_1에서 변화가 발생할 수 있다고 가정하는 로그 가능도의 두 배입니다. l_0 값은 변화가 없다고 가정하는 로그 가능도의 두 배입니다. l_0이 $l_1 + l_2$보다 많이 작으면 공정이 불안정하다고 가정합니다.

관측값 $m_1 + 1$ 에서 변화가 시작되는지 여부를 검정하는 경우 가능도비 검정 통계량은 다음과 같이 정의됩니다.

$$\begin{aligned} \text{lrt}[m_1] &= (l_1 + l_2 - l_0) \\ &= (m_1(p-k_1) + m_2(p-k_2))(1+\log(2\pi)) \\ &\quad + m\log[|S|] - m_1\log\left[\left|S_1\right|_{k_1}\right] - m_2\log\left[\left|S_2\right|_{k_2}\right] \end{aligned}$$

가능도비 검정 통계량은 자유도가 $p(p+3)/2$인 점근적 카이제곱 분포를 따릅니다. 로그 가능도비 값이 크면 공정이 불안정하다는 것을 나타냅니다.

관리도 통계량

시뮬레이션에 따르면 $\text{lrt}[m_1]$의 기대값은 계열에서 관측값의 위치를 기반으로 하며, 특히 p와 m에 따라 달라집니다. 자세한 내용은 Sullivan and Woodall 연구 자료(2000)에서 확인하십시오.

$\text{lrt}[m_1]$의 기대값을 구하는 근사 계산식은 시뮬레이션을 통해 도출됩니다. p에 대한 기대값의 종속성을 줄이기 위해 $\text{lrt}[m_1]$을 점근적 기대값 $p(p+3)/2$로 나눕니다.

$p(p+3)/2$ 로 나눈 $\mathrm{lrt}[m_1]$ 의 근사 기대값을 구하는 계산식은 다음과 같이 정의됩니다 .

$$\mathrm{ev}[m,p,m_1] = \begin{cases} a_p + m_1 b_p, & \text{if } m_1 < p+1 \\ a_p + (m-m_1)b_p, & \text{if } m - m_1 < p+1 \\ 1 + \dfrac{m-2p-1}{(m_1-p)(m-p-m_1)}, & \text{otherwise} \end{cases}$$

여기서 ,

$$a_p = -\frac{0.08684(p-14.69)(p-2.036)}{(p-2)}$$

및

$$b_p = \frac{0.1228(p-1.839)}{(p-2)}$$

$p = 2$ 의 경우 m_1 또는 $m_2 = 2$ 일 때 $\mathrm{ev}[m,p,m_1]$ 값은 1.3505 입니다.

참고 : $p > 12$ 또는 $m < (2p+4)$ 일 때는 위의 계산식이 정확하지 않습니다 . 이러한 경우 시뮬레이션을 사용하여 근사 기대값을 구해야 합니다 .

공정이 안정적이라고 가정할 때 약 0.05 의 확률로 잘못된 관리이탈 신호를 생성하는 근사 관리 상한은 다음과 같이 계산됩니다 .

$$\mathrm{UCL}[m,p] \cong (3.338 - 2.115\log[p] + 0.8819(\log[p])^2 - 0.1382(\log[p])^3)$$

$$+ \ (0.6389 - 0.3518\log[p] + 0.01784(\log[p])^3)\log[m].$$

이 계산식은 m 과 p 에 따라 달라집니다.

관리도 통계량은 가능도비 검정 통계량 로그의 두 배를 $p(p+3)$, 근사 기대값 및 관리 한계 근사값으로 나눈 값으로 정의됩니다 . UCL 의 근사값으로 나누었으므로 관리 상한 1 에 대해 관리도 통계량을 표시할 수 있습니다 . 근사 관리도 통계량은 다음과 같이 계산됩니다 .

$$\hat{y}[m_1] = \frac{2\mathrm{lrt}[m_1]}{p(p+3)(\mathrm{ev}[m,p,m_1]\mathrm{UCL}[m,p])}$$

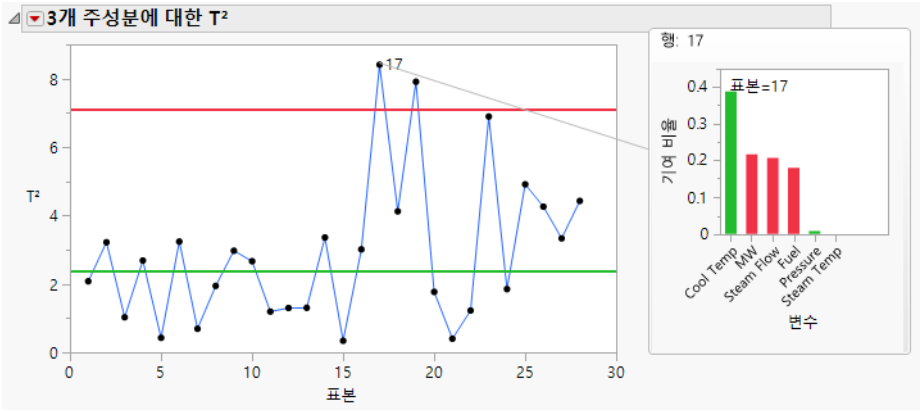
11 장

모형 기반 다변량 관리도

복합 공정 모니터링 및 진단

모형 기반 다변량 관리도는 단일 관리도에서 여러 공정에 대한 모수를 모니터링하는 데 사용됩니다. MDMVCC(모형 기반 다변량 관리도) 플랫폼에서는 주성분 또는 부분 최소 제곱 모형을 기반으로 관리도를 생성할 수 있습니다. 연속형 변수 집합의 경우 MDMVCC 플랫폼은 주성분을 사용하여 관리도를 생성합니다. 저장된 주성분 또는 부분 최소 제곱 스코어 함수의 경우 MDMVCC 플랫폼은 제공된 모형을 기반으로 관리도를 생성합니다. MDMVCC 플랫폼을 사용하면 관리이탈 신호를 발생시키는 기본 성분을 대화식으로 탐색하고 이해할 수 있습니다.

그림 11.1 모형 기반 다변량 관리도



목차

모형 기반 다변량 관리도 개요 289
모형 기반 다변량 관리도의 예 289
모형 기반 다변량 관리도 플랫폼 시작 292
데이터 형식 292
모형 기반 다변량 관리도 보고서 293
모형 기반 다변량 관리도 플랫폼 옵션 294
그림 옵션 295
스코어 그림 296
모형 기반 다변량 관리도 플랫폼의 추가 예 297
과거 데이터를 사용하는 MDMVCC의 예 297
PLS 모형을 사용하는 MDMVCC의 예 298
모형 기반 다변량 관리도 플랫폼에 대한 통계 상세 정보 299
모니터링 통계량 299
한계 301
기여도 303
스코어 그림 그룹 비교 306
스코어 그림 적재 306

모형 기반 다변량 관리도 개요

MDMVCC(모형 기반 다변량 관리도) 플랫폼에는 모니터링과 진단이라는 두 가지 주요 기능이 있습니다.

- 다변량 관리도를 사용하여 다변량 공정을 모니터링할 수 있습니다.
- 대화식으로 드릴다운하여 전체 신호에 대한 개별 변수의 기여도를 조사한 후 공정을 진단할 수 있습니다.

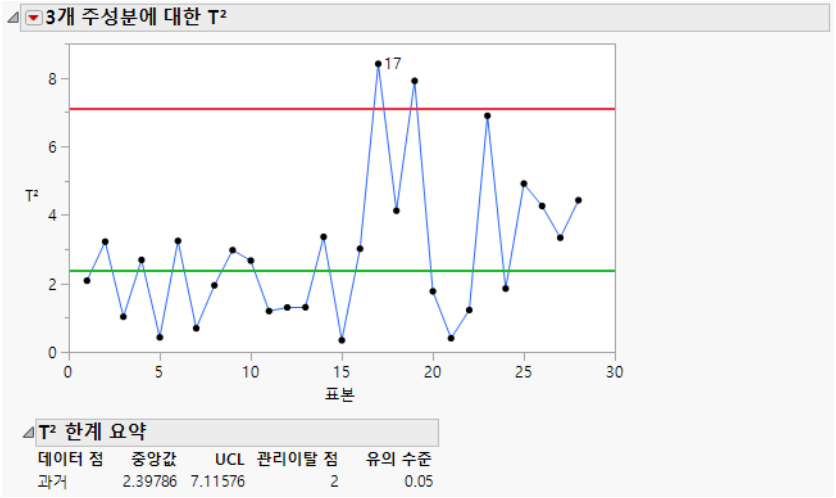
다변량 관리도에 대한 자세한 내용은 Kourti and MacGregor 연구 자료(1996)에서 확인하십시오.

모형 기반 다변량 관리도의 예

이 예에서는 증기 터빈 시스템의 공정 변수가 포함된 Steam Turbine Historical.jmp 샘플 데이터 테이블을 사용합니다. 모니터링된 6개 변수에 대한 관리도를 생성하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Steam Turbine Historical.jmp 를 엽니다.
2. **분석 > 품질 및 공정 > 모형 기반 다변량 관리도**를 선택합니다.
3. 6 개 열을 모두 선택하고 **공정**을 클릭한 후 **확인**을 클릭합니다.

그림 11.2 Steam Turbine(증기 터빈) 보고서

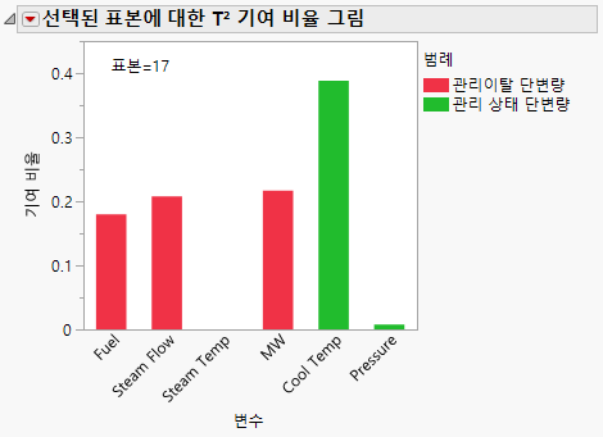


표본 16 이후에 공정 변화가 시작됩니다.

4. 표본 17 데이터 점을 선택합니다. 마우스 오른쪽 버튼을 클릭하고 **행 > 행 라벨**을 선택합니다.

5. 표본 17 데이터 점을 커서로 가리켜 해당 점에 대한 T^2 기여 비율 그림을 봅니다. 그림을 클릭하면 보고서 창에서 그림이 열립니다.

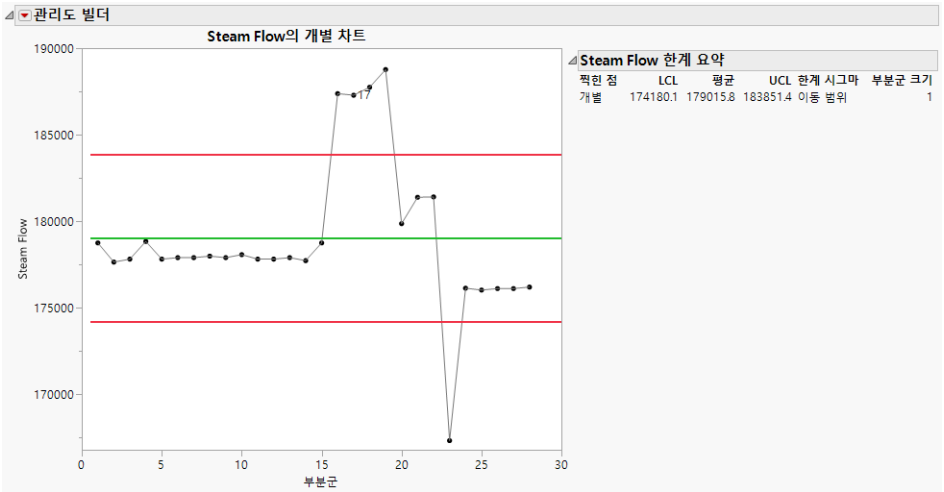
그림 11.3 표본 17 에 대한 기여 비율 그림



Cool Temp 의 기여 비율은 T^2 값의 40% 입니다. Cool Temp 막대는 녹색으로 표시되어 표본 17 이 Cool Temp 에 대한 단변량 관리 한계 내에 있음을 나타냅니다. Steam Flow 와 MW 의 기여 비율은 각각 T^2 값의 20% 입니다. 둘 다 빨간색으로 표시되어 표본 17 이 각 변수에 대한 단변량 관리 한계를 벗어남을 나타냅니다. T^2 값에 대한 Steam Temp 의 기여도는 0 입니다. 이 예에서 다변량 관리이탈 표본을 관리이탈 단변량 변수로 추적할 수 있는 변수를 찾았습니다. 그러나 항상 그런 것은 아닙니다. 다변량 공정 관리에서 T^2 차트에 관리이탈 점이 표시되지만 해당 표본이 모든 변수에 대한 단변량 수준에서는 관리 상태일 수 있습니다.

6. 기여 비율 그림에서 Steam Flow 막대를 커서로 가리켜 Steam Flow 에 대한 단변량 관리도를 봅니다. 차트를 클릭하면 새 보고서 창에서 차트가 열립니다.

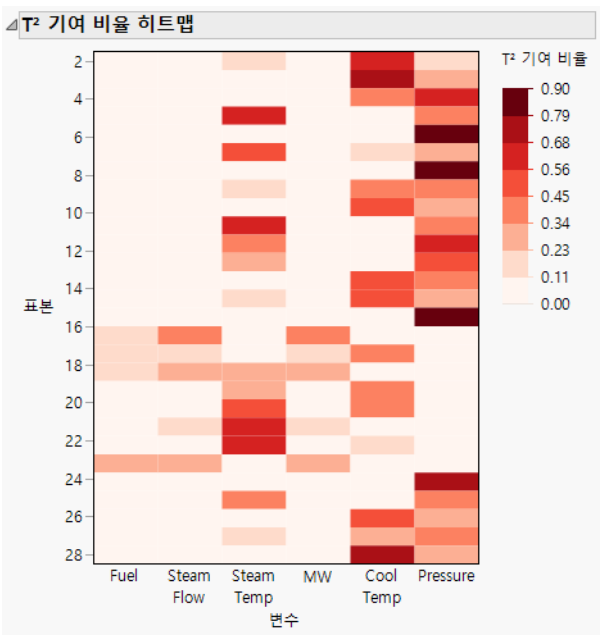
그림 11.4 Steam Flow 에 대한 개별 차트



개별 차트를 보면 대략 표본 17에서 증기 유량(Steam Flow)이 역전되는 것을 볼 수 있습니다. "관리 한계" 열 특성에 관리 한계를 지정한 경우 해당 값이 개별 차트에 적용됩니다.

7. "PCA 모형 기반 다변량 관리도" 보고서 창에서 "3 개 주성분에 대한 T^2" 의 빨간색 삼각형을 클릭하고 **기여 비율 히트맵**을 선택합니다.

그림 11.5 기여 비율 히트맵

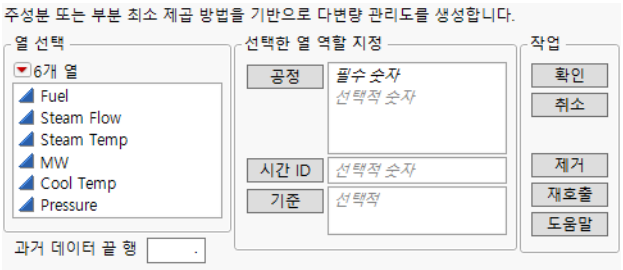


기여 비율 히트맵에서는 다른 행과 비교하여 16, 17, 18 및 23 행에 대한 기여 비율에 변화가 있음을 보여 줍니다. 일반적으로 Steam Temp, Cool Temp 및 Pressure 가 각 행의 T^2 값에 가장 많이 기여합니다.

모형 기반 다변량 관리도 플랫폼 시작

분석 > 품질 및 공정 > 모형 기반 다변량 관리도를 선택하여 MDMVCC(모형 기반 다변량 관리도) 플랫폼을 시작할 수 있습니다.

그림 11.6 모형 기반 다변량 관리도 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

모형 기반 다변량 관리도 플랫폼 시작 창에는 다음 옵션이 포함되어 있습니다.

공정 공정 열을 할당합니다. 자세한 내용은 " 데이터 형식 " 에서 확인하십시오.

시간 ID 표본을 식별하는 데 사용되는 열을 할당합니다. " 시간 ID" 를 할당하지 않은 경우 행 번호가 관측값을 식별합니다. " 시간 ID" 열이 시간인 경우 시간이 각 표본을 식별합니다. 그렇지 않은 경우에는 " 시간 ID" 의 숫자 값이 각 표본을 식별합니다.

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 조합에 대해 개별 보고서가 생성됩니다.

과거 데이터 끝 행 과거 데이터 끝 위치를 나타내는 행 번호를 지정합니다. 이렇게 하면 과거 데이터를 기반으로 차트 한계를 계산할 수 있습니다. 과거 데이터와 현재 데이터가 모두 차트에 표시됩니다. 과거 데이터는 단계 I 데이터라고 하고 현재 데이터는 단계 II 데이터라고도 합니다.

데이터 형식

MDMVCC 플랫폼에서는 다음과 같은 세 가지 형식의 데이터가 허용됩니다.

원시 데이터 연속형 공정 데이터를 사용하여 데이터의 주성분을 기반으로 하는 관리도를 생성합니다. 관리도의 기본 차원은 공정 변동의 85% 를 차지하는 주성분 수를 기반으로 합니다. 이 값은 주성분 고유값의 누적 백분율을 기반으로 합니다.

주성분 이전에 PCA(주성분 분석) 에서 저장한 주성분 열을 사용합니다 . 관리도의 기본 차원은 공정 변수로 지정된 성분 수입니다 .

부분 최소 제곱 스코어 데이터 이전에 PLS(부분 최소 제곱) 분석에서 저장한 스코어 열을 사용하여 스코어 열을 기반으로 하는 관리도를 생성합니다 . 관리도의 기본 차원은 공정 변수로 지정된 스코어 수입니다 .

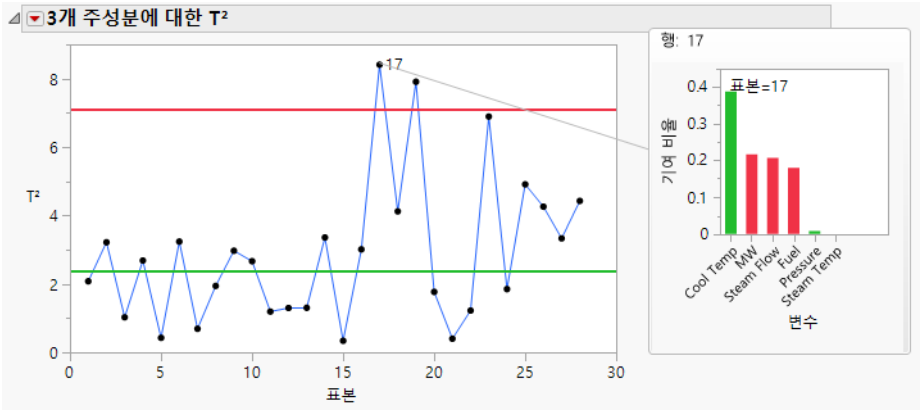
참고 :

- 빈도 열이나 가중치 열로 생성된 PCA 또는 PLS 모형은 지원되지 않습니다 .
- 과거 데이터로 생성된 PCA 또는 PLS 모형은 MDMVCC 시작 창에 지정된 것과 동일한 수의 과거 데이터 행을 사용해야 합니다 .
- 다변량 플랫폼 내에서 생성된 PCA 모형은 지원되지 않습니다 .

모형 기반 다변량 관리도 보고서

초기 모형 기반 다변량 관리도 보고서에는 T^2 관리도가 표시됩니다 . 차트의 가리키기 라벨 자체가 차트입니다 . 가리키기 라벨 차트를 클릭하면 더 큰 버전의 차트가 열립니다 . 차트에 따라 별도의 보고서 창이나 MDMVCC 보고서의 " 공정 진단 " 섹션에서 열립니다 . Graphlet 을 사용하여 데이터를 대화식으로 드릴다운할 수 있습니다 .

그림 11.7 가리키기 Graphlet 이 있는 MDMVCC 보고서



모형 기반 다변량 관리도 플랫폼 옵션

과거 요약 통계량 표시 과거 데이터로 지정된 행 또는 모든 행 (과거 데이터 행이 지정되지 않은 경우) 에 대한 요약 통계량을 표시하거나 숨깁니다 . 요약 통계량에는 공정 변수에 대한 단변량 평균 및 표준편차가 포함됩니다 . PCA 기반 차트의 경우 고유값과 고유 벡터가 표시됩니다 . PLS 스코어 기반 차트의 경우 스코어 표준편차와 스코어 적재가 표시됩니다 .

공정 모니터링

모니터링 그림 표시 선택된 공정 모니터링 그림을 표시하거나 숨깁니다 .

성분 설정 T^2, DModX 또는 SPE 그림의 성분 수를 설정할 수 있습니다 . 성분 수의 범위는 1 에서 유효한 고유 벡터 수 (PCA 기반 분석의 경우) 또는 1 에서 PLS 모형 요인 수 (PLS 기반 분석의 경우) 까지입니다 .

α 수준 설정 모든 관리도 한계에 사용되는 유의 수준을 조정할 수 있습니다 .

T^2 그림 T^2 그림을 표시하거나 숨깁니다 . T^2 통계량은 관측값이 PCA 또는 PLS 모형의 중심에서 얼마나 멀리 떨어져 있는지를 측정하는 다변량 변동의 요약 통계량입니다 .

정규화된 DModX 그림 정규화된 DModX 값의 그림을 표시하거나 숨깁니다 . DModX 는 각 관측값에서 PCA 또는 PLS 모형까지의 거리를 측정합니다 . DModX 값이 크면 관측값이 데이터의 기본 상관 구조에서 벗어났음을 나타냅니다 .

제곱 예측 오차 그림 SPE(제곱 예측 오차) 그림을 표시하거나 숨깁니다 . SPE 는 PCA 또는 PLS 모형의 잔차 제곱합을 측정합니다 . SPE 값이 크면 관측값이 데이터의 기본 상관 구조에서 벗어났음을 나타냅니다 .

스코어 그림 주성분 또는 부분 최소 제곱 요인의 스코어 그림을 표시하거나 숨깁니다 . 자세한 내용은 " 스코어 그림 " 에서 확인하십시오 .

공정 진단 (하나 이상의 진단 그림이 활성화된 경우 사용 가능) 진단 그림을 표시하거나 숨깁니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

그림 옵션

다음 옵션은 T^2, 정규화된 DModX, 제곱 예측 오차 및 스코어 그림에 적용됩니다. 그림을 선택하면 보고서 창의 "공정 진단" 섹션에 나타납니다.

한계 요약 표시 (스코어 그림에는 사용 불가능) 차트 아래의 보고서 테이블에서 관리도 한계 및 요약 데이터를 표시하거나 숨깁니다 .

기여도 히트맵 각 관측값의 변수 기여도를 기준으로 색상이 지정된 히트맵을 표시하거나 숨깁니다 .

기여 비율 히트맵 각 관측값이 전체 값에서 차지하는 비율로 표현된 변수 기여도를 기준으로 색상이 지정된 히트맵을 표시하거나 숨깁니다 .

선택된 표본에 대한 기여도 그림 (하나 이상의 점을 선택한 경우에만 사용 가능) 선택된 각 표본에 대해 전체 값에 대한 개별 성분의 기여도를 보여 주는 막대 차트를 표시하거나 숨깁니다 .

선택된 표본에 대한 기여 비율 그림 (하나 이상의 점을 선택한 경우에만 사용 가능) 선택된 각 표본에 대해 전체 값에서 차지하는 기여 비율로 표현된 개별 성분 기여도를 보여 주는 막대 차트를 표시하거나 숨깁니다 .

선택된 표본에 대한 평균 기여 비율 그림 (두 개 이상의 점을 선택한 경우에만 사용 가능) 선택된 각 표본에 대해 전체 값에 대한 개별 성분의 기여도 평균을 보여 주는 막대 차트를 표시하거나 숨깁니다 .

참고 : 값 또는 평균 값이 평균의 3 시그마 내에 있으면 막대가 녹색입니다 . 값 또는 평균 값이 평균의 3 시그마를 벗어나면 막대가 빨간색입니다 . " 관리 한계 " 열 특성에 관리 한계를 지정한 경우 해당 값에 따라 막대 색상이 적용됩니다 .

기여도 그림에는 다음 옵션을 사용할 수 있습니다 .

막대 정렬 막대를 기여도가 가장 큰 것부터 순서대로 정렬하거나 여러 그림에 대한 평균 기여도가 가장 큰 것부터 순서대로 정렬할 수 있습니다 .

막대 라벨 지정 값 또는 열 이름별로 막대에 라벨을 지정하거나 라벨을 제거 (" 라벨 없음 ") 할 수 있습니다 .

선택한 항목에 대한 관리도 선택한 열에 대한 관리도를 표시합니다 . " 관리 한계 " 열 특성에 관리 한계를 지정한 경우 해당 값이 개별 차트에 적용됩니다 .

산점도 행렬 (두 개 이상의 막대를 선택한 경우 사용 가능) 선택한 변수에 대한 산점도 행렬을 표시합니다 .

그림 제거 보고서 창에서 그림을 제거합니다 .

선택된 표본에 대한 정규화된 스코어 그림 (하나 이상의 점을 선택한 경우 스코어 그림에만 사용 가능) 선택한 각 표본에 대한 정규화된 스코어를 보여 주는 막대 차트를 표시하거나 숨깁니다 .

스코어 타원 범위 (성분이 두 개인 스코어 그림에만 사용 가능) 범위가 지정된 타원을 스코어 그림에 추가합니다 . 타원은 과거 데이터를 기반으로 합니다 . 단계 I 데이터와 단계 II 데이터가 모두 있는 경우 각 단계에 대한 타원이 추가되고 단계 II 타원은 파선으로 표시됩니다 .

점 연결 (스코어 그림에만 사용 가능) 스코어 그림의 데이터 점을 연결합니다 .

적재 표시 (스코어 그림에만 사용 가능) 행렬도 화살표를 사용하여 스코어 그림에 PCA 적재를 표시합니다 .

열 저장 각 그림에 대해 데이터 테이블에 값을 저장하는 다음 세 가지 옵션이 있습니다 .

값 저장 값 (T^2, 정규화된 DModX, SPE 또는 스코어) 을 새 데이터 테이블 열에 저장합니다 .

기여도 저장 기여도를 새 데이터 테이블 열에 저장합니다 .

기여 비율 저장 기여 비율을 새 데이터 테이블 열에 저장합니다 .

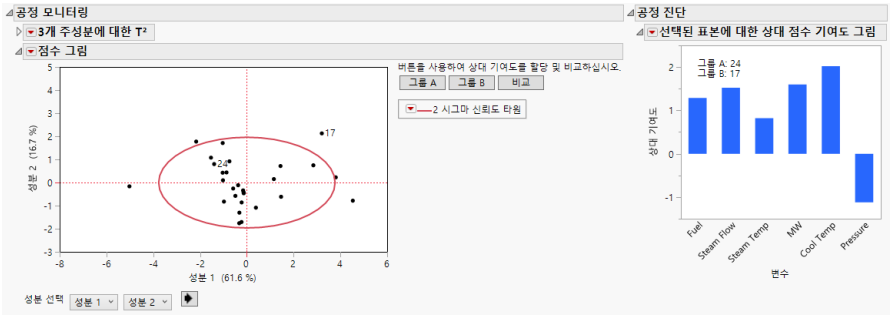
스코어 그림

스코어 그림에는 주성분 또는 부분 최소 제곱 요인에 대한 그림이 표시됩니다. 그림 아래의 컨트롤을 사용하여 스코어 그림에 표시된 성분을 변경할 수 있습니다.

그림 오른쪽의 버튼을 사용하여 두 점 그룹 간의 상대 기여도를 할당하고 비교할 수 있습니다. 상대 기여도는 둘 이상의 표본이 서로 어떻게 다른지 보여 줍니다. 상대 기여도는 기본 공정 변수의 어떤 변화로 인해 표본 그룹의 차이가 발생하는지 보여 줍니다. 한 가지 용도는 관리 상태 표본과 관리이탈 표본 간의 차이를 조사하는 것입니다.

그룹 A는 참조 그룹이고 그룹 B는 비교 그룹입니다. 각 그룹은 하나 이상의 점으로 구성될 수 있습니다. 참조 그룹을 할당하려면 하나 이상의 점을 선택한 후 "그룹 A"를 클릭합니다. 비교 그룹을 할당하려면 하나 이상의 점을 선택한 후 "그룹 B"를 클릭합니다. "비교"를 클릭하면 상대 스코어 기여도 그림이 표시됩니다.

그림 11.8 24 행을 기준으로 17 행에 대한 상대 기여도 그림이 있는 스코어 그림



모형 기반 다변량 관리도 플랫폼의 추가 예

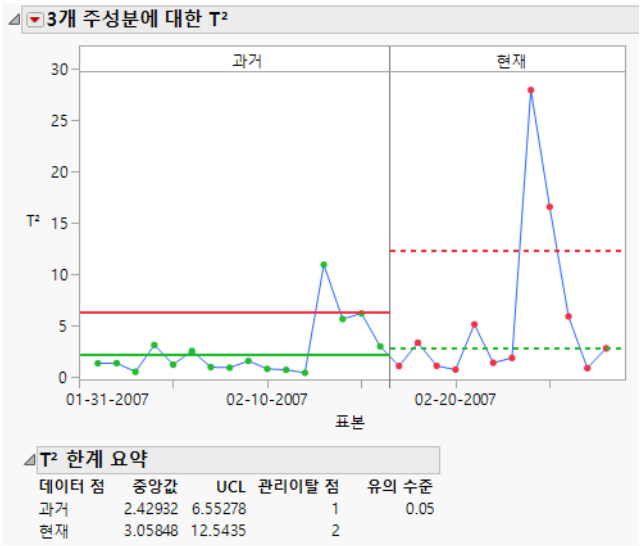
- " 과거 데이터를 사용하는 MDMVCC 의 예 "
- "PLS 모형을 사용하는 MDMVCC 의 예 "

과거 데이터를 사용하는 MDMVCC 의 예

이 예에서는 과거 데이터를 사용하여 현재 데이터에 대한 모니터링 한계를 설정하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Flight Delays.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 모형 기반 다변량 관리도**를 선택합니다 .
3. AA ~ WN 을 선택하고 **공정**을 클릭합니다 .
4. Flight Date 를 선택하고 **시간 ID** 를 클릭합니다 .
5. **과거 데이터 끝 행**에 16 을 입력합니다 .
6. **확인**을 클릭합니다 .

그림 11.9 과거 및 현재 데이터에 대한 T^2 차트



두 개의 한계 집합이 있습니다. 하나는 과거 데이터에 적용됩니다. 두 번째 한계 집합은 현재 데이터에 적용됩니다. 과거 데이터를 사용하여 두 한계 집합을 계산하는 방법에 대한 자세한 내용은 "한계"에서 확인하십시오.

팁 : 데이터 테이블에 관측값을 더 추가할 때 차트가 자동으로 업데이트되게 하려면 " 자동 재계산 " 을 설정하십시오 . " 자동 재계산 " 은 "PCA 모형 기반 다변량 관리도 " 의 빨간색 삼각형을 클릭하면 나타나는 메뉴에서 " 다시 실행 " 의 하위 옵션입니다 .

PLS 모형을 사용하는 MDMVCC 의 예

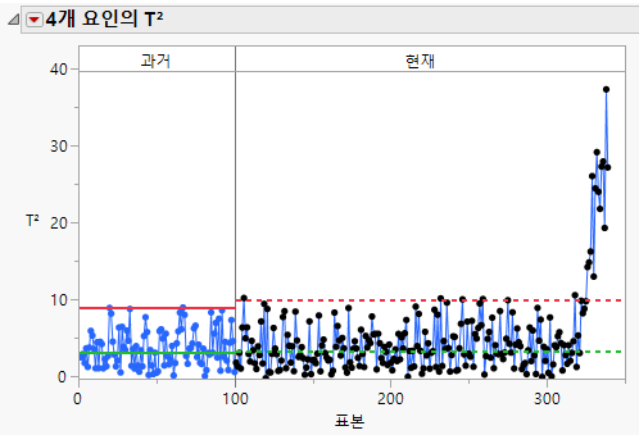
이 예에서는 PLS 모형을 사용하여 다변량 공정을 모니터링하는 방법을 보여 줍니다 . 14 개의 입력과 5 개의 품질 변수가 있는 공정을 예로 들어보겠습니다 . 공정을 설명하는 PLS 모형이 있고 이 모형을 사용하여 공정의 편차를 모니터링하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Polyethylene Process.jmp 를 엽니다 .

   이 데이터 테이블에는 14 개의 공정 변수와 5 개의 품질 (출력) 변수가 포함되어 있습니다 . 처음 100 개 행은 공정을 설명하기 위해 PLS 모형을 생성하는 데 사용된 과거 데이터입니다 . 이러한 행은 파란색으로 표시됩니다 . 나머지 239 개 행은 모형이 생성된 후 수집된 데이터입니다 .

   부분 최소 제곱 모형은 공정을 설명하는 4 개의 스코어 함수를 찾습니다 . 이러한 함수는 데이터 테이블의 X Score 1 Formula 열 ~ X Score 4 Formula 열에 저장됩니다 . PLS 모형을 생성하기 위해 Set current data as excluded 테이블 스크립트를 사용하여 과거 데이터 이후 수집된 239 개의 데이터 행을 제외합니다 . 그런 다음 Fit Partial Least Squares 테이블 스크립트를 사용하여 품질 변수를 공정 변수와 관련짓는 PLS 모형을 생성합니다 .
2. **분석 > 품질 및 공정 > 모형 기반 다변량 관리도**를 선택합니다 .
3. X Score 1 Formula ~ X Score 4 Formula 를 선택하고 **공정**을 클릭합니다 .
4. **과거 데이터 끝 행**을 100 으로 설정합니다 .
5. **확인**을 클릭합니다 .

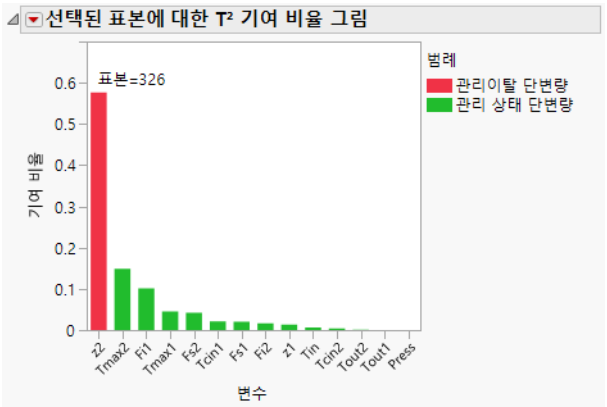
그림 11.10 T^2 차트



파란색 데이터 점은 과거 데이터를 나타냅니다. 검은색 데이터 점은 관리도가 만들어진 후 수집된 데이터 점입니다. 공정은 대략 표본 번호 326에서 역전이 시작되었습니다.

6. 관리 상한 위에 있는 표본 데이터 점을 커서로 가리켜 기여도 그림을 봅니다.
7. 관리 상한 위에 있는 데이터 점의 첫 번째 군집을 선택합니다. "4 개 요인의 T^2" 의 빨간색 삼각형을 클릭하고 **선택된 표본에 대한 평균 기여 비율 그림**을 선택합니다.
8. " 선택된 표본에 대한 T2 평균 기여 비율 그림 " 옆의 빨간색 삼각형을 클릭하고 **막대 정렬**을 선택합니다.

그림 11.11 평균 기여 비율 그림



기여도 그림은 PLS 모형 입력 변수의 관점에서 작성되었습니다. z2 와 Tmax2 가 공정 역전을 일으키는 것으로 보입니다. Tmax2 와 z2 는 관련되어 있습니다. Tmax2 는 반응기 온도이고 z2 는 Tmax2 온도의 위치입니다.

참고 : 요인에 대한 설명은 " 노트 " 열 특성에 기록되어 있습니다.

모형 기반 다변량 관리도 플랫폼에 대한 통계 상세 정보

모니터링 통계량

T^2

i개 관측값 각각의 T_i^2 값이 T^2 관리도에 표시됩니다. 과거 및 현재 데이터에 대해 k개의 성분이 있는 PCA 또는 PLS 모형의 T^2 값은 다음과 같이 정의됩니다.

$$T_i^2 = t_i^T S_k^{-1} t_i$$

다음은 각 요소에 대한 설명입니다.

t_i = i 번째 관측값에 대한 k 개 스코어의 벡터

$\mathbf{S}_k$ = 과거 관측값에 대한 k 개 스코어의 주대각 표본 공분산 행렬

PCA 모형의 경우 $\mathbf{S}_k$는 주대각 고유값 행렬입니다.

데이터 전처리 단계에서 데이터가 중심화되는 경우 k개 과거 스코어 벡터의 평균은 각각 0입니다. 이 단계는 PCA에서 상관계수 행렬 또는 공분산 행렬을 이용하거나, PLS에서 중심화와 함께 발생합니다. **X**가 중심화되지 않은 전처리 옵션의 경우 데이터가 사용자에 의해 중심화되었다고 가정하므로 k 개 스코어 벡터의 평균은 각각 0입니다. Hotelling T^2에 대한 자세한 내용은 Montgomery 연구 자료 (2013)에서 확인하십시오.

SPE

PCA 모형과 PLS 모형 둘 다 전처리된 **X** 행렬을 다음과 같이 분해할 수 있습니다.

$$X = T_k P_k^T + E$$

여기서 $T_k = (t_1,...,t_k)$는 k차원 스코어 행렬이고, $P_k = (p_1,...,p_k)$는 첫 번째 k 고유 벡터가 있는 행렬 (PCA 모형의 경우) 또는 적재 행렬 (PLS 모형의 경우)입니다. 이 PCA 또는 PLS 모형의 제곱 예측 오차는 SPE 관리도에 사용됩니다.

i개 관측값 각각의 SPE_i 값이 SPE 관리도에 표시됩니다. 제곱 예측 오차는 다음과 같이 정의됩니다.

$$SPE_i = e_i^T e_i = \sum_{j=1}^{p} e_{ij}^2$$

다음은 각 요소에 대한 설명입니다.

e_i = 관측값 i 에 대한 잔차 벡터

p = 변수 수

DModX

i개 관측값 각각의 $DModX_i$ 값이 DModX 관리도에 표시됩니다. 정규화된 DModX(모형까지의 거리)는 다음과 같이 정의됩니다.

$$DModX_i = \frac{SPE_i/(df_1)}{\left(\sum_{i=1}^{n} SPE_i\right)/(df_2)} = \frac{\sum_{j=1}^{p} e_{ij}^2/(df_1)}{\sum_{i=1}^{n}\sum_{j=1}^{p} e_{ij}^2/(df_2)} = \frac{\sum_{j=1}^{p} e_{ij}^2}{d}$$

다음은 각 요소에 대한 설명입니다.

e_{ij} = 관측값 i 와 변수 j 에 대한 잔차

$df_1 = p-k$

df_2 = 데이터가 중심화된 경우 $(n-k-1)(p-k)$, 데이터가 중심화되지 않은 경우 $(n-k)(p-k)$

n = 과거 데이터 관측값 수

k = PCA/PLS 성분 수

p = 변수 수

참고 : $DModX_i$ 는 $1/d$ 로 척도화된 SPE_i 와 같습니다 .

한계

시작 창에서 과거 행 수를 지정하지 않은 경우 모든 데이터가 과거 데이터로 처리됩니다. 자세한 내용은 "모형 기반 다변량 관리도 플랫폼 시작"(292 페이지)에서 확인하십시오.

T^2

과거 데이터의 UCL(관리 상한)은 베타 분포를 기반으로 하며 다음과 같이 정의됩니다.

$$UCL = \frac{(n-1)^2}{n}\beta\left[1-\alpha;\frac{k}{2};\frac{n-k-1}{2}\right]$$

다음은 각 요소에 대한 설명입니다.

n = 과거 데이터 관측값 수

k = PCA 또는 PLS 성분 수

$\beta\left[1-\alpha;\frac{k}{2};\frac{n-k-1}{2}\right]$ = 베타$\left[\frac{k}{2};\frac{n-k-1}{2}\right]$ 분포의 $(1-\alpha)$ 번째 분위수

현재 데이터의 UCL은 F 분포를 기반으로 하며 다음과 같이 정의됩니다.

$$UCL = \frac{k(n+1)(n-1)}{n(n-k)}F[1-\alpha;k;(n-k)]$$

다음은 각 요소에 대한 설명입니다.

n = 과거 데이터 관측값 수

k = PCA 또는 PLS 성분 수

$F(1-\alpha; k; n-k) = F(k; n-k)$ 분포의 $(1-\alpha)$ 번째 분위수

DModX

PCA 및 PLS 모형의 UCL은 F 분포를 기반으로 합니다. PCA 모형의 DModX UCL은 다음과 같이 정의됩니다.

$$UCL = F[1-\alpha; df_1; df_2]$$

다음은 각 요소에 대한 설명입니다.

$df_1 = p-k$

df_2 = 데이터가 중심화된 경우 $(n-k-1)(p-k)$, 데이터가 중심화되지 않은 경우 $(n-k)(p-k)$

n = 과거 데이터 관측값 수

k = PCA 성분 수

p = 변수 수

$F(1-\alpha; n-p-1; p-k) = F(n-p-1; p-k)$ 분포의 $(1-\alpha)$ 번째 분위수

PLS 모형의 DModX UCL은 다음과 같이 정의됩니다.

$$UCL = F[1-\alpha; h; nh]$$

다음은 각 요소에 대한 설명입니다.

$h = (2\hat{\mu}_{SPE}^2)/(\hat{\sigma}_{SPE}^2)$

$\hat{\mu}_{SPE}$ = SPE 의 과거 표본 평균

$\hat{\sigma}_{SPE}^2$ = SPE 의 과거 표본 분산

n = 과거 데이터 관측값 수

$F(1-\alpha; h; nh) = F(h; nh)$ 분포의 $(1-\alpha)$ 번째 분위수

SPE

PCA 모형의 SPE UCL은 다음과 같이 정의됩니다.

$$UCL = \theta_1 \left[1 - \frac{\theta_2 h_0 (1-h_0)}{\theta_1^2} + \frac{z_{1-\alpha}(2\theta_2 h_0^2)^{1/2}}{\theta_1} \right]^{1/h_0}$$

다음은 각 요소에 대한 설명입니다.

$$h_0 = 1 - 2\theta_1\theta_3/(3\theta_2^2)$$

$$\theta_1 = \sum_{a=1}^{k} \lambda_a$$

$$\theta_2 = \sum_{a=1}^{k} \lambda_a^2$$

$$\theta_3 = \sum_{a=1}^{k} \lambda_a^3$$

$\lambda_a = a$ 번째 고유값

k = PCA 성분 수

$z_{1-\alpha}$ = 표준 정규 분포의 (1–α) 번째 분위수

PCA 모형의 SPE 관리 한계에 대한 자세한 내용은 Jackson and Mudholkar 연구 자료(1979)에서 확인하십시오.

PLS 모형의 UCL은 카이제곱 분포를 기반으로 하며 다음과 같이 정의됩니다.

$$UCL = g\chi^2_{[1-\alpha;h]}$$

다음은 각 요소에 대한 설명입니다.

$$g = (\hat{\sigma}^2_{SPE})/(2\hat{\mu}_{SPE})$$

$$h = (2\hat{\mu}^2_{SPE})/(\hat{\sigma}^2_{SPE})$$

$\hat{\mu}_{SPE}$ = SPE 의 과거 표본 평균

$\hat{\sigma}^2_{SPE}$ = SPE 의 과거 표본 분산

$\chi^2(1-\alpha; h) = \chi^2(h)$ 분포의 (1–α) 번째 분위수

g 및 h 모수는 적률 방법으로 추정됩니다. PLS 모형의 SPE 관리 한계에 대한 자세한 내용은 Nomikos 연구 자료(1995)에서 확인하십시오.

기여도

T^2

p개 변수와 k개 성분이 있는 PCA 또는 PLS 모형의 T^2 기여도는 다음과 같이 계산됩니다.

$$
\begin{aligned}
T_i^2 &= t_i' S_k^{-1} t_i \\
&= \sum_{a=1}^{k} \frac{t_{ia}^2}{s_a} \\
&= \sum_{a=1}^{k} \frac{t_{ia}}{s_a} \sum_{j=1}^{p} r_{ja} x_{ij} \\
&= \sum_{j=1}^{p} \left(\sum_{a=1}^{k} \frac{t_{ia}}{s_a} r_{ja} x_{ij} \right)
\end{aligned}
$$

다음은 각 요소에 대한 설명입니다.

$t_i = i$ 번째 관측값에 대한 k 개 스코어의 벡터

$\mathbf{S}_k$ = 과거 관측값에 대한 k개 스코어의 주대각 표본 공분산 행렬. PCA 모형의 경우 $\mathbf{S}_k$는 주대각 고유값 행렬입니다 .

$s_a = \mathbf{S}_k$ 의 a 번째 주대각 요소

r_{ja} = PCA 모형의 경우 a번째 고유 벡터의 j번째 요소, PLS 모형의 경우 $\mathbf{R}_k$ 적재 행렬의 a번째 열. $\mathbf{R}_k$ 는 $\mathbf{T}_k = \mathbf{X}\mathbf{R}_k$ 가 되도록 스코어 행렬 $\mathbf{T}_k$ 와 $\mathbf{X}$ 행렬을 관련시키는 데 사용되는 행렬입니다 .

$x_{ij} = i$ 번째 관측값에 대한 j 번째 변수의 값

참고 : 마지막 합의 p 항은 변수 기여도입니다 .

각 변수의 기여도는 각 스코어에 대한 기여도 합이며, 정규화된 스코어 값에 따라 가중치가 부여됩니다. 정규화된 스코어 값이 크고 변수 기여도가 큰 경우 해당 변수는 T_i^2 에 대한 기여도가 큰 것으로 간주됩니다.

$$
\sum_{j=1}^{p} Cont(T_i^2)_j = T_i^2
$$

변수 j의 기여 비율은 다음과 같이 정의됩니다.

$$
ContProp(T_i^2)_j = \frac{Cont(T_i^2)_j}{\sum_{j=1}^{p} Cont(T_i^2)_j}
$$

참고 : JMP 에서는 T^2 기여 비율을 계산할 때 음의 기여도를 0 으로 설정합니다 . PCA 및 PLS 모형에서 **X** 투영 중에 변수의 교호작용으로 인해 음의 기여도가 발생할 수 있습니다 . 전체 양의 기여도에서 큰 비율을 나타내는 변수 기여도를 식별하기 위해 음의 기여도를 영점화합니다 .

PCA 기여도 및 음의 기여도에 대한 자세한 내용은 Kourti and MacGregor 연구 자료 (1996) 에서 확인하십시오 . PLS 기여도에 대한 자세한 내용은 Li et al. 연구 자료 (2009) 에서 확인하십시오 .

DModX

PCA 및 PLS 모형에서 $DModX_i$에 대한 변수 j의 기여도는 다음과 같이 정의됩니다.

$$Cont(DModX_i)_j = \frac{e_{ij}}{\sqrt{d}}$$

다음 식에 따라

$$DModX_i = \sum_{j=1}^{p} \frac{e_{ij}^2}{d}$$

변수 j의 기여 비율은 다음과 같이 정의됩니다.

$$ContProp(DModX_i)_j = \frac{e_{ij}^2/d}{DModX_i}$$

SPE

PCA 및 PLS 모형에서 SPE_i에 대한 변수 j의 기여도는 다음과 같이 정의됩니다.

$$Cont(SPE_i)_j = e_{ij}$$

다음 식에 따라

$$SPE_i = \sum_{j=1}^{p} e_{ij}^2$$

변수 j의 기여 비율은 다음과 같이 정의됩니다.

$$ContProp(SPE_i)_j = \frac{e_{ij}^2}{SPE_i}$$

스코어 기여도

스코어 기여도 계산은 T^2 기여도와 동일하지만 스코어 그림에서 선택한 차원에 대해서만 계산됩니다.

스코어 그림 그룹 비교

스코어 그림 그룹 비교의 경우 변수 j에 대한 상대 스코어 기여도는 그룹 B의 평균 기여도와 그룹 A의 평균 기여도 간의 차이입니다.

$$\sum_{i \in B} \frac{Cont(T_i^*)_j}{n_b} - \sum_{i \in A} \frac{Cont(T_i^*)_j}{n_a}$$

다음은 각 요소에 대한 설명입니다.

T^*_i = 스코어 그림에 표시된 차원에 해당하는 열이 있는 스코어 행렬의 i 번째 행

A = 그룹 A 의 관측값 집합

B = 그룹 B 의 관측값 집합

n_a = 그룹 A 의 관측값 수

n_b = 그룹 B 의 관측값 수

스코어 그림 적재

스코어 그림에 표시된 적재는 PCA 고유 벡터 또는 PLS X 스코어 적재(**R** 행렬)를 기반으로 합니다. 이러한 적재는 스코어의 최대 절대값으로 척도화됩니다. 적재를 스코어 그림에 그래프로 나타내기 위해 척도화가 수행됩니다. 적재는 각 변수가 각 스코어에 미치는 영향을 대략적으로 보여 줍니다.

12 장

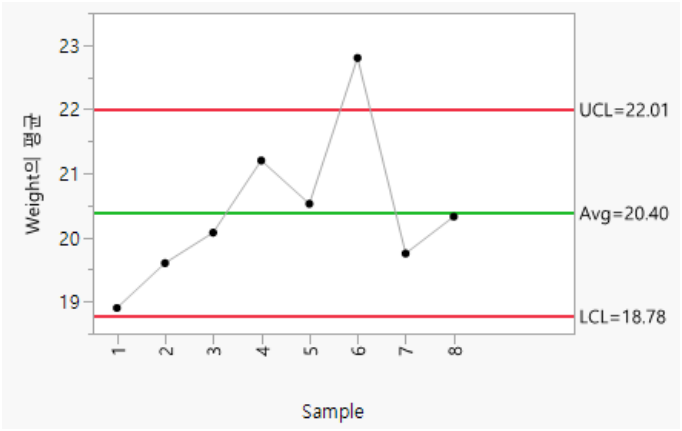
레거시 관리도

계량형 및 계수형 관리도 생성

관리도는 공정 변동을 모니터링하기 위한 그래픽 및 분석 도구입니다. 공정 자체에서 발생하는 고유 변동은 관리 한계 집합을 사용하여 정량화할 수 있습니다. 관리 한계는 우연 원인 변동과 특수 원인 변동을 구별하는 데 도움이 됩니다. 일반적으로 특수 원인 변동을 식별 및 제거하기 위한 작업이 수행됩니다. 공정의 우연 원인 변동을 정량화하는 것도 중요합니다. 이는 공정 능력을 결정하기 때문입니다.

JMP 의 레거시 관리도 플랫폼에서는 다양한 관리도는 물론 런 차트 , V- 마스크 CUSUM 차트 및 가중 이동 평균 차트도 제공합니다 . 공정 개선 이니셔티브를 지원하기 위해 대부분의 관리도 옵션은 프로젝트의 여러 단계에 대한 개별 관리도를 동일한 차트에 표시합니다 .

그림 12.1 관리도 예제



목차

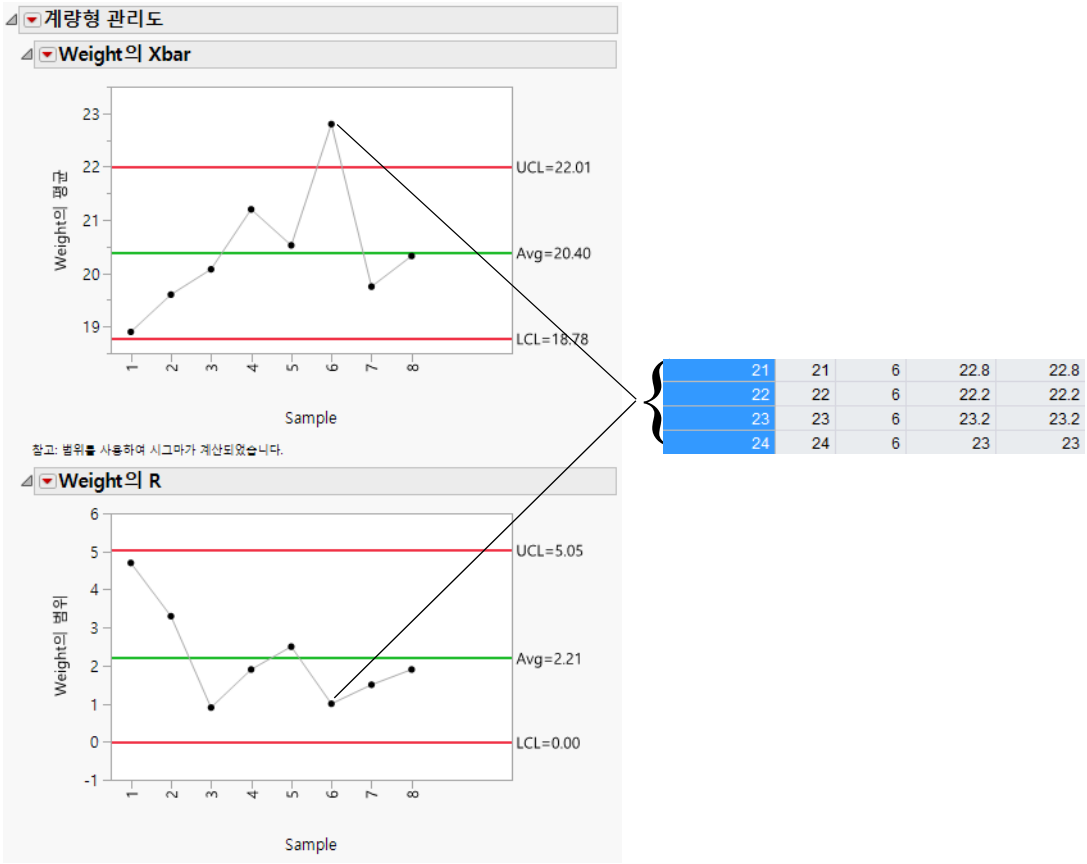
레거시 관리도의 예 309
레거시 관리도 유형 310
계량형 관리도 310
계수형 관리도 313
레거시 관리도 플랫폼 시작 314
공정 정보 315
한계 규격 317
지정된 통계량 317
레거시 관리도 보고서 318
V- 마스크 CUSUM 차트 보고서 320
양측 V-마스크 CUSUM 차트 해석 320
단측 CUSUM 차트 해석 322
레거시 관리도 플랫폼 옵션 322
레거시 관리도에 대한 창 옵션 322
레거시 관리도에 대한 차트 옵션 325
V-마스크 CUSUM 관리도에 대한 차트 옵션 327
한계 저장 및 검색 328
제외 , 삭제 및 숨겨진 표본 331
관리도 플랫폼의 추가 예 333
사전 집계 차트 예제 333
V-마스크 CUSUM 차트 예제 335
단측 CUSUM 차트 예제 337
UWMA 차트 예제 338
관리도 플랫폼에 대한 통계 상세 정보 339
중앙값 이동 범위 차트에 대한 관리 한계 339
공정 능력 분석에 대한 통계 상세 정보 339
V-마스크 CUSUM 관리도에 대한 통계 상세 정보 345
가중 이동 평균 차트에 대한 통계 상세 정보 348

레거시 관리도의 예

이 예에서는 Quality Control 샘플 데이터 폴더의 Coating.jmp 샘플 데이터 테이블을 사용합니다 (ASTM Manual on Presentation of Data and Control Chart Analysis 에서 가져옴). 관심 있는 품질 특성은 Weight 열입니다 . 4 개의 부분군 표본이 선택됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Coating.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 레거시 관리도 > Xbar** 를 선택합니다 .
   Xbar 및 **R** 차트 유형이 선택되어 있습니다 .
3. Weight 를 선택하고 **공정**을 클릭합니다 .
4. Sample 을 선택하고 **표본 라벨**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 12.2 Coating(코팅) 데이터에 대한 계량형 차트



공정에 대한 Xbar 차트와 R 차트가 그림 12.2에 표시되어 있습니다. 표본 6은 공정이 통계적 관리 상태가 아님을 나타냅니다. 표본 값을 확인하려면 두 관리도 중 하나에서 표본 6 요약 점을 클릭합니다. 데이터 테이블에서 해당 행이 강조 표시됩니다.

참고 : S 차트를 Xbar 차트와 함께 선택한 경우 표준편차를 기반으로 Xbar 차트의 한계가 결정됩니다 . 그렇지 않은 경우 Xbar 차트의 한계는 범위를 기반으로 합니다 .

레거시 관리도 유형

레거시 관리도는 크게 다음 두 가지 범주로 분류됩니다.

- " 계량형 관리도 "— IR, Xbar, 런 차트 , Levey-Jennings, 사전 집계 , CUSUM, UWMA 및 EWMA
- " 계수형 관리도 "— P, NP, C 및 U

계량형 관리도

계량형 관리도는 차트에 표시된 부분군 요약 통계량에 따라 다음과 같이 분류됩니다 .

- **IR** 을 선택하면 다음과 같은 추가 차트 유형이 제공됩니다 .
  - 개별 측정값 차트에는 개별 측정값이 표시됩니다 . 각 부분군 표본에 대해 하나의 측정값만 사용 가능한 경우 이 차트가 적합합니다 .
  - 이동 범위 차트에는 두 개 이상의 연속 측정값에 대한 이동 범위가 표시됩니다 . 자세한 내용은 " 이동 범위 (평균) 차트 " 에서 확인하십시오 .
- **Xbar** 차트에는 부분군 평균이 표시됩니다 . 이 항목을 선택하면 다음과 같은 추가 차트 유형이 제공됩니다 .
  - **R** 차트에는 부분군 범위 (최대값 – 최소값) 가 표시됩니다 .
  - **S** 차트에는 부분군 표준편차가 표시됩니다 .

  연속형 척도로 측정된 품질 특성에 대해 일반적인 분석을 수행하면 해당하는 R 또는 S 차트 위에 평균 차트를 정렬하여 공정 평균과 변동성이 모두 표시됩니다 .
- **런 차트**는 데이터를 연결된 일련의 점으로 표시합니다. 또한 창에서 또는 스크립트를 통해 **표본 라벨** 역할이 사용되는 경우 그룹 평균도 런 차트에 표시될 수 있습니다 .
- **Levey-Jennings** 차트에는 장기 시그마를 기반으로 한 관리 한계와 공정 평균이 표시됩니다 . 관리 한계는 중심선에서 3*s* 떨어진 위치에 배치됩니다 . Levey-Jennings 차트에 대한 표준편차 *s* 는 분포 플랫폼의 표준편차와 동일한 방식으로 계산됩니다 .
- **사전 집계** 차트에는 부분군 평균과 표준편차가 표시됩니다 . 자세한 내용은 " 사전 집계 차트 " 에서 확인하십시오 .
- **CUSUM** 차트에는 부분군 또는 개별 측정값과 목표값 간의 누적합이 표시됩니다 . 자세한 내용은 "V- 마스크 CUSUM 관리도 " 에서 확인하십시오 .

- UWMA 차트에는 지정된 수의 측정값에 대한 균등 가중 이동 평균이 표시됩니다 . 자세한 내용은 " 균등 가중 이동 평균 차트 " 에서 확인하십시오 .
- EWMA 차트에는 가중치가 지정된 모든 측정값의 지수 가중 이동 평균이 표시됩니다 . 자세한 내용은 " 지수 가중 이동 평균 차트 " 에서 확인하십시오 .

이동 범위 (평균) 차트

이동 평균 차트에서는 부분군 평균 대신 개별 관측값을 사용하여 평균을 구할 수 있습니다. 그러나 개별 측정값에 대한 이동 평균 차트는 개별 측정값에 대한 관리도 또는 개별 측정값이 표시된 이동 범위와 다릅니다.

이동 범위(평균) 차트에는 두 개 이상의 연속 측정값에 대한 이동 범위가 표시됩니다. 이동 범위는 **범위 길이** 상자에 입력하는 연속 측정값 수에 대해 계산됩니다. 기본 범위 길이는 2입니다. 이동 범위는 상관관계가 있으므로 차트를 해석할 때 주의해야 합니다.

중앙값 이동 범위 차트도 사용할 수 있습니다. 중앙값 이동 범위 차트와 개별 측정값 차트를 선택하는 경우 개별 측정값 차트의 한계는 평균 이동 범위 대신 중앙값 이동 범위를 시그마로 사용합니다.

사전 집계 차트

데이터가 동일한 공정 단위의 반복 측정값으로 구성된 경우 이를 해당 단위에 대한 하나의 측정값으로 결합할 수 있습니다. 데이터에 각 공정 또는 측정 단위에 대한 반복 측정값이 포함된 경우 외에는 사전 집계하지 않는 것이 좋습니다.

사전 집계는 선택한 표본 크기 또는 표본 라벨을 기반으로 공정 열을 표본 평균 또는 표준편차로 요약합니다. 그런 다음 시작 창에서 선택한 옵션에 따라 요약 데이터를 차트에 표시합니다. 시작 창에서 해당하는 상자를 선택하여 공정 능력 분석을 추가할 수도 있습니다.

사전 집계 시작 창에는 다음과 같은 차트 유형 옵션이 있습니다 .

- 그룹 평균에 대한 개별값
- 그룹 표준편차에 대한 개별값
- 그룹 평균에 대한 이동 범위
- 그룹 표준편차에 대한 이동 범위
- 그룹 평균에 대한 중앙값 이동 범위
- 그룹 표준편차에 대한 중앙값 이동 범위

이동 범위 차트 유형에 사용되는 범위 길이를 설정하는 옵션도 있습니다.

V- 마스크 CUSUM 관리도

V-마스크 CUSUM(누적합) 관리도에는 부분군 또는 개별 측정값과 목표값 간의 누적합이 표시됩니다. V-마스크 CUSUM 차트는 공정 평균에서 지속적으로 발생하는 작은 변화를 감지하여 공정이 통계적 관리 상태인지 여부를 판단하는 데 도움이 될 수 있습니다. 이에 비해 표준

Shewhart 관리도는 2 또는 3 시그마 변화와 같이 측정값에서 갑작스럽게 발생하는 큰 변화를 감지할 수 있지만 1 시그마 변화와 같은 작은 변화를 발견하는 데는 효과가 떨어집니다.

CUSUM 메뉴 선택 항목에는 V- 마스크 누적합 차트에 대한 옵션이 있습니다 . "K 시그마 " 외에 다음 항목도 지정합니다 .

- 양측 차트의 경우 V- 마스크 원점과 V- 마스크 상위 또는 하위 기준선 사이의 세로 거리 *h*. 단측 차트의 경우 H 는 결정 구간입니다 . H 를 표준 오차의 배수로 선택합니다 .
- 참조 값 $k(k > 0)$

누적합 관리도의 또 다른 유형에는 테이블 형식 CUSUM 차트가 있습니다 . 테이블 형식 CUSUM 차트를 생성하려면 "CUSUM 관리도 " 의 내용을 참조하십시오 . 다음을 포함하여 다양한 이유로 V- 마스크 차트 대신 테이블 형식 CUSUM 차트가 권장됩니다 .

- V-마스크는 단순히 마지막 관측값 위에 추가되는 것이 아니라 각 관측값과 함께 이동해야 합니다 .
- V- 마스크 절차의 누적합은 목표에 적합한 공정의 경우에도 그래프 중심에서 멀리 떨어질 수 있습니다 .

주의 : Montgomery(2013) 에서는 "V- 마스크 절차를 사용하지 말 것을 " 권장합니다 .

균등 가중 이동 평균 차트

UWMA(균등 가중 이동 평균) 차트의 각 점은 현재 부분군 평균을 포함하여 *w* 개의 가장 최근 부분군 평균의 평균입니다. 새 부분군 표본을 얻으면 *w* 개의 이전 부분군 평균 중 가장 오래된 것을 삭제하고 최신 부분군 평균을 포함하여 다음 이동 평균이 계산됩니다. 상수 *w* 를 이동 평균의 **거리**라고 합니다.

UWMA 시작 창에서 "K 시그마 " 와 " 알파 " 외에 다음 항목도 지정합니다 .

- 이동 평균을 구하기 위해 포함할 부분군 수를 나타내는 이동 평균 거리 (*w* 라고도 함). 이동 평균 거리 (*w*) 가 클수록 UWMA 선이 더 평활해지고 변화의 크기를 적게 반영합니다 . 즉 , *w* 값이 클수록 더 작은 변화에 대비할 수 있습니다 . 자세한 내용은 "UWMA 차트에 대한 관리 한계 " 에서 확인하십시오 .

지수 가중 이동 평균 차트

EWMA(지수 가중 이동 평균) 차트의 각 점은 현재 부분군 표본의 평균을 포함하여 이전의 모든 부분군 평균에 대한 가중 평균입니다. 가중치는 시간에 따라 지수적으로 감소합니다.

참고 : EWMA(지수 가중 이동 평균) 차트는 GMA(기하 이동 평균) 차트라고도 합니다 .

EWMA 시작 창에서 "K 시그마 " 와 " 알파 " 외에 다음 항목도 지정합니다 .

- 가중치 모수 , 즉 현재 부분군 표본 평균에 할당된 가중치 ($0 < \text{weight} \leq 1$). 가중치 값이 작으면 작은 변화가 발생하지 않도록 대비할 수 있습니다 . 자세한 내용은 "EWMA 차트에 대한 관리 한계 " 에서 확인하십시오 .

팁 : 최신 EWMA 관리도 플랫폼에 대한 내용은 "EWMA 관리도 " 에서 확인하십시오 .

계수형 관리도

이전 유형의 차트에서는 측정 데이터가 공정 변수였습니다. 이 유형의 데이터는 주로 연속형이며 차트는 연속형 데이터에 대한 이론을 기반으로 생성됩니다. 개수 데이터라는 다른 유형의 데이터도 있습니다. 이 경우 관심 변수는 부분군당 불량 또는 결함 수를 나타내는 이산형 개수입니다. 이산형 개수 데이터의 경우 이항 및 Poisson 모형을 기반으로 하므로 계수형 차트를 적용할 수 있습니다. 개수는 부분군별로 측정되므로 차트를 비교할 때 차트 간에 부분군의 항목 수가 비슷한지 확인해야 합니다. 계수형 차트는 계량형 차트와 마찬가지로 차트에 표시된 부분군 표본 통계량에 따라 분류됩니다.

사용할 계수형 차트 결정

각 항목은 다음과 같이 적합 또는 부적합으로 판단됩니다 .

p 차트 불량품 **비율**을 표시합니다 .

np 차트 불량품 **개수**를 표시합니다 .

각 항목에 대해 다음과 같이 결함 수가 계산됩니다 .

c 차트 결함 **수**를 표시합니다 .

u 차트 결함 **비율**을 표시합니다 .

계수형 차트의 경우 결함 수 또는 불량품 비율이 포함된 열을 공정 변수로 지정합니다 . 열에 0 에서 1 사이의 정수가 아닌 값이 포함된 경우 외에는 데이터가 개수로 해석됩니다 .

- P 차트에는 부분군 크기가 변동 가능한 표본에서 부적합품 (불량) 의 비율이 표시됩니다 . P 차트의 각 부분군은 N_i 개 항목으로 구성되며 각 항목이 적합 또는 부적합으로 평가되므로 부분군의 최대 부적합품 수는 N_i 개입니다 .
- NP 차트에는 부분군 표본의 부적합품 (불량) 수가 표시됩니다 . NP 차트의 각 부분군은 N_i 개 항목으로 구성되며 각 항목이 적합 또는 부적합으로 평가되므로 부분군 i 의 최대 부적합품 수는 N_i 개입니다 .

  참고 : P 또는 NP 차트에 " 시그마 " 열 특성을 사용하려면 값이 비율과 같아야 합니다 . JMP 에서는 시그마를 비율과 표본 크기의 함수로 계산합니다 .

- C 차트에는 대개 하나의 검사 단위로 구성된 (반드시 그런 것은 아님) 부분군 표본의 결점 (결함) 수가 표시됩니다 .

주의 : C 차트의 경우 " 표본 크기 " 또는 " 상수 크기 " 를 지정하지 않으면 " 표본 라벨 " 이 표본 크기로 사용됩니다 .

- U 차트에는 검사 단위 수가 변할 수 있는 각 부분군 표본의 결점 (결함) 수가 표시됩니다 .

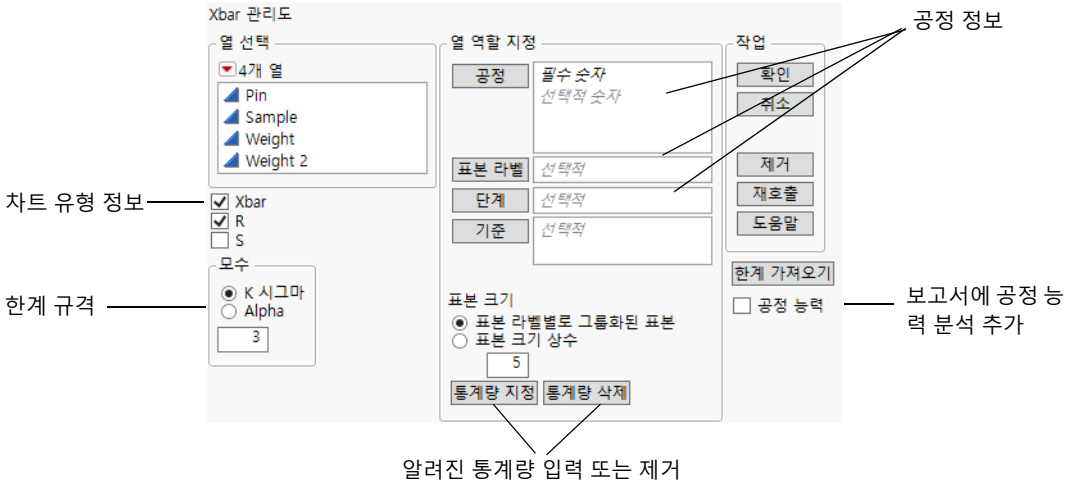
주의 : U 차트의 경우 " 단위 크기 " 또는 " 상수 크기 " 를 지정하지 않으면 " 표본 라벨 " 이 단위 크기로 사용됩니다 .

레거시 관리도 플랫폼 시작

분석 > 품질 및 공정 > 레거시 관리도를 선택하여 레거시 관리도 플랫폼을 시작하면 그림 12.3 에 나오는 것과 유사한 시작 창이 나타납니다 . 특정 컨트롤은 선택한 차트 유형에 따라 다릅니다 . 초기에는 다음과 같은 유형의 정보가 창에 표시됩니다 .

- 측정 변수 선택을 위한 " 공정 정보 "
- 차트 유형 정보 (자세한 내용은 " 레거시 관리도 유형 " 참조)
- " 한계 규격 "
- " 지정된 통계량 "

그림 12.3 Xbar 관리도 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

공정 정보

시작 창에 현재 데이터 테이블의 열 목록이 표시됩니다 . 여기서 분석할 변수와 부분군 표본 크기를 지정합니다 .

공정

공정 역할은 차트 생성에 사용할 변수를 선택합니다 .

- 계량형 차트의 경우 측정값을 공정으로 지정합니다 .
- 계수형 차트의 경우 결함 수 또는 불량품 비율을 공정으로 지정합니다 . 0 에서 1 사이의 정수가 아닌 값이 포함된 경우 외에는 데이터가 개수로 해석됩니다 .

참고 : 데이터 테이블의 행은 관측값을 수집한 순서대로 정렬되어야 합니다 . **표본 라벨** 변수를 지정한 경우에도 이에 따라 관측값을 정렬해야 합니다 .

표본 라벨

표본 라벨 역할을 사용하면 해당 값으로 가로 축에 라벨을 지정하고 , 동일하지 않은 부분군 크기도 식별할 수 있는 변수를 지정할 수 있습니다 . 표본 라벨 변수를 지정하지 않으면 표본이 부분군 표본 번호로 식별됩니다 .

- 표본 부분군 크기가 같으면 **표본 크기 상수** 옵션을 선택하고 텍스트 상자에 크기를 입력합니다 . 표본 라벨 변수를 입력한 경우 해당 값은 가로 축에 라벨을 지정하는 데 사용됩니다 . 표본 크기는 표본에 결측값이 있는지 여부에 관계없이 한계 계산에 사용됩니다 .
- 표본 부분군의 행 수가 같지 않거나 결측값이 있고 각 표본을 식별하는 열이 있으면 **표본 라벨별로 그룹화된 표본** 옵션을 선택한 후 표본 식별 열을 표본 라벨로 입력합니다 .

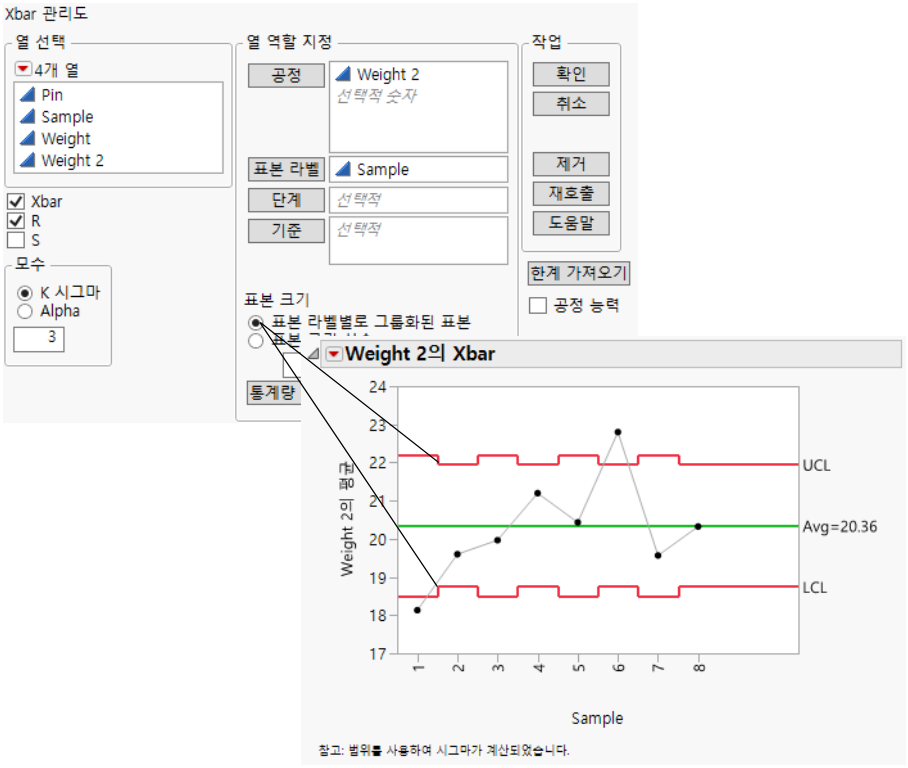
계수형 차트(P, NP, C 및 U 차트)의 경우 이 변수는 부분군 표본 크기입니다. 선택 항목에 따라 **표본 크기**, **상수 크기** 또는 **단위 크기**를 포함한 추가 옵션이 시작 창에 나타납니다. 계량형 차트에서는 표본을 식별합니다. 차트 유형이 **IR**이면 **범위 길이** 텍스트 상자가 나타납니다. 범위 길이는 이동 범위가 계산되는 연속 측정값 수를 지정합니다.

참고 :

- 데이터 테이블의 행은 관측값을 수집한 순서대로 정렬되어야 합니다 . **표본 라벨** 변수를 지정한 경우에도 이에 따라 관측값을 정렬해야 합니다 .
- **상수 크기** 값의 정수가 아닌 부분은 잘립니다 . 부분군 표본 크기가 정수가 아닌 상수인 경우 상수 값 열을 지정해야 합니다 .

그림 12.4 의 그림에서는 Quality Control 샘플 데이터 폴더의 Coating.jmp 샘플 데이터를 사용하여 부분군 표본 크기가 같지 않은 공정에 대한 Xbar 차트를 보여 줍니다 .

그림 12.4 부분군 표본 크기가 같지 않은 계량형 차트



단계

단계 역할을 사용하면 여러 단계 또는 섹션을 식별하는 열을 지정할 수 있습니다. **단계**는 데이터 테이블의 연속 관측값 그룹입니다. 예를 들어 단계는 새 공정이 생산에 도입된 후 연속된 변화를 겪는 기간에 해당할 수 있습니다. 단계는 지정된 단계 변수의 각 수준에 대해 새 시그마, 한계 집합, 영역 및 검정 결과를 생성합니다.

Xbar, R, S, IR, P, NP, C, U, 사전 집계 및 Levey-Jennings 차트의 창에 **단계** 변수 버튼이 나타납니다. 단계 변수가 지정되면 행별로 단계 변수를 검토하여 각 행이 속한 단계를 식별합니다. 한계 파일에 저장하면 각 단계에 대해 계산된 시그마 및 특정 한계가 표시됩니다.

기준

기준 역할은 열에 나타나는 각 값에 대해 별도의 분석을 생성하기 위해 변수를 식별합니다.

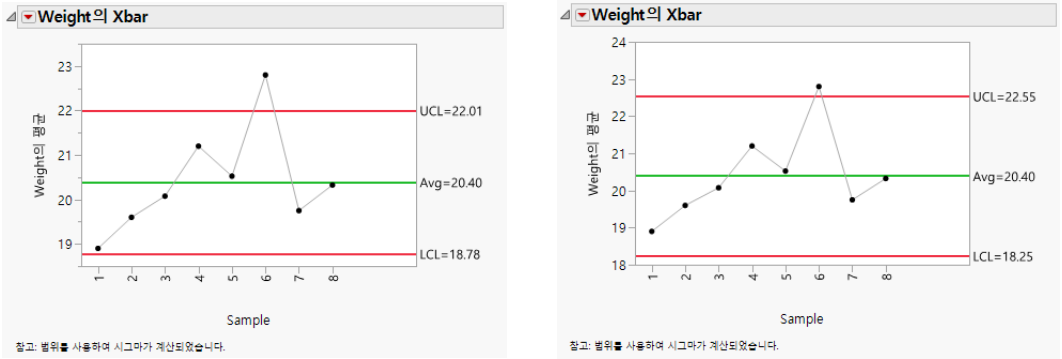
한계 규격

*k*에 대한 값(**K 시그마**) 또는 α에 대한 확률(**알파**)을 입력하거나, 공정 열 특성 또는 이전에 생성한 한계 테이블에서 한계 값을 검색하여 관리 한계 계산을 지정할 수 있습니다. 한계 테이블 및 **한계 가져오기** 버튼에 대한 내용은 "한계 저장 및 검색" 섹션에서 설명합니다. **K 시그마** 또는 **알파** 규격이 있어야 합니다. **K 시그마**의 창 기본값은 3입니다.

K 시그마

K 시그마 모수 옵션을 사용하면 표본 표준 오차의 배수로 관리 한계를 지정할 수 있습니다. **K 시그마**는 중심선으로 표시되는 기대값의 위/아래에 있는 *k* 표본 표준 오차로 관리 한계를 지정합니다. *k*(시그마 수)를 지정하려면 **K 시그마** 라디오 버튼을 클릭하고 텍스트 상자에 양수 *k* 값을 입력합니다. *k* 에 일반적으로 사용되는 값은 3 표준편차를 나타내는 3입니다. 그림 12.5에 표시된 예에서는 Coating.jmp 데이터에 대한 Xbar 차트를 **K 시그마** = 3 및 **K 시그마** = 4로 그린 관리 선과 비교합니다.

그림 12.5 K 시그마 = 3(왼쪽) 및 K 시그마 = 4(오른쪽) 관리 한계



알파

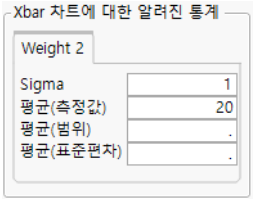
알파 모수 옵션은 공정이 관리 상태라고 가정할 때 단일 부분군 통계량이 관리 한계를 벗어날 확률 α의 관점에서 관리 한계(**확률 한계**라고도 함)를 지정합니다. 알파를 지정하려면 **알파** 라디오 버튼을 클릭하고 원하는 확률을 입력합니다. α의 경우 0.01 또는 0.001을 선택하는 것이 합리적입니다. 정규성 및 알려진 관리 상태 모수를 가정한 Xbar 차트의 경우 **K 시그마** = 3과 동등한 **알파** 값은 0.0027입니다.

지정된 통계량

공정 변수를 지정한 후 관리도 시작 창에서 **통계량 지정**(사용 가능한 경우) 버튼을 클릭하면 편집 가능한 필드가 있는 탭이 창 아래쪽에 추가됩니다. 이를 통해 공정 변수에 대한 과거 통계량, 즉 과거 데이터에서 얻은 통계량을 입력할 수 있습니다. 관리도 플랫폼에서는 이러한 항목을 사

용하여 관리도를 생성합니다. 여기에 나오는 예에서는 공정 변수의 표준편차를 1, 평균 측정값을 20으로 각각 보여 줍니다.

그림 12.6 통계량 지정의 예



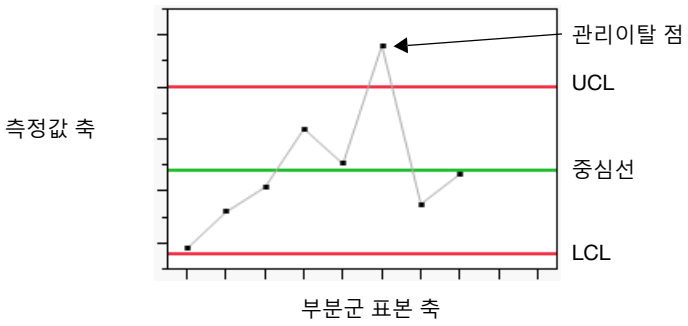
참고 : 사용자가 지정한 평균의 경우 그림에서 μ0 라벨이 지정됩니다 .

관리도 시작 창에서 **공정 능력** 옵션을 선택하면 (그림 12.3) 플랫폼이 시작될 때 규격 한계를 입력할 수 있는 창이 나타납니다. 선택한 관리도의 표준편차가 창으로 전달되고 기본 옵션인 "지정된 시그마" 값으로 나타납니다. 규격 한계를 입력하고 "확인"을 클릭하면 관리도 옆의 동일한 창에 공정 능력 분석 결과가 나타납니다. 공정 능력 지수를 계산하는 방법에 대한 자세한 내용은 "정규 분포에 대한 공정 능력 지수"에서 확인하십시오.

레거시 관리도 보고서

분석을 통해 공정이 통계적 관리 상태인지 여부를 판별하는 데 사용할 수 있는 차트가 생성됩니다 . 보고서는 선택한 차트 유형에 따라 달라집니다 . 그림 12.7 에는 간단한 관리도의 일부가 표시되어 있습니다 . 데이터 테이블에서 데이터가 추가 또는 변경되면 관리도가 동적으로 업데이트됩니다 .

그림 12.7 관리도의 예



참고 : 데이터 테이블에서 제외된 행은 런 차트 , P 차트 , U 차트 및 C 차트에서도 숨겨집니다 .

관리도에는 다음과 같은 특성이 있습니다 .

- 차트에 표시된 각 점은 개별 공정 측정값 또는 요약 통계량을 나타냅니다 . 그림 12.7 에서 점은 측정값 표본의 평균을 나타냅니다 .

  부분군은 추론에 근거하여 **합리적으로** 선택해야 합니다 . 즉 , 부분군 **간의** 공정 변화가 참일 확률을 최대화하도록 선택해야 합니다 . 대개 가장 효과적인 그룹화 전략을 결정하려면 공정에 대한 지식이 필요합니다 . 자세한 내용은 Wheeler(2004), Woodall and Adams(1998) 연구 자료에서 확인하십시오 .
- 관리도의 세로 축은 요약 통계량과 동일한 단위로 척도화됩니다 .
- 관리도의 가로 축은 부분군 표본을 식별하고 시간 순서로 지정됩니다 . 시간별 공정을 관측하는 것은 공정의 변화 여부를 평가하는 데 중요합니다 .
- 녹색 선은 중심선 , 즉 데이터의 평균입니다 . 중심선은 공정이 통계적 관리 상태일 때 요약 통계량의 평균 값 (기대값) 을 나타냅니다 . 측정값은 중심선의 양쪽에 균등하게 나타나야 합니다 . 그렇지 않으면 공정 평균이 변하고 있다는 증거가 될 수 있습니다 .
- 빨간색 선 두 개는 각각 UCL 및 LCL 라벨이 지정된 관리 상한과 관리 하한입니다 . 관리 한계는 공정이 통계적 관리 상태일 때 요약 통계량에 예상되는 변동 범위를 제공합니다 . 공정에서 일상적 변동만 나타나는 경우 모든 점이 해당 범위 내에 무작위로 포함되어야 합니다 . 그림 12.7 에서 하나의 측정값이 관리 상한 위에 있습니다 . 이것은 해당 측정값이 특수 원인의 영향을 받았거나 결함 가능성이 있다는 증거입니다 .
- 관리 한계 또는 CUSUM 차트의 V- 마스크를 벗어난 점은 특수 원인에 의한 변동이 있음을 나타냅니다 .

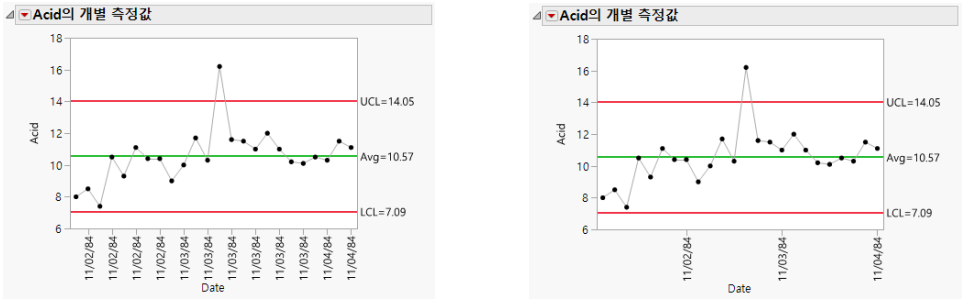
각 플랫폼 내의 옵션은 표본을 받고 기록하거나 데이터 테이블에 추가할 때 동적으로 업데이트할 수 있는 관리도를 생성합니다 .

관리도에 비정상적인 변동이 나타날 때 공정 성능이 저하된 경우에는 공정을 통계적 관리 상태로 되돌리는 조치를 취해야 합니다 . 비정상적인 변동이 공정 개선을 나타낼 경우에는 변동 원인을 연구하고 구현해야 합니다 .

가로 또는 세로 축을 두 번 클릭하면 형식 , 축 값 , 눈금 수 , 격자선 , 참조선 및 축에 표시할 기타 옵션을 지정할 수 있는 적절한 축 규격 창이 나타납니다 .

예를 들어 Pickles.jmp 에는 3 일 동안 매일 측정한 데이터가 나열되어 있습니다 . 그림 12.8 에서는 기본적으로 가로 축에 눈금 두 개마다 하나씩 라벨이 지정되어 있습니다 . 이 경우 그림 12.8 의 왼쪽에서 볼 수 있듯이 라벨이 중복될 때도 있습니다 . 8 개 단위로 증분하여 라벨을 지정하면 오른쪽 차트와 같이 가로 축 라벨이 날짜별로 하나씩 지정됩니다 .

그림 12.8 라벨이 지정된 x 축 눈금 표식의 예



팁 : 경고와 규칙에 대한 자세한 내용은 " 검정 " 및 "Westgard 규칙 " 에서 확인하십시오 .

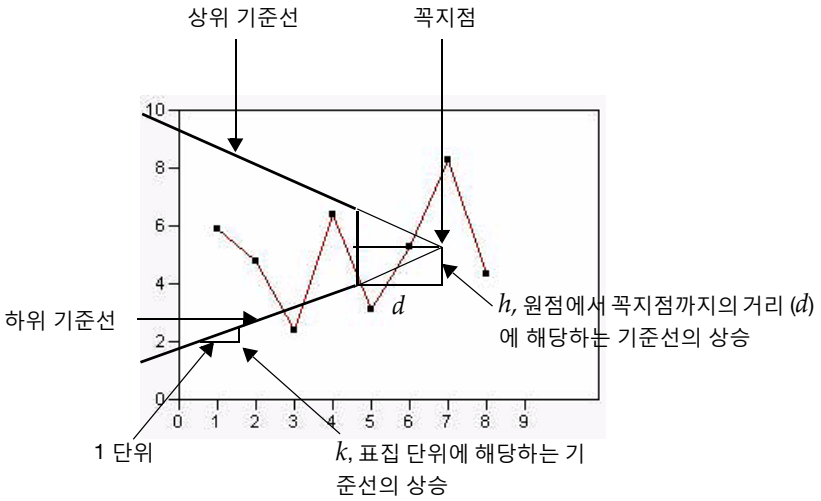
V- 마스크 CUSUM 차트 보고서

양측 V- 마스크 CUSUM 차트 해석

참고 : 자세한 내용은 "V- 마스크 CUSUM 차트 예제 " 에서 확인하십시오 .

양측 CUSUM 차트를 해석하려면 V-마스크를 구성하는 한계와 점을 비교합니다. V-마스크는 누적합 그래프에 양측 면이 겹쳐져 있는 V 형태의 모양입니다. V-마스크는 V자 모양의 한계를 표시하여 구성됩니다. V-마스크의 원점은 가장 최근에 표시된 점이며, 그림 12.9에서와 같이 기준선이 가로 축에서 뒤로 길어집니다. 데이터가 수집되면 누적합 수열이 업데이트되고 가장 최근 점으로 원점이 재배치됩니다.

그림 12.9 양측 CUSUM 차트의 V- 마스크



공정 평균이 변화하면 표시된 점의 기울기가 바뀌므로 CUSUM 차트에서 공정 평균 변화를 시각적으로 쉽게 감지할 수 있습니다. 기울기가 변하는 지점이 변화가 발생하는 지점입니다. 이전에 표시된 점 중 하나 이상이 V-마스크의 상위 또는 하위 기준선과 교차하면 관리이탈 상태입니다. 하위 기준선과 교차하는 점은 공정 평균이 증가한다는 신호이고 상위 기준선과 교차하는 점은 하향 변화를 나타내는 신호입니다.

CUSUM 차트와 Shewhart 차트 사이에는 다음과 같은 중요한 차이가 있습니다 .

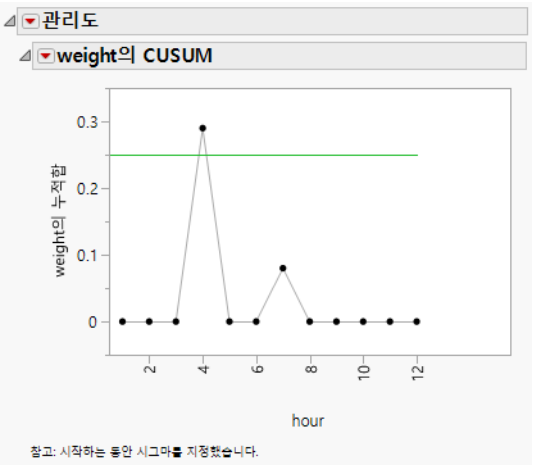
- Shewhart 관리도는 단일 부분군 표본의 정보를 기반으로 점을 표시합니다 . CUSUM 차트의 각 점은 현재 부분군을 포함하여 수집한 모든 표본의 정보를 기반으로 합니다 .
- Shewhart 관리도의 가로 관리 한계는 점이 관리이탈 상태를 나타내는지 여부를 정의합니다 . CUSUM 차트의 한계는 V- 마스크 형태 또는 가로 결정 구간일 수 있습니다 .
- 일반적으로 Shewhart 관리도의 관리 한계는 3σ 한계로 지정됩니다 . CUSUM 차트의 한계는 평균 런 길이를 기준으로 결정됩니다 .

CUSUM 차트는 공정 평균의 작은 변화를 감지하는 데 더 효율적입니다. Lucas(1976)에서는 V-마스크가 Shewhart 관리도보다 약 4 배 빠르게 1σ 변화를 감지한다고 설명합니다.

단측 CUSUM 차트 해석

단측 CUSUM 차트를 사용하여 관심 측면에 근접하거나 초과하는 데이터를 식별할 수 있습니다.

그림 12.10 단측 CUSUM 차트의 예



결정 구간 또는 가로선은 시작 창에 입력한 H 값으로 설정됩니다. 이 예에서는 0.25입니다. 결정 구간 0.25를 초과하는 값은 변화 또는 관리이탈 상태를 나타냅니다. 이 예에서 관측값 4는 변화 발생 지점으로 보입니다. 또한 단측 CUSUM 차트에는 V-마스크가 나타나지 않습니다.

레거시 관리도 플랫폼 옵션

레거시 관리도에는 플랫폼의 여러 부분에 영향을 주는 빨간색 삼각형 메뉴가 있습니다.

- 맨 위 제목 표시줄의 메뉴는 전체 플랫폼 창에 영향을 줍니다. 포함된 항목은 선택한 차트 유형에 따라 다릅니다. 자세한 내용은 " 레거시 관리도에 대한 창 옵션 " 에서 확인하십시오.
- 차트 유형 제목 표시줄에는 각 차트에 개별적으로 영향을 주는 옵션이 포함된 항목 메뉴가 있습니다. 자세한 내용은 " 레거시 관리도에 대한 차트 옵션 " 에서 확인하십시오.

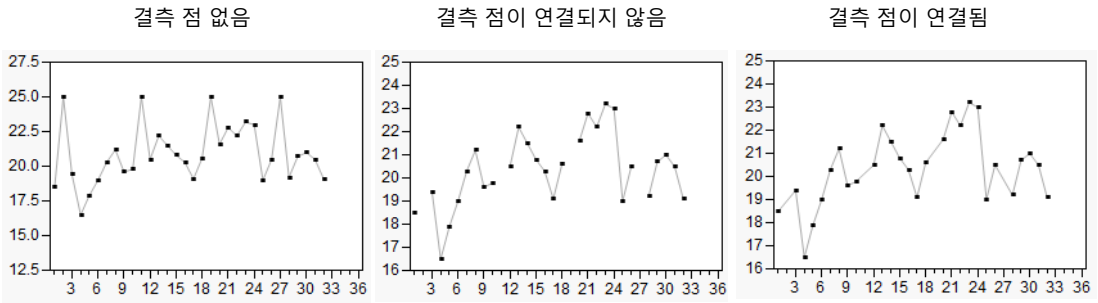
레거시 관리도에 대한 창 옵션

창 제목 표시줄의 빨간색 삼각형 메뉴에는 보고서 창에 영향을 주는 옵션이 나열됩니다. **Xbar** 와 **R** 을 동시에 요청하면 각 차트 유형을 확인하여 표시하거나 숨길 수 있습니다. 사용할 수 있는 특정 옵션은 요청한 관리도 유형에 따라 다릅니다. 사용할 수 없는 옵션은 회색 메뉴 항목으로 표시됩니다.

한계 범례 표시 차트 오른쪽에 평균, UCL 및 LCL 값을 표시하거나 숨깁니다.

결측값을 통과하여 연결 일부 표본에 결측값이 있는 경우 점을 연결합니다. 그림 12.11 의 왼쪽 차트에는 결측 점이 없습니다. 중간 차트에는 표본 2, 11, 19, 27 이 결측이며 해당 점이 연결되어 있지 않습니다. **결측값을 통과하여 연결** 옵션 (기본값) 을 선택하면 오른쪽 차트가 나타납니다.

그림 12.11 결측값을 통과하여 연결 옵션의 예



중앙값 사용 런 차트의 경우 개별 런 차트의 빨간색 삼각형 메뉴에서 **중심선 표시** 옵션을 선택하면 열의 중심 값을 통과하는 선이 그려집니다. 중심선은 메인 런 차트의 빨간색 삼각형 메뉴에 있는 **중앙값 사용** 설정에 의해 결정됩니다. **중앙값 사용**을 선택하면 중앙값이 중심선으로 사용됩니다. 그렇지 않으면 평균이 사용됩니다. 한계를 파일에 저장할 때 전체 평균과 중앙값이 모두 저장됩니다.

공정 능력 (단계 변수가 지정된 경우에는 사용 불가능) 데이터에 대한 공정 능력 분석을 수행합니다. 먼저 공정 변수에 대한 규격 하한, 목표값 및 규격 상한 값을 입력할 수 있는 팝업 창이 표시됩니다.

그림 12.12 공정 능력 분석 창

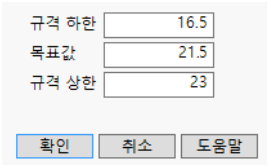
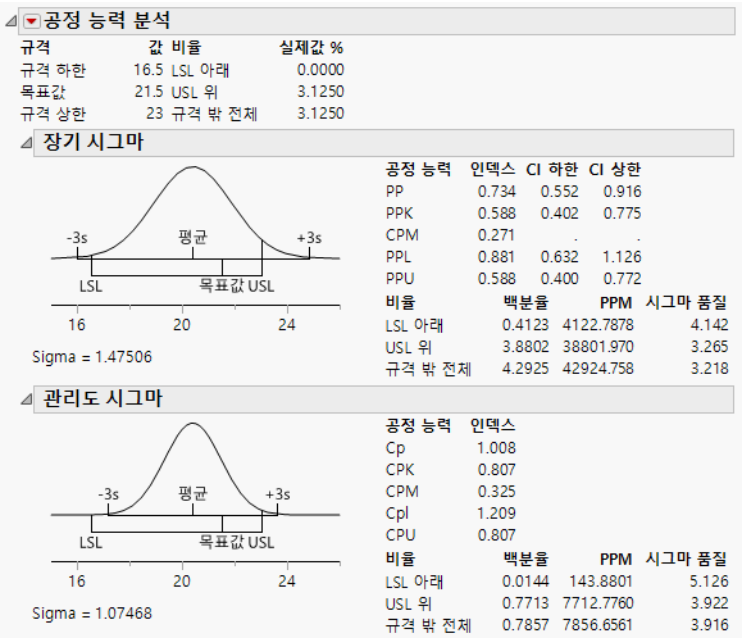


그림 12.13 에서는 규격 하한이 16.5, 목표값이 21.5, 규격 상한이 23 으로 각각 설정된 Coating.jmp 에 대한 공정 능력 분석 보고서의 예를 보여 줍니다.

그림 12.13 Coating.jmp 에 대한 공정 능력 분석 보고서



자세한 내용은 " 공정 능력 분석에 대한 통계 상세 정보 " 에서 확인하십시오 .

시그마 저장 계산된 시그마 값을 JMP 데이터 테이블의 공정 변수 열에 열 특성으로 저장합니다 .

한계 저장 다음 중 한 가지 방법으로 관리 한계를 저장합니다 .

열에 관리 한계를 반응 변수의 기존 데이터 테이블에 열 특성으로 저장합니다 . 한계가 일정하면 보고서의 각 차트 유형에 대한 LCL, 평균 및 UCL 값이 저장됩니다 . 단계 차트에는 이 옵션을 사용할 수 없습니다 . 또한 각 차트에 대해 표본 크기가 일정하지 않으면 이 옵션을 사용해도 아무 효과가 없습니다 .

새 테이블에 각 차트의 표준편차 및 평균을 새 데이터 테이블에 저장합니다 . 한계가 일정하면 각 차트의 LCL, 평균 및 UCL 도 저장됩니다 . 단계가 있는 경우 각 단계에 대한 새 값 집합이 저장됩니다 . 나중에 한계를 사용할 때 이 데이터 테이블을 사용할 수 있습니다 . 관리도 시작 창에서 **한계 가져오기**를 클릭한 후 저장된 데이터 테이블을 선택합니다 . 자세한 내용은 " 한계 저장 및 검색 " 섹션에서 확인하십시오 .

요약 저장 표본 라벨 , 표본 크기 , 표시된 통계량 , 중심선 및 관리 한계를 포함하는 새 데이터 테이블을 생성합니다 . 테이블에 포함되는 특정 통계량은 차트 유형에 따라 다릅니다 .

경보 스크립트 데이터가 특수 원인 검정에 실패했음을 나타내는 스크립트를 작성하고 실행할 수 있습니다 . 결과를 로그에 쓰거나 음성으로 전달할 수 있습니다 . 자세한 내용은 이 가이드의 " 검정 " 에서 확인하십시오 . 사용자 경보 스크립트 작성에 대한 자세한 내용은 Scripting Guide 의 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

레거시 관리도에 대한 차트 옵션

차트 이름 옆의 아이콘을 클릭하면 차트 옵션의 빨간색 삼각형 메뉴가 나타납니다 . 일부 옵션은 차트를 마우스 오른쪽 버튼으로 클릭하면 표시되는 **차트 옵션**에서도 사용할 수 있습니다 . 아래의 모든 옵션을 모든 관리도 유형에 사용할 수 있는 것은 아닙니다 .

상자 그림 평균 차트에 표시된 부분군 평균에 상자 그림을 겹쳐 놓습니다 . 상자 그림에는 부분군 최대값 , 최소값 , 75 번째 백분위수 , 25 번째 백분위수 및 중앙값이 표시됩니다 . **점 표시** 옵션을 선택 취소하지 않으면 부분군 평균에 대한 표식이 나타납니다 . **상자 그림** 옵션은 $\bar{X}$ 차트에만 사용할 수 있습니다 . 이 옵션은 부분군 표본 크기가 더 클 때 가장 적합합니다 (부분군의 표본 수가 10 개보다 많을 때).

바늘 표시된 점을 세로 선분으로 중심선에 연결합니다 .

점 연결 데이터 점을 연결하는 선을 표시하거나 숨깁니다 .

점 표시 요약 통계량을 나타내는 점을 표시하거나 숨깁니다 . 처음에는 점이 표시됩니다 . **상자 그림** 옵션이 활성화된 경우 이 옵션을 사용하여 부분군 평균을 나타내는 표식을 표시하지 않을 수 있습니다 .

연결 색상 점을 연결하는 데 사용되는 선분의 색상을 선택할 수 있도록 JMP 색상 팔레트를 표시합니다 .

중심선 색상 중심선을 그리는 데 사용되는 선분의 색상을 선택할 수 있도록 JMP 색상 팔레트를 표시합니다 .

한계 색상 상한 및 하한 선에 사용되는 선분의 색상을 선택할 수 있도록 JMP 색상 팔레트를 표시합니다 .

선 너비 관리도 선의 너비를 선택할 수 있습니다 . 옵션은 **줄임** , **중간** 또는 **굵게**입니다 .

점 표식 차트에 사용되는 표식을 선택할 수 있습니다 .

중심선 표시 관리도에 중심선을 표시하거나 숨깁니다 .

관리 한계 표시 차트 관리 한계 및 범례를 표시하거나 숨깁니다 .

한계 정밀도 관리 한계 라벨에 대한 소수점 자릿수를 설정합니다 .

검정 검정이 양성일 때 차트에 표시할 검정을 선택할 수 있는 하위 메뉴를 표시합니다 . 검정은 한계가 3σ 인 차트에만 적용됩니다 . 검정 1 ~ 4 는 평균 , 개별값 및 계수형 차트에 적용됩니다 . 검정 5 ~ 8 은 평균 차트 , 사전 집계 및 개별 측정값 차트에만 적용됩니다 . 차트에 검정이

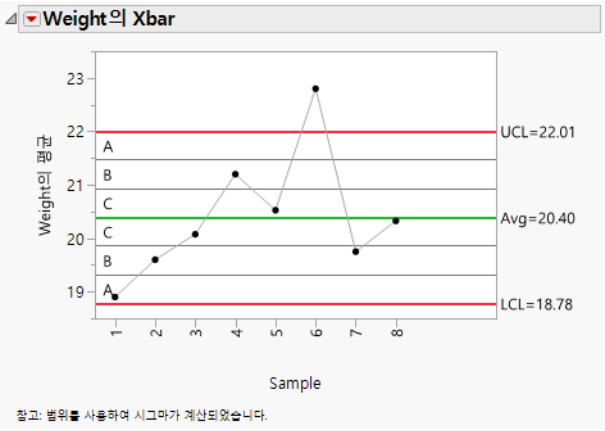
적용되지 않으면 " 검정 " 옵션이 흐리게 표시됩니다 . 표본 크기가 다르면 " 검정 " 옵션이 회색으로 표시됩니다 . 차트가 열린 상태에서 표본이 변경되어 크기가 같아진 경우 영역 또는 검정 옵션이 선택되면 해당 영역 또는 검정이 즉시 적용된 후 차트에 나타납니다 . 이러한 특수 검정을 Western Electric 규칙이라고도 합니다 . 특수 원인 검정에 대한 자세한 내용은 " 검정 " 에서 확인하십시오 .

Westgard 규칙 Westgard 규칙은 공정이 관리 상태인지 아니면 관리 이탈인지 결정하는 데 도움이 되는 관리 규칙입니다 . 특정 검정에 대한 결정 규칙을 사용하여 약어 형태로 표현된 여러 검정이 있습니다 . 자세한 내용은 "Westgard 규칙 " 의 텍스트 및 차트를 참조하십시오 .

한계를 벗어난 경우 검정 한계를 벗어난 점에 "*" 로 플래그를 지정합니다 . 이 검정은 한계가 있는 모든 차트에서 작동합니다 . 이때 표본 크기가 일정한지 여부 및 k 의 크기나 한계 너비는 상관없습니다 . 예를 들어 표본 크기가 같지 않은 경우 R 차트의 한계를 벗어나는 점에 플래그를 지정하려면 이 명령을 사용할 수 있습니다 .

영역 표시 영역 선을 표시하거나 숨깁니다 . Coating.jmp 샘플 데이터의 weight 평균 그림에 표시된 것처럼 영역에 A, B, C 라벨이 지정됩니다 . 관리도 검정에서는 영역 선을 경계로 사용합니다 . 7 개의 영역 선이 중심선에서부터 1 시그마 간격으로 떨어져 있습니다 .

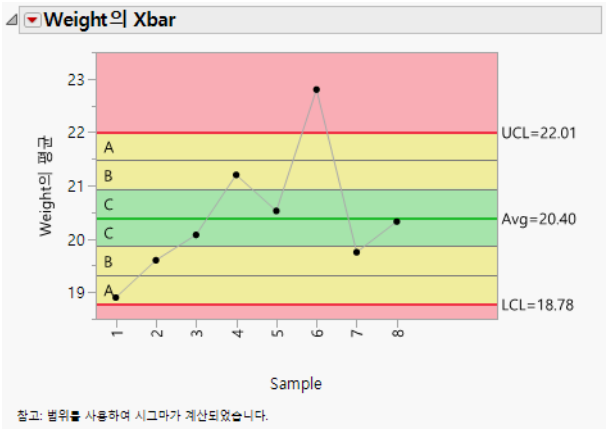
그림 12.14 영역 표시



영역에 음영 적용 세 개의 영역과 이 영역의 외부에 대해 기본 녹색 , 노란색 및 빨간색을 표시하거나 숨깁니다 . 녹색은 중심선에서 1 시그마 떨어진 영역 , 노란색은 중심선에서 2 시그마 및 3 시그마 떨어진 영역 , 빨간색은 3 시그마를 벗어난 영역을 각각 나타냅니다 . 영역 선을 포함하거나 포함하지 않고 음영을 표시할 수 있습니다 .

팁 : 영역에 음영을 적용하는 데 사용되는 색상을 변경하려면 관리도를 마우스 오른쪽 버튼으로 클릭하고 " 사용자 정의 " 를 선택합니다 . " 그래프 사용자 정의 " 창에서 세 영역에 대한 색상을 각각 지정할 수 있습니다 .

그림 12.15 영역에 음영 적용



OC 곡선 (부분군 크기가 같은 경우 사용 가능) 활성 관리도에서 직접 계산된 모든 값을 사용하여 OC(검사 특성) 곡선이 포함된 새 창을 엽니다 . 자세한 내용은 " 검사 특성 곡선 유틸리티 " 에서 확인하십시오 .

V- 마스크 CUSUM 관리도에 대한 차트 옵션

마스크 색상 ("V 마스크 표시 " 옵션을 선택한 경우에만 사용 가능) V- 마스크의 선 색상을 선택할 수 있습니다 .

변화 표시 시작 창에 입력한 변화를 표시하거나 숨깁니다 .

V 마스크 표시 CUSUM 관리도 시작 창에 지정한 통계량을 기반으로 V- 마스크를 표시하거나 숨깁니다 .

모수 표시 CUSUM 차트 생성 모수를 요약하는 보고서를 표시하거나 숨깁니다 .

ARL 표시 ARL(평균 런 길이) 정보를 표시하거나 숨깁니다 . 평균 런 길이는 관리이탈 상태 신호가 발생하기 전에 수집한 기대 표본 수입니다 .

- ARL(델타) 또는 ARL1은 지정된 델타 크기의 변화를 감지하기 위한 평균 런 길이입니다.
- ARL(0) 또는 ARL0 은 지정된 모수에 대한 관리상태 평균 런 길이입니다 (Montgomery 2013 참조).

한계 저장 및 검색

JMP 에서는 관리도에 대해 이전에 설정한 관리 한계를 사용할 수 있습니다 .

- 관리 상한 , 관리 하한 및 중심선 값
- 평균 및 표준편차와 같은 한계를 계산하기 위한 모수

관리 한계 또는 한계 모수 값은 JMP 데이터 테이블 (**한계 테이블**이라고도 함) 에 있거나 , 공정 열에 열 특성으로 저장되어야 합니다 . **관리도** 명령을 지정할 때 관리도 시작 창의 **한계 가져오기** 버튼을 사용하여 한계 테이블을 검색할 수 있습니다 .

팁 : 규격 한계를 여러 열에 한 번에 추가하는 방법은 " 규격 한계 관리 유틸리티 " 에서 확인하십시오 .

한계 데이터 테이블을 생성하는 가장 쉬운 방법은 관리도 플랫폼에서 계산된 결과를 저장하는 것입니다 . 각 관리도에 대한 빨간색 삼각형 메뉴의 **한계 저장** 명령은 표본 값의 한계를 자동으로 저장합니다 . 테이블에 저장되는 데이터 유형은 분석 창의 관리도 유형에 따라 다릅니다 . 원하는 소스의 값을 사용하고 고유한 한계 테이블을 직접 생성할 수도 있습니다 .

모든 한계 테이블에는 다음 항목이 있어야 합니다 .

- 각 행을 식별하는 특수 키워드 열
- 알려진 표준 모수 또는 한계를 값으로 사용하는 각 변수에 대한 열 . 이 열 이름은 관리도 플랫폼에서 분석할 데이터 테이블의 해당 공정 변수 이름과 동일해야 합니다 .

다음 표에서는 레거시 관리도와 관리도 빌더에서 생성한 차트 둘 다에 대해 한계 키워드 및 연관된 관리도를 설명합니다 .

표 12.1 한계 테이블 키 (해당 차트 및 의미 포함)

키워드	차트	의미
_KSigma	관리도 빌더와 V-마스크 CUSUM 을 제외한 모든 차트	관리 한계를 계산하기 위한 통계량의 표준편차의 배수 . 유의수준을 기준으로 하는 한계의 경우 결측으로 설정됩니다 .
_Alpha	관리도 빌더를 제외한 모든 차트	관리 한계를 계산하는 데 사용되는 제 1 종 오류 확률
_Range Span	IM, MR, MMR	이동 범위가 계산되는 연속 측정값 수 . 범위 길이가 항상 2 인 관리도 빌더 플랫폼에는 적용되지 않습니다 .

표 12.1 한계 테이블 키 (해당 차트 및 의미 포함)(계속)

키워드	차트	의미
_Sample Size	Levey-Jennings 및 사전 집계를 제외한 모든 차트	부분군 크기
_Std Dev	Xbar, R, S, IM, MR, G, T, V- 마스크 CUSUM, Levey-Jennings	알려진 공정 표준편차
_U	C, U	단위별 알려진 평균 결함 수
_P	NP, P	평균 부적합 비율의 알려진 값
_LCL, _UCL	Xbar, IM, P, NP, C, U, G, T, Levey-Jennings	평균 차트, 개별 측정값 차트, 계수형 차트 또는 희귀 사건 차트의 관리 하한 및 상한
_AvgR	R, MR	평균 범위 또는 평균 이동 범위
_LCLR, _UCLR	R, MR	R 또는 MR 차트의 관리 하한 R 또는 MR 차트의 관리 상한
_AvgS, _LCLS, _UCLS	S 차트	S 차트의 평균 표준편차, 관리 상한 및 하한
_AvgR_PreMeans _AvgR_PreStdDev _LCLR_PreMeans _LCLR_PreStdDev _UCLR_PreMeans _UCLR_PreStdDev _Avg_PreMeans _Avg_PreStdDev _LCL_PreMeans _LCL_PreStdDev _UCL_PreMeans _UCL_PreStdDev	IM, MR	사전 집계된 그룹 평균 또는 표준편차를 기반으로 하는 평균, 관리 상한 및 하한

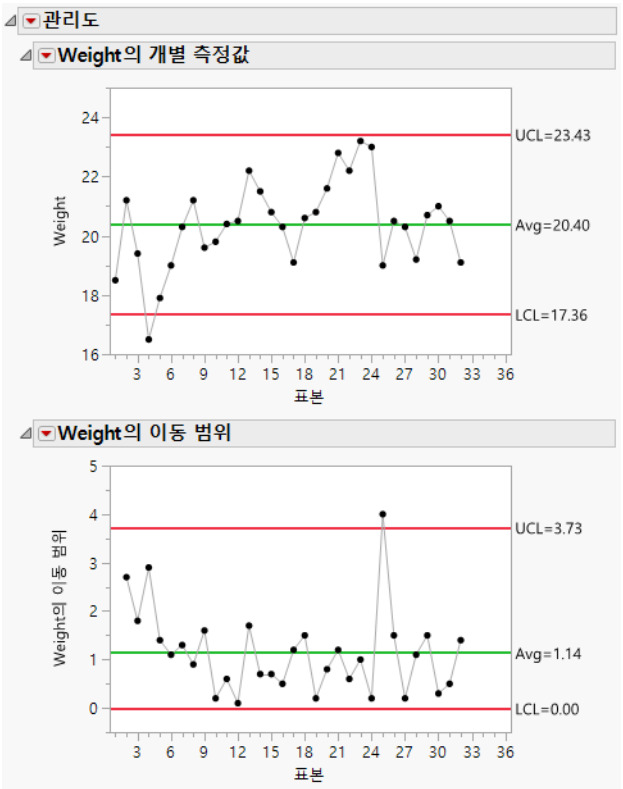
표 12.1 한계 테이블 키 (해당 차트 및 의미 포함)(계속)

키워드	차트	의미
_Data Units	V-마스크 CUSUM	V-마스크 CUSUM 차트의 규격
_Two Sided		
_Headstart		
_Beta		
_Delta		

한계를 새 데이터 테이블에 저장하거나 반응 열의 특성으로 저장할 수 있습니다. **새 테이블에** 명령을 사용하여 관리 한계를 저장하는 경우 테이블에 작성되는 한계 키워드는 현재 표시된 차트 유형에 따라 결정됩니다.

그림 12.16에서는 Coating.jmp를 사용하여 데이터 테이블에 저장된 관리 한계의 예를 보여 줍니다. _Mean, _LCL 및 _UCL 값이 있는 행은 개별 측정값 차트가 대상입니다. R 접미사가 있는 값 (_AvgR, _LCLR, _UCLR)은 이동 범위 차트가 대상입니다. 이 한계 테이블을 사용하여 이러한 차트를 다시 생성하는 경우 관리도 플랫폼은 _LimitsKey 열의 키워드에서 적절한 한계를 식별합니다.

그림 12.16 데이터 테이블에 한계를 저장하는 예



	_LimitsKey	Weight
1	_KSigma	3
2	_Alpha	0.0026997961
3	_Range Span	2
4	_Std Dev	1.0120139729
5	_Mean	20.396875
6	_LCL	17.360833081
7	_UCL	23.432916919
8	_AvgR	1.1419354839
9	_LCLR	0
10	_UCLR	3.7301687077

_KSigma, _Alpha 및 _Range Span 값은 관리도 시작 창에서 지정할 수 있습니다. JMP에서는 항상 이 창의 값을 먼저 검토합니다. 창에 지정된 값이 활성 한계 테이블의 값보다 우선합니다.

알 수 없는 키워드가 있는 행과 '제외됨' 행 상태가 표시된 행은 무시됩니다. _Range Span, _KSigma, _Alpha 및 _Sample Size를 제외하고 필요한 값이 지정되지 않으면 데이터에서 추정됩니다.

제외 , 삭제 및 숨겨진 표본

다음 표에는 표본 및 부분군의 다양한 조건이 미치는 영향에 대해 요약되어 있습니다.

표 12.2 제외 , 삭제 및 숨겨진 표본

차트를 생성하기 전에 표본의 모든 행이 제외됩니다.	표본이 한계 계산에 포함되지 않지만 그래프에는 나타납니다.

표 12.2 제외 , 삭제 및 숨겨진 표본 (계속)

차트를 생성한 후 표본이 제외됩니다.	표본이 한계 계산에 포함되며 그래프에도 나타납니다. 그래프가 열린 상태에서 표본을 제외해도 출력에 아무 변화가 없습니다.
차트를 생성하기 전에 표본을 숨겼습니다.	표본이 한계 계산에 포함되지만 그래프에는 나타나지 않습니다.
차트를 생성한 후 표본을 숨겼습니다.	표본이 한계 계산에 포함되지만 그래프에는 나타나지 않습니다. 표본 표식이 그래프에서 사라지고 표본 라벨은 축에 계속 표시되지만 한계는 그대로 유지됩니다.
차트를 생성하기 전에 표본의 모든 행이 제외되고 숨겨집니다.	표본이 한계 계산에 포함되지 않고 그래프에도 나타나지 않습니다.
차트를 생성한 후 표본의 모든 행이 제외되고 숨겨집니다.	표본이 한계 계산에 포함되지만 그래프에는 나타나지 않습니다. 표본 표식이 그래프에서 사라지고 표본 라벨은 축에 계속 표시되지만 한계는 그대로 유지됩니다.
차트를 생성하기 전에 표본이 삭제된 상태에서 데이터 집합이 부분집합으로 구성됩니다.	표본이 한계 계산에 포함되지 않고 축에 표본 값이 포함되지 않으며 표본 표식이 그래프에 나타나지 않습니다.
차트를 생성한 후 표본이 삭제된 상태에서 데이터 집합이 부분집합으로 구성됩니다.	표본이 한계 계산에 포함되지 않고 그래프에도 나타나지 않습니다. 표본 표식이 그래프에서 사라지고 표본 라벨이 축에서 제거되며 그래프가 이동하고 한계가 변경됩니다.

다음 내용도 참조하십시오 .

- 숨기기 및 제외하기는 표본에서 첫 번째 관측값의 행 상태에만 작동합니다 . 예를 들어 표본의 두 번째 관측값이 숨겨져 있지만 첫 번째 관측값은 숨겨지지 않은 경우 표본이 차트에 나타납니다 .

  참고 : 사전 집계 차트의 제외된 행은 표본 내에서의 위치에 관계없이 계산에서 제외됩니다 .

- 제외 / 숨기기 규칙의 예외 : 숨겨진 행과 제외된 행은 특수 원인 검정을 위한 점 개수에 모두 포함됩니다 . 제외된 행은 특수 원인 플래그로 라벨을 지정할 수 있습니다 . 숨겨진 점에는 라벨을 지정할 수 없습니다 . 숨겨진 점에 특수 원인 검정에 대한 플래그가 있는 경우 차트에 표시되지 않습니다 .
- 적용된 특정 규칙 (표 12.2) 때문에 관리도는 자동 재계산 스크립트를 지원하지 않습니다 .

관리도 플랫폼의 추가 예

- " 사전 집계 차트 예제 "
- "V- 마스크 CUSUM 차트 예제 "
- " 단측 CUSUM 차트 예제 "
- "UWMA 차트 예제 "

사전 집계 차트 예제

이 예에서는 Coating.jmp 데이터 테이블을 사용합니다 .

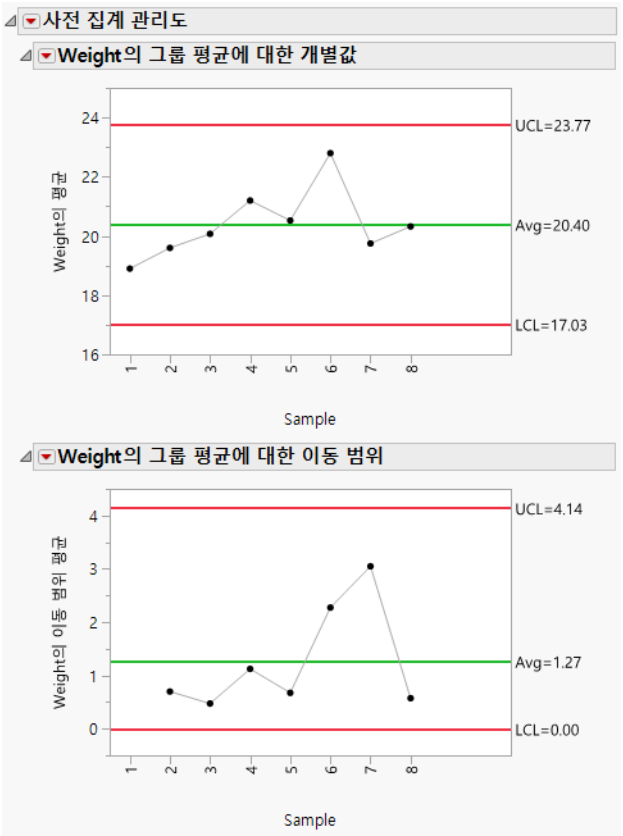
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Coating.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 레거시 관리도 > 사전 집계**를 선택합니다 .
3. Weight 를 선택하고 **공정**을 클릭합니다 .
4. Sample 을 선택하고 **표본 라벨**을 클릭합니다 .
5. **그룹 평균에 대한 개별값**과 **그룹 평균에 대한 이동 범위**를 둘 다 선택합니다 . **표본 라벨별로 그룹화된 표본** 버튼은 표본 라벨 변수를 선택하면 자동으로 선택됩니다 .

   사전 집계 차트를 사용할 때 **그룹 평균에 대한** 옵션 또는 **그룹 표준편차에 대한** 옵션 중 하나를 선택하거나 둘 다 선택할 수 있습니다 . IR 차트 유형 두 개를 모두 선택한 경우 각 옵션은 두 개의 차트 (X 차트라고도 하는 개별 측정값 차트 , 이동 범위 차트) 를 생성합니다 .

   그룹 평균에 대한 옵션은 각 표본 평균을 계산한 후 평균을 표시하고 평균에 대한 개별 측정값 차트 및 이동 범위 차트를 생성합니다 .

   그룹 표준편차에 대한 옵션은 각 표본 표준편차를 계산하고 표준편차를 개별 점으로 표시합니다 . 그런 다음 표준편차에 대한 개별 측정값 차트와 이동 범위 차트가 나타납니다 .
6. **확인**을 클릭합니다 .

그림 12.17 사전 집계 데이터로 차트 생성의 예



Xbar 차트와 S 차트의 점은 그룹 평균에 대한 개별값 차트 및 그룹 표준편차에 대한 개별값 차트와 동일하지만 개별 차트로 계산되므로 한계가 다릅니다.

Coating.jmp 데이터 테이블을 사용하여 사전 집계 차트를 생성하는 또 다른 방법은 다음과 같습니다.

1. **테이블 > 요약**을 선택합니다.
2. Sample 을 선택하고 **그룹**을 클릭합니다.
3. Weight 를 선택한 후 **통계량** > **평균** 및 **통계량** > **표준편차**를 클릭합니다.
4. **확인**을 클릭합니다.
5. **분석 > 품질 및 공정 > 레거시 관리도 > IR** 을 선택합니다.
6. 평균 (Weight) 및 표준편차 (Weight) 를 선택하고 **공정**을 클릭합니다.
7. **확인**을 클릭합니다.

결과 차트는 사전 집계 차트와 일치합니다.

V- 마스크 CUSUM 차트 예제

기계 하나가 8온스 캔에 2사이클 엔진 오일 첨가제를 주입합니다. 주입 공정은 통계적 관리 상태를 유지한다고 간주됩니다. 가득 주입된 캔의 평균 무게(μ_0)가 8.10 온스가 되도록 공정을 설정합니다. 이전 분석에 따르면 주입 무게의 표준편차(σ_0)는 0.05온스입니다.

4 개 캔이 포함된 부분군 표본을 선택하고 12 시간 동안 1 시간에 한 번씩 무게를 측정했습니다 . Oil1 Cusum.jmp 데이터 테이블의 각 관측값은 무게 (weight) 값 하나와 관련 시간 (hour) 값을 포함합니다 . 관측값은 시간 값이 작은 것부터 순서대로 정렬됩니다 .

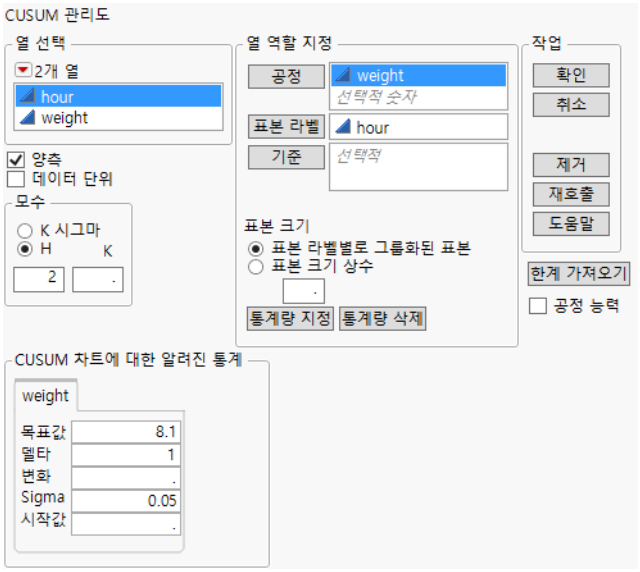
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Oil1 Cusum.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 레거시 관리도 > CUSUM** 을 선택합니다 .
3. weight 를 선택하고 **공정**을 클릭합니다 .
4. hour 를 선택하고 **표본 라벨**을 클릭합니다 .
5. **양측** 체크박스를 선택합니다 (선택되어 있지 않은 경우).
6. " 모수 " 영역에서 **H** 버튼을 클릭하고 2 를 입력합니다 .
7. **통계량 지정**을 클릭합니다 .
8. **목표값** 옆에 8.1 을 입력합니다 .

   8.1 은 가득 주입된 캔의 평균 무게 (온스) 입니다 . 이 값은 목표 평균입니다 .
9. **델타** 옆에 1 을 입력합니다 .

   1 은 공정 표준편차 또는 표준 오차의 배수로 감지할 최소 변화의 절대값입니다 .
10. **시그마** 옆에 0.05 를 입력합니다 .

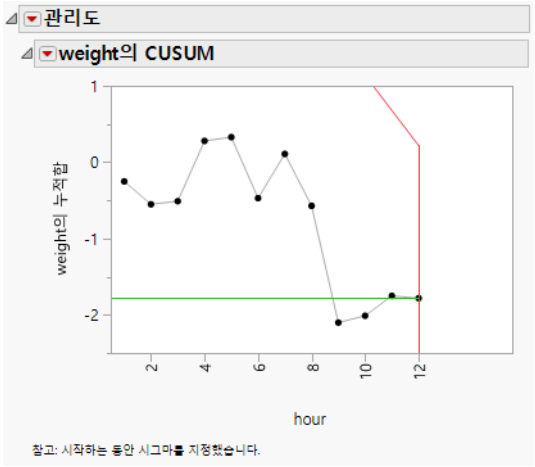
    0.05 는 주입 무게 (온스) 의 알려진 표준편차 (σ_0) 입니다 .

그림 12.18 완료된 시작 창



11. **확인**을 클릭합니다.

그림 12.19 Oil1 Cusum.jmp 데이터에 대한 양측 CUSUM 차트



점을 V-마스크와 비교하여 차트를 해석할 수 있습니다. V-마스크의 오른쪽 모서리는 가장 최근 점(12번째 시간)의 중심에 위치하게 됩니다. V-마스크의 기준선을 벗어나는 점이 없으므로 공정에서 변화가 발생했다는 증거가 없습니다. 자세한 내용은 "V-마스크 CUSUM 차트 보고서"에서 확인하십시오.

단측 CUSUM 차트 예제

기계 하나가 8 온스 캔에 엔진 오일을 주입하는 "V- 마스크 CUSUM 차트 예제 " 에 사용된 데이터를 예로 들어보겠습니다 . 현재 제조업체는 비용 절감을 위해 과다 주입에 대해 우려하고 있습니다 (주입 부족에 대해서는 크게 염려하지 않음). 단측 CUSUM 차트를 사용하여 과다 주입 사례를 식별합니다 . 평균 8.1 을 0.25 온스 초과하는 경우는 모두 문제로 간주됩니다 .

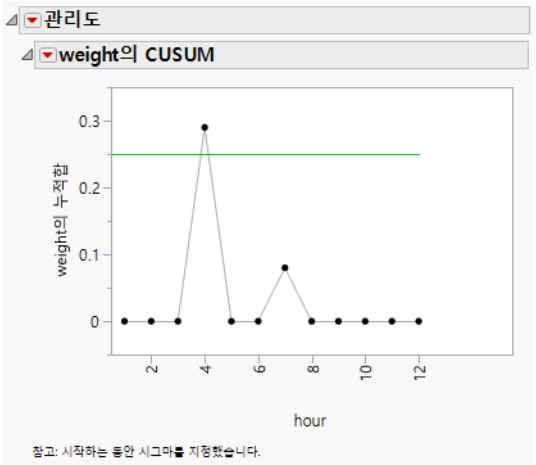
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Oil1 Cusum.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 레거시 관리도 > CUSUM** 을 선택합니다 .
3. **양측**을 선택 취소합니다 .
4. weight 를 선택하고 **공정**을 클릭합니다 .
5. hour 를 선택하고 **표본 라벨**을 클릭합니다 .
6. **H** 를 클릭하고 0.25 를 입력합니다 .
7. **통계량 지정**을 클릭합니다 .
8. **목표값** 옆에 8.1 을 입력합니다 .

   8.1 은 가득 주입된 캔의 평균 무게 (온스) 입니다 . 이 값은 목표 평균입니다 .
9. **델타** 옆에 1 을 입력합니다 .

   1 은 공정 표준편차 또는 표준 오차의 배수로 감지할 최소 변화의 절대값입니다 .
10. **시그마** 옆에 0.05 를 입력합니다 .

    0.05 는 주입 무게 (온스) 의 알려진 표준편차 (σ_0) 입니다 .
11. **확인**을 클릭합니다 .

그림 12.20 Oil1 Cusum.jmp 데이터에 대한 단측 CUSUM 차트



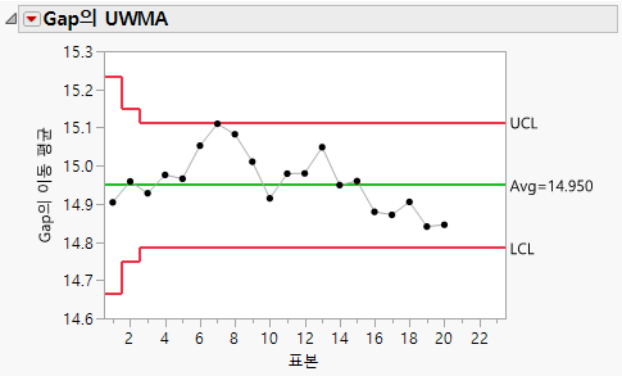
결정 구간은 입력한 H 값(0.25)으로 설정됩니다. 네 번째 시간에서 상당한 과다 주입이 발생한 것을 볼 수 있습니다.

UWMA 차트 예제

Clips1.jmp 샘플 데이터 테이블에서 관심 측도는 제조된 금속 클립의 끝부분 사이 간격입니다 . 공정에서 평균 간격의 변화를 모니터링하기 위해 클립 5 개로 구성된 부분군 표본을 매일 선택했습니다 . 이동 평균 거리가 3 인 UWMA 차트를 검토합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Clips1.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 레거시 관리도 > UWMA** 를 선택합니다 .
3. Gap 을 선택하고 **공정**을 클릭합니다 .
4. Sample 을 선택하고 **표본 라벨**을 클릭합니다 .
5. **이동 평균 거리**를 3 으로 변경합니다 .
6. **확인**을 클릭합니다 .

그림 12.21 Clips1 데이터에 대한 UWMA 차트



첫째 날의 점은 해당 날짜에 대한 5 개 부분군 표본 값의 평균입니다 . 둘째 날에 표시된 점은 첫째 날과 둘째 날에 대한 부분군 표본 평균의 평균입니다 . 나머지 날의 점은 각 날짜와 두 개의 이전 날짜에 대한 하위 표본 평균의 평균입니다 .

평균 클립 간격이 감소하는 것으로 보이지만 3σ 한계를 벗어나는 표본 점은 없습니다 .

관리도 플랫폼에 대한 통계 상세 정보

- " 중앙값 이동 범위 차트에 대한 관리 한계 "
- " 공정 능력 분석에 대한 통계 상세 정보 "
- "V- 마스크 CUSUM 관리도에 대한 통계 상세 정보 "
- " 가중 이동 평균 차트에 대한 통계 상세 정보 "

참고 : 다른 차트 유형 (예 : Xbar-R 차트 , P 차트 , NP 차트 등) 에 대한 자세한 내용은 " 관리도 빌더에 대한 통계 상세 정보 " 에서 확인하십시오 .

중앙값 이동 범위 차트에 대한 관리 한계

중앙값 이동 범위 차트에 대한 관리 한계는 다음과 같이 계산됩니다 .

$\text{LCL}_{\text{MMR}} = \max(0, \text{MMR} - kd_3(n)\hat{\sigma})$

$\text{UCL}_{\text{MMR}} = \text{MMR} + kd_3(n)\hat{\sigma}$

다음은 각 요소에 대한 설명입니다 .

MMR = 비결측 이동 범위의 중앙값

$\hat{\sigma}$ = MMR/0.954

$d_3(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 표준편차

공정 능력 분석에 대한 통계 상세 정보

이 섹션에는 공정 능력 분석 보고서의 통계량 계산에 대한 자세한 정보가 포함되어 있습니다.

변동 통계량

모든 공정 능력 분석에서 사용하는 계산식은 동일합니다. 옵션은 다음과 같이 시그마(σ) 계산 방식에 차이가 있습니다 .

장기 시그마 전체 표준편차를 사용합니다 . 이 옵션은 P_{pk} 통계량에 사용되며 다음과 같이 시그마를 계산합니다 .

$$\hat{\sigma} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

참고 : 기본적으로 장기 시그마 보고서의 공정 능력 지수는 다른 시그마 보고서에 사용되는 Cp 라벨을 사용합니다 . 장기 시그마 보고서에 Ppk 라벨을 사용하려면 **파일 > 환경 설정 > 플랫폼 > 분포 > PpK 공정 능력 라벨** 환경 설정을 선택하십시오 .

관리도 시그마 관리도 설정에 따라 결정된 시그마를 사용합니다 .

- 관리도 시작 창의 " 통계량 지정 " 버튼을 사용하여 " 시그마 " 값을 지정하면 지정된 값이 공정 능력 지수 계산에 사용됩니다 .
- "이동 범위 (평균)" 옵션을 사용하는 IR 차트에서는 시그마 값이 다음과 같이 계산됩니다.

$$\hat{\sigma} = \frac{\bar{R}}{d_2(n)}$$

다음은 각 요소에 대한 설명입니다 .

$\bar{R}$ = 이동 범위의 평균

$d_2(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 기대값 . 여기서 n 은 " 범위 길이 " 옵션의 값입니다 .

- "중앙값 이동 범위" 옵션을 사용하는 IR 차트에서는 시그마 값이 다음과 같이 계산됩니다.

$$\hat{\sigma} = \frac{MMR}{d_4(n)}$$

다음은 각 요소에 대한 설명입니다 .

MMR = 비결측 이동 범위의 중앙값

$d_4(n)$ = 단위 표준편차를 갖는 n 개 독립 정규 분포 변수의 범위 중앙값 . 여기서 n 은 " 범위 길이 " 옵션의 값입니다 .

- "R" 옵션을 사용하는 Xbar 차트에서는 시그마 값이 다음과 같이 계산됩니다 .

$$\hat{\sigma} = \frac{\frac{R_1}{d_2(n_1)} + \ldots + \frac{R_N}{d_2(n_N)}}{N}$$

다음은 각 요소에 대한 설명입니다 .

$R_i = i$ 번째 부분군의 범위

$n_i = i$ 번째 부분군의 표본 크기

$d_2(n_i)$ = 단위 표준편차를 갖는 n_i 개 독립 정규 분포 변수의 범위 기대값

N = 부분군 수 ($n_i \geq 2$)

- "S" 옵션을 사용하는 Xbar 차트에서는 시그마 값이 다음과 같이 계산됩니다 .

$$\hat{\sigma} = \frac{\frac{s_1}{c_4(n_1)} + \ldots + \frac{s_N}{c_4(n_N)}}{N}$$

다음은 각 요소에 대한 설명입니다 .

n_i = i 번째 부분군의 표본 크기

$c_4(n_i)$ = 단위 표준편차를 갖는 n_i 개 독립 정규 분포 변수의 표준편차 기대값

N = 부분군 수 ($n_i \geq 2$)

s_i = i 번째 부분군의 표본 표준편차

정규 분포에 대한 공정 능력 지수

이 섹션에서는 정규 데이터에 대한 공정 능력 지수의 계산에 대해 자세히 설명합니다 .

평균이 μ 이고 표준편차가 σ 인 공정 특성에 대한 모집단 기반 공정 능력 지수는 다음과 같이 정의됩니다 .

$$\text{Cp} = \frac{USL - LSL}{6\sigma}$$

$$\text{Cpl} = \frac{\mu - LSL}{3\sigma}$$

$$\text{Cpu} = \frac{USL - \mu}{3\sigma}$$

$$\text{Cpk} = \min(\text{Cpl}, \text{Cpu})$$

$$\text{Cpm} = \frac{\min(T - LSL, USL - T)}{3\sigma\sqrt{1 + \left(\frac{T - \mu}{\sigma}\right)^2}}$$

다음은 각 요소에 대한 설명입니다 .

LSL = 규격 하한

USL = 규격 상한

T = 목표값

표본 기반 공정 능력 지수의 경우 모수는 추정값으로 대체됩니다 . σ 의 추정값은 공정 능력 분석 창에서 지정한 방법을 사용합니다 . 자세한 내용은 "변동 통계량"에서 확인하십시오 .

규격 한계가 하나뿐인 경우 결측 규격 한계를 포함하는 공정 능력 지수가 결측값으로 보고됩니다 .

팁 : 공정 능력 지수 1.33 은 허용 가능한 최소값으로 간주되는 경우가 많습니다 . 정규 분포의 경우 공정 능력 지수 1.33 은 100,000 개당 약 6 개의 기대 부적합 단위 수에 해당합니다 .

공정 능력 지수에 대한 신뢰 구간

참고 : 공정 능력 지수에 대한 신뢰 구간은 장기 시그마 보고서에만 나타납니다 .

Cp 에 대한 100(1 - α)% 신뢰 구간은 다음과 같이 계산됩니다 .

$$\left(\hat{\mathrm{CP}} \sqrt{\frac{\chi^2_{\alpha/2,\,n-1}}{n-1}}, \hat{\mathrm{CP}} \sqrt{\frac{\chi^2_{1-\alpha/2,\,n-1}}{n-1}} \right)$$

다음은 각 요소에 대한 설명입니다 .

$\hat{\mathrm{CP}}$ = Cp 에 대한 추정값

$\chi^2_{\alpha/2,\,n-1}$ = 자유도가 n - 1 인 카이제곱 분포의 (α/2) 번째 분위수

n = 관측값 수

Cpk 에 대한 100(1 - α)% 신뢰 구간은 다음과 같이 계산됩니다 .

$$\left(\hat{C}_{pk}\left[1 - \Phi^{-1}{}_{1-\alpha/2} \sqrt{\frac{1}{9n\hat{C}^2_{pk}} + \frac{1}{2(n-1)}} \right], \hat{C}_{pk}\left[1 + \Phi^{-1}{}_{1-\alpha/2} \sqrt{\frac{1}{9n\hat{C}^2_{pk}} + \frac{1}{2(n-1)}} \right] \right)$$

다음은 각 요소에 대한 설명입니다 .

$\hat{C}_{pk}$ = Cpk 에 대한 추정값

$\Phi^{-1}{}_{1-\alpha/2}$ = 표준 정규 분포의 (1 - α/2) 번째 분위수

n = 관측값 수

CPM 에 대한 100(1 - α)% 신뢰 구간은 다음과 같이 계산됩니다 .

$$\left(\hat{\mathrm{CPM}} \sqrt{\frac{\chi^2_{\alpha/2,\,\gamma}}{\gamma}}, \hat{\mathrm{CPM}} \sqrt{\frac{\chi^2_{1-\alpha/2,\,\gamma}}{\gamma}} \right)$$

다음은 각 요소에 대한 설명입니다 .

$\hat{\mathrm{CPM}}$ = CPM 에 대한 추정값

$\chi^2_{\alpha/2,\,\gamma}$ = 자유도가 γ 인 카이제곱 분포의 (α/2) 번째 분위수

$$\gamma = \frac{n\left(1 + \left(\frac{\bar{x} - T}{s}\right)^2\right)^2}{1 + 2\left(\frac{\bar{x} - T}{s}\right)^2}$$

n = 관측값 수

$\bar{x}$ = 관측값의 평균

T = 목표값

s = 장기 시그마 추정값

참고 : CPM 에 대한 신뢰 구간은 목표값이 규격 하한과 규격 상한 사이에서 중심화되는 경우에만 계산됩니다 .

CPL 과 CPU 의 신뢰 하한 및 상한은 Chou et al. (1990) 의 방법을 사용하여 계산됩니다 .

CPL 에 대한 100(1 - α)% 신뢰 한계 (CPL_L, CPL_U 로 표시) 는 다음 방정식을 충족합니다 .

$$\Pr[t_{n-1}(\delta_L) \geq 3\hat{C}_{pl}\sqrt{n}] = \alpha/2 \qquad \text{여기서 , } \delta_L = 3CPL_L\sqrt{n}$$

$$\Pr[t_{n-1}(\delta_U) \leq 3\hat{C}_{pl}\sqrt{n}] = \alpha/2 \qquad \text{여기서 , } \delta_U = 3CPL_U\sqrt{n}$$

다음은 각 요소에 대한 설명입니다 .

$t_{n\text{-}1}(\delta)$ 은 n - 1 자유도와 비중심성 모수 δ 를 사용하는 비중심 t 분포를 따릅니다 .

$\hat{C}_{pl}$ = Cpl 에 대한 추정값

CPU 에 대한 100(1 - α)% 신뢰 한계 (CPU_L, CPU_U 로 표시) 는 다음 방정식을 충족합니다 .

$$\Pr[t_{n-1}(\delta_L) \geq 3\hat{C}_{pu}\sqrt{n}] = \alpha/2 \qquad \text{여기서 , } \delta_L = 3CPU_L\sqrt{n}$$

$$\Pr[t_{n-1}(\delta_U) \leq 3\hat{C}_{pu}\sqrt{n}] = \alpha/2 \qquad \text{여기서 , } \delta_U = 3CPU_U\sqrt{n}$$

다음은 각 요소에 대한 설명입니다 .

$t_{n\text{-}1}(\delta)$ 은 n - 1 자유도와 비중심성 모수 δ 를 사용하는 비중심 t 분포를 따릅니다 .

$\hat{C}_{pu}$ = Cpu 에 대한 추정값

비정규 분포에 대한 공정 능력 지수

이 섹션에서는 비정규 분포에 대한 공정 능력 지수를 계산하는 방법을 설명합니다 . 이러한 일반화된 공정 능력 지수는 다음과 같이 정의됩니다 .

$$\text{Cp} = \frac{USL - LSL}{P_{0.99865} - P_{0.00135}}$$

$$\text{Cpk} = \min(C_{pl}, C_{pu})$$

$$\text{Cpm} = \frac{\min\left(\frac{T - LSL}{P_{0.5} - P_{0.00135}}, \frac{USL - T}{P_{0.99865} - P_{0.5}}\right)}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

$$\text{Cpl} = \frac{P_{0.5} - LSL}{P_{0.5} - P_{0.00135}}$$

$$\text{Cpu} = \frac{USL - P_{0.5}}{P_{0.99865} - P_{0.5}}$$

다음은 각 요소에 대한 설명입니다 .

LSL = 규격 하한

USL = 규격 상한

T = 목표값

P_α = 적합 분포의 α*100 번째 백분위수

Cpm 계산에서 μ와 σ는 적합 분포의 기대값 및 분산 제곱근을 사용하여 추정됩니다. 모수 추정값 보고서의 모수와 적합 분포의 기대값 및 분산 간의 관계에 대한 자세한 내용은 **기본 분석**의 에서 확인하십시오.

시그마 품질 통계량

각 부분(LSL 아래, USL 위, 규격 밖 전체)에 대한 시그마 품질 통계량은 다음과 같이 계산됩니다.

$$\text{시그마 품질} = \Phi^{-1}{}_{1 - \text{Pct}/100} + 1.5$$

다음은 각 요소에 대한 설명입니다 .

Pct = 보고서에서 " 백분율 " 열의 값

$\Phi^{-1}{}_{1 - \text{Pct}/100}$ = 표준 정규 분포의 (1 - Pct/100) 번째 분위수

참고 : LSL 아래 백분율과 USL 위 백분율을 합하면 규격 밖 전체 백분율 값이 되지만 LSL 아래 시그마 품질과 USL 위 시그마 품질을 합하면 규격 밖 전체 시그마 품질 값이 되지는 않습니다 . 시그마 품질을 계산할 때는 정규 분포 분위수 찾기가 포함되어 가산되지 않기 때문입니다 .

벤치마크 Z 통계량

벤치마크 Z 통계량은 정규 분포를 기반으로 하는 공정 능력 분석에만 사용할 수 있습니다 . 벤치마크 Z 통계량은 다음과 같이 계산됩니다 .

$$\text{Z Bench} = \Phi^{-1}{}_{1 - P(LSL) - P(USL)}$$

$$\text{Z LSL} = \frac{\mu - LSL}{\sigma} = 3 * \text{Cpl}$$

$$\text{Z USL} = \frac{USL - \mu}{\sigma} = 3 * \text{Cpu}$$

다음은 각 요소에 대한 설명입니다 .

LSL = 규격 하한

USL = 규격 상한

μ = 표본 평균

σ = 표본 표준편차

$\Phi^{-1}_{1 - P(LSL) - P(USL)}$ = 표준 정규 분포의 (1 - *P(LSL)* -*P(USL)*) 번째 분위수

P(LSL) = Prob(*X* < *LSL*) = 1 - Φ(Z LSL).

P(USL) = Prob(*X* > *USL*) = 1 - Φ(Z USL).

Φ = 표준 정규 누적 분포 함수

V- 마스크 CUSUM 관리도에 대한 통계 상세 정보

이러한 계산식에는 다음 표기가 사용됩니다 .

- μ 는 모집단의 평균을 나타내며 공정 평균 또는 공정 수준이라고도 합니다 .
- μ_0 은 모집단의 목표 평균 (목표) 을 나타냅니다 . 경우에 따라 $\bar{X}_0$ 기호가 μ_0 에 사용됩니다 . 자세한 내용은 American Society for Quality Statistics Division(2004) 에서 확인하십시오 . 시작 창의 "CUSUM 차트에 대한 알려진 통계" 영역에서 μ_0을 "목표값"으로 제공할 수 있습니다.
- σ 는 모집단 표준편차를 나타내고 $\hat{\sigma}$ 는 σ 의 추정값을 나타냅니다 .
- σ_0 은 알려진 표준편차를 나타냅니다 . 시작 창의 "CUSUM 차트에 대한 알려진 통계 " 영역에서 σ_o 을 " 시그마 " 로 제공할 수 있습니다 .
- *n* 은 CUSUM 차트의 명목형 표본 크기를 나타냅니다 .
- δ 는 감지할 μ 변화를 나타내며 , 표준편차의 배수로 표현됩니다 . 시작 창의 "CUSUM 차트에 대한 알려진 통계 " 영역에서 δ 를 " 델타 " 로 제공할 수 있습니다 .
- Δ 는 감지할 μ 변화를 나타내며 , 데이터 단위로 표현됩니다 . 전체 부분군에서 표본 크기 *n* 이 일정하면 다음 계산이 적용됩니다 .

  $\Delta = \delta\sigma_{\bar{X}} = (\delta\sigma)/\sqrt{n}$

  시작 창의 "CUSUM 차트에 대한 알려진 통계 " 영역에서 Δ 를 " 변화 " 로 제공할 수 있습니다 .

참고 : 일부 작성자는 Δ 대신 D 기호를 사용합니다 .

단측 CUSUM 차트

양의 변화

감지할 변화 (δ) 가 양수이면 t 번째 부분군의 CUSUM 이 다음과 같이 계산됩니다 .

$S_t = \max(0, S_{t-1} + (z_t - k))$

$t = 1, 2,..., n$. 여기서 $S_0 = 0$ 이고 z_t 는 양측 차트에 대해 정의됩니다 . **참조 값**이라고 하는 k 모수는 양수입니다 . 시작 창에서 k 모수가 지정되지 않은 경우 k 는 δ/2 로 설정됩니다 . CUSUM S_t 를 **상위 누적합**이라고 합니다 . S_t 는 다음과 같이 계산할 수 있습니다 .

$$\max\left(0, S_{t-1} + \frac{\overline{X}_i - (\mu_0 + k\sigma_{\overline{X}_t})}{\sigma_{\overline{X}_t}}\right)$$

S_t 수열은 μ_0으로부터 k 표준 오차보다 큰 부분군 평균의 편차를 누적합니다. S_t가 양수 값 h(**결정 구간**이라고 함)를 초과하면 변화 또는 관리이탈 상태 신호가 표시됩니다.

음의 변화

감지할 변화가 음수이면 t 번째 부분군의 CUSUM 이 다음과 같이 계산됩니다 .

$$S_t = \max(0, S_{t-1} - (z_t + k))$$

$t = 1, 2,..., n$. 여기서 $S_0 = 0$ 이고 z_t 는 양측 차트에 대해 정의됩니다 . **참조 값**이라고 하는 k 모수는 양수입니다 . 시작 창에서 k 모수가 지정되지 않은 경우 k 는 δ/2 로 설정됩니다 . CUSUM S_t 를 **하위 누적합**이라고 합니다 . S_t 는 다음과 같이 계산할 수 있습니다 .

$$\max\left(0, S_{t-1} - \frac{\overline{X}_i - (\mu_0 - k\sigma_{\overline{X}_t})}{\sigma_{\overline{X}_t}}\right)$$

S_t 수열은 μ_0으로부터 k 표준 오차보다 작은 부분군 평균의 편차 절대값을 누적합니다. S_t가 양수 값 h(**결정 구간**이라고 함)를 초과하면 변화 또는 관리이탈 상태 신호가 표시됩니다.

δ가 양수 또는 음수인지에 관계없이 S_t는 항상 양수이고 h도 항상 양수입니다. 음의 변화를 감지하기 위한 차트의 경우 일부 작성자는 S_t가 음수 한계보다 작을 때 변화 신호가 발생하는 S_t의 반사형 버전을 정의합니다.

Lucas and Crosier(1982)에서는 초기 CUSUM S_0이 "시작값"으로 설정된 CUSUM 차트에 대한 FIR(Fast Initial Response) 기능의 특성을 설명합니다. 이 차트에서 제공하는 평균 런 길이 계산을 통해 공정이 관리 상태일 때는 FIR 기능이 미치는 효과가 거의 없으며, 결과적으로 초기 관리이탈 상태에 대한 반응이 표준 CUSUM 차트보다 빠르다는 것을 알 수 있습니다. 시작 창의 "CUSUM 차트에 대한 알려진 통계" 영역에서 "시작값"을 제공할 수 있습니다.

상수 표본 크기

부분군 표본 크기가 일정하면 (n) 데이터와 동일한 단위로 척도화된 CUSUM 을 계산하는 것이 더 좋을 수 있습니다 . 이 경우 CUSUM 은 다음과 같이 계산됩니다 .

$$S_t = \max(0, S_{t-1} + (\overline{X}_t - (\mu_0 + k\sigma/\sqrt{n})))$$

여기서 , $\delta > 0$

$$S_t = \max(0, S_{t-1} - (\overline{X}_t - (\mu_0 - k\sigma/\sqrt{n})))$$

여기서, $\delta < 0$입니다. 두 경우 모두 k 모수의 척도가 $k' = k\sigma/\sqrt{n}$로 조정됩니다. 시작 창에서 k 모수가 지정되지 않은 경우 k'는 $\delta/2$로 설정됩니다. S_t가 $h' = h\sigma/\sqrt{n}$를 초과하면 변화 신호가 발생합니다. 일부 작성자는 h'에 H 기호를 사용합니다.

양측 CUSUM 차트

CUSUM 차트가 양측인 경우 t번째 부분군에 대해 표시되는 누적합 S_t는 다음과 같이 정의됩니다.

$$S_t = S_{t-1} + z_t$$

$t = 1, 2, \ldots, n$. 여기서 $S_0=0$ 이고 z_t 항은 다음과 같이 계산됩니다.

$$z_t = (\bar{X}_t - \mu_0)/(\sigma/\sqrt{n_t})$$

여기서 $\bar{X}_t$는 t 번째 부분군 평균이고 n_t는 t 번째 부분군 표본 크기입니다. 부분군 표본이 개별 측정값 x_t로 구성된 경우 z_t 항 계산은 다음과 같이 단순화됩니다.

$$z_t = (x_t - \mu_0)/\sigma$$

첫 번째 방정식은 다음과 같이 다시 쓸 수 있습니다.

$$S_t = \sum_{i=1}^{t} z_i = \sum_{i=1}^{t} (\bar{X}_i - \mu_0)/\sigma_{\bar{X}_i}$$

여기서 S_t 수열은 목표 평균 μ_0과 부분군 평균 간의 표준화된 편차를 누적합니다.

대부분의 경우 부분군 표본 크기 n_i는 일정하고 $(n_i = n)$ S_t에 대한 방정식은 다음과 같이 단순화될 수 있습니다.

$$S_t = (1/\sigma_{\bar{X}}) \sum_{i=1}^{t} (\bar{X}_i - \mu_0) = (\sqrt{n}/\sigma) \sum_{i=1}^{t} (\bar{X}_i - \mu_0)$$

경우에 따라 S_t를 다음과 같이 계산하는 것이 더 좋을 수 있습니다.

$$S_t = \sum_{i=1}^{t} (\bar{X}_i - \mu_0)$$

이렇게 하면 데이터와 동일한 단위로 척도화됩니다. 이 경우 V-마스크 모수 h와 k는 각각 $h' = h\sigma/\sqrt{n}$ 및 $k' = k\sigma/\sqrt{n}$로 척도가 조정됩니다. 일부 작성자는 k'에 F 기호를 사용하고 h'에 H 기호를 사용합니다.

공정이 관리 상태이고 평균 μ가 목표값 μ_0이거나 이 값에 가까우면 확률 보행 모형이 적용됩니다. 따라서 점이 0에서 멀어질 수 있지만 μ_0에서부터 양의 이동과 음의 이동은 서로 상쇄되는 경향이 있으므로 큰 추세를 보이지는 않습니다. μ가 양의 방향으로 이동하면 점이 상향 추세를 나타내고 μ가 음의 방향으로 이동하면 점이 하향 추세를 나타냅니다.

가중 이동 평균 차트에 대한 통계 상세 정보

UWMA 차트에 대한 관리 한계

각 부분군 i 에 대한 UWMA 차트의 관리 한계는 다음과 같이 계산됩니다 .

$$LCL_i = \mu_0 - k\frac{\hat{\sigma}}{min(i,w)}\sqrt{\frac{1}{n_i} + \frac{1}{n_{i-1}} + \ldots + \frac{1}{n_{1+max(i-w,0)}}}$$

$$UCL_i = \mu_0 + k\frac{\hat{\sigma}}{min(i,w)}\sqrt{\frac{1}{n_i} + \frac{1}{n_{i-1}} + \ldots + \frac{1}{n_{1+max(i-w,0)}}}$$

다음은 각 요소에 대한 설명입니다 .

w = 범위 모수 (이동 평균의 항 수)

n_i = i 번째 부분군의 표본 크기

k = 표준편차의 수

μ_0 = 부분군 평균의 가중 평균

$\hat{\sigma}$ = 추정된 공정 표준편차

EWMA 차트에 대한 관리 한계

EWMA 차트의 관리 한계는 다음과 같이 계산됩니다 .

$$LCL = \mu_0 - k\hat{\sigma}r\sqrt{\sum_{j=0}^{i-1}\frac{(1-r)^{2j}}{n_{i-j}}}$$

$$UCL = \mu_0 + k\hat{\sigma}r\sqrt{\sum_{j=0}^{i-1}\frac{(1-r)^{2j}}{n_{i-j}}}$$

다음은 각 요소에 대한 설명입니다 .

r = EWMA 가중치 모수 ($0 < r \leq 1$)

x_{ij} = i 번째 부분군의 j 번째 측정값 ($j = 1, 2, 3, \ldots, n_i$)

n_i = i 번째 부분군의 표본 크기

k = 표준편차의 수

μ_0 = 부분군 평균의 가중 평균

$\hat{\sigma}$ = 추정된 공정 표준편차

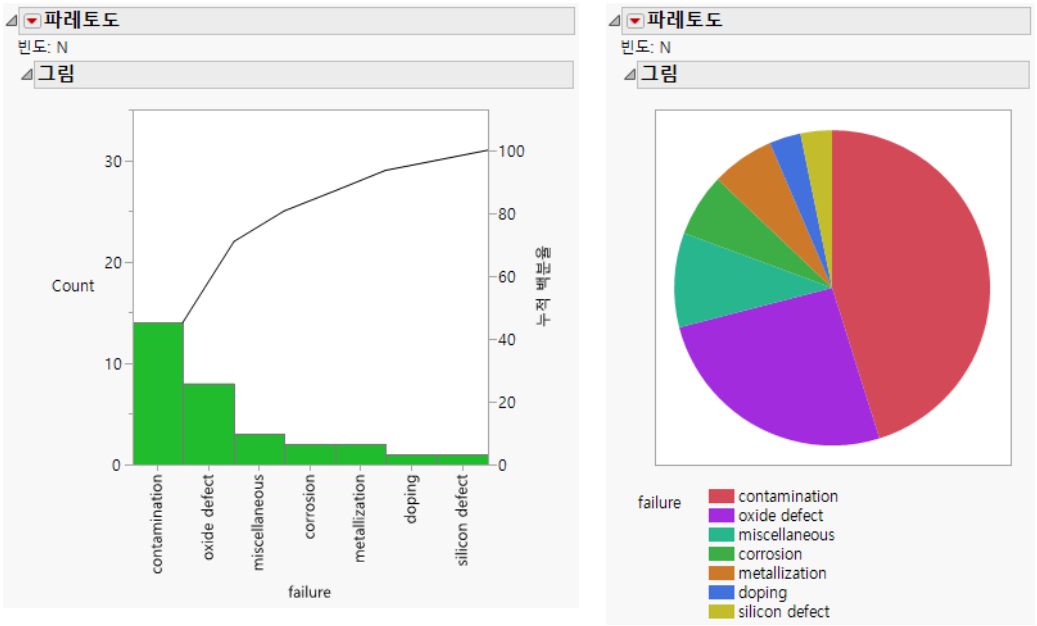
13 장

파레토도

근본 원인을 개선하기 위한 노력에 집중

파레토도를 사용하여 공정 또는 작업의 통계적 품질을 개선합니다. 파레토도는 품질 관련 공정 또는 작업에서 문제의 심각도(빈도)를 보여 주는 차트입니다. 파레토도는 문제의 빈도와 심각도를 강조 표시하여 어떤 문제를 먼저 해결해야 할지 결정하는 데 도움이 됩니다.

그림 13.1 파레토도 예제



목차

파레토도 플랫폼 개요 351
파레토도 플랫폼의 예 351
파레토도 플랫폼 시작 354
파레토도 보고서 355
파레토도 플랫폼 옵션 356
원인 옵션 357
파레토도 플랫폼의 추가 예 358
결합 원인의 임계 예제 358
그룹 전체에서 상수 크기 사용 예제 360
그룹 전체에서 상수가 아닌 표본 크기 사용 예제 361
일원 비교 파레토도 예제 363
이원 비교 파레토도 예제 364
파레토도 플랫폼에 대한 통계 상세 정보 366
가능도비 카이제곱 검정 366

파레토도 플랫폼 개요

파레토도 플랫폼은 품질 관련 공정 또는 작업에서 문제의 상대적 빈도나 심각도를 표시하는 차트를 생성합니다. 처음에 파레토도는 문제 분류를 내림차순으로 정렬하여 보여 주는 막대 차트로 표시됩니다. 값이 문제의 원인인 열에는 *Y* 역할이 할당되고 **공정 변수**라고 합니다.

동일한 공정 변수에 대해 둘 이상의 파레토도를 결합하여 비교 파레토도를 생성할 수도 있습니다. *X* 역할이 할당된 열의 각 값 또는 두 *X* 변수의 수준 조합에 대한 그림을 단일 표시에서 보여 줍니다. *X* 역할이 할당된 열을 **분류 변수**라고 합니다.

파레토도는 *X* 분류 변수가 없거나, 단일 *X* 변수 또는 두 개의 *X* 변수와 함께 단일 *Y*(공정) 변수를 차트로 작성할 수 있습니다. Pareto 함수는 수치형 변수와 문자형 변수를 구분하거나 모델링 유형을 구분하지 않습니다. 막대 차트와 파이 차트 사이를 전환할 수는 있습니다. 모든 값은 이산형으로 처리되며 막대와 파이는 개수 또는 백분율을 나타냅니다.

파레토도 플랫폼의 예

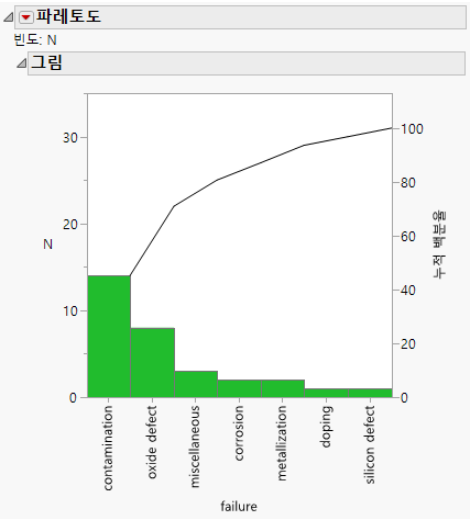
이 예에서는 고장 데이터와 빈도 열이 포함된 Failure.jmp 샘플 데이터 테이블을 사용합니다 . 여기에는 집적 회로 제작 중의 고장 원인과 각 결함 유형별 발생 횟수가 나열되어 있습니다 . 분석을 통해 공정 고장에 가장 많이 기여하는 요인을 확인할 수 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Failure.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 파레토도**를 선택합니다 .
3. failure 를 선택하고 **Y, 원인**을 클릭합니다 .

   이 열에는 고장 원인이 나열됩니다 . 이 변수를 검사하려고 합니다 .
4. N 을 선택하고 **빈도**를 클릭합니다 .

   이 열에는 각 유형의 고장이 발생한 횟수가 나열됩니다 .
5. **확인**을 클릭합니다 .

그림 13.2 파레토도 보고서 창



왼쪽 축은 고장 횟수를 나타내고 오른쪽 축은 각 범주의 고장 비율을 나타냅니다 . 막대는 가장 자주 발생하는 고장이 왼쪽에 있는 내림차순으로 정렬됩니다. 곡선은 누적된 고장을 왼쪽에서 오른쪽으로 나타냅니다 .

6. " 파레토도 " 의 빨간색 삼각형을 클릭하고 **누적 백분율 점 라벨 지정**을 선택합니다 .

   Contamination(오염) 이 고장의 약 45% 를 차지합니다 . Oxide Defect(산화물 결함) 막대 위의 점은 Contamination과 Oxide Defect가 고장의 약 71%를 차지한다는 것을 보여 줍니다.

7. " 파레토도 " 의 빨간색 삼각형을 클릭하고 **누적 백분율 점 라벨 지정** 및 **누적 백분율 곡선 표시**를 선택 취소합니다 .

8. **N** 라벨로 표시된 y 축의 라벨을 클릭하고 **Count** 로 이름을 바꿉니다 .

9. y 축을 두 번 클릭하여 **Y 축 설정** 창을 표시합니다 .
   - **최대값** 필드에 15 를 입력합니다 .
   - **증분** 필드에 2 를 입력합니다 .
   - **축 라벨 행** 패널에서 **주** 격자선에 대해 **격자선**을 선택합니다 .
   - **확인**을 클릭합니다 .

10. " 파레토도 " 의 빨간색 삼각형을 클릭하고 **범주 범례**를 선택합니다 .

그림 13.3 표시 옵션이 있는 파레토도

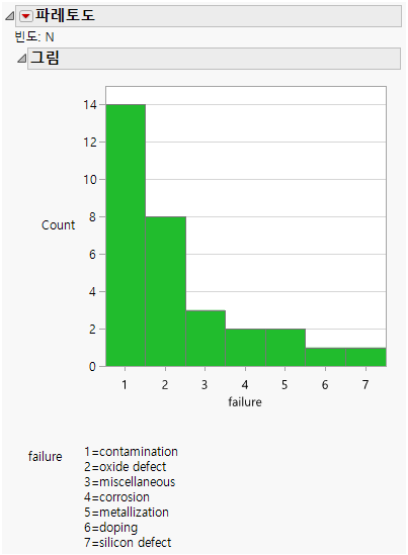
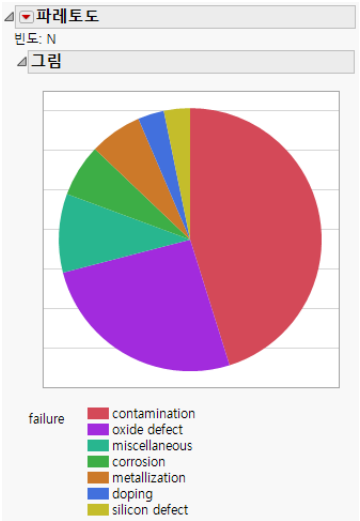


그림 13.3 에는 다양한 유형의 고장 수가 표시되고 범주 범례가 있습니다 . 세로 축 (Count) 의 척도가 재조정되고 주 눈금 표식에 격자선이 있습니다 .

11. 데이터를 파이 차트로 보려면 " 파레토도 " 의 빨간색 삼각형을 클릭하고 **파이 차트**를 선택합니다 .

그림 13.4 파이 차트로 표시한 파레토도

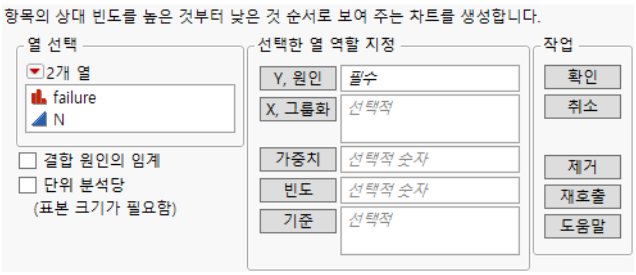


Contamination 과 Oxide Defect 가 고장의 대부분을 차지한다는 것을 명확하게 나타냅니다.

파레토도 플랫폼 시작

분석 > 품질 및 공정 > 파레토도를 선택하여 파레토도 플랫폼을 시작할 수 있습니다 .

그림 13.5 파레토도 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

파레토도 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y, 원인 값이 문제의 원인인 열을 식별합니다 . 이 변수를 공정 변수라고 하며 검사 대상입니다 .

X, 그룹화 그룹화 요인을 식별합니다 . 그룹화 변수는 각 값에 대한 그림이 나란히 배치된 하나의 파레토도 창을 생성합니다 . 그룹화 변수가 없거나 , 하나의 그룹화 변수 (" 일원 비교 파레토도 예제 " 참조) 또는 두 개의 그룹화 변수 (" 이원 비교 파레토도 예제 " 참조) 를 사용할 수 있습니다 .

가중치 관측값에 서로 다른 가중치를 지정하기 위해 변수를 할당합니다 .

빈도 빈도 값을 포함하는 열을 식별합니다 .

기준 열에 나타나는 각 값에 대해 별도의 분석을 생성하기 위해 변수를 식별합니다 .

결합 원인의 임계 최소 비율 또는 개수를 지정하여 원인 결합을 위한 임계값을 지정할 수 있습니다 . 이 옵션을 선택한 후 **꼬리 %** 또는 **개수**를 선택하고 임계값을 입력합니다 ." 꼬리 %" 옵션은 지정된 총계 백분율에 대해 더 작은 수의 그룹을 결합합니다 (결합된 작은 그룹 수 / 총 그룹 수)." 개수 " 옵션을 사용하면 특정 개수 임계값을 지정할 수 있습니다 . 예는 " 결합 원인의 임계 예제 " 에서 확인하십시오 .

단위별 분석 그룹 간 결함 비율을 비교할 수 있습니다 . JMP 에서는 결함 비율 및 결함 비율의 95% 신뢰 구간을 계산합니다 . 이 옵션을 선택한 후 **상수** 또는 **빈도 열의 값**을 선택하고 표본 크기 값 또는 원인 코드를 각각 입력합니다 ." 상수 " 옵션을 사용하면 시작 창에서 상수 표본 크기를 지정할 수 있습니다 ." 빈도 열의 값 " 옵션을 사용하면 특수 원인 코드를 통해 그룹에 고유 표본 크기를 지정하여 행을 원인 행으로 지정할 수 있습니다 .

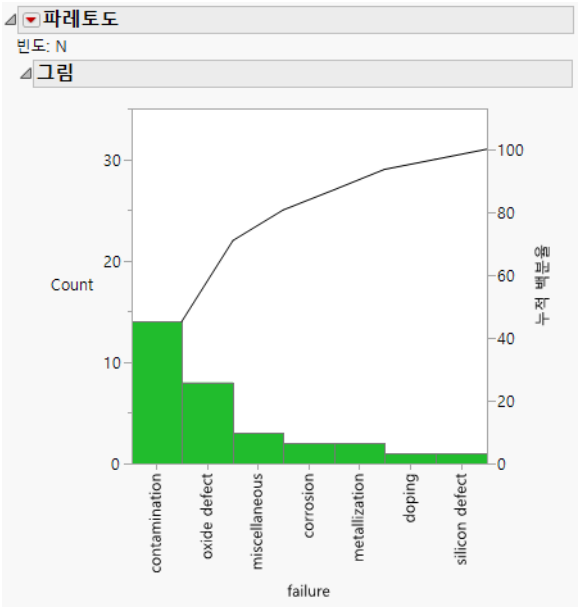
파레토도에서 원인을 결합할 수 있지만 이러한 분석을 위한 계산은 이에 대응하여 변경되지 않습니다 .

예는 " 그룹 전체에서 상수 크기 사용 예제 " 및 " 그룹 전체에서 상수가 아닌 표본 크기 사용 예제 " 에서 확인하십시오 .

파레토도 보고서

파레토도는 데이터의 변수 백분율을 표시하는 막대 차트와 변수의 누적 백분율을 표시하는 선 그래프를 결합합니다 .

그림 13.6 파레토도 예제



파레토도는 *X* 분류 변수가 없거나 , 단일 *X* 변수 또는 두 개의 *X* 변수와 함께 단일 *Y*(공정) 변수를 차트로 작성할 수 있습니다 . 파레토도는 수치형 변수와 문자형 변수를 구분하거나 모델링 유형을 구분하지 않습니다 . 모든 값은 이산형으로 처리되며 막대는 개수 또는 백분율을 나타냅니다 . 다음 목록에서는 파레토도의 배열 방식을 설명합니다 .

- *X* 분류 변수가 없는 *Y* 변수는 *Y* 변수의 각 값에 대한 막대가 포함된 단일 차트를 생성합니다 . 예는 " 파레토도 플랫폼의 예 " 에서 확인하십시오 .
- 하나의 *X* 분류 변수가 있는 *Y* 변수는 일렬로 표시되는 파레토도를 생성합니다 . 각 *Y* 수준에 대한 막대가 포함된 각 *X* 변수 수준의 그림이 있습니다 . 이러한 그림을 비교 파레토도의 **셀** 이라고 합니다 . *X*(분류) 변수의 각 수준에 대한 셀이 있습니다 . *X* 변수가 하나뿐이므로 이를 **일원 비교 파레토도**라고 합니다 . 예는 " 일원 비교 파레토도 예제 " 에서 확인하십시오 .
- 두 개의 *X* 변수가 있는 *Y* 변수는 행과 열로 구성된 파레토도를 생성합니다 . 첫 번째 *X* 변수의 각 수준에 대한 행과 두 번째 *X* 변수의 각 수준에 대한 열이 있습니다 . *X* 변수가 두 개 있으므로 이를 **이원 비교 파레토도**라고 합니다 . 앞에서 설명한 대로 행에는 첫 번째 *X* 변수의 각 값에 대한 파레토도가 있습니다 . 왼쪽 위의 셀을 **키 셀**이라고 합니다 . 이 셀의 막대는 내림차순으로 정렬됩니다 . 다른 셀의 막대는 키 셀과 동일한 순서로 나타납니다 . 셀의 행과 열

을 재정렬할 수 있습니다. 왼쪽 위로 이동하는 셀이 새 키 셀이 되고 이에 따라 다른 모든 셀의 막대가 재정렬됩니다. 예는 " 이원 비교 파레토도 예제 " 에서 확인하십시오.

- 각 막대는 연결된 데이터 테이블에서 해당 *Y* 수준에 대한 행에 할당된 색상으로 표시됩니다. 그렇지 않으면 *Y* 수준에 색상이 할당된 행이 없는 모든 막대에 단일 색상이 사용됩니다. *Y* 수준의 행에 서로 다른 색상이 할당된 경우 해당 *Y* 수준에 대한 막대는 데이터 테이블에서 해당 *Y* 수준의 첫 번째 행에 할당된 색상으로 표시됩니다.

"파레토도"의 빨간색 삼각형 메뉴에 포함된 옵션을 사용하여 척도 유형 및 막대 배열을 변경하고 막대를 파이 차트로 변환할 수 있습니다. 자세한 내용은 "파레토도 플랫폼 옵션"에서 확인하십시오.

파레토도 플랫폼 옵션

" 파레토도 " 의 빨간색 삼각형 메뉴에는 그림 모양을 사용자 정의하는 옵션이 포함되어 있습니다. 또한 **원인** 하위 메뉴에는 파레토도 내의 개별 막대에 영향을 주는 옵션도 있습니다. 다음 명령은 파레토도의 전체 모양에 영향을 줍니다.

백분율 척도 개수 및 백분율 왼쪽 세로 축 표시를 표시하거나 숨깁니다.

범례 수 그림 영역에 총 표본 크기를 표시하거나 숨깁니다.

범주 범례 라벨이 지정된 막대 및 별도의 범주 범례를 표시하거나 숨깁니다.

파이 차트 막대 차트 및 파이 차트 표현을 표시하거나 숨깁니다.

가로 재정렬 , 세로 재정렬 하나 이상의 그룹화 변수가 있을 때 그룹화된 파레토도를 재정렬합니다.

그림 그룹 해제 파레토 차트 그룹을 개별 그림으로 분할합니다.

가산 분석 단위별 결함 분석을 수행합니다. 결함 비율을 비교하고 그룹 내 / 그룹 간 비율 검정을 수행할 수 있습니다.

단위별 비율 그룹 간 결함 비율을 비교합니다 . 표본 크기가 지정된 경우 DPU(Defects Per Unit) 및 PPM(Parts Per Million) 열이 보고서에 추가됩니다.

그룹 내 비율 검정 그룹 내에서 원인 간 비율이 동일한지 여부를 확인하기 위해 가능도비 카이제곱 검정을 수행합니다. 자세한 내용은 " 파레토도 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오.

그룹 간 비율 검정 각 원인에 대한 비율이 그룹 간에 동일한지 여부를 확인하기 위해 가능도비 카이제곱 검정을 수행합니다 . 자세한 내용은 " 파레토도 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오.

누적 백분율 곡선 표시 막대 위의 누적 백분율 곡선과 오른쪽 세로 축의 누적 백분율 축을 표시하거나 숨깁니다.

누적 백분율 축 표시 오른쪽 세로 축의 누적 백분율 축을 표시하거나 숨깁니다.

누적 백분율 점 표시 누적 백분율 곡선의 점을 표시하거나 숨깁니다.

누적 백분율 점 라벨 지정 누적 곡선의 점에 라벨을 표시하거나 숨깁니다.

누적 백분율 곡선 색상 누적 백분율 곡선의 색상을 변경합니다.

원인 하나 이상의 개별 차트 막대에 영향을 주는 옵션을 포함합니다. 이러한 옵션에 대한 설명은 "원인 옵션"에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

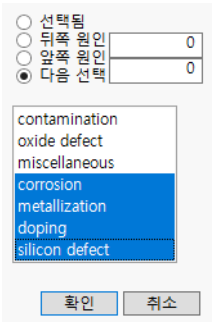
그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

원인 옵션

막대를 클릭하여 강조 표시할 수 있습니다. 인접하지 않은 여러 막대를 선택하려면 Ctrl 키를 누른 채 막대를 클릭하십시오. 막대를 선택하면 빨간색 삼각형 메뉴에서 파레토도 막대에 영향을 주는 명령에 액세스할 수 있습니다. 이러한 명령은 빨간색 삼각형 메뉴의 **원인** 하위 메뉴에 있습니다. 그림 영역의 아무 곳에서나 마우스 오른쪽 버튼을 클릭해도 해당 옵션을 사용할 수 있습니다. 다음 옵션은 차트 전체가 아니라 강조 표시된 막대에 적용됩니다.

원인 결합 선택한 (강조 표시된) 막대를 결합합니다. **선택됨**, **뒤쪽 원인**, **앞쪽 원인**을 선택하거나 변수 목록에서 선택할 수 있습니다.

그림 13.7 원인 결합 창



원인 분리 선택한 막대를 원래 성분 막대로 분리합니다.

처음으로 이동 하나 이상의 강조 표시된 막대를 왼쪽 (첫 번째) 위치로 이동합니다.

마지막으로 이동 하나 이상의 강조 표시된 막대를 오른쪽 (마지막) 위치로 이동합니다.

색상 하나 이상의 강조 표시된 막대에 색상을 적용하기 위한 색상 팔레트를 표시합니다 .

표식 **누적 백분율 점 표시** 명령이 활성화된 경우 누적 백분율 곡선의 점에 표식을 할당하기 위한 표식 팔레트를 표시합니다 .

라벨 강조 표시된 모든 막대의 맨 위에 막대 값을 표시합니다 .

파레토도 플랫폼의 추가 예

- " 결합 원인의 임계 예제 "
- " 그룹 전체에서 상수 크기 사용 예제 "
- " 그룹 전체에서 상수가 아닌 표본 크기 사용 예제 "
- " 일원 비교 파레토도 예제 "
- " 이원 비교 파레토도 예제 "

결합 원인의 임계 예제

이 예에서는 고장 데이터와 빈도 열이 포함된 Failure.jmp 샘플 데이터 테이블을 사용합니다 . 여기에는 집적 회로 제작 중의 고장 원인과 각 결함 유형별 발생 횟수가 나열되어 있습니다 . 이 예에는 임계값 2 가 지정됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Failure.jmp 를 엽니다 .
2. **분석** > **품질 및 공정 > 파레토도**를 선택합니다 .
3. failure 를 선택하고 **Y, 원인**을 클릭합니다 .
4. N 을 선택하고 **빈도**를 클릭합니다 .
5. **결합 원인의 임계**를 선택한 후 **개수**를 선택합니다 .
6. 임계값으로 2 를 입력합니다 .
7. **확인**을 클릭합니다 .

그림 13.8 임계 개수가 2 인 파레토도

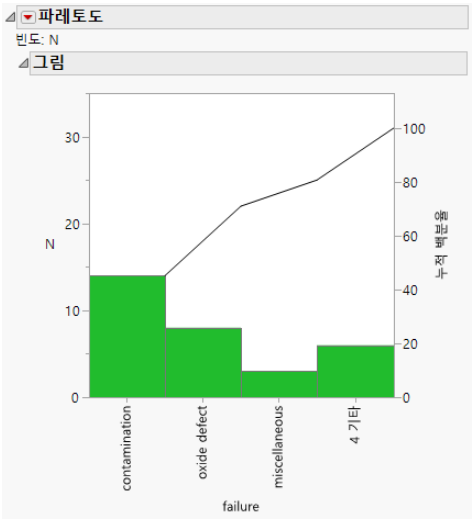
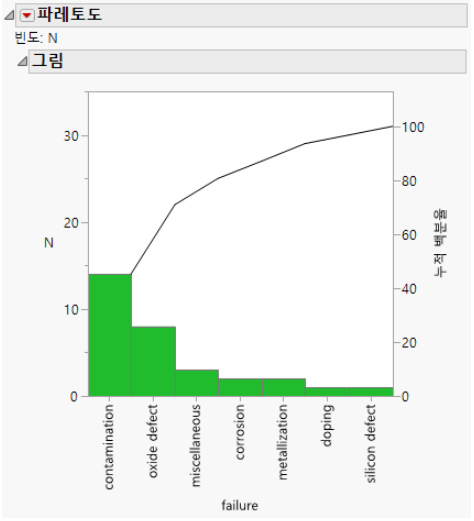


그림 13.8 에서는 개수 2 를 지정한 후의 그림을 보여 줍니다 . 개수가 2 이하인 모든 원인이 "4 기타 " 라벨이 지정된 마지막 막대로 결합됩니다 .

8. 결합된 막대를 그림 13.9 에 표시된 것과 같이 원래 범주로 분리하려면 **원인** > **원인 분리**를 선택합니다 .

그림 13.9 원인이 분리된 파레토도

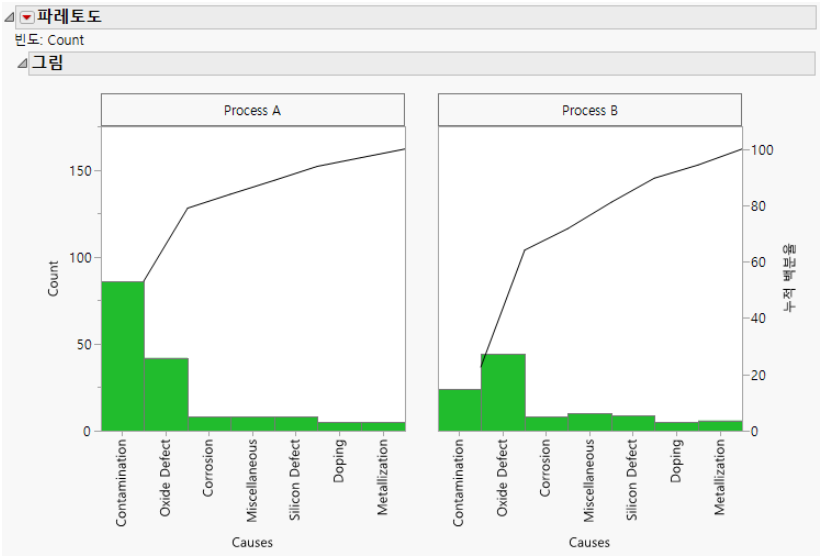


그룹 전체에서 상수 크기 사용 예제

이 예에서는 고장 데이터와 빈도 열이 포함된 Failures.jmp 샘플 데이터 테이블을 사용합니다 . 여기에는 두 공정에 대해 집적 회로 제작 중의 고장 원인과 각 결함 유형별 발생 횟수가 나열되어 있습니다 . 이 예에는 상수 표본 크기 1000 이 지정됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Failures.jmp 를 엽니다 .
2. **분석** > **품질 및 공정 > 파레토도**를 선택합니다 .
3. Causes 를 선택하고 **Y, 원인**을 클릭합니다 .
4. Process 를 선택하고 **X, 그룹화**를 클릭합니다 .
5. Count 를 선택하고 **빈도**를 클릭합니다 .
6. **단위별 분석**을 선택한 후 **상수**를 선택합니다 .
7. **표본 크기**에 1000 을 입력합니다 .
8. **확인**을 클릭합니다 .

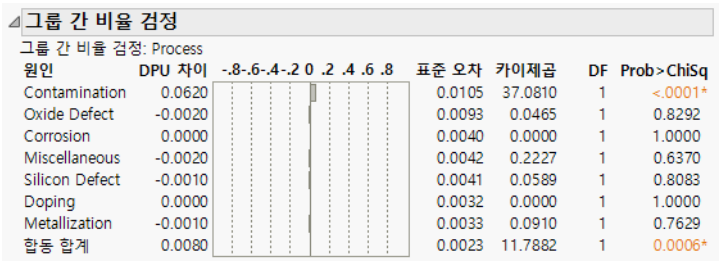
그림 13.10 파레토도 보고서 창



Process A 에서는 Contamination(오염) 이 가장 큰 고장 원인으로 나타나고 Process B 에서는 Oxide Defect(산화물 결함) 가 가장 중요한 고장 원인으로 나타납니다 .

9. " 파레토도 " 의 빨간색 삼각형을 클릭하고 **가산 분석** > **그룹 간 비율 검정**을 선택합니다 .

그림 13.11 그룹 간 비율 검정 결과



그룹 간 비율 검정

그룹 간 비율 검정: Process

원인	DPU 차이	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8	표준 오차	카이제곱	DF	Prob>ChiSq
Contamination	0.0620		0.0105	37.0810	1	<.0001*
Oxide Defect	-0.0020		0.0093	0.0465	1	0.8292
Corrosion	0.0000		0.0040	0.0000	1	1.0000
Miscellaneous	-0.0020		0.0042	0.2227	1	0.6370
Silicon Defect	-0.0010		0.0041	0.0589	1	0.8083
Doping	0.0000		0.0032	0.0000	1	1.0000
Metallization	-0.0010		0.0033	0.0910	1	0.7629
합동 합계	0.0080		0.0023	11.7882	1	0.0006*

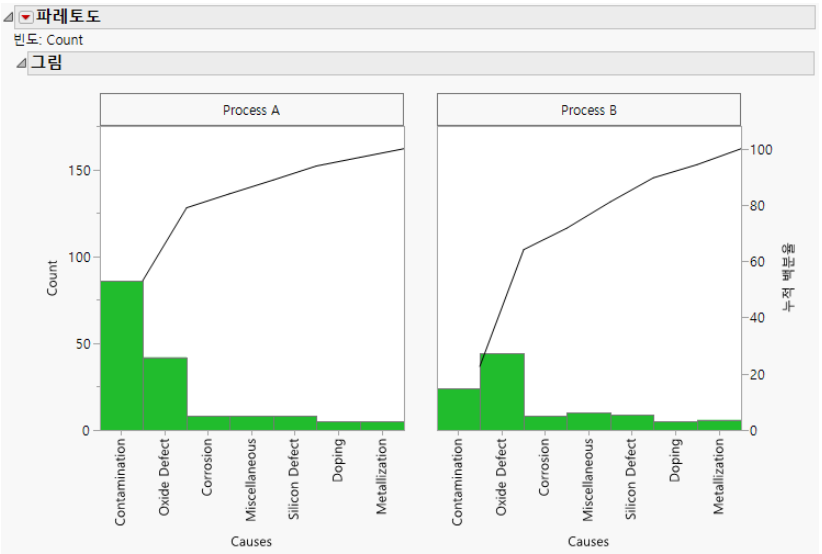
그룹 (Process A, B) 간 Contamination 의 DPU 는 약 0.06 입니다 .

그룹 전체에서 상수가 아닌 표본 크기 사용 예제

이 예에서는 고장 데이터와 빈도 열이 포함된 Failuressize.jmp 샘플 데이터 테이블을 사용합니다 . 여기에는 두 공정에 대해 집적 회로 제작 중의 고장 원인과 각 결함 유형별 발생 횟수가 나열되어 있습니다 . 다른 원인 (Oxide Defect, Silicon Defect 등) 중에 size 라벨이 지정된 원인이 있습니다 . 크기를 원인 코드로 지정하면 행이 크기 행으로 지정됩니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Failuressize.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 파레토도**를 선택합니다 .
3. Causes 를 선택하고 **Y, 원인**을 클릭합니다 .
4. Process 를 선택하고 **X, 그룹화**를 클릭합니다 .
5. Count 를 선택하고 **빈도**를 클릭합니다 .
6. **단위별 분석**을 선택한 후 **빈도 열의 값**을 선택합니다 .
7. **원인 코드**에 size 를 입력합니다 .
8. **확인**을 클릭합니다 .

그림 13.12 파레토도 보고서 창



9. " 파레토도 " 의 빨간색 삼각형을 클릭하고 **가산 분석** > **단위별 비율** 및 **가산 분석** > **그룹 간 비율 검정**을 선택합니다 .

그림 13.13 단위별 비율 및 그룹 간 비율 검정 결과

단위 비율당

Process	원인	개수	DPU	PPM	95% 하한	95% 상한
Process A	Contamination	86	0.8515	851485.15	0.6811	1.0516
	Oxide Defect	42	0.4158	415841.58	0.2997	0.5621
	Corrosion	8	0.0792	79207.92	0.0342	0.1561
	Miscellaneous	8	0.0792	79207.92	0.0342	0.1561
	Silicon Defect	8	0.0792	79207.92	0.0342	0.1561
	Doping	5	0.0495	49504.95	0.0161	0.1155
	Metallization	5	0.0495	49504.95	0.0161	0.1155
	합동 합계	162	0.2291	229137.20	0.1952	0.2673
	size	101				
Process B	Contamination	24	0.1655	165517.24	0.1061	0.2463
	Oxide Defect	44	0.3034	303448.28	0.2205	0.4074
	Corrosion	8	0.0552	55172.41	0.0238	0.1087
	Miscellaneous	10	0.0690	68965.52	0.0331	0.1268
	Silicon Defect	9	0.0621	62068.97	0.0284	0.1178
	Doping	5	0.0345	34482.76	0.0112	0.0805
	Metallization	6	0.0414	41379.31	0.0152	0.0901
	합동 합계	106	0.1044	104433.50	0.0855	0.1263
	size	145				

그룹 간 비율 검정

그룹 간 비율 검정: Process

원인	DPU 차이	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8	표준 오차	카이제곱	DF	Prob>ChiSq
Contamination	0.6860		0.0978	63.0776	1	<.0001*
Oxide Defect	0.1124		0.0788	2.1195	1	0.1454
Corrosion	0.0240		0.0341	0.5202	1	0.4707
Miscellaneous	0.0102		0.0355	0.0847	1	0.7710
Silicon Defect	0.0171		0.0348	0.2500	1	0.6171
Doping	0.0150		0.0270	0.3251	1	0.5685
Metallization	0.0081		0.0278	0.0871	1	0.7679
합동 합계	0.1247		0.0207	40.7524	1	<.0001*

그룹 A 의 원인에 대한 DPU 를 계산할 때 표본 크기 101 이 사용되었지만 그룹 B 의 원인에 대한 DPU 를 계산할 때는 표본 크기 145 가 사용되었습니다 .

두 개의 그룹 변수 (예 : Day, Process) 가 있는 경우 " 단위별 비율 " 에는 각 원인에 대한 Day 와 Process 의 모든 조합을 대상으로 DPU 또는 비율이 나열됩니다 . 그러나 " 그룹 간 비율 검정 " 에서는 그룹 간의 전체적인 차이만 검정합니다 .

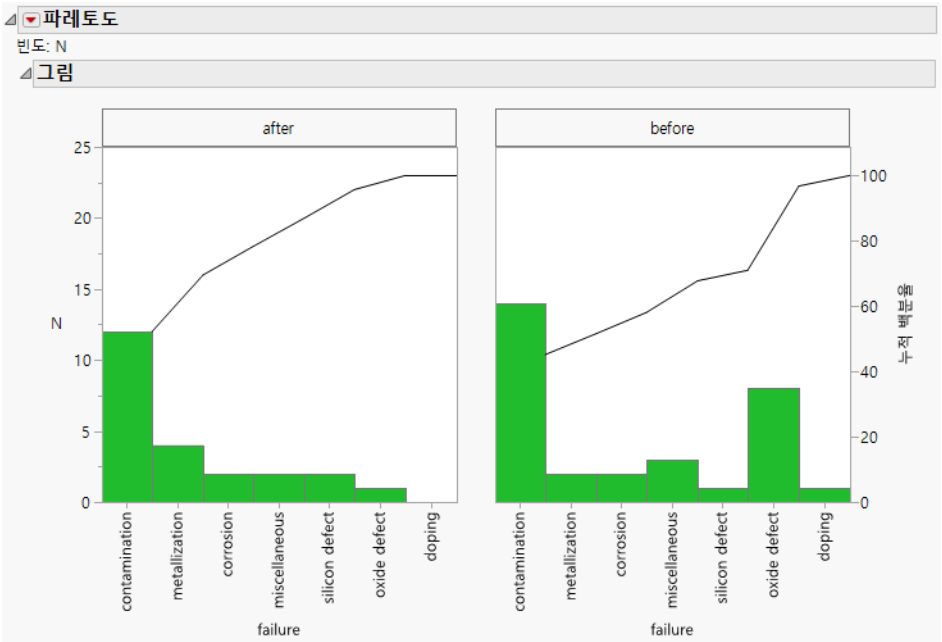
일원 비교 파레토도 예제

이 예에서는 Failure2.jmp 샘플 데이터 테이블을 사용합니다 . 이 테이블에는 확산로에서 튜브를 세척하기 전에 제조된 콘덴서 표본과 확산로를 세척한 후에 제조된 표본의 고장이 기록되어 있습니다 . 각 고장 유형에 대해 clean 변수는 "before" 또는 "after" 값으로 표본을 식별합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Failure2.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 파레토도**를 선택합니다 .
3. failure 를 선택하고 **Y, 원인**을 클릭합니다 .
4. clean 을 선택하고 **X, 그룹화**를 클릭합니다 .
5. N 을 선택하고 **빈도**를 클릭합니다 .
6. **확인**을 클릭합니다 .

   그림 13.14 에는 clean 변수의 각 값에 대한 그림이 나란히 표시되어 있습니다 .

그림 13.14 일원 비교 파레토도

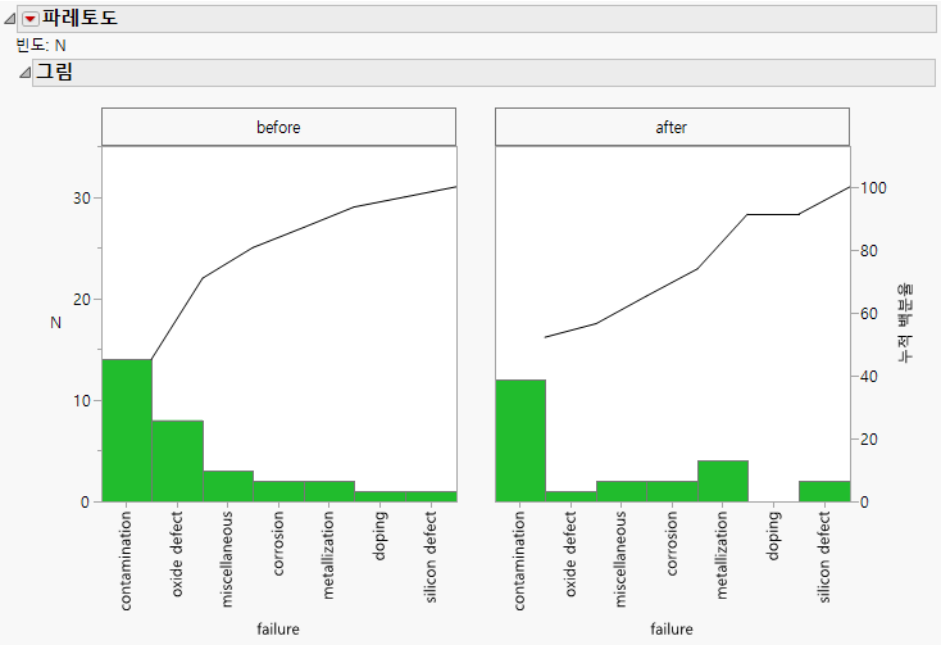


두 그림의 가로 축과 세로 축은 동일하게 척도화됩니다 . 첫 번째 그림의 막대는 *y* 축 값에 따라 내림차순으로 정렬되며 모든 셀의 순서를 결정합니다 .

7. 첫 번째 타일에서 제목 (after) 을 클릭한 후 다음 타일의 제목 (before) 으로 드래그하여 그림 순서를 재정렬합니다 .

   이 두 그림을 비교하면 세척 후 oxide defect(산화물 결함) 가 감소했음을 알 수 있습니다 . 그러나 그림 13.15 에 표시된 before-after 그림으로 나타내면 해석하기가 더 쉽습니다 . 첫 번째 셀을 기준으로 하는 순서를 반영하도록 원인 순서가 변경됩니다 .

그림 13.15 셀이 재정렬된 일원 비교 파레토도



이원 비교 파레토도 예제

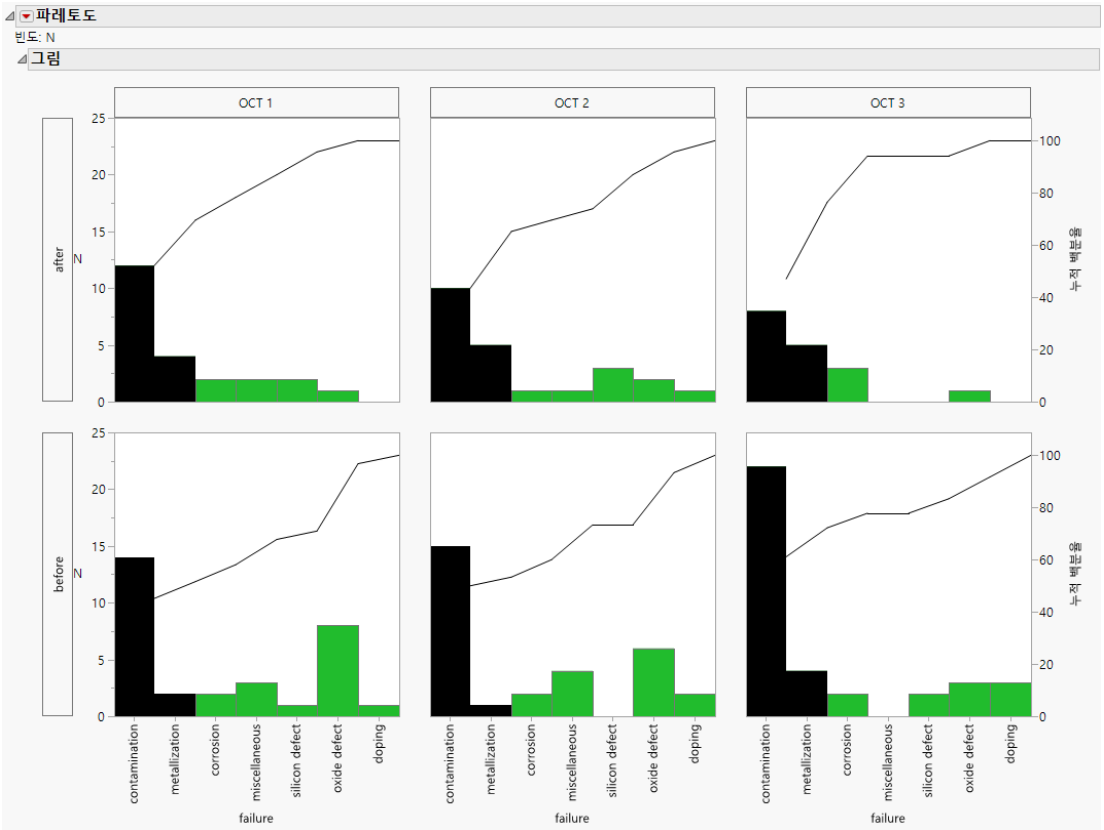
이 예에서는 Failure3.jmp 샘플 데이터 테이블을 사용합니다 . 이 데이터는 3 일 동안 콘덴서 제조 공정의 용광로 세척 전 / 후 생산 표본을 모니터링합니다 . 데이터 테이블에는 OCT 1, OCT 2 및 OCT 3 값이 포함된 date 열이 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Quality Control/Failure3.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 파레토도**를 선택합니다 .
3. failure 를 선택하고 **Y, 원인**을 클릭합니다 .
4. clean 과 date 를 선택하고 **X, 그룹화**를 클릭합니다 .
5. N 을 선택하고 **빈도**를 클릭합니다 .
6. **확인**을 클릭합니다 .

그림 13.16 에는 두 X 변수의 각 수준을 보여 주는 이원 레이아웃 그림이 포함된 파레토도가 표시되어 있습니다 . 왼쪽 위의 셀을 **키 셀**이라고 합니다 . 이 셀의 막대는 내림차순으로 정렬됩니다 . 다른 셀의 막대는 키 셀과 동일한 순서로 나타납니다 .

7. 키 셀에서 Contamination 과 Metallization 을 클릭하면 다른 모든 셀에서 해당 범주에 대한 막대가 강조 표시됩니다 .

그림 13.16 이원 비교 파레토도



파레토도에서는 **근본 원인**을 강조 표시하여 보여 줍니다. 이원 비교 그림의 각 셀에서 가장 자주 발생하는 두 가지 문제를 나타내는 막대가 선택됩니다. Contamination 과 Metallization 은 모든 셀에서 중요한 두 범주입니다. 용광로 세척 후에는 Contamination 이 문제가 되지 않습니다.

파레토도 플랫폼에 대한 통계 상세 정보

가능도비 카이제곱 검정

표기

파레토도 플랫폼에서 계산된 가능도비 카이제곱 검정 통계량에서는 다음과 같은 표기를 사용합니다 .

- n_{ij} = 그룹 j 에서 원인 i 의 개수
- E_j = 그룹 j 에 대한 기대 개수 . 이 값은 원인 전체에서 각 그룹의 평균 개수입니다 .
- E_i = 원인 i 에 대한 기대 개수 . 이 값은 그룹 전체에서 각 원인의 평균 개수입니다 .

그룹 내의 가능도비 카이제곱 검정 통계량

$$G_j^2 = 2 \sum_{i=1}^{K} n_{ij} \ln(n_{ij}/E_j)$$

그룹 간의 가능도비 카이제곱 검정 통계량

$$G_i^2 = 2 \sum_{j=1}^{J} n_{ij} \ln(n_{ij}/E_i)$$

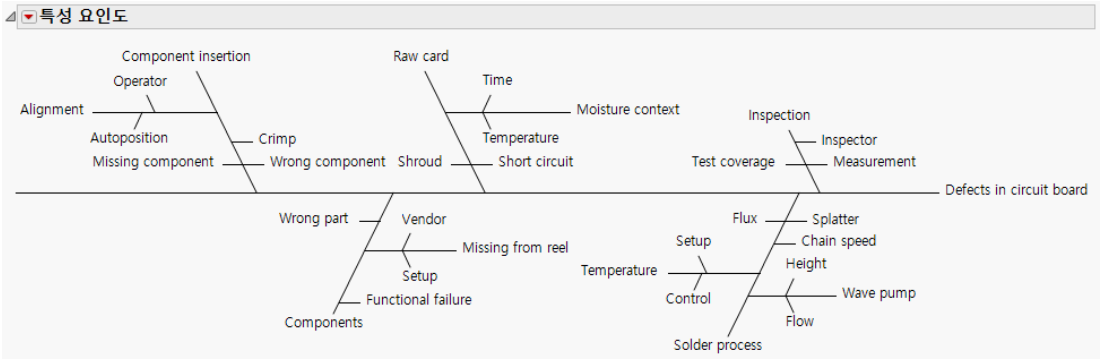
14 장

특성 요인도
근본 원인 식별

다이어그램 플랫폼을 사용하여 Ishikawa 차트 또는 Fishbone 차트라고도 하는 특성 요인도를 생성합니다. 이러한 다이어그램을 사용하여 다음을 수행할 수 있습니다.

- 효과의 원인 구성 (문제의 원인)
- 브레인스토밍
- 추가 실험을 위한 변수 식별

그림 14.1 특성 요인도의 예



목차

특성 요인도 개요 369
특성 요인도의 예 369
데이터 준비 370
다이어그램 플랫폼 시작 370
특성 요인도 371
마우스 오른쪽 버튼 클릭 메뉴 371
특성 요인도 메뉴 옵션 374
다이어그램 저장 374
다이어그램을 데이터 테이블로 저장 374
다이어그램을 저널로 저장 375
다이어그램을 스크립트로 저장 375

특성 요인도 개요

다이어그램 플랫폼을 사용하여 Ishikawa 차트 또는 Fishbone 차트라고도 하는 특성 요인도를 생성합니다. 이러한 다이어그램을 사용하여 다음을 수행할 수 있습니다.

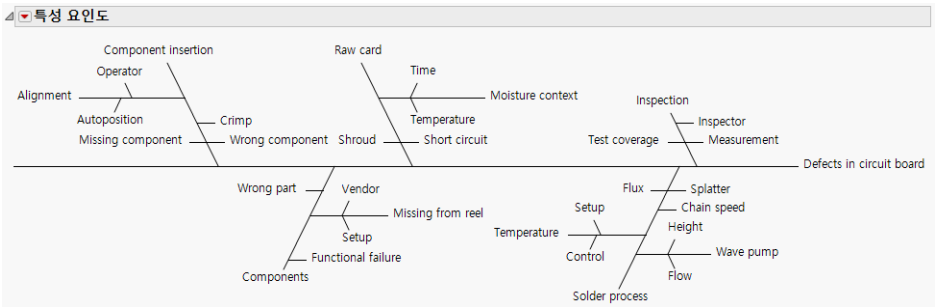
- 효과의 원인 구성 (문제의 원인)
- 브레인스토밍
- 추가 실험을 위한 변수 식별

특성 요인도의 예

회로기판의 결함 데이터가 있습니다. 다이어그램에서 결함의 주 요인 및 가능한 원인을 검토하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ishikawa.jmp 를 엽니다.
2. **분석 > 품질 및 공정 > 다이어그램**을 선택합니다.
3. Parent 를 선택하고 **X, 상위**를 클릭합니다.
4. Child 를 선택하고 **Y, 하위**를 클릭합니다.
5. **확인**을 클릭합니다.

그림 14.2 Ishikawa.jmp 다이어그램



주 요인은 Inspection, Solder process, Raw card, Components 및 Component insertion 입니다. Inspection 요인에 대한 Inspector, Measurement 및 Test coverage 와 같이 각 주 요인에서 가능한 원인이 갈라집니다.

한 번에 하나의 영역에 집중하여 각 주 요인에 대해 가능한 변동 원인 또는 소스를 자세히 검토할 수 있습니다.

데이터 준비

다이어그램을 생성하기 전에 데이터 테이블의 두 열에 있는 데이터로 시작합니다.

그림 14.3 Ishikawa.jmp 데이터 테이블의 예

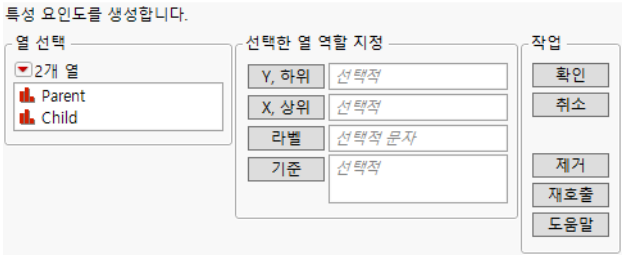
	Parent	Child
1	Defects in circuit board	Inspection
2	Defects in circuit board	Solder process
3	Defects in circuit board	Raw card
4	Defects in circuit board	Components
5	Defects in circuit board	Component insertion
6	Inspection	Measurement
7	Inspection	Test coverage
8	Inspection	Inspector
9	Solder process	Splatter
10	Solder process	Flux
11	Solder process	Chain speed
12	Solder process	Temperature
13	Solder process	Wave pump
14	Temperature	Setup

Parent 값 "Defects in circuit board"(효과)에는 Child 열에 나열된 다섯 가지 주 요인이 있습니다. 이러한 주 요인 중 하나는 "Inspection"이며 이 요인의 원인은 Child 열에 나열되어 있습니다. 상위 값은 하위 항목을 포함하고, 하위 항목도 자신만의 하위 항목을 포함하므로 Parent 열과 Child 열에 모두 나열됩니다.

다이어그램 플랫폼 시작

분석 > 품질 및 공정 > 다이어그램을 선택하여 다이어그램 플랫폼을 시작할 수 있습니다.

그림 14.4 다이어그램 시작 창



"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

팁 : 데이터 테이블을 기반으로 하지 않는 기본 다이어그램을 생성하려면 **Y, 하위** 및 **X, 상위** 필드를 비워 두고 **확인**을 클릭합니다. 그런 다음 마우스 오른쪽 버튼 클릭 메뉴의 옵션을 사용하여 노드를 편집합니다. 자세한 내용은 " 마우스 오른쪽 버튼 클릭 메뉴 " 에서 확인하십시오.

Y, 하위 상위 요인에 기여하는 하위 요인을 나타냅니다.

X, 상위 하위 요인이 있는 상위 요인 (효과 포함) 을 나타냅니다.

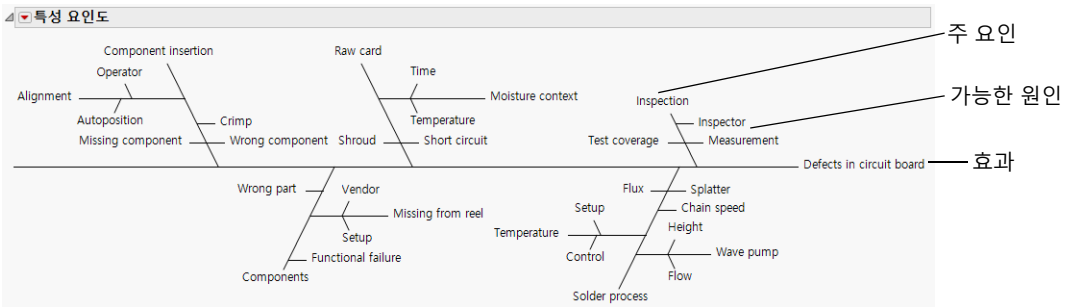
라벨 다이어그램 노드에 " 라벨 " 열의 텍스트를 포함합니다.

기준 기준 변수의 각 값에 대해 개별 다이어그램을 생성합니다.

특성 요인도

그림 14.5에서 효과 또는 문제(Defects in circuit board)가 오른쪽에 중심선으로 나타납니다. 영향을 주는 주 기여 요인은 분지의 끝에 나타납니다(Inspection, Solder process, Raw Card 등). 각 주 요인에서 가능한 원인이 갈라집니다.

그림 14.5 특성 요인도



마우스 오른쪽 버튼 클릭 메뉴

강조 표시된 노드를 마우스 오른쪽 버튼으로 클릭하여 텍스트 수정, 새 노드 삽입, 다이어그램 유형 변경 등을 수행할 수 있습니다. 다음 사항에 유의하십시오.

- 제목을 마우스 오른쪽 버튼으로 클릭하여 글꼴 및 색상, 위치, 표시 여부 또는 형식을 변경합니다.
- 노드를 클릭하고 강조 표시하여 이름을 바꿉니다.
- 노드를 클릭하고 드래그하여 이동합니다.

텍스트 메뉴

" 텍스트 " 메뉴에는 다음 옵션이 포함되어 있습니다.

글꼴 텍스트 또는 숫자 문자의 글꼴을 선택합니다.

색상 텍스트 또는 숫자 문자의 색상을 선택합니다 .

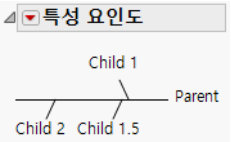
왼쪽으로 회전 , 오른쪽으로 회전 , 수평 텍스트나 숫자를 수평으로 , 왼쪽으로 90 도 또는 오른쪽으로 90 도 회전합니다 .

삽입 메뉴

삽입 메뉴를 사용하여 기존 노드에 항목을 삽입합니다 . " 삽입 " 메뉴에는 다음 옵션이 포함되어 있습니다 .

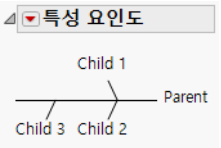
앞 강조 표시된 노드의 오른쪽에 새 노드를 삽입합니다 . 예를 들어 그림 14.6 에서는 Child 1.5 를 Child 2 앞에 삽입합니다 .

그림 14.6 앞에 삽입



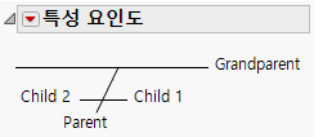
뒤 강조 표시된 노드의 왼쪽에 새 노드를 삽입합니다 . 예를 들어 그림 14.7 에서는 Child 3 을 Child 2 뒤에 삽입합니다 .

그림 14.7 뒤에 삽입



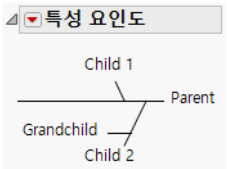
위 현재 노드 위의 수준에 새 노드를 삽입합니다 . 예를 들어 그림 14.8 에서는 Grandparent 를 Parent 위의 수준에 삽입합니다 .

그림 14.8 위에 삽입



아래 현재 노드 아래의 수준에 새 노드를 삽입합니다 . 예를 들어 그림 14.9 에서는 Grandchild 를 Child 2 아래의 수준에 삽입합니다 .

그림 14.9 아래에 삽입



이동 메뉴

" 이동 " 메뉴를 사용하여 노드 또는 분지를 이동합니다 . " 이동 " 메뉴에는 다음 옵션이 포함되어 있습니다 .

첫 번째 강조 표시된 노드를 상위 항목 아래의 첫 번째 위치로 이동합니다 .

마지막 강조 표시된 노드를 상위 항목 아래의 마지막 위치로 이동합니다 .

반대쪽 강조 표시된 노드를 상위 항목 선의 반대쪽으로 이동합니다 .

왼쪽으로 가로로 그려진 모든 요소를 상위 항목의 왼쪽에 표시합니다 .

오른쪽으로 가로로 그려진 모든 요소를 상위 항목의 오른쪽에 표시합니다 .

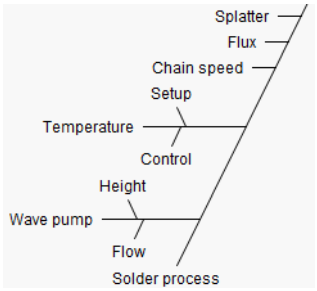
위로 세로로 그려진 모든 요소를 상위 항목의 위에 표시합니다 .

아래로 세로로 그려진 모든 요소를 상위 항목의 아래에 표시합니다 .

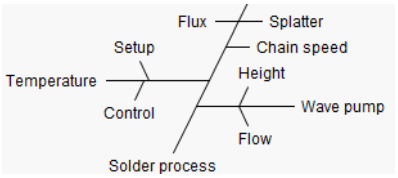
교대로 하위 항목을 상위 항목 선의 좌우에 교대로 그립니다 .

그림 14.10 강제 적용 옵션

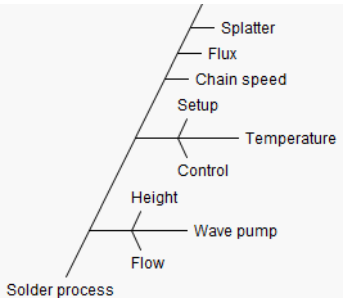
왼쪽으로



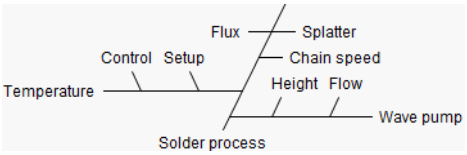
교대로



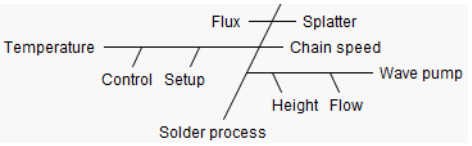
오른쪽으로



위로



아래로



기타 메뉴 옵션

강조 표시된 노드의 마우스 오른쪽 버튼 클릭 메뉴에는 다음 옵션도 포함되어 있습니다 .

유형 변경 전체 차트 유형을 **Fishbone**, **계층** 또는 **내포됨**으로 변경합니다 .

편집할 수 없음 **이동** 및 **유형 변경**을 제외한 다른 모든 명령을 비활성화합니다 .

텍스트 줄바꿈 너비 텍스트 줄바꿈이 발생하는 라벨 너비를 지정합니다 .

데이터 테이블로 만들기 현재 강조 표시된 노드를 데이터 테이블로 변환합니다 . 전체 다이어그램 (효과) 을 강조 표시하면 모든 노드가 변환됩니다 .

닫기 강조 표시된 노드를 숨깁니다 .

삭제 강조 표시된 노드와 모든 하위 항목을 삭제합니다 .

특성 요인도 메뉴 옵션

" 특성 요인도 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

다이어그램 저장

다이어그램을 저장할 수 있는 여러 가지 방법이 있습니다 . 다음 중 하나를 선택합니다 .

- 다이어그램을 데이터 테이블로 저장
- 다이어그램을 저널로 저장
- 다이어그램을 스크립트로 저장

다이어그램을 데이터 테이블로 저장

이 방법의 경우 다음 사항에 유의하십시오 .

- 데이터 테이블을 업데이트해야 하는 다른 공정이 있는 경우 이 방법을 선택하는 것이 좋습니다 .
- 데이터 테이블에서는 사용자 정의를 나타낼 수 없으므로 사용자 정의가 거의 불가능합니다 .

다이어그램을 데이터 테이블로 저장하려면 다음을 수행하십시오 .

1. 전체 다이어그램을 강조 표시합니다 .
2. 마우스 오른쪽 버튼을 클릭하고 **데이터 테이블로 만들기**를 선택합니다 .
3. 새 데이터 테이블을 저장합니다 .

다이어그램을 저널로 저장

이 방법의 경우 다음 사항에 유의하십시오 .

- 이 옵션은 즉석 작업에 적합합니다. 예를 들어 수동으로 다이어그램을 작성하고 저널로 저장한 후 나중에 저널을 다시 열어 다이어그램 작성 및 편집을 계속할 수 있습니다 .
- 사용자 정의는 저널에만 있으며 저널은 데이터 테이블에 연결되지 않습니다 .

다이어그램을 저널로 저장하려면 다음을 수행하십시오 .

1. 전체 다이어그램을 강조 표시합니다 .
2. 마우스 오른쪽 버튼을 클릭하고 **편집 > 저널**을 선택합니다 .
3. 새 저널을 저장합니다 .

다이어그램을 스크립트로 저장

이 방법의 경우 다음 사항에 유의하십시오 .

- 데이터 테이블을 업데이트해야 하는 다른 공정이 있는 경우 이 방법을 선택하는 것이 좋습니다.
- 데이터 테이블에서 다이어그램을 생성한 경우 사용자 정의 없이 해당 데이터 테이블에 대해 다시 실행되는 간단한 스크립트가 나타납니다 .
- 데이터 테이블을 사용하지 않거나 저널에서 다이어그램을 생성한 경우 다이어그램의 각 영역을 추가하고 사용자 정의하는 데 필요한 모든 명령이 포함된 좀 더 복잡한 스크립트가 나타납니다 .

다이어그램을 스크립트로 저장하려면 다음을 수행하십시오 .

1. "특성 요인도" 옆의 빨간색 삼각형을 클릭하고 **스크립트 저장 > 스크립트 창에**를 선택합니다.
2. 새 스크립트를 저장합니다 .

15 장

품질 유틸리티
규격 한계 관리 및 OC 곡선 생성

이 장에서는 분석 > 품질 및 공정 메뉴의 유틸리티에 대한 내용을 다룹니다. 특히 규격 한계 관리 및 OC 곡선 유틸리티에 대해 설명합니다.

목차

규격 한계 관리 유틸리티 379
규격 한계 관리 유틸리티의 예 379
규격 한계 관리 옵션 382
검사 특성 곡선 유틸리티 383
OC 곡선 유틸리티 시작 383
관리도에 대한 OC 곡선 옵션 383
합격 표집에 대한 OC 곡선 옵션 384

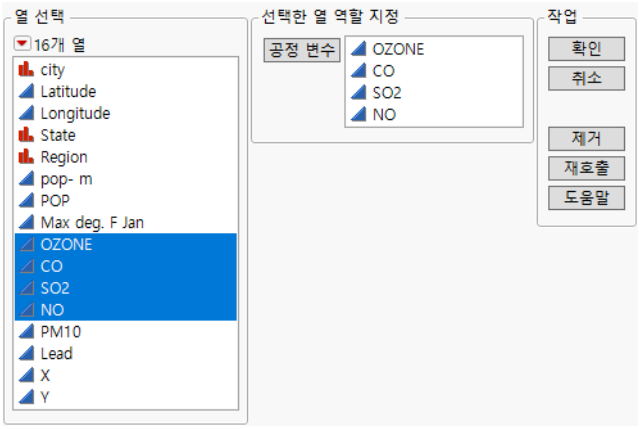
규격 한계 관리 유틸리티

규격 한계 관리 유틸리티를 사용하면 여러 열에 대한 많은 규격 한계를 한 번에 빠르게 추가하거나 편집할 수 있습니다. 그런 다음 향후 분석에 규격 한계가 사용됩니다. 또한 각 공정에 대한 중요도 값을 지정하고 한계를 그래프에 참조선으로 표시할지 여부를 나타낼 수도 있습니다.

규격 한계 관리 유틸리티의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cities.jmp 를 엽니다 .
2. **분석 > 품질 및 공정 > 규격 한계 관리**를 선택합니다 .
3. 규격 한계를 설정할 열을 지정합니다 . 이 예에서는 OZONE, CO, SO2 및 NO 를 선택하고 **공정 변수**를 클릭합니다 .

그림 15.1 열 지정



4. **확인**을 클릭합니다 .
5. 규격 한계를 추가합니다 . JMP 데이터 테이블에서 기존 한계를 로드하거나 (**한계 테이블에서 로드**) 한계를 수동으로 입력할 수 있습니다 . 이 예에서는 다음 한계를 수동으로 입력합니다 .
   - OZONE: LSL 0.12, USL 0.2
   - CO: LSL 6, USL 12
   - SO2: LSL 0.015, USL 0.06
   - NO: LSL 0.02, USL 0.04
6. " 규격 한계 관리 " 옆의 빨간색 삼각형을 클릭하고 **한계 모두 표시**를 선택합니다 .

   모든 열에 대한 규격 한계가 향후 분석의 그래프에 나타납니다 . 개별 열에 대해서만 규격 한계를 표시하려면 해당 열 옆의 **한계 표시** 상자를 선택합니다 .

그림 15.2 규격 한계 설정

7. 규격 한계를 저장할 방법을 선택합니다 . 이 예에서는 **열 특성에 저장**을 클릭합니다 . 이렇게 하면 규격 한계가 해당 데이터 테이블에 열 특성으로 저장됩니다 . 세로형 또는 가로형 형식의 새 데이터 테이블에 저장할 수도 있습니다 .

   Cities.jmp 데이터 테이블의 " 열 " 패널에서 OZONE, CO, SO2 및 NO 옆에 " 규격 한계 " 열 특성을 나타내는 별표가 표시됩니다 .

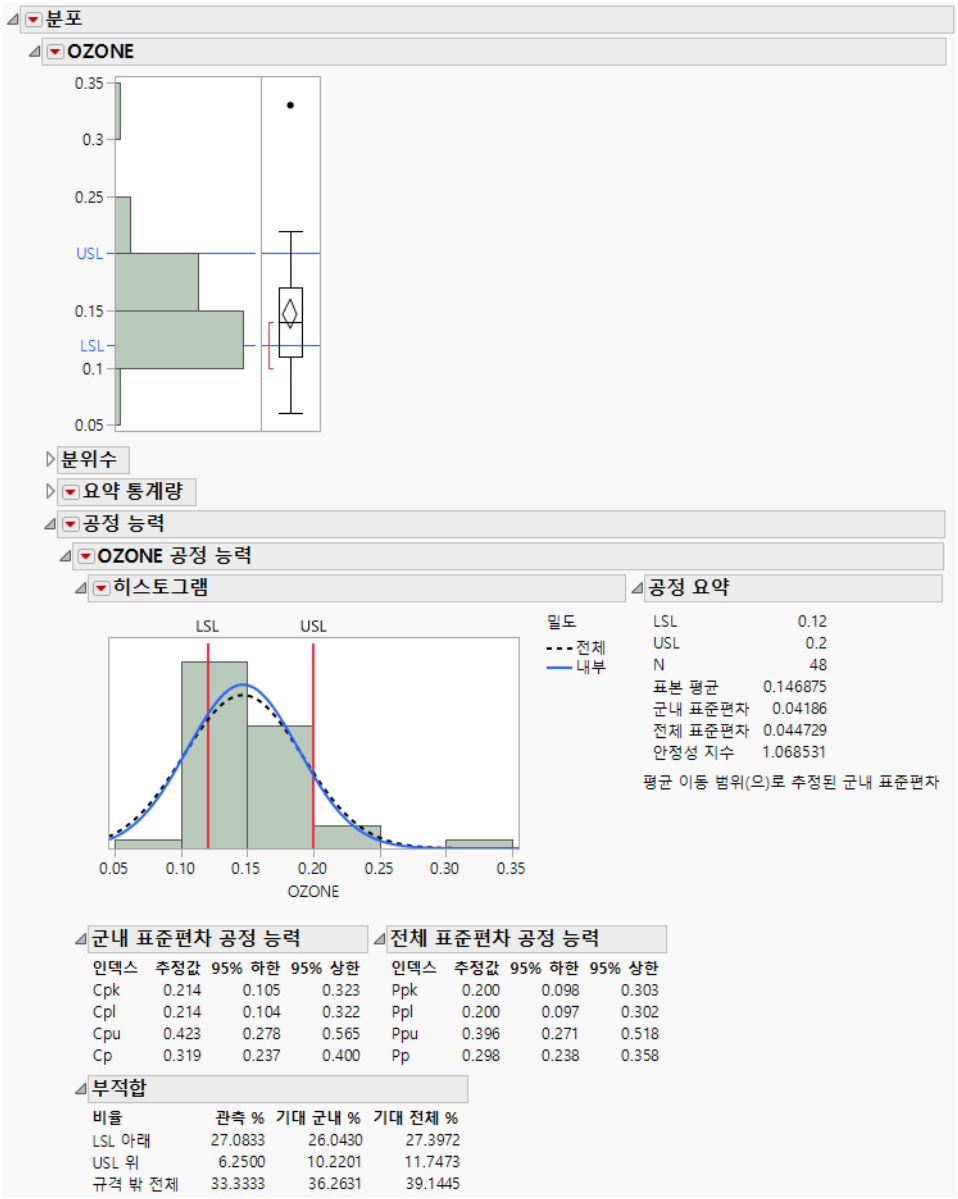
8. 데이터 테이블에서 한계를 벗어나는 값을 보려면 " 규격 한계 관리 " 옆의 빨간색 삼각형을 클릭하고 **규격 이탈 값에 색상 적용**을 선택합니다 . Cities.jmp 데이터 테이블로 이동하면 한계를 벗어나는 값에 색상이 지정되어 있음을 확인할 수 있습니다 .

9. 이제 분석을 실행할 수 있습니다 . 이 예에서는 **분석 > 분포**를 선택합니다 .

10. OZONE, CO, SO2 및 NO 를 선택하고 **Y, 열**을 클릭합니다 .

11. **확인**을 클릭합니다 .

그림 15.3 분포에서 OZONE 에 대한 규격 한계



OZONE 열에 추가한 규격 한계가 히스토그램에 나타납니다. 열에 "규격 한계" 열 특성이 포함되어 있으므로 "분포" 보고서에 공정 능력 분석 보고서도 포함됩니다.

규격 한계 관리 옵션

규격 한계를 설정하는 창에는 규격 한계를 저장 및 로드하는 버튼과 " 규격 한계 관리 " 의 빨간색 삼각형 메뉴 옵션이 있습니다 .

보고서 테이블 열

" 규격 한계 관리 " 보고서 맨 위의 테이블에는 시작 창에 지정된 각 공정 열에 대한 행이 포함되어 있습니다 . 이 테이블을 사용하여 각 열에 대해 규격 한계와 다음 옵션을 지정할 수 있습니다 .

한계 표시 지정된 열에 대한 " 규격 한계 " 열 특성에서 " 그래프 참조선으로 표시 " 옵션이 선택되도록 지정합니다 .

공정 중요도 각 열에 대한 공정 중요도 값을 지정합니다 . 공정 중요도 값은 원하는 순서대로 공정을 정렬하는 메커니즘을 제공합니다 . 공정 중요도 값은 많은 플랫폼 그래프에서 표식 크기를 지정하는 데 사용됩니다 .

단위 각 열의 단위를 지정합니다 .

버튼

한계 테이블에서 로드 JMP 데이터 테이블에서 규격 한계를 로드합니다 .

열 특성에 저장 규격 한계를 연결된 데이터 테이블에 열 특성으로 저장합니다 .

세로형 규격 한계 테이블에 저장 규격 한계를 세로형 형식의 새 데이터 테이블에 저장합니다 .

가로형 규격 한계 테이블에 저장 규격 한계를 가로형 형식의 새 데이터 테이블에 저장합니다 .

빨간색 삼각형 옵션

한계 모두 표시 모든 열의 " 한계 표시 " 아래에 있는 상자를 선택합니다 . 열에 대해 " 한계 표시 " 를 선택한 경우 " 규격 한계 " 열 특성에서 " 그래프 참조선으로 표시 " 옵션이 선택됩니다 . " 그래프 참조선으로 표시 " 옵션은 지정된 규격 한계와 목표값을 선택한 분석 그림에 참조선으로 표시합니다 .

참고 : " 한계 표시 " 아래의 모든 상자가 선택된 경우 " 한계 모두 표시 " 옵션을 선택하면 " 한계 표시 " 아래의 모든 상자가 선택 취소됩니다 .

소수점 근사값 규격 한계를 반올림할 소수 자릿수를 설정합니다 .

규격 이탈 값에 색상 적용 데이터 테이블에서 열에 대한 규격 한계를 벗어나는 값에 색상을 적용합니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

검사 특성 곡선 유틸리티

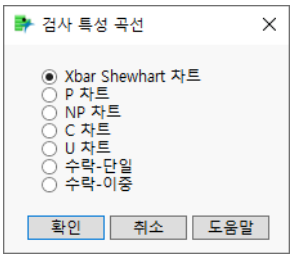
OC(검사 특성) 곡선 유틸리티를 사용하면 관리도 및 계수형 합격 표집 계획에 대한 OC 곡선을 생성할 수 있습니다 . OC 곡선은 Xbar, P, NP, C 및 U 차트에 사용할 수 있습니다 . 지정된 관리도에 대해 OC 곡선은 특정 크기의 변화를 감지하지 못할 확률을 보여 줍니다. 또한 단일 및 이중 계수형 합격 표집 계획을 위한 OC 곡선이 있습니다 . 지정된 합격 표집 계획에 대해 OC 곡선은 로트 품질에 따른 로트 합격 확률의 변화를 보여 줍니다 .

참고 : 관리도에 대한 OC 곡선은 양측 곡선입니다 . 음의 변화와 양의 변화에 대한 곡선을 그립니다 . 대개 OC 곡선은 절대 변화에 대해 하나의 곡선을 표시합니다 .

OC 곡선 유틸리티 시작

분석 > 품질 및 공정 > OC 곡선 을 선택하여 검사 특성 곡선을 시작할 수 있습니다 . 관심 OC 곡선을 선택하고 " 확인 " 을 클릭하여 시작합니다 .

그림 15.4 OC 곡선 시작 창



관리도에 대한 OC 곡선 옵션

관리도에서는 OC 곡선을 사용하여 차트 모수 설정이 β, 즉 지정된 관심 변화를 감지하지 못할 확률에 미치는 영향을 탐색할 수 있습니다 . 텍스트 상자 , 슬라이더 또는 OC 곡선의 LCL, 변화 및 UCL 핸들을 사용하여 차트 모수를 설정하고 조정합니다 . 차트 모수는 다음과 같습니다 .

관리 하한 관리도의 관리 하한을 지정합니다 .

관리 상한 관리도의 관리 상한을 지정합니다 .

표본 크기 (C 차트에는 사용 불가능) 관리도 측도에 사용되는 표본 크기를 지정합니다 .

시그마 (XBar 차트에만 사용 가능) 관리도 시그마를 지정합니다 .

변화 감지할 변화를 지정합니다 .

베타 관리도 규격이 주어진 경우 지정된 변화를 감지하지 못할 확률을 지정합니다 . 규격이 변경되면 베타가 업데이트됩니다 .

합격 표집에 대한 OC 곡선 옵션

계수형 합격 표집의 경우 OC 곡선을 사용하여 표집 계획 및 가정된 제품 품질이 로트 합격 확률에 미치는 영향을 탐색할 수 있습니다. 텍스트 상자, 슬라이더 또는 OC 곡선의 불량률 핸들을 사용하여 표집 계획을 설정하고 조정합니다 .

단일 표집 OC 곡선 옵션

표집 유형 " 로트 표집 " 또는 " 이항 표집 " 을 선택할 수 있습니다 .

로트 표집 고정된 로트 크기를 기반으로 합격 계획을 지정하고 탐색할 수 있습니다 .

이항 표집 연속형 공정이거나 이항 분포가 적절한 기타 상황에 대한 합격 계획을 지정하고 탐색할 수 있습니다 .

로트 크기 (N) (로트 표집에만 사용 가능) 표집할 로트 크기를 지정합니다 .

표본 크기 (n) 검사할 단위 수를 지정합니다 .

허용 가능한 고장 수 (c1) 허용 가능한 고장 수를 지정합니다. 관측된 결함 수가 c1 보다 크면 로트가 기각됩니다 .

불량률 로트의 예상 불량률을 지정합니다 .

합격률 정의된 표집 계획에 따라 로트를 수락할 확률을 지정합니다 . 규격이 변경되면 합격률이 업데이트되지만 이 값을 직접 조정할 수도 있습니다 . 합격률을 직접 조정하면 불량률 값이 업데이트됩니다 .

이중 표집 OC 곡선 옵션

첫 번째 표본 첫 번째 표본에 대한 다음 설정을 포함합니다 .

검사 개수 (n1) 첫 번째 표본의 검사 단위 수를 지정합니다 .

허용 가능한 고장 수 (c1) 첫 번째 표본의 허용 가능한 고장 수를 지정합니다 .

두 번째 표본 두 번째 표본에 대한 다음 설정을 포함합니다 .

검사 개수 (n2) 두 번째 표본의 검사 단위 수를 지정합니다 .

허용 가능한 총 고장 수 (c1+c2) 허용 가능한 총 고장 수를 지정합니다 .

불량률 로트의 예상 불량률을 지정합니다 .

합격률 정의된 표집 계획에 따라 로트를 수락할 확률을 지정합니다 . 규격이 변경되면 합격률이 업데이트되지만 이 값을 직접 조정할 수도 있습니다 . 합격률을 직접 조정하면 불량률 값이 업데이트됩니다 .

A 부록

참조 자료

Agresti, A., and Coull, B. (1998). "Approximate Is Better than 'Exact' for Interval Estimation of Binomial Proportions." *The American Statistician* 52:119–126.

American Society for Quality Statistics Division (2004). *Glossary and Tables for Statistical Quality Control*. 4th ed. Milwaukee, WI: ASQC Quality Press.

American Society for Testing and Materials (1976). *ASTM Manual on Presentation of Data and Control Chart Analysis*. Philadelphia: ASTM.

Automotive Industry Action Group (2002). *Measurement Systems Analysis Reference Manual*. 3rd ed. Troy, MI: Automotive Industry Action Group.

AIAG (2005). *Statistical Process Control*. 2nd ed. Troy, MI: Automotive Industry Action Group.

Barrentine, L. B. (1991). *Concepts for R&R Studies*. Milwaukee, WI: ASQC Quality Press.

Bissell, A. F. (1990). "How Reliable Is Your Capability Index?" *Journal of the Royal Statistical Society, Series C* 39:331–340.

Box, G. E. P, Luceño, A., and Paniagua-Quinones, M. d. C. (2009). *Statistical Control by Monitoring and Adjustment*. 2nd ed. New York: John Wiley & Sons.

Burdick, P. K., Borror, C. M., and Montgomery, D. C. (2005). *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random Mixed ANOVA Models.* Philadelphia, PA and Alexandria, VA: SIAM and ASA.

Chou, Y.-M., Owen, D. B., and Borrego, S. A. (1990). "Lower confidence limits on process capability indices." *Journal of Quality Technology* 22:223–229.

Crowder, S. V. (1987). "Average Run Lengths of Exponentially Weighted Moving Average Charts." *Journal of Quality Technology* 29:401–408.

David, H. A. (1951). "Further Applications of Range to the Analysis of Variance." *Biometrika* 38:393–407.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons.

Goel, A. L., and Wu, S. M. (1971). "Determination of A.R.L. and a Contour Nomogram for Cusum Charts to Control Normal Mean." *Technometrics* 13:221–230.

Harter, H. L. (1960). "Tables of Range and Studentized Range." *Annals of Mathematical Statistics* 31:1122–1147.

Hoffman, D. (2003). "Negative Binomial Control Limits for Count Data with Extra-Poisson Variation." *Pharmaceutical Statistics* 2:127–132.

Jackson, J. E., and Mudholkar, G. S. (1979). "Quality Control Procedures for Residuals Associated with Principal Component Analysis." *Annals of Mathematical Statistics* 25:290–302.

Knoth, S. (2004). "Fast Initial Response Features for EWMA Control Charts." *Statistical Papers* 46:47–64.

Kourti, T., and MacGregor, J. F. (1996). "Multivariate SPC Methods for Process and Product Monitoring." *Journal of Quality Technology* 28:409–428.

Levey, S., and Jennings, E. R. (1950). "The Use of Control Charts in the Clinical Laboratory." *American Journal of Clinical Pathology* 20:1059–1066.

Li, G., Qin, S., Ji, Y., Zhou, D.(2009). "Total PLS Based Contribution Plots for Fault Diagnosis." *Acta Automatica Sinica* 35:759 - 765.

Lucas, J. M. (1976). "The Design and Use of V–Mask Control Schemes." *Journal of Quality Technology* 8:1–12.

Lucas, J. M., and Crosier, R. B. (1982). "Fast Initial Response for CUSUM Quality Control Schemes: Give Your CUSUM a Head Start." *Technometrics* 24:199–205.

Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control.* 7th ed. New York: John Wiley & Sons.

Nair, V. N. (1984). "Confidence Bands for Survival Functions with Censored Data: A Comparative Study." *Technometrics* 26:265–275.

Nelson, L. (1984). "The Shewhart Control Chart—Tests for Special Causes." *Journal of Quality Technology* 15:237–239.

Nelson, L. (1985). "Interpreting Shewhart X Control Charts." *Journal of Quality Technology* 17:114–116.

Nomikos, P., and MacGregor, J. F., (1995). "Multivariate SPC Charts for Monitoring Bach Processes." *Technometrics* 37:41–59.

Pearn, W. L., and Kotz, S. (2006). *Encyclopedia and Handbook of Process Capability Indices*. Vol. 12 of Series on Quality, Reliability, and Engineering Statistics. New Jersey: World Scientific.

Portnoy, S. (1971). "Formal Bayes Estimation with Application to a Random Effects Model." *The Annals of Mathematical Statistics* 42:1379–1402.

Sahai, H. (1974). "Some Formal Bayes Estimators of Variance Components in the Balanced Three-Stage Nested Random Effects Model." *Communication in Statistics – Simulation and Computation* 3:233–242.

Slifker, J. F., and Shapiro, S. S. (1980). "The Johnson System: Selection and Parameter Estimation." *Technometrics* 22:239–246.

Sullivan, J. H., and Woodall, W. H. (2000). "Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations." *IIE Transactions* 32:537–549.

Tracy, N. D., Young, J. C., and Mason, R. L. (1992). "Multivariate Control Charts for Individual Observations." *Journal of Quality Technology* 24:88–95.

Westgard, J. O. (2002). *Basic QC Practices.* 2nd ed. Madison, WI: Westgard QC.

Wheeler, D. J. (2004). *Advanced Topics in Statistical Process Control.* 2nd ed. Knoxville, TN: SPC Press.

Wheeler, D. J. (2006). *EMP III Using Imperfect Data*. Knoxville, TN: SPC Press.

Wludyka, P., and Sa, P. (2004). "A robust I-Sample analysis of means type randomization test for variances for unbalanced designs." *Journal of Statistical Computation and Simulation* 74:701–726.

Woodall, W. H., and Adams, B. M. (1998). *Handbook of Statistical Methods for Engineers and Scientists*. 2nd ed. Chapter 7. New York: McGraw-Hill.

B 부록

기술 라이선스 고지 사항

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

Version 16

Reliability and Survival Methods

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. *JMP® 16 Reliability and Survival Methods*. Cary, NC: SAS Institute Inc.

JMP® 16 Reliability and Survival Methods

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

August 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Get the Most from JMP

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

https://www.jmp.com/getstarted

Contents

Reliability and Survival Methods

1 Learn about JMP 13
Documentation and Additional Resources
Formatting Conventions in JMP Documentation 15
JMP Help 16
JMP Documentation Library 16
Additional Resources for Learning JMP 22
JMP Tutorials 22
Sample Data Tables 22
Learn about Statistical and JSL Terms 23
Learn JMP Tips and Tricks 23
JMP Tooltips 23
JMP User Community 24
Free Online Statistical Thinking Course 24
JMP New User Welcome Kit 24
Statistics Knowledge Portal 24
JMP Training 24
JMP Books by Users 25
The JMP Starter Window 25
JMP Technical Support 25

2 Introduction to Reliability and Survival 27
Lifetime and Failure Analysis

3 Life Distribution 29
Fit Distributions to Lifetime Data
Overview of the Life Distribution Platform 31
Example of the Life Distribution Platform 31
Launch the Life Distribution Platform 34
Life Distribution Report 37
Event Plot 38
Compare Distributions 41
Life Distribution: Statistics 45
Life Distribution Report Options 52

Mixture ... 54
Fit Competing Risk Mixture ... 57
Weibayes Report ... 58
Competing Cause Report ... 59
Competing Cause Workflow ... 60
Competing Cause Model ... 60
Cause Combination ... 61
Competing Cause: Statistics ... 62
Individual Causes ... 63
Competing Cause Report Options ... 63
Life Distribution - Compare Groups Report ... 65
Compare Groups: Statistics ... 66
Individual Group ... 67
Life Distribution - Compare Groups Report Options ... 67
Additional Examples of the Life Distribution Platform ... 68
Omit Competing Causes ... 69
Change the Scale ... 70
Examine the Same Distribution across Groups ... 72
Weibayes Estimates ... 74
Fit Mixture Example ... 76
Fit Competing Risk Mixture Example ... 79
Statistical Details for the Life Distribution Platform ... 81
Distributions ... 81
Competing Cause Details ... 96
Notes Regarding Median Rank Regression ... 101

4 Fit Life by X ... 103
Fit Single-Factor Models to Time-to Event Data

Overview of the Fit Life by X Platform ... 105
Example of the Fit Life by X Platform ... 106
Launch the Fit Life by X Platform ... 108
The Fit Life by X Report ... 110
Summary of Data ... 111
Scatterplot ... 111
Nonparametric Overlay ... 113
Comparisons ... 114
Results ... 118
Custom Relationship ... 129
Fit Life by X Platform Options ... 130
Additional Examples of the Fit Life by X Platform ... 131

Capacitor Example ... 131
Custom Relationship Example ... 133

5 Cumulative Damage ... 137
Model Product Deterioration with Variable Stress over Time

Overview of the Cumulative Damage Platform ... 139
Example of the Cumulative Damage Platform ... 139
Launch the Cumulative Damage Platform ... 143
Time to Event ... 144
Stress Pattern ... 145
The Cumulative Damage Report ... 148
Event Plot ... 148
Stress Patterns Report ... 148
Model List ... 149
Model Results ... 149
Cumulative Damage Platform Options ... 150
Stress Patterns Options ... 150
Additional Example of the Cumulative Damage Platform ... 152
Example of Simulating New Data ... 152

6 Recurrence Analysis ... 155
Model the Frequency or Cost of Recurrent Events over Time

Overview of the Recurrence Analysis Platform ... 157
Example of the Recurrence Analysis Platform ... 157
Launch the Recurrence Analysis Platform ... 160
Recurrence Analysis Platform Options ... 161
Fit Model ... 162
Additional Examples of the Recurrence Analysis Platform ... 166
Bladder Cancer Recurrences Example ... 167
Diesel Ship Engines Example ... 170

7 Degradation ... 175
Model Product Deterioration over Time

Overview of the Degradation Platform ... 177
Example of the Degradation Platform ... 177
Launch the Degradation Platform ... 179
The Degradation Reports ... 181
Model Specification ... 184
Simple Linear Path ... 184
Nonlinear Path ... 186
Inverse Prediction ... 194

Prediction Graph ... 196
Degradation Platform Options ... 197
Model Reports ... 199
Model Lists ... 199
Reports ... 200
Destructive Degradation ... 200
Stability Analysis ... 205

8 Destructive Degradation ... 209
Model Product Deterioration over Time

Example of the Destructive Degradation Platform ... 211
Launch the Destructive Degradation Platform ... 217
Specify Two Y Columns ... 218
The Destructive Degradation Plot Options and Models ... 219
Plot Options ... 220
Models ... 221
The Destructive Degradation Report ... 225
Model List ... 226
Model Outlines ... 226
Destructive Degradation Platform Options ... 229
Statistical Details for the Destructive Degradation Platform ... 229

9 Reliability Forecast ... 231
Forecast Product Failure Using Production and Failure Data

Overview of the Reliability Forecast Platform ... 233
Example Using the Reliability Forecast Platform ... 233
Launch the Reliability Forecast Platform ... 237
The Reliability Forecast Report ... 240
Observed Data Report ... 240
Life Distribution Report ... 240
Forecast Report ... 241
Reliability Forecast Platform Options ... 245
Additional Example Using the Reliability Forecast Platform ... 246

10 Reliability Growth ... 249
Model System Reliability as Changes Are Implemented

Overview of the Reliability Growth Platform ... 251
Example Using the Reliability Growth Platform ... 252
Launch the Reliability Growth Platform ... 256
Launch Window Roles ... 257
Data Table Structure ... 258

The Reliability Growth Report ... 261
Model Descriptions and Feasibilities ... 261
Observed Data Report ... 261
Reliability Growth Platform Options ... 264
Model Reports ... 265
Crow AMSAA ... 266
Crow AMSAA with Modified MLE ... 272
Fixed Parameter Crow AMSAA ... 273
Piecewise Weibull NHPP ... 274
Reinitialized Weibull NHPP ... 278
Piecewise Weibull NHPP Change Point Detection ... 281
Additional Examples of the Reliability Growth Platform ... 282
Piecewise NHPP Weibull Model Fitting with Interval-Censored Data ... 282
Piecewise Weibull NHPP Change Point Detection with Time in Dates Format ... 285
Statistical Details for the Reliability Growth Platform ... 287
Statistical Details for the Crow-AMSAA Report ... 287
Statistical Details for the Piecewise Weibull NHPP Change Point Detection Report ... 288

11 Reliability Block Diagram ... 289
Engineer System Reliabilities

Overview of the Reliability Block Diagram Platform ... 291
Example Using the Reliability Block Diagram Platform ... 291
Add Components ... 293
Align Shapes ... 295
Connect Shapes ... 296
Configure Components ... 297
The Reliability Block Diagram Window ... 299
Preview Window ... 300
Reliability Block Diagram Platform Options ... 301
Workspace Options ... 304
Configuration Settings ... 305
Distribution Configurations ... 305
Specify a Nonparametric Distribution ... 307
Options for Design and Library Items ... 308
Profilers ... 309
Component Importance and Time to Failure ... 311
Component Plots ... 313
Print Algebraic Reliability Formula ... 316
Generate Algebraic Expression Data Table ... 317
Clone and Delete ... 317

12 Repairable Systems Simulation 319
Estimate Outage Time for a Complex System
Overview of the Repairable Systems Simulation Platform 321
Example Using the Repairable Systems Simulation Platform 321
The Repairable Systems Simulation Window 323
System Diagram 324
Shapes Toolbar 325
Configuration Panel 326
Add Event Panel 329
Add Action Panel 330
Repairable Systems Simulation Platform Options 333
Options for Block Items 334
Distribution Options 334
Specify a Nonparametric or Estimated Distribution 335
Event Settings 336
Action Settings 337
Repairable Systems Simulation Results 338
Results Table 338
Results Explorer 339
Repairable Systems Simulation Results Options 341
RSS Results Explorer Point Estimation Profilers 342

13 Survival Analysis 343
Analyze Survival Time Data
Overview of the Survival Analysis Platform 345
Example of Survival Analysis 346
Launch the Survival Platform 348
The Survival Plot 349
Survival Platform Options 350
Exponential, Weibull, and Lognormal Plots and Fits 352
Fitted Distribution Plots 356
Competing Causes 357
Additional Examples of the Survival Platform 358
Example of Survival Analysis 358
Example of Competing Causes 361
Example of Interval Censoring 363
Statistical Reports for Survival Analysis 365

14 Fit Parametric Survival 369
Fit Survival Data Using Regression Models
Overview of the Fit Parametric Survival Platform 371

Example of Parametric Regression Survival Fitting ... 371
Launch the Fit Parametric Survival Platform ... 373
The Parametric Survival Fit Report ... 375
The Parametric Survival - All Distributions Report ... 376
Parametric Competing Cause Report ... 378
Fit Parametric Survival Options ... 378
Nonlinear Parametric Survival Models ... 380
Additional Examples of Fitting Parametric Survival ... 381
Arrhenius Accelerated Failure LogNormal Model ... 381
Interval-Censored Accelerated Failure Time Model ... 385
Analyze Censored Data Using the Nonlinear Platform ... 386
Left-Censored Data ... 388
Weibull Loss Function Using the Nonlinear Platform ... 389
Fitting Simple Survival Distributions Using the Nonlinear Platform ... 392
Statistical Details for the Fit Parametric Survival Platform ... 395
Loss Formulas for Survival Distributions ... 395

15 Fit Proportional Hazards ... 397
Fit Survival Data Using Semiparametric Regression Models

Overview of the Fit Proportional Hazards Platforms ... 399
Example of the Fit Proportional Hazards Platform ... 399
Risk Ratios for One Nominal Effect with Two Levels ... 401
Launch the Fit Proportional Hazards Platform ... 402
The Fit Proportional Hazards Report ... 404
Fit Proportional Hazards Options ... 405
Example Using Multiple Effects and Multiple Levels ... 406
Risk Ratios for Multiple Effects and Multiple Levels ... 408

A References ... 411

B Technology License Notices ... 415

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation 15
JMP Help 16
JMP Documentation Library 16
Additional Resources for Learning JMP 22
JMP Tutorials 22
Sample Data Tables 22
Learn about Statistical and JSL Terms 23
Learn JMP Tips and Tricks 23
JMP Tooltips 23
JMP User Community 24
Free Online Statistical Thinking Course 24
JMP New User Welcome Kit 24
Statistics Knowledge Portal 24
JMP Training 24
JMP Books by Users 25
The JMP Starter Window 25
JMP Technical Support 25

Formatting Conventions in JMP Documentation

These conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit https://www.jmp.com/software/pro.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from https://community.jmp.com.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Model Screening • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Legacy Control Charts • Pareto Plot • Diagram • Manage Spec Limits • OC Curves

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "JMP Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "JMP Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "JMP New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\16\Samples\Data

On macOS: \Library\Application Support\JMP\16\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit https://jmp.com/tools.

Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > Initial JMP Starter Window**.

JMP Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at https://www.jmp.com/support, including the technical support phone number.

Chapter 2

Introduction to Reliability and Survival

Lifetime and Failure Analysis

Reliability and Survival Methods describes a number of methods and tools that are available in JMP to help you evaluate and improve reliability in a product or system and analyze survival data for people and products:

- The Life Distribution platform enables you to analyze the lifespan of a product, component, or system to improve quality and reliability. This analysis helps you determine the best material and manufacturing process for the product, thereby increasing the quality and reliability of the product. See "Life Distribution".
- The Fit Life by X platform helps you analyze lifetime events when only one factor is present. You can choose to model the relationship between the event and the factor using various transformations, or create a custom transformation of your data. See "Fit Life by X".
- The Cumulative Damage platform enables you to analyze an accelerated life test where the stress levels might have changed over time. See "Cumulative Damage".
- The Recurrence Analysis platform analyzes event times where the events can recur several times for each unit. Typically, these events occur when a unit breaks down, is repaired, and then put back into service after the repair. See "Recurrence Analysis".
- The Degradation platform analyzes degradation data to predict pseudo failure times. These pseudo failure times can then be analyzed by other reliability platforms to estimate failure distributions. You can include an explanatory factor. You can perform stability analysis to set product expiration dates. You can also fit custom destructive degradation models. See "Degradation".
- The Destructive Degradation platform models failure data for product characteristics whose measurement requires that the product be destroyed. This results in a single observation per product unit. You can also include an acceleration factor. A wide range of common degradation models is available. See "Destructive Degradation".
- The Reliability Forecast platform helps you predict the number of future failures. The analysis estimates the parameters for a life distribution using production dates, failure dates, and production volume. See "Reliability Forecast".
- The Reliability Growth platform models the change in reliability of a single repairable system over time as improvements are incorporated into its design. See "Reliability Growth".

- JMP PRO The Reliability Block Diagram platform displays the reliability relationship between a system's components and, if reliability distributions are given to the components, analytically obtains the reliability behavior. See "Reliability Block Diagram".
- JMP PRO The Repairable Systems Simulation platform enables you to interactively define the relationships between the components of a repairable system. It can also simulate the down time of the system. See "Repairable Systems Simulation".
- The Survival platform computes survival estimates for one or more groups. It can be used as a complete analysis or is useful as an exploratory analysis to gain information for more complex model fitting. See "Survival Analysis".
- The Fit Parametric Survival platform fits the time to event variable using linear regression models that can involve both location and scale effects. The fit is performed using several distributions. See "Fit Parametric Survival".
- The Fit Proportional Hazards platform fits the Cox proportional hazards model, which assumes a multiplying relationship between predictors and the hazard function. See "Fit Proportional Hazards".

Chapter 3

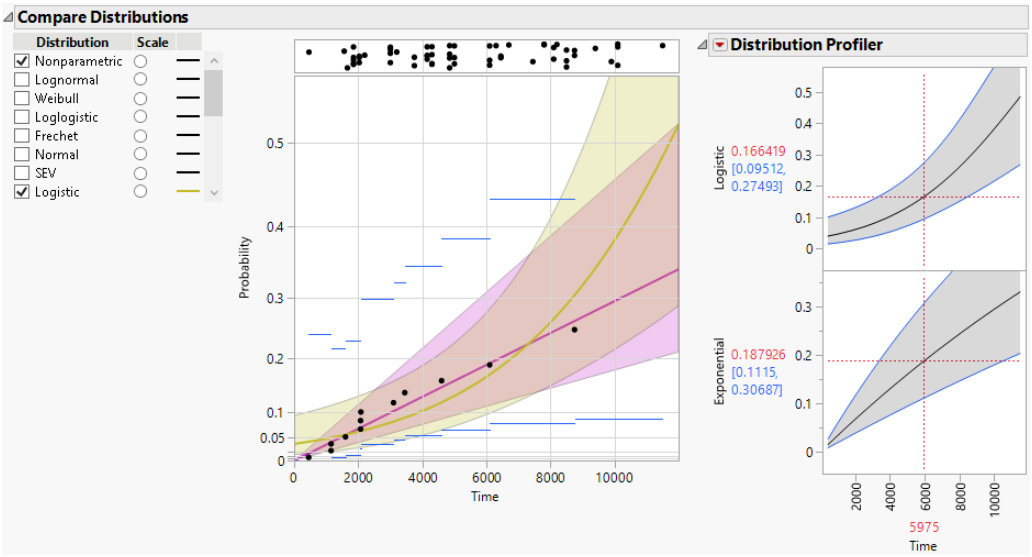
Life Distribution

Fit Distributions to Lifetime Data

The Life Distribution platform models time-to-event data. The platform accommodates both right-censored and interval-censored data. Use Life Distribution to do the following:

- Compare multiple distributional fits to determine which distribution best fits your data.
- Construct Bayesian fits.
- Model zero-failure data.
- Compare groups to analyze group differences.
- Analyze multiple causes of failure.
- Estimate the components of a mixture and estimate the probability that an observation comes from a given component.
- Estimate the components of a competing risk mixture and estimate the impact of a component on an observation.

Figure 3.1 Distributional Fits and Comparisons



Contents

Overview of the Life Distribution Platform . . . 31
Example of the Life Distribution Platform . . . 31
Launch the Life Distribution Platform . . . 34
Life Distribution Report . . . 37
Event Plot . . . 38
Compare Distributions . . . 41
Life Distribution: Statistics . . . 45
Life Distribution Report Options . . . 52
Mixture . . . 54
Fit Competing Risk Mixture . . . 57
Weibayes Report . . . 58
Competing Cause Report . . . 59
Competing Cause Workflow . . . 60
Competing Cause Model . . . 60
Cause Combination . . . 61
Competing Cause: Statistics . . . 62
Individual Causes . . . 63
Competing Cause Report Options . . . 63
Life Distribution - Compare Groups Report . . . 65
Compare Groups: Statistics . . . 66
Individual Group . . . 67
Life Distribution - Compare Groups Report Options . . . 67
Additional Examples of the Life Distribution Platform . . . 68
Omit Competing Causes . . . 69
Change the Scale . . . 70
Examine the Same Distribution across Groups . . . 72
Weibayes Estimates . . . 74
Fit Mixture Example . . . 76
Fit Competing Risk Mixture Example . . . 79
Statistical Details for the Life Distribution Platform . . . 81
Distributions . . . 81
Competing Cause Details . . . 96
Notes Regarding Median Rank Regression . . . 101

Overview of the Life Distribution Platform

Life data analysis, or *life distribution analysis*, is the process of modeling the lifespan of a product, component, or system to predict lifetime or time to failure. For example, you can observe failure rates over time to predict when a computer component might fail. This technique enables you to compare materials and manufacturing processes for the product, enabling you to increase the quality and reliability of the product. Decisions on warranty periods and advertising claims can also be more accurate.

With the Life Distribution platform, you can analyze *censored* data in which some time observations are unknown. And if there are potentially multiple causes of failure, you can analyze the *competing causes* to estimate which cause is more influential.

You can use the **Reliability Test Plan** and **Reliability Demonstration** calculators to choose the appropriate sample sizes for reliability studies. These calculators are found at **DOE > Sample Size and Power**. See the *Design of Experiments Guide*.

Example of the Life Distribution Platform

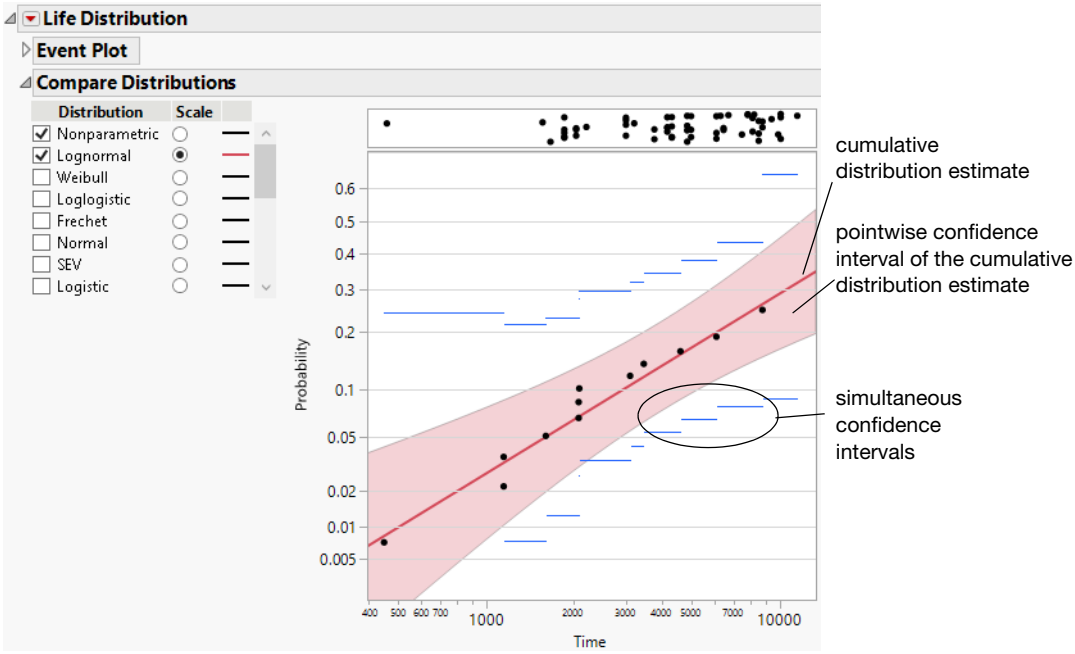
Suppose you have failure times for 70 engine fans, with some of the failure times censored. You want to fit a distribution to the failure times and then estimate various measurements of reliability.

1. Select **Help > Sample Data Library** and open Reliability/Fan.jmp.
2. Select **Analyze > Reliability and Survival > Life Distribution**.
3. Select Time and click **Y, Time to Event**.
4. Select Censor and click **Censor**.
5. Click **OK**.

   The Life Distribution report window appears.
6. In the Compare Distribution report, select **Lognormal** distribution and the corresponding **Scale** radio button.

   A probability plot appears in the report window.

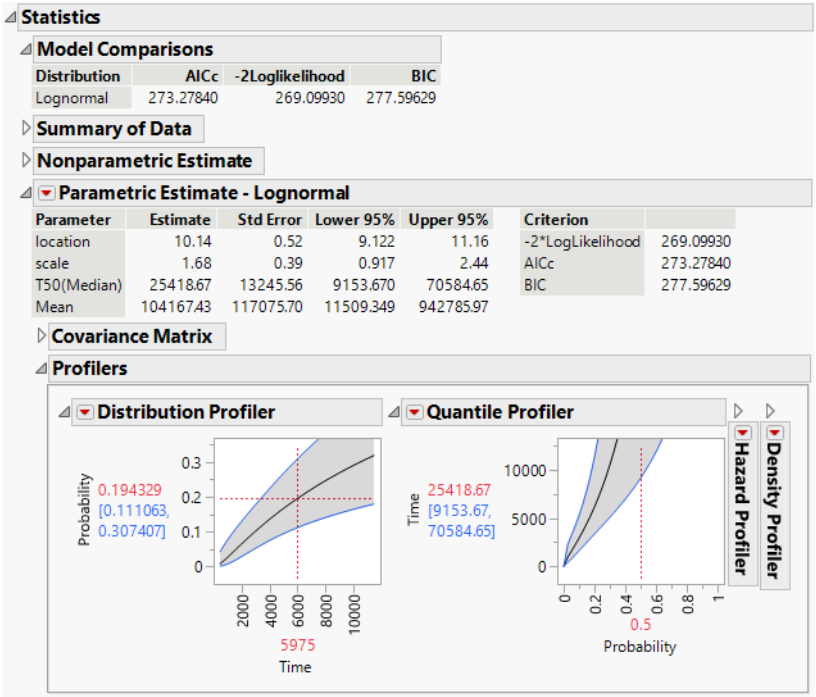
Figure 3.2 Probability Plot



In the probability plot, the data points generally fall along the red line, indicating that the lognormal fit is reasonable.

Below the Compare Distributions report, the Statistics report appears. This report provides a Model Comparison report, nonparametric and parametric estimates, profilers, and more.

Figure 3.3 Statistics Report

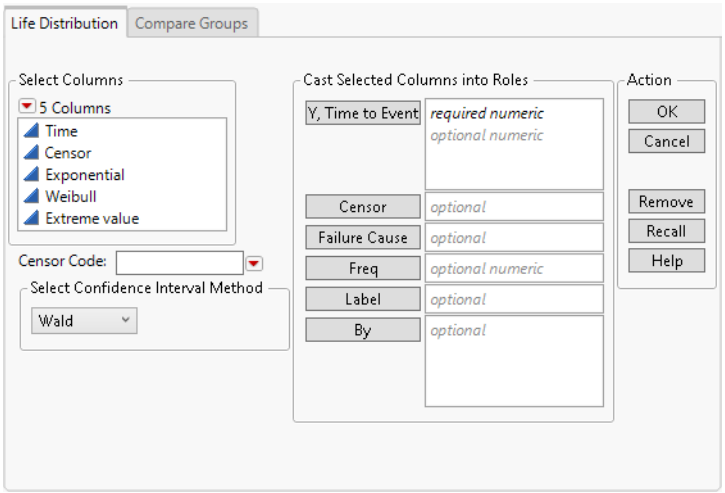


The parameter estimates for the lognormal distribution are provided. The profilers are useful for visualizing the fitted distribution and for estimating probabilities and quantiles. For example, the Quantile Profiler indicates that the estimated median time to failure is 25,418.67 hours.

Launch the Life Distribution Platform

Launch the Life Distribution platform by selecting **Analyze > Reliability and Survival > Life Distribution**. For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Figure 3.4 The Life Distribution Launch Window



Launch Window Tabs

The launch window includes two tabs:

- The Life Distribution tab models ungrouped data. The following types of reports can result:
  - The Life Distribution report appears when you do not specify a Failure Cause role. From this report, you can compare common distributions and examine statistics. See "Life Distribution Report".
  - The Weibayes report appears when you have zero failures in your data. See "Weibayes Report".
  - The Competing Cause report appears when you specify a Failure Cause role. In addition to the features in the Life Distribution report, you can also compare individual failure causes. See "Competing Cause Report".

  Note: You can examine Fixed Parameter and Bayesian models in the Life Distribution and Competing Cause reports.

- The Compare Groups tab enables you to specify a Grouping variable. The Compare Groups report compares different groups using a single specified distribution. For

example, you might compare Weibull fits for components grouped by supplier. In contrast, the Life Distribution tab compares several fitted distributions for a single group. See "Life Distribution - Compare Groups Report".

Launch Window Options

The launch window contains the following options:

Y, Time to Event One or more response columns. The number of response columns specified depends on the censoring structure in the data table.

- If one variable is specified, it is interpreted as the time to event (such as the time to failure) or time to right censoring. Use the Censor column to indicate right-censored responses. For more information about right censoring, see "Single Time to Event Column".
- If two variables are specified, they are interpreted as interval-censored observations. The first Y variable gives the lower limit and the second Y variable gives the upper limit for each unit. For an example of using two response columns to represent various types of censoring, see Figure 3.5. For more information about censoring with two response columns, see "Two Time to Event Columns".

Figure 3.5 Censored Data Types for Two Response Variables

	Start Time	End Time	Count	Censoring Type
1	•	50	50	Left
2	•	100	6	Left
3	25	•	30	Right
4	75	•	10	Right
5	80	150	4	Interval
6	100	250	7	Interval

- If three or more variables are specified, the report contains a separate analysis computed using each specified variable as time to event data.

Grouping (Appears only in the Compare Groups tab.) A column containing the groups that you want to compare. For an example, see "Examine the Same Distribution across Groups".

Censor A column that identifies right-censored observations. Select the value that identifies right-censored observations from the Censor Code menu beneath the Select Columns list. The Censor column is used only when one Y is entered.

Failure Cause A column that contains multiple failure causes. If a Failure Cause column is selected, then a section is added to the window. This section contains check boxes that allow the failure mode to use ZI distributions, TH distributions, DS distributions, fixed parameter models, or Bayesian models for the analysis. The following options are also available:

Distribution Specifies the initial distribution to fit for each failure cause. Select one distribution to fit for all causes; select **Individual Best** to let the platform automatically choose the best fit for each cause; or select **Manual Pick** to manually choose the distribution to fit for each failure cause after JMP creates the Life Distribution report. You can also change the distribution fits in the Life Distribution report itself.

Comparison Criterion (Appears only when you choose the **Individual Best** distribution fit.) Specifies the method by which JMP chooses the best distribution: Corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC), or twice the negative log-likelihood (-2Loglikelihood). See *Fitting Linear Models*. You can change the method later in the Model Comparisons report. See "Model Comparisons".

Censor Indicator in Failure Cause Column Identifies the value used in the **Failure Cause** column to indicate that an observation did not fail. To specify such an indicator, select this option and then enter the indicator in the box that appears. You can also select a value from the list to the right of the box.

See Meeker and Escobar (1998, ch. 15) for a discussion of multiple failure causes. "Omit Competing Causes" illustrates how to analyze multiple causes.

Freq A column that contains frequencies or observation counts when the information in a row represents multiple units. If the value in a row is 0 or a positive number, then the value represents the frequencies or counts of observations for that row.

Label A column that contains identifiers other than the row number. These labels appear on the *Y* axis in the event plot.

By An optional variable whose levels define rows used to create separate models.

Censor Code Identifies the value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

Select Confidence Interval Method Defines the method used for computing confidence intervals for the parameters. The default is Wald, but you can select Likelihood instead. However, all confidence intervals provided in the profilers are based on the Wald method. This is done to reduce computation time. See "Estimation and Confidence Intervals".

Failure Distribution by Cause (Appears only in the Life Distribution tab when a Cause is specified.) Specify which families of distributions should be available to model the life distributions for individual causes. Select an initial distribution, Individual Best, or Manual Pick from the Distribution menu. See "Failure Cause".

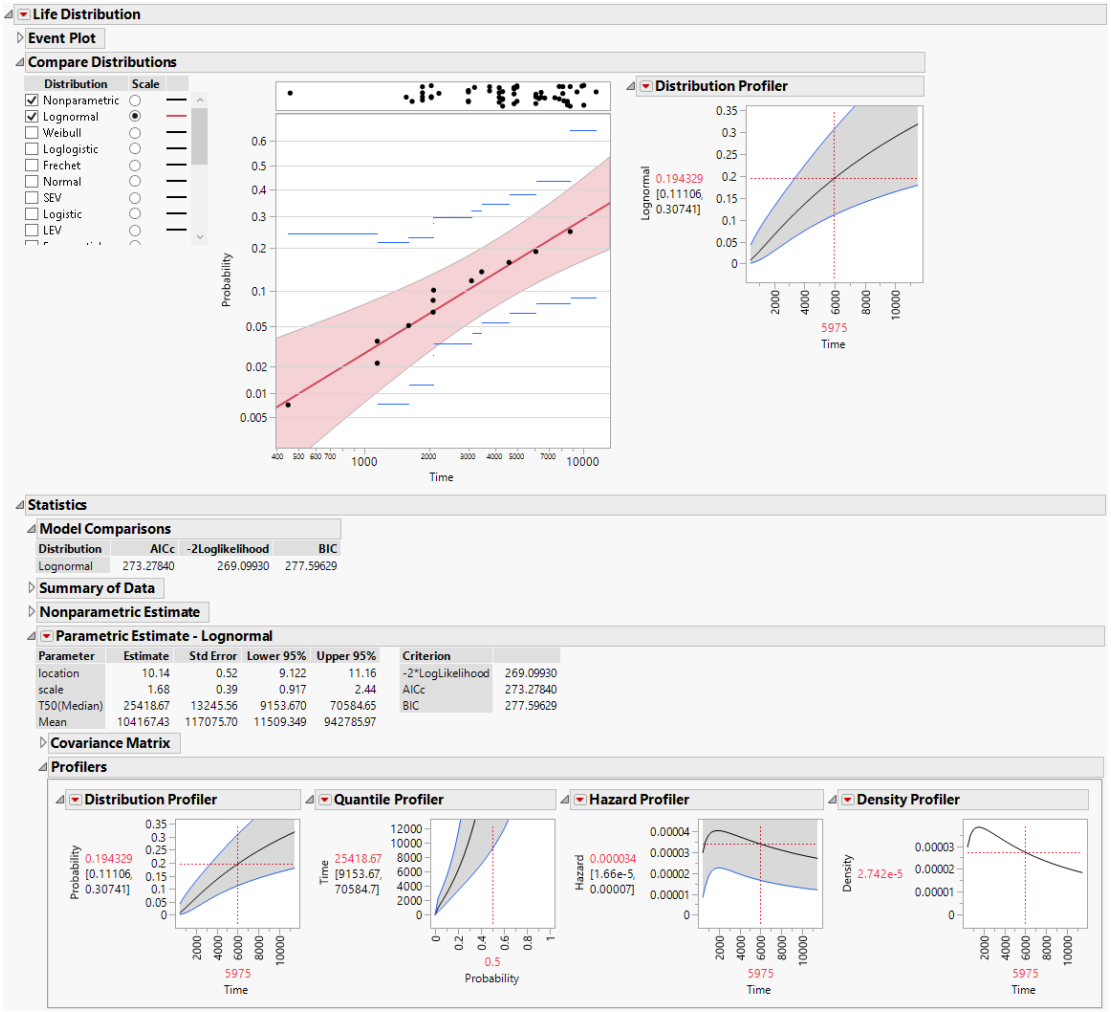
Life Distribution Report

Tip: If you find that the report window is too long, select **Tabbed Report** from the Life Distribution red triangle menu.

If you have not selected a Failure Cause in the launch window, the Life Distribution report appears. Use this report to analyze lifetime data where some time observations might be censored. The Life Distribution report contains the following content and options:

- "Event Plot"
- "Compare Distributions"
- "Life Distribution: Statistics"
- "Life Distribution Report Options"

Figure 3.6 Example of the Life Distribution Report for Fan.jmp



Event Plot

Click the Event Plot disclosure icon to see a plot of the failure or censoring times. For each row in the data table, the Event Plot shows a horizontal line indicating whether the units in the row have been censored. When units have been censored, the line indicates the nature of the censoring.

- The time period when the units in the row are known to be functioning is indicated with a solid line.
- The time period when it is *not* known if the units in the row are functioning is indicated with a dashed line.

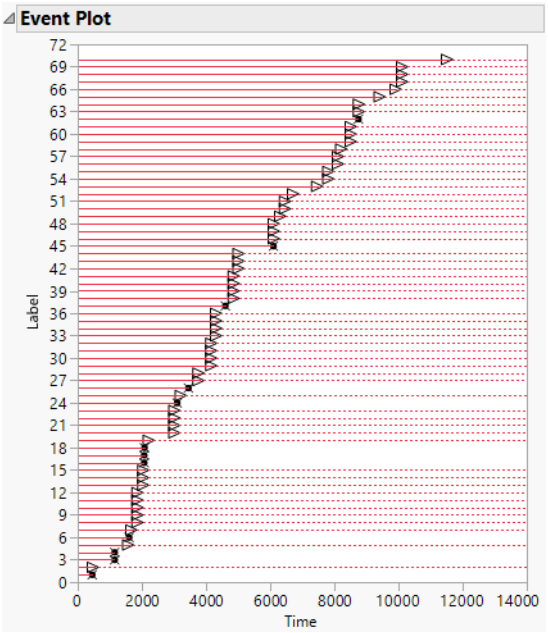
- The line terminates once the units in the row are known to have failed.

Single Time to Event Column

In the Fan.jmp sample data table there is a single Time column indicating failure time. When the failure time is unknown, the value Censored is recorded in the Censor column. All censored units are assumed to be right-censored. Figure 3.7 shows the Event Plot for this data.

Note: To construct the plot in Figure 3.7, select **Help > Sample Data Library** and open Reliability/Fan.jmp. Click the green triangle next to the **Life Distribution - Exponential** script. Click the **Event Plot** disclosure icon.

Figure 3.7 Event Plot for Right-Censored Data



The unit in row 3 failed at Time 1150. Its lifetime is represented by a solid horizontal line that ends at Time 1150. The failure time is marked with an “x”.

The unit in row 5 is right censored. It was last know to be functioning at Time 1560. The time period during which the unit is known to be functioning is represented by a solid horizontal line that ends at Time 1560. At Time 1560, a right arrow is plotted. The line continues as a dashed line, indicating that the failure time is unknown, but greater than 1560.

Two Time to Event Columns

In the Censor Labels.jmp sample data table, there are two columns, Start Time and End Time. Start Time indicates when units in a row were last known to be functioning. End Time indicates when units in that row were first known to have failed.

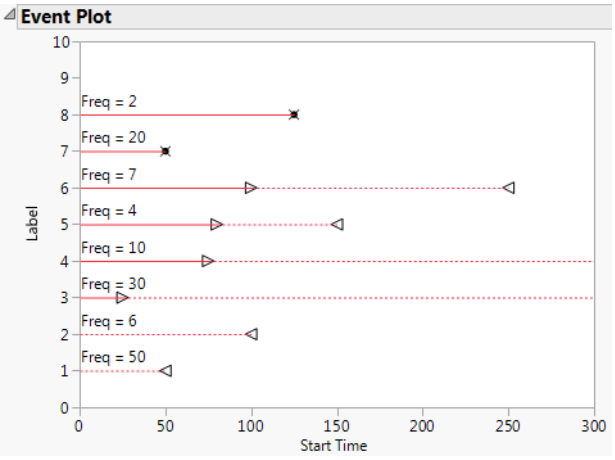
The Start Time and End Time values indicate the following about the units:

- Units in rows 1 and 2 are left censored. They were known to fail before the time in the End Time column, but their exact failure times are unknown.
- Units in rows 3 and 4 are right censored. They were known to be last functioning at the time in the Start Time column, but their failure times are unknown.
- Units in rows 5 and 6 are interval censored. They were known to fail within the interval defined by the Start Time and End Time.
- Units in rows 7 and 8 are not censored. Their failure times are given by the values in the Start Time and End Time columns, which are identical.

Figure 3.8 shows the Event Plot for this data.

Note: To construct the plot in Figure 3.8, select **Help > Sample Data Library** and open Censor Labels.jmp. Click the green triangle next to the **Life Distribution** script.

Figure 3.8 Event Plot for Mixed-Censored Data



The line patterns in the Event Plot represent the various types of censoring:

- The pattern → ····· indicates right censoring. The unit failed after its last inspection.
- The pattern ····· ◁ indicates left censoring. The unit failed after being put on test and prior to the indicated time, but it is not known when it was last functioning.

- The pattern indicates that the unit failed during the time interval marked by the two arrow heads.
- The pattern indicates no censoring. The unit failed at the time marked by the x.

Compare Distributions

The Compare Distributions report lets you fit and compare different failure time distributions. Two lists appear:

Distribution Select a distribution for the response. Different distributions appear based on characteristics of the data. For more information about which distributions are available, see "Available Parametric Distributions".

Scale Select a scale for the probability axis. The probability scale corresponds to the distribution listed to the left of the Scale button. Using this scale, the fitted model is represented by a line. Suppose that you fit a given distribution and then scale the axis using that distribution. If the points generally fall along a line, this indicates that the distribution provides a reasonable fit.

The default plot shows the nonparametric estimates (Kaplan-Meier-Turnbull) for the uncensored data values and their confidence intervals. The confidence intervals are indicated by horizontal blue lines.

Tip: To customize the plot of the nonparametric estimates, select **File > Preferences > Platforms > Life Distribution** and select one or more of the following preferences: **Show Shaded Pointwise Intervals**, **Show Shaded Simultaneous Intervals**, or **Show Staircase Style Function**.

By default, there is a panel at the top of the plot that displays times of right-censored observations.

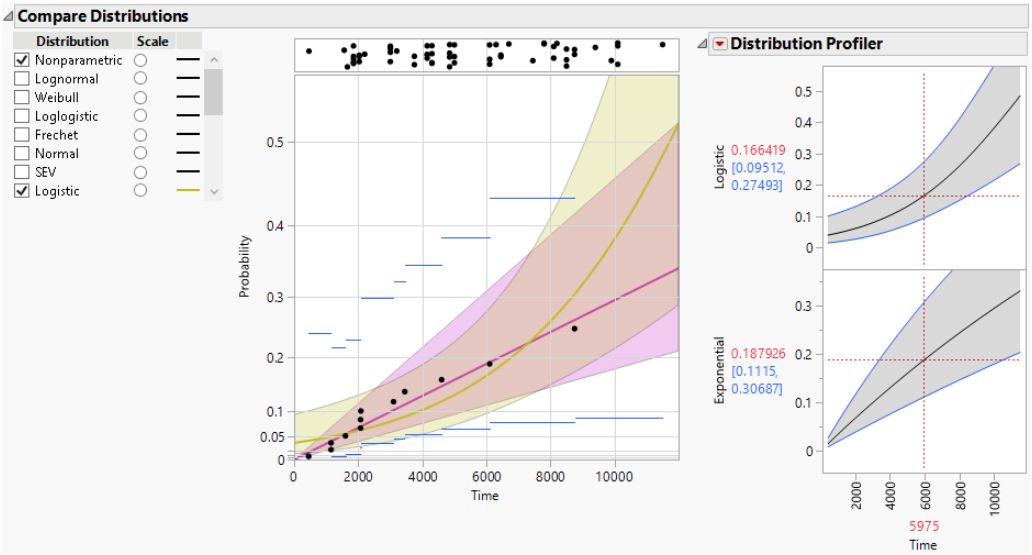
Tip: To hide the panel that contains markers for right-censored observations, select **File > Preferences > Platforms > Life Distribution** and uncheck **Show Markers for Right Censored Observations**.

For each distribution that you select, the Compare Distributions report is updated to show the following:

- the estimated cumulative distribution curve, which appears on the probability plot
- a shaded region that indicates confidence intervals for the cumulative distribution
- a Distribution Profiler that shows the cumulative probability of failure for a given period of time

Figure 3.9 shows an example of the Compare Distributions report. The Logistic (yellow) and Exponential (magenta) distributions are shown. The plot is scaled using the Exponential distribution.

Figure 3.9 Compare Distributions Report and Distribution Profiler



Available Parametric Distributions

This section addresses the distributions available in the Compare Distributions report.

Note: Distributions for the Competing Cause report are covered in "Available Distributions for Competing Cause Compare Distributions Reports".

The available distributions are listed and described in detail in "Parametric Distributions". There are four major groupings of parametric distributions:

- "Basic Failure-Time Distributions"
- "Threshold Distributions"
- "Defective Subpopulation Distributions"
- "Zero-Inflated Distributions"

Tip: To restrict which distributions are available by default, select **File > Preferences > Platforms > Life Distribution** and uncheck the distributions that you do not want to appear. The distributions listed include the Threshold, Defective Subpopulation, Zero-Inflated, LogGenGamma, and GenGamma distributions. By default, all of the distributions are checked and available.

The rules that determine which distributions appear in the Compare Distributions panel depend on the particular implementation. As a general guide, distributions are available if they are not disabled in Preferences and if they are appropriate in the given situation.

Basic Failure-Time Distributions

The basic failure-time distributions are available whenever all failure times are positive. They include the following:

- Lognormal
- Weibull
- Loglogistic
- Fréchet
- Normal
- SEV
- Logistic
- LEV
- Exponential
- LogGenGamma
- GenGamma

Note: When there are negative or zero failure times, only the Normal, SEV, Logistic, LEV, and LogGenGemma are available.

Threshold Distributions

The threshold (TH) distributions are always available. Threshold distributions are log-location-scale distributions with threshold parameters. The threshold parameter shifts the distribution away from 0. These distributions assume that all units survive until the threshold value. Threshold distributions are useful for fitting moderate to heavily shifted distributions. The threshold distributions are the following:

- TH Lognormal
- TH Weibull
- TH Loglogistic
- TH Fréchet

Defective Subpopulation Distributions

The defective subpopulation (DS) distributions are available when all failure times are positive. These distributions are useful when only a fraction of the population has a particular defect leading to failure. Use the DS distribution options to model failures that occur on only a subpopulation. The DS distributions are the following:

- DS Lognormal
- DS Weibull
- DS Loglogistic
- DS Fréchet

Zero-Inflated Distributions

When the time-to-event data contain zero as the minimum value in the Life Distribution platform, the following zero-inflated distributions are available:

- Zero-Inflated Lognormal (ZI Lognormal)
- Zero-Inflated Weibull (ZI Weibull)
- Zero-Inflated Loglogistic (ZI Loglogistic)
- Zero-Inflated Fréchet (ZI Fréchet)

Zero-inflated distributions are used when some proportion of units fails at time zero. When the data contain more zeros than expected by a standard model, the number of zeros is inflated.

Zero-Failure Data

In the case of zero-failure data, none of the above distributions are available by default. To obtain Bayesian fits for those distributions where the Bayesian Estimate option is available, select **File > Preferences > Platforms > Life Distribution** and uncheck **Weibayes Only for Zero Failure Data**. See "Weibayes Only for Zero Failure Data".

Parametric Distributions That Allow Bayesian Estimation

Bayesian estimation is available for the following parametric distributions:

- Lognormal
- Weibull
- Loglogistic
- Fréchet
- Normal
- SEV

- Logistic
- LEV

A list of distributions that are available as priors for hyperparameters of these distributions is given in "Prior Distributions for Bayesian Estimation".

Life Distribution: Statistics

The Statistics report includes the following sub-reports:

- "Model Comparisons"
- "Summary of Data"
- "Nonparametric Estimate"
- "Parametric Estimate - <Distribution Name>" (one report appears for each distribution that you select in the Compare Distributions report)

Model Comparisons

The Model Comparisons report provides the AICc, -2Loglikelihood, and BIC statistics for each fitted distribution. Smaller values of each of these statistics indicate a better fit. For more information about these statistics, see *Fitting Linear Models*.

Initially, the rows are sorted by AICc. To change the statistic used to sort the report, click the Life Distribution red triangle and select **Comparison Criterion**. See "Life Distribution Report Options" for more information about this option.

Summary of Data

The Summary of Data report shows the total number of units observed, the number of uncensored units, and the numbers of right-censored, left-censored, and interval-censored units.

Nonparametric Estimate

The Nonparametric Estimate report shows nonparametric estimates for each observation. For right-censored data specified as a single Time to Event column, the report gives the following:

Midpoint Estimate Midpoint-adjusted Kaplan-Meier estimates.

Lower 95%, Upper 95% Pointwise 95% confidence intervals. You can change the confidence level by selecting Change Confidence Level from the report options.

Simultaneous Lower 95% (Nair), Simultaneous Upper 95% (Nair) Simultaneous 95% confidence intervals. You can change the confidence level by selecting Change Confidence Level from the report options. See Nair (1984) and Meeker and Escobar (1998).

Kaplan-Meier Estimate Standard Kaplan-Meier estimates.

If failure times are represented by two Time to Event columns, the report gives Turnbull estimates (in a column called Estimate), pointwise confidence intervals, and simultaneous confidence intervals (Nair).

See "Nonparametric Fit" for more information about nonparametric estimates.

Parametric Estimate - <Distribution Name>

A report called Parametric Estimate - <Distribution Name> appears for each distribution that is fit. The report gives the distribution's parameter estimates, their standard errors, and confidence intervals. The criteria that appear in the Model Comparisons report are shown under Criterion.

Note: Whenever an estimate of the mean is provided, its confidence interval is computed as a Wald interval even if you select Likelihood as the Confidence Interval Method in the launch window. In this case, the notation Mean (Wald CI) appears in the Parameter column to indicate that the confidence interval for the mean is a Wald interval.

For more information about how the distributions are parametrized, see "Parametric Distributions".

The Parametric Estimate report contains the following reports:

- "Covariance Matrix"
- "Profilers"
- Additional reports can be added by selecting report options from the Parametric Estimate red triangle menu. These include the Fix Parameter, Bayesian Estimates, Custom Estimation (Estimate Probability, Estimate Quantile), and Mean Remaining Life reports. See "Parametric Estimate Options".

Covariance Matrix

For each distribution, the Covariance Matrix report shows the covariance matrix for the estimates.

Profilers

Four types of profilers appear for each distribution:

- The Distribution Profiler shows cumulative failure probability as a function of time.

- The Quantile Profiler shows failure time as a function of cumulative probability.
- The Hazard Profiler shows the hazard rate as a function of time.
- The Density Profiler shows the density function for the distribution.

The profilers contain the following red triangle menu options:

Confidence Intervals The Distribution, Quantile, and Hazard profilers show Wald-based confidence curves for the plotted functions. This option shows or hides the confidence curves.

Reset Factor Grid Displays a window for each factor enabling you to enter a specific value for the factor's current setting, to lock that setting, and to control aspects of the grid. See *Profilers*.

Factor Settings Provides a menu that consists of several options. See *Profilers*.

Note: The confidence intervals provided in the profilers are based on the Wald method even if the Likelihood Confidence Interval Method is selected in the launch window. This is done to reduce computation time.

Parametric Estimate Options

The Parametric Estimate red triangle menu contains the following options:

Save Probability Estimates Saves the estimated failure probabilities and confidence intervals to the data table.

Save Quantile Estimates Saves the estimated quantiles and confidence intervals to the data table.

Save Hazard Estimates Saves the estimated hazard values and confidence intervals to the data table.

Show Likelihood Contour Shows or hides a contour plot of the log-likelihood function. If you have selected the Weibull distribution, a second contour plot appears for the alpha-beta parameterization. This option is available only for distributions with two parameters.

Show Likelihood Profiler Shows or hides a profiler of the log-likelihood function. This option is not available for the threshold (TH) distributions.

Fix Parameter Opens a report where you can specify the value of parameters. Enter your fixed parameter values, select the appropriate check box, and then click **Update**. JMP re-estimates the other parameters, covariances, and profilers based on the new parameters, and shows them in the Fix Parameter report. A distribution profiler of the unconstrained model is shown below the distribution profiler for the fixed parameter

model. For an example in a competing cause situation, see "Specify a Fixed Parameter Model as a Distribution for a Cause".

For the Weibull distribution, the **Fix Parameter** option lets you select the Weibayes method. For an example, see "Weibayes Estimates". The Weibayes option is not available for interval-censored data.

Bayesian Estimates Performs Bayesian estimation of parameters for certain distributions based on three methods of specifying prior distributions (Location and Scale Priors, Quantile and Parameter Priors, and Failure Probability Priors). See "Bayesian Estimation - <Distribution Name>". This option is available only for the following distributions: Lognormal, Weibull, Loglogistic, Fréchet, Normal, SEV, Logistic, LEV.

Custom Estimation Provides calculators that enable you to predict failure probabilities, survival probabilities, and quantiles for specific time and failure probability values. Each calculated quantity includes confidence intervals, which can be two-sided or one-sided (in either direction). Two reports appear: Estimate Probability and Estimate Quantile. See "Custom Estimation".

Mean Remaining Life Provides a calculator that enables you to estimate the mean remaining life of a unit. In the Mean Remaining Life Calculator, enter a Time and press **Enter** to see the estimate. Click the plus sign to enter additional times. This calculator is available for the following distributions: Lognormal, Weibull, Loglogistic, Fréchet, Normal, SEV, Logistic, LEV, and Exponential.

Bayesian Estimation - <Distribution Name>

For certain distributions, you can fit Bayesian models. This is done using rejection sampling or a Markov Chain Monte Carlo (MCMC) algorithm. More specifically, the platform attempts a basic rejection sampler. If the rejection sampler produces valid results, these results are reported. If the rejection sampler cannot produce valid results, the platform uses a random walk Metropolis-Hastings algorithm and adds a note to the top of the Bayesian Estimation report. See Robert and Casella (2004).

From the Parametric Estimate - <Distribution Name> report outline, select **Bayesian Estimates**. This opens an outline called Bayesian Estimation - <Distribution Name>. The initial report is a control panel where you can specify the parameters for the priors and control aspects of the simulation.

The following steps describe the workflow:

- Select a prior specification method from the Bayesian Estimation red triangle menu and set values for the parameters of the priors. See "Bayesian Estimation Red Triangle Options".
- Specify the simulation options. See "Bayesian Estimates - Result <N>".
- Select Fit Model to fit a model. See "Bayesian Estimates - Result <N>".

Bayesian Estimation Red Triangle Options

You can choose from the following prior specification methods in the Bayesian Estimation red triangle menu:

Location and Scale Priors Enables you to specify hyperparameters for prior distributions on generic parameters (location and scale parameters). Select the Prior Distribution red triangle menu to select a distribution for each parameter. You can enter new values for the hyperparameters of the priors. The initial values that are provided are estimates consistent with the MLEs. See "Prior Distributions for Bayesian Estimation".

Quantile and Parameter Priors Enables you to specify prior information about a quantile and the scale parameter (or Weibull β if the parametric fit is Weibull). The quantile is defined by the value next to Probability. The default Probability value is 0.10, but you can specify a value that corresponds to the quantile of interest. Specify information about the prior information in terms of Lower and Upper 99% limits on the range of each prior distribution. See Meeker and Escobar (1998). The initial values that are provided are estimates consistent with the MLEs. See "Prior Distributions for Bayesian Estimation".

Failure Probability Priors Enables you to specify prior information about failure probabilities at two distinct time points. You can specify the two time points. The prior distribution for each time point is Beta. You can specify the prior distributions using either of two synchronized approaches:

1. Specify failure probability by estimates and error percentages. The prior information for each Beta prior distribution can be specified using a probability estimate and an estimate error. See Kaminskiy and Krivtsov (2005).
2. Specify failure probability estimate ranges. You can specify the 99% range for the two Beta distributions in the following ways:

– For each failure time, enter an initial value for the Lower and Upper 99% Limits.

– Click the vertical line segments in the graph and drag them to your two time points. Adjust the vertical spread of each marker to specify the 99% limits.

Simulation Options

For any of the prior specification methods that you select in the Bayesian Estimation red triangle menu, the following options appear at the bottom of the panel:

Number of Monte Carlo Iterations Controls the sample size that will be drawn from the posterior distribution after a burn-in procedure.

Random Seed Sets the initial state of the simulation. By default, it is the clock time. The number should be a positive integer greater than 1. If you specify 1, the current clock time is used.

Show Prior Scatter Plot Select this option to draw random samples from the prior distributions and to plot results on a scatter plot. After you select Fit Model, the scatter plot appears in an outline entitled Prior Scatter Plot in the Bayesian Estimates - Results <N> report.

Overlay Likelihood Contour Overlays likelihood-based contours on scatter plots in the Bayesian Estimates Results report.

Fit Model Estimates the posterior lifetime distribution based on prior distributions that JMP fits using the values that you specified. Adds a report entitled Bayesian Estimates - Results <N>, where N is an integer that consecutively numbers the Bayesian Results reports.

Bayesian Estimates - Result <N>

Once you have specified priors using one of the red triangle menu options, select Fit Model. A Bayesian Estimates - Result <N> report is provided for each selection of priors. This report contains these headings:

Priors Documents the specifications that you entered in the Bayesian Estimation report to fit the Bayesian model. The Priors report also specifies the random seed.

Posterior Estimates Shows five marginal statistics that describe the posterior distribution of the generic parameters (location and scale parameters). The marginal statistics are the median, 0.025 quantile (Lower Bound), 0.975 quantile (Upper Bound), mean, and standard deviation computed from the Monte Carlo samples. When the posterior estimates are generated using the Quantile and Parameter Priors specification, this table also includes the posterior estimate of the quantile and Slope β (for the Weibull distribution).

To compute statistics for other derived variables based on the posterior estimates of the generic parameters, click the Export Monte Carlo Samples link.

Prior Scatter Plot Appears when you select Show Prior Scatter Plot before clicking Fit Model. Shows prior scatter plots of parameters or equivalent quantities associated with the prior specification method for the distribution.

Posterior Scatter Plot Shows posterior scatter plots of parameters or equivalent quantities associated with the prior specification method for the distribution.

Profilers Shows two profilers based on samples from the posterior distribution.

The values shown in the Distribution Profiler, at a given time *t*, are calculated as follows:

- For each set of sampled parameter values from the posterior distribution, the value of the cumulative distribution function at time *t* is calculated.
- The predicted value is the median of these calculated values.

– The upper and lower confidence limits are the 0.025 and 0.975 quantiles of these calculated values.

The plot and confidence limits shown in the Quantile Profiler are obtained in a similar fashion. For a given Probability value p, the quantiles corresponding to p are calculated from the distributions associated with the posterior parameter values.

Weibayes Only for Zero Failure Data

In a zero-failure situation, no units fail. All observations are right censored. If you have zero-failure data, it is possible to conduct either Bayesian estimation or Weibayes inference. See "Weibayes Report".

Note: By default, zero-failure data is analyzed using the Weibayes method. If you want to conduct a broader Bayesian analysis on zero-failure data, select **File > Preferences > Platforms > Life Distribution** and uncheck **Weibayes Only for Zero Failure Data**.

Custom Estimation

The Custom Estimation option produces two reports: Estimate Probability and Estimate Quantile. The Estimate Probability report contains a calculator that enables you to predict failure and survival probabilities for specific time values. The Estimate Quantile report contains a calculator that enables you to predict quantiles for specific failure probability values. Both Wald-based and likelihood-based confidence intervals appear for each estimated quantity. The confidence level for these intervals is determined by the Change Confidence Level option in the Life Distribution red triangle menu.

Estimate Probability

In the Estimate Probability calculator, enter a value for Time. Press **Enter** to see the estimates of failure probabilities, survival probabilities, and corresponding confidence intervals. To calculate multiple probability estimates, click the plus sign, enter another Time value in the box, and press **Enter**. Click the minus sign to remove the last entry.

The Estimate Probability calculator contains an option, Side, that enables you to change the form of the intervals. Select one of the following suboptions:

Two Sided Provides two-sided confidence intervals for failure probability and survival probability.

Upper Failure Probability Provides one-sided confidence intervals that contain upper limits for the failure probability and lower limits for the survival probability.

Lower Failure Probability Provides one-sided confidence intervals that contain lower limits for the failure probability and upper limits for the survival probability.

Estimate Quantile

In the Estimate Quantile report, enter a value for Failure Probability. Press **Enter** to see the quantile estimates and corresponding confidence intervals. To calculate multiple quantile estimates, click the plus sign, enter another Failure Probability value in the box, and press **Enter**. Click the minus sign to remove the last entry.

The Estimate Quantile calculator contains an option, Side, that enables you to change the form of the intervals. Select one of the following suboptions:

Two Sided Provides two-sided confidence intervals for quantiles.

Lower Provides one-sided confidence intervals that contain lower limits for quantiles.

Upper Provides one-sided confidence intervals that contain upper limits for quantiles.

Life Distribution Report Options

The Life Distribution red triangle menu contains the following options:

Fit All Distributions Fits all distributions other than the threshold (TH) distributions. The distributions are compared in the Model Comparisons report. See "Compare Distributions".

> **Tip:** Select the **Comparison Criterion** option to change the criterion for finding the best distribution.

Fit All Nonnegative Fits all nonnegative distributions (Exponential, Lognormal, Loglogistic, Fréchet, Weibull, and Generalized Gamma). The distributions are compared in the Model Comparisons report. See "Compare Distributions". Note the following:

- The option does not fit DS or TH distributions.
- If the data have negative values, then the option produces no results.
- If the data have zeros, the option fits the four zero-inflated (ZI) distributions: ZI Lognormal, ZI Weibull, ZI Loglogistic, and ZI Fréchet. For more information about zero-inflated distributions, see "Zero-Inflated Distributions".

Fit All DS Distributions Fits all defective subpopulation (DS) distributions: DS Lognormal, DS Weibull, DS Loglogistic, and DS Fréchet. For more information about defective subpopulation distributions, see "Distributions for Defective Subpopulations".

Fit Mixture Fits a distribution that is a mixture of the distributions other than the threshold (TH) distributions. See "Mixture".

Fit Competing Risk Mixture Fits a competing risk mixture distribution to the data. See "Fit Competing Risk Mixture".

Show Points Shows or hides data points in the probability plot. The Life Distribution platform uses the midpoint estimates of the step function to construct probability plots. When you deselect **Show Points**, the midpoint estimates are replaced by Kaplan-Meier estimates.

Show Event Plot Frequency Label (Appears only if you have specified a Freq variable.) Shows or hides the Frequency label in the Event Plot.

Show Survival Curve Switches between the failure probability and the survival curve on the Compare Distributions probability plot and the Distribution Profiler plots.

Show Quantile Functions Shows or hides a Quantile Profiler that overlays the plots for the selected distributions. The Quantile plot also shows points plotted at time values. The plot appears beneath the Compare Distributions report. If you select distributions in any of the Compare Distributions, Quantile Profiler, and Hazard Profiler plots, they appear in the other two plots.

Show Hazard Functions Shows or hides a Hazard Profiler that overlays the plots for the selected distributions. The plot appears above the Statistics report. If you select distributions in any of the Compare Distributions, Quantile Profiler, and Hazard Profiler plots, they appear in the other two plots.

Show Statistics Shows or hides the Statistics report. See "Life Distribution: Statistics".

Tabbed Report Shows graphs and data on individual tabs rather than in the default outline style.

Show Confidence Area Shows or hides the shaded confidence regions in the plots.

Interval Type Determines the type of confidence interval shown for the Nonparametric fit in the Compare Distributions plot. Select either pointwise or simultaneous confidence intervals.

Change Confidence Level Enables you to change the confidence level for the entire platform. All plots and reports update accordingly.

Comparison Criterion Enables you to select the criterion used to rank models in the Model Comparison report. For all three criteria, smaller values indicate better fit. Burnham and Anderson (2002) and Akaike (1974) discuss using AICc and BIC for model selection. See *Fitting Linear Models*.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

If you have specified a By variable, separate Life Distribution reports appear for each level of the By variable.

Save By Group Results For each By group, saves the estimates that appear in all Parametric Estimate reports for that group as a separate row in a new table.

Do Same Analyses for All Groups Applies all of the selected options for the current group to all other By groups.

Mixture

The Fit Mixture option adds the Mixture outline to the report where you can fit a mixture distribution to the data. For an example, see "Fit Mixture Example".

The mixture distribution's probability function F(x) is defined as follows:

$$F(x) = \sum_{i=1}^{k} w_i F_i(x)$$

where $F_i(x)$ is one of the supported distributions, k is the number of components in the mixture, and the w_i are positive weights that sum to 1. The Fit Mixture option attempts to identify clusters of observations that are drawn from each of the component distributions, $F_i(x)$. It estimates the parameters of the mixture and the probability that an observation is drawn from any given component.

Model Fit and Mixture Starting Value Methods

The fitting methodology is based on assumptions about the underlying clusters, called the Starting Value Method. Suppose that you designate k distributions. There are three Starting Value Methods:

- Single Cluster assumes that all observations are affected by all of the ingredient distributions to some extent. None of the densities stand out as affecting only a portion of the observations.

- Separable Clusters assumes that the ingredient distributions affect some observations more profoundly than others. For separable clusters, each of the *k* densities has an identifiable mode and defines a cluster.
- Overlapping Clusters assumes a situation that is intermediate between Single Cluster and Separable Clusters. Some densities stand out, but others jointly affect a portion of the observations. In this case, there are *m* clusters in the data, where *m* is less than *k*, the total number of densities.

The fitting process consists of these steps:

1. Clusters of observations are defined.
2. Assignment of clusters to densities is based on the Starting Value Method:
   - For Separable Clusters, the highest likelihood assignment of clusters to the specified ingredient densities is determined by examining the possible permutations.
   - For Overlapping Clusters, the highest likelihood assignment of clusters to the specified ingredient densities is determined by examining the possible permutations of clusters and combinations of observations.

Note: Suppose that you fit a model using a given Starting Value Method and then select another Starting Value Method. If a better fit based on the likelihood value cannot be achieved, no new model is added.

Mixture Control Panel

The control panel consists of these items:

Ingredient Lists distributions that you can use as components of the fitted mixture distribution.

Quantity Select the number of components in the mixture distribution that have the given distribution. The sum of the Quantity values is *k*, the number of densities in the mixture.

Starting Value Methods Select a method that reflects your assumptions about the mixture. See "Model Fit and Mixture Starting Value Methods".

Overlay Shows the nonparametric estimates (Kaplan-Meier-Turnbull) for the uncensored data values. When you fit a mixture, the plot is updated to show the model and 95% level confidence bands. The confidence level for these bands is determined by the Change Confidence Level option in the Life Distribution red triangle menu. A Legend appears to the right of the plot.

Go Click **Go** to fit the desired mixture. The Model List is updated with the model that you fit, and a report with the name of the mixture model is added.

Fit Mixture Reports

Model List

The Model List report lists the mixture distributions that you fit. The report provides the number of parameters, the number of actual observations, and the AICc, -2Loglikelihood, and BIC statistics for each mixture distribution. For more information about these statistics, see *Fitting Linear Models*.

Note the following:

- Smaller values of each of these statistics indicate a better fit.
- The rows are sorted by AICc.
- The **Comparison Criterion** red triangle option does not affect the order of models in the Model List.
- The AICc, -2Loglikelihood, and BIC statistics also appear in the Model Comparisons table. This enables you to compare mixture distribution to other distributions for your data. See "Model Comparisons".

Mixture Reports

The Model List report is followed by reports for each of the mixture distributions that you have fit. The title of each report describes the corresponding mixture using the specified ingredients and their quantities. The report lists the parameters, their estimates, standard errors, and 95% Wald confidence intervals. These intervals are not affected by the selection of Likelihood as the Confidence Interval Method in the launch window.

Parameter estimates are given for each distribution in the mixture. The Parameter column also includes parameters called Portion <i>, where i = 1, 2,..., k-1. These are estimates of the weights w_i for the mixture. Since the weights sum to 1, the k^{th} weight can be computed from the first k - 1 weights.

Density Overlay Plot

The Density Overlay plot shows estimates of the density functions for each of the components in the mixture. A legend to the right of the plot enables you to select which density functions appear.

Mixture Report Options

The red triangle menu contains the following options:

Remove Removes the model report and the entry for the model in the Model List.

Show Profilers Shows four types of profilers for the combined mixture distribution *F*. See "Mixture Profiler Options" for a description of their red triangle options.

- The Distribution Profiler shows cumulative failure probability as a function of time.
- The Quantile Profiler shows failure time as a function of cumulative probability.
- The Hazard Profiler shows the hazard rate as a function of time.
- The Density Profiler shows the density function for the distribution.

Save Predictions For each mixture density, saves a column to the data table containing the probability that an observation belongs to that density. For the formulas used in the calculation, see "Fit Mixture Save Predictions Formulas".

Mixture Profiler Options

The profilers for each mixture report contain the following red triangle options:

Confidence Intervals The Distribution, Quantile, and Hazard profilers show 95% Wald-based confidence curves for the plotted functions. This option shows or hides the confidence curves. The confidence level for these curves is determined by the Change Confidence Level option in the Life Distribution red triangle menu.

Note: To reduce computation time, the confidence intervals provided in the profilers are based on the Wald method, even if the Likelihood Confidence Interval Method is selected in the launch window.

Reset Factor Grid Displays a window for each factor enabling you to enter a specific value for the factor's current setting, to lock that setting, and to control aspects of the grid. See *Profilers*.

Factor Settings Provides a menu that consists of options relating to profiler settings, scripts, and linking profilers. See *Profilers*.

Fit Competing Risk Mixture

The Fit Competing Risk Mixture option enables you to fit competing risk mixture distributions. It estimates the probability that a given observation fails due to the cause represented by each of the component mixture distributions. For an example, see "Fit Competing Risk Mixture Example".

The competing risk mixture probability distribution function *F(x)* is defined as follows:

$$F(x) = 1 - \prod_{i=1}^{k} (1 - F_i(x))$$

where $F_i(x)$ represents the cumulative failure distributions for the i^{th} risk and k is the number of components (or risks) in the mixture. The Fit Competing Risk Mixture option attempts to identify clusters of observations that are drawn from each of the component distributions, $F_i(x)$. It estimates the parameters of the mixture and the probability that an observation is drawn from any given component.

The Competing Risk Mixture report is structured in a fashion that mirrors the Mixture report. See "Mixture". However, the Fit Competing Risk Mixture reports do not show a Density Overlay plot. They show a Distribution Overlay plot instead.

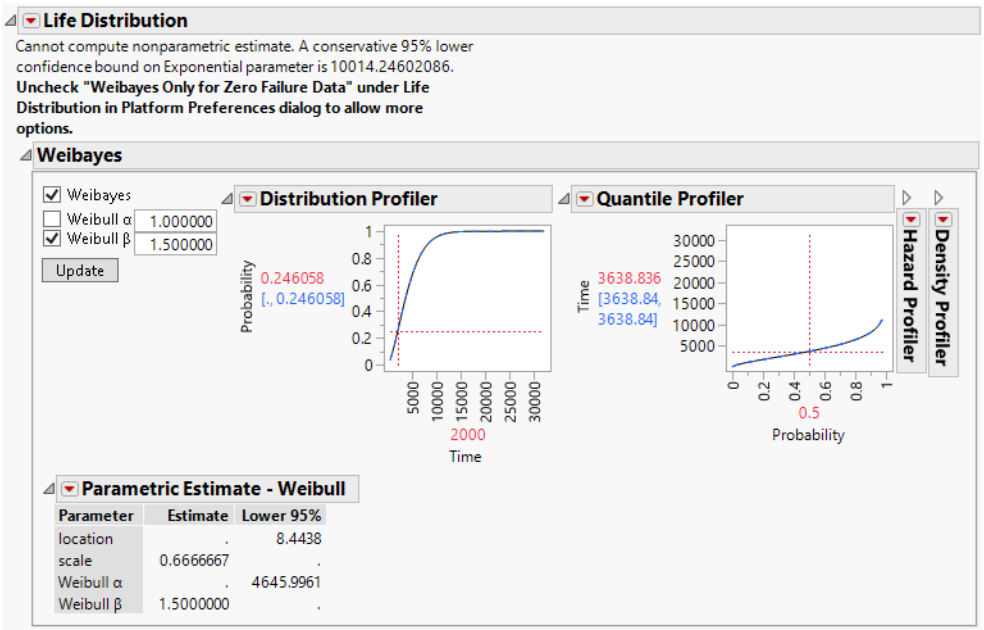
Distribution Overlay Plot

The Distribution Overlay plot shows the cumulative distribution functions for each of the mixture components and for the combined mixture (Aggregated). A legend to the right of the plot enables you to select which cumulative distribution functions appear.

Weibayes Report

If you have data with zero failures (right-centered) and you have not turned off the preference **Weibayes Only for Zero Failure Data**, a special Weibayes report appears. Figure 3.10 shows the Weibayes report for the Weibayes No Failures.jmp sample data table, found in the Reliability folder.

Figure 3.10 Weibayes Report



The analysis begins by assuming an exponential model. The note at the top of the report provides a lower confidence bound for the parameter of an exponential distribution. This lower confidence bound is computed using the method described in Meeker and Escobar (1998, sec. 7.7).

The Weibayes section of the report conducts the Weibayes analysis. For a description of the procedure, see Nelson (1985).

To obtain Weibayes estimates, make sure that Weibayes and Weibull β options are selected. Change the Weibull β value and click Update. The estimates and profilers are updated. The values shown in the profilers use the conservative confidence bound. For an example, see "Weibayes Example for Data with No Failures".

Note: If you deselect the Weibayes option, JMP shows the fixed parameter MLE.

When your data table contains at least one failure, a full Life Distribution report appears, but the maximum likelihood inference might not produce useful results. In this instance, a Weibayes analysis might be preferable. For an example, see "Weibayes Example for Data with One Failure".

Note: To conduct Bayesian inference using the usual Life Distribution options when you have zero-failure data, select **File > Preferences > Platforms > Life Distribution** and deselect **Weibayes Only for Zero Failure Data** before launching the analysis.

Competing Cause Report

Tip: If you find that the report window is too long, select **Tabbed Report** from the Competing Cause red triangle menu.

The Competing Cause report appears if you have assigned a Failure Cause column in the launch window. Use this report to analyze the competing causes to determine which causes are influential. For an example of this report, see "Omit Competing Causes". For technical details, see "Competing Cause Details".

The Competing Cause report contains the following content and options:

- "Cause Combination"
- "Competing Cause: Statistics"
- "Individual Causes"
- "Competing Cause Report Options"

Competing Cause Workflow

Follow these steps to facilitate your use of the Competing Cause report:

1. For convenience, click the Competing Cause red triangle and select the Tabbed Report option.
2. Select the Individual Causes tab. For each failure cause, use the options in its individual Life Distribution outline to select a distributional fit. See "Individual Causes".
3. Select the Cause Combination tab. For each failure cause, specify the desired distribution in each **Distribution** list. See "Cause Combination".
4. Click **Update Model**.
5. Select the Statistics tab to explore and save results for the model. See "Competing Cause: Statistics".

Note: Customizations made to the competing cause model report in the Statistics outline might be lost if you change the model and click **Update Model** again.

Competing Cause Model

In a competing cause situation, the aggregated failure function can be written as follows:

$$F(x) = 1 - \prod_{i=1}^{k} [1 - F_i(x)]$$

where $F_i(x)$ is the cumulative failure distribution for the i^{th} cause and k is the total number of causes. The function $F_i(x)$ is cause-specific. It reflects the probability of failure due to cause i alone and does not account for other causes of failure.

An alternative formulation is defined as follows:

$$F(x) = \sum_{i=1}^{k} \tilde{F}_i(x)$$

where each $\tilde{F}_i$ is a monotone increasing function with values in the interval [0, 1]. The function $\tilde{F}_i$ is called a subdistribution. This form of the aggregated failure distribution is used to predict proportions of failures that are associated with individual causes, accounting for failure due to other causes.

Cause Combination

The Cause Combination report lets you fit and compare different failure time distributions ($F_i(x)$) for the various causes. Different distributions are available based on selections that you have made in the launch window. Negative and zero-failure times are not allowed.

The default plot is based on a Linear scale and shows the following:

- Nonparametric estimates (Kaplan-Meier-Turnbull) for the uncensored data values and their confidence intervals. The confidence intervals are represented by horizontal blue lines.
- Fits of cumulative failure distributions ($F_i(x)$) to each of the causes. The initial distribution is the one that you selected in the launch window. If you selected Individual Best, the best distribution is computed for each group and these fits are shown. (This selection can be time-intensive.) If you selected Manual Pick, the initial Distribution is Nonparametric for all groups and nonparametric fits are shown. A legend appears to the right of the plot.
- The Aggregated cumulative failure distribution, $F(x)$, represented by a black line. This function is computed based on the selected cause distribution. If a nonparametric distribution is specified for a cause, the aggregated cumulative failure distribution extends only as far as the final time observation for that cause.

As you interact with the report, statistics for the aggregated model are re-evaluated.

The Cause Combination report contains these elements:

Scale Specifies the probability scale for the plot's vertical axis. "Change the Scale" illustrates how changing the scale affects the distribution fit.

Omit Check a box to remove the fit for the corresponding cause. Use this when a particular cause has been fixed. The Aggregated model updates to reflect the removal of the failures due to that cause. "Omit Competing Causes" illustrates the effect of omitting causes.

Cause Lists the causes in the Cause column.

Distribution Lists the available distributions for each cause. To change the distribution for a specific failure cause, select the distribution from the **Distribution** list. Click **Update Model** to show the new distribution fit on the plot. The Cause Summary report is also updated.

Count Gives the number of observations with the given failure cause.

Update Model Shows the selected distributions in the plot; updates the Cause Summary report with the selected models; adds the selected model as the most recent one in the Individual Causes report.

Competing Cause: Statistics

The Statistics report for Competing Cause contains the following reports:

- "Cause Summary Report"
- "Profilers"

Cause Summary Report

The Cause Summary report shows information for the fit defined by the current selections under Distribution in the Cause Summary report. The report shows the number of failures for each cause and the parameter estimates for the distribution fit to each failure cause. When you change distribution fits in the Cause Combination report and click Update Model, the Cause Summary report is updated.

The following information is provided:

- The Cause column shows either labels of causes or the censor code.
- The Counts column lists the number of failures for each cause.
- A numerical entry in the Counts column indicates that the cause has enough failure events to consider. A cause with fewer than two events is considered right censored. The column also identifies missing causes.
- The Distribution column specifies the selected distribution for each cause.
- Depending on the selected distributions, various columns display the parameters of the distributions:
  - The location column specifies location parameters for various distributions.
  - The scale column specifies scale parameters for various distributions.
  - The Weibull α and Weibull β columns show Weibull estimates of alpha and beta.
  - Other columns show parameters of other selected distributions.
  - A Convergence column appears if there are convergence issues.

The Cause Summary red triangle menu options enable you to save the probability, quantile, hazard, and density estimates for the aggregated failure distribution to the data table.

Profilers

The Distribution, Quantile, Hazard, and Density profilers help you visualize the aggregated failure distribution. The Distribution, Quantile, and Hazard profilers show 95% level confidence bands. See "Profilers".

Note: If a nonparametric distribution is specified for a cause, the Hazard and Density profilers are not provided. Also, confidence limits are not provided in the Distribution and Quantile profilers.

Individual Causes

The Individual Causes report contains Life Distribution - Failure Cause: <Distribution Name> reports for each of the individual causes. Each Life Distribution - Failure Cause: <Distribution Name> report shows plots and distributional fit statistics for the individual failure cause indicated in the report title.

Whenever you click Update Model in the Cause Combination report, any new cause distribution that you select is added to the Life Distribution - Failure Cause: <Distribution Name> report for that cause. In the Life Distribution - Failure Cause <Distribution Name> report, the following occur:

- The distribution is selected in the Compare Distributions report.
- The distribution is added to the Model Comparisons report.
- A Parametric report is added for that distribution.

Each Life Distribution - Failure Cause <Distribution Name> report is a Life Distribution: Statistics report as described in "Life Distribution: Statistics". However, all confidence intervals are Wald intervals. These intervals are not affected by the selection of Likelihood as the Confidence Interval Method in the launch window.

Available Distributions for Competing Cause Compare Distributions Reports

If you specify a Failure Cause in the launch window, you can specify which groupings of distributions and models you want to allow in the resulting Compare Distributions reports for Individual Causes. You can select ZI (Zero-Inflated), TH (Threshold), and DS (Defective Subpopulation) distributions. You can also select fixed parameter and Bayesian models.

Note: If you have disallowed any distributions in Preferences, these do not appear. Also, rules that govern which distributions appear for Life Distribution apply. See "Available Parametric Distributions".

Competing Cause Report Options

The Competing Cause red triangle menu contains the following options:

Tabbed Report Shows graphs and data on individual tabs rather than in the default outline style.

Tabbed Report for Individual Causes Shows the Life Distribution - Failure Cause: <Distribution Name> reports in tabs, rather than as a stack of Life Distribution reports.

Show Points Shows or hides data points in the Cause Combination plot. The Life Distribution platform uses the midpoint estimates of the step function to construct probability plots. When you deselect **Show Points**, the midpoint estimates are replaced by Kaplan-Meier estimates.

Show Subdistributions Shows the profiler for each individual cause subdistribution $\tilde{F}_i$. See "Individual Subdistribution Profiler for Cause".

Show Remaining Life Distribution Shows the remaining life distribution through the Distribution Profiler. This is conditional upon the unit surviving through a given time.

Mean Remaining Life Estimates the mean remaining life of a unit, given a survival time. In the Mean Remaining Life Calculator, enter a time value and press **Enter** to see the estimate and confidence limits. Click the plus sign to enter additional times. Click the minus sign to remove the most recent entry.

To obtain a confidence interval, select the **Configuration** option from the Mean Remaining Life Calculator. Check **Use bootstrap to construct confidence intervals**. Enter appropriate values, keeping in mind that computation can be time-intensive. See "Mean Remaining Life Calculator".

Export Bootstrap Results Appears only when a Bayesian model is selected and when **Update Model** has been applied.

Bootstrap Sample Size When Bayesian Estimates or Weibayes results are used for any cause, the confidence limits for aggregated functions that appear in the Distribution Profiler must be simulated using parametric bootstrap. Use this option to specify the number of samples to be used in the bootstrap. See "Specify a Bayesian Model for a Cause".

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Individual Subdistribution Profiler for Cause

For a given cause, the Individual Subdistribution Profiler for Cause shows the estimated probability of failure, $\tilde{F}_i$, from that cause at time t. The estimated probability takes into account failures from competing causes. See "Competing Cause Model".

To show a profiler of the subdistribution for each cause, click the Competing Cause red triangle and select **Show Subdistributions**. The Individual Subdistribution Profiler for Cause report appears beneath the other profilers. It consists of a profiler and a calculator.

Note: When you select **Show Subdistributions**, the Cause Combination plot is updated to show the subdistribution functions for all causes.

Select a Cause from the list to the right of the profiler to see its profiler. Options that apply to the profiler are provided in the Individual Subdistribution Profiler for Cause red triangle menu. See "Profilers".

Use the Calculator panel to find values of the subdistribution functions for all causes at one or more Time to Event values. Enter a value for the Time to Event variable. When you press Enter (or click outside the text box), values for each of the causes are updated. To add a Time to Event value, click the plus sign. To remove the most recent value, click the minus sign.

Life Distribution - Compare Groups Report

Tip: If you find that the report window is getting too long, select **Tabbed Report** from the red triangle menu next to Life Distribution - Compare Groups.

If you selected the Compare Groups tab in the launch window, the Life Distribution - Compare Groups report appears. This report compares different groups using a single specified distribution. For example, you might compare Weibull fits for components grouped by supplier. In contrast, the Life Distribution tab compares several fitted distributions for a single group.

You can compare the CDF, Quantile, Hazard, and Density functions. You can also consolidate the probability and quantile predictions of all groups.

For an example of this report, see "Examine the Same Distribution across Groups".

The Compare Groups report can contain the following content and options:

- "Compare Distributions" (no Distribution Profiler)
- "Compare Groups: Statistics"
- "Individual Group"

- "Life Distribution - Compare Groups Report Options"

Compare Groups: Statistics

The Statistics report for Compare Groups contains the following reports:

- "Summary"
- "Group Homogeneity Tests"
- "Model Comparisons"
- "Parameter Estimates"

Summary

The Summary report contains a row for each group and for the combined data. Each row shows the number of units that failed and the numbers that were left, interval, or right censored. Each row also gives the corresponding mean and standard error. For more information about how the mean and standard error are computed, see "Statistical Reports for Survival Analysis".

Group Homogeneity Tests

The Group Homogeneity Tests report contains results of three tests for equality of the hazard functions across groups. The tests differ in how the observations are weighted over time.

Log-Rank The Log-Rank test uses equal weights for all time points.

Peto-Peto The Peto-Peto test uses the survival proportion at each time point for weights.

Wilcoxon The generalized Wilcoxon test uses the number at risk at each time point for weights. This test is referred to as the Gehan test in Klein and Moeschberger (1997).

For each of the above tests, the report includes the chi-square approximation, the associated degrees of freedom, and the *p*-value for each test. A small *p*-value suggests that the groups differ. For more information about these tests and comparisons of failure curves, see Klein and Moeschberger (1997, ch. 7).

Parameter Estimates

The Parameter Estimates outline contains reports entitled Parametric Estimate - <Distribution Name> for each distribution that is fit. For each group, the Parametric Estimate - <Distribution Name> report gives the distribution's parameter estimates and their 95% level confidence intervals. The confidence level for these intervals is determined by the Change Confidence Level option in the Life Distribution - Compare Groups menu.

Individual Group

The tabs within the Individual Group report contain Life Distribution reports for each individual group. For more information about these reports, see "Life Distribution Report" and "Life Distribution Report Options".

Life Distribution - Compare Groups Report Options

Many of the options in the Compare Groups red triangle menu can also be found in the Life Distribution red triangle menu. See "Life Distribution Report Options".

However, the following options are specific to Compare Groups:

Show Quantile Functions Shows or hides the Compare Quantile report. Select a distribution. For each group, a curve is plotted showing the estimated quantiles for the time variable. Confidence bands are displayed. A legend is shown to the right of the plot. Only one distribution can be specified at a time.

Show Hazard Functions Shows or hides the Compare Hazard report. Select a distribution. For each group, a curve is plotted showing the hazard function. Confidence bands are displayed. A legend is shown to the right of the plot. Only one distribution can be specified at a time.

Show Density Functions Shows or hides the Compare Density report. Select a distribution. For each group, the density function is displayed. A legend is shown to the right of the plot. Only one distribution can be specified at a time.

Estimate Probability Adds an Estimate Probability report corresponding to the most recently selected distribution under Compare Distribution. Enter a value for the Time to Event variable in the text box and press **Enter**. To add a Time to Event value, click the plus sign. To remove the most recent value, click the minus sign. See "Estimate Probability Report".

Estimate Quantile (Appears only if Compare Quantile is selected.) Adds an Estimate Quantile report corresponding to the most recently selected distribution under Compare Quantile. Enter a value for the probability of interest in the text box and press **Enter**. To add a Prob value, click the plus sign. To remove the most recent value, click the minus sign. For each group and probability value, the Time to Event quantile and 95% Wald and Likelihood confidence intervals are shown.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Estimate Probability Report

For each group and Time to Event value, this report provides the following:

Midpoint Estimate Midpoint-adjusted Kaplan-Meier estimate of failure by the specified time.

Lower 95%, Upper 95% Pointwise 95% confidence intervals for the probability of failure by the specified time.

Simultaneous Lower 95% (Nair), Simultaneous Upper 95% (Nair) Simultaneous 95% confidence intervals for the probability of failure by the specified time. See Nair (1984) and Meeker and Escobar (1998).

Survival Probability Midpoint-adjusted estimate of survival beyond the specified time.

Survival Probability Lower 95%, Upper 95% Pointwise 95% confidence intervals for the probability of survival beyond the specified time.

Survival Probability Simultaneous Lower 95% (Nair), Simultaneous Upper 95% (Nair)

Simultaneous 95% confidence intervals for the probability of survival beyond the specified time. See Nair (1984) and Meeker and Escobar (1998).

Additional Examples of the Life Distribution Platform

- "Omit Competing Causes"
- "Change the Scale"
- "Examine the Same Distribution across Groups"
- "Weibayes Estimates"
- "Fit Mixture Example"
- "Fit Competing Risk Mixture Example"

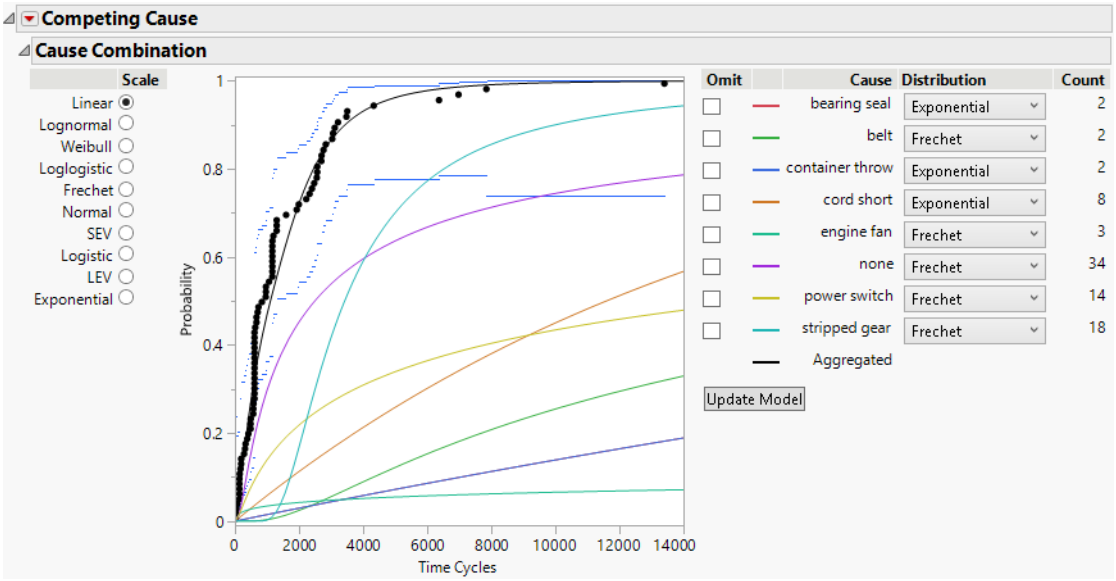
Omit Competing Causes

This example illustrates how to decide on the best fit for competing causes.

1. Select **Help > Sample Data Library** and open Reliability/Blenders.jmp.
2. Select **Analyze** > **Reliability and Survival** > **Life Distribution**.
3. Select Time Cycles and click **Y, Time to Event**.
4. Select Causes and click **Failure Cause**.
5. Select Censor and click **Censor**.
6. Select **Individual Best** as the Distribution.
7. Make sure that **AICc** is the Comparison Criterion.
8. Click **OK**.

   On the Competing Cause report, JMP shows the best distribution fit for each failure cause.

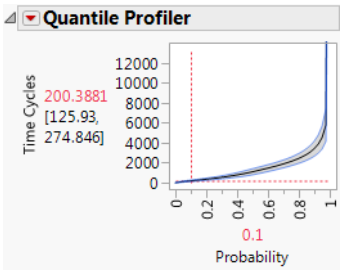
Figure 3.11 Initial Competing Cause Report



9. In the Quantile Profiler, type 0.1 for the probability.

   The estimated time by which 10% of the failures occur is 200.

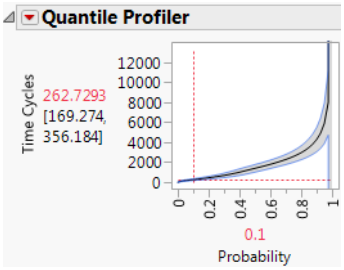
Figure 3.12 Estimated Failure Time for 10% of the Units



10. Select **Omit** for bearing seal, belt, container throw, cord short, and engine fan (the causes with the fewest failures).

    The estimated time by which 10% of the failures occur is now 263.

Figure 3.13 Updated Failure Time



When power switch and stripped gear are the only causes of failure, the estimated time by which 10% of the failures occur increases by approximately 31%.

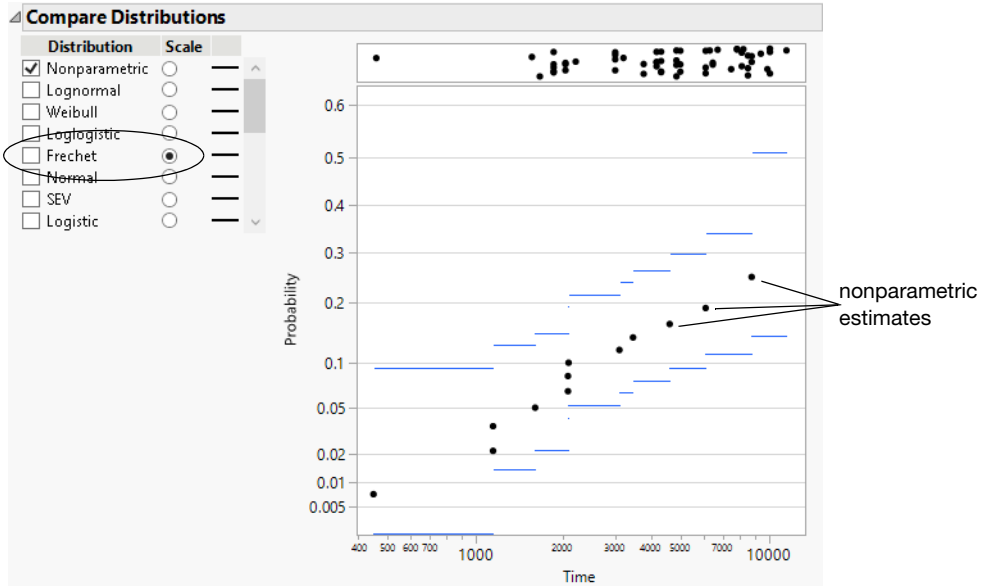
Change the Scale

In the initial Compare Distributions report, the probability and time axes are linear. But suppose that you want to see distribution estimates on a Fréchet scale.[1]

1. Follow step 1 through step 5 in "Example of the Life Distribution Platform".
2. In the Compare Distributions report, select **Fréchet** in the Scale column.
3. Click the Life Distribution red triangle and select **Interval Type > Pointwise**.

1. Using different scales is sometimes referred to as drawing the distribution on different types of *probability paper*.

Figure 3.14 Nonparametric Estimates with a Fréchet Probability Scale

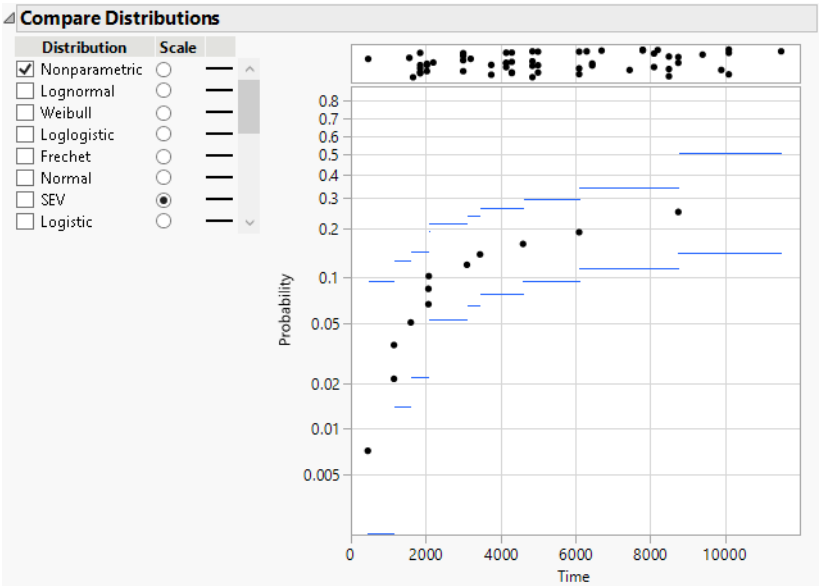


Using a Fréchet scale, the nonparametric estimates approximate a straight line, meaning that a Fréchet fit might be reasonable.

4. Select **SEV** in the Scale column.

The nonparametric estimates no longer approximate a straight line. You now know that the SEV distribution is not appropriate.

Figure 3.15 Nonparametric Estimates with a SEV Probability Scale

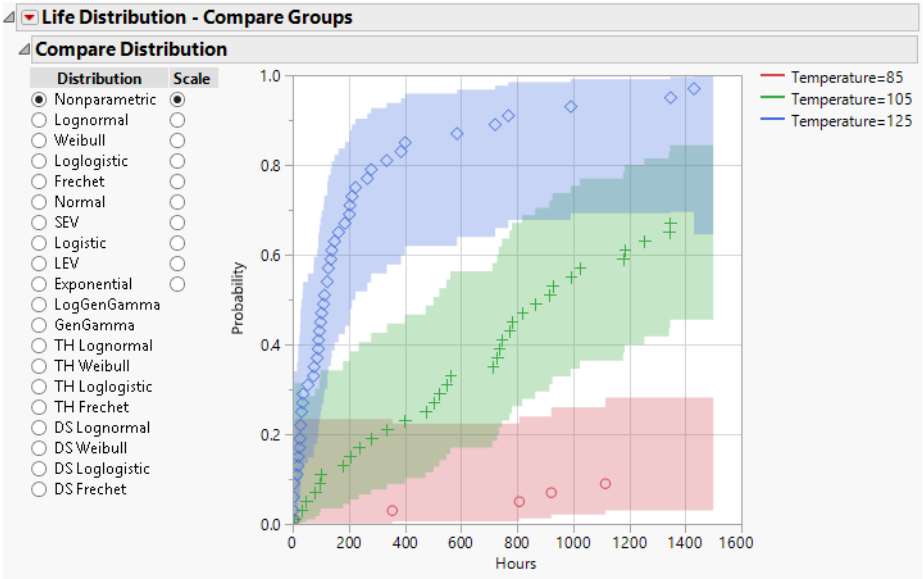


Examine the Same Distribution across Groups

Suppose you want to compare the same distribution across different groups. You want to examine estimates of failure probabilities for a single type of capacitor operating at three different temperatures.

1. Select **Help > Sample Data Library** and open Reliability/Capacitor ALT.jmp.
2. Select **Analyze** > **Reliability and Survival** > **Life Distribution**.
3. Click the Compare Groups tab.
4. Select Hours and click **Y, Time to Event**.
5. Select Temperature and click **Grouping**.
6. Select Censor and click **Censor**.
7. Select Freq and click **Freq**.
8. Click **OK**.

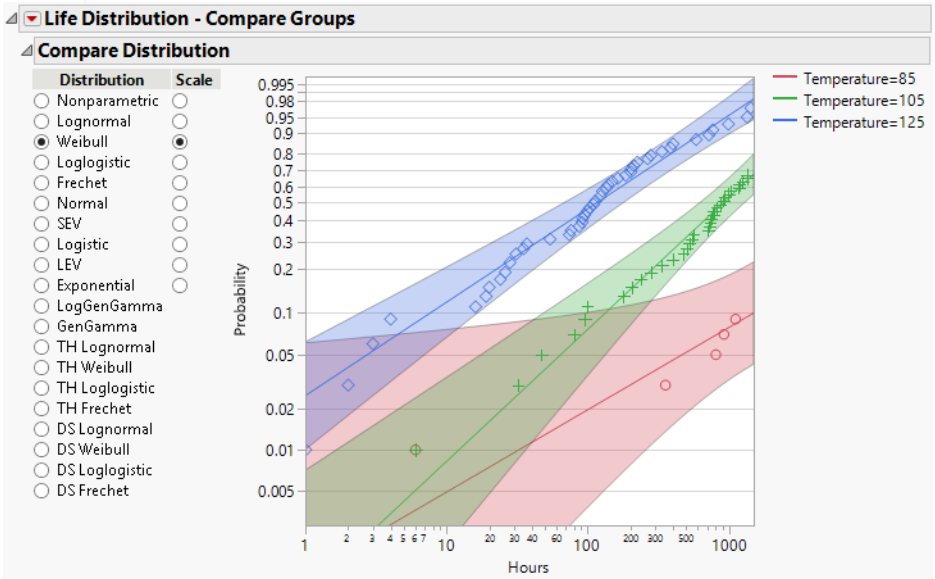
Figure 3.16 Compare Distribution for Groups



The default graph shows the nonparametric estimates. At a higher temperature, the capacitor has a higher probability of failure. You want to try fitting a parametric distribution.

9. Select **Weibull** for Distribution and Scale.

Figure 3.17 Compare Weibull Distribution for Groups



When plotted against a Weibull probability scale, the points come close to following three lines. This indicates that a Weibull distribution provides a reasonable fit for each of the Temperature groups.

Weibayes Estimates

There are two possible ways to obtain a Weibayes analysis:

- You have no failures (all observations are right-censored) and the preference **Weibayes Only for Zero Failure Data** is checked. Then the Weibayes report appears. See "Weibayes Example for Data with No Failures".
- You have few failures. A full Life Distribution report is presented. Fit a Weibull distribution. In the Parametric Estimate - Weibull report, select the Fix Parameter option. Then select the Weibayes option in the Fixed Parameter report. See "Weibayes Example for Data with One Failure".

Weibayes Example for Data with No Failures

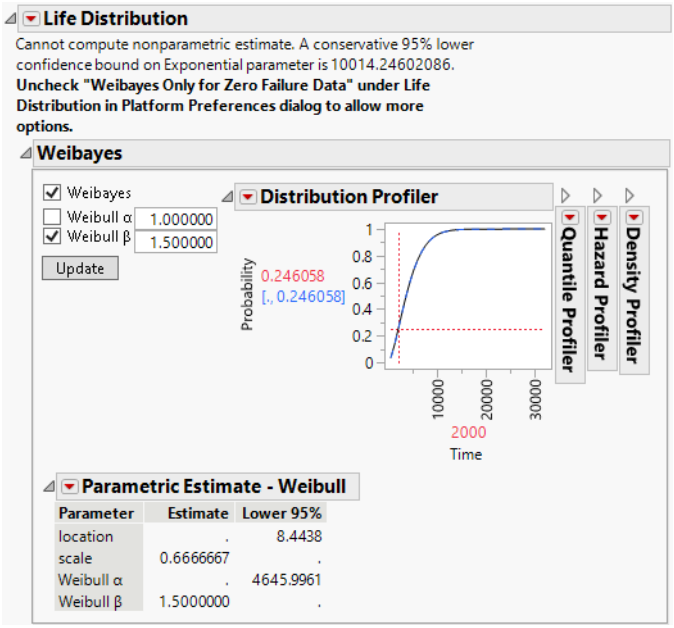
You have data for a product that is mostly reliable. Thirty were tested for 1,000 hours with no failures occurring. You want to predict the failure probability at 2,000 hours.

1. Select **Help > Sample Data Library** and open Reliability/Weibayes No Failures.jmp.
2. Select **Analyze** > **Reliability and Survival** > **Life Distribution**.
3. Select Time and click **Y, Time to Event.**
4. Select Censor and click **Censor**.
5. Select Freq and click **Freq**.
6. Select **Likelihood** as the Confidence Interval Method.
7. Click **OK.**

   A special Life Distribution report appears. **Weibayes** and **Weibull beta** should be selected.
8. Type 1.5 as the known Weibull β value.

   The value 1.5 is considered appropriate for this example.
9. Click **Update**.
10. In the Distribution Profiler, type 2000 for Time.

Figure 3.18 Life Distribution Report for Zero Failures



From the Distribution Profiler, you can see that at 2,000 hours, the conservative probability is 24.6058%. That means that the one-tailed conservative 95% confidence limit for the failure probability is 24.6058%.

Weibayes Example for Data with One Failure

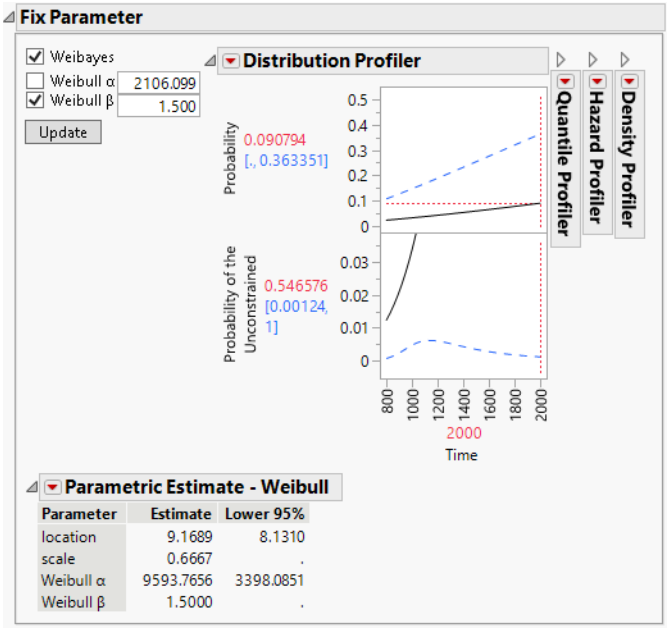
Suppose you have the same data, but this time, one failure occurred at 800 hours. Again, you want to predict the failure probability at 2,000 hours.

1. Select **Help > Sample Data Library** and open Reliability/Weibayes One Failure.jmp.
2. Select **Analyze** > **Reliability and Survival** > **Life Distribution**.
3. Select Time and click **Y, Time to Event.**
4. Select Censor and click **Censor**.
5. Select Freq and click **Freq**.
6. Select **Likelihood** as the Confidence Interval Method.
7. Click **OK.**

   The Life Distribution report appears.
8. Select the **Weibull** distribution in the Compare Distributions plot.
9. Click the red triangle next to Parametric Estimate - Weibull and select **Fix Parameter**.
10. Select **Weibayes** and **Weibull beta** in the Fix Parameter report.

11. Type 1.5 as the known Weibull β value.
12. Click **Update**.
13. In the Distribution Profiler, type 2000 for Time.
14. Hover over the top of the *Y* axis. The cursor becomes a hand. Drag the axis downward until it reaches 0.5 as the top number.

Figure 3.19 Life Distribution Report for One Failure



In the Distribution Profiler, the solid line shows the MLE. The dashed line shows the Weibayes conservative limit. You can see that at 2,000 hours, the conservative probability is 36.3351%. That means that the one-tailed conservative 95% confidence limit for the failure probability is 36.3351%.

Fit Mixture Example

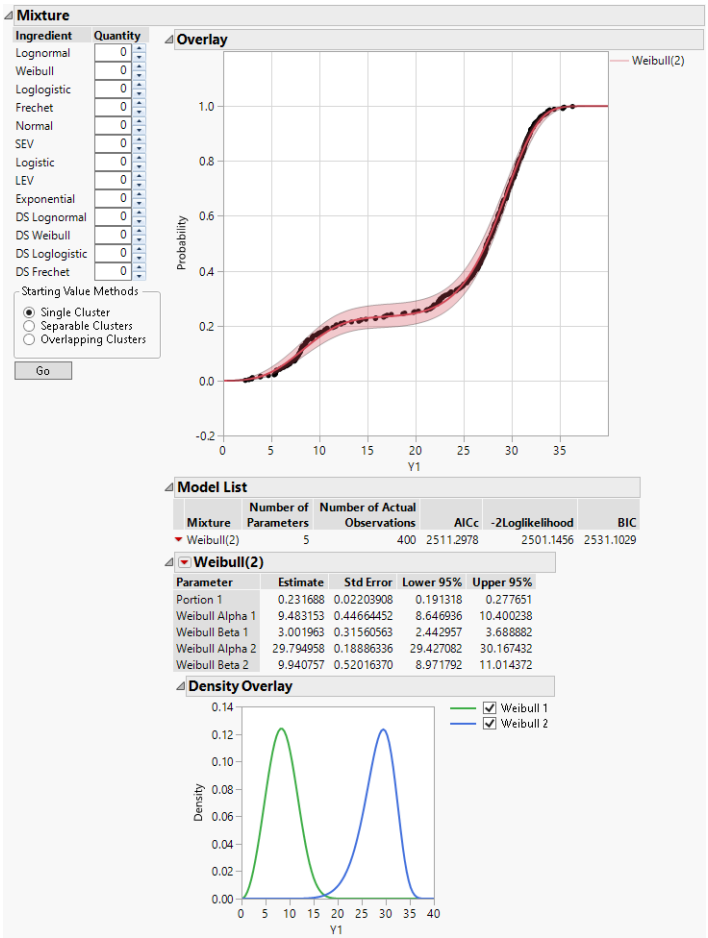
In this example, you fit two mixture distributions and then identify observations belonging to one of the clusters for the second mixture.

Fit Two Mixture Distributions

1. Select **Help > Sample Data Library** and open Reliability/Mixture Demo.jmp.
2. Select **Analyze > Reliability and Survival > Life Distribution**.
3. Select Y1 and click **Y, Time to Event**.

4. Click **OK**.
5. Click the Life Distribution red triangle and select **Fit Mixture**.
6. Type 2 in the **Quantity** box next to **Weibull**.
7. Select **Separable Clusters** in the Starting Value Methods panel.
8. Click **Go**.

Figure 3.20 Fit Mixture for Weibull (2)

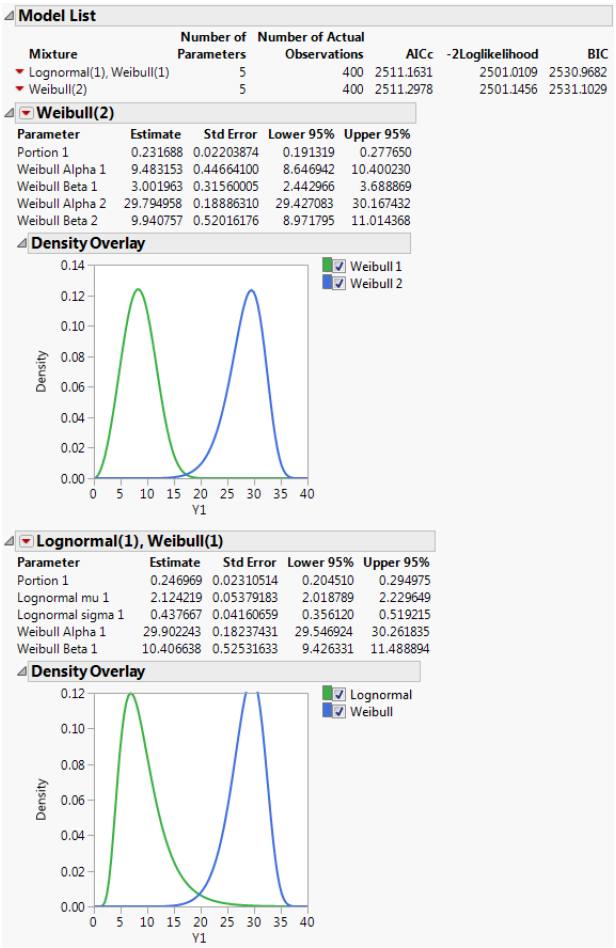


JMP fits a mixture model consisting of two Weibull components. Portion 1 is estimated as 0.231688, indicating that approximately 23% of observations have the Weibull distribution with alpha = 9.483153 and beta = 3.001963. The remaining 77% are estimated to come from the second Weibull distribution.

To compare this model to another, you can change the Ingredient selections and the Quantity of components.

9. Type 1 next to **Lognormal** and 1 next to **Weibull**.
10. Click **Go**.

Figure 3.21 Fit Mixture for Lognormal(1), Weibull(1)



The Overlay plot is updated to show both mixture models. The plots and statistics in the Model List indicate that the Lognormal(1), Weibull(1) mixture seems to give a fit that is very similar to the Weibull(2) mixture.

Identify Observations Belonging to a Cluster

1. Click the red triangle next to Lognormal(1), Weibull(1) and select **Save Predictions**.

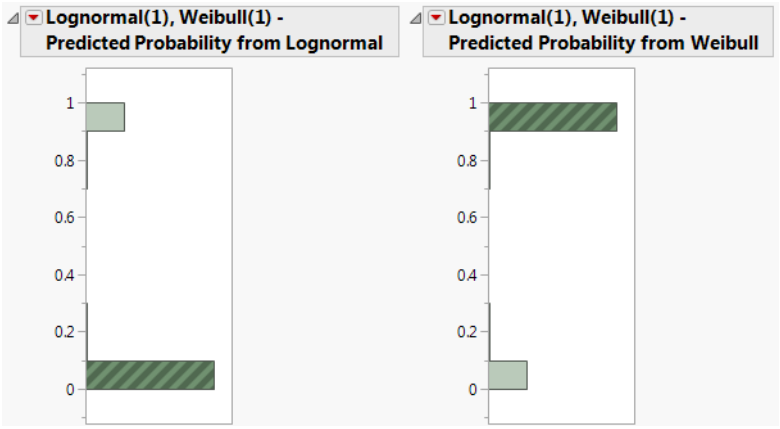
   Two columns are added to the data table:

   – Lognormal(1), Weibull(1) - Predicted Probability from Lognormal

– Lognormal(1), Weibull(1) - Predicted Probability from Weibull

2. Select **Analyze > Distribution**.
3. Select the two new columns from the Select Columns list and click **Y, Columns**.
4. Check **Histograms Only**.
5. Click **OK**.
6. In the histogram for Lognormal(1), Weibull(1) - Predicted Probability from Weibull, click in the bar corresponding to the value near 1.

Figure 3.22 Histograms for Mixture Probabilities



In the data table, the 297 corresponding rows are selected. These are the observations that are likely to have come from the Weibull distribution with parameters alpha = 29.90 and beta = 10.41.

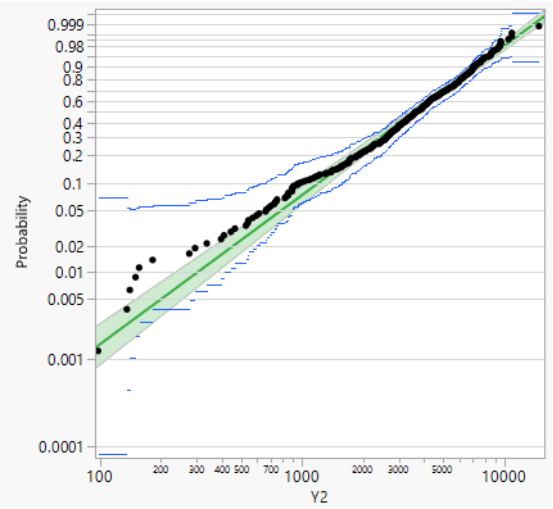
Fit Competing Risk Mixture Example

In this example, you compare the fit of a single Weibull distribution to a competing risk mixture distribution with two Weibull fits.

1. Select **Help > Sample Data Library** and open Reliability/Mixture Demo.jmp.
2. Select **Analyze > Reliability and Survival > Life Distribution**.
3. Select Y2 and click **Y, Time to Event**.
4. Click **OK**.
5. In the Compare Distributions report, select **Weibull** distribution and the corresponding **Scale** radio button.

   A probability plot for a single Weibull distribution fit appears. Note that the fit is not very good in the lower part of the range of Y2.

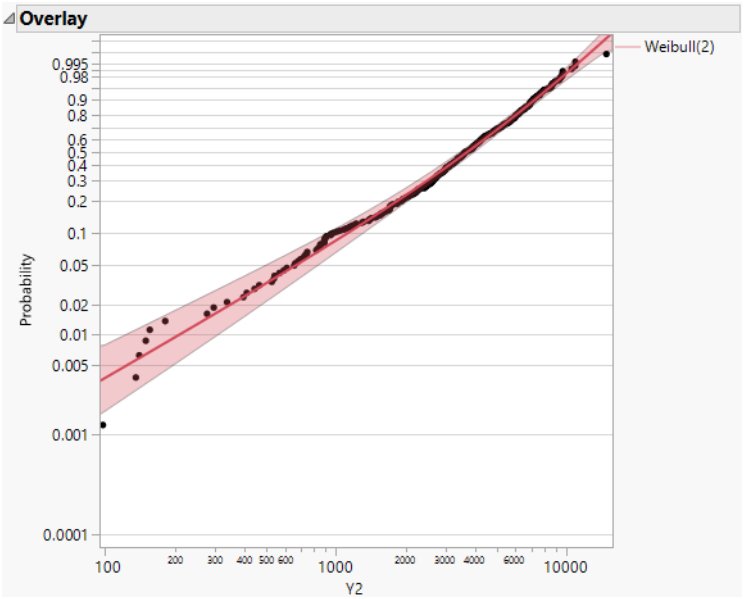
Figure 3.23 Weibull Distribution Fit



6. Click the Life Distribution red triangle and select **Fit Competing Risk Mixture**.
7. Scroll down to the Competing Risk Mixture report. Next to **Weibull**, type 2 in the **Quantity** box.
8. Click **Go**.
9. Scroll up to the Compare Distributions report. In the probability plot, right-click the vertical axis and select **Edit > Copy Axis Settings**.
10. Scroll down to the Competing Risk Mixture report. In the overlay plot, right-click the vertical axis and select **Edit > Paste Axis Settings**.
11. Do the same for the horizontal axis. In the probability plot, right-click the horizontal axis and select **Edit > Copy Axis Settings**.
12. In the overlay plot, right-click the horizontal axis and select **Edit > Paste Axis Settings**.

    The probability plot for the Weibull(2) distribution fit appears. Note that the mixture of two Weibull distributions helps better capture the distribution in the lower part of the range of Y2.

Figure 3.24 Weibull(2) Competing Risk Mixture Distribution Fit



Statistical Details for the Life Distribution Platform

This section covers the following topics:

- The distributions used in the Life Distribution platform. See "Distributions".
- Technical information about the Competing Cause report. See "Competing Cause Details".
- Technical information about median rank regression. See "Notes Regarding Median Rank Regression".

Distributions

This section provides details for the distributional fits in the Life Distribution platform. Meeker and Escobar (1998, ch. 2-5) is an excellent source of theory, application, and discussion for both the nonparametric and parametric details that follow.

Estimation and Confidence Intervals

The parameters of all distributions, unless otherwise noted, are estimated using maximum likelihood estimates (MLEs). The only exceptions are the threshold distributions. If the smallest observation is an exact failure, then this observation is treated as interval-censored with a small interval. The parameter estimates are the MLEs estimated from this slightly modified data set. Without this modification, the likelihood can be unbounded, so an MLE might not exist. This approach is similar to that described in Meeker and Escobar (1998, p. 275), except that only the smallest exact failure is censored. This is the minimal change to the data that guarantees boundedness of the likelihood function.

The Life Distribution platform offers two methods for calculating confidence intervals for the distribution parameters. These methods are labeled as Wald or Likelihood and can be selected in the launch window for the Life Distribution platform. Wald confidence intervals are used as the default setting. The computations for the confidence intervals for the cumulative distribution function (cdf) start with Wald confidence intervals on the standardized variable. Next, the intervals are transformed to the cdf scale (Nelson 1982, pp. 332–333 and pp. 346-347). The confidence intervals given in the other graphs and profilers are transformed Wald intervals (Meeker and Escobar 1998, ch. 7). Joint confidence intervals for the parameters of a two-parameter distribution are shown in the log-likelihood contour plots. They are based on approximate likelihood ratios for the parameters (Meeker and Escobar 1998, ch. 8).

Nonparametric Fit

A nonparametric fit describes the basic curve of a distribution. For data with no censoring (failures only) and for data where the observations consist of both failures and right-censoring, JMP uses Kaplan-Meier estimates. For mixed, interval, or left censoring, JMP uses Turnbull estimates. When your data set contains only right-censored data, the Nonparametric Estimate report indicates that the nonparametric estimate cannot be calculated.

The Life Distribution platform uses the midpoint estimates of the step function to construct probability plots. The midpoint estimate is halfway between (or the average of) the current and previous Kaplan-Meier estimates.

Parametric Distributions

Parametric distributions provide a more concise distribution fit than nonparametric distributions. The estimates of failure-time distributions are also smoother. Parametric models are also useful for extrapolation (in time) to the lower or upper tails of a distribution.

Note: Many distributions in the Life Distribution platform are parameterized by location and scale. For lognormal fits, the median is also provided. A threshold parameter is also included in threshold distributions. Location corresponds to μ, scale corresponds to σ, and threshold corresponds to γ.

Lognormal

Lognormal distributions are used commonly for failure times when the range of the data is several powers of 10. This distribution is often considered as the multiplicative product of many small positive identically independently distributed random variables. It is reasonable when the log of the data values appears normally distributed. Examples of data appropriately modeled by the lognormal distribution include hospital cost data, metal fatigue crack growth, and the survival time of bacteria subjected to disinfectants. The pdf curve is usually characterized by strong right-skewness. The lognormal pdf and cdf are:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma}\phi_{nor}\left[\frac{\log(x)-\mu}{\sigma}\right], \quad x > 0$$

$$F(x;\mu,\sigma) = \Phi_{nor}\left[\frac{\log(x)-\mu}{\sigma}\right]$$

where

$$\phi_{nor}(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$$

and

$$\Phi_{nor}(z) = \int_{-\infty}^{z}\phi_{nor}(w)dw$$

are the pdf and cdf, respectively, for the standardized normal, or nor($\mu = 0$, $\sigma = 1$) distribution.

Weibull

The Weibull distribution can be used to model failure time data with either an increasing or a decreasing hazard rate. It is used frequently in reliability analysis because of its tremendous flexibility in modeling many different types of data, based on the values of the shape parameter, β. This distribution has been successfully used for describing the failure of electronic components, roller bearings, capacitors, and ceramics. Various shapes of the Weibull distribution can be revealed by changing the scale parameter, α, and the shape parameter, β. The Weibull pdf and cdf are commonly represented as follows:

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha^{\beta}} x^{(\beta-1)} \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]; \qquad x > 0, \alpha > 0, \beta > 0$$

$$F(x;\alpha,\beta) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]$$

where α is a scale parameter, and β is a shape parameter. The Weibull distribution is particularly versatile because it reduces to an exponential distribution when β = 1. An alternative parameterization commonly used in the literature and in JMP is to use σ as the scale parameter and μ as the location parameter. These are easily converted to an α and β parameterization using the following definitions:

$$\alpha = \exp(\mu)$$

and

$$\beta = \frac{1}{\sigma}$$

The pdf and the cdf of the Weibull distribution are also expressed as a log-transformed smallest extreme value distribution (SEV). This uses a location scale parameterization, with μ = log(α) and σ = 1/β,

$$f(x;\mu,\sigma) = \frac{1}{x\sigma}\phi_{\text{sev}}\left[\frac{\log(x)-\mu}{\sigma}\right], \qquad x > 0, \sigma > 0$$

$$F(x;\mu,\sigma) = \Phi_{\text{sev}}\left[\frac{\log(x)-\mu}{\sigma}\right]$$

where

$$\phi_{\text{sev}}(z) = \exp[z - \exp(z)]$$

and

$$\Phi_{sev}(z) = 1 - \exp[-\exp(z)]$$

are the pdf and cdf, respectively, for the standardized smallest extreme value ($\mu = 0$, $\sigma = 1$) distribution.

Loglogistic

The pdf of the loglogistic distribution is similar in shape to the lognormal distribution but has heavier tails. It is often used to model data exhibiting non-monotonic hazard functions, such as cancer mortality and financial wealth. The loglogistic pdf and cdf are:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma}\phi_{logis}\left[\frac{\log(x)-\mu}{\sigma}\right]$$

$$F(x;\mu,\sigma) = \Phi_{logis}\left[\frac{\log(x)-\mu}{\sigma}\right]$$

where

$$\phi_{logis}(z) = \frac{\exp(z)}{[1+\exp(z)]^2}$$

and

$$\Phi_{logis}(z) = \frac{\exp(z)}{[1+\exp(z)]} = \frac{1}{1+\exp(-z)}$$

are the pdf and cdf, respectively, for the standardized logistic or logis distribution($\mu = 0$, $\sigma = 1$).

Fréchet

The Fréchet distribution is known as a log-largest extreme value distribution or sometimes as a Fréchet distribution of maxima when it is parameterized as the reciprocal of a Weibull distribution. This distribution is commonly used for financial data. The pdf and cdf are:

$$f(x;\mu,\sigma) = \exp\left[-\exp\left(-\frac{\log(x)-\mu}{\sigma}\right)\right]\exp\left(-\frac{\log(x)-\mu}{\sigma}\right)\frac{1}{x\sigma}$$

$$F(x;\mu,\sigma) = \exp\left[-\exp\left(-\frac{\log(x)-\mu}{\sigma}\right)\right]$$

and are more generally parameterized as follows:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma}\phi_{lev}\left[\frac{\log(x)-\mu}{\sigma}\right]$$

$$F(x;\mu,\sigma) = \Phi_{lev}\left[\frac{\log(x)-\mu}{\sigma}\right]$$

where

$$\phi_{lev}(z) = \exp[-z-\exp(-z)]$$

and

$$\Phi_{lev}(z) = \exp[-\exp(-z)]$$

are the pdf and cdf, respectively, for the standardized largest extreme value LEV($\mu = 0$, $\sigma = 1$) distribution.

Normal

The normal distribution is the most widely used distribution in most areas of statistics because of its relative simplicity and the ease of applying the central limit theorem. However, it is rarely used in reliability. It is most useful for data where $\mu > 0$ and the coefficient of variation (σ / μ) is small. Because the hazard function increases with no upper bound, it is particularly useful for data exhibiting wear-out failure. Examples include incandescent light bulbs, toaster heating elements, and mechanical strength of wires. The pdf and cdf are:

$$f(x;\mu,\sigma) = \frac{1}{\sigma}\phi_{nor}\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < x < \infty$$

$$F(x;\mu,\sigma) = \Phi_{nor}\left(\frac{x-\mu}{\sigma}\right)$$

where

$$\phi_{nor}(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$$

and

$$\Phi_{nor}(z) = \int_{-\infty}^{z} \phi_{nor}(w)dw$$

are the pdf and cdf, respectively, for the standardized normal, or nor($\mu = 0$, $\sigma = 1$) distribution.

Smallest Extreme Value (SEV)

This non-symmetric (left-skewed) distribution is useful in two cases. The first case is when the data indicate a small number of weak units in the lower tail of the distribution (the data indicate the smallest number of many observations). The second case is when σ is small relative to μ, because probabilities of being less than zero, when using the SEV distribution, are small. The smallest extreme value distribution is useful to describe data whose hazard rate becomes larger as the unit becomes older. Examples include human mortality of the aged and rainfall amounts during a drought. This distribution is sometimes referred to as a Gumbel distribution. The pdf and cdf are:

$$f(x;\mu,\sigma) = \frac{1}{\sigma}\phi_{sev}\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < \mu < \infty, \quad \sigma > 0$$

$$F(x;\mu,\sigma) = \Phi_{sev}\left(\frac{x-\mu}{\sigma}\right)$$

where

$$\phi_{sev}(z) = \exp[z - \exp(z)]$$

and

$$\Phi_{sev}(z) = 1 - \exp[-\exp(z)]$$

are the pdf and cdf, respectively, for the standardized smallest extreme value, SEV($\mu = 0$, $\sigma = 1$) distribution.

Logistic

The logistic distribution has a shape similar to the normal distribution, but with longer tails. The logistic distribution is often used to model life data when negative failure times are not an issue. Logistic regression models for a binary or ordinal response assume the logistic distribution as the latent distribution. The pdf and cdf are:

$$f(x;\mu,\sigma) = \frac{1}{\sigma}\phi_{logis}\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < \mu < \infty \text{ and } \sigma > 0$$

$$F(x;\mu,\sigma) = \Phi_{logis}\left(\frac{x-\mu}{\sigma}\right)$$

where

$$\phi_{logis}(z) = \frac{\exp(z)}{[1 + \exp(z)]^2}$$

and

$$\Phi_{\text{logis}}(z) = \frac{\exp(z)}{[1+\exp(z)]} = \frac{1}{1+\exp(-z)}$$

are the pdf and cdf, respectively, for the standardized logistic or logis distribution ($\mu = 0$, $\sigma = 1$).

Largest Extreme Value (LEV)

This right-skewed distribution can be used to model failure times if σ is small relative to $\mu > 0$. This distribution is not commonly used in reliability but is useful for estimating natural extreme phenomena, such as a catastrophic flood heights or extreme wind velocities. The pdf and cdf are:

$$f(x;\mu,\sigma) = \frac{1}{\sigma}\phi_{\text{lev}}\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < \mu < \infty \text{ and } \sigma > 0$$

$$F(x;\mu,\sigma) = \Phi_{\text{lev}}\left(\frac{x-\mu}{\sigma}\right)$$

where

$$\phi_{\text{lev}}(z) = \exp[-z-\exp(-z)]$$

and

$$\Phi_{\text{lev}}(z) = \exp[-\exp(-z)]$$

are the pdf and cdf, respectively, for the standardized largest extreme value LEV($\mu = 0$, $\sigma = 1$) distribution.

Exponential

Both one- and two-parameter exponential distributions are used in reliability. The pdf and cdf for the two-parameter exponential distribution are:

$$f(x;\theta,\gamma) = \frac{1}{\theta}\exp\left(-\frac{x-\gamma}{\theta}\right), \quad \theta > 0$$

$$F(x;\theta,\gamma) = 1-\exp\left(-\frac{x-\gamma}{\theta}\right)$$

where θ is a scale parameter and γ is both the threshold and the location parameter. Reliability analysis frequently uses the one-parameter exponential distribution, with $\gamma = 0$. The exponential distribution is useful for describing failure times of components exhibiting wear-out far beyond their expected lifetimes. This distribution has a constant failure rate, which means that for small time increments, failure of a unit is independent of the unit's age. The exponential distribution should not be used for describing the life of mechanical components that can be exposed to fatigue, corrosion, or short-term wear. This distribution is, however, appropriate for modeling certain types of robust electronic components. It has been used successfully to describe the life of insulating oils and dielectric fluids (Nelson 1990, p. 53).

Log Generalized Gamma (LogGenGamma)

The log generalized gamma distribution contains the SEV, LEV, and Normal. The pdf and cdf are:

$$f(x;\mu,\sigma,\lambda) = \begin{cases} \frac{|\lambda|}{\sigma}\phi_{lg}[\lambda\omega + \log(\lambda^{-2});\lambda^{-2}] & \text{if } \lambda \neq 0 \\ \frac{1}{\sigma}\phi_{nor}(\omega) & \text{if } \lambda = 0 \end{cases}$$

$$F(x;\mu,\sigma,\lambda) = \begin{cases} \Phi_{lg}[\lambda\omega + \log(\lambda^{-2});\lambda^{-2}] & \text{if } \lambda > 0 \\ \Phi_{nor}(\omega) & \text{if } \lambda = 0 \\ 1 - \Phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda < 0 \end{cases}$$

where $-\infty < x < \infty$, $\omega = [x - \mu]/\sigma$, and

$$-\infty < \mu < \infty, \quad -12 < \lambda < 12, \text{ and } \sigma > 0.$$

Note that

$$\phi_{lg}(z;\kappa) = \frac{1}{\Gamma(\kappa)}\exp[\kappa z - \exp(z)]$$

$$\Phi_{lg}(z;\kappa) = \Gamma_I[\exp(z);\kappa]$$

are the pdf and cdf, respectively, for the log-gamma variable and $\kappa > 0$ is a shape parameter. The standardized distributions above are dependent upon the shape parameter κ.

Note: In JMP, the shape parameter, λ, for the generalized gamma distribution is bounded between [-12,12] to provide numerical stability.

Extended Generalized Gamma (GenGamma)

The extended generalized gamma distribution can include many other distributions as special cases, such as the generalized gamma, Weibull, lognormal, Fréchet, gamma, and exponential. It is particularly useful for cases with little or no censoring. This distribution has been successfully modeled for human cancer prognosis. The pdf and cdf are:

$$f(x;\mu,\sigma,\lambda) = \begin{cases} \frac{|\lambda|}{x\sigma}\phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda \neq 0 \\ \frac{1}{x\sigma}\phi_{nor}(\omega) & \text{if } \lambda = 0 \end{cases}$$

$$F(x;\mu,\sigma,\lambda) = \begin{cases} \Phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda > 0 \\ \Phi_{nor}(\omega) & \text{if } \lambda = 0 \\ 1 - \Phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda < 0 \end{cases}$$

where $x > 0$, $\omega = [\log(x) - \mu]/\sigma$, and

$$-\infty < \mu < \infty, \quad -12 < \lambda < 12, \text{ and } \sigma > 0.$$

Note that

$$\phi_{lg}(z;\kappa) = \frac{1}{\Gamma(\kappa)}\exp[\kappa z - \exp(z)]$$

$$\Phi_{lg}(z;\kappa) = \Gamma_I[\exp(z);\kappa]$$

are the pdf and cdf, respectively, for the standardized log-gamma variable and $\kappa > 0$ is a shape parameter.

The standardized distributions above are dependent upon the shape parameter κ. Meeker and Escobar (1998, ch. 5) give a detailed explanation of the extended generalized gamma distribution.

Note: In JMP, the shape parameter, λ, for the generalized gamma distribution is bounded between [-12,12] to provide numerical stability.

Distributions with Threshold Parameters

Threshold Distributions are log-location-scale distributions with threshold parameters. Some of the distributions above are generalized by adding a threshold parameter, denoted by γ. The addition of this threshold parameter shifts the left endpoint of the distribution away from 0. Threshold parameters are sometimes called shift, minimum, or guarantee parameters because all units survive at least until threshold time. Note that while adding a threshold parameter shifts the distribution on the time axis, the shape, and spread of the distribution are not affected. Threshold distributions are useful for fitting moderately to heavily shifted distributions. The general forms for the pdf and cdf of a log-location-scale threshold distribution are:

$$f(x;\mu,\sigma,\gamma) = \frac{1}{\sigma(x-\gamma)}\phi\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right], \quad x > \gamma$$

$$F(x;\mu,\sigma,\gamma) = \Phi\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right]$$

where ϕ and Φ are the pdf and cdf, respectively, for the specific distribution. Examples of specific threshold distributions are shown below for the Weibull, lognormal, Fréchet, and loglogistic distributions, where, respectively, the SEV, Normal, LEV, and logis pdfs and cdfs are appropriately substituted.

Note: If the smallest observation is a failure (not censored), JMP creates a small interval around the point and treats the observation as interval censored. This padding around the failure bounds the log-likelihood function and improves estimation. If the smallest observation is censored, then no extra padding is added to the observation.

TH Weibull

The pdf and cdf of the three-parameter Weibull distribution are:

$$f(x;\mu,\sigma,\gamma) = \frac{1}{(x-\gamma)\sigma}\phi_{sev}\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right], \quad x > \gamma,\ \sigma > 0$$

$$F(x;\mu,\sigma,\gamma) = \Phi_{sev}\left(\frac{\log(x-\gamma)-\mu}{\sigma}\right) = 1 - \exp\left[-\left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right], \quad x > \gamma$$

where $\mu = \log(\alpha)$, and $\sigma = 1/\beta$ and where

$$\phi_{sev}(z) = \exp[z - \exp(z)]$$

and

$$\Phi_{sev}(z) = 1 - \exp[-\exp(z)]$$

are the pdf and cdf, respectively, for the standardized smallest extreme value, SEV($\mu = 0$, $\sigma = 1$) distribution.

TH Lognormal

The pdf and cdf of the three-parameter lognormal distribution are:

$$f(x;\mu, \sigma, \gamma) = \frac{1}{\sigma(x-\gamma)}\phi_{nor}\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right], \quad x > \gamma$$

$$F(x;\mu,\sigma,\gamma) = \Phi_{nor}\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right]$$

where

$$\phi_{nor}(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$$

and

$$\Phi_{nor}(z) = \int_{-\infty}^{z}\phi_{nor}(w)dw$$

are the pdf and cdf, respectively, for the standardized normal, or N($\mu = 0$, $\sigma = 1$) distribution.

TH Fréchet

The pdf and cdf of the three-parameter Fréchet distribution are:

$$f(x;\mu, \sigma, \gamma) = \frac{1}{\sigma(x-\gamma)}\phi_{lev}\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right], \quad x > \gamma$$

$$F(x;\mu,\sigma,\gamma) = \Phi_{lev}\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right]$$

where

$$\phi_{lev}(z) = \exp[-z - \exp(-z)]$$

and

$$\Phi_{lev}(z) = \exp[-\exp(-z)]$$

are the pdf and cdf, respectively, for the standardized largest extreme value LEV($\mu = 0$, $\sigma = 1$) distribution.

TH Loglogistic

The pdf and cdf of the three-parameter loglogistic distribution are:

$$f(x;\mu, \sigma, \gamma) = \frac{1}{\sigma(x-\gamma)}\phi_{\text{logis}}\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right], \quad x > \gamma$$

$$F(x;\mu,\sigma,\gamma) = \Phi_{\text{logis}}\left[\frac{\log(x-\gamma)-\mu}{\sigma}\right]$$

where

$$\phi_{\text{logis}}(z) = \frac{\exp(z)}{[1+\exp(z)]^2}$$

and

$$\Phi_{\text{logis}}(z) = \frac{\exp(z)}{[1+\exp(z)]} = \frac{1}{1+\exp(-z)}$$

are the pdf and cdf, respectively, for the standardized logistic or logis distribution ($\mu = 0$, $\sigma = 1$).

Distributions for Defective Subpopulations

In reliability experiments, there are times when only a fraction of the population has a particular defect leading to failure. Because all units are not susceptible to failure, using the regular failure distributions is inappropriate and might produce misleading results. Use the DS distribution options to model failures that occur on only a subpopulation. The following DS distributions are available:

- DS Lognormal
- DS Weibull
- DS Loglogistic
- DS Fréchet

The pdf and cdf for defective subpopulation distributions are defined as follows:

$$f(t) = \left[p\frac{1}{t\sigma}\right]\phi\left[\frac{(\log(t)-\mu)}{\sigma}\right]$$

$$F(t) = p\Phi\left[\left(\frac{(\log(t)-\mu)}{\sigma}\right)\right]$$

where:

p is the defective subpopulation fraction

t is the time of measurement for the lifetime event

μ and σ are estimated by calculating the usual maximum likelihood estimations using the pdf and cdf of the corresponding defective subpopulation

$\phi(z)$ and $\Phi(z)$ are the density and cumulative distribution function, respectively, for a standard distribution. For example, for a Weibull distribution,

$\phi(z) = \exp(z\text{-}\exp(z))$ and $\Phi(z) = 1 - \exp(\text{-}\exp(z))$.

See Tobias and Trindad (2012, p. 321) for more information about defective subpopulation models.

The defective subpopulation model is also known as a *limited failure population* model in Meeker and Escobar (1998, ch. 11).

Zero-Inflated Distributions

Zero-inflated distributions are used when some proportion (p) of the data fail at $t = 0$. When the data contain more zeros than expected by a standard model, the number of zeros is inflated. When the time-to-event data contain zero as the minimum value in the Life Distribution platform, four zero-inflated distributions are available. These distributions include:

- Zero-Inflated Lognormal (ZI Lognormal)
- Zero-Inflated Weibull (ZI Weibull)
- Zero-Inflated Loglogistic (ZI Loglogistic)
- Zero-Inflated Fréchet (ZI Fréchet)

The pdf and cdf for zero-inflated distributions are defined as follows:

$$f(t) = \left[(1-p)\frac{1}{t\sigma}\right]\phi\left[\frac{(\log(t)-\mu)}{\sigma}\right]$$

$$F(t) = p + (1-p)\Phi\left[\left(\frac{(\log(t)-\mu)}{\sigma}\right)\right]$$

where:

p is the proportion of zero data values

t is the time of measurement for the lifetime event

μ and σ are estimated by calculating the usual maximum likelihood estimations after removing zero values from the original data

$\phi(z)$ and $\Phi(z)$ are the density and cumulative distribution function, respectively, for a standard distribution. For example, for a Weibull distribution,

$\phi(z) = \exp(z\text{-}\exp(z))$ and $\Phi(z) = 1 - \exp(\text{-}\exp(z))$.

See Lawless (2003, p. 34) for more information about zero-inflated distributions. Substitute $p = 1 - p$ and $S_1(t) = 1 - \Phi(t)$ to obtain the form shown above.

See Tobias and Trindade (1995, p. 232) for more information about reliability distributions. This reference gives the general form for mixture distributions. Using the parameterization in Tobias and Trindade, the form above can be found by substituting $\alpha = p$, $F_d(t) = 1$, and $F_N(t) = \Phi(t)$.

Prior Distributions for Bayesian Estimation

The following distributions are available for Location Scale Priors:

- Normal/Lognormal, with hyperparameters Location (mu) and Scale (sigma). For a definition, see "Lognormal" and "Normal".
- Uniform, with hyperparameters Low and End, which define the support of a Uniform distribution.
- Gamma, with hyperparameters Shape and Scale. The k/theta parameterization and the probability density function is used.
- Point Mass, with hyperparameter Location. This is a degenerated prior; there is only one possible value for the parameter that we are assigning a prior distribution to. The only possible value equals the value that is entered to this Location hyperparameter.

The following distributions are available for Quantile Parameter Priors:

- Normal/Lognormal, with a 99% probability range, specifies the prior distribution using the 0.005 and 0.995 percentiles of the distribution. JMP backs out the mu and sigma.
- Uniform, with hyperparameters Lower and Upper Limits, which define the support of a Uniform distribution.
- Log-Uniform, with Lower (a) and Upper (b) Limits. This distribution is uniform on the log scale between Log(a) and Log(b).
- Point Mass, with hyperparameter Location. This is a degenerated prior; there is only one possible value for the parameter that we are assigning a prior distribution to. The only possible value equals the value that is entered to this Location hyperparameter.

The following distributions are available for Failure Probability Priors:

- Beta, characterized by the probability density function.

- Specify the Beta prior using estimates and error percentages (mean and variance). The mean equals the number entered in to the Estimate, and the variance equals (Error Percentage / 100 * Estimate)^2.
- Specify the Beta prior using 0.005 and 0.995 percentiles of the distribution. JMP backs out the hyperparameters.

Competing Cause Details

For a competing cause model, a closed form for the aggregated distribution is defined as follows:

$$F(x) = 1 - \prod_{i=1}^{k} [1 - F_i(x)]$$

where the $F_i(x)$, $i = 1, \ldots, k$, are individual failure distributions corresponding to causes. Confidence limits are readily available, because all involved estimates are MLEs.

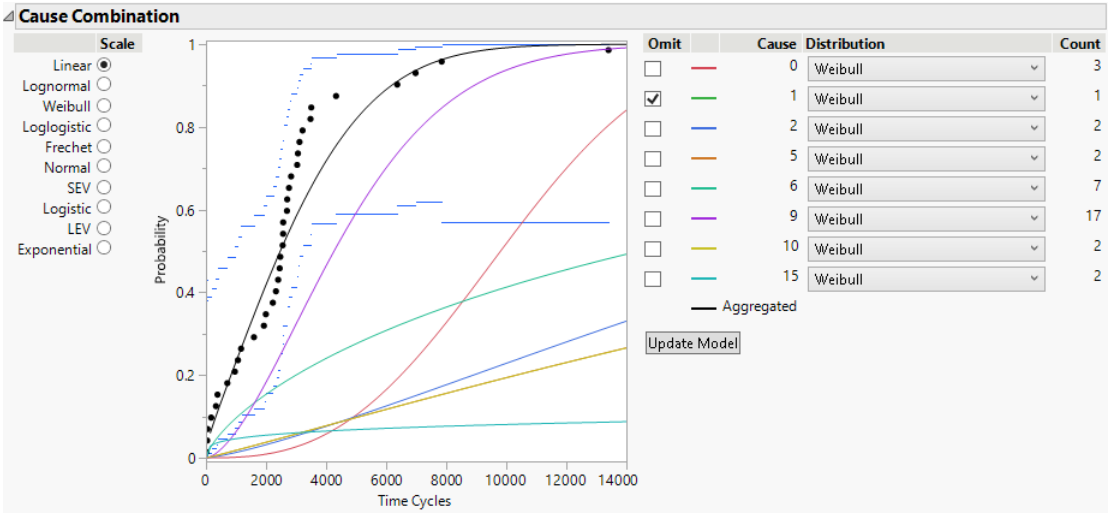
Specify a Fixed Parameter Model as a Distribution for a Cause

If a Fixed Parameter model is specified for a cause, you must fix the parameter in the Individual Causes report for that cause. Fix the parameter in the desired Parametric Estimate report in the Life Distribution - Failure Cause: <Name> report. The fixed parameter becomes part of the aggregated distribution when you click Update Model.

This example illustrates how to include the Fixed Parameter model into the aggregated distribution:

1. Select **Help > Sample Data Library** and open Reliability/Appliance.jmp.
2. Select **Analyze** > **Reliability and Survival** > **Life Distribution**.
3. Select Time Cycles and click **Y, Time to Event.**
4. Select Cause Code and click **Failure Cause**.
5. Select **Likelihood** as the Confidence Interval Method.
6. Select **Allow failure mode to use fixed parameter models**.
7. Click **OK.**

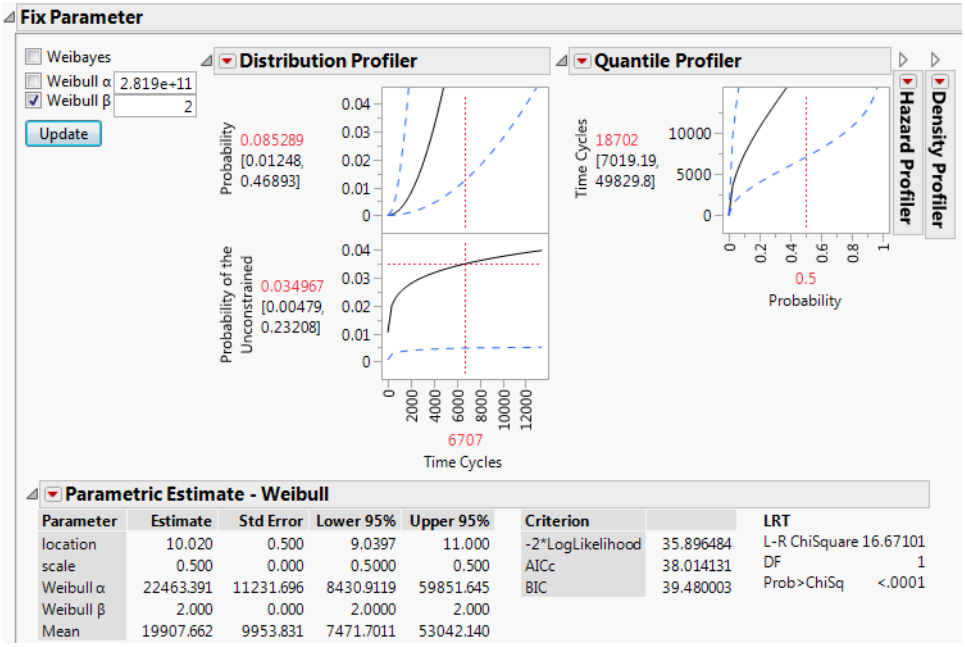
Figure 3.25 Fixed Parameter Model with Cause 1 Omitted



By default, Cause = 1 is omitted, because there are not enough data. However, you do not want this cause to be omitted.

8. Open the Individual Causes report for Cause 1. The report is called Life Distribution - Failure Cause: 1 Failure Counts: 1.
9. Click the red triangle next to Parametric Estimate - Weibull and select **Fix Parameter**.
10. Select **Weibull beta** and type 2.
11. Click **Update**.

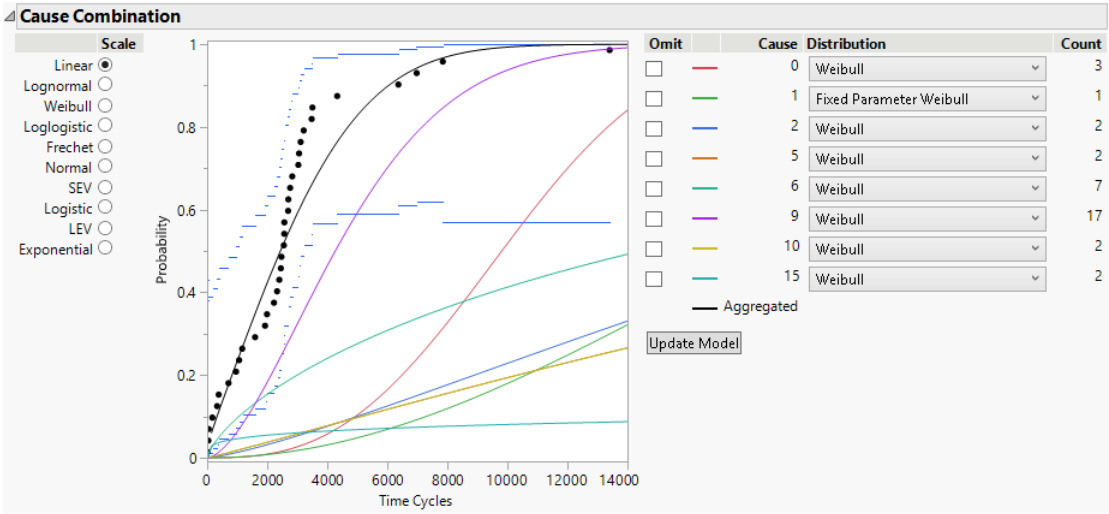
Figure 3.26 Fixed Parameter Model with Weibull Beta Specified



In the Parametric Estimate - Weibull report, assuming β equals 2, the alpha parameter is estimated to be 22463.391. Now you can use this for the failure distribution for Cause=1.

12. Scroll up to Cause Combination at the top of the report window.
13. Deselect **Omit** for Cause 1.
14. For the distribution for Cause 1, select **Fixed Parameter Weibull**.
15. Click **Update Model**.

Figure 3.27 Updated Model Showing Cause 1



Now the aggregated model uses the Fixed Parameter Weibull results for Cause 1 in the overall competing cause model.

Specify a Bayesian Model for a Cause

The steps for specifying a Bayesian model for a cause are similar to those described in "Specify a Fixed Parameter Model as a Distribution for a Cause". Define the model in the desired Bayesian Estimation report found in the corresponding Parametric Estimate outline under Statistics in the Life Distribution report for the individual cause. See "Bayesian Estimation - <Distribution Name>".

To incorporate a Bayesian model into the aggregated model, non-Bayesian distributions for other causes must be amenable to a simulation-based framework. For example, suppose that a model has two failure causes. One is modeled using a Weibull distribution and the other using a Bayesian approach for estimating the parameters of a second Weibull. The parameters for the first Weibull distribution, denoted by the vector θ_1, are estimated using maximum likelihood. The parameters for the second Weibull, θ_2, are estimated using the Bayesian approach.

The quantiles and median of the aggregated mixture distribution, denoted $F(x, \theta_1, \theta_2)$, are obtained as follows:

- A parametric bootstrap is performed for the first Weibull, yielding random samples from the asymptotic distribution of the maximum likelihood estimate $\hat{\theta}_1$. Denote a sampled value from the asymptotic distribution of $\hat{\theta}_1$ by θ_1^*.
- A sample is drawn from the posterior distribution of θ_2, denoted by θ_2^*.

- For each set of values θ_1^* and θ_2^*, an estimate of $F(x, \theta_1, \theta_2)$, denoted by $F^*(x, \theta_1, \theta_2)$, is obtained.
- The values $F^*(x, \theta_1, \theta_2)$ are used to obtain estimates of the quantiles and median of the aggregated distribution. These are the values displayed in the Distribution profiler at a given value of x.

Specify a Weibayes Model for a Cause

The steps for specifying a Weibayes model for a cause are similar to those described in "Specify a Fixed Parameter Model as a Distribution for a Cause". Select the Fix Parameter option in the Parametric Estimate - Weibull outline under Statistics in the Life Distribution report for the cause. In the Fix Parameter report, check the Weibayes option. The Weibayes model is treated as a Bayesian model and a bootstrap sample is drawn from the posterior distribution of the parameter alpha. See Liu and Wang (2013).

Mean Remaining Life Calculator

Use the Configuration option in the red triangle menu to set a value for the number of simulated failure times used in computing the mean remaining life. Denote this value by m.

To obtain an estimate of the mean remaining life at time t, m samples are drawn from the aggregated distribution conditioned on survival to time t. Their average is computed.

To compute the confidence limits for the mean remaining life, you must select the box in the Configuration window. You then have the option to set the number of bootstrap samples. Denote this value by n.

To compute the confidence interval, n samples of parameter estimates are drawn from either the asymptotic distributions of the MLEs, or the posterior distributions derived using Bayesian inference. For each sample of parameter values, an aggregated distribution is formed, from which m samples are drawn to compute a mean remaining life. The samples of n mean remaining life values are used to construct the confidence interval.

Fit Mixture Save Predictions Formulas

This section gives the formulas used in calculating values in the columns saved by the Fit Mixture report option Save Predictions.

Consider the following notation:

$\hat{p}_i$ is an estimate of the mixture proportion, w_i

$\hat{F}_i$ is the estimated probability distribution function F_i

$\hat{f}_i$ is the estimated probability density function for F_i

- If the observation y is not censored, the saved value is given by the following:

$$\frac{\hat{p}_i \hat{f}_i(y)}{\sum_{i=1}^{k} \hat{p}_i \hat{f}_i(y)}$$

- If the observation is censored, the saved value is obtained by replacing the estimated density values in the formula for an uncensored observation by the following:

$\hat{F}_i(y)$ for right censoring

$1 - \hat{F}_i(y)$ for left censoring

$\hat{F}_i(y_{high}) - \hat{F}_i(y_{low})$ for interval censoring

Notes Regarding Median Rank Regression

When there are no censored rows and no Weight column, median rank regression (MRR) for Weibull parameter estimates is available in the Life Distribution platform. The following conditions must be met before MRR estimates appear in the Life Distribution report:

- This Life Distribution platform preference is selected: Report Median Rank Regression Based Weibull Parameter Estimates When There Are No Censored Observations.
- There are no censored observations in the analysis. In other words, all observations represent failure times.
- There is no Weight column specified.

If all of the above conditions are met, the Parametric Estimate - Weibull report is modified to show MRR estimates as well as the usual maximum likelihood estimates (MLE). The column headings in the parameter estimates table are prefixed with MLE or MRR to denote which estimation method was used to produce each column of estimates. There are no confidence intervals available for the MRR estimates.

Caution: The use of MRR estimates is not recommended. Median rank regression uses least squares estimation to produce estimates, instead of maximum likelihood. There are no commonly accepted methods to apply least squares estimation to censored data. Further, Genschel and Meeker (2010) show that the MLE method is more precise. They provide simulation results that are based on the fact that the true estimates are known, but in reality, the accuracy of MRR estimates is unknown. Because MRR is a loosely defined and ad hoc estimation procedure, different software packages produce different MRR estimates for the same data.

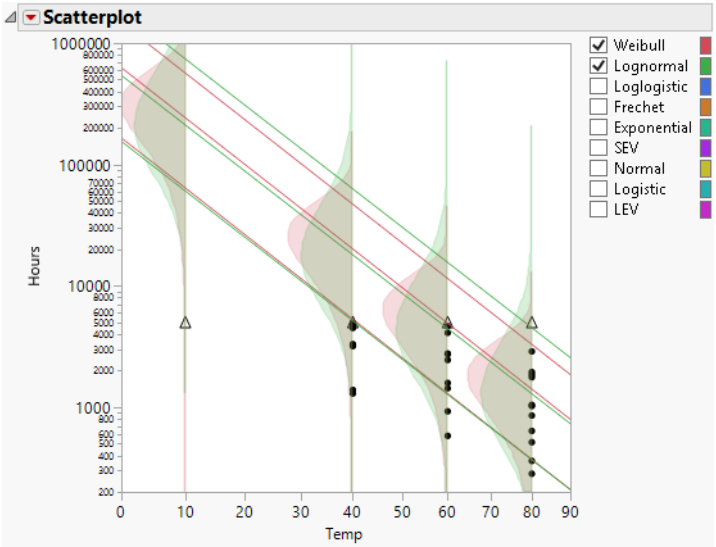
Chapter 4

Fit Life by X

Fit Single-Factor Models to Time-to Event Data

The Fit Life by X platform helps you analyze lifetime events when only one factor is present. You can choose to model the relationship between the event and the factor using various transformations, or create a custom transformation of your data. You also have the flexibility of comparing different distributions at the same factor level and comparing the same distribution across different factor levels.

Figure 4.1 Scatterplot Showing Varying Distributions and Factor Levels



Contents

Overview of the Fit Life by X Platform 105
Example of the Fit Life by X Platform 106
Launch the Fit Life by X Platform 108
The Fit Life by X Report 110
Summary of Data 111
Scatterplot 111
Nonparametric Overlay 113
Comparisons 114
Results 118
Custom Relationship 129
Fit Life by X Platform Options 130
Additional Examples of the Fit Life by X Platform 131
Capacitor Example 131
Custom Relationship Example 133

Overview of the Fit Life by X Platform

The Fit Life by X platform provides the tools needed for accelerated life-testing analysis. Accelerated tests are routinely used in industry to provide failure-time information about products or components in a relatively short time frame. Common accelerating factors include temperature, voltage, pressure, and usage rate. Results are extrapolated to obtain time-to-failure estimates at lower, normal operating levels of the accelerating factors. These results are used to assess reliability, detect and correct failure modes, compare manufacturers, and certify components.

The Fit Life by X platform includes many commonly used transformations to model physical and chemical relationships between the event and the factor of interest. Examples include transformation using Arrhenius (Celsius, Fahrenheit, and Kelvin) relationship time-acceleration factors and Voltage-acceleration mechanisms. Linear, Log, Logit, Reciprocal, Square Root, Box-Cox, Location, Location and Scale, and Custom acceleration models are also included in this platform.

You can use the DOE > Accelerated Life Test Design platform to design accelerated life test experiments. See the *Design of Experiments Guide*.

Meeker and Escobar (1998, p. 495) offer the following strategy for analyzing accelerated lifetime data:

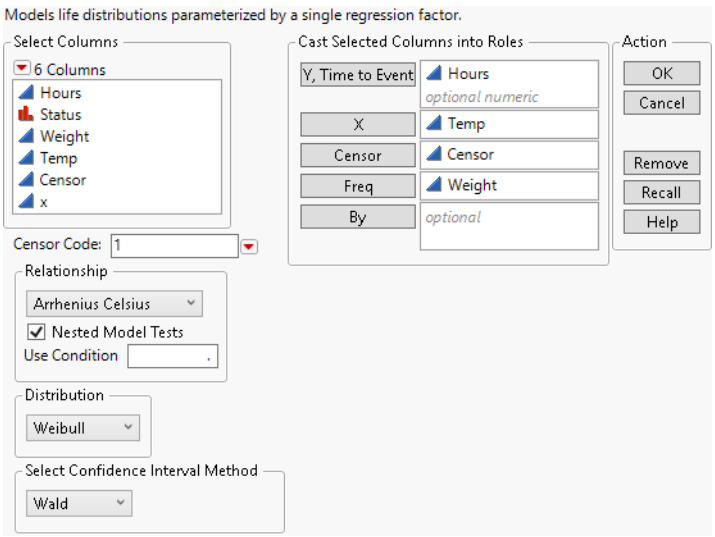
1. Examine the data graphically. One useful way to visualize the data is by examining a scatterplot of the time-to-failure variable versus the accelerating factor.
2. Fit distributions individually to the data at different levels of the accelerating factor. Repeat for different assumed distributions.
3. Fit an overall model with a plausible relationship between the time-to-failure variable and the accelerating factor.
4. Compare the model in Step 3 with the individual analyses in Step 2, assessing the lack of fit for the overall model.
5. Perform residual and various diagnostic analyses to verify model assumptions.
6. Assess the plausibility of the data to make inferences.

Example of the Fit Life by X Platform

This example uses the Devalt.jmp sample data table, from Meeker and Escobar (1998), and can be found in the Reliability folder of the sample data. It contains time-to-failure data for a device at accelerated operating temperatures. No time-to-failure observation is recorded for the normal operating temperature of 10 degrees Celsius; all other observations are shown as time-to-failure or censored values at accelerated temperature levels of 40, 60, and 80 degrees Celsius.

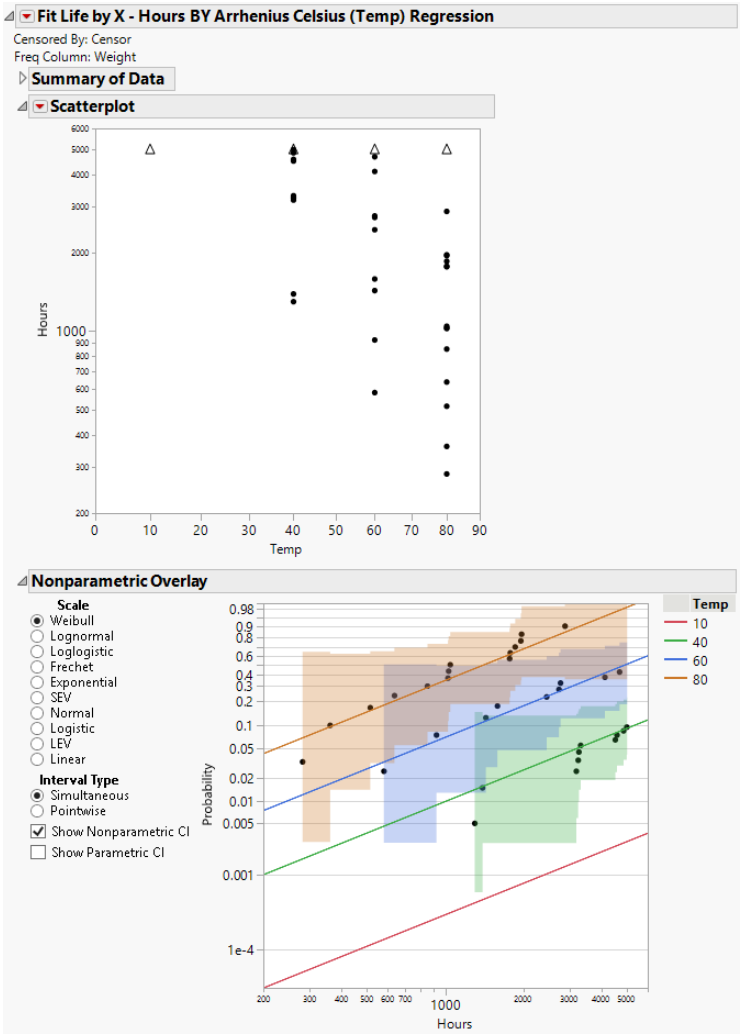
1. Select **Help > Sample Data Library** and open Reliability/Devalt.jmp.
2. Select **Analyze > Reliability and Survival > Fit Life by X**.
3. Select Hours and click **Y, Time to Event**.
4. Select Temp and click **X**.
5. Select Censor and click Censor.
6. Leave the **Censor Code** as 1.
7. Select Weight and click **Freq**.
8. Keep **Arrhenius Celsius** as the relationship, and keep the **Nested Model Tests** option selected.
9. Select **Weibull** as the distribution.
10. Keep **Wald** as the confidence interval method.

Figure 4.2 Fit Life by X Launch Window



11. Click **OK**.

Figure 4.3 Fit Life by X Report Window for Devalt.jmp Data

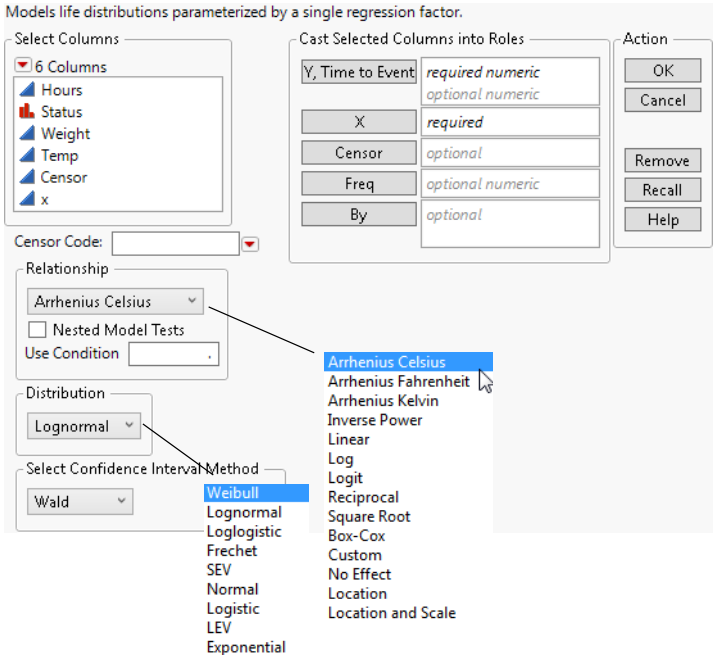


The report window shows summary data, diagnostic plots, comparison data and results, including detailed statistics and prediction profilers. Separate result sections are shown for each selected distribution. Distribution, Quantile, Hazard, Density, and Acceleration Factor Profilers are included for each of the specified distributions.

Launch the Fit Life by X Platform

Launch the Fit Life by X platform by selecting **Analyze > Reliability and Survival > Fit Life by X**. For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Figure 4.4 The Fit Life by X Launch Window



The Fit Life by X launch window contains the following options:

Y, Time to Event Identifies the time to event (such as the time to failure) or time to censoring. With interval censoring, specify two Y variables, where one Y variable gives the lower limit and the other Y variable gives the upper limit for each unit. For more information about censoring, see "Event Plot".

X Identifies the accelerating factor.

Censor Identifies censored observations. Select the value that identifies right-censored observations from the Censor Code menu beneath the Select Columns list. The Censor column is used only when one Y is entered.

Freq Identifies frequencies or observation counts when there are multiple units. If the value is 0 or a positive integer, then the value represents the frequencies or counts of observations for each row when there are multiple units recorded.

By Identifies a column that creates a report consisting of separate analyses for each level of the variable.

Censor Code Identifies the value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

Relationship Identifies the relationship between the event and the accelerating factor. Table 4.1 defines the model for each relationship.

Table 4.1 Models for Relationship Options

Relationship	Model
Arrhenius Celsius	μ = b0 + b1 * 11604.5181215503 / (X + 273.15)
Arrhenius Fahrenheit	μ = b0 + b1 * 11604.5181215503 / ((X + 459.67) / 1.8)
Arrhenius Kelvin	μ = b0 + b1 * 11604.5181215503 / X
Voltage	μ = b0 + b1 * log(X)
Linear	μ = b0 + b1 * X
Log	μ = b0 + b1 * log(X)
Logit	μ = b0 + b1 * log(X / (1 - X))
Reciprocal	μ = b0 + b1 / X
Square Root	μ = b0 + b1 * sqrt(X)
Box-Cox	μ = b0 + b1 * BoxCox(X)
Location	means that μ is different for every level of X
Location and Scale	means that μ and σ are both different for every level of X (equivalent to a Life Distribution fit with X as a By variable)
Custom	user-defined μ and σ

If you select Box-Cox, a text edit box appears below the Use Condition option. Use this box to specify a lambda value. The BoxCox(X) transformation for a specified λ is defined as follows:

$$x_i^{(\lambda)} = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x_i) & \text{if } \lambda = 0 \end{cases}$$

If you want to use a Custom relationship for your model, see "Custom Relationship".

Nested Model Tests Appends a nonparametric overlay plot, nested model tests, and a multiple probability plot to the report window.

Use Condition Enables you to enter a value for the explanatory variable, X, of the acceleration factor. You can also set the use condition value after launching the platform using the Set Time Acceleration Use Condition option from the Fit Life by X red triangle menu.

Distribution Specifies one distribution (Weibull, Lognormal, Loglogistic, Fréchet, SEV, Normal, Logistic, LEV, or Exponential distributions) at a time. Lognormal is the default setting.

Select Confidence Interval Method Displays the method for computing confidence intervals for the parameters. The default is Wald, but you can select Likelihood instead. The Wald method is an approximation and runs faster. The Likelihood method provides more precise parameters but takes longer to compute.

Note: The Confidence Interval Method preference enables you to select the Likelihood method as the default confidence interval method. You can change this preference in **Preferences > Platforms > Fit Life by X**.

The Fit Life by X Report

The initial report window includes the following sections:

- "Summary of Data"
- "Scatterplot"
- "Nonparametric Overlay"
- "Comparisons"

  Distribution, Quantile, Hazard, Density, and Acceleration Factor profilers, along with criteria values under Comparison Criterion can be viewed and compared.
- "Results"

Parametric estimates, covariance matrices, nested model tests, and diagnostics can be examined and compared for each of the selected distributions. You can also perform custom estimation and obtain Bayesian estimates for the distribution parameters.

Summary of Data

The Summary of Data section gives the total number of observations, the number of uncensored values, and the number of censored values (right, left, and interval). Figure 4.5 shows the summary data for the Devalt.jmp sample data table.

Figure 4.5 Summary of Data Example

Summary of Data

Observation Used	165
Uncensored Values	33
Right Censored Values	132

Scatterplot

The Scatterplot of the lifetime event versus the explanatory variable is shown at the top of the report window. For the Devalt.jmp sample data, the Scatterplot shows Hours versus Temp. Table 4.2 indicates how each type of failure is represented on the Scatterplot in the report window. To increase the size of the markers on the graph, right-click the graph, select **Marker Size**, and then select one of the marker sizes listed.

Figure 4.6 Scatterplot of Hours versus Temp

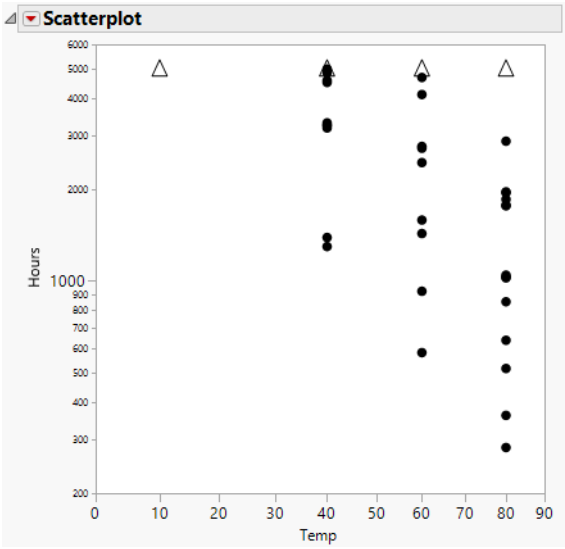


Table 4.2 Scatterplot Representation for Failure and Censored Observations

Event	Scatterplot Representation
failure	dots
right-censoring	upward triangles
left-censoring	downward triangles
interval-censoring	downward triangle on top of an upward triangle, connected by a solid line

Scatterplot Options

The Scatterplot red triangle menu contains the following options:

Add Density Curve Specify the density curve that you want, one at a time, by entering any value within the range of the accelerating factor. You can then select different distributions by selecting the appropriate check box(es) that appear after you add a curve.

Remove Density Curves Displays previously entered density curve values. Remove curves by selecting the appropriate check box.

Show Density Curves Displays the density curves. If the **Location** or the **Location and Scale** model is fit, or if **Nested Model Tests** is selected in the launch window, then the density curves for all of the given explanatory variable levels are shown. After the curves have been created, the **Show Density Curves** option toggles the curves on and off the plot.

Add Quantile Lines Specify the quantile lines that you want, three at a time. You can add more quantiles by continually selecting **Add Quantile Lines**. Default quantile values are 0.1, 0.5, and 0.9. Invalid quantile values, such as missing values, are ignored. If desired, you can enter just one quantile value, leaving the other entries blank.

Show Quantile Line CI Bands Shows or hides confidence intervals around the quantile lines.

Set Level of Quantile Line CI Bands Specifies the confidence level for the confidence intervals around the quantile lines.

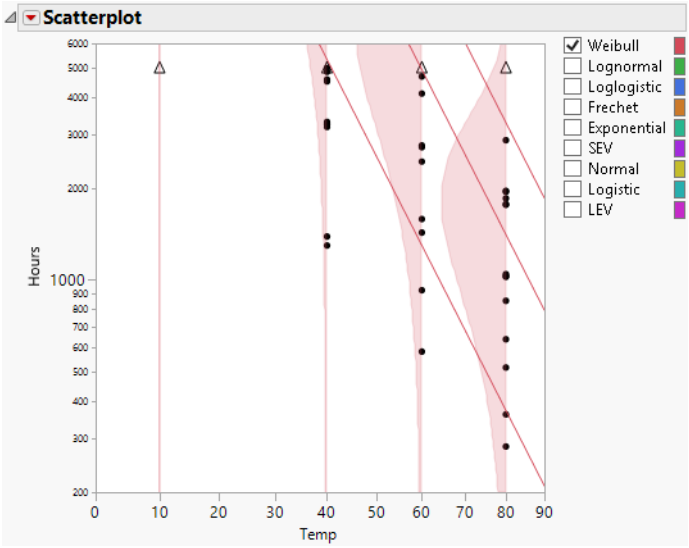
Remove Quantile Lines Displays previously entered quantile values. Remove lines by selecting the appropriate check box.

Transposed Axes Swaps the horizontal and vertical axes.

Use Transformation Scale The default view of the scatterplot incorporates the transformation scale. Select this option to switch between the linear and nonlinear scales for the horizontal axis.

Figure 4.6 shows the initial scatterplot; Figure 4.7 shows the resulting scatterplot with the **Show Density Curves** and **Add Quantile Lines** options selected displaying the curves and the lines for the various Temp levels for the Weibull distribution. You can also view density curves across all the levels of Temp for the other distributions. These distributions can be selected one at a time or can be viewed simultaneously by checking the boxes to the left of the desired distribution name(s).

Figure 4.7 Scatterplot with Density Curve and Quantile Line Options



Nonparametric Overlay

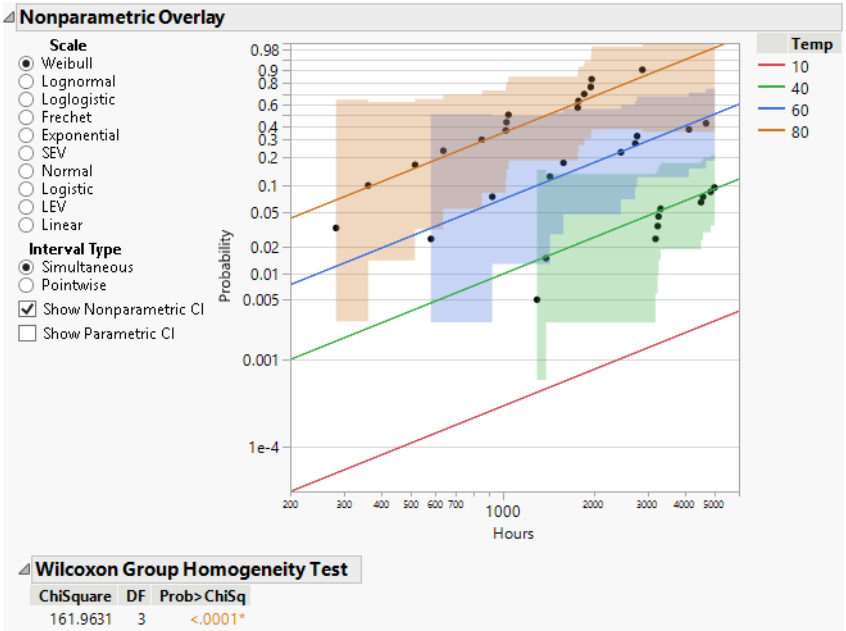
The Nonparametric Overlay plot is displayed after the scatterplot. Differences among groups can readily be detected by examining this plot. You can choose different scales for viewing these differences. You can change the interval type on a Nonparametric fit probability plot between **Simultaneous** and **Pointwise** (results displayed when **Show Nonparametric CI** is selected). You can also select whether to show parametric or nonparametric confidence intervals.

Pointwise estimates show the pointwise 95% confidence bands on the plot while simultaneous confidence intervals show the simultaneous confidence bands for all groups on the plot. Meeker and Escobar (1998, ch. 3) discuss pointwise and simultaneous confidence intervals and the motivation for simultaneous confidence intervals in a lifetime analysis.

Wilcoxon Group Homogeneity Test

For this example, the **Wilcoxon Group Homogeneity Test**, shown in Figure 4.8, indicates that there is a difference among groups. The high chi-square value and low *p*-value are consistent with the differences seen among the Temp groups in the Nonparametric Overlay plot.

Figure 4.8 Nonparametric Overlay Plot and Wilcoxon Test for Devalt.jmp



Comparisons

The Comparisons report section, shown in Figure 4.9, shows profilers for the selected distributions in the Nonparametric Overlay section, and includes the following tabs:

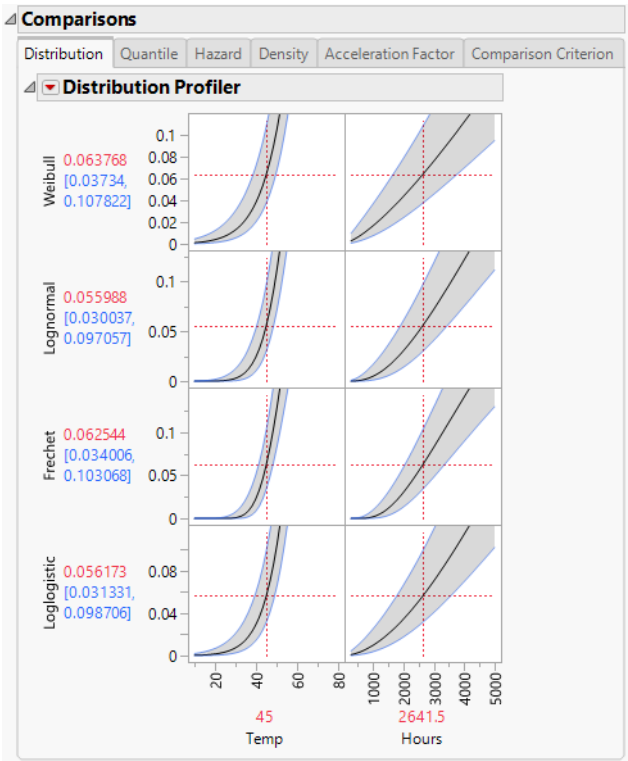
- Distribution
- Quantile
- Hazard
- Density
- Acceleration Factor
- Comparison Criterion

To show a specific profiler, select the appropriate distribution option in the Nonparametric Overlay section.

Profilers

The first five tabs show profilers for the selected distributions. Curves shown in the first four profilers correspond to both the time-to-event and explanatory variables. The Acceleration Factor profiler tab corresponds only to the acceleration factor (explanatory variable). Figure 4.9 shows the Distribution Profiler for the Weibull, Lognormal, Fréchet, and Loglogistic distributions.

Figure 4.9 Distribution Profiler

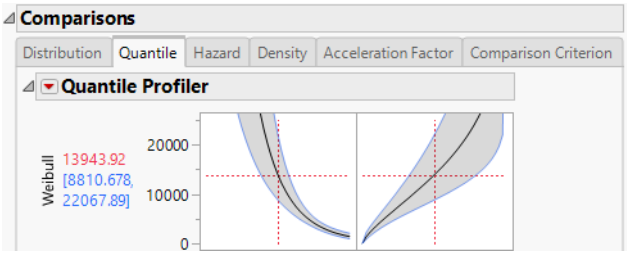


Comparable results appear on the **Quantile**, **Hazard**, and **Density** tabs. The Distribution, Quantile, Hazard, Density, and Acceleration Factor Profilers behave similarly to the Prediction Profiler in other platforms. For example, the vertical lines of Temp and Hours can be dragged to see how each of the distribution values change with temperature and time. For a detailed explanation of the Prediction Profiler, see *Profilers*.

Quantile

You can use the Quantile profiler for extrapolation. Suppose that the data are represented by a Weibull distribution. From viewing the Weibull Acceleration Factor Profiler in Figure 4.11, you see that the acceleration factor at 45 degrees Celsius is 17.42132 for a use condition temperature of 10 degrees Celsius. Select the **Quantile** tab to see the Quantile Profiler for the Weibull distribution. Select and drag the vertical line in the probability plot so that the probability reads 0.5. From viewing Figure 4.10, where the Probability is set to 0.5, you find that the quantile for the failure probability of 0.5 at 45 degrees Celsius is 13943.92 hours. So, at 10 degrees Celsius, you can expect that 50% of the units fail by 13943.92 * 17.42132 = 242921 hours.

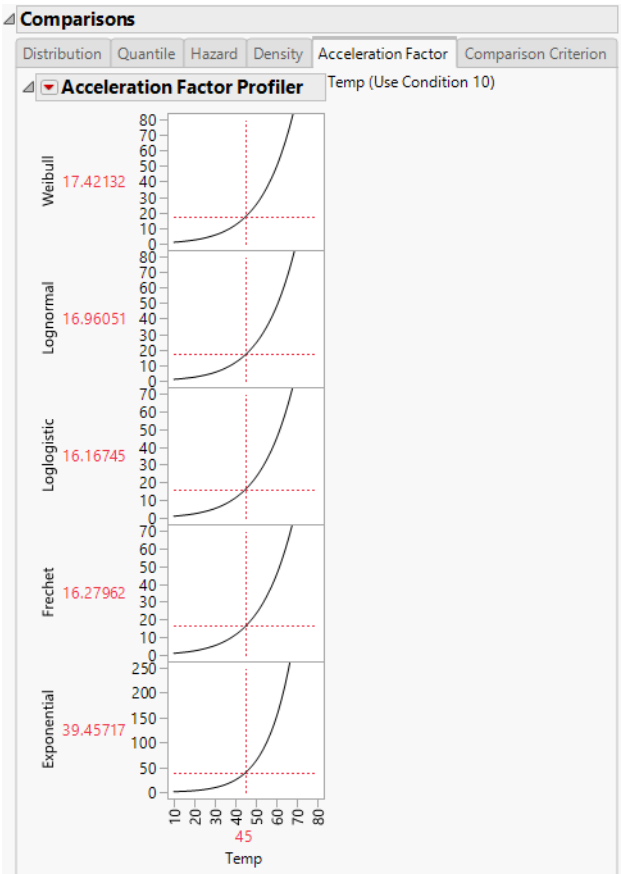
Figure 4.10 Weibull Quantile Profiler for Devalt.jmp



Acceleration Factor

Selecting the **Acceleration Factor** tab shows the Acceleration Factor Profiler for the time-to-event variable for each specified distribution. To produce Figure 4.11, select **Fit All Distributions** from the Fit Life by X red triangle menu. Modify the use condition value for the explanatory variable by selecting **Set Time Acceleration Use Condition** from the Fit Life by X red triangle menu and entering the desired value. Note that the explanatory variable and the use condition value appear beside the profiler title.

Figure 4.11 Acceleration Factor Profiler for Devalt.jmp



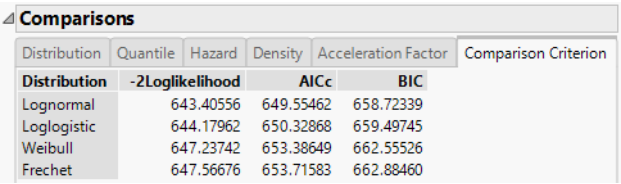
The Acceleration Factor Profiler lets you estimate time-to-failure for accelerated test conditions when compared with the use condition value and a parametric distribution assumption. The interpretation of a time-acceleration plot is generally the ratio of the p^{th} quantile of the use condition to the p^{th} quantile of the accelerated test condition. This relation applies only when the distribution is Lognormal, Weibull, Loglogistic, or Fréchet, and the scale parameter is constant for all levels. For more information about the parameterizations of the distributions, see "Distributions".

Note: The Acceleration Factor Profiler does not appear in the following instances: when the explanatory variable is discrete; the explanatory variable is treated as discrete; a customized formula does not use a unity scale factor; or the distribution is Normal, SEV, Logistic, or LEV.

Comparison Criterion

The **Comparison Criterion** tab shows the -2Loglikelihood, AICc, and BIC criteria for the distributions of interest. Figure 4.12 shows these values for the Weibull, Lognormal, Loglogistic, and Fréchet distributions. Distributions providing better fits to the data are shown at the top of the Comparisons report, sorted by AICc.

Figure 4.12 Comparison Criterion Report Tab



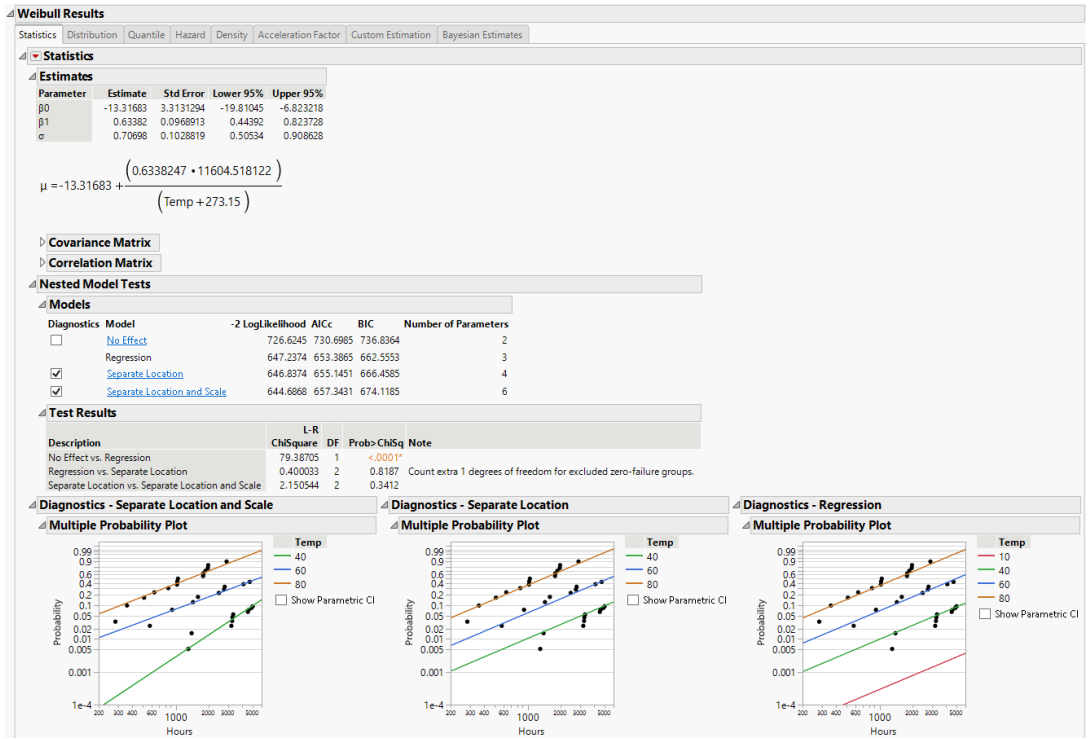
This report suggests that the Lognormal and Loglogistic distributions provide the best fits for the data, because the lowest criteria values are seen for these distributions. For more information about the criteria, see *Fitting Linear Models*.

Results

The Results section of the report window shows more detailed statistics and prediction profilers than those shown in the Comparisons report. Separate result sections are shown for each selected distribution. Figure 4.13 shows a portion of the Weibull results, Nested Model Tests, and Diagnostics plots for Devalt.jmp.

Statistical results, diagnostic plots, and Distribution, Quantile, Hazard, Density, and Acceleration Factor Profilers are included for each of your specified distributions. The Custom Estimation tab lets you estimate specific failure probabilities and quantiles, using both Wald and Profile interval methods. When the Box-Cox Relationship is selected on the platform launch window, the Sensitivity tab appears. This tab shows how the Relative Likelihood and B10 Life change as a function of Box-Cox lambda.

Figure 4.13 Weibull Distribution Nested Model Tests for Devalt.jmp Data



Statistics

For each parametric distribution, there is a Statistics section that shows parameter estimates, a covariance matrix, confidence intervals, summary statistics, and diagnostic plots. You can save probability, quantile, and hazard estimates by selecting any or all of these options from the Statistics red triangle menu for each parametric distribution. The estimates and the corresponding lower and upper confidence limits are saved as columns in your data table.

Nested Model Tests

Nested Model Tests are included, if you selected the option on the platform launch window. The Nested Model Tests include statistics and diagnostic plots for the following models:

Separate Location and Scale Assumes that the location and scale parameters are different for all levels of the explanatory variable. This option is equivalent to fitting the distribution by the levels of the explanatory variable. The Separate Location and Scale model has multiple location parameters and multiple scale parameters (Figure 4.14).

Separate Location Assumes that the location parameters are different, but the scale parameters are the same for all levels of the explanatory variable. The Separate Location model has multiple location parameters and only one scale parameter (Figure 4.15).

Regression The default model shown in the initial Fit Life by X report window (Figure 4.16).

No Effect Assumes that the explanatory variable does not affect the response. This option is equivalent to fitting all of the data values to the selected distribution. The No Effect Model has one location parameter and one scale parameter (Figure 4.17).

Separate Location and Scale, Separate Location, and Regression analyses results are shown by default. Regression parameter estimates and the location parameter formula are shown under the Estimates section, by default. The Diagnostics plots for the No Effect model can be displayed by selecting the check box to the left of No Effect under the Nested Model Tests title.

To see results for each of the models (independently of the other models), click the underlined model of interest (listed under Nested Model Tests) and then uncheck the check boxes for the other models.

If the Nested Model Tests option was not checked in the launch window, then the Separate Location and Scale and Separate Location models are not assessed. In this case, estimates are given for the regression model for each distribution that you select and the Cox-Snell Residual P-P Plot is the only diagnostic plot.

Note: When Separate Location and Scale or Separate Location models are fit for the Weibull distribution, both parameterizations of the Weibull distribution are shown in the Estimates table, as is the case in Figure 4.14 and Figure 4.15. For more information about the Weibull parameterizations, see "Weibull".

Diagnostics

The Multiple Probability Plots shown in Figure 4.13 are used to validate the distributional assumption for the different levels of the accelerating variable. If the line for each level does not run through the data points for that level, the distributional assumption might not hold. Side-by-side comparisons of the diagnostic plots provide a visual comparison for the validity of the different models. See Meeker and Escobar (1998, sec. 19.2.2) for a discussion of multiple probability plots. Each multiple probability plot has an option below the legend that enables you to show or hide shaded parametric confidence intervals for each line in the plot.

The Cox-Snell Residual P-P Plots are used to validate the distributional assumption for the data. If the data points deviate far from the diagonal, then the distributional assumption might be violated. The Cox-Snell Residual P-P Plot red triangle menu has an option called **Save Residuals** that enables you to save the residual data to the data table. See Meeker and Escobar (1998, sec. 17.6.1) for a discussion of Cox-Snell residuals.

The Residuals versus Fitted Plots are used to validate the distributional assumption for the different levels of the accelerating variable. The plots show standardized residuals on the vertical axis and the X variable on the horizontal axis. The Residuals vs Fitted red triangle menu has an option called **Save Residuals** that enables you to save the standardized residual data to the data table.

Figure 4.14 Separate Location and Scale Model with the Weibull Distribution for Devalt.jmp Data

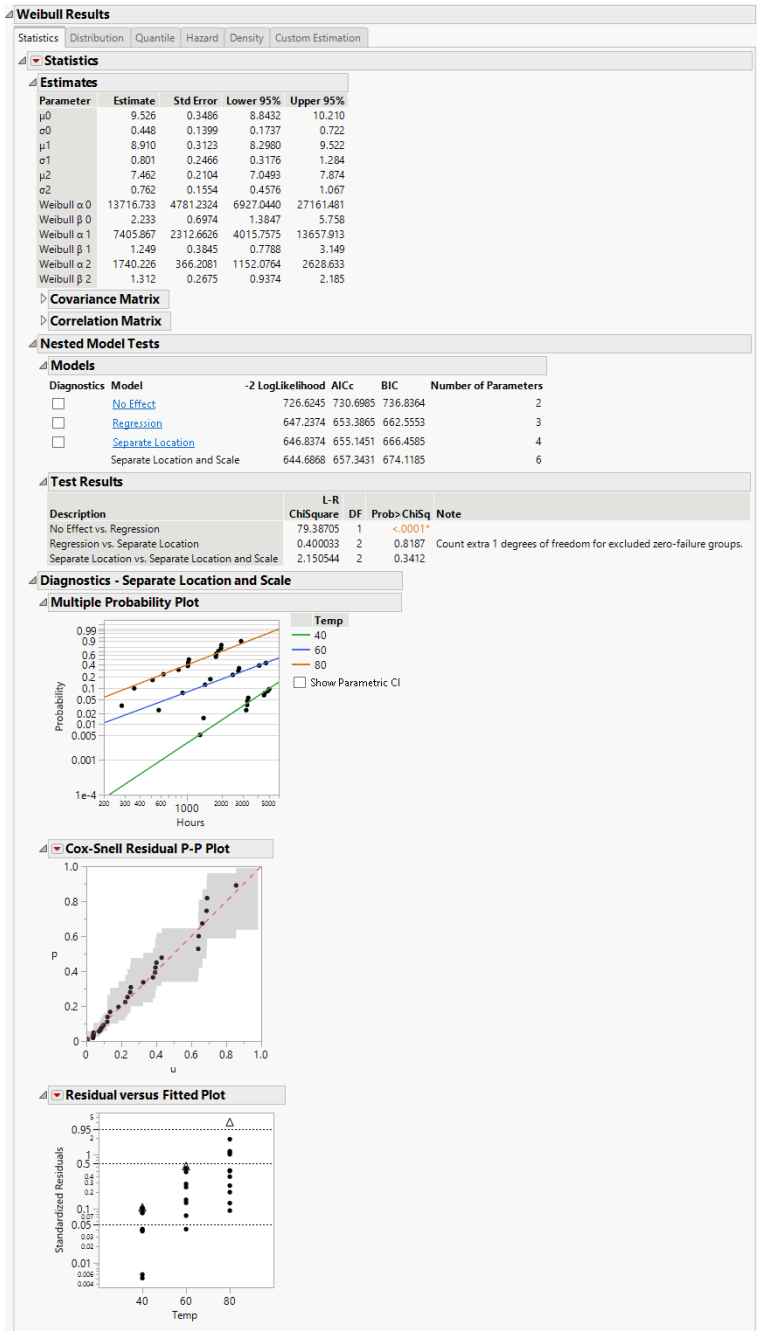


Figure 4.15 Separate Location Model with the Weibull Distribution for Devalt.jmp Data

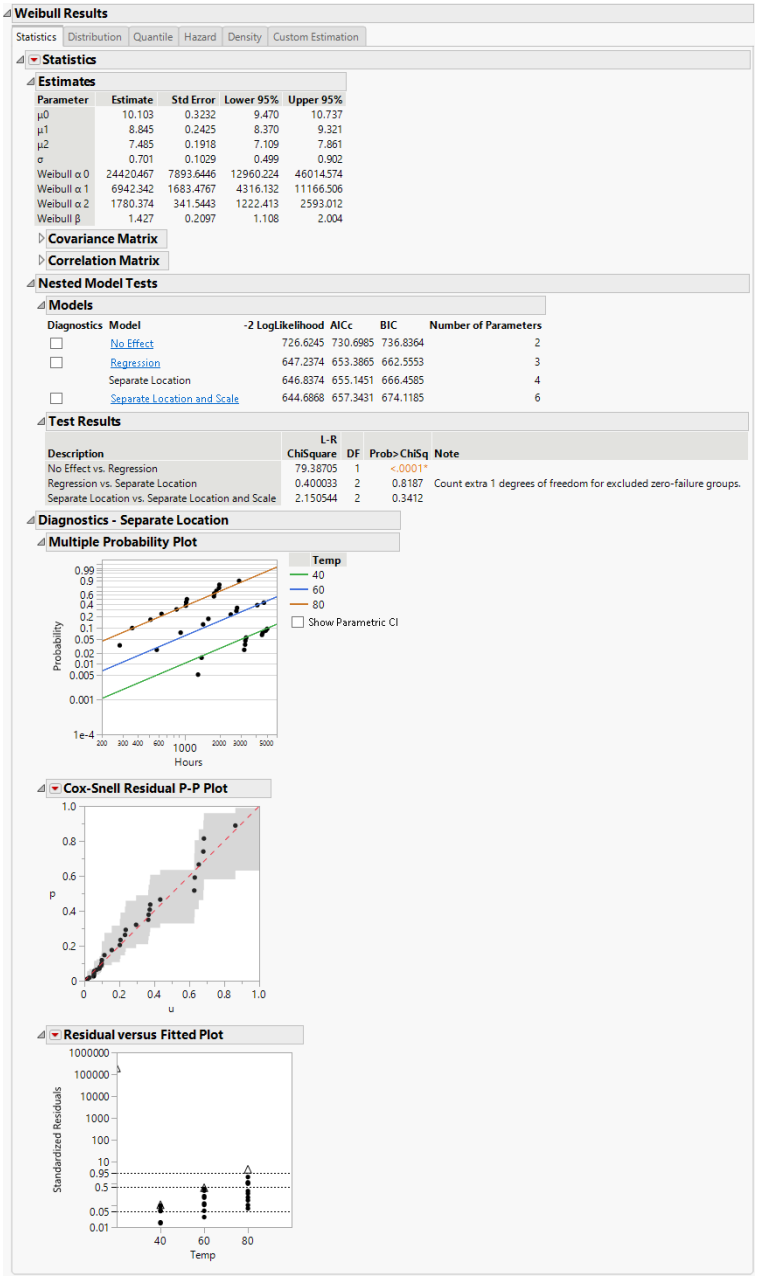


Figure 4.16 Regression Model with the Weibull Distribution for Devalt.jmp Data

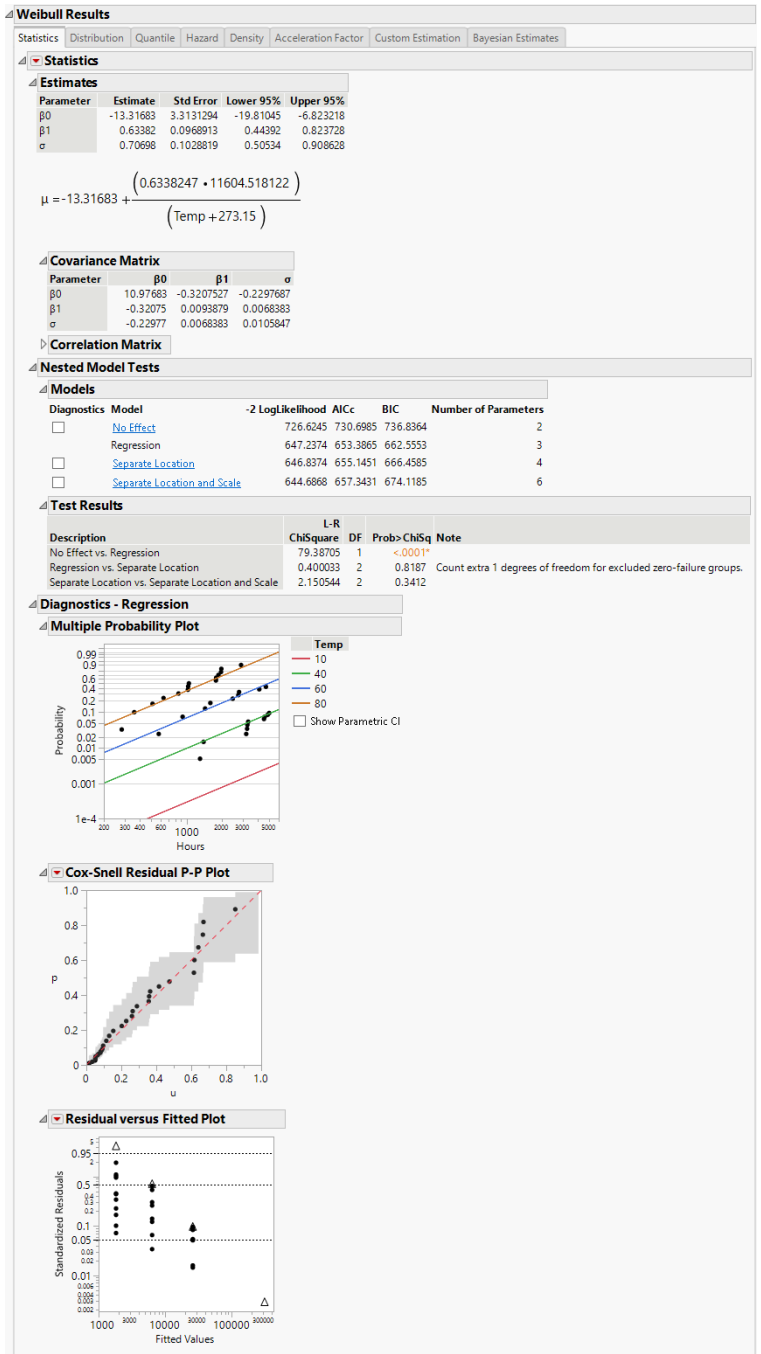
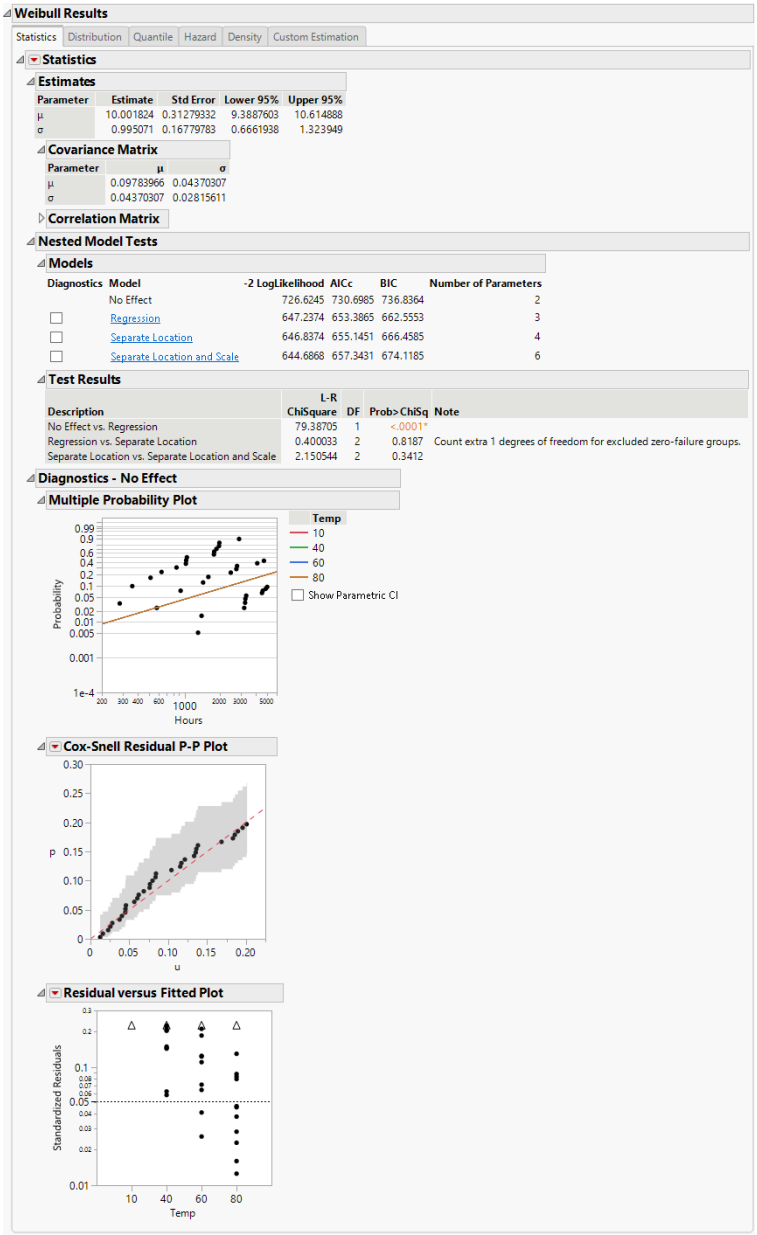


Figure 4.17 No Effect Model with the Weibull Distribution for Devalt.jmp Data

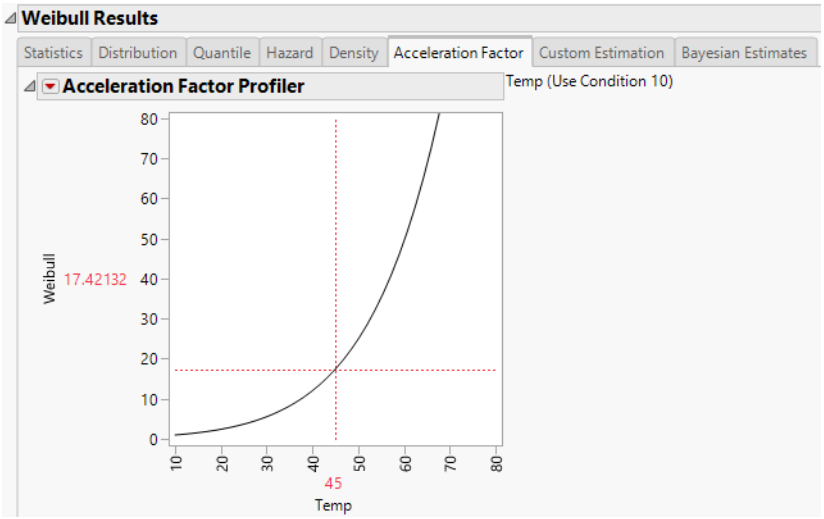


Profilers and Surface Plots

In addition to a statistical summary and diagnostic plots, the Fit Life by X report window also includes profilers and surface plots for each of your specified distributions. To view the Weibull time-accelerating factor and explanatory variable profilers, click the **Distribution** tab under Weibull Results. To see the surface plot, click the disclosure icon to the left of the Weibull title (under the profilers). The profilers and surface plot behave similarly to other platforms. See *Profilers*.

The report window also includes a tab labeled **Acceleration Factor**. Clicking the **Acceleration Factor** tab shows the Acceleration Factor Profiler. This profiler is an enlargement of the Weibull plot shown under the **Acceleration Factor** tab in the Comparisons section of the report window. Figure 4.18 shows the Acceleration Factor Profiler for the Weibull distribution of Devalt.jmp. The use condition level for the explanatory variable can be modified by selecting the **Set Time Acceleration Use Condition** option in the Fit Life by X red triangle menu.

Figure 4.18 Weibull Acceleration Factor Profiler for Devalt.jmp



Custom Estimation

For each parametric distribution, there is a Custom Estimation section that contains two reports: Estimate Quantile and Estimate Probability. For distributions with positive support, the Custom Estimation section also contains an Estimate Mean Remaining Life (MRLF) report.

Estimate Quantile

The Estimate Quantile report contains a calculator that enables you to predict quantiles for specific failure probability values. In the Estimate Quantile calculator, enter a value for Prob and the X variable. Press **Enter** to see the quantile estimates and corresponding confidence intervals. To calculate multiple quantile estimates, click the plus sign, enter another Prob value, another X value, or both, and press **Enter**. Click the minus sign to remove the last entry. If you enter more than one value in either column, the table contains all combinations of Prob and X values.

By default, Wald-based intervals are shown. Click **Likelihood CI** in the Interval Type outline to switch to likelihood-based confidence intervals. The confidence level for these intervals is determined by the **Change Confidence Level** option in the Fit Life by X red triangle menu.

Estimate Probability

The Estimate Probability report contains a calculator that enables you to predict failure and survival probabilities for specific time values. In the Estimate Probability calculator, enter a value for Time and the X variable. Press **Enter** to see the failure probability estimates and corresponding confidence intervals. To calculate multiple failure probability estimates, click the plus sign, enter another Time value, another X value, or both, and press **Enter**. Click the minus sign to remove the last entry. If you enter more than one value in either column, the table contains all combinations of Time and X values.

By default, Wald-based intervals are shown. Click **Likelihood CI** in the Interval Type outline to switch to likelihood-based confidence intervals. The confidence level for these intervals is determined by the **Change Confidence Level** option in the Fit Life by X red triangle menu.

Estimate Mean Remaining Life (MRLF)

The Estimate Mean Remaining Life (MRLF) report contains a calculator that enables you to predict mean remaining life for specific time values. In the Estimate Mean Remaining Life (MRLF) calculator, enter a value for Survival Time and the X variable. Press **Enter** to see the mean remaining life estimates and corresponding Wald-based confidence intervals. The confidence level for these intervals is determined by the **Change Confidence Level** option in the Fit Life by X red triangle menu. To calculate multiple mean remaining life estimates, click the plus sign, enter another Survival Time value, another X value, or both and press **Enter**. Click the minus sign to remove the last entry. If you enter more than one value in either column, the table contains all combinations of Survival Time and X values.

Note: When the survival time equals zero, the calculated mean remaining life (MRLF) value is equivalent to the mean time to failure (MTTF) value.

Bayesian Estimates

For each parametric distribution other than Exponential, there is a Bayesian Estimates section that enables you to obtain Bayesian parameter estimates. The Bayesian Estimates section is not available if the Relationship in the Fit Life by X launch window is Custom, No Effect, Location, or Location and Scale.

Bayesian estimation in the Fit Life by X platform is done using rejection sampling or a Markov Chain Monte Carlo (MCMC) algorithm. More specifically, the platform attempts a basic rejection sampler. If the rejection sampler produces valid results, these results are reported. If the rejection sampler cannot produce valid results, the platform uses a random walk Metropolis-Hastings algorithm and adds a note to the top of the Bayesian Estimation report. See Robert and Casella (2004).

The initial report is a control panel where you can specify prior distributions for the parameters and control aspects of the simulation. To obtain posterior estimates of the parameters, specify the prior distributions and the simulation options, and then click Fit Model.

To specify the prior distributions of the parameters, you must specify information about a quantile of the distribution and the slope β_1 and scale σ parameters. (For the Weibull distribution, you specify the Weibull β rather than σ.) The quantile is defined by two values: the probability of the quantile and the value of the X variable at the specified quantile. The default Probability value is 0.10, but you can specify a value that corresponds to the quantile of interest. Specify information about the range of the prior distribution. For Normal and Lognormal prior distributions, the range is specified in terms of Lower and Upper 99% limits. For Uniform and Log-Uniform prior distributions, the range is specified in terms of the Lower and Upper limits. See Meeker and Escobar (1998). The initial values that are provided are estimates consistent with the maximum likelihood estimates in the Statistics section of the report.

The following options for the simulation appear below the prior distribution specification table:

Number of Monte Carlo Iterations Sets the sample size that will be drawn from the posterior distribution after a burn-in procedure is completed.

Random Seed Sets the initial state of the simulation. By default, it is the clock time. The number should be a positive integer greater than 1. If you specify 1, the current clock time is used.

Bayesian Estimates - Result <N> Report

After you specify prior distributions and the simulation options, click **Fit Model** to perform the simulation. A Bayesian Estimates - Result <N> report is provided for each simulation. This report contains the following headings:

Priors Shows the specifications that you entered in the Bayesian Estimates report to run the simulation. The Prior report also contains the random seed.

Posterior Estimates Shows five marginal statistics that describe the posterior distribution of β_0, β_1, σ, and the quantile. The marginal statistics are the median, 0.025 quantile (Lower Bound), 0.975 quantile (Upper Bound), mean, and standard deviation computed from the Monte Carlo samples. If the Weibull distribution is specified, this table contains the posterior estimate of the Weibull β instead of σ.

To compute statistics for other derived variables based on the posterior estimates of the generic parameters, click the Export Monte Carlo Samples link.

Posterior Scatter Plot Shows two scatter plots of values from the Monte Carlo simulation. The scatter plot on the left shows the values of the posterior parameters as they are specified in the Priors report. The scatter plot on the right shows the values of the posterior parameters as they are specified in the Posterior Estimates report.

Profilers Shows two profilers based on samples from the posterior distribution. The values shown in the profilers, at the specified values of the X and Time variables, are calculated using the following steps:

- For each set of sampled parameter values from the posterior distribution, the values of the cumulative distribution function and the quantile function are calculated at the specified values of the X and Time variables.
- The predicted values of the cumulative distribution function and the quantile function are the medians of the calculated values.
- The upper and lower confidence limits are the 0.025 and 0.975 quantiles of the calculated values. The confidence level for these limits is determined by the **Change Confidence Level** option in the Fit Life by X red triangle menu.

Distribution Profiler Shows the parametric cumulative distribution function as a function of the X variable and Time.

Quantile Profiler Shows the parametric quantile function as a function of the X variable and a specified probability.

Bayesian Estimates - Result <N> Options

The Bayesian Estimates - Result <N> red triangle menu contains the following options:

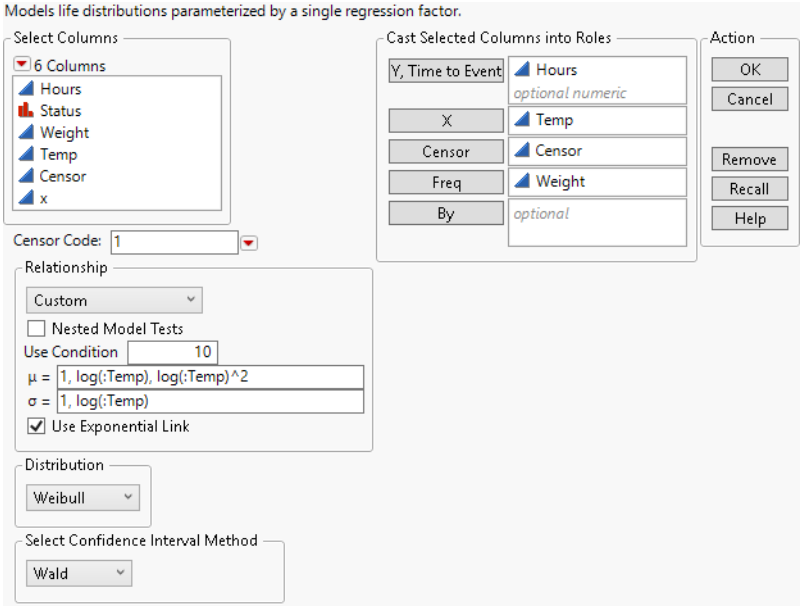
Remove Removes the current Bayesian Estimates report from the Fit Life by X report.

Export Monte Carlo Samples Saves the results of the Monte Carlo simulation to a new data table. You can use this table to compute statistics of the posterior estimates.

Custom Relationship

If you want to use a custom transformation to model the relationship between the lifetime event and the accelerating factor, use the **Custom** option. This option is found in the list under Relationship in the launch window. Enter comma delimited values into the entry fields for the location (μ) and scale (σ) parameters. For the Devalt.jmp sample data, an example entry for μ could be "1, log(:Temp), log(:Temp)^2", and an entry for σ could be "1, log(:Temp)", where 1 indicates that an intercept is included in the model. Select the **Use Exponential Link** check box to ensure that the sigma parameter is positive.

Figure 4.19 Custom Relationship Specification in Fit Life by X Launch Window



After selecting **OK**, location and scale transformations are created and included at the bottom of the Estimates report section.

Figure 4.20 Weibull Estimates and Formulas for Custom Relationship

Estimates

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
β0	-0.603333	9.9924940	-20.18826	18.98160
β1	8.047095	4.8377341	-1.43469	17.52888
β2	-1.412340	0.5989816	-2.58632	-0.23836
λ0	-2.317331	2.0821317	-6.39823	1.76357
λ1	0.474054	0.5015868	-0.50904	1.45715

$$\mu = -0.6033328 + 8.047095 \cdot \text{Log}(\text{Temp}) + -1.41234 \cdot \text{Log}(\text{Temp})^2$$

$$\sigma = \text{Exp}\left(-2.317331 + 0.4740541 \cdot \text{Log}(\text{Temp})\right)$$

For an example of how to use a custom transformation, see "Custom Relationship Example". Analysis proceeds similarly to the "Example of the Fit Life by X Platform", where the **Arrhenius Celsius** Relationship was specified.

Fit Life by X Platform Options

The Fit Life by X red triangle menu contains the following options:

Fit Lognormal Fits a lognormal distribution to the data.

Fit Weibull Fits a Weibull distribution to the data.

Fit Loglogistic Fits a loglogistic distribution to the data.

Fit Fréchet Fits a Fréchet distribution to the data.

Fit Exponential Fits an exponential distribution to the data.

Fit SEV Fits a smallest extreme value (SEV) distribution to the data.

Fit Normal Fits a normal distribution to the data.

Fit Logistic Fits a logistic distribution to the data.

Fit LEV Fits a largest extreme value (LEV) distribution to the data.

Fit All Distributions Fits all distributions to the data.

Set Time Acceleration Use Condition Enables you to enter a use condition value for the explanatory variable of the acceleration factor in a pop-up window.

Change Confidence Level Enables you to enter a desired confidence level, for the plots and statistics, in a pop-up window. The default confidence level is 0.95.

Tabbed Report Lets you specify how you want the report window displayed. Two options are available: Tabbed Overall Report and Tabbed Individual Report. Tabbed Individual Report is checked by default. You can select one, both or none.

Show Surface Plot Shows or hides the surface plot for the distribution on and off in the individual distribution results section of the report. The surface plot is shown in the Distribution, Quantile, Hazard, and Density sections for the individual distributions, and it is on by default.

Show Points Shows or hides the data points on and off in the Nonparametric Overlay plot and in the Multiple Probability Plots. The points are shown in the plots by default. If this option is unchecked, step functions are shown instead.

Scatterplot Shows a scatterplot of a lifetime event versus an explanatory variable.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Additional Examples of the Fit Life by X Platform

- "Capacitor Example"
- "Custom Relationship Example"

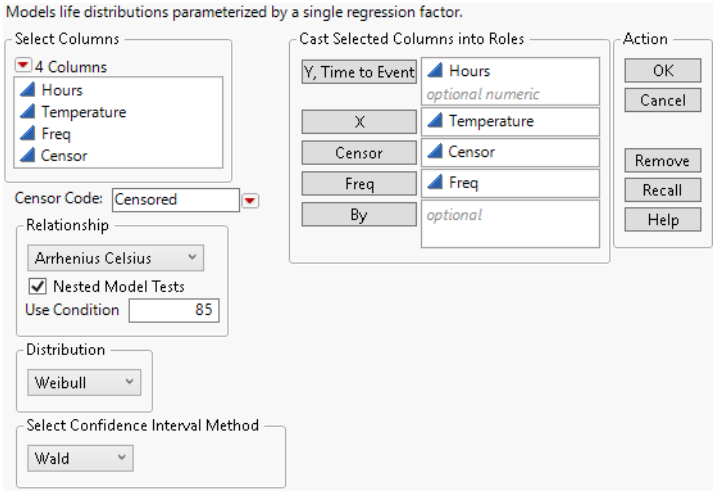
Capacitor Example

This example uses Capacitor ALT.jmp and can be found in the Reliability folder of the sample data. It contains simulated data that gives censored observations for three levels of temperature, for a reliability study. Observations are shown as censored values at temperature levels of 85, 105, and 125 degrees Celsius.

1. Select **Help > Sample Data Library** and open Reliability/Capacitor ALT.jmp.

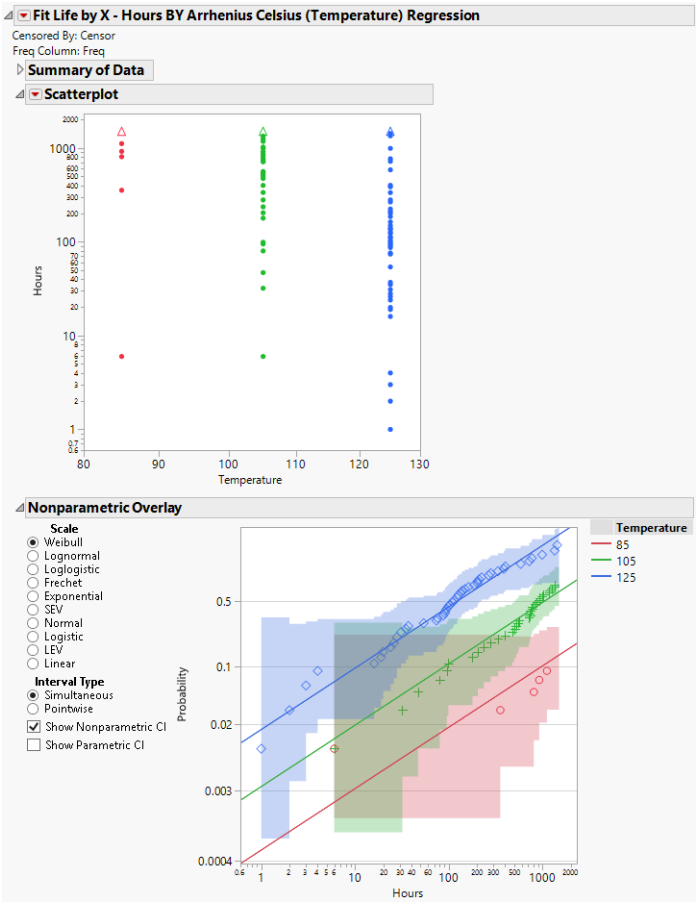
2. Select **Analyze** > **Reliability and Survival > Fit Life by X**.
3. Select Hours and click **Y, Time to Event**.
4. Select Temperature and click **X**.
5. Select Censor and click Censor.
6. Leave the **Censor Code** as Censored.
7. Select Freq as **Freq**.
8. Keep **Arrhenius Celsius** as the relationship, and keep the **Nested Model Tests** option selected.
9. Select **Weibull** as the distribution.
10. Keep **Wald** as the confidence interval method.

Figure 4.21 Fit Life by X Launch Window



11. Click **OK**.

Figure 4.22 Fit Life by X Report Window for Capacitor ALT.jmp Data



The report window shows summary data, diagnostic plots, comparison data and results, including detailed statistics and prediction profilers. Separate result sections are shown for each selected distribution. Distribution, Quantile, Hazard, Density, and Acceleration Factor Profilers are included for each of the specified distributions.

Custom Relationship Example

To create a quadratic model with Log(Temp) for the Weibull location parameter and a log-linear model with Log(Temp) for the Weibull scale parameter, follow these steps:

1. Select **Help > Sample Data Library** and open Reliability/Devalt.jmp.
2. Select **Analyze** > **Reliability and Survival** > **Fit Life by X**.
3. Select Hours and click **Y, Time to Event**.
4. Select Temp and click **X**.

5. Select Censor and click **Censor**.
6. Select Weight and click **Freq**.
7. Select **Custom** as the Relationship from the list.
8. In the entry field for μ, enter 1, log(:Temp), log(:Temp)^2.

   (The 1 indicates that an intercept is included in the model.)
9. In the entry field for σ, enter 1, log(:Temp).
10. Select the check box for **Use Exponential Link**.
11. Deselect the check box for **Nested Model Tests**.
12. In the entry field for Use Condition, enter 10.
13. Select **Weibull** as the Distribution.

    Figure 4.23 shows the completed launch window using the **Custom** option.

Note: The Nested Model Tests check box is not checked for non-constant scale models. Nested Model test results are not supported for this option.

14. Click **OK**.

Figure 4.23 Custom Relationship Specification in Fit Life by X Launch Window

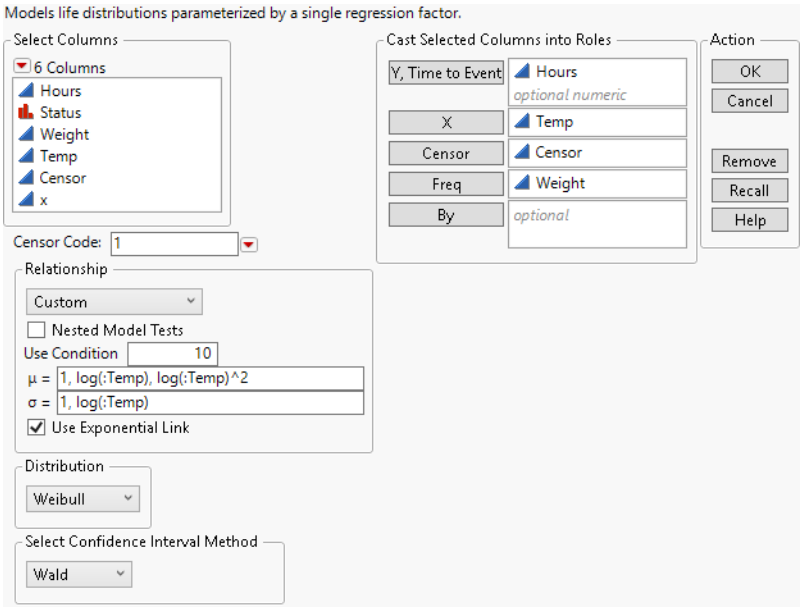


Figure 4.24 shows the location and scale transformations, which are subsequently created and included at the bottom of the Estimates report section.

Analysis proceeds similarly to the "Example of the Fit Life by X Platform", where the **Arrhenius Celsius** Relationship was specified.

Figure 4.24 Weibull Estimates and Formulas for Custom Relationship

Estimates

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
β0	-0.603333	9.9924940	-20.18826	18.98160
β1	8.047095	4.8377341	-1.43469	17.52888
β2	-1.412340	0.5989816	-2.58632	-0.23836
λ0	-2.317331	2.0821317	-6.39823	1.76357
λ1	0.474054	0.5015868	-0.50904	1.45715

$$\mu = -0.6033328 + 8.047095 \cdot \text{Log}\left(\text{Temp}\right) + -1.41234 \cdot \text{Log}\left(\text{Temp}\right)^2$$

$$\sigma = \text{Exp}\left(-2.317331 + 0.4740541 \cdot \text{Log}\left(\text{Temp}\right)\right)$$

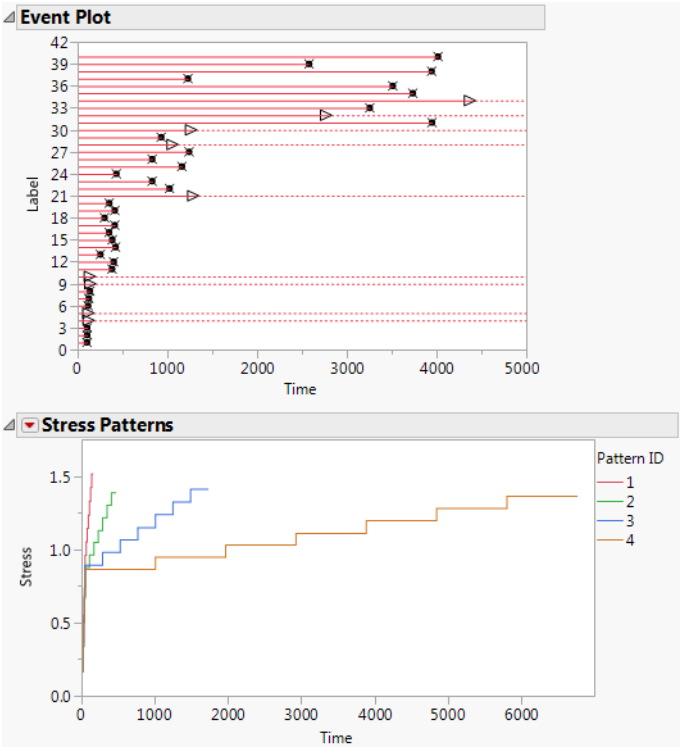
Chapter 5

Cumulative Damage

Model Product Deterioration with Variable Stress over Time

Cumulative damage models, which include step-stress models, enable you to analyze an accelerated life test where the stress levels might be changed over time. The stress can be applied by many different forces: load, temperature, pressure. A typical cumulative damage experiment consists of multiple test units. Each unit has an initial stress level, and the stress level can be changed throughout the experiment. The response is the failure time or time-to-event. The platform plots the failure events and enables you to fit multiple distributions to your data.

Figure 5.1 Example of Cumulative Damage Report



Contents

Overview of the Cumulative Damage Platform 139
Example of the Cumulative Damage Platform 139
Launch the Cumulative Damage Platform 143
Time to Event 144
Stress Pattern 145
The Cumulative Damage Report 148
Event Plot 148
Stress Patterns Report 148
Model List 149
Model Results 149
Cumulative Damage Platform Options 150
Stress Patterns Options 150
Additional Example of the Cumulative Damage Platform 152
Example of Simulating New Data 152

Overview of the Cumulative Damage Platform

A cumulative damage experiment, also called a *varying-stress experiment*, is an accelerated life test where the stress levels can change over time. The stress can be applied by many different forces: load, temperature, pressure. A typical cumulative damage experiment consists of multiple test units. Each unit has an initial stress level, and the stress level can be changed throughout the experiment.

The most common cumulative damage experiment is a *step-stress experiment*. A step-stress experiment uses multiple units with varying levels of stress applied. Stress can be applied using factors such as temperature, pressure, or voltage. For each unit, there is an initial stress level. At specified time points, the stress levels are adjusted based on different patterns of stress levels. Between stress level changes, the stress level remains constant.

The Cumulative Damage platform also includes three other varying-stress pattern models:

- In a ramp-stress experiment, the stress levels start at an initial value and then increase linearly over time at a specified slope.
- In a sinusoid-stress experiment, the stress levels fluctuate in a periodic fashion that is defined by a sine wave.
- In a piecewise ramp-stress experiment, the stress levels are defined at specified time points similar to the step-stress case. However, the stress level is not required to stay constant between time points. Rather, it changes linearly from a starting stress level to an ending stress level between time points. If a pair of starting and ending stress levels are equal, the interval is equivalent to a step-stress interval.

For more information about varying-stress and step-stress models, see Nelson (2004, ch. 10).

Example of the Cumulative Damage Platform

A step-stress experiment is conducted on 40 test units at varying stress conditions. Your goal is to estimate the probability of failure at 10,000 time units given a stress level of 0.75. These data tables are based on data from Nelson (2004, ch. 10).

The Reliability/CD Step Stress Pattern.jmp data table contains a column called Pattern ID that identifies four different stress patterns. The stress level at a particular step is the ratio of Voltage to Thickness. (Note that these two columns are hidden.) Thickness is held constant for each stress pattern. However, Voltage is set to different levels and increases within each pattern.

The Reliability/CD Step Stress.jmp data table contains the time to failure data.

1. Select **Help > Sample Data Library** and open Reliability/CD Step Stress.jmp and Reliability/CD Step Stress Pattern.jmp.

   The CD Step Stress table contains failure time data:

   – The Time column gives the failure times.

   – The Pattern ID column identifies the stress pattern.

   – The Censor column indicates whether the failure time is exact or censored.

   Each row of the table corresponds to one test unit.

   The CD Step Stress Pattern table contains the four stress patterns (identified as 1 through 4). The levels of the stress factor, Stress, are varied within each value of the Pattern ID column. The Duration column represents how many time units a particular level of the stress factor lasted.

2. Select **Analyze > Reliability and Survival > Cumulative Damage**.

   The launch window has two sections: one for the failure time data (Time-to-Event) and one for the stress pattern data (Stress Pattern).

3. Click **Select Table** in the Time-to-Event panel.

   A Time-to-Event Data Table window appears, which prompts you to specify the data table for the failure time data.

4. Select CD Step Stress and click **OK**.

   The columns from this table now populate the **Select Columns** list in the Time-to-Event panel.

5. Select Time and click **Time to Event**.

6. Select Censor and click **Censor**.

7. Select Pattern ID for **Pattern ID**.

8. Click **Select Table** in the Stress Pattern panel.

9. Select CD Step Stress Pattern and click **OK**.

   The columns from this table now populate the **Select Columns** list in the Stress Pattern panel.

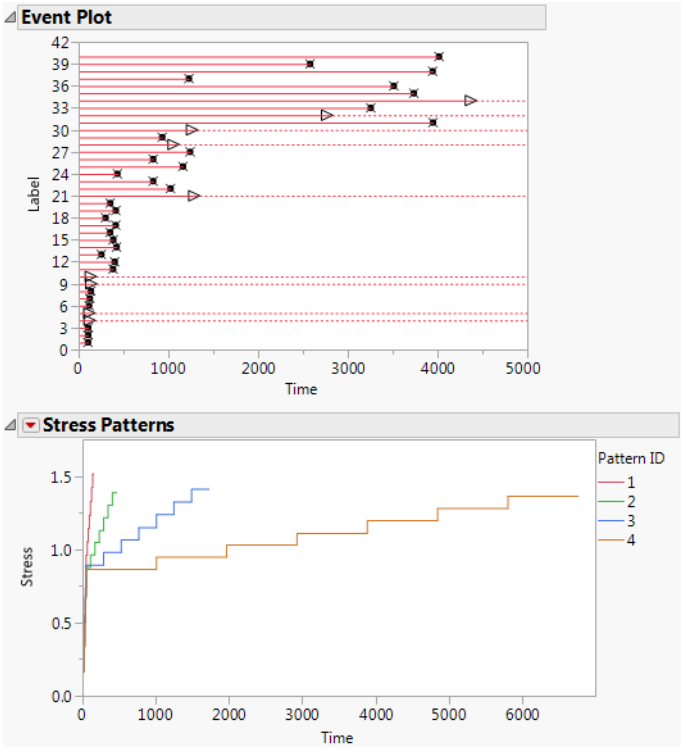
10. Select Duration and click **Stress Duration**.

11. Select Stress and click **Stress**.

12. Select Pattern ID and click **Pattern ID**.

13. Click **OK**.

Figure 5.2 Event Plot and Stress Patterns Plot



The initial report contains the Event Plot and a plot of the defined stress patterns. All four stress patterns increase the stress level quickly over the first 40 time units, after which they increase at much different rates.

14. Click the Cumulative Damage red triangle and select **Fit All**.

Figure 5.3 Model List Report

Model List

Distribution	-2 Log Likelihood	Nparm	AICc	BIC
Exponential	398.48312	2	402.80745	405.86088
Weibull	398.4822	3	405.14887	409.54884
Loglogistic	400.93264	3	407.59931	411.99928
Lognormal	402.3488	3	409.01547	413.41544
Frechet	410.92577	3	417.59243	421.99241

From the Model List report, you determine that the best fitting distribution is the Exponential distribution.

15. In the Results report, scroll to the Exponential report.
16. In the Distribution Profiler report, set the current value of Stress to 0.75.
17. Set the current value of Time to 10000.

Figure 5.4 Distribution Profiler for Exponential Distribution at Specified Settings

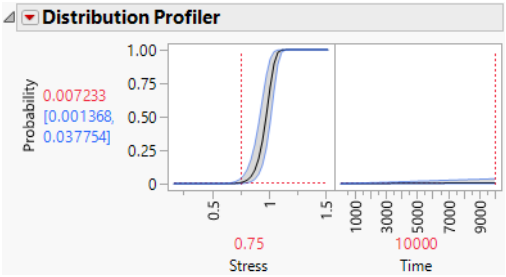
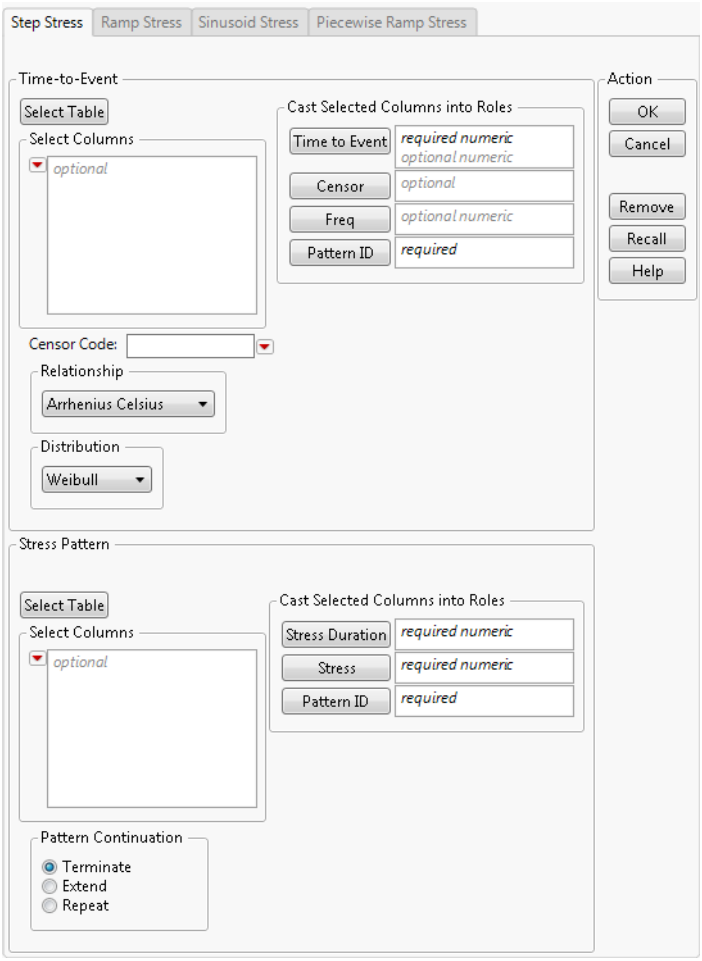


Figure 5.4 shows that the predicted probability of failure for a test unit under constant stress of 0.75 at 10000 time units is 0.007233, with a 95% confidence interval of 0.001368 to 0.037754.

Launch the Cumulative Damage Platform

Launch the Cumulative Damage platform by selecting **Analyze > Reliability and Survival > Cumulative Damage**. You must supply the Cumulative Damage platform with two data tables as input. The first data table is time-to-event data for each unit under test. The second data table defines the stress patterns used for each unit.

Figure 5.5 The Cumulative Damage Launch Window



The launch window includes a separate tab for each step-stress data format. For more information about the various stress patterns, see "Stress Pattern". For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Each of the step-stress format tabs contains two panels for specifying variables for the model:

- The Time-to-Event panel is common to all of the step-stress formats. This panel is similar to the Fit Life by X platform launch window.
- The Stress Pattern panel is used to describe the type of stress. Consequently, its format depends on the selected step-stress tab.

Time to Event

The Time to Event panel in the Cumulative Damage launch window contains the following options:

Time to Event Identifies the time to event (such as the time to failure) or time to censoring. For interval censoring, your data table should contain two columns, where one gives the lower bound and the other gives the upper bound of the failure time for each unit. Enter the two censoring columns as Time to Event. For more information about censoring, see "Event Plot".

Censor Identifies right-censored observations. Select the value that identifies right-censored observations from the Censor Code menu beneath the Select Columns list. The Censor column is used only when one Y is entered.

Freq Identifies frequencies or observation counts when there are multiple units. If the value is 0 or a positive integer, then the value represents the frequencies or counts of observations with the given row's settings.

Pattern ID Contains values that specify the stress pattern in the Stress Pattern data table that was used for the given row.

Censor Code Identifies the value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

Relationship Identifies the relationship between the event and the stress factor. Table 5.1 defines the model for each relationship.

Table 5.1 Models for Relationship Options

Relationship	Model
Arrhenius Celsius	$\mu = b0 + b1 * 11604.5181215503 / (X + 273.15)$

Table 5.1 Models for Relationship Options *(Continued)*

Relationship	Model
Arrhenius Fahrenheit	μ = b0 + b1 * 11604.5181215503 / ((X + 459.67) / 1.8)
Arrhenius Kelvin	μ = b0 + b1 * 11604.5181215503 / X
Inverse Power (default)	μ = b0 + b1 * log(X)
Linear	μ = b0 + b1 * X
Log	μ = b0 + b1 * log(X)
Logit	μ = b0 + b1 * log(X / (1 - X))
Reciprocal	μ = b0 + b1 / X
Square Root	μ = b0 + b1 * sqrt(X)
Box-Cox	μ = b0 + b1 * BoxCox(X)
Custom	user-defined (Available only in the Step Stress panel.)

The BoxCox(X) transformation is defined as follows:

$$x_i^{(\lambda)} = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x_i) & \text{if } \lambda = 0 \end{cases}$$

If you select **Custom**, additional controls appear that require you to define the Custom transformation that models the relationship between the lifetime event and the stress factor.

If you want to use a Custom relationship for your model, see "Custom Relationship".

Distribution Specifies an initial time-to-failure distribution. Select from Weibull, Lognormal, Loglogistic, Fréchet, or Exponential. Weibull is the default setting. For more information about the distributions, see "Distributions".

Stress Pattern

Specify the stress patterns used in the experiment using the second panel.

Step Stress Pattern

The Step Stress pattern has stress levels that are changed at arbitrary time points. The duration of each stress step and associated stress level must be specified in ascending time order.

The Stress Pattern panel in the Step Stress tab contains the following options:

Stress Duration The column that contains the length in time units of each stress step.

Stress The column that contains the level of the stress setting.

Pattern ID The column that contains a unique identifier for the stress pattern. This column is used to match stress patterns in the Stress Pattern data table with the observations in the Time-to-Event data table.

Pattern Continuation Specifies how to handle failures that occur after the final time period in the defined stress pattern. This panel contains the following options:

Terminate A failure that occurs at a time beyond the final time period in the defined stress pattern produces an error, and the model is not fit.

Extend A failure that occurs at a time beyond the final time period in the defined stress pattern assumes the same stress level as the level in the final time period.

Repeat A failure that occurs at a time beyond the final time period in the defined stress pattern assumes the same stress level as if the stress pattern were being repeated. For example, if a failure occurs 10 time units after the final time period in the defined stress pattern, then the stress level at that failure time is set to the stress level at 10 time units after the beginning of the defined stress pattern.

Note: The default Pattern Continuation setting for the Step Stress Pattern is Terminate.

Ramp Stress Pattern

The Ramp Stress pattern defines stress as a linear function of time. Each pattern is defined by an intercept (the stress level at time zero) and a slope (the increase in the stress level for every one time unit). Each pattern is described in a single row in the stress pattern data table.

Intercept The column that contains the intercept for each pattern.

Slope The column that contains the slope for each pattern.

Pattern ID The column that contains a unique identifier for the stress pattern. This column is used to match stress patterns in the Stress Pattern data table with the observations in the Time-to-Event data table.

Sinusoid Stress Pattern

The Sinusoid Stress pattern defines stress as a periodic function. The pattern is defined by a level, an amplitude, a period, and a phase. Each pattern is described in a single row in the stress pattern data table. The pattern is defined as follows:

$$S(t) = \text{level} + \text{amplitude} * \sin(\text{phase} + (2*\pi*t) / \text{period})$$

Level The column that contains the level for each pattern.

Amplitude The column that contains the amplitude for each pattern.

Period The column that contains the period for each pattern.

Phase The column that contains the phase for each pattern.

Pattern ID The column that contains a unique identifier for the stress pattern. This column is used to match stress patterns in the Stress Pattern data table with the observations in the Time-to-Event data table.

Piecewise Ramp Stress Pattern

The Piecewise Ramp Stress pattern defines stress as a piecewise linear function of time. The line segments for the stress level over time can be disjoint or continuous. Line segments can also be flat, so that step stress and ramp stress can be combined. The line segments are defined in the stress pattern data table by the time duration of the segment and the start and end levels of the stress setting.

Stress Duration The column that contains the length in time units of each stress step.

Stress Ramp The two columns that contain the stress levels at the start and end of the step.

Pattern ID The column that contains a unique identifier for the stress pattern. This column is used to match stress patterns in the Stress Pattern data table with the observations in the Time-to-Event data table.

Pattern Continuation The Pattern Continuation panel enables you to specify the stress levels that occur after the final time period in the defined stress pattern. This panel contains the following options:

Terminate A failure that occurs at a time beyond the final time period in the defined stress pattern produces an error, and the model is not fit.

Extend A failure that occurs at a time beyond the final time period in the defined stress pattern assumes the same stress level as the level in the final time period.

Repeat A failure that occurs at a time beyond the final time period in the defined stress pattern assumes the same stress level as if the stress pattern were being repeated. For example, if a failure occurs 10 time units after the final time period in the defined stress pattern, then the stress level at that failure time is set to the stress level at 10 time units after the beginning of the defined stress pattern.

Note: The default Pattern Continuation setting for the Piecewise Ramp Stress Pattern is Terminate.

The Cumulative Damage Report

After you click OK, the Cumulative Damage report window appears. By default, the Cumulative Damage report contains an event plot, a stress patterns report, a model list, and model results.

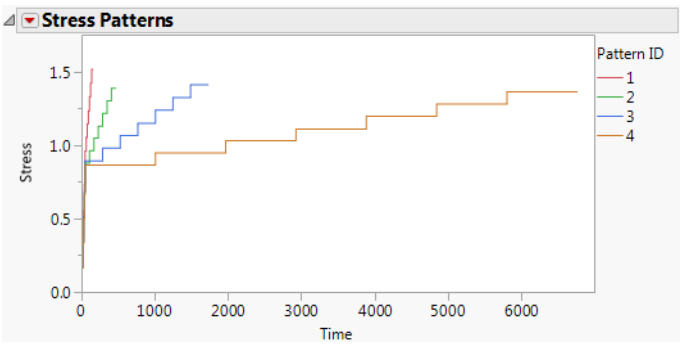
Event Plot

The Event Plot in Cumulative Damage displays time to failure or censoring. See "Event Plot".

Stress Patterns Report

The Stress Patterns report shows a plot of the stress level over time for each of the stress pattern IDs. The Simulate option in this report enables you to simulate new data. Figure 5.6 shows an example of the Stress Patterns report plot. See "Stress Patterns Options".

Figure 5.6 Stress Patterns Report



Model List

The Model List report provides the -2Loglikelihood, number of parameters, AICc, and BIC statistics for each fitted distribution. Smaller values of each of these statistics (other than number of parameters) indicate a better fit. For more information about these statistics, see *Fitting Linear Models*.

Model Results

The Results report contains a separate report for each fitted distribution. Each report contains the following:

Parameter Estimates Shows the estimates, standard errors, and Wald-based 95% confidence intervals.

The fitted equation appears below the Parameter Estimates table. This equation takes into account the fitted parameter estimates and the relationship specified in the launch window.

Distribution Profiler Shows cumulative failure probability as a function of time.

Quantile Profiler Shows failure time as a function of cumulative probability.

Hazard Profiler Shows the hazard rate as a function of time.

Density Profiler Shows the density function for the distribution.

Probability Plot of Standardized Residuals Shows a plot of standardized residuals for the model on a probability scale axis.

The Probability Plot of Standardized Residuals red triangle menu has a Save Residuals option that saves the standardized residuals to three new columns in the failure time data table. The three columns are Left Residuals, Right Residuals, and Residual Weight.

Cox-Snell Residual P-P Plot Shows a residual plot that is used to validate the distributional assumption for the data.

The Cox-Snell Residual P-P Plot red triangle menu has a Save Residuals option that saves the Cox-Snell residuals to three new columns in the failure time data table. The three columns are Left Residuals, Right Residuals, and Residual Weight. See Meeker and Escobar (1998, sec. 17.6.1) for a discussion of Cox-Snell residuals.

Cumulative Damage Platform Options

The Cumulative Damage red triangle menu contains the following options:

Fit All Fits the distributions that were not selected in the launch window. The available distributions are listed under "Distribution".

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

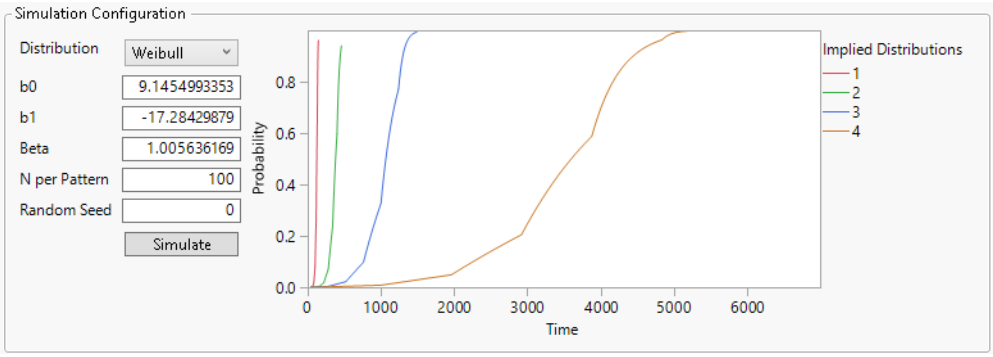
Stress Patterns Options

The Simulate option in the Stress Pattern red triangle menu shows or hides the Simulation Configuration panel.

The Simulation Control Panel

Figure 5.7 shows the Simulation Configuration Panel. The initial values for Distribution and the parameter settings are determined by the parameter estimates for the distribution specified in the Cumulative Damage launch window. The graph shows the estimated failure distribution functions over time for the stress patterns defined in the Stress Pattern data table.

Figure 5.7 Simulation Configuration Panel



The Simulation Configuration panel enables you to simulate new failure time data based on a distribution and stress pattern. The stress pattern defined in the Stress Pattern data table used in the launch of the platform is also used for the simulation. This panel contains the following options:

Distribution The distribution to be used for the simulation. The available distributions are the same as in the Cumulative Damage launch window. For more information about the distributions, see "Distributions".

b0 The intercept for the location parameter of the distribution.

b1 The slope for the location parameter of the distribution.

lambda (Available only for the Box-Cox relationship.) The lambda value for the Box-Cox relationship.

b2, s0, s1, and so on (Available only for the Custom relationship.) Other parameters that are defined in the Custom relationship in the Cumulative Damage launch window.

Beta (Available only for the Weibull distribution.) The Beta parameter of the Weibull distribution.

sigma (Available only for the Lognormal, Loglogistic, and Fréchet distributions.) The sigma parameter of the distribution.

N per Pattern The number of points generated in the simulation for each stress pattern.

Random Seed (Optional) A nonzero random seed that ensures the reproducibility of simulation results.

Termination (Not available when the specified Pattern Continuation in the Cumulative Damage launch window is Terminate.) A time beyond which surviving test units are censored.

Simulate

The plot in the Simulation Configuration panel shows the implied distributions for each of the stress patterns over time. Click the Simulate button to generate a new JMP data table that contains the results of the simulation.

Additional Example of the Cumulative Damage Platform

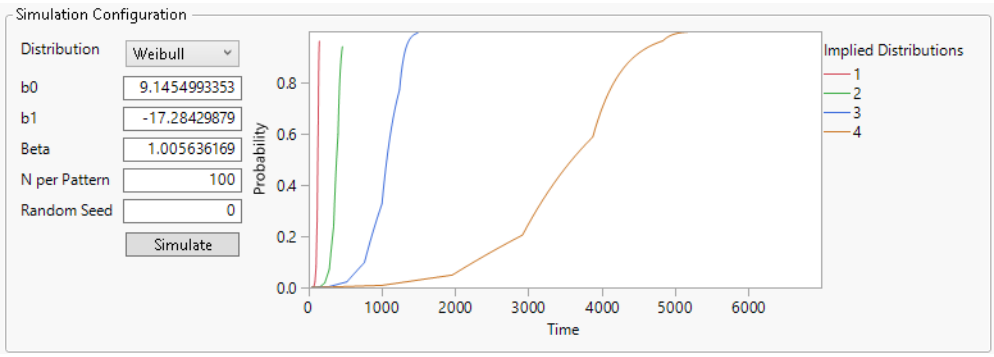
This section contains additional examples using the Cumulative Damage platform.

Example of Simulating New Data

This example illustrates using the Simulation Configuration panel in the Cumulative Damage report window to generate new step stress data. This example uses the same data as the example in "Example of the Cumulative Damage Platform".

1. Select **Help > Sample Data Library** and open Reliability/CD Step Stress.jmp and Reliability/CD Step Stress Pattern.jmp.
2. In the CD Step Stress data table, run the script Cumulative Damage.
3. Click the Stress Patterns red triangle and select **Simulate**.

Figure 5.8 Simulation Configuration Panel



The Simulation Configuration panel appears in the Stress Patterns report. The selection for Distribution is Weibull. The fitted values for b0 and b1 are used as initial values for the simulation.

4. Select Exponential for **Distribution**.
5. Enter 10 for **b0**.

6. Enter -18 for **b1**.
7. (Optional) Enter 14678 for **Random Seed**.
8. Click **Simulate**.

Figure 5.9 Partial Results of Simulation

	Time Left	Time Right	Pattern ID
1	140.67948106	140.67948106	1
2	139.34770497	139.34770497	1
3	136.95430716	136.95430716	1
4	136.30050091	136.30050091	1
5	145	•	1
6	118.5123739	118.5123739	1
7	119.03562567	119.03562567	1
8	145	•	1
9	127.19420127	127.19420127	1
10	118.48391055	118.48391055	1
11	103.54004804	103.54004804	1
12	145	•	1
13	121.04292162	121.04292162	1
14	136.15486483	136.15486483	1
15	143.65666248	143.65666248	1
16	123.04501594	123.04501594	1
17	145	•	1
18	124.26692664	124.26692664	1
19	120.96407573	120.96407573	1

Figure 5.9 shows a partial listing of the simulated data table. The stress pattern with Pattern ID equal to 1 is defined only up to 145 time units. Since the Pattern Configuration setting in the launch window was set to Terminate, the simulation censors any simulated values at 145 for stress pattern 1.

Chapter 6

Recurrence Analysis

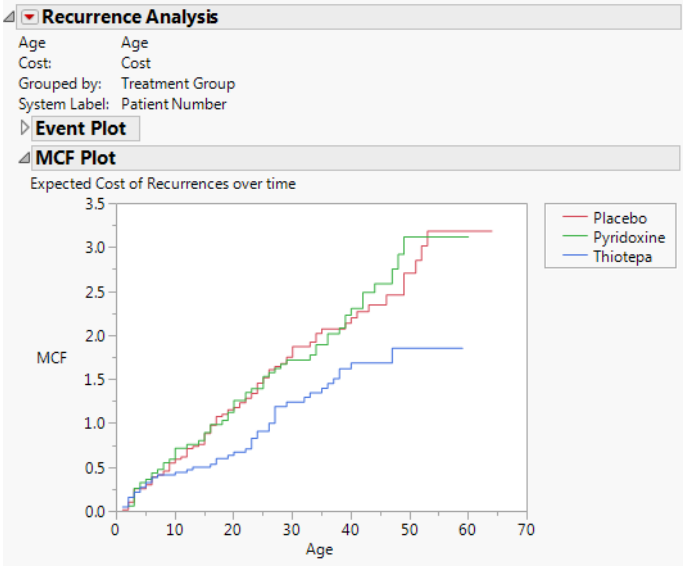
Model the Frequency or Cost of Recurrent Events over Time

The Recurrence Analysis platform analyzes event times, where the events can recur several times for each unit, item, or person. In an industrial setting, these events can occur when a unit breaks down, is repaired, and then put back into service after the repair. The units are followed until they are ultimately taken out of service.

In a medical setting, recurrence analysis can be used to analyze data from continuing treatments of a long-term disease, such as the recurrence of tumors in patients receiving treatment for bladder cancer.

The goal of the analysis is to obtain the mean cumulative function (MCF), which shows the total cost per unit as a function of time. Cost can be a count of the number of repairs, or it can be the actual cost of repair.

Figure 6.1 Recurrence Analysis Example



Contents

Overview of the Recurrence Analysis Platform 157
Example of the Recurrence Analysis Platform 157
Launch the Recurrence Analysis Platform 160
Recurrence Analysis Platform Options 161
Fit Model 162
Additional Examples of the Recurrence Analysis Platform 166
Bladder Cancer Recurrences Example 167
Diesel Ship Engines Example 170

Overview of the Recurrence Analysis Platform

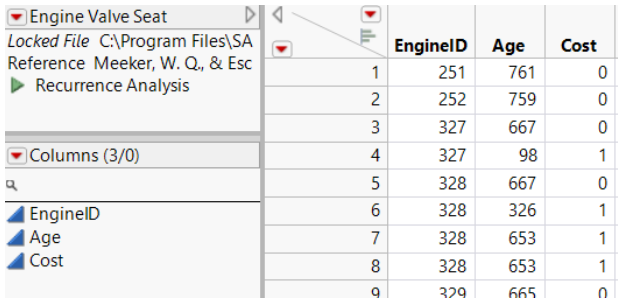
Recurrent event data involves the cumulative frequency or cost of repairs as units age. In JMP, the Recurrence Analysis platform analyzes recurrent events data.

The data for recurrence analysis have one row for each observed event and a closing row with the last observed age of a unit. Any number of units or systems can be included. In addition, these units or systems can include any number of recurrences.

Example of the Recurrence Analysis Platform

A typical unit might be a system, such as a component of an engine or appliance. For example, consider the sample data table Engine Valve Seat.jmp, which records valve seat replacements in locomotive engines. See Meeker and Escobar (1998, p. 395) and Nelson (2003). A partial listing of this data is shown in Figure 6.2. The EngineID column identifies a specific locomotive unit. Age is time in days from beginning of service to replacement of the engine valve seat. Note that an engine can have multiple rows with its age at each replacement and its cost, corresponding to multiple repairs. Here, Cost=0 indicates the last observed age of a locomotive.

Figure 6.2 Partial Engine Valve Seat Data Table

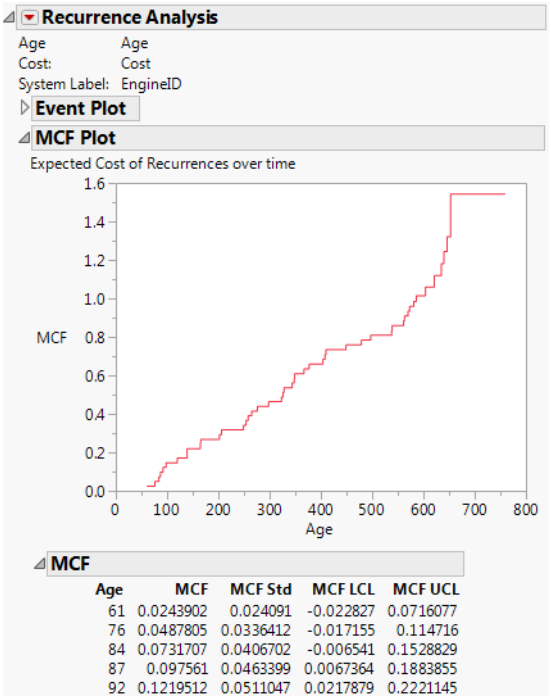


	EngineID	Age	Cost
1	251	761	0
2	252	759	0
3	327	667	0
4	327	98	1
5	328	667	0
6	328	326	1
7	328	653	1
8	328	653	1
9	329	665	0

Complete the launch window as shown in Figure 6.5.

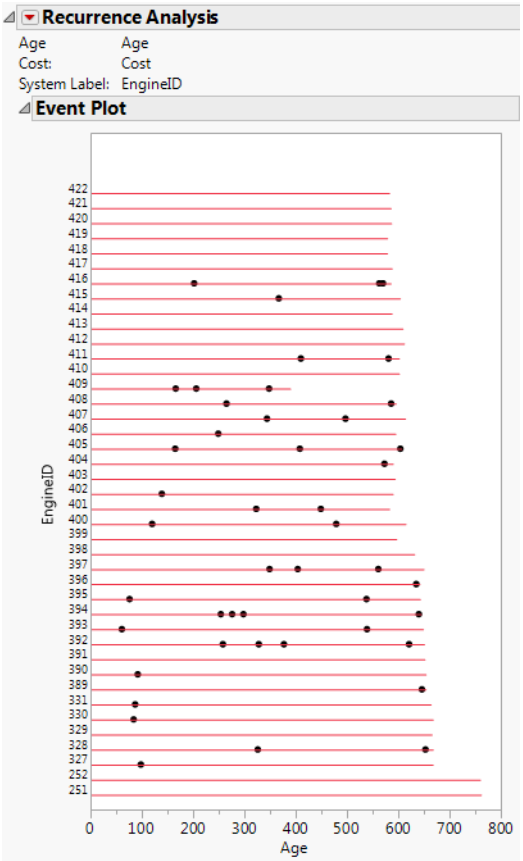
When you click **OK**, the Recurrence platform shows the reports in Figure 6.3 and Figure 6.4. The MCF plot shows the sample mean cumulative function. For each age, this is the nonparametric estimate of the mean cumulative cost or number of events per unit. This function goes up as the units get older and total costs grow. The plot in Figure 6.3 shows that about 580 days is the age that averages one repair event.

Figure 6.3 MCF Plot and Partial Table for Recurrence Analysis



The event plot in Figure 6.4 shows a time line for each unit. There are markers at each time of repair, and each line extends to that unit's last observed age. For example, unit 409 was last observed at 389 days and had three valve replacements.

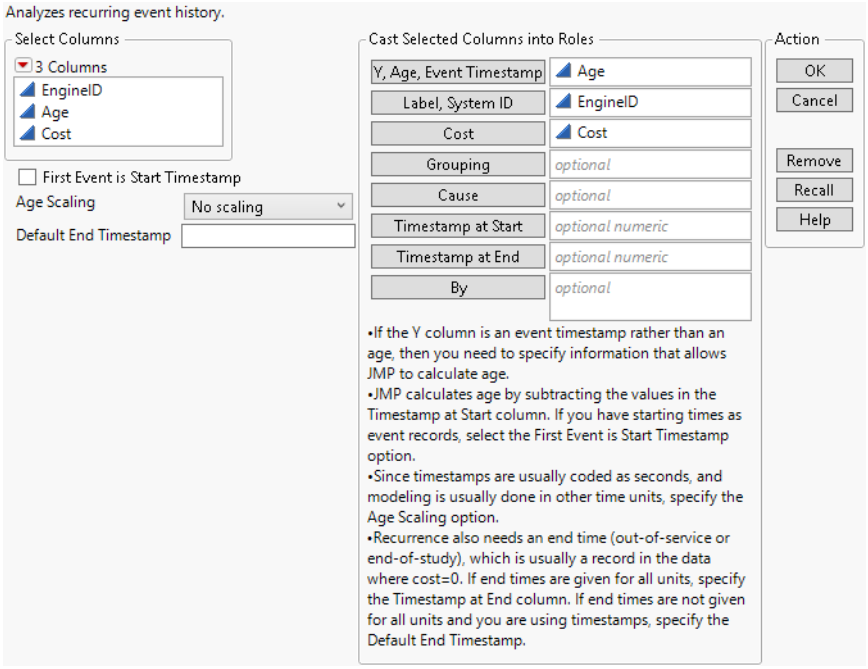
Figure 6.4 Event Plot for Valve Seat Replacements



Launch the Recurrence Analysis Platform

Launch the Recurrence Analysis platform by selecting **Analyze > Reliability and Survival > Recurrence Analysis**.

Figure 6.5 Recurrence Analysis Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Y, Age, Event Timestamp Specifies either the unit's age at the time of an event or the timestamp of the event. If the Y column is an event timestamp, then you must specify the start and the end timestamp so that JMP can calculate age.

Label, System ID Identifies the unit for each event and censoring age.

Cost Identifies a column that must contain one of the following values:

- A 1, indicating that an event has occurred (a unit failed or was repaired, replaced, or adjusted). When indicators (1s) are specified, the MCF is the mean cumulative count of events per unit as a function of age.
- A cost for the event (the cost of the repair, replacement, or adjustment). When costs are specified, the MCF is a mean cumulative cost per unit as a function of age and the markers in the Event Plot are sized by the cost values.

– A zero, indicating that the unit went out-of-service, or is no longer being studied. All units (each System ID) must have one row with a zero for this column, with the **Y, Age, Event Timestamp** column containing the final observed age. If each unit does not have exactly one last observed age in the table (where the Cost column cell is zero), then an error message appears.

Note: Cost indicators for **Recurrence Analysis** are the reverse of censor indicators seen in **Life Distribution** or **Survival Analysis**. For the cost variable, the value of 1 indicates an event, such as repair; the value of 0 indicates that the unit is no longer in service. For the censor variable, the value of 1 indicates censored values, and the value of 0 indicates the event or failure of the unit (non-censored value).

Grouping Produces separate MCF estimates for the different groups that are identified by this column.

Cause Specifies multiple failure modes.

Timestamp at Start Specifies the column with the origin timestamp. If you have starting times as event records, select the **First Event is Start Timestamp** option instead. JMP calculates age by subtracting the values in this column.

Timestamp at End Specifies the column with the end-of-service timestamp. If end times are given for all units, specify that column here. If end times are not given for all units, specify the **Default End Timestamp** option instead. But if you have a record in which Cost is equal to zero, JMP uses that record as the end timestamp and you do not need to specify this role.

Age Scaling Specifies the time units for modeling. For example, if your timestamps are coded in seconds, you can change them to hours.

Recurrence Analysis Platform Options

The Recurrence Analysis red triangle menu contains the following options:

MCF Plot Shows or hides the mean cumulative function (MCF) plot.

MCF Confid Limits Shows or hides lines corresponding to the approximate 95% confidence limits of the mean cumulative function (MCF).

Event Plot Shows or hides the Event Plot. If a Cost column is specified in the launch window, the markers in the Event Plot are sized by the values in the Cost column.

Calendar Event Plot (Available only when the events are designated by a timestamp rather than an age.) Shows or hides the Calendar Event Plot, which shows events with calendar

date on the horizontal axis. This plot is next to the Event Plot and the units in each plot are aligned vertically.

Plot Interarrival by Age Shows or hides the Interarrival by Age plot, which plots the time between successive events on the vertical axis and the event times on the horizontal axis. You can use this plot to determine whether there are changes in the time between events in your data. For a recurrence analysis, the interarrival times should be independent and identically distributed over time. For more information about interarrival plots, see Tobias and Trindade (2012, p. 420).

Plot MCF Differences (Available only when you specify a grouping variable.) Shows or hides a plot of the difference of MCFs, including a 95% confidence interval for that difference. The MCFs are significantly different where the confidence interval lines do not cross the zero line. This option is available only when you specify a grouping variable.

MCF Plot Each Group (Available only when you specify a grouping variable.) Shows or hides a report that contains a mean cumulative function (MCF) plot for each level of the grouping variable.

This option can be used to get an MCF Plot for each unit if the **Label, System ID** variable is also specified as the **Grouping** variable.

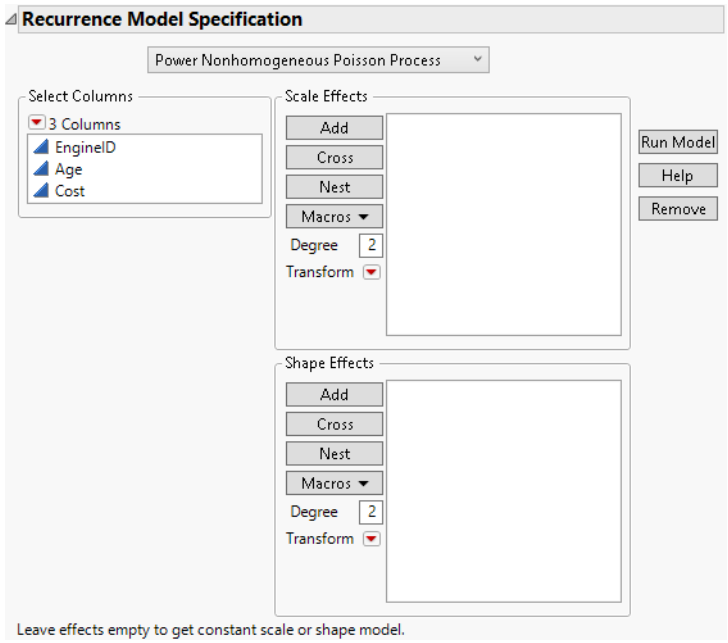
Fit Model Enables you to fit models for the Recurrence Intensity and Cumulative functions. See "Fit Model".

Fit Model

The Fit Model option is used to fit models for the Recurrence Intensity and Cumulative functions. There are four models available for describing the intensity and cumulative functions. You can fit the models with constant parameters, or with parameters that are functions of effects.

Select Fit Model from the Recurrence Analysis red triangle menu to produce the Recurrence Model Specification window shown in Figure 6.6.

Figure 6.6 Recurrence Model Specification



You can select one of four models, with the following Intensity and Cumulative functions:

Power Nonhomogeneous Poisson Process

$$I(t) = \left(\frac{\beta}{\theta}\right)\left(\frac{t}{\theta}\right)^{\beta - 1}$$

$$C(t) = \left(\frac{t}{\theta}\right)^{\beta}$$

Proportional Intensity Poisson Process

$$I(t) = \delta t^{\delta - 1} e^{\gamma}$$

$$C(t) = t^{\delta} e^{\gamma}$$

Loglinear Nonhomogeneous Poisson Process

$$I(t) = e^{\gamma + \delta t}$$

$$C(t) = \frac{I(t) - I(0)}{\delta} = \frac{e^{\gamma + \delta t} - e^{\gamma}}{\delta}$$

Homogeneous Poisson Process

$$I(t) = e^{\gamma}$$

$$C(t) = te^{\gamma}$$

where t is the age of the product.

Table 6.1 defines each model parameter as a scale parameter or a shape parameter.

Table 6.1 Scale and Shape Parameters

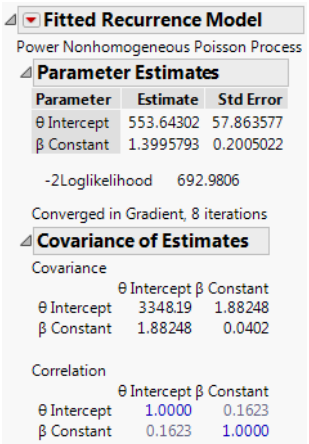
Model	Scale Parameter	Shape Parameter
Power NHPP	θ	β
Proportional Intensity PP	γ	δ
Loglinear NHPP	γ	δ
Homogeneous PP	γ	none

Note the following:

- For the Recurrence Model Specification window (Figure 6.6), if you include Scale Effects or Shape Effects, the scale and shape parameters in Table 6.1 are modeled as functions of the effects. To fit the models with constant scale and shape parameters, do not include any Scale Effects or Shape Effects.
- The Homogeneous Poisson Process is a special case compared to the other models. The Power NHPP and the Proportional Intensity Poisson Process are equivalent for one-term models, but the Proportional Intensity model seems to fit more reliably for complex models.

Click **Run Model** to fit the model and see the model report.

Figure 6.7 Model Report



The Fitted Recurrence Model red triangle menu contains the following options:

Profiler Shows or hides the Profiler showing the Intensity and Cumulative functions.

Effect Marginals Evaluates the parameter functions for each level of the categorical effect, holding other effects at neutral values. This helps you see how different the parameter functions are between groups. This is available only when you specify categorical effects.

Test Homogeneity Tests if the process is homogeneous. This option is not available for the Homogeneous Poisson Process model.

Effect Likelihood Ratio Test Produces a test for each effect in the model. This option is available only if there are effects in the model.

Specific Intensity and Cumulative Computes the intensity and cumulative values associated with particular time and effect values. The confidence intervals are profile likelihood intervals.

Specific Time for Cumulative Computes the time associated with a particular number of recurrences and effect values.

Save Intensity Formula Saves the Intensity formula to the data table.

Save Cumulative Formula Saves the Cumulative formula to the data table.

Publish Intensity Formula Creates the Intensity formula and saves it as a formula column script in the Formula Depot platform. If a Formula Depot report is not open, this option creates a Formula Depot report. See *Predictive and Specialized Modeling*.

Publish Cumulative Formula Creates the Cumulative formula and saves it as a formula column script in the Formula Depot platform. If a Formula Depot report is not open, this option creates a Formula Depot report. See *Predictive and Specialized Modeling*.

Simulate from Model Enables you to simulate new data from the estimated recurrence model. This option creates a new data table of simulated observations that are based on options that are specified in the Simulate from Model window. See "Simulate from Model".

Remove Fit Removes the model report.

Simulate from Model

When you select the Simulate from Model option from the Fitted Recurrence Model red triangle menu, the Simulate from Model window appears. This window contains the following specifications for the simulation:

Maximum Number of Events Specifies the maximum number of events to be simulated in each level of the System ID.

Maximum Age Specifies the maximum time for a simulated event.

Number of Units Specifies the number of units in each level of the System ID.

> **Note:** If the model contains terms other than the Intercept and Constant terms, the simulated data table contains this number of units for all level combinations of the regression terms. If the regression term is continuous, it is divided into 5 levels.

After you click OK, a new data table that contains the results appears. This table contains a script that enables you to analyze the simulated observations in the Recurrence platform.

Additional Examples of the Recurrence Analysis Platform

- "Bladder Cancer Recurrences Example"
- "Diesel Ship Engines Example"

Bladder Cancer Recurrences Example

The sample data file Bladder Cancer.jmp contains data on cancerous bladder tumor recurrences from the Veteran's Administration Co-operative Urological Research Group. See Andrews and Herzberg (1985, table 45). All patients presented with superficial bladder tumors, which were removed upon entering the trial. Each patient was then assigned to one of three treatment groups: placebo pills, pyridoxine (vitamin B6) pills, or periodic chemotherapy with Thiotepa. The following analysis of tumor recurrence explores the progression of the disease, and whether there is a difference among the three treatments.

Launch the platform with the options shown in Figure 6.8.

Figure 6.8 Bladder Cancer Launch Window

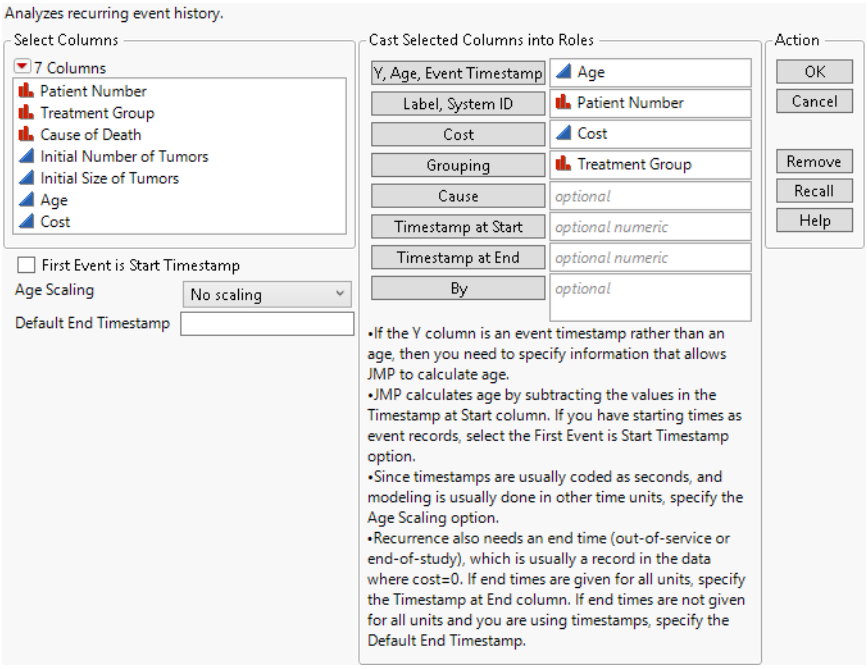
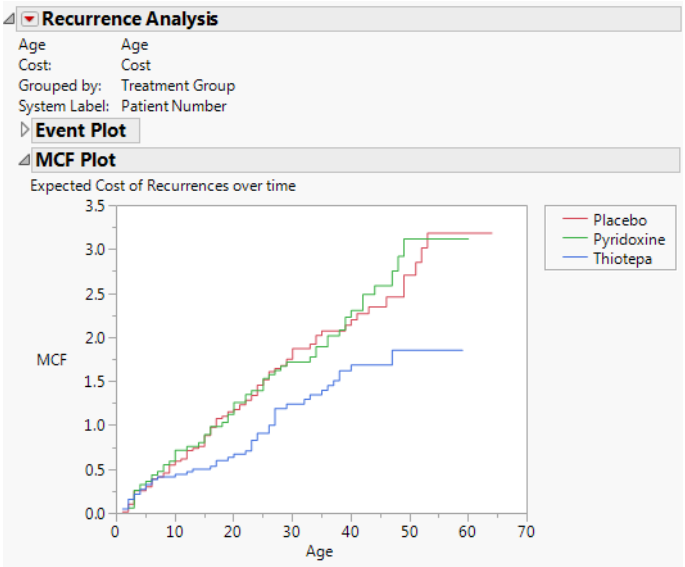


Figure 6.9 shows the MCF plots for the three treatments.

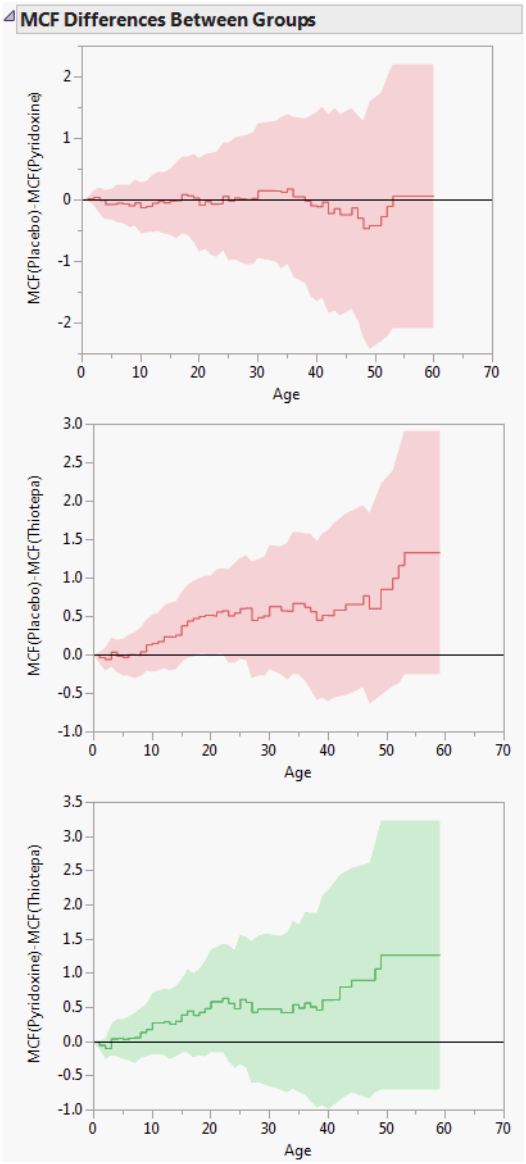
Figure 6.9 Bladder Cancer MCF Plot



Note that all three of the MCF curves are essentially straight lines. The slopes (rates of recurrence) are therefore constant over time, implying that patients do not seem to get better or worse as the disease progresses.

To examine if there are differences among the treatments, click the Recurrence Analysis red triangle and select **Plot MCF Differences**.

Figure 6.10 MCF Differences



To determine whether there is a statistically significant difference between treatments, examine the confidence limits on the differences plot. If the limits do not include zero, the treatments are convincingly different. The graphs in Figure 6.10 show there is no significant difference among the treatments.

Diesel Ship Engines Example

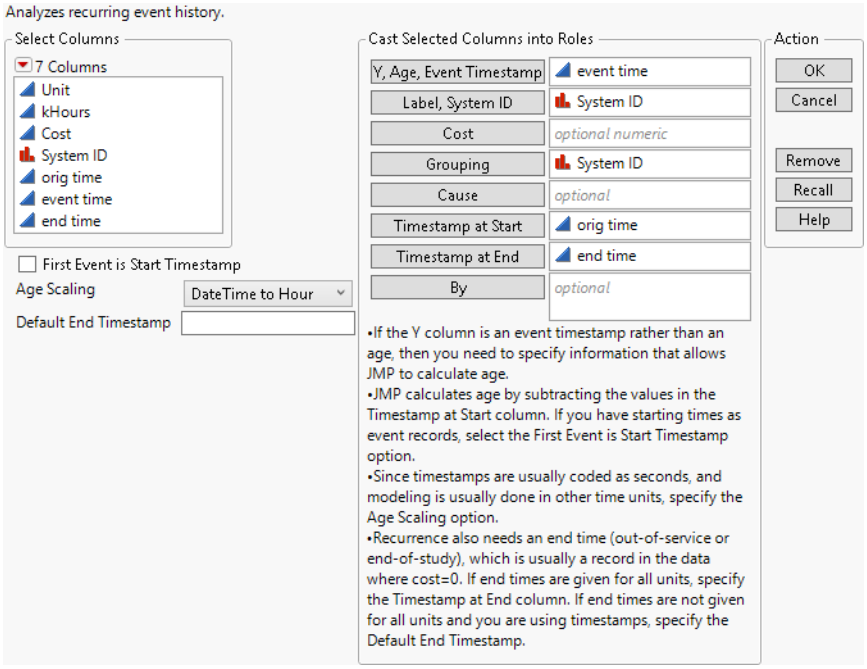
The sample data table Diesel Ship Engines.jmp contains data on engine repair times for two ships (Grampus4 and Halfbeak4) that have been in service for an extended period of time. See Meeker and Escobar (1998). You want to examine the progression of repairs and gain a sense of how often repairs might need to be done in the future. These observations can help you decide when an engine should be taken out of service.

1. Select **Help > Sample Data Library** and open Reliability/Diesel Ship Engines.jmp.
2. Ensure that rows 57 and 129 are set as Excluded.

> **Note:** If they are not set to Excluded, select rows 57 and 129 and select **Rows > Exclude/Unexclude**.

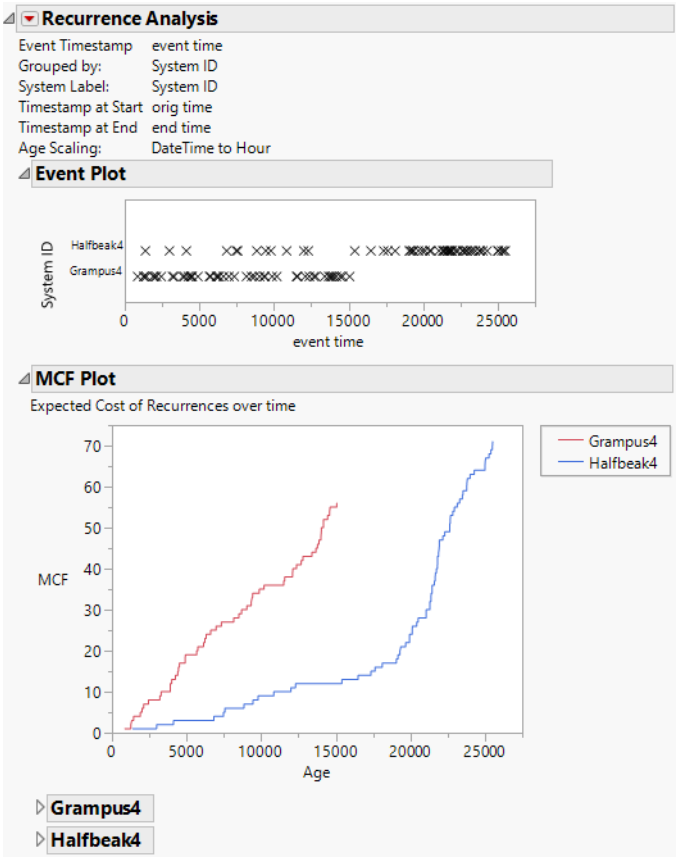
3. Select **Analyze > Reliability and Survival > Recurrence Analysis**.
4. Complete the launch window as shown in Figure 6.11.

Figure 6.11 Diesel Ship Engines Launch Window



5. Click **OK**.

Figure 6.12 Diesel Ship Engines Report

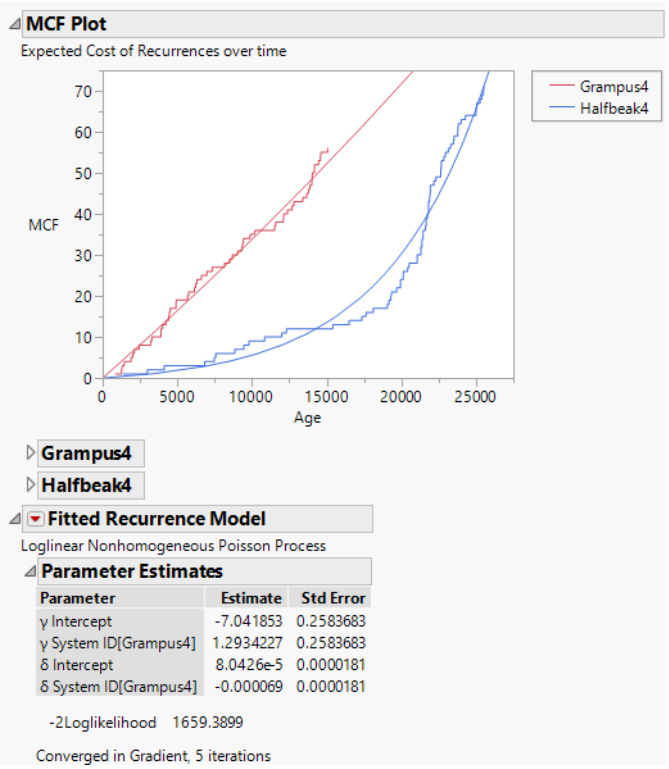


Looking at the Event Plot, you can see that repairs for the Grampus4 engine have been relatively consistent. Repairs for the Halfbeak4 engine have been more sporadic, and there appears to be an abrupt increase in repairs somewhere around the 19,000 hour mark. This increase is even more obvious in the MCF Plot.

Continue your analysis by fitting a parametric model to help predict future performance.

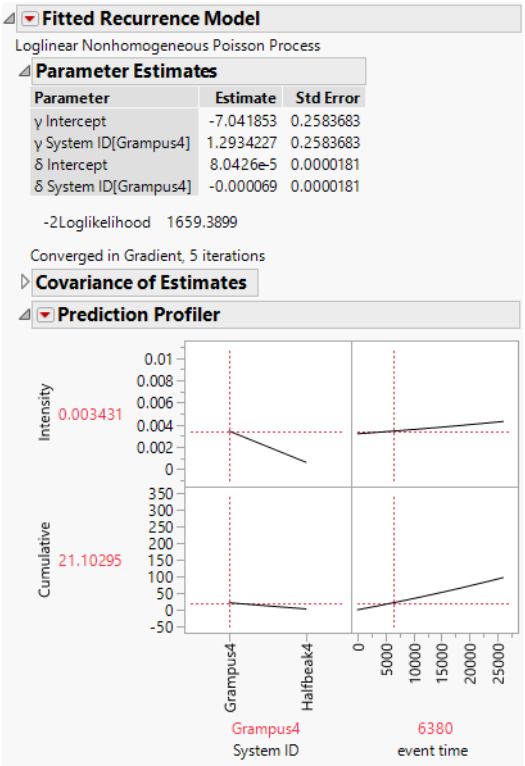
6. Click the Recurrence Analysis red triangle and select **Fit Model**.
7. In the Recurrence Model Specification, select the **Loglinear Nonhomogeneous Poisson Process**.
8. Add the System ID column as both a Scale Effect and a Shape Effect.
9. Click **Run Model**.

Figure 6.13 Diesel Ship Engines Fitted Model



10. Click the Fitted Recurrence Model red triangle and select **Profiler**.

Figure 6.14 Diesel Ship Profiler



Compare the number of future repairs for the Grampus4 engine to the Halfbeak4 engine. Change the event time value to see the effect on the cumulative number of future repairs.

- To see how many repairs will be needed after 30,000 hours of service, type 30,000 for the event time. The Grampus4 engine will require about 114 repairs. To see the values for Halfbeak4, click and drag the dotted line from Grampus4 to Halfbeak4. The Halfbeak4 engine will require about 140 repairs.
- To see how many repairs will be needed after 80,000 hours of service, type 80,000 for the event time. The Halfbeak4 engine will require about 248,169 repairs. Click and drag the dotted line from Halfbeak4 to Grampus4. The Grampus4 engine will require about 418 repairs.

You can conclude that in the future, the Halfbeak4 engine will require many more repairs than the Grampus4 engine.

Chapter 7

Degradation

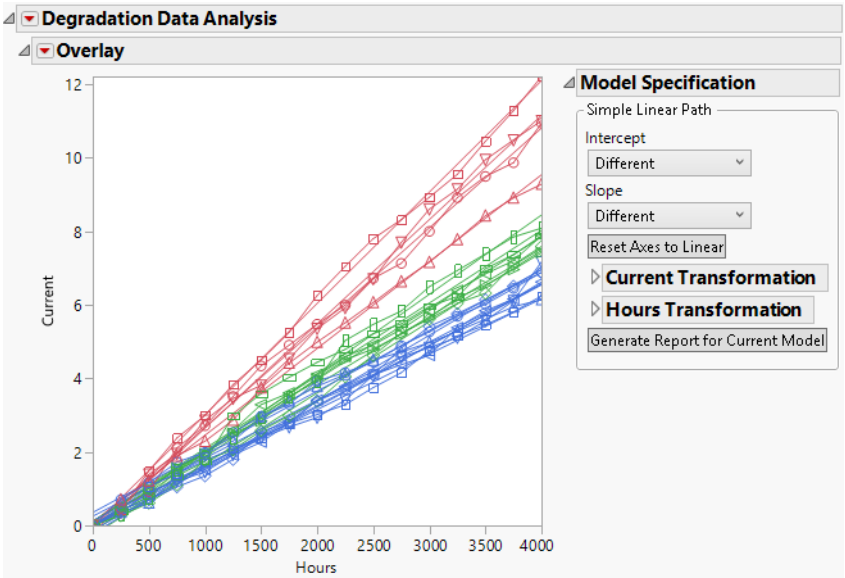
Model Product Deterioration over Time

Using the Degradation platform, you can analyze degradation data to predict pseudo failure times. These pseudo failure times can then be analyzed by other reliability platforms to estimate failure distributions.

Both linear and nonlinear degradation paths can be modeled. You can specify an accelerating factor to analyze accelerated degradation data.

You can also perform stability analysis, which is useful when setting pharmaceutical product expiration dates.

Figure 7.1 Degradation Analysis Example



Contents

Overview of the Degradation Platform 177
Example of the Degradation Platform 177
Launch the Degradation Platform 179
The Degradation Reports 181
Model Specification 184
Simple Linear Path 184
Nonlinear Path 186
Inverse Prediction 194
Prediction Graph 196
Degradation Platform Options 197
Model Reports 199
Model Lists 199
Reports 200
Destructive Degradation 200
Stability Analysis 205

Overview of the Degradation Platform

In reliability analyses, the primary objective is to model the failure times of the product under study. In many situations, these failures occur because the product degrades (weakens) over time. But, sometimes failures do not occur. In these situations, modeling the product degradation over time is helpful in making predictions about failure times. When an accelerating factor is included in the data, you can fit an accelerated degradation model.

The Degradation platform can model data that follows linear or nonlinear degradation paths. If a path is nonlinear, transformations are available to linearize the path. If linearization is not possible, you can specify a nonlinear model.

You can also use the Degradation platform to perform stability analysis. Three types of linear models are fit, and an expiration date is estimated. Stability analysis is used in setting pharmaceutical product expiration dates.

Example of the Degradation Platform

This example uses the GaAs Laser.jmp data table from Meeker and Escobar (1998), which contains measurements of the percent increase in operating current taken on several gallium arsenide lasers. When the percent increase reaches 10%, the laser is considered to have failed.

1. Select **Help > Sample Data Library** and open Reliability/GaAs Laser.jmp.
2. Select **Analyze > Reliability and Survival > Degradation**.
3. Select Current and click **Y, Response**.
4. Select Hours and click **Time**.
5. Select Unit and click **Label, System ID**.
6. Type 10 in the text box for **Upper Spec Limit**.
7. Click **OK**.

Figure 7.2 Initial Degradation Report

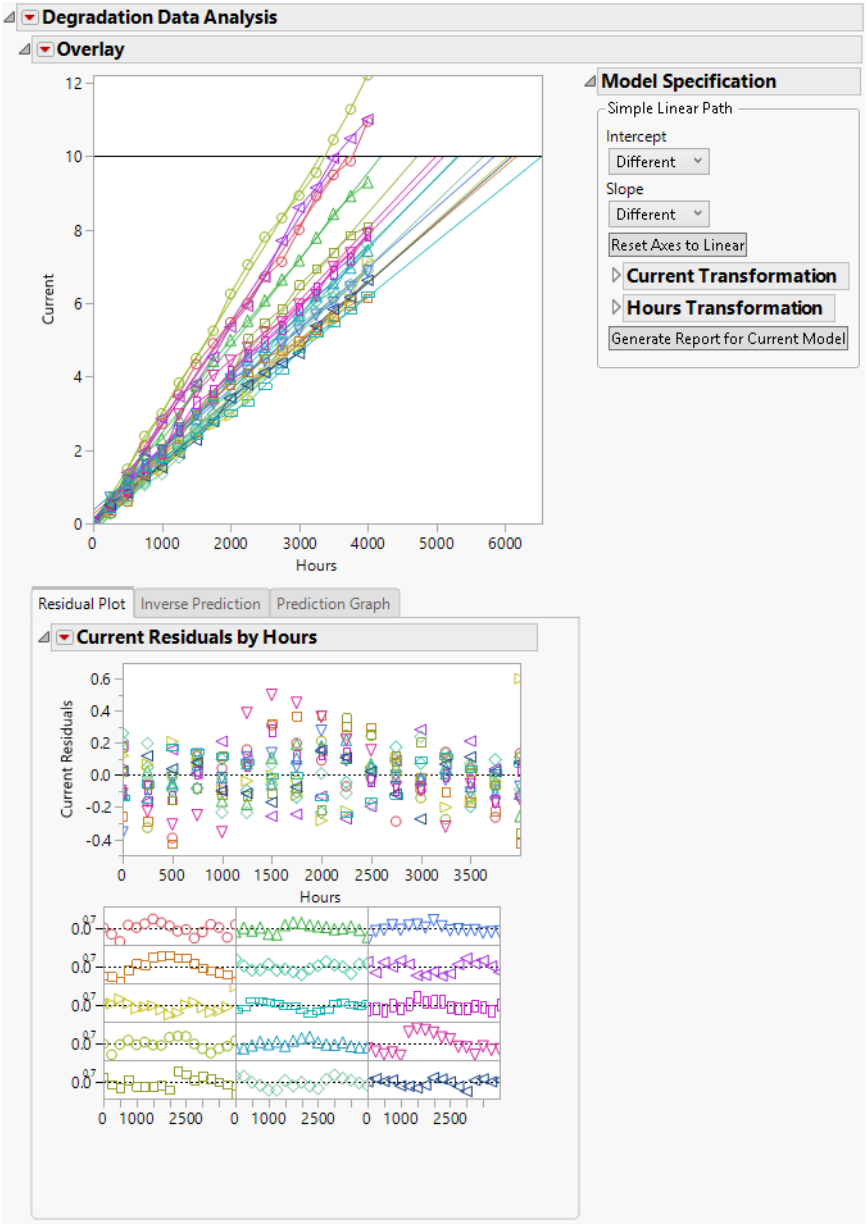


Figure 7.2 shows the initial Degradation report. The Overlay plot shows the measurements of Current versus Time for each unit in the data. The horizontal line at Current = 10 corresponds to the upper specification limit at 10%. Units with values above this limit are considered to have failed. Three of the fifteen units have reached that point by the end of the study period. The Inverse Prediction outline shows the predicted Hours value for which each unit fails, based on your specified model.

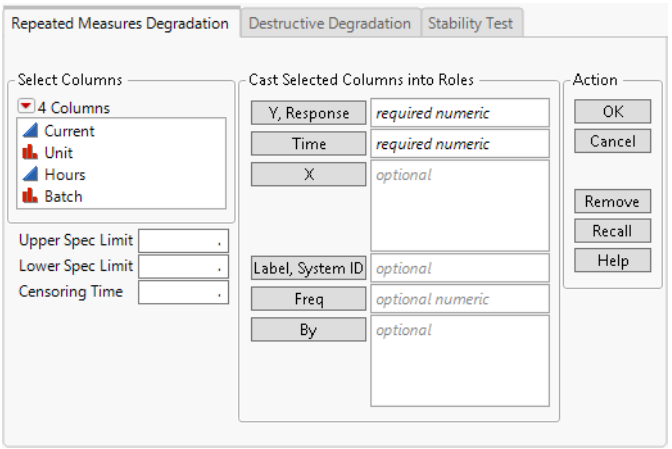
The default model fits a separate slope and intercept for each unit, using linear transformations. You can fit other models using the Model Specification outline.

The Residual Plot tab in Figure 7.2 shows residuals based on your specified model. The top plot shows residuals for all units plotted against Hours and overlaid on one plot. The bottom plot shows individual plots of the residuals for each unit in a rectangular array.

Launch the Degradation Platform

Launch the Degradation platform by selecting **Analyze > Reliability and Survival > Degradation**. Figure 7.3 shows the Degradation launch window using the GaAs Laser.jmp data table (located in the Reliability folder). For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Figure 7.3 The Degradation Launch Window



Analysis Types

The launch window is split into three tabs, representing three different types of analyses:

Repeated Measures Degradation Performs linear or nonlinear degradation analysis. This option allows only one Y, Response variable. It does not allow censoring.

Destructive Degradation Choose this type of analysis if units are destroyed during the measurement process. This option allows censoring. See "Destructive Degradation".

Note: The Destructive Degradation platform provides a flexible collection of predefined models for destructive testing. See "Destructive Degradation".

Stability Test Performs a stability analysis for setting pharmaceutical product expiration dates. This option allows only one Y, Response variable. See "Stability Analysis".

Launch Window Options

The launch window contains the following options:

Y, Response Assign the column with degradation measurements.

Time Assign the column containing the time values.

X (Available only in the Repeated Measures Degradation and Destructive Degradation tabs.) Assign an explanatory variable. Use this role to specify the accelerating factor in an accelerated degradation model.

Label, System ID (Available only in the Repeated Measures Degradation and Stability Test tabs.) Assign the column that designates the unit IDs.

Freq Assign a column giving a frequency for each row.

Censor (Available only in the Destructive Degradation tab.) Assign a column that designates if a unit is censored.

By Assign a variable to produce an analysis for each level of the variable.

Censor Code (Available only in the Destructive Degradation tab.) Specify the value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

Upper Spec Limit Assign an upper specification limit. (Optional except for Stability Test tab.)

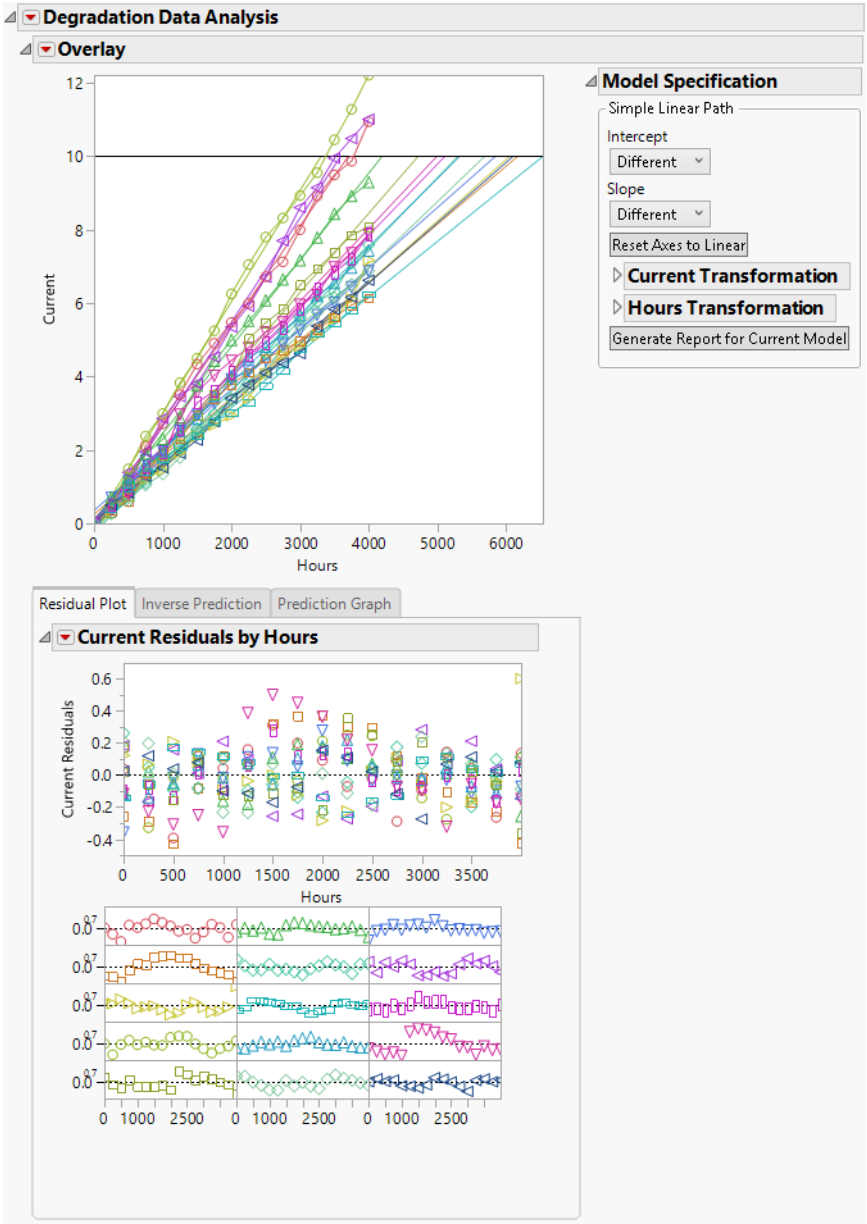
Lower Spec Limit Assign a lower specification limit. (Optional except for Stability Test tab.)

Censoring Time (Available only in the Repeated Measures Degradation and Destructive Degradation tabs.) Assign a Time value that represents censoring of pseudo failures when you use Inverse Prediction. See "Inverse Prediction".

The Degradation Reports

The Repeated Measures Degradation and Destructive Degradation methods show a report that fits a default model. As shown in the Model Specification outline in Figure 7.4, this model fits each unit with its own intercept and slope, using a linear transformation of the response and time columns. Separate intercepts and slopes are fit for each value of the Label, System ID variable, or, if only an X variable is specified, separate intercepts and slopes are fit for each level of the X variable. The Stability Test method fits three models.

Figure 7.4 Initial Repeated Measures Degradation Report with Transformation Outlines Open



To reproduce this example, see "Example of the Degradation Platform".

The reports for Repeated Measures Degradation, Destructive Degradation, and Stability Test include the following:

Overlay

An Overlay plot of the Y, Response variable versus the Time variable. In this example, the plot is of Current versus Hours. The Save Estimates option in the Overlay plot red triangle menu creates a new data table containing the estimated slopes and intercepts for all units.

Model Specification

Specify your model and generate a report for that model. See "Model Specification". (Available only for the Repeated Measures Degradation and Destructive Degradation methods.)

Stability Tests Outline

Compare models and estimate expiration dates. See "Stability Analysis". (Available only for the Stability Test method.)

Reports

Shows analysis results for three different models and the best model. See "Stability Analysis". (Available for the Stability Test method by default. Also available for the Repeated Measures Degradation and Destructive Degradation methods after you click Generate Report for Current Model.)

Tabbed Reports

Residual Plot Shows a single residual plot with all the units overlaid and separate residual plots for each unit arranged in a rectangular grid. The red triangle menu contains the following options:

Save Residuals Saves the residuals of the current model to a new data table.

Jittering Adds random noise to the points in the time direction. This is useful for visualizing the data if there are a lot of points clustered together.

Separate Groups Adds space between the groups to visually separate the groups. This option appears only when an X variable is specified on the platform launch window.

Jittering Scale Changes the magnitude of the jittering and group separation. This option appears only if Jittering is selected.

Inverse Prediction Enables you to predict the time at which the Y variable reaches a specified value. See "Inverse Prediction".

Prediction Graph Enables you to predict the Y variable for a specified Time value. See "Prediction Graph".

Model Specification

You can use the Model Specification outline to specify the model that you want to fit to the degradation data. There are two types of Model Specifications:

Simple Linear Path Used to model linear degradation paths, or nonlinear paths that can be transformed to linear. See "Simple Linear Path".

Nonlinear Path Used to model nonlinear degradation paths, especially those that cannot be transformed to linear. See "Nonlinear Path".

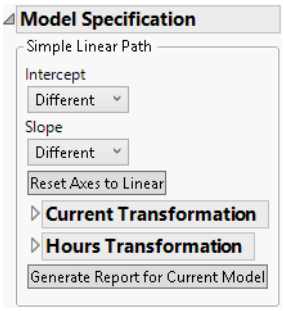
To change between the two specifications, use the Degradation Path Style submenu from the Degradation Data Analysis red triangle menu.

Simple Linear Path

To model linear degradation paths, select **Degradation Path Style > Simple Linear Path** from the Degradation Data Analysis red triangle menu.

Use the Simple Linear Path Model specification to specify the form of the linear model that you want to fit to the degradation path. You can model linear paths, or nonlinear paths that can be transformed to linear.

Figure 7.5 Simple Linear Path Model Specification



The Simple Linear Path model specification contains the following options:

Intercept Specifies the form of the intercept in the model:

Different Fits a different intercept for each level of the ID variable.

Common in Group Fits the same intercept for each level of the ID variable in the same level of the X variable, and fits different intercepts between levels.

Common Fits the same intercept for all levels of the ID variable.

Zero Restricts the intercept to be zero for all levels of the ID variable.

Slope Specifies the form of the slope in the model:

Different Fits a different slope for each level of the ID variable.

Common in Group Fits the same slope for each level of the ID variable in the same level of the X variable, and fits different slopes between levels.

Common Fits the same slope for all levels of the ID variable.

Reset Axes to Linear Resets the Overlay plot axes to their initial settings.

<Y, Response> Transformation If a transformation on the Y variable can linearize the degradation path, select the transformation (Linear, ln(x), exp(x), x^2, sqrt(x) or Custom) here. For more information about the Custom option, see "Custom Transformations".

<Time> Transformation If a transformation for the Time variable can linearize the degradation path, select the transformation (Linear, ln(x), x^2, sqrt(x) or Custom) here. For more information about the Custom option, see "Custom Transformations".

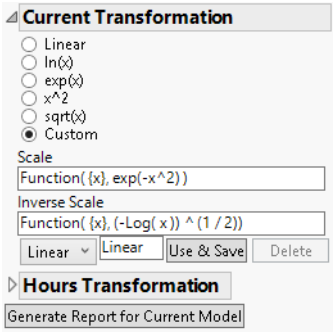
Generate Report for Current Model Creates a report for the current model settings. This includes a Model Summary report, and Estimates report giving the parameter estimates. See "Model Reports".

Custom Transformations

If you need to perform a transformation that is not given, use the Custom option. For example, to transform the response variable using $\exp(-x^2)$, enter the transformation as shown in the Scale box in Figure 7.6. Also, enter the inverse transformation in the Inverse Scale box.

Note: JMP automatically attempts to solve for the inverse transformation. If it can solve for the inverse, it automatically enters it in the Inverse Scale box. If it cannot solve for the inverse, you must enter it manually.

Figure 7.6 Custom Transformation Options



Name the transformation using the text box. When finished, click the **Use & Save** button to apply the transformation. Select a transformation from the menu if you have created multiple custom transformations. Click the **Delete** button to delete a custom transformation.

Nonlinear Path

To model nonlinear degradation paths, select **Degradation Path Style > Nonlinear Path** from the Degradation Data Analysis red triangle menu. This is useful if a degradation path cannot be linearized using transformations, or if you have a custom nonlinear model that you want to fit to the data.

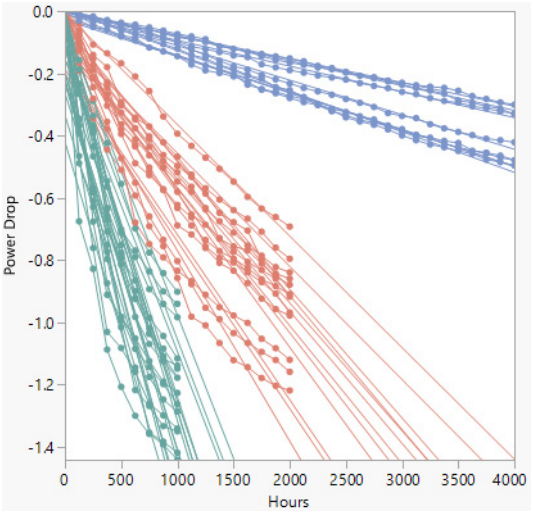
To facilitate explaining the Nonlinear Path Model Specification, open the Device B.jmp data table. The data consists of power decrease measurements taken on 34 units, across four levels of temperature. Follow these steps:

1. Select **Help > Sample Data Library** and open Reliability/Device B.jmp.
2. Select **Analyze > Reliability and Survival > Degradation**.
3. Select Power Drop and click **Y, Response**.
4. Select Hours and click **Time**.
5. Select Degrees C and click **X**.

   The temperature setting is the accelerating factor in the experiment.
6. Select Device and click **Label, System ID**.
7. Click **OK**.

The initial overlay plot of the data appears.

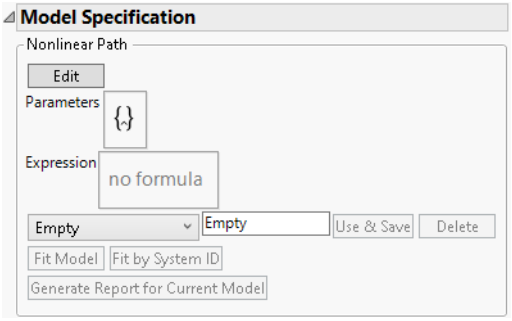
Figure 7.7 Device B Overlay Plot



The degradation paths appear linear for the first several hundred hours, but then start to curve. To fit a nonlinear model, select **Degradation Path Style > Nonlinear Path** from the Degradation Data Analysis red triangle menu to show the Nonlinear Path Model Specification outline (Figure 7.8).

Note: To view the Edit button displayed in Figure 7.8, you must select the interactive formula editor preference (**File > Preferences > Platforms > Degradation > Use Interactive Formula Editor**) before launching the Degradation platform.

Figure 7.8 Initial Nonlinear Model Specification Outline



The first step to create a model is to select one of the options on the menu initially labeled **Empty**:

- For more information about Reaction Rate models, see "Reaction Rate Models".
- For more information about Constant Rate models, see "Constant Rate Models".

- For more information about using a Prediction Column, see "Prediction Columns".

Reaction Rate Models

The Reaction Rate options are applicable when the degradation occurs from a single chemical reaction, and the reaction rate is a function of temperature only. Select **Reaction Rate** or **Reaction Rate Type 1** from the menu shown in Figure 7.8. Although similar to the Reaction Rate model, the Reaction Rate Type 1 model contains an offset term that changes the basic assumption concerning the response value's sign.

The Setup window prompts you to select the temperature scale, and the baseline temperature. The baseline temperature is used to generate initial estimates of parameter values. The baseline temperature should be representative of the temperatures used in the study.

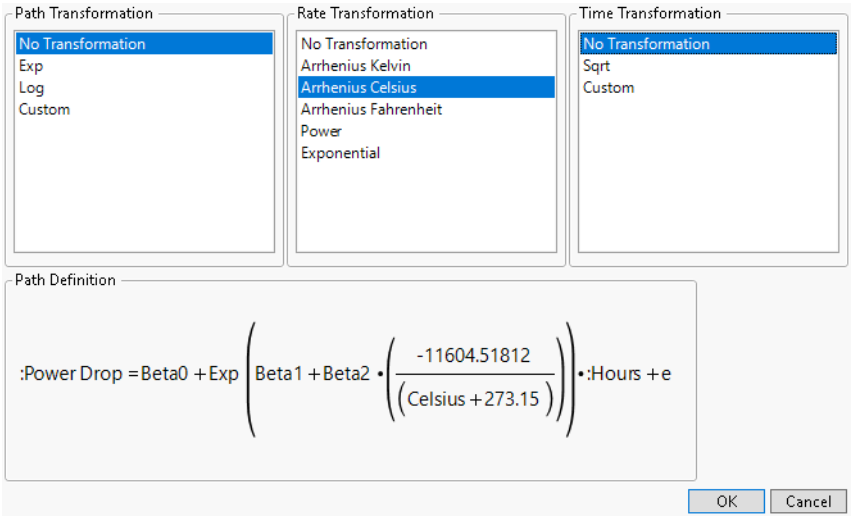
For this example, select **Reaction Rate** and then select Celsius as the Temperature Unit. Click **OK** to return to the report. For more information about all the features for Model Specification, see "Model Specification Details".

Constant Rate Models

The Constant Rate option is for modeling degradation paths that are linear with respect to time (or linear with respect to time after transforming the response or time). The reaction rate is a function of temperature only.

Select **Constant Rate** from the menu shown in Figure 7.8. The Constant Rate Model Settings window prompts you to enter transformations for the Path, Rate, and Time.

Figure 7.9 Constant Rate Transformation



Once a selection is made for each transformation, the associated formula appears in the lower left corner as shown in Figure 7.9.

After all selections are made, click **OK** to return to the report. For more information about all the features for Model Specification, see "Model Specification Details".

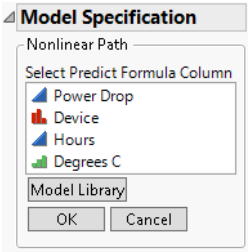
Prediction Columns

The Prediction Column option enables you to use a custom model that is stored as a formula in a data table column. The easiest approach is to create the formula column before launching the Degradation platform. You can also create the formula column from within the Degradation platform if you want to use one of the built-in models in the Nonlinear Model Library.

For more information about how to create a custom model and store it as a column formula, see "Fit a Custom Model" or *Predictive and Specialized Modeling*.

Select **Prediction Column** from the list that appears beneath the Expression area in Figure 7.8. The Model Specification outline changes to prompt you to select the column that contains the model.

Figure 7.10 Column Selection



At this point, do one of three things:

- If the model that you want to use already exists as a formula in a column of the data table, select the corresponding column here, and then click **OK**. You are returned to the Nonlinear Path Model Specification. For more information about all the features for that specification, see "Model Specification Details".
- If the model that you want to use does not already exist in the data table, you can click the **Model Library** button to use one of the built-in models. For more information about using the Model Library button, see "Model Library" or *Predictive and Specialized Modeling*. After the model is created, select **Redo > Redo Analysis** from the Degradation Data Analysis red triangle menu. Then, return to the column selection shown in Figure 7.10. Select the column that contains the model, and then click **OK**. You are returned to the Nonlinear Path Model Specification. For more information about all the features for that specification, see "Model Specification Details".

- If the model that you want to use is not in the data table, and you do not want to use one of the built-in models, then you are not ready to use this model specification. First, create the model, relaunch the Degradation platform, and then return to the column selection (Figure 7.10). Select the column containing the model, and then click **OK**. You are returned to the Nonlinear Path Model Specification. For more information about all the features for that specification, see "Model Specification Details".

Model Specification Details

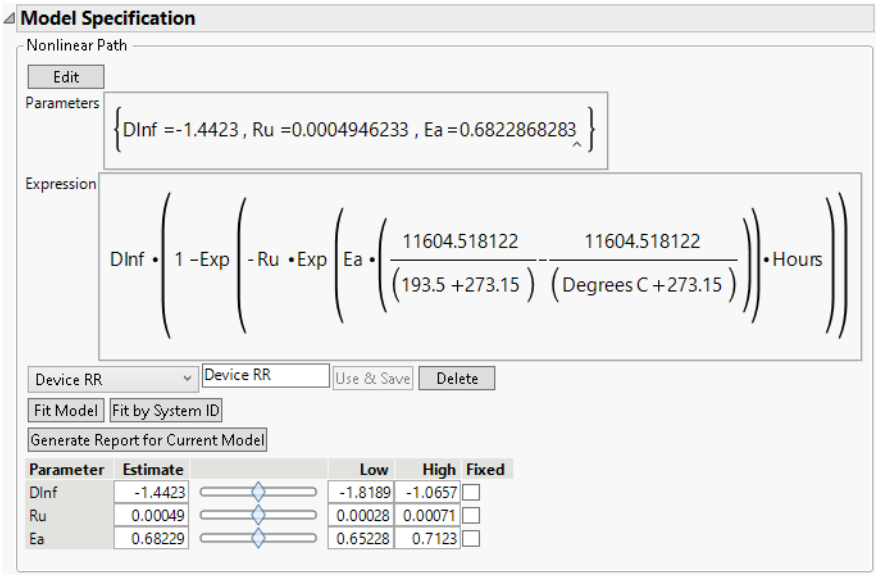
After you select one of the model types and supply the required information, you are returned to the Nonlinear Path Model Specification window.

Note: To view the Edit button displayed in Figure 7.11, you must select the interactive formula editor preference (**File > Preferences > Platforms > Degradation > Use Interactive Formula Editor**) before launching the Degradation platform.

A model is now shown in the script box that uses the `Parameter` statement. Initial values for the parameters are estimated from the data. For more information about creating models that use parameters, see "Fit a Custom Model" or *Predictive and Specialized Modeling*.

If desired, enter a name in the text box to name the model. For this example, use the name "Device RR". After that, click the **Use & Save** button to enter the model and activate the other buttons and features. Figure 7.11 shows the Model Specification window after clicking the Use & Save button.

Figure 7.11 Model Specification



- The **Fit Model** button is used to fit the model to the data.
- The **Fit by System ID** is used to fit the model to every level of **Label, System ID**.
- The **Delete** button is used to delete a model from the model menu.
- The **Generate Report for Current Model** button creates a report for the current model settings. See "Model Reports".

The initial parameter values are shown at the bottom, along with sliders for visualizing how changes in the parameters affect the model. The fitted lines are shown on the Overlay plot. Move the parameter sliders to see how changes affect the fitted lines.

Here are the parameters for the Reaction Rate model (Meeker and Escobar 1998):

- **Dinf** (D_∞) - asymptotic degradation level
- **Ru** (R_U) - reaction rate at use temperature ($temp_U$)
- **Ea** (Ea) - reaction-specific activation energy

The above parameters are calculated as follows:

$$D(t; \text{temp}) = D_\infty \text{ x } \{1\text{-exp}[-R_U \text{ x } AF(\text{temp}) \text{ x } t]\}$$

where R_U is the reaction rate at use temperature $temp_U$, R_U x AF(temp) is the reaction rate at a general temperature temp, and for temp > $temp_U$, AF(temp) > 1

and

$$AF(\text{temp}) = \frac{R(temp)}{R(temp_U)} = \exp\left[Ea\left(\frac{11604.5181215503}{(temp_U K)} - \frac{11604.5181215503}{(tempK)}\right)\right]$$

where $temp_U K$ and $tempK$ are temperatures expressed on the Kelvin scale.

To compute the optimal values for the parameters, click the **Fit Model** or **Fit by System ID** button.

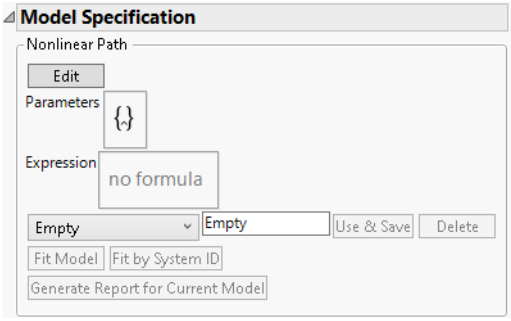
To fix a value for a parameter, check the box under **Fixed** for the parameter. When fixed, that parameter is held constant in the model fitting process.

Entering a Model with the Formula Editor

You can use the Formula Editor to enter a model. Click the **Edit** button to open the Formula Editor to enter parameters and the model. For more information about entering parameters and formulas in the Formula Editor, see *Using JMP*.

Note: To view the Edit button displayed in Figure 7.12, you must select the interactive formula editor preference (**File > Preferences > Platforms > Degradation > Use Interactive Formula Editor**) before launching the Degradation platform.

Figure 7.12 Alternate Model Specification Report



Fit a Custom Model

If you want to fit a custom model, you must first create a formula column with initial parameter estimates. This method requires a few more steps than fitting a built-in model, but it allows any nonlinear model to be fit. Also, you can provide a custom loss function, and specify several other options for the fitting process.

1. Open your data table.
2. Create a new column in the data table.
3. Open the Formula Editor for the new column.
4. Select **Parameters** from the list in the lower left corner.
5. Click **New Parameter**.
6. Enter the name of the parameter.
7. Enter the initial value of the parameter.

   Repeat steps 4 to 6 to create all the parameters in the model.
8. Build the model formula using the data table columns, parameters, and formula editor functions.
9. Click **OK**.

Parameters for Models with a Grouping Variable

In the formula editor, when you add a parameter, note the check box for **Expand Into Categories, selecting column**. This option is used to add several parameters (one for each level of a categorical variable for example) at once. When you select this option, a window appears that enables you to select a column. After selection, a new parameter appears in the Parameters list with the name D_*column*, where D is the name that you gave the parameter. When you use this parameter in the formula, a Match expression is inserted, containing a separate parameter for each level of the grouping variable.

Model Library

The Model Library can assist you in creating a formula column with parameters and initial values. Click **Model Library** under Model Specification to open the library. Select a model in the list to see its formula in the **Formula** box.

Click **Show Graph** to show a 2-D theoretical curve for one-parameter models and a 3-D surface plot for two-parameter models. No graph is available for models with more than two explanatory (*X*) variables. Use the slider bars to change the default starting values for the parameters. You can also click the values and enter new values directly.

The **Reset** button sets the starting values for the parameters back to their default values.

Click **Show Points** to overlay the actual data points to the plot. A window opens, asking you to assign columns into *X* and *Y* roles, and an optional Group role. The Group role allows for fitting the model to every level of a categorical variable. If you specify a Group role here, also specify the same column in the Label, System ID role in the platform launch window.

For most models, the starting values are constants. Showing points enables you to adjust the parameter values to see how well the model fits for different values of the parameters.

Click **Make Formula** to create a new column in the data table. This column has the formula as a function of the specified *X* variable and uses the parameter values specified in the graph window.

Note: If you click **Make Formula** before you click the **Show Graph** or **Show Points** buttons, you are asked to provide the *X* and *Y* roles, and an optional Group role. After that, you are brought back to the plot so that you have the option to adjust the starting values for the parameters. Once the starting values for the parameters are satisfactory, click **Make Formula** again to create the new column.

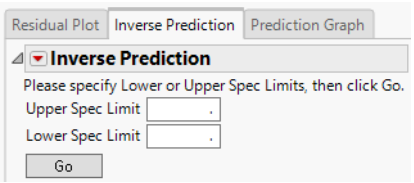
Once the formula is created in the data table, click the Degradation Data Analysis red triangle and select **Redo > Redo Analysis**. Then, return to the column selection shown in Figure 7.10. Select the column that contains the model, and then click **OK**. You are returned to the Nonlinear Path Model Specification. For more information about all the features for that specification, see "Model Specification Details".

Note: You can customize the models included in the Nonlinear Model Library by modifying the built-in script named NonlinLib.jsl. This script is located in the Resources/Builtins folder in the folder that contains JMP (Windows) or in the Application Package (macOS).

Inverse Prediction

Use the Inverse Prediction tab to predict the time at which the Y variable reaches a specified value. These times are sometime called pseudo failure times.

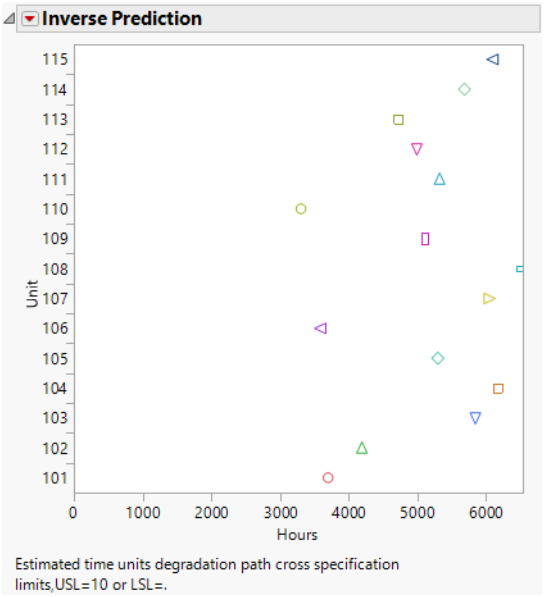
Figure 7.13 Inverse Prediction Tab



Enter either the Lower or Upper Spec Limit. Generally, if your Y variable decreases over time, then enter a Lower Spec Limit. If the Y variable increases over time, then enter an Upper Spec Limit.

For the GaAs Laser example, enter 10 for the Upper Spec Limit and click **Go**. A plot is produced showing the estimated times until the units reach a 10% increase in operating current.

Figure 7.14 Inverse Prediction Plot



The Inverse Prediction red triangle menu contains the following options:

Save Crossing Time Saves the pseudo failure times to a new data table. The table contains a Life Distribution or Fit Life by X script that can be used to fit a distribution to the pseudo failure times. When one of the Inverse Prediction Interval options is enabled, the table also includes the intervals.

Set Upper Spec Limit Sets the upper specification limit.

Set Lower Spec Limit Sets the lower specification limit.

Set Censoring Time Sets the censoring time. The plot updates to show the Censoring Time as a dotted vertical line. If Inverse Prediction Interval > No Interval is selected, observations that exceed the Censoring Time are displayed on horizontal lines starting at the Censoring Time. If Confidence Interval or Prediction Interval is selected, horizontal lines extend indefinitely to the right of observations whose upper limits exceed the Censoring Time. The Censoring Time is reflected in data tables constructed using Save Crossing Time and Generate Pseudo Failure Data.

Use Interpolation through Data Specifies the use of linear interpolation between points (instead of the fitted model) to predict when a unit crosses the specification limit. The behavior depends on whether a unit has observations that exceed the specification limit.

- If a unit has observations exceeding the specification limit, the inverse prediction is the linear interpolation between the observations that surround the specification limit.
- If a unit does not have observations exceeding the specification limit, the inverse prediction is censored and has a value equal to the maximum observed time for that unit.

Inverse Prediction Interval Specifies that confidence or prediction intervals for the pseudo failure times are shown on the Inverse Prediction plot. When intervals are enabled, the intervals are also included in the data table that is created when using the Save Crossing Time option.

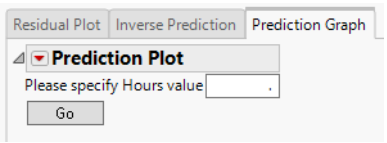
Inverse Prediction Alpha Specifies the alpha level used for the intervals.

Inverse Prediction Side Specifies one or two sided intervals.

Prediction Graph

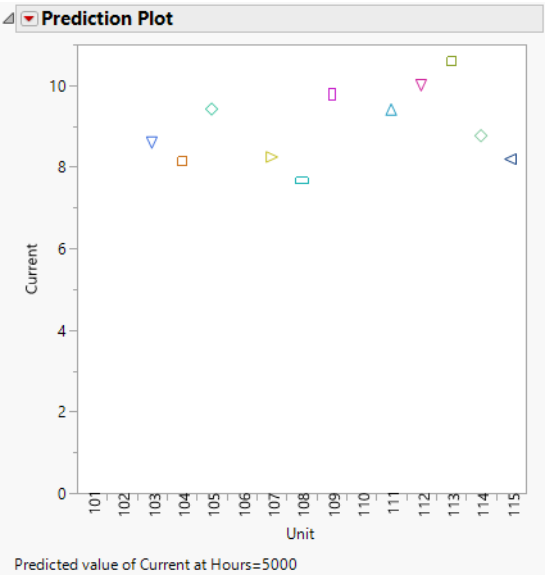
Use the Prediction Graph tab to predict the Y variable for a specified Time value.

Figure 7.15 Prediction Plot Tab



For the GaAs Laser example, no data was collected after 4000 hours. If you want to predict the percent increase in operating current after 5000 hours, enter 5000 and click **Go**. A plot is produced showing the estimated percent decrease after 5000 hours for all the units.

Figure 7.16 Prediction Plot



The Prediction Plot red triangle menu contains the following options:

Save Predictions Saves the predicted Y values to a data table. When one of the Longitudinal Prediction Interval options is enabled, the table also includes the intervals.

Longitudinal Prediction Interval Shows or hides confidence or prediction intervals for the estimated Y on the Prediction Plot. When intervals are enabled, the intervals are also included in the data table that is created when using the Save Predictions option.

Longitudinal Prediction Time Specifies the time value for which you want to predict the Y.

Longitudinal Prediction Alpha Specifies the alpha level used for the intervals.

Degradation Platform Options

The Degradation Data Analysis red triangle menu contains the following options:

Path Definition The Y variable at a given time is assumed to have a distribution. You can model the mean, location parameter, or median of that distribution.

Mean Path Specifies that the mean is the parameter to be modeled.

Location Parameter Path Specifies that the location parameter of the distribution is the parameter to be modeled.

Median Path Specifies that the median of the distribution is the parameter to be modeled.

When the Location Parameter or Median Path option is selected, a menu appears in the Model Specification that enables you to select the distribution of the response.

Degradation Path Style Contains options for selecting the style of degradation path to fit.

Simple Linear Path Enables you to fit linear degradation paths and nonlinear paths that can be transformed to linear paths. See "Simple Linear Path".

Nonlinear Path Enables you to fit nonlinear degradation paths. See "Nonlinear Path".

Graph Options Contains options for modifying the appearance of graphs in the report.

Connect Data Markers Shows or hides lines connecting the points on the Overlay plot.

Show Fitted Lines Shows or hides the fitted lines on the Overlay plot.

Show Spec Limits Shows or hides the specification limits on the Overlay plot.

Show Residual Plot Shows or hides the residual plot.

Show Inverse Prediction Plot Shows or hides the inverse prediction plot

Show Curve Interval Shows or hides the confidence intervals on the fitted lines on the Overlay plot.

Curve Interval Alpha Enables you to change the alpha used for the confidence interval curves.

Show Median Curves Shows or hides median lines on the plot when the Path Definition is set to Location Parameter Path.

Show Legend Shows or hides a legend for the markers used on the Overlay plot.

No Tab List Shows or hides the Residual Plot, Inverse Prediction, and Prediction Graph in tabs or in stacked reports.

Prediction Settings Contains options for modifying the settings used in the model predictions.

Upper Spec Limit Specifies the upper specification limit.

Lower Spec Limit Specifies the lower specification limit.

Censoring Time Specifies the censoring time. See "Inverse Prediction".

Baseline Specifies the normal use conditions for an X variable in nonlinear degradation paths. A path with this value usually results in an Overlay plot.

Inverse Prediction Specifies the interval type, alpha level, and one- or two-sided intervals for inverse prediction. To do inverse prediction, you must also specify the lower or upper specification limit. See "Inverse Prediction".

Longitudinal Prediction Specifies the Time value, interval type, and alpha level for longitudinal prediction. See "Prediction Graph".

Applications Contains options for further analysis of the degradation data.

Generate Pseudo Failure Data Creates a data table giving the predicted time each unit crosses the specification limit. The table contains a Life Distribution or Fit Life by X script that can be used to fit a distribution to the pseudo failure times.

Test Stability Performs stability analysis. See "Stability Analysis".

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Model Reports

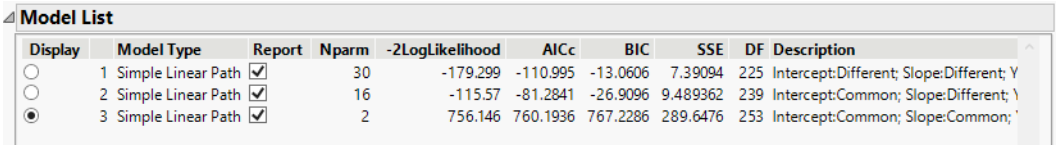
When the **Generate Report for Current Model** button is clicked, summary reports are added in two places:

- An entry is added to the Model List report. See "Model Lists".
- An entry is added to the Reports report. See "Reports".

Model Lists

The Model List report gives summary statistics and other options for every fitted model. Figure 7.17 shows an example of the Model List with summaries for three models.

Figure 7.17 Model List



Model List

Display		Model Type	Report	Nparm	-2LogLikelihood	AICc	BIC	SSE	DF	Description
○	1	Simple Linear Path	☑	30	-179.299	-110.995	-13.0606	7.39094	225	Intercept:Different; Slope:Different; Y
○	2	Simple Linear Path	☑	16	-115.57	-81.2841	-26.9096	9.489362	239	Intercept:Common; Slope:Different; \
◉	3	Simple Linear Path	☑	2	756.146	760.1936	767.2286	289.6476	253	Intercept:Common; Slope:Common;

Display Select the model that you want represented in the Overlay plot, Residual Plot, Inverse Prediction plot, and Prediction Graph.

Model Type Gives the type of path, either linear or nonlinear.

Report Select the check boxes to display the report for a model. For more information about the reports, see "Reports".

Nparm Gives the number of parameters estimated for the model.

-2LogLikelihood Gives twice the negative of the log-likelihood. See *Fitting Linear Models*.

AICc Gives the corrected Akaike Criterion. See *Fitting Linear Models*.

BIC Gives the Bayesian Information Criterion. See *Fitting Linear Models*.

SSE Gives the error sums-of-squares for the model.

DF Gives the error degrees of freedom.

Description Gives a description of the model.

Reports

The Reports outline node gives details about each model fit. The report includes a Model Summary report, and an Estimate report.

The Model Summary report contains the following information:

<Y, Response> Scale The transformation on the response variable.

<Time> Scale The transformation on the time variable.

SSE The error sums-of-squares.

Nparm The number of parameters estimated for the model.

DF The error degrees of freedom.

RSquare The R-square.

MSE The mean square error.

The Estimate report contains the following information:

Parameter The name of the parameter.

Estimate The estimate of the parameter.

Std Error The standard error of the parameter estimate.

t Ratio The t statistic for the parameter, computed as the Estimate divided by the Std Error.

Prob>|t| The *p*-value for a two-sided test for the parameter.

Destructive Degradation

To measure a product characteristic, sometimes the product must be destroyed. For example, when measuring breaking strength, the product is stressed until it breaks. Regular degradation analysis no longer applies in these situations. You can handle these situations in one of two ways:

- If your failure time model is a standard one, it might be covered by the Destructive Degradation platform. See "Destructive Degradation".
- If you are using custom transformations of the time or response variables and custom nonlinear models, use the Degradation platform. In the launch window, select the Destructive Degradation tab.

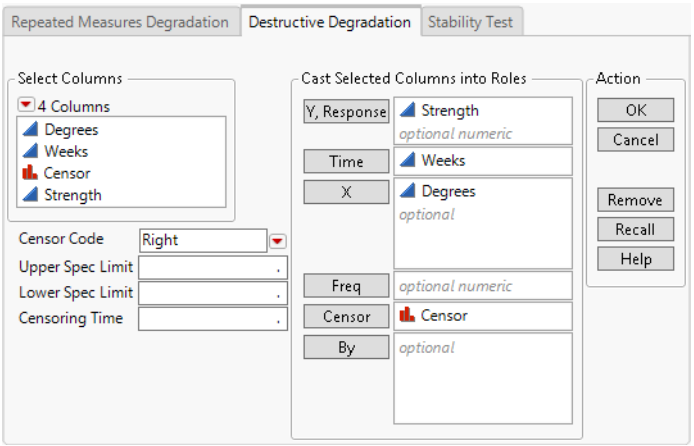
Example of Accelerated Destructive Degradation

This example fits a custom nonlinear model. The data consist of measurements on the strength of an adhesive bond. The product is stressed until the bond breaks, and the required breaking strength is recorded. Because units at normal use conditions are unlikely to break, the units were tested at several levels of an acceleration factor (temperature). You want to estimate the strength (in newtons) of units after 52 weeks (one year) at use conditions of 25°C.

Complete the Launch Window

1. Select **Help > Sample Data Library** and open Reliability/Adhesive Bond.jmp.
2. Select **Analyze > Reliability and Survival > Degradation**.
3. Select the **Destructive Degradation** tab.
4. Select Strength and click **Y, Response**.
5. Select Weeks and click **Time**.
6. Select Degrees and click **X**.
7. Select Censor and click **Censor**.

Figure 7.18 Completed Launch Window



8. Click **OK**.

Define and Fit the Model

1. Select **Lognormal** from the menu in the Location Parameter Path Specification panel in the Model Specification outline.

   This specifies a lognormal transformation for the response, Strength.
2. Click the Degradation Data Analysis red triangle and select **Degradation Path Style > Nonlinear Path**.

This adds a script window to the report. You insert a script that specifies your model in this window.

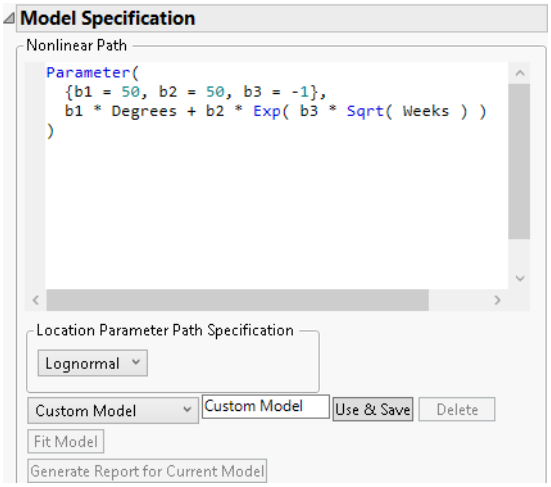
3. Copy the JSL formula below and paste it into the script window under Nonlinear Path:

```
Parameter(
   {b1 = 50, b2 = 50, b3 = -1},
   b1 * :Degrees + b2 * Exp( b3 * Sqrt( :Weeks ) )
)
```

The script defines a model for Strength in terms of parameters b1, b2, and b3. The script specifies initial values for the parameters.

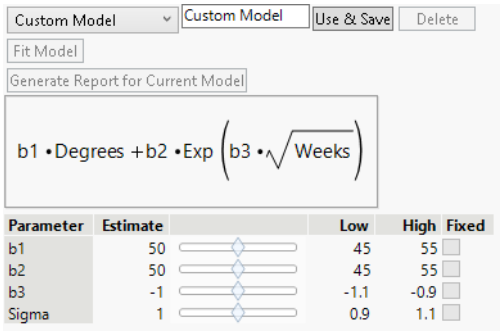
4. In the text box at the bottom of the report, change Empty to Custom Model.

Figure 7.19 Nonlinear Path Script



5. Click **Use & Save**.

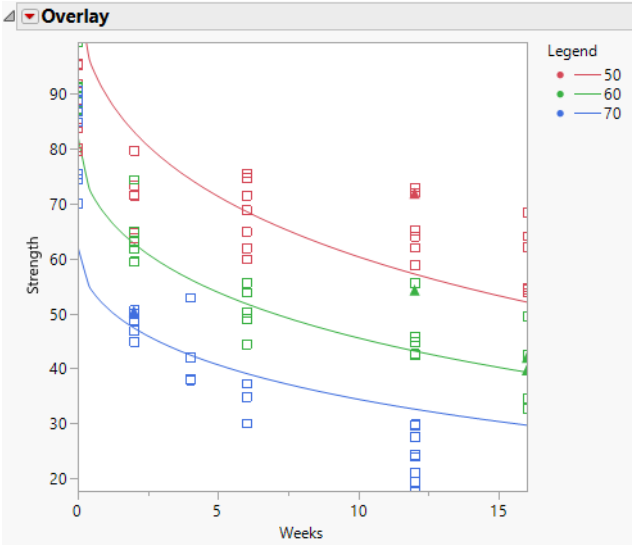
Figure 7.20 Updated Model Specification Outline



The report is updated to include controls for changing initial values for the parameters. Initial values for the parameters are set to the values specified in the formula in the script editor.

6. Click **Fit Model**.

Figure 7.21 Plot of Fitted Model



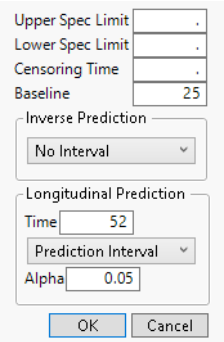
The Parameter panel in the Model Specification outline is updated to show the parameter estimates for the fitted model. The model fit is shown in the Overlay plot. You can drag the axes to show the points, as is done in Figure 7.21. The legend identifies the curves.

Obtain Predicted Values and Prediction Intervals

Next, you obtain a predicted value and prediction interval for Strength after 52 weeks at the baseline use condition of 25°C.

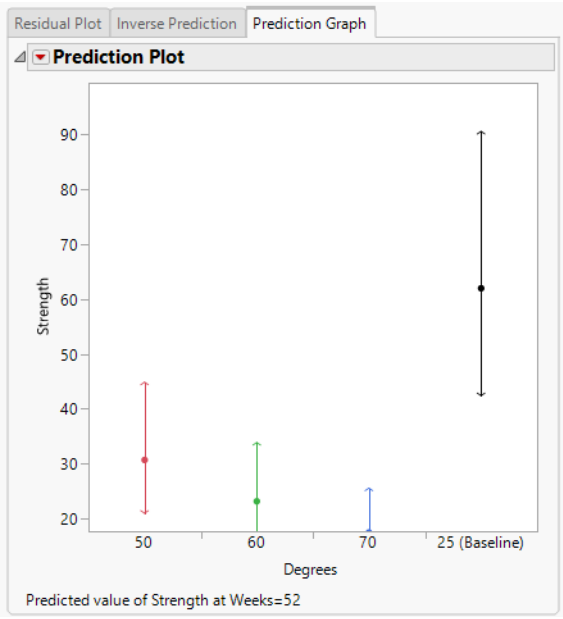
1. Click the Degradation Data Analysis red triangle and select **Prediction Settings**.
2. In the **Prediction Settings** window:
   - Enter 25 for Baseline.
   - Enter 52 for Time in the Longitudinal Prediction panel.
   - Select Prediction Interval from the menu in the Longitudinal Prediction panel.

Figure 7.22 Prediction Settings



3. Click **OK**.
4. Select the **Prediction Graph** tab in the report.

Figure 7.23 Prediction Plot at Weeks = 52



The Prediction Plot shows the predicted value and its 95% level prediction interval above the axis label of 25 (Baseline).

5. Click the Prediction Plot red triangle and select **Save Predictions**.

   Predicted values for the three values of Degrees and for the desired baseline of 25 degrees are saved to a data table. The predicted strength of the adhesive bond after 1 year at 25° C is 61.96, with a prediction interval of 42.42 to 90.50.

Stability Analysis

Stability analysis is used in setting pharmaceutical product expiration dates. Three linear degradation models are fit, and an expiration date is estimated following International Conference on Harmonisation (ICH) guidelines. The ICH guidelines are used for the general framework of determining if batches can be pooled for expiration dating (ICH Q1E 2003). For specific implementation details, see the STAB macro and FDA guidelines in Chow (2007, Appendix B).

The Stability Tests report summarizes three degradation models and their corresponding earliest crossing times. The best model is selected and displayed in the Overlay plot. The models are listed in the order of complexity.

The first model fits different intercepts and different slopes to each batch. (In the procedure steps and model comparisons described below, this is the *full model*.) When this model is used for estimating the expiration date, the mean square error (MSE) is not pooled across batches. Prediction intervals are computed for each batch using individual mean squared errors, and the interval that crosses the specification limit first is used to estimate the expiration date. The earliest crossing times are based on 95% two-sided prediction intervals when there are two specification limits provided; the earliest crossing times are based on 95% one-sided prediction intervals when there is only one specification limit provided.

The second model fits different intercepts to each batch, but fits a common slope across all batches. The third model fits a common intercept and a common slope across all batches. When this model is appropriate, it provides the expiration date furthest into the future.

Figure 7.24 Stability Tests Summary

Stability Tests

The best model accepted at the significance level of 0.25 has Different intercepts and Common slopes. The model suggests the earliest crossing time at 23.47505 with 95 percent confidence. ICH Guidelines indicate an expiration time of 23.47505.

Display	Intercept	Slope	Earliest Crossing Time
○	Different	Different	23.11596
◉	Different	Common	23.47505
○	Common	Common	29.98567

The Model Comparisons section of the Stability Tests report summarizes the tests of significance for each of the stability models. The Legend describes each source. The procedure for determining the best model for expiration dating considers the *p*-values for Sources C and B. The sources are listed below in reverse order to accommodate the order of steps in the procedure.

Tip: The ICH guidelines specify using a significance level of 0.25 to determine the appropriate model. You can follow the steps below and compare the *p*-values at each source to a different significance level. If this results in a different model being selected, you can use the radio buttons in the Display column of the table in the Stability Tests summary report.

Source E Specifies the sum of squared responses minus the Source D sum of squares. The value in the Mean Square column for Source E is the Source E SS value divided by the Source E degrees of freedom, which are equal to the number of parameters in the full model (different intercepts and different slopes).

Source D Specifies the error sum of squares and corresponding MSE for the full model (different intercepts and different slopes).

Source C Specifies the test of equal slopes. This is a test of the second model (different intercepts and common slope) versus the full model (different intercepts and different slopes).

- If the *p*-value is less than 0.25, the slopes are assumed to be different across batches. The procedure stops and the full model (different intercepts and different slopes) is used to estimate the expiration date.
- If the *p*-value is greater than or equal to 0.25, the slopes are assumed to be common across batches and you then evaluate the intercepts using Source B.

Source B Specifies the test for equal intercepts. This is a test of the third model (common intercepts and common slopes) versus the second model (different intercepts and common slope).

- If the *p*-value is less than 0.25, the intercepts are assumed to be different across batches, and the second model (different intercepts and common slope) is used to estimate the expiration date.
- If the *p*-value is greater than or equal to 0.25, the intercepts are assumed to be common across batches, and the third model (common intercepts and common slopes) is used to estimate the expiration date.

Source A Specifies the test of the third model (common intercepts and common slopes) versus the full model (different intercepts and different slopes). This test is not used in the procedure to determine the best model for expiration dating.

Figure 7.25 Stability Model Comparisons

Model Comparisons

Source	DF	SS	Mean Square	F Statistic	Prob>F
A	6	61.48956	10.24826	10.10836	<.0001*
B	3	60.48866	20.16289	19.88764	<.0001*
C	3	1.000906	0.333635	0.329081	0.8043
D	28	28.38752	1.01384		
E	8	360214.5	45026.81		

Legend

Source		Intercept	Slope	Intercept	Slope
A		Different	Different	Common	Common
B		Different	Common	Common	Common
C		Different	Different	Different	Common
D	Residual				
E	Whole Model				

The Reports section contains models fit by the Test Stability option. The following four models are fit by the Test Stability option:

- A different intercept, different slope model where the MSE (mean squared error) is pooled across batches. (This is an alternative to the first model in the Summary report.)
- A different intercept, common slope model. (This is the second model in Summary report.)
- A common intercept, common slope model. (This is the third model in Summary report.)
- A different intercept, different slope model where the MSE (mean squared error) is not pooled across batches. (This is the first model in the Summary report.)

Caution: In addition to the four models fit by the Test Stability option, the Reports section also contains any models fit in the Degradation platform prior to running the Test Stability option.

Example

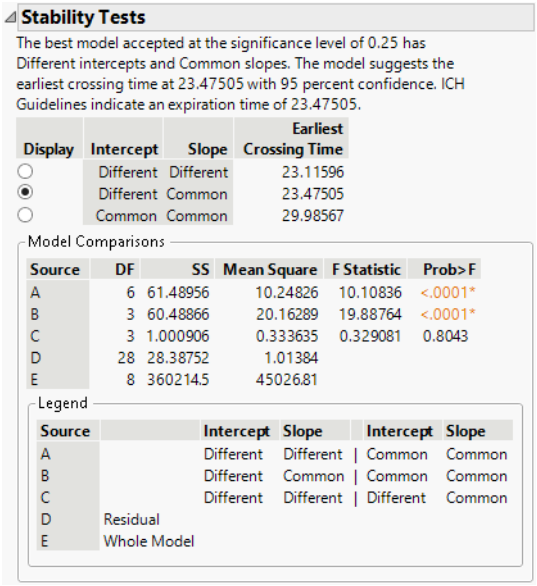
Use the data in the Stability.jmp sample data table to establish an expiration data for a new product. The data consists of product concentration measurements on four batches. A concentration of 95 is considered the end of the product's usefulness.

To perform the stability analysis, do the following steps:

1. Select **Help > Sample Data Library** and open Reliability/Stability.jmp.
2. Select **Analyze > Reliability and Survival > Degradation**.
3. Select the **Stability Test** tab.
4. Select Concentration (mg/Kg) and click **Y, Response**.
5. Select Time and click **Time**.
6. Select Batch Number and click **Label, System ID**.
7. Enter 95 for the Lower Spec Limit.

8. Click **OK**.

Figure 7.26 Stability Models



Stability Tests

The best model accepted at the significance level of 0.25 has Different intercepts and Common slopes. The model suggests the earliest crossing time at 23.47505 with 95 percent confidence. ICH Guidelines indicate an expiration time of 23.47505.

Display	Intercept	Slope	Earliest Crossing Time
○	Different	Different	23.11596
◉	Different	Common	23.47505
○	Common	Common	29.98567

Model Comparisons

Source	DF	SS	Mean Square	F Statistic	Prob>F
A	6	61.48956	10.24826	10.10836	<.0001*
B	3	60.48866	20.16289	19.88764	<.0001*
C	3	1.000906	0.333635	0.329081	0.8043
D	28	28.38752	1.01384		
E	8	360214.5	45026.81		

Legend

Source		Intercept	Slope	Intercept	Slope
A		Different	Different	Common	Common
B		Different	Common	Common	Common
C		Different	Different	Different	Common
D	Residual				
E	Whole Model				

The test for equal slopes has a *p*-value of 0.8043. Because this is larger than a significance level of 0.25, the test is not rejected, and you conclude the degradation slopes are equal between batches.

The test for equal intercepts and slopes has a *p*-value of <.0001. Because this is smaller than a significance level of 0.25, the test is rejected, and you conclude that the intercepts are different between batches.

Because the test for equal slopes was not rejected, and the test for equal intercepts was rejected, the chosen model is the one with Different Intercepts and Common Slope. This model is the one selected in the report, and gives an estimated expiration date of 23.475.

Chapter 8

Destructive Degradation

Model Product Deterioration over Time

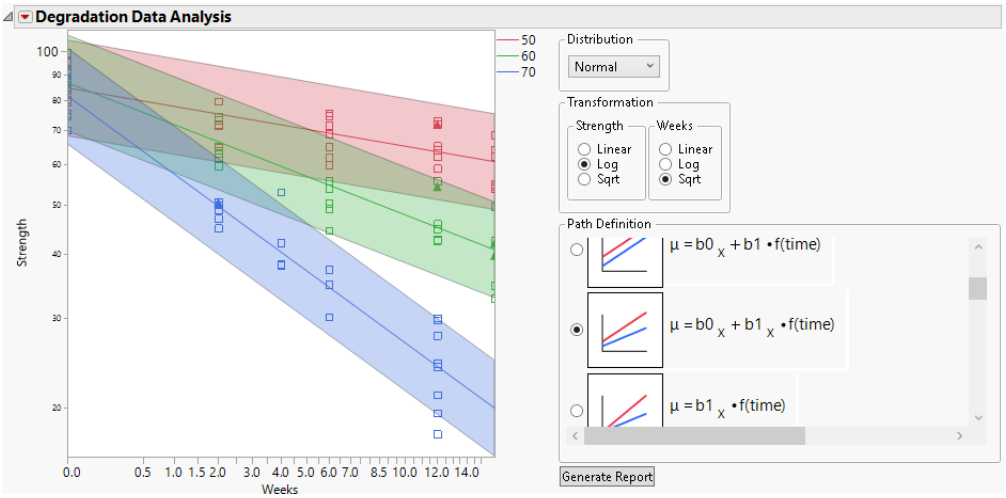
To measure a product characteristic, sometimes the product must be destroyed. For example, when measuring breaking strength, the product is stressed until it breaks. Because the test is destructive, there is only one observation per product unit. In such a situation, you can model product reliability using the Destructive Degradation platform.

The platform models how a (typically) nonnegative response changes over time. Observations are assumed to be independent and measure the value of the response and the time at failure. A large and flexible collection of pre-defined models is provided. The models include location-scale and log-location-scale distributions whose location parameters are functions of time. The models allow explanatory variables and additional parameters. When an explanatory variable is specified, the platform fits an accelerated destructive degradation model.

Note: If you require a model that is not represented among the models provided, you can use the Degradation platform. See "Destructive Degradation".

For more information about destructive degradation and reliability, see Escobar et al. (2003) and Meeker and Escobar (1998).

Figure 8.1 Destructive Degradation Example of Model for Adhesive Bond.jmp



Contents

Example of the Destructive Degradation Platform 211
Launch the Destructive Degradation Platform 217
Specify Two Y Columns 218
The Destructive Degradation Plot Options and Models 219
Plot Options 220
Models 221
The Destructive Degradation Report 225
Model List 226
Model Outlines 226
Destructive Degradation Platform Options 229
Statistical Details for the Destructive Degradation Platform 229

Example of the Destructive Degradation Platform

This example of an accelerated destructive degradation model is patterned after an example from Escobar et al. (2003). The data consist of measurements on the strength (measured in newtons) of an adhesive bond. Temperature is considered to be an acceleration factor. The product is stressed until the bond breaks and the required breaking stress is recorded. Because units at normal temperatures are unlikely to break, the units were tested at several levels over a wide range of temperatures. Strength less than 50 newtons is considered failure. You want to estimate the proportion of units with a strength below 50 newtons after 260 weeks (5 years) at a reference temperature of 35 degrees Celsius.

This example has three stages:

- "Perform the Initial Analysis"
- "Change the Model and Generate the Report"
- "Using the Profilers for Prediction"

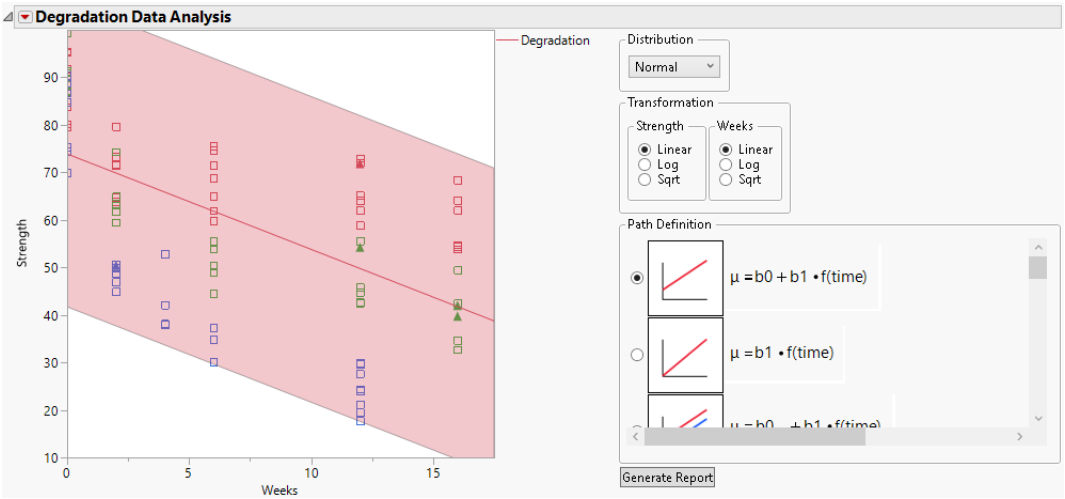
Perform the Initial Analysis

1. Select **Help > Sample Data Library** and open Reliability/Adhesive Bond.jmp.
2. Select **Analyze > Reliability and Survival > Destructive Degradation**.
3. Select Strength and click **Y, Response**.
4. Select Weeks and click **Time**.
5. Select Degrees and click **X**.

   The temperature is the accelerating factor in the experiment.
6. Select Censor and click **Censor**.

   Notice that the **Censor Code** is set to Right.
7. Click **OK**.

Figure 8.2 Initial Degradation Plot



The platform specifies a default model. The default model assumes that the data are described by a single Normal distribution, whose location parameter is a linear function of time.

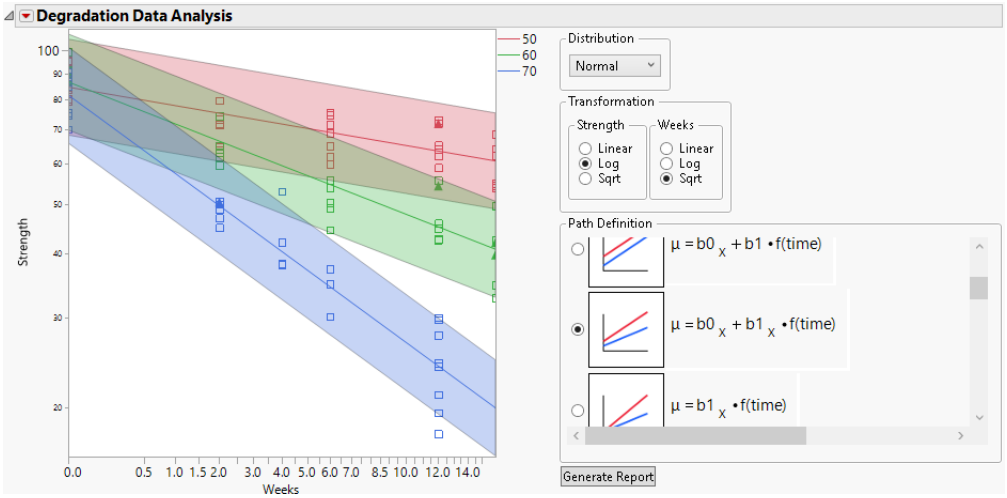
Change the Model and Generate the Report

1. Select **Log** for the Y (Strength) Transformation.
2. Select **Sqrt** for the Time (Weeks) Transformation.
3. Select $\mu = b0_x + b1_x * f(time)$ for the Path Definition.

   The subscript "x" denotes the accelerating variable, which is Degrees in this example.

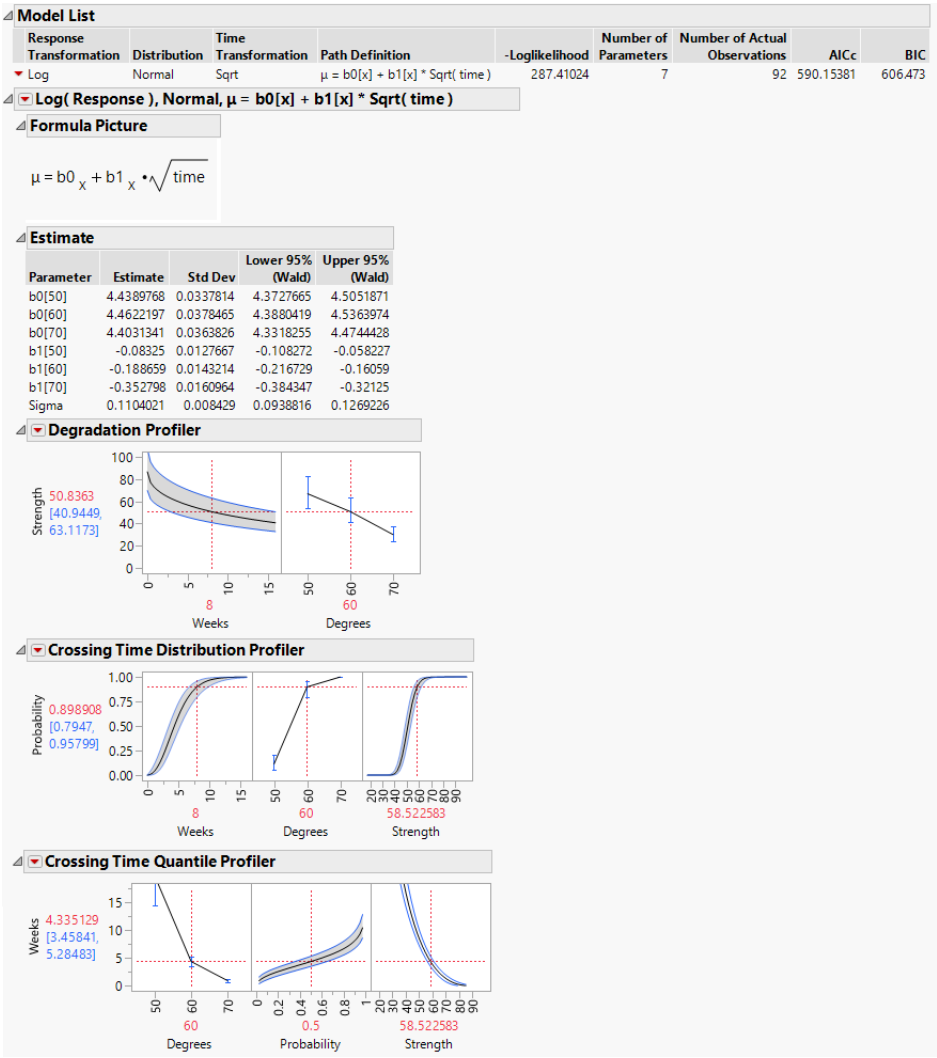
Note: This model is linear in all parameters.

Figure 8.3 Plot Showing Model



4. Click **Generate Report**.

Figure 8.4 Report for Basic Model



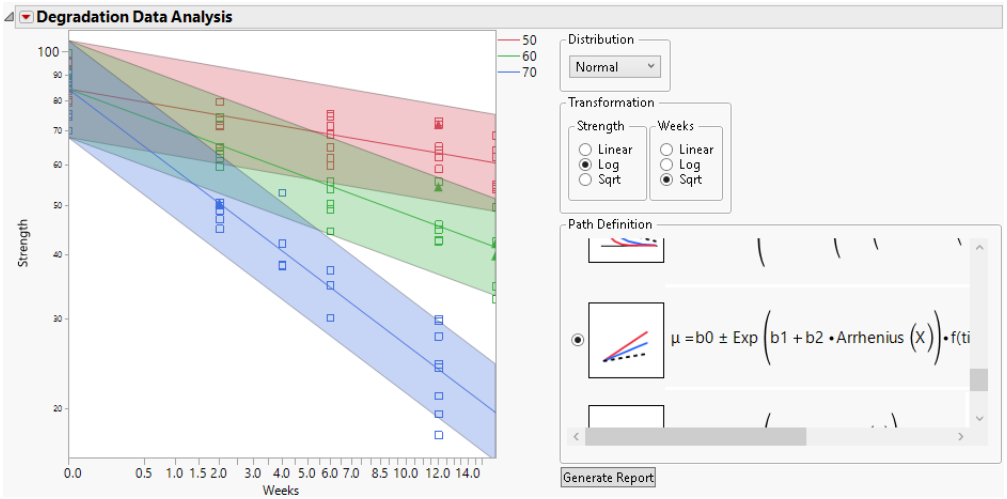
The estimates of the slope b1 at the three values of Degrees suggest that degradation occurs more quickly at higher temperatures. Failure mechanisms that depend on chemical processes are often well modeled using the Arrhenius model for temperature. For this reason, you now fit a model where an Arrhenius transformation is applied to Degrees, which is measured on a Celsius scale.

5. Select μ = b0 ± Exp(b1 + b2*Arrhenius(X))*f(time) for the Path Definition.

Note: This model is not linear in the parameters.

6. Select **Celsius** and click **OK**.

Figure 8.5 Plot Showing Model with Arrhenius Transformation



7. Click **Generate Report**.

Figure 8.6 Report Including Second Model with Arrhenius Transformation

Model List

Response Transformation	Distribution	Time Transformation	Path Definition	-Loglikelihood	Number of Parameters	Number of Actual Observations	AICc	BIC
Log	Normal	Sqrt	μ = b0[x] + b1[x] * Sqrt(time)	287.41024	7	92	590.15381	606.473
Log	Normal	Sqrt	μ = b0 - Exp(b1 + b2 * Arrhenius(X)) * Sqrt(time)	288.09559	4	92	584.65095	594.27834

Log(Response), Normal, μ = b0[x] + b1[x] * Sqrt(time)

Log(Response), Normal, μ = b0 - Exp(b1 + b2 * Arrhenius(X)) * Sqrt(time)

Formula Picture

$$\mu = b0 - \text{Exp}\left(b1 + b2 \bullet \text{Arrhenius}\left(X\right)\right) \bullet \sqrt{\text{time}}$$

Estimate

Parameter	Estimate	Std Dev	Lower 95% (Wald)	Upper 95% (Wald)
b0	4.4354686	0.0205347	4.3952214	4.4757158
b1	22.825503	1.4029332	20.075805	25.575202
b2	-0.704728	0.0414658	-0.786	-0.623457
Sigma	0.1114956	0.0085011	0.0948337	0.1281574

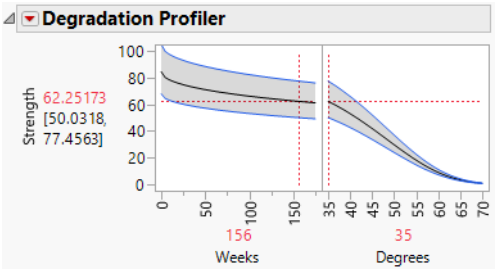
Using the Profilers for Prediction

Because the Arrhenius model shows a better fit, as indicated by its smaller AICc and BIC values (Figure 8.6), you continue your analysis using this model.

Recall that strength less than 50 newtons is considered failure. You are interested in units lasting 156 weeks (three years) at a reference temperature of 35 degrees Celsius. Change the settings in the profilers to reflect these values. Click the value in red beneath each plot's horizontal axis and enter the new value.

1. In the Degradation profiler for the Arrhenius model, set **Weeks** to 156 and **Degrees** to 35.

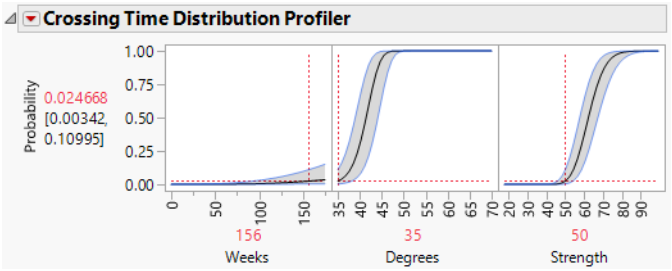
Figure 8.7 Degradation Profiler



The predicted **Strength** at these settings is 62.25173, with a 95% prediction interval ranging from 50.0318 to 77.4563. Failures are not very likely at these or less extreme settings.

2. In the Crossing Time Distribution Profiler, set **Weeks** to 156, **Degrees** to 35, and **Strength** to 50.

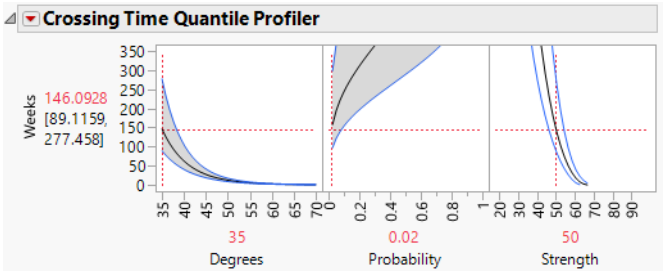
Figure 8.8 Crossing Time Distribution Profiler



At 156 weeks at a temperature of 35 degrees Celsius, the probability that the value of Strength is less than 50 is 0.024668. The 95% confidence interval ranges from 0.00342 to 0.10995. The probability of failure at these or less extreme conditions is about 2%.

3. In the Crossing Time Quantile Profiler, set **Degrees** to 35, **Probability** to 0.02, and **Strength** to 50.
4. Adjust the vertical axis of the Crossing Time Quantile Profiler so that the maximum value is about 350.

Figure 8.9 Crossing Time Quantile Profiler

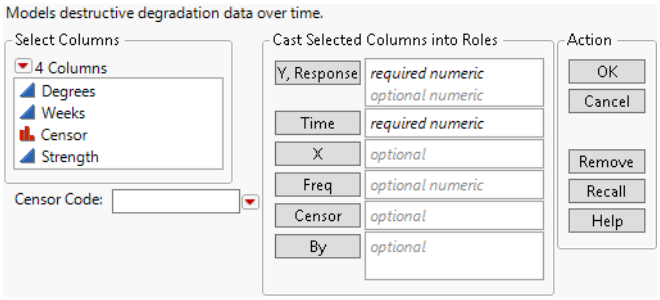


The number of weeks within which 2% of units fail at a temperature of 35 degrees Celsius is estimated to be 146.0928.The 95% confidence interval ranges from 89.1159 to 277.458.

Launch the Destructive Degradation Platform

Launch the Destructive Degradation platform by selecting **Analyze** > **Reliability and Survival** > **Destructive Degradation**. Figure 8.10 shows the Destructive Degradation launch window using the Adhesive Bond.jmp data table (located in the Reliability folder).

Figure 8.10 The Destructive Degradation Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

The Destructive Degradation launch window contains the following options:

Y, Response Identifies the column containing the degradation measurements. When your response values are interval-censored, you can enter two columns. See "Specify Two Y Columns".

Time Identifies the column containing the time values.

X Identifies an optional explanatory variable. If the distribution of Y changes not only over time, but is also impacted by some other variable, that additional variable can be supplied

as X. Use this role to specify the accelerating factor in an accelerated destructive degradation model.

Freq Identifies frequencies or observation counts when there are multiple units. If the value is 0 or a positive integer, then the value represents the frequencies or counts of observations with the given row's settings.

Censor Identifies an optional column that identifies censored response measurements. When there is only one column in the Y role, this column indicates whether the response Y in a given row is exact or right censored. A right-censored observation is one where the exact measurement is unknown, but is known to be larger than the Y value in the corresponding row.

Select the value that identifies right-censored observations from the Censor Code menu. Rows corresponding to other values in the Censor column are treated as uncensored. Rows with missing censor code values are excluded from the analysis. JMP attempts to detect the censor code and display it in the list.

By Identifies an optional By variable. A separate analysis is produced for each level of this variable.

Censor Code Identifies the value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

Specify Two Y Columns

You can specify two Y columns when some of the degradation measurements are interval censored or left censored. For a given row, the values in the two Y columns determine the type of censoring.

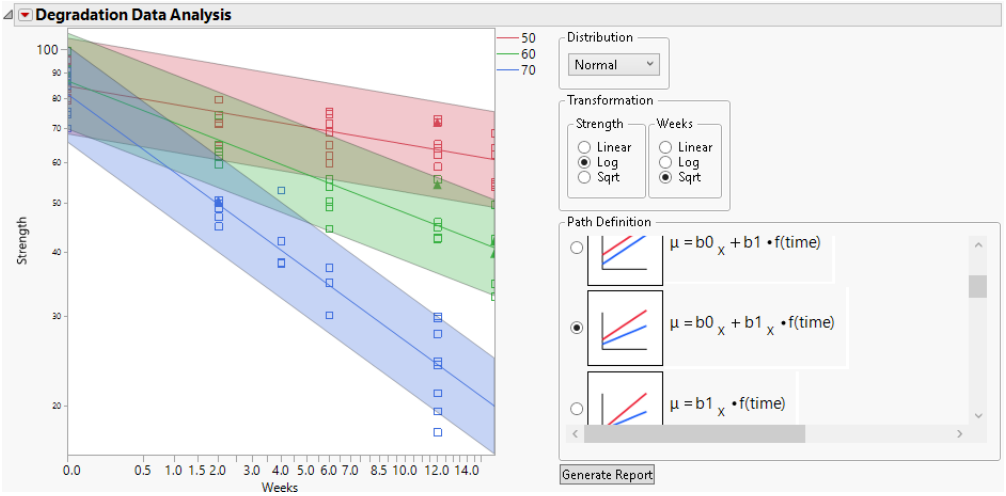
- If the two Y values are equal and neither is missing, then the common measurement is treated as exact.
- If the two Y values are not equal and neither is missing, then the measurement is interval censored and assumed to be between the two values.
- If only the first value is missing, then the measurement is left censored and assumed to be smaller than the second value.
- If only the second value is missing, then the measurement is right censored and assumed to be larger than the first value.

Note: The only way to fit left-censored measurements in the Destructive Degradation platform is through the use of two Y columns.

The Destructive Degradation Plot Options and Models

The Degradation Data Analysis plot shows the data and a graphical representation of the model that is currently specified based on selections of Distribution, Transformation, and Path Definition. The plot shown in Figure 8.11 is for the Adhesive Bond.jmp data table and represents a model that includes an optional X variable.

Figure 8.11 Destructive Degradation Plot and Options



Note the following about the plot:

- Data values are represented by markers.

Table 8.1 Marker Descriptions

Icon	Description
□	An exact measurement.

Table 8.1 Marker Descriptions *(Continued)*

Icon	Description
▼	Left-censored observation, indicating that the censored measurement is below the triangle. **Note:** In the Degradation platform, left-censoring arises only when observations are interval censored. See "Specify Two Y Columns".
▲	Right-censored observation, indicating that the censored measurement is above the triangle.
▼ ▲	Interval-censored observation, indicating that the censored value is within the specified interval.

Note: By default, the markers are not colored. To color them to match the model color scheme, select **Rows > Color or Mark by Column**. Select the X column. Deselect **Continuous Scale**. Select **JMP Default** from the Colors menu.

- For each level of X, a colored band appears. If the model does not include an X variable, then a single band appears. For a given value of Time, the upper and lower bounds of the band are the 0.025 and 0.975 percentiles of the fitted distribution of Y. The colors of the bands correspond to the values of X, as indicated by the legend to the upper right of the plot.

Note: Marker colors correspond to the color states assigned in the data table.

- The solid curve in the center of a band is the median of the fitted distribution of Y for the corresponding value of X over time. If the model does not include an X variable, then the curve plots the median of Y over time.

Plot Options

Distribution Choose a location-scale or a log-location-scale distribution.

Note: It is not recommended to fit a log-location-scale distribution model when you specify a Log transformation for the response column.

Transformation Choose a transformation function for the response Y and for the Time variable.

Note: If you apply the Log transformation to a column that contains nonpositive values, the rows with nonpositive values are omitted from the model fit. If you apply the Sqrt transformation to a column that contains negative values, the rows with negative values are omitted from the model fit.

Path Definition Choose a linear or a nonlinear path for the regression model. For more information about each model, see "Models".

Generate Report Creates a report for the specified model. The first time you select Generate Report, a Model List outline is created. When you select Generate Report to fit other models, the Model List outline is updated and an outline is added for each model.

Models

The following table provides the equations for each model in the Path Definition list. For a description of each model, follow the link.

Note: The thumbnail sketch shown to the left of each equation shows a generic plot of the behavior of the location parameter, μ, over time. In the report's main plot, the plot of the estimated median can differ from the thumbnail based on your selections for Distribution and Transformation.

Table 8.2 Model Equations

Model	Equation
"Common Path with Intercept"	$\mu = b0 + b1 * f(time)$
"Common Path without Intercept"	$\mu = b1 * f(time)$
"Common Slope"	$\mu = b0_X + b1 * f(time)$
"Individual Path with Intercept"	$\mu = b0_X + b1_X * f(time)$
"Individual Path without Intercept"	$\mu = b1_X * f(time)$

Table 8.2 Model Equations *(Continued)*

Model	Equation
"Common Intercept"	$\mu = b0 + b1_X * f(time)$
"First-Order Kinetics Type 1"	$\mu = b0 - b1 * Exp[-b2 * Exp[b3 * [Arrhenius(X_0) - Arrhenius(X)]] * f(time)]$
"First-Order Kinetics Type 2"	$\mu = b0 * [1 - Exp[-b1 * Exp[b2 * [Arrhenius(X_0) - Arrhenius(X)]] * f(time)]]$
"First-Order Kinetics Type 3"	$\mu = b0 + b1 * Exp[-b2 * Exp[b3 * [Arrhenius(X_0) - Arrhenius(X)]] * f(time)]$
"First-Order Kinetics Type 4"	$\mu = b0 * Exp[-b1 * Exp[b2 * [Arrhenius(X_0) - Arrhenius(X)]] * f(time)]$
"Arrhenius Rate"	$\mu = b0 \pm Exp[b1 + b2 * Arrhenius(X)] * f(time)$
"Polynomial Rate"	$\mu = b0 \pm Exp[b1 + b2 * Log(X)] * f(time)$
"Exponential Rate"	$\mu = b0 \pm Exp[b1 + b2 * X] * f(time)$

Models That Are Linear in Transformed Time

Common Path with Intercept

This model fits a single distribution whose location parameter changes linearly over transformed time. This model fits a common intercept and a common slope, regardless of whether there is an X variable.

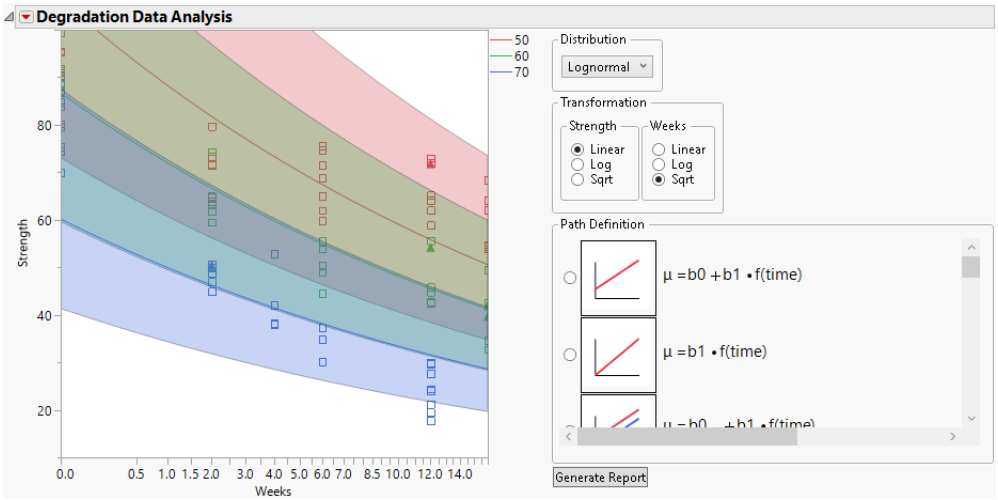
Common Path without Intercept

This model fits a single distribution whose location parameter changes linearly over transformed time but whose value at time zero is zero. Based on selections for Distribution and Transformation, the median curve might not be a straight line and might not pass through origin. This model fits a zero intercept and a common slope, regardless of whether there is an X variable.

Common Slope

In this model, the location parameters are linear functions of transformed time with separate intercepts for the values of X but a common slope. Based on selections for Distribution and Transformation, the model fits might appear as curves. For example, selecting a Lognormal distribution gives the plot in Figure 8.12.

Figure 8.12 Common Slope Model Using a Lognormal Distribution



Individual Path with Intercept

In this model, the location parameters are linear functions of transformed time with separate intercepts and separate slopes for the values of X.

Individual Path without Intercept

In this model, the location parameters are linear functions of transformed time with zero intercepts and separate slopes for the values of X.

Common Intercept

In this model, the location parameters are linear functions of transformed time with a common intercept and separate slopes for the values of X.

First-Order Kinetics Models

Four first-order kinetics models where the location parameter is a nonlinear function based on an Arrhenius transformation of temperature are provided. Each of these location models fit separate models for each value of the optional explanatory variable X.

When you first select any of these models, you must specify the measurement scale for temperature and a reference temperature value, X_0. The specified reference temperature affects the interpretation of *b2* in these models. The *b2* parameter is the rate constant at X_0. The value for X_0 is used to construct a time acceleration factor (Meeker and Escobar 1998). If you subsequently select another of the first-order kinetics models or the Arrhenius Rate model (see "Arrhenius Rate"), the platform remembers and uses these specifications.

First-Order Kinetics Type 1

In this model, *b1* and *b2* are positive. On a linear scale, the curves have an upper asymptote at *b0* as time approaches infinity.

First-Order Kinetics Type 2

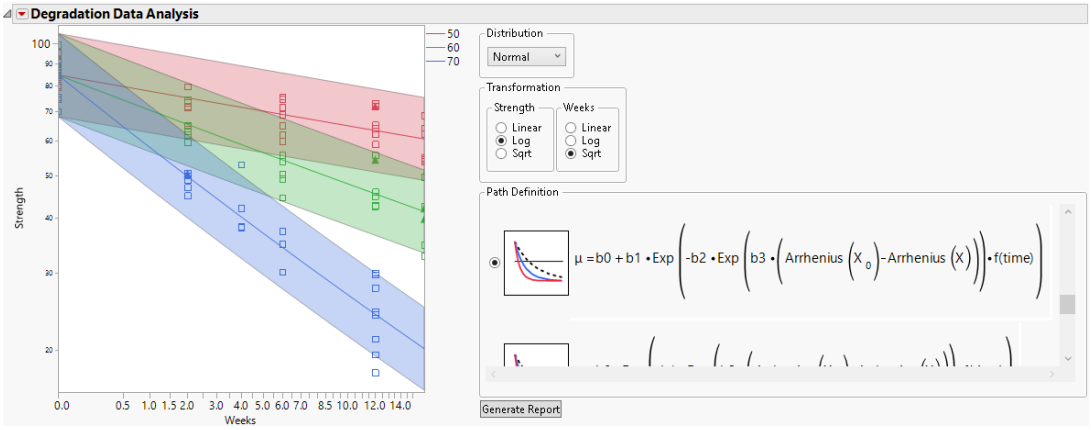
In this model, the location parameter is zero at time zero. Both *b0* and *b1* are positive. On a linear scale, the curves have an upper asymptote at *b0* as time approaches infinity. You can think of the Type 2 model as a vertically shifted version of the Type 1 model.

First-Order Kinetics Type 3

In this model, both *b1* and *b2* are positive. Because of the sign preceding *b1* is the opposite of the sign preceding *b1* in the Type 1 model, this model is an inverted version of the Type 1 model. On a linear scale, it has a lower asymptote at *b0* as time approaches infinity.

Given data exhibiting a negative slope over time, the fitted model can produce a plot similar to Figure 8.13. The figure is for Adhesive Bond.jmp. The selected temperature measurement scale is Celsius and the specified reference temperature under typical use conditions is 35 degrees.

Figure 8.13 Example of First-Order Kinetics Model Type 3



First-Order Kinetics Type 4

This model is a vertically shifted version of the Type 3 model. On a linear scale, the curves have a lower asymptote at 0 as time approaches infinity.

Rate Models

Three models where the location parameter is an exponential function of the transformed X variable are provided. Each of these location models fits common intercept and separate slope models for each value of the optional explanatory variable X. For each of these models, on a linear scale, the location parameter is linear in the transformed time values.

Arrhenius Rate

This model involves an exponential function of the Arrhenius transformation multiplied by transformed time. When you select this model, you are asked to specify the measurement scale for temperature, unless you have already supplied this information.

Polynomial Rate

This model involves an exponential function of a linear function of the log of X multiplied by transformed time.

Exponential Rate

This model involves an exponential function of a linear function of X multiplied by transformed time.

The Destructive Degradation Report

The Destructive Degradation report contains an outline for each model that you fit. When you fit a model, the Model List is updated with a row for that model.

Note: All models are fit using the maximum likelihood method.

Figure 8.14 Model List and Model Outlines

Model List

Response Transformation	Distribution	Time Transformation	Path Definition	-Loglikelihood	Number of Parameters	Number of Actual Observations	AICc	BIC
Log	Normal	Sqrt	μ = b0[x] + b1[x] * Sqrt(time)	287.41024	7	92	590.15381	606.473
Log	Normal	Sqrt	μ = b0 - Exp(b1 + b2 * Arrhenius(X)) * Sqrt(time)	288.09559	4	92	584.65095	594.27834

Log(Response), Normal, μ = b0[x] + b1[x] * Sqrt(time)

Log(Response), Normal, μ = b0 - Exp(b1 + b2 * Arrhenius(X)) * Sqrt(time)

Model List

The first four columns in the Model List reflect the choices that you made in the plot options. The -Loglikelihood, AICc, and BIC statistics are information-based measures that can be used for model comparisons. For descriptions of these measures, see *Fitting Linear Models*.

The three information-based measures in the Model List are comparable across models as long as the models being compared have the same Number of Actual Observations. If this is not the case, exercise caution because different models might use different subsets. The Number of Actual Observations might be reduced due to the choice of distribution or the choice of transformation. Choosing a log-location-scale distribution excludes all non-positive Y values. Also, the Log and Sqrt transformations exclude all non-positive values.

Each row of the Model List table has a red triangle menu with the following options:

Scroll To Scrolls the report window to the corresponding model outline.

Remove Removes the model from the Model List and removes the corresponding model outline from the report.

Model Outlines

The outline for each model contains a red triangle menu with the following option:

Remove Removes the model outline from the report and removes the model from the Model List.

The outline for each model contains the following reports:

Formula Picture Shows the equation for the location parameter.

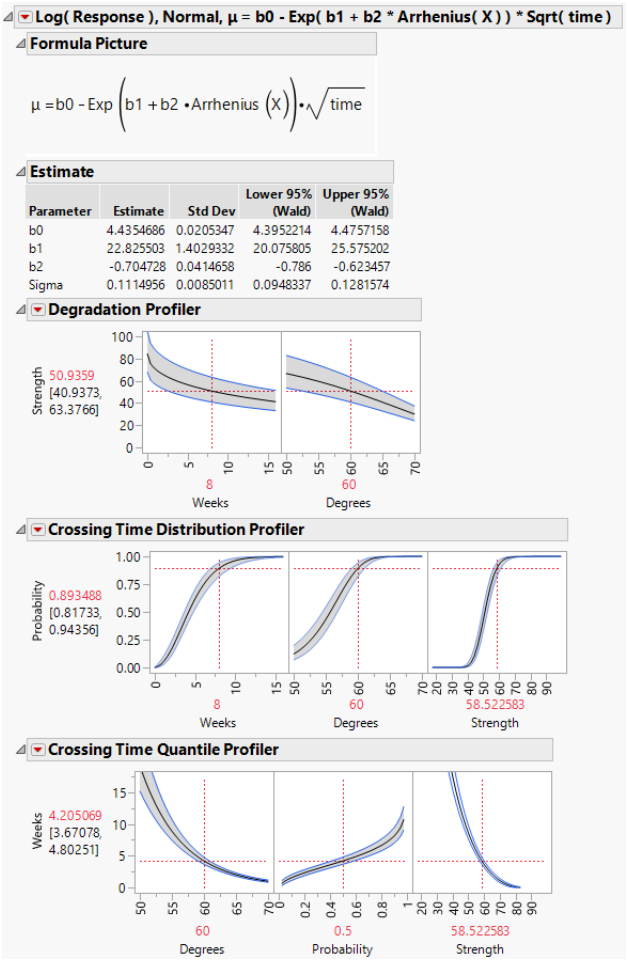
Estimate Shows parameter estimates and their standard deviations. This report also contains 95% Wald-based confidence intervals for the parameters.

Note: If you specify a reference temperature (X_0), this value is also displayed at the top of the Estimate report.

Distribution, Quantile, and Inverse Prediction For more information about the profilers, see "Profilers".

Residual Plots For more information about the residual plots, see "Residual Plots".

Figure 8.15 Reports within a Model Outline



Profilers

Three profilers appear in the model report.

Degradation Profiler

The Degradation Profiler shows a profiler view of the Degradation Data Analysis plot for the given model. The response is the degradation response Y. The profiler includes a plot against the Time variable and a plot against the optional explanatory variable X (if you have specified one). The plot against Time shows the median of the fitted distribution of Y as a solid curve. The dashed curves show the 0.025 and 0.975 percentiles of the fitted distribution of Y.

Crossing Time Distribution Profiler

Use the Crossing Time Distribution Profiler to determine the probability that the degradation measurement falls below a given threshold at some point in time.

The profiler plots the estimated cumulative distribution function of the response Y as a function of Time, the optional X variable, and Y. The plots for Time and Y show Wald confidence intervals.

Crossing Time Quantile Profiler

Use the Crossing Time Quantile Profiler to determine the time at which a specified proportion of measurements falls below a given threshold value.

The profiler plots the estimated Time as a function of the optional X variable, quantile values for Y (Probability), and Y. The plots for Probability and Y show Wald confidence intervals.

Residual Plots

Four residual plots appear in the model report. Use these plots to validate the distributional assumption for the model. For more information about the standardized residuals, see "Statistical Details for the Destructive Degradation Platform".

Cox-Snell Residual P-P Plot

If the points deviate far from the diagonal, then the distributional assumption might be violated. The Cox-Snell Residual P-P Plot red triangle menu has an option called Save Residuals that enables you to save the residual data to the data table. See Meeker and Escobar (1998, sec. 17.6.1) for a discussion of Cox-Snell residuals.

Probability Plot of Standardized Residuals

If the points deviate far from the diagonal, then the distributional assumption might be violated. The Probability Plot of Standardized Residuals red triangle menu has an option called Save Residuals that enables you to save the residual data to the data table.

Standardized Residuals versus Time

Use the Standardized Residuals versus Time plot to examine differences in variability over time. The Standardized Residuals versus Time red triangle menu has an option called Save Residuals that enables you to save the residual data to the data table.

Standardized Residuals versus Predicted

Use the Standardized Residuals versus Predicted plot to examine differences in variability over the range of predicted values. The Standardized Residuals versus Predicted red triangle menu has an option called Save Residuals that enables you to save the residual data to the data table.

Destructive Degradation Platform Options

The Degradation Data Analysis red triangle menu contains the following options:

Graph Options Provides options for modifying the platform graphs.

Shade Turns the shading on or off. By default, the upper and lower bounds of each shaded band at a given value of time correspond to the 0.025 and 0.975 quantiles of the distribution of Y.

Shade Coverage If shading is on, you can enter a proportion to increase or decrease the amount of shading coverage.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Statistical Details for the Destructive Degradation Platform

The destructive degradation model can be expressed as follows:

$$g(Y) \sim F(\mu, \sigma)$$

$$\mu = h(f(Time), X)$$

where:

- $g(Y)$ is the transformed Y variable
- F is the selected probability distribution
- μ is the location parameter, defined by h
- h is a function that relates the transformed Time variable and the explanatory variable X
- σ is the scale parameter of the distribution
- f(Time) is the transformed Time variable
- X is an optional explanatory variable

The standardized residuals are obtained as follows:

- For location-scale distributions, the standardized residuals are $(y\text{-}\mu)/\sigma$.
- For log-location-scale distributions, the standardized residuals are $(log(y)\text{-}\mu)/\sigma$.

Chapter 9

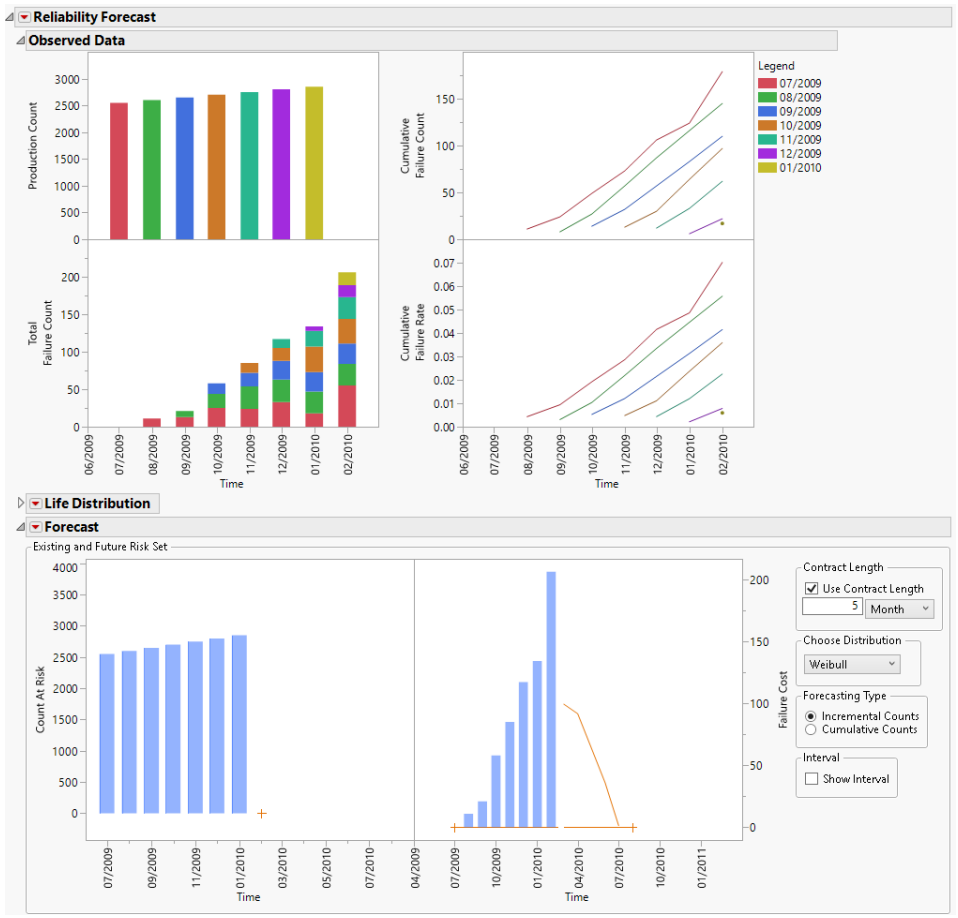
Reliability Forecast

Forecast Product Failure Using Production and Failure Data

The Reliability Forecast platform helps you predict the number of future failures. JMP estimates the parameters for a life distribution using production dates, failure dates, and production volume.

Using the interactive graphs, you can adjust factors such as future production volumes and contract length to estimate future failures. Repair costs can be incorporated into the analysis to forecast the total cost of repairs across all failed units.

Figure 9.1 Example of a Reliability Forecast



Contents

Overview of the Reliability Forecast Platform 233
Example Using the Reliability Forecast Platform 233
Launch the Reliability Forecast Platform 237
The Reliability Forecast Report 240
Observed Data Report 240
Life Distribution Report 240
Forecast Report 241
Reliability Forecast Platform Options 245
Additional Example Using the Reliability Forecast Platform 246

Overview of the Reliability Forecast Platform

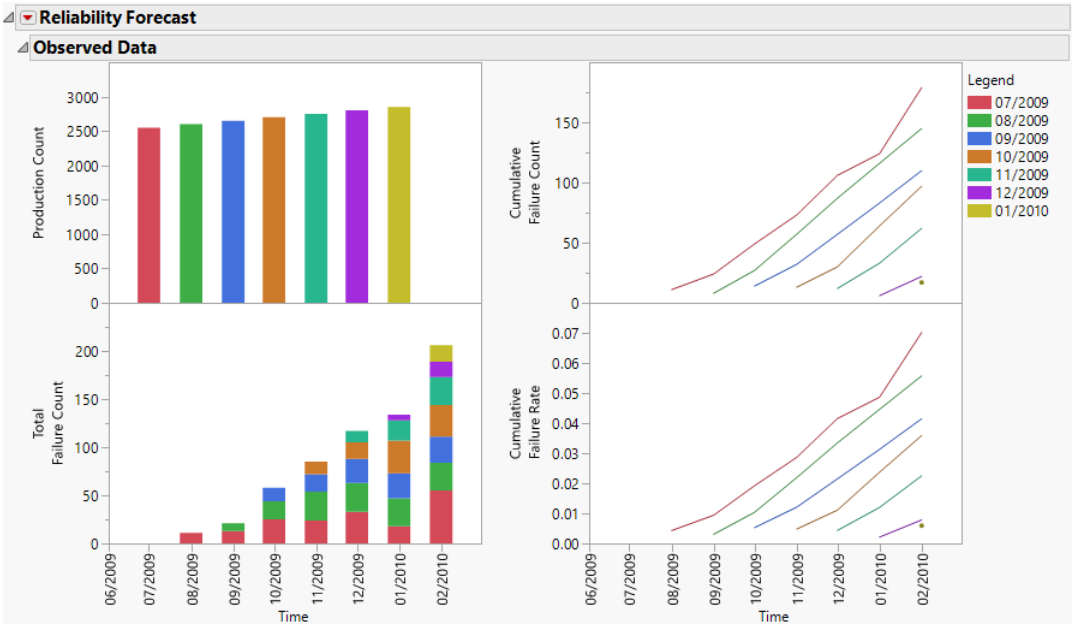
The Reliability Forecast platform helps you predict the number of future failures. JMP estimates the parameters for a life distribution using production dates, failure dates, and production volume. Using the interactive graphs, you can adjust factors such as future production volumes and contract length to estimate future failures. Repair costs can be incorporated into the analysis to forecast the total cost of repairs across all failed units.

Example Using the Reliability Forecast Platform

You have data on seven months of production and returns. You want to use this information to forecast the total number of units that will be returned for repair through February 2011. The product has a 12-month contract.

1. Select **Help > Sample Data Library** and open Reliability/Small Production.jmp.
2. Select **Analyze > Reliability and Survival > Reliability Forecast**.
3. On the **Nevada Format** tab, select Sold Quantity and click **Production Count**.
4. Select Sold Month and click **Timestamp**.
5. Select the other columns and click **Failure Count**.
6. Click **OK**.

Figure 9.2 Observed Data Report

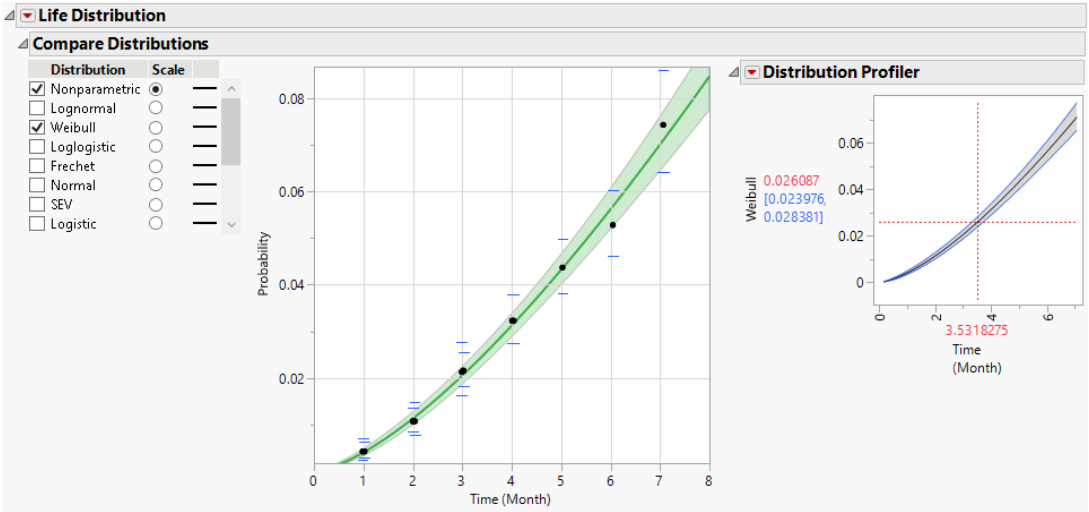


In the Observed Data report, the bottom left shows bar charts of previous failures. Cumulative failures are shown in the line graphs on the right. Note that production levels are fairly consistent. As production accumulates over time, more units are at risk of failure, so the cumulative failure rate gradually increases. The consistent production levels also result in similar cumulative failure rates and counts from month to month.

7. Click the **Life Distribution** disclosure icon.

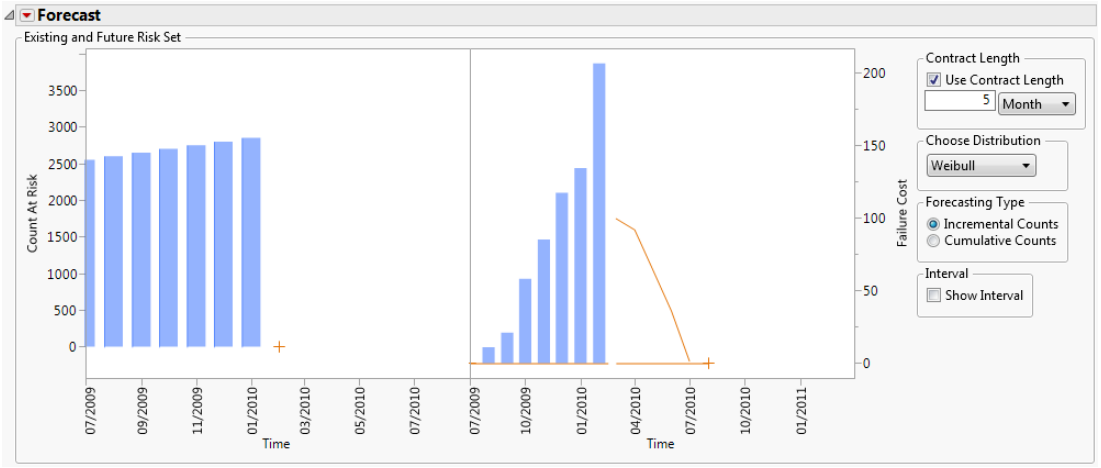
   JMP fits production and failure data with a Weibull distribution using the Life Distribution platform (Figure 9.3). JMP then uses the fitted Weibull distribution to forecast returns for the next five months (Figure 9.4).

Figure 9.3 Life Distribution Report



The Forecast report shows previous production in the left graph (Figure 9.4). In the right graph, you see that the number of previous failures increased steadily over time.

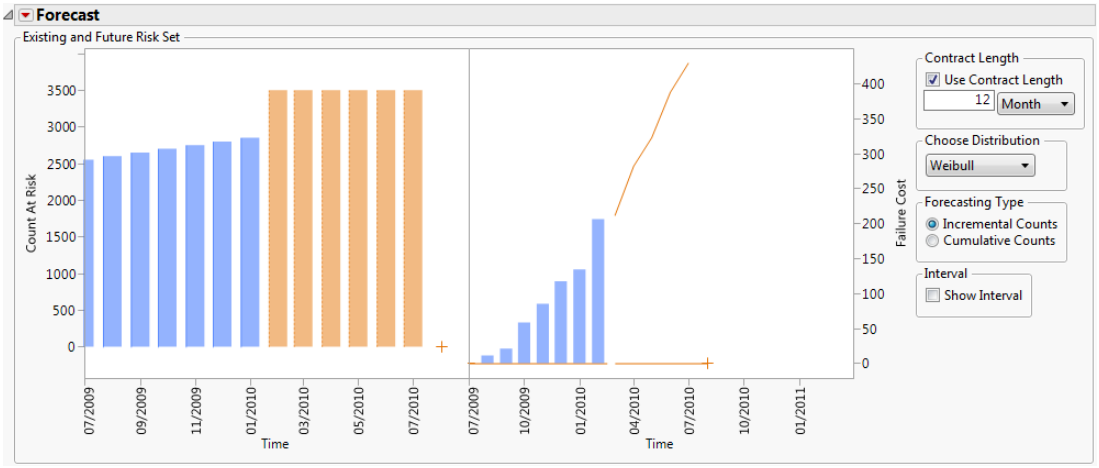
Figure 9.4 Forecast Report



8. In the Forecast report, type 12 for the Contract Length.
9. In the left Forecast graph, drag the animated hotspot over to July 2010 and upward to approximately 3500.

   Orange bars appear in the left graph to represent future production. The monthly returned failures in the right graph increase gradually through August 2010.

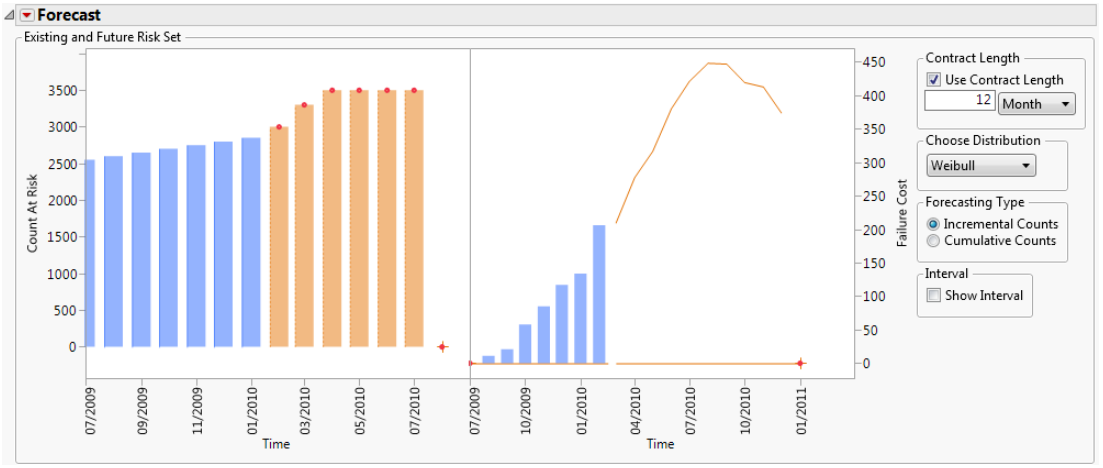
Figure 9.5 Production and Failure Estimates



10. In the left graph, drag the February 2010 hotspot to approximately 3000 and then drag the March 2010 hotspot to approximately 3300.
11. In the right graph, drag the right hotspot to February 2011.

    JMP estimates that the number of returns will gradually increase through August 2010 and decrease by February 2011.

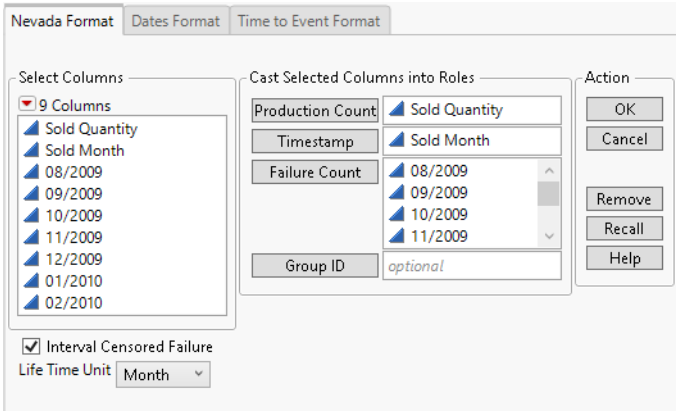
Figure 9.6 Future Production Counts and Forecasted Failures



Launch the Reliability Forecast Platform

Launch the Reliability Forecast platform by selecting **Analyze > Reliability and Survival > Reliability Forecast**.

Figure 9.7 The Reliability Forecast Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

The launch window includes a tab for each contract data format: Nevada, Dates, and Time to Event (for time to failure data). The following sections describe these formats.

Nevada Format

Contract data is commonly stored in the Nevada format: shipment or production dates and failure counts within specified periods are shaped like the state of Nevada. Figure 9.8 shows the Small Production.jmp sample data table.

Figure 9.8 Example of the Nevada Format

	Sold Quantity	Sold Month	08/2009	09/2009	10/2009	11/2009	12/2009	01/2010	02/2010
1	2550	07/2009	11	13	25	24	33	18	55
2	2600	08/2009	0	8	19	30	30	29	29
3	2650	09/2009	0	0	14	18	25	26	27
4	2700	10/2009	0	0	0	13	17	34	33
5	2750	11/2009	0	0	0	0	12	21	29
6	2800	12/2009	0	0	0	0	0	6	16
7	2850	01/2010	0	0	0	0	0	0	17

The Nevada Format tab contains the following options:

Interval Censored Failure Considers the returned quantity to be interval censored. The interval is between the last recorded time and the time that the failure was observed. This option is selected by default.

Life Time Unit Physical date-time format of all time stamps, including column titles for return counts. The platform uses this setting in forecasting step increments.

Production Count Column that identifies the number of units produced.

Timestamp Column that contains the production date.

Failure Count Column that contains the number of failed units.

Group ID Column by which observations are grouped. Each group has its own distribution fit and forecast. A combined forecast is also included.

Dates Format

The Dates format focuses on production and failure dates. One data table specifies the production counts for each time period. The other table provides failure dates, failure counts, and the corresponding production times of the failures.

Figure 9.9 shows the Small Production part1.jmp and Small Production part2.jmp sample data tables.

Figure 9.9 Example of the Dates Format

production data

	Sold Quantity	Sold Month
1	2550	07/2009
2	2600	08/2009
3	2650	09/2009
4	2700	10/2009
5	2750	11/2009
6	2800	12/2009
7	2850	01/2010

failure data

	Return Month	Return Quantity	Sold Month
1	08/2009	11	07/2009
2	09/2009	13	07/2009
3	10/2009	25	07/2009
4	11/2009	24	07/2009
5	12/2009	33	07/2009
6	01/2010	18	07/2009
7	02/2010	55	07/2009

The Dates Format tab is divided into Production Data and Failure Data sections.

Production Data

Select Table Select the table that contains the number of units and the production dates.

Failure Data

Select Table Select the table that contains failure data, such as the number of failed units, production dates, and failure dates.

Left Censor Column that identifies censored observations.

Timestamp Column that links production observations to failure observations, indicating which batch a failed unit came from.

Censor Code Value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

Other options are identical to those on the Nevada Format tab. See "Nevada Format".

Time to Event Format

The Time to Event format shows production and failure data. Unlike the Nevada and Dates formats, Time to Event data does not include date-time information, such as production or failure dates. This format can also accommodate arbitrary censoring schemes in the data. See "Defining Risk Sets with Time to Event Data".

Figure 9.10 shows the Small Production Time to Event.jmp sample data table. For an example that uses time-to-event data, see "Additional Example Using the Reliability Forecast Platform".

Figure 9.10 Example of the Time to Event Format

start time | end time | failure counts

	Time (Month)	Time Right	Freq	Timestamp
1	0	1.0184804928	11	07/2009
2	1.0184804928	2.0369609856	13	07/2009
3	2.0369609856	3.022587269	25	07/2009
4	3.022587269	4.0410677618	24	07/2009
5	4.0410677618	5.0266940452	33	07/2009
6	5.0266940452	6.045174538	18	07/2009
7	6.045174538	7.0636550308	55	07/2009
8	7.0636550308	•	2371	07/2009

The Time to Event Format tab contains the following options:

Forecast Start Time Time at which the forecast begins. Enter the first value that you want on the horizontal axis. To enable forecasting, you must select Numeric for the Life Time Unit and set the Forecast Start Time to zero.

Censor Code Code for censored observations. Available only when you assign a **Censor** variable.

Other options are identical to those on the Nevada Format tab. See "Nevada Format".

The Reliability Forecast Report

The Reliability Forecast report contains the Observed Data report, Life Distribution report, and Forecast report. Use these reports to view the current data, compare distributions to find the right fit, and then adjust factors that affect forecasting. You can also save the forecast to a data table for use in financial forecasts. See Figure 9.1 for an example of the report.

The Reliability Forecast red triangle menu contains options for filtering the observed data by date and saving the data in another format. See "Reliability Forecast Platform Options".

Observed Data Report

The Observed Data report gives you a quick view of Nevada and Dates data (Figure 9.11).

- Bar charts show the production and failure counts for the specified production periods.
- Line charts show the cumulative forecast by production period.

Note: Time-to-event data does not include date-time information, such as production or failure dates, so the platform does not create an Observed Data report for this format.

Life Distribution Report

The Life Distribution report lets you compare distributions and work with profilers to find the right fit. When you select a distribution in the Forecast report, the Life Distribution report is updated. For more information about this report, see "Life Distribution". (Some of the options described in the Life Distribution chapter are not available in the Life Distribution report in the Reliability Forecast platform.)

Forecast Report

The Forecast report provides interactive graphs that help you forecast failures. By dragging hotspots, you can add anticipated production counts and see how they affect the forecast.

Adjusting Risk Sets

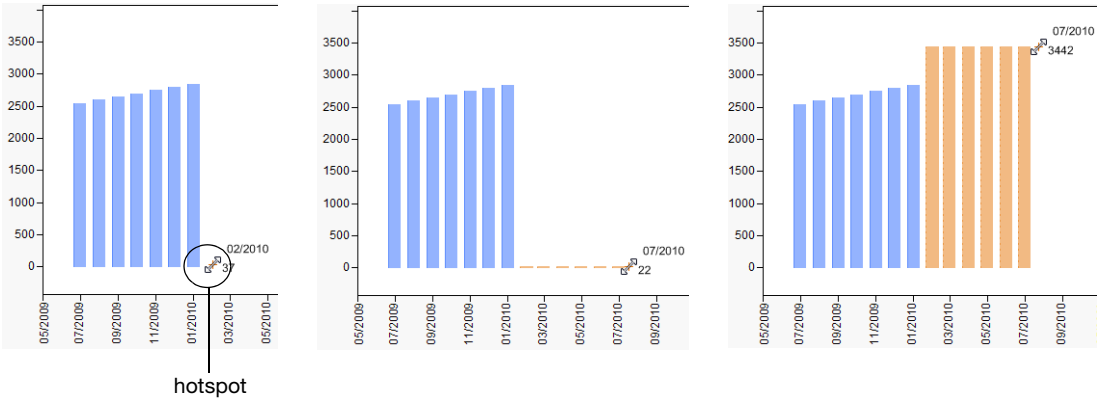
Future Production

In the left graph, the blue bars represent previous production counts. To add anticipated production, follow these steps:

1. Drag a hotspot to the right to add one or more production periods.

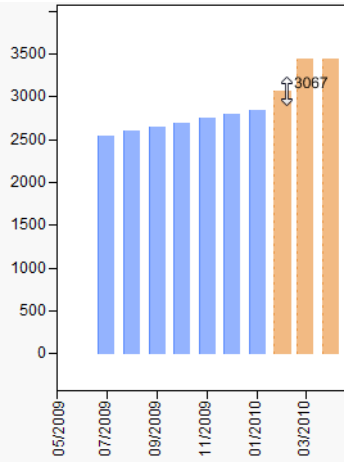
   The orange bars represent future production.

Figure 9.11 Add Production Periods



2. Drag each bar upward or downward to change the production count for each period.

Figure 9.12 Adjust Production Counts



Tip: To adjust future production counts, click the Forecast red triangle and select **Spreadsheet Configuration of Risk Sets**. See "Spreadsheet Configuration of Risk Sets".

Existing Production

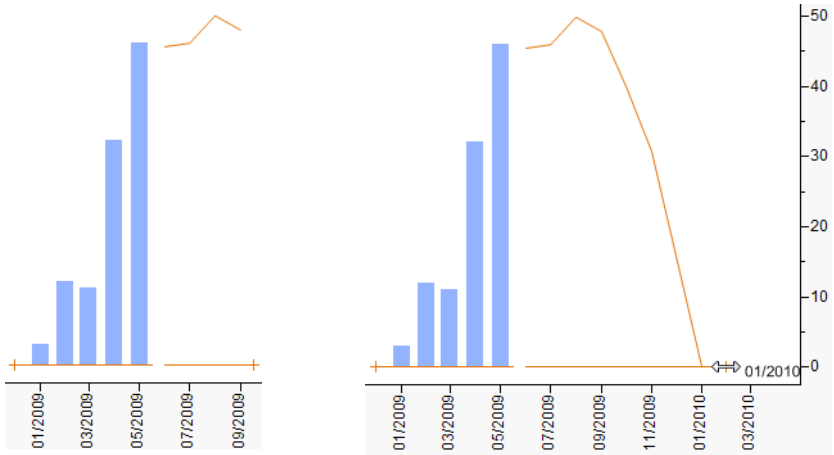
To remove a risk set from the forecast results, right-click a blue bar and select **Exclude**. To return that data to the risk set, right-click and select **Include**.

Tip: To adjust existing production counts, click the Forecast red triangle and select **Spreadsheet Configuration of Risk Sets**. See "Spreadsheet Configuration of Risk Sets".

Forecasting Failures

When you adjust production in the left graph, the right graph is updated to estimate future failures (Figure 9.13). Dragging a hotspot lets you change the forecast period. The orange line then shortens or lengthens to show the estimated failure counts.

Figure 9.13 Adjust the Forecast Period



Defining Risk Sets with Time to Event Data

You can obtain forecasts for arbitrary risk sets using the Time to Event Format tab. In the launch window's Time to Event Format tab, you must select Numeric for the Life Time Unit and set the Forecast Start Time to zero. Enter appropriate columns for Time to Event, Censor, Freq, and Group ID.

The plot on the left provides locations for bars representing existing production. Drag the hotspot at 0 to the left to create existing risk sets. Drag the hotspot at 1 to the right to create future production risk sets.

You can specify existing and future risk sets using the Spreadsheet Configuration of Risk Sets option. For time-to-event data, you must enter negative time values for the existing risk set in the Future Risk area. See "Spreadsheet Configuration of Risk Sets".

For an example that uses time-to-event data, see "Additional Example Using the Reliability Forecast Platform".

Forecast Graph Options

To explore your data further, you can change the contract length, distribution type, and other options.

- To forecast failures for a different contract period, enter the number next to **Use Contract Length**. Change the time unit if necessary.
- To change the distribution fit, select a distribution from the **Choose Distribution** list. The distribution is then fit to the future graph of future risk. The distribution fit appears in the Life Distribution report plot, and a new profiler is added. Changing the distribution fit in the Life Distribution report does not change the fit in the Forecast graph.

- If you are more interested in the total number of failures over time, select **Cumulative Counts**. Otherwise, JMP shows failures incrementally, which can make trends easier to identify.
- To show 95% confidence limits for the anticipated number of failures, select **Show Interval**.

Forecast Report Options

The Forecast red triangle menu contains the following options:

Animation Controls the flashing of the hotspots. You can also right-click a blue bar in the existing risk set and select or deselect **Animation**.

Interactive Configuration of Risk Sets Determines whether you can drag hotspots in the graphs.

Spreadsheet Configuration of Risk Sets Lets you specify production counts and periods (rather than adding them to the interactive graphs). You can also exclude production periods from the analysis.

- To remove an existing time period from analysis, highlight the period in the Existing Risk area, click, and then select **Exclude**. Or select **Include** to return the period to the forecast.
- To edit production, double-click in the appropriate Future Risk field and enter the new values.
- To add a production period to the forecast, right-click in the Future Risk area and select one of the **Append** options. (**Append Rows** adds one row; **Append N Rows** lets you specify the number of rows.)

As you change these values, the graphs update accordingly.

Note: If you launched the platform using the Time to Event Format tab, values for the existing risk set must be entered in the Future Risk area using negative time values.

Import Future Risk Set Lets you import future production data from another open data table. The new predictions then appear in the future risk graph. The imported data must have a column for timestamps and for quantities.

Show Interval Shows or hides 95% confidence limits in the graph. This option works the same as selecting **Show Interval** next to the graphs.

After you select **Show Interval**, the **Forecast Interval Type** option appears in the menu. Select one of the following interval types:

Plugin Interval Considers only forecasting errors given a fixed distribution.

Prediction Interval Considers forecasting errors when a distribution is estimated with estimation errors (for example, with a non-fixed distribution).

If you select Prediction Interval, the **Prediction Interval Settings** option appears in the menu. Approximate intervals are initially shown in the graph. Select **Monte Carlo Sample Size** or **Random Seed** to specify those values instead. To use the system clock, enter a missing number.

Use Contract Length Determines whether the specified contract length is considered in the forecast. This option works the same as selecting **Use Contract Length** next to the graphs.

Use Failure Cost Shows failure cost instead of the failure count in the future risk graph.

After you select **Use Failure Cost**, the **Set Failure Cost** option appears in the menu. This option enables you to set a cost for each failure. If you specified a Group variable in the launch window, the Set Failure Cost window enables you to specify separate costs for failures in each group.

Save Forecast Data Table Saves the cumulative and incremental number of returns in a new data table, along with the variables that you selected in the launch window. For grouped analyses, table names include the group ID and the word "Aggregated". Existing returns are also included in the aggregated data tables.

Reliability Forecast Platform Options

The Reliability Forecast red triangle menu contains the following options:

Save Data in Time to Event Format Saves Nevada or Dates data in a Time to Event formatted table.

Show Legend Shows or hides a legend for the Observed Data report. Unavailable for Time to Event data.

Show Graph Filter Shows or hides the Graph Filter so that you can select which production periods to show in the Observed Data graphs. Bars fade for deselected periods. Deselect the periods to show the graph in its original state. Unavailable for Time to Event data.

See *Using JMP* for more information about the following options:

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Additional Example Using the Reliability Forecast Platform

You have seven months of production and returns data that are stored in time-to-event format. You want to use this information to forecast the total number of units that will be returned for repair in the next 6 months. The product has a 4-month contract.

1. Select **Help > Sample Data Library** and open Reliability/Small Production Time to Event.jmp.
2. Select **Analyze > Reliability and Survival > Reliability Forecast**.
3. On the **Time to Event Format** tab, select Time (Month) and Time Right. Click **Time to Event**.
4. Select Freq and click **Freq**.
5. Click **OK**.
6. In the Forecast report, type 4 for the Contract Length.
7. (Optional.) Click the **Life Distribution** disclosure icon.

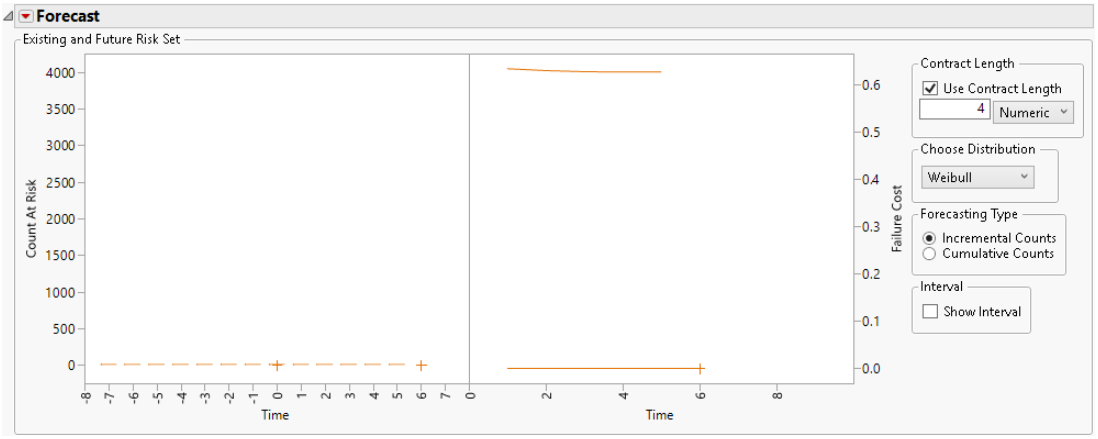
   The Life Distribution report shows a Weibull fit for the production and failure data. The Reliability Forecast platform then uses the fitted Weibull distribution to forecast returns.

Specify the Current and Future Production Counts

8. Set the vertical axis settings to prepare for past and future production counts. Right-click the vertical axis in the left Forecast graph and select **Axis Settings**. Specify the following axis settings:
   1. Type -250 for the Minimum.
   2. Type 4250 for the Maximum.
   3. Type 500 for the Increment.
   4. Click **OK**.
9. Set the horizontal axis to cover the previous 8 and next 8 months. Right-click the horizontal axis in the left Forecast graph and select **Axis Settings**. Specify the following axis settings:
   1. Type -8 for the Minimum.
   2. Type 8 for the Maximum.
   3. Click **OK**.

10. Set the past 7 months of production counts. In the left Forecast graph, drag the leftmost animated hotspot left to -7.
11. Set the future 5 months of production estimates. In the left Forecast graph, drag the rightmost animated hotspot right to 5.

Figure 9.14 Risk Set after Dragging Animated Hotspots



12. In the Forecast red triangle menu, select **Spreadsheet Configuration of Risk Sets**.
13. Drag the bottom of the Future Risk panel down so that you can see all 13 rows of the Future Risk table.

> **Note:** When using time-to-event formatted data, the existing production quantities must be specified in the Future Risk table using negative numeric time values.

14. Type in the Count values that are shown in Figure 9.15.

Figure 9.15 Future Risk Count Specifications

Future Risk

	Time	Count
1	-7	2550
2	-6	2600
3	-5	2650
4	-4	2700
5	-3	2750
6	-2	2800
7	-1	2850
8	0	3022
9	1	3307
10	2	3502
11	3	3502
12	4	3502
13	5	3502

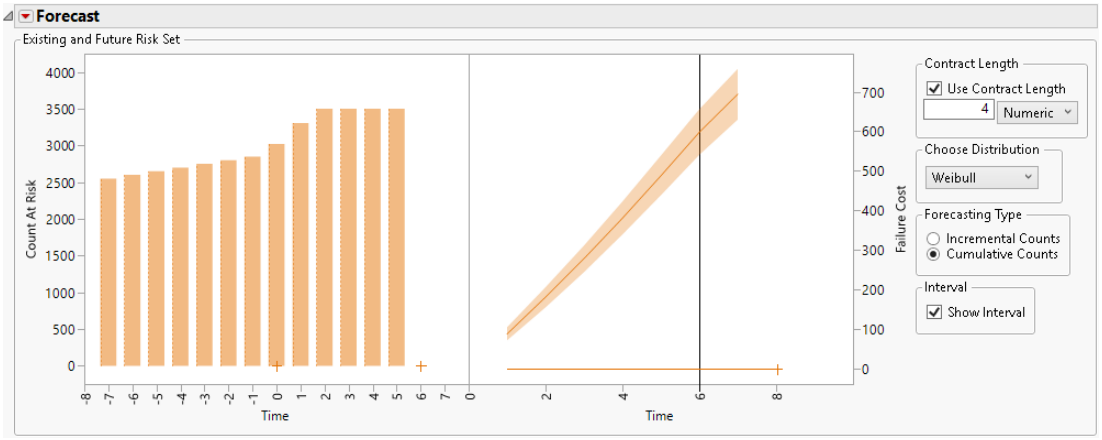
optional

15. In the right Forecast graph, drag the animated hotspot right to 8.
16. In the Forecasting Type panel, select **Cumulative Counts**.
17. In the Interval panel, select **Show Interval**.
18. Right-click the horizontal axis in the right Forecast graph and select **Axis Settings**. Add a reference line at month 6 using the following steps:
    1. Type 6 for the Value.
    2. Click **Add**.
    3. Click **OK**.

The forecast for the total number of units that will be returned for repair in the next 6 months is approximately 600 with confidence interval about 550 to 650.

Tip: You can use the **Save Forecast Data Table** option in the Forecast red triangle menu to see exact values for the forecasts and intervals.

Figure 9.16 Forecast Results



Chapter 10

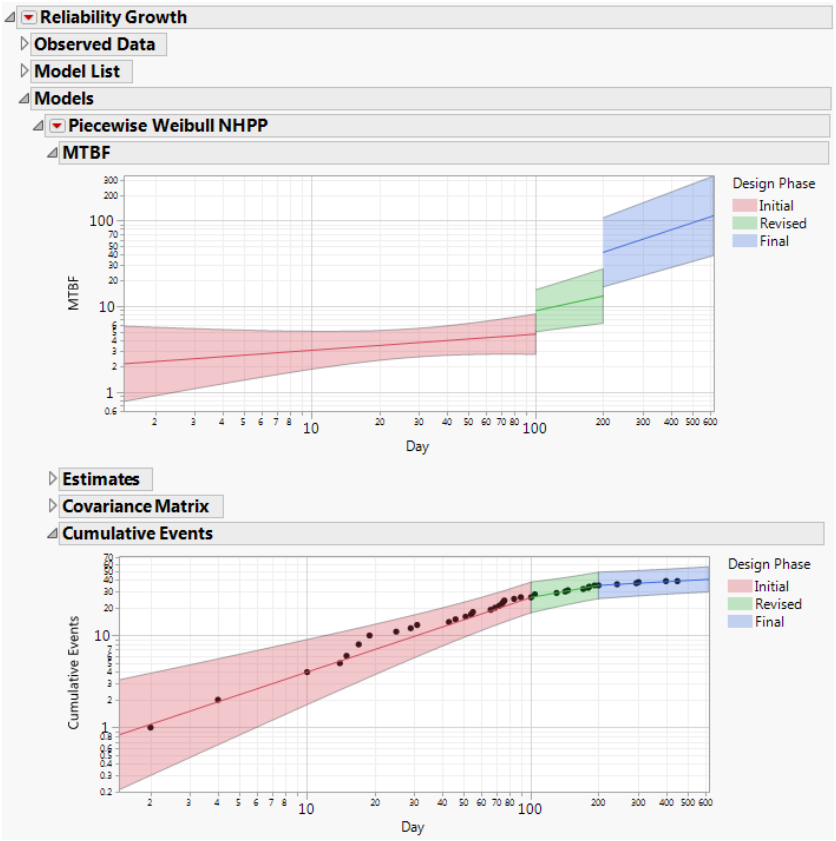
Reliability Growth

Model System Reliability as Changes Are Implemented

The Reliability Growth platform models the change in reliability of a single repairable system over time as improvements are incorporated into its design. A reliability growth testing program attempts to increase the system's mean time between failures (MTBF) by integrating design improvements as failures are discovered.

The Reliability Growth platform fits Crow-AMSAA models. These are non-homogeneous Poisson processes with Weibull intensity functions. Separate models can accommodate various phases of a reliability growth program. The platform also fits models for multiple systems.

Figure 10.1 Example of Plots for a Three-Phase Reliability Growth Model



Contents

Overview of the Reliability Growth Platform 251
Example Using the Reliability Growth Platform 252
Launch the Reliability Growth Platform 256
Launch Window Roles 257
Data Table Structure 258
The Reliability Growth Report 261
Model Descriptions and Feasibilities 261
Observed Data Report 261
Reliability Growth Platform Options 264
Model Reports 265
Crow AMSAA 266
Crow AMSAA with Modified MLE 272
Fixed Parameter Crow AMSAA 273
Piecewise Weibull NHPP 274
Reinitialized Weibull NHPP 278
Piecewise Weibull NHPP Change Point Detection 281
Additional Examples of the Reliability Growth Platform 282
Piecewise NHPP Weibull Model Fitting with Interval-Censored Data 282
Piecewise Weibull NHPP Change Point Detection with Time in Dates Format 285
Statistical Details for the Reliability Growth Platform 287
Statistical Details for the Crow-AMSAA Report 287
Statistical Details for the Piecewise Weibull NHPP Change Point Detection Report 288

Overview of the Reliability Growth Platform

The Reliability Growth platform fits Crow-AMSAA models, described in MIL-HDBK-189 (1981). A Crow-AMSAA model is a non-homogeneous Poisson process (NHPP) model with Weibull intensity; it is also known as a power law process. Such a model allows the failure intensity to vary over time. The failure intensity is defined by the two parameters beta and lambda.

For single-prototype data, the platform fits four classes of models and performs automatic change-point detection. The following reports are available:

- Simple Crow-AMSAA model, where both parameters are estimated using maximum likelihood.
- Crow-AMSAA with Modified MLE, where the maximum likelihood estimate for beta is corrected for bias.
- Fixed Parameter Crow-AMSAA model, where the user is allowed to fix either or both parameters.
- Piecewise Weibull NHPP model, where the parameters are estimated for each test phase, taking failure history from previous phases into account.
- Reinitialized Weibull NHPP model, where both parameters are estimated for each test phase in a manner that ignores failure history from previous phases.
- Automatic estimation of a change-point and the associated piecewise Weibull NHPP model, for reliability growth situations where different failure intensities can define two distinct test phases.

For multiple-prototype data, the platform fits the following classes of models:

- Piecewise Weibull NHPP model, where each system in a multi-phase study follows the same piecewise Weibull NHPP model. The differences among the systems are assumed to be due to the randomness of individual realizations of the same model. This model contains one beta parameter for each phase and one lambda parameter.
- Piecewise Weibull NHPP with Different Intercepts model, where each system in a multi-phase study follows a separate piecewise Weibull NHPP model. This model contains one beta parameter for each phase and one lambda parameter for each system.
- Distinct Phase Weibull NHPP model, where each system in a multi-phase study follows the same Crow-AMSAA model in each phase. This model contains one beta parameter and one lambda parameter for each phase.
- Distinct Weibull NHPP model, where each system in a multi-phase study follows a separate Crow-AMSAA model in each phase. This model contains one beta parameter and one lambda parameter for each combination of system and phase in the study.

- Distinct System Weibull NHPP model, where each system in the study follows a separate Crow-AMSAA model with different parameters.
- Identical System Weibull NHPP model, where each system in the study follows a single Crow-AMSAA model. The differences among the systems are assumed to be due to the randomness of individual realizations of the same model.

Interactive profilers enable you to explore changes in MTBF, failure intensity, and cumulative failures over time. When you suspect a change in intensity over the testing period, you can use the change-point detection option to estimate a change-point and its corresponding model.

Example Using the Reliability Growth Platform

Suppose that you are testing a prototype for a new class of turbine engines. The testing program has been ongoing for over a year and has been through three phases.

The data are given in the TurbineEngineDesign1.jmp sample data table, found in the Reliability subfolder. For each failure that occurred, the number of days since test initiation was recorded in the column Day. The number of failures on a given day, or equivalently, the number of required fixes, was recorded in the column Fixes.

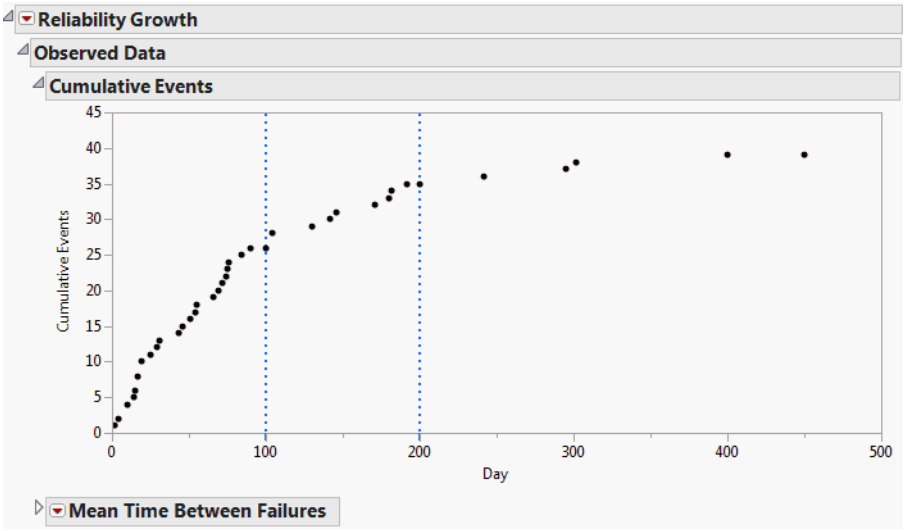
The first 100-day phase of the program was considered the initial testing phase. Failures were addressed with aggressive design changes, resulting in a substantially revised design. This was followed by another 100-day phase, during which failures in the revised design were addressed with design changes to subsystems. The third and final testing phase ran for 250 days. During this final phase, failures were addressed with local design changes.

Each phase of the testing was terminated based on the designated number of days, so that the phases are time terminated. Specifically, a given phase is time terminated at the start time of the next phase. However, the failure times are exact (not censored).

1. Select **Help > Sample Data Library** and open Reliability/TurbineEngineDesign1.jmp.
2. Select **Analyze > Reliability and Survival > Reliability Growth**.
3. On the **Time to Event Format** tab, select Day and click **Time to Event**.
4. Select Fixes and click **Event Count**.
5. Select Design Phase and click **Phase**.
6. Click **OK**.

   The Reliability Growth report appears. The Cumulative Events plot shows the cumulative number of failures by day. Vertical dashed blue lines show the two transitions between the three phases.

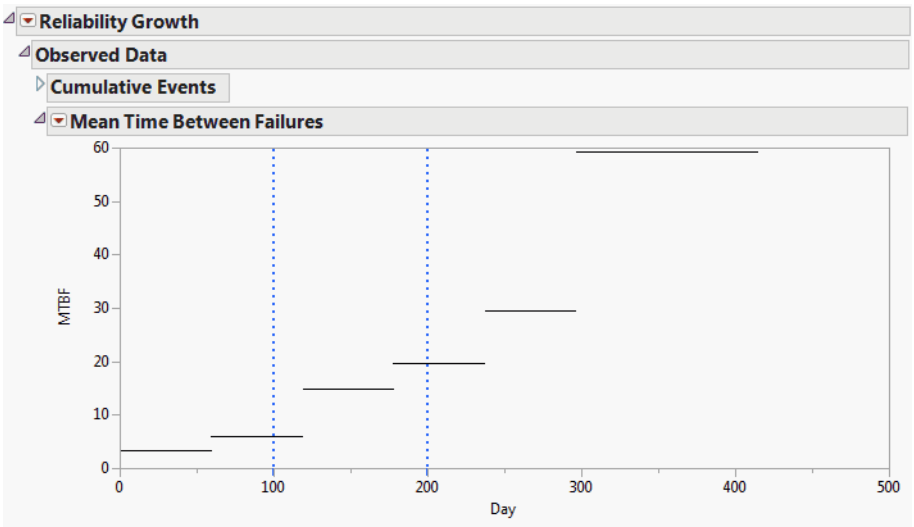
Figure 10.2 Observed Data Report



7. Click the **Mean Time Between Failures** disclosure icon.

   This provides a plot with horizontal lines at the mean times between failures computed over intervals of a predetermined size. An option in the red triangle menu enables you to specify the interval size.

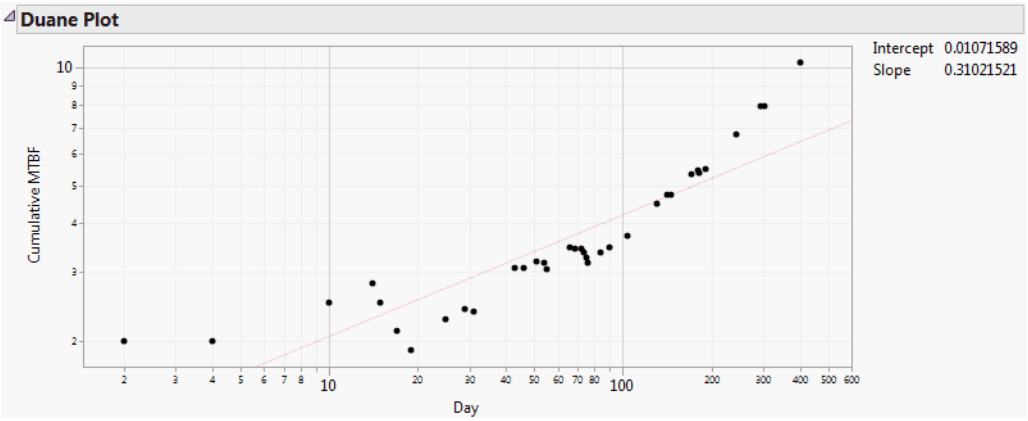
Figure 10.3 Mean Time between Failures Plot



8. Click the **Duane Plot** disclosure icon.

This provides a plot that displays the Cumulative MTBF estimates on the vertical axis versus the time to event variable on the horizontal axis. If the data follow the Crow-AMSAA model, the points should fall along a line when plotted on log-log paper.

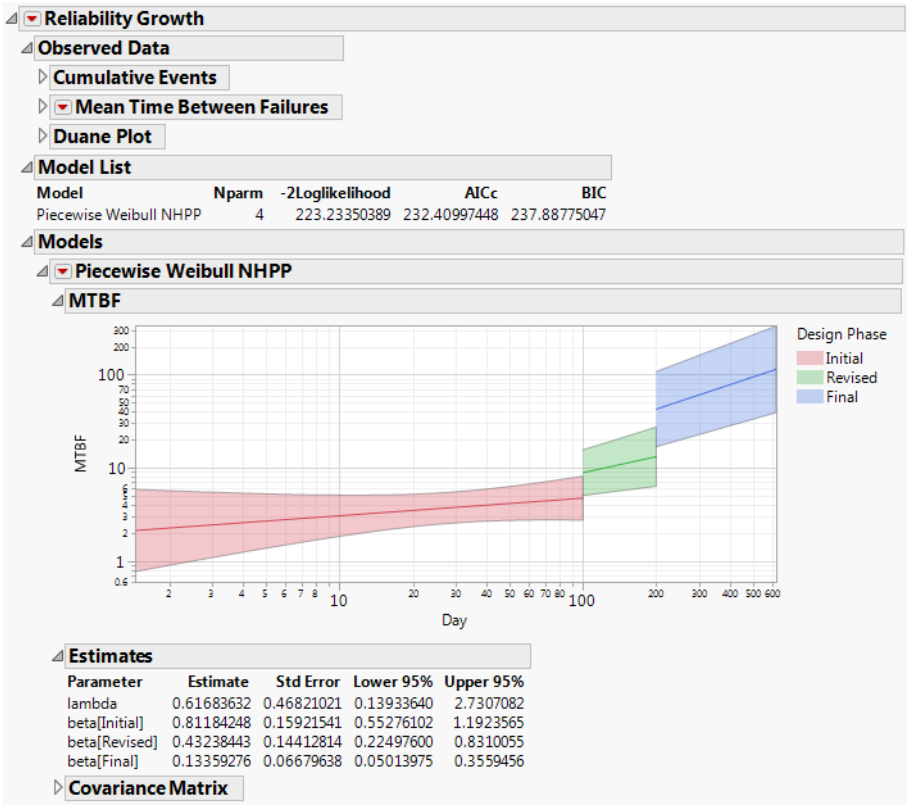
Figure 10.4 Duane Plot



9. Click the Reliability Growth red triangle and select **Fit Model > Piecewise Weibull NHPP**.

   This fits Weibull NHPP models to the three phases of the testing program, treating these phases as multiple stages of a single reliability growth program (Figure 10.5). Options in the Piecewise Weibull NHPP red triangle menu provide various other plots and reports.

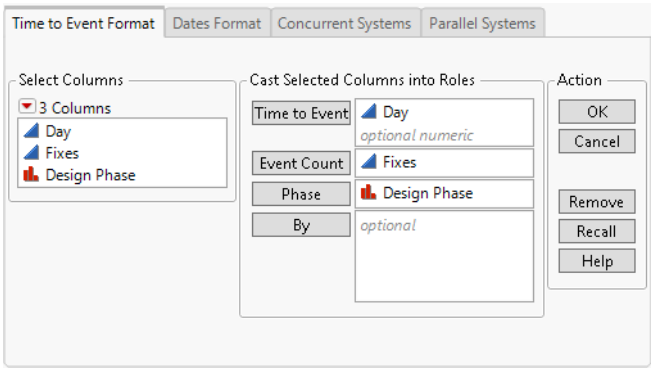
Figure 10.5 Piecewise Weibull NHPP Report



Launch the Reliability Growth Platform

Launch the Reliability Growth platform by selecting **Analyze > Reliability and Survival > Reliability Growth**. The launch window, using data from TurbineEngineDesign1.jmp, is shown in Figure 10.6.

Figure 10.6 Reliability Growth Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

The launch window includes a tab for each of four data formats: Time to Event Format, Dates Format, Concurrent Systems, and Parallel Systems.

- The Dates Format assumes that time is recorded in a date/time format, indicating an absolute date or time. On the Dates Format tab, the time column or columns are given the Timestamp role.
- The other three data formats assume that time is recorded as the number of time units; for example, days or hours, since initial start-up of the system. The test start time is assumed to be time zero. On the Time to Event Format tab, the time column or columns are given the Time to Event role.

Rows with missing data in the columns Time to Event, Timestamp, Event Count, Phase, or System ID are not included in the analysis.

Note: All data formats for the Reliability Growth platform require that times or time intervals be specified in non-decreasing order.

Launch Window Roles

The roles that are available for variables in the Reliability Growth launch window are determined by the specified data format tab. The roles are explained in this section.

Time to Event

Time to Event is the number of time units that elapse between the start of the test and the occurrence of an event (a failure or test termination). The test start time is assumed to be time zero. Note that the Time to Event role is available only on the Time to Event Format tab.

Two conventions are allowed (See "Exact Failure Times versus Interval Censoring".):

- A single column can be entered. In this case, it is assumed that the column gives the exact elapsed times at which events occurred.
- Two columns can be entered, giving interval start and end times. If an interval's start and end times differ, it is assumed that the corresponding events given in the Event Count column occurred at some unknown time within that interval. We say that the data are interval-censored. If the interval start and end times are identical, it is assumed that the corresponding events occurred at that specific time point, so that the times are exact (not censored).

The platform requires that the time columns be sorted in non-decreasing order. When two columns giving interval start and end times are provided, these intervals must not overlap (except at their endpoints). Intervals with zero event counts that fall strictly within a phase can be omitted, as they do not affect the likelihood function.

Timestamp

Timestamp is an absolute time (for example, a date). As with Time to Event, Timestamp allows times to be entered using either a single column or two columns. Note that the Timestamp role is available only on the Dates Format tab.

For times entered as Timestamp, the first row of the table is considered to give the test start time:

- When a single column is entered, the timestamp corresponding to the test start, with an event count of 0, should appear in the first row.
- When two columns, giving time intervals, are entered, the first entry in the first column should be the test start timestamp. See "Phase".

Other details are analogous to those described for Time to Event Format in the section "Time to Event". See also "Exact Failure Times versus Interval Censoring".

Event Count

Event Count is the number of events, usually failures addressed by corrective actions (fixes), occurring at the specified time or within the specified time interval. If no column is entered as Event Count, it is assumed that the Event Count for each row is one.

System ID

System ID identifies the prototype for an observation for multiple-prototype data. The System ID role is available only in the Concurrent Systems and Parallel Systems data formats. It is a required role for both data formats.

Phase

Reliability growth programs often involve several periods, or phases, of active testing. These testing phases can be specified in the optional Phase column. For the Time to Event Format and Dates Format data formats, the Phase variable can be of any data or modeling type. For the Parallel Systems data format, the data type of the Phase variable must be Numeric. For more information about structuring multi-phase data, see "Test Phases". For an example, see "Piecewise NHPP Weibull Model Fitting with Interval-Censored Data".

By

This produces a separate analysis for each value that appears in the column.

Data Table Structure

The Time to Event Format and the Dates Format enable you to enter either a single column *or* two columns as Time to Event or Timestamp, respectively. The Concurrent Systems and the Parallel Systems enable you to enter one column corresponding to each level of the System ID variable. There are examples of multiple-prototype data tables in the Reliability folder of the Sample Data folder: Concurrent Systems.jmp for Concurrent Systems and four tables with the prefix Parallel Systems for Parallel Systems.

This section describes how to use these two approaches to specify the testing structure.

Exact Failure Times versus Interval Censoring

In some testing situations, the system being tested is checked periodically for failures. In this case, failures are known to have occurred within time intervals, but the precise time of a failure is not known. We say that the failure times are *interval-censored*.

The Reliability Growth platform accommodates both exact, non-censored failure-time data, and interval-censored data. When a single column is entered as Time to Event or Timestamp, the times are considered exact times (not censored).

When two columns are entered, the platform views these as defining the start and end points of time intervals. If an interval's start and end times differ, then the times for failures occurring within that interval are considered to be interval-censored. If the end points are identical, then the times for the corresponding failures are assumed to be exact and equal to that common time value. So, you can represent both exact and interval-censored failure times by using two time columns.

In particular, *exact failures times* can be represented in one of two ways: As times given by a single time column, or as intervals with identical endpoints, given by two time columns.

Model-fitting in the Reliability Growth platform relies on the likelihood function. The likelihood function takes into account whether interval-censoring is present or not. So, mixing interval-censored with exact failure times is permitted.

Failure and Time Termination

A test plan can call for test termination once a specific number of failures has been achieved or once a certain span of time has elapsed. For example, a test plan might terminate testing after 50 failures occur. Another plan might terminate testing after a six-month period.

If testing terminates based on a specified number of failures, we say that the test is *failure terminated*. If testing is terminated based on a specified time interval, we say that the test is *time terminated*. The likelihood functions used in the Reliability Growth platform reflect whether the test phases are failure or time terminated.

Test Phases

Reliability growth testing often involves several *phases* of testing. For example, the system being developed or the testing program might experience substantial changes at specific time points. The data table conveys the start time for each phase and whether each phase is failure or time terminated, as described below.

Single Test Phase

When there is a single test phase, the platform infers whether the test is failure or time terminated from the time and event count entries in the last row of the data table.

- If the last row contains an exact failure time with a nonzero event count, the test is considered failure terminated.
- If the last row contains an exact failure time with a zero event count, the test is considered time terminated.

- If the last row contains an interval with nonzero width, the test is considered time terminated with termination time equal to the right endpoint of that interval.

Note: To indicate that a test has been time terminated, be sure to include a last row in your data table showing the test termination time. If you are entering a single column as Time to Event or Timestamp, the last row should show a zero event count. If you are entering two columns as Time to Event or Timestamp, the right endpoint of the last interval should be the test-termination time. In this case, if there were no failures during the last interval, you should enter a zero event count.

Multiple Test Phases

When using Time to Event Format, the start time for any phase other than the first should be included in the time column(s). When using Dates Format, the start times for all phases should be included in the time column(s). If no events occurred at a phase start time, the corresponding entry in the Event Count column should be zero. For times given in two columns, it might be necessary to reflect the phase start time using an interval with identical endpoints and an event count of zero.

In a multi-phase testing situation, the platform infers whether each phase, other than the last, is failure or time terminated from the entries in the last row preceding a phase change. Suppose that Phase A ends and that Phase B begins at time t_B. In this case, the first row corresponding to Phase B contains an entry for time t_B.

- If the failure time for the last failure in Phase A is exact and if that time differs from t_B, then Phase A is considered to be time terminated. The termination time is equal to t_B.
- If the failure time for the last failure in Phase A is exact and is equal to t_B, then Phase A is considered to be failure terminated.
- If the last failure in Phase A is interval-censored, then Phase A is considered to be time terminated with termination time equal to t_B.

The platform infers whether the final phase is failure or time terminated from the entry in the last row of the data table.

- If the last row contains an exact failure time with a nonzero event count, the test is considered failure terminated.
- If the last row contains an exact failure time with a zero event count, or an interval with nonzero width, the test is considered time terminated. In the case of an interval, the termination time is taken as the right endpoint.

The Reliability Growth Report

The Observed Data report appears by default. If the Parallel Systems data format is specified in the launch window, the Model Descriptions and Feasibilities report also appears by default.

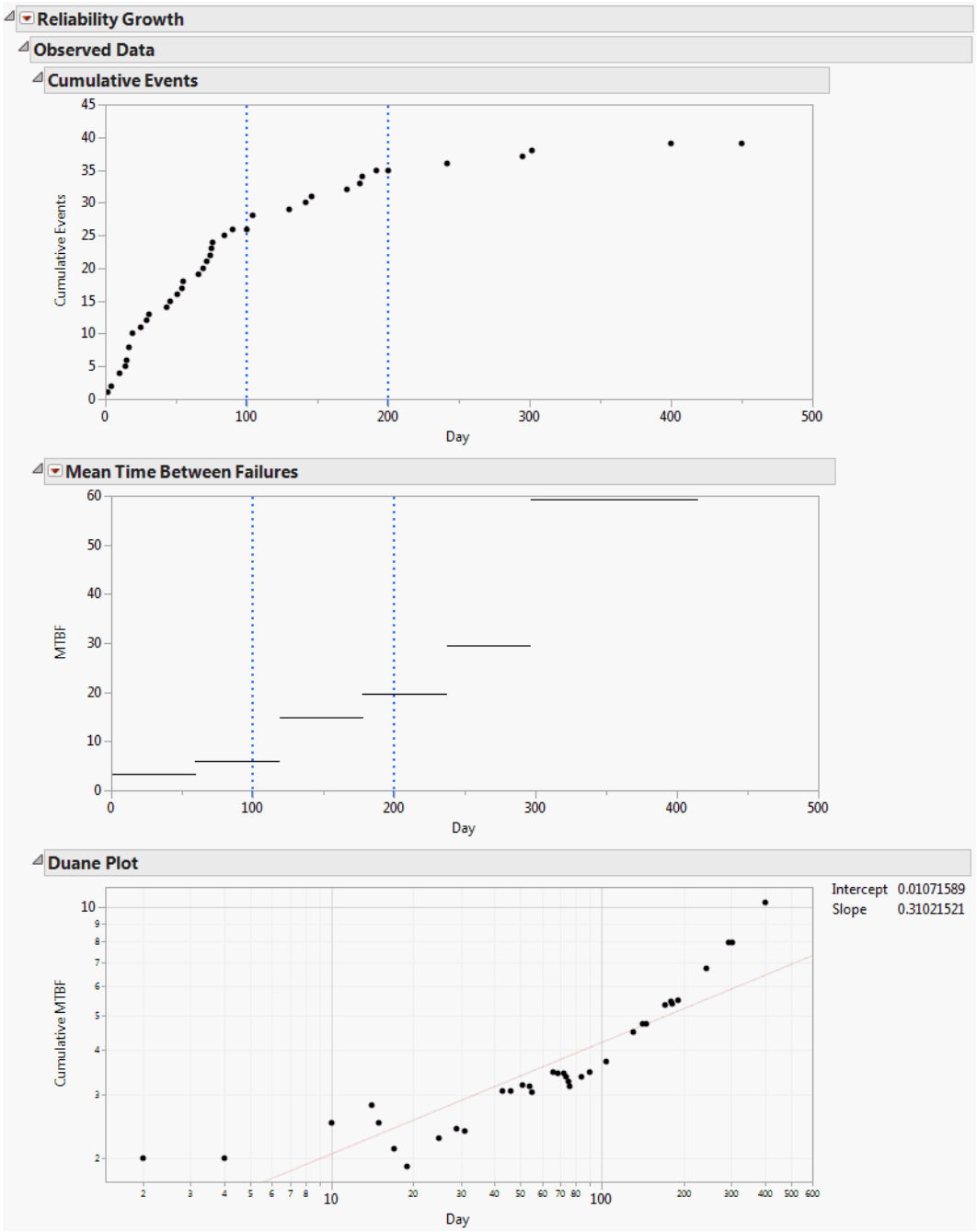
Model Descriptions and Feasibilities

The Model Description and Feasibilities report lists the name and descriptions of the parallel systems models. It also contains a column that shows if a model is available for the current data. If a model is listed as not available, the reason is reported in the right-most column of this table.

Observed Data Report

The Observed Data report contains the Cumulative Events plot, the Mean Time Between Failures plot, and the Duane plot. These are shown in Figure 10.7, where we have opened the Mean Time Between Failures and Duane Plot reports. To produce this report, follow the instructions in "Example Using the Reliability Growth Platform".

Figure 10.7 Observed Data Report



Cumulative Events Plot

This plot shows how events are accumulating over time. The vertical coordinate for each point on the Cumulative Events plot equals the total number of events that have occurred by the time given by the point's horizontal coordinate.

Whenever a model is fit, that model is represented on the Cumulative Events plot. Specifically, the cumulative events estimated by the model are shown by a curve, and 95% confidence intervals are shown by a solid band. Check boxes to the right of the plot enable you to control which models are displayed.

For the Parallel Systems data format, the Cumulative Events plot has a separate panel in the plot for each level of the System ID variable. It also contains a panel at the bottom of the plot that overlays the cumulative events for all levels of the System ID variable. The levels of the System ID variable are designated by colors.

Mean Time Between Failures Plot

The Mean Time Between Failures plot shows mean times between failures averaged over small time intervals of equal length. These are not tied to the Phases. The default number of equal length intervals is based on the number of rows.

Mean Time Between Failures Plot Options

Click the Mean Time Between Failures red triangle and select Options to open a window that enables you to specify intervals over which to average.

Two types of averaging are offered:

- Equal Interval Average MTBF (Mean Time Between Failures) enables you to specify a common interval size.
- Customized Average MTBF enables you to specify cut-off points for time intervals.
  - Double-click within a table cell to change its value.
  - Right-click in the table to open a menu that enables you to add and remove rows.

Duane Plot

The Duane Plot displays Cumulative MTBF estimates plotted against the Time to Event variable, with both axes on a log10 scale. If the data follow the Crow-AMSAA model, the points should follow a line when plotted on log-log paper.

Note: The Duane Plot is available only if failure times are exact and Time to Event Format is used. The plot is not available for interval-censored data or data entered in Dates Format.

The line displayed on the plot is the least squares regression line for the regression of log10 of Cumulative MTBF on log10 of the Time to Event variable.

Note: The Duane Plot does not reflect Phase variables. Rows where the Time to Event variable defines Phase changes are ignored in constructing the plot and fitting the regression line.

Intercept and Slope

Intercept and Slope values are displayed to the right of the plot.

- The Intercept value is given, for historical reasons, as the intercept for a fit in the *natural* logarithmic scale. Specifically, the natural logarithm of Cumulative MTBF is regressed on the natural logarithm of Time to Event. The value of Intercept is the value predicted by this regression equation at log(1) = 0, where *log* is the natural logarithm. To obtain the intercept for a fit in terms of base 10 logarithms, divide the Intercept value by log(10). See Tobias and Trindade (2012, ch. 13).
- The Slope value is the slope for a fit in either the natural or the logarithmic scale. This follows from the properties of logarithms.

Reliability Growth Platform Options

The Reliability Growth red triangle menu contains the following options:

Fit Model If the Time to Event Format, Dates Format, or Concurrent Systems data formats are specified in the launch window, this menu contains options to fit various Non-Homogeneous Poisson Process (NHPP) models. These options are described in detail below. Depending on the choices made in the launch window, the possible options are:

- "Crow AMSAA"
- "Crow AMSAA with Modified MLE"
- "Fixed Parameter Crow AMSAA"
- "Piecewise Weibull NHPP"
- "Reinitialized Weibull NHPP"
- "Piecewise Weibull NHPP Change Point Detection"

Fit Parallel System Model If the Parallel Systems data format is specified in the launch window, this menu contains options to fit various models for multiple-prototype data. The possible options in this menu are dependent on choices made in the launch window.

See *Using JMP* for more information about the following options:

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Model List

Once a model is fit, the Model List report appears. This report provides various statistical measures that describe the fit of the model. As additional models are fit, they are added to the Model List, which provides a convenient summary for model comparison. The models are sorted in ascending order based on AICc. The Model List report contains the following statistics:

Nparm The number of parameters in the model.

-2Loglikelihood The likelihood function is a measure of how probable the observed data are, given the estimated model parameters. In a general sense, the higher the likelihood, the better the model fit. It follows that smaller values of twice the negative of the log-likelihood (-2Loglikelihood) indicate better model fits.

AICc The Corrected Akaike's Information Criterion.

BIC The Bayesian Information Criterion.

See *Fitting Linear Models* for more information about -2Loglikelihood, AICc, and BIC.

Model Reports

This section describes the reports generated when the available models are fit as well as the options available in those model reports.

- "Crow AMSAA"
- "Crow AMSAA with Modified MLE"
- "Fixed Parameter Crow AMSAA"
- "Piecewise Weibull NHPP"
- "Reinitialized Weibull NHPP"
- "Piecewise Weibull NHPP Change Point Detection"

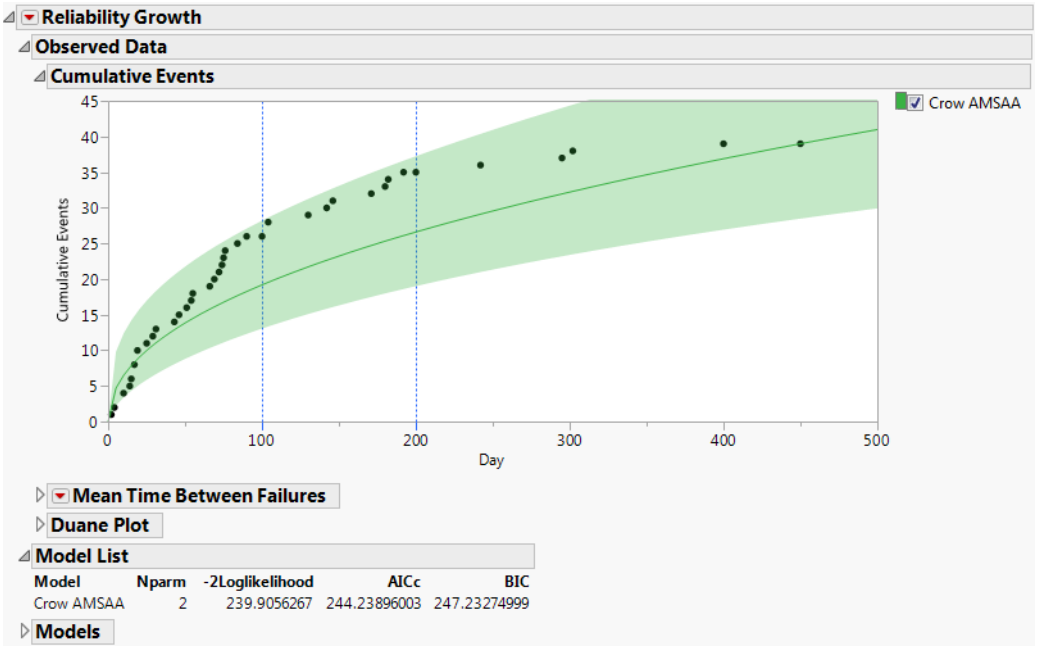
Crow AMSAA

This option fits a Crow-AMSAA model (MIL-HDBK-189 1981). A Crow-AMSAA model is a nonhomogeneous Poisson process with failure intensity as a function of time t given by $\rho(t) = \lambda\beta t^{\beta-1}$. Here, λ is a scale parameter and β is a growth parameter. This function is also called a Weibull intensity, and the process itself is also called a power law process (Rigdon and Basu 2000; Meeker and Escobar 1998). Note that the Recurrence platform fits the Power Nonhomogeneous Poisson Process. The Power Nonhomogeneous Poisson Process is equivalent to the Crow-AMSAA model, although it uses a different parameterization. See "Fit Model".

The intensity function is a concept applied to repairable systems. Its value at time t is the limiting value of the probability of a failure in a small interval around t, divided by the length of this interval; the limit is taken as the interval length goes to zero. You can think of the intensity function as measuring the likelihood of the system failing at a given time. If $\beta < 1$, the system is improving over time. If $\beta > 1$, the system is deteriorating over time. If $\beta = 1$, the rate of occurrence of failures is constant.

When the Crow AMSAA option is selected, the Cumulative Events plot updates to show the cumulative events curve estimated by the model. For each time point, the shaded band around this curve defines a 95% confidence interval for the true cumulative number of events at that time. The Model List report also updates. Figure 10.8 shows the Observed Data report for the data in TurbineEngineDesign1.jmp.

Figure 10.8 Crow AMSAA Cumulative Events Plot and Model List Report



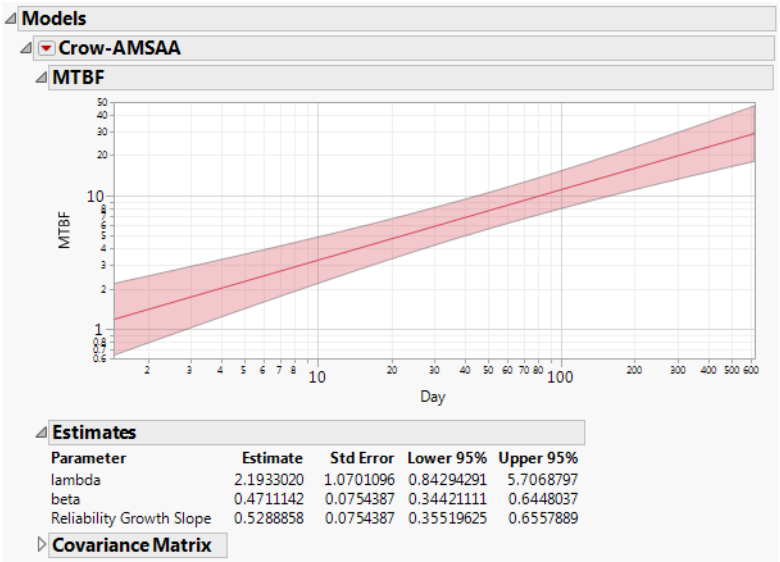
Crow-AMSAA Report

A Crow-AMSAA report opens within the Models report. If Time to Event Format is used, the Crow-AMSAA report shows an MTBF plot with both axes scaled logarithmically. See "MTBF Plot".

MTBF Plot

The mean time between failures (MTBF) plot is displayed by default (Figure 10.9). For each time point, the shaded band around the MTBF plot defines a 95% confidence interval for the true MTBF at time *t*. The plot is shown with both axes logarithmically scaled. With this scaling, the MTBF plot is linear.

Figure 10.9 MTBF Plot



To see why the MTBF plot is linear when logarithmic scaling is used, consider the following. The mean time between failures is the reciprocal of the intensity function. For the Weibull intensity function, the MTBF is $1/(\lambda\beta t^{\beta-1})$, where t represents the time since testing initiation. It follows that the logarithm of the MTBF is a linear function of $\log(t)$, with slope $1 - \beta$. The estimated MTBF is defined by replacing the parameters λ and β by their estimates. So the log of the estimated MTBF is a linear function of $\log(t)$.

Estimates

Maximum likelihood estimates for lambda (λ), beta (β), and the Reliability Growth Slope ($1 - \beta$), appear in the Estimates report below the plot (Figure 10.9). Standard errors and 95% confidence intervals for λ, β, and $1 - \beta$ are given. For more information about the calculations, see "Parameter Estimates for Crow-AMSAA Models".

Covariance Matrix

Estimated covariance matrix for the estimates of the parameters of the fitted model. This report is closed by default.

Crow-AMSAA Options

This section describes the options that are available in the Crow-AMSAA red triangle menu when a Crow AMSAA model is fit.

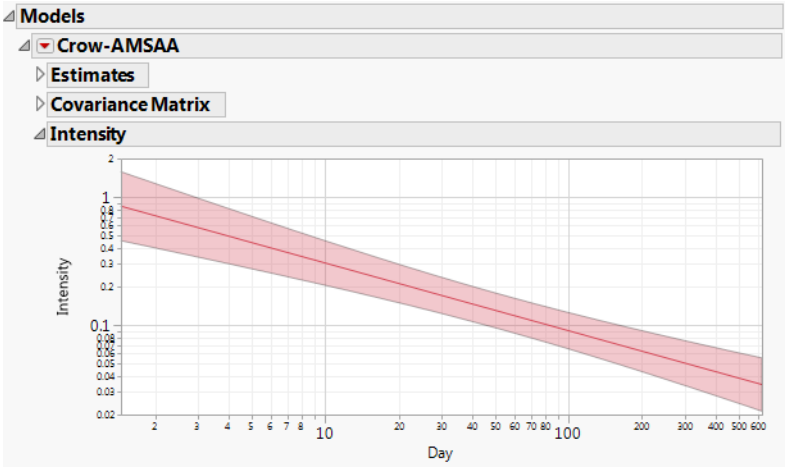
Show MTBF Plot

This option shows or hides the MTBF Plot. See "MTBF Plot".

Show Intensity Plot

This plot shows the estimated intensity function. The Weibull intensity function is given by $\rho(t) = \lambda\beta t^{\beta-1}$, so it follows that log(Intensity) is a linear function of log(t). Both axes are scaled logarithmically.

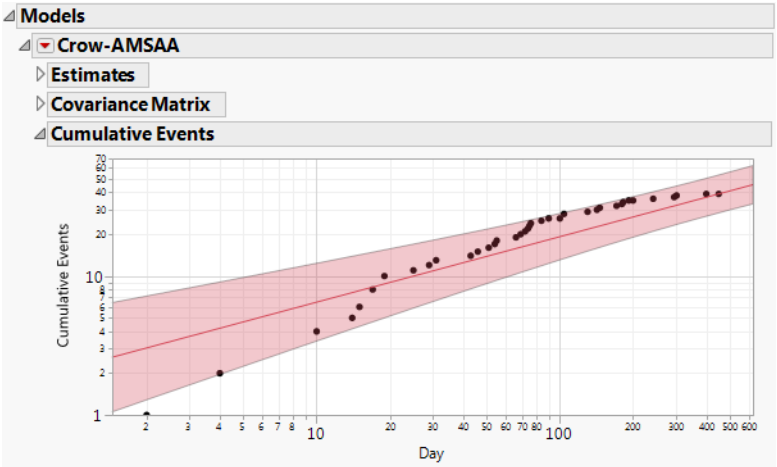
Figure 10.10 Intensity Plot



Show Cumulative Events Plot

This plot shows the estimated cumulative number of events. The observed cumulative numbers of events are also displayed on this plot. Both axes are scaled logarithmically.

Figure 10.11 Cumulative Events Plot



For the Crow-AMSAA model, the cumulative number of events at time t is given by λt^{β}. It follows that the logarithm of the cumulative number of events is a linear function of $\log(t)$. So, the plot of the estimated Cumulative Events is linear when plotted against logarithmically scaled axes.

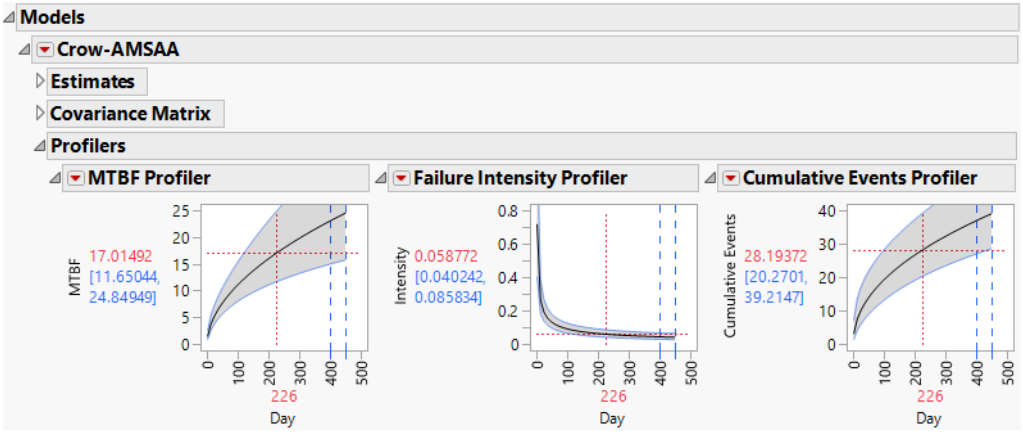
Show Profilers

Three profilers are displayed, showing estimated MTBF, Failure Intensity, and Cumulative Events. These profilers do not use logarithmic scaling. By dragging the red vertical dashed line in any profiler, you can explore model estimates at various time points; the value of the selected time point is shown in red beneath the plot. Also, you can set the time axis to a specific value by pressing Ctrl while you click in the plot. A blue vertical dashed line denotes the time point of the last observed failure.

The profilers also display 95% confidence bands for the estimated quantities. For the specified time setting, the estimated quantity (in red) and 95% confidence limits (in black) are shown to the left of the profiler. See "Profilers".

Note that you can link these profilers by selecting **Factor Settings > Link Profilers** from any of the profiler red triangle menus. For more information about the use and interpretation of profilers, see *Fitting Linear Models*. See also *Profilers*.

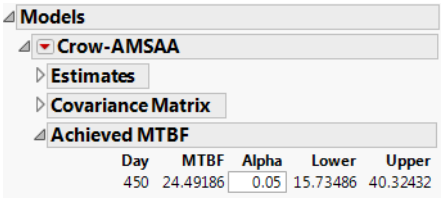
Figure 10.12 Profilers



Achieved MTBF

A confidence interval for the MTBF at the point when testing concludes is often of interest. For uncensored failure time data, this report gives an estimate of the Achieved MTBF and a 95% confidence interval for the Achieved MTBF. You can specify a 100*(1-α)% confidence interval by entering a value for Alpha. The report is shown in Figure 10.13. For censored data, only the estimated MTBF at test termination is reported.

Figure 10.13 Achieved MTBF Report



There are infinitely many possible failure-time sequences from an NHPP; the observed data represent only one of these. Suppose that the test is failure terminated at the n^{th} failure. The confidence interval computed in the Achieved MTBF report takes into account the fact that the n failure times are random. If the test is time terminated, then the number of failures as well as their failure times are random. Because of this, the confidence interval for the Achieved MTBF differs from the confidence interval provided by the MTBF Profiler at the last observed failure time. See Crow (1982) and Lee and Lee (1978).

When the test is failure terminated, the confidence interval for the Achieved MTBF is exact. However, when the test is time terminated, an exact interval cannot be obtained. In this case, the limits are conservative in the sense that the interval contains the Achieved MTBF with probability at least 1-α.

Goodness of Fit

The Goodness of Fit report tests the null hypothesis that the data follow a Crow-AMSAA model. Depending on whether one or two time columns are entered, either a Cramér-von Mises (see "Cramér-von Mises Test for Data with Uncensored Failure Times") or a chi-squared test (see "Chi-Squared Goodness of Fit Test for Interval-Censored Failure Times") is performed.

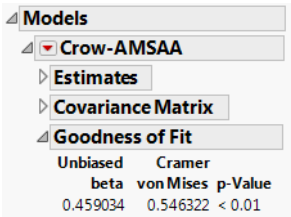
Cramér-von Mises Test for Data with Uncensored Failure Times

When the data are entered in the launch window as a single Time to Event or Timestamp column, the goodness of fit test is a Cramér-von Mises test. For the Cramér-von Mises test, large values of the test statistic lead to rejection of the null hypothesis and the conclusion that the model does not fit adequately. The test uses an unbiased estimate of beta, given in the report. The value of the test statistic is found below the Cramér-von Mises heading.

The entry below the p-Value heading indicates how unlikely it is for the test statistic to be as large as what is observed if the data come from a Crow-AMSAA model. The platform computes *p*-values up to 0.25. If the test statistic is smaller than the value that corresponds to a *p*-value of 0.25, the report indicates that its *p*-value is >=0.25. For more information about this test, see Crow (1975).

Figure 10.14 shows the goodness-of-fit test for the fit of a Crow-AMSAA model to the data in TurbineEngineDesign1.jmp. The computed test statistic corresponds to a *p*-value that is less than 0.01. We conclude that the Crow-AMSAA model does not provide an adequate fit to the data.

Figure 10.14 Goodness of Fit Report - Cramér-von Mises Test



Chi-Squared Goodness of Fit Test for Interval-Censored Failure Times

When the data are entered in the launch window as two Time to Event or Timestamp columns, a chi-squared goodness of fit test is performed. The chi-squared test is based on comparing observed to expected numbers of failures in the time intervals defined. Large values of the test statistic lead to rejection of the null hypothesis and the conclusion that the model does not fit.

In the Reliability Growth platform, the chi-squared goodness of fit test is intended for interval-censored data where the time intervals specified in the data table cover the entire time period of the test. This means that the start time of an interval is the end time of the preceding interval. In particular, intervals where no failures occurred should be included in the data table. If some intervals are not consecutive, or if some intervals have identical start and end times, the algorithm makes appropriate accommodations. But the resulting test is only approximately correct.

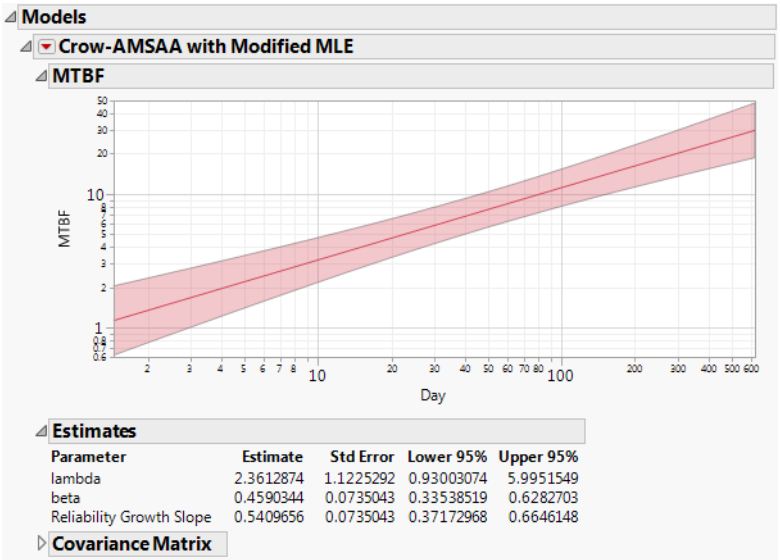
Crow AMSAA with Modified MLE

In the Crow-AMSAA model, the maximum likelihood estimate (MLE) of β is biased. This option fits a Crow AMSAA model where β is adjusted for bias.

Note: This option is available only when the data are entered in the launch window as a single Time to Event or Timestamp column. It is not available for interval-censored data.

Figure 10.15 shows a Crow-AMSAA with Modified MLE fit to the data in TurbineEngineDesign1.jmp.

Figure 10.15 Crow-AMSAA with Modified MLE Report



The formula for the bias-corrected estimate of β depends on whether the test is failure terminated or time terminated. See "Parameter Estimates for Crow-AMSAA with Modified MLE".

When the Crow-AMSAA with Modified MLE option is selected, the Cumulative Events Plot updates to display this model. The Model List also updates. The Crow-AMSAA with Modified MLE report opens to show the MTBF plot, Estimates, and Covariance Matrix for the Crow-AMSAA with Modified MLE fit; this plot is described in the section "MTBF Plot".

In addition to Show MTBF plot, available options are Show Intensity Plot, Show Cumulative Events Plot, Show Profilers, Achieved MTBF, and Goodness of Fit. These reports are described under "Crow AMSAA". For more information about how the modified MLEs are used to construct these reports, see "Parameter Estimates for Crow-AMSAA with Modified MLE". Details about the Goodness of Fit and Achieved MTBF reports specific to the modified MLE option are given below.

Goodness of Fit

Because the Crow-AMSAA with Modified MLE option is available only when the data are entered as a single Time to Event or Timestamp column, the Goodness of Fit test is a Cramér-von Mises test. Because the estimate of β is used in this test is bias-corrected, the test results are identical to those of the Goodness of Fit test for the Crow-AMSAA model.

Achieved MTBF

The achieved MTBF is estimated using the modified MLEs. However, the confidence interval for the achieved MTBF uses the true MLE and is identical to the interval given by the Crow AMSAA model.

Fixed Parameter Crow AMSAA

This option enables you to specify parameter values for a Crow-AMSAA fit. If a Crow-AMSAA report has not been obtained before choosing the Fixed Parameter Crow-AMSAA option, then both a Crow-AMSAA report and a Fixed Parameter Crow-AMSAA report are provided.

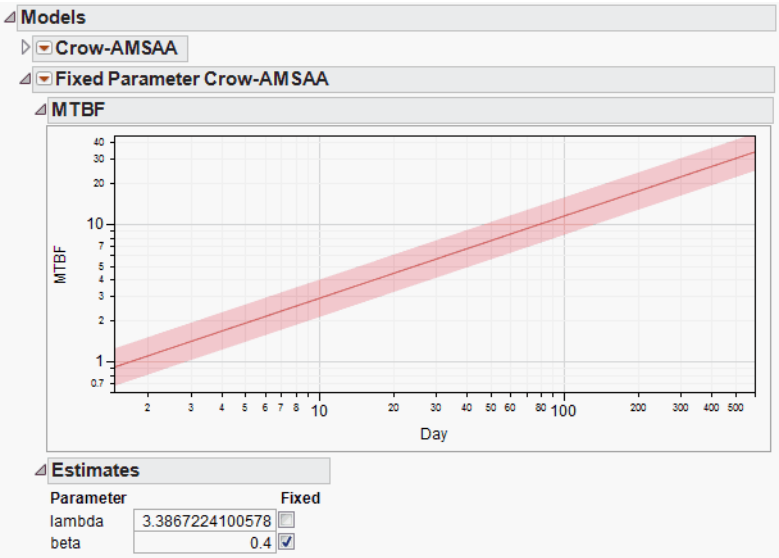
When the Fixed Parameter Crow-AMSAA option is selected, the Cumulative Events Plot updates to display this model. The Model List also updates. The Fixed Parameter Crow-AMSAA report opens to show the MTBF plot for the Crow-AMSAA fit; this plot is described in the section "MTBF Plot".

In addition to Show MTBF plot, available options are Show Intensity Plot, Show Cumulative Events Plot, and Show Profilers. The construction and interpretation of these plots is described under "Crow AMSAA".

Estimates

The initial parameter estimates are the MLEs from the Crow-ASMAA fit. Either parameter can be fixed by checking the box next to the desired parameter and then entering the desired value. The model is re-estimated and the MTBF plot updates to describe this model. Figure 10.16 shows a fixed-parameter Crow-AMSAA fit to the data in TurbineEngineDesign1.jmp, with the value of beta set at 0.4.

Figure 10.16 Fixed Parameter Crow-AMSAA Report

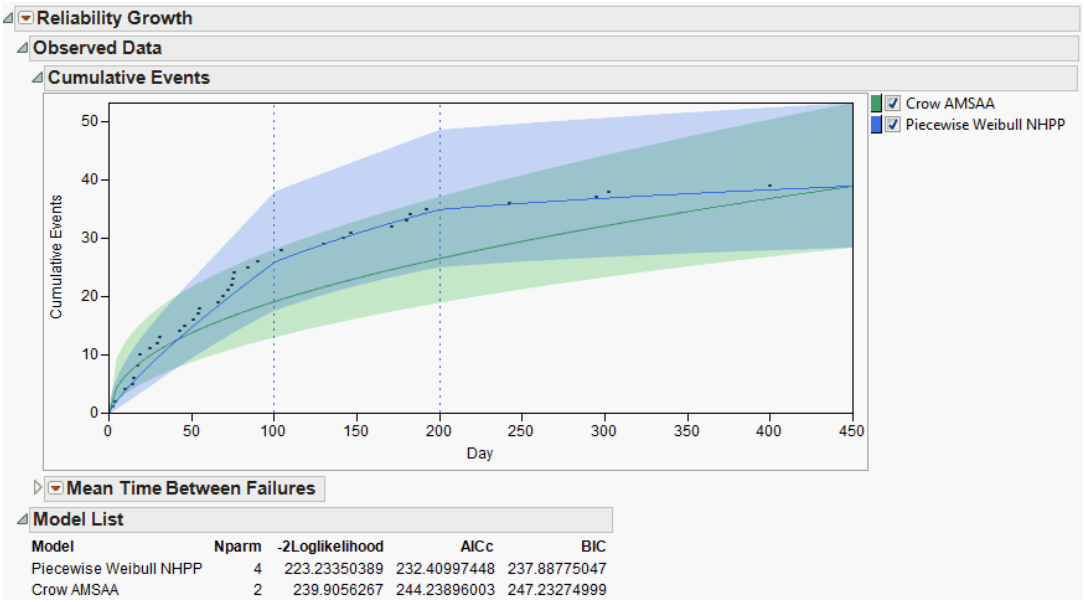


Piecewise Weibull NHPP

The Piecewise Weibull NHPP model can be fit when a Phase column specifying at least two values has been entered in the launch window. Crow-AMSAA models are fit to each of the phases under the constraint that the cumulative number of events at the start of a phase matches that number at the end of the preceding phase. For proper display of phase transition times, the first row for every Phase other than the first must give that phase's start time. See "Multiple Test Phases".

When the report is run, the Cumulative Events plot updates to show the piecewise model. Blue vertical dashed lines show the transition times for each of the phases. The Model List also updates. See Figure 10.17, where both a Crow-AMSAA model and a piecewise Weibull NHPP model have been fit to the data in TurbineEngineDesign1.jmp. Note that both models are compared in the Model List report.

Figure 10.17 Cumulative Events Plot and Model List Report

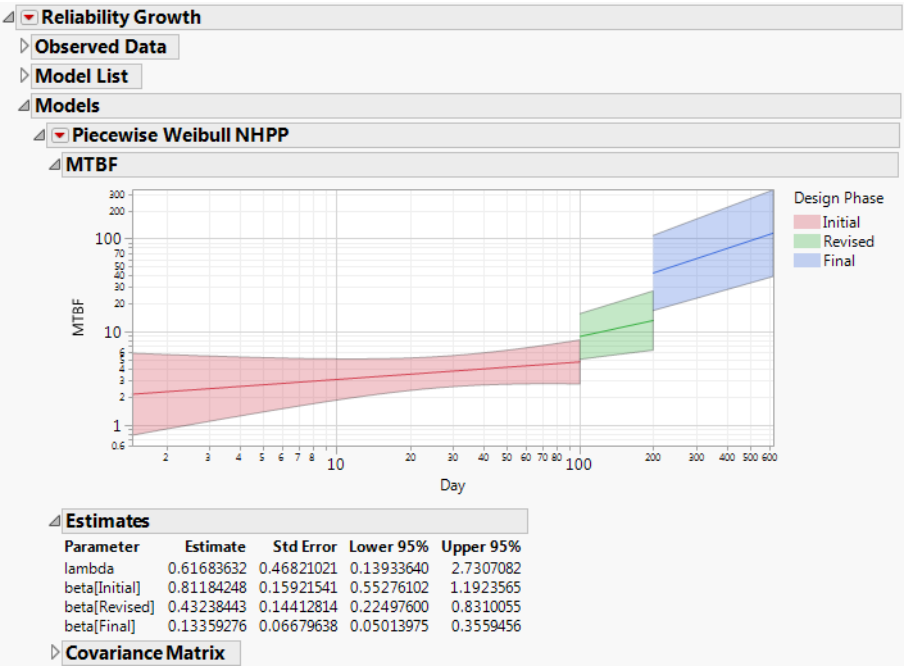


Model	Nparm	-2Loglikelihood	AICc	BIC
Piecewise Weibull NHPP	4	223.23350389	232.40997448	237.88775047
Crow AMSAA	2	239.9056267	244.23896003	247.23274999

Piecewise Weibull NHPP Report

By default, the Piecewise Weibull NHPP report shows the estimated MTBF plot. The phases are differentiated with different colors. The Estimates and Covariance Matrix reports are shown below the plot.

Figure 10.18 Piecewise Weibull NHPP Report



MTBF Plot

The MTBF plot and the Estimates report appear by default when the Piecewise Weibull NHPP option is chosen (Figure 10.18). When Time to Event Format is used, the axes are logarithmically scaled. For more information about the plot, see "MTBF Plot".

Estimates

The Estimates report gives estimates of the model parameters. Note that only the estimate for the value of λ corresponding to the first phase is given. In the piecewise model, the cumulative events at the end of one phase must match the number at the beginning of the subsequent phase. Because of these constraints, the estimate of λ for the first phase and the estimates of the βs determine the remaining λs.

The method used to calculate the estimates, their standard errors, and the confidence limits is similar to that used for the simple Crow-AMSAA model. See "Parameter Estimates for Crow-AMSAA Models". The likelihood function reflects the additional parameters and the constraints on the cumulative numbers of events.

Covariance Matrix

Estimated covariance matrix for the estimates of the parameters of the fitted model. This report is closed by default.

Piecewise Weibull NHPP Options

This section describes the options available in the Piecewise Weibull NHPP red triangle menu.

Show MTBF Plot

This option shows or hides the MTBF plot. See "MTBF Plot".

Show Intensity Plot

The Intensity plot shows the estimated intensity function and confidence bounds over the design phases. The intensity function is generally discontinuous at a phase transition. Color coding facilitates differentiation of phases. If Time to Event Format is used, the axes are logarithmically scaled. See "Show Intensity Plot".

Show Cumulative Events Plot

The Cumulative Events plot shows the estimated cumulative number of events, along with confidence bounds, over the design phases. The model requires that the cumulative events at the end of one phase match the number at the beginning of the subsequent phase. Color coding facilitates differentiation of phases. If Time to Event Format is used, the axes are logarithmically scaled. See "Show Cumulative Events Plot".

Show Profilers

Three profilers are displayed, showing estimated MTBF, Failure Intensity, and Cumulative Events. These profilers do not use logarithmic scaling. For more information about interpreting and using these profilers, see the section "Show Profilers".

It is important to note that, due to the default resolution of the profiler plot, discontinuities are not displayed clearly in the MTBF or Failure Intensity Profilers. In the neighborhood of a phase transition, the profiler trace shows a nearly vertical, but slightly sloped, line; this line represents a discontinuity (Figure 10.19). Such a line at a phase transition should not be used for estimation. Obtain a higher-resolution display to make these lines appear more vertical:

1. Press Ctrl while clicking in the profiler plot.
2. Enter a larger value for Number of Plotted Points in the window. (See Figure 10.20, where we have specified 500 as the Number of Plotted Points.)

Figure 10.19 Profilers

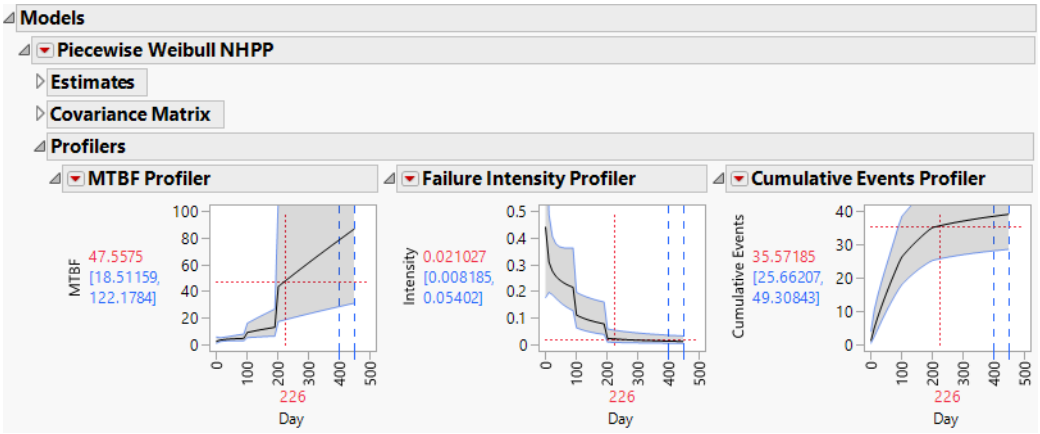
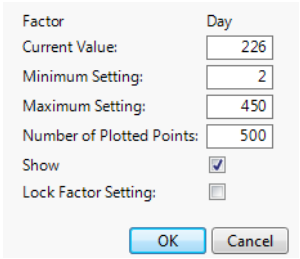


Figure 10.20 Factor Settings Window



Reinitialized Weibull NHPP

This option fits an independent growth model to the data from each test phase. Fitting models in this fashion can be useful when the factors influencing the growth rate, either in terms of testing or engineering, have changed substantially between phases. In such a situation, you might want to compare the test phases independently. The Reinitialized Weibull NHPP option is available when a Phase column specifying at least two phases has been entered in the launch window.

For the algorithm to fit this model, each row that contains the first occurrence of a new phase must contain the start date.

- Suppose that a single column is entered as Time to Event or Timestamp. Then the start time for a new phase, with a zero Event Count, must appear in the first row for that phase. See the sample data table ProductionEquipment.jmp, in the Reliability subfolder, for an example.
- If two columns are entered, then an interval whose left endpoint is that start time must appear in the first row, with the appropriate event count. The sample data table

TurbineEngineDesign2.jmp, found in the Reliability subfolder, provides an example. Also, see "Piecewise NHPP Weibull Model Fitting with Interval-Censored Data".

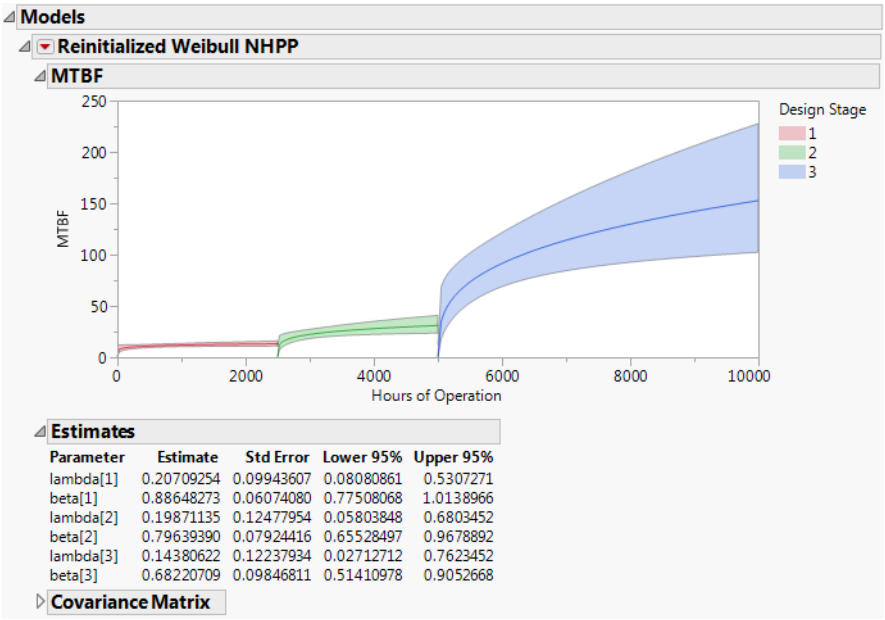
See "Multiple Test Phases".

Independent Crow-AMSAA models are fit to the data from each of the phases. When the report is run, the Cumulative Events plot updates to show the reinitialized models. Blue vertical dashed lines show the transition points for each of the phases. The Model List also updates.

Reinitialized Weibull NHPP Report

By default, the Reinitialized Weibull NHPP report shows the estimated MTBF plot. The phases are differentiated with different colors. The Estimates and Covariance Matrix reports are shown below the plot. (See Figure 10.21, which uses the ProductionEquipment.jmp sample data file from the Reliability subfolder.)

Figure 10.21 Reinitialized Weibull NHPP Report



MTBF Plot

The MTBF plot opens by default when the Reinitialized Weibull NHPP option is chosen. For more information s about the plot, see "MTBF Plot".

Estimates

The Estimates report gives estimates of λ and β for each of the phases. For a given phase, λ and β are estimated using only the data from that phase. The calculations assume that the phase begins at time 0 and reflect whether the phase is failure or time terminated, as defined by the data table structure. See "Test Phases". Also shown are standard errors and 95% confidence limits. These values are computed as described in "Parameter Estimates for Crow-AMSAA Models".

Covariance Matrix

Estimated covariance matrix for the estimates of the parameters of the fitted model. This report is closed by default.

Reinitialized Weibull NHPP Options

This section describes the options available in the Reinitialized Weibull NHPP red triangle menu.

Show MTBF Plot

This option shows or hides the MTBF plot. See "MTBF Plot".

Show Intensity Plot

The Intensity plot shows the estimated intensity functions for the phases, along with confidence bands. Because the intensity functions are computed based only on the data within a phase, they are discontinuous at phase transitions. Color coding facilitates differentiation of phases. See "Show Intensity Plot".

Show Cumulative Events Plot

The Cumulative Events plot for the Reinitialized Weibull NHPP model portrays the estimated cumulative number of events, with confidence bounds, over the design phases in the following way. Let t represent the time since the first phase of testing began. The model for the phase that is in effect at time t is evaluated at time t. In particular, the model for the phase that is in effect is not evaluated at the time since the beginning of the specific phase; rather it is evaluated at the time since the beginning of the *first* phase of testing.

At phase transitions, the cumulative events functions are discontinuous. The Cumulative Events plot matches the estimated cumulative number of events at the beginning of one phase to the cumulative number at the end of the previous phase. This matching allows the user to compare the observed cumulative events to the estimated cumulative events functions. Color coding facilitates differentiation of phases.

Show Profilers

Three profilers are displayed, showing estimated MTBF, Failure Intensity, and Cumulative Events. Note that the Cumulative Events Profiler is constructed as described in the Cumulative Events Plot section. In particular, the cumulative number of events at the beginning of one phase is matched to the number at the end of the previous phase. See "Show Profilers".

Piecewise Weibull NHPP Change Point Detection

The Piecewise Weibull NHPP Change Point Detection option attempts to find a time point where the reliability model changes. This might be useful if you suspect that a change in reliability growth has occurred over the testing period. Note that detection seeks only a single change point, corresponding to two potential phases.

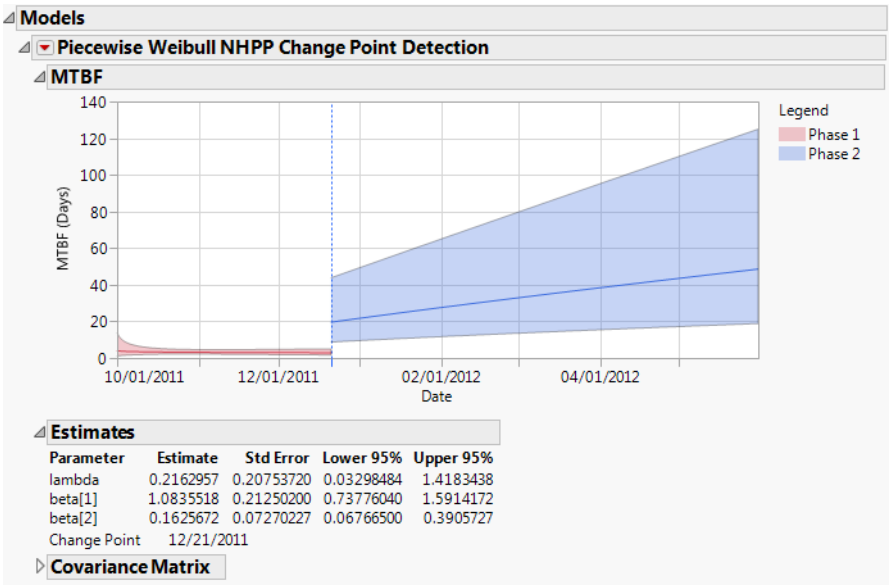
This option is available only when a Phase has *not* been entered in the launch window and one of the following is true:

- a single column has been entered as Time to Event or Timestamp in the launch window (indicating that failure times are exact), and
- two columns have been entered as Time to Event in the Concurrent Systems panel of the launch window.

When the Piecewise Weibull NHPP Change Point Detection option is selected, the estimated model plot and confidence bands are added to the Cumulative Events report under Observed Data. The Model List updates, giving statistics that are conditioned on the estimated change point. Under Models, a Piecewise Weibull NHPP Change Point Detection report is provided.

The default Piecewise Weibull NHPP Change Point Detection report shows the MTBF plot, Estimates, and Covariance Matrix. (See Figure 10.22, which uses the data in BrakeReliability.jmp, found in the Reliability subfolder.) Note that the Change Point, shown at the bottom of the Estimates report, is estimated as 12/21/2011. The standard errors and confidence intervals consider the change point to be known. The plot and the Estimates report are described in the section "Piecewise Weibull NHPP".

Figure 10.22 Piecewise Weibull NHPP Change Point Detection Report



Available options are: Show MTBF Plot, Show Intensity Plot, Show Cumulative Events Plot, Show Profilers. These options are described in the section "Reinitialized Weibull NHPP Options".

The procedure used in estimating the change point is described in "Statistical Details for the Piecewise Weibull NHPP Change Point Detection Report".

Additional Examples of the Reliability Growth Platform

- "Piecewise NHPP Weibull Model Fitting with Interval-Censored Data"
- "Piecewise Weibull NHPP Change Point Detection with Time in Dates Format"

Piecewise NHPP Weibull Model Fitting with Interval-Censored Data

The sample data file TurbineEngineDesign2.jmp, found in the Reliability subfolder, contains data on failures for a turbine engine design over three phases of a testing program. The first two columns give time intervals during which failures occurred. These intervals are recorded as days since the start of testing. The exact failure times are not known; it is known only that failures occurred within these intervals.

The reports of failures are provided generally at weekly intervals. Intervals during which there were no failures and which fell strictly within a phase are not included in the data table (for example, the interval 106 to 112 is not represented in the table). Because these make no contribution to the likelihood function, they are not needed for estimation of model parameters.

However, to fit a Piecewise Weibull NHPP or Reinitialized Weibull NHPP model, it is important that the start times for all phases be provided in the Time to Event or Timestamp columns.

Here, the three phases began at days 0 (Initial phase), 91 (Revised phase), and 200 (Final phase). There were failures during the weeks that began the Initial and Revised phases. However, no failures were reported between days 196 and 231. For this reason, an interval with beginning and ending days equal to 200 was included in the table (row 23), reflecting 0 failures. This is necessary so that JMP knows the start time of the Final phase.

The test was terminated at 385 days. This is an example of interval-censored failure times with time terminated phases.

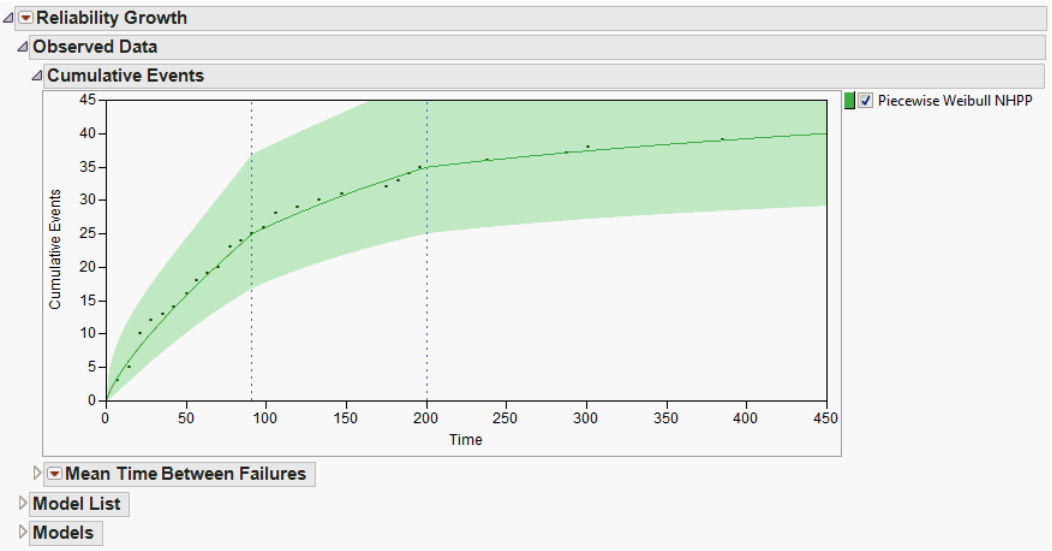
Note: The phase start times are required for proper display of the transition times for the Piecewise Weibull NHPP model; they are required for estimation of the Reinitialized Weibull NHPP model. For interval-censored data, the algorithm defines the beginning time for a phase as the start date recorded in the row containing the first occurrence of that phase designation. In our example, if row 23 were not in the table, the beginning time of the Final phase would be taken as 231.

---

1. Select **Help > Sample Data Library** and open Reliability/TurbineEngineDesign2.jmp.
2. Select **Analyze > Reliability and Survival > Reliability Growth**.
3. On the **Time to Event Format** tab, select the columns Interval Start and Interval End, and click **Time to Event**.
4. Select Fixes and click **Event Count**.
5. Select Design Phase and click **Phase**.
6. Click **OK**.
7. Click the Reliability Growth red triangle and select **Fit Model > Piecewise Weibull NHPP**.

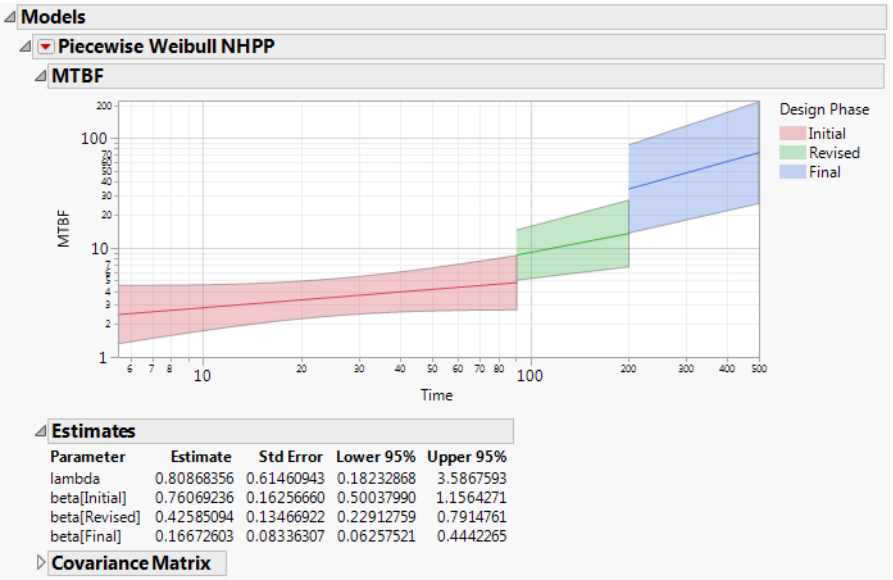
   The Cumulative Events plot from the Observed Data report is shown in Figure 10.23. The vertical dashed blue lines indicate the phase transition points. The first occurrence of Revised in the column Design Phase is in row 14. So, the start of the Revised phase is taken to be the Interval Start value in row 14, namely, day 91. Similarly, the first occurrence of Final in the column Design Phase is in row 23. So, the start of the Final phase is taken to be the Interval Start value in row 23, namely, day 200.

Figure 10.23 Cumulative Events Plot



The Piecewise Weibull NHPP report is found under the Models outline node. Here we see the mean time between failures increasing over the three phases. From the Estimates report, we see that the estimates of beta decrease over the three testing phases.

Figure 10.24 MTBF Plot



Parameter	Estimate	Std Error	Lower 95%	Upper 95%
lambda	0.80868356	0.61460943	0.18232868	3.5867593
beta[Initial]	0.76069236	0.16256660	0.50037990	1.1564271
beta[Revised]	0.42585094	0.13466922	0.22912759	0.7914761
beta[Final]	0.16672603	0.08336307	0.06257521	0.4442265

Piecewise Weibull NHPP Change Point Detection with Time in Dates Format

The file BrakeReliability.jmp, found in the Reliability subfolder, contains data on fixes to a braking system. The Date column gives the dates when Fixes, given in the second column, were implemented. For this data, the failure times are known. Note that the Date column must be in ascending order.

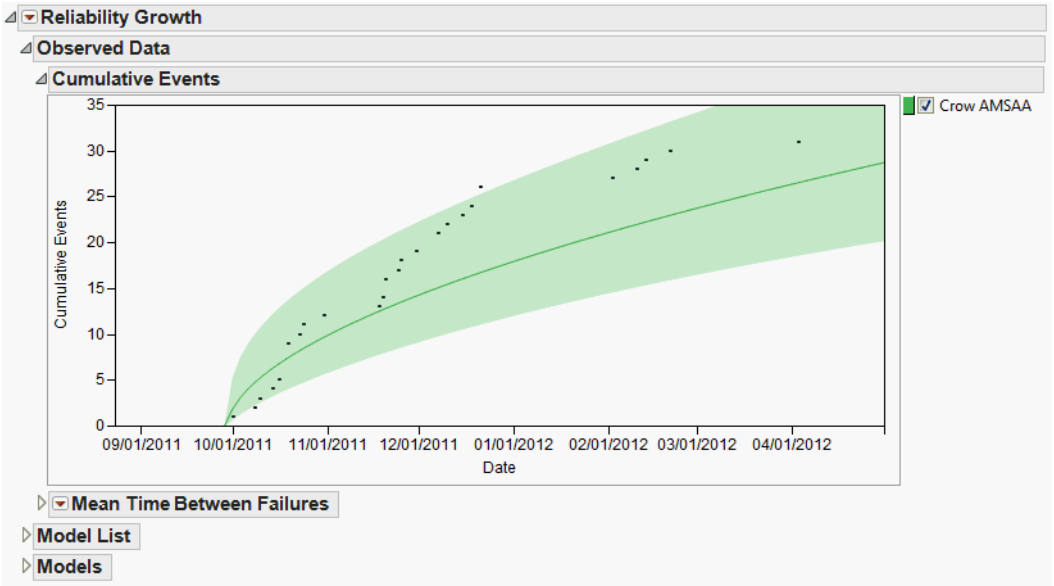
The test start time is the first entry in the Date column, 09/29/2011, and the corresponding value for Fixes is set at 0. This is needed in order to convey the start time for testing. If there had been a nonzero value for Fixes in this first row, the corresponding date would have been treated as the test start time. However, the value of Fixes would have been treated as 0 in the analysis.

The test termination time is given in the last row as 05/31/2012. Because the value in Fixes in the last row is 0, the test is considered to be time terminated on 5/31/2012. If there had been a nonzero value for Fixes in this last row, the test would have been considered failure terminated.

1. Select **Help > Sample Data Library** and open Reliability/BrakeReliability.jmp.
2. Select **Analyze > Reliability and Survival > Reliability Growth**.
3. Select the **Dates Format** tab.
4. Select Date and click **Timestamp**.
5. Select Fixes and click **Event Count**.
6. Click **OK**.
7. Click the Reliability Growth red triangle and select **Fit Model > Crow AMSAA**.

   The Cumulative Events plot in the Observed Data report updates to show the model. The model does not seem to fit the data very well.

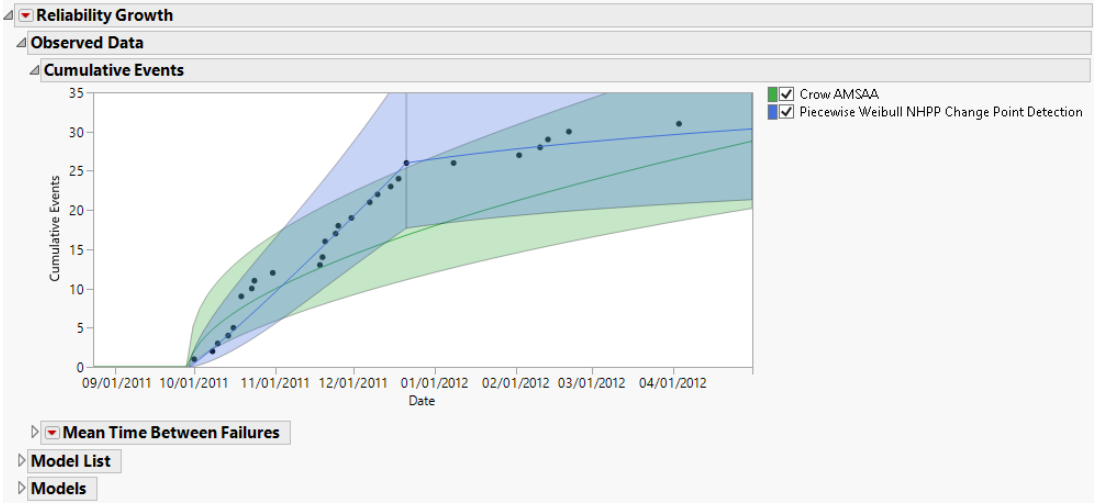
Figure 10.25 Cumulative Events Plot with Crow AMSAA Model



8. Click the Reliability Growth red triangle and select **Fit Model > Piecewise Weibull NHPP Change Point Detection**.

   The Cumulative Events plot in the Observed Data report updates to show the piecewise model fit using change-point detection. Both models are shown in Figure 10.26. Though the data are rather sparse, the piecewise model seems to provide a better fit to the data.

Figure 10.26 Cumulative Events Plot with Two Models



Statistical Details for the Reliability Growth Platform

- "Statistical Details for the Crow-AMSAA Report"
- "Statistical Details for the Piecewise Weibull NHPP Change Point Detection Report"

Statistical Details for the Crow-AMSAA Report

This section contains details for the parameter estimates and profilers that appear in the Crow-AMSAA Report.

Parameter Estimates for Crow-AMSAA Models

With the exception of the Crow-AMSAA with Modified MLE option, the estimates for λ and β are maximum likelihood estimates, which are computed as follows. The likelihood function is derived using the methodology in Meeker and Escobar (1998). It is reparametrized in terms of $param_1 = \log(\lambda)$ and $param_2 = \log(\beta)$. This is done to enable the use of an unconstrained optimization algorithm, namely, an algorithm that searches from -∞ to +∞. The MLEs for $param_1$ and $param_2$ are obtained.

The standard errors for λ and β are obtained from the Fisher information matrix. Confidence limits for $param_1$ and $param_2$ are calculated based on the asymptotic distribution of the MLEs, using the Wald statistic. These estimates and their confidence limits are then transformed back to the original units using the exponential function.

Parameter Estimates for Crow-AMSAA with Modified MLE

For the Crow AMSAA with Modified MLE option, the estimate for β is corrected for bias. The formula for the bias-corrected estimate of β depends on whether the test is failure terminated or time terminated.

Denote the MLE for β by $\hat{\beta}$, let n be the number of observations, and let T be the total test time.

The bias-corrected estimate (*modified MLE*) of β is $\bar{\beta}$, where:

$$\bar{\beta} = \left(\frac{n-2}{n}\right)\hat{\beta} \quad \text{for a failure-terminated test}$$

$$\bar{\beta} = \left(\frac{n-1}{n}\right)\hat{\beta} \quad \text{for a time-terminated test}$$

The modified MLE for λ, denoted $\bar{\lambda}$, is calculated according to the expression given by the likelihood function, but based on the adjusted value of beta:

$$\bar{\lambda} = n / T^{\bar{\beta}}$$

The covariance matrix for the parameters is estimated using the Fisher information matrix. See "Parameter Estimates for Crow-AMSAA Models". However, the bias-corrected estimates for λ and β are substituted for the MLEs in the resulting formulas. All confidence bands in plots and confidence intervals in reports are based on this procedure.

Profilers

For the Crow-AMSAA models, the estimates for the MTBF, Intensity, and Cumulative Events given in the profilers are obtained by replacing the parameters λ and β in their theoretical expressions by their MLEs. In the case of the Crow-AMSAA with Modified MLE option, the modified MLEs are used. Confidence limits are obtained by applying the delta method to the log of the expression of interest.

For example, consider the cumulative events function. The cumulative number of events at time t since testing initiation is given by $N(t) = \lambda t^{\beta}$. It follows that $\log(N(t)) = \log(\lambda) + \beta\log(t)$. The parameters λ and β in $\log(N(t))$ are replaced by their MLEs (or modified MLEs) to estimate $\log(N(t))$. The delta method is applied to this expression to obtain an estimate of its variance. This estimate is used to construct a 95% Wald-based confidence interval. The resulting confidence limits are then transformed using the exponential function to give confidence limits for the estimated cumulative number of events at time t.

Statistical Details for the Piecewise Weibull NHPP Change Point Detection Report

The change point is estimated as follows:

- Using consecutive event times, disjoint intervals are defined.
- Each point within such an interval can be considered to be a change point defining a piecewise Weibull NHPP model with two phases. So long as the two phases defined by that point each consist of at least two events, the algorithm can compute MLEs for the parameters of that model. The log-likelihood for that model can also be computed.
- Within each of the disjoint intervals, a constrained optimization routine is used to find a local optimum for the log-likelihood function.
- These local optima are compared, and the point corresponding to the largest is chosen as the estimated change point.

Note that this procedure differs from the grid-based approach described in Guo et al. (2010).

Chapter 11

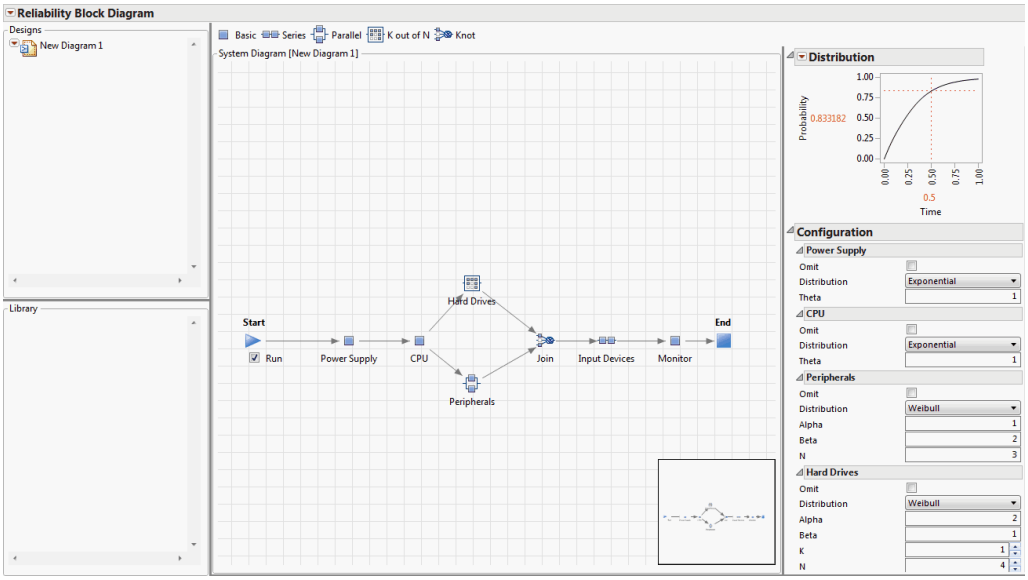
JMP PRO Reliability Block Diagram

Engineer System Reliabilities

The Reliability Block Diagram platform is available only in JMP Pro.

The Reliability Block Diagram platform displays the reliability relationships among a system's components. If reliability distributions are assigned to the components, the platform analytically models the reliability behavior. A Reliability Block Diagram (also known as a dependence diagram) illustrates how component reliability contributes to the success or failure of a system.

Figure 11.1 Example of a Reliability Block Diagram



Contents

Overview of the Reliability Block Diagram Platform 291
Example Using the Reliability Block Diagram Platform 291
Add Components 293
Align Shapes 295
Connect Shapes 296
Configure Components 297
The Reliability Block Diagram Window 299
Preview Window 300
Reliability Block Diagram Platform Options 301
Workspace Options 304
Configuration Settings 305
Distribution Configurations 305
Specify a Nonparametric Distribution 307
Options for Design and Library Items 308
Profilers 309
Component Importance and Time to Failure 311
Component Plots 313
Print Algebraic Reliability Formula 316
Generate Algebraic Expression Data Table 317
Clone and Delete 317

JMP PRO Overview of the Reliability Block Diagram Platform

The Reliability Block Diagram platform enables you to diagram a system and its related components to show how component reliability affects the success of the whole system. Each block in a Reliability Block Diagram represents a component in the system and is connected to other components in the system.

A Reliability Block Diagram can include components connected either in series or in parallel. A failure in any series component causes the entire system to fail. In a parallel-connected (or redundant) system, all parallel components must fail for the entire system to fail. In addition, the diagram can include K out of N components where *k*-out-of-*n* components must function for the system to function.

The Reliability Block Diagram template places a Start block on the left side and an End block on the right side of the diagram. Using the Shape tools, you diagram the system to be analyzed beginning at the Start block and connecting components to reach the End block.

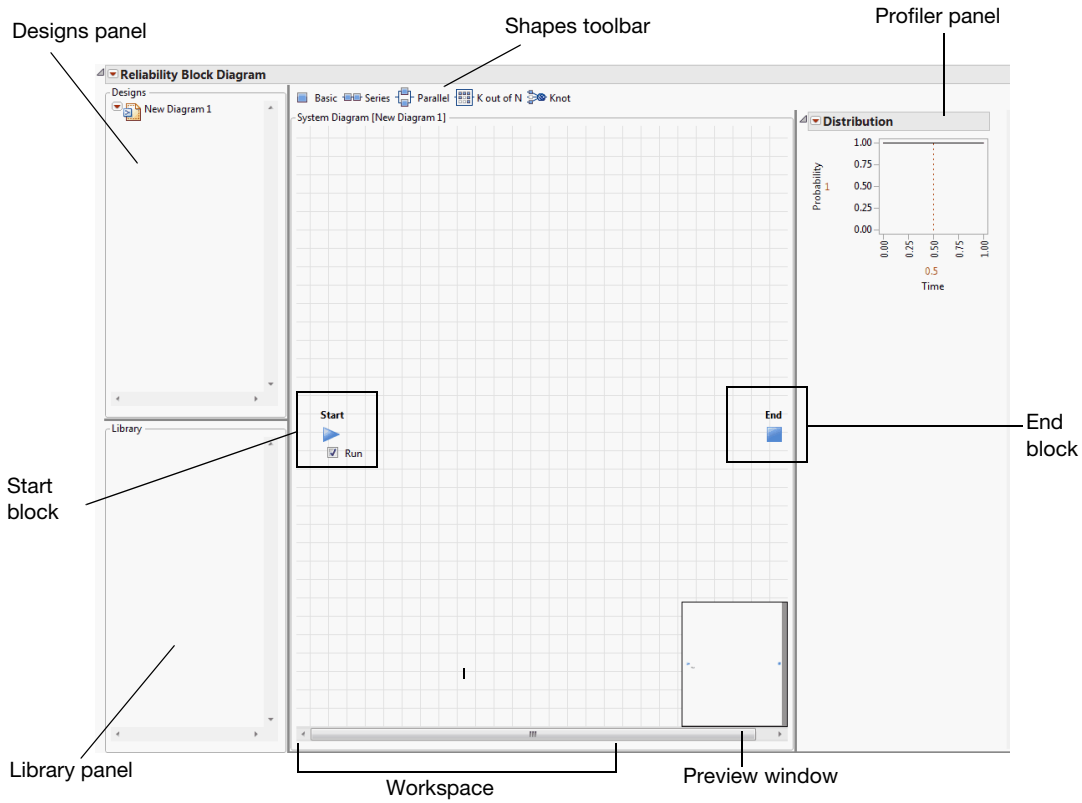
JMP PRO Example Using the Reliability Block Diagram Platform

In this example, you learn how to create a new Reliability Block Diagram.

1. Select **Analyze > Reliability and Survival > Reliability Block Diagram**.

   A blank Reliability Block Diagram window appears.

Figure 11.2 New Reliability Block Diagram



Note: The Distribution profiler appears by default.

2. In the Designs panel, select and rename New Diagram 1 to Computer.

   The Workspace is now named System Diagram [Computer].

3. Deselect **Run** in the Start block.

   With **Run** selected, the platform updates the diagram's reliability calculations after each change to the system diagram. These changes can include adding or deleting components, changing a component's configuration, and adding or deleting a connection.

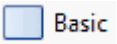
   With **Run** deselected, the platform does not update the reliability calculations after any changes.

Tip: Deselect **Run** when you are diagramming large systems. Select **Run** when the diagram is complete.

4. Proceed with "Add Components".

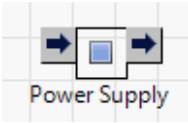
JMP PRO Add Components

The Reliability Block Diagram drawing elements that are located in the toolbar are called *shapes*. The term *component* refers to a shape that represents a constituent part of the system.

1. Click the Basic icon 

on the Shape toolbar and drag the shape to the System Diagram to the right of the Start block.
2. Select the label, replace New Basic 1 by Power Supply, and press Enter.

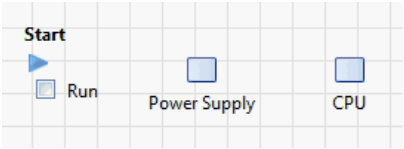
Figure 11.3 Basic Shape



When you click the label or the shape, connection arrows appear. The arrows disappear when you click elsewhere in the template.

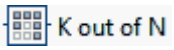
3. Drag a second Basic shape to the right of the Power Supply shape.
4. Select the label and type CPU.

Figure 11.4 Example System Diagram



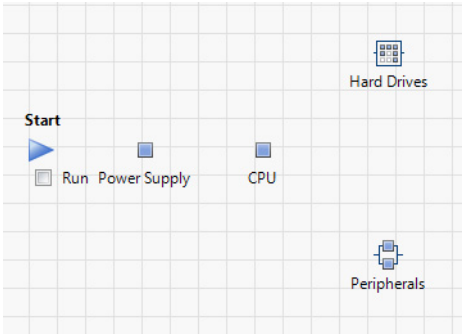
Note: You will align the shapes later, in the section "Align Shapes".

5. Drag a Parallel shape

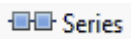
to a position to the right and below the CPU shape.
6. Select the label and type Peripherals.
7. Drag a K out of N shape 

to a position to the right and above the CPU shape.
8. Select the label and type Hard Drives.

Figure 11.5 Partial System Diagram

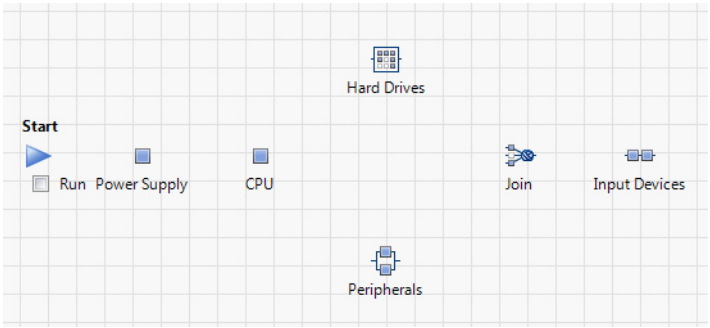


9. Drag a Knot shape

to the right of the previous shapes.
10. Select the label and type Join.
11. Drag a Series shape 

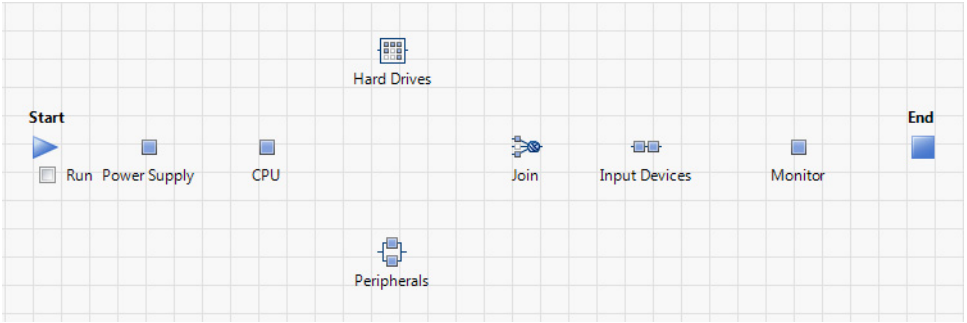
to a position to the right of the Knot shape.
12. Select the label and type Input Devices.

Figure 11.6 Partial System Diagram



13. Drag a Basic shape to a position to the right of the Input Devices shape.
14. Select the label and type Monitor.

Figure 11.7 System Diagram Showing All Shapes



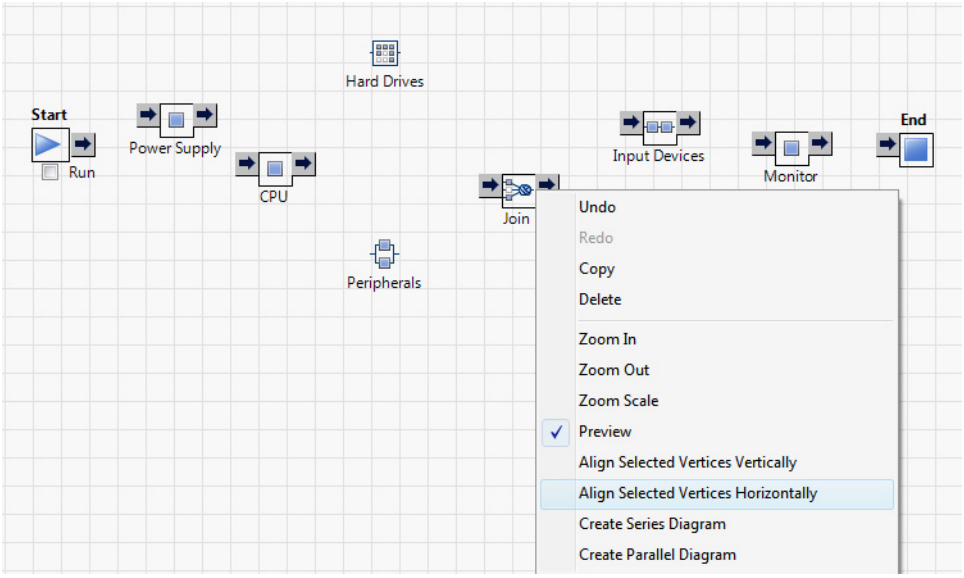
15. Proceed with "Align Shapes".

JMP PRO Align Shapes

1. To vertically align the shapes for Hard Drives and Peripherals, select the components:
   - Hard Drives
   - Peripherals

   Tip: To select shapes, drag the cursor around the shapes or press Shift and click each shape.
2. With the shapes selected, right-click one of the shapes and select **Align Selected Vertices Vertically**.
3. To horizontally align the remaining shapes, select the following components:
   - Start
   - Power Supply
   - CPU
   - Join
   - Input Devices
   - Monitor
   - End
4. With the shapes selected, right-click one of the shapes and select **Align Selected Vertices Horizontally**.

Figure 11.8 Align Shapes Horizontally



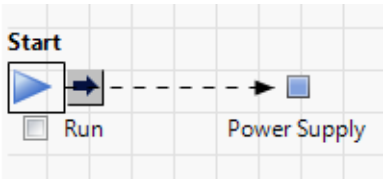
5. Proceed with "Connect Shapes".

JMP PRO Connect Shapes

To connect shapes, select a shape to display its connection arrows. Suppose you want to connect shape A to shape B. Select shape A. Drag the right arrow to shape B to indicate that shape A precedes shape B. Drag the left arrow to shape B to indicate that shape B precedes shape A. To connect the shapes in your diagram, select the right arrows to connect to the next shape in the sequence.

1. Select the Start block (blue arrow) to display the connection arrow.
2. Select the single connection arrow and drag it to the Power Supply component.

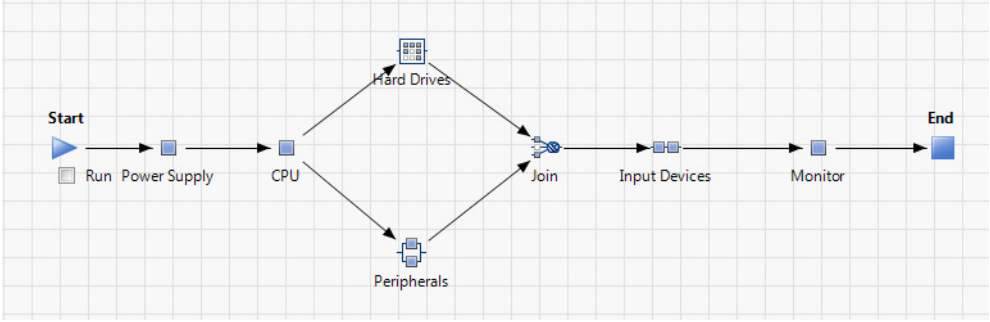
Figure 11.9 Connecting Shapes



3. For each of the following components, click the first component, select its right connection arrow, and drag the arrow to the second component:
    1. Power Supply → CPU

2. CPU → Hard Drive
3. CPU → Peripherals
4. Hard Drives → Join
5. Peripherals → Join
6. Join → Input Devices
7. Input Devices → Monitor
8. Monitor → End block

Figure 11.10 Completed System Diagram



4. Proceed with "Configure Components".

JMP PRO Configure Components

1. In the Configuration panel, enter the Configuration settings for the components. See "Configuration Settings".

Table 11.1 Configuration Settings

Component	Settings
Power Supply	• Distribution—Exponential • Theta—1
CPU	• Distribution—Exponential • Theta—1

Table 11.1 Configuration Settings *(Continued)*

Component	Settings
Peripherals	• Distribution—Weibull • Alpha—1 • Beta—2 • N—3
Hard Drives	• Distribution—Weibull • Alpha—2 • Beta—1 • K—1 • N—4
Join	Minimum available—1
Input Devices	• Distribution—Fréchet • Location—0 • Scale—1 • N—2
Monitor	• Distribution—Exponential • Theta—1

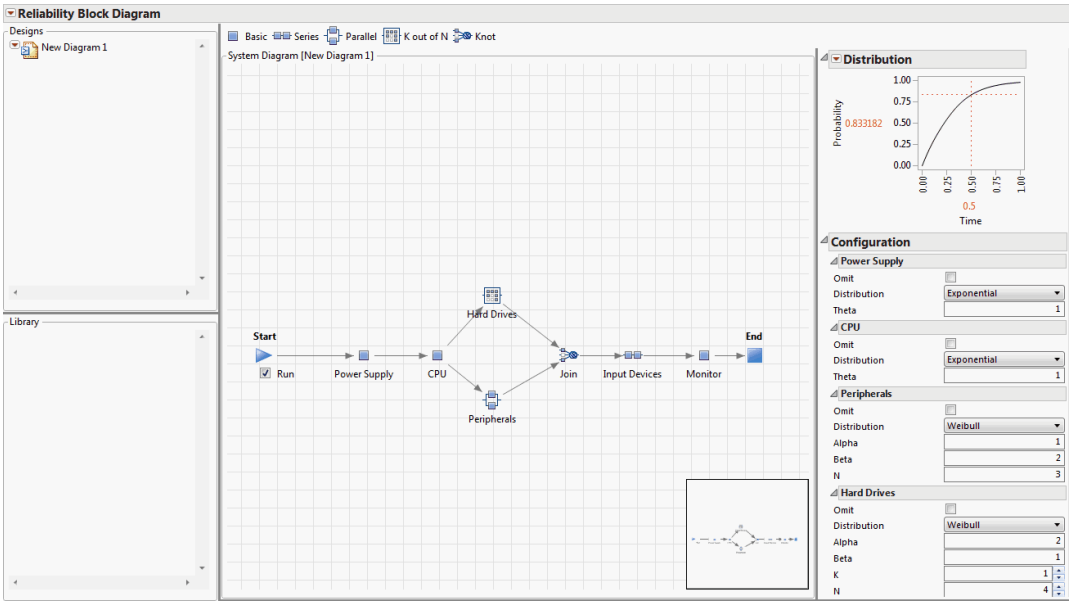
The Reliability Block Diagram is complete (Figure 11.11).

2. Select **Run**.

   The system's reliability information is updated. This is reflected in the Distribution plot in the Profiler pane.

3. To save the Reliability Block Diagram as a JMP Scripting Language (JSL) file, select **File > Save** and name it exampleRBDcomplete.jsl.

Figure 11.11 Example Reliability Block Diagram



The Reliability Block Diagram Window

The Reliability Block Diagram window is divided into the following panels:

- Designs: Lists system diagrams that were created using the Reliability Block Diagram platform.
- Library: Lists sub-system designs that are available for reuse in creating new system diagrams.

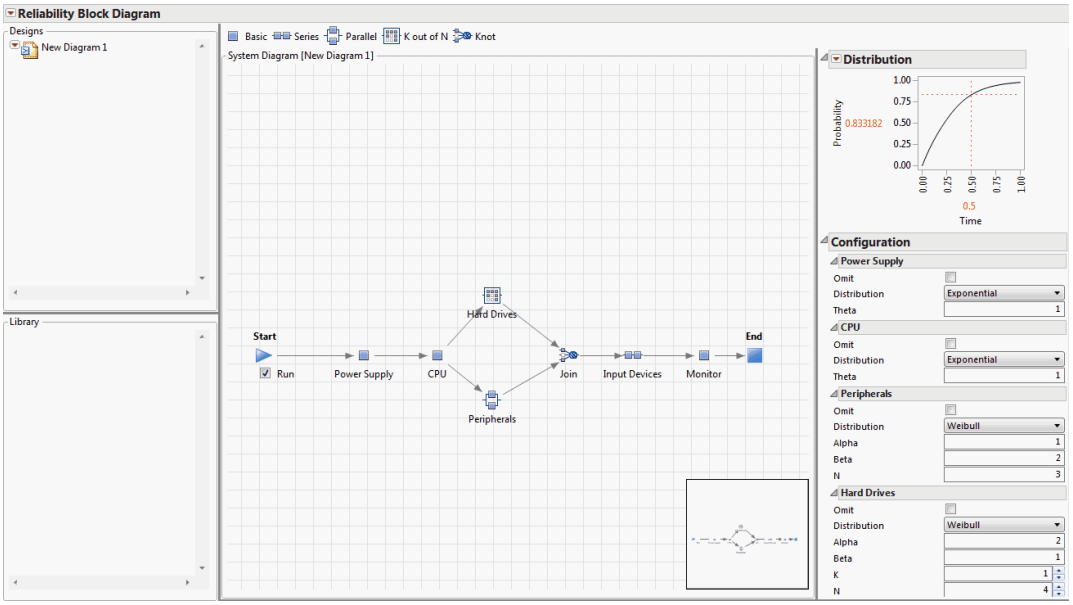
Note: Each system diagram that appears in the Designs and the Library panels contains its own red triangle menu. The options in each of those red triangle menus are described in "Options for Design and Library Items".

- Workspace: Displays the Shape toolbar, the System Diagram, and the Preview window.

Tip: To hide the Preview window, right-click in the Workspace and deselect Preview.

- Profiler Panel: Displays the Distribution profiler, Configuration settings, and various system and component profilers and plots that you select from the Diagram red triangle menu.

Figure 11.12 Example of a Reliability Block Diagram



The Shape toolbar includes the following drawing tools:

Basic Adds a single block shape to the system diagram.

Series Adds a series block shape that represents a group of components connected in a series. All components must function for the system to function.

Parallel Adds a parallel block shape that represents a group of components that are in parallel. At least one of the components must function for the system to function.

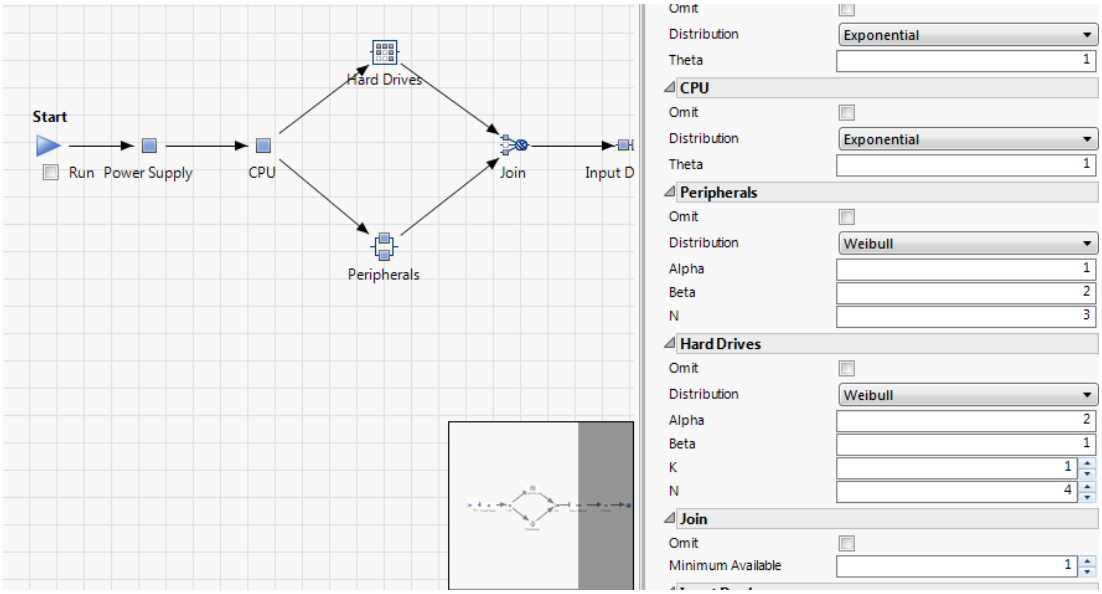
K out of N Adds a *k*-out-of-*n* block shape where you specify *k* and *n*. At least k of the components must function for the system to function.

Knot Adds a knot shape to the system diagram that enables you to join Parallel or K out of N components that have different Distribution property settings.

JMP PRO Preview Window

When the system diagram is too large to appear in the workspace, the Preview window enables you to reposition the portion of the system diagram that appears in the workspace. The viewing area is indicated in the Preview window by a white background. Drag the viewing area to reposition the view of the Workspace to another part of the system diagram.

Figure 11.13 Preview Window with Visible Part of Diagram on White Background



JMP PRO Reliability Block Diagram Platform Options

The Reliability Block Diagram red triangle menu contains the following options for the Reliability Block Diagram window:

Save and Save As Enables you to save a Reliability Block Diagram, or save an existing Reliability Block Diagram with a new name, to a JMP Scripting Language (JSL) script that is automatically executed when it is opened in JMP. See the *Scripting Guide* for more information about Auto-Submit scripts.

Note: The Save and Save As red triangle options are equivalent to File > Save and File > Save As. They are available in the red triangle menu for convenience.

Show Design Diagram If **Show Design Comparisons** is selected, this option enables you to hide or show the system diagram in the Workspace.

Show Design Comparisons Displays a Distribution Overlay profiler and Remaining Life Distribution Overlay profiler for the selected system diagrams. See "Show Design Comparisons".

Import Component Distribution Settings Enables you to import configuration settings for the system diagram from a data table. The table must contain columns for the diagram category, diagram name, component name, distribution, and one or more parameters. The number of parameters depends on the specified distribution.

Note: The strings in the imported table must be exact matches to the strings in the system diagram.

New Design Item Enables you to add a new design diagram to the Designs panel. See "Add a New Design Item".

New Library Item Enables you to add a new item to the Library panel. The Library panel lets you create subsystem diagrams for use in multiple system designs. See "Add a New Library Item".

Note: Library items are available only for use in diagrams that are contained in the Designs panel of a JSL file. They are not available to other diagrams in other JSL files.

JMP PRO Show Design Comparisons

The **Show Design Comparisons** option displays a Distribution Overlay plot and a Remaining Life Distribution Overlay plot. These plots appear in the Workspace below the System Diagram. See Figure 11.14 and Figure 11.15.

- The Distribution Overlay plot shows the probability of system failure for each of the designs in the Designs panel.
- The Remaining Life Distribution Overlay plot shows the conditional probability of failure for each design in the Designs panel, given that the systems have survived to a specified survival time. Enter the Survival Time in the box to the right of the plot. Alternatively, select the small rectangle at the origin and drag it to the right to dynamically set the Survival Time.

Check boxes enable you to select which designs are represented in the plots. This enables you to compare a subset of designs with respect to failure probabilities.

Tip: To hide the system diagram in the Workspace, click the Reliability Block Diagram red triangle and deselect **Show Design Diagram**. Alternatively, to show more of the Design Comparisons, reposition the horizontal splitter upward.

Figure 11.14 Distribution Overlay Example

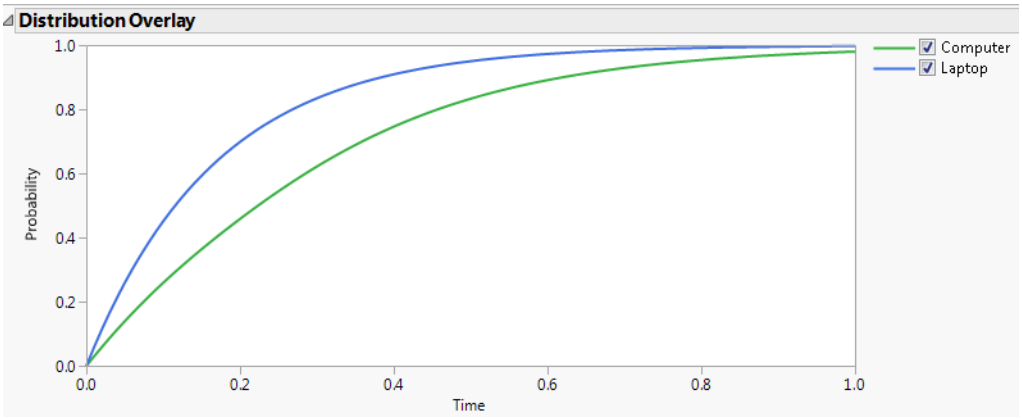
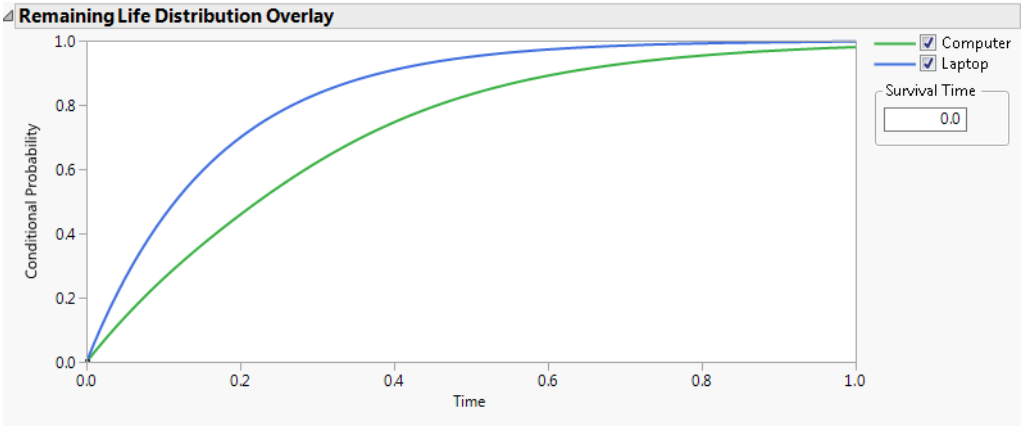


Figure 11.15 Remaining Life Distribution Overlay Example



JMP PRO Add a New Design Item

The **New Design Item** option adds a new system design to the Designs panel list.

- Click the Reliability Block Diagram red triangle and select **New Design Item**.
- A design named New Diagram X, where X is a number that identifies the diagram name, is added to the Designs panel list.
- To name the design, select the label and enter a name.
- Use the Shape toolbar to draw the system diagram. See "Example Using the Reliability Block Diagram Platform".

Tip: You can also copy and paste the body of a diagram from another block diagram template. The Start and End shapes are not copied. After you paste the diagram, you must reconnect the Start and End shapes.

- Configure the components.
- Save the file.

To display a design in the System Diagram window and to view its Profiler panel, double-click its icon in the Designs panel.

JMP PRO Add a New Library Item

The **New Library Item** option adds a new sub-system to the Library panel list.

Tip: Drag a system design from the Designs panel to the Library panel to add it to the Library as a sub-system.

- Click the Reliability Block Diagram red triangle and select **New Library Item**.
- A sub-system called New Diagram X, where X is a number that identifies the diagram name, is added to the Library panel list.
- To name the sub-system, select the label and enter a name.
- Use the Shape toolbar to draw the sub-system diagram.
- Configure the components.
- Create connections from the Start block to the End block.
- Save the file.

JMP PRO Workspace Options

The Workspace panel supports several right-click commands for adjusting the view of the panel. Here are some of the available commands:

Zoom In Enables you to zoom in to the system diagram.

Zoom Out Enables you to zoom out from the system diagram.

Tip: On Windows, you can press Ctrl and use the mouse scroll wheel to zoom in to and out from the diagram.

Zoom Scale Opens the Set Zoom Scale window enabling you to zoom in or out using a scaling factor that is specified as a decimal. By default, the zoom scale is set to 1 (100%).

Preview Enables you to hide or show the Preview window in the Workspace panel.

Align Selected Vertices Vertically Enables you to align shapes on a vertical line.

Align Selected Vertices Horizontally Enables you to align shapes on a horizontal line.

Create Series Diagram Enables you to create a series diagram using the nodes in the diagram. The order of the series is determined by the order of node creation. This option is available only when the diagram does not contain any arrows.

Create Parallel Diagram Enables you to create a parallel diagram using the nodes in the diagram. This option is available only when the diagram does not contain any arrows.

JMP PRO Configuration Settings

Each component in a Reliability Block Diagram can be assigned a failure distribution. The available failure distributions are listed in Figure 11.2. To see the formulas and parameterization for these failure distributions, see "Distributions".

JMP PRO Distribution Configurations

When a component is added to the diagram, an outline for that component appears beneath the Configuration outline.

Note: You can omit a component from the analysis by checking the box next to Omit.

Select a Distribution and enter the required parameter values.

Table 11.2 Distributions and Parameters

Property Type	Required Inputs
Exponential	Theta
Weibull	Alpha, Beta
Lognormal	location, scale
Loglogistic	location, scale
Fréchet	location, scale
GenGamma	mu, sigma, lambda
DS Weibull	Alpha, Beta, Defective Probability
DS Lognormal	location, scale, Defective Probability

Table 11.2 Distributions and Parameters *(Continued)*

Property Type	Required Inputs
DS Loglogistic	location, scale, Defective Probability
DS Fréchet	location, scale, Defective Probability
Nonparametric	data or data file

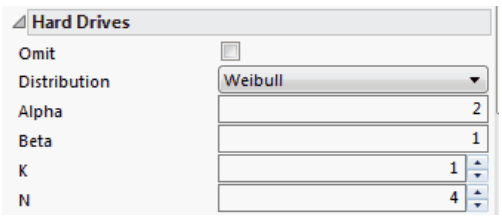
To view Configuration settings for the components in a selected design or subsystem, do the following:

- To view the Configuration settings for a specific component, select the component's shape.
- To view Configuration settings for more than one component, select multiple components' shapes using the Arrow tool or pressing Control and clicking.
- To view Configuration settings for all components, deselect any shapes by clicking in a blank portion of the Workspace.

For each component in the diagram, you can either omit the component from calculations by checking the box next to Omit or:

- From the Distribution list, select the appropriate distribution. For all selections other than Nonparametric, enter parameter values for the distribution. For Nonparametric, see "Specify a Nonparametric Distribution".
- For Series and Parallel components, you must also enter a value for N, the total number of components contained in the series or parallel shape.
- For K out of N components, you must also enter K, the minimum number of components that must function for the system to function, and N, the total number of components in the shape.
- For Knot components, enter the Minimum Available Dependences. The Knot shape enables you to configure a *k*-out-of-*n* shape where the shapes that are joined have different distributions. The Minimum Number of Dependencies is *k*, the minimum number of paths leading to the Knot that need to function in order for the system to function.

Figure 11.16 Example of a Weibull Configuration for a K out of N Shape

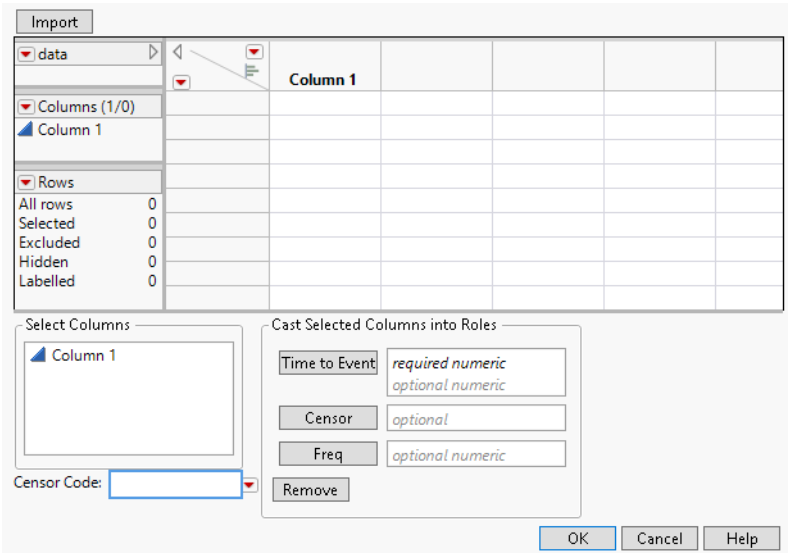


JMP PRO Specify a Nonparametric Distribution

The Nonparametric option under Distribution enables you to approximate an arbitrary distribution. Enter data or import a file containing a large set of data. This data is used to approximate the distribution.

After selecting Nonparametric, click the icon next to Data. The Provide Data window appears, enabling you to either enter data or import a data file. Once you have imported or entered your data, the data is used to calculate a nonparametric distribution for the component.

Figure 11.17 Provide Data Window



To import data from a file, do the following.

1. Open the JMP data table that contains the data to import.
2. In the Provide Data window, click **Import**.

   The Select Data Table window appears.
3. From the Data Table list, select the data table.
4. Click **OK**.
5. In the panel beneath the data grid, specify which columns represent Time to Event data, as well as Censor and Freq data if appropriate.
6. In the Provide Data window, click **OK**.

To enter data manually, do the following.

1. Create columns for Time to Event data, as well as Censor and Freq data, if appropriate.

2. Enter the data into the columns.
3. In the panel beneath the data grid, specify which columns represent Time to Event data, as well as Censor and Freq data if appropriate.
4. In the Provide Data window, click **OK**.

JMP PRO Options for Design and Library Items

The available options for Design and Library items are listed below:

- Show Configuration shows or hides the Configuration outline. See "Configuration Settings".
- Options to show various profilers. See "Profilers" and the following:
  - "Distribution Profiler"
  - "Remaining Life Distribution Profiler"
  - "Reliability Profiler"
  - "Quantile Profiler"
  - "Density Profiler"
  - "Hazard Profiler"
  - "Cumulative Hazard Profiler"
- Options to show plots for component importance measures and mean time to failure. See "Component Importance and Time to Failure" and the following:
  - "Birnbaum's Component Importance"
  - "Remaining Life BCI"
  - "Component Integrated Importance"
  - "Mean Time to Failure"
- Options to show overlay plots for the components in a system diagram. See "Component Plots" and the following:
  - "Component Distribution Functions"
  - "Component Reliability Functions"
  - "Component Density Functions"
  - "Component Hazard Functions"
  - "Component Cumulative Hazard Functions"
- "Print Algebraic Reliability Formula"
- "Generate Algebraic Expression Data Table"

- "Clone and Delete"

JMP PRO Profilers

Various profilers are provided to help you analyze the reliability properties of a system. You can view profilers for each system diagram that is listed in the Designs and the Library panels. This section describes the profilers. The profilers appear in the Profilers panel of the report.

Red Triangle Options for Profilers

Each profiler has a red triangle menu that contains the following commands:

Reset Factor Grid Displays a window where you can enter a current setting and values that control the display. See *Profilers* for more information about setting the factor grid.

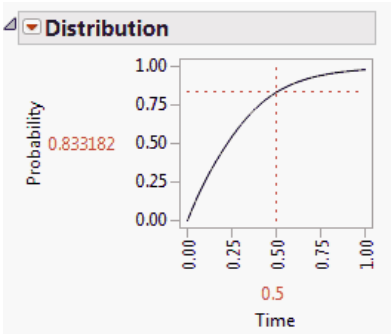
Factor Settings Select this option to configure the profiler's settings and to link the profilers. See *Profilers* for more information about Factor Settings.

Note: Adjust the X and Y axes of a profiler to view the desired portion of the graph.

JMP PRO Distribution Profiler

The Distribution Profiler displays the probability that the system fails as a function of time. The Distribution profiler for the system appears by default.

Figure 11.18 Distribution

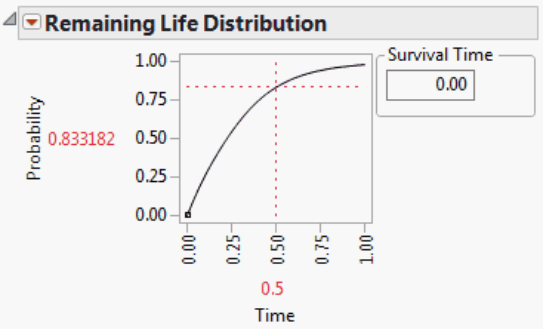


JMP PRO Remaining Life Distribution Profiler

The Remaining Life Distribution profiler for the system shows the probability that the system fails given that it has survived a specified amount of time, designated as Survival Time. By default, Survival Time is set to zero, and the remaining life distribution function is equivalent to the distribution function.

Enter a value for **Survival Time** to indicate the time to which the system has survived without failing. As an alternative to entering a value, select the small rectangle at the origin of the graph and drag it to the right to dynamically set the Survival Time.

Figure 11.19 Remaining Life Distribution



JMP PRO Reliability Profiler

The Reliability profiler for the system shows the probability that the system will function as a function of time. The reliability function is also known as the survival function.

JMP PRO Quantile Profiler

The Quantile profiler for the system shows time as a function of the failure probability. Note that the Quantile function is the inverse of the Distribution function.

JMP PRO Density Profiler

The Density profiler for the system shows the probability density function associated with the system's failure distribution function.

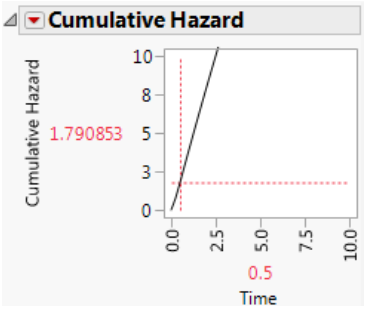
JMP PRO Hazard Profiler

The Hazard profiler for the system shows the instantaneous failure rate at a given time.

JMP PRO Cumulative Hazard Profiler

The Cumulative Hazard profiler for the system shows the cumulative hazard function as a function of time.

Figure 11.20 Cumulative Hazard



JMP PRO Component Importance and Time to Failure

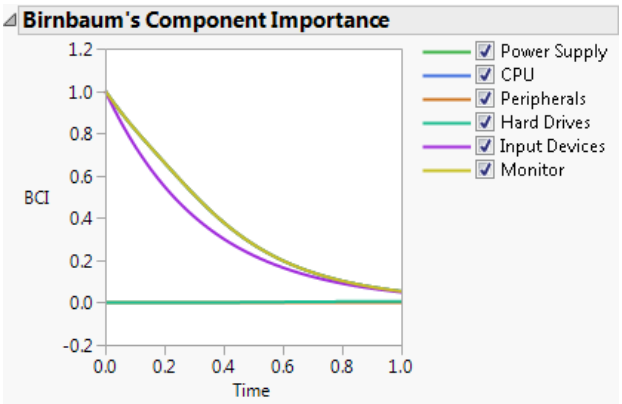
Plots that analyze component importance and mean time to failure are provided for each system diagram that is listed in the Designs and the Library panels. This section describes the component importance measures and mean time to failure. These plots and statistics appear in the profilers panel of the report.

Note: Check boxes in the legend of each component importance plot enable you to select which components to view in that plot.

JMP PRO Birnbaum's Component Importance

Select **Show BCI** to show an overlay plot of the Birnbaum's Component Importance measures for each component of the selected system diagram. A component's BCI at a given time is the probability that the system fails if the component fails. A component with a large BCI is critical to system reliability.

Figure 11.21 Birnbaum's Component Importance

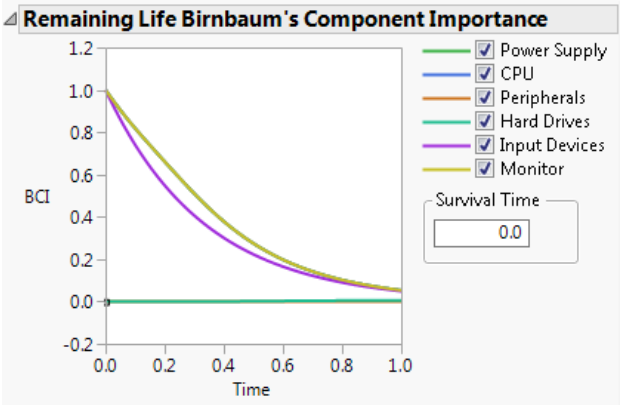


JMP PRO Remaining Life BCI

Select **Show Remaining Life BCI** to show an overlay plot of the Birnbaum's Component Importance for Remaining Life. The BCI for Remaining Life is the probability that the system fails if the component fails, given that the system has survived a specified amount of time, designated as Survival Time. By default, Survival Time is set to zero, and the BCI for Remaining Life is equivalent to the Birnbaum's Component Importance.

Enter a value for **Survival Time** to indicate the time to which the system has survived without failing. As an alternative to entering a value, select the small rectangle at the origin of the graph and drag it to the right to dynamically set the Survival Time.

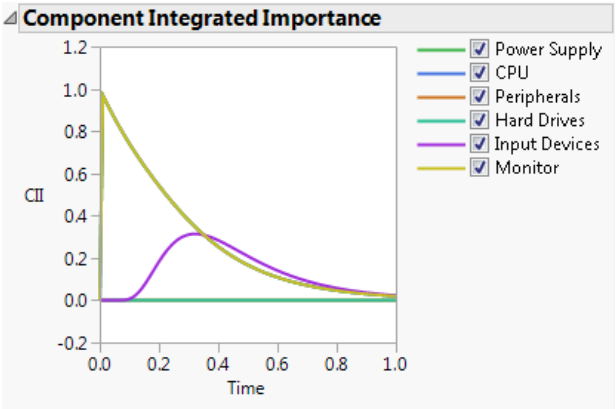
Figure 11.22 Birnbaum's Component Importance for Remaining Life



JMP PRO Component Integrated Importance

Select **Show Component Integrated Importance** to show an overlay plot of the integrated importance measures for the components of the Reliability Block Diagram. The integrated importance measure for each component takes into account the failure rate of the component as well as the likelihood of failing instantaneously. See Si et al. (2012).

Figure 11.23 Component Integrated Importance



JMP PRO Mean Time to Failure

Select **Show MTTF** to view the Mean Time to Failure (MTTF) for the system.

Note: The formula that is used to calculate the Mean Time to Failure depends on the specified failure distributions and Configuration settings for each component in the system.

JMP PRO Component Plots

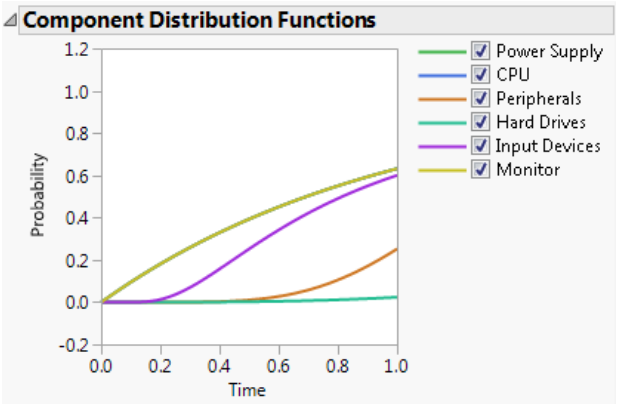
A system diagram usually contains many individual components. The reliability functions of each of these components can be examined in overlay plots. You can view component overlay plots for each system diagram that is listed in the Designs and the Library panels. This section describes the component overlay plots. These plots appear in the profilers panel of the report.

Note: Check boxes in the legend of each component plot enable you to select which components to view in that plot.

JMP PRO Component Distribution Functions

Select **Show Component Distribution Functions** to show an overlay plot of the distribution functions for the components of the Reliability Block Diagram.

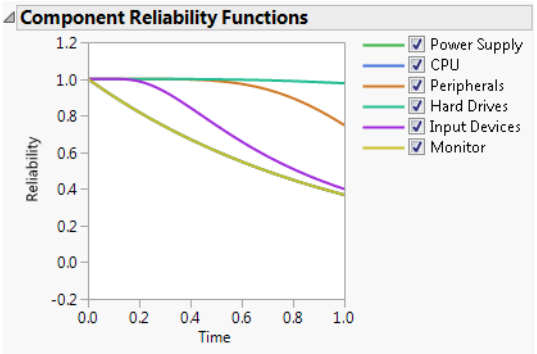
Figure 11.24 Component Distribution Functions



JMP PRO Component Reliability Functions

Select **Show Component Reliability Functions** to show an overlay plot of the reliability functions for the components of the Reliability Block Diagram.

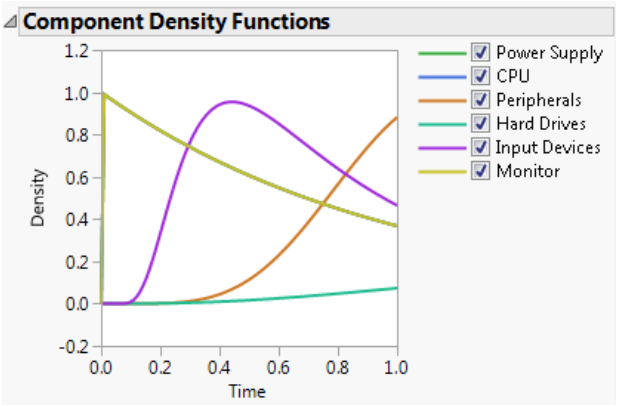
Figure 11.25 Component Reliability Functions



JMP PRO Component Density Functions

Select **Show Component Density Functions** to show an overlay plot of the density functions for the components of the Reliability Block Diagram.

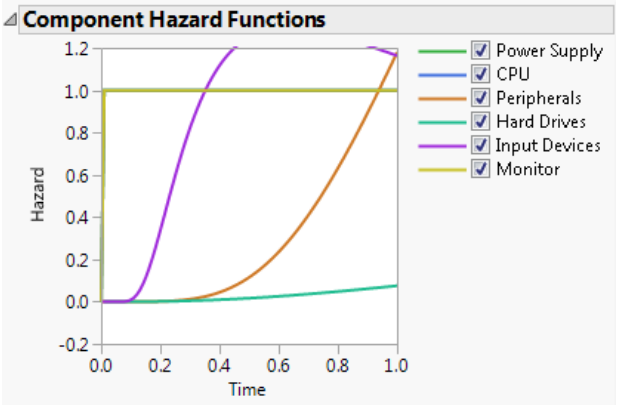
Figure 11.26 Component Density Functions



JMP PRO Component Hazard Functions

Select **Show Component Hazard Functions** to show an overlay plot of the hazard functions for the components of the Reliability Block Diagram.

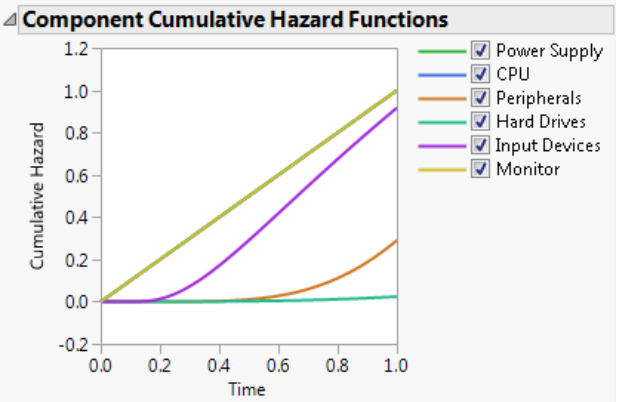
Figure 11.27 Component Hazard Functions



JMP PRO Component Cumulative Hazard Functions

Select **Show Component Cumulative Hazard Functions** to show an overlay plot of the cumulative hazard functions for the components of the Reliability Block Diagram.

Figure 11.28 Component Cumulative Hazard Functions



JMP PRO Print Algebraic Reliability Formula

The Print Algebraic Reliability to the Log Window option prints the reliability formula for the selected system diagram to the Log window.

1. (Windows only) Select **View > Log**.

   Or

   (macOS only) Select **Window > Log**.
2. Open the exampleRBDcomplete.jsl file that you created.
3. In the Designs panel, select **Computer**.
4. Click the Computer red triangle and select **Print Algebraic Reliability to the Log Window**.

   The Log window displays the formula for the Algebraic Reliability of the selected block diagram.

Figure 11.29 Algebraic Reliability in Log Window

```
/*:

Reliability Block Diagram[]Algebraic Reliability:
R["Power Supply"] * R["CPU"] * R["Peripherals"] * R["Input Devices"] * R["Monitor"]
 + R["Power Supply"] * R["CPU"] * F["Peripherals"] * R["Hard Drives"] * R[
"Input Devices"] * R["Monitor"]
```

Note: Reliability formulas use “R” to represent a component’s reliability and “F” to represent probability of a component’s failure.

JMP PRO Generate Algebraic Expression Data Table

The Generate Algebraic Expression Data Table option creates a new data table that contains the algebraic reliability expression as a column formula. The table contains a column for each component in the system. The final column contains the system reliability algebraic expression as a JSL column formula that use the preceding columns as arguments.

JMP PRO Clone and Delete

Each system diagram that is listed in the Designs and the Library panels has options that aid in managing the design and library items. This section describes the Clone and Delete operations.

JMP PRO Clone a Design or Library Item

The Clone option creates a new system design or library sub-system identical to the selected system design or library sub-system.

- Select the **Clone** option from the red triangle menu for the design or library item to be copied.

  A new design or library item is added to the Designs or Library list.
- Save the file.

JMP PRO Delete a Design or Library Item

The Delete option removes the selected system design or a library sub-system from the panel.

- Select the **Delete** option from the red triangle menu for the design or library item to be deleted.

  The specified design or library item is deleted from the Designs or Library list.
- Save the file.

Chapter 12

JMP PRO Repairable Systems Simulation

Estimate Outage Time for a Complex System

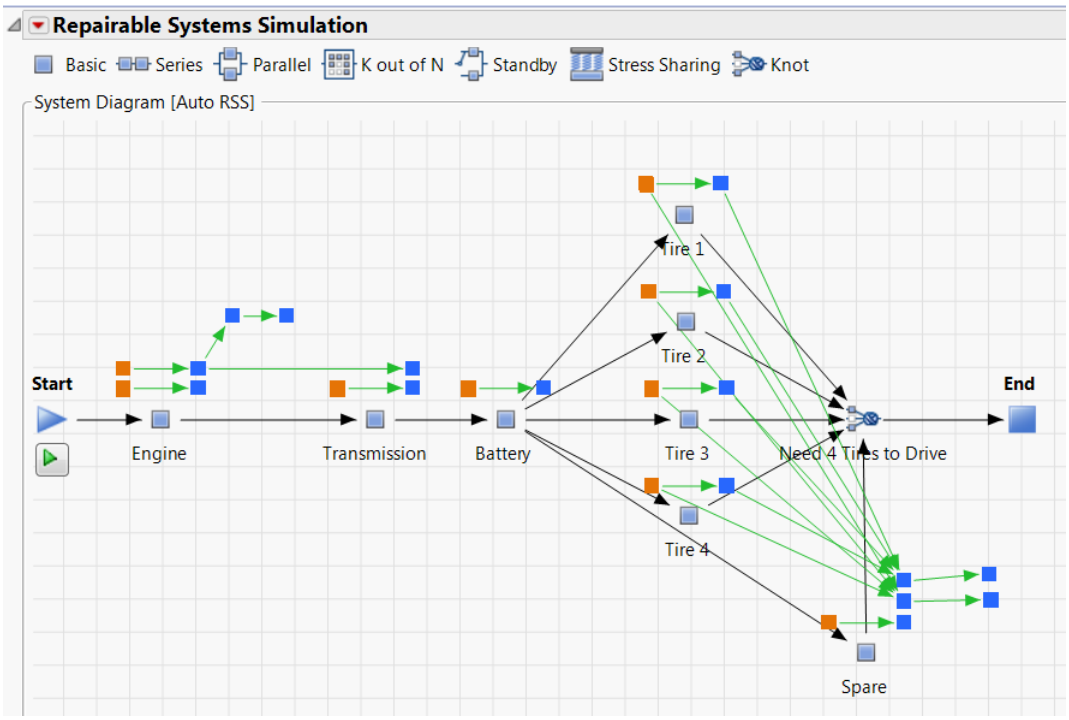
The Repairable Systems Simulation platform is available only in JMP Pro.

The Repairable Systems Simulation (RSS) platform enables you to explore the reliability within complex repairable systems. A *repairable system* consists of individual components that age and require maintenance. The state of a repairable system changes as its components fail and are subsequently repaired.

Repairable systems are involved in many aspects of modern life. They can range in size from refrigerators and houses to power plants and telephone communication networks.

The RSS platform enables you to simulate different system configurations in order to optimize your repairable system. For example, you can extend the useful life of costly components or maximize production outputs while maintaining a safe system operation.

Figure 12.1 Example of a Repairable Systems Simulation Diagram



Contents

Overview of the Repairable Systems Simulation Platform . . . 321
Example Using the Repairable Systems Simulation Platform . . . 321
The Repairable Systems Simulation Window . . . 323
System Diagram . . . 324
Shapes Toolbar . . . 325
Configuration Panel . . . 326
Add Event Panel . . . 329
Add Action Panel . . . 330
Repairable Systems Simulation Platform Options . . . 333
Options for Block Items . . . 334
Distribution Options . . . 334
Specify a Nonparametric or Estimated Distribution . . . 335
Event Settings . . . 336
Action Settings . . . 337
Repairable Systems Simulation Results . . . 338
Results Table . . . 338
Results Explorer . . . 339
Repairable Systems Simulation Results Options . . . 341
RSS Results Explorer Point Estimation Profilers . . . 342

JMP PRO Overview of the Repairable Systems Simulation Platform

The Repairable Systems Simulation (RSS) platform enables you to simulate a repairable system and to analyze its *outage time*. Outage time is the total time a system is not in the On state as a result of either planned maintenance or unintentional failure.

A repairable system is represented by a system diagram. Block shapes in the system diagram represent one or more components that connect to other components. A component that is performing work in the block is said to be *functional*. Components can be connected in series or in parallel. When components are connected in series, the system fails if any of the components in the series fails. When components are connected in parallel, the system fails if all of the parallel components fail.

A component pathway is said to be *uninterrupted* when at least one path between the Start and End blocks does not pass through a block shape failure. During a simulation, the system remains in the On state as long as there is an uninterrupted component pathway. If a block shape failure interrupts the component pathway, the system is set to the Down state. If a block shape fails but the component pathway is not interrupted, the system remains in the On state.

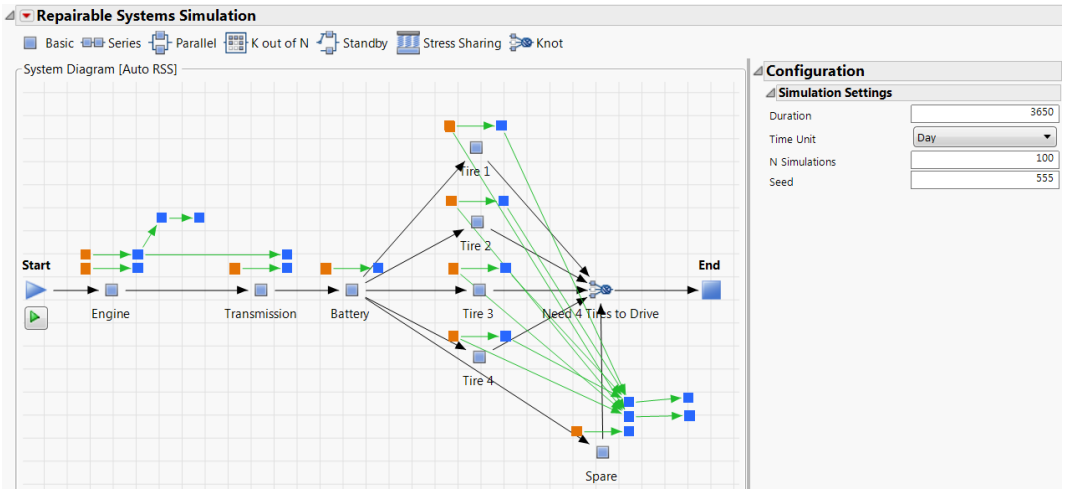
You can add unique events and actions to characterize the repairable nature of individual components. An *event* is a specific occurrence within the system, such as component failure, scheduled maintenance, or the start of a simulation iteration. Events trigger the execution of one or more *actions*, which express how components behave. Actions can alter the state of a component or the entire system.

JMP PRO Example Using the Repairable Systems Simulation Platform

In this example, you run 100 iterations of a simulation of a car system and analyze the system's estimated outage time over the course of 10 years.

1. Select **Help > Sample Data**, click **Open the Sample Scripts Directory**, and open Car Repair Simulation.jsl.

Figure 12.2 Car Repair System Diagram



An RSS window that diagrams a car system appears. In the diagram, the first block on the left is the Start block. Notice in the Configuration panel, on the right side of the RSS window, that the simulation is set to run for 3650 days.

2. (Optional) Enter 555 next to Seed.

   Because the simulation involves random component failures, this action ensures that you obtain the exact results shown below.

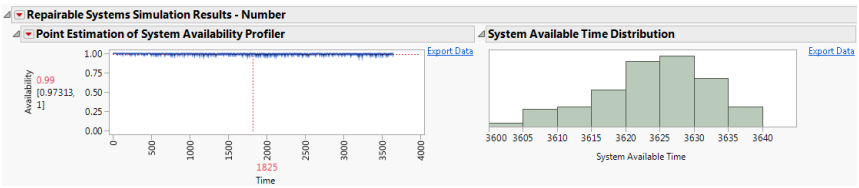
3. Click the green triangle below the Start block to simulate the car system.

   A data table that contains the simulation results appears.

4. Click the green arrow next to the Launch Repairable Systems Simulation Results Explorer script.

   A window appears containing the Repairable Systems Simulation Results report. For more information about how to interpret these results, see "Repairable Systems Simulation Results".

Figure 12.3 Partial RSS Report

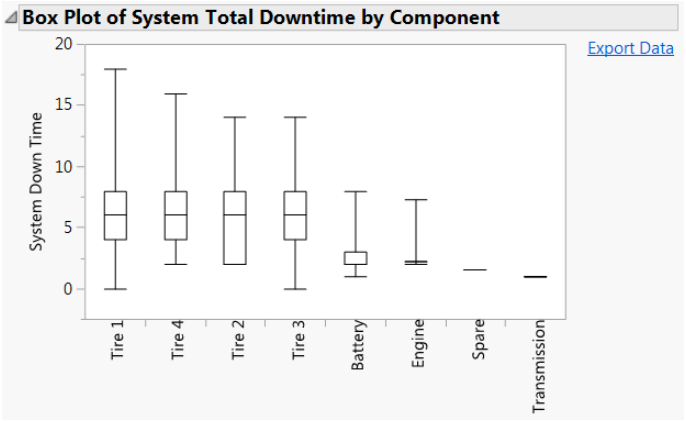


You predict that the car will be available between 3,600 and 3,640 days over the next 10 years. Because the values shown in the Point Estimation of System Availability graph are

close to one, you conclude that the car will be mostly available to drive over the next 10 years. You are interested in which components cause the most downtime for the system.

5. Click the red triangle next to Repairable Systems Simulation Results - Number and select **Box Plot of Total System Downtime by Component**.

Figure 12.4 Partial Box Plot of Total System Downtime by Component Report



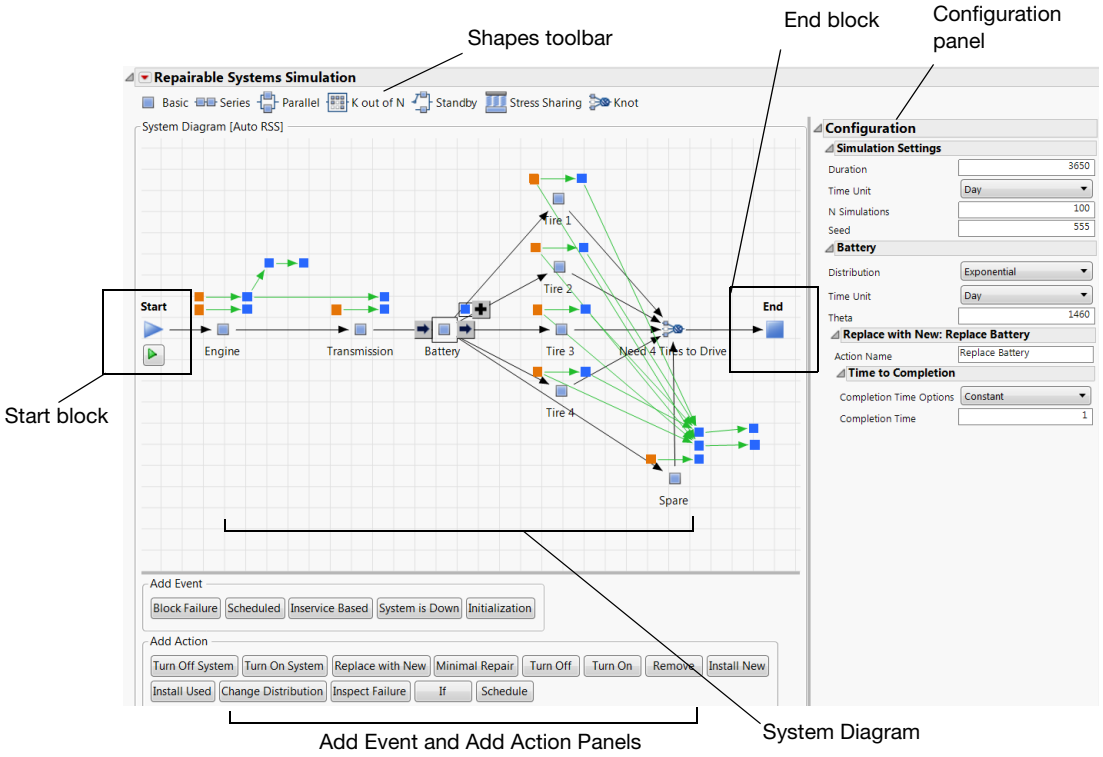
You conclude that the tires cause the most downtime for the car system. To increase the average total system availability, you might consider using more durable tires or carrying another spare tire.

JMP PRO The Repairable Systems Simulation Window

Launch the Repairable Systems Simulation platform by selecting **Analyze > Reliability and Survival > Repairable Systems Simulation**. The Repairable Systems Simulation window is divided into the following panels:

- "System Diagram"
- "Shapes Toolbar"
- "Configuration Panel"
- "Add Event Panel"
- "Add Action Panel"

Figure 12.5 The Repairable Systems Simulation Window



JMP PRO System Diagram

The System Diagram is the space where you can diagram a repairable system. A new System Diagram contains a Start block and an End block.

The following options are available in a pop-up menu when you right-click in the System Diagram:

Run Simulation Runs the simulation using the current Simulation Settings in the Configuration panel. The results of the simulation are reproducible if a nonzero value is specified for Seed.

Run Multithreaded Simulation Runs a multithreaded simulation of the system. The results of the simulation are not reproducible when you run a multithreaded simulation.

Diagram Operation Contains options to control the appearance of the diagram:

Delete (Available only if a block, event, or action is selected.) Removes the selected items from the diagram.

Show Block Names Shows or hides the names that appear below the blocks in the diagram.

Zoom In Increases the Zoom Scale value by a factor of 0.9.

Zoom Out Decreases the Zoom Scale by a factor of 0.9.

Tip: On Windows, you can press Ctrl and use the mouse scroll wheel to zoom in to and out from the diagram.

Zoom Scale Enables the user to set the scale of the zoom on a scale of 0 to 10000. The original Zoom Scale value is 1. The Zoom Scale value increases as you zoom into the diagram.

Preview Shows or hides the Preview window in the lower right corner of the System Diagram.

Show Event Action Link Across Blocks Shows or hides the green arrows that represent action links from one block to another. If this option is turned off, action links for selected events and actions still appear.

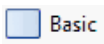
Align Selected Vertices Vertically (Available only if more than one block is selected.) Updates the horizontal position of the selected blocks so that they are aligned vertically.

Align Selected Vertices Horizontally (Available only if more than one block is selected.) Updates the vertical position of the selected blocks so that they are aligned horizontally.

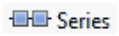
When you right-click an item in the System Diagram, additional submenu items appear in the pop-up menu. The submenu items that appear depend on the type of item. These options enable you to add, remove, or change the events and actions attached to a block item.

JMP PRO Shapes Toolbar

The Shapes toolbar contains the block shapes used to represent components in the System Diagram. Add block shapes to the System Diagram by clicking the icon for the block shape and dragging it onto the System Diagram. The Shapes toolbar includes the following block shape icons:

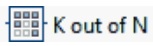


Adds a single block shape that represents a single component.

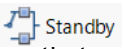


Adds a series block shape that represents a group of identical components that are connected in series. All of the components must be functional for the block to remain functional.

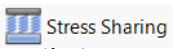
Adds a parallel block shape that represents a group of identical components that are connected in parallel. At least one of the components must be functional for the block to remain functional.



Adds a *k*-out-of-*n* block shape that represents a group of *n* identical components that are connected in parallel. At least *k* of the *n* components must be functional for the block to remain functional.



Adds a standby block shape that represents a group of *n* identical components that are connected in parallel. Only *k* components are *active*, or performing work in the system. The remaining *inactive* components act as standby components that are activated when any of the *k* active components fail. The block must have at least *k* functional and activated components for the block to remain functional.



Adds a stress sharing block shape that represents *n* identical components that are connected in parallel. Components fail one at a time, and remaining components fail at a quicker rate. The block must have at least one functional component and the block must successfully reallocate stress across remaining components for the block to remain functional.

Adds a knot block shape that represents a combination of the block shapes that point to the knot. At least a specified number of the block shapes that point to the knot must be functional for the knot to remain functional.

For information about the settings available for block shapes, see "Configuration Panel".

JMP PRO Configuration Panel

The Simulation Settings report appears at the top of the Configuration panel. The component settings for every selected component appear below the Simulation Settings outline. Configuration settings for selected events and actions appear below the block shape to which the action or event belongs.

JMP PRO Simulation Settings

Alter the settings for the simulation in the Simulation Settings report, which is available in the Configuration panel. The following settings are available:

Duration The length of time that is simulated in each iteration.

Time Unit The unit of time to use for the simulation.

Note: Recurring events and non-immediate actions use the Time Unit specified in Simulation Settings.

N Simulations The number of iterations in the simulation.

Seed (Optional) A random seed that ensures the reproducibility of simulation results. By default, the Seed is set to zero, which does not produce reproducible results. When you save the analysis to a script, the random seed that you enter is saved to the script.

Caution: If you run a simulation by right-clicking and selecting **Run Multithreaded Simulation**, the results are not reproducible, even if you specify a random seed.

JMP PRO Block Settings

Select a block shape in the System Diagram to see its settings in the Configuration panel.

Each block shape, except the knot block, has a Turn On System Exemption option. If this option is selected for a block, the block is not turned on when the Turn On System action is invoked.

Each block shape, except the knot block, has a failure distribution that determines the rate at which the block shape's individual components randomly fail. The failure distribution for a basic block determines the rate at which the block fails, because basic blocks represent only one component. For more information about the available failure distribution options, see "Distribution Options". Under the Distribution option, you can specify the Time Unit and distribution-dependent parameters for the block.

Series and Parallel

A series block fails when one of its components fails. A parallel block fails when all of its components fail. The following option is available for series and parallel blocks:

N Specifies the number of identical components contained in the block.

K-out-of-N

K-out-of-*N* blocks contain *n* identical components. The block fails when fewer than *k* of the components are functional. The following options are available for *K*-out-of-*N* blocks:

K Specifies the minimum number of functional components required for the block to remain functional.

N Specifies the number of identical components contained in the block.

Standby

Standby blocks have secondary components, called standby components, that are inactive. Active components perform work within a standby block. Inactive components do not perform work within a standby block, and are activated one at a time as active components fail. Occasionally, the activation process is not successful. A component switch might fail when activating a standby component. A standby block fails when less than k of its n identical components are active. The following options are available for standby blocks:

K Specifies the number of components that are initially active. This is also the minimum number of active components that is required for the block to remain functional.

N Specifies the total number of identical components within the block. The difference between k and n is equal to the number of standby components.

Switch Type Specifies the mechanism that activates a single standby component if any active component fails.

Single Switch A single switch exists in the block. If the activation of a standby component fails, then the block also fails.

Individual Switches A switch exists for each standby component. If the activation of a standby component fails, that standby component cannot be activated. The standby block attempts to activate the next standby component until a standby component is activated. If no remaining switches are functional and fewer than k of the components are active, then the block fails.

Switch Reliability Specifies the probability of success of activating a standby component when any active component fails.

Standby Type Specifies the state and failure distribution of the standby components.

Cold Standby components do not age until they are activated.

Warm Standby components age according to a secondary failure distribution while they are inactive. When standby components are activated, they age according to the primary failure distribution. Use the secondary failure distribution to mimic reduced stress on standby components that are not performing work in the standby block.

Stress Sharing

A stress sharing block distributes stress equally among its components. As components fail, the components that remain functional experience increased stress and subsequently fail at an increased rate.

N Specifies the total number of identical components contained in the block.

Switch Reliability Specifies the probability of successfully reallocating stress among the remaining functional components. The block fails if the reallocation of stress fails.

Stress Sharing Type Specifies how stress is shared among functional components.

Basic (default) Specifies that stress is shared equally among the remaining functional components. This type of stress sharing is referred to as *Load Sharing*. The characteristic life of individual components is proportional to the number of components that share the work load.

Custom Specifies that components share stress according to the JSL code that appears in the Sharing Formula option. The Sharing Formula defines how stress changes when components fail.

Knot

Knot blocks do not have a failure distribution. Knot blocks fail only if the number of connected blocks shapes that are functional falls below the specified minimum number. The following option is available for knot blocks:

Minimum Available Specifies the minimum number of functional blocks that must point to the knot block for the knot block to remain functional.

JMP PRO Add Event Panel

Events represent discrete occurrences in a simulation. You can use events to trigger actions. There is no limit to how many actions a single event can trigger. The following options are available in the Add Event Panel:

Note: Some block shapes do not have all of the available event types.

Block Failure Occurs when the block fails. If the component pathway is interrupted, the system is set to the Down state.

Scheduled Occurs at a specified recurring interval. The Max Occurrence option sets a limit on the number of occurrences of this event in a simulation. By default, Max Occurrence is set to missing. In this case, the event continues to occur until the end of the simulation. See "Event Settings".

Inservice Based Occurs when the component reaches a specified age. A component's age is the cumulative time that it has been functional and active. Age stops increasing when a component fails or its block is turned off. The Max Occurrence sets a limit on how many times this event can take place in a simulation. By default, the event continues to occur until the end of the simulation. See "Event Settings".

System is Down Occurs when the system is set to either the Down or the Off state. The system state can be changed intentionally or unintentionally.

Initialization Occurs when each simulation iteration begins. Use this event to arrange actions that have to complete before the system runs.

Create an Event

1. To create an event, select a block shape in the System Diagram to display the Add Event and Add Action panels at the bottom of the System Diagram.
2. Select one of the options in the Add Event panel to define an event for the selected block shape.

Figure 12.6 Block Failure Event

An orange square that represents the new event appears above the component. Notice that a connection arrow appears on the right side of the selected event.

JMP PRO Add Action Panel

An action defines component and system behavior that is triggered by either a connected event or a connected action. Connected actions are triggered upon completion of the prior action. The following options are available in the Add Action Panel:

Note: Some block shapes do not have all of the available action types.

Table 12.1 Action Options

Action Name	Blocks that can use this Action	Starting behavior	Completion Behavior
Turn off System	All		All blocks are turned off. The system is set to the Off state if it was not already in the Down state.

Table 12.1 Action Options *(Continued)*

Action Name	Blocks that can use this Action	Starting behavior	Completion Behavior
Turn on System	All		All blocks that were not already failed are turned on. The system is set to the On state if there is an uninterrupted component pathway. Blocks with the Turn On System Exemption option selected are not turned on by this action.
Replace with New	All	Turns off the original block, and sets the system to the Down state if the component pathway is interrupted.	The block becomes new. Its age is reset and the block is turned on. The system is set to the On state if there is an uninterrupted component pathway.
Minimal Repair	Basic and Series	Equivalent to starting behavior for the Replace with New action.	Turns the block on.The system is set to the On state if there is an uninterrupted component pathway. The age of the block is not reset.
Turn On Block	All	Cancels the action if the block is in the Down state or is currently removed from the system.	Turns the block on increments Turn On Count by one.
Turn Off Block	All		Turns the block off. The system is set to the Down state if the component pathway is interrupted.
Remove Block	All		The block is removed from the system. The system is set to the Down state if the component pathway is interrupted.

Table 12.1 Action Options *(Continued)*

Action Name	Blocks that can use this Action	Starting behavior	Completion Behavior
Install New	All	Equivalent to starting behavior for Replace with New action.	Turns the block off. The age of the block is reset.
Install Used	Basic	Equivalent to starting behavior for Replace with New action.	Turns the block off. You specify a new age and failure distribution for the block.
Change Distribution	Basic	Turns the block off. The system is set to the Down state if the component pathway is interrupted.	Changes the failure distribution of the block. Use this to mimic operation changes over time or cumulative damage to the block.
Inspect Failure	All		Triggers connected actions if the block has failed or is removed from the system.
If	All		Triggers connected actions if the specified condition script is true.
Schedule	All		Triggers connected actions at a specified interval. You can limit the number of scheduled intervals or allow the action to continue through the end of the simulation iteration.

Create an Action

1. To create an action, select a block shape in the System Diagram to display the Add Event and Add Action panels at the bottom of the System Diagram.
2. Select one of the options from the Add Action panel to define an action for the selected block shape.

A blue action square is created above the selected block shape. Actions are triggered when a connected event occurs. Create an event that triggers the action that you defined.

3. Select one of the options from the Add Event panel to define an event for the selected block shape.

   An orange event square is created above the block shape. Notice the connection arrow on the right side of the event.

4. Click the connection arrow and drag it to the blue action square that you created in step 2.

   A green arrow connects the event and action squares. When the event occurs in a simulation iteration, it triggers the connected action.

5. Select the blue action square that you created in step 2.

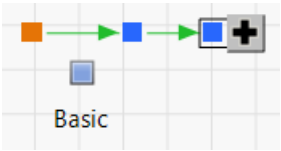
   You can connect additional actions that are triggered upon completion of the previous action by using the addition sign that appears on the right side of the selected action.

6. Click the addition sign and drag it to an empty area in the System Diagram.

   A list of the available actions appears.

7. Select one of the options from the action list to create an action that is connected to the first action.

Figure 12.7 Create an Action



A green arrow connects the two actions. The second action is triggered upon completion of the first action.

JMP PRO Repairable Systems Simulation Platform Options

The Repairable Systems Simulation red triangle menu contains the following options:

Save and Save As Enables you to save a Repairable Systems Simulation to a JMP Scripting Language (JSL) script that is automatically executed when it is opened in JMP. See the *Scripting Guide* for more information about Auto-Submit scripts.

Note: The Save and Save As red triangle options are equivalent to selecting **File > Save** and **File > Save As**, respectively. They are available in the red triangle menu for convenience.

Import Component Distribution Settings Enables you to import configuration settings for the system diagram from a data table. The table must contain columns for the component name, distribution, and one or more parameters. The number of parameters depends on the specified distribution.

Note: The strings in the imported table must be exact matches to the strings in the system diagram.

JMP PRO Options for Block Items

To view settings for the components in a selected design or subsystem, do the following:

- To view the Configuration settings for a block shape, select the block shape in the System Diagram.
- To view Configuration settings for more than one block shape, select multiple block shapes using the Arrow tool, press Ctrl, and click multiple block shapes.

JMP PRO Distribution Options

Each block shape randomly fails according to its specified failure distribution. In addition, you specify a time unit for the distribution. The block shape's time unit can be different from the time unit option that you specify in the simulation settings. The Turn On Count option is based on the number of times the individual block shape has been set to the On state.

The available failure distributions are listed in Table 12.2.

Table 12.2 Distributions and Additional Parameters

Property Type	Required Inputs
Exponential	Theta
Weibull	Alpha, Beta
Lognormal	location, scale
Loglogistic	location, scale
Fréchet	location, scale
GenGamma	mu, sigma, lambda
DS Weibull	Alpha, Beta, Defective Probability

Table 12.2 Distributions and Additional Parameters *(Continued)*

Property Type	Required Inputs
DS Lognormal	location, scale, Defective Probability
DS Loglogistic	location, scale, Defective Probability
DS Fréchet	location, scale, Defective Probability
Nonparametric	data or data file
Estimated	estimated distribution, data, or data file

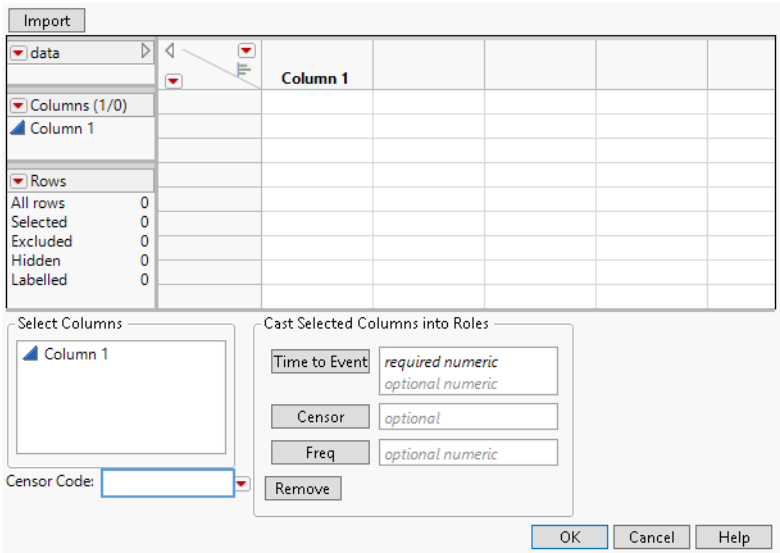
To see the formulas and parameterization for these failure distributions, see "Distributions".

JMP PRO Specify a Nonparametric or Estimated Distribution

The Nonparametric and Estimated options under Distribution enable you to approximate an arbitrary distribution. You can enter data manually or import a file that contains data. These data are used to approximate the distribution.

After selecting Nonparametric or Estimated, click the ... icon next to Data. The Provide Data window appears, which enables you to either enter data or import a data file. After you have imported or entered your data, the data are used to calculate a distribution for the component.

Figure 12.8 Provide Data Window



To import data from a file, do the following:

1. Before clicking the icon, open the JMP data table that contains the data to import.
2. Click the icon next to Data.
3. In the Provide Data window, click **Import**.

   The Select Data Table window appears.
4. From the Data Table list, select a data table.
5. Click **OK**.
6. In the panel beneath the data grid, specify the columns that represent Time to Event data.
7. Specify Censor and Freq columns, if appropriate.
8. In the Provide Data window, click **OK**.

To enter data manually, do the following:

1. Create columns for Time to Event data.
2. Create columns for Censor and Freq data, if appropriate.
3. Enter the data in the columns.
4. In the panel beneath the data grid, specify which columns represent Time to Event data, as well as Censor and Freq data if appropriate. See step 5 and step 6 above.
5. In the Provide Data window, click **OK**.

JMP PRO Event Settings

Select an event in the System Diagram to see its settings in the Configuration panel. You can change the Event Name setting to distinguish the event from others. Only the Scheduled and Inservice events have additional settings, which are listed here:

Scheduled

Scheduled events have the following options:

Recurring Interval Specifies the length of the time interval between event occurrences.

Max Occurrences Specifies the maximum number of times this event can occur.

Inservice

Inservice events have the following options:

Inservice Specifies the interval of component run time to wait before this event occurs.

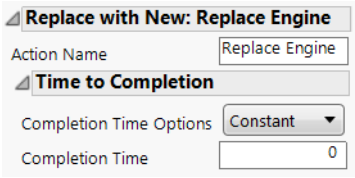
Note: Component run time accumulates only when the system is in the On state and the component is functional.

Max Occurrences Specifies the maximum number of times this event can occur.

JMP PRO Action Settings

Select an action in the System Diagram to see its settings in the Configuration panel. You can change the Action Name setting to distinguish the action from others. By default, actions are completed immediately. Non-immediate completion time options are available in the Completion Time Options menu.

Figure 12.9 Action Settings



The following options appear in the Completion Time Options list:

Immediate (default) Specifies that no time passes between starting and completion behavior.

Constant Specifies that the time lapse is always the specified Completion Time.

Choice Specifies that the time lapse is randomly chosen from the specified list of comma-separated values.

Uniform Specifies that the time lapse is randomly chosen from a uniform distribution with the specified Minimum and Maximum.

Triangle Specifies that the time lapse is randomly chosen from a triangular distribution with the specified Minimum, Mode, and Maximum.

Normal Specifies that the time lapse is randomly chosen from a normal distribution with the specified Mean and Standard Deviation.

JMP PRO Repairable Systems Simulation Results

- "Results Table"
- "Results Explorer"
- "Repairable Systems Simulation Results Options"
- "RSS Results Explorer Point Estimation Profilers"

JMP PRO Results Table

When you click the green arrow under the Start block to run the simulation, the results of the simulation appear in a data table. This table describes the events and subsequent actions that occurred in each simulation iteration.

The results table contains the following columns:

Sim ID Identifies the simulation iteration to which the event or action belongs.

Time Gives the exact time that the event or action took place in the simulation.

Subject Gives the name of the block shape to which the event or action is connected or System in the case of system events and actions.

Predicate Gives the name of the event or action that took place.

State Gives the state of the Subject at that exact Time. The initialization and termination of each action are denoted with Start and Finish, respectively.

Note Gives an additional description of an action. The end of each simulation iteration is denoted with End.

Figure 12.10 RSS Results Table

	Sim ID	Time	Subject	Predicate	State	Note
1	1	0	Spare	Initialization Spar...	Start	
2	1	0	Spare	Initialization Spar...	Off	
3	1	0	Spare	Initialization Spar...	Finish	
4	1	65.024584796	Tire 2	Tire 2 is unrepair...	Down	
5	1	65.024584796	System	Turn Down System	Down	Unintended
6	1	65.024584796	Engine	Turn Off Block by...	Off	
7	1	65.024584796	Transmission	Turn Off Block by...	Off	
8	1	65.024584796	Battery	Turn Off Block by...	Off	
9	1	65.024584796	Tire 1	Turn Off Block by...	Off	
10	1	65.024584796	Tire 3	Turn Off Block by...	Off	
11	1	65.024584796	Tire 4	Turn Off Block by...	Off	
12	1	65.024584796	Tire 2	Replace Tire 2	Removed	
13	1	65.024584796	Tire 2	Replace Tire 2	Start	

Notice that the first entry in Figure 12.10 is the start of the Initialization Spare in Trunk action, which has an immediate completion time. The action turns the Spare component off and then finishes while Time is still 0.

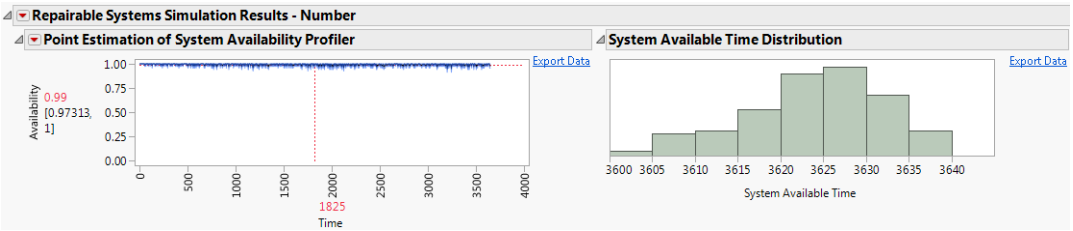
The next event that occurs is Tire 2 is Unrepairable when approximately 65 days have passed in the first simulation iteration. In row five, the system is unintentionally set to the Down state because only three tires are functional. The Need 4 Tires to Drive knot requires at least four tires, including the Spare, to be functional. The Tire 2 is Unrepairable event simultaneously triggers the Replace Tire 2 action and the Use Spare action. The Drive with Spare action is triggered in row eleven, and the system is set to the On state.

Notice that Tire 2 failed and was replaced by the Spare at the same time. No time elapsed between the time that the system was set to the Down state and the time that the system was returned to the On state. Because the subsequent actions had an immediate completion time, the Tire 2 is Unrepairable event did not cause any system outage time. The results explorer shows the system outage time that accumulates when actions are non-immediate.

JMP PRO Results Explorer

When you click the green arrow next to the Launch Repairable Systems Simulation Results Explorer script in the results data table, a report appears that contains an analysis of the simulation results.

Figure 12.11 Partial RSS Explorer Report



By default, the report contains two types of graphs. The first type of graph, shown on the left side of Figure 12.11, is a point estimate graph. The point estimate graph displays aggregated simulation results on the vertical axis plotted against simulation iteration time on the horizontal axis. The range of the horizontal axis is the duration that you specify in the simulation settings. The estimated probability of the system being available at a given time is shown in red next to the vertical axis. Below the point estimate is a 95% confidence interval for the point estimate.

Tip: To see the point estimation of system availability at a specific time in the simulation iteration, click the red number below the horizontal axis. Specify a time within the range of the simulation duration and press Enter.

The second type of graph, shown on the right side of Figure 12.11, is a histogram of the system available time from the simulation iterations. The bins in the histograms are representative of one of the following:

- Total simulation time.
- Proportion of the simulation duration.

Tip: Select the **Export Data** option beside a graph to create a data table with the data that the graph uses.

By default, the report contains the following graphs:

Point Estimation of System Availability Profiler Shows a profiler graph over time of the estimated probability that the system is in the On or the Off state during a simulation iteration.

System Available Time Distribution Shows a histogram of the total time that the system was available for each simulation iteration.

System Availability Distribution Shows a histogram of the total time that the system was available for each simulation iteration divided by the duration of the simulation. This distribution is the same as the System Available Time Distribution, except that it is shown as a proportion of the duration of the simulation.

Point Estimation of System In Service Probability Profiler Shows a profiler graph over time of the estimated probability that the system is in the On state during a simulation iteration.

System In Service Time Distribution Shows a histogram of the total time that the system was in the On state for each simulation iteration.

System In Service Probability Distribution Shows a histogram of the total time that the system was in the On state for each simulation iteration divided by the duration of the simulation. This distribution is the same as System In Service Time Distribution, except that the distribution is shown as a proportion of the duration of the simulation.

Point Estimation of System Unplanned Outage Profiler Shows a profiler graph over time of the estimated probability that the system is in the Down state during a simulation iteration.

System Unplanned Outage Distribution Shows a histogram of the total time that the system was in the Down state for each simulation iteration.

System Unplanned Outage Percentage Distribution Shows a histogram of the total time that the system was in the Down state for each simulation iteration divided by the duration of the simulation. This distribution is the same as System Unplanned Outage Distribution, except that the distribution is shown as a proportion of the duration of the simulation.

JMP PRO Repairable Systems Simulation Results Options

The Repairable Systems Simulation Results red triangle menu contains the following options:

Point Estimation of Component Availability For each block shape in the system diagram, this option shows the following reports:

- Point Estimation of Availability Profiler
- Available Time Distribution
- Availability Distribution
- Point Estimation of Unplanned Outage Profiler
- Unplanned Outage Time Distribution
- Unplanned Outage Distribution

System Availability by Period Shows a graph that enables you to examine the average system availability during specified time periods. Use the By Time panel to define the time period as an interval of time. Use the By Event panel to define the time period as a specified number of occurrences of one event. When you click **Go**, the report shows a bar chart of the average system availability during each period.

Box Plot of System Total Downtime by Component Shows a box plot of the total time the system was in the Down or Off states caused by individual components. The components are ordered by contribution from the most outage time to the least outage time, similar to a Pareto chart.

Box Plot of System Unplanned Total Outage Time by Component Shows a box plot of total time the system was in the Down state caused by individual components. The components are ordered by contribution from the most outage time to the least outage time, similar to a Pareto chart.

See *Using JMP* for more information about the following options:

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

JMP PRO RSS Results Explorer Point Estimation Profilers

Each point estimation profiler within the Repairable Systems Simulation Results Explorer report has a red triangle menu that contains the following options:

Confidence Intervals Shows or hides the 95% confidence intervals on the point estimation graphs.

Reset Factor Grid Opens the Factor Settings window, which enables you to modify the parameters of the point estimation graph. See *Profilers* for more information about setting the factor grid.

Factor Settings Provides additional options affecting the Factor Grid. See *Profilers* for more information about Factor Settings.

Chapter 13

Survival Analysis

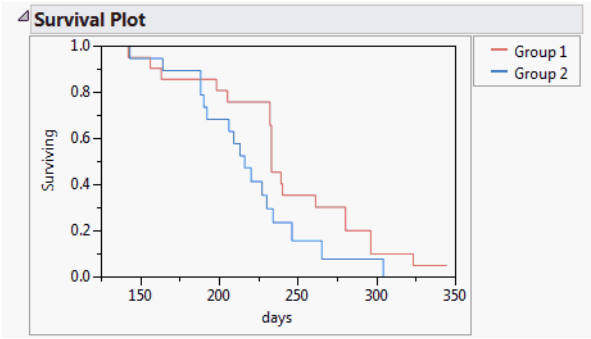
Analyze Survival Time Data

Survival data contain duration times until the occurrence of a specific event and are sometimes referred to as *event-time* response data. The event is usually failure, such as the failure of an engine or death of a patient. If the event does not occur before the end of a study for an observation, the observation is said to be *censored*.

The Survival platform fits a single Y that represents time to event (or time to failure). Use the Survival platform to examine the distribution of the failure times.

Tip: To fit explanatory models, use the Fit Parametric Survival platform or the Fit Proportional Hazards platform.

Figure 13.1 Example of a Survival Plot



Contents

Overview of the Survival Analysis Platform . . . 345
Example of Survival Analysis . . . 346
Launch the Survival Platform . . . 348
The Survival Plot . . . 349
Survival Platform Options . . . 350
Exponential, Weibull, and Lognormal Plots and Fits . . . 352
Fitted Distribution Plots . . . 356
Competing Causes . . . 357
Additional Examples of the Survival Platform . . . 358
Example of Survival Analysis . . . 358
Example of Competing Causes . . . 361
Example of Interval Censoring . . . 363
Statistical Reports for Survival Analysis . . . 365

Overview of the Survival Analysis Platform

Survival data need to be analyzed with specialized methods for two reasons:

1. The survival times usually have specialized non-normal distributions, like the exponential, Weibull, and lognormal.
2. Some of the data could be censored.

Survival functions are calculated using the nonparametric Kaplan-Meier method for one or more groups of either *complete* or *right-censored* data. Complete data have no censored values. Right-censoring is when you do not know the exact survival time, but you know that it is greater than the specified value. Right-censoring occurs when the study ends without all the units failing, or when a patient has to leave the study before it is finished. The censored observations cannot be ignored without biasing the analysis. The elements of a survival model are:

- A time indicating how long until the unit (or patient) either experienced the event or was censored. Time is the model response (*Y*).
- A censoring indicator that denotes whether an observation experienced the event or was censored. JMP uses the convention that the code for a censored unit is 1 and the code for a non-censored event is zero.
- Explanatory variables (if a regression model is used.)
- Interval censoring is when a data point is somewhere on an interval between two values. If interval censoring is needed, then two Y variables hold the lower and upper limits bounding the event time.

Common terms used for reliability and survival data include lifetime, life, survival, failure-time, time-to-event, and duration.

The Survival platform computes product-limit (Kaplan-Meier) survival estimates for one or more groups. It can be used as a complete analysis or is useful as an exploratory analysis to gain information for more complex model fitting. The Kaplan-Meier Survival platform does the following:

- Shows a plot of the estimated survival function for each group. A plot for the whole sample is optional.
- Calculates and lists survival function estimates for each group and for the combined sample.
- Shows exponential, Weibull, and lognormal diagnostic failure plots to graphically check the appropriateness of using these distributions for further regression modeling. Parameter estimates are available on request.
- Computes the Log Rank and generalized Wilcoxon Chi-square statistics to test homogeneity of the estimated survival function across groups.

- Analyzes competing causes, prompting for a cause of failure variable, and estimating a Weibull failure time distribution for censoring patterns corresponding to each cause.

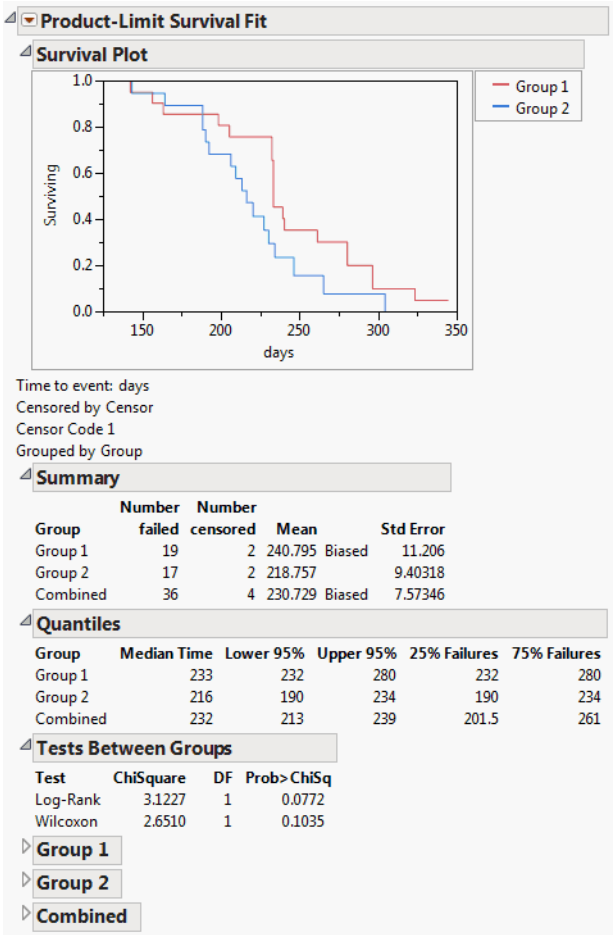
Example of Survival Analysis

An experiment was undertaken to characterize the survival time of rats exposed to a carcinogen in two treatment groups. The data are in the Rats.jmp sample data table. The event in this example is death. The objective is to see whether rats in one treatment group live longer (more days) than rats in the other treatment group.

1. Select **Help > Sample Data Library** and open Rats.jmp.

   The data in the days column is the survival time. Notice that some observations are censored.
2. Select **Analyze > Reliability and Survival > Survival**.
3. Select days and click **Y, Time to Event**.
4. Select Group and click **Grouping**.
5. Select Censor and click **Censor**.
6. Click **OK**.

Figure 13.2 Survival Plot for Rats.jmp Data

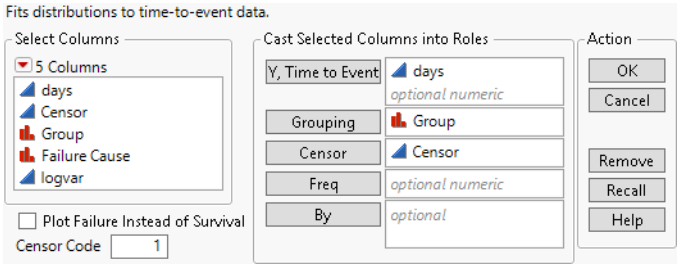


It appears that the rats in treatment group 1 are living longer than the rats in treatment group 2.

Launch the Survival Platform

Launch the Survival platform by selecting **Analyze > Reliability and Survival > Survival**.

Figure 13.3 The Survival Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

The Survival launch window contains the following options:

Y, Time to Event Identifies the time to event or time to censoring. If you have interval censoring, specify two Y variables, representing the lower and upper limits.

Grouping Classifies the data into groups that are fit separately.

Censor Identifies censored values. Enter the value that identifies censoring in the Censor Code box. This column can contain more than two distinct values under the following conditions:

- All censored rows contain the value that is entered in the Censor Code box.
- Non-censored rows have a value other than what is in the Censor Code box.

Freq Indicates the column whose values are the frequencies of observations for each row when there are multiple units recorded. If the value is 0 or a positive integer, then the value represents the frequencies or counts of observations for each row.

By Performs a separate analysis for each level of a classification or grouping variable.

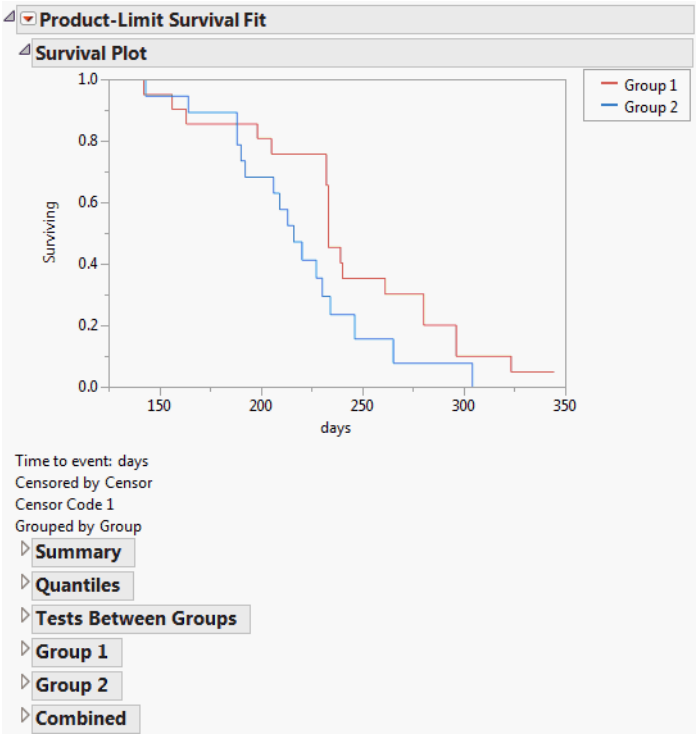
Plot Failure Instead of Survival Shows a failure probability plot instead of its reverse (a survival probability plot).

Censor Code Identifies the value in the Censor column that designates right-censored observations. The default value is 1.

The Survival Plot

The Survival platform shows overlay step plots of estimated survival functions for each group. A legend identifies groups by color and line type.

Figure 13.4 The Survival Plot



Reports beneath the plot show summary statistics and quantiles for survival times. Estimated survival times for each observation are computed within groups. Survival times are computed from the combined sample. When there is more than one group, statistical tests compare the survival curves.

If there are any failures that occur at time zero, the Failures at Time Zero report appears when you request a distribution fit from the Product-Limit Survival Fit red triangle menu. This report contains a table of counts of zero-time failures for each level of the Grouping variable. If there is no Grouping variable specified, there is one row in the table labeled Combined. The table also contains a column labeled Prob Time>0. This column is the proportion of observations in each group that has a nonzero failure time.

Survival Platform Options

The Product-Limit Survival Fit red triangle menu contains the following options:

Survival Plot Shows the overlaid survival plots for each group.

Failure Plot Shows the overlaid failure plots (proportion failing over time) for each group (in the tradition of the reliability literature.) A failure plot reverses the vertical axis to show the number of failures rather than the number of survivors.

Note: The Failure Plot option replaces the Reverse Y Axis option found in older versions of JMP (which is still available in scripts).

Plot Options Contains the following options:

Note: The first seven options (Show Points, Show Kaplan Meier, Show Combined, Show Confid Interval, Show Simultaneous CI, Show Shaded Pointwise CI, and Show Shaded Simultaneous CI) and the last two options (Fitted Survival CI, Fitted Failure CI) pertain to the initial survival plot and failure plot. The other five (Midstep Quantile Points, Connect Quantile Points, Fitted Quantile, Fitted Quantile CI Lines, Fitted Quantile CI Shaded) pertain only to the distributional plots.

Show Points Shows the sample points at each step of the survival plot. Failures appear at the bottom of the steps, and censorings are indicated by points above the steps.

Show Kaplan Meier Shows the Kaplan-Meier curves. This option is on by default.

Show Combined Shows the survival curve for the combined groups in the Survival Plot.

Show Confid Interval Shows the pointwise 95% confidence bands on the survival plot for groups and for the combined plot when it appears with the Show Combined option.

Show Points, Show Combined When you select the Show Points and Show Combined options, the survival plot for the total or combined sample appears as a gray line. The points also appear at the plot steps of each group.

Show Simultaneous CI Shows the simultaneous confidence bands for all groups on the plot. Meeker and Escobar (1998, ch. 3) discuss pointwise and simultaneous confidence intervals and the motivation for simultaneous confidence intervals in survival analysis.

Midstep Quantile Points Changes the plotting positions to use the *modified Kaplan-Meier* plotting positions. These plotting positions are equivalent to taking mid-step positions of the Kaplan-Meier curve, rather than the bottom-of-step positions. This option is recommended, so it is on by default.

Connect Quantile Points Shows the lines in the plot. This option is on by default.

Fitted Quantile Shows the straight-line fit on the fitted Weibull, lognormal, or exponential plot. This option is on by default.

Fitted Quantile CI Lines Shows the 95% confidence bands for the fitted Weibull, lognormal, or exponential plot.

Fitted Quantile CI Shaded Shows the display of the 95% confidence bands for a fit as a shaded area or dashed lines.

Fitted Survival CI Shows the confidence intervals (on the survival plot) of the fitted distribution.

Fitted Failure CI Shows the confidence intervals (on the failure plot) of the fitted distribution.

Exponential Plot Plots the cumulative exponential failure probability by time for each group. Lines that are approximately linear empirically indicate the appropriateness of using an exponential model for further analysis. For example, in Figure 13.5, the lines for Group 1 and Group 2 in the Exponential Plot are curved rather than straight. This indicates that the exponential distribution is not appropriate for this data. See "Exponential, Weibull, and Lognormal Plots and Fits".

Exponential Fit Produces the Exponential Parameters table and the linear fit to the exponential cumulative distribution function in the Exponential Plot (Figure 13.5). The parameter Theta corresponds to the mean failure time. See "Exponential, Weibull, and Lognormal Plots and Fits".

Weibull Plot Plots the cumulative Weibull failure probability by log(time) for each group. A Weibull plot that has approximately parallel and straight lines indicates a Weibull survival distribution model might be appropriate to use for further analysis. See "Exponential, Weibull, and Lognormal Plots and Fits".

Weibull Fit Produces the linear fit to the Weibull cumulative distribution function in the Weibull plot and two popular forms of Weibull estimates. These estimates are shown in the Extreme-value Parameter Estimates table and the Weibull Parameter Estimates tables (Figure 13.5). The Alpha parameter is the 0.632 quantile of the failure-time distribution. The Extreme-value table shows a different parameterization of the same fit, where Lambda = ln(Alpha) and Delta = 1/Beta. See "Exponential, Weibull, and Lognormal Plots and Fits".

LogNormal Plot Plots the cumulative lognormal failure probability by log(time) for each group. A lognormal plot that has approximately parallel and straight lines indicates a lognormal distribution is appropriate to use for further analysis. See "Exponential, Weibull, and Lognormal Plots and Fits".

LogNormal Fit Produces the linear fit to the lognormal cumulative distribution function in the lognormal plot and the LogNormal Parameter Estimates table shown in Figure 13.5. Mu and Sigma correspond to the mean and standard deviation of a normally distributed natural logarithm of the time variable. See "Exponential, Weibull, and Lognormal Plots and Fits".

Fitted Distribution Plots Use in conjunction with the fit options to show three plots corresponding to the fitted distributions: Survival, Density, and Hazard. If you have not performed a fit, no plot appears. See "Fitted Distribution Plots".

Competing Causes Performs an estimation of the Weibull model using the specified causes to indicate a failure event and other causes to indicate censoring. The fitted distribution appears as a dashed line in the Survival Plot. See "Competing Causes".

Estimate Survival Probability Estimates survival probabilities for the time values that you specify.

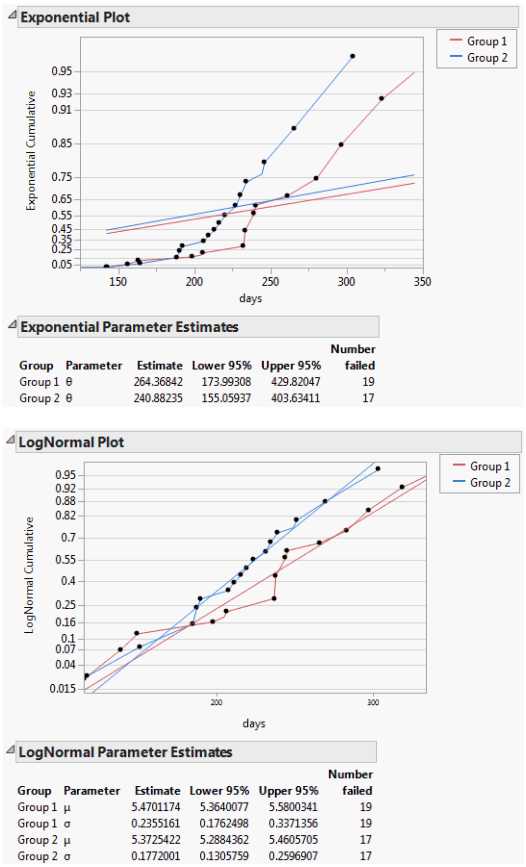
Estimate Time Quantile Estimates a time quantile for each survival probability that you specify.

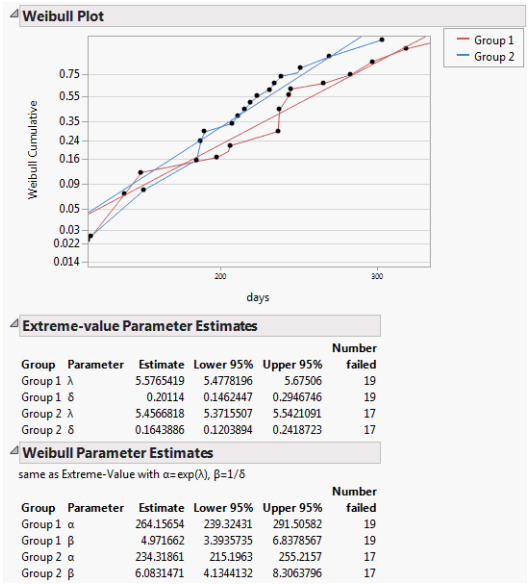
Save Estimates Creates a data table containing survival and failure estimates, along with confidence intervals, and other distribution statistics.

Exponential, Weibull, and Lognormal Plots and Fits

For each of the three supported distributions in the Survival platform, there is a plot command and a fit command. Use the plot command to see whether the event markers seem to follow a straight line. The markers tend to follow a straight line when the distributional fit is suitable for the data. Then, use the fit commands to estimate the parameters.

Figure 13.5 Exponential, Weibull, and Lognormal Plots and Reports





The following table shows what to plot to make a straight line fit for that distribution:

Table 13.1 Straight Line Fits for Distribution

Distribution Plot	Horizontal Axis	Vertical Axis	Interpretation
Exponential	time	-log(S)	slope is 1/theta
Weibull	log(time)	log(-log(S))	slope is beta
Lognormal	log(time)	Probit(1-S)	slope is 1/sigma

Note: S = product-limit estimate of the survival distribution.

Exponential

The exponential distribution is the simplest distribution for modeling time-to-event data. The exponential distribution has only one parameter, theta. It is a constant-hazard distribution, with no memory of how long it has survived to affect how likely an event is. The parameter theta is the expected lifetime.

Weibull

The Weibull distribution is the most popular distribution for modeling time-to-event data. The Weibull distribution can have two or three parameters. The Survival platform fits the two-parameter Weibull distribution. Authors parameterize this distribution in many different ways (Table 13.2). JMP reports two of these parameterizations: the Weibull alpha-beta parameterization and a parameterization based on the smallest extreme value distribution.

The alpha-beta parameterization, shown in the Weibull Parameter Estimates report, is widely used in the reliability literature (Nelson 1990). The *alpha* parameter is interpreted as the quantile at which 63.2% of the units fail. The *beta* parameter determines how the hazard rate changes over time. If *beta* > 1, the hazard rate increases over time; if *beta* < 1, the hazard rate decreases over time; and if *beta* = 1, the hazard rate is constant over time. A Weibull distribution with a constant hazard function is equivalent to an exponential distribution.

The lambda-delta extreme value parameterization is shown in the Extreme-Value Parameter Estimates report. This parameterization is sometimes desirable in a statistical sense because it places the Weibull distribution in a location-scale setting (Meeker and Escobar 1998, p. 86). The location parameter is *lambda*, and the scale parameter is *delta*. In relation to the alpha-beta parameterization, *lambda* is equal to the natural log of alpha, and *delta* is equal to the reciprocal of beta. Therefore, the delta parameter determines how the hazard rate changes over time. If *delta* > 1, the hazard rate decreases over time; if *delta* < 1, the hazard rate increases over time; and if *delta* = 1, the hazard rate is constant over time. A Weibull distribution with a constant hazard function is equivalent to an exponential distribution.

Table 13.2 Various Weibull Parameters in Terms of alpha and beta in JMP

JMP Weibull	*alpha*	*beta*
Wayne Nelson	alpha=*alpha*	beta=*beta*
Meeker and Escobar	eta=*alpha*	beta=*beta*
Tobias and Trindade	c = *alpha*	m = *beta*
Kececioglu	eta=*alpha*	beta=*beta*
Hosmer and Lemeshow	exp(X beta)=*alpha*	lambda=*beta*
Blishke and Murthy	beta=*alpha*	alpha=*beta*

Table 13.2 Various Weibull Parameters in Terms of alpha and beta in JMP *(Continued)*

Kalbfleisch and Prentice	lambda = 1/*alpha*	p = *beta*
JMP Extreme Value	lambda=log(*alpha*)	delta=1/*beta*
Meeker and Escobar s.e.v.	mu=log(*alpha*)	sigma=1/*beta*

Lognormal

The lognormal distribution is also very popular for modeling time-to-event data. The lognormal distribution is equivalent to the distribution where if you take the log of the values, the distribution is normal. If you want to fit a normal distribution to your data, you can take the exp() of it and model your data with a lognormal distribution. See "Additional Examples of Fitting Parametric Survival".

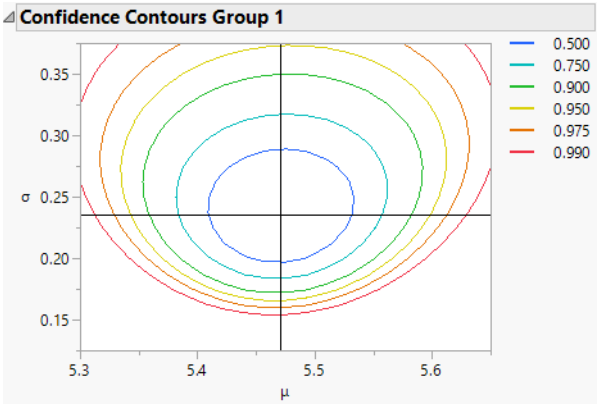
Additional Options

To see additional options for the exponential, Weibull, and lognormal fits, press Shift, click the red triangle next to Product-Limit Survival Fit, and select the desired fit.

Use these options to do the following tasks:

- Set the confidence level for the limits.
- Set the constrained value for theta (in the case of an exponential fit), sigma (in the case of a lognormal fit) or beta (in the case of a Weibull fit). See "WeiBayes Analysis".
- Obtain a Confidence Contour Plot for the Weibull and lognormal fits (when there are no constrained values).

Figure 13.6 Confidence Contour Plot



WeiBayes Analysis

JMP can constrain the values of the Theta (Exponential), Beta (Weibull), and Sigma (LogNormal) parameters when fitting these distributions. This feature is needed in *WeiBayes* situations, for example:

- Where there are few or no failures
- There are existing historical values for beta
- There is still a need to estimate alpha

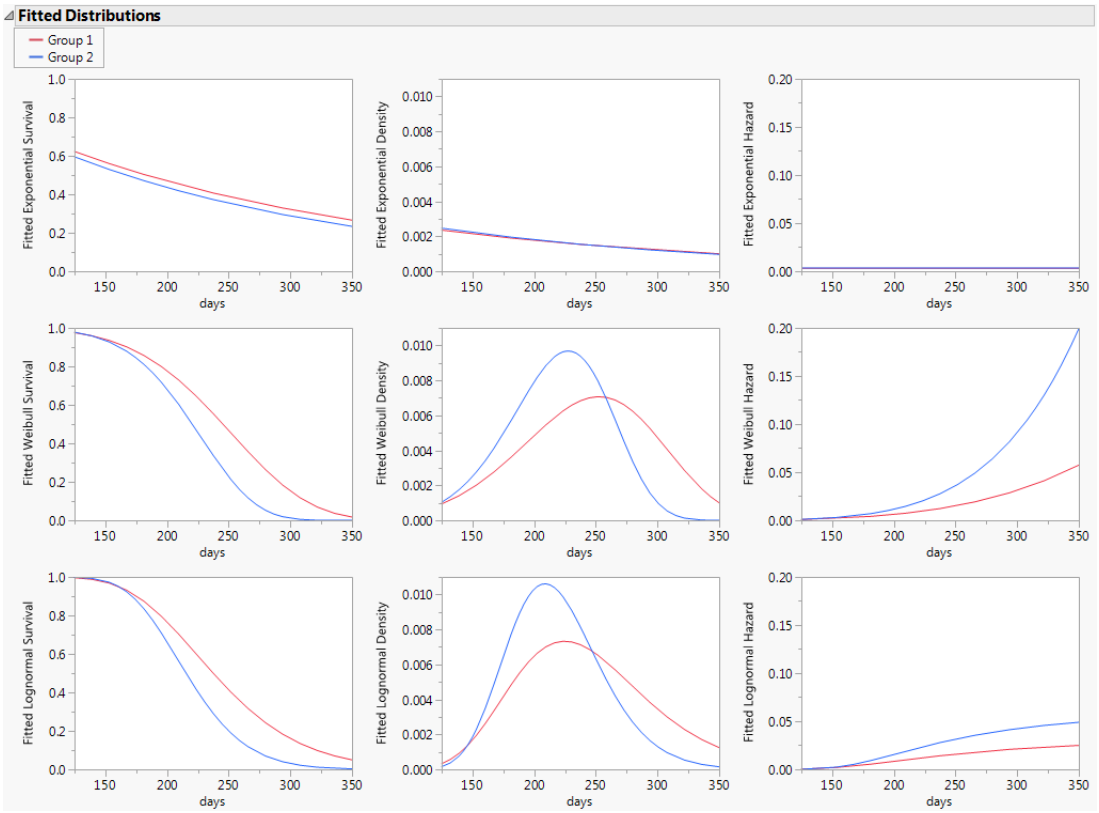
For more information about WeiBayes situations, see Abernethy (1996).

With no failures, the standard technique is to add a failure at the end. Then, the estimates reflect a type of lower bound on the alpha value, rather than a real estimate. However, the WeiBayes feature allows for a true estimation.

Fitted Distribution Plots

Use the Fitted Distribution Plots option to see Survival, Density, and Hazard plots for the exponential, Weibull, and lognormal distributions. The plots share the same axis scaling so that the distributions can be easily compared.

Figure 13.7 Fitted Distribution Plots for Three Distributions



These plots can be transferred to other graphs through the use of graphic scripts. To copy the graph, right-click in the plot to be copied and select **Edit > Copy Frame Contents**. Right-click in the destination plot and select **Edit > Paste Frame Contents**.

Competing Causes

Sometimes there are multiple causes of failure in a system. For example, suppose that a manufacturing process has several stages and the failure of any stage causes a failure of the whole system. If the different causes are independent, the failure times can be modeled by an estimation of the survival distribution for each cause. A censored estimation is undertaken for a given cause by treating all the event times that are not from that cause as censored observations.

The Competing Causes red triangle menu contains the following options:

Omit Causes Removes the specified cause value and recalculates the survival estimates.

Save Cause Coordinates Adds a new column to the current table called log(–log(Surv)). This information is often used to plot against the time variable with a grouping variable, such as the code for type of failure.

Weibull Lines Adds Weibull lines to the plot.

Hazard Plot Adds a Hazard Plot.

Simulate Creates a new data table containing time and cause information from the Weibull distribution, as estimated by the data.

Additional Examples of the Survival Platform

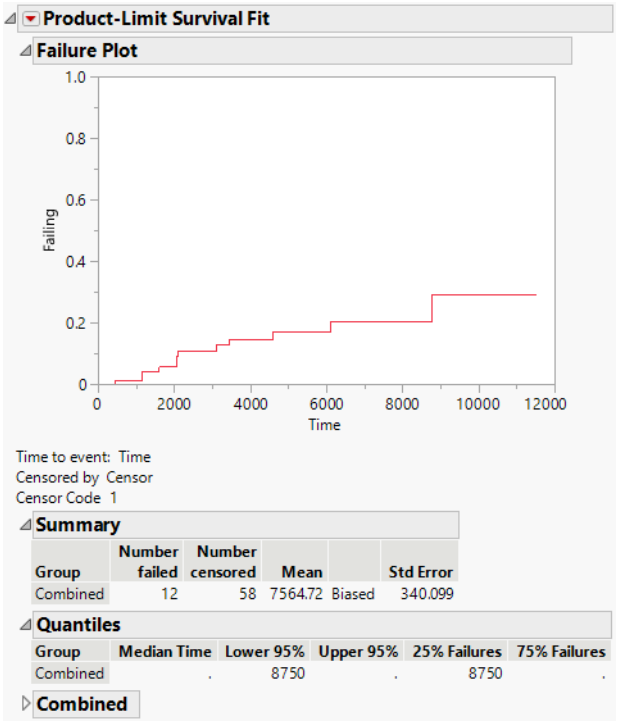
- "Example of Survival Analysis"
- "Example of Competing Causes"
- "Example of Interval Censoring"

Example of Survival Analysis

The failure of diesel generator fans was studied by Nelson (1982, p. 133) and Meeker and Escobar (1998, app. C1).

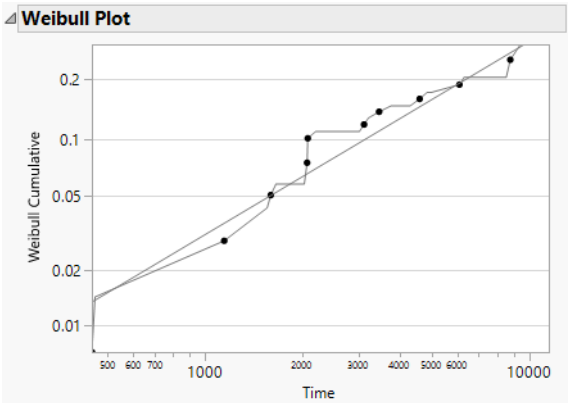
1. Select **Help > Sample Data Library** and open Reliability/Fan.jmp.
2. Select **Analyze > Reliability and Survival > Survival**.
3. Select Time and click **Y, Time to Event**.
4. Select Censor and click **Censor**.
5. Select the check box for **Plot Failure instead of Survival**.
6. Click **OK**.

Figure 13.8 Fan Initial Output



Notice that the probability of failure increases over time. Often the next step is to explore distributional fits, such as a Weibull model. Click the red triangle next to Product-Limit Survival Fit and select **Weibull Plot** and **Weibull Fit**.

Figure 13.9 Weibull Output for Fan Data



Extreme-Value Parameter Estimates

Parameter	Estimate	Lower 95%	Upper 95%	Number failed
λ	10.177204	9.5201194	11.572014	12
δ	0.9447814	0.6031578	1.6502353	12

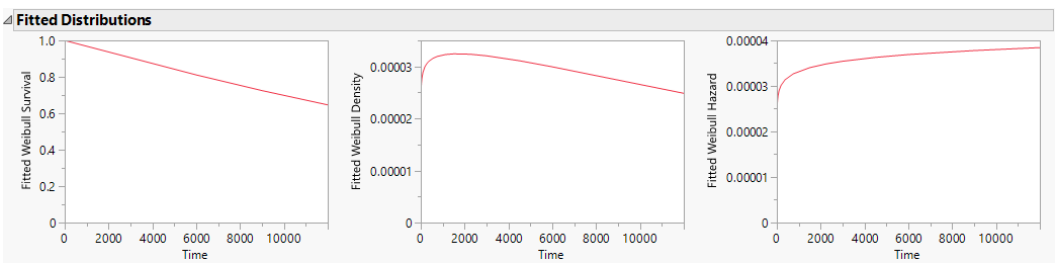
Weibull Parameter Estimates

same as Extreme-Value with α=exp(λ), β=1/δ

Parameter	Estimate	Lower 95%	Upper 95%	Number failed
α	26296.845	13631.239	106086.92	12
β	1.0584459	0.6059742	1.6579409	12

Because the fit is reasonable and the Beta estimate is near 1, you can conclude that this looks like an exponential distribution, which has a constant hazard rate. Click the red triangle next to Product-Limit Survival Fit and select **Fitted Distribution Plots**. Three views of the Weibull fit appear.

Figure 13.10 Fitted Distribution Plots

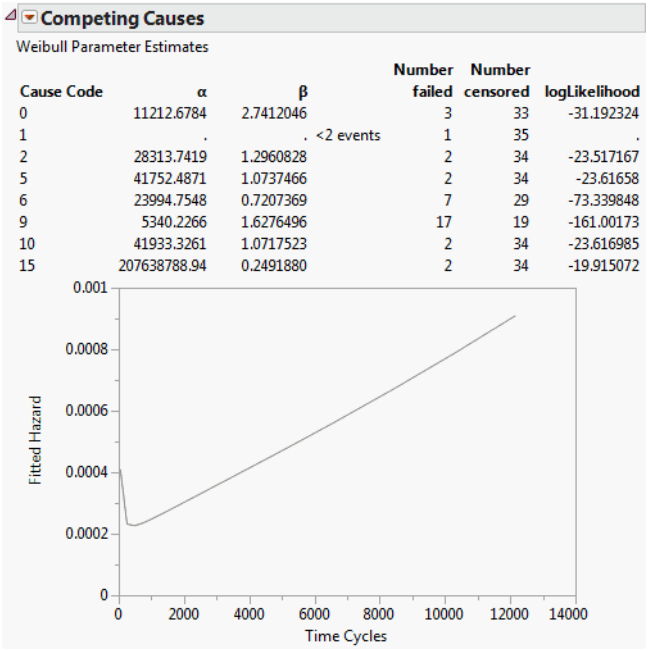


Example of Competing Causes

Nelson (1982) discusses the failure times of a small electrical appliance that has a number of causes of failure. One group (Group 2) of the data is represented in the JMP sample data table Appliance.jmp.

1. Select **Help > Sample Data Library** and open Reliability/Appliance.jmp.
2. Select **Analyze > Reliability and Survival > Survival**.
3. Select Time Cycles and click **Y, Time to Event**.
4. Click **OK**.
5. Click the red triangle next to Product-Limit Survival Fit and select **Competing Causes**.
6. Click Cause Code, and click **OK**.
7. Click the Competing Causes red triangle and select **Hazard Plot**.

Figure 13.11 Competing Causes Report and Hazard Plot

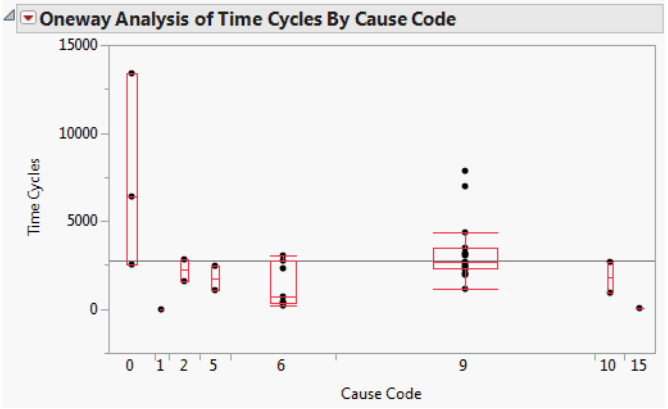


The survival distribution for the whole system is simply the product of the survival probabilities. The Competing Causes table shows the Weibull estimates of Alpha and Beta for each failure cause.

In this example, most of the failures were due to cause 9. Cause 1 occurred only once and could not produce good Weibull estimates. Cause 15 happened for very short times and resulted in a small beta and large alpha. Recall that alpha is the estimate of the 63.2% quantile of failure time, which means that causes with early failures often have very large alphas. If these causes do not result in early failures, then these causes do not usually cause later failures.

Figure 13.12 shows the Fit Y by X plot of Time Cycles by Cause Code with the Quantiles option in effect. This plot further illustrates how the alphas and betas relate to the failure distribution.

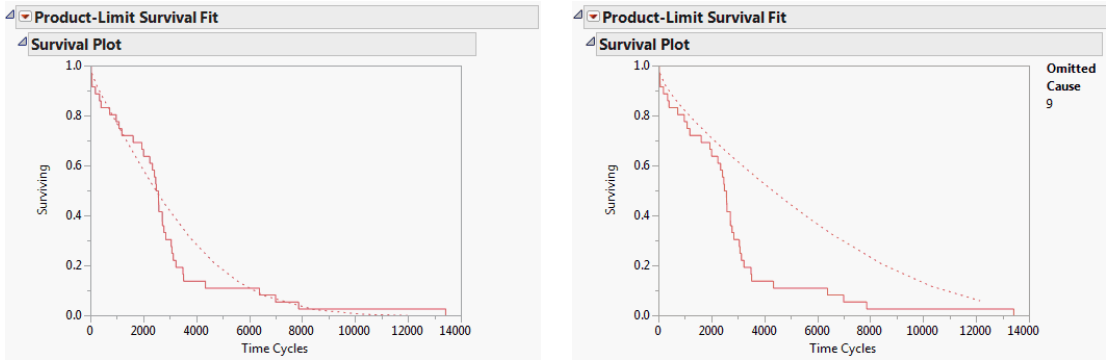
Figure 13.12 Fit Y by X Plot of Time Cycles by Cause Code



In this example, recall that cause 9 was the source of most of the failures. If cause 9 was corrected, how would that affect the survival due to the remaining causes? Select the **Omit Causes** option to remove a cause value and recalculate the survival estimates.

Figure 13.13 shows the survival plots with all competing causes and without cause 9. You can see that the survival rate (represented by the dashed line) without cause 9 does not improve much until 2,000 cycles. It then becomes much better and remains improved, even after 10,000 cycles.

Figure 13.13 Survival Plots with Omitted Causes



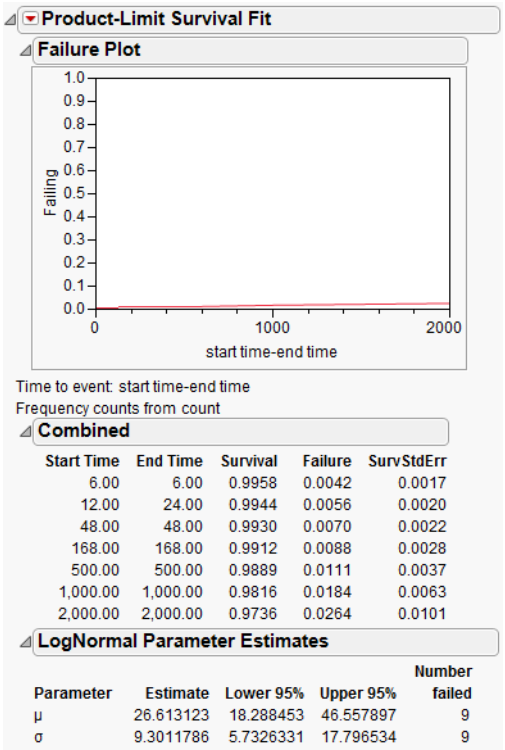
Example of Interval Censoring

With interval censored data, you know only that the events occurred in some time interval. The Turnbull method is used to obtain nonparametric estimates of the survival function.

In this example from Nelson (1990, p. 147), microprocessor units are tested and inspected at various times and the failed units are counted. Missing values in one of the columns indicate that you do not know the lower or upper limit, and therefore the event is left or right censored, respectively.

1. Select **Help > Sample Data Library** and open Reliability/Microprocessor Data.jmp.
2. Select **Analyze > Reliability and Survival > Survival**.
3. Select start time and end time and click **Y, Time to Event**.
4. Select count and click **Freq**.
5. Select the check box next to **Plot Failure instead of Survival**.
6. Click **OK**.
7. Click the red triangle next to Product-Limit Survival Fit and select **LogNormal Fit**.

Figure 13.14 Interval Censoring Output



The resulting Turnbull estimates are shown. Turnbull estimates might have gaps in time where the survival probability is not estimable. In this example, such gaps occur between 6 and 12, 24 and 48, 48 and 168, and so on.

At this point, select a distribution to see its fitted estimates —in this case, a Lognormal distribution is fit. Notice that the failure plot shows very small failure rates for these data.

Statistical Reports for Survival Analysis

For data that is not interval censored, the initial reports show Summary and Quantiles data (Figure 13.15). The Summary data shows the number of failed and number of censored observations for each group (when there are groups) and for the whole study. The mean and standard deviations are also adjusted for censoring. For computational details about these statistics, see the LIFETEST Procedure chapter in SAS Institute Inc. (2020).

The Quantiles data shows time to failure statistics for individual and combined groups. These include the median survival time, with upper and lower 95% confidence limits. The median survival time is the time (number of days) at which half the subjects have failed. The quartile survival times (25% and 75%) are also included.

Figure 13.15 Summary Statistics for the Univariate Survival Analysis

Product-Limit Survival Fit

Time to event: days
Censored by Censor
Censor Code 1
Grouped by Group

Summary

Group	Number failed	Number censored	Mean		Std Error
Group 1	19	2	240.795	Biased	11.206
Group 2	17	2	218.757		9.40318
Combined	36	4	230.729	Biased	7.57346

Quantiles

Group	Median Time	Lower 95%	Upper 95%	25% Failures	75% Failures
Group 1	233	232	280	232	280
Group 2	216	190	234	190	234
Combined	232	213	239	201.5	261

The Summary report gives estimates for the mean survival time, as well as the standard error of the mean. The estimated mean survival time is defined as follows:

$$\hat{\mu} = \sum_{i=1}^{D} \hat{S}(t_{i-1})(t_i - t_{i-1}) \quad \text{with a standard error of} \quad \hat{\sigma}(\hat{\mu}) = \sqrt{\frac{m}{m-1}\sum_{i=1}^{D-1}\frac{A_i^2}{n_i(n_i-d_i)}}$$

where

$$\hat{S}(t_i) = \prod_{j=1}^{i}\left(1 - \frac{d_j}{n_j}\right)$$

$$A_i = \sum_{j=i}^{D-1} \hat{S}(t_j)(t_{j+1} - t_j)$$

$$m = \sum_{j=1}^{D} d_j$$

$\hat{S}(t_i)$ is the survival distribution at time t_i

D is the number of distinct event times

n_i is the number of surviving units just prior to t_i

d_i is the number of units that fail at t_i

t_0 is defined to be 0

When there are multiple groups, the Tests Between Groups table provides statistical tests for homogeneity among the groups. Kalbfleisch and Prentice (1980, ch. 1), Hosmer and Lemeshow (1999, ch. 2), and Klein and Moeschberger (1997, ch. 7) discuss statistics and comparisons of survival curves.

Figure 13.16 Tests between Groups

Tests Between Groups

Test	ChiSquare	DF	Prob>ChiSq
Log-Rank	3.1227	1	0.0772
Wilcoxon	2.6510	1	0.1035

Test Names two statistical tests of the hypothesis that the survival functions are the same across groups.

Chi-Square Provides the Chi-square approximations for the statistical tests.

The **Log-Rank** test places more weight on larger survival times and is more useful when the ratio of hazard functions in the groups being compared is approximately constant. The hazard function is the instantaneous failure rate at a given time. It is also called the *mortality rate* or *force of mortality*.

The **Wilcoxon** test places more weight on early survival times and is the optimum rank test if the error distribution is logistic. See Kalbfleisch and Prentice (1980).

DF Provides the degrees of freedom for the statistical tests.

Prob>ChiSq Lists the probability of obtaining a Chi-square value greater than the one computed if the survival functions are the same for all groups.

Figure 13.17 shows an example of the product-limit survival function estimates for one group.

Figure 13.17 Example of Survival Estimates Table

Group 1

days	Survival	Failure	SurvStdErr	Number failed	Number censored	At Risk
0.000	1.0000	0.0000	0.0000	0	0	21
142.000	0.9524	0.0476	0.0465	1	0	21
156.000	0.9048	0.0952	0.0641	1	0	20
163.000	0.8571	0.1429	0.0764	1	0	19
198.000	0.8095	0.1905	0.0857	1	0	18
204.000	0.8095	0.1905	0.0857	0	1	17
205.000	0.7589	0.2411	0.0941	1	0	16
232.000	0.6577	0.3423	0.1053	2	0	15
233.000	0.4554	0.5446	0.1114	4	0	13
239.000	0.4048	0.5952	0.1099	1	0	9
240.000	0.3542	0.6458	0.1072	1	0	8
261.000	0.3036	0.6964	0.1031	1	0	7
280.000	0.2024	0.7976	0.0902	2	0	6
296.000	0.1012	0.8988	0.0678	2	0	4
323.000	0.0506	0.9494	0.0493	1	0	2
344.000	0.0506	0.9494	0.0493	0	1	1

Note: When the final time recorded is a censored observation, the report indicates a *biased* mean estimate. The biased mean estimate is a lower bound for the true mean.

Chapter 14

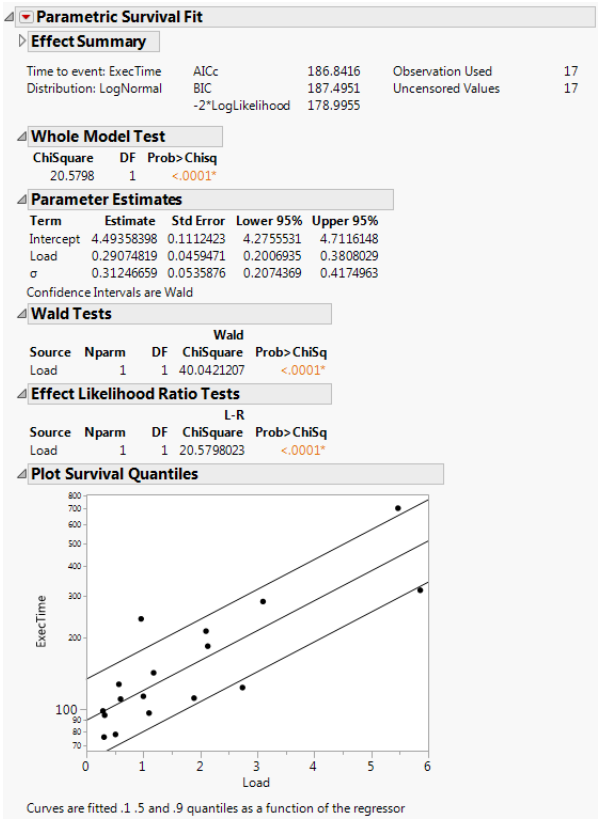
Fit Parametric Survival

Fit Survival Data Using Regression Models

Survival times can be expressed as a function of one or more variables. When this is the case, use a regression platform that fits a linear regression model while taking into account the survival distribution and censoring. The Fit Parametric Survival platform fits the time to event Y (with censoring) using linear regression models that can involve both location and scale effects. The fit is performed using the Weibull, lognormal, exponential, Fréchet, loglogistic, smallest extreme value (SEV), normal, largest extreme value (LEV), and logistic distributions.

Note: The Fit Parametric Survival platform is a slightly customized version of the Fit Model platform. You can also fit parametric survival models using the Nonlinear platform.

Figure 14.1 Example of a Parametric Survival Fit



Contents

Overview of the Fit Parametric Survival Platform . . . 371
Example of Parametric Regression Survival Fitting . . . 371
Launch the Fit Parametric Survival Platform . . . 373
The Parametric Survival Fit Report . . . 375
The Parametric Survival - All Distributions Report . . . 376
Parametric Competing Cause Report . . . 378
Fit Parametric Survival Options . . . 378
Nonlinear Parametric Survival Models . . . 380
Additional Examples of Fitting Parametric Survival . . . 381
Arrhenius Accelerated Failure LogNormal Model . . . 381
Interval-Censored Accelerated Failure Time Model . . . 385
Analyze Censored Data Using the Nonlinear Platform . . . 386
Left-Censored Data . . . 388
Weibull Loss Function Using the Nonlinear Platform . . . 389
Fitting Simple Survival Distributions Using the Nonlinear Platform . . . 392
Statistical Details for the Fit Parametric Survival Platform . . . 395
Loss Formulas for Survival Distributions . . . 395

Overview of the Fit Parametric Survival Platform

Survival times can be expressed as a function of one or more variables. When this is the case, use a regression platform that fits a linear regression model while taking into account the survival distribution and censoring. The Fit Parametric Survival platform fits the time to event Y (with censoring) using linear regression models that can involve both location and scale effects. The fit is performed using the Weibull, lognormal, exponential, Fréchet, loglogistic, SEV, normal, LEV, and logistic distributions.

Example of Parametric Regression Survival Fitting

The data table Comptime.jmp contains data on the analysis of computer program execution time whose lognormal distribution depends on the effect Load.

Note: The data in Comptime.jmp comes from Meeker and Escobar (1998, p. 434).

1. Select **Help > Sample Data Library** and open Reliability/Comptime.jmp.
2. Select **Analyze > Reliability and Survival > Fit Parametric Survival**.
3. Select ExecTime and click **Time to Event**.
4. Select Load and click **Add**.
5. Change the **Distribution** from **Weibull** to **Lognormal**.
6. Click **Run**.

Figure 14.2 Computing Time Output

Parametric Survival Fit

Effect Summary

Time to event: ExecTime	AICc	186.8416	Observation Used	17
Distribution: LogNormal	BIC	187.4951	Uncensored Values	17
	-2*LogLikelihood	178.9955		

Whole Model Test

ChiSquare	DF	Prob>Chisq
20.5798	1	<.0001*

Parameter Estimates

Term	Estimate	Std Error	Lower 95%	Upper 95%
Intercept	4.49358398	0.1112423	4.2755531	4.7116148
Load	0.29074819	0.0459471	0.2006935	0.3808029
σ	0.31246659	0.0535876	0.2074369	0.4174963

Confidence Intervals are Wald

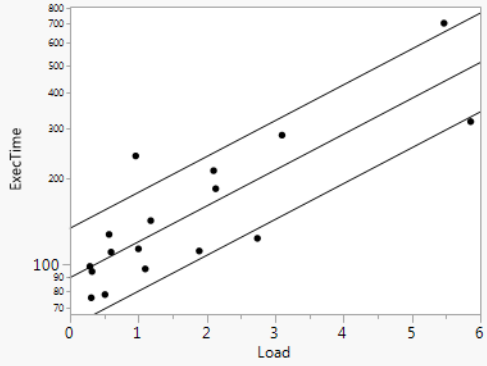
Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
Load	1	1	40.0421207	<.0001*

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Load	1	1	20.5798023	<.0001*

Plot Survival Quantiles



Curves are fitted .1 .5 and .9 quantiles as a function of the regressor

When there is only one effect, a plot of the survival quantiles for three survival probabilities are shown as a function of the effect.

Time quantiles are desired for when 90% of jobs are finished under a system load of 5. See Meeker and Escobar (1998, p. 438).

7. Click the Parametric Survival Fit red triangle and select **Estimate Quantile**.
8. Type 5 in the first row beneath **Load**.
9. Type 0.9 in the first row beneath **p**.
10. Click **Go**.

Figure 14.3 Estimates of Time Quantile

Estimates of Quantile

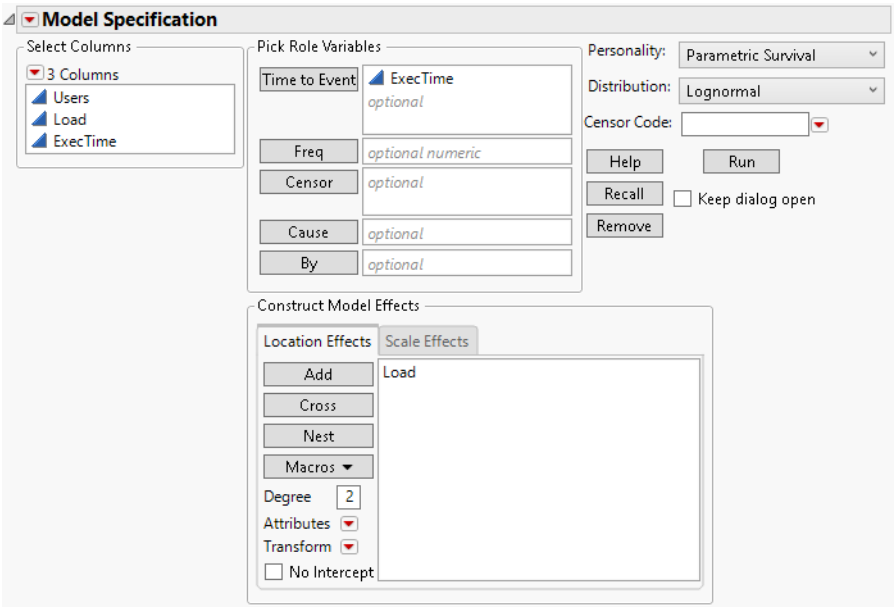
Load	p	Quantile	Lower 95%	Upper 95%
5	0.9	571.21575	401.29076	813.09482

The report estimates that 90% of the jobs will be done by 571 seconds of execution time under a system load of 5.

Launch the Fit Parametric Survival Platform

Launch the Fit Parametric Survival platform by selecting **Analyze > Reliability and Survival > Fit Parametric Survival**.

Figure 14.4 The Fit Parametric Survival Launch Window



Tip: To change the alpha level, click the Model Specification red triangle and select **Set Alpha Level**.

For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

The Fit Parametric Survival launch window contains the following options:

Time to Event Contains the time to event or time to censoring. With interval censoring, specify two *Y* variables, where one *Y* variable gives the lower limit and the other *Y* variable gives the upper limit for each unit.

Censor Specifies a column with indicators to identify right-censored observations. Select the value that identifies right-censored observations from the Censor Code menu. The Censor column is used only when one Time to Event column is entered.

Freq Specifies a column that contains the frequencies or counts of observations for each row when there are multiple units recorded.

Cause Specifies a column that contains multiple failure causes. This column is particularly useful for estimating competing causes. A separate parametric fit is performed for each cause value. Failure events can be coded with either numeric or categorical (labels) values.

By Performs a separate analysis for each level of a classification or grouping variable.

Location and Scale Effects Specifies location and scale effects. For more information about the Construct Model Effects options, see *Fitting Linear Models*.

Personality Indicates the fitting method. Parametric Survival should always be selected.

Distribution Choose the desired response distribution that is appropriate for your data. Choose the All Distributions option to fit all the distributions and compare the fits. If you choose All Distributions, the report shows a comparison of the distribution fits. See "The Parametric Survival - All Distributions Report".

Note: By default, the All Distributions option fits a model for the log-location-scale distributions that appear above All Distributions in the Distribution menu. Select **Preferences > Platforms > Fit Parametric Survival > Include location-scale distributions in All Distributions** to change the behavior of the All Distributions option to also include the location-scale distributions.

Censor Code Identifies the value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

The Parametric Survival Fit Report

If you select All Distributions in the launch window, a Parametric Survival Fit report appears for each distribution. If you specify a Cause column in the launch window, a Parametric Survival Fit report appears for each cause. Otherwise, only one Parametric Survival Fit report appears. Each Parametric Survival Fit report contains the following:

Effect Summary Shows an interactive report that enables you to add or remove effects from the model. See *Fitting Linear Models*.

Model Fit Details The Time to event shows which Y column is specified, and the Distribution shows which distribution is fit. AICc, BIC, and -2Loglikelihood are all measures of the model fit. These measures allow for comparisons to other model fits. Observation Used and Uncensored Values are summary statistics for the data. See *Fitting Linear Models*.

Whole Model Test Compares the complete fit with an intercept-only fit. If there is only an intercept term, the fit is the same as that from the Life Distribution platform.

Parameter Estimates Shows the estimates of the regression parameters.

JMP PRO A link to launch the Generalized Regression platform appears below the Parameter Estimates table. The link enables you to perform variable selection using the Generalized Regression platform and appears under the following circumstances:

- The model has no scale effects.
- No Cause column is specified in the launch window.
- The Distribution specified in the launch window is Normal, Lognormal, or Weibull.

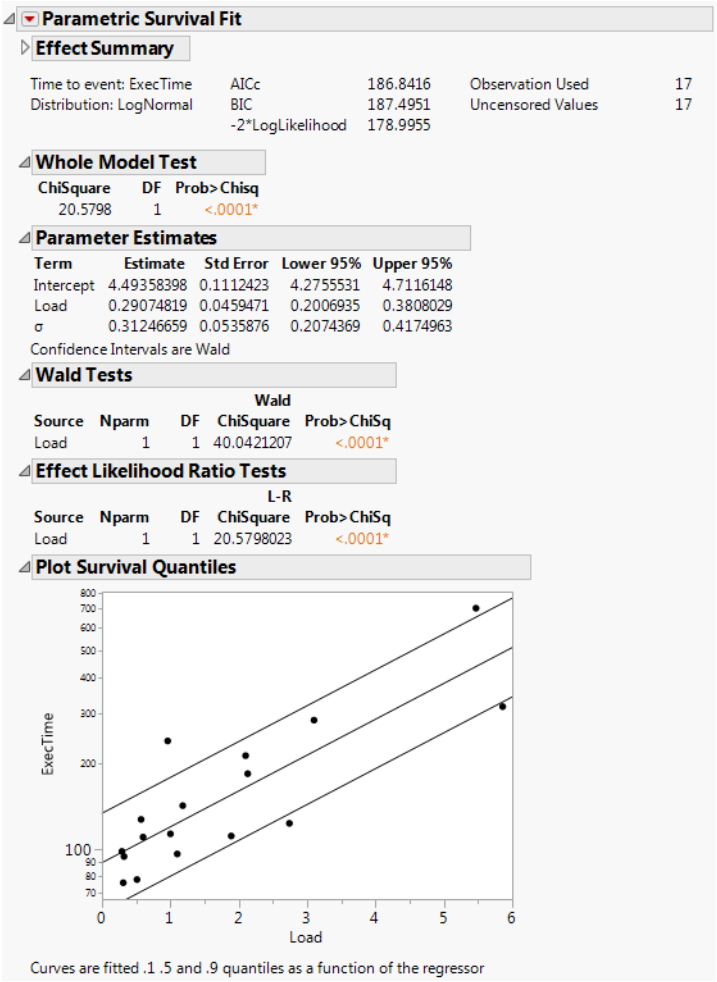
Alternate Parameterization (Available only for the Weibull distribution.) Shows the parameter estimates for the α and β parameterization of the Weibull distribution. For more information about this parameterization, see "Weibull".

Wald Tests Shows a Wald Chi-square test for each term in the model.

Effect Likelihood Ratio Tests Compare the log-likelihood from the fitted model to one that removes each term from the model individually.

Plot Survival Quantiles Shows the data points plotted with the 0.1, 0.5, and 0.9 quantiles.

Figure 14.5 The Parametric Survival Fit Report



The Parametric Survival - All Distributions Report

The Parametric Survival - All Distributions report appears only when you select All Distributions in the launch window. By default, this report contains a Model Comparison report and Distribution Overlay plot. The Quantile Function Overlay plot is available in the red triangle menu next to Parametric Survival - All Distributions.

Model Comparison Table that lists fit statistics (AICc and BIC) for the fitted distributions. The distributions with the smallest AICc and BIC values are labeled in the right-most column. If one distribution has the smallest value for both AICc and BIC, that distribution is labeled “Best”. The Parametric Survival Fit report corresponding to the distribution with

the smallest AICc is open by default. For more information about these statistics, see *Fitting Linear Models*.

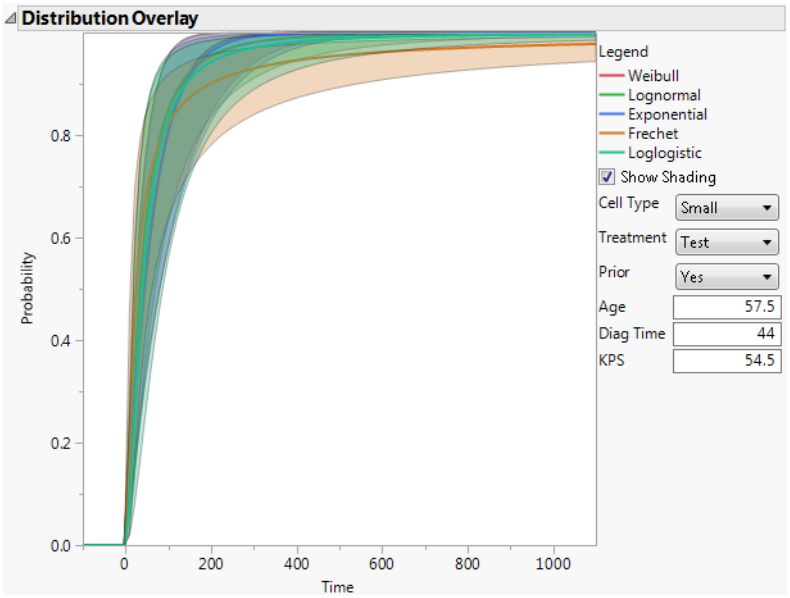
Distribution Overlay Plot of overlaid distribution functions for the fitted distributions at specified values of the effects.

Quantile Function Overlay Plot of overlaid quantile functions for the fitted distributions at specified values of the effects.

Both the Distribution Overlay and Quantile Function Overlay plots show curves for each of the fitted distributions overlaid on the same graph. By default, each curve has a shaded Wald-based confidence interval. To the right of each plot, there is legend, an option to show shading of confidence intervals, and controls that enable you to specify different values of the effects. Figure 14.6 shows an example of a Distribution Overlay plot.

Note: You can change the α level for the shaded confidence intervals by selecting Set Alpha Level from the red triangle menu in the Fit Model launch window. The default α level is 0.05.

Figure 14.6 Distribution Overlay Plot



Parametric Competing Cause Report

If you specify a Cause column in the launch window, the Parametric Competing Cause report appears. If you also specify All Distributions in the launch window, this report is labeled Parametric Competing Cause - All Distributions. The Parametric Competing Cause report contains the following:

Summary by Cause Table that lists fit statistics (AICc and BIC) for each cause. If All Distributions is selected for the Distribution option in the launch window, this table includes fit statistics for each cause within each distribution fit.

Model Comparison (Available only when All Distributions is selected for the Distribution option in the launch window.) Shows a table that lists fit statistics (AICc and BIC) for the fitted distributions. The distributions with the smallest AICc and BIC values are labeled in the right-most column. If one distribution has the smallest value for both AICc and BIC, that distribution is labeled "Best". The Parametric Survival Fit report corresponding to the distribution with the smallest AICc is open by default.

For more information about the AICc and BIC statistics, see *Fitting Linear Models*.

Fit Parametric Survival Options

The Parametric Survival Fit red triangle menu contains the following options:

Likelihood Ratio Tests Produces tests that compare the log-likelihood from the fitted model to one that removes each term from the model individually.

Wald Tests Produces chi-square test statistics and *p*-values for Wald tests of whether each parameter is zero.

Likelihood Confidence Intervals Specifies the type of confidence intervals shown in the Parameter Estimates table for each parameter. When this option is selected, a profile likelihood confidence interval appears. Otherwise, a Wald interval is shown. In the report, the interval type is noted below the Parameter Estimates table. This option is on by default when the computational time for the profile likelihood confidence intervals is not large.

Note: You can change the α level for the confidence intervals by selecting Set Alpha Level from the red triangle menu in the Fit Model launch window. The default α level is 0.05.

Correlation of Estimates Produces a correlation matrix for the model effects with each other and with the parameter of the fitting distribution.

Covariance of Estimates Produces a covariance matrix for the model effects with each other and with the parameter of the fitting distribution.

Estimate Survival Probability Estimates the failure and survival probabilities for the given time values. Specify effect values and one or more time values. JMP then calculates the survival and failure probabilities with 95% confidence limits for all possible combinations of the entries.

Estimate Quantile Estimates the quantiles for the given probabilities. Specify effect values and one or more quantile probabilities. JMP then calculates the time quantiles and 95% confidence limits for all possible combinations of the entries.

Note: For the Estimate Survival Probability and Estimate Quantile options, you can change the alpha level from the default of 0.05.

Residual Probability Plot Shows a probability plot of the standardized residuals. For location-scale distributions, the standardized residuals are defined as $(Y-\mu(X))/\sigma(X)$. For log-location-scale distributions, the standardized residuals are defined as $\exp(\log(Y)-\mu(X))/\sigma(X)$. For all distributions, $\mu(X)$ and $\sigma(X)$ are linear functions of the covariates defined by the model specification for the location and scale effects, respectively. These standardized residuals are a type of Cox-Snell residuals. See Meeker and Escobar (1998, sec. 17.6.1) for a discussion of Cox-Snell residuals.

Save Residuals Saves the residuals to a new column in the data table. For interval-censored observations, two columns of residuals are saved to the data table.

Distribution Profiler Shows the response surfaces of the failure probability versus individual explanatory and response variables.

Quantile Profiler Shows the response surfaces of the response variable versus the explanatory variable and the failure probability.

Distribution Plot by Level Combinations Shows three probability plots for assessing model fit. The plots show different lines for each combination of the X levels.

Separate Location A probability plot assuming equal scale parameters and separate location parameters. This is useful for assessing the parallelism assumption.

Separate Location and Scale A probability plot assuming different scale and location parameters. This is useful for assessing if the distribution is adequate for the data. This plot is not shown for the Exponential distribution.

Regression A probability plot for which the distribution parameters are functions of the X variables.

Save Probability Formula Saves the estimated probability formula to a new column in the data table.

Save Quantile Formula Saves the estimated quantile formula to a new column in the data table. Selecting this option displays a pop-up window, asking you to enter a probability value for the quantile of interest.

Publish Probability Formula Creates a probability formula and saves it as a formula column script in the Formula Depot platform. If a Formula Depot report is not open, this option creates a Formula Depot report. See *Predictive and Specialized Modeling*.

Publish Quantile Formula Creates a quantile formula and saves it as a formula column script in the Formula Depot platform. If a Formula Depot report is not open, this option creates a Formula Depot report. See *Predictive and Specialized Modeling*.

Model Dialog Relaunches the launch window.

Effect Summary Shows the interactive Effect Summary report that enables you to add or remove effects from the model. See *Fitting Linear Models*.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Nonlinear Parametric Survival Models

Use the Nonlinear platform for survival models in the following instances:

- The model is nonlinear.
- You need a distribution other than Weibull, lognormal, exponential, Fréchet, loglogistic, SEV, normal, LEV, or logistic.
- You have censoring that is not the usual right, left, or interval censoring.

With the ability to estimate parameters in specified loss functions, the Nonlinear platform becomes a powerful tool for fitting maximum likelihood models. For complete information about the Nonlinear platform, see *Predictive and Specialized Modeling*.

To fit a nonlinear model when data are censored, you must first use the formula editor to create a parametric equation that represents a loss function adjusted for censored observations. Then use the Nonlinear platform to estimate the parameters using maximum likelihood.

Additional Examples of Fitting Parametric Survival

- "Arrhenius Accelerated Failure LogNormal Model"
- "Interval-Censored Accelerated Failure Time Model"
- "Analyze Censored Data Using the Nonlinear Platform"
- "Left-Censored Data"
- "Weibull Loss Function Using the Nonlinear Platform"
- "Fitting Simple Survival Distributions Using the Nonlinear Platform"

Arrhenius Accelerated Failure LogNormal Model

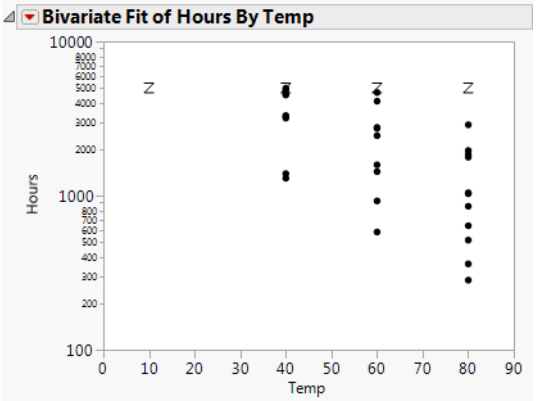
In the Devalt.jmp data, units are stressed by heating, in order to make them fail soon enough to obtain enough failures to fit the distribution.

Note: The data in Devalt.jmp comes from Meeker and Escobar (1998, p. 493).

1. Select **Help > Sample Data Library** and open Reliability/Devalt.jmp.

   First, use the Bivariate platform to see a plot of hours by temperature using the log scale for time.
2. Select **Analyze > Fit Y by X**.
3. Select Hours and click **Y, Response**.
4. Select Temp and click **X, Factor**.
5. Click **OK**.

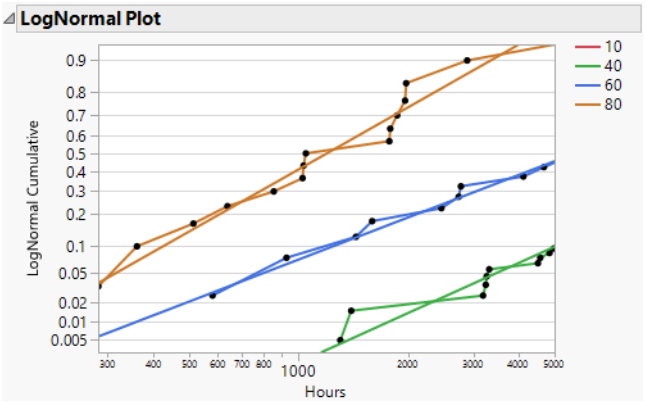
Figure 14.7 Bivariate Plot of Hours by Log Temp



Next, use the survival platform to produce a LogNormal plot of the data for each temperature.

6. Select **Analyze > Reliability and Survival > Survival**.
7. Select Hours and click **Y, Time to Event**.
8. Select Censor and click **Censor**.
9. Select Temp and click **Grouping**.
10. Select Weight and click **Freq**.
11. Click **OK**.
12. Click the red triangle next to Product-Limit Survival Fit and select **LogNormal Plot** and **LogNormal Fit**.
13. Click **OK**.

Figure 14.8 Lognormal Plot



Next, use the Fit Parametric Survival platform to fit one model using an effect for temperature.

14. Select **Analyze > Reliability and Survival > Fit Parametric Survival**.
15. Select Hours and click **Time to Event**.
16. Select x and click **Add**.
17. Select Censor and click **Censor**.
18. Select Weight and click **Freq**.
19. Change the **Distribution** type to **Lognormal**.
20. Click **Run**.

Figure 14.9 Devalt Parametric Output

Parametric Survival Fit

Effect Summary

Time to event: Hours
Distribution: Lognormal
Censored By: Censor
Freq Column: Weight

AICc	649.5546	Observation Used	165
BIC	658.7234	Uncensored Values	33
-2*LogLikelihood	643.4056	Right Censored Values	132

Whole Model Test

ChiSquare	DF	Prob>Chisq
80.7141	1	<.0001*

Parameter Estimates

Term	Estimate	Std Error	Lower 95%	Upper 95%
Intercept	-13.468649	2.8871953	-19.12745	-7.809851
x	0.62787903	0.0828422	0.4655112	0.7902468
σ	0.97782331	0.1326468	0.7178403	1.2378063

Confidence Intervals are Wald

Variable Selection using Generalized Regression

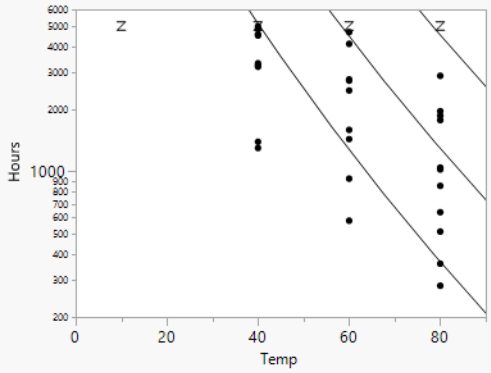
Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
x	1	1	57.4444814	<.0001*

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
x	1	1	80.714074	<.0001*

Plot Survival Quantiles



Curves are fitted .1 .5 and .9 quantiles as a function of the regressor
Some of the points are censored times

The result shows the regression fit of the data:

– If there is only one effect and it is continuous, then a plot of the survival as a function of the effect is shown. Lines are at 0.1, 0.5, and 0.9 survival probabilities.

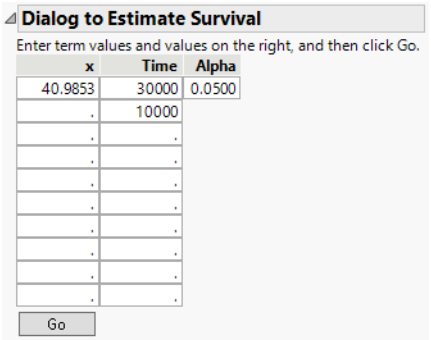
– If the effect column has a formula in terms of one other column, as in this case, the plot is done with respect to the inner column. In this case, the effect was the column x, but the plot is done with respect to Temp, of which x is a function.

Finally, get estimates of survival probabilities extrapolated to a temperature of 10 degrees Celsius for the times 30000 and 10000 hours.

21. Click the Parametric Survival Fit red triangle and select **Estimate Survival Probability**.
22. Enter the values shown in Figure 14.10 into the Dialog to Estimate Survival.

    The Arrhenius transformation of 10 degrees is 40.9853, the effect value.

Figure 14.10 Estimating Survival Probabilities



23. Click **Go**.

Figure 14.11 Survival Probabilities

Estimates of Survival

x	Time	Prob Failure	Lower 95%	Upper 95%	Prob Survival
40.9853	30000	0.0227191	0.0024223	0.1182138	0.9772809
40.9853	10000	0.0008917	4.3289e-5	0.0100954	0.9991083

The Estimates of Survival report shows the estimates and a confidence interval.

Interval-Censored Accelerated Failure Time Model

The ICdevice02.jmp data shows failures that were found to have happened between inspection intervals. The model uses two *y*-variables, containing the upper and lower bounds on the failure times. Right-censored times are shown with missing upper bounds.

Note: The data in ICdevice02.jmp comes from Meeker and Escobar (1998, p. 640).

1. Select **Help > Sample Data Library** and open Reliability/ICdevice02.jmp.
2. Select **Analyze > Reliability and Survival > Fit Parametric Survival**.
3. Select HoursL and HoursU and click **Time to Event**.
4. Select Count and click **Freq**.
5. Select x and click **Add**.
6. Click **Run**.

Figure 14.12 ICDevice Output

Parametric Survival Fit

Effect Summary

Time to event: HoursL .. HoursU
Distribution: Weibull
Freq Column: Count

AICc	185.9584	Observation Used	250
BIC	196.4252	Right Censored Values	194
-2*LogLikelihood	179.8608	Interval Censored Values	56

Whole Model Test

ChiSquare	DF	Prob>Chisq
195.5928	1	<.0001*

Parameter Estimates

Term	Estimate	Std Error	Lower 95%	Upper 95%
Intercept	-10.533672	2.0139264	-14.48089	-6.586449
x	0.85579005	0.0977613	0.6641815	1.0473986
δ	0.4376781	0.0500839	0.3395155	0.5358407

Confidence Intervals are Wald
Variable Selection using Generalized Regression

Alternate Parameterization

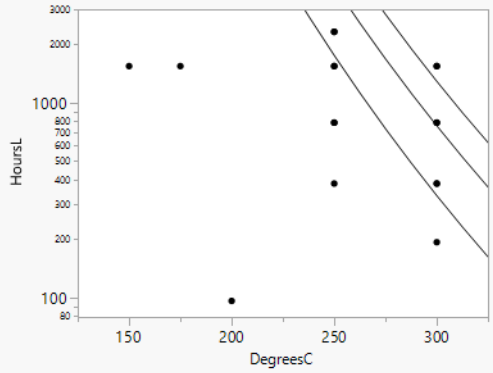
Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
x	1	1	76.6303486	<.0001*

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
x	1	1	195.592803	<.0001*

Plot Survival Quantiles



Curves are fitted .1 .5 and .9 quantiles as a function of the regressor
Points are from lower and upper values of interval

The resulting regression shows a plot of time by degrees.

Analyze Censored Data Using the Nonlinear Platform

You can analyze left-censored data using the Nonlinear platform. For the left-censored data, zero is the censored value because it also represents the smallest known time for an observation.

Note: The Tobit model is popular in economics for responses that must be positive or zero, with zero representing a censored point.

Note the following about the Tobit2.jmp data table:

- The response variable is a measure of the durability of a product and cannot be less than zero (Durable, is left-censored at zero).
- Age and Liquidity are independent variables.
- The table also includes the model and Tobit loss function. The model in residual form is `durable-(b0+b1*age+b2*liquidity)`. To see the formula associated with Tobit Loss, right-click the column and select **Formula**.

Fit the Tobit model:

1. Select **Help > Sample Data Library** and open Reliability/Tobit2.jmp.
2. Select **Analyze > Specialized Modeling > Nonlinear**.
3. Select Model and click **X, Predictor Formula**.
4. Select Tobit Loss and click **Loss**.
5. Click **OK**.
6. Click **Go**.
7. Click **Confidence Limits**.

Figure 14.13 Solution Report

Solution

Loss	DFE	Avg Loss	Sqrt Avg Loss
28.92596097	16	1.8078726	1.3445715

Parameter	Estimate	ApproxStdErr	Lower CL	Upper CL
b0	15.277120063	16.0327167	-22.059913	54.7012259
b1	-0.134007501	0.21893135	-0.7365041	0.32841861
b2	-0.045135584	0.05826851	-0.1858906	0.09330941
Sigma	5.5693492202	1.72814365	3.31613117	11.5291865

Solved By: Analytic NR

Correlation of Estimates

Left-Censored Data

The Tobit model from the previous section assumes a normal distribution that is censored at zero. An observation of zero is considered to be left-censored. In the Fit Parametric Survival platform, left-censored observations are specified using two response columns. In this example, you create a new column that indicates observations for which left censoring has occurred. For left-censored observations, the new column contains a missing value. Otherwise, it contains the observed value of durable. The new column is then used as the left side of interval-censored observations, and the existing column durable is used as the right side of the intervals.

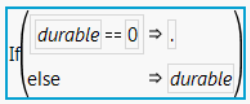
1. Select **Help > Sample Data Library** and open Reliability/Tobit2.jmp.

Create the Left Censoring Column:

2. Select **Cols > New Columns**.
3. Type durable0 for **Column Name**.
4. Select **Column Properties** and click **Formula**.
5. Select **Conditional > If** and select durable.
6. Select **Comparison > a == b**, type 0, and press Enter.
7. Select the box labeled "else clause" and select durable.

   Figure 14.14 shows the completed formula.

Figure 14.14 Column Formula for durable0



8. Click **OK**.
9. Click **OK**.

Fit the Tobit Model:

10. Select **Analyze > Reliability and Survival > Fit Parametric Survival**.
11. Select durable0 and click **Time to Event**.
12. Select durable and click **Time to Event**.

    You must use two response columns to specify left-censored observations. The direction of censoring is determined by the order of the columns in the Time to Event role.
13. Select age and liquidity and click **Add**.
14. Change the **Distribution** type to **Normal**.

15. Click **Run**.

Figure 14.15 Tobit Model Results

Parametric Survival Fit

Effect Summary

Time to event: durable0 .. durable
Distribution: Normal

AICc	68.51859	Observation Used	20
BIC	69.83485	Uncensored Values	7
-2*LogLikelihood	57.85192	Left Censored Values	13

Whole Model Test

ChiSquare	DF	Prob>Chisq
1.1325	2	0.5677

Parameter Estimates

Term	Estimate	Std Error	Lower 95%	Upper 95%
Intercept	15.2771208	16.032721	-16.14643	46.700676
age	-0.1340075	0.2189314	-0.563105	0.2950901
liquidity	-0.0451356	0.0582685	-0.15934	0.0690686
σ	5.56935051	1.7281446	2.1822494	8.9564517

Confidence Intervals are Wald
Variable Selection using Generalized Regression

Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
age	1	1	0.37466429	0.5405
liquidity	1	1	0.60002606	0.4386

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
age	1	1	0.3931753	0.5306
liquidity	1	1	0.57102176	0.4499

The report shows the estimated Tobit model fit. Note in the upper right part of the report that there are 13 left-censored observations. This is a good way to check that you have correctly specified the left censoring.

Weibull Loss Function Using the Nonlinear Platform

In this example, models are fit to the survival time using the Weibull, lognormal, and exponential distributions. Model fits include a simple survival model containing only two effects, a more complex model with all the effects, and the creation of dummy variables for the discrete effect Cell Type to be included in the full model.

Nonlinear model fitting is often sensitive to the initial values that you give to the model parameters. In this example, one way to find reasonable initial values is to first use the Nonlinear platform to fit only the linear model. When the model converges, the solution values for the parameters become the initial parameter values for the nonlinear model.

1. Select **Help > Sample Data Library** and open VA Lung Cancer.jmp.

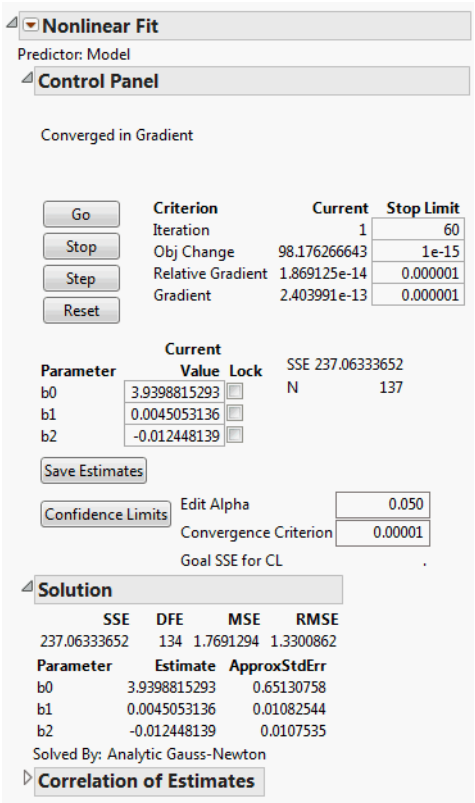
   The first model and all the loss functions have already been created as formulas in the data table. The Model column has the following formula:

   ```
   Log(:Time) - (b0 + b1 * Age + b2 * Diag Time)
   ```

2. Select **Analyze > Specialized Modeling > Nonlinear**.

3. Select Model and click **X, Predictor Formula**.
4. Click **OK**.
5. Click **Go**.

Figure 14.16 Initial Parameter Values in the Nonlinear Fit Control Panel



The report computes the least squares parameter estimates for this model.

6. Click **Save Estimates**.

   The parameter estimates in the column formulas are set to those estimated by this initial nonlinear fitting process.

   The Weibull column contains the Weibull formula, explained in "Weibull Loss Function".

   To continue with the fitting process:

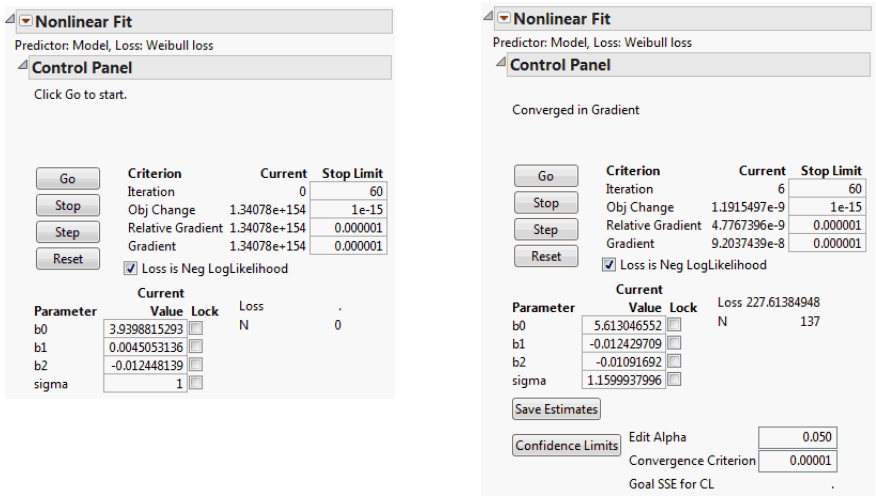
7. Select **Analyze > Specialized Modeling > Nonlinear** again.
8. Select Model and click **X, Predictor Formula**.
9. Select Weibull loss and click **Loss**.
10. Click **OK**.

The Nonlinear Fit Control Panel on the left in Figure 14.17 appears. There is now the additional parameter called sigma in the loss function. Because it is in the denominator of a fraction, a starting value of 1 is reasonable for sigma. When using any loss function other than the default, the **Loss is Neg LogLikelihood** box on the Control Panel is checked by default.

11. Click **Go**.

    The fitting process converges as shown on the right in Figure 14.17.

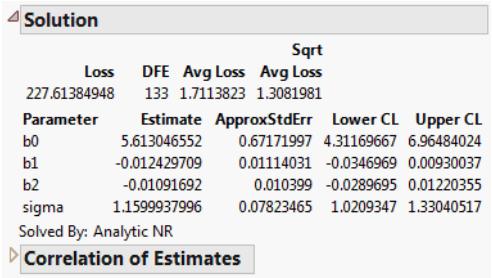
Figure 14.17 Nonlinear Model with Custom Loss Function



The fitting process estimates the parameters by maximizing the negative log of the Weibull likelihood function.

12. (Optional) Click **Confidence Limits** to show lower and upper 95% confidence limits for the parameters in the Solution table.

Figure 14.18 Solution Report



Note: Because the confidence limits are profile likelihood confidence intervals instead of the standard asymptotic confidence intervals, they can take time to compute.

You can also run the model with the predefined exponential and lognormal loss functions. Before you fit another model, reset the parameter estimates to the least squares estimates, as they might not converge otherwise. To reset the parameter estimates:

13. (Optional) Click the Nonlinear Fit red triangle and select **Revert to Original Parameters**.

Fitting Simple Survival Distributions Using the Nonlinear Platform

The following examples show how to use maximum likelihood methods to estimate distributions from time-censored data when there are no effects other than the censor status.

The Loss Function Templates folder has templates with formulas for exponential, extreme value, loglogistic, lognormal, normal, and one-and two-parameter Weibull loss functions. To use these loss functions, copy your time and censor values into the Time and censor columns of the loss function template. To run the model, select **Nonlinear** and assign the loss column as the **Loss** variable. Because both the response model and the censor status are included in the loss function and there are no other effects, you do not need a prediction column (model variable).

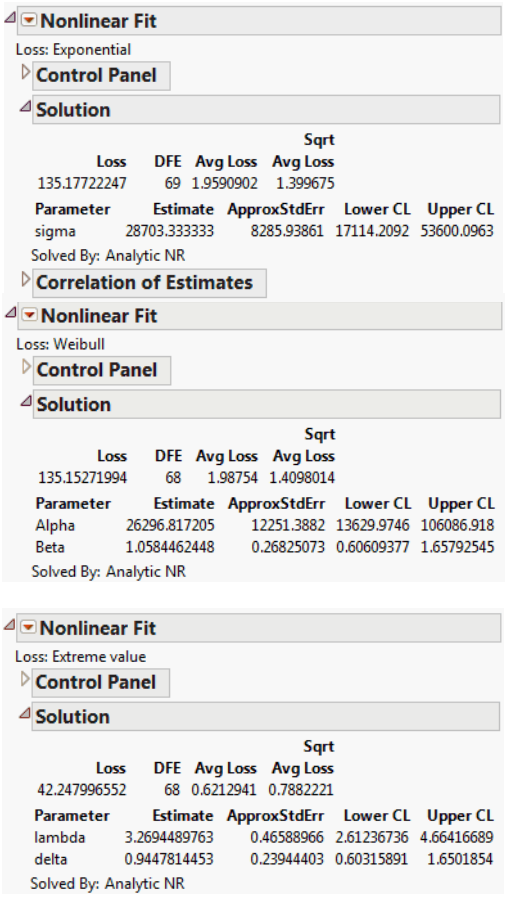
Exponential, Weibull, and Extreme-Value Loss Function

The Fan.jmp data table can be used to illustrate the Exponential, Weibull, and Extreme value loss functions discussed in Nelson (1982). The data are from a study of 70 diesel fans that accumulated a total of 344,440 hours in service. The fans were placed in service at different times. The response is failure time of the fans or run time, if censored.

Tip: To view the formulas for the loss functions, in the Fan.jmp data table, right-click the Exponential, Weibull, and Extreme value columns and select **Formula**.

1. Select **Help > Sample Data Library** and open Reliability/Fan.jmp.
2. Select **Analyze > Specialized Modeling > Nonlinear**.
3. Select Exponential and click **Loss**.
4. Click **OK**.
5. Make sure that the Loss is Neg LogLikelihood check box is selected.
6. Click **Go**.
7. Click **Confidence Limits**.
8. Repeat these steps, but select Weibull and Extreme value instead of Exponential.

Figure 14.19 Nonlinear Fit Results



Lognormal Loss Function

The Locomotive.jmp data can be used to illustrate a lognormal loss. The lognormal distribution is useful when the range of the data is several powers of *e*.

Tip: To view the formula for the loss function, in the Locomotive.jmp data table, right-click the logNormal column and select **Formula**.

The lognormal loss function can be very sensitive to starting values for its parameters. Because the lognormal distribution is similar to the normal distribution, you can create a new variable that is the natural log of Time and use **Distribution** to find the mean and standard deviation of this column. Then, use those values as starting values for the Nonlinear platform. In this example, the mean of the natural log of Time is 4.72 and the standard deviation is 0.35.

1. Select **Help > Sample Data Library** and open Reliability/Locomotive.jmp.

2. Select **Analyze > Specialized Modeling > Nonlinear**.
3. Select logNormal and click **Loss**.
4. Click **OK**.
5. Type 4.72 in the box next to Mu.
6. Type 0.35 in the box next to sigma.
7. Click **Go**.
8. Click **Confidence Limits**.

Figure 14.20 Solution Report

Solution

Loss	DFE	Avg Loss	Sqrt Avg Loss
237.09354534	94	2.5222718	1.5881662

Parameter	Estimate	ApproxStdErr	Lower CL	Upper CL
Mu	5.1169247193	0.10416697	4.93735853	5.35966436
sigma	0.7054940302	0.0932136	0.55446171	0.93294904

Solved By: Numerical NR

Correlation of Estimates

The maximum likelihood estimates of the lognormal parameters are 5.11692 for Mu and 0.7055 for Sigma. The corresponding estimate of the median of the lognormal distribution is the antilog of 5.11692, ($e^{5.11692}$), which is approximately 167. This represents the typical life for a locomotive engine.

Statistical Details for the Fit Parametric Survival Platform

This section contains statistical details for the Fit Parametric Survival platform.

Loss Formulas for Survival Distributions

The following formulas are for the negative log-likelihoods to fit common parametric models. Each formula uses the calculator `if` conditional function with the uncensored case of the conditional first and the right-censored case as the `Else` clause. You can copy these formulas from tables in the Loss Function Templates folder in Sample Data and paste them into your data table.

Exponential Loss Function

In the exponential loss function shown here, sigma represents the mean of the exponential distribution and Time is the age at failure.

$$-\text{If}\left(\begin{array}{ll} \text{Censor} == 0 & \Rightarrow -\text{Log}\left(\text{sigma}\right) - \dfrac{\text{Time}}{\text{sigma}} \\ \text{else} & \Rightarrow -\left(\dfrac{\text{Time}}{\text{sigma}}\right) \end{array}\right)$$

A characteristic of the exponential distribution is that the instantaneous failure rate remains constant over time. This means that the chance of failure for any subject during a given length of time is the same regardless of how long a subject has been in the study.

Weibull Loss Function

The Weibull density function often provides a good model for the lifetime distributions. You can use the Survival platform for an initial investigation of data to determine whether the Weibull loss function is appropriate for your data.

$$-\text{If}\left(\begin{array}{ll} \text{Censor} == 0 & \Rightarrow \dfrac{\text{Model}}{\text{sigma}} - \text{Exp}\left(\dfrac{\text{Model}}{\text{sigma}}\right) - \text{Log}\left(\text{sigma}\right) \\ \text{else} & \Rightarrow -\text{Exp}\left(\dfrac{\text{Model}}{\text{sigma}}\right) \end{array}\right)$$

There are examples of one-parameter, two-parameter, and extreme-value functions in the Loss Function Templates folder.

Lognormal Loss Function

The formula shown below is the lognormal loss function where `Normal Distribution(model/sigma)` is the standard normal distribution function. The hazard function has value 0 at $t = 0$, increases to a maximum, and then decreases. The hazard function approaches zero as t becomes large.

$$-\text{If}\left(\begin{array}{ll} censor==0 & \Rightarrow -0.5*\left(\frac{Model}{\mathbf{sigma}}\right)^2 - 0.5*\text{Log}(2*\text{Pi}) - \text{Log}(\mathbf{sigma}) \\ \text{else} & \Rightarrow \text{Log}\left(1 - \text{Normal Distribution}\left(\frac{Model}{\mathbf{sigma}}\right)\right) \end{array}\right)$$

Loglogistic Loss Function

If Y is distributed as the logistic distribution, Exp(Y) is distributed as the loglogistic distribution.

$$-\text{IfMZ}\left(\begin{array}{ll} \text{censor} == 0 & \Rightarrow \frac{\text{Model}}{\text{sigma}} - 2 * \text{Log}\left(1 + \text{Exp}\left(\frac{\text{Model}}{\text{sigma}}\right)\right) - \text{Log}(\text{sigma}) \\ \text{else} & \Rightarrow -\text{Log}\left(1 + \text{Exp}\left(\frac{\text{Model}}{\text{sigma}}\right)\right) \end{array}\right)$$

Chapter 15

Fit Proportional Hazards

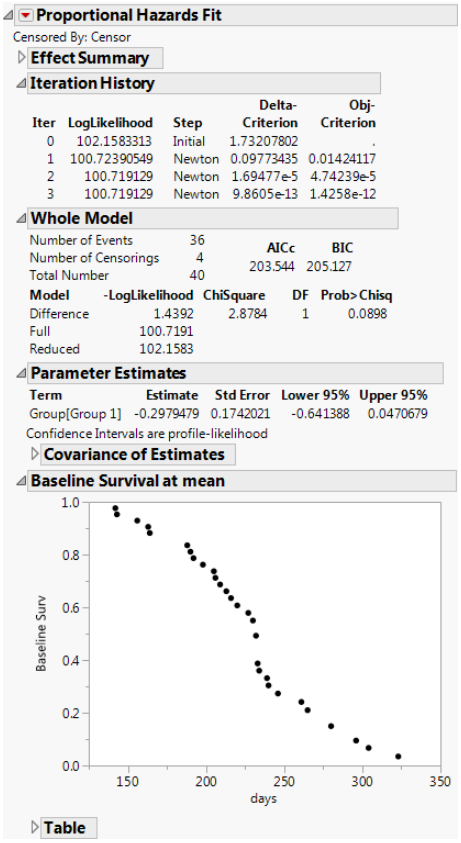
Fit Survival Data Using Semiparametric Regression Models

The Fit Proportional Hazards platform fits the Cox proportional hazards model, which assumes a multiplying relationship between covariates (predictors) and the hazard function.

Proportional hazards models are popular regression models for survival data with covariates. This model is semiparametric. The linear model is estimated, but the form of the hazard function is not. Time-varying covariates are not supported.

Note: The Fit Proportional Hazards platform is a slightly customized version of the Fit Model platform.

Figure 15.1 Example of a Proportional Hazards Fit



Contents

Overview of the Fit Proportional Hazards Platforms 399
Example of the Fit Proportional Hazards Platform 399
Risk Ratios for One Nominal Effect with Two Levels 401
Launch the Fit Proportional Hazards Platform 402
The Fit Proportional Hazards Report 404
Fit Proportional Hazards Options 405
Example Using Multiple Effects and Multiple Levels 406
Risk Ratios for Multiple Effects and Multiple Levels 408

Overview of the Fit Proportional Hazards Platforms

The proportional hazards model is a special semiparametric regression model proposed by D. R. Cox (1972) to examine the effect of explanatory variables on survival times. The survival time of each member of a population is assumed to follow its own hazard function.

The proportional hazards model is nonparametric in the sense that it involves an unspecified arbitrary baseline hazard function. It is parametric because it assumes a parametric form for the covariates. The baseline hazard function is scaled by a function of the model's (time-independent) covariates to give a general hazard function. Unlike the Kaplan-Meier analysis, proportional hazards computes parameter estimates and standard errors for each covariate. The regression parameters (β) associated with the explanatory variables and their standard errors are estimated using the maximum likelihood method. A conditional risk ratio (or hazard ratio) is also computed from the parameter estimates.

The survival estimates in proportional hazards are generated using an empirical method. See Lawless (1982). They represent the empirical cumulative hazard function estimates, $H(t)$, of the survivor function, $S(t)$, and can be written as $S_0 = \exp(-H(t))$. The hazard function is defined as follows:

$$\mathrm{H}(t) = \sum_{j:t_j < t} \frac{d_j}{\sum_{l \in R_j} e^{x_l \beta}}$$

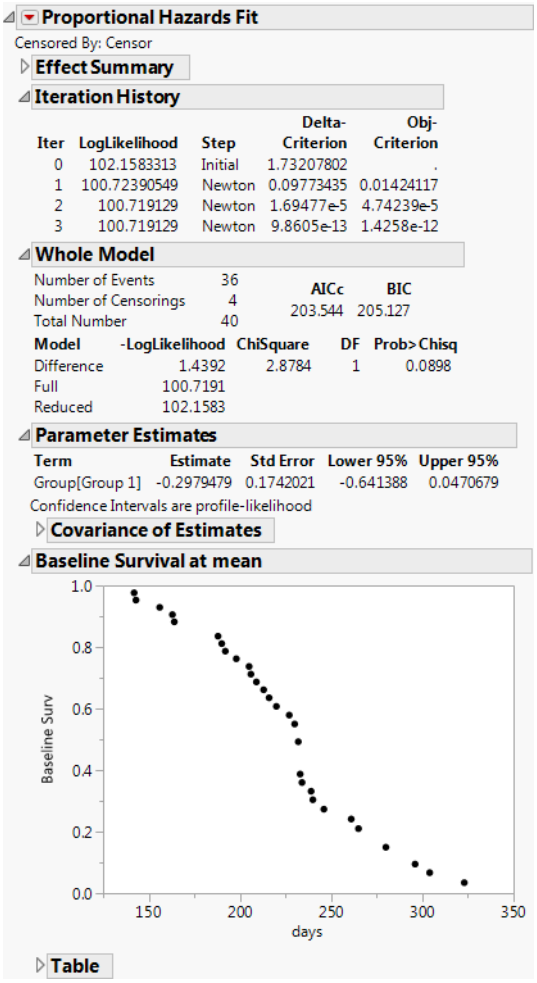
When there are ties in the response, meaning there is more than one failure at a given time event, the Breslow likelihood is used.

Example of the Fit Proportional Hazards Platform

This example illustrates one nominal effect with two levels. For an example with multiple effects and multiple levels, see "Example Using Multiple Effects and Multiple Levels".

1. Select **Help > Sample Data Library** and open Rats.jmp.
2. Select **Analyze > Reliability and Survival > Fit Proportional Hazards**.
3. Select days and click **Time to Event**.
4. Select Censor and click **Censor**.
5. Select Group and click **Add**.
6. Click **Run**.

Figure 15.2 Proportional Hazards Fit Report for Rats.jmp Data



In the Rats.jmp data, there are only two groups. Therefore, in the Parameter Estimates report, a confidence interval that does not include zero indicates an alpha-level significant difference between groups. Also, in the Effect Likelihood Ratio Tests report, the test of the null hypothesis for no difference between the groups shown in the Whole Model Test table is the same as the null hypothesis that the regression coefficient for Group is zero.

Risk Ratios for One Nominal Effect with Two Levels

To show risk ratios for effects, select the **Risk Ratios** option from the red triangle menu. In this example, there is only one effect, and there are only two levels for that effect. The risk ratio for Group 2 is compared with Group 1 and appears in the Risk Ratios for Group report. See Figure 15.3. The risk ratio in this table is determined by computing the exponential of the parameter estimate for Group 2 and dividing it by the exponential of the parameter estimate for Group 1.

Note the following:

- The Group 1 parameter estimate appears in the Parameter Estimates table (Figure 15.2).
- The Group 2 parameter estimate is calculated by taking the negative value for the parameter estimate of Group 1.
- Reciprocal shows the value for 1/Risk Ratio.

Tip: To see the Reciprocal values, right-click in the Risk Ratios report and select **Columns > Reciprocal**.

For this example, the risk ratio for Group2/Group1 is calculated as follows:

exp[-(-0.2979479)] /exp(-0.2979479) = 1.8146558

This risk ratio value suggests that the risk of death for Group 2 is 1.81 times that for Group 1.

Figure 15.3 Risk Ratios for Group Table

Risk Ratios

Risk Ratios for Group

Level1	/Level2	Risk Ratio	Prob>Chisq	Lower 95%	Upper 95%
Group 2	Group 1	1.8146558	0.0872	0.9167104	3.5921661
Group 1	Group 2	0.5510687	0.0872	0.2783836	1.0908571

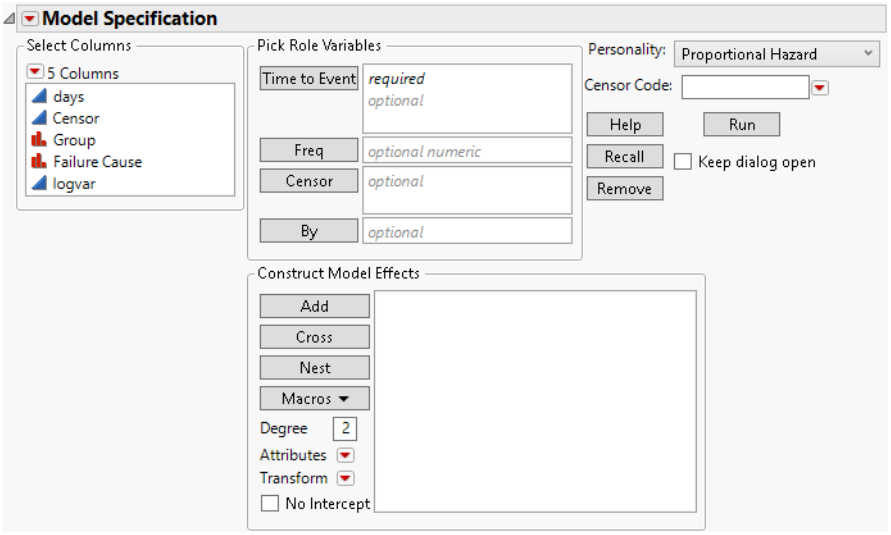
Normal approximations used for ratio confidence limits effects: Group

For information about calculating risk ratios when there are multiple effects, or categorical effects with more than two levels, see "Risk Ratios for Multiple Effects and Multiple Levels".

Launch the Fit Proportional Hazards Platform

Launch the Fit Proportional Hazards platform by selecting **Analyze > Reliability and Survival > Fit Proportional Hazards**.

Figure 15.4 The Fit Proportional Hazards Launch Window



Tip: To change the alpha level, click the Model Specification red triangle and select **Set Alpha Level**.

For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

The Fit Proportional Hazards launch window contains the following options:

Time to Event Contains the time to event or time to censoring.

Censor Specifies a column with indicators to identify right-censored observations. Select the value that identifies right-censored observations from the Censor Code menu.

Freq Column whose values are the frequencies or counts of observations for each row when there are multiple units recorded.

By Performs a separate analysis for each level of a classification or grouping variable.

Construct Model Effects Enters effects into your model. For more information about the Construct Model Effects options, see *Fitting Linear Models*.

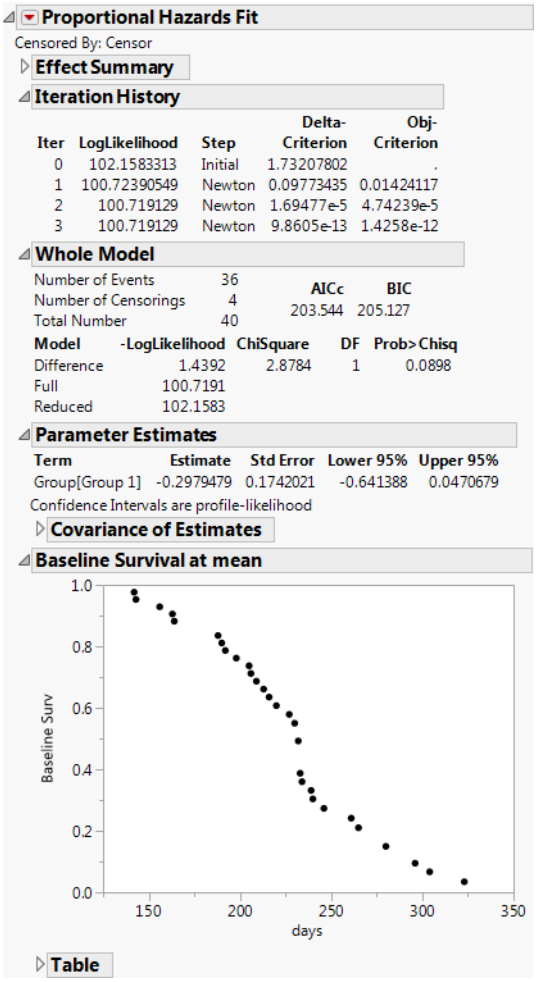
Personality Indicates the fitting method. Proportional Hazard should always be selected.

Censor Code Identifies the value in the Censor column that designates right-censored observations. After a Censor column is selected, JMP attempts to automatically detect the censor code and display it in the box. To change this, you can click the red triangle and select from a list of values. You can also enter a different value in the box. If the Censor column contains a Value Labels column property, the value labels appear in the list of values. Missing values are excluded from the analysis.

The Fit Proportional Hazards Report

Finding parameter estimates for a proportional hazards model is an iterative procedure. When the fitting is complete, the report in Figure 15.5 appears.

Figure 15.5 The Proportional Hazards Fit Report



Iteration History Lists iteration results occurring during the model calculations.

Whole Model Shows the negative of the log-likelihood function (–LogLikelihood) for the model with and without the covariates. Twice the positive difference between them gives a chi-square test of the hypothesis that there is no difference in survival time among the effects. The degrees of freedom (DF) are equal to the change in the number of parameters between the full and reduced models. See *Fitting Linear Models*.

Parameter Estimates Shows the parameter estimates for the covariates, their standard errors, and 95% upper and lower confidence limits. A confidence interval for a continuous column that does not include zero indicates that the effect is significant. A confidence interval for a level in a categorical column that does not include zero indicates that the difference between the level and the average of all levels is significant.

Effect Likelihood Ratio Tests Shows the likelihood ratio chi-square test of the null hypothesis that the parameter estimates for the effects of the covariates is zero.

Baseline Survival at mean Plots the baseline function estimates at each event time in the data. The values in the Table report are plotted here.

Fit Proportional Hazards Options

The Proportional Hazards Fit red triangle menu contains the following options:

Likelihood Ratio Tests Produces tests that compare the log-likelihood from the fitted model to one that removes each term from the model individually.

Wald Tests Produces chi-square test statistics and *p*-values for Wald tests of whether each parameter is zero.

Likelihood Confidence Intervals Specifies the type of confidence intervals shown in the Parameter Estimates table for each parameter. When this option is selected, a profile likelihood confidence interval appears. Otherwise, a Wald interval is shown. In the report, the interval type is noted below the Parameter Estimates table. This option is on by default when the computational time for the profile likelihood confidence intervals is not large.

Note: You can change the α level for the confidence intervals by selecting Set Alpha Level from the red triangle menu in the Fit Model launch window. The default α level is 0.05.

Risk Ratios Shows the risk ratios for the effects. For continuous columns, unit risk ratios and range risk ratios are calculated. The Unit Risk Ratio is Exp(estimate) and the Range Risk Ratio is Exp[estimate*($x_{Max} - x_{Min}$)]. The Unit Risk Ratio shows the risk change over one unit of the regressor, and the Range Risk Ratio shows the change over the whole range of the regressor. For categorical columns, risk ratios are shown in separate reports for each effect. Note that for a categorical variable with *k* levels, only *k* -1 design variables, or levels, are used.

Tip: To see Reciprocal values in the Risk Ratio report, right-click in the report and select **Columns > Reciprocal**.

Model Dialog Shows the completed launch window for the current analysis.

Effect Summary Shows the interactive Effect Summary report that enables you to add or remove effects from the model. See *Fitting Linear Models*.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Example Using Multiple Effects and Multiple Levels

This example uses a proportional hazards model for the sample data, VA Lung Cancer.jmp. The data were collected from a randomized clinical trial, where males with inoperable lung cancer were placed on either a standard or a novel (test) chemotherapy (Treatment). The primary interest of this trial was to assess if the treatment type has an effect on survival time; special interest is given to the type of tumor (Cell Type). See Prentice (1973) and Kalbfleisch and Prentice (2002) for more information about this data set.

For the proportional hazards model, covariates include the following:

- Whether the patient had undergone previous therapy (Prior)
- Age of the patient (Age)
- Time from lung cancer diagnosis to beginning the study (Diag Time)
- A general medical status measure (KPS)

Age, Diag Time, and KPS are continuous measures and Cell Type, Treatment, and Prior are categorical (nominal) variables. The four nominal levels of Cell Type include Adeno, Large, Small, and Squamous.

This example illustrates the results for a model with more than one effect and a nominal effect with more than two levels. Risk ratios are also demonstrated, with example calculations for risk ratios for a continuous effect and risk ratios for an effect that has more than two levels.

1. Select **Help > Sample Data Library** and open VA Lung Cancer.jmp.
2. Select **Analyze > Reliability and Survival > Fit Proportional Hazards**.

3. Select Time as **Time to Event**.
4. Select censor as **Censor**.
5. Select Cell Type, Treatment, Prior, Age, Diag Time, and KPS and click **Add**.
6. Click **Run**.
7. (Optional) Click the disclosure icon on the Baseline Survival at mean title bar to close the plot, and click the disclosure icon on Effect Summary to close the report.

Figure 15.6 Report Window for Proportional Hazards Model with Multiple Effects and Levels

Proportional Hazards Fit

Censored By: censor

Effect Summary

Iteration History

Whole Model

		AICc	BIC
Number of Events	128	967.484	989.719
Number of Censorings	9		
Total Number	137		

Model	-LogLikelihood	ChiSquare	DF	Prob>Chisq
Difference	30.7046	61.4091	8	<.0001*
Full	475.1794			
Reduced	505.8840			

Parameter Estimates

Term	Estimate	Std Error	Lower 95%	Upper 95%
Cell Type[Adeno]	0.57719588	0.1849861	0.2036319	0.9316196
Cell Type[Large]	-0.2114757	0.1741197	-0.56618	0.1191116
Cell Type[Small]	0.24538322	0.1592518	-0.070502	0.5550322
Treatment[Standard]	-0.1449679	0.1036051	-0.348984	0.0578618
Prior[No]	-0.0361633	0.1160663	-0.259353	0.1970409
Age	-0.0085494	0.0093042	-0.026447	0.0100846
Diag Time	-0.000092	0.0091251	-0.020233	0.016114
KPS	-0.0326217	0.0055052	-0.043442	-0.021834

Confidence Intervals are profile-likelihood

Covariance of Estimates

Baseline Survival at mean

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Cell Type	3	3	18.6003364	0.0003*
Treatment	1	1	1.96251687	0.1612
Prior	1	1	0.09639285	0.7562
Age	1	1	0.82813759	0.3628
Diag Time	1	1	0.00010177	0.9920
KPS	1	1	34.616163	<.0001*

Note the following about the results:

- In the Whole Model report, the low Prob>Chisq value (<.0001) indicates that there is a difference in survival time when at least one of the effects is included in the model.
- In the Effect Likelihood Ratio Tests report, the Prob>ChiSq values indicate that KPS and at least one of the levels of Cell Type are significant; however, the Treatment, Prior, Age, and Diag Time effects are not significant.

8. Click the Proportional Hazards Fit red triangle and select **Risk Ratios**.

Figure 15.7 Risk Ratios Report

Risk Ratios

Unit Risk Ratios

Per unit change in regressor

Term	Risk Ratio	Lower 95%	Upper 95%	Reciprocal
Age	0.991487	0.97357	1.009733	1.0085861
Diag Time	0.999908	0.982184	1.017952	1.000092
KPS	0.967905	0.957517	0.978405	1.0331596

Range Risk Ratios

Per change in regressor over entire range

Term	Risk Ratio	Lower 95%	Upper 95%	Reciprocal
Age	0.669099	0.283964	1.576584	1.4945466
Diag Time	0.992119	0.213096	4.619042	1.0079435
KPS	0.05484	0.020991	0.143271	18.23482

Risk Ratios for Cell Type

Level1	/Level2	Risk Ratio	Prob>Chisq	Lower 95%	Upper 95%
Large	Adeno	0.4544481	0.0092*	0.2511038	0.8224612
Small	Adeno	0.7176217	0.2286	0.4181316	1.2316242
Squamous	Adeno	0.3047391	<.0001*	0.1690124	0.5494621
Small	Large	1.579106	0.0862	0.9370433	2.6611104
Squamous	Large	0.6705696	0.1574	0.3853373	1.166935
Adeno	Large	2.2004712	0.0092*	1.2158629	3.9824176
Squamous	Small	0.4246514	0.0019*	0.2476225	0.7282408
Adeno	Small	1.3934918	0.2286	0.811936	2.3915917
Large	Small	0.6332697	0.0862	0.375783	1.0671865
Adeno	Squamous	3.2814957	<.0001*	1.8199618	5.916725
Large	Squamous	1.4912695	0.1574	0.8569458	2.5951289
Small	Squamous	2.3548727	0.0019*	1.3731721	4.0384051

Risk Ratios for Treatment

Level1	/Level2	Risk Ratio	Prob>Chisq	Lower 95%	Upper 95%
Test	Standard	1.3363418	0.1617	0.8903074	2.0058347
Standard	Test	0.7483115	0.1617	0.4985456	1.1232076

Risk Ratios for Prior

Level1	/Level2	Risk Ratio	Prob>Chisq	Lower 95%	Upper 95%
Yes	No	1.0750063	0.7554	0.6820551	1.6943478
No	Yes	0.9302271	0.7554	0.5901976	1.4661572

Normal approximations used for ratio confidence limits effects: Cell Type Treatment Prior

Risk Ratios for Multiple Effects and Multiple Levels

Figure 15.7 shows the Risk Ratios for the continuous effects (Age, Diag Time, KPS) and the nominal effects (Cell Type, Treatment, Prior). For illustration, focus on the continuous effect, Age, and the nominal effect with four levels (Cell Type) for the VA Lung Cancer.jmp sample data.

For the continuous effect, Age, in the VA Lung Cancer.jmp sample data, the risk ratios are calculated as follows:

Unit Risk Ratios

$$\exp(\beta) = \exp(-0.0085494) = 0.991487$$

Range Risk Ratios

$\exp[\beta(xmax - xmin)] = \exp(-0.0085494 * 47) = 0.669099$

For the nominal effect, Cell Type, all pairs of levels are calculated and are shown in the Risk Ratios for Cell Type table. Note that for a categorical variable with *k* levels, only *k* -1 design variables, or levels, are used. In the Parameter Estimates table, parameter estimates are shown for only three of the four levels for Cell Type (Adeno, Large, and Small). The Squamous level is not shown, but it is calculated as the negative sum of the other estimates. Here are two example Risk Ratios for Cell Type calculations:

$\text{Large/Adeno} = \exp(\beta_{\text{Large}})/\exp(\beta_{\text{Adeno}}) = \exp(-0.2114757)/\exp(0.57719588) = 0.4544481$

$\text{Squamous/Adeno} = \exp[-(\beta_{\text{Adeno}} + \beta_{\text{Large}} + \beta_{\text{Small}})]/\exp(\beta_{\text{Adeno}})$

$= \exp[-(0.57719588 + (-0.2114757) + 0.24538322)]/\exp(0.57719588) = 0.3047391$

Reciprocal shows the value for 1/Risk Ratio.

Appendix A

References

Abernethy, R. B. (1996). *The New Weibull Handbook*. 2nd ed. North Palm Beach, FL: Robert B. Abernethy.

Akaike, H. (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* AC-19:716–723.

Andrews, D. F., and Herzberg, A. M. (1985). *A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer-Verlag.

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. New York: Springer-Verlag.

Chow, S.-C. (2007). *Statistical Design and Analysis of Stability Studies*. Boca Raton, FL: Chapman & Hall/CRC.

Cox, D. R. (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society, Series B* 34:187–220.

Crow, L. H. (1975). *Reliability Analysis for Complex, Repairable Systems*. Technical Report No. 138, December 1975, US Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, MD.

Crow, L. H. (1982). "Confidence Interval Procedures for the Weibull Process with Applications to Reliability Growth." *Technometrics* 24:67–72.

Escobar, L. A., Meeker, W. Q., Kugler, D. L., and Kramer, L. L. (2003). "Accelerated Destructive Degradation Tests: Data, Models, and Analysis." In *Mathematical and Statistical Methods in Reliability*, edited by B. H. Lindqvist and K. A. Doksum, 319–338. London: World Scientific Publishing Company.

Genschel, U., and Meeker, W. Q. (2010). "A Comparison of Maximum Likelihood and Median-Rank Regression for Weibull Estimation." *Quality Engineering* 22:236–255.

Guo, H., Mettas, A., Sarakakis, G., and Niu, P. (2010). "Piecewise NHPP Models with Maximum Likelihood Estimation for Repairable Systems." In *2010 Proceedings - Annual Reliability and Maintainability Symposium (RAMS).* New York: IEEE Press.

Hosmer, D. W., Jr., and Lemeshow, S. (1999). A*pplied Survival Analysis: Regression Modeling of Time-to-Event Data*. New York: John Wiley & Sons.

ICH Q1E. (2003). *Evaluation for Stability Data*. Tripartite International Conference on Harmonization Guideline Q1E, Geneva. https://database.ich.org/sites/default/files/Q1E_Guideline.pdf.

Kalbfleisch, J. D., and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons.

Kalbfleisch, J. D., and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Kaminskiy, M. P., and Krivtsov, V. V. (2005). "A Simple Procedure for Bayesian Estimation of the Weibull Distribution." *IEEE Transactions on Reliability* 54:612–616.

Klein, J. P., and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. 2nd ed. New York: John Wiley & Sons.

Lee, L., and Lee S. K. (1978). "Some Results on Inference for the Weibull Process." *Technometrics* 20:41–45.

Liu, P., and Wang, P. (2013). "Competing Failure Modes Modeling with Limited Wearout Failures." In *2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS)*. New York: IEEE Press.

Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.

Nair, V. N. (1984). "Confidence Bands for Survival Functions with Censored Data: A Comparative Study." *Technometrics* 26:265–275.

Nelson, W. B. (1982). *Applied Life Data Analysis*. New York: John Wiley & Sons.

Nelson, W. B. (1985). "Weibull Analysis of Reliability Data with Few or No Failures." *Journal of Quality Technology* 17:140–146.

Nelson, W. B. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*. New York: John Wiley & Sons.

Nelson, W. B. (2003). *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*. Philadelphia: Society for Industrial Mathematics.

Nelson, W. B. (2004). *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*. New York: John Wiley & Sons.

Prentice, R. L. (1973). "Exponential Survivals with Censoring and Explanatory Variables." *Biometrika* 60:279–288.

Rigdon, S. E., and Basu, A. P. (2000). *Statistical Methods for the Reliability of Repairable Systems*. New York: John Wiley & Sons.

Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods*. 2nd ed. New York: Springer-Verlag.

SAS Institute Inc. (2020). "The LIFETEST Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/lifetest.pdf.

Si, S., Dui, H., Zhao, X., Zhang, S., and Sun, S. (2012). "Integrated Importance Measure of Component States Based on Loss of System Performance." *IEEE Transactions on Reliability*. 61:192–202.

Tobias, P. A., and Trindade, D. C. (1995). *Applied Reliability*. 2nd ed. New York: Van Nostrand Reinhold.

Tobias, P. A., and Trindade, D. C. (2012). *Applied Reliability*. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.

US Department of Defense (1981). *Military Handbook: Reliability Growth Management* (MIL-HDBK-00189). Washington, DC: US Department of Defense.

Appendix B

Technology License Notices

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

  - You must give any other recipients of the Work or Derivative Works a copy of this License; and
  - You must cause any modified files to carry prominent notices stating that You changed the files; and
  - You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
  - If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
  - You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

Version 16

Consumer Research

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. *JMP® 16 Consumer Research*. Cary, NC: SAS Institute Inc.

JMP® 16 Consumer Research

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

August 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Get the Most from JMP

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

https://www.jmp.com/getstarted

Contents

Consumer Research

1 Learn about JMP 9
Documentation and Additional Resources
Formatting Conventions in JMP Documentation 11
JMP Help 12
JMP Documentation Library 12
Additional Resources for Learning JMP 18
JMP Tutorials 18
Sample Data Tables 18
Learn about Statistical and JSL Terms 19
Learn JMP Tips and Tricks 19
JMP Tooltips 19
JMP User Community 20
Free Online Statistical Thinking Course 20
JMP New User Welcome Kit 20
Statistics Knowledge Portal 20
JMP Training 20
JMP Books by Users 21
The JMP Starter Window 21
JMP Technical Support 21

2 Introduction to Consumer Research 23
Overview of Customer and Behavioral Research Methods

3 Categorical Response Analysis 25
Analyze Survey and Other Counting Data
Example of the Categorical Platform 27
Launch the Categorical Platform 33
Response Roles 33
Columns Roles 38
Other Launch Window Options 38
The Categorical Report 40
Categorical Platform Options 41
Report Options 41

Statistical Testing Options ... 42
Additional Categorical Platform Options ... 44
Crosstab Table Options ... 47
Comparison Letters ... 48
Supercategories ... 49
Supercategories Options ... 50
Additional Examples of the Categorical Platform ... 51
Example of the Test for Response Homogeneity ... 51
Example of the Multiple Response Test ... 52
Example of Supercategories ... 54
Example of the Cell Chisq Test ... 57
Example of Compare Each Sample with Comparison Letters ... 58
Example of Compare Each Cell with Comparison Letters ... 59
Example of User-Specified Comparison with Comparison Letters ... 61
Example of Aligned Responses ... 62
Example of Conditional Association and Relative Risk ... 64
Example of Rater Agreement ... 66
Example of Repeated Measures ... 67
Example of the Multiple Responses ... 68
Example of Mean Score with Comparison Letters ... 73
Example of a Structured Report ... 75
Example of a Multiple Response with Nonresponse ... 76
Set Preferences ... 79
Statistical Details for the Categorical Platform ... 80
Rao-Scott Correction ... 80

4 Choice Models ... 81
Fit Models for Choice Experiments

Overview of the Choice Modeling Platform ... 83
Examples of the Choice Platform ... 85
One Table Format with No Choice ... 85
Multiple Table Format ... 88
Launch the Choice Platform ... 95
Launch Window for One Table, Stacked ... 96
Launch Window for Multiple Tables, Cross-Referenced ... 98
Choice Model Report ... 104
Effect Summary ... 104
Parameter Estimates ... 105
Likelihood Ratio Tests ... 106
Bayesian Parameter Estimates ... 106

Choice Platform Options ... 108
Willingness to Pay ... 111
Additional Examples ... 114
Example of Making Design Decisions ... 114
Example of Segmentation ... 122
Example of Logistic Regression Using the Choice Platform ... 126
Example of Logistic Regression for Matched Case-Control Studies ... 129
Example of Transforming Data to Two Analysis Tables ... 131
Example of Transforming Data to One Analysis Table ... 136
Statistical Details for the Choice Platform ... 138
Special Data Table Rules ... 138
Utility and Probabilities ... 139
Gradients ... 140

5 MaxDiff ... 141
Fit Models for MaxDiff Experiments
Overview of the MaxDiff Modeling Platform ... 143
Examples of the MaxDiff Platform ... 143
One Table Format ... 144
Multiple Table Format ... 147
Launch the MaxDiff Platform ... 150
Launch Window for One Table, Stacked ... 151
Launch Window for Multiple Tables, Cross-Referenced ... 153
MaxDiff Model Report ... 157
Effect Summary ... 157
MaxDiff Results ... 158
Parameter Estimates ... 159
Bayesian Parameter Estimates ... 161
Likelihood Ratio Tests ... 162
MaxDiff Platform Options ... 163
Comparisons Report ... 165
Save Bayes Chain ... 166

6 Uplift ... 167
Model the Incremental Impact of Actions on Consumer Behavior
Overview of the Uplift Platform ... 169
Example of the Uplift Platform ... 170
Launch the Uplift Platform ... 172
The Uplift Model Report ... 173
Uplift Model Graph ... 173
Uplift Report Options ... 176

7 Multiple Factor Analysis 177
Analyze Agreement among Panelists
Overview of the Multiple Factor Analysis Platform 179
Example of Multiple Factor Analysis 180
Launch the Multiple Factor Analysis Platform 183
Data Format 184
The Multiple Factor Analysis Report 186
Summary Plots 186
Consensus Map 187
Multiple Factor Analysis Platform Options 187
Statistical Details for the Multiple Factor Analysis Platform 190
A References 193
B Technology License Notices 195

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation . . . 11
JMP Help . . . 12
JMP Documentation Library . . . 12
Additional Resources for Learning JMP . . . 18
JMP Tutorials . . . 18
Sample Data Tables . . . 18
Learn about Statistical and JSL Terms . . . 19
Learn JMP Tips and Tricks . . . 19
JMP Tooltips . . . 19
JMP User Community . . . 20
Free Online Statistical Thinking Course . . . 20
JMP New User Welcome Kit . . . 20
Statistics Knowledge Portal . . . 20
JMP Training . . . 20
JMP Books by Users . . . 21
The JMP Starter Window . . . 21
JMP Technical Support . . . 21

Formatting Conventions in JMP Documentation

These conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit https://www.jmp.com/software/pro.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from https://community.jmp.com.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Model Screening • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Legacy Control Charts • Pareto Plot • Diagram • Manage Spec Limits • OC Curves

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "JMP Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "JMP Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "JMP New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\16\Samples\Data

On macOS: \Library\Application Support\JMP\16\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit https://jmp.com/tools.

Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > Initial JMP Starter Window**.

JMP Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at https://www.jmp.com/support, including the technical support phone number.

Chapter 2

Introduction to Consumer Research

Overview of Customer and Behavioral Research Methods

- The Categorical platform enables you to tabulate, plot, and compare categorical responses in your data, including multiple response data. You can use this platform to analyze data from surveys and other categorical response data, such as defect records and study participant demographics. Using the Categorical platform, you can analyze responses from data tables that are organized in many different ways. See "Categorical Response Analysis".
- The Choice platform is designed for use in market research experiments, where the ultimate goal is to discover the preference structure of consumers. Then, this information is used to design products or services that have the attributes most desired by consumers. See "Choice Models".
- The MaxDiff platform is an alternative to using standard preference scales to determine the relative importance of items being rated. A MaxDiff experiment forces respondents to report their most and least preferred options, thereby forcing respondents to rank options in terms of preference. See "MaxDiff".
- The Uplift platform enables you to maximize the impact of your marketing budget by sending offers only to individuals who are likely to respond favorably. It can do this even when you have large data sets and many possible behavioral or demographic predictors. You can use uplift models to make such predictions. This method has been developed to help optimize marketing decisions, define personalized medicine protocols, or, more generally, to identify characteristics of individuals who are likely to respond to some action. See "Uplift".
- The Multiple Factor Analysis platform enables you to analyze agreement among panelists in sensory data analysis. You can use MFA to analyze studies where items are measured on the same or different attributes by different instruments, individuals, or under different circumstances. See "Multiple Factor Analysis".

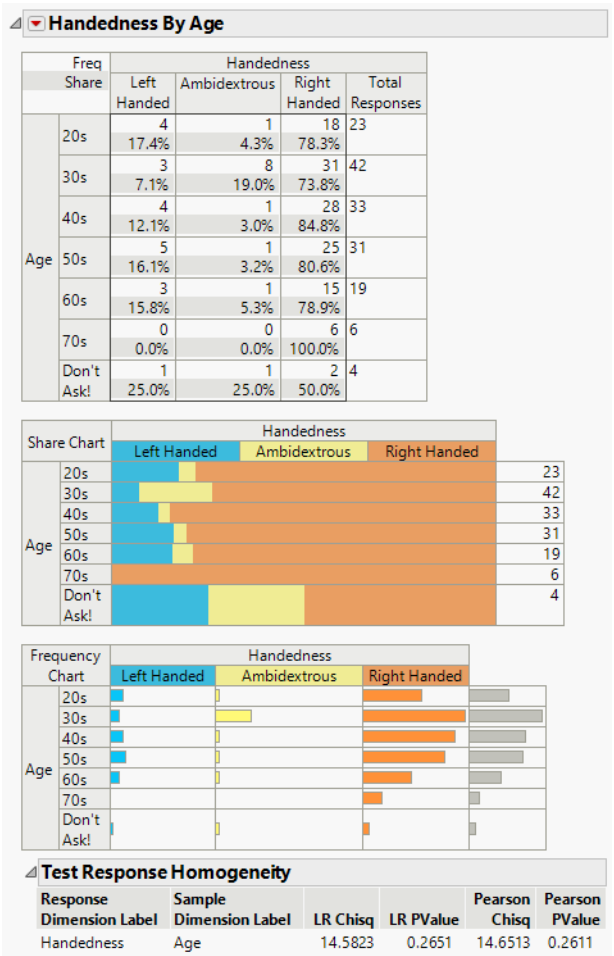
Chapter 3

Categorical Response Analysis

Analyze Survey and Other Counting Data

The Categorical platform enables you to tabulate, chart, and compare categorical response data, including multiple response data. You can analyze data from surveys and other categorical response data, such as defect records and study participant demographics. Many different data types and formats are supported.

Figure 3.1 Categorical Analysis Example



Handedness By Age

Freq / Share		Handedness: Left Handed	Ambidextrous	Right Handed	Total Responses
Age	20s	4 17.4%	1 4.3%	18 78.3%	23
	30s	3 7.1%	8 19.0%	31 73.8%	42
	40s	4 12.1%	1 3.0%	28 84.8%	33
	50s	5 16.1%	1 3.2%	25 80.6%	31
	60s	3 15.8%	1 5.3%	15 78.9%	19
	70s	0 0.0%	0 0.0%	6 100.0%	6
	Don't Ask!	1 25.0%	1 25.0%	2 50.0%	4

Share Chart — Handedness: Left Handed, Ambidextrous, Right Handed

Age: 20s 23; 30s 42; 40s 33; 50s 31; 60s 19; 70s 6; Don't Ask! 4

Frequency Chart — Handedness: Left Handed, Ambidextrous, Right Handed

Age: 20s, 30s, 40s, 50s, 60s, 70s, Don't Ask!

Test Response Homogeneity

Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
Handedness	Age	14.5823	0.2651	14.6513	0.2611

Contents

Example of the Categorical Platform . . . 27
Launch the Categorical Platform . . . 33
Response Roles . . . 33
Columns Roles . . . 38
Other Launch Window Options . . . 38
The Categorical Report . . . 40
Categorical Platform Options . . . 41
Report Options . . . 41
Statistical Testing Options . . . 42
Additional Categorical Platform Options . . . 44
Crosstab Table Options . . . 47
Comparison Letters . . . 48
Supercategories . . . 49
Supercategories Options . . . 50
Additional Examples of the Categorical Platform . . . 51
Example of the Test for Response Homogeneity . . . 51
Example of the Multiple Response Test . . . 52
Example of Supercategories . . . 54
Example of the Cell Chisq Test . . . 57
Example of Compare Each Sample with Comparison Letters . . . 58
Example of Compare Each Cell with Comparison Letters . . . 59
Example of User-Specified Comparison with Comparison Letters . . . 61
Example of Aligned Responses . . . 62
Example of Conditional Association and Relative Risk . . . 64
Example of Rater Agreement . . . 66
Example of Repeated Measures . . . 67
Example of the Multiple Responses . . . 68
Example of Mean Score with Comparison Letters . . . 73
Example of a Structured Report . . . 75
Example of a Multiple Response with Nonresponse . . . 76
Set Preferences . . . 79
Statistical Details for the Categorical Platform . . . 80
Rao-Scott Correction . . . 80

Example of the Categorical Platform

This example uses the Color Preference Survey.jmp sample data table, which contains survey data on color preferences. You can use the Categorical platform to summarize results from different question types:

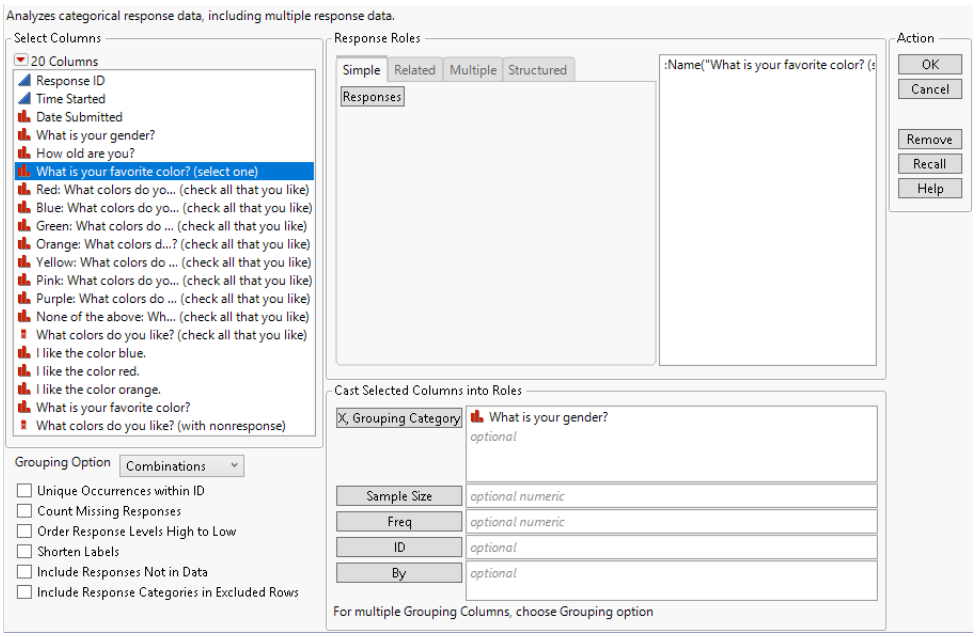
- A single response question
- Three aligned ranking questions
- A multiple response question
- A single response question segmented by a secondary question

You will set up four analyses (one for each type of question) in the Categorical launch window. Alternatively, you can use one launch tab at a time.

1. Select **Help > Sample Data Library** and open Color Preference Survey.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select What is your favorite color? (select one) and click **Responses**.
4. Select What is your gender? and click **X, Grouping Category**.

> **Note:** If you want to run the single analysis of favorite color by gender, click **OK** now.

Figure 3.2 Completed Simple Tab in the Categorical Launch Window

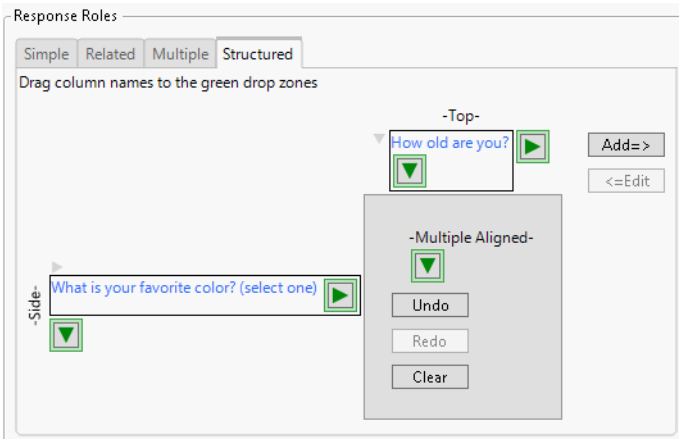


5. Select the **Related** tab.
6. Select I like the color blue. through I like the color orange. and click **Aligned Responses**.

   This enters three questions with rating scales to be analyzed together. The gender grouping category applies to this analysis.
7. Select the **Multiple** tab.
8. Select Red: What colors do you like? (check all that you like) through None of the above: What colors do you like? (check all that you like) and click **Multiple Response**.

   This enters the multiple response question "What colors do you like?" where each response is in an individual column for analysis. The gender grouping category applies to this analysis.
9. Select the **Structured** tab.
10. Drag What is your favorite color? (select one) to the Side green arrow and drag How old are you? to the Top green arrow.

Figure 3.3 Structured Tab in Categorical Launch Window

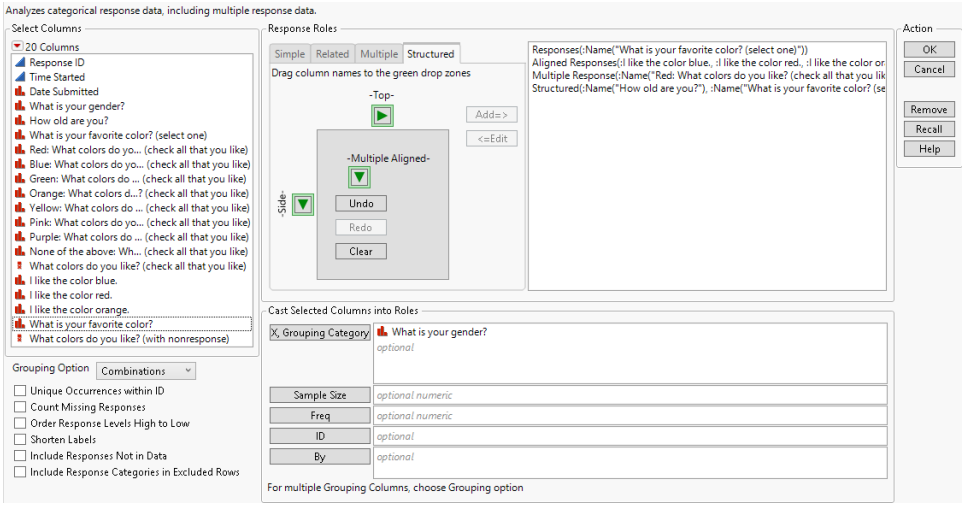


11. Click **Add**.

    The What is your gender? column specified in the X. Grouping Category role does not apply to the structured analysis.

Tip: Click on the green arrows in the structured tab for a column list, where you can click to add a column. You can nest multiple columns by using the down arrow on the top or the right arrow on the side.

Figure 3.4 Completed Launch Window for Full Example



12. Click **OK**.

The results for each analysis are stacked vertically in a single report window.

Simple Tab Report

The first section shows the results for the simple response survey question, “What is your favorite color?”, grouped by gender. Each respondent was asked to select one color from the list: Red, Blue, Green, Orange, Yellow, Pink, Purple, or None of the above.

Figure 3.5 Simple Response: Favorite Color by Gender

Categorical

What is your favorite color? (select one) By What is your gender?

Freq / Share		What is your favorite color? (select one)								
		Blue	Green	None of the above	Orange	Pink	Purple	Red	Yellow	Total Responses
What is your gender?	Female	11 29.7%	4 10.8%	6 16.2%	0 0.0%	1 2.7%	10 27.0%	3 8.1%	2 5.4%	37
	Male	19 44.2%	8 18.6%	5 11.6%	2 4.7%	1 2.3%	3 7.0%	5 11.6%	0 0.0%	43

The analysis shows that blue is the favorite color for both genders. For males, the frequency of the 43 male respondents selecting blue is 19. This corresponds to a share of 44.2% or 19/43. For females, 11 of 37 female respondents (frequency) or 29.7% (share) selected blue.

Tip: You can define the colors in the charts using the Value Colors column property. See *Using JMP*.

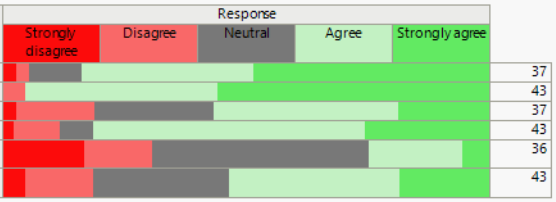
Related Tab Report

The second section of the report shows the results for three related questions that asked respondents to rank how much they liked a color. The responses to these questions are aligned, because the questions used the same rating scale of Strongly agree to Strongly disagree.

Figure 3.6 Aligned Response: Ranking Colors by Gender

Aligned Responses(I like the color blue., I like the color red., I like the color orange.) By What is your gender?

				Freq / Share	Response: Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total Responses
Response	I like the color blue.	What is your gender?	Female		1 2.7%	1 2.7%	4 10.8%	13 35.1%	18 48.6%	37
			Male		0 0.0%	2 4.7%	0 0.0%	17 39.5%	24 55.8%	43
	I like the color red.	What is your gender?	Female		1 2.7%	6 16.2%	9 24.3%	14 37.8%	7 18.9%	37
			Male		1 2.3%	4 9.3%	3 7.0%	24 55.8%	11 25.6%	43
	I like the color orange.	What is your gender?	Female		6 16.7%	5 13.9%	16 44.4%	7 19.4%	2 5.6%	36
			Male		2 4.7%	6 14.0%	12 27.9%	15 34.9%	8 18.6%	43



The analysis shows that blue is the color with the highest rankings for both males and females. Orange has a high number of neutral results for both genders as compared to red and blue. The value colors column property is used to define the colors used in the share chart.

Note: If you analyze aligned questions individually using the Simple tab, you obtain three individual reports containing the same information as the related report. Using the Related tab aligns the reports so that the results are easier to compare to one another.

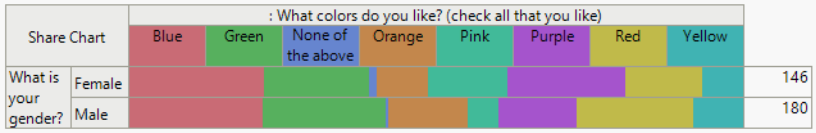
Multiple Tab Report

The third section in the report shows the results from a multiple choice survey question grouped by gender. The question was “What colors do you like? (check all that you like)”. Each possible response was collected in an individual column. If the subject selected a response, there is a value in the corresponding column. Otherwise, the column is empty.

Figure 3.7 Multiple Response: Colors Liked by Gender

Structured

Freq / Share / Rate		: What colors do you like? (check all that you like)										
		Blue	Green	None of the above	Orange	Pink	Purple	Red	Yellow	Total Responses	Total Cases	Total Cases Responding
What is your gender?	Female	32 21.9% 86.5%	25 17.1% 67.6%	2 1.4% 5.4%	12 8.2% 32.4%	19 13.0% 51.4%	28 19.2% 75.7%	18 12.3% 48.6%	10 6.8% 27.0%	146	37	37
	Male	39 21.7% 90.7%	36 20.0% 83.7%	1 0.6% 2.3%	23 12.8% 53.5%	9 5.0% 20.9%	23 12.8% 53.5%	34 18.9% 79.1%	15 8.3% 34.9%	180	43	43



From the results, we observe that blue and green are highly liked by both genders. Females like pink and purple at higher rates than males.

In a multiple response question, more than one answer is allowed. A case represents a single responder, and Total Cases is the total number of responders. In our data, we have 37 female cases and 43 male cases. The total number of cases responding are the number of cases who selected one or more responses (which in our data, was everyone).

The Total Responses column lists the total number of selections made. For females, there were 146 colors selected by the 37 females. For males, there were 180 colors selected by the 43 male responders. The tabulation includes the number of responders selecting each color, the share or the percent of the total responses, and the rate or the percent of the total cases.

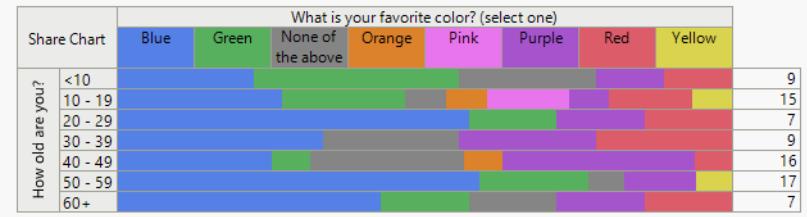
Structured Tab Report

The final section shows the table generated from the structured tab where you set the favorite color question as the response of interest grouped by the age group. In the structured format, the levels of the primary response of interest form the rows of the table and the levels of the grouping variable form the columns.

Figure 3.8 Structured: Favorite Color by Age

What is your favorite color? (select one) By How old are you?

What is your favorite color? (select one)	Freq / Share	<10	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60+
	Blue	2 22.2%	4 26.7%	4 57.1%	3 33.3%	4 25.0%	10 58.8%	3 42.9%
	Green	3 33.3%	3 20.0%	1 14.3%	0 0.0%	1 6.3%	3 17.6%	1 14.3%
	None of the above	2 22.2%	1 6.7%	0 0.0%	2 22.2%	4 25.0%	1 5.9%	1 14.3%
	Orange	0 0.0%	1 6.7%	0 0.0%	0 0.0%	1 6.3%	0 0.0%	0 0.0%
	Pink	0 0.0%	2 13.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
	Purple	1 11.1%	1 6.7%	1 14.3%	2 22.2%	5 31.3%	2 11.8%	1 14.3%
	Red	1 11.1%	2 13.3%	1 14.3%	2 22.2%	1 6.3%	0 0.0%	1 14.3%
	Yellow	0 0.0%	1 6.7%	0 0.0%	0 0.0%	0 0.0%	1 5.9%	0 0.0%
	Total Responses	9	15	7	9	16	17	7

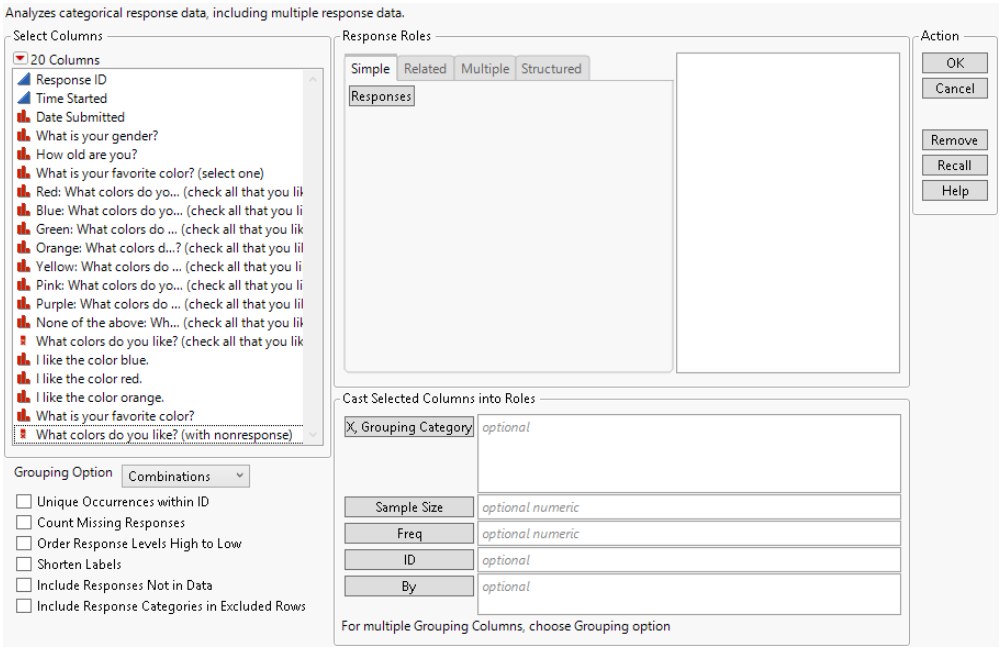


The 10 to 19 age group has two responders with pink as their favorite color. No other age groups have responders with pink as their favorite color. You can see from the bottom row of the table that the number of responders in each age group is small. You could gather more data to draw conclusions about the favorite colors across age groups.

Launch the Categorical Platform

Launch the Categorical platform by selecting **Analyze > Consumer Research > Categorical**.

Figure 3.9 Categorical Platform Launch Window for the Simple Tab



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Response Roles

The launch window includes tabs for three groups of response roles (Simple, Related, and Multiple) and a Structured tab where you can create custom data summaries. The response role corresponds to the type of responses you want to analyze. Options on each tab correspond to how the responses are organized in your data table.

Simple Tab

Use the Simple tab to analyze a single response, such as a survey question where only a single answer is allowed. The data to be analyzed is contained in a single column.

Responses Summarizes data from a single column. If multiple columns are selected, the categorical report contains a separate report for each individual column.

Related Tab

Use the Related tab to analyze a set of related responses across multiple columns.

Figure 3.10 Categorical Platform Launch Window Related Tab

Aligned Responses Summarizes data from multiple columns that have the same response levels in a single report. This option is useful for survey data where you have many questions with the same set of responses. You can quickly summarize and compare response trends for all of the questions at once.

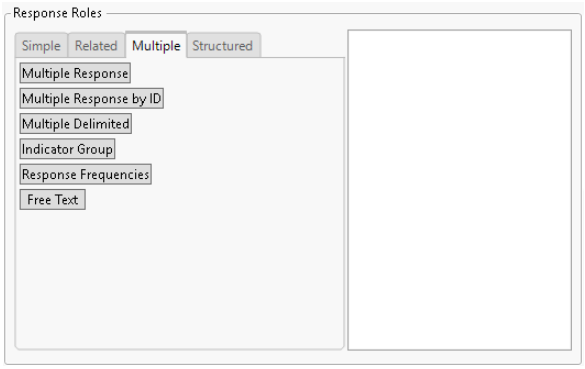
Repeated Measures Summarizes data from multiple columns where each column contains responses to the same question made at different time points. If an individual responds at multiple time points, the samples are called overlapping. When there are overlapping samples, the Kish correction is used. See Kish (1965, sec. 12.4).

Rater Agreement Summarizes data from multiple columns where each column is a rating for the same question or item, but is given by a different individual (rater).

Multiple Tab

Use the Multiple tab to analyze multiple responses where the responses are recorded in one or more columns. The options on the Multiple tab are specific to how the data is organized in your data table. A set of multiple responses could be from a survey where the response set allows for more than one choice (check-all-that apply questions), or from a set of defect data where an item can have multiple defects.

Figure 3.11 Categorical Platform Launch Window Multiple Tab



Multiple Response Summarizes data from multiple responses where each possible response is recorded in its own individual column. Each column can contain blanks, which correspond to the item not being selected.

Figure 3.12 Multiple Response Column Format

	Red: What colors do you like? (check all that you like)	Blue: What colors do you like? ...	Green: What colors do you ...	Orange: What colors do you ...	Yellow: What colors do you ...	Pink: What colors do you like? ...	Purple: What colors do you ...	None of the above: What colors do you like? (check all that you like)
1	Red	Blue				Pink	Purple	
2	Red	Blue	Green					
3	Red	Blue	Green	Orange			Purple	
4								None of the above
5		Blue	Green	Orange			Purple	
6	Red	Blue	Green				Purple	
7	Red	Blue		Orange				
8	Red		Green	Orange				
9	Red	Blue	Green				Purple	
10	Red	Blue	Green	Orange	Yellow	Pink	Purple	

Multiple Response by ID Summarizes data from multiple responses where the data are recorded in a stacked format. There is a single column of responses with a second column containing an ID for the subject.

Figure 3.13 Multiple Response by ID Format

	Response ID	What colors do you like?
1	2	Red
2	2	Blue
3	2	
4	2	
5	2	
6	2	Pink
7	2	Purple
8	2	
9	3	Red
10	3	Blue

Multiple Delimited Summarizes data from multiple responses where the responses are in a single column and each response is separated by a comma, semicolon, or tab.

Figure 3.14 Delimited Multiple Response Format

	What colors do you like? (check all that you like)
1	Red, Blue, , , , Pink, Purple,
2	Red, Blue, Green, , , , ,
3	Red, Blue, Green, Orange, , , Purple,
4	, , , , , , , None of the above
5	, Blue, Green, Orange, , , Purple,
6	Red, Blue, Green, , , , Purple,
7	Red, Blue, , Orange, , , ,
8	Red, , Green, Orange, , , ,
9	Red, Blue, Green, , , , Purple,
10	Red, Blue, Green, Orange, Yellow, Pink, Purple,

Tip: Use the Multiple Responses column property for columns that contain delineated multiple responses. See *Using JMP*.

Indicator Group Summarizes multiple responses that are stored in indicator columns. The data table contains a column for each possible response, and each column is an indicator (for example, 0 and 1). A blank value indicates a missing response. See "Indicator Group".

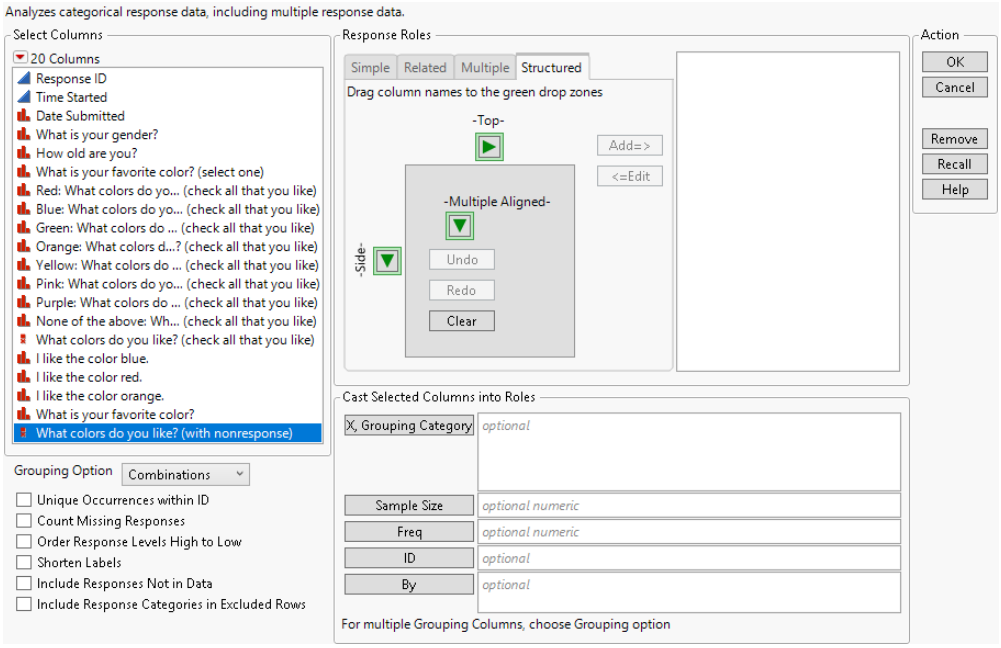
Response Frequencies Summarizes multiple responses that are stored in columns that contain frequency counts. This data format is the summarized version of the Indicator Group format. See "Response Frequencies".

Free Text Summarizes text data. The Free Text option launches a Text Explorer report inside the Categorical report window. See *Basic Analysis*.

Structured Tab

Use the Structured tab to construct custom cross tabulations.

Figure 3.15 Categorical Platform Launch Window for the Structured Tab



The Structured tab has three drop zones for assigning columns to roles. Drag column names to the green drop zone arrows. Alternatively, click the green arrows for a column list to select columns. The resulting structured table considers the innermost terms on the side of the table as responses and all other terms as grouping factors.

Drop Zones

Top Assigns one or more data table columns to the columns of the cross tabulation table. After one column has been assigned, use the down arrows to nest additional columns within already assigned columns or use the right arrows to add additional column groups.

Side Assigns one or more data table columns to the rows of the cross tabulation table. After one column has been assigned, use the right arrows to nest additional columns nested within already assigned rows or use the down arrows to add additional row groups.

Multiple Aligned Assigns two or more columns with aligned response.

Controls

Undo Click to undo column assignments.

Redo Click to redo the most recent assignment that was undone.

Clear Click to remove all assignments in the drop zones.

Add=> Click to add the constructed table to the structured table list.

<=Edit Click to make changes to the selected table from the structured table list.

Columns Roles

The categorical platform launch window has the following column roles:

X, Grouping Category (Not applicable for the Structured tab.) Assigns a column as a grouping category. The responses are summarized for each group. If more than one grouping column is used, then the tabulation is nested by default. Use the Grouping Option in the launch window to change the summarization.

Sample Size Assigns a column whose values define the number of individual units in the group to which that frequency is applicable. The sample size is used for multiple response roles with summarized data. See "Example of the Multiple Responses".

Freq Assigns a column whose values define a frequency to each row for the analysis. The frequency role is used for summarized data.

ID Assigns a column that identifies the respondent. This option is required when Multiple Response by ID is selected, and it is not used for any other response types.

By Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

Other Launch Window Options

Additional options are located in the lower left of the launch window. Alternatively, these options can be selected from the Categorical red triangle menu in the platform report window.

Grouping Option Defines how to use grouping variables in the analysis when more than one grouping column is specified.

Combinations Analyzes the response for combinations of the grouping variables. The first column in the grouping list is the outermost group in the cross tabulation table.

Each Individually Analyzes the response for each grouping variable individually.

Both Provides reports for combinations of the grouping variables as well as for each grouping variable individually.

Unique Occurrences within ID Limits the counts to unique response levels within a participant. Specify a column as the ID using the ID role. This option is applicable only when Multiple Response by ID is selected, and it is not used for any other response types.

Count Missing Responses Specifies that missing values be included as a category. Missing values can be either empty cells or a defined missing code in the Missing Value Codes column property. If a column contains only missing values, the missing values are counted regardless of this option. For multiple responses, a response is considered missing if all response categories are missing or for indicators, if all are zero.

Note: If this option is not selected, missing values are excluded from the analysis.

Order Response Levels High to Low Orders the responses from high to low. (The default ordering is low to high.) This option applies only to the response, and does not apply to grouping categories.

Tip: Use the Value Order column property to define a specific category ordering. See *Using JMP*.

Shorten Labels Shortens value labels by removing prefixes and suffixes that are common to all labels.

Note: This option applies only to value labels, and does not apply to column names.

Include Responses Not in Data Specifies that categories with no responses be included in the report. The categories with no responses must be specified in the Value Labels column property. This option applies only to responses. For grouping variables tabulations, include only categories with responses.

Include Response Categories in Excluded Rows Specifies that response categories that appear only in excluded rows be included in the report. The counts for these categories are zero.

The Categorical Report

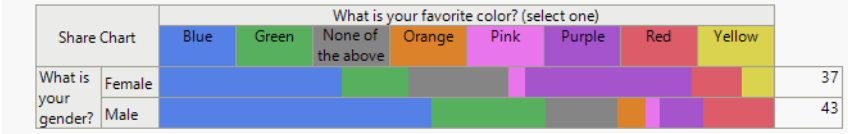
The initial Categorical report shows a cross tabulation and a share chart for each set of selected responses.

Figure 3.16 The Initial Categorical Report

Categorical

What is your favorite color? (select one) By What is your gender?

Freq / Share		What is your favorite color? (select one)								
		Blue	Green	None of the above	Orange	Pink	Purple	Red	Yellow	Total Responses
What is your gender?	Female	11 29.7%	4 10.8%	6 16.2%	0 0.0%	1 2.7%	10 27.0%	3 8.1%	2 5.4%	37
	Male	19 44.2%	8 18.6%	5 11.6%	2 4.7%	1 2.3%	3 7.0%	5 11.6%	0 0.0%	43



The upper left corner of the table lists the quantities (Freq, Share, and Rate when applicable) that are included in each cell of the table. Remove or add these quantities using the options in the Categorical red triangle menu.

- The Frequencies count (labeled Freq) is provided for each category with the total frequency (Total Responses) at the right of the table. When there are multiple responses, the summary columns at the right of the table also include the number of cases or the number of rows (Total Cases) and the number of responders (Total Cases Responding).
- The Share of Responses (Share) is determined by dividing each count (Freq) by the total number of responses.
- The Rate is the frequency of response (Freq) divided by the total number of cases (Total Cases). This quantity appears only for multiple responses and is not shown in Figure 3.16.

In Figure 3.16, the number of responses to the question What is your favorite color? are tabulated by gender. Consider the first row of the table with the results for females:

- The first cell of the table contains 11 responses. This is the count of the female respondents who selected blue as their favorite color.
- There are 37 total responses for females. Of the 37 female responses, 29.7% (11/37) selected blue as their favorite color.

Categorical Platform Options

The Categorical red triangle menu contains options that enable you to customize the report and request various statistical tests. The specific options that are available are determined by the response roles, the use of grouping categories, and the options selected in the launch window.

- "Report Options"
- "Statistical Testing Options"
- "Additional Categorical Platform Options"
- "Crosstab Table Options"

Report Options

The following options enable you to customize the appearance of the report:

Cross Tabulation Options

Frequencies Shows or hides the frequency (labeled Freq) in the Crosstab table. The frequency is the count of the responses in each category.

Share of Responses Shows or hides the share of responses (labeled Share) in the Crosstab table. The share of responses is the percent of responses in each category.

Rate Per Case (Available only for multiple responses.) Shows or hides the rate per case (labeled Rate) in the Crosstab table. The rate per case is the percent of responses in each category based on the total number of cases (regardless of whether they were a respondent).

Rate per Case Responding (Available only for multiple responses.) Shows or hides the Rate per Case Responding in the Crosstab table. The rate per case responding is the percent of responses in each category based on the cases that responded.

Chart Options

Share Chart Shows or hides a divided bar chart. The bar length is proportional to the percentage of responses for each type. The column on the right shows the number of responses in each grouping category. If no grouping category is used, the column on the right shows the total number of responses.

Frequency Chart Shows or hides a Frequency Chart. The bars reflect the frequency of responses within each group. The scale is consistent across the chart. The gray bars at the far right represent the total number of responses in each grouping category.

Transposed Freq Chart Shows or hides a transposed Frequency Chart. The bars reflect the frequency of responses within each group. The responses are the rows, and the grouping levels are the columns in chart. The totals for each grouping level are represented by gray bars in the bottom row of the chart.

Tip: You can change the colors in the share chart using the Value Colors column property. See *Using JMP*.

Crosstab Viewing Options

Crosstab Shows or hides the Crosstab table. The Crosstab table displays the response categories as column headings and the grouping levels (when used) as row labels. The upper left cell of the table shows the labels for the items in each cell of the table (Freq, Share, Rate, and Rate per Case Responding). If the report contains a transposed Crosstab table, this option also removes the transposed Crosstab table from the report.

Crosstab Transposed Shows or hides a transposed Crosstab table. The transposed Crosstab table displays the response categories as row labels and the grouping levels (when used) as column headings. The upper left cell of the table shows the labels for the items in each cell of the table (Freq, Share, Rate, and Rate per Case Responding). If the report contains a Crosstab table, this option also removes the Crosstab table from the report.

Statistical Testing Options

The statistical testing options that are available depend on the response roles and the use of grouping variables in the analysis. Options include tests for multiple responses, response homogeneity, association, relative risk, and agreement. Options depend on both the response data type and the grouping variable (X, Grouping Category) data type.

Test Multiple Response (Available only for multiple response data with one or more grouping categories.) See "Example of the Multiple Response Test". Contains tests for independence of responses across each grouping category:

Count Test, Poisson Shows or hides a test of independence of rates that uses Poisson regression. The frequency per unit is modeled by the sample categorical variable. The result is a likelihood ratio chi-square test of whether the rate of each individual response differs across grouping levels.

Homogeneity Test, Binomial Shows or hides the likelihood ratio chi-square test of independence for each individual response level. Each response category has a binomial distribution (selected or not selected).

Exclude Nonresponses Excludes nonresponses for count and homogeneity tests of multiple response categories. If a row is missing data in all response categories, it is treated as a nonresponse.

Test Response Homogeneity (Available only for response variables that do not have a multiple response modeling type and when one or more groupings variables are specified.) Shows or hides a report that contains tests for response homogeneity that depend on your grouping variable:

- When the grouping variable is not a multiple response, then one tests for independence of the response across the grouping variable levels. The likelihood ratio and Pearson chi-square tests are provided. See "Example of the Test for Response Homogeneity".
- When the grouping variable is a multiple response, then one tests for independence of the response across *each* of the grouping variable levels. A Rao-Scott Chi-square test is provided.

Cell Chisq Shows or hides *p*-values for each cell in the table for a chi-square test of independence. A small *p*-value indicates a cell with an observed value that is larger or smaller than expected under the assumption that the rows are independent of the columns. The *p*-values are colored and shaded according to whether the count is larger or smaller than expected. See "Example of the Cell Chisq Test".

Compare Each Sample (Available only for single responses with one or more grouping variables.) Shows or hides a report that contains pairwise likelihood ratio and Pearson chi-square tests for independence of responses across levels of a grouping variable. See "Example of Compare Each Sample with Comparison Letters".

Compare Each Cell (Available only for single and multiple responses with one or more grouping variables.) Shows or hides pairwise likelihood ratio chi-square, Pearson chi-square, and Fisher's exact tests for independence of each level of the response versus all other levels combined across levels of a grouping variable. See "Example of Compare Each Cell with Comparison Letters".

Relative Risk (Available when the grouping variable has two levels and either the response has two levels or is a multiple response and the Unique occurrences within ID option has been selected.) Shows or hides the relative risks for a two-level grouping variable for each level of the response. See "Example of Conditional Association and Relative Risk".

Conditional Association (Available only when the Unique occurrence within ID option has been selected.) Shows or hides the conditional probability of one response level given a second response level. See "Example of Conditional Association and Relative Risk".

Agreement Statistic (Available only for Rater Agreement responses.) Shows or hides the Kappa coefficient of agreement and the Bowker test of symmetry. See "Example of Rater Agreement".

Transition Report (Available only for Repeated Measures responses.) Shows or hides transition counts and rates matrices for changes in responses across time. See "Example of Repeated Measures".

Test Options Options available in this menu depend on your selected analysis.

ChiSquare Test Choices Specifies which chi-square tests of homogeneity are calculated for single responses. You can choose between Both LR and Pearson, LR Only, or Pearson Only, where LR refers to likelihood ratio.

Show Warnings Shows warnings for small sample sizes in chi-square tests.

Order by Significance Reorders the reports so that the most significant reports are at the top.

Hide Nonsignificant Suppresses reports that are non-significant.

FDR Adjusted PValues Shows or hides false discovery rate *p*-values based on the method of Benjamini and Hochberg (1995).

Additional Categorical Platform Options

The following options enable you to add summary statistics to the report, save reports, and set report formats.

Summary Statistics Options

Total Responses Shows or hides the sum of the frequency counts for the response in crosstab tables and share charts. When a grouping variable is specified, the total is across each grouping category.

Response Levels Shows or hides the categories for the response column in crosstab tables and share charts.

Show Supercategories (Available only when one or more supercategories is defined.) Shows or hides columns for supercategories in the crosstab table and the Frequency Chart. For more information about supercategories, see "Supercategories".

Tip: This option shows or hides the Supercategories. To hide the individual categories within the supercategory, use the Hide option in the Supercategories column property. Alternatively, use the Response Levels option to hide all response levels so that only Supercategories remain unhidden.

Total Cases (Available only for multiple response columns.) Shows or hides a column in the crosstab table that contains the number of cases (participants) in each group.

Total Cases Responding (Available only for multiple response columns.) Shows or hides a column in the crosstab table that contains the number of cases (participants) who responded at least once. Participants who did not respond at all are not included. The total cases responding is less than or equal to the total cases.

Mean Score Shows or hides a column in the crosstab table and share chart that contains the overall mean of the response or the mean for each grouping category. The mean is calculated based on a numerical value assigned to each response category.

- For numeric categories, the numeric value is the actual value.
- For non-numeric categories the value is the value assigned to the categories by the Value Scores column property.
- For categories without value scores, the value is based on a default assignment of 1 to the number of categories.

See "Example of Mean Score with Comparison Letters".

Mean Score Comparisons Shows or hides the Compare Means column in the crosstab table. This column compares the mean scores across grouping categories using the unpooled Satterthwaite *t* test for pairwise comparisons. See the TTEST Procedure chapter in SAS Institute Inc. (2020). The results of the comparison are shown using letters. For more information about comparison letters, see "Comparison Letters". For more information about specifying comparison groups, see "Example of User-Specified Comparison with Comparison Letters".

Std Dev Score Shows or hides a column in the crosstab table that contains the overall standard deviation of the response or the standard deviation of each grouping category.

Order by Mean Score (Appears only when more than one response is specified and there are no grouping variables in the analyses.) Orders the response reports by the mean score.

Save Options

Save Tables Contains options to save specific portions of the reports to a new data table. Each option creates an individual data table for each report. The options available in this menu depend on your selected analysis. The saved tables all include a Source script.

Note: Supercategories are not included in the new tables.

Save Frequencies Saves the frequency counts from the crosstab table to a new data table.

Save Share of Responses Saves the share of responses from the crosstab table to a new data table.

Save Contingency Table Saves the complete crosstab table to a new data table.

Save Rate Per Case Saves the rates per case from the crosstab table to a new data table.

Save Transposed Frequencies Saves the transposed frequency counts from the crosstab table to a new data table.

Save Transposed Share of Responses Saves the transposed share of responses from the crosstab to a new data table.

Save Transposed Rate Per Case Saves the transposed rates per case from the crosstab table to a new data table.

Save Test Rates Saves the results of the Test Multiple Response option to a new data table.

Save Test Homogeneity Saves the results of the Test Response Homogeneity option to a new data table.

Save Mean Scores Saves the mean scores for each sample group to a new data table.

Save tTests and pValues Save *t* tests and *p*-values from the Mean Score Comparisons report to a new data table.

Save Excel File (Available only on Windows.) Creates a Microsoft Excel spreadsheet with the structure of the crosstab table. This option maps all of the tables to one sheet, with the response categories as rows, the sample levels as columns, and sharing the headings for sample levels across multiple tables. When there are multiple elements in each table cell, you have the option to make them multiple or single cells in Microsoft Excel.

Report Format Options

Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Contents Summary Shows or hides a Contents Summary report at the top of the Categorical report. The Contents Summary report contains all of the tests and mean scores in a summary that has links to the associated report.

Show Columns Used in Report (Available only with SPSS or SAS names.) Shows or hides Columns Used in Report information. This option affects only columns that have an SPSS or SAS Name or SPSS or SAS Label column property.

Tip: When you import survey data from SAS or SPSS, the Name and Label column properties are automatically added to your JMP table. You can add a SAS or SPSS Name or Label column property using the Other column property. For example, if you use the SAS or SPSS Name column property to store a survey question, the column name can be a short name.

Format Elements Enables you to specify formats for Frequencies, Shares, Rates, and Means.

Arrange in Rows Enables you to arrange multiple reports across the window. Enter the number of reports that you want to view across the window.

Set Preferences Enables you to set preferences for future launches of the Categorical platform in the current JMP session and in future JMP sessions. See "Set Preferences".

Category Options Contains options (Grouping Option, Count Missing Responses, Order Response Levels High to Low, Shorten Labels, and Include Responses Not in Data) that are also available on the launch window. If these options are selected here, the platform updates with the new setting. For more information about the Category Options, see "Other Launch Window Options".

Force Crosstab Shading Forces shading on Crosstab reports even if the preference is set to no shading. If this option is not selected, the Crosstab reports are shaded according to the current setting of the Shade Alternate Table Rows preference.

Relaunch Dialog Enables you to return to the launch window and edit the specifications for an analysis.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Crosstab Table Options

Show Letters Shows or hides the column letter IDs in the Crosstab table. These letters are used in many of the tests of homogeneity and are displayed automatically for those tests.

Specify Comparison Groups Enables you to specify specific comparison groups for tests of homogeneity. Use group comparison letters separated by a slash to represent each group. Separate multiple groups by commas. For example, to test A with E, B with D, and C with F, specify the groups as "A/E, B/D, C/F". A Compare Each Cell report is provided for the

defined comparison groups. See "Example of User-Specified Comparison with Comparison Letters".

Remove Removes the report from the report window.

Caution: The Remove option cannot be undone.

Comparison Letters

The Compare Each Cell, Compare Each Sample, and Mean Score Comparisons options use comparison letters to identify sample levels. For more than 26 levels, numbers are appended to the letters. The letters are shown to the right of the sample level headings of the crosstab table when a comparison option is turned on. These letters are used to identify levels in the Compare column.

Figure 3.17 Crosstab Table with Comparison Letters

country By type

			Freq / Share	country: American	country: European	country: Japanese	Total Responses	Compare
type	Family	A		74 47.7%	15 9.7%	66 42.6%	155	B
	Sporty	B		23 23.0%	21 21.0%	56 56.0%	100	c
	Work	C		18 37.5%	4 8.3%	26 54.2%	48	*

Default Comparison Groups: A/B/C

Shows letter of the category it is significantly different from at the higher share level

* Base count warning	100	Uppercase Alpha Level	0.05
** Base count minimum	30	Lowercase Alpha Level	0.1

A letter in the Compare column indicates a difference between two levels. The row containing the letter is one level and the letter in the Compare column indicates the second level. When two sample levels are significantly different, the letter of the sample level with a smaller share of responses is placed into the comparison cell of the other level. An uppercase letter indicates a stronger difference between levels than a lowercase letter. The default alpha level (significance level) for an uppercase letter is 0.05 and 0.10 for a lowercase letter. For example, in Figure 3.17, notice the following:

- The first row of the table for Family cars has a B in the Compare column. The letter B is associated with Sporty cars. This indicates that there is a difference in the country of origin for Sporty and Family cars at the 0.05 significance level. The B is in the row for Family cars

because the total responses for Sporty cars (100) is less than the total responses for Family cars (155).

- The c in the Compare column of the Sporty row indicates a significant difference at the 0.10 level between the country of origin when comparing Sporty to Work cars. The c is in the row for Sporty cars because the total responses for Sporty cars (100) is greater than the total responses for Work cars (48).

Warnings for small counts are indicated by asterisks in the comparison cells. One asterisk indicates that the level has fewer than 100 responses. Two asterisks indicate fewer than 30 responses. In Figure 3.17, notice that the row for Work has an asterisk in the Compare column. This warning that the count of responses is small appears because the total responses for Work cars (48) is less than100.

You can change the alpha levels and thresholds for the warning counts in the Categorical platform preferences. For more information about changing preferences, see "Set Preferences".

Tip: If you want only one set of comparison letters in your report, set the **Lowercase Alpha Level** to 0 in the preferences.

The following examples illustrate the use of comparison letters:

- "Example of Compare Each Sample with Comparison Letters"
- "Example of Compare Each Cell with Comparison Letters"
- "Example of Mean Score with Comparison Letters"

Supercategories

The term *supercategories* refers to the aggregation of response categories. For example, when using a five-point rating scale, you might want to know the percent of responses from the top two ratings (top two boxes). Use the Supercategories column property to define groups of responses.

When you add a Supercategories column property to a response column, no additional columns are added to your data table. Instead, the Supercategories column property adds an additional category column to crosstab tables and Frequency Charts in the Categorical platform report. You can create multiple supercategories for a single response column. Share Charts do not show supercategories, and supercategories are not applied to grouping columns.

To create a supercategory:

1. Select a column in your data table that contains categories that you would like to aggregate.
2. Select **Cols > Column Info**.
3. Click **Column Properties** and select **Supercategories**.
4. (Optional) To change the default name of the supercategory, enter a Supercategory Name.
5. Select one or more categories from the Column's Categories list.
6. Click **Add**.
7. (Optional) Select the supercategory and click the Supercategories red triangle for additional options.

Supercategories Options

The following options are available in the Supercategories red triangle menu in the Column Properties window:

Hide Hides categories within a supercategory in the crosstab table and frequency chart.

> **Tip:** If you want the flexibility to show or hide the individual categories in your reports, then do not use the Hide option. Use the Response Level option in the Categorical red triangle menu.

Net (Available only for a multiple response column.) Prevents individual respondents from being counted twice when they appear in more than one supercategory.

Add Mean Includes mean statistics in the report.

Add Std Dev Includes standard deviation statistics in the report.

Add All Includes total responses in the report. By default, the Total Responses column is always included.

Note: Supercategories are supported for all response effects except Repeated Measures and Rater Agreement.

Additional Examples of the Categorical Platform

- "Example of the Test for Response Homogeneity"
- "Example of the Multiple Response Test"
- "Example of Supercategories"
- "Example of the Cell Chisq Test"
- "Example of Compare Each Sample with Comparison Letters"
- "Example of Compare Each Cell with Comparison Letters"
- "Example of User-Specified Comparison with Comparison Letters"
- "Example of Aligned Responses"
- "Example of Conditional Association and Relative Risk"
- "Example of Rater Agreement"
- "Example of Repeated Measures"
- "Example of the Multiple Responses"
- "Example of Mean Score with Comparison Letters"
- "Example of a Structured Report"
- "Example of a Multiple Response with Nonresponse"

Example of the Test for Response Homogeneity

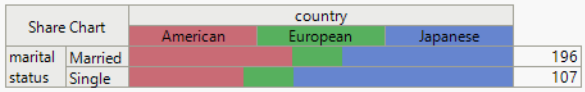
This example uses the Car Poll.jmp sample data table, which contains data collected from a survey about car ownership. The data include demographics about the individuals polled and information about their car. You want to explore the relationship between marital status and the origin of the car. You also want to test for the homogeneity of the responses. That is, you want to test to see whether the distribution of the origin of cars is the same for married and single respondents.

1. Select **Help > Sample Data Library** and open Car Poll.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select country and click **Responses** on the Simple tab.
4. Select marital status and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Test Response Homogeneity**.

Figure 3.18 Test Response Homogeneity

country By marital status

Freq / Share		country: American	country: European	country: Japanese	Total Responses
marital status	Married	83 42.3%	26 13.3%	87 44.4%	196
	Single	32 29.9%	14 13.1%	61 57.0%	107



Test Response Homogeneity

Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
country	marital status	5.14007	0.0765	5.08146	0.0788

The Share Chart indicates that the married group is evenly split between ownership of American and Japanese cars. In the single group, Japanese cars are the most frequently owned.

The test for response homogeneity provides results from two versions of the test. The Pearson Test and the Likelihood Ratio Test both have chi-square test statistics and associated *p*-values. The test for response homogeneity has a *p*-value of about 0.08 for either method.

Example of the Multiple Response Test

This example uses the Consumer Preferences.jmp sample data table. This table contains data from a survey about people's attitudes and opinions, as well as questions concerning oral hygiene. You can use the Test Multiple Response option to test if the response rates for each brushing time (Brush Delimited) are the same across groups (Brush). The groups are defined by the frequency that responders brush their teeth.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Scroll to the right until you see the Brush Delimited column.

Figure 3.19 Consumer Preferences Data Table

	Brush After Waking Up	Brush After Meal	Brush Before Sleep	Brush Another Time	Brush Other	Brush Delimited
1	1	0	0	0		Wake,
2	0	1	0	0		After Meal,
3	1	0	1	0		Wake,Before Sleep,
4	0	1	0	0		After Meal,
5	1	0	1	0		Wake,Before Sleep,
6	1	1	1	0		Wake,After Meal,Before Sleep,
7	1	1	1	0		Wake,After Meal,Before Sleep,

Note that the Brush Delimited column contains the responses to a multiple response question. Each response is separated by a comma. The four columns preceding Brush Delimited contain the same information in a different data format. Each column (Brush after Waking Up, Brush After Meal, Brush Before Sleep, and Brush Another Time) contains one response. If the response was selected the column value is a 1, and otherwise it is 0.

3. Select **Analyze > Consumer Research > Categorical**.
4. Select the **Multiple** tab.
5. Select Brush Delimited and click **Multiple Delimited** on the Multiple tab.

Tip: Alternatively, you can select Brush after Waking Up, Brush After Meal, Brush Before Sleep, and Brush Another Time and click **Indicator Group** on the Multiple tab

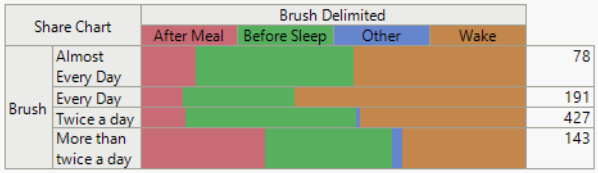
---

6. Select Brush and click **X, Grouping Category**.
7. Click **OK**.
8. Click the Categorical red triangle and select **Test Multiple Response > Count Test, Poisson**.

Figure 3.20 Test Multiple Response, Poisson

Multiple Delimited(Brush Delimited) By Brush

	Freq / Share / Rate	Brush Delimited: After Meal	Before Sleep	Other	Wake	Total Responses	Total Cases	Total Cases Responding
Brush	Almost Every Day	11	32	0	35	78	48	48
		14.1%	41.0%	0.0%	44.9%			
		22.9%	66.7%	0.0%	72.9%			
	Every Day	21	55	0	115	191	142	142
		11.0%	28.8%	0.0%	60.2%			
		14.8%	38.7%	0.0%	81.0%			
	Twice a day	49	189	4	185	427	209	209
		11.5%	44.3%	0.9%	43.3%			
		23.4%	90.4%	1.9%	88.5%			
	More than twice a day	46	47	4	46	143	49	49
		32.2%	32.9%	2.8%	32.2%			
		93.9%	95.9%	8.2%	93.9%			



Test Each Response, Poisson

Brush, Brush Delimited

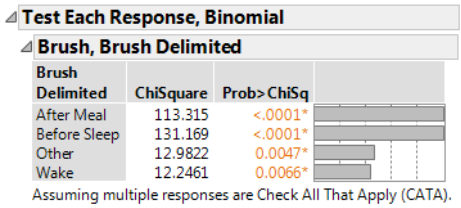
Brush Delimited	ChiSquare	Prob>ChiSq
After Meal	59.5436	<.0001*
Before Sleep	39.1097	<.0001*
Other	12.7131	0.0053*
Wake	1.8793	0.5978

Chi-squared tests use Poisson rates.

The *p*-values show that the response rates for After Meal, Before Sleep, and Other are significantly different across brushing groups.Wake is not significantly different across brushing groups. The bar graph to the right of the Prob>ChiSq column plots the *p*-values on a -Log10(p) scale. From the crosstab table, you can see that most people brush their teeth when they wake up regardless of how frequently they brush their teeth.

9. Click the Categorical red triangle and select **Test Multiple Response > Homogeneity Test, Binomial**.

Figure 3.21 Test Multiple Response, Binomial



The Homogeneity Test, Binomial option always produces a larger test statistic (and therefore a smaller *p*-value) than the Count Test, Poisson option. The binomial distribution compares not only the rate at which the response occurred (the number of people who reported that they brush upon waking) but also the rate at which the response did not occur (the number of people who did not report that they brush upon waking).

In this example, the proportion of responders for each response (After Meal, Before Sleep, Wake, and Other) differs across the groups. The *p*-value for each response is less than 0.05.

Tip: JMP detects a multiple response column by the Multiple Response modeling type or the Multiple Response column property.

Example of Supercategories

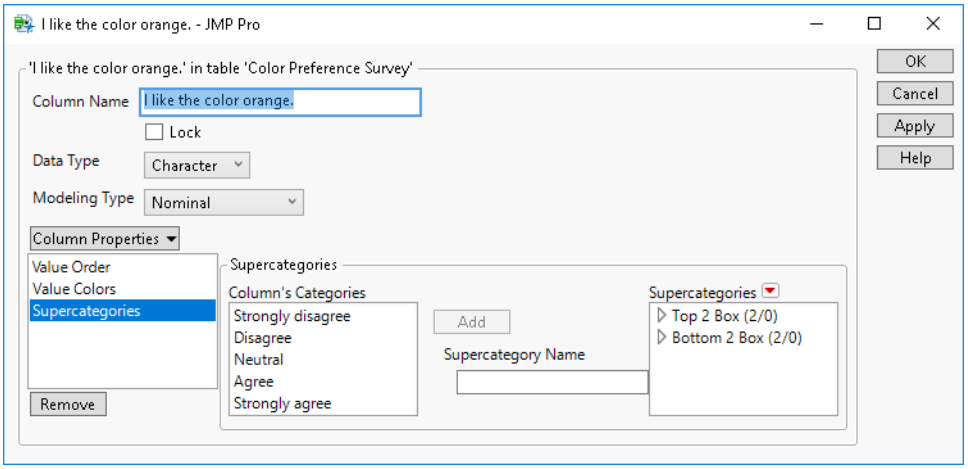
This example uses the Color Preference Survey.jmp sample data table, which contains survey data on people's color preferences. This example illustrates the use of supercategories.

1. Select **Help > Sample Data Library** and open Color Preference Survey.jmp.
2. Select the column I like the color orange. Right-click the column heading and select **Column Properties > Supercategories**.
3. Under Supercategories, select Agree and Strongly agree for the column's categories.
4. Under Supercategory Name, enter Top 2 Box.
5. Click **Add**.

Tip: Click on the triangle to the left of the name to show the categories included in the supercategory.

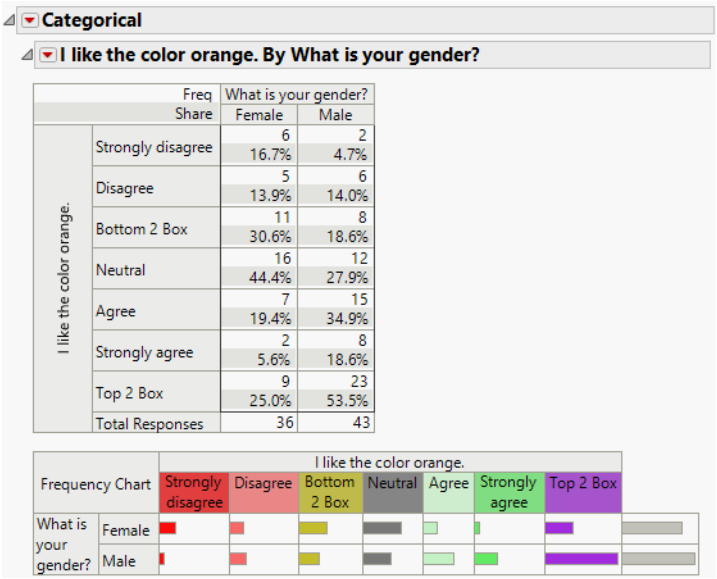
6. Under Supercategories, select Disagree and Strongly disagree for the column's categories.
7. Under Supercategory Name, enter Bottom 2 Box.
8. Click **Add**.

Figure 3.22 Supercategory Column Property



9. Click **OK**.
10. Select **Analyze > Consumer Research > Categorical**.
11. Select the **Structured** tab.
12. Click the Side green triangle and select I like the color orange.
13. Click the Top green triangle and select What is your gender?
14. Click **Add=>** and then click **OK**.
15. Click the Categorical red triangle and select **Transposed Freq Chart**.

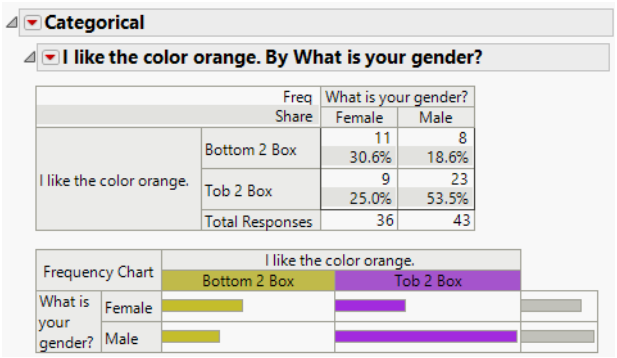
Figure 3.23 Structured Categorical Report with Supercategories



The crosstab table includes two additional rows, one for each supercategory. The supercategories are also included in the Frequency Chart. Note that the frequency counts in the Top 2 Box row are the sums of the counts in the Agree and Strongly agree categories. The frequency counts in the Bottom 2 Box row are the sums of the counts in the Disagree and Strongly disagree categories.

16. Click the Categorical red triangle and deselect **Response Levels**.

Figure 3.24 Structured Categorical Report with Supercategories and No Response Levels



By removing the response levels, your output now contains only the supercategories. Note that the totals are for all levels. In this example, the neutral responses are not included in the supercategories.

Example of the Cell Chisq Test

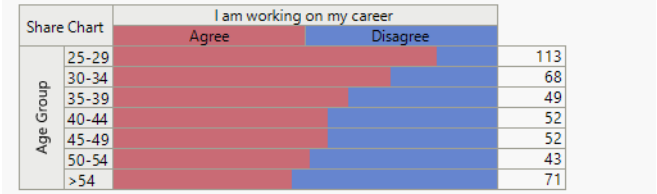
This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about people's attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of responses to the statement "I am working on my career" across age groups.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select I am working on my career and click **Responses** on the Simple tab.
4. Select Age Group and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Crosstab Transposed**.
7. Click the Categorical red triangle and select **Cell Chisq**.

Figure 3.25 Cell Chisq

I am working on my career By Age Group

Freq / Share / Cell Chisq PVal		Age Group						
		25-29	30-34	35-39	40-44	45-49	50-54	>54
I am working on my career	Agree	95	49	30	29	29	22	33
		84.1%	72.1%	61.2%	55.8%	55.8%	51.2%	46.5%
		0.00788	0.41003	0.80398	0.45495	0.45495	0.29058	0.06415
	Disagree	18	19	19	23	23	21	38
		15.9%	27.9%	38.8%	44.2%	44.2%	48.8%	53.5%
		0.00039	0.27136	0.74035	0.31848	0.31848	0.15823	0.01345
	Total Responses	113	68	49	52	52	43	71



Small *p*-values indicate that there is a significant difference between the observed cell count and the expected cell count. The *p*-values are colored by significance level from dark red for cells with significantly higher counts than expected to dark blue for cells with significantly lower counts than expected. The expected cell count is based on the observed row and column totals. The expected cell count is calculated as the row total times the column total divided by the overall count.

For example, the expected number of responses under the null hypothesis that the rates are equal in the 25 - 29 group who agree is (287*113)/448 = 72.4; the observed value was 95. This observed value, with a *p*-value of 0.00788, is significantly larger than the expected value. The number of responses in the 25 - 29 group who agree with "I am working on my career" is higher than expected if the response to this question was independent of age.

Example of Compare Each Sample with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about people's attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of responses to the statement "I am working on my career" between each age group.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select I am working on my career and click **Responses** on the Simple tab.
4. Select Age Group and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Compare Each Sample**.

Figure 3.26 Compare Each Sample

I am working on my career By Age Group

			Freq / Share	I am working on my career: Agree	Disagree	Total Responses	Compare
Age Group	25-29	A		95 84.1%	18 15.9%	113	b,C,D,E,F,G
	30-34	B		49 72.1%	19 27.9%	68	d,e,F*
	35-39	C		30 61.2%	19 38.8%	49	*
	40-44	D		29 55.8%	23 44.2%	52	*
	45-49	E		29 55.8%	23 44.2%	52	*
	50-54	F		22 51.2%	21 48.8%	43	*
	>54	G		33 46.5%	38 53.5%	71	B*

Default Comparison Groups: A/B/C/D/E/F/G

Shows letter of the category it is significantly different from at the higher share level
* Base count warning 100 Uppercase Alpha Level 0.05
** Base count minimum 30 Lowercase Alpha Level 0.1

Compare Each Sample

Letter comparisons use Pearson Chisq
Age Group, I am working on my career

LR PValues
LR Chi-square p-value on pairs

	A	B	C	D	E	F	G
A	1.0000	0.0552	0.0020	0.0001	0.0001	<.0001	<.0001
B	0.0552	1.0000	0.2183	0.0641	0.0641	0.0261	0.0020
C	0.0020	0.2183	1.0000	0.5781	0.5781	0.3312	0.1108
D	0.0001	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083
E	0.0001	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083
F	<.0001	0.0261	0.3312	0.6540	0.6540	1.0000	0.6276
G	<.0001	0.0020	0.1108	0.3083	0.3083	0.6276	1.0000

Pearson PValues
Pearson Chi-square p-value on pairs

	A	B	C	D	E	F	G
A	1.0000	0.0523	0.0015	<.0001	<.0001	<.0001	<.0001
B	0.0523	1.0000	0.2170	0.0638	0.0638	0.0255	0.0022
C	0.0015	0.2170	1.0000	0.5783	0.5783	0.3314	0.1119
D	<.0001	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
E	<.0001	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
F	<.0001	0.0255	0.3314	0.6540	0.6540	1.0000	0.6276
G	<.0001	0.0022	0.1119	0.3087	0.3087	0.6276	1.0000

The crosstab table summarizes the statement "I am working on my career" across age groups. The cells of the table contain the frequency (count) and share (percent) of those who agree or disagree with the statement for each age group. In addition, the crosstab includes comparison letters. Each group is labeled with a letter in the Compare column, which is to the right of the group label. The letters in the Compare column enable you to interpret the outcomes of the statistical test of independence among groups.

The Compare Each Sample report provides *p*-values from the pairwise Pearson and Chi-square likelihood ratio chi-square tests. The *p*-values are reported in symmetric matrices labeled by the comparison letters.

For this example we make the following observations:

- The comparison column for the 25 - 29 group contains all letters b through g. Thus, the 25 - 29 group has significantly different response rates to the statement "I am working on my career" as compared to all other groups. Because the letter b is lowercase, the difference between the 25 - 29 group and the 30 - 34 group is significant at the 0.10 level. All other letters are uppercase, which indicate differences that are significant at the 0.05 level.
- The >54 group, denoted by letter G, is significantly different from the 30 - 34 group, denoted by B. The letter for the comparison is in the cell for group G because group G has a higher number of responders (71 versus 68) than group B.
- The single asterisks in the comparison cells are small sample size warnings. A single asterisk indicates that a group has more than 30 but fewer than 100 responses.
- A double asterisk, not observed in this example, would indicate a group size of fewer than 30.

Example of Compare Each Cell with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about people's attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of the responses about job satisfaction between employee tenure groups.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select Job Satisfaction and click **Responses** on the Simple tab.
4. Select Employee Tenure and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Compare Each Cell**.
7. Click the **Compare Each Cell - Details** gray disclosure icon.

Figure 3.27 Compare Each Cell

Job Satisfaction By Employee Tenure

			Job Satisfaction			
Freq / Share / Comparisons			Not at all satisfied	Somewhat satisfied	Extremely satisfied	Total Responses
Employee Tenure	less than 5 years	A	11 6.9% 	97 60.6% d	52 32.5% 	160
	5 to 10 years	B	13 9.4% 	72 51.8% 	54 38.8% 	139
	10 to 20 years	C	4 4.7% *	43 50.0% *	39 45.3% a*	86
	more than 20 years	D	4 6.3% *	30 47.6% *	29 46.0% a*	63

Default Comparison Groups: A/B/C/D

Shows letter of the category it is significantly different from at the higher share level
* Base count warning 100 Uppercase Alpha Level 0.05
** Base count minimum 30 Lowercase Alpha Level 0.1

Share Chart		Job Satisfaction			
		Not at all satisfied	Somewhat satisfied	Extremely satisfied	
Employee Tenure	less than 5 years				160
	5 to 10 years				139
	10 to 20 years				86
	more than 20 years				63

Compare Each Cell - Details

Letter comparisons use Fisher's Exact Test
Employee Tenure, Job Satisfaction

LR Pairs

		AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.4323	1.0000	0.4781	0.1806	1.0000	0.8872	0.4658	0.6515	1.0000
Somewhat satisfied	1.0000	0.1246	1.0000	0.1092	0.7931	1.0000	0.0783	0.5820	0.7739	1.0000
Extremely satisfied	1.0000	0.2526	1.0000	0.0476	0.3366	1.0000	0.0607	0.3377	0.9341	1.0000

Pearson Pairs

		AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.4316	1.0000	0.4870	0.1948	1.0000	0.8878	0.4763	0.6496	1.0000
Somewhat satisfied	1.0000	0.1246	1.0000	0.1086	0.7931	1.0000	0.0774	0.5820	0.7740	1.0000
Extremely satisfied	1.0000	0.2524	1.0000	0.0465	0.3360	1.0000	0.0585	0.3364	0.9341	1.0000

Fisher Exact Pairs

		AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.5235	1.0000	0.5853	0.2990	1.0000	1.0000	0.5912	0.7221	1.0000
Somewhat satisfied	1.0000	0.1302	1.0000	0.1373	0.8909	1.0000	0.0983	0.6493	0.8685	1.0000
Extremely satisfied	1.0000	0.2764	1.0000	0.0531	0.4034	1.0000	0.0649	0.3574	1.0000	1.0000

The *p*-values for pairwise likelihood ratio chi-square, Pearson chi-square, and Fisher's exact tests for independence are provided in tables. The tables are labeled by comparison letters. The comparison letters are shown in the crosstab table to the right of the group labels. Response rates that differ between groups are indicated with a comparison letter in the crosstab table cells.

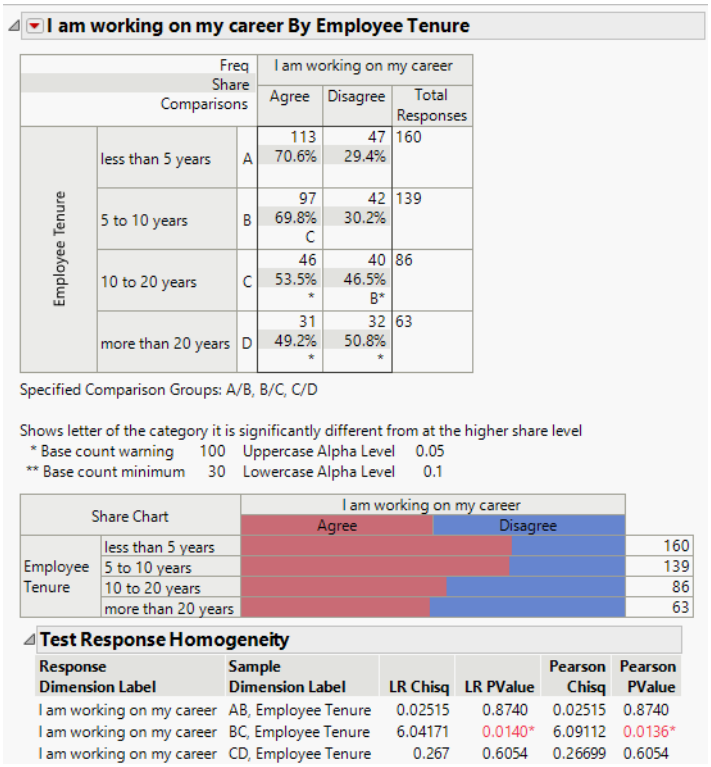
Employees with fewer than 5 years of tenure are somewhat satisfied at a greater rate than those with 20 years of tenure. This finding is noted by the letter d in the Somewhat satisfied cell in the first row of the Crosstab table. In addition, the employees with fewer than 5 years of tenure are Extremely satisfied at a lower rate than the group with 20 years of tenure. This finding is noted by the letter a in the Extremely satisfied cell of the last row of the crosstab table. The letters are placed in the cell with the highest share of responses.

Example of User-Specified Comparison with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about people's attitudes and opinions. You define specific comparison groups across which to compare the responses to the statement "I am working on my career".

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select I am working on my career and click **Responses** on the Simple tab.
4. Select Employee Tenure and click **X, Grouping Category**.
5. Click **OK**.
6. Click the red triangle next to I am working on my career By Employee Tenure and select **Show Letters**.
7. Click the red triangle next to I am working on my career By Employee Tenure and select **Specify Comparison Groups**.
8. Enter A/B, B/C, C/D and click **OK**.
9. Click the Categorical red triangle and select **Test Response Homogeneity**.

Figure 3.28 Specify Comparison Example



I am working on my career By Employee Tenure

			Freq / Share / Comparisons	I am working on my career: Agree	I am working on my career: Disagree	Total Responses
Employee Tenure	less than 5 years	A		113 70.6%	47 29.4%	160
	5 to 10 years	B		97 69.8% C	42 30.2%	139
	10 to 20 years	C		46 53.5% *	40 46.5% B*	86
	more than 20 years	D		31 49.2% *	32 50.8% *	63

Specified Comparison Groups: A/B, B/C, C/D

Shows letter of the category it is significantly different from at the higher share level
* Base count warning 100 Uppercase Alpha Level 0.05
** Base count minimum 30 Lowercase Alpha Level 0.1

Share Chart		I am working on my career: Agree / Disagree	
Employee Tenure	less than 5 years		160
	5 to 10 years		139
	10 to 20 years		86
	more than 20 years		63

Test Response Homogeneity

Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
I am working on my career	AB, Employee Tenure	0.02515	0.8740	0.02515	0.8740
I am working on my career	BC, Employee Tenure	6.04171	0.0140*	6.09112	0.0136*
I am working on my career	CD, Employee Tenure	0.267	0.6054	0.26699	0.6054

The test of response homogeneity compares Group A to Group B, Group B to Group C, and Group C to Group D. Group B (5 to 10 years) respondents agree with the statement "I am working on my career" more often than those in Group C (10 to 20 years). The Pearson *p*-value for this difference in agreement rates is 0.0136.

Example of Aligned Responses

This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about attitudes and opinions.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Scroll to see the column I am working on my career.

Figure 3.29 Consumer Preferences Data Table (Partial Table)

	I am working on my career	I want to see the world	My home needs some major ...	I have vast interests outside of work	I want to get my debt under control	I come from a large family
1	Agree	Disagree	Disagree	Agree	Disagree	Agree
2	Agree	Agree	Agree	Agree	Disagree	Agree
3	Disagree	Agree	Agree	Agree	Agree	Agree
4	Agree	Agree	Agree	Agree	Agree	Agree
5	Disagree	Agree	Disagree	Agree	Agree	Disagree

Note that there are six columns, all with the same responses: Agree and Disagree. The responses for these six columns are aligned. For this example, we analyze the columns I am working on my career and I want to see the world. First, use standardize attributes to set the value colors, value order, and modeling type for these two columns:

3. Select the column headings I am working on my career and I want to see the world. Right-click and select **Standardize Attributes**.
4. Select **Column Properties > Value Colors.** Right-click the Agree color oval and set it to green and set Disagree to red.
5. Select **Column Properties > Value Order.** Click **Reverse**.
6. Select **Attributes > Modeling Type** and **Modeling Type > Ordinal**.
7. Click **OK**.

   Next, use the Categorical platform to analyze these two columns.
8. Select **Analyze > Consumer Research > Categorical**.
9. Select I am working on my career and I want to see the world.
10. Select the **Related** tab, and then click **Aligned Responses** on the Related tab.

> **Tip:** To do the same thing using the Structured tab, drag the two columns into the Multiple Aligned drop zone (green arrow).

11. Click **OK**.

Figure 3.30 Aligned Response Report

Aligned Responses(I am working on my career, I want to see the world)

		Freq / Share	Response: Disagree	Response: Agree	Total Responses
Response	I am working on my career		161	287	448
			35.9%	64.1%	
Response	I want to see the world		78	370	448
			17.4%	82.6%	

		Share Chart: Response Disagree	Response Agree	
Response	I am working on my career			448
Response	I want to see the world			448

Notice that the Share Chart is a directional bar chart. The type of bar chart you see depends on the modeling type of the columns:

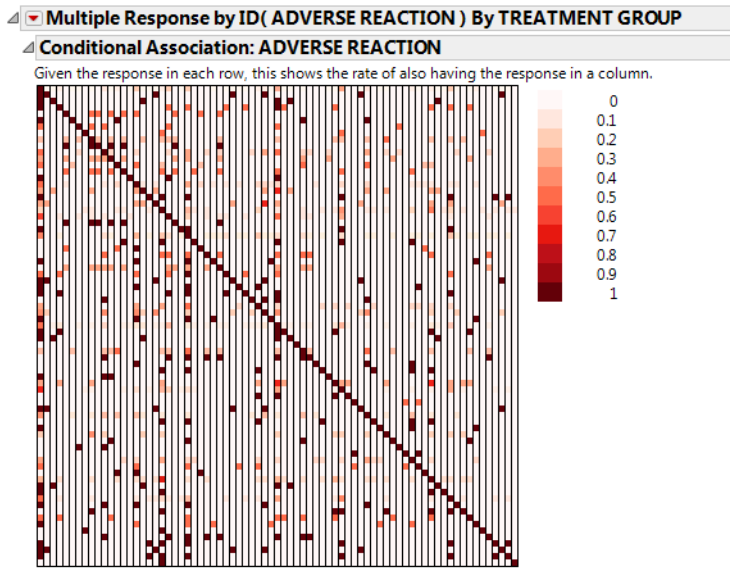
- Ordinal columns result in a directional bar chart. The columns in this example are both ordinal.
- Nominal columns result in a stacked bar chart.

Example of Conditional Association and Relative Risk

This example uses the AdverseR.jmp sample data table, which contains adverse reactions from a clinical trial. Use this data to explore the conditional association of adverse events and then the relative risk of the events in the treatment group as compared to the control.

1. Select **Help > Sample Data Library** and open AdverseR.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select ADVERSE REACTION and click **Multiple Response by ID** on the Multiple tab.
5. Select TREATMENT GROUP and click **X, Grouping Category**.
6. Select PATIENT ID and click **ID**.
7. Under the other launch window options, select **Unique Occurrences within ID** and click **OK**.
8. Click the Categorical red triangle and select **Conditional Association**.

Figure 3.31 Conditional Association Report (Partial Report)



Conditional Assoc

	ABDOMINAL PAIN	ABNORMAL VISION	ALOPECIA	AMBLYOPIA	ANEMIA
ABDOMINAL PAIN	1.0000	0.0323	0.0323	0.0645	0.0000
ABNORMAL VISION	1.0000	1.0000	0.0000	0.0000	0.0000
ALOPECIA	1.0000	0.0000	1.0000	0.0000	0.0000
AMBLYOPIA	1.0000	0.0000	0.0000	1.0000	0.0000
ANEMIA	0.0000	0.0000	0.0000	0.0000	1.0000
ANGINA PECTORIS	1.0000	0.0000	0.0000	0.0000	0.0000
ANOREXIA	0.5000	0.0000	0.0000	0.0000	0.0000
ANXIETY	0.0000	0.0000	0.0000	0.0000	0.0000
ASTHENIA	0.2500	0.0000	0.0000	0.0000	0.2500
ASTHMA	0.0000	0.0000	0.0000	0.0000	1.0000
BACK PAIN	0.5000	0.0000	0.0000	0.0000	0.0000
BRONCHITIS	0.3333	0.0000	0.0000	0.0000	0.3333
CHEST PAIN	0.5000	0.0000	0.0000	0.0000	0.0000

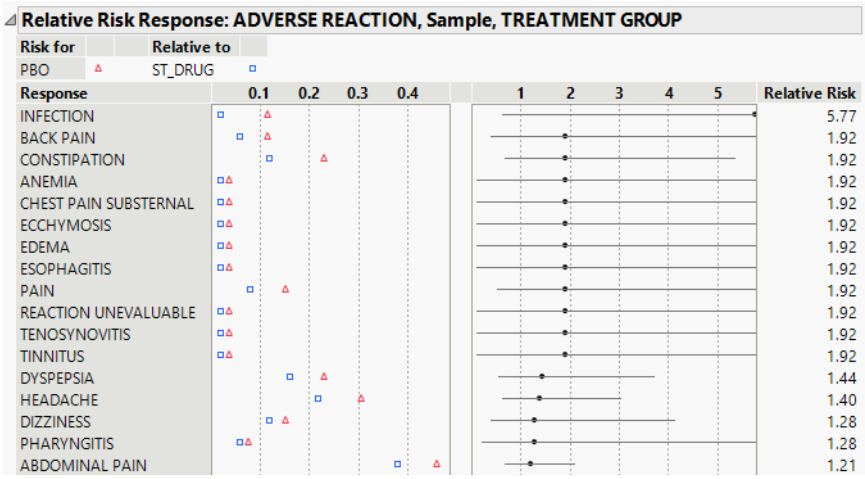
The conditional association matrix provides the conditional probability of one adverse reaction given the presence of another reaction. The probabilities are across all groups. The probability of abnormal vision given that a patient has abdominal pain is 0.0323.

Tip: Hover over the heat map for conditional probabilities.

9. Click the Categorical red triangle and select **Relative Risk.**
10. Select PBO in the window and click **OK**.

    Right-click the Relative Risk Report in the window and select **Sort by Column**.
11. Select Relative Risk and click **OK**.

Figure 3.32 Relative Risk Report (Partial Report)



The Relative Risk option computes relative risks of different responses as the ratio of the risk for each level of the grouping variable. The default Relative Risk report lists the response name, the risk (rate) for each level of the grouping variable, a plot of the relative risk with 95% confidence intervals, and the relative risk estimate. Here you can compare the relative risk of the adverse reactions by treatment group. The relative risk of an infection is 5.7 times greater for PBO relative to ST_DRUG. However, the confidence interval is very wide and includes a relative risk of 1.0. A relative risk of 1.0 occurs when the risk is equal for each level of the grouping variable.

Right-click and select **Columns > Lower 95%** and **Columns > Upper 95%** to add 95% confidence intervals on the relative risk estimates to the report table.

Example of Rater Agreement

This example uses the Attribute Gauge.jmp sample data, which has the ratings (0/1) from three operators rating 50 parts three times.

1. Select **Help > Sample Data Library** and open Attribute Gauge.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Related tab.
4. Select A, B, and C, and click **Rater Agreement** on the Related tab.
5. Click **OK**.

Figure 3.33 Rater Agreement Report

Rater Agreement(A, B, C)

Freq / Share		Response 0	Response 1	Total Responses
Response	A	50 33.3%	100 66.7%	150
	B	47 31.3%	103 68.7%	150
	C	51 34.0%	99 66.0%	150

Share Chart		Response 0	Response 1	
Response	A			150
	B			150
	C			150

Rater Agreement Statistics

Degree of Agreement

Degree of Agreement

Response1	Response2	Kappa	Std Err	Bowker Symmetry	Bowker P-Value
A	B	0.862944	0.044198	1	0.3173
A	C	0.776119	0.054671	0.066667	0.7963
B	C	0.788007	0.053702	1.142857	0.2850

Two by two tables are the same as Bowker's Test

Details

Rater Row	Rater Col	Level	0	1
B	A	0	44	3
B	A	1	6	97
C	A	0	43	8
C	A	1	7	92
C	B	0	42	9
C	B	1	5	94

The rater agreement is strong as shown by the Kappa statistics. The Kappa statistic can take on a value between 0 (no agreement) to 1.0 (perfect agreement). The details section provides 2x2 tables for each pair of raters. The Bowker test of symmetry tests the null hypothesis that cell proportions are symmetric for all pairs of cells ($p_{ij} = p_{ji}$ for all i, j). Here, p-values for the Bowker test are all greater than 0.05, indicating no strong evidence of asymmetry between raters.

Example of Repeated Measures

This example uses the Presidential Elections.jmp sample data table, which contains United States presidential election results for each state from 1980 through 2012. Use this data table to explore repeated measures where we consider the election results as repeated measures.

1. Select **Help > Sample Data Library** and open Presidential Elections.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Related tab.
4. Select 1980 Winner through 2012 Winner, and click **Repeated Measures** on the Related tab.
5. Click **OK**.

6. Near the bottom of the report window, click the gray Transition Report disclosure icon to open the Transition Report.

Figure 3.34 Repeated Measures Transition Report

Transition Report

All From	to	Transition Counts			Transition Rates			
All 1980 Winner	1984 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1984 Winner**
		Democrat	1	5	Democrat	0.1667	0.8333	
		Republican	0	44	Republican	0.0000	1.0000	
All 1984 Winner	1988 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1988 Winner**
		Democrat	1	0	Democrat	1.0000	0.0000	
		Republican	9	40	Republican	0.1837	0.8163	
All 1988 Winner	1992 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1992 Winner**
		Democrat	10	0	Democrat	1.0000	0.0000	
		Republican	22	18	Republican	0.5500	0.4500	
All 1992 Winner	1996 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1996 Winner**
		Democrat	29	3	Democrat	0.9063	0.0938	
		Republican	2	16	Republican	0.1111	0.8889	
All 1996 Winner	2000 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2000 Winner**
		Democrat	20	11	Democrat	0.6452	0.3548	
		Republican	0	19	Republican	0.0000	1.0000	
All 2000 Winner	2004 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2004 Winner**
		Democrat	18	2	Democrat	0.9000	0.1000	
		Republican	1	29	Republican	0.0333	0.9667	
All 2004 Winner	2008 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2008 Winner**
		Democrat	19	0	Democrat	1.0000	0.0000	
		Republican	9	22	Republican	0.2903	0.7097	
All 2008 Winner	2012 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2012 Winner**
		Democrat	26	2	Democrat	0.9286	0.0714	
		Republican	0	22	Republican	0.0000	1.0000	

The Transition Report is unique to the repeated measures analysis. This report includes counts and rates of differences between subsequent time points. Between 1980 and 1984, there were 5 Democratic states that transitioned to Republican states at a rate of 0.8333 or 5 out of 6 states. In 1980, they voted Democratic but voted Republican in 1984. Between 2008 and 2012, there were 2 out of 28 Democratic states that transitioned to Republican at a rate of 0.0714. All other states voted the same way in both the 2008 and 2012 elections.

Example of the Multiple Responses

The examples in this section use sample data tables that contain the same information organized in five different data table layouts. The data come from testing a fabrication line on three different occasions under two different conditions. Each set of operating conditions (or batch) yielded 50 units for inspection. Inspectors recorded seven types of defects. Each unit could have zero, one, or more than one defect. A unit could have more than one defect of the same kind.

Multiple Response

The Failure3MultipleField.jmp sample data table has a row for each unit and multiple columns for defects, where defects are entered one per column. In this example, there are three columns for defects. Thus, any one unit had at most three defects.

1. Select **Help > Sample Data Library** and open Quality Control/Failure3MultipleField.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select Failure1, Failure2, and Failure3, and click **Multiple Response** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Click **OK**.

Figure 3.35 Multiple Response Report

Multiple Response(Failure1, Failure2, Failure3) By clean*date

			Freq / Share / Rate	Failure									
				contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect	Total Responses	Total Cases	Total Cases Responding
clean	after	date	OCT 1	12	2	0	4	2	1	2	23	50	23
				52.2%	8.7%	0.0%	17.4%	8.7%	4.3%	8.7%			
				24.0%	4.0%	0.0%	8.0%	4.0%	2.0%	4.0%			
			OCT 2	10	1	1	5	1	2	3	23	50	23
				43.5%	4.3%	4.3%	21.7%	4.3%	8.7%	13.0%			
				20.0%	2.0%	2.0%	10.0%	2.0%	4.0%	6.0%			
			OCT 3	8	3	0	5	0	1	0	17	50	17
				47.1%	17.6%	0.0%	29.4%	0.0%	5.9%	0.0%			
				16.0%	6.0%	0.0%	10.0%	0.0%	2.0%	0.0%			
	before	date	OCT 1	14	2	1	2	3	8	1	31	50	28
				45.2%	6.5%	3.2%	6.5%	9.7%	25.8%	3.2%			
				28.0%	4.0%	2.0%	4.0%	6.0%	16.0%	2.0%			
			OCT 2	15	2	2	1	4	6	0	30	50	30
				50.0%	6.7%	6.7%	3.3%	13.3%	20.0%	0.0%			
				30.0%	4.0%	4.0%	2.0%	8.0%	12.0%	0.0%			
			OCT 3	22	2	3	4	0	3	2	36	50	36
				61.1%	5.6%	8.3%	11.1%	0.0%	8.3%	5.6%			
				44.0%	4.0%	6.0%	8.0%	0.0%	6.0%	4.0%			

Share Chart				Failure							
				contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect	
clean	after	date	OCT 1								23
			OCT 2								23
			OCT 3								17
	before	date	OCT 1								31
			OCT 2								30
			OCT 3								36

The crosstab table has a row for each batch and a column for each defect type. The frequency, share, and rate of each defect within each batch are shown in the table cells. For example, for the batch after cleaning on OCT 1, there were 12 contamination defects representing 12/23 or 52.2% of the defects for that batch. The 12 contamination defects were from 50 units. For each clean and date combination, there were 50 total units. Each unit could have one or more defects. Therefore, the rate of contamination per unit was 24%. For the batch before cleaning on OCT 1, the Total Cases Responding is 28. The Total Responses count is 31, because three of the cases reported two defects.

Multiple Response by ID

The Failure3ID.jmp sample data table has a row for each defect type within each batch, a column for the number of occurrences of each defect type, and an ID column for each batch.

Figure 3.36 Failure3ID Data Table (Partial Table)

	failure	N	clean	date	SampleSize	ID
1	contamination	14	before	OCT 1	50	OCT 1 before
2	corrosion	2	before	OCT 1	50	OCT 1 before
3	doping	1	before	OCT 1	50	OCT 1 before
4	metallization	2	before	OCT 1	50	OCT 1 before
5	miscellaneous	3	before	OCT 1	50	OCT 1 before
6	oxide defect	8	before	OCT 1	50	OCT 1 before
7	silicon defect	1	before	OCT 1	50	OCT 1 before
8	doping	0	after	OCT 1	50	OCT 1 after
9	corrosion	2	after	OCT 1	50	OCT 1 after
10	metallization	4	after	OCT 1	50	OCT 1 after

1. Select **Help > Sample Data Library** and open Quality Control/Failure3ID.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select failure and click **Multiple Response by ID** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Select SampleSize and click **Sample Size**.
7. Select N and click **Freq**.
8. Select ID and click **ID**.
9. Click **OK**.

The resulting report is the same as the report shown in Figure 3.35 with the exception of the Total Cases Responding column in the crosstab table. Here, the defect counts were summarized. From the summarized table, there is no record of the number of units with zero defects. Thus, the Total Cases Responding is the full batch size of 50 for each batch.

Multiple Delimited

The Failures3Delimited.jmp sample data table has a row for each unit with a single column in which the defects are recorded, delimited by a comma. Note in the partial data table, shown in Figure 3.37, that some units did not have any observed defects, so the failures column is empty.

Figure 3.37 Failure3Delimited.jmp Data Table (Partial Table)

	failures	clean	date	ID	ID Label
1		before	OCT 1	1	OCT 1 before
2	oxide defect	before	OCT 1	1	OCT 1 before
3	contamination,oxide defect	before	OCT 1	1	OCT 1 before
4		before	OCT 1	1	OCT 1 before
5	contamination	before	OCT 1	1	OCT 1 before
6	oxide defect	before	OCT 1	1	OCT 1 before
7	contamination	before	OCT 1	1	OCT 1 before
8		before	OCT 1	1	OCT 1 before
9		before	OCT 1	1	OCT 1 before
10	metallization,contamination	before	OCT 1	1	OCT 1 before

1. Select **Help > Sample Data Library** and open Quality Control/ Failures3Delimited.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select failures and click **Multiple Delimited** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Click **OK**.

When you click **OK**, you also get the report shown in Figure 3.35.

Note: If you specify more than one delimited column, separate analyses are produced for each column.

Indicator Group

The Failures3Indicators.jmp sample data table has a row for each unit and an indicator column for each defect type. The data entry in each defect column is a 0 if that defect was not observed and a 1 if the defect was observed for the unit.

Figure 3.38 Faliure3Indicators.jmp Data Table (Partial Table)

	clean	date	ID	ID Label	contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect
1	before	OCT 1	1	OCT 1 before	0	0	0	0	0	0	0
2	before	OCT 1	1	OCT 1 before	0	0	0	0	0	1	0
3	before	OCT 1	1	OCT 1 before	1	0	0	0	0	1	0
4	before	OCT 1	1	OCT 1 before	0	0	0	0	0	0	0
5	before	OCT 1	1	OCT 1 before	1	0	0	0	0	0	0

1. Select **Help > Sample Data Library** and open Quality Control/Failures3Indicators.jmp.

2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select contamination through silicon defect and click **Indicator Group** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Click **OK**.

When you click **OK**, you get the report shown in Figure 3.35.

Response Frequencies

The Failure3Freq.jmp sample data table has a row for each batch, a column for each defect type, and a column for the batch size. The data entries in the defect columns are the number of occurrences of each defect in the batch.

Figure 3.39 Failure3Freq.jmp Data Table

	clean	date	contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect	SampleSize
1	after	OCT 1	12	2	0	4	2	1	2	50
2	after	OCT 2	10	1	1	5	1	2	3	50
3	after	OCT 3	8	3	0	5	0	1	0	50
4	before	OCT 1	14	2	1	2	3	8	1	50
5	before	OCT 2	15	2	2	1	4	6	0	50
6	before	OCT 3	22	2	3	4	0	3	2	50

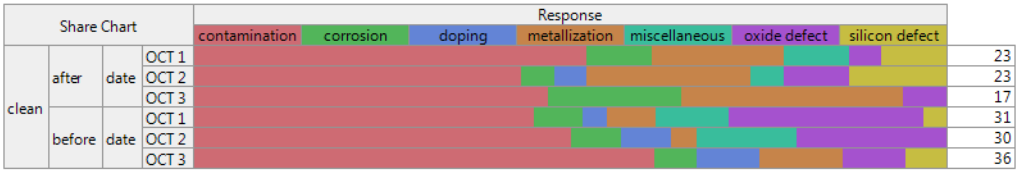
1. Select **Help > Sample Data Library** and open Quality Control/Failure3Freq.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select the frequency variables (contamination through silicon defect).
5. On the Multiple tab, click **Response Frequencies**.
6. Select clean and date and click **X, Grouping Category**.
7. Select Sample Size and click **Sample Size**.
8. Click **OK**.

Figure 3.40 Defect Rate Output

Sample Size: SampleSize

Response Frequencies(contamination, corrosion, doping, metallization, miscellaneous, oxide defect, silicon defect) By clean*date

				Freq / Share / Rate	contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect	Total Responses	Total Cases	Total Cases Responding
clean	after	date	OCT 1	12	2	0	4	2	1	2	23	50	50
				52.2%	8.7%	0.0%	17.4%	8.7%	4.3%	8.7%			
				24.0%	4.0%	0.0%	8.0%	4.0%	2.0%	4.0%			
			OCT 2	10	1	1	5	1	2	3	23	50	50
				43.5%	4.3%	4.3%	21.7%	4.3%	8.7%	13.0%			
				20.0%	2.0%	2.0%	10.0%	2.0%	4.0%	6.0%			
			OCT 3	8	3	0	5	0	1	0	17	50	50
				47.1%	17.6%	0.0%	29.4%	0.0%	5.9%	0.0%			
				16.0%	6.0%	0.0%	10.0%	0.0%	2.0%	0.0%			
	before	date	OCT 1	14	2	1	2	3	8	1	31	50	50
				45.2%	6.5%	3.2%	6.5%	9.7%	25.8%	3.2%			
				28.0%	4.0%	2.0%	4.0%	6.0%	16.0%	2.0%			
			OCT 2	15	2	2	1	4	6	0	30	50	50
				50.0%	6.7%	6.7%	3.3%	13.3%	20.0%	0.0%			
				30.0%	4.0%	4.0%	2.0%	8.0%	12.0%	0.0%			
			OCT 3	22	2	3	4	0	3	2	36	50	50
				61.1%	5.6%	8.3%	11.1%	0.0%	8.3%	5.6%			
				44.0%	4.0%	6.0%	8.0%	0.0%	6.0%	4.0%			



The resulting output is the same as that in Figure 3.35 with the exception of the Total Cases Responding column in the crosstab table. Here, the defect counts were summarized. From the summarized table, there is no record of the number of units with zero defects. Thus, the Total Cases Responding is the full batch size of 50 for each batch.

Example of Mean Score with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table to explore the relationship between employee tenure and having school age children. The Employee Tenure column is a numeric column with values 1, 2, 3, and 4. These values have been assigned Value Labels using the Value Labels column property. To evaluate an average employee tenure using the Mean Score option in the categorical platform, assign Value Scores to the column values. For more information about column properties, see *Using JMP*.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. In the data table, right-click the Employee Tenure column heading and select **Column Properties > Value Scores**.
3. Enter 1 for **Value** and 3 for **Score** and click **Add**.
4. Enter 2 for **Value** and 7.5 for **Score** and click **Add**.

5. Enter 3 for **Value** and 15 for **Score** and click **Add**.
6. Enter 4 for **Value** and 25 for **Score** and click **Add**.
7. Click **OK**.
8. Select **Analyze > Consumer Research > Categorical**.
9. Select Employee Tenure and click **Responses** on the Simple tab.
10. Select School Age Children and click **X, Grouping Category**.
11. Click **OK**.
12. Click the Categorical red triangle and select **Mean Score**.
13. Click the Categorical red triangle and select **Mean Score Comparisons**.

Figure 3.41 Categorical Report with Mean Scores

Employee Tenure By School Age Children

			Freq / Share	Employee Tenure: less than 5 years	5 to 10 years	10 to 20 years	more than 20 years	Total Responses	Mean	Compare Means
School Age Children	Yes	A		53	65	44	23	185	10.170	
				28.6%	35.1%	23.8%	12.4%			
	No	B		107	74	42	40	263	9.529	
				40.7%	28.1%	16.0%	15.2%			

Default Comparison Groups: A/B

Shows letter of the category it is significantly different from at the higher share level
* Base count warning 100 Uppercase Alpha Level 0.05
** Base count minimum 30 Lowercase Alpha Level 0.1

Share Chart		Employee Tenure: less than 5 years	5 to 10 years	10 to 20 years	more than 20 years	
School Age Children	Yes					185
	No					263

The mean employee tenure for those with school age children is 10.17 and 9.53 for those without school age children. Because the means are not statistically different, the Compare Means column in the Crosstab table is empty. If there were a difference, a letter would indicate the difference. If you had not used value scores, then the mean for those with school age children would be 2.20 and 2.057 for those without school age children.

Tip: Be aware of how your data are recorded when using the mean score option. If your data are recorded as coded numeric data with value labels, the mean value calculations are based on the numeric data. If the numeric values do not have meaning, use the Value Score column property to assign meaningful values to the response levels.

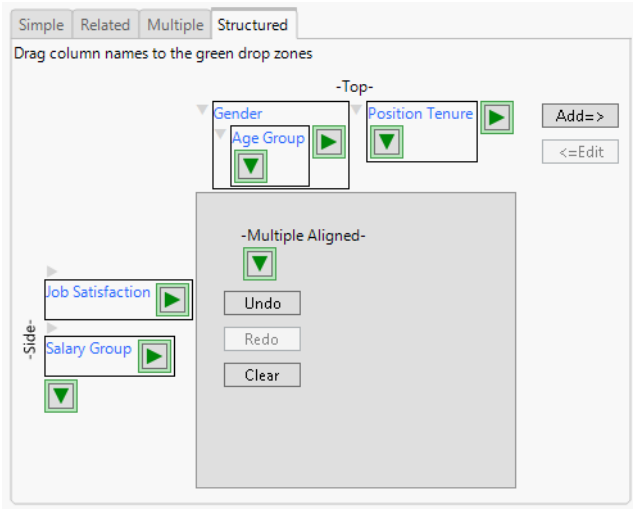
Example of a Structured Report

This example uses the Consumer Preferences.jmp sample data table to compare job satisfaction and salary against gender by age group and position tenure. Use the Structured tab to create the report.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Structured tab.
4. Drag Gender to the green drop zone at the **Top** of the table on the Structured tab.
5. Drag Age Group to the green drop zone just below Gender.
6. Drag Position Tenure to the green drop zone at the **Top** of the table next to Gender.
7. Drag Job Satisfaction to the green drop zone at the **Side** of the table.
8. Drag Salary Group to the green drop zone at the **Side** of the table under Job Satisfaction.

Tip: Click on the green drop zone arrows to select columns.

Figure 3.42 Structured Tab Report Setup



9. Click **Add=>**.
10. Click **OK**.
11. Click the Categorical red triangle and select **Test Response Homogeneity**.

Figure 3.43 Structured Tab Report Example

Job Satisfaction+Salary Group By Gender*Age Group+Position Tenure

		Freq	Gender																
		Share	M							F									
			Age Group							Age Group							Position Tenure			
			25-29	30-34	35-39	40-44	45-49	50-54	>54	25-29	30-34	35-39	40-44	45-49	50-54	>54	less than 5 years	5 to 10 years	10 to 20 years	more than 20 years
Job Satisfaction	Not at all satisfied		3	3	1	3	3	2	2	3	1	0	5	2	1	3	15	11	6	0
			5.5%	10.3%	3.3%	8.8%	7.9%	8.3%	4.1%	5.2%	2.6%	0.0%	27.8%	14.3%	5.3%	13.6%	7.2%	8.2%	6.8%	0.0%
	Somewhat satisfied		32	14	15	19	19	10	26	34	30	12	6	5	11	9	122	70	40	10
			58.2%	48.3%	50.0%	55.9%	50.0%	41.7%	53.1%	58.6%	76.9%	63.2%	33.3%	35.7%	57.9%	40.9%	58.7%	52.2%	45.5%	55.6%
	Extremely satisfied		20	12	14	12	16	12	21	21	8	7	7	7	7	10	71	53	42	8
			36.4%	41.4%	46.7%	35.3%	42.1%	50.0%	42.9%	36.2%	20.5%	36.8%	38.9%	50.0%	36.8%	45.5%	34.1%	39.6%	47.7%	44.4%
	Total Responses		55	29	30	34	38	24	49	58	39	19	18	14	19	22	208	134	88	18
Salary Group	less than 40000		20	6	4	8	6	5	10	35	19	6	6	2	7	4	80	40	15	3
			36.4%	20.7%	13.3%	23.5%	15.8%	20.8%	20.4%	60.3%	48.7%	31.6%	33.3%	14.3%	36.8%	18.2%	38.5%	29.9%	17.0%	16.7%
	40000 to 60000		16	14	11	12	14	7	13	15	10	8	6	7	6	12	67	38	37	9
			29.1%	48.3%	36.7%	35.3%	36.8%	29.2%	26.5%	25.9%	25.6%	42.1%	33.3%	50.0%	31.6%	54.5%	32.2%	28.4%	42.0%	50.0%
	60000 to 80000		8	5	7	7	8	8	10	5	5	3	3	3	4	5	32	25	19	5
			14.5%	17.2%	23.3%	20.6%	21.1%	33.3%	20.4%	8.6%	12.8%	15.8%	16.7%	21.4%	21.1%	22.7%	15.4%	18.7%	21.6%	27.8%
	80000 to 120000		9	3	6	4	7	3	7	1	2	1	2	1	1	0	19	17	10	1
			16.4%	10.3%	20.0%	11.8%	18.4%	12.5%	14.3%	1.7%	5.1%	5.3%	11.1%	7.1%	5.3%	0.0%	9.1%	12.7%	11.4%	5.6%
	greater than 120000		2	1	2	3	3	1	9	2	3	1	1	1	1	1	10	14	7	0
			3.6%	3.4%	6.7%	8.8%	7.9%	4.2%	18.4%	3.4%	7.7%	5.3%	5.6%	7.1%	5.3%	4.5%	4.8%	10.4%	8.0%	0.0%
	Total Responses		55	29	30	34	38	24	49	58	39	19	18	14	19	22	208	134	88	18

Test Response Homogeneity

Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
Job Satisfaction	Gender = M, Age Group	4.65556	0.9685	4.59049	0.9703
Job Satisfaction	Gender = F, Age Group	23.8406	0.0214*	24.9917	0.0149*
Salary Group	Gender = M, Age Group	22.0707	0.5750	23.8074	0.4727
Salary Group	Gender = F, Age Group	27.6209	0.2764	26.3992	0.3332
Job Satisfaction	Position Tenure	8.00461	0.2378	6.76025	0.3436
Salary Group	Position Tenure	25.8336	0.0113*	24.0752	0.0199*

The structured tab report contains the table that you specified in the Structured tab. The tests for response homogeneity are for each combination of grouping variables. We see that, for males, there is no difference in job satisfaction across age groups (Pearson p-value = 0.9703). For females, there is a difference in job satisfaction across age groups (Pearson p-value = 0.0149). The middle-aged females tend to be the least satisfied with their jobs. Share and frequency charts can be added to your report to visualize your results.

Example of a Multiple Response with Nonresponse

This example uses the Color Preference Survey.jmp sample data table, which contains survey data about people's color preferences. You can use the Categorical platform to summarize results from a multiple response question with nonresponses. A nonresponse is a participant who did not answer the question. Surveys often use "none of the above" or "not applicable" as a response choice. This provides a method for distinguishing responders who do not see an appropriate response from those who did not answer the question.

1. Select **Help > Sample Data Library** and open Color Preference Survey.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Click the **Structured** tab.
4. Select What colors do you like? (with nonresponse) and drag it to the green arrow drop zone at the **Side** of the table.

Note: This column contains multiple response data where the response levels are delimited with commas. There are three rows that contain no responses. This column was constructed for this example and does not align with the responses in the individual "What colors do you like" columns.

5. Click the **Top** green arrow and select What is your gender?
6. Click **Add=>**.
7. Click **OK**.

Figure 3.44 Initial Cross Tabulation

Categorical

What colors do you like? (with nonresponse) By What is your gender?

		Freq / Share / Rate	What is your gender?	
			Female	Male
What colors do you like? (with nonresponse)	Blue		31 21.5% 83.8%	38 21.5% 88.4%
	Green		24 16.7% 64.9%	36 20.3% 83.7%
	None of the above		2 1.4% 5.4%	0 0.0% 0.0%
	Orange		12 8.3% 32.4%	23 13.0% 53.5%
	Pink		19 13.2% 51.4%	9 5.1% 20.9%
	Purple		28 19.4% 75.7%	23 13.0% 53.5%
	Red		18 12.5% 48.6%	33 18.6% 76.7%
	Yellow		10 6.9% 27.0%	15 8.5% 34.9%
	Total Responses		144	177
	Total Cases		37	43
	Total Cases Responding		36	41

Note that the number of total cases is higher than the number of total cases responding. There are 37 Females who filled out the survey but only 36 responded to this question.

8. Click the Categorical red triangle and select **Test Multiple Responses > Exclude Nonresponses**.
9. Click the Categorical red triangle and select **Test Multiple Responses > Homogeneity Test, Binomial**.

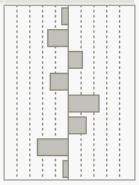
Figure 3.45 Binomial Homogeneity Test Results

Test Each Response, Binomial

Note: Nonresponse rows are excluded.

What is your gender?, What colors do you like? (with nonresponse)

What colors do you like? (with nonresponse)	ChiSquare	Prob>ChiSq
Blue	0.8918	0.3450
Green	5.0612	0.0245*
None of the above	3.1020	0.0782
Orange	4.0516	0.0441*
Pink	7.9930	0.0047*
Purple	4.1132	0.0426*
Red	8.0999	0.0044*
Yellow	0.6819	0.4089



Assuming multiple responses are Check All That Apply (CATA).

The analysis supports that blue and yellow have similar preference levels across genders while other colors tend to be favored more or less across genders. For example, Red appears to be liked more by males than by females. The *p*-value for the test is 0.0044, which supports a conclusion that there is a difference in preference across genders. From the crosstab in Figure 3.44 you find that 76.7% of the males liked the color red as compared to 48.6% of the females.

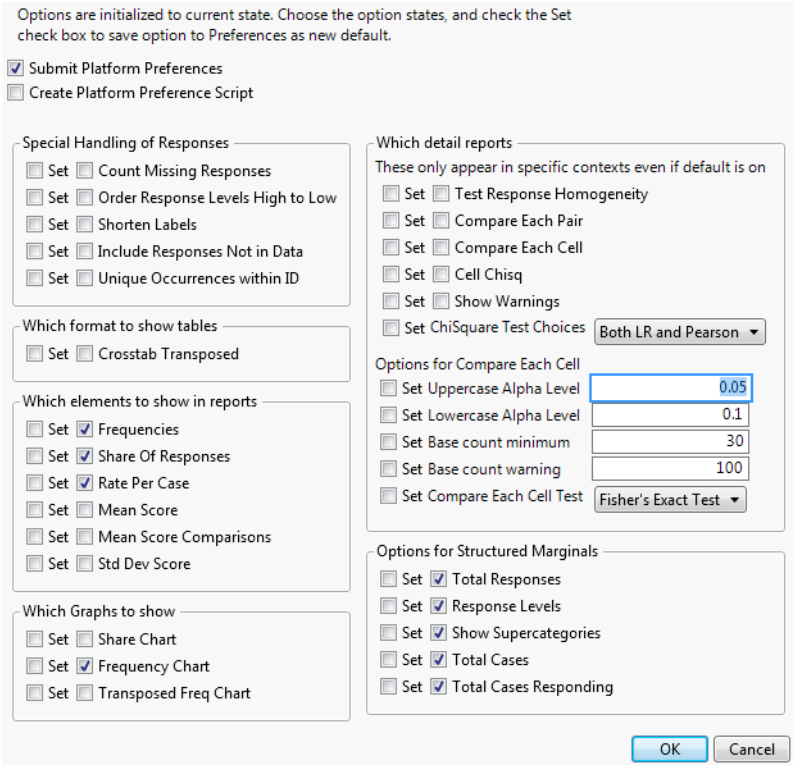
The note at the top of the output indicates that nonresponse rows are excluded from the analysis. The note at the bottom of the output indicates that it is assumed that the multiple response is a check all that applies type question. This means that a respondent can select more than one response to the question. The tests are performed using only those rows with responses. The three rows that did not have responses for the question are not included.

Tip: You can obtain the same output using the Multiple tab. Select What colors do you like? (with nonresponse) as a Multiple Delimited response and What is your Gender? as a grouping variable.

Set Preferences

The Categorical red triangle menu has a Set Preferences option to enable you to specify settings and preferences.

Figure 3.46 Set Preferences Window



Select the **Set** box for the options that you want to set. Select the option box if you want the option to appear by default, or deselect the option box if you do not want the option to appear by default. To submit the changes that you make to the platform preferences, select the **Submit Platform Preferences** box. To save the changes that you make as a preference script, select the **Create Platform Preference Script** box. When the Categorical platform is launched, the preferences associated with the current preference set are used to create the Categorical report.

Running the saved script submits the preferences to the platform preferences. You can use the platform preference script to share a preference set among multiple users, or to save the settings for specific projects.

Statistical Details for the Categorical Platform

This section contains statistical details for the Categorical platform.

Rao-Scott Correction

The Rao-Scott correction is applied to the test of response homogeneity for multiple responses. See Lavassani et al. (2009).

In the case of a multiple response, you can have overlapping samples, meaning a single participant can provide more than one response. The Pearson chi-square test is not appropriate for multiple responses, because the multiple responses violate the Pearson chi-square test assumption of independence. In addition, expected values calculated using the marginal totals are influenced by the multiple responses because the totals are larger than if multiple responses were not allowed.

The Rao-Scott chi-square statistic is defined as follows:

$$\chi_C^2 = \frac{\chi^2}{\bar{\delta}}$$

where

χ^2 is the standard Pearson Chi-squared statistic

$\bar{\delta} = 1 - \frac{m_{++}}{n_+ C}$ is the correction factor

The correction factor contains the following quantities:

m_{++} is the total count of the multiple responses

n_+ is the total number of participants and

C is the number of response levels (number of columns in the Crosstab table).

The degrees of freedom are $(R\text{-}1)C$ or the number of rows minus 1 times the number of columns.

Chapter 4

Choice Models

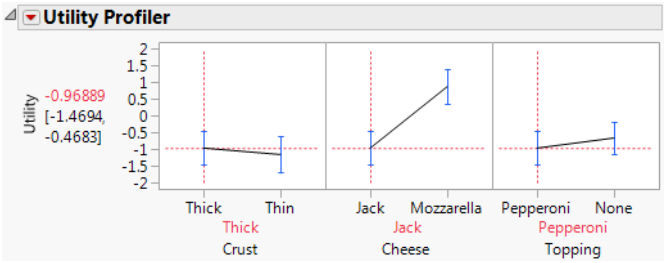
Fit Models for Choice Experiments

Use the Choice platform to analyze the results of choice experiments conducted in the course of market research. Choice experiments are used to help discover which product or service attributes your potential customers prefer. You can use this information to design products or services that have the attributes that your customers most desire.

The Choice platform enables you to do the following:

- Use information about subject (customer) traits as well as product attributes.
- Analyze choice experiments where respondents were allowed to select “none of these”.
- Integrate data from one, two, or three sources.
- Use the integrated profiler to understand, visualize, and optimize the response (utility) surface.
- Obtain subject-level scores for segmenting or clustering your data.
- JMP PRO Estimate subject-specific coefficients using a Bayesian approach.
- Use bias-corrected maximum likelihood estimators (Firth 1993).

Figure 4.1 Choice Platform Utility Profiler



Contents

Overview of the Choice Modeling Platform 83
Examples of the Choice Platform 85
One Table Format with No Choice 85
Multiple Table Format 88
Launch the Choice Platform 95
Launch Window for One Table, Stacked 96
Launch Window for Multiple Tables, Cross-Referenced 98
Choice Model Report 104
Effect Summary 104
Parameter Estimates 105
Likelihood Ratio Tests 106
Bayesian Parameter Estimates 106
Choice Platform Options 108
Willingness to Pay 111
Additional Examples 114
Example of Making Design Decisions 114
Example of Segmentation 122
Example of Logistic Regression Using the Choice Platform 126
Example of Logistic Regression for Matched Case-Control Studies 129
Example of Transforming Data to Two Analysis Tables 131
Example of Transforming Data to One Analysis Table 136
Statistical Details for the Choice Platform 138
Special Data Table Rules 138
Utility and Probabilities 139
Gradients 140

Overview of the Choice Modeling Platform

Choice modeling, pioneered by McFadden (1974), is a powerful analytic method used to estimate the probability of individuals making a particular choice from presented alternatives. Choice modeling is also called conjoint choice modeling, discrete choice analysis, and conditional logistic regression.

A choice experiment studies customer preferences for a set of product or process (in the case of a service) attributes. Respondents are presented sets of product attributes, called *profiles*. Each respondent is shown a small set of profiles, called a *choice set*, and asked to select the preference that he or she most prefers. Each respondent is usually presented with several choice sets. Use the Choice platform to analyze the results of a choice experiment.

Note: You can design your choice experiment using the Choice Design platform. See the *Design of Experiments Guide*.

Because customers vary in how they value attributes, many market researchers view market segmentation as an important step in analyzing choice experiments. Otherwise, you risk designing a product or process that pleases the "average" customer, who does not actually exist, and ignoring the preferences of market segments that *do* exist.

For background on choice modeling, see Louviere et al. (2015), Train (2009), and Rossi et al. (2005).

The Choice Platform

The Choice Modeling platform uses a form of conditional logistic regression to estimate the probability that a configuration is preferred. Unlike simple logistic regression, choice modeling uses a linear model to model choices based on response attributes and not solely upon subject characteristics. In choice modeling, a respondent might choose between two cars that are described by a combination of ten attributes, such as price, passenger load, number of cup holders, color, GPS device, gas mileage, anti-theft system, removable-seats, number of safety features, and insurance cost.

The Choice platform allows respondents to *not* make a choice from among a set of profiles. The *no choice* option is treated as a product with a single attribute ("Select none of these") that respondents are allowed to select. The parameter estimate for the No Choice attribute can then be interpreted in many ways, depending on the assumptions of the model. The Choice platform also enables you to obtain subject-level information, which can be useful in segmenting preference patterns.

You can obtain bias-corrected maximum likelihood estimators as described by Firth (1993). This method has been shown to produce better estimates and tests than MLEs without bias correction. In addition, bias-corrected MLEs improve separation problems that tend to occur in logistic-type models. See Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

Note: The Choice platform is not appropriate to use for fitting models that involve ranking, scoring, or nested hierarchical choices. You can use PROC MDC in SAS/ETS for these analyses.

Choice Designs in Developing Products and Services

Although customer satisfaction surveys can disclose what is wrong with a product or service, they fail to identify consumer preferences with regard to specific product attributes. When engineers design a product, they routinely make hundreds or thousands of small design decisions. If customer testing is feasible and research participants (subjects) are available, you can use choice experiments to guide some design decisions.

Decreases in survey deployment, modeling, and prototyping costs facilitate the customer evaluation of many attributes and alternatives as a product is designed. Choice modeling can be used in Six Sigma programs to improve consumer products, or, more generally, to make the products that people want. Choice experiments obtain data on customer preferences, and choice modeling analysis reveals such preferences.

Segmentation

Market researchers sometimes want to analyze the preference structure for each subject separately in order to see whether there are groups of subjects that behave differently. However, there are usually not enough data to do this with ordinary estimates. If there are sufficient data, you can specify the subject identifier as a "By groups" in the Response Data or you could introduce a subject identifier as a subject-side model term. This approach, however, is costly if the number of subjects is large.

If there are not sufficient data to specify "By groups," you can segment in JMP by clustering subjects using the Save Gradients by Subject option. The option creates a new data table containing the average Hessian-scaled gradient on each parameter for each subject. For an example, see "Example of Segmentation". For more information about the gradient values, see "Gradients".

JMP PRO In JMP Pro, you can request that the Choice platform use a Hierarchical Bayes approach in order to facilitate market segmentation. Bayesian modeling provides subject-specific estimates of model parameters (also called part-worths). These parameters i can be analyzed with hierarchical clustering or some other type of cluster analysis to reveal market segments.

Examples of the Choice Platform

In a study of pizza preferences, each respondent is presented with four choice sets, each containing two profiles. The Choice platform can analyze data that is in a one table format or a multiple data format. In the multiple table format, information about responses, choice sets, and subjects is saved in different data tables. In the one table format, that information is contained in a single data table.

- "One Table Format with No Choice" shows how to analyze a subset of the available data in a one table format.
- "Multiple Table Format" shows how to bring together information from different tables into one Choice analysis

One Table Format with No Choice

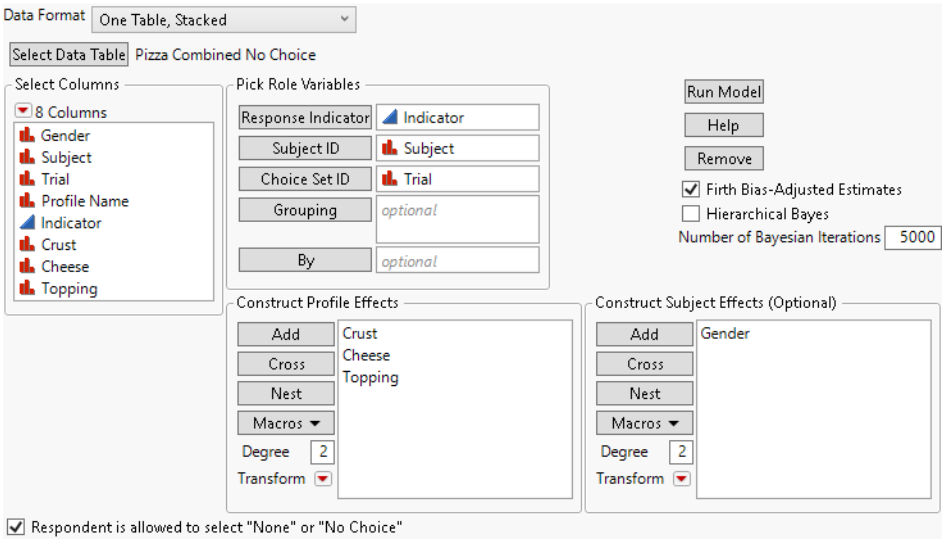
In this example, some respondents do not express a preference for either profile. The respondent makes "no choice". When a respondent does not express a preference, the respondent's choice indicator is entered as missing.

1. Select **Help > Sample Data Library** and open Pizza Combined No Choice.jmp.

   Choice sets are defined by the combination of Subject and Trial. Notice that there are missing values in the Indicator column for some choice sets.
2. Select **Analyze > Consumer Research > Choice**.

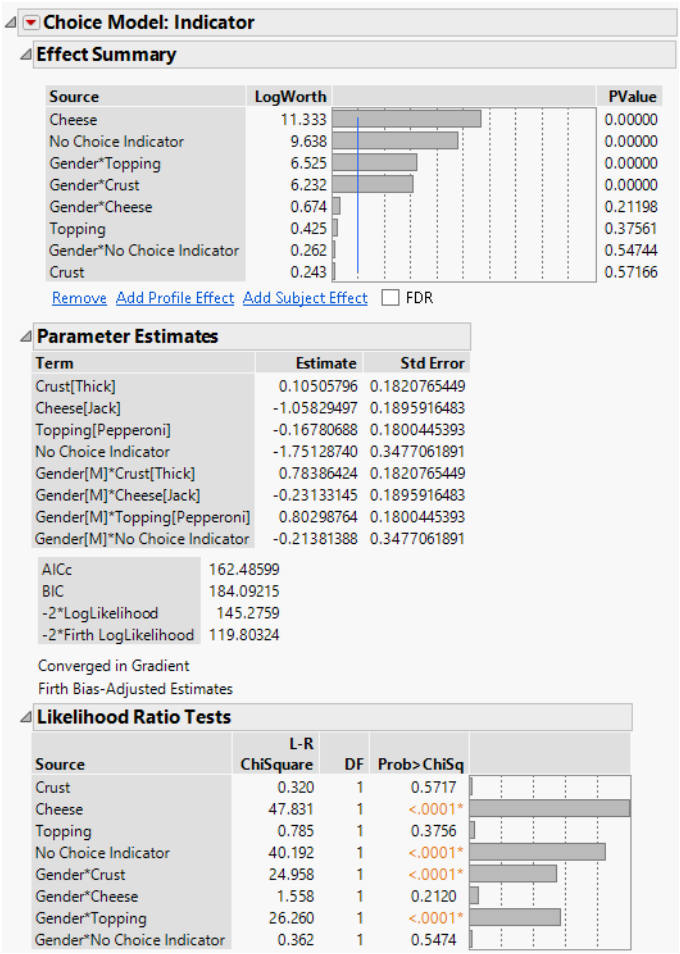
   The One Table, Stacked data format is the default.
3. Click **Select Data Table**.
4. Select **Pizza Combined No Choice** and click **OK**.
5. Complete the launch window:
   - Select Indicator and click **Response Indicator**.
   - Select Subject and click **Subject ID**.
   - Select Trial and click **Choice Set ID**.
   - Select Crust, Cheese, and Topping and click **Add** in the Construct Profile Effects panel.
   - Select Gender and click **Add** in the Construct Subject Effects (Optional) panel.

Figure 4.2 Completed Launch Window



6. Check the box next to **Respondent is allowed to select “None” or “No Choice”**.
7. Click **Run Model**.

Figure 4.3 Report Showing No Choice as an Effect



The Effect Summary report shows the effects in order of significance. Cheese is the most significant effect, followed by the No Choice Indicator, which is treated as a model effect. The subject effect interactions Gender*Topping and Gender*Crust are also significant, indicating that preferences for Topping and Crust depend on Gender market segments.

To get some insight on the nature of the No Choice responses, select and view those choice sets that resulted in No Choice.

8. In the data table, right-click in a cell in the Indicator column where the response is missing and select **Select Matching Cells**.
9. In the Rows panel, right-click **Selected** and select **Data View**.

Figure 4.4 Choice Sets with No Choice Responses

	Gender	Subject	Trial	Profile Name	Indicator	Crust	Cheese	Topping
1	F	2	4	TrimPepperjack	•	Thin	Jack	Pepperoni
2	F	2	4	TrimOni	•	Thin	Mozzarella	Pepperoni
3	M	7	2	Trimella	•	Thin	Mozzarella	None
4	M	7	2	TrimJack	•	Thin	Jack	None
5	M	7	3	Trimella	•	Thin	Mozzarella	None
6	M	7	3	TrimJack	•	Thin	Jack	None
7	F	8	2	ThickElla	•	Thick	Mozzarella	None
8	F	8	2	ThickJack	•	Thick	Jack	None
9	M	11	4	ThickOni	•	Thick	Mozzarella	Pepperoni
10	M	11	4	ThickJackoni	•	Thick	Jack	Pepperoni
11	F	14	2	TrimOni	•	Thin	Mozzarella	Pepperoni
12	F	14	2	TrimPepperjack	•	Thin	Jack	Pepperoni
13	F	18	3	ThickJack	•	Thick	Jack	None
14	F	18	3	ThickElla	•	Thick	Mozzarella	None
15	F	24	3	ThickJack	•	Thick	Jack	None
16	F	24	3	TrimOni	•	Thin	Mozzarella	Pepperoni
17	M	29	1	TrimPepperjack	•	Thin	Jack	Pepperoni
18	M	29	1	ThickJackoni	•	Thick	Jack	Pepperoni

In the table in Figure 4.4, consider the profiles in the first seven choice sets, which are defined by the Subject and Trial combinations in rows 1 to 14. The only difference within each choice set is the Cheese. There is an indication that some respondents might not be able to detect the difference in cheeses. However, the analysis takes the No Choice Indicator into account and concludes that, despite this behavior, Cheese is significant.

To see how to further analyze data of this type, see "Find Optimal Profiles".

Multiple Table Format

In this example, you examine pizza choices where three attributes, with two levels each, are presented to the respondents. The study was designed such that the respondents had to make a choice. The analysis uses three data tables: Pizza Profiles.jmp, Pizza Responses.jmp, and Pizza Subjects.jmp. Although you can always arrange your data into a single table, a multi-table approach can be more convenient than a one table analysis when you have additional profile and subject variables that you want to include in your analysis.

1. Select **Help > Sample Data Library** and open Pizza Profiles.jmp, Pizza Responses.jmp, and Pizza Subjects.jmp.
   - The profile data table, Pizza Profiles.jmp, lists all the pizza choice combinations that you want to present to the subjects. Each choice combination is given an ID.
   - The responses data table, Pizza Responses.jmp, contains the design and results. For the experiment, each subject is given four choice sets, where each choice set consists of two

choice profiles (Choice1 and Choice2). The subject selects a preference (Choice) for each choice set. For information about how to construct a choice design, see the *Design of Experiments Guide*. Notice that each value in the Choice column is an ID value in the Profile data table that contains the attribute information.

– The subjects data table, Pizza Subjects.jmp, includes a Subject ID column and a single characteristic of the subject, Gender. Each value of Subject in the Pizza Subjects.jmp data table corresponds to values in the Subject column in the Pizza Responses.jmp data table.

2. Select **Analyze > Consumer Research > Choice** to open the launch window.

Note: This can be done from any of the three open data tables.

3. From the Data Format menu, select **Multiple Tables, Cross-Referenced**.

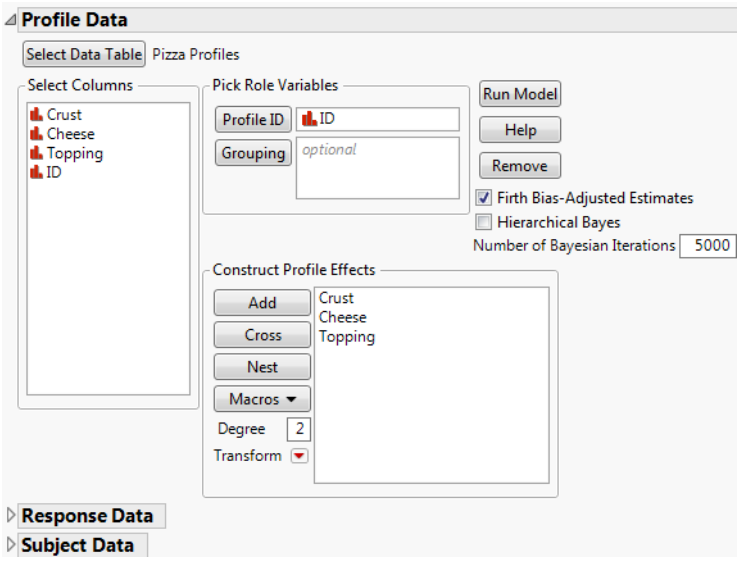
   There are three separate sections, one for each of the data sources.

4. Click **Select Data Table** under Profile Data.

   A Profile Data Table window appears, which prompts you to specify the data table for the profile data.

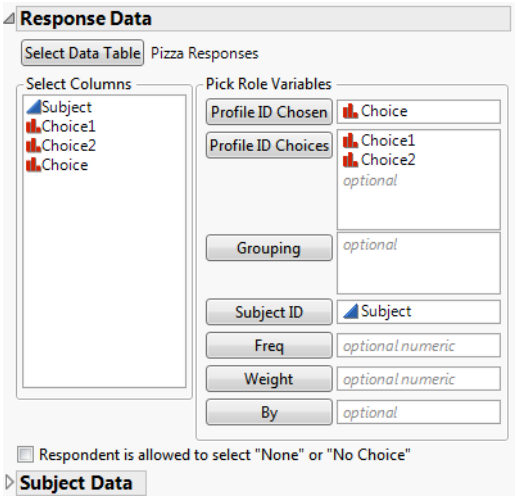
5. Select Pizza Profiles and click **OK**.
6. Select ID and click **Profile ID**.
7. Select Crust, Cheese, and Topping and click **Add**.

Figure 4.5 Profile Data



8. Click the disclosure icon next to Response Data to open the outline and click **Select Data Table**.
9. Select Pizza Responses and click **OK**.
10. Enter Profile ID selections:
    - Select Choice and click **Profile ID Chosen**.
    - Select Choice1 and Choice2 and click **Profile ID Choices**.
    - Select Subject and select **Subject ID**.

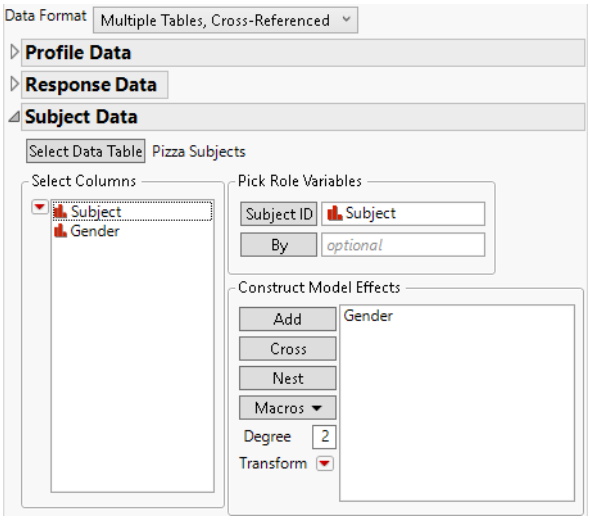
Figure 4.6 Response Data Window



Choice1 and Choice2 are the profiles presented to a subject in each of four choice sets. The Choice column contains the chosen preference between Choice1 and Choice2.

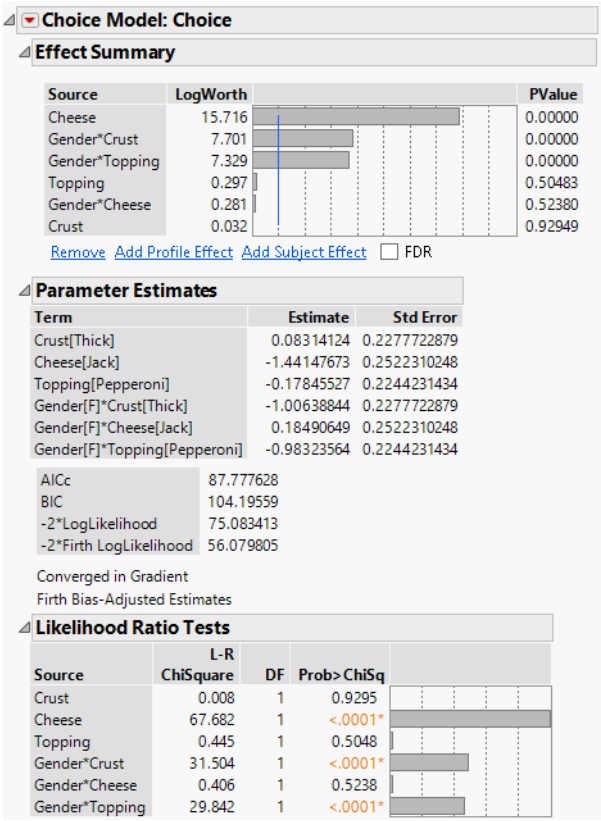
11. Click the disclosure icon next to Subject Data to open the outline and click **Select Data Table**.
12. Select Pizza Subjects and click **OK**.
13. Select Subject and click **Subject ID**.
14. Select Gender and click **Add**.

Figure 4.7 Subject Data Window



15. Click **Run Model**.

Figure 4.8 Choice Model Results



Choice Model: Choice

Effect Summary

Source	LogWorth	PValue
Cheese	15.716	0.00000
Gender*Crust	7.701	0.00000
Gender*Topping	7.329	0.00000
Topping	0.297	0.50483
Gender*Cheese	0.281	0.52380
Crust	0.032	0.92949

Remove Add Profile Effect Add Subject Effect FDR

Parameter Estimates

Term	Estimate	Std Error
Crust[Thick]	0.08314124	0.2277722879
Cheese[Jack]	-1.44147673	0.2522310248
Topping[Pepperoni]	-0.17845527	0.2244231434
Gender[F]*Crust[Thick]	-1.00638844	0.2277722879
Gender[F]*Cheese[Jack]	0.18490649	0.2522310248
Gender[F]*Topping[Pepperoni]	-0.98323564	0.2244231434

AICc	87.777628
BIC	104.19559
-2*LogLikelihood	75.083413
-2*Firth LogLikelihood	56.079805

Converged in Gradient
Firth Bias-Adjusted Estimates

Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Crust	0.008	1	0.9295
Cheese	67.682	1	<.0001*
Topping	0.445	1	0.5048
Gender*Crust	31.504	1	<.0001*
Gender*Cheese	0.406	1	0.5238
Gender*Topping	29.842	1	<.0001*

Six effects are entered into the model. The effects Crust, Cheese, and Topping are product attributes. The interaction effects, Gender*Crust, Gender*Cheese, and Gender*Topping are subject-effect interactions with the attributes. These interaction effects enable you to construct products that meet market-segment preferences.

Note: For Choice models, subject effects cannot be entered as main effects. They appear only as interaction terms.

The Effect Summary and Likelihood Ratio Tests reports show strong interactions between Gender and Crust and between Gender and Topping. Notice that the main effects of Crust and Topping are not significant. If you had not included subject-level effects, you might have overlooked important information relative to market segmentation.

Find Optimal Profiles

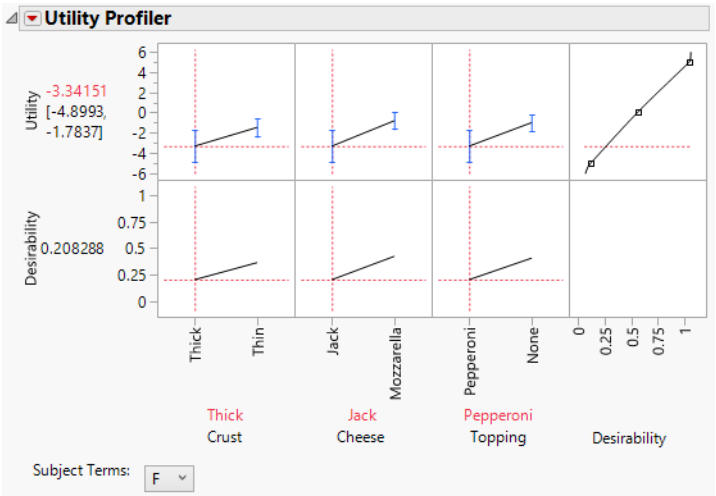
Next, you use the Utility Profiler to explore your results and find optimal settings for the attributes.

1. Click the Choice Model: Choice red triangle and select **Utility Profiler**.

   The Subject Terms menu beneath the profiler indicates that it is showing results for females.

2. Click the Utility Profiler red triangle and select **Optimization and Desirability > Desirability Functions**.

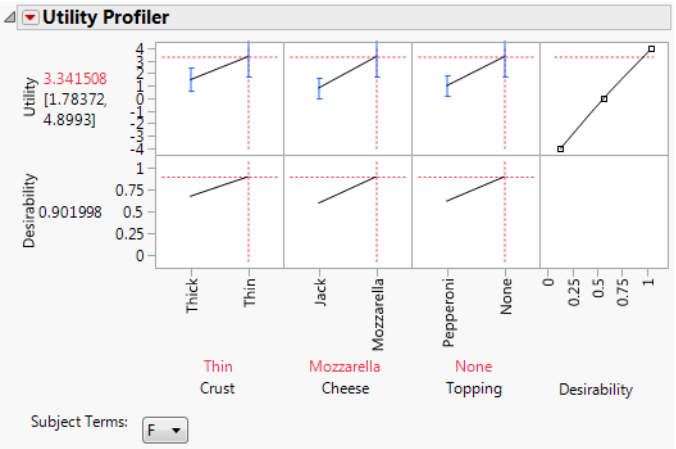
Figure 4.9 Utility Profiler with Desirability Function



   A desirability function that maximizes utility is added to the profiler. See *Profilers*.

3. Click the Utility Profiler red triangle and select **Optimization and Desirability > Maximize Desirability**.

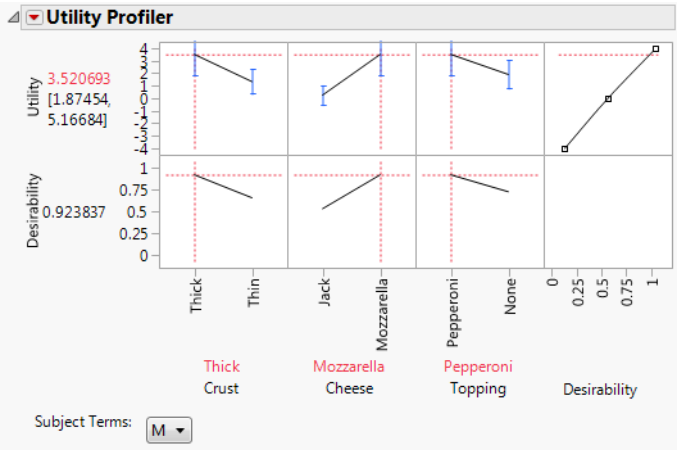
Figure 4.10 Utility Profiler with Optimal Settings for Females



The optimal settings for females are a thin crust, Mozzarella cheese, and no topping.

4. From the Subject Terms menu, select **M**.
5. Click the Utility Profiler red triangle and select **Optimization and Desirability > Maximize Desirability.**

Figure 4.11 Utility Profiler with Male Level Factor Setting



The optimal settings for males are a thick crust, Mozzarella cheese, and a Pepperoni topping.

In this example, understanding the preferences of gender-defined market segments enables you to provide two pizza choices that appeal to two segments of customers.

Launch the Choice Platform

Launch the Choice platform by selecting **Analyze > Consumer Research > Choice**.

Your data for the Choice platform can be combined in a single data table or it can reside in two or three separate data tables. In the Choice launch window specify whether you are using one or multiple data tables in the Data Format list.

One Table, Stacked

For the One Table, Stacked format, the data are in a single data table. There is a row for every profile presented to a subject and an indicator of whether that profile was selected. The Pizza Combined No Choice.jmp sample data table contains the results of a choice experiment in a single table format. See "One Table Format with No Choice".

For more information about the launch window for this format, see "Launch Window for One Table, Stacked".

Multiple Tables, Cross-Referenced

For the Multiple Tables, Cross-Referenced format, the data are in two or three separate data tables. A profile data table and a response data table are required. A subject data table is optional. Note the following:

- The profile data table must contain a column with a unique identifier for each profile and columns for the profile level variables. The profile identifier is used in the response data table to identify the profiles presented and the profile selected.
- The optional subject data table must contain a column with a unique subject identifier for each subject and columns for the subject level variables. The subject identifier is used in the response table to identify the subjects.

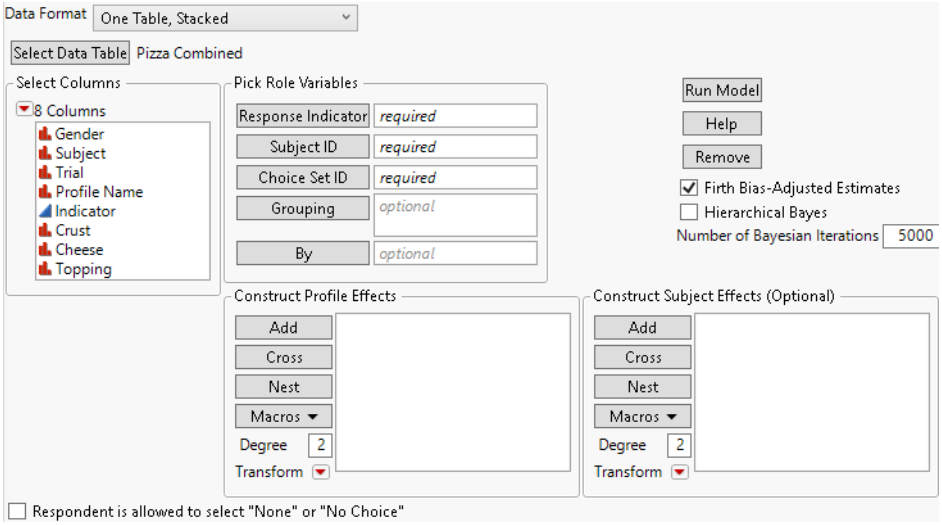
The launch window for this format contains three sections: Profile Data, Response Data, and Subject Data. Each section corresponds to a different data table. You can expand or collapse each section as needed.

The Pizza Profiles.jmp, Pizza Responses.jmp, and Pizza Subjects.jmp sample data tables contain the results of a choice experiment using three tables. There is one table for the profiles, one for the responses, and one for the subject information. See "Multiple Table Format".

For more information about the launch window for this format, see "Launch Window for Multiple Tables, Cross-Referenced".

Launch Window for One Table, Stacked

Figure 4.12 Launch Window for One Table, Stacked Data Format



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Select Data Table Select or open the data table that contains the combined data. Select Other to open a file that is not already open.

Response Indicator A column that contains values that indicate the preferred choice. A 1 indicates the preferred profile and a 0 indicates the other profiles. If respondents are given an option to select no preference, enter missing values for choice sets where no preference is indicated. See "Respondent is allowed to select "None" or "No Choice"".

Subject ID An identifier for the study participant.

Choice Set ID An identifier for the choice set presented to the subject for a given preference determination.

Grouping A column which, when used with the Choice Set ID column, uniquely designates each choice set. For example, if a choice set has Choice Set ID = 1 for Survey = A, and another choice set has Choice Set ID = 1 for Survey = B, then Survey should be used as a Grouping column.

By Produces a separate report for each level of the By Variable. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

Construct Profile Effects Add effects constructed from the attributes in the profiles.

For information about the Construct Profile Effects panel, see *Fitting Linear Models*.

Note: The choice model observes the column coding property of continuous profile and subject effects.

Construct Subject Effects (Optional) Add effects constructed from subject-related factors.

For information about the Construct Subject Effects panel, see *Fitting Linear Models*.

Firth Bias-adjusted Estimates Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. See Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

JMP PRO **Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See "Bayesian Parameter Estimates".

JMP PRO **Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject-specific parameters. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

Respondent is allowed to select "None" or "No Choice" Enters a No Choice Indicator into the model for response rows containing missing values. For the One Table, Stacked data format, the No Choice rows must contain (numeric) missing values in the Response Indicator column. The option appears at the bottom of the launch window.

Launch Window for Multiple Tables, Cross-Referenced

Figure 4.13 Launch Window for Multiple Tables, Cross-Referenced Data Format

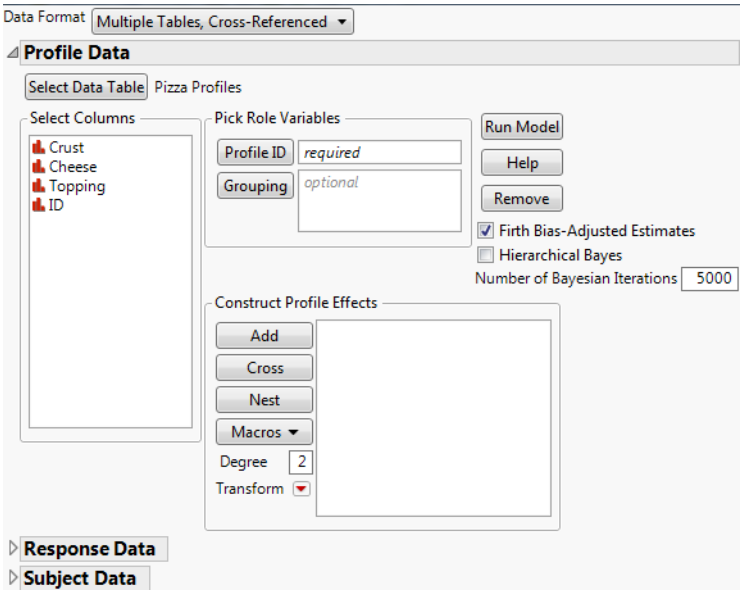


Figure 4.13 shows the launch window for Multiple Tables, using Pizza Profiles.jmp as the Profile table.

In the case of Multiple Tables, Cross-referenced, the launch window has three sections:

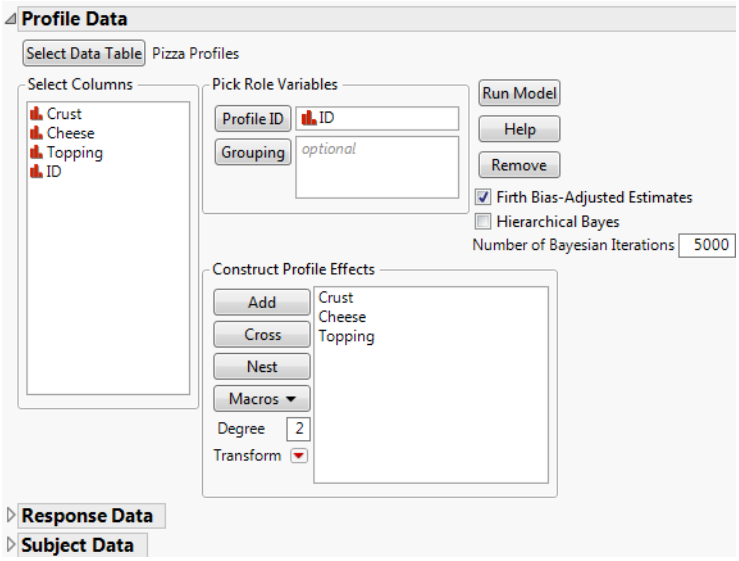
- "Profile Data"
- "Response Data"
- "Subject Data"

Profile Data

The profile data table describes the attributes associated with each choice. Each attribute defines a column in the data table. There is a row for each profile. A column in the table contains a unique identifier for each profile. Figure 4.14 shows the Pizza Profiles.jmp data table and a completed Profile Data panel.

Figure 4.14 Profile Data Table and Completed Profile Data Outline

	Crust	Cheese	Topping	ID
1	Thick	Mozzarella	Pepperoni	ThickOni
2	Thick	Mozzarella	None	ThickElla
3	Thick	Jack	Pepperoni	ThickJackoni
4	Thick	Jack	None	ThickJack
5	Thin	Mozzarella	Pepperoni	TrimOni
6	Thin	Mozzarella	None	Trimella
7	Thin	Jack	Pepperoni	TrimPepperjack
8	Thin	Jack	None	TrimJack



Select Data Table Select or open the data table that contains the profile data. Select Other to open a file that is not already open.

Profile ID Identifier for each row of attribute combinations (profile). If the **Profile ID** column does not uniquely identify each row in the profile data table, you need to add **Grouping** columns. Add **Grouping** columns until the combination of **Grouping** and **Profile ID** columns uniquely identify the row, or profile.

Grouping A column which, when used with the Profile ID column, uniquely designates each choice set. For example, if Profile ID = 1 for Survey = A, and a different Profile ID = 1 for Survey = B, then Survey would be used as a **Grouping** column.

Construct Profile Effects Add effects constructed from the attributes in the profiles.

For information about the Construct Profile Effects panel, see *Fitting Linear Models*.

Note: The choice model observes the column coding property of continuous profile and subject effects.

Firth Bias-adjusted Estimates Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. See Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

JMP PRO **Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See "Bayesian Parameter Estimates".

JMP PRO **Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject effects. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

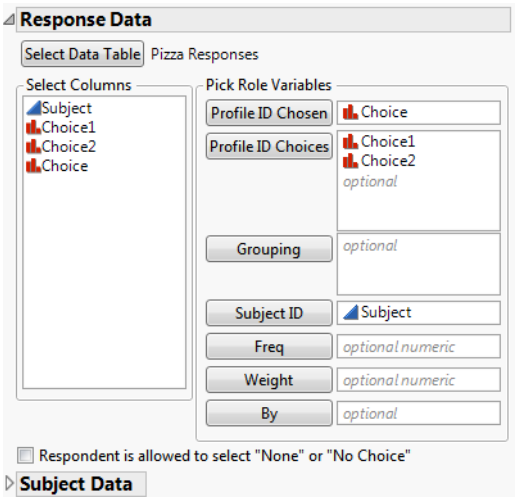
Response Data

The response data table includes a subject identifier column, columns that list the profile identifiers for the profiles in each choice set, and a column containing the preferred profile identifier. There is a row for each subject and choice set. Grouping variables can be used to distinguish choice sets when the data contain more than one group of choice sets. Figure 4.15 shows the Pizza Responses.jmp data table and a completed Response Data panel.

Grouping variables can be used to align choice indices when more than one group is contained within the data.

Figure 4.15 Response Data Table and Completed Responses Data Outline

	Subject	Choice1	Choice2	Choice
1	1	ThickJack	TrimPepperjack	TrimPepperjack
2	1	TrimPepperjack	ThickElla	ThickElla
3	1	TrimOni	Trimella	TrimOni
4	1	ThickElla	ThickJack	ThickElla
5	2	Trimella	ThickJackoni	Trimella
6	2	TrimJack	ThickElla	ThickElla
7	2	Trimella	TrimPepperjack	Trimella
8	2	TrimPepperjack	TrimOni	TrimOni
9	3	TrimOni	ThickJackoni	TrimOni
10	3	TrimPepperjack	ThickElla	ThickElla
11	3	ThickJackoni	TrimPepperjack	ThickJackoni
12	3	ThickOni	Trimella	ThickOni
13	4	ThickElla	ThickOni	ThickElla
14	4	TrimPepperjack	ThickJack	ThickJack



Select Data Table Select or open the data table that contains the response data. Select Other to open a file that is not already open.

Profile ID Chosen The Profile ID from the Profile data table that represents the subject's selected profile.

Grouping A column which, when used with the Profile ID Chosen column, uniquely designates each choice set.

Profile ID Choices The Profile IDs of the set of possible profiles. There must be at least two profiles.

Subject ID An identifier for the study participant.

Freq A column containing frequencies. If *n* is the value of the Freq variable for a given row, then that row is used in computations n times. If it is less than 1 or missing, then JMP does not use it to calculate any analyses.

Weight A column containing a weight for each observation in the data table. The weight is included in analyses only when its value is greater than zero.

By Produces a separate report for each level of the By Variable. If more than one By variable is assigned, a separate analysis is produced for each possible combination of the levels of the By variables.

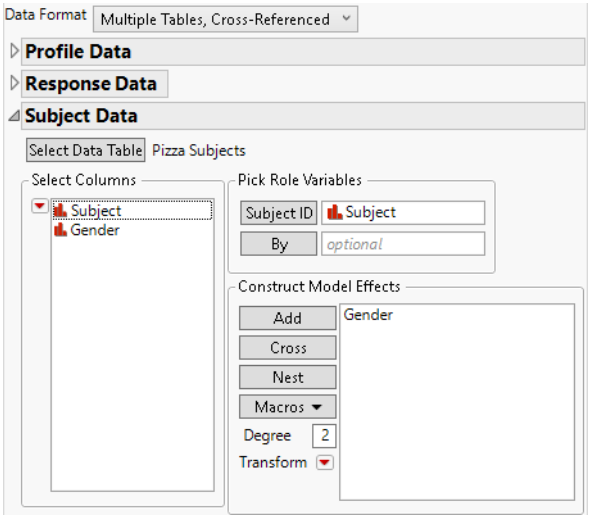
Respondent is allowed to select "None" or "No Choice" Enters a No Choice Indicator into the model for response rows containing missing values. For the Multiple Tables, Cross-Referenced data format, the No Choice rows must contain (categorical) missing values in the Profile ID Chosen column in the Response Data table. The option appears at the bottom of the Response Data panel.

Subject Data

The subject data table is optional and depends on whether you want to model subject effects. The table contains a column with the subject identifier used in the response table, and columns for attributes or characteristics of the subjects. You can put subject data in the response data table, but you should specify the subject effects in the Subject Data outline. Figure 4.16 shows the Pizza Subjects.jmp data table and a completed Subject Data panel.

Figure 4.16 Subject Data Table and Completed Subject Data Outline

	Subject	Gender
1	1	M
2	2	F
3	3	M
4	4	F



Select Data Table Select or open the data table that contains the subject data. Select Other to open a file that is not already open.

Subject ID Unique identifier for the subject.

By Produces a separate report for each level of the By variable. If more than on By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

Construct Model Effects Add effects constructed from columns in the subject data table.

For information about the Construct Model Effects panel, see *Fitting Linear Models*.

Choice Model Report

- "Effect Summary"
- "Parameter Estimates"
- "Likelihood Ratio Tests"
- "Bayesian Parameter Estimates"

Effect Summary

The Effect Summary report appears if your model contains more than one effect and if it can be calculated quickly. (If the report does not appear, select Likelihood Ratio Tests from the red triangle menu to make both reports appear.) It lists the effects estimated by the model and gives a plot of the LogWorth (or FDR LogWorth) values for these effects. The report also provides controls that enable you to add or remove effects from the model. The model fit report updates automatically based on the changes made in the Effects Summary report. See *Fitting Linear Models*.

Note: The Effect Summary report is not applicable to models fit with Hierarchical Bayes.

Effect Summary Table Columns

The Effect Summary table contains the following columns:

Source Lists the model effects, sorted by ascending *p*-values.

LogWorth Shows the LogWorth for each model effect, defined as $-\log_{10}(p\text{-value})$. This transformation adjusts *p*-values to provide an appropriate scale for graphing. A value that exceeds 2 is significant at the 0.01 level (because $-\log_{10}(0.01) = 2$).

FDR LogWorth Shows the False Discovery Rate LogWorth for each model effect, defined as $-\log_{10}(\text{FDR PValue})$. This is the best statistic for plotting and assessing significance. Select the **FDR** check box to replace the LogWorth column with the **FDR LogWorth** column.

Bar Chart Shows a bar chart of the LogWorth (or FDR LogWorth) values. The graph has dashed vertical lines at integer values and a blue reference line at 2.

PValue Shows the *p*-value for each model effect. This is the *p*-value corresponding to the significance test displayed in the Likelihood Ratio Tests report.

FDR PValue Shows the False Discovery Rate *p*-value for each model effect calculated using the Benjamini-Hochberg technique. This technique adjusts the *p*-values to control the false

discovery rate for multiple tests. Select the **FDR** check box to replace the **PValue** column with the **FDR PValue** column.

For more information about the FDR correction, see Benjamini and Hochberg (1995). For more information about the false discovery rate, see *Predictive and Specialized Modeling* or Westfall et al. (2011).

Effect Summary Table Options

The options below the summary table enable you to add and remove effects:

Remove Removes the selected effects from the model. To remove one or more effects, select the rows corresponding to the effects and click the Remove button.

Add Profile Effect Opens a panel that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Profile Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

Add Subject Effect Opens a panel that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Subject Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

Parameter Estimates

The Parameter Estimates report gives estimates and standard errors of the coefficients of utility associated with the effects listed in the Term column. The coefficients associated with attributes are sometimes referred to as *part-worths*. When the Firth Bias-Adjusted Estimates option is selected in the launch window, the parameter estimates are based on the Firth bias-corrected maximum likelihood estimators. These estimates considered to be more accurate than MLEs without bias correction. For more information about utility, see "Utility and Probabilities".

Comparison Criteria

The AICc (corrected Akaike's Information Criterion), BIC (Bayesian Information Criterion), –2Loglikelihood, and –2Firth Loglikelihood fit statistics are shown as part of the report and can be used to compare models. See *Fitting Linear Models*.

The –2Firth Loglikelihood fit statistic is included in the report when the Firth Bias-Adjusted Estimates option is selected in the launch window. Note that this option is checked by default. The decision to use or not use the Firth Bias-Adjusted Estimates does not affect the AICc score or the –2Loglikelihood results.

Note: For each of these statistics, a smaller value indicates a better fit.

Likelihood Ratio Tests

The Likelihood Ratio Test report appears by default if the model is fit in less than five seconds. If the report does not appear, you can select the Likelihood Ratio Tests option from the Choice Model red triangle menu. The report gives the following:

Source Lists the effects in the model.

L-R ChiSquare The value of the likelihood ratio ChiSquare statistic for a test of the corresponding effect.

DF The degrees of freedom for the ChiSquare test.

Prob>ChiSq The *p*-value for the ChiSquare test.

Bar Graph Shows a bar chart of the L-R ChiSquare values.

JMP PRO Bayesian Parameter Estimates

(Available only for Hierarchical Bayes.) The Bayesian Parameter Estimates report gives results for model effects. The estimates are based on a Hierarchical Bayes fit that integrates the subject-level covariates into the likelihood function and estimates their effects on the parameters directly. The subject-level covariates are estimated using a Bayesian procedure combined with the Metropolis-Hastings algorithm. See Train (2001). Posterior means and variances are calculated for each model effect. The algorithm also provides subject-specific estimates of the model effect parameters. See "Save Subject Estimates".

During the estimation process, each individual is assigned his or her own vector of parameter estimates, essentially treating the estimates as random effects and covariates. The vector of coefficients for an individual is assumed to come from a multivariate normal distribution with arbitrary mean and covariance matrix. The likelihood function for the utility parameters for a given subject is based on a multinominal logit model for each subject's preference within a choice set, given the attributes in the choice set. The prior distribution for a given subject's vector of coefficients is normal with mean equal to zero and a diagonal covariance matrix with the same variance for each subject. The covariance matrix is assumed to come from an inverse Wishart distribution with a scale matrix that is diagonal with equal diagonal entries.

For each subject, a number of burn-in iterations at the beginning of the chain is discarded. By default, this number is equal to half of the Number of Bayesian Iterations specified on the launch window.

Figure 4.17 Bayesian Parameter Estimates Report

Choice Model: Choice

Bayesian Parameter Estimates

Term	Posterior Mean	Posterior Std Dev	Subject Std Dev
Crust[Thick]	0.27529781	0.724891965	2.815958945
Cheese[Jack]	-7.01110292	4.697169974	2.593184268
Topping[Pepperoni]	-1.06702410	0.941602303	2.260232580

Total Iterations	5000
Burn-In Iterations	2500
Number of Respondents	32
Avg Log Likelihood After Burn-In	-14.19342

Term The model term.

Posterior Mean The parameter estimate for the term's coefficient. For each iteration after the burn-in period, the mean of the subject-specific coefficient estimates is computed. The Posterior Mean is the average of these means.

Tip: Select the red triangle option Save Bayes Chain to see the individual estimates for each iteration.

Posterior Std Dev The standard deviation of the means of the subject-specific estimates over the iterations after burn-in.

Subject Std Dev The standard deviation of the subject-specific estimates.

Tip: Select the red triangle option Save Subject Estimates to see the individual estimates.

Total Iterations The total number of iterations performed, including the burn-in period.

Burn-In Iterations The number of burn-in iterations. This number is equal to half of the Number of Bayesian Iterations specified on the launch window.

Number of Respondents The number of subjects.

Avg Log Likelihood After Burn-In The average of the log-likelihood function, computed on values obtained after the burn-in period.

Choice Platform Options

Choice Model red triangle menu contains the following options.

Note: When you use Hierarchical Bayes, the subject-level estimates are based on Monte Carlo sampling. For this reason, results obtained for the options below vary from run to run.

Likelihood Ratio Tests See "Likelihood Ratio Tests".

JMP PRO **Show MLE Parameter Estimates** (Available for Hierarchical Bayes) Shows non-Firth maximum likelihood estimates and standard errors for the coefficients of model terms. These estimates are used as starting values for the Hierarchical Bayes algorithm.

Joint Factor Tests (Not available for Hierarchical Bayes) Tests each factor in the model by constructing a likelihood ratio test for all the effects involving that factor. For more information about Joint Factor Tests, see *Fitting Linear Models*.

Confidence Intervals (Not available for Hierarchical Bayes) Shows or hides a confidence interval for each parameter in the Parameter Estimates report.

Confidence Limits (Available for Hierarchical Bayes) Shows or hides confidence limits for each parameter in the Bayesian Parameter Estimates report. The limits are constructed based on the 2.5 and 97.5 quantiles of the posterior distribution.

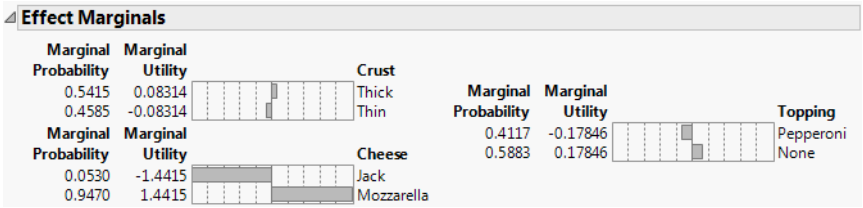
Correlation of Estimates If Hierarchical Bayes was not selected, shows the correlations between the maximum likelihood parameter estimates.

For Hierarchical Bayes, shows the correlation matrix for the posterior means of the parameter estimates. The correlations are calculated from the iterations after burn-in. The posterior means from each iteration after burn-in are treated as if they are columns in a data table. The Correlation of Estimates table is obtained by calculating the correlation matrix for these columns.

Effect Marginals Shows or hides marginal probabilities and marginal utilities for each main effect in the model. The marginal probability is the probability that an individual selects attribute A over B with all other attributes set to their mean or default levels.

In Figure 4.18, the marginal probability of any subject choosing a pizza with mozzarella cheese, thick crust and pepperoni, over that same pizza with Monterey Jack cheese instead of mozzarella, is 0.9470.

Figure 4.18 Example of Marginal Effects



Utility Profiler Shows or hides the predicted utility for different factor settings. The utility is the value predicted by the linear model. See "Find Optimal Profiles" for an example of the Utility Profiler. For more information about utility, see "Utility and Probabilities". For more information about the Utility Profiler options, see *Profilers*.

Probability Profiler Enables you to compare choice probabilities among a number of potential products. This predicted probability is defined as follows:

$$(\exp(U))/(\exp(U) + \exp(U_b))$$

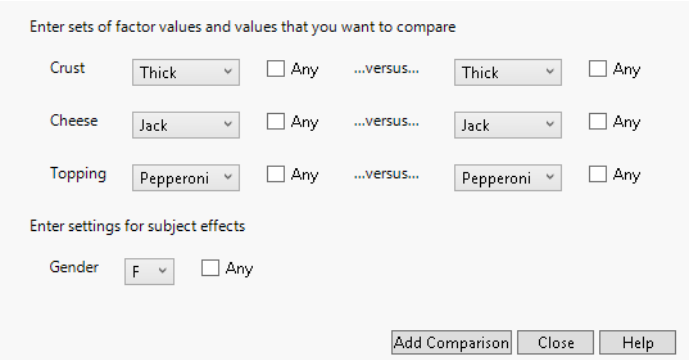
where U is the utility for the current settings and U_b is the utility for the baseline settings. This implies that the probability for the baseline settings is 0.5. See "Utility and Probabilities".

See "Comparisons to Baseline" for an example of using the Probability Profiler. For more information about the Probability Profiler options, see *Profilers*.

Multiple Choice Profiler Provides the number of probability profilers that you specify. This enables you to set each profiler to the settings of a given profile so that you can compare the probabilities of choosing each profile relative to the others. See "Multiple Choice Comparisons" for an example of using the Multiple Choice Profiler. For more information about the Multiple Choice Profiler options, see *Profilers*.

Comparisons Performs comparisons between specific alternative choice profiles. Enables you to select the factors and the values that you want to compare. You can compare specific configurations, including comparing all settings on the left or right by selecting the **Any** check boxes. If you have subject effects, you can select the levels of the subject effects to compare. Using Any does not compare all combinations across features, but rather all combinations of comparisons, one feature at a time, using the left settings as the settings for the other factors.

Figure 4.19 Utility Comparisons Window



Willingness to Pay Requires that your model includes a continuous price column. Calculates the maximum price increase (decrease) that a customer is willing to pay for a new feature over the baseline feature cost. The result is calculated using the Baseline settings for each background setting.

Save Utility Formula When the analysis is on multiple data tables, creates a new data table that contains a formula column for utility. The new data table contains a row for each subject and profile combination, and columns for the profiles and the subject effects. When the analysis is on one data table, a new Utility Formula column is added.

Save Gradients by Subject (Not available for Hierarchical Bayes.) Constructs a new table that has a row for each subject containing the average (Hessian-scaled-gradient) steps for the likelihood function on each parameter. This corresponds to using a Lagrangian multiplier test for separating that subject from the remaining subjects. These values can later be clustered, using the built-in-script, to indicate unique market segments represented in the data. See "Gradients". For an example, see "Example of Segmentation".

JMP PRO **Save Subject Estimates** (Available for Hierarchical Bayes.) Creates a table where each row contains the subject-specific parameter estimates for each effect. The distribution of subject-specific parameter effects for each effect is centered at the estimate for the term given in the Bayesian Parameter Estimates report. The Subject Acceptance Rate gives the rate of acceptance for draws of new parameter estimates during the Metropolis-Hastings step. Generally, an acceptance rate of 0.20 is considered to be good. See "Bayesian Parameter Estimates".

JMP PRO **Save Bayes Chain** (Available for Hierarchical Bayes.) Creates a table that gives information about the chain of iterations used in computing subject-specific Bayesian estimates. See "Save Bayes Chain".

Model Dialog Shows the Choice launch window, which can be used to modify and re-fit the model. You can specify new data sets, new IDs, and new model effects.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Willingness to Pay

The term *willingness to pay* refers to the price that a customer is willing to pay for new features, calculated to match a customer's utility for baseline features. For example, suppose that a customer is willing to pay $1,000 for a computer with a 40 GB hard drive. Willingness to Pay for an 80 GB hard drive is calculated by setting the Hard drive feature to 80 GB and then solving for the price that delivers the same utility as the $1000 40 GB hard drive.

Willingness to Pay Launch Window Options

When you select the Willingness to Pay option, the Willingness to Pay launch window is shown. The launch window in Figure 4.20 is obtained by selecting the Willingness to Pay option in the report that results from running the **Choice** data table script in Laptop Profile.jmp.

Factor The variables from the analysis. These can be product features or subject-specific attributes.

Baseline The baseline setting for each factor. If the factor is categorical, select the baseline value from a list. If the factor is numeric, enter the baseline value.

Role The type of factor.

Feature Factor A product or service feature from the experiment that you want to price.

Price Factor A price factor in the experiment. The price factor must be continuous, and there can be only one specified price factor for each Willingness to Pay analysis.

Background Constant A factor that you want to hold constant in the Willingness to Pay calculation. Generally, these are subject-specific variables.

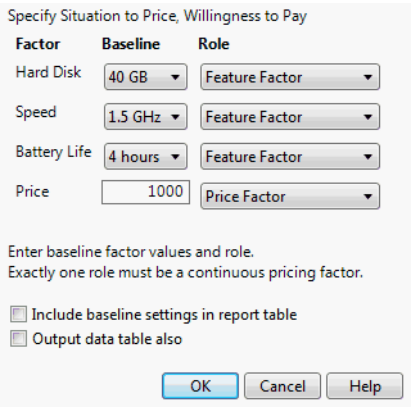
Background Variable A factor that you want to hold constant, at each of its levels, in the Willingness to Pay calculation. Generally, these are subject-level factors. Specifying a subject factor as a Background Variable rather than a Background Constant provides Willingness to Pay estimates for all levels of the variable.

Include baseline settings in report table Adds the baseline settings with a price change of zero to the Willingness to Pay report.

Tip: If you make an output table, use this option to display all the baseline settings as well as the attribute settings.

Output data table also Creates a data table containing the Willingness to Pay report.

Figure 4.20 Willingness to Pay Launch Window



Once you complete your first Willingness to Pay calculation, the platform remembers the baseline values and assigned roles that you selected. This enables you to do multiple Willingness to Pay comparisons without having to re-enter the baseline information. If there is no factor called Price, but there is a continuous factor used in the analysis, the continuous factor is automatically assigned as the Price factor in the Willingness to Pay window. Common cost variables that are not prices in the traditional sense include factors such as travel time or distance.

Willingness to Pay Report

The Willingness to Pay report displays the baseline value for each factor, as well as baseline utility values. For each factor, the report shows the feature setting, estimated price change, and new price. If there are no interaction or second-order effects, standard errors and confidence intervals are also shown. These are calculated using the delta method.

Figure 4.21 Willingness to Pay Report

Willingness to Pay

Factor	Baseline Value
Hard Disk	40 GB
Speed	1.5 GHz
Battery Life	4 hours
Price	1000

Baseline Utility
-3.4736

Factor	Feature Setting	Price Change	Std Error	Lower 95%	Upper 95%	New Price
Hard Disk	80 GB	$959.67	381.743	$211.47	$1,707.87	$1,959.67
Speed	2.0 GHz	$496.86	218.515	$68.58	$925.14	$1,496.86
Battery Life	6 hours	$387.88	191.088	$13.35	$762.40	$1,387.88

Standard deviations for Price Change calculated by Delta method.

JMP PRO Save Bayes Chain

(Available for models fit with Hierarchal Bayes.) You can use the Bayes Chain data table to determine whether your estimates have stabilized. The table that is created has a number of rows equal to the Number of Bayesian Iterations (specified on the launch window) plus one. The first row, Iteration 1, gives the starting values. Subsequent rows show the results of the iterations, in order. The table has a column for the iteration counts, the model Log Likelihood, and columns corresponding to each model effect:

Iteration Gives the iteration number, where the first row shows starting values.

Log Likelihood The log-likelihood of the model for that iteration. You can plot the Log Likelihood against Iteration to view behavior over the burn-in and tuning periods.

Adaptive Sigma for <model effect> Gives the estimate of the square root of the diagonal entries of the inverse Wishart distribution scale matrix for the corresponding effect.

Acceptance for <model effect> Gives the sampling acceptance rate for the corresponding effect.

Mean of <model effect> Gives the estimated mean for the corresponding effect.

Variance of <model effect> Gives the estimated variance for the corresponding effect.

Additional Examples

- "Example of Making Design Decisions"
- "Example of Segmentation"
- "Example of Logistic Regression Using the Choice Platform"
- "Example of Logistic Regression for Matched Case-Control Studies"
- "Example of Transforming Data to Two Analysis Tables"
- "Example of Transforming Data to One Analysis Table"

Example of Making Design Decisions

You can use the Choice Modeling platform to determine the relative importance of product attributes. Even if the attributes of a particular product that are important to the consumer are known, information about preference trade-offs with regard to these attributes might be unknown. By gaining such information, a market researcher or product designer is able to incorporate product features that represent the optimal trade-off from the perspective of the consumer. This example illustrates the advantages of this approach to product design.

It is already known that four attributes are important for laptop design: hard-disk size, processor speed, battery life, and selling price. The data gathered for this study are used to determine which of four laptop attributes (Hard Disk, Speed, Battery Life, and Price) are most important. It also assesses whether there are Gender or Job effects associated with these attributes.

This example is described in four sections:

- "Complete the Launch Window"
- "Analyze the Model"
- "Comparisons to Baseline"
- "Multiple Choice Comparisons"

Complete the Launch Window

1. Select **Help > Sample Data Library** and open Laptop Runs.jmp.

Note: If you prefer not to follow the manual steps in this section, click the green triangle next to the script **Choice with Gender** to run the model, and go to "Analyze the Model".

2. Click the green triangle next to the **Open Profile and Subject Tables** script.

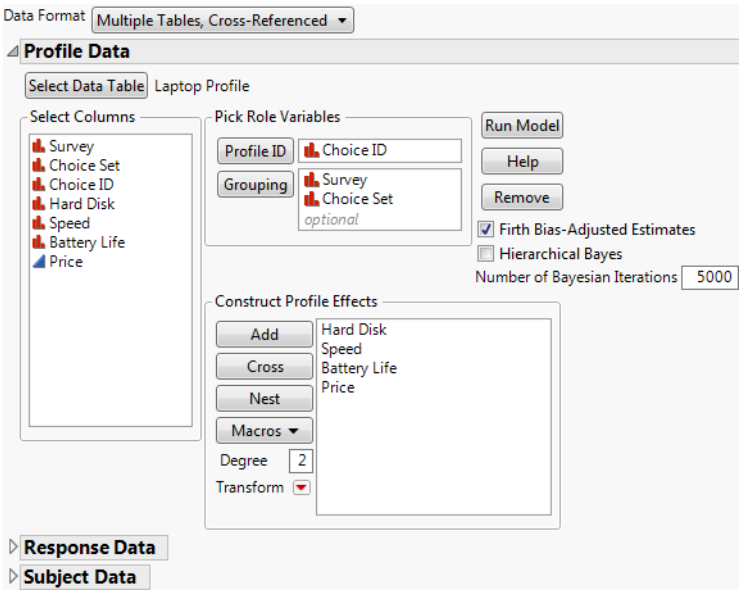
   The script opens the Laptop Profile.jmp and Laptop Subjects.jmp data tables.

3. Select **Analyze** > **Consumer Research** > **Choice**.

Note: This can be done from any of the three open data tables.

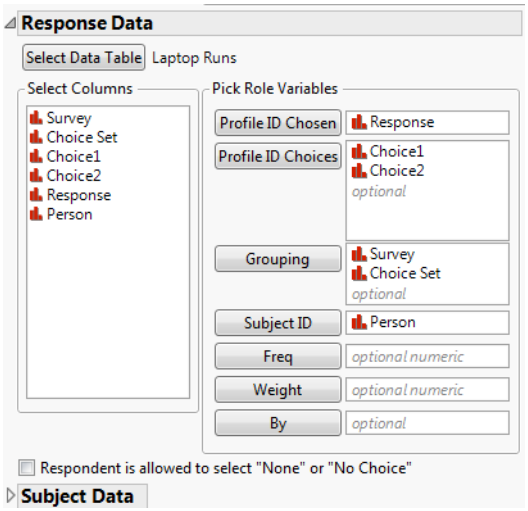
4. From the Data Format list, select **Multiple Tables, Cross-Referenced**.
5. Click **Select Data Table** under Profile Data and select Laptop Profile. Select Choice ID and click Profile ID.
6. Select Hard Disk, Speed, Battery Life, and Price and click **Add**.
7. Select Survey and Choice Set and click **Grouping**.

Figure 4.22 Profile Data Window for Laptop Study



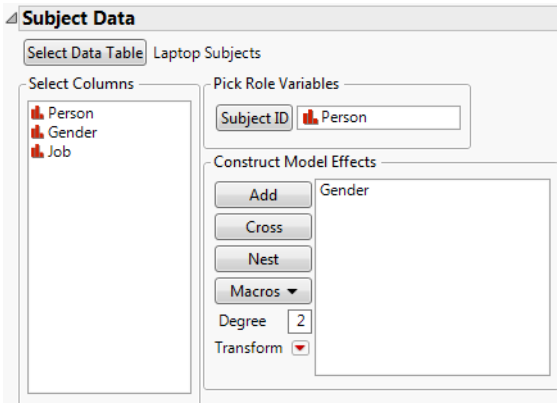
8. Open the **Response Data** outline.
9. From the **Select Data Table** list, select Laptop Runs.
10. Complete the Response Data table:
    - Select Response and click **Profile ID Chosen**.
    - Select Choice1 and Choice2 and click **Profile ID Choices**.
    - Select Survey and Choice Set and click **Grouping**
    - Select Person and click **Subject ID**.

Figure 4.23 Response Data Window for Laptop Study



11. Open the **Subject Data** outline.
12. From the **Select Data Table** list, select Laptop Subjects.
13. Select Person and click **Subject ID**.
14. Select Gender click **Add**.

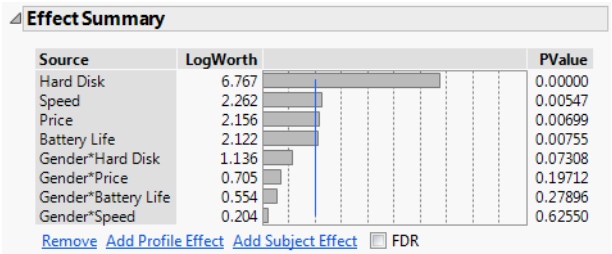
Figure 4.24 Subject Data Window for Laptop Study



Analyze the Model

1. Click **Run Model**.

Figure 4.25 Laptop Effect Summary



Effect Summary

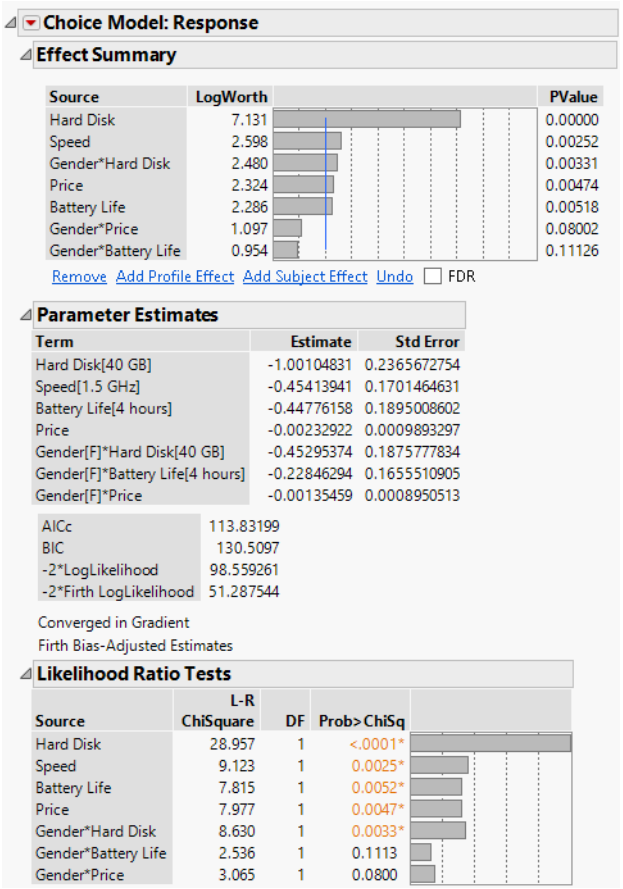
Source	LogWorth	PValue
Hard Disk	6.767	0.00000
Speed	2.262	0.00547
Price	2.156	0.00699
Battery Life	2.122	0.00755
Gender*Hard Disk	1.136	0.07308
Gender*Price	0.705	0.19712
Gender*Battery Life	0.554	0.27896
Gender*Speed	0.204	0.62550

Remove Add Profile Effect Add Subject Effect FDR

The Effect Summary report shows that Hard Disk is the most significant effect. You can reduce the model by removing terms with a *p*-value greater than 0.15. This process should be done one term at a time. Here, Gender*Speed is the least significant effect, with a *p*-value of 0.625.

2. In the Effect Summary report, select Gender*Speed and click **Remove**.

Figure 4.26 Laptop Results



Choice Model: Response

Effect Summary

Source	LogWorth	PValue
Hard Disk	7.131	0.00000
Speed	2.598	0.00252
Gender*Hard Disk	2.480	0.00331
Price	2.324	0.00474
Battery Life	2.286	0.00518
Gender*Price	1.097	0.08002
Gender*Battery Life	0.954	0.11126

Remove Add Profile Effect Add Subject Effect Undo FDR

Parameter Estimates

Term	Estimate	Std Error
Hard Disk[40 GB]	-1.00104831	0.2365672754
Speed[1.5 GHz]	-0.45413941	0.1701464631
Battery Life[4 hours]	-0.44776158	0.1895008602
Price	-0.00232922	0.0009893297
Gender[F]*Hard Disk[40 GB]	-0.45295374	0.1875777834
Gender[F]*Battery Life[4 hours]	-0.22846294	0.1655510905
Gender[F]*Price	-0.00135459	0.0008950513

AICc	113.83199
BIC	130.5097
-2*LogLikelihood	98.559261
-2*Firth LogLikelihood	51.287544

Converged in Gradient
Firth Bias-Adjusted Estimates

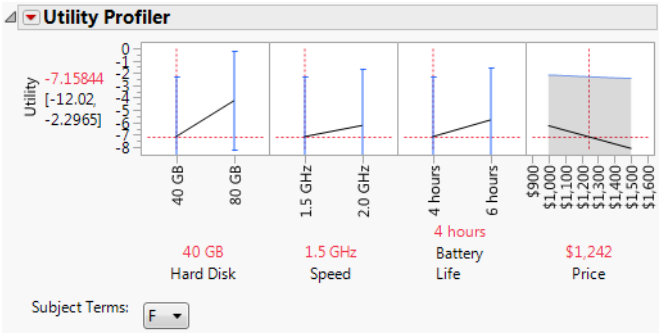
Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Hard Disk	28.957	1	<.0001*
Speed	9.123	1	0.0025*
Battery Life	7.815	1	0.0052*
Price	7.977	1	0.0047*
Gender*Hard Disk	8.630	1	0.0033*
Gender*Battery Life	2.536	1	0.1113
Gender*Price	3.065	1	0.0800

Once Gender*Speed is removed from the model, all effects have a *p*-value of 0.15 or less. Therefore, you use this as your final model.

3. Click the Choice Model: Response red triangle and select **Utility Profiler**.

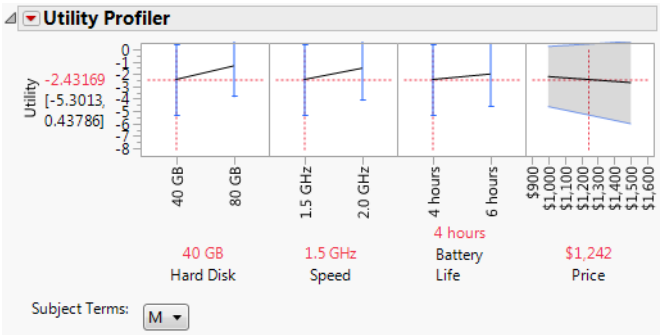
Figure 4.27 Laptop Profiler Results for Females



Tip: If your utility profiler does not look like Figure 4.27, click the Utility Profiler red triangle and select **Appearance > Adapt Y Axis**.

4. From the list next to Subject Terms, select **M**.

Figure 4.28 Laptop Profiler Results for Males in Development



The interaction effect between Gender and Hard Disk is highly significant, with a *p*-value of 0.0033 (Figure 4.26). In the Utility Profilers, check the slope for Hard Disk for both levels of Gender. You see that the slope is steeper for females than for males.

Comparisons to Baseline

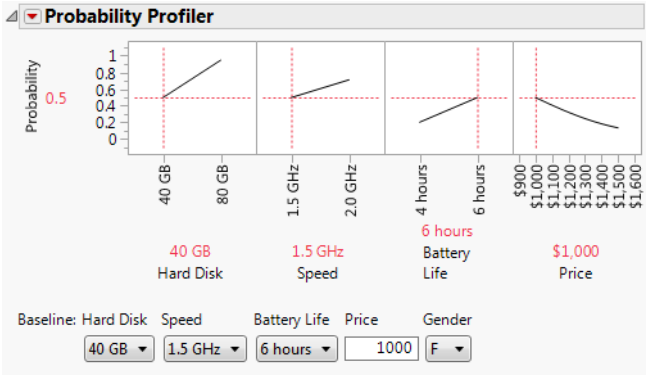
Suppose you are developing a new product. You want to explore the likelihood that a customer selects the new product over the old product, or over a competitor's product. Use the Probability Profiler to compare profiles to a baseline profile.

In this example, your company is currently producing laptops with 40 GB hard drives, 1.5 GHz processors, and 6-hour battery life, that cost $1,000. You are looking for a way to make your product more desirable by changing as few factors as possible. You set the current product configuration as the baseline. JMP adjusts the probabilities so that the probability of preference for the baseline configuration is 0.5. Then you compare the probabilities of other configurations to the baseline probability.

1. Do one of the following:
   - Follow the steps in "Complete the Launch Window". Then complete step 1 and step 2 in "Analyze the Model".
   - In the Laptop Runs.jmp sample data table, click the green triangle next to the **Choice Reduced Model** script.
2. Click the Choice Model: Response red triangle and select **Probability Profiler**.

   Note that the Probability Profiler is for Gender = F. You can change this later.
3. Using the menus and text box below the profiler, in the Baseline area, specify the Baseline settings as 40 GB, 1.5 GHz, 6 hours, and 1000.
4. Now set these as the values in the Probability Profiler. To set the Price at $1000, click $1242 above Price under the rightmost profiler cell, and type 1000. Then click outside the text box.

Figure 4.29 Probability Profiler with Text Entry Area for Price

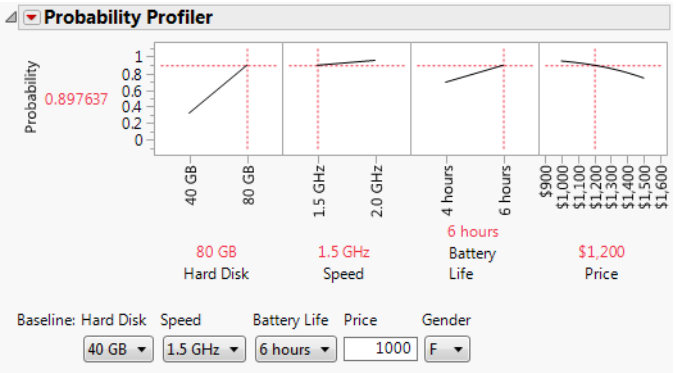


This configuration has probability 0.5.

5. In the Probability Profiler, move the slider for HardDisk to 80 GB.

   Notice that, with this change, the probability is relatively insensitive to increases in Price.
6. Click the $1000 label above the Price cell in the profiler, type **$1,200**, and click outside the text box.

Figure 4.30 Laptop Probability Profiler Results with Baseline Effects



An increase in Hard Disk size from 40 GB to 80 GB and an increase in price to $1200 coincides with an increased probability of preference, from 0.50 to 0.90 for females. Change the Gender effect in the Baseline to **M**. The probability of preference is 0.71.

Multiple Choice Comparisons

Use the Multiple Choice Profiler to compare product profiles.

- You currently produce a low-end laptop with a small hard drive, a slow processor, and low battery life. You charge $1000.
- Company A produces a product with a fast processor speed and high battery life at a reasonable price of $1200.
- Company B makes the biggest hard drives with the fastest speed, but at a high price of $1500 and low battery life.

You want to gain market share by increasing only one area of performance, and price.

1. Do one of the following:
   - Follow the steps in "Complete the Launch Window". Then complete step 1 and step 2 in "Analyze the Model".
   - In the Laptop Runs.jmp sample data table, click the green triangle next to the **Choice Reduced Model** script.
2. Click the Choice Model: Response red triangle and select **Multiple Choice Profiler**.

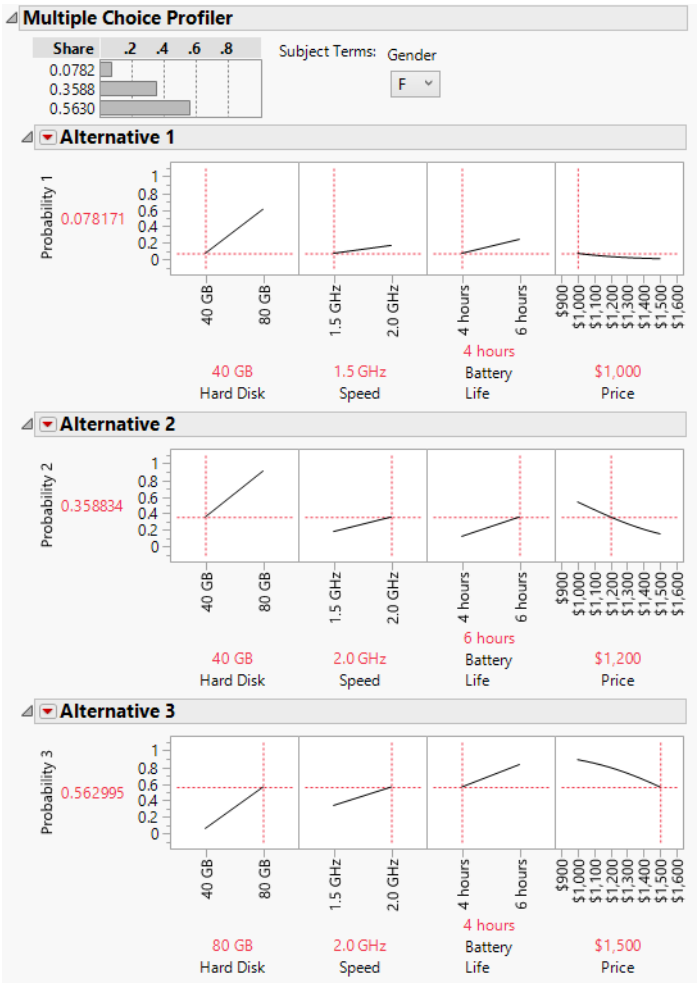
   A window appears, asking for the number of alternative choices to profile. Accept the default number of 3.
3. Click **OK**.

   Three Alternative profilers appear. Notice that the profilers are set for Gender = F.

Each factor in each profiler is set to its default values. Alternative 1 indicates the product that you want to develop. Alternative 2 indicates Company A's product. Alternative 3 indicates Company B's product.

4. For Alternative 1, set Hard Disk to 40 GB, Speed to 1.5 GHz, Battery Life to 4hours, and Price to $1,000.
5. For Alternative 2, set Hard Disk to 40 GB, Speed to 2.0 GHz, Battery Life to 6 hours, and Price to $1,200.
6. For Alternative 3, set Hard Disk to 80 GB, Speed to 2.0 GHz, Battery Life to 4 hours, and Price to $1,500.

Figure 4.31 Multiple Choice Profiler for Females

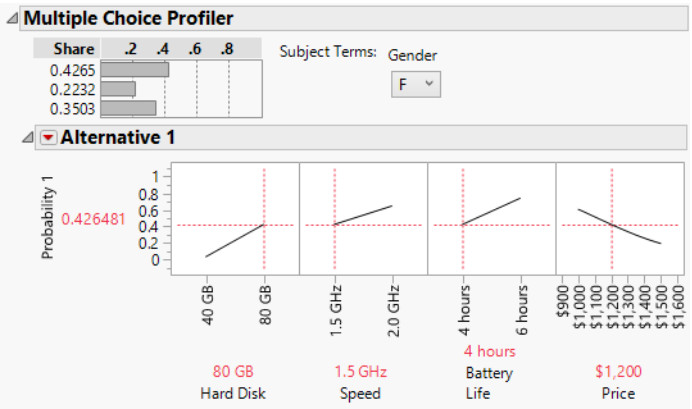


You can see that Company B has the greatest Share of 0.5630. It is obvious that with your company's settings, very few females buy your product.

You want to increase your market share by upgrading your company's laptop in one of the performance areas while increasing price. The slope of the line in Alternative 1's Hard Disk profile suggests increasing hard disk space increases market share the most.

7. For Alternative 1, set Hard Disk to **80 GB** and Price to **$1,200**.

Figure 4.32 Multiple Choice Profiler with Improved Laptop



By increasing hard disk space, you can increase the price of your laptop and expect a market share among females of about 43%. This share exceeds that of Company B's high-performance laptop and is much better than the market share with the initial low-end settings seen in Figure 4.31.

Explore the settings that increase your market share for males. If you increase both Hard Disk size and Speed, you can capture a 44% market share among males.

Example of Segmentation

In this example, you attempt to identify market segments for pizza preferences.

To see how to complete the launch window for this example, see step 1 to step 15 in the example "Multiple Table Format". Otherwise, follow the instructions below.

Define Clusters

1. Select **Help > Sample Data Library** and open Pizza Responses.jmp.
2. Click the green triangle next to the **Choice** script.
3. Click the Choice Model: Choice red triangle and select **Save Gradients by Subject**.

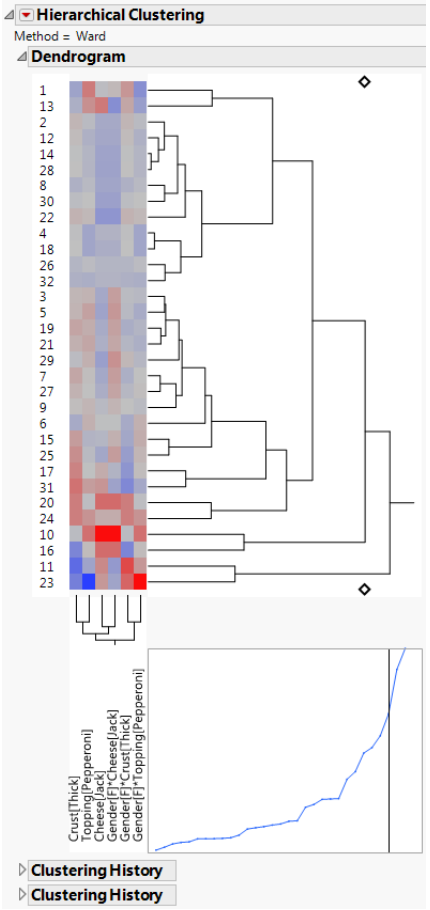
A data table appears with gradient forces saved for each main effect and subject interaction.

Figure 4.33 Gradients by Subject for Pizza Data, Partial View

	Subject	Crust[Thick]	Cheese[Jack]	Topping[Pepperoni]	Gender[F]*Crust[Thick]	Gender[F]*Cheese[Jack]	Gender[F]*Topping[Pepperoni]
1	1	-0.00959	-0.00168	0.014876	0.009585	0.001685	-0.01488
2	2	0.002373	-0.00758	-0.00239	0.002373	-0.00758	-0.00239
3	3	0.002129	-0.0079	0.003031	-0.00213	0.007899	-0.00303
4	4	-0.00106	-0.00485	-0.00901	-0.00106	-0.00485	-0.00901
5	5	0.002828	-0.00945	0.00725	-0.00283	0.009453	-0.00725
6	6	-0.0073	-0.00089	0.003761	-0.0073	-0.00089	0.003761

4. Click the green triangle next to the **Hierarchical Cluster** script.

Figure 4.34 Dendrogram of Subject Clusters for Pizza Data



The script runs a hierarchical cluster analysis on all columns in the gradient table, except for Subject. Click either diamond to see that the rows have been placed into three clusters.

5. Click the Hierarchical Clustering red triangle and select **Save Clusters**.

   A new column called Cluster is added to the data table containing the gradients. Each subject has been assigned a Cluster value that is associated with other subjects having similar gradient forces. See *Multivariate Methods* for a discussion of other Hierarchical Clustering options.

   You can delete the gradient columns because they were used only to obtain the clusters.

6. Select all columns except Subject and Cluster. Right-click the selected columns and select **Delete Columns**.

7. Click the green triangle next to the **Merge Data Back** script (Figure 4.33).

   The cluster information is merged into the Subject data table. The columns in the Subject data table are now Subject, Gender, and Cluster.

Figure 4.35 Subject Data with Cluster Column

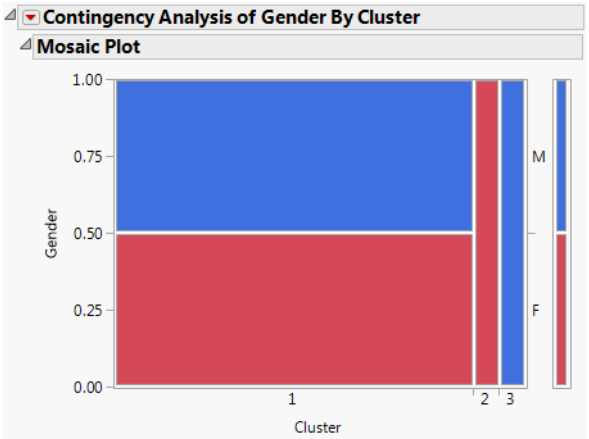
	Subject	Crust[Thick]	Cheese[Jack]	Topping[Pepperoni]	Gender[F]*Crust[Thick]	Gender[F]*Cheese[Jack]	Gender[F]*Topping[Pepperoni]	Cluster
1	1	-0.00959	-0.00168	0.014876	0.009585	0.001685	-0.01488	1
2	2	0.002373	-0.00758	-0.00239	0.002373	-0.00758	-0.00239	1
3	3	0.002129	-0.0079	0.003031	-0.00213	0.007899	-0.00303	1
4	4	-0.00106	-0.00485	-0.00901	-0.00106	-0.00485	-0.00901	1
5	5	0.002828	-0.00945	0.00725	-0.00283	0.009453	-0.00725	1
6	6	-0.0073	-0.00089	0.003761	-0.0073	-0.00089	0.003761	1
7	7	0.006003	-0.00815	0.000308	-0.006	0.008151	-0.00031	1
8	8	-0.0055	-0.00887	-0.00274	-0.0055	-0.00887	-0.00274	1
9	9	0.000438	-0.00271	0.002402	-0.00044	0.00271	-0.0024	1
10	10	-0.00217	0.032641	0.017043	-0.00217	0.032641	0.017043	2

This table can now be used for further analysis.

Explore the Clusters

1. Click the icon to the left of the Cluster variable in the columns panel and select **Nominal**.
2. Select **Analyze** > **Fit Y by X**.
3. Select Gender and click **Y, Response**.
4. Select Cluster and click **X, Factor**.
5. Click **OK**.

Figure 4.36 Contingency Analysis of Gender by Cluster



Contingency Table

Gender			
Count / Total % / Col % / Row %	F	M	
1	14	14	28
	43.75	43.75	87.50
	87.50	87.50	
	50.00	50.00	
2	2	0	2
	6.25	0.00	6.25
	12.50	0.00	
	100.00	0.00	
3	0	2	2
	0.00	6.25	6.25
	0.00	12.50	
	0.00	100.00	
	16	16	32
	50.00	50.00	

Cluster

Tests

N	DF	-LogLike	RSquare (U)
32	2	2.7725887	0.1250

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.545	0.0625
Pearson	4.000	0.1353

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

Observe patterns in the clusters:

- Cluster 1 is evenly divided between males and females
- Cluster 2 consists of only females
- Cluster 3 consists of only males

If desired, you could now refit and analyze the model with the addition of the Cluster variable.

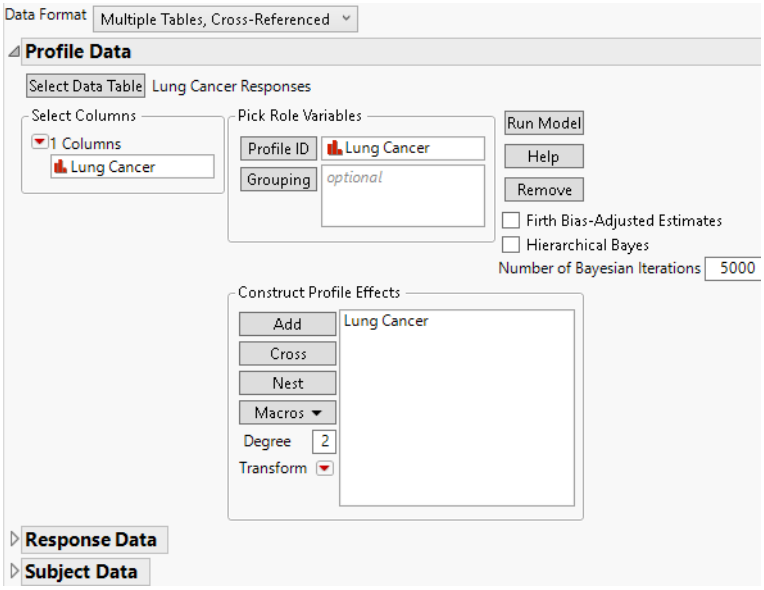
Example of Logistic Regression Using the Choice Platform

Use the Choice Platform

1. Select **Help > Sample Data Library** and open Lung Cancer Responses.jmp and Lung Cancer Choice.jmp.

   Notice Lung Cancer Responses.jmp has only one column (Lung Cancer) with two rows (Cancer and NoCancer).
2. Select **Analyze** > **Consumer Research** > **Choice**
3. Select **Multiple Tables, Cross-Referenced** from the list next to Data Format.
4. Click **Select Data Table**, select Lung Cancer Responses, and click **OK**.
5. Select Lung Cancer and click **Profile** ID.
6. Select Lung Cancer and click **Add**.
7. Uncheck the Firth Bias-Adjusted Estimates box.

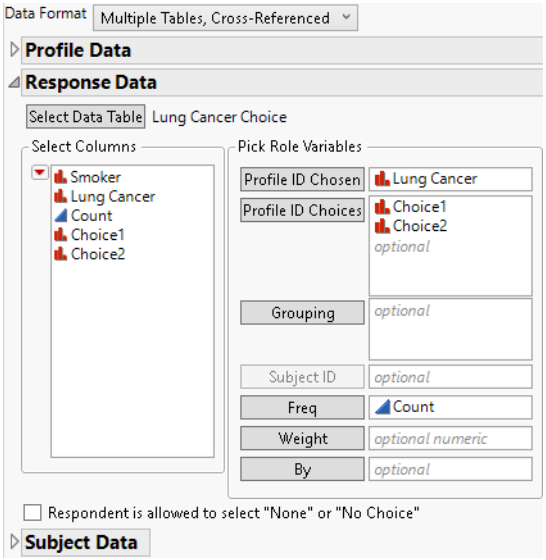
Figure 4.37 Completed Profile Data Panel



8. Open the **Response Data** outline.
9. Click **Select Data Table**, select Lung Cancer Choice, and click **OK**.
10. Complete the **Response Data** outline:
    - Select Lung Cancer and click **Profile ID Chosen**.
    - Select Choice1 and Choice2 and click **Profile ID Choices**.

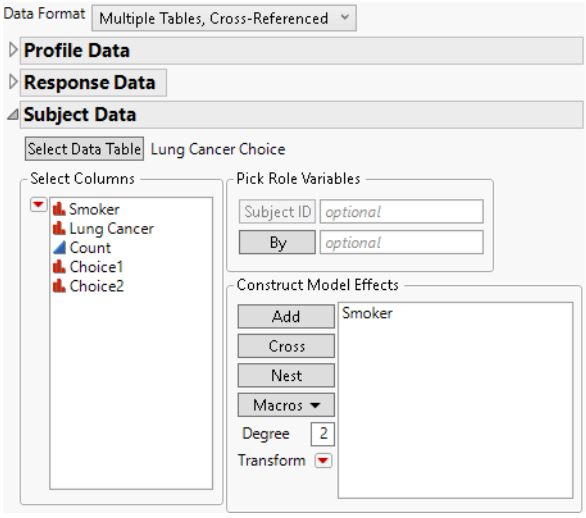
– Select Count and click **Freq**.

Figure 4.38 Completed Response Data Panel



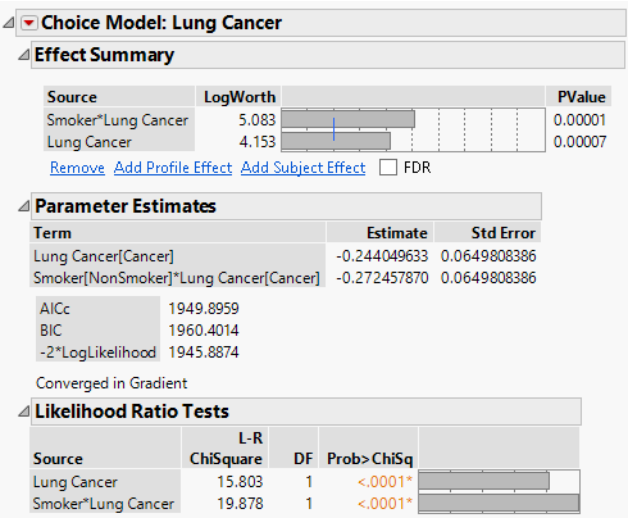
11. Open the **Subject Data** outline.
12. Click **Select Data Table**, select Lung Cancer Choice, and click **OK**.
13. Select Smoker and click **Add**.

Figure 4.39 Completed Subject Data Panel



14. Click **Run Model**.

Figure 4.40 Choice Modeling Logistic Regression Results

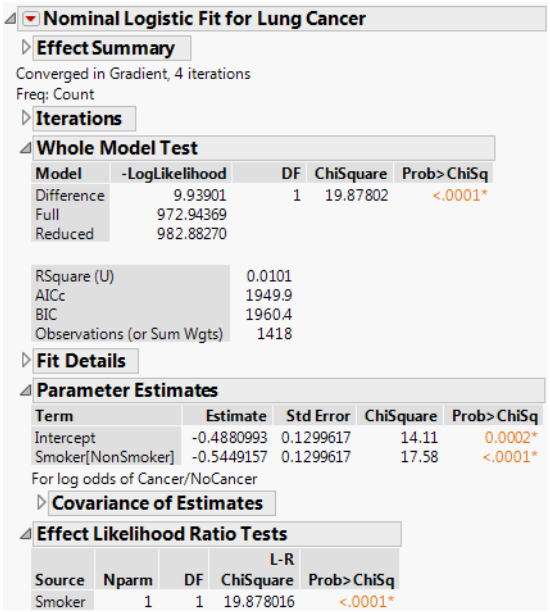


Use the Fit Model Platform

1. Select **Help > Sample Data Library** and open Lung Cancer.jmp.
2. Select **Analyze** > **Fit Model**.

   Because the data table contains a model script, the Model Specification window is automatically completed. The **Nominal Logistic** personality is selected.
3. Set **Target Level** to Cancer.
4. Click **Run**.

Figure 4.41 Fit Model Nominal Logistic Regression Results



Nominal Logistic Fit for Lung Cancer

Effect Summary

Converged in Gradient, 4 iterations
Freq: Count

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	9.93901	1	19.87802	<.0001*
Full	972.94369			
Reduced	982.88270			

RSquare (U)	0.0101
AICc	1949.9
BIC	1960.4
Observations (or Sum Wgts)	1418

Fit Details

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-0.4880993	0.1299617	14.11	0.0002*
Smoker[NonSmoker]	-0.5449157	0.1299617	17.58	<.0001*

For log odds of Cancer/NoCancer

Covariance of Estimates

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Smoker	1	1	19.878016	<.0001*

Notice that the likelihood ratio chi-square test for Smoker*Lung Cancer in the Choice model matches the likelihood ratio chi-square test for Smoker in the Logistic model. The reports shown in Figure 4.40 and Figure 4.41 support the conclusion that smoking has a strong effect on developing lung cancer. See *Fitting Linear Models*.

Example of Logistic Regression for Matched Case-Control Studies

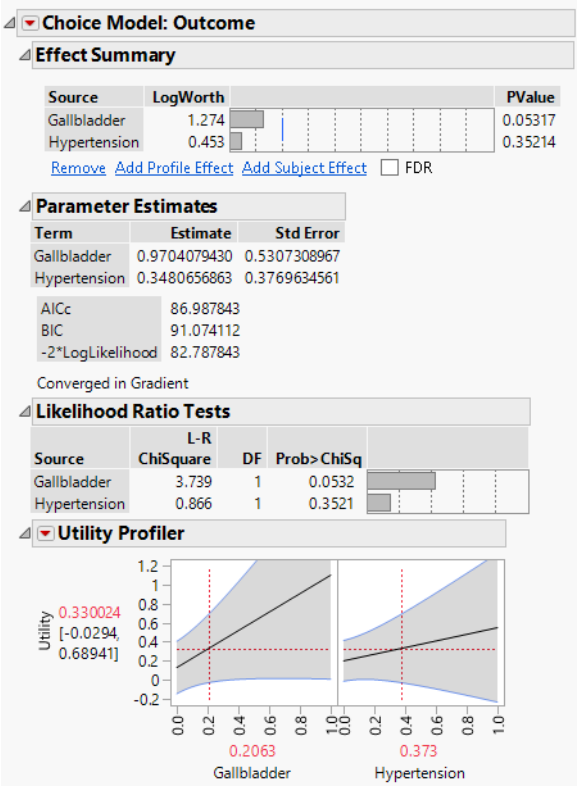
This section provides an example using the Choice platform to perform logistic regression on the results of a study of endometrial cancer with 63 matched pairs. The data are from the Los Angeles Study of the Endometrial Cancer Data reported in Breslow and Day (1980). The goal of the case-control analysis was to determine the relative risk for gallbladder disease, controlling for the effect of hypertension. The Outcome of 1 indicates the presence of endometrial cancer, and 0 indicates the control. Gallbladder and Hypertension data indicators are also 0 or 1.

For more information about performing logistic regression using the Choice platform, see "Logistic Regression".

1. Select **Help > Sample Data Library** and open Endometrial Cancer.jmp.
2. Select **Analyze > Consumer Research > Choice**.
3. Check that the Data Format selected is **One-Table, Stacked**.
4. Click the **Select Data Table** button.

5. Select Endometrial Cancer as the profile data table. Click **OK**.
6. Select Outcome and click **Response Indicator**.
7. Select Pair and click **Grouping**.
8. Select Gallbladder and Hypertension and click **Add** in the Construct Profile Effects window.
9. Deselect the **Firth Bias-Adjusted Estimates** check box.
10. Click **Run Model**.
11. Click the Choice Model: Outcome red triangle and select **Utility Profiler**.

Figure 4.42 Logistic Regression on Endometrial Cancer Data



Likelihood Ratio tests are given for each factor. Note that Gallbladder is nearly significant at the 0.05 level (p-value = 0.0532). Use the Utility Profiler to visualize the impact of the factors on the response.

Example of Transforming Data to Two Analysis Tables

Consider the data from Daganzo, found in Daganzo Trip.jmp. This data set contains the travel time for three transportation alternatives and the preferred transportation alternative for each subject.

Add Choice Mode and Subjects

1. Select **Help > Sample Data Library** and open the Daganzo Trip.jmp data table.

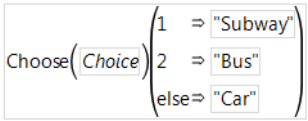
Figure 4.43 Partial Daganzo Trip Table

	Subway	Bus	Car	Choice
1	16.481	16.196	23.89	2
2	15.123	11.373	14.182	2
3	19.469	8.822	20.819	2
4	18.847	15.649	21.28	2
5	12.578	10.671	18.335	2

Each Choice number listed must first be converted to one of the travel mode names. This transformation is easily done by using the **Choose** function in the formula editor:

2. Select **Cols > New Columns**.
3. Specify the Column Name as Choice Mode and the modeling type as **Nominal**.
4. Click the **Column Properties** and select **Formula**.
5. Click **Conditional** in the functions list, select **Choose**, and press comma twice to obtain additional arguments for the function.
6. Click Choice for the Choose expression (expr), and double-click each clause entry box to enter “Subway”, “Bus”, and “Car” (with the quotation marks).

Figure 4.44 Choose Function for Choice Mode Column of Daganzo Data



7. Click **OK** in the Formula Editor window.
8. Click **OK** in the New Column window.

   The new Choice Mode column appears in the data table. Because each row contains a choice made by each subject, another column containing a sequence of numbers should be created to identify the subjects.
9. Select **Cols > New Columns**.

10. Specify the Column Name as Subject.
11. Click **Missing/Empty** next to Initialize Data and select **Sequence Data**.
12. Click **OK**.

Figure 4.45 Partial Daganzo Trip Data with New Choice Mode and Subject Columns

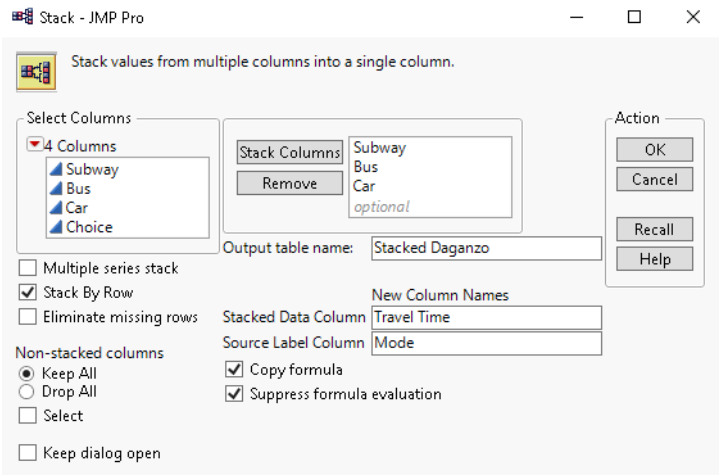
	Subway	Bus	Car	Choice	Choice Mode	Subject
1	16.481	16.196	23.89	2	Bus	1
2	15.123	11.373	14.182	2	Bus	2
3	19.469	8.822	20.819	2	Bus	3
4	18.847	15.649	21.28	2	Bus	4
5	12.578	10.671	18.335	2	Bus	5
6	11.513	20.582	27.838	1	Subway	6
7	10.651	15.537	17.418	1	Subway	7

Stack the Data

In order to construct the Profile data, each alternative needs to be expressed in a separate row.

1. Select **Tables > Stack**.
2. Select Subway, Bus, and Car and click **Stack Columns**.
3. For the Output table name, type Stacked Daganzo. Type Travel Time for the Stacked Data Column and Mode for the Source Label Column.

Figure 4.46 Stack Dialog for Daganzo Data



4. Click **OK**.

Figure 4.47 Partial Stacked Daganzo Table

	Choice	Choice Mode	Subject	Mode	Travel Time
1	2	Bus	1	Subway	16.481
2	2	Bus	1	Bus	16.196
3	2	Bus	1	Car	23.89
4	2	Bus	2	Subway	15.123
5	2	Bus	2	Bus	11.373
6	2	Bus	2	Car	14.182
7	2	Bus	3	Subway	19.469

Make the Profile Data Table

For the Profile Data Table, you need the Subject, Mode, and Travel Time columns.

1. Select the Subject, Mode, and Travel Time columns and select **Tables** > **Subset**.
2. Select **All Rows** and **Selected Columns** and click **OK**.

   A partial data table is shown in Figure 4.48. Note the default table name is Subset of Stacked Daganzo.

Figure 4.48 Partial Subset Table of Stacked Daganzo Data

	Subject	Mode	Travel Time
1	1	Subway	16.481
2	1	Bus	16.196
3	1	Car	23.89
4	2	Subway	15.123
5	2	Bus	11.373
6	2	Car	14.182
7	3	Subway	19.469

Make the Response Data Table

For the Response Data Table, you need the Subject and Choice Mode columns, but you also need a column for each possible choice.

3. From the Daganzo Trip.jmp data, select the Subject and Choice Mode columns.
4. Select **Tables > Subset**.
5. Select **All Rows** and **Selected Columns** and click **OK**.

   Note that the default table name is Subset of Daganzo Trip.
6. Select **Cols > New Columns**.
7. For the Column prefix, type Choice.
8. Select **Data Type>Character**.

9. Enter 3 for the Number of columns to add.
10. Click **OK**.

    The columns Choice 1, Choice 2, and Choice 3 have been added.
11. Type "Bus" (without quotation marks) in the first row of Choice 1. Right-click the cell and select **Fill > Fill to end of table**.
12. Type "Subway" (without quotation marks) in the first row of Choice 2. Right-click the cell and select **Fill > Fill to end of table**.
13. Type "Car" (without quotation marks) in the first row of Choice 3. Right-click the cell and select **Fill > Fill to end of table**.

Figure 4.49 Partial Subset Table of Daganzo Data with Choice Set

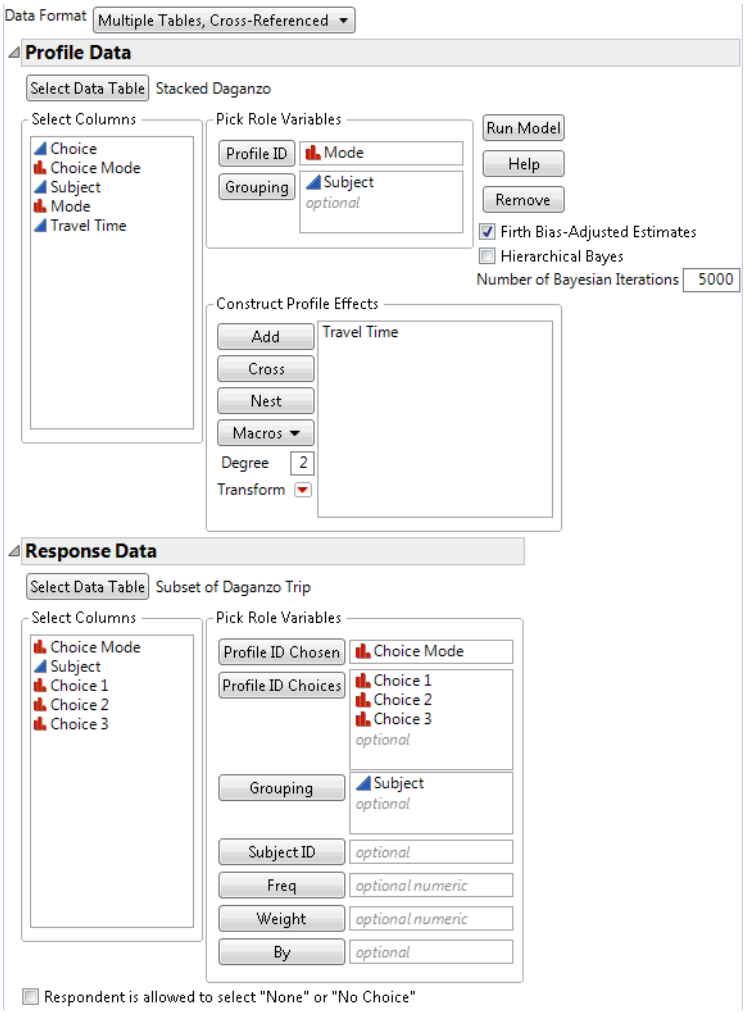
	Choice Mode	Subject	Choice 1	Choice 2	Choice 3
1	Bus	1	Bus	Subway	Car
2	Bus	2	Bus	Subway	Car
3	Bus	3	Bus	Subway	Car
4	Bus	4	Bus	Subway	Car
5	Bus	5	Bus	Subway	Car
6	Subway	6	Bus	Subway	Car
7	Subway	7	Bus	Subway	Car

Fit the Model

Now that you have separated the original Daganzo Trip.jmp table into two separate tables, you can run the Choice Platform.

1. Select **Analyze > Consumer Research > Choice**.
2. From the Data Format list, select **Multiple Tables, Cross-Referenced**.
3. Specify the model, as shown in Figure 4.50.

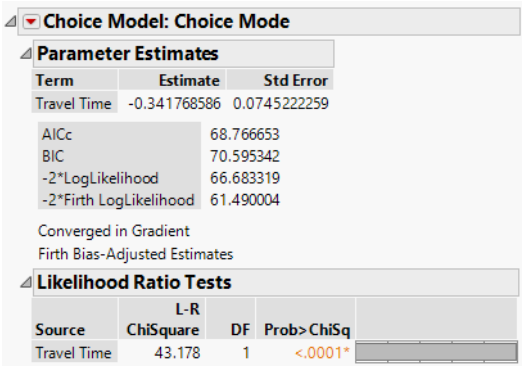
Figure 4.50 Choice Dialog Box for Daganzo Data with Multiple Tables



4. Click **Run Model**.

   The resulting parameter estimate now expresses the utility coefficient for Travel Time.

Figure 4.51 Parameter Estimate for Travel Time of Daganzo Data



The negative coefficient implies that increased travel time has a negative effect on consumer utility or satisfaction. The likelihood ratio test result indicates that the Choice model with the effect of Travel Time is significant.

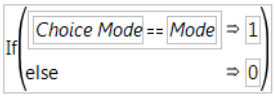
Example of Transforming Data to One Analysis Table

Rather than creating two or three tables, it can be more practical to transform the data so that only one table is used. For the one-table format, the subject effect is added as in the previous example. A response indicator column is added instead of using three different columns for the choice sets (Choice 1, Choice 2, Choice 3). Transform data for use with the one-table scenario:

1. Create or open Stacked Daganzo.jmp from the "Stack the Data" steps shown in "Example of Transforming Data to Two Analysis Tables".
2. Select **Cols > New Columns**.
3. Type Response as the Column Name.
4. Click **Column Properties** and select **Formula**.
5. Select **Conditional** in the functions list and then select **If**.
6. Select the column Choice Mode for the expression (expr).
7. Enter "=" and select Mode.
8. Type 1 for the **Then Clause** and 0 for the **Else Clause**.
9. Click **OK** in the Formula Editor window. Click **OK** in the New Column window.

   The completed formula should look like Figure 4.52.

Figure 4.52 Formula for Response Indicator for Stacked Daganzo Data



10. Select the Subject, Travel Time, and Response columns and then select **Tables** > **Subset**.
11. Select **All Rows** and **Selected Columns** and click **OK**.

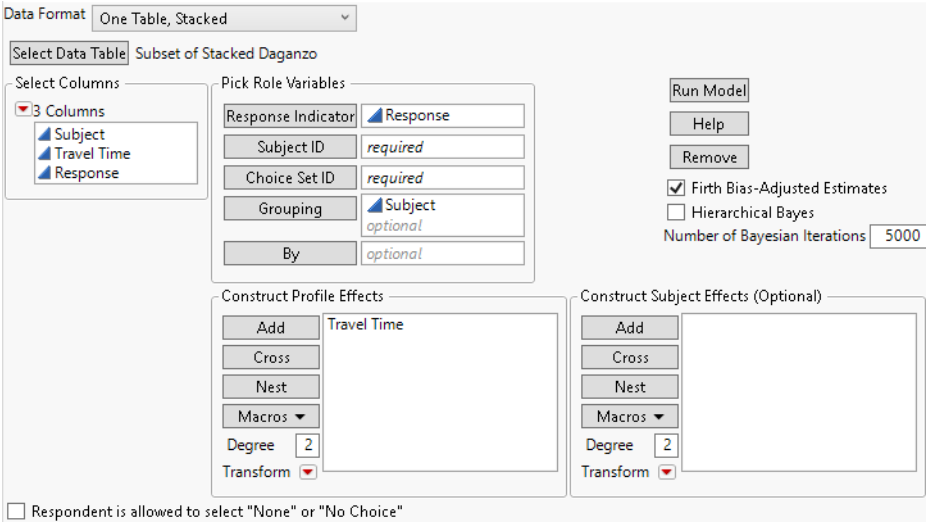
    A partial listing of the new data table is shown in Figure 4.53.

Figure 4.53 Partial Table of Stacked Daganzo Data Subset

	Subject	Travel Time	Response
1	1	16.481	0
2	1	16.196	1
3	1	23.89	0
4	2	15.123	0
5	2	11.373	1
6	2	14.182	0
7	3	19.469	0

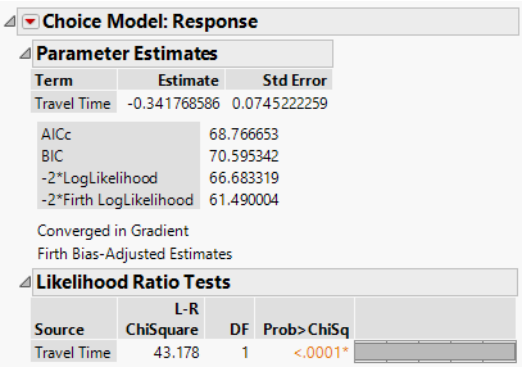
12. Select **Analyze** > **Consumer Research** > **Choice** to open the launch window and specify the model as shown in Figure 4.54.

Figure 4.54 Choice Dialog Box for Subset of Stacked Daganzo Data for One-Table Analysis



13. Click **Run Model.**

Figure 4.55 Parameter Estimate for Travel Time of Daganzo Data from One-Table Analysis



Notice that the result is identical to that obtained for the two-table model, shown earlier in Figure 4.51.

This chapter illustrates the use of the Choice Modeling platform with simple examples. This platform can also be used for more complex models, such as those involving more complicated transformations and interaction terms.

Statistical Details for the Choice Platform

- "Special Data Table Rules"
- "Utility and Probabilities"
- "Gradients"

Special Data Table Rules

Default Choice Set

If in every trial, you can choose any of the response profiles, you can omit the **Profile ID Choices** selection under **Pick Role Variables** in the Response Data section of the Choice launch window. The Choice Model platform then assumes that all choice profiles are available on each run.

Subject Data with Response Data

If you have subject data in the Response data table, select this table as the **Select Data Table** under the Subject Data. In this case, a **Subject ID** column does not need to be specified. In fact, it is not used. It is generally assumed that the subject data repeats consistently in multiple runs for each subject.

Logistic Regression

Ordinary logistic regression can be performed with the Choice Modeling platform.

Note: The Fit Y by X and Fit Model platforms are more convenient to use than the Choice Modeling platform for logistic regression modeling. This section is used only to demonstrate that the Choice Modeling platform can be used for logistic regression, if desired.

If your data are already in the choice-model format, you might want to use the steps given below for logistic regression analysis. However, three steps are needed:

- Create a trivial Profile data table with a row for each response level.
- Put the explanatory variables into the Response data.
- Specify the Response data table, again, for the Subject data table.

For examples of conducting Logistic Regression using the Choice Platform, see "Example of Logistic Regression Using the Choice Platform" and "Example of Logistic Regression for Matched Case-Control Studies".

Utility and Probabilities

Parameter estimates from the choice model identify consumer *utility*, or marginal utilities in the case of a linear utility function. Utility is the level of satisfaction consumers receive from products with specific attributes and is determined from the parameter estimates in the model.

The choice statistical model is expressed as follows:

Let $X[k]$ represent a subject attribute design row, with intercept

Let $Z[j]$ represent a choice attribute design row, without intercept

Then, the probability of a given choice for the $k'th$ subject to the $j'th$ choice of m choices is:

$$P_i[jk] = \frac{\exp(\beta'(X[k] \otimes Z[j]))}{\sum_{l=1}^{m} \exp(\beta'(X[k] \otimes Z[l]))}$$

where:

- $\otimes$ is the Kronecker rowwise product
- the numerator calculates for the $j'th$ alternative actually chosen
- the denominator sums over the m choices presented to the subject for that trial

Gradients

The gradient values that you obtain when you select the Save Gradients by Subject option are the subject-aggregated Newton-Raphson steps from the optimization used to produce the estimates. At the estimates, the total gradient is zero, and $\Delta = H^{-1}g = 0$, where g is the total gradient of the log-likelihood evaluated at the MLE, and H^{-1} is the inverse Hessian function or the inverse of the negative of the second partial derivative of the log-likelihood.

But, the disaggregation of Δ results in the following:

$$\Delta = \Sigma_{ij}\Delta_{ij} = \Sigma H^{-1}g_{ij} = 0,$$

Here i is the subject index, j is the choice response index for each subject, Δ_{ij} are the partial Newton-Raphson steps for each run, and g_{ij} is the gradient of the log-likelihood by run.

The mean gradient step for each subject is then calculated as follows:

$$\bar{\Delta}_i = \Sigma_j \frac{\Delta_{ij}}{n_i} ,$$

where n_i is the number of runs per subject. The $\bar{\Delta}_i$ are related to the force that subject i is applying to the parameters. If groups of subjects have truly different preference structures, these forces are strong, and they can be used to cluster the subjects. The $\bar{\Delta}_i$ are the gradient forces that are saved. You can then cluster these values using the Clustering platform.

Chapter 5

MaxDiff

Fit Models for MaxDiff Experiments

Use MaxDiff (maximum difference scaling) as an alternative to standard preference scales to determine the relative importance of items being rated. MaxDiff forces respondents to report their most and least preferred options. This often results in rankings that are more definitive than rankings obtained using standard preference scales.

The MaxDiff platform enables you to do the following:

- Use information about respondent (subject) traits as well as product attributes.
- Integrate data from one, two, or three sources.
- Obtain subject-level scores for segmenting or clustering your data.
- JMP PRO Estimate subject-specific coefficients using a Bayesian approach.
- Use bias-corrected maximum likelihood estimators (Firth 1993).

Figure 5.1 MaxDiff All Comparisons Report

All Levels Comparison Report

Difference (Row-Column) Standard Error of Difference Wald p-Value	All Dressed	Barbecue	Biscuits and Gravy	Dill Pickle	Gyro	Ketchup	Reuben	Sour Cream and Onion	Southern Barbecue	Truffle Fries
All Dressed	0	-1.4951	-0.3222	-0.0804	0.94665	0.12371	0.24744	-0.228	-1.0068	-0.3615
	0	0.36838	0.29974	0.29447	0.33571	0.29728	0.30148	0.28158	0.3418	0.31136
		9.22e-5	0.28477	0.78538	0.00569	0.67812	0.41355	0.41975	0.00393	0.24809
Barbecue	1.49507	0	1.1729	1.41468	2.44172	1.61877	1.7425	1.26703	0.48824	1.13355
	0.36838	0	0.36267	0.3539	0.41499	0.36911	0.37117	0.37654	0.39164	0.39213
	9.22e-5		0.00161	0.00012	4.32e-8	2.64e-5	7.69e-6	0.00105	0.21514	0.00462
Biscuits and Gravy	0.32217	-1.1729	0	0.24179	1.26882	0.44588	0.56961	0.09413	-0.6847	-0.0393
	0.29974	0.36267	0	0.28249	0.34456	0.29752	0.30639	0.30716	0.34038	0.31345
	0.28477	0.00161		0.39388	0.00036	0.1368	0.06567	0.75983	0.0467	0.90032
Dill Pickle	0.08038	-1.4147	-0.2418	0	1.02703	0.20409	0.32782	-0.1477	-0.9264	-0.2811
	0.29447	0.3539	0.28249	0	0.35251	0.27938	0.30328	0.30707	0.35532	0.30871
	0.78538	0.00012	0.39388		0.00432	0.46661	0.28208	0.63156	0.01038	0.36444
Gyro	-0.9467	-2.4417	-1.2688	-1.027	0	-0.8229	-0.6992	-1.1747	-1.9535	-1.3082
	0.33571	0.41499	0.34456	0.35251	0	0.32802	0.34758	0.33327	0.38939	0.35477
	0.00569	4.32e-8	0.00036	0.00432		0.01356	0.04668	0.00062	2.01e-6	0.00035
Ketchup	-0.1237	-1.6188	-0.4459	-0.2041	0.82294	0	0.12373	-0.3517	-1.1305	-0.4852
	0.29728	0.36911	0.29752	0.27938	0.32802	0	0.30893	0.31915	0.34808	0.31157
	0.67812	2.64e-5	0.1368	0.46661	0.01356		0.68956	0.27279	0.00154	0.12223
Reuben	-0.2474	-1.7425	-0.5696	-0.3278	0.69921	-0.1237	0	-0.4755	-1.2543	-0.609
	0.30148	0.37117	0.30639	0.30328	0.34758	0.30893	0	0.30932	0.35324	0.31817
	0.41355	7.69e-6	0.06567	0.28208	0.04668	0.68956		0.12709	0.00056	0.05821
Sour Cream and Onion	0.22804	-1.267	-0.0941	0.14766	1.17469	0.35175	0.47548	0	-0.7788	-0.1335
	0.28158	0.37654	0.30716	0.30707	0.33327	0.31915	0.30932	0	0.34896	0.31379
	0.41975	0.00105	0.75983	0.63156	0.00062	0.27279	0.12709		0.02764	0.67137
Southern Barbecue	1.00683	-0.4882	0.68465	0.92644	1.95348	1.13053	1.25426	0.77879	0	0.6453
	0.3418	0.39164	0.34038	0.35532	0.38939	0.34808	0.35324	0.34896	0	0.34904
	0.00393	0.21514	0.0467	0.01038	2.01e-6	0.00154	0.00056	0.02764		0.06715
Truffle Fries	0.36152	-1.1335	0.03935	0.28114	1.30817	0.48523	0.60896	0.13348	-0.6453	0
	0.31136	0.39213	0.31345	0.30871	0.35477	0.31157	0.31817	0.31379	0.34904	0
	0.24809	0.00462	0.90032	0.36444	0.00035	0.12223	0.05821	0.67137	0.06715	

Contents

Overview of the MaxDiff Modeling Platform 143
Examples of the MaxDiff Platform 143
One Table Format 144
Multiple Table Format 147
Launch the MaxDiff Platform 150
Launch Window for One Table, Stacked 151
Launch Window for Multiple Tables, Cross-Referenced 152
MaxDiff Model Report 156
Effect Summary 156
MaxDiff Results 157
Parameter Estimates 158
Bayesian Parameter Estimates 160
Likelihood Ratio Tests 161
MaxDiff Platform Options 162
Comparisons Report 164
Save Bayes Chain 165

Overview of the MaxDiff Modeling Platform

MaxDiff, also known as *best-worst scaling* (BWS), is a choice-based measurement method. Rather than asking a respondent to report one favorite choice among several alternative profiles, MaxDiff asks a respondent to report both a *best* and a *worst* choice. The MaxDiff approach can provide more information about preferences than an approach where a respondent reports only a favorite choice. For background on MaxDiff studies, see Louviere et al. (2015). For background on choice modeling, see Louviere et al. (2015), Train (2009), and Rossi et al. (2005).

MaxDiff analysis uses the framework of random utility theory. A choice is assumed to have an underlying value, or *utility*, to respondents. The MaxDiff platform estimates these utilities. The MaxDiff platform also estimates the probabilities that a choice is preferred over other choices. This is done using conditional logistic regression. See McFadden (1974).

Note: One-factor MaxDiff studies can be designed using the MaxDiff Design platform. See the *Design of Experiments Guide*.

Segmentation and Bayesian Subject-Level Effects

Market researchers sometimes want to analyze the preference structure for each subject separately in order to see whether there are groups of subjects that behave differently. If there are sufficient data, you can specify "By groups" in the Response Data or you could introduce a Subject identifier as a subject-side model term. This approach, however, is costly if the number of subjects is large. Other segmentation techniques discussed in the literature include Bayesian and mixture methods.

If there are not sufficient data to specify "By groups," you can segment in JMP by clustering subjects using response data and the **Save Gradients by Subject** option. The option creates a new data table containing the average Hessian-scaled gradient on each parameter for each subject. For an example, see "Example of Segmentation". For more information about the gradient values, see "Gradients".

JMP PRO MaxDiff also provides a Hierarchical Bayesian approach to estimating subject-level effects. This approach can be useful in market segmentation.

Examples of the MaxDiff Platform

Thirty respondents participated in a MaxDiff study to compare seven flavors of potato chips. Each choice set consisted of three profiles (potato chip flavors). For each choice set, a respondent's favorite choice was recorded as 1 and his or her least favorite choice was recorded as -1. Intermediate choices were recorded as 0.

The MaxDiff platform can analyze data that is presented in a one-table format or in a multiple-table format. In the multiple table format, information about responses, choice sets, and subjects is saved in different data tables. In the one-table format, that information is contained in a single data table.

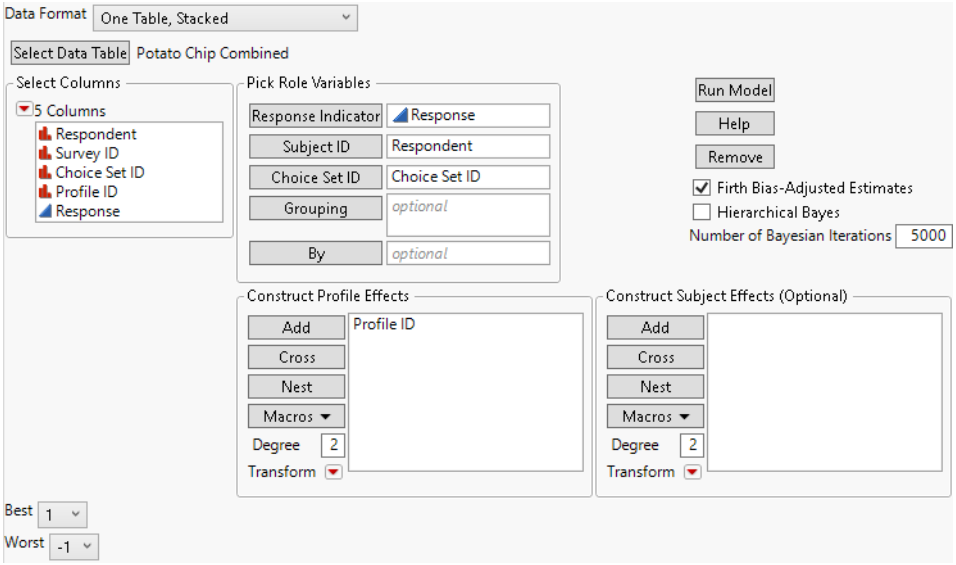
- "One Table Format" shows how to analyze a subset of the available data in a one-table format. Note that you could add additional profile and subject data to the single table for a more complete analysis.
- "Multiple Table Format" shows how to bring together information from different tables into one MaxDiff analysis.

One Table Format

1. Select **Help > Sample Data Library** and open Potato Chip Combined.jmp.
2. Select **Analyze > Consumer Research > MaxDiff**.

   Note that the default Data Format is set to One Table, Stacked.
3. Click **Select Data Table**.
4. Select Potato Chip Combined and click **OK**.
5. Assign roles to columns to complete launch dialog:
   - Select Response and click **Response Indicator**.
   - Select Respondent and click **Subject ID**.
   - Select Choice Set ID and click **Choice Set ID**.
   - Select ProfileID and click **Add** in the Construct Profile Effects panel.

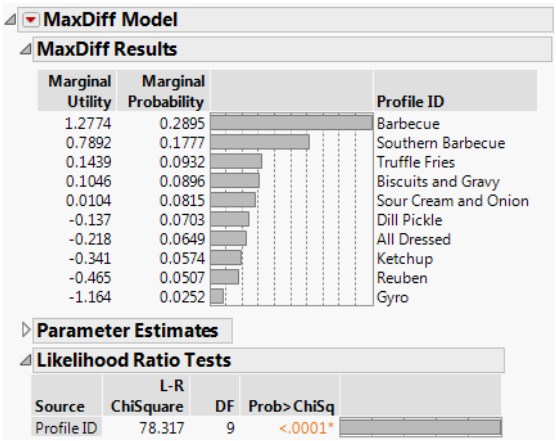
Figure 5.2 Completed MaxDiff Launch Window



Note that the setting for Worst choice changed to -1 when you specified the Response column as the Response Indicator variable.

6. Click **Run Model**.

Figure 5.3 MaxDiff Report for Potato Chip Combined.jmp



The report indicates that Profile ID is significant, indicating that preferences for the various chip types differ significantly. The highest Marginal Utility is for Barbecue chips. The estimated probability that Barbecue chips are preferred to other chip types is 0.2895.

7. Click the MaxDiff Model red triangle and select **All Levels Comparison Report.**

Figure 5.4 All Comparisons Report

All Levels Comparison Report

Difference (Row-Column) Standard Error of Difference Wald p-Value	All Dressed	Barbecue	Biscuits and Gravy	Dill Pickle	Gyro	Ketchup	Reuben	Sour Cream and Onion	Southern Barbecue	Truffle Fries
All Dressed	0	-1.4951	-0.3222	-0.0804	0.94665	0.12371	0.24744	-0.228	-1.0068	-0.3615
	0	0.36838	0.29974	0.29447	0.33571	0.29728	0.30148	0.28158	0.3418	0.31136
		9.22e-5	0.28477	0.78538	0.00569	0.67812	0.41355	0.41975	0.00393	0.24809
Barbecue	1.49507	0	1.1729	1.41468	2.44172	1.61877	1.7425	1.26703	0.48824	1.13355
	0.36838	0	0.36267	0.3539	0.41499	0.36911	0.37117	0.37654	0.39164	0.39213
	9.22e-5		0.00161	0.00012	4.32e-8	2.64e-5	7.69e-6	0.00105	0.21514	0.00462
Biscuits and Gravy	0.32217	-1.1729	0	0.24179	1.26882	0.44588	0.56961	0.09413	-0.6847	-0.0393
	0.29974	0.36267	0	0.28249	0.34456	0.29752	0.30639	0.30716	0.34038	0.31345
	0.28477	0.00161		0.39388	0.00036	0.1368	0.06567	0.75983	0.0467	0.90032
Dill Pickle	0.08038	-1.4147	-0.2418	0	1.02703	0.20409	0.32782	-0.1477	-0.9264	-0.2811
	0.29447	0.3539	0.28249	0	0.35251	0.27938	0.30328	0.30707	0.35532	0.30871
	0.78538	0.00012	0.39388		0.00432	0.46661	0.28208	0.63156	0.01038	0.36444
Gyro	-0.9467	-2.4417	-1.2688	-1.027	0	-0.8229	-0.6992	-1.1747	-1.9535	-1.3082
	0.33571	0.41499	0.34456	0.35251	0	0.32802	0.34758	0.33327	0.38939	0.35477
	0.00569	4.32e-8	0.00036	0.00432		0.01356	0.04668	0.00062	2.01e-6	0.00035
Ketchup	-0.1237	-1.6188	-0.4459	-0.2041	0.82294	0	0.12373	-0.3517	-1.1305	-0.4852
	0.29728	0.36911	0.29752	0.27938	0.32802	0	0.30893	0.31915	0.34808	0.31157
	0.67812	2.64e-5	0.1368	0.46661	0.01356		0.68956	0.27279	0.00154	0.12223
Reuben	-0.2474	-1.7425	-0.5696	-0.3278	0.69921	-0.1237	0	-0.4755	-1.2543	-0.609
	0.30148	0.37117	0.30639	0.30328	0.34758	0.30893	0	0.30932	0.35324	0.31817
	0.41355	7.69e-6	0.06567	0.28208	0.04668	0.68956		0.12709	0.00056	0.05821
Sour Cream and Onion	0.22804	-1.267	-0.0941	0.14766	1.17469	0.35175	0.47548	0	-0.7788	-0.1335
	0.28158	0.37654	0.30716	0.30707	0.33327	0.31915	0.30932	0	0.34896	0.31379
	0.41975	0.00105	0.75983	0.63156	0.00062	0.27279	0.12709		0.02764	0.67137
Southern Barbecue	1.00683	-0.4882	0.68465	0.92644	1.95348	1.13053	1.25426	0.77879	0	0.6453
	0.3418	0.39164	0.34038	0.35532	0.38939	0.34808	0.35324	0.34896	0	0.34904
	0.00393	0.21514	0.0467	0.01038	2.01e-6	0.00154	0.00056	0.02764		0.06715
Truffle Fries	0.36152	-1.1335	0.03935	0.28114	1.30817	0.48523	0.60896	0.13348	-0.6453	0
	0.31136	0.39213	0.31345	0.30871	0.35477	0.31157	0.31817	0.31379	0.34904	0
	0.24809	0.00462	0.90032	0.36444	0.00035	0.12223	0.05821	0.67137	0.06715	

Each comparison is the difference in estimated utilities between the chip type labeling the row and the chip type labeling the column. Small *p*-values are colored with an intense blue or red color, depending on the sign of the difference. For example, based on the blue colors across the Gyro row, you can see that Gyro chips have significantly lower utility than all other chip types. Barbecue chips have higher utility than all other chip types, though they do not differ significantly from Southern Barbecue chips.

Note: Because the All Comparisons Report *p*-values are not corrected for multiple comparisons, use them as a guide.

Multiple Table Format

This version of the potato chip study uses three data tables: Potato Chip Profiles.jmp, Potato Chip Responses.jmp, and Potato Chip Subjects.jmp. Although you can always arrange your data into a single table, a multi-table approach can be more convenient than a one-table analysis when you have additional profile and subject variables that you want to include in your analysis.

Complete the Launch Window

1. Select **Help > Sample Data Library** and open the Potato Chip Responses.jmp sample data table.

Note: If you prefer not to follow the steps for completing the launch window, click the green triangle next to the **MaxDiff for Flavor** script. Then proceed to "Explore the Model".

2. Click the green triangle next to the **Open Profile and Subject Tables** script.
   - The profile data table, Potato Chip Profiles.jmp, lists all the potato chip types in the study (Flavor) along with information about the country of origin (Product Of). Each choice has a Profile ID.
   - The subjects data table, Potato Chip Subjects.jmp, lists the respondents. It also gives additional information about each respondent: Citizenship and Gender.
   - The responses data table, Potato Chip Responses.jmp, lists the respondents. For each respondent, the Survey ID and Choice Set ID for each set of profiles is listed, along with the Profile ID values for each choice set. The table also contains response data in the Best Profile and Worst Profile columns.
3. From any of the three data tables, select **Analyze > Consumer Research > MaxDiff**.
4. From the Data Format list, select Multiple Tables, Cross-Referenced.

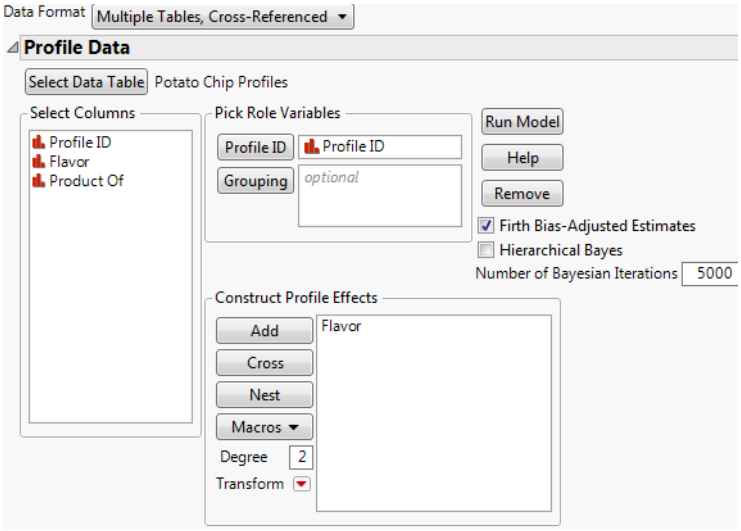
   There are three separate outlines, one for each of the data sources.
5. Click **Select Data Table** under Profile Data.

   A Profile Data Table window appears, which prompts you to specify the data table for the profile data.
6. Select Potato Chip Profiles and click **OK**.

   The columns from this table appear in the **Select Columns**.
7. Select Profile ID from the Select Columns list and click **Profile ID** under **Pick Role Variables**.
8. Select Flavor and click **Add** under **Construct Model Effects.**

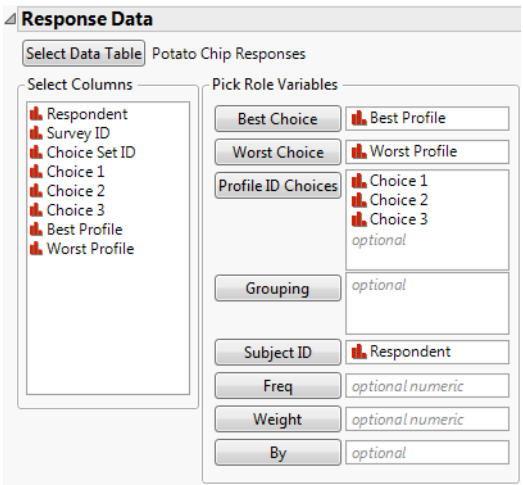
   Note that Product Of is another profile effect that you could add to the effects list.

Figure 5.5 Complete Profile Data Outline



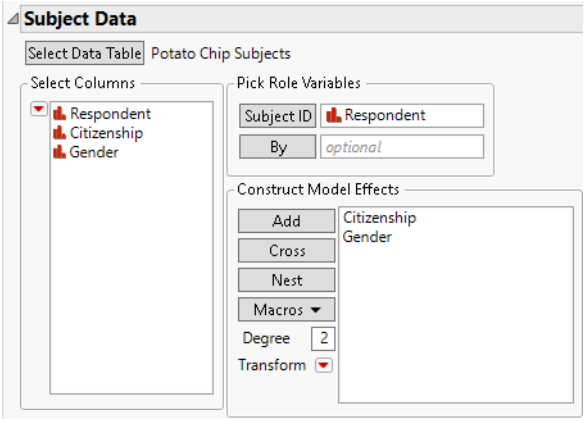
9. Open the Response Data outline. Click **Select Data Table**.
10. Select Potato Chip Responses and click **OK**.
11. Assign roles to columns to complete the launch dialog:
    - Select Best Profile and click **Best Choice**.
    - Select Worst Profile and click **Worst Choice**.
    - Select Choice 1, Choice 2, and Choice 3 and click **Profile ID Choices**.
    - Select Respondent and click **Subject ID**.

Figure 5.6 Completed Response Data Outline



12. Open the Subject Data outline. Click **Select Data Table**.
13. Select Potato Chip Subjects and click **OK**.
14. Select Respondent and click **Subject ID**.
15. Select Citizenship and Gender and click **Add** under **Construct Model Effects**.

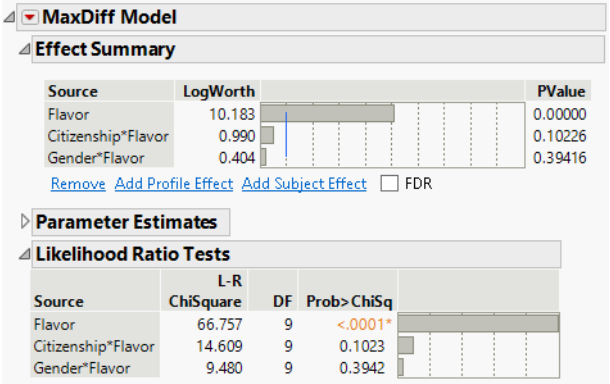
Figure 5.7 Completed Subject Data Outline



Explore the Model

1. Click **Run Model**.

Figure 5.8 MaxDiff Model Report



The Effect Summary report shows the terms in the model and gives *p*-values for their significance. Notice that Flavor is a profile effect, and that each of Citizenship*Flavor and Gender*Flavor is an interaction of a subject and a profile effect.

The Likelihood Ratio Tests report indicates that Flavor is significant.

Launch the MaxDiff Platform

Launch the MaxDiff platform by selecting **Analyze > Consumer Research > MaxDiff**.

Your data for the MaxDiff platform can be combined in a single data table or it can reside in two or three separate data tables. In the MaxDiff launch window, specify whether you are using one or multiple data tables in the Data Format list.

One Table, Stacked

For the One Table, Stacked format, the data are in a single data table. There is a row for every profile presented to a subject within a choice set and an indicator for the best and worst profiles in that choice set. The Potato Chip Combined.jmp sample data table contains the results of a MaxDiff experiment in a single table format. See "One Table Format".

For more information about the launch window for this format, see "Launch Window for One Table, Stacked".

Multiple Tables, Cross-Referenced

For the Multiple Tables, Cross-Referenced format, the data are in two or three separate data tables. A profile data table and a response data table are required. A subject data table is optional. Note the following:

- The profile data table must contain a column with a unique identifier for each profile and columns for the profile level variables. The profile identifier is used in the response data table to identify best and worst profile responses for each choice set.
- The optional subject data table must contain a column with a unique subject identifier for each subject and columns for the subject level variables. The subject identifier is used in the response data table to identify the subjects.

The launch window for this format contains three sections: Profile Data, Response Data, and Subject Data. Each section corresponds to a different data table. You can expand or collapse each section as needed.

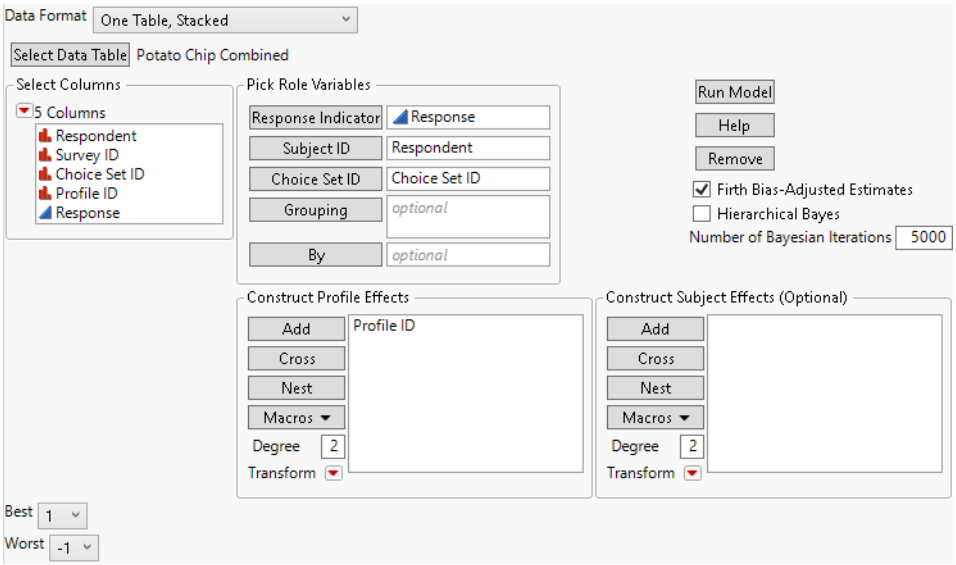
The Potato Chip Profiles.jmp, Potato Chip Responses.jmp, and Potato Chip Subjects.jmp sample data tables contain the results of a MaxDiff experiment using three tables. See "Multiple Table Format".

For more information about the launch window for this format, see "Launch Window for Multiple Tables, Cross-Referenced".

Launch Window for One Table, Stacked

Launch the MaxDiff platform by selecting **Analyze > Consumer Research > MaxDiff**. For one table select **One Table, Stacked** from the Data Format menu.

Figure 5.9 Launch Window for One Table, Stacked Data Format



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Select Data Table Select or open the data table that contains the combined data. Select Other to open a file that is not already open.

Response Indicator A column containing the preference data. Use two of the values 1, -1, or 0 for the Best and Worst choices, and the third value for profiles that are not Best or Worst. The default coding is a 1 to indicate the Best choice and a -1 for the Worst choice.

Subject ID An identifier for the study participant.

Choice Set ID An identifier for the set of profiles presented to the subject for a given preference determination.

Grouping A column which, when used with the Choice Set ID, uniquely designates each choice set. For example, if a choice set has Choice Set ID = 1 for Survey = A, and another choice set has Choice Set ID = 1 for Survey = B, then Survey should be used as a **Grouping** column.

By Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of By variables.

Construct Profile Effects Add effects constructed from the attributes for the profiles.

For information about the Construct Profile Effects panel, see *Fitting Linear Models*.

Construct Subject Effects (Optional) Add effects constructed from subject-related factors.

For information about the Construct Subject Effects panel, see *Fitting Linear Models*.

Firth Bias-adjusted Estimates Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. See Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

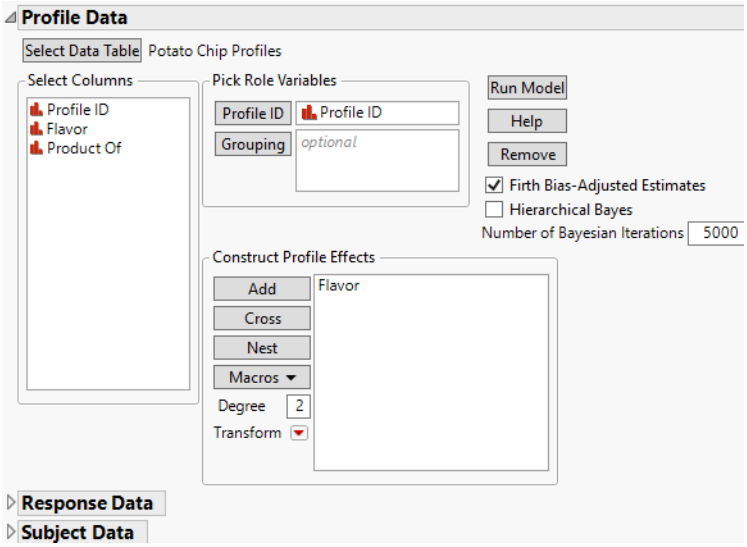
JMP PRO **Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See "Bayesian Parameter Estimates".

JMP PRO **Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject effects. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

Launch Window for Multiple Tables, Cross-Referenced

Launch the MaxDiff platform by selecting **Analyze > Consumer Research > MaxDiff**. For multiple tables select **Multiple Tables, Cross-Referenced** form the Data Format menu.

Figure 5.10 Launch Window for Multiple Tables, Cross-Referenced Data Format



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

In the case of Multiple Tables, Cross-Referenced, the launch window has three sections:

- "Profile Data"
- "Response Data"
- "Subject Data"

Profile Data

The profile data table describes the attributes associated with each choice. Each choice can comprise many different attributes, and each attribute is listed as a column in the data table. There is a row for each possible choice, and each possible choice contains a unique ID.

Select Data Table Select or open the data table that contains the profile data. Select Other to open a file that is not already open.

Profile ID Identifier for each row of choice combinations. If the **Profile ID** column does not uniquely identify each row in the profile data table, you need to add **Grouping** columns. Add **Grouping** columns until the combination of **Grouping** and **Profile ID** columns uniquely identifies the row, or profile.

Grouping A column which, when used with the Choice Set ID column, uniquely designates each choice set. For example, if Profile ID = 1 for Survey = A, and a different Profile ID = 1 for Survey = B, then Survey would be used as a **Grouping** column.

Construct Profile Effects Add effects constructed from the attributes in the profiles.

For information about the Construct Profile Effects panel, see *Fitting Linear Models*.

Firth Bias-adjusted Estimates Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. See Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

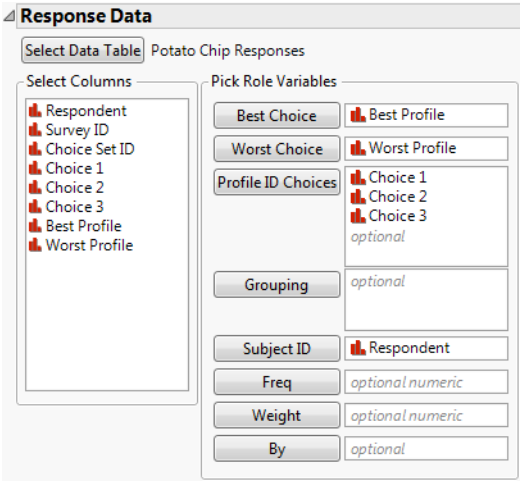
JMP PRO **Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See "Bayesian Parameter Estimates".

JMP PRO **Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject effects. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

Response Data

Figure 5.11 shows the Response Data outline populated using Potato Chip Responses.jmp.

Figure 5.11 Response Data Outline



The response data table contains the study results. It gives the choice set IDs for each trial as well as the profiles selected as best and worst by the subject. The Response data are linked to the Profile data through the choice set columns and the choice response column. Grouping variables can be used to align choice indices when more than one group is contained within the data.

Select Data Table Select or open the data table that contains the response data. Select Other to open a file that is not already open.

Best Choice The Response table column containing the Profile ID of the profile that the study participant designated as Best.

Worst Choice The Response table column containing the Profile ID of the profile that the study participant designated as Worst.

Profile ID Choices The columns that contain the **Profile ID**s of the set of possible choices for each choice set. There must be at least three profiles.

Grouping A column which, when used with the Profile ID Chosen column, uniquely designates each choice set.

Subject ID A unique identifier for the study participant.

Freq A column containing frequencies. If n is the value of the Freq variable for a given row, then that row is used in computations n times. If it is less than 1 or missing, then JMP does not use it to calculate any analyses.

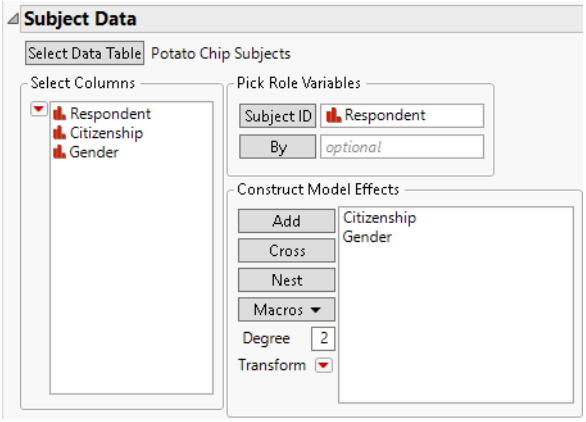
Weight A column containing a weight for each observation in the data table. The weight is included in analyses only when its value is greater than zero.

By Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of By variables.

Subject Data

Figure 5.12 shows the Subject Data outline populated using Potato Chip Subjects.jmp.

Figure 5.12 Subject Data Outline



Note: A subject data table is optional, depending on whether subject effects are to be modeled.

The subject data table contains the Subject ID and one or more columns of attributes or characteristics for each subject. The subject data table contains the same number of rows as subjects and has an identifier column that matches a similar column in the Response data table.

Note: You can include subject data in the response data table, but you need to specify subject effects in the Subject Data outline.

Select Data Table Select or open the data table that contains the subject data. Select Other to open a file that is not already open.

Subject ID Unique identifier for the subject.

By Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of By variables.

Construct Model Effects Add effects constructed from columns in the subject data table.

For information about the Construct Model Effects panel, see *Fitting Linear Models*.

MaxDiff Model Report

The MaxDiff Model window shows some of the following reports by default, depending on your selections in the launch window.

- "Effect Summary"
- "MaxDiff Results"
- "Parameter Estimates"
- "Bayesian Parameter Estimates"
- "Likelihood Ratio Tests"

For descriptions of the platform options, see "MaxDiff Platform Options".

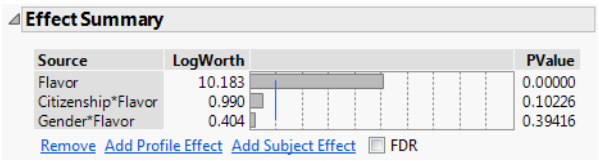
Effect Summary

The Effect Summary report appears if your model contains more than one effect. It lists the effects estimated by the model and gives a plot of the LogWorth (or FDR LogWorth) values for these effects. The report also provides controls that enable you to add or remove effects from the model. The model fit report updates automatically based on the changes made in the Effects Summary report. See *Fitting Linear Models*.

Note: The Effect Summary report is not applicable to models fit with Hierarchical Bayes.

Figure 5.13 shows the Effect Summary report obtained by running the script **MaxDiff for Flavor** in Potato Chip Responses.jmp.

Figure 5.13 Effect Summary Report



Effect Summary Table Columns

The Effect Summary table contains the following columns:

Source Lists the model effects, sorted by ascending *p*-values.

LogWorth Shows the LogWorth for each model effect, defined as $-\log_{10}(p\text{-value})$. This transformation adjusts p-values to provide an appropriate scale for graphing. A value that exceeds 2 is significant at the 0.01 level (because $-\log_{10}(0.01) = 2$).

FDR LogWorth Shows the False Discovery Rate LogWorth for each model effect, defined as $-\log_{10}(\text{FDR PValue})$. This is the best statistic for plotting and assessing significance. Select the **FDR** check box to replace the LogWorth column with the **FDR LogWorth** column.

Bar Chart Shows a bar chart of the LogWorth (or FDR LogWorth) values. The graph has dashed vertical lines at integer values and a blue reference line at 2.

PValue Shows the p-value for each model effect. This is the p-value corresponding to the significance test displayed in the Likelihood Ratio Tests report.

FDR PValue Shows the False Discovery Rate p-value for each model effect calculated using the Benjamini-Hochberg technique. This technique adjusts the p-values to control the false discovery rate for multiple tests. Select the **FDR** check box to replace the **PValue** column with the **FDR PValue** column.

For more information about the FDR correction, see Benjamini and Hochberg (1995). For more information about the false discovery rate, see *Predictive and Specialized Modeling* or Westfall et al. (2011).

Effect Summary Table Options

The options below the summary table enable you to add and remove effects:

Remove Removes the selected effects from the model. To remove one or more effects, select the rows corresponding to the effects and click the Remove button.

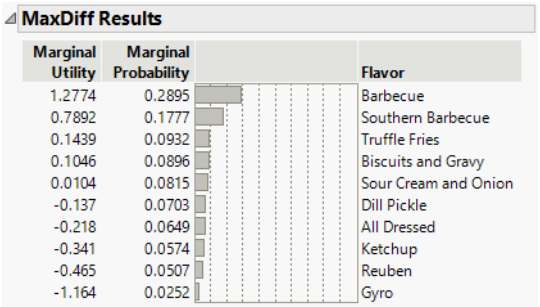
Add Profile Effect Opens a column dialog that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Profile Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

Add Subject Effect Opens a column dialog that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Subject Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

MaxDiff Results

Figure 5.14 shows the MaxDiff Results report obtained by running the script **MaxDiff with No Subject Effects** in Potato Chip Responses.jmp.

Figure 5.14 MaxDiff Results Report



MaxDiff Results

Marginal Utility	Marginal Probability	Flavor
1.2774	0.2895	Barbecue
0.7892	0.1777	Southern Barbecue
0.1439	0.0932	Truffle Fries
0.1046	0.0896	Biscuits and Gravy
0.0104	0.0815	Sour Cream and Onion
-0.137	0.0703	Dill Pickle
-0.218	0.0649	All Dressed
-0.341	0.0574	Ketchup
-0.465	0.0507	Reuben
-1.164	0.0252	Gyro

For each Profile effect specified in the launch window, the following are displayed:

Marginal Utility An indicator of the perceived value of the corresponding level of the effect. Larger values suggest that the feature is of greater value.

Marginal Probability The estimated probability that a subject expresses a preference for the corresponding level of the effect over all other levels. For each effect, the marginal probabilities sum to one.

Bar Chart Shows a bar chart of the marginal probabilities.

Effect Column Gives the name of the effect and a list of its levels. The levels define the features to which the Marginal Utility and Marginal Probability estimates apply.

Parameter Estimates

This report gives details about parameter estimates, fit criteria, and the fitting algorithm.

Figure 5.15 shows the Parameter Estimates report obtained by running the script **MaxDiff for Flavor** in Potato Chip Responses.jmp.

Figure 5.15 Parameter Estimates Report

Parameter Estimates

Term	Estimate	Std Error
Flavor[All Dressed]	-0.15616234	0.2227718641
Flavor[Barbecue]	1.22210326	0.2978972284
Flavor[Biscuits and Gravy]	0.16398758	0.2249076601
Flavor[Dill Pickle]	-0.17356828	0.2151668801
Flavor[Gyro]	-1.11927509	0.2827016652
Flavor[Ketchup]	-0.47308393	0.2319900823
Flavor[Reuben]	-0.50927309	0.2294979936
Flavor[Sour Cream and Onion]	0.21115573	0.2450361368
Flavor[Southern Barbecue]	0.70149945	0.2695322771
Citizenship[Canadian]*Flavor[All Dressed]	-0.04368106	0.2239442590
Citizenship[Canadian]*Flavor[Barbecue]	-0.16180196	0.2978036763
Citizenship[Canadian]*Flavor[Biscuits and Gravy]	0.05734172	0.2233312021
Citizenship[Canadian]*Flavor[Dill Pickle]	-0.09824391	0.2188896544
Citizenship[Canadian]*Flavor[Gyro]	0.43257276	0.2907051874
Citizenship[Canadian]*Flavor[Ketchup]	-0.38035261	0.2349720398
Citizenship[Canadian]*Flavor[Reuben]	-0.34677939	0.2342643217
Citizenship[Canadian]*Flavor[Sour Cream and Onion]	0.56678250	0.2355493683
Citizenship[Canadian]*Flavor[Southern Barbecue]	-0.00712532	0.2720518372
Gender[Female]*Flavor[All Dressed]	-0.26955535	0.2106170508
Gender[Female]*Flavor[Barbecue]	0.39881410	0.2961368976
Gender[Female]*Flavor[Biscuits and Gravy]	0.09184462	0.2245258406
Gender[Female]*Flavor[Dill Pickle]	-0.11461679	0.2131955199
Gender[Female]*Flavor[Gyro]	-0.40538827	0.2950786212
Gender[Female]*Flavor[Ketchup]	-0.07317652	0.2135446256
Gender[Female]*Flavor[Reuben]	0.20780054	0.2227994637
Gender[Female]*Flavor[Sour Cream and Onion]	0.27353428	0.2337547848
Gender[Female]*Flavor[Southern Barbecue]	0.07570279	0.2627935035

AICc	397.44423
BIC	456.27172
-2*LogLikelihood	327.00945
-2*Firth LogLikelihood	244.39836

Converged in Gradient
Firth Bias-Adjusted Estimates

Term Lists the terms in the model.

Estimate An estimate of the parameter associated with the corresponding term. In discrete choice experiments, parameter estimates are sometimes referred to as *part-worths*. Each part-worth is the coefficient of utility associated with the given term. By default, these estimates are based on the Firth bias-corrected maximum likelihood estimators and therefore are considered to be more accurate than MLEs without bias correction.

Std Error An estimate of the standard deviation of the parameter estimate.

Comparison Criteria

The AICc (corrected Akaike's Information Criterion), BIC (Bayesian Information Criterion), –2Loglikelihood, and –2Firth Loglikelihood fit statistics are shown as part of the report and can be used to compare models. See *Fitting Linear Models* for more information about the first three of these measures.

The –2Firth Loglikelihood value is included in the report only when the Firth Bias-adjusted Estimates check box is checked in the launch window. This option is checked by default.

For each of these statistics, a smaller value indicates a better fit.

JMP PRO Bayesian Parameter Estimates

(Appears only if Hierarchical Bayes is selected on the launch window.) The Bayesian Parameter Estimates report gives results for model effects. The estimates are based on a Hierarchical Bayes fit that integrates the subject-level covariates into the likelihood function and estimates their effects on the parameters directly. The subject-level covariates are estimated using a version of the algorithm described in Train (2001), which incorporates Adaptive Bayes and Metropolis-Hastings approaches. Posterior means and variances are calculated for each model effect. The algorithm also provides subject-specific estimates of the model effect parameters. See "Save Subject Estimates".

During the estimation process, each individual is assigned his or her own vector of parameter estimates, essentially treating the estimates as random effects and covariates. The vector of coefficients for an individual is assumed to come from a multivariate normal distribution with arbitrary mean and covariance matrix. The likelihood function for the utility parameters for a given subject is based on a multinominal logit model for each subject's preference within a choice set, given the attributes in the choice set. The prior distribution for a given subject's vector of coefficients is normal with mean equal to zero and a diagonal covariance matrix with the same variance for each subject. The covariance matrix is assumed to come from an inverse Wishart distribution with a scale matrix that is diagonal with equal diagonal entries.

For each subject, a number of burn-in iterations at the beginning of the chain is discarded. By default, this number is equal to half of the Number of Bayesian Iterations specified on the launch window.

Figure 5.16 Bayesian Parameter Estimates Report

Bayesian Parameter Estimates

Term	Posterior Mean	Posterior Std Dev	Subject Std Dev
Product Of[Canada]	-0.444803533	0.1857067541	0.2930798802
Citizenship[Canadian]*Product Of[Canada]	-0.103783512	0.2338637067	0.1945040440
Gender[Female]*Product Of[Canada]	-0.258577340	0.2205559716	0.1846953243

Total Iterations	5000
Burn-In Iterations	2500
Number of Respondents	30
Avg Log Likelihood After Burn-In	-196.5618

Term The model term.

Posterior Mean The parameter estimate for the term's coefficient. For each iteration after the burn-in period, the mean of the subject-specific coefficient estimates is computed. The Posterior Mean is the average of these means.

Tip: Select the red triangle option Save Bayes Chain to see the individual estimates for each iteration.

Posterior Std Dev The standard deviation of the means of the subject-specific estimates over the iterations after burn-in.

Subject Std Dev The standard deviation of the subject-specific estimates around the posterior mean.

Tip: Select the red triangle option Save Subject Estimates to see the individual estimates.

Total Iterations The total number of iterations performed, including the burn-in period.

Burn-In Iterations The number of burn-in iterations, which are discarded. This number is equal to half of the Number of Bayesian Iterations specified on the launch window.

Number of Respondents The number of subjects

Avg Log Likelihood After Burn-In The average of the log-likelihood function, computed on values obtained after the burn-in period.

Likelihood Ratio Tests

Figure 5.17 shows the Likelihood Ratio Tests report obtained by running the script **MaxDiff for Flavor** in Potato Chip Responses.jmp.

Figure 5.17 Likelihood Ratio Tests

Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Flavor	66.757	9	<.0001*
Citizenship*Flavor	14.609	9	0.1023
Gender*Flavor	9.480	9	0.3942

Source Lists the effects in the model.

L-R ChiSquare The value of the likelihood ratio ChiSquare statistic for a test of the corresponding effect.

DF The degrees of freedom for the ChiSquare test.

Prob>ChiSq The *p*-value for the ChiSquare test.

Bar Chart Shows a bar chart of the L-R ChiSquare values.

MaxDiff Platform Options

The MaxDiff Model red triangle menu contains the following options:

JMP PRO **Show MLE Parameter Estimates** (Available for Hierarchical Bayes.) Shows non-Firth maximum likelihood estimates and standard errors for the coefficients of model terms. These estimates are used as starting values for the Hierarchical Bayes algorithm.

Joint Factor Tests (Not available for Hierarchical.) Tests each factor in the model by constructing a likelihood ratio test for all the effects involving that factor. For more information about Joint Factor Tests, see *Fitting Linear Models*.

Confidence Intervals (Not available for Hierarchical Bayes) Shows or hides a confidence interval for each parameter in the Parameter Estimates report.

Confidence Limits (Available for Hierarchical Bayes) Shows or hides confidence limits for each parameter in the Bayesian Parameter Estimates report. The limits are constructed based on the 2.5 and 97.5 quantiles of the posterior distribution.

Correlation of Estimates If Hierarchical Bayes was not selected, shows or hides the correlations between the maximum likelihood parameter estimates.

For Hierarchical Bayes, shows or hides the correlation matrix for the posterior means of the parameter estimates. The correlations are calculated from the iterations after burn-in. The posterior means from each iteration after burn-in are treated as if they are columns in a data table. The Correlation of Estimates table is obtained by calculating the correlation matrix for these columns.

Comparisons Performs comparisons between specific alternative choice profiles. Enables you to select factor values and the values that you want to compare. You can compare specific configurations, including comparing all settings on the left or right by selecting the **Any** check boxes. Using **Any** does not compare all combinations across features, but rather all combinations of comparisons, one feature at a time, using the left settings as the settings for the other factors. See "Comparisons Report".

All Levels Comparison Report Shows the All Levels Comparison Report, which contains a table with information about all pairwise comparisons of profiles. If you are modeling subject effects, you must specify a combination of subject effects and the table is specific to that combination of subject effects. Each cell of the table shows the difference in utilities for the row level and column level, the standard error of the difference, and a Wald *p*-value for a test of no difference.

Caution: The *p*-values are not corrected for multiple comparisons. Use these results as a guide.

The Wald *p*-values are colored. A saturated blue (respectively, red) color indicates that the Difference (Row - Column) is negative (respectively positive). The intensity of the red and blue coloring indicates the degree of significance.

Save Utility Formula When the analysis is on multiple data tables, creates a new data table that contains a formula column for utility. The new data table contains a row for each subject and profile combination, and columns for the profiles and the subject effects. When the analysis is on one data table, a new Utility Formula column is added.

Save Gradients by Subject (Not available for Hierarchical Bayes.) Constructs a new table that has a row for each subject containing the average (Hessian-scaled-gradient) steps for the likelihood function on each parameter. This corresponds to using a Lagrangian multiplier test for separating that subject from the remaining subjects. These values can later be clustered, using the built-in-script, to indicate unique market segments represented in the data. See "Example of Segmentation".

Note: When a subject gradient is non-estimable it is set to missing in the Gradient by Subject table.

JMP PRO **Save Subject Estimates** (Available for Hierarchical Bayes.) Creates a table where each row contains the subject-specific parameter estimates for each effect. The distribution of subject-specific parameter effects for each effect is centered at the estimate for the term given in the Bayesian Parameter Estimates report. The Subject Acceptance Rate gives the rate of acceptance for draws of new parameter estimates during the Metropolis-Hastings step. Generally, an acceptance rate of 0.20 is considered to be good. See "Bayesian Parameter Estimates".

JMP PRO **Save Bayes Chain** (Available for Hierarchical Bayes.) Creates a table that gives information about the chain of iterations used in computing subject-specific Bayesian estimates. See "Save Bayes Chain".

Model Dialog Shows the MaxDiff launch window that resulted in the current analysis, which can be used to modify and re-fit the model. You can specify new data sets, new IDs, and new model effects.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Comparisons Report

The Comparisons report is shown when you specify pairwise comparisons. It contains the following columns:

Factor Shows the levels of the subject factors that you specified.

Compared 1 Shows the factor and levels for the profile variables in the first component of the comparison.

Compared 2 Shows the factor and levels for the profile variables in the second component of the comparison.

Utility 1 Shows the estimated utility of the first component for the subjects specified in the Factor column.

Utility 2 Shows the estimated utility of the second component for the subjects specified in the Factor column.

Probability 1 Shows the predicted probability that the first component is preferred to the second for the subjects specified in the Factor column.

Probability 2 Shows the predicted probability that the second component is preferred to the first for the subjects specified in the Factor column.

Odds 1 Probability 1 divided by Probability 2.

Odds 2 Probability 2 divided by Probability 1.

Comparison Difference Utility 1 minus Utility 2.

Standard Deviation The sample standard error of the estimated Comparison Difference.

JMP PRO Save Bayes Chain

(Available for models fit with Hierarchal Bayes.) You can use the Bayes Chain data table to determine whether your estimates have stabilized. The table that is created has a number of rows equal to the Number of Bayesian Iterations (specified on the launch window) plus one. The first row, Iteration 1, gives the starting values. Subsequent rows show the results of the iterations, in order. The table has a column for the iteration counts, the model Log Likelihood, and columns corresponding to each model effect:

Iteration Gives the iteration number, where the first row shows starting values.

Log Likelihood The log-likelihood of the model for that iteration. You can plot the Log Likelihood against Iteration to view behavior over the burn-in and tuning periods.

Adaptive Sigma for <model effect> Gives the estimate of the square root of the diagonal entries of the inverse Wishart distribution scale matrix for the corresponding effect.

Acceptance for <model effect> Gives the sampling acceptance rate for the corresponding effect.

Mean of <model effect> Gives the estimated mean for the corresponding effect.

Variance of <model effect> Gives the estimated variance for the corresponding effect.

Chapter 6

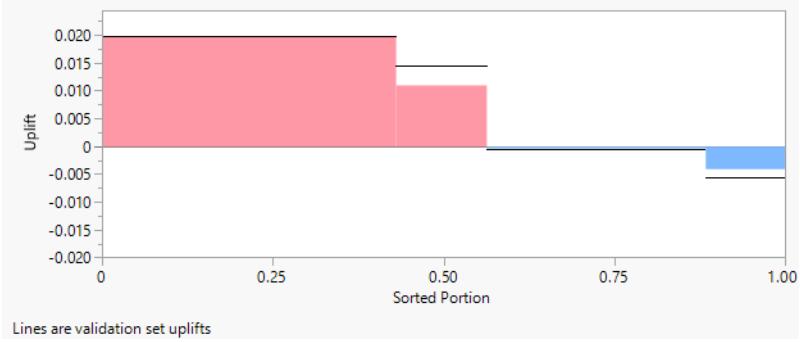
JMP PRO Uplift

Model the Incremental Impact of Actions on Consumer Behavior

The Uplift platform is available only in JMP Pro.

Use uplift modeling to optimize marketing decisions, to define personalized medicine protocols, or, more generally, to identify characteristics of individuals who are likely to respond to an intervention. Also known as incremental modeling, true lift modeling, or net modeling, uplift modeling differs from traditional modeling techniques in that it finds the interactions between a treatment and other variables. It directs focus to individuals who are likely to react positively to an action or treatment.

Figure 6.1 Example of Uplift for a Hair Product Marketing Campaign



Contents

Overview of the Uplift Platform 169
Example of the Uplift Platform 170
Launch the Uplift Platform 172
The Uplift Model Report 173
Uplift Model Graph 173
Uplift Report Options 176

JMP PRO Overview of the Uplift Platform

Use the Uplift platform to model the incremental impact of an action, or *treatment*, on individuals. An uplift model helps identify groups of individuals who are most likely to respond to the action. Identification of these groups leads to efficient and targeted decisions that optimize resource allocation and impact on the individual. See Radcliffe and Surry (2011).

The Uplift platform fits partition models. Although traditional partition models select splits to optimize classification, uplift models select splits to maximize treatment differences.

The uplift partition model accounts for the fact that individuals are grouped by a treatment factor. To determine splits, models are fit to all possible binary splits of each factor. The type of model that is fit is dependent on the type of response. A continuous response is modeled as a linear function of the split, the treatment, and the interaction of the split and treatment. A categorical response is expressed as a logistic function of the split, the treatment, and the interaction of the split and treatment. In either case, the interaction term measures the difference in uplift between the groups of individuals in the two splits. The most significant split is selected and the process repeats.

The Uplift platform selects the most significant split based on the significance of interaction terms in each of the binary split models. However, predictor selection based solely on *p*-values introduces bias in favor of predictors with many levels that result in many models for the single predictor. For this reason, JMP adjusts *p*-values to account for the number of levels or models considered. The correction used is based on Monte Carlo simulation. See Sall (2002). The splits are determined by the minimum adjusted *p*-values for tests of the significance of the interaction effect across models for all possible binary splits across all predictors. The logworth for each adjusted *p*-value, namely $-\log_{10}(\text{adj } p\text{-value})$, is reported.

JMP PRO Example of the Uplift Platform

The Hair Care Product.jmp sample data table results from a marketing campaign designed to increase purchases of a hair coloring product targeting both genders. For purposes of designing the study and tracking purchases, 126,184 "club card" members of a major beauty supply chain were identified. Approximately half of these members were randomly selected and sent a promotional offer for the product. Purchases of the product over a subsequent three-month period by all club card members were tracked.

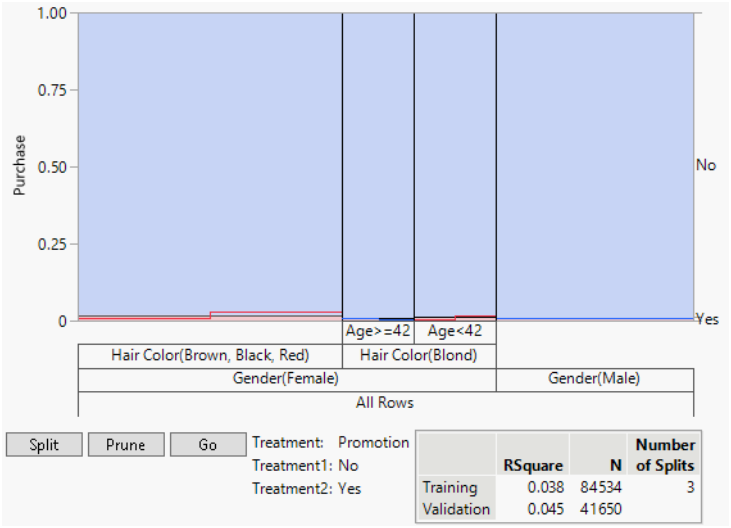
The data table shows a Promotion column, indicating whether the member received promotional material. The column Purchase indicates whether the member purchased the product over the test period. For each member, Gender, Age, Hair Color (natural), U.S. Region, and Residence (whether the member is located in an urban area) was assembled. Also shown is a Validation column consisting of about 33% of the subjects.

For a categorical response, the Uplift platform interprets the first level in its value ordering as the response of interest. This is why the column Purchase has the Value Order column property. This property ensures that "Yes" responses are first in the ordering.

1. Select **Help > Sample Data Library** and open Hair Care Product.jmp.
2. Select **Analyze > Consumer Research > Uplift**.
3. From the Select Columns list:
   - Select Promotion and click **Treatment**.
   - Select Purchase and click **Y, Response**.
   - Select Gender, Age, Hair Color, U.S. Region, and Residence, and click **X, Factor**.
   - Select Validation and click **Validation**.
4. Click **OK**.
5. Below the Graph in the report that appears, click **Go**.

   Based on the validation set, the optimal Number of Splits is determined to be three. Note that the left vertical scale is locked in order to maintain the overall rates of outcomes displayed on the right vertical axis.

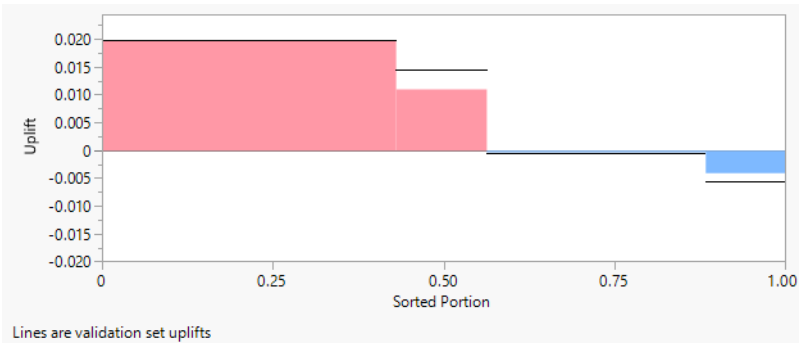
Figure 6.2 Graph after Three Splits



The right hand vertical axis in the graph indicates that the proportion of purchases is small compared to non-purchases. The graph shows that uplift in purchases occurs for females with black, red, or brown hair and for younger females (Age < 42) with blond hair. For older blond-haired women (Age ≥ 42) and males, the promotion has a negative effect.

6. Click the Uplift Model for Purchase red triangle and select **Uplift Graph**.

Figure 6.3 Uplift Graph

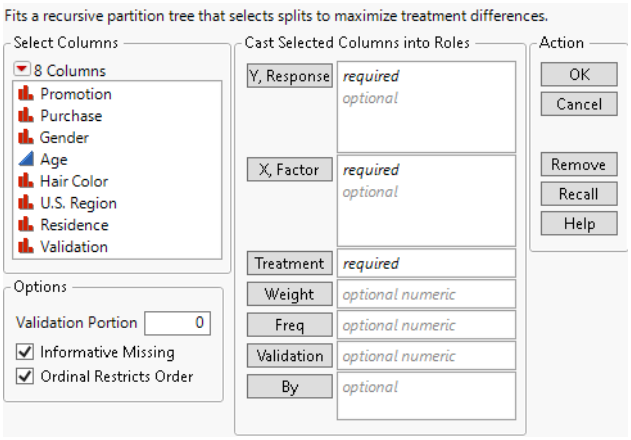


Notice that for two groups of subjects (males and non-blond women in the Age ≥ 42 group), the promotion has a negative effect. The horizontal lines shown on the Uplift Graph delineate the graph for the validation set. Specifically, the decision tree is evaluated for the validation set and the Uplift Graph is constructed from the estimated uplifts.

JMP PRO Launch the Uplift Platform

Launch the Uplift platform by selecting **Analyze > Consumer Research > Uplift**.

Figure 6.4 Uplift Platform Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Y, Response Assigns one or more columns to be analyzed.

X, Factor Assigns one or more columns to be used as factors.

Treatment Assigns a categorical treatment column. If the treatment column contains more than two levels, the first level is treated as one treatment level and the remaining levels are combined into a second treatment level.

Weight Assigns a numeric column that contains a weight for each observation in the data table. A row is included in the analysis only when its weight is greater than zero.

Freq Assigns a frequency variable to this role. This is useful for summarized data.

Validation Assigns a numeric column that defines the validation sets. This column should contain at most three distinct values:

- If there are two values, the smaller value defines the training set and the larger value defines the validation set.
- If there are three values, these values define the training, validation, and test sets in order of increasing size.

– If the validation column has more than three levels, the rows that contain the smallest three values define the validation sets. All other rows are excluded from the analysis.

The Uplift platform uses the validation column to train and tune the model or to train, tune, and evaluate the model. For more information about validation, see *Predictive and Specialized Modeling*.

If you click the Validation button with no columns selected in the Select Columns list, you can add a validation column to your data table. See *Predictive and Specialized Modeling*.

By Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variable.

Validation Portion The portion of the data to be used as a validation set. Enter a value between 0 and 1.

Informative Missing If selected, enables missing value categorization for categorical predictors and informative treatment of missing values for continuous predictors.

Ordinal Restricts Order If selected, restricts consideration of splits to those that preserve the ordering.

JMP PRO The Uplift Model Report

- "Uplift Model Graph"
- "Uplift Report Options"

JMP PRO Uplift Model Graph

The graph represents the response on the vertical axis. The horizontal axis corresponds to observations, arranged by nodes. For each node, a black horizontal line shows the mean response. Within each split, there is a subsplit for treatment shown by a red or blue line. These lines indicate the mean responses for each of the two treatment groups within the split. The value ordering of the treatment column determines the placement order of these lines. As nodes are split, the graph updates to show the splits beneath the horizontal axis. Vertical lines divide the splits.

Beneath the graph are the control buttons: **Split**, **Prune**, and **Go**. The Go button appears only if there is a validation set. Also shown is the name of the Treatment column and its two levels, called Treatment1 and Treatment2. If more than two levels are specified for the Treatment column, all levels except the first are treated as a single level and combined into Treatment2.

To the right of the Treatment column information is a report showing summary values relating to prediction. (Keep in mind that prediction is not the objective in uplift modeling.) The report updates as splitting occurs. If a validation set is used, values are shown for both the training and the validation sets.

RSquare The RSquare for the regression model associated with the tree. Note that the regression model includes interactions with the treatment column. An RSquare closer to 1 indicates a better fit to the data than does an RSquare closer to 0.

Note: A low RSquare value suggests that there might be variables not in the model that account for the unexplained variation. However, if your data are subject to a large amount of inherent variation, even a useful uplift model can have a low RSquare value.

RMSE The root mean square error (RMSE) for the regression model associated with the tree. RMSE is given only for continuous responses. See *Fitting Linear Models*.

N The number of observations.

Number of Splits The number of times splitting has occurred.

AICc The Corrected Akaike Information Criterion (AICc), computed using the associated regression model. AICc is given only for continuous responses. See *Fitting Linear Models*.

JMP PRO Uplift Decision Tree

The decision tree shows the splits used to model uplift. See Figure 6.5 for an example using the Hair Care Product.jmp sample data table. Each node contains the following information:

Treatment The name of the treatment column is shown, with its two levels.

Rate (Appears only for two-level categorical responses.) For each treatment level, the proportion of subjects in this node who responded.

Mean (Appears only for continuous responses.) For each treatment level, the mean response for subjects in this node.

Count The number of subjects in this node in the specified treatment level.

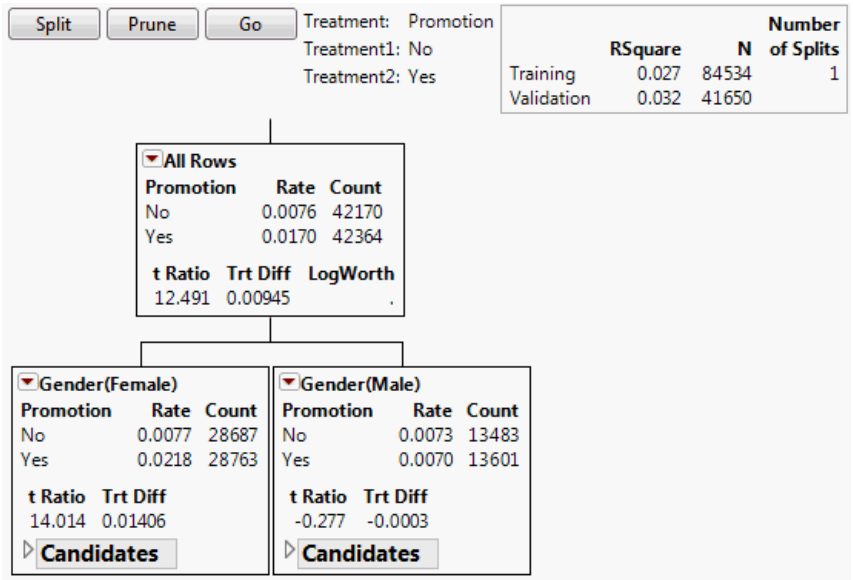
t Ratio The *t* ratio for the test for a difference in response across the levels of Treatment for subjects in this node. If the response is categorical, it is treated as continuous (values 0 and 1) for this test.

Trt Diff The difference in response means across the levels of Treatment. This is the uplift, with the following assumptions:

- The first level in the treatment column's value ordering represents the treatment.
- The response is defined so that larger values reflect greater impact.

LogWorth The value of the logworth for the subsequent split based on the given node.

Figure 6.5 Nodes for First Split



Candidates Report

Each node also contains a Candidates report with additional information:

Term The model term.

LogWorth The maximum logworth over all possible splits for the given term. The logworth corresponding to a split is $-\log_{10}$ of the adjusted p-value.

F Ratio When the response is continuous, this is the F Ratio associated with the interaction term in a linear regression model. The regression model specifies the response as a linear function of the treatment, the binary split, and their interaction. When the response is categorical, this is the ChiSquare value for the interaction term in a nominal logistic model.

Gamma When the response is continuous, this is the coefficient of the interaction term in the linear regression model used in computing the F ratio. When the response is categorical, this is an estimate of the interaction constructed from Firth-adjusted log-odds ratios.

Cut Point If the term is continuous, this is the point that defines the split. If the term is categorical, this describes the first (left) node.

JMP PRO Uplift Report Options

With the exception of the options described below, all of the red triangle options for the Uplift report are described in the documentation for the Partition platform. For more information about these options, see *Predictive and Specialized Modeling*.

JMP PRO Minimum Size Split

This option presents a window where you enter a number or a fractional portion of the total sample size to define the minimum size split allowed. To specify a number, enter a value greater than or equal to 1. To specify a fraction of the sample size, enter a value less than 1. The default value for the Uplift platform is set to 25 or the floor of the number of rows divided by 2,000, whichever value is greater.

JMP PRO Column Uplift Contributions

This table and plot address a column's contribution to the uplift tree structure. A column's contribution is computed as the sum of the F Ratio values associated with its splits. Recall that these values measure the significance of the treatment-by-split interaction term in the regression model.

JMP PRO Uplift Graph

Consider the observations in the training set. Define uplift for an observation as the difference between the predicted probabilities or means across the levels of Treatment for the observation's terminal node. These uplift values are sorted in descending order. On its vertical axis, the Uplift Graph shows the uplift values. On its horizontal axis, the graph shows the proportion of observations with each uplift value.

JMP PRO Save Columns

Save Difference Saves the estimated difference in mean responses across levels of Treatment for the observation's node. This is the estimated uplift.

Save Difference Formula Saves the formula for the Difference, or uplift.

Publish Difference Formula Creates the difference formula and saves it as a formula column script in the Formula Depot platform. If a Formula Depot report is not open, this option creates a Formula Depot report. See *Predictive and Specialized Modeling*.

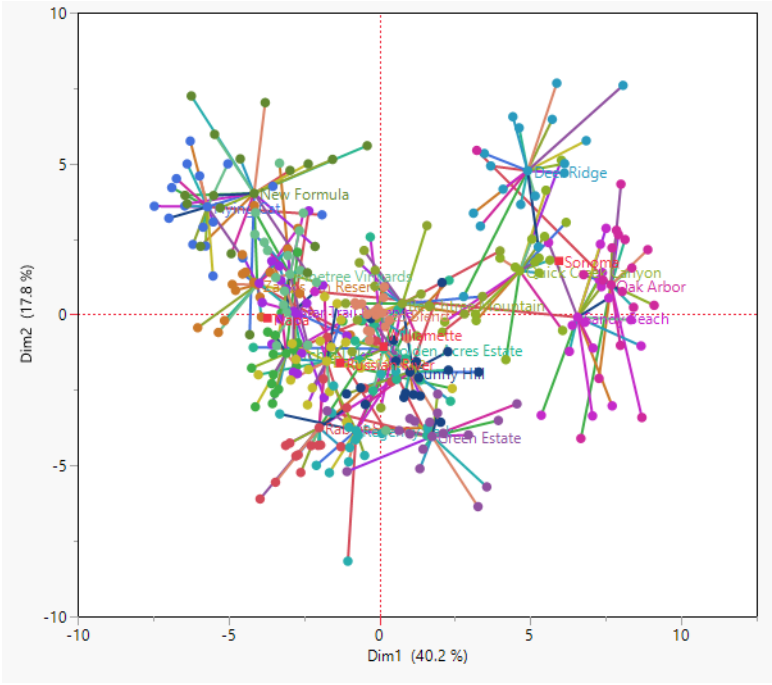
Chapter 7

Multiple Factor Analysis

Analyze Agreement among Panelists

Multiple factor analysis (MFA) is an analytical method closely related to principal components analysis (PCA). MFA uses eigenvalue decomposition to transform multiple measurements on the same items into orthogonal principal components. These components can help you understand how the items are similar and how they are different. MFA uses multiple table or consensus PCA techniques.

Figure 7.1 Consensus Map in Multiple Factor Analysis



Contents

Overview of the Multiple Factor Analysis Platform 179
Example of Multiple Factor Analysis 180
Launch the Multiple Factor Analysis Platform 183
Data Format 184
The Multiple Factor Analysis Report 186
Summary Plots 186
Consensus Map 187
Multiple Factor Analysis Platform Options 187
Statistical Details for the Multiple Factor Analysis Platform 190

Overview of the Multiple Factor Analysis Platform

Multiple factor analysis (MFA) is an analytical method that is closely related to principal components analysis (PCA). However, MFA differs from PCA in that it combines measurements from more than one table. Such tables are sometimes called sub-tables or sub-matrices. Each sub-table has the same number of rows, which represent the items or products being tested. In JMP, sub-tables are represented as groups of columns in a single data table. Each column group is called a block. Note the following about blocks:

- The number of columns in a block can vary. For example, in sensory analysis, a block represents a panelist. Some panelists might rate fewer attributes of a product than other panelists.
- Each block of columns can represent different measurements entirely. MFA scales each block to enable global analysis of all measurements.

The primary goal of MFA is to find groupings of products (rows in a data table) that are similar. A secondary goal is to identify outlier panelists. An outlier panelist results are so different from the rest of the group that they change the study results. Supplementary variables can be used investigate why items group together.

You can use MFA to analyze studies where items are measured on the same or different attributes by different instruments, individuals, or under different circumstances. MFA is frequently used in sensory analysis to account for different measurements among panelists. Traditional sensory analysis can entail hours of up-front training to ensure that panelists' measurements are consistent with each other. For example, consider a juice product with sensory measurements described as "fruity", "sweet", and "refreshing". In traditional sensory analysis, each panelist would have to be trained and tested to make sure reporting on distinct sensory measurements was consistent across panelists. MFA enables the researcher to perform a PCA-like analysis with untrained panelists.

When you use MFA, the same items are measured each time and the measurements can be arranged into internally consistent groups or blocks. For sensory analysis, the rows are the items measured, and the columns are the sensory aspects recorded by each panelist (there is a block for each panelist). Missing observations are replaced by the column mean.

For more information about multiple factor analysis, see Abdi et al. (2013).

Example of Multiple Factor Analysis

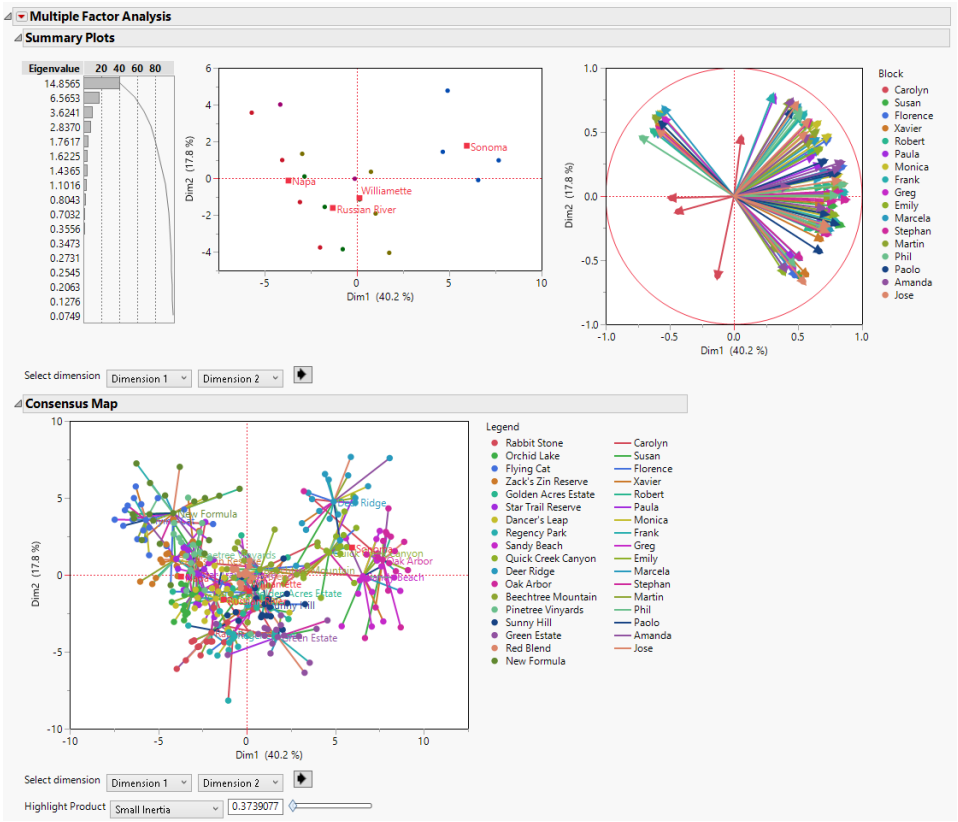
This example uses data from a simulated sensory panel study of wine characteristics. Participants rated 16 wines on a number of characteristics from 1 (no intensity) to 10 (prominent intensity). You want to better understand how the 16 wines are similar or different.

1. Select **Help > Sample Data Library** and open Wine Sensory Data.jmp.
2. Select **Analyze > Consumer Research > Multiple Factor Analysis**.
3. Select Vineyard and click **Product ID**.
4. Select Region and click **Z, Supplementary**.
5. Select all of the column groups from Carolyn to Jose and click **Add Block**.

   Note: The columns in this data table are grouped into one block for each panelist. For ungrouped data, select the columns for a block, click **Add Block**, and repeat for each block.

6. Click **Run Model**.

Figure 7.2 Initial Multiple Factor Analysis Report



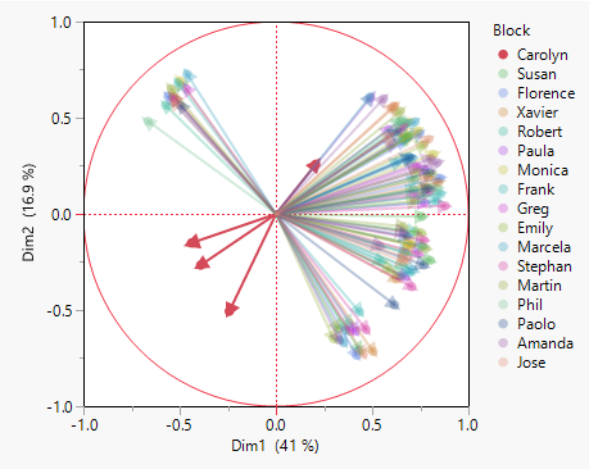
Tip: To set the legend in the Consensus Map to two columns, double-click the legend and set the Item Wrap in the Legend Settings to 18.

Notice the following in the Summary Plots:

- In the plot of the factor scores in the first two dimensions, the wines tend to cluster together according to their regions.
- In the loading plot, the rays in the lower left quadrant correspond to Carolyn. They indicate a difference between Carolyn and the other raters.

7. In the legend next to the loading plot, click **Carolyn** to highlight her results.

Figure 7.3 Loading Plot with Results for Carolyn Highlighted

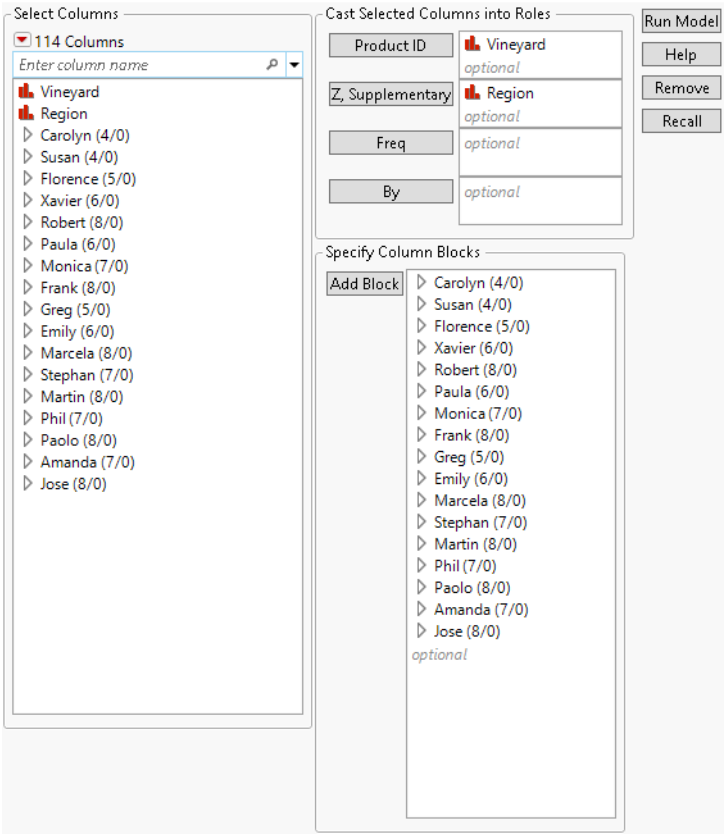


Carolyn's results differ from the other panelists. You might want to re-run the analysis without her results. See Figure 7.6 for the results of the analysis without Carolyn.

Launch the Multiple Factor Analysis Platform

Launch the Multiple Factor Analysis Platform by selecting **Analyze > Consumer Research > Multiple Factor Analysis**.

Figure 7.4 The Multiple Factor Analysis Launch Window



For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

Product ID Specifies columns of items or products to be analyzed.

Z, Supplementary Specifies the columns to be used as supplementary variables. These variables are those with which you are interested in identifying associations, but they are not included in the calculations.

Freq Identifies one column whose numeric values assign a frequency to each row in the analysis.

By A column or columns whose levels define separate analyses. For each level of the specified column, the corresponding rows are analyzed using the other variables that you have specified. The results are presented in separate reports. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

Add Block Use to add one or more columns to a block:

- Adds individual columns as a single block.
- Adds a column group as a block.
- Adds individual columns to a selected block.

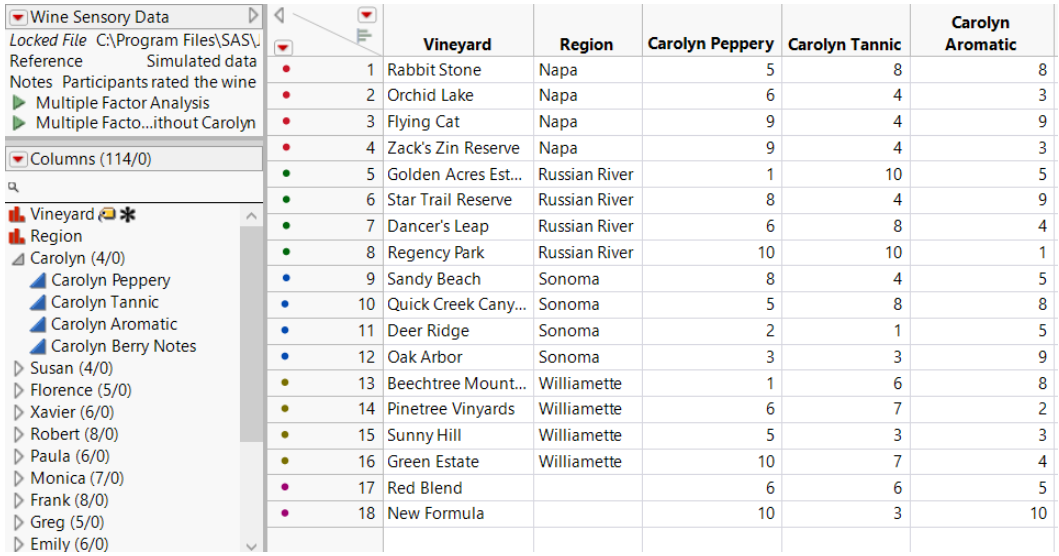
Tip: If you group the columns into blocks before running the platform, you can select multiple column groups and cast them as blocks in a single action. Otherwise, you must select each group of columns for each block and click Add Block, one block at a time. Double-click a block name to change it.

Data Format

The Multiple Factor Analysis platform uses a data table that contains column groups, or sub-tables. Each column group, referred to as a block, can have a different number of columns. Each block of columns can represent different measurements. The columns or blocks do not have to be in JMP column groups. However, the platform is easier to launch when the columns are grouped into blocks in the data table.

The data table rows represent the items that are being measured. Observations for each item must be in a single row. For example, Figure 7.5 shows a table that is measuring attributes of 16 wines from different vineyards. The column panel shows the column groups, or sub-tables, for each panelist. The Vineyards in rows 17 and 18 are not assigned a Region. The analysis could be used to explore which region the vineyards are most aligned to.

Figure 7.5 Partial View of a Data Table for Multiple Factor Analysis



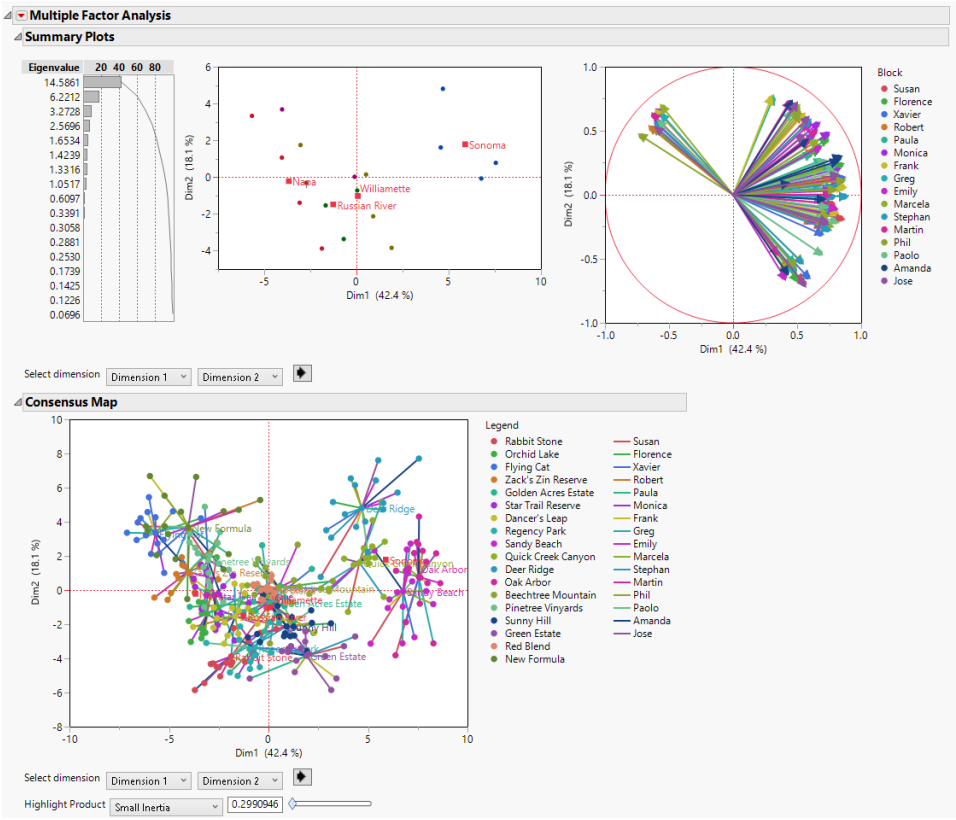
	Vineyard	Region	Carolyn Peppery	Carolyn Tannic	Carolyn Aromatic
1	Rabbit Stone	Napa	5	8	8
2	Orchid Lake	Napa	6	4	3
3	Flying Cat	Napa	9	4	9
4	Zack's Zin Reserve	Napa	9	4	3
5	Golden Acres Est...	Russian River	1	10	5
6	Star Trail Reserve	Russian River	8	4	9
7	Dancer's Leap	Russian River	6	8	4
8	Regency Park	Russian River	10	10	1
9	Sandy Beach	Sonoma	8	4	5
10	Quick Creek Cany...	Sonoma	5	8	8
11	Deer Ridge	Sonoma	2	1	5
12	Oak Arbor	Sonoma	3	3	9
13	Beechtree Mount...	Williamette	1	6	8
14	Pinetree Vinyards	Williamette	6	7	2
15	Sunny Hill	Williamette	5	3	3
16	Green Estate	Williamette	10	7	4
17	Red Blend		6	6	5
18	New Formula		10	3	10

Note: Missing observations are replaced by the column mean. When missing observations result in no variation for a column, the column is excluded from the analysis. Missing rules are applied to all variables, including supplementary variables.

The Multiple Factor Analysis Report

The initial Multiple Factor Analysis report shows a table of eigenvalues, summary plots, and a consensus map.

Figure 7.6 Multiple Factor Analysis Report



Summary Plots

The Summary Plot report has three sections:

- The first section shows the eigenvalues of the consensus PCA with a plot of the cumulative percent of variance explained by each component. A consensus PCA is used to obtain a common representation of the blocks of data. Consensus PCA refers to the principal component solution of the weighted sub-tables and is used to obtain a common representation of the blocks of data.

- The middle section is a plot of factor scores with a marker for each row (item). If supplementary variables are used, there is a labeled marker for each level of the variables. Items that cluster together in this plot are considered to be similar.
- The third section is a loading plot of factor loadings for each block.

Tip: In the loading plot legend, click a block to highlight it in the plot.

Select dimension Controls the dimensions plotted on the score and loading plots. The first control selects the horizontal dimension, and the second control selects the vertical dimension.

Consensus Map

The Consensus Map report contains a plot that overlays the individual panelist responses with the average response among all panelists for each item. You might use this map to investigate response consistency among panelists. For example, if a given panelist's points fall consistently farther from the average of the other panelists, then that panelist might be a candidate to exclude from the analysis. Hover over a data point to view the block label for that point. Click a product ID in the legend to highlight it in the consensus map.

Select dimension Controls the dimensions plotted on the consensus map. The first control selects the horizontal dimension, and the second control selects the vertical dimension.

Highlight Product Controls the transparency of the items according to their inertia score. Small inertia indicates items that panelists have good agreement on. Large inertia indicates items that panelists do not agree on.

Small Inertia Highlights items with inertia less than or equal to the value in the text box. To adjust the cutoff for the inertia value, use the slider or enter a value in the text box.

Large Inertia Highlights items with inertia greater than or equal to the value in the text box. To adjust the cutoff for the inertia value, use the slider or enter a value in the text box.

Min and Max Inertia Highlights the items with the smallest and largest inertia. The text box and slider have no impact on the results when this option is selected.

Multiple Factor Analysis Platform Options

The Multiple Factor Analysis red triangle menu includes the following options.

Block Weights Shows or hides the first eigenvalue for each block as well as the weight of that eigenvalue. The weight is the inverse of the square root of the first eigenvalue.

Eigenvalues Shows or hides a table of eigenvalues that correspond to the consensus dimensions, in order, from largest to smallest.

Eigenvectors Shows or hides a table of the eigenvectors for each of the consensus dimensions, in order, from left to right. Using these coefficients to form a linear combination of the original variables produces the consensus principal component variables.

Variable Loadings Shows or hides the loadings for each column. As in principal components analysis, loadings represent correlations of variables with components. Values near zero indicate the variable has little effect on the consensus dimension.

Variable Partial Contributions Shows or hides a table that contains the partial contributions of variables. The partial contributions represent the percentage of variance that each variable contributes to the consensus dimension.

Variable Squared Cosines Shows or hides a table that contains the squared cosines of variables. The sum of the squared cosine values across consensus dimensions is equal to 1 (100%) for each variable. The squared cosines represent the overlap in variance between variables and dimensions.

Tip: For the variable loadings, variable partial contributions, and variable squared cosines, values near zero indicate the variable is weakly related to the consensus dimension. Values far from zero indicate a strong association. The degree of transparency for the table values highlights these effects.

Summary Plots Shows or hides the summary plots. The summary plots include the plot of the eigenvalues or score plot and the loading plot.

Consensus Map Shows or hides the consensus map. See "Consensus Map".

Biplot Shows or hides a plot that is an overlay of the score and loadings plots. Use the controls to select any two dimensions for the plot.

Partial Axes Plot Shows or hides a partial axes plot. This plot displays correlations between PCA scores from separate block analyses and the consensus principal component. Use the controls to select any two dimensions for the plot. Click a block in the legend to highlight that block in the plot.

Display Options

Arrow Lines Enables you to show or hide arrows on the loading plot, and the partial axes plot. Arrows are shown if the number of variables is 1000 or fewer. If there are more than 1000 variables, the arrows are off by default.

Show Labels Shows or hides block name labels on all points in the consensus map and bi-plot. Shows or hides column name labels on all points in the partial axes plot.

Tip: Use row labels to identify centroids on the consensus map and data points on the loading plot.

RV Correlations Shows or hides a matrix of squared correlation coefficients between blocks.

Lg Coefficients Shows or hides a matrix of similarity measures between blocks.

Block Partial and Consensus Correlations Shows or hides a matrix of correlation coefficients between block partial scores and consensus principal component scores. The matrix is rectangular because only correlations between concordant dimensions are displayed.

Block Partial Contributions Shows or hides the sum of the variable contributions within the block.

Block Partial Inertias Shows or hides the block contribution multiplied by the eigenvalue for the principal component and then divided by 100.

Block Squared Cosines Shows or hides the block inertia squared and divided by the sum of squares used to calculate the eigenvalues. The values have a range between 0 and 1. The Block Squared Cosines can be considered as the percentage of the block variance explained by each principal component.

Save Individual Scores Saves the item consensus principal components to new columns in the data table. If one or more categorical supplementary variables are used, this option also saves individual scores for each level of the supplementary variables to a new data table.

Save Individual Squared Cosines Saves the item squared cosines to new columns in the data table. If one or more categorical supplementary variables are used, this option also saves categorical supplementary variable squared cosines to a new data table.

Save Individual Partial Contributions Saves the item partial contributions to new columns in the data table. If one or more categorical supplementary variables are used, this option also saves categorical supplementary partial contributions to a new data table.

Save Block Partial Scores Saves the block partial scores to a new data table.

Save Partial Axes Coordinates Saves the partial axes coordinates to a new data table.

See *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Statistical Details for the Multiple Factor Analysis Platform

Multiple factor analysis combines information from sub-tables into a set of orthogonal columns that describe the items in the rows of the table. The basic procedure is as follows:

- Perform PCA on each sub-table.
- Record the first eigenvalue of each sub-table to create a matrix of weights.
- Concatenate the sub-tables side-by-side, center and normalize the matrix.
- Perform a generalized PCA on the concatenated table via the singular value decomposition. Generalized PCA is used to constrain the solution using the sub-table weights.

This results in three matrices of generalized right and left singular vectors and singular values. These are then used to derive component scores, eigenvalues, and component loadings for the consensus across sub-tables. These three matrices are the result of decomposing the many columns from the original measurements into a few interpretable dimensions that explain the similarities and differences between the objects being measured.

Calculations

For MFA, a singular value decomposition of the **X** matrix can be defined as follows:

$$X = P\Delta Q^T \quad \text{with the constraint} \quad P^T MP = Q^T AQ = I$$

The matrices use are as follows:

X is an n x p centered and normalized matrix of sub-tables. In consumer research there are n products and p panelists' ratings.

Q is a p x q matrix of right singular vectors, which are weighted by the MFA singular values to obtain the loadings on q principal components.

Δ is a q x q diagonal matrix of singular values from the generalized PCA. As with PCA, the magnitude of the squared singular values, or eigenvalues, represent the importance of each principal component in the combined analysis.

P is an n x q matrix of left singular vectors, which are weighted by the MFA singular values to obtain the q principal components of the compromise.

M is the n x n diagonal matrix of mass weights.

A is the p x p diagonal matrix of block or panelist weights.

For more information about multiple factor analysis, see Abdi et al. (2013).

Mass Weight

JMP calculations use $N - 1$ for mass weight calculations. These calculations affect individual and block partial scores.

Appendix A

References

Abdi, H., Williams, L. J., and Valentin, D. (2013). "Multiple factor analysis: principal component analysis for multitable and multiblock data sets." *WIREs Comp Stat*, 5:149–179. doi: 10.1002/wics.1246.

Benjamini, Y., and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57:289–300.

Breslow, N. E., and Day, N. E. (1980). *The Analysis of Case-Control Studies*. Statistical Methods in Cancer Research, IARC Scientific Publications, vol. 1, no. 32. Lyon: International Agency for Research on Cancer.

Firth, D. (1993). "Bias Reduction of Maximum Likelihood Estimates." Biometrika 80:27–38.

Heinze, G., and Schemper, M. (2002). "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21:2409–2419.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

Lavassani, K. M., Movahedi, B., and Kumar, V. (2009). "Developments in Analysis of Multiple Response Survey Data in Categorical Data Analysis: The Case of Enterprise System Implementation in Large North American Firms." *Journal of Applied Quantitative Methods* 4:45–53.

Louviere, J. J., Flynn, T. N., and Marley, A. A. (2015). *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge: Cambridge University Press.

McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior." in *Frontiers in Econometrics* edited by P. Zarembka, 105–142. New York: Academic Press. Accessed April 25, 2016. https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf.

Radcliffe, N. J., and Surry, P. D. (2011). *Real-World Uplift Modeling with Significance-Based Uplift Trees*. Portrait Technical Report TR-2011-1, Stochastic Solutions. Accessed January 13, 2017. http://www.stochasticsolutions.com/pdf/sig-based-up-trees.pdf.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. Chichester, UK: John Wiley & Sons.

Sall, J. (2002). "Monte Carlo Calibration of Distributions of Partition Statistics." SAS Institute Inc.,Cary, NC. Accessed July 29, 2015. https://www.jmp.com/content/dam/jmp/documents/en/white-papers/montecarlocal.pdf.

SAS Institute Inc. (2020). "The TTEST Procedure." In *SAS/STAT 15.2 User's Guide.* Cary NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/ttest.pdf.

Train, K. E. (2001). "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit." Department of Economics, University of California, Berkley. Accessed January 13, 2017. https://eml.berkeley.edu/~train/compare.pdf.

Train, K. E. (2009). *Discrete Choice Methods and Simulation*. 2nd ed. Cambridge University Press.

Westfall, P. H., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple Comparisons and Multiple Tests Using SAS*. 2nd ed. Cary, NC: SAS Institute Inc.

Appendix B

Technology License Notices

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

Version 16

Scripting Guide

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. *JMP® 16 Scripting Guide*. Cary, NC: SAS Institute Inc.

JMP® 16 Scripting Guide

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

August 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Get the Most from JMP

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

https://www.jmp.com/getstarted

Contents

Scripting Guide

1 Learn about JMP 23

Documentation and Additional Resources

Formatting Conventions in JMP Documentation 25
JMP Help 26
JMP Documentation Library 26
Additional Resources for Learning JMP 32
JMP Tutorials 32
Sample Data Tables 32
Learn about Statistical and JSL Terms 33
Learn JMP Tips and Tricks 33
JMP Tooltips 33
JMP User Community 34
Free Online Statistical Thinking Course 34
JMP New User Welcome Kit 34
Statistics Knowledge Portal 34
JMP Training 34
JMP Books by Users 35
The JMP Starter Window 35
JMP Technical Support 35

2 Introduction 37

Welcome to the JMP Scripting Language

What JSL Can Do for You 39
Help with Learning JSL 39
The Scripting Guide 39
The Scripting Index 40
Let JMP Teach You JSL 42
Terminology 43
Basic JSL Syntax 46
Book Conventions 47

3 Getting Started 49
Let JMP Write Your Scripts

Capturing a Script for an Analysis Report 51
Capturing a Script for a Data Table 52
Capturing a Script to Import a File 53
Gluing Scripts Together 54

4 Scripting Tools 59
Using the Script Editor, Log Window, Debugger and Profiler

Work with the Script Editor 61
Introduction to the Script Editor 61
Run a Script 62
Stop a Script 63
Edit a Script 63
Color Coding and Themes 63
Autocomplete Functions 64
Tooltips 64
Split a Window 65
Match Parentheses, Brackets, and Braces 66
Select a Rectangular Block of Text 67
Select Fragmented Text 68
Drag and Drop Text 68
Find and Replace 68
Automatic Formatting 69
Errors Are Marked by Special Characters 69
Typing in Multiple Rows at Once 69
Add Code Folding Markers 70
Advanced Options 71
Set Preferences for the Script Editor 72
Work with the Log 72
Enhanced Log in JMP 73
Text Log in JMP 78
Debug or Profile Scripts 79
Debugger and Profiler Window 80
Work with Breakpoints 84
View Variables 87
Work with Watches 87
Modify Preferences in Debugger 88
Persistent Debugger Sessions 89
Examples of Debugging and Profiling Scripts 89

5 JSL Building Blocks 97

Learning the Basics of JSL

JSL Syntax Rules 100
- Value Separators 100
- Numbers 103
- Names 103
- Comments 104

Operators 105
Boolean Values 109
Global and Local Variables 110
- Local Namespaces 110
- Named Namespaces 111
- Show Symbols, Clear Symbols, and Delete Symbols 111
- Lock and Unlock Symbols 112
- Hide a Global Variable 113

Rules for Name Resolution 113
- Variable Names Used as Arguments 114
- Rules for Resolving Names 114
- Scoping Operators 116
- Troubleshoot Variables and Column Names 119
- Troubleshoot Variables and Keywords 120
- Frequently Asked Questions about Name Resolution 120

Alternatives for Gluing Expressions Together 121
Iterate 122
- For 122
- While 123
- Summation 125
- Product 125
- Break and Continue 126

Conditional Functions 128
- If 128
- Match 130
- Choose 132
- Interpolate 133
- Step 135

Compare Incomplete or Mismatched Data 135
- Missing Values 137

Inquiry Functions 138
- General Element Types 139
- Specific Element Types 139

Object Attributes ... 141
Host Information ... 141
Version Information ... 141

6 Types of Data ... 143
Working with Numbers, Strings, Dates, Currency, and More

Numbers and Strings ... 145
Formatting Numbers ... 146
Unicode Characters ... 147
Path Variables ... 147
Create and Customize Path Variables ... 150
Relative Paths ... 151
File Path Separators ... 151
Date-Time Functions and Formats ... 152
Date-Time Values ... 152
Program with Date-Time Functions ... 153
Date-Time Values in Data Tables ... 160
Custom Date-Time Formatting with Format Patterns ... 164
Currency ... 166
Hexadecimal and BLOB Functions ... 167
Work with Character Functions ... 170
Concat ... 170
Munger ... 171
Repeat ... 172
Substitute and Substitute Into ... 173
Regular Expressions ... 174
Regex ... 174
Regex Match ... 176
Special Characters in Regular Expressions ... 178
Escaped Characters in Regular Expressions ... 179
Greedy and Reluctant Regular Expressions ... 181
Backreferences and Capturing Groups ... 183
Lookaround Assertions ... 184
Pattern Matching ... 185
Parsing Strings in Fixed Fields ... 186
Patterns and Case ... 188
Troubleshooting Patterns ... 188

7 Data Structures ... 191
Working with Collections of Data

Lists ... 193

Evaluate Lists ... 193
Assignments with Lists ... 195
Perform Operations in Lists ... 195
Find the Number of Items in a List ... 195
Subscripts ... 195
Locate Items in a List ... 197
List Operators and Functions ... 198
Iterate through a List ... 201
Concatenate Lists ... 202
Insert a Nested List in an Existing List ... 203
Index into a List with Another List ... 203
Send Messages to Lists ... 203
Matrices ... 204
Construct Matrices ... 204
Subscripts ... 206
Inquiry Functions ... 210
Comparisons, Range Checks, and Logical Operators ... 210
Numeric Operations ... 211
Concatenation ... 214
Transpose ... 214
Indexing into a Matrix or List with Another Matrix or List ... 215
Matrices and Data Tables ... 216
Matrices and Reports ... 218
Loc Functions ... 219
Ranking and Sorting ... 221
Special Matrices ... 221
Inverse Matrices and Linear Systems ... 226
Decompositions and Normalizations ... 230
Build Your Own Matrix Functions ... 235
Statistical Examples ... 236
Associative Arrays ... 240
Create Associative Arrays ... 240
Work with Associative Arrays ... 243
Applications for Associative Arrays ... 246
Associative Arrays in Graph Theory ... 248
Associative Arrays in Set Operations ... 251

8 Programming Methods ... 253
Complex Scripting Techniques and Additional Functions

Lists and Expressions ... 255

Expressions 255
Macros 267
Manipulating Lists 267
Manipulating Expressions 270
Advanced Scoping and Namespaces 274
Names Default To Here 274
Scoped Names 278
Namespaces 282
Referencing Namespaces and Scopes 286
Resolving Named Variable References 289
Best Practices for Advanced Scripting 290
Advanced Programming Concepts 291
Throw and Catch Exceptions 291
Functions 293
Recursion 296
Includes 296
Load and Save Text Files 297
Run External Programs 297
Classes 299
File and Directory Operations 302
Select a Directory or File 302
Get a List of Filenames 303
Connect to a Network Directory 304
Scripting By Groups 305
Encrypt and Decrypt Scripts 306
Encryption and Global Variables 308
Encrypting Scripts in Data Tables 308
Working with Expressions in an Encrypted Script 309
Additional Numeric Functions 310
Derivatives 310
Algebraic Manipulations 312
Maximize and Minimize 314
Identifying Differences Between Strings, Lines, or Sequences 315
Scheduling Actions 317
Functions that Communicate with Users 318
Writing to the Log 319
Send Information to the User 320

9 Data Tables ... 323

Working with Data Table Objects

Get Started with Data Tables ... 327
Basic Data Table Scripting ... 330
- Open a Data Table ... 330
- Create a New Data Table ... 332
- Import Data ... 333
- Set the Current Data Table ... 357
- Name a Data Table ... 358
- Save a Data Table ... 358
- Hide a Data Table ... 359
- Print a Data Table ... 360
- Resize a Data Table ... 360
- Close a Data Table ... 360
- Set and Get a Data Table ... 361
- Perform Actions on All Open Data Tables ... 361
- Create Journals ... 362
- Create Excel Workbooks ... 363
- Prevent Changes to a Data Table ... 364

Advanced Data Table Scripting ... 365
- Select Columns in a Data Table ... 365
- Store Summary Statistics in Global Variables ... 366
- Create a Table of Summary Statistics ... 369
- Subset a Data Table ... 371
- Add a Data Filter ... 372
- Sort a Data Table ... 379
- Stack Values in a Data Table ... 379
- Split Values in a Stacked Data Table ... 380
- Transpose a Data Table ... 382
- Vertically Concatenate Data Tables ... 383
- Horizontally Concatenate Data Tables ... 383
- Virtually Join Data Tables ... 384
- Replace Data in Data Tables ... 387
- Create a Table Using Tabulate ... 389
- Copy Data Tables ... 390
- Find Missing Data Patterns ... 390
- Compare Data Tables ... 391
- Create a Summary Table ... 391
- Subscribe to a Data Table ... 392
- Move Data Between Matrices and Data Tables ... 394

Add Hyperlinks to Text Columns . . . 394
Columns . . . 396
Send Messages to Data Column Objects . . . 397
Create Columns . . . 398
Create a New Column by Text Matching . . . 400
Add Several Columns at Once . . . 401
Group Columns . . . 402
Use Set Selected to Select Columns . . . 404
Rearrange and Move Columns . . . 405
Recode Columns . . . 406
Add a Column Switcher . . . 407
Compress Selected Columns . . . 408
Create Custom Functions, Transforms, and Formats . . . 408
Custom Transform Example . . . 411
Delete Columns . . . 415
Get Column Names . . . 415
Column Attributes . . . 416
Column Properties . . . 423
Specify Engineering SI Units . . . 430
Rows . . . 431
Add Rows . . . 431
Delete Rows . . . 433
Select Rows . . . 433
Find Rows . . . 437
Move Rows . . . 437
Assign Colors and Markers to Rows . . . 438
Color Cells . . . 438
Hide, Exclude, and Label Rows . . . 439
Iterate on Rows in a Table . . . 440
Prevent Editing Rows . . . 443
Row States . . . 444
Accessing Data Values . . . 460
Set or Get Values by Column Name . . . 461
Additional Ways to Access Data Values . . . 462
Add Metadata to a Data Table . . . 463
Data Table Variables . . . 463
Table Scripts . . . 463
Formulas . . . 468
Delete Metadata . . . 469
Calculations . . . 469

Pre-Evaluated Statistics . . . 469
Calculator Formulas . . . 471

10 Scripting Platforms . . . 473
Create, Repeat, and Modify Analyses

Example of Scripting a Platform . . . 477
Send Messages to a Platform . . . 480
Conventions for Messages and Arguments . . . 480
Send Multiple Messages . . . 481
Find Messages for Objects . . . 482
Interpret the Show Properties List . . . 483
Send Messages to an Object . . . 484
Specify Which Columns to Analyze . . . 485
Create Column References . . . 485
Specify Multiple Column Names at Once . . . 486
Enable Users to Specify Columns . . . 486
Specify a By Variable . . . 486
Filter by Value using a Where Statement . . . 490
Enable User Input . . . 490
Execute Embedded Red Triangle Options . . . 491
Make Platforms Invisible . . . 491
Specify Report Titles . . . 493
Common Messages for Platform Windows . . . 493
Scripting Considerations by Platform . . . 494
Categorical . . . 494
Control Charts . . . 495
Distribution . . . 498
DOE . . . 499
Fit Model . . . 501
Formula Depot . . . 505
Graph Builder . . . 506
Neural and Neural Net . . . 508
Partial Least Squares and PLS . . . 508
Process Capability . . . 508
Scatterplot 3D . . . 509
Text Explorer . . . 510
Explore Outliers . . . 510
Scripting-Only Messages and Arguments by Platform . . . 511
Bivariate Script-Only Messages . . . 513
Boosted Tree Script-Only Messages . . . 514

Bootstrap Forest Script-Only Messages . . . 515
Bubble Plot Script-Only Messages . . . 516
Choice Script-Only Messages . . . 516
Clustering Script-Only Messages . . . 517
Contingency Script-Only Messages . . . 517
Contour Plot Script-Only Messages . . . 517
Contour Profiler Script-Only Messages . . . 517
Control Chart Script-Only Messages . . . 518
Control Chart Builder Script-Only Messages . . . 518
Cumulative Damage Script-Only Messages . . . 518
Custom Profiler Script-Only Messages . . . 519
Degradation Script-Only Messages . . . 519
Destructive Degradation Script-Only Messages . . . 519
Discriminant Script-Only Messages . . . 519
Distribution Script-Only Messages . . . 520
EMP Measurement Systems Analysis Script-Only Messages . . . 520
Explore Missing Values Script-Only Messages . . . 521
Explore Outliers Script-Only Messages . . . 521
Factor Analysis Script-Only Messages . . . 521
Fit Life by X Script-Only Messages . . . 521
Fit Model Script-Only Messages . . . 522
Fit Parametric Survival Script-Only Messages . . . 524
Functional Data Explorer Script-Only Messages . . . 525
Gaussian Process Script-Only Messages . . . 525
K Means Cluster Script-Only Messages . . . 525
K Nearest Neighbors Script-Only Messages . . . 525
Latent Class Analysis Script-Only Messages . . . 525
Life Distribution Script-Only Messages . . . 526
Make Validation Column Script-Only Messages . . . 527
MaxDiff Script-Only Messages . . . 527
Mixture Profile Script-Only Messages . . . 527
Model Screening Script-Only Messages . . . 528
Multivariate Script-Only Messages . . . 528
Naive Bayes Script-Only Messages . . . 528
Neural Script-Only Messages . . . 528
Nonlinear Script-Only Messages . . . 529
Normal Mixtures Script-Only Messages . . . 530
Oneway Script-Only Messages . . . 530
Parallel Plot Script-Only Messages . . . 530
Pareto Plot Script-Only Messages . . . 531

Partial Least Squares Script-Only Messages 531
Partition Script-Only Messages 531
Principal Components Script-Only Messages 532
Process Capability Script-Only Messages 532
Process Screening Script-Only Messages 532
Profiler Script-Only Messages 533
Reliability Forecast Script-Only Messages 533
Reliability Growth Script-Only Messages 533
Response Screening Script-Only Messages 534
Scatterplot 3D Script-Only Messages 534
Structural Equation Models Script-Only Messages 535
Support Vector Machines Script-Only Messages 536
Surface Plot Script-Only Messages 537
Tabulate Script-Only Messages 540
Ternary Plot Script-Only Messages 540
Text Explorer Script-Only Messages 540
Time Series Script-Only Messages 541
Uplift Script-Only Messages 541
Variability Chart Script-Only Messages 542

11 Display Trees 543
Create and Interact with Windows

Navigate JMP Reports 545
Examples of Common Display Boxes 546
View the Display Tree 547
Show Display Box Properties 549
Display Box Object References 550
Send Messages to Display Boxes 555
Example of Creating a Report 562
Construct Custom Windows 565
Example of Creating a Graph Box 565
Extract Values from a Window 566
Constructors for New Windows 568
Interact with a Closing Window 608
Update an Existing Display 610
Set Function and Set Script 614
Get and Set Selected Values of Display Elements That Return Lists 616
Send Messages to Constructed Displays 618
Construct Display Boxes That Contain Platforms 619
Counting the Number of Child Boxes 621

Examples of Creating a Dashboard from Two Reports . . . 622
Example of Creating a Cluster Platform Launch Window . . . 628
Example of Constructing a Custom Platform . . . 631
Example of Specifying Text and Background Colors . . . 634
Modal Windows . . . 635
Construct a Modal Window . . . 635
Construct a Column Dialog . . . 637
Comparison of Column Dialog and New Window . . . 639
Constructors for Column Dialogs . . . 640
Convert Deprecated Dialog to New Window . . . 643
New Window Example . . . 643
Example of Deprecated Dialog . . . 644
Differences Between New Window and the Deprecated Dialog . . . 646
Optional Scripts in New Window . . . 653
Scripting the Script Editor . . . 654
Technical Details . . . 655
Write Tab Box and Tab Page Box Scripts . . . 655

12 Scripting Graphs . . . 657
Editing and Creating Two-Dimensional Graphs

Add Scripts to Graphs . . . 659
Specify the Order of Graphical Elements . . . 660
Create New Graphs from Scratch and Customize Them . . . 666
Customize Graphs . . . 667
Create Shapes for a Bubble Plot . . . 669
How to Find Shapes . . . 672
Add Lines, Arrows, Points, Shapes, and Text . . . 673
Draw Lines . . . 673
Draw Arrows . . . 675
Draw Markers . . . 677
Draw Pies and Arcs . . . 678
Draw Regular Shapes: Circles, Rectangles, and Ovals . . . 680
Draw Irregular Shapes: Polygons and Contours . . . 682
Add Text . . . 685
Specify Colors . . . 686
Specify Transparency . . . 690
Add Fill Patterns . . . 691
Specify Line Types . . . 694
Draw with Pixels . . . 695
Interactive Graphs . . . 696

Handle() ... 696
Mousetrap() ... 701
Drag Functions ... 703
Troubleshoot Interactive Graphs ... 706
Create a Color Picker ... 706
Clip Display Segs in Graphics ... 707
Create Background Maps ... 709
Graph Elements ... 710
Plot Functions ... 711
Get the Properties of a Graphics Frame ... 717
Add a Legend ... 717
Hover Labels ... 718
Add Rich Text to Hover Labels Using Textlets ... 719
Add Graphs or Images to Hover Labels Using Graphlets ... 722
Update Names and Values in Hover Labels Using Gridlets ... 732
Create Drill-Down Graphs ... 737
Work with the Hover Label Execution Context ... 740

13 Three-Dimensional Scenes ... 749
Scripting in Three Dimensions

About JSL 3-D Scenes ... 751
JSL 3-D Scene Boxes ... 751
Setting the Viewing Space ... 754
Setting Up a Perspective Scene ... 755
Setting up an Orthographic Scene ... 756
Changing the View ... 757
The Translate Command ... 757
The Rotate Command ... 757
The Look At Command ... 759
The ArcBall ... 760
Graphics Primitives ... 761
Primitives Example ... 764
Controlling the Appearance of Primitives ... 765
Other Uses of Begin and End ... 771
Drawing Spheres, Cylinders, and Disks ... 771
Construction ... 771
Lighting ... 772
Drawing Text ... 773
Using Text with Rotate and Translate ... 773
Using the Matrix Stack ... 774

Lighting and Normals ... 778
Creating Light Sources ... 778
Lighting Models ... 780
Normal Vectors ... 781
Shading Model ... 781
Material Properties ... 782
Alpha Blending ... 783
Fog ... 783
Example ... 783
Bézier Curves ... 784
One-Dimensional Evaluators ... 785
Two-Dimensional Evaluators ... 787
Using the Mouse ... 788
Pick Commands ... 789
Three-Dimensional Arguments ... 790
Improve the Display of Graphics ... 790

14 Extending JMP ... 793
External Data Sources, Analytical Tools, and Automation

Real-Time Data Capture on Windows ... 797
Create a Data Feed Object ... 797
Read in Real-Time Data ... 799
Manage a Data Feed with Messages ... 800
Examples of Data Feed ... 802
Dynamic Link Libraries (DLLs) ... 804
Use Sockets in JSL ... 806
Socket-Related Commands ... 807
Messages for Sockets ... 808
Database Access ... 809
Run Query Builder Queries ... 809
Open Database Function ... 810
Creating a Database Connection and Executing SQL ... 812
Write a SQL Query ... 813
Work with SAS ... 814
Make a SAS DATA Step ... 814
Create SAS DATA Step Code for Formula Columns ... 814
SAS Variable Names ... 815
Connect to a SAS Metadata Server ... 815
Get the Values of SAS Macro Variables ... 818
Preferences ... 819

Sample Scripts . . . 819
Work with MATLAB . . . 821
Install MATLAB . . . 821
Work with R . . . 822
Installing R . . . 822
JMP to R Interfaces . . . 824
R JSL Scriptable Object Interfaces . . . 825
Conversion Between JMP Data Types and R Data Types . . . 825
Troubleshooting . . . 828
Examples . . . 829
Work with Python . . . 831
Install Python . . . 831
Troubleshooting the JMP Python Integration . . . 832
JMP-to-Python Interfaces . . . 836
Equivalent Data Types for Python Send() . . . 836
Equivalent Data Types for Python Get() . . . 837
JMP Scoping Operators and Python . . . 838
Python Name() and Python Send() Examples . . . 839
Additional Python Integration Examples . . . 840
Work with Microsoft Excel . . . 842
Parse XML . . . 843
OLE Automation . . . 845
Share Reports on JMP Live . . . 845
JMP Live Prerequisites . . . 845
Get an API Key . . . 845
Publish a Single Report to JMP Live . . . 846
Publish Several Reports to a JMP Live Folder . . . 847
Delete Reports in a JMP Live Folder . . . 847
Replace the Contents of a JMP Live Report . . . 848
Search for a JMP Live Report or Folder ID . . . 848
Update Data in JMP Live Reports . . . 849
Work with Other Content in JMP Live Reports . . . 850
HTTP Requests in JMP Live . . . 850
Communicate with REST Web Service . . . 850
Examples of HTTP Requests . . . 850
OAuth 2.0 for Web APIs . . . 853
Security . . . 853
Using Postman for Redirect URLs . . . 853
Authorization Code Grant . . . 854
Implicit Grant . . . 855

Resource Owner Grant 855
Using a Token 856
OAuth 2.0 Resources 856
Communicate with SAS Cloud Analytic Services 856
About CAS Actions 857
Connect to the CAS Server 858
Create and Submit a CAS Action 859
Create and Submit a CAS DATA Step Action 859
Create a New CAS Server 859
Import Data from the CAS Server 859
View a CAS Dataset 860
Get a List of Available CAS Libraries 860
Export Data from a Data Table to a CAS Server 860
Convert a CAS Table to a Data Table 860
Delete a CAS Table 863
Remove a CAS Table 863
Verify the CAS Server Connection 863
Get Details about the Current Connection 863
Disconnect from the CAS Server 863
Terminate All CAS Sessions 863

15 Creating Projects 865
Organizing Project Files in JMP

Create a Simple Project 867
Create a Project with a Bookmarks Pane and Log 868
Running Scripts in a Project 871
Create a Bookmark Group 872
Close Tabs in Projects 873
Get a Project 873
Reset the Window Layout 874
Run Startup Scripts to Control Workspaces 875

16 Creating Applications 877
Application Builder

Create Applications with Application Builder 879
Example of Creating an Application 879
Application Builder Terminology 881
Design an Application 882
Application Builder Window 883
Application Builder Red Triangle Options 884
Create an Application 886

Edit or Run an Application ... 897
Options for Saving Applications ... 898
Additional Examples of Creating Applications ... 899
Compile Add-Ins with the JMP Add-In Builder ... 915
Compile Scripts into an Add-In Using Add-In Builder ... 915
Edit an Add-In ... 920
Share an Add-In ... 920
Installing Multiple Add-Ins ... 920
Register an Add-In Using JSL ... 921
Create an Add-In Manually ... 921
Manage JMP Add-Ins ... 923

17 Common Tasks ... 925
Getting Started with Sample Scripts

Run a Script at Start Up ... 927
Convert Character Dates to Numeric Dates ... 928
Format Date-Time Values and Subset Data ... 929
Create a Formula Column ... 930
Extract Values from an Analysis into a Report ... 932
Create an Interactive Program ... 936

A Compatibility Notes ... 939
Changes in Scripting in JMP 16

Compatibility Issues in JMP 16.1 ... 939
Compatibility Issues in JMP 16.0 ... 939

B Efficient Scripts ... 941
Tips for Writing Scripts That Run Faster

Using Matrices ... 943
Design of Experiments ... 944
Scripting Data Tables ... 944
Scripting Graphs and Display Boxes ... 945
Scripting Numbers, Strings, Arrays, and Lists ... 946
Programming Tips ... 946
Troubleshooting ... 946

C References ... 947

D Glossary ... 949
Terms, Concepts, and Placeholders

E Technology License Notices ... 953

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation . . . 25
JMP Help . . . 26
JMP Documentation Library . . . 26
Additional Resources for Learning JMP . . . 32
JMP Tutorials . . . 32
Sample Data Tables . . . 32
Learn about Statistical and JSL Terms . . . 33
Learn JMP Tips and Tricks . . . 33
JMP Tooltips . . . 33
JMP User Community . . . 34
Free Online Statistical Thinking Course . . . 34
JMP New User Welcome Kit . . . 34
Statistics Knowledge Portal . . . 34
JMP Training . . . 34
JMP Books by Users . . . 35
The JMP Starter Window . . . 35
JMP Technical Support . . . 35

Formatting Conventions in JMP Documentation

These conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit https://www.jmp.com/software/pro.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from https://community.jmp.com.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Model Screening • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Legacy Control Charts • Pareto Plot • Diagram • Manage Spec Limits • OC Curves

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "JMP Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "JMP Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "JMP New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\16\Samples\Data

On macOS: \Library\Application Support\JMP\16\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit https://jmp.com/tools.

Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > Initial JMP Starter Window**.

JMP Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at https://www.jmp.com/support, including the technical support phone number.

Chapter 2

Introduction

Welcome to the JMP Scripting Language

The JMP Scripting Language, or *JSL*, lets you write scripts to re-create results in JMP. Power users often develop scripts to extend JMP's functionality and automate a regularly scheduled analysis in production settings. If you do not want to learn JSL, JMP can write the scripts for you.

JSL is used to perform many actions:

- implements column formulas
- launches platforms
- interactively modifies platforms
- creates graphics

Contents

What JSL Can Do for You 39
Help with Learning JSL 39
The Scripting Guide 39
The Scripting Index 40
Let JMP Teach You JSL 42
Terminology 43
Basic JSL Syntax 46
Book Conventions 47

What JSL Can Do for You

JMP can automatically save scripts to reproduce any data table or analysis in its current state. You can pause anytime in your analysis to save a script to a script window (or *script editor*), in a data table, or in an analysis report. You can then modify the script as needed for future projects. When you are finished with your work, you can then save a script to reproduce your final results.

Here are some examples where JSL scripts can be helpful:

- Suppose you need to describe an analysis process in detail, from beginning to end. An example is to create an audit trail for a governing agency, or for peers reviewing your journal article.
- Suppose you have a set of analysis steps that should be followed routinely by your lab technicians.
- Suppose you fit the same model to new data every day, and the steps are always the same.

You can use JMP interactively as usual, save scripts to reproduce your work, and in the future run those scripts to reproduce your results.

There are a few things that JSL is not designed to do:

- JMP cannot record scripts while you are working. Though script-recording is a useful feature in some other scripting languages, it is less important for software like JMP, where the results are what matter. You cannot use script-recording to observe how a sequence of interactive steps is performed.
- JSL is not an alternative command-line interface for using the program.

Help with Learning JSL

There are several places within JMP to get help with writing or understanding a JSL script.

The Scripting Guide

The *Scripting Guide* begins with basic information (such as terminology and syntax) for JMP users who are not familiar with the scripting language. The *Scripting Guide* then progresses to more advanced information.

Chapters 2 through 4	Includes information about learning JSL, producing basic scripts, and introduces you to the JSL scripting environment.
Chapters 5 through 8	Introduces the building blocks of the language; working with basic data types, such as numbers and strings; writings lists, matrices, and associate arrays; namespaces; and the fundamentals of programming in JSL.
Chapters 9 through 13	Covers using JSL with objects in JMP, such as data tables, platforms, windows, and graphics.
Chapter 14	Describes how to write scripts that work with external programs, such as SAS, R, and Microsoft Excel.
Chapter 15	Describes how to organize files that you use in an analysis, perform analyses, and run scripts from one workspace.
Chapter 16	Introduces creating JMP applications in Application Builder, a drag-and-drop environment for visually designing windows with buttons, lists, graphs, and other objects. The chapter also describes how to use Add-In Builder to compile scripts into one easily shared file.
Chapter 17	Contains a collection of recipes, or script examples, that you can copy and modify for your own use.
Appendices A, B, C, and D	Provides information about compatibility issues with the previous version of JMP, provides tips for writing more efficient scripts, gives information about references, and defines JSL concepts and terminology.

The Scripting Index

The Scripting Index on the **Help** menu provides a brief description and the syntax for JSL functions, objects, and display boxes. Each entry includes an example that you can run and modify to test your own code. And an embedded log window lets you see messages as examples are run.

Notes:

- To display the Scripting Index entry for a function in a script, run the script, press Alt, and then double-click the function. You can also run the script, right-click the function, and select **Help Scripting Index.**
- To copy the syntax or description from the Scripting Index window, right-click the area that you want to copy and select **Copy Text**.

The Scripting Index window includes the following buttons:

Click the **Clear** button to clear the search text box to begin a new search.

Click the arrow to set search filter options and parameters.

After you edit the sample script in the Scripting Index, click this button to revert to the original script.

Note: When you edit an example script, the changes persist when you view other entries in the Scripting Index. To revert the script, click **Reset**, which appears above the script after you edit it.

Search Filter Options

Click the down arrow button next to the search box to refine your search.

Contains Terms Returns items that contain a part of the search criteria. A search for “ease oom” returns messages such as “Release Zoom”.

Contains Phrase Returns items that contain the exact search criteria. A search for “text box” returns entries that contain “text” followed directly by “box” (for example, “Context Box” and “Text Box”).

Starts With Phrase Returns items that start with the search criteria.

Ends With Phrase Returns items that end with the search criteria.

Whole Phrase Returns items that consist of the entire string. A search for “text box” returns entries that contain only “text box”.

Regular Expression Enables you to use the wildcard (*) and period (.) in the search box. Searching for “get.*name” looks for items that contain “get” followed by one or more words. It returns “Get Color Theme Names”, “Get Name Info”, and “Get Effect Names”, and so on.

Invert Result Returns items that do not match the search criteria.

Match All Terms Returns items that contain both strings. A search for “t test” returns elements that contain either or both of the search strings: “Pat Test”, “Shortest Edit Script” and “Paired t test”.

Ignore Case Ignores the case in the search criteria.

Match Whole Words Returns items that contain each word in the string based on the Match All Terms setting. If you search for “data filter”, and Match All Terms is selected, entries that contain both “data” and “filter” are returned.

Search Category and Item Names Returns category names (in the left column) and item names (in the middle column) that match the search criteria.

Search Prototypes and Descriptions Returns prototypes (the JSL syntax) and descriptions that match the search criteria.

Search Examples Returns examples that match the search criteria.

Let JMP Teach You JSL

The best JSL writer is JMP. You can work in JMP interactively and then save the results as a script to reuse later. With simple modifications, your script can serve as a template for speeding up routine tasks.

Because JSL is a very flexible language, you can reach your goals in many different ways. Here is an example. Typically, the script that JMP saves for you specifies every detail of your analysis, even if most of the details happen automatically by default. Does that mean that the scripts that you write have to be just as complete and detailed? Not at all. You usually need to specify only those details that you would select in the graphical user interface (GUI). For example, if you open Big Class.jmp from the sample data folder and want to launch Distribution for height, weight, and sex, the following script is all that is necessary:

```
Distribution( Y( :height, :weight, :sex ) );
```

Suppose you run the Distribution platform in the GUI and then select **Save Script > To Script Window** from the red triangle menu for the report. The following script appears:

```
Distribution(
   Continuous Distribution( Column( :height ) ),
   Continuous Distribution( Column( :weight ) ),
   Nominal Distribution( Column( :sex ) ),
);
```

Both scripts give the same result.

Feel free to experiment with JSL. If you think something ought to be possible, it probably is. Give it a try, and see what happens.

Terminology

Before you begin creating scripts, you should become familiar with basic JSL terms used throughout the *Scripting Guide*.

Operators and Functions

An *operator* is one- or two-character symbol (such as + or =) for features such as common arithmetic actions, scoping names, regular expressions, concatenating, and subscripting.

A *function* is a command that might contain additional information for the function to use.

Certain JSL functions work the same as operators but provide access to more complex actions. For example, the following two lines are equivalent:

```
2 + 3; // returns 5
Add( 2, 3 ); // returns 5
```

The first line uses the + operator. The second line uses the `Add()` function equivalent.

Although all JSL operators have function equivalents, not all functions have operator equivalents. For example, `Sqrt(a)` can be represented only by the `Sqrt()` function.

Note: In previous versions of JMP and its documentation, the terms *operators* and *functions* were used interchangeably. Now each term has a specific meaning.

Objects and Messages

An *object* is a dynamic entity in JMP, such as a data table, a data column, a platform results window, a graph, and so on. Most objects can receive messages that instruct the object to perform some action on itself.

A *message* is a JSL expression that is directed to an object. That object knows how to evaluate the message. In the following example, `dt` is the data table object. << indicates that a message follows. In the following example, the message tells JMP to create a summary table with the specified variables.

```
dt << Summary( Group( :age ), Mean( :height ) )
```

In this expression, `dt` is the name of a variable that contains a reference to a data table. You could use any name for this variable. The *Scripting Guide* commonly uses `dt` to represent data table references. Here are some of the more common names used to represent references to certain objects:

Abbreviation	Object
`dt`	data table
`col`	column in a data table
`colname`	the name of a column in a data table
`obj`	an object
`db`	display box

These variables are not pre-assigned references. Each one must be assigned prior to its use. In the following example, the global variable named `A` is assigned the value “Hello, World”. When the `Show( A )` command is processed, the result is the value of `A`.

```
A = "Hello, World";
Show( A );
   A = "Hello, World";
```

Arguments and Parameters

An *argument* is additional information that you can provide to a function or message. For example, in `Root(25)`, `25` is an argument to the `Root()` function. `Root()` acts on the argument that you provide and returns the result: 5.

Programming and scripting books commonly talk about parameters as well. A *parameter* is a description of the argument that a function accepts. For example, the general specification for `Root()` might be `Root( number )`, where *`number`* is the parameter.

Parameter and *argument* express two perspectives of the same concept: information that a function needs.

For simplicity in the *Scripting Guide*, we use the word *argument* in both cases.

A *named argument* is an optional argument that you select from a predetermined set and explicitly define. For example, `title("My Line Graph")` in the `Graph Box()` function is a named argument because the title is explicitly defined as such.

```
Graph Box( title( "My Line Graph" ),
      Frame Size( 300, 500 ),
      Marker( Marker State( 3 ), [11 44 77], [75 25 50] );
      Pen Color( "Blue" );
      Line( [10 30 70], [88 22 44] ));
```

Note that the `Frame Size()` arguments `300` and `500` are not named. The position of these arguments implies meaning; the first argument is always the width, the second argument is always the height.

Optional Arguments

Functions and messages require certain arguments, and other arguments are optional. You can include them, but you do not have to. In specifications, optional arguments are enclosed in angle brackets. For example:

```
Root( x, <n> )
```

The *x* argument is required. The *n* argument is optional.

Optional arguments often have a default value. For example, for `Root()`, the default value of *n* is 2:

Code	Output	Explanation
`Root( 25 )`	`5`	Returns the square root of 25.
`Root( 25, 2 )`	`5`	Returns the square root of 25.
`Root( 25, 3 )`	`2.92401773821287`	Returns the cube root of 25.

Or and the Vertical Bar Symbol

A single vertical bar (`|`) represents a logical `OR`. For brevity, `|` represents the word *or* when referring to alternative values.

For example, a pathname can be either absolute or relative. When you see an argument such as `absolute|relative`, this means that you enter *either* one of the following two options:

- `absolute` indicates an absolute pathname.
- `relative` indicates a relative pathname.

More than two options can also be strung together with a vertical bar in this way.

Script Formatting

Whitespace characters (such as spaces, tabs, and newlines) and capitalization are ignored in JSL. This means that the following two expressions are equivalent:

Expression 1:

```
sum = 0;
For( i = 1, i <= 10, i++,
   sum += i;
   Show( i, sum );
);
```

Expression 2:

```
Sum = 0;
For( i = 1, i <= 10, i++,
   Sum += i;
   Show( i, Sum );
);
```

You can format your script in any way that you like. However, the script editor can also format your script for you. The *Scripting Guide* uses the script editor's default formatting for capitalization, spaces, returns, tabs, and so on. See "Work with the Script Editor" for more information about using the script editor.

Note: The only white space exception is two-character operators (such as <= or ++). The operators cannot be separated by a space.

Basic JSL Syntax

A JSL script is a series of expressions. Each expression is a section of JSL code that accomplishes a task. JSL expressions hold data, manipulate data, and send commands to objects.

Many expressions are nested message names. Message contents are enclosed in parentheses:

```
Message Name ( argument 1, argument 2, ... );
```

The meaning of JSL names depends on the context. The same name might mean one thing in a data table context and something entirely different in a function context. See "Rules for Name Resolution".

Almost anything that follows certain punctuation rules, such as matching parentheses, is a valid JSL expression. For example:

```
win = New Window( "Window Example",
   <<Modal,
   Text Box( "Hello, World" ),
   Text Box( "-----" ),
   Button Box( "OK" )
);
```

Notes:

- Names can have embedded spaces. See "Names".
- Message contents are enclosed in parentheses, which must be balanced. See "Parentheses".
- Items are separated by commas. See "Commas".
- JSL is not case sensitive; you can type "`text box();`" or "`Text Box()`".

- Messages are commonly nested inside other messages.

Book Conventions

In the *Scripting Guide*, the function names that are case sensitive are capitalized as you must use them. Arguments that are placeholders for actual choices are in `lowercase`. For example, `Connect Color` is a function that you need to type as is, and `color` stands for some color choice that you make yourself.

```
Connect Color(color);
```

In this case, the argument in parentheses must be some color value (for example, a JMP color number, or a supported color name like `"red"`, `"blue"`, and so on, or an RGB value given as a list, such as `{.75, .50, .50}`). Sometimes alternatives like these are shown with the vertical bar (`|`) character for "or," like this:

```
Connect Color( number | "color name" | {r,g,b} );
```

Syntax coloring is applied to scripts that you can paste into a script editor and then run.

Chapter 3

Getting Started

Let JMP Write Your Scripts

You often have to produce the same reports for the same data on a regular basis. This chapter shows you how to let JMP write scripts for common tasks like importing text data, opening Microsoft Excel files, and producing reports. A final tutorial shows you how to put it all together into a single script to open an Microsoft Excel file and produce three reports automatically.

The *Scripting Guide* is written for users who are familiar with JMP but might not be familiar with JSL. For more information about performing common tasks, see *Using JMP*. *Discovering JMP* is also a good resource for learning basic concepts and understanding the JMP workflow.

Contents

Capturing a Script for an Analysis Report 51
Capturing a Script for a Data Table 52
Capturing a Script to Import a File 53
Gluing Scripts Together 54

Capturing a Script for an Analysis Report

Here are the basic steps for capturing a script to reproduce an analysis:

1. Launch a platform, such as Distribution.
2. Make any changes or additions that you need. For example, add tests and other graphs.
3. Capture the script to recreate your results.

You can save the script in the data table, so that if you send the data table to others, they can run your script and duplicate your reports.

Example

Follow these steps to produce a distribution report, capture the script to reproduce it, and save it to the data table.

Note: The data tables that you use in examples are located in JMP's Samples/Data folder.

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze > Distribution** to open the Distribution launch window.
3. Select Profits ($M) in the Select Columns box and click the **Y, Columns** button.
4. Click **OK**.

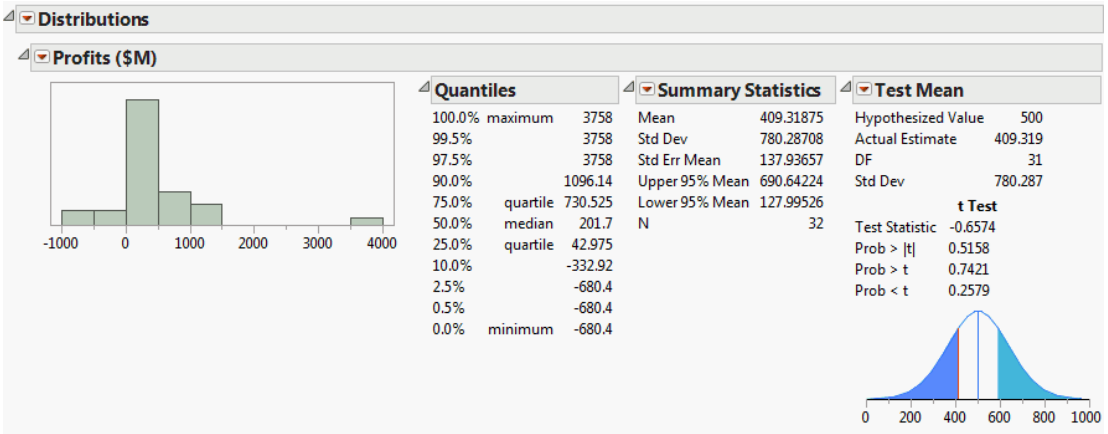
   The Distribution report window appears.
5. Click the Distributions red triangle and select **Stack** to make your report horizontal.
6. Click the red triangle next to Profits ($M) and deselect **Outlier Box Plot** to turn the option off.
7. Click the red triangle next to Profits ($M) and select **Test Mean**.

   The Test Mean window appears.
8. Type 500 in the **Specify Hypothesized Mean** box.
9. Click **OK**.

   The test for the mean is added to the report window.

   Now you have your customized report.

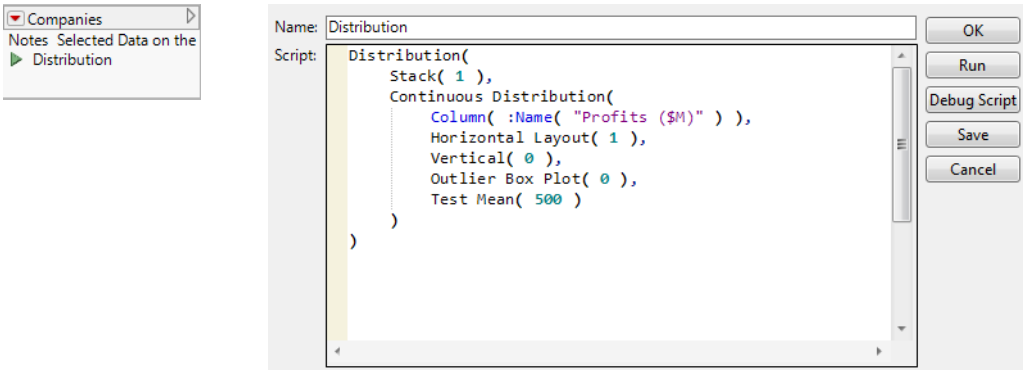
Figure 3.1 Customized Distribution Report



10. Click the Distributions red triangle and select **Save Script > To Data Table**.

    Your data table now has a script named Distribution saved to it. Right-click the Distribution script and select **Edit** to see the script.

Figure 3.2 Distribution Script Saved to the Data Table



11. To run the script and reproduce your final report exactly, click the green triangle next to the script.

Capturing a Script for a Data Table

Here are the basic steps for capturing a script to reproduce a data table:

1. Open the data table.

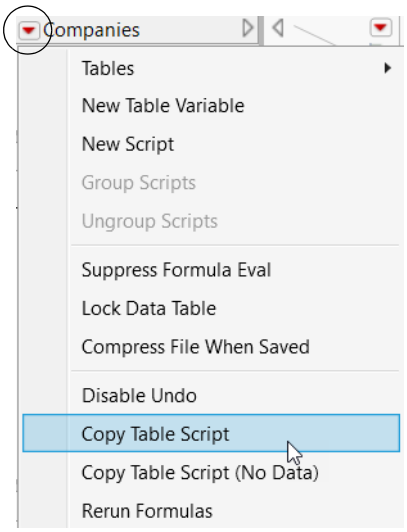
2. Make any changes that you need. For example, add a script, correct values, add new columns.
3. Capture the script to recreate your data table.

Example

Use the data table from the previous example, where you saved a script to it.

1. In the data table, select the red triangle next to the data table's name.
2. Select **Copy Table Script**.

Figure 3.3 Copy the Table Script



3. Open a script window by selecting **File > New > Script**.
4. Select **Edit > Paste**.

You now have a script that duplicates your data table. You can save this script and run it at any time to recreate your data table, with all its scripts attached.

Capturing a Script to Import a File

To capture a script that imports a file, you open the file in JMP. JMP automatically records the steps that occurred when you opened the file.

Import a Text File

1. Select **File > Open**.

The Open Data File window appears.

2. Select **Text Files** from the list next to **File name**.
3. In the Open as section, select **Data (Best Guess).**

   JMP formats the data based on tabs, commas, white space, and other characters in the text file.
4. Browse to select the file, and then select **Open**.

   The file is opened as a data table. The data table includes a script named Source. This JSL script imports your text file with the text import rules that you used.
5. Right-click the Source script and select **Edit**.

   You can copy this script, paste it into a new script window, and save it. Then you can run this script later to reimport the text file.

Tip: The import script is an `Open()` expression that specifies the text file and the import options to correctly import the file into JMP. The first part of this expression is the pathname to the specific file that you imported. If you save this script and want to run it a different place, you might need to edit the pathname so that it points to the text file. Pathnames are discussed in greater detail in "Path Variables".

Gluing Scripts Together

Suppose new data is saved out to an Microsoft Excel file once a week, and you need to produce the same reports every week. You could open the file and perform the same steps every week. However, creating a script that imports the new Microsoft Excel file into JMP and runs all analyses automatically is more efficient. The following example shows you how to set up your script and run it each week.

Import the Microsoft Excel File

1. Open a new script window (**File > New > Script**).
2. In your script window, enter the `Open()` expression to open the Solubil.xlsx sample import data file. The file is located in JMP's Samples/Import Data folder.

   ```
   dt = Open( "$SAMPLE_IMPORT_DATA/Solubil.xlsx" );
   ```

   Be sure to put the semicolon at the end of this expression, because you will add more expressions. The semicolon glues expressions together.
3. Run your script to import the Microsoft Excel file by selecting **Edit > Run Script**.

   The Microsoft Excel file opens as a data table.

Notes:

- You can also include the `Excel Wizard` argument in the `Open()` expression to preview the worksheet before importing it. See "Import Data from a Microsoft Excel File".
- You can specify an absolute or relative path to the file rather than using a path variable. For relative links, the script and file being opened must be in the same relative location each time you run the script. With absolute links, make sure that other users running the script have access to the file's location. See "Path Variables" for more information about using pathnames.

Run Your Reports and Capture Their Scripts

You have three reports to produce: a distribution report, a 3D scatterplot, and a multivariate report. Perform each one using the GUI, and add its script to the script window.

1. With your new data table open, select **Analyze > Distribution**.
2. Select all the columns except Labels and click **Y, Columns**.
3. Click **OK**.
4. Press Ctrl, click the eth red triangle, and then select **Histogram Options > Show Counts**.

   Bar counts are added to all six histograms.
5. In the Distribution window, click the Distributions red triangle and select **Save Script > To Clipboard**.
6. Click a line or two after your `Open()` expression and select **Edit > Paste**.
7. Type a semicolon after the last close parenthesis.
8. Select **Graph > Scatterplot 3D**.
9. Select all the columns except Labels and click **Y, Columns**.
10. Click **OK**.
11. Copy and paste the script for Scatterplot 3D into the script window just like you did for your Distribution report. Be sure to add the semicolon at the end.
12. Select **Analyze > Multivariate Methods > Multivariate**.
13. Select all the columns except Labels and click **Y, Columns**.
14. Click **OK**.
15. Copy and paste the script for Multivariate into the script window just like you did for Distributions and Scatterplot 3D.

    You should see the following script:

```
dt = Open( "$SAMPLE_IMPORT_DATA/Solubil.xlsx" );
Distribution(
  Continuous Distribution( Column( :eth ), Show Counts( 1 ) ),
  Continuous Distribution( Column( :oct ), Show Counts( 1 ) ),
```

```
    Continuous Distribution( Column( :cc14 ), Show Counts( 1 ) ),
    Continuous Distribution( Column( :c6c6 ), Show Counts( 1 ) ),
    Continuous Distribution( Column( :hex ), Show Counts( 1 ) ),
    Continuous Distribution( Column( :chc13 ), Show Counts( 1 ) ),
);
Scatterplot 3D(
     Y( :eth, :oct, :cc14, :c6c6, :hex, :chc13 ),
     Frame3D( Set Grab Handles( 0 ), Set Rotation( -54, 0, 38) )
);
Multivariate(
   Y( :eth, :oct, :cc14, :c6c6, :hex, :chc13 ),
   Estimation Method( "Row-wise" ),
   Scatterplot Matrix( Density Ellipses( 0 ), Shaded Ellipses( 0 ) )
);
```

Save the Script

You now have a script that reproduces all of the steps that you performed manually. Save the script, and close your data table and all its report windows.

1. In the script window that contains your script, select **File > Save** or **File > Save As**.
2. Specify a filename (for example, Weekly Report).
3. Click **Save**.

Run the Script

As long as your weekly updated Microsoft Excel file is saved in the same place and contains the same columns, you can run your script and automatically produce all your reports.

1. Open the script that you saved.
2. Select **Edit > Run Script**.

   Your Microsoft Excel file is opened in JMP, and all three of your reports appear.

You can send this script to others. As long as they have access to the same Microsoft Excel file in the same location, they can also run the script in JMP and see your reports.

Advanced Note: Auto-Submit

If you want a particular script to always be executed instead of opened into the script window, put the following command on the first line of the script:

```
//!
```

If this is not the very first line, with nothing else on the same line, this command does nothing.

You can override this command when opening the file.

1. Select **File > Open**.

2. Press Ctrl, select the JSL file, and then click **Open**.

   The script opens into a script window instead of being executed.

The command is also ignored when you right-click the file in the Home Window and select **Edit Script**.

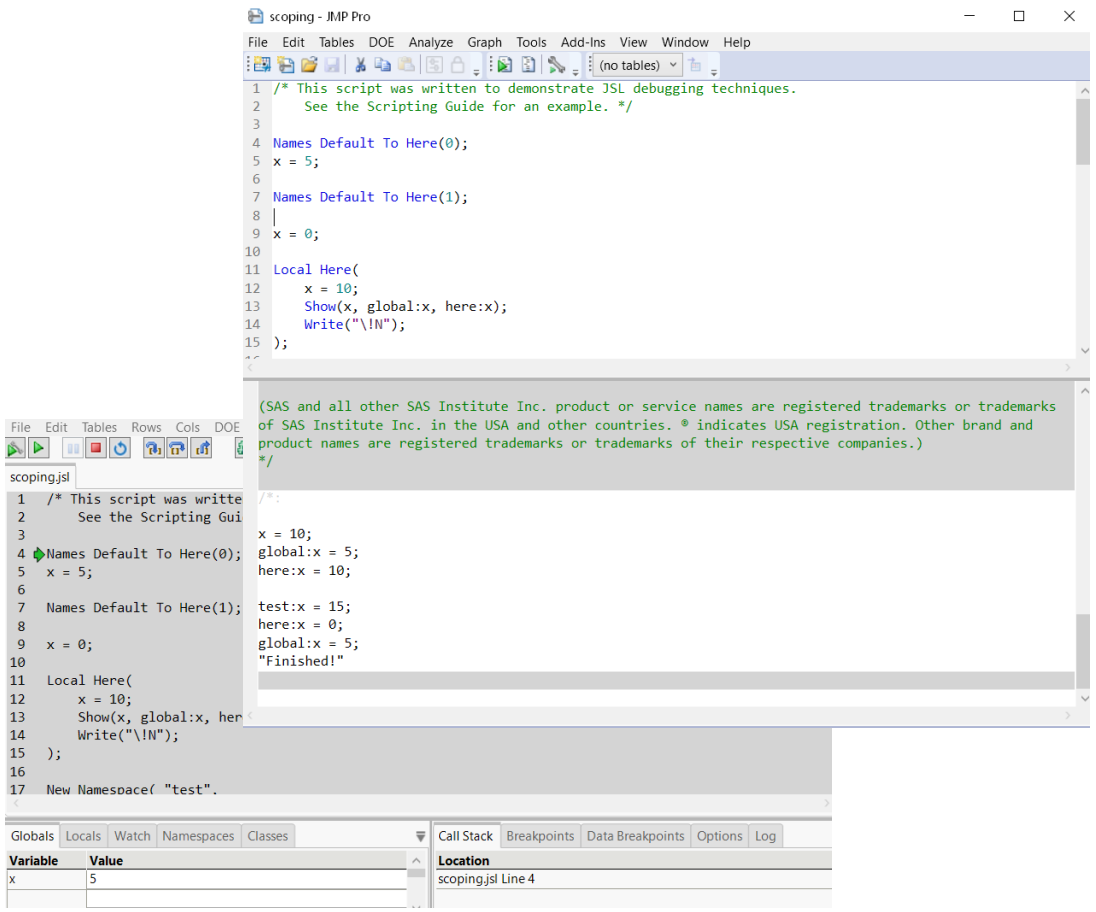
Chapter 4

Scripting Tools

Using the Script Editor, Log Window, Debugger and Profiler

JMP provides several programming tools for script writers. The script editor supports syntax coloring, autocompletes functions as you type, highlights matching braces, allows for code folding, and has additional features to help you develop scripts more quickly. Error messages and output are shown in the log window, which can be displayed inside the script editor. The JMP Scripting Language (JSL) Debugger and Profiler can help you troubleshoot your scripts.

Figure 4.1 Script Editor with Embedded Log and the Debugger



Contents

Work with the Script Editor . . . 61
Introduction to the Script Editor . . . 61
Run a Script . . . 62
Stop a Script . . . 63
Edit a Script . . . 63
Color Coding and Themes . . . 63
Autocomplete Functions . . . 64
Tooltips . . . 64
Split a Window . . . 65
Match Parentheses, Brackets, and Braces . . . 66
Select a Rectangular Block of Text . . . 67
Select Fragmented Text . . . 68
Drag and Drop Text . . . 68
Find and Replace . . . 68
Automatic Formatting . . . 69
Errors Are Marked by Special Characters . . . 69
Typing in Multiple Rows at Once . . . 69
Add Code Folding Markers . . . 70
Advanced Options . . . 71
Set Preferences for the Script Editor . . . 72
Work with the Log . . . 72
Enhanced Log in JMP . . . 73
Text Log in JMP . . . 78
Debug or Profile Scripts . . . 79
Debugger and Profiler Window . . . 80
Work with Breakpoints . . . 84
View Variables . . . 87
Work with Watches . . . 87
Modify Preferences in Debugger . . . 88
Persistent Debugger Sessions . . . 89
Examples of Debugging and Profiling Scripts . . . 89

Work with the Script Editor

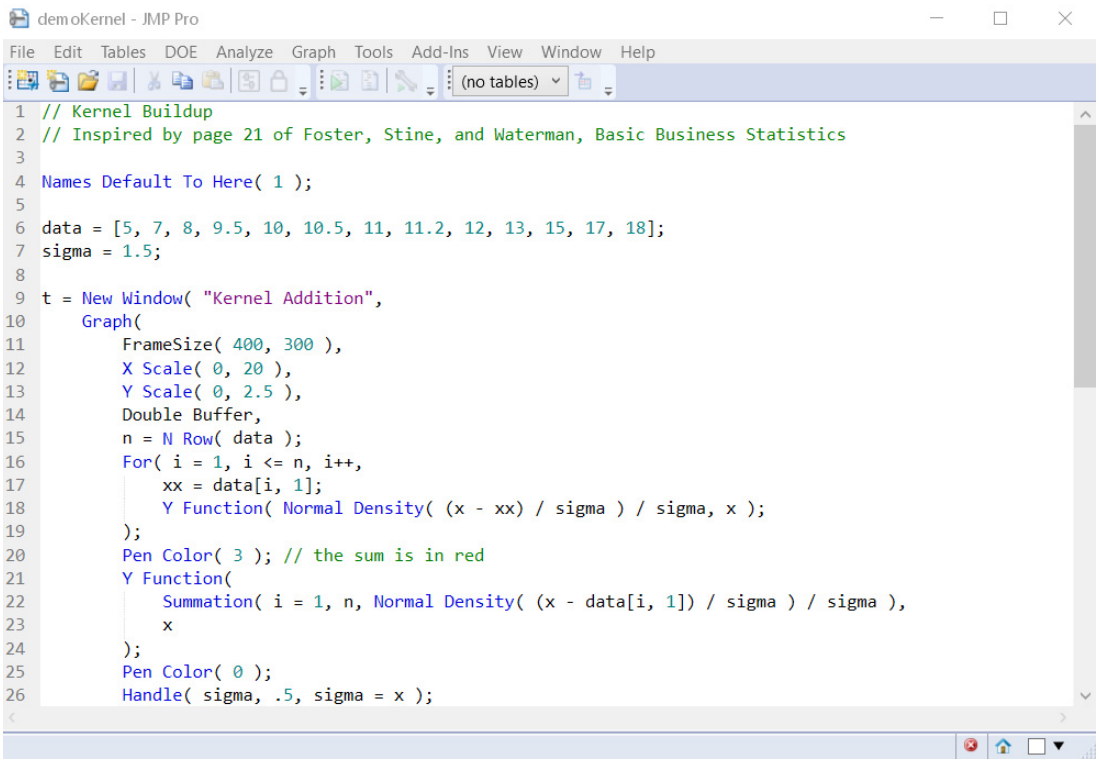
- "Introduction to the Script Editor"
- "Run a Script"
- "Stop a Script"
- "Edit a Script"
- "Color Coding and Themes"
- "Autocomplete Functions"
- "Tooltips"
- "Split a Window"
- "Match Parentheses, Brackets, and Braces"
- "Select a Rectangular Block of Text"
- "Select Fragmented Text"
- "Drag and Drop Text"
- "Find and Replace"
- "Automatic Formatting"
- "Errors Are Marked by Special Characters"
- "Typing in Multiple Rows at Once"
- "Add Code Folding Markers"
- "Advanced Options"
- "Set Preferences for the Script Editor"

Introduction to the Script Editor

The script editor provides a friendly environment for writing and reading JSL scripts. Figure 4.2 shows basic features such as syntax coloring, inline commenting, and automatic formatting. Other common programming options are described later in this section.

Script editor features are also available in the log window and anywhere else that you can edit or write a script (for example, in the Scripting Index or Application Builder).

Figure 4.2 The Script Editor



Tip: Consider setting the *Autosave timeout* value in the General preferences to automatically save open scripts at the specified number of minutes. This autosave value also applies to data tables, journals, scripts, projects, and reports.

Run a Script

To run an entire script, select **Edit > Run Script.**

To run specific lines in a script, select those lines and then select **Edit > Run Script**.

To run specific lines that are not adjacent, press Ctrl, select the lines, and then select **Edit > Run Script**.

On Windows, you can also click in a line or select several lines and press Enter on your numeric keypad.

Run the script automatically when you open it by using one of the following methods:

- Type `//!` on the first line.
- Include `Run JSL( 1 )` in the `Open()` statement:

```
Open( "$SAMPLE_SCRIPTS/scoping.jsl", Run JSL( 1 ) );
```

Stop a Script

To stop the script, press Esc on Windows (or Command-period on macOS). You can also select **Edit > Stop Script**. On macOS, **Edit > Stop Script** is available only when the script is running.

Edit a Script

To edit a script on Windows, press Insert. You can then type directly over (overwrite) any existing JSL code. Note that this feature is not available on macOS.

Color Coding and Themes

On Windows, the default font in the script editor is 10-point Consolas. On macOS, the default script editor font is 13-point Menlo.

JMP applies the following colors in the script window:

- black for text, identifiers (JSL functions), braces, and user macros
- white for the script editor background
- gray for the disabled background and guides
- green for comments
- purple for strings
- maroon for platform names

  Note: If a platform name is used in a message, the JSL messages color is used.

- teal blue for numbers
- dark blue for operator symbols, the first keyword, and JSL messages
- medium blue for function names, second and third keywords, and macros
- red for unknown objects

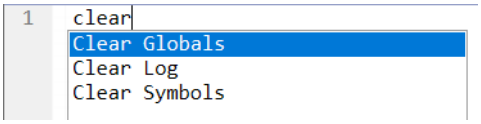
You can customize colors and fonts in the Customize Styles section of the Script Editor preferences. Syntax coloring for JSL, the log, SAS code, JavaScript, Text, JMP Text Analytics, R, Matlab, Python, C, JSON, and SQL are supported. In addition, options for dark and light themes are provided.

Autocomplete Functions

If you do not remember the exact name of a function, use autocompletion to see a list of functions that match what you have typed so far. Type part of the name, and then press Ctrl+Space on Windows (Option+Esc on macOS).

Suppose that you want to clear your JSL variables, but do not remember the command. You can type `clear` and then press Ctrl+Space, to see a list of possible clear commands. Select the command that you want to insert.

Figure 4.3 Autocomplete Example

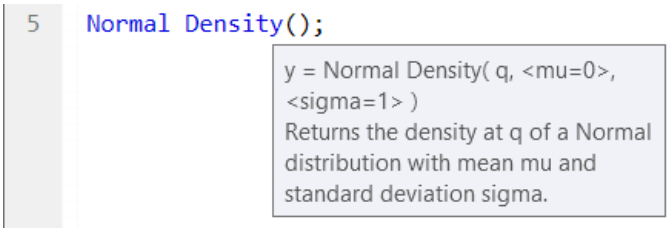


Tooltips

If you are using a function and do not remember the syntax or need more information about it, place the cursor over it to see a brief explanation. This works only with JSL function names, not platform commands, messages, or user-created functions. JSL function names are colored blue in the script editor.

The tooltip shows the syntax, arguments, and a brief explanation of the function (Figure 4.4). The tip also appears in the script editor window status bar.

Figure 4.4 Tooltip for a JSL Function



After running a script, you can also place the cursor over variable names to see their current value. To turn off variable tooltips, deselect **Preferences > Script Editor > Show Variable Value Tips**.

To turn off function tooltips, deselect **Preferences > Script Editor > Show Operator Tips**.

Tip: In addition to showing the tooltip for a function, you can display its Scripting Index entry. Run the script, press Alt, and then double-click the function. A description, sample script, and information about syntax is provided. You can also read help for the topic by running the script, right-clicking the function, and selecting **Help Browser**.

Example of a Tooltip for a JSL Variable

1. Enter and run the following line in a script window:

   ```
   my_variable = 8;
   ```

2. Hover over the variable name after you run the line.

   A tooltip shows the name of the variable and its value: 8.

3. Enter and run the following line:

   ```
   my_variable = "eight";
   ```

4. Hover over the variable name after you run the line.

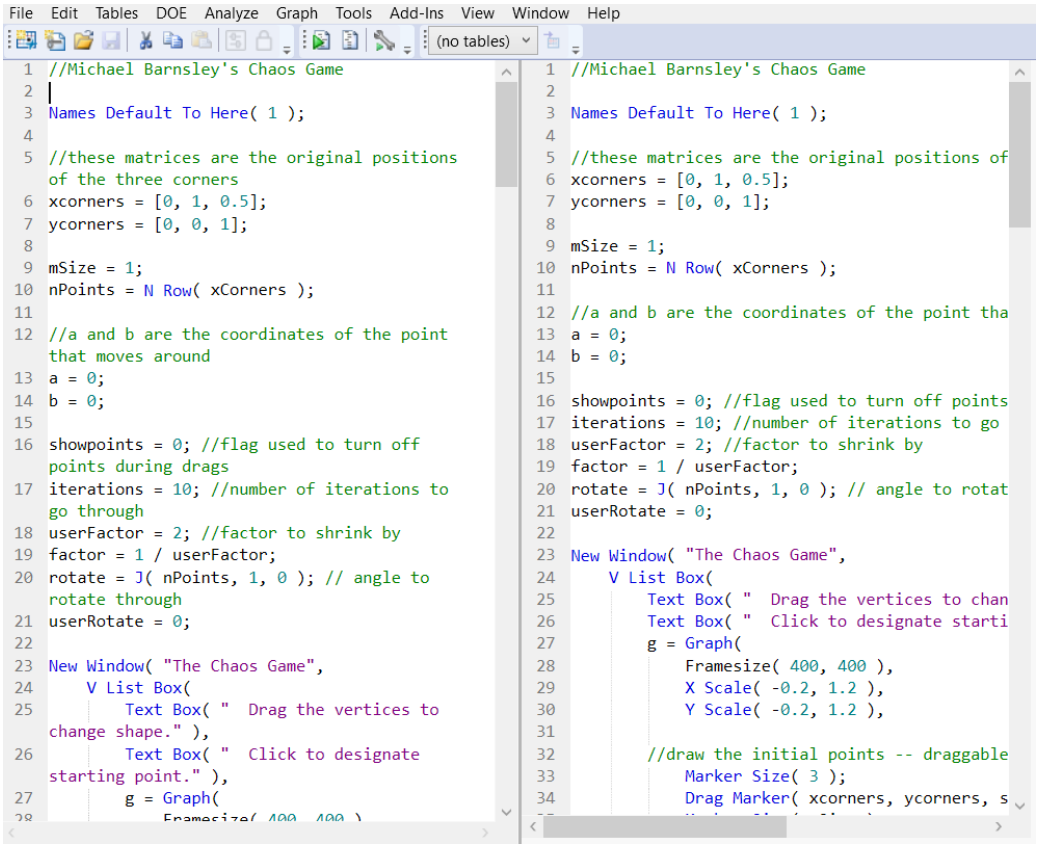
   A tooltip shows the name of the variable and its value: “eight”.

Split a Window

You can split the Script Editor window into two vertical or horizontal windows. This feature enables you to independently scroll through your code in two different places and edit the contents in both. When you make a change in one window, the change is immediately reflected in the other window.

- To split an open Script Editor window, right-click in the window and select **Split > Horizontal** or **Vertical**.
- To revert to a single window, right-click and select **Remove Split**.

Figure 4.5 Example of Splitting a Window Horizontally

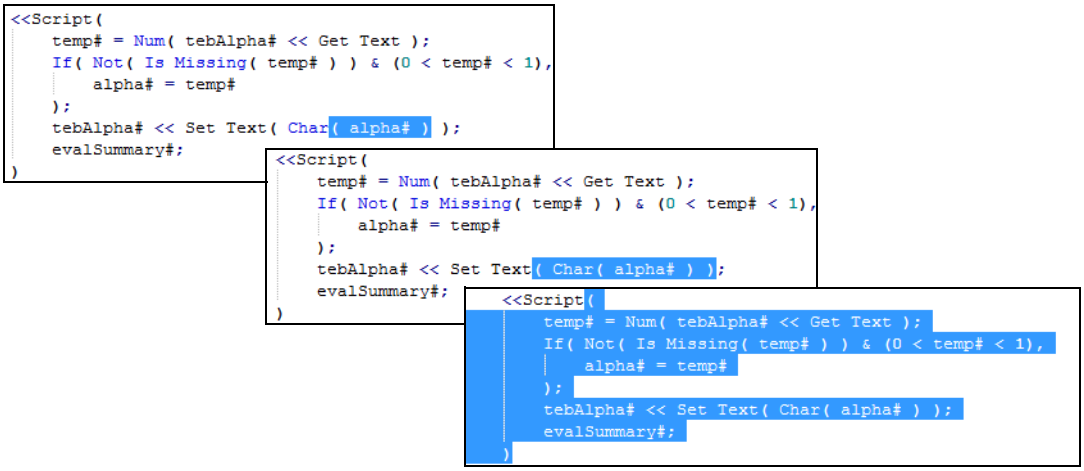


Match Parentheses, Brackets, and Braces

The script editor helps you match *fences* (or parentheses, square brackets, and braces) in the following ways:

- The matching closing fence is added when you type an opening fence.
- When you click next to either an opening or closing fence, the fence and its match are highlighted in blue. If the fence does not have a match, it is highlighted in red.
- If you double-click a fence, everything between the matching fence is selected (including the fences).
- If you put your cursor within an expression and press Ctrl+] on Windows (Command+B on macOS), the entire expression is selected. Fences that enclose the expression are included. Repeat this process to highlight the next-higher expression. Figure 4.6 shows an example.

Figure 4.6 Each Step in Matching Fences



When you type an opening brace, JMP adds the closing brace. Enter code between the braces, type the closing brace, and then your cursor automatically moves after the closing brace that JMP added. This prevents you from accidentally adding an unnecessary closing brace.

You can turn on and off the autocompletion of braces in the JMP preferences. See "Set Preferences for the Script Editor".

Select a Rectangular Block of Text

To select a rectangular block of text, press Alt and drag your cursor from the starting point to the end of the block. You can either copy or cut the text enclosed in the block.

Suppose that you want to the select all of the following code except for the comment marks.

```
// Y( :Y ),
// X( :X ),
```

Select a rectangular portion beginning with Y. When you paste, you get the following code:

```
Y( :Y ),
X( :X ),
```

The rectangular selection inserts returns where needed to maintain the structure of the text. Select `Get Menu Item State` on both lines in the following example.

```
bb << Get Menu Item State(1),
bb << Get Menu Item State(2),
```

When you paste, a return is inserted at the end of each line.

```
Get Menu Item State
Get Menu Item State
```

Select Fragmented Text

To select text that is not contiguous, press Ctrl on Windows or Command on macOS and drag your cursor over the text. Continue this action for any other text that you want to select. You can then copy and paste your selection into a new script or you can run the selected text. Text will be pasted or run in the order it was selected.

Drag and Drop Text

You can drag and drop text within a script editor window or between windows or from a data table into a script editor window. On Windows, pressing Ctrl before dragging and dropping copies the text. On macOS, the text is copied by default.

Drag and drop text in any of these ways:

- Select a row or column, pause, and then drop it into the script editor window.
- Double-click text in a text field and then drop it into the script editor window. Examples are text in a data table cell and any other selectable text.

On Windows, you can also drag and drop text into a minimized window.

1. Drag the text over the Home Window button in the lower right corner of the window.
   The Home Window appears.
2. In the Home Window list, drag the text over the destination window.
   That window appears.
3. Drop the text where you want it.

Find and Replace

Many find and replace options are available in the script editor, including the support of regular expressions. For example, searching with the following regular expression:

```
get.*name
```

returns messages such as “Get Button Name”, and “GetFontName”

Basic regular expressions such as `^` and `$` (which match the start of line and end of line) and `\n` (which matches a carriage return) are also supported.

See *Using JMP* for more information about the Search options.

Automatic Formatting

The script editor can format a script for easier reading. Any generated script (for example, by saving a platform script) is automatically formatted with tabs and returns in appropriate places.

You can also reformat individual scripts that are difficult to read (for example, scripts in which all commands are strung together with no whitespace characters). From the **Edit** menu, select **Reformat Script**.

Tip: This command alerts you if your script is badly formed (for example, if your parentheses are not matched).

Errors Are Marked by Special Characters

After you run a script, the /*###*/ characters in the log indicate the location of an error in the script execution. The following example shows a script in which the height column name is misspelled.

```
Names Default To Here( 1 );
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt:heihgt[3]; // misspelled column name
```

After you run the script, the log reads:

```
Scoped data table access requires a data table column or variable in access or
   evaluation of 'dt:heihgt' , dt:heihgt/*###*/
In the following script, error marked by /*###*/
Names Default To Here( 1 );
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt:heihgt/*###*/[3];
```

/*###*/ directly follows the misspelled height argument.

Typing in Multiple Rows at Once

When you edit a script, you can type in multiple rows at once instead of manually updating the same code on multiple lines. For example, in the following script, you want to change the s variable to aString. You can easily change the first instance of s to aString. To change the other instances at once, hold down Alt and select the column of s variables. Type aString once.

```
s = "";
For( i = 1, i <= n, i++,
   s += "Hello";
   s += "Number";
```

```
    s += Char( i );
);
```

The script then looks like this:

```
aString = "";
For( i = 1, i <= n, i++,
   aString += "Hello";
   aString += "Number";
   aString += Char( i );
);
```

Note: You can type in multiple rows at once, but you can type only in the same column. In the preceding example, the `s` variables are in the same column.

Add Code Folding Markers

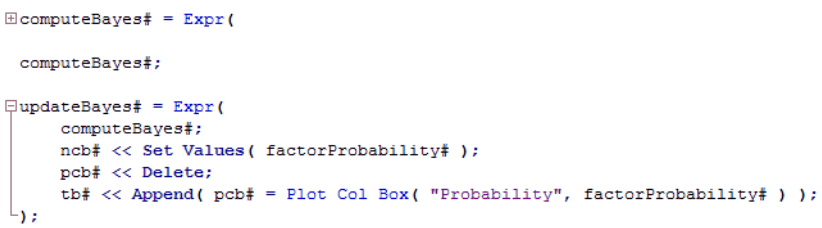
You can add code folding markers that show the beginning and the end of the code block, enabling you to collapse and expand code inside stand-alone functions.

To turn on this feature, select **Code folding** in the Script Editor preferences. Then you can expand and collapse blocks of code by right-clicking on a script and selecting **Advanced > Expand All** or **Collapse All**.

Tip: To fold or expand all code in the entire script, press Shift+Ctrl and then click a code marker.

After you select this preference, `Function` and `Expr` expressions are foldable. See "Add More Folding Keywords" for more information about adding folding markers to other expressions.

Figure 4.7 Code Folding Markers Shown in a Script



By default, code does not remain collapsed after you save the script and restart JMP. To save the state of the folded code, select **Save and restore document state information** in the Script Editor preferences.

Add More Folding Keywords

Custom code folding is supported for other stand-alone functions as shown in the following example:

```
{"If", "For", "For Each Row", "While", "Try", "New Window", "V List Box",
  "H List Box"}
```

JMP supports multiple keyword lists. A system administrator can define a set of keywords in jmpKeywords.jsl and save the script in C:/ProgramData/SAS/JMP/ or designated directory listed below. You save your version of jmpKeywords.jsl in your C:/Users/<user>/Documents/ folder. JMP merges all keyword lists from the designated directories.

On Windows, the following directories are examined in the order listed:

- C:/ProgramData/SAS/JMP/16/
- C:/ProgramData/SAS/JMP/
- C:/Users/<user>/AppData/Roaming/SAS/JMP/16/
- C:/Users/<user>/AppData/Roaming/SAS/JMP/
- C:/Users/<user>/Documents/

On macOS, the following directories are examined in the order listed:

- /Library/Application Support/JMP/16/
- /Library/Application Support/JMP/
- ~/Library/Application Support/JMP/16/
- ~/Library/Application Support/JMP/
- ~/Documents/

Note that jmpKeywords.jsl is stored in the designated JMP directory, even if you are using JMP Pro.

Notes:

- The list in jmpKeywords.jsl is case insensitive.
- Code folding is not supported for messages, platforms, user-defined functions, and comments.
- After you edit and save the list in jmpKeywords.jsl, turn the **Allow additional code folding keywords** preference off and then back on for the changes to take effect. Messages in the log indicate that the keywords were loaded.

Advanced Options

Right-clicking on selected text in the Script Editor provides the following Advanced options:

Expand All (Appears only if JSL code folding is on.) Expands all blocks of code.

Collapse All (Appears only if JSL code folding is on.) Collapses all blocks of code.

Comment Block Makes the selected text comments.

Uncomment Block Uncomments the selected comments.

Make Uppercase Changes all selected text to uppercase.

Make Lowercase Changes all selected text to lowercase.

Set Preferences for the Script Editor

In the JMP preferences, customize the script editor settings such as the font, colors, and spacing options.

Setting the Fonts

1. Select **File > Preferences**.
2. Select the **Fonts** group.
3. Click **Mono** to set the font for the script editor.

For more information about font preferences, see *Using JMP*.

Set Script Editor Preferences

Select **File > Preferences > Script Editor** to further customize the editor. For more information about script editor preferences, see *Using JMP*.

Work with the Log

The Log in JMP has two modes:

- "Enhanced Log in JMP"
- "Text Log in JMP"

You can switch between the two modes by selecting **File > Preferences > Log**.

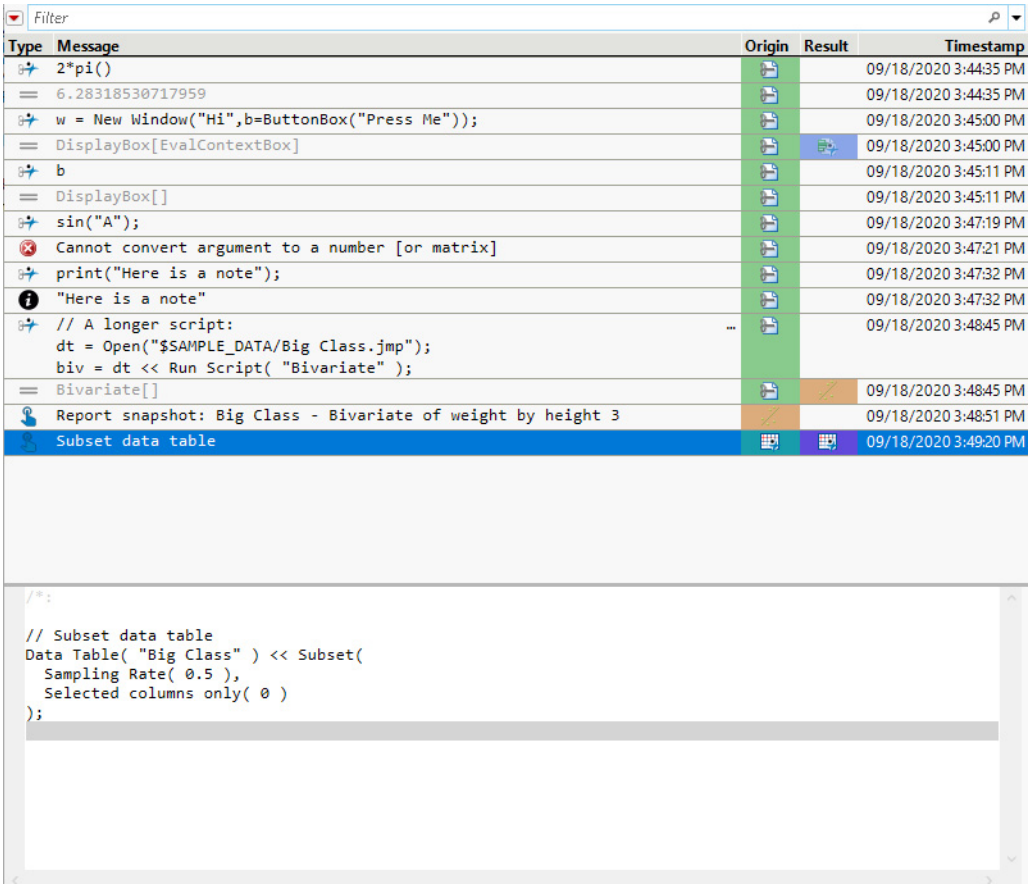
Note: The log shows the steps you take during data cleanup. For example, if you add a column property or import the data into Excel, then the JSL expressions for those actions are captured in the log. This enables you to clean up the data more efficiently when you receive a new data set.

Enhanced Log in JMP

The Enhanced Log is a tabular view of the events in a JMP session. To view the JMP Log, select **View > Log** (**Window > Log** on macOS). By default, the log in JMP appears in Enhanced mode. The log window contains events, or actions, in the rows of the table and the corresponding JSL code and output in a separate text pane. This tabular view of the events is easy to navigate, displays errors and warnings, and can help you create a JSL script of the actions in a JMP session. For an example of using the log to generate a JSL script for recorded actions, see "Example of Recording Actions in the Enhanced Log".

Note: When you used the embedded log in a script window, the embedded log does not use the Enhanced Log mode. Further, any script output from a script window with an embedded log does not appear in the Enhanced Log.

Figure 4.8 Enhanced Log



Each row of the tabular view contains information in the following columns:

Type Contains an icon that represents message type. For more information about the types of messages, see "Filter".

> **Tip:** You can hover over an icon in the Type column to see the type of message that a particular icon represents.

Message Contains a description of the event or result. Among other things, this could be a recorded action, JSL that was submitted, the result of JSL that was submitted, or an error message. Each description of a logged item or result is limited to three lines of text; longer items are denoted with an Ellipsis icon ··· in the upper right corner of the Message section of the row. The full text for longer items appears in the text pane at the bottom of the log window.

Origin Contains an icon that shows the type of JMP window from which the logged item or result was generated. If the Color by Window option is selected in the red triangle menu next to Filter, logged items generated by the same JMP window are specified by a common color. You can click on the icon in the Origin column to bring the corresponding window to the front.

> **Note:** Not all rows in the log table have an icon in the Origin column.

Result Contains an icon that shows the type of JMP window that contains the result. If the Color by Window option is selected, logged items generated by the same JMP window are specified by a common color. You can click on the icon in the Result column to bring the corresponding window to the front.

> **Note:** Not all rows in the log table have an icon in the Result column.

Timestamp Contains a date-time value that indicates when the logged item was generated.

Enhanced Log Options

The red triangle menu in the Enhanced Log contains the following options:

Go To (Available only when one or more rows of the log table are selected.) Brings the window corresponding to the selected row in the log table to the front. If an Origin window and a Result window both appear in the selected row, both windows are brought to the front with the Result window on top of the Origin window.

Run Scripts (Available only when one or more rows of the log table that contain a scriptable action are selected.) Runs the JSL script that corresponds to the selected rows in the log table.

Color by Window Specifies whether the rows in the log table are colored by the JMP window that created the row of the log output.

Clear (Available only when one or more rows of the log table are selected.) Removes the selected row or rows from the log table.

Caution: This action cannot be undone.

Clear Log Removes all rows from the log table.

Caution: This action cannot be undone.

Filter Contains the following options to filter what types of logged events appear in the log table.

Result Shows or hides rows that correspond to the results of running JSL.

Script Shows or hides rows that correspond to submitted JSL.

Log Shows or hides rows that correspond to informational log messages.

Warning Shows or hides rows that correspond to warning messages in the log.

Error Shows or hides rows that correspond to error messages in the log.

Action Shows or hides rows in the log table that correspond to interactive JMP actions. These actions include data table operations, platform launches, and closing report windows. You can customize which actions are shown in the log table by selecting specific options in the Action Recording submenu.

Action Recording Contains the following options to filter what types of interactive JMP actions appear in the log table.

Data Table Operations Shows or hides rows that correspond to manipulations of a data table. These actions can include row and column operations, as well as actions in the Tables, Rows, and Cols menus.

Platform Launch Shows or hides rows that correspond to launching a JMP platform.

Report Snapshot on Close Shows or hides rows that correspond to a JSL script that is saved when a report window is closed.

Save Script Contains the following options that save a script that reproduces all selected rows in the log table. If no rows are selected, these options save a script to reproduce all events recorded in the log table.

To Script Window Saves the script to a script window.

To Clipboard Saves the script to the clipboard.

Make Into Data Table Creates a data table that contains all rows from the log table. If rows are selected in the log table, they are also selected in the data table that is created.

Filter the Results in the Enhanced Log

To the right of the red triangle menu, use the search field to filter the items in the log table view.

Click the down arrow button next to the search box to refine your search.

Contains Terms Returns items that contain a part of the search criteria. A search for "ease oom" returns messages such as "Release Zoom".

Contains Phrase Returns items that contain the exact search criteria. A search for "text box" returns entries that contain "text" followed directly by "box" (for example, "Context Box" and "Text Box").

Starts With Phrase Returns items that start with the search criteria.

Ends With Phrase Returns items that end with the search criteria.

Whole Phrase Returns items that consist of the entire string. A search for "text box" returns entries that contain only "text box".

Regular Expression Enables you to use the wildcard (*) and period (.) in the search box. Searching for "get.*name" looks for items that contain "get" followed by one or more words. It returns "Get Color Theme Names", "Get Name Info", and "Get Effect Names", and so on.

Invert Result Returns items that do not match the search criteria.

Match All Terms Returns items that contain both strings. A search for "t test" returns elements that contain either or both of the search strings: "Pat Test", "Shortest Edit Script" and "Paired t test".

Ignore Case Ignores the case in the search criteria.

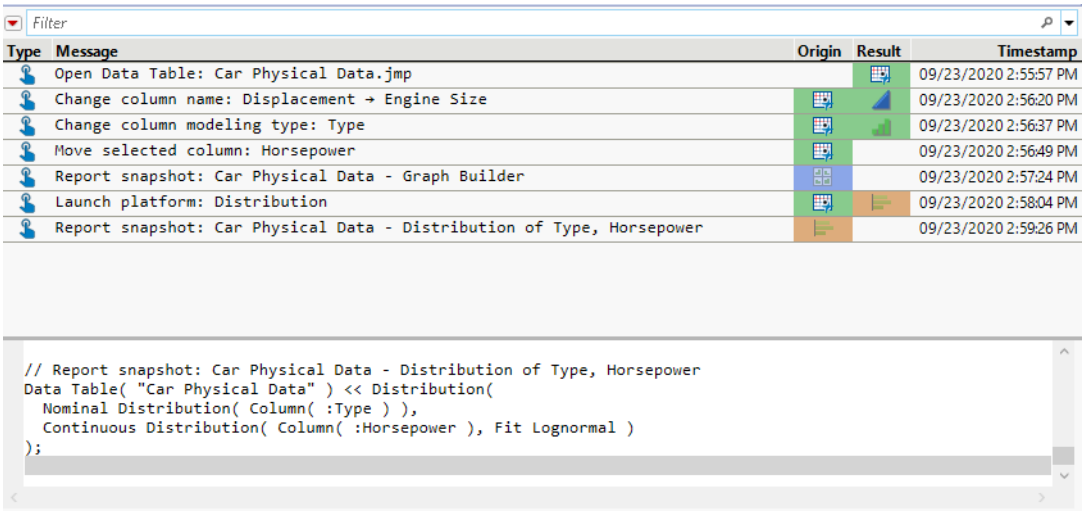
Match Whole Words Returns items that contain each word in the string based on the Match All Terms setting. If you search for "data filter", and Match All Terms is selected, entries that contain both "data" and "filter" are returned.

Example of Recording Actions in the Enhanced Log

You can use the Enhanced Log to record many of the interactive actions that you perform for data cleanup and analysis. In this example, the Enhanced Log is used to generate a JSL script for some data manipulation and analysis actions.

1. Select **View > Log**.
2. Click the red triangle next to the Filter box and select **Clear Log**.
3. Select **Help > Sample Data Library** and open Car Physical Data.jmp.
4. Right-click Displacement in the Columns panel in the data table, and select **Column Info**.
5. Type Engine Size for the Column Name. Click **OK**.
6. Right-click Type in the Columns panel in the data table, and select **Column Info**.
7. Select Ordinal for the Modeling Type. Click **OK**.
8. Click Horsepower in the Columns panel in the data table and drag it above Weight.
9. Select **Graph > Graph Builder**.
10. Drag Horsepower to the **Y** drop zone.
11. Drag Engine Size to the **X** drop zone.
12. Close the Graph Builder window.
13. Select **Analyze > Distribution**.
14. Select Type and Horsepower and click **Y, Columns**.
15. Click **OK**.
16. Click the red triangle next to Horsepower and select **Continuous Fit > Fit Lognormal**.
17. Close the Distribution window.
18. Select **View > Log**.

Figure 4.9 Enhanced Log



19. Press Ctrl and select the first, second, third, and last rows of the log table.
20. Click the red triangle next to the Filter box and select **Save Script > To Script Window**.

The script to reproduce the selected actions appears in a new script window. You can save this script to reuse later or use it as a starting point for a more complex analysis.

Text Log in JMP

The Text Log in JMP is a text-based log of JSL scripts that have been run during a JMP session and the results of those scripts. When the log is open, output appears in the log after you run a script. The script that was run is shaded in the log, and the corresponding output appears beneath it. To view the log in Text mode, select Text for the Mode preference after selecting **File > Preferences > Log**.

Syntax and compatibility errors are reported in the log, including the line number, script filename, and code that JMP could not process. Many interactive JMP actions are also recorded in the log.

Tips:

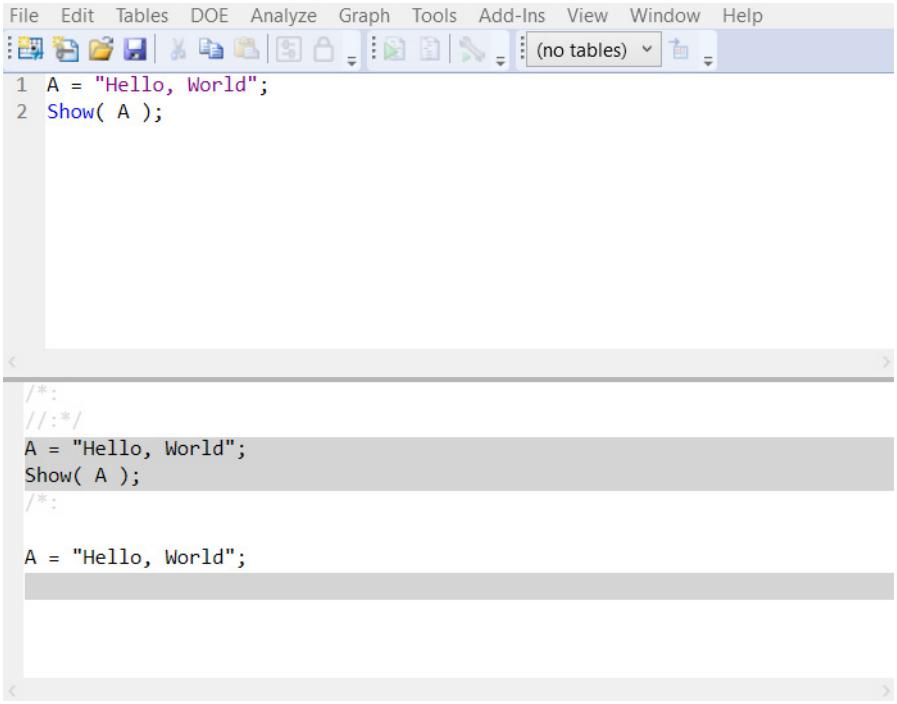
- To omit compatibility warnings from the log, deselect **Show log warnings for JSL compatibility changes in 12** in the JMP General preferences.
- To choose which actions are recorded in the log, you can use the Action Recording preferences after selecting **File > Preferences > Log**.

Show the Text Log in the Script Window

You can view the log inside the script window by right-clicking and selecting **Show Embedded Log**. To view the embedded log by default, select **Show embedded log on script window open** in the Script Editor preferences. This option makes it easy to edit and run a script, quickly see the results of your changes, and then continue to develop the script. The embedded log uses the Text log mode. See "Text Log in JMP".

The embedded log always appears in the Scripting Index script window, but it is not available in Application Builder and the Debugger.

Figure 4.10 The Script Window with an Embedded Text Log



Debug or Profile Scripts

In an open script, click the **Debug Script** button (or select **Edit > Debug Script**) to show the script in the JSL Debugger window. You can also use a keyboard shortcut:

- Press Ctrl+Shift+R (Windows).
- Press Shift+Command+R (macOS).

The JSL Debugger helps identify the point at which a script causes an error or fails. Rather than commenting out portions of the script or adding `Print()` expressions, you can use the Debugger to find the problem.

Once the JSL Debugger appears, you can continue in this mode, or you can click the **Profile JSL Script** button to move into the JSL Profiler mode. The JSL Profiler helps with optimization. You can profile your scripts during execution to see how much time is spent executing particular lines or how many times a particular line is executed.

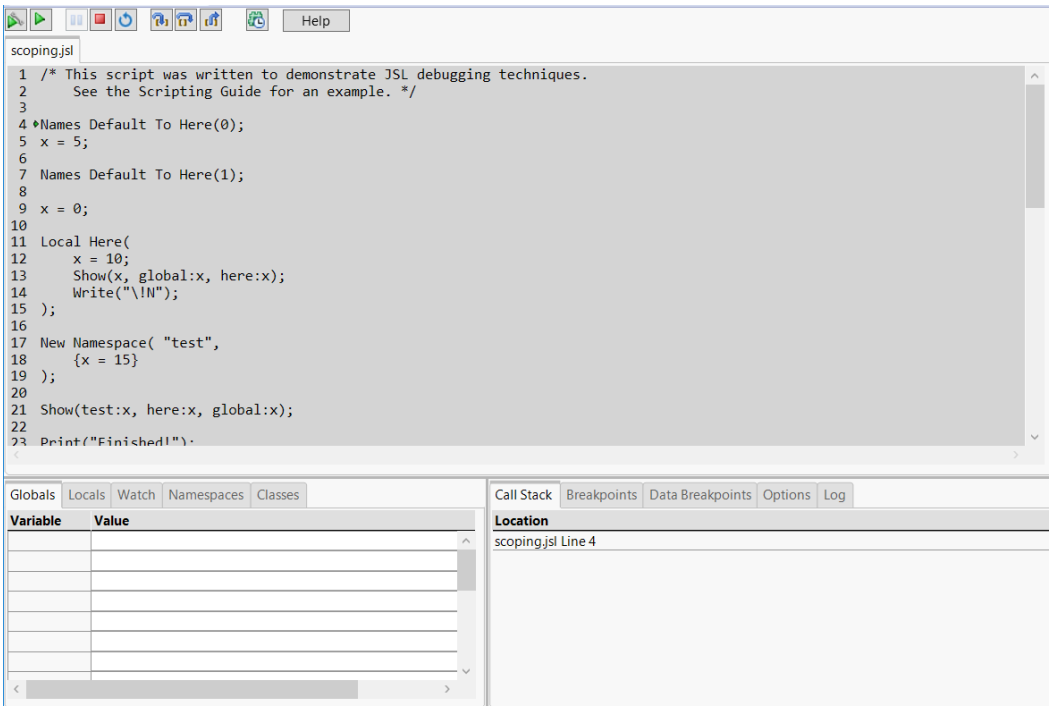
Tip: To debug a script automatically when you open it, include `Debug JSL( 1 )` in the `Open()` statement:

```
Open( "$SAMPLE_SCRIPTS/scoping.jsl", Debug JSL( 1 ) );
```

Debugger and Profiler Window

The Debugger opens in a new instance of JMP (Figure 4.11). The original instance is inoperable until the script produces something that requires interaction. At that point, the Debugger window becomes inoperable until you perform whatever action is required. Then control is returned to the Debugger. Close the Debugger to work again in the original instance of JMP.

Figure 4.11 The Debugger Window



Use the buttons at the top to control the Debugger or the JSL Profiler. One or more scripts that you are debugging or profiling are shown in tabs. If your script includes other scripts, each one opens in a new tab.

Tabs in the bottom portion of the Debugger provide options to view variables, namespaces, the log, and the current execution point; work with breakpoints; and set options.

Tip: To view or edit long values on a tab in the Debugger, right-click the value and select **Edit**. The code opens in a script editor.

Using the Execution Buttons

Use the buttons at the top to control execution of the script within the Debugger or JSL Profiler.

Table 4.1 Description of the Debugger Buttons

Button	Button Name	Action
	Run	Runs the script in the Debugger until it reaches either a breakpoint or the end of the script. **Note:** If there is an error in the script, an abbreviated error message is displayed at the top of the window. Click **More** to see the complete error message.
	Run without breakpoints Run profiler	Runs the script through the end without stopping.
	Break All	If the script is busy, click Break All to stop all action in the script and return to the Debugger or JSL Profiler (for example, if you are in a very long loop). The Debugger or JSL Profiler might not be able to break execution if the executing script is waiting on some interactive user action, such as completing a window or interacting with an opened window.
	Stop	Stops debugging the script and exits the Debugger or the JSL Profiler.
	Restart	Closes the current Debugger session and opens a new session.

Table 4.1 Description of the Debugger Buttons *(Continued)*

Button	Button Name	Action
	Step Into	Lets you step into a function or an included file. Otherwise, it behaves the same as Step Over.
	Step Over	Runs all expressions on a single line, or a complex expression that spans multiple lines, without stepping into a called expression, function, or `Include()` file.
	Step Out	Runs the current script or function to a breakpoint or the end and returns to the calling point. If you are in the main script and the Debugger reaches the end, a message appears: Program execution terminated. The Debugger remains open in order for you to inspect the final program conditions.
	Profile JSL Script	Opens the JSL Profiler. (Press the Run profiler button to start the JSL Profiler.) Use the JSL Profiler to see how much time is spent executing particular lines or how many times a particular line is executed. **Notes:** • You can switch back and forth between the Debugger and JSL Profiler modes only prior to the start of the program. • Some of the debugger buttons are disabled when profiling. • All breakpoints are disabled when running in the JSL Profiler mode.
	Show Profile by Line Count	Shows the number of times each line is executed.
	Show Profile by Time	Shows how much time is spent executing a line.
	Show Profile by Count	For line counts, shows the number of times the line is executed. For time, shows the number of microseconds (or milliseconds or seconds) the line takes to complete.
	Show Profile by Percent	For line counts, shows the individual line count divided by the total line count. For time, shows the percentage of time spent on an individual line (line time/total time*100).
	Time Units	Sets the time unit to microseconds, milliseconds, or seconds. Available in the JSL Profiler after you click the Run profiler button .

Table 4.1 Description of the Debugger Buttons *(Continued)*

Button	Button Name	Action
	Color Theme	Sets the color of the shading for the JSL Profiler. Available in the JSL Profiler after you click the Run profiler button .

Variable Lists

The tabs on the bottom left of the Debugger let you examine global variables, local variables, watch variables, and variables within namespaces.

Globals The Globals tab lists all global variables and updates their values as you step through the script. Each variable is added as it is initialized. If there are already global variables defined from running earlier scripts, they will be listed with their current values when you start the Debugger. See "View Variables".

Locals The Locals tab lists all variables by scope and updates their values as you step through the script. Select a scope in the menu. See "View Variables".

Watch If there is a particular variable or value of an expression whose values you want to watch as you step through the code, you can add them here. This is particularly useful if your script uses many variables that might be difficult to watch in the Globals or Locals lists. See "Work with Watches".

Namespaces As namespaces are defined, they are added to the menu. Select a namespace to view any variables and their values used within the namespace. See "View Variables".

Classes The Classes tab lists classes that are defined in the script. See "Classes".

Debugger Options

The tabs on the bottom right of the Debugger let you view the call stack, work with breakpoints, set options, and view the log.

Call Stack The call stack lists the current execution point in scripts and functions. The main script is always the first script listed. If you call a function, the function is added on top of the calling script. Likewise, any included files are added to the top of the list as you step through them. When you exit any function or script, it is removed from the list and you return to the next one in the list. The current line numbers are updated as you step through.

Double-click a row in the call stack to move the cursor to the specified line.

Breakpoints Add, edit, delete, and disable or enable breakpoints on lines. See "Work with Breakpoints". You can also double-click a row on the Breakpoints tab to move the cursor to the specified line.

Data Breakpoints Add, edit, delete, and disable or enable breakpoints in variables. Data breakpoints watch for changes in variables, unlike breakpoints, which watch for changes in lines.

Options Set the Debugger preferences interactively on this tab. See "Modify Preferences in Debugger".

Log The log from the script that you are debugging is shown on this tab.

Work with Breakpoints

A *breakpoint* interrupts the execution of a script. Although you can step through a script line by line, this can be tedious and lengthy for a long or complex script. You can set breakpoints at places of interest and simply run the script in the Debugger. The script is run normally until a breakpoint is reached. At the breakpoint, the Debugger stops executing the script so that you can look at the values of variables or start stepping line by line.

Breakpoints that appear on the Breakpoints tab in the lower right corner watch for changes in a line. Breakpoints that appear on the Data Breakpoints tab watch for changes in variables.

JMP preserves breakpoints across sessions. So when you close and reopen JMP, the breakpoints still appear.

Tip: Turn on line numbers by right-clicking in the script and selecting **Show Line Numbers**. You can also show line numbers by default in all scripts by modifying the Script Editor preferences.

Create a Breakpoint

When creating a breakpoint, you can specify settings such as conditions and break behavior. To do so, click on the Breakpoints tab or Data Breakpoints tab, and then enter the breakpoint information.

Otherwise, create a quick breakpoint by doing one of the following:

- In the Debugger margin, click the appropriate line (to the right of the line number if displayed).
- In the Debugger margin, right-click in the margin where you want the breakpoint and select **Set Breakpoint**.

The red breakpoint icon appears where you inserted the breakpoint and on the Breakpoints tab.

Delete Breakpoints

Do one of the following:

- In the Debugger margin, click the breakpoint icon.
- In the Debugger margin, right-click the breakpoint icon and select **Clear Breakpoint**.
- On the Breakpoints tab, select the breakpoint and then click .
- On the Breakpoints tab, click to delete all breakpoints (not just the selected breakpoints).

The red breakpoint icon is removed, and the breakpoint no longer appears on the Breakpoints tab.

Disable and Enable Breakpoints

Disabling a breakpoint is helpful when you potentially fix a problem and then want to see whether the script will run correctly past that breakpoint. You can then enable the breakpoint when necessary rather than re-creating it.

Do one of the following:

- In the Debugger margin, right-click the breakpoint icon and select **Enable Breakpoint** or **Disable Breakpoint**.
- On the Breakpoints tab, select or deselect the breakpoint's check box.
- On the Breakpoints tab, click to disable or enable all breakpoints.

A disabled breakpoint turns white; enabled breakpoints are shaded red.

Specify and Clear Conditional Expressions on Breakpoints

Setting a condition on a breakpoint is an alternative to single-stepping through code. Rather than single-step and view the variables for each expression, you specify that the script break only when a condition is met. Then you can step through the code and figure out where the problem arises.

Suppose that a calculation in your script is incorrect, and you suspect the problem occurs when `i==19`. Set a conditional breakpoint for `i==18`. The Debugger will run until that condition is met, then you can step through the code to identify the problem.

Specify a Breakpoint Condition

1. Right-click the breakpoint icon and select **Edit Breakpoint**.
2. On the Condition tab, select **Condition** and enter the conditional expression.
3. Specify whether to break when the expression **Is true** or **Has Changed**.
4. Click **OK**.

Disable or Enable a Condition

1. Right-click the breakpoint icon and select **Edit Breakpoint**.
2. On the Condition tab, deselect or select **Condition**.

Delete a Condition

On the Breakpoints tab, click in the breakpoint's Condition field and press Delete.

Specify Break Options

Right-clicking the breakpoint and selecting **Edit Breakpoint** provides a quick way to manage breakpoint behavior. Alternatively, select the breakpoint on the Breakpoints tab and click . Both methods display the Breakpoint Information window, where you customize settings on the Hit Count and Action tabs.

Change the Hit Count

You can control the number of times a breakpoint must be hit and when the break occurs. For example, to break when the condition is met twice, select **break when the hit count is equal to** and type 2 on the Hit Count tab.

Define an Action

You also have the option of defining a JSL expression or script that the Debugger executes when a breakpoint is hit and execution has stopped. This script is called an *action*. On the Action tab, enter the JSL expression to be executed.

Run the Script to the Cursor

When you right-click and select **Run To Cursor**, all expressions before the location of the cursor are executed. Select this option when you want to see values only up to the current line. To see values when each expression is executed, use the stepping options.

Tips for Setting Breakpoints

- If you do not want to watch for errors in a specific loop, set a breakpoint after the loop ends. The Debugger will hit the next breakpoint rather than stepping through each line of the loop.
- Avoid inserting a breakpoint in lines that do not trigger an action (such as comments, blank lines, and end parentheses). Debugger will not break on these lines.
- When you insert breakpoints, close Debugger, and edit the script, the breakpoints remain on the original line numbers. You might need to delete and then reinsert the breakpoints.

- Breakpoints are remembered across Debugger sessions. This means that your breakpoints list includes breakpoints that have been set in all scripts, not just the script that you are currently debugging.
- Breakpoints are remembered by the Debugger session, not by each script. This means that breakpoints are listed even for scripts that have been moved or deleted.
- On the Breakpoints tab, click to remove all breakpoints in scripts whether they are currently open or not, or for scripts that no longer exist.

View Variables

The variables lists are populated as they appear in the script. Their values are updated every time the script changes them. If you are uncertain why a variable has a particular value when you run your script, you can watch its value at every step to see what happens.

You can also assign the variables whatever value you want. For example, if you are stepping through a `For()` loop but are interested only in what happens starting with a particular iteration, you can assign your iterating variable that value. Step through the first part of the loop that initializes the iteration variable and then assign it the value that you want in the variable list at the bottom. Then when you step through, the loop begins executing at that point.

Tips for Managing Variables

- If you have run several scripts using the global scope, you might want to clear or delete global variables. This makes the list of variables in the Debugger shorter and relevant. Use the `Delete Symbols()` function to do so. You can also close JMP and restart to clear the space.
- If your script uses so many variables that they are difficult to find and watch in the variable lists, add watches for the specific variables in which you are interested.
- To view or edit long values on a tab in the Debugger, right-click the value and select **Edit**. The code opens in a script editor.

Work with Watches

JMP preserves the Watch variables across sessions. So when you close and reopen JMP, the Watch variables are still listed on the Watch tab.

Create a Watch

- On the Watch tab, click and enter the value in the window.

- In the Debugger, right-click the line that you want to watch, select **Add Watch**, and then enter the variable name in the window.
- In the Debugger, place the cursor in or next to a variable name (or select the variable name), right-click, and select **Add Watch**.
- On the Watch tab, enter the variable in an empty **Variable** field.

Modify a Watch

Do one of the following on the Watch tab to enter a new value:

- Select the watch and click .
- Click in the Variable field and enter a new value.

Delete Watches

Do one of the following on the Watch tab to delete watches:

- Select the watch and then click .
- Click to remove all watches.

Modify Preferences in Debugger

The Debugger lets you change preferences as you work in the Debugger. Select the Options tab to find the following settings:

Show Line Numbers Shows or hides the line numbers for the script in the Debugger.

Break on Multiple Statements Per Line Stops executing the script between each expression in a single line.

Break On Throw Breaks when the script executes the `Throw()` function. For example, `Throw()` might be enclosed in a `Try()` expression. The Debugger breaks on `Throw()` instead of continuing through the rest of the expression. This lets you identify where the problem occurred in the script and then return to debugging.

Break On Execution Error Stops executing the script when the error occurs rather than closing the Debugger.

Warn On Assignment In Condition Shows a warning when you enter a breakpoint condition that contains the assignment. For example, if you have a breakpoint on `x = 1` and add the condition `x = 1` to the breakpoint, you are prompted to verify the assignment of `x`.

Enter Debugger Upon Termination Keeps the Debugger open after a JSL program terminates execution. On by default, this option lets you examine attributes of the executed program.

Break For Compatibility Warnings Breaks when a JSL compatibility issue is found.

Persistent Debugger Sessions

JMP saves all breakpoints and watches until you delete them. Other user-specific settings, such as column widths on the tabs and the Debugger window size, persist between sessions of JMP.

These settings are stored in a file named JMP.jdeb, the location of which is defined in the USER_APPDATA variable:

- Windows 7 and higher: "/C:/Users/<username>/AppData/Roaming/SAS/JMP/16/"
- macOS: "/Users/<username>/Library/Application Support/JMP/16/"

As usual, the values of local variables, global variables, and namespaces clear when you close and reopen JMP.

Note: On Windows, the paths differ based on the JMP edition. In JMP Pro, the path refers to "JMPPro".

Examples of Debugging and Profiling Scripts

This section includes examples of setting breakpoints to watch variables; stepping into, over, and out of expressions; watching variables in different scopes and namespaces; debugging interactive scripts; and profiling scripts with the JSL Profiler.

Example scripts are located in the JMP Samples/Scripts folder.

Tip: Make sure that **Show Line Numbers** is selected on the Debugger Options tab before proceeding.

Using Breakpoints and Watching Global Variables

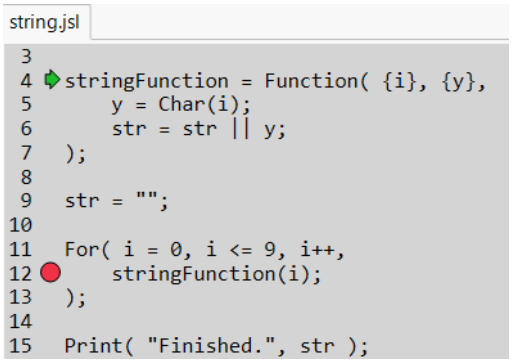
The following example shows how to set a breakpoint in a loop and watch variables change through each iteration of the loop.

1. Open the string.jsl sample script and click the **Debug Script** button.
2. Click in the margin for line 12 to add a breakpoint.

   You should have a breakpoint for the following expression inside the For() loop:

   ```
   stringFunction(i);
   ```

Figure 4.12 Set the Breakpoint



3. Click **Run**.

   The first two expressions are evaluated:

   – `stringFunction` is defined as a function.

   – `str` is defined as an empty string.

   Both variables and their types and values have been added to the Globals list. In addition, the `For()` loop has been evaluated up to the line with the breakpoint, shown in Figure 4.12.

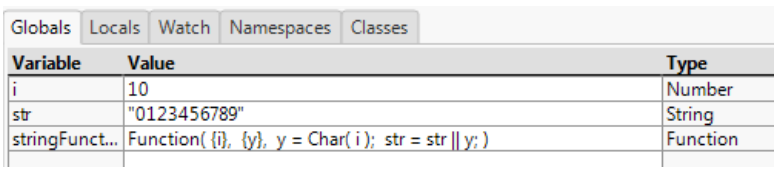
   – `i` has been assigned to 0.

   – `i` and its value and type have been added to the Globals list.

   – `i` has been determined to be less than or equal to 9.

   – `stringFunction()` has not yet been called.

Figure 4.13 View the Initial Global Variables



Tip: The initial value of `i` is 10 because the script was run before entering the Debugger; it's showing the value of `i` after running the script (10). Start a new instance of JMP, open the script without running it, and enter the Debugger immediately.

4. Click **Run** again.

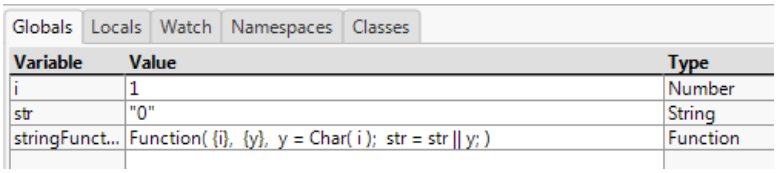
   The script runs until it hits the breakpoint. The results are shown in Figure 4.14.

   – `stringFunction()` is called, evaluated, and returns to the loop.

   – `i` is incremented and determined to be less than or equal to 9.

– In the Globals list, `i` is now 1 and `str` is now "0".

Figure 4.14 Global Variables at First Breakpoint



5. Click **Run** again.

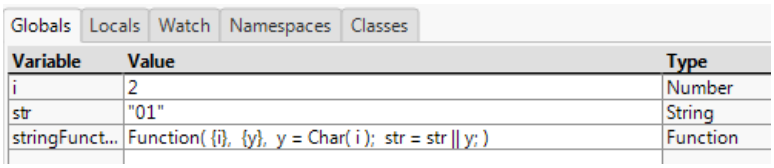
   The script runs until it hits a breakpoint. The results are shown in Figure 4.15.

   – `stringFunction()` is called, evaluated, and returns to the loop.

   – `i` is incremented and determined to be less than or equal to 9.

   In the Globals list, `i` is now 2, and `str` is now "01".

Figure 4.15 Global Variables at Second Breakpoint



You can continue to click **Run** and watch `i` and `str` change with each iteration of the loop. Or, click **Run without breakpoints** to complete running the script and exit the Debugger.

Stepping Into, Over, and Out

Step Into, **Step Over**, and **Step Out** offer flexibility when your script contains expressions, functions, or includes other JSL files.

1. Open the scriptDriver.jsl sample script and click the **Debug Script** button.
2. This script writes information to the log, so select the Log tab at the bottom of the Debugger to view the messages.
3. Click **Step Over**.

   The first line in the script is evaluated.

4. Click **Step Over** again.

   The current expression is evaluated, and the Debugger moves to the following line. In this case, the expression is a few lines long, and it assigns an expression to a variable.

5. Click **Step Over** again.

   This expression is several lines long, and assigns a function to a variable.

Line 30 calls the expression that was created earlier.

6. Click **Step Over**.

   The Debugger steps into the expression, running it line by line.

7. Continue clicking **Step Over** until the expression ends.

   The Debugger returns to the line following the expression call.

   Line 31 calls the function defined earlier.

8. Click **Step Over** to run the function without stepping into it. The Debugger runs the entire function, and returns to the line following the function call.

   Line 33 includes another script.

9. Click **Step Into**.

   The Debugger opens the script in another tab and waits.

10. Click **Step Over**.

    The next line in the included script is run.

11. Click **Step Out**.

    The Debugger runs the rest of the included script and returns to the line following the `Include()` function.

Watching Variables in Different Scopes and Namespaces

Tabs at the bottom of the Debugger window let you watch variables as they are created and changed. This example shows variables in several scopes and a namespace.

1. Open the scoping.jsl sample script and click the **Debug Script** button.
2. Click **Step Over**.

   The fourth line turns off `Names Default To Here`. If you run this script again in the same JMP session, this line resets the scoping so that the first variable that is created is in the global scope.

3. Click **Step Over**.

   A global variable named *x* is created. On the Globals tab, *x* has been added to the list, showing its value as 5 and its type as number.

4. Select the **Locals** tab, and then select **Global** from the list of scopes.

   The global variable *x* is also shown here.

5. Click **Step Over** twice.

   `Names Default To Here` is turned on, which places the rest of the script into a Here scope. Then a new variable *x* is created in that scope.

   Notice that the value of the global variable *x* has not changed.

6. Select **Here** from the list on the Locals tab.

   The local *x* is listed under **Here**, with its value and type.

7. Click **Step Over**.

   A `Local Here` scope is created. A second Here scope is shown in the Locals list.

8. Click **Step Over**.

   A new *x* variable is created in this Here scope. On the Locals tab, select each of the three scopes from the list (**Here**, **Here**, and **Global**) to see three different *x* variables.

9. Click **Step Over**.

   Look in the Debugger's log to see the output. Notice that *here:x* scopes to the local here, not the script window's here.

10. Click **Step Over**.

    After writing an empty line to the log, the script exits the Local Here scope. The second Here, along with its' *x* variable, has disappeared from the Locals list.

11. Click **Step Over**.

    A namespace called "test" is created, with another variable named *x*. Select the **Namespaces** tab to see it.

12. Click **Step Over** and look at the log.

13. Click **Step Over** to exit the Debugger.

Using the Debugger with Interactive Scripts

When your script creates interactive elements, the Debugger hands control back to the main instance of JMP so that you can interact with it. When you are finished, control returns to the Debugger.

1. Open the interactive.jsl sample script and click the **Debug Script** button.
2. Click **Step Over** twice.

   The `New Window` expression is evaluated, and a modal window waiting for input is created. You might need to move the Debugger window to see the new modal window.

3. Enter two numbers in the Assign X and Y window and click **OK**.

   Control is given back to the Debugger.

4. Click **Step Over** three times and look at the log in the Debugger.

   The log shows the two new numbers that you entered in the window.

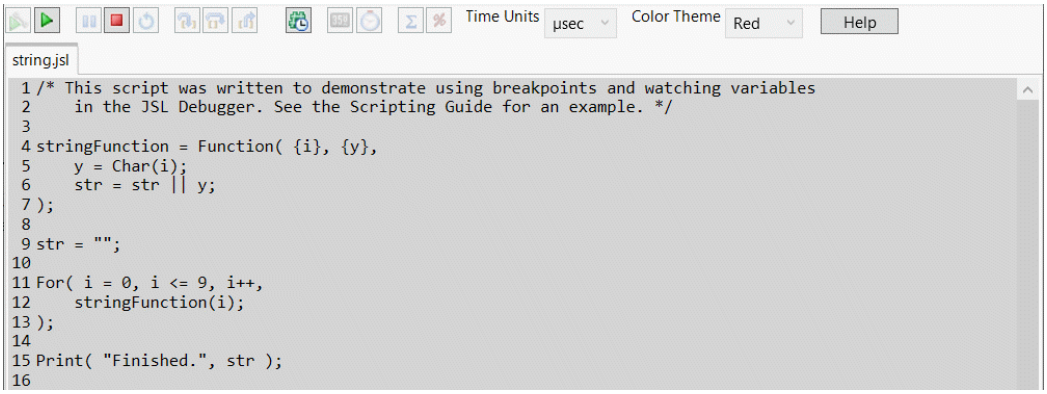
5. Click **Step Into** to exit the Debugger.

Using the JSL Profiler

Use the JSL Profiler to see how much time is spent executing particular lines or how many times a particular line is executed.

1. Open the string.jsl sample script.
2. Click the **Debug Script** button .
3. Click the **Profile JSL Script** button .

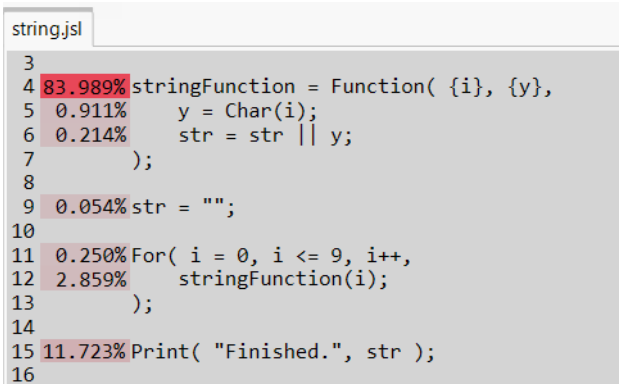
Figure 4.16 Initial JSL Profiler Window



4. Click the **Run** button to start profiling.

   The profiler collects information about the number of times a statement is executed and the time it takes to execute it. Time is cumulative and collected each time a JSL statement is executed.

Figure 4.17 Profiled Script Window



In the left margin, the selected statistics are displayed. Percent of time is displayed by default. Click the **Show Profile by Count** button to switch to percent of statement counts instead. The left margin is color-coded to allow for quick identification of problematic performance areas.

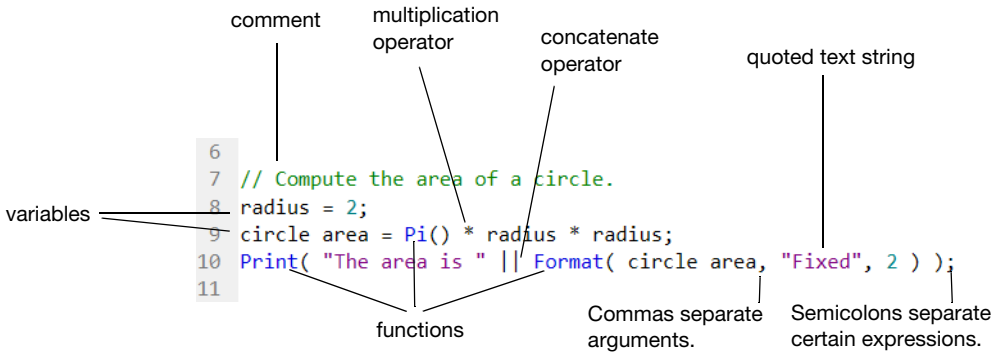
Chapter 5

JSL Building Blocks

Learning the Basics of JSL

Studying the syntax and basics of JSL is crucial, whether you are a beginning or an advanced user. Some concepts (such as loops and variables) are common among many scripting languages, but punctuation rules differ in JSL.

Figure 5.1 Example of a JSL Script



This chapter introduces you to the basic concepts of JSL, from syntax rules and file path conventions to conditions and namespaces.

Contents

JSL Syntax Rules . . . 100
Value Separators . . . 100
Numbers . . . 103
Names . . . 103
Comments . . . 104
Operators . . . 105
Boolean Values . . . 109
Global and Local Variables . . . 110
Local Namespaces . . . 110
Named Namespaces . . . 111
Show Symbols, Clear Symbols, and Delete Symbols . . . 111
Lock and Unlock Symbols . . . 112
Hide a Global Variable . . . 113
Rules for Name Resolution . . . 113
Variable Names Used as Arguments . . . 114
Rules for Resolving Names . . . 114
Scoping Operators . . . 116
Troubleshoot Variables and Column Names . . . 119
Troubleshoot Variables and Keywords . . . 120
Frequently Asked Questions about Name Resolution . . . 120
Alternatives for Gluing Expressions Together . . . 121
Iterate . . . 122
For . . . 122
While . . . 123
Summation . . . 125
Product . . . 125
Break and Continue . . . 126
Conditional Functions . . . 128
If . . . 128
Match . . . 130
Choose . . . 132
Interpolate . . . 133
Step . . . 135
Compare Incomplete or Mismatched Data . . . 135
Missing Values . . . 137
Inquiry Functions . . . 138
General Element Types . . . 139
Specific Element Types . . . 139
Object Attributes . . . 141

Host Information 141
Version Information 141

JSL Syntax Rules

All scripting and programming languages have their own syntax rules. JSL looks familiar if you have programmed in languages such as C and Java. However, rules for punctuation and spaces differ in JSL.

The following sections describe JSL syntax rules for the following basic components of the language.

Value Separators

Words in JSL are separated by parentheses, commas, semicolons, spaces, and various operators (such as +, –, and so on). This section describes the rules for using these separators and delimiters in JSL.

Commas

A comma separates items, such as items in a list, rows in a matrix, or arguments to a function.

```
my list = {1, 2, 3};
your list = List( 4, 5, 6) ;
my matrix = [3 2 1, 0 -1 -2];
If( Y < 20, X = Y );
Table Box( String Col Box( "Age", a ) );
```

Note: To glue a sequence of commands into a single argument inside a function, separate each sequence with a semicolon. See "Semicolons".

Parentheses

Parentheses have several purposes in JSL:

- Parentheses group operations in an expression. The following parentheses group the operations in the `If()` expression.

  ```
  y = 10;
  If(
     Y < 20, X = Y,
     X = 20
  );
  ```

- Parentheses delimit arguments to a function. In the following example, parentheses enclose the argument to the `Open()` function.

  ```
  Open( "$SAMPLE_DATA/Big Class.jmp" );
  ```

- Parentheses also mark the end of a function name, even when arguments are not needed. For example, the `Pi` function has no arguments. However, the parentheses are required so that JMP can identify *Pi* as a function.

  ```
  Pi();
  ```

Note: Be careful that parentheses match. Every `(` needs a `)`, or errors result.

The script editor can match fences (parentheses, brackets, and braces). Press Ctrl+] (Command+B on macOS) with your cursor in any part of a script. The editor searches for fences, highlighting the text between the first set of opening and closing fences that it finds. Repeat this process to highlight the next-higher fence. See "Match Parentheses, Brackets, and Braces" for an example.

Semicolons

Expressions separated by a semicolon are evaluated in succession, returning the result of the last expression. In the following code, 0 is assigned to the variable `i` and then 2 is assigned the variable `j`.

```
i = 0;
j = 2;
```

You can also use semicolons to join arguments that are separated by commas as shown in the following `If()` expression.

```
If( x < 5, y = 3; z++; );
```

The semicolon in other languages is used as an expression terminator character. In JSL, the semicolon is a signal to continue because more commands might follow. For more information about separating expressions with semicolons, see "Alternatives for Gluing Expressions Together".

Semicolons at the end of a script or at the end of a line of arguments are harmless, so you can also think of semicolons as terminating characters. Trailing semicolons are allowed at the end of a script stream and after a closing parenthesis or closing curly brace. In fact, terminating each complete JSL expression with a semicolon helps avoid errors when you copy and paste small scripts into a larger one.

The semicolon is equivalent to the `Glue()` function. See "Operators" for more information about semicolons and `Glue()`.

Double Quotes

Double quotes enclose text strings. Anything inside double quotes is taken literally, including spaces and upper- or lower-case; nothing is evaluated. If you have `"Pi() ^ 2"` (inside double quotes), it is just a sequence of characters, not a value close to ten.

You do have to be a careful with text strings. Extra spaces and punctuation do affect the output, because text strings inside double quotes are used literally, exactly as you enter them.

To have double quotes inside a quoted string, precede each quotation mark with the escape sequence \! (backslash-bang). For example, run the following script and look at the title of the window:

```
New Window( "\!"Hello\!" is a quoted string.",
  Text Box( Char( Repeat( "*", 70 ) ) )
);
```

Table 5.1 Escape Sequences for Quoted Strings

\!b	blank
\!t	tab
\!r	carriage return only
\!n	linefeed (newline) only
\!N	inserts line breaking characters appropriate for the host environment[a]
\!f	formfeed (page break)
\!0	null character **Note:** The null character is dangerous to use, because it is typically treated as the end of the string. Be sure to type the number zero, not the letter O.
\!\	backslash
\!"	double quotation mark

a. On macOS, this escape sequence is CR (carriage return character, hexadecimal '0D'). On Windows, this sequence is CR LF (carriage return followed by a linefeed, hexadecimal '0D0A').

Sometimes, long passages require a lot of escaped characters. In these cases, use the notation \[...]\ and everything between the brackets does not need or support escape sequences. Here is an example where \[...]\ is used inside a double-quoted string.

```
jslPhrase = "The JSL to do this is :\[
  a = "hello";
  b = a|| " world.";
  Show(b);
  ]\ and you use the Submit command to run it.";
```

Spaces

JSL allows whitespace characters inside names; spaces, tabs, returns, and blank lines inside or between JSL words are ignored. This is because most JSL words come from the user interface, and most of those commands have spaces in them. For example, the JSL expression for the Fit Model platform is `Fit Model()` or `FitModel()`.

Spaces inside an operator or between digits in a single number are not allowed. In these cases, the script generates errors. For example, you cannot put spaces between the two plus signs in `i++` (`i+ +`) or in numbers (`4 3` is not the same as `43`).

Note: Why does JSL allow whitespace characters inside names? For one reason, the names of commands and options match the equivalent commands in JMP menus and windows. Another reason is that data table column names often include spaces.

Numbers

Numbers can be written as integers, decimal numbers, dates, times, or date-time values. They can also be included in scientific notations with an `E` preceding the power of ten. For example, these are all numbers:

```
.   1   12   1.234  3E3  0.314159265E+1 1E-20 01JAN98
```

Note: A single period by itself is considered a missing numeric value (sometimes called *NaN* for "not a number").

For more information about dates, times, and date-time values, see "Date-Time Functions and Formats". See "Currency" for more information about combining numbers with currency symbols.

Names

A name is simply something to call an item. When you assign the numeric value 3 to a variable in the expression `a = 3`, `a` is a name.

Commands and functions have names, too. In the expression `Log( 4 )`, `Log` is the name of the logarithmic function.

Names have a few rules:

- Names *must* start with an alphabetic character or underscore and can continue with the following:
  - alphabetic characters (a-z A-Z)
  - numeric digits (0-9)

  - whitespace characters (spaces, tabs, line endings, and page endings)
  - mathematical symbols in Unicode (such as the Greek small alpha letter α)
  - a few punctuation marks or special characters (apostrophes (‘), percent signs (%), periods (.), backslashes (\), and underscores (_))
- When comparing names, JMP ignores whitespace characters (such as spaces, tabs, and line endings). Upper case and lower case characters are not distinguished. For example, the names `Forage` and `for age` are equivalent, despite the differences in white space and case.

You can still have a name that is any other sequence of characters. If the name does not follow the rules above, it needs to be quoted and `followed by "n"`. For example, to use a global variable with the name `taxable income(2011)`, you must use `"taxable income(2011)"n` every time the variable appears in a script:

```
"taxable income( 2011 )"n = 456000;
tax = .25;
Print( tax * "taxable income( 2011 )"n) ;
   114000
```

Note: Previously, the `Name()` parser directive wrapped the column name that contained illegal characters. `Name()` is deprecated.

For more information about how JMP interprets names, see “Rules for Name Resolution”.

Comments

Comments are notes in the code that are ignored by the JSL processor (or *parser*). You include comments to describe sections of the script. Comments are also convenient for removing portions of a script temporarily. For example, you can insert comment symbols around code that might be causing an error and then rerun the script.

Type the comment symbols around code that you want to comment. The following example shows code commented with `/* */` in the middle of a line. When the script is run, JMP considers both expressions to be identical.

```
tax /*percentage*/ = .25;
tax = .25;
```

Table 5.2 describes the comment symbols.

Table 5.2 Comment Symbols

Symbol	Syntax	Explanation
`//`	`// comment`	Begins a comment. The comment does not have to be at the beginning of a line, but everything up to the end of the line is a comment.
`/* */`	`/* comment */`	Begins and ends a comment. This comment can appear in the middle of a line. Script text before and after the comment is parsed normally.
`//!`	`//!`	Add `//!` to the first line of the script, and the script runs automatically when opened in JMP. (In other words, the script editor does not open.)

Operators

An operator is a one- or two-character symbol used to perform common arithmetic actions, compare values, construct lists, subscript into a data element, send messages, concatenate lists, scope names, glue expressions, and end expressions.

Arithmetic operators come in several varieties:

- *infix* (with arguments on either side, such as + in `3 + 4`, or = in `a = 7`)
- *prefix* (with one argument on its right side, such as `!a` for logical negation)
- or *postfix* (with one argument on its left side, such as `a++` for incrementing `a`

To make writing expressions easier, JSL uses certain special character operators that are alternatives to using functions. These operators have the same meaning as if the phrase had been written as a function. For example, the following two expressions are equivalent.

```
Net Income After Taxes = Net Income - Taxes;
Assign( Net Income After Taxes, Subtract( Net Income, Taxes ) );
```

The assignment operation can be written either as the `Assign()` function or as an infix operator =. Similarly, subtraction can be done with the `Subtract()` function or the infix minus sign; they are equivalent inside JMP.

Note: Usually white space is ignored in JSL, so that "netincomeaftertaxes" and "net income after taxes" are the same thing. There is one exception: the two-character operators must *not* have a space between characters. Always enter the following operators without spaces in between:

```
||, |/, <=, >=, !=, ==, +=, -=, *=, /=,  ++, --, <<, ::, :*, :/, /*, */
```

Another common operator is the semicolon (`;`). You use the semicolon to:

- Separate yet join one expression to another in a programming sequence. The semicolon returns the result of its last argument, so `a;b` is the same as `Glue(a,b)`.
- End an expression. Though the semicolon is permitted at the end of an expression, it is not the expression terminator used in other languages.

An expression can contain more than one operator. In these instances, the operators are grouped in order of precedence with decreasing priority. For example, * takes precedence over +:

```
a + b * c
```

So `b * c` is evaluated first, and then the result is added to `a`.

+ takes precedence over -:

```
a + b * c - d
```

So `b * c` is evaluated, and then the result is added to `a`. `d` is then subtracted from the result of `a + b * c`.

Table 5.3 shows operators shaded in order of precedence and each operator's function equivalent.

Table 5.3 Operators and Their Function Equivalents in Order of Precedence

Operator		Function Syntax	Explanation
`{ }`	`List`	`{a,b}` `List(a,b)`	Construct a list.
`[ ]`	`Subscript`	`a[b,c]` `Subscript(a,b,c)`	Subscripts identify specific elements within a data element *a*, where *a* could be a list, a matrix, a data column, a platform object, a display box, and so on.

Table 5.3 Operators and Their Function Equivalents in Order of Precedence *(Continued)*

Operator		Function Syntax	Explanation
++	Post Increment	*a*++ Post Increment(*a*)	Adds one (1) to *a*, in place. JMP does not have a pre-increment operator. Instead, use the Add To() operator, which is (+=).
--	Post Decrement	*a*-- Post Decrement(*a*)	Subtracts one (1) from *a*, in place. JMP does not have a pre-decrement operator. Instead, use the Subtract To() operator, which is (-=).
^	Power	*a*^*b* Power(*a*,*b*) Power(*x*)	Raise *a* to exponent power *b*. With only one argument, 2 is assumed as the power, so Power(*x*) computes x^2.
-	Minus	-*a* Minus(*a*)	Reverses sign of *a*.
!	Not	!*a* Not(*a*)	Logical Not. Maps nonzero (or true) values to 0 (which means false). Maps 0 (or false) values to 1 (which means true).
*	Multiply	*a**b* Multiply(*a*,*b*)	Multiplies *a* by *b*.
:*	EMult	*a*:**b* EMult(*a*,*b*)	Elementwise multiplication for matrices *a* and *b*. (Each element in matrix *a* is multiplied by each element in matrix *b*.)
/	Divide	*a*/*b* Divide(*a*,*b*) Divide(*x*)	Divide(*a*, *b*) divides *a* by *b*. Divide(*x*) interprets the argument as a denominator and implies 1 as the numerator, yielding the reciprocal 1/*x*.
:/	EDiv	*a*:/*b* EDiv(*a*,*b*)	Elementwise division for matrices *a* and *b*. (Each element in matrix *a* is divided by each element in matrix *b*.)

Table 5.3 Operators and Their Function Equivalents in Order of Precedence *(Continued)*

Operator		Function Syntax	Explanation
+	Add	*a*+*b* Add(*a*,*b*)	Adds *a* and *b*.
–	Subtract	*a*–*b* Subtract(*a*,*b*)	Subtracts *b* from *a*.
||	Concat	*a*||*b* Concat(*a*,*b*)	Joins two or more strings; two or more lists; and horizontally concatenates matrices. See "Concatenate Lists" or the *JSL Syntax Reference*.
|/	VConcat	*matrix1*|/*matrix2* VConcat(*matrix1*, *matrix2*)	Vertically concatenate matrices. (Use || or Concat() to horizontally concatenate matrices.)
::	Index	*a*::*b* Index(*a*,*b*)	For matrices, generates the integers from *a* to *b*. (Colons are also used as prefix operators for scoping, where :a means data table column a, and ::a means JSL global variable *a*. See "Scoping Operators".)
<<	Send	object << message Send(*object*, *message*)	Send *message* to *object*.
==	Equal	*a*==*b* Equal(*a*,*b*)...	Boolean values for comparisons. They all return 1 if true, 0 if false. Missing values in either *a* or *b* causes a return value of missing, which evaluates as neither true nor false. See "Missing Values", for treatment of missing values.
!=	Not Equal	*a*!=*b* Not Equal(*a*,*b*)...	
<	Less	*a*<*b* Less(*a*,*b*)...	
<=	Less or Equal	*a*<=*b* Less or Equal(*a*,*b*)	
>	Greater	*a*>*b* Greater(*a*,*b*)	
>=	Greater or Equal	*a*>=*b* Greater or Equal(*a*,*b*)	

Table 5.3 Operators and Their Function Equivalents in Order of Precedence *(Continued)*

Operator		Function Syntax	Explanation
<=, <	Less Equal Less	*a*<=*b*<*c* Less Equal Less(*a*,*b*,*c*)	Range check. Return 1 if true, 0 if false. Missing values in either *a* or *b* propagate missing values.
<, <=	Less Less Equal	*a*<*b*<=*c* Less Less Equal(*a*,*b*,*c*)	
&	And	*a*&*b* And(*a*,*b*)	Logical And. Returns true if both are true. If the value on the left is false, the value on the right is not evaluated. See "Missing Values", for treatment of missing values.
|	Or	*a*|*b* Or(*a*,*b*)	Logical Or. Returns true if either or both are true. See "Missing Values", for treatment of missing values.
=	Assign	*a*=*b* Assign(*a*,*b*)	Put the value of *b* into *a*. Replaces the current value of *a*.
+=	Add To	*a*+=*b* AddTo(*a*,*b*)	Add the value of *b* into *a*.
-=	Subtract To	*a*-=*b* SubtractTo(*a*,*b*)	Subtract *b* from *a*, and put back into *a*.
*=	Multiply To	*a**=*b* MultiplyTo(*a*,*b*)	Multiply *b* with *a*, and put back into *a*.
/=	Divide To	*a*/=*b* DivideTo(*a*,*b*)	Divide *b* into *a*, and put back into *a*.
;	Glue	*a*;*b* Glue(*expr*, *expr*, ...)	First do *a*, and then do *b*.

Boolean Values

A Boolean value enables or disables an option. To enable the option, use `1`, `true`, or `yes` as the argument.

```
obj << Unequal Variances( 1 );
obj << Unequal Variances( "true" );
obj << Unequal Variances( "yes" );
```

Note: Quoting the value prevents it from being evaluated as a variable.

Omitting the argument also enables the option. This is the default for all but row state messages.

```
obj << Unequal Variances;
obj << Unequal Variances();
```

To disable the option, use `0`, `false`, or `no` as the argument.

To toggle an option (turn off an enabled option or turn on a disabled option), use `"toggle".`

Notes:

- The quoted `"present"`, `"absent"`, `"on"`, `"off"`, `"switch"`, and `"flip"` are supported for backwards compatibility.
- For row state messages, an empty argument always toggles the option.

Global and Local Variables

Variables are names that hold values, which you reference later in scripts. There are two types of variables:

- *Global variables* are shared among all scripts that you run in a JMP session.
- *Local variables* apply only to the script context in which you define them. They can also be local to only a piece of a script, as with variables local to a particular function.

To limit the scope of variables, you can define them in a *namespace*, which is a collection of variables, functions, and other unique names. JMP has a single global variable namespace that all scripts use by default. When you use a name plainly, without a qualifying syntax, the name is an *unscoped* variable and therefore in the global namespace.

```
x = 1;
```

Local Namespaces

Putting variables in the global namespace can cause conflicts. When two scripts have variables with the same names, the value of the variable in the last script that you run last overwrites the variable's value in the first script.

To prevent this problem, we recommend that you begin each script with the following line:

```
Names Default To Here( 1 );
```

The `Names Default To Here( 1 );` function makes all unscoped variables in the script local to that script and does not affect the global variable namespace. See "Advanced Scoping and Namespaces".

Notes:

- The `Names Default to Here` setting is true by default for custom menus and toolbar buttons. A script that runs when you select a custom menu item or click a custom toolbar button does not affect global variables.
- The `Names Default to Here` setting is false by default for table scripts. If a table script specifies `Names Default to Here(1)`, this setting lasts only as long as the table script runs. After the script finishes running, the setting is reset to whatever it was before the script ran.

Named Namespaces

You can also create a variable in a specific namespace. In the following example, the `x` variable is created in the `aa` namespace:

```
aa:x = 1;
```

Preceding local variables with the `Local()` function is another option. Both `a` and `b` are local variables in the following expression:

```
Local( {a = 1, b}, ... );
```

Scoping operators also distinguish a global variable from a local variable. See "Rules for Name Resolution".

The following sections describe functions that help you manage variables.

Show Symbols, Clear Symbols, and Delete Symbols

The `Show Symbols()` function lists all variables and namespaces that are defined both globally and in the local script, along with their current values. Here is an example of `Show Symbols()` messages that are shown in the log:

```
Show Symbols();
   // Here
   a = 5;
   b = 6;
   // 2 Here

   // Global
   c = 10;
   // 1 Global
```

Tip: The JSL debugger also shows you the values of variables and namespaces. See "Debug or Profile Scripts".

The `Clear Symbols()` function erases the values set for variables that are defined both globally and in the local script. For example, after you clear and then show symbols, the variables are empty.

```
Clear Symbols();
Show Symbols();
   // Here
   a = Empty;
   b = Empty;
   // 2 Here

   // Global
   c = Empty;
   // 1 Global
```

Note: The older `Show Globals()` and `Clear Globals()` functions are aliases of the newer `Show Symbols()` and `Clear Symbols()` functions.

To remove *all* global variables and namespaces, use the function `Delete Symbols()`. After the last `Show Symbols()` in the following script is run, nothing shows up in the log. All variables have been completely removed from memory.

```
Delete Symbols();
Show Symbols();
```

To list variables in all namespaces, use `Show Namespaces()`. To delete only a specific namespace, use `ns << Delete. Clear Symbols()` and `Delete Symbols()` do not clear or delete variables in each namespace, although they do clear and delete variables that contain references to namespaces. See "Rules for Name Resolution" for more information about unscoped variables.

Note: `Clear Symbols()` and `Delete Symbols()` break all scripts that are currently in use. These functions can be very useful in a programming and debugging environment, but do not include them in any script that you plan to distribute. If you include `Names Default To Here(1)` in your scripts, clearing and deleting global symbols is unnecessary.

Lock and Unlock Symbols

If you want to lock a variable to prevent it from being changed, use the `Lock Symbols()` function. (`Lock Globals()` is an alias.)

```
Lock Symbols( name1, name2, ... );
```

To release the lock and enable the global to be changed, use the `Unlock Symbols()` function. (`Unlock Globals()` is an alias.)

```
Unlock Symbols( name1, name2, ... );
```

The primary use of these two commands is to prevent inadvertent changes to variables. For example, locking a variable prevents `Clear Symbols`() from clearing a variable that is being used by another script.

Note: You cannot use `Lock Symbols()` to lock a namespace. Instead, use `ns <<Lock Namespace.`

Hide a Global Variable

To hide (or protect) a global variable, put two underscores (__) before the name. Protecting a global variable means that it will be hidden and can not be examined or shown. However, the behavior is different depending upon the context.

This example will not print the variable because it is protected:

```
Show(::__xyz);
```

This example returns a NULL value for the protected variable:

```
::a = Name Expr( ::__xyz );
```

This example does return a value for the protected variable:

```
Show( Eval( ::__xyz ) );
```

Rules for Name Resolution

In JMP, you identify the following types of objects by a name:

- Columns and table variables in a data table
- Global variables, which hold values for the remainder of a JMP session
- Scriptable object types
- Arguments and local variables inside formulas

Most of the time, you can just use an object's name directly to refer to the object. Consider the following example:

```
ratio = height / weight;
```

Depending on the complexity of your script, it might be obvious that `ratio` is a variable and `height` and `weight` are data table column names. But what if the meanings are ambiguous? A script might use `ratio` as a global variable and as column names.

Variable Names Used as Arguments

Message arguments are evaluated as a rule. In previous versions of JMP, some messages interpreted a variable name as a string value. For example, the following script results in a linear scale in the current version of JMP and a log scale in JMP 11.

```
Log = "Linear";
axis box << Scale( Log );
```

The syntax `Scale( "Log" )` is preferred for setting a string literal. To catch ambiguous instances, set the "Allow Unquoted Strings in JSL" General preference to "Yes (with a warning)" or "No".

Rules for Resolving Names

JMP interprets object names using *name resolution*. The following rules are applied sequentially to unscoped names:

1. If the variable is followed by a pair of parentheses `( )`, look it up as a function.
2. If the variable is prefixed by `:` scope operator or an explicit data table reference, look it up as a data table column or table variable.
3. If the variable is prefixed by `::` scope operator, look it up as a global variable.
4. If the variable is an explicit scope reference (such as `group:vowel`), look it up in the user-defined `group` namespace.
5. If the variable is in a `Local` or `Parameter` function, look it up as a local variable. If it is nested, repeat until a function call boundary is found.
6. If the variable is in a user-defined function or method, look it up as a function or method argument or local variable.
7. Look the variable up in the current scope and its parent scope. Repeat until the `Here` scope is encountered.
8. Look the variable up as a variable in the `Here` scope.
9. Look the variable up as a global variable.
10. If `Names Default to Here(1)` is at the top of the script, stop looking. The scope is local.
11. Look the variable up as a data table column or table variable. Columns are preferred over table variables.

12. Look the variable up as an operator or a platform launch name (for example, Distribution, Bivariate, Chart, and so on).
13. When the name cannot be found:
    - If the name is used in a reference, print and error to the log.
    - If the name is used as the target of an assignment (as an L-value), test the following:

      If the variable is preceded by `::` scope operator, create and use a global variable.

      If the variable is an explicit scope reference, create and use the variable in the specified namespace or scope.

      If `Names Default to Here(0)` is at the top of the script, create a global variable.

      If `Names Default to Here(1)` is at the top of the script, create a `Here` namespace variable.

Exceptions

- Some names are variables that refer to an object such as a data table, data column, or platform; they are not used for getting or setting a value. These names are passed through (or interpreted literally) rather than resolved.
- For function definitions, column formulas, and Nonlinear platform formulas, the scope is the same for each row in a column.
- If a name is a direct reference to a column in a data table that has been closed, the name is resolved again to that column when the table is reopened.

  ```
  dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
  Show( :weight << Get As Matrix ); // weight  resolves to a column name
  Close( dt, NoSave );
  Show( :weight << Get As Matrix ); // weight  cannot be resolved
  dt = Open( "$SAMPLE_DATA/Big Class.jmp" ); // reopen data table
  Show( :weight << Get As Matrix ); // weight  resolves to a column name
  ```

  However, the following example does not resolve the variable to the second instance of the data table:

  ```
  dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
  col = Column( dt, 5); // col is Column( "weight" );
  Close( dt, NoSave );
  dt = Open( "$SAMPLE_DATA/Big Class.jmp" ); // reopen data table
  Show( col << Get As Matrix ); // the reference to the first data table no
    longer exists
  ```

The following sections describe how JMP resolves the names of data table columns. For more information about name resolution, see “Advanced Scoping and Namespaces”.

Scoping Operators

Using *scoping operators* is an easy way to help JMP resolve ambiguous names (for example, when a name refers to both a variable and a column name).

In the following example, the prefix double-colon operator (`::`) identifies `z` as a global variable. The single-colon prefix operator (`:`) identifies `x` and `y` as column names.

```
::z = :x + :y;
```

Tip: The `Names Default to Here(1)` function also affects name resolution. See "Names Default To Here and Global Variables".

Two JSL functions are interchangeable with scoping operators. Table 5.4 describes the functions and syntax.

Table 5.4 Scoping Operators

Operator and Equivalent Function	Function Syntax	Explanation

`:` `As Column`	`:name`
	`dt:name`
	`As Column(`*`dt`*`,`
	`name``)`
		Forces *`name`* to be evaluated as a data table column. The optional data table reference argument, *`dt`*, sets the current data table. See "Scoped Column Names" for examples.

`::` `As Global`	`::name` `As Global(name)`	Forces *`name`* to be evaluated as a global variable. **Note:** The double-colon is also used as an infix operator to represent ranges.

Scoped Column Names

Scoping column names is the simplest way to prevent conflicts with variable names. Use scoping operators to force names in a script to refer to columns.

1. The prefix colon (`:`) means that the name refers to a table column or table variable only, never a global variable. The prefix colon refers to the current data table context.

   ```
   :age;
   ```

2. The infix colon (`:`) operator extends this notion by using a data table reference to specify which data table has the column. This is particularly important when multiple data tables are referenced in a script.

   In the following example, the `dt` variable sets the data table reference to `Big Class.jmp`. The infix colon separates the data table reference and `age` column.

```
dt = Data Table( "Big Class.jmp" );
dt:age // The colon is an infix operator.
```

`As Column()` achieves the same results:

```
dt = Data Table( "Big Class.jmp" );
As Column( dt, age );
```

Therefore, the following expressions are equivalent when only Big Class.jmp is open:

```
:age;
As Column( dt, age );
dt:age;
```

The `Column` function can also identify a column. For Big Class.jmp, the following expressions all refer to age, the second column in the table:

```
Column( "age" );
Column( 2 );
Column( dt, 2 );
Column( dt, "age" );
```

Prevent Column Name and Variable Name Conflicts

To avoid conflicts, use unique column and variable names or scope the names.

- When a global variable and a column have the same name, the global variable name takes precedence. In this situation, you must scope the column name.

  ```
  ::age = [];
  age = :age << Get As Matrix;
  ```

- To avoid ambiguity between the global variable and column name, scope both variables.

  ```
  ::age = :age << Get As Matrix;
  ```

- If more than one data table might be open, assign data table references to variables. Scope your columns to the appropriate table.

  ```
  dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
  dt2 = Open( "$SAMPLE_DATA/Students.jmp" );
  ::age = dt1:age << Get As Matrix;
  ::height = dt2:height << Get As Matrix;
  ```

Note that JMP evaluates column formulas through each consecutive cell of the column, so scoping the column name is usually unnecessary. However, if a variable assigned in a formula has the same name as a column, you must scope the column name. See "Scoped Names".

Tip: The `Names Default to Here(1)` function also affects name resolution. See "Names Default To Here and Global Variables".

Unscoped Column Names

Sometimes an unscoped name gets or sets a value. JMP resolves it as a column in a data table (rather than a global variable) under these circumstances:

- if no global variable, local variable, or an argument using that name already exists,
- and the data table in context has a column of that name,
- and
  - either the current row is set to a positive value
  - or the name is subscripted (for example, the subscript `[1]` in `weight[1]` selects the first value in the weight column).

Exception

In column formulas and Nonlinear formulas, column names take precedence over global variables.

Set the Current Data Table Row

By default, the current row is 0, an illegal row number. So the following expression assigns a missing value to the *ratio* global variable:

```
ratio = height / weight;
```

Specify the row number with the `Row()` function. In the following example, the row is set to 3. The height in that row is divided by the weight, and the result is assigned to the *ratio* global variable.

```
Row() = 3;
ratio = height / weight;
```

Another possibility is to use subscripts to specify the row number. The following expression divides the height in row 3 by the weight in row 4.

```
ratio = height[3] / weight[4];
```

Specifying the row number is unnecessary when the script iterates a row at a time down the whole column. The following example creates the ratio column. For each row, the height is divided by the weight.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "ratio" );
For Each Row( :ratio = height / weight );
```

JMP evaluates formulas and calculates pre-evaluated statistics iteratively down a column. In these instances, identifying the row number is also unnecessary. (Pre-evaluated statistics are single numbers computed from the data table, as discussed in "Pre-Evaluated Statistics".)

Name Resolution in Automatically Run Scripts

In an `Open()` function, you can specify that the script run automatically when it's opened.

```
Open( "$MY_SCRIPTS/A.jsl", Run JSL( 1 ) );
Open( "$MY_SCRIPTS/B.jsl", Run JSL( 1 ) );
```

You can also type `//!` on the first line of the script to run the script automatically.

In these instances, the only shared names between `A.jsl` and `B.jsl` are globals and namespaces (not here and not local).

To apply the names in `A.jsl` to `B.jsl`, include `B.jsl` in `A.jsl` instead of opening and running both scripts.

```
Include( "$MY_SCRIPTS/B.jsl" );
```

Troubleshoot Variables and Column Names

When you reference a column name using `As Name()`, and `Names Default To Here( 1 )` is set, JMP returns a variable reference. That reference is then processed using the standard reference rules.

In the following example, there is no height variable in the `Here:` scope, so JMP returns an error.

```
Names Default To Here( 1 );
Open( "$SAMPLE_DATA/Big Class.jmp" );
As Name( "height" )[3]; // converts height to a variable reference
  As Name( "height" )[/*###*/3];
```

The `/*###*/` characters in the log indicate the location of an error in the script execution.

To prevent this problem, use one of the following methods:

- Use `As Column()` instead of `As Name()`:

  ```
  Names Default To Here( 1 );
  Open( "$SAMPLE_DATA/Big Class.jmp" );
  As Column( "height" )[3]; // converts height to a data column reference
  ```

- Explicitly scope height with `As Name():`

  ```
  Names Default To Here( 1 );
  dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
  dt:( As Name( "height" ) )[3]; // scopes height as a data column reference
  ```

These scripts return 55, the value of height in the third row of Big Class.jmp.

Troubleshoot Variables and Keywords

Name resolution errors can also occur when a variable and unquoted keyword have the same name. For example, one argument for `<<Preselect Role()` is "Y". Quote this argument if your script also uses Y as a variable.

Frequently Asked Questions about Name Resolution

Should you always scope?

Yes. When in doubt, scope. Scoping is especially important in scripts that might be used by many people on a variety of data tables; you will not necessarily know whether a name is used in two contexts (such as for both a global variable and column name).

If you are writing such scripts, consider using explicit scoping and namespaces. See "Advanced Scoping and Namespaces".

Prefix scope operators do not take run-time overhead after the first resolution. Infix scope operators, which follow data table references, always take run-time overhead.

What is the difference between a column reference and a column referred to by name? If I have a column reference in a global variable, how do I assign a value to a cell in the column?

With a column reference, you can send messages to change specific characteristics of the column or to access its values (for example, coloring cells or setting a formula).

When a column has been assigned to a global variable, assign a value to a cell in the column using a subscript. Suppose that the name of the column height has been assigned to the x variable:

```
x = Column( "height" );
```

Assign a value to the third row in the height column:

```
x[3] = 64 // sets the third row of height to 64
```

Note: The current row in a JSL script is *not* determined by selecting rows or positioning your cursor in the row. The current row is defined to be zero (no row) by default. You can set a current row with `Row()` (for example, `Row() = 3`). Please note that such a setting only lasts for the duration of that script. Then, `Row()` reverts to its default value, zero. This behavior means that submitting a script all at once can produce different results than submitting a script a few lines at a time.

Another way to establish a current row for a script is to enclose it in `For Each Row()`. This method executes the script once for each row of the current data table. For an example, see "If". See "Data Tables" for more information about working with data tables.

Which Has Precedence When Scoping, ":" or "[ ]"?

Scoping occurs before subscripting. This means that these two lines of code are equal:

```
dataTable:colName[i]
(dataTable:colName)[i]
```

Alternatives for Gluing Expressions Together

You can separate expressions with a semicolon, either on the same line or on different lines. JMP then evaluates each expressions in succession, returning the result of the last one. Here is an expression that first sets `a` to 2 and then sets `b` to 3:

```
a = 2;
b = 3;
```

The semicolon joins the two expressions and returns the value of the last one. So if `x = (a = 2; b = 3)`, the value of `x` is 3.

The `Glue()` function returns the result of the last expression. This function is equivalent to using semicolons. The following expressions both return 3:

```
Glue( a = 2, b = 3 );
a = 2; b = 3;
```

The `First()` function also evaluates each argument sequentially but returns the result of the first expression. The following expression returns 2:

```
First( a = 2, b = 3);
```

Example

What does `First()` do in the following script?

```
x = 1000;
```

```
First( x, x = 2000 );
```

The `First()` function returns the value of `x` (1000). 2000 is then assigned to `x`.

Iterate

JSL provides the `For()`, `While()`, `Summation()`, and `Product()` functions to repeat (or *iterate*) actions according to the conditions that you specify.

Note: A similar function called `For Each Row()` is for iterating actions over rows of a data table. See "If" for an example. "Additional Ways to Access Data Values" also describes iterating through table rows.

- "For"
- "While"
- "Summation"
- "Product"
- "Break and Continue"

For

The `For()` function expects four arguments separated by commas. The first three arguments are rules for how many times to repeat the loop, and the fourth is what to do each time the loop is executed.

Here is the basic syntax for `For()`:

```
For( initialization, while, iteration, body );
```

For example, the following script sums the numbers from 0 to 20:

```
s = 0;
For( i = 0, i <= 20, i++, s += i );
```

The script works like this:

`s = 0;`	Sets the `s` variable to 0. This variable holds the sum.
`For(`	Begins the `For()` loop.
`i = 0,`	Sets the initialization variable (`i`) to 0. This expression is performed only once.

`i <= 20,`	Each time the loop begins, compares `i` to 20. As long as `i` is less than or equal to 20, continue evaluating the loop. If `i` is greater than 20, immediately break out of the loop.
`i++,`	At the end of the loop, increments `i` by 1. Note that this step is done after the body of the loop (next line) is evaluated.
`s += i`	Evaluates the body of the loop. Adds the value of `i` to `s`. After the body is finished, *`i`* is incremented (previous line).
`);`	Ends the loop.

Infinite Loops

`For` loops that always evaluate as true create an infinite loop, which never stops. To stop the script, press Esc on Windows (or Command+period on macOS). You can also select **Edit > Stop Script**. On macOS, **Edit > Stop Script** is available only when the script is running.

Comparing For Loops in JSL to C and C++

The JSL `For()` loop works just like it does in the C (and C++) programming language, although the punctuation is different.

Tip: If you know C, watch out for the difference between JSL and C in the use of semicolons and commas. In JSL, `For()` is a function where commas separate arguments and semicolons join expressions. In C, `for` is a special clause where semicolons separate arguments and commas join them.

While

A related function is `While()`, which repeatedly tests the condition and evaluates its body script as long as the condition is true. The syntax is:

```
While( condition, body );
```

For example, here are two different programs that use a `While()` loop to find the least power of 2 that is greater than or equal to `x` (287). The result of both programs is 512.

```
x = 287;

// loop 1:
y = 1;
While( y < x, y *= 2 );
Show( y );
```

```
// loop 2:
k = 0;
While( 2 ^ k < x, k++ );
Show( 2 ^ k );
```

The scripts work like this:

`x = 287;`	Sets `x` to 287.
`// loop 1`	
`y = 1;`	Sets `y` to 1.
`While(`	Begins the `While()` loop.
`y < x,`	As long as `y` is less than `x`, continues evaluating the loop.
`y *= 2`	Multiplies 1 by 2 and then assigns the result to `y`. The loop then repeats while `y` is less than 287.
`);`	Ends the loop.
`Show(y);`	Shows the value of `y` (512).
`// loop 2`	
`k = 0;`	Sets `k` to 0.
`While(`	Begins the `While()` loop.
`2 ^ k < x,`	Raises 2 to the exponent power of `k` and continues evaluating the loop as long as the result is less than 287.
`k++`	Increments `k` to 1. The loop then repeats while `2 ^ k` is less than 287.
`);`	Ends the loop.
`Show(2 ^ k);`	Shows the value of `2 ^ k` (512).

As with `For()` loops, `While()` loops that always evaluate as true create an infinite loop, which never stops. To stop the script, press Esc on Windows (or Command-period on macOS). You can also select **Edit > Stop Script**. On macOS, **Edit > Stop Script** is available only when the script is running.

Summation

The Summation() function adds the body results over all i values. The syntax is:

```
Summation( initialization, limitvalue, body );
```

For example:

```
s = Summation( i = 1, 10, i );
```

returns 55, the result of 1+2+3+4+5+6+7+8+9+10.

The script works like this:

s =	Sets the s variable to the value of the function.
Summation(	Begins the Summation() loop.
i = 1,	Sets i to 1.
10,	Sets the limit of i to 10.
i	All values of i from 1 to 10 are added together, resulting in 55.
);	Ends the loop.

This behavior is similar to Σ in the Formula Editor. The following expression:

```
Summation( i = 1, N Row(), x ^ 2 );
```

is equivalent to the following formula in the Formula Editor:

$$\sum_{i=1}^{\text{NRow}} x^2$$

Product

The Product() function is similar to Summation() except that it multiplies the body results rather than adding them. The syntax is the same as for Summation(). For example:

```
p = Product( i = 1, 5 , i );
```

returns 120, the result of 1*2*3*4*5.

In this example, the initial value of i is 1, the upper limit is 5, then all integer values of i up to 5 are multiplied.

Here is the equivalent in the Formula Editor:

$$\prod_{i=1}^{5} i$$

Break and Continue

The `Break()` and `Continue()` functions give you more control over looping. `Break()` immediately stops the loop and proceeds to the next expression that follows the loop. `Continue()` is a gentler form of `Break()`. It immediately stops the current iteration of the loop and continues with the next iteration of the loop.

Break

`Break()` is typically used inside a conditional expression. For example:

```
For( i = 1, i <= 5, i++,
   If( i == 3, Break() );
   Print( "i=" || Char( i ) );
);
```

results in:

```
"i=1"
"i=2"
```

The script works like this:

`For(`	Begins the `For()` loop.
`i = 1,`	Sets `i` to 1.
`i <= 5,`	As long as `i` is less than or equal to 5, continues evaluating the loop.
`i++,`	Increments `i` by 1. Note that this step is done after the `If` loop is evaluated.
`If(`	Begins the `If()` loop.
`i == 3, Break()`	If `i` is equal to 3, breaks the loop.
`);`	Ends the loop.
`Print(`	When `i` is not equal to 3, opens the `Print()` loop.
`"i="`	Prints the string `"i="` to the log.
`||`	Places `"i="` on the same line as the value that follows.

`Char(i));`	Prints the value of `i` to the log. The `For()` loop then repeats until the value of i is less than or equal to 5, breaking when i equals 3, and breaking and printing only when i is less than 3.
`);`	Ends the loop.

Note that when the `If()` and `Break()` expressions follow `Print()`, the script prints the values of `i` from 1 to 3, because the loop breaks after "i=3" is printed.

```
"i=1"
"i=2"
"i=3"
```

Continue

As with `Break()`, `Continue()` is typically used inside a conditional expression. For example:

```
For( i = 1, i <= 5, i++,
   If( i < 3, Continue() );
   Print( "i=" || Char( i ) );
);
```

results in:

```
"i=3"
"i=4"
"i=5"
```

The script works like this:

`For(`	Begins the `For()` loop.
`i = 1,`	Sets `i` to 1.
`i <= 5,`	Evaluates 1 as less than or equal to 5.
`i++,`	Increments `i` by 1. Note that this step is done after the `If` loop is evaluated.
`If(`	Begins the `If()` loop.
`i < 3, Continue()`	Evaluates `i` as 1 and continues as long as `i` is less than 3.
`);`	Ends the `If()` loop.

`Print(`	When `i` is no longer less than 3, opens the `Print()` loop.
`"i="`	Prints the string `"i="` to the log.
`||`	Places `"i="` on the same line as the value that follows.
`Char(i));`	Prints the value of `i` to the log. The `For()` loop then repeats until the value of `i` is less than or equal to 5, continuing and printing only when i is equal to or greater than 3.
`);`	Ends the loop.

Conditional Functions

JSL provides five functions to evaluate an expression conditionally:

- "If"
- "Match"
- "Choose"
- "Interpolate"
- "Step"

If

The `If()` function evaluates the first result expression when its condition evaluates as true (a nonzero or nonmissing value). Otherwise, it evaluates the second result expression.

The syntax is:

```
If ( condition, result1, result2 );
```

For example, the following script returns "Young" when the age is less than 12. Otherwise, the script returns "Young at Heart".

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "age group", "Character" );
For Each Row(
   :age group = If( :age <= 12,
      "Young",
      "Young at Heart"
   )
);
```

You can also string together multiple conditions and results. The syntax is:

```
If( condition1, result1,
  condition2, result2,
  ...,
  resultElse );
```

In the preceding example, if condition1 is not true, the function continues evaluating until it finds a true condition. Then that condition's result is returned.

The last result is returned when all conditions are false. And when a value is missing, the missing value is returned. For these reasons, it's very important to include a default result at the end of the expression. Consider the following example, which recodes gender abbreviations in Big Class.jmp:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row( sex =
  If(
     sex == "F", "Female",
     sex == "M", "Male",
     "Unknown" );
);
```

The script works like this:

`For Each Row( sex =`	For each row in the table, sex is the column that is recoded.
`If(`	Begins the `If()` loop.
`sex == "F", "Female",`	If the value of sex is F, replaces the value with Female.
`sex == "M", "Male",`	If the value of sex is M, replaces the value with Male.
`"Unknown" );`	If neither of the above conditions are true, replaces the value with Unknown. If this result were omitted and the value of sex were missing, the script would return a missing value.
`);`	Ends the loop.

You can also put actions and assignments in the result expression. The following example assigns 20 to `x`, because the first condition (`y < 20`) is false:

```
y = 25;
z = If( y < 20, x = y, x = 20 );
```

Note: Be careful to use two equal signs (==) for equality tests, not one equal sign (=). An `If` with an argument such as *`name=value`* assigns rather than tests the value.

Avoiding Memory Issues with Multiple If Statements

Running a script that contains dozens of If() Statements can cause memory problems. We recommend reducing the If() nesting. Suppose that your script contains 100 `If()` statements in this format:

```
If( val, If( val, If( val, If( val, If( val,
If( val, If( val, If( val, If( val, If( val,
...
) ) ) ) ) ) ) ) ) );
```

Rewrite the script:

```
If( val, val1, If(val, val2, If( val, val3,
If( val, val4, If( val, val5, If( val, val6, ...) ) ) ) ) );
```

You can also try setting the maximum call depth. The `Maximum Call Depth` preference sets the default for the maximum call depth (or stack size) in which JSL built-in functions, user-defined functions, or `Recurse()` function calls can be made. By default, the maximum call depth is set to 256.

In JMP, your JSL script has a limited amount of physical runtime stack memory in which to perform JSL function calls. By default, this size is set to 2MB. Increasing the maximum call depth can cause a physical runtime stack overflow, so incrementally increase this preference in small amounts until you find the best value that works for your JSL script.

```
Preferences[1] << Set( Maximum JMP call depth( 64 ) );
```

Match

You can use the `Match()` function to make several equality comparisons without needing to rewrite the value to be compared. The syntax is:

```
Match( x, value1, result1, value2, result2, ..., resultElse );
```

For example, the following script recodes gender abbreviations in Big Class.jmp:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row( sex =
  Match(
     sex,
     "F", "Female",
     "M", "Male",
     "Unknown");
  );
```

The script works like this:

`For Each Row( sex =`	For each row in the table, `sex` is the column that is recoded.
`Match(`	Begins the `Match()` loop.
`sex,`	Specifies `sex` as the match argument.
`"F", "Female",`	If the value matches "F" replace it with “Female”.
`"M", "Male",`	If the value matches "M", replace it with "Male".
`"Unknown" );`	If F or M are not matched, replaces the value with "Unknown".
`);`	Ends the loop.

This `Match()` example is a simplified version of the example in “If”. The advantage of `Match()` is that you define the comparison value once rather than repeat it in each condition. The disadvantage is that you cannot use expressions with operators as you can with `If`; the argument `sex == "F"` returns an error in a `Match()` expression.

With more groups of conditions and results, the value of `Match()` becomes more apparent. The following script would require many additional lines of code with `If()`.

```
dt = Open( "$SAMPLE_DATA/Travel Costs.jmp" );
For Each Row(
   Booking Day of Week = Match( Booking Day of Week,
      "Sunday", "SUN",
      "Monday", "MON",
      "Tuesday", "TUE",
      "Wednesday", "WED",
      "Thursday", "THU",
      "Friday", "FRI",
      "Saturday", "SAT",
      "Not Specified"
   )
);
```

Be careful with the data type of the condition and result. In the preceding example, the conditions and results are both character data. If the data types do not match, JMP automatically changes the column’s data type. The results might not be what you want.

Note: The `Match()` function explicitly checks to see if the compare expression *x* is missing and if the value of *value1* is missing, then it returns the value of `result1`; otherwise it continues to compare the expression *x* to each *valueN* value in each *valueN/resultN* pair, ignoring any missing values. If the expression *x* is equal to any of the *valueN* value, then the corresponding *resultN* value is returned. If no matching *valueN* value is found, then the *resultElse* value is returned.

Choose

The `Choose()` function shortens scripts even more than `Match()`, provided the arguments are tested against integers. The syntax is:

```
Choose( expr, result1, result2, result3, ..., resultElse );
```

Suppose you have a data table with a group column of numeric values from 1 through 7. If the first cell contains the number 1, the following script returns x = "Low".

```
x = ( Choose( group[1], "Low", "Medium", "High", "Unknown" ); );
Show( x );
```

The script works like this:

`x =`	Creates the x variable.
`Choose(`	Begins the `Choose()` loop.
`group[1],`	Evaluates the value of group in the first row.
`"Low",`	If the value of group is 1, return "Low".
`"Medium",`	If the value of group is 2, return "Medium".
`"High",`	If the value of group is 3, return "High".
`"Unknown"`	Otherwise, return "Unknown".
`)`	Ends the loop.
`);`	Closes the x variable.
`Show( x );`	Returns the value of x.

If the expression evaluates to an out-of-range integer (such as 7 when only 4 replacement values are listed), the last result is returned. In the preceding example, "Unknown" is returned.

Notice that `If()` and `Match()` require more code to achieve the same results as the `Choose` function:

```
If(
  group[1] == 1, "Low",
  group[1] == 2, "Medium",
  group[1] == 3, "High",
  "Unknown"
);
Match( group, 1, "Low", 2, "Medium", 3, "High", "Unknown" );
```

Note: If the data types in the expression do not match, JMP automatically changes the column's data type.

Interpolate

The `Interpolate()` function performs linear interpolation for continuous data or bilinear interpolation for two-dimensional continuous data.

In the simplest case, the function finds the `y` value that corresponds to a given `x` value between two sets of points. You might use `Interpolate()` to calculate missing values between data points.

The data points can be specified as alternating ordered arguments:

```
Interpolate( x, x1, y1, x2, y2, ... );
```

or as matrices containing the `x` and `y` values:

```
Interpolate( x, xmatrix, ymatrix );
```

Suppose that your data set includes the height of individuals from age 20 through 25. However, there is no data point for age 23. To estimate the height for 23-year-olds, use interpolation. The following example shows the value that you want to evaluate (age 23), followed by matrices for ages (20 through 25) and heights (59 through 75).

```
Interpolate( 23, [20 21 22 24 25], [59 62 56 69 75] );
```

returns:

```
62.5
```

The value 62.5 is halfway between the `y` values 56 and 69, just as 23 is halfway between the `x` values 22 and 24.

You can also interpolate multiple points by specifying a matrix or list of values in the first argument.

```
Interpolate( [23, 24.5], [20 21 22 24 25], [59 62 56 69 75] );
```

returns:

```
[62.5, 72]
```

Bilinear Interpolation

You can also interpolate in two-dimensions:

```
Interpolate( {x, y}, xvector, yvector, zmatrix );
```

Here, the first argument is a list of two points, the second and third arguments are vectors that define the grid of `x` and `y` values, and the fourth argument is a matrix of data points. The function then finds the interpolated `z` value within the appropriate quadrant of the `zmatrix`. The appropriate quadrant is found by comparing the `x` and `y` values to the `xvector` and `yvector` arguments.

Suppose you have a 2x3 matrix and you want to interpolate a point that is halfway between the first and second rows of points and halfway between the second and third columns of points.

```
Interpolate( {0.5, 0.75}, [0 1], [0 0.5 1], [10 15 20, 12 16 18] );
```

returns:

```
17.25
```

Note that 0.5 is halfway between 0 and 1 for the x values (corresponding to the rows of the matrix) and 0.75 is halfway between 0.5 and 1 for the y values (corresponding to the columns of the matrix). You can find the bilinear interpolation in two steps:

1. Find the halfway points between the second and third columns for the each row of the matrix:
   - Row 1: 17.5 is halfway between 15 and 20.
   - Row 2: 17 is halfway between 16 and 18.
2. Find the halfway point between the two points in the previous step: 17.25.

Notes:

- The `x` values must be in ascending order. For example, `Interpolate(2,1,1,3,3)` returns 2. However, `Interpolate(2,3,3,1,1)` returns a missing value (.).
- `Interpolate()` is best used for continuous data, but `Step()` is designed for discrete data. See "Step".

Step

The `Step()` is like `Interpolate()` except that it finds the corresponding `y` for a given `x` from a step-function fit rather than a linear fit. Use `Step()` with discrete `y` values (that is, when the `y` value's corresponding `x` value can be only y_1 or y_2). However, when the `y` value's corresponding `x` value can fall between y_1 and y_2, use `Interpolate()`.

As with `Interpolate`, the data points can be specified as a list:

```
Step( x, x1, y1, x2, y2, ... );
```

or as matrices containing the `x` and `y` values:

```
Step( x, xmatrix, ymatrix );
```

Suppose that your data table shows the discount percentage for purchases of $25, $50, $75, and $100. You want to create a graph that shows the discount for a $35 purchase, which the data table does not specify. The following example shows the value that you want to evaluate, 35, followed by matrices for purchases from $25 to $100.

```
Step( 35, [25 50 75 100], [5 10 15 25] );
```

returns:

```
5
```

If the discounts were on a sliding scale (in this example, between 5 and 10, you would use `Interpolate()`:

```
Interpolate( 35, [25 50 75 100], [5 10 15 25] );
```

returns:

```
7
```

As with `Interpolate()`, the data points must create a positive slope.

Compare Incomplete or Mismatched Data

Comparing data that contains missing values can return misleading results unless you specify a condition that is always true or use functions such as `Is Missing()` or `Zero Or Missing()`. Comparisons of data with mismatched types (numeric versus character) or data in matrices can also be confusing.

Table 5.5 shows examples of such comparisons and matrices and explanations of the results. For a review of operators used in comparisons, see "Operators". For an overview of comparing data, see the *JSL Syntax Reference*.

Note: Matrices must include the same number of columns or rows.

Table 5.5 Some Special-Case Comparison Tests

Test	Result	Explanation
`m=.; m==1`	.	An equality test with a missing value returns missing.
`m=.; m!=1`	.	An inequality test with a missing value returns missing.
`m=.; m<1; m>1; and so on`	.	A comparison with a missing value returns missing (unless it could not possibly be true, see next).
`m=.; 1<m<0`	0	A comparison involving a missing value that could not possibly be true returns false; false takes precedence over missing for comparisons with more than two arguments (as with logical operators).
`{a, b}==List(a, b)`	1	An equality test of list arguments returns a single result.
`{a, b}<{a, c}`	.	A comparison test of list arguments is not allowed.
`1=="abc"`	0	An equality test with mixed data types returns false.
`1<="abc"`	.	A comparison with mixed data types returns missing.
`[1 2 3]==[2 2 5]`	`[0 1 0]`	An equality test of matrices returns a matrix of elementwise results. When a matrix is compared to a matrix, comparison is done element-by-element and returns a matrix of 1s and 0s.
`[1 2 3]==2`	`[0 1 0]`	An equality test of a matrix and a matrix filled with 2s. If a matrix is compared to a number, the number is treated as a matrix filled with that number.
`[1 2 3] < [2 2 5]`	`[1 0 1]`	A comparison of matrices returns a matrix of elementwise results.
`[1 2 3] < 2`	`[1 0 0]`	A comparison of a matrix and a matrix filled with 2s.
`Is Missing(m)`	1	Returns 1 for a missing value and returns 0 otherwise. For missing character values, you can also use empty quotes for the comparison, as in m == "".
`Zero Or Missing(m)`	1	Returns 1 when the value is 0 or missing. The argument must be numeric or a matrix and not a string.
`All([2 2]==[1 2])`	0	Summarizes elementwise comparisons; returns 1 only if *all* comparisons are true and returns 0 otherwise.

Table 5.5 Some Special-Case Comparison Tests *(Continued)*

Test	Result	Explanation
`Any([2 2]==[1 2])`	1	Summarizes elementwise comparisons; returns 1 if *any* comparison is true and returns 0 otherwise.

Missing Values

In a comparison, missing values typically return missing, not true or false. For this reason, it is very important to include a result that is always true. Suppose that a data table column contains the values 1, 2, 3, and a missing value in column A. A formula in column B sets up the comparison.

```
New Table( "Testing Comparisons",
   Add Rows( 4 ),
   New Column( "A",
      Numeric,
      "Continuous",
      Format( "Best", 10 ),
      Set Values( [1, 2, 3, .] )
   ),
   New Column( "B", Character, "Nominal", Formula( If( :A, "true", 1, "false"
   ) ) )
);
```

The following values are stored in column B:

```
"true"
"true"
"true"
"false"
```

The script works like this:

`If(`	Begins the comparison.
`:A, "true",`	If the value of A is nonmissing and nonzero, the result is "true". This comparison is true for the first three rows.
`1, "false"`	The value of 1 is always true, so the missing value returns "false".
`);`	Closes the comparison.

Notes:

- When a cell contains a non-empty string, the comparison fails with the first comparison. The following script returns a missing value if a is a string. It works for numeric values.

```
a = "a";
If( a, "true", "false" );
```

- When one value is false and another one is missing, `Or()` returns missing. Use `OrMZ()` to return false instead. See the *JSL Syntax Reference*.
- When one value is true and another one is missing, `And()` returns missing. Use `AndMZ()` to return false instead. See the *JSL Syntax Reference*.
- `MatchMZ()` is similar to `Match()` except that missing values are treated as 0.
- `IfMZ()` is similar to `If()` except that missing values are treated as 0.

Is Missing

If you know that some values are missing, you can also compare with `Is Missing()`. The comparison in the preceding example can be rewritten to return "missing" for missing values:

```
If( :A, "true", Is Missing( :A ), "missing", "false" );
```

The preceding expression returns "true" when A is nonmissing and nonzero, "missing" when A is missing, and "false" otherwise.

Zero Or Missing

If the missing value could be 0, use the `Zero Or Missing()` function instead:

```
Zero Or Missing( A );
```

This expression returns 1 when A is 0 or missing.

Tip: You cannot compare a known value with an explicitly defined missing value, only with variables, matrices, or other things that could *contain* missing values.

Inquiry Functions

Inquiry functions identify the type of an element, such as a string, list, or matrix. You can then write a script specific to that element type.

JMP also uses inquiry functions to determine the writability of a directory or file and to identify a computer's operating system and the JMP version.

General Element Types

The Type() function returns a string naming the type of the resulting value. For example:

```
Show( Type( 1 ), Type( "hi" ), Type( {"a", 2} ), Type( [10 24 325] ) );
```

results in:

```
Type(1) = "Integer"
Type("hi") = "String"
Type({"a", 2}) = "List"
Type([10 24 325]) = "Matrix";
```

Specific Element Types

Other inquiry functions (such as Is Matrix(), Is List(), Is Scriptable(), and so on) let you test for specific types of objects. In the following example, Is Matrix() evaluates as true, then the specified calculations are run:

```
a = [2 3];
b = [1, 1];
c = a * b;
If( Is Matrix( c ),
   (c ^ a) / (a * b),
   Print( "c is not a matrix." )
);
   [5 25]
```

Is Scriptable() returns 1 when the object is scriptable. Four variables in the following example refer to a data table, column, platform, and report. All four objects are scriptable, so Is Scriptable() returns 1 for each example.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = Column( "weight" );
plat = Bivariate( Y( :weight ), X( :height ) );
rep = Report( plat );
Show( Is Scriptable( dt ) );
   Is Scriptable(dt) = 1;
Show( Is Scriptable( col ) );
   Is Scriptable(col) = 1;
Show( Is Scriptable( plat ) );
   Is Scriptable(plat) = 1;
Show( Is Scriptable( rep ) );
   Is Scriptable(rep) = 1;
```

Is Empty() tests to see whether a variable has a value, a function, an expression, or a reference to an object. It returns 1 if the variable is undefined or holds the Empty() value.Otherwise, you get errors when referring to something that has not been created or assigned a value yet. Programmers call this an *uninitialized variable*.

Here is an example of a test to see whether a data table is opened and therefore assigned to the `dt` variable. If a data table is not opened, the `Open()` function prompts the user to open the table.

```
If( Is Empty( dt = Current Data Table() ),
   dt = Open()
);
```

You can use `Is Empty()` for any variable (such as global variable, local variable, and columns).

Table 5.6 shows functions that identify object types.

Table 5.6 Inquiry Functions That Identify Object Types

Syntax	Explanation
`Is Associative Array(`*x*`)`	Returns 1 if the evaluated argument is an associative array or 0 otherwise.
`Is Directory(x)`	Returns 1 if the x argument is a directory and 0 otherwise.

`Is Empty(`*global*`)`
`Is Empty(`*dt*`)`
`Is Empty(`*col*`)`
	Returns 1 if the global variable, data table, or data column does not have a value (is uninitialized) or 0 otherwise.

`Is Expr(`*x*`)`	Returns 1 if the evaluated argument is an expression or 0 otherwise.
`Is File(x)`	Returns 1 if the x argument is a file and 0 otherwise.
`Is List(`*x*`)`	Returns 1 if the evaluated argument is a list or 0 otherwise.
`Is Matrix(`*x*`)`	Returns 1 if the evaluated argument is a matrix or 0 otherwise.
`Is Name(`*x*`)`	Returns 1 if the evaluated argument is a name or 0 otherwise. See "Retrieve a Stored Expression Rather than its Result".
`Is Namespace(`*x*`)`	Returns 1 if the evaluated argument is a namespace or 0 otherwise.
`Is Number(`*x*`)`	Returns 1 if the evaluated argument is a number or missing numeric value or 0 otherwise.
`Is Scriptable(`*x*`)`	Returns 1 if the evaluated argument is a scriptable object or 0 otherwise.
`Is String(`*x*`)`	Returns 1 if the evaluated argument is a string or 0 otherwise.
`Type(`*x*`)`	Returns a string naming the type of x.

Object Attributes

JMP provides the following functions to determine whether a file or directory is writable before attempting to write to them. Use these functions in combination with `Is Directory(path)` and `Is File(path)` to verify a script destination and attribute. See the *JSL Syntax Reference* for additional information.

The `Is Directory Writable(path)` function returns 1 if the directory specified in the path argument is writable and 0 otherwise.

The `Is File Writable(path)` function returns 1 if the file specified in the path argument is writable and 0 otherwise.

For example, the following code verifies the path refers to a directory and then checks to ensure the directory is writable:

```
If( Is Directory( "$SAMPLE_DATA/Loss Function Templates" ),
  If( Is Directory Writable( "$SAMPLE_DATA/Loss Function Templates" ),
     "Directory is writable.",
     "Directory is read only!"
  ),
  "Is a read only directory."
);
```

Host Information

The `Host Is()` inquiry function identifies the current operating system. Then actions specific to that operating system can be performed.

The following example specifies text sizes in reports for different operating systems. If you commonly write your scripts on Windows and share them with macOS users, the results can look different from what you intended. For example, the following expression sets the text to a larger size (12) on macOS and a smaller size (10) on Windows:

```
textsize = If( Host is( "Mac" ), 12, 10 );
```

Version Information

The `JMP Version()` inquiry function returns the JMP version as a string. You might use this function to determine the JMP version and then run a script compatible with that version.

```
JMP Version(); // returns "16.0.0" in JMP 16
JMP Version(); // returns " 9.0.0" in JMP 9
```

Notice that a leading blank is inserted before versions less than 10.0.0. This blank helps when comparing version numbers. Without the leading blank, 9.0.0 is interpreted as greater than 10.0.0.

Chapter 6

Types of Data

Working with Numbers, Strings, Dates, Currency, and More

This chapter discusses basic data types:

- numbers and strings
- paths, which are a special type of string
- dates and times, which can be either special numbers or special strings
- currency
- hexadecimal values and blobs

At the end of the chapter are two sections that show more advanced methods of interacting with strings and pattern matching with regular expressions.

Contents

Numbers and Strings 145
Formatting Numbers 146
Unicode Characters 147
Path Variables 147
Create and Customize Path Variables 150
Relative Paths 151
File Path Separators 151
Date-Time Functions and Formats 152
Date-Time Values 152
Program with Date-Time Functions 153
Date-Time Values in Data Tables 160
Custom Date-Time Formatting with Format Patterns 164
Currency 166
Hexadecimal and BLOB Functions 167
Work with Character Functions 170
Concat 170
Munger 171
Repeat 172
Substitute and Substitute Into 173
Regular Expressions 174
Regex 174
Regex Match 176
Special Characters in Regular Expressions 178
Escaped Characters in Regular Expressions 179
Greedy and Reluctant Regular Expressions 181
Backreferences and Capturing Groups 183
Lookaround Assertions 184
Pattern Matching 185
Parsing Strings in Fixed Fields 186
Patterns and Case 188
Troubleshooting Patterns 188

Numbers and Strings

Numbers can be written as integers, decimal numbers, in scientific notation with an E preceding the power of ten, and as date-time values. A single period by itself is the missing numeric value.

For example, these are all numbers:

```
.   1   12   1.234  3E3  0.314159265E+1 1E-20
```

One or more characters placed within double quotation marks constitute a *string*. For example, these are all strings:

```
"Green" "Hello,\NWorld!" "54"
```

Notice that if a number is in quotation marks, it is a string, not a number. There are two functions you can use to change a number into a string or a string into a number.

- Use `Num()` to convert a string into a number. For example:

```
Num( "54" );
   54
```

Note: `Num()` cannot convert non-numeric characters, so it produces a missing value.

```
Num( "Hello" );
   .
```

- Use `Char()` to convert a number into a string. For example:

```
Char( 54 );
   "54"
Char( 3E3 )
   "3000"
```

To preserve locale-specific numeric formatting in `Num()` or `Char()` output, include the `<<Use Locale(1)` option as shown in the following example:

```
Char( 42, 5, 2, << Use Locale( 1 ) );
   // results in the character value "42,00" in the French locale
```

To look at each character in a string, use the `Substr()` function. This example looks for the letter "a" in the string and prints a message to the log:

```
ch = Substr( "alphabetic", 1, 1 ); // start and end with the first character
If( ch == "a",
   Print( "First letter is a." )
);
   "First letter is a."
```

Formatting Numbers

The `Format()` function provides options for formatting numbers with thousands separators, decimal points, geographic symbols, *p*-values, and so on. The following example shows a basic format for a fixed decimal column.

```
Format( x, "Fixed Dec", 10, 2, "Use thousands separator");
```

The format starts with the number to be formatted. Then you can specify the width of the number, the number of decimal places, and whether to use a thousands separator.

Here are other examples:

```
Format( x, "Currency", 20, <<Use Locale(0) );
  // ignores computer locale and uses a period for the decimal separator
Format( x, "Currency", 20, "Use thousands separator" );
Format( x, "m/d/y", 10 );
Format( x, "Precision", 10, 2, "Keep trailing zeroes", "Keep all whole digits"
  );
Format( x, "Latitude DDD", "PUNDIR"); // "PUN" for punctuation, "DIR" for
  direction, PUNDIR for both
Format( x, "Custom", Formula( Abs( value ) ), 15 );
```

The syntax for `Function()` arguments is shown in the *JSL Syntax Reference*. You can also find the argument names and options in the data table Column Info window.

See "Date-Time Functions and Formats" and "Currency" for more information about formatting date-time values and currency.

Note: If the date format is unknown, an error is written to the log.

Matrix Box, Number Col Box, Number Col Edit Box, Number Edit Box

To format numbers in `Matrix Box()`, `Number Col Box()`, `Number Col Edit Box()`, and `Number Edit Box()`, send the `Set Format` message to the display box. You can then specify the same arguments as for the `Format()` function, such as the following:

```
<<Set Format( 10, 2, "Use thousands separator");
<<Set Format( "Currency", "EUR", 20 );
```

The syntax for `Set Format` arguments is shown in the *JSL Syntax Reference*. You can also find the argument names and options in the data table Column Info window.

Note: If you specify the number of decimal places, that number must be preceded with the width.

Unicode Characters

JMP supports both Unicode UTF-8 and UTF-16 standards for encoding and representing text for most of the world languages. See the The Unicode Consortium for code charts and more information on the Unicode standard.

To display Unicode characters in JMP, precede the Unicode code for the character with '\!'. For example:

- Greek letter sigma (σ) in Unicode = U+03C3; in JMP, use \!U03C3
- Greek letter mu (μ) in Unicode = U+03BC; in JMP, use \!U03BC

To use Unicode to express superscripts and subscripts:

- subscript 1 ($_1$) in Unicode = U+2081; in JMP, use \!U2081
- superscript 2 (^{2}) in Unicode = U+00B2; in JMP, use \!U00B2

To express x^2 in Unicode, in JMP, use \!U0078\!U00B2.

To use Unicode surrogate pairs, specify the Unicode escape sequence twice. If you specify an emoji Unicode character, you might need to change the font for it to appear correctly. The following example shows how to include an emoji in a text box.

```
New Window( "Emoji Example",
   V List Box(
      Text Box( "Surrogate Pair with space: \!UD83D \!UDE00", Set Font( "Segoe
   UI Emoji" ) ),
      Text Box( "Surrogate Pair without space: \!UD83D\!UDE00", Set Font(
   "Segoe UI Emoji" ) )
   )
);
```

Path Variables

Path variables are shortcuts to directories or files. Rather than enter the entire path to the directory or file, you use the path variable in a script. A path variable is a special type of string and is always contained within double quotation marks.

One common predefined path variable in JMP is $SAMPLE_DATA. This variable points to the sample data folder in your JMP or JMP Pro installation folder. The following example opens the Big Class.jmp sample data table.

```
Open( "$SAMPLE_DATA/Big Class.jmp" );
```

Several path variables are predefined in JMP. The following table shows the definitions for the current JMP version. Variables in previous versions of JMP might differ.

Table 6.1 Predefined Path Variable Definitions

Variable	Path
ADDIN_HOME(com.your.addin.id) Defines the folder where the specified add-in is stored.	Add-In Builder stores the specified add-in in the following folders, based on your operating system: • Windows: "C:/Users/<username>/AppData/Roaming/SAS/JMP/Addins/" • macOS: "/Users/<username>/Library/Application Support/JMP/Addins/" For example, `Convert File Path( "$ADDIN_HOME(com.your.addin.id)` evaluates to the location of the `com.your.addin.id` add-in. To install your add-ins in a shared network location, use the `Register Addin()` function. See "Register an Add-In Using JSL".
ALL_HOME Defines a folder that can be accessed by all users on a machine.	• Windows (JMP): "/C:/ProgramData/SAS/JMP/16/" • Windows (JMP Pro): "/C:/ProgramData/SAS/JMPPro/16/" • macOS: "/Library/Application Support/JMP/16/" To see if the directory exists, run the script `Is Directory("$ALL_HOME");`. If the folder exists, 1 is returned.
BUILTIN_SCRIPTS	• Windows (JMP): "/C:/Program Files/SAS/JMP/16/Resources/Builtins/" • Windows (JMP Pro): "/C:/Program Files/SAS/JMPPRO/16/Resources/Builtins/" • macOS: "/Applications/JMP 16.app/Contents/Resources/Builtins/"
DESKTOP	• Windows: "/C:/Users/<username>/Desktop/" • macOS "/Users/<username>/Desktop/"
DOCUMENTS	• Windows: "/C:/Users/<username>/Documents/" • macOS: "/Users/<username>/Documents/"
DOWNLOADS	• Windows: "/C:/Users/<username>/Downloads/" • macOS: "/Users/<username>/Downloads/"
GENOMICS_HOME	"/<JMP Genomics installation directory>/"

Table 6.1 Predefined Path Variable Definitions *(Continued)*

Variable	Path
HOME	• Windows (JMP): "/C:/Users/<username>/AppData/Roaming/SAS/JMP/16/" • Windows (JMP Pro): "/C:/Users/<username>/AppData/Roaming/SAS/JMPPro/16/" • macOS: "/Users/<username>/"
JMP_HOME	• Windows: /C:/<JMP installation directory> **Note:** The variable is not available on macOS.
SAMPLE_APPS	• Windows: /C:/<JMP installation directory>/Samples/Apps/" • macOS: "/Library/Application Support/JMP/16/Samples/Apps/"
SAMPLE_DASHBOARDS	• Windows: /C:/<JMP installation directory>/Samples/Dashboards/" • macOS: "/Library/Application Support/JMP/16/Samples/Dashboards/"
SAMPLE_DATA	• Windows: "/C:/<JMP installation directory>/Samples/Data/" • macOS: "/Library/Application Support/JMP/16/Samples/Data/"
SAMPLE_IMAGES	• Windows: "/C:/<JMP installation directory>/Samples/Images/" • macOS "/Library/Application Support/JMP/16/Samples/Images/"
SAMPLE_IMPORT_DATA	• Windows: "/C:/<JMP installation directory>/Samples/Import Data/" • macOS: "/Library/Application Support/JMP/16/Samples/Import Data/"
SAMPLE_PROJECTS	• Windows: "/C:/<JMP installation directory>/Samples/Projects/" • macOS: "/Library/Application Support/JMP/16/Samples/Projects/"
SAMPLE_SCRIPTS	• Windows: "/C:/<JMP installation directory>/Samples/Scripts/" • macOS: "/Library/Application Support/JMP/16/Samples/Scripts/"
TEMP	• Windows: "/C:/Users/<username>/AppData/Local/Temp/" • macOS: "var/folders/..."

Table 6.1 Predefined Path Variable Definitions *(Continued)*

Variable	Path
USER_APPDATA Changes to JMP preferences, menus, and the Home Window are stored here, along with Debugger session settings.	• Windows (JMP): "/C:/Users/<username>/AppData/Roaming/SAS/JMP/16/" • Windows (JMP Pro): "/C:/Users/<username>/AppData/Roaming/SAS/JMPPro/16/" • macOS: "/Users/<username>/Library/Application Support/JMP/16/"

Path variable definitions are updated automatically based on the version of JMP you are using. For example, when you run a JMP 14 script in JMP 16, the JMP 16 path variable definitions are used.

To see the definition of any path variable, use the function `Get Path Variable`:

```
Get Path Variable( "HOME" );
   "/C:/Users/<username>/AppData/Roaming/SAS/JMP/16/"
```

Note that you don't include a dollar sign for `Set Path Variable()` or `Get Path Variable()`. But you must include the dollar sign when using the variable in a script.

Trailing Slashes

Make sure to include a trailing slash after the path variable. In the following example, the root name `"Big Class"` is assigned to the dtName variable. The `Open()` expression evaluates `$SAMPLE_DATA` and the trailing slash and then appends the `dtName` value along with the file extension `.jmp`.

```
dtName = "Big Class";
dt = Open( "$SAMPLE_DATA/" || dtName || ".jmp" );
```

The path is interpreted as:

```
C:/Program Files/SAS/JMP/16/Samples/Data/Big Class.jmp
```

Without the slash that follows `$SAMPLE_DATA`, the path is interpreted as:

```
C:/Program Files/SAS/JMP/16/Samples/DataBig Class.jmp
```

Create and Customize Path Variables

You can create your own path variables or override some of the built-in variables with the `Set Path Variable()`. In the following example, the path variable is called *root*. The variable points to the `c:/` directory.

```
Set Path Variable( "root", "c:/" );
```

To get the value of the new variable, use Get Path Variable().

```
Get Path Variable( "root" );
  "c:/"
```

Use your path variable as you would other variables. The following expression opens the myimportdata.txt file in the c:/ directory.

```
Open( "$root/myimportdata.txt" );
```

As with getting path variables, omit the dollar sign when setting path variables.

Relative Paths

If you plan to use relative paths in variables, you must set the default directory. Then any path not preceded by a drive letter is relative to the default directory. Here is an example:

```
Set Default Directory( "c:/users/smith/data" );
```

To return the value of the default directory, use Get Default Directory().

```
Get Default Directory(); // returns "c:/users/smith/data"
```

So the following expression resolves as C:/users/smith/data/cleansers.jmp:

```
Open("cleansers.jmp");
```

File Path Separators

In JMP, the preferred file path format is the Portable Operating System Interface (POSIX), or UNIX, format with forward slashes (/) as separators. This means that you do not have to identify the current operating system in scripts run on both Windows and macOS. However, each host still accepts its native format for compatibility.

You can a convert file path format from Windows to POSIX (and vice versa) using Convert File Path(). Converting from a POSIX to a Windows path might be useful when you need to output a path to a file or to another application. The syntax is:

```
Convert File Path (path, <absolute|relative>, <POSIX|windows>, <base(path)>);
```

For example, the following script converts a POSIX path to a Windows path:

```
Convert File Path( "c:/users/smith", windows);
  c:\users\smith
```

You can substitute a path variable (such as $HOME) for the path inside quotes.

Date-Time Functions and Formats

A date-time value consists of any portion of a date or time. The value can be seconds (3388594698), a complete date (such as "Wednesday, May 18, 2011"), the date and time ("05/18/2011 8:18:18 PM"), the week number (3), and so on.

JMP lets you convert date-time values to common formats, perform arithmetic on the values, and manipulate the data in a number of ways.

Tip: For descriptions of all date-time functions and their arguments, see the *JSL Syntax Reference*.

Date-Time Values

Date-time values are stored and calculated as the number of seconds since midnight, January 1, 1904. For example:

```
Today(); // returns 3388649872 on May 19, 2011 at 12:00:00 AM
```

As with `Today()`, the `Date DMY()` and `Date MDY()` functions also return month, day, and year arguments as seconds. For example, if it were 12:00:00 a.m. on May 19, 2011, all of the following statements would return the same value:

```
Date DMY( 19, 5, 2011 );
Date MDY( 5, 19, 2011 );
Today();
   3388608000
```

The `As Date()` function takes the number of seconds and displays it as a date or duration.

- Values that represent one year or more are returned as dates:

```
As Date( 3388608000 );
   19May2011
```

- Values that represent less than a year are returned as durations.

```
As Date( 50000 );
   :0:13:53:20
```

You can use date-time values in two ways:

- a literal value, for example `19May2011:10:10`
- a string, for example `"Thursday, May 19, 2011"`

You can perform arithmetic with date-time literals, which use the number of seconds as the base number.

```
As Date( 19May2011 + 1 );
```

19May2011:00:00:01

Program with Date-Time Functions

Table 6.2 shows functions that convert seconds into date-time values and date-time values into seconds.

Table 6.2 Date-Time Functions

Function	Explanation
Abbrev Date(*date*)	Returns a string representation for the *date* supplied. The format is based on your computer's regional setting. So for the English (United States) locale, the date is formatted like "02/29/2004". Even if you are running JMP in English with a different locale, the locale format is applied.

As Date(*expression*)	Formats a number or expression so that it shows as a date or duration in a text window. For example, values that represent one year or more are returned as dates.
	`x = As Date( 8Dec2000 + In Days( 2 ) );`
	shows as:
	10Dec2000
	Values that represent less than a year are returned as durations.
	`As Date( 50000 );`
	shows as:
	:0:13:53:20

Date DMY(*day*, *month*, *year*)	Returns the specified date expressed as the number of seconds since midnight, 1 January 1904. For example, the second Leap Day of the third millennium is DateDMY(29,2,2004), which returns 3160857600.
Date MDY(*month*, *day*, *year*)	Returns the specified date expressed as the number of seconds since midnight, 1 January 1904. For example, the second Leap Day of the third millennium is DateMDY(2,29,2004), which returns 3160857600.
Day Of Week(*date*)	Returns an integer representation for the day of the week of the *date* supplied. Weeks are Sunday–Saturday.

Table 6.2 Date-Time Functions *(Continued)*

Function	Explanation
`Day Of Year(date)`	Returns an integer representation for the day of the year of the *date* supplied.
`Day(date)`	Returns an integer representation for the day of the month of the *date* supplied.
`Format(date, "format")`	Returns the *value* in the *format* specified in the second argument. Most typically used for formatting datetime values from a number of seconds to a formatted date. Format choices are those shown in the **Column Info** dialog box. Also see Table 6.3.
`Hour(datetime)`	Returns an integer representation for the hour part of the *date-time* value supplied.
`In Days(n)`	These functions return the number of seconds per *n* minutes, hours, days, weeks, or years. Divide by these functions to express an interval in seconds as an interval in other units.
`In Hours(n)`	
`In Minutes(n)`	
`In Weeks(n)`	
`In Years(n)`	
`Long Date(date)`	Returns a string representation for the specified *date*. The format is based on your computer's regional setting. So for the English (United States) locale, the date is formatted like `"Sunday, February 29, 2004"`. Even if you are running JMP in English with a different locale, the locale format is applied.
`MDYHMS(date)`	Returns a string representation for the *date* supplied, formatted like `"2/29/2004 00:02:20 AM"`.
`Minute(date-time)`	Returns an integer representation for the minute part of the *date-time* value supplied.
`Month(date)`	Returns an integer representation for the month of the *date* supplied.
`Num(date-time)`	Returns an integer representation for the *date-time* value supplied.
`InFormat(string, "format")` `Parse Date(string, "format")`	Parses a *string* of a given *format* and returns datetime value expressed as if surrounded by `As Date()`, returning the date in `dd`Monyyyy format.

Table 6.2 Date-Time Functions *(Continued)*

Function	Explanation
`Second(`*`date-time`*`)`	Returns an integer representation for the second part of the *`date-time`* value supplied.
`Short Date(`*`date`*`)`	Returns a string representation for the *`date`* supplied, in the format MM/DD/YYYY, regardless of locale (for example, "02/29/2004").
`Time Of Day(`*`date`*`)`	Returns an integer representation for the time of day of the *`date-time`* supplied.
`Today()`	Returns the current date and time expressed as the number of seconds since midnight, 1 January 1904. No arguments are accepted, but the parentheses are still needed.
`Week Of Year(`*`date`*`, <rule_n>)`	Returns the week of the year as a date-time value. Three rules determine when the first week of the year begins. • With rule 1 (the default), weeks start on Sunday, with the first Sunday of the year being week 2. Week 1 is a partial week or empty. • With rule 2, the first Sunday begins with week 1, with previous days being week 0. • With rule 3, the ISO-8601 week number is returned. Weeks start on Monday. Week 1 is the first week of the year with four days in that year. It is possible for the first or last three days of the year to belong to the neighboring year's week number.
`Year(`*`date`*`)`	Returns an integer representation for the year of the specified *`date`*.

Examples of Common Date-Time Functions

You can use any function that returns seconds within a function that returns a date-time.

For example, if today is May 19, 2011 and the time is 11:37:52 AM, `Today()` returns the number of seconds, and the functions that follow show that number of seconds since the base time in different date-time formats:

```
Today()
   3666955847
Short Date( Today() );
   "03/13/2020"
```

```
Long Date( Today() );
   "Friday, March 13, 2020"
Abbrev Date( Today() );
   "3/13/2020"
MDYHMS( Today() );
   "03/13/2020 2:50:17 PM"
```

The date argument in parentheses can be seconds (or any function that returns seconds), or any date-time literal value. For example, both of the following expressions return the same value:

```
Long Date( 3388649872 );
Long Date( 19May2011 );
   "Friday, March 13, 2020"
```

Note: `Long Date()` and `Abbrev Date()` values are formatted according to your computer's regional settings.

Extract Parts of Dates

You can extract parts of date values using the functions `Month()`, `Day()`, `Year()`, `Day Of Week()`, `Day Of Year()`, `Week Of Year()`, `Time Of Day`, `Hour()`, `Minute()`, and `Second()`, which all return integers. If today is May 24th, 2011, each of the following examples returns the 144th day of the year:

```
Day of Year( Today() );
Day of Year( 24May2011 );
Day of Year( Date MDY( 5, 24, 2011 ) );
   144
```

Example

A data table column named Date contains date-time values that are formatted as `"m/d/y"`. You want to create a column that shows only the time. In the following script, the second column's formula extracts the time of day from the Date value in the first column.

```
New Table( "Assembly Tests",
   Add Rows( 1 ),
   New Column( "Date",
      Numeric, Continuous,
      Format( "m/d/y" ),
      Set Values( [3389083557] )
   ),
   New Column( "Time",
      Numeric, Continuous,
      Formula( Format( Time Of Day( :Date ), "h:m:s") )
   )
);
```

Figure 6.1 shows the result. Note that the time of day does not appear in the Date column, because the `Format` function applies the "m/d/y" format.

Figure 6.1 Example of Extracting the Time

	Date	Time
1	05/24/2011	12:05:57 PM

Rules for Determining the Week of the Year

`Week of Year()` returns the week of the year as a date-time value. Three rules determine when the first week of the year begins.

- With rule 1 (the default), weeks start on Sunday, with the first Sunday of the year being week 2. Week 1 is a partial week or empty.

```
Week Of Year( Date DMY( 19, 6, 2013 ), 1 );
  25
```

- With rule 2, the first Sunday begins with week 1, with previous days being week 0.

```
Week Of Year( Date DMY( 19, 6, 2013), 2 ) ;
  24
```

- With rule 3, the ISO-8601 week number is returned. Weeks start on Monday. Week 1 is the first week of the year with four days in that year. It is possible for the first or last three days of the year to belong to the neighboring year's week number.

```
Week Of Year( Date DMY( 19, 6, 2013 ), 3 );
  25
```

Arithmetic on Dates

You can perform the usual arithmetic operations with date-time data as with any other numeric data. One option is simple arithmetic, such as subtracting a number from a date-time value.

Another option is writing a formula to perform the arithmetic.

Example

The Date column in your data table shows when a customer uses his credit card to buy gas. You want to know how many days elapse between purchases. The following script creates a Days elapsed column. The formula in that column subtracts the Date value in the current row from that of the previous row.

```
New Table( "Gas Purchases",
  Add Rows( 3 ),
  New Column( "Date",
```

```
        Numeric,
        "Continuous",
        Format( "m/d/y" ),
        Set Values( [3392323200 3393532800 3394828800] )
    ),
    New Column( "Days elapsed",
        Formula(
        // returns a missing value for the first row
            If( Row() == 1,
                .,
                (:Date[Row()] - :Date[Row() - 1]) / In Days()
            )
        )
    )
);
```

Figure 6.2 shows the result.

Figure 6.2 Example of Calculating Date-Time Values

	Date	Days elapsed
1	07/01/2011	•
2	07/15/2011	14
3	07/30/2011	15

Time Intervals

The `In Minutes`, `In Hours`, `In Days`, `In Weeks`, and `In Years` functions are used to express time intervals in units rather than seconds. Each of these functions returns the number of seconds associated with a particular period of time. For example, the following expression returns the number of weeks between now and July 4, 2012.

```
(Date DMY( 04, 07, 2012 ) - Today()) / In Weeks();
  -208.225444775132
```

When the argument for the interval function is empty, JMP counts by 1. You can enter another number to change the count. For example, `In Years(10)` converts the interval to decades. The following expression returns the number of decades between now and December 31, 2037.

```
(Date DMY( 31, 12, 2037) - Today() ) / In Years( 10 );
  2.18927583529799
```

Two-Digit and Four-Digit Years

JMP applies its own algorithms for interpreting and displaying datetime strings rather than supporting operating system-specific datetime formats. However, JMP uses the date-time separators selected in the Region and Language control panel (Windows) or the Date & Time preferences (macOS) to interpret and display dates.

Two-digit years are interpreted according to the current system clock year and JMP rules. For example, when the year in a script is 11, and you run the script after 1990, the year shows as 2011.

```
Long Date( 25May11 );
  "Wednesday, May 25, 2011"
```

To avoid ambiguity, enter four-digit years. The following expression returns 1911 (rather than 2011) as indicated:

```
Long Date( 25May1911 );
  "Thursday, May 25, 1911"
```

Table 6.3 explains how JMP interprets two-digit years.

Table 6.3 How JMP Interprets Two-Digit Years

Two-Digit Year Value	When it is Evaluated	Result	Examples	Result
00–10	before 1990 (on Windows)	19__	enter 5 in year 1979	1905
	before or during 1990 (on macOS)			
	during or after 1990 (on Windows)	20__	enter 5 in year 1991	2005
	after 1990 (on macOS)			
11–89 (on Windows)	any time	current century	enter 13 in year 1988	1913
11–90 (on macOS)			enter 13 in year 2024	2013
90–99 (on Windows)	before 2011	19__	enter 99 in year 1999	1999
91–99 (on macOS)	during or after 2011	20__	enter 99 in year 2015	2099

Note: JMP always displays four-digit years regardless of the regional settings. If you need to show two-digit years, use character string functions. See "Types of Data".

Date-Time Values in Data Tables

Change Date-Time Input and Display Formats

In data tables, JMP can accept the input of date-time values in one format (the *input format*), store them internally as the number of seconds since the base date, and display them in a different date-time format. The `Informat()` and `Format()` functions give you this control.

- `Informat()` takes a string date-time value, defines the date format used in that string, and returns the date in `ddMonyyyy` format.

```
Informat( "19May2011 11:37:52 AM", "ddMonyyyy h:m:s" );
   19May2011:11:37:52
```

- `Format()` takes the number of seconds since the base date (or a date-time function that returns that number) and returns the date in the specified format.

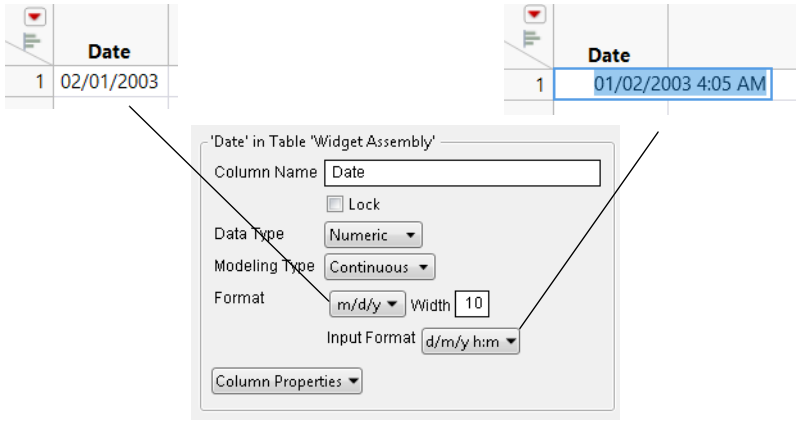
```
Format( 3388649872, "ddMonyyyy h:m:s" );
   "19May2011 11:37:52 AM"
Format( Today(), "ddMonyyyy h:m:s" );
   "19May2011 11:37:52 AM"
```

Suppose that you are entering dates into a column using the `d/m/y h:m` format, but you want to see the dates in the `m/d/y` format. `Input Format` defines the input format, and `Format` defines the display format. For example,

```
New Table( "Widget Assembly",
   Add Rows( 1 ),
   New Column( "Date",
      Numeric,
      "Continuous",
      Format( "m/d/y" ),
      Input Format( "d/m/y h:m" ),
      Set Values( [3126917100] )
   )
);
```

The `Format` and `Input Format` values are shown in the data table's column properties (Figure 6.3). Note that when you click in the cell to edit it, the date-time value appears in the input format. When you edit the value, or add a new value, the format specified in the data table column Format list is used to display the value.

Figure 6.3 Example of Date-Time Display and Input Values



Notes:

- In a script that converts a column from character to numeric, specify `Format()` and `Informat()` to prevent missing values. See "Convert Character Dates to Numeric Dates".
- The date-separator character on your computer might differ from the forward slash (/) character shown in the *Scripting Guide*.
- You can enter time values in 24-hour format (military time) or with AM or PM designators.

Table 6.4 describes the formats used as arguments in date-time functions or as data table formats. You can also use the formats for the *`format`* argument to a `Format` message to a data column. See "Set or Get Formats".

For descriptions of specific date-time functions, see the *JSL Syntax Reference*.

Notes:

- Date-time formats in which a colon or T separate the date and time always use a 24-hour format. Other formats use your computer's regional settings.
- If the format is unknown, an error is written to the log.

Table 6.4 Date-Time Formats

Type	Format argument	Example
Date only	`"m/d/y"`	`"01/02/1999"`
	`"mmddyyyy"`	`"01021999"`
	`"m/y"`	`"01/1999"`
	`"d/m/y"`	`"02/01/1999"`

Table 6.4 Date-Time Formats *(Continued)*

Type	Format argument	Example
	"ddmmyyyy"	"02011999"
	"ddMonyyyy"	"02Jan1999"
	"Monddyyyy"	"Jan021999"
	"y/m/d"	"1999/01/02"
	"yyyymmdd"	"19990102"
	"yyyy-mm-dd"	“1999-01-02“
	"yyyyQq"	1999Q1
Date and time	"m/d/y h:m"	"01/02/1999 13:01" "01/02/1999 1:01 PM"
	"m/d/y h:m:s"	"01/02/1999 13:01:55" "01/02/1999 1:01:55 PM"
	"d/m/y h:m"	"02/01/1999 13:01" "02/01/1999 1:01 PM"
	"d/m/y h:m:s"	"02/01/1999 13:01:55" "02/01/1999 1:01:55 PM"
	"y/m/d h:m"	‘1999/01/02 13:01’ ‘1999/01/02 1:01 PM’
	"y/m/d h:m:s"	‘1999/01/02 13:01:02 ‘1999/01/02 1:01:02 PM’
	"ddMonyyyy h:m"	"02Jan1999 13:01" "02Jan1999 1:01 PM"
	"ddMonyyyy h:m:s"	"02Jan1999 13:01:02" "02Jan1999 1:01:02 PM"
	"ddMonyyyy:h:m"	"02Jan1999:13:01" "02Jan1999:1:01 PM"
	"ddMonyyyy:h:m:s"	"02Jan1999:13:01:02" "02Jan1999:1:01:02 PM"
	"Monddyyyy h:m"	"Jan021999 13:01" "Jan021999 1:01 PM"
	"Monddyyyy h:m:s"	"Jan021999 13:01:02" "Jan021999 1:01:02 PM"

Table 6.4 Date-Time Formats *(Continued)*

Type	Format argument	Example
Day number and time	":day:hr:m"	"34700:13:01" ":33:001:01 PM"
	":day:hr:m:s"	"34700:13:01:02" ":33:001:01:02 PM"
	"h:m:s"	"13:01:02" "01:01:02 PM"
	"h:m"	"13:01" "01:02 PM"
	"yyyy-mm-ddThh:mm"	1999-01-02T13:01
	"yyyy-mm-ddThh:mm:ss"	1999-01-02T13:01:02
Duration	":day:hr:m"	"52:03:01" reads fifty-two days, three hours, and one minute
	":day:hr:m:s"	"52:03:01:30" reads fifty-two days, three hours, one minute, and thirty seconds
	"hr:m"	"17:37" reads seventeen hours and thirty-seven minutes
	"hr:m:s"	"17:37:04" reads seventeen hours, thirty-seven minutes, and 4 seconds
	"min:s"	"37:04" reads thirty-seven minutes and 4 seconds
Note: The following formats display the date-time according to your computer's regional settings. They are available only for the display of dates, not for date input in a data table. Examples are shown for the United States locale.		
Abbreviated date	"Date Abbrev"	(Display only) "01/02/1999"
Long date	"Date Long"	(Display only) "Saturday, January 02, 1999"
Locale date	"Locale Date"	(Display only) "01/02/1999"

Table 6.4 Date-Time Formats *(Continued)*

Type	Format argument	Example
Locale date and time	"Locale Date Time h:m"	(Display only) "01/02/1999 13:01" or "01/02/1999 01:01 PM"
	"Locale Date Time h:m:s"	(Display only) "01/02/1999 13:01:02" or "01/02/1999 01:01:02 PM"

Custom Date-Time Formatting with Format Patterns

Format patterns are strings that define a date-time format. Here's one example:

```
<YYYY></><MM></><DD> <hh><:><mm><:><ss><ampm>
```

The parts of the pattern in angle brackets are called *field descriptors*. They represent a value (such as <YYYY>, which is a four-digit year) or other date-time text (such as </>, which is a locale-specific date separator).

The conversion of a date-time value to a string and a string converted to a date-time value work as follows:

- When a date-time value is converted to a string, the field descriptors are replaced by the appropriate value or symbols.
- When a string is converted to a date-time value, a value is or matching symbols are expected to be found where each field descriptor is. Anything outside a format descriptor is treated as literal text. The literal text is replaced as-is when it is converted *to* a string, and is expected to be found where specified when converting *from* a string.

Table 6.5 Output for Custom Date-Time Formats

Format	Output
<YYYY></><MM></><DD> <hh><:><mm><:><ss><ampm>	2020/02/10 11:54:33 AM
<DayOfWeek>, <Month> <D>, <YYYY>	Wednesday, February 5, 2020
<Day>d <hh24>h <mm>m <ss>s	123d 02h 34m 56s

Note: <Day> is the day count, used as the most significant field of a duration. <Hour> and <Minute> field descriptors are also available.

Example of Outputting the Date-Time Value to the Log

The following script uses `Informat()` to read a string date-time value, defines the format pattern used in that string, and returns the date in `ddMonyyyy` format. The script then outputs the date-time value to the log.

```
dateVal = Informat( "June 16, 2016", "Format Pattern", "<Month> <D>, <YYYY>"
  );
  16Jun2016
```

Here's another way to specify the format pattern:

```
date2 = 16Jun2021;

  dayofwk2 = Format( date2, "Format Pattern", "<DayOfWeek>" );
  monthnew2 = Format( date2, "Format Pattern", "<Month>" );

  YYYYMMDD = Format( date2, "Format Pattern", "<YYYY></><MM></><DD>" );
  YYMMMDD = Format( date2, "Format Pattern", "<YY></><MMM></><DD>" );
  "21/Jun/16"
```

Example of Changing the Input Format of a Column

The following script shows how to change the input format of a column. The Input Format appears when you edit the value in a data table.

```
New Table( "temp_date",
  Add Rows( 2 ),
  New Column( "date1",
     Numeric,
     "Continuous",
     // specify how the value appears in the data table
     Format( "Format Pattern", "<Month> <D>, <YYYY>", 35 ),
     // specify how the value appears when you edit the value
     Input Format( "Format Pattern", "<YYYY>*<MM>*<DD>" ),
     Set Values( {"2018*10*31", "2018*09*09"} )
  ),
  New Column( "date2",
     Numeric,
     "Continuous",
     Format( "Format Pattern", "<YYYY>*<MM>*<DD>", 10 ),
     Set Values( {"2018*02*05", "2018*07*03"} )
  )
);
```

Note: See *Using JMP* for details about creating a format pattern in a data table rather than in a script.

Currency

JMP displays numbers as currency using the `Format()` function, which uses the following syntax:

```
Format(x,"Currency", <"currency code">, <decimal>, < << Use Locale(Boolean));
```

Where:

- *x* is a column or a number
- *"currency code"* is an International Standards Organization (ISO) 4217 code
- *decimal* is the number of decimal places

To illustrate the `Format` function:

```
Format( 12345.6, "Currency", "GBP", 3 );
  "£12,345.600"
```

By default, the operating system locale is used for the currency if the currency code symbol is omitted. For example, running the following script in a Japanese operating system formats the number with the yen symbol.

```
Format( 12345.6, "Currency", 3);
  "¥12,345.600"
```

To ignore the computer's locale and use a period for the decimal separator, include the `Use Locale(0)` argument:

```
Format( 12345.6, "Currency", 3, <<Use Locale( 0 ) );
```

Tip: If you change the region when JMP is open, restart JMP for the `Use Locale` setting to work properly.

If the currency code is not supported by JMP, the currency code string appears before the number.

```
Format( 12345.6, "Currency", "BBD", 3 );
  "BBD 12,345.600"
```

Table 6.6 lists the currencies supported in JMP.

Table 6.6 Currencies Supported in JMP

Code	Currency	Code	Currency	Code	Currency
AUD	Australian dollar	ILS	Israeli new shekel	RUB	Russian ruble
BRL	Brazilian real	INR	Indian rupee	SEK	Swedish krone

Table 6.6 Currencies Supported in JMP *(Continued)*

Code	Currency	Code	Currency	Code	Currency
CAD	Canadian dollar	JPY	Japanese yen	SGD	Singapore dollar
CHF	Swiss franc	KRW	South Korean won	THB	Thai baht
CNY	Chinese yuan	MXN	Mexican peso	TRY	New Turkish lira
COP	Colombian peso	MYR	Malaysian ringgit	TWD	New Taiwan dollar
DKK	Danish krone	NOK	Norwegian krone	USD	US dollar
EUR	Euro	NZD	New Zealand dollar	ZAR	South African rand
GBP	British pound	PHP	Philippine peso		
HKD	Hong Kong dollar	PLN	Polish zloty		

For more information about how to format currency, see "Formatting Numbers".

Hexadecimal and BLOB Functions

JMP can also handle binary (large) objects, commonly called BLOBs. The functions below convert between hexadecimal values, numbers, characters, and BLOBs. Some of the functions are covered in more detail following Table 6.7.

See the *JSL Syntax Reference* for more information.

Table 6.7 Hexadecimal and BLOB Functions

Syntax	Explanation

`Hex("text")`
`Hex("num")`
`Hex("blob")`
	Returns the hexadecimal codes for the characters in *text*, *number*, or *blob*.
	`Char To Hex` is an alias.

`Hex To BLOB("hexstring")`	Returns a BLOB representation of the hexadecimal code supplied as a quoted string.

Table 6.7 Hexadecimal and BLOB Functions *(Continued)*

Syntax	Explanation
`Hex To Char("hexstring", "encoding")`	Returns a character string that corresponds to the hexadecimal code supplied as a quoted string. The default encoding for the character string is `utf-8`. `utf-16le`, `utf-16be`, `us-ascii`, `iso-8859-1`, `ascii~hex`, `shift-jis`, and `euc-jp` are also supported.
`Hex to Number`	Returns the number that corresponds to the hexadecimal code suppled as a quoted string.
`Char To BLOB("string")` `Char To BLOB("string", "encoding")`	Converts a character string into a binary (BLOB). The default encoding for the blob is `utf-8`. `utf-16le`, `utf-16be`, `us-ascii`, `iso-8859-1`, `ascii~hex`, `shift-jis`, and `euc-jp` are also supported.
`BLOB To Char("blob")` `BLOB To Char("blob", "encoding")`	Converts binary data to a character string. The default encoding for the character string is `utf-8`. `utf-16le`, `utf-16be`, `us-ascii`, `iso-8859-1`, `ascii~hex`, `shift-jis`, and `euc-jp` are also supported.
`BLOB Peek("blob", offset, length)`	Returns a new BLOB that is a subset of the given BLOB that is *length* bytes long and begins at the *offset*. Note that the offset is 0-based.
`Matrix to BLOB(matrix, type, bytesEach, endian)`	Makes a BLOB from a matrix by converting the matrix elements to 1-byte, 2-byte, or 4-byte signed or unsigned integers or 4-byte or 8-byte floating point numbers.
`BLOB to Matrix(blob, "type", bytes, "endian", <nCols>)`	Creates a matrix by converting each byte in the `blob` to numbers.

`Hex(string)` returns the hexadecimal codes for each character in the argument. For example,

```
Hex( "Abc" );
```

returns

```
"416263"
```

since 41, 62, and 63 are the hexadecimal codes (in ASCII) for “A”, “b”, and “c”.

`Hex to Char(string)` converts hexadecimal to characters. The resulting character string might not be valid display characters. All the characters must be in pairs, in the ranges 0-9, A-Z, and a-z. Blanks and commas are allowed, and skipped. For example,

```
Hex To Char( "4142" );
```

returns

```
"AB"
```

since 41 and 42 are the hexadecimal equivalents of "A" and "B".

`Hex` and `Hex To Char` are inverses of each other, so

```
Hex To Char( Hex( "Abc" ) );
```

returns

```
"Abc"
```

`Hex To BLOB(string)` takes a string of hexadecimal codes and converts it to a binary object.

```
a = Hex To BLOB( "6A6B6C" );
Show( a );
   a = Char To BLOB("jkl", "ascii~hex")
```

`BLOB Peek(blob,offset,length)` extracts bytes as defined by the arguments from a blob.

```
b = BLOB Peek( a, 1, 2 );
Show( b );
   b = Char To BLOB("kl", "ascii~hex")
b = BLOB Peek( a, 0, 2 );
Show( b );
   b = Char To BLOB("jk", "ascii~hex")
b = BLOB Peek( a, 2 );
Show( b );
   b = Char To BLOB("l", "ascii~hex")
```

`Hex(blob)` converts a blob into hexadecimal.

```
c = Hex( a );
Show( c );
   c = "6A6B6C"
d = Hex To Char( c );
Show( d );
   d = "jkl"
```

`Concat(blob1,blob2)` or `blob1 || blob2` concatenates two blobs.

```
e = Hex To BLOB( "6D6E6F" );
Show( e );
f = a||e;
Show( f );
   e = Char To BLOB("mno", "ascii~hex")
   f = Char To BLOB("jklmno", "ascii~hex")
```

`Length(blob)` returns the number of bytes in a blob.

```
g = Length( f );
```

```
Show( g );
   g = 6
```

Note: When blobs are listed in the log, they are shown with the constructor function `Char To BLOB("...")`.

Any hex code outside the ASCII range (`space` to }, or hex 20 - 7D) is encoded as the three-character sequence [~][hexdigit][hexdigit]. For example,

```
h = Hex To BLOB( "19207D7E" );
Show( h );
i = Hex( h );
Show( i );
   h = Char To BLOB("~19 }~7E", "ascii~hex")
   i = "19207D7E"
```

`Char To BLOB(string)` creates a blob from a string, converting ~hex codes.
`BLOB To Char(blob)` creates a string with ~hex codes to indicate non-visible and non-ASCII codes.

Work with Character Functions

This section shows how to use some of the more complex character functions that are described in the *JSL Syntax Reference*.

Concat

In the `Concat` function, expressions yielding names are treated like character strings, but globals that have the name values are evaluated. The following example demonstrates that if you have a stored name value, you need to either use `Char` before storing it in a global, or `Name Expr` on the global name.

```
n = {abc};
c = n[1] || "def";
Show( c );
   "abcdef"

m = Expr( mno );
c = m || "xyz";
Show( c );
   Name Unresolved: mno in access or evaluation of 'mno' , mno/*###*/
// The /*###*/ characters indicate the location of an error in the script
   execution.

m = Expr( mno );
```

```
c = Name Expr( m ) || "xyz";
Show( c );
   "mnoxyz"

m = Char( Expr( mno ) );
c = m || "xyz";
Show( c );
   "mnoxyz"
```

`Concat Items()` converts a list of string expressions into a single string, with each item separated by a delimiter. If unspecified, the delimiter is a blank. Its syntax is

```
resultString = Concat Items ({list of strings}, <"delimiter string">);
```

For example,

```
a = {"ABC", "DEF", "HIJ"};
result = Concat Items(a, "/");
```

returns

```
"ABC/DEF/HIJ"
```

Alternatively,

```
result = Concat Items( a );
```

returns

```
"ABC DEF HIJ"
```

Munger

`Munger` works many different ways, depending on what you specify for its arguments:

```
Munger(string, offset, find | length, <replace>);
```

Table 6.8 Munger behaviors for various types of arguments

***Find*, *length*, and *replace* arguments**	**Example**
If you specify a string as the *find* and specify no *replace* string, `Munger` returns the position (after *offset*) of the first occurrence *find* string.	`Munger( "the quick brown fox", 1, "quick" );` `5`
If you specify a positive integer as the *length* and specify no *replace* string, `Munger` returns the characters from *offset* to *offset* + *length*.	`Munger( "the quick brown fox", 1, 5 );` `"the q"`

Table 6.8 Munger behaviors for various types of arguments *(Continued)*

***Find*, *length*, and *replace* arguments**	**Example**
If you specify a string as the *find* and specify a *replace* string, Munger replaces the first occurrence after *offset* of *text* with *replace*.	`Munger( "the quick brown fox", 1, "quick", "fast" );` `"the fast brown fox"`
If you specify a positive integer as the *length* and specify a *replace* string, Munger replaces the characters from *offset* to *offset* + *length* with *replace*.	`Munger( "the quick brown fox", 1, 5, "fast");` `"fastuick brown fox"`

If you specify a positive integer as the *length*, and *offset* + *length* exceeds the length of *text*, Munger either returns *text* from *offset* to the end or replaces that portion of *text* with the *replace* string, if it exists.	`Munger( "the quick brown fox", 5, 25);`
	`"quick brown fox"`
	`Munger( "the quick brown fox", 5, 25, "fast" );`
	`"the fast"`

If you specify zero as the *length* and specify no *replace* string, Munger returns a blank string.	`Munger( "the quick brown fox", 1, 0 );` `""`
If you specify zero as the *length* and specify a *replace* string, the string is inserted before the *offset* position.	`Munger( "the quick brown fox", 1, 0, "see " );` `"see the quick brown fox"`
If you specify a negative integer as the *length* value and specify no *replace* string, Munger returns all characters from the offset to the end of the string.	`Munger( "the quick brown fox", 5, -5 );` `"quick brown fox"`
If you specify a negative integer for *length* and specify a *replace* string, Munger replaces all characters from the offset to the end with the *replace* string.	`Munger( "the quick brown fox", 5, -5, "fast" );` `"the fast"`

Repeat

The `Repeat()` function makes copies of its first argument into a result. The second (and sometimes a third) argument is the number of repeats, where 1 means a single copy.

If the first argument evaluates to a character value or list, the result is that many copies.

```
Repeat( "abc", 2 );
```

```
   "abcabc"
Repeat( {"A"}, 2 );
   {"A","A"}
Repeat( {1, 2, 3}, 2 );
   {1,2,3,1,2,3}
```

If the first argument evaluates to a number or matrix, the result is a matrix. The second argument is the number of row repeats, and a third argument can specify the number of column repeats. If only two arguments are specified, the number of column repeats is 1.

```
Repeat( [1 2, 3 4], 2, 3 );
   [  1 2 1 2 1 2,
      3 4 3 4 3 4,
      1 2 1 2 1 2,
      3 4 3 4 3 4]
Repeat( 9, 2, 3 );
   [  9 9 9,
      9 9 9]
```

The `repeat` function is compatible with the function of the same name in the SAS/IML language, but is incompatible with the SAS character DATA step function, which repeats one more time than this function.

Substitute and Substitute Into

The `Substitute()` function returns a copy of a string with a replacement expression. The `Substitute Into()` function changes the original string with a replacement expression.

For example, consider the following script:

```
str1 = str2 = "All things considered";

// str3 holds the result of the Substitute, and str1 is not changed
str3 = Substitute( str1, "All", "Some" );

/* Substitute Into returns nothing, so str4 is missing,
and str2 is changed and now holds the result of Substitute Into */
str4 = Substitute Into( str2, "All", "Some" );
Show( str1, str2, str3, str4 );
```

This script returns the following output:

```
str1 = "All things considered";
str2 = "Some things considered";
str3 = "Some things considered";
str4 = .;
```

Regular Expressions

A regular expression is a specification of a pattern frequently used to clean up or extract pieces of data. You can search for a pattern and replace it with a different string or extract specific parts of the string. Define the pattern in the `Regex()` or `Regex Match()` function.

- "Regex"
- "Regex Match"
- "Special Characters in Regular Expressions"
- "Escaped Characters in Regular Expressions"
- "Greedy and Reluctant Regular Expressions"
- "Backreferences and Capturing Groups"
- "Lookaround Assertions"

Regex

`Regex()` searches for a *`pattern`* within a *`source`* string and returns a string. It simply identifies a pattern in a string or transforms a string into another string.

```
Regex(source, pattern, (<replacement string>, <GLOBALREPLACE>), <format>,
  <IGNORECASE>);
```

`IGNORECASE` disregards case. `GLOBALREPLACE` repeats the match until the entire string is processed. `format` is a backreference to the matched group. `Regex()` returns missing if the match fails.

Example of Matching a String

`bus|car` is the regular expression (in quotation marks because it is a string). The expression means match "bus" or "car".

```
sentence = "I took the bus to work.";
vehicle = Regex( sentence, "bus|car" );
  "bus"
```

Examples of Replacing a String

The third optional argument in `Regex()` is a specification of the result string. The default value, `\0`, is a backreference to everything that was matched by the regular expression. In the preceding example, the word "bus" is matched in `sentence`. The default third argument, `\0`, replaces the entire sentence with "bus".

A more interesting variation uses parentheses to create additional backreferences.

```
sentence = "I took the bus to work.";
Regex( sentence, "(.*) bus (.*)", "\1 car \2" );
   "I took the car to work."
```

The `(.*)` before and after `bus` are part of the regular expression. The parentheses create a capturing group. The `.` matches any character. The * matches zero or more of the previous expression. As a result, the first parenthesis pair matches everything before `bus`, and the second parenthesis pair matches everything after `bus`. The third argument, `\1 car \2`, reassembles the text; it leaves out `bus` and substitutes `car`.

See "Backreferences and Capturing Groups".

Example of Global Replacement

`GLOBALREPLACE` changes the behavior of `Regex()`. If the match succeeds, the entire source string is returned with substitutions made for each place where the pattern matches. If there are no matches, an unchanged source string is returned.

```
sentence = "I took the red bus followed by the blue bus to get to work
   today.";
Regex( sentence, "bus", "car", GLOBALREPLACE);
   "I took the red car followed by the blue car to get to work today."
```

You can also use backreferences. This example starts with a different sentence.

```
sentence = "I took the red bus followed by the blue car to get to work
   today.";
Regex(
   sentence,
   "(\w*) (bus|car)",
   "bicycle (not \2) that was \1",
   GLOBALREPLACE
);
   "I took the bicycle (not bus) that was red followed by the bicycle (not car)
      that was blue to get to work today."
```

The `\w*` matches zero or more word characters and becomes backreference 1 because of the parentheses. `bus|car` becomes backreference 2 because of the parentheses. The third argument, `bicycle (not \2) that was \1`, describes how to build the substitution text for the part of the source text that was matched.

Notice how the backreferences can be used to swap data positions. This might be useful for swapping the position of first names and last names.

Regex Match

`Regex Match()` returns an empty list with zero elements if the match fails. If the match succeeds, the first list is the text of the entire match (backreference `0`). The second list is the text that matches backreference 1, and so on.

```
Regex Match(source, pattern, <replacement>|<MATCHCASE>, <NULL>)
```

Unlike `Regex()`, `Regex Match()` is case insensitive. Include `MATCHCASE` for a case-sensitive match. Include `NULL` if you want to match case but there is no replacement text.

Example of Parsing Name-Value Pairs

The following example parses pairs of names and values.

```
Regex Match(
   "person=Fred id=77 friend= favorite=tea",
   "(\w+)=(\S*) (\w+)=(\S*) (\w+)=(\S*) (\w+)=(\S*)"
);
   {"person=Fred id=77 friend= favorite=tea", "person", "Fred", "id", "77",
      "friend", "", "favorite", "tea"}
```

The `\w+` matches one or more word characters. The `\S*` matches zero or more characters that are not spaces. In the resulting JSL list, the field names (`person, id, friend, favorite`) and their corresponding values (`Fred, 77, "", tea`) are separate strings.

If the first argument to `Regex Match()` is a variable and a third argument specifies the replacement value, the matched text is replaced in the variable.

Comparing Regex and Regex Match

`Regex()` and `Regex Match()` match a pattern in a given string but return different results. To transforms your string into another string, use `Regex()`. To identify the substrings that match specific parts of the pattern, use `Regex Match()`.

This example shows the efficiency of `Regex Match()` compared to `Regex()`. The `source` is a list of six strings. The goal is to extract portions of those six strings into the subject, verb, and object columns of a data table.

Figure 6.4 Final Data Table

	subject	verb	object
1	cat	ate	chicken
2	dog	chased	cat
3	ralph	like	mary
4	girl	pets	dog
5	cat	chased	dog

```
source = {"the cat ate the chicken", "the dog chased the cat", "did ralph like
   mary", "the girl pets the dog",
"these words are strange", "the cat was chased by the dog"};

dt = New Table( "English 101", // create the data table
   New Column( "subject", character ),
   New Column( "verb", character ),
   New Column( "object", character )
);
// iterate through the strings in the list
For( i = 1, i <= N Items( source ), i++,

   // assign the result of each match to matchList
   matchList = Regex Match(
      source[i],

   /* scan each string, match zero or more characters
   and one item in each group */

   ".*?(cat|dog|ralph|girl).*?(ate|chased|like|pets).*?(chicken|cat|mary|dog)
   "
   );

/* If matchList has zero items (string 5), don't add a row
to the table. Put each matched string in separate
data table cells. */
   If( N Items( matchList ) > 0,
      dt << Add Rows( 1 );
      dt:subject = matchList[2]; // match the first open parenthesis
      dt:verb = matchList[3]; // match the second open parenthesis
      dt:object = matchList[4]; // match the third open parenthesis
   );
);
```

`Regex Match()` returns `{"the cat was chased by the dog", "cat", "chased", "dog"}` in a single try with each answer in a separate string. Compare this example to a similar one using `Regex()`, which returns one answer at a time and builds the final string using backreferences.

```
For( i = 1, i <= N Items( source ), i++,
   s = Regex( source[i],
   ".*?(cat|dog|ralph|girl).*?(ate|chased|like|pets).*?(chicken|cat|mary|dog)
   ", "\1" ); // match an item in the first group
   v = Regex( source[i],
   ".*?(cat|dog|ralph|girl).*?(ate|chased|like|pets).*?(chicken|cat|mary|dog)
   ", "\2" ); // match an item in the second group
```

```
      o = Regex( source[i],
      ".*?(cat|dog|ralph|girl).*?(ate|chased|like|pets).*?(chicken|cat|mary|dog)
      ", "\3" ); // match an item in the third group
      If( !Is Missing( s ) & !Is Missing( v ) & !Is Missing( o ),
         dt << Add Rows( 1 );
         dt:subject = s; // return the match for \1
         dt:verb = v; // return the match for \2
         dt:object = o; // return the match for \3
      );
   );
```

Backreferences are discussed in "Backreferences and Capturing Groups".

Special Characters in Regular Expressions

Special characters are commonly used in regular expressions. The period is a special character that matches one instance of the specified character. It must be escaped with a backslash to be interpreted as a period. In the following expression, the period is replaced with an exclamation point.

```
Regex( "Bicycling makes traveling to work fun.", "\.", "!", GLOBALREPLACE );
  "Bicycling makes traveling to work fun!"
```

Table 6.9 describes the special characters and provides examples.

Table 6.9 Special Characters in Regular Expressions

\	• Precedes a literal character. `<\/a>` interprets the forward slash literally in the end HTML anchor tag. • Precedes an escape sequence. `\n` matches a newline character.
^	Matches the beginning of a string, not including the newline character. `^apple` matches "apple" at the beginning of a string.
$	Matches the end of a string, not including the newline character. `apple$` matches "apple" at the end of a string.
.	Matches any single character including a newline character. `.apple` matches any single character and then "apple".
|	Represents a logical OR to separate alternative values. `(apple|orange|banana)` matches "apple", "orange", or "banana".

Table 6.9 Special Characters in Regular Expressions *(Continued)*

?	Matches zero or one instance. `apple (pie)?` matches one or more instances of "pie".
*	Matches zero or more instances.
+	Matches one or more instances.
( )	Encloses a sub-expression. `(apple|orange|banana)` matches "apple", "orange", or "banana". `^(\w+)` matches the beginning of a line and then one or more word characters.
[ ]	Encloses an expression that matches set of characters. `[\s]` matches a whitespace character or a digit. `[a-z0-9]` matches "a" through "z" and numbers "0" through "9".
{ }	Encloses an expression that represents repetition. `apple{3}` repeats three times. `apple{3,}` repeats at least three times as many times as possible. `apple{3, 10}` repeats three times but no more than 10 times. Append a question mark to indicate repeating as few times as possible. For example, `apple{3,}?` repeats at least three times as few times as possible.

Escaped Characters in Regular Expressions

The backslash in a regular expression precedes a literal character. You also escape certain letters that represent common character classes, such as `\w` for a word character or `\s` for a space. The following example matches word characters (alphanumeric and underscores) and spaces.

```
Regex(
   "Are you there, Alice?, asked Jerry.", // source
   "(here|there).+(\w+).+(said|asked)(\s)(\w+)\." ); // regular expression
   "there, Alice?, asked Jerry."
```

`(here|there).+`	Matches "there", a comma, and a space.
`(\w+)`	Matches "Alice".

`.+`	Matches "?, ".
`(said|asked)(\s)`	Matches "asked" followed by a space. Without the space, the match would end here; "asked" is followed by a space in the source string.
`(\w+)\.`	Matches "Jerry" and a period.

Table 6.10 describes the escaped characters supported in JMP. `\C`, `\G`, `\X`, and `\z` are not supported.

Table 6.10 Escaped Characters

`\\`	single backslash
`\A`	start of a string
`\b`	word boundary. The zero-length string between `\w` and `\W` or `\W` and `\w`.
`\B`	not at a word boundary
`\c`*X*	ASCII control character
`\d`	single digit `[0-9]`
`\D`	single character that is NOT a digit `[^0-9]`
`\E`	stop processing escaped characters
`\l`	match a single lowercase letter `[a-z]`
`\L`	single character that is not lowercase `[^a-z]`
`\Q`	ignore escaped characters until `\E` is found
`\r`	carriage return
`\s`	single whitespace character
`\S`	single character that is NOT white space
`\u`	single uppercase character `[A-Z]`
`\U`	single character that is not uppercase `[^A-Z]`
`\w`	word character `[a-zA-Z0-9_]`
`\W`	single character that is NOT a word character `[^a-zA-Z0-9_]`

Table 6.10 Escaped Characters *(Continued)*

\x*00*-\x*FF*	hexadecimal character
\x{*0000*}-\x{*FFFF*}	Unicode code point
\Z	end of a string before the line break

Greedy and Reluctant Regular Expressions

The ?, *, and + operators are greedy by default. They match as many of the preceding character as possible. The ? operator makes them reluctant; ?? matches 0, then 1 if needed; +? matches 1 and then additional characters; *? matches 0 and then additional characters.

The following example starts at the letter n and compares it to the first \d (digits) in the pattern. No digit matches. Because the pattern does not begin with ^ (start of line), the matcher advances to u. The process repeats until the 3 matches the first \d and the 2 matches the second \d.

```
Regex( "number=32.5", "\d\d" );
   "32"
```

Change the pattern to use the greedy + (match one or more).

```
Regex( "number=324.5", "\d+" );
   "324"
```

The preceding example begins much the same, but as soon as the 3 is found and the \d matches, the + greedily matches the 2 and the 4.

Usually, the greedy behavior makes pattern matching faster because the string is consumed sooner. Sometimes a reluctant behavior is better. Adding the ? after the * or + changes them from greedy to reluctant.

```
Regex( "number=324.5", "\d+?" );
   "3"
```

Here, + requires at least one match of a digit character, but ? changes it from "as many as possible" to "as few as possible". It stopped after the 3 because the pattern was satisfied.

Compare the following results.

Greedy:

```
Regex( "number=324.5", "(\d+)(\d+)\.", "first=\1 second=\2" );
   "first=32 second=4"
```

Reluctant:

```
Regex( "number=324.5", "(\d+?)(\d+?)\.", "first=\1 second=\2" );
   "first=3 second=24"
```

In the greedy example above, the matcher greedily matched 3, 2, and 4 for the first \d+. The matcher then had to give back the 4 so that the second \d+ could match something. The reluctant example followed a different path to get a different answer. Initially, the second value was 2, but the pattern could not match the period to the 4, so the second \d+? reluctantly matched the 4 as well.

Use the Reluctant Match for Speed

The greedy and reluctant matches usually produce the same result but not always. See the previous section. One reason you might need the reluctant match is for speed. Suppose that you have a million-character string that begins "The quick fox…" and you want to find the word before "fox". You might write the following expression and expect \1 to contain "quick".

```
The (.+) fox
```

\1 might contain "quick" eventually, after the .+ grabs the million characters to the end of the string and then gives them up, one at a time, until "fox" is found. If there is more than one "fox", it will be the last fox, not this one. To speed it up and make sure we get the first fox, add the ? operator.

```
The (.+?) fox
```

The ? advances one character at a time to get past "quick" and find the first "fox". This method is much faster than going too far.

Typically, the + or * operator is applied to a more restrictive expression such as \d* to match a run of digits, and greedy is faster than reluctant.

Aside from the multiple fox possibility, greedy and reluctant eventually get the same answer. Using the right operator speeds up the match. The right one might be greedy, or it might be reluctant. It depends on what is being matched.

The greedy .* finds the last fox after backing up.

```
Regex(
   "The quick fox saw another fox eating grapes",
   "The (.*) fox",
   "\1"
);
   "quick fox saw another"
```

The reluctant .*? stops on the first fox.

```
Regex(
   "The quick fox saw another fox eating grapes",
   "The (.*?) fox",
   "\1"
);
   "quick"
```

The greedy `.*` has to back up a lot. There is no second fox.

```
Regex(
  "The quick fox saw another animal eating grapes",
  "The (.*) fox",
  "\1"
);
  "quick"
```

The greedy word character match is an even better choice for this problem.

```
Regex(
  "The quick fox saw another fox eating grapes",
  "The (\w*) fox",
  "\1"
);
  "quick"
```

Backreferences and Capturing Groups

A regular expression can consist of patterns grouped in parentheses, also known as capturing groups. In `([a-zA-Z])\s([0-9])`, `([a-zA-Z])` is the first capturing group; `([0-9])` is the second capturing group.

Use a backreference to replace the pattern matched by a capturing group. In Perl, these groups are represented by the special variables `$1`, `$2`, `$3`, and so on. (`$1` indicates text matched by the first parenthetical group.) In JMP, use a backslash followed by the group number (`\1`, `\2`, `\3`).

The following example includes a third argument that specifies the replacement text and backreferences.

```
Regex(
  "   Are you there, Alice?, asked Jerry.", // source
  " (here|there).+ (\w+).+(said|asked) (\w+)\.", // regular expression
"  I am \1, \4, replied \2." ); // optional format argument
  "  I am there, Jerry, replied Alice."
```

`" I am \1,`	Creates the text “I am”, a space, and then the first matched pattern, “there”.
`\4,`	Creates the text “Jerry” with the fourth matched pattern `(\w+)`.
`replied \2."`	Creates the text “replied” and a space. Matches “Alice.” with the second matched pattern `(\w+)`.

Lookaround Assertions

Lookaround assertions check for a pattern but do not return that pattern in the results. *Lookaheads* look forward for a pattern. *Lookbehinds* look back for a pattern.

Negative Lookahead Example

Negative lookaheads check for the absence of a pattern before a specific pattern. `?!` indicates a negative lookahead. The following expression matches a comma *not* followed by a number or space and replaces the pattern with a comma and space:

```
Regex( "one,two 1,234 cat,dog,duck fish, and chips, to go",
   ",(?!\d|\s)", ", ", GLOBALREPLACE );
   "one, two 1,234 cat, dog, duck fish, and chips, to go"
```

Positive Lookahead Example

Positive lookaheads check for the presence of a pattern before a specific pattern. `?=` indicates a positive lookahead. The following expression has the same result as the preceding negative lookahead but matches a comma followed by any lowercase character:

```
Regex( "one,two 1,234 cat,dog,duck fish, and chips, to go",
   ",(?=[a-z])", ", ",GLOBALREPLACE );
   "one, two 1,234 cat, dog, duck fish, and chips, to go"
```

Positive Lookbehind Example

In this example, the positive lookbehind regular expression matches the “ssn=” or “salary=” keywords without including the keyword in the matched text. The matched text is the string of characters that consists of zero or more dollar signs, digits, and hyphens.

```
data = "name=bill salary=$5 ssn=123-45-6789 age=13,name=mary salary=$6
   ssn=987-65-4321 age=14";
redacted = Regex(data, "(?<=(ssn=)|(salary=))[$\d-]*", "###", GLOBALREPLACE);
   "name=bill salary=### ssn=### age=13,name=mary salary=### ssn=### age=14"
```

Here is another way to get the same result using a backreference substitution. `((ssn=)|(salary=))` is the capturing group. `"\1"` is the backreference to that group.

```
data = "name=bill salary=$5 ssn=123-45-6789 age=13,name=mary salary=$6
   ssn=987-65-4321 age=14";
redacted = Regex(data, "((ssn=)|(salary=))[$\d-]*", "\1###", GLOBALREPLACE);
   "name=bill salary=### ssn=### age=13,name=mary salary=### ssn=### age=14"
```

Backreferences are discussed in “Backreferences and Capturing Groups”.

Pattern Matching

Pattern matching in JSL is a flexible method for searching and manipulating strings.

You define and use pattern variables just like any JMP variable:

```
i = 3; // a numeric variable
a = "Ralph"; // a character variable
t = textbox("Madge"); // a display box variable
p = ( "this" | "that" ) + patSpan(" ")
     + ( "car" | "bus" ); // a pattern variable
```

When the above statement executes, `p` is assigned a pattern value. The pattern value can be used either to construct another pattern or to perform a pattern match. The `patSpan` function returns a pattern that matches a span of characters specified in the argument; `patSpan("0123456789")` matches runs of digits.

```
p2 = "Take " + p + "."; // using p to build another pattern
If( Pat Match( "Take this bus.", p2 ), // performing a match
   Print( "matches" ),
   Print( "no match" )
);
```

Sometime all you need to know is that the pattern matched the source text, as above. Other times, you might want to know what matched; for example, was it a bus or a car?

```
p = ("this" | "that") + Pat Span( " " ) + ("car" | "bus") >?
vehicleType; // conditional assignment ONLY if pattern matches
If( Pat Match( "Take this bus.", p ),
   // do not use vehicleType in the ELSE because it is not set
   Show( vehicleType ),
   Print( "no match" )
);
```

You could pre-load `vehicleType` with a default value if you do not want to check the outcome of the match with an `if`. The >? conditional assignment operator has two arguments, the first being a pattern and the second a JSL variable. >? constructs a pattern that matches the pattern (first argument) and stores the result of the match in the JSL variable (second argument) after the pattern succeeds. Similarly, >> does not wait for the pattern to succeed. As soon (and as often) as the >> pattern matches, the assignment is performed.

```
findDelimString = Pat Len( 3 ) >> beginDelim + Pat Arb() >? middlePart
+Expr( beginDelim );
testString = "SomeoneSawTheQuickBrownFoxJumpOverTheLazyDog'sBack";
rc = Pat Match( testString, findDelimString, "<<<" || middlePart || ">>>" );
Show( rc, beginDelim, middlePart, testString );
```

The above example shows a third argument in the `patMatch` function: the replacement string. In this case, the replacement is formed from a concatenation (`||` operator) of three strings. One of the three strings, `middlePart`, was extracted from the `testString` by `>?` because the replacement cannot occur unless the pattern match succeeds (`rc == 1`).

Look at the pattern assigned to `findDelimString`. It is a concatenation of 3 patterns. The first is a `>>` operator that matches 3 characters and assigns them to `beginDelim`. The second is a `>?` operator that matches an arbitrary number of characters and, when the entire match succeeds, assigns them to `middlePart`. The last is an unevaluated expression, consisting of whatever string is in `beginDelim` at the time the pattern is executing, *not* at the time the pattern is built. Just like `expr()`, the evaluation of its argument is postponed. That makes the pattern hunt for two identical three letter delimiters of the middle part.

Other pattern functions might be faster and represent the problem that you are trying to solve better than writing a lot of alternatives; for example, `"a"|"b"|"c"` is the same as `Pat Any("abc")`. The equivalent example for `Pat Not Any("abc")` is much harder. Similar to `Pat Span` (above), `Pat Break("0123456789")` matches up to, but not including, the first number.

Here is a pattern that matches numbers with decimals and exponents and signs. It also matches some degenerate cases with no digits; look at the pattern assigned to digits.

```
digits = Pat Span( "0123456789" ) | "";

number = (Pat Any( "+-" ) | "") >? signPart + (digits) >? wholePart + ("."
+digits | "") >? fractionPart + (Pat Any( "eEdD" ) + (Pat Any( "+-" ) | "")
 + digits | "") >? exponentPart;

If( Pat Match( "-123.456e-78", number ),
   Show( signPart, wholePart, fractionPart, exponentPart )
);
```

Parsing Strings in Fixed Fields

Sometimes data is in fixed fields. The `Pat Tab()`, `Pat R Tab()`, `Pat Len()`, `Pat Pos()`, and `Pat R Pos()` functions make it easy to split out the fields in a fixed field string. `Pat Tab()` and `Pat R Tab()` work from the left and right end of the string and take a number as their argument. They succeed by matching forward to the specified tab position. For example:

```
p = Pat Pos(10) + Pat Tab(15);
```

`Pat Pos(10)` matches the null string if it is in position 10. So at match time, the matcher works its way forward to position 10, then `Pat Tab(15)` matches text from the current position (10) forward to position 15. This pattern is equivalent to `Pat Pos(10)+patLen(5)`. Another example:

```
p = Pat Pos(0) + Pat R Tab(0);
```

This example matches the entire string, from 0 characters from the start to 0 characters from the end. the `Pat Rem()` function takes no argument and is shorthand for `Pat R Tab(0)`; it means the remainder of the string. Pattern matching can also be anchored to the beginning of the string like this:

```
Pat Match( "now is the time", Pat Len(15) + Pat R Pos(0), NULL, ANCHOR );
```

The above pattern uses `NULL` rather than a replacement value, and `ANCHOR` as an option. Both are uppercase, as shown. `NULL` means that no replacement is done. `ANCHOR` means that the match is anchored to the beginning of the string. The default value is `UNANCHORED`.

Patterns can be built up like this, but this is *not* recursive:

```
p = "a" | "b"; // matches one character
p = p + p; // two characters
p = p + p; // four characters
Pat Match( "babb", Pat Pos(0) + p + Pat R Pos(0) );
```

A recursive pattern refers to its current definition using `Expr()`:

```
p = "<" + Expr(p) + "*" + Expr(p) + ">" | "x";
Pat Match( "<<x*<x*x>>*x>", Pat Pos(0) + p + Pat R Pos(0) );
```

Remember, `expr()` is the procrastination function; when the pattern is assigned to the variable `p`, `expr()` delays evaluating its argument (`p`) until later. In the next statement, `patMatch` performs the pattern match operation, and each time it encounters `expr()`, it looks for the current value of the argument. In this example, the value does not change during the match). So, if `p` is defined in terms of itself, how can this possibly work?

`p` consists of two alternatives. The right hand choice is easy: a single letter x. The left side is harder: <p*p> . Each `p` could be a single letter x, since that is one of the choices `p` could match, or it could be <p*p>. The last few example have used `patPos(0) + ... + patRPos(0)` to make sure the pattern matches the entire source text. Sometimes this is what you want, and sometimes you would rather the pattern match a subtext. If you are experimenting with these examples by changing the source text, you probably want to match the entire string to easily tell what was matched. The result from `Pat Match` is 0 or 1.

This example uses “Left” recursion:

```
x =  Expr(x) + "a" | "b"; // + binds tighter than |
```

If the pattern is used in `FULLSCAN` mode, it eventually uses up all memory as it expands. By default, the `patMatch` function does not use `FULLSCAN`, and makes some assumptions that allow the recursion to stop and the match to succeed. The pattern matches either a “b”, or anything the pattern matches followed by an “a”.

```
rc = Pat Match( "baaaaa", x );
```

Patterns and Case

Unlike regular expressions, pattern matching is case insensitive. To force case sensitivity, you can add the named argument MATCHCASE to either `Pat Match()` or `Regex Match()`. For example:

```
string = "abcABC";

result = Regex Match( string, Pat Regex( "[aBc]+" ) );
Show( string, result );
      / *returns string = "abcABC"
   result = {"abcABC"}*/

result = Regex Match( string, Pat Regex( "[cba]+" ), NULL, MATCHCASE );
Show( string, result );
      /* returns string = "abcABC"
   result = {"abc"}*/
```

Troubleshooting Patterns

When a pattern isn't working the way you expect, use `Log()` to send pieces of the match to the JMP log. The following example shows how to debug pattern matching and rewrite a script to account for the white space between words in the phrase "the quick brown fox".

```
// greedy match for one or more characters except for white space
word = Pat Repeat( Pat Not Any( " " ) );

Pat Match(
   "the quick brown fox",
   word >> a +
   word >> b +
   word >> c +
   word >> d
);

Show( a, b, c, d );
   a = "qu";
   b = "i";
   c = "c";
   d = "k";
```

The output is probably not what you want, given the name of the pattern *word*. None of the words in the pattern were matched.

The `Show( a, b, c, d )` expression is a good first step. If it doesn't tell you how JMP got that answer, try inserting `Log()`:

```
word = Pat Repeat( Pat Not Any( " " ) );
```

```
Pat Match(
   "the quick brown fox",
   (word >> a) >> Log( "a" ) +
   (word >> b) >> Log( "b" ) +
   (word >> c) >> Log( "c" ) +
   (word >> d) >> Log( "d" )
);

Show( a, b, c, d );
   0(a) the (the number 0 is the position of the match)

   0(a) th (the "a" is an aribtrary identifier you supplied to the log)

   2(b) e

   ...
   4(a) quick
   4(a) quic
   ...
```

With this extra information, you can now see into the workings of the match, how the word pattern does, in fact, match a word, at least initially. But then it backs up and tries a smaller amount of the word so that the next word pattern can match some non-whitespace characters. However, there's no provision in this pattern to match the white space between words. And that's exactly what it did (find four consecutive words of one or more non-whitespace characters). Maybe the following script is what you need:

```
word = Pat Repeat( Pat Not Any( " " ) );
gap = Pat Repeat( Pat Any( " " ), 1, 999, GREEDY );

Pat Match(
   "the quick brown fox",
   word >> a + gap +
   word >> b + gap +
   word >> c + gap +
   word >> d
);

Show( a, b, c, d );
   a = "the";
   b = "quick";
   c = "brown";
   d = "fox";
```

Chapter 7

Data Structures

Working with Collections of Data

JSL provides these basic data structures that can hold a variety of data in a single variable:

- A list holds a number of other values, including nested lists and expressions.
- A matrix is a row-by-column table of numbers.
- An associative array maps keys to values, which can be almost any other type of data.

Contents

Lists . . . 193
Evaluate Lists . . . 193
Assignments with Lists . . . 195
Perform Operations in Lists . . . 195
Find the Number of Items in a List . . . 195
Subscripts . . . 195
Locate Items in a List . . . 197
List Operators and Functions . . . 198
Iterate through a List . . . 201
Concatenate Lists . . . 202
Insert a Nested List in an Existing List . . . 203
Index into a List with Another List . . . 203
Send Messages to Lists . . . 203
Matrices . . . 204
Construct Matrices . . . 204
Subscripts . . . 206
Inquiry Functions . . . 210
Comparisons, Range Checks, and Logical Operators . . . 210
Numeric Operations . . . 211
Concatenation . . . 214
Transpose . . . 214
Indexing into a Matrix or List with Another Matrix or List . . . 215
Matrices and Data Tables . . . 216
Matrices and Reports . . . 218
Loc Functions . . . 219
Ranking and Sorting . . . 221
Special Matrices . . . 221
Inverse Matrices and Linear Systems . . . 226
Decompositions and Normalizations . . . 230
Build Your Own Matrix Functions . . . 235
Statistical Examples . . . 236
Associative Arrays . . . 240
Create Associative Arrays . . . 240
Work with Associative Arrays . . . 243
Applications for Associative Arrays . . . 246
Associative Arrays in Graph Theory . . . 248
Associative Arrays in Set Operations . . . 251

Lists

Lists are containers to store items, such as the following:

- numbers
- variables
- character strings
- expressions (for example, assignments or function calls)
- matrices
- nested lists

Create a list in one of the following ways:

- use the `List` function
- use `{ }` curly braces

Examples

Use the `List()` function or curly braces to create a list that includes numbers and variables:

```
x = List(1, 2, b);
x = {1, 2, b};
```

A list can contain text strings, nested lists, and function calls:

```
{"Red", "Green", "Blue", {1, "true"}, sqrt( 2 )};
```

You can place a variable into a list and assign it a value at the same time:

```
x = {a = 1, b = 2};
```

Evaluate Lists

When you run a script that contains a list, a copy of the list is returned. The items inside the list are not evaluated.

```
b = 7;
x = {1, 2, b, Sqrt( 3 )};
Show( x );
   x = {1, 2, b, Sqrt(3)};
```

To evaluate items in a list, use the `Eval List()` function.

```
b = 7;
x = {1, 2, b, Sqrt( 3 )};
c = Eval List( x );
   {1, 2, 7, 1.73205080756888}
```

When using a list of variables that reference lists, you will need to use an Eval() function.

Consider the following example, where you have a variable called fullMonth that is a list of 12 items.

```
::fullMonth = {January, February, March, April, May, June, July, August,
  September, October, November, December};
::abbrevMonth = {Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec};
::dow = {Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday};
::levels = {Very Low, Low, Medium Low, Medium, Medium High, High, Very High};
::feel = {Strongly Disagree, Disagree, Neutral, Indifferent, Agree, Strongly
  Agree};
::rating = {Failing, Unacceptable, Very Poor, Poor, Bad, Acceptable, Average,
  Good, Better, Very Good, Excellent, Best};
::mlist = {::fullMonth, ::abbrevMonth, ::dow, ::levels, ::feel, ::rating};
N Items(::fullMonth); // returns 12 because fullMonth has 12 items.
N Items(::mlist[1]); /* returns this message: "N Items() argument must be a
  list" because ::mlist[1] is a variable (fullMonth) and fullMonth is a
  list.*/
```

When you add the Eval() function:

```
N Items( Eval( ::mlist[1] ) );
  12
/* returns 12 because::mlist[1] holds ::fullMonth, and evaluating it
returns ::fullMonth's list.*/
```

In a loop, you need the Eval() function in the nested loop to evaluate the variable's contents. For example:

```
For( g = 1, g <= N Items( ::mlist ), g++,
    For( i = 1, i <= N Items(::mlist[g]), i++,
        Show( ::mlist[g][i] )
    )
);
// returns the message "N Items() argument must be a list."
```

To fix the problem, add the Eval() function:

```
::mlist = {::fullMonth, ::abbrevMonth, ::dow, ::levels, ::feel, ::rating};
For( g = 1, g <= N Items( ::mlist ), g++,
    For( i = 1, i <= N Items( Eval( ::mlist[g] ) ), i++,
        Show( ::mlist[g][i] )
    )
);
```

Assignments with Lists

Create a list to assign values to variables.

Examples

```
{a, b, c} = {1, 2, 3}; // assign 1 to a, 2 to b, and 3 to c
{a, b, c}--; // decrement a, b, and c
{{a}, {b, c}}++; // increment a, b, and c by 1
mylist = {1, log( 2 ), e()^pi(), height[40]}; // store the expressions
```

Perform Operations in Lists

In lists, you can perform operations.

```
a = {{1, 2}, 3, {4, 5}};
b = {{10, 20}, 30, {40, 50}};
c = a + b;
   c = {{11, 22}, 33, {44, 55}}
```

Find the Number of Items in a List

To determine the number of items in a list, use the `N Items()` function.

```
x = {1, 2, y, Sqrt( 3 ), {a, b, 3}};
N = N Items( x );
Show( n );
   n = 5;
```

Subscripts

Subscripts extract specified items from a list. Use a list as a subscript to return multiple items from a list.

Note: JSL starts counting from 1, so the first element in a list is [1], not [0] as in some other languages.

Examples

List a contains four items.

```
a = {"bob", 4, [1, 2, 3], {x, y, z}};
Show( a[1] );
   a[1] = "bob";
Show( a[{1, 3}] );
```

```
   a[{1, 3}] = {"bob", [1, 2, 3]};
a[2] = 5; // assign 5 to the second list item
```

You can also use subscripts to select or change items in a list:

```
Show( a );
   a = {"bob", 5, [1, 2, 3], {x, y, z}};
c = {1, 2, 3};
c[{1, 2}] = {4, 4};
// c[{1, 2}] = 4 produces the same result
Show( c );
   c = {4, 4, 3};
```

When you have assignments or functions in a list, you can use a quoted name for the subscript to extract the value.

```
x={sqrt( 4 ), log( 3 )};
xx= {a = 1, b = 3, c = 5};
x["sqrt"];
   4
xx["b"];
   3
```

The name must be in quotation marks, or else JMP tries to evaluate it and use its value. The following example shows the values of the second item in the list, rather than the value of `a` in the list.

```
a = 2;
Show( xx[a] );
   xx[a] = b = 3;
```

Notes:

- Multiple left-side subscripts (for example, `a[i][j] = value` where `a` contains a list of things that are subscriptable) are allowed in the following circumstances:
  - Each level except the outermost level must be a list. So, in the example above, `a` must be a list but `a[i]` can be anything subscriptable.
  - Each subscript except the last must be a number. So, in the example above, `i` must be a number, but `j` could be a matrix or list of indices.
- Subscripting can be done to any level of nesting, such as the following:

  ```
  a[i][j][k][l][m][n] = 99;
  ```

Note: To get a value in the matrix of an expression column, use double subscripts, as in `:Discrim Data Matrix[row()][i]`.

Locate Items in a List

Use the `Loc()` function or the `Contains()` function to find values in a list:

```
Loc( list, value );
Contains( list, value );
```

`Loc()` and `Contains()` return the positions of the values. `Loc()` returns the results in a matrix, and `Contains()` returns the results as a number.

Notes:

- The `Loc` function returns each occurrence of a repeated value. `Contains()` returns only the first occurrence of a repeated value.
- If the value is not found, the `Loc` function returns an empty matrix and `Contains()` returns a zero.
- To assess whether an item is in a list, use `Loc()` and `Contains()` with `>0`. A returned value of zero means that the item is not in the list. A returned value of 1 means that the item is in the list at least once.

Note: For more information about matrix manipulation and a description of the equivalent `Loc()` command for matrices, see "Matrices".

Examples

```
nameList = {"Katie", "Louise", "Jane", "Jane"};
numList = {2, 4, 6, 8, 8};
```

Search for the value `"Katie"` in the `nameList`:

```
Loc( nameList, "Katie" );
   [1]
Contains( nameList, "Katie" );
   1
```

Search for the value `"Erin"` in the `nameList`:

```
Loc( nameList, "Erin" );
   []
Contains( nameList, "Erin" );
   0
```

Search for the number `8` in the `numList`:

```
Loc( numList, 8 );
   [4, 5]
Contains( numList, 8 );
   4
```

Find out if the number `5` exists in the `numList`:

```
NRow( Loc( numList, 5 ) ) > 0;
   0
Contains( numList, 5 ) > 0;
   0
```

List Operators and Functions

Table 7.1 describes the list operators and their syntax.

Table 7.1 List Operators

Operator and Function		Syntax	Explanation
	As List()	As List(*matrix*)	Returns the matrix as a list. A matrix with multiple columns is returned as a list of lists, one list per row.
= += -= *= /= ++ --	Assign() Add To() SubtractTo() MultiplyTo() DivideTo() Post Increment() Post Decrement()	{*list*} = {*list*} {*list*} += *value* {*list*} -= {*list*} ...	If the target of an assignment operator is a list and the value to be assigned is a list, then it assigns item by item. The ultimate values in the left list must be L-values (in other words, names capable of being assigned values). **Notes:** • If you want to test equality of lists, use ==, not =. • JMP does not have pre-increment or pre-decrement operators. Instead, use the Add To() operator, which is (+=), or the Subtract To() operator, which is (-=).
||=	Concat To()	Concat To(*list1*, *list2*, ...)	Inserts the second and subsequent lists at the end of the first list.
||	Concat()	Concat(*list1*, *list2*, ...);	Returns a copy of the first list with any additional lists inserted at after it.

Table 7.1 List Operators *(Continued)*

Operator and Function	Syntax	Explanation
Eval List()	Eval List(*list*)	Returns a list of the evaluated expressions inside list. See "Evaluate Lists".
Insert Into()	Insert Into(*list*, *x*, <*i*>)	Inserts a new item (*x*) into the *list* at the given position (*i*). If *i* is not given, the item is added to the end of the list. This function does change the original list.
Insert()	*list* = Insert(*list*, *x*, <*i*>)	Returns a copy of the *list* with a new item (*x*) inserted into the *list* at the given position (*i*). If *i* is not given, the item is added to the end of the list. This function does not change the original list.
Is List()	Is List(*arg*)	Returns true (1) if *arg* is a classical list (in other words, one that would result from the construction by List(*items*) or {*items*}) and returns false (0) otherwise. An empty list is still a list, so IsList({ }) returns true. If miss=., then IsList(miss) returns false, not missing.
{ } List	List(*a*, *b*, *c*) {*a*, *b*, *c*}	Constructs a list from a set of items. An item can be any expression, including other lists. Items must be separated by commas. Text should either be enclosed in double quotation marks (" ") or stored in a variable and called as that variable.
N Items	N Items(*list*)	Returns the number of elements in the *list* specified. Can be assigned to a variable.

Table 7.1 List Operators *(Continued)*

Operator and Function	Syntax	Explanation
`Remove From()`	`Remove From(list, <i>, <n>)`	Deletes *n* items from the *list*, starting from the indicated position (*i*). If *n* is omitted, the item at *i* is deleted. If *n* and *i* are omitted, the item at the end is removed. This function does change the original list.
`Remove()`	`Remove(list, <i>, <n>)`	Returns a copy of the *list* with the *n* items deleted, starting from the indicated position (*i*). If *n* is omitted, the item at *i* is deleted. If *n* and *i* are omitted, the item at the end is removed. This function does not change the original list.
`Reverse Into()`	`Reverse Into(list)`	Reverses the order of the items in the *list*. This function does change the original list.
`Reverse()`	`Reverse(list)`	Returns a copy of the *list* with the items in reverse order. This function does not change the original list.
`Shift Into()`	`Shift Into(list, <n>)`	Shifts *n* items from the front of the *list* to the end of the *list*. If *n* is omitted, the first item is moved to the end of the list. This function does change the original list.
`Shift()`	`Shift(list, <n>)`	Returns a copy of the *list* with *n* items shifted from the front of the list to the end of the *list*. If *n* is omitted, the first item is moved to the end of the list. This function does not change the original list.
`Sort Ascending()`	`Sort Ascending(list)`	Returns a copy of the *list* sorted in ascending order. This function does not change the original list.

Table 7.1 List Operators *(Continued)*

Operator and Function		Syntax	Explanation
	`Sort Descending()`	`Sort Descending(`*`list`*`)`	Returns a copy of the *`list`* sorted in descending order. This function does not change the original list.
	`Sort List Into()`	`Sort List Into(`*`list`*`)`	Sorts the *`list`* in ascending order. This function does change the original list.
	`Sort List()`	`Sort List(`*`list`*`)`	Returns a copy of the *`list`* sorted in ascending order. This function does not change the original list.
`[ ]`	`Subscript()`	*`list`*`[`*`i`*`]` *`x`* `=` *`list`*`[`*`i`*`]` *`list`*`[`*`i`*`]` `=` *`value`* *`a`*`[`*`b`*`,` *`c`*`]` `Subscript(`*`a`*`,` *`b`*`,` *`c`*`)`	Subscripts for lists extract the i^{th} item from the *`list`*. Subscripts can in turn be lists or matrices.
	`Substitute()`	`Substitute(`*`list`*`,` *`pattExpr1`*`,` *`replExpr1`*`, ...)`	Returns a copy of a string, list, or expression, replacing instances of each pattern expression with the corresponding replacement expression. See "Substitute and Substitute Into".
	`Substitute Into()`	`Substitute Into(`*`list`*`,` *`pattExpr1`*`,` *`replExpr1`*`, ...)`	Changes a string, list, or expression, replacing instances of each pattern expression with the corresponding replacement expression. **Note:** The *`list`* or expression must be a variable. See "Substitute and Substitute Into".

Iterate through a List

Iterate through a list to do something with each value or look for a particular value. The following script looks at each item in the list. If the item in the list is less than or equal to 10, it is replaced with its square.

```
x = {2, 12, 8, 5, 18, 25};
n = N Items( x );
For( i = 1, i <= n, i++,
```

```
    If( x[i] <= 10,
       x[i] = x[i] ^ 2
    )
);
Show( x );
   x = {4, 12, 64, 25, 18, 25};
```

You can use `Loc()` to locate the items in the new list that are equal to 25:

```
Loc( x, 25 );
   [4, 6] // The fourth and sixth items in the list are equal to 25.
```

Concatenate Lists

Join two or more lists into one list using `Concat()` or the `||` operator. The original lists are not changed.

The following example uses `Concat()` to join lists *a* and *b*:

```
a = {1, 2};
b = {7, 8, 9};
Concat( a, b );
   {1, 2, 7, 8, 9}
```

The following example joins the same lists using the `||` operator:

```
{1, 2} || {7, 8, 9}
   {1, 2, 7, 8, 9}
```

Lists of different types can be concatenated (for example, lists that contain character strings and numbers).

```
d = {"apples", "bananas"};
e = {"oranges", "grapes"};
f = {1, 2, 3};
Concat( d, e, f);
   {"apples", "bananas", "oranges", "grapes", 1, 2, 3}
```

Join two or more lists and replace the first list with the combined list using `Concat to()` or the `||=` operator. The following example performs concatenation in place using `Concat to()`:

```
d = {"apples", "bananas"};
e = {"peaches", "pears"};
Concat to(d,e);
Show( d );
   d = {"apples", "bananas", "peaches", "pears"}
```

The following example joins the same lists using the `||=` operator:

```
d = {"apples", "bananas"};
e = {"peaches", "pears"};
d||=e;
```

```
Show( d );
  d = {"apples", "bananas", "peaches", "pears"}
```

Insert a Nested List in an Existing List

To insert a nested list in an existing list, specify the placement of the list:

```
list2 = {{"a", "b"}, {"c", "d"}};
list2[3] = {"apple", "banana"}; // specify a list as the third item
Show( list2 );
  list2 = {{"a", "b"}, {"c", "d"}, {"apple", "banana"}};
```

If you do not know the number of items in a list, use `N Items()`. In this example, an empty list is inserted at the end of the list.

```
list1 = {"a", "b"};
list1[N Items( list1 ) + 1] = {};
Show( list1 );
  list1 = {"a", "b", {}};
```

Index into a List with Another List

Indexing into a a list with another list is a powerful technique for making scripts fast and small. See “Indexing into a Matrix or List with Another Matrix or List”.

Send Messages to Lists

You can send a single message to a list instead of having to send individual messages. Suppose that you want to create a Bivariate report with a By variable. The script returns a list of the bivariate objects, one for each level of the By variable. Send a single `Fit Line` message to the list to fit a line to all of the Bivariate reports.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
/* return a list of bivariate objects:
{Bivariate[], Bivariate[], Bivariate[], Bivariate[], Bivariate[],
Bivariate[]}; */
biv = dt << Bivariate( X( dt:weight ), Y( dt:height ), By( dt:age ) );

biv << Fit Line; /* send the Fit Line message to
each Bivariate object in the list */
```

Troubleshooting Messages Sent to a List

Any message that is sent to a list is sent to each object in the list. If the message saves a file, it's performing the save for each object in the list to the *same file*. This means that only the last object is actually saved in a file.

Suppose that you have a list of three control charts and you send a message (such as `Save Interactive HTML`) to the list. The message is sent to each member of the list in turn. First, it saves an HTML page with the first control chart; then it saves a second HTML page with the same name that contains the second control chart; and then it saves a third HTML page with the same name that contains only the third control chart. You end up with a page that contains only the third control chart.

The solution is to send the message to a single container higher up the tree that contains all control charts. To do this, rewrite the following line of code:

```
obj << Save Interactive HTML( "Control Charts.html" );
```

`Top Parent` treats the first object as the single container:

```
tp = obj[1] << Top Parent();
tp << Save Interactive HTML( "Control Charts.html" );
```

Matrices

A matrix is a rectangular array of numbers that are arranged in rows and columns. Use matrices to store numbers and perform calculations on those numbers using matrix algebra.

Note the following for this section:

- Matrices are represented with an uppercase bold variable (for example, **A**).
- A matrix with one row or one column is a *vector* (or more specifically, a *row vector* or a *column vector* respectively).
- For clarity, we represent matrices that are vectors with lowercase bold letters (such as **x**).
- A *scalar* is a numeric value that is not in a matrix.

Construct Matrices

Note the following when creating matrices:

- Place matrix literals in square brackets. `[...]`
- Matrix values can contain decimal points, can be positive or negative, and can be in scientific notation.

- Separate items across a column with blank spaces. You can use any number of blank spaces.
- Separate rows with a comma.

For constructing more advanced matrices, see "Special Matrices".

Examples

Create matrix **A** with 3 rows and 2 columns:

```
A = [1 2, 3 4, 5 6];
```

R is a row vector and **C** is a column vector:

```
R = [10 12 14];
C = [11, 13, 15];
```

B is a 1-by-1 matrix, or a matrix with one row and one column:

```
B = [20];
```

E is an empty matrix:

```
E = [];
```

The following examples show two ways of creating empty matrices:

```
J( 0,4 );
  [](0,4)

[]( 5,0 )
  [](5,0)
```

Specifying the number of rows and columns in an empty matrix is optional. JMP creates the matrix as necessary. For example, the following script defines an empty matrix, iterates over each row in the data table, and returns the heights and weights in a matrix.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
data = []( 0, 2 );
For Each Row( data |/= Matrix( {{dt:height, dt:weight}} ) );
Show( data );
  data = [59 95, 61 123, 55 74,...]
```

A script can return an empty matrix. In Big Class.jmp, the following expression looks for rows in which age equals 8, finds none, and returns an empty matrix:

```
a = dt << Get Rows Where( age == 8 );
Show( a );
  a = [](0,1);
```

Construct Matrices from Lists

If you want to convert lists into a matrix, use the `Matrix()` function. A single list is converted into a column vector. Two or more lists are converted into rows.

Create a column vector from a single list:

```
A = Matrix( {1, 2, 3} );
   [1,2,3]
```

Create a matrix from a list of lists. Each list is a row in the matrix.

```
A = Matrix( {{1, 2, 3}, {4, 5, 6}, {7, 8, 9}} );
   [1 2 3,
   4 5 6,
   7 8 9]
```

Construct Matrices from Expressions

To construct matrices from expressions, use `Matrix()`. Elements must be expressions that resolve to numbers.

```
A = Matrix( {4 * 5, 8 ^ 2, Sqrt( 9 )} );
   [20, 64, 3]
```

Subscripts

Use the subscript operator (`[ ]`) to pick out elements or submatrices from matrices. The `Subscript()` function is usually written as a bracket notation after the matrix to be subscripted, with arguments for rows and columns.

Single Element

The expression `A[i,j]` extracts the element in row `i`, column `j`, returning a scalar number. The equivalent functional form is `Subscript(A,i,j)`.

```
P = [1 2 3, 4 5 6, 7 8 9];
P[2, 3]; // returns 6
Subscript( P, 2, 3 ); // returns 6
```

Assign the value that is in the third row and the first column in the matrix **A** (which is 5) to the variable `test`.

```
A = [1 2, 3 4, 5 6];
test = A[3, 1];
Show( test );
   test = 5;
```

Matrix or List Subscripts

To extract a sub-matrix, use matrix or list subscripts. The result is a matrix of the selected rows and columns. The following expressions select the second and third rows with the first and third columns.

```
P = [1 2 3, 4 5 6, 7 8 9];
P[[2 3],[1 3]]; // matrix subscripts
P[{2, 3},{1, 3}]; // list subscripts
```

Both of these methods provide the following output:

```
[4 6,
7 9]
```

Single Subscripts

A single subscript addresses matrices as if all the rows were connected end-to-end in a single row. This makes the double subscript `A[i,j]` the same as the single subscript `A[(i-1)*ncol(A)+j]`.

Examples

```
Q = [2 4 6, 8 10 12, 14 16 18];
Q[8]; // same as Q[3,2]
  16
```

The following examples all return the column vector [10, 14, 18]:

```
Q = [2 4 6, 8 10 12, 14 16 18];
Q[{5, 7, 9}];
Q[[5 7 9]];
ii = [5 7 9];
Q[ii];
ii = {5, 7, 9};
Q[ii];
Subscript( Q, ii );
```

This script returns the values 1 through 9 from the matrix **P** in order:

```
P = [1 2 3, 4 5 6, 7 8 9];
For( i = 1, i <= 3, i++,
  For( j = 1, j <= 3, j++,
    Show( P[i, j] )
  )
);
```

Delete Rows and Columns

Deleting rows and columns is accomplished by assigning an empty matrix to that row or column.

```
A[k, 0] = [];  // deletes the kth row
A[0, k] = [];  // deletes the kth column
```

Select Whole Rows or Columns

A subscript argument of zero selects all rows or columns.

```
P = [1 2 3, 4 5 6, 7 8 9];
```

Select column 2:

```
P[0, 2];
   [2, 5, 8]
```

Select columns 3 and 2:

```
P[0, [3, 2]];
   [3 2, 6 5, 9 8]
```

Select row 3:

```
P[3, 0];
   [7 8 9]
```

Select rows 2 and 3:

```
P[[2, 3], 0];
   [4 5 6, 7 8 9]
```

Select all columns and rows (the whole matrix):

```
P[0, 0];
   [1 2 3, 4 5 6, 7 8 9]
```

Assignment through Subscripts

You can change values in matrices using subscripts. The subscripts can be single indices, matrices or lists of indices, or the zero index representing all rows or columns. The number of selected rows and columns for the insertion must either match the dimension of the inserted argument, or the argument can be inserted repeatedly into the indexed positions.

Examples

Change the value in row 2, column 3 to 99:

```
P = [1 2 3, 4 5 6, 7 8 9];
P[2, 3] = 99;
Show( P );
```

```
P=[1 2 3, 4 5 99, 7 8 9]
```

Change the values in four locations:

```
P = [1 2 3, 4 5 6, 7 8 9];
P[[1 2], [2 3]] = [66 77, 88 99];
Show( P );
   P=[1 66 77, 4 88 99, 7 8 9]
```

Change three values in one column:

```
P = [1 2 3, 4 5 6, 7 8 9];
P[0, 2] = [11, 22, 33];
Show( P );
   P=[1 11 3, 4 22 6, 7 33 9]
```

Change three values in one row:

```
P = [1 2 3, 4 5 6, 7 8 9];
P[3, 0] = [100 102 104];
Show( P );
   P=[1 2 3, 4 5 6, 100 102 104]
```

Change all the values in one row to the same value:

```
P = [1 2 3, 4 5 6, 7 8 9];
P[2, 0] = 99;
Show( P );
   P=[1 2 3, 99 99 99, 7 8 9]
```

Operator Assignment

You can use operator assignments (such as +=) on matrices or subscripts of matrices. For example, the following statement adds 1 to the *i* - *j*th element of the matrix:

```
P = [1 2 3, 4 5 6, 7 8 9];
P[1, 1] += 1;
Show( P );
   P=[2 2 3,
      4 5 6,
      7 8 9];

P[1, 1] += 1;
Show( P );
   P=[3 2 3,
   4 5 6,7 8 9];
```

Ranges of Rows or Columns

If you are working with a range of subscripts, use the `Index()` function or `::` to create matrices of ranges.

```
T1 = 1 :: 3; // create the vector [1 2 3]
T2 = 4 :: 6; // create the vector [4 5 6]
T3 = 7 :: 9; // create the vector [7 8 9]
T = T1 |/ T2 |/ T3; // concatenate the vectors into a single matrix
T[1 :: 3, 2 :: 3]; // refer to rows 1 through 3, columns 2 and 3
   [2 3, 5 6, 8 9]
T[Index( 1, 3 ), Index( 2, 3 )]; // equivalent Index function
   [2 3, 5 6, 8 9]
```

Inquiry Functions

The `NCol()` and `NRow()` functions return the number of columns and rows in a matrix (or data table), respectively:

```
NCol( [1 2 3, 4 5 6] ); // returns 3 (for 3 columns)
NRow( [1 2 3, 4 5 6] ); // returns 2 (for 2 rows)
```

To determine whether a value is a matrix, use the `Is Matrix()` function, which returns a 1 if the argument evaluates to a matrix.

```
A = [20, 64, 3];
B = {20, 64, 3};
Is Matrix( A ); // returns 1 for yes
Is Matrix( B ); // returns 0 for no
```

Comparisons, Range Checks, and Logical Operators

JMP's comparison, range check, and logical operators work with matrices and produce matrices of elementwise Boolean results. You can compare conformable matrices.

```
A < B; // less than
A <= B; // less or equal
A > B; // greater than
A >= B; // greater or equal
A == B; // equal to
A != B; // not equal to
A < B < C; // continued comparison (range check)
A | B; // logical OR
A & B; // logical AND
( A & !B ) | ( !A & B ); // exlusive OR
!( A & B ) & ( A | B ); // exlusive OR
```

You can use the `Any()` or `All()` functions to summarize matrix comparison results. `Any()` returns a 1 if any element is nonzero. `All()` returns a 1 if all elements are nonzero.

```
[2 2] == [1 2]; // returns [0 1], therefore:
All( [2 2] == [1 2] ); // returns 0
```

```
Any( [2 2] == [1 2] ); // returns 1
```

`Min()` or `Max()` return the minimum or maximum element from the matrix or matrices given as arguments.

```
A = [1 2 3, 4 5 6, 7 8 9, 10 11 12];
B = [0 1 2, 2 1 0, 0 1 1, 2 0 0];
Min( A ); // returns 1
Max( A ); // returns 12
Min( A, B ); // returns 0
```

Numeric Operations

You can perform numeric operations (such as subtraction, addition, and multiplication) on matrices. Most statistical methods are expressed in compact matrix notation and can be implemented in JSL.

For example, the following expression uses matrix multiplication and inversion to illustrate least squares regression:

$$b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Implement this equation through the following JSL expression:

```
b = Inv( X'*X )*X'*y;
```

Basic Arithmetic

You can perform the following basic arithmetic on matrices:

- addition
- subtraction
- multiplication
- division (multiplying by the inverse)

Note: The standard multiply operator is a matrix multiplier, not an elementwise multiplier.

To perform matrix multiplication, use one of the following methods:

- `*` operator
- `Multiply()` function
- `Matrix Mult()` function

To perform matrix division, use one of the following methods:

- `/` operator

- `Divide()` function

Note the following about matrix multiplication and division:

- Remember that while multiplication or division of scalars is commutative (ab = ba), multiplication or division of matrices is *not* commutative.
- When one of the two elements is a scalar, elementwise multiplication or division is performed.
- To use elementwise multiplication, use `:*` or the `EMult()` function.
- To use elementwise division, use `:/`, or the equivalent `EDiv()` function.

Examples

```
A = [1 2 3, 4 5 6, 7 8 9, 10 11 12];
B = [0 1 2, 2 1 0, 0 1 1, 2 0 0];
C = [1 2 3 4, 4 3 2 1, 0 1 0 1];
D = [0 1 2, 2 1 0, 1 2 0];
```

Matrix addition:

```
R = A + B;
   [1 3 5,
   6 6 6,
   7 9 10,
   12 11 12]
```

Matrix subtraction:

```
R = A - B;
   [1 1 1,
   2 4 6,
   7 7 8,
   8 11 12]
```

Matrix multiplication (inner product of rows of **A** with columns of **C**):

```
R = A * C;
   [9 11 7 9,
   24 29 22 27,
   39 47 37 45,
   54 65 52 63]
```

Matrix division (equivalent to `A*Inverse(D)`):

```
R = A / D;
   [1.5 0.5 0,
   3 2 0,
   4.5 3.5 0,
   6 5 0]
```

Matrix elementwise multiplication:

```
R = A :* B;
```

```
[0 2 6,
8 5 0,
0 8 9,
20 0 0]
```

Matrix scalar multiplication:

```
R = C * 2;
   [2 4 6 8,
   8 6 4 2,
   0 2 0 2]
```

Matrix scalar division:

```
R = C / 2;
   [0.5 1 1.5 2,
   2 1.5 1 0.5,
   0 0.5 0 0.5]
```

Matrix elementwise division (division by zero results in missing values):

```
R = A :/ B;
   [. 2 1.5,
   2 5 .,
   . 8 9,
   5 . .]
```

Numeric (Scalar) Functions on Matrices

Numeric functions work elementwise on matrices. Most of the pure numeric functions can be applied to matrices, resulting in a matrix of results. You can mix conformable matrix and scalar arguments.

Examples of numeric functions include the following:

- `Sqrt()`, `Root()`, `Log()`, `Exp()`, `^ Power()`, `Log10()`
- `Abs()`, `Mod()`, `Floor()`, `Ceiling()`, `Round()`, `Modulo()`
- `Sine,()`, `Cosine()`, `Tangent()`, `ArcSine()`, and other trigonometry functions.
- `Normal Distribution()`, and other probability functions.

Example

```
A = [1 2 3, 4 5 6, 7 8 9, 10 11 12];
B = Sqrt( A ); // elementwise square root
   [1 1.414213562373095 1.732050807568877,
      2 2.23606797749979 2.449489742783178,
      2.645751311064591 2.82842712474619 3,
      3.16227766016838 3.3166247903554 3.464101615137754]
```

Concatenation

The `Concat()` function combines two matrices side by side to form a larger matrix. The number of rows must agree. A double vertical bar (`||`) is the infix operator, equivalent for horizontal concatenation.

```
Identity( 2 ) || J( 2, 3, 4 );
   [1 0 4 4 4, 0 1 4 4 4]

B = [1, 1];
B || Concat( Identity( 2 ), J( 2, 3, 4 ) );
   [1 1 0 4 4 4, 1 0 1 4 4 4]
```

The `VConcat()` function stacks two matrices on top of each other to form a larger matrix. The number of columns must agree. A vertical-bar-slash (`|/`) is the infix operator, equivalent for vertical concatenation.

```
Identity( 2 ) |/ J( 3, 2, 1 );
// or VConcat( Identity( 2 ),J( 3, 2, 1 ) );
   [1 0, 0 1, 1 1, 1 1, 1 1]
```

Both `Concat()` and `VConcat()` support concatenating to empty matrices, scalars, and lists.

```
a=[];
a || [1]; // returns [1]
a || {2}; // returns [2]
a || [3 4 5]; // returns [3 4 5]
```

There are two in place concatenation operators: `||=` and `|/=` . They are equivalent to the `Concat To()` and `V Concat To()` functions, respectively.

- `a || = b` is equivalent to `a = a || b`
- `a |/ = b` is equivalent to `a = a | /b`

Transpose

The `Transpose()` function transposes the rows and columns of a matrix. A back-quote (`` ` ``) is the postfix operator, equivalent to `Transpose()`. In matrix notation, `Transpose()` is expressed as the common prime or superscript-T notation (A′ or A^T).

```
A = [1 2 3, 4 5 6, 7 8 9, 10 11 12];
A`
   [1 4 7 10,
   2 5 8 11,
   3 6 9 12]
Transpose( [1 2, 3 4] );
   [1 3, 2 4]
```

Indexing into a Matrix or List with Another Matrix or List

Indexing into a matrix (or a list) with another matrix (or list) is a powerful technique for making scripts fast and small. JMP 13 and later handles one edge case differently.

Consider the following example of using a matrix as an index into a data column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
tall rows = Loc( :height << Get Values() >= 67 ); // [25, 27, 30, 37, 39, 40]
:height[tall rows]; // [68, 69, 67, 68, 68, 70]
N Rows( :height[tall rows] ); // 6
```

The second line is the key. `:height << Get Values()` returns the values of the height column as a matrix. The comparison `>= 67` returns a matrix of 0s and 1s corresponding to where the expression is false (0) or true (1). The `Loc()` function returns the locations of the 1s within the previous matrix.

In this case, you get [25, 27, 30, 37, 39, 40] as the row numbers where the `height` column is greater than or equal to 67. That's pretty handy by itself, but the real power comes from being able to use that matrix as the index into a column or other matrix as in the third and fourth lines.

The alternative to this technique is to write a `For()` loop to iterate through every row explicitly. JMP still loops through the rows in either case, but the looping is much faster if done internally rather than in JSL. And the JSL is more compact without the explicit loop.

You can do the same thing with a matrix variable instead of a data column reference:

```
m = :height << Get Values();
tall rows = Loc( m >= 67 ); // [25, 27, 30, 37, 39, 40]
m[tall rows]; // [68, 69, 67, 68, 68, 70]
N Rows( m[tall rows] ); // 6
```

Consider the same examples with a different height cutoff (70 instead of 67):

```
tall rows = Loc( :height << Get Values() >= 70 ); // [40]
:height[tall rows]; // [70]
N Rows( :height[tall rows] ); // 1

m = :height << Get Values();
tall rows = Loc( m >= 70 ); // [40]
m[tall rows]; // [70]
N Rows( m[tall rows] ); /* error in JMP 12 and prior versions
otherwise, returns 1 */
```

Notice that `m[tall rows]` returned 70 instead of [70]. In many cases, a 1x1 matrix and a number are treated the same way, but not always. The `N Rows()` function is one example where it makes a difference. The indexing is too aggressively simplifying the 1x1 matrix into a number.

One work-around is to call `Is Matrix()` before using `N Rows()`. An alternative is to concatenate an empty matrix to the result, which creates a matrix whether the source is a number or a matrix:

```
m = m |/ []
```

In JMP 13 and later, the 1x1 matrix is maintained so that the result of the index operation is a matrix whenever the index itself is a matrix. The same principle applies to list indexing. Data table column indexing has always followed this principle.

Matrices and Data Tables

You can move information between a matrix and a JMP data table. You can use matrix algebra to perform calculations on numbers that are stored in JMP data tables, and you can save the results back to JMP data tables.

Move Data into a Matrix from a Data Table

These sections describe how to move data from a data table into a matrix.

Move All Numeric Values

The `Get As Matrix()` function generates a matrix containing all of the numeric values in a data table or column:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
A = dt << Get As Matrix;
   [12 59 95,
   12 61 123,
   12 55 74, ...]
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = Column( "Height" );
A = col << Get As Matrix;
   [59, 61, 55, 66, 52, ...]
```

Move All Numeric Values and Character Columns

The `Get All Columns As Matrix()` function returns the values from all columns of the data table in a matrix, including character columns. Character columns are numbered according to the alphanumeric order, starting at 1.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
A = dt << Get All Columns As Matrix;
   [21 12 1 59 95,
   28 12 1 61 123, ...]
```

Move Only Certain Columns

To get certain columns of a data table, use column list arguments (names or characters).

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
x = dt << Get As Matrix( {"height", "weight"} );
```

or

```
x = dt << Get As Matrix( {height, weight} );
   [59 95, 61 123, 55 74, ...]
```

Currently Selected Rows

To get a matrix of the currently selected row numbers in the data table:

```
dt << Get Selected Rows;
```

Note: If no rows are selected, the output is an empty matrix.

Find Rows Where

To see a matrix of row numbers where the expression is true:

```
dt << Get Rows Where( expression );
```

For example, the following script returns the row numbers where the sex is male (M):

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
A = dt << Get Rows Where( Sex == "M" );
   [6, 7, 8, 12, 13, 14, ...]
```

Move Data into a Data Table from a Matrix

This section describes how to move data from a matrix into a data table.

Move a Column Vector

The `Set Values()` function copies values from a column vector into an existing data table column:

```
col << Set Values( x );
```

`col` is a reference to the data table column, and **x** is a column vector.

For example, the following script creates a new column called test and copies the values of vector **x** into the test column:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "test" );
col = Column( "test" );
x = 1::40;
```

```
col << Set Values( x );
```

Move All Matrix Values

The `Set Matrix()` function copies values from a matrix into an existing data table, making new rows and new columns as needed to store the values of the matrix. The new columns are named Col1, Col2, and so on.

```
dt = New Table( "B" );
dt << Set Matrix([1 2 3 4 5, 6 7 8 9 10]);
```

This script creates a new data table called B containing two rows and five columns.

To create a new data table from a matrix argument, use the `As Table(matrix)` command. The columns are named Col1, Col2, and so on. For example, the following script creates a new data table containing the values of **A**:

```
A = [1 2 3, 4 5 6, 7 8 9, 10 11 12];
dt = As Table( A );
```

Matrices and Reports

You can extract matrices of values from reports. First, you need to locate the items that you want to extract. This information is in the tree structure of the report.

Run the following script to create a table of parameter estimates in a Bivariate report:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" ) ;
biv = dt << Bivariate( X( height ), Y( weight ), Fit Line );
```

Now, open the tree structure to identify which item contains the parameter estimates:

- Right-click a gray disclosure icon and select **Edit > Show Tree Structure**.

  The parameter estimates are contained in `NumberColBox(13)`. Continue with the script:

  ```
  colBox = Report( biv )[Number Col Box( 13 )];
  beta = colBox << Get As Matrix;
    [-127.145248610915, 3.711354893859555]
  ```

Notes:

- When a variable contains a reference to a table box, `Get As Matrix()` creates a matrix **A** that contains the values from all numeric columns in the table:

  ```
  A = tableBox << Get As Matrix;
  ```

- When a variable contains a reference to a numeric column in a report table, `Get As Matrix()` creates a matrix **A** as a column vector of values from that column.

  ```
  A = colBox << Get As Matrix;
  ```

Loc Functions

The Loc(), Loc Nonmissing(), Loc Min(), Loc Max(), and Loc Sorted() functions all return matrices that show positions of certain values in a matrix.

Loc

The Loc() function creates a matrix of positions that locate where **A** is true (nonzero and nonmissing).

```
A = [0 1 . 3 4 0];
B = [2 0 0 2 5 6];
```

The following example returns the indices for the values of **A** that are nonmissing and nonzero.

```
I = Loc( A );
   [2, 4, 5]
```

The following example returns the indices for the values of **A** that are less than the corresponding values of **B**. Note that the two matrices must have the same number of rows and columns.

```
I = Loc( A < B );
   [1, 5, 6]
```

The following script replaces all values less than 4 in **A** with 0.

```
A = [0 1 0 3 4 0];
A[Loc( A < 4 )] = 0;
Show( A );
   A = [0 0 0 0 4 0];
```

The following script returns an empty matrix:

```
Loc( [2 3 4 5 6] > 7 );
   [](0, 1)
```

Loc Nonmissing

The Loc Nonmissing() function returns a vector of row numbers in a matrix that do not contain any missing values. For example,

```
A = [1 2 3, 4 . 6, 7 8 ., 8 7 6];
Loc Nonmissing( A );
   [1, 4]
```

Loc Min and Loc Max

The `Loc Min()` and `Loc Max()` functions return the position of the first occurrence of the minimum and maximum elements of a matrix. Elements of a matrix are numbered consecutively, starting in the first row, first column, and proceeding left to right.

```
A = [1 2 2, 2 4 4, 1 1 1];
B = [6, 12, 9];
Show( Loc Max( A ) );
Show( Loc Min( B ) );
   Loc Max(A) = 5;
   Loc Min(B) = 1;
```

Loc Sorted

The `Loc Sorted()` function is mainly used to identify the interval that a number lies within. The function returns the position of the highest value in **A** that is less than or equal to the value in **B**. The resulting vector contains an item for each element in **B**.

```
A = [10 20 30 40];
B = [35];
Loc Sorted( A, B );
   [3]
A = [10 20 40];
B = [35 5 45 20];
Loc Sorted( A, B );
   [2, 1, 3, 2]
```

Here's an example that defines a list of strings. `Loc Sorted()` identifies the string at the seventh interval and puts it in a list.

```
{"No Fat", "Some Fat", "Low Fat", "Normal Fat", "Medium Fat", "High Fat"}
[Loc Sorted( [0, 1, 5, 10, 20, 25], 7)]
   {"Low Fat"}
```

To convert the returned value to a string, include the subscript [1] at the end of the expression:

```
{"No Fat", "Some Fat", "Low Fat", "Normal Fat", "Medium Fat", "High Fat"}
[Loc Sorted( [0, 1, 5, 10, 20, 25], 7)][1]
   "Low Fat"
```

Notes:

- **A** must be sorted in ascending order.
- The returned values are always 1 or greater. If the element in **B** is smaller than all of the elements in **A**, then the function returns a value of 1. If the element in **B** is greater than all of the elements in **A**, then the function returns *n*, where *n* is the number of elements in **A**.

Ranking and Sorting

The `Rank()` function returns the positions of the numbers in a vector or list, as if the numbers were sorted from lowest to highest.

```
E = [1 -2 3 -4 0 5 1 8 -7];
R = Rank( E );
   [9, 4, 2, 5, 7, 1, 3, 6, 8]
```

If **E** were sorted from lowest to highest, the first number would be -7. The position of -7 in **E** is 9.

The original matrix **E** can then be sorted using the matrix **R** as subscripts to **E**.

```
sortedE = E[R];
   [-7,- 4, -2, 0, 1, 1, 3, 5, 8]
```

The `Ranking Tie()` function returns ranks for the values in a vector or list, with ranks for ties averaged. Similarly, `Ranking()` returns ranks for the values in a vector or list, but the ties are ranked arbitrarily.

```
E = [1 -2 3 -4 0 5 1 8 -7];
Ranking Tie( E );
   [5.5, 3, 7, 2, 4, 8, 5.5, 9, 1]

E = [1 -2 3 -4 0 5 1 8 -7];
Ranking( E );
   [5, 3, 7, 2, 4, 8, 6, 9, 1]
```

The `Sort Ascending()` and `Sort Descending()` functions sort vectors.

```
E = [1 -2 3 -4 0 5 1 8 -7];
Sort Ascending( E );
   [-7 -4 -2 0 1 1 3 5 8]

E = [1 -2 3 -4 0 5 1 8 -7];
Sort Descending( E );
   [8 5 3 1 1 0 -2 -4 -7]
```

If the argument is not a vector or list, an error message is generated.

Special Matrices

Construct an Identity Matrix

The `Identity()` function constructs an identity matrix of the dimension that you specify. An identity matrix is a square matrix of zeros except for a diagonal of ones. The only argument specifies the dimension.

```
Identity( 3 );
```

```
[1 0 0,
0 1 0,
0 0 1]
```

Construct a Matrix with Specific Values

The `J()` function constructs a matrix with the number of rows and columns that you specify as the first two arguments, whose elements are all the third argument, for example:

```
J( 3, 4, 5 );
   [5 5 5 5,
   5 5 5 5,
   5 5 5 5]
J( 3, 4, Random Normal() ); // your results will differ
   [0.407709113182904 1.67359154091978 1.00412665221308 0.240885679837327,
   -0.557848036549455 -0.620833861982722 0.877166783247633 1.50413740148892,
   -2.09920574748608 -0.154797501010655 0.0463943433032137 0.0640418263933316]
```

Create a Diagonal Matrix

The `Diag()` function creates a diagonal matrix from a square matrix (having an equal number of rows and columns) or a vector. A diagonal matrix is a square matrix whose nondiagonal elements are zero.

```
D = [1 -1 1];
Diag( D );
   [1 0 0,
   0 -1 0,
   0 0 1]
Diag([1, 2, 3, 4]);
   [1 0 0 0,
   0 2 0 0,
   0 0 3 0,
   0 0 0 4]

A = [1 2,3 4];
f = [5];
D = Diag( A, f );
   [1 2 0
   3 4 0
   0 0 5]
```

In the third example, at first glance, not all of the nondiagonal elements are zero. Using matrix notation, the matrix can be expressed as follows:

```
[A 0,
0` f]
```

Where `A` and `f` are the matrices from the example, and **`0`** is a column vector of zeros.

Create a Column Vector from Diagonal Elements

The `VecDiag()` function creates a column vector from the diagonal elements of a matrix.

```
v = Vec Diag(
[1 0 0 1, 5 3 7 1, 9 9 8 8, 1 5 4 3]);
   [1, 3, 8, 3]
```

Calculate Diagonal Quadratic Forms

The `Vec Quadratic()` function calculates the hats in regression that go into the standard errors of prediction or the Mahalanobis or T2 statistics for outlier distances. `Vec Quadratic(Sym, X)` is equivalent to calculating `Vec Diag(X*Sym*X`)`.The first argument is a symmetric matrix, usually an inverse covariance matrix. The second argument is a rectangular matrix with the same number of columns as the symmetric matrix argument.

Return the Sum of Diagonal Elements

The `Trace()` function returns the sum of the diagonal elements for a square matrix.

```
D = [0 1 2, 2 1 0, 1 2 0];
Trace( D ); // returns 1
```

Generate a Row Vector of Integers

The `Index()` function generates a row vector of integers from the first argument to the last argument. A double colon `::` is the equivalent infix operator.

```
6::10;
   [6 7 8 9 10]
Index( 1, 5 );
   [1 2 3 4 5]
```

The optional *`increment`* argument changes the default increment of +1.

```
Index( 0.1, 0.4, 0.1 );
   [0.1, 0.2, 0.3, 0.4]
```

The increment can also be negative.

```
Index( 6, 0, -2 );
   [6, 4, 2, 0]
```

The default value of the increment is 1, or -1 if the first argument is higher than the second.

Reshape a Matrix

The `Shape()` function reshapes an existing matrix across rows to be the specified dimensions. The following example changes the 3x4 matrix `a` into a 12x1 matrix:

```
a = [1 1 1, 2 2 2, 3 3 3, 4 4 4];
Shape( a, 12, 1 );
   [1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4]
```

Summarize Columns of a Matrix

There are several functions that return a row vector based on summary values for each column in a matrix.

```
mymatrix = [11 22, 33 44, 55 66];
V Max( mymatrix ); // returns the maximum of each column
   [55 66]
V Min( mymatrix ); // returns the minimum of each column
   [11 22]
V Mean( mymatrix ); // returns the mean of each column
   [33 44]
V Median( mymatrix ); // returns the median of each column
   [33 44]
V Quantile( mymatrix, 0.75 ); // returns the 0.75 quantile of each column
   [55 66]
V Sum( mymatrix ); // returns the sum of each column
   [99 132]
V Std( mymatrix ); // returns the standard deviations of each column
   [22 22]
```

Create Design Matrices

The `Design()` function creates a matrix of design columns for a vector or list. There is one column for each unique value in the vector or list. The design columns have the elements 0 and 1. For example, **x** below has values 1, 2, and 3, then the design matrix has a column for 1s, a column for 2s, and a column for 3s. Each row of the matrix has a 1 in the column representing that row's value. So, the first row (1) has a 1 in the 1s column (the first column) and 0s elsewhere; the second row (2) has a 1 in the 2's column and 0s elsewhere; and so on.

```
x = [1, 2, 3, 2, 1];
Design( x );
   [1 0 0,
   0 1 0,
   0 0 1,
   0 1 0,
   1 0 0]
```

A variation is the `Design Nom()` or `Design F()` function, which removes the last column and subtracts it from the others. Therefore, the elements of `Design Nom()` or `Design F()` matrices are 0, 1, and –1. And the `Design Nom()` or `Design F()` matrix has one less column than the vector or list has unique values. This function makes full-rank versions of design matrices for effects.

```
x = [1, 2, 3, 2, 1];
Design Nom( x );
   [1 0, 0 1, -1 -1, 0 1, 1 0]
```

`Design Nom()` is further demonstrated in the "ANOVA Example".

To facilitate ordinal factor coding, use the `Design Ord()` function. This function produces a full-rank coding with one less column than the number of unique values in the vector or list. The row for the lowest value in the vector or list is all zeros. Each succeeding value adds an additional 1 to the row of the design matrix.

```
x = [1, 2, 3, 4, 5, 6];
Design Ord( x );
   [0 0 0 0 0,
   1 0 0 0 0,
   1 1 0 0 0,
   1 1 1 0 0,
   1 1 1 1 0,
   1 1 1 1 1]
```

`Design()`, `Design Nom()`, and `Design Ord()` support a second argument that specifies the levels to be looked up and their order. This feature allows design matrices to be created one row at a time.

- `Design( values, levels )` creates a design matrix of indicator columns.
- `DesignNom( values, levels )` creates a full-rank design matrix of indicator columns.

Notes:

- The *values* argument can be a single element or a matrix or list of elements.
- The *levels* argument can be a list or matrix of levels to be looked up.
- The result has the same number of rows as there are elements in the *values* argument.
- The result always has the same number of columns as there are items in the *levels* argument. In the case of `Design Nom()` and `Design Ord()`, there is one less column than the number of items in the *levels* argument.
- If a value is not found, the whole row is zero.

Examples

```
Design( 20, [10 20 30] );
   [0 1 0]
Design( 30, [10 20 30] );
```

```
   [0 0 1]
Design Nom( 20, [10 20 30] );
   [0 1]
Design Nom( 30, [10 20 30] );
   [-1 -1]
Design Ord( 20, [10 20 30] );
   [1 0]
Design( [20, 10, 30, 20], [10 20 30] );
   [0 1 0,
   1 0 0,
   0 0 1,
   0 1 0]
Design({"b", "a", "c", "b"}, {"a", "b", "c"});
   [0 1 0,
   1 0 0,
   0 0 1,
   0 1 0]
```

Find the Direct Product

The `Direct Product()` function finds the direct product (or Kronecker product) of two square matrices. The direct product of a *m* x *m* matrix and a *n* x *n* matrix is the *mn* x *mn* matrix whose elements are the products of numbers, one from **A** and one from **B**.

```
G = [1 2, 3 5];
H = [2 3, 5 7];
Direct Product( G, H );
   [2 3 4 6,
   5 7 10 14,
   6 9 10 15,
   15 21 25 35]
```

The `H Direct Product()` function finds the row-by-row direct product of two matrices with the same number of rows.

```
H Direct Product( G, H );
   [2 3 4 6, 15 21 25 35]
```

`HDirect Product()` is useful for constructing the design matrix columns for interactions.

```
XA = Design Nom( A );
XB = Design Nom( B );
XAB = HDirect Product( XA, XB );
X = J( NRow( A ), 1 ) || XA || XB || XAB;
```

Inverse Matrices and Linear Systems

JMP has the following functions for computing inverse matrices: `Inverse()`, `GInverse()`, and `Sweep()`. The `Solve()` function is used for solving linear systems of equations.

Inverse or Inv

The `Inverse()` function returns the matrix inverse for the square, nonsingular matrix argument. `Inverse()` can be abbreviated `Inv`. For a matrix **A**, the matrix product of **A** and `Inverse`(**A**) (often denoted $\mathbf{A}(\mathbf{A}^{-1})$) returns the identity matrix.

```
A = [5 6, 7 8];
AInv = Inv( A );
A*A Inv;
   [1 0,0 1]

A = [1 2, 3 4];
AInv = Inverse( A );
A*AInv;
   [1 1.110223025e-16,0 1]
```

Note: There can be small discrepancies in the results due to floating point limitations, as illustrated in the second example.

GInverse

The (Moore-Penrose) generalized inverse of a matrix **A** is any matrix **G** that satisfies the following conditions:

AGA = **A**

GAG = **G**

(**AG**)′ = **AG**

(**GA**)′ = **GA**

The `GInverse()` function accepts any matrix, including non-square ones, and uses singular-value decomposition to calculate the Moore-Penrose generalized inverse. The generalized inverse can be useful when inverting a matrix that is not full rank. Consider the following system of equations:

$$\begin{aligned} x + 2y + 2z &= 6 \\ 2x + 4y + 4z &= 12 \\ x + y + z &= 1 \end{aligned}$$

Find the solution to this system using the following script:

```
A = [1 2 2, 2 4 4, 1 1 1];
B = [6, 12, 1];
Show( GInverse( A ) * B );
   GInverse(A) * B = [-4, 2.5, 2.5];
```

Solve

The `Solve()` function solves a system of linear equations. `Solve()` finds the vector **x** so that $x = A^{-1}b$ where *A* equals a square, nonsingular matrix and *b* is a vector. The matrix **A** and the vector **b** must have the same number of rows. `Solve(A,b)` is the same as `Inverse(A)*b`.

```
A = [1 -4 2, 3 3 2, 0 4 -1];
b = [1, 2, 1];
x = Solve( A, b );
   [-16.9999999999999, 4.99999999999998, 18.9999999999999]
A*x;
   [1, 2, 0.999999999999997]
```

Note: There can be small discrepancies in the results due to floating point limitations, as illustrated in the example.

Sweep

The `Sweep()` function inverts parts (or *pivots*) of a square matrix. If you sequence through all of the pivots, you are left with the matrix inverse. Normally the matrix must be positive definite (or negative definite) so that the diagonal pivots never go to zero. `Sweep()` does not check whether the matrix is positive definite. If the matrix is *not* positive definite, then it still works, as long as no zero pivot diagonals are encountered. If zero (or near-zero) diagonal pivots are encountered on a full sweep, then the result is a g2 generalized inverse if the zero pivot row and column are zeroed.

About the Sweep Function

Suppose matrix **E** consists of smaller matrix partitions, **A**, **B**, **C**, and **D**:

$$E = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

The syntax for `Sweep()` appears as follows:

```
Sweep( E, [...] );
```

[...] indicates partition A.

This produces the matrix result equivalent to the following:

$$\begin{bmatrix} A^{-1} & A^{-1}B \\ -CA^{-1} & D - CA^{-1}B \end{bmatrix}$$

Notes:

- The submatrix in the **A** position becomes the inverse.
- The submatrix in the **B** position becomes the solution to **A**x = **B**.
- The submatrix in the **C** position becomes the solution to x**A** = **C**.

Use of the Sweep Function

`Sweep()` is sequential and reversible:

- `A = Sweep( A, {i,j} )` is the same as `A = Sweep( Sweep( A, i ), j )`. It is sequential.
- `A = Sweep( Sweep( A, i ), i )` restores **A** to its original values. It is reversible.

If you have a cross-product matrix partitioned as follows:

$$C = \begin{bmatrix} X'X & X'y \\ y'X & y'y \end{bmatrix}$$

Then after sweeping through the indices of **X′X**, the result appears as follows:

$$\begin{bmatrix} (X'X)^{-1} & (X'X)^{-1}X'y \\ -yX(X'X)^{-1} & y'y - y'X(X'X)^{-1}X'y \end{bmatrix}$$

The partitions are recognizable to statisticians as follows:

- the least squares estimates for the model Y = Xb + e in the upper right
- the sum of squared errors in the lower right
- a matrix proportional to the covariance of the estimates in the upper left

The `Sweep` function is useful in computing the partial solutions needed for stepwise regression.

The *`index`* argument is a vector that lists the rows (or equivalently the columns) on which you want to sweep the matrix. For example, if **E** is a 4x4 matrix, to sweep on all four rows to get $\mathbf{E}^{-1}$ requires these commands:

```
E = [ 5 4 1 1, 4 5 1 1, 1 1 4 2, 1 1 2 4];
Sweep( E, [1, 2, 3, 4] );
   [0.56 -0.44 -0.02 -0.02,
   -0.44 0.56 -0.02 -0.02,
   -0.02 -0.02 0.34 -0.16,
   -0.02 -0.02 -0.16 0.34]
Inverse( E ); // notice that these results are the same
   [0.56 -0.44 -0.02 -0.02,
```

```
-0.44 0.56 -0.02 -0.02,
-0.02 -0.02 0.34 -0.16,
-0.02 -0.02 -0.16 0.34]
```

Note: For a tutorial on `Sweep()`, and its relation to the Gauss-Jordan method of matrix inversion, see Goodnight, J.H. (1979) "A tutorial on the SWEEP operator." *The American Statistician*, August 1979, Vol. 33, No. 3. pp. 149–58.

`Sweep()` is further demonstrated in the "ANOVA Example".

Determinant

The `Det()` function returns the determinant of a square matrix. The determinant of a 2 × 2 matrix is the difference of the diagonal products, as demonstrated below. Determinants for $n \times n$ matrices are defined recursively as a weighted sum of determinants for $(n - 1) \times (n - 1)$ matrices. For a matrix to have an inverse, the determinant must be nonzero.

$$\begin{vmatrix} 1 & 2 \\ 3 & 5 \end{vmatrix} = (1 \cdot 5) - (3 \cdot 2) = -1$$

```
F = [1 2, 3 5];
Det( F );
   -1
```

Decompositions and Normalizations

This section contains functions that calculate eigenvalues and eigenvectors and functions that decompose matrices.

Eigenvalues

The `Eigen()` function performs eigenvalue decomposition of a symmetric matrix. Eigenvalue decompositions are used in many statistical techniques, most notably in principal components and canonical correlation, where the transformation associated with the largest eigenvalues are transformations that maximize variances.

`Eigen()` returns a list of matrices. The first matrix in the returned list is a column vector of eigenvalues; the second matrix contains eigenvectors as the columns.

```
A = [ 5 4 1 1, 4 5 1 1, 1 1 4 2, 1 1 2 4];
Eigen( A );
   {[10, 5, 2, 1][0.632455532033676 - 0.316227766016838 - 2.77555756156289e-16
      -0.707106781186547, 0.632455532033676 - 0.316227766016837 -
      1.66533453693773e-16 0.707106781186547, 0.316227766016838
      0.632455532033676 0.707106781186548 0, 0.316227766016837
      0.632455532033676 - 0.707106781186547 0]}
```

Since the function returns a list of matrices, you might want to assign it to a list of two global variables. That way, the column vector of eigenvalues is assigned to one variable, and the matrix of eigenvectors is assigned to another variable:

```
{evals, evecs} = Eigen( A );
```

For some $n \times n$ matrix **A**, eigenvalue decomposition finds all λ (lambda) and vectors **x**, so that the equation Ax = λx has a nonzero solution **x**. The λ's are called eigenvalues, and the corresponding **x** vectors are called eigenvectors. This is equivalent to solving (A - λI)**x** = 0. You can reconstruct **A** from eigenvalues and eigenvectors by a statement like the following:

```
newA = evecs * Diag( evals ) * evecs`;
  [5.000000000000001 4 1 1,
  4 5 1 1,
  1 1 4 2,
  1 1 2 4]
```

Note the following about eigenvalues and eigenvectors:

- The eigenvector matrices are orthonormal, such that the inverse is the transpose: $E'E = EE' = I$.
- Eigenvectors are uniquely determined only if the eigenvalues are unique.
- Zero eigenvalues correspond to singular matrices.
- Inverses can be obtained by inverting the eigenvalues and reconstituting with the eigenvectors. Moore-Penrose generalized inverses can be formed by inverting the nonzero eigenvalues and reconstituting. (See "GInverse".)

  Note: You must decide whether a very small eigenvalue is effectively zero.

- The eigenvalue decomposition enables you to view any square-matrix multiplication as the following:
  - a rotation (multiplication by an orthonormal matrix)
  - a scaling (multiplication by a diagonal matrix)
  - a reverse rotation (multiplication by the orthonormal inverse, which is the transpose), or in notation:

  ```
  A * x = E` * Diag( M ) * E * x;
  ```

  `E` rotates, `Diag( M )` scales, and `E`` reverse-rotates.

Cholesky Decomposition

The `Cholesky()` function performs Cholesky decomposition. A positive semi-definite matrix **A** is re-expressed as the product of a nonsingular, lower-triangular matrix **L** and its transpose: **L***L' = **A**.

```
E = [ 5 4 1 1, 4 5 1 1, 1 1 4 2, 1 1 2 4];
```

```
L = Cholesky( E );
   [2.23606797749979 0 0 0,
   1.788854381999832 1.341640786499874 0 0,
   0.447213595499958 0.149071198499986 1.9436506316151 0,
   0.447213595499958 0.149071198499986 0.914659120760047 1.71498585142509]
```

To verify the results, enter the following:

```
L*L`;
   [5 4 1 1,
   4 5 1 1,
   1 1 4 2,
   1 1 2 4]
```

About Cholesky Decomposition

`Cholesky()` is useful for reconstituting expressions into a more manageable form. For example, eigenvalues are available only for symmetric matrices in JMP, so the eigenvalues of the product **AB** could not be obtained directly. (Although **A** and **B** can be symmetric, the product is not symmetric.) However, you can usually rephrase the problem in terms of eigenvalues of **L′BL** where **L** is the Cholesky root of **A**, which has the same eigenvalues.

Another use of `Cholesky()` is in reordering matrices within `Trace()` expressions. Expressions such as `Trace(A*B*A`)` might involve huge operations counts if **A** has many rows. However, if **B** is small and can be factored into LL′ by Cholesky, then it can be reformulated to `Trace(A*L*L`*A`)`. The resulting matrix is equal to `Trace(L`A`*AL)`. This expression involves a much smaller number of operations, because it consists of only the sum of squares of the **AL** elements.

Use the function `Chol Update()` to update a Cholesky decomposition. If **L** is the Cholesky root of a $n \times n$ matrix **A**, then after using `cholUpdate(L, C, V)`, **L** will be replaced with the Cholesky root of `A+V*C*V'`. **C** is an $m \times m$ symmetric matrix and **V** is an $n \times m$ matrix.

Examples

Manually update the Cholesky decomposition, as follows:

```
exS = [16 1 0 11 -1 12, 1 11 -1 1 -1 1, 0 -1 12 -1 1 0, 11 1 -1 11 -1 9, -1 -1
  1 -1 9 -1, 12 1 0 9 -1 12];
exAchol = Cholesky( exS ); // conducts the Cholesky decomposition

// add two column vectors to the design matrix
exV = [1 1, 0 0, 0 1, 0 0, 0 0, 0 1];

/* The first column vector is added to one of the rows in the design matrix.
The second column vector is subtracted from one of the rows in the design
  matrix. */
exC = [1 0, 0 -1];
exAnew = exS + exV * exC * exV`;
```

```
// update the Cholesky decomposition manually
exAcholnew = Cholesky( exAnew );
```

Instead of manual updating, use `Chol Update()` to update the Cholesky decomposition, as follows:

```
// update the Cholesky decomposition more efficiently
exAcholnew_test = Chol Update( exAchol, exV, exC );
Show( exAcholnew_test );
   exAcholnew_test =
   [ 4 0 0 0 0 0,
   0.25 3.30718913883074 0 0 0 0, ...]

// results are the same as the manual process
Show( exAcholnew );
   exAcholnew=
   [ 4 0 0 0 0 0,
   0.25 3.30718913883074 0 0 0 0, ...]
```

Singular Value Decomposition

The `SVD()` function finds the singular value decomposition of a matrix. That is, for a matrix **A**, `SVD()` returns a list of three matrices **U**, **M**, and **V**, so that `U*diag(M)*V`=A`.

Notes:

- When **A** is taller than it is wide, **M** is more compact, without extra zero diagonals.
- Singular value decomposition re-expresses **A** in the form USV′, where:
  - **U** and **V** are matrices that contain orthogonal column vectors (perpendicular, statistically independent vectors)
  - **S** is a $n \times n$ diagonal matrix containing the nonnegative square roots of the eigenvalues of **A′A**, the singular values of **A**.
- Singular value decomposition is the basis of correspondence analysis.

Example

```
A = [1 2 1 0, 2 3 0 1, 1 0 1 5, 0 1 5 1];
{U, M, V} = SVD( A );
newA = U * Diag( M ) * V`;
   [1 2 0.999999999999997 -2.99456640040496e-15,
   2 3 -1.17505831453979e-15 1,
   0.999999999999997 -2.16851276518826e-15 0.999999999999999 5,
   2.22586706011274e-15 1 5 0.999999999999997]
```

Orthonormalization

The `Ortho()` function orthogonalizes the columns and then divides the vectors by their magnitudes to normalize them. This function uses the Gram-Schmidt method. The column vectors of orthogonal matrices are unit-length and are mutually perpendicular (their dot products are zero).

```
B = Ortho( [1 -1, 1 0, 0 1] );
   [0.408248290463863 -0.707106781186548, 0.408248290463863 0.707106781186548,
      -0.816496580927726 3.14018491736755e-16]
```

To verify that these vectors are orthogonal, multiply **B** by its transpose, which should yield the identity matrix.

```
C = B` * B;
   [1 -3.119061760824e-16, -3.119061760824e-16 1]
```

By default, vectors are normalized, meaning that they are divided by their magnitudes, which scales them to have length 1. Include the option `Scaled(0)` to turn scaling off:

```
Ortho( [1 -1, 1 0, 0 1], Scaled( 0 ) );
   [0.408248290463863 -0.353553390593274, 0.408248290463863 0.353553390593274,
      -0.816496580927726 1.57009245868377e-16]
```

To create vectors whose elements sum to zero, include the `Centered(1)` option. This option is useful when constructing a matrix of contrasts.

```
result = Ortho( [1 -1, 1 0, 0 1], Centered( 1 ) );
   [0.408248290463863 -0.707106781186548, 0.408248290463863 0.707106781186548,
      -0.816496580927726 3.14018491736755e-16]
```

To verify that the elements of each column sum to zero, premultiply by a vector of ones to sum the columns.

```
J(1, 3) * result;
   [1.11022302462516e-16 2.02996189274239e-16]
```

Orthogonal Polynomials

The `Ortho Poly()` function returns orthogonal polynomials for a vector of indices. The polynomial order is specified as a function argument. Orthogonal polynomials can be useful for fitting polynomial models where some regression coefficients might be highly correlated.

```
Ortho Poly( [1 2 3], 2 );
   [-0.707106781186548 0.408248290463862, 0 -0.816496580927726,
      0.707106781186548 0.408248290463864]
```

The polynomial order must be less than the dimension of the vector. Use the `Scaled(1)` option to produce vectors of unit length, as described in "Orthonormalization".

QR Decomposition

`QR()` factorization is useful for numerically stable matrix work. `QR()` returns a list of two matrices. Here is an example:

```
{Q, R} = QR( [11 22, 33 44] );
```

`Q` and `R` hold the results. For a *m* × *n* matrix, `QR()` creates an orthogonal *m* × *m* matrix `Q` and an upper triangular *m* × *n* matrix `R`, so that `X = Q*R`.

Update Inverse Matrices

To add or drop one or more rows in an inverse of an **M`M** matrix, use the `Inv Update(S, X, w)` function. Updating inverse matrices is helpful in drop-1 influence diagnostics and also in candidate design evaluation.

Notes:

- The first argument, **S**, is the matrix to be updated.
- The second argument, **X**, is the matrix of rows to be added or dropped.
- The third argument, `w`, is either `1` to add or `-1` to delete the row or rows.
- Multiple rows can be added or deleted.

Using the `Inv Update(S, X, w)` function is equivalent to calculating the following:

```
S - w * S * X` * Inv( I + w * X * S * X` ) * X * S
```

Where **I** is an identity matrix and `Inv(A)` is an inverse matrix of **A**.

Build Your Own Matrix Functions

You can store your own operations in macros. See "Maximize and Minimize". Similarly, you can create custom matrix functions. For example, you can make your own matrix operation called `Mag()` to find the magnitude of a vector:

```
mag = Function( {x},
   Sqrt( x` * x )
);
```

Similarly, you could create a function called `Normalize` to divide a vector by its magnitude:

```
normalize = Function( {x},
   x / Sqrt( x` * x )
);
```

Statistical Examples

This section contains statistical examples of using matrices.

Regression Example

Suppose that you want to implement your own regression calculation, rather than use the facilities built into JMP. Because of the compact matrix notation, it might require only a few lines of code:

```
Y = [98, 112.5, 84, 102.5, 102.5, 50.5, 90, 77, 112, 150, 128, 133, 85, 112];
X = [65.3, 69, 56.5, 62.8, 63.5, 51.3, 64.3, 56.3, 66.5, 72, 64.8, 67, 57.5,
  66.5];
X = J( N Row( X ), 1 ) || X; // put in an intercept column of 1s
beta = Inv( X` * X ) * X` * Y; // the least square estimates
resid = Y - X * beta; // the residuals, Y - predicted
sse = resid` * resid; // sum of squared errors
Show( beta, sse );
```

This could be expanded into a script that gets its data from a data table, calculates additional results, and shows the results in a report window:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );

// get data into matrices
x = (Column( "Age" ) << Get Values ) || (Column( "Height" ) << Get Values);
x = J(N Row( x ), 1, 1) || x;
y = Column( "Weight" ) << Get Values;

// regression calculations
xpxi = Inv( x` * x );
beta = xpxi * x` * y; // parameter estimates
resid = y - x * beta; // residuals
sse = resid` * resid; // sum of squared errors
dfe = N Row( x ) - N Col( x );  // degrees of freedom
mse = sse / dfe; // mean square error, error variance estimate

// additional calculations on estimates
stdb = Sqrt( Vec Diag( xpxi ) * mse ); // standard errors of estimates
alpha = .05;
qt = Students t Quantile( 1 - alpha / 2, dfe );
betau95 = beta + qt * stdb; // upper 95% confidence limits
betal95 = beta - qt * stdb; // lower 95% confidence limits
tratio = beta :/ stdb; // Student's T ratios
probt = ( 1 - TDistribution( Abs( tratio ), dfe ) ) * 2; // p-values
```

```
// present results
New Window( "Big Class Regression",
   Table Box(
      String Col Box( "Term",{"Intercept", "Age", "Height"} ),
      Number Col Box( "Estimate", beta ),
      Number Col Box( "Std Error", stdb ),
      Number Col Box( "TRatio", tratio ),
      Number Col Box( "Prob>|t|", probt ),
      Number Col Box( "Lower95%", betal95 ),
      Number Col Box( "Upper95%", betau95 ) ) );
```

ANOVA Example

You can implement your own one-way ANOVA. This example presents a problem involving a three-level factor indicating Low, Medium, and High doses and a response measurement. Therefore, this example solves the general linear model, as follows:

$$Y = a + bX + e$$

Where:

- **Y** is a vector of responses
- *a* is the intercept term
- **b** is a vector of coefficients
- **X** is a design matrix for the factor
- *e* is an error term

```
factor = [1, 2, 3, 1, 2, 3, 1, 2, 3];
y = [1, 2, 3, 4, 3, 2, 5, 4, 3];
```

First, build a design matrix for the factor:

```
Design Nom( factor );
   [1 0, 0 1, -1 -1, 1 0, 0 1, -1 -1, 1 0, 0 1, -1 -1]
```

Next, add a column of 1s to the design matrix for the intercept term. You can do this by concatenating `J` and `Design Nom()`:

```
x = J( 9, 1, 1) || Design Nom( factor );
   [1 1 0,
   1 0 1,
   1 -1 -1,
   1 1 0,
   1 0 1,
   1 -1 -1,
   1 1 0,
   1 0 1,
   1 -1 -1]
```

Now, to solve the normal equation, you need to construct a matrix **M** with partitions:

$$\begin{bmatrix} X'X & X'y \\ y'X & y'y \end{bmatrix}$$

You can construct matrix **M** in one step by concatenating the pieces:

```
M = ( x` * x || x` * y )
        |/
  ( y` * x || y` * y );
   [9 0 0 27,
      0 6 3 2,
      0 3 6 1,
      27 2 1 93]
```

Now, sweep **M** over all the columns in **X′X** for the full fit model, and over the first column only for the intercept-only model:

```
FullFit = Sweep( M, [1, 2, 3] ); // full fit model
InterceptOnly = Sweep( M, [1] ); // model with intercept only
```

Recall that some of the standard ANOVA results are calculated by comparing the results of these two models. This example focuses on the full fit model, which produces this swept matrix:

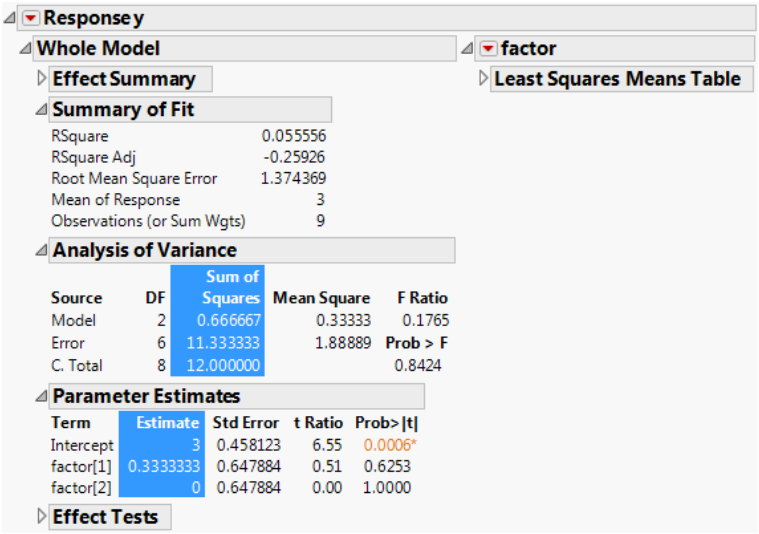
```
[0.111111111111111 0 0 3,
0 0.222222222222222 -0.11111111111111 0.333333333333333,
0 -0.11111111111111 0.222222222222222 0,
-3 -0.33333333333333 0 11.33333333333333]
```

Examine the model coefficients from the upper right partition of the matrix. The lower left partition is the same, except that the signs are reversed: 3, 0.333, 0. The results can be interpreted as follows:

- The coefficient for the intercept term is 3.
- The coefficient for the first level of the factor is 0.333.
- The coefficient for the second level is 0.
- Because of the use of `Design Nom()`, the coefficient for the third level is the difference, –0.333.
- The lower right partition of the matrix holds the sum of squares, 11.333.

You could modify this into a generalized ANOVA script by replacing some of the explicit values in the script with arguments. These results match those from the Fit Model platform.

Figure 7.1 ANOVA Report Within Fit Model



Here is how to construct the report in Figure 7.1:

1. Build a data table (described in "Data Tables"):

```
dt = New Table( "Data" );
dt << New Column( "y", Set Values( [1, 2, 3, 4, 3, 2, 5, 4, 3] ) );
dt << New Column( "factor", "Nominal", Values( [1, 2, 3, 1, 2, 3, 1, 2, 3] )
  );
```

2. Run a model (described in "Launch the Analysis and Save the Script to a Window"):

```
obj = Fit Model(
   Y( :y ),
   Effects( :factor ),
   Personality( "Standard Least Squares" ),
   Run(
      y << {Plot Actual by Predicted( 0 ), Plot Residual by Predicted( 0 ),
      Plot Effect Leverage( 0 )}
   )
);
```

3. Use JSL techniques for navigating displays (described in "Display Box Object References"):

```
ranova = obj << Report;
ranova[Outline Box( 6 )] << Close( 0 );
ranova[Outline Box( 7 )] << Close( 1 );
ranova[Outline Box( 9 )] << Close( 1 );
ranova[Outline Box( 5 ), Number Col Box( 2 )] << Select;
ranova[Outline Box( 6 ), Number Col Box( 1 )] << Select;
```

The result is shown in Figure 7.1.

Associative Arrays

An associative array maps unique keys to values that can be non-unique. An associative array is also called a dictionary, a map, a hash map, or a hash table.

Note the following about keys:

- Keys are placed in quotes.
- The value associated with a key can be a number, date, matrix, list, and so on.
- A key cannot be a floating point number. Only integers are supported.
- You can use matrices as both keys and values, or lists as both keys and values, but you cannot mix the two. In other words, you cannot use a matrix as a key and a list as a value, or the other way around.
- Though associative arrays are not usually ordered, in JMP, keys are returned in alphanumeric order for the purpose of iteration and serialization.

For very large lists, using an associative array is more efficient and faster.

Create Associative Arrays

To create an empty associative array, use the `Associative Array()` function or `[=>]`.

```
cary = Associative Array();
cary = [=> ];
   Associative array( 0 elements ) assigned.
```

Keys and values can be any JSL objects. Items can be added and changed with subscripting:

```
cary = Associative Array();
cary["state"] = "NC";
cary["population"] = 116234;
cary["weather"] = "cloudy";
cary["population"] += 10;
cary["weather"] = "sunny";
cary["high schools"] = {"Cary", "Green Hope", "Panther Creek"};
```

Default Values

A default value determines the value of a key that does not exist in an associative array. If you try to access a key that does not exist in an associative array, an error results. If you define a default value for your associative array, accessing a key that does not exist results in the following:

- adds the key to the associative array
- assigns the default value to the new key
- returns the new key's (default) value instead of an error

If you construct an associative array from a list of strings without assigning values to the keys, then the keys are assigned values of 1. The default value for the associative array is set to 0.

To set the default value:

```
cary = Associative Array();
cary << Set Default Value( "Cary, NC" );
```

To determine whether there is a default value set for an associative array, use the `<<Get Default Value` message.

```
cary << Get Default Value;
   "Cary, NC"
```

If there is no default value, `Empty()` is returned.

Besides the `Set Default Value` message, a default value can be set in the literal constructor using `=>value` without a key.

```
counts = ["a" => 10,
"b" => 3,
=> 0]; // default value of 0
counts["c"] += 1;
Show( counts );
   counts = ["a" => 10, "b" => 3, "c" => 1, => 0];
```

In the first line, the default value is set to 0. In the second line, the key `"c"` does not exist in `counts`. In the output, the key `"c"` is created with the default value of 0 and then incremented by 1.

Note: If a key's value is the default value, then the key is dropped because any key will return the default value.

Associative Array Constructors

Create an empty associative array:

```
map = [=>];
map = Associative Array();
```

Create an empty associative array with a default value:

```
map = [=>0];
map = Associative Array( 0 );
```

Create an associative array with specific values:

```
map = ["yes" => 0, "no" => 1];
```

Create an associative array with specific values with a default value:

```
map = ["yes" => 0, "no" => 1, => 2];
```

Create an associative array with specific values that contains an associative array:

```
map = ["yes" => 0, "no" => 1, "alt" => ["a" => "1", "b" => "2"]];
```

Create an associative array from a list that contains two lists of a key-value pair:

```
map = Associative Array( {{"yes", 0}, {"no", 1}} );
```

Create an associative array from a list that contains two lists of a key-value pair with a default value:

```
map = Associative Array( {{"yes", 0}, {"no", 1}}, 2 );
```

Create an associative array from a list of keys and a list of values:

```
map = Associative Array( {"yes", "no"}, {0, 1} );
```

Create an associative array from a list of keys and a list of values with a default value:

```
map = Associative Array( {"yes", "no"}, {0, 1}, 2 );
```

Create an associative array from two column references. The first column contains the keys and the second contains the values.

```
map = Associative Array( :name, :height );
```

Create an associative array from two column references with a default value:

```
map = Associative Array(:name, :height, .);
```

Create an associative array from a single list of keys or a single column reference of keys with a default value of 0:

```
set = Associative Array( {"yes", "no"} );
set = Associative Array( :name );
```

Work with Associative Arrays

Find the Number of Keys

To determine the number of keys that an associative array contains, use the `N Items()` function.

```
cary = Associative Array();
cary["state"] = "NC";
cary["population"] = 116234;
cary["weather"] = "cloudy";
cary["population"] += 10;
cary["weather"] = "sunny";
cary["high schools"] = {"Cary", "Green Hope", "Panther Creek"};
N Items( cary );
   4
```

Add and Delete Keys and Values

To add or delete key-value pairs from an associative array, use the following functions:

- `Insert()`
- `Insert Into()`
- `Remove()`
- `Remove From()`

Notes:

- `Insert()` and `Remove()` return a named copy of the given associative array with the key-value pair added or removed.
- `Insert Into()` and `Remove From()` add or remove the key-value pairs directly from the given associative array.
- `Insert()` and `Insert Into()` take three arguments: the associative array, a key, and a value.
- `Remove()` and `Remove From()` take two arguments: the associative array and a key.
- If you insert a key with no value provided, the key is assigned a value of 1.

Examples

The following examples illustrate `Insert()` and `Insert Into()`:

```
cary = Associative Array();
cary["state"] = "NC";
cary["population"] = 116234;
cary["weather"] = "cloudy";
cary["population"] += 10;
```

```
cary["weather"] = "sunny";
cary["high schools"] = {"Cary", "Green Hope", "Panther Creek"};
newcary = Insert( cary, "time zone", "Eastern" );
Show( cary, newcary );
   cary = ["high schools" => {"Cary", "Green Hope", "Panther Creek"},
      "population" => 116244, "state" => "NC", "weather" => "sunny"];
   newcary = ["high schools" => {"Cary", "Green Hope", "Panther Creek"},
      "population" => 116244, "state" => "NC", "time zone" => "Eastern",
      "weather" => "sunny"];

cary = Associative Array();
cary["state"] = "NC";
cary["population"] = 116234;
cary["weather"] = "cloudy";
cary["population"] += 10;
cary["weather"] = "sunny";
cary["high schools"] = {"Cary", "Green Hope", "Panther Creek"};
Insert Into(cary, "county", "Wake");
Show( cary );
   cary = ["county" => "Wake", "high schools" => {"Cary", "Green Hope", "Panther
      Creek"}, "population" => 116244, "state" => "NC", "weather" => "sunny"];
```

Note that `cary <<Insert` is a message sent to an associative array that does the same thing as the function `Insert Into()`. For example, these two statements are equivalent:

```
cary << Insert( "county", "Wake" );
Insert Into( cary, "county", "Wake" );
```

The following examples illustrate `Remove()` and `Remove From()`:

```
cary = Associative Array();
cary["state"] = "NC";
cary["population"] = 116234;
cary["weather"] = "cloudy";
cary["population"] += 10;
cary["weather"] = "sunny";
cary["high schools"] = {"Cary", "Green Hope", "Panther Creek"};
newcary = Remove( cary, "high schools" );
Show( cary, newcary );
   cary = ["high schools" => {"Cary", "Green Hope", "Panther Creek"},
      "population" => 116244, "state" => "NC", "weather" => "sunny"];
   newcary = ["population" => 116244, "state" => "NC", "weather" => "sunny"];
Remove From( cary, "weather" );
Show( cary );
   cary = ["high schools" => {"Cary", "Green Hope", "Panther Creek"},
      "population" => 116244, "state" => "NC", "weather" => "sunny"];
   newcary = ["population" => 116244, "state" => "NC", "weather" => "sunny"];
   cary = ["high schools" => {"Cary", "Green Hope", "Panther Creek"},
      "population" => 116244, "state" => "NC"];
```

Note that `aa << Remove` is a message sent to an associative array that does the same thing as the function `Remove From()`. For example, these two statements are equivalent:

```
cary << Remove( "weather" );
Remove From( cary, "weather" );
```

Find Keys or Values in an Associative Array

To determine whether a certain key is contained within an associative array, use the `Contains()` function.

```
cary = Associative Array();
cary["state"] = "NC";
cary["population"] = 116234;
cary["weather"] = "cloudy";
cary["population"] += 10;
cary["weather"] = "sunny";
cary["high schools"] = {"Cary", "Green Hope", "Panther Creek"};
Contains(cary, "high schools");
  1
Contains(cary, "lakes");
  0
```

To obtain a list of all keys contained in an associative array, use the `<< Get Keys` message.

```
cary << Get Keys;
  {"high schools", "population", "state", "weather"}
```

To obtain a list of all values contained in an associative array, use the `<<Get Values` message.

```
cary << Get Values;
  {{"Cary", "Green Hope", "Panther Creek"}, 116244, "NC", "sunny"}
```

If you want to see only the values for certain keys, you can specify them as arguments. The keys must be given as a list.

```
cary << Get Values({"state", "population"});
  {"NC", 116244}
```

To see a value for a single key, use the `<<Get Value` message. Specify only one key and do not place it in a list.

```
cary << Get Value("weather");
  "sunny"
```

To obtain a list of all key-value pairs in an associative array, use the `<<Get Contents` message.

```
cary << Get Contents;
  {{"high schools", {"Cary", "Green Hope", "Panther Creek"}},
  {"population", 116244},
  {"state", "NC"},
  {"weather", "sunny"}}
```

Note: Using the `<<Get Contents` message, the returned list does not include the default value. Keys are listed alphabetically.

Iterate through an Associative Array

To iterate through as associative array, use the `<<First` and `<<Next` messages. `<<First` returns the first key in the associative array. `<<Next(key)` returns the key that follows the *key* that is given as the argument.

The following example removes all key-value pairs from the associative array `cary`, leaving an empty associative array:

```
currentkey = cary << First;
total = N Items( cary );
For( i = 1, i <= total, i++,
   nextkey = cary << Next( currentkey );
   Remove From( cary, currentkey );
   currentkey = nextkey;
);
Show( cary );
   cary = [=>];
```

Applications for Associative Arrays

You can use associative arrays to quickly and efficiently perform other tasks.

Get the Unique Values from a Data Table Column

Note: The strategy described here is not useful for columns with floating point numbers. Use Summarize instead. See "Store Summary Statistics in Global Variables".

A key can exist only once in an associative array, so putting a column's values into one automatically results in the unique values. For example, the Big Class.jmp sample data table contains 40 rows. To see how many unique values are in the column height, run this script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
unique heights = Associative Array( dt:height );
nitems( unique heights );
   17
```

There are only 17 unique values for height. You can use those unique values by getting the keys:

```
unique heights << Get Keys;
   {51, 52, 55, 56, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70}
```

Note: This is possible because you can use any JMP data type as keys in an associative array, not only strings.

Using an associative array to discover unique values in a column is efficient and fast. The following script takes some time to create a data table with 100,000 rows. Finding the 39 unique values takes very little time.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
nms = dt:name << Get Values;
dtbig = New Table( "Really Big Class",
   New Column( "name",
      Character,
      Set Values( nms[J( 100000, 1, Random Integer( N Items( nms ) ) )] )
   )
);
Wait( 0 );
t1 = Tick Seconds();
Write(
   "\!N# names from Really Big Class = ",
   N Items( Associative Array( dtbig:name ) ),
   ", elapsed time=",
   Tick Seconds() - t1
);
   # names from Really Big Class = 39, elapsed time=0.116666666639503
```

Sort a Column's Values in Lexicographic Order

Because keys are ordered lexicographically, putting the values into an associative array also sorts them. For example, the `<<Get Keys` message returns the keys (unique values of the names column) in ascending order:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
unique names = Associative Array( dt:name );
unique names << Get Keys;
   {"ALFRED", "ALICE", "AMY", "BARBARA", "CAROL", "CHRIS", "CLAY", "DANNY",
      "DAVID", "EDWARD", "ELIZABETH", "FREDERICK", "HENRY", "JACLYN", "JAMES",
      "JANE", "JEFFREY", "JOE", "JOHN", "JUDY", "KATIE", "KIRK", "LAWRENCE",
      "LESLIE", "LEWIS", "LILLIE", "LINDA", "LOUISE", "MARION", "MARK",
      "MARTHA", "MARY", "MICHAEL", "PATTY", "PHILLIP", "ROBERT", "SUSAN",
      "TIM", "WILLIAM"}
```

Compare Columns in Two Different Data Tables

Using associative arrays, determining which values in one column are not in another column (or determining which values are in both columns) is fast. For example, given two data tables with information about countries, which countries are in both data tables?

Place the columns of each data table that contain country names into associative arrays:

```
dt1 = Open( "$SAMPLE_DATA/BirthDeathYear.jmp" );
dt2 = Open( "$SAMPLE_DATA/World Demographics.jmp" );
aa1 = Associative Array( dt1:Country );
aa2 = Associative Array( dt2:Territory );
```

Use `N Items()` to see how many countries appear in each data table:

```
N Items(aa1);
   23
N Items(aa2);
   239
```

Use the `<<Intersect` message to find the common values:

```
aa1 = Associative Array( dt1:Country );
aa1 << Intersect( aa2 );
```

Look at the results:

```
Show(N Items(aa1), aa1 << Get Keys);
   N Items(aa1) = 21;
   aa1 << get keys = {"Australia", "Austria", "Belgium", "France", "Greece",
      "Ireland", "Israel", "Italy", "Japan", "Mauritius", "Netherlands", "New
      Zealand", "Norway", "Panama", "Poland", "Portugal", "Romania",
      "Switzerland", "Tunisia", "United Kingdom", "United States"};
```

This example uses a set operation called intersection. For more examples of using set operations with associative arrays to compare values, see "Associative Arrays in Set Operations".

Associative Arrays in Graph Theory

You can use associative arrays for graph theory data structures, such as the following directed graph example:

```
g = Associative Array();
g[1] = Associative Array({1, 2, 4});
g[2] = Associative Array({1, 3});
g[3] = Associative Array({4, 5});
g[4] = Associative Array({4, 5});
g[5] = Associative Array({1, 2});
```

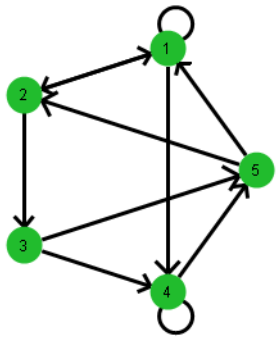
This is a two-level associative array. The associative array `g` contains five associative arrays (`1`, `2`, `3`, `4`, and `5`). In the containing array `g`, both the keys (1-5) and the values (the arrays that define the map) are important. In the inner associative arrays, the values do not matter. Only the keys are important.

The associative array represents the graph shown in Figure 7.2:

- node 1 is incident on nodes 1, 2, and 4
- node 2 is incident on nodes 1 and 3

- node 3 is incident on nodes 4 and 5
- node 4 is incident on nodes 4 and 5
- node 5 is incident on nodes 1 and 2

Figure 7.2 Example of a Directed Graph



The following depth-first search function can be used to traverse this graph, or any other directed graph represented as an associative array:

```
dfs = Function( {ref, node, visited},
    {chnode, tmp},
    Write( "\!NNode: ", node, ", ", ref[node] << Get Keys );
    visited[node] = 1;
    tmp = ref[node];
    chnode = tmp << first;
    While( !Is Missing( chnode ),
        If( !visited[chnode],
            visited = Recurse( ref, chnode, visited )
        );
        chnode = tmp << Next( chnode );
    );
    visited;
);
```

Notes:

- The first argument is the associative array that contains the map structure.
- The second argument is the node that you want to use as the starting point.
- The third argument is a vector that the function uses to keep track of nodes visited.

To see how the preceding example works, add the following expression to the end of the script:

```
dfs( g, 2, J( N Items( g << Get Keys ), 1, 0 ) );
```

Here is the output:

```
Node 2: {1, 3}
Node 1: {1, 2, 4}
Node 4: {4, 5}
Node 5: {1, 2}
Node 3: {4, 5}
[1, 1, 1, 1, 1]
```

The first five output lines show that starting from node 2, you can reach all the other nodes in the order in which they are listed. Each node also lists the nodes it is incident on (the keys). The value for each key is 1. The final line is a matrix that shows that you can reach each node from 2. If there were nodes that could not be reached from node 2, their values in the matrix would be 0.

Here is how to read the traversal of the nodes:

1. Start at node 2 and go to node 1.
2. From node 1, go to node 4.
3. From node 4, go to node 5.
4. From node 5, go back to node 2, and then to node 3.

Map Script

Here is the script that produced the map shown in Figure 7.2.

```
New Window( "Directed Graph",
   Graph Box(
      Frame Size( 300, 300 ),
      X Scale( -1.5, 1.5 ),
      Y Scale( -1.5, 1.5 ),
      Local( {n = N Items( g ), k = 2 * Pi() / n, r, i, pt, from, to,
         edge, v, d},
         Fill Color( "green" );
         Pen Size( 3 );
         r = 1 / (n + 2);
         For( i = 1, i <= n, i++,
            pt = Eval List( {Cos( k * i ), Sin( k * i )} );
            edges = g[i];
            For( edge = edges << First, !Is Empty( edge ),
               edge = edges << Next( edge ),
               to = Eval List( {Cos( k * edge ), Sin( k * edge )} );
               If( i == edge,
                  Circle( Eval List( 1.2 * pt ), 0.9 * r ), // else
                  v = pt - to;
                  d = Sqrt( Sum( v * v ) );
                  {from, to} = Eval List(
                     {pt * (d - r) / d + to * r / d, pt * r / d + to *
```

```
                        (d - r) / d}
                    );
                    Arrow( from, to );
                );
            );
            Circle( pt, r, "fill" );
            Text( Center Justified, pt - {0, 0.05}, Char( i ) );
        );
    )
  )
);
```

Associative Arrays in Set Operations

You can also use associative arrays to perform set operations. The following examples show how to take a union of two sets, a difference of two sets, and an intersection of two sets.

First, create three sets and view them in the log:

```
set_y = Associative Array( {"chair", "person", "relay", "snake", "tripod"} );
set_z = Associative Array( {"person", "snake"} );
set_w = Associative Array( {"apple", "orange"} );

// write the sets to the log
Write(
  "\!NExample:\!N\!tset_y = ",
  set_y << Get Keys,
  "\!N\!tset_z = ",
  set_z << Get Keys,
  "\!N\!tset_w = ",
  set_w << Get Keys
);
  Example:
    set_y = {"chair", "person", "relay", "snake", "tripod"}
    set_z = {"person", "snake"}
    set_w = {"apple", "orange"}
```

Union Operation

To find the union of two sets, insert one set into the other:

```
set_z << Insert( set_w );
Write( "\!N\!NUnion operation (set_w, set_z):,\!N\!tset_z = ", set_z << Get
  Keys );
  Union operation (set_w, set_z):,
    set_z = {"apple", "orange", "person", "snake"}
```

Difference Operation

To find the difference of two sets, remove one set from the other:

```
set_y << Remove( set_z );
Write( "\!N\!NDifference operation (set_z from set_y):\!N\!tset_y = ", set_y
   << Get Keys );
   Difference operation (set_z from set_y):
      set_y = {"chair", "relay", "tripod"}
```

Intersect Operation

To find the intersection of two sets, use the `aa << Intersect` message.

```
set_w << Intersect( set_z );
Write( "\!N\!NIntersect operation (set_w, set_z):\!N\!tset_w = ", set_w << Get
   Keys );
   Intersect operation (set_w, set_z):
      set_w = {"apple", "orange"}
```

Example of Using Set Operations

Given a list of names, which of them are not contained in Big Class.jmp? You can find the answer by taking the difference of the two sets of names.

1. Get the list of names and open the data table:

   ```
   names list = {"ROBERT", "JEFF", "CHRIS", "HARRY"};
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   ```

2. Put the list of names into an associative array:

   ```
   names = Associative Array( names list );
   ```

3. Perform a difference operation by removing the column values from your list:

   ```
   names << Remove( Associative Array( dt:name ) );
   ```

4. Look at the result:

   ```
   Write( "\!NWhich of {ROBERT, JEFF, CHRIS, HARRY}, is not in Big Class = ",
      names << Get Keys );
      Which of {ROBERT, JEFF, CHRIS, HARRY}, is not in Big Class = {"HARRY", "JEFF"}
   ```

Chapter 8

Programming Methods

Complex Scripting Techniques and Additional Functions

This chapter includes advanced techniques, such as throwing and catching exceptions, encrypting scripts, and using complex expressions.

Contents

Lists and Expressions 255
Expressions 255
Macros 267
Manipulating Lists 267
Manipulating Expressions 270
Advanced Scoping and Namespaces 274
Names Default To Here 274
Scoped Names 278
Namespaces 282
Referencing Namespaces and Scopes 286
Resolving Named Variable References 289
Best Practices for Advanced Scripting 290
Advanced Programming Concepts 291
Throw and Catch Exceptions 291
Functions 293
Recursion 296
Includes 296
Load and Save Text Files 297
Run External Programs 297
Classes 299
File and Directory Operations 302
Select a Directory or File 302
Get a List of Filenames 303
Connect to a Network Directory 304
Scripting By Groups 305
Encrypt and Decrypt Scripts 306
Encryption and Global Variables 308
Encrypting Scripts in Data Tables 308
Working with Expressions in an Encrypted Script 309
Additional Numeric Functions 310
Derivatives 310
Algebraic Manipulations 312
Maximize and Minimize 314
Identifying Differences Between Strings, Lines, or Sequences 315
Scheduling Actions 317
Functions that Communicate with Users 318
Writing to the Log 319
Send Information to the User 320

Lists and Expressions

Expressions

This section delves into JSL expressions and introduces several JSL functions that can be used to manipulate JSL expressions. These functions provide JSL programmers with a rich, powerful, meta programming toolkit. One of the benefits of such a toolkit is that it provides JSL programmers with a powerful macro system capability that is similar in functionality to macro systems available in other modern programming languages.

In order to understand JSL expressions, it is useful to keep in mind that everything that happens in JSL is a result of a function call. Function calls come in three varieties:

prefix The function name comes before its arguments (for example, `Foo( m, n )`). Most function calls in JSL are of this form.

infix The function name is between its arguments (for example, `m + n`). This form is common for mathematical operators.

special Functions like `[ ]` (the subscript operator) fall into this category. Since there are only a handful of special functions, they are not discussed further in this section.

Although function calls come in three forms, all function calls can be written in prefix form. It turns out that JMP actually uses the prefix form internally and so, in general, it is very useful to think of JSL expressions in this way.

To illustrate this idea, begin by examining the syntactic structure of JSL expressions a bit more closely. All JSL expressions must conform to a very simple syntactic structure. Each expression consists of a *head* expression and a sequence of zero or more argument expressions. Consider the following JSL expressions, written first in infix form and then in prefix form, to see how this syntactic structure applies:

```
x = 100.1
m & n
1 <= z <= 20
```

Using the prefix form, these expressions are written equivalently:

```
Assign( x, 100.1 )
And( m, n )
Less or Equal( 1, z, 20 )
```

The head of the first example is `Assign` and its arguments are `x` and 100.1. For the second example, the head is `And` and its arguments are `m` and `n`, whereas for the third example, the head is `Less or Equal` and its arguments are 1, `z`, and 20. For these examples, the "=", "&", and "<=" operators are just human friendly ways of representing `Assign`, `And`, and `Less or Equal`.

Examine a few additional simple examples.

```
Assign( 100.1, x )
And( )
Less or Equal( z )
```

These examples also conform to the required syntactic structure, but they differ from the previous examples in that they all produce an error message when they are evaluated. This point leads to an important distinction that you should make when thinking about JSL expressions. That is, you should separate the idea of the syntactic structure of a JSL expression from what happens when the expression is evaluated.

Consider the following examples that present both the human friendly form and the prefix form in each case:

```
x = 100.1 + Log( z )                    // human friendly form
Assign( x, Add( 100.1, Log( z ) ) )     // prefix form
```

This example illustrates that the arguments of a head expression can also be expressions.

```
// human friendly form
x = 0.8;
x ||= Random Uniform( 0.2, 0.6 );

// prefix form
Glue(
   Assign( x, 0.8 ),
   Concat To( x, Random Uniform( 0.2, 0.6 ) )
)
```

This example would typically be referred to as a JSL script, but it is just another case of a JSL expression. Here, the *head* expression is `Glue`, which has two arguments: the expressions `Assign( x, 0.8 )` and `Concat To( x, Random Uniform( 0.2, 0.6 ) )`. This example illustrates two important points:

1. The ";" symbol is another instance of a human friendly operator. In this case, it is for the `Glue` function.
2. There is no distinction between what are typically referred to as a JSL script and a JSL expression.

The remainder of this section covers how to assign an expression to a variable and then how to manipulate, or execute, that expression when needed. For information about how JMP resolves names when evaluating JSL expressions, see the section "Rules for Name Resolution".

Quoting and Unquoting Expressions

The operators to control when expressions are evaluated are `Expr` and `Eval`, which you should think of as the procrastination and eager operators. `Expr` just copies its argument as an expression, rather than evaluating it. `Eval` does the opposite: it evaluates its argument, and then takes that result and evaluates it again.

`Expr` and `Eval` can be thought of as quoting and unquoting operators, telling JMP when you mean the expression itself, and when you mean the result of evaluating the expression.

The following examples all assume these two assignments:

```
x = 1; y = 20;
```

If you assign the expression `x+y` to *a*, quoting it as an expression with `Expr`, then whenever *a* is evaluated, it evaluates the expression using the current values of *x* and *y* and returns the result. (Exceptions are the utilities `Show`, `Write`, and `Print`, which do not evaluate expressions for pure name arguments.)

```
x = 1; y = 20;
a = Expr( x + y );
a;
   21
```

If you want the expression that is stored in the name, rather than the result of evaluating the expression, use the `NameExpr` function. See "Retrieve a Stored Expression Rather than its Result".

```
x = 1; y = 20;
Show( Name Expr( a ) );
   NameExpr(a) = x + y
```

If you assign an extra level of expression-quoting, then when *a* is evaluated, it unpacks one layer and the result is the expression `x+y`.

```
x = 1; y = 20;
a = Expr( Expr( x + y ) );
Show( a );
   a = Expr(x + y)
```

If you want the value of the expression, then use `Eval` to unpack all layers:

```
x = 1; y = 20;
Show( Eval( a ) );
   Eval(a) = 21
```

You can do this to any level, for example:

```
x = 1; y = 20;
a = Expr( Expr( Expr( Expr( x + y ) ) ) );
b = a;
   Expr( Expr( x + y ) )
c = Eval( a );
   Expr( x + y )
d = Eval( Eval( a ) );
   x+y
e = Eval( Eval( Eval( a ) ) );
   21
```

For more information about expression handling, see Morgan (2010).

Quote an expression as a string

The `JSL Quote()` function returns the contents of an expression as a quoted string. Comments and white space in the string are preserved.

The following expression is an example:

```
x = JSL Quote( /* Begin quote. */
For (i = 1, i <= 5, i++,
// Print the value of i.
Print( i );
);
// End expression.
);
Show( x );
```

In the output, the contents of the `JSL Quote()` function are enclosed in quotes.

```
x = " /* Begin quote. */
For (i = 1, i <= 5, i++,
   // Print the value of i.
Print(i);
);
   // End expression.
";
```

Quoting Expressions

A common use of the `Expr()` function is to store an expression in a JSL variable. This could be considered as a macro. Consider the following script:

```
dist = Expr( Distribution( Column( height ) ) );
```

To execute the above expression, just use the following:

```
dist;
```

You can use expressions in a loop to execute it:

```
For( i = 0, i < 3, i = i + 1, dist );
```

This loop results in three Distribution platform reports for height.

Use `Eval()` to evaluate an expression explicitly:

```
Eval( dist );
```

Note, however, that in column formulas, `Eval()` only works if it is outermost in the formula. So, for example,

```
Formula( Log( Eval( Column Name( i ) ) ) );
```

would generate an error. Instead, use:

```
Formula( Eval( Substitute( Expr( Log( xxx ) ), Expr( xxx ), Column Name( i ) )
  ) );
```

As another example,

```
Formula( Eval( Column Name( i ) ) + 10 );
```

generates an error, since `Eval()` is actually under the `Add` function. Instead, use:

```
Formula(Eval(Substitute(Expr(xxx+10), Expr(xxx), column name(i))))
```

Retrieve a Stored Expression Rather than its Result

What if you wanted the symbolic value of a variable (such as the expression `Distribution(Column(height))` stored in `dist` above), rather than the evaluation of it (the actual launched platform)? The `Name Expr` function does this. `Name Expr` retrieves its argument as an expression without evaluating it.

`Expr` returns its argument exactly, whereas `Name Expr` looks up the expression stored in its argument. `Name Expr` "unpacks" just one layer to get the expression, but does not keep unpacking to get the result.

For example, you would need to use this if you had an expression stored in a name and you wanted to edit the expression:

```
popVar = Expr( Summation( i = 1, N Row(), (y[i] - Col Mean( y )) ^ 2 / N Row()
  ) );
  Summation( i = 1, N Row(), (y[i] - Col Mean( y )) ^ 2 / N Row() )

unbiasedPopVar = Substitute( Name Expr( popVar ), Expr( Wild()/N Row() ),
  Expr( (y[i] - Col Mean( y )) ^ 2 / ( N Row() - 1 ) ) );
  Summation( i = 1, N Row(), (y[i] - Col Mean( y )) ^ 2 / (N Row() - 1) )
```

Compare `x`, `Expr(x)`, `NameExpr(x)`, and `Eval(x)` after submitting this script:

```
a = 1; b = 2; c = 3;
x = Expr( a + b + c );
```

Table 8.1 Compare Eval, Name Expr, and Expr

Command and result	Explanation
`x;` *6*	Evaluates *x* to the expression `a+b+c`, and then evaluates the expression, returning the result, 6 (unpacks *all* layers).
`Eval( x );` *6*	Equivalent to simply calling *x*. Evaluates *x* to the expression `a+b+c`, and then evaluates the expression, returning the result, 6 (unpacks *all* layers).
`NameExpr( x );` *a+b+c*	Returns the expression that was stored in *x*, which is `a+b+c` (unpacks the *outside* layer).
`Expr( x );` *x*	Returns the expression `x` (packs *one* layer).

JSL also supports functions to access and traverse expressions, all of them either a name or a literal expression as an argument. In the following, *expressionArg* is either a single name or a compound expression to be taken literally.

`NArg(`*expressionArg*`)` finds the number of arguments in *expressionArg*.

The *expressionArg* can be a name holding an expression, an expression evaluated to an expression, or a literal expression quoted by `Expr()`.

`NArg(`*name*`)` obtains the expression held in `name` (it is not evaluated) and returns the number of arguments

`NArg(`*expression*`)` evaluates `expression` and returns the number of arguments

`NArg(Expr(`*expression*`))` returns the number of arguments to literal expression.

For example, if `aExpr = {a+b,c,d,e+f+g};`

- `NArg(aExpr)` results in 4.
- `NArg(Arg(aExpr,4))` results in 3.
- `NArg(Expr({1,2,3,4}))` results in 4.

`Head(`*expressionArg*`)` returns the head of the expression without any arguments. If the expression is an infix, prefix, or postfix special character operator, then it is returned as the functional equivalent.

The *expressionArg* can be a name holding an expression, an expression evaluated to an expression, or a literal expression quoted by `Expr()`.

For example, if `aExpr = expr(a+b);`

- `r = Head(aExpr)` results in `Add()`.
- `r = Head (Expr(sqrt(r)))` results in `Sqrt()`.
- `r = Head({1,2,3})` results in `{}` .

`Arg(`*`expressionArg`*`,`*`indexArg`*`)` extracts the specified argument of the symbolic expression, resulting in an expression.

For example, `Arg(`*`expressionArg`*`,i)` extracts the i^{th} argument of *`expressionArg`*.

The *`expressionArg`* can be a name holding an expression, an expression evaluated to an expression, or a literal expression quoted by `Expr()`.

- `Arg(`*`name`*`,i)` obtains the expression held in *`name`* (it is not evaluated) and finds the `i`th argument.
- `Arg(`*`expression`*`,i)` evaluates *`expression`* and finds the `i`th argument.
- `Arg(Expr(`*`expression`*`),i)` finds the i^{th} argument of *`expression`*.

As another example, if `aExpr = Expr(12+13*sqrt(14-15));`

- `Arg(aExpr,1)` yields `12`.
- `Arg(aExpr,2)` yields `13*sqrt(14-15)`.
- `Arg(Expr(12+13*sqrt(14-15)),2)` yields `13*sqrt(14-15)`.

To extract an argument of an argument inside an expression, you can nest Arg commands:

- `Arg(Arg(aExpr,2),1)` yields the first argument within the second argument of `aExpr`, or 13.
- `Arg(Arg(aExpr,2),2)` yields `Sqrt(14-15)`.
- `Arg(Arg(Arg(aExpr,2),2),1)` yields `14-15`.
- `Arg(Arg(Arg(aExpr,2),2),3)` yields `Empty()`

Here is a description of how the last example line unwraps itself:

1. The inner `Arg` statement is evaluated.

   ```
   Arg(aExpr,2)
     13 * Sqrt( 14 - 15 )
   ```

2. Then the next one is evaluated.

   ```
   Arg(Arg(aExpr,2),2)
   // this is equivalent to Arg(Expr (13 * Sqrt( 14 - 15 ) ), 2)
     Sqrt( 14 - 15 )
   ```

3. Finally, the outer Arg is evaluated.

   ```
   Arg(Arg(Arg(aExpr,2),2),3)
   // this is equivalent to Arg (Expr (Sqrt( 14 - 15 ) ), 3)
   ```

```
Empty()
```

There is only one element to the `Sqrt` expression, so a request for the third argument yields `Empty()`. To access the two arguments inside the `Sqrt` expression, try this:

```
Arg(Arg(Arg(Arg(aExpr,2),2),1),2);
  15
```

`HeadName(`*`expressionArg`*`)` returns the name of the head of the expression as a string. If the expression is an infix, prefix, postfix, or other special character operator, then it is returned as the functional equivalent.

The *`expressionArg`* can be a name holding an expression, an expression evaluated to an expression, or a literal expression quoted by `Expr()`.

For example, if `aExpr = expr(a+b);`

- `r = HeadName (aExpr)` results in `"Add"`.
- `r = HeadName (Expr(sqrt(r)))` results in `"Sqrt"`.
- `r = HeadName ({1,2,3})` results in `"List"`.

In previous versions of JMP, other versions of `Arg`, `Narg`, `Head`, and `HeadName` were implemented, called `ArgExpr`, `NArgExpr`, `HeadExpr`, and `HeadNameExpr`, respectively. These did the same thing, but did not evaluate their argument. These forms are now deprecated and will not be documented in future versions.

Making Substitutions

`Eval Insert` is for the situation where you want to make substitutions, by evaluating expressions inside a character string.

With `Eval Insert`, you specify characters that delimit the start and end of an expression, and everything in between is evaluated and expanded.

There are two functions, one to return the result, the other to do it in-place.

```
resultString = EvalInsert( string with embedded
  expressions,startDelimiter,endDelimiter )
EvalInsertInto( string l-value with embedded
  expressions,startDelimiter,endDelimiter )
```

The delimiter is optional. The default start delimiter is `"^"`. The default end delimiter is the start delimiter.

```
xstring = "def";
r = Eval Insert( "abc^xstring^ghi" );  // returns "abcdefghi";

r = "abc^xstring^ghi"; // in-place evaluation
Eval Insert Into( r ); // r now has "abcdefghi";
```

```
// with specified delimiter
r = Eval Insert( "abc%xstring%ghi","%" );  // returns "abcdefghi";

// with different start and end delimiters
r = Eval Insert( "abc[xstring]ghi","[","]" );  // returns "abcdefghi";
```

When a numeric value contains locale-specific formatting, include the <<Use Locale(1) option. The following example substitutes a comma for the decimal point based on the computer's locale setting.

```
Eval Insert( "^1.2^", <<Use Locale( 1 ) );
  1,2
```

Evaluate expressions inside lists

`Eval List` evaluates expressions inside a list and returns a list with the results:

```
x = { 1 + 2, 3 + 4 };
y = Eval List( x );   // returns y is {3,7}
```

`Eval List` is useful for loading the list of user choices returned by `Column Dialog` or `New Window` with the `Modal` argument.

Evaluate expressions inside expressions

`Eval Expr()` evaluates only the inner expressions and returns an expression that contains the results of the evaluation. By comparison, `Eval` evaluates the inner expressions, takes the results, and then evaluates it again.

Suppose that a data table contains a column named X3. Here is an example of using `Eval Expr()` to evaluate the inner expression first:

```
x = Expr( Distribution( Column( Expr("X"||Char( i ) ) ) ) );
i = 3;
y = Eval Expr( x );   // returns Distribution( Column( "X3" ) )
```

To evaluate further, you need to either call the result in a subsequent step, or else put `Eval()` around the `Eval Expr()`. The following examples create a Distribution report.

```
// two-step method
x = Expr( Distribution( Column( Expr( "X" || Char( i ) ) ) ) );
i = 3;
y = Eval Expr( x );
y;

// one-step method
x = Expr( Distribution( Column( Expr( "X" || Char( i ) ) ) ) );
i = 3;
Eval( Eval Expr( x ) );
```

See Table 8.3 to learn what would happen if you tried to use Eval directly on *x* without first doing Eval Expr.

Parsing Strings into Expressions, and Vice Versa

Parsing is the syntactic scanning of character strings into language expressions. Suppose that you have read in a valid JSL expression into a character string, and now want to evaluate it. The Parse function returns the expression. To evaluate it, use the Eval function.

```
x = Parse( "a=1" ) ;  // x now has the expression a=1
Eval( Parse( "a=1" ) ); // a now has the value 1
```

To go in the reverse, use the Char function, which converts an expression into a character string. Usually the argument to a Char function is an Expr function (or a NameExpr of a variable), since Char evaluates its argument before deparsing it.

```
y = Char( Expr( a = 1 ) ); // results in y having the character value "a=1"
z = Char( 42 ); // returns "42"
```

The Char function allows arguments for field width and decimal places if the argument is a number. The default is 18 for width and 99 for decimal (Best format).

```
Char( 42, 5, 2 ); // returns the character value "42.00"
```

To preserve locale-specific formatting in the numeric value, include the <<Use Locale(1) option as shown in the following example:

```
Char( 42, 5, 2, <<Use Locale(1) ); // returns the character value "42,00" in
  the French locale
```

The reverse of Char is not quite as simple. To convert a character string into an expression, you use Parse, but to convert a character string into a number value, you use Num.

```
Parse( y );
Num( z );
```

Table 8.2 Functions to Store or Evaluate Expressions

Function	Syntax	Explanation
Char	Char(Expr(*expression*)) Char(*name*)	Converts an *expression* into a character string. The expression must be quoted with Expr; otherwise its evaluation is converted to a string.
	string = char(*number*, *width*, *decimal*)	Converts a *number* into its character representation. *Width* and *decimal* are optional arguments to specify formatting; the default is 18 for width and 99 for decimal.

Table 8.2 Functions to Store or Evaluate Expressions *(Continued)*

Function	Syntax	Explanation
Eval	Eval(*x*)	Evaluates *x*, and then evaluates the result of *x* (unquoting).
Eval Expr	Eval Expr(*x*)	Returns an expression with all the expressions inside *x* evaluated.
Eval List	Eval List(*list*)	Returns a list of the evaluated expressions inside *list*.
Expr	Expr(*x*)	Returns the argument unevaluated (expression-quoting).
NameExpr	NameExpr(*x*)	Returns the unevaluated expression of *x* rather than the evaluation of *x*. NameExpr is like Expr except that if *x* is a name, NameExpr returns the unevaluated expression stored in the name rather than the unevaluated name *x*.
Num	Num("*string*")	Converts a character string into a number.
Parse	Parse("*string*")	Converts a character string into a JSL expression.

Summary

Table 8.3 compares various ways that you can use the evaluation-control functions with *x*. Assume that a data table contains a column named X3, and *x* and *i* have been assigned:

```
x = Expr( Distribution(Column( Expr("X"||Char( i ) ) ) ) );
i = 3;
```

Table 8.3 Functions for Controlling Evaluation

Commands and results	Explanation

x; // or
Eval(x);
Not Found in access or evaluation of 'distribution' , Bad Argument({"X" || Char(i)})
	Eval(x) and simply calling x are equivalent.
	Evaluates the expression distribution(column(expr("X" || Char(i)))). This results in errors. The column name is recognized as "X"||Char(i) because it is packed by the Expr() function.

Expr(x); *x*	Returns the expression x (packs *an additional* layer).

Table 8.3 Functions for Controlling Evaluation *(Continued)*

Commands and results	Explanation
`Name Expr(x);` *`Distribution(Column(Expr("X" || Char(i))))`*	Returns the expression stored in *x* exactly as is: `Distribution(Column(Expr("X" || Char(i))))`.
`y=Eval Expr(x);` *`Distribution(Column("X3"))`*	Evaluates the inner expression but leaves the outer expression unevaluated, so that *y* is `Distribution(Column("X3"))`.

`y; //or`
`Eval(Eval Expr(x));`
`Distribution[]`
	`Eval(eval expr(x))` and simply calling `y` are equivalent.
	Evaluates `Distribution(Column("X3"))` to launch the platform.

`z = Char(nameexpr(x));` *`"Distribution(Column(Expr (\!"X\!" || Char(i))))"`*	Quotes the entire expression as a text string, adding `\!"` escape characters as needed. Note that `Char(x)` would first attempt to evaluate *x*, producing an error and ultimately returning a quoted missing value: `"."`
`Parse(z);` *`Distribution(Column(Expr("X" || Char(i))))`*	Unquotes the text string and returns an expression.

`a = Parse(Char( NameExpr(x)));`
`Eval(EvalExpr(a));`
`Distribution[]`
	Evaluation control taken to its logical extreme.
	Note that you must break this into at least two steps as shown. Combining it into one giant step produces different results because the `Eval Expr` layer causes the `Parse` layer to be copied literally, not executed.
	`Eval(`
	`EvalExpr(`
	`Parse(`
	`Char(`
	`NameExpr(x)))));`
	`Distribution(Column(Expr("X" || Char(i))))`

Macros

Stored expressions can serve as a macro feature. You can store a generalized action as an expression in a JSL variable, and then call that variable wherever you need that action to be performed. This example has four macros as the arguments to `If`:

```
lastStdzdThickness=expr(
  (thickness[nrow()]-col mean(thickness)) / col std dev(thickness));
continue=expr(...<script to read in more data>...);
log=expr(print("In control at "||char( long date(today()))));
break=expr(...<script to shut down process>...); limitvalue=1;

if(lastStdzdThickness<limitvalue,log;continue,break);
```

Storing the expression with `Expr` delays its evaluation until the variable is evaluated. Any variables, data points, or expressions included in that expression are evaluated on the fly when the expression is evaluated. See "Expressions" for detailed rules for storing expressions and later evaluating them.

Manipulating Lists

The following functions manipulate lists. They can also be used to manipulate expressions, as shown in the next section, "Manipulating Expressions". A summary of commands with explanations is in Table 8.4.

Most of the functions have two variants, one that produces a new value, and one that works in-place directly on its arguments. Here are some examples:

```
// In-place example
A = {2, 3, 4, 2, 1, 2, 1};
A = Remove( A, 3 ); // delete the third item in the list A, storing result in A
Remove From( A, 3 ); // delete the third item in the list A, in place
Show( A ); // A = {2, 3, 2, 1, 2, 1};

// New value example
B = {2, 3, 4, 2, 1, 2, 1};
onetwo = Insert( {1}, 2 );  // {1, 2}
Insert Into( B, {1, 2}, 4 ); // put 1, 2 before the current 4th item
Show( B ); // B = {2, 3, 4, 1, 2, 2, 1, 2, 1};
```

Note: If the position is omitted in the `Insert Into` command, items are placed at the end of the list.

```
a = Shift( {1,2,3,4}, 1 ); // store the list {2,3,4,1} in a
Shift Into( a, -1 );
Show( a ); // a = {1, 2, 3, 4};
```

```
b = Reverse( a ); // b is now {4,3,2,1}
Reverse Into( a ); // a is now {4,3,2,1}
Show( a ); // a = {4, 3, 2, 1};

s = Sort List( {1,4,2,5, -7.2, Pi(), -11, cat, apple, cake} );
Show( s ); // s = {-11, -7.2, 1, 2, 4, 5, apple, cake, cat, Pi()};

c = {5, pie, 2, Pi(), -2};
Sort List Into( c );
Show( c ); // c = {-2, 2, 5, Pi(), pie};
```

In-place functions

In-place functions are those that operate on lists or expressions directly. They have `From` or `Into` in their names (for example, `Remove From` and `Insert Into`). They do not return a result; you have to show the list to see the result. The first argument for an in-place function must be an L-value. An *L-value* is an entity such as a variable whose value can be set.

```
myList = {a, b, c, d};
Insert Into( myList, 2, 3 );
Show( myList );
   myList = {a, b, 2, c, d}
```

These examples show how to use `Insert Into` and `Remove From` with nested lists:

```
a = {{1, 2, 3}, {"A", "B", "C"}};
Show( a );
   a = {{1, 2, 3}, {"A", "B", "C"}}

Insert Into( a[1], 99, 1 );
Show( a );
   a = {{99, 1, 2, 3}, {"A", "B", "C"}}

Remove From( a[1], 1 );
Show( a );
   a = {{1, 2, 3}, {"A", "B", "C"}}
```

Not in-place functions

For the *not-in-place functions,* you must either state the list directly or else quote a name that evaluates to a list. Such functions do *not* have `From` or `Into` in their names. They *return* manipulated lists or expressions without changing the original list or expression given in the first argument.

```
myNewList = Insert( {a, b, c, d}, 2, 3 );
   {a, b, 2, c, d}

oldList = {a, b, c, d};
newList = Insert( oldList, 2, 3 );
```

```
{a, b, 2, c, d}
```

Substituting

`Substitute()` and `Substitute Into()` merit further discussion. Both functions find all matches to a pattern in a list (or expression) and replace them with another expression. Each pattern must be a name. The arguments are evaluated before they are applied, so most of the time you must quote them with an `Expr()` function.

```
Substitute( {a,b,c}, Expr( a ), 23); // returns {23,b,c}
Substitute( Expr( Sine( x ) ), Expr( x ), Expr( y ) ); // returns Sine(y)
```

To delay evaluating an argument, use `NameExpr` instead of `Expr`:

```
a = {quick, brown, fox, jumped, over, lazy, dogs};
b = Substitute( a, Expr( dogs ), Expr( cat ) );
canine = Expr( dogs );
equine = Expr( horse );
c = Substitute( a, Name Expr( canine ), Name Expr( equine ) );
Show( a, b, c );
   a = {quick,brown,fox,jumped,over,lazy,dogs}
   b = {quick,brown,fox,jumped,over,lazy,cat}
   c = {quick,brown,fox,jumped,over,lazy,horse}
```

`Substitute Into` does the same work, in place:

```
Substitute Into( a, Expr( dogs ), Expr( horse ) );
```

You can list multiple pattern and replacement arguments to do more than one replacement in a single step:

```
d = Substitute( a,
   Name Expr( quick ), Name Expr( fast ),
   Name Expr( brown ), Name Expr( black ),
   Name Expr( fox ), Name Expr( wolf )
);
   {fast,black,wolf,jumped,over,lazy,dogs}
```

Note that substitutions are done repeatedly over multiple instances of the expression pattern. For example:

```
Substitute( Expr( a + a ), Expr( a ), Expr( aaa ) );
```

results in:

```
aaa + aaa
```

Manipulating Expressions

The functions for manipulating lists can also operate on most expressions. Be sure to quote the expression with `Expr()`. For example:

```
Remove( Expr( A + B + C + D ), 2 ); // results in the expression A + C + D
b = Substitute( Expr( Log( 2 ) ^ 2 / 2 ), 2, 3 ); /* returns the expression
  Log( 3 ) ^ 3 / 3*/
```

As with lists, remember that the first argument for in-place functions must be an L-value. An L-value is an entity such as a variable whose value can be set. In-place functions are those that operate on lists or expressions directly. They have `From` or `Into` in their names (for example, `Remove From` and `Insert Into`). They do not return a result; you have to show the expression to see the result.

```
polynomial = Expr( a * x ^ 2 + b * x + c );
Insert Into( polynomial, Expr( d * x ^ 3 ), 1 );
Show( polynomial );
  polynomial = d * x ^ 3 + a * x ^ 2 + b * x + c
```

For the not-in-place functions, you must either state the expression directly or else quote a name that evaluates to an expression using `NameExpr`. Such functions do *not* have `From` or `Into` in their names. They *return* manipulated lists or expressions without changing the original list or expression given in the first argument.

```
cubic = Insert( Expr( a * x ^ 2 + b * x + c ), Expr( d * x ^ 3 ), 1 );
  d * x ^ 3 + a * x ^ 2 + b * x + c

quadratic = Expr( a * x ^ 2 + b * x + c );
cubic = Insert( Name Expr( quadratic ), Expr( d * x ^ 3 ), 1 );
  d * x ^ 3 + a * x ^ 2 + b * x + c
```

Substituting

Substituting is extremely powerful; please review the earlier discussion "Substituting". Here are a few notes regarding substituting for expressions.

`Substitute(`*`pattern`*`,`*`name`*`,`*`replacement`*`)` substitutes for names in expressions

`NameExpr()` looks through the name but copies instead of evaluates:

```
a = Expr(
  Distribution( Column( x ), Normal Quantile Plot )
);
Show( Name Expr( a ) );
  Name Expr(a) = Distribution(Column(x), Normal Quantile Plot);
```

`Substitute()` evaluates all its arguments, so they must be quoted correctly:

```
b = Substitute( Name Expr( a ), Expr( x ), Expr( :weight ) );
```

```
Show( Name Expr( b ) );
  Name Expr(b) = Distribution(Column(:weight), Normal Quantile Plot);
```

SubstituteInto() needs an L-value, so the first argument is not quoted:

```
Substitute Into( a, Expr( x ), Expr( :weight ) );
Show( Name Expr( a ) );
  Name Expr(a) = Distribution(Column(:weight), Normal Quantile Plot);
```

Substitute() is useful for changing parts of expressions, such as in the following example that tests the Is functions:

```
data = {1, {1, 2, 3}, [1 2 3], "abc", x, x( y )};
ops = {is number, is list, is matrix, is string, is name, is expr};
m = J( N Items( data ), N Items( ops ), 0 );
test = Expr(
  m[r, c] = _op( data[r] )
);
For( r = 1, r <= N Items( data ), r++,
  For( c = 1, c <= N Items( ops ), c++,
    Eval( Substitute( Name Expr( test ), Expr( _op ), ops[c] ) )
  )
);
Show( m );
  m =
  [1 0 0 0 0 0,
  0 1 0 0 0 1,
  0 0 1 0 0 0,
  0 0 0 1 0 0,
  0 0 0 0 1 1,
  0 0 0 0 0 1];
```

You can use SubstituteInto() to have JMP solve quadratic equations. The following example solves $4x^2 - 9 = 0$:

```
/* FIND THE ROOTS FOR THE EQUATION:
       a*x^2 + b*x + c = 0
The quadratic formula is x=(-b + - sqrt(b^2 - 4ac))/2a.
Use a list to store both the + and - results of the +- operation */
x = {Expr(
  (-b + Sqrt( b ^ 2 - 4 * a * c )) / (2 * a)
), Expr(
  (-b - Sqrt( b ^ 2 - 4 * a * c )) / (2 * a)
)};
Substitute Into( x, Expr( a ), 4, Expr( b ), 0, Expr( c ), -9 );

// plug in the coefficients
Show( x ); // see the result of substitution
Show( Eval Expr( x ) ); // see the solution
  x = {Expr(((-0) + Sqrt(0 ^ 2 - 4 * 4 * -9)) / (2 * 4)), Expr(((-0) - Sqrt(0
    ^ 2 - 4 * 4 * -9)) / (2 * 4))};
```

```
Eval Expr(x) = {1.5, -1.5};
```

The functions for manipulating lists and expressions are discussed in the previous section, "Manipulating Lists", and summarized in Table 8.4.

Table 8.4 Functions for Manipulating Lists or Expressions

Function	Syntax	Explanation

`Remove`	`x = Remove(list|expr)`
	`x = Remove(list|expr, position)`
	`x = Remove(list|expr, {positions})`
	`x = Remove(list|expr, position, n)`
		Copies the *list* or *expression*, deleting the item(s) at the indicated *position*. If *position* is omitted, items are deleted from the end. *Position* can be a list of positions. An extra argument, *n*, deletes *n* items instead of just 1.

`Remove From`	`Remove From(list|expr, position)`
	`Remove From(list|expr)`
	`Remove From(list|expr, position, n)`
		Remove items in place. The function returns the removed item(s), but you do not have to assign them to anything. The first argument must be an L-value.

`Insert`	`x = Insert(list|expr, item, position)` `x = Insert(list|expr, item)`	Inserts a new item into the *list* or *expression* at the given position. If position is not given, it is inserted at the end.
`Insert Into`	`Insert Into(list|expr, item, position)` `Insert Into(list|expr, item)`	Same as `Insert`, but does it in place. *List* or *expression* must be an L-value.
`Shift`	`x = Shift(list|expr)` `x = Shift(list|expr, n)`	Shift an item or *n* items from the front to the back of the *list* or *expression*. Shift items from back to front if *n* is negative.

Table 8.4 Functions for Manipulating Lists or Expressions *(Continued)*

Function	Syntax	Explanation
Shift Into	Shift Into(*list*|*expr*) Shift Into(*list*|*expr*, *n*)	Shift items in place.
Reverse	*x*=Reverse(*list*|*expr*)	Reverse the order of elements of a *list* or terms of an *expression*.
Reverse Into	Reverse Into(*list*|*expr*)	Reverse the order of elements of a *list* or terms of an *expression* in place.
Sort List	*x*=Sort List(*list*|*expr*)	Sort the elements of a *list* or the terms of an *expression*. Numbers sort low, followed by the name value of names, strings, or operators. For example 1+2 is lower than 1-2 because the name value Add sorts lower than the name value Subtract. {1,2} sorts lower than {1,3}, which sorts lower than {1,3,0}. {1000} sorts lower than {“a”}, but {a} and {“a”} sort as equal.
Sort List Into	Sort List Into(*list*|*expr*)	Sort the elements of a *list* or terms of an *expression* in place.
Sort Ascending	Sort Ascending(*list*|*matrix*)	Returns a copy of a *list* or *matrix* with the items in ascending order.
Sort Descending	Sort Descending(*list*|*matrix*)	Returns a copy of a *list* or *matrix* with the items in descending order.
Loc Sorted	Loc Sorted(*A*, *B*)	Creates a matrix of subscript positions where the values in matrix **A** match the values in matrix **B**. **A** must be a matrix sorted in ascending order.
Substitute	R = Substitute(*list*|*expr*, Expr(*pattern*), Expr(*replacement*), ...)	Finds all matches to the pattern in the *list* or *expression*, and replaces them with the replacement expression. Each pattern must be a name. The second and third arguments are evaluated before they are applied, so most of the time you must quote them with an Expr function. To delay evaluating an argument, use Name Expr instead of Expr. You can list multiple pattern-replacement pairs for multiple substitutions in one statement.

Table 8.4 Functions for Manipulating Lists or Expressions *(Continued)*

Function	Syntax	Explanation
Substitute Into	Substitute Into(*list|expr*, Expr(*pattern*), Expr(*replacement*), ...)	Substitute in place.

Advanced Scoping and Namespaces

Scripts that are used in production environments need to use more advanced scoping techniques to avoid collisions between scripts. JMP provides three progressively more advanced techniques:

- The `Names Default To Here()` function. If you have simple scripting needs, this single command might be sufficient. See "Names Default To Here".
- Scopes that are pre-defined by JMP. See "Scoped Names".
- Namespaces that you can create for your scripts. See "Namespaces".

Names Default To Here

If you write production scripts, you need to insulate the script from the current user environment. Otherwise, the variables that you use might interact with variables used by the user and by other scripts. The way to do this is to keep your names in a local environment, which you can do by setting an execution mode with the statement:

```
Names Default To Here( 1 );
```

Unqualified names in a script with the *Names Default To Here* mode turned on are private to that script. However, the names persist as long as the script persists, or as long as objects created by or holding the script are still active. It is recommend that all production scripts start with `Names Default To Here(1)` unless there is a specific reason not to do so. When the script uses an unqualified name in this mode, that name is resolved in the local namespace.

To refer to global variables, scope the name specifically as a global variable (for example, `::global_name`). To refer to columns in a data table, scope with name specifically as a data table column (for example, `:column_name`).

Note: `Names Default To Here( 1 )` defines a mode for a particular script. It is not a global function. One script can have this mode turned on, while another script can have it turned off. The default setting is off.

In JMP 8 and earlier, the only method to the insulate scripts was to use lengthy names that were less likely to collide with names in other scripts. Using `Names Default To Here(1)` makes this technique unnecessary.

If you have simple scripting needs, `Names Default To Here(1)` might be sufficient.

Local()

`Local()` creates local scopes only in specific contexts within a script and cannot enclose a longer script with interacting functions, while `Names Default To Here(1)` creates a local scope for an entire script. In `Local()`, you list the variables that are local to the local block. Anything not explicitly listed is not local to that block.

Local Here()

`Local Here()` provides a namespace block with `Names Default to Here(1)`. Use `Local Here()` to prevent name collisions when multiple scripts are executed from the same root namespace (for example, when a script executes two button scripts that have the same variables or an included script is run from the main script).

`Local()` does not always work due to the lifetime of the local block, but `Local Here()` is persistent across the call.

Here's an example that runs an included script from a main script. The scripts define x and y variables as different values. Variables in the included script are in the Local Here namespace.

Main.jsl:

```
Names Default To Here( 1 );

x = 5;
y = 0;

Print( "Main, Before: x=" || Char( x ) || ", y=" || Char( y ) );
Include( "Inc1.jsl" );
Print( "Main, After: x=" || Char( x ) || ", y=" || Char( y ) );
```

Inc1.jsl:

```
Names Default To Here( 1 );

Local Here( // variables are local to this script
  x = 100;
  y = 100;

Print( "Inc1: x=" || Char( x ) || ", y=" || Char( y ) );
);
Print( "Inc1 outside: x=" || Char( x ) || ", y=" || Char( y ) );
```

Return:

```
"Main, Before: x=5, y=0" // defined in Main.jsl, before running Inc1.jsl
"Inc1: x=100, y=100" // defined in Inc1.jsl
"Inc1 outside: x=5, y=0" // from Inc1.jsl, referring to Main.jsl
"Main, After: x=5, y=0" // from Main.jsl, after running Inc1.jsl
```

Notice that the x and y variables in inc1.jsl don't change because they're in the Local Here namespace.

Handling Unqualified Named Variable References

The `Names Default To Here()` function determines how *unqualified* named variable references are resolved. Explicitly scoping a variable using `here:var_name` always works, whether `Names Default To Here()` is on or off. See "Scoped Names" for more information about `here` and other scopes.

Enabling the *Names Default To Here* mode associates a scope called `Here` with an executing script. The `Here` scope contains all of the unqualified named variables that are created when they are the target of an assignment (as an L-value). In JMP 8 and earlier, these variables normally would have been placed in the `Global` scope. Using a `Here` scope keeps variables in multiple executing scripts separate from each other, avoiding name collisions and simplifying the scripting and management of variable name collisions. You can still share information using the `Global` scope.

Names Default To Here and Global Variables

Run this example script one line at a time to see how the `Names Default To Here()` function changes the resolution of variable names.

Example Script

```
a = 1;
Names Default To Here( 1 );
a = 5;
Show( global:a, a, here:a );
   global:a = 1;
   a = 5;
   here:a = 5;
```

1. Run the first line to create a global variable named *a* that holds the value 1.
2. Run the second line to turn on the *Names Default To Here* mode.
3. Run the third line to create a new variable named *a* in the local space that holds the value 5. This line does *not* change the value assigned to the global variable *a*.
4. Run the fourth line to see how scoped and unscoped variables are resolved.

The unqualified *a* is resolved to `here:a`. If `Names Default To Here()` were not on, *a* would be resolved to the global variable named *a*.

Note that if you use `::a` instead of `global:a` in the `Show()` function, your output is a little different:

```
Show(::a, a, here:a);
   a = 1;
   a = 5;
   here:a = 5;
```

Example of Using the Names Default To Here() Function

You have two scripts with the following definitions, and`Names Default To Here()` is turned *off* (the default condition) in both scripts.

Note: Both scripts must be in separate script windows for this example.

```
a = 1; // script 1
Show( a );

a = 3; // script 2
Show( a );
```

1. Run Script 1. Here is the result:

   ```
   a = 1
   ```

2. Run Script 2. Here is the result:

   ```
   a = 3
   ```

3. Run only the `show(a);` line in Script 1. Here is the result:

   ```
   a = 3
   ```

The log shows `a = 3` because variable *a* is global, and was last modified by Script 2. This is the default behavior in JMP 9 and later, and it is the only possible behavior in JMP 8 and earlier.

4. Now turn on `Names Default To Here()` in *both* scripts.

   ```
   Names Default To Here(1);
   ```

   Note: `Names Default To Here()` is local to a particular script. It is *not* a global setting.

5. Run Script 1. Here is the result:

   ```
   a = 1
   ```

6. Run Script 2. Here is the result:

   ```
   a = 3
   ```

7. Run only the `Show( a );` line in Script 1. Here is the result:

   ```
   a = 1
   ```

The log shows `a = 1`, because a copy of variable *a* is maintained for each script.

Note: Problems using this function are generally due to the mixing of unqualified and qualified references to global variables. Always explicitly scoping a name prevents accessing an unintended variable.

Scoped Names

Specify where a name is to be resolved by using a scope in the form `scope:name` where `scope` indicates how to resolve the name. For example, `here:name` indicates that the name should be resolved locally. Using the *Names Default To Here* mode, `here:name` is equivalent to `name`. The scope instructs how to look up the name.

The syntax is to use the colon scope operator:

```
scope:name
```

There are several types of scopes:

- Scope can be a resolution rule. For example, `here:x` means that `x` should be resolved to a name that is local to the script. `Global:x` means that `x` should be resolved to a global name.
- Scope can be a namespace reference variable. For example. `ref:a` means that `a` should be resolved within the namespace that `ref` refers to.
- Scope can be a data table reference to look up names as column names. For example, `dt:height` means that `height` should be resolved as a column in the data table that `dt` references.
- Scope can be the name of a namespace that you created. For example, `myNamespace:b` where `myNamespace` is a namespace that you created. `"myNamespace":b` is equivalent. See "Namespaces".

Note: If you create a function, make sure you scope it to avoid conflicts with built-in functions.

Examples of Scoping Column Formulas

The following examples demonstrate how to scope columns that contain formulas. In both scripts, *x* is a global variable, local variable, and column name.

In the first script, the column name x is unscoped. the formula in the second column multiplies the value in column `x` by 100. In this case, The result is a column with the values 100, 200, and 300.

```
::x = 5;
```

```
New Table( "Test",
   New Column( "x", Values( [1, 2, 3] ) ),
   New Column( "y", Formula( 100 * x ) ),
);
```

In the following script, the formula in column y assigns 500 to x and then adds 50 to x. Each cell in the column contains the value 550.

```
::x = 5;
New Table( "Test",
   New Column( "x", Values( [1, 2, 3] ) ),
   New Column( "y", Formula( Local( {x = 500}, x + 50 ) ) ),
);
```

Predefined Scopes

JMP provides predefined that cannot be removed or replaced. Each of these scopes has specific roles, depending on its associated object.

Table 8.5 Predefined Scopes

Scope	Description
`Global`	Global names are shared throughout the JMP environment.
`Here`	Scope of the executing script.
`Builtin`	JMP built-in functions. For example, `Builtin:Sqrt()`. These names are shared throughout the JMP environment. If you over-ride a JSL function with a custom function, you can still access the built-in JSL function by using this scope.
`Local`	Nearest local scope. Can be nested within the user-defined functions, `Local` and `Parameter`.
`Local Here`	Provides a namespace block inside `Names Default to Here(1)`. Use `Local Here()` to prevent name collisions when multiple scripts are executed from the same root namespace (for example, when a script executes two button scripts that have the same variables). The argument can be any valid JSL expression. `Local( {Default Local}, )` does not always work due to the lifetime of the local block, but `Local Here()` is persistent across the call.
`Window`	Scope of the containing user-defined window. (Rare.)
`Platform`	Scope of the current platform. (Rare.)

Table 8.5 Predefined Scopes *(Continued)*

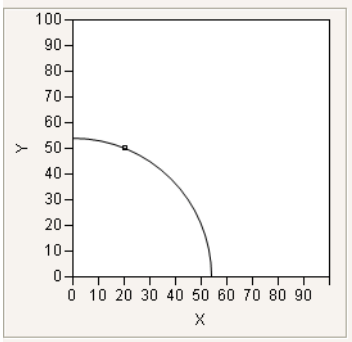
Scope	Description
Box	Scope of the containing context box. A context box is nested within a user-defined window. (Rare.)

Example of Using the Window Scope

This example uses the `Window` scope to pass information during execution. Explicitly scoping the variables `x` and `y` to this window ensures that JMP does not try to scope `x` and `y` in other contexts, such as a data table. The variables `x` and `y` are created and used solely inside the `Window` environment. The `Window` scope is similar to using `Local()`, but more useful because `Local()` is limited in the places that it can be used.

```
New Window( "Example",
   window:gx = 20;
   window:gy = 50;
   Graph Box(
      Frame Size( 200, 200 ),
      Handle(
         window:gx,
         window:gy,
         Function( {x, y},
            window:gx = x;
            window:gy = y;
         )
      );
      Circle( {0, 0}, Sqrt( window:gx * window:gx + window:gy * window:gy ) );
   );
);
```

Figure 8.1 Example of Current Window Namespace



Example of Using the Here Scope

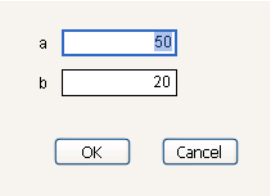
This example uses the `Here` scope to pass information between windows that are created by the same script. Scoping a variable using `Here:` is not dependent on turning `Names Default To Here()` on. The `Here:` scope is always available.

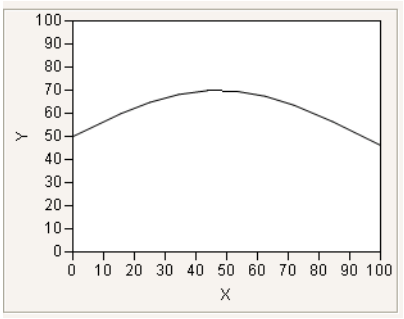
This script produces two windows and uses two different scopes.

The Launcher window asks the user for two values. Those two values are passed to the Output window, which uses them to graph a function. The Launcher window scopes `aBox` and `bBox` to the window: essentially, those two variables (pointers to Number Edit Boxes) exist only in the Launcher window and are not available to the Output window. The values from those two boxes are then copied into variables that are scoped to `Here`, and so are available to both windows that are produced by this script.

```
launchWin = New Window( "Launcher",
   <<Modal,
   V List Box(
      Lineup Box( N Col( 2 ), Spacing( 10 ),
         Text Box( "a" ),
         window:aBox = Number Edit Box( 50 ),
         Text Box( "b" ),
         window:bBox = Number Edit Box( 20 ),

      ),
      Lineup Box( N Col( 2 ), Spacing( 20 ),
         Button Box( "OK",
            // copy values before window goes away
            here:a = window:aBox << Get;
            here:b = window:bBox << Get;
         ),
         Button Box( "Cancel", Throw( 1 ) )
      ),

   )
);

New Window( "Output",
   Graph Box( Y Function( here:a + here:b * Sin( x / 30 ), x ) )
);
```

Figure 8.2 Launcher and Output





Namespaces

A *namespace* is a collection of unique names and corresponding values. You can store references to namespaces in variables. Namespace names are global, because JMP has only one namespace map. Namespace references are variables like any other variable that references an object, so they must be unique within their scope or namespace. The members of a namespace are referenced with the `:` scoping operator, such as `my_namespace:x` to refer to the object that is named `x` within the namespace called `my_namespace`. See "User-Defined Namespace Functions" for more information about creating and managing your own namespaces. Namespaces are especially useful for avoiding name collisions between different scripts.

User-Defined Namespace Functions

Create your own namespaces to hold related sets of variables and function definitions. There are several functions that you can use to manage namespaces.

New Namespace

```
nsref = New Namespace( <"nsname">, <{ name = expr, ... }> );
```

Creates a new namespace called `nsname` (a string expression) and returns a reference to the namespace. All arguments are optional.

`Nsname` is the name of the namespace in the internal global list of namespace names. `Nsname` can be used as the prefix in a scoped variable. The function returns a reference to the namespace, and can also be used as the prefix in a scoped variable reference. If `nsname` is absent, the namespace is anonymous and is given a unique name created by JMP. `Show Namespace()` shows all namespaces and their names, whether assigned or anonymous.

Important: If you already have a namespace named `nsname`, it is replaced. This behavior means that while you are developing your script, you can make your changes and re-run the script without having to clear or delete your namespace. To avoid unintentional replacement, you can either use anonymous namespaces, or test to see whether a particular namespace already exists:

```
If( !Namespace Exists( "nsname" ), New Namespace( "nsname" ) );
```

A list of named expressions is optional. The names are JMP variables that exist only within the namespace.

Note: The named expressions must be a comma-separated list. Separating the expressions with semi-colons causes the list to be ignored.

These namespaces must be uniquely named to prevent collisions in situations where multiple user-defined namespaces are being used. Using anonymous namespace names prevents collisions.

Namespace

```
nsref = Namespace( "nsname" | nsref);
```

Returns a namespace reference. The argument might be either of the following:

- a quoted string that contains a namespace
- a reference to a namespace

Note: `Namespace()` returns a reference to a namespace that already exists. It does not create a new namespace.

Is Namespace

```
b = Is Namespace( nsref );
```

Returns `1` (`true`) if *`nsref`* is a namespace or `0` (`false`) otherwise.

As Scoped

```
b = As Scoped( "nsname", var_name);
nsname:var_name;
```

`As Scoped()` is the function form of a scoped reference. The function returns a reference to the specified variable in the specified scope.

Namespace Exists

```
b = Namespace Exists( "nsname" );
```

Returns `1` (`true`) if *`nsname`* exists in the list of global namespaces, or `0` (`false`) otherwise.

Show Namespaces

```
Show Namespaces();
```

Shows the contents of all namespaces contained in the list of global namespaces. Namespaces are not visible unless a reference is made to one, using either the `New Namespace` or `Namespace` functions.

Namespace Messages

In addition to the namespace management functions, a namespace also supports a set of messages to access and manipulate its contents.

Note that these messages, as with all message, must be sent to a scriptable object. A namespace name is not a defined scriptable object and cannot be used in a Send operation. However, you can use the name of a namespace in variable references. For example, `nsmane::var` is equivalent to `nsref::var`.

See the *JSL Syntax Reference* for descriptions of the messages that are supported by user-defined namespace references.

Using Namespace References

The following are all equivalent references to a variable that is named `b` in the namespace that is named `nsname` that has a reference `nsref`:

```
nsref:b
nsname:b
"nsname":b
nsref["b"]
nsref<<Get Value("b") // used as an r-value
```

Namespaces and Included Scripts

An included script runs in the namespace of the parent script. If the included script has its own namespace definitions, you need to do one of the following:

- manage the namespace names to avoid name collisions
- use anonymous names created by the `New Namespace` function

In either case, you still need to manage variable references to namespaces.

There is also an option for the `Include` function (`New Context`) that creates a namespace that the included script runs in. This namespace is anonymous and it is independent from the parent script's namespace. For example:

```
Include("file.jsl", <<New Context);
```

This anonymous namespace can be referenced using `Here`.

If both the parent and included scripts use the global namespace, add `Names Default To Here` to the `New Context` option to avoid name collisions. For example:

```
Include("file.jsl", <<New Context, <<Names Default To Here);
```

See "Includes" for more information about the `Include` function.

Examples of User-Defined Namespaces

Creating and Using a Basic Namespace with Expressions

This example shows creating an anonymous namespace and using functions and variables within it.

```
new_emp = New Namespace(
   {name_string = "Hello, *NAME*!",

   print_greeting = Function( {a},
      Print( Substitute( new_emp:name_string, "*NAME*", Char( a ) ) )
   )}
);
```

Note that you must use the fully qualified name for variables defined within the namespace.

```
new_emp:print_greeting( 6 );
   "Hello, 6!"
```

Complex Number Operations

This example creates a namespace that contains functions to support using two-element lists to represent complex numbers, and then locks the namespace.

```
If( !Namespace Exists( "complex" ),
   New Namespace(
      "complex"
   );
   complex:makec = Function( {a, b},
      Eval List( {a, b} )
   );
   complex:addc = Function( {a, b}, a + b );
   complex:subtractc = Function( {a, b}, a - b );
   complex:multiplyc = Function( {a, b},
```

```
      Eval List( {a[1] :* b[1] - a[2] :* b[2], a[1] :* b[2] + a[2] :* b[1]} )
   );
   complex:dividec = Function( {a, b},
      d = b[1] ^ 2 + b[2] ^ 2;
      Eval List(
         {a[1] :* b[1] - a[2] :* b[2] / d, a[2] :* b[1] - a[1] :* b[2] / d}
      );
   );
   complex:charc = Function( {a},
      Char( a[1] ) || "+" || Char( a[2] ) || "i"
   );
);
Namespace( "complex" ) << Lock;
```

Here are examples using functions that are within the above user-defined namespace.

```
c1 = complex:makec( 3, 4 );
c2 = complex:makec( 5, 6 );
cadd = complex:addc( c1, c2 ); // returns {8, 10}
csum = complex:subtractc( c1, c2 ); // returns {-2, -2}
cmul = complex:multiplyc( c1, c2 ); // returns {-9, 38}
cdiv = complex:dividec( c1, c2 ); // returns {14.6065573770492,
  19.7049180327869}
Show( complex:charc( c1 ) ); // returns complex:char(c1) = "3+4i";
cm1 = complex:makec( [1, 2, 3], [4, 5, 6] ); // returns {[1, 2, 3], [4, 5, 6]}
```

Referencing Namespaces and Scopes

There are a number of factors in resolving a named variable reference. Table 8.6 describes the named variable references that are resolved for specific situations.

Notes:

- In JMP 9 and later, `a`, `:a`, and `::a` have the same meaning with the *Names Default To Here* mode turned off.
- If the current point of execution is in a user-defined function, or a Local or Parameter JSL function body, then the Local namespace is used for creation when `Default Local` is specified. However, the reference rule is unchanged.

Table 8.6 Namespace References

Form	Reference Type	Reference Rule	Creation Rule
`a`	Unqualified	If the *Names Default To Here* mode is on, JMP looks for the variable in these locations: • `Local` namespace • `Here` namespace • current data table If the *Names Default To Here* mode is off, JMP looks for the variable in these locations: • `Local` namespace • `Here` namespace • `Global` namespace • current data table	• If the *Names Default To Here* mode is on, then JMP creates the variable in the `Local` namespace or in the `Here` namespace. • If the *Names Default To Here* mode is off, then JMP creates the variable in the `Local` namespace or in the `Global` namespace.
`:a`	Current data table	JMP looks for the variable in the current data table.	(Not applicable)
`::a` `Global:a`	Global	JMP looks for the variable in the `Global` namespace.	JMP creates the variable in the `Global` namespace.

`ns:a`
`dt:a`
`Here:a`
`"name":a`
`expr:a`
	Qualified	JMP looks for the variable in the specified namespace. If the variable is not found, an error results.	JMP creates the variable in the specified namespace. Any previous values are replaced.

`ns["a"]` `ns[expr]`	Subscript	JMP looks for the variable in the specified namespace. If the variable is not found, an error results.	JMP creates the variable in the specified namespace. Any previous values are replaced.
`Platform:a`	Qualified	JMP looks for the variable in the encapsulating platform.	JMP creates the variable in the encapsulating platform.

Table 8.6 Namespace References *(Continued)*

Form	Reference Type	Reference Rule	Creation Rule
`Local:a`	Qualified	JMP looks for the variable within any nested local function, up to and including a function call boundary. See "Example of Local:a".	JMP creates the variable in the innermost nested local function or function call boundary.
`Window:a`	Qualified	JMP looks for the variable in the encapsulating `New Window` window namespace.	JMP creates the variable in the encapsulating `New Window` window namespace.
`Box:a`	Qualified	JMP looks for the variable in the encapsulating `Context Box` namespace contained in a `New Window` window.	JMP creates the variable in the encapsulating `Context Box` namespace contained in a `New Window` window.

Example of Local:a

Sample Script	Log Output

`Delete Symbols();`
`Local( {d1l1 = 12},`
`  local:f1f1 = Function( {fa1, fa2},`
`  {fl1 = 99},`
`  local:fa12 = fa1 + fa2;`
`  Local( {d2l1 = 56},`
`  local:l2l2 = 78;`
`  Show( fa12 );`
`  Show( fl1 );`
`  Try( Show( d1l1 ), Write(`
`"\!n\!n***Error=" || Char( exception_msg`
`) || "\!n" ) );`
`  Show Symbols();`
`  );`
`  local:fa12;`
`  );`
`  f1f1( 2, 3 );`
`);`
	fa12 = 5;
	fl1 = 99;
	dil1 = 12;
	
	// Local
	
	d2l1 = 56;
	l2l2 = 78;
	
	// 2 Local
	
	// Local
	
	fa1 = 2;
	fa12 = 5;
	fa2 = 3;
	fl1 = 99;
	
	// 4 Local
	
	// Local
	
	d1l1 = 12;
	
	// 1 Local
	
	5

Resolving Named Variable References

When variables are referenced within a JMP script, JMP resolves the storage location of the variable using a specific set of rules. If the variable is referenced by a qualified name, then the resolution is based on the specific qualification specification. If the variable is referenced by an unqualified name, the situation is a bit more complex. JMP looks through a hierarchy of scopes representing the point of execution with the executing script. This section describes the rules that are used to resolve named variable references.

By default, variable name resolution in JMP 9 and later worked the same way as in JMP 8 and earlier, allowing your current JSL scripts to be executed unchanged. For JMP 9 and later, the difference between qualified and unqualified variable named references is important to understand.

Qualified Named References

A qualified named reference uses the `:` and `::` operators to provide specific information about where a referenced variable resides, or where it is created. Examples of qualified named references include the following:

```
:var
::globalvar
datatable:var
nsref:var
"nsname":var
```

Unqualified Named References

An unqualified named reference provides no explicit information to completely identify where a variable resides or where it is created. No scoping operator (`:` or `::`) is specified in the reference. To change the behavior of JMP when resolving unqualified named variable references, use the `Names Default To Here(1)` function. For more information about variable name resolution, see the "Rules for Name Resolution".

Rules for Resolving Variable References

JMP applies rules in sequence to resolve variable references. See "Rules for Resolving Names".

Best Practices for Advanced Scripting

Minimize Polluting the Global Namespace and Prevent Scripts from Interacting

Always start your script with this line:

```
Names Default To Here(1);
```

Share Variables Across Scripts

Use named namespaces. Namespace names are placed in the global scope.

Use Anonymous Namespaces

Using namespace references to anonymous namespaces avoids possible conflicts with other namespaces.

Advanced Programming Concepts

This section covers some more advanced programming techniques that can be useful for developing scripts.

- "Throw and Catch Exceptions"
- "Functions"
- "Recursion"
- "Includes"
- "Load and Save Text Files"
- "Run External Programs"
- "Classes"

Throw and Catch Exceptions

A script can stop itself by executing the `Throw()` function. If you want to escape from part of a script when it is in an error condition, you can enclose it in a `Try()` expression.

If you include a character-valued expression as a `Throw()` argument, one of the following occurs:

- Throwing stores the string in a global variable named `exception_msg`. This is illustrated in the first example below.
- Throwing stores the string in a local variable named `exception_msg` if you define `exception_msg` in the Here namespace using `Names Default to Here(1)`.

`Try()` takes two expression arguments. It starts by evaluating the first expression, and if or when the first expression throws an exception by evaluating `Throw`, it does the following:

1. Immediately stops evaluating that first expression.
2. Returns nothing
3. Evaluates the second expression.

Examples

You can use `Try()` and `Throw()` to catch an exception that JMP itself throws.

```
Try( Throw( "Hello." ), Show( exception_msg ) );
  exception_msg = "Hello.";
```

The following example prints a message to the log when the `Try()` expression cannot be executed:

```
Try(
```

```
    dt = Open( "Mydata.jmp" ); // a file that cannot be opened
    Summarize( a = by( age ), c = count, meanHt = Mean( Height ) );
    Show( a, c, meanHt );
,
    Print( "This script does not work without the data set." );
    Throw();
);
```

If the Throw() string begins with "!" and is inside a Try() expression, throwing creates an error message about where the exception was caught.

```
Try(
    Throw( "!This message has an exclamation point." );
,
    Write( "\!Nexception_msg: ", exception_msg );
    Print( "Hello from the catch block (WITH exclamation)." );
);
Print( "This AFTER message will NOT print due to error." );
```

When writing the Try() error handler code, consider wrapping the exception variable exception_msg in a Char() function. If your script calls Concat() without wrapping it in a Char() call, you might get an error depending on what was actually returned in this variable. Though some commands return a string in exception_msg, other commands (such as Open()) return a list or another data structure. This would cause an error in the error handler if it were passed directly as an argument to Concat().

```
Try(
    Open( "no file" ),
    If( Is Empty( exception_msg ),
        // use Write to print "no error"if exception_msg is empty
        Write( "\!Nexception_msg (0): no error" ),
        // use Print to print the exception message
        Print( "\!Nexception_msg (1): ", exception_msg );
        /* or first make sure that exception message is a string
        and then use Concat to build the message that is passed to Print(). */
        Print( "\!Nexception_msg (2): " || Char( exception_msg ) );
    )
);
    exception_msg (1): "
    {"Cannot open table"(1, 2, "Open", Open /*###*/("no file"))}
    "
    exception_msg (2): {\!"Cannot open table\!"(1, 2, \!"Open\!", Open
        /*###*/(\!"no file\!"))}"
```

You can also use Try() and Throw() to escape from deep inside For loops.

```
a = [1 2 3, 4 5 ., 7 8 9];
b = a;
nr = N Row( a );
```

```
nc = N Col( a );
// a[2, 3] = 2; // uncomment this line to see the "Missing b" outcome

Try(
   sum = 0;
   For( i = 1, i <= nr, i++,
      For( j = 1, j <= nc, j++,
         za = a[i, j];
         If( Is Missing( za ),
            Throw( "Missing a" )
         );
         zb = b[j, i];
         If( Is Missing( zb ),
            Throw( "Missing b" )
         );
         sum += za * zb;
      )
   );
,
   Show( i, j, exception_msg );
   Throw();
);
   i = 2;
   j = 3;
   exception_msg = "Missing a";
```

You do not have to use `Try()` to make use of `Throw()`. In this example, `Throw()` is not caught by `Try()` but still stops a script that cannot proceed:

```
dt = New Table(); // to get an empty data table
If( N Row( dt ) == 0,
   Throw( "!Empty Data Table" )
);
```

Functions

JSL also has a function called `Function` to extend the macro concept with a local context arguments. Suppose that you want to create a function that takes the square root but tolerates negative arguments, returning zero rather than errors. You first specify the local arguments in a list with braces { } and then state the expression directly. You do not need to enclose the expression in `Expr` because `Function` stores it as an expression implicitly.

```
myRoot = Function( {x},
   If( x > 0, Sqrt( x ), 0 )
);
a = myRoot( 4 );  // returns 2
b = myRoot( -1 ); // returns 0
```

Functions are stored in variables, the same as values. This means that you cannot have both a root function and a root value. It also means that you can redefine a function anytime except when you are inside the function itself.

When a function is called, its arguments are evaluated and given to the local variables specified in the list forming the first argument. Then the body of the function, the second argument, is evaluated.

The values of the arguments are for the temporary use of the function. When the function is exited, the values are discarded. The only value returned is the return value. If you want to return several values, then return a list instead of a single value.

In defined functions, the stored function is not accessible directly, even by the `Name Expr` command. If you need to access the function expression in your script, you have to create the function within an `expr()` clause. For example,

```
makeFunction = Expr(
   myRoot = Function( {x},
      If( x > 0, Sqrt( x ), 0 )
   )
);
d = Substitute( Name Expr( MakeFunction ), Expr( x ), Expr( y ) );
Show( d );
makeFunction;
```

Optional Arguments

You can create optional arguments. For example:

```
f1 = Function( {x, y, z}); // all arguments are required
f2 = Function( {x, y=2, z=4}); // x is required, y and z are optional
```

Notes:

- To make an argument optional, give it a default value in the function's definition (x=1 rather than just x).
- Optional arguments always must follow a required argument. For example, this is not allowed:

  ```
  Function( {x = 1, y}, ... )
  ```

  It will work when defined, but you will always have to pass in a value for x and y, because the default value will never be used.
- When you call the function, you cannot skip optional arguments. In this example, you have to supply a value for y to supply a value for z when you call the function:

  ```
  ex = Function( {x, y = 2, z = 3},
    Return( x + y + z )
  );
  ex( 1, 4 ); // passes in 1 for x and 4 for y; z will be 3. returns 8.
  ```

Local Symbols

You can declare variables as local to a function so that they do not affect the global symbol space. This is particularly useful for recursive functions, which need to keep separate the values of the local variables at each level of function call evaluation.

As shown above, a function definition looks like this:

```
functionName=Function({arg1, ...}, body);
```

You can also have the function definition default all the unscoped names to be local.

```
functionName=Function({arg1, ...}, {Default Local}, body);
```

For the purposes of creation, the use of `Default Local` localizes all the names that meet the following qualifications:

- Are not scoped as globals (for example, `::name`)
- Are not scoped as data table column names (for example, `:name`)
- Occur without parentheses after them (for example, are not of the form `name(...)`)

For example, the following function sums three numbers.

```
add3 = Function( {a, b, c},
   {temp},
   temp = a + b;
   temp + c;
);
X = add3( 1, 5, 9 );
   15
```

The following function does the same thing, automatically finding locals.

```
add3 = Function( {a, b, c},
   {Default Local},
   temp = a + b;
   temp + c;
);
X = add3( 1, 5, 9 );
   15
```

In both cases, the variable `temp` is not a global, or, if it is already a global, remains untouched by evaluating the functions.

Using `Default Local` in user-defined functions can cause some confusion because it is context-sensitive. That is, the same function may behave differently in different contexts (depending on whether same-named outer variables are in scope) because the reference rules for unqualified names are unchanged when using `Default Local`. To make a function less context-sensitive, enumerate each and every variable that you want to be local. This reduces the confusion and the potential incorrect use of outer scope variable values.

Recursion

The `Recurse()` function makes a recursive call of the defining function. For example, you can make a function to calculate factorials. A factorial is the product of a number, the number minus 1, the number minus 2, and so on, down to 1.

```
myfactorial = Function( {a},
   If( a == 1,
      1,
      a * Recurse( a - 1 )
   )
);
myfactorial( 5 );
   120
```

You can define recursive calculations without using `Recurse()`. For example, you could replace `Recurse()` by `myfactorial`, and the script would still work. However, `Recurse()` offers these advantages:

- It avoids name conflicts when a local variable has the same name as the function.
- You can recurse even if the function itself has not been named (for example, assigned to a global variable, such as *myfactorial* above).

Includes

The `Include()` function opens a script file, parses the script in it, and executes the JSL in the specified file.

```
Include( "pathname" );
```

For example,

```
Include( "$SAMPLE_SCRIPTS/myStartupScript.jsl" );
```

There is an option to obtain the parsed expression from the file, rather than evaluating the file.

```
Include( "pathname", <<Parse Only );
```

Another named option creates a namespace that the included script runs in. This namespace is an anonymous namespace and it is independent from the parent script's namespace.

```
Include( "file.jsl", <<New Context );
```

See "Advanced Scoping and Namespaces" for information about using namespaces with your scripts.

Note the following about included files:

- JMP files aside from JSL cannot be used.

- Other recognized file types, such as image files, SAS data sets, and Microsoft Excel files cannot be used.
- Unrecognized file types are treated as a JSL file.
- Files with the .txt extension are treated as a JSL file. A text file that contains data can be included, however an error will appear since this is not valid JSL.

Load and Save Text Files

The `Load Text File()` and `Save Text File()` commands allow manipulation of text files from JSL. Note that the paths in the following code are strings.

```
text = Load Text File( "path" );
Save Text File( "path", text );
```

You can load a text file from a Web site:

```
Load Text File( "URL", <blob> );
```

The URL is a quoted string that contains the URL for the text file. The text file is returned as a string. If you add the optional named argument *blob*, a blob is returned instead.

Run External Programs

`Run Program()` enables you to control a program that uses standard input and standard output. Line-mode programs read commands from stdin and write answers to stdout. cmd.exe, the Windows command prompt, is one example. You can also use it to launch GUI programs, such as Windows Notepad.

```
Run Program(
   Executable( "cmd.exe" ),
   Options( {"/c", "TaskList\!n"} ),
   Read Function( "text" ) // returns all text
);
```

The preceding script writes the Windows task list and details to the log.

Figure 8.3 Log Output

```
"
Image Name                     PID Session Name        Session#    Mem Usage
========================= ======== ================ =========== ============
System Idle Process              0 Services                   0         24 K
System                           4 Services                   0     64,340 K
smss.exe                       516 Services                   0      1,568 K
```

You can also print the task list directly without using cmd.exe and then import the task list into a data table.

```
blob = Run Program(
   Executable( "tasklist.exe" ),
   Read Function( "blob" )
);

Open(
   blob,
   Columns(
      Column( "Image Name", Character, "Nominal" ),
      Column(
         "PID",
         Numeric,
         "Continuous",
         Format( "Best", 10 )
      ),
      Column( "Session Name", Character, "Nominal" ),
      Column(
         "Session#",
         Numeric,
         "Continuous",
         Format( "Best", 10 )
      ),
      Column( "Mem Usage", Character, "Nominal" ),
      Omitted Column( . )
   ),
   Import Settings(
      Fixed Column Widths( 26, 9, 17, 12, 12, 63 ),
      Labels( 1 ),
      Column Names Start( 2 ),
      Data Starts( 4 )
   )
);
```

Figure 8.4 Imported Data from Run Program()

	Image Name	PID	Session Name	Session#	Mem Usage
1	System Idle Proc...	0	Services	0	8 K
2	System	4	Services	0	42,560 K
3	smss.exe	456	Services	0	1,228 K

See the *JSL Syntax Reference*.

Note: If you use `Run Program()` to launch a GUI program, JMP is inoperable until you close the program.

Classes

Classes enable you to create objects based on templates. You can use JSL to create a definition of a class. A class definition is a template, which contains definitions for member variables, methods, and functions. After you have defined a class in JSL, you can instantiate an object that has all of the characteristics defined in the class definition.

Use the JSL function `Define Class()` to create a JSL class definition. Use the JSL function `New Object()` to instantiate an instance of a JSL class.

The JSL function `Define Class()` has a number of arguments that define the contents of the class:

```
Define Class( "class name",
   <Base Class( "base class name", <"base class name", ...> ),>
   <Show( All( Boolean ) ) | Show( <Members( Boolean ),> <Methods( Boolean ),>
   <Functions( Boolean )> ),>
   <Assignment Statements>
);
```

Where:

- `"class name"` is the name of the class being defined.
- `"base class name"` is one or more class names that make the set of base classes that are used as the foundation of the class being created.
- `Show` defines the value that appears when a class object is written out using the `Show()`, `Print()`, or `Write()` functions. Options include:
  - `All( Boolean )` shows all members, methods, and functions defined in a class object instance. The default value for All is true. If you specify the `All` option, that setting overrides any other arguments in the `Show` option.
  - `Members( Boolean )` shows all member variables contained in a class object instance.
  - `Methods( Boolean )` shows all methods contained in a class object instance.
  - `Functions( Boolean )` shows all functions contained in a class object instance.

  Note: You can specify either the `All` option or any number of the `Members`, `Methods`, and `Functions` options, but you cannot specify both options at the same time.

  Note: If a class contains a `_show_` method, that method overrides the `Show` option.

- `Assignment Statements` are zero or more of the following assignment statements, separated by semicolons:
  - `Member = JSL Expression`
  - `Method Name = Method( {<argument, ...>}, body )`

- `Function Name = Function( {<argument, ...}, {<local variable, ...>}, body )`

Note: The Assignment Statements can alternatively be specified as a list of JSL expressions.

There are three specialized methods that you can use in a class definition:

init Initializes the object at the time it is instantiated. This is considered the constructor for the class. You can have only one constructor per class.

show Defines the output of the `Show()`, `Print()`, and `Write()` JSL functions when an instance of the class is used as the argument.

to string Defines the output of the `Char()` JSL function when an instance of the class is used as the argument.

Notes:

- All methods, member variables, and functions are public.
- Functions defined within a class cannot access method variables or methods within the class.

Example of Defining a Class

The following example defines a simple class for a pet, containing the following information: name, breed, and color.

In this example, you define the constructor to require all information to create a new Pet object. Also, you define a set of accessor methods to obtain each of the attributes of a Pet object.

```
Define Class( "Pet",
   name = ""; breed = ""; color = "";
   _init_ = Method( { name, breed, color },
      this:name = name;
      this:color = color;
      this:breed = breed;
   );
   _show_ = Method( {},
      Return( name || " is a " || color || " " || breed || "." )
   );
   Get Pet Name = Method( {}, Return( name ) )
);
```

Instantiating a Class

To instantiate an instance of a JSL class object, use `New Object()` JSL function with any of the following functions and arguments:

```
New Object( "Class Name"( constructor arguments ) );
New Object( Class Name( constructor arguments ) );
New Object( Class Reference( constructor arguments ) );
New Object( "Class Name", { constructor arguments } );
New Object( Class Name, { constructor arguments } );
New Object( Class Reference, { constructor arguments } );
```

Examples of Instantiating and Using a Class

The following example shows how to instantiate a class to create and access JSL class objects.

```
class = "Pet";
xa = New Object( Pet( "Oliver", "Schnauzer", "Black" ) );
xb = New Object( class, { "Flynn", "Cockapoo", "Brown" } );
xc = New Object( class( "Mandi", "King Charles Cavalier", "Ruby" ) );
xd = New Object( "Pet"( "Charlie", "Asian", "White" ) );
xd1 = New Object( xd( "Fred", "Siamese", "Silver" ) );
xd2 = New Object( xd, { "Sam", "German Shepard", "Black and White" } );

Show( xa );
Show( xb );
Show( xc );
Show( xd );
Show( xd1 );
Show( xd2 );
xd2 = xd;                  // returns a second reference to xd
xdclone = xd << Clone;     // copy xd to a new instance, named xdclone
xd2:name = "Jerry";
Show( xd:Get Pet Name(), xd2:Get Pet Name(), xdclone:Get Pet Name() );
```

This example creates multiple instances of the Pet class. The output of the `Show()` JSL function is determined by the `_show_` method that is defined in the Pet class definition. This example also illustrates the difference between creating a second JSL reference to a class instance and cloning a class instance. When using the `Clone` message, you are creating a new class instance that is independent of the original. However, the new instance starts out as a copy of the original.

File and Directory Operations

`Pick Directory()` and `Pick File()` enable the user to select a file or directory when the script is run. To get a list of files in a directory, use `Files In Directory()`.

- "Select a Directory or File"
- "Get a List of Filenames"
- "Connect to a Network Directory"

Select a Directory or File

You can prompt the user to select a directory using the `Pick Directory()` function. The command displays a platform-specific window in which the user selects a folder. On Windows, the optional prompt string appears at the top of the Browse window (Windows) or below "Open" on the Finder window (macOS). You can also specify the path for an initial directory that appears, and specify whether files should appear in the Pick Directory window.

```
Pick Directory ( "Select a directory." );
```

To let the user select a file, use the `Pick File()` function:

```
Pick File(
  <"prompt message">, <"initial directory"> <{filter list}>,
  <first filter>, <save flag>, <"default file">,
  <multiple>)
```

The `"prompt message"` is used as the window title. The `"initial directory"` defines which folder initially appears. If a directory is defined as an empty string, the default directory is used.

You can also define the `{filter list}` used for the `Open()` window, forcing it to show only certain file types. This list must use the following syntax:

```
{"Label1|suffix1;suffix2;suffix3", "Label2|suffix4;suffix5"}
```

Each quoted string adds an entry to the **File name** list in the `Open()` window. `Label` defines the text that is displayed for each menu option. The following list of suffixes defines the file types that are displayed if its corresponding label is selected. Note the use of `"*"` to list all files in the window.

```
Pick File(
  "Select JMP File", // prompt message
  "$SAMPLE_DATA", // initial directory
  {"JMP Files|jmp;jsl;jrn", "All Files|*"}, // file filter list
  1, // intially selected item
  0, // doesn't prompt the user to save the file
```

```
    "Analgesics.jmp" // file that is selected by default
);
```

Tip: All arguments are optional, however, they are also positional. This means that you can leave out arguments only at the end of the script. Use empty strings for the arguments that you want to omit from the beginning of the script.

The script below does not set the default directory or the default file:

```
Pick File(
   "Select JMP File",
   "", // no default directory
   {"JMP Files|jmp;jsl;jrn", "All Files|*"},
   1,
   0,
   "" // no default file
);
```

The <first filter> argument sets the default selection where n is the index for the list item. In the script above, the <first filter> is the first item in the list: "JMP Files|jmp;jsl;jrn".

If <Save Flag> is false, the Multiple argument can be added to allow the user to select multiple files using the one window:

```
Pick File(
   "Select JMP File",
   "",
   {"JMP Files|jmp;jsl;jrn", "All Files|*"},
   1,
   0,
   "",
   multiple // save flag is 0, allows multiple file selection
);
```

Get a List of Filenames

To obtain a list of filenames in a specific directory, use the Files In Directory command.

```
Files In Directory( path, <recursive( 0 | 1 )> );
```

Both filenames and subdirectory names are returned as shown in the following example:

```
Files In Directory( "$SAMPLE_DATA" );
   { "2D Gaussian Process Example.jmp", "Abrasion.jmp", ... "Design
     Experiment", "Detergent.jmp", ... }
```

Notice that the files within the Design Experiment subdirectory are not included. And only files in the root $SAMPLE_DATA directory are listed.

To return a list of all file names, add the optional `recursive(Boolean)` argument to `Files In Directory`:

```
Files In Directory( "$SAMPLE_DATA", recursive( 1 ) );
   { "2D Gaussian Process Example.jmp", "Abrasion.jmp", ... "Design
     Experiment/2x3x4 Factorial.jmp", "Design Experiment/Algorithm Data.jmp",
     ... }
```

To get the full pathnames, recurse the directories and concatenate the file paths and file names. The following example loops through each file in the $SAMPLE_DATA directory and subdirectories. The file path is concatenated to each file name.

```
names = Files In Directory( "$SAMPLE_DATA", recursive( 1 ) );
For( i = 1, i <= N Items( names ), i++,
   names[i] = Convert File Path( "$SAMPLE_DATA" ) || names[i]
);
names;
   {"/C:/Program Files/SAS/JMP/16/Samples/Data/2D Gaussian Process
     Example.jmp",
   "/C:/Program Files/SAS/JMP/16/Samples/Data/Abrasion.jmp", ... }
```

The `Files in Directory` command accepts native and POSIX paths, as well as paths using path variables. See "Path Variables" for more information about working with paths.

Connect to a Network Directory

On Windows, you can connect to a network directory without connecting to the shared drive first.

```
Files in Directory( "\\myserver.company.com/source/Users/Smith" );
```

On macOS, you have to mount the volume first and then connect to the directory using a combination of JSL and AppleScript. Note that you can also modify this script to work on Windows. On macOS only, the user is prompted for a password.

```
mount = Function( {server},
   {Default Local},

   If( Host is( "Mac" ),

      /* Use osascript to mount the volume.
      The user will be prompted for credentials. */
      Run Program(
         Executable( "/usr/bin/osascript" ),
         Options( {"-e", Eval Insert( "mount volume \!"smb://^server^\!"" )} ),
         Read Function( "text" )
      );
      server = Concat Items( Remove( Words( server, "/" ), 1 ), "/" );
      Eval Insert( "/Volumes/^server^" );
```

```
    ,
       // Windows can use UNC to auto-mount the server.
       Eval Insert( "\!\\!\^server^" )
    )
);


users = Mount( "myserver.company.com/Users" );
Files In Directory( Eval Insert( "^users^/Smith" ) );
```

Scripting By Groups

By group arguments are supported for these functions: ColMean(), ColStdDev(), ColNumber(), ColNMissing(), ColMinimum(), ColMaximum().

Any number of By arguments can be specified, and you can use expressions for the By arguments. By arguments must be used in a column formula, or in the context of ForEachRow(). The first argument can also be a general numeric expression.

Here is an example:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Mean of height by sex", Numeric, Formula( Col Mean(
   :height, :sex ) ) );

dt << New Column( "Minimum of height by sex and age",
   Numeric,
   Formula( Col Minimum( :height, :sex, :age ) )
);

dt << Distribution( Continuous Distribution( Column( :height ) ), By( :sex )
   );

dt << Tabulate(
   Show Control Panel( 0 ),
   Add Table(
      Column Table(
         Analysis Columns( :height ),
         Statistics( Mean, N, Std Dev, Min, Max, N Missing )
      ),
      Row Table( Grouping Columns( :age, :sex ) )
   )
);
```

Encrypt and Decrypt Scripts

To add a basic level of protection to scripts, you can encrypt it so only someone who knows the password can view it; you can also require a password to run it. This is useful in situations when you want to implement controlled sharing of a script.

To encrypt a script:

1. Open the script that you want to encrypt.
2. Select **Edit > Encrypt Script**.
3. Enter a decrypt password so that the user needs a password to view the script.
4. (Optional) Enter a run password to require the user to enter a password before running the encrypted script.

   Note: The passwords must consist of single-byte characters; using a text Input Method Editor (IME) does not work.

5. Click **OK**.
6. If you entered only a decrypt password, click **Yes** to confirm that you do not want to assign a run password.

   The encrypted script opens in a new window. For example:

```
//-e6.0.2
S@FTQ;VGMUTF?J<;LS;B<=IRLXCU=BV;@NS<TW;LR<ZFOP=JJS>NNDA@T<V><DZA>SU@MG;LR<ZFO
   P=JJS>NNDA@T<V><DZA>SU@MG;LR<ZFOP=JJS>NNDA@T<V><DZA>SU@MG;LR<ZFOP=JJS>NNDA
   @T<HNIZ;WDN?RMJ;FR>KYAXTEPPF?;XFJJOP=RQGBIAGXOYNNZ>PLIF>SW>L>ACL<KGP;=QQTC
   EG??U<PUXLV?TRBO?J>QGWTJCFJA@BNHWLVORNNGQYPIKL<IM<>JX>@G?LJ>=;RBODH@PTKK@S
   IUE;IJOR<TUTRMTGSYRSVGOR<XK<F=IWQYE=LVZFP;AUHA?YJLL;EIT?ZJZC;*
```

7. Save the encrypted script.

To encrypt a column formula in a script:

Column formulas created by an encrypted script are visible by default. If you want a formula to be encrypted, follow these steps:

1. Paste just the formula into a new script window.
2. Encrypt the script by clicking **Edit > Encrypt Script**.
3. Provide a password only.
4. Copy the encrypted script.
5. Paste the encrypted script in a new script as a string argument for `JSL Encrypted()` as shown below.

```
dt = Open( "$SAMPLE_DATA\Big Class.jmp" );

dt << New Column( "Encrypted",
   Numeric,
   Continuous,
   Format( "Best", 12 ),
   Formula(
      JSL Encrypted(
         "//-e6.0.2
0,
136,4112eJwFwYkRhCAMAMB+roUQcITwDcoP4fovxN3RiOB+MFqvqXB8ZOhZdFC+dgqclJo0dwrlZ
   bFjgyPlZRuR2hp399ObhpxhQXP8
GPtftOMN6Rwz8pDnjmK1KOqq1bEG+HOxvyBb
,0,"
      )
   )
);
```

To decrypt a JSL script:

1. Open the encrypted script in JMP.
2. Select **Edit > Decrypt Script**.
3. Enter the decrypt password and click **OK**.

   The decrypted script opens in a new window.

To run an encrypted JSL script:

Note: You must know which data table the script runs on before running an encrypted script. If you do not know the name of the data table, you must decrypt the script before running it.

1. Open the encrypted script in JMP.
2. Select **Edit > Run Script**.
3. Enter the run password and click **OK**.

   The script runs:

   – If the script references a data table, you are prompted to open the data table, and then the script runs.

   – If the script requires an empty data table, you must create the table and then run the encrypted script.

Note that entering the run password runs the script, but does not show the script: you must supply the decrypt password to actually view the script.

Encryption and Global Variables

Encryption alone does not hide global variables and their values. A `Show Globals()` command displays them normally. If you want to hide global variables in an encrypted script, you can give them special names.

Any global variable whose name begins with two underscore characters (__) is hidden, and `Show Globals()` displays neither its name nor its value. For example:

```
myvar = 2;
__myvar = 5;
Show Symbols();
   // Globals
   myvar = 2;
   // 2 Global (1 Hidden)
```

This strategy works whether your script is encrypted or not.

Encrypting Scripts in Data Tables

You can also encrypt a script that is saved to a data table using the `JSL Encrypted()` or `Include( Char to Blob() )` functions.

- `JSL Encrypted()` is more straightforward, because it involves one function. You can include comments inside the encrypted script.
- `Include( Char to Blob() )` lets you include comments, but not inside the script.

Note: Column formulas in encrypted scripts are not encrypted. To encrypt them, include them in the `JSL Encrypted()` function.

Follow these steps to encrypt a data table script:

1. Place the script in a script window.

   You cannot directly encrypt a script that is already saved to a data table.
2. In the script window, select **Edit > Encrypt**.
3. Enter a decrypt password.
4. (Optional) Enter a run password to require the user to enter a password before running the script.
5. If you entered only a decrypt password, click **Yes** to confirm that you do not want to assign a run password.

   The encrypted script opens in a new script window.
6. Copy the entire encrypted script.
7. Create a new data table script or open an existing script.

8. In the script portion of the window, type one of the following functions:

   ```
   JSL Encrypted( "" );
   ```

   ```
   Include( Char to Blob( "" ) );
   ```

9. Paste the encrypted script inside the quotation marks in the function.
10. Click **OK**.

Figure 8.5 Example of an Encrypted Data Table Script

```
JSL Encrypted(
    "//-e6.0.2
WE@GSACGT<?CKEG=NG;B<=IRLXCU=BV;@NS<TW;LR<ZFOP=JJS>NNDA@T<V><DZA
>SU@MG;LR<ZFOP=JJS>NNDA@T<V><DZA>SU@MG;LR<ZFOP=JJS>NNDA@T<V><DZA
>SU@MG;LR<ZFOP=JJS>NNDA@T<FFHG=R;GYNBNME>ZBMIB?O;G<>?;UTG@=HVYCC
@MFYCMVQEOTNWASNOEV=Y<TCELCQHW>AFY>;PGSSF<IBIU?H<YLQABWEGSQBMLHW
BJE?HUC=BTP?;?;LVTT>ZQI=IXTJ<P=IHYJIPAC?PGZH;OCCMXPOJHLIRHERSW;V
;@GZDHDNF><YCN@WLBTOFZRAILT?L;NT<OOCLLLWYAIVCSPBMVL???KK=TBZJ@KU
BD>=@;EOUFMUUFLSP<HMZWL?VLHXKK?S@WAFXC>EZODQOHREMNYKRBKAWS=PSXSQ
CRYFOPVJRBNHIVLE?BNPVNM=LO>BAT<?CQF?JNNQAUEHMM;BTU><=WFVOO<;WVKB
;RSBEXTS;QGBEKNDCANPFP;DYDHTSMQNAGVESB;TCWJKGWNBG>JBESVNKOVOBPBE
;?ZPYS;X?=VE=>YFHD;;;*"
)
```

Working with Expressions in an Encrypted Script

An encrypted script might not run the same as the unencrypted script if the script attempts to convert parts of itself from expressions back to character strings. This is necessary to keep the script encrypted. You can avoid this problem by using the expression manipulation functions such as `Arg()`, `Head()`, `Substitute()`, and so on. You can also use `Parse()` to make expressions from character strings that are part of the encrypted script.

For example, here is a script that modifies an expression the wrong way for encryption:

```
ex = Expr(
   aaa = 20;
   Show( aaa );
);
// Wrong, don't do this.
temp = (Parse( Substitute( Char( Name Expr( ex ) ), "aaa", "bbb" ) ));
temp;
```

The script fails when encrypted because `Char()` is unable to break the encryption of the expression stored in `ex`. A second example follows:

```
ex = Expr(
   aaa = 20;
   Show( aaa );
);
// Best choice if it works for you.
temp = Substitute( Name Expr( ex ), Expr( aaa ), Expr( bbb ) );
temp;
```

A third example, though not optimal, follows:

```
ch = "aaa = 20; Show(aaa);";
/* Leaves the ch variable holding unencrypted text
and Temp holds an unencrypted expression.*/
temp = Parse( Substitute( ch, "aaa", "bbb" ) );
temp;
```

You might prefer the third choice over the second choice because it looks easier to understand; be aware that you might want to clear variables like `ch` and `Temp` when you are done with them. You probably encrypted the script to keep other folks from reading it, and they will definitely be able to see the content of variables when you are done.

Additional Numeric Functions

JSL also offers several categories of operations that do not make much sense in the context of the formula editor: matrix operations and numeric derivatives of functions. Algebraic derivatives are also available.

The basic arithmetic operators can also be used with matrix arguments for matrix-wise addition, subtraction, and so on. Matrices also have a few special operators for elementwise multiplication and division, concatenation, and indexing. See "Matrices".

Derivatives

JSL has three internal functions (not all available in the calculator) for taking derivatives.

`Derivative()` takes the first derivative of an expression with respect to names you specify in the second argument. A single name might be entered as this second argument; or multiple values can be specified in a list, in other words, surrounded by braces.

Note: `Derivative()` is also available as an editing command inside the formula editor (calculator), located on the drop-down list in the top center of the formula editor (above the keypad). To use it, highlight a single variable in the expression (to designate which variable the derivative should be taken with respect to), then select the Derivative command from the menu. The whole formula is replaced by its derivative with respect to the highlighted name.

In scripts, the easiest way to use the function is with a single name. In this example, the mathematical notation is shown first followed by the JSL equivalent.

For $f(x) = x^3$, the first derivative is $f'(x)$ or $\frac{d}{dx}x^3 = 3x^2$.

```
result = Derivative( x ^ 3, x );
```

```
Show( result );
  result = 3 * x ^ 2
```

If you want an efficient expression to take the derivative with respect to several variables, then the variables are specified in a list. The result is a list containing a *threaded* version of the original expression, followed by expressions for the derivatives. The expression is *threaded* by inserting assignments to temporary variables of expressions that are needed in several places for the derivatives.

Here is an example involving an expression involving three variables. Listing all three variables returns the first derivatives with respect to each. The result is a list with the original expression and then the derivatives in the order requested. However, note here that JMP creates a temporary variable `T#1` for storing the subexpression `x^2`, and then uses that subexpression subsequently to save calculations.

```
result2 = Derivative( 3 * y * x ^ 2 + z ^ 3, {x, y, z} );
Show( result2 );
  result2 = {3 * y * (T#1 = x ^ 2) + z ^ 3, 6 * x * y, 3 * T#1, 3 * z ^ 2}
```

To take second derivatives, specify the variable as a third argument. Both the second and third arguments must be lists. JMP returns a list with the original expression, the first derivative(s), and then the second derivative(s) in the order requested.

```
second = Derivative( 3 * y * x ^ 2, {x}, {x} );
Show( second );
  second = {3 * y * x ^ 2, 6 * x * y, 6 * y}

second = Derivative( 3 * y * x ^ 2, {y}, {y} );
Show( second );
  second = {3 * y * (T#1 = x ^ 2), 3 * T#1, 0}

second = Derivative( 3 * y * x ^ 2, {y}, {x} );
Show( second );
  second = {3 * y * (T#2 = x ^ 2), 3 * T#2, 6 * x}
```

`Num Deriv()` takes the first numeric derivative of a function with respect to the value of the first argument by calculating the function at that value and at that value plus a small delta (Δ) and dividing the difference by the delta. `Num Deriv2()` computes the second numeric derivative in a similar fashion. These are used internally for nonlinear modeling but are not frequently useful in JSL. Note that these functions do not differentiate using a variable, but only with respect to arguments to a function. In order to differentiate with respect to x, you have to make x one of the immediate arguments, not a symbol buried deep into the expression.

Suppose to differentiate $y = 3x^2$ at the value of $x = 3$. The *incorrect* way would be to submit the following script:

```
x = 3;
n = Num Deriv( 3 * x ^ 2 );
```

The *correct* way is to make x an argument in the function.

```
x = 3;
f = Function( {x}, 3 * x ^ 2 );
n = Num Deriv( f( x ), 1 );
   18.000029999854
```

Consider both the mathematical notation and the JSL equivalent for another example:

For $f(x) = x^2$, it calculates $\frac{d}{dx}x^2 = \frac{(x+\Delta)^2 - x^2}{\Delta}$. At $x_0 = 3$, $\frac{d}{dx}x^2 = 6.00001$.

```
x = 3;
y = Num Deriv( x ^ 2 ); // or equivalently: y = Num Deriv( 3 ^ 2 );
   6.0000099999513
```

And here are a few more examples:

```
x = Num Deriv( Sqrt( 7 ) ); // returns 0.188982168980445
y = Num Deriv( Normal Distribution( 1 ) ); // returns 0.241969514669371
z = Num Deriv2( Normal Distribution( 1 ) ); // returns -0.241969777547979
```

Table 8.7 Derivative functions

Function	Syntax	Explanation
Derivative	Derivative(*expr*, {*name*, ...})	Returns the derivative of the *expr* with respect to *name*. Note that the second argument can be specified in a list with braces { } or simply as a variable if there is only one. Give two lists of names to take second derivatives.
NumDeriv	NumDeriv(*expr*)	Returns the first numeric derivative of the *expr* with respect to the first argument in the expression.
NumDeriv2	NumDeriv2(*expr*)	Returns the second numeric derivative of the *expr* with respect to the first argument in the expression.

Algebraic Manipulations

JSL provides a way of algebraically unwinding an expression (essentially, solving for a variable). It is accomplished through the `Invert Expr()` function.

`Invert Expr(`*expression*`, `*name*`, `*y*`)`

where

- *expression* is the expression to be inverted, or the name of a global containing the expression
- *name* is the name inside expression to unwind the expression around

- *y* is what the expression was originally equal to

For example,

```
Invert Expr( Sqrt( log( x ) ), x, y );
```

is wound around the name x (which should appear in the expression only once), and results in

```
Exp( y ^ 2 )
```

It is performed exactly as you would when doing the algebra by hand.

$y = \text{Sqrt}(\log(x));$

$y^2 = \text{Log}(x);$

$\text{Exp}(y^2) = x;$

`Invert Expr` supports most basic operations that are invertible, and makes assumptions as necessary, such as assuming you are interested only in the positive roots, and that the trigonometric functions are in invertible areas so that the inverse functions are legal.

F, Beta, Chi-square, *t*, Gamma, and Weibull distributions are supported for the first arguments in their Distribution and Quantile functions. If it encounters an expression that it cannot convert, `Invert Expr()` returns `Empty()`.

JSL provides a `Simplify Expr` command that takes a messy, complex formula and tries to simplify it using various algebraic rules. To use it, submit

```
result = Simplify Expr(expr(expression));
```

or

```
result = Simplify Expr(nameExpr(global));
```

For example,

```
Simplify Expr( Expr( 2 * 3 * a + b * (a + 3 - c) - a * b ) );
```

results in

```
6*a + 3*b + -1*b*c
```

`Simplify Expr()` also unwinds nested `If` expressions. For example:

```
r = Simplify Expr( Expr( If( cond1, result1, If( cond2, result2, If( cond3,
  result3, resultElse ) ) ) ) );
```

results in

```
If(cond1, result1, cond2, result2, cond3, result3, resultElse);
```

Maximize and Minimize

The `Maximize()` and `Minimize()` functions find the factor values that optimize an expression. The expression is assumed to be a continuous function of the factor values.

Here is the form of the call:

```
result = Maximize(objectiveExpression,{list of factor names},
  <<option(value))
result = Minimize(objectiveExpression,{list of factor names},
  <<option(value))
```

objectiveExpression is the expression whose value is to be optimized, and can either be the expression itself, or the name of a global containing a stored expression.

{*list of factor names*} is an expression yielding a list of names involved in *objectiveExpression*.

The name can be followed by limits that bound the permitted values, for example name(`lowerBound`,`upperBound`).

If you want to limit the values on one side, make the other side a missing value, for example:

```
{beta} // unconstrained
{beta (0,1)} // constrained between 0 and 1
{beta (.,1)} // upper limit of 1
{beta (0,.)} or {beta (0)} // lower limit of 0
```

Factor values can be either numbers or matrices.

Options available, shown with their default value, include:

```
<< Tolerance(.00000001) // convergence criterion
<< Max Iter( 250) // maximum number of iterations
<< Limits() //
```

Initial values are assumed to be already supplied the factor values before calling the function.

These functions are not expected to find global optima for functions that have multiple local optima; they are useful only for taking an initial value and moving it to either a local or global optimum.

The return value is currently the value of the objective function, if the optimization was successful, or `Empty()` if not.

Least Squares Example

The following example uses `Minimize` to find the least squares estimates of this exponential model, with data taken from the Nonlinear Example/US Population.jmp sample data table.

```
xx = [1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900,
  1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990];
yy = [3.929, 5.308, 7.239, 9.638, 12.866, 17.069, 23.191, 31.443, 39.818,
  50.155, 62.947, 75.994, 91.972, 105.71, 122.775, 131.669, 151.325, 179.323,
  203.211, 226.5, 248.7];
b0 = 3.9;
b1 = .022;
sseExpr = Expr(
  Sum( (yy - (b0 * Exp( b1 * (xx - 1790) ))) ^ 2 )
);
sse = Minimize( sseExpr, {b0, b1}, <<Tolerance( .00001 ) );
Show( b0, b1, sse );
  b0 = 13.9991388055261;
  b1 = 0.0147104409355048;
  sse = 1862.14141218875;
```

Identifying Differences Between Strings, Lines, or Sequences

The `Shortest Edit Script()` function compares two strings, lines, or sequences and returns a list of changes or a matrix that describes the changes. You might use `Shortest Edit Script()` to identify differences between two columns, lists, or matrices. The function describes *a* (not *the*) shortest list of instructions to convert sequence A into sequence B. There might be more than one shortest script.

Example of Sequences

```
t1 = New Table( "t1",
  New Column( "Column 1",
     Numeric,
     Continuous,
     Format( "Best", 12 ),
     Set Values( [1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, 7] )
  )
);

t2 = New Table( "t2",
  New Column( "Column 1",
     Numeric,
```

```
      Continuous,
      Format( "Best", 12 ),
      Set Values( [2, 3, 4, 5, 2, 3, 4, 5, 2, 3, 4, 5, 6] )
   )
);
EditScript = Shortest Edit Script( // compares column 1 in each table
   sequences(
      N Rows( t1 ),
      N Rows( t2 ),
      Function( {a, b}, // subscripts the columns
         t1:column 1[a] == t2:column 1[b] // in data tables t1 and t2
      )
   )
);
   [-1 1 . 1, // delete 1 item at position 1 in string a
   0 2 1 4, // keep 4 items at position 2 in string a, position 1 in string b
   -1 6 . 2, // delete 2 items at position 6 in string a
   1 . 5 4, // add 4 items at position 5 in string b
   0 8 9 5, // keep 5 items at position 8 in string a, position 9 in string b
   -1 13 . 1] // delete 1 item at position 13 in string a
```

Here are the columns in the matrix:

- Column 1: delete the items (-1), keep the common items (0), or add the items (1)
- Column 2: the position in string a
- Column 3: the position in string b
- Column 4: the number of items the instruction uses

A missing value indicates that the position isn't used in the comparison.

Example of Separators

The following example considers "@" and "$" to be separators.

```
aa = "this is$a test of@shortest$edit script lines$with several words";
bb = "this is a$test of$shortest$edit script lines with several@words";

// @ and $ separators
Shortest Edit Script( lines( aa, bb, separators( "@$" ) ) );
   {{"Remove", "this is$a test of@"}, {"Insert", "this is a$test of$"},
      {"Common",
   "shortest$"}, {"Remove", "edit script lines$with several words"}, {"Insert",
   "edit script lines with several@words"}}
```

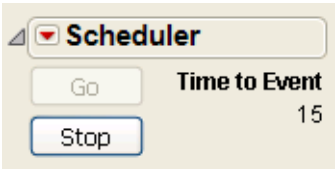
See the *JSL Syntax Reference*.

Scheduling Actions

A `Schedule()` function lets you set up a script to be executed some number of seconds later.

```
Schedule( 15, Print( "hello" ) );
```

Figure 8.6 JMP Scheduler



A Scheduler window shows the time until the next event and has buttons for restarting (Go) or stopping (Stop) the schedule. Its pop-up menu has a command Show Schedule, which echoes the current schedule to the log. For example, if you checked the schedule several times during the "hello" example, you would see something like this:

```
Scheduled at 11.55000000000018 :Print("hello")
Scheduled at 4.716666666666697 :Print("hello")
Scheduled at 3.083333333333485 :Print("hello")
```

The script might also be a name referring to a stored expression. For example, try submitting this script, which calls itself:

```
quickieScript = Expr(
   Show( "Hi there" );
   Schedule( 15, quickieScript );
);
quickieScript;
```

This script should show the string "Hi there" in a log window after 15 seconds, and reschedule itself for another 15 seconds, continuing until the Stop button is clicked.

More typically, in a production setting you might want to set up a schedule like this:

```
FifteenMinuteCheck = Expr(
   Show( "Checking data" );
   Open( "my file", options... );
   distribution( Column( column1 ), capability( spec limits ) );
   Schedule( 15 * 60, FifteenMinuteCheck );
);
FifteenMinuteCheck;
```

`Schedule()` initiates an event queue, but once it has the event queued, JMP proceeds with the next statement in the script. For example, the following has results that might surprise you:

```
Schedule( 3, Print( "one" ) );
```

```
Print( "two" );
   "two"
   "one"
```

If you want the script to wait until the scheduled events are finished before proceeding, one solution would be to use `Wait( )` with a suitable pause. Another is to embed the subsequent actions into the schedule queue. `Schedule()` accepts a series of arguments to queue many events in sequence. Each event is a separate call to the schedule. Each event time is an absolute time relative to "now" (or the instant that **Go** is clicked). Therefore, the following sequence finishes in five seconds, not in twelve:

```
Schedule( 3, Print( "hello" ) );
Schedule( 4, Print( ", world" ) );
Schedule( 5, Print( "--bye" ) );
```

To cancel all events in a schedule queue, use `Clear Schedule`.

```
scheduler[1] << Clear Schedule( );
```

Note: It is not possible to create multiple threads using `Schedule`.

Table 8.8 Schedule commands

Message	Syntax	Explanation
`Schedule`	*sc*=Schedule(*n*, *script*)	Queues an event to run the *script* after *n* seconds.
`Clear Schedule`	*sc*<<Clear Schedule()	Cancels all events in a schedule queue.

Functions that Communicate with Users

`Show()`, `Print()`, and `Write()` put messages in the log window. `Speak()`, `Caption()`, `Beep()`, and `StatusMsg()` provide ways to say something to a viewer. `Mail()` can send an e-mail alert to a process operator.

The JMP scripting language has methods for constructing dialog boxes to ask for data column choices and other types of information. See "Modal Windows".

Tip: To preserve locale-specific numeric formatting in `Show()`, `Print()`, or `Write()` output, include `<<Use Locale(1)`.

Writing to the Log

Show

`Show()` displays the items that you specify in the log. Notice that when you show variables, the resulting message is the variable name, an equal sign, and its current value.

```
X = 1;
A = "Hello, World";
Show( X, A, "foo" );
   x = 1
   a = "Hello, World"
   "foo"
```

Print

`Print()` sends the message that you specify to the log. `Print()` is the same as `Show()` except that it prints only the value of each variable without the variable name and an equal sign.

```
X = 1;
A = "Hello, World";
Print( X, A, "foo" );
   1
   "Hello, World"
   "foo"
```

Write

`Write()` sends the message that you specify to the log. `Write()` is the same as `Print()` except that it suppresses the quotation marks around the text string, and it does not start on a new line unless you include a return character yourself with the `\!N` escape sequence.

```
myText = "Here is a message.";
Write( "Here is a message." );
   Here is a message.

Write( myText || " Do not forget to buy milk." ); // use || to concatenate
Write( "\!NAnd bread." ); // use \!N for return
   Here is a message. Do not forget to buy milk.
   And bread.
```

The sequence `\!N` inserts the line breaking characters that are appropriate for the host environment. For an explanation of the three line breaking escape sequences, see "Double Quotes".

Send Information to the User

Beep

`Beep()` causes the user's computer to make an alert sound.

Speak

`Speak()` reads text aloud. On macOS, `Speak()` has one Boolean option, `Wait()`, to specify whether JMP should wait for speaking to finish before proceeding with the next step. The default is not to wait, and you need to issue `Wait(1)` each time. For example, here is a script certain to drive anybody crazy. With `Wait(1)`, you probably want to interrupt execution before too long. If you change it to `Wait(0)`, the iterations proceed faster than the speaking possibly can and the result sounds strange. On Windows, you can use a `Wait(n)` command to accomplish the same effect.

```
For( i = 99, i > 0, i--,
   Speak(
      Wait( 1 ),
      Char( i ) || " bottles of beer on the wall, " || Char( i ) || " bottles
   of beer; " ||
      "If one of those bottles should happen to fall, " || Char( i - 1 ) || "
   bottles of beer on the wall. "
   ) );
```

A more practical example has JMP announce the time every sixty seconds:

```
script = Expr(
   tod = Mod( Today(), In Days( 1 ) );
   hr = Floor( tod / In Hours( 1 ) );
   min = Floor( Mod( tod, In Hours( 1 ) ) / 60 );
   timeText = "time, " || Char( hr ) || " " || Char( min );
   text = Long Date( Today() ) || ", " || timeText;
   Speak( text );
   Show( text );
   Schedule( 60, script );    // seconds before next script
);
script;
```

You might use a similar technique to have JMP alert an operator that a process has gone out of control.

Caption

`Caption()` brings up a small window with a message to the viewer. Captions are a way to annotate demonstrations without adding superfluous objects to results windows. The first argument is an optional `{h,v}` screen location given in pixels from the upper left; the second argument is the text for the window. If the location argument is omitted, windows appear in the upper left corner.

You can include pauses in the playback by including the named argument `Delayed` and a time in seconds. Such a setting causes that caption *and all subsequent* caption windows to be delayed by that number of seconds, until a different `Delayed` setting is issued in a `Caption` statement. Use `Delayed(0)` to stop delaying altogether.

Specify the font type, font size, text color, or background color with the following arguments:

```
Font( font );
Font Size( size );
Text Color("color");
Back Color("color");
```

The `Spoken` option causes captions to be read aloud by the operating system's speech system (if available). `Spoken` takes a Boolean argument, and the current setting (on or off) remains in effect until switched by another `Caption` statement that includes a `Spoken` setting.

This script turns speaking on and leaves it on until the last caption. In the first caption, the font type, color, and background color is specified. Run the script and notice that the font and color settings apply only to the first caption.

```
Caption(
   {10, 30},
   "A Tour of the JMP Analyses",
   Font( "Arial Black" ),
   Font Size( 16 ),
   Text Color( "blue" ),
   Back Color( "yellow" ),
   Spoken( 1 ),
   Delayed( 5 )
);
Caption( "Open a data table." );
bigClass = Open( "$SAMPLE_DATA/Big Class.jmp" );
Caption( "A data table consists of rows and columns of data." );
Caption( "The rows are numbered and the columns are named." );
Caption( {250, 50}, "The data itself is in the grid on the right" );
Caption(
   {5, 30},
   Spoken( 0 ),
   "A panel along the left side shows columns and other attributes."
);
```

Each new `Caption` hides the previous one. In other words, there is only one caption window available at a time. To close a caption without displaying a new one, use the named argument `Remove`.

```
Caption( remove );
```

StatusMsg

This command sends a message to the status bar.

```
StatusMsg( "string" );
```

Mail

`Mail()` sends an e-mail message to a user. For example, a process control manager might include a test alert script in a control chart to trigger an e-mail warning to her pager:

```
Mail(
   "JaneDoe@company.com",
   "out of control",
   "Process 12A out of control at " || Format( Today(), "d/m/y h:m:s" )
);
```

`Mail()` can also send an attachment with the e-mail. An optional fourth argument specifies the attachment. The attachment is transferred in binary format after its existence on the disk is verified. For example, to attach the Big Class.jmp data table, submit

```
Mail(
   "JohnDoe@company.com",
   "Interesting Data Set",
   "Have a look at this class data.",
   "$SAMPLE_DATA\Big Class.jmp" );
```

Note: On macOS, `Mail()` works on Mountain Lion and later operating systems. On Mountain Lion, you must enter the e-mail address and subject in the e-mail due to operating system limitations. Click the **Send message** button to send the e-mail.

Chapter 9

Data Tables

Working with Data Table Objects

Before you can work with a data table or with objects in a data table, you must first open the data table and assign a *reference* to it. In the *Scripting Guide*, *dt* represents a reference to a data table object.

To manipulate an object in JSL, you send a message to the reference that represents the object, asking the object to perform one of its tasks. *Messages* are commands that can be understood only in context by a particular type of object. For example, messages for data table objects include Save, New Column, Sort, and so on. Most of these messages would not make sense for another object, such as a platform object.

In this chapter, you can learn the following:

- How to assign a reference to a data table
- How to send messages to the reference, resulting in specific actions
- About the different types of messages that data table objects can understand

Most of this chapter focuses on the different types of messages that you can send to data table objects, such as data tables, columns, rows, and values.

Tip: This chapter contains most, but not all of the JSL commands that you can use with data tables. For an exhaustive list, see the JMP Scripting Index. Select **Help > Scripting Index**. Select **Objects** from the menu, and then select **Data Table**.

Contents

Get Started with Data Tables . . . 327
Basic Data Table Scripting . . . 330
Open a Data Table . . . 330
Create a New Data Table . . . 332
Import Data . . . 333
Set the Current Data Table . . . 357
Name a Data Table . . . 358
Save a Data Table . . . 358
Hide a Data Table . . . 359
Print a Data Table . . . 360
Resize a Data Table . . . 360
Close a Data Table . . . 360
Set and Get a Data Table . . . 361
Perform Actions on All Open Data Tables . . . 361
Create Journals . . . 362
Create Excel Workbooks . . . 363
Prevent Changes to a Data Table . . . 364
Advanced Data Table Scripting . . . 365
Select Columns in a Data Table . . . 365
Store Summary Statistics in Global Variables . . . 366
Create a Table of Summary Statistics . . . 369
Subset a Data Table . . . 371
Add a Data Filter . . . 372
Sort a Data Table . . . 379
Stack Values in a Data Table . . . 379
Split Values in a Stacked Data Table . . . 380
Transpose a Data Table . . . 382
Vertically Concatenate Data Tables . . . 383
Horizontally Concatenate Data Tables . . . 383
Virtually Join Data Tables . . . 384
Replace Data in Data Tables . . . 387
Create a Table Using Tabulate . . . 389
Copy Data Tables . . . 390
Find Missing Data Patterns . . . 390
Compare Data Tables . . . 391
Create a Summary Table . . . 391
Subscribe to a Data Table . . . 392
Move Data Between Matrices and Data Tables . . . 394
Add Hyperlinks to Text Columns . . . 394

Columns 396
Send Messages to Data Column Objects 397
Create Columns 398
Create a New Column by Text Matching 400
Add Several Columns at Once 401
Group Columns 402
Use Set Selected to Select Columns 404
Rearrange and Move Columns 405
Recode Columns 406
Add a Column Switcher 407
Compress Selected Columns 408
Create Custom Functions, Transforms, and Formats 408
Custom Transform Example 411
Delete Columns 415
Get Column Names 415
Column Attributes 416
Column Properties 423
Specify Engineering SI Units 430
Rows 431
Add Rows 431
Delete Rows 432
Select Rows 433
Find Rows 436
Move Rows 437
Assign Colors and Markers to Rows 437
Color Cells 438
Hide, Exclude, and Label Rows 439
Iterate on Rows in a Table 440
Prevent Editing Rows 443
Row States 443
Accessing Data Values 460
Set or Get Values by Column Name 461
Additional Ways to Access Data Values 462
Add Metadata to a Data Table 463
Data Table Variables 463
Table Scripts 463
Formulas 468
Delete Metadata 469
Calculations 469
Pre-Evaluated Statistics 469

Calculator Formulas 471

Get Started with Data Tables

Tip: Keep the log window open to see the output of each script that you run. Select **View > Log** to open the Log window. See "Work with the Log".

The typical way to work with values in a data table is to follow these steps:

1. Set up the data table whose values you want to access as the current data table. Or, if you already have a data table reference, you can simply use that reference.
2. Specify the row or rows whose values you want to access and specify the column name that contains the values that you want to access.

The following example opens the Big Class.jmp sample data table (making it the current data table), and then specifies row 2 in the weight column. A value of 123 is returned in the log, which is the weight for Louise in row 2.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt:weight[2];
   123
```

If the data table that you want to work with is already open, proceed using one of the following examples:

```
dt = Data Table( "My Table" ); // the open table named My Table
dt = Current Data Table(); // the table in the active window
dt = Data Table( 3 ); // the third open table
```

Once you have an open data table with a reference, you can send it messages using either the << operator or the Send function. The following example illustrates both methods:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Save();
Send( dt, Save() );
```

Note the following about messages:

- You can send messages to any data table object (a data table, column, row, and so on).
- Messages sent to data table objects always have the same pattern:

  ```
  Reference << Message( Arguments );
  ```

- Usually, creating a reference and sending many messages to the reference is the easiest approach. However, you can also use the direct route instead of using a reference if you have only one message to send. The following example saves the current data table:

  ```
  Current Data Table() << Save();
  ```

- You can stack up a series of messages in one statement. Commands are evaluated from left to right, and each returns a reference to the affected object. The following example creates a new data table called My Table, adds two columns to it, and prompts you to save it:

  ```
  dt = New Table( "My Table" );
  dt << New Column( "Column 1" ) << New Column( "Column 2" ) << Save( "" );
  ```

  In this example, each message returns a reference to the data table (`dt`).
- If you specify too few arguments for a message, JMP presents a window to get the necessary information from you. JMP often presents windows when your script is incomplete, a behavior that you can use to your advantage when writing scripts that need to query users for their choices.
- Some messages come in pairs; one to "set" or assign each attribute, and one to "get" or query the current setting of each attribute.

Why are Some Commands Sent to Objects and Others Used Directly?

`New Table` and `Open` are commands to create objects that do not exist yet. Once created, you send them messages requesting changes. To close such objects, you must close the objects' *container*, because the objects cannot delete themselves.

The following example creates a table and assigns the data table reference to the *dt* variable. New Column messages are sent to the data table reference. To delete one of those columns, the Delete Columns message is sent to the data table reference, not to the column itself.

```
dt = New Table( "Airline Data" );
dt << New Column( "Date" );
dt << New Column( "Airline" );
dt << Delete Columns( "Date" );
```

How Can I See All of the Messages that Can be Sent to a Data Table Object?

To see all of the messages that can be sent to a data table object, see the Scripting Index:

1. Select **Help > Scripting Index**.
2. Select **All Categories** from the list.
3. Select **Data Table** in the list.

You can also use the `Show Properties()` command. This is a good approach if you want to print or copy the list of messages. The `Show Properties()` command lists the messages that you can send to a data table in the Log window. `Show Properties()` is a command that takes any scriptable object (such as a data table or column) as its argument. To show properties for a data table:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Show Properties( dt );
```

The resulting message list is hierarchical.

`[Subtable]` refers to a set of JMP menu commands. For example, the Tables subtable represents the JMP Tables menu, and the indented messages in this subtable correspond to commands in the Tables menu. Many messages include a short description.

Messages labeled as `[Action]` all result in some action being taken. Some messages are available for `[Scripting Only]`. Some messages take a Boolean argument, labeled as `[Boolean]`.

The platforms in the Analyze and Graph menus appear in the properties list because you can send a platform name as a message to a data table. This launches the platform for the data table through the usual launch window. For more information about writing platform-specific scripts, see "Scripting Platforms".

```
dt << Distribution( Y( height ) );
```

Note: In addition to data tables, `Show Properties` also works with platforms and display boxes.

The JMP Scripting Index provides more information about these properties. You can also run and modify sample scripts from the Scripting Index. Select **Help > Scripting Index** and search for the property in the Objects list.

Basic Data Table Scripting

Before you can work with data table objects, you must open or create a data table, and assign a reference to the data table.

- "Open a Data Table"
- "Create a New Data Table"
- "Import Data"
- "Set the Current Data Table"
- "Name a Data Table"
- "Save a Data Table"
- "Hide a Data Table"
- "Print a Data Table"
- "Resize a Data Table"
- "Close a Data Table"
- "Set and Get a Data Table"
- "Perform Actions on All Open Data Tables"
- "Create Journals"
- "Create Excel Workbooks"
- "Prevent Changes to a Data Table"

Open a Data Table

Use the `Open()` function to open a data table.

- To simply open a data table without returning a reference to it:

```
Open( "$SAMPLE_DATA/Big Class.jmp" ); // open the data table
```

- To open a data table and retain a reference to it:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" ); // open and store a reference
```

The path to the data table can be a quoted literal path (absolute or relative) or an unquoted expression that yields a pathname. Relative paths are interpreted relative to the location of the .jsl file (for a saved script). For unsaved scripts, the path is relative to your primary partition (Windows) or your <*username*>/Documents folder (macOS).

```
Open( "../My Data/Repairs.jmp" ); // relative path on Windows and macOS
Open( "::My Data:Repairs.jmp" ); // relative path on macOS
Open( "C:/My Data/Repairs.jmp" ); // absolute path
```

JMP provides shortcuts (*path variables*) to directories or files. Instead of entering the entire path to the directory or file, you include a path variable in the `Open()` expression. For example, JMP sample scripts typically use the $SAMPLE_DATA path variable to open files in the Samples/Data folder. For more information about path variables, see "Path Variables".

If you do not want to specify the entire path every time you open a data table, define a filepath string and concatenate the path with the filename:

```
myPath = "C:/My Data/Store25/Maintenance/Expenses/";
Open( myPath || "Repairs.jmp" );
```

Upon opening a data table, JMP stores the data table in memory. This means that if a script attempts to open a data table that is already open, the opened version appears. JMP does not read the version that is saved on your computer.

Test for an Open Data Table

In a script that depends on an opened data table, you can test to see whether the table is open using `Is Empty()` or `Is Scriptable()`. In the following example, the script performs a Bivariate analysis on Big Class.jmp and then closes the data table. Before proceeding to the Oneway analysis later in the script, `Is Scriptable()` tests for the open data table. If `1` (true) is returned, the table opens, and the script continues.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
obj = dt << Bivariate( Y( :height ), X( :age ), Fit Line );
Close( dt );
If( Not( Is Scriptable( dt ) ),
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" ),
);
obj = dt << Oneway( Y( :height ), X( :age ), Means( 1 ), Mean Diamonds( 1 ) );
```

You might want to see if the data table is open and then close it to discard any changes before running another script on the data table. To test for and close Big Class.jmp, run the following script:

```
Try( Close( Data Table( "Big Class" ) ) );
```

Prompt Users to Open a Data Table

You can use an `If()` expression to prompt a user to open a data table, if no open data table is found. And if they do not select a table, the script should end. The following script shows an example:

```
dt = Current Data Table();
If( Is Empty( dt ),
   Try( dt = Open(), Throw( "No data table found" ) )
);
```

The user is prompted to open a data table. If the user clicks Cancel instead of opening a data table, an error appears in the log.

Show Only Specific Columns

To open a data table and show only a specific set of columns, identify those columns in the `Open()` expression. This is particularly helpful with a large data table in which only a few columns are necessary.

The following example opens Big Class.jmp and includes only the age, height, and weight columns.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp", Select Columns( "age", "height",
  "weight" ) );
```

You can also specify the columns to leave out of the open data table:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp", Ignore Columns( "name", "sex" ) );
```

Create a New Data Table

You can start a new data table, or start a new data table and store its reference in a global variable. In either case, specify the table name as an argument.

```
New Table( "My Table" );
```

or

```
dt = New Table( "My Table" );
```

The following sections describe the optional arguments for the `New Table()` function.

Private, Invisible, and Visible

The optional `visibility` argument specifies whether the new data table is shown. `visibility( "invisible" )` hides the data table from view; it appears only in the JMP Home Window and the Window menu. `visibility( "private" )` avoids opening the data table. `visibility( "visible" )` shows the data table. `"visible"` is the default value.

```
New Table( "My Table", visibility( "private" ) );
```

A private data table enables you to avoid exposing the table to general use. Sometimes a script creates a private data table to store private data generated by a process. It is not for saving memory.

When you do not need a view of the data, or when you do not need to interact with the data, use an invisible data table.

Notes:

- The advantages of using private data tables is based on the size of the data in the table. If the data table has only five columns and 20 rows, the savings for a private data table is substantial. However, for a table with 100 columns and 1,000,000 rows, the overhead of the actual window is small in comparison to the actual data.
- Private tables are not counted when `N Table()` returns the number of open data tables.
- Private tables cannot be made current with `Current Data Table()`.

Actions

An optional argument that can define the new table. For example, the following script creates a new data table named *Little Class*. It adds three rows and two named columns as well as entering values for each cell in the table:

```
dt = New Table( "Little Class",
   Add Rows( 3 ),
   New Column( "name",
      Character,
      "Nominal",
      Set Values( {"KATIE", "LOUISE", "JANE"} )
   ),
   New Column( "height",
      "Continuous",
      Set Values( [59, 61, 55] )
   )
);
```

Import Data

Upon importing, JMP converts file types such as text files and Microsoft Excel files to the data table format. On Windows, JMP relies on filename extensions to identify the type of file and how to interpret its contents. Here are some examples of filename extensions identified by JMP:

```
Open( "$SAMPLE_IMPORT_DATA/Bigclass.xlsx" ); // Microsoft Excel file
Open( "$SAMPLE_IMPORT_DATA/Bigclass.txt" ); // text file
Open( "$SAMPLE_IMPORT_DATA/Carpoll.xpt" ); // SAS transport file
```

On macOS, JMP relies on the macOS Type and Creator codes (if present) and secondarily on three-letter filename extensions. Be sure to add the file extension before importing the file on macOS.

- Type and Creator codes are invisible data that enable the Finder to display a file with the correct icon corresponding to the application that created it.

- Files with generic icons should have the filename extensions (as with files created on other operating system).

```
Open( "$SAMPLE_IMPORT_DATA/Bigclass.txt", text );
```

Additional supported formats include .csv, .jsl, .dat, .tsv, and .jrn.

For more detailed information about import options, see the *JSL Syntax Reference*.

Import Multiple Files

When you save a script from the Multiple File Import (MFI) window into a script window, you get a compact form of the script that includes all of the parameters, default or otherwise, as messages. You could pare down the script as follows:

```
mfi = Multiple File Import(
   <<Set Folder( "$SAMPLE_IMPORT_DATA" ),
   <<Set Name Filter( "UN*.csv" ), // import files with this name
   <<Set Name Enable( 1 ),
   <<Set Add File Name Column( 1 ) // display the file names in a column
)
<<Import Data();
```

Note the slight distinction between the following expressions:

```
mfi = Multiple File Import();
```

`mfi` holds the Multiple File Import object.

```
dtlist = Multiple File Import(…) << Import Data;
```

`dtlist` holds the list of files produced by sending the `Import Data` message to the MFI object. The MFI object is gone because it was not stored in a variable or window. You can keep the MFI object by sending `Create Window` to it or storing the object in a variable, such as `mfi`, and by sending messages such as `Import Data` to that variable. Sending `Import Data` to an MFI object always returns a list, which could affect tasks such as creating reports.

`Create Window` can be useful for debugging. The window is visible *and* the MFI object is available. Run the following lines one at a time to see the results:

```
mfi = Multiple File Import();
mfi <<Create Window;
mfi <<Set Folder( "$DESKTOP" );
```

Here's an example of using a callback function to get a list of imported files after the user clicks Import on the Multiple File Import window.

```
Delete Directory( "$TEMP/test" );
Create Directory( "$TEMP/test" );
Save Text File( "$TEMP/test/argylesocks1.txt", "a1\!n1" ); // make some
Save Text File( "$TEMP/test/argylesocks2.txt", "a2\!n1" ); // data
```

```
global:importedFiles={};

mfi = Multiple File Import( // make an MFI object
     <<Set Folder( "$TEMP/test/" ),
     <<Set Name Filter( "argylesocks?.txt;" ),
     <<Set Name Enable( 1 ),
     <<Set Add File Name Column( 1 ),
     <<Set Import Callback(
           Function( {this, files}, // the callback function
                global:importedFiles = files;
/* optional: omit Close Window to keep the Multiple File Import window open so
   that the user can modify the settings. */
                win << Close Window;
           )
     )
);
win = mfi << Create Window; // open the window
While(
   Try(
      win << Parent;
      1;
   ,
      0
   ),
   Wait( 1 )
); // wait for the dialog to close
Show( global:importedFiles ); // proceed
```

Notes:

- See *Using JMP* for more information about MFI.
- See "Import Images from a Directory" for more information about using MFI to import images from a web page into a data table.

Import a PDF File

In a PDF file, data is laid out in tables. JMP auto-detects these tables to determine where the rows and columns begin and end.

You write scripts that specify the dimensions of each table on each page, the number of column headers, whether the tables should be concatenated or individual, and more.

Note: A PDF file consists of tables that contain data. In the PDF preview of the PDF Import Wizard, the tables are outlined and selected on each page.

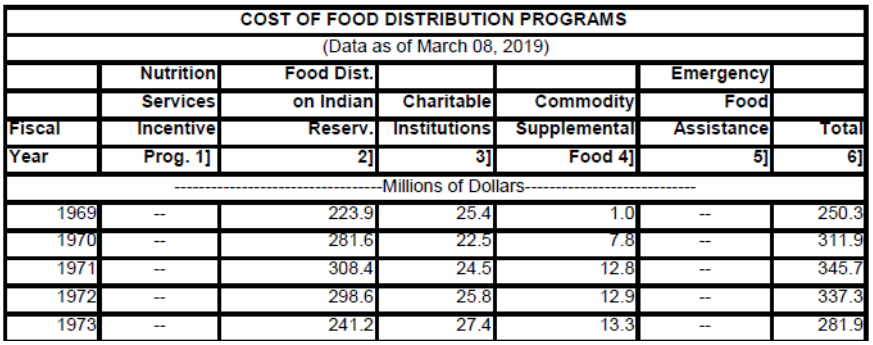
The following options are available for importing a PDF file:

```
Open( "PDF file path" ),
   Tables(
      Table( "data table name",
      Add Rows( Header Rows( number ),
      (Rect( left, right, top, bottom ),
      Page( number ), Row Borders(), Column Borders())) |
      Tables( Combine( All | Matching Headers | None),
         Minimum Rows( number ), Minimum Columns(#))))
   | PDF wizard );
```

Concatenate Tables Into One Data Table

This section shows how to import and concatenate the tables into one data table. The script that you create appears in a Source variable in the final data table, so you can run it again anytime.For this example, Figure 9.1 shows the sample import data file called Food Distribution.pdf, which contains one table per page.

Figure 9.1 Initial PDF File



COST OF FOOD DISTRIBUTION PROGRAMS

(Data as of March 08, 2019)

Fiscal Year	Nutrition Services Incentive Prog. 1]	Food Dist. on Indian Reserv. 2]	Charitable Institutions 3]	Commodity Supplemental Food 4]	Emergency Food Assistance 5]	Total 6]
----------Millions of Dollars----------						
1969	--	223.9	25.4	1.0	--	250.3
1970	--	281.6	22.5	7.8	--	311.9
1971	--	308.4	24.5	12.8	--	345.7
1972	--	298.6	25.8	12.9	--	337.3
1973	--	241.2	27.4	13.3	--	281.9

```
Open(
   "$SAMPLE_IMPORT_DATA/Food Distribution.pdf",
   PDF Tables(
      Table( // import tables into one table
         Table Name( "Food Distribution" ), // data table name
         Add Rows( // rows on page 1
            Header Rows( 7 ), // number of header rows
            Page( 1 ),
            Rect( 0.6922, 0.7155, 7.2842, 10.1124 ) // size of the table
         ), // rows and specifications on page 2
         Add Rows( Page( 2 ), Rect( 0.6922, 0.7156, 7.2842, 10.0708 ) )
      )
   )
);
```

Tip: Import a PDF file in the PDF Import Wizard and copy the script from the Source variable into a standalone script. This is a quick way to get the settings and avoid having to guess at the `Rect` settings.

Save as Individual Data Tables

The following syntax shows how to save tables in a PDF file into individual data tables. Each resulting data table contains a Source script like the one below.

```
Open(
   "PDF file path",
   PDF Tables(
      Table(
         Table Name( "data table name" ),
         Add Rows( Page( number ), Rect( left, right, top, bottom ) )
      )
   )
);
```

Concatenate Data Based on Column Headers

The following example shows how to concatenate data when column names on each page match.

Figure 9.2 Data with Matching Column Headers on each Page

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa

```
pdftable = Open(
   "$SAMPLE_IMPORT_DATA/Iris.pdf",
   PDF All Tables( Combine( Matching Headers ) )
);
```

Figure 9.3 Final Data Table

Iris Page(1) T...
Source

Columns (7/0)
Page
Table
Sepal length
Sepal width
Petal length
Petal width
Species

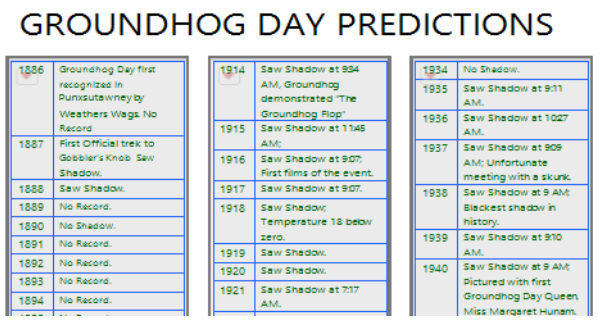
Rows
All rows 150
Selected 0
Excluded 0
Hidden 0
Labelled 0

	Page	Table	Sepal length	Sepal width	Petal length	Petal width	Species
1	1	1	5.1	3.5	1.4	0.2	setosa
2	1	1	4.9	3	1.4	0.2	setosa
3	1	1	4.7	3.2	1.3	0.2	setosa
4	1	1	4.6	3.1	1.5	0.2	setosa
5	1	1	5	3.6	1.4	0.2	setosa
6	1	1	5.4	3.9	1.7	0.4	setosa
7	1	1	4.6	3.4	1.4	0.3	setosa
8	1	1	5	3.4	1.5	0.2	setosa
9	1	1	4.4	2.9	1.4	0.2	setosa
10	1	1	4.9	3.1	1.5	0.1	setosa
11	1	1	5.4	3.7	1.5	0.2	setosa
12	1	1	4.8	3.4	1.6	0.2	setosa
13	1	1	4.8	3	1.4	0.1	setosa
14	1	1	4.3	3	1.1	0.1	setosa
15	1	1	5.8	4	1.2	0.2	setosa
16	1	1	5.7	4.4	1.5	0.4	setosa
17	1	1	5.4	3.9	1.3	0.4	setosa
18	1	1	5.1	3.5	1.4	0.3	setosa

Import PDF Pages that Contain Multiple Tables

A PDF file might contain data divided into multiple columns on each page. Figure 9.4 shows the sample import data PDF file called Groundhog Day Predictions.pdf, which contains multiple tables per page.

Figure 9.4 Multiple Columns on a Page


GROUNDHOG DAY PREDICTIONS

1886	Groundhog Day first recognized in Punxsutawney by Weathers Wags. No Record
1887	First Official trek to Gobbler's Knob. Saw Shadow.
1888	Saw Shadow.
1889	No Record.
1890	No Shadow.
1891	No Record.
1892	No Record.
1893	No Record.
1894	No Record.

1914	Saw Shadow at 9:34 AM. Groundhog demonstrated "The Groundhog Flop"
1915	Saw Shadow at 11:45 AM;
1916	Saw Shadow at 9:07; First films of the event.
1917	Saw Shadow at 9:07.
1918	Saw Shadow; Temperature 18 below zero.
1919	Saw Shadow.
1920	Saw Shadow.
1921	Saw Shadow at 7:17 AM.

1934	No Shadow.
1935	Saw Shadow at 9:11 A.M.
1936	Saw Shadow at 10:27 A.M.
1937	Saw Shadow at 9:09 AM; Unfortunate meeting with a skunk.
1938	Saw Shadow at 9 AM; Blackest shadow in history.
1939	Saw Shadow at 9:10 A.M.
1940	Saw Shadow at 9 AM; Pictured with first Groundhog Day Queen, Miss Margaret Hunam.

The following script imports all columns into a single data table.

```
Open(
  "$SAMPLE_IMPORT_DATA/Groundhog Day Predictions.pdf",
  PDF Tables(
    Table(
      Table Name( "Groundhog Day Predictions" ),
      // specifications for page 1
      Add Rows( Page( 1 ), Rect( 0.9726, 1.4325, 3.0351, 9.9974 ) ),
      Add Rows( Page( 1 ), Rect( 3.3063, 1.4325, 5.3688, 9.9608 ) ),
```

```
            Add Rows( Page( 1 ), Rect( 5.6399, 1.4325, 7.7025, 8.7925 ) ),
            // specifications for page 2
            Add Rows( Page( 2 ), Rect( 0.9726, 1.4325, 3.0351, 9.9441 ) ),
            Add Rows( Page( 2 ), Rect( 3.3063, 1.4325, 5.3688, 9.9274 ) ),
            Add Rows( Page( 2 ), Rect( 5.6399, 1.4325, 7.7025, 8.8225 ) ),
            // specifications for page 3
            Add Rows( Page( 3 ), Rect( 0.9726, 1.4325, 3.0351, 7.6404 ) ),
            Add Rows( Page( 3 ), Rect( 3.3063, 1.4325, 5.3688, 7.1254 ) ),
            Add Rows( Page( 3 ), Rect( 5.6399, 1.4325, 7.7025, 7.4721 ) )
        )
    )
);
```

Figure 9.5 Final Data Table

Groundhog Day P...
Source
Columns (4/0)
Page
Table
Column 3
Column 4
Rows
All rows 132

	Page	Table	Column 3	Column 4
1	1	1	1886	Groundhog Day first
2	1	1	1887	First Official trek to
3	1	1	1888	Saw Shadow.
4	1	1	1889	No Record.
5	1	1	1890	No Shadow.
6	1	1	1891	No Record.
7	1	1	1892	No Record.
8	1	1	1893	No Record.
9	1	1	1894	No Record.
10	1	1	1895	No Record.
11	1	1	1896	No Record.
12	1	1	1897	No Record.
13	1	1	1898	Saw Shadow.
14	1	1	1899	No Record. Groundh...
15	1	1	1900	Saw Shadow.

Export Data to a Text String

You can export data from a PDF file into a text string. You might want to do this to extract a pattern from the data.

Here is a basic example that imports all text from the Iris.pdf sample import data.

```
Open( "$SAMPLE_IMPORT_DATA/Iris.pdf", PDF text );
```

The following script finds the word “Figure” followed by a space, a digit, a period, and another digit and then prints the data to the log.

```
txt = Open( "$JMP_HOME/Documentation/JMP Documentation Library.pdf",
  PDFText() );
figures = {};
For( page = 1, page < N Items( txt ), page++,
  // search for the string
  next = Regex Match( txt[page], "Figure\s+\d+\.\d+", "" );
  While( N Items( next ) > 0,
```

```
        figures = Insert( figures, next );
        // extract the string
        next = Regex Match( txt[page], "Figure\s+\d+\.\d+", "" );
    );
);
Print( figures );
    {"Figure 2.1", "Figure 2.2", "Figure 2.3", "Figure 2.4", "Figure 2.4",
        "Figure 2.4", "Figure 2.4", "Figure 2.5", "Figure 2.6", "Figure 2.7",
        "Figure 2.8", "Figure 2.9", "Figure 2.10", "Figure 2.10", ...}
```

Import Data from a Text File

The Import Settings in the Text Data Files preferences determine how text files are imported. For example, column names begin on line one and data begin on line two by default. To use different settings, specify the import settings as `Open()` options in your script.

The default Import Settings and your custom import settings are saved in the data table Source script, so you can reimport the data using the same settings. However, the default Import Settings are optional in the script.

The following `Open()` options are available:

```
/* Charset options: "Best Guess", "utf-8", "utf-16", "us-ascii",
   "windows-1252", "x-max-roman", "x-mac-japanese", "shift-jis", "euc-jp",
   "utf-16be", "gb2312" */
CharSet("option")
Number of Columns(Number)

/* colType is Character|Numeric
colWidth is an integer specifying the width of the column */
Columns(colName=colType(colWidth),... )
Treat Empty Columns as Numeric(Boolean)
Scan Whole File(Boolean)
End Of Field(Tab|Space|Comma|Semicolon|Other|None)
EOF Other("char")
End Of Line(CRLF|CR|LF|Semicolon|Other)
EOL Other("Char")
Strip Quotes|Strip Enclosing Quotes(Boolean)
Labels | Table Contains Column Headers(Boolean)
Year Rule | Two digit year rule ("Decade Start")
Column Names Start | Column Names are on line(Number)
Data Starts | Data Starts on Line(Number)
Lines to Read(Number)
Use Apostrophe as Quotation Mark
CompressNumericColumns(Boolean)
CompressCharacterColumns(Boolean)
CompressAllowListCheck(Boolean)
```

The following script opens a text file of comma-delimited text, which includes no column names. The script defines the column names and the column widths.

```
Open(
   "$SAMPLE_IMPORT_DATA/EOF_comma.txt",
   End of Field( comma ),
   Labels( 0 ),
   Columns(
      name = Character( 12 ),
      age = Numeric( 5 ),
      sex = Character( 5 ),
      height = Numeric( 3 ),
      weight = Numeric( 3 )
   )
);
```

Here is an example of opening a text file in which the field separator is a space and the text file does not contain column headings:

```
Open(
   "$SAMPLE_IMPORT_DATA/EOF_space.txt",
   Labels( 0 ),
   End of Field( Space )
);
```

In the preceding example of a text file with space field separators, JMP can select the best file format so that you do not have to specify the labels or separator:

```
Open(
   "$SAMPLE_IMPORT_DATA/EOF_space.txt",
   Import Settings( Guess File Format() )
);
```

To set the import options interactively, include the `Text Wizard` argument. A preview of the text file opens in the text import window.

```
Open( "$SAMPLE_IMPORT_DATA/EOF_space.txt", "Text Wizard" );
```

The following sections describe each argument in more detail. For more information about import options, see the *JSL Syntax Reference*.

Number of Columns

Specifies the total number of columns in the source file. This option is important if data is not clearly delimited.

Columns

Identifies column names, column types, and column widths with a `Columns` argument as shown in the preceding examples.

If you specify settings for a column other than the first column in the file, you must also specify settings for all the columns that precede it. Suppose that you want to open a text file that has four columns (name, sex, and age, and ID, in that order). age is a numeric column, and the width should be 5. You must also set the name and sex column types and widths, and list them in the same order:

```
Columns(
   name = Character( 15 ),
   sex = Character( 5 ),
   age = Numeric( 5 )
);
```

You are not required to provide settings for any columns that follow the one that you want to set (in this example, ID).

After the data is imported, you use the modeling type for a column. See "Set or Get Data and Modeling Types".

Note: Most of the following arguments are defined in the JMP preferences. To override the preference, include the corresponding argument described below in your import scripts.

Treat Empty Columns as Numeric

Imports columns of missing data as numeric rather than character data. A period, Unicode dot, NaN, or a blank string are possible missing value indicators. This is a Boolean value. The default value is false.

Scan Whole File

Specifies how long JMP scans the file to determine data types for the columns. This is a Boolean value. The default value is true; the entire file is scanned until the data type is determined. To import large files, consider setting the value to false, which scans the file for five seconds.

Strip Quotes | Strip Enclosing Quotes

Specifies whether to include or remove the double quotation marks (") that surround string values. This is a Boolean value. The default value is `true`.

For example, suppose that the field delimiter is a space:

- *John Doe* is interpreted as two separate strings (*John* and *Doe)*.
- "*John Doe*" is interpreted as a single string. Most programs (including JMP) read a quotation mark and ignore other field delimiters until the second quotation occurs.
- If you include `Strip Quotes(1),` "*John Doe*" is interpreted as *John Doe* (one string without quotation marks).

Note that many word processors have a “smart quotation marks” feature that automatically converts double quotation marks (") into left and right curled quotation marks (“ ”). Smart quotation marks are interpreted literally as characters when the text file is imported, even when JMP strips double quotation marks.

End of Line(CR | LF | CRLF | Semicolon | Other)

Specifies the character or characters that separate rows:

- CR for carriage returns (typical for text files created on macOS up to version 9)
- LF for linefeeds (typical for UNIX and Mac OS X text files)
- CRLF for both a carriage return followed by a linefeed (typical for Windows text files).

All three characters are line delimiters by default.

Use the `Other` option to use an additional character for the row separator, which you must specify in the `EOLOther` argument. JMP interprets either this character or the default character as a row separator.

End of Field(Tab | Space | Spaces | Comma | Semicolon | Other | None)

Specifies the character or characters used to separate fields.

Notes:

- The default field delimiter is `Tab`.
- Use the `Other` option to use a different character for the field separator, which you must specify in the `EOFOther` argument.
- The `Space` option uses a single space as a delimiter.
- The `Spaces` option uses two or more spaces.

EOFOther, EOLOther

Specifies the character or characters used to separate fields or rows. For example, `EOLOther("*")` indicates that an asterisk separates rows in the text file.

Labels | Table Contains Column Headers

Indicates whether the first line of the text file contains column names. This is a Boolean value. The default value is true.

Year Rule | Two Digit Year Rule

Specifies how to import two-digit year values. If the earliest date is 1979, specify "1970". If the earliest date is 2001, specify "20xx".

Column Names Start | Column Names Are on Line

Specifies the starting line for column names. The following example specifies that the column names in the text file start on line three.

```
Open(
    "$SAMPLE_IMPORT_DATA/Animals_line3.txt",
    Columns(
        Column( "species", Character, "Nominal" ),
        Column( "subject", Numeric, "Continuous", Format( "Best", 10 ) ),
        Column( "miles", Numeric, "Continuous", Format( "Best", 10 ) ),
        Column( "season", Character, "Nominal" )
    ),
   Column Names Start( 3 )
);
```

Data Starts | Data Starts on Line

Specifies the starting line for data.

The following example specifies that the data in the text file start on line five.

```
Open(
   "$SAMPLE_IMPORT_DATA/Bigclass_L.txt",
   Columns(
      Column( "name", Character, "Nominal" ),
      Column( "age", Numeric, "Continuous", Format( "Best", 10 ) ),
      Column( "sex", Character, "Nominal" ),
      Column( "height", Numeric, "Continuous", Format( "Best", 10 ) ),
      Column( "weight", Numeric, "Continuous", Format( "Best", 10 ) )
   ),
   Data Starts( 5 )
);
```

Lines to Read

Specifies the number of lines to include in the data table. JMP starts counting after column names are read.

The following example includes only the first 10 lines in the data table.

```
Open(
    "$SAMPLE_IMPORT_DATA/Bigclass_L.txt",
    Columns(
        Column( "name", Character, "Nominal"),
        Column( "age", Numeric, "Continuous", Format( "Best", 10 ) ),
        Column( "sex", Character, "Nominal" ),
        Column( "height", Numeric, "Continuous", Format( "Best", 10 ) ),
        Column( "weight", Numeric, "Continuous", Format( "Best", 10 ) )
```

```
    ),
  Lines To Read( 10 )
);
```

Use Apostrophe as Quotation Mark

For data that are enclosed in apostrophes, this option treats apostrophes as quotation marks and omits them. For example, '2010' is imported as 2. This is a Boolean value. The default value is false.

This option is not recommended unless your data comes from a nonstandard source that places apostrophes around data fields rather than quotation marks.

Set Text Import Preferences

If you want to set preferences for importing text, it can be helpful to first see a list of all preferences. To do so, use the `Show Preferences (All)` function.

Then, copy and paste only the section you want into a Preferences function. For example, to specify import settings, write the following expression:

```
Preferences(
  Import Settings(
    End Of Field( Tab, Spaces, Comma )
  )
);
```

Import Data from a Microsoft Excel File

When you open a Microsoft Excel workbook in JMP, the file is converted to a data table. JMP supports .xls, .xlsm, and .xlsx formats. See *Using JMP* for more information about Microsoft Excel support.

Excel Preferences

In the JMP preferences, settings in the Third Party Data group can help determine how worksheets are imported:

Excel Open Method Specifies how a Microsoft Excel file should be opened by default, when using a non-specific open statement.

Use Excel Wizard Opens the Excel Import Wizard to import the file. This is the default setting.

Open All Sheets Opens all worksheets in the Microsoft Excel file.

Select Individual Worksheets Prompts users to select one or more worksheets when they open the file.

Use Excel Labels as Headings Determines whether text in the first row of the worksheet is converted to column headings in the data table.

By default, JMP takes the best guess. If names have been defined for all cells in the first row, the text in those cells is converted to column heading. Otherwise, columns are named *Column 1*, *Column 2*, and so on.

To override a preference, include the corresponding argument described later in this section.

Note: See "Create Excel Workbooks" for information about exporting data tables to an Excel workbook.

Open() Function

Using the `Open()` function without additional arguments to open an Excel file has different behaviors depending on the context:

- If the `Open()` function is a direct part of the script, the Excel files open into data tables using your Excel preferences. The following example opens both worksheets into data tables without using the wizard:

```
Path = Convert File Path( "$SAMPLE_IMPORT_DATA/Team Results.xlsx", absolute );
dt = Open( Path );
```

Note: To use the Excel Wizard, you must specify the `Excel Wizard` option in the argument as in `Open( "$SAMPLE_IMPORT_DATA/Team Results.xlsx", "Excel Wizard" );`.

- However, if the `Open()` function is part of a script that is run from clicking a button, the Preview window opens and requires the user to interact with it. Run the following example and click the button to see the Excel Import Wizard:

```
Path = Convert File Path( "$SAMPLE_IMPORT_DATA/Team Results.xlsx", absolute );
New Window( "button", Button Box( "Open", dt = Open( Path ) ) );
```

- To prevent the button script from opening the Preview window and importing the Excel file directly, provide additional arguments to the `Open()` function. Run the example and click the button. Both worksheets are opened into data tables without using the Excel Import Wizard.

```
Path = Convert File Path( "$SAMPLE_IMPORT_DATA/Team Results.xlsx", absolute );
New Window( "button", Button Box( "Open", dt = Open( Path, Use for all
   sheets(1) ) ) );
```

Alternatively, you can set the Excel Open Method preference to open all sheets:

```
Preference( Excel Open Method( "Open All Sheets" ) );
Path = Convert File Path( "$SAMPLE_IMPORT_DATA/Team Results.xlsx", absolute );
New Window( "button", Button Box( "Open", dt = Open( Path ) ) );
```

Open a Workbook in the Excel Import Wizard

The Excel Import Wizard shows a preview of the data and lets you modify the settings before importing the data. Specify `"Excel Wizard"` as the argument.

```
dt = Open( "$SAMPLE_IMPORT_DATA/Team Results.xlsx", Excel Wizard );
```

Customize Import Settings

Customize settings such as the number of rows in the column headers.

```
Open(
   "$SAMPLE_IMPORT_DATA/Team Results.xlsx",
   Worksheets( "Grouped Team Results" ),
   Use for all sheets( 1 ),
   Concatenate Worksheets( 0 ),
   Create Concatenation Column( 0 ),
   Worksheet Settings(
      1,
      Has Column Headers( 1 ),
      Number of Rows in Headers( 1 ),
      // customization based on Team Results structure
      Headers Start on Row( 3 )
   )
);
```

When you click a custom button to open a workbook, the Excel Open Method preference applies. To force the worksheet to open directly, specify the Excel Open Method in your script:

```
Preference( Excel Open Method( "Open All Sheets" ) );
```

Import Specific Worksheets

Suppose that you want to import data from specific worksheets in your workbook. Specify those worksheets using the `Worksheets` argument. In the following example, the worksheet named *small* is imported into JMP.

```
Open( "C:\My Data\cars.xlsx", Worksheets( "small" ) );
```

Or specify the number of the worksheet, the third worksheet in the following example:

```
Open( "C:\My Data\cars.xlsx", Worksheets( "3" ) );
```

Import multiple or all worksheets by including the worksheet names in a list:

```
Open( "C:\My Data\cars.xlsx", Worksheets( {"small", "medium", "large"} ) );
```

Import XML and JSON Files

The XML Import Wizard imports two types of nested text data: XML and JSON files. Both kinds of files contain structured text that can be nested to represent hierarchical relationships. This wizard helps you pick out the elements of the hierarchy that hold values and elements that determine rows to build a data table.

A useful way to imagine a nested text file: the document is a book of short stories. Some of the short stories might have chapters, others only have paragraphs. Chapters have paragraphs. Paragraphs have sentences. Here's an example of the full tree for such a book. This content is also available in the JMP Samples/Import Data directory as Book.xml.

```
<book>
  <story name="car poll">
    <wheels>4</wheels>
    <para><name>chev</name></para>
    <para><name>ford</name></para>
    <para><name>volk</name></para>
  </story>
  <story name="big class">
    <para><name>ralph</name><height>6</height></para>
    <para><name>billy</name><height>5</height></para>
  </story>
  <story name="cheese" location="NC">
    <chapter name="american">
      <para>1<price>2.00</price><quantity>2</quantity></para>
      <para>2<price>2.10</price><quantity>4</quantity></para>
      <para>3<price>2.20</price><quantity>3</quantity></para>
    </chapter>
    <chapter name="swiss">
      <para>1<price>3.00</price><quantity>3</quantity></para>
      <para>2<price>3.10</price><quantity>2</quantity></para>
    </chapter>
  </story>
</book>
```

This XML document (the book) contains three short stories; each short story is a data table, all different. This particular XML file uses paragraphs to represent rows in a table in all of the tables. That is not required, nor are any of the names special. If you wanted to import the cheese table, which is slightly more complicated than the other two, you'd probably want a variety column, containing "american" or "swiss" for each row. You would also want price and quantity columns. And you might want a column that doesn't have a clearly defined name for the 1, 2, 3, 1, 2 values of <para>. You'd expect five rows. Finally, you might want a table variable or a column for the location NC value.

The following example shows how to import the preceding data into a data table.

```
Open(
```

```
    "$SAMPLE_IMPORT_DATA/Book.xml",
    XML Settings(
       Row( "/book/story/chapter/para" ),
       // create a row for each value in <para>
       Col(
          "/book/story/@location",
          // create a column for the location attribute
          Column Name( "location" ),
          // rename the column
          Fill( "Use Forever" ),
          // fill an empty cell with the preceding value
          Type( "Character" ), // data type
          Format( "best" ) // column format
       ),
       Col(
          "/book/story/chapter/@name",
          Column Name( "variety" ),
          Fill( "Use Once" ),
          Type( "Character" ),
          Format( "best" )
       ),
       Col(
          "/book/story/chapter/para",
          Column Name( "interesting" ),
          Fill( "Use Once" ),
          Type( "Numeric" ),
          Format( "best" )
       ),
       Col(
          "/book/story/chapter/para/price",
          Column Name( "price" ),
          Fill( "Use Once" ),
          Type( "Character" ),
          Format( "best" )
       ),
       Col(
          "/book/story/chapter/para/quantity",
          Column Name( "quantity" ),
          Fill( "Use Once" ),
          Type( "Character" ),
          Format( "best" )
       )
    ),
    XML Wizard( 0 ) /* omit or change to 1 to preview the data
)                      in the XML Import Wizard */
```

Import from and Export to Google Sheets

A Google Sheet is a spreadsheet (which consists of sheets, like tabs) that is saved in your Google account. This enables you to continue to edit the spreadsheet and to share it with others.

To write a script to import data from a Google spreadsheet, log in to your Google account and get the URL for the spreadsheet. Also notice how the data is structured. For example, you must know which row the data begin on if there are column headers.

Give Google access to JMP when running the script for the first time.

To import data from a Google spreadsheet

```
dt = Google Sheet Import(
   Email( "yourgoogleaccount@gmail.com" ), // modify the email address
   // modify the sheet URL
   Spreadsheet(
   "https://docs.google.com/spreadsheets/d/1JO1M4ScGsAHb-bCUaOi575WWvQjXciQIV
   NOwVTrXDSe/edit#gid=2303744702" ),
   Sheets( "May","June" ), // specify the sheet name(s)
   Sheet Settings( // specify the structure of the data
      Has Column Headers( 1 ),
      Headers Start on Row ( 1 ),
      Data Starts on Row( 2 ),
      Cell Range( "" ),
      Import Cell Colors( 1 ),
      Suppress Empty Columns( 1 )
   )
);
```

Note: See *Using JMP* for more information about security, country restrictions, and more.

To export a JMP data table to a Google spreadsheet

Export a JMP data table to a new spreadsheet in Google Sheets by following this example:

```
email = "yourgoogleaccount@gmail.com"; // modify the Google account
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Google Sheet Export(
       dt,
       Email( email ),
      New Spreadsheet( "JSL Example" ), // specify the name of the spreadsheet
       Sheet Name( "Example 1" ) // specify the sheet name
);
```

“1” is returned in the log when the export is successful.

Export a JMP data table to an existing spreadsheet in Google Sheets by following this example. You'll need the URL of the existing spreadsheet.

```
id =
  "https://docs.google.com/spreadsheets/d/1J01M4ScGsAHb-bCUa0i575WWvQjXciQIV
  N0wVTrXDSe/edit#gid=2303744702"; // modify the URL
email = "yourgoogleaccount@gmail.com"; // modify the Google account
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Google Sheet Export(
       dt,
       Email( email ),
       Spreadsheet( id ),
       Sheet Name( "Example 1" )
);
```

"1" is returned in the log when the export is successful.

Note: See *Using JMP* for more information about security, country restrictions, and more.

Import SAS Data Sets

Open a SAS file as a data table without connecting to a SAS server.

```
sasxpt = Open( "$SAMPLE_IMPORT_DATA/carpoll.xpt" );
```

To convert the labels to column headings, include the `Use Labels for Var Names` argument.

```
sasdbf = Open( "$SAMPLE_IMPORT_DATA/Bigclass.sas7bdat", Use Labels for Var
  Names( 1 ));
```

.xpt and .stx file formats are also supported.

On Windows, you can also open SAS data sets from a SAS server. See "Connect to a SAS Metadata Server".

Import Web Pages and Remote Files

You can import data from websites or from other computers. The data might be a JMP data table, a table defined in a web page, or another file type that JMP supports.

Open or Import Data from a Website

In the `Open()` command, specify the quoted URL to open a file from a website. You can open JMP data tables or other supported file types this way.

```
Open( "http://company1.com/Repairs.jmp" );
Open( "http://company1.com/My Data.txt", text); // specify text on macOS
```

Import a Web Page

A web page can include data in tabular format. Import the table as a JMP data table:

```
Open( "http://company1.com", HTML Table( n ) );
```

n identifies which table you want to import. For example, to import the fourth table on the page, specify `HTML Table(4)`. If you omit the value, only the first table on the page is imported.

JMP attempts to preserve the table header defined in a `<th>` HTML tag. The table header is converted to column headings in the data table. If the `<th>` tag is wrong or missing, use `ColumnNames(n)` to specify the *n*th row. By default, `DataStarts(n)` will be the next row, or you can specify the `DataStarts` row.

Import Images from a Web Page

You can import images on the web into a data table, using the Expression data type. The following example creates a new data table and adds two images to it.

```
dt = New Table( "test", New Column( "Image", Expression ) );
dt << Add Rows( 1 );
dt:Image = New Image( Open(
  "https://www.jmp.com/support/help/en/16.0/jmp/images/iris_first.gif" )
  );
dt << Add Rows( 1 );
dt:Image = New Image( Open(
  "https://www.jmp.com/support/help/en/16.0/jmp/images/trimeth.png" ) );
```

Import Images from a Directory

You can use `Multiple File Import()` and an Event Handler script to create a data table of thumbnail images that link to a directory of full-size pictures. This example sets the cell height so that it fits the height of the tallest thumbnail.

Note: Typically, the Multiple File Import option is for importing multiple text files. However, this example shows how to use the option to import images.

```
path = "$SAMPLE_IMAGES/"; // images are in this directory

{dtx} = Multiple File Import(
  <<Set Folder( path ),
  <<Set Name Filter( "*.jpg; " ),
  <<Set Name Enable( 1 ), /* enable the ability to filter by name,
  in this example, files with the .jpg extension */
  <<Set Add File Name Column( 1 ), // must be included but can be hidden
  <<Set Add File Size Column( 1 ) // for Show Sizes at end
```

```
) << Import Data;

/* FileName is the column that contains the image file name.
The column is required for the image links to work, but you don't need
to see it. */
dtx:FileName << Hide;
maxheight = 0;

For Each Row( // recreate each image to minimize size of data table
   x = Eval( dtx:Picture );
   x << Scale( .1 ); // scale the image by 10%
   m = x << Get Pixels; // compress the image by recreating it
   dtx:Picture = New Image( m );
   maxheight = Max( maxheight, (dtx:Picture[] << Get Size)[2] );
   // width x height is returned, [2] gets the height
);

Eval( /* insert the value in the path variable into the Click handler. The
   path variable won't be available later, but the path value is needed to
   concatenate with the file name column's value to build a link to the image
   on the computer. */
   Eval Expr(
      dtx:picture << Set Property(
         "Event Handler",
         Event Handler( // Event Handler column property
            Click( JSL Quote( Function( {thisTable, thisColumn, iRow},
            // insert links to images
            Open( Expr( path )||Char( thisTable:FileName[ iRow ] ) ); ); ) ),

            Tip( JSL Quote( Function( {thisTable, thisColumn, iRow}, "Open " ||
   Char( thisTable:FileName[ iRow ] ) || " in your viewer."; ); ) ),

            // color of the link -- blue
            Color( JSL Quote( Function( {thisTable, thisColumn, iRow}, 5; );) )
         )          )
   )
);

// add two pixels to the maximum height
dtx << Set Cell Height( maxheight + 2 );

/* save the file to show that the data table is much smaller than the
original file */
dtx << Save( "$TEMP/Sample Images.jmp" );
Show( File Size( "$TEMP/Sample Images.jmp" ), Col Sum( dtx:FileSize ) );
```

Import a File from a Shared Computer

JMP can import files stored on a shared computer, such as another computer or a network drive. The following example shows how to open a file from Windows:

```
Open( "\\Data\Repairs.jmp" ); // UNC name
Open( "z:/Data/Repairs.jmp" ); // mapped letter drive
Open( "Repairs.jmp" ); // relative path in which the script is in the same
  directory as the file that you are importing
```

macOS doesn't have a concept like UNC naming or drive letters. Instead, you need to mount the volume through the user interface. It is possible to do it from a script, but you would need to have connected interactively at least once and stored the password in your keychain.

Here is an example that connects to a remote volume from either macOS or Windows:

```
mount = Function( {server},
  {Default Local},

  If( Host is( Mac ),
    Run Program(
      Executable( "/usr/bin/osascript" ),
      // replace ^server^ with your server name throughout the script
      Options( {"-e", Eval Insert( "mount volume \!"smb://^server^\!"" )} ),
      Read Function( "text" )
    );
    server = Concat Items( Remove( Words( server, "/" ), 1 ), "/" );
    Eval Insert( "/Volumes/^server^" );
  ,
    Eval Insert( "\!\\!\^server^" )
  )
);


users = mount( "data.company.com/Users" );
Files In Directory( Eval Insert( "^users^/Smith" ) );
```

Import JSON Data

JSON (JavaScript Object Notation) is a data interchange format that is based on JavaScript. It shares many of the same principles as languages based on C and is easy to read and write. Because of these characteristics, JSON is highly portable and can be imported into JMP.

Table 9.1 Representations of JSON in JSL

JSON	JSL
array	list (sequence of values)
object	associative array (set of named values)
number	number
string	string
true	1.0
false	0.0
null	. (missing value)

JSON consists of name-value pairs in an associative array. For example, here is JSON code that shows name-value pairs:

```
"[ { \!"key1\!": \!"a\!", \!"key2\!": \!"b\!"}, {\!"key1\!": \!"c\!",
  \!"key2\!": \!"d\!" }, {\!"key2\!": \!"f\!", \!"key1\!":\!"e\!",
  \!"key3\!": \!"g\!"}]";
```

Each list is imported as a column in the data table. “key1” is the name of the first column. The first value in the column is “a”. The second value is “c”, and the third value is “e”.

Use `JSON To Data Table()` to convert the JSON into a data table.

```
js = "[ { \!"key1\!": \!"a\!", \!"key2\!": \!"b\!"}, {\!"key1\!": \!"c\!",
  \!"key2\!": \!"d\!" }, {\!"key2\!": \!"f\!", \!"key1\!":\!"e\!",
  \!"key3\!": \!"g\!"}]";
d = JSON To Data Table( js ); // convert JSON to a data table
d << New Data View; // make a view of the data table
```

Figure 9.6 Imported JSON Data

	key1	key2	key3
1	a	b	
2	c	d	
3	e	f	g

You can also use `JSON To List()` to convert a string containing JSON into a list.

```
list = JSON To List(
```

```
    "[ { \!"name\!": \!"KATIE\!", \!"age\!": 12, \!"sex\!": \!"F\!",
    \!"height\!": 59, \!"weight\!": 95 }, { \!"name\!": \!"LOUISE\!",
    \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!": 61, \!"weight\!": 123 }, {
    \!"name\!": \!"JANE\!", \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!":
    55, \!"weight\!": 74 } ]"
);
Show( list );
   list = {{{"name", "KATIE"}, {"age", 12}, {"sex", "F"}, {"height", 59},
      {"weight", 95}}, ...}};
```

Exporting JSL to JSON

`As JSON Expr()` converts an associative array to a JSON string. You can then export the data into a JSON document.

```
list_of_associative_arrays = {["key1" => "a", "key2" => "b"], ["key1" => "c",
"key2" => "d"], ["key2" => "f", "key1" => "e", "key3" => "g"]};
jx = As JSON Expr( list_of_associative_arrays );
   "[{\!"key1\!":\!"a\!",\!"key2\!":\!"b\!"}..."[{\!"key1\!":\!"a\!",\!"key2\
      !":
   \!"b\!"}
```

Note: To import the data table as invisible, include the argument `Invisible(1)` in `JSON To Data Table()` and omit `New Data View`.

Import HDF5 Files

Hierarchical Data Format, Version 5 (HDF5) is a portable file format for storing data. An HDF5 file consists of groups and datasets. When you import the file, JMP opens a group to present the names of the inner datasets.

JMP handles only tables with numeric (integer, float, double) and string types, and compound files with three or fewer dimensions that contain only simple types.

You can import up to 1,000,000 columns and an unbounded number of rows.

Here is the syntax:

```
Open( "filename.h5", {"list_of", "dataset_names"});
```

Errors are written to the log when invalid data set names are passed.

Import ESRI Shapefiles

An ESRI shapefile is a geospatial vector data format used to create maps. JMP imports shapefiles as data tables. A .shp shapefile consists of coordinates for each shape. A .dbf shapefile includes values that refer to regions. To create maps in JMP, you modify the structure of the data and save the files with specific suffixes.

The following example imports a .shp file and saves it with the -XY suffix.

```
dt = Open( "$SAMPLE_IMPORT_DATA/Parishes.shp"),
:X << Format( "Longitude DDD", 14, 4 );
:Y << Format( "Latitude DDD", 14, 4 ) );
dt << Save( "c:/Parishes-XY.jmp" );
```

Save the .dbf file with the -Name suffix.

```
dt = Open( "$SAMPLE_IMPORT_DATA/Parishes.dbf" );
dt << Save( "c:/Parishes-Name.jmp" );
```

Restructuring the data requires several steps, including adding a Map Role column property to names in the -Name.jmp file. See *Essential Graphing*.

Import a Database

`Open Database()` opens a database using Open Database Connectivity (ODBC) and extracts data into a JMP data table. See the "Database Access".

JMP also converts DataBase Files (.dbf) files to data table format.

```
sasdbf = Open( "$SAMPLE_IMPORT_DATA/Bigclass.dbf",
  Use Labels for Var Names( 1 )
);
```

Set the Current Data Table

Tip: Be careful about assuming that the data table that you want is the current data table, even if you set it to be earlier in your script, because interim actions could change that.

A data table becomes the current data table in the following instances:

- When you open it like this:

  ```
  dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
  ```

- When you create a new table like this:

  ```
  dt2 = New Table( "Cities" );
  ```

To switch to another open data table, specify the name of the data table in `Current Data Table()`:

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt2 = New Table( "Cities" );
Current Data Table( dt1 ); // make Big Class.jmp the current data table
```

`Current Data Table()` can also take a scriptable object reference as an argument. The following expression would make the data table in use by the second Bivariate object the current data table:

```
Current Data Table( Bivariate[2] );
```

For more information about using references to analysis platform objects, see "Send Messages to a Platform".

Name a Data Table

Assign a name to a data table by sending it the `Set Name` message. The argument is a filename in quotation marks, or something that evaluates to a filename.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Set Name( "New Big Class.jmp" );

s = "New Big Class";
dt << Set Name( s );
```

To retrieve the name, send a `Get Name` message to the data table:

```
dt << Get Name();
   "New Big Class"
```

Save a Data Table

To save a data table, send a `Save` message to the data table. Here are some examples:

```
dt << Save(); // save using the current name
dt << Save( "Newest Big Class.jmp" ) // save as a new file
dt << Save( "c:/My Data/New Big Class.jmp" );  // save as a new file
dt << Save( "My Table", JMP( 5 ) ); // save as JMP 5 table
dt << Save("") // prompt to select the directory and save in the desired
   format
dt << Save( "Big Class.xlsx" ); // save as a Microsoft Excel file
Create Excel Workbook( "c:/MyData/data.xlsx", {"Abrasion", "Big Class"},
{"Abrasive", "Class"} ); // save several data table as a workbook
```

Note: If you specify the filename with no path and have not set the default directory, the file is saved on your primary partition (Windows) or in your *<username>*/Documents folder (macOS). For more information about setting the default directory, see "Relative Paths".

On Windows, saving with a `.txt` extension exports according to the Text Export preferences. On macOS, add `Text` as a second argument to the `Save` function:

```
dt << Save( "New Big Class.txt", Text );
```

If you plan to set the name of a data table and later send the `Save` message, you can just specify the name in a `Save` message.

```
dt << Set Name( "New Big Class.jmp" );
```

```
dt << Save();
```

works the same as

```
dt << Save( "New Big Class.jmp" );
```

Including Save and the pathname is also an alternative to using Save As along with the pathname.

Revert to a Saved Data Table

To return to the most recently saved data table, send a Revert message to the data table.

```
dt << Revert();
```

Hide a Data Table

There are two ways to hide a data table if the user does not need to see the table: by opening it as invisible or private.

Invisible Data Tables

An *invisible* data table is hidden from view but linked to analyses that you run on it. Open a data table as an invisible file:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp", "invisible" );
```

Here is an example of creating an invisible table, called Abc, that has ten rows. One column is named X.

```
dt = New Table( "Abc", "invisible", New Column( "X" ), Add Rows( 10 ) );
```

To find out if an open data table is invisible, pass the Boolean Has Data View message to the data table object. The following expression returns 0 (false) if the data table is invisible:

```
dt << Has Data View();
```

After you are finished with an invisible data table, do not forget to close it with the Close() function. Otherwise, the data table remains in memory until you quit JMP.

Showing an Invisible Data Table

The Show Window() function shows an invisible data table.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp", "invisible" );
dt << Show Window( 1 );
```

The user has the option of opening the table from the JMP interface. On Windows, an invisible data table appears in the Window List on the Home Window and the **Window > Unhide** menu. On macOS, the data table appears in the JMP Home window and the **Window > Hidden** menu.

Private Data Tables

Completely hide the data table from view by including the `private` argument:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp", "private" );
```

A private data table enables you to avoid exposing the table to general use. Sometimes a script creates a private data table to store private data generated by a process. It is not for saving memory.

See "Private, Invisible, and Visible".

Print a Data Table

Print a data table by sending it a `Print Window` message. JMP uses your computer's default printer settings.

```
dt << Print Window();
```

This message also applies to other display boxes.

Resize a Data Table

dt `<< Optimize Display` forces the data table to remeasure all of its columns and zoom to the best-sized window.

```
dt << Optimize Display;
```

Close a Data Table

To close a data table, use a `Close` command with the data table's reference as the argument. Data tables with unsaved changes are saved automatically, including any linked tables. Reports and graphs generated from the data tables are also closed.

To save and close the data table:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Close( dt );
```

To close the data table without saving changes:

```
Close( dt, No Save );
```

To save a copy with a new name and then close the original version:

```
Close( dt, Save("c:/My File.jmp" ) );
```

You can also close all data tables at once and either save or discard changes:

```
Close All( Data Tables );
```

```
Close All( Data Tables, No Save );
```

To close all private or invisible data tables, specify the appropriate argument:

```
Close All( Data Tables, private );
Close All( Data Tables, invisible );
```

Tip: `Close All()` also works with journals. Specify "Journals" in the argument.

Set and Get a Data Table

To specify the data table in a display box, for example, in `New Data Box()`, use `Set Data Table()` and `Get Data Table()` to set and return a data table.

The following example uses the Big Class.jmp and Cities.jmp sample data tables. The sample data tables are opened invisibly. In a new window, a new data box is created containing Big Class.jmp. After waiting 1 second, the data box contents change to display cities.jmp.

```
dtC = Open( "$SAMPLE_DATA/Cities.jmp", "invisible" );
dtA = Open( "$SAMPLE_DATA/Big Class.jmp", "invisible" );
nw = New Window( "My Table", H List Box( dtBox = dtA << New Data Box() ) );
Wait( 1.0 );
dtBox << Set Data Table( Data Table( "Cities.jmp" ) );
```

Perform Actions on All Open Data Tables

If you want to perform an action on all of the data tables that are currently open (except private tables), use `N Table()` to get a list of references to each one:

```
dt1 = Open( "$SAMPLE_DATA/Cities.jmp" );
dt2 = Open( "$SAMPLE_DATA/Big Class.jmp" );
openDTs = List();
For( i = 1, i <= N Table(), i++,
  Insert Into( openDTs, Data Table( i ) );
);
```

`openDTs` now is a list of references to all open data tables. You can send messages to any one by using `openDTs(n)`. You can use a for loop to send messages to all of the open data tables. This loop adds a new column named My Column to each open data table.

```
For( i = 1, i <= N Items( openDTs ), i++,
  openDTs[i] << New Column( "My Column" );
);
```

If you just want a list of table names and not references, use the `Get Name` message:

```
For( i = 1, i <= N Table(), i++,
  Insert Into( openDTs, Data Table( i ) << Get Name());
```

```
);
```

You can then use the list of table names to put the names of all of the open data tables in a list box (window) so that the user can choose and perform tables operations. Or, you can write the names out to a file.

Create Journals

Journals consist of JMP graphs and reports, graphics, text, and links to items such as web pages and files. They make it easy to reuse content in presentations and import into other documents.

Here is the basic syntax for journals:

```
dt << Journal;
```

You can also create a new journal window and immediately add to it within the `New Window()` command. The following examples create an empty untitled journal:

```
New Window( <<  Journal);
```

The following examples create an empty journal named “Sales”:

```
New Window( "Sales", << Journal );
```

Here is an example of creating a journal named “Test Buttons” that contains two buttons.

```
New Window( "Test Buttons", << Journal,
   Button Box( "Test One", New Window( "Hi there1", << Modal ) ),
   Button Box( "Test Two", New Window( "Hi there2", << Modal ) ) );
```

Capturing Journals

There is an optional argument in the journal and journal window commands. This argument captures (or freezes) the current display. That is, it converts the display tree with just a picture.

Here are the four options for arguments:

- Makes a bitmap “snapshot” of the display when it is sent to the journal, rather than sending a clone of the more editable display box structure to the journal. By freezing the display into a bitmap, references to variables from scripts in graphs are less problematic.

  ```
  << Journal("Freeze All");
  ```

- Similar to `Freeze All` but only within areas of a report called pictures.

  ```
  << Journal("Freeze Pictures");
  ```

- Similar to `Freeze Pictures` but only within frame boxes (which are within pictures).

  ```
  << Journal("Freeze Frames");
  ```

- Similar to `Freeze Frames` but only for the frame boxes that have scripts to draw something on them.

```
<< Journal("Freeze Frames with Scripts");
```

Current Journal

`Current Journal()` returns a reference to the display box at the top of the current journal display window. If no journal is open, one is created. There are no arguments.

You can add to the journal using the Append command. The following example adds a text box to the bottom of the current journal:

```
Current Journal() <<  Append( Text Box( "Hello World" ) );
```

You can also find items in an existing journal by enclosing the search specification in square brackets. Suppose that the current journal includes the string "Parameter Estimates". The following script appends a text box to the bottom of that journal:

```
Current Journal()["Parameter Estimates"] << Append( Text Box( "Asterisks show
  items significant at 0.05" ) );
```

Create Excel Workbooks

You can save open data tables in a new or existing Excel workbook by using the `Create Excel Workbook()` function. The first example adds all open data tables without changing the worksheet names:

```
dt1 = Open( "$SAMPLE_DATA/Big Class Families.jmp" );
dt2 = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
Create Excel Workbook( "c:/MyWorkbook1.xlsx" );
```

The following example creates MyWorkbook2.xlsx by combining the open Big Class Families.jmp and San Francisco Crime.jmp sample data tables.

```
dt1 = Open( "$SAMPLE_DATA/Big Class Families.jmp" );
dt2 = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );

// specify the data tables
tableList = {"Big Class Families", "San Francisco Crime"};

// specify the worksheets (optional)
sheetList = {"Families", "Crime"};
Create Excel Workbook( "c:/MyWorkbook2.xlsx", tableList, sheetList );
```

Use data table references to create the workbook.

```
dt1 = Open( "$SAMPLE_DATA/Big Class Families.jmp" );
dt2 = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
Create Excel Workbook(
  "c:/MyWorkbook3.xlsx",
  {dt1, dt2}, // data table references
```

```
    {"Families", "Crime"}
);
```

`Create Excel Workbook()` also accepts a mix of data table references and data table names in the list. For example, `{dt1, dt2}` above can also be expressed as `{dt1, "San Francisco Crime"}`.

Notes:

- The saved file is an .xlsx file; the .xls format is not supported.
- The JMP data tables must conform to the maximum Excel limit of 1 million rows and 16,535 columns, or the data will be truncated in Excel.
- Data table references or data table names must always be in a list, even if the list has only one item.

Prevent Changes to a Data Table

To prevent any changes to the data table (such as changing column names, changing the data type, deleting rows, and so on), use `Lock Data(1)`.

```
dt << Lock Data Table( 1 );
```

To prevent modifying adding or deleting columns and rows and modifying cells, use `Set Edit Lock`:

```
dt << Set Edit Lock( "Add Columns" );
dt << Set Edit Lock( "Delete Columns" );
dt << Set Edit Lock( "Add Rows" );
dt << Set Edit Lock( "Delete Rows" );
dt << Set Edit Lock( "Modify Cells" );
```

To get a list of data table operations that are locked, use `Get Edit Lock`:

```
dt << Get Edit Lock();
   {"Modify Cells", "Add Rows", "Delete Rows", "Add Columns", "Delete Columns"}
```

To allow a specific data table operation again, use `Clear Edit Lock`:

```
dt << Clear Edit Lock( "Modify Cells" );
```

Advanced Data Table Scripting

This section covers more advanced actions that you can perform on a data table.

- "Select Columns in a Data Table"
- "Store Summary Statistics in Global Variables"
- "Create a Table of Summary Statistics"
- "Subset a Data Table"
- "Add a Data Filter"
- "Sort a Data Table"
- "Stack Values in a Data Table"
- "Split Values in a Stacked Data Table"
- "Transpose a Data Table"
- "Vertically Concatenate Data Tables"
- "Horizontally Concatenate Data Tables"
- "Virtually Join Data Tables"
- "Replace Data in Data Tables"
- "Create a Table Using Tabulate"
- "Copy Data Tables"
- "Find Missing Data Patterns"
- "Compare Data Tables"
- "Create a Summary Table"
- "Subscribe to a Data Table"
- "Move Data Between Matrices and Data Tables"
- "Add Hyperlinks to Text Columns"

Select Columns in a Data Table

To select one or more columns, send the `Select Columns` message to the data table.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
clist = {:Height, :Weight};
dt << Select Columns( clist );
```

To select all columns, send the Select Columns message with the All argument to the data table.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Columns( all );
```

Store Summary Statistics in Global Variables

The `Summarize` command collects summary statistics for a data table and stores them in global variables. The `Summarize` command is different from the `Summary` command, which also calculates summary statistics, but presents them in a new data table.

The first argument is an optional data table reference. Include it if more than one data table might be open.

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt2 = Open( "$SAMPLE_DATA/Animals.jmp" );
Summarize( dt1,
   exg = By( :sex ),
   exm = Mean( :height )
);
Show( exg );
Show( exm );
```

Named arguments include the following: `Count`, `Sum`, `Mean`, `Min`, `Max`, `StdDev`, `First`, `Corr`, and `Quantile`. These statistics can be calculated only for numeric columns. Each argument takes a data column argument.

Notes:

- If a *name*=`By(`*groupvar*`)` statement is included, then a list of subgroup statistics is assigned to each *name*.
- `Count` does not require a column argument, but it is often useful to specify a column to count the number of nonmissing values.
- `Quantile` also takes a second argument for specifying which quantile, such as 0.1 for the 10th percentile.
- When you summarize data that contain row states, the summary statistics are not calculated. An error appears in the log instead.

Note: Excluded rows are excluded from `Summarize` calculations. If all data are excluded, Summarize returns lists of missing values. If all data have been deleted (there are no rows), Summarize returns empty lists.

The following example uses the Big Class sample data table:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Summarize(
   a = By( :age ),
   c = Count,
   sumHt = Sum( :height ),
   meanHt = Mean( :height ),
   minHt = Min( :height ),
```

```
    maxHt = Max( :height ),
    sdHt = Std Dev( :height ),
    q10Ht = Quantile( :height, .10 )
);
Show( a, c, sumHt, meanHt, minHt, maxHt, sdHt, q10Ht );
```

Because the script included a By group, the results are a list and six matrices:

```
a = {"12", "13", "14", "15", "16", "17"}
c = [8, 7, 12, 7, 3, 3]
sumHt = [465, 422, 770, 452, 193, 200]
meanHt = [58.125, 60.28571428571428, 64.16666666666667, 64.57142857142857,
   64.33333333333333, 66.66666666666667]
minHt = [51, 56, 61, 62, 60, 62]
maxHt = [66, 65, 69, 67, 68, 70]
sdHt = [5.083235752381126, 3.039423504234876, 2.367712103711172,
   1.988059594776032, 4.041451884327343, 4.163331998932229]
q10Ht = [51, 56, 61.3, 62, 60, 62]
```

You can format the results using `Table Box`.

```
New Window( "Summary Results",
  Table Box(
     String Col Box( "Age", a ),
     Number Col Box( "Count", c ),
     Number Col Box( "Sum", sumHt ),
     Number Col Box( "Mean", meanHt ),
     Number Col Box( "Min", minHt ),
     Number Col Box( "Max", maxHt ),
     Number Col Box( "SD", sdHt ),
     Number Col Box( "Q10", q10Ht )
  )
);
```

Figure 9.7 Results from Summarize

Age	Count	Sum	Mean	Min	Max	SD	Q10
12	8	465	58.125	51	66	5.08324	51
13	7	422	60.2857	56	65	3.03942	56
14	12	770	64.1667	61	69	2.36771	61.3
15	7	452	64.5714	62	67	1.98806	62
16	3	193	64.3333	60	68	4.04145	60
17	3	200	66.6667	62	70	4.16333	62

You can add totals to the window:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );

Summarize(
  a = By( :age ),
  c = Count,
  sumHt = Sum( :height ),
```

```
    meanHt = Mean( :height ),
    minHt = Min( :height ),
    maxHt = Max( :height ),
    sdHt = Std Dev( :height ),
    q10Ht = Quantile( :height, .10 )
);

Summarize(
    tc = Count,
    tsumHt = Sum( :height ),
    tmeanHt = Mean( :height ),
    tminHt = Min( :height ),
    tmaxHt = Max( :height ),
    tsdHt = Std Dev( :height ),
    tq10Ht = Quantile( :height, .10 )
);

Insert Into( a, "Total" );
c = c |/ tc;
sumHt = sumHt |/ tsumHt;
meanHt = meanHt |/ tmeanHt;
minHt = minHt |/ tminHt;
maxHt = maxHt |/ tmaxHt;
sdHt = sdHt |/ tsdHt;
q10Ht = q10Ht |/ tq10Ht;

New Window( "Summary Results",
    Table Box(
        String Col Box( "Age", a ),
        Number Col Box( "Count", c ),
        Number Col Box( "Sum", sumHt ),
        Number Col Box( "Mean", meanHt ),
        Number Col Box( "Min", minHt ),
        Number Col Box( "Max", maxHt ),
        Number Col Box( "SD", sdHt ),
        Number Col Box( "Q10", q10Ht )
    )
);
```

Figure 9.8 Summarize with Total

Age	Count	Sum	Mean	Min	Max	SD	Q10
12	8	465	58.125	51	66	5.08324	51
13	7	422	60.2857	56	65	3.03942	56
14	12	770	64.1667	61	69	2.36771	61.3
15	7	452	64.5714	62	67	1.98806	62
16	3	193	64.3333	60	68	4.04145	60
17	3	200	66.6667	62	70	4.16333	62
Total	40	2502	62.55	51	70	4.24234	56.2

If you do *not* specify a By group, the result in each name is a single value:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Summarize(
// a = By( :age ),
   c = Count,
   sumHt = Sum( :height ),
   meanHt = Mean( :height ),
   minHt = Min( :height ),
   maxHt = Max( :height ),
   sdHt = Std Dev( :height ),
   q10Ht = Quantile( :height, .10 )
);
Show( c, sumHt, meanHt, minHt, maxHt, sdHt, q10Ht );
   c = 40;
   sumHt = 2502;
   meanHt = 62.55;
   minHt = 51;
   maxHt = 70;
   sdHt = 4.24233849397192;
   q10Ht = 56.2;
```

Summarize supports multiple By groups. For example, in Big Class.jmp, proceed as follows:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Summarize( g = By( :age, :sex ), c = Count() );
Show( g, c );
   g = {"12", "12", "13", "13", "14", "14", "15", "15", "16", "16", "17", "17"},
      {"F", "M", "F", "M", "F", "M", "F", "M", "F", "M", "F", "M"}}
   c = [5,3,3,4,5,7,2,5,2,1,1,2]
```

If you specify a By group, the results are always matrices. Otherwise, the results are scalars.

Create a Table of Summary Statistics

The Summary command creates a new table of summary statistics according to the grouping columns that you specify. Do not confuse Summary with Summarize, which collects summary statistics for a data table and stores them in global variables. See "Store Summary Statistics in Global Variables".

```
summDt = dt << Summary(
   Group( groupingColumns ),
   Subgroup( subGroupColumn ),
   Statistic( columns ),// where statistic is Mean, Min, Max, Std Dev, and so
   on.
   Output Table Name( newName ) );
```

The following example creates a new table with columns for the mean of height and weight by age, and the maximum height and minimum weight by age:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
summDt = dt << Summary(
  Group( :age ),
  Mean( :height, :weight ), Max( :height ), Min( :weight ),
  Output Table Name( "Height-Weight Chart" ) );
```

Tip: `Output Table Name` can take a quoted string or a variable that is a string.

By default, a summary table is linked to the original data table. If you want to produce a summary that is not linked to the original data table, add this option to your `Summary` message:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
summDt = dt << Summary( Group( :age ), Mean( :height ),
  Link to Original Data Table( 0 )
);
```

You can choose to add marginal statistics for grouping variables to the output columns of your data table. JMP also adds a row to the end of the data table that summarizes each level of the grouping variables. To add marginal statistics, add this option to your `Summary` message:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
summDt = dt << Summary( Group( :age ), Mean( :height ),
  Include marginal statistics
);
```

You can specify the format of the statistics column name using `statistics column name format()`. Table 9.2 shows the option and examples:

Table 9.2 Statistics Column Name Format Options and Examples

Option	Example
stat (column)	Mean (Profits ($M))
column	Profits ($M)
stat of column	Mean of Profits ($M)
column stat	Profits ($M) Mean

For example, add an option like this to your `Summary` message:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
summDt = dt << Summary( Group( :age ), Mean( :height ),
  statistics column name format( "stat of column" )
);
```

The height column is named Mean of height.

Subset a Data Table

Subset() creates a new data table from rows that you specify. If you specify no rows, Subset uses the selected rows. If no rows are selected or specified, it uses all rows. If no columns are selected or specified, it uses all columns. And if Subset has no arguments, the Subset window appears.

```
dt << Subset((<"Private">|<"Invisible">), <Selected Columns>, <Columns(column
   list)>, <Selected Rows>, <Rows([number, number, ...])>, <By(column
   list)>, <Sampling Rate(fraction)>, <Sample Size(integer)>,
   <Stratify(column list)>, <Link to Original Data Table(Boolean)>, <Copy
   Formula(Boolean)>, <Suppress Formula Evaluation(Boolean)>, <Keep by
   Column>);
```

Note: For more arguments, see the Scripting Index in the Help menu.

For example, using Big Class.jmp, to select the columns for all rows in which the age is 12:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row( Selected( Row State() ) = (:age == 12) );
subdt = dt << Subset( Output Table Name( "Subset" ) );
```

To select three columns and all rows, run this script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
subDt1 = dt << Subset(
   Columns( :name, :age, :height ),
   Output Table Name( "Big Class 2" )
);
```

To select specified rows of two columns and link the two data tables, run this script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
subDt2 = dt << Subset(
   Columns( :name, :weight ),
   Rows( [2, 4, 6, 8] ),
   Linked
);
```

To select the columns for all rows in which the age is 12, run this script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age == 12 );
dt << Subset( ( Selected Rows ), Output Table Name( "Subset" ) );
```

To verify that rows were selected and then conditionally proceed to subset the current data table or open another data table, run this script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age == 12 );

/* If the condition selected no rows, open Fitness.jmp.
   Else, generate the desired subset table of Big Class.jmp.*/
If( N Rows( dt << Get Selected Rows ) == 0,

newdt = Open( "$SAMPLE_DATA/Fitness.jmp" ),
   newdt = dt << Subset( All Columns( 1 ), Selected Rows( 1 ) )
);
```

Add a Data Filter

The Data Filter provides a variety of ways to identify subsets of data. Using Data Filter commands and options, you can select complex subsets of data, hide these subsets in plots, or exclude them from analyses.

Here is the basic syntax:

```
dt << Data Filter(<Columns>, <Mode(...)>, <Add Filter(...)>, <Show Columns(),
   <Selector>, <Animation>, <Animate>, <Clear>, <Delete All>, <Delete>,
   <Filter Columns>, <Filter Column>, <Filter Group>, <Add Filter>, <Add>,
   <Match>, <Mode>, <Display>, <Where>, <Make Subset>, <Display>, <Show
   Histograms and Bars>)
```

You can send an empty `Data Filter` message to a data table, and the initial Data Filter window appears, showing the Add Filter Columns panel that lists all the variables in the data table.

`Mode` takes three arguments, all of which are optional: `Select(`*`Boolean`*`) Show(`*`Boolean`*`), Include(`*`Boolean`*`)`. These arguments turn on or off the corresponding options. The default value for `Select` is true (1). The default value for `Show` and `Include` is false (0).

`Add Filter()` adds rows and builds the WHERE clauses that describe a subset of the data table. Here is the basic syntax:

```
Add Filter( Columns( col, ... ), Where( ... ), ... )
```

To add columns to the data filter, list the columns names separated by commas. Note that this is not a list data structure.

You can define one or more WHERE clauses to specify the filtered columns:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
df = dt << Data Filter(
   Mode( Show( 1 ) ),
```

```
    Add Filter(
       Columns( :age, :sex, :height ),
       Where( :age == {13, 14, 15} ),
       Where( :sex == "M" ),
       Where( :height >= 52 & :height <= 65 )
    )
);
```

This script selects 13, 14, and 15-year-old males with a height greater than or equal to 52 and less than or equal to 65. In this filtered data, you can instead select the excluded ages by adding the `Invert Selection` message to the preceding script.

```
df << ( Filter Column( :age ) << Invert Selection );
```

In this example, ages other than 13, 14, and 15 in the filtered data are selected.

You can also use `Add Filter()` to select matching strings from columns with the Multiple Response property or Multiple Response modeling type.

```
dt = Open( "$SAMPLE_DATA/Consumer Preferences.jmp" );
df = dt << Data Filter(
   Location( {437, 194} ),
   Add Filter(
      Columns( :Brush Delimited ),
      Match None( Where( :Brush Delimited == {"Before Sleep", "Wake"} ) ),

      /* Define the display options. "Brush Delimited" appears above
      the list of values. Size is the size of the filter list. The
      values appear next to check boxes. */
      Display( :Brush Delimited, Size( 121, 70 ), Check Box Display )
      )
);
```

This script selects rows with values in the Brush Delimited column that do not match either of the specified values ("Before Sleep", "Wake"). Other available scripting options include `Match Any`, `Match All`, `Match Exactly`, and `Match Only`. See *Using JMP* for more information about the Multiple Response property and the Multiple Response modeling type.

You can also send messages to an existing Data Filter object:

```
Clear(), Display( ... ), Animate(), Mode(), ...
```

`Clear` takes no arguments and clears the data filter.

To prevent the data filter from appearing when the script is run, set `Show Window` to 0:

```
obj  = ( dt << Data Filter( Show Window( 0 ) ) );
```

To specify which values to exclude, use the `!=` operator. The following example excludes ages 16 and 17 from the filtered values.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
```

```
df = dt << Data Filter(
   Add Filter( Columns( :age, :sex ) ),
   Match( Columns( :age, :sex ),
      Where( :sex = "M" ), Where( :age != {16, 17} )
   )
);
```

To omit the Data Filter outline and red triangle menu, specify `No Outline Box`.

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
obj = dt << Data Filter(
   Add Filter( Columns( :Region, :POP ), Where( :Region == {"C", "N"} ) ),
   No Outline Box( 1 )
);
```

Add a Local Data Filter

A local data filter enables you to filter data in an individual report window without affecting other reports. When the local data filter is added to the platform, the reference to the data filter is returned. This provides an easy way to get access to the filter.

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
biv = dt << Bivariate(
   Y( :x ),
   X( :y ),
   Automatic Recalc( 1 ),
   Fit Line( {Line Color( {213, 72, 87} )} )
);

dfObj = biv << Local Data Filter(
// Include Show Histograms and Bars(0) to hide them in the filter.
// Enable Include to include rows selected by the local data filter in
// calculations that are required to build graphs controlled by the local data
// filter.  Deselected rows are excluded from calculations.
   Location( {566, 47} ),
   Mode( Include( 1 ) )
);

dfObj << Add Filter(
   Columns( :Region, :POP ),
   Where( :Region == {"C", "N"} )
);
```

Track Data Filter Changes

Make Filter Change Handler enables you to track changes to the data filter.

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
```

```
dist = Distribution(
  Automatic Recalc( 1 ),
  Continuous Distribution( Column( :POP ) )
);
filter = dist << Local Data Filter(Add Filter( columns( :Region ) ));
f = Function( {a}, Print( a ) );
rs = filter << Make Filter Change Handler( f );
```

Define the Context of a Data Filter

A data filter lets you interactively select complex subsets of data, hide the subsets in plots, or exclude them from analyses. Your selections affect all analyses of the data table.

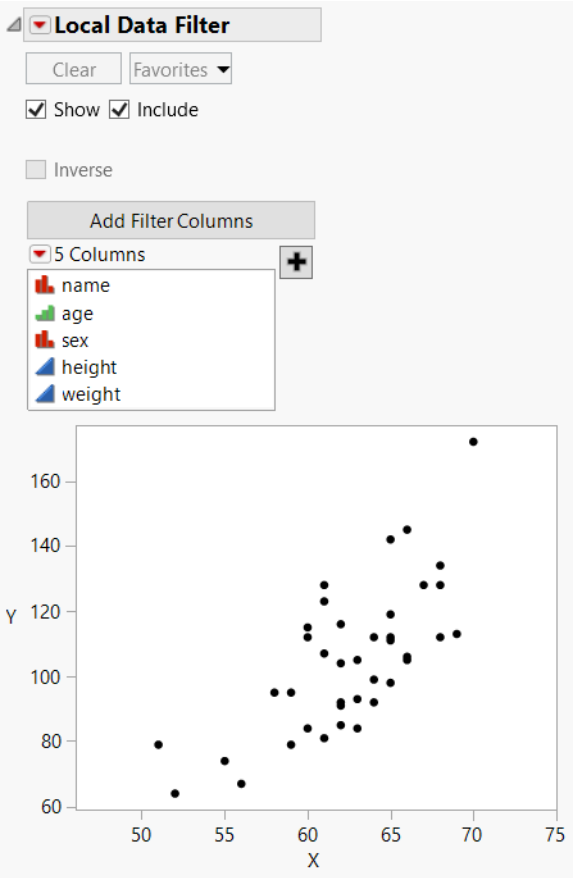
Another option is to filter data from specific platforms or display boxes. Create a local data filter inside the `Data Filter Context Box()` function. This defines the context as the current platform or display box rather than the data table.

The following example creates a local data filter for a Graph Box. See Figure 9.9 for the output.

```
New Window( "Marker Seg Example",
  Data Filter Context Box(
    V List Box(
      dt << Data Filter( Local ),
      g = Graph Box(
        Frame Size( 300, 240 ),
        X Scale( Min( xx ) - 5, Max( xx ) + 5 ),
        Y Scale( Min( yy ) - 5, Max( yy ) + 5 ),
        Marker Seg( xx, yy, Row States( dt, rows ) )
      )
    )
  )
);
```

Tip: The preceding script is part of a larger script that first builds the arrays required for a marker seg. To experiment with this script, open the Local Data Filter for Custom Graph.jsl sample script.

Figure 9.9 Local Data Filter and Graph



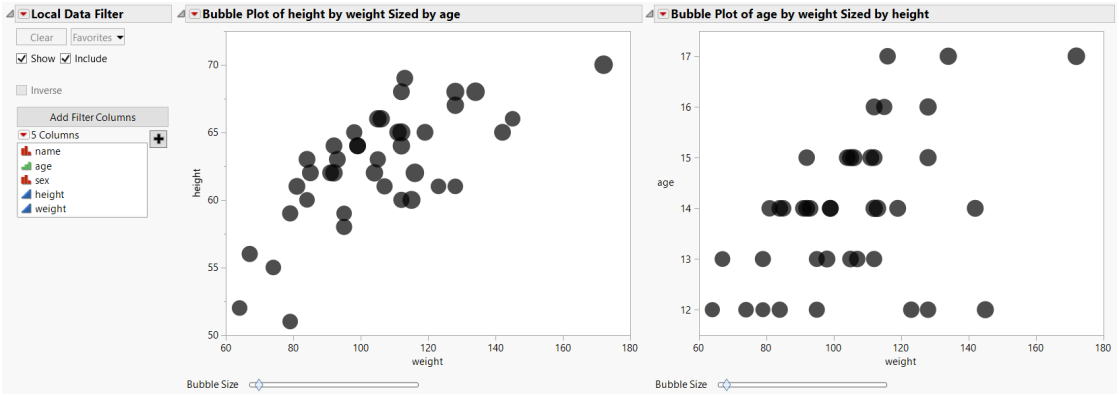
A context box can also contain several graphs with one local filter. The filtering then applies to both graphs. The following script creates two bubble plots and one data filter in one context box. See Figure 9.10 for the output.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
New Window( "Shared Local Filter",
   Data Filter Context Box(
      H List Box(
         dt << Data Filter( Local ),
         Platform(
            dt,
            Bubble Plot( X( :weight ), Y( :height ), Sizes( :age ) )
         ),
         Platform(
            dt,
            Bubble Plot( X( :weight ), Y( :age ), Sizes( :height ) )
         )
```

```
      )
   )
);
```

Tip: To experiment with this script, open the Local Data Filter Shared.jsl sample script.

Figure 9.10 Local Filter with Two Bubble Plots



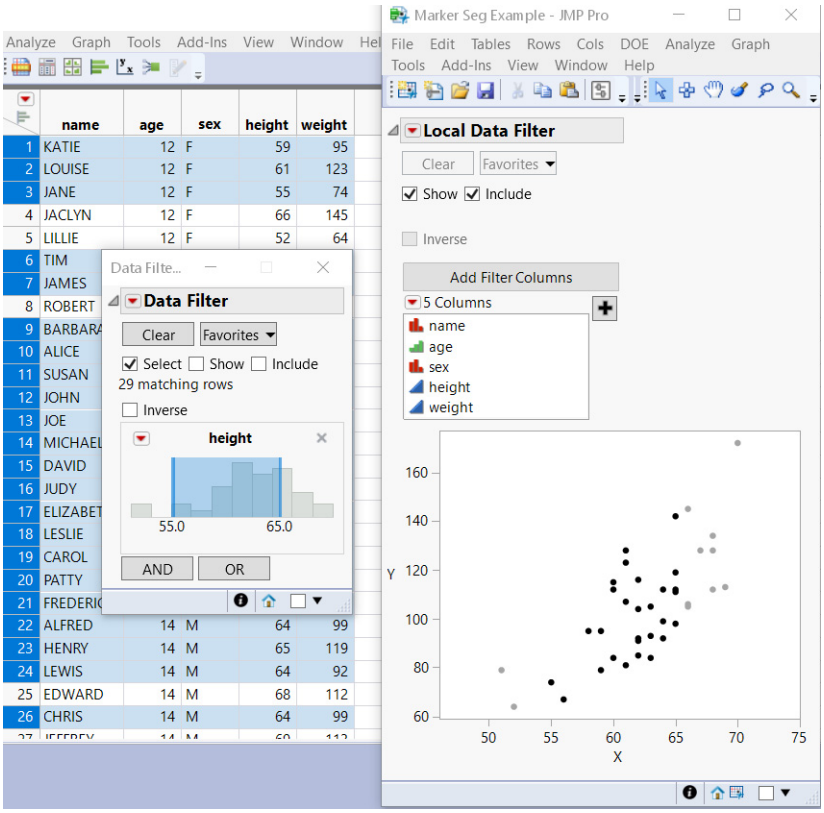
Data filters can be hierarchical. One script might generate a data filter and a local data filter. The outer data filter for the data table determines which data is available for the local data filter. To illustrate this point, the following script produces both types of data filters as shown in Figure 9.11.

Tip: The following script is part of a larger script that first builds the arrays required for a marker seg. See the Local Data Filter for Custom Graph.jsl sample script for code that builds those arrays.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
...
New Window( "Marker Seg Example",
   Data Filter Context Box(
      V List Box(
         dt << Data Filter( Local ),
         g = Graph Box(
            Frame Size( 300, 240 ),
            X Scale( Min( xx ) - 5, Max( xx ) + 5 ),
            Y Scale( Min( yy ) - 5, Max( yy ) + 5 ),
            Marker Seg( xx, yy, Row States( dt, rows ) )
         )
      )
   )
);
```

Heights greater than or equal to 55 and less than or equal to 65 are initially filtered. Then the user can work with the local data filter to filter columns on the graph.

Figure 9.11 Data Filter Hierarchy



Export a Report as Interactive HTML with a Live Data Filter

The preceding examples of adding a data filter enabled the `Include` mode. This mode includes rows selected by the local data filter in calculations that are required to build graphs controlled by the local data filter. Deselected rows are excluded from calculations.

For an Interactive HTML export, disable the `Include` mode, because Interactive HTML doesn't exclude rows from calculations.

```
dt = Open( "$SAMPLE_DATA/Quality Control/Coating.jmp" );
dt << Graph Builder(
   Size( 532, 455 ),
   Show Control Panel( 0 ),
   Variables( X( :Sample ), Y( :Weight ) ),
   Elements( Points( X, Y, Legend( 3 ) ), Line( X, Y, Legend( 5 ) ) ),
   Local Data Filter( Mode( Include( 0 ) ), Add Filter( columns( :Weight ) ) )
```

```
);
Current Report() << Save Interactive HTML( "$DESKTOP/example1.html" );
Web( "$DESKTOP/example1.html" );
```

Sort a Data Table

Sort() rearranges the rows of a table according to the values of one or more columns, either replacing the current table or creating a new table with the results. Specify ascending or descending sort for each By column.

```
dt << Sort(("<Private>"|"<Invisible>"), <Replace Table>, By(columns),
  Order(Descending | Ascending), <Output Table Name(name)>)>
```

The following example creates a new data table based on Big Class.jmp that sorts the data in descending order by age and by name:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
sortedDt = dt << Sort(
  By( :age, :name ),
  Order( Descending, Ascending ),
  Output Table Name( "Sorted age name" ) );
```

You can also use an associative array to sort values in a column. The associative array gives you a quick look at unique values in a column. See "Sort a Column's Values in Lexicographic Order".

Stack Values in a Data Table

Stack() combines values from several columns into one column.

```
dt << Stack(
  Columns ( columns ),  // the columns to stack together
  Source Label Column ( "name" ), // to identify source columns
  Stacked Data Column ( "name" ), // name for the new stacked column
  Number of Series( n ), // stacks selected columns into two or more columns
  Contiguous, // specifies that the series consists of adjacent columns
  Keep ( columns ), // the columns to keep in the data table
  Drop ( columns ), // the columns to drop in the data table
  Output Table ( "name" ), // the name for the new data table
  Columns( columns ) ), // specify which columns to include in the stacked
  table
  Copy Formula( Boolean ), // copy the formula into the stacked column
  Drop All Other Columns( Boolean ), // omit the non-stacked columns from the
  new table
  "non-stacked columns"n )( Keep( col1, ... ) // the non-stacked columns to
  keep
```

```
"non-stacked columns"n )( Drop( col1, ... ) // the non-stacked columns to
omit
Suppress Formula Evaluation( Boolean ) // do not evaluate the formula
```

For example, the following script stacks the weight and height columns in Big Class.jmp:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
StackedDt = dt << Stack(
    Columns( :weight, :height ),
    Source Label Column( "ID" ),
    Stacked Data Column( "Y" ),
   "Non-stacked columns"n )( Keep( :age, :sex ) ),
    Output Table( "Stacked Table" )
);
```

The `Columns(columns)` argument can take a list of columns, or an expression that evaluates to a list.

You can also stack multiple sets of columns into multiple columns.

```
dt = Open( "$SAMPLE_DATA/Blood Pressure.jmp" );
dt << Stack(
  Columns(
     :BP 8M,
     :BP 12M,
     :BP 6M,
     :BP 8W,
     :BP 12W,
     :BP 6W,
     :BP 8F,
     :BP 12F,
     :BP 6F
  ),
  Stacked Data Column( "BP" ),
  Source Label Column( "Day" ),
  Number of Series( 3 ), // number of sets in each series
  Contiguous, /* If you want to stack vertically. Otherwise, omit the
  argument. */

);
```

Split Values in a Stacked Data Table

`Split()` breaks a stacked column into several columns.

```
dt << Split(
  Split( columns ),
  // the column to split (required)
```

```
Split by( column ),
// the column to split by (required)

Group(column),
// split data within groups

<Private>|<Invisible>,
// resulting table is private or invisible

Remaining Columns( Keep All | Drop All | Select( columns ) ),
/* specify what to do with the remaining columns in the resulting table
(Keep All by default) */

<Copy formula( 0|1 )>,
/* include column formulas from the source table in the resulting table
(default is 1, true) */

<Suppress formula evaluation( 0|1 )>,
// stop any copied formulas from being evaluated (default is 1, true)

Sort by Column Property,
/* Sort the order of the output columns by the sort column property that is
defined for the Split by column. */

Output Table( "name" ));
// generate the output to the table name specified
```

The following example reverses the previous example for Stack(), returning essentially the original table, except that the height and weight columns now appear in alphabetic order:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Stack(
    Columns( :weight, :height ),
    Source Label Column( "ID" ),
    Stacked Data Column( "Y" ),
   "Non-stacked columns"n( Keep( :age, :sex ) ),
    Output Table( "Stacked Table" )
);
dt2 = Data Table( "Stacked Table" );
dt2 << Split(
/* Split adds the column or columns that you want to split into several
new columns. */
  Split( Y ),
/* Split By adds the column whose values you want to use as the new column
   names, and as the basis for splitting the column. */
  Split By( ID ),
```

```
  Output Table( "Split" ) );
```

When multiple rows are mapped to the same row, only one value is kept.

```
dtabc = New Table( "abc",
 New Column( "Column 1", Character, Set Values( {"a", "a", "b", "b", "c"} )),
 New Column( "Column 2", Character, Set Values( {"x", "y", "x", "y", ""} )),
 New Column( "Column 3", Set Values( [1, 2, 3, 4, 5] ))
);
/* Split adds Column 3. You want to split the values (1, 2, 3..) into several
   new cols.
Split By adds Column 2. You want to use the values (x, y) as the new column
   names. */
dtabc << Split( Split By( :Column 2 ), Split( :Column 3 ));
```

Transpose a Data Table

`Transpose` creates a new data table by flipping a data table on its side, interchanging rows for columns and columns for rows. If you specify no rows, `Transpose()` uses the selected rows. If no rows are selected, it uses all rows.

```
dt << Transpose(
   "private", "invisible",
   columns( columns ),
   Rows( row matrix ),
   By ( column ),
   Label column name( "name" ),
   Output Table( "name" )
);
dt << Transpose(
   Columns( columns ),
   Rows( row matrix ),
   Output Table Name( "name" )
);
```

The following example transposes the height and weight columns in the Big Class.jmp sample data table:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
tranDt = dt << Transpose( Columns( :height, :weight ),
   Output Table Name( "Transposed Columns" ) );
```

Note: The simple transpose command `dt << Transpose` brings up the Transpose window. If you do not want the window to appear, invoke the transpose as `dt << Transpose(no option)`.

Vertically Concatenate Data Tables

Concatenate, also known as a vertical join, combines rows of several data tables top to bottom.

```
dt << Concatenate( DataTableReferences,...,Keep Formulas,
  Output Table Name( "name" ) );
```

For example, if you have subsetted tables for males and females, you can put them back together using Concatenate:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :sex == "M" );
m = dt << Subset( Output Table Name( "M" ) );
dt << Invert Row Selection;
f = dt << Subset( Output Table Name( "F" ) );
both = m << Concatenate( f, Output Table Name( "Both" ) );
```

Or, instead of creating a new table containing all the concatenated data, you can append all the data to the current data table:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :sex == "M" );
m = dt << Subset( Output Table Name( "M" ) );
dt << Invert Row Selection;
f = dt << Subset( Output Table Name( "F" ) );
both = dt << Concatenate( m, f, "Append to First Table" );
```

Horizontally Concatenate Data Tables

Join, also known as horizontal join or concatenate, combines data tables side to side.

```
dt << Join( // message to the first table
  With( dataTable ), // the secondary data table
  Select( columns ), /* select columns from the main table
                      to add to the output table */
  Select With( columns ), /* select columns from the secondary table to
                           add to the output table */
  // join type; alternatives are
  Cartesian join, By Row Number, By matching columns(col1=col2, ...)
  Merge Same Name Columns, // merge columns with the same name
  Copy Formula( 0 ), // on by default; 0 turns it off
  Suppress Formula Evaluation( 0 ), // on by default; 0 turns it off
  Match Flag( 1 ), /* add the Match Flag column from the joined data table
  when you are matching by column */
  Update, /* replace the data in the main table with the corresponding data
  from the secondary table */
  // options for each table:
```

```
Drop Multiples( Boolean, Boolean ), // include all rows from the data table
Include Non Matches( Boolean, Boolean ), // include non-matching rows
Preserve Main Table Order(), // maintain the order of the original data
Output Table Name( "name" )); // the resulting table
```

To try this, first break Big Class.jmp into two parts:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
part1 = dt << Subset(
  Columns( :name, :age, :height ),
  Output Table Name( "Big Class 1" )
);
part2 = dt << Subset(
  Columns( :name, :sex, :weight ),
  Output Table Name( "Big Class 2" )
);
```

To make it a realistic experiment, rearrange the rows in part 2:

```
sortedPart2 = part2 << Sort( By( :name ), Output Table Name( "Sorted_Big
  Class" ) );
```

Now you have a data set in two separate chunks, and the rows are not in the same order, but you can join them together by matching on the column that the two chunks have in common.

```
part1 << Join(
  With( sortedPart2 ),
  By Matching Columns( :name == :name ),
  Preserve Main Table Order();
  Output Table Name( "Joined Parts" );
);
```

The resulting table has two copies of the name variable, one from each part, and you can inspect these to see how `Join` worked. Notice that you now have four Robert rows, because each part had two Robert rows (there were two Roberts in the original table) and `Join` formed all possible combinations.

Tip: To maintain the order of the original data table in the joined table (instead of sorting by the matching columns), include `Preserve Main Table Order()`. This function speeds up the joining process.

See the *JSL Syntax Reference* for more information about the Join arguments.

Virtually Join Data Tables

Virtual Join links a main data table to one or more auxiliary data tables. The feature enables the main data table to access data from the auxiliary data tables without physically joining the tables. See *Using JMP*.

The following example shows how to virtually join data tables and create distributions from the joined data:

```
dt1 = Open( "$SAMPLE_DATA/Pizza Profiles.jmp" );
dt1:ID << Set Property( "Link ID", 1 );// add Link ID and turn it on

dt2 = Open( "$SAMPLE_DATA/Pizza Responses.jmp" );

// add Link Reference to the Choice1, Choice2, and Choice columns
dt2:Choice1 << Set Property( "Link Reference", Reference Table(
  "$SAMPLE_DATA/Pizza Profiles.jmp" ) );
dt2:Choice2 << Set Property( "Link Reference", Reference Table(
  "$SAMPLE_DATA/Pizza Profiles.jmp" ) );
dt2:Choice << Set Property( "Link Reference", Reference Table(
  "$SAMPLE_DATA/Pizza Profiles.jmp" ) );

obj = dt2 << Distribution( // create a distribution of Crust[Choice1]
  Weight( :Subject ),
  Nominal Distribution(
    Column(
      Referenced Column(
        "Crust[Choice1]", // specify the joined column name
        Reference( Column( :Choice1 ), Reference( Column( :Crust ) ) )
      )
    )
  ),
  Nominal Distribution( Column( :Choice1 ) )
);
```

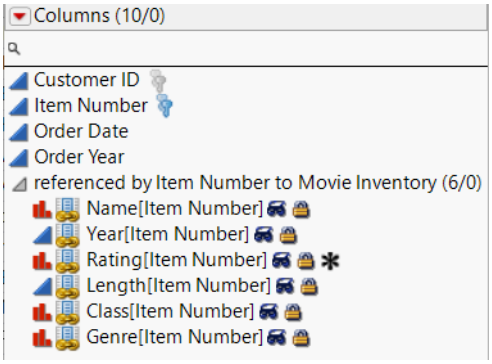
For another example of virtual join, search the Help > Scripting Index for “virtual join”.

Specifying Linked Column Names

The Columns panel in a data table shows you the names of the linked columns. For example, open the Movie Inventory.jmp and Movie Rentals.jmp sample data tables to see the linked columns shown in Figure 9.12.

The column name of the source column is followed by the name of the referencing column in brackets. For example, Figure 9.12 shows the Name source column followed by the referencing column [Item Number] in brackets.

Figure 9.12 Linked Columns in Movie Rentals.jmp



You change the name of the virtually linked column name with `Set Title`. For example, the following script changes the name of the virtually linked column Rating[Item Number] to Movie Rating in the Distribution report.

```
dt1 = Open( "$SAMPLE_DATA/Movie Inventory.jmp" );
dt2 = Open( "$SAMPLE_DATA/Movie Rentals.jmp" );
dt2 << Distribution(
   Nominal Distribution(
      Column(
         Referenced Column( // specify the joined column name
            "Rating[Item Number]",
            Reference( Column( :Item Number ), Reference( Column( :Rating ) ) )
         )
      )
   ),
   SendToReport( Dispatch( {}, "Rating[Item Number]", OutlineBox,
      {Set Title( "Movie Rating" )} ) ) /* change the name of the
      virtually linked column */
);
```

Note that running the script automatically opens the referenced table (the one with the Link ID property), just like running any other platform script, before the referenced table is opened.

You can also specify that the referenced column name be used in linked columns. This option lets you define a shorter column name (for example, Rating instead of Rating[Item Number]). Include the `Use Linked Column Name` option:

```
dt1 = Open( "$SAMPLE_DATA/Movie Inventory.jmp" );
dt2 = Open( "$SAMPLE_DATA/Movie Rentals.jmp" );

Column( dt2, "Item Number" ) << Set Property(
   "Link Reference",
   {Reference Table( "Movie Inventory.jmp" ), Options( "Use Linked Column
   Name"( 1 ) )}
```

```
);
dt2 << Distribution(
  Nominal Distribution(
     Column(
        Referenced Column( // specify the joined column name
           "Rating",
           Reference( Column( :Item Number ), Reference( Column( :Rating ) ) )
        )
     )
  )
);
```

Merging Referenced Data into One Data Table

To share the data from virtually joined data tables with another user, you might want to merge the data so that the data tables are permanently joined. This option is helpful because you provide one data table, not the main and auxiliary data tables.

When you merge referenced data, the data are replicated from the auxiliary data table to the main data table. When you save the main data table, the actual data are also saved.

The following example shows how to merge two referenced columns:

```
dt1 = Open( "$SAMPLE_DATA/Pizza Profiles.jmp" );
dt2 = Open( "$SAMPLE_DATA/Pizza Responses.jmp" );

dt1:ID << Set Property( "Link ID", 1 );

dt2:Choice << Set Property( "Link Reference", Reference Table( dt1 ) );
dt2:Choice1 << Set Property( "Link Reference", Reference Table( dt1 ) );
dt2:Choice2 << Set Property( "Link Reference", Reference Table( dt1 ) );

// select the columns to merge
dt2 << Select Columns( :Choice1, :Choice2 );

dt2 << Merge Referenced Data();
/* linked data from the referenced columns are now embedded in
Pizza Responses.jmp */
```

Note: If you don't select the columns to merge, all columns are merged.

Replace Data in Data Tables

Note: `Merge Update()` is an alias for `Update()`.

`Update()` replaces data in one table with data from a second table.

```
dt << Update( // message to first table
  With( dataTable ), // the other data table
  By Row Number, /* default join type; alternative is
                  By Matching Columns( col1 == col2 ) */
  Ignore Missing, // optional, does not replace values with missing values
  );
```

To try this, make a subset of Big Class.jmp:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
NewHt = dt << Subset( Columns( :name, :height), Output Table Name( "hts" ) );
```

Next, add 0-6 inches to each student's height:

```
diff = Random Uniform( 0, 6 );
For Each Row( NewHt, :height = :height + diff );
```

Finally, update the heights of students in Big Class.jmp with the new heights from the subset table:

```
dt << Update(
  With( NewHt ),
  By Matching Columns( :name == :name),
  );
```

Controlling the Columns Added to an Updated Table

Your updated table might contain more columns than your original table. You can select which columns are included in your updated table using the option `Add Column from Update Table()`.

To add no additional columns:

```
Data Table( "table" ) << Update(
  With( Data Table( "update data" ) ),
  Match Columns( :ID = :ID ),
  Add Columns from Update Table( None )
);
```

To add some columns:

```
Data Table( "table" ) << Update(
  With( Data Table( "update data" ) ),
  Match Columns( :ID = :ID ),
  Add Columns from Update table( :col1, :col2, :col3 )
);
```

Create a Table Using Tabulate

Tabulate constructs tables of descriptive statistics. The tables are built from grouping columns, analysis columns, and statistics keywords. The following example creates a table containing the standard deviation and mean for the height and weight of male and female students:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Tabulate( // message to data table
  Add Table( // start a new table
     Column Table( Grouping Columns( :sex ) ), // group using the column sex
     Row Table( // add rows to the table

        // use the height and weight columns for the analysis
        Analysis Columns( :height, :weight ),

        // show the standard deviation and mean
        Statistics( Std Dev, Mean )
     )
  )
);
```

You can apply a transformation to a column in the Tabulate table and set the format of the column at the same time. Use the `Transform Column()` function inside `Analysis Column()`. For example, the following script applies the Log transformation to height and sets the column format for the Mean and % of Total columns.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Tabulate(
  Set Format(
     Mean(
        :height( 10, 1 ),
        Analysis Column(
           Transform Column(
              "Log[height]",
              Formula( Log( :height ) )
           ),
           Format( 10, "Best" )
        )
     ),
     "% of Total"n(:height( 12, 2 ),
     Analysis Column(
        Transform Column(
           "Log[height]",
           Formula( Log( :height ) )
        ),
        Format( 12, 2 )
     ))
```

```
    ),
    Add Table(
       Column Table(
          Analysis Columns(
             :height,
             Transform Column(
                "Log[height]",
                Formula( Log( :height ) )
             )
          ),
          Statistics( Mean, "% of Total"n )
       ),
       Row Table( Grouping Columns( :sex ) )
    )
);
```

Note: You can also change the width of columns and perform additional operations within Tabulate. See the JMP Scripting Index in the Help menu.

Copy Data Tables

The most efficient way to copy a data table is with the `Subset()` function. The following example subsets all rows and columns into New Table.jmp.

```
sourceDT = Open( "$SAMPLE_DATA/Big Class.jmp" );
destDT = sourceDT << Subset( Columns(), Rows(), Output Table Name( "New
  Table.jmp" ) );
```

You can also use `Copy File()`.

```
Copy File( "$SAMPLE_DATA/Big Class.jmp", "c:/MyData/New Table.jmp" );
```

`Copy File()` is equivalent to copying the file on your computer. However, you have to consider whether the destination data table with the specified name already exists. And the source file has to be saved. If the source file is not saved, or the destination file already exists, the copy fails. It's safer to use the `Subset()` method shown above.

Find Missing Data Patterns

If your data table contains a lot of missing data in different columns, you might want to see whether there is a pattern in the distribution of the missing data relative to the different columns.

```
dt << Missing Data Pattern( // message to data table
  Columns( :miss ), // find missing data in this column
  Output Table( "Missing Data Pattern" ) // name the output table
```

```
);
```

Compare Data Tables

JMP can compare two open data tables and report the differences between data, scripts, table variables, column names, column properties, and column attributes. To compare data tables using JSL:

```
obj = dt << Compare Data Tables(
Compare With( Data Table ( "Data Table Name" ) );
```

For example, to compare the data tables Students1.jmp and Students2.jmp, proceed as follows:

```
dt = Open( "$SAMPLE_DATA/Students1.jmp" );
dt2 = Open( "$SAMPLE_DATA/Students2.jmp" );
obj = dt << Compare Data Tables( Compare With( Data Table( "Students2" ) ) );
```

To view the difference summary matrix for the comparison results, use this function:

```
mtx = ( obj << Get Difference Summary Matrix );
```

The resulting matrix appears in the log window:

```
[-1 1 2 2,
-1 2 4 3,
-1 1 7 4,
1 3 8 4,
1 1 10 9,
0 1 11 11,
-1 1 14 14,
-1 1 16 15,
1 1 18 16,
-1 1 19 18,
0 1 22 20,
0 1 26 24,
1 1 29 27,
1 1 34 33]
```

In the comparison summary matrix in the example, the first column represents the action performed on a column: -1 for a Delete, 0 for a Replace, and 1 for an Add. The second column represents the number of rows affected by the action. The third and fourth columns represent the row numbers affected in the two data tables in their respective order (Students1.jmp, Students2.jmp).

Create a Summary Table

A Summary table includes summary statistics for a data table, such as the mean, median, standard deviation, minimum value, and maximum value. By default, the summary table is linked to its source table. When you select rows in the summary table, the corresponding rows are highlighted in its source table.

```
dt << Summary( <private>, <invisible>, <Group( column )>, <Weight( column )>,
   <Freq( column )>, <N>, <Mean( column )>, <Std Dev( column )>, <Min( column
   )>, <Max( column )>...
```

Include columns for any statistic listed in the Summary launch window.

The following example creates a summary table from the Fitness.jmp sample data table. The N Rows column, which specifies the number of rows for each level, is included in the Summary table by default. This script creates a column of the mean for Runtime by Sex.

```
dt = Open( "$SAMPLE_DATA/Fitness.jmp" );
dt << Summary(
       Group( :Sex ),
       Mean( :Runtime ),
);
```

You can override the preselected Freq or Weight columns in the data table by specifying the Freq or Weight columns in the script. To omit the preselected Freq or Weight, use the keyword "none".

```
dt = Open( "$SAMPLE_DATA/Fitness.jmp" );
dt << Summary(
       Group( :Sex ),
       Mean( :Runtime ),
       Weight( "none" )
);
```

To unlink the Summary table, include `Link to Original Data Table( 0 )` in the script.

```
dt = Open( "$SAMPLE_DATA/Fitness.jmp" );
dt << Summary(
     Group( :Sex ),
     Mean( :Runtime ),
     Link to Original Data Table( 0 )
);
```

Subscribe to a Data Table

If you want to receive a message when a data table changes, use the `Subscribe` message. For example, you might want a message sent to the log when columns are added or deleted.

Here is the basic syntax:

```
dt << Subscribe( keyname(<client>), On Delete Columns|On Add Columns|On Add
   Rows|On Delete Rows|On Rename Column|On Close|On Save|On Rename(function)
   );
```

The first argument is a name for the subscription (or “client”), so that it can also be removed.

Note: The subscription is to the data table. The subscriber subscribes to the broadcast from the data table; it does not subscribe to a program or to script.

The application can also subscribe as a client to the data table (for example, most built-in platforms such as Distribution). If a data table has clients when the data table is closed, the user is warned that there are applications open that might need the data table.

Each subscription remains in effect until you unsubscribe:

```
dt << Unsubscribe("keyname", On Delete Columns | On Add Columns | On Add Rows
   | On Delete Rows | On Close | On Col Rename | All);
```

The following example sends messages to the log when a row is added or deleted:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
delRowsFn = Function( {a, b, rows},
   dtname = ( a << Get Name());
   Print( dtname );
   Print( b );
   Print Matrix( rows );
);
addRowsFn = Function( {a, b, insert},
   dtname = ( a << Get Name() );
   Print( dtname );
   Print( b );
   Print( insert );
);
dt << Subscribe( "Test Delete", onDeleteRows( delRowsFn, 3 ) );
dt << Subscribe( "Test Add", onAddRows( addRowsFn, 3 ) );
```

`Subscribe to Data Table List()` subscribes to a list of data tables to be notified when a new data table has been added, closed, or renamed. The following example shows how to subscribe and unsubscribe:

```
f2 = Function( {x}, Show( x ) );
Show( Subscribe to Data Table List( "My Data", OnClose( f2 ) ) );
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Wait( 2 );
Close( dt );
Show( Unsubscribe to Data Table List( "My Data" ) );
   Subscribe to Data Table List( "My Data", OnClose( f2 ) ) = "My Data";
   x = Data Table( "Big Class" );
   Unsubscribe to Data Table List( "My Data" ) = Empty();
```

Empty Application Names

If `Subscribe` is called with an empty application name, JMP generates a unique name that is returned to the caller. In the following example, `appname2` is subscribed to the data table as a client.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
appname1 = dt << Subscribe( "", On Close( Print( "Closing Data Table" ) ) );
appname2 = dt << Subscribe(
   ""( "client" ),
   On Close(
      Function( {dtab},
         dtname = ( dtab << Get Name() );
         Print( dtname );
      )
   )
);
dt << Unsubscribe( appname1, On Close );
dt << Run Script( "Distribution" )
```

Move Data Between Matrices and Data Tables

For information about moving information between a matrix and a data table, see "Matrices and Data Tables".

Add Hyperlinks to Text Columns

You can define a text column as a column of hyperlinks using the Event Handler column property. For example, you can create a column of geographical data as hyperlinks to a Google map, or you might want to include links to a Wikipedia web pages.

Note: To learn more about the Event Handler column property, add the column property to a column that contains a hyperlink and study the property in the JMP interface. Tips about the property are provided. To see the script that creates the property, select the data table red triangle menu; select **Copy Script (No Data)**; and copy the script to an empty script.

Use the Event Handler column property to build hyperlinks based on one or more columns in your data table. This example shows how the hyperlinks in the SAS Offices.jmp sample data table were created.

1. Select **Help > Sample Data Library** and open SAS Offices.jmp.
2. Right-click the City column and select **Column Info**.

   Notice that the Event Handler column property has been added.
3. The script on the Click tab inserts the code for the URL.

```
Function( {thisTable, thisColumn, iRow},
Web(
   "https://www.google.com/maps/@"||Char(thisTable:latitude[irow])||","||Char
   (thisTable:longitude[irow])||",12z" ) );
```

- `thisTable` is the data table object.
- `thisColumn` is the column object.
- `iRow` is the index of the row in `thisColumn`.
- `Web` opens the URL in the default browser.
- The `@` sign in "https://www.google.com/maps/@" is specific to the Google API for maps. Google expects coordinates to follow it.
- `||Char(thisTable:latitude[irow])` concatenates the latitude defined in the Latitude column.
- `||","||Char(thisTable:longitude[irow])||` concatenates the longitude defined in the Longitude column.
- The Google API uses `",12z"` to specify a zoom factor.

- The script on the Tip tab defines the tooltip that appears when you hover over the hyperlink. The tip accurately describes what happens after the link is clicked. Consider using this function to set the tooltip to the URL the Click script uses or a description of the platform.

  ```
  Function( {thisTable, thisColumn, iRow},
  "Open " ||
     "https://www.google.com/maps/@"||char(thisTable:latitude[irow])||","||char
     (thisTable:longitude[irow])||",12z" || " in your browser.";);
  ```

  - Returns the tool tip string.
- The script on the Color tab defines the color of the hyperlink (in this example, blue).

  ```
  Function( {thisTable, thisColumn, iRow}, 5;);
  ```

  See "JMP Colors" for a diagram of JMP colors.

You might also want to embed links to other files in a data table column. The following example shows how to use Event Handler scripts and set the values to the audio and video locations on your computer.

```
New Table( "Wildlife",
   New Column( "Column 1", Character, "Nominal",
      Set Property( "Event Handler",
         Event Handler(
            Click( JSL Quote(Function( {thisTable, thisColumn, iRow}, Open(
   Char(
thisTable:thisColumn[ iRow ] ) ); );) ),
            Tip( JSL Quote(Function( {thisTable, thisColumn, iRow}, "Open " ||
   Char(
```

```
thisTable:thisColumn[ iRow ] ) || " in a media player."; );) ),
           Color( JSL Quote(Function( {thisTable, thisColumn, iRow}, 5; );) )
        )
     ),
     Set Selected,
     Set Values( {
"C:\Users\Public\Videos\Sample Videos\Wildlife.wmv",
"C:\Users\Public\Music\Sample Music\Kalimba.mp3"
     } )
  )
)
```

The SAS Country Office and Photo columns in SAS Offices.jmp also contain examples of Event Handler scripts.

Columns

This section covers actions that you can perform on columns in a data table, such as creating, grouping, setting and getting attributes and properties, and so on.

- "Send Messages to Data Column Objects"
- "Create Columns"
- "Create a New Column by Text Matching"
- "Add Several Columns at Once"
- "Group Columns"
- "Use Set Selected to Select Columns"
- "Rearrange and Move Columns"
- "Recode Columns"
- "Add a Column Switcher"
- "Compress Selected Columns"
- "Create Custom Functions, Transforms, and Formats"
- "Custom Transform Example"
- "Delete Columns"
- "Get Column Names"
- "Column Attributes"
- "Column Properties"

Note: JMP can display math symbols and Greek letters (controlled in Preferences in the Font category). This means that if you save a column (such as T square limits), the column name could either be "T Square Limits" (no special characters) or "T^2 Limits" (with special characters). Any reference using the column name must match the name exactly, or it fails.

Send Messages to Data Column Objects

Just as you send data table messages to a data table reference, you can send column messages to a reference to a data column object. The `Column` function returns a data column reference. Its argument is either a name in quotation marks, something that evaluates to a name in quotation marks, or a number.

```
Column( "age" );       // a reference to the age column
col = Column( 2 );     // assign a reference to the second column
```

Note: You can also use the syntax `"Profits ($M)"n` to evaluate to a name that contains special characters.

The *Scripting Guide* uses *col* to represent data column references. To see the messages that you can send to data column objects, see the Scripting Index (**Help > Scripting Index > Objects > Data Tables > Column Scripting**). Alternatively, you can use the `Show Properties()` command:

```
Show Properties( col );
```

Note: With column references, you must include a subscript to refer to the individual data values. Without a subscript, the reference is to the column object as a whole.

Once you have stored a data column reference in a global variable, to modify columns, you send messages to the data column reference.

Sending messages to columns is comparable to sending messages to data tables. Either state the object, a double-angle operator <<, and then the message with its arguments in parentheses, or use the `Send()` function with the object and then the message. In some cases the messages themselves need no arguments, so the trailing parentheses are optional.

```
col << message( arg, arg2, ... );
Send( col, message(arg, arg2, ... ) );
```

As with data tables and other types of objects, you can stack or list messages:

```
col << message << message2 << ...
col << {message, message2, ...};
```

Tip: To delete a column, you must send the message to a data *table* reference, because objects cannot delete themselves, only their containers can delete them.

Access Cell Values through Column References

Always use a subscript on a column reference to access the values in the cells of columns. Without a subscript, the reference is to the column object as a whole.

```
x = col[irow]    // specific row
x = col[]        // current row
col[irow] = 2;   // as an l-value for assignment
dt << Select Where( col[] < 14 );  // in a WHERE clause
```

Access Cell Values that have Value Labels

If the column whose values you want to access has a Value Label property, you might want to access the value label (formatted value) instead of the actual data value. To do this, use the formatted option. Consider the following example:

```
dt = Open( "$SAMPLE_DATA\Big Class.jmp" );
dt << Run Script( "Set Age Value Labels" );
x = Column( dt, "age", formatted )[1];
     Show( x );
y = Column( dt, "age" )[1];
     Show( y );
```

The log shows the formatted value for *x*, which is `Twelve`, and the actual data value for *y*, which is `12`.

Create Columns

To add a new column to a data table, send a `New Column` message to a data table reference or use the `New Column()` function. The first argument, the column's name, is required. Either enclose the name in quotation marks or specify an expression evaluating to the name.

```
dt = New Table( "MyData1.jmp" );
dt << New Column( "wafer" );
```

or

```
dt = New Table( "MyData2.jmp" );
a = "wafer";
dt << New Column( a );
```

or

```
dt = New Table( "MyData3.jmp" );
col = New Column( "X", Formula( Random Uniform() ) );
```

If the table already includes a column by the same name, a sequential numeric value is appended to the new column name (wafer, wafer 2, wafer 3, and so on).

You can specify the following properties when creating a new column:

- Data type (numeric, character, row state, or expression)
- Modeling type (continuous, nominal, ordinal, multiple response, unstructured text, or vector)
- Column width (only for numeric columns)
- Numeric format

Unless otherwise specified, columns are numeric, continuous, and 12 characters wide.

The following example creates a new column whose data type is numeric, modeling type is continuous, width is set to 5, and numeric format is set to Best:

```
dt << New Column( "wafer", Numeric, "Continuous", Format( "Best", 5 ) );
```

The following example creates a character column and automatically assigns the nominal modeling type.

```
dt << New Column( "Last Name", Character );
```

You can also add formulas and other script messages appropriate to the column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Ratio", Numeric, "Continuous", Formula( :height/:weight )
  );
dt << New Column( "myMarkers",
   Row State,
   Set Formula( Marker State( age - 12 ) )
);
```

You can also create new columns that have the same properties as an existing column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Like Name", Like( :name ) );
```

If you plan to work with specific columns later (for example, grouping or changing data types), create a column reference:

```
myCol = dt << New Column( "Birth Date" );
```

Fill a Column with Data

To fill the column with data, use `Values` or its equivalent, `Set Values`. Include the value for each cell in a list.

The following example adds a new column called Last Name to a new data table. Three values are added: Smith, Jones, and Anderson.

```
dt = New Table( "My Data");
dt << New Column( "Last Name", Character, Values( {"Smith", "Jones",
   "Anderson"} ) );
```

The column can also be filled with numeric values. The following example adds a new column called Row Number to a new data table. The function N Row returns the number of rows in the table, and numeric values populate all of the rows in the column, beginning with 1.

```
dt = New Table( "My Data");
dt << New Column( "Last Name", Character, Values( {"Smith", "Jones",
   "Anderson"} ) );
dt << New Column( "Row Number",
   Numeric,
   Values( 1 :: N Row() ),
   Format( "Best", 5 )
);
```

To add a column of random numbers, insert a formula along with the random function.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Random", Numeric, Formula( Random Uniform() ) );
```

Another option is to add a constant numeric value to each cell. In the following example, the number 5 is added to each cell.

```
dt << New Column( "Number" );
:Number << Set Each Value( 5 );
```

New Column() can also be used as a built-in function, in other words, without the << (Send) command applied to the data table reference. When used in this way, the column is added to the current data table.

```
dt << New Column( "Address" ); // command is sent to the data table reference
New Column( "Address" ); // column is added to the current data table
```

Create a New Column by Text Matching

The New Column by Text Matching option enables you to use regular expressions (which identify patterns in text) to match text in a column and then create a new column from the matched text. This feature is an alternative to using the Text Explorer platform to write regular expressions and then save the results to a column.

The following example shows how to select *cat*, *cats*, *dog*, and *dogs* in a column and put the values in new columns.

```
dt = Open( "$SAMPLE_DATA/Pet Survey.jmp" );
dt << New Column by Text Matching(
```

```
   Column( :Survey Response ),
   Set Regex(
      Custom(
         Title( "Cat" ),
         /* finds "cat", optionally followed by "s",
            with either a space or period afterward. */
         Regex( "(cat(?:s?))([\s\.])" ),
         Result( "\[\1]\" ), // results in the value matched by (cat...)
      ),
   ),
   Output Column Name( "Instances of Cats" )
);
dt << New Column by Text Matching(
   Column( :Survey Response ),
   Set Regex(
      Custom(
         Title( "Dog" ),
         Regex( "(dog(?:s?))([\s\.])" ),
         Result( "\[\1]\" ),
      ),
   ),
   Output Column Name( "Instances of Dogs" )
);
```

For more information about using regular expressions in JSL, see "Regular Expressions". To learn more about regular expressions, visit Regular-Expressions.info.

Add Several Columns at Once

The `Add Multiple Columns` message creates several columns at once. Arguments include the column name prefix, the number of columns, where to insert them (`Before First`, `After Last`, or `After(col)`), and the data type (`Numeric`, `Row State`, or `Character(width)`). An additional argument, field width, is optional for a numeric column.

```
// add two row state columns named "beginning <n>" before the first column
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Add Multiple Columns( "beginning", 2, Before First, Row State );

// add three numeric columns named "middle <n>" after the height column
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Add Multiple Columns( "middle", 3, After( :height ), Numeric );

// add four character columns named "end <n>" after the last column
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Add Multiple Columns( "end", 4, After Last, Character( 4 ) );
```

The column name is a prefix. If multiple columns with the same name are added, a sequential number is appended to each column name (beginning 1, beginning 2, and so on).

Group Columns

You can group columns by sending the data table the `Group Columns` message, which takes a list of columns to group as an argument. For example, the following code opens the Big Class.jmp sample data table and groups the age and sex columns.

```
dt = Open( "$SAMPLE_DATA\Big Class.jmp" );
dt << Group Columns( {:age, :sex} );
   "age etc."
```

You can also send a column name followed by the number of columns to include in the group. The group includes the first column named and the *n*-1 columns that follow it. This line is equivalent to the line above, which groups age and sex:

```
dt << Group Columns( :age, 2 );
```

The group name is based on the first column specified in the argument. So in the preceding example, the group is automatically named age etc. To customize the name, include the group name as the first argument.

```
dt << Group Columns( "My group", :age, 2 );
```

To ungroup grouped columns, use the `Ungroup Columns` message, which takes a list of columns in a group as an argument. For example, the following line ungroups the two columns grouped in the previous example.

```
dt << Ungroup Columns( {:age, :sex} );
```

Note the following about grouping and ungrouping columns:

- Both messages take a single list as an argument. The list must be enclosed in braces.
- You cannot create more than one group in a single message (for example, by giving the `Group Columns` message two lists of columns). Instead, you must send the data table two separate `Group Columns` messages.
- The `Ungroup Columns` message takes a list of columns to ungroup, not the name of a group of columns. You can remove a partial list of columns from a group. For example, this line creates a group of four columns:

  ```
  dt << Group Columns( {:age, :sex, :height, :weight} );
  ```

  And this line removes two of the columns, while leaving the other two in the group:

  ```
  dt << Ungroup Columns( {:age, :sex} );
  ```

  Notice that the column group's name does not change automatically. You have to change the name.

Get Column Groups or Names

To get column groups or names, use `Get Column Group` or `Get Column Groups Names`. The first part of this example creates two sets of grouped columns as shown in the previous section. `Get Column Group()` returns the column names in the specific group. `Get Column Groups Names` returns the names of the column groups.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Group Columns( {:age, :sex} );
   "age etc."
dt << Group Columns( {:weight, :height} );
   "height etc."
dt << Get Column Group( "age etc." );
   {:age, :sex}
dt << Get Column Groups Names();
   {"age etc.", "height etc."}
```

Select a Column Group

To select or deselect a column group, use the following messages:

```
dt << Select Column Group( "name" ); // select the columns in the group
dt << Deselect Column Group( "name" ); // deselect the columns in the group
```

The following example groups columns X and Y and groups columns Ozone through Lead, and then selects those groups in the data table:

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
dt << Group Columns( "xy", {:X, :Y} );
dt << Group Columns(
   "pollutants",
   :Ozone :: :Lead
);
dt << Select Column Group( "xy", "pollutants" );
```

Rename a Column Group

To rename a column group, use the following message:

```
dt << Rename Column Group( "name", "toname" );
```

The following example renames the column group from xy to coordinates:

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
dt << Group Columns( "xy", {:X, :Y} );
dt << Group Columns(
   "pollutants",
   :Ozone :: :Lead
);
Wait( 3 );
```

```
dt << Rename Column Group( "xy", "coordinates" );
```

Move a Column Group

To move a column group, use the following message:

```
dt << Move Column Group (To First | To Last | After(column) | After(group));
```

The following example moves the columns in the pollutants group to the end:

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
dt << Group Columns( "xy", {:X, :Y} );
dt << Group Columns(
   "pollutants",
   :Ozone :: :Lead
);
dt << Move Column Group( To Last, "pollutants" );
```

Use Set Selected to Select Columns

Send the `Set Selected` message to select a column.

```
col << Set Selected( 1 );
```

To select all columns with a specific string in the name, run the following script:

```
dt = Open( "$SAMPLE_DATA/Semiconductor Capability.jmp" );

colList = dt << Get Column Names( String );

For( i = 1, i <= N Cols( dt ), i++,
   If( Contains( colList[i], "NPN" ), /* select only columns with NPN
   in the name */
      Column( colList[i] ) << Set Selected( 1 )
   )
);
```

If you want to select many columns, consider sending the `Select Columns` message to the data table. See the *Scripting Guide*.

Get Selected Columns

To get a list of currently selected columns, use the `Get Selected Columns` message.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
:age << Set Selected( 1 );
:sex << Set Selected( 1 );
dt << Get Selected Columns();
   {:age, :sex}
```

Return the list of selected columns as a string using the `string` argument:

```
dt << Get Selected Columns( string );
   {"age", "sex", "height"}
```

Once you know what columns are selected, you can then write a script that acts upon these columns. Or your script can iteratively select columns and act upon them one at a time.

To actually select the columns before getting the columns, send the `Set Selected` message to a column. See "Column Attributes".

Go To Column

To select and move to a specific column, use the `Go To` message.

```
dt << Go To( column name | column number );
```

For data tables with many columns, you can use this message to scroll the data table all the way to the left, so that the first column comes into view and is selected:

```
dt = Open( "$SAMPLE_DATA/Tiretread.jmp" );
dt << Go To( 1 );
```

Invert Column Selections

To select any column that is currently deselected and deselect any column that is currently selected, use the `Invert Column Selection` message.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt:age << Set Selected( 1 );
dt:height << Set Selected( 1 );
Wait( 1 );
b = dt << Invert Column Selection;
```

If a list of columns is provided, the columns that are not in the list are selected.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
a = {:height, :weight};
b = dt << Invert Column Selection( a );
```

Rearrange and Move Columns

These messages enable you to rearrange columns in a data table:

```
dt << Reorder By Name;           // alphanumeric order
dt << Reverse Order;             // reverse current order
dt << Reorder By Data Type;      // row state, character, and then numeric
dt << Reorder By Modeling Type;  // continuous, ordinal, nominal
dt << Original Order;            // saved order
```

These commands move the currently selected column or columns to the indicated destination point.

```
dt << Move Selected Column( to first );
dt << Move Selected Column( to last );
dt << Move Selected Column( after( "name" ) );
```

Note that `Move Selected Column` is an alias for `Move Selected Columns`.

You can also move columns without first selecting them in the data table by using the following syntax.

```
dt << Move Selected Columns( {"name"}, to first );
dt << Move Selected Columns( {"name"}, to last );
dt << Move Selected Columns( {"name"}, after( "name" ) );
```

You may use a list to specify the column names.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
list = {"name", "sex"};
dt << Move Selected Columns( list, To Last );
```

Recode Columns

Use `Recode Column` to change all of the values in a column at once. Specify the original column, or reference column, a list of transformations, and a target column. The function applies the transformations to the reference column, updating the target column. If the `Update Properties` argument is turned on, the following column properties are updated when a column is recoded: Value Labels, Value Scores, Value Order, Value Colors, Supercategories, Coding, and Missing Value Codes.

```
obj << Recode Column(<column reference>, {<transform>, ...}, <Update
  Properties(0|1)>, Target Column(<column reference> | <column name>))
```

For example, in Consumer Preferences.jmp, change the coding of the Gender column to 0|1 instead of 1|2, but keep the value labels the same.

```
dt = Open( "$SAMPLE_DATA/Consumer Preferences.jmp" );
col = New Column( :Gender );
dt << Recode Column(
   :Gender,
   {if(
      _rcNow == 1, 0,
      1,
   )},
   Update Properties( 1 ),
   Target Column( col )
);
```

You can also make the transformations to the original column. Recode the age column to age groups in Big Class.jmp.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Recode Column(
   :age,
   {if(
      _rcNow >= 17, "Older",
      _rcNow >= 15, "Middle",
      "Younger"
   )},
   Target Column( :age)
);
```

Add a Column Switcher

Use JSL to add a Column Switcher panel to a report window. The Column Switcher lets you quickly analyze different variables without having to re-create your analysis.

```
// add a Column Switcher to the report
obj << Column Switcher(<default_col>(<col1>, <col2>, ...);

// remove the Column Switcher from a report
obj << Remove Column Switcher();
```

For example, add the Column Switcher to a Contingency report for the Car Poll.jmp data:

```
dt = Open( "$SAMPLE_DATA/Car Poll.jmp" );
obj = dt << Contingency(
   Y( :size ),
   X( :marital status )
);
ColumnSwitcherObject = obj << Column Switcher(
   :marital status,
   {:sex, :country, :marital status}
);
// number of pixels for the switcher width
ColumnSwitcherObject << Set Size( 200 );

// number of columns to display in the switcher
ColumnSwitcherObject << Set NLines( 6 );
```

You can also share a Column Switcher between two platform reports:

```
dt = Open( "$SAMPLE_DATA/Nicardipine.jmp" );
New Window( "Column Switcher Example",
   H List Box(
      sexCS = dt << Column Switcher( :Sex, {:Sex, :Causality, :Race} ),
```

```
        nDist = Distribution( Nominal Distribution( Column( :Sex ) ) ),
        bar = Graph Builder(
           Show Control Panel( 0 ),
           Variables( X( :Sex ), Y( :Age ) ),
           Elements( Box Plot( X, Y, Legend( 5 ) ) )
        )
     )
  );

  sexCS << Link Platform( nDist );
  sexCS << Link Platform( bar );
```

Note: If you are publishing reports that share a Column Switcher to JMP Public or JMP Live, add them to a dashboard instead of using JSL. See *Using JMP*.

For more information about the Column Switcher, see *Using JMP*.

Compress Selected Columns

To minimize the size of a large data table, use the `Compress Selected Columns` message. Each column is compressed into the most compact form.

```
dt << Compress Selected Columns( {column name, column name} );
```

For example, compress the age, sex, height, and weight columns in Big Class.jmp:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Compress Selected Columns(
{:age, :sex, :height, :weight}
);
```

For more information about this feature, see *Using JMP*.

Create Custom Functions, Transforms, and Formats

You can create user-defined custom functions and then use them in the Formula Editor's function list, in JSL scripting, as custom formats, and as custom transforms. Note that these custom functions can only be used in the context of the current JMP session. Restarting JMP removes previously defined custom functions. They must be defined and added for each session. An easy way to do this is to define your functions in an add-in or a startup script that runs each time you start JMP.

To define a custom function, use `New Custom Function()`. After a function is defined, you can add or activate it in the current running instance of JMP using `Add Custom Functions()`. Similarly, you can remove or deactivate a function using `Remove Custom Functions()`.

`New Custom Function()` requires three parameters: namespace, functionName, functionDefinition. Additional messages may be passed to define the following:

- Formula Editor values: Formula category, Result type, Parameters
- Custom transform values: Transform category
- Custom format values: Custom format category
- Scripting Index values: Description, Prototype, Scripting Index category, Examples

Tip: If you receive a data table that uses a custom transform, function, or format, you'll see "###" in the column. The data table author must give you a script to add that custom item to your computer. Run the script and then reopen the data table.

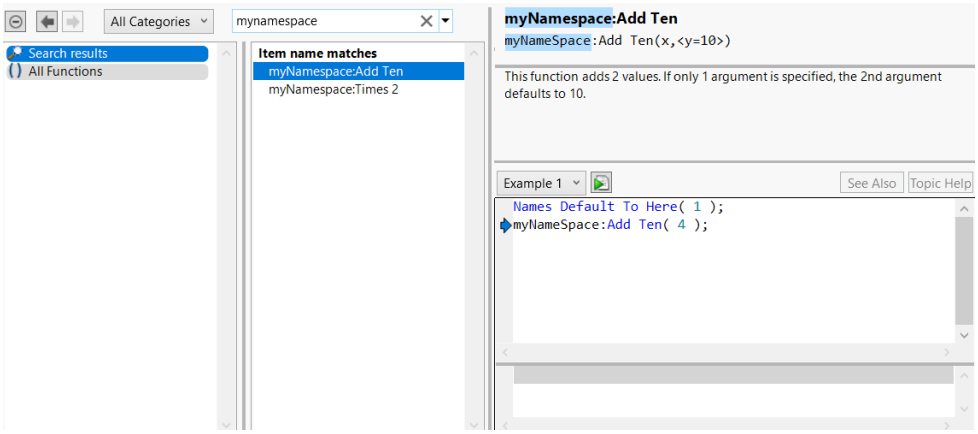
Custom Function Example

Let's create a completely specified function (including Scripting Index information and hint text). We will define the function first, then activate it.

```
funcDef = Function( {x, y = 10}, x + y );
description = "This function adds 2 values. If only 1 argument is specified,
   the 2nd argument defaults to 10.";

// examples can be multiple statements inside expr()
ex2 = Expr(
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   myNameSpace:Add Ten( :age[4], 20 );
);
parmHint1 = "x value";
parmHint2 = "<y=10>";

/* create function using variables
the commands can also be accepted as arguments in New Custom Function */
newAdd = New Custom Function( "myNamespace", "Add Ten", funcDef );
newAdd << Description( description );
newAdd << Prototype( "myNameSpace:Add Ten(x,<y=10>)" );
newAdd << Example( "myNameSpace:Add Ten(4)" );

// expr() Examples need to be passed inside NameExpr()
newAdd << Example( Name Expr( ex2 ) );
newAdd << Parameter( "Number", parmHint1 );
newAdd << Parameter( "Number", parmHint2 );
newAdd << Formula Category( "NumberStuff" );

// add myNamespace:Add Ten function to system
Add Custom Functions( newAdd );
```

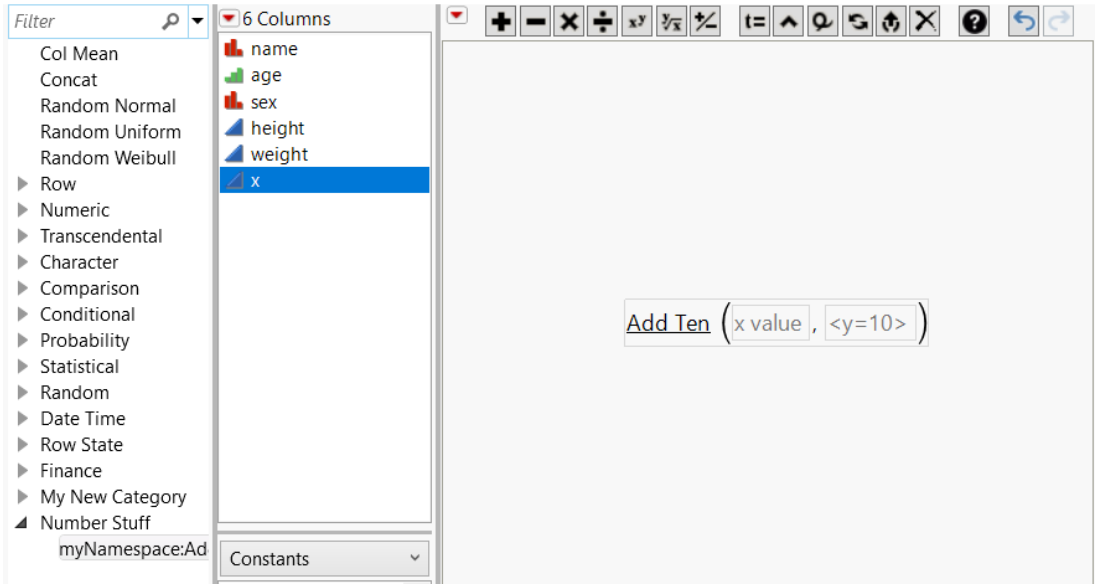
Running the above example will write “Deploying function: myNamespace:Add Ten” in the JMP log. Search for this function in the Scripting Index to see the results of using the Description, Prototype, and Example messages.

Figure 9.13 Custom Function in the Scripting Index



The `Formula Category` message results in a new category in the Formula Editor’s list of formula categories. Custom functions are shown with an underline in the Formula Editor. Also see the hint text from the `Parameter` messages.

Figure 9.14 Custom Function in the Formula Editor



Delete a Custom Function Example

To delete a custom function, use `Remove Custom Functions( {"name"} )`.

```
Remove Custom Functions( {"myNamespace:Add Ten"} );
```

Custom Transform Example

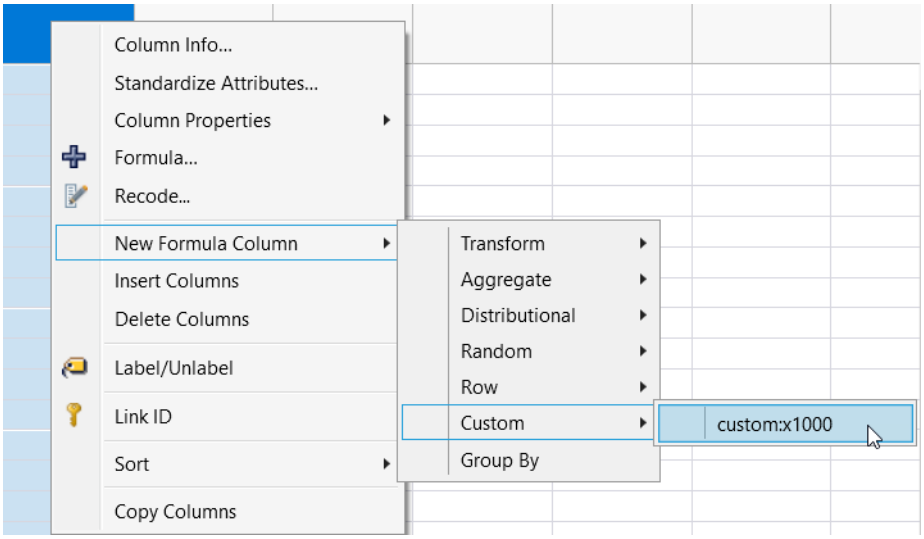
Let's create a simple Custom Transform to multiply a value by 1,000. We embed the function definition inside `Add Custom Functions()`.

```
Add Custom Functions(
  New Custom Function(
    "custom", // namespace in which new function resides
    "x1000", /* function name. The completely scoped name is
      "custom:x1000".*/

    Function( {x}, x * 1000 ),// function definition

    // optional message to enables this as a custom transform
    <<Transform Category( 1 )
    )
);
```

Running the above example writes "Deploying function: custom:x1000" in the JMP log. Search for this function in the Scripting Index to show that `custom:x1000` is an existing function.

When you create a new formula column in a data table, the custom transform appears as shown in Figure 9.15.

Figure 9.15 Custom Transform in a Data Table



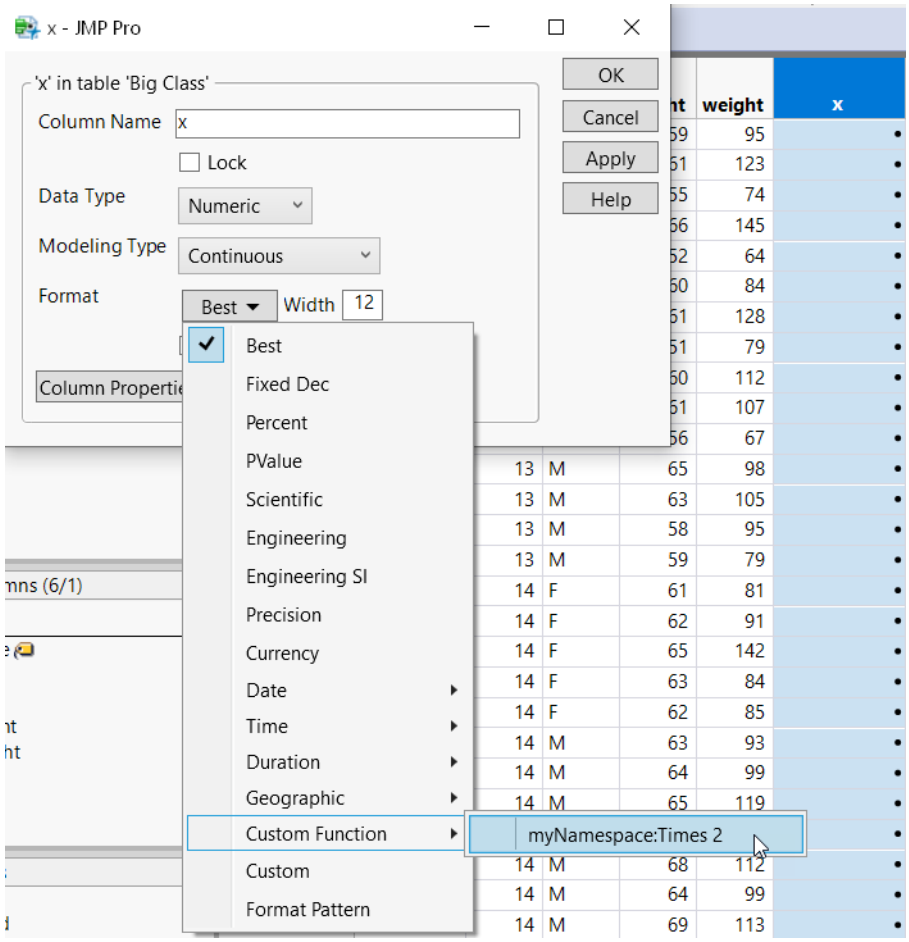
Custom Format Example: Multiplication

The last type of custom function is a custom format. You create custom formats using `New Custom Function()` and by specifying the `Custom Format Category(1)` message. Here's an example that multiplies an input variable by 2 and also displays `(x2)` after the value. Note that, in this example, this function persists as a custom format and in the formula category "My New Category". This allows you to use the format in a column formula or as a custom format.

```
/* add a New Custom Function, specifying it as a Custom Format
show "(x2)" after the value in the data table */
Add Custom Functions(
   New Custom Function(
      "myNamespace",
      "Times 2",
      Function( {inputVar},
         {Default Local},
         Char( inputVar * 2 ) || " (x2)"
      ),
      <<Description( "Multiply input by 2. Display (x2) after the new value."
   ),
      <<Formula Category( "My New Category" ),
      <<Custom Format Category( 1 ) // enable this as a custom format
   )
);
```

By specifying the `Custom Format Category(1)` message, this format now persists under the Custom Function category in the Format menu. Here's what this looks like when applying this new Times 2 custom format in the Column Info window:

Figure 9.16 Custom Format in the Col Info Window



Custom Format Example: Dates

Suppose that you want to display dates in a format that isn't provided in JMP. The following script formats a data to show the month, year, day, and day of week.

```
Add Custom Functions(
  New Custom Function(
    "myNamespace",
    "DateYMD_Day",
    Function( {inputVar},
      {Default Local},
```

```
            Char(
               Match( Month( inputVar ),
                  1, "January ",
                  2, "February ",
                  3, "March ",
                  4, "April ",
                  5, "May ",
                  6, "June ",
                  7, "July ",
                  8, "August ",
                  9, "September ",
                  10, "October ",
                  11, "November ",
                  12, "December "
               )
            ) || Char( Year( inputVar ) ) || ", Day#" || Char( Day( inputVar ) )
      || "-" ||
            Char(
               Match( Day Of Week( inputVar ),
                  1, "Sunday",
                  2, "Monday",
                  3, "Tuesday",
                  4, "Wednesday",
                  5, "Thursday",
                  6, "Friday",
                  7, "Saturday"
               )
            )

         ),
         <<Description(
            "Date format showing Month, Year, Day and Day of Week \!nfor example,
      April 1930, Day#29-Tuesday"
         ),
         <<Custom Format Category( 1 )
      )
   );
```

After you run the script, you can assign the format to a column of date values. In the Cols > Column Info window, the new date format is in the Format > Custom Function menu. You need to widen the data table column to be wide enough to display all of the characters.

Tip: See "Display Numbers as Four Digits" for another example of creating a custom format.

Delete Columns

To delete columns, send a `Delete Columns` message and specify which column or columns to delete. To delete more than one column, list the columns as multiple arguments or as a list.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Delete Columns( "weight" );
dt << Delete Columns( "weight", "age", "sex" );
dt << Delete Columns( {"weight", "age", "sex"} );
```

Without an argument, `Delete Columns` deletes columns that were previously selected. See "Column Attributes".

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column("age") << Set Selected;
dt << Delete Columns();
```

Get Column Names

`Column Name(`*n*`)` returns the name of the *n*th column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column Name( 2 );
   age
```

The returned value is a name value, not a quoted string. What this means is you can use it anywhere you would normally use the actual name in a script. For example, you could subscript it:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column Name( 2 )[1];
   12
```

If you want the name as a text string, quote it with `Char`:

```
Char( Column Name( 2 ) );
   "age"
```

To retrieve a list of the names of all columns in a data table, submit `Get Column Names`.

```
dt << Get Column Names( argument );
```

where the optional *`argument`* controls the output of the `Get Column Names` function, as follows:

- Specify `Numeric`, `Character`, `Row State`, or `Expression` to include only those column data types.
- Specify `Continuous`, `Ordinal`, `Nominal`, `Multiple Response`, `Unstructured Text`, `None`, and `Vector` to include only those modeling types.
- Specify `String` to return a list of strings rather than column names.

For example, if you want to get numeric and continuous columns in the Big Class.jmp sample data table, proceed as follows:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
names = dt << Get Column Names(Numeric, "Continuous")
   {height, weight}
```

Column Attributes

Use a collection of message pairs for data table columns to control all of the various attributes or characteristics of a column, including its name, data, states, and metadata. The messages come in pairs, one to "set" or assign each attribute and one to "get" or query the current setting of each attribute.

For example, you can hide, exclude, label, and turn on or off the scroll lock for a column through scripting. The value is Boolean; enter a one to turn the column attribute on, and a zero to turn it off.

In the following examples, the name column is unhidden, unexcluded, labeled, and locked from horizontal scrolling.

```
Column( "name" ) << Hide( 0 );
Column( "name" ) << Exclude( 0 );
Column( "name" ) << Label( 1 );
Column( "name" ) << Set Scroll Locked( 1 );
```

Note: All the messages to set various arguments (for example, `Set Name`, `Set Values`, `Set Formula`) start with `Set`. The word `Set` is optional for all messages except `Set Name` (recall that `Name` is already used for something else, the command that lets you use unusual characters in a name). Use whichever form you prefer or find easier to remember. The corresponding messages to retrieve the current value of an argument (for example, `Get Formula`) are the same, except that they start with `Get` instead of `Set`, and the word `Get` is *not* optional.

To deselect all selected columns, send a `Clear Column Selection` message to the data table object.

```
dt << Clear Column Selection;
```

Set or Get a Column Name

`Set Name` lets you name or rename a column, and `Get Name` returns the name for a column. The following example changes the name of the second column from age to ratio. It then returns the current column name to the log.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = Column( 2 );
```

```
col << Set Name( "ratio" );
col << Get Name;
  "ratio"
```

Set or Get Column Values

Similarly, `Set Values` sets values for a column. If the variable is character, the argument should be a list. If the variable is numeric, the argument should be a matrix (vector). If the number of values is greater than the current number of rows, the necessary rows are added to the data table. `Get Values` returns the values in list or matrix form. `Get As Matrix` is similar to `Get Values` but returns values in the numeric columns.

```
col << Set Values( myMatrix ); // for a numeric variable
col << Set Values( myList ); // for a character variable
col << Get Values; // returns a matrix, or list if character
col << Get Matrix(<list of column names>|<list of column numbers>|<column
  range>; // returns the specified columns as a matrix
```

The following example returns a list and a matrix of values:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column( "name" ) << Values( {Fred, Wilma, Fred, Ethel, Fred, Lamont} );
myList = :name << Get Values;
  {"Fred", "Wilma", "Fred", "Ethel", "Fred", "Lamont", "JAMES", "ROBERT",
    "BARBARA", ...}

Column( "age" ) << Values( [28, 27, 51, 48, 60, 30] );
myVector = :age << Get Values;
  [28, 27, 51, ... ]
myMatrix = :weight << Get as Matrix;
  [95, 123, 74, ... ]
```

Set or Get Value Labels

Note: For more information about value labels, see *Using JMP*.

Value labels provide a method of displaying a descriptive label for abbreviated data. For example, you might have a column of 0 and 1 values, where 0 represents a male and 1 represents a female. The value label “male” for 0 and “female” for 1 are more readable.

You can specify value labels in any one of the following three ways. Using the Big Class.jmp sample data table, assume that M maps to Male, and F maps to Female.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
:sex << Value Labels( {"F", "M"}, {"Female", "Male"} ); // use two lists
:sex << Value Labels( {"F", "Female", "M", "Male"} ); // use a list of pairs
:sex << Value Labels( {"F" = "Female", "M" = "Male"} ); // use a list of
  assignments
```

You can activate value labels by sending `Use Value Labels` as a message to the column.

```
:sex << Use Value Labels( 1 );
```

To revert back to showing the column's actual values:

```
:sex << Use Value Labels( 0 );
```

The same message can be used for the data table to turn value labels on and off for all columns.

```
dt << Use Value Labels( 1 );
```

Set or Get Data and Modeling Types

You can set or get the data type of a column using JSL. The choices are character, numeric, and row state. The following example adds a new column that has a character data type to the Big Class.jmp sample data table.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "New" );
Column( "New" ) << Data Type( Character );
Column( "New" ) << Get Data Type;
   "Character"
```

To set or get the modeling type of a column:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = New Column( "New" );
col << Modeling Type( "Continuous" );
col << Get Modeling Type;
   "Continuous"
```

You can specify the format of a column when changing its data type:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = New Column( "Date" );
Column( "Date" ) << Data Type( Numeric, Format( "ddMonYYYY" ) );
```

Set or Get Formats

The `Format` message controls numeric and date/time formatting. The first argument is a quoted string from the list of format choices shown in the Column Info window. Subsequent arguments depend on the format choice. You can also set the field width by itself.

```
col << Format( "Best", 5 ); // width is 5
col << Format( "Fixed Dec", 9, 3 ); // width is 9, with 3 decimal places
col << Format( "PValue", 6 );
col << Format( "d/m/y", 10 );
col << Set Field Width( 30 );
```

For date formats, the `Format` message sets how dates appear in a data table column. To set the format that you use for entering data, or for displaying the current cell when you have it selected for entry or editing, use the `Input Format` message.

```
col << Format( "d/m/y", 10 ); // display the date in day-month-year order
col << Input Format( "m/d/y" ); // enter the date in month-day-year order
```

For more information about date/time formatting choices, see "Date-Time Functions and Formats".

Note: Do not confuse the `Format` *message* for columns with the `Format` *function* for converting numeric values to strings according to the format specified (typically used for date/time notation as described in "Date-Time Functions and Formats"). Sending a message to an object has a very different effect from using a function that might happen to have the same name.

To get the current format of a column, submit a `Get Format` message:

```
col << Get Format;
```

Display Numbers as Four Digits

The `Format()` function and the `Format` message have a custom format capability that enables you to provide your own script to do the formatting. The name of the format is "Custom" (rather than "Best", "Fixed", and so on) and the argument is a script that works with the variable named "value" to return a string to represent that value. To make positive or negative integers always have at least four digits by adding zeros, you could do this:

```
New Table( "Untitled",
  New Column( "Column 1",
    Numeric,
    "Continuous",
    Format(
      "Custom", // make a custom format
      Formula( // using a JSL formula
        Local( {sign, length = 4, result}, // local variables
          sign = If( value < 0, -1, 1 ); // handle negative numbers
          value = Char( Abs( value ) );
          If( Length( value ) > length, // handle too big
            value = Repeat( "#", length )
          );
          result = Right( value, length, "0" ); // add zeros
          If( sign < 0,
            result = "-" || result // re-attach sign
          );
          result; // return the modified value
        )
```

```
        )
      ),
      // sample data to test positive, negative, zero
      Set Values( [0, 3, -5, 33, -55, 333, -555, 3333, -5555, 33333, -55555] )
    )
);
```

See "Create Custom Functions, Transforms, and Formats" for more information about creating custom formats.

Specify the Significant Digits in a Numeric Column

Significant digits are the first nonzero digit and all succeeding digits. For example, 3.14159 has six significant digits, and 0.00314 has three significant digits.

Set the number of significant digits with the Precision column format:

```
Format( "Precision", 12, 3 ); // 12 wide with 3 significant digits
```

Set, Get, or Evaluate a Formula

The following examples show how to set, get, and evaluate a formula:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = New Column( "Ratio" ); // create column and stores its reference
col << Set Formula( :height/:weight ); // set the formula
col << Eval Formula; // evaluate the formula
col << Get Formula; // return the expression :height/:weight
```

To use the values from columns in scripts, be sure to add commands to evaluate the formula. Formula evaluation timing can differ between different versions of JMP.

Notes:

- When formulas are added, they are scheduled to be evaluated in a background task. This can be a problem for scripts if they depend on the column having the values while the script is running.
- To force a single column to evaluate, you can send an `Eval Formula` command to the column. You can do this inside the command to create the column, right after the Formula clause:

  ```
  dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
  dt << New Column( "Ratio",
     Numeric,
     Formula( :height / :weight ),
     Eval Formula
  );
  ```

  where `Formula()` is an alias for `Set Formula()`.

However, it is best to wait until you are finished adding a set of formulas, and then use the command `Run Formulas` to evaluate all of the formulas in their proper order, as follows:

```
dt << Run Formulas;
```

- The `Run Formulas` command is preferable to the `Eval Formula` command, because while it is evaluating the formulas, `Eval Formula` does not suppress the background task from evaluating them again. The formula dependency system background task takes great care to evaluate the formulas in the right order, and `RunFormulas` simply calls this task until all the formulas are finished evaluating.
- If you use random numbers and use the `Random Reset(`*`seed`*`)` feature to make a reproducible sequence, then you have another reason to use `Run Formulas`, in order to avoid a second evaluation in the background.
- If formula suppression is enabled with `Supress Eval`, turn it off to enable `Eval Formula` to run.
- The `Add Custom Function()` function enables you to create your own functions and display them in the Formula Editor's Functions list. See "Create Custom Functions, Transforms, and Formats".

Set and Get Range and List Checks

You can manipulate list and range check properties using JSL. The following examples use the Big Class.jmp sample data table.

Set and clear the list check property in the sex column:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column( "sex" ) << List Check( {"M", "F"} ); // set the property
Column( "sex" ) << List Check(); // clear the property
```

Range checks require the specification of a range using the syntax in Table 9.3.

Table 9.3 Range Check Syntax

To specify this range	Use this function
$a \leq x \leq b$	`LELE(a, b)`
$a \leq x < b$	`LELT(a,b)`
$a < x \leq b$	`LTLE(a,b)`
$a < x < b$	`LTLT(a,b)`

The following example specifies that the values in the age column must be greater than zero and less than 120:

```
Column( "age" ) << Range Check( LTLT( 0, 120 ) );
```

All of the functions can be preceded by `Not` and one of them can be missing. The following example specifies that the values in the age column should be greater than or equal to 12:

```
Column( "age" ) << Range Check( not( LT( 12 ) ) );
```

To clear a range check state, submit an empty `Range Check()`:

```
Column( "age" ) << Range Check();
```

To retrieve the list or range check assigned to a column, send a `Get List Check` or `Get Range Check` message to the column:

```
Column( "sex" ) << Get List Check;
Column( "age" ) << Get Range Check;
```

Here is an example of `Get Range Check` for the age column in Big Class.jmp:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column( "age" ) << Range Check( LTLT( 0, 120 ) );
Column( "age" ) << Get Range Check;
   Range Check( LTLT( 0, 120 ) )
```

Note that you can also use `Set Property`, `Get Property`, and `Delete Property` to set, retrieve, and remove list checks and range checks. See "Column Properties".

Note: Operations sent through JSL that involve range check columns show any warnings in the log rather than in interactive windows.

Get a Column Script

`Get Script` returns a script to create the column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
New Column( "Ratio", Set Formula( :height/ :weight) );
Column( "Ratio" ) << Get Script;
   New Column( "Ratio",
      Numeric,
      "Continuous",
      Format( "Best", 10 ),
      Formula( :height / :weight )
   );
```

Preselect Roles

To preselect a role on a column, use the `Preselect Role` message. Choices include No Role, X, Y, Weight, Freq, and Validation. The `Get Role` message returns the current setting.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = New Column( "New" );
col << Preselect Role( X );
col << Get Role;
```

Lock a Column or Cells

To lock or unlock a column, use `Lock` or `Set Lock` with a Boolean argument. `Get Lock` returns the current setting.

```
col << Lock( 1 ); // lock
col << Set Lock( 0 ); // unlock
col << Get Lock; // show current state
```

Column Properties

Tip: JSL choices for properties are the same as those in the **Column Properties** menu in the Column Info window. The arguments for each property correspond to the settings in the Column Info window. An easy way to learn the syntax is to establish the property that you want in the Column Info window first, and then use `Get Property` to view the JSL. See *Using JMP* for more information about each property.

Data columns have numerous optional metadata attributes that can be set, queried, or cleared using the messages `Get Property`, `Set Property`, and `Delete Property`.

```
col << Set Property( "propertyName", {argument list} );
col << Get Property( "propertyName" );
col << Delete Property( "propertyName" );
```

The name of the property in question is always the first argument for `Set Property`, and what is expected for subsequent arguments depends on which property you set:

- `Get Property` and `Delete Property` always take a single argument, which is the name of the property.
- `Get Property` returns the property's settings. `Delete Property` completely removes the property from the column.
- To get a list of all column property names for a column, specify the column and then use `Get Properties List`.

If you want to set several properties, you need to send several separate `Set Property` messages. You can stack several messages in a single JSL statement if you want.

```
col << Set Property( "Axis",{Min(50), Max(180)} )<< Set Property( "Notes", "to
  get proportions" );
```

To get a property's value, send a `Get Property` message whose argument is the name of the property that you want:

```
Column("ratio") << Get Property( "axis" ); // return axis settings
```

To set columns as label columns:

```
dt << Set Label Columns( col1, col2, col3 );
```

To clear all label columns:

```
dt << Set Label Columns();
```

The same syntax works for `Set Scroll Lock Columns`, and `Scroll Lock`.

Set Column Properties with Variables

When writing JSL to script a column property, it is important to understand that the column is a repository for all column properties. The column stores the column property with the values supplied without evaluation or validation of the supplied arguments. In order to add column properties that contains variables using JSL code, the script must convert any variables to their actual values.

One way to do this is to use expressions. In the following example, the `Eval Expr()` function evaluates and replaces any variable wrapped in the `Expr()` function with its value. The outer `Eval()` function evaluates the entire statement after the replacement has occurred.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
// assign desired spec limits
lLimit = 55;
uLimit = 70;
tLimit = 62.5;

Eval(
   Eval Expr(

      // store the limits in the Spec Limits column property
      :height << Set Property("Spec Limits",
         {LSL( Expr( lLimit ) ), USL( Expr( uLimit ) ), Target( Expr( tLimit )
   ),
         Show Limits( 0 )}
      )
   )
);
```

Set Value Order Column Properties

The Value Order column property specifies the order in which you want the data to appear in reports. Data can be sorted by numerical order, custom order, or by the order in which the data appear in the data table.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );

// numerical ordering sorts character column values by splitting
the data into string and numeric segments. The following example turns off
numerical order, which you might specify if the column contains character
data. Hex data ("A1FE2", "A13E4", "BEEEE") sort differently with numerical
ordering on or off).
:age << Set Property( "Value Order", {Numerical Order( 0 )} );

// specifies a custom order for the age column
:age << Set Property( "Value Order", {Custom Order( {15, 17, 12, 13, 14, 16}
  )} );

// sorts "ELIZABETH" first, the remaining items are sorted in alphabetical
  order
:name << Set Property( "Value Order", {Custom Order( {"ELIZABETH"} )} )

// specifies that if all values have a predefined sort order
// (for example, the days of the week), the remaining items are
// sorted on that basis. This option is available only for character data.
:age << Set Property( "Value Order", {Common Order( 1 )} )

// sorts the column values by their occurrence in the data table,
// rather than sorting by value
:name << Set Property( "Value Order", {Row Order Levels( 1 )} );
```

When you set the column property, you are setting the entire property. Suppose that you specify the row order levels. If the data happen to match a Common Order (for example, the days of the week), then Common Order takes precedence. In that case, you need to also include `Common Order(0)` in the property; you cannot send separate `Set Property` messages.

Examples of Setting and Getting Popular Column Properties

Table 9.4 shows examples for setting and getting popular column properties. See *Using JMP* for more information about each property.

Table 9.4 Common Properties for Data Table Columns

Property	Description	Example Using Arguments

Notes	Stores notes about a column. Is a quoted text string.	`col<<Set Property( "Notes", "Extracted from Fisher iris data" );`
		
		`col<<Get Property( "Notes" );`

List Check
Range Check
	Prescribes the possible values that can be entered in a column.	`col<<Set Property( List Check,{"F","M"} );`
		
		`col<<Set Property( "Range Check",LTLT(0, 120) );`
		
		`col<<Get Property( "List Check" );`
		
		`col<<Delete Property( "Range Check" );`

Missing Value Codes	Specifies column values that should be treated as missing.	`col<<Set Property( "Missing Value Codes", {0, 1} );`

Value Labels	Specifies labels to be displayed in place of the values.	`col<<Value Labels( {0 = "Male", 1 = "Female"} );`
		
		Use Boolean values to turn value labels on or off:
		
		`col<<Use Value Labels( 1 );.`

Value Order	Specifies the order in which you want the data to appear in reports.	`col<<Set Property( "Value Order", {"Spring", "Summer", "Fall", "Winter"} );`

Table 9.4 Common Properties for Data Table Columns *(Continued)*

Property	Description	Example Using Arguments

`Value Colors and Color Gradient`	Specifies colors for either categorical or continuous data, respectively.	Specify the color by assigning a JMP number to each value:
		`col<<Set Property( "Value Colors", {"Female" = 3, "Male" = 5} );`
		Specify a color theme:
		`col<<Set Property( "Value Colors", Color Theme( "White to Blue" ) );`
		If you omit the color theme, JMP uses the color theme preferences.
		Specify the gradient and the range and midpoint for Value Gradient:
		`col<<Set Property( "Color Gradient", {"White to Blue", Range( {18, 60, 25} )} );`

`Axis`	Most platforms use this (if it exists) when constructing axes. Mostly Boolean.	`col<<Set Property( "Axis",{Min(50), Max(180), Inc(0), Minor Ticks(10), Show Major Ticks(1), Show Minor Ticks(1), Show Major Grid(0), Show Labels(1), Scale(Linear)} );`
`Units`	Provided for custom uses. Specify the units of measure.	`col<<Set Property( "units", "grams" );` `col<<Get Property( "units" );`
`Coding`	Used for DOE and fitting. List with low and high values.	`col<<Set Property( "Coding", {59,172} );` `col<<Get Property( "Coding" );`
`Mixture`	Used for DOE, fitting, and profiling. Specifies the Mixture column properties in a list.	`col<<Set Property( "Mixture", {0.2, 0.8, 1, L PseudoComponent Coding} );` `col<<Get Property( "Mixture" );`
`Row Order Levels`	Specify to sort levels by their occurrence in the data instead of by value.	`col<<Set Property( "Row Order Levels", 1 );`
`Spec Limits`	Used for capability analysis and variability charts.	`col<<Set Property( "Spec Limits", {LSL(-1), USL(1), Target(0)} );` `col<<Get Property( "Spec Limits" );`

Table 9.4 Common Properties for Data Table Columns *(Continued)*

Property	Description	Example Using Arguments

`Control Limits`	Used for control charts.	`col<<Set Property( "Control Limits", {XBar(Avg(44), LCL(29), UCL(69))} );`
		
		`col<<Get Property( "Control Limits" );`

`Detection Limits`	Set in DOE and used in generalized regression.	`col<<Set Property( "Detection Limits", {10, 50} );`

`Response Limits`	Set in DOE and used in desirability profiling.	`col<<Set Property( "Response Limits", {Goal("Match Target"), Lower(1,1), Middle(2,2), Upper(3,3)} );`
		
		`col<<Get Property("Response Limits");`
		
		Choices for `Goal` are `Maximize`, `Match Target`, `Minimize`, `None`. Other arguments take numeric value and desirability arguments.

`Design Role`	Used for DOE. Specify a single role.	`col<<Set Property( "Design Role", "Covariate" );`
		
		`col<<Get Property( "Design Role" );`
		
		Choices for role are `Continuous`, `Discrete Numeric`, `Categorical`, `Blocking`, `Covariate`, `Mixture`, `Constant`, `Uncontrolled`, `Random Block`, `Signal`, and `Noise`.

`Factor Changes`	Sets the difficulty of changing a factor (Easy, Hard, or Very Hard).	`col<<Set Property( "Factor Changes", Hard) ;`
		
		`col<<Get Property( "Factor Changes" );`

`Sigma`	Used for control charts. Specify known sigma value.	`col<<Set Property( "Sigma",1.332 );`
		
		`col<<Get Property( "Sigma" );`
		
		Each type of chart uses a different sigma calculation.

`Distribution`	Set the distribution type to fit to the column.	`col<<Set Property( "Distribution", Distribution( GLog ) );`
		
		`col<<Get Property( "Distribution" );`

Table 9.4 Common Properties for Data Table Columns *(Continued)*

Property	Description	Example Using Arguments

`Time Frequency`	Set the type of time frequency.	`col<<Set Property( "Time Frequency", Time Frequency( "Annual" ) );`
		
		`col<<Get Property( "Time Frequency" );`

`Map Role`	Set how the column is used to connect map shape data with name data. Specify the role and other information as necessary.	`col<<Set Property( "Map Role", Map Role( Shape Name Use( "filepath to data table", "column name" ) ) );`
		
		`col<<Get Property( "Map Role" );`

`Supercategories`	Groups specific categories into one category. Supported only in the Categorical platform.	`col<<Set Property( "Supercategories", {Group( "Genders", {"F", "M"} )} );`
		
		`col<<Get Property( "Supercategories" );`

`Multiple Response`	Specifies the character that separates the responses within a cell.	`col<<Set Property( "Multiple Response", Multiple Response( Separator( "." ) ) );`
		
		`col<<Get Property( "Multiple Response" ) );`
		
		Note: We recommend assigning the Multiple Response modeling type if the delimiter is a comma. Then you don't have to specify the delimiter in the column property. See *Using JMP*.

Table 9.4 Common Properties for Data Table Columns *(Continued)*

Property	Description	Example Using Arguments

`Profit Matrix`	Assigns weights that define a decision model.
	
	Specify a list that contains the weights in a matrix and categories in another list. In the matrix, each row contains weights that reflect the profit for predicting the category indicated by the column of the matrix.
	
	Caution: The matrix argument is the transpose of the profit matrix as displayed in the Column Properties window.
		`col<<Set Property( "Profit Matrix", {[5 -1, -1 4, -2 -2], {"F", "M", "Undecided"}} );`
		
		`col<<Get Property( "Profit Matrix" );`

`Expression Role`	Applies to columns that contain expressions. Specifies whether the expression should be interpreted as a Picture, a Matrix, or an Expression.	`col<<Set Property( "Expression Role", Expression Role( "Picture", MaxSize( 640, 480 ), StretchToMaxSize( 1 ), PreserveAspectRatio( 1 ), Frame( 0 ) )`

`[Custom property]`	Provided for custom uses. Corresponds to **Column Properties > Other** in the Column Info window.
	
	The first argument is a name for the custom property, and the second argument is an expression.
		`col<<Set Property( "Date recorded",12Dec1999 );`
		
		`long date( col<<Get Property( "Date recorded") );`

Specify Engineering SI Units

The Engineering SI format displays the value with an appropriate SI suffix. For example, 3,100 is displayed as 3.1k. 0.0055 is displayed as 5.5m.

Magnitude	Suffix	Magnitude	Suffix	Magnitude	Suffix	Magnitude	Suffix
103	k	1015	P	10-3	m	10-15	f
106	M	1018	E	10-6	µ	10-18	a
109	G	1021	Z	10-9	n	10-21	z
1012	T	1024	Y	10-12	p	10-24	y

Rows

This section describes the messages for adding and manipulating the rows in a data table. Row messages are directed to a data table reference, and most act on the currently selected rows. A number of row messages might not be practical in scripting (for example, `Move Rows`).

- "Add Rows"
- "Delete Rows"
- "Select Rows"
- "Find Rows"
- "Move Rows"
- "Assign Colors and Markers to Rows"
- "Color Cells"
- "Hide, Exclude, and Label Rows"
- "Iterate on Rows in a Table"
- "Prevent Editing Rows"
- "Row States"

Add Rows

To add rows, send an `Add Rows` message and specify how many rows. You can also specify after which row to insert the new rows. The arguments can either be numbers or expressions that evaluate to numbers.

```
dt << Add Rows( 3 ); // add 3 rows to bottom of data table
dt << Add Rows( 3, 10 ); /* add 3 rows after the 10th row, moving the 11th and
   lower rows farther down */
```

A variation of `Add Rows` lets you specify an argument yielding a list of assignments. Assignments can be separated with commas or semicolons.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Add Rows(
   {:name = "Peter", :age = 14, :sex = "M", :height = 61, :weight = 124}
);
```

or

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Add Rows(
   {:name = "Peter"; :age = 14; :sex = "M"; :height = 61; :weight = 124}
);
```

You can send several arguments yielding lists, or even a list of lists. The following script creates a data table with `Add Rows` commands of each variety:

```
dt = New Table( "Cities" );
dt << New Column( "xx", Numeric );
dt << New Column( "cc", Character, width( 12 ) );

dt << Add Rows( {xx = 12, cc = "Chicago"} ); // single list
dt << Add Rows( // several lists
   {xx = 13, cc = "New York"}, {xx = 14, cc = "Newark"} );
dt << Add Rows( // list of lists
   {{xx = 15, cc = "San Francisco"}, {xx = Sqrt( 256 ), cc = "Oakland"}}
);
a = {xx = 20, cc = "Miami"};
dt << Add Rows( a ); // evaluate as single list

b={{xx = 17, cc = "San Antonio"},{xx = 18, cc = "Houston"}, {xx = 19, cc =
   "Dallas"}};
dt << Add Rows( b ); // evaluate as list of lists
```

Further information for rows can be specified with messages described under "Row States".

Delete Rows

To delete rows, send a `Delete Rows` message and specify which row or rows to delete. To delete more than one row, give a list or matrix as the *rownum* argument, or combine `Delete Rows` with other commands such as `For`. The *rownum* argument can be a number, list of numbers, range of numbers, matrix, or an expression that yields one of these. Without an argument, `Delete Rows` returns the number of selected rows and deletes those rows. When an argument is not specified or rows are not selected, `Delete Rows` only returns the number of selected rows: 0.

```
dt << Delete Rows( 10 ); // delete row 10
```

```
dt << Delete Rows( {11, 12, 13} ); // delete rows 11-13
myList = {11, 12, 13};
dt << Delete Rows( mylist ); // delete rows 11-13
dt << Delete Rows( 1 :: 20 ); // delete first 20 rows
dt << Delete Rows( [1 2 3] );  // delete first 3 rows
```

For example, the following script opens Big Class.jmp and deletes row 10:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Selected( Row State( 10 ) ) = 1; // select row 10
dt << Delete Rows;  // delete row 10
```

You can list duplicate rows, and you can list rows in any order with no consequence.

Here is a general way to remove the bottom *x* rows of a data table of any size. The following example removes five rows from the bottom of the data table:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
x = 5;
n = N Row( dt );
For( i = n, i > n - x, i--,
   dt << Delete Rows( i )
);
```

`NRow` counts the rows in the table. See "Iterate a Script on Each Row".

Select Rows

`Select All Rows` selects (or highlights) all of the rows in a data table.

```
dt << Select All Rows;
```

If all rows are selected, you can deselect them all by using `Invert Row Selection`. This command reverses the selection state for each row, so that any selected rows are deselected, and any deselected rows are selected.

```
dt << Invert Row Selection;
```

Note: With the exception of `Invert Row Selection`, whose result depends on the current selection, any new selection message starts over with a new selection. If you already have certain rows selected and you then send a new message to select rows, all rows are first deselected.

To select a specific row, use `Go To Row`:

```
dt << Go To Row( 9 );
```

To select specific rows in a data table based on their row number, use the `Select Rows` command. The argument to the command is a list of row numbers. For example, to select rows 1, 3, 5, and 7 of a data table:

```
dt << Select Rows( {1, 3, 5, 7} );
```

To select a range of rows, specify one of the following messages:

```
dt << Select Rows( Index( 7, 10 ) );
dt << Select Where( Any( Row() == Index( 7, 10 ) ) );
```

Both of these examples select rows seven through 10 in the current data table.

Select Where

To select rows according to data values, use `Select Where`, specifying a logical test inside the parentheses.

Tip: For a description of the functions and operators that you can use within a `Select Where` message, see "Operators".

For example, using the Big Class.jmp sample data table, select the rows where the students' age is greater than 13:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age > 13 );
```

Or, select the rows where the students' ages are less than 14:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
col = Column( dt, 2 );
dt << Select Where( col[] < 14 );
```

The following example selects the rows where the student's ages are less than 15 and the sex is "F":

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age < 15 & :sex == "F" );
```

To select a row without deselecting a previously selected row, combine `<< Select Where` with `<< Select Where` and the `Current Selection("extend")` argument. This is an alternative to using an `OR` statement.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age == 14 );
dt << Select Where( :sex == "F", Current Selection( "extend" ) );
```

To select rows that contain specific strings, follow this example:

```
dt = Open( "$SAMPLE_DATA/Hollywood Movies.jmp" );
my_genres = {"Romance", "Comedy"};
dt << Select Where( Contains( my_genres, :Genre ) );
```

You can also create more complex conditions by using regular expressions.

```
// select rows where :name begins with "j" or "J".
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
/* \0 is a group reference that causes the expression to return the entire
   matched text. !Is Missing is needed because the regex returns the matched
   text if it successfully matched or missing if it failed to match. So !Is
   Missing means the string successfully matched */
dt << Select Where( !Is Missing( Regex( :name, "^j", "\0", IGNORECASE ) ) );

// select rows where :name ends with "N" or "E".
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( !Is Missing( Regex( :name, "(N|E)$", "\0" ) ) );
```

Select Excluded, Hidden, or Labeled Rows

To select rows that are currently excluded, hidden, or labeled:

```
dt << Select Excluded;
dt << Select Hidden;
dt << Select Labeled;
```

To select rows that are *not* excluded, hidden, or labeled, stack a select message and an invert selection message together in the same statement, or send the two messages sequentially:

```
dt << Select Hidden << Invert Row Selection;
dt << Select Hidden;
dt << Invert Row Selection;
```

Select Duplicate Rows

To select duplicate rows in a data table, use the `Select Duplicate Rows` message:

```
dt = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
dt << Select Duplicate Rows(); // selects rows 301 and 8864
```

The second and subsequent duplicate rows are selected. Note that the duplicate values are case sensitive.

Note: By default, if no matching column is specified, all of the columns are matched.

In the preceding example, the duplicate rows might have been data input errors. Exclude the duplicate rows to remove them from the analysis:

```
dt = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
dt << Select Duplicate Rows();
dt << Exclude();
```

To select duplicate values in the rows of selected columns, specify the column names. The following example finds duplicate values in the rows of the Incident Number and Time columns.

```
dt = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
dt << Select Duplicate Rows( Match( Column( "Incident Number" ), :Time ) );
```

To select duplicate rows according to data values, select the duplicate rows, specify the data values, and then restrict the duplicate rows to the current selection.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Duplicate Rows( Match( :age ) );
dt << Select Where( :age > 15, Current Selection( "restrict" ) );
```

Refer to a Specific Cell in the Row

To refer to a specific cell, assign a subscript to the cell's row number. In the following example, the subscript `[1]` is used with the weight column. The formula then calculates the ratio between each height and the first value in the weight column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
New Column( "ratio", Formula( height / weight[1] ) );
```

Select Random Rows

To obtain a random selection, use this syntax:

```
dt << Select Randomly( number );
dt << Select Randomly( Sample Size( number ) );
dt << Select Randomly( probability );
dt << Select Randomly( Sampling Rate( probability ) );
```

These commands use a conditional probability to obtain the exact count requested.

Select Matching Cells

The row menu command `Select Matching Cells` is also implemented in JSL.

```
// select matching cells in the current data table
dt << Select Matching Cells;
// select matching cells in all open data tables
dt << Select All Matching Cells;
```

For more complicated selections, or to store selections permanently as row state data, see "Row States".

Find Rows

`Get Rows` returns a matrix of rows that match the specified condition. The following example select rows in which the age is greater than or equal to 16:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
```

```
dt << Get Rows Where( :age >= 16 );
   [35, 36, 37, 38, 39, 40]
```

`Get Selected Rows` returns a matrix of the currently selected rows. The following example selects rows 1, 3, 5, and 7 and then returns the row numbers in a matrix:

```
dt << Select Rows( {1, 3, 5, 7} );
dt << Get Selected Rows;
   [1, 3, 5, 7]
```

Find Selected Rows

`Next Selected` and `Previous Selected` scroll the data table window up or down so that the next selected row that is not already in view moves into view. The table wraps, so `Next Selected` jumps from the bottom-most selected row to the top-most, and *vice versa* for `Previous Selected`.

```
dt << Next Selected;
dt << Previous Selected;
```

Clear Selected Rows

To cancel a selection, leaving no rows selected, use `Clear Select`:

```
dt << Clear Select;
```

Move Rows

These commands move the currently selected rows to the indicated destination point.

```
dt << Move Rows( At Start );
dt << Move Rows( At End );
dt << Move Rows( After( rowNumber ) );
```

Assign Colors and Markers to Rows

You can use the `Colors` and `Markers` messages to assign (or change) colors and markers used for rows. These settings mostly affect graphs produced from the data table. Both messages expect numeric arguments to choose which color or marker to use. For more information about how numbers correspond to colors and markers, see "Colors and Markers".

```
dt << Colors( 3 );  // set selected rows to red
dt << Markers( 2 ); // pick the X marker for selected rows
```

As with other row messages, you can stack selection and other messages together:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age == 12 ) // select the youngest subjects
   << Colors( 8 ) << Markers( 8 ); // and use purple open circles for them
```

`Color by Column` sets colors according to the values of a column that you specify, and `Marker by Column` works similarly:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Color by Column( :age );
dt << Marker by Column( :age );
```

Here are additional, named arguments:

- `Continuous Scale` (`Color by Column` only) Assigns colors in a chromatic sequential fashion based on the values in the highlighted column.
- `Reverse Scale` Reverses the color scheme in use.
- `Make Window with Legend` Creates a separate window with a legend.
- `Excluded Rows` Applies the row states to excluded rows.
- `Marker Theme` Specifies the marker type.
- `Color Theme` Specifies the color theme.

Color Cells

You can color individual cells in the data grid. For example, this line uses the row state color to color the cells:

```
dt << Color Rows by Row State;
```

You can also specify a color theme for either categorical or continuous columns:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
:height << Set Property(
  "Color Gradient",
  {"White to Blue", Range( 40, 80 )}
);
:height << Color Cell By Value( 1 ); // turn on the cell coloring

:age << Set Property(
  "Value Colors", // assign the value colors
  {12 = Red, 13 = Yellow, 14 = Green, 15 = Blue, 16 = Magenta, 17 =
  Gray}
);

:age << Color Cell By Value( 0 ); // turn off the cell coloring
```

You can also color specific cells. The following example sets rows 1, 5, 8 of the “name” column to red:

```
:name << Color Cells( red, {1, 5, 8} );
```

To remove the color from specific cells, set the color to black. The following example removes the color for row 1 of the “name” column.

```
:name << Color Cells( black, {1} );
```

You can color cells based on specific values. The following example colors the cells in the height column. All cells containing a value greater than 60 are blue. All cells containing a value equal to or less than 60 are purple.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For( i = 1, i <= N Row( dt ), i++,
   If( Column( dt, "height" )[i] > 60,
      Column( dt, "height" ) << Color Cells( 5, {i} ),
      Column( dt, "height" ) << Color Cells( 8, {i} )
   )
);
```

Note: The first argument for `Color Cells` represents the color value. The second argument contains the row numbers.

Hide, Exclude, and Label Rows

Note: For more information about hiding, excluding, and labeling rows using row state functions, see “Row States”.

Use the `Hide`, `Exclude`, and `Label` messages to hide, exclude, and label rows. These messages are toggles, meaning that to turn the messages on you send them once, and to turn them off you send the message a second time. They also accept Boolean arguments to explicitly turn them off and on.

For example, to hide all rows in Big Class where age is greater than 13, you could do the following:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age > 13 );
dt << Hide( 1 );
```

Messages to the same object can be stacked together in a single statement, so you could simplify the preceding script.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age > 13 ) << Hide( 1 );
```

Iterate on Rows in a Table

In addition to built-in programming functions for iterating, JSL provides functions for iterating through data table rows, groups, or conditional selections of rows.

Generally, an expression is executed on the current row of the data table only. Some exceptions are the expressions inside formula columns, `Summarize` and the pre-evaluated statistics functions, and any use of data table columns by analysis platforms.

Set the Current Row

Note: The current row for scripting is *not* related to rows being selected (or highlighted) in the data table or to the current cursor position in the data table window. The current row for scripting is defined to be zero (no row) by default.

You can set the current row for a script using `For Each Row` or `Row() = X`.

```
Row() = 3; ...
For Each Row( ... );
```

`For Each Row` executes the script once for each row of the current data table. Note that `Row()=1` only lasts for the duration of the script, then `Row()` reverts to its default value, which is zero. This means that submitting a script all at once can produce different results than submitting a script a few lines at a time.

Throughout this chapter, examples without an explicit current row should be assumed to take place within a context that establishes a current row. See "What is the Current Row?".

What is the Current Row?

By default, the current row number is 0. The first row in a table is row 1, so row 0 is essentially not a row. In other words, *by default, an operation is done on no rows*. Unless you take action to set a current row or to specify some set of rows, you get missing values due to the lack of data. For example, a column name returns the value of that column on the current row. Scope the column name with the prefix `:` operator to avoid ambiguity (to force the name to be interpreted as a column name).

```
:sex; // returns ""
:age; // returns .
```

Scoping names prevents you from getting a result that might look reasonable for the whole data table but is actually based on only one row. It also protects you from accidentally overwriting data values when making assignments to ambiguous names under most circumstances. You can have even more complete protection by using the prefix or infix `:` operator to refer specifically to a data column and the prefix :: operator to refer specifically to a global script variable. See "Advanced Scoping and Namespaces".

You can use the Row() function to get or set the current row number. Row() is an example of an *L-value* expression in JSL: a function that returns its value unless you place it before an assignment operator (=, +=, and so on.) to set its value.

```
Row(); // returns the number of the current row (0 by default)
x = Row(); // store the current row number in x
Row() = 7; // make the 7th row current
Row() = 7; :age; // make the 7th row current and returns 12
```

Note that the current row setting only lasts for the portion of a script that you select and submit. After the script executes, the current row setting resets to the default (row 0, or no row). Therefore, a script submitted all at once can produce different results from the same script submitted a few lines at a time.

How Many Rows and Columns?

The N Rows() and N Cols() functions return the rows and columns in a data table.

```
N Rows( dt ); // number of rows
N Cols( dt ); // number of columns
```

N Rows and N Cols also count the number of rows in matrices. Note that NRow and NCol are synonyms. See the "Inquiry Functions".

Iterate a Script on Each Row

To iterate a script on each row of the current data table, put For Each Row around the script.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row( If( :age > 15, Show( :age ) ) );
```

To specify the open data table, include a data table reference as the first argument.

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt2 = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
For Each Row( dt1, If( :age > 15, Show( :age ) ) );
```

You can use For Each Row to set row states instead of creating a new formula column in the data table. The scripts below are similar, except that the first one creates a row state column, and the For Each Row script simply sets the row state without creating a column.

```
New Column( "My Row State", Row State, Formula( Color State( :age - 9 ) ) );
For Each Row( Color of( Row State() ) = :age - 9 );
```

To iterate a script on each row that meets a specified condition, combine For Each Row and If:

```
For Each Row( Marker of( Row State() ) = If( :sex == "F", 2, 6 ) );
```

You can use Break and Continue to control the execution of a For Each Row loop. See "Break and Continue".

Return Row Values

`Dif()` and `Lag()` are special functions that can be useful for statistical computations, particularly when working with time series or cumulative data.

- `Lag()` returns the value of a column *n* rows before the current row.
- `Dif()` returns the difference between the value in the current row and the value *n* rows previous.

The following lines are equivalent:

```
dt  <<  New Column( "htDelta" );
For Each Row( :htDelta = :height - Lag( :height, 1 ) );
For Each Row( :htDelta = Dif( :height, 1 ) );
```

Add Sequence Data

`Sequence()` corresponds to the Sequence function in the Formula Editor and is used to fill the cells in a data table column. It takes four arguments and the last two are optional:

```
Sequence( from, to, stepsize, repeat );
```

`From` and `to` are not optional. They specify the range of values to place into the cells. If `from = 4` and `to = 8`, the cells are filled with the values 4, 5, 6, 7, 8, 4, ...

`Stepsize` is optional. If you do not specify a stepsize, the default value is 1. Stepsize increments the values in the range. If `stepsize = 2` with the above `from` and `to` values, the cells are filled with the values 4, 6, 8, 4, 6, ...

`Repeat` is optional. If you do not specify a Repeat, the default value is 1. Repeat specifies how many times each value is repeated before incrementing to the next value. If `repeat = 3` with the above `from`, `to`, and `stepsize` values, the cells are filled with the values 4, 4, 4, 6, 6, 6, 8, 8, 8, 4, ... If you specify a `Repeat` value, you must also specify a `Stepsize` value.

The sequence is always repeated until each cell in the column is filled.

```
dt = New Table( "Sequence Example" ); // create a new data table
dt << New Column( "Count to Five" ); // add two columns
dt << New Column( "Count to Seventeen by Fours" );
dt << Add Rows( 50 ); // add fifty rows

/* fill the first column with the data sequence 1, 2, 3, 4, 5, ...
fill the second column with the data sequence 1, 1, 5, 5, 9, 9, 13, 13, 17,
  17, ... */
For Each Row (
  Column( 1 )[ ] = Sequence( 1, 5 );
  Column( 2 )[ ] = Sequence( 1, 17, 4, 2 );
);
```

Because `Sequence()` is a formula function, you can also set a column's formula to use `Sequence()` to fill the column. The following example creates a new column named Formula Sequence and adds a formula to it. The formula is a sequence that fills the column with values between 25 and 29, incremented by 1, and repeated twice (25, 25, 26, 26, 27, 27, 28, 28, 29, 29, 25, ...).

```
dt = New Table( "Formula Sequence Example" );
dt << Add Rows( 20 );
dt << New Column( "Formula Sequence", Formula( Sequence( 25, 29, 1, 2 ) ) );
```

The following are more examples of `Sequence()` results:

- `Sequence(1,5)` produces 1,2,3,4,5,1,2,3,4,5,1, ...
- `Sequence(1,5,1,2)` produces 1,1,2,2,3,3,4,4,5,5,1,1, ...
- `Sequence(10,50,10)` produces 10,20,30,40,50,10, ...
- `10*Sequence(1,5,1)` also produces 10,20,30,40,50,10, ...
- `Sequence(1,6,2)` produces: 1,3,5,1,3,5, ... The limit is never reached exactly.

Note: If you want a matrix of values, then use the `Index` function, not `Sequence`.

Prevent Editing Rows

To prevent adding or deleting rows, use `Set Edit Lock`:

```
dt << Set Edit Lock( "add rows", "delete rows");
```

See "Prevent Changes to a Data Table" for more information about preventing other data table operations.

Row States

There is a special data element type called a row state that stores various attributes in the data table. Row states can indicate the following:

- Whether a row is selected, excluded, hidden, or labeled
- Which marker type, color, shade, and hue to use for graphs

In JSL, you can use row state functions to manipulate row states.

About Row States

Row states change how JMP works with your data. Table 9.5 explains each row state. Remember that you can use several row states at once to get the combination of effects that you want.

Table 9.5 Row States

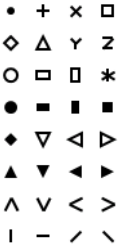
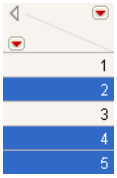
Row states	How they affect results
Excluded	If rows are excluded, JMP omits them from calculations for statistical analyses (text reports and charts). Results are the same as if the data was not entered. However, points are still included in *plots*. (To omit points from plots, use `Hide`. To omit points from all results, use both `Exclude` and `Hide`.)
Hidden	If rows are hidden, JMP does not show them in plots. However, the rows *are* still included in text reports and *charts*. (To omit points from reports and charts, use `Exclude`. To omit points from all results, use both `Exclude` and `Hide`.)
Labeled	If rows are labeled, JMP places row number labels, or the values from a designated Label column, on points in scatterplots.
Color	If rows have colors, JMP uses those colors to distinguish the points in scatterplots.
Marker 	If rows have markers, JMP uses those markers to distinguish the points in scatterplots.

Table 9.5 Row States *(Continued)*

Row states	How they affect results
Selected 	If rows are selected, JMP highlights the corresponding points and bars in plots and charts.

Note: For row state messages, omitting the argument *toggles* the option: if the option is off, the message turns it on, and if the option is on, the message turns it off. You can also use the `"toggle"`, `"switch"`, or `"flip"` argument, as in `r << Exclude( "toggle" )`.

About Row State Functions

When you give a row state function a number as its argument (or an expression that evaluates to a number), the function interprets that number as an index to its possible values.

Table 9.6 shows a comparison chart of the different row state functions, so that you can see which functions convert row states to numbers and numbers to row states. It also includes the numbers that you can use with each function.

Table 9.6 Row State Functions

Numbers	Convert from numbers to row states →	Row states	Convert from row states to numbers →	Numbers
1 or 0	`Excluded State(`*`n`*`)`	Excluded	`Excluded(`*`rowstate`*`)`	1 or 0
1 or 0	`Hidden State(`*`n`*`)`	Hidden	`Hidden(`*`rowstate`*`)`	1 or 0

Table 9.6 Row State Functions *(Continued)*

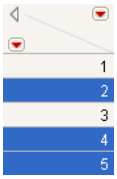
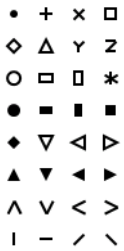
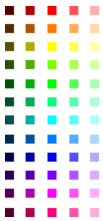
Numbers	Convert from numbers to row states →	Row states	Convert from row states to numbers →	Numbers
1 or 0	`Labeled State(n)`	Labeled	`Labeled(rowstate)`	1 or 0
1 or 0	`Selected State(n)`	Selected 	`Selected(rowstate)`	1 or 0
0 to 31	`Marker State(n)`	Marker 	`Marker Of(rowstate)`	0 to 31
0 to 84 (0–15 basics, 16–31 dark, 32–47 light, 48–63 very dark, 64–79 very light, 80–84 grays)	`Color State(n)`	Color	`Color Of(rowstate)`	0 to 84 (0–15 basics, 16–31 dark, 32–47 light, 48–63 very dark, 64–79 very light, 80–84 grays)
0–11 (rainbow order)	`Hue State(n)`	Hue		

Table 9.6 Row State Functions *(Continued)*

Numbers	Convert from numbers to row states →	Row states	Convert from row states to numbers →	Numbers
–2 to 2 (dark to light)	`Shade State(n)`	Shade 		

Assign Row States

To assign a row state using JSL, use `Select Where` to indicate which rows are affected, and then specify which row state to assign to the rows. In the following example, marker type 5 (a triangle) is assigned to rows in which sex is “F”.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :sex == "F" ) << Markers( 5 );
```

You can also use a formula to assign row states. Here is how you would do it in the data table:

1. Create a row state column.
2. Add a formula to the column that assigns a marker to each row in which sex is “F”.
3. In the Column panel, right-click the star next to the row state column and select **Copy to Row States**.

The JSL equivalent would be something like this:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Row State",
   Row State,
   Set Formula( If( :sex == "F", Marker State( 4 ) ) )
);
Column( "Row State" ) << Copy To Row States();
```

Store Row State Information

A row state column is a dedicated column that *stores* row state information, but does not put the information into effect. You can then use `For Each Row` to put the row state column into effect. For more information about row state columns, see *Using JMP*.

The following example creates a row state column then puts the row states into effect:

1. Submit this line to start a new data table:

```
dt = New Table( "Row State Testing" );
```

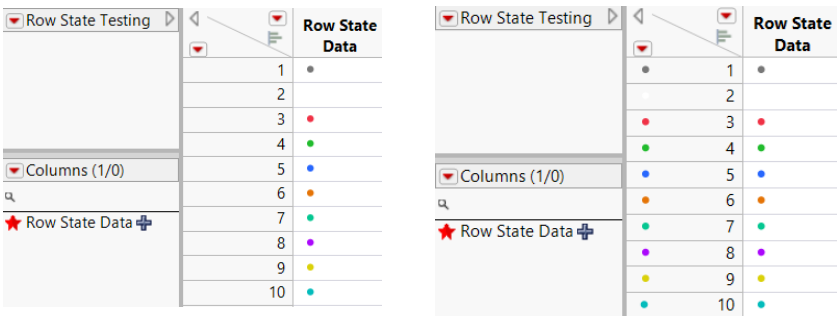
2. Submit these lines to add row states to a column and add ten rows:

```
dt << New Column( "Row State Data", Row State, Set Formula( Color State( Row()
   ) ) );
dt << Add Rows( 10 );
```

3. Submit this line to put the row states into effect:

```
For Each Row( Row State() = :Row State Data );
```

Figure 9.17 Table with No Row States (left) and Table with Row States (right)



This action replaces any row states in effect with the row state combination from the row state column. For more information about how to change selected attributes of a row state without changing or canceling others, see "Set One Characteristic and Cancel Others".

Set or Get Row States

From JSL, you can set or get row states directly using the `Row State` function. You can set or get the state for row *n* with `Row State(n)`. If you do not supply an argument, you set or get the current row. To work with all rows, either use a `For Each Row` loop or work with formula columns.

To set a row state, place the row state expression on the left of an assignment (L-value):

```
Row State( 1 ) = Color State( 3 ); // make row 1 red
Row() = 8; Row State() = Color State( 3 ); // make the 8th row red
For Each Row( Row State() = Color State( 3 ) ); // make each row red
```

To specify the data table, include a data table reference as the first argument:

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt2 = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
Row State( dt1, 1 ) = Color State( 3 ); // make row 1 red;
```

To get a row state, place the row state expression on the right side of an assignment:

```
x = Row State( 5 );  // put the row state of row 5 into x
x = Row State();   // row state of current row
```

Notes on Setting Row States

Be careful whether you set every aspect of `Row State()` or just one aspect of it, such as `Color Of( Row State() )`. To see how this works, first color and mark all the rows:

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row(
  Row State() = Combine States( Color State( Row() ), Marker State( Row() ) )
  );
```

And now observe the difference between setting one attribute of a row state:

```
Color Of( Row State( 1 ) ) = 3; // make row 1 red without changing marker
```

And setting every aspect of a row state to a single state:

```
Row State( 1 ) = Color State( 5 ); // make row 1 blue and remove its marker
```

To copy all the current row states into a row state column:

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row(
       dt1,
       Row State() = Combine States(
              Color State( Row() ),
              Marker State( Row() )
       )
);
dt1 << New Column( "rscol", "Row State", Set Formula( Row State() ) );
```

To copy several but not all of the current row states into a row state column, use a script like the following (commenting out or omitting any states that you do not want):

```
New Column( "rscol2",
  Set Formula(
     Combine States(
        Color State( Color Of() ),
        Excluded State( Excluded() ),
        Hidden State( Hidden() ),
        Labeled State( Labeled() ),
        Marker State( Marker Of() ),
```

```
            Selected State( Selected() )
        )
    )
);
```

To set a component of a row state:

```
Color Of( Row State( i ) ) = 3; // change color to red for row i
Selected( Row State( i ) ) = 1; // select the ith row state and set to 1
```

You can see from the example above that some of the functions convert numbers into states, and others convert states into numbers. Here are some helpful hints for remembering which are which:

Number-to-state functions have the word "State" Functions that take number arguments and either return states or accept state assignments all have the word "State" in their names: `Row State`, `As Row State`, `Color State`, `Combine States`, `Excluded State`, `Hidden State`, `Hue State`, `Labeled State`, `Marker State`, `Selected State`, `Shade State`.

State-to-number functions are one word or have the word "Of" Functions that take row state arguments (and assume that the argument `Row State()` if none is given) and functions that return or are set to numbers are either one word, or their second word is "Of": `Color Of`, `Excluded`, `Hidden`, `Labeled`, `Marker Of`, `Selected`.

Table 9.6 is a helpful comparison chart for these functions.

The following lines are equivalent to their interactive commands:

```
Copy From Row States
Add From Row States
Copy To Row States
Add To Row States
```

Examples of Getting Row States

For example, create the following table containing row states:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Row State Data", Row State, Set Formula( Color State( Row()
  ) ) );
dt << Add Rows( 10 );
For Each Row( Row State() = :Row State Data );
```

To get row states, submit the following lines:

```
Row State( 1 ); // returns row state for row 1
Row() = 8; Row State();  // returns row state for current (8th) row
For Each Row( Print( Row State()) ); // returns the row states for each row
```

To get a row state and store it in a global, place `Row State()` on the right side of an assignment. Add the following lines to the preceding script:

```
::x = Row State( 1 ); // put the row state of row 1 in x, a global
Row() = 8; ::x = Row State();  // put the row state of the 8th row in x
Show( x ); // returns x = Color State( 8 )
dt << New Column( "rscol", Row State) ;
// put row states in rscol (a row state column)
For Each Row( :rscol = Row State() );
```

To get a component of a row state, place the row state expression on the right side of an assignment and also use one of the L-value operators:

```
x = Selected( Row State() ); // selection index of current row selected
```

Get and Set Multiple Characteristics at Once

You can get or set many characteristics at once by combining state settings inside `Combine States()`. You can also get or set each characteristic one at a time, the ultimate row state being the accumulation of characteristics. The following example sets green Y markers for males, but hides them in plots for now. It also sets red X markers for females but shows them in plots.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row(
  If( sex == "M",
     /* then */ Row State() = Combine States(
        Color State( 4 ), Marker State( 6 ), Hidden State( 1 ) ),
     /* else */ Row State() = Combine States(
        Color State( 3 ), Marker State( 2 ), Hidden State( 0 ) ) ) );
```

Get the row state for one row, such as the 6th:

```
Row State( 6 );
  Combine States(Hidden State(1), Color State(4), Marker State(6))
```

Notice that JMP returns a `Combine State()` combination. This is because a row state datum is not just the state of one characteristic, such as color, but the cumulative state of *all* the characteristics that have been set: exclusion, hiding, labeling, selection, markers, colors, hues, and shades. A list of such characteristics is called a *row state combination*.

Just as there can be many row state characteristics in effect, a row state *column* can have multiple characteristic row states as its values.

Set One Characteristic and Cancel Others

In addition to the overall `Row State()` function for getting or setting all the characteristics of a row state, there are separate functions to get or set one characteristic at a time preemptively. That is, to give a row one characteristic, canceling any other characteristics that might be in effect. The functions that set one characteristic and cancel others are `Color State`, `Combine States`, `Excluded State`, `Hidden State`, `Hue State`, `Labeled State`, `Marker State`, `Selected State`, `Shade State`.

For example, to make row 4 be hidden only:

```
Row State( 4 ) = Hidden State( 1 );
```

Set or Get One Characteristic at a Time

A row state is not just one characteristic, but many. To work with just one characteristic at a time, use one of the L-value operators with `Row State` on either side of the equal sign. The side of the equal sign depends on whether you want to get or set a characteristic. There is an L-value operator for each of the following characteristics: `Color Of`, `Excluded`, `Hidden`, `Labeled`, `Marker Of`, `Selected`.

This example hides row 4 without affecting any other characteristics:

```
Hidden( Row State( 4 ) ) = 1
```

This example stores the color of row 3 without getting any other characteristics:

```
::color = Color Of( Row State( 3 ) );
```

Identify Row State Changes

The `MakeRowStateHandler` message (sent to a data table object) obtains a callback when the row states change. For example,

```
f = Function( {X}, Show( x ) );
obj = dt << Make Row State Handler( f );
```

Then when you select a group of rows, the row numbers of any row whose row state changed are sent to the log. For example:

```
x:[3, 4, 28, 40, 41]
```

When a group is highlighted, it might call the handler twice, once for rows whose selection is cleared, then again for the new selection.

Exclude, Hide, Label, and Select

This section discusses conditions that have Boolean states (either on or off). These conditions include: excluding, hiding, labeling, and selecting rows.

- `Excluded` gets or sets an excluded index. The index is 1 for true or 0 for false, indicating whether each row is excluded.
- `Hidden` gets or sets a hidden index, which is 1 for hidden or 0 for not hidden.
- `Labeled` gets or sets a labeled index, which is 1 for labeled or 0 for not labeled.
- `Selected` gets or sets a selected index, which is 1 for selected or 0 for not selected.

The following examples illustrate these conditions:

```
Excluded( Row State() ); // returns 1 if current row is excluded, 0 if not
Hidden();                // returns 1 if current row is hidden, 0 if not
Labeled( Row State() );  // returns 1 if current row is labeled, 0 if not
Selected();              // returns 1 if current row is selected, 0 if not

Excluded( Row State() ) = 1; // exclude current row
Hidden() = 0;                // unhide current row
Labeled( Row State() ) = 1;  // label current row
Selected() = 0;              // deselect current row
```

Remember that these functions assume the argument `Row State()` if none is given.

`Excluded State`, `Hidden State`, `Labeled State`, and `Selected State` do the reverse; they get or set a row state condition as true or false according to the argument. Nonzero values set the row state to true, and zero values set it to false. Missing values result in no change of state.

```
Row State() = Excluded State( 1 ); // change current row state to excluded
Row State() = Hidden State( 0 );   // change current row state to not hidden
Row State() = Labeled State( 1 );  // change current row state to labeled
Row State() = Selected State( 0 );
// change current row state to not selected
```

Notice that the first two expressions above replace the row state with just the exclusion or just the unhiding, so that any preexisting row state characteristics are lost. More commonly, you would issue the `state` commands for all of the characteristics that you want inside a `Combine States`. The following example changes the third row to hidden and a green square marker:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Row State( 3 ) = Combine States(
  Color State( 4 ),
  Marker State( 3 ),
  Hidden State( 1 ) );
```

Another common way to use a `-State` command would be in a row state data column whose values could be added to the row state (for cumulative characteristics). The following example excludes each odd numbered row:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "myExcl",
  Row State,
```

```
    Set Formula( Excluded State( Modulo( Row(), 2 ) ) )
);
For Each Row( Row State() = Combine States( :myExcl, Row State() ) );
```

Clear Row States

`Clear Row States()` removes row states from the data table. The following script assigns a row state to each row and then removes the row state.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row( Row State() = Color State( Row() ) );
Wait( 3 ); // pause for demonstration purposes
dt << Clear Row States();
```

To clear row states for selected rows, select the rows first and then clear the selected row states.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row( Row State() = Color State( Row() ) );
dt << Select Rows( [ 5 ] );
Wait( 3 ); // pause for demonstration purposes
dt << Clear Selected Row States();
```

Colors and Markers

This section discusses conditions that have many choices. These conditions include: coloring, adding markers, different hues, or different shades to rows.

`Color Of()` returns or sets the color index. The color index is a number from the JMP color map that corresponds to the *`row state`*, or a missing value if there is no assigned color.

```
Color Of( Row State() );   // returns color index for current row
Color Of() = 4;           // set current row to Color 4
```

Similarly, `Marker Of` returns or sets the marker index. The marker index is a number from the JMP marker map that corresponds to the active marker, or a missing value if there is no assigned marker.

```
Marker Of();                  // returns marker index for current row
Marker Of( Row State() ) = 4; // set current row to Marker 4
```

Both `Color Of()` and `Marker Of()` accept any row state expression or column or `Row State()` as arguments. They also assume the argument `Row State()` if none is given (some examples are shown with, and some without).

`Color State()` and `Marker State()` are similar to `Color Of()` and `Marker Of()`, except they work in the opposite direction. Where the `-Of` functions turn actual states into indices, the `-State` functions turn indices into states.

```
Row State() = Color State( 4 );  // change current row to green
```

```
Row State() = Marker State( 4 ); // change current row to the diamond marker
```

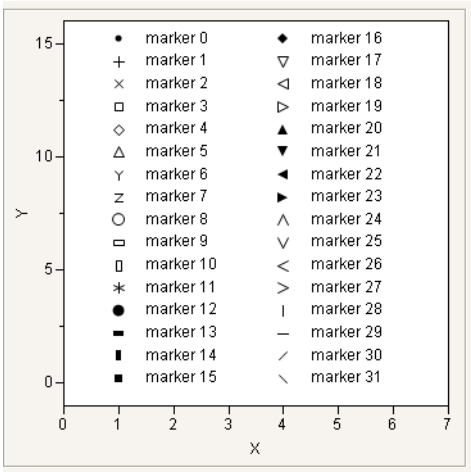
Notice that the last two commands replace the row state with just the color or just the marker, so that any preexisting row state characteristics are lost. More commonly you would issue the `-State` commands for all the characteristics that you want inside a `Combine States()`. The following example puts a green square marker in the third row and hides the row:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Row State( 3 ) = Combine States(
   Color State( 4 ),
   Marker State( 3 ),
   Hidden State( 1 ) );
```

The following script shows the standard JMP markers, which are numbered 0–31. Indices outside the range 0–31 have undefined behavior.

```
New Window( "Markers",
   Graph Box(
      FrameSize( 300, 300 ),
      Y Scale( -1, 16 ),
      X Scale( 0, 7 ),
      For(
         i = 0;
         jj = 15;,
         i < 16;
         jj >= 0;,
         i++;
         jj--;, // 16 rows, 2 columns
         Marker Size( 3 );
         Marker( i, {1, jj + .2} ); // markers 0-15
         Marker( i + 16, {4, jj + .2} ); // markers 16-31
         Text( {1.5, jj}, "marker ", i ); // marker labels 0-15
         Text( {4.5, jj}, "marker ", i + 16 ); // marker labels 16-31
      )
   )
);
```

Figure 9.18 JMP Markers



Tip: For more information about this script, see "Display Trees".

Note the following about colors:

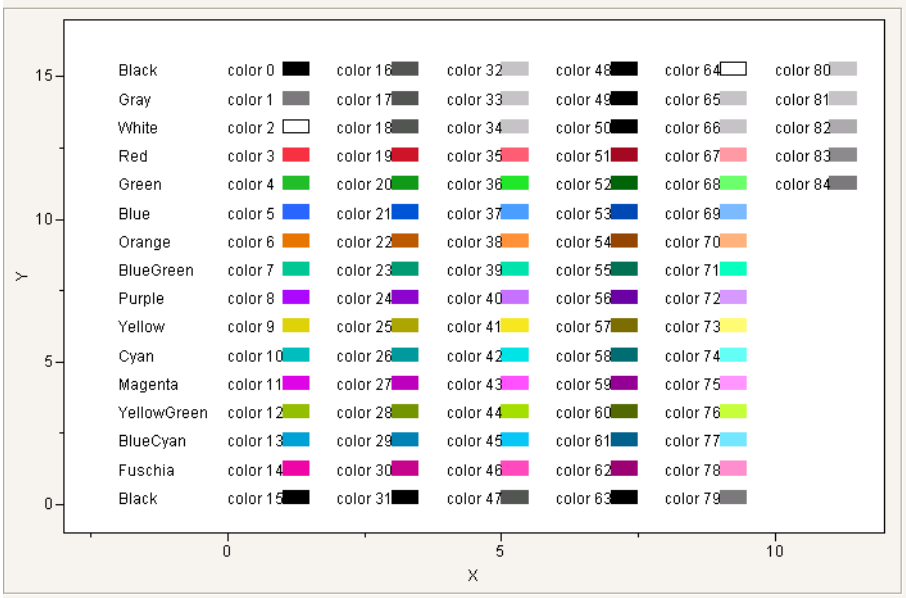
- JMP colors are numbered 0 through 84.
- The first 16 named colors are the basic colors. See the script below.
- Numbers higher than 16 are darker or lighter shades of the basic colors.
- Indices outside the range 0–84 have undefined behavior.
- For more information about using JMP colors, see "Specify Colors".

The following script illustrates the standard JMP colors:

```
Text Color( 0 );
New Window( "Colors",
   Graph Box(
      FrameSize( 640, 400 ),
      Y Scale( -1, 17 ),
      X Scale( -3, 12 ),
      k = 0;
      For( jj = 1, jj <= 12, jj += 2,
         l = 15;
         For( i = 0, i <= 15 & k < 85, i++,
            thiscolor = Color To RGB( k );
            Fill Color( k );
            thisfill = 1;
            If( thiscolor == {1, 1, 1},
               Pen Color( 0 );
               thisfill = 0;
```

```
            ,
               Pen Color( k )
            );
            Rect( jj, l + .5, jj + .5, l, thisfill );
            Text( {jj - 1, l}, "color ", k );
            k++;
            l--;
         );
      );
      jj = -2;
      color = {"Black", "Gray", "White", "Red", "Green", "Blue", "Orange",
"BlueGreen",
      "Purple", "Yellow", "Cyan", "Magenta", "YellowGreen", "BlueCyan",
"Fuschia", "Black"};
      For(
         i = 0;
         l = 15;, i <= 15 & l >= 0,
         i++;
         l--;,
         Text( {jj, l}, color[i + 1] )
      );
) );
```

Figure 9.19 JMP Colors



If you prefer to use RGB values, each color should be a list with percentages for each color in red, green, blue order. For example, the following percentages produce a teal color:

```
Pen Color( {.38,.84,.67} ); // a lovely teal
```

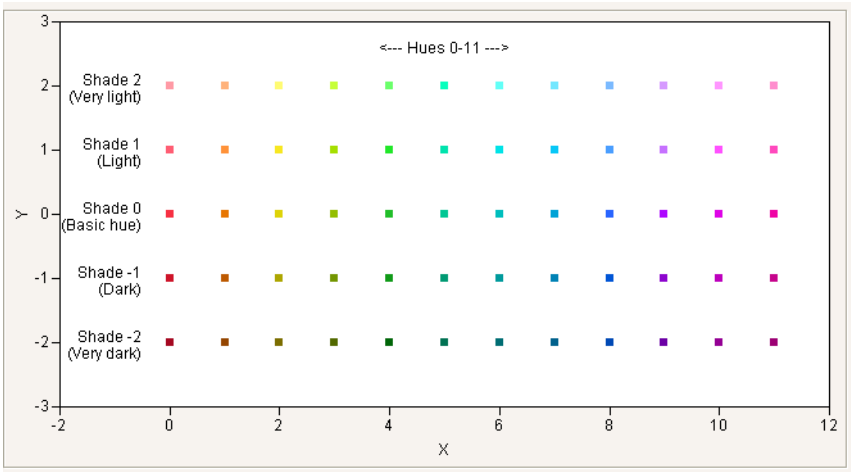
Hue and Shade Example

`Hue State` and `Shade State` together are an alternative to `Color State` for choosing colors. You cannot select black, white, or the shades of gray when you use `Hue State`. For these, you must use `Shade State` alone, or `Color State`.

The following script demonstrates how hue and shade values relate to colors:

```
New Window( "Hues and Shades",
   Graph Box(
      FrameSize( 600, 300 ),
      Y Scale( -3, 3 ),
      X Scale( -2, 12 ),
      k = 0;
      For( h = 0, h < 12, h++,
         For( s = -2, s < 3, s++,
            myMk = Combine States( Hue State( h ), Shade State( s ), Marker
   State( 15 ) );
            Marker Size( 3 );
            Marker( myMk, {h, s} );
         )
      );
      Text( Center Justified, {5, 2.5}, " <--- Hues 0-11 ---> " );
      Text( Right Justified,
         {-.5, -2}, "Shade -2", {-.5, -2.25}, "(Very dark)",
         {-.5, -1}, "Shade -1", {-.5, -1.25}, "(Dark)",
         {-.5, 0}, "Shade 0", {-.5, -.25}, "(Basic hue)",
         {-.5, 1}, "Shade 1", {-.5, .75}, "(Light)",
         {-.5, 2}, "Shade 2", {-.5, 1.75}, "(Very light)"
      );
   )
);
```

Figure 9.20 Hues and Shades



There are no `-Of` functions for `Hue` and `Shade`. `Color Of` returns the equivalent `Color State` index for a color row state that has been set with `Hue State` or `Shade State`. For example, the following example gives rows 4 and 5 the same dark red marker:

```
Row State( 4 ) = Combine States( Hue State( 0 ), Shade State( -1 ), Marker
   State( 12 ) );
Row State( 5 ) = Combine States(
   Color State( Color Of( Row State( 4 ) ) ),
   Marker State( Marker Of( Row State( 4 ) ) )
);
```

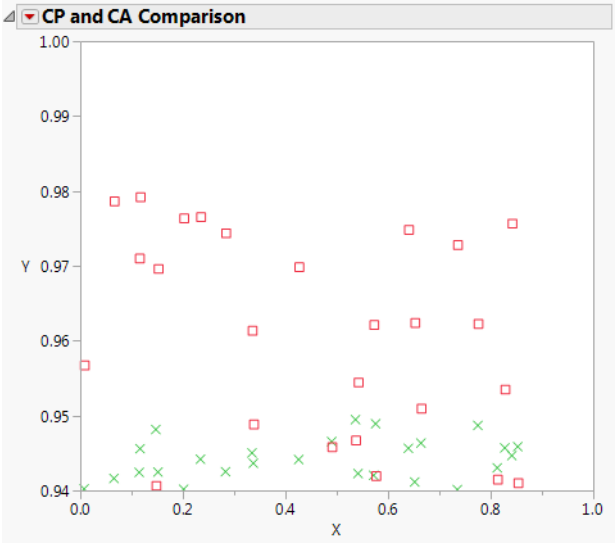
Row State and Matrices Example

In the following example, row state values are prepared ahead and passed to the `Marker` routine, along with matrices of coordinates.

```
dt = New Table( "Artificial CP and CA data",
   Add Rows( 26 ),
   New Column( "cover_cp",
      Numeric,
      "Continuous",
      Formula( Random Uniform() / 100 + 0.94 )
   ),
   New Column( "cover_ca",
      Numeric,
      "Continuous",
      Formula( Random Uniform() * 0.04 + 0.94 )
   ),
   New Column( "p", Numeric, "Continuous", Formula( Random Uniform() ) )
);
```

```
dt << Run Formulas;
greenMark = Combine States( Marker State( 2 ), Color State( 4 ) );
redDiamond = Combine States( Marker State( 3 ), Color State( 3 ) );
New Window( "CP and CA Comparisons",
   Graph Box(
      Title( "CP and CA Comparison" ),
      FrameSize( 400, 350 ),
      X Scale( 0, 1 ),
      Y Scale( 0.94, 1 ),
      For Each Row(
         Marker( greenMark, {p, cover_cp} );
         Marker( redDiamond, {p, cover_ca} );
      )
   )
);
```

Figure 9.21 Example of Row States Values and Matrices



Accessing Data Values

The typical way to work with values in a data table is to follow these steps:

1. Set up the data table whose values you want to access as the current data table. Or, if you already have a data table reference, you can simply use that reference. See "Set the Current Data Table".

2. Specify the row or rows whose values you want to access and specify the column name that contains the values that you want to access. See "Set or Get Values by Column Name".

The following example opens the Big Class.jmp sample data table (making it the current data table), and then specifies row 2 in the weight column. A value of 123 is returned in the log, which is the weight for Louise in row 2.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt:weight[2];
   123
```

Set or Get Values by Column Name

The easiest way to refer to a column is by its name. If you have a global variable and a column with the same name, to prevent ambiguity, scope the column name with the `:` prefix operator.

To set the value of a cell in the current row, provide the column name and the new value.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Row() = 5; // select row 5
dt:age = 19; // set the age in row 5 to 19
dt:name = "Sam"; // set the name in row 5 to Sam
```

To set the value of a cell in row 10, specify the row number in a subscript.

```
dt:age[10] = 20; // set the value of age in row 10 to 20
```

An empty subscript refers to the current row, so :age[ ] is the same as :age.

To get the value of a cell in row 16, specify the column name.

```
Row() = 16; // select row 16
myGlobal = :age; // store the age as a variable
:age; // returns 14, the age in row 16
Show( :age ); // returns age = 14;
```

To get the value of a cell in a specific row, include the column name and row number. Both of the following examples return 13, the value of age in row 12 of Big Class.jmp:

```
:age[12];
myGlobal = :age[12];
```

To get a value in a data column reference, use `Column()` and `As Column()` to get the value in a data column reference. See "Access Cell Values through Column References".

Notes:

- If you do not specify the row number, the current row is selected.
- An empty subscript, such as `:age[ ]`, refers to the current row.

- Be careful that you are subscripting to a table row that exists. The default row number is zero, so statements like `:name` that refer to row zero generate an `Invalid Row Number` error.

Additional Ways to Access Data Values

There are other ways to specify a data table, row, and column. You can specify all three items in one expression by using an infix operator and a subscript:

```
dt:age[2] = 12; // table, column, and row
```

If you want to target multiple rows, you can use subscripts with a list or matrix of row numbers.

```
age[i] = 3;
age[{3, 12, 32}] = 14;
list = age[{3, 12, 32}]; // results in a list
vector = age[1 :: 20]; // results in a matrix
dt[1,1] = dt[2,1]; // in Big Class.jmp, “KATIE” (row 1, column 1) is now
  “LOUISE” (from row 2, column 1)
dt[0,{height}] = dt[0,{weight}]; // everybody’s height is their weight
dt[1,{height,weight}] = dt[2,4::5]; // row 1 height and weight is from Louise
dt[[5 3 1], 0] = .; /* set rows 5, 3, and 1 to missing. Using out-of-order
  subscripts makes more sense if the right hand side is a matrix or a data
  table with subscripts */
```

Note about Changing Values

Whenever you change values in a data table, messages are sent to the displays to keep them up-to-date. However, if you have thousands of changes in a script, this increases the time it takes to complete the updates.

In order to speed up changes, use `Begin Data Update` before the changes to block these update messages. Use `End Data Update` after the changes have been completed to release the messages and update the displays.

```
dt << Begin Data Update;
...<many changes>...
dt << End Data Update;
```

Be sure to always send the `End Data Update` message, otherwise the display is not updated until forced to do so in some other way.

Note: `Begin Data Update` does not affect the data refresh due to some other table manipulations. For example, when you delete or add columns, the data table is updated and then the data update begins.

Add Metadata to a Data Table

Data tables store observation data, or measurements of various variables on a specific set of subjects. However, JMP data tables can also store *metadata*, or data about the data.

- "Data Table Variables"
- "Table Scripts"
- "Formulas"
- "Delete Metadata"

Data Table Variables

Table variables are for storing a single text string value, such as "Notes". To understand how variables work, first get its existing value by sending a `Get Table Variable` message:

```
dt = Open( "$SAMPLE_DATA/Solubility.jmp" );
dt << Get Table Variable( "Notes" );
   "Chemical compounds were measured for solubility in different solvents. This
      table shows the log of the partition coefficient (logP) of 72 organic
      solutes in 6 aqueous/nonpolar systems."
```

Now change the existing value of the string using `Set Table Variable` and then use `Get Table Variable` again to check that the string has been updated:

```
dt << Set Table Variable( "Notes", "Solubility of chemical compounds" );
dt << Get Table Variable( "Notes" );
   "Solubility of chemical compounds"
```

The following example adds two new table variables to a data table:

```
dt = Open( "$DOCUMENTS/Big Class.jmp" );
myvar = "This is my version of the JMP Big Class sample data.";
dt << Set Table Variable( "key1", myvar );
dt << Set Table Variable( "key2", myvar );
```

Notice that setting the value creates a new variable only if one by the given name does not already exist. If you add two table variables with the same name, only one variable is created.

Table Scripts

You can add a new script, run a script, or get a script using JSL.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Script( "Bivariate Example", Bivariate( Y( :weight), X( :height ) )
   );
dt << Get Table Variable( "Bivariate Example" );
```

```
    // returns Bivariate( Y( weight ), X( height ) )
dt << Run Script( "Bivariate Example" )
```

The following example replaces the contents of the table script that was created by the preceding script:

```
dt << Set Property( "Bivariate Example", Bivariate( Y( :weight ), X( :height
  ), Fit Line ) ); // include Fit Line in the Bivariate Example script
```

Suppose you want a text representation of a data table, perhaps to e-mail to a colleague or to use as part of a script. You can obtain a script that reconstructs the information in a data table with `Get Script`. The following example opens Big Class.jmp and prints the data, table variables, and column properties to the log. A portion of the output is shown here:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Get Script;
   New Table( "Big Class",
      Add Rows( 40 ),
      New Script(
         ["en" => "Distribution",...],
         Distribution(
            Continuous Distribution( Column( :weight ) ),
            Nominal Distribution( Column( :age ) )
         ) ),
```

`Delete Scripts` deletes one or more scripts.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Delete Scripts( "Distribution" );
```

In JMP versions prior to 14, use `Delete Property` to delete a script.

Note: The `Names Default to Here` setting is false by default for table scripts. If a table script specifies `Names Default to Here(1)`, this setting lasts only as long as the table script runs. After the script finishes running, the setting is reset to whatever it was before the script ran.

Options for Downloaded Data Tables that Contain Scripts

You are prompted to open or examine downloaded data tables that contain scripts. This help you prevent potentially damaging scripts from running on your computer. In the `Open()` command, you can control this behavior.

To open the data table and enable scripts to run:

```
Open( "$DOWNLOADS/file.jmp", Quarantine Action( "Allow Scripts" ) );
```

To open the data table and block scripts:

```
Open( "$DOWNLOADS/file.jmp", Quarantine Action( "Block Scripts" ) );
```

To display the default prompt window to open or examine the file:

```
Open( "$DOWNLOADS/file.jmp", Quarantine Action( "Show Dialog" ) )
```

To enable a `Try()` expression to capture the error:

```
Try( Open( "$DOWNLOADS/file.jmp", Quarantine Action( "Do Not Open" ) ),
Show( exception_msg ) );
```

If the user selects Examine, the scripts are disabled.

Notes:

- The Block Scripts and Show Dialog options don't work if the domain that you downloaded the data table from is on your Windows Internet Options Trusted sites list.
- After you select Open for a quarantined data table, the data table is trusted until the next time you open JMP.

Prompting to Run a Script upon Opening a Data Table

You can prompt to open a table script named `On Open` or `OnOpen` when the data table opens. The choice is remembered each time you open the data table in the current JMP session.

To create an `On Open` script, perform one of the following actions:

- Create the script using the **Save Script > To Data Table** option, double-click the script name, and change the name to `On Open`.
- Store the script using a `New Script` message.

In this example, you create an OnOpen script in Big Class.jmp.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Script(
   "OnOpen", // create the script
   sortedDt = Data Table( "Big Class" ) << Sort( By( :name ), Output Table
   Name( "Sorted Big Class" ) ) // sort the data and put it in a new data table
);
```

The JMP preference called Evaluate OnOpen Scripts determines when the script runs. By default, the user is prompted to run the script. You can set the preference to always run the `On Open` script or to prevent it from running:

```
Preference( Evaluate OnOpen Scripts( "always" ) );
Preference( Evaluate OnOpen Scripts( "never" ) );
Preference( Evaluate OnOpen Scripts( "prompt" ) ); // default setting
```

`On Open` scripts that execute other programs are never run. As a safety precaution, you might consider suppressing automatic execution when opening data tables that you receive from others.

Notes:

- If the OnOpen script is in a downloaded data table, you must specify `Quarantine Action( "Allow Scripts" )` in the `Open()` expression to run the OnOpen script automatically. See "Options for Downloaded Data Tables that Contain Scripts".
- When you create a new data table in a script and include the `On Open()` function, `On Open()` is called after the data table is created.
- Some operations in an OnOpen script are considered unsafe, such as saving the current data table. Run an external script to save that table instead of opening a table to run the OnOpen script.

Work with Groups of Table Scripts

Suppose that you want to put related table scripts in a group to save room in the list of data table scripts. Use `Group Scripts`, specify the group name, and then specify the script names to put in the group.

```
dt = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
dt << Group Scripts(
   "Maps",
   {"Bubble Plot Street Map", "Graph Builder Street Map: Traffic Incidents",
   "Graph Builder Street Map Zoomed"}
);
dt << Group Scripts(
   "Tabulate",
   {"Tabulate: Category and Description", "Tabulate: Category, Summary,
   Percentage",
   "Tabulate: Category and Resolution", "Tabulate: Resolution and
   Description"}
);
```

Move a Script Group

`Move Script Group` rearranges the table script groups using the `to first`, `to last`, `after(script)`, and `after(group)` arguments:

```
dt << Move Script Group( "Tabulate", after( "Maps" ) );
dt << Move Script Group( "Tabulate", to last);
```

Get Groups and Group Names

To get the list of table scripts in the group, use `Get Script Group`. Then you can perform an operation such as running a script in the group.

```
tb = dt << Get Script Group( "Tabulate" );
dt << Run Script( tb[2] ); // runs the second script in the Tabulate group
```

`Get Script Groups Names` returns the group names in a list:

```
gn = dt << Get Script Groups Names; // returns {"Maps", "Tabulate"}

// move the Maps group to the end of the table scripts
dt << Move Script Group(gn[1], to last);
```

Select a Script Group

`Select Script Group` selects the named table script group, a list of scripts in a group, or all scripts.

```
dt << Select Script Group; // select all scripts groups
dt << Select Script Group( "Tabulate" ); // select Tabulate group
dt << Select Script Group( {"Tabulate", "Maps"} );// select both groups
```

Ungroup Scripts

`Ungroup Scripts` removes specified table scripts or group from the group. You can send the message to the name of the script group or a list of scripts.

```
dt << Ungroup Scripts( "Tabulate" );
dt << Ungroup Scripts(
   {"Tabulate: Category and Description", "Tabulate: Category, Summary,
   Percentage",
   "Tabulate: Category and Resolution", "Tabulate: Resolution and
   Description"}
);
```

Rename a Script Group

You can rename a table script group using `Rename Script Group`, specifying the existing group name, and specifying the new group name.

```
dt << Rename Script Group( "Maps", "Street Maps" );
```

Delete a Script Groups

To delete a table script group, you get it first and then delete it.

```
a = dt << Get Script Group( "Street Maps" );
dt << Delete Scripts( a );
```

Formulas

The message `Suppress Formula Eval` takes a Boolean argument to specify whether formula evaluation should be suppressed or not. You might want to suppress evaluation if you plan to make numerous changes to the data table and do not want to wait for formula updates between steps.

```
dt << Suppress Formula Eval( 1 );
dt << Suppress Formula Eval( 0 ) ;
```

To accomplish the same effect for all data tables, use the `Suppress Formula Eval` command to turn off formulas globally. This is the same as the message above, except that you do not send it to a data table object.

```
Suppress Formula Eval( 1 ); // make formulas static globally
Suppress Formula Eval( 0 ); // make formulas dynamic globally
```

Note that formulas are not evaluated when they are installed in the column. Even when you force evaluation, they end up being evaluated again in the background. This can be a problem for scripts if they depend on the column having the values while the script is running. If you need a mechanism to control evaluation, use the `EvalFormula` command or the `Run Formulas` command.

To force a single column to evaluate, send an `Eval Formula` command to the column. You can even do this inside the command to create the column, after the formula clause:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Ratio",
   Numeric,
   Formula( :height / :weight ),
   Eval Formula
);
```

`dt << Run Formulas` performs all pending formula evaluations, including evaluations that are pending as a result of evaluating other formulas. This function is useful when you have a whole series of columns to run.

Tip: This method is preferred over `EvalFormula`. Although `EvalFormula` evaluates the formulas, it does not suppress the background task from evaluating them again. The background task takes great care to evaluate the formulas in the right order.

If you send the `Run Formulas` command to a data column, the evaluation is done at the time of the command, but it does not suppress the scheduled evaluations that are pending. Therefore, formulas might end up being evaluated twice if you also send the command to the data table and the data column. Being evaluated twice might be desirable for formulas that have random function in them, or it might be undesirable if they depend on randomization seeds being set. If you use random numbers and use the `Random Reset(seed)` feature to make a replicable sequence, then use the `Run Formulas` command, because it avoids a second evaluation.

Note: All platforms send a `Run Formulas` command to the data table to assure that all formulas have finished evaluating before analyses start.

Set Values without a Formula

`col << Set Each Value(expression)` evaluates the expression for each row of the data table and assigns the result to the column. It does not store the expression as a formula.

Delete Metadata

You can delete table variables, table properties (such as a script or a variable), and formulas from the data table, using the following commands:

```
dt << Delete Table Variable( name );
dt << Delete Table Property( name|{property1, property2, ...});
col << Delete Formula;
col << Delete Property( name|{property1, property2, ...} );
col << Delete Column Property( name );
```

Calculations

This section discusses functions for pre-evaluated columnwise and rowwise statistics and shows how JSL expressions work behind the scenes in the JMP formula calculator.

Pre-Evaluated Statistics

The following functions are special, pre-evaluated functions: `Col Maximum`, `Col Mean`, `Col Minimum`, `Col N Missing`, `Col Number`, `Col Quantile`, `CV` (Coefficient of Variation), `Col Standardize`, `Col Std Dev`, `Col Sum`, `Maximum`, `Mean`, `Minimum`, `NMissing`, `Number`, `Std Dev`, and `Sum`.

Note: Statistics are also computed with `Summarize` ("Store Summary Statistics in Global Variables"). Although the named arguments to `Summarize` have the same names as these pre-evaluated statistic functions, they are *not* calling the pre-evaluated statistic functions. The resemblance is purely coincidental.

All the statistics are *pre-evaluated*. That is, JMP calculates them once over the rows or columns specified and thereafter uses the results as constants. Because they are computed once and then used over and over again, they are more efficient to use in calculations than the equivalent formula-calculated results.

When JMP encounters a pre-evaluated function in a script, it immediately evaluates the function and then uses the result as a constant thereafter. Therefore, pre-evaluated functions enable you to use columnwise results for rowwise calculations. For example, if you use `Col Mean` inside a column formula, it first evaluates the mean for the column specified and then uses that result as a constant in evaluating the rest of the formula on each row. A formula might standardize a column using its pre-evaluated mean and standard deviation:

```
( Height-Col Mean( Height ) ) / Col Std Dev( Height )
```

For the Big Class.jmp data, `Col Mean(Height)` is 62.55 and `Col Std Dev(Height)` is 4.24. So for each row, the formula above would subtract 62.55 from that row's height value and then divide by 4.24.

Note: Pre-evaluated functions disregard the excluded row state, meaning that any excluded rows are included in calculations. For summary statistics that obey row exclusion, use the Distribution platform.

Columnwise Functions

The functions whose names begin with "`Col`" all work *columnwise*, or down the values in the specified column, and return a single number. For example, `Col Mean(height)` finds the mean of the values in all the rows of the column height and returns it as a scalar result. Some examples include the following:

```
Average Student Height = Col Mean( height );
Height Sigma = Col Std Dev( height );
```

With the `Col` functions, column properties such as Missing Value Codes assign data values that produce incorrect calculations. Suppose that the Missing Value Codes column property is assigned to the `x1` column to treat "999" as a missing value. Another column includes a formula that calculates the mean. To use the value "999" instead of a missing value to calculate the mean, refer to `Col Stored Value()` in the formula:

```
Mean( Col Stored Value( :x1 ), :x2, :x3 )
```

Rowwise Functions

The functions *without* "Col" listed below work *rowwise* across the values in the variables specified and return a column result. For example, `Mean(height, weight)` finds the mean of the height and weight for the current row of the data table. The rowwise statistics are valid only when used in an appropriate data table row context. The following are some possibilities:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );

// scalar result for row 7 assigned to JSL global variable
Row() = 7;
::scalar = Mean( height, weight );

// formula column created in data table
dt << New Column( "Scaled Ht-Wt Ratio",
   Formula( Mean( height, weight ) / age )
);

vector = J( 1, 40 ); // create a 1x40 matrix to hold results
For Each Row( vector[Row()] = Mean( height, weight ) ); // fill the vector
```

Rowwise functions can also take vector (column matrix) or list arguments:

```
myMu = Mean( [1 2 3 4] );
mySigma = Std Dev( {1, 2, 3} );
```

Calculator Formulas

You can store formulas in columns that are automatically evaluated to create the values in the cells of the column. If you open the formula, you get a calculator interface to edit the formula structurally. However, the formula is implemented with JSL, and you can obtain the text JSL form of any expression in the calculator by double-clicking it. The text can be edited, and when it is de-focused, it is compiled back into the structural form.

Note: There is no difference between a formula column created through the calculator window and one created directly through JSL with commands such as `New Column(..., Formula(...))` or `Col << Formula(...)`.

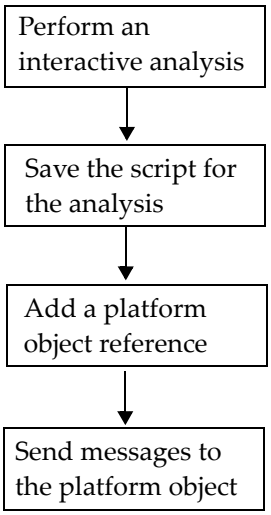
Chapter 10

Scripting Platforms

Create, Repeat, and Modify Analyses

If you run the same analysis frequently, you can script it to automate the process. Anyone can then run the script, ensuring the same results every time. To get started, perform your analysis interactively as you normally would with JMP, and then save a script that re-creates that analysis. You can modify the script to further customize it.

Figure 10.1 Typical Workflow for Scripting Platforms



This chapter is about scripting platforms, not reports. *Platform* object references and *report* object references receive different types of JSL messages. Platforms can run tests, draw plots, and so on. Reports can copy pictures, select display boxes, or close outline nodes. To learn about scripting reports, see the section "Navigate JMP Reports".

Tip: For additional scripting help, see the JMP Scripting Index (**Help > Scripting Index**) and the *JSL Syntax Reference*.

Contents

Example of Scripting a Platform . . . 477
Send Messages to a Platform . . . 480
Conventions for Messages and Arguments . . . 480
Send Multiple Messages . . . 481
Find Messages for Objects . . . 482
Interpret the Show Properties List . . . 483
Send Messages to an Object . . . 484
Specify Which Columns to Analyze . . . 485
Create Column References . . . 485
Specify Multiple Column Names at Once . . . 485
Enable Users to Specify Columns . . . 486
Specify a By Variable . . . 486
Filter by Value using a Where Statement . . . 489
Enable User Input . . . 490
Execute Embedded Red Triangle Options . . . 490
Make Platforms Invisible . . . 491
Specify Report Titles . . . 493
Common Messages for Platform Windows . . . 493
Scripting Considerations by Platform . . . 493
Categorical . . . 493
Control Charts . . . 494
Distribution . . . 498
DOE . . . 498
Fit Model . . . 500
Formula Depot . . . 505
Graph Builder . . . 506
Neural and Neural Net . . . 508
Partial Least Squares and PLS . . . 508
Process Capability . . . 508
Scatterplot 3D . . . 509
Text Explorer . . . 509
Explore Outliers . . . 510
Scripting-Only Messages and Arguments by Platform . . . 510
Bivariate Script-Only Messages . . . 512
Boosted Tree Script-Only Messages . . . 513
Bootstrap Forest Script-Only Messages . . . 514
Bubble Plot Script-Only Messages . . . 515
Choice Script-Only Messages . . . 515
Clustering Script-Only Messages . . . 516

Contingency Script-Only Messages 516
Contour Plot Script-Only Messages 516
Contour Profiler Script-Only Messages 516
Control Chart Script-Only Messages 517
Control Chart Builder Script-Only Messages 517
Cumulative Damage Script-Only Messages 517
Custom Profiler Script-Only Messages 518
Degradation Script-Only Messages 518
Destructive Degradation Script-Only Messages 518
Discriminant Script-Only Messages 518
Distribution Script-Only Messages 519
EMP Measurement Systems Analysis Script-Only Messages 519
Explore Missing Values Script-Only Messages 520
Explore Outliers Script-Only Messages 520
Factor Analysis Script-Only Messages 520
Fit Life by X Script-Only Messages 520
Fit Model Script-Only Messages 521
Fit Parametric Survival Script-Only Messages 523
Functional Data Explorer Script-Only Messages 524
Gaussian Process Script-Only Messages 524
K Means Cluster Script-Only Messages 524
K Nearest Neighbors Script-Only Messages 524
Latent Class Analysis Script-Only Messages 524
Life Distribution Script-Only Messages 525
Make Validation Column Script-Only Messages 526
MaxDiff Script-Only Messages 526
Mixture Profile Script-Only Messages 526
Model Screening Script-Only Messages 527
Multivariate Script-Only Messages 527
Naive Bayes Script-Only Messages 527
Neural Script-Only Messages 527
Nonlinear Script-Only Messages 528
Normal Mixtures Script-Only Messages 529
Oneway Script-Only Messages 529
Parallel Plot Script-Only Messages 529
Pareto Plot Script-Only Messages 530
Partial Least Squares Script-Only Messages 530
Partition Script-Only Messages 530
Principal Components Script-Only Messages 531
Process Capability Script-Only Messages 531

Process Screening Script-Only Messages 531
Profiler Script-Only Messages 532
Reliability Forecast Script-Only Messages 532
Reliability Growth Script-Only Messages 532
Response Screening Script-Only Messages 533
Scatterplot 3D Script-Only Messages 533
Structural Equation Models Script-Only Messages 534
Support Vector Machines Script-Only Messages 535
Surface Plot Script-Only Messages 536
Tabulate Script-Only Messages 539
Ternary Plot Script-Only Messages 539
Text Explorer Script-Only Messages 539
Time Series Script-Only Messages 540
Uplift Script-Only Messages 540
Variability Chart Script-Only Messages 541

Example of Scripting a Platform

If you are not familiar with JSL, it can be helpful to start by interactively performing your analysis in JMP, and then saving the analysis to a script. JMP creates the script for you, and you can then add to it or modify it as desired, by creating a platform object and sending it messages.

Overview

This example takes you through these steps:

1. "Launch the Analysis and Save the Script to a Window"
2. "Add a Platform Reference Object and Send Messages"
3. "Add a Data Table Reference and Save the Script"

Launch the Analysis and Save the Script to a Window

1. Select **Help > Sample Data Library** and open Big Class.jmp.

   This fictional data set contains the names, ages, sex, heights, and weights for 40 students.
2. Select **Analyze** > **Fit Y by X**.
3. Select height and click **Y, Response**.
4. Select age and click **X, Factor**.
5. Click **OK**.

   A Oneway plot of students' height by age appears.
6. Click the Oneway Analysis red triangle and select the following options:
   - **Means/Anova**
   - **Compare Means > Each Pair, Student's t**
7. Click the Oneway Analysis red triangle and select **Save Script > To Script Window**.

Here is the resulting script:

```
Oneway(
   Y( :height ), X( :age ),
   Each Pair( 1 ),
   Means( 1 ),
   Mean Diamonds( 1 )
);
```

Notice the following:

- The script begins with a call to the `Oneway()` function, which returns a Oneway object.

- The parentheses after the `Oneway()` function contain arguments that tell the Oneway function how to make the object.
  - The first two arguments, `Y` and `X`, are required at launch.
  - The next three arguments are optional: `Each Pair`, `Means`, and `Mean Diamonds`.
  - When you selected the Each Pair, Student's t option, the Each Pair feature was turned on.
  - When you selected the Means/Anova option, the Means and Mean Diamonds features were turned on.
- The scripting equivalent of turning an option on or off is the Boolean argument 0 (off) or 1 (on).

Add a Platform Reference Object and Send Messages

1. (Optional) Right-click in the script and select **Show Embedded Log**.

   Output that appears in the log is now easily visible below the script.
2. Choose a JSL variable to remember the Oneway platform object.

   In this example, use the name `oneObj`:

   ```
   oneObj = Oneway(
     Y( :height ),
     X( :age ),
     Each Pair( 1 ),
     Means( 1 ),
     Mean Diamonds( 1 )
   );
   ```

   This gives you a way to address the platform object and send messages to it.
3. Click **Run Script** .

   This creates the Oneway object and sets the variable `oneObj`. Now, send a message to tell the platform to turn on the Unequal Variances report.
4. At the end of the script, add the following line: `oneObj << Unequal Variances(1)`, as shown here:

   ```
   oneObj = Oneway(
     Y( :height ),
     X( :age ),
     Each Pair( 1 ),
     Means( 1 ),
     Mean Diamonds( 1 )
   );
   oneObj << Unequal Variances( 1 );
   ```
5. Highlight the Unequal Variances line of JSL and click **Run Script** .

Tip: You can also press Ctrl+R to run a highlighted line of JSL.

At the bottom of the report window, the Tests That the Variances Are Equal report now appears. Now, suppose that you want to see only that report (and the graph), and you want to close the other reports.

6. Add the following lines to the script:

```
rep = Report( oneObj );
rep["Oneway Anova"] << Close( 1 );
rep["Means Comparisons"] << Close( 1 );
```

The `Report()` function returns the report object for the Oneway platform, and stores a reference to the report in the JSL variable called `rep`. You can send messages or perform actions on the report object, such as closing specific report outlines.

7. Highlight these last 3 lines of the script and click **Run Script** .

In the report window, the Oneway Anova and the Means Comparisons reports are now closed. Only the initial graph and the Tests that the Variances are Equal report (which includes the Welch's Test report) appear.

If you close the Big Class.jmp sample data table at this point and then try to run the script, you are prompted to open a data table. A best practice is to precede the call to the Oneway platform with an `Open()` function to open the associated data table each time the script is run.

Add a Data Table Reference and Save the Script

1. Add the following line as the first line in the script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
```

If the data table is not open when the script is run, this line opens the associated data table.

Your finished script should look like this:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
oneObj = Oneway(
   Y( :height ),
   X( :age ),
   Each Pair( 1 ),
   Means( 1 ),
   Mean Diamonds( 1 )
);
oneObj << Unequal Variances( 1 );
rep = Report( oneObj );
rep["Oneway Anova"] << Close( 1 );
rep["Means Comparisons"] << Close( 1 );
```

Once you complete your script, save it for future access.

2. From the script window, select **File** > **Save As**. Enter a name for your script and save it to any directory.

Send Messages to a Platform

Before you can send messages to a platform, you need to create a platform object that you can send the messages to. Use the Send operator (<<).

The following line keeps a reference to a Oneway platform in the JSL variable called `oneObj`:

```
oneObj = Oneway( Y( height ), X( age ) );
```

Then, add the Send operator to send the platform object a message. In this example, the message is to turn on the Unequal Variances report:

```
oneObj << Unequal Variances( 1 );
```

To see which messages can be sent to a specific object, do one of the following:

- Use the autocomplete shortcut. On a new line, enter the object name, then the Send operator (<<), and press Ctrl+space (Windows) or Ctrl+Option+space (macOS). The following example shows the messages that you can send to the platform object `oneObj` on Windows:

  ```
  oneObj << (Ctrl+space)
  ```

  A list of messages appears. Click one to insert it.
- Use the `Show Properties ()` function. Specify the object in the parenthesis. The following example shows the messages that you can send to the platform object `oneObj`:

  ```
  Show Properties (oneObj)
  ```

  A list of messages appears in the log.

Conventions for Messages and Arguments

- For most messages, omitting the argument for a Boolean command enables the option. For example, all of these messages create a test for unequal variances:

  ```
  oneObj << Unequal Variances;
  oneObj << Unequal Variances( 1 );
  oneObj << Unequal Variances( "true" );
  oneObj << Unequal Variances( "yes" );
  oneObj << Unequal Variances( "present" );
  oneObj << Unequal Variances( "on" );
  ```

The first three approaches are recommended.

Notes:

- The quoted arguments "present", "absent", "on", "off", "switch", and "flip" are supported for backwards compatibility.
- When the name of the JMP option contains several options separated by a comma or slash (such as the **Means/Anova** option or the **T Square, T^2** option) you can use any *one* of the menu names. Or, you can use the full menu name, including the comma or slash, if the string is followed by n. For example, any of these messages create a Means/Anova test:

  ```
  oneObj << Means( 1 )
  oneObj << Anova( 1 )
  oneObj << "Means/Anova"n( 1 )
  ```

- When an option appears under a menu in JMP, the corresponding script message is the option itself without the parent menu. For example, in the Oneway red triangle menu, the Nonparametric menu has multiple options under it, including Wilcoxon Test. Simply use Wilcoxon Test as a function or message in a script:

  ```
  oneObj = Oneway( Y( height ), X( age ), Wilcoxon Test( 1 ));
  oneObj << Wilcoxon Test( 1 );
  ```

- When a menu in JMP contains values rather than options, in the script, specify the parent menu and the value as an argument. For example, in the Oneway red triangle menu, the Set Alpha Level menu has values, such as 0.10, 0.05, 0.01, and Other. To specify 0.01 in a script, add a line like the following:

  ```
  oneObj << SetAlphaLevel(0.01);
  ```

Send Multiple Messages

To send several messages, just add more << operators or more Send arguments:

```
dist << Quantiles( 1 ) << Moments( 1 ) << More Moments( 1 ) << Horizontal
  Layout( 1 );
Send( dist, Quantiles( 1 ), Moments( 1 ), More Moments( 1 ), Horizontal
  Layout(1));
```

Because << is an *eliding operator*, it combines arguments and works differently than if its arguments were grouped. You can stack multiple messages with extra << symbols to perform them all in order, from left to right.

Another way to stack messages is to send a list of messages:

```
dist << {Quantiles( 1 ), Moments( 1 ), More Moments( 1 ), Horizontal
  Layout( 1 )};
```

These approaches work well assuming that no value is returned as a result of one of your messages. However, if your goal is to send a message to the *result* of another message, add grouping parenthesis. The second line shows the wrong way to do it; the third line shows how to do it correctly:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
/* the following line is incorrect, and sends the Fit Mean message to the
   original data table object */
bv = dt << Run Script( "Bivariate" ) << Fit Mean( 1 );
/* the following line is correct, and sends the Fit Mean message to the
   resulting Bivariate object/report */
bv2 = ( dt << Run Script( "Bivariate" ) ) << Fit Mean( 1 );
```

Find Messages for Objects

Some messages apply to all platform objects, and others are specific to certain platforms. Once you have created a platform object, find out what messages you can send to the object, using any of the following approaches:

Use the Scripting Index in JMP

1. Select **Help** > **Scripting Index**.
2. Enter the object type that you are interested in.
3. Click the correct object in the resulting list. Messages for the object appear in the Items list.

Once you have created a platform object, use the Show Properties(obj) function

Replace `obj` with your named platform object. A list and brief description of all the messages the object can receive appears in the Log. To understand more about the Show Properties output, see "Interpret the Show Properties List".

Tip: The `Show Properties()` function also works with data tables and display boxes. See "How Can I See All of the Messages that Can be Sent to a Data Table Object?" and "The << Operator".

- Examine the platform in JMP launch windows and report windows. Most options that you find there have JSL equivalents with the same names and arguments.

Interpret the Show Properties List

Show Properties lists the messages that can be sent to a scriptable object, such as a data table, an analysis platform, or a display, producing a text report in the log. Each type of message is categorized and appears in brackets, and can be any of the following:

Message Type	Description	Example
`[Action]`	Correspond to options in the JMP interface.	`Redo Analysis [Action] (Rerun this same analysis in a new window. The analysis will be different if the data has changed.)`
`[Action Choice]` `[Enum]`	List specific choices for their arguments.	`Fit Polynomial [Action Choice] {2, quadratic, 3, cubic, 4, quartic, 5, 6}` `Draw [Enum] {Filled, Outlined, Filled and Outlined}`
`[Boolean]`	Turn options on or off. Arguments are usually 1 or 0. If a message is specified without an argument, sending the message changes it to the opposite state. `[Default On]` indicates that the option is on by default.	`Show Points [Boolean] [Default On]`

`[Subtable]`	List options within submenus. You use the option itself, not the parent item.	`Script [Subtable]`
		`Redo Analysis [Action] (Rerun this same analysis in a new window. The analysis will be different if the data has changed.)`
		`Save Script to Datatable [Action] (Return to the launcher for this analysis.)`

`[New Entity]`	Open a new window in the interface.	`Prediction Interval [New Entity] (Prediction Interval to contain a single future observation or the mean of m future observations.)`

For example, run the following JSL:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = Bivariate( Y( :height ), X( :age ) );
Show Properties( biv );
```

The object `biv` supports `Subtable`, `Boolean`, `Action`, and `Action Choice` type messages. `Enum` and `New Entity` messages are not supported.

Send Messages to an Object

Suppose that you want to send a message to an object in the output. For example, in Fit Model, you want to send the Response Limits message to the Profiler object, Use the subscript operator (`[ ]`) to refer to the Profiler object.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );

fm = dt << Fit Model(
   Y( :height ),
   Effects( :age, :sex, :weight, :age * :weight, :sex * :weight ),
   Personality( Standard Least Squares ),
   Emphasis( Effect Screening ),
   Run
);

obj = fm << Profiler(
   1,
   Confidence Intervals( 1 ),
   Desirability Functions( 1 ),
   Term Value( age( 12 ), sex( "F" ), weight( 105 ) )
);

Show( Profiler[1] << Get Desirability ); // show the limits before

Profiler[1] << (height << Response Limits(
   {Lower( 50, 0.95 ), Middle( 58, 0.7 ), Upper( 70, 0.066 ),
   Goal( "Minimize" ), Importance( 1 )}
));
Wait( 0.1 ); // for demonstration purposes
Show( Profiler[1] << Get Desirability ); // show the limits after
```

Specify Which Columns to Analyze

If you created your script interactively, you already specified columns in the launch window, so there is no need to specify them again. However, when creating a script from scratch, you normally want to specify the columns to analyze. For example, the following line launches the Distribution platform and specifies the height and weight columns as the Y variables:

```
Distribution( Y( height,weight ));
```

Create Column References

Tip: To see all the messages you can send to a column, go to **Help** > **Scripting Index** > **Data Table** > **Column Scripting**.

Creating a column reference enables you to access and send messages to the column. If you store each column reference in a JSL variable, you can use the variables instead of column names in the Y and X roles and elsewhere in the script.

For example, using Big Class.jmp, the following lines create Ycol as a column reference for weight, and Xcol as a column reference for height. The last line sends a message to Xcol to get the data type.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Ycol = Column( "weight" );
Xcol = Column( "height" );
Xcol << Get Data Type;
```

In the log, “Numeric” is output for the data type of the Xcol (height column).

Referring to a column reference with the : scoping operator (rather than plain text) is convenient in scripts saved to a data table. If the columns are renamed in the data table, the column references in the script are updated.

Specify Multiple Column Names at Once

If you have a lot of column names, instead of listing each one individually, you can use lists as column arguments. The following example gets the Y column names from a list and creates a Oneway analysis on each column in the list.

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
dt << Fit Group( Oneway( Y( {:SO2, :POP} ), X( :State ) ) );
```

If you do not know the column names when you are writing the script, use column indices, as done in the following example.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Distribution( Y( 1::N Col( dt ) ) );
```

You can also use indices when specifying effects as shown here:

```
dt << Fit Model( Y( :y ), Effects( 1 :: 4 ) );
```

Here's another example of specifying multiple column names. `5::31` selects columns 5 through 31.

```
Names Default To Here( 1 );
dt = Open( "$SAMPLE_DATA/Baltic.jmp" );
dt << Partial Least Squares( Y( :ls, :ha, :dt ), X( 5 :: 31 ), Go );
```

Enable Users to Specify Columns

Instead of specifying analysis columns yourself, you can enable a user to specify them, but you can still control what happens after they do so. In this case, use an empty specification. The following example runs the Distribution launch window:

```
Distribution();
```

In this scenario, the Distribution launch window appears and the user can choose the columns to analyze. After the user selects the columns and clicks **OK**, the Distribution report is shown as specified.

Specify a By Variable

In many JMP platforms, you can specify columns as By variables. To do this in a script, include a `By` argument in the platform command, listing each column as an argument.

The following example uses the Big Class.jmp data table, which contains the names, ages, sex, heights, and weights for 40 students. Create a bivariate report of weight by height, using sex as a By variable, and adding a variety of fits.

Create the Bivariate Report Using a By Variable

1. Select **File** > **New** > **Script**.
2. Add the following lines:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( weight ), X( height ), By( sex ));
```

The first line opens the Big Class.jmp sample data table. The second line creates a platform object called biv that runs a bivariate report of weight and height by sex.

3. Click **Run Script** .

   The report window shows two graphs, one where sex is F and one where sex is M.

Send Messages to the Entire Platform or to a Single By Level

1. If you do not have the log showing, turn it on. Right-click in the script and select **Show Embedded Log**.
2. Add this line to the script:

   ```
   Show( biv );
   ```

3. Highlight the `Show( biv )` line and click **Run Script** .

   In the log, rather than returning a single reference to a platform, `Bivariate[]`, the platform object returns a list of references: `biv = {Bivariate[], Bivariate[]}`. The two references correspond to the two levels (F and M) of the By variable, sex.

   You can direct messages to each By level individually or to all By levels.

4. Send a message to all By levels to add a linear regression fit. Add this line to the script:

   ```
   biv << Fit Line;
   ```

5. Highlight the line you just created and click **Run Script** .

   In the report window, a linear regression line is added to both graphs, with the corresponding Linear Fit reports.

6. Send a message to only the F level, adding a cubic polynomial fit. Add this line to the script:

   ```
   biv[1] << Fit Polynomial( 3 );
   ```

   Tip: The number for each By level corresponds to their order in the report window, which is usually alphanumeric. If the By column contains a Value Order column property, that ordering is followed.

7. Highlight the line you just created and click **Run Script** .

   In the report window, a polynomial fit line is added to the F graph, with the corresponding Polynomial Fit reports.

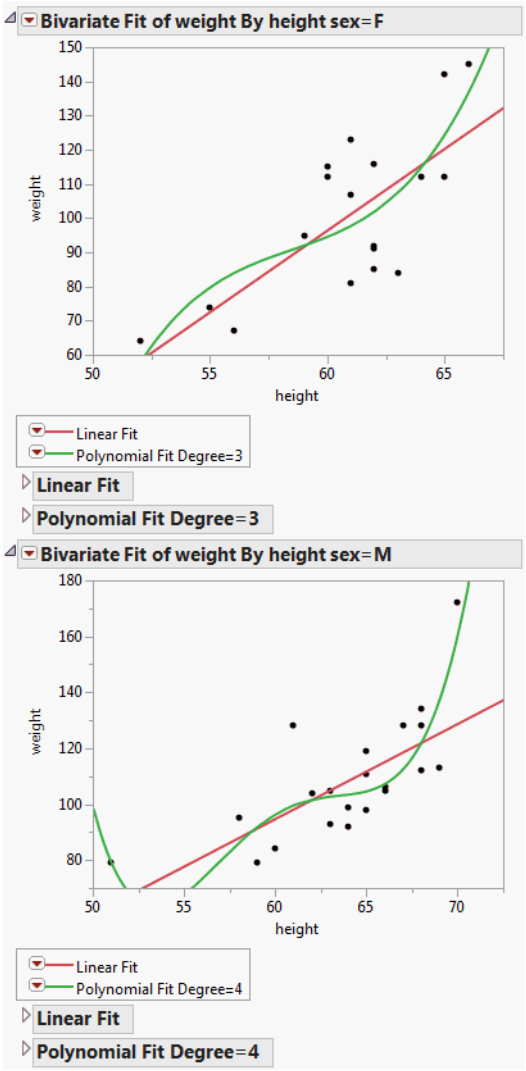
8. Send a message only to the M level, adding a quartic polynomial fit. Add this line to the script:

   ```
   biv[2] << Fit Polynomial( 4 );
   ```

9. Highlight the line you just created and click **Run Script** .

   In the report window, a quartic fit line is added to the M graph, with the corresponding Polynomial Fit reports.

Figure 10.2 By Group Reports



If you specify more than one column in `By` argument, graphs appear for each subgroup of each By variable. In this example, `By( sex, age )` would produce graphs for females age 12, females age 13, and so on, up to age 17. It would also produce graphs for males age 12 up to age 17.

Extract Results from a Report

The following example shows how to launch a platform with `By` groups and extract results (in this case, variances) from each group:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
```

```
// onew is the JSL variable holding a list of platforms
onew = dt << Oneway( x( :age ), y( :height ), by( :sex ), anova );

// r is a list of reports
r = onew << Report;

// nBy is the number of reports generated
nBy = N Items( r );

// vc is an array of as many rows as reports, one column, all set to zero
vc = J( nBy, 1, 0 );

// for each report, do the following:
For( i = 1, i <= nBy, i++,
  vc[i] = r[i] // vc[i] is the variance of the ith report
[Outline Box( "Analysis of Variance" ), // look for this outline

// then look for this column and get the second value
  Column Box( "Sum of Squares" )][2]
);
Show( vc ); // debugging, look in log to see this value

// byValues becomes a list of the values in the sex column
Summarize( byValues = By( :sex ) );

// make a new table with two rows (M.F) and two columns
New Table( "Variances" )
<< New Column( "Sex", // create a new column called Sex
  character,
  width( 8 ),
  values( byValues )

// create a new column called Variance
) << New Column( "Variance", Numeric, "Continuous", Values( vc ) );
```

Filter by Value using a Where Statement

You can use a `Where` statement to filter by a specific value. If you are adding a Where clause to a script that contains column references, you must first resolve the references. The following example limits the analysis to females and fits a line on the Bivariate report.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Ycol = Column( "weight" );
```

```
Xcol = Column( "height" );
biv2 = dt << Bivariate(
Y( Ycol ),
X( Xcol ),
Fit Line( ),
Where( :sex == "F" ) );
```

Enable User Input

When you send an `Action()` function to a platform, it evaluates an expression. For example, use `Action` messages when you want the launch window to appear. A user chooses columns in the launch window and the script continues running after the user runs the analysis.

The following example launches four platforms: Distribution, Bivariate, Oneway, and Contingency. The launch window for each platform appears, and the user chooses the analysis columns and clicks OK. The report window for each analysis appears, and then the next platform launch window comes up, until all platforms have been launched.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Distribution( Action( doit ) );
doit = Expr(
  New Window( "Bivariate", Bivariate( Action( doit2 ) ) )
);
doit2 = Expr(
  New Window( "Oneway", Oneway( Action( doit3 ) ) )
);
doit3 = Expr(
  New Window( "Contingency", Contingency( Action( doit4 ) ) )
);
doit4 = Expr();
```

Execute Embedded Red Triangle Options

In addition to the red triangle menu at the top of a report, some report windows contain embedded red triangle menus. These appear when you select an option that has sub-options, such as profilers. To execute options in embedded red triangle menus, use the `Get Scriptable Object` message. The following example creates a Life Distribution report then turns off the Confidence Intervals option in the Hazard Profiler red triangle menu:

```
dt = Open( "$SAMPLE_DATA/Reliability/Fan.jmp" );

// create a Life Distribution report with a lognormal distribution
obj = dt << Life Distribution(
```

```
        Y( :Time ),
        Censor( :Censor ),
        << Fit Lognormal
);

// get a reference to the Hazard Profiler red triangle menu
haz_prof = ( report( obj )[ "Hazard Profiler" ] << Get Scriptable Object);

// turn off confidence intervals in the Hazard Profiler
haz_prof << Confidence Intervals( 0 );
```

Make Platforms Invisible

In certain circumstances, you might want to make the platform analysis invisible. This means that no platform windows appear, and the platform runs in the background.

If you are doing a simulation or bootstrap analysis in your script, the script might call a platform hundreds or thousands of times. Perhaps you just want to pull one or two results out of each platform report. The following example performs a simulation and then creates a table of results and shows them in a Distribution report:

```
dt = As Table( (0 :: 10)`, << Column Names( {"X"} ) ); // create X variable

// set initial random seed to get reproducible results
Random Reset( 12345 );
dt << New Column( "Y",
   "Continuous",
   "Numeric",
   Set Formula( :X * 2 + 1 + Random Normal() )
);

// Y values have a random component
res = []; // define results matrix (empty to start)
For( i = 1, i <= 100, i++, // loop through simulation runs

   // for each iteration, fit a regression line and save parameter estimates
   bv = dt << Bivariate( Y( :Y ), X( :X ), Fit Line( 1 ), Invisible );
   res |/= (Report( bv )["Parameter Estimates"][

   // concatenate the results to the bottom of the results matrix
   Number Col Box( "Estimate" )] << Get as Matrix)`;

   // close the window (even though it's invisible)
   bv << Close Window;
      dt:Y << Eval Formula; // generate new Y values
```

```
);
// create table of results
dtres = As Table( res, << Column Names( {"Intercept", "Slope"} ) );

// show results in Distribution
dtres << Distribution( Y( :Intercept, :Slope ) );
```

If you want to run an analysis as an intermediate step in your script, and do not need the user to see the analysis, hide it with the `invisible` option. In this case, you might want to show only specific results from the report, which you can output to a journal or to the log.

- The following example runs the Distribution platform as invisible, and outputs only the CDF plot to a new journal:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dist = dt << Distribution( Y( :height ), CDF Plot( 1 ), invisible );
Report( dist )["CDF Plot"] << Journal;
dist << Close Window;
```

- The following example extracts the *F*-Ratio from an invisible Bivariate report, and outputs it to the log:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( x( :height ), y( :weight ), invisible );
biv << Fit Line;
r = biv << Report;
fratio = r[ColumnBox( "F Ratio" )][1];
r << Close Window;
Show( fratio );
  fratio = 1.15609545178219;
```

Tip: Invisible windows use resources that must be manually freed. Be sure to close the invisible window when the script is done with it.

You can also use the `invisible` option on the options in the Tables menu, such as Subset, Sort, Stack, and so on. The following example makes the Subset operation invisible:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Where( :age == 14 );
dt = dt << Subset( invisible );
subDt << Bivariate( x( :height ), y( :weight ), Fit Line );
```

Note that the preceding line could also be handled using a `WHERE` clause:

```
subDt << Bivariate( x( :height ),y( :weight ), Where( :age == 14 ), Fit Line
  );
```

Specify Report Titles

You can specify the title (shown in the title bar of a platform's report) by adding the `title` command to the launch request. The following example replaces the standard bivariate report's title with a user-specified one (My Title).

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Bivariate( x( :height ), y( :weight ), Title( "My Title" ) );
```

Common Messages for Platform Windows

For a list of common messages that you can send to most platforms, see the *JSL Syntax Reference*.

Scripting Considerations by Platform

This section explains specific considerations when scripting the following platforms:

- "Categorical"
- "Control Charts"
- "Distribution"
- "DOE"
- "Fit Model"
- "Formula Depot"
- "Graph Builder"
- "Neural and Neural Net"
- "Partial Least Squares and PLS"
- "Process Capability"
- "Scatterplot 3D"
- "Text Explorer"
- "Explore Outliers"

Categorical

The Free Text option in Categorical displays a Text Explorer report. Scripts that call the Free Text option in Categorical should be updated to use the Text Explorer platform.

Scripting for Supercategories

When ratings are involved in a data set (for example, a five point scale), you might want to know the percent of the responses in the top two or another subset of ratings. You can define a group of ratings in the data using the supercategories scripting option (instead of the column property).

The following example creates a new column called Floss Delimited 2, and adds a supercategory called Sleep Related, which contains two values: Before Sleep and Wake.

```
dt = Open( "$SAMPLE_DATA/Consumer Preferences.jmp" );
dt << New Column( "Floss Delimited 2",
   Character,
   Set Property(
      "Supercategories",
      {Group( "Sleep Related", {"Before Sleep,", "Wake,"} )}
   ), ...);
```

In the data table, under the Columns panel at left, scroll down to the new Floss Delimited 2 column. Click the column property indicator ✻ and select Supercategories. You can see the new Sleep Related supercategory with its two values.

Supercategories can also be specified in a Categorical launch command, where the properties are listed inside parentheses after the column name:

```
dt = Open( "$SAMPLE_DATA/Consumer Preferences.jmp" );
dt << Categorical(
   Supercategories(
      :Floss Delimited(
         {Group( "Sleep Related", {"Before Sleep,", "Wake,"} )}
      )
   ),
   Multiple Response( :Floss Delimited )
);
```

In the Categorical report, notice that the Sleep Related category has 203 cases. If you had run the Categorical platform without creating this supercategory, Before Sleep and Wake would appear as separate categories with 143 and 60 cases, respectively.

For more information about supercategories, see *Consumer Research*.

Control Charts

Using JSL, you can create control charts, customize tests, run alarm scripts, and more.

Reference Lines in Control Chart Builder

If you add a custom reference line to a Control Chart Builder chart through JSL, the reference line label must be something other than “LSL”, “USL”, or “AVG”. These names are keywords used by the Control Chart Builder platform.

The following example creates custom reference lines with unique labels.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
obj = dt << Control Chart Builder(
  Variables( Y( :height ) ),
  SendToReport(
    Dispatch(
      {},
      "height",
      ScaleBox,
      {Add Ref Line( 70, "Solid", "Black", "Upper Limit", 1 ),
      Add Ref Line( 55, "Solid", "Black", "Lower Limit", 1 )}
    )
  )
);
```

Customize Tests in Control Chart Builder

You can design custom tests and select or deselect multiple tests at once using the `Customize Tests` function. You specify the description, desired number, and label. This option is available only for Variables and Attribute chart types.

The following example creates a custom test called Test 1 and turns it on.

```
dt = Open( "$SAMPLE_DATA/Quality Control/Diameter.jmp" );
dt << Control Chart Builder(
  Variables( Subgroup( :DAY ), Y( :DIAMETER ) ),
  Customize Tests( Test 1( 2, "1" ) ), // test number, n, label
  Chart( Position( 1 ), Warnings( Test 1( 1 ) ) ), // turn Test 1 on
  Chart( Position( 2 ) )
);
```

Run Alarm Scripts

An alarm script alerts you when the data fail one or more tests. As an Alarm Script is invoked, the following variables are available, both in the issued script and in subsequent JSL scripts:

`qc_col` is the name of the column

`qc_test` is the test that failed

`qc_sample` is the sample number

`qc_phase` is the label of the phase during which the failure occurred

`qc_firstRow` is the first row in the sample

`qc_lastRow` is the last row in the sample

Example 1: Automatically Writing to a Log

One way to generate automatic alarms is to make a script and store it with the data table as a data table script or column property named QC Alarm Script.

The following example automatically writes a message to the log whenever a test fails:

```
dt = Open( "$SAMPLE_DATA/Quality Control/Coating.jmp" );
obj = dt << Control Chart(
   Sample Size( :Sample ),
   KSigma( 3 ),
   Chart Col( :Weight, XBar, R ),
   Alarm Script(
      Write(
         "Out of Control for test ",
         qc_test,
         " in column ",
         qc_col,
         " in sample ",
         qc_sample
      )
   )
);
obj << Test 1( 1 );
```

Example 2: Running a Chart with Spoken Tests

You can create a control chart with tests that are spoken aloud using the Speak function. Try the following example:

```
dt = Open( "$SAMPLE_DATA/Quality Control/Coating.jmp" );
dt << Control Chart(
   Alarm Script(
      Speak(
         Match( QC_Test,
            1, "One point beyond Zone A",
            QC_Test, 2,
            "Nine points in a row in zone C or beyond", QC_Test,
            5,
               "Two out of three points in a row in Zone A
               or beyond"
         )
      )
   ),
```

```
   Sample Size( :Sample ),
   Ksigma( 3 ),
   Chart Col( :Weight, Xbar( Test 1( 1 ), Test 2( 1 ), Test 5( 1 ) ), R )
);
```

You can have either of these scripts use any of the JSL alert commands, such as `Speak`, `Write`, or `Mail`.

Set Phase Limits

A *phase* is a group of consecutive observations in the data table. For example, phases might correspond to time periods during which a new process is brought into production and then put through successive changes. Phases generate, for each level of the specified Phase variable, a new sigma, set of limits, zones, and resulting tests.

The following example illustrates setting the limits for the different phases of Diameter.jmp:

```
dt = Open( "$SAMPLE_DATA/Quality Control/Diameter.jmp" );
dt << Control Chart(
   Phase( :Phase ),
   Sample Size( :DAY ),
   KSigma( 3 ),
   Chart Col(
      :DIAMETER,
      XBar(
         Phase Level(
            "1",
            Sigma( .29 ),
            Avg( 4.3 ),
            LCL( 3.99 ),
            UCL( 4.72 )
         ),
         Phase Level(
            "2",
            Sigma( .21 ),
            Avg( 4.29 ),
            LCL( 4 ),
            UCL( 4.5 )
         )
      ),
      R( Phase Level( "1" ), Phase Level( "2" ) )
   )
);
```

Distribution

You can set specification limits using the Spec Limits column property or using the `Spec Limits` argument in JSL:

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
dt << Distribution(
   Column( :OZONE ),
   Fit Distribution(
      Weibull(
         Spec Limits( LSL( 0.075 ), Target( 0.15 ), USL( 0.25 ) )
      )
   )
);
```

To set specification limits using K Sigma normal probabilities, use the `Set Spec Limits for KSigma` message. The first argument to the message is the value of K Sigma. By default, the limits are two-sided. To make them one-sided, use the optional `sided` argument. Specify `sided=1` for LSL only and `sided=2` for USL only.

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
obj = Distribution( Column( :CO ) );
obj << Fit Distribution(
   LogNormal( Set Spec Limits for KSigma( 3 ) )
);
```

DOE

Tip: For a complete list of DOE scripting options, select **Help > Scripting Index**, select **Objects** from the menu, and search for DOE.

To reproduce your DOE work, most DOE platforms provide these approaches:

- In the red triangle menu for the DOE window, the Save Script to Script Window option provides a script that reproduces the work that you have entered. This option is not available for the Nonlinear Design and Taguchi Array platforms.
- In the output design tables created by most DOE platforms, a script called DOE Dialog reproduces the work in the DOE window that created the design table. This script also contains the random seed (and number of starts or column starts, where appropriate) that enables you to reproduce your design exactly.

For many DOE platforms, the design table that is created also contains one or more scripts that you can run to conduct an appropriate analysis. These scripts are prefilled with the appropriate settings for your design.

The Random Seed

In constructing designs using the DOE platform, JMP uses a random seed to control certain actions that have a random component. A random seed can control the following:

- initializing search algorithms for design generation
- randomizing Run Order for design construction
- selecting a starting design for designs based on random starts

To reproduce a design using a script, you need to specify the random seed that generated them. For designs using random starts, set a random seed in the script before making the design.

The following example creates a Custom Design, sets the random seed, and makes the design:

```
DOE(
  Custom Design,
  Add Factor( Continuous, -1, 1, "X1", 0 ),
  Add Factor( Continuous, -1, 1, "X2", 0 ),
  Set Random Seed( 34067086 ),
  Make Design
);
```

Reproduce a Custom Design or a Main Effects Screening Design

Custom designs and main effects screening designs (constructed in the Screening Design platform) are generated by allocating a maximum number of seconds (Design Search Time) spent searching for a design. The default setting for Design Search Time is based on the complexity of the design. Even when two custom or main effects screening designs are constructed using the same random seed, the designs might not be identical because of different machine processing capabilities.

When you save the script for either of these designs using the Save Script to Script Window option, or when the DOE Dialog script is saved to the data table, the script specifies the Random Seed and the Number of Starts. The Number of Starts is the number of random starts used during the Design Search Time. The combination of the Set Random Seed and the Number of Starts options enables you to reproduce the design.

Note: If you write your own script, be aware that the order in which you specify options in your script can affect your results. For example, you should specify the design options before you make the design. In particular, if your design is constructed using a random seed, set it before you make the design.

Global Variables

Here are some global variables used to initiate or tune search algorithms.

Starting Design Enables you to specify a starting design. The following example replaces the random starting design with a specified matrix:

```
DOE Starting Design = matrix;
```

If a starting design is supplied, the Custom Design platform has only one start using this design.

K Exchange Value The coordinate exchange algorithm considers every row of factor settings for possible replacement in every iteration. You can limit the number of rows considered for exchange in an iteration so that only a small number of most likely rows is considered. The following example tells the algorithm to consider only the three most likely rows for exchange in each iteration:

```
DOE K Exchange Value = 3;
```

Bayes Diagonal Defines a vector that is added to the diagonal elements of the $X'X$ matrix. This new matrix is used in finding the D-optimal design. The following example adds the elements of `vector` to the diagonal elements of the $X'X$ matrix:

```
Bayes Diagonal = vector:
```

Fit Model

When scripting the Fit Model platform, here are a few general tips for controlling the behavior of the launch window:

- To keep the launch window open and fit the model at the same time, do one of the following (they are equivalent):
  - Include the `Run Model` message in your script.
  - Include both the `Run` message and the `Keep Dialog Open(1)` message.
- To run the model without showing the launch window, include the `Run` message in your script.

Model Scripts in the Data Table

When a data table contains a *Model* script and you launch Fit Model interactively, the Fit Model launch window is pre-populated according to the roles assigned by the table script.

Fit Group

If you want to fit different models and see a shared Profiler so that you can compare the models, use the `Fit Group` function. You can fit models such as least squares, nonlinear, neural, Gaussian processing, and mixed, in the same window with a shared Profiler. The following example creates both a Standard Least Squares model and a Gaussian Process model.

```
dt = Open( "$SAMPLE_DATA/Tiretread.jmp" );
obj = dt << Fit Group(
  Fit Model(
     Y( :ABRASION ),
     Effects(
        :SILICA & RS,
        :SILANE & RS,
        :SULFUR & RS,
        :SILICA * :SILICA,
        :SILICA * :SILANE,
        :SILANE * :SILANE,
        :SILICA * :SULFUR,
        :SILANE * :SULFUR,
        :SULFUR * :SULFUR
     ),
     Personality( "Standard Least Squares" ),
     Emphasis( Minimal Report ),
     Run()
  ),
  Gaussian Process(
     Y( :HARDNESS ),
     X( :SILICA, :SILANE, :SULFUR ),
     Set Correlation Function( "Cubic" )
  )
);
```

In the report, click the Fit Group red triangle and select **Profiler**. Scroll down to the Prediction Profiler. You can see the ABRASION row for the SLS model, and the HARDNESS row for the Gaussian Process model appearing in the same profiler.

Effects

Effects in the model can be more than just a list of columns, and can have a specialized syntax:

```
Effects( list of effects, list of effect macros, or both lists );
```

An effect can be a column name, a crossing of several column names with asterisk (*) notation, or nested columns specified with subscript bracket ([ ]) notation. Additional effect options can appear after an ampersand (&) character. Here are some examples:

```
A,                  // a column name alone is a main effect
A*B,                // a crossed effect, interaction, or polynomial
A[B],               // nested
A*B[C D],           // crossed and nested
effect&Random,      // a random effect
effect&LogVariance, // a variance model term
effect&RS,          // a response surface term
effect&Mixture,     // for an effect participating in a mixture
effect&Excluded,    // for an effect that with no model arguments
effect&Knotted,     // for a knotted spline effect
```

Effect macros are:

```
Factorial( columns ),   // for a full factorial design
Factorial2( columns ),  // for up to 2nd-degree interactions only
Polynomial( columns ),  // for a 2nd-degree polynomial only
```

Responses and Effects for MANOVA

To address an individual response function analysis, use a subscripted `Response`:

```
manovaObj << ( Response[ 1 ] << {response options} );
manovaObj << ( Response[ "Contrast" ] << {response options} );
```

Each response function supports the `Custom Test` message:

```
Custom Test( matrix, <Power Analysis( ... )>, <Label( "..." )> )
```

where each row of the *matrix* specifies coefficients for all the arguments in the model.

To address an individual Effect test, use subscripted `Effect` with a name or number:

```
manovaObj << ( Response[1] << ( Effect["Whole Model"] << {effect options} ) );
manovaObj << ( Response[1] << ( Effect[i] << {effect options} ) );
```

The effects are numbered:

- 0 for the intercept
- 1, 2, and so on, for regular effects
- *n*+1 for the "Whole Model" test, where *n* is the number of effects *not* including the intercept

Each effect in each response function supports the following messages, where each row of the *matrix* has coefficients for all the levels in the effect:

```
Test Details( 1 ),
Centroid Plot( 1 ),
```

```
Save Canonical Scores,
Contrast( matrix, <Power Analysis(...)> );
```

For example, the following JSL script adds test details for an effect to the report window:

```
dt = Open( "$SAMPLE_DATA/Dogs.jmp" );
manObj = dt << Fit Model(
   Y( logHist0, logHist1, logHist3, logHist5 ),
   Effects( dep1, drug, drug * dep1 ),
   Personality( "MANOVA" ),
   Run Model
);
manObj << Response Function( "Contrast" );
manObj << (Response["Contrast"] << (Effect["Whole Model"] <<

/* send the Test Details message to the Whole Model outline
under Contrast in the report window */
Test Details( 1 )));

/* send the Test Details message to response 1 (Contrast)
and effect 3 (drug*dep1) */
manObj << (Response[1] << (Effect[3] << Test Details( 1 )));
```

Send Function

Tip: Recall that the << operator is equivalent to the `Send()` function.

When you have multiple responses, you can send messages to a specific response column's fit. Use the following JSL (where *responseName* is the specific response):

```
fitObj << ( responseName << {options, ...});
```

The second `Send()` function finds the named response and sends the list of messages to it.

Note: If you send the messages directly to the *fitObj* with a single `Send()` function, the messages are sent to all responses.

In the following script, the last line sends a message to the AICc report for the ABRASION response:

```
dt = Open( "$SAMPLE_DATA/Tiretread.jmp" );
fitObj = dt << Fit Model(
   Y( :ABRASION, :MODULUS, :ELONG, :HARDNESS ),
   Effects(
      :SILICA & RS,
      :SILANE & RS,
      :SULFUR & RS,
```

```
        :SILICA * :SILICA,
        :SILANE * :SILICA,
        :SILANE * :SILANE,
        :SULFUR * :SILICA,
        :SULFUR * :SILANE,
        :SULFUR * :SULFUR
    ),
    Personality( "Standard Least Squares" ),
    Run
);
/* shows up under the Studentized Residuals plot*/
fitObj << (:ABRASION << {AICc( 1 )});
```

To send messages to an individual effect, nest even further:

```
fitObj << ( responseName << ((effectName) << effectOption ));
```

Standard Least Squares

For Standard Least Squares models that contain only fixed effects and more than one Y response, you can choose to fit the models for the Y responses together or separately. If some of the rows have missing values, the following apply:

- In a script, the Y responses are fit together by default. The model fits each Y using only those rows that are nonmissing for all of the Y variables. For example, a row that includes one or more missing Y values is *excluded* from the model. You can explicitly include `Run( "Fit Together" )` in a script to get the same result.
- If you include the `Run( "Fit Separately" )` option, the model fits each Y using all rows that are nonmissing for that particular Y. For example, a row that includes one or more missing Y values is *included* in the model.

The following script fits the Y responses together:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Fit Model(
    Y( :height, :weight ),
    Effects( :sex ),
    Personality( "Standard Least Squares" ),
    Emphasis( "Minimal Report" ),
    Run // or Run( "Fit Together" )
);
```

The following script fits the Y responses separately:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Fit Model(
    Y( :height, :weight ),
    Effects( :sex ),
```

```
    Personality( "Standard Least Squares" ),
    Emphasis( "Minimal Report" ),
    Run( "Fit Separately" )
);
```

When running the model in JMP, the user is prompted to select fitting the responses separately or together. The Fit Model launch window contains a Fit Separately option, which is deselected by default. For more information about how missing Y values are handled, see *Fitting Linear Models*.

Note: In models that contain a random effect, Y values are fit separately by default.

Formula Depot

Sending commands via JSL to a Formula Depot platform report requires that you first create an empty Formula Depot. Once you have a reference to the Formula Depot report, you can send commands to it. The following script opens a formula depot report and adds a model to the depot.

```
dt = Open( "$SAMPLE_DATA/Liver Cancer.jmp" );
fd = Formula Depot(); // open a new Formula Depot
obj1 = dt << Run Script( "Lasso Poisson, Validation Column" );
model = obj1 << (Fit[1] << Publish Prediction Formula );

/* Publish a model to the Formula Depot. The
following script opens a formula depot report, adds a model to
the depot, generates scoring code for the formula, saves the code to a file,
and closes all windows.*/
case_name = "fitLS_prediction";
dest_dir = "c:\"; //CHANGE ME to a better location
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Print( "Running " || case_name );
fm_ls = Fit Model(
   Y( :weight ),
   Effects( :age, :sex, :height ),
   Personality( "Standard Least Squares" ),
   Emphasis( "Minimal Report" ),
   Run
);
fd = fm_ls << Publish Prediction Formula;
Try(
   Print( "Generate Python Code" );
   code = fd << Generate Python Code( No Editor );
   Save Text File( dest_dir || case_name || ".py", code );
   Print( "Done." );
```

```
,
  Print( "Code generation failed for " || case_name || ", reason: " );
  Show( exception_msg );
);
fdw = Window( "Report: Formula Depot" );
fdw << Close Window( No Save );
Close( dt, No Save );
```

Graph Builder

The following is an example of the basic structure of a Graph Builder script:

```
Graph Builder(
Size( x, y ),
// Use any commands from the red triangle menu - Size is an example
Variables( role( column ), ...), //Required
Elements( element name( element options, ...), //Required
Data Filter( ... ),
// Use any display customizations here - Data Filter is an example
);
```

The following is an example of a simple Graph Builder script that creates a bar chart:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Graph Builder(Variables(
  X( :age ),
  Y( :weight ) ),
Elements( Bar( X, Y, Legend( 8 ) ) )
// Legend is not required, but provides an internal legend ID that you can use
  for later customization
);
```

The following is a more detailed Graph Builder script that creates a bar chart:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Graph Builder(
Size( 373, 332 ),
Show Control Panel( 0 ),
Variables(
  X( :age ),
  Y( :weight ),
  Y( Transform Column( "weight25",
       Formula( Col Quantile( :weight, 0.25, :age ) )
       ),
     Position( 1 )
/* By default each Y variable gets its own axis, but the Position(1)
option keeps each additional Y in the first slot and merges the axes. */
  ),
```

```
    Y( Transform Column( "weight75", Formula( Col Quantile( :weight, 0.75, :age
    ) ) ),
       Position( 1 )
    )
  ),
  Elements(
    Bar( X, Y( 2 ), Y( 3 ), Legend( 10 ), Bar Style( "Range" ) ),
  // This bar element uses only the second and third Y variables and
  // draws a Range bar between the values.*/
    Bar( X, Y( 1 ), Legend( 8 ),
       Bar Style( "Float" ),Summary Statistic( "Median" )
  // This bar element uses the first Y variable and draws its median with the
  // float style, which is a short horizontal line
    ),
    Points( X, Y( 1 ), Legend( 9 ) )
  // The Points element shows the raw data as markers.
  ),
  SendToReport(
  // Contains display customizations
    Dispatch( {}, "400", ScaleBox,
  // 400 is the fixed name
       {Legend Model( 10,
       Level Name( 0, "1st to 3rd
       quartiles",
       Item ID( "Mean(weight25)..Mean(weight75)", 1 )
  // Renames the legend, moves the legend name onto several lines, and changes
  // the color for legend 10 (the Range bar)
    ),
       Properties( 0, {Fill Color( 32 )},
       Item ID( "Mean(weight25)..Mean(weight75)", 1 )
  // Changes the color for legend 10 (the range bar)
    )
       ), Legend Model( 9,
       Properties( 0, {Marker Size( 4 )}, Item ID( "weight",1 ) )
  // Changes the marker size for legend 9 (the Points element)
    )}
  ),
  Dispatch( {}, "graph title", TextEditBox,
    {Set Text( "Weights with quartile ranges" )}
  // Changes the graph title
  ),
  Dispatch( {}, "Y title", TextEditBox, {Set Text( "weight" )} ),
  // Changes the Y axis title
  Dispatch( {}, "400", LegendBox,
    {Legend Position( {10, [0], 8, [1], 9, [-1]} )}
  // Hides the legend item for legend 9 by giving it a position of -1
```

```
    )
)
);
```

Neural and Neural Net

The Neural platform replaces the Neural Net platform. Scripts that use Neural Net are deprecated and might not work in the future. It is recommended to use the Neural platform for new scripts and update any old scripts using Neural Net to Neural.

Partial Least Squares and PLS

The Partial Least Squares (PLS) platform has replaced the PLS platform. Scripts that use PLS might not work in the future. It is recommended to use the Partial Least Squares platform for new scripts and update any old scripts using PLS to Partial Least Squares.

Process Capability

The Process Capability platform has replaced the Capability platform. Scripts that use Capability might not work in the future. It is recommended to use the Process Capability platform for new scripts and update any old scripts using Capability to Process Capability.

Specification Limits in JSL Scripts

Specification limits can be read from JSL scripts, column properties, or from a specification limits data table. As an example of reading in specification limits from JSL, consider the following example:

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
dt << Process Capability(
   Process Variables( :OZONE, :CO, :SO2, :NO ),
   Spec Limits(
      OZONE( LSL( 0.075 ), Target( 0.15 ), USL( 0.25 ) ),
      CO( LSL( 5 ), Target( 7 ), USL( 12 ) ),
      SO2( LSL( 0.01 ), Target( 0.04 ), USL( 0.09 ) ),
      NO( LSL( 0.01 ), Target( 0.025 ), USL( 0.04 ) )
   )
);
```

Using a Limits Data Table and JSL

There is no extra syntax needed to differentiate between the two table types (wide and tall) when they are read using JSL. This example is based on the CitySpecLimits.jmp sample data table. It places the specification limits data table inside an `Import Spec Limits()` expression.

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
dt << Process Capability(
   Process Variables( :OZONE, :CO, :SO2, :NO ),
   Spec Limits(
      Import Spec Limits(
         "$SAMPLE_DATA/CitySpecLimits.jmp"
)));
```

Scatterplot 3D

If your machine supports acceleration, and you want to display the 3-D scatterplot faster, you can turn on the Use Hardware Acceleration option, or in JSL, specify the `Scene3DHardwareAcceleration` global variable. A value of 1 turns it on, and a value of zero turns it off.

```
dt = Open( "$SAMPLE_DATA/Solubility.jmp" );
Scene3DHardwareAcceleration = 1;
dt << Scatterplot 3D(
   Y( :"1-Octanol"n, :Ether, :Chloroform, :Benzene, :Carbon Tetrachloride,
   :Hexane )
);
```

Text Explorer

By default, text is converted to all lowercase in the Text Explorer platform. You can use the `"From Source"` keyword in the Save Regex Column message to save the results of a custom regex that preserve the original case of the text. The first argument is the name of the column that is created in the data table. The second argument is the string `"From Source"` or `"From Result"`. The default behavior is equivalent to specifying the `"From Result"` keyword.

```
dt = Open( "$SAMPLE_DATA/Aircraft Incidents.jmp" );
obj = dt << Text Explorer(
   Text Columns( :Airport Name),
   Tokenizing( "Regex" ),
   Save Regex Column( "Output Column", "From Source" )
);
obj << Close Window();
```

Explore Outliers

The Explore Outliers platform provides options to identify, explore, and manage outliers in your univariate or multivariate data.

In the Probe.jmp sample data table, several columns contain outlier values of 9999. Many industries use nines as a missing value code. The following example selects VDP_PCOLL, VDP_PINNBASE, and VDP_PINPBASE, the columns with the highest number of 9s, and adds the Missing Value Codes column property to the columns. The script then rescans the columns and removes these columns from the report.

```
dt = Open( "$SAMPLE_DATA/Probe.jmp" );
obj = Explore Outliers(
   Y(
      :VDP_M1,
      :VDP_M2,
      :VDP_NBASE,
      :VDP_NEMIT,
      :VDP_NENBNI,
      :VDP_NSINK,
      :VDP_PBASE,
      :VDP_PBL,
      :VDP_PCOLL,
      :VDP_PEMIT,
      :VDP_PINNBASE,
      :VDP_PINPBASE,
      :VDP_PSINK,
      :VDP_SICR
   ),
   Quantile Range Outliers( 1 ),
   Show only columns with outliers( 1 ),
);
obj << Add Highest Nines to Missing Value Codes( :VDP_PCOLL, :VDP_PINNBASE,
   :VDP_PINPBASE );
obj << Rescan;
```

Scripting-Only Messages and Arguments by Platform

Some messages and arguments that can be sent to a platform object are scripting-only. This means that these options do not have corresponding options in JMP launch windows or reports. However, note that scripting-only options can appear as Preferences.

This section describes scripting-only messages and arguments for specific platforms.

Tip: For more information about any of these messages or arguments, and for specific examples, go to the Scripting Index (**Help** > **Scripting Index**).

- "Bivariate Script-Only Messages"
- "Boosted Tree Script-Only Messages"
- "Bootstrap Forest Script-Only Messages"
- "Bubble Plot Script-Only Messages"
- "Choice Script-Only Messages"
- "Clustering Script-Only Messages"
- "Contingency Script-Only Messages"
- "Contour Plot Script-Only Messages"
- "Contour Profiler Script-Only Messages"
- "Control Chart Script-Only Messages"
- "Control Chart Builder Script-Only Messages"
- "Cumulative Damage Script-Only Messages"
- "Custom Profiler Script-Only Messages"
- "Degradation Script-Only Messages"
- "Destructive Degradation Script-Only Messages"
- "Discriminant Script-Only Messages"
- "Distribution Script-Only Messages"
- "EMP Measurement Systems Analysis Script-Only Messages"
- "Explore Missing Values Script-Only Messages"
- "Explore Outliers Script-Only Messages"
- "Factor Analysis Script-Only Messages"
- "Fit Life by X Script-Only Messages"
- "Fit Model Script-Only Messages"
- "Fit Parametric Survival Script-Only Messages"
- "Functional Data Explorer Script-Only Messages"
- "Gaussian Process Script-Only Messages"
- "K Means Cluster Script-Only Messages"
- "K Nearest Neighbors Script-Only Messages"
- "Latent Class Analysis Script-Only Messages"
- "Life Distribution Script-Only Messages"
- "Make Validation Column Script-Only Messages"
- "MaxDiff Script-Only Messages"

- "Mixture Profile Script-Only Messages"
- "Model Screening Script-Only Messages"
- "Multivariate Script-Only Messages"
- "Naive Bayes Script-Only Messages"
- "Neural Script-Only Messages"
- "Nonlinear Script-Only Messages"
- "Normal Mixtures Script-Only Messages"
- "Oneway Script-Only Messages"
- "Parallel Plot Script-Only Messages"
- "Pareto Plot Script-Only Messages"
- "Partial Least Squares Script-Only Messages"
- "Partition Script-Only Messages"
- "Principal Components Script-Only Messages"
- "Process Capability Script-Only Messages"
- "Process Screening Script-Only Messages"
- "Profiler Script-Only Messages"
- "Reliability Forecast Script-Only Messages"
- "Reliability Growth Script-Only Messages"
- "Response Screening Script-Only Messages"
- "Scatterplot 3D Script-Only Messages"
- "Structural Equation Models Script-Only Messages"
- "Support Vector Machines Script-Only Messages"
- "Surface Plot Script-Only Messages"
- "Tabulate Script-Only Messages"
- "Ternary Plot Script-Only Messages"
- "Text Explorer Script-Only Messages"
- "Time Series Script-Only Messages"
- "Uplift Script-Only Messages"
- "Variability Chart Script-Only Messages"

Bivariate Script-Only Messages

The Fit Where message for a Bivariate object fits a curve or line to a single level of a categorical variable.

```
biv_object << Fit Where( WHERE_clause, fit_command );
```

The Curve message for a Bivariate object enables you to send a message to a curve or line that has already been fit in the Bivariate report. The curves are numbered sequentially in the order they appear in the report.

```
biv_object << ( Curve[n] << curve_message );
```

The Set Kernel message for a Nonpar Density curve in a Bivariate object enables you to set the kernel standard deviations for the X and Y values:

```
biv_object << ( Curve[n] << Set Kernel( x, y ) );
```

Boosted Tree Script-Only Messages

When you launch the Boosted Tree platform via JSL, you can use the following arguments that are available only in JSL:

```
Maximum Depth( number )
Use Excluded Rows for Validation( 0|1 )
```

The following messages for a Boosted Tree object are available only in JSL:

```
boosted_tree_object << Get Average Absolute Error Test;
boosted_tree_object << Get Average Absolute Error Training;
boosted_tree_object << Get Average Absolute Error Validation;
boosted_tree_object << Get Average Log Error Test;
boosted_tree_object << Get Average Log Error Training;
boosted_tree_object << Get Average Log Error Validation;
boosted_tree_object << Get Confusion Matrix Test;
boosted_tree_object << Get Confusion Matrix Training;
boosted_tree_object << Get Confusion Matrix Validation;
boosted_tree_object << Get Confusion Rates Test;
boosted_tree_object << Get Confusion Rates Training;
boosted_tree_object << Get Confusion Rates Validation;
boosted_tree_object << Get Gen RSquare Test;
boosted_tree_object << Get Gen RSquare Training;
boosted_tree_object << Get Gen RSquare Validation;
boosted_tree_object << Get Measures;
boosted_tree_object << Get Microseconds;
boosted_tree_object << Get Misclassification Rate Test;
boosted_tree_object << Get Misclassification Rate Training;
boosted_tree_object << Get Misclassification Rate Validation;
boosted_tree_object << Get MM SAS DATA Step;
boosted_tree_object << Get MM Tolerant SAS DATA Step;
boosted_tree_object << Get Prediction Formula;
boosted_tree_object << Get RMS Error Test;
boosted_tree_object << Get RMS Error Training;
boosted_tree_object << Get RMS Error Validation;
boosted_tree_object << Get ROC Area Test;
```

```
boosted_tree_object << Get ROC Area Training;
boosted_tree_object << Get ROC Area Validation;
boosted_tree_object << Get RSquare Test;
boosted_tree_object << Get RSquare Training;
boosted_tree_object << Get RSquare Validation;
boosted_tree_object << Get SAS DATA Step;
boosted_tree_object << Get Seconds;
boosted_tree_object << Get Tolerant Prediction Formula;
boosted_tree_object << Get Tolerant SAS DATA Step;
```

Bootstrap Forest Script-Only Messages

When you launch the Boosted Tree platform via JSL, you can use the following arguments that are available only in JSL:

```
Portion Bootstrap( fraction )
Time Limit( number )
Use Excluded Rows for Validation( 0|1 )
```

The following messages for a Bootstrap Forest object are available only in JSL:

```
bootstrap_forest_object << Get Average Absolute Error Test;
bootstrap_forest_object << Get Average Absolute Error Training;
bootstrap_forest_object << Get Average Absolute Error Validation;
bootstrap_forest_object << Get Average Log Error Test;
bootstrap_forest_object << Get Average Log Error Training;
bootstrap_forest_object << Get Average Log Error Validation;
bootstrap_forest_object << Get Confusion Matrix Test;
bootstrap_forest_object << Get Confusion Matrix Training;
bootstrap_forest_object << Get Confusion Matrix Validation;
bootstrap_forest_object << Get Confusion Rates Test;
bootstrap_forest_object << Get Confusion Rates Training;
bootstrap_forest_object << Get Confusion Rates Validation;
bootstrap_forest_object << Get Gen RSquare Test;
bootstrap_forest_object << Get Gen RSquare Training;
bootstrap_forest_object << Get Gen RSquare Validation;
bootstrap_forest_object << Get Measures;
bootstrap_forest_object << Get Microseconds;
bootstrap_forest_object << Get Misclassification Rate Test;
bootstrap_forest_object << Get Misclassification Rate Training;
bootstrap_forest_object << Get Misclassification Rate Validation;
bootstrap_forest_object << Get MM SAS DATA Step;
bootstrap_forest_object << Get MM Tolerant SAS DATA Step;
bootstrap_forest_object << Get Prediction Formula;
bootstrap_forest_object << Get RMS Error Test;
bootstrap_forest_object << Get RMS Error Training;
bootstrap_forest_object << Get RMS Error Validation;
```

```
bootstrap_forest_object << Get ROC Area Test;
bootstrap_forest_object << Get ROC Area Training;
bootstrap_forest_object << Get ROC Area Validation;
bootstrap_forest_object << Get RSquare Test;
bootstrap_forest_object << Get RSquare Training;
bootstrap_forest_object << Get RSquare Validation;
bootstrap_forest_object << Get SAS DATA Step;
bootstrap_forest_object << Get Seconds;
bootstrap_forest_object << Get Tolerant Prediction Formula;
bootstrap_forest_object << Get Tolerant SAS DATA Step;
```

Bubble Plot Script-Only Messages

The following messages for a Bubble Plot object are available only in JSL:

```
bubble_plot_object << Auto Stretching( string );
bubble_plot_object << Bubble Size( number );
bubble_plot_object << Color Levels( vector );
bubble_plot_object << Get Custom Path;
bubble_plot_object << Get Draw;
bubble_plot_object << Get Label;
bubble_plot_object << Get Shape;
bubble_plot_object << Go;
bubble_plot_object << Label Offset( list );
bubble_plot_object << Prev;
bubble_plot_object << Set Custom Path( ... );
bubble_plot_object << Speed( number );
bubble_plot_object << Step;
bubble_plot_object << Stop;
bubble_plot_object << Time Index( number );
bubble_plot_object << Title Position( x, y );
bubble_plot_object << Toggle Animation;
```

Choice Script-Only Messages

When you launch the Choice platform via JSL, you can set the acceptable criterion for parameter convergence. In addition, when using Hierarchical Bayes, you can set the number of burn-in iterations.

```
Convergence Criterion( fraction )
Number of Burn In Iterations ( number)
```

Clustering Script-Only Messages

The following messages for a Hierarchical Cluster object are available only in JSL:

```
hierarchical_cluster_object << Get Column Names;
hierarchical_cluster_object << Get Clusters;
hierarchical_cluster_object << GetDisplayOrder;
hierarchical_cluster_object << GetColumnDisplayOrder;
hierarchical_cluster_object << Get Distance Matrix;
hierarchical_cluster_object << Standardize( 0|1 );
```

The following messages for a KDTable object are available only in JSL:

```
kdtable_object << Distance between rows( row1, row2 );
kdtable_object << Insert rows( number|[ vector ] );
kdtable_object << K nearest rows( stop, <position> );
kdtable_object << Remove rows( number|[ vector ] );
```

Contingency Script-Only Messages

The following message is available only in JSL. It creates a data table that contains values from the Correspondence Analysis plot in a Contingency report.

```
contingency_obj << Correspondence Analysis( Make Table );
```

Contour Plot Script-Only Messages

The following message for a Contour Plot object is available only in JSL:

```
contour_plot_object << Set Alpha( number );
```

Contour Profiler Script-Only Messages

The following messages for a Contour Profiler object are available only in JSL:

```
contour_profile_object << Contour Value( y1( ... ), y2( ... ), ... );
contour_profile_object << Get Constraints;
contour_profile_object << Get Factor Settings;
contour_profile_object << Get Factor Settings Script;
contour_profile_object << Get Simulator;
contour_profile_object << Horizontal Factor( column );
contour_profile_object << Number of Plots Across( number );
contour_profile_object << Reset;
contour_profile_object << Term Value( x1( number ), x2( number ), ... );
contour_profile_object << Vertical Factor( column );
contour_profile_object << Y Colors( color1, color2, ... );
```

Control Chart Script-Only Messages

When you launch the Control Chart platform via JSL, you can use the following arguments that are available only in JSL:

```
Chart Col( column, chart_type( Avg( number ), UCL( number ), LCL( number ) )
Chart Type( string )
Get Limits( string )
Get Targets( string ) //Multivariate Control Charts
Use Excluded Points on MR
```

The following messages for a Control Chart object are available only in JSL:

```
control_chart_object << Point Marker( string ); //CUSUM charts
control_chart_object << Save Limits( string );
```

Control Chart Builder Script-Only Messages

When you launch the Control Chart Builder platform via JSL, you can use the following arguments that are available only in JSL:

```
Chart( position( i ), Add Limits( ... ) )
Chart( position( i ), Add Spec Limits( ... ) )
Chart( position( i ), Limits( ... ) )
Chart( position( i ), Points( ... ) )
Chart( position( i ), Set Control Limits( ... ) )
Class( string )
Customize Tests( ... )
KSigma( number )
Show Lower Control Limit( 0|1 )
Show Two Shewhart Charts( 0|1 )
Variables( ... )
Variables( ..., nTrials( column|number ) )
```

The following messages for a Control Chart Builder object are available only in JSL:

```
control_chart_builder_object << Chart( position( i ), ... );
control_chart_builder_object << Rerun All Tests;
control_chart_builder_object << Save Limits( string );
control_chart_builder_object << Size( width, height );
```

Cumulative Damage Script-Only Messages

The following message for a Cumulative Damage object is available only in JSL. It returns a list of model fit results.

```
cumulative_damage_object << Get Results;
```

Custom Profiler Script-Only Messages

The following messages for a Custom Profiler object are available only in JSL:

```
custom_profiler_object << Get Constraints;
custom_profiler_object << Get Factor Settings;
custom_profiler_object << Get Factor Settings Script;
custom_profiler_object << Get Objective;
custom_profiler_object << Get Objective Formula;
custom_profiler_object << Get Simulator;
custom_profiler_object << Objective Formula;
custom_profiler_object << Reset;
custom_profiler_object << Term Value( x1( number ), x2( number ), ... );
```

Degradation Script-Only Messages

The following messages for a Degradation object are available only in JSL:

```
degradation_object << Get Inverse Prediction Results;
degradation_object << Get Prediction Results;
degradation_object << Get Residuals;
degradation_object << Get Results;
degradation_object << Residual Plot;
degradation_object << Set Baseline;
```

Destructive Degradation Script-Only Messages

The following messages for a Destructive Degradation object are available only in JSL:

```
destructive_degradation_object << Control( yTrans, xTrans, distribution, path
  );
destructive_degradation_object << Model( yTrans, xTrans, distribution, path
  );
```

Discriminant Script-Only Messages

The following messages for a Discriminant object are available only in JSL:

```
discriminant_object << Get Descrim Matrices;
discriminant_object << Get Measures;
discriminant_object << Score Data( 0|1 );
```

Distribution Script-Only Messages

The following message for a Distribution object is available only in JSL. The message creates an array of handles to the fitted distributions, so you can send commands to specific distributions:

```
dist_object << ( Fit Handle[n] << ...);
```

You can also get a reference to a fitted distribution directly if you send the Fit Distribution message to a Distribution object. This enables you to send commands to specific distributions:

```
fit_object = dist_object << Fit Distribution( distribution_name );
fit_object << ...;
```

The following message for a Distribution object is available only in JSL. The message specifies that PpK labeling be used in the Process Capability reports within the Distribution report object.

```
dist_object << PpK Capability Labeling( 1 );
```

The following message for a Distribution object is available only in JSL:

```
dist_object << Outlier Box Plot Row Cutoff( number );
```

The following messages for a Distribution object use the name of a fitted distribution, such as `Fit Lognormal`, and they are available only in JSL:

```
dist_object << Fit Distribution Name( Process Capability( Set Proabilities for
  Quantile Spec Limits( ... ) );
dist_object << Fit Distribution Name( Process Capability( Set Sigma Multiplier
  for Quantile Spec Limits( ... ) );
```

EMP Measurement Systems Analysis Script-Only Messages

The following messages for an EMP Measurement Systems Analysis object are available only in JSL:

```
emp_msa_object << Conv Limit( number );
emp_msa_object << Dispersion Chart( 0|1 );
emp_msa_object << Include Interactions in Reproducibility( 0|1 );
emp_msa_object << Max Iter( number );
emp_msa_object << Set Alpha Level( number );
emp_msa_object << Set Random Seed( number );
emp_msa_object << Show Monitor Classification Legend( 0|1 );
emp_msa_object << Show Part Legend( 0|1 );
emp_msa_object << Show Shift Detection Profiler Legend( 0|1 );
```

Explore Missing Values Script-Only Messages

When you launch Automated Data Imputation through the Explore Missing Values utility via JSL, you can use the following arguments that are available only in JSL:

```
Maximum Dimension( number )
Validation Portion( number )
```

The following messages for an Explore Missing Values report are available only in JSL:

```
explore_missing_values_object << Get U V Sigma ADI Matrices;
```

Explore Outliers Script-Only Messages

When you launch K Nearest Neighbors Outliers through the Explore Outliers utility via JSL, you can use the following arguments that are available only in JSL:

```
Impute Missing( 0|1 )
```

The following messages for a Quantile Range Outliers object in Explore Outliers object are available only in JSL:

```
explore_outliers_obj << Get Quantile Outliers;
explore_outliers_obj << Save Quantile Outlier Limits;
```

When you launch Robust PCA Outliers through the Explore Outliers utility via JSL, you can use the following arguments that are available only in JSL:

```
Tolerance( number )
```

Factor Analysis Script-Only Messages

The following messages for a Factor Analysis model fit are available only in JSL. The model fits are numbered sequentially in the order they appear in the report.

```
factor_analysis_object << ( Fit[n] << Save Rotated Components );
factor_analysis_object << ( Fit[n] << Save Rotated Components with Imputation
  );
```

Fit Life by X Script-Only Messages

The following messages for a Fit Life by X object are available only in JSL:

```
fit_life_object << Add Density Curve to Scatterplot( number );
fit_life_object << Add Quantile Line to Scatterplot( number );
fit_life_object << Density( ... );
fit_life_object << Get Results;
fit_life_object << Hazard( ... );
```

```
fit_life_object << Maximum Iterations( number );
fit_life_object << Probability( ... );
fit_life_object << Quantile( ... );
fit_life_object << Rejection Sample Maximum Trials( number );
fit_life_object << Set Level of Quantile Line CI Bands( number );
fit_life_object << Set Scale( distribution );
fit_life_object << Set Scriptables( ... );
fit_life_object << TAF( ... );
fit_life_object << Time Acceleration Baseline( number );
```

Fit Model Script-Only Messages

The following messages for a Model Dialog object are available only in JSL:

```
model_dialog_object << Cauchy Fit( 0|1 ); //Response Screening
model_dialog_object << Choose High Target( 0|1 ); //Nominal Logistic
model_dialog_object << Generalized Distribution( string ); //Generalized
  Regression
model_dialog_object << GLM Distribution( string ); //Generalized Linear Model
model_dialog_object << Max Iterations( number ); //Partial Least Squares
model_dialog_object << Power Link Parameter( number ); //Generalized Linear
  Model
model_dialog_object << Robust Fit( 0|1 ); //Response Screening
model_dialog_object << Run Model;
```

The following sections describe scripting-only messages for MANOVA, Generalized Linear Models, Generalized Regression, Nominal and Ordinal Logistic, Standard Least Squares, and Stepwise.

MANOVA

The following message for a Fit MANOVA object is available only in JSL:

```
fit_model_object << ( Response[1] << (Effect[1] << ... ) );
```

The third message can be one of the following:

```
Test Details // show or hide the test details for an individual effect
Centroid Plot // show or hide the centroid plot for an individual effect
Save Canonical Scores // save the canonical scores for an individual effect
Contrast /* run a customized F test contrasting different levels for
        an effect in the model*/
```

The following example saves canonical scores for the Response Function, which in this data, is Sum:

```
dt = Open( "$SAMPLE_DATA/Dogs.jmp" );
obj = dt << Fit Model(
  Y( :LogHist0, :LogHist1, :LogHist3, :LogHist5 ),
```

```
    Effects( :drug, :dep1, :drug * :dep1 ),
    Personality( "Manova" ),
    Run( Response Function( Sum ) )
);
obj << (Response[1] << (Effect[1] << Save Canonical Scores) );
```

Generalized Linear Models

The following messages for a Fit Generalized Linear Model object are available only in JSL:

```
fit_model_object << GLM Distribution( distribution );
fit_model_object << Parametric Formula;
fit_model_object << Power Link Parameter( number );
```

Generalized Regression

The following messages for a Fit Generalized Regression object are available only in JSL:

```
fit_model_object << Fit( ... );
fit_model_object << Get X Matrix;
```

The following messages for a Fit Generalized Regression object are available only in JSL and must be sent to an individual model fit. The model fits are numbered sequentially in the order they appear in the report.

```
fit_model_object << (Fit[n] << Get Prediction Formula);
fit_model_object << (Fit[n] << Set Solution ID( number ));
fit_model_object << (Fit[n] << Step Backward);
fit_model_object << (Fit[n] << Step Forward);
```

Nominal and Ordinal Logistic

The following messages for Fit Nominal Logistic and Fit Ordinal Logistic objects are available only in JSL:

```
fit_model_object << Get Confusion Matrix Test;
fit_model_object << Get Confusion Matrix Training;
fit_model_object << Get Confusion Matrix Validation;
fit_model_object << Get Confusion Rates Test;
fit_model_object << Get Confusion Rates Training;
fit_model_object << Get Confusion Rates Validation;
fit_model_object << Get MM SAS Data Step;
fit_model_object << Get SAS Data Step;
fit_model_object << Get Measures;
fit_model_object << Get Probability Formulas;
```

The following message for a Fit Nominal Logistic object is available only in JSL:

```
fit_model_object << Positive Level( level );
```

Standard Least Squares

The following messages for a Fit Least Squares object are available only in JSL:

```
fit_model_object << Get Conditional Formula;
fit_model_object << Get Effect Names;
fit_model_object << Get Effect PValues;
fit_model_object << Get Estimates;
fit_model_object << Get Indiv Confid Limit Formula;
fit_model_object << Get Mean Confid Limit Formula;
fit_model_object << Get Measures;
fit_model_object << Get MM SAS DATA Step;
fit_model_object << Get Parameter Names;
fit_model_object << Get Parameterized Formula;
fit_model_object << Get Prediction Formula;
fit_model_object << Get Random Effect Names;
fit_model_object << Get Std Errors;
fit_model_object << Get Standard Error Formula;
fit_model_object << Get Variance Components;
fit_model_object << Get X Matrix;
fit_model_object << Get XPX Inverse;
fit_model_object << Get Y Matrix;
fit_model_object << Profiler( Extrapolation Control Criterion( string ) );
fit_model_object << Show VIF;
```

Stepwise

The following messages for a Fit Stepwise object are available only in JSL:

```
fit_stepwise_object << Backward Step;
fit_stepwise_object << Enter( column );
fit_stepwise_object << Finish; // preferred in scripts over << Go
fit_stepwise_object << Forward Step;
fit_stepwise_object << Get Measures;
fit_stepwise_object << Lock( column );
fit_stepwise_object << Prob to Enter( number );
fit_stepwise_object << Prob to Leave( number );
fit_stepwise_object << Remove( column );
fit_stepwise_object << Unlock( column );
```

Fit Parametric Survival Script-Only Messages

The following messages for a Fit Parametric Survival object are available only in JSL:

```
fit_parametric_survival_object << Estimate Time Quantile( ... );
fit_parametric_survival_object << Get Parameter Names;
fit_parametric_survival_object << Get Estimates;
```

```
fit_parametric_survival_object << Get Std Errors;
fit_parametric_survival_object << Get Effect Names;
fit_parametric_survival_object << Get Effect PValues;
```

Functional Data Explorer Script-Only Messages

The following messages for a Functional Data Explorer FDOE object are available only in JSL:

```
fdoe_object << Generalized Regression FPC Model( FPC Number( n ), Commands );
```

Gaussian Process Script-Only Messages

When you launch the Gaussian Process platform via JSL, you can use the following arguments that are available only in JSL:

```
Set Nugget( number )
```

K Means Cluster Script-Only Messages

The followings messages for a K Means Cluster object are available only in JSL:

```
kmeans_cluster_object << Get Statistics;
```

K Nearest Neighbors Script-Only Messages

When you launch the K Nearest Neighbors platform via JSL, you can use the following arguments that are available only in JSL:

```
Use Excluded Rows for Validation( 0|1 )
```

The following messages for a K Nearest Neighbors object are available only in JSL:

```
knn_object << Get Measures;
knn_object << ( Response[number] << Get Prediction Formula );
knn_object << ( Response[number] << Get Best K );
knn_object << ( Response[number] << Set K( number ) );
```

Latent Class Analysis Script-Only Messages

The following message for a Latent Class Analysis model fit constructs a script to create probability formula columns and returns them to the log. This message is available only in JSL.

```
latent_class_analysis_object << ( Fit[number] << Get Probability Formulas );
```

The following example sets a random seed in order for the model to be reproducible.

```
dt = Open( "$SAMPLE_DATA/Car Poll.jmp" );
obj = dt << Latent Class Analysis(
   Y( :sex, :marital status, :country, :size, :type ),
   Set Random Seed( 12345 ),
   Number of Clusters( 3 ),
);
```

Life Distribution Script-Only Messages

The following messages for a Life Distribution object are available only in JSL:

```
life_dist_obj << Compute Remaining Life Distribution( time0, time1 );
life_dist_object << Confidence Invterval Method( string );
life_dist_object << Density( number );
life_dist_object << Fit DS Frechet;
life_dist_object << Fit DS Loglogistic;
life_dist_object << Fit DS Lognormal;
life_dist_object << Fit DS Weibull;
life_dist_object << Fit Exponential;
life_dist_object << Fit Frechet;
life_dist_object << Fit GenGamma;
life_dist_object << Fit LEV;
life_dist_object << Fit LogGenGamma;
life_dist_object << Fit Logistic;
life_dist_object << Fit Loglogistic;
life_dist_object << Fit Lognormal;
life_dist_object << Fit Model( ... );
life_dist_object << Fit Normal;
life_dist_object << Fit SEV;
life_dist_object << Fit TH Frechet;
life_dist_object << Fit TH Loglogistic;
life_dist_object << Fit TH Lognormal;
life_dist_object << Fit TH Weibull;
life_dist_object << Fit Weibull;
life_dist_object << Fit ZI Frechet;
life_dist_object << Fit ZI Loglogistic;
life_dist_object << Fit ZI Lognormal;
life_dist_object << Fit ZI Weibull;
life_dist_object << Get Causes;
life_dist_object << Get Estimates;
life_dist_object << Get Formula;
life_dist_object << Get Life Distribution( i );
life_dist_object << Get Model Specification;
life_dist_object << Get Results;
```

```
life_dist_object << Hazard( number );
life_dist_object << Mean Remaining Life( Compute( vector ) );
life_dist_object << Omit( i, 0|1 );
life_dist_object << Probability( number );
life_dist_object << Quantile( number );
life_dist_object << Rejection Sample Maximum Trials( number );
life_dist_object << Select Distribution( ... );
life_dist_object << Select Scale( distribution );
life_dist_object << Set Scale( distribution );
life_dist_object << Subdistribution( ... );
life_dist_object << Suppress Plot( plot_name );
```

Make Validation Column Script-Only Messages

When you launch the Make Validation Column platform via JSL, you can use the following arguments that are available only in JSL:

```
Test Set( number )
Training Set( number )
Validation Set( number )
```

MaxDiff Script-Only Messages

When you launch the MaxDiff platform via JSL, you can set the acceptable criterion for parameter convergence. In addition, when using Hierarchical Bayes, you can set the number of burn-in iterations:

```
Convergence Criterion( fraction );
Number of Burn In Iterations ( number);
```

Mixture Profile Script-Only Messages

The following messages for a Mixture Profiler object are available only in JSL:

```
mixture_profiler_object << Contour Value( y1( ... ), y2( ... ), ... );
mixture_profiler_object << Get Constraints;
mixture_profiler_object << Get Factor Settings;
mixture_profiler_object << Get Factor Settings Script;
mixture_profiler_object << Get Simulator;
mixture_profiler_object << Left Factor( column );
mixture_profiler_object << Reset;
mixture_profiler_object << Right Factor( column );
mixture_profiler_object << Term Value( x1( ... ), x2( ... ), ... );
mixture_profiler_object << Top Factor( column );
```

Model Screening Script-Only Messages

When you launch the Model Screening platform via JSL, you can use the following arguments that are available only in JSL:

```
Model NParm Limit( number )
Set Probability Threshold( number )
SVM NRow Limit( number )
```

The following message for a Model Screening object is available only in JSL:

```
model_screening_object << Select Fit( ... );
```

Multivariate Script-Only Messages

The following message for a Multivariate object is available only in JSL:

```
multivariate_object << Get Correlation Matrix;
multivariate_object << Get Inv Correlation Matrix;
```

Naive Bayes Script-Only Messages

When you launch the Naive Bayes platform via JSL, you can use the following arguments that are available only in JSL:

```
Prior Bias( fraction )
Use Excluded Rows for Validation( 0|1 )
```

The following messages for a Naive Bayes object are available only in JSL:

```
naive_bayes_object << Get Measures;
naive_bayes_object << Get Probability Formulas;
```

Neural Script-Only Messages

Note: The Neural platform replaces the Neural Net platform. Scripts that use Neural Net are deprecated and might not work in the future. It is recommended to use the Neural platform for new scripts and update any old scripts using Neural Net to Neural.

When you launch the Neural platform via JSL, you can use the following arguments that are available only in JSL:

```
Multithreading( 0|1 )
N Boost( number )
Validation Method( ... )
```

The following messages for a Neural Fit object are available only in JSL:

```
neural_object << Fit( ... );
neural_object << ( Fit[n] << Get Average Absolute Error Test );
neural_object << ( Fit[n] << Get Average Absolute Error Training );
neural_object << ( Fit[n] << Get Average Absolute Error Validation );
neural_object << ( Fit[n] << Get Average Log Error Test );
neural_object << ( Fit[n] << Get Average Log Error Training );
neural_object << ( Fit[n] << Get Average Log Error Validation );
neural_object << ( Fit[n] << Get Confusion Matrix Test);
neural_object << ( Fit[n] << Get Confusion Matrix Training);
neural_object << ( Fit[n] << Get Confusion Matrix Validation);
neural_object << ( Fit[n] << Get Confusion Rates Test);
neural_object << ( Fit[n] << Get Confusion Rates Training);
neural_object << ( Fit[n] << Get Confusion Rates Validation);
neural_object << ( Fit[n] << Get Gen RSquare Test );
neural_object << ( Fit[n] << Get Gen RSquare Training );
neural_object << ( Fit[n] << Get Gen RSquare Validation );
neural_object << ( Fit[n] << Get Measures );
neural_object << ( Fit[n] << Get Misclassification Rate Test );
neural_object << ( Fit[n] << Get Misclassification Rate Training );
neural_object << ( Fit[n] << Get Misclassification Rate Validation );
neural_object << ( Fit[n] << Get MM SAS DATA Step );
neural_object << ( Fit[n] << Get NBoost );
neural_object << ( Fit[n] << Get Prediction Formula );
neural_object << ( Fit[n] << Get RMS Error Test );
neural_object << ( Fit[n] << Get RMS Error Training );
neural_object << ( Fit[n] << Get RMS Error Validation );
neural_object << ( Fit[n] << Get ROC Area Test);
neural_object << ( Fit[n] << Get ROC Area Training );
neural_object << ( Fit[n] << Get ROC Area Validation);
neural_object << ( Fit[n] << Get RSquare Test );
neural_object << ( Fit[n] << Get RSquare Training );
neural_object << ( Fit[n] << Get RSquare Validation );
neural_object << ( Fit[n] << Get SAS DATA Step );
neural_object << ( Fit[n] << Get Seconds);
```

Nonlinear Script-Only Messages

When you launch the Nonlinear platform via JSL, you can use the following arguments that are available only in JSL:

```
Second Deriv Method( 0|1 )
```

The following messages for a Nonlinear object are available only in JSL:

```
nonlinear_object << CL Alpha( number );
```

```
nonlinear_object << CL Limit( number );
nonlinear_object << Delta( number );
nonlinear_object << Finish;
nonlinear_object << Get CI;
nonlinear_object << Get Corr;
nonlinear_object << Get Cov;
nonlinear_object << Get Estimates;
nonlinear_object << Get Parameter Names;
nonlinear_object << Get SSE;
nonlinear_object << Get Std Errors;
nonlinear_object << Gradient Limit( number );
nonlinear_object << Iteraction Limit( number );
nonlinear_object << Loss is Neg LogLikelihood( 0|1 );
nonlinear_object << Numeric Chain Deriv Delta;
nonlinear_object << Obj Change Limit( number );
nonlinear_object << Relative Gradient( number );
nonlinear_object << Set Parameter( Name = Expr, ... );
nonlinear_object << Unlock Parameter( Name, ... );
```

Normal Mixtures Script-Only Messages

When you launch the Normal Mixtures platform via JSL, you can use the following arguments that are available only in JSL:

```
Mixtures N Starts( number )
Mixtures Tolerance( number )
```

The following messages for a Normal Mixtures object are available only in JSL:

```
normal_mixtures_object << Get Statistics;
```

Oneway Script-Only Messages

The following message for a Oneway object is available only in JSL:

```
oneway_object << Select Group( Xvalue );
```

Parallel Plot Script-Only Messages

The following message for a Parallel Plot object is available only in JSL:

```
parallel_plot_object << Reversed( columns );
```

Pareto Plot Script-Only Messages

The following message for a Pareto Plot object is available only in JSL:

```
No Plot( 0|1 );
```

Partial Least Squares Script-Only Messages

The following messages for a Partial Least Squares object are available only in JSL:

```
pls_object << ( Fit[ number ] << Get Measures ) );
```

Partition Script-Only Messages

When you launch the Partition platform via JSL, you can use the following arguments that are available only in JSL:

```
Initial Splits( ... )
Missing Value Order( Low( list of numeric columns ), High( list of numeric
  columns ) )
Multithreading( 0|1 )
Set Random Seed( number )
```

The following messages for a Partition object that return the requested item from the fitted model are available only in JSL:

```
partition_object << Get Average Absolute Error Test;
partition_object << Get Average Absolute Error Training;
partition_object << Get Average Absolute Error Validation;
partition_object << Get Average Log Error Test;
partition_object << Get Average Log Error Training;
partition_object << Get Average Log Error Validation;
partition_object << Get Confusion Matrix Test;
partition_object << Get Confusion Matrix Training;
partition_object << Get Confusion Matrix Validation;
partition_object << Get Confusion Rates Test;
partition_object << Get Confusion Rates Training;
partition_object << Get Confusion Rates Validation;
partition_object << Get Difference Formula;
partition_object << Get Gen RSquare Test;
partition_object << Get Gen RSquare Training;
partition_object << Get Gen RSquare Validation;
partition_object << Get Measures;
partition_object << Get Microseconds;
partition_object << Get Misclassification Rate Test;
partition_object << Get Misclassification Rate Training;
partition_object << Get Misclassification Rate Validation;
```

```
partition_object << Get MM SAS Data Step;
partition_object << Get MM Tolerant SAS Data Step;
partition_object << Get Prediction Formula;
partition_object << Get RMS Error Test;
partition_object << Get RMS Error Training;
partition_object << Get RMS Error Validation;
partition_object << Get ROC Area Test;
partition_object << Get ROC Area Training;
partition_object << Get ROC Area Validation;
partition_object << Get RSquare Test;
partition_object << Get RSquare Training;
partition_object << Get RSquare Validation;
partition_object << Get SAS Data Step;
partition_object << Get Seconds;
partition_object << Get Tolerant Prediction Formula;
partition_object << Get Tolerant SAS Data Step;
```

Principal Components Script-Only Messages

The following message for a Principal Components object is available only in JSL:

```
prin_comp_object << Save Principal Component Script( number );
```

Process Capability Script-Only Messages

The following messages for a Process Capability object are available only in JSL:

```
process_capability_object << Goal Plot( 1, Capability Lines( number ) );
process_capability_object << Capability Index Plot( 1, Capability Lines(
  number ) );
process_capability_object << Process Performance Plot( 1, Capability
  Boundary( number ) );
process_capability_object << Process Performance Plot( 1, Stability Boundary(
  number ) );
```

Process Screening Script-Only Messages

When you launch the Process Screening platform via JSL, you can use the following arguments that are available only in JSL:

```
Drift Alpha( number )
Make Detailed Shift Data( 0|1 )
Ppk Capability Boundary( number )
Remove( column )
Stability Index Boundary( number )
```

```
Stability Ratio Boundary( number )
```

The following messages for a Process Screening object are available only in JSL:

```
process_screening_object << RowStates;
process_screening_object << Select All;
```

Profiler Script-Only Messages

The following messages for a Profiler object are available only in JSL:

```
profiler_object << Colorize;
profiler_object << Converge Limit( number );
profiler_object << General Extrapolation Control Multiplier( number );
profiler_object << Get Constraints;
profiler_object << Get Desirability;
profiler_object << Get Factor Settings;
profiler_object << Get Factor Settings Script;
profiler_object << Get Main Indices;
profiler_object << Get Simulator;
profiler_object << Get Total Indices;
profiler_object << Max Cycles( number );
profiler_object << MaxIter( number );
profiler_object << Reset;
profiler_object << Term Value( factor( ... ) );
profiler_object << Trips( number )
```

Reliability Forecast Script-Only Messages

The following messages for a Reliability Forecast object are available only in JSL:

```
reliability_forecast_object << Forecast Options( ... );
reliability_forecast_object << Get Results;
```

Reliability Growth Script-Only Messages

The following messages for a Reliability Growth object are available only in JSL:

```
reliability_growth_object << Get Results;
(reliability_growth_object << Mean Time Between Failures Plot) << Options(
  Interval Size( number ) );
(reliability_growth_object << Mean Time Between Failures Plot) << Options(
  Sample MTBF Type( string ) );
```

Response Screening Script-Only Messages

The following message for a Response Screening object is available only in JSL:

```
response_screening_object << Get PValues;
```

Scatterplot 3D Script-Only Messages

The following messages for a Scatterplot 3D object are available only in JSL:

```
scatterplot_3D_object << Frame3D( Add Ellipsoid( ... ) );
scatterplot_3D_object << Frame3D( Add Markers( ... ) );
scatterplot_3D_object << Frame3D( Add Vector( ... ) );
scatterplot_3D_object << Frame3D( Get Axes );
scatterplot_3D_object << Frame3D( Get Box );
scatterplot_3D_object << Frame3D( Get Grab Handles );
scatterplot_3D_object << Frame3D( Get Graph Size );
scatterplot_3D_object << Frame3D( Get Grids );
scatterplot_3D_object << Frame3D( Get Hide Lights Border );
scatterplot_3D_object << Frame3D( get light active( number ) );
scatterplot_3D_object << Frame3D( get light color( number ) );
scatterplot_3D_object << Frame3D( get light position( number ) );
scatterplot_3D_object << Frame3D( Get Line Scale );
scatterplot_3D_object << Frame3D( Get Marker Quality );
scatterplot_3D_object << Frame3D( Get Marker Scale );
scatterplot_3D_object << Frame3D( Get Marker Transparency );
scatterplot_3D_object << Frame3D( Get Rotation );
scatterplot_3D_object << Frame3D( Get Text Scale );
scatterplot_3D_object << Frame3D( Get View Ortho );
scatterplot_3D_object << Frame3D( Get View Perspective );
scatterplot_3D_object << Frame3D( Get View Zoom );
scatterplot_3D_object << Frame3D( Get Wall Color );
scatterplot_3D_object << Frame3D( Get Walls );
scatterplot_3D_object << Frame3D( Get X Axis Color );
scatterplot_3D_object << Frame3D( Get X Axis Label );
scatterplot_3D_object << Frame3D( Get Y Axis Color );
scatterplot_3D_object << Frame3D( Get Y Axis Label );
scatterplot_3D_object << Frame3D( Get Z Axis Color );
scatterplot_3D_object << Frame3D( Get Z Axis Label );
scatterplot_3D_object << Frame3D( Set Axes( 0|1 ) );
scatterplot_3D_object << Frame3D( Set Box( 0|1 ) );
scatterplot_3D_object << Frame3D( Set Graph Size( x, y ) );
scatterplot_3D_object << Frame3D( Set Grids( 0|1 ) );
scatterplot_3D_object << Frame3D( Set Hide Lights Border( 0|1 ) );
scatterplot_3D_object << Frame3D( set light active( ... ) );
scatterplot_3D_object << Frame3D( set light color( ... ) );
```

```
scatterplot_3D_object << Frame3D( set light position( ... ) );
scatterplot_3D_object << Frame3D( Set Line Scale( number ) );
scatterplot_3D_object << Frame3D( Set Marker Quality( number ) );
scatterplot_3D_object << Frame3D( Set Marker Scale( number ) );
scatterplot_3D_object << Frame3D( Set Marker Transparency( number ) );
scatterplot_3D_object << Frame3D( Set Oscillation( ... ) );
scatterplot_3D_object << Frame3D( Set Rotation( ... ) );
scatterplot_3D_object << Frame3D( Set Spin( ... ) );
scatterplot_3D_object << Frame3D( Set Text Scale( number ) );
scatterplot_3D_object << Frame3D( Set View Ortho( 0|1 ) );
scatterplot_3D_object << Frame3D( Set View Perspective( number ) );
scatterplot_3D_object << Frame3D( Set View Zoom( number ) );
scatterplot_3D_object << Frame3D( Set Wall Color( number ) );
scatterplot_3D_object << Frame3D( Set Walls( 0|1 ) );
scatterplot_3D_object << Frame3D( Set X Axis Color( color ) );
scatterplot_3D_object << Frame3D( Set X Axis Label( string ) );
scatterplot_3D_object << Frame3D( Set Y Axis Color( color ) );
scatterplot_3D_object << Frame3D( Set Y Axis Label( string ) );
scatterplot_3D_object << Frame3D( Set Z Axis Color( color ) );
scatterplot_3D_object << Frame3D( Set Z Axis Label( string ) );
scatterplot_3D_object << Frame3D( XAxis( ... ) );
scatterplot_3D_object << Frame3D( YAxis( ... ) );
scatterplot_3D_object << Frame3D( Z Axis( ... ) );
scatterplot_3D_object << Nonpar Density Contour Settings( ... );
```

Structural Equation Models Script-Only Messages

The following messages for a Structural Equation Models object enable you to specify aspects of the model specification, the fitted model, and the path diagram, respectively. These messages are available only in JSL.

```
structural_equation_models_object << Model Specification( ... );
structural_equation_models_object << Fit( ... );
structural_equation_models_object << Path Diagram Properties( ... );
```

The following messages are available within a Model Specification() or Fit() message sent to a Structural Equation Models object.

```
structural_equation_models_object << Fit( Covariances( ... ) );
structural_equation_models_object << Fit( Loadings( ... ) );
structural_equation_models_object << Fit( Means( ... ) );
structural_equation_models_object << Fit( Model Name( string ) );
structural_equation_models_object << Fit( New Latent( string ) );
structural_equation_models_object << Fit( Regressions( ... ) );
structural_equation_models_object << Fit( Variances( ... ) );
```

The following messages are available within a Path Diagram Properties message sent to a Structural Equation Models object.

```
SEM_object << Path Diagram Properties( Constant Border Color( string ) );
SEM_object << Path Diagram Properties( Constant Fill Color( string ) );
SEM_object << Path Diagram Properties( Constant Font( font ) );
SEM_object << Path Diagram Properties( Constant Height( number ) );
SEM_object << Path Diagram Properties( Constant Shape( ... ) );
SEM_object << Path Diagram Properties( Constant Size Option( string ) );
SEM_object << Path Diagram Properties( Constant Text Color( string ) );
SEM_object << Path Diagram Properties( Constant Width( number ) );
SEM_object << Path Diagram Properties( Diagram Size( {x, y} ) );
SEM_object << Path Diagram Properties( Latent Border Color( string ) );
SEM_object << Path Diagram Properties( Latent Fill Color( string ) );
SEM_object << Path Diagram Properties( Latent Font( font ) );
SEM_object << Path Diagram Properties( Latent Height( number ) );
SEM_object << Path Diagram Properties( Latent Shape( ... ) );
SEM_object << Path Diagram Properties( Latent Size Option( string ) );
SEM_object << Path Diagram Properties( Latent Text Color( string ) );
SEM_object << Path Diagram Properties( Latent Width( number ) );
SEM_object << Path Diagram Properties( Lock Diagram( 0|1 ) );
SEM_object << Path Diagram Properties( Manifest Border Color( string ) );
SEM_object << Path Diagram Properties( Manifest Fill Color( string ) );
SEM_object << Path Diagram Properties( Manifest Font( font ) );
SEM_object << Path Diagram Properties( Manifest Height( number ) );
SEM_object << Path Diagram Properties( Manifest Shape( ... ) );
SEM_object << Path Diagram Properties( Manifest Size Option( string ) );
SEM_object << Path Diagram Properties( Manifest Text Color( string ) );
SEM_object << Path Diagram Properties( Manifest Width( number ) );
SEM_object << Path Diagram Properties( Path Styles( ... ) );
SEM_object << Path Diagram Properties( Paths Alpha Level( number ) );
SEM_object << Path Diagram Properties( Paths Color( string ) );
SEM_object << Path Diagram Properties( Paths Font( font ) );
SEM_object << Path Diagram Properties( Paths Opacity( number ) );
SEM_object << Path Diagram Properties( Paths Thickness( number ) );
SEM_object << Path Diagram Properties( Place Nodes( ... ) );
SEM_object << Path Diagram Properties( Rotate Loops( ... ) );
```

Support Vector Machines Script-Only Messages

The following messages for a Support Vector Machines object are available only in JSL:

```
svm_object << Get Measures;
svm_object << Get Prediction Formula;
```

Surface Plot Script-Only Messages

The following messages for a Surface Plot object are available only in JSL:

```
surface_plot_object << Frame3D( Get Axes );
surface_plot_object << Frame3D( Get Box );
surface_plot_object << Frame3D( Get Grab Handles );
surface_plot_object << Frame3D( Get Graph Size );
surface_plot_object << Frame3D( Get Grids );
surface_plot_object << Frame3D( Get Hide Lights Border );
surface_plot_object << Frame3D( get light active );
surface_plot_object << Frame3D( get light color );
surface_plot_object << Frame3D( get light position );
surface_plot_object << Frame3D( Get Line Scale );
surface_plot_object << Frame3D( Get Marker Quality );
surface_plot_object << Frame3D( Get Marker Scale );
surface_plot_object << Frame3D( Get Marker Transparency );
surface_plot_object << Frame3D( Get Rotation );
surface_plot_object << Frame3D( Get Text Scale );
surface_plot_object << Frame3D( Get View Ortho );
surface_plot_object << Frame3D( Get View Persepctive );
surface_plot_object << Frame3D( Get View Zoom );
surface_plot_object << Frame3D( Get Wall Color );
surface_plot_object << Frame3D( Get Walls );
surface_plot_object << Frame3D( Get X Axis Color );
surface_plot_object << Frame3D( Get X Axis Label );
surface_plot_object << Frame3D( Get Y Axis Color );
surface_plot_object << Frame3D( Get Y Axis Label );
surface_plot_object << Frame3D( Get Z Axis Color );
surface_plot_object << Frame3D( Get Z Axis Label );
surface_plot_object << Frame3D( Set Axes( 0|1 ) );
surface_plot_object << Frame3D( Set Box( 0|1 ) );
surface_plot_object << Frame3D( Set Graph Size( x, y ) );
surface_plot_object << Frame3D( Set Grids( 0|1 ) );
surface_plot_object << Frame3D( Set Hide Lights Border( 0|1 ) );
surface_plot_object << Frame3D( set light active( light number, 0|1 ) );
surface_plot_object << Frame3D( set light color( light number, red value,
   green value, blue value ) );
surface_plot_object << Frame3D( set light position( light number, X, Y, Z ) );
surface_plot_object << Frame3D( Set Line Scale( number ) );
surface_plot_object << Frame3D( Set Marker Quality( number ) );
surface_plot_object << Frame3D( Set Marker Scale( number ) );
surface_plot_object << Frame3D( Set Marker Transparency( fraction ) );
surface_plot_object << Frame3D( Set Oscillation( X, Y, Z, duration ) );
surface_plot_object << Frame3D( Set Rotation( X, Y, Z ) );
surface_plot_object << Frame3D( Set Spin( dx, dy, sx, sy ) );
```

```
surface_plot_object << Frame3D( Set Text Scale( number ) );
surface_plot_object << Frame3D( Set View Ortho( 0|1 ) );
surface_plot_object << Frame3D( Set View Perspective( fraction ) );
surface_plot_object << Frame3D( Set View Zoom( number ) );
surface_plot_object << Frame3D( Set Wall Color( number ) );
surface_plot_object << Frame3D( Set Walls( 0|1 ) );
surface_plot_object << Frame3D( Set X Axis Color( color ) );
surface_plot_object << Frame3D( Set X Axis Label( string ) );
surface_plot_object << Frame3D( Set Y Axis Color( color ) );
surface_plot_object << Frame3D( Set Y Axis Label( string ) );
surface_plot_object << Frame3D( Set Z Axis Color( color ) );
surface_plot_object << Frame3D( Set Z Axis Label( string ) );
surface_plot_object << Frame3D( XAxis( Min( number ), Max( number ), Inc(
  number ), Format( ) );
surface_plot_object << Frame3D( YAxis( Min( number ), Max( number ), Inc(
  number ), Format( ) );
surface_plot_object << Frame3D( ZAxis( Min( number ), Max( number ), Inc(
  number ), Format( ) );
surface_plot_object << Clip Sheet( 0|1 );
surface_plot_object << Data points Color( color );
surface_plot_object << Datapoints Choice( Off | Points | Needles | Mesh |
  Surface );
surface_plot_object << Dependent Variables Points( 0|1 );
surface_plot_object << Dependent Variables Response Grid( 0|1 );
surface_plot_object << Equation( column, <column>, <column>, <column> );
surface_plot_object << Frame3D( Scatterplot 3D options );
surface_plot_object << Iso Value( id, value );
surface_plot_object << Mode( "Sheet, points" | "Isosurface" | "Density grid" )
surface_plot_object << Resolution( number );
surface_plot_object << Response( column, ... );
surface_plot_object << Response Column Color Theme( Green to Black to Red |
  Green to White to Red | White to Black | Blue to Gray to Red | Blue to Green
  to Red | Spectral | Jet | White to Blue | White to Green | White to Red );
surface_plot_object << Response Column Fill( Solid | Continuous | Gradients |
  Discrete Gradients );
surface_plot_object << Response Column Gradient Lines( 0|1 );
surface_plot_object << Response Column Gradients( number );
surface_plot_object << Set Z Variable( column );
surface_plot_object << SetVariableAxis( column, <Current Value( number )>,
  <Axis Data( axis options )> );
surface_plot_object << SetXVariable( column );
surface_plot_object << SetYVariable( column );
surface_plot_object << SetZAxis( column, <Current Value( number )>, <Axis
  Data( axis options )> );
surface_plot_object << Surface Alpha( number );
```

```
surface_plot_object << Surface Color Method( "Solid" | formula, <"Solid" |
  formula>, <"Solid" | formula>, <"Solid" | formula> );
surface_plot_object << Surface Gradient Type( Solid | Continuous Gradients |
  Discrete Gradients );
surface_plot_object << Surface Gradients( number );
surface_plot_object << Surface Selector( 0|1 );
surface_plot_object << X Grid( 0|1 );
surface_plot_object << X Resolution( number );
surface_plot_object << X Rotate( degrees );
surface_plot_object << Y Grid( 0|1 );
surface_plot_object << Y Resolution( number );
surface_plot_object << Y Rotate( degrees );
surface_plot_object << Z Grid( 0|1 );
surface_plot_object << Z Resolution( number );
surface_plot_object << Z Rotate( degrees );
```

The following messages for a Surface Plot object act on one of the four surfaces or dependent variables and therefore need a number, <n>, where n=1, 2, 3, or 4.

```
surface_plot_object << Clip Sheet<n>( 0|1 );
surface_plot_object << Contour Color<n>( color );
surface_plot_object << Data points Color<n>( color );
surface_plot_object << Datapoints Choice<n>( Off | Points | Needles | Mesh |
  Surface );
surface_plot_object << Mesh Color<n>( color );
surface_plot_object << Response Column Color Theme<n>( Green to Black to Red |
  Green to White to Red | White to Black | Blue to Gray to Red | Blue to Green
  to Red | Spectral | Jet | White to Blue | White to Green | White to Red );
surface_plot_object << Response Column Fill<n>( Solid | Continuous | Gradients
  | Discrete Gradients );
surface_plot_object << Response Column Gradient Lines<n>( 0|1 );
surface_plot_object << Response Column Gradients<n>( number );
surface_plot_object << Show Contour<n>( Off | Below | Above | On Surface );
surface_plot_object << Show Mesh<n>( Off | X | Y | X and Y );
surface_plot_object << Show Surface<n>( Off | Both Sides | Above | Below Only
  );
surface_plot_object << Surface Alpha<n>( number );
surface_plot_object << Surface Color Range<n>( Data | Axis );
surface_plot_object << Surface Color Theme<n>( Green to Black to Red | Green
  to White to Red | White to Black | Blue to Gray to Red | Blue to Green to
  Red | Spectral | Jet | White to Blue | White to Green | White to Red );
surface_plot_object << Surface Color<n>( color );
surface_plot_object << Surface Gradient Type<n>( Solid | Continuous Gradients
  | Discrete Gradients );
surface_plot_object << Surface Gradients<n>( number );
surface_plot_object << Surface Lighting<n>( None | Low Reflection | Normal );
```

Tabulate Script-Only Messages

The following messages for a Tabulate object are available only in JSL:

```
tabulate_object << Add Table( ... );
MC: I decided to ignore the following that must be specified within a Column
  Table message OR a Row Table message OR within Add Table or Modify Table:
  Aggregate Statistics, Analysis Columns, Columns by Categories, Groupign
  Columns, Statistics,
tabulate_object << Change Item Label( ... );
tabulate_object << Display Column Width( ... );
tabulate_object << Modify Table( ... );
tabulate_object << Order by Count( ... );
tabulate_object << Plot Scale( min, max );
tabulate_object << Remove Column Label( Grouping Columns( column ) );
tabulate_object << Retore Column Label( Grouping Columns( column ) );
tabulate_object << Set Format( ... );
tabulate_object << Test Build( Sample Size( number ) );
tabulate_object << Test Data View;
tabulate_object << Uniform plot scale( 0|1 );
```

Ternary Plot Script-Only Messages

The following message for a Ternary Plot object is available only in JSL:

```
ternary_plot_object << Contour Value( [c1, c2, c3, c4, c5] );
```

Text Explorer Script-Only Messages

The following messages for a Text Explorer object are available only in JSL:

```
text_explorer_object << Add Delimiters( string )
text_explorer_object << Add Phrases( list )
text_explorer_object << Add Phrase Exceptions( list )
text_explorer_object << Add Stem Exceptions( list )
text_explorer_object << Add Stem Overrides( list )
text_explorer_object << Add Stop Words( list )
text_explorer_object << Add Stop Word Exceptions( list )
text_explorer_object << Add Recodes( list )
text_explorer_object << Add Recode Exceptions( list )
text_explorer_object << Cloud Width( pixels )
text_explorer_object << Custom Stemmer( function( {string, dot}, ... ) )
text_explorer_object << Latent Class Analysis( Set Random Seed( number ) );
text_explorer_object << Minimum Frequency for Phrase( n )
text_explorer_object << Phrases Alphabetical( Boolean )
text_explorer_object << Save Regex Column( string, <"From Source"> )
```

```
text_explorer_object << Set Delimiters( string )
text_explorer_object << Set Regex( ... )
text_explorer_object << Terms Alphabetical( Boolean )
```

The following messages for a Sentiment Analysis object in the Text Explorer platform are available only in JSL:

```
sentiment_analysis_object << Add Feature Words
sentiment_analysis_object << Add Intensifier Exception Words
sentiment_analysis_object << Add Intensifier Words
sentiment_analysis_object << Add Negation Exception Words
sentiment_analysis_object << Add Negation Words
sentiment_analysis_object << Add Sentiment Exception Words
sentiment_analysis_object << Add Sentiment Words
```

Time Series Script-Only Messages

The following messages for a Time Series object are available only in JSL:

```
time_series_object << Fit Recommended ETS( ... );
time_series_object << Get Model Specs;
time_series_object << Get Models;
time_series_object << Input Series( ... );
time_series_object << Keep Best Models( "AIC" | "SBC" );
time_series_object << Simple Moving Average Centering Method( "No Centering" |
   "Centered" | "Centered and Double Smoothed for Even Number of Terms" );
```

The following messages for a Simple Moving Average object in the Time Series platform are available only in JSL:

```
sma_object << Get Results;
sma_object << Remove Model( ... );
```

Uplift Script-Only Messages

When you launch the Uplift platform via JSL, you can use the following arguments that are available only in JSL:

```
Multithreading( 0|1 )
Set Random Seed( number )
```

The following messages for an Uplift object are available only in JSL:

```
uplift_object << Get Average Absolute Error Test;
uplift_object << Get Average Absolute Error Training;
uplift_object << Get Average Absolute Error Validation;
uplift_object << Get Average Log Error Test;
uplift_object << Get Average Log Error Training;
```

```
uplift_object << Get Average Log Error Validation;
uplift_object << Get Confusion Matrix Test;
uplift_object << Get Confusion Matrix Training;
uplift_object << Get Confusion Matrix Validation;
uplift_object << Get Confusion Rates Test;
uplift_object << Get Confusion Rates Training;
uplift_object << Get Confusion Rates Validation;
uplift_object << Get Difference Formula;
uplift_object << Get Gen RSquare Test;
uplift_object << Get Gen RSquare Training;
uplift_object << Get Gen RSquare Validation;
uplift_object << Get Measures;
uplift_object << Get Microseconds;
uplift_object << Get Misclassification Rate Test;
uplift_object << Get Misclassification Rate Training;
uplift_object << Get Misclassification Rate Validation;
uplift_object << Get Prediction Formula;
uplift_object << Get RMS Error Test;
uplift_object << Get RMS Error Training;
uplift_object << Get RMS Error Validation;
uplift_object << Get ROC Area Test;
uplift_object << Get ROC Area Training;
uplift_object << Get ROC Area Validation;
uplift_object << Get RSquare Test;
uplift_object << Get RSquare Training;
uplift_object << Get RSquare Validation;
uplift_object << Get Seconds;
uplift_object << Get Tolerant Prediction Formula;
```

Variability Chart Script-Only Messages

The following messages for a Variability Chart object are available only in JSL:

```
variability_chart_object << Analysis Type( string );
variability_chart_object << Conv Limit( number );
variability_chart_object << Max Iter( number );
variability_chart_object << Model( string );
variability_chart_object << Number Function Evals( number );
variability_chart_object << "Reduced Gauge R&R Report"n( 0|1 );
variability_chart_object << Set Random Seed( number );
```

Chapter 11

Display Trees

Create and Interact with Windows

After you learn the basics of scripting, you're ready to manipulate JMP reports and create new JMP windows. This chapter focuses on the following topics:

- navigating JMP display trees to manipulate items in report windows
- building new windows with custom results
- building modal windows
- interacting with the script editor
- converting deprecated `Dialog()` scripts to `New Window()` scripts

Display Tree Compatibility Issues

JMP platforms are generally stable from release to release. However, sometimes the structure of reports changes in major releases. When this happens, old scripts can fail because of the new structure. To avoid the continual updating of old scripts, write scripts so that they avoid dependence on the report structure. Using a string to index a box, followed by a numbered subscript to pick a specific child box, is most reliable across releases. See "Create a Reference by Subscripting". See "Compatibility Notes" for compatibility issues in JMP 16.

Note: This chapter shows how to create and use the most common display boxes. All display boxes are documented in the JMP Scripting Index, which is available in the Help menu.

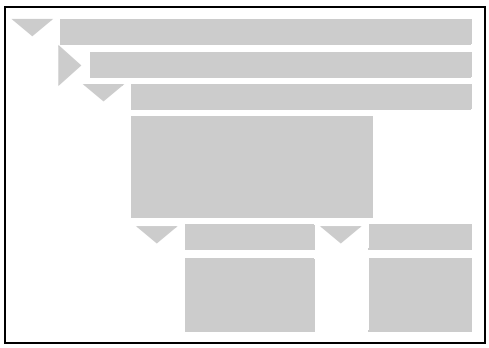
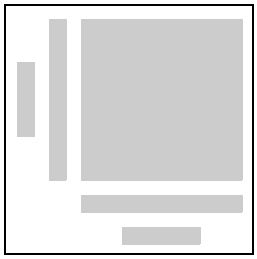
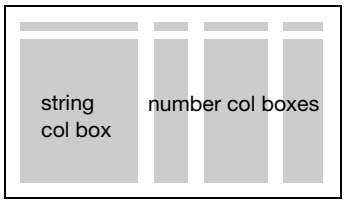
Contents

Navigate JMP Reports 545
Examples of Common Display Boxes 546
View the Display Tree 547
Show Display Box Properties 549
Display Box Object References 550
Send Messages to Display Boxes 555
Example of Creating a Report 562
Construct Custom Windows 565
Example of Creating a Graph Box 565
Extract Values from a Window 566
Constructors for New Windows 568
Interact with a Closing Window 608
Update an Existing Display 610
Set Function and Set Script 614
Get and Set Selected Values of Display Elements That Return Lists 616
Send Messages to Constructed Displays 618
Construct Display Boxes That Contain Platforms 619
Counting the Number of Child Boxes 621
Examples of Creating a Dashboard from Two Reports 622
Example of Creating a Cluster Platform Launch Window 628
Example of Constructing a Custom Platform 631
Example of Specifying Text and Background Colors 634
Modal Windows 635
Construct a Modal Window 635
Construct a Column Dialog 637
Comparison of Column Dialog and New Window 639
Constructors for Column Dialogs 640
Convert Deprecated Dialog to New Window 643
New Window Example 643
Example of Deprecated Dialog 644
Differences Between New Window and the Deprecated Dialog 646
Optional Scripts in New Window 653
Scripting the Script Editor 654
Technical Details 655
Write Tab Box and Tab Page Box Scripts 655

Navigate JMP Reports

As shown in Table 11.1, reports in JMP are built in a hierarchical manner by nesting and concatenating different types of rectangular boxes called *display boxes*. This section shows you how to navigate reports, extract data from them, and customize them.

Table 11.1 Common Display Boxes in JMP Reports

`Outline Box()` creates an outline hierarchy.	
`Picture Box()` combines axes, frames, and labels to make graphs.	
`Table Box()` is a special `H List Box()` whose elements are string and number columns.	

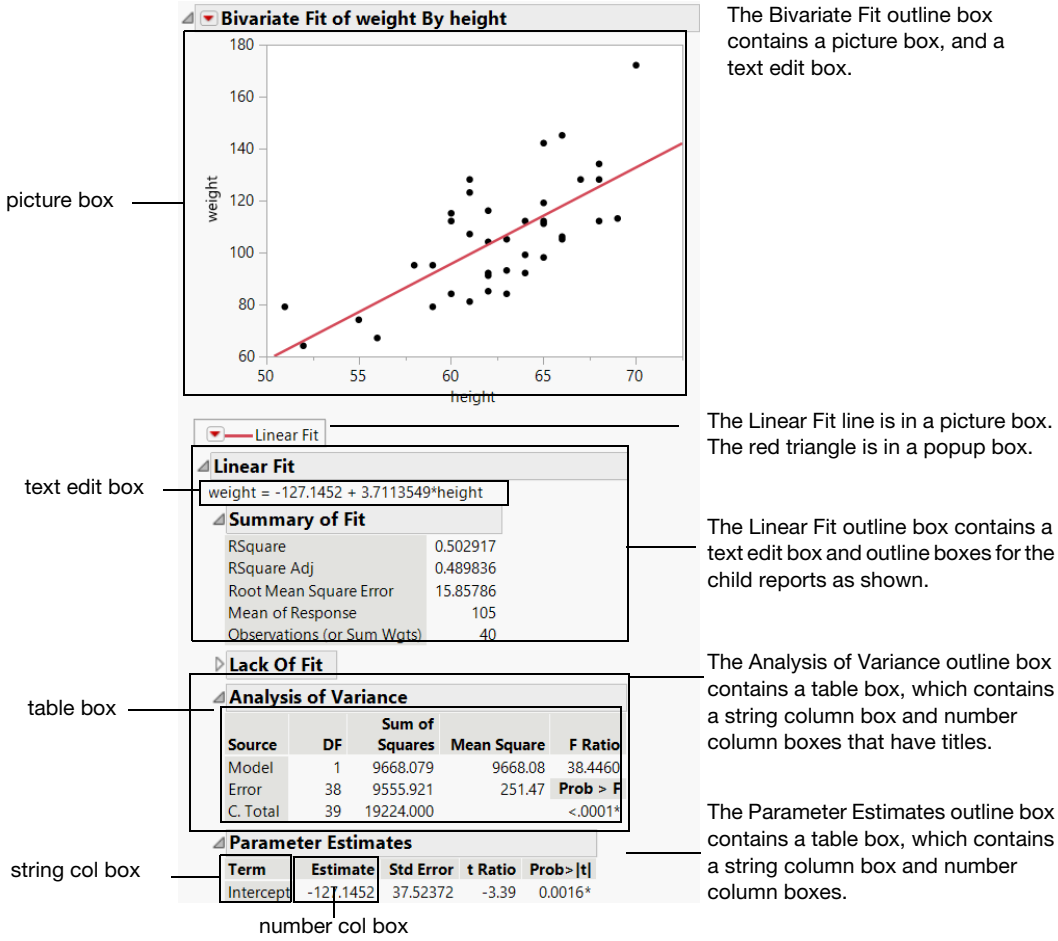
The hierarchical (or parent-child) relationships between display boxes help determine the sizing. The outermost display box is the parent; the inner display boxes are children. When a report is displayed, the parent display box detects the size of the last child box (the leaf) and then detects the size of the child box that precedes the last child box. The size is detected for each box on the way back up the tree. Then the positioning of the boxes takes places from the top down so that the boxes can be shown in an orderly arrangement.

The hierarchical (or parent-child) relationships between display boxes form a tree of display boxes, called the *display tree*. For (parent) boxes that contain other (child) boxes, the parent box's size is determined by the sizes of its child boxes rather than some fixed size, which keeps the whole display resizable. Similarly, positions of boxes are determined by positions of the boxes before them rather than using fixed positions. In tree terminology, the sizing is bottom-up and the positioning is top-down.

Examples of Common Display Boxes

Figure 11.1 shows a Bivariate report that consists of several common display boxes.

Figure 11.1 Display Boxes in a Report



Run the following script to create the report:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( :weight ), X( :height ), Fit Line() );
```

View the Display Tree

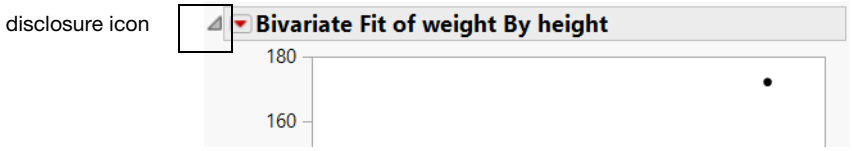
View the report structure to see how display boxes are arranged in the display tree.

1. Run the following script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate(
   Y( :weight ),
   X( :height ),
   Fit Mean( {Line Color( {57, 177, 67} )} )
);
```

2. Right-click the top disclosure icon and select **Edit > Show Tree Structure**.

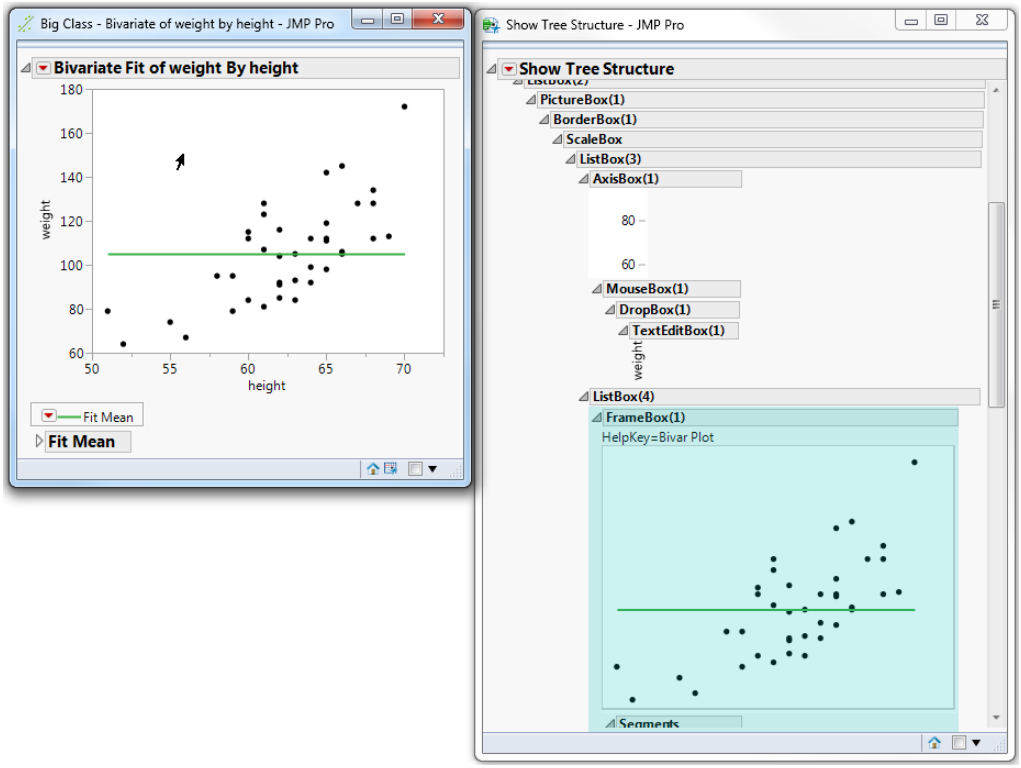
Figure 11.2 Top Disclosure Icon



In the Show Tree Structure window, you see the display box structure of the report as nested outline boxes. Each outline box corresponds to a display box in the original report.

3. Click the Bivariate graph. In Figure 11.3, the frame box that contains the graph is highlighted.

Figure 11.3 Show Tree Structure

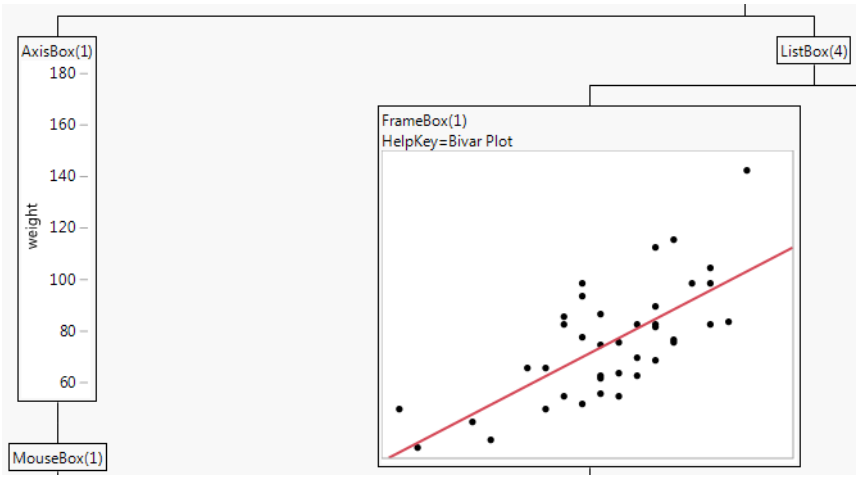


You can also obtain the tree structure through a script. Send the `Show Tree Structure()` message to any report. Or, send the message to a piece of the report (any display box object) to see the tree structure for just that part of the report.

The newer JMP tree structure provides a flat view of the display box tree rather than an outline view. To view the classic structure, press Shift before you right-click the disclosure icon and then select **Edit > Show Tree Structure**. Note the following differences between the newer and classic tree structure views:

- The classic tree structure is not linked to the original report; you cannot highlight specific portions of the tree.
- Siblings appear to the left or right of each other in the classic tree structure. In the newer tree structure, siblings are above or below each other.

Figure 11.4 Partial View of the Classic Tree Structure

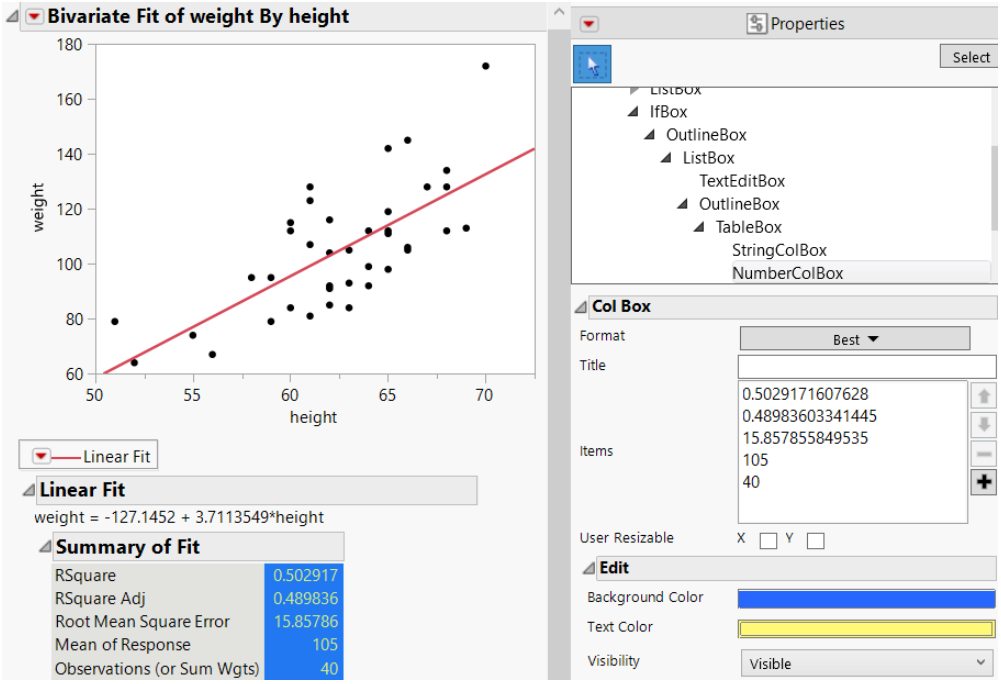


Show Display Box Properties

There is an easy way to format reports before you save them to a script. You can edit the display box properties such as color and position in the Show Tree Structure window. In the report window, select **Edit > Properties** (Windows) or **View > Show Properties** (macOS). Click the node that you want to edit.

In Figure 11.5, the Summary of Fit statistics were selected below the graph. The Background Color and Text Color were then changed in the Properties. These changes are preserved when you save the report as a script or journal the report.

Figure 11.5 Example of Showing Properties for a Plot



Display Box Object References

When you launch a platform, the results are in two places: a visible report, made of display boxes, and an invisible object, called a platform, that accepts commands (or messages).

One of the messages that you can send a platform is the `Report` message. The `Report` message retrieves the display box that holds the report.

The red triangle menu at the top of the report is connected to the platform object and is the way you interactively send messages to the platform.

To fit the mean in the Bivariate platform, assign a reference to the Bivariate platform.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( // biv references the analysis layer
   X( height ),
   Y( weight ),
   Fit Mean,
   Fit Polynomial( 4 )
);
```

To get a reference to the display box that represents the view of the platform, use the `Report` message. For example, add this expression to the preceding script:

```
rbiv = biv << Report; // rbiv references the report layer
```

Now you can manipulate the display boxes in the report layer as described in the following sections.

See What You Can Do with a Display Box

The following example creates a Bivariate report and stores its reference in the `rbiv` report object. It then puts the `rbiv` report object in a frame box and assigns the `fbx` variable.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( // biv references the analysis layer
   Y( :weight ),
   X( :height ),
   Fit Mean( {Line Color( {57, 177, 67} )} )
);
rbiv = biv << Report; // assign rbiv to the report layer
fbx = rbiv[Frame Box( 1 )];
```

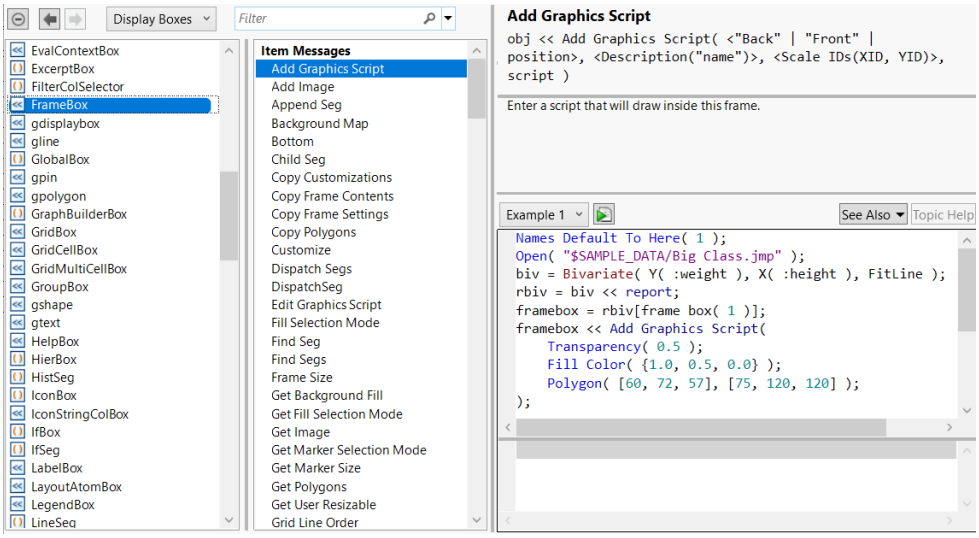
After fbx has been defined, you can send it a message.

```
fbx << Set Background Color( blue ); // colors the frame box in blue
```

The `Set Background Color` message works the same as right-clicking in the graph and selecting **Background Color > Blue**.

To find out which messages you can send to a frame box, select **Help > Scripting Index** and search for “frame box”.

Figure 11.6 Scripting Index Display Boxes List



Running the `Show Properties()` function is an alternative to using the Scripting Index. The messages for the specified display box are printed to the log.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( :weight ), X( :height ) );
Show Properties( biv );
   Show Points [Boolean] [Default On]
   Histogram Borders [Boolean]
   Fit Mean [Action](Fits a flat line at the mean.)
   ...
```

Many messages are the same as items in the pop-up menu for a given object. The next section, "Construct Custom Windows", discusses messages for common display boxes in more detail.

Create a Reference by Subscripting

Use the subscript operator to navigate to a specific display box in the report. The best practice is to refer to the display box with an index number. The following expression identifies the second outline box in the display tree:

```
var = Outline Box[2];
```

Another way to navigate a report is to search for a text string. The following expression identifies the report that has the specified text string:

```
var = report[text];
```

This expression finds the display box in *rpt* that has the title *text*. Note that *text* must be the complete title, not just a substring of the title. The "*text*" argument can also be any expression that evaluates to a text string. This approach is less reliable than referring to the display box with a subscript because the string in a report can easily change.

Typically, you want to assign the identified part of the report to a variable so that you can send messages to it easily. For example, to find the Analysis of Variance outline box in the Bivariate report, use this expression:

```
var = rbiv["Analysis of Variance"];
```

The sections below describe other methods for using a subscript operator to locate display boxes.

Wildcard String

You can use a wildcard character (such as "?") with a substring to match the rest of the outline title. The wildcard character "?" represents places in the search string where you want to match any sequence of any characters. The following script searches for a string that ends with "Mean", finds "Fit Mean", and then opens the Fit Mean outline:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate(
   Y( :weight ),
   X( :height ),
   Fit Mean( {Line Color( {57, 177, 67} )} ),
   Fit Polynomial( 4 )
);
rbiv = biv << Report; // return a reference to the Bivariate report
out1 = rbiv["? Mean"]; // find "Fit Mean" in the Bivariate report
out1 << Close( 0 ); // open the Fit Mean outline
```

Box Type and Wildcard String

Another method for locating display boxes is to include the display box and wildcard search string in the substring operator.

```
var = rbiv[Display Box( "?text" )];
```

This expression finds the outline box that contains the specified text string in the `rbiv` report. The `out1` variable assigns a reference to the outline box.

```
out1 = rbiv[Outline Box( "? Mean" )];
```

Box Type and Index Number

Another method for locating display boxes uses the *boxType* and an index number, *n*.

```
rpt[Display Box ( n )];
```

The expression above finds the *n*th of the specified display box of the type *boxType*. For example, this expression finds the first outline box in the `rbiv` report and assigns it to the `out1` variable:

```
out1 = rbiv[Outline Box( 1 )];
```

This method is less reliable across releases because the number of a display box might change.

Tip: To determine the index number of a display box, view the tree structure as described in "View the Display Tree".

Nested Display Boxes

To reference a display box that appears several layers down in a report, use this format:

```
var = rpt[arg1][arg2][arg3][...];
```

The expression above matches the last argument to a display box that is *contained by* the next to last argument's outline node, which is in turn contained by the argument before the next to last, and so on. In other words, it is a way of digging down into the hierarchy of an outline tree to identify a nested display box.

If you want to identify an item nested several layers down in an outline box, use a subscript with more than one argument of any of the above types. After locating the first item, JMP looks *inside* that item for the next, and so on.

```
rpt[arg1][arg2][arg3] // finds the first item then the next starting after
  that location, and so on
rpt[arg1, arg2, arg3] // different syntax, same result
```

Note that you can string together subscripts that have text or index numbers. The expression below finds the "Parameter Estimate" inside the outline node "Polynomial Fit Degree=4" and selects the third display box in the first table of the outline node.

```
out = rbiv["Polynomial Fit Degree=4"]["Parameter Estimates"][1][3] << Select;
```

"Example of Creating a Report" shows how to extract parameter estimates from the report.

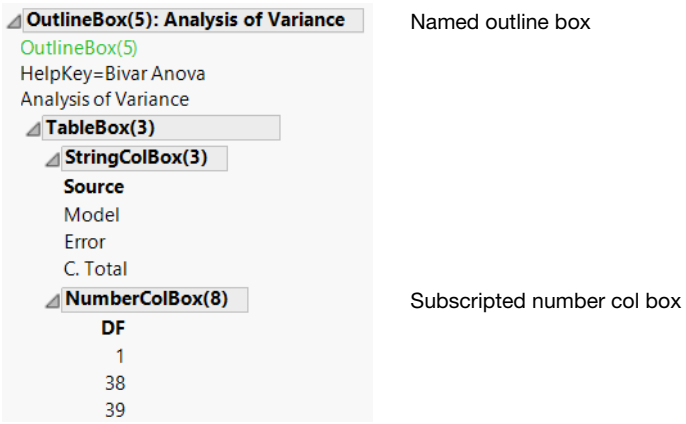
Subscripting Display Boxes for Compatibility

JMP might add or remove display boxes in a report from release to release, causing numbered subscripts to choose a different box. Using a string to index a box, followed by a numbered subscript to pick a specific child box, is more reliable across releases.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
rp = dt << Bivariate(
  Y( :weight ),
  X( :height ),
  Fit Line( {Line Color( {213, 72, 87} )} )
);
```

```
Print(
  "C.Total DF=",

  // string references the parent box
  Report( rp )[Outline Box( "Analysis of Variance" )][
  Number Col Box( 1 )][3] // subscript references the child box
);
```

Figure 11.7 Subscripted Display Box



The numbered subscripts (in the preceding example) are relative to the named box, which makes them less likely to change. However, this also means that the absolute numbers in Show Tree Structure can't be used directly.

Send Messages to Display Boxes

Use the display reference to send messages to the display elements using the `Send` or `<<` operator. For example, if `out2` is a reference to an outline node, you can ask the node to close itself if it is open:

```
out2 << Close(); // close the outline node
```

`Close()` toggles an outline node back and forth between closed and open states, just like selecting items in the red triangle menu. You can include a Boolean argument (1 or 0) for "close or leave closed" or "open or leave open."

```
rbiv["Fit Mean"] << Close;  // toggle open and closed
rbiv["Fit Mean"] << Close( 1 ); // close or leave closed
rbiv["Fit Mean"] << Close( 0 ); // open or leave open
```

Tip: The Report Invalid Display Box Messages option in the General preferences helps you identify invalid display box messages.

- The preference is off by default. When a display box receives an invalid message, JMP ignores those errors and continues executing the script.
- When this preference is on, an error appears in the log.

This option can be useful during script development but can cause unwanted log messages for existing scripts.

The << Operator

You can send messages to display boxes using the send (<<) operator. The operator makes it clear when the script is evaluating child arguments and optional arguments. It also helps make it clearer which argument is an option and which is a script to run inside the graph.

Here is an example of sending a Graph Box a message:

```
win = New Window( "Messages",
   gb = Graph Box(
      Frame Size( 400, 400 ),
      X Scale( 0, 25 ),
      Y Scale( 0, 25 )
   )
);
gb << Background Color( "red" );
```

Nesting and Precedence

When you stack (or *nest*) messages, each message is evaluated left to right.

- The following expression sends `message1` to `box`, then `message2` to `box`, then `message3` to `box`. Note that any of the messages can change `box` before the next message is sent.

  ```
  box << message1 << message2 << message3;
  ```

- The following expression sends `message1` to `box` and gets the result. `Message2` is sent to that result, and `message3` is sent to the result of `message2`.

  ```
  ( (box << message1) << message2) << message3
  ```

- The result of `message3` is assigned to the variable `x`.

  ```
  x = box << message1 << message2 << message3;
  ```

Here is an example of using nested messages:

```
win = New Window( "Messages", gb = Graph Box() );
```

```
// message 1 sets the background color
sz = gb << Background Color( "red" )

// message 2 saves the graph box as a PNG file
<<Save Picture( "$DOCUMENTS/red.png", "png" )

// message 3 sets the graph box background color to white
<<Background Color( "white" )

// message 4 returns the graph box size and prints it to the log
<<Get Size();
```

If you nest several messages, you might want to use parentheses to group the messages to be sure your script does what you want it to.

Customize Reports with the Send To Report and Dispatch Functions

The `Send To Report()` and `Dispatch()` functions provide a way for JMP to record display box modifications inside an analysis script in a self-contained way. For example, arguments within the two functions open and close outline nodes, resize graphics frames, or customize the colors in a graphics frame. You can also specify the arguments in separate JSL statements.

`Send To Report()` contains a list of options that change the display tree.

`Dispatch()` contains four arguments that change the display tree:

- The first argument is a list of outline nodes that need to be traversed to find the desired part of the display tree.
- The second and third arguments work together. The second argument is the name of a display element, and the third argument is the display element's type. These two arguments specify which particular part of the display tree is to be customized.
- The fourth argument is a list of options.

For example, open Big Class.jmp and run the attached Bivariate script. This generates a report with a fitted line. Open the Lack of Fit outline node, and close the Analysis of Variance outline node. Select **Save Script > To Script Window**. The following script appears in the script editor window:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate(
   Y( :weight ),
   X( :height ),
   Fit Line(),
   SendToReport(
      Dispatch(
         {"Linear Fit"},
         "Lack Of Fit",
```

```
            OutlineBox,
            {Close( 0 )} ),
         Dispatch(
            {"Linear Fit"},
            "Analysis of Variance",
            OutlineBox,
            {Close( 1 )}
         )
      )
   );
```

The `Send To Report()` function contains two `Dispatch()` functions that customize the default report.

- The first argument finds an outline node named “Linear Fit”.
- The second and third arguments finds an outline box named “Lack of Fit” below the “Linear Fit” outline.
- The fourth argument is the argument to send to this outline box. In this case, the message is `Close(0)`, in other words, open the node.

Note: If there are several outline nodes with identical names, subscripts are assigned to them. For example, if you have a Bivariate analysis with two quadratic fits (resulting in identical titles), when you dispatch a command to the second fit, the subscript [2] is added to the duplicated title.

The best way to deal with `Send to Report()` and `Dispatch()` is to first create a customized report interactively. Save the report as a script and then examine the script that JMP generates. Remember: the best JSL writer is JMP itself.

Create a Journal

Creating a journal enables you to capture the contents of a JMP window at a fixed moment in time. Suppose that a JMP window contains a graph and selected check boxes. The journal shows the graph and the check boxes that were selected when you created the journal.

The following examples show what you can do with a journal.

Suppose that you start with a Bivariate report.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( weight ), X( height ) );
rbiv = biv << Report;
```

Journal the results:

```
rbiv << Journal Window;
```

Save the journal to a file:

```
rbiv << Save Journal(
   "$DOCUMENTS/test.jrn"
);
```

Note that the above example uses macOS pathname conventions. On Windows, you can use forward slashes (which are appropriate for all platforms) or backslashes (which are appropriate only for Windows).

When a script uses a By variable, the result is a list of references to the analysis results for each By group. In order to journal all By member parts of the report, you need to use parent messages to get to the top of the report.

For example, the following script creates a Bivariate report with a By group and then journals the entire report:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate(
   Y( weight ),
   X( height ),
   By( sex )
);
( ( Report( biv[1] ) << Parent ) << Parent ) <<
Save Journal( "test.jrn" );
```

Journals with Grid Lines

When axes are collapsed, grid lines disappear when the report is journaled, pasted, or saved as a PDF file. Saving the report from JMP as a picture file or a Microsoft PowerPoint presentation retains the grid lines in the graph.

Using Common Messages

Update a Numeric Column

To update a numeric column in a display box table, use the `Set Values` message.

```
win = New Window( "Numeric Column",
   num = Number Col Box( "Values", [9, 10, 11] )
);
num << Set Values( [1, 2, 3] );
```

The matrix argument specifies the new numbers for the table.

Add the Thousands Separator in a Numeric Column

To add the thousands separator in a numeric column, use the `Set Format` message.

```
New Window( "Example",
   ncb = Number Col Box( "Random Numbers",
      {Random Uniform(), Random Uniform() * 10,
      Random Uniform() * 100, Random Uniform() * 1000,
      Random Uniform() * 10000}
   )
);
ncb << Set Format(10,3, "Fixed Dec", "Use thousands separator");
```

Update a String Column

To update a string column in a display box table, use the `Set Values` message.

```
win = New Window( "String Column",
   str = String Col Box( "Values", {"a", "b", "c"} )
);
str << Set Values( {"A", "B", "C"} );
```

The list argument specifies the new strings for the table.

Wrap Text

Generally, JMP automatically wraps text within `Text Box()`. However, the default wrap point can be overridden with a `Set Wrap` (*n*) message. The following script wraps the text at 200 pixels.

```
win = New Window( "Set Wrap",
   tb = Text Box(
      "Hover over a data point for more information."
   )
);
tb << Set Wrap( 200 );
```

Include Bullet Points in Text Boxes

You can add bullet points using the `Bullet Point( 1 )` message.

```
win = New Window( "Bullet List",
   text1 = Text Box( "Hover over a data point for more information." ),
   text2 = Text Box( "Draw a circle around a statistic for more information." )
);
text1 << Bullet Point( 1 );
text2 << Bullet Point( 1 );
```

Sending the `Bullet Point( 1 )` message to a text box places a bullet in front of the text and indents subsequent lines within that text box.

You can also send the `Bullet Point( 1 )` message inline as shown in the following example:

```
text1 = Text Box( "Hover over a data point for more information." ), << Bullet
  Point( 1 ) ),
```

Blink a Selected Display Box

You can use `Select`, `Reshow`, and `Deselect` to blink the selection highlight on a display box. As previously described, you can string together several << clauses to send an object several messages in a row. The results read left to right; for example, here `Select` is done first, then `Reshow`, then `Deselect`, then the other `Reshow`.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate(
  Y( :weight ),
  X( :height ),
  Fit Mean( )
);

// assign the bivariate analysis layer to rbiv
rbiv = biv << Report;

/* find outline box with the specified string in the
rbiv report and assigns it to the out2 variable */
out2 = rbiv[Outline Box( "? Mean" )];
out2 << Close( 0 ); // open the outline box

// find the first string col box in rbiv and
assign the scbx variable */
scbx = rbiv[String Col Box( 1 )];

Wait( .25 ); // wait .25 seconds
For( i = 0, i < 20, i++, // set i to 0, iterate through the loop 20 times

  /* select the string col box and reshow it,
  deselect it and reshow it */
  scbx << Select << Reshow << Deselect << Reshow
);
```

Add Shading or Borders to Table Headings

Headings are shaded and have column borders by default. To turn them off, use the `Set Shade Headings ( 0 )` message and the `Set Heading Column Borders ( 0 )` message. The following example shows the effect of turning shading and borders off:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Summarize(
  meanHt = Mean( Height ),
  minHt = Min( Height ),
```

```
   maxHt = Max( Height )
);
win = New Window( "Summary Results",
   tb = Table Box(
      Number Col Box( "Mean", meanHt ),
      Number Col Box( "Min", minHt ),
      Number Col Box( "Max", maxHt )
   )
);
Wait( 2 );
tb << Set Shade Headings( 0 );
Wait( 1 );
tb << Set Heading Column Borders( 0 );
Wait( 2 );
tb << Set Shade Headings( 1 );
Wait( 1 );
```

Example of Creating a Report

Here is a script to build a report from start to finish.

1. Open a data table.

   ```
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   ```

2. Launch a Bivariate platform and assign it to the platform reference `biv`.

   ```
   /* a reference to the platform */
   biv = dt << Bivariate( Y( weight ), X( height ) );
   ```

3. To find out what you can do with the Bivariate platform, search for Bivariate in the Objects list of the Help > Scripting Index. Here are a few options:

   ```
   biv << Fit Spline( 1000000 ) << Fit Mean;

   // degree 4, orange curve
   biv << Fit Polynomial( 4, {Line Color( RGBColor( 1, 0.5, 0 ) )});
   biv << ( Curve[1] << Line Color( red ) ); // first curve is red
   ```

4. Assign `rbiv` to the report layer:

   ```
   rbiv = biv << Report;
   ```

5. Get the frame box to a comfortable size and scroll up to the top to see the graph.

   ```
   rbiv[Frame Box( 1 )] << Size Window( 500, 700 );
   biv << Scroll Window( {0, 0} );
   ```

6. Get to work on the report. First, you need to create a reference, and then see what you can do with it.

   ```
   Show Properties( rbiv ); // show the properties of the report
   ```

Note: The log lists the messages to use with reports. Using `Show Properties()` is an alternative to the Scripting Index.

```
Close [Boolean]
GetOpen [Action] [Scripting Only]
SetOpen [Boolean] [Scripting Only]
...
```

7. Open the Fit Mean node of the outline:

```
rbiv["Fit Mean"] << Close( 0 );
```

`Close()` is a Boolean function (true or false), and passing 0 indicates that the outline is not closed.

8. Practice selecting some results (submit each line alone to see its result):

```
rbiv["Summary of Fit"] << Select;
rbiv["Parameter Estimates"] << Select;
rbiv["Analysis of Variance"] << Select;
rbiv["Polynomial Fit Degree=4", "Parameter ?", Column Box( "Estimate" )] <<
  Select; // dig down in the outline tree
rbiv << Deselect;
```

9. Change the format of the Estimate column in the 4-degree polynomial Parameter Estimates report.

```
pe = rbiv["Polynomial Fit Degree=4", "Parameter ?"];
ests = pe[Number Col Box( "Estimate" )];
ests << Set Format( 12, 6 );
```

The first argument to `Set Format` sets the column width by the number of characters to display. The second argument sets how many decimal places are shown in the table.

Figure 11.8 Applying Changes to a Report

before

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-17.60008	84.82175	-0.21	0.8368
height	1.9521428	1.346006	1.45	0.1559
(height-62.55)^2	-0.220179	0.29996	-0.73	0.4678
(height-62.55)^3	0.0703948	0.035633	1.98	0.0561
(height-62.55)^4	0.0073899	0.004231	1.75	0.0895

after
Set Format(12,6)

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-17.600085	84.82175	-0.21	0.8368
height	1.952143	1.346006	1.45	0.1559
(height-62.55)^2	-0.220179	0.29996	-0.73	0.4678
(height-62.55)^3	0.070395	0.035633	1.98	0.0561
(height-62.55)^4	0.007390	0.004231	1.75	0.0895

Note: You can even get a single number out of the table. For example, get the estimate for the cubic term by adding the following script to the preceding script:

```
terms = rbiv[Outline Box( "Parameter Estimates" )][String Col Box( "Term"
  )]; // find the Term column
ests = rbiv[Outline Box( "Parameter Estimates" )][Number Col Box(
  "Estimate" )]; // find the Estimate column
estimate = .; // in case there is no cubic term
Try(
    For( i = 1, i < 10, i++,

        // get(i) goes off the end of the table handled by Try
        If( Contains( terms << Get( i ), "^3" ),
         estimate = ests << Get( i );
            Break();
        )
    )
);
Show( estimate );
  0.070394822744608
```

10. Use a `For()` loop to count down to the row for the term that you want. The second argument to `For()` is a condition; as long as the condition tests true, looping continues. Here the test is "when the string in the `Terms` column is not `"^3"` and we have not reached the tenth row." As soon as the string *does* match, looping stops and `i`'s value is the number for the matching row. You then use `i` as a subscript to the `Get` message on the Estimates column.
11. You can also get values from boxes as a matrix, which you can then use for further computations or write to a data table. The following example also shows how to make a data table:

```
// get the values in the Sum of Squares column as a matrix
myVector = rbiv[Table Box( 5 )][Number Col Box( "Sum of Squares" )] << Get as
  Matrix;

/* create a new column in Big Class.jmp named "Sum of Squares"
insert the values of myVector in the column */
dt << New Column( "Sum of Squares", Values( myVector ) );

// put the values of the fifth table box into a new data table
rbiv[Table Box( 5 )] << Make Data Table( "ANOVA table" );
```

12. Now adjust the scales on the axes.

```
rbiv[Axis Box( 1 )] << Min( 70 ) << Max( 170 ); // set the Y axis
rbiv[Axis Box( 2 )] << Min( 50 ) << Max( 70 );  // set X axis
```

13. Copy the graph at the top of the report. Note that you need to select the picture box that contains the graph; selecting just the graph would leave its axes behind.

```
rbiv[Picture Box( 1 )] << Copy Picture;
```

Construct Custom Windows

You can use constructor functions to create objects and construct your own displays. For example, you might prompt the user for information, create custom reports, create custom launch windows, or add check box and radio button items to a window.

Start with a `New Window()`, including a title in quotes, and then construct display boxes inside the window. Make a reference to the new window by assigning it a variable so that you can send messages to it. When display objects are created or referred to by JSL, they are freely shared references until they are copied into another display box or until you close the window and they disappear. When you plug a display object into another display tree, JMP makes a copy of it that the new box owns.

The names of all display box constructors end in “Box”. There are similar objects that live inside graphics frames called display segs (segments such as markers and lines). The names of all display seg objects end in “Seg”.

Example of Creating a Graph Box

The following example creates a new window that contains a graph box:

```
win = New Window( "Simple Earth Map",
   gb = Graph Box(
      X Scale( -180, -60 ),
      Y Scale( 20, 80 ), // end of graph box parameters
      <<Background Map( Images( "Simple Earth" ) ) // beginning of script
   )
);
```

The parameters set the x-axis and y-axis scale and the frame size. The `Background Map` message then shows the map. Note that the parameters for `Graph Box()`, `X Scale` and `Y Scale`, come first. After the first comma, the script begins. If you put the script in the middle of the parameters in a `New Window()` function, the script might run but not produce the results you expect. Scripts are also easier for the users to read if the parameters are at the top of the new window.

Extract Values from a Window

After the user selects options in a window, you can extract those selections from the window using the `Return Result` message, `Get` message, or `Get Selected` message.

Extract Values: Method One

To have a `New Window()` script automatically return the results after the user clicks OK, include the `Return Result` message after the `Modal` message.

```
win = New Window( "Set a Value",
   <<Modal,
   <<Return Result,
   Text Box( "Set this value" ),
   variablebox = Number Edit Box( 42 ),
   Button Box( "OK" ),
   Button Box( "Cancel" )
);
Write( win["variablebox"] ); // create a subscript to the variablebox variable
   33 // the user typed "33" in the number edit box
```

Extract Values: Method Two

Include the `Get` message to return 1 for a selected check box and 0 for a deselected check box. To view the selection, add a `Show()` expression at the end of the script.

```
win = New Window( "V List Box",
   <<Modal,
   V List Box(
      kb1 = Check Box( "a" ),
      kb2 = Check Box( "b" ),
      kb3 = Check Box( "c" )
   ),
   Button Box( "OK",
      val1 = kb1 << Get; // get the value of the first check box
      val2 = kb2 << Get;
      val3 = kb3 << Get;
   )
);
Show( val1, val2, val3 ); // return variables after window closes
   val1 = 1; // first and second checkboxes are selected
   val2 = 1;
   val3 = 0; // third checkbox is not selected
```

Extract Values: Method Three

Include the `Get Selected` message to return the selected column names and then insert the columns in a Bivariate plot.

Note: Method two and method three of extracting values are alternate examples of the same method. The method, in both examples, uses a JSL callback from one of the controls in the window. The JSL callback captures a value from a display box in the window into a global variable that will be available after the window closes. In method two, the JSL callback is on a button box. In method three, the JSL callback is on a list box.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
xvar = .;
yvar = .;
win = New Window( "Return Values",
   <<Modal,
   // require the user to select two variables before clicking OK
   <<On Validate(
      Show( xvar, yvar );
         If( Is Missing( xvar ) | Is Missing( yvar ),
            0, // if xvar or yvar are missing, do nothing when OK is clicked
         1
      );
   ),
   Text Box( " Select two numeric columns. " ),
   H List Box(
      Text Box( " X, Factor " ),
      x = Col List Box(
         dt, // data table reference
         all, // display all columns from the data table
         xvar = (x << Get Selected)[1];
         // get the name of the selected column before the window closes
         Show( xvar );
      ),
      Text Box( "Y, Response" ),
      y = Col List Box(
         dt,
         all,
         yvar = (y << Get Selected)[1];
         Show( yvar );
      )
   )
);
xcol = Column( dt, xvar ); // get the columns
ycol = Column( dt, yvar );
```

```
dt << Bivariate( Y( ycol ), X( xcol ) ); // create a Bivariate plot
```

Constructors for New Windows

JSL supports both modal and non-modal windows.

- Modal windows must be closed before other windows can be used: clicking outside the modal window produces an error sound, and almost all script execution stops until the user closes the modal window.
- JMP reports and platform launch windows are examples of non-modal windows. Many non-modal windows can be open at once and you can switch between them freely. The Transparency dialog in a graph is an example of a modal window. After the Transparency dialog is opened, you must close it before continuing.

This section describes common display boxes that you can construct in a window. There are many other display boxes (such as axis boxes) in JMP that cannot be constructed. Display boxes that cannot be constructed are used by platforms in JMP. Even though they can't be constructed in JSL, they still have functions and messages that can be used from JSL.

Tip: In the Scripting Index (**Help** > **Scripting Index**), display boxes that cannot be constructed are marked by . Display boxes that you can construct are marked by .

Notes:

- The `Dialog()` function, which creates modal windows, is deprecated and will not work at all in future versions of JMP. Use either `New Window()` with the `Modal` message or `Column Dialog()` for column selection. See "Construct a Column Dialog" and "Convert Deprecated Dialog to New Window".
- In JMP, the following display boxes support only one direct child box: `Border Box()`, `Center Box()`, `If Box()`, `Mouse Box()`, `Scroll Box()`, and `Sheet Panel Box()`. To make a border box that appears to have multiple children, use a `V List Box()` or `H List Box()` as the only child and put multiple children in the `V List Box()` or `H List Box()`.

Container Display Boxes

Container display boxes enclose child display boxes.

Border Box

`Border Box()` adds space around the *displaybox* argument.

```
Border Box (Left( pix ), Right( pix ), Top( pix ), Bottom( pix ),
  Sides( int ), display box args);
```

`Left`, `Right`, `Top`, and `Bottom` add space around the *`displaybox`* argument. `Sides` draws a border around the box, as described in Table 11.2. Additional effects can also be applied to the borders using `Sides`, as described in Table 11.2. To add both an effect and a border, add the two numbers.

For example, the following example produces a text box with a border at the top and bottom (Draw Border value of 5) colored blue, the default color (Effect value of 16):

```
win = New Window( "Borders",
   bb = Border Box( Sides( 5 ), // top and bottom border
      Text Box( "Hello World!" )
   )
);
bb << Set Color( "red" ); // set the border color to red
```

The `bb` variable is assigned to the border box so that the color can be set on the last line.

Note: Border boxes support only one display box argument. To make a border box that appears to have multiple children, use a `V List Box()` or `H List Box()` as the only child and put multiple children in the `V List Box()` or `H List Box()`.

In the `Sides()` function, add the following values to get the borders that you want. For example, "5" draws a top and a bottom border.

1: top

2: left

4: bottom

8: right

`Border Box()` provides the additional effects described in Table 11.2.

Table 11.2 Border Box Sides() Additional Effects

Add to Number Above	Additional Effect
16	Make the border lines the same color as the operating system's highlight color. Not recommended because this color is different on different computers.
32	Fill the border box with white before drawing the border box's child.

Table 11.2 Border Box Sides() Additional Effects *(Continued)*

Add to Number Above	Additional Effect
64	Fill the border box with the background color before drawing the border box's child. This option makes a difference only if the border box's parent draws something behind the border box first.

Col List Box

`Col List Box()` returns a display box that shows a list of current data table columns.

```
Col List Box( <Data Table( <name> ), <all>|<character>|<numeric>,
   <Width( n )>, <Max Selected( n )>, <NLines( n )>, <Max Items( n )), <Min
   Items( n )>, <On Change( expression )>, <script> );
```

Note: To add a transform column to a custom display box, use `Filter Col Selector()` instead of `Col List Box()`. See "Filter Col Selector".

`All` specifies that all columns in the current data table should be included. `Width` is measured in pixels. `Max Selected` is the maximum number of items that might be selected in the list box. `NLines` is the number of lines displayed in the box. `Max Items` is the maximum number of items displayed in col list box. `Min Items` is the minimum number of items displayed in the col list box. `On Change` is the script that runs when the selection changes.

You can send a `Get Items` message to a col list box to retrieve a list of all columns selected. Here is an example script showing `Get Items` in use. See the Scripting Index in the Help menu for other messages and examples.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Get Items Demonstration",
   H List Box(
      chooseme =

      // Col List Box shows all columns, 100 pixels wide with 6 lines
      Col List Box( "All", width( 100 ), NLines( 6 ) ),

      /* line up box arranges button boxes in 1 column
      with 2 pixels of space around the boxes*/
      Line Up Box( N Col( 1 ), Spacing( 2 ),

         Button Box( "Add Column >>",
```

```
            /* append the selected column to previously
            selected columns */
            listcols << Append( chooseme << Get Selected );

            // get a list of selected columns from the col list box
               Chosen Columns = listcols << Get Items;
                  ),
         Button Box( "<< Remove Column",
            listcols << Remove Selected; // remove the selected column
            Chosen Columns = listcols << Get Items; // get the selected column
         ),
      ),
         // create another col list box with no columns inside
         listcols = Col List Box( width( 100 ), NLines( 6 ) ),
   ),
   Text Box( " " ),

   /* at the bottom of the h list box,
   show which columns Get Items returned */
   stuff = Global Box( Chosen Columns )

);
```

The `All` argument for `Col List Box()` includes all data table columns in the display box. Because all columns are required, you cannot remove selected rows using `Remove Selected` or `Remove All`. Instead, create an empty column list box and append all columns. Then the columns that you select can be removed.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
colNames = dt << Get Column Names; // get a list of column names

New Window( "Example",
// create Col List Box with all the column names from the data table
myBox = Col List Box( width( 200 ), nlines( 10 ) ) );
myBox << Append( colNames );

myBox << Set Selected( 3 ); // select one or all of the entries in myBox
myBox << Remove Selected(); // remove selected entries in myBox
```

Note: To show only the character or numeric columns from the data table in the col list box, send `Set Data Type(numeric|character)` to the col list box.

Col Span Box

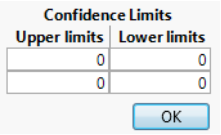
`Col Span Box()` creates spanned columns headers inside a table box. The top column header spans two child column headers.

```
Col Span Box( "title", display box args);
```

In the following script, the Confidence Limits header spans the Upper limits and Lower limits columns below it.

```
win = New Window( "Col Span Box",
   <<Modal,
   Table Box(
      Col Span Box(
         "Confidence Limits",
         neb = Number Col Edit Box( "Upper limits", [0, 0] ),
         Number Col Edit Box( "Lower limits", [0, 0] )
      )
   )
);
```

Figure 11.9 Confidence Limits



Filter Col Selector

`Filter Col Selector()` creates a display box that shows a list of items, for example, columns in a data table. The display box is similar to `Col List Box()` except that `Filter Col Selector()` enables the user to filter columns by name, modeling type, data type, and so on. And `Filter Col Selector()` supports transform columns; `Col List Box()` does not.

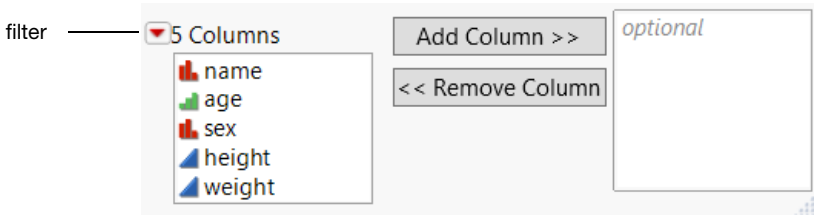
```
Filter Col Selector( <Data Table( name )>, <Width( pixels )>, <Nlines( n )>,
   <script>, <On Change( expr )>, display box args, ... );
```

Here's an example that shows how to use `Get Items` to display a list of selected columns from the filter col selector.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Filter Col Selector Example",
   H List Box(
      chooseme =

      // filter col selector shows all columns, 100 pixels wide with 6 lines
      Filter Col Selector( width( 100 ), NLines( 6 ) ),
```

```
        /* line up box arranges button boxes in 1 column
        with 2 pixels of space around the boxes */
        Line Up Box( N Col( 1 ), Spacing( 2 ),
                Button Box( "Add Column >>",

             /* append the selected column to previously
             selected columns */
             listcols << Append( chooseme << Get Selected );

                // get a list of selected columns from the filter col selector
                Chosen Columns = listcols << Get Items;
                     ),
           Button Box( "<< Remove Column",
              listcols << Remove Selected; // remove the selected column
              Chosen Columns = listcols << Get Items; // get the selected column
           ),
        ),
           // create another col list box with no columns inside
           listcols = Col List Box( width( 100 ), NLines( 6 ) ),
     )
  );
```

Figure 11.10 Filter Col Selector



H List Box and V List Box

`H List Box()` aligns boxes horizontally.

```
H List Box( <Align( center|bottom )>, display box args, ...);
```

`Align` specifies center or bottom alignment of the contents in the display box. The contents are top aligned by default.

The following script aligns the check boxes horizontally:

```
win = New Window( "H List Box",
   <<Modal,
   H List Box(
      kb1 = Check Box( "a" ),
      kb2 = Check Box( "b" ),
      kb3 = Check Box( "c" )
```

```
   ),
);
```

`V List Box()` aligns boxes vertically.

```
V List Box( <Align( center|right )>, display box args, ...);
```

`Align` specifies left, right, or center alignment of the contents in the display box. The contents are left aligned by default.

```
win = New Window( "V List Box",
   Text Box( "Select your favorite fruit." ),
   V List Box(
      Check Box( "apple" ),
      Check Box( "banana" ),
      Check Box( "orange" ),
   ),
);
```

Note: The default maximum width is 180 pixels. To set the width, set the width of the display box that encloses the V List Box.

Suppose that you want to enable the user to add a transform column to custom dialog box. Here is an example of how to create a horizontal list box with an Add and Remove button. The user right-clicks a column to create a transform column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
New Window( "Example",
     H List Box(
           ll1 = Filter Col Selector( ),
           Button Box( "Add", ll2 << Append( ll1 << Get Selected() ) ),
           ll2 = Col List Box( "Numeric", MaxItems( 1 ), nlines( 1 ) ),
           Button Box( "Remove", ll2 << Remove Selected )
     )
);
```

Note: Transform columns use formulas or calculations to define the column values. Closing the window deletes any transform columns. See *Using JMP*.

If Box

`If Box()` creates a display box whose contents are conditionally displayed.

```
If Box( 0|1, display box args );
```

The Boolean argument 1 displays the display boxes inside the if box. 0 does not display them.

The following script creates a splitter box with three scroll boxes. The second scroll box is enclosed in `If Box()`. Because `If Box()` is set to 1, the second scroll box appears in the window.

```
win = New Window( "If Box",
   H Splitter Box(
      Size( 500, 250 ),
      Scroll Box(),
      If Box( 1, Scroll Box() ), // if box sets the scroll box to show
      Scroll Box()
   )
);
```

When you set an if box to be hidden, then anything within the if box disappears from the report window. In the tree structure view, the hidden display box is no longer numbered. Hide the if box in this example, and the second scroll box is no longer numbered in the tree structure view. `Scroll Box(3)` becomes `Scroll Box(2)`. Change the argument for `If Box()` in this script from 1 to 0 and then send the `win << Show Tree Structure` message to see the structure view.

Note: If boxes support only one display box argument.

H Splitter Box and V Splitter Box

`H Splitter Box()` and `V Splitter Box()` create a display box that arranges the display boxes provided by the arguments in a horizontal or vertical layout (or panel). The splitter enables the user to interactively resize the panel.

```
H Splitter Box( <Size( h,v )>, display box args, <arguments> );
V Splitter Box( <Size( h,v )>, display box args, <arguments> );
```

`Size( h, v )` specifies the size of the splitter box in pixels. Inner display boxes are proportionately sized according to the width and height of the splitter box. For a list of arguments, see the Scripting Index in the JMP Help menu.

The following example puts a Bivariate and Oneway graph inside an H splitter box. Splitter boxes are especially useful when the content is set to auto-stretching as shown in the example.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "H Splitter Box",
   H Splitter Box(
      Size( 900, 600 ),
      biv = dt << Bivariate( Y( :weight ), X( :height ), Fit Line(), Line
   Style( Dotted ) ),
      one = Oneway( Y( :height ), X( :sex ), "Means/Anova"n( 1 ), Mean
   Diamonds( 1 ) )
   )
);
```

```
/* make a reference to the report layers
set the x and y axes on the frame boxes to auto stretch
when the user drags the splitter */
(biv << Report)[FrameBox( 1 )] << Set Stretch( "Window", "Window" );
(one << Report)[FrameBox( 1 )] << Set Stretch( "Window", "Window" );
```

Line Up Box

`Line Up Box()` shows the *display box* arguments in *n* columns. You can specify optional spacing, in pixels, for the space between columns.

```
Line Up Box( NCol( n ), <Spacing( pixels )>, display box args, ... );
```

`Spacing` specifies the number of pixels around the box.

In the following example, the Distribution reports are lined up in three columns:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Line Up Box",
   Line Up Box( NCol( 3 ), // line up the Distributions in 3 columns
      dist = Distribution(
         Continuous Distribution( Column( :height ) ),
         By( :age )
      )
   )
);
```

Outline Box

`Outline Box()` creates a display box to hold other display boxes and show them in an outline view.

```
Outline Box("title", display box args, ... );
```

In the following example, each outline contains a Distribution:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Outline Box",
   obj1 = Outline Box( "Age and Weight",
      Distribution( Column( :age, :weight ) ),

   ),
   obj2 = Outline Box( "Weight and Height",
      Distribution( Column( :weight, :height ) )
   )
);
```

To close an outline by default, send the `Close( 1 )` message to the outline object.

```
obj2 << Close( 1 );
```

The first argument for an outline box is the title. You can specify the title in the script, or the title can be based on another component, such as the value of a slider box. In the following example, as the user adjusts the slider, the title of the outline box is updated:

```
sliderValue = .6;
win = New Window( "Outline Box",
   Panel Box( "Slider Box",
      tb = Outline Box( "Value: " || Char( sliderValue ) ),
      sb = Slider Box(
         0,
         1,
         sliderValue,
         tb << Set Title( "Value: " || Char( sliderValue ) )
      )
   )
);
```

To make your own red triangle menu within an outline box, send the `Set Menu Script` message to the outline box reference. The following example also uses the `Set Submenu` message to create nested menu items.

```
win = New Window( "Red Triangle Menu Items",
ob = Outline Box( "Outline Box" ) );
ob << Set Menu Script(
   {"A", "", "A1", Print( "A1" ), "A2", Print( "A2" ), "B", "",
   "B1", Print( "B1" ), "B2", Print( "B2" ), "B3",
   Print( "B3" ), "C", Print( "C" ), "Selected", Print( "Selected" )}
);
ob << Set Submenu( 1, 2 ); // menu A with 2 items in submenu A1 and A2
ob << Set Submenu( 4, 3 ); // menu B with 3 items in submenu B1, B2, and B3
```

To indicate whether an item is selected, unavailable, or available, include `Set Menu Item State`. The following expression shows a check mark beside the ninth menu item (“Selected”) in the preceding script.

```
ob << Set Menu Item State( 9, 1 );
```

Specify the value -1 to disable the menu item or 0 to make it available.

Panel Box

`Panel Box()` encloses the *`displaybox`* argument in a labeled border.

```
Panel Box( "title", display box args );
```

In the following example, the panel box appears around the edge of the Distribution launch window:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Panel Box",
   H List Box( Panel Box( "My Distribution", Distribution() ) )
);
```

Display Element Boxes

Graph Box

`Graph Box()` constructs a display box that contains a frame box and axes.

```
Graph Box( "title", <X Scale( low, high ), <Y Scale( low, high )>, <Frame
   Size( h, v )>,< <XName( "name" )>, <YName( "name" )>, <SuppressAxes>,
   script );
```

`X Scale()` and `Y Scale()` specify the lower and upper limits of the x and y axes. `Frame Size()` specifies the size of the frame around the graph. `XName` and `YName` specify the name of the x and y axes.

The following example creates a simple graph:

```
win = New Window( "Graph Box",
   Graph Box(
      Frame Size( 300, 300 ),
      Marker( Marker State( 3 ), [11 44 77], [75 25 50] );
      Pen Color( "Blue" );
      Line( [10 30 70], [88 22 44] );
   )
);
```

To remove tick marks, add two arguments to the axes.

```
win = New Window( "Graph Box",
   Graph Box(
      Frame Size( 300, 300 ),
      xaxis( // remove tick marks from the x axis
         Show Major Ticks( false ),
         Show Minor Ticks( false ),
         Show Labels( false )
      ),
      yaxis( // remove tick marks from the y axis
         Show Major Ticks( false ),
         Show Minor Ticks( false ),
         Show Labels( false )
      ),
      Marker( Marker State( 3 ), [11 44 77], [75 25 50] );
```

```
        Pen Color( "Blue" );
        Line( [10 30 70], [88 22 44] );
    )
);
```

JMP provides a robust set of features to make your graphs interactive (such as dragging objects or capturing mouse clicks). "Scripting Graphs" describes graph box options in detail.

H Sheet Box and V Sheet Box

`Sheet Box()` lets you create a grid of plots. `H Sheet Box()` and `V Sheet Box()` contain display boxes and arrange them in columns and rows.

```
H Sheet Box( <<Hold( report ), display box args );
V Sheet Box( <<Hold( report ), display box args );
```

The general approach is to first consider which display boxes you want and in which arrangement. Create either an `H` or `V Sheet Box` and send it a `Hold` message for each graph to specify which element the sheet box should hold. Finally, create interior `H` or `V Sheet Box`es and tell each one which graph the sheet box should hold.

Here is an example of creating a sheet with four plots: a Bivariate plot, a Distribution, a Treemap, and a Bubble Plot.

First, open the data table and create a new window.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
nw = New Window( "Sheet Box Example",
```

Use a `V Sheet Box` to organize the window into two columns.

```
    V Sheet Box(
```

Send the `V Sheet Box()` four `Hold` messages, one for each graph. The order matters.

```
        <<Hold(
            dt << Bivariate( // graph 1
                Y( :weight ),
                X( :height ),
                Fit Line()
            )
        ),
        <<Hold(
            Distribution( // graph 2
                Continuous Distribution(
                    Column( :height ),
                    Horizontal Layout( 1 ),
                    Outlier Box Plot( 0 )
                )
            )
```

```
        ),
        <<Hold( Treemap( Categories( :age ) ) ), // graph 3
        <<Hold(
           Bubble Plot( // graph 4
              X( :height ),
              Y( :weight ),
              Sizes( :age ),
              Coloring( :sex ),
              Circle Size( 6.226 ),
              All Labels( 0 )
           )
        ),
```

Finally, add two `H Sheet Box`es to the `V Sheet Box` and tell each one which graph it should hold. Each H `Sheet Box` holds two side-by-side graphs. They are held by a `V Sheet Box`, so the `H Sheet Box`es are displayed vertically.

```
        H Sheet Box(
           Sheet Part(
              "",
```

`Sheet Part` displays a previously defined graph, held by `Excerpt Box`. The first argument is the number of the graph, determined by the order in which you defined the graphs. The first `H Sheet Box` contains the Bivariate graph on the left and the Distribution on the right. `{Picture Box( 1 )}` designates which picture box from the report to display. Typically, use `1`, the first picture box in the report.

```
              Excerpt Box( 1, {Picture Box( 1 )} ) // excerpt box 1
           ),
           Sheet Part(
              "Distribution of height",
              Excerpt Box( 2, {Picture Box( 1 )} ) // excerpt box 2
           )
        ),
        H Sheet Box(
           Sheet Part(
              "",
              Excerpt Box( 3, {Picture Box( 1 )} ) // excerpt box 3
           ),
           Sheet Part(
              "My Title Here",
              Excerpt Box( 4, {Picture Box( 1 )} ) // excerpt box 4
           )
        )
     )
  );
```

The `Sheet Part()` title is required. If you include an empty string as the title, the sheet part's title is the default report title, for example, "Bivariate Fit of weight By height".

Journal Box

The `Journal Box()` constructs a display box from instructions that are stored in a journal.

```
box = Journal Box("journal text");
```

where *"journal text"* is text that has been extracted from a journal file.

Since journal text has lots of rules about which boxes can be with other boxes, we recommend that you get journal text by highlighting an area of the report, selecting Edit > Journal, and save it as a journal. Open the journal in a text editor, and then paste it into your script as the `Journal Box()` argument.

We highly recommend that you use the `"\[ ... ]\"` quoting mechanism so that you do not have to escape double quotes within the journal text. This excerpt shows how to escape the journal text:

```
win = New Window( "Mosaic Plot",
   Journal Box(
      "\[ // beginning of quoting mechanism
   ...display box arguments...
   ]\"
   ) // end of quoting mechanism
);
```

Another way to get journal text is to send the `Get Journal` message to a display box. For example, run the following script and copy the journal text from the log into the `Journal Box()` function:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( y( :weight ), x( :height ) );
rbiv = biv << Report;
Print( rbiv << Get Journal ); // print the journal to the log
```

Picture Object

JSL has a Picture Object, which stores pictures of JMP output or formulas. You can take a picture of anything in a display box or create a picture of a text formula as it would look in the Formula Editor.

To create picture data, send a `Get Picture` message to a *display box*.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( y( :weight ), x( :height ) );
rbiv = biv << Report;
rbiv << Get Picture;
```

The `Expr As Picture()` function evaluates its argument and creates a picture of the expression, using the same formatting mechanism as the Formula Editor. If you have a literal expression as the argument, remember to enclose it in `Expr()` so that JMP just takes a picture of the result. The expression then is not evaluated.

```
win = New Window( "Formula",
   Expr As Picture( Expr( a + b * c + r / d + Exp( x ) ) )
);
```

After you have the picture, there are two ways of using it:

- Add the picture to a new display tree using a display box constructor.
- Write the picture to a file using `Save Picture()`.

  ```
  picture << Save Picture( "path", type );
  ```

  where *type* can be EMF (Windows), JPEG or JPG, GIF, PDF, or PNG.

On Windows, the Windows Specific preferences determine the resolution (or DPI), or you can run this expression:

```
Pref( Save Image DPI( <number> ) );
```

On macOS, the operating system determines the DPI.

Input Selector Display Boxes

Input selector display boxes enable users to enter information and interact with the display box. The display box script includes a script that runs when the user interacts with the display box. See "Set Function and Set Script" for more information about writing the scripts.

Button Box

`Button Box()` draws a button that contains text.

```
Button Box("text", <script>);
```

The following example creates a new window with an OK button and a Cancel button:

```
win = New Window( "Set the Value",
   <<Modal,
   Text Box( "Set this value" ),
   variablebox = Number Edit Box( 42 ),
   Button Box( "OK" ),
   Button Box( "Cancel" ) );
```

Note that a modal window must have at least one button. If you omit the `Button Box()` from your script, an OK button is automatically included. To omit the OK button, set the visibility to collapse.

```
win = New Window( "No OK Button",
```

```
    <<Modal,
    b = Button Box( "Close Window", b << Close Window ),
    bb = Button Box( "OK", <<Visibility( "collapse" ) )
);
```

You can add a tooltip for a button by sending a `Set Tip` message to it. The following script creates a button named "Submit". A tooltip that contains "Send my information." appears when you hover over the button.

```
Button Box( "Submit", << Set Tip( "Send my information." ) );
```

You can also send the message `Open Next Outline` as a script command, which causes the next outline box to open. If you are sending more than one message to the button box, this must be the first command listed. When the user clicks the next button in this example, the second outline box opens.

```
win = New Window( "Open Next Outline",
   Outline Box( "First",
      ex = Button Box( "Next", ex << Open Next Outline )
   ),
   Outline Box( "Second", Text Box( "Last" ), << Close )
);
```

Using the `Set Menu Items` message, you can create a button that contains a menu:

```
win = New Window( "Set and Get Menu Choice",
   bb = Button Box( "Select a dessert",

   // get menu choices when the button is clicked
      choice = bb << Get Menu Choice;
      Show( choice );
   )
);
bb << Set Menu Items( {"fruit", "no thank you", "-", "chocolate bar", "ice
   cream"} );
```

The `"-"` in the preceding script creates a menu separator. The separator counts as an item in the list. If you select `chocolate bar` from the menu, `4` is returned, not `3`.

To indicate whether an item is selected, unavailable, or available, include `Set Menu Item State`. The following example makes the fifth menu item ("ice cream") in the preceding script unavailable:

```
bb << Set Menu Item State( 5, -1 );
```

Specify 0 to make the menu item available or 1 to show a check mark before the menu item.

Note that, after you click the button, the button box doesn't know where the script is stored because the button is *disconnected* from the script. For example, relative paths don't work when executed from a button box because of this disconnection. An `Open()` script must evaluate the current directory and concatenate it to the relative file path to construct an absolute path.

```
nw = New Window( "Open File Example",
   Button Box( "Open",
      Open(
         Eval( Get Path Variable( "$SAMPLE_DATA" ) ) ||
         "Quality Control/Coating.jmp"
      )
   )
);
```

In the following example, the path is interpreted as an absolute path beginning with the root.

```
nw = New Window( "Example",
Button Box( "Open", Open( "/MyData/Big Class.jmp" ) ) );
```

Note: Line-break characters are ignored in button boxes.

Check Box

`Check Box()` constructs a display box to show one or more check boxes. Any number of the check box items can be selected simultaneously.

```
Check Box ({"item 1", "item 2", ...}, <script>);
```

The following example shows how to get the user's selection by sending the `Get Selected` message to the check box. Each selection is printed to the log.

```
win = New Window( "Check Box",
   cb = Check Box(
      {"apple", "banana", "orange"},
      scb = cb << Get Selected();
      Show( scb );
   )
);
   scb = {"banana", "orange"};
```

The following example shows one way to enable the user to click OK to close the window.

```
win = New Window( "Check Box",
   <<Modal,
   <<On Close(
      flagbox1 = ckbox <<Get( 1 );
      flagbox2 = ckbox <<Get( 2 );
      1; // On Close returns non-zero, so the window closes.
   ),
```

```
   ckbox = Check Box( {"Check Box 1", "Check Box 2"} ),
   ckbox << Set( 1, 1 ),
   ckbox << Set( 2, 0 )
);
```

Combo Box

`Combo Box()` creates a drop-down list.

```
Combo Box( {"item 1", "item 2", ...}, <script> );
```

"item1" and "item2" are strings in the drop-down list.

The `Editable` argument allows the user to enter text in the combo box. The following example creates a new window and an editable combo box with the list items "one", "two", and "three". The script prints the selected item name and index number to the log.

```
win = New Window( "Combo Box",
   cb = Combo Box(
      {"One", "Two", "Three"},
      Editable,
      selection = cb << GetSelected();
      Print( "Selected: " || selection );
      Print( "Index: " || Char( cb << Get() ) );
   )
);
```

Note that closing the window for an editable combo box can produce unexpected results. On Windows, the `Print()` script runs after you select an item from the list or enter the item name. On macOS, the `Print()` script runs after you select an item from the list or enter the item name and then press Return.

For cross-platform compatibility, include `!Is Empty` to test for an existing combo box before running the script.

```
win = New Window( "Combo Box",
   cb = Combo Box(
      {"One", "Two", "Three"},
      Editable,
      If( !Is Empty( cb ),
         selection = cb << GetSelected();
         Print( "Selected: " || selection );
         Print( "Index: " || Char( cb << Get() ) );
      )
   )
);
```

To make an item selectable or unselectable, send the `<<Enable Item` message to `Combo Box()`.

```
cb << Enable Item( 2, 0 ) // makes the second item unselectable
```

Note: If the editable combo box has no empty strings, nothing is selected by default. If the editable combo box has one or more empty strings, the first empty string is selected by default

Global Box

`Global Box` shows the name and current value of a JSL global variable.

```
Global Box( name );
```

The user can assign a new value to the global variable by editing the value directly in the window and pressing Enter or Return to commit the change. Global boxes automatically update the displayed value of a variable when the variable changes.

```
ex = Sqrt( 4 );
win = New Window( "Global", Global Box( ex ) );
```

The preceding script creates a new window named “Global” that displays the following result:

ex=2

The following example shows the value of the myGlobal variable on the graph:

```
myGlobal = 6;
   win = New Window( "Global Box",
      V List Box(
          gr = Graph Box(
              Frame Size( 300, 300 ),
              X Scale( 0, 10 ),
              Y Scale( 0, 10 ),
              Y Function( x, x ),
              Text( {5, 5}, "My global is ", myGlobal )
          ),
          Global Box( myGlobal ) // value of myGlobal
      )
   );
```

Note: When you are using `Global Box()`, remember that every time the variable is changed, JMP remeasures, reshows, and updates the window. Depending on the number of `Global Box()` objects, changes to the variable can affect the refresh rate of the window. JMP recommends avoiding numerous global boxes in deliverable scripts. Global boxes can be replaced with text boxes that the script can manually update when necessary.

List Box

`List Box()` creates a display box that shows a list box of items that can be selected.

```
List Box({"item 1", "item 2", ...}, <Width( n )>, <MaxSelected( n )>,
  <NLines( n )>, <script>);
```

Item names are case sensitive by default. `Width` is measured in pixels. `MaxSelected` is the maximum number of items that might be selected in the list box. `NLines` is the number of lines to display in the box, with a default value of 3.

Add items to a list box by either using `Append` or `Insert`:

```
win = New Window( "test", lb = List Box( {"a", "e"} ) );
lb << Append( {"f", "g"} ); // result is a, e, f, g
lb << Insert( {"b", "c", "a", "d"}, 1 ); // result is a, b, c, d, e, f, g
```

`Append` always adds the list to the end of the list box. `Insert` adds the list after the position specified.

You can have two mutually exclusive list boxes, so that the item you select in one box is deselected when you select an item in the other box. Send the `Clear Selection` message to deselect the item.

```
win = New Window( "Each Box Clears the Other Box",
   window:la = List Box(
      {"broccoli", "pepper", "spinach"},
      <<Set Script( window:lb << Clear Selection )
   ),
   window:lb = List Box(
      {"avocado", "pumpkin", "tomato"},
      <<Set Script( window:la << Clear Selection )
   )
);
```

A list box can also contain an image. In the following example, an image from the user's computer appears in the "first" list item. The JMP nominal icon appears in the "second" list item.

```
win = New Window( "List Box with Image",
    List Box(
        {{"first", "$SAMPLE_IMAGES/pi.gif"}, {"second", "nominal"}},
        width( 200 ),
    )
);
```

To set the width of the list box, use the `width` argument.

```
listA = {"a", "b", "c",
"This is a very, very long name like Mary Ann Johnston."};
New Window( "Example",
```

```
// set the width to the length of the string
  V List Box( listBoxA = List Box( listA, width( 0 ) ) ) );
```

To set the font of the list box, use the `Set Base Font` message.

```
New Window( "Example",
   fontobj = lb = List Box(
      {"First Item", "Second Item", "Third Item"},
      width( 200 ),
      max selected( 2 ),
      nlines( 6 )
   )
);
Wait( 2 );
fontobj << Set Base Font( "Title" );
```

`Set Base Font` also works with `Bar Seg`, `Col List Box()`, `Contour Seg`, `Filter Col Selector`, `Hier Box()`, `Number Edit Box()`, `Outline Box()`, `Poly Seg`, `Text Box()`, `Text Edit Box()`, and `Text Seg`.

Mouse Box

`Mouse Box()` creates a box that can make JSL callbacks for dragging and dropping, marking, or clicking and tracking mouse actions.

```
Mouse Box( display box args, messages );
```

The following example is from the DragDrop.jsl sample script in the Samples/Scripts folder. One list box contains “a”, “b”, and “c”. The other list box contains “d”, “e”, “f”. You can drag a letter from one list box to the other list box.

Note: Mouse boxes support only one display box argument.

```
win = New Window( "Drag and drop between list boxes",
   H List Box(
      MouseBox(
         one = List Box( {"a", "b", "c"} ),

         // if the mouse is over the preceding list box, drop on the other box
         <<setTrack(
            (two << parent) << setDropEnable( 1 );
            (one << parent) << setDropEnable( 0 );
         ),

         // watch the mouse movement (not clicks)
         <<setTrackEnable( 1 ),
```

```
            // when a drop happens...
            <<setDropCommit(
               Function( {this, clickPt, text},
               // find out what is selected in the other box
                  x = two << getSelected;
                  If( Words( text, "\!t" ) == x,
                     two << removeSelected; // and take it...
                     one << append( x ); // and add it to me
                  ,
                     Print( "Drag from the other box please." )
                  );
               )
            )
         ),
         Spacer Box( size( 20, 20 ) ),

         /* same comments as above with reversed logic
         for current box versus other box */
         MouseBox(
            two = List Box( {"d", "e", "f"} ),
            <<setTrack(
               (one << parent) << setDropEnable( 1 );
               (two << parent) << setDropEnable( 0 );
            ),
            <<setTrackEnable( 1 ),
            <<setDropCommit(
               Function( {this, clickPt, text},
                  x = one << getSelected;
                  If( Words( text, "\!t" ) == x,
                     one << removeSelected;
                     two << append( x );
                  ,
                     Print( "Drag from the other box please." )
                  );
               )
            )
         )
      )
   );
```

Number Col Edit Box

`Number Col Edit Box()` creates a display box that shows editable numbers in a column. The numeric entries can be in list or matrix form.

```
Number Col Edit Box("title", numbers);
```

The following script creates number edit boxes in which the user enters values. The `Return Result` message extracts the values.

```
x = y = z = 0;
win = New Window( "Number Col Edit Box",
   <<Modal,
   <<Return Result,
   Table Box(
      neb = Number Col Edit Box( "values", {x, y, z} )
   )
);
   {neb = {2, 4, 6}, Button( 1 )}
```

Number Edit Box

`Number Edit Box()` creates an editable number box that initially contains the value argument. You can also set the width of the box in characters.

```
Number Edit Box( initValue, <width>);
```

The following script shows an editable box that initially contains the number "42". The number that the user types is printed to the log.

```
win = New Window( "Set a Value in Number Edit Box",
   <<Modal,
   <<Return Result,
   Text Box( "Set this value" ),
   variablebox = Number Edit Box( 42 ),
   Button Box( "OK" ),
   Button Box( "Cancel" )
);
Write( win["variablebox"] ); // create a subscript to the variablebox variable
   33 // the user typed "33" in the number edit box
```

The following script shows how to make the initial value a date, and enables the user to choose a different date.

```
New Window( "Date/Time Selector",
      <<Modal,
      neb = Number Edit Box(
            Today(),
            23,
            <<Set Format( Format( "m/d/y h:m:s", 23, 0 ) )
      )
);
```

Popup Box

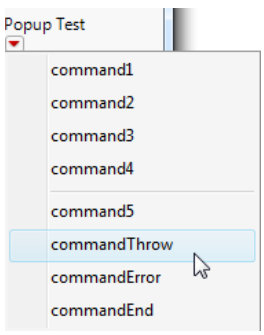
`Popup Box()` creates a red triangle menu that contains menu item commands.

```
Popup Box( {"command1", script1, "command2", script2, ...} );
```

The following example stores a command list in a variable and then uses `Popup Box()` to display the items:

```
List = {
"command1", Print( "command1" ),
"command2", Print( "command2" ),
"command3", Print( "command3" ),
"command4", Print( "command4" ),
"", Empty(), // make a separator line
"command5", Print( "command5" ),
"commandThrow", Throw( "commandThrow1" ),
"commandError", Sqrt( 1, 2, 3 ),
"commandEnd", Print( "commandEnd" )};

win = New Window( "Example",
  Text Box( "Popup Test" ),
  Popup Box( List );
);
```

Figure 11.11 Red Triangle Menu



Note that you can disable and re-enable the menu using the `Enable( Boolean )` message. An argument of 1 turns the menu on, and an argument of 0 turns the menu off. Using the previous example, assign the popup box to a variable and then send messages to it:

```
win = New Window( "Example",
  Text Box( "Popup Test" ),
  mymenu = Popup Box( List );
);

mymenu << Enable( 0 ); // disabled menu
```

You can also add a popup menu to an outline box. See "Outline Box" for an example.

Radio Box

`Radio Box()` creates a list of radio buttons.

```
Radio Box({"item 1", "item 2", ...}, <script>);
```

Only one item in the radio box can be selected at any time.

The following script creates "A", "B", and "C" radio buttons and selects the second box.

```
win = New Window( "Radio Box",
   V List Box(
      rb = Radio Box( {"A", "B", "C"}, <<Set( 2 ) ),
   )
);
```

"New Window Example" also provides an example of creating radio boxes.

Slider Box

`Slider Box()` creates a slider control for picking any value for the variable, within the minimum and maximum range that you specify.

```
Slider Box( min, max, global variable, script, <Set Width( n )>, <Rescale
   Slider( min, max )> );
```

`Set Width` specifies the width of the slider box. `Rescale Slider` specifies the minimum and maximum values.

When the slider is moved, the value given by the current position of the slider is assigned to the global variable. Thus, `Slider Box()` is another way to parameterize a graph.

```
ex = .5;
win = New Window( "Slider Box",
   tb = Text Box( "Value: " || Char( ex ) ),
   sb = Slider Box(
      0,
      1,
      ex,
      tb << Set Text( "Value: " || Char( ex ) )
   )
);
sb << Set Width( 100 ) << Rescale Slider( 0, .8 );
```

The following script creates multiple slider boxes without using a unique global variable to hold the current value for each one:

```
sb1 = Slider Box( 1, 10 );
```

```
// create a slider box with only the min and max value
sb2 = Slider Box( -10, 10 );
sb1 << Set Function( // set the script or the function
   Function( {this},
      {},
      Show( this << Get, sb2 << Get )
   )
);
sb2 << Set Script( Show( sb2 << Get, sb1 << Get ) );
New Window( "Values", sb1, sb2 ); // place the slider boxes in a window
```

As you move the slider, the value is printed to the log.

See "Example of Slider Boxes and Range Slider Boxes" for another example of slider boxes.

Spacer Box

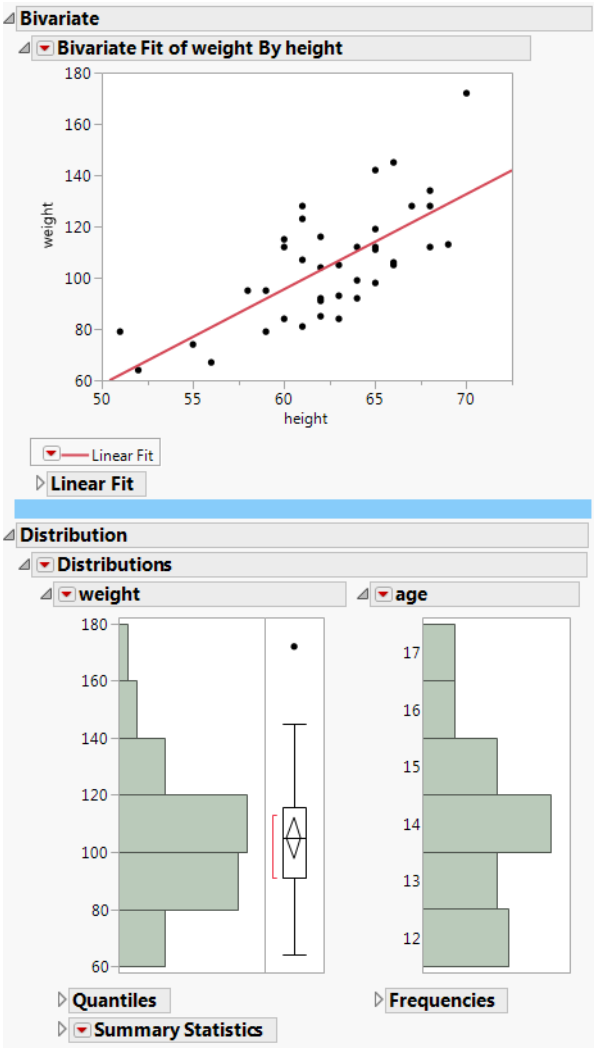
`Spacer Box()` creates space between display boxes.

```
Spacer Box(<size(horizontal_pixels,vertical_pixels)>, <color(color)>)
```

The following script places a 450 pixel wide and 15 pixel tall blue spacer box between the two outline boxes.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Spacer Box",
   V List Box(
      Outline Box( "Bivariate",
         biv = dt << Run Script( "Bivariate" ),
         // Color string consists of percentages of RGB values
         Spacer Box( Size( 450, 15 ), Color( {.53, .80, .98} ), )
      ),
      Outline Box( "Distribution",
         V List Box( dist = dt << Run Script( "Distribution" ), )
      )
   )
);
```

Figure 11.12 Spacer Box Example



String Col Edit Box

`String Col Edit Box()` creates a column named title in a table containing the string items listed. The string boxes are editable.

```
String Col Edit Box( "title", {strings} );
```

The following script creates a column of editable text boxes, preceded by the title “Names”:

```
a = b = c = "";
win = New Window( "String Col Edit Box",
   <<Modal,
   <<Return Result,
```

```
   Table Box(
      seb =
      String Col Edit Box(
         "Names",
         {a, b, c}
         )
      )
);
   {seb = {"apple", "orange", "banana"}, Button( 1 )}
```

Tab Box and Tab Page Box

`Tab Page Box()` organizes the title and contents of the page in one display box. When `Tab Page Box()` is inside `Tab Box()`, a tabbed window with multiple tabbed pages appears.

```
Tab Box( Tab Page Box( Title( "page title 1" ), <options>, contents of
   page 1), Tab Page Box( "page title 2", <options>, contents of page 2),
   ...);
```

The following script creates three tab boxes:

```
win = New Window( "Tab Box",
   Tab Box(
      Tab Page Box(
         Title( "First page" ),
         Text Box( "first line of first page" ),
         Text Box( "second line of first page" )
      ),
      Tab Page Box(
         Title( "Second page" ),
         Text Box( "first line of second page" ),
         Text Box( "second line of second page" )
      ),
      Tab Page Box(
         Title( "Third page" ),
         Text Box( "first line of third page" ),
         Text Box( "second line of third page" )
      )
   )
);
```

Figure 11.13 Tab Boxes

You can specify which tab should be selected by sending the `Set Selected(n)` message, where *n* is the tab number, to the tab box object.

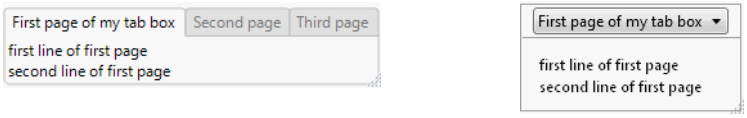
The `Set Style` message lets you select the visual appearance of the tab box. The default value is `tab`. Other options include the following:

- `combo` creates a combo box.
- `outline` creates an outline node.
- `vertical spread` displays the tab title vertically.
- `horizontal spread` displays the tab title horizontally.
- `minimize size` bases the tab style on the width of the tab title. In the following example, the first tab title is fairly long:

```
win = New Window( "Tab Box",
   tb = Tab Box(
      Tab Page Box(
         Title( "First page of my tab box" ),
         Text Box( "first line of first page" ),
         Text Box( "second line of first page" )
      ),
      Tab Page Box(
         Title( "Second page" ),
         Text Box( "first line of second page" ),
         Text Box( "second line of second page" )
      ),
      Tab Page Box(
         Title( "Third page" ),
         Text Box( "first line of third page" ),
         Text Box( "second line of third page" )
      )
   )
);
tb << Set Style( "minimize size" );
```

Minimizing the size of the tab box converts the tab box into a combo box. Figure 11.14 compares the default and minimized tab boxes.

Figure 11.14 Default Tab Box (Left) and Minimized Tab Box (Right)



For `Tab Page Box()` examples, see "Examples of Creating a Dashboard from Two Reports".

Tips:

- Unlike with other display boxes, you must define the title in the `Title()` function.
- The `Set Style` message works for both bare words and quoted words. However, you cannot assign the style to a variable and then pass the variable as the argument.
- You can combine `Tab Box()` with `Splitter Box()` to create docked tabs, which can be dragged to drop zones in a window to create a different layout. See "Examples of Creating a Dashboard from Two Reports".
- To delete a tab, you can either send `Delete(index)` to the tab box, or you can send `Delete Box()` to a tab page box. Note that the entire display box (all tabs) is deleted, not just a single tab.
- Use the `Visibility("Collapsed")` message to hide a tab.
- When tabs in a tab list are not wide enough for the titles, send the `Set Overflow Enabled(1)` message to `Tab Box()`. A triangle symbol is shown to the right of a tab list, and the tab titles are truncated using ellipses.
- When `Tab Page Box()` is created or dragged outside a `Tab Box()`, it is a stand-alone container. The title appears in a shaded box at the top of the page, similar to the title in `Sheet Part()`; the title does not appear in an interactive tab.
- See "Write Tab Box and Tab Page Box Scripts" for more information about the differences between `Tab Box()` and `Tab Page Box()`.
- In Bubble Plot and Graph Builder scripts, set the `Fit to Window` message to `"On"` to resize the window when you resize the graph in `Tab Box()`.
- To close a tab, first make a tab closeable by sending `Closeable(1)` to the tab page box and then send the `Close` message.
- Closeable tabs can be closed with the middle mouse button.

Text Box

`Text Box()` creates a non-editable text box. Text boxes are frequently used as labels of other controls.

```
Text Box( "text" );
```

You can format the text with HTML tags. For example, the following script formats text in bold:

```
win = New Window( "Formatted Text",
  Text Box( "This is <b>bold</b> text.",
  <<Markup) );
```

Make sure you close nested tags correctly and escape the quotes as shown here:

```
"This is <font color=\!"Blue\!" face=\!"Lucida Sans Unicode\!"
  size=\!"24\!">blue</font> text. This is a <background
```

```
color=\!"yellow\!">yellow</background> background. This is <font
color=\!"Blue\!"><background color=\!"yellow\!">blue text on a yellow
background</background></font>."
```

See "Example of Specifying Text and Background Colors" for another example.

Note: Rotated text boxes support `Markup` text except for larger text boxes with multiple lines, multiple formats or justifications applied.

Text Edit Box

`Text Edit Box()` creates an editable text box.

```
Text Edit Box ( "text" );
```

You can add a script to a text edit box by sending it a script message. This is usually most convenient at the time the box is created. Simply add the script message as the last argument.

The following example prints "Changed" in the log each time the text edit box is changed:

```
win = New Window( "Text Edit Box",
   Text Edit Box( "Change Me", <<Script( Print( "Changed" ) ) )
);
```

Assign a reference to the text edit box object to access its contents. The following example prints the value of the text edit box to the log each time the box changes:

```
win = New Window( "Text Edit Box",
   teb = Text Edit Box( "Change Me",
      <<Script( Print( teb << Get Text ) )
   )
);
```

Text edit boxes are empty. However, you can specify placeholder text as a visual cue for the user to enter a value in the box. The placeholder text is only a hint and does not affect the value of the text field.

The following example produces the text edit box shown in Figure 11.15. The `Hint` option returns `"mm/dd/yyyy"` and formats it as light gray text.

```
win = New Window( "Text Edit Box",
   Text Edit Box( "Current Date" ),
   Text Edit Box( "", Hint( "mm/dd/yyyy" ) )
);
```

Figure 11.15 Text Edit Box with Placeholder Text

Tip: To omit the border around the editable box and the white background color, send the `White Box Style( 0 )` message to the text edit box.

If you need to use a text edit box for your user to enter passwords, you can apply a style to the text edit box to replace all characters entered with an asterisk. For example, the following example creates a text edit box whose value is the string "a". When a user types a new string into the text edit box, each character is displayed as an asterisk, and the message "Changed!" is printed to the log.

```
win = New Window( "Text Edit Box",
   teb = Text Edit Box( "a",
      Password Style( 1 ),
      Set Script( Print( "Changed!" ) )
   )
);
```

You can also send a text edit box a message to either use password style or to stop using password style.

```
q << Password Style( 1 ); // set the text edit box to password style
q << Password Style( 0 ); // set the text edit box to standard style
```

Filtering Display Boxes

`Data Filter Source Box()` and `Data Filter Context Box()` work together to filter data in multiple graphs. `Data Filter Source Box()` defines which graph is the "source" of the selection filter. When data are selected in the source graph, graphs in the common `Data Filter Context Box()` are updated.

Data Filter Context Box

`Data Filter Context Box()` creates a display box that enables you to filter data in multiple graphs. The function determines the number of graphs that the filter applies to, for example, all graphs in a report.

```
Data Filter Context Box( display box args );
```

In the following example, the `H List Box()` that contains the graphs is wrapped in `Data Filter Context Box()`. After you run the script, select values in the filter to see both graphs update.

```
dt = Open( "$SAMPLE_DATA/Hollywood Movies.jmp" );
win = New Window( "Shared Local Filter",
   Data Filter Context Box(

   // enclose the h list box that contains distribution and graph builder
      H List Box(
```

```
            dt << Data Filter( Local ), // add the local data filter
            dt << Distribution(
               Weight( :Profitability ),
               Nominal Distribution( Column( :Lead Studio Name ) ),
               Nominal Distribution( Column( :Genre ) ),
               Histograms Only
            ),
            dt << Graph Builder(
               Variables(
                  X( :Genre ),
                  Y( :Domestic Gross ),
                  Y( :Foreign Gross, Position( 1 ) )
               ),
               Show Control Panel( 0 ),
               Elements(
                  Bar(
                     X,
                     Y( 1 ),
                     Y( 2 ),
                     Legend( 2 ),
                     Bar Style( "Side by side" ),
                     Summary Statistic( "Mean" )
                  ),
                  Frequencies( 0 ),
               )
            )
         )
      )
   );
```

Data Filter Source Box

`Data Filter Source Box()` and `Data Filter Context Box()` work together to filter data in multiple graphs. `Data Filter Source Box()` defines which graph is the "source" of the selection filter. When data are selected in the source graph, graphs in the common `Data Filter Context Box()` are updated.

`Data Filter Source Box(` *`display box args`* `);`

The following script lets you select data in Distribution graphs, and only that data is shown in the Graph Builder graph:

```
dt = Open( "$SAMPLE_DATA/Hollywood Movies.jmp" );
win = New Window( "Shared Local Filter",
   Data Filter Context Box(

   // enclose data filter source box and the graph builder platform
      H List Box(
```

```
            Data Filter Source Box(
               dt << Distribution( // use distribution as the source graph
                  Weight( :Profitability ),
                  Nominal Distribution( Column( :Lead Studio Name ) ),
                  Nominal Distribution( Column( :Genre ) ),
                  Histograms Only
               )
            ),
            dt << Graph Builder(
               Variables(
                  X( :Genre ),
                  Y( :Domestic Gross ),
                  Y( :Foreign Gross, Position( 1 ) )
               ),
               Show Control Panel( 0 ),
               Elements(
                  Bar(
                     X,
                     Y( 1 ),
                     Y( 2 ),
                     Legend( 2 ),
                     Bar Style( "Side by side" ),
                     Summary Statistic( "Mean" )
                  ),
                  Frequencies( 0 ),
               )
            )
         )
      )
   );
```

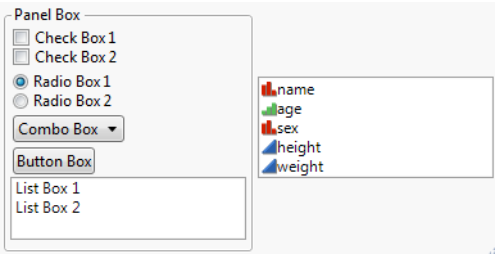
Examples of Combined Display Boxes

The following example generates a sample of many controls illustrated in the previous sections.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Window Controls",
   Line Up Box( N Col( 2 ), Spacing( 3 ),
      Panel Box( "Panel Box",
         Check Box( {"Check Box 1", "Check Box 2"} ),
         Radio Box( {"Radio Box 1", "Radio Box 2"} ),
         Combo Box( {"Combo Box"} ),
         Button Box( "Button Box" ),
         List Box( {"List Box 1", "List Box 2"} )
      ),
      Col List Box( "all" )
```

```
    )
);
```

Figure 11.16 Example of Many Interactive Display Elements



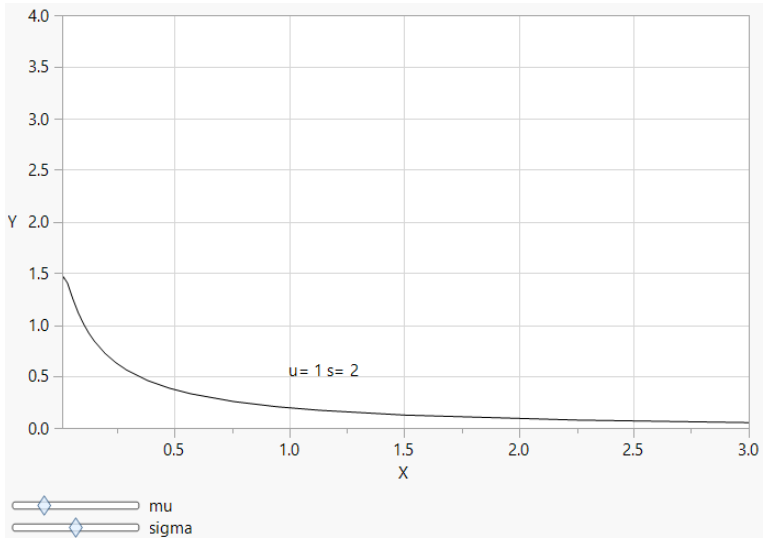
Example of Slider Boxes and Range Slider Boxes

The following example combines a graph with two sliders and a button by gluing the graph box, two horizontal boxes, and a button together in a vertical list box:

```
// Slider LogNormal
lU = 1;
lS = 2;
win = New Window( "LogNormal Density",
   V List Box(
      gr = Graph Box( // define the arguments and script for the graph box
         Frame Size( 500, 300 ),
         X Scale( 0.01, 3 ),
         Y Scale( 0, 4 ),
         XAxis( Show Major Grid ),
         YAxis( Show Major Grid ),
         Y Function(
            Exp( -(Log( x ) - Log( lU )) ^ 2 / (2 * lS ^ 2) ) / (lS * x *
            Sqrt( 2 * Pi() )),
            x
         );

         /* display the specified text at 1 on the x axis
         and .5 on the y axis */
         Text( {1, .5}, "u= ", lU, " s= ", lS );
               ),

      /* define the minimum and maximum range
      insert the value of lU and updates the display as the slider moves */
      H List Box( Slider Box( 0, 4, lU, gr << Reshow ), Text Box( "mu" ) ),
         H List Box(
         Slider Box( 0, 4, lS, gr << Reshow ),
         Text Box( "sigma" )
```

```
        ),
        Button Box( "Reshow", gr << Reshow )
    )
);
Show( gr );
gr << Reshow; // update the display of the Graph Box
```

Figure 11.17 Example of Sliders and Buttons in a Report Window

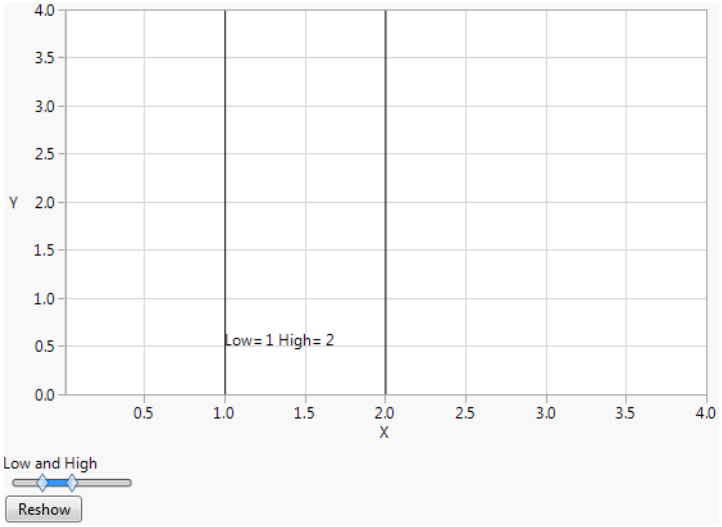


The following script uses a `Range Slider Box()` instead of a `Slider Box()` to perform the same function:

```
lLow = 1;
lHigh = 2;
win = New Window( "Range Slider",
   V List Box(
      gr = Graph Box(
         Frame Size( 500, 300 ),
         X Scale( 0.01, 4 ),
         Y Scale( 0, 4 ),
         XAxis( "Show Major Grid" ),
         YAxis( "Show Major Grid" ),
         X Function( lLow, x );
         X Function( lHigh, x );
         Text( {1, .5}, "Low= ", lLow, " High= ", lHigh );
      ),
      V List Box(
         Text Box( "Low and High" ),
         sb = Range Slider Box( 0, 4, lLow, lHigh, gr << Reshow )
```

```
      )
   ),
   Button Box( "Reshow",
      lLow = 1;
      lHigh = 2;
      gr << Reshow;
      sb << Reshow;
   )
);
Show( gr );
gr << Reshow;
```

Figure 11.18 Example of Using Range Slider Box

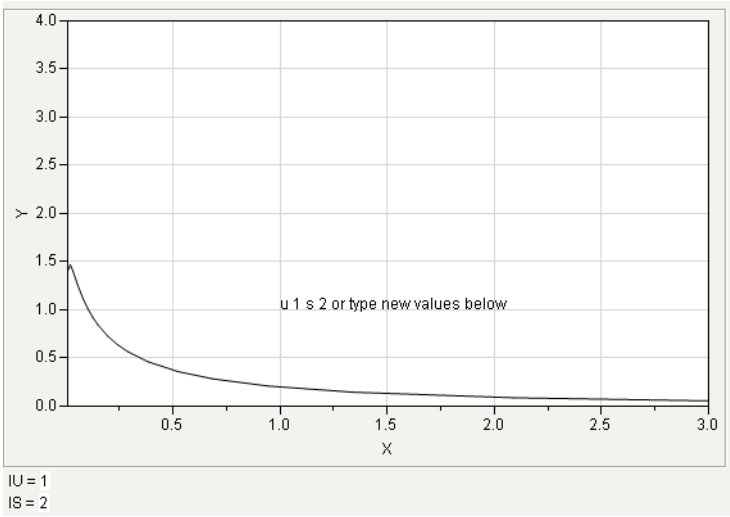


The script below is similar but uses `Global Box()` as an editable text box instead of two `Slider Box()` controls:

```
// Global LogNormal
lU = 1;
lS = 2;
win = New Window( "LogNormal Density",
   V List Box(
      gr = Graph Box( // define the arguments and script for the Graph Box
         Frame Size( 500, 300 ),
         X Scale( 0.01, 3 ),
         Y Scale( 0, 4 ),
         XAxis( "Show Major Grid" ),
         YAxis( "Show Major Grid" ),
         Y Function(
            Exp( -(Log( x ) - Log( lU )) ^ 2 / (2 * lS ^ 2) ) / (lS * x *
```

```
            Sqrt( 2 * Pi() )),
            x
         );

         // display the specified text at 1 on the x and y axes
         Text( {1, 1}, "u= ", lU, " s= ", lS, " or type new values below" );
      ),

      // display the lU variable value in the first H List Box
      H List Box( Global Box( lU ) ),

      // display the lS variable value in the second H List Box
      H List Box( Global Box( lS ) )
   )
);
```

Figure 11.19 Example of Using a Global Box Instead of Sliders



Example of a Calendar and Date Selector

You can create a calendar or date selector using `Calendar Box()` and `Number Edit Box()`.

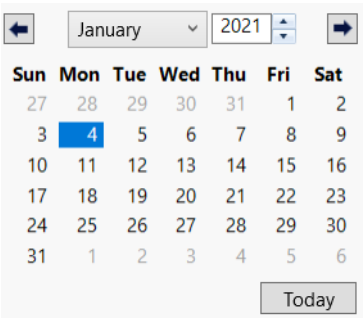
Calendar Box

With `Calendar Box()`, you can select dates from a calendar and input the year. The following example creates a calendar with October 5, 1989 initially selected. You can select any date 60 days before or 60 days after October 5, 1989.

```
win = New Window( "Calendar Box", cal = Calendar Box() );
date = Date MDY( 01, 04, 2021 );
```

```
cal << Date( date );
cal << Show Time( 0 ); // omit the time

/* earliest date that can be selected is 60 days before 01/04/2021
"start" truncates the value so the time is not included */
cal << Min Date( Date Increment(date, "Day", -60, "start" ) );

// latest date that can be selected is 60 days after 01/04/2021
cal << Max Date( Date Increment(date, "Day",  60, "start" ) );
// print the abbreviated date to the log
cal << Set Function( Function( {this}, Print( Abbrev Date(this << Get Date())
  ) ) );
```

Figure 11.20 Creating a Calendar



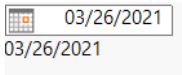
Number Edit Box

With `Number Edit Box()`, you can enter a date in the box or click the calendar button to select the date from a calendar.

The following example shows how to create an interactive calendar using `Number Edit Box()`. The user enters a date in the box or clicks the calendar to select a date from the calendar. The number appears in a text box below the number edit box. “01/04/2021” is initially selected.

```
f = Function( {this}, // callback
   textbox << Set Text( Format( this << Get, "m/d/y" ) ) // get the date
);
win = New Window( "Date Selector",
   numbox = Number Edit Box( 0, 20, << Set Function( f ) ),
   textbox = Text Box( "" ) // specified date
);
numbox << Set Format(
   Format( "m/d/y", 12 ) // date format and number of characters displayed
   );
numbox << Set( Date MDY( 01, 04, 2021 ) ); // default date
```

Figure 11.21 Creating a Date Selector



Note that a callback is assigned to the `f` variable. In the callback, `this` refers to the number edit box, the box that calls the function later. Referring to `this` prevents the need for using the `numbox` variable within the callback. The `Get` message is sent to the `this` argument to get the specified date.

Example of Constructing a Summary Results Report

The following example uses the `Summarize()` function to collect summary statistics on the Height column of Big Class.jmp and then uses display box constructors to show the results in a nicely formatted window.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Summarize(
   a = by( age ),
   c = count,
   sumHt = Sum( Height ),
   meanHt = Mean( Height ),
   minHt = Min( Height ),
   maxHt = Max( Height )
);
win = New Window( "Summary Results",
   Table Box(
      String Col Box( "Age", a ),
      Number Col Box( "Count", c ),
      Number Col Box( "Sum", sumHt ),
      Number Col Box( "Mean", meanHt ),
      Number Col Box( "Min", minHt ),
      Number Col Box( "Max", maxHt )
   )
);
```

This produces the summary results shown in Figure 11.22.

Figure 11.22 Producing a Customized Summary Report

Age	Count	Sum	Mean	Min	Max
12	8	465	58.125	51	66
13	7	422	60.2857	56	65
14	12	770	64.1667	61	69
15	7	452	64.5714	62	67
16	3	193	64.3333	60	68
17	3	200	66.6667	62	70

The following script saves the report as a Microsoft PowerPoint presentation.

```
Show Properties( win );
win << Save Presentation( "$DOCUMENTS/Summary.pptx" );
```

Interact with a Closing Window

A user can close a window by clicking the OK or Cancel button or by clicking the red Close button in the top corner of the window. A message can also be sent to close the window. `On Validate` and `On Close` expressions determine what happens you interact with the window.

- `On Validate` evaluates its expression *when the OK button is clicked*. The expression is not evaluated for any other button click or action. If the expression evaluates as true, the window closes. If the expression evaluates as false, the window is not closed. The script for an OK button runs when you click the button, no matter what the `On Validate` expression does.
- `On Close` evaluates its expression *right before the window closes*. The `On Close` return value is like the `On Validate` return value and can prevent the window from closing. However, the best practice for `On Close` is to always return 1 so that you do not accidentally create a window that cannot be closed.

  If the window closes, `On Close` is evaluated. `On Close` is not evaluated if the `On Validate` expression returns false, because the window is not closed.

Tips:

- `On Close` works for modal or non-modal windows. `On Validate` works only for modal windows.
- Do not use `On Close` for validation because it does not distinguish between OK and Cancel. Use `On Close` to clean up resources (such as invisible data tables) because it always runs when the window closes.
- Do not use `On Validate` to clean up resources. Use `On Close` instead or in addition to `On Validate`, because `On Validate` does not run for the Cancel button or red Close button.

On Close() Example

Suppose that a non-modal window is a report window, and the resource that needs cleaning up is an invisible data table. The `On Close` script closes the data table. This is good because the user might not know that the data table is open and has no easy way to close it.

```
dt = New Table( "Untitled", << Invisible,
    New Column( "x", Set Values( [1, 2, 3, 4] ) ),
    New Column( "y", Set Values( [4, 2, 3, 1] ) ) )
);

dt << Bivariate(
    y( dt:y ), x( dt:x ), <<On Close( Close( dt, "nosave" ) )
```

```
);
```

Before the script is run, no data tables are open. The Untitled invisible table is created and displayed in the Home Window and in the Data Tables list above the script. Then the graph is created. After you close the report window, the invisible table is closed by `On Close`.

On Validate() Example

Suppose that a modal window contains a question that the user must answer before closing the window. The following example shows how to validate the input when the user closes the window.

```
win = New Window( "Validate Example",
   <<Modal,
   <<Return Result,
   <<On Validate(
      If( (!Is Missing( variablebox << Get ) & 40 <= variablebox << Get <= 50),
         ( 1 ), // else the user did not enter the correct answer
         tb << Set Text(
            Match( Random Integer( 1, 3 ),
               1, "Are you sure?",
               2, "Guess again.",
               3, "Please read the question."
            )
         );
         ( 0 );
      )
   ),
   Text Box( "I'm thinking of a number between 40 and 50. What's your guess?"
   ),
   variablebox = Number Edit Box( . ),
   H List Box( Button Box( "OK" ), Button Box( "Cancel" ) ),
   tb = Text Box( "" )
);

New Window( "Modal Window",
   <<Modal,
   Text Box(
      If(

         // user clicked Cancel or the red Close button
         win["Button"] == -1, "Try again when you feel better.",
         win["variablebox"] == 42, "Good job.", // user entered 42
         "Take the test again." // user did not enter 42 and clicked OK
      )
   )
);
```

Update an Existing Display

Sometimes, you do not know how many display boxes will appear in a future report. For example, you might write a generic script that analyzes and reports on one or more variables. You do not know how many display boxes are needed, because the number of variables can change from one run of the script to the next.

The following sections describe how to add and delete display boxes from a report using `Append()`, `Prepend()`, `Delete()`, and `Sib Append()`.

Append

Use the `Append` message to add a display box to the bottom of an existing display. In the script, construct a single, empty box, then append boxes to it for each variable in the analysis. You can also use the `Append` message to modify an existing display with custom information or organization.

The following code example assumes that there is a list of effect names in the variable `effectsList`, and that each one corresponds to a column in a matrix `varprop`. In other words, `effectsList[1]` is the label for `varprop[0,1]`; `effectsList[2]` is the label for `varprop[0,2]`; and so on.

```
varprop = [0 1 2, 3 4 5, 6 7 8];
effectsList = {"one", "two", "three"};
```

First, an empty `Outline Box` containing an `H List Box` is created. The interior empty container is given the name `hb`:

```
win = New Window( "H List Box Example",
   Outline Box( "Variance Proportions", hb = H List Box() )
);
```

Then, a `for` loop steps through the `effectsList` and adds a `Number Col Box` for each element of `effectsList`:

```
For( i = 1, i <= N Items( effectsList ), i++,
   Eval(
      Substitute(
            Expr(
               hb << Append(
                  Number Col Box( effectslist[i], varprop[0, i] )
               )
            ),
         Expr( i ), i
      )
   )
);
```

`Sib Append()` is similar to `Append()` but makes the display box a sibling of the last child display box. For an example that compares `Append()` to `Sib Append()`, see Figure 11.24.

Prepend

The `Prepend` message works just like `Append` but adds the item at the beginning of the display box rather than at the end. If the display box is one of several that do not allow appending, then `Prepend` delegates the command to a child display box that can accept the command. It is fine to apply it to the top of the tree.

For example, the following example creates a Bivariate report with a button box at the top.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( height ), X( weight ), Fit Line );

/* select the first outline box in the biv report layer
send the Prepend message to the first outline box */
(biv << Report)(Outline Box( 1 )) << Prepend(

     // prepend a button box that fits a quadratic curve
     Button Box( "Click here for curve", biv << Fit Polynomial( 2 ) )
);
```

Click the **Click here for curve** button to add a quadratic curve to the graph.

You can accomplish the same thing by appending the button to the top of the tree:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( height ), X( weight ), Fit Line );
biv << Report << Prepend( // send the prepend message to the report

  // prepend the button box to the top of the report
  Button Box( "Click here for curve", biv << Fit Polynomial( 2 ) )
);
```

Delete

The `Delete` message removes the specified display box and all its children from the report. This is useful with the `Append` and `Prepend` messages for building completely dynamic displays. In the example below, a text box is replaced with another text box. In this case, the script could have used `Set Text`, but many display boxes cannot change their content.

```
win = New Window( "X",
   list = V List Box(
      t1 = Text Box( "t1" ),
      t2 = Text Box( "t2" )
   )
);
```

```
t1 << Delete;
list << Append( t1 = Text Box( "t1new" ) );
```

Note: Avoid deleting display boxes from a JMP platform. Send the `Hide( 1 )` message or the `Visibility( "Hidden" )` message to the display box that you want to hide.

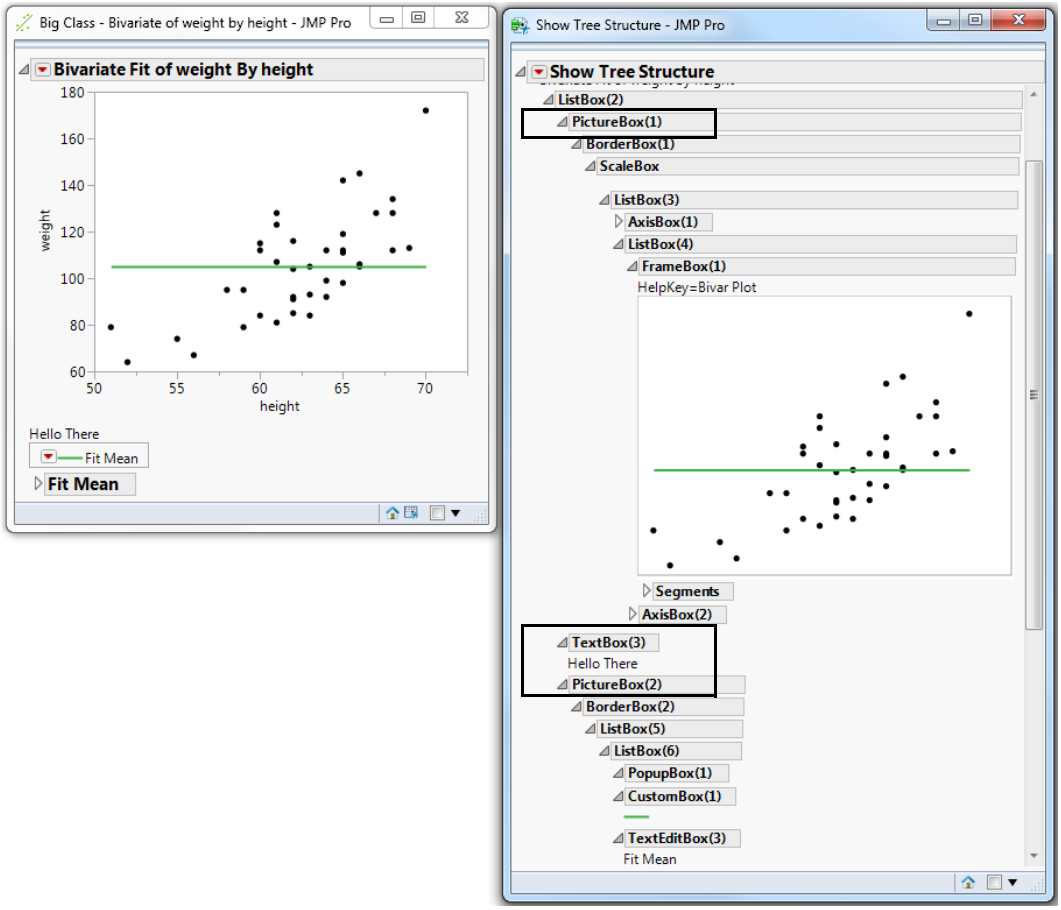
Sib Append

You can use the `Sib Append` message to add a display box immediately after an existing display box. Figure 11.23 shows two picture box trees. `Picture Box( 1 )` holds the Bivariate scatterplot. `Picture Box( 2 )` holds the Fit Mean menu, determined by seeing the green line and the Fit Mean text box.

Suppose you wanted to insert a text box in between these two boxes. You want to append a sibling to `Picture Box( 1 )`, so you send it the `Sib Append` message:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate(
   Y( :weight ),
   X( :height ),
   Fit Mean( {Line Color( {57, 177, 67} )} )
);

/* the report function takes the biv object and returns a report object.
 append the text box immediately after the first picture box */
Report( biv )[Picture Box( 1 )] << Sib Append( Text Box( "Hello There" ) );
```

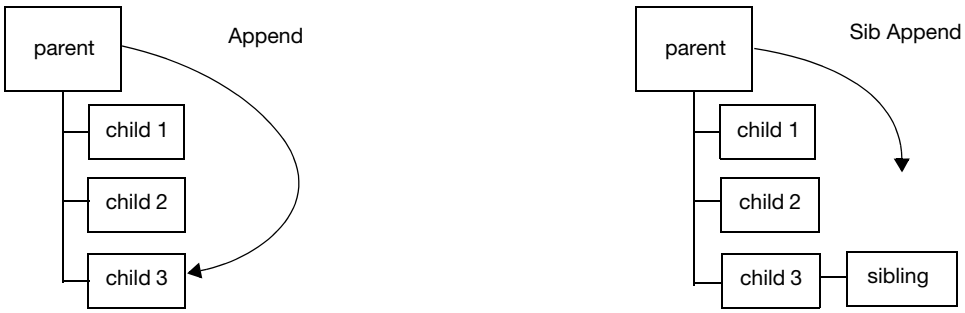
Figure 11.23 Appending a Sibling Text Box



Note: In Figure 11.23, the appended sibling text box appears below the first picture box in the Show Tree Structure View window. In the classic Show Tree Structure View window, siblings appeared next to each other in the display tree. See "View the Display Tree".

Figure 11.24 shows the difference between and `Append()` and `Sib Append()`. `Append()` makes the display box a *child* of the parent display box. `Sib Append()` makes the display box a *sibling* of the last display box.

Figure 11.24 Append (Left) and Sib Append (Right)



Note: If the parent display box is not a `V List Box()`, `Sib Append()` wraps that display box in `V List Box()`.

Set Function and Set Script

You can use the `Set Script` message to have a display box control (for example, a button box or combo box) run a script when it is clicked with the mouse.

```
win = New Window( "Set Script Example",
   ex = Button Box( "Press me" )
);
ex << Set Script( Print( "Pressed" ) );
```

The preceding script prints the following text to the log when the button box is clicked:

```
"Pressed"
```

Alternatively, you can use the `Set Function` message to have a display box control run a specific function where the first argument is the specific display box. `Set Function` enables you to build more object-oriented scripts and to create a larger system.

```
win = New Window( "Set Function Example",
   Button Box( "Press me",

      // functions specified with Set Function get 'this' display box
      << Set Function(
         Function(
            {this},
            this << Set Button Name( "Thanks" )
         )
      )
   )
);
```

The preceding script creates a button box with the name “Press me”. When the button is clicked, the function called by `Set Function` changes the name to “Thanks”.

If you have multiple buttons that can be serviced by the same script, `Set Function` is much simpler than `Set Script` because it tells you which button was pressed. In this example, both check boxes use the script that begins on the first line.

```
f = Function( {this, idx}, // idx just changed
   Write(

   /* <<Get Items returns a list of all items; indexing that list with idx
   is the name of the item that is changing */
      "\!n changing item=" || (this << Get Items())[idx] || " to " || Char(

   // <<Get returns the new value of that item
   this << Get( idx ) ) ||
      "\!n new selection=" || Char( this << Get Selected ) || "\!n"
   )// <<Get Selected returns a list of the currently checked (selected) items
);
New Window( "Example",
   H List Box( // 'f' is the 'named' function, used twice
      V List Box( Text Box( "Column 1" ), Check Box( {"a", "b"}, <<Set
   Function( f ) ) ),
      Spacer Box( size( 50, 50 ) ),
      V List Box( Text Box( "Column 2" ), Check Box( {"c", "d"}, <<Set
   Function( f ) ) )
   )
);
```

In the following example of `Set Function`, the same function is used to create the buttons.

```
New Window( "Cash Box",
   V List Box(
      H Center Box( TB = Text Box() ), // TB will show total of coins
      LB = Lineup Box( N Col( 3 ) ), // LB will hold buttons, added below
      H Center Box( // CLEAR button will reset the total
         Button Box( "CLEAR",
            total = 0;
            TB << Set Text( Char( total ) );
         )
      )
   )
);
coins = {1, 5, 10, 25, 50, 100}; // coins is a list of the button labels
total = 0;
```

```
/* loop creates the buttons and shows that the same function
is used for each */
For( iButton = 1, iButton <= N Items( coins ), iButton++,
   LB << Append( Button Box( Char( coins[iButton] ), Set Function(
   buttonFunction ) ) )
);
buttonFunction = Function( {this},
   total = total + Num( this << Get Button Name );
   TB << Set Text( Char( total ) );
);
```

The function called for any button, except CLEAR, uses the button's name to determine what to do. Rather than using the button's name, you might want to use a sibling display box by `this<<sib` or perhaps `(this<<parent)<<child` to find the first sibling of this button.

Notes:

- The `Set Script` and `Set Function` messages work for button boxes, calendar boxes, check boxes, combo boxes, list boxes, popup boxes, radio boxes, range slider boxes, slider boxes, and spin boxes.
- You cannot use both `Set Script` and `Set Function` at the same time. Use `Set Function` if you need to reference a specific display box object.

Get and Set Selected Values of Display Elements That Return Lists

You can use the `Set Selected(Item Number, <State>, <Run Script(0|1)>)` message to pre-select an item. You can use this message separately to a saved display box reference, or you can specify it inline as a list box << message.

To retrieve the selected value, use the `Get Selected` message, which returns the value of the selected item. The `Get Selected Indices` message returns the index number of the selected item.

```
antennaList = {"Dish","Helical","Polarizing","Radiant Array"};

win = New Window( "Test List", // method 1: display box reference
   listObj = List Box(
      antennaList,
      Print(
         "iList",
         listObj << Get Selected,
         listObj << Get Selected Indices
      )
   )
);
listObj << Set Selected( 2, 1 );
```

```
win = New Window( "Test List", // method 2: inline
   listObj = List Box(
      antennaList,
      <<Set Selected( 2, 1 ),
      Print(
         "iList",
         listObj << Get Selected,
         listObj << Get Selected Indices
      )
   )
);
```

Both of these scripts print the following text to the log:

```
"iList"
{"Helical"}
{2}
```

In the preceding examples, the `Print` expression is executed when the `Set Selected` message is completed. To prevent the script from running, include `Run Script(0)` as the last argument. `Run Script( 0|1 )` controls whether a display box on-change script runs after a `Set` or `Set Selected` message.

```
antennaList = {"Dish","Helical","Polarizing","Radiant Array"};
win = New Window( "Test List",
   listObj = List Box(
      antennaList,
      Print(
         "iList",
         listObj << Get Selected,
         listObj << Get Selected Indices
      )
   )
);
listObj << Set Selected( 2, 1, Run Script( 0 ) );
```

With `Run Script( 1 )`, the script is executed when the `Set` message is completed, even if the value is unchanged. (The script does not run again if the user selects the same value.) With `Run Script( 0 )`, the script does not run.

On most interactive display boxes, the script does not run on most interactive display boxes if you leave out `Run Script()`. However, on `List Box()`, the script runs by default, which is consistent with previous behavior.

Send Messages to Constructed Displays

When you assign a constructed display to a name, that name becomes a reference to the window, which in turn owns the display boxes inside it. Using subscripts, you can then send messages to the display boxes inside the window.

For example, the following script creates an interactive sine wave graph. The script automates the interaction by sending messages to the frame box inside the window. Note the frame box assignment to *tf* in the middle of the script).

```
amplitude = 1;
freq = 1;
phase = 0;
win = New Window( "Wiggle Wave",
   Graph Box(
      Frame Size( 500, 300 ),
      X Scale( -5, 5 ),
      Y Scale( -5, 5 ),
      Y Function( amplitude * Sine( x / freq + phase ), x );
      Handle(

         /* current values of phase and amplitude position
         the square handle.
         script begins after the last comma after amplitude */
         phase,
         amplitude,

         /* x and y have been set to the handle's position
         but the handle doesn't move unless the values specified
         in the first two arguments are updated */
         phase = x;
         amplitude = y;
      );

      // Handle works similarly except that the y value is .5
      Handle( freq, .5, freq = x );

      Text( // display a text string in the graph
         {3, 4},
         "amplitude: ",
         Round( amplitude, 4 ), // display the current value
         {3, 3.5},
         "frequency: ",
         Round( freq, 4 ), // display the current value
         {3, 3},
         "phase: ",
         Round( phase, 4 ) // display the current value
```

```
      );
   )
);
tf = win[Frame Box( 1 )]; // get the frame box (the graph)
For( amplitude = -4, amplitude < 4, amplitude += .1,// animate the amplitude
   tf << Reshow // force the graph to update
);
amplitude = 1; // use for loops for more complex movement
freq = 1;
phase = 0;
For( i = 0, i < 1000, i++,
   amplitude += (Random Uniform() - .5);
   amplitude = If(
      amplitude > 4, 4,
      amplitude < -4, -4,
      amplitude
   );
   freq += (Random Uniform() - .5) / 20;
   phase += (Random Uniform() - .5) / 10;
   tf << Reshow;
   Wait( .05 );
);
```

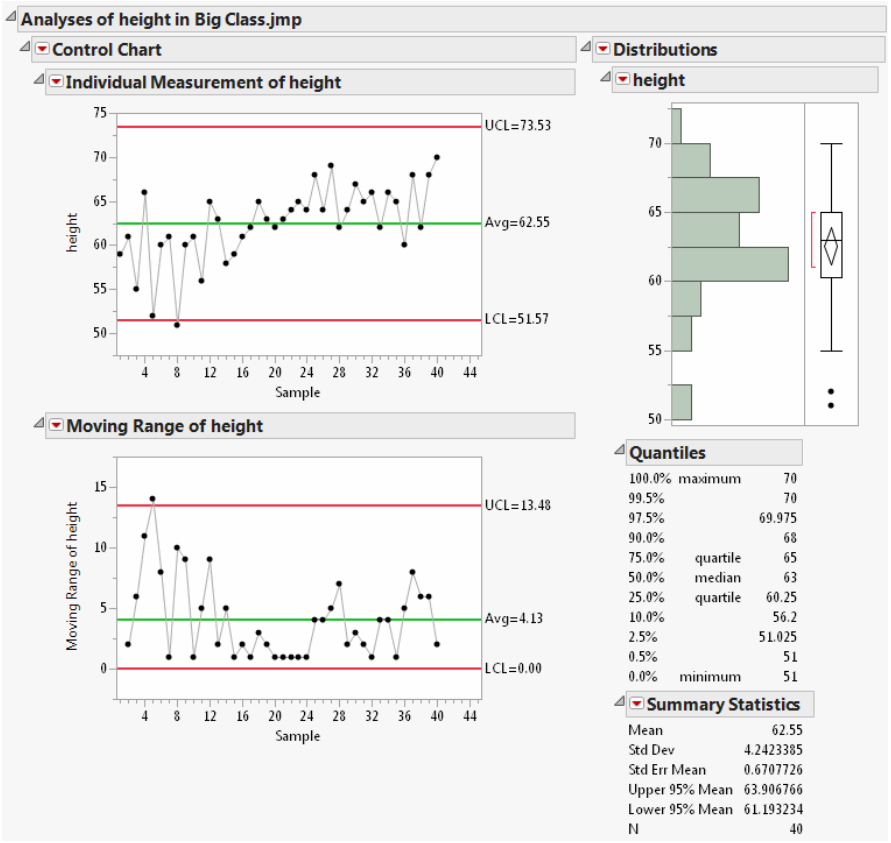
Construct Display Boxes That Contain Platforms

Another type of display that you might want to construct is your own combination of results from the analysis platforms in JMP. Write the platform script inside a display box and assemble the display boxes into a window. For ease in routing messages to it later, assign the whole script to a reference.

The following example creates several graphs and reports in one window:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
csp = New Window( "Platform Example",
   Outline Box( "Analyses of height in Big Class.jmp",
      H List Box(
         cc = Control Chart(
            Chart Col( Height, "Individual
            Measurement", "Moving Range" ),
            K Sigma( 3 )
         ),
         dist = Distribution( Columns( Height ) )
      )
   )
);
```

Figure 11.25 Example: Multiple Graphs in One Report Window



Now you can work with the window by sending messages to the reference *csp*. This is a display box reference, whose capabilities are similar to those of a `Report` for a platform. You can use multiple-argument subscripting to locate specific items within the outline tree:

```
csp["Control ?", "moving range ?"] << Close;
csp["Dist?", "quantiles"] << Close;
```

Notice that the preceding script not only assigned the whole window to a reference (`csp`) but also assigned the platform launch scripts to names (`cc` and `dist`) within their display boxes. This makes it easy to route messages to the platforms. You could, in turn, get the reports for these and have yet another way to manipulate display boxes. The following are equivalent messages that reopen the nodes:

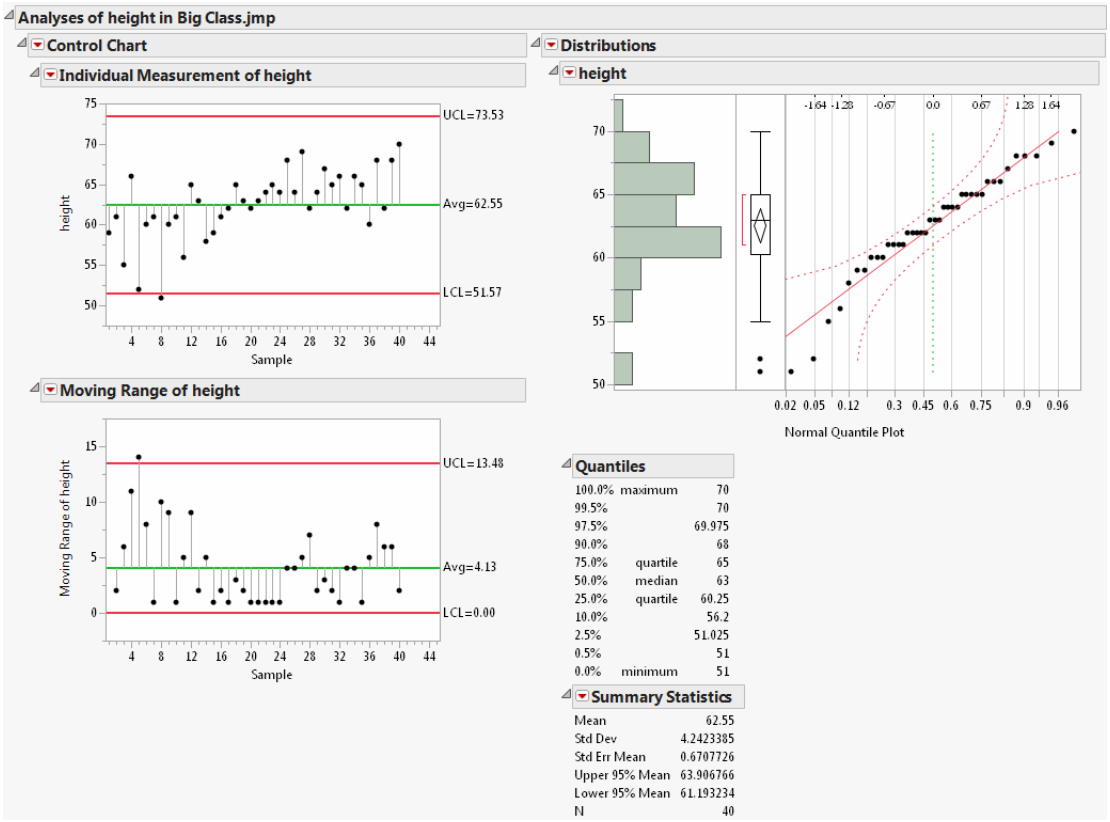
```
rcc = cc << Report;
rdist = dist << Report;
rcc["moving range ?"] << Close;
rdist["quantiles"] << Close;
```

You can send messages directly to the platform references themselves. Find out your options by searching for "Distribution" and "Control Chart" in the JMP Scripting Index. For example, Control Chart provides a message called `Needle` that creates a needle chart. Distribution provides a message called `Normal Quantile Plot`.

To execute these options from JSL, send them as messages to the platform references:

```
cc << Needle;
dist << Normal Quantile Plot;
```

Figure 11.26 Changing a Custom Report



Counting the Number of Child Boxes

Suppose that you want to count the number of child text boxes in a list box. The most reliable way is to use XPath to find a specific string in the text boxes and then use `N Items()` to count the number of text boxes.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = Bivariate(
```

```
    Y( :weight ),
    X( :height ),
    Group By( :age ),
    Fit Line( {Line Width( 3 )} )
);

/* start with the root list box, find a text edit box
that starts with “Linear Fit age==”, and count the number of children */
N Items( biv << XPath( "//ListBox[starts-with(TextEditBox,'Linear Fit
  age==')]" ) );
```

Examples of Creating a Dashboard from Two Reports

In JMP, you create dashboards in Dashboard Builder. See *Using JMP*. This section shows how to write scripts to create dashboards.

Dashboard with One Row of Reports

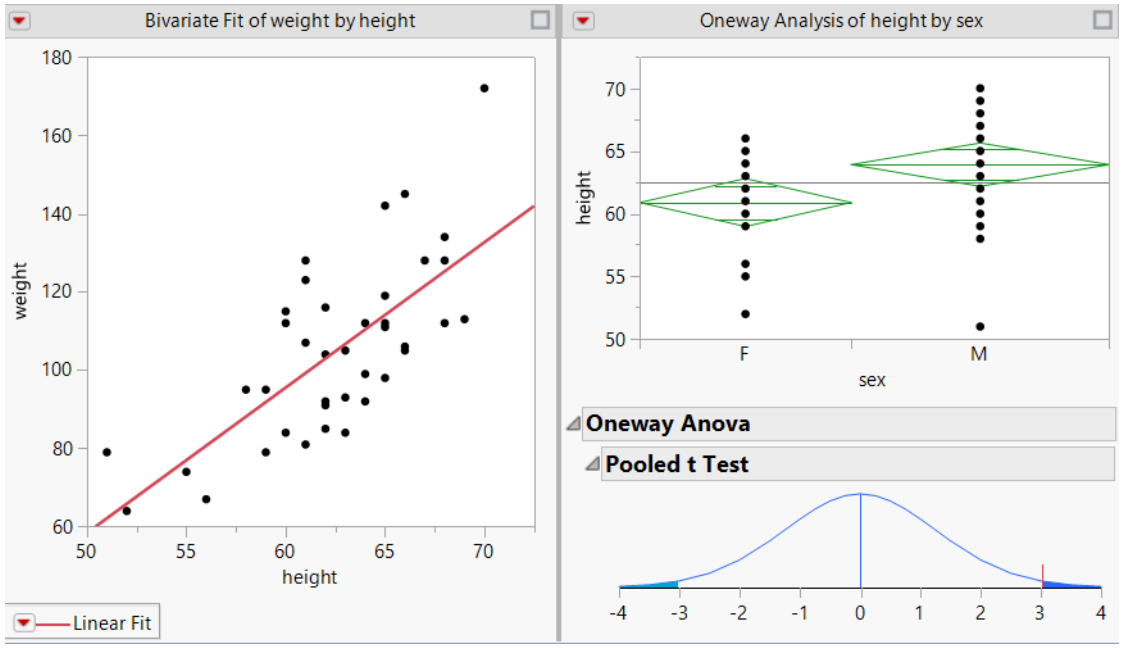
In the report window, reports appear in tab page boxes. You reposition reports in a dashboard by clicking the tab title and dragging the tab page to a highlighted drop zone. The highlighted drop zones indicate how to arrange reports horizontally, vertically, or in tabs. If a valid drop zone is not selected, the box returns to its original location.

This example shows how to write a script that creates a dashboard with two reports in a row:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
New Window( "My Dashboard",
  H Splitter Box( // place reports in a splitter box
  Size( 700, 400 ),
     Tab Page Box( // place report in a tab page box
        Title( "Bivariate Fit of weight by height" ),
        dt << Bivariate( // Bivariate report
           Y( :weight ),
           X( :height ),
           Fit Line,
           Report View( "Summary" ) // show only the graph
        ),
        <<Moveable( 1 ) // make the report moveable
     ),
     Tab Page Box(
        Title( "Oneway Analysis of height by sex" ),
        dt << Oneway( // Oneway report
           Y( :height ),
           X( :sex ),
           Means( 1 ),
           Mean Diamonds( 1 ),
```

```
            Report View( "Summary" )
         ),
         <<Moveable( 1 )
      ),
      /* each tab box can be dragged to a different location
      required for moveable display boxes */
      <<Dockable( 1 )
   )
);
```

Figure 11.27 Dashboard with Two Reports



Note: For dockable and moveable tabs that contain a single platform, the top-level outline box is subsumed into the tab title. This means that the red triangle options now appear in the title bar of the tabs.

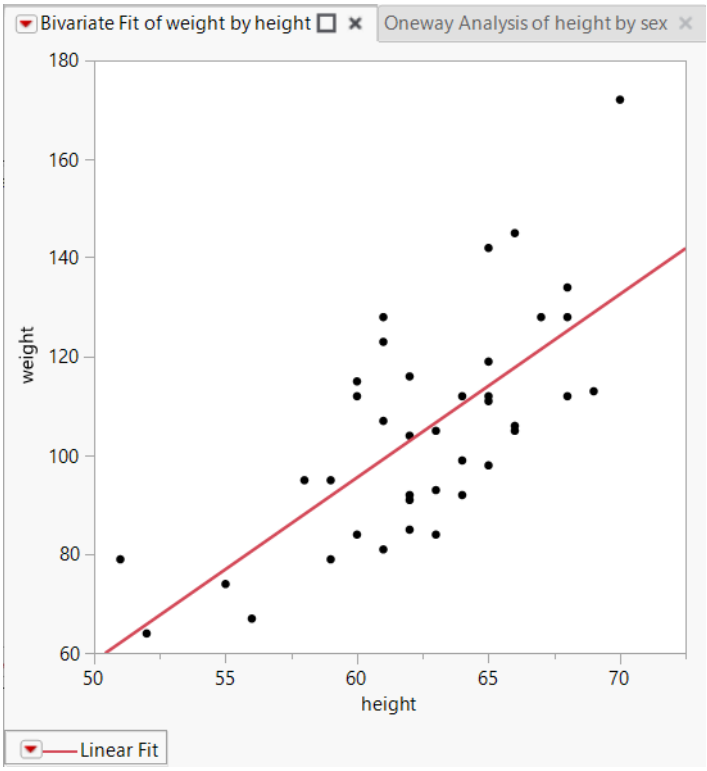
Dashboard with Tabs

The following script creates a dashboard that places each report on a tab:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
New Window( "My Dashboard",
   Tab Box( // place reports in a tab box
      Tab Page Box( // place report in a tab page box
         Title( "Bivariate Fit of weight by height" ),
```

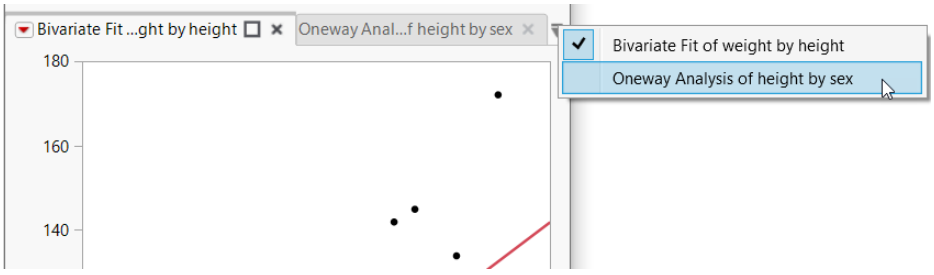
```
            dt << Bivariate( // Bivariate report
               Y( :weight ),
               X( :height ),
               Fit Line,
               Report View( "Summary" ) // show only the graph
            ),
            <<Moveable( 1 ), // make the report moveable
            <<Closeable( 1 ) // make the report closeable
         ),
         Tab Page Box(
            Title( "Oneway Analysis of height by sex" ),
            dt << Oneway( // Oneway report
               Y( :height ),
               X( :sex ),
               Means( 1 ),
               Mean Diamonds( 1 ),
               Report View( "Summary" )
            ),
            <<Moveable( 1 ),
            <<Closeable( 1 )
         ),
         /* each tab box can be dragged to a different location
         required for moveable display boxes */
         <<Dockable( 1 ),
         <<Set Overflow Enabled( 0 ) // both tab pages appear
      )
   );
```

Figure 11.28 Dashboard with Tabs



In a dashboard with several tabs, the window widens to show all tabbed reports. You might want to show only the first report and then let the user select the other tabbed reports from a list. Change `Set Overflow Enabled( 0 )` near the end of the script to `Set Overflow Enabled( 1 )`. Figure 11.29 shows the result.

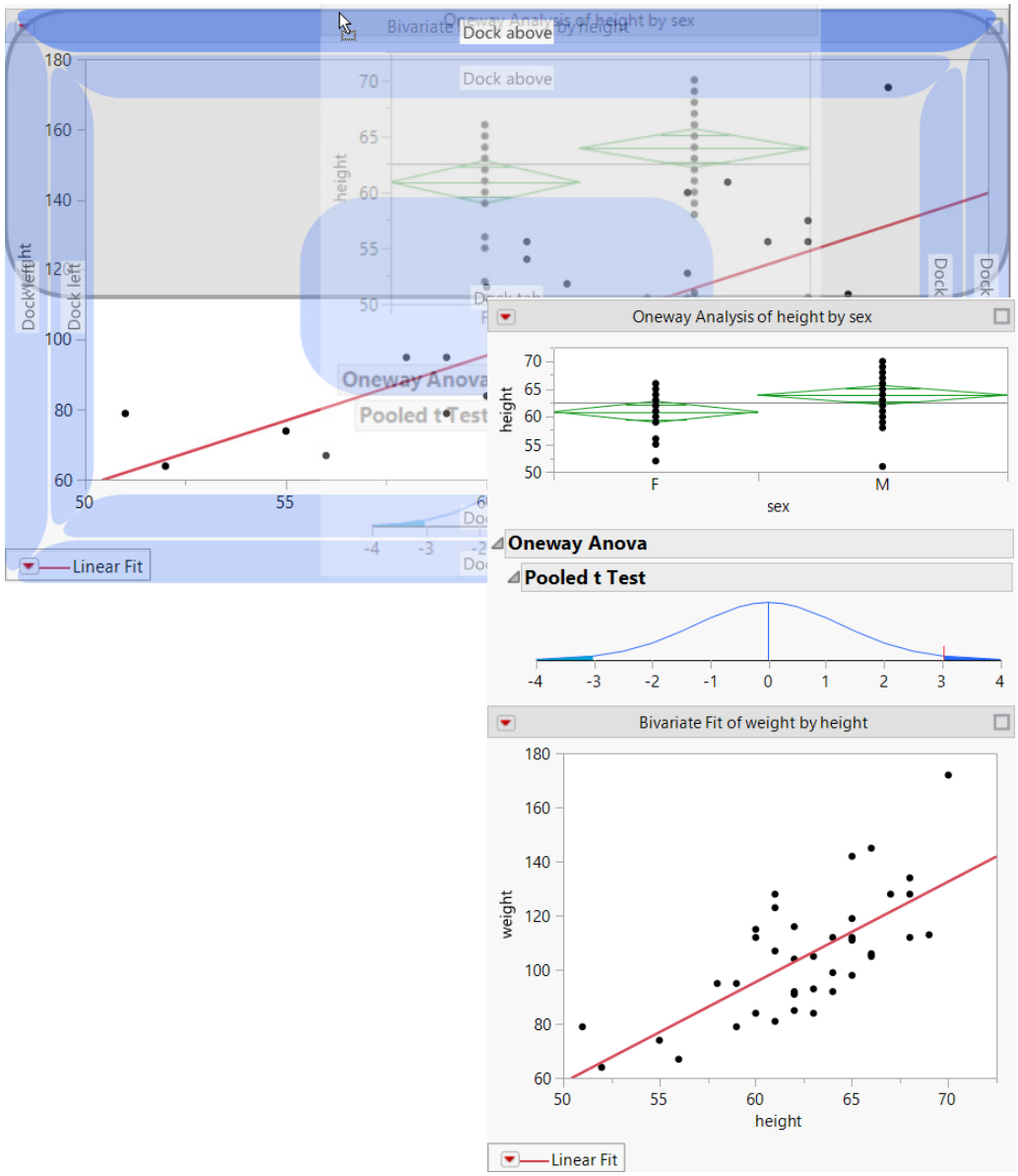
Figure 11.29 Dashboard with Overflow Tabs



How to Dock Tabs in Tab Boxes

A docking layout is a collection of `Tab Page Box()` objects and docking containers that enable reports to be rearranged. In the report window, you drag a docked tab to a highlighted drop zone to create a new layout. Figure 11.27 shows two docked reports. In Figure 11.30, the Oneway report was dragged to the highlighted drop zone to display the reports in one column.

Figure 11.30 Docking a Tab in a Dashboard



You combine `Tab Box()` with `Splitter Box()` to create docked tabs. Tab boxes and splitter boxes must directly contain other docking containers or the tab pages that define the content. Note the stacked boxes from a portion of the preceding example:

```
H Splitter Box( // place reports in a splitter box
  Size( 700, 400 ),
  Tab Page Box( // place report in a tab page box
```

The following expressions specify that the tab box and splitter box are docking containers, within which tab pages can be docked:

```
Tab Box << Set Dockable();
Splitter Box << Set Dockable();
```

In the report window, you reposition tab pages by clicking the tab title and dragging the tab page to a highlighted drop zone. The highlighted drop zones indicate how to arrange reports horizontally, vertically, or in tabs. If a valid drop zone is not selected, the box returns to its original location.

Closing Tabs

The following expression creates a close button to the right of the report title on a closeable tab page:

```
Tab Page Box << Set Closeable();
```

Clicking the close button deletes the tab and its contents. Parent containers might also be removed if they are no longer necessary (for example, if there are no more tabs in a tab box or only a single child in a splitter box).

Use the `Tab Page Box <<Set Close` message to run a JSL script when the close button is clicked. A return value of "1" indicates that the tab is closed. A return value of "0" indicates that closing the tab is not allowed at that time. Closing the tab might not be allowed as the result of a user prompt or based on the state of the program.

Controlling Wide Tabs

When a docking layout includes tab page boxes, you may place many pages in a single `Tab Box()` container. In this case, the titles may become wider than the actual content or wider than desired in the overall display. The `Tab Box <<Set Overflow Enabled( 0|1 )` message specifies that the tab titles should not cause the tab box to grow wider. If the titles do not all fit, a pop-up button appears in the upper right corner of the tab box to select tabs that are in the "overflow list". See Figure 11.29 for an example.

Example of Creating a Cluster Platform Launch Window

The following example creates a simplified replica of the Cluster platform launch window, which then launches the platform with the given arguments.

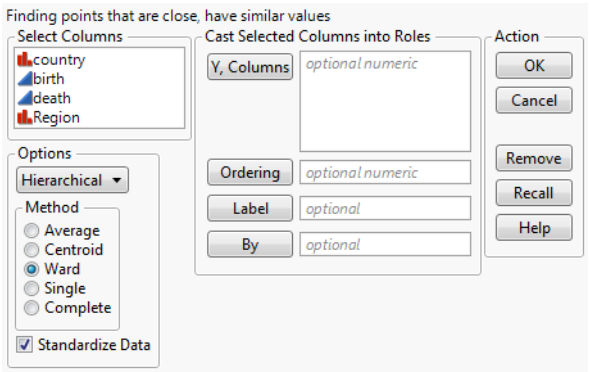
Note: Some functions (`Recall` and `Help`) are not implemented in the script, so an alert window is shown when the Recall and Help button are clicked. In addition, switching from Hierarchical to K-Means clustering does not change anything, unlike the real Cluster launch window.

```
dt = Open( "$SAMPLE_DATA/Birth Death.jmp" );
nc = N Col( dt ); // number of columns in the window
lbWidth = 130; // width of the window

// define the method list
methodList = {"Average", "Centroid", "Ward", "Single", "Complete"};
notImplemented = Expr(
   win = New Window( "Feature Not Implemented Yet", <<Modal, Button Box( "OK"
   ) )
);
clusterDlg = New Window( "Clustering", // create the window
   <<Modal,
   Border Box( Left( 3 ), Top( 2 ),
      V List Box(
         Text Box( "Finding points that are close, have similar values" ),
         H List Box(
            V List Box(
               Panel Box( "Select Columns",
                  colListData = Col List Box(
                     All,
                     width( lbWidth ),
                     NLines( Min( nc, 10 ) )
                  )
               ),
               Panel Box( "Options",
                  V List Box(
                     comboObj = Combo Box(
                        {"Hierarchical", "K-Means"},
                        <<Set( 1 )
                     ),
                     Panel Box( "Method",
                        methodObj = Radio Box( methodList, <<Set( 3 ) )
                     ),
                     checkObj = Check Box( {"Standardize Data"}, <<Set( 1, 1 ) )
                  )
               )
            ),
            Panel Box( "Cast Selected Columns into Roles",
               Line Up Box( N Col( 2 ), Spacing( 3 ),
                  Button Box( "Y, Columns",
                     colListY << Append( colListData << GetSelected )
                  ),
                  colListY = Col List Box(
                     width( lbWidth ),
                     NLines( 5 ),
                     "numeric"
```

```
            ),
            Button Box( "Ordering",
               colListO << Append( colListData << GetSelected )
            ),
            colListO = Col List Box(
               width( lbWidth ),
               NLines( 1 ),
               "numeric"
            ),
            Button Box( "Label",
               colListL << Append( colListData << GetSelected )
            ),
            colListL = Col List Box( width( lbWidth ), NLines( 1 ) ),
            Button Box( "By",
               colListB << Append( colListData << GetSelected )
            ),
            colListB = Col List Box( width( lbWidth ), NLines( 1 ) )
         )
      ),
      Panel Box( "Action",
         Line Up Box( N Col( 1 ),
            Button Box( "OK",
               If( (comboObj << Get) == 1,
                  Hierarchical Cluster(
                     Y( Eval( colListY << GetItems ) ),
                     Order( Eval( colListO << GetItems ) ),
                     Label( Eval( colListL << GetItems ) ),
                     By( Eval( colListB << GetItems ) ),
                     Method( methodList[methodObj << Get]
                     ),
                     Standardize( checkObj << Get( 1 ) )
                  ),
                  KMeansCluster( Y( colListY << GetItems ) )
               );
               clusterDlg << Close Window;
            ),
            Button Box( "Cancel", clusterDlg << Close Window ),
            Text Box( " " ),
            Button Box( "Remove",
               colListY << RemoveSelected;
               colListO << RemoveSelected;
               colListL << RemoveSelected;
               colListB << RemoveSelected;
            ),
            Button Box( "Recall", notImplemented ),
            Button Box( "Help", notImplemented )
```

```
            )
         )
      )
   )
)
);
```

Figure 11.31 The Cluster Launch Window



Example of Constructing a Custom Platform

An example in "Manipulating Expressions", showed how to use the `Substitute Into()` function to input coefficients for a quadratic polynomial into the quadratic formula and then use the formula to calculate the roots of the polynomial. That example required specifying the coefficients as arguments to `Substitute Into()`.

The section "Construct a Column Dialog" shows an example to collect coefficients from the user using a modal dialog box.

This section further develops the example into a complete customized platform that first displays a dialog box to ask for coefficients, finds the roots, and then displays the results along with a graph in a custom window.

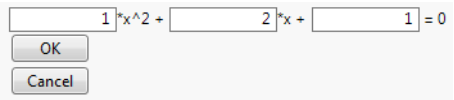
```
// create a window to collect coefficients from the user
myCoeffs = New Window( "Find the roots for the equation",
   <<Modal,
   H List Box(
      a = Number Edit Box( 1 ),
      Text Box( "*x^2 + " ),
      b = Number Edit Box( 2 ),
      Text Box( "*x + " ),
      c = Number Edit Box( 1 ),
      Text Box( " = 0" )
   ),
```

```
      Button Box( "OK",
         a = a << Get;
         b = b << Get;
         c = c << Get;
         Show( a, b, c );
      ),
      Button Box( "Cancel" )
   );

   /* calculate the results: The quadratic formula is
   x=(-b + - sqrt(b^2 - 4ac))/2a. Plug the coefficients into
   the quadratic formula */
   x = {Expr(
      (-b + Sqrt( b ^ 2 - 4 * a * c )) / (2 * a)
   ), Expr(
      (-b - Sqrt( b ^ 2 - 4 * a * c )) / (2 * a)
   )};
   xx = Eval Expr( x );


   results = Expr( // store the solution list
   /* test whether real roots were found and make an appropriate display
   if yes (for example, with the window's defaults), show roots and a graph */
      xmin = xx[1] - 5;
      xmax = xx[2] + 5;
      ymin = -20;
      ymax = 20;
      win = New Window( "The roots of a quadratic function",
         V List Box(
            Text Box( "The real roots for the equation " ),
            Text Box( "    " || Expr( po ) || " = 0" ),
            H List Box( Text Box( "are x=" ), Text Box( xxx ) ),
            Text Box( " " ), // to get a blank line
            Graph Box(
               Frame Size( 200, 200 ),
               X Scale( xmin, xmax ),
               Y Scale( ymin, ymax ),
               Line Style( 2 ),
               H Line( 0 ),
               Line Style( 0 ),
               Y Function( polynomial, x ),
               Line Style( 3 ),
               Pen Color( 3 ),
               V Line( xx[1] ),
               V Line( xx[2] ),
               Marker Size( 2 ),
```

```
            Marker( 0, {xx[1], 0}, {xx[2], 0} )
         )
      )
   );
);
error = Expr(

/* if no (for example, with a=3, b=4, c=5), put up an error
window with a helpful graph */
   win = New Window( "Error",
      V List Box(
         Text Box( " " ),
         Text Box( "  Polynomial " || po || " has no real roots.  " ),
         Text Box( " " ),
         Text Box( "Examine a graph of the function to get an idea why." ),
         Graph Box(
            Frame Size( 200, 200 ),
            X Scale( -20, 20 ),
            Y Scale( -20, 20 ),
            Line Style( 2 ),
            H Line( 0 ),
            Line Style( 0 ),
            Y Function( polynomial, x )
         )
      )
   )
);

/* either way, the script needs to have some strings ready
rewrite the polynomial with the coefficients specified */
polynomial = Expr( a * x ^ 2 + b * x + c );

// store this instance of the polynomial as a string
po = Char( Eval Expr( polynomial ) );

// store the solution list as a string
xxx = Char( Eval Expr( x ) );

If( Is Missing( xx[1] ) | Is Missing( xx[2] ), // now it's ready for a test
   error,
   results
);
```

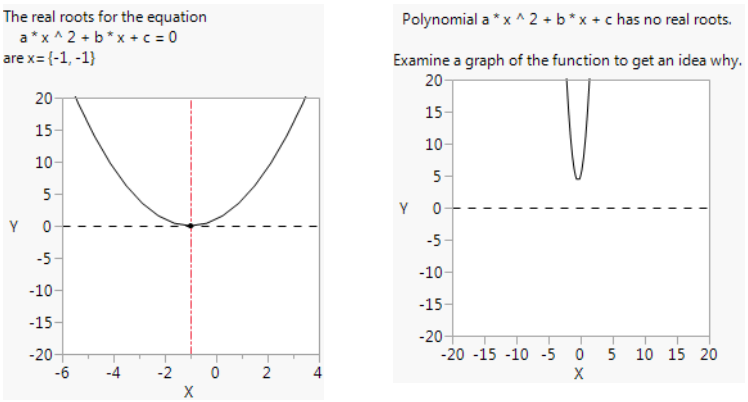
When you run the preceding script, you first see a window like this:

Figure 11.32 A Custom Platform Launch Window



Clicking **OK**, displays a results window (Figure 11.33, left), with either the roots or an error message. Rerun the script, input 5, 4, and 5 respectively and click **OK**. Note that JMP displays an error message (Figure 11.33, right).

Figure 11.33 The Custom Platform's Report

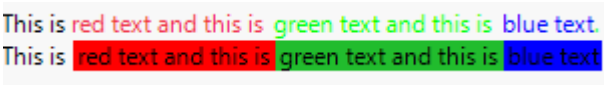


Example of Specifying Text and Background Colors

The `markup` argument in a display box enables you to write HTML code to specify text and background colors. The color can be a hex value (starting with "0x"or "#") or a known JMP color. Here is an example:

```
New Window( "Colors",
   Text Box(
      "This is <font color=\!"Red\!">red text and this is <font
   color=\!"0x00ff00\!">green text and this is <font color=\!"#0000FF\!">blue
   text</font>.",
      <<markup
   ),
   Text Box(
      "This is <background color=\!"#ff0000\!"> red text and this
   is<background color=\!"Green\!"> green text and this is<background
   color=\!"0x0000FF\!"> blue text</background>",
      <<markup
   )
);
```

Figure 11.34 HTML Text and Background Colors



Modal Windows

As described in "Constructors for New Windows", *modal* means that the user must interact with the window immediately. Clicking outside the window produces an error sound. Script execution stops until the user responds to the window.

JMP provides two types of modal window functions:

- `New Window()` with the `Modal` message. This creates a new window in which you arrange other display boxes.
- `Column Dialog()` constructs a window for column role assignments as found in many launch windows. The specialized constructors for `Column Dialog()` mean that you have less flexibility than with `New Window()`. However, `Column Dialog()` lets you easily create a launch window.

Construct a Modal Window

When you submit a script with a modal window, JMP draws the window, waits for the user to make choices and click **OK**, and then stores a list of the variables with their values. Any attempt to click outside the window produces an error sound, and script execution is suspended until the user clicks **OK** or **Cancel**.

The `Column Dialog()` function is specifically intended to prompt users to choose columns from the current (topmost) data table. Windows created by `Column Dialog()` are also modal windows.

Tips:

- Put all the modal windows near the beginning of the script, if possible. This way, all the user interaction can be accomplished at once, and then users can leave JMP to finish its work unattended.
- Make sure your modal windows give the user enough information. Do not just present a number field. Tell users how the number is used. If there are limits for valid responses, say so.

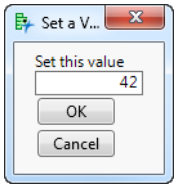
A very simple modal window might request a value for one variable:

```
win = New Window( "Set a Value",
   <<Modal,
```

```
    Text Box( "Set this value" ),
    variablebox = Number Edit Box( 42 ),
    Button Box( "OK" ),
    Button Box( "Cancel" )
);
```

Notice that the argument to `Number Edit Box()` is the default value, 42.

Figure 11.35 Sample Modal Display Box



If you click **OK**, the window closes and {`Button(1)`} is returned. To get the value entered in the box, use the `Return Result` message and a subscript into the window. Note that the deprecated `Dialog()` returns the list of variable assignments the same way.

```
win = New Window( "Set a Value",
    <<Modal,
    <<Return Result,
    Text Box( "Set this value" ),
    variablebox = Number Edit Box( 42 ),
    Button Box( "OK" ),
    Button Box( "Cancel" )
);

// create a subscript to the variablebox variable
Write( win["variablebox"] );
```

In this example, the user typed 33 in the number edit box. That value is stored in `win`.

If you click **Cancel**, the window closes, it returns {`Button(-1)`}, and the script continues.

To detect whether the user clicked Cancel, add the following expression:

```
If( win["Button"] == 1,
    Print( win["variablebox"] );
    ,
    Print("Canceled")
);
```

Note that a modal window must have at least one button. If you omit the `Button Box()` from your script, an OK button is automatically included. To omit the OK button, set the visibility to collapse.

```
win = New Window( "No OK Button",
```

```
    <<Modal,
    b = Button Box( "Close Window", b << Close Window ),
    bb = Button Box( "OK", <<Visibility( "collapse" ) )
);
```

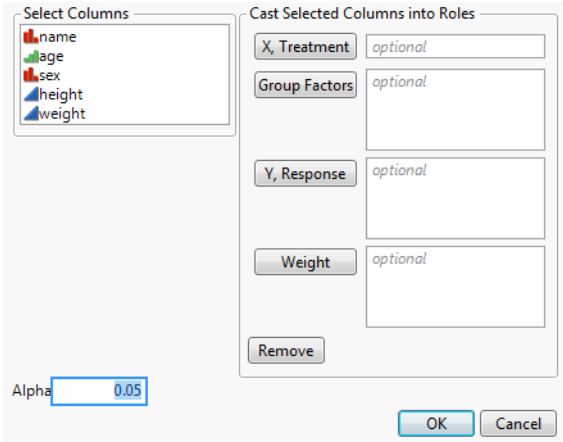
Note: Modal windows require at least one button for dismissing the window. You can use **OK**, **Yes**, **No**, or **Cancel** to label modal window buttons. If you do not include at least one of these buttons in a modal window, JMP adds an **OK** button.

Construct a Column Dialog

`Column Dialog()`, a variant of the deprecated `Dialog()` function, lets you prompt for column selections from the current data table or contextual data table. The `OK`, `Cancel`, and `Remove` buttons and the list of columns to choose from the data table are added automatically.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dlg = Column Dialog(
   Col ID = Col List( "X, Treatment", Max Col( 1 ) ),
   Group = Col List( "Group Factors" ),
   Split = Col List( "Y, Response" ),
   w = Col List( "Weight" ),
   H List( "Alpha", alpha = Edit Number( .05 ) )
);
```

Figure 11.36 Column Dialog



The following example returns a list similar to this one, depending on the user's choices:

```
{Col ID = {}, Group = {}, Split = {}, w = {}, alpha = 0.05, Button( -1 )}
```

For each destination list, a `Col List` clause must be a direct argument of `Column Dialog()` (not nested inside some other argument). An optional `MaxCol(`*`n`*`)` argument restricts the number of data columns that can be chosen to *n*. Lists are *always* returned, although they can sometimes be empty lists. You can include as many as twelve `Col List` clauses.

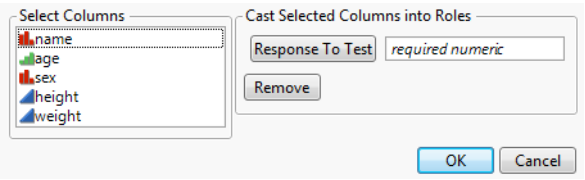
Other items permitted in the deprecated `Dialog()` are permitted in `Column Dialog()` and have the same functionality. You can specify the minimum and maximum number of columns that are allowed in a column dialog box with the `MinCol` and `MaxCol` arguments.

You can also specify the modeling type of the columns that are allowed to be selected (`Ordinal`, `Nominal` or `Continuous`). You can set the width of the list using `Select List Width(`*`pixels`*`)` argument. To set the width of the column list, use `Width(`*`pixels`*`)` inside the `Col List()` function.

The following example generates a column dialog box that only allows the selection of exactly one numeric column:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
rt_dlg = Column Dialog(
   cv = Col List( "Response To Test", MaxCol( 1 ), MinCol( 1 ), DataType(
   "Numeric" ) )
);
```

Figure 11.37 Restricting Selection of Columns



The `Data Type` choices are `Numeric`, `Character`, and `Any`.

In addition, use the `Columns` specification to pre-fill some column selections. For example, the following script assigns height to the X role and weight and age to the Y role:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dlg = Column Dialog(
   xCols = Col List( "X, Factors", Columns( :height ) ),
   yCols = Col List( "Y, Response", Columns( :weight, :age ) )
);
```

Comparison of Column Dialog and New Window

`New Window()` and `Column Dialog()` enable you to construct a launch window that has a column selector. Use `Column Dialog()` to construct a simple launch window. Several items are added automatically to column dialogs: the Remove and Cancel buttons, the list of columns, and an OK button (if no buttons are explicitly defined).

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dlg = Column Dialog(
   x = ColList( "X", Max Col( 1 ) ),
   y = ColList( "Y", Max Col( 1 ) )
);

// returns dlg = {x = {:weight}, y = {:height}, Button(1)};
Show( dlg );

If( dlg["Button"] == 1,
   xCol = dlg["x"];
   yCol = dlg["y"];
);

Show( xCol, yCol );
   xCol = {:height};
   yCol = {:weight};
```

`New Window()` is more flexible because of numerous optional arguments. In the following example, the user must select two columns before clicking the OK button. A Bivariate graph is then created. Validating the user's selections is not possible in `Column Dialog()`.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
xvar = .;
yvar = .;
win = New Window( "Example of Returning Values",
   <<Modal,

   // require the user to select two variables before clicking OK
   <<On Validate(
      Show( xvar, yvar );
         If( Is Missing( xvar ) | Is Missing( yvar ),

         // if xvar or yvar are missing do nothing when OK is clicked
         0,
         1
      );
   ),
   Text Box( " Select two numeric columns. " ),
   H List Box(
```

```
      Text Box( " X, Factor " ),
      x = Col List Box(
         dt, // data table reference
         all, // display all columns from the data table

         // get the name of the selected column before the window closes
         xvar = (x << Get Selected)[1];
         Show( xvar );
      ),
      Text Box( "Y, Response" ),
      y = Col List Box(
         dt,
         all,
         yvar = (y << Get Selected)[1];
         Show( yvar );
      )
   )
);
If (win["Button"] == 1, // if the user clicks OK...
xcol = Column( dt, xvar ); //  get the columns
ycol = Column( dt, yvar );
dt << Bivariate( Y( ycol ), X( xcol ) ); // create a Bivariate plot
);
```

Constructors for Column Dialogs

Constructors such as `Edit Number()` and `Check Box()` can be used in column dialogs, which construct launch windows with a column selector. The following example shows a typical column dialog with an `Edit Number()` box:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column Dialog(
   ex x = ColList( "X", Max Col( 1 ) ),
   HList( "Alpha",
   ex = Edit Number( .05 ) )
);
```

Table 11.3 describes the column dialog constructors.

Notes:

- A column dialog automatically includes the list of columns in the data table.
- A column dialog also automatically includes OK, Cancel, and Remove buttons.

Table 11.3 Column Dialog Constructors

Constructor	Syntax	Explanation
`Button`	`Button( "OK" ), Button( "Cancel" )`	Draws an OK or a Cancel button. If OK is clicked, `Button(1)` is returned. If Cancel is clicked, `Button(-1)` is returned.
`Check Box`	*var*`( CheckBox( "`*Text after box*`", <1|0> ) )`	When OK is clicked, a selected check box assigns 1 to the variable. A check box that is not selected assigns 0 to the variable. Add an optional 1 for the check box to be selected (on), or 0 for it to be not selected (off) when the window first appears. The default value is 0 (off).
`Col List`	*var*`=Col List( "`*role*`", <MaxCol( `*n*` )>, <Datatype(` *type* `)> )`	Creates a selection destination with a *role* button; the user's choices are returned in a list item of the form *var*={*choice 1*, *choice 2*, …, *choice n*}. You can specify the minimum and maximum number of columns using `MinCol(` *n* `)` or `MaxCol(` *n* `)`. You can specify the required data type of the column using `Datatype(`*type*`)`. The choices for *type* are `Numeric`, `Character`, or `Rowstate`.
`Combo Box`	*var*`( ComboBox( "`*choice1*`", "`*choice2*`", ...) )`	Produces a menu with the choices listed. The first choice is the default. Choices evaluate to quoted text strings. Choices can also be inside a list.
`Edit Number`	*var*`( Edit Number(` *number* `) )`	Produces an edit field for a number with *number* as the default value. When OK is clicked, the number entered in the field is assigned to the variable.

Table 11.3 Column Dialog Constructors *(Continued)*

Constructor	Syntax	Explanation
Edit Text	*var*(Edit Text("*string*", <width(*x*)>))	Produces an edit field for a string with *string* as the default value. You can also specify the minimum width of the box in pixels. The default width is 72 pixels. When OK is clicked, the text entered in the field is assigned to the variable.
HList	HList(*item*, *item*, ...)	Top aligns and spaces the *items* in a horizontal row. Placing a pair of VLists within an HList produces a top aligned, spaced pair of columns.
Line Up Box	Line Up Box(*n*, *item_11*, *item_12*, ..., *item_1n*, ..., *item_nn*)	Lines up the *items* listed in *n* columns, where *item_ij* is the *j*th item of the *i*th row.
List Box	var=List Box({"item", "item", ...}, width(50), max selected(2), NLines(6))	Creates a display box that contains a list of items.
Radio Buttons	*var*(RadioButtons("*choice1*", "*choice2*", ...))	Produces a vertical, left-justified list of radio buttons with the choices specified. The first choice is the default. When OK is clicked, the button that is selected is assigned to the variable. Choices must evaluate to quoted text strings.
string	"*string*"	Draws text in the window. For example, can include a labeling string before an Edit Number field. All strings must be quoted.
VList	VList(*item*, *item*, ...)	Left aligns and spaces the *items* in a vertical column. Placing a pair of HLists within a VList produces a left-aligned, spaced pair of rows.

Convert Deprecated Dialog to New Window

Because `New Window()` provides more display options and better control over the content and functions of the window, rewrite your `Dialog()` scripts.

Note: `Dialog()` scripts will not run in the future.

The following sections illustrate how to rewrite `Dialog()` scripts as `New Window()` scripts.

- "New Window Example"
- "Example of Deprecated Dialog"
- "Differences Between New Window and the Deprecated Dialog"
- "Optional Scripts in New Window"

New Window Example

The following example creates a window with multiple vertical and horizontal list boxes:

```
win = New Window( "New Window Example",
   <<Modal,
   <<ReturnResult, // ensure that you get the same result as with Dialog
   V List Box(
      V List Box(
         Text Box( "Radio Frequency Embolism Projection" ),
         Line Up Box(
            NCol( 2 ),
            Text Box( "Lower Spec Limit" ),
            lsl_box = Number Edit Box( 230 ),
            Text Box( "Upper Spec Limit" ),
            usl_box = Number Edit Box( 340 ),
            Text Box( "Threshold" ),
            threshold_box = Number Edit Box( 275 )
         ),
         H List Box(
            Panel Box( "Type of Radio",
                  rb_box1 = Radio Box( {"RCA", "Matsushita",
                  "Zenith", "Sony"} )),
            Panel Box( "Type of Antenna" ,
                  rb_box2 = Radio Box( {"Dish", "Helical", "Polarizing",
                  "Radiant Array"} ) )
         ),
         cb_box1 = Check Box( "Emission Synchronization" ),
         Text Box( "Title for plot" ),
         title_box = Text Edit Box( "My projection" ),
```

```
            H List Box( Text Box( "Quality" ),
            cb_box2 = Combo Box( {"Fealty", "Loyalty", "Piety", "Obsequiousness"}
      ) )
         ),
         H List Box(
            Align( Right ),
            Spacer Box(),
            Button Box( "OK",
               lsl = lsl_box << Get;
               usl = usl_box << Get;
               threshold = threshold_box << Get;
               radio_type = rb_box1 << Get;
               antenna = rb_box2 << Get;
               synch = cb_box1 << Get;
               title = title_box << Get Text;
               quality = cb_box2 << Get;
            ),
            Button Box( "Cancel" )
         )
      )
   );
   If(win["Button"] == 1,

   // if the user clicks OK, show the selections in the log
      Show( "OK", lsl, usl, threshold, radio_type, antenna, synch, title,
      quality);
      ,
      Show( "Canceled" ); // if the user clicked Cancel, print "Canceled"
   );
```

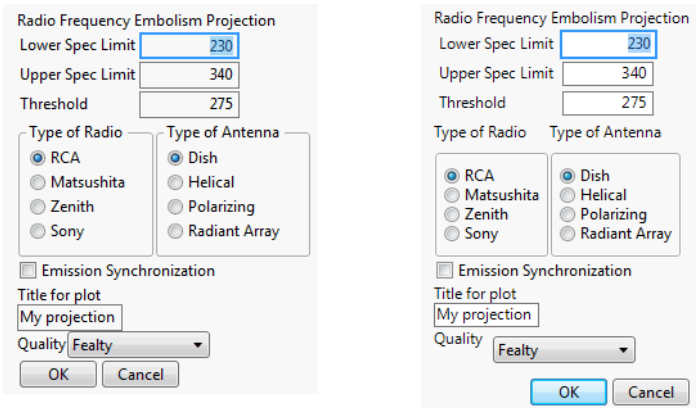
Example of Deprecated Dialog

Compare the previous `New Window()` example to this example of the deprecated `Dialog()`.

```
dlg = Dialog(
   Title( "Dialog Example" ),
   H List(
      V List(
         "Radio Frequency Embolism Projection",
         Lineup( 2,
            "Lower Spec Limit", lsl = Edit Number( 230 ),
            "Upper Spec Limit", usl = Edit Number( 340 ),
            "Threshold", threshold = Edit Number( 275 )
         ),
         H List(
            V List(
```

```
                "Type of Radio",
                type = Radio Buttons( "RCA", "Matsushita", "Zenith", "Sony" )
             ),
             V List(
                "Type of Antenna",
                antenna = Radio Buttons( "Dish", "Helical", "Polarizing",
                   "Radiant Array" )
             )
          ),
          synch = Check Box( "Emission Synchronization", 0 ),
          "Title for plot",
          title = Edit Text( "My projection" ),
          H List(
             "Quality",
             quality = Combo Box( "Fealty", "Loyalty", "Piety", "Obsequiousness"
             )
          )
       ),
       V List( Button( "OK" ), Button( "Cancel" ) )
    )
);
If( dlg["Button"] == 1,
   Show(
      "OK",
      dlg["lsl"],
      dlg["usl"],
      dlg["threshold"],
      dlg["type"],
      dlg["antenna"],
      dlg["synch"],
      dlg["title"],
      dlg["quality"]
   ),
   Show( "Canceled" )
);
```

Figure 11.38 Results from New Window (left) and the Deprecated Dialog (right)



Note: Use `Return Result` to get the same result in `New Window()` as with the deprecated `Dialog()`. See "Construct a Modal Window" for an example.

Differences Between New Window and the Deprecated Dialog

This section shows the differences between display boxes in `New Window()` and the deprecated `Dialog()` scripts. Table 11.4 summarizes those differences. More information can be found in the sections that follow the table.

Table 11.4 Display Boxes in New Window and Deprecated Dialog Scripts

Box Type	New Window	Deprecated Dialog

"Strings and Lists"	`win = New Window( "Combo Box",`
	`  <<Modal,`
	`  <<ReturnResult,`
	`  cb = Combo Box( {"True", "False"} ),`
	`);`
		`Dialog( Title( "Combo Box" ),`
		`  cb = Combo Box( "True", "False" ),`
		`);`

"List Boxes and Check Boxes"	`win = New Window( "H List Box",`
	`  <<Modal,`
	`  <<ReturnResult,`
	`  H List Box(`
	`  kb1 = Check Box( "a" ),`
	`  kb2 = Check Box( "b" ),`
	`  kb3 = Check Box( "c" )`
	`  ),`
	`);`
		`dlg = Dialog( Title( "H List" ),`
		`  H List(`
		`  kb1 = Check Box( "a", 0 ),`
		`  kb2 = Check Box( "b", 0 ),`
		`  kb3 = Check Box( "c", 0 )`
		`  ),`
		`);`

Table 11.4 Display Boxes in New Window and Deprecated Dialog Scripts *(Continued)*

Box Type	New Window	Deprecated Dialog
"Line Up Items"	win = New Window("Line Up Box", <<Modal, <<ReturnResult, V List Box(Line Up Box(NCol(2), Text Box("Set this value"), var1=Number Edit Box(42), Text Box("Set another value"), var2=Number Edit Box(86),),), H List Box(Button Box("OK"), Button Box("Cancel")); // returns: {Button(1)}	dlg = Dialog(V List(Line Up(2, "Set this value", variable=Edit Number(42), "Set another value", var2=Edit Number(86)), H List(Button("OK"), Button("Cancel")))); // returns: {variable = 42, var2 = 86, Button(1)}
"Radio Boxes"	win = New Window("Radio Box", <<Modal, <<ReturnResult, Panel Box("Select", rbox = Radio Box({"a", "b", "c"})));	dlg = Dialog(Title("Radio Buttons"), rb = Radio Buttons({"a", "b", "c"}),);
"Text Edit Boxes"	win = New Window("Text Edit Box", <<Modal, <<ReturnResult, V List Box(Text Box("Strings"), str1 = Text Edit Box("The"), str2 = Text Edit Box("quick"),));	dlg = Dialog(Title("Edit Text"), V List("Strings", str1 = Edit Text("The"), str2 = Edit Text("quick"),));

Table 11.4 Display Boxes in New Window and Deprecated Dialog Scripts *(Continued)*

Box Type	New Window	Deprecated Dialog

"Number Edit Boxes"	`win = New Window( "Number Edit Box",`
	`<<Modal,`
	`<<ReturnResult,`
	`V List Box(`
	`Text Box("Random Numbers"),`
	`num1 = Number Edit Box(`
	`Random Uniform()),`
	`num2 = Number Edit Box(`
	`Random Uniform() * 10),`
	`...);`
		`dlg = Dialog(`
		`Title( "Edit Number" ),`
		`V List(`
		`"Random Numbers",`
		`num1 = Edit Number(`
		`Random Uniform() ),`
		`num2 = Edit Number(`
		`Random Uniform() * 10 ),`
		`...);`

Strings and Lists

A major difference between the `New Window()` and the deprecated `Dialog()` functions is how you specify lists and strings as arguments. In `New Window()`, the items must be placed in a list. In the deprecated `Dialog()`, items are generally separated by commas. For example, compare the following two instances of a combo box.

```
win = New Window( "Combo Box", // modal New Window
   <<Modal,
   <<Return Result,
   cb = Combo Box( {"True", "False"} ), // items are in a bracketed list
);

Dialog( Title( "Combo Box" ),// modal Dialog
   cb = Combo Box( "True", "False" ), // items are separated by commas
   );
```

Note: The script does not explicitly define the OK button. For modal windows, JMP automatically adds the OK button. Explicitly define each additional button.

Additionally, the deprecated `Dialog()` could include `" "` to indicate empty text. `New Window()` requires that empty text appear in a `Text Box (" ")`.

List Boxes and Check Boxes

In `New Window()`, boxes are horizontally aligned with `H List Box()` and vertically aligned with `V List Box()`. In the deprecated `Dialog()`, you use `H List` and `V List` instead.

Note: Boxes are vertically aligned by default if `H List Box()` or `V List Box()` are omitted.

The following example creates a window with three horizontal check boxes and an OK button:

```
win = New Window( "H List Box",
   <<Modal,
   <<Return Result,
   H List Box(
      kb1 = Check Box( "a" ),
      kb2 = Check Box( "b" ),
      kb3 = Check Box( "c" )
   ),
);
```

Here is the same window created with the deprecated `Dialog()` and an `H List()`:

```
dlg = Dialog( Title( "H List" ),
   H List(
         kb1 = Check Box( "a", 0 ),
         kb2 = Check Box( "b", 0 ),
         kb3 = Check Box( "c", 0 )
     ),
);
```

Line Up Items

In `New Window()`, use `Line Up Box()` to arrange items in the number of columns that you specify. In the deprecated `Dialog()`, you used `Line Up`.

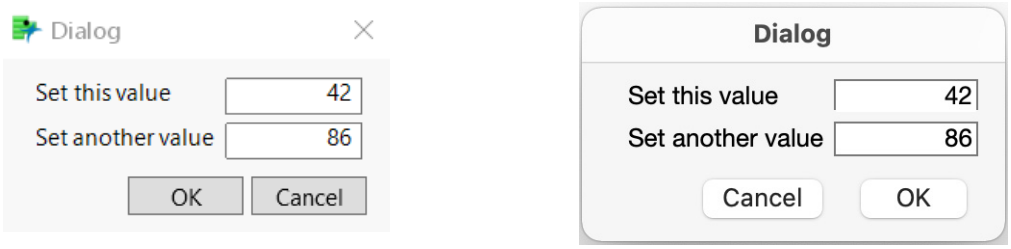
The following example arranges the text boxes in one column and the number edit boxes in another column:

```
win = New Window( "Line Up Box",
   <<Modal,
   <<Return Result,
   V List Box(
      Lineup Box( N Col( 2 ),
         Text Box( "Set this value" ),
         var1 = Number Edit Box( 42 ),
         Text Box( "Set another value" ),
         var2 = Number Edit Box( 86 ),

      ),
   ),
   H List Box( Button Box( "OK" ), Button Box( "Cancel" ) ) )
);
```

Clicking **OK** returns the following result in the log window:

```
List( 3 elements ) assigned.
```

Figure 11.39 Default Dialog Arrangements for Windows (left) and macOS (right)



Here's the same window created with the deprecated `Dialog()` and `Line Up`:

```
dlg = Dialog(
   V List(
      Line Up( 2,
         "Set this value", variable = Edit Number( 42 ),
         "Set another value", var2 = Edit Number( 86 )
      ),
      H List( Button( "OK" ), Button( "Cancel" ) )
   )
);
```

Clicking **OK** returns the following result in the log window:

```
{variable = 42, var2 = 86, Button(1)}
```

Note: JMP does exert some control over OK and Cancel button positions to ensure that dialog boxes are consistent with what the operating system expects. In certain cases, JMP needs to override your `H List Box()`, `V List Box()`, and `Line Up Box()` settings for `Button( "OK" )` and `Button( "Cancel" )`. Do not be alarmed if the result is slightly different from what you expect.

Radio Boxes

Another difference between `New Window()` and the deprecated `Dialog()` is the usage of `Radio Box`.

In `New Window()`, you must define the `Panel Box()` container for `Radio Box()` if you want the radio buttons in a box.

```
win = New Window( "Radio Box",
   <<Modal,
   <<Return Result,
   Panel Box( "Select",
      rbox = Radio Box( {"a", "b", "c"} )
   )
);
```

In the deprecated `Dialog()`, the `Radio Buttons()` automatically appear in a panel box.

```
dlg = Dialog( Title( "Radio Buttons" ),
   rb = Radio Buttons( {"a", "b", "c"} ),
);
```

Text Edit Boxes

In `New Window()`, use `Text Edit Box()` to create editable boxes that contain specified strings. In the deprecated `Dialog()`, you used `Edit Text()`.

```
win = New Window( "Text Edit Box",
   <<Modal,
   <<Return Result,
   V List Box(
      Text Box( "Strings" ),
      str1 = Text Edit Box( "The" ),
      str2 = Text Edit Box( "quick" ),
      str3 = Text Edit Box( "brown" ),
      str4 = Text Edit Box( "fox" ),
      str5 = Text Edit Box( "jumps" ),
      str6 = Text Edit Box( "over" ),
      str7 = Text Edit Box( "the" ),
      str8 = Text Edit Box( "lazy" ),
      str9 = Text Edit Box( "dog" ) ) );
```

Here is the same window created with the deprecated `Dialog()` and `Edit Text()`:

```
dlg = Dialog(
   Title( "Edit Text" ),
   V List(
      "Strings",
      str1 = Edit Text( "The" ),
      str2 = Edit Text( "quick" ),
      str3 = Edit Text( "brown" ),
      str4 = Edit Text( "fox" ),
      str5 = Edit Text( "jumps" ),
      str6 = Edit Text( "over" ),
      str7 = Edit Text( "the" ),
      str8 = Edit Text( "lazy" ),
      str9 = Edit Text( "dog" )
   )
);
```

Alternatively, use `String Col Edit Box()` to create editable boxes within in a table structure that contains specified strings. For example, the following script creates a window named "String Col Edit Box". The window contains a column (labeled "Strings") of editable boxes, each of which contains the specified string:

```
win = New Window( "String Col Edit Box",
   <<Modal,
   <<Return Result,
   steb = String Col Edit Box(
      "Strings",
      {"The", "quick", "brown", "fox", "jumps", "over", "the", "lazy", "dog"}
   )
);
```

Note: In the preceding example, `New Window()` assigns only two elements to the list (the specified list items and the button).

Number Edit Boxes

In `New Window()`, use `Number Edit Box()` to create editable boxes that contains the specified numbers. In the deprecated `Dialog()`, you used `Edit Number()`.

```
win = New Window( "Number Edit Box",
   <<Modal,
   <<ReturnResult,
   V List Box(
      Text Box( "Random Numbers" ),
      num1 = Number Edit Box( Random Uniform() ),
      num2 = Number Edit Box( Random Uniform() * 10 ),
      num3 = Number Edit Box( Random Uniform() * 100 ),
      num4 = Number Edit Box( Random Uniform() * 1000 )
   ),
);
```

Here is the same window created with the deprecated `Dialog()` and `Edit Number()`:

```
dlg = Dialog(
   Title( "Edit Number" ),
   V List(
      "Random Numbers",
      num1 = Edit Number( Random Uniform() ),
      num2 = Edit Number( Random Uniform() * 10 ),
      num3 = Edit Number( Random Uniform() * 100 ),
      num4 = Edit Number( Random Uniform() * 1000 )
   ),
);
```

Another method for creating the same window uses `Number Col Edit Box()` to create editable boxes within in a table structure that contains specified numbers. For example, the following script creates a window named "Number Col Edit Box". The window contains a column (labeled "Random Numbers") of editable boxes, each of which shows the specified type of random number:

```
win = New Window( "Number Col Edit Box",
   <<Modal,
   <<ReturnResult,
   nceb = Number Col Edit Box(
      "Random Numbers",
      {num1 = Random Uniform(), num2 = Random Uniform() * 10, num3 =
      Random Uniform() * 100, num4 = Random Uniform() * 1000}
   ),
);
```

Note: In the preceding example, the `Number Col Edit Box()` assigns only two elements to the list. The `New Window()` and deprecated `Dialog()` examples for `Number Col Edit Box()` assign five elements to the list.

Optional Scripts in New Window

One of the benefits of using `New Window()` is the ability to add optional scripts to display boxes. In the deprecated `Dialog()`, the following combo box could not include an optional script. If you wanted an action associated with any of the display box controls, you had to place a script as the control's last argument. For example:

```
win = New Window( "Combo Box",
<<Modal,
<<ReturnResult,
   comboObj = Combo Box(
      {"True", "False"},
      << Set( 1 ),
      Print( comboObj << Get )
   )
);
```

When the user selects a different value, the selected item number (1 or 2 in this case, because there are two items in the combo box) is printed to the log.

Scripting the Script Editor

Even the script editor window is a display tree in JMP, which means you can write a JSL script to write and save another JSL script.

There is no `New Script` command. Instead, to open a new script window, you use the `New Window()` function and then send it a message to tell it that it's a script window:

```
ww = New Window( "Window Title", <<Script, "Initial Contents" );
```

The last argument is optional. If you include a string, the new script window contains that string.

In the `New Window` example above, `ww` is a reference to the display box that is the entire window. To write to a script window, you first need to get a reference to the part of the display box that you can write to, which is called a script box:

```
ed = ww[Script Box(1)];
```

Using the reference `ed`, you can add text, remove text, and get the text that is already there.

```
ed << Get Text();
   "Initial Contents"
```

Use `Set Text` to set all the text in the script window. The following command clears all text in the Script Window and then adds `aaa=3;` followed by a return:

```
ed << Set Text( "aaa=3;\!N" );
```

Use `Append` to add additional text to the end of the script window.

```
ed << Append Text( "bbb=1/10;" );
ed << Append Text( "\!Ncc=4/100;" );
```

Use the `Get Line Text` command to get the text at the line of a specified line number. Use the `Set Line Text` command to replace a specified line of text with new text.

```
ed << Get Line Text( 2 );
ed << Set Line Text( 2, "bbb = 0.1;" );
```

Use the `Get Line Count` message to get the total number of lines in the script. The `Get Lines` message returns a list of each line in the script as a string.

```
ed << Get Line Count();
ed << Get Lines();
```

Use the `Reformat` message to automatically format a script for easier reading.

```
ed << Reformat();
```

To run an entire script in a script window, use it the `Run` message.

```
ed << Run();
```

To close the script window without saving the script, send the window the `Close Window` message, just like you can do with any JMP window.

```
ww << Close Window( nosave );
```

To save and view the script, use the `Save Text` and `Load Text File` messages:

```
ww = New Window( "Test", <<Script, "Open(\!"$SAMPLE_DATA/Big
  Class.jmp\!");");
ww << Save Text( "$TEMP/Test.jsl" );
Write( Load Text File( "$TEMP/Test.jsl" ) );
```

Note: When the script is loaded into the script window, the window is considered "dirty" or modified. Use `<<Set Dirty(0)` to remove this property. Otherwise, you are prompted to save the script when you close the window.

To save the script and close the window, follow this example:

```
ww = New Window( "Test", <<Script, "Open(\!"$SAMPLE_DATA/Big
  Class.jmp\!");");
ww << Save Text( "$TEMP/Test.jsl" );
ww << Close Window();
```

Technical Details

Write Tab Box and Tab Page Box Scripts

`Tab Page Box()` organizes the title and contents of the page in one display box. When `Tab Page Box()` is inside `Tab Box()`, a tabbed window with multiple tabbed pages appears. See "Tab Box and Tab Page Box".

In previous versions of JMP, display boxes such as `V List Box()` contained tab contents. Consider the following script:

```
win = New Window( "Tab Box",
   tb = Tab Box(
      "First page", // name of the tab
      V List Box( // tab contents
         Text Box( "first line of first page" ),
         Text Box( "second line of first page" )
      ),
      "Second page", // name of the tab
      V List Box( // tab contents
         Text Box( "first line of second page" ),
         Text Box( "second line of second page" )
      ),
```

```
    )
);
```

We recommend that `Tab Box()` contain a `Tab Page Box()` for each tab. The preceding example is rewritten as follows:

```
win = New Window( "Tab Box",
   tb = Tab Box(
      Tab Page Box( // tab contents
         Title( "First page of my tab box" ), // name of the tab
         Text Box( "first line of first page" ),
         Text Box( "second line of first page" )
      ),
      Tab Page Box( // tab contents
         Title( "Second page" ), // name of the tab
         Text Box( "first line of second page" ),
         Text Box( "second line of second page" )
      )
   )
);
```

Note: `Tab Page Box()` requires that the title be defined in a `Title()` function.

When a tab page box is created or dragged outside a tab box, the tab page box is a stand-alone container. The title appears in a shaded box at the top of the page, similar to the title in `Sheet Part()`; the title does not appear in an interactive tab. "Dashboard with One Row of Reports" provides an example.

Deprecated Tab Box and Tab Page Box Messages

Some messages previously used for `Tab Box()` are deprecated. Their counterparts are compatible with `Tab Page Box()`.

Table 11.5 Deprecated Tab Box() Messages

Deprecated Tab Box Message	Current Tab Page Box Message
`Set Title`	`Title`
`Set Tip`	`Tip`
`Set Icon`	`Icon`
`Set Closeable`	`Closeable`

See "Tab Box and Tab Page Box" for more information about the two display boxes.

Chapter 12

Scripting Graphs

Editing and Creating Two-Dimensional Graphs

The JSL script for a graph contains expressions that create shapes and lines, apply fill patterns, assign background colors, and so on. You can also right-click a graph and select Customize to add the same expressions.

This chapter also describes how to customize hover labels with rich text, images, and drill-down graphs.

For information about scripting three-dimensional graphs, see the chapter "Three-Dimensional Scenes".

Contents

Add Scripts to Graphs 659
Specify the Order of Graphical Elements 660
Create New Graphs from Scratch and Customize Them 666
Customize Graphs 667
Create Shapes for a Bubble Plot 669
How to Find Shapes 672
Add Lines, Arrows, Points, Shapes, and Text 673
Draw Lines 673
Draw Arrows 675
Draw Markers 677
Draw Pies and Arcs 678
Draw Regular Shapes: Circles, Rectangles, and Ovals 680
Draw Irregular Shapes: Polygons and Contours 682
Add Text 685
Specify Colors 686
Specify Transparency 690
Add Fill Patterns 691
Specify Line Types 694
Draw with Pixels 695
Interactive Graphs 696
Handle() 696
Mousetrap() 701
Drag Functions 703
Troubleshoot Interactive Graphs 706
Create a Color Picker 706
Clip Display Segs in Graphics 707
Create Background Maps 709
Graph Elements 710
Plot Functions 711
Get the Properties of a Graphics Frame 717
Add a Legend 717
Hover Labels 718
Add Rich Text to Hover Labels Using Textlets 719
Add Graphs or Images to Hover Labels Using Graphlets 722
Update Names and Values in Hover Labels Using Gridlets 732
Create Drill-Down Graphs 737
Work with the Hover Label Execution Context 740

Add Scripts to Graphs

When you right-click a graphics frame, you can enter or paste JSL commands as shown in the following example. The script usually contains drawing commands that run in the context of the graphics frame. The context of the graphics frame includes the data range of the axes and the order that the data and scripts are drawn in.

1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Select **Analyze** > **Fit Y by X**.
3. Choose weight for Y and height for X, and then click **OK**.
4. Right-click inside the graph and select **Customize**.

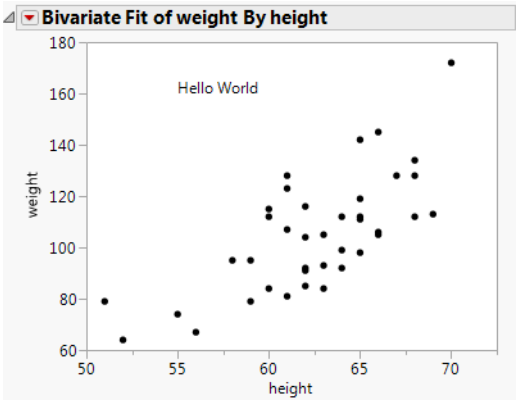
Figure 12.1 Customizing the Graph Options

5. Click the **Add** button ➕ to add a new graphics script.
6. Type the following text and click **OK**.

```
Text( {55, 160}, "Hello World" );
```

Now the graph has a text element at the graph's *x*-coordinate 55 and *y*-coordinate 160.

Figure 12.2 Adding a Script to a Graph Interactively



By default, the data points appear on top of graphics.

For example, add a new script in the Customize Graph window:

```
Fill Color( "Green" ); Rect( 57, 175, 65, 110, 1 );
```

A solid green rectangle appears. The list of scripts shows the order in which each script is drawn, so the first item on the list is drawn first.

- If you want the rectangle behind everything, move the rectangle script above the text script: select the rectangle script and click the up button .
- If you want the rectangle on top of everything, select the rectangle script and click the down button .

You can arrange all the scripts in a graph into the drawing order that you prefer. Any new script is initially added directly after the item in the list that is selected.

Tip: To use a script that references a column name, use `Column( colname )` or a colon (`:colname`) to scope it properly.

Hint: To see the JSL for the above actions, select **Save Script > To Script Window** from the red triangle menu.

Specify the Order of Graphical Elements

You can also add graphics elements using JSL instead of doing so interactively:

```
Frame Box <<Add Graphics Script(<order>, <"description">, script)
```

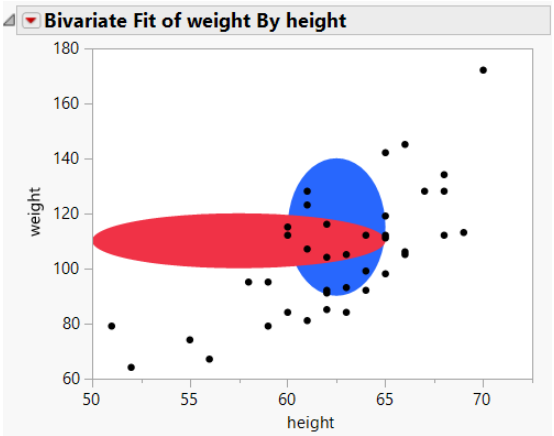
When you add graphical elements using JSL, they are drawn on top of whatever is already in the graph.

- The optional `order` argument specifies in what order to draw the graphics element. The order can be the keyword `Back` or `Front` *or* an integer that specifies the drawing order for a number of graphics elements. For example, if you add an oval to a scatterplot, the oval is drawn on top of the markers. The keyword `Back` or `1` cause the oval to be drawn last. `Front` or `2` means the object is drawn first.
- The optional `Description` argument is a string that appears in the Customize Graph window next to the graphics script.

To specify the drawing order for a number of graphics elements, use an integer for the `order` argument to determine where each is drawn in relation to the others. The following script first adds a blue oval and then adds a red oval over the blue oval. Both ovals are behind the points.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
op = dt << Bivariate(
  X( :height ),
  Y( :weight ),
);

Report( op )[Frame Box( 1 )] << Add Graphics Script(
  1, // draw the blue oval
  Description( "Blue Oval" ),
  Fill Color( "Blue" );
  Oval( 60, 140, 65, 90, 1 );
);

Report( op )[Frame Box( 1 )] << Add Graphics Script(
  2, // draw the red oval over the blue oval
  Description( "Red Oval" ),
  Fill Color( "Red" );
  Oval( 50, 120, 65, 100, 1 );
);
```

Figure 12.3 Specifying the Drawing Order



Copy and Paste Frame Contents or Settings

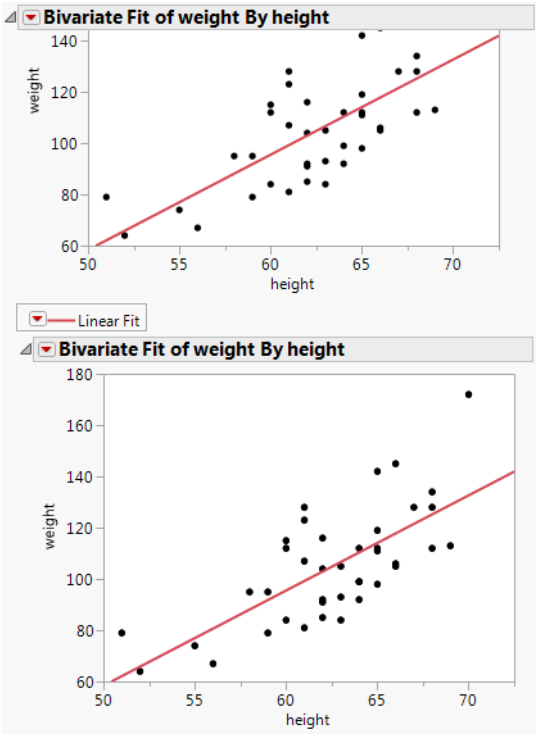
You can copy and paste the contents or the settings of a frame with these JSL commands:

```
obj << Copy Frame Contents // contents such as a graph or line of fit
obj << Paste Frame Contents
obj << Copy Frame Settings // settings such as background color
obj << Paste Frame Settings
```

The following example creates two Bivariate graphs, adds a line of fit to the first graph, and then copies and pastes the line to the second graph.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( y( :weight ), x( :height ) );
rbiv1 = biv << Report; // create a report object
rbiv2 = rbiv1 << Clone Box; // create a copy of the Bivariate graph

/* place the cloned box below the Bivariate graph
to create a second Bivariate graph */
rbiv1 << Append( rbiv2 );

framebox1 = rbiv1[Frame Box( 1 )]; /* assign references to
the reports' frame boxes*/
framebox2 = rbiv2[Frame Box( 1 )];
biv << Fit Line; // fit a line to the first Bivariate graph

// copy the frame contents from framebox1
framebox1 << Copy Frame Contents;

// paste the frame contents to framebox2
framebox2 << Paste Frame Contents;
```

Figure 12.4 Copying and Pasting Frame Contents

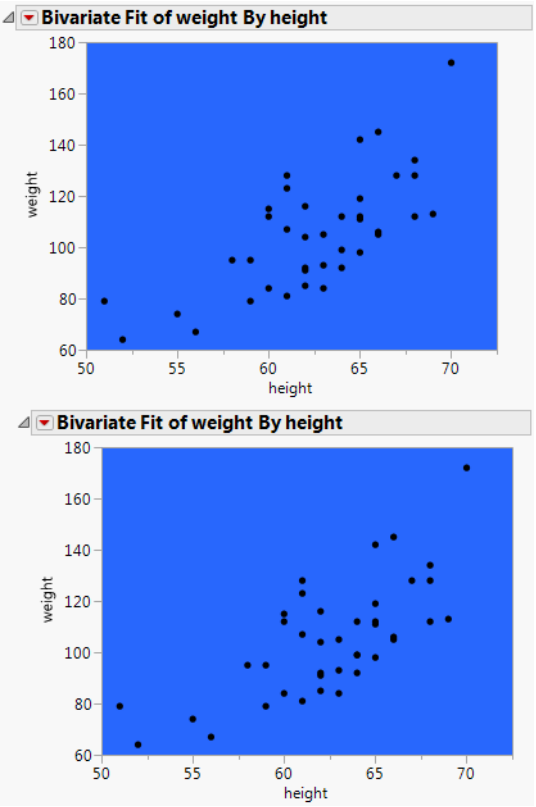


The following example creates two Bivariate graphs, sets the background of the first graph to blue, and then copies and pastes the blue background setting to the second graph:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );

biv1 = dt << Bivariate( y( :weight ), x( :height ) );

// create the first Bivariate graph
rbiv1 = biv1 << Report; // create a report object
rbiv2 = rbiv1 << Clone Box; // create a copy of the Bivariate graph

/* place the cloned box below the Bivariate graph
to create a second Bivariate graph */
rbiv1 << Append( rbiv2 );

framebox1 = rbiv1[Frame Box( 1 )]; /* assign a reference to the
                                      reports' frame boxes */
framebox2 = rbiv2[Frame Box( 1 )];

// assign the background color to framebox1
framebox1 << Background Color( "Blue" );
```

```
// copy the frame settings from framebox1
framebox1 << Copy Frame Settings;

// paste the frame settings to framebox2
framebox2 << Paste Frame Settings;
```

Figure 12.5 Copying and Pasting the Background



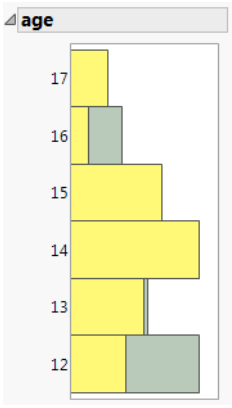
Copy and Paste Histograms

The following example shows how to create two histograms and copy the second histogram into the first histogram. A platform will not save a pasted-in segment, so you'll need to run a script like this to reproduce the graph.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dist = dt << Distribution(
   SendToByGroup( {:sex == "F"} ),
   Nominal Distribution( Column( :age ) ),
   Histograms Only,
   By( :sex ),
```

```
   SendToByGroup(
      {:sex == "M"},
      SendToReport(
         Dispatch(
            {"Distributions sex=M", "age"},
            "Distrib Nom Hist",
            FrameBox,
            {DispatchSeg(
               Hist Seg( 1 ),

               // set the histogram color for males
               Fill Color( "Light Yellow" ),
            )}
         )
      )
   )
);

For( i = 2, i <= N Items( dist ), i++,
   Report( dist[i] )[FrameBox( 1 )] << Copy Frame Contents;

   // copy the the second histogram into the first histogram
   Report( dist[1] )[FrameBox( 1 )] << Paste Frame Contents;
);

New Window( "Distribution", Outline Box( "age", Report( dist[1] )[Picture Box(
   1 )] ) );
```

Figure 12.6 Pasted Histogram



Interactively Copy and Paste Frame Contents and Settings

To interactively copy and paste the frame settings, follow these steps:

1. Right-click the graph and select **Edit > Copy Frame Settings**.
2. Right-click the frame to which you want to paste the settings and select **Edit > Paste Frame Settings**.

To interactively copy and paste the frame contents, follow these steps:

1. Right-click the graph and select **Edit > Copy Frame Contents**.
2. Right-click the frame to which you want to paste the contents and select **Edit > Paste Frame Contents**.

Create New Graphs from Scratch and Customize Them

Graphics scripts are set up inside the `Graph Box()` command within a `New Window()` command.

```
New Window("title", <arguments>, Graph Box( named arguments,..., script));
```

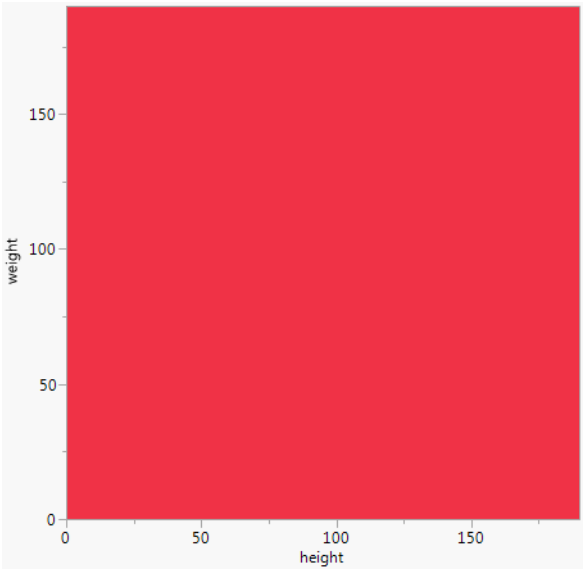
The following arguments are named arguments for `Graph Box()`:

```
Frame Size( horizontal, vertical ), // size in pixels
X Scale( xmin, xmax ), Y Scale( ymin, ymax ), // range of x and y axes
X Name( "x" ), Y Name( "y" ), // names for x and y axes
Suppress Axes // omit the axes
```

For example, `Graph Box()` accepts named arguments as shown in the following example, which creates a graph with a red background.

```
win = New Window( "Named arguments",
   Graph Box(
      Frame Size( 300, 300 ), // construct the graph box
      X Scale( 0, 190 ), // set the x-axis range
      Y Scale( 0, 190), // set the y-axis range
      X Name( "height" ),
      Y Name( "weight" ),
      <<Background Color( "Red" ) // set the background color
   )
);
```

Figure 12.7 Creating a Graph



Alternatively, you can use the send << operator to send commands to the `Graph Box` instead of using the named arguments. This example also creates a graph with a red background.

```
win = New Window( "Messages",
   Graph Box(
      <<Frame Size( 300, 300 ),
      <<XAxis( 0, 190 ),
      <<YAxis( 0, 190 ),
      <<Background Color( "Red" )
   )
);
```

Customize Graphs

You can also make changes to graphs in a script. For example, create a window with a graph, get a reference to the report, and then set the frame box size.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( :weight ), X( :height ), Fit Line );
rbiv = biv << Report;
rbiv[Frame Box( 1 )] << Frame Size( 400, 400 );
```

To see a list of possible messages for any given display box object, select **Help > Scripting Index** and select **Display Boxes** from the list. Running the `Show Properties()` command is an alternative to using the Scripting Index. For example, here is a partial list of messages that you can send to an axis:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( :weight ), X( :height ), Fit Line );
rbiv = biv << Report;
Show Properties( rbiv[Axis Box( 1 )] );
   Axis Settings [Action](Provides a way to bring up the axis dialog or set
      configurable axis options for a given axis.)
   Revert Axis [Action](Restore the settings that this axis had originally.)
   Add Axis Label [Action]
   Remove Axis Label [Action]
   ...
```

The following expressions change the font of both axis labels (weight and height in the example above) to 12-point, italic Arial Black:

```
rbiv[Text Edit Box( 1 )] << Set Font( "Arial Black", 12, Italic );
rbiv[Text Edit Box( 2 )] << Set Font( "Arial Black", 12, Italic );
```

or

```
rbiv[Text Edit Box( 1 )] << Set Font( "Arial Black" );
rbiv[Text Edit Box( 1 )] << Set Font Style( "Italic" );
rbiv[Text Edit Box( 1 )] << Set Font Size( 12 );
rbiv[Text Edit Box( 1 )] << Set Font( "Arial Black" );
rbiv[Text Edit Box( 2 )] << Set Font Style( "Italic" );
rbiv[Text Edit Box( 2 )] << Set Font Size( 12 );
rbiv[Text Edit Box( 2 )] << Set Font Name( "Arial Black" ) << Set Font Size(
  12 ) << Set Font Style( Italic )
```

The default font specified in the Preferences is applied if the font is not installed on the computer.

Specifying a Missing Color Value

When you specify a color theme in a script, you can specify the color of a missing value. Use the `Missing` argument. In this example, the missing color is yellow. Refer to an online RGB value table for the color values.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Graph Builder(
   Size( 528, 450 ),
   Show Control Panel( 0 ),
   Variables(
      X( :height ),
      Y( :weight ),
      Color( Transform Column( "wc", Formula( Root( :weight - 80 ) ) ) ) )
```

```
    ),
    Elements( Points( X, Y, Legend( 8 ) ) ),
    SendToReport(
       Dispatch( {}, "400", ScaleBox,
          {Legend Model( 8,
             Properties( 0, {gradient(
                {Color Theme(
                   {"Blue to Gray to Red Copy", 4099, {{42, 63, 255}, {192,
                   192, 192}, {252, 11, 11}, Missing( {255, 211, 0} )}}
                )}
             )},
             Item ID( "wc", 1 )
             )
          )}
       )
    )
);
```

Create Shapes for a Bubble Plot

Bubble Plot supports predefined shapes and custom shapes. Shapes can give visual cues about the data or provide a more interesting shape than a circle. For example, you can show arrows (Figure 12.8) or a hurricane symbol on a map (Figure 12.9).

Specify a Predefined Shape

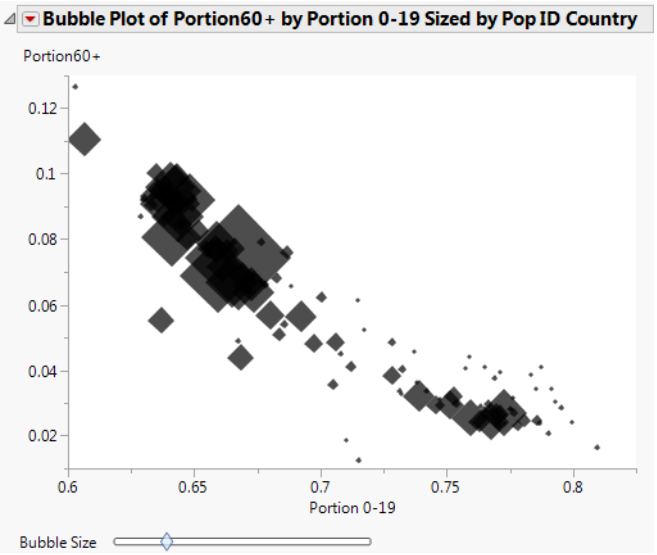
The `Set Shape` message specifies the predefined shape:

```
bp << Set Shape( "Circle"|"Triangle"|"Square"|"Diamond"|"Arrow" );
```

The following example creates the bubble plot and diamond shape shown in Figure 12.8:

```
dt = Open( "$SAMPLE_DATA/PopAgeGroup.jmp" );
bp = dt << Bubble Plot(
   X( :"Portion 0-19"n ),
   Y( :"Portion60+"n),
   Sizes( :Pop ),
   ID( :Country )
);
bp << Set Shape( "Diamond" );
```

Figure 12.8 Diamond Shapes



Create a Custom Shape

You can create custom shapes only in JSL through a combination of two messages:

```
bp << Set Custom Path( pathMatrix | pathText );
bp << Set Shape( "Custom" );
```

- In `Set Custom Path()`, the path can consist of an *N*x3 matrix or a string that contains SVG code. The default value is 0.

  ```
  bp << Set Custom Path(
     "M-1,-1 L-1,1 L0,0.5 L1,1 L1,-1 L0,-0.5 L-1,-1 Z, ..." );
  ```

- In `Set Shape()`, specify `"Custom"`.

  ```
  bp << Set Shape( "Custom" );
  ```

The following script creates the map and custom hurricane shape shown in Figure 12.9.

```
dt = Open( "$SAMPLE_DATA/Hurricanes.jmp" );
dt << Bubble Plot(
   X( :Longitude ),
   Y( :Latitude ),
   Sizes( "Wind (Knots)"n ),
   Time( :Date ),
   Coloring( :Landfall in USA ),
   ID( :Name and ID ),
   Speed( 220 ),
   Bubble Size( 51.43 ),
   Time Index( 729.9 ),
```

```
   Trail Bubbles( 1 ),
   All Labels( 0 ),
   No Labels( 0 ),
/* rotate shapes in the direction of the movement
   good for animated plots */
   Orient Shapes( 1 ),
   Set Custom Path( // specify the SVG string
      "M-0.99999898,0.00921400000000006 C-0.99999898,-0.51266322
   -0.55204076,-0.93622284 -0.000119060000000104,-0.93622284
   C0.1835362,-0.93622284 0.3556076,-0.88949112 0.5035232,-0.80766502
   C0.3236248,-1.36413554 0.0260398000000001,-1.815694896 -0.3555946,-2
   C0.4067376,-1.7567781 0.9372286,-0.91116206 1.0000004,0.00829720000000009
   C1.0000004,0.5301742 0.5518026,0.9548002 -0.00011905999999982,0.9547998
   C-0.18139054,0.9547998 -0.35150496,0.9090838 -0.49816916,0.8292164
   C-0.31759686,1.3771358 -0.0174124399999999,1.8176124 0.3602494,1.999999
   C-0.38991788,1.7606584 -0.92375834,0.908779 -0.99999898,0.00921400000000034
   z"
   ),
   Set Shape( "Custom" ), // set the custom shape
   Title Position( -88.679, 59.73 ),
   SendToReport(
      Dispatch(
         {},
         "Bubble Plot",
         FrameBox,
         {Frame Size( 743, 603 ), Background Map( Images( "Simple Earth" ) ),
         Grid Line Order( 3 ), Reference Line Order( 2 )}
      )
   )
);
```

Figure 12.9 Custom Hurricane Shape

How to Find Shapes

Search Wikimedia Commons or Google Images and include the keyword *SVG*. Look for SVG examples that are composed of stroke paths. SVG can be used to represent very complex drawings, with gradients and other elements. Only a path representation, which provides both outline and fill modes, is supported.

You might need to translate and scale the SVG that you find in order to get correct coordinates. Inkscape or similar tools can be useful for doing this manipulation. To copy the path from Inkscape, select **Edit > XML Editor**, find the `svg:path` element, and copy the path string from the `d Attribute` into `Set Custom Path()`.

Add Lines, Arrows, Points, Shapes, and Text

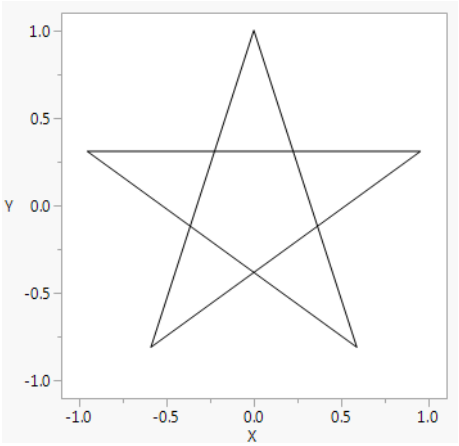
- "Draw Lines"
- "Draw Arrows"
- "Draw Markers"
- "Draw Pies and Arcs"
- "Draw Regular Shapes: Circles, Rectangles, and Ovals"
- "Draw Irregular Shapes: Polygons and Contours"
- "Add Text"
- "Specify Colors"
- "Specify Transparency"
- "Add Fill Patterns"
- "Specify Line Types"
- "Draw with Pixels"

Draw Lines

`Line()` draws lines between points.

```
win = New Window( "Five-Point Star",
   Graph Box(
      Frame Size( 300, 300 ),
      X Scale( -1.1, 1.1 ),
      Y Scale( -1.1, 1.1 ),
      Line(
         {Cos( 1 * Pi() / 10 ), Sin( 1 * Pi() / 10 )},
         {Cos( 9 * Pi() / 10 ), Sin( 9 * Pi() / 10 )},
         {Cos( 17 * Pi() / 10 ), Sin( 17 * Pi() / 10 )},
         {Cos( 5 * Pi() / 10 ), Sin( 5 * Pi() / 10 )},
         {Cos( 13 * Pi() / 10 ), Sin( 13 * Pi() / 10 )},
         {Cos( 1 * Pi() / 10 ), Sin( 1 * Pi() / 10 )}
      )
   )
);
```

Figure 12.10 Using Lines to Draw a Star



You can either specify the points in two-item lists as demonstrated above or as matrices of *x* and then *y* coordinates. Matrices are flattened by rows. You can use either row or column vectors, as long as you have the same number of elements in each matrix. The following expressions have the same effect:

```
Line( {1,2}, {3,0}, {2,4} ); // several {x,y} lists
Line( [1 3 2],[2 0 4] ); // row vectors
Line( [1,3,2], [2,0,4] ); // column vectors
Line( [1 3 2], [2,0,4] ); // one of each
```

The star example could also be drawn this way. Note that the script must include the full `Matrix( {...} )` notation rather than `[ ]` shorthand because the entries are expressions. The following example uses the `Matrix()` function.

```
win = New Window( "Five-Point Star",
   Graph Box(
      Frame Size( 300, 300 ),
      X Scale( -1.1, 1.1 ),
      Y Scale( -1.1, 1.1 ),
      Line(
         Matrix(
            { // the x coordinates
            Cos( 1 * Pi() / 10 ), Cos( 9 * Pi() / 10 ), Cos( 17 * Pi() / 10 ),
            Cos( 5 * Pi() / 10 ), Cos( 13 * Pi() / 10 ), Cos( 1 * Pi() / 10 )}
         ),
         Matrix(
            { // the y coordinates
            Sin( 1 * Pi() / 10 ), Sin( 9 * Pi() / 10 ), Sin( 17 * Pi() / 10 ),
            Sin( 5 * Pi() / 10 ), Sin( 13 * Pi() / 10 ), Sin( 1 * Pi() / 10 )}
         )
      ) ) );
```

`HLine()` draws a horizontal line across the graph at the *y*-value that you specify. Similarly, `VLine()` draws a vertical line down the graph at the *x*-value that you specify. Both functions support drawing multiple lines by using a matrix of values in the *y* argument. See "Mousetrap()".

Draw Arrows

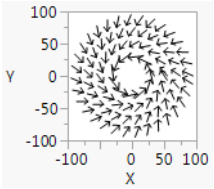
`Arrow()` draws an arrow from the first point to the second point. The default arrowhead is scaled to 1 plus the square root of the length of the arrow. To set the length of the arrowhead, add an optional first argument, specifying the length of the arrowhead in pixels. The following example draws simple arrows.

```
win = New Window( "Arrowheads",
   Graph Box(
      Pen Size( 4 );

      /* matrix 1 defines the x coordinates of an arrow
      matrix 2 defines the y coordinates of an arrow */
      Arrow( 20, [10 30 90], [88 22 44] );
   )
);
```

The following example draws arrows in a circular pattern.

```
win = New Window( "Hurricane",
   Graph Box(
      Frame Size( 100, 100 ),
      X Scale( -100, 100 ),
      Y Scale( -100, 100 ),
      For( r = 35, r < 100, r += 20,
         ainc = 2 * Pi() * 3 / r;
         For( a = 0, a < 2 * Pi(), a += ainc,
            x = r * Cosine( a );
            y = r * Sine( a );
            aa = a + ainc * 45 / r;
            rr = r - r / 6;
            x2 = rr * Cosine( aa );
            y2 = rr * Sine( aa );

            /* list 1 defines the start point for the arrow
            list 2 defines the end point for the arrow */
            Arrow( {x, y}, {x2, y2} );
         );
      )
   )
);
```

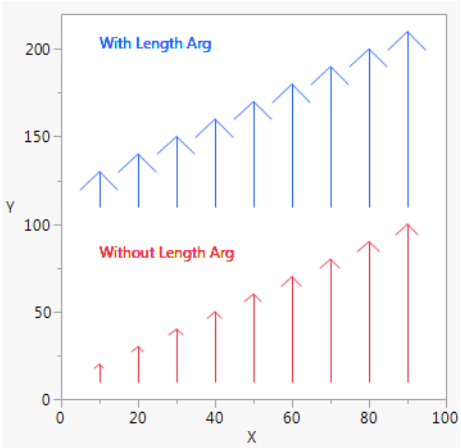
Figure 12.11 Drawing Arrows



The following example compares drawing with a specified length (19 pixels) and drawing with the default arrow head size.

```
win = New Window( "Arrowheads",
   Graph Box(
      Frame Size( 300, 300 ),
      X Scale( 0, 100 ),
      Y Scale( 0, 220 ),
      x = 10;
      y1 = 10;
      y2 = y1 + 10;
      For( i = 1, i < 10, i++,
         Pen Color( "Red" );
         Arrow( {x, y1}, {x, y2} );

         y2 += 10;
         y1 += 100;
         y2 += 100;
         Pen Color( "Blue" );
         Arrow( 20, {x, y1}, {x, y2} );

         x += 10;
         y1 -= 100;
         y2 -= 100;

         Text Color( "Red" );
         Text( {10, 80}, "Without Length Arg" );

         Text Color( "Blue" );
         Text( {10, 200}, "With Length Arg" );
      );
   )
);
```

Figure 12.12 Arrowhead Sizes



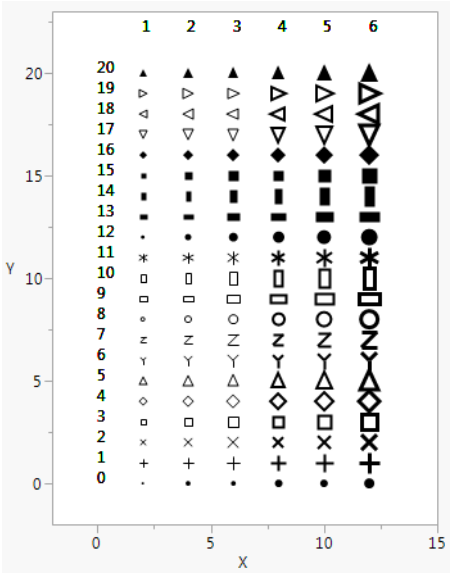
As with `Line()`, you can either specify the points in two-item lists as demonstrated above or as matrices of *x* and then *y* coordinates.

Draw Markers

`Marker()` draws a marker of the type that you specify (1–20) in the first argument at the point that you specify in the second argument. `Marker Size()` scales markers from 0 to 6 (dot to XXXL). To set markers to the size that is specified in the Preferences, use a value of `-1`.

```
ymax = 20;
win = New Window( "Markers",
   Graph Box(
      Frame Size( 300, 400 ),
      X Scale( -2, ymax - 5 ),
      Y Scale( -2, ymax + 3 ),
      For( j = 1, j < 7, j++,
         Marker Size( j );
         For( i = 0, i < (ymax + 1), i++,
            Marker( i, {j * 2, i} );
            Text( {0, i}, i );
            Text( {j * 2, ymax + 2}, j );
         );
      )
   )
);
```

Figure 12.13 Drawing Markers



You can also include a row state argument before, after, or instead of the marker ID argument. By using `Combine States()`, you can set multiple row states inside `Marker()`. Try substituting each of the following lines in the preceding script:

```
Marker( i, Color State( i ), {j * 2, i} );
Marker( Color State( i ), i, {j * 2, i} );
Marker(
   Combine States( Color State( i ), Marker State( i ), Hidden State( i ) ),
   {j * 2, i}
);
```

As with `Line()` and `Arrow()`, you can also specify points as matrices of *x* and then *y* coordinates.

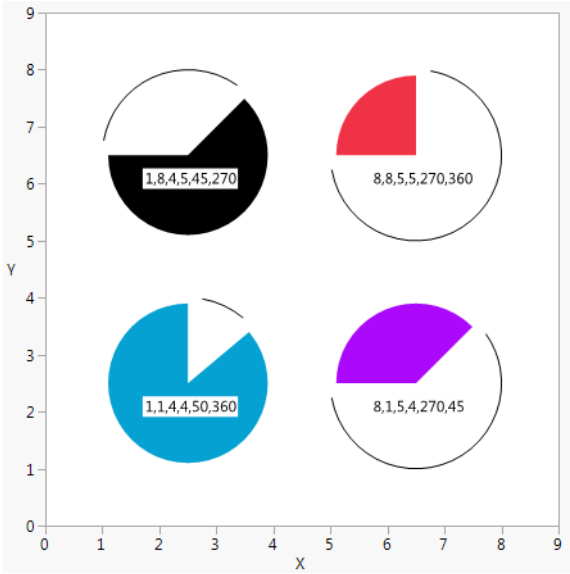
Draw Pies and Arcs

`Pie()` and `Arc()` draw wedges and arc segments. The first four arguments are *x1, y1, x2,* and *y2,* the coordinates of the rectangle to inscribe. The last two arguments are the starting and ending angle in degrees, where 0 degrees is 12 o'clock and the arc or slice is drawn clockwise from start to finish.

```
win = New Window( "Pies and Arcs",
   Graph Box(
      Frame Size( 400, 400 ),
      X Scale( 0, 9 ),
      Y Scale( 0, 9 ),
```

```
        Fill Color( "Black" ), // top left
        Pie( 1.1, 7.9, 3.9, 5.1, 45, 270 ),
        Text( Erased, {1.75, 6}, "1,8,4,5,45,270" ),
        Arc( 1, 8, 4, 5, 280, 35 ),
        Fill Color( "Red" ), // top right
        Pie( 7.9, 7.9, 5.1, 5.1, 270, 360 ),
        Text( Erased, {5.75, 6}, "8,8,5,5,270,360" ),
        Arc( 8, 8, 5, 5, 370, 260 ),
        Fill Color( "BlueCyan" ), // bottom left
        Pie( 1.1, 1.1, 3.9, 3.9, 50, 360 ),
        Text( Erased, {1.75, 2}, "1,1,4,4,50,360" ),
        Arc( 1, 1, 4, 4, 370, 40 ),
        Fill Color( "Purple" ), // bottom right
        Pie( 7.9, 1.1, 5.1, 3.9, 270, 45 ),
        Text( Erased, {5.75, 2}, "8,1,5,4,270,45" ),
        Arc( 8, 1, 5, 4, 55, 260 )
    )
);
```

Figure 12.14 Drawing Pies and Arcs



Draw Regular Shapes: Circles, Rectangles, and Ovals

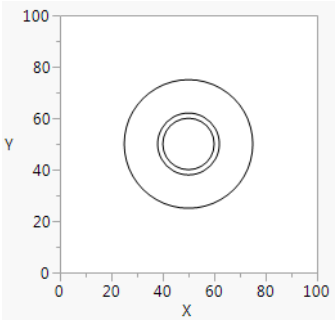
Circles

`Circle()` draws a circle with the center point and radius given. Subsequent arguments specify additional radii.

```
win = New Window( "Circles",
   Graph Box( Frame Size( 200, 200 ),
      Circle( {50, 50}, 10, 12, 25 )
   )
);
```

`Circle()` also has a final optional argument, `"fill"`. This string indicates that all circles defined in the function are filled with the current fill color. If `"fill"` is omitted, the circle is empty.

Figure 12.15 Drawing Circles



Note that a circle is always a circle, even if you resize the graph to a different aspect ratio. If you want your circle to change aspect ratios (in other words, cease being a circle) when the graph is resized, use an oval instead.

If you do not want your circle to resize if the graph is resized, specify the radius in pixels instead:

```
win = New Window( "Circles",
   Graph Box(
      Frame Size( 200, 200 ),
      Circle( {50, 50}, Pixel Radius( 10 ), Pixel Radius( 12 ), Pixel Radius(
   25 ) )
   )
);
```

Rectangles

`Rect()` draws a rectangle from the diagonal coordinates you specify. The coordinates can be specified either as four arguments in order (*left*, *top*, *right*, *bottom*), or as a pair of lists ({*left*, *top*}, {*right*, *bottom*}).

```
win = New Window( "Rectangles",
   Graph Box(
      Frame Size( 200, 200 ),
      Pen Color( 1 );
      Rect( 0, 40, 60, 0 );
      Pen Color( 3 );
      Rect( 10, 60, 70, 10 );
      Pen Color( 4 );
      Rect( 50, 90, 90, 50 );
      Pen Color( 5 );
      Rect( 0, 80, 70, 70 );
   )
);
```

`Rect()` has an optional fifth argument, *fill*. Specify 0 to get an unfilled rectangle and 1 to get a filled rectangle. The default value for *fill* is 0. The rectangle is filled with the current `fill color`.

Any negative fill argument produces an unfilled frame inset by one pixel:

```
win = New Window( "Framed Rectangle",
   Graph Box( Frame Size( 200, 200 ), Rect( 0, 40, 60, 0, -1 ) )
);
```

Ovals

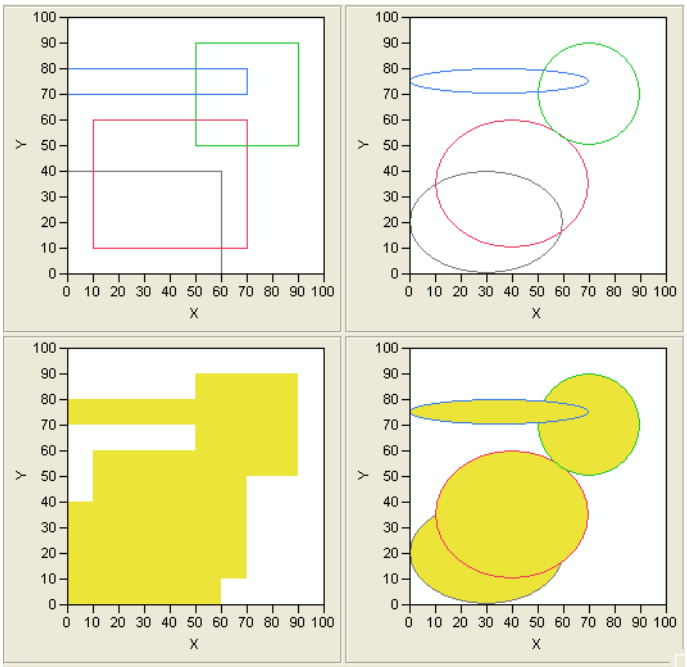
`Oval()` draws an oval inside the rectangle given by its *x1*, *y1*, *x2*, and *y2* arguments:

```
win = New Window( "Ovals",
   Graph Box(
      Frame Size( 200, 200 ),
      Pen Color( 1 );
      Oval( 0, 40, 60, 0 );
      Pen Color( 3 );
      Oval( 10, 60, 70, 10 );
      Pen Color( 4 );
      Oval( 50, 90, 90, 50 );
      Pen Color( 5 );
      Oval( 0, 80, 70, 70 );
   )
);
```

`Oval()` also has an optional fifth argument, *`fill`*. Specify 0 to get an unfilled rectangle and 1 to get a filled oval. The default value for *`fill`* is 0. The oval is filled with the current `fill color`.

Figure 12.16 shows rectangles and ovals, drawn both filled and unfilled. Notice that filled rectangles do not have outlines. Ovals do have outlines. If you want a filled rectangle with an outline, you must draw the filled rectangle and then draw an unfilled rectangle with the same coordinates.

Figure 12.16 Rectangles and Ovals, Unfilled and Filled



Draw Irregular Shapes: Polygons and Contours

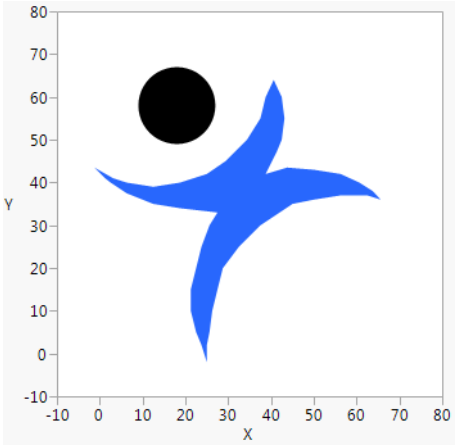
Polygons

`Polygon()` works similarly to `Line()`. They both connect points, but `Polygon()` returns to the first point to close the polygon and filling the resulting area. You can specify the points as individual points in two-item lists (as shown in "Draw Markers") or as matrices of *x* and then *y* coordinates. Matrices are flattened by rows, so you can use either row or column vectors, as long as you have the same number of elements in each matrix. First set up the matrices of points and then call them inside `Polygon()`.

```
gCoordX = [25, 23.75, 22.5, 21.25, 21.25, 22.5, 23.75, 25.625, 27.5, 18.75,
  12.5, 6.25, 2.5,
```

```
1.25, -1.25, 3.125, 6.25, 12.5, 18.75, 25, 29.375, 34.375, 37.5, 38.75,
  40.625, 42.5, 43.125,
42.5, 41.25, 38.75, 43.75, 50, 56.25, 60.625, 63.75, 65.625, 62.5, 56.25, 50,
  45, 37.5, 32.5,
28.75, 27.5, 26.25, 25.625, 25];
gCoordY = [-2, 2, 5, 10, 15, 20, 25, 30, 33, 34, 35, 37.5, 40, 41, 43.5, 41,
  40, 39, 40, 42, 45,
50, 55, 60, 64, 60, 55, 50, 47, 42, 43.5, 43, 42, 40, 38, 36, 37, 37, 36, 35,
  30, 25, 20, 15, 10, 5, 2];
win = New Window( "The JMP Man",
  Graph Box(
     Frame Size( 300, 300 ),
     X Scale( -10, 80 ),
     Y Scale( -10, 80 ),
     Pen Color( "Black" );
     Fill Color( "Blue" );
     Polygon( gCoordX, gCoordY );
     Fill Color( "Black" );
     Circle( {18, 58}, 9, "FILL" );
  )
);
```

Figure 12.17 Drawing a Polygon



A related command, `In Polygon()`, tells whether a given point falls inside the specified polygon. This code checks some points from the JMP man pictured in Figure 12.17:

```
In Polygon( 0,60, GcoordX,GCoordY );  // return 0
In Polygon( 30,38, GcoordX,GCoordY ); // return 1
```

You can add `In Polygon()` to the JMP man script. Run the following script, and then click various locations in the picture and watch the log window. "In" and "out" are printed to the log when you click the body.

```
gCoordX = [25, 23.75, 22.5, 21.25, 21.25, 22.5, 23.75, 25.625, 27.5, 18.75,
   12.5, 6.25, 2.5, 1.25, -1.25, 3.125, 6.25, 12.5, 18.75, 25, 29.375, 34.375,
   37.5, 38.75, 40.625, 42.5, 43.125, 42.5, 41.25, 38.75, 43.75, 50, 56.25,
   60.625, 63.75, 65.625, 62.5, 56.25, 50, 45, 37.5, 32.5, 28.75, 27.5, 26.25,
   25.625, 25];
gCoordY = [-2, 2, 5, 10, 15, 20, 25, 30, 33, 34, 35, 37.5, 40, 41, 43.5, 41,
   40, 39, 40, 42, 45, 50, 55, 60, 64, 60, 55, 50, 47, 42, 43.5, 43, 42, 40,
   38, 36, 37, 37, 36, 35, 30, 25, 20, 15, 10, 5, 2];
win = New Window( "The JMP Man",
   Graph Box(
      Frame Size( 300, 300 ),
      X Scale( -10, 80 ),
      Y Scale( -10, 80 ),
      Pen Color( "Black" );
      Fill Color( "Black" );
      Polygon( gCoordX, gCoordY );
      Mousetrap(
         {},
         Print( If( In Polygon( x, y, gCoordX, gCoordY ), "in", "out" ) )
      );
   )
);
```

Contours

`Contour()` draws contour lines using a grid of coordinates.

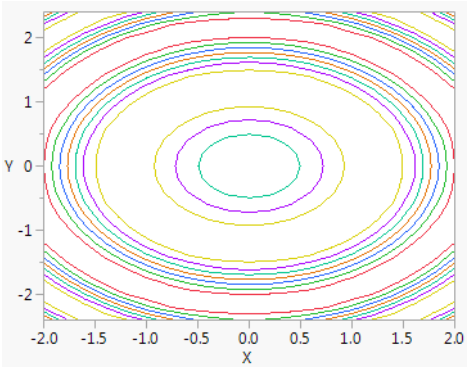
```
Contour( xVector, yVector, zGridMatrix, zContour, <zColors> );
```

Suppose that you have an **n** by **m** matrix *zGridMatrix* of values on some surface (defined across the *n* values of *xVector* by the **m** values of *yVector*). `Contour()` draws the contour lines that are defined by the values in *zContour* in the colors defined by *zColors*. Here is an example:

```
x = (-10 :: 10) / 5;
y = (-12 :: 12) / 5;
grid = J( 21, 25, 0 );
z = [-.75, -.5, -.25, 0, .25, .5, .75];
zcolor = [3, 4, 5, 6, 7, 8, 9];
For( i = 1, i <= 21, i++,
   For( j = 1, j <= 25, j++,
      grid[i, j] = Sin( (x[i]) ^ 2 + (y[j]) ^ 2 )
   )
);
```

```
Show( grid );
win = New Window( "Hat",
   Graph Box(
      X Scale( -2, 2 ),
      Y Scale( -2.4, 2.4 ),
      Contour( x, y, grid, z, zcolor )
   )
);
```

Figure 12.18 Drawing Contour Lines



Add Text

You can use `Text()` to draw text at a given location.

```
Text( <properties>, ( {x, y}|{left, bottom, right, top} ), "text" );
```

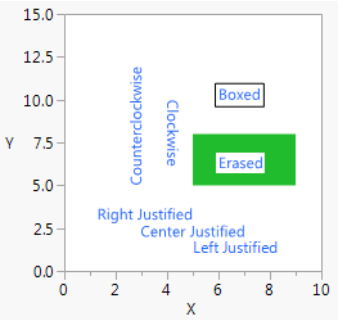
The point and text can be in any order and repeated. You can precede the point and text with an optional first argument: `Center Justified`, `Right Justified`, `Erased`, `Boxed`, `Counterclockwise`, or `Clockwise`. `Erased` is for removing whatever would otherwise obscure the text in a graph. It paints a background-colored rectangle behind the text.

In the following example, notice how the erased text appears inside a white box over the green `Rect()`.

```
mytext = New Window( "Adding Text",
   Graph Box(
      Frame Size( 200, 200 ),
      Y Scale( 0, 15 ),
      X Scale( 0, 10 ),
      Text Size( 9 );
      Text Color( "blue" );
      Text( {5, 1}, "Left Justified" );
      Text( Center Justified, {5, 2}, "Center Justified" );
      Text( Right Justified, {5, 3}, "Right Justified" );
```

```
        Fill Color( 4 );
        Rect( 5, 8, 9, 5, 1 );
        Text( Erased, {6, 6}, "Erased" );
        Text( Boxed, {6, 10}, "Boxed" );
        Text( Clockwise, {4, 10}, "Clockwise" );
        Text( Counterclockwise, {3, 5}, "Counterclockwise" );
    )
);
```

Figure 12.19 Drawing Text in a Graph Box



There is a variant of the `Text()` function that draws a string inside the rectangle that is specified by four coordinates. Here is the syntax:

```
Text( {left, top, right, bottom}, string);
```

Specify Colors

The following commands control colors:

- `Fill Color()` for solid areas
- `Level Color()` for categorical data
- `Pen Color()` for lines and points
- `Back Color()` for the background of text (similar to the box around the erased text in Figure 12.19)
- `Background Color()` for the graph's background color
- `Font Color()` for added text

Fill colors overwrite pen colors for drawn shapes. You do not get both, as in some drawing packages. To get both a fill and a pen line, draw two shapes, one with fill and one without. You can select a color with a single numeric argument, a color name in quotation marks, or an RGB value. The standard colors are specified with numbers 0–15 (both 0 and 15 are black) or by their names (Table 12.1).

Tip: For more shades, add the word Dark, Medium Dark, Medium Light, or Light before a color name in quotation marks.

Table 12.1 Standard JMP Colors

Number	Name
0	Black
1	Gray
2	White
3	Red
4	Green
5	Blue
6	Orange
7	BlueGreen
8	Purple
9	Yellow
10	Cyan
11	Magenta
12	YellowGreen
13	BlueCyan
14	Fuchsia

The following script creates a graph that shows all JMP colors:

```
Text Color( 0 );
New Window( "Colors",
  Graph Box(
     FrameSize( 640, 400 ),
     Y Scale( -1, 17 ),
     X Scale( -3, 12 ),
     k = 0;
     For( jj = 1, jj <= 12, jj += 2,
        l = 15;
```

```
            For( i = 0, i <= 15 & k < 85, i++,
               thiscolor = Color To RGB( k );
               Fill Color( k );
               thisfill = 1;
               If( thiscolor == {1, 1, 1},
                  Pen Color( 0 );
                  thisfill = 0;
               ,
                  Pen Color( k )
               );
               Rect( jj, l + .5, jj + .5, l, thisfill );
               Text( {jj - 1, l}, "color ", k );
               k++;
               l--;
            );
         );
         jj = -2;
         color = {"Black", "Gray", "White", "Red", "Green", "Blue", "Orange",
      "BlueGreen",
         "Purple", "Yellow", "Cyan", "Magenta", "YellowGreen", "BlueCyan",
      "Fuschia", "Black"};
         For(
            i = 0;
            l = 15;, i <= 15 & l >= 0,
            i++;
            l--;,
            Text( {jj, l}, color[i + 1] )
         );
      ) );
```

Larger numbers cycle through shading variations of the same color sequence. A script demonstrating this appears under "Colors and Markers". Values outside the range 0–84 are not accepted.

Table 12.2 How Numbers Map to Colors

Number	Result
16–31	Medium Dark
32–47	Medium Light
48–63	Dark
64–79	Light

Table 12.2 How Numbers Map to Colors *(Continued)*

Number	Result
80–84	Light, Medium Light, Gray, Medium Dark, Dark

If you prefer to use RGB values, create a list with the fraction for each color in red, green, and blue order.

```
Pen Color( {.38,.84,.67} ); // teal
```

`RGB Color()` and `Color to RGB()` convert color values between JMP color numbers and the red-green-blue system. The following example finds the RGB values for JMP color 3 (red):

```
Color to RGB( 3 );
  {0.941176470588235, 0.196078431372549, 0.274509803921569}
```

Likewise, `HLS Color()` and `Color to HLS()` convert color values between JMP color numbers and the hue-lightness-saturation system.

```
win = New Window( "Color Wheel",
   Graph(
      Frame Size( 200, 200 ),
      For( hue = 0, hue < 360, hue += 30,
         y = 50 - 40 * Cos( hue * 2 * Pi() / 360 );
         x = 50 + 40 * Sin( hue * 2 * Pi() / 360 );
         Fill Color( HLS Color( hue / 360, 0.5, 1 ) );
         Oval( x - 10, y - 10, x + 10, y + 10, 1 );
      )
   )
);
```

Finally, `Heat Color()` returns the JMP color that corresponds to a value in any color theme that is supported by Cell Plot, Treemap, and so on. The syntax is:

```
Heat Color( n, <<"theme" )
```

The theme message is optional, and the default value is `"Blue to Gray to Red"`. You can specify any color theme, including custom color themes. You can also create and use an anonymous color theme as shown in the following examples:

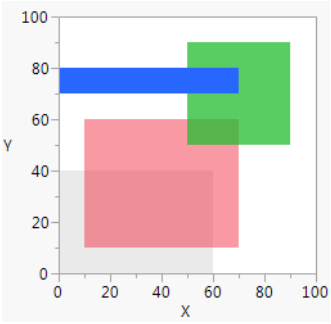
```
Heat Color( z, <<{"", {{1, 1, 0}, {0, 0, 1}} } );
Heat Color( z, <<{"", {blue, green, yellow} } );
```

Specify Transparency

In a graphics environment (like a Frame Box), use the Transparency function to set the level of transparency. The argument, alpha, can be any number between zero and one. The value 0 means clear and drawing has no effect. The value 1 means completely opaque and is the usual drawing mode. Intermediate values build semi-transparent color layers on top of what has already been drawn below it. The following example script illustrates transparency with rectangles.

```
win = New Window( "Transparency",
   Graph Box(
      Frame Size( 200, 200 ),
      Pen Color( "gray" );
      Fill Color( "gray" );
      Transparency( 0.25 );
      Rect( 0, 40, 60, 0, 1 );

      Pen Color( "red" );
      Fill Color( "red" );
      Transparency( 0.5 );
      Rect( 10, 60, 70, 10, 1 );

      Pen Color( "green" );
      Fill Color( "green" );
      Transparency( 0.75 );
      Rect( 50, 90, 90, 50, 1 );

      Pen Color( "blue" );
      Fill Color( "blue" );
      Transparency( 1 );
      Rect( 0, 80, 70, 70, 1 );
   )
);
```

Figure 12.20 Transparency and Rectangles



Add Fill Patterns

`Fill Pattern()` sets the pattern for filled areas.

Matrix Example

Specify a mask (a matrix of values between 0 and 1) or an image. Each value in the matrix creates a pixel.

```
win = New Window( "Example",
   Graph Box(
      Fill Pattern(
         [1 0.5 0 0, 0.5 0 0 1, 0 0 1 0.5, 0 1 0.5 0]
      );
      Polygon( [10 30 90], [88 22 44] );
   )
);
```

Named Pattern Example

Specify a name in quotation marks that corresponds to a pattern shown in Table 12.3.

```
win = New Window( "Named Pattern Example",
   Graph Box(
      Fill Pattern( "vertical light" );
      Polygon( [10 30 90], [88 22 44] );
   )
);
```

Graphic Example

Specify an image in quotation marks. If the image is not installed on the viewer's computer, question marks appear in the shape.

```
win = New Window( "Graphic Example",
   Graph Box(
      Fill Pattern( Open( "$SAMPLE_IMAGES/pi.gif", "gif" ) );
      Polygon( [10 30 90], [88 22 44] );
   )
);
```

Table 12.3 Fill Patterns

Pattern Name	Pattern
left slant light	

Table 12.3 Fill Patterns *(Continued)*

Pattern Name	Pattern
right slant light	
vertical light	
horizontal light	
grid light	
hatch light	
h wave light	
v wave light	
hollow circle	
left slant medium	
right slant medium	
vertical medium	
horizontal medium	
grid medium	
hatch medium	
h wave medium	
v wave medium	
filled circle	

Table 12.3 Fill Patterns *(Continued)*

Pattern Name	Pattern
left slant heavy	
right slant heavy	
vertical heavy	
horizontal heavy	
grid heavy	
hatch heavy	
h wave heavy	
v wave heavy	
diamond	
left slant heavy b	
right slant heavy b	
random	
square	
square offset	
wide	
tall	
checkerboard	

Table 12.3 Fill Patterns *(Continued)*

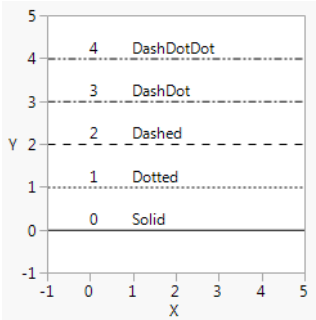
Pattern Name	Pattern
grid dots	
triangle up	
triangle down	
triangle left	
triangle right	
weave light	
weave heavy	
honeycomb	

Specify Line Types

You can also control `Line Style()` by number (0–4) or name (`Solid`, `Dotted`, `Dashed`, `DashDot`, `DashDotDot`). Figure 12.21 shows the numerical values for each line type.

```
linestyles = {"Solid", "Dotted", "Dashed", "DashDot", "DashDotDot"};
win = New Window( "Line Styles",
   Graph Box(
      Frame Size( 200, 200 ),
      X Scale( -1, 5 ),
      Y Scale( -1, 5 ),
      For( i = 0, i < 5, i++,
         Line Style( i );
         H Line( i );
         Text( {0, i + .1}, i );
         Text( {1, i + .1}, linestyles[i + 1] );
      )
   )
);
```

Figure 12.21 Line Styles



To control the thickness of lines, set a `Pen Size` and specify the line width in pixels. The default is 1 for single-pixel lines. For printing, think of `Pen Size` as a multiplier for the default line width, which varies according to your printing device.

```
win = New Window( "Pen Size",
   Graph Box(
      Pen Size( 2 ); // double-width lines
      Line( [10 30 90], [88 22 44] );
   )
);
```

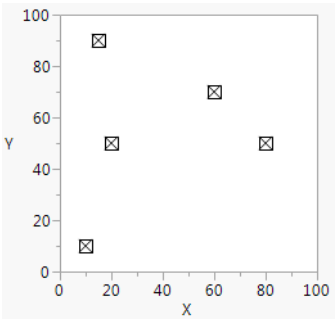
Draw with Pixels

You can also draw using pixel coordinates. First, set the `Pixel Origin()` in terms of graph coordinates, and then use `Pixel Move To()` or `Pixel Line To()` in pixel coordinates relative to that origin. The main use for `Pixel()` functions is for drawing custom markers that do not vary with the size or scale of the graph. You can store a marker in a script and then call it within any graph. This example uses `Function()` to store pixel commands in a script with its own arguments, *x* and *y*.

```
ballotBox = Function( {x, y},
   Pixel Origin( x, y );
   Pixel Move To( -5, -5 );
   Pixel Line To( -5, 5 );
   Pixel Line To( 5, -5 );
   Pixel Line To( -5, -5 );
   Pixel Line To( 5, 5 );
   Pixel Line To( -5, 5 );
   Pixel Move To( 5, 5 );
   Pixel Line To( 5, -5 );
);
win = New Window( "Custom Markers",
   Graph Box(
```

```
        Frame Size( 200, 200 ),
        ballotBox( 10, 10 );
        ballotBox( 15, 90 );
        ballotBox( 20, 50 );
        ballotBox( 80, 50 );
        ballotBox( 60, 70 );
    )
);
```

Figure 12.22 Drawing Custom Markers



Interactive Graphs

`Handle()` and `Mousetrap()` are functions for making interactive graphs that respond to clicking and dragging. `Handle()` lets you parametrize a graph by adding a handle-marker that can be dragged around with the mouse, executing the graph's script at each new location. `Mousetrap()` is similar, but it takes its arguments from the coordinates of a click without dragging a handle. The main difference is that `Handle()` only catches mousedowns at the handle-marker's location, but `Mousetrap()` catches mousedowns at any location.

Another approach is to place buttons or slider controls outside the graph with `Button Box()`, `Slider Box()`, or `Global Box()`.

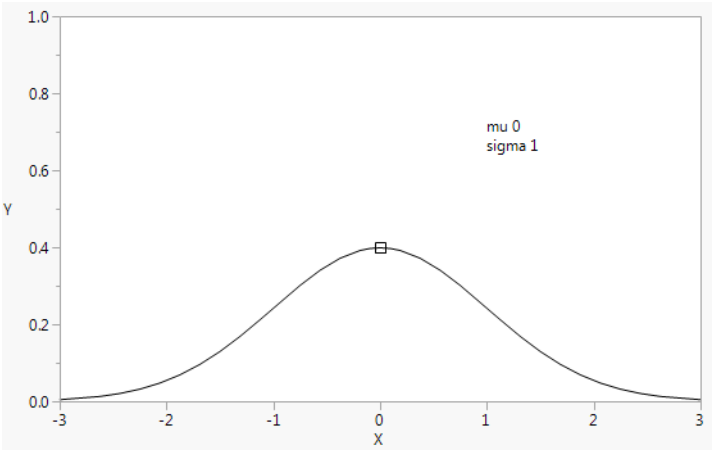
Handle()

`Handle()` places a marker at the coordinates given by the initial values of the first two arguments and draws the graph using the initial values of the arguments. You can then click and drag the marker to move the handle to a new location. The first script is executed at each mousedown to update the graph dynamically, according to the new coordinates of the handle. The second script (optional, and not used here) is executed at each mouseup, similarly; see the example for "Mousetrap()".

```
// Normal Density
mu = 0;
sigma = 1;
rsqrt2pi = 1 / Sqrt( 2 * Pi() );
win = New Window( "Normal Density",
  Graph Box(
     Frame Size( 500, 300 ),
     X Scale( -3, 3 ),
     Y Scale( 0, 1 ),
     Y Function( Normal Density( (x - mu) / sigma ) / sigma, x );
     Handle(
        mu,
        rsqrt2pi / sigma,
        mu = x;
        sigma = rsqrt2pi / y;
     );
     Text( {1, .7}, "mu ", mu, {1, .65}, "sigma ", sigma );
  )
);
```

Run demoPlotProb.jsl in the JMP Samples/Scripts folder to see Beta Density, Gamma Density, Weibull Density, and LogNormal Density graphs. The output for Normal Density is shown in Figure 12.23. Because we cannot show you the picture in motion, be sure to run this script yourself.

Figure 12.23 Normal Density Example for `Handle()`



To avoid errors, be sure to set the initial values of the handle's coordinates, as in the first line of this example.

If you want to use some function of a handle's coordinates, such as in the normal density example, you should adjust the arguments for `Handle()`. Otherwise, the handle marker would run away from the mouse.

```
Y Function( a * x ^ b );
Handle( a, b, a = 2 * x, b = y );
```

Suppose you drag the marker from its initial location to (3,4). The argument *a* is set to 6 and *b* to 4; the graph is redrawn as $Y = 6x^4$; and the handle is now drawn at (6,4), several units away from the mouse. To compensate, you would adjust the first argument to handle, for example.

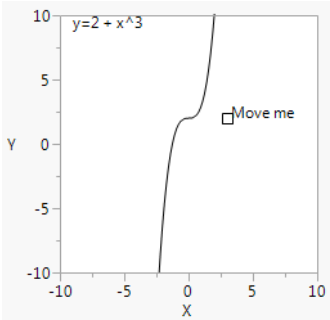
```
Handle( a / 2, b, a = 2 * x; b=y );
```

To generalize, suppose you define the `Handle()` arguments as functions of the handle's (x, y) coordinates. For example, `a=f(x)` and `b=g(y)`. If `f(x)=x` and `g(y)=y`, then you would specify simply `a,b` as the first two arguments. If not, you would solve $a = f(x)$ for x and solve $b = g(y)$ for y to get the appropriate arguments.

You can use other functions to constrain `Handle()`. For example, here is an interactive graph to demonstrate power functions that uses `Round()` to prevent bad exponents and to keep the intercepts simple.

```
a = 3;
b = 2;
win = New Window( "Intercepts and Powers",
   Graph Box(
      Frame Size( 200, 200 ),
      X Scale( -10, 10 ),
      Y Scale( -10, 10 ),
      Y Function( Round( b ) + x ^ (Round( a )), x );
      Handle(
         a,
         b,
         a = x;
         b = y;
      );
      Text( {a, b}, " Move me" );
      Text( {-9, 9}, "y=", Round( b ), " + x^", Round( a ) );
   )
);
```

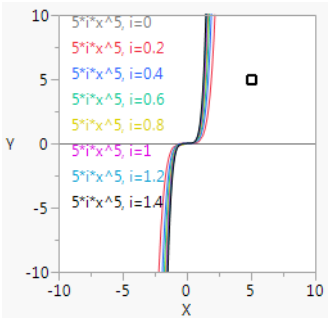
Figure 12.24 Intercepts and Powers for `Handle()`



`Handle()` and `For()` can be nested for complex graphs.

```
a = 5;
b = 5;
win = New Window( "Powers",
   Graph Box(
      Frame Size( 200, 200 ),
      X Scale( -10, 10 ),
      Y Scale( -10, 10 ),
      For( i = 0, i < 1.5, i += .2,
         Pen Color( 1 + 10 * i );
         Text Color( 1 + 10 * i );
         Y Function( i * x ^ Round( a ), x );
         Handle(
            a,
            b,
            a = x;
            b = y;
         );
         h = 9 - 10 * i;
         Text( {-9, h}, b, "*i*x^", Round( a ), ", i=", i );
      ) ) );
```

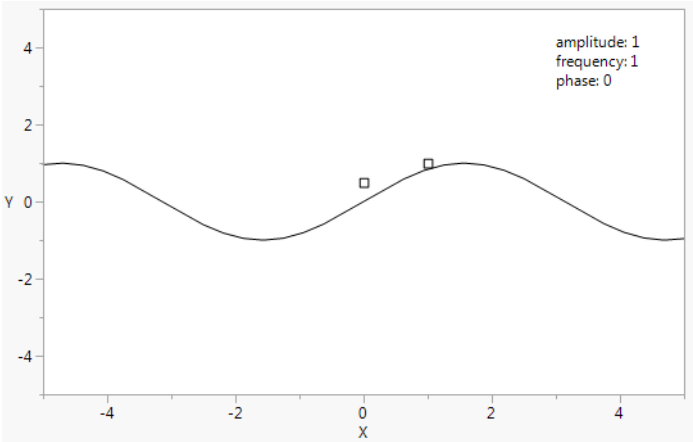
Figure 12.25 Nested Handle() and For()



You can use more than one handle in a graph:

```
amplitude = 1;
freq = 1;
phase = 0;
win = New Window( "Sine Wave",
   Graph Box(
      Frame Size( 500, 300 ),
      X Scale( -5, 5 ),
      Y Scale( -5, 5 ),
      Y Function( amplitude * Sine( x / freq + phase ), x );
      Handle( // first handle
         freq,
         amplitude,
         freq = x;
         amplitude = y;
      );
      Handle( phase, .5, phase = x ); // second handle
      Text(
         {3, 4},
         "amplitude: ",
         Round( amplitude, 4 ),
         {3, 3.5},
         "frequency: ",
         Round( freq, 4 ),
         {3, 3},
         "phase: ",
         Round( phase, 4 )
      );
   )
);
```

Figure 12.26 Two Handles



Mousetrap()

`Mousetrap()` takes arguments for a graph from the coordinates of a mouse click. The first script is executed after each mousedown and the second script after each mouseup to update the graph dynamically, according to the new coordinates of the handle. As with `Handle`, it is important to set the initial values for the `Mousetrap's` coordinates. If you include both `Mousetrap()` and `Handle()` in a graph, put the `Handle()` before the `Mousetrap()`. Then `Handle()` has a chance to catch clicks before a `Mousetrap()` does.

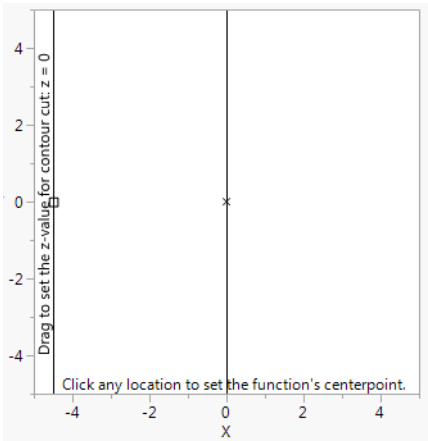
This example uses both `Mousetrap()` and `Handle()` to draw a three-dimensional function centered on the `Mousetrap()` coordinates, where the single contour line takes its value from a `Handle()`.

```
x0 = 0;
y0 = 0;
z0 = 0;
win = New Window( "Viewing a 3-D Function in Flatland",
   Graph Box(
      Frame Size( 300, 300 ),
      X Scale( -5, 5 ),
      Y Scale( -5, 5 ),
      Contour Function(
         Exp( -(x - x0) ^ 2 ) * Exp( -(y - y0) ^ 2 ) * (x - x0),
         x,
         y,
         z0 / 10
      );
```

```
        // get the z-cut values from a handle
        Handle( -4.5, z0, z0 = Round( y * 10 ) / 10 );
              V Line( -4.5 );
        Text Size( 9 );
        Text(
           Counterclockwise,
           {-4.6, -4},
           "Drag to set the z-value for contour cut: z = " || Char( z0 / 10 )
        );
        Marker Size( 2 );
        Marker( 2, {x0, y0} );
        Mousetrap( // set the origin to the click-point
           x0 = x;
           y0 = y;
        );
        Text(
           {-4.25, -4.9},
           "Click any location to set the function's centerpoint."
        );
     )
  );
```

Note: Instead of setting `x0` to `x` and `y0` to `y` in the `Mousetrap()` expression, you might want to use `Function( {xx, yy}, Show( xx, yy ); x0 = xx; y0 = yy );`. `xx` and `yy` are formal parameters to the function. They come from arguments that the mousetrap passed when the user clicked on the graph. The `Function( {xx, yy}...)` alternative makes the `xx` and `yy` names explicit, and makes it clear the you can name them anything, not just `x`, `y`.

Figure 12.27 Mousetrap() and Handle()



You might use Mousetrap() to collect points in a data table, such as for visually interpolating points in a graph. The following example illustrates a script that could be adapted and added to a data plot (such as a scatterplot from Fit Y by X) for that purpose. After you run the script and click in the graph, the point at which you click is recorded in the data table.

```
dt = New Table( "dat1" );
New Column( "xx", Numeric );
New Column( "yy", Numeric );
x = 0;
y = 0;
Add Point = Expr(
   dt << Add Rows( 1 );
   Row() = N Row();
   :xx = x;
   :yy = y;
);
win = New Window( "Add Points",
   Graph Box(
      Frame Size( 500, 300 ),
      X Scale( -5, 5 ),
      Y Scale( -5, 5 ),
      For Each Row( Marker( {xx, yy} ) );
      Mousetrap( {}, Add Point );
   )
);
```

Notice that the first script argument is empty. At mousedown, nothing happens. The second script, add point, is executed at mouseup to add a data point. If you click, drag, and release, then the point that is added to your data set is the point where you let go of the mouse button; it is not the point where you pressed down the mouse button.

Drag Functions

There are five Drag() functions to perform similar functions to Handle() and Mousetrap() but with more than one point at a time. For *n* coordinates in matrices listed as the first two arguments:

- Drag Marker() draws n markers.
- Drag Line() draws a connected line with n vertices and $n-1$ segments.
- Drag Rect() draw a filled rectangle using the first two coordinates, ignoring any further coordinates.
- Drag Polygon() draws a filled polygon with n vertices.

- `Drag Text()` draws a text item at the coordinates. Or if there is a list of text items, it draws the *i*th list item at the *i*th (*x*,*y*) coordinate. If there are fewer list items than coordinate pairs, the last item is repeated for remaining points.

The syntax for these commands:

```
Drag Marker( xMatrix, yMatrix, dragScript, mouseupScript );
Drag Line( xMatrix, yMatrix, dragScript, mouseupScript );
Drag Rect( xMatrix, yMatrix, dragScript, mouseupScript );
Drag Polygon( xMatrix, yMatrix, dragScript, mouseupScript );
Drag Text( xMatrix, yMatrix, "text", dragScript, mouseupScript );
```

They all must have L-value arguments for the coordinates. The coordinates are literal matrices or names of matrix values that are modified if you click a vertex and drag it to a new position. The script arguments are optional and behave the same as with `Handle()`. However, there is no *x* nor *y* that is modified as in `Handle()`.

`Drag()` functions are ways to display data that the user can adjust and then capture the adjusted values. Consider the earlier script to draw the JMP man. `Drag Polygon()` makes it possible to draw an editable JMP man; using a matching `Drag Marker()` statement makes the vertices more visible. And, similar to the `Mousetrap()` example, you can save the new coordinates to a data table. Notice how `:` and `::` operators avoid ambiguity among matrices and data table columns with the same names.

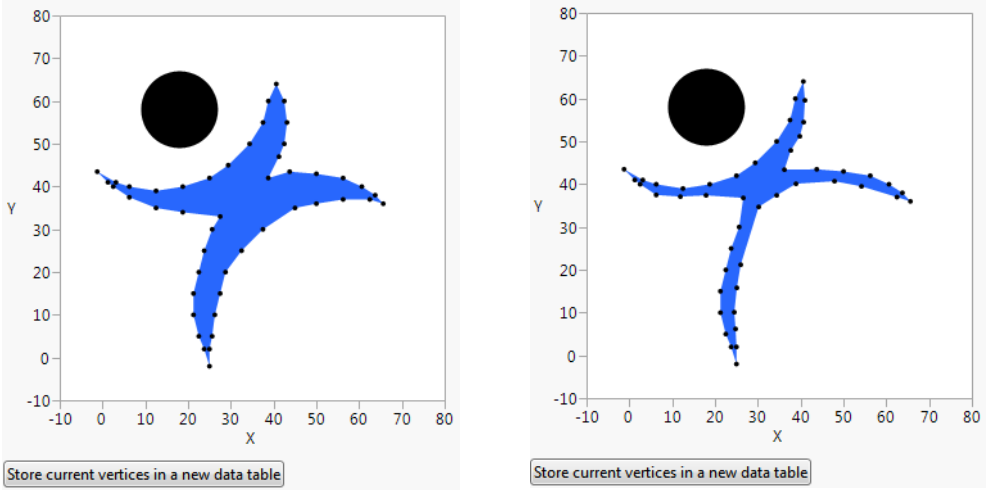
You could just as easily put *`storepoints`* in the fourth argument of `Drag Polygon()` or `Drag Marker()`, but that would create a data table after each drag. You probably just want a single data table when you are finished. Regardless, the values in *`gCoordX`* and *`gCoordY`* update with each drag.

```
::i = 1;
storepoints = Expr(
  mydt = New Table( "My coordinates" || Char( i ) );
  i++;
  New Column( "GCoordX", Numeric );
  New Column( "GCoordY", Numeric );
  mydt << Add Rows( N Row( GcoordX ) );
  :GCoordX << Values( ::GcoordX );
  :GCoordY << Values( ::GcoordY );
);

::GcoordX = [25, 23.75, 22.5, 21.25, 21.25, 22.5, 23.75, 25.625, 27.5, 18.75,
  12.5, 6.25, 2.5, 1.25, -1.25, 3.125, 6.25, 12.5, 18.75, 25, 29.375, 34.375,
  37.5, 38.75, 40.625, 42.5, 43.125, 42.5, 41.25, 38.75, 43.75, 50, 56.25,
  60.625, 63.75, 65.625, 62.5, 56.25, 50, 45, 37.5, 32.5, 28.75, 27.5, 26.25,
  25.625, 25];
::GcoordY = [-2, 2, 5, 10, 15, 20, 25, 30, 33, 34, 35, 37.5, 40, 41, 43.5, 41,
  40, 39, 40, 42, 45, 50, 55, 60, 64, 60, 55, 50, 47, 42, 43.5, 43, 42, 40,
  38, 36, 37, 37, 36, 35, 30, 25, 20, 15, 10, 5, 2];
```

```
win = New Window( "Redraw the JMP Man!",
   V List Box(
      Graph Box(
         Frame Size( 300, 300 ),
         X Scale( -10, 80 ),
         Y Scale( -10, 80 ),
         Fill Color( "blue" );
         Drag Polygon( GcoordX, GCoordY );
         Pen Color( "gray" );
         Drag Marker( GcoordX, GCoordY );
         Fill Color( {0, 0, 0} );
         Circle( {18, 58}, 9, "FILL" );
      ),
      Button Box( "Store current vertices in a new data table", storepoints )
   )
);
```

Suppose you want the body of the JMP man to be smaller. Here is how he looks before and after some judicious vertex-dragging. Clicking the button after re-shaping the JMP Man executes the *storepoints* script to save his new, slender figure in a data table of coordinates.

Figure 12.28 Redraw the JMP Man



This example uses two functions discussed under "Constructors for New Windows":

- `Button Box()`, which creates controls outside the graph.
- `V List Box()`, which combined the graph box and the button box in the same graph window.

Troubleshoot Interactive Graphs

If your interactive graphs do not work as expected, make sure that you supply initial values for the `Handle()` or `Mousetrap()` coordinates (and other globals as needed). Also make sure that the values make sense for the graph.

Create a Color Picker

Let users select a color to apply to graphs by creating a color picker. Users can select a predefined color or create their own color. You can also specify a default color in your script.

- Use the named or numbered colors shown in "Colors and Markers".

  or

- Specify values between 0 and 1 for each of the red, blue, and green components in `RGB Color()`.

Examples of Creating Color Pickers

The following example does not name the window, so the default title is used. Because the script does not specify a color, Black is selected by default. If the user selects a color in the color picker window, then the lines are drawn in the selected color.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Run Script( "Bivariate" );
Wait( 0 );
pickColor = Pick Color();
Report( biv )[AxisBox( 1 )] <<
Add Ref Line( 105, "Dashed", pickColor, "Mean" );
```

The following example creates a color picker with the window title "Select a Line Color". BlueGreen is preselected.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Run Script( "Bivariate" );
Wait( 0 );
pickColor = Pick Color( "Select a Line Color", "BlueGreen" );
Report( biv )[AxisBox( 1 )] <<
Add Ref Line( 105, "Dashed", pickColor, "Mean" );
```

The following example creates a color picker with the window title "Select a Background Color". BlueGreen is preselected.

```
picked color = Pick Color( "Select a Background Color", "BlueGreen" );
New Window( "Example",
        gb = Graph Box(
```

```
                Frame Size( 300, 300 ),
                Marker( Marker State( 3 ), [11 44 77], [75 25 50] );
                Line( [10 30 70], [88 22 44] );
        )
);
gb[FrameBox(1)] << Background Color ( picked color );
```

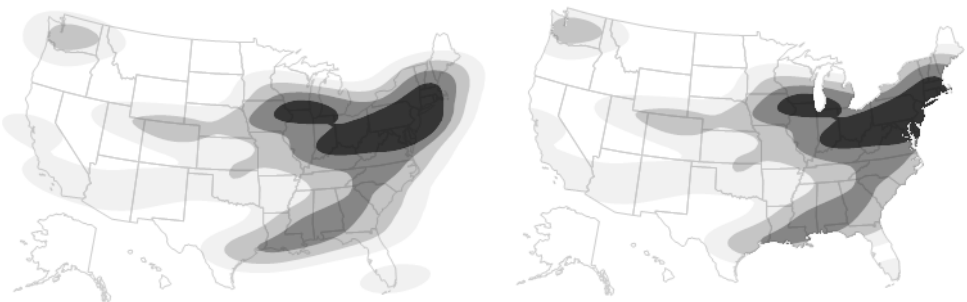
The following example specifies the RGB values for red, green, and blue to create pink. If you do not know the JMP color name or number, you might consider this option.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Run Script( "Bivariate" );
Wait( 0 );
pickColor = Pick Color( "RGB", RGBCOLOR( 1/*red*/, 0/*green*/, 1/*blue*/ ) );
Report( biv )[AxisBox( 1 )] <<
Add Ref Line( 105, "Dashed", pickColor, "Mean" );
```

Clip Display Segs in Graphics

Display segs (such as reference lines, grid lines, and contours) can be clipped to conform to the natural boundaries of the underlying data. This gives you a better view of the boundaries. For example, Figure 12.29 shows contours that are clipped to the boundaries of a map of the United States.

Figure 12.29 Unclipped (Left) and Clipped (Right) Boundaries



Send the `Clip Shape` message to the Contour Seg object to clip the display segs.

```
Open( "$SAMPLE_DATA/Cities.jmp" );
gb = Graph Builder(
   Size( 653, 396 ),
   Show Control Panel( 0 ),
   Variables( X( :Longitude ), Y( :Latitude ) ),
   Elements( Contour( X, Y, Legend( 2 ) ) ),
   SendToReport(
```

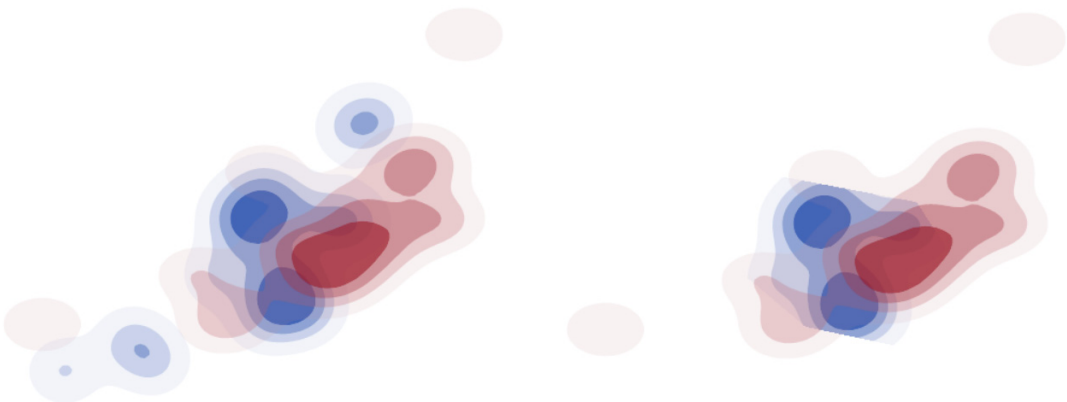
```
      Dispatch(
         {},
         "Graph Builder",
         FrameBox,
         {Background Map( Boundaries( "US States" ) ), Grid Line Order( 2 ),
   Reference Line Order( 3 )}
         )
      )
);
cs = (gb << Report)[FrameBox( 1 )] << Find Seg( Contour Seg( 1 ) );

Wait( 2 );

cs << Clip Shape( Boundaries( "US States" ) );
```

Tip: `Clip Shape(Empty())` removes the clipping path.

Here's an example of defining the clip shape path with a matrix:

```
Open( "$SAMPLE_DATA/Big Class.jmp" );
gb = Graph Builder(
   Show Control Panel( 0 ),
   Variables( X( :height ), Y( :weight ), Overlay( :sex ) ),
   Elements( Contour( X, Y, Legend( 3 ) ) )
);
r = ( gb << Report );
cs = r[FrameBox( 1 )] << Find Seg( Contour Seg( 1 ) );
cs << Clip Shape( Path( [60 80 1, 50 140 2, 65 120 2, 70 65 -2] ) );
```

Figure 12.30 Unclipped (Left) and Clipped (Right) Contours



The matrix [60 80 1, 50 140 2, 65 120 2, 70 65 -2] indicates the following:

1. Start at 60, 80.

2. Draw a line to 50, 140.
3. Draw a line to 65, 120.
4. Draw a line to 70, 65.
5. Connect back to the first point.

Create Background Maps

Background maps can be scripted in JSL. You can write a script that creates a graph and then turns on the background map in the script.

There are two types of background maps: `Images()` and `Boundaries()`. Each of these takes a parameter, which is the name of the map to use. The name is one of the maps listed in the window.

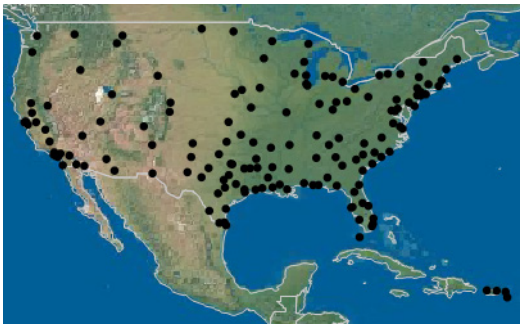
- For `Images()`, the choices are Simple Earth, Detailed Earth, NASA, Street Map Service, and Web Map Service. If you use Web Map Service, then there are two additional parameters: the WMS URL and the layer supported by the WMS server.
- For `Boundaries()`, the choices vary since boundaries can be user-defined. A typical choice is World.

The following example uses the Simple Earth as an image and World as a boundary:

```
dt = Open( "$SAMPLE_DATA/Air Traffic.jmp" );
dt << Graph Builder(
   Size( 1101, 603 ),
   Show Control Panel( 0 ),
   Variables( X( :Longitude ), Y( :Latitude ) ),
   Elements( Points( X, Y, Legend( 8 ) ) ),
   SendToReport(
      Dispatch(
         {},
         "Graph Builder",
         Frame Box,
         {Background Map( Images( "Simple Earth" ), Boundaries( "World" ) ),
         Grid Line Order( 3 ), Reference Line Order( 4 )}
      )
   )
);
```

Figure 12.31 shows an excerpt from the map:

Figure 12.31 JSL Scripting Example



To change the script to use a WMS server, the command would look like this:

```
Background Map ( Images ( "Web Map Service",
  "http://sedac.ciesin.columbia.edu/geoserver/wms",
  "gpw-v3:gpw-v3-population-density_2000" ), Boundaries ( "US States" ) )
```

To see which layers are available on a WMS server, install the WMS Explorer Add-In. Download the add-in from the JMP File Exchange at https://community.jmp.com/docs/DOC-6095. Note that some WMS servers are not reliable. If the server is down, or if the user does not have an Internet connection, the WMS map does not appear.

Tips:

- To see the syntax for the JSL script, add a background map through the user interface. Then from the red triangle menu, select **Save Script > To Script Window** to see the script that is generated.
- To specify image and boundary names in the JSL script, use the names shown in the Set Background Map window. (Right-click a map and select **Graph > Background Map**).
- Look at sample data to find examples of background map scripts. Select **Help > Sample Data Library** and open Napoleons March.jmp, Pollutants Map.jmp, or San Francisco Crime.jmp. Right-click the map table script and select **Edit** to see the `Background Map()` function.

Graph Elements

You can use the following commands inside `Graph Box()` statements. This chapter focuses on the JSL that is specific to graphing, but you can also use general script commands such as `For` and `While`.

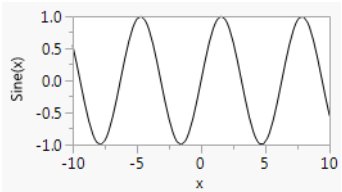
Note: `New Window()` expects window and display box commands, and platforms expect platform commands. Therefore, the general script commands such as `For` and `While` might not work inside platforms or `New Window()`, or you might get unexpected results.

Plot Functions

A `Y Function()` function is used to draw smooth functions. The first argument is the expression to be plotted. The second argument is the name of the X variable in the expression.

```
win = New Window( "Sine Function",
   Graph Box(
      Frame Size( 200, 100 ),
      X Scale( -10, 10 ),
      Y Scale( -1, 1 ),
      X Name( "x" ),
      Y Name( "Sine(x)" ),
      Y Function( Sine( x ), x ) ) );
```

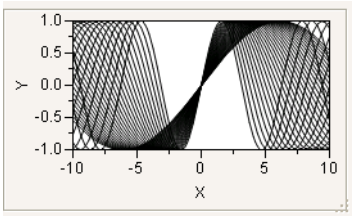
Figure 12.32 Sine Wave



You can use `For` to overlap several sine waves:

```
win = New Window( "Overlapping Sine Waves",
   Graph Box(
      Frame Size( 200, 100 ),
      X Scale( -10, 10 ),
      Y Scale( -1, 1 ),
      For( i = 1, i <= 4, i += .1,
         Y Function( Sine( x / i ), x )
      )
   )
);
```

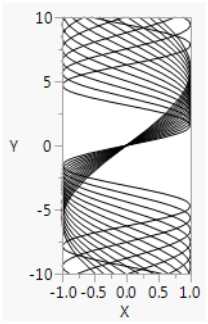
Figure 12.33 Overlapping Sine Waves



Similarly, `X Function()` is for drawing a graph where the symbol is varied on the Y variable.

```
win = New Window( "Overlapping Sine Waves",
   Graph Box(
      Frame Size( 100, 200 ),
      X Scale( -1, 1 ),
      Y Scale( -10, 10 ),
      For( i = 1, i <= 4, i += .2,
         X Function( Sine( y / i ), y )
      )
   )
);
```

Figure 12.34 Overlapping Sine Waves along the X-Axis



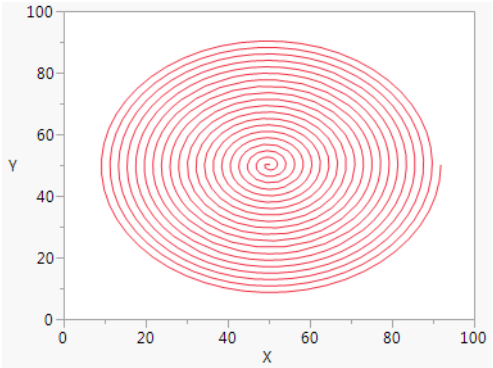
An `XY Function()` draws a smooth curve using a pair of formulas (parametric equations) that depend on a third variable. The third variable's value is incremented from a minimum value to a maximum value to generate the X-Y pairs.

```
win = New Window( "Spiral",
   Graph Box(
      Pen Color( "red" ); // red line color
      xCenter = 50; // location of the center of the curve on the x axis
      yCenter = 50; // location of the center of the curve on the y axis
      minAngle = 0;
      maxAngle = Pi() * 2 * 20;
      XY Function(
```

```
            xCenter + ((ta / 3) * Cos( ta )),
            yCenter + ((ta / 3) * Sin( ta )),
            ta,
            Min( minAngle ),
            Max( maxAngle ),
            Inc( Pi() / 100 )
         );
      )
   );
```

In this example, `Sin()` and `Cos()` use `ta` as an argument (rotates) *and* as a factor (expands). (`Sin()` and `Cos()` use radians, not degrees.)

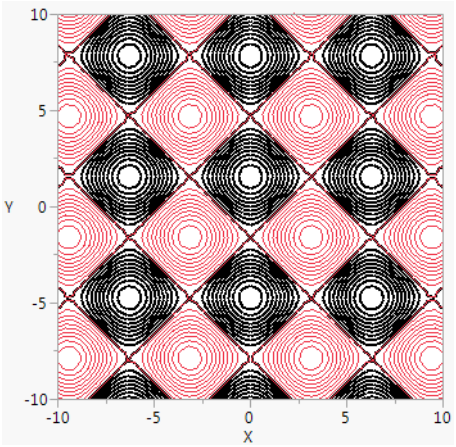
Figure 12.35 Spiral Parametric Plot



`Contour Function()` is an analogous way to represent a three-dimensional function in a two-dimensional space. The final argument specifies the value(s) for the contour line(s). The argument can be a value, an indexed range of values using `::`, or a matrix of values.

```
win = New Window( "Bird's eye view of the egg carton function",
   Graph Box(
      Frame Size( 300, 300 ),
      X Scale( -10, 10 ),
      Y Scale( -10, 10 ),
      Pen Color( "black" );
      Pen Size( 2 );
      Contour Function( Sine( y ) + Cosine( x ), x, y, (0 :: 20) / 5 );
      Pen Color( "red" );
      Pen Size( 1 );
      Contour Function( Sine( y ) + Cosine( x ), x, y, (-20 :: 0) / 5 );
   )
);
```

Figure 12.36 Egg Carton Function



`Normal Contour()` draws normal probability contours for *k* populations and two variables. The first argument is a scalar probability or a matrix of probability values for the contours. Subsequent arguments are matrices to specify means, standard deviations, and correlations. The mean and standard deviation matrices have dimension *k* × 2. The correlation matrix should be *k* × 1, where the first row pertains to the first contour, the second row to the second contour, and so on. The first column is for *x* and the second column for *y*. Consider the following example:

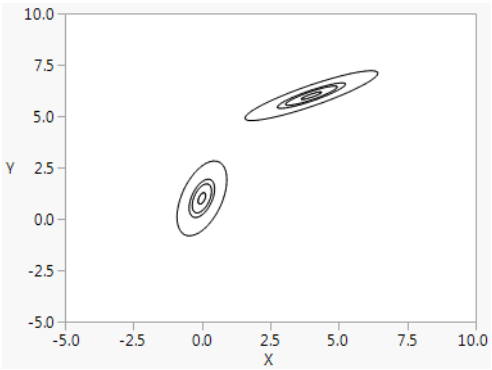
```
Normal Contour(
   [ prob1,
     prob2,
     prob3, ...],
   [ xmean1 ymean1,
     xmean2 ymean2,
     xmean3 ymean3, ...],
   [ xsd1 ysd1,
     xsd2 ysd2,
     xsd3 ysd3, ...],
   [ xycorr1,
     xycorr2,
    xycorr3, ...]);
```

The following script draws contours at probabilities 0.1, 0.5, 0.7, and 0.99 for two populations and two variables. The first population has *x* mean 0 and *y* mean 1, with standard deviation 0.3 along the *x* axis and 0.6 along the *y*-axis, and with correlation 0.5. The second has *x* mean 4 and *y* mean 6, with standard deviation 0.8 along the *x* axis and 0.4 along the *y*-axis, and with correlation 0.9.

```
win = New Window( "Normal Contours",
   Graph Box(
```

```
        X Scale( -5, 10 ),
        Y Scale( -5, 10 ),
        Normal Contour( [.1, .5, .7, .99], [0 1, 4 6], [.3 .6, .8 .4], [.5, .9] )
    )
);
```

Figure 12.37 Normal Contour Function



`Normal Contour()` is a general way to accomplish effects such as Bivariate's density ellipses. The Bivariate script in the Football.jmp sample data creates an example.

Gradient Function

`Gradient Function()` fills a set of rectangles on a grid according to a color determined by the expression value as it crosses a range corresponding to a range of colors.

`Gradient Function( `*`expression`*`, `*`xname`*`, `*`yname`*`, [`*`zlow`*`, `*`zhigh`*`], ZColor( [`*`colorLow`*`, `*`colorHigh`*`] ), <XGrid( `*`min`*`, `*`max`*`, `*`incr`*` )>, <YGrid( `*`min`*`, `*`max`*`, `*`incr`*` )>, <Transparency( t )>;`

Here are explanations of the elements in the `Gradient Function()` syntax:

```
GradientFunction(
    /* the expression to be contoured, which is a function in terms of
    the two variables that follow expression*/
    Expression,

    // the two variable names used in the expression
    xname, yname,

    // the low and high expression values the gradient is scaled between
    [zlow, zhigh],

    // the colors that correspond to the low value and the high value
    ZColor([colorLow, colorHigh])
```

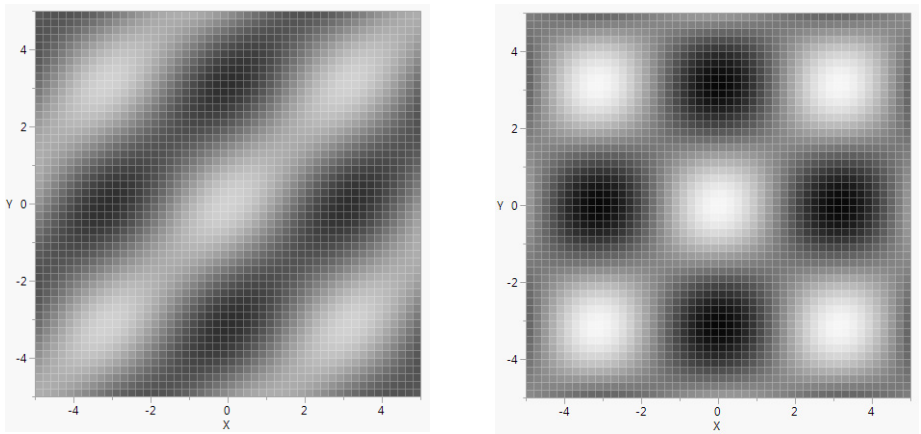
```
// optional specification for the grid of values
<XGrid(min, max, incr),>
<YGrid(min, max, incr)> );

/* the value 0 means clear. The value 1 means completely opaque and
is the usual drawing mode. */
<Transparency>
```

The `ZColor()` values must be numeric codes rather than names. You can use the color menu indices (such as 0=black, 1=grey, 2=white, 3=red, 4=green, and 5=blue) found in "Specify Colors".

The following example uses `Gradient Function()` to create an animated graph.

```
phase = 0.7;
win = New Window( "Gradient Function",
   a = Graph(
      Frame Size( 400, 400 ),
      X Scale( -5, 5 ),
      Y Scale( -5, 5 ),
      Gradient Function(
         phase * Sine( x ) * Sine( y ) + (1 - phase) * Cosine( x ) * Cosine( y
   ),
         x,
         y,
         [-1 1],
         zcolor( [0, 2] )
      )
   )
);
b = a[FrameBox( 1 )];
For( i = 1, i <= 5, i++,
   For( phase = 0, phase < 1, phase += 0.05,
      b << Reshow;
      Wait( 0.01 );
   );
   For( phase = 1, phase > 0, phase -= 0.05,
      b << Reshow;
      Wait( 0.01 );
   );
);
```

Figure 12.38 Gradient Function



Get the Properties of a Graphics Frame

Several functions are useful for getting properties of an existing graphics frame:

H Size Returns the horizontal size of the graphics frame in pixels.

V Size Returns the vertical size of the graphics frame in pixels.

X Origin Returns the *x*-value for the left edge of the graphics frame.

X Range Returns the distance from the left to right edges of a display box.

Y Origin Returns the *y*-value for the bottom edge of the graphics frame.

Y Range Returns the distance from the bottom to top edges of a display box.

In the following expressions, the first line calculates the right edge. The second line calculates the top edge.

```
Oval(
   ...,
   rightEdge = X Origin() + X Range();
   topEdge = Y Origin() + Y Range();
);
```

Add a Legend

To add a legend to a graph, send the `Row Legend` message to the frame box. In the `Row Legend` message, specify which column you want to base the legend on and whether the legend affects colors and markers.

For example, using Big Class.jmp, submit the following JSL to turn on a legend based on the age column. Set both colors and markers by values in the age column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( y( :height ), x( :weight ) );
rbiv = biv << Report;
framebox = rbiv[Frame Box( 1 )];

/* make a legend with the "age" title
display colors and markers in the legend */
framebox << Row Legend( "age", Color( 1 ), Marker( 1 ) );
```

The `Color()` and `Marker()` arguments are optional. Colors are on and markers are off by default.

To use a continuous scale if your variable is nominal or ordinal, you can use a continuous color theme with the `Continuous Scale( 1 )`.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( y( :weight ), x( :height ) );
rbiv = biv << Report;
framebox = rbiv[Frame Box( 1 )];
framebox << Row Legend( "age", Color( 1 ), Continuous Scale( 1 ) );
```

To interactively add a legend, right-click the graph, select **Row Legend**, and change the column settings.

Hover Labels

When you hover over a visual element in a graph, usually you see a basic description of the data point on a hover label.

You can customize what appears on the hover label using the following methods:

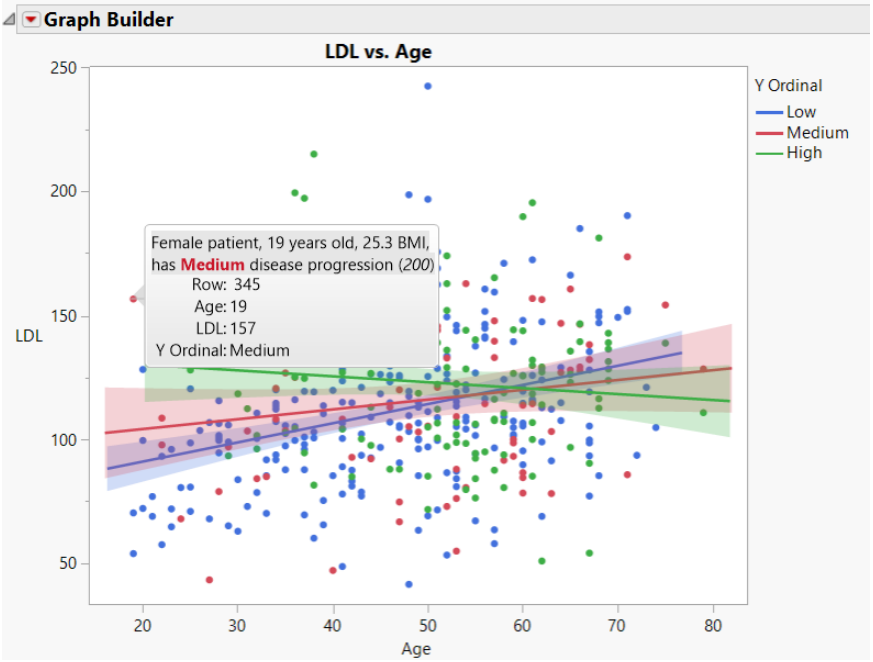
- "Add Rich Text to Hover Labels Using Textlets"
- "Add Graphs or Images to Hover Labels Using Graphlets"
- "Update Names and Values in Hover Labels Using Gridlets"
- "Create Drill-Down Graphs"
- "Work with the Hover Label Execution Context"

Add Rich Text to Hover Labels Using Textlets

A *textlet* adds richly formatted text to the hover label. You might use a textlet to describe data in an explanatory paragraph or to emphasize important information by making it bold or applying a color.

The textlet in Figure 12.39 shows text that is formatted in a gray box with red text.

Figure 12.39 Textlet Example



The following script creates the textlet shown in Figure 12.39.

```
Names Default To Here( 1 );
dt = Open( "$SAMPLE_DATA/Diabetes.jmp" );

// create the baseline visualization
gb = dt << Graph Builder(
   Size( 534, 456 ),
   Show Control Panel( 0 ),
   Variables( X( :Age ), Y( :LDL ), Overlay( :Y Ordinal ) ),
   Elements( Points( X, Y, Legend( 8 ) ), Line Of Fit( X, Y, Legend( 10 ) ) )
);

// get a handle to the graph frame box
rpt = gb << Report;
```

```
frame = rpt[Framebox( 1 )];

/* configure the textlet -- a hover label extension that adds
   a rich text paragraph to the hover label. */
frame << Set Textlet(
    /* "Setup" is a JSL fragment that uses Hover Label Execution Contect
     variables to define content that will be embedded in the
     rich text paragraph.*/
   Setup(
      // map numerical values to text
      local:gender = If( Column( local:_dataTable, "Gender" )[local:_firstRow]
         == "1",
         "Male",
         "Female"
      );
      // map a categorical value to text color
      local:color = Match( local:_Y Ordinal,
         "Low", "Medium Dark Blue",
         "Medium", "Medium Dark Red",
         "High", "Medium Dark Green"
      );
      // capture values from other columns for the current row
      local:progression = Column( local:_dataTable, "Y" )[local:_firstRow];
      local:BMI = Column( local:_dataTable, "BMI" )[local:_firstRow];
   ),
   /* markup defines the rich text template. It supports HTML-like markup such
   as bold, italic and other font tags. Variable references between braces
   ({}) are substituted with their actual value. */
   Markup(
      "{local:gender} patient, {local:_age} years old, {local:BMI} BMI,
      has <font color='{local:color}'><b>{local:_Y Ordinal}</b></font>
      disease progression (<i>{local:progression}</i>)"
   )
);
// pin a hover label to display a textlet-created rich text paragraph.
frame << Add Pin Annotation(
   Seg( Marker Seg( 2 ) ),
   Index( 64 ),
   Index Row( 344 ),
   UniqueID( 568154464 ),
   FoundPt( {96, 222} ),
   Origin( {18.68125, 156.93706122449} ),
   Tag Line( 1 )
);
```

Texlet Commands

See the example in "Add Rich Text to Hover Labels Using Textlets" for more information about writing texlet expressions.

Setup

`Setup()` defines variables used in the rich text that is displayed on the hover label.

```
...
frame << Set Textlet(
    /* "Setup" is a JSL fragment that uses Hover Label Execution Contect
     variables to define content that will be embedded in the
     rich text paragraph.*/
  Setup(
     // map numerical values to text
     local:gender = If( Column( local:_dataTable, "Gender" )[local:_firstRow]
        == "1",
        "Male",
        "Female"
     );
     // map a categorical value to text color
     local:color = Match( local:_Y Ordinal,
        "Low", "Medium Dark Blue",
        "Medium", "Medium Dark Red",
        "High", "Medium Dark Green"
     );
     // capture values from other columns for the current row
     local:progression = Column( local:_dataTable, "Y" )[local:_firstRow];
     local:BMI = Column( local:_dataTable, "BMI" )[local:_firstRow];
  )
...
```

In the interface, Setup is called "JSL Variables" on the Textlets Markup tab.

Markup

`Markup()` specifies a rich text paragraph template based on a subset of HTML tags. Can contain JSL variable references enclosed in delimiters. The default delimiters are braces ({}).

```
...
Markup(
     "{local:gender} patient, {local:_age} years old, {local:BMI} BMI,
     has <font color='{local:color}'><b>{local:_Y Ordinal}</b></font>
     disease progression (<i>{local:progression}</i>)"
  )
```

In the interface, Markup is called "HTML Markup" on the Textlets Markup tab.

See *Using JMP* for more information about the supported font formatting tags.

Delimiters

`Delimiters()` specifies the delimiters that indicate variables for replacement in the Markup area. For example, you might use square bracket delimiters for JMP variables if the script contains C++ code.

```
Delimiters( "[]" )
```

In the interface, Delimiters is on the Textlets Other tab.

Width

`Width()` enables you to specify a preferred width in pixels for the hover label.

```
Width( 300 )
```

In the interface, Width is on the Textlets Other tab.

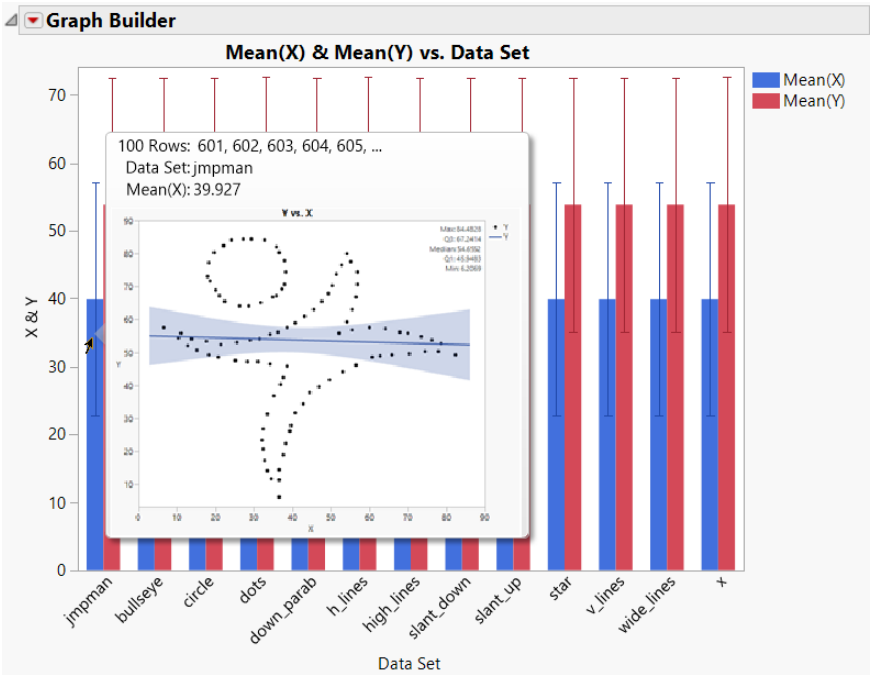
Add Graphs or Images to Hover Labels Using Graphlets

You can add a thumbnail of a related graph or an image to a hover label using a *graphlet*. These graphs and images can have actions associated with them that are triggered by a mouse click. For example, you can click a graph to launch it, or click an image to open a related link.

Example of a Graphlet in Graph Builder

Click the JMP man graph shown in Figure 12.40 to open it in Graph Builder and customize the graph. See *Using JMP* for details about opening a graph in a platform, customizing the graph, and then saving the updated graph in the original graph.

Figure 12.40 Graphlet Example



The following script creates the graph and hover label that are shown in Figure 12.40.

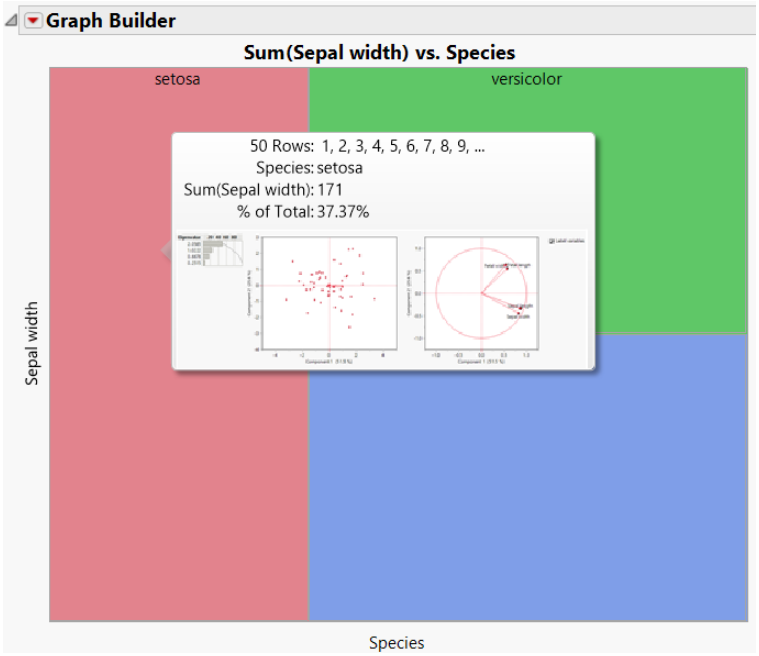
```
Names Default To Here( 1 );
// open the data table
dt = Open( "$SAMPLE_DATA/JMP Man Dozen.jmp" );

// launch the baseline visualization
gb = Graph Builder(
    Size( 529, 466 ),
    Show Control Panel( 0 ),
    Variables( X( :Data Set ), Y( :X ), Y( :Y, Position( 1 ) ) ),
    Elements(
        Bar(
            X,
            Y( 1 ),
            Y( 2 ),
            Legend( 12 ),
            Error Interval( "Standard Deviation" )
        )
    )
);

// get a reference to the visualization frame
```

```
frame = (gb << Report)[FrameBox(1)];

/* use the "Set Graphlet" command to define the visualization
that will be rendered inside the hover label using the
"Picture" attribute */
frame << Set Graphlet(
    /* The value of the "Picture" attribute is a JSL fragment
     that is evaluated to generate a thumbnail. The thumbnail is
     displayed inside the hover label. In this case, the hover
     label shows a scatterplot with a caption box. */
  Picture(
       Graph Builder(
           Show Control Panel( 0 ),
           Variables( X( :X ), Y( :Y ) ),
           Elements(
               Points( X, Y, Legend( 1 ) ),
               Line Of Fit( X, Y, Legend( 3 ) ),
               Caption Box( X, Y, Legend( 4 ),
                Summary Statistic( "5 Number Summary" ) )
           )
       )
    /* Because the optional "Click" attribute is not defined, the
        "Picture" attribute script is re-used to launch the associated
        platform when the user clicks the graphlet thumbnail. */
  )
);
```

Example of a Graphlet in Principal Components

Create a graphlet in the Principal Components platform. When the user clicks the hover label, the Principal Components report window appears. The user can then explore and filter the data.

Figure 12.41 Prinicipal Components Graphlet



The following script creates the graphlet shown in Figure 12.41:

```
dt = Open( "$SAMPLE_DATA/Iris.jmp" );
gb = dt << Graph Builder(
   Size( 531, 451 ),
   Show Control Panel( 0 ),
   Variables( X( :Species ), Y( :Sepal width ) ),
   Elements( Treemap( X, Y, Legend( 4 ) ) ),
   SendToReport(
      Dispatch(
         {},
         "Graph Builder",
         FrameBox,
         {Set Graphlet(
            Picture(
               Principal Components(
                  Y( :Sepal length, :Sepal width, :Petal length, :Petal width
   ),
                  Z( :Species ),
                  Estimation Method( "Default" ),
                  "on Correlations"
               )
            )
         )}
```

```
        )
    )
);
```

Graphlet Commands

Picture

A required script that adds a thumbnail of a graph or an image to the hover label.

`Picture()` is always executed in a Hover Label Execution Context derived from the current visual element. The first `Picture Box` (outside of a local data filter) found will be used to generate the displayed image.

If `Picture()` is a call to a platform, a local data filter is included.

```
...
Picture(
        Graph Builder(
            Show Control Panel( 0 ),
            Variables( X( :X ), Y( :Y ) ),
            Elements(
                Points( X, Y, Legend( 1 ) ),
                Line Of Fit( X, Y, Legend( 3 ) ),
                Caption Box( X, Y, Legend( 4 ),
                 Summary Statistic( "5 Number Summary" ) )
            )
        )...
```

See "Add Graphs or Images to Hover Labels Using Graphlets" for a complete example.

Click

An optional script for user-defined actions to be executed when the user clicks the associated image.

- Clicking runs the main graph's platform (for example, Graph Builder), which launches in its own window.
- If Click is not provided, the script provided in `Picture` is evaluated in a Hover Label Execution Context scope.

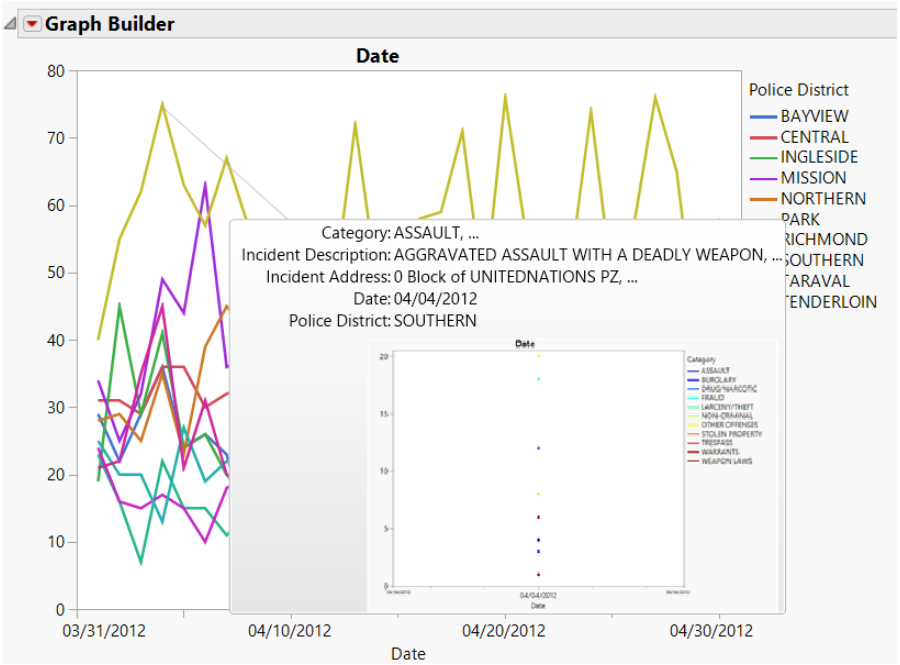
Skip Filters

An optional script. By default, any code executed in the context of a hover label customization (such as a graphlet) executes in a context defined in the hovered visual element (for example, a bar in a bar chart). This context includes a local data filter based on the current visual role assignment, and it is usually what you want. However, in some rare cases, you don't want the filter, or maybe part of the filter, so that the next graph generated by the graphlet can include more of the original data.

- If the Skip Filters field evaluates to a non-0 value (usually a positive integer such as 1) the selected Skip Filters Columns will be removed from the filter that will be applied to the new graphlet.
- Otherwise (for example, if it evaluates to 0), the selected columns are ignored and the filter is created as usual.
- If the Skip Filters Flag field is empty and no columns are selected in Skip Filters Columns (the default), then no columns are skipped when creating filters for the new window. Note that Skip Filters Flag is a JSL expression, which can be as complex as needed.
- If Skip Filters Flag field is not empty, but Skip Filter Columns is empty (no columns are selected), then all columns are filtered; the graphlet or new visualization includes all the data used by the original visualization. See "Example of Defining the Context of a Scatterplot Matrix and Ellipse".

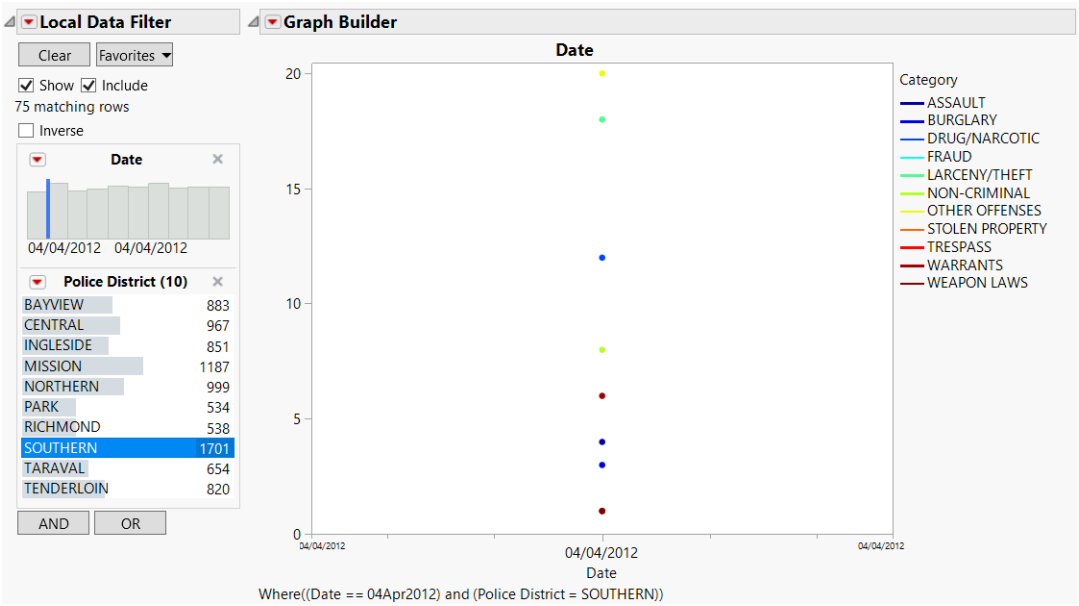
Suppose that you are looking at crime over time in the San Francisco Crime.jmp sample data table. You use a line chart where X = Date, an overlay of Police District, and no Y assignment. Lines represent counts over time per district. Now you want to use a graphlet to look at crime over time inside each district, broken down by category. Create this new graph and select **Save > to Clipboard** and then **Paste Graphlet** on the original graph to create the graphlet. However, the hovered graph shown in Figure 12.42 is not what you expect.

Figure 12.42 Initial Filtered Data



JMP created a filter based on both roles, X and the overlay. This is easier to see if you click on the graphlet to launch the new graph, and then expand the local data filter (on the left, collapsed by default). The filter on a single day causes the line chart to collapse into what looks like a marker plot as shown in Figure 12.43.

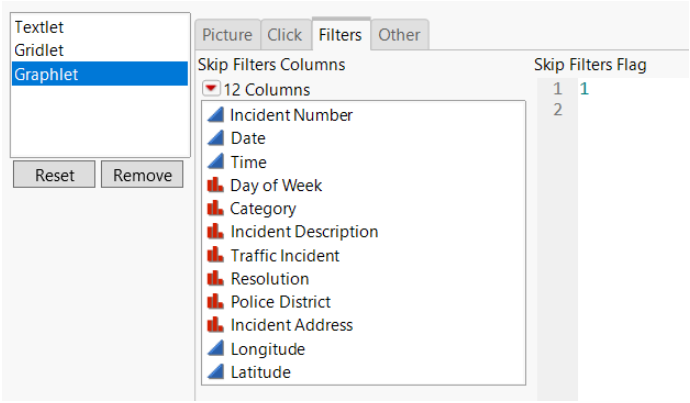
Figure 12.43 Local Data Filter with a Line Chart that Looks like a Marker Plot



You want the overlay value to restrict what you see in the new graph. However, you don't want the date value to do so. You still want to filter by the "SOUTHERN" Police District, but you do not want to filter on "04/14/2012" Date.

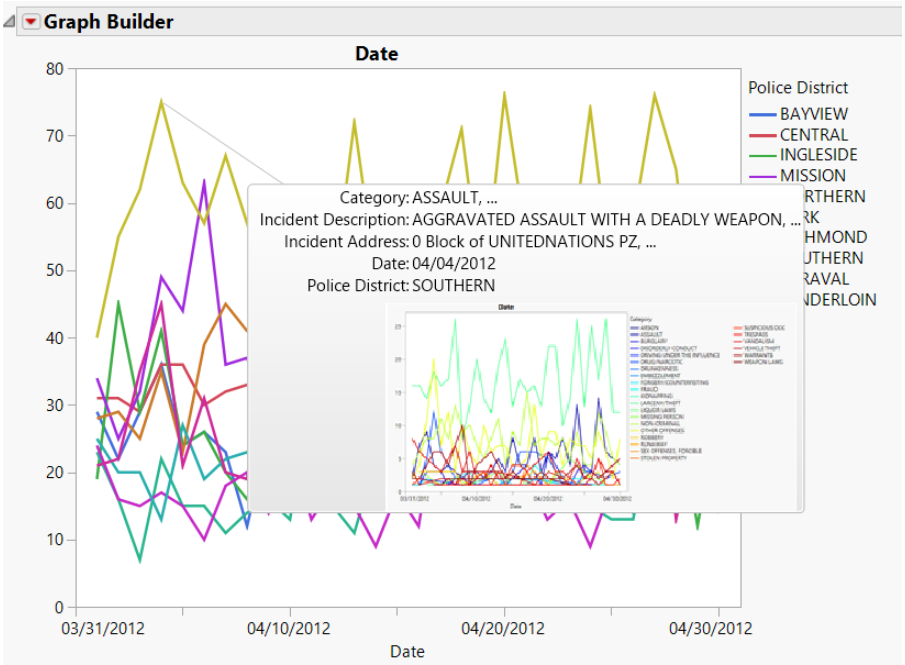
The solution is to use the Skip Filters Flag to remove the filter on Date from your original graph. On the Hover Label Editor, the script looks like this:

Figure 12.44 Example of Skip Filters Flag script in the Hover Label Editor



After the filter is removed, the end result is what you expect as shown in Figure 12.45.

Figure 12.45 Line Chart where Skip Filters Flag = 1



The following script creates the graph shown in Figure 12.45.

```
Names Default To Here( 1 );

dt = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
gb = dt << Graph Builder(
   Size( 528, 456 ),
   Show Control Panel( 0 ),
   Variables( X( :Date ), Overlay( :Police District ) ),
   Elements( Line( X, Legend( 4 ) ) ),
   SendToReport(
      Dispatch(
         {},
         "Graph Builder",
         FrameBox,
         {Set Graphlet(
            Picture(
               Graph Builder(
                  Size( 528, 456 ),
                  Show Control Panel( 0 ),
                  Variables( X( :Date ), Overlay( :Category ) ),
                  Elements( Line( X, Legend( 1 ) ), Smoother( X, Legend( 2 ) ) )
               )
```

```
            ),
            Skip Filters( 1, {:Date} )
          )}
        )
      )
    );
```

The Skip Filters Flag field holds an expression that is evaluated in the hovered visual element context. If expression evaluates to a truthful value (usually a positive integer such as 1), the selected “Skip Filters Columns” will be removed from the filter that will be applied to the new graphlet. Otherwise (for example, if it evaluates to zero), the selected columns are ignored and the filter is created as usual.

Title

Title is an expression that evaluates to a string. The result is the title that appears on the title bar of the window that contains the graph. Variables defined in the Hover Label Evaluation Context can be used in the Title expression.

For example, the following expression indicates the baseline visualization type and how many drilling levels have been used so far:

```
Title("Graphlet for " || local:_displaySegName ||
       ", depth is " || Char( local:_drillDepth ))
```

If the Title expression is not specified, the new window is named “Graphlet”.

Note: In the JMP interface, Title is on the Hover Label Editor Other tab of a graphlet.

Reapply

Reapply is an expression that indicates whether the graphlet definition should be reapplied to new graphs or platforms created by clicking on a graphlet image. This is the core mechanism behind the ability to drill down using graphlets.

```
Reapply( local:_underlyingRows > 1 )
```

Note: In the JMP interface, Reapply is on the Hover Label Editor “Other” tab of a graphlet.

If Reapply is empty (the default setting), the graphlet definition is not reapplied. New windows launched by clicking on this graphlet doesn’t show graphlets on its hover labels.

Update Names and Values in Hover Labels Using Gridlets

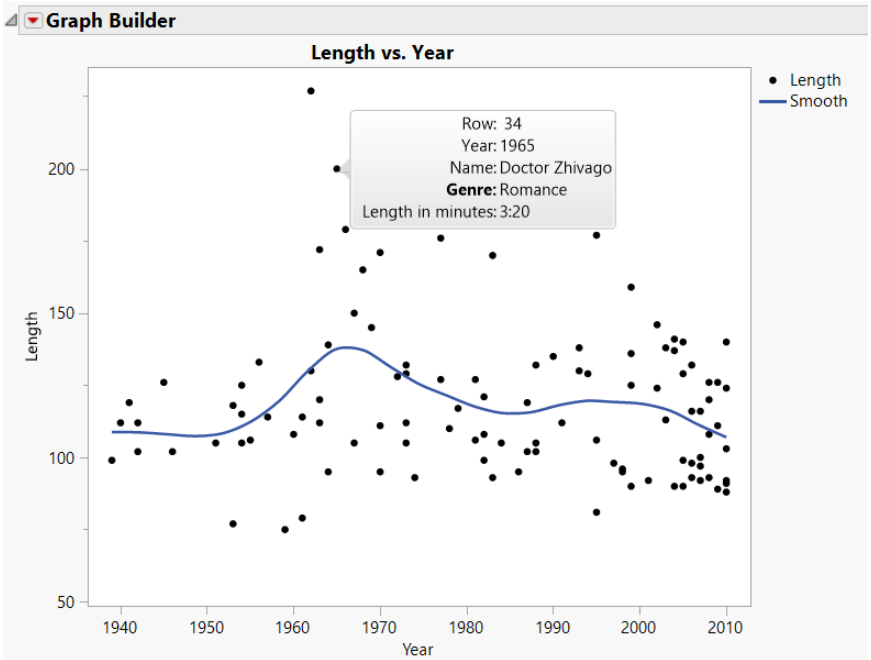
A *gridlet* supports the full customization of the hover label name and value pairs that reflect the underlying visual data. Data can be added, removed, or renamed. You can also change the style and numeric formats.

When you hover over a visualization in a graph, you see content in the hover label for that marker. A grid entry is text that appears on the hover label by default, either from the label columns or from the columns associated with the visualization. By default, the column names that are assigned to roles (and their corresponding values) are shown, along with columns that are labeled. You can customize these column names or values or add other content. Here are a few things you can do:

- add or remove a column name and its value
- rename a grid label
- add a new grid entry with a dynamic value and an optional click action. For example, you can insert an entry named “See also” with a clickable link to a Wikipedia page as the value.
- applying a style to text (for example, formatting text in bold)
- formatting numeric values (for example, setting the number of decimal places to 2)

Figure 12.46 shows adding Genre, Name, and Length in minutes, reformatting the minutes, and formatting “Genre” in bold.

Figure 12.46 Customized Gridlet



Example of Customizing a Gridlet

The following script creates the graph and hover label shown in Figure 12.46.

```
Names Default To Here( 1 );
dt = Open( "$SAMPLE_DATA/Movie Inventory.jmp" );

// launch the baseline visualization
dt << Graph Builder(
   Size( 534, 456 ),
   Show Control Panel( 0 ),
   Variables( X( :Year ), Y( :Length ) ),
   Elements( Points( X, Y, Legend( 3 ) ), Smoother( X, Y, Legend( 4 ) ) ),
   SendToReport(
      Dispatch(
         {},
         "Graph Builder",
         FrameBox,
         {Set Gridlet(
            // delete the Length column
            Expunge( {{Matcher( "Length" )}} ),
            // add the Name, Genre, and Length in minutes columns
            Annex(
               {{Matcher( "Name" ), value( :Name[local:_firstRow] )},
               {Matcher( "Genre" ), value( :Genre[local:_firstRow] )},
               {Matcher( "Length in minutes" ), value( local:_Length * 60 )}}
            ),
            Reformat(
            // reformat the Length in minutes format
               {{Matcher( "Length in minutes" ), Format( "hr:m", 80 ), 80}}
            ),
            Style(
            // format "Genre" in bold nine-point text
               {{Matcher( "Genre", Scope( Name ) ), Font("Segoe UI", 9,
                  "Bold")
               }}
            )
         ), Add Pin Annotation(
            Seg( Marker Seg( 1 ) ),
            Index( 33 ),
            Index Row( 33 ),
            UniqueID( 610643713 ),
            FoundPt( {508, 199} ),
            Origin( {1964.8951552795, 199.00456244898} ),
            Tag Line( 1 )
         )}
      )
```

```
    )
);
```

Gridlet Commands

The first argument in all `Set Gridlet` commands is a matcher, which identifies its scope of action (the grid entries it affects). A matcher can refer to the following:

- A single grid entry: by name, matching against a string. A grid entry is text that appears on the hover label by default, either from the label columns or from the columns associated with the visualization.
- Multiple grid entries: by column reference, regular expression or using the argument `All`.

For example, a grid entry for Mean(height) can be matched against the following:

- the name `"Mean(height)"n`
- the column reference `:height`
- the regular expression `PatReg("^.*\(height\)$")`
- the special argument `All`, which always matches all rows in the grid.

If there is a match, the command-associated action is carried against the matching entries using the arguments.

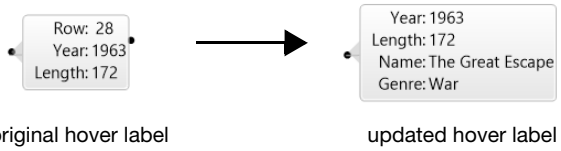
The exception is the `Annex` command, which adds a new entry to the grid if there are no matches with existing entries (to avoid unplanned naming collisions).

Example of Adding and Deleting Columns

This example describes how to use the `Expunge` command to remove the first entry added to the hover label when the underlying table has no Label columns, followed by two instances of the `Annex` command to add new rows. Note that you could have used a single `Annex` instance and passed to it a list of lists instead.

Figure 12.47 shows the original hover label and the updated hover label. Notice that Row was removed from the label. Name and Genre were added.

Figure 12.47 Example of the Expunge and Annex Commands



The following script creates the updated hover label shown in Figure 12.47.

```
Names Default To Here( 1 );
```

```
dt = Open( "$SAMPLE_DATA/Movie Inventory.jmp" );

// create the baseline visualization
gb = dt << Graph Builder(
        Size( 526, 454 ),
        Show Control Panel( 0 ),
        Variables( X( :Year ), Y( :Length ) ),
        Elements( Points( X, Y, Legend( 3 ) ), Smoother( X, Y, Legend( 4 ) ) )
    );

rpt = gb << Report();
box = rpt[Framebox( 1 )];
box << Set Gridlet(
     // delete Row
      Expunge( {{Matcher( 1 )}} ),
      Annex(
        // add the Name and Genre columns to the hover label
        {{Matcher( "Name" ), value( :Name[local:_firstRow] )},
         {Matcher( "Genre" ), value( :Genre[local:_firstRow] )}}
      )
);
```

Annex (Append)

`Annex({`*`matcher`*`, `*`value`*`(<script>), <Click(<script>)>})` adds a new row to the hover label grid. The `Annex` command corresponds to the Append tab in the Hover Label Editor Gridlet panel.

This command takes three arguments:

- `Matcher`: The new name. `Matcher` should not match existing names.
- `Value`: Required. The script is evaluated in the Hover Label Execution Context. The result is the value that appears for the new row. See "Work with the Hover Label Execution Context" for details.

  ```
  box << Set Gridlet(
         Annex( { "LDL, HDL", value(
         Char(:LDL[local:_firstRow]) || ", " || Char(:HDL[local:_firstRow])
         ) })
     );
  ```

- `Click`: Optional. `Click` is evaluated in the Hover Label Execution Context when the user clicks on annexed entry value. This option is equal to the `Click` attribute in graphlets. If present, the value portion of the new entry is rendered using a link style (blue text color with an underline) to indicate that it can be clicked.

Expunge (Delete)

`Expunge({Matcher})` deletes the matcher. The `Expunge` command corresponds to the Delete tab in the Hover Label Editor Gridlet panel.

All matching entries are removed from the grid.

```
box << Set Gridlet( Expunge( {"age"} ) );
```

Rename

`Rename({`*`Matcher`*`, Value(<string>)})` replaces the name portion of the matched entry or entries with the given value. The `Rename` command corresponds to the Rename tab in the Hover Label Editor Gridlet panel.

```
box << Set Gridlet( Rename( {Matcher( "name" ), Value( "students" )} ), );
```

Reformat

`Reformat({`*`Matcher`*`, <`*`format`*`>})` takes a matcher and an argument that specifies a format to be applied to the matched row (or rows) grid numerical value or values. Matching entries are changed to reflect the specified format.

```
box << Set Gridlet( Reformat( {:height, Format( "Scientific", 10, 4 )} ) );
```

Style

`Style({`*`Matcher`*`, <`*`style1`*`>, <`*`style2`*`>, ...})` takes a matcher and one or more style arguments. Here are the supported styles:

- Background Color
- Text Color
- Font (which further supports font family, size and style)
- Justification (left, center, or right)
- Matching entries are changed to reflect the specified styles.

```
box << Set Gridlet(
   Style(
      {{Matcher( "sex" ), Text Color( "Red" ), Font("Papyrus", 14,
      "Bold Underline")}}
   )
);
```

Create Drill-Down Graphs

A drill-down graph is created when you click on the thumbnail. Clicking on the hover graph opens it in a new window, enabling you to see the graph in more detail and also make modifications such as changing the data filter. For example, click a cell on a treemap to see details about the data in a new window (Figure 12.48). Press Ctrl and click to replace the contents of the last launched window instead of opening a new window.

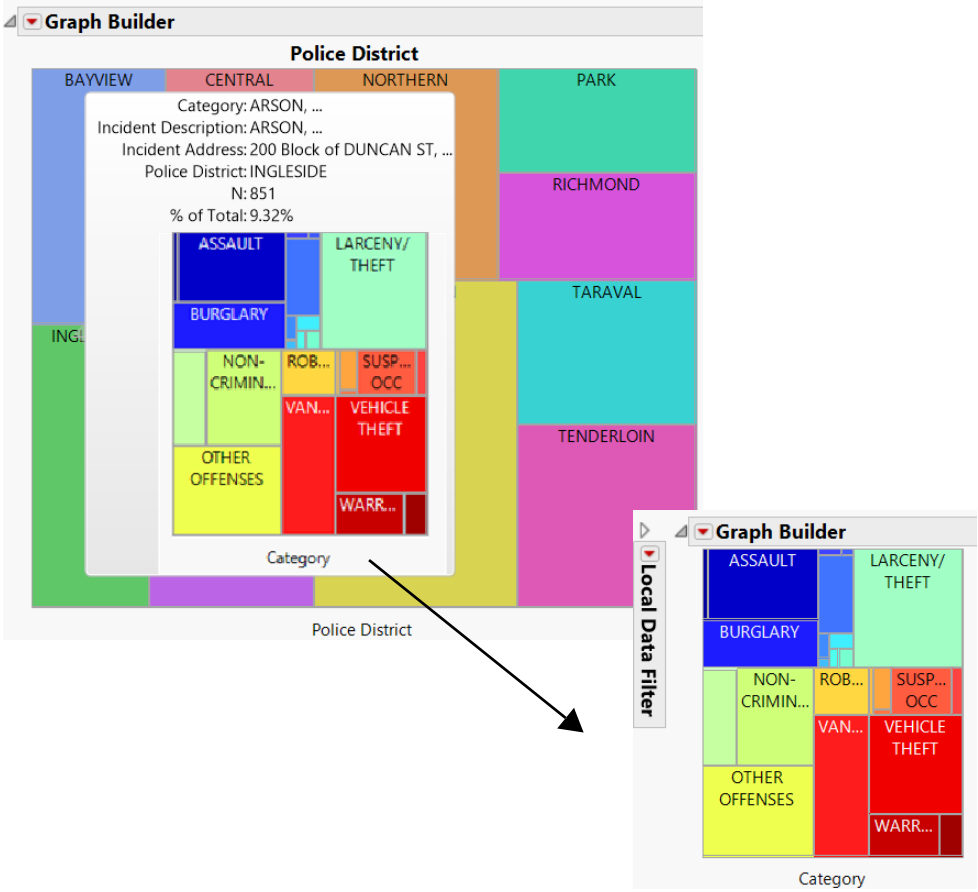
- The Next in Hierarchy column property enables you to define the behavior of categorical drill downs in preset graphs. This column property defines only one level of the drill down, from a category in a grouping role in the current visualization to the category in the grouping role in the next one, launched by the graphlet.
- If the Next in Hierarchy column property is not defined, the list of categorical or nominal columns from the data table that have not been added to a filter is used.

Notes:

- The Next in Hierarchy column property is not supported in graphlets defined using Paste Graphlet. See *Using JMP* for details about the Paste Graphlet command.
- If there is no measurement, JMP tries to get one from the data table and uses Count (or no measurement) if there are no numeric columns in the data table.

Figure 12.48 shows drilling-down on a cell in a treemap.

Figure 12.48 Example of a Drill-Down Graph on a Hover Label



The following script creates the simplified versions of the treemap presets shown in Figure 12.48.

```
Names Default To Here( 1 );

/* traverse the given list of column names, defining the Next in Hierarchy
   column property so that it creates a chain to enable the drill down */
setHierarchy = Function( {dt, cols},
   {Default Local},
   nCols = N Items( cols );
   If( nCols < 2,
      Write( "\!NsetHierarchy: Insufficient number of columns!" ),
      col = Column( dt, cols[1] );
      For( i = 2, i <= nCols, i++,
```

```
            nextCol = cols[i];
            Try(
               Column( dt, nextCol ),
               Write( "Invalid column name: " || nextCol );
               Throw();
            );
            // Eval() is needed for now
            col << Set Property( "Next in Hierarchy", Eval( nextCol ) );
            nih_name = col << Get Property( "Next in Hierarchy" );
            col = Column( dt, nextCol );
         );
      );
   );

   // open San Francisco Data.jmp
   dt = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );

   // add a Next in Hierarchy chain to the data table
   setHierarchy( dt, {"Police District", "Category", "Incident Description"} );

   /* Graph Builder treemap based on the first level of our hierarchy, "Police
      District" */
   gb = dt << Graph Builder(
      Size( 495, 440 ),
      Show Control Panel( 0 ),
      Variables( X( :Police District ) ),
      Elements( Treemap( X ) )
   );

   // get a handle to the graph framebox
   rpt = gb << Report;
   frame = rpt[Framebox( 1 )];

   /* configure the graphlet -- a hover label extension that adds a visualization
      thumbnail to the hover label. Clicking on the thumbnail launches the
      associated visualization in its own window (or replaces the graph in the
      current window on a Ctrl-click). */
   frame << Set Graphlet(
      // define the thumbnail and launch treemap
      Picture(
         // next-level grouping is retrieved from the Next in Hierarchy property
         local:propColName = local:_groupings[1] << GetProperty( "Next In
      Hierarchy" );
         local:nextGrouping = Column( local:_dataTable, local:propColName );

         // define the Graph Builder graph for the thumbnail
```

```
      Graph Builder(
         Size( 200, 250 ),
         Show Legend( 0 ),
         Show Title( 0 ),
         Show Footer( 0 ),
         Show Control Panel( 0 ),
         Variables( X( local:nextGrouping ) ),
         Elements( Treemap( X ) )
      );
   )
);

// add a pinned annotation to display the graphlet
frame << Add Pin Annotation(
   Seg( TreeMapSeg( 1 ) ),
   Index( 2 ),
   Index Row( 7033 ),
   UniqueID( 581248018 ),
   FoundPt( {52, 248} ),
   Origin( {27.6544421487603, 200.14987244898} ),
   Tag Line( 1 )
);
```

Work with the Hover Label Execution Context

Hover Label Execution Context variables define the environment in which your graphlet and textlet scripts are executed.

Local Data Filter

When defining a graphlet, you specify how to visualize the data that appears as a thumbnail. But how do you specify which rows should be displayed?

This information is specified indirectly you hover over different visual elements of the baseline graph (the graph to which a graphlet definition was added). Suppose that the baseline graph is a bar chart. Hovering over one of the bars indicates that all the rows behind that bar (and only those rows) should be included in the generation of the visualization. This visualization shows up as a thumbnail on the hover label. In other words, the hovered visual element provides the data context in which the visualization is generated.

This data context is expressed in a where clause that is computed by JMP based on:

- the type of graph used in the baseline visualization
- the columns assigned to its different roles
- the current values for the particular visual element being inspected

The resulting where clause is then used to create a local data filter that is assigned to the platform specified in the graphlet definition.

The following graph roles are supported as sources for the creation of local data filters:

- X, including nested columns
- Y, if not representing a summary statistic or other calculation
- Group X
- Group Y
- Wrap
- Page
- overlays

Numeric columns are used in comparison to a single value or checked for inclusion in a range, depending on the underlying visualization type. Visualizations that group numeric data in bins such as histograms and heat maps (and the box in box plots) result in the use of corresponding ranges in the where clause.

If present, the where clause of a Local Data Filter associated with the baseline visualization is captured as well.

Local Scope Context

The JSL expressions that configure different aspects of a graphlet and other hover label extensions (such as the textlet `Setup` and the gridlet `Append` and `Click` scripts) are executed inside a local namespace. This namespace is initialized with local variables that reveal information from the execution context: the hover label grid entries and references to the underlying visual element and associated data. These variables are used extensively by the graphlet presets.

Figure 12.4 describes variables that are provided in a graphlet local namespace. See "Complete Example of the Hover Label Execution Context Variables".

Note: To see the variables at work, right-click on a graph, select **Hover Label Editor**, select the Graphlet panel, and then select one of the presets. The code that appears in the Picture field illustrates how many of the Hover Label Excecution Context variables are used to control the graphlet's behavior.

Table 12.4 Hover Label Execution Context Variables

Variable	Description

`_mode`	A string indicating the current evaluation context. Values include `Picture`, `Click`, `Textlet`, `Graphlet`, and `Other`. You can use `_mode` to customize a Graph Builder expression to show more or less information.
	From the JMP log:
	`{"_mode", "Textlet"},`
	Example from the Bar preset:
	`If( local:_mode == "Picture",`
	`  local:sizeDef = "Size( 200, 250 ),";`
	`  local:l = 0;`
	`,`
	`  local:sizeDef = "";`
	`  local:l = 1;`
	`);`
	Note: `Other` means the other attributes that take expressions (for example, `SkipFiltersFlag`).

`_groupings`	A list of the current grouping (X or overlay roles) data column references. See "Example of Defining the Context of a Scatterplot Matrix and Ellipse". `{"_groupings", {:sex}}`
`_measurements`	A list of the current measurements (Y role) data column references. `{"_measurements", {:height}}`
`_filters`	A list of the current local data filter data column references. `{"_filters", {:sex}}`

Table 12.4 Hover Label Execution Context Variables *(Continued)*

Variable	Description

_where	A textual representation of the where clause associated with the hovered visual element. The expression is parsed to be used with JSL statements that take where clauses such as `Get Rows Where`. `_where` is useful to cache complex calculations that are based on context.
	`expr = Parse( local:_where );`
	`r = local:_dataTable << Get Rows Where( expr );`
	`ss = dt << Subset( Private, Invisible, Rows( r ) );`
	`Summarize( ss, exg = By( :sex ), exm = Mean( :height ) );`
	Another example:
	`{"_where", ":sex == \!"F\!""}`

_whereExpr	The where clause that is associated with the hovered visual element as a JSL expression.
	`local:_h = Round(Mean( :height[local:_dataTable << Get Rows Where( local:_whereExpr )] ), 2);`
	Another example:
	`{"_whereExpr", :sex == "F"}`

_displaySegName	The class name of the underlying seg. Useful when creating descriptive text or in the definition of graph-dependent drill-down logic.
	`Title("Graphlet for " || local:_displaySegName || ", depth is " || Char( local:_drillDepth ))`
	Another example:
	`{"_displaySegName", "BarSeg"}`

_dataTable	A reference to the underlying data table. Useful for creating subsets. See "Add Rich Text to Hover Labels Using Textlets" for an example. `{"_dataTable", Data Table( "Big Class" )}`

Table 12.4 Hover Label Execution Context Variables *(Continued)*

Variable	Description
`_summaryStatistic`	A string with the current summary statistic. Useful for passing to new graphs in a drill down. `{"_summaryStatistic", "Mean"}`
`_drillDepth`	An integer that indicates how many drill-down levels have been opened before. Useful to limit drill-down chains or to control `SkipFilters` logic. See "Example of Defining the Context of a Scatterplot Matrix and Ellipse" `{"_drillDepth", 1}`
`_firstRow`	The 1-based index of the first row associated with the hovered visual element. See "Example of Adding and Deleting Columns". `{"_firstRow", 1}`
`_underlyingRows`	The count of rows associated with the visual element that appears in the hover label. You can use this variable to decide the type of the visualization based on the number of rows to display, going from the low granularity of a heat map to a dense marker plot if the number of rows is small enough (and vice versa). See "Reapply" for an example. `{"_underlyingRows", 18}`

In addition to these variables, a variable is defined for each of the hover label grid entries. The name of this variable is taken for the grid entry name, with a '_' prefix.

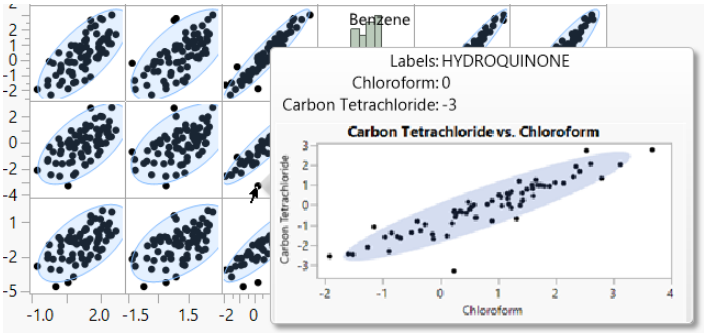
Note: See "Example of Defining the Context of a Scatterplot Matrix and Ellipse" for examples of these variables.

As an example, using Hover Label Execution Context variables, you can create a scatterplot matrix, hover over a point to show a close-up ellipse, and then click the ellipse to modify it in Graph Builder. The graphlet shows the same visualization as the one in the underlying scatterplot matrix cell (a scatterplot of the two columns associated with that cell).

Clicking on the thumbnail then has the effect of "extracting" the cell from the matrix and opening it up as an independent visualization.

The script shown in "Example of Defining the Context of a Scatterplot Matrix and Ellipse" enables the user to extract the matrix cells in the main scatterplot matrix.

Figure 12.49 Viewing an Ellipse in a Scatterplot Matrix



Example of Defining the Context of a Scatterplot Matrix and Ellipse

The following script shows an example of using the context in which the script runs. This script creates the scatterplot matrix shown in Figure 12.49.

```
Names Default To Here( 1 );

dt = Open( "$SAMPLE_DATA/Solubility.jmp" );
win = New Window( "Scatterplot Matrix Graphlet Example",
   plt = dt << Multivariate(
      Y(
         "1-Octanol"n,
         :Ether,
         :Chloroform,
         :Benzene,
         :Carbon Tetrachloride,
         :Hexane
      ),
      Estimation Method( "Row-wise" ),
      Scatterplot Matrix(
         Density Ellipses( 1 ),
         Shaded Ellipses( 1 ),
         Ellipse Color( 69 ),
         Horizontal( 1 )
      )
   )
);
frame = win << xpath( "//FrameBox" );
frame << Set Graphlet(
   Picture(
      Show( local:_displaySegName );
```

```
        If( local:_displaySegName == "MarkerSeg",
           // scatterplot cells
           Graph Builder(
              Size( 400, 200 ),
              Show Control Panel( 0 ),
              Show Legend( 0 ),
              Variables(
                 X( local:_groupings[1] ),
                 Y( local:_groupings[2] )
              ),
              Elements( Points( X, Y ), Ellipse( X, Y ) )
           ),
           // else histograms in matrix diagonal
           If( local:_N > 5, /* Note: only present if baseline graph
              is a histogram.*/
              Graph Builder(
                 Size( 400, 400 ),
                 Show Control Panel( 0 ),
                 Show Legend( 0 ),
                 Variables( X( local:_groupings[1] ) ),
                 Elements( Histogram( X ) )
              ),
              Graph Builder(
                 Size( 400, 400 ),
                 Show Control Panel( 0 ),
                 Show Legend( 0 ),
                 Variables( X( local:_groupings[1] ) ),
                 Elements( Points( X ) )
              )
           )
        );
     ),
     // skip filters if you are at the first drill level -- the original graph
     SkipFilters( local:_drillDepth == 1 ), // no list -- all filters
     Title( "Graphlet" ),
     Reapply( local:_displaySegName == "HistSeg" & local:_N > 5 )
  );
```

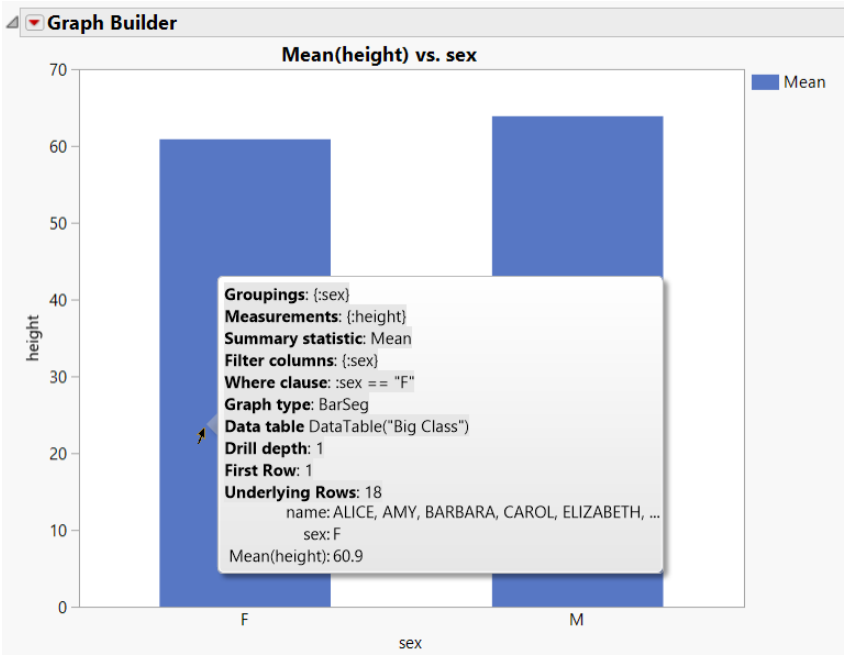
Complete Example of the Hover Label Execution Context Variables

The following script produces the graph shown in Figure 12.50. When you hover over a bar, the hover label appears. Details about the data and the Hover Label Context Execution variables appear in the log.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Graph Builder(
   Size( 528, 450 ),
```

```
   Show Control Panel( 0 ),
   Variables( X( :sex ), Y( :height ) ),
   Elements( Bar( X, Y, Legend( 3 ) ) ),
   SendToReport(
      Dispatch(
         {},
         "Graph Builder",
         FrameBox,
         {Set Textlet(
            Setup(
               local:encodedWhere = XML Encode( local:_where );
               /* Show() provides additional information about the
               HLEC variables in the log */
               Show( Namespace( "local" ) << Get Contents() );
            ),
            Markup(
               "
<b>Groupings</b>: {local:_groupings}
<b>Measurements</b>: {local:_measurements}
<b>Summary statistic</b>: {local:_summaryStatistic}
<b>Filter columns</b>: {local:_filters}
<b>Where clause</b>: {local:encodedWhere}
<b>Graph type</b>: {local:_displaySegName}
<b>Data table</b> {local:_dataTable}
<b>Drill depth</b>: {local:_drillDepth}
<b>First Row</b>: {local:_firstRow}
<b>Underlying Rows</b>: {local:_underlyingRows}
"
            )
         )}
      )
   )
);
```

Figure 12.50 Bar Chart with Textlet



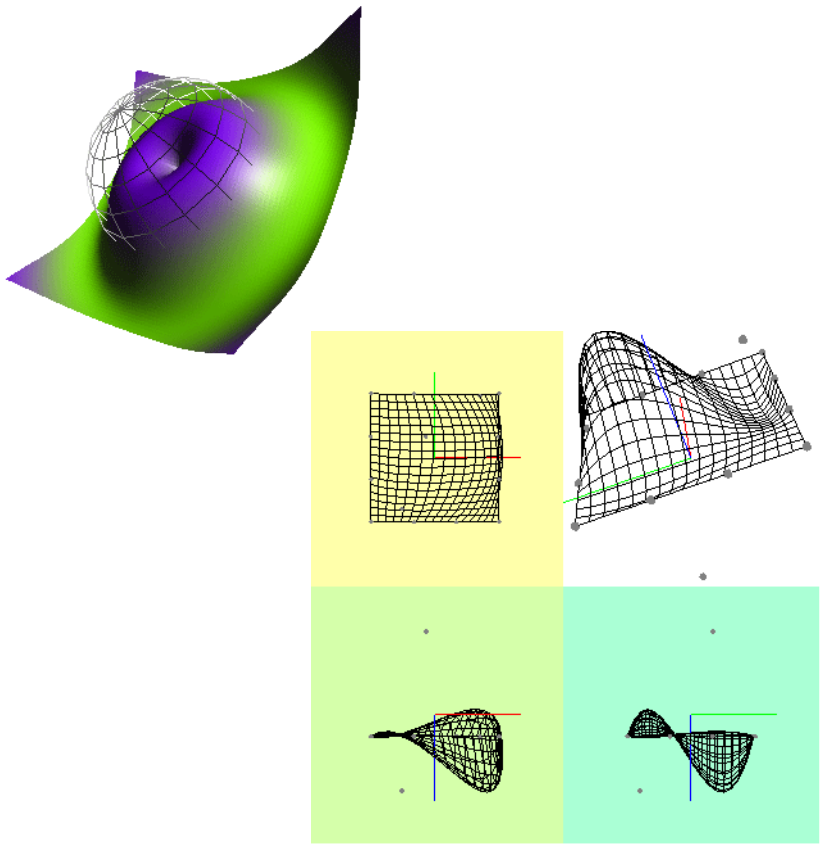
Chapter 13

Three-Dimensional Scenes

Scripting in Three Dimensions

JSL includes commands for scripting three-dimensional scenes derived from OpenGL. Although not a complete OpenGL® implementation, JSL's 3-D scene commands enable complex, interactive plots to be constructed and viewed. The Surface Plot platform in JMP is built using JSL scene commands.

Figure 13.1 Sample Three-Dimensional Shapes



Contents

About JSL 3-D Scenes . . . 751
JSL 3-D Scene Boxes . . . 751
Setting the Viewing Space . . . 754
Setting Up a Perspective Scene . . . 755
Setting up an Orthographic Scene . . . 756
Changing the View . . . 757
The Translate Command . . . 757
The Rotate Command . . . 757
The Look At Command . . . 759
The ArcBall . . . 760
Graphics Primitives . . . 761
Primitives Example . . . 764
Controlling the Appearance of Primitives . . . 765
Other Uses of Begin and End . . . 771
Drawing Spheres, Cylinders, and Disks . . . 771
Construction . . . 771
Lighting . . . 772
Drawing Text . . . 773
Using Text with Rotate and Translate . . . 773
Using the Matrix Stack . . . 774
Lighting and Normals . . . 778
Creating Light Sources . . . 778
Lighting Models . . . 780
Normal Vectors . . . 781
Shading Model . . . 781
Material Properties . . . 782
Alpha Blending . . . 783
Fog . . . 783
Example . . . 783
Bézier Curves . . . 784
One-Dimensional Evaluators . . . 785
Two-Dimensional Evaluators . . . 787
Using the Mouse . . . 788
Pick Commands . . . 789
Three-Dimensional Arguments . . . 790
Improve the Display of Graphics . . . 790

About JSL 3-D Scenes

JMP's 3-D scene language is built on top of the OpenGL API, extending, replacing, and leaving out various parts of the OpenGL API. As such, the 3-D scene language is not certified or licensed by Silicon Graphics, Inc. under the OpenGL API.

This chapter documents JMP's JSL commands for creating 3-D scenes but is not a tutorial on OpenGL programming. If you are not familiar with OpenGL programming, you might want supplemental material. If you are familiar with OpenGL programming, you still need to read this chapter because some items are nonstandard.

JMP is shipped with sample files in the Scene3D subfolder of Sample Scripts to get you started and give you some ideas. Some of the example scripts are similar to some of the examples in this chapter; some are almost complete applications.

The website https://opengl.org is a good source for information, as is your favorite search engine.

JMP's Scene 3-D language does some work for you that the OpenGL API requires you to do for yourself. JMP makes text easy, gives you a built-in arcball controller, and makes sure the matrix operations that belong on the model view stack and projection stack go on their respective stacks. JMP uses its own display list manager so that your scenes can be journaled and played back later, and provides a select mechanism that calls back to your JSL code to tell you what object in your scene is under the mouse, with almost no extra effort on your part. At this time, JMP does not provide access to some features, like texturing.

JSL 3-D Scene Boxes

These commands are necessary to set up and configure a 3-D scene.

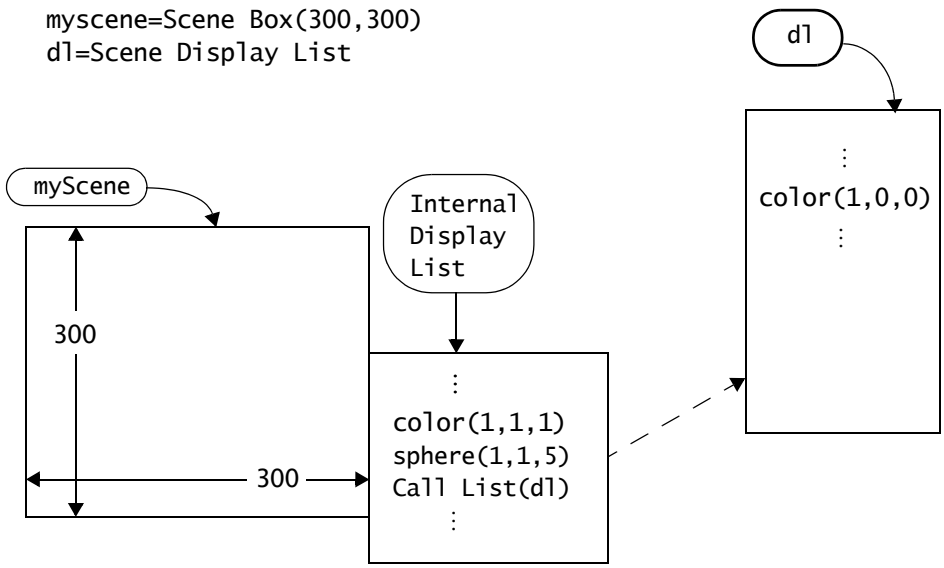
Like all displays in JMP (detailed in "Display Trees"), 3-D scenes must be placed in a display box (in this case, a Scene Box). This box is then placed in a window. Therefore, a simple 3-D scene script has the following form.

```
myScene = Scene Box(300, 300); // create a 300 by 300 pixel scene box
...(commands to set up the scene)...
New Window("3-D Scene", myScene); // draw the scene in a window
...(commands that manipulate the scene)
```

The scene can be sent messages that construct elements in the scene. Typical messages alter the viewer's vantage point, construct physical elements in the scene itself, or manipulate lights and textures. These messages are maintained in a display list and are manipulated in one of two ways:

- They are sent as messages to the scene, which immediately adds them to the scene's internal display list.
- They are sent as messages to a display list stored in a global variable, which is called by the scene's display list later.

Figure 13.2 Constructing Scenes



For example, of commands sent to the scene's display list directly, consider the following small script. Each of the commands is explained in detail later in the chapter.

```
scene = Scene Box( 400, 400 ); // make a scene box
New Window( "Example 1", scene ); // put the scene in a window
scene << Perspective( 45, 3, 7 ); // define the camera
scene << Translate( 0.0, 0.0, -4.5 ); // move to (0,0,-4.5) to draw
scene << Color( 1, 0, 0 ); // set the RGB color of the text
scene << Text( center, baseline, 0.2, "Hello, World." ); // add text
scene << Update; // update the scene
```

The first two lines create a scene and place it in a window. The `Perspective` command defines the viewing angle and field depth. By sending it as a message to the scene, it is immediately added to the scene's display list. Because the "Hello World" text is drawn at the origin (0, 0, 0), the `Translate` command is added to the display list. This command moves the camera back a bit so that the origin is in the field of vision. The color is set to red with the `Color` command, the text is drawn, and the `Update` command causes the scene to be rendered (in other words, causes the display list that contains the commands to be drawn.)

Figure 13.3 Hello World

Equivalently, the commands to construct the display can be accumulated in a display list stored in a global variable, which is then sent to the scene all at once. To define a global variable as a display list, assign it using the `Scene Display List` function. For example, to use the global `greeting` as a display list, issue the command

```
greeting=Scene Display List();
```

Display commands can then be sent as messages to `greeting`. An equivalent "Hello World" example using a display list follows.

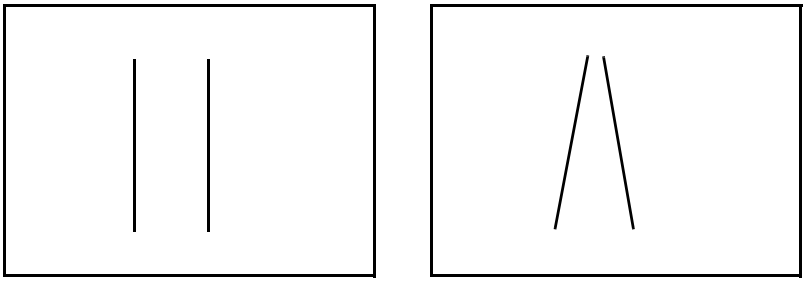
```
greeting = Scene Display List();
greeting << Color( 1, 0, 0 ); // set the RGB color of the text
greeting << Text( center, baseline, 0.2, "Hello, World." ); // add text

// draw the window and send it the stored display list
scene = Scene Box( 400, 400 ); // make a scene box
New Window( "Example 1", scene ); // put the scene in a window
scene << Perspective( 45, 3, 7 ); // define the camera
scene << Translate( 0.0, 0.0, -4.5 ); // move to (0,0,-4.5) to draw
scene << Call List( greeting ); // send the display list to the scene
scene << Update; // update the scene
```

Note which commands were separated into the display list, and which were applied to the scene directly. Those that manipulate the camera (`Translate` and `Rotate`) are applied to the scene. Those that define the object (`Color` and `Text`) were relegated to the display list. This is done so that the display list can be called many times to replicate the object at different positions.

Setting the Viewing Space

3-D scenes can be rendered in two ways. *Orthographic* projections place the elements in a box, where coordinates are not changed to accommodate the perspective of the viewer. *Perspective* projections modify the display to simulate the position of the elements in relation to the position of the viewer. For example, two parallel lines (like railroad tracks) stay parallel in orthographic projections, but seem to connect at a distance in perspective projections.

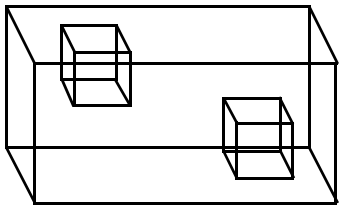
Figure 13.4 Parallel Lines in an Orthographic Projection (left) and a Perspective Projection (right)



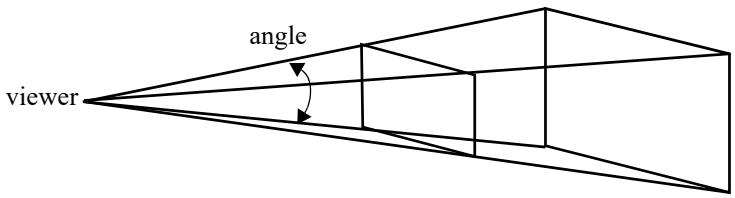
As another example, imagine looking at a tube edge-on (like a telescope). In an orthographic projection, the tube would appear as a thin circle. In a perspective projection, the circle would have a thickness; the hole at the far end of the tube would appear smaller than the close hole, and the interior of the tube is visible.

Therefore, the viewable space of an orthographic projection is a rectangular shape. That of a perspective projection is the frustum of a pyramid (that is, a pyramid whose top has been sliced off).

Figure 13.5 Comparing Projections



Orthographic Scene



Perspective Scene

In general, perspective projections give a more realistic view of the world, since it mimics the way an eye or a camera sees. Orthographic projections are important when it is essential to preserve dimensions, such as an architectural CAD program.

Setting Up a Perspective Scene

To set up a perspective scene in JSL, send the `Perspective` command to a display list.

```
Perspective (angle, near, far)
```

where *`angle`* is the viewing angle, *`near`* is the distance to the near plane, and *`far`* is the distance to the far plane, as illustrated in the drawing above. A couple of things need to be remembered when defining the viewing space.

- Items outside the viewing space (for example, closer than the *`near`* plane or farther than the *`far`* plane) are not drawn. They are clipped off.
- The ratio of *`far`* to *`near`* needs to be small. This enables the rendering engine to effectively determine which items should be drawn "on top of" other items, simulating closeness of items. The *near* argument must be greater than zero.

The "Hello World" example contains the line

```
scene << Perspective( 45, 3, 7 ); // define the camera
```

This defines a 45-degree viewing angle, with a near plane 3 units from the viewer and a far plane 7 units from the viewer.

The viewing angle functions in the same way as a wide angle or telephoto lens on a camera. Small viewing angles zoom into a drawing; wide angles zoom out. In other words, a small viewing angle maps the screen space onto a small portion of the scene, resulting in apparently larger scene elements. A large viewing angle maps the screen space onto a large portion of the scene, resulting in apparently small screen elements. The size of scene elements can therefore be manipulated using the *angle* argument of the `Perspective` function. The picture here shows the hello world script with perspective angles of 45 and 90 degrees.

Figure 13.6 Changing the Perspective

As an alternative to the `Perspective` command, you can define the actual viewing frustum with the `Frustum` command.

```
Frustum(left, right, bottom, top, near, far);
```

The frustum's viewing volume is defined by (*left, bottom, near*) and (*right, top, near*), which specify the (*x, y, z*) coordinates of the lower left and upper right corners of the near clipping plane; *near* and *far* give the distances from the viewpoint to the near and far clipping planes.

Setting up an Orthographic Scene

Orthographic scenes are specified in ways similar to perspective scenes. Issue the command

```
Ortho(left, right, bottom, top, near, far)
```

which specifies the four corners of the near plane, the distance to the near plane, and the distance to the far plane.

If you are dealing with a simple 2-D environment, you can set up a two-dimensional orthographic scene with the command

```
Ortho2D (left, right, bottom, top)
```

which specifies the corners of the two-dimensional view.

Changing the View

One of the advantages of creating a 3-D scene is the ease that they can be viewed from different angles and positions. The `Translate` and `Rotate` commands let you set the position from which you view the scene.

In addition, you can use the `ArcBall` command to enable the user to change the viewing angle interactively.

The Translate Command

You have actually seen the `Translate` command in earlier sample scripts. It sets the position from which the scene is viewed. The arguments give the amount to move from the current position in the *x*, *y*, and *z* direction.

```
Translate (x, y, z)
```

For example,

```
Translate( 0.0, 0.0, -2 );
```

moves the origin two units in the negative *z* direction.

Initially, the origin and camera were at the same place. Now, the camera can see the origin because the camera faces down the negative *z* axis.

The Rotate Command

The `Rotate` command is used to modify the viewing angle of a scene. It has the following format.

```
Rotate (degrees, xAxis, yAxis, zAxis)
```

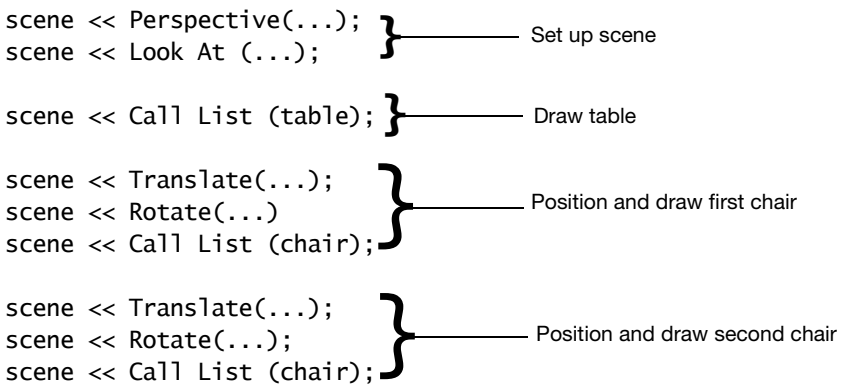
This rotates by degrees around the axis described by the vector (*xAxis*, *yAxis*, *zAxis*). For example, to rotate a model 90 degrees about the *x*-axis, use `Rotate( 90, 1, 0, 0 )`.

You can also specify the three axis values in a matrix. For example, `Rotate( 90, [1, 0, 0] )`.

Note: The `Rotate` command uses degrees, in contrast to JMP's trigonometric functions, which use radians.

`Translate` and `Rotate` are also used to position objects with respect to each other. The first `Translate` or `Rotate` can be thought of as positioning everything that follows with respect to the camera. Subsequent `Translate` and `Rotate` commands are used to position objects, such as spheres, cylinders, disks, and display lists in `Call List` and `ArcBall` commands. For example, suppose you have a display list named `table` and another named `chair`. Your scene might look like this:

Figure 13.7 Using Translate and Rotate

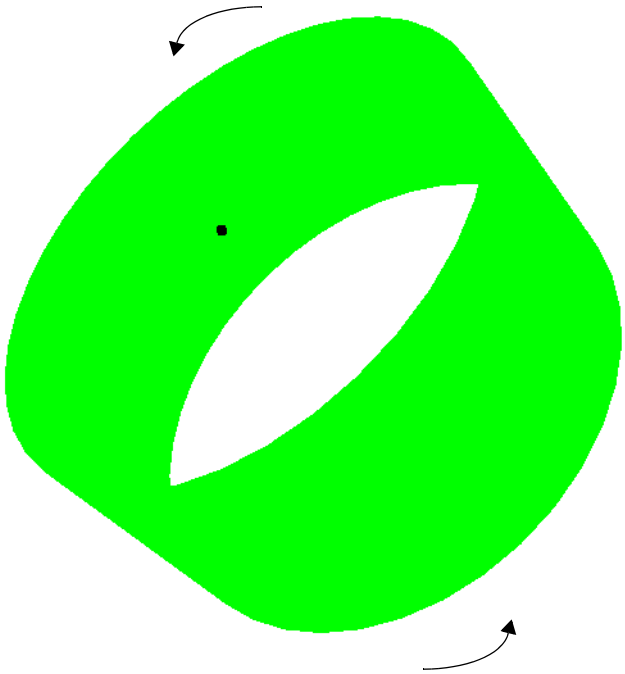


The following example uses the `Rotate` command inside a `For` loop to continuously change the viewing angle of a scene. It draws a cylinder that swings around a central point. This central point is shown by a small sphere.

```
scene = Scene Box( 600, 600 ); // make a scene box...holds an OpenGL scene.

New Window( "Example 1", scene ); // put the scene in a window

For( i = 1, i < 360, i++,
   scene << Clear;

   // the lens is 50 degrees, near is 1 units from the camera, far is 10.
   scene << Perspective( 50, 1, 10 );
   scene << Translate( 0.0, 0.0, -2 );

   scene << Rotate( i, 1, 0, 0 );
   scene << Rotate( i * 3, 0, 1, 0 );
   scene << Rotate( i * 3 / 2, 0, 0, 1 );
```

```
scene << Color( 0, 1, 0 ); // green for cylinder
scene << Cylinder( 0.5, 0.5, 0.5, 40, 10 );
scene << Color( 0, 0, 0 ); // black for sphere
scene << Sphere( 0.01, 10, 5 );
scene << Update;
Wait( 0.01 ); );
```

Figure 13.8 Rotating a Cylinder



Note the use of the `Update` command at the end of the scene messages. This command tells JMP to make the displayed screen agree with the current state of the display list. It is important to clear the list at the beginning (so that the list does not contain the old angles as well as the current) and update the scene after each change.

The Look At Command

The `Look At` command is an alternative way to set the camera view.

`Look At(` *eyeX*, *eyeY*, *eyeZ*, *centerX*, *centerY*, *centerZ*, *upX*, *upY*, *upZ* `)`

The Look At command puts the camera at the *eye* coordinates and points it toward the *center* coordinates. The *up* vector describes how the camera is rotated on its line of sight. The model is typically constructed at the origin. Therefore, a JMP scene should have either a `Look At` or a `Translate` command near its beginning to move the camera away from the origin.

First, clear the scene box of any commands from the previous frame.

```
scene << Clear;
```

Then use one of these projections:

```
scene << Perspective( 45, 2, 10 );
scene << Frustum( -.5, .5, -.5, .5, 1, 10 );
scene << Ortho( -2, 2, -2, 2, 1, 10 );
scene << Ortho2d( -2, 2, -2, 2 );
```

Note: If you use the `ortho2d` projection, you should not also set the camera position using either `Translate` or `Look At`.

Finally, use either `Translate` or `Look At` to set the camera position:

```
scene << Translate(0.0, 0.0, -4.5);
/* the camera faces down the negative Z axis.
   move it back so that 0,0,0 is in view. */
scene << Look At( /*eye*/ 3,3,3, /*center*/ 0,0,0, /*up*/ 1,0,0 );
/*this is much easier. */
```

Once the scene and camera position are set, add your model.

The ArcBall

Sometimes you want a scene to rotate based on the movements of the mouse. The Surface Plot platform in JMP is an example of a 3-D scene that rotates based on mouse movements.

An *ArcBall* creates a sphere around the 3-D scene and enables the user to click the sphere's surface and drag it around, thus causing the scene to rotate.

Use an `ArcBall` instead of a `Call List` command to place the scene in an ArcBall. Scenes that are attached to an ArcBall automatically respond to clicks and drags of the mouse. Custom programming is not needed. However, rotations made in the arcball are not saved. (Technically, the ArcBall is surrounded by an implicit `Push Matrix` and `Pop Matrix` block, so the movements are gone after it returns. See "Using the Matrix Stack" for more information about pushing and popping.)

For example, examine the script from "Primitives Example". Change the single line

```
scene << CallList(shape); // send the display list to the scene
```

so that it reads

```
scene << ArcBall(shape,2); // send the display list to an arcball
```

This displays the script with an associated arcball with diameter 2. When you run the script and the window appears, right-click and select **Show ArcBall > Always** from the menu that appears.

Note: ArcBall comes from an article by Shoemake (1994) found in *Graphics Gems IV*, published by Academic Press.

This sets the display so that the ArcBall is always showing. Click and drag on the ArcBall to rotate the scene. The right-click menu with Background Color, Use Hardware Acceleration, and Show ArcBall is always available, whether the scene is displayed through a platform, in a journal, or through JSL.

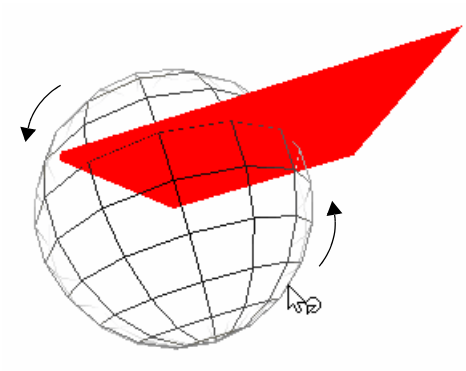
Note: The ArcBall does not have to be showing to react to mouse commands. It is shown here for display purposes only.

You can also set the display state of the ArcBall in JSL using the `Show ArcBall` command.

```
scene << Show Arcball(state)
```

where *state* is `During Drag`, `Always`, or `Never`.

Figure 13.9 Showing the Arc Ball



Graphics Primitives

All scenes in JSL are built with a small number of graphics primitives. These fundamental elements function as the building blocks for complicated scenes.

Every graphics primitive involves specifying vertices. In some cases, the vertices are simply drawn as points. In others, the vertices are connected to form polygons. To draw a primitive, you must specify the type of primitive and the coordinates and properties of the vertices involved. In JSL, this specification is accomplished through the `Begin` and `End` statements.

```
scene<<Begin(primitive type);
...(commands specifying vertices and their properties)...
scene<<End();
```

To specify the coordinates of the vertices, use the `vertex` command.

```
scene<<Begin(primitive type);
scene<<Vertex(x, y, z);
...
scene<<End();
```

The options for *`primitive type`* are the following. In these examples, assume that `v0`, `v1`, and so on, have been specified between a `Begin` and `End` pair, similar to the following.

```
scene<<Begin(primitive type);
scene<<Vertex(x0,y0,z0) // specify vertex v0
scene<<Vertex(x1,y1,z1) // specify vertex v1
...
scene<<Vertex(xn,yn,zn) // specify vertex vn
scene<<End();
```

Table 13.1 Primitive Types

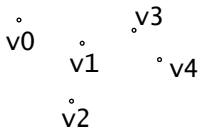
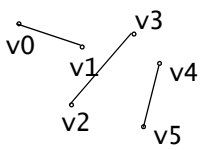
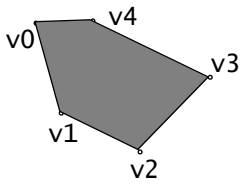
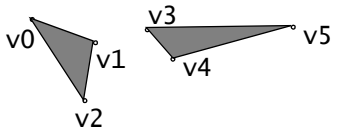
primitive type=POINTS	Draws a point at each of the vertices.	
primitive type=LINES	Draws a series of (unconnected) line segments. Segments are drawn between `v0` and `v1`, between `v2` and `v3`, and so on. If `n` is odd, the last vertex is ignored.	
primitive type=POLYGON	Draws a polygon using the points `v0`,..., `vn` as vertices. Three vertices must exist, or nothing is drawn. In addition, the polygon specified must not intersect itself and must be convex. If the vertices do not satisfy these conditions, the results are unpredictable.	
primitive type=TRIANGLES	Draws a series of (disconnected) triangles using vertices `v0`, `v1`, `v2`, then `v3`, `v4`, `v5`, and so on. If the number of vertices is not an exact multiple of 3, the final one or two vertices are ignored.	

Table 13.1 Primitive Types *(Continued)*

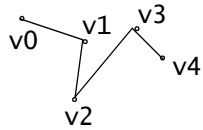
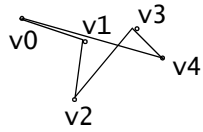
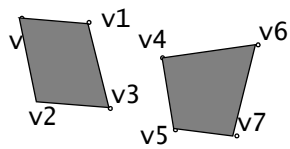
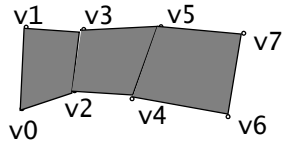
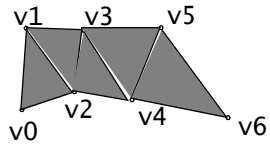
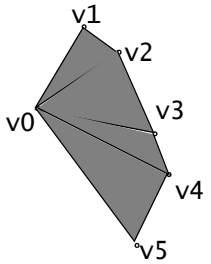
primitive type=LINE_STRIP	Draws a line segment from v0 to v1, then from v1 to v2, and so on. Therefore, n vertices specify n–1 line segments. Nothing is drawn unless there is more than one vertex. There are no restrictions on the vertices describing a line strip; the lines can intersect arbitrarily.	
primitive type=LINE_LOOP	Same as LINE_STRIP, except that a final line segment is drawn from the last vertex to the first, completing a loop.	
primitive type=QUADS	Draws a series of quadrilaterals (four-sided polygons) using vertices v0, v1, v2, v3, then v4, v5, v6, v7, and so on. If the number of vertices is not a multiple of 4, the final one, two, or three vertices are ignored.	
primitive type=QUAD_STRIP	Draws a series of quadrilaterals (four-sided polygons) beginning with v0, v1, v3, v2, then v2, v3, v5, v4, then v4, v5, v7, v6, and so on. The number of vertices must be at least 4 before anything is drawn, and if odd, the final vertex is ignored.	
primitive type=TRIANGLE_STRIP	Draws a series of triangles (three-sided polygons) using vertices v0, v1, v2, then v2, v1, v3 (note the order), then v2, v3, v4, and so on. The ordering is to ensure that the triangles are all drawn with the same orientation so that the strip can correctly form part of a surface. There must be at least three vertices for anything to be drawn.	

Table 13.1 Primitive Types *(Continued)*

primitive type=TRIANGLE_FAN	Same as TRIANGLE_STRIP, except that the vertices are v0, v1, v2, then v0, v2, v3, then v0, v3, v4, and so on.	

Primitives Example

The following short example illustrates the use of a graphics primitive.

```
shape = Scene Display List(); // create a display list and send it commands
shape << Color( 1, 0, 0 ); // set the RGB color of the text
shape << Begin( POLYGON );
shape << Vertex( 0, 0, 0 );
shape << Vertex( 0, 3, 0 );
shape << Vertex( 3, 3, 0 );
shape << Vertex( 5, 2, 0 );
shape << Vertex( 4, 0, 0 );
shape << Vertex( 2, -1, 0 );
shape << End();

// draw the window and send it the stored display list
scene = Scene Box( 400, 400 ); // make a scene box
New Window( "Primitive", scene ); // put the scene in a window
scene << Perspective( 90, 3, 7 ); // define the camera
scene << Translate( 0.0, 0.0, -5 ); // move to (0,0,-5) to draw
scene << Call List( shape ); // send the display list to the scene
scene << Update; // update the scene
```

The first section of the script creates a display list named `shape`. Inside this display list, a polygon is defined using six vertices.

The second section of the script creates a scene box and a new window. It then uses the `Call List` message to put the list in the display.

Note that all the *z*-coordinates are zero, which makes sure the polygon lies in a plane. Polygons that do not lie in a plane can cause unpredictable results.

Experiment with the line

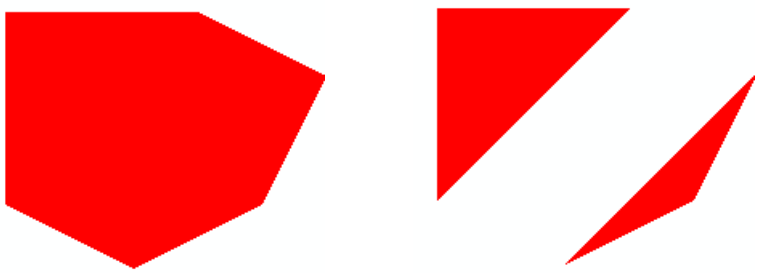
```
shape <<Begin(POLYGON);
```

by changing it to some of the other primitive types. For example, changing it to

```
shape <<Begin(TRIANGLES);
```

results in a different picture.

Figure 13.10 Polygon (left) and Triangles (right)



Controlling the Appearance of Primitives

JSL has several commands that let you tailor-make the appearance of primitive drawing objects. You can also specify the widths of lines and their stippling pattern (that is, whether they are dashed, dotted, and so on.)

Size and Width

To set the point size of rendered objects, use the `Point Size` command.

```
Point Size( n )
```

where *n* is the number of pixels. Note that this might not be the actual number of pixels rendered, depending on other settings such as anti-aliasing and your hardware configuration.

Set the line width using the `Line Width` command

```
Line Width( n)
```

where *n* is the number of pixels. The argument *n* must be larger than zero and is, by default, one.

Stippling Pattern

To make stippled lines, use the `Line Stipple` command.

```
Line Stipple( factor, pattern )
```

Factor is a stretching factor. *Pattern* is a 16-bit integer that turns pixels on or off. Use `Enable(LINE_STIPPLE)` to turn the effect on.

To construct a line stippling pattern, write a 16-bit binary number that represents the stippling pattern that you desire. Note that the pattern should read from right to left, so your representation might seem backward to the way it is rendered. Convert the binary number to an integer and use this as the *`pattern`* argument.

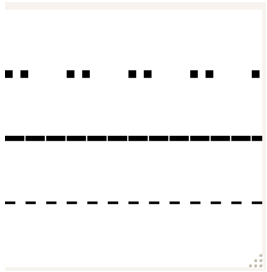
For example, imagine you want the dotted line pattern 0000000011111111. This is equal to 255 in decimal notation, so use the command `Line Stipple(1, 255)`.

The *`factor`* argument expands each binary digit to two digits. In the example above, `Line Stipple(2, 255)` would result in 0000000000000000111111111111111.

For example, the following script draws three lines, each of different widths (the `Line Width` commands) and stippling patterns.

```
scene = Scene Box( 200, 200 ); // make a scene box...holds an OpenGL scene

New Window( "Stipples", scene ); // put the scene in a window
scene << Ortho( -2, 2, -2, 2, -1, 1 );
scene << Color( 0, 0, 0 ); // set the RGB color of the text

scene << Enable( LINE_STIPPLE );
scene << Line Width( 2 );
scene << Line Stipple( 1, 255 );

scene << Begin( LINES );
scene << Vertex( -2, -1, 0 );
scene << Vertex( 2, -1, 0 );
scene << End();

scene << Line Width( 4 );
scene << Line Stipple( 1, 32767 );

scene << Begin( LINES );
scene << Vertex( -2, 0, 0 );
scene << Vertex( 2, 0, 0 );
scene << End();

scene << Line Width( 6 );
scene << Line Stipple( 3, 51 );

scene << Begin( LINES );
scene << Vertex( -2, 1, 0 );
scene << Vertex( 2, 1, 0 );
scene << End();
scene << Update;
```

Figure 13.11 Stipples



Note: Stipple patterns "crawl" on rotating models because they are in screen pixels, not model units. Lines in the model change length on the screen even though nothing changes in model units.

Fill Pattern

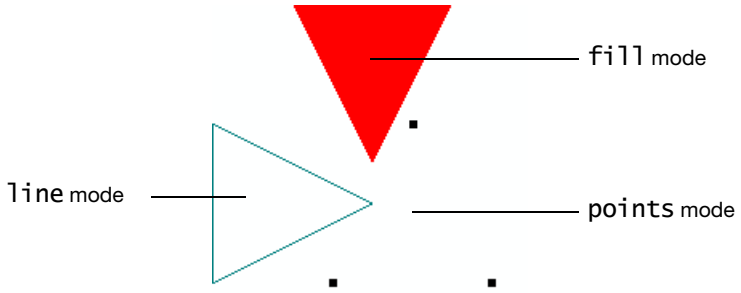
Polygons are rendered with both a front and a back, and the drawing mode of each side is customizable. This enables the user to see the difference between the back and front of the polygon.

To set the drawing mode of a polygon, use the `Polygon Mode` command.

```
Polygon Mode( face, mode )
```

where *`face`* can be `FRONT`, `BACK`, or `FRONT_AND_BACK`, and *`mode`* can be `POINT`, `LINE`, or `FILL`.

Figure 13.12 Points, Line, and Fill Modes



For example, the following script creates a display list that defines a triangle. This display list is used three times in conjunction with `Translate`, `Rotate`, and `Color` commands to draw triangles in three positions. In addition, the `Polygon Mode` command changes the drawing mode of each triangle. Note there is no explicit call to the `FILL` mode, since it is the default.

The following table dissects the script, showing how the `Translate` and `Rotate` commands accumulate to manipulate a single display list.

Table 13.2 Translate and Rotate Commands

Code from above script	comments
`shape = Scene Display List();`	Creates a display list

`shape << Begin( TRIANGLES );`
`shape << Vertex( 0, 0, 0 );`
`shape << Vertex( -1, 2, 0 );`
`shape << Vertex( 1, 2, 0 );`
`shape << End();`
	Creates a display list named shape that holds vertices for the triangles. All the z vertices are zero since this is a two dimensional scene

`scene = Scene Box( 200, 200 );`
`New Window( "Fill Modes", scene );`
`scene << Ortho2d( -2, 2, -2, 2 );`
	Put the scene in a display box, and create a new window.

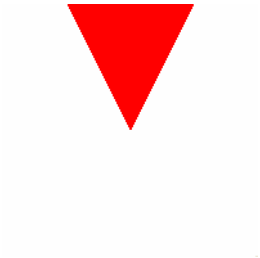
`scene << Color(1,0,0);`
`scene << Call List( shape );`
 	
`// update the scene to see the triangle`
`scene << Update;`
	Draw the first triangle in red.
	 

Table 13.2 Translate and Rotate Commands *(Continued)*

Code from above script	comments

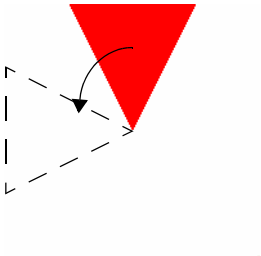
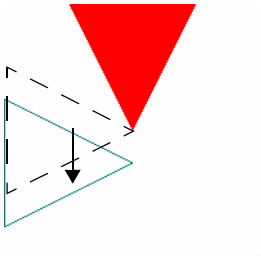
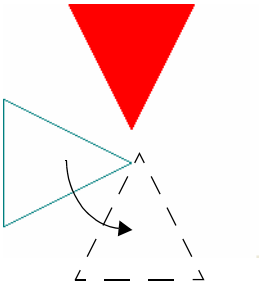
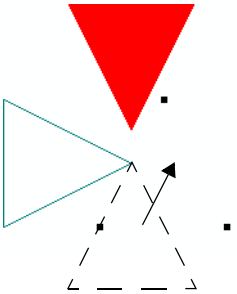
`scene << Rotate ( 90, 0, 0, 1 );`
`scene << Translate ( -0.5, 0, 0 );`
`scene << Color( 0, 0.5, 0.5 );`
`scene << Polygon Mode( FRONT_AND_BACK, LINE );`
`scene << Call List( shape );`
 	
`// update the scene to see the triangle`
`scene << Update;`
	Draw the second triangle in teal. Note that we first rotate the triangle.
	 
	And then translate it.
	 

Table 13.2 Translate and Rotate Commands *(Continued)*

Code from above script	comments

`scene << Rotate( 90, 0, 0, 1 );`
`scene << Translate( -0.5, -1, 0 );`
`scene << Color( 0, 0, 0 );`
 	
`// large points so that they are visible`
`scene << Point Size( 5 );`
`scene << Polygon Mode( FRONT_AND_BACK, POINT );`
`scene << Call List( shape );`
 	
`// update the scene to see the triangle`
`scene << Update;`
	Draw the third triangle as black points. First rotate.
	 
	And then translate to get the final picture.
	 

Some developers use the fill mode in concert with the line mode to draw a filled polygon with a differently colored border. However, due to the way the figures are rendered, they sometimes do not line up correctly. The `Polygon Offset` command is used to correct for this so-called "stitching" problem.

`Polygon Offset (`*`factor`*`, `*`units`*`)`

To enable offsetting, use `Enable(POLYGON_OFFSET_FILL)`, `Enable(POLYGON_OFFSET_LINE)`, or `Enable(POLYGON_OFFSET_POINT)`, depending on the desired mode. The actual offset values are calculated as m*(*`factor`*)+r*(*`units`*). m is the maximum depth slope of the polygon and r is the smallest value guaranteed to produce a resolvable difference in window coordinate depth values. Start with `Polygon Offset(1,1)` if you need this.

An example of `Polygon Offset` is in the Surface Plot platform, when a surface and a mesh are displayed on top of each other, or a surface and contours displayed on top of each other. In either case, the surface would interfere with the lines if the lines were not moved closer or the surface moved farther from the viewer.

Other Uses of Begin and End

Although vertices are typically specified between begin and end statements, there are other commands that are valid. These commands are discussed in other sections of this chapter.

- `Vertex` adds a vertex to the list
- `Color` changes the current color
- `Normal` sets the normal vector coordinates
- `Edge Flag` controls drawing of edges
- `Material` sets material properties
- `Eval Coord` and `Eval Point` generate coordinates
- `Call List` executes a display list.

Drawing Spheres, Cylinders, and Disks

There are several pre-defined commands that allow for quick rendering of spheres, cylinders, and disks. The advantage of these commands is not only their ease-of-use, but that they have special lighting properties (their "normals") built in.

Construction

The following commands are used to construct cylinders, disks, partial disks, and spheres.

Cylinders

```
Cylinder( baseRadius, topRadius, height, slices, stacks )
```

`baseRadius` is the radius of the cylinder's base. Similarly, *`topRadius`* is the radius of the top. *`height`* is the height of the cylinder.

`Slices` can be 10 for a reasonably accurate cylindrical shape. Using `QuadricNormals(Smooth)` helps the appearance.

`Stacks` sets the number of vertices available for lighting reflections. Use a larger value for *`Stacks`* for accurate "hot-spots".

Disks

The following command draws a paper-thin disk with an *`innerRadius`* hole in the middle.

```
Disk( innerRadius, outerRadius, slices, loops )
```

Like `Cylinder`, *`slices`* controls the accuracy of the curve and *`loops`* makes more vertices (for lighting accuracy).

```
Partial Disk( innerRadius, outerRadius, slices, loops, startAngle,
  sweepAngle )
```

The `Partial Disk` command works like `Disk`, but with a slice of the disk removed. Specify the part of the disk that is showing using *`startAngle`* and *`sweepAngle`*.

Spheres

The following command draws a sphere with the specified *`radius`*.

```
Sphere( radius, slices, stacks )
```

The *`slices`* can be thought of as longitudes and *`stacks`* as latitudes. About 10 of each make a nicely drawn sphere.

Lighting

It is not necessary to make specific calculations of normal vectors (as is the case for customized surfaces) for spheres, disks, and cylinders. However, you can use the following commands to customize the automatic lighting.

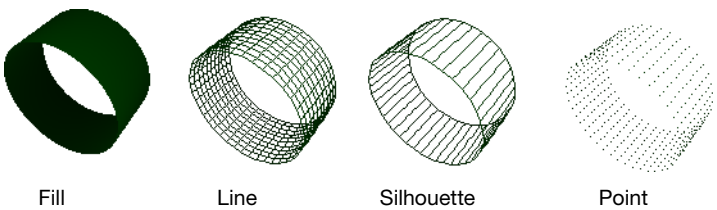
`Quadric Normals(`*`mode`*`)` tells what type of normal should be automatically generated. The argument *`mode`* can be `None`, `Flat`, or `Smooth`. `Flat` makes faceted surfaces. `Smooth` makes the normals at each vertex be the average of the adjacent polygons.

`Quadric Orientation(`*`mode`*`)` determines which way the normals point. The argument *`mode`* can be `Inside` or `Outside`.

`Quadric Draw Style(`*`mode`*`)` specifies the drawing mode. The argument *`mode`* can be `Fill`, `Line`, `Silhouette`, or `Point`.

JMP uses the values that you set for `Quadric Normals`, `Quadric Orientation`, and `Quadric Draw Style` for subsequently generated cylinders, disks, and spheres.

Figure 13.13 Draw Styles



Note: Other OpenGL documentation refers to quadric objects. JMP has only one and always uses it.

Drawing Text

As shown in the "Hello World" example above, text is added to a scene using the `Text` command.

`Text( horz, vert, size, string, <billboard>)`

- *horz* can be `Left`, `Center`, or `Right` justification.
- *vert* can be `Top`, `Middle`, `Baseline`, or `Bottom` justification.
- *size* represents the height of a capital letter M in model coordinates.
- *string* is the text to draw.
- *billboard* is an optional argument that causes the text to rotate with the model. Text with this option always faces the viewer.

The font is always the JMP Text font. You can change the text font from the preferences menu, but because of the way JMP caches fonts for scenes, changes might not take effect until JMP is restarted.

Note: Text is not part of the standard OpenGL definition.

Using Text with Rotate and Translate

The following example uses the text command in conjunction with the `Translate` and `Rotate` commands.

```
scene = Scene Box( 600, 600 ); // make a scene box...holds an OpenGL scene
New Window( "Example 2", scene ); // put the scene in a window

// the "lens" is 45 degrees, near is 3 units from the camera, far is 7
scene << Perspective( 45, 3, 7 );

// move the world so that 0,0,0 is visible in the camera
scene << Translate( 0.0, 0.0, -4.5 );

// rotate the first text about the Y (vertical on screen) axis
scene << Rotate( 30, 0, 1, 0 );
scene << Color( 1, 0, 2 ); // magenta
scene << Text( "center", "baseline", .2, "Top magenta string" );
```

```
// the next string is even farther away from the camera
scene << Translate( 0.0, 0.0, -2.0 );

// rotate the second text about the Y (vertical on screen) axis
scene << Rotate( 30, 0, 1, 0 );
scene << Color( 0, 1, 1); // aqua blue
scene << Text( "center", "baseline", .2, "Back aqua blue very long string" );

// update the display box in the window using the current display list
scene << Update;
```

Figure 13.14 Rotating and Translating Text Strings

Note that the aqua blue string extends backward beyond the far clipping plane. Change the 7 to 10 in the `Perspective` command to see the complete string.

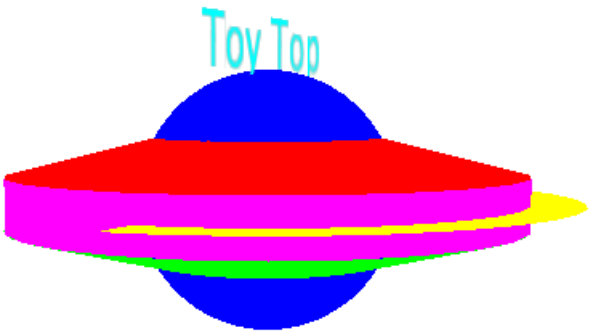
Using the Matrix Stack

JMP 3-D scenes use a matrix stack to keep track of the current transform. The stack is initialized to the identity matrix, and each time a translate, rotate, or scale command is given, the top matrix on the stack is changed.

Note: Unlike many OpenGL implementations, JMP does not use a transposed matrix.

The JSL example below uses `Push Matrix` and `Pop Matrix` to position pieces of the toy top and then return to the origin. This is faster than using the `Translate` command a second time in reverse.

Figure 13.15 Drawing With a Matrix Stack




```
toyTop = Scene Display List();
toyTop << Push Matrix;
toyTop << Translate( 0, 0, .1 );
toyTop << Color( 1, 0, 0 ); // red

// baseRadius, topRadius, height, slices, stacks
toyTop << Cylinder( 1, .2, .2, 25, 5 );
toyTop << Pop Matrix;
toyTop << Push Matrix;
toyTop << Translate( 0, 0, -.1 );
toyTop << Rotate( 180, 1, 0, 0 );
toyTop << Color( 0, 1, 0 ); // green
toyTop << Cylinder( 1, .2, .2, 25, 5 );
toyTop << Pop Matrix;

toyTop << Color( 0, 0, 1 ); // blue
toyTop << Sphere( .5, 30, 30 ); // radius, slices, stacks

toyTop << Color( 1, 1, 0 ); // yellow

// innerRadius, outerRadius, slices, rings, startAngle, sweepAngle
toyTop << Partial Disk( 1, 1.2, 25, 2, 0, 270 );

toyTop << Push Matrix;
toyTop << Translate( 0, 0, -.1 );
toyTop << Color( 1, 0, 1 ); // magenta
toyTop << Cylinder( 1, 1, .2, 25, 3 );

// baseRadius, topRadius, height, slices, stacks
toyTop << Pop Matrix;
```

```
toyTop << Push Matrix;
toyTop << Rotate( 90, 1, 0, 0 );
toyTop << Translate( 0, .5, 0 );
toyTop << Color( 0, 1, 1 ); // cyan
toyTop << Text( "center", "baseline", .2, "Toy Top" );
toyTop << Pop Matrix;

scene = Scene Box( 600, 600 ); // make a scene box...holds an OpenGL scene

New Window( "Example 3", scene ); // put the scene in a window
scene << Perspective( 45, 3, 7 );
scene << Translate( 0.0, 0.0, -4.5 );
scene << Rotate( -85, 1, 0, 0 );
scene << Rotate( 65, 0, 0, 1 );
scene << Call List( toyTop );

scene << Update; // update the display box
```

There are some cases where you want to replace the current matrix on the stack. For these cases, use the `Load Matrix` command.

```
Load Matrix(m)
```

where *m* is a 4x4 JMP matrix that is loaded onto the current matrix stack.

Similar is the `Mult Matrix` command

```
Mult Matrix(m)
```

When the `Mult Matrix` command is issued, the matrix on the top of the current matrix stack is multiplied by *m*.

The following matrices perform some simple commands.

Translation:

```
[1  0  0  x
0  1  0  y
0  0  1  z
0  0  0  1]
```

In the following rotation matrices, `c` = cos(angle) and `s`=sin(angle).

Rotation about X axis:

```
[1  0  0  0
0  c -s  0
0 -s  c  0
0  0  0  1]
```

Rotation about Y axis:

```
[c  0  s  0
 0  1  0  0
-s  0  c  0
 0  0  0  1]
```

Rotation about Z axis:

```
[c -s  0  0
 s  c  0  0
 0  0  1  0
 0  0  0  1]
```

For example, here are two equivalent (except for the translation being opposite) ways to translate and rotate a display list.

```
// first way uses matrix
gl << Push Matrix;
xt = Identity( 4 ); // translate this one left by .75
xt[1, 4] = -.75;
xr = Identity( 4 ); // rotate this one, cos needs radians, not degrees
xr[2, 2] = Cos( 3.14159 * a / 180 );
xr[2, 3] = -Sin( 3.14159 * a / 180 );
xr[3, 2] = Sin( 3.14159 * a / 180 );
xr[3, 3] = Cos( 3.14159 * a / 180 );
yr = Identity( 4 );
yr[1, 1] = Cos( 3.14159 * a / 180 );
yr[1, 3] = Sin( 3.14159 * a / 180 );
yr[3, 1] = -Sin( 3.14159 * a / 180 );
yr[3, 3] = Cos( 3.14159 * a / 180 );
zr = Identity( 4 );
zr[1, 1] = Cos( 3.14159 * a / 180 );
zr[1, 2] = -Sin( 3.14159 * a / 180 );
zr[2, 1] = Sin( 3.14159 * a / 180 );
zr[2, 2] = Cos( 3.14159 * a / 180 );

// order of multiplication matters with matrices
gl << Mult Matrix( xt * xr * yr * zr );
gl << Arcball( dl, 1 );
gl << Pop Matrix;

// second way uses functions
gl << Push Matrix;
gl << Translate( .75, 0, 0 ); // translate this one right by .75
gl << Rotate( a, 1, 0, 0 ); // rotate this one in degrees
gl << Rotate( a, 0, 1, 0 ); // order of operations also matters here
gl << Rotate( a, 0, 0, 1 );
```

```
gl << Arcball( dl, 1 );
gl << Pop Matrix;
```

It is not possible to read back the current transform matrix, because the matrix exists only while the display list is drawing, not while your JSL script is creating it. If you must know its content, create it in JSL and use `Load Matrix` to put it on the stack.

Lighting and Normals

The following methods enable you to add lighting, materials, and normal vectors to your shapes. Using these methods, models can appear shiny or light-absorbing.

- "Creating Light Sources"
- "Lighting Models"
- "Normal Vectors"
- "Shading Model"
- "Material Properties"
- "Alpha Blending"
- "Fog"
- "Example"

Creating Light Sources

Light sources are specifications of a color, position, and direction. JSL allows for up to eight lights (numbered 0 to 7) defined by the `Light` command, where *n* is the number of the light.

```
Light( n, argument, value, ... value )
```

Note: To turn each light on, issue an `Enable(`*`Lighting`*`)` and an `Enable(light`*`n`*`)` command, where *n* is the light number. Then, move the light to a position in the scene with a `Light(`*`n`*`, `*`POSITION`*`, `*`x`*`, `*`y`*`, `*`z`*`)` command.

The value of *argument* can be any one of those shown in Table 13.3. The table shows default values for each argument.

Table 13.3 Light Arguments and Default Values

Argument	Default Value	Meaning
AMBIENT	(0, 0, 0, 1)	Ambient RGBA intensity
DIFFUSE	(1, 1, 1, 1)	diffuse RGBA intensity
SPECULAR	(1, 1, 1, 1)	specular RGBA intensity
POSITION	(0, 0, 1, 0)	(x, y, z, w) position
SPOT_DIRECTION	(0, 0, -1)	(x, y, z) direction of spotlight
SPOT_EXPONENT	0	spotlight exponent
SPOT_CUTOFF	180	spotlight cutoff angle
CONSTANT_ATTENUATION	1	constant attenuation factor
LINEAR_ATTENUATION	0	linear attenuation factor
QUADRATIC_ATTENUATION	0	quadratic attenuation factor

Note: The default values for DIFFUSE and SPECULAR in this table apply only to Light 0. For other lights, the default value is (0, 0, 0, 1) for both arguments.

The first three arguments (AMBIENT, DIFFUSE, and SPECULAR) are used to color the light. DIFFUSE is the argument that is most closely associated with the physical color of the light. AMBIENT refers to the property of the light when it functions as a background light. SPECULAR alters the way a light is reflected off a surface.

Specify the position of the light using the POSITION argument. Nonzero values of the fourth (w) coordinate position the light in homogenous object coordinates.

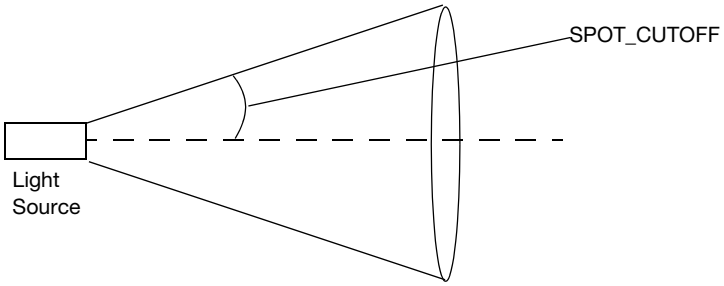
Light in the real-world decreases in intensity as distance from the light increases. Since a directional light is infinitely far away, it does not make sense to attenuate its intensity as a function of distance. However, JSL attenuates a light source by multiplying the contribution of the source by an attenuation factor

$$\text{attenuation factor} = \frac{1}{c + ld + qd^2}$$

where c = CONSTANT_ATTENUATION, l = LINEAR_ATTENUATION, and q = QUADRATIC_ATTENUATION.

To create a spotlight, limit the shape of the light to a cone. Use the `SPOT_CUTOFF` argument to define the side of the cone, as shown in the following illustration.

Figure 13.16 Spotlight



In addition to the cutoff angle, you can control the intensity and direction of the light distribution in the cone. `SPOT_DIRECTION` specifies the direction for the spotlight to point; `SPOT_EXPONENT` influences how concentrated the light is.

Lighting Models

Lighting models are specified with the `Light Model` command.

```
Light Model( argument, value,...,value )
```

Light models specify three attributes of lights.

- The global ambient light intensity
- Whether the viewpoint is local or is an infinite distance away
- Whether lighting calculations should be performed differently for the front and back faces of objects.

Table 13.3 shows the three valid arguments for the `Light Model` command.

Table 13.4 Light Model Arguments and Default Values

Argument	Default Value	Meaning
`LIGHT_MODEL_AMBIENT`	(0.2, 0.2, 0.2, 1)	Ambient RGBA intensity of the entire scene
`LIGHT_MODEL_LOCAL_VIEWER`	0 (false)	how specular reflection angles are computed
`LIGHT_MODEL_TWO_SIDE`	0 (false)	nonzero values imply two-sided lighting

Normal Vectors

Normal vectors point in a direction perpendicular to a surface. For a plane, all normals are the same. For a more complicated surface, normals are more complicated. JSL enables you to specify the normal vector for each vertex. These normals specify the orientation of the surface in space, necessary for lighting calculations. Accurate normals assure accurate lighting.

The normal vector is of length 1 and is perpendicular to the vertex. Typically, a vertex is shared between several polygons. A smooth shaded effect is desired, so the perpendicular at the vertex is calculated as a (possibly weighted) average of the polygon's normals. It is important to calculate the "outward" normal for polygons unless two-sided shading is enabled because only the outer face of the polygon is illuminated. With a scaled polygon, the normal's length is not 1 after scaling, and the lighting is wrong.

Normal vectors are set at the same time the surface is constructed, and are specified with the `Normal` command. Use the `Enable(NORMALIZE)` command to have the normals re-normalized to 1 each time the scene is drawn.

Shading Model

The shading model of a polygon is set using the `Shade Model` command.

```
Shade Model (mode)
```

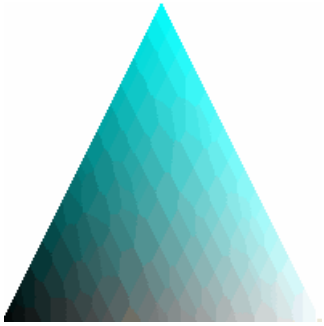
where *mode* can be `SMOOTH` (the default) or `FLAT`. `SMOOTH` shading interpolates the colors of the primitive from one vertex to the next. `FLAT` mode duplicates the color of one vertex across the entire primitive.

The following script changes the color at each of a triangle's vertices. The `FILL` shade model interpolates the color of the interior automatically.

```
scene = Scene Box( 200, 200 ); // make a scene box...holds an OpenGL scene.

New Window( "Shade Model", scene ); // put the scene in a window
scene << Clear;
scene << Ortho2D( -1, 1, -1, 1 );

scene << Shade Model( SMOOTH );
scene << Polygon Mode( FRONT_AND_BACK, FILL );
scene << Begin( TRIANGLES );
scene << Color( 0, 0, 0 ); // black
scene << Vertex( -1, -1, 0 );
scene << Color( 0, 1, 1 ); // cyan
scene << Vertex( 0, 1, 0 );
scene << Color( 1, 1, 1 ); // white
scene << Vertex( 1, -1, 0 );
scene << End();
```

```
scene << Update;
```

Figure 13.17 Shading



Material Properties

To set the material properties of a surface, use the `Material` command.

```
Material( face, argument, value,...value )
```

face can be `Front`, `Back`, or `Front_and_back`. (Note that the material properties can be set separately for the front and back faces of a polygon.)

Table 13.5 shows the arguments and default values for `Material` arguments.

Table 13.5 Material Arguments and Default Values

Argument	Default Value	Meaning
`AMBIENT`	(0.2, 0.2, 0.2, 1.0)	Ambient color of material
`DIFFUSE`	(0.8, 0.8, 0.8, 1.0)	Diffuse color of material
`AMBIENT_AND_DIFFUSE`		Both `AMBIENT` and `DIFFUSE`
`SPECULAR`	(0.0, 0.0, 0.0, 1.0)	Specular color of material
`SHININESS`	0	Specular exponent that can range from 0 to 128.
`EMISSION`	(0, 0, 0, 1)	Emissive color of material

Alpha Blending

The BlendFunc command allows for alpha blending. To use it, send a BlendFunc message to a scene, for example:

```
scene << BlendFunc(SRC_ALPHA, ONE_MINUS_SRC_ALPHA)
```

SRC_ALPHA and ONE_MINUS_SRC_ALPHA are OpenGL constants that tell BlendFunc to use alpha to blend against the existing display buffer. Disabling *z*-buffer testing or rendering primitives from back to front might be needed for some applications. By default, the *z*-buffer tests prevent anything from drawing behind a transparent polygon after it is drawn.

More information about all the constants available to BlendFunc (many of which are not useful to the JSL programmer) are available in the OpenGL documentation at opengl.org.

Fog

Fog enables figures to fade into the distance, making for more realistic models. All types of geometric figures can be fogged. To turn fog on, enable the FOG argument.

```
Enable (FOG)
```

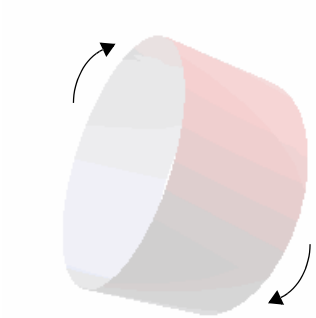
Example

The following example uses several of the concepts presented in this section, including lighting, fog, and normalization. It draws a spinning cylinder that is affected by two lights.

```
scene = Scene Box( 300, 300 ); // make a scene box
New Window( "Cylinder", scene ); // put the scene in a window

For( i = 1, i < 360, i++,
  scene << Clear;

  // the lens is 45 degrees, near is 3 units from the camera, far is 7.
  scene << Perspective( 50, 1, 10 );

  // move the world so that 0,0,0 is visible in the camera
  scene << Translate( 0.0, 0.0, -2 );
  scene << Enable( Lighting );
  scene << Enable( Light0 );
  scene << Enable( Light1 );
  scene << Light( Light0, POSITION, 1, 1, 1, 1 ); // near viewer
  scene << Light( Light0, DIFFUSE, 1, 0, 0, 1 ); // red light
  scene << Light( Light1, POSITION, -1, -1, -1, 1 ); // behind object
  scene << Light( Light1, DIFFUSE, .5, .5, 1, 1 ); // blue-gray light
  scene << Enable( Fog );
```

```
    scene << Enable( NORMALIZE );

    scene << Rotate( i, 1, 0, 0 );
    scene << Rotate( i * 3, 0, 1, 0 );
    scene << Rotate( i * 3 / 2, 0, 0, 1 );
    scene << Cylinder( 0.5, 0.5, 0.5, 40, 10 );
    scene << Update;
    Wait( 0.01 );
);
```

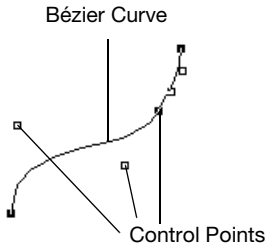
Figure 13.18 Fog



Bézier Curves

A complete discussion of Bézier curves is beyond the scope of the *Scripting Guide*. JSL has several commands for defining and drawing curves and their associated meshes.

Figure 13.19 Bézier Curve



One-Dimensional Evaluators

To define a one-dimensional map, use the `Map1` command.

```
Map1( target, u1, u2, stride, order, matrix )
```

The *`target`* argument defines what the control points represent. Values of the *`target`* argument are shown in Table 13.6. Note that you must use the `Enable` command to enable the argument.

Table 13.6 Map1 *`Target`* Arguments and Default Values

`target` Argument	Meaning
`MAP1_VERTEX_3`	(x, y, z) vertex coordinates
`MAP1_VERTEX_4`	(x, y, z, w) vertex coordinates
`MAP1_INDEX`	color index
`MAP1_COLOR_4`	R, G, B, A
`MAP1_NORMAL`	normal coordinates
`MAP1_TEXTURE_COORD_1`	s texture coordinates
`MAP1_TEXTURE_COORD_2`	s, t texture coordinates
`MAP1_TEXTURE_COORD_3`	s, t, r texture coordinates
`MAP1_TEXTURE_COORD_4`	s, t, r, q texture coordinates

The second two arguments (*`u1`* and *`u2`*) define the range for the map. The *`stride`* value is the number of values in each block of storage. In other words, the value is the offset between the beginning of one control point and the beginning of the next control point. The *`order`* should equal the degree of the curve plus one. The *`matrix`* holds the control points.

For example, `Map1( MAP1_VERTEX_3, 0, 1, 3, 4, <4x3 matrix> )` is typical for setting the two end points and two control points to define a Bézier line.

You use the `MapGrid1` and `EvalMesh1` commands to define and apply an evenly spaced mesh.

```
MapGrid1( un, u1, u2 )
```

sets up the mesh with *`un`* divisions spanning the range *`u1`* to *`u2`*. Code is simplified by using the range 0 to 1.

```
EvalMesh1( mode, i1, i2 )
```

actually generates the mesh from *i1* to *i2*. The *mode* can be either POINT or LINE. The EvalMesh1 command makes its own Begin and End clause.

The following example script demonstrates a one-dimensional outlier. A random set of control points draws a smooth curve. Only the first and last points are on the curve. Using NPOINTS=4 results in a cubic Bézier spline.

```
boxwide = 500;
boxhigh = 400;

gridsize = 100; // bigger for finer divisions

/* We suggest you use only values between 2 and 8 (inclusively). Numbers
   beyond these might be interpreted differently, depending on implementation.
   This value is the degree+1 of the fitted curve */
NPOINTS = 4;

point = J( NPOINTS, 3, 0 );

// create an array of x,y,z triples
For( x = 1, x <= NPOINTS, x++,
   point[x, 1] = (x - 1) / (NPOINTS - 1) - .5;
      // x from -.5 to +.5
   point[x, 2] = Random Uniform() - .5; // y is random in same range

   // z is always zero, which causes the curve to stay in a plane
   point[x, 3] = 0;
);

spline = Scene Box( boxwide, boxhigh );

// data from -.5 to .5 in x and y; this is a little larger
spline << Ortho( -.6, .6, -.6, .6, -2, 2 );

spline << Enable( MAP1_VERTEX_3 );
spline << MapGrid1( gridsize, 0, 1 );
spline << Color( .2, .2, 1 ); // blue curve

spline << Map1( MAP1_VERTEX_3, 0, 1, 3, NPOINTS, point );
spline << Line Width( 2 ); // not-so-skinny curve
spline << EvalMesh1( LINE, 0, gridsize ); // also try LINE, POINT

spline << Color( .2, 1, .2 );
spline << Point Size( 4 ); // big fat green points

// show the points and label them
For( i = 1, i <= NPOINTS, i++,
```

```
    spline << Begin( "POINTS" );
    spline << Vertex( point[i, 1], point[i, 2], point[i, 3] );
    spline << End;
    spline << Push Matrix;
    spline << Translate( point[i, 1], point[i, 2], point[i, 3] );
    spline << Text( center, bottom, .05, Char( i ) );
    spline << Pop Matrix;
);

New Window( "Spline", spline );
```

https://www.tinaja.com/glib/bezconn.pdf offers an explanation of connecting cubic segments so that both the slope and the rate of change match at the connection point. This example does not illustrate doing so; there is only one segment here.

Two-Dimensional Evaluators

Two-dimensional evaluators follow their one dimensional counterparts, and are used in a similar way.

```
Map2(target, u1, u2, ustride, uorder, v1, v2, vstride, vorder, matrix)
Eval Coord2(u, v)
```

Values for the *target* argument are the same as those shown in Table 13.6 with *Map1* replaced with *Map2* appropriately. The *u1, u2, v1,* and *v2* values specify the range of the two-dimensional mesh.

For example, `Map2(MAP2_VERTEX_3, 0, 1, 3, 4, 0, 1, 12, 4, <16x3 matrix>)` is typical for setting the 16 points that define a Bézier surface.

Use the `MapGrid2` and `EvalMesh2` commands to define and apply an evenly spaced mesh.

```
MapGrid2(un, u1, u2, vn, v1, v2)
```

sets up the mesh with *un* and *vn* divisions spanning the range *u1* to *u2* and *v1* to *v2*. Code is simplified by using ranges that span 0 to 1.

```
EvalMesh2(mode, i1, i2, j1, j2)
```

actually generates the mesh from *i1* to *i2* and *j1* to *j2*. The *mode* can be `POINT`, `LINE`, or `FILL`. The `EvalMesh2` command makes its own `Begin` and `End` clause.

Using the Mouse

Mouse activity is supported through two feedback functions. The Patch Editor.jsl sample script uses these functions to support the dragging and dropping of points. Part of that script, the call-back function for mouse activity, is explained below. To run the script, open PatchEditor.jsl in the Scene3D folder inside the JMP Sample Scripts folder.

```
topClick2d = Function( {x, y, m, k},
   dragfunc( x, boxhigh - y, m, 1, 2 );
   1;
);
frontClick2d = Function( {x, y, m, k},
   dragfunc( x, boxhigh - y, m, 1, 3 );
   1;
);
rightClick2d = Function( {x, y, m, k},
   dragfunc( x, boxhigh - y, m, 2, 3 );
   1;
);

Click3d = Function( {x, y, m, k, hitlist},
   If( m == 1,
      If( N Items( hitlist ) > 0,
         CurrentPoint = hitlist[1][3], /* first matrix in the list is the
   closest; 3rd element of matrix is ID*/
         CurrentPoint = 0
      );
      makePatch();
   );
   0; /* cares only about initial mouse down. return 1 if drag, release is
   needed, but then arcball does not happen. */
);

/* after one of the 3 Click2d functions figures out which axis of the model is
   represented by the screen X, Y, pass in to this common code */
dragfunc = Function( {x, y, m, ix, iy}, /* ix, iy are the index of the X, Y,
   or Z part of the coord in the points matrix */
   If( CurrentPoint > 0,
      points[CurrentPoint, ix] = (x / boxwide) * (orthoright - ortholeft) +
   ortholeft;
      points[CurrentPoint, iy] = (y / boxhigh) * (orthotop - orthobottom) +
   orthobottom;
      makepatch();
   )
);
```

When a 2-D function is called, the arguments are X, Y, M, K.

- X and Y are the coordinates of the mouse.
- M shows the state of the mouse and button. M=0 says that the mouse button is up. M=1 says that the button was just pressed. M=2 says that the button is down and the mouse is moving. M=3 says that the button was just released.
- K is related to the keys Shift, Alt, and Ctrl. K=1 for the Shift key, K=2 for the Ctrl (Command) key, and K=3 for the Alt (Option) key.

The 3-D function is called similarly. The arguments are X, Y, M, K, `hitlist`, where `hitlist` is a list of matrices

```
[znear, zfar, id1, id2, id3, ...]
```

`znear`, `zfar` is the Z distance from the camera of the near and far edge of the object. The matrices are sorted from near to far by the midpoint of `znear`, `zfar`. The `id`s in the list are the `pushname`, `loadname`, and `popname` values you just put in the display list.

The drag functions use a return value to tell if mouse processing should continue. That is the trailing "1" you see in the functions. Anything else stops the mouse tracking. This is needed because the 2-D and 3-D functions do not run in parallel. You might want the 2-D to return 0 and the 3-D to return 1 so the tracking would happen in 3-D rather than 2-D.

Pick Commands

This `SceneBox` callback gets the 2-D mouse coordinates, and then uses `Pick` to determine the "named" object under the mouse. For example, `hitlist` is a 5x5 pixel pick box around `x,y`; up to 1000 items returned, but just the leaf names. The format of the return is determined by the last argument (1 returns a simple array, 0 returns a sorted (by depth) list of arrays).

```
Track2d=function({x, y, m, k},
hitlist = theSceneBox<<pick( x, y, 5, 5, 1000, 1 );
  if ( nrow(hitlist) > 0, // something IS in the pick box
     ... hitlist[1..n] // are names that you put in the display list
  ) );
```

Contrast this with a call back `Track3d` function, where the pick rectangle is always 1x1 and picking only happens when the mouse moves. This is almost always what you want, but points are difficult to pick because the 1x1 pick area is the same small size as the point. This function lets you pick without a mouse move.

The `Track3d` function always provides a depth-sorted list of arrays; each array can describe multiple names in a hierarchy (pushname and popname construct a hierarchy of objects). The sorting can be very slow when thousands of objects are selected. The final argument (1, above) controls whether the pick function replaces the sorted list of arrays with a simple array. The simple array contains only the "leaf" names, not higher level names.

Three-Dimensional Arguments

Arguments enable you to specify special modes and settings. To enable an argument, use the `Enable(argument)` command. To disable an argument, use the `Disable(argument)` command. Available arguments are shown in Table 13.7.

Table 13.7 Arguments

ALPHA_TEST	LIGHT5	MAP2_TEXTURE_COORD_3
AUTO_NORMAL	LIGHT6	MAP2_TEXTURE_COORD_4
BLEND	LIGHT7	MAP2_VERTEX_3
CLIP_PLANE0	LIGHTING	MAP2_VERTEX_4
CLIP_PLANE1	LINE_SMOOTH	NORMALIZE
CLIP_PLANE2	LINE_STIPPLE	POINT_SMOOTH
CLIP_PLANE3	MAP1_COLOR_4	POLYGON_OFFSET_FILL
CLIP_PLANE4	MAP1_INDEX	POLYGON_OFFSET_LINE
CLIP_PLANE5	MAP1_NORMAL	POLYGON_OFFSET_POINT
COLOR_LOGIC_OP	MAP1_TEXTURE_COORD_1	POLYGON_SMOOTH
COLOR_MATERIAL	MAP1_TEXTURE_COORD_2	POLYGON_STIPPLE
CULL_FACE	MAP1_TEXTURE_COORD_3	SCISSOR_TEST
DEPTH_TEST	MAP1_TEXTURE_COORD_4	STENCIL_TEST
DITHER	MAP1_VERTEX_3	
FOG	MAP1_VERTEX_4	
LIGHT0	MAP2_COLOR_4	
LIGHT1	MAP2_INDEX	
LIGHT2	MAP2_NORMAL	
LIGHT3	MAP2_TEXTURE_COORD_1	
LIGHT4	MAP2_TEXTURE_COORD_2	

Improve the Display of Graphics

Hardware acceleration determines how fast and smoothly graphics are drawn. On Windows, hardware acceleration is turned off by default.

Note: JMP uses OpenGL 1.1 to perform 3-D rendering.

If you need OpenGL and have the hardware support, turn on hardware acceleration in one of the following ways:

- Right-click the graph and select **Use Hardware Acceleration**.
- Send the `Use Hardware Acceleration(1)` message to the scene box.

- In your script, set the global variable `Scene3DHardwareAcceleration = 1;` If you want hardware acceleration to be the default across runs of JMP, add it to your jmpStartAdmin.jsl file. Because it is a global variable, it will remain in effect until changed.

Chapter 14

Extending JMP

External Data Sources, Analytical Tools, and Automation

This chapter discusses scripting features that are particularly useful for production settings, such as the following:

- a data feed for capturing real-time data from a laboratory instrument
- using SAS, MATLAB, R, or Python through JMP Scripting Language (JSL)
- using JMP with Microsoft Excel
- connecting to databases
- controlling JMP externally by OLE automation
- parsing XML
- publishing to JMP Public or JMP Live
- communicating with REST web services
- communicating with web APIs

Some general JSL commands that might be of particular interest for use in a production setting include `Caption`, `Speak`, `Print`, `Write`, and `Mail`. These commands are described in "Functions that Communicate with Users".

Contents

Real-Time Data Capture on Windows . . . 797
Create a Data Feed Object . . . 797
Read in Real-Time Data . . . 799
Manage a Data Feed with Messages . . . 800
Examples of Data Feed . . . 802
Dynamic Link Libraries (DLLs) . . . 804
Use Sockets in JSL . . . 806
Socket-Related Commands . . . 807
Messages for Sockets . . . 808
Database Access . . . 809
Run Query Builder Queries . . . 809
Open Database Function . . . 810
Creating a Database Connection and Executing SQL . . . 812
Write a SQL Query . . . 813
Work with SAS . . . 814
Make a SAS DATA Step . . . 814
Create SAS DATA Step Code for Formula Columns . . . 814
SAS Variable Names . . . 815
Connect to a SAS Metadata Server . . . 815
Get the Values of SAS Macro Variables . . . 818
Preferences . . . 819
Sample Scripts . . . 819
Work with MATLAB . . . 821
Install MATLAB . . . 821
Work with R . . . 822
Installing R . . . 822
JMP to R Interfaces . . . 824
R JSL Scriptable Object Interfaces . . . 825
Conversion Between JMP Data Types and R Data Types . . . 825
Troubleshooting . . . 828
Examples . . . 829
Work with Python . . . 831
Install Python . . . 831
Troubleshooting the JMP Python Integration . . . 832
JMP-to-Python Interfaces . . . 836
Equivalent Data Types for Python Send() . . . 836
Equivalent Data Types for Python Get() . . . 837
JMP Scoping Operators and Python . . . 838
Python Name() and Python Send() Examples . . . 839

Additional Python Integration Examples . . . 840
Work with Microsoft Excel . . . 842
Parse XML . . . 842
OLE Automation . . . 844
Share Reports on JMP Live . . . 845
JMP Live Prerequisites . . . 845
Get an API Key . . . 845
Publish a Single Report to JMP Live . . . 846
Publish Several Reports to a JMP Live Folder . . . 847
Delete Reports in a JMP Live Folder . . . 847
Replace the Contents of a JMP Live Report . . . 848
Search for a JMP Live Report or Folder ID . . . 848
Update Data in JMP Live Reports . . . 849
Work with Other Content in JMP Live Reports . . . 850
HTTP Requests in JMP Live . . . 850
Communicate with REST Web Service . . . 850
Examples of HTTP Requests . . . 850
OAuth 2.0 for Web APIs . . . 853
Security . . . 853
Using Postman for Redirect URLs . . . 853
Authorization Code Grant . . . 854
Implicit Grant . . . 855
Resource Owner Grant . . . 855
Using a Token . . . 856
OAuth 2.0 Resources . . . 856
Communicate with SAS Cloud Analytic Services . . . 856
About CAS Actions . . . 857
Connect to the CAS Server . . . 858
Create and Submit a CAS Action . . . 859
Create and Submit a CAS DATA Step Action . . . 859
Create a New CAS Server . . . 859
Import Data from the CAS Server . . . 859
View a CAS Dataset . . . 860
Get a List of Available CAS Libraries . . . 860
Export Data from a Data Table to a CAS Server . . . 860
Convert a CAS Table to a Data Table . . . 860
Delete a CAS Table . . . 863
Remove a CAS Table . . . 863
Verify the CAS Server Connection . . . 863
Get Details about the Current Connection . . . 863

Disconnect from the CAS Server 863
Terminate All CAS Sessions 863

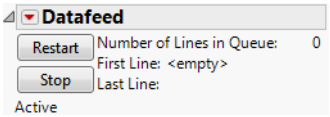
Real-Time Data Capture on Windows

A data feed is a real-time method to read data continuously, such as from a laboratory measurement device connected to a serial port. A data feed object sets up a concurrent thread with a queue for input data lines that arrive in real time and are processed with background events. You set up scripts to interpret the data lines and push the data to data tables, or do whatever else your process requires.

For example, submit this to get records from com1: and list them in the log.

```
feed = Open Datafeed(
  Connect( Port( "com1:" ), Baud( 9600 ), DataBits( 7 ) ),
  Set Script( Print( feed << GetLine ) )
);
```

Figure 14.1 A Datafeed Window Shows the Status and Offers Controls



Create a Data Feed Object

To create a data feed object, use the `Open Datafeed()` function with arguments specifying details about the connection:

```
feed = Open Datafeed( options );
```

No arguments are required. You can simply create the data feed object and then send messages to it. Messages might include connecting to a port or setting up a script to process the data coming in. However, you would typically set up the basic operation of the data feed in the `Open Datafeed()` function and subsequently send messages as needed to manage the data feed. Any of the options below work both as options inside `Open Datafeed()` or as messages sent to a data feed object.

It's a good idea to store a reference to the object in a global variable, such as *feed* above, so that you can easily send messages to the object. You can store a reference to an existing object by using a subscript; for example, to store a reference to the second data feed created:

```
feed2 = Datafeed[2];
```

Data Feed Options

To connect to a live data source, use `Connect( )` and specify details for the port. Each setting takes only one argument; in this syntax summary the symbol | between argument choices means "or." The `Port` specification is needed if you want to connect, otherwise the object still works, but is not connected to a data feed. The last three items, `DTR_DSR`, `RTS_CTS`, and `XON_XOFF`, are Boolean to specify which control characters are sent back and forth to indicate when the data feed is ready to get data. Typically, you would turn on at most one of them.

```
feed = Open Datafeed(
  Connect(
     Port( "com1:" | "com2:" | "lpt1:" | ... ),
     Baud( 9600 | 4800 | ... ),
     Data Bits( 8 | 7 ),
     Parity( None | Odd | Even ),
     Stop Bits( 1 | 0 | 2 ),
     DTR_DSR( 0 | 1 ), // Data Terminal Ready
     RTS_CTS( 0 | 1 ), // Request To Send | Clear To Send
     XON_XOFF( 1 | 0 ) // Transmitter On | Transmitter Off
  )
);
```

This command creates a scriptable data feed object and stores a reference to it in the global variable *feed*. The `Connect()` argument starts up a thread to watch a communications port and collect lines. The thread collects characters until it has a line, and then appends it to the line queue and schedules an event to call the script.

Note: A line of data from the feed might generate multiple rows in a data table or might be a fraction of one row. You provide JSL to parse the data lines and to add rows to a data table as needed.

`Set Script()` attaches a script to the data feed object. This script is triggered by the `Open Datafeed` message whenever a line of data arrives. The argument for `Set Script()` is simply the script to run, or a global containing a script.

```
feed = Open Datafeed( Set Script( myScript ) );
feed = Open Datafeed( Set Script( Print( feed << Get Line ) ) );
```

A data feed script typically uses `Get Line` to get a copy of one line and then does something with that line. Often it parses the line for data and adds it to some data table.

Read in Real-Time Data

The term *live data feed* describes the way an external data source sends information via a physical or a logical communication link to another device. You can connect JMP to a live data feed through the serial port of your Windows computer to read a stream of incoming data in real time. Remember the following:

- The data feed must come through a standard nine-pin serial port. Data cannot be read through a USB port unless there is a driver that can simulate a serial port.
- You need to know the exact baud rate, parity, stop bits, and data bits for the attached device.

Once you obtain the numbers for your device, enter them into the `Open Datafeed()` command in the script below. (The `4800`, `even`, `2`, and `7` in the script below are examples, so replace them with your information). Then connect the data feed to your computer and open and run the script:

```
streamScript = Expr(
   line = feed << Get Line;
   Show( line );
   len = Length( line );
   Show( len );
   If( Length( line ) >= 1,
      Show( "Hi" );
      Show( line );
      field = Substr( line, 5, 8 );
      Show( field );
      x = Num( field );
      Show( x );
      If( !Is Missing( x ),
         dt << Add Row( {:Column1 = x} );
         Show( x );
      );
   );
);
feed = Open Datafeed(
   Baud Rate( 4800 ),
   parity( even ),
   Stop bits( 2 ),
   Data bits( 7 )
);
feed << Set Script( streamScript );
feed << Connect;
```

To ensure harmony between the communications settings for JMP and the instrument reading data from an external source, select **File > Preferences > Communications**. Refer to the documentation for your instrument to find the appropriate settings.

Manage a Data Feed with Messages

A data feed object responds to several messages, including `Connect` and `Set Script`. These are detailed above as arguments for `Open Datafeed()`. They can also be sent as messages to a data feed object that already exists:

```
feed << Connect( port( "com1:" ), baud( 4800 ), databits( 7 ), parity( odd ),
  stopbits( 2 ) );
feed << Set Script( myScript );
```

The following messages could also be used as arguments to `On Datafeed()`. However, it would be more common to send them as messages to a data feed object that is already present.

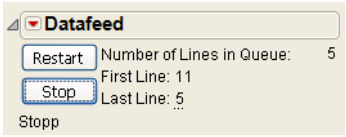
You can send lines to a data feed from a script. This is a quick way to test a data feed. If you do not have a device available, `Queue Line` does not send data to a device; it simulates getting data from a device. Include a text argument or a global that stores text:

```
feed << Queue Line( "14" );
feed << Queue Line( myValue );
```

Here is a test script to queue five lines of data:

```
feed << Queue Line( "11" );
feed << Queue Line( "22" );
feed << Queue Line( "33" );
feed << Queue Line( "44" );
feed << Queue Line( "55" );
```

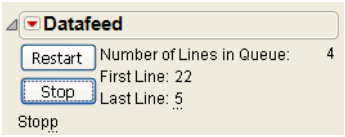
Figure 14.2 Datafeed: 5 Lines Queued



To get the first line currently waiting in the queue, use a `Get Line` (singular) message. When you get a line, it is removed from the queue. Five lines were queued with the test script above, and `Get Line` returns the first line and removes it from the queue:

```
feed << Get Line
   "11"
```

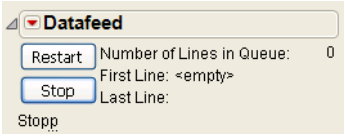
Figure 14.3 Datafeed: 4 Lines Queued



To empty all lines from the queue into a list, use `Get Lines` (plural). This returns the next four lines from the test script in list `{ }` format.

```
myList = feed << GetLines;
   { "22", "33", "44", "55" }
```

Figure 14.4 Datafeed: 0 Lines Queued



To stop and later restart the processing of queued lines, either click the **Stop** and **Restart** buttons in the Datafeed window, or send the equivalent messages:

```
feed << Stop;
feed << Restart;
```

To close the data feed and its window:

```
feed << Close;
```

To disconnect from the live data source:

```
feed << Disconnect
```

Table 14.1 Datafeed Messages

Message	Syntax	Explanation
`Open Datafeed` `Datafeed`	`feed = Open Datafeed( commands )`	Creates a data feed object. Any of the following can be used as commands inside `Open Datafeed` or sent as messages to an existing data feed object. `Datafeed` is a synonym.
`Set Script`	`feed << Set Script( script )`	Assigns the *`script`* that is run each time a line of data is received.
`Get Line`	`feed << Get Line`	Returns and removes one line from the data feed queue.

Table 14.1 Datafeed Messages *(Continued)*

Message	Syntax	Explanation
Get Lines	`feed << Get Lines`	Returns as a list and removes all lines from the data feed queue.
Queue Line	`feed << Queue Line( string )`	Sends one line to the end of the data feed queue.
Stop	`feed << Stop`	Stops processing queued lines.
Restart	`feed << Restart`	Restarts processing queued lines.
Close	`feed << Close`	Closes the data feed object and its window.
Connect	`feed << Connect( Port( "com1:" | "lpt1:" | ... ), Baud( 9600 | 4800 | ... ), Data Bits( 8 | 7 ), Parity( None | Odd | Even ), Stop Bits( 1 | 0 | 2 ), DTR_DSR( 0 | 1 ), RTS_CTS( 0 | 1 ), XON_XOFF( 1 | 0 ) )`	(Windows only) Sets up port settings for the connection to the device. The symbol | between arguments represents "or"; choose one argument for each setting.
Disconnect	`feed << Disconnect`	(Windows only) Disconnects the device from the data feed queue but leaves the data feed object active.

Examples of Data Feed

Reading Data

Here is a typical data feed script. It expects to find a string 3 characters long starting in column 11. If it does, it uses it as a number and then adds a row to the data table in the column "thickness."

```
feed = Open Datafeed( );
myScript = Expr(
   line = feed << Get Line;
   If( Length( line ) >= 14,
      x = Num( Substr( line, 11, 3 ) );
```

```
        If( !Is Missing( x ),
           dt << Add Row( {thickness = x} )
        );
    );
);
```

Assign the script to the data feed object by using `Set Script`:

```
feed << Set Script( myScript );
```

Set Up a Live Control Chart

Here is a sample script that sets up a new data table and starts a control chart based on the data feed.

```
dt = New Table( "Gap Width" ); // make a data table with one column
dc = dt << New Column( "gap", Numeric, Best );

// set up control chart properties
dc << Set Property(
  "Control Limits",
  {XBar( Avg( 20 ), LCL( 19.8 ), UCL( 20.2 ) )}
);
dc << Set Property( "Sigma", 0.1 );

// make the data feed
feed = Open Datafeed();
feedScript = Expr(
  line = feed << Get Line;
  z = Num( line );
  Show( line, z ); // if logging or debugging
  If( !Is Missing( z ),
     dt << Add Row( {:gap = z} )
  );
);
feed << Set Script( feedScript );

// start the control chart
Control Chart Builder(
       Show Capability( 0 ),
       Variables( Y( :gap ) ),
       Set Subgroup Size( 5 )
);
```

```
/* Either start the feed from the device or test-feed some data
to see it work (comment out one of the lines):
feed << Connect( Port( "com1:" ), Baud( 9600 ) );*/
For( i = 1, i <= 20, i++,
  feed << Queue Line( Char( 20 + Random Uniform() * .1 ) );
  Wait( .1 ); // simulate a delay in the feed
);
```

Store the Script in a Data Table

You can further automate the production setting by placing a data feed script such as the one above in an `On Open` data table property. A property with this name is run automatically each time the table is opened (unless you set a preference to suppress execution). If you save such a data table as a template, opening the template runs the data feed script and creates a new data table.

Dynamic Link Libraries (DLLs)

Note: 64-bit JMP cannot load 32-bit DLLs. You must recompile a 32-bit DLL for JMP to be able to load it.

You can extend JMP functionality by using JMP Scripting Language (JSL) to load a DLL and call functions exported by that DLL. There is one JSL command and six messages that implement this functionality.

```
dll_obj = Load DLL("path" <, AutoDeclare(Boolean | Quiet | Verbose) |Quiet |
  Verbose > )
```

`Load DLL()` loads the DLL in the specified *`path`*. Use the `AutoDeclare(Quiet)` argument to suppress log window messaging.

Use the `Declare Function` message to declare a function that is defined in the DLL.After you declare the function, you can call it.

```
dll_obj <<Declare Function("name", Convention(named_argument),
  Alias("string"), Arg(type, "string"), Returns(type), other_named_arguments)
```

The `Alias` defines an alternate name that you can use in JSL. For example, if you declared `Alias("MsgBox")` for a function that is named `"Message Box"` in the DLL, then you would call it as follows:

```
result = dll_obj <<MsgBox(...)
```

Here are the named arguments for `Convention`:

- STDCALL or PASCAL

- CDECL

The type argument for both `Arg` and `Returns` can be one of the following:

Table 14.2 Types for Arg and Returns

`Int8`	`UInt8`	`Int16`	`UInt16`
`Int32`	`UInt32`	`Int64`	`UInt64`
`Float`	`Double`	`AnsiString`	`UnicodeString`
`Struct`	`IntPtr`	`UIntPtr`	`ObjPtr`

See the *JSL Syntax Reference* for the `Declare Function` message arguments.

Finally, use the `UnLoadDLL` message to unload the DLL:

```
dll_obj << UnLoadDLL
```

Note: Refer to the documentation for that function provided by the DLL author when you declare a function. If the argument types and calling convention declared do not match the actual function in the DLL, calling the function could cause JMP to terminate.

Example

In the following example, the DLL contains a function named `netPresentValue`, which takes four parameters. `Declare Function()` defines these parameters and the data type that the function returns so that JMP can successfully call the function.

```
xftdll = Load Dll( "c:\ExtFunctionTests\ExtFunctionTests_x64.dll" ),

Show( xftdll );

xftdll << Declare Function(
  "netPresentValue",
 Convention( CDECL ),
  Alias( "npv" ), // use this name to call the function in JMP
  Arg( Double, "discount rate" ),
  Arg( Double, "cash flows per period" ),
  Arg( Int32, "number of cash flows" ),
  Arg( Double, Array, "array of cash flow values" ),
  Returns( Double )
);

result = xftdll << npv( 0.10, 1, 10, [ 100, 100, 100, 100, 100,
```

```
    100, 100, 100, 100, 100 ]
);
```

Other DLL Messages

The `Show Functions` message sends any functions that have been declared using `Declare Function()` to the log:

```
dll_obj << Show Functions;
```

If you are writing your own DLL, you can create functions in it using JSL. The `Get Declaration JSL` message sends any JSL functions in your DLL to the log:

```
dll_obj << Get Declaration JSL;
```

Use Sockets in JSL

Another tool that can be useful in establishing a live data feed is JSL Sockets. You can create two types of sockets using JSL:

Stream Stream sockets create a reliable connection between JMP and another computer. The other computer might be running JMP, or it might be a vending machine, data collector, printer, or other device that is capable of socket communication. Some devices implement their interface as an HTTP web server.

Datagram Datagram sockets create a less reliable connection between JMP and another computer. Datagrams are connectionless and information might arrive multiple times, not at all, and out of order. A datagram connection does not include all the overhead of a stream connection that does guarantee reliability. Because datagrams are connectionless, the destination address must be supplied each time (and for the same socket that can be different each time).

Once a socket is created, it can do two things: wait for a connection from another socket, or make a connection to another socket. Here is a simple example program that makes a connection to another computer's web server to get some data:

```
tCall = Socket( );
tCall << Connect( "www.jmp.com", "80" );
tCall << Send( Char To Blob( "GET / HTTP/1.0~0d~0a~0d~0a", "ASCII~HEX" ) );
tMessage = tCall << Recv( 1000 );
text = Blob To Char( tMessage[3] );
Show( text );
tCall << Close( );
```

The first line creates a socket and gives it a reference name (`tCall`). By default, a stream socket is created. You can designate the type of socket to create with an optional argument: `socket(STREAM)` or `socket(DGRAM)`.

The second line connects the `tCall` socket to port 80 (which is generally the HTTP port) of the JMP website.

The third line sends a `GET` request to the JMP web server; this message tells the JMP web server to send the JMP home page back. The / that follows the word GET should be the path to the page to be opened. A / opens the root page.

The fourth line receives up to 1000 bytes from the JMP web server and stores a list of information in `tMessage`. Each socket call returns a list. The first element of the list is the name of the call, and the second is a text message, which might be `ok` or a longer diagnostic message. Additional elements, if present, are specific to the call. In this case, the third element in the list is the data received.

The fifth line converts the binary information received into a character string. `tMessage[3]` is the third item in the list returned by `Recv`; it is the data from the JMP web server.

The sixth line displays the data in the log.

The last line closes the socket. The web server has already closed the far end, so this socket either needs reconnecting or proper disposal (`close`).

See the Scripting Index in the Help menu for more examples. The JMP Samples/Scripts folder also contains several examples of scripts using sockets.

Socket-Related Commands

Before creating and using a socket, you might need to retrieve information about the end that you want to connect to. `GetAddrInfo( )` and `GetNameInfo( )` each takes an address argument and an optional port argument and returns a list of information. For example:

```
Print( Get Addr Info( "www.sas.com" ) );
Print( Get Addr Info( "www.sas.com", "80" ) );
Print( Get Name Info( "149.173.5.120" ) );
   {"Get Addr Info", "ok", {"PF_INET", "SOCK_0", "IPPROTO_0", "149.173.5.120",
      "0"}}
   {"Get Addr Info", "ok", {"PF_INET", "SOCK_0", "IPPROTO_0", "149.173.5.120",
      "80"}}
   {"Get Name Info", "ok", {"PF_INET", "SOCK_0", "IPPROTO_0",
      "chess.exnet.sas.com", "0"}}
```

Sometimes there can be more than one answer. In that case, the sublist might be repeated one or more times. These functions can be quite slow; you probably should not try to build a data table of every website name with it. For IPV6 compatibility, you should generally use names like “www.sas.com” rather than the numerical form of an address.

Messages for Sockets

Once you have created a socket with `Socket( )`, there are many messages that you can send to it.

connect Connects to a listening socket. Returns a list: `{"connect", "ok"}` if the connection was successful; or an error if not (for example, `{"connect", "CONNREFUSED: The attempt to connect was forcefully rejected. "}`).

close Closes the connection when you are finished with it. Returns a list (for example, `{"Close", "ok"}`).

send Sends a STREAM message to the other end of the socket.

sendto Sends a DGRAM message to the other end of the socket.

recv Receives a STREAM message. The data comes back in a list, along with some other information. `Recv` takes a required numeric argument that specifies the number of bytes to accept.

recvfrom Receives a DGRAM message.

ioctl Controls the socket's blocking behavior. By default, JMP sockets block if no data is available; the socket does not return control to the JSL program until data *is* available. This makes scripts easy to write, but not particularly robust if the remote end of the connection fails to supply the data. A socket that is set for non-blocking behavior always returns immediately, either with an `ok` return code and some data, or with a "`WOULDBLOCK: ...`" return code, which means if it were a blocking socket, it would have to wait (block progress of the next JSL statement) until data became available.

Important: Background operations that use a JSL callback avoid this issue; a socket used in a background `recv`, `recvfrom`, or `accept` is set to non-blocking and is polled during `wait` statements and when JMP is otherwise idle.

`Ioctl` returns a list. For example, `{"ioctl", "ok"}`, or `{"ioctl", "NOTCONN: The socket is not connected. "}` if the socket has not been bound (see `bind`, below) or connected already.

bind Tells the server socket what address the client socket listens on. `Bind` associates a port on the local machine with the socket. This is required before a socket can `Listen`. (See below). `Bind` is not usually used on sockets that connect; the operating system selects an unused port for you. `Bind` is needed for a server because anyone that wants to connect to the server needs to know what port is being used. A common port is `80`, the HTTP port. `Bind` returns a list. For example, `{"bind", "ok"}`, or `{"bind", "ADDRNOTAVAIL: The specified address is not available from the local machine. "}` if you try binding to a name that is not on your machine. Another socket can connect to this socket if it knows your machine name and the number.

listen Tells the server socket to listen for connections. A listening socket is listening for connections from other sockets. You need only to put the socket into listen mode once. `Accept` (see below) is used to accept a connection from another socket. `Listen` returns a list. For example, `{"listen", "ok"}`, or `{"listen", "INVAL: The socket is (or is not, depending on context) already bound to an address. or, Listen was not invoked prior to accept. or, Invalid host address. or, The socket has not been bound with Bind. "}` if your bind call did not succeed.

accept Tells the server socket to accept a connection and return a new connected socket. `Accept` returns a list stating what happened, and if successful, a new socket that is connected to the socket at the other end. For example, the returned list might be `{"Accept", "ok", "localhost", socket( )}`. In this case `localhost` is the name of the machine that just connected, and the fourth argument is a socket that you can use to `send` or `recv` a message.

getpeername Asks about the other end of a connection. `GetPeerName` returns a list with information about the other end's socket: `{"getpeername", "ok", "127.0.0.1", "4087"}`. If this is a server socket, you can discover the address and port of the client that connected. If this is a client socket, you re-discover the name and port of the server that you used in the connect request.

getsockname Asks about this end of a connection. `GetSockName` returns a list with information about this socket: `{"getsockname", "ok", "localhost", "httpd"}`. If this is a client socket, you can discover the port the operating system assigned. If this is a server socket, you already know this information from a `bind` you already made.

Database Access

Run Query Builder Queries

Query Builder is the preferred method for selecting and importing data from a SQL database without writing SQL statements. You can preview the data before importing it into a data table. Share your queries so that other users can customize and run the queries.

After you create the query, you can run the query with a JSL script. Use `Include()` to include the query in your script without opening it.

```
query = Include( "$DOCUMENTS/Airline.jmpquery" );
query << Run Foreground( );
```

Instead of running the query in the foreground, you can use `Run Background()` to run the query in the background or `Run()` to use the Query Builder preference for running queries. By default, queries run in the background.

See the *JSL Syntax Reference* for more information about query messages.

Open Database Function

JMP supports ODBC access to SQL databases through JSL with the `Open Database` function.

```
dt = Open Database(
   "Connect Dialog" | "DSN=...", // data source
   "sqlStatement" | "dataTableName" | "SQLFILE=...", // SQL statement
   Invisible,  //Optional keyword to hide the table upon importing it
   "outputTableName" // new table name
);
```

Note: The characters `$# -+/%()&|;?` in a database table name must be quoted with square brackets [].

The first argument is a quoted connection string to specify the data source. It should be either of the following:

- `"Connect Dialog"` to display the Select Data Source window (Windows) or Choose DSN window (macOS).
- `"DSN="` and then the data source name and any other information needed to connect to the data source. On Windows, the data source name is shown in the name column of the User DSN or System DSN tab of the ODBC Data Source Administrator. On macOS, the DSN is shown in the ODBC Manager or iODBC Driver Manager. The rest of the string depends on the ODBC data source.

For example:

```
"DSN=dBASE Files;DBQ=C:/Program Files/SAS/JMP/16/Samples/Import Data;"
```

The second argument is a double-quoted string that can be one of the following:

1. An SQL statement to execute. For example, the second argument might be a `SELECT` statement in a quoted string like the following:

   ```
   "SELECT AGE, SEX, WEIGHT FROM BIGCLASS"
   ```

   The SQL must conform to the SQL that the data source supports. That is, a table named "Big Class" would need to be appropriately quoted because of the space between “Big” and “Class” (if whitespace is supported). The method for quoting depends on the data source and is usually done with ", ’, ‘, or [] characters.
2. The name of a data table. In this case, the effect is an SQL `"SELECT * FROM"` statement for the data table indicated. For example, `Open Database` would in effect execute the statement `"SELECT * FROM BIGCLASS"` if you specify this for the second argument:

   ```
   "BIGCLASS"
   ```

3. "SQLFILE=" and a path to a text file containing an SQL statement to be executed. For example, with the following argument, JMP would attempt to open the file mySQLFile.txt from the C:\ directory and then execute the SQL statement in the file.

   ```
   "SQLFILE=C:\mySQLFile.txt"
   ```

The optional `Invisible` argument creates a hidden data table. Hidden data tables remain in memory until they are explicitly closed, reducing the amount of memory that is available to JMP. To close the hidden data table, call `Close(dt)`, where `dt` is the data table reference.

The optional *outputTableName* argument is optional and specifies the name of the output table to be created, if any. Note that `Open Database` does not always return a data table. The return value might be null. Whether it returns a data table depends on the type of SQL statement executed. For example, a `SELECT` statement would return a data table, but a `DROP TABLE` statement would not.

To save a table back to a database through JSL, send the data table reference a `Save Database( )` message:

```
dt << Save Database( "connectInfo", "TableName" );
```

The first argument works the same way as it does in `Open Database`. Note that some databases do not allow you to save a table over one that already exists. If you want to replace a table in a database, use a DROP TABLE SQL statement in an Open Database command:

```
Open Database ( "connectinfo", "DROP TABLE TableName" );
```

Note: JMP 13 or later can save tables with spaces and mixed case in the names if the data source supports them. Most data sources do support spaces, except Apache Hive and Apache Hadoop. Mixed case is preserved, but for the most part, SQL is case insensitive.

The following script opens a database with an SQL query, saves it back to the database under a new name, and then deletes the new table.

```
dt = Open Database(
   "Connect Dialog",
   "SELECT age, sex, weight FROM \!"Bigclass$\!"",
   "My Big Class"
);
dt << Save Database( "Connect Dialog", "MY_BIG_CLASS" );
Open Database( "Connect Dialog", "DROP TABLE BIGCLASS.MY_BIG_CLASS" );
```

Note: When you import data from an ODBC database, a table variable is added that can contain user ID and password information. To prevent this from happening, set the following preference: `pref(ODBC Hide Connection String(1))`. You can also select the **File** menu (Windows) or **JMP** menu (macOS), **Preferences > Tables,** and then select **ODBC Hide Connection String**. See *Using JMP*.

Creating a Database Connection and Executing SQL

You can use the following functions to handle more complex database operations:

```
db = Create Database Connection( "Connection String With Password" );
Execute SQL( db, "SQL statement", <invisible>, <"New Table Title"> );
Close Database Connection( db );
```

Using these three functions, you can open a connection, call `Execute SQL` several times, and then close the connection. `Create Database Connection` returns a handle for use in `Execute SQL` and `Close Database Connection`.

Depending on the SQL submitted, a table might or might not be returned. A SELECT statement typically returns a JMP table. INSERT INTO would not return a table, because it is modifying one in the database.

Examples

Open a connection to your database:

```
dbc = Create Database Connection(
  "DSN=dBASE Files;DBQ=$SAMPLE_IMPORT_DATA/;"
);
```

Execute one or more SQL statements using this connection:

```
dt = Execute SQL( dbc,
  "SELECT HEIGHT, WEIGHT FROM Bigclass", "NewTable"
);
```

When you are finished, close your connection.

```
Close Database Connection( dbc );
```

You might want to create a database connection without selecting a table. The script would open the Database Open window with the connection already showing. We strongly recommend that you encrypt this script. Here's an example:

```
dbc = Create Database Connection(
  "DSN=dBASE Files;DBQ=$SAMPLE_IMPORT_DATA/;UID=MyDatabase;pwd=MyPassword");
Main Menu( "Open Table" );
```

Note: `Execute SQL()` creates a connection string that includes the cleartext password in the resulting data table. We encourage you to create the SQL query in Query Builder and then run the query in JSL. Then the password is not included in the data table.

Write a SQL Query

The `Query()` function enables you to manipulate JMP data tables using SQL statements. The SQL statement is the only required argument to `Query()`. However, most interesting SQL statements refer to one or more data tables. Any data table that the SQL statement refers to must be passed in using a *table reference* argument.

A table reference argument can be a data table reference (`dt` in the following example):

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
result = Query( dt, "SELECT * FROM 'Big Class' WHERE age > 14;" );
```

The SQL statement in the preceding example includes the name of the JMP data table. If the data table path is long, you might want to use an alias for the data table in the SQL statement. To use an alias, pass the table reference as a `Table()` argument:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
result = Query( Table( dt, "t1" ), "SELECT * FROM t1 WHERE age > 14;" );
```

Using `Table()` to alias your data tables also enables you to write SQL statements for multiple queries with tables that have different names without rewriting the statements.

Private and Hidden Data Tables

If you want the data table that is produced by `Query()` to be either private or hidden, you can pass either `Private` or `Invisible` as arguments to `Query()`.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
result = Query( dt, invisible, "SELECT * FROM 'Big Class' WHERE age > 14;" );
```

Returning a Single Value

The `Query()` function normally returns the result of the query as a JMP data table. However, sometimes you might write a SQL query that returns a single value; you would rather have `Query()` return that value instead of placing the value in a JMP data table. You can do that by passing the `Scalar` argument.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
result = Query(
  Table( dt, "t1" ),
  Scalar,
  "SELECT AVG(height) FROM t1 WHERE age > 14;"
);
```

The preceding query returns 65, the average height of students older than 14.

Work with SAS

Make a SAS DATA Step

Sending `Make SAS Data Step` to a data table returns the text for a SAS DATA Step that can re-create the data table in SAS. For example,

```
dt << Make SAS Data Step
```

prints a DATA Step to the log that can be used in the SAS Program Editor.

Sending `Make SAS Data Step Window` produces this code in a window with a .SAS suffix, so that it can be easily sent to SAS.

Create SAS DATA Step Code for Formula Columns

Sending `Get SAS Data Step for Formula Columns` to a data table includes column formulas in the SAS data step code. Here is an example that outputs the formula for the Ratio column:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Ratio", Formula( :height / :weight ));
dt << Get SAS Data Step for Formula Columns;
```

This script output the following code to the log:

```
/*%PRODUCER: JMP - DataTable Formulas */
/*%TARGET: Ratio */
/*%INPUT: height */
/*%INPUT: weight */
/*%OUTPUT: Ratio */
/* Code to score Ratio */
Ratio =height/weight
drop ;
```

To get formulas for all columns, omit the column names as shown here:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Get SAS Data Step for Formula Columns;
```

You can also include column formulas in scoring code for SAS Model Manager. Send `Get MM SAS Data Step for Formula Columns` to the data table.

```
dt = Open( "$SAMPLE_DATA/Tiretread.jmp" );
dt << Get MM SAS Data Step for Formula Columns;
```

The results of this script are also shown in the log.

As with `Get SAS Data Step for Formula Columns`, specifying column names is optional.

SAS Variable Names

`SAS Open For Var Names()` opens a SAS data set only to obtain the names of its variables, returning those names as a list of strings.

The rules for SAS variable names are more strict than those of JMP. The `SAS Name` function converts JMP variable names to SAS variable names, changing special characters and blanks to underscores, and various other transformations to produce a valid SAS name.

```
result = SAS Name(name);
result = SAS Name({list of names});
```

If the argument is a list of names, the result is a blank-separated character string of names. For example,

```
SAS Name({“x 1”, “x 2”})
```

produces

```
“ x_1 x_2”
```

Connect to a SAS Metadata Server

You can connect to a SAS server and work directly with SAS data sets. Making connections and interacting with SAS data sets is scriptable through JSL. For information about connecting to a metadata server in the JMP interface, see *Using JMP*.

Note: If you connect to a physical workspace server, there is no metadata server involved, so metadata security is never applied. You must connect to a SAS Metadata Server and then connect to a logical workspace server. Then metadata security is enforced on the metadata-defined libraries you access.

Make the Connection

First, use the `Meta Connect()` function to connect to a SAS metadata server:

```
connected = Meta Connect( "MyMetadataServer", port );
```

To specify the SAS version for the metadata server, use the `SASVersion` named argument:

```
connected = Meta Connect( "MyMetadataServer", port, SASVersion("9.4") );
```

If you supply only the machine name (for example, myserver.mycompany.com) and the port, you are prompted to provide the authentication domain, your user name, and your password. You can also specify all that in JSL:

```
connected = Meta Connect( "MyMetadataServer", port, "authdomain", "user
  name", "password" );
```

When you are finished using the SAS metadata server, use `Meta Disconnect()` to disconnect the connection. No arguments are necessary; the command closes the current metadata server connection.

You can see the repositories that are available on a metadata server and set the one that you want to use:

```
Meta Get Repositories();
   {"Foundation"}
Meta Set Repository( "Foundation" );
```

Note that if there is only one repository available, it is selected automatically and you do not need to explicitly set it.

Once your repository is set, you can view the servers that are available:

```
mylist = Meta Get Servers();
   {"SASMain", "Schroedl", "SASMain_ja", "SASMain_zh", "SASMain_ko",
      "SASMain_fr", "SASMain_de", "SASMain_Unicode"}
```

Next, set your SAS connection. You can also use this command to connect directly to a local or remote server instead of using a metadata server.

```
conn = SAS Connect( "SASMain" );
```

Now you can send `Disconnect` and `Connect` messages to the `conn` object to close and open the SAS connection.

```
conn << Disconnect();
conn << Connect();
```

This is an example of using an object and messages with SAS server connections. You might have also connected to SAS servers using global functions. If so, the `Disconnect` and `Connect` messages do not affect those global connections. However, if there is no active global connection, that global connection is set to the connection opened by the object.

Automatically Connect SAS Libraries

When you connect to a SAS server, use `Connect Libraries` to automatically connect metadata-defined libraries.

```
conn = SAS Connect( "SASMain", Connect Libraries( 1 ) );
```

All metadata-defined libraries are connected, which can slow down your connection.

To connect specific libraries later, use the `SAS Connect Libref()` function or `Connect Libref` message to a SAS server object.

View SAS Libraries

After you connect to a SAS server, use the `Get Lib Ref()` function to view the libraries on that server:

```
conn << Get Lib Refs();
   {"BOOKS", "EGSAMP", "GENOMICS", "GISMAPS", "JMPSAMP", "JMPTEST",
   "MAILLIB", "MAPS", "OR_GEN", "ORION_RE", "ORSTAR", "SASHELP",
   "SASUSER", "TEMPDATA", "TSERIES", "V6LIB", "WORK", "WRSTEMP"}
```

If the library containing the data that you want is not assigned, assign it:

```
conn << Assign Lib Refs( "MyLib", "$DOCUMENTS/data" );
```

Open SAS Data Sets

First, assign a SAS library reference:

```
conn << Assign Lib Refs( "MyLib", "$DOCUMENTS" );
```

The first argument is any name that you want to use to refer to the library reference. The second is the path on the server where the data sets are located.

Next, get the list of data sets in the selected library:

```
datasets << Get Data Sets( "MyLib" );
   {"ANDORRA", "ANDORRA2", "ANYVARNAME", "BOOKS", "BOOKSCOPYNOT", "BOOKS_VIEW",
      "CATEGORIES", "DATETIMETESTS", "MOREUGLY", "NOTTOOUGLY", "PAYPERVIEW",
      "PUBLISHERS", "PURCHASES", "PURCHASES_FULL",
```

Now you can open a data set:

```
conn << Import Data( "MyLib", "PURCHASES" );
```

or

```
conn << Import Data( librefs[1], datasets[12] );
```

or

```
conn << Import Data( "MyLib.PURCHASES" );
```

Now you can get information about any SAS data set in that library. For example, you can get a list of variables:

```
conn << Get Var Names( "MyLib.PURCHASES" );
   {"purchaseyear", "purchasemonth", "purchaseday", "bookid", "catid",
   "pubid", "price", "cost"}
```

With that information, you can choose to import only part of the data set by specifying the variables to import.

```
conn << Import Data( librefs[1], datasets[12], columns( bookvars[1],
   bookvars[2], bookvars[4] ) );
```

Note: The name specified in `Import Data` can be a friendly metadata library name or a SAS 8-character library name.

Save SAS Data Sets

To save a JMP data table or an imported SAS Data Set, use the `SAS Export Data()` command:

```
conn << Export Data( dt, librefs[1], datasets[4], ReplaceExisting );
```

Run a Stored Process

To get a reference to a stored process:

```
stp = Meta Get Stored Process( "Samples/Stored Processes/Sample: Hello World"
  );
```

There is no way to acquire a list of stored processes through JSL; you must know the path to the stored process that you want to run.

To run it, send the stored process a message:

```
stp << Run();
```

Submit SAS Code from JMP

You can also directly submit SAS code and get back SAS results. For example:

```
conn << Submit( "proc print data=sashelp.class; run;" );
```

Two optional arguments control whether you see the output and the SAS log in JMP:

```
conn << Submit( "SAS Code" <,No Output Window(True|False)> <,Get
  Log(True|False )> );
```

You can also see the SAS Log at any time using the command

```
conn << Get Log();
```

`Get Log()` returns the contents, which can be placed in a JSL variable and used like any JSL string.

Get the Values of SAS Macro Variables

JMP provides several methods for querying SAS macro variables.

To show the `systime` value:

```
systime = sas << Get Macro Var("SYSTIME");
show(systime);
```

To show the defined SAS macro variables:

```
macro_names = sas << Get Macro Var Names();
show(macro_names);
```

To iterate through the SAS macro variables and print out the values:

```
macro_names = sas << Get Macro Var Names();
For( i = 1, i <= N Items( macro_names ), i++,
  macro_value = sas << Get Macro Var( macro_names[i] );
  output = macro_names[i] || " " || Char( macro_value );
  Show( output );
);
```

To submit SAS code that defines “test” as a SAS macro variable and then gets the value from SAS:

```
sas << Submit( "%let test = 1;" );
test = sas << Get Macro Var( "test" );
Show( test );
```

All macro variable values will evaluate to numbers, if possible, otherwise they will be characters.

Preferences

To get the current SAS version preference, use:

```
Get SAS Version Preference();
```

To set the current SAS version preference, use:

```
Preference( SAS Integration Settings( SASVersion( "9.4" ) ) );
```

Sample Scripts

The JMP Samples/Scripts/SAS Integration folder contains sample scripts. To run the stored process scripts successfully, the stored processes need to be placed on your SAS Metadata Server. The stored processes can be found in the sampleStoredProcesses.spk file, also in this folder.

To import sampleStoredProcesses.spk into your SAS Metadata Server

Caution: We recommend that you import these stored processes into a SAS Metadata Server that is used for testing rather than into a production system.

1. Run SAS Management Console.

2. Connect to your SAS Metadata Server using an account with administrative privileges.
3. Expand the BI Manager node in the left pane of the SAS Management Console.
4. Navigate to the folder in the tree under which you would like to create the imported sample stored processes.
5. Right-click that folder in either the left pane or the right pane of the SAS Management Console and select **Import**.

   The Import Wizard appears.
6. Enter the full path to sampleStoredProcesses.spk or use the **Browse** button to navigate to it.
7. Select **All Objects** in the Import Options section of the wizard.
8. Click **Next**.

   The next panel reports that during the import process, you must specify values for **Application servers** and **Source code repositories**.
9. Click **Next**.

   Select which of the application servers defined in your SAS Metadata Server that you would like to use to execute the imported stored processes.
10. Select an application server from the drop-down list under Target.
11. Click **Next**.

    Select the source code repository (directory) defined on your SAS Metadata Server where you would like the SAS code for the imported stored processes to be placed.
12. Select a source code repository from the drop-down list under Target Path.
13. Click **Next**.

    The next panel gives a summary of what occurs if you click **Import**.
14. Review the information about the panel, and if it looks correct, click **Import**.
15. During the Import process, you might be asked to provide login credentials for connecting to the metadata server to perform the import. Provide credentials with administrative privileges and click **OK**.

After the import completes, you will find a folder named BIP Tree under the folder that you imported the stored processes into. Under BIP Tree is a folder named JMP Samples, and in the JMP Samples folder are two sample stored processes: Shoe Chart and Diameter.

Please note that the paths to the sample stored processes needs to be adjusted in the sample scripts storedProcessHTML.jsl and storedProcessJSL.jsl to match the folder into which you imported the sample stored processes. Otherwise, these scripts will not work correctly.

Work with MATLAB

MATLAB, a product of MathWorks Inc., provides an interactive working environment for analyzing and visualizing computational models. MATLAB is available for Windows, Mac OS X, and Linux (64-bit). JMP supports MATLAB on both Windows and macOS platforms.

You can interact with MATLAB using JMP Scripting Language (JSL):

- Submit statements to MATLAB from within a JSL script.
- Exchange data between JMP and MATLAB.
- Display graphics produced by MATLAB.

See the *JSL Syntax Reference* for more information about using MATLAB functions in JMP.

Textual output and error messages from MATLAB appear in the log window.

Install MATLAB

MATLAB must be installed on the same computer as JMP. For the supported version of MATLAB, see the JMP website: https://www.jmp.com/system/.

How JMP Finds MATLAB

JMP delays loading MATLAB until a JSL-based script requires access to it. When you run a JSL script that calls MATLAB, JMP locates the software based on the operating system's `PATH` environment variable (for example, `C:\Program Files\MATLAB\R2012a\`).

Test Your Install

To verify that your computer is able to run JSL-based scripts using MATLAB:

1. Run the following JSL script:

```
MATLAB Init();
MATLAB Submit( "m = magic(3)" );
magicMat = MATLAB Get( m );
Show( magicMat );
MATLAB Term();
```

   The MATLAB function `M = magic(3)` returns a 3-by-3 matrix using integers in the range of 1 to 3^2 with equal row and column sums. This matrix is called a *magic square*.

2. Select **View > Log**.

   You should see the following response in the log window:

```
m =
```

```
      8     1     6
      3     5     7
      4     9     2


magicMat =
[ 8 1 6,
  3 5 7,
  4 9 2];
0
```

If you see the following message in the log window:

```
An installation of MATLAB cannot be found on this system.
```

1. Add a new environment variable with the `name` of `MATLABROOT` and a value of C:\Program Files\MATLAB\R2012a\ or C:\Program Files (x86)\MATLAB\R2012a\.

   Note: The path entered depends on the MATLAB installation path.

2. Verify that the `MATLAB` path is included in the `PATH` variable.
3. Run the script again to ensure JMP can access MATLAB.

Work with R

You can interact with R using JSL:

- Submit statements to R from within a JSL script.
- Exchange data between JMP and R.
- Display graphics produced by R.

Text output and error messages from R appear in the log window.

Installing R

Install R on the same computer as JMP. You can download R from the Comprehensive R Archive Network website:

https://cran.r-project.org

Note: For macOS Catalina and later versions, download the notarized and signed package.

Because JMP is a 64-bit application, you must install the corresponding 64-bit version of R. For the supported version of R, see the system requirements on the JMP website: https://www.jmp.com/system.

Override Default R Install Location on Windows

Normally JMP determines the R_HOME environment variable internally if it is not defined in the Windows system registry:

computer\HKEY_LOCAL_MACHINE\SOFTWARE\R-code\R\InstallPath

You might have installed more than one version of R on your computer. To override the default R installation location, define the R_HOME environment variable using either of the two following methods:

1. Create the variable in your system environment variables using the Control Panel, select **Start > Control Panel > System > Advanced system settings**.
2. Click **Environment Variables**.
3. In the System variables pane, click **New**.
4. Type R_HOME for the **Variable name**.
5. Type the path to the R .exe file (for example, C:\Program Files\R\R-2.15.3).
6. Click **OK** and click **OK** again to close the System Properties window.

or

Create the variable using the JSL `Set Environment Variable()` function:

```
Set Environment Variable( "R_HOME", "C:\Program Files\R\R-2.15.3" );
```

Override the Default R Install Location on macOS

You might have installed more than one version of R on your computer. To use an older version, change the R_HOME environment variable to point to the R version that you want to use.

To specify the R_HOME location, follow these steps:

1. Open a terminal window.
2. Type the following code, where *value* is the location of R.

   `export R_HOME=value`
3. Press Enter.

How JMP Finds R on Windows

JMP delays loading R until a JSL-based script requires access to it. When JMP needs to load R, it follows the standard steps for finding R on a Windows computer:

1. Look up the environment variable `R_HOME`.

   If the variable exists, load R from the specified directory.
2. If the environment variable `R_HOME` does not exist, look up the `InstallPath` value in the Windows registry under the following key:

   `HKEY_LOCAL_MACHINE\SOFTWARE\R-core\R`

   If the `InstallPath` value exists, load R from the specified directory.
3. If the `InstallPath` value does not exist, an error message states that R could not be found.

How JMP Finds R on macOS

JMP uses the value of the `R_HOME` environment variable if the variable exists. Otherwise, it looks in `/Library/Frameworks/R.framework/Versions/Current/Resources` for R.

Testing Your Setup

To test that your computer is able to run JSL-based scripts that use R, run the following JSL script:

```
R Init( );
R Submit( "
   x <- 1:5
   x
" );
R Term( );
```

You should see the following output in the log:

```
[1] 1 2 3 4 5
```

JMP to R Interfaces

The following JMP interfaces are provided to access R. The basic execution model is to first initialize the R connection, perform the required R operations, and then terminate the R connection. In most cases, these functions return 0 if the R operation was successful, or an error code if it was not. If the R operation is not successful, a message is written to the log. The single exception to this is `R Get()`, which returns a value.

R JSL Scriptable Object Interfaces

The R interfaces are also scriptable using an R connection object. A scriptable R connection object can be obtained using the `R Connect()` JSL function.

Conversion Between JMP Data Types and R Data Types

Table 14.3 shows what JMP data types can be exchanged with R using the `R Send( )` function. Sending lists to R recursively examines each element of the list and sends each base JMP data type. Nested lists are supported.

Table 14.3 Equivalent JMP and R Data Types for R Send( )

R Data Type	JMP Data Type
Double	Numeric
String	String
Double Matrix	Matrix
List	List
Data Frame	Data Table
Integer	Row State
Date and Time	Datetime
Time/Duration	Duration
Paired List	Associative Array

Example

```
R Init();
X = 1;
R Send( X );
S = "Report Title";
R Send( S );
M = [1 2 3, 4 5 6, 7 8 9];
R Send( M );
R Submit( "
X
S
M
" );
```

```
R Term();
```

Table 14.4 shows what JMP data types can be exchanged with R using the `R Get()` function. Getting lists from R recursively examines each element of the list and sends each base R data type. Nested lists are supported.

Table 14.4 Equivalent JMP and R Data Types for R Get( )

R Data Type	JMP Data Type
Double	Numeric
Logical (Boolean)	Numeric (0 | 1)
String	String
Integer	Numeric
Date and Time	Datetime
Time/Duration	Duration
Factor	Expanded to a list of Strings or a Numeric matrix
Data Frame	Data Table
List	List of converted R data types
Matrix	Numeric Matrix
Numeric Vector	Numeric Matrix
String Vector	List of Strings
Graph	Picture
Time Series	Matrix
Paired List	Associative Array

JMP Scoping Operators and R

A JMP object sent to R using `R Send()` uses the same JMP reference as the name of the R object that gets created. For example, sending the JMP variable `dt` to R creates an R object named `dt`. The colon and double colon scoping operators (`:` and `::`) are not valid in R object names, so these are converted as follows:

- A single colon scoping operator is replaced with a period (`.`).

  For example, sending `nsref:dt` to R creates a corresponding R object named `nsref.dt`.

- A double colon scoping operator (designating a global variable) is ignored.

  For example, sending `::dt` to R creates a corresponding R object named `dt`.

Using R Name() with R Send()

The `R Name()` option to `R Send()` has an argument that is a quoted string that contains a valid R object name. The JMP object sent to R becomes an R object with the name specified. For example:

```
R Send( jmp_var_name, R Name( "r_var_name") );
R Submit( "print(r_var_name)" )
```

Example

This example creates a variable `x` in the `Here` namespace, a variable `y` in the global namespace, and a variable `z` that is not explicitly referenced to any namespace. The variable `z` defaults to Global unless `Names Default To Here(1)` is on. These variables are then passed to R.

```
Here:x = 1;
::y = 2;
z = 3;

R Init(); // initiate the R connection

/* send the Here variable to R
Here:x creates the R object Here.x */
R Send( Here:x );

/* note that the JMP log labels the output with the original JMP variable
   reference Here:x */
R Submit( "print(Here.x)" );

R Send( ::y ); // ::y create the R object y
R Submit( "print(y)" );

// to use a different name for the R object, use the R Name() option
R Send( Here:x, R Name( "localx" ) );

/* The R Name option to the R Send() command creates the R object named
   "localx"" corresponding to the JMP variable "Here:x". Again the log shows
   the original corresponding JMP variable name. */
R Submit( "print(localx)" );
R Send( z ); // z creates the R object z
R Submit( "print(z)" );
```

Troubleshooting

Recording Output

On Windows, if you want to record output to the graphics window, send the following R code using `R Submit( )`.

```
windows.options( record = TRUE );
```

Character Vectors

A JMP list of strings is not the same as an R character vector. If you send a list of strings to R, it becomes a list of strings in R, not a character vector. You can use the R function Unlist to convert it:

```
R Init();
X = {"Character", "JMP", "List"};
R Send( X );

/* R output is:
[1] "list"
*/
R Submit( "class(X)" );

/* Object Y is now a character vector. The R output is:
[1] "character"
*/
R Submit( "Y<-unlist(X)
     class(Y)" );
R Term();
```

Element Names

A feature of an R list (called attributes) lets you associate a name with each element of the list. You can use the name to access that element instead of having to know the position of it in the list

In the following example, the list that is created in R has two elements named x and y that are created using the `List()` function of R. When your bring the R list into JMP and then send it back to R, the names are lost. Therefore in R, you cannot access the first matrix using `pts$x`. Instead, you must use the index using `pts[[1]]`.

```
R Init();
R Submit("
   pts <- list(x=cars[,1], y=cars[,2])
   summary(pts)
");
```

```
JMP_pts = R Get( pts );

R Send( JMP_pts );
R Submit("
   Summary( JMP_pts )
");
R Term();
```

Examples

Sending a Data Table to R

This example initiates an R connection, sends a data table to R, prints it to the log, and closes the R connection.

```
R Init();
dt = Open( "$SAMPLE_DATA/Big Class.jmp", invisible );
R Send( dt ); // send the opened data table represented by dt to R;
R Submit( "print( dt )" );
R Term();
```

Creating Objects in R

This example initiates an R connection, creates an R object, retrieves the object into JMP, and closes the R connection.

```
R Init();
R Submit(
   "
   L3 <- LETTERS[1:3]
   d <- data.frame(cbind(x=1, y=1:15), Group=sample(L3, 15, repl=TRUE))
"
);
R Get( d ) << NewDataView;
R Term();
```

Using R Functions and Graphics

This example initiates an R connection and plots the normal density function in R using the R graphics device,. Then the graph is retrieved from R and displayed in JMP. Finally, the R connection is closed.

```
R Init();
R Submit( "\[plot(function(x) dnorm(x), -5, 5, main = "Normal(0,1) Density")
   ]\" );
```

```
picture = R Get Graphics( "png" );
New Window( "Picture", picture );
Wait( 10 );
R Term();
```

Simple Matrix Addition in R

This example initiates an R connection, sends a matrix to R, creates a matrix in R, adds them together, returns the new matrix to JMP, and closes the R connection.

```
R Init();
X = J( 2, 2, 1 );
R Send( X );
R Submit(
  "
  X                                    #prints X to the log
  Y <- matrix(1:4, nrow=2, byrow=TRUE)  #makes a 2x2 matrix object Y
  Y                                    #prints Y to the log
  Z <- X + Y #matrix object Z is addition of X and Y
"
);
Z = R Get( Z );
R Term();
Show( Z );
```

A Bootstrap Sample

See the file JMPtoR_bootstrap.jsl in the sample scripts folder for an example script.

This script performs a bootstrap simulation by using the JMP to R Project integration.

The script produces a JMP window that asked the user to specify the variable to perform bootstrapping over. Then the user selects a statistic to compute for each bootstrap sample. Finally, the data is sent to R using the R interface in JSL.

The boot package in R is used to call the `boot()` function and the `boot.ci()` function to calculate the sample statistic for each bootstrap sample and the bootstrap confidence interval.

The results are brought back to JMP and displayed using the JMP Distribution platform.

Work with Python

The JMP-Python integration allows JMP users to interact with a large amount of scientific, statistical, and data analysis tools written in Python. JMP implements this capability by loading the Python shared library into the running JMP process. JSL gives Python access to JMP.

The JMP installer does not contain Python. Several Python implementations are available. Visit https://www.python.org/download/alternatives to view the options.

JMP has several methods for finding an instance of Python installed on the computer. You should not need to set any environment variables or additional arguments to the JSL `Python Init()` statement. See "Troubleshooting the JMP Python Integration" for more information.

The Python integration enables you to perform the following tasks:

- submit statements to Python from within a JSL script
- exchange data between JMP and Python
- display graphics produced by Python

Text output and error messages from Python appear in the log window.

Install Python

JMP 16 is a 64-bit program; only 64-bit Python is supported

- Python.org 3.6 through 3.9.1 distributions are supported on Windows and macOS. On Apple silicon, you will not have full functionality until third-party packages have been updated.
- Anaconda 3.6 distribution is supported on Windows only.
- Anaconda 3.7 distribution is not supported on Windows.
- The way Python is built from versions 3.6 to 3.7 changed. In Python 3.7, Python must be installed for all users. Otherwise, the libraries that the Python DLL depend on cannot be located at DLL load time. Other Python distributions might work fine installed for a single user or might also require a system-wide install.
- JMP does not currently support Python virtual environments.

Notes:

- Python is not distributed with JMP. Download Python from the Python web site at https://python.org.
- Several Python implementations are available. Visit https://www.python.org/download/alternatives to view the alternatives.

- On macOS, installing Python 3.6.x from https://python.org installs in /Library/Frameworks/Python.framework.

Required Python Packages

For the JMP-to-Python interfaces to work properly, you must install certain Python packages and their prerequisites. In many Python installations, such as Anaconda, these packages are already installed. If you install Python from Python.org, also install the following packages:

numpy https://www.numpy.org

pandas https://pandas.pydata.org

matplotlib https://matplotlib.org

scipy https://www.scipy.org

sqlite3 https://www.sqlite.org

PyQt5 https://pypi.org/project/PyQt5/

Use the standard Python update tools to install these packages for your Python configuration. You must use the correct pip-tool for your installation. The pip-tool name is dependent on the what product has installed Python. The installation instructions for Python have instructions on using pip tools.

Troubleshooting the JMP Python Integration

Sometimes JMP has trouble locating the correct version of Python. The first thing to try is running Python from the command line. If Python doesn't run from a terminal or console window, something is wrong with the Python installation.

Note: Before you proceed with troubleshooting, verify that you're running a supported version of Python. See "Install Python".

How JMP Finds Python on Windows

JMP delays loading Python until a JSL-based script requires access to it. When JMP loads Python, it needs to be able to find the Python shared library and any of the shared libraries that the Python DLL depends on. Once the Python interpreter is loaded, Python needs to be able to find the distribution's standard Python modules. To guide this process, JMP uses registry entries, environment variables, and user-specified paths.

Python also looks for certain environment variables. Avoid using any user-specified paths or environment variables, and let JMP find Python in the registry. The following methods help JMP locate Python if the automatic location doesn't work:

- JMP_PYTHON_MODULE_PATH is a JMP-specific version of PYTHONPATH. This variable is the replacement for Python Sys Path. In JMP_PYTHON_MODULE_PATH, specify standard OS paths. Separate the paths with a colon on macOS. Separate the paths with a semicolon on Windows.
- JMP_LIB_PYTHON_PATH is a JMP-specific version of PYTHONHOME. The variable contains the full path to the Python shared library, or the parent directory to the shared library.

Windows finds Python in the following order of precedence:

1. JMP looks at the environment variables Python Sys Path (deprecated), PYTHONINSTALLPATH (deprecated), and JMP_PYTHON_MODULE_PATH, in that order. JMP uses the first variable that is found.
2. JMP checks JMP_LIB_PYTHON_PATH for a path to the Python shared library.
3. JMP checks for options that are specified in the Python Init() or Python Connect() function calls. Path() corresponds to the location of the Python shared library. Use Python Version() allows you to specify the version of Python to look for.

   Python Sys Path provides the mechanism to set Python's sys.path variable. If present, script parameters take precedence over the environment variables. There is one caveat. After JMP loads the Python shared library, the library is unloaded only when JMP shuts down. As a result, the Path() portion is only effective on the first load of Python, and does not allow for running different versions of Python in different scripts without shutting down JMP and restarting.
4. JMP reads the Windows registry to determine if Python has been installed. It looks at appropriate locations based on whether the installation was for an individual user or for all users on the Windows machine. If you have specified a Python version, other versions of Python will not be considered.

If JMP cannot locate Python, try the following option in the Python Init() expression:

```
Python Init( Init Trace( "TRUE" ) );
```

JMP prints details to the log about where it detects Python, including paths and registry entries for Python.

Typically, JMP determines the equivalent to the PYTHONHOME environment variable automatically if it is defined in the Windows registry.

Computer\<root>\SOFTWARE\Python\PythonCore\<version>\InstallPath

<root> is either the root key HKEY_LOCAL_MACHINE or HKEY_CURRENT_USER depending on where and how you installed Python. *<version>* is the Python version number.

If multiple versions of Python are installed on your computer, you must specify which installation for JMP to use.

- The easiest method is to specify the version using the Use Python Version() argument in Python Init() or Python Connect() functions.

- Use the `Path("path to an installed Python DLL or shared library")` argument with the `Python Init()` or `Python Connect()` function.
- Define the `JMP_LIB_PYTHON_PATH` environment variable as described below.

Change the Default Python Install Location on Windows

Create the variable in your system environment variables using the Control Panel.

1. Select **Start > Control Panel > System > Advanced system settings**.
2. Click **Environment Variables**.
3. In the System variables pane, click **New**.
4. Type `JMP_LIB_PYTHON_PATH` for the Variable name.
5. Type the path to the Python<version>.dll file, for example:
   `C:/Program Files/Python/<version>/Python<version>.dll`
6. Click **OK** and click **OK** again to close the System Properties window.

macOS: Override How JMP Finds Python

Locating Python on the Mac works the same as on Windows except for the step of looking in the registry. JMP looks in a couple of typical installation directories so see if it can locate Python. Python typically installs in `/Library/Frameworks/Python.framework`. A compatible version of Python located in the standard system location should work without any changes to environment variables or parameters passed to JSL.

If Python is installed in an alternate location, set configuration variables in your shell environment as needed. JMP on macOS uses the same environment variables as described in "How JMP Finds Python on Windows".

`JMP_LIB_PYTHON_PATH` is the path to the Python shared library `libpython3.x.dylib`, where *x* is the minor version. If environment variables are needed, you typically only need to set `JMP_LIB_PYTHON_PATH`.

If JMP is unable to find install Python packages, you might also need to set the variable `JMP_PYTHON_MODULE_PATH`. The `JMP_PYTHON_MODULE_PATH` contains the same paths that the Python call print(sys.path) outputs. When you set an environment variable, use host OS path conventions. When the variable is passed in JSL, use a JSL list.

One way is to have Python tell you what it thinks its `sys.path` should be. In the example in this section, the same Python commands can be run on Windows to determine `sys.path`.

In the terminal or console window, enter the following boldface commands to get a string that specifies the location of the Python packages:

```
MDz320:~ shlori$ python --version
Python 3.6.4 :: Anaconda, Inc.
MDz320~ shlori$ python3
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 12:04:33) [GCC 4.2.1
  Compatible Clang 4.0.1 (tags/RELEASE_401/final)] on darwin Type "help",
  "copyright", "credits" or "license" for more information.
>>> import sys
>>> print(sys.path)
  ['', '/anaconda3/lib/python36.zip', '/anaconda3/lib/python3.6',
     '/anaconda3/lib/python3.6/lib-dynload',
     '/anaconda3/lib/python3.6/site-packages',
     '/anaconda3/lib/python3.6/site-packages/aeosa']
```

Test Your Setup

To test that your computer is able to run JSL scripts that use Python, run the following script:

```
Python Init();
Python Submit( "\[basket = ['apple', 'orange', 'pear']
print(basket)]\" );
Python Term();
```

You should see the following output in the log:

```
['apple', 'orange', 'pear']
0
```

JMP-to-Python Interfaces

JMP provides the interfaces to access Python that are described in the following sections. The basic execution model is to first initialize the Python connection, perform the required Python operations, and then terminate the Python connection. In most cases, these functions return 0 if the Python operation was successful or an error code if it was not. If the Python operation is not successful, a message is written to the log. The single exception to this is `Python Get()`, which returns a value.

Python JSL-Scriptable Object Interfaces

The Python interfaces are also scriptable using a Python connection object. A scriptable Python connection object can be obtained using the `Python Connect()` function.

```
// return a Python connection scriptable object
PythonConnection = Python Connect();

// return 1 if a Python connection is active. Otherwise, returns 0.
x = PythonConnection << Is Connected;
Show( x );
PythonConnection << Disconnect;
   x = 1;
   0
```

Equivalent Data Types for Python Send()

Table 14.5 shows what JMP data types can be exchanged with Python using the `Python Send()` function. Sending lists or associative arrays to Python recursively examines each element of the list or associative array and sends each base JMP data type. Nested lists and associative arrays are supported.

Table 14.5 Equivalent JMP and Python Data Types for Python Send()

JMP Data Type	Python Data Type
Boolean	Boolean
Data Table	Pandas.DataFrame
Associative Array	Dictionary

Table 14.5 Equivalent JMP and Python Data Types for Python Send() *(Continued)*

JMP Data Type	Python Data Type
Numeric	Float
Matrix	Float Matrix
List	List
String	Unicode String

Python Send() Example

```
Python Init();
X = 1;
Python Send( X );
S = "Report Title";
Python Send( S );
M = [1 2 3, 4 5 6, 7 8 9];
Python Send( M );
Python Submit( "\[
print(X)
print(S)
print(M)
]\" );
Python Term();
```

```
1.0
Report Title
[[ 1.  2.  3.]
 [ 4.  5.  6.]
 [ 7.  8.  9.]]
0
```

Equivalent Data Types for Python Get()

Table 14.6 shows which JMP data types can be exchanged with Python using the `Python Get()` function. Getting lists from Python recursively examines each element of the list and gets each base Python data type. Nested lists are supported.

Table 14.6 Equivalent Python and JMP Data Types for Python Get()

Python Data Type	JMP Data Type
Array	Numeric Matrix
Date and time	Numeric

Table 14.6 Equivalent Python and JMP Data Types for Python Get() *(Continued)*

Python Data Type	JMP Data Type
Pandas.DataFrame	Data Table
Dictionary	Associative Array
Graph	Picture Object
Integer	Integer
List	List of converted Python data types
Matrix	Numeric Matrix
Set	List
Unicode String	String

Python Get() Example

```
Python Init();
x1 = [1, 2, 3];
Python Send( x1 );
x2 = Python Get( x1 );
Show( x1, x2 );
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
Python Send( dt1 );
dt2 = Python Get( dt1 );
dt2 << New Data View;
Close( dt1 );
Python Term();
```

JMP Scoping Operators and Python

A JMP object sent to Python with `Python Send()` uses the same JMP reference as the name of the Python object that gets created. For example, sending the JMP variable `dt` to Python creates a Python object named `dt`.

The colon and double colon JSL scoping operators (`:` and `::`) are not valid in Python object names, so the colons are converted as follows:

- A single colon scoping operator is replaced with an underscore (_).

  For example, sending `nsref:dt` to Python creates a corresponding Python object named `nsref_dt`.
- A double colon scoping operator (designating a global variable) is ignored.

For example, sending `::dt` to Python creates a corresponding Python object named `dt`.

Note: The `Python Name` option for `Python Send()` has a quoted string argument that contains a valid Python object name. The JMP object sent to Python becomes a Python object with the specified name.

Python Name() and Python Send() Examples

The following example creates the `jmp var name` variable, assigns it to the Python name `python_var_name`, submits the `print` statement, and closes the connection.

```
Python Init();
jmp var name = 25;
Python Send( jmp var name, Python Name( "python_var_name" ) );
Python Submit( "print(python_var_name)" );
Python Term();
```

```
25.0

0
```

The following example creates a variable *x* in the Here namespace, a variable *y* in the global namespace, and a variable *z* that is not explicitly referenced to any namespace. The variable *z* defaults to Global unless `Names Default To Here(1)` is on. These variables are then passed to Python.

```
Here:x = 1;
::y = 2;
z = 3;

Python Init(); // initiate the Python connection

/* send the Here variable to Python
Here:x creates the Python object Here_x */
Python Send( Here:x );

Python Submit( "print(Here_x)" );

Python Send( ::y ); // ::y creates the Python object y
Python Submit( "print(y)" );

// to use a different name for the Python object, use the Python Name() option
Python Send( Here:x, Python Name( "localx" ) );

/* The Python Name option to the Python Send() command creates the Python
   object named "localx", which corresponds to the JMP variable "Here:x". */
Python Submit( "print(localx)" );
```

```
Python Send( z ); // z creates the Python object z
Python Submit( "print(z)" );
Python Term();
  1.0

  2.0

  1.0

  3.0

  0.0
```

Additional Python Integration Examples

Send a Data Table to Python

This example initiates a Python connection, sends a data table to Python, prints it to the log, and closes the Python connection.

```
Python Init();
dt = Open( "$SAMPLE_DATA/Big Class.jmp", invisible );
Python Send( dt ); // send the opened data table represented by dt to Python
Python Submit( "print( dt )" );
Python Term();
          name     age   sex   height   weight
  0       KATIE    12    F       59       95
  1      LOUISE    12    F       61      123
  2        JANE    12    F       55       74
  ...
  0
```

Create Objects in Python

This example initiates a Python connection, creates a Python object, retrieves the object into JMP, and closes the Python connection.

```
Python Init();
Python Submit("\[
import numpy as np
import pandas as pd
# Generate basic series
ss = pd.Series(np.random.randn(5), index=['a', 'b', 'c', 'd', 'e'])
print(ss)
sn = pd.Series(np.random.randn(5))
print(sn)

# Generate Data Frames from Series
```

```
s1 = pd.Series(np.random.randn(5), index=['a', 'b', 'c', 'd', 'e'])
print(s1)
dfs1 = pd.DataFrame(s1)
print(dfs1)

d1 = {'one' : pd.Series([1., 2., 3.], index=['a', 'b', 'c']),
      'two' : pd.Series([1., 2., 3., 4.], index=['a', 'b', 'c', 'd'])}
print(d1)
dfd1 = pd.DataFrame(d1)
print(dfd1)

d2 = {'one' : [1., 2., 3., 4.],
      'two' : [4., 3., 2., 1.]}
print(d2)
dfd2 = pd.DataFrame(d2)
print(dfd2)
]\"
);
ss = Python Get( ss );
Show( ss );
sn = Python Get( sn );
Show( sn );
dfs1 = Python Get( dfs1 );
dfs1 << New Data View;
dfd1 = Python Get( dfd1 );
dfd1 << New Data View;
dfd2 = Python Get( dfd2 );
dfd2 << New Data View;
Python Term();
   a   -0.384649
   b    0.455864
   c   -2.214405
   d    0.260955
   e    1.342286
   dtype: float64
   ...
```

Matrix Operations in Python

In JMP, a simple matrix looks like this:

```
xx = [1 2 3, 4 5 6, 7 8 9 ];
```

In numpy, the matrix is an array:

```
xx = numpy.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
```

In order to submit a matrix as a numpy array, you must translate the JMP matrix.

```
q = "new = numpy.array ([";
```

```
For( i = 1, i <= N Items( xx ), i++,
   zz = Matrix( xx[i] );
   q = q || Char( zz ) || ",";
);
q = q || "])";
Show( q );

Python Submit( "import numpy
" || q );
Python Submit( "print(new)" );
```

Work with Microsoft Excel

You can script the Profiler interface to Microsoft Excel, although not the Transfer to JMP interface. Here is the basic syntax:

```
excel_obj = Excel Profiler(
   Workbook( "excel_workbook_path" ),
   Model( "name_of_model" )
);
```

The `Model` argument is optional. If you supply only the workbook, you are prompted to select the model. You do not need to specify the worksheet, because the model is found no matter where it is located.

Once you have an Microsoft Excel Profiler object, you can send the following messages to run the JMP prediction profiler using the Microsoft Excel model. For example:

```
excel_obj <<Prediction Profiler( 1 );
```

Parse XML

JSL has several commands available to parse XML.

`Parse XML( `*`string`*`, On Element( "`*`tagname`*`", Start Tag( `*`expr`*` ), End Tag( `*`expr`*` ) ) );`

parses an XML expression using the `On Element()` expression for specified XML tags.

`value = XML Attr( "`*`attribute name`*`" );`

extracts the string value of an XML argument when evaluating a `Parse XML()` expression.

```
value = XML Text();
```

extracts the string text of the body of an XML tag when evaluating a `Parse XML()` expression.

Example of Parsing XML

Suppose that a Microsoft Excel file contains one row of data from Big Class.jmp. The file is saved as the valid XML document BigclassExcel.xml, shown here and also saved in the JMP Samples/Import Data folder.

```
<?xml version="1.0" encoding="UTF-8"?>
<Workbook xmlns="urn:schemas-microsoft-com:office:spreadsheet"
 xmlns:o="urn:schemas-microsoft-com:office:office"
 xmlns:x="urn:schemas-microsoft-com:office:excel"
 xmlns:ss="urn:schemas-microsoft-com:office:spreadsheet"
 xmlns:html="http://www.w3.org/TR/REC-html40">
 <Worksheet ss:Name="Bigclass">
  <Table ss:ExpandedColumnCount="5" ss:ExpandedRowCount="41"
   x:FullColumns="1"
   x:FullRows="1">
   <Row>
    <Cell><Data ss:Type="String">name</Data></Cell>
    <Cell><Data ss:Type="String">age</Data></Cell>
    <Cell><Data ss:Type="String">sex</Data></Cell>
    <Cell><Data ss:Type="String">height</Data></Cell>
    <Cell><Data ss:Type="String">weight</Data></Cell>
   </Row>
   <Row>
    <Cell><Data ss:Type="String">KATIE</Data></Cell>
    <Cell><Data ss:Type="Number">12</Data></Cell>
    <Cell><Data ss:Type="String">F</Data></Cell>
    <Cell><Data ss:Type="Number">59</Data></Cell>
    <Cell><Data ss:Type="Number">95</Data></Cell>
   </Row>
  </Table>
 </Worksheet>
</Workbook>
```

The following script reads BigclassExcel.xml and creates a JMP data table with the information in it. This script, named ParseXML.jsl, is in the JMP Samples/Scripts folder.

```
file contents = Load Text File( "$SAMPLE_IMPORT_DATA/BigclassExcel.xml" );
Parse XML( file contents,
   OnElement(
      "urn:schemas-microsoft-com:office:spreadsheet^Worksheet",
      StartTag(
         sheetname = XML Attr(
            "urn:schemas-microsoft-com:office:spreadsheet^Name",
            "Untitled"
         );
         dt = New Table( sheetname );
         row = 1;
```

```
            col = 1;
         )
      ),
      OnElement(
         "urn:schemas-microsoft-com:office:spreadsheet^Row",
         StartTag(
            If( row > 1, // first row has column names
               dt << Add Rows( 1 )
            )
         ),
         EndTag(
            row++;
            col = 1;
         )
      ),
      OnElement( "urn:schemas-microsoft-com:office:spreadsheet^Cell", EndTag(
      col++ ) ),
      OnElement(
         "urn:schemas-microsoft-com:office:spreadsheet^Data",
         EndTag(
            data = XML Text( collapse );
            If( row == 1,
               New Column( data, Character( 10 ) ), // first row has column names
               Column( col )[row - 1] = data // and other rows have data
            );
         )
      )

   );
```

OLE Automation

Most of JMP can be driven through OLE automation. Please see the Automation Reference.pdf at https://www.jmp.com/en_us/support/jmp-documentation.html for more information about automating JMP. This document introduces how to automate JMP through Visual Basic and using Visual C++ with MFC. It also contains more information about the methods and properties that JMP exposes to automation clients like Visual Basic and Visual C++.

The JMP Samples/Automation folder contains several sample Visual Basic .Net, and Visual C# .Net files.

Share Reports on JMP Live

Write JSL scripts to automate the generation of JMP analyses and publish them to JMP Live.

- "JMP Live Prerequisites"
- "Get an API Key"
- "Publish a Single Report to JMP Live"
- "Publish Several Reports to a JMP Live Folder"
- "Delete Reports in a JMP Live Folder"
- "Replace the Contents of a JMP Live Report"
- "Search for a JMP Live Report or Folder ID"
- "Update Data in JMP Live Reports"
- "Work with Other Content in JMP Live Reports"
- "HTTP Requests in JMP Live"

JMP Live Prerequisites

Before you can use JSL to publish reports to JMP Live, do the following:

1. Get an API key (if required by a JMP Live administrator). See "Get an API Key".
2. Establish a connection to a JMP Live server. See *Using JMP*. When you add a JMP Live server, you need to enter your API key (if required).

   When you establish a connection, a JMP Live Connection is created in JSL:

   ```
   liveconnection = New JMP Live( Connection( "My JMP Live" ), Prompt( If Needed
     ));
   ```

   where `"My JMP Live"` is the name of the JMP Live server connection you created. If authentication is required, an interactive prompt appears. You can change the `Prompt` to `No`, but if authentication is required, the connection fails.

Get an API Key

Before you can script reports and publish them to JMP Live, you might need an API key. JMP Live administrators determine whether an API key is required for scripting.

An API key is unique for a user and a JMP Live server. That means that if you obtain a new API key, it overrides the previous one, and any scripts using the previous API key will need to be updated to use the new API key.

To obtain an API key:

1. Log into JMP Live.
2. At the top, click *your user icon* and select **Settings**.
3. Click **Generate API Key**. Make a note of your key.

Publish a Single Report to JMP Live

Publish a single report to JMP Live using a JMP Live Report.

The following example creates a Bivariate report and an empty JMP Live Report called `webreport`, adds the Bivariate report to the `webreport`, and publishes the `webreport` to JMP Live:

```
dt = Open( "$SAMPLE_DATA\Big Class.jmp" );
// Create a Bivariate report.
biv = Bivariate( X( :Height ), Y( :Weight ));
// Create an empty JMP Live Report.
webreport = New Web Report();
/* Add the Bivariate report to the webreport. A title is required and a
   description is optional.*/
webreport << Add Report( biv, Title( "Bivariate Report" ), Description( "A
   test of the new scripting" ));
/* Publish the webreport to JMP Live as a private report. Use the JMP Live
   Connection you established and return a JMP Live object called liveresult.
   */
liveresult = liveconnection << Publish( webreport );
```

You can use JSL to make the report visible to everyone or share the report with groups in JMP Live. See the `Set Public` and `Share With Groups` messages in the JMP Scripting Index at Help > Scripting Index.

Work with a JMP Live Result

The `Publish` function returns a JMP Live Result (`liveresult`). Use JSL scripting functions on a JMP Live Result to get information, such as whether the publishing succeeded or failed, failure codes, and more:

```
// Return the outcome of the last action (1 for success, 0 for failure).
liveresult << Succeeded ();
// Return any error messages from the last action.
liveresult << Get Error Message();
```

The JMP Live Result can be made scriptable and turned into a JMP Live Report (for a single report) or a JMP Live Folder object (for multiple reports) so that you can send commands to it. The following example changes a JMP Live Result into a scriptable JMP Live Report and then changes the report access from private to public:

```
livereport = liveresult << As Scriptable();
livereport << Set Public( 1 );
```

See the JMP Live Report object in the JMP Scripting Index at Help > Scripting Index.

Publish Several Reports to a JMP Live Folder

Publish several reports to JMP Live by creating a JMP Live Folder and adding the JMP Live Report to the folder. A JMP Live Report object can hold a single report or multiple reports.

The following example creates a scriptable JMP Live Folder and a JMP Live Report called `webreport`, adds a Bivariate and a Oneway report to the `webreport`, and adds the reports to a folder:

```
/* Create a folder in JMP Live. A title is required and a description is
   optional. Use the JMP Live Connection you established and return a JMP Live
   object called liveresult. */
liveresult = liveconnection << Create Folder( Title( "Test Folder"
  ),Description( "Folder created for script testing" ));
// Make the liveresult scriptable.
folder = liveresult << As Scriptable();
// Create an empty JMP Live Report.
webreport = New Web Report();
// Add a Bivariate report to the webreport.
webreport << Add Report( biv, Title( "Bivariate Report" ));
// Add a Oneway report to the webreport.
webreport << Add Report( oneway, Title( "Oneway Report" ));
// Add the reports to the folder.
liveresult = folder << Add Reports To Folder( webreport );
```

You can use JSL to share the folder with JMP Live groups. See the `Share With Groups` object in the JMP Scripting Index at **Help** > **Scripting Index**.

Delete Reports in a JMP Live Folder

You can delete reports and folders using JSL. Folders must be empty before they can be deleted.

```
returncode = liveconnection << Delete Report( report << Get ID() );
returncode = liveconnection << Delete Folder( folder << Get ID() );
```

In both of these cases, a value of 1 (True) or 0 (False) will indicate if the operation was successful.

Replace the Contents of a JMP Live Report

You can replace the contents of an existing report rather than creating a new report with updated content. You need to know the ID of the report you want to replace. If you have the report object available, you can use the `Get ID( )` function to return the ID.

The following example creates a new JMP Live Report called `webreport2`, adds an updated Oneway report, and replaces an existing Oneway report using the report ID:

```
// Create an empty JMP Live Report.
webreport2 = New Web Report();
/* Add a Oneway report to webreport3. A title is required and a description is
   optional. */
webreport2 << Add Report( oneway, Title( "New Oneway Report" ), Description(
   "Replace a report" ));
/* Use the JMP Live Connection you established to replace the report with the
   specified ID with webreport2. */
liveconnection << Replace( webreport2, ID(report << Get ID() ));
```

Search for a JMP Live Report or Folder ID

For many JSL functions, you need to know the report or folder ID. Using JSL, you can search by keyword (in the title or description) or by the publisher of the report or folder. If your search may return many results, you can specify how many results to return at a time.

The following example searches all reports that contain the word "Production" in the title or description, and are published by a user with the display name "dianesmith". It also specifies that the search will return up to ten items at a time.

```
/* Search for reports with the keyword Production that are published by the
   user dianesmith. Return no more than 10 results at a time. */
liveresult = liveconnection << Find Reports( Search( "Production" ),
   Publisher( "dianesmith" ), ( PageSize( 10 ));
// Change the liveresult into JMP Live Reports and make them scriptable.
livereports = liveresult << As Scriptable();
// Show how many results are returned.
show( livereports << Get Number Of Items());
// Return the first report in list of results.
report1 = livereports[ 1 ];
// Show the title of the first report.
show( report1 << Get Title());
```

In the example above, if you might have more than ten results, you can page through them using the Next and Previous functions:

```
// Put the next 10 results into the next page.
nextpage = livereports << Next();
```

```
// Show the current page number.
show( livereports << Get Current Page Number());
// Get the first item in the list of items in the next page.
next1 = nextpage[ 1 ];
// Show the title of the first report.
show( next1 << Get Title());
// Show the number of items in the current page.
show( nextpage << Get Number Of Items());
```

If `Next()` finds no further pages, an error appears, which you should catch to proceed with further processing.

Tip: If you know the number of pages to expect, you can jump directly to a page. See the `Get Page` function in the JMP Scripting Index at **Help** > **Scripting Index**.

Update Data in JMP Live Reports

If the data behind a JMP Live report changes, you can update the data and regenerate the report. You can update the data for a standalone report (using the report ID) or for all the reports in a folder (using the folder ID).

The following example updates data for an individual report. Suppose that in the Big Class.jmp data table, a subject's age changed and you want to update an associated age report on JMP Live.

```
dt:Age[1] = 77;
/* Use the JMP Live Connection you established to update the data for the
   report with the specified ID based on the updated Big Class.jmp data table.
   */
updateresult = liveconnection << Update Data( report << Get ID( ), Data( dt,
   "Big Class.jmp" ));
// Show the outcome of the update.
show( updateresult << Succeeded );
```

Similarly, you can also specify the folder that the report resides in:

```
updateresult = liveconnection << Update Data( folder << Get ID( ), Data( dt,
   "Big Class.jmp" ));
```

Only reports in the folder using the specified data table update.

Work with Other Content in JMP Live Reports

Primarily in JSL, you work with JMP Live folders and reports. However, you can publish images to JMP Live reports. In JSL, you can use a generic object called a JMP Live Post to manage images or other content that does not fall under reports or folders. See the `JMP Live Post` function in the JMP Scripting Index at **Help** > **Scripting Index**.

HTTP Requests in JMP Live

To invoke HTTP methods in JMP Live, use `Get HTTP Request()`:

```
lc = New JMP Live(Connection("JMP Live Connection"), Prompt(If Needed));
hrequest = lc << Get HTTP Request();
hrequest << URL(url);
hrequest << Method("GET");
data = hrequest << Send();
json_data = Parse JSON(data);
show(json_data);
```

For more details, see "Communicate with REST Web Service".

Communicate with REST Web Service

Representational state transfer (REST) or RESTful web services is a way of providing interoperability between computer systems on the Internet. Use JSL to communicate with a RESTful application program interface (API) that uses HTTP requests to GET, POST, PUT, and DELETE data. For example, you can get data from a web service and return it as a data table. Some web services are Amazon Web Services (AWS), DropBox, and Text-Processing.com.

Notes:

- Strings that contain "&" are truncated. "+" is converted to a space.
- `New HTTP Request()` supports redirects, for example, from HTTP to HTTPS.

Examples of HTTP Requests

The following examples communicate with the National Aeronautics and Space Administration (NASA) API. You must request a key from https://api.nasa.gov/index.html#apply-for-an-api-key and assign it to the `key` variable to run the scripts successfully.

Write Data to the Log

```
url = "http://technology.nasa.gov/api/query/patent/all";
request = New HTTP Request( URL( url ), Method( "Get" ) );
data = request << Send();
Write( data );
   {"results":[["59fa04859600022b4d2e076f","LAR-TOPS-48","All-Organic
      Electroactive Device","NASA's Langley Research Center offers you an
      all-organic electroactive device system fabricated with single-wall
      carbon nanotube (SWCNT)....}
   // JSON output written to the log
```

Notes:

- The query string is at the end of the URL, following "?". For example, in https://www.google.com/search?q=jmp, "q=jmp" is the query string. `New HTTP Request()` automatically URI-encodes key/value pairs in a query string (and FORM).
- When you put an associative array in either a query string or a form, the key/value pairs get URI encoded. Spaces, ampersands, and other characters get encoded to ASCII, and the ampersand is escaped. The encoding is a "%" encoding.
- The web service defines the valid key/value pairs that it accepts. Refer to the API documentation for valid key/value pairs.
- For Kerberos authentication, specify the password as follows:

  `Password(":")`

Return Data as an Associative Array and Write it to the Log

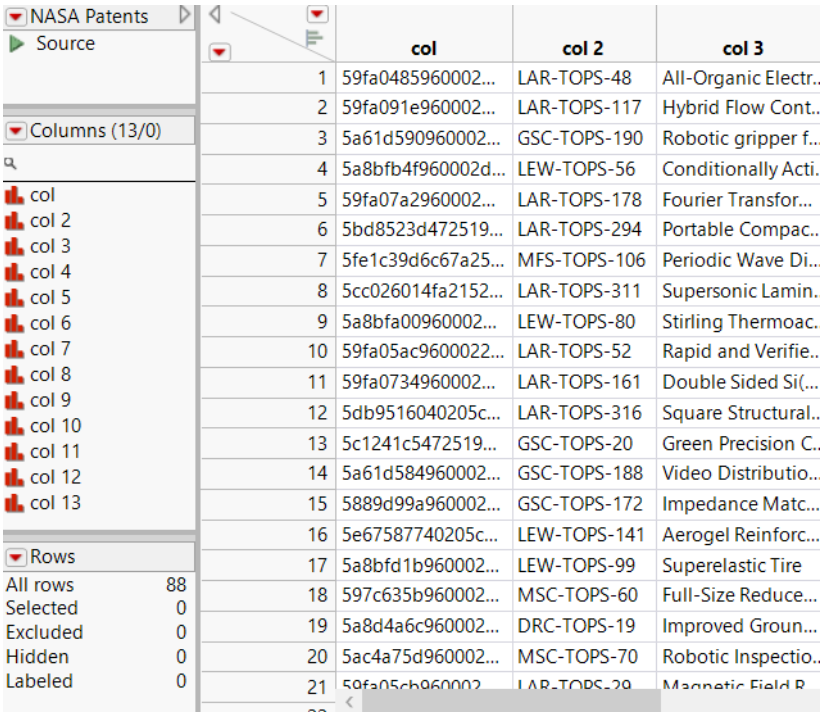
```
url = "http://technology.nasa.gov/api/query/patent/all";
request = New HTTP Request( URL( url ), Method( "Get" ) );
json = request << Send();
jsl_json = Parse JSON( json );
// turns JSON data into a JSL associative array
results = jsl_json["results"];
Write( results );
   {{"59fa04859600022b4d2e076f", "LAR-TOPS-48", "All-Organic Electroactive
      Device", "NASA's Langley Research Center offers you an all-organic
      electroactive device system fabricated with single-wall carbon nanotube
      (SWCNT)....}}
   // associative array
```

Return Data as a JMP Data Table

```
url = "http://technology.nasa.gov/api/query/patent/all";
request = New HTTP Request( URL( url ), Method( "Get" ) );
json = request << Send();
jsl_json = Parse JSON( json );
json = As JSON Expr( jsl_json["results"] );
dt = JSON To Data Table(
```

```
    json,
    JSON Settings(
       Stack( 0 ),
       Col(
          "/root",
          Column Name( "root" ),
          Fill( "Use Once" ),
          Type( "Pandas Values" ),
          Format( {"Best"} ),
          Modeling Type( "Continuous" )
       )
    )
  );
  dt << Set Name( "NASA Patents" );
```

Figure 14.5 Data Imported in a Data Table



OAuth 2.0 for Web APIs

OAuth 2.0 is an authorization standard for accessing personal user information through web APIs. An authorization standard is an agreed upon set of rules by many major web companies that are responsible for authorizing who can access information, and what scope of information they can access. This authorization is granted in the form of a limited time access *token* (a unique identifier specific to the individual and the scope of information they can access). Due to this two-pronged approach of limiting both users and scopes, OAuth is the current standard for securing personal information. Web sites such as Google, Salesforce, Reddit, and Dropbox use OAuth 2.0.

Since OAuth secures your personal information, it can appear to be fairly complex. OAuth is even more complicated due to the variety of ways it is implemented, which are called *grant types*. We'll walk through the different ways you can get an OAuth token in JSL. Please note that not all methods will work with all companies, so you'll have to do a little research to find out which grant type to use. Usually, the company will have documentation of this either with their APIs or possibly in a separate OAuth section.

We recommend that you familiarize yourself with the OAuth protocol at the official website (https://oauth.net/2/).

Note: Only OAuth 2.0 is supported in JSL. OAuth 1.0 and 1.0a are still used by a handful of companies (such as Twitter) but are largely regarded as inferior to 2.0.

Security

OAuth tokens can contain sensitive information. Depending on the grant type, you might need to provide your password or a client secret. JMP automatically handles this information for you.

Information that Google uses to authenticate your account is stored in C:/Users/<user name>/AppData/Roaming/SAS/JMP/OAuthTokens.jmp. (Your Google password is not stored.) After using an OAuth token on your personal account, you can delete all columns except for User and Client ID. Keep in mind that if you need to share your script, others won't have the User and Client ID columns saved on their machine. We recommend moving this information to an encrypted JSL script before sharing the OAuth script.

Using Postman for Redirect URLs

JMP does not provide a redirect URL for free. However, using Postman, it is easy to set one up yourself. Some examples that follow use Postman to redirect requests.

1. Go to https://www.getpostman.com.

2. Click the orange sign-in button in the top right corner.
3. Sign in with your Google account, or proceed to create your own account.
4. You can then use https://app.getpostman.com/oauth2/callback as a redirect URL.

Authorization Code Grant

The Authorization Code grant usually sends *refresh tokens*. Refresh tokens are unique in that they act as a sort of pseudo, permanent access token. Although you won't access sensitive information with them, a refresh token is all you need to get more access tokens. Treat it like you would your password.

The following code snippet creates an OAuth token using the Authorization Code grant. Notice this requires `Request Auth()`, `Client Secret()`, and `Redirect URL()`. Remember, after you make the token once, you can remove everything but the User and Client ID from OAuthTokens.jmp.

```
token = New OAuth2 Token(
  User( "Test User" ),
  Client ID( "12ab" ),
  Client Secret( "3456dEfG" ),
  Request Auth(
     Scope( "history" ),
     Auth URL( "https://example.com/services/oauth2/authorize" ),
  ),
  Redirect URL( "https://www.getpostman.com/oauth2/callback" ) ,
  Token URL( "https://example.com/services/oauth2/token" )
);
```

Note: See your API documentation for more information about how to get values such as the client secret and token URL.

After you run this code once, clean it up and make it secure.

```
token = New OAuth2 Token(
  User( "Test User" ),
  Client ID( “12ab” ),
);
```

Existing Refresh Tokens

You might already be familiar with OAuth and have saved a refresh token to your script. The following snippet creates an OAuth token with that existing token.

```
token = New OAuth 2 Token (
  User( "yourgoogleaccount@gmail.com" ),
```

```
  Refresh Token( "1a2b3c4e5F" ),
  Token URL( "https://www.example.com/oauth2/token" ),
  Client ID( "12ab" ),
  Client Secret( "3456dEfG" )
);
```

After you run this code once, clean it up and make it secure.

```
token = New OAuth2 Token(
  User( "yourgoogleaccount@gmail.com" ),
  Client ID( "12ab" )
);
```

Implicit Grant

The Implicit grant type doesn't require any sensitive information. If you're going to share a script with other people, it would be easiest to use this grant type. Note, however, that because sensitive information isn't required, services are less likely to implement this flow.

The following code snippet creates an OAuth token using the Implicit grant. This requires `Redirect URL()`.

```
token = New OAuth2 Token(
  User( "Implicit Flow"),
  Client ID( "123abc" ),
  Redirect URL( "https://www.getpostman.com/oauth2/callback") ,
  Token URL( "https://example.com/services/oauth2/token" )
);
```

Resource Owner Grant

The Resource Owner grant is the only type that doesn't require a redirect URL. If setting up Postman (the API development platform) is proving difficult, try this grant type.

The following snippet will create an OAuth token using the Resource Owner grant. This requires `Client Secret()` and `Password()`.

```
token = New OAuth2 Token(
  User( "yourgoogleaccount@gmail.com" ),
  Password( "wordP@ss0192" ),
  Client ID( "1234abcd" ),
  Client Secret( "5678EfGh" ),
  Token URL( "https://example.com/services/v1/token" )
);
```

After you run this code once, clean it up and make it secure.

```
token = New OAuth2 Token(
```

```
    User( "yourgoogleaccount@gmail.com" ),
    Client ID( "1234abcd" ),
);
```

Using a Token

In order to access a web API that requires OAuth, pass the token to `HTTP Request()` using the OAuth2 field.

```
request = HTTPRequest(
    URL( apiEndpoint ),
    Method( "POST" ),
    OAuth2( token ),
);
```

OAuth 2.0 Resources

The official website for OAuth 2.0 is https://oauth.net/2/.

https://auth0.com/docs/protocols/oauth2 provides more information about the OAuth 2 protocol.

The OAuth 2.0 standard, RFC6749, is at https://tools.ietf.org/html/rfc6749.

Google provides comprehensive documentation and examples of OAuth 2.0 at https://developers.google.com/identity/protocols/OAuth2.

Communicate with SAS Cloud Analytic Services

SAS Viya uses Cloud Analytic Services (CAS) code to perform calculations in the cloud. This feature provides a way for SAS Viya users and JMP users to collaborate. You can write CAS code in JMP to import data from and export data to the CAS server, convert CAS JSON text to a JMP data table, get a list of available CAS libraries, and more.

- See the SAS CAS documentation at https://developer.sas.com/apis/cas/rest/current/apidoc.html for more information.
- The SAS CAS documentation for CAS actions cover CASL, R, Python, Lua.

This section covers the following topics:

- "About CAS Actions"
- "Connect to the CAS Server"
- "Create and Submit a CAS Action"
- "Create and Submit a CAS DATA Step Action"
- "Create a New CAS Server"
- "Import Data from the CAS Server"
- "View a CAS Dataset"
- "Get a List of Available CAS Libraries"
- "Export Data from a Data Table to a CAS Server"
- "Convert a CAS Table to a Data Table"
- "Delete a CAS Table"
- "Remove a CAS Table"
- "Verify the CAS Server Connection"
- "Get Details about the Current Connection"
- "Disconnect from the CAS Server"
- "Terminate All CAS Sessions"

About CAS Actions

The smallest unit of work for the CAS server is a CAS action. CAS actions can load data, transform data, compute statistics, perform analytics, and create output.

You configure each action by specifying a set of input parameters. Running a CAS action on the CAS server processes the action's parameters and the data, and creates an action result. See *CAS Actions* at https://developer.sas.com/apis/cas/actions.html.

The JSON parameters are specified in the CAS documentation *CAS REST API*. See https://developer.sas.com/apis/cas/rest/current/apidoc.html. For JSON, you can look at the Lua format and make the appropriate adjustments.

For example, in Figure 14.6, notice the difference between the JSON syntax and the Lua syntax. JSON uses the `"name" : value` format. Lua uses the `name = value` syntax.

Figure 14.6 Differences Between JSON and LUA

JSON Syntax

```
{
    "attrTable": "string",
    "attrs": [ {
        column": "string",
        key": "string",
        value": "string" |
                number
}, {...} ],
...
}
```

LUA Syntax

```
{
    attrTable="string",
    attrs={ {
        column="string",
        key="string",
        value="string" |
              64-bit integer |
              integer |
              double |
             binary-large-object
      }, {...} },
,,,
}
```

Connect to the CAS Server

You can show the login window and prompt the user to provide the URL for the CAS server and their login information.

```
CAS Connect( Prompt( "Never"|"Always"|"IfNeeded" ) );
```

To add the login details to the script, follow this example:

```
CAS Connect( URL( URL ), Username( username ), Password( password ) );
```

Information about the CAS server appears in the log after the user logs in.

To specify a trusted certificate for https connections to the CAS server, use the optional Certificates argument.

```
CAS Connect( URL( URL ), Certificates( string ) );
```

To specify whether a certificate is verified, add either `Verify Certificates()` or `No Verify Certificates()` to the script:

```
Verify Certificates( 0|1 )
No Verify Certificates( 0|1 )
```

To specify a proxy server, add `Proxy Server()` to the script:

```
Proxy Server( "https://mycompany.com:80" )
Proxy User( username )
```

To bypass a proxy server, add `Bypass Proxy()` to the script:

```
Bypass Proxy( "https://localhost:80" )
```

Create and Submit a CAS Action

```
New CAS Action( Action( "builtins.echo" ), JSON( echo ) );
```

`Action` is the JSL variable name. builtins.echo is the name of the CAS action that SAS created. `echo` contains the JSL arguments.

Create and Submit a CAS DATA Step Action

```
New CAS DATA Step action( Code( code ) );
```

code defines the contents of the DATA step. Action is the JSL variable name.

```
cas = Current CAS Connection();
code =
"\[
   data temp;
   x = 9.1; y = 6; z = sqrt(x**2 + y**2);
   A = "SAS"; B = "Statistics";
   put _ALL_;              /* display all variables and values */
   run;
]\";
action = New CAS DATA Step action( Code( code ) );
cas << Submit( action );
```

Create a New CAS Server

`New CAS Server()` creates a new CAS server.

```
url = "URL";
cas = New CAS Server( Connect( URL( url ), Prompt(
   "Never"|"Always"|"IfNeeded" ) ) );
```

Import Data from the CAS Server

`CAS Import Data()` imports data from the CAS server:

```
CAS Import Data( libref.dataset, <Private|Invisible>, <"Use Labels for Var
   Names">);
```

- `invisible` hides the data table from view but shows it in the JMP Home Window. `private` hides the data table completely. Specify `private` if the user doesn't need to interact with the data table.
- `"Use Labels for Var Names"` indicates that the table name and the column names will be set to the SAS dataset label and the column labels. The appropriate column property, SAS Name SAS Label, is used.

View a CAS Dataset

CAS datasets are on the CAS file system. Datasets in memory are not listed.

To see the datasets on the CAS server that you want to import, use `CAS Get Data Sets()` after you connect to the server.

The following example shows how to list datasets that are only in casuser:

```
datasets = CAS Get Data Sets( casuser ); // quote the argument
Show( datasets );
```

The following example shows how to lists datasets that you have permissions to view:

```
datasets = CAS Get Data Sets();
Show( datasets );
```

Get a List of Available CAS Libraries

Libraries are namespaces, just like SAS libraries, librefs, or libnames. Use `CAS Get Libraries()` to get libraries from the CAS server.

```
libraries = CAS Get Libraries();
Show( libraries );
```

Export Data from a Data Table to a CAS Server

`CAS Export Data()` exports a data table to a CAS server to run an action on it.

```
CAS Export Data(
  Open( "$SAMPLE_DATA/Big Class.jmp" ),
  "casuser",
  "Big Class"
);
```

`CAS Export Data()` also has two arguments:

- `Save(0)` saves the data in memory.
- `Save(1)` saves the data to the file system.

Convert a CAS Table to a Data Table

`CAS Table to Data Table()` converts JSON data to a data table. A JSON element in the results of an action contains "`_ctb`". This is a marker indicating that it is formatted to represent a CAS table. You can send the JSON results to `CAS Table To Data Table()`. JMP will create a JMP data table from the JSON.

If there's more than one result, you might need to perform a little extra work as shown with `Get Results` at the end of the following example. This example shows how to export data, run an action, and save the data in several data tables.

```
cas = Current CAS Connection();
If( IsEmpty(cas),
   cas = CAS Connect( URL( url ), Username( username ), Prompt( "IfNeeded"
   ));
);
dt = Open( "$SAMPLE_DATA/Animals.jmp" );
cas << Export Data( dt, "CASUSER.Animals" );
json = "\[
{
  "table": {
   "name": "Animals",
   "caslib": "CASUSER"
  },
  "class": [
    "species",
    "subject",
    "season"
  ],
  "model": {
    "depVar": "miles",
    "effects": [
      {
        "vars": [
          "species",
          "season"
        ],
        "interaction": "BAR"
      }
    ],
    "printsol": true,
    "cl": false,
    "dfmethod": "RESIDUAL"
  },
  "random": [
    {
      "depVars": "miles",
      "effects": [
        {
          "vars": [
            "subject"
          ],
          "nest": [
            "species"
```

```
            ]
          }
        ]
      }
    ],
    "method": "REML"
}
]\";
action = New CAS Action(
   Action( "mixed.mixed" ),
   JSON( json )
);
rc = cas << Submit( action );
if(rc,
   results = action << Get Results;
   keys = results << Get Keys();
   For(i = 1, i <= N Items( keys ), i++,
      dt = CAS Table To Data Table( results[keys[i]] );
   );
);
cas << Disconnect();
```

`Get Results` enables you to examine the results of a CAS action; it's a JSL associative array. `Get JSON` gives you the raw JSON; it's a JSL string.

The JSON data is in the format discussed at https://developer.sas.com/apis/cas/rest/current/apidoc.html. Search for "ResultsTable".

In `CAS Table to Data Table()`, specify the JSON string, open the data table as invisible or private, and control the table name and column names.

```
json = "jsonstring";
CAS Table to Data Table( json, <Invisible(Boolean)|Private(Boolean)>, <"Use
   Labels For Var Names">);
```

- `Invisible` hides the data table from view but shows it in the JMP Home Window. `private` hides the data table completely. Specify `Private` if the user doesn't need to interact with the data table.
- `"Use Labels for Var Names"` indicates that the table name and the column names will be set to the SAS dataset label and the column labels. The appropriate column property, SAS Name SAS Label, is used.

Note: CAS ResultsTable attribute values (peers to "`_ctb`") found in the "`_ctb`" JSON are also set as table variables. The "CREATETIME" attribute creates a character table variable with a locale-formatted timestamp (for example, 04/03/2019 4:38:05 PM).

Delete a CAS Table

`CAS Delete Table()` deletes the underlying file from the CAS file system.

```
CAS Delete Table( table name, <remove> ); // quote the table name
```

The `CAS Delete Table()` function has a remove option, which removes an in-memory table from the CAS server. `CAS Remove Table()` serves the same purpose.

Remove a CAS Table

`Remove CAS Table()` removes an in-memory table from the CAS server.

```
CAS Remove Table( casuser, Big Class ) // quote both arguments
```

Verify the CAS Server Connection

`CAS Is Connected()` determines whether the CAS server is connected.

```
connected = CAS Is Connected();
Show( connected );
  connected = 1;
```

Get Details about the Current Connection

`Current CAS Connection()` obtains details about the current connection: the URL of the CAS server, your user name, and the session ID.

```
connection = Current CAS Connection();
Show( connection );
```

Disconnect from the CAS Server

`CAS Disconnect()` disconnects from the current CAS server and optionally terminates the session. By default, the session is terminated when disconnected.

```
CAS Disconnect(<"Terminate"|"NoTerminate">);
```

Terminate All CAS Sessions

`CAS Terminate Sessions()` ends all sessions owned by the current user except for the current session.

```
CAS Teminate Sessions();
```

You can also specify the session ID to terminate the specified session.

```
CAS Terminate Sessions(session ID);
```

Use `Get Session()` to get the session ID.

```
url = "http://myCASserver"; // quoted string
cas = CAS Connect(
   URL( url ),
   Username( "my_username" ), // quoted string
   Prompt( "IfNeeded" )
);
session_id = cas << Get Session;
```

To terminate the current session, use `CAS Disconnect()`.

Chapter 15

Creating Projects

Organizing Project Files in JMP

A JMP project provides a way to organize files that you use in an analysis. You can add JMP files (reports, data tables, scripts, journals, and so on) and non JMP files, such as Microsoft Word or Adobe PDF files. After you perform an analysis, the data table, report, and graph appear in the project on tabs. You can display multiple reports or graphs, display the project or script editor log, and run scripts from the project. Reports and graphs remain linked to the data table.

Projects help you avoid the clutter of several open JMP windows. You can maximize the project window to have a bigger view of the project. And when you save the project, the state of the project is saved (for example, the open reports and the layout of the window).

Contents

Create a Simple Project 867
Create a Project with a Bookmarks Pane and Log 868
Running Scripts in a Project 870
Create a Bookmark Group 872
Close Tabs in Projects 872
Get a Project 873
Reset the Window Layout 874
Run Startup Scripts to Control Workspaces 875

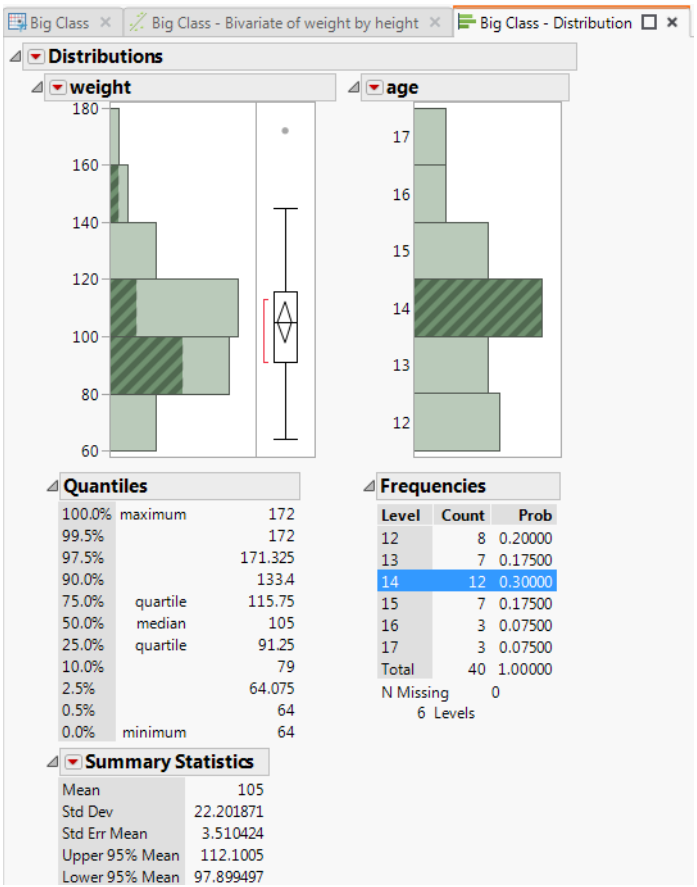
Create a Simple Project

A JMP project provides a way to organize files that you use in an analysis and display the output. A simple project might display the data table, graphs, and reports on separate tabs in the project window.

```
project = New Project();
project << Run Script(
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   dt << Run Script( "Bivariate" );
   dt << Run Script( "Distribution" );
);
```

Figure 15.1 shows the project with selected bars, showing that the graph is still linked to the data table.

Figure 15.1 Example of a Simple Project



Note: The best way to learn how to write a script for a project is to change the project extension to .zip, open the zip file, and then look for a file called project.jsl in the root directory.

Create a Project with a Bookmarks Pane and Log

A more complicated project might include a Bookmarks pane (to show file, folders, and groups that use frequently), and a project log. The data table and output are in separate tabs.

Begin by creating a new project and creating a bookmark.

```
project = New Project();
project << Add Bookmarks({
   File( "$SAMPLE_DATA/Big Class.jmp" )
});
```

`Run Script` opens the data table and creates output inside the project.

```
project << Run Script(
   Open(
      "$SAMPLE_DATA/Big Class.jmp",
      Set Window ID( "data" ) // unique window ID
   );
   Distribution(
      Set Window ID( "distrib" ),
      Continuous Distribution( Column( :weight ) ),
      Nominal Distribution( Column( :age ) )
   );
   Graph Builder(
      Set Window ID( "graphbuilder" ),
      Size( 443, 409 ),
      Show Control Panel( 0 ),
      Fit to Window( "Maintain Aspect Ratio" ),
      Variables( X( :age ), Y( :height ), Y( :weight ) ),
      Elements( Position( 1, 1 ), Line( X, Y, Legend( 1 ) ) ),
      Elements( Position( 1, 2 ), Bar( X, Y, Legend( 2 ) ) )
   );
);
```

`Set Layout` specifies the layout of the tabs that show the data table and output.

```
project << Set Layout(
   V Splitter Box(
      Size( 1215, 700 ),
      <<Set Sizes( {0.77, 0.23} ),
      <<Dockable(),
      V Splitter Box(
```

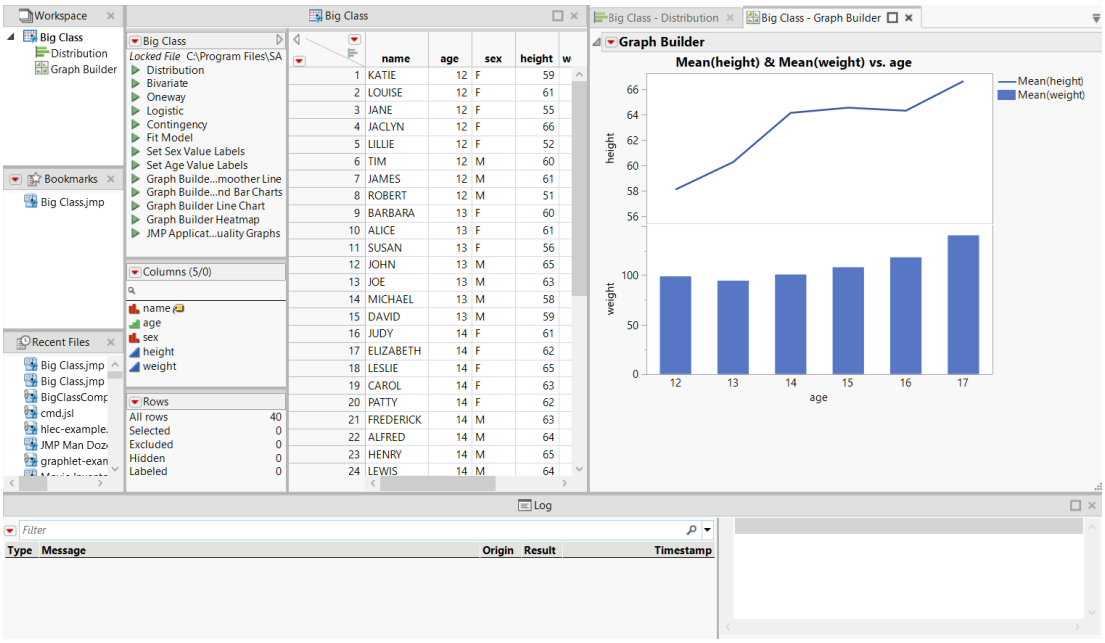
```
            Size( 1215, 540 ),
            <<Set Sizes( {1} ),
            <<Dockable(),
            H Splitter Box(
               Size( 1215, 540 ),
               <<Set Sizes( {0.11, 0.42, 0.47} ),
               <<Dockable(),
               V Splitter Box(
                  Size( 128, 540 ),
                  <<Set Sizes( {0.33, 0.33, 0.33} ),
                  <<Dockable(),
                  Tab Page Box(
                     Title( "Window List" ), // show the Window List
                     Window ID( "Window List" )
                  ),
                  Tab Page Box(
                     Title( "Bookmarks" ), // show the Bookmarks pane
                     Window ID( "Bookmarks" )
                  ),
                  Tab Page Box(
                     Title( "Recent Files" ), // show the Recent Files pane
                     Window ID( "Recent Files" )
                  )
               ),
               Tab Page Box( // show the data table
                  Title( "Big Class" ),
                  Window ID( "data" )
               ),
               Tab Box(
                  <<Dockable(),
                  Tab Page Box( // show the Distribution graph and reports
                     Title( "Big Class - Distribution" ),
                     Window ID( "distrib" )
                  ),
                  Tab Page Box( // show the Graph Builder graph
                     Title( "Big Class - Graph Builder" ),
                     Window ID( "graphbuilder" )
                  )
               )
            )
         ),
         Tab Page Box( Title( "Log" ), Window ID( "Log" ) )
      )
   );
```

Save saves the project.

```
// if the project has already been saved, the Save message
// with no argument saves the file to the existing location
project << Save( "$DOCUMENTS/Big Class.jmpprj" );
```

Note: If you're saving a project with unsaved documents in JSL, the project doesn't close and you get a log message that the project has unsaved documents. All files in a project must be saved or closed before the project can be saved. However, if you interactively save a project with unsaved documents, you are prompted to save the documents.

Figure 15.2 Example of a Complex Project



Open the project from another script:

```
project = Open( "Big Class.jmpprj" );
```

In this example, the script that you are running to open the project is in the same folder as the project.

Running Scripts in a Project

When you run a script outside of a project, the reports generated by the script appear in different windows. When you run a script inside a project, the reports appear in different tabs inside the project.

Running a JSL script outside of a project

Open a Script Editor window by selecting File > New > Script. Paste and run the following script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Run Script( "Bivariate" );
dt << Run Script( "Distribution" );
Wait( 2 );
Get Window( "Big Class - Distribution" ) << Close Window();
```

Note that this script opens three JMP windows: one for the data table and one for each report. Close the reports and then the data table.

Running a JSL script inside of a project

Create a new project by selecting File > New > Project. From the project window, create a new script editor window by selecting File > New > Script. Note that the script editor window opens as a tab in the project. Paste the same JSL script you previously ran into the script editor and run it:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Run Script( "Bivariate" );
dt << Run Script( "Distribution" );
Wait( 2 );
Get Window( "Big Class - Distribution" ) << Close Window();
```

Note that this time, the data table and the reports all opened as tabs in the project, and that `Get Window()` correctly found the distribution tab. In general, all JSL functions that open, find, manipulate, and close JMP windows, tables, and reports work from within a project the same as they would outside projects.

Running a JSL script using a project's Run Script message

You can use the `Run Script` message to run JSL in the project's context. Open a Script Editor window, then paste and run the following code:

```
project = New Project();
project << Run Script(
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   dt << Run Script( "Bivariate" );
   dt << Run Script( "Distribution" );
   Wait( 2 );
   Get Window( "Big Class - Distribution" ) << Close Window();
);
```

Note that the script inside the `Run Script` message runs the same as if it had been in a script editor window inside the project.

Scoping Variables

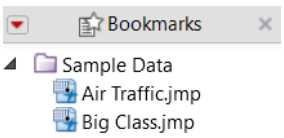
JMP can have several projects open at the same time, as well as items not open in any project. Each project has its own global scope, which prevents conflicts caused by identically named variables in different projects. In the rare case that you do want to share specific variables among all opened projects, use the `:::` root operator.

Create a Bookmark Group

You can create bookmark groups to organize the bookmarks in a project. A bookmark group can contain bookmarked files, folder, and additional bookmark groups. The bookmark group exists only in the project, not on your computer. Use `Group()` to create the bookmark group.

```
project = New Project();
project << Add Bookmarks(
        Group(
           "Sample Data",
           Expanded( 1 ), // open the group
           {File( "$SAMPLE_DATA/Air Traffic.jmp" ),
           File( "$SAMPLE_DATA/Big Class.jmp" )}
        )
);
project << Set Layout(
   H Splitter Box(
      <<Set Sizes( {0.15, 0.85} ),
      Tab Page Box( Window ID( "Bookmarks" ), )
   )
);
project << Save( "$DOCUMENTS/My Project.jmpprj" );
```

Figure 15.3 "Sample Data" Group in a Project



Close Tabs in Projects

Close a tab in a project by identifying the window with `Get Window()` and then closing it with `Close Window`.

```
project = New Project();
```

```
project << Run Script(
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   dt << Run Script( "Bivariate" );
   dt << Run Script( "Distribution" );
);

Wait( 2 ); // for demonstration purposes

// returns a reference to the open "Big Class - Distribution" window
dist = Get Window( Project( project ), "Big Class - Distribution" ) ;
dist << Close Window();
```

To close all tabs in a project, use `Get Window List << Close Window()`.

```
project = New Project();
project << Run Script(
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   dt << Run Script( "Bivariate" );
   dt << Run Script( "Distribution" );
);

Wait( 2 ); // for demonstration purposes

Get Window List( Project( project ) ) << Close Window();
```

Get a Project

When running a JSL script from within a project, you can use the `This Project()` function to obtain that project. The following example displays the name of the current project in a new window.

```
project = This Project();
If( Is Empty( project ),

   /* if the script is not run from a project, display this
   sentence in the new window */
   New Window( "My Project", Text Box( "Script isn't running in a project." )
   ),

   /* if the script is run from a project, display the window title
   in the new window */
   New Window( "My Project", Text Box( project << Get Window Title() ) )
);
```

`Get Project()` takes the same arguments as the `Window()` function. Omit the arguments to get a list of the currently open projects, include a string to find a specific project by title, or include a number to get the *i*th open project.

```
Open( "$SAMPLE_PROJECTS/Big Class.jmpprj" );
Open( "$SAMPLE_PROJECTS/Sports.jmpprj" );

// print the title of Sports.jmpprj
Print( Get Project( 2 ) << Get Window Title() );
```

`Get Project()` is also helpful when you want to save a project or set the layout from within the project.

```
Get Project() << Save(...);
Get Project() << Set Layout(...);
```

Reset the Window Layout

Suppose that you create a project and customize the layout of the windows. For example, you might hide the Window List. Use the `Reset Layout` message to return the windows to their original layout.

```
project = New Project();
project << Run Script( // open data table and create a graph
       Open(
              "$SAMPLE_DATA/Big Class.jmp",
             Set Window ID( "data" )
       );
       New Window( "Big Class - Bivariate of weight by height",
              Set Window ID( "bivariate" ),
              Bivariate(Y( :weight ), X( :height ),
              )
       ) << Set Window Icon( "Bivariate" );
);

Wait(3); // for demonstration purposes
project << Set Layout( // set a custom layout
       H Splitter Box(
              <<Set Sizes( {0.2, 0.8} ),
              Tab Page Box( Window ID( "Bookmarks" ) ),
                V Splitter Box(
                     <<Set Sizes( {0.53, 0.47} ),
                     Tab Page Box(
                         Window ID( "bivariate" ) ),
                     Tab Page Box(
                         Window ID( "data" ) )
```

```
                )
            )
);

Wait( 3 ); // for demonstration purposes
project << Reset Layout(); // return to the original layout
```

Run Startup Scripts to Control Workspaces

A workspace is either a project window or the windows that are not in a project (for example, the home window). A default workspace loads when JMP starts, and then each time you create or open a project, JMP creates another workspace.

Consider writing a workspace startup script to customize variables for a specific project. You might also construct a new customized window in a workspace startup script to put a copy of that window in each project. This is an alternative to having a single copy of the script that runs outside of projects.

The workspace startup script runs each time a new workspace loads. This means that the workspace startup script runs once when JMP starts (when the "default workspace" loads) and then runs again each time a new project is loaded. The workspace startups script runs in the context of the project, so any windows it opens or variables it sets are for that project only.

Name the script workspaceStart.jsl and place it in one of the following folders, as appropriate for your operating system. When JMP starts, JMP looks for the workspaceStart.jsl script in these folders in the order in which they are listed here. The first one that is found is run, and the search immediately stops.

Note: Some path names in this section refer to the "JMP" folder. On Windows, in JMP Pro, the "JMP" folder is named "JMPPro".

On Windows:

1. C:/Users/<username>/AppData/Roaming/SAS/JMP/16
2. C:/Users/<username>/AppData/Roaming/SAS/JMP

On macOS:

1. /Users/<username>/Library/Application Support/JMP/16
2. /Users/<username>/Library/Application Support/JMP

The workspaceStart.jsl script runs only for a particular user on a computer. You can add a script named workspaceStartAdmin.jsl in one of the following places, as appropriate for your operating system. This script is run for every user on a computer. JMP runs workspaceStartAdmin.jsl first if found. Then JMP runs workspaceStart.jsl if found.

On Windows:

1. C:/ProgramData/SAS/JMP/16
2. C:/ProgramData/SAS/JMP

On macOS:

1. /Library/Application Support/JMP/16
2. /Library/Application Support/JMP

Chapter 16

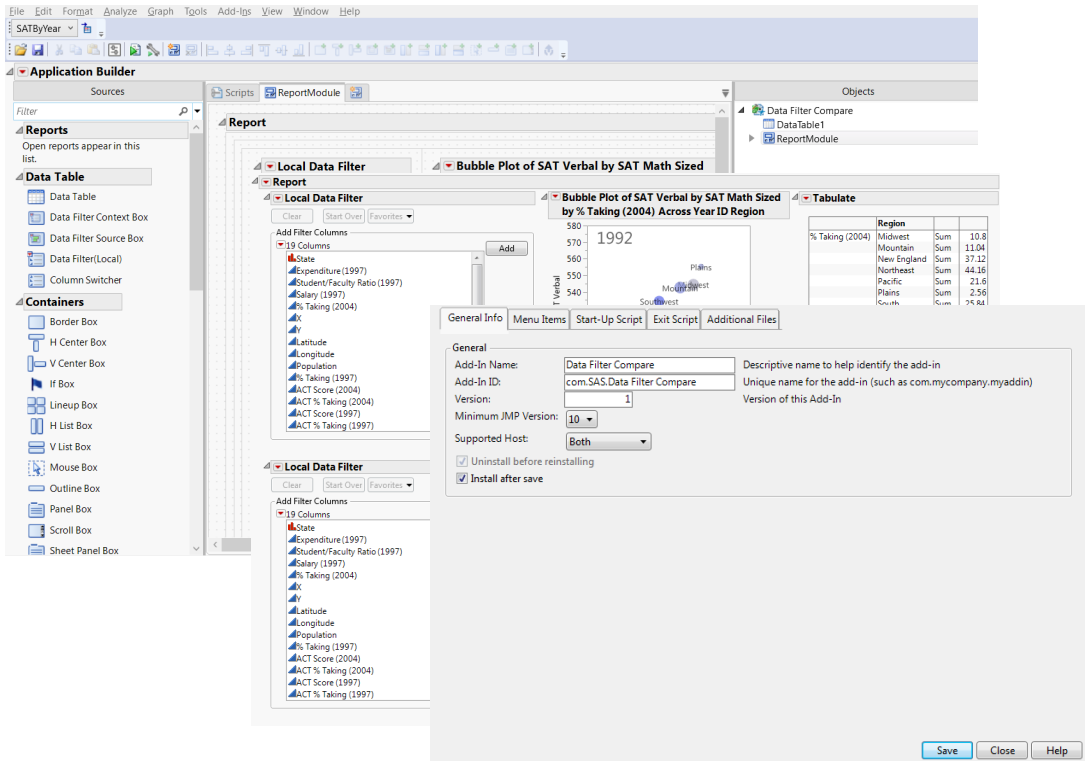
Creating Applications

Application Builder

In addition to providing platforms for quickly analyzing data, JMP lets you create applications to support your specific needs. You often perform the same tasks every day (such as running Distribution and Fit Model analyses on a data table and viewing the results). In JMP's Application Builder, you can create an application that shows the results for both analyses in one window. Application Builder's drag-and-drop interface reduces the amount of scripting required to create an application.

Add-In Builder visually guides you through creating an add-in that users can install in JMP. And after creating an application in Application Builder, use Add-In Builder to save the application as an add-in. This allows users to install the application and quickly run it inside JMP.

Figure 16.1 Creating Customized Reports in Application Builder



Contents

Create Applications with Application Builder 879
Example of Creating an Application 879
Application Builder Terminology 881
Design an Application 882
Application Builder Window 883
Application Builder Red Triangle Options 884
Create an Application 886
Edit or Run an Application 897
Options for Saving Applications 898
Additional Examples of Creating Applications 899
Compile Add-Ins with the JMP Add-In Builder 915
Compile Scripts into an Add-In Using Add-In Builder 915
Edit an Add-In 920
Share an Add-In 920
Installing Multiple Add-Ins 920
Register an Add-In Using JSL 921
Create an Add-In Manually 921
Manage JMP Add-Ins 923

Create Applications with Application Builder

Application Builder's drag-and-drop interface lets you visually design windows with buttons, lists, graphs, and other objects. This saves you the step of writing scripts to create these objects.

Application Builder also helps boost the productivity of experienced JMP Scripting Language (JSL) programmers. You can start designing a program in Application Builder and edit the automatically generated scripts. Integrate your own scripts to create even more powerful custom applications.

You can use Add-In Builder with Application Builder. First, use Application Builder to write complex scripts that create new platforms and interact with users and JMP. Then, use Add-In Builder to easily distribute your complex scripts so that any JMP user can access them without having to open and run the scripts. See "Compile Add-Ins with the JMP Add-In Builder"

Notes:

- Creating simple applications in Application Builder requires little to no JSL programming. However, to create most simple applications, Dashboard Builder might be sufficient. Dashboard Builder provides a subset of features that are found in Application Builder. See *Using JMP*.
- In each version of JMP, Application Builder takes advantages of new features that are not supported in earlier releases. For maximum compatibility, we recommend that you create applications in the earliest version of JMP that the application will be run in.

Example of Creating an Application

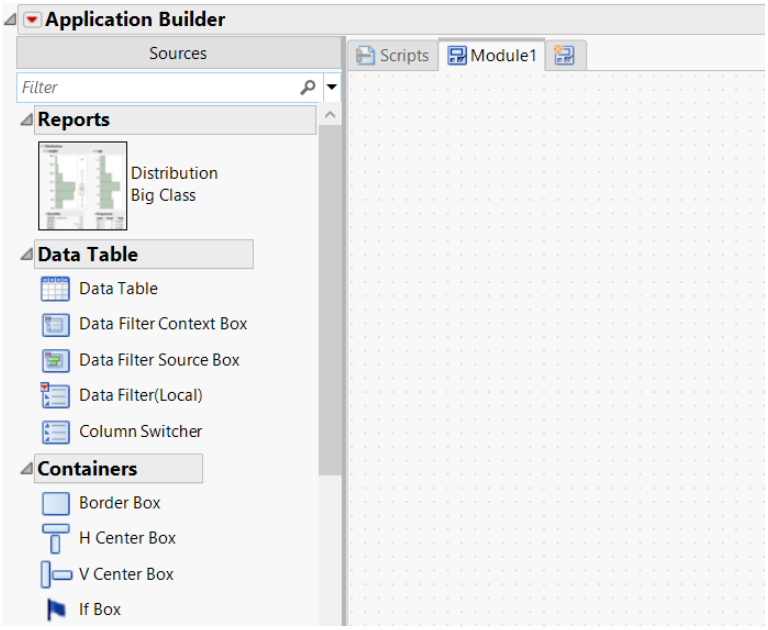
Application Builder helps you create instant applications without writing JSL. One example is a JMP window in which you arrange reports, graphs, and other objects to make them easy to interact with at once.

This example shows how to create a simple instant application for a Distribution report. This application lets you analyze preselected variables rather than having to select them in a launch window.

1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Run the **Distribution** table script to generate the report.
3. Select **File > New > Application** (or **File > New > New Application** on macOS).
4. Click the Blank Application box.

   A new blank application appears. Notice the Distribution report in the left column.

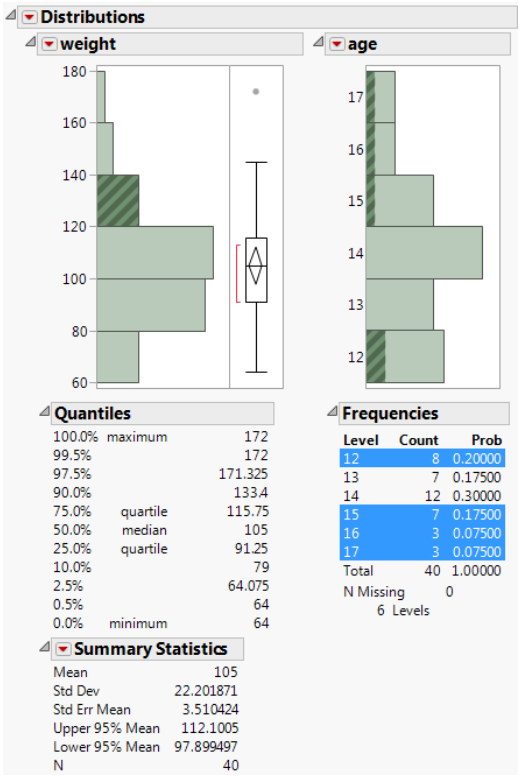
Figure 16.2 New Application with Distribution Report in the Sources Panel



5. In the Sources panel, drag the Distribution report onto the application workspace (the blank area on the Module1 tab).
6. Click the Application Builder red triangle and select **Run Application**.

   The Distribution report appears in a new window. Note that the histograms are interactive and still associated with the data table. When the data changes, you run the application again, and the Distribution report is updated.

Figure 16.3 Example of an Instant Application



	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107
11	SUSAN	13	F	56	67
12	JOHN	13	M	65	98
13	JOE	13	M	63	105
14	MICHAEL	13	M	58	95
15	DAVID	13	M	59	79
16	JUDY	14	F	61	81

Application Builder Terminology

An *application* consists of modules. A *module* contains the objects and scripts that are compiled into an application. When you run an application, JMP can show a separate window for each module. For example, one window might show a **Create Graphs** button. After the user clicks the button, the graphs appear in a new window.

There are terms you should become familiar with before using Application Builder. Knowing these terms helps you understand how a module works, even if you need to minimally edit an automatically generated script.

Application The top-level file that consists of one or more modules.

Module A collection of objects, messages, instances, and other JSL statements that are compiled into an application to create new JMP features.

Instant Application An application that typically consists of reports and does not require JSL scripting. You can also create an instant application, or “dashboard”, in Dashboard

Builder. Similar to Application Builder, Dashboard Builder has a drag-and-drop interface but with pre-configured display boxes.

Custom Application An application that demonstrates custom behavior through JSL scripting.

Objects and Messages An *object* is a dynamic entity in JMP, such as a data table, a data column, a platform results window, a graph, and so on. Most objects can receive messages that instruct the object to perform some action on itself.

A *message* is a JSL statement that is directed to an object. That object knows how to evaluate the message.

Module Instance The occurrence of a module in the application. In a complex application, you might create a start-up script that creates multiple instances of one or more modules.

Namespace A collection of unique variable names and corresponding values that prevents collisions among scripts.

JMP creates the `Application` namespace and the `ModuleInstance` namespace automatically. Symbols in the `Application` namespace are visible only to scripts in the application; the namespace is not available to scripts not created in the application. For more information about namespaces, see "Advanced Scoping and Namespaces".

Variable A unique name to which you assign a value. JMP creates the `thisApplication` and `thisModuleInstance` variables automatically.

- The `thisApplication` variable (used in the `Application` namespace) contains module names, which are available in any script in the application.
- The `thisModuleInstance` variable (used in the `ModuleInstance` namespace) contains box and script names, which are valid only in their own modules.

Container A display box object (such as a tab or outline) into which you drag other objects.

Design an Application

Before creating the application, you must identify its purpose, decide on which graphical elements to include, and figure out which JSL scripts you need to write.

Purpose What do you want the application to accomplish? Which features of JMP do you need to customize (for example, a launch window, report, or graph)?

Graphical Elements How many modules do you need? Look at the list of graphical objects in Application Builder and decide which objects to include in your application. The Sources panel shows dozens of boxes and icons that you can drag and drop into a module. Also look at the preinstalled samples to see whether a similar module exists.

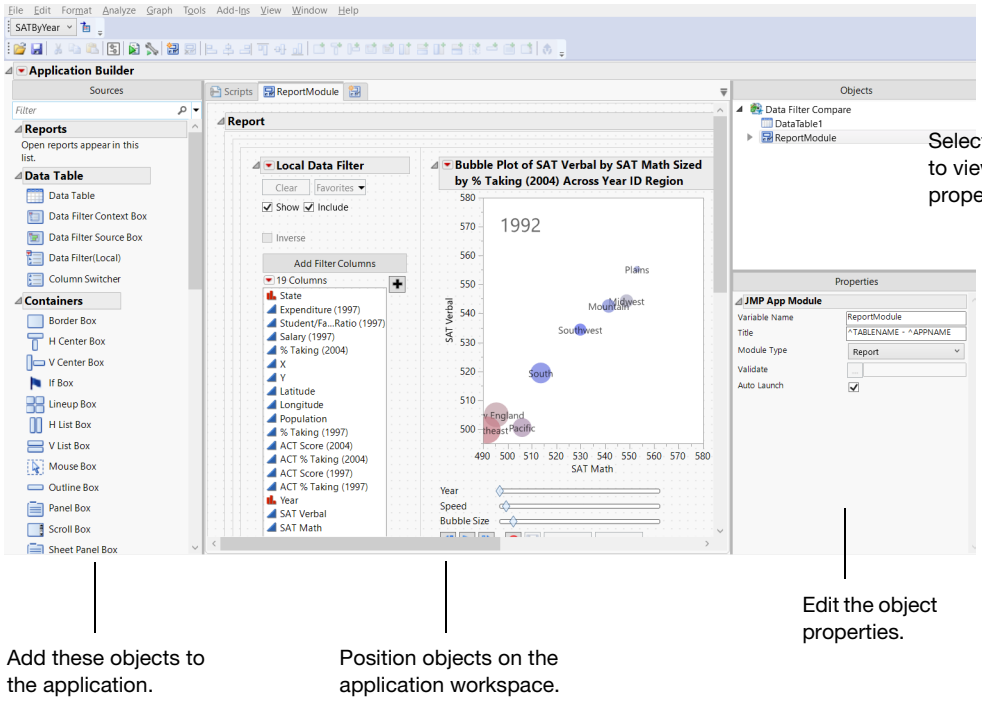
Scripting Depending on the complexity of the application, you write JSL scripts. See "Write Scripts" for more information about writing scripts for an application.

A non-interactive application that only shows objects does not require scripts. One example is a window showing reports that you have already generated in JMP.

Application Builder Window

Figure 16.4 shows an application under development in Application Builder.

Figure 16.4 The Application Builder Window



The Application Builder window provides the following features:

- The toolbar provides quick access to many features, such as aligning objects and inserting common display boxes. (Select **View > Toolbars > Application Builder** to show the toolbar.)
- The Sources panel shows objects that you can include in the application, including open reports and graphs. Right-click an object and select **Scripting Help** for information about the object and its properties.
- The workspace in the middle shows dotted grid lines, which help you line up objects. Drag objects onto one or more Module tabs in the workspace and write scripts on the Scripts tab to control the objects' functionality.

- The Objects panel shows a hierarchical view of the application. Along with the application, each module and its nested objects are listed. You can click an object to show its properties or to select it on the workspace.
- In the Properties panel, you set properties for each object, such as the location, width, or name of an object. The properties vary depending on the type of object.

Tip: To hide the grid and turn off snap-to-grid for all applications that you create, select **File > Preferences > Platforms > JMP App** and deselect these options. You can also deselect these features in the Application Builder red triangle menu.

Application Builder Red Triangle Options

The Application Builder red triangle menu provides options for running and debugging the application, opening sample applications, showing the grid, and more.

Run Application Starts the application. A window for each module opens, and you can interact with the application as the user will.

Debug Application Opens the script in the JSL Debugger to troubleshoot errors. See "Debug or Profile Scripts".

Snap to Grid Aligns (or snaps) objects to the nearest dotted grid lines as you drag them on the workspace. Selected by default.

Show Grid Displays dotted grid lines on the workspace. Selected by default.

Show Properties Shows or hides the Objects and Properties panels.

Auto Scroll Automatically scrolls horizontally or vertically as you drag an object near the edges of the workspace. Selected by default.

Dashboard Mode Displays the application in the simpler Dashboard Builder window. When you drag a report, data filter, or data table to the application, it is automatically placed in splitter and tab page boxes. You cannot add scripts in Dashboard Mode.

Be aware that switching back and forth does not change anything that you have already added to the application. Switching modes changes the way the interface works when you add new content.

Source Panel Shows or hides the Sources panel (the left column in which application components are displayed) and categorizes the icons. Both settings are selected by default.

Save Script Lets you save the application to a data table, journal, script window, template, or add-in. See "Options for Saving Applications".

Table 16.1 Sample Applications Installed with JMP

Six Quality Graphs.jmpappsource	Creates three Control Charts, a Distribution report, and a Capability Analysis report.
Data Filter Compare.jmpappsource	Creates two Bubble Plots. Each plot has its own data filter and Tabulate report. See this application for an example of the `Data Filter Context Box` function.
Data Table Application.jmpappsource	Lets the user select data table columns to stratify and the sampling rate. Display the data table to see the results.
Graph Launcher.jmpappsource	Lets the user enter an equation and axes settings and then create a graph. The initial window remains open, so the user can modify the equation and create new graphs.
Instant App.jmpappsource	Combines two Multivariate reports (Principal Components/Factor Analysis and T Square with All Principal Components).
Instant App Customized.jmpappsource	A modified version of Instant App.jmpappsource. Options for selecting the principal components report, changing the marker size, and showing means are included.
Launcher With Report.jmpappsource	Lets the user select data table columns in a launch window and then creates a graph.
Parameterized Instant App.jmpappsource	Lets the user select data table columns and then creates two Multivariate reports. An argument is assigned to the Y role. This means that the reports can be created from any open data table, not just the table specified in the application.
Parameterized MSA Combo Chart.jmpappsource	Creates a collection of reports for Measurement System Analysis (MSA).
Presentation.jmpappsource	Creates an onscreen presentation, similar to a slideshow, with navigation buttons and an embedded script.
R Application.jmpappsource	Lets the user select columns and then shows multivariate data in the shape of a face (the Chernoff faces). Requires the R TeachingDemos package.

Table 16.1 Sample Applications Installed with JMP *(Continued)*

SAS Application.jmpappsource	Runs a SAS script and then adds the output to a report. Prompts you to log on to a SAS server if you are not already connected.

Create an Application

After writing the application specifications, you create a blank application in JMP and begin adding objects and scripts.

This section provides the basic steps for creating an application:

- "Create a New Application"
- "Arrange and Remove Objects"
- "Customize Object Properties"
- "Write Scripts"

Note: In each version of JMP, Application Builder takes advantages of new features that are not supported in earlier releases. For maximum compatibility, we recommend that you create applications in the earliest version of JMP that the application will be run in.

Create a New Application

To create a new blank application, follow these steps:

1. Select **File > New > Application** (or **File > New > New Application** on macOS).
2. Select the Blank Application.

   The Application Builder window appears.
3. Select **File > Save** and save the file as a JMP Source File. The extension is .jmpappsource.

After you create the blank application, you add open reports, display boxes, and other objects to the workspace. See "Example of Creating an Application".

Manage Modules

When you create a new application, a Module tab appears by default. You can add objects to this module to create interface items. To create additional windows for your application, you can add and customize new modules.

Add Objects to a Module

Follow these steps to add objects to a module:

1. In the Sources panel, select the type of object that you want to add.
2. Drag the object onto the **Module1** tab (or double-click the object).
3. Select the object and update its properties in the Properties panel. See "Customize Object Properties".
4. Add scripts to scriptable objects. See "Write Scripts".
5. (Optional) To prevent a module from launching when you run the application, select the module in the Objects panel and deselect **Auto Launch**. (You might do this while testing one out of multiple modules.)
6. Click the Application Builder red triangle and select **Run Application** to test your application.

   Depending on how the modules are used, it may not be possible to test one module without creating the other modules.

Create a Module

1. Select **Format > Add Module.**
2. In the Properties panel, select the **Module Type**:
   - Dialog
   - Dialog with Menu
   - Modal Dialog
   - Launcher
   - Report
   - Display Box

Tip: When you run an application, the window title is the name of the data table, if applicable, followed by a hyphen and the application name. Change the application name by modifying text in the JMP App object's **Name** property.

Rename a Module

Change the **Variable Name** in the module's object properties.

Delete a Module

Select **Format > Delete Module.**

Delete a Data Table

Select the Data Table object, right-click and select **Delete.**

Note: You cannot delete the data table if it has any associated objects. Delete any objects and then delete the data table.

Modal Dialog Modules

Modal Dialog modules have some special behaviors to be aware of when using them.

- `ret = Module1 << Create Instance()` will not return until the dialog is complete. By this time, the dialog has been destroyed. Therefore, the return value will not be a handle to the Module Instance. Like `New Window()`, the return will instead be of the form `{Button(1 | -1), User Data}`. (1 | -1) indicates whether the OK (1) or Cancel (-1) button was pressed.
- `User Data()` is a new property on the Module Instance. During execution of the dialog, the Module script can set the user data:

  ```
  thisModuleInstance << Set User Data(...);
  ```

  to store something to be parsed by the caller. The user data will be evaluated at the time that it is set. There is also a corresponding `Get User Data()`.
- Like `New Window()`, if there is no OK or Cancel button included, an OK button will be added.
- There is an optional `Validate` script property on the Module Instance. This script is only used for Modal dialogs and behaves like `On Validate()` in `New Window()`. It is called when the OK button is pressed, and returns 1 to accept the input or 0 to disallow the close.

  Tip: If you use the Validate property, you can call `thisModuleInstance << Set User Data()`.

- Unlike other module types, a Modal Dialog will not show until the module script has completed. For other module types, the window is created during the call to `thisModuleInstance << Create Objects`. JMP cannot display a Modal Dialog here because control would stop and there would be no way to initialize the contents of the boxes in the display. Since the window is not created yet, you will not be able to do actions like set the window title or set an on-close script.

Arrange and Remove Objects

Drag objects from the Sources panel onto the workspace or double-click them. A blue border appears around the edge of a selected object. Then you can position the selected object in several ways. You can also change the container of an object or insert an object into a new container.

Tip: To quickly select an object, select it in the Objects panel. This method is particularly helpful when the internal object completely covers the container, as with a border that has an internal text box.

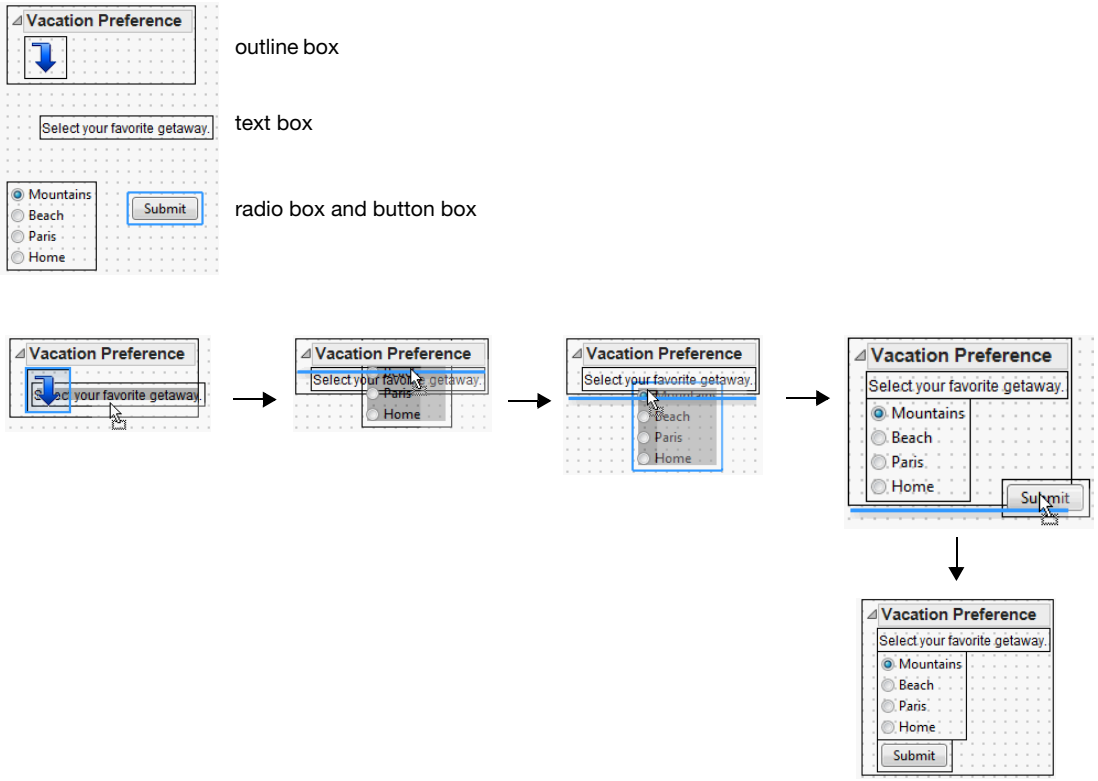
Drag an Object

One way to position objects is by dragging them around the workspace.

- The top and left edges of an object align (or snap) to the nearest dotted grid lines as you drag them on the workspace. For a more precise placement, click the Application Builder red triangle and deselect **Snap to Grid**. You can also turn off **Snap to Grid** for all applications in the JMP Preferences. Select **File (Windows) or JMP (macOS)> Preferences > Platforms > JMP App** to change the option.
- Position objects in containers instead of positioning them independently on the workspace. For example, place two button boxes in a V List Box to align them vertically.
- When you drag an object, press Shift to restrict movement to a single axis (for example, to move the object horizontally without moving the vertical position).
- When you drag an object into a container, the arrow indicates where you can drop the object.
- When you drag an object over another one, a blue line shows where you can drop the object.

Figure 16.5 shows the various methods for dragging objects inside a container.

Figure 16.5 Examples of Dragging Objects into a Container



Change the X and Y Positions on a Container Object

- To position a container object precisely, select the object and then change the **X Position** and **Y Position** properties. After you enter a new X position, press Tab to see the object move to the new position and then enter the new Y position.
- To place an object in the upper left corner, right-click the object and select **Move to Corner**. This is a shortcut to setting the X and Y positions to 0. On macOS, press Ctrl+Command and then select **Move to Corner**.

Line Up Multiple Objects

To line up objects horizontally or vertically, select the objects and then select an option from the **Format > Align Boxes** menu.

Tips:

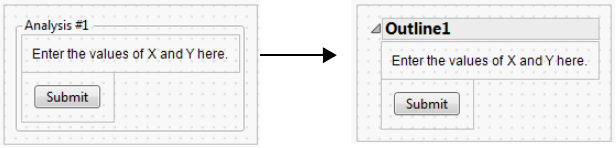
- The **Align Boxes** options are unavailable when you select a container rather than the internal objects.

- If you right-click a container, you might inadvertently select one of the internal objects instead. Right-click in the workspace instead.

Change the Type of Container

After inserting some containers, you can change them to a different type of container without having to re-create the object. Suppose that you create a panel that contains text and a button. To see what the panel looks like as an outline, select the panel, right-click in the module, and then select **Change Container > Outline Box**.

Figure 16.6 Change a Panel to an Outline



In this example, you would also change the outline title to match that of the panel.

Tip: To change the orientation of a list box quickly, select or deselect **Horizontal** in the object properties.

Insert Objects into a New Container

The toolbar at the top of the Application Builder window provides buttons for containers such as border boxes and mouse boxes (Figure 16.7).

To insert selected objects into a container, click the appropriate button on the toolbar or select **Format > Add Container**.

Figure 16.7 Container Toolbar

Duplicate Objects

When you copy and paste an object, the second instance of the object is renamed. The scripts attached to the object are also renamed. Pressing Ctrl (or Command on macOS) and dragging the object also creates duplicates of the originals.

Delete an Object

- Remove one or more objects by selecting them and then pressing Delete.
- Drag the objects outside of the module.
- On macOS, select the objects, press Ctrl+Command, and then select **Delete**.

- On Windows, select **Edit > Clear** to delete all objects in a module.
- Select the object and then press Backspace.

If you no longer need the object's script, you can delete the script also.

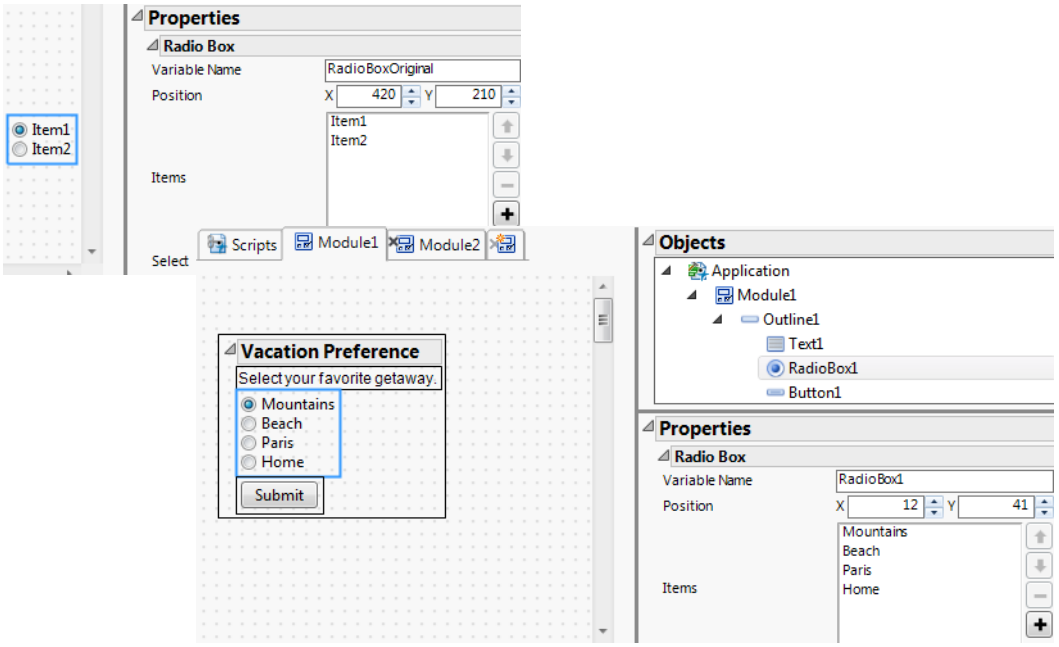
Customize Object Properties

After you drag an object from the Sources panel onto a module, customize the object properties in the Properties panel. Editing the object properties saves you the step of writing JSL for list items, button names, and so on.

Variable names are case- and space-insensitive. When a script contains an object named "Button1", Application Builder warns if you try to rename another object "Button 1".

To create the radio box list items in Figure 16.8, double-click the placeholder items "Item1" and "Item2" and enter the new list item.

Figure 16.8 Radio Box Object Properties



For more information about an object, right-click the object and select **Scripting Help**. The JMP Scripting Index appears, and the object that you selected is highlighted. The Scripting Index often includes a script that you can run to see an example of the object.

To read more about values in the Properties panel, hover over the value.

Application Properties

The Application properties identify the run password, data table name, and more. Select the application in the Objects list.

Name Appears after the data table name in the application's title bar.

Description Enter a description of the application.

Auto Launch Shows a launch window in which the user selects the arguments that have been defined in the application. This launch window is not one of the defined modules; the modules will be instanced based on their own Auto Launch property.

Encrypt Prevents users from editing the application in a text editor. Only applications that the user runs are encrypted (the .jmpapp file and the application in a JMP add-in). Scripts that you save to a data table, journal, and add-in are encrypted. For more information about encryption, see "Encrypt and Decrypt Scripts".

Run Password Enter the password required to run the application. To test the password, run the application outside Application Builder.

Table Module Properties

Inserting the Data Table object creates a data table object. To define properties such as the data table path, select the data table module in the Objects panel. Then modify the following options in the Properties panel.

Variable Name Specifies the name of the data table object. This name appears in the Properties and Objects panels and in the application's JSL script.

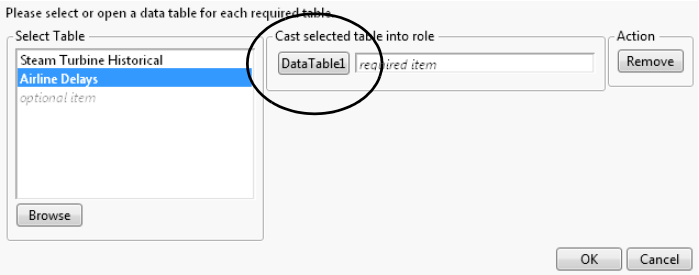
Path Specifies the absolute or relative filepath for the data table used in the application. You can precede the data table name with a path variable (such as $HOME or $USER_APPDATA).

Application Builder opens the specified data table when you edit the application. If the Path property is empty or the data table cannot be found, you are prompted to open the table.

When you close a data table and objects in the application depend on the data table, the objects are removed from the application and a warning message appears. To restore the objects, reopen the application.

Label Specifies the string used when prompting the user to open a data table. Figure 16.9 shows the default value.

Figure 16.9 Label in Data Table Prompt Window



Location Determines how the data table used in the application is selected when the user runs the application.

- Current Data Table: Uses the current data table. If no data tables are open, the user is prompted to open one. This option gives users the flexibility to use their own data.
- Full Path: Uses the data table specified in the Path property.
- Name: Uses the first open data table with the specified name. Otherwise, JMP uses the data table specified in the Path property. Note that if the application is moved to another computer, the link to the data table is broken. Consider storing the data table in a shared location.
- Prompt: Asks the user to select an open data table or browse to select a data table.
- Script: Uses the data table defined in the application or module script.
- Embedded Script: Reads the Source script in the data table that was used to generate the report. The Source script is run when you open the application.

Invisible Hides the data table from view but lists it in the JMP home window. This option is available for the Full Path and Name locations.

Write Scripts

After adding an object to the module, you write a script for the object to provide functionality. For example, you might want the user to click a button, select a directory, and then select a data table. Or you might have an application that displays a different graph based on which radio button the user selects.

Application and Module Namespaces

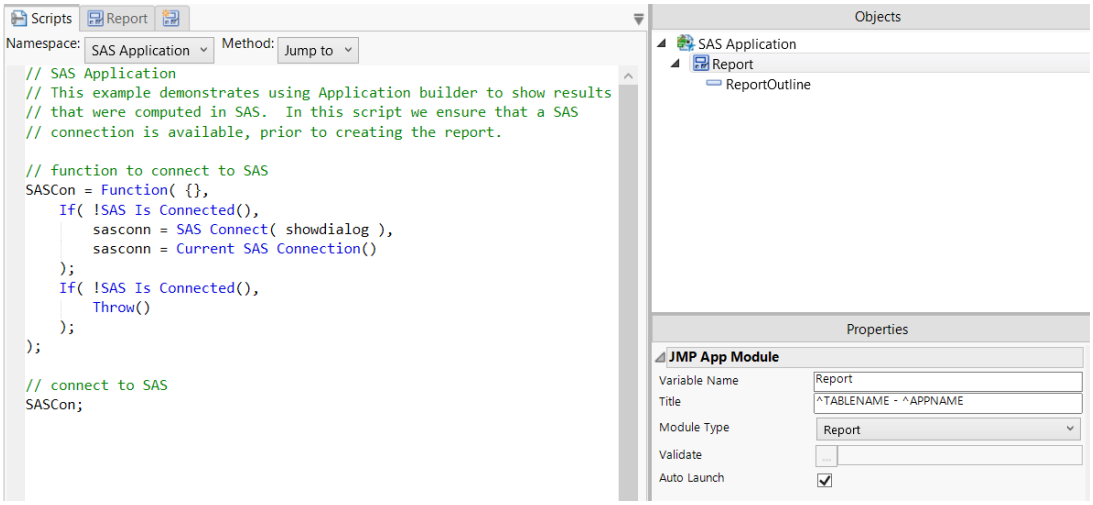
To prevent variable names and values from conflicting among scripts, Application Builder automatically creates the Application namespace and a namespace for each module:

- In the Application namespace, you write scripts that are executed when the user runs the application. Functions defined in the Application namespace can be used in any module.

- In the Module namespaces, you write scripts that are run when an instance of that module is created. If clicking a button in the application opens a new launch window, that launch window is an instance of the module. Two instances of the same module have their own copies of any variables or functions.

To see the scripts in these namespaces, click the **Scripts** tab and then select the namespace in the Namespace list (or in the Objects panel).

Figure 16.10 Application and Module Namespaces



There are two types of scripts: named scripts and anonymous scripts.

Write a Named Script

A *named script* is a function that several controls can use. The `this` argument tells which control is calling the function. In the following example, `Get Button Name` is sent to the `this` argument to print the button name to the log when the button is clicked:

```
Button1Press = Function({this}, Print(this <<Get Button Name))
```

On another button, use `Button1Press` script to produce the same results.

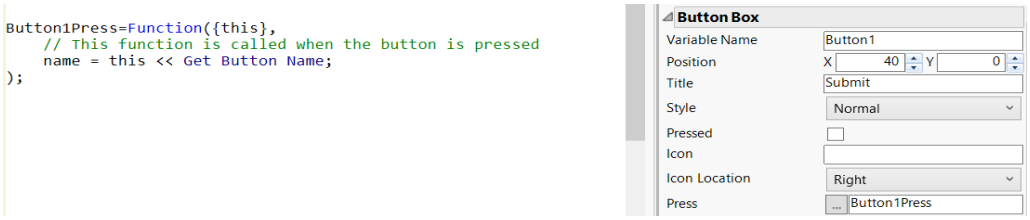
Two other advantages of named scripts: when you add a named script to an object, JMP inserts a placeholder script on the Scripts tab, which you then edit. On the Scripts tab, you can also select a script from the Method list to navigate to that code, which is particularly helpful in long scripts.

Add a named script:

1. Right-click the object, select **Scripts**, and then select the script that you want to add. In this example of creating a button, select the **Press** script. (On macOS, press Control, and then select **Scripts > Press**.)

The object's placeholder script appears on the **Scripts** tab (Figure 16.11). The name of the named script, *Button1Press*, shows up in both the script and the object properties.

Figure 16.11 New Script and Script Properties



2. Edit the object's placeholder script and properties to provide the necessary functionality. For example, the button box shown in Figure 16.11 was renamed. The Title property was changed to *Submit*, the text that appears on the button.

Tip: After you add a script to an object and then delete the object, delete the object's script from the **Scripts** tab if you no longer need the script. This feature prevents scripts that you might want in the future from being deleted.

If you rename the script in the object properties, rename it on the **Scripts** tab also. And if the script is used in another part of the application, rename it there as well.

Write an Anonymous Script

An *anonymous script* is available only to the object that defines it. For example, you might want a simple `Print` statement for one button that is not used elsewhere. By writing an anonymous script, you reduce the number of names to manage in the script. Anonymous scripts also reduce clutter among more important named scripts in the Scripts tab, because you add them to the object properties.

The following examples show two types of anonymous scripts:

```
Print(Button1 <<GetButtonName) // simple anonymous
Function({this}, Print(this <<GetButtonName)) // parameterized anonymous
```

Notice that the simple script sends a message to the "Button1" variable, but `this` is the control in the parameterized script.

Objects such as check boxes might provide additional arguments after `this` that are not otherwise available. One example is an argument that tells which check box in a column of check boxes just changed.

Write an anonymous script:

1. Select the object and click ... in the object properties.

   The anonymous script editor appears.

2. Enter the script and click **OK**.

   The text of the anonymous script appears in the object's properties (rather than the name of a named script).

Tip: Avoid copying and pasting anonymous functions to simplify code maintenance; use a named script instead if the script is needed in more than one place.

Show Specific Scripts

There are several ways to view a script for a specific object:

- Double-click the object on the **Module** tab.
- When an object has multiple scripts, right-click the object, select **Scripts**, and then select the script name. You can also select the **Scripts** tab, select the module name from the **Namespace** list and then select the script name from the **Method** list. Likewise, to see the application scripts, select **Application** from the **Namespace** list.

In each case, the **Scripts** tab appears with the cursor in the first line of the object's script.

Tip: To make a poorly formatted script easier to read, right-click and select **Reformat Script**.

Copy and Paste Objects with Scripts

When you copy and paste an object that has a script, the second instance of the object and script are renamed.

Edit a Platform Script

You can select the object for a platform in an application and edit the script for that platform. For example, you might edit the script for the Multivariate platform to show the scatterplot matrix or to use different columns.

Right-click the platform object on the workspace and select **Edit Platform Script**. After you edit the script and click OK, the object on the workspace is updated.

Edit or Run an Application

To open an application for editing, select **File > Open**, select the .jmpappsource file, and then select **Open**.

Note: If the Table property is empty or the specified data table cannot be found, the application runs anyway but objects requiring the data table will fail to create and JMP displays a warning.

To run an open application, click the Application Builder red triangle and select **Run Application**.

To run a closed application, select **File > Open**, select the .jmpapp file, and then select **Open**.

On Windows, you can also open or run an application from the JMP Home Window by doing one of the following:

- Drag the file from Windows Explorer onto the JMP Home Window or onto a blank application window.
- Double-click the file in Recent Files list.
- Right-click either the .jmpappsource or the .jmpapp file and select either **Edit Application** or **Run Application**.

Options for Saving Applications

JMP provides several options for saving application files.

- When you select **File > Save As** on Windows, you can save the application as an application source file (.jmpappsource), an application (.jmpapp), or a script.
- Selecting **File > Export** on macOS enables you to save the application as an application file or a script. To save the application as an application source file on macOS, select **File > Save As**.

The .jmpappsource file allows an application developer to save the application as is, in order to continue editing the application later. When reopened, the App Builder interface remains in same state as when it was saved. The data tables used to create the application must still be available.

The .jmpapp file is used to distribute applications for others to use. The application runs immediately after opening the file. The application resolves the data tables using the current data tables, files on disk, or by prompting the user, depending on the application settings.

The Save Script red triangle menu provides additional options for saving scripts. When an encrypted script is saved, JMP encloses the script in the `JSL Encrypted()` function to preserve white space and comments.

To Data Table Saves the script to the data table that was used to produce the report. This enables you to run the script again from the data table to recreate the results. When you edit this script from the data table, the application opens in Application Builder, not the script editor.

To Journal Saves a button that runs the script in a journal. The script is added to the current journal. The script contains the path to the data table. Note that if the data table cannot be found, the script does not run.

To Script Window Opens a script editor window and adds the script to it. If you have already saved a script to a script window, additional scripts are added to the bottom of the same script window. Note that if you modify the application significantly in a script window, you might not be able to edit the application in Application Builder. This option is useful when you want to embed the application in a larger JSL process.

Tip: This can also be accomplished by saving as a .jmpapp file and using `Open()` to run the application. However, saving the script can reduce the number of files.

To Add-In Lets JMP users install the application and launch it from a JMP menu. See "Compile Add-Ins with the JMP Add-In Builder".

To Template Saves the application as a template from which you can create new applications. The templates are saved in the following locations based on your operating system and JMP version:

- Windows (JMP): "/C:/Users/<username>/AppData/Roaming/SAS/JMP/<version number>/ApplicationTemplates"
- Windows (JMP Pro): "/C:/Users/<username>/AppData/Roaming/SAS/JMPPro/<version number>/ApplicationTemplates"
- macOS: "/Users/<username>/Library/Application Support/JMP/<version number>/ApplicationTemplates"

Additional Examples of Creating Applications

The following examples illustrate various uses for applications in JMP.

Parameterizing Variables

The following example shows how to create an application with parameterized variables. Users select the variables in a launch window, and then predefined reports are generated.

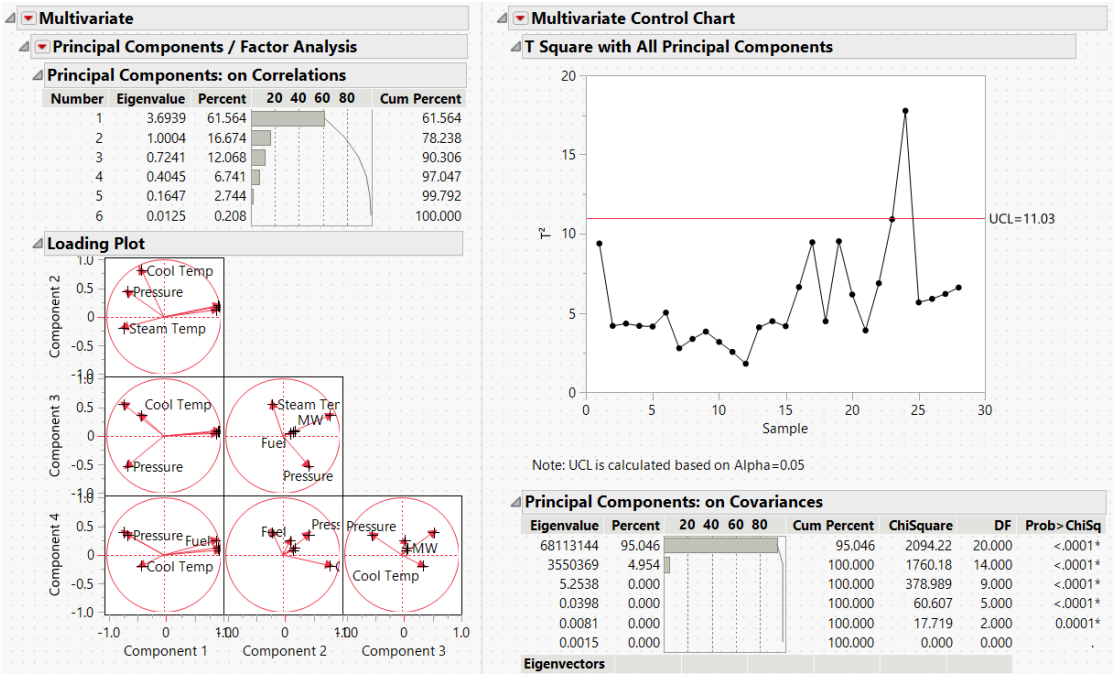
1. Select **Help > Sample Data Library** and open Quality Control/Steam Turbine Historical.jmp.
2. Run the Principal Component Analysis and Loading plot table scripts to generate the reports.
3. Select **File > New > Application**.
4. Select **Blank Application**.

   The Application Builder window appears.
5. Enlarge the window.
6. In the Sources panel, drag each Multivariate report onto the application workspace in a single row.

7. Select **Edit > Select All**.
8. Select **Format > Add Container > H List Box** to align the reports horizontally.
9. Select each report and type `yvar` next to **Y Variable** in the Objects panel.
10. Click the Application red triangle and select **Run Application**.
11. Select the Fuel, Steam Flow, Steam Temp, and MW variables and then click the **Y** button.
12. Click **OK**.

    New Multivariate reports appear in one window.

Figure 16.12 Multivariate Application



Tip: The absolute path to the data table is inserted automatically in the application's Table property. If the data is sample data, the `$SAMPLE_DATA` path variable is automatically used in the path. You can also enter an absolute or relative path. Remember that the user must have access to this path.

Filtering Data in Multiple Reports

In an application that contains several reports, you can select data in one report and then view only that data in all other reports contained in the same window.

To set up the filter, follow these steps:

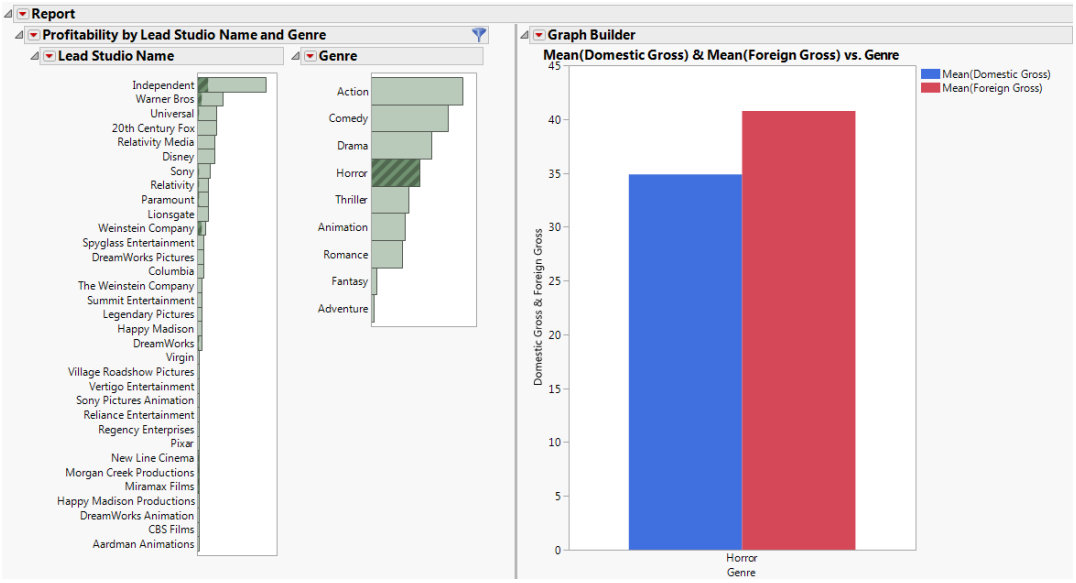
1. Create the application and add two reports to the workspace.
2. Select both reports. From the toolbar, select the HListBox icon.
3. Right-click the primary report and select **Use as Selection Filter**.

   This places the primary report display box in a Data Filter Source Box and the parent report display box in a Data Filter Content Box.
4. Click the Application Builder red triangle and select **Run Application**.

   The reports appear in one window (Figure 16.13).
5. Test the application by selecting a histogram bar in the primary report.

   Only data for the selected bar appears in the second report.

Figure 16.13 Example of Filtered Content



Filtering Data in Separate Reports Using a Local Data Filter

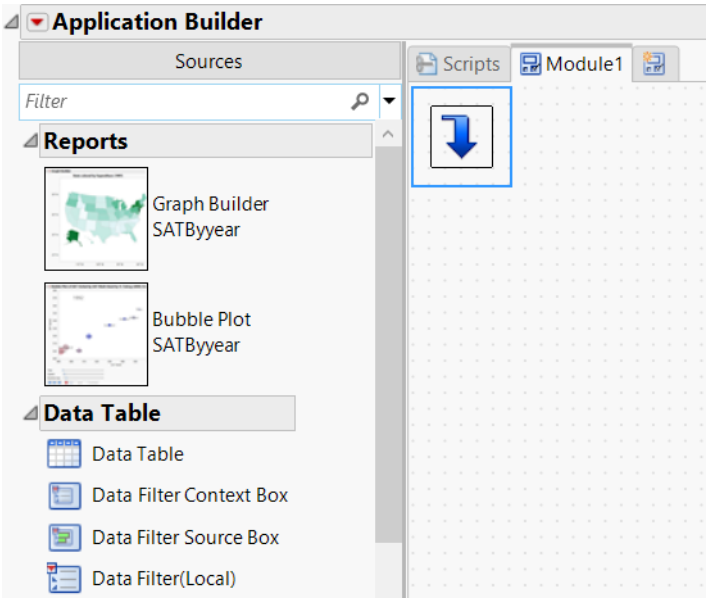
The following example illustrates how to add a Local Data Filter to multiple reports and then filter each report separately.

1. Select **Help** > **Sample Data Library** and open SATByYear.jmp.
2. Run these scripts: Bubble Plot by Region and Graph Builder Map.
3. Resize the report windows to make them smaller, so that a 2x2 arrangement will fit on the screen.
4. Select **File** > **New** > **Application** and choose **Blank Application**.

5. Double-click **Data Filter Context Box**.

   This adds a data filter context box to the workspace. Everything placed in this box will share the same Local Data Filter.

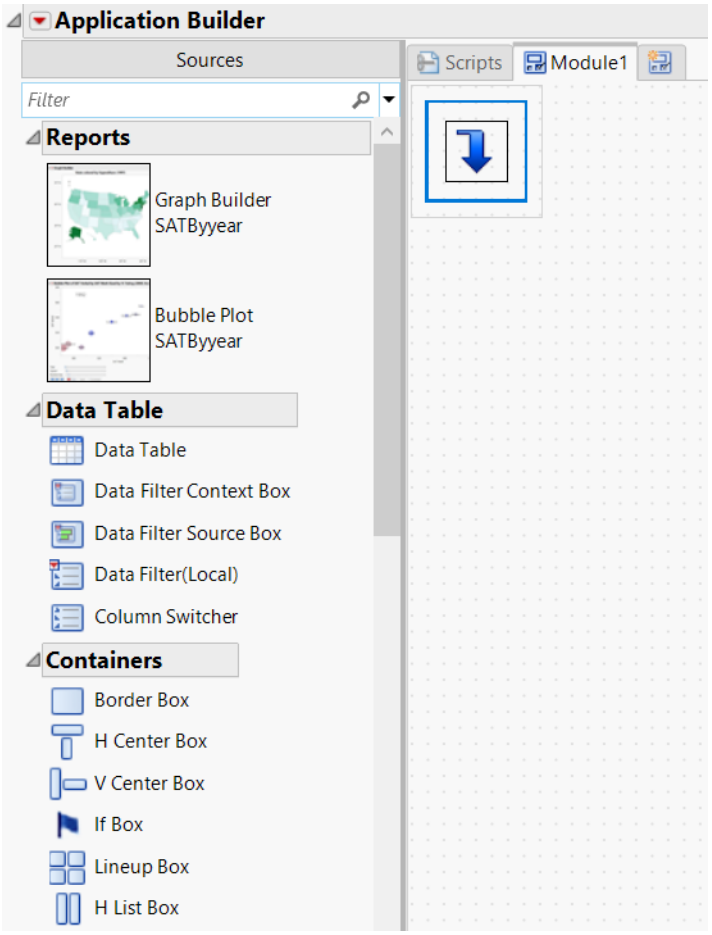
Figure 16.14 Adding a Data Filter Context Box



6. Under Containers, drag a **H List Box** into the Data Filter Context Box.

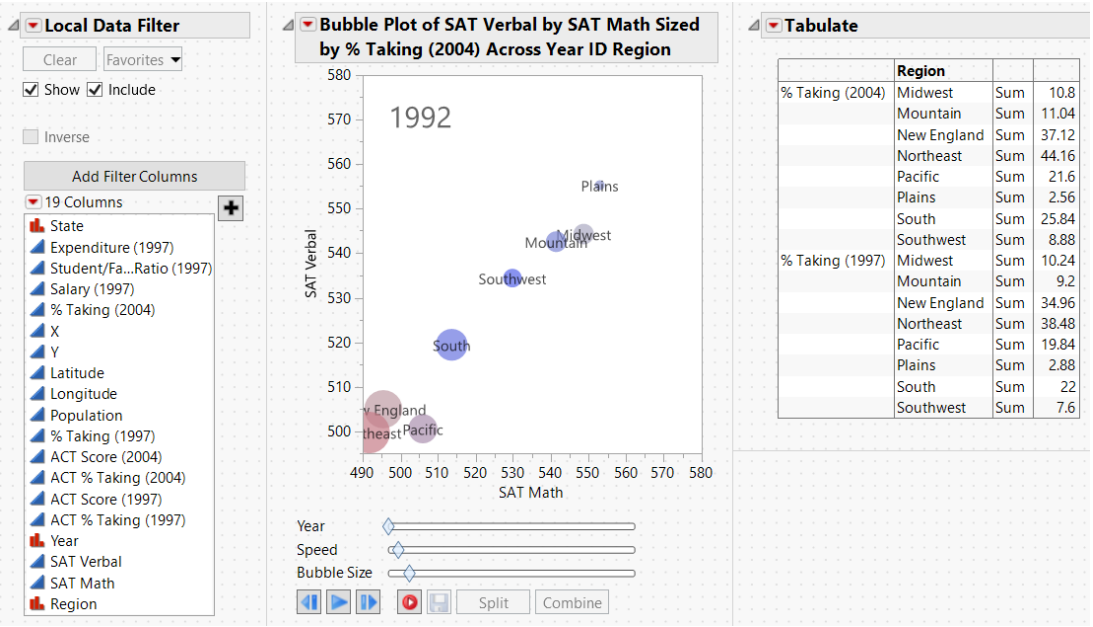
   This allows you to put multiple objects in this box and arrange them horizontally.

Figure 16.15 Adding a H List Box



7. Now drag a Data Filter (local), the Bubble Plot, and the Tabulate report into the H List Box.

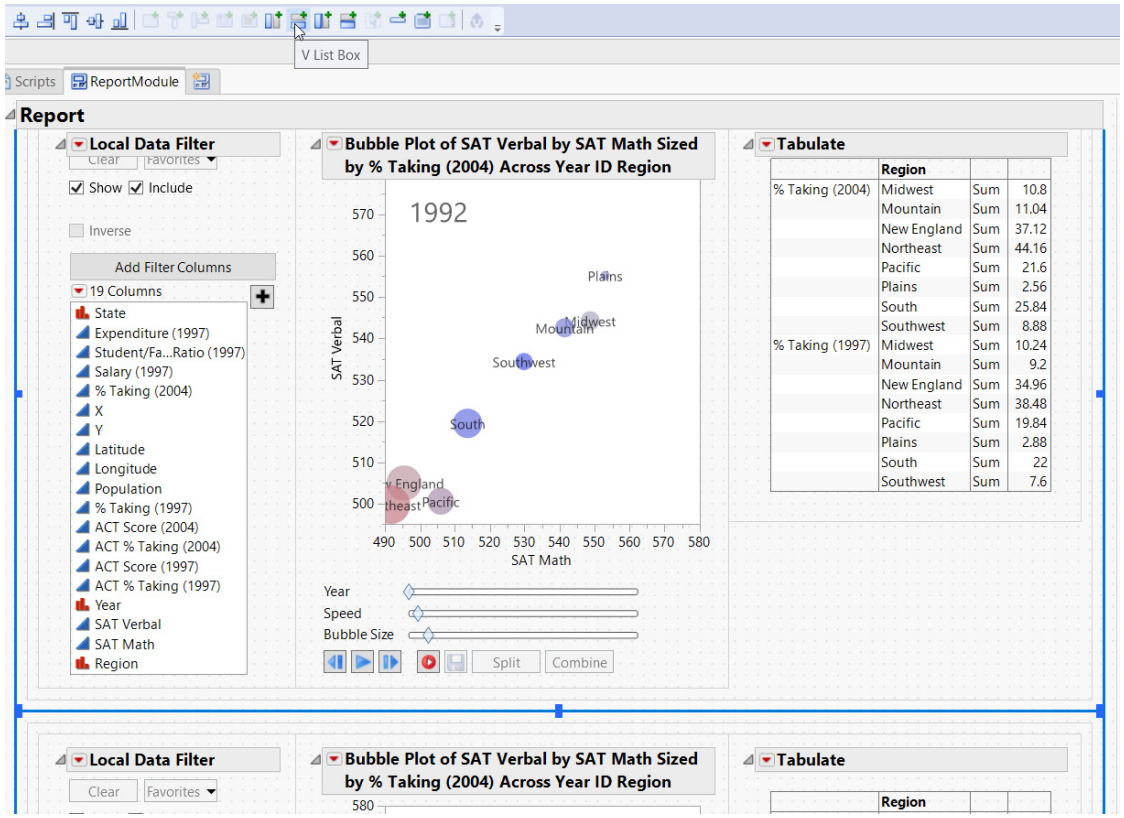
Figure 16.16 Adding Local Data Filter and Two Reports



8. Select the entire Data Filter Context Box. Press Ctrl and drag it below itself, to place a copy below the original set of reports.
9. Select both Data Filter Context boxes, then on the toolbar, click the V List Box icon.

   This arranges both reports into a single list box for display purposes.

Figure 16.17 Selecting Both Data Filter Context Boxes and the V List Box Icon



10. Click the Application Builder red triangle and select **Run Application**.
11. In the Local Data Filter on the top, click SAT Math, then click the Add button . Move the slider to see how it only impacts the reports at top. Do the same for the bottom set of reports.

Selecting Dates

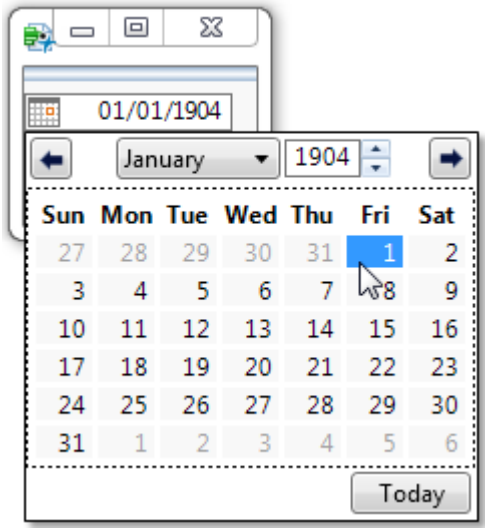
To insert a date selector window into an application, follow these steps:

1. Drag a Number Edit Box from the Sources panel to the workspace.
2. Select the Number Edit Box and then click the button next to Format in the Properties panel.
3. Select **Date** from the list and then select **m/d/y**.
4. Click outside the list.
5. Click the Application Builder red triangle and select **Run Application**.

   The Number Edit Box, which shows a date, appears in a new window.

6. Click the calendar icon to view the date selector window.

Figure 16.18 Example of a Date Selector



The date selector enables you to select the month and year as well as the date and time for the box.

Creating a Launch Window and Report

The Launcher with Report sample application combines a platform launch window with the resulting report. The sample application is displayed when you select **File > New > Application**.

The following example shows you how to arrange objects and include scripts that make up the application.You will recreate the Launcher with Report sample application so that you can create your own applications similarly.

Note: In the Launcher with Report sample application, the variable names of display boxes are customized to describe the display box. For example, “Text1” was renamed “Description” in the sample application. To reduce the number of steps in this example, you do not rename the variables unless one of the scripts refers to the variable name.

1. Select **File > New > Application** and click the **Launcher with Report** sample.

   You will copy and paste JSL from this sample into your copy. Notice that the Iris.jmp sample data table opens when you open the application.
2. Select **File > New > Application** and click **Blank Application**.

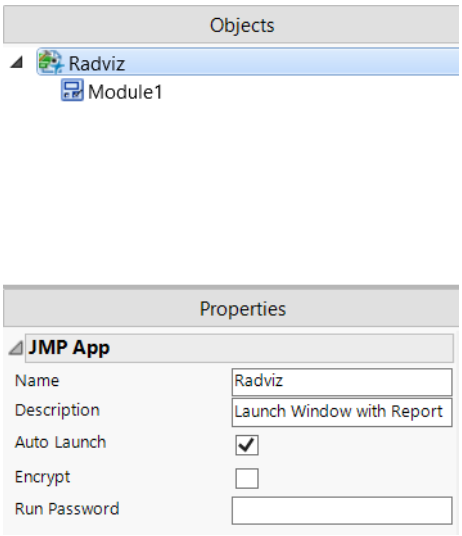
Create the Launch Module

The Launch Module contains objects that create a launch window.

1. In the Objects panel of the blank application, select Application.
2. In the Properties panel:
   - Change the Name from Application to “Radviz” and press Enter.

     This name appears in the title bar when you run the application.
   - Change the Description to “Launch Window with Report”.

Figure 16.19 Customizing the Application

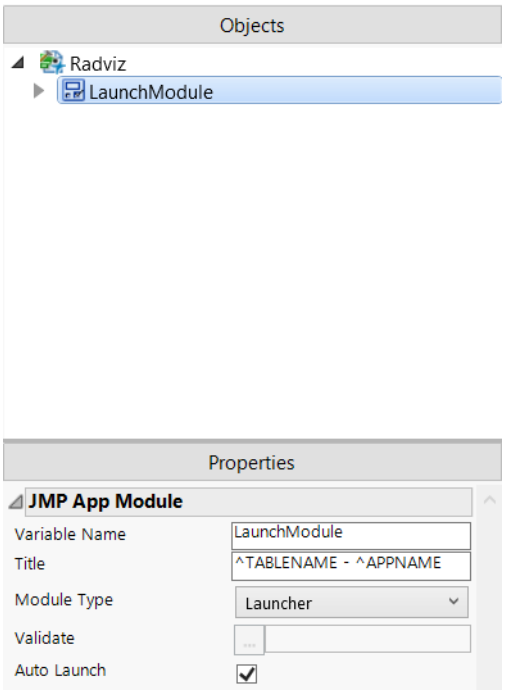


3. Click the **Scripts** tab and select Radviz from the Namespace list.
4. After the green comments, on a new line, type the following expression:

   ```
   dt = Current Data Table();
   ```

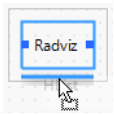
   One of the scripts that you add to this application refers to `dt` to identify the data table.
5. In the Objects panel, select **Module1**.
6. In the Properties panel:
   - Change the Variable Name to “LaunchModule” and press Enter.
   - Change the Module Type to Launcher.
   - Make sure that Auto Launch is selected so that the launch window opens when you run the application.

Figure 16.20 Customizing the LaunchModule



7. Click the **LaunchModule** tab and drag a V List Box (under Containers) onto the workspace.
8. Drag a Text Box (under Display) over the arrow of the V List Box.
9. Double-click Text1, change it to “Radviz”, and then press Enter.
10. Drag an H List Box (under Containers) below the border of the Text Box until a blue line appears.

Figure 16.21 Adding the H List Box



11. Because you have made many changes, select **File > Save** and rename the file. Make sure that .jmpappsource is selected as the file type.

Create the Select Columns Box

1. Drag a Panel Box (under Containers) over the H List Box arrow.
2. In the Properties panel, change the Title to “Select Columns” and press Enter.
3. Drag a Col List Box(All) (under Input) over the Panel Box arrow.

4. In the Properties panel, change the Variable Name to “ColumnList” and press Enter.

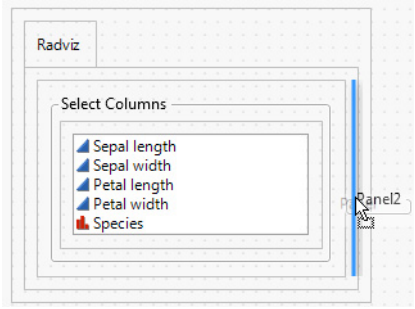
   The script you add later refers to this variable name.
5. Select **DataTable1** in the Objects panel.

   The current data table, Iris.jmp, appears in the Path box.

Create the Cast Selected Columns into Roles Box

1. Drag a Panel Box (under Containers) to the right of the existing Panel Box until a blue line appears.

Figure 16.22 Adding the Panel Box



Note: Make sure that you drag the Panel Box to the H List Box shown in Figure 16.22.

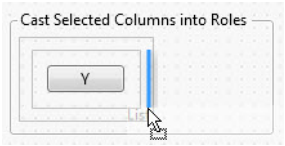
2. In the Properties panel, change the Title to Cast Selected Columns into Roles and press Enter.
3. Drag an H List Box (under Containers) over the Panel Box arrow.

Create the Contents of the Cast Selected Columns into Roles Box

1. Drag a Button Box (under Buttons) over the H List Box arrow.
2. In the Properties panel:
   - Change the Title to “Y” and press Enter.
   - In the Press box, type “SetY” and press Enter.

   You copy and paste the script for this button later.
3. Drag a Col List Box (under Input) inside the right border of the Y Button Box.

Figure 16.23 Adding the Col List Box



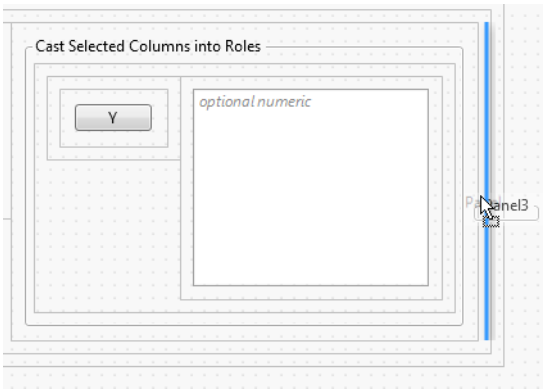
4. In the Properties panel:
   – Change the Variable Name to "YList" and press Enter.
     The script that you add later refers to this variable name.
   – Change the Data type to Numeric.
     The words "optional numeric" appear in the Col List Box.

Tip: Change Min Items to display the minimum number of required items in the Col List Box. If you type 3, then "required numeric" appears three times in the box as a hint to the user.

Create the Action Panel

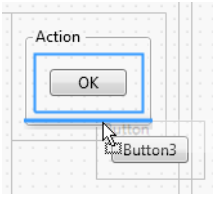
1. Drag a Panel Box (under Containers) to the right of the Panel Box that you just added.

Figure 16.24 Adding a Panel Box



2. In the Properties panel, change the Title to "Action" and press Enter.
3. Drag a Button Box (under Buttons) over the Panel Box arrow.
4. In the Properties panel:
   – Change the Title to OK and press Enter.
   – In the Press box, type "OnOK" and press Enter.
5. Drag another Button Box (under Buttons) below the OK button.

Figure 16.25 Adding a Cancel Button



6. In the Properties panel:
   - Change the Title to “Cancel” and press Enter.
   - In the Press box, type “OnCancel”.
7. Drag a Spacer Box (under Spacers) below the Cancel button.
8. Drag a Button Box (under Buttons) below the Spacer Box.
9. In the Properties panel:
   - Change the Title to “Remove” and press Enter.
   - In the Press box, type “OnRemove” and press Enter.
10. Select the outer border of the objects on the workspace.
11. Right-click and select **Move to Corner**.

    The extra space around the selected object is removed.
12. Select **File > Save** to save your changes in the.jmpappsource file.

Create the Launch Module Script

1. Go back to the Launcher with Report sample application window.
2. Click the Scripts tab.
3. Select LaunchModule as the Namespace and copy the script.
4. Go back to your application window and click the Scripts tab.
5. Delete the script in the LaunchModule namespace and paste the script from the sample application.
6. Select **File > Save** to save your changes in the.jmpappsource file.

Create the Report Module

The Report Module contains an object that builds a graph.

1. In your application window, click the button to the right of the LaunchModule tab to add a new module.
2. In the Properties panel, change the Variable Name to “ReportModule” and press Enter.
3. Make sure that the Module Type is Report.

4. Deselect Auto Launch to prevent the report from opening when the application is run.
5. Drag an Outline Box (under Containers) to the workspace.
6. In the Properties panel, change the Title to “Radviz” and press Enter.
7. Drag a Graph Box (under Display) over the Outline Box arrow.
8. In the Properties panel, change the Variable Name to “Graph” and press Enter.
9. Select the outer border of the objects on the workspace.
10. Right-click and select Move to Corner.

    The extra space around the selected object is removed.
11. Select **File > Save** to save your changes in the.jmpappsource file.

Create the Report Module Script

1. Go back to the Launcher with Report sample application window.
2. Make sure that the Scripts tab is displayed.
3. Select ReportModule as the Namespace and copy the script.
4. Go back to your application window and click the Scripts tab.
5. Delete the script in the ReportModule namespace and paste the script from the sample application.
6. Select **File > Save** to save your changes in the.jmpappsource file.

Your application should look like the Launcher with Report sample application.

Create the Radviz Script

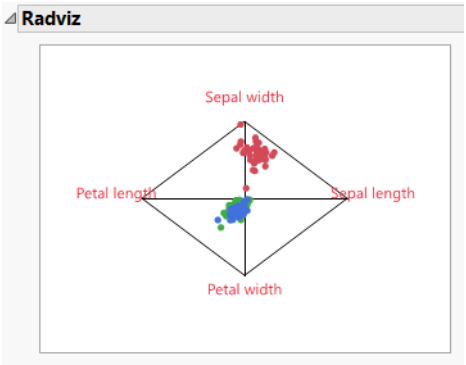
1. Go back to the Launcher with Report sample application window.
2. Make sure that the Scripts tab is displayed.
3. Select Radviz as the Namespace and copy the script.
4. Go back to your application window and click the Scripts tab.
5. Delete the script in the Radviz namespace and paste the script from the sample application.
6. Select **File > Save** to save your changes in the.jmpappsource file.

Test and Save the Final Application

1. In your application window, click the Application Builder red triangle and select **Run Application**.
2. Select the Sepal length and width and Petal length and width columns, click **Y**, and then click **OK**.

A graph of the data should appear.

Figure 16.26 Iris.jmp Graph



If the graph does not appear or errors are generated, make sure that you typed in the correct variable names when necessary. Also make sure that you added the scripts to the Scripts tab for each module.

3. On Windows, select **File > Save As**, select **JMP Application File** from the Save as type list, and then click **Save**.

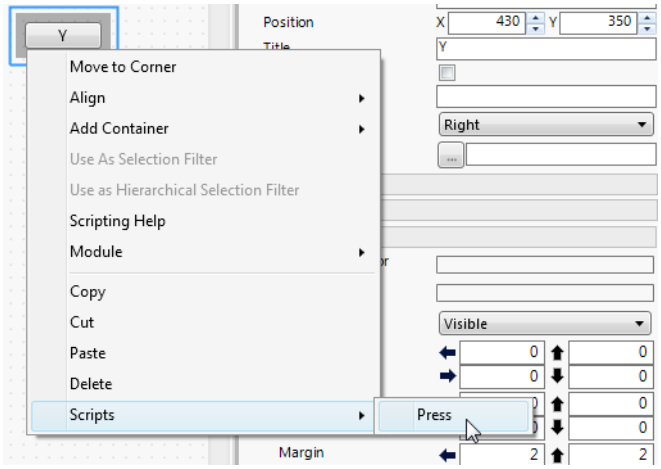
   On macOS, select **File > Export**, select JMP Application, and click **Next** to save the application.

   Your newly created application should match the Launcher with Report sample application.
4. Distribute the .jmpapp file to users. This file cannot be modified.
5. Save the .jmpappsource file so you can edit the application and resave it as a JMP application.

About the Application Scripts

When you create an application, you write scripts to provide functionality to display boxes. You can right-click a display box and select a script from the Scripts menu.

Figure 16.27 Adding a Script Interactively



Selecting a script from the Scripts menu adds a placeholder script to the Scripts tab for the selected module. For example, the Press script shown in Figure 16.27 adds the following script to the Scripts tab:

```
Button5Press=Function({this},
   // This function is called when the button is pressed
   name = this << Get Button Name;
);
```

Fill in the placeholder script with your own JSL. In the application that you previously created, you named two variables `SetY` and `ColumnList`. Rewrite the preceding placeholder script:

```
SetY = Function({},
   // This function will be called when the Y button is
   // pressed

   items = ColumnList << Get Selected;
   YList << Append(items);
);
```

`Get Selected` is passed to the `ColumnList` because you gave the Col List Box this variable name (`ColumnList`) in the object properties. `Get Selected` returns a list of the selected columns. `Append(items)` appends the selected columns to YList, the variable assigned to the Col List Box in the Cast Selected Columns into Roles window.

As shown when you previously created the application, you can write the script from scratch and indicate the script name in the object properties. To get help with syntax for a selected object, right-click the object and select Scripting Index.

Compile Add-Ins with the JMP Add-In Builder

A JMP add-in is a JSL script that you can run anytime from the JMP Add-Ins menu. You can create submenus to group your JMP add-ins and have many levels of menus, if desired.

Experienced JSL script writers can create scripts that extend JMP in many ways (for example, add a custom analytical tool or a user interface to communicate with a database). The JMP add-in architecture simplifies deploying and using these complicated scripts.

You can send co-workers a suite of scripts and tell them how to run them. Another option is sending a single add-in file that any JMP user can install and then use just like any other part of JMP.

As a JMP user, you might be given add-ins to use by co-workers. You can also find add-ins on the JMP website at https://www.jmp.com/addins.

Compile Scripts into an Add-In Using Add-In Builder

To create a JMP add-in, you write the scripts and gather files that are used in the add-in. Then you compile the scripts into an add-in.

- On Windows, select **File > New > Add-In**.
- On macOS, select **File > New > New Add-In**.
- In Dashboard Builder, click the red triangle and select **Save Script > To Add-In**.

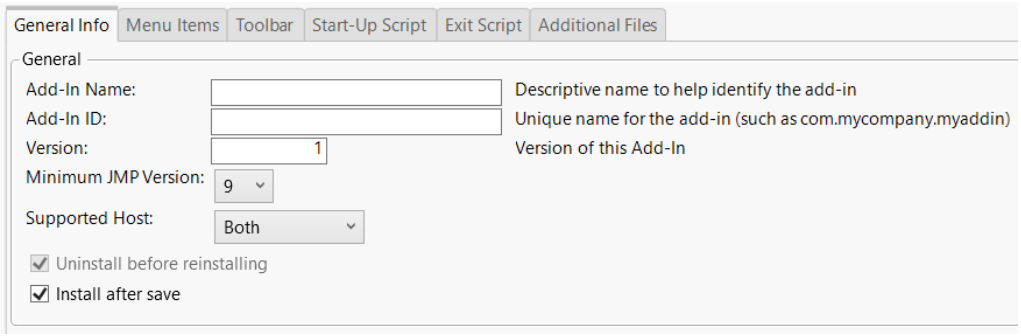
The process of compiling scripts into an add-in involves the following steps:

- "Add General Information"
- "Create Menu Items"
- "Create Toolbars (Windows Only)"
- "Specify Start-up and Exit Scripts (Optional)"
- "Add Additional Files"
- "Save the Add-In"
- "Test the Add-In"

Add General Information

First, in the **General Info** tab, add the general information that identifies and sets up your add-in.

Figure 16.28 Add-In Builder General Info Tab



1. Enter a name for the add-in.

   This is the name of the registered add-in, which appears in the **View > Add-Ins** window.

2. Enter a unique identifier string.

   Unique ID strings are case-insensitive. To ensure uniqueness, it is strongly recommended to use reverse-DNS names (for example, `com.mycompany.myaddin`). The ID string must meet the following requirements:

   – It can be up to 64 characters in length.
   – It must begin with a letter.
   – It should consist only of letters, numbers, periods, and underscores.
   – It should contain no spaces.

   In JSL, use this string to refer to the add-in (for example, in calls to `Get Addin()` or as an argument to $ADDIN_HOME).

3. Enter the version of the add-in.

   If you decide to make changes to the add-in at a later date, you can update the version number and then verify that users have the correct version.

4. Select the minimum version of JMP that the add-in works on.

   Note: Add-ins were introduced in JMP 9, so no prior versions are supported. And when saving an application as an add-in, select 10 or 11 as the JMP minimum version.

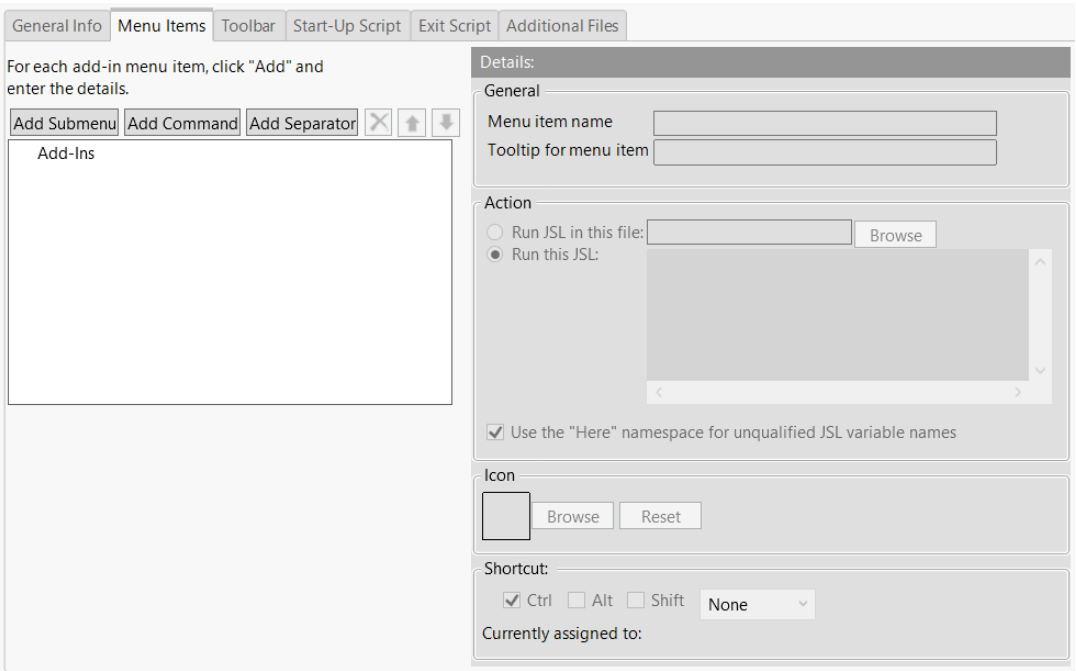
5. Select whether you want the add-in to be supported on Windows, macOS, or both.
6. (Optional) Select the check box next to **Install after save** if you want to install the add-in after saving it.

   If you do not select this option, the add-in is not installed once you save it, and it does not appear as a selectable menu item in the **Add-Ins** menu.

Create Menu Items

1. Click the **Menu Items** tab.

Figure 16.29 Add-In Builder Menu Items Tab



2. (Optional) Click **Add Submenu**.

   If you have multiple add-ins, you can group them under a submenu.
3. If you add a submenu, next to **Menu item name**, enter the name of the submenu.

   This name appears in the **Add-Ins** menu.
4. Click **Add Command**.
5. Next to **Menu item name**, enter the name of the add-in command.
6. (Optional) Next to **Tooltip for menu item**, enter the content that appears as a tooltip when the users place their cursor over the menu item.
7. Add the script. Select either **Run this JSL** and copy and paste in your script, or select **Run JSL in this file** and click **Browse** to find the file containing your script.
8. (Optional) Select **Use the "Here" namespace for unqualified JSL variable names** to ensure that all unqualified JSL variables are in the `Here` namespace, and local only to the script.

Notes:

- If your script creates a custom menu or toolbar, the variables are in the `Here` namespace by default.
- For more information about the `Here` namespace, see "Advanced Scoping and Namespaces".

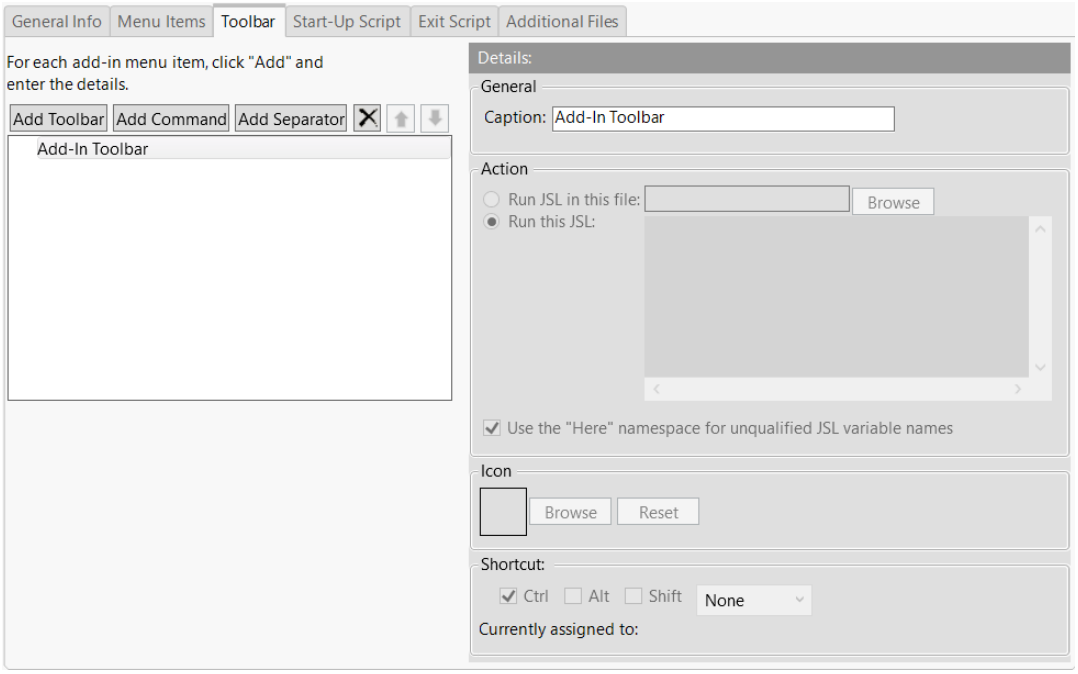
9. (Optional) Browse to add an icon that appears next to the menu item in the **Add-Ins** menu.
10. (Optional, Windows only) Create a keyboard shortcut for the add-in.
11. To add multiple menu items, repeat the steps.

    You can add multiple levels of submenus and add-in commands.
12. Click **Save** and save the add-in to the desired directory.
13. Click **Close**.

Create Toolbars (Windows Only)

1. Click the **Toolbar** tab.

Figure 16.30 Add-In Builder Toolbar Tab



2. Click **Add Command**.
3. Next to Caption, type the name of the toolbar item. This name is only for internal use. The user does not see it.

4. Add the script. Select either **Run this JSL** and copy and paste in your script, or select **Run JSL in this file** and click **Browse** to find the file containing your script.
5. (Optional) Select **Use the "Here" namespace for unqualified JSL variable names** to ensure that all unqualified JSL variables are in the `Here` namespace, and local only to the script.
6. Click **Save** and save the add-in to the desired directory.
7. Click **Close**.

Note: For more information about customizing menus in JMP, see *Using JMP*.

Specify Start-up and Exit Scripts (Optional)

Click the **Start-Up Script** tab to add a script that runs when JMP starts up (and the add-in starts). You can select an existing script (**Run JSL in this file**) or copy and paste in a script (**Run this JSL**). For example, you could provide a message telling the user that the add-in is installed upon start-up.

Click the **Exit Script** tab to add a script that runs when JMP exits or when you disable the add-in. You can select an existing script (**Run JSL in this file**) or copy and paste in a script (**Run this JSL**). For example, you could provide a prompt for the user to export a JMP data table upon exiting or disabling the add-in.

Add Additional Files

If your script calls other scripts, or contains graphics or data tables, add those files here.

Save the Add-In

Save the add-in by clicking the **Save** button on any tab. This effectively creates the add-in.

- If you selected the **Install after save** option in the General Info tab, then the add-in menu item appears in the **Add-Ins** menu immediately.
- If you did not select the **Install after save** option, when you open the saved add-in file, you are prompted to install the add-in.

Test the Add-In

Once your add-in is installed, test your add-in:

1. Select **View > Add-Ins**.
2. Select your add-in and click **Unregister**.
3. Reinstall your add-in by either selecting **File > Open** in JMP, or by double-clicking your .jmpaddin file.
4. Ensure that the menu and toolbar button run your script correctly and that the script itself runs correctly.

Edit an Add-In

To edit a saved add-in:

1. Select **File > Open**.
2. Navigate to and select the add-in file.
3. Select one of the following options:
   - On Windows click the arrow to the right of the **Open** button and select **Open using Add-In Builder**.
   - On macOS, select the **Edit after opening** option.
4. Click **Open**.

   The file opens in the Add-In Builder. Update the arguments and then save the changes.

Share an Add-In

Once you have a .jmpaddin file, you can share the file with other users. Email the file or you place it in a shared location, such as a network folder, or on the JMP File Exchange (located online in the JMP User Community).

When JMP users open the file, the add-in files are extracted into the appropriate folder, and the add-in is registered and installed. The add-in now appears in the user's JMP Add-Ins menu.

Installing Multiple Add-Ins

If you want to install multiple add-ins, copy the add-ins into the following location:

- Add-In files on Windows are located here:
  - %ALLUSERSPROFILE%/SAS/JMP/AddIns (any user on this machine can access the add-in)
  - %APPDATA%/SAS/JMP/AddIns (only the current user on this machine can access the add-in)
- Add-In files on macOS are located here:
  - /Library/Application Support/JMP/AddIns (any user on this machine can access the add-in)
  - ~/Library/Application Support/JMP/AddIns (only the current user on this machine can access the add-in)

When JMP starts, the addinRegistry.xml file is read, which contains information about previously registered JMP add-ins. Then JMP looks in the add-in folders for any other add-ins and installs them automatically.

Notes:

- The Home Folder for discovered add-ins does not have to be the AddIns subfolder in which the addin.def file was found. The addin.def file can be the only file in that subfolder and have a Home setting that points to some other location where the add-in files actually reside.
- If an automatically discovered add-in has the same unique ID as an add-in that was explicitly registered, the automatically discovered add-in is used.

Register an Add-In Using JSL

If your add-in files are not contained within a .jmpaddin file, you can use the `Register Addin()` JSL function to manually register the addin.def file. This installs and registers the add-in.

- For information about the JSL functions, see the *JSL Syntax Reference*.
- For information about creating the addin.def file, see "Create an Add-In Manually".

Notes:

- JMP might find a file named addin.def in the specified home folder. If so, values from that file are used for any optional arguments that are not included in the `Register Addin()` function.
- The addin.def file is used only for values that are not provided in the `Register Addin()` function. This function is useful while developing, but not necessary, since the addin.def file is enough to register an add-in.

Create an Add-In Manually

The addin.def file is a simple text file containing name-value pairs that provide registration information about a JMP add-in. Here are the name-value pairs to include in the addin.def file:

`name` Optional. The name that can be displayed in the JMP user interface wherever add-in names are displayed, instead of the unique ID. This name is displayed if no localized names are provided or when JMP is run under a language for which you did not provide a localized name. Named *Add-In Name* in Add-In Builder.

`name_xx` Optional. Allows the user-friendly name to be localized for different languages, where *xx* is the two-letter ISO 639-1 code for the language. If you include localized names,

you should still include a language-neutral name in case JMP is running under regional settings for which you do not have a localized name. Not available in Add-In Builder.

`id` Required. The unique ID for your add-in. The string can contain up to 64 characters. The string must begin with a letter and contain only letters, numbers, periods, and underscores. Reverse-DNS names are recommended to increase the likelihood of uniqueness. Named *Add-In ID* in Add-In Builder.

`home` Optional. The path to the add-in files. The Home Folder for the add-in is assumed to be the folder where `addin.def` is located. You need to include a setting for home only if the Home Folder is somewhere else (for example, a network shared folder). Not available in Add-In Builder.

`home_win` Optional. The path to the add-in files to be used when JMP is running on Windows. Overrides the value specified for home on Windows, if any. Not available in Add-In Builder.

`home_mac` Optional. The path to the add-in files to be used when JMP is running on the macOS. Overrides the value specified for home on macOS, if any. Not available in Add-In Builder.

`autoLoad` Optional, Boolean. The default value is True (1). Determines whether this add-in is initially configured to load automatically during JMP start-up. Named *Install after save* in Add-In Builder.

`host` Optional. Valid values are `Win` and `Mac`. Named *Supported Host* in Add-In Builder.

`minJMPVersion` Optional. Valid values are integers corresponding to the JMP major version that is the minimum version that the add-in supports. Named *Minimum JMP Version* in Add-In Builder.

`maxJMPVersion` Optional. Valid values are integers corresponding to the JMP major version that is the maximum version that the add-in supports. Use this setting *only* if there is a known incompatibility between your add-in and a specific version of JMP. You should provide a new version of the add-in for later versions of JMP. Not available in Add-In Builder.

Example of a addin.def File

```
id="com.mycompany.myaddin"
name="My Add-In's Friendly Name"
name_fr="My Add-In's French Name"
name_de="My Add-In's German Name"
home="\\server\share\myjmpaddin"
Autoload=1
MinJMPVersion=9
```

Example of a JMP Add-In

A sample add-in named Simple Calculator.jmpaddin is located in one of the following folders:

- On Windows: C:/Program Files/SAS/JMP/16/Samples/Scripts
- On macOS: /Library/Application Support/JMP/16/Sample/Scripts

Note: On Windows, in JMP Pro, the "JMP" folder is named "JMPPro".

To see what the add-in contains, change the extension to .zip and unzip it into a new folder. To see how it works, change the extension back to .jmpaddin and install.

The add-in contains the following files:

addin.def

Provides the specification for JMP to register the add-in. It contains only these two lines:

```
id="com.jmp.sample.calculator"
name="Simple Calculator"
```

addin.jmpcust

Provides the menu customization file that is created when you interactively create a custom menu. Your customizations are shown only to the current user; other users who log on to the computer and open JMP do not see your personal toolbars and menus. This example places the add-in menu item into the default **Add-Ins** menu.

calculator.jsl

A JSL script that creates a basic calculator.

calc_icon.gif

The image used as the calculator's icon.

To download more add-ins, visit
https://community.jmp.com/t5/File-Exchange/ct-p/FileExchange.

Manage JMP Add-Ins

Install Add-Ins

A JMP add-in is a file that has the .jmpaddin extension. You can install it in one of two ways:

1. Select **File > Open**.
2. Navigate to the .jmpaddin file and select it.

3. Click **Open**.

You can also double-click the .jmpaddin file or drag it onto the JMP Home Window.

View Your Add-Ins

Select **View > Add-Ins** to see the add-ins that you have already installed.

Reorder Your Add-Ins

Select **View** > **Add-Ins**, then use the Load Order arrows to move each add-in where you want it to appear in the Add-Ins menu.

Update Add-Ins

If you are given an update to an add-in you already have installed, you can just install the update as you did the original add-in. Doing so overwrites the old version with the new one.

Disable, Enable, and Remove Add-Ins

To temporarily disable an add-in without removing it

1. Select **View > Add-Ins**.
2. Select the add-in that you want to disable.
3. Deselect the **Enabled** check box.

To enable a disabled add-in

1. Select **View > Add-Ins**.
2. Select the add-in that you want to enable.
3. Select the **Enabled** check box.

To unregister an add-in

1. Select **View > Add-Ins**.
2. Select the add-in that you want to remove.
3. Click **Unregister**.

Chapter 17

Common Tasks

Getting Started with Sample Scripts

Examining working scripts line-by-line is one of the best ways to learn JSL. This chapter describes common tasks in JMP, such as converting date/time values and extracting specific values from reports. Sample scripts that address these issues are installed with JMP in the Samples/Scripts folder so you can run them yourself.

Contents

Run a Script at Start Up 927
Convert Character Dates to Numeric Dates 928
Format Date-Time Values and Subset Data 929
Create a Formula Column 930
Extract Values from an Analysis into a Report 932
Create an Interactive Program 936

Run a Script at Start Up

You can run the same script every time you start JMP. For example, you might include the definitions of some utility functions or a namespace that you want available throughout your JMP session. You might also set some preferences in a startup script to keep preferences consistent. Preferences persist across JMP sessions, but you could explicitly reset some preferences in case they were changed at some point in the previous session.

Name the script jmpStart.jsl and place it in one of the following folders, as appropriate for your operating system. When JMP starts, JMP looks for the jmpStart.jsl script in these folders in the order in which they are listed here. The first one that is found is run, and the search immediately stops.

Note: Some path names in this section refer to the "JMP" folder. On Windows, in JMP Pro, the "JMP" folder is named "JMPPro".

On Windows:

1. C:/Users/<username>/AppData/Roaming/SAS/JMP/16
2. C:/Users/<username>/AppData/Roaming/SAS/JMP

On macOS:

1. /Users/<username>/Library/Application Support/JMP/16
2. /Users/<username>/Library/Application Support/JMP

The jmpStart.jsl script runs only for a particular user on a computer. You can add a script named jmpStartAdmin.jsl in one of the following places, as appropriate for your operating system. This script is run for every user on a computer. JMP runs jmpStartAdmin.jsl first if found. Then JMP runs jmpStart.jsl if found.

On Windows:

1. C:/ProgramData/SAS/JMP/16
2. C:/ProgramData/SAS/JMP

On macOS:

1. /Library/Application Support/JMP/16
2. /Library/Application Support/JMP

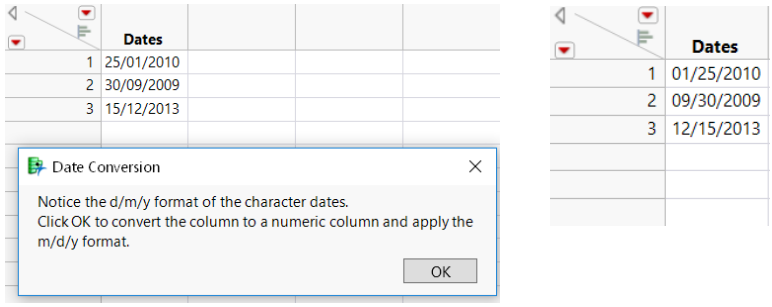
Convert Character Dates to Numeric Dates

Data might appear to be numeric in the data table. However, the column properties may specify a character data type. To manipulate the data as date/time values, convert the column to a numeric column and specify how you want the values to appear.

Convert Dates.jsl creates a data table, specifies the data input format, changes the column to a numeric continuous column, and applies the m/d/y format (Figure 17.1).

```
// Create a data table with character dates.
dt = New Table( "Example",
   Add Rows( 3 ),
   New Column( "Dates",
      Character,
      Nominal,
      Set Values( {"25/01/2010", "30/09/2009", "15/12/2013"} )
   )
);

// Display a modal dialog for the user to confirm the format conversion.
nw = New Window( "Date Conversion",
   <<Modal,
   tb = Text Box(
      "Notice the d/m/y format of the character dates.
Click OK to convert the column to a numeric column and apply the m/d/y
   format."
   )
);

/* Apply the Numeric data type.
Specify the Informat (input format) value "d/m/y".
Specify the Format (display format) value "m/d/y".
Apply the Continuous modeling type */
col = Column( dt, "Dates" );
col << Data Type( "Numeric", Informat( "d/m/y" ), Format( "m/d/y" ) );
col << Modeling Type( "Continuous" );
```

Figure 17.1 Converting Character Dates (Before and After)



When you change the column's data type from character to numeric, defining the format in which the data were entered is important. In this example, `Informat( "d/m/y" )` defines the input format. `Format( "m/d/y" )` defines the new display format. If `Informat()` is omitted, the `Format()` value is applied as both the input and display format. This results in missing values for some data.

Modify Convert Dates.jsl to see for yourself.

1. Open Convert Dates.jsl from the sample scripts folder.
2. Right-click the script window and select **Show Line Numbers**.
3. On line 9, change "25/01/2010" to "01/25/2010".
4. On line 27, delete `Informat( "d/m/y" ),` (including the comma).
5. Run the script.

   `Format( "m/d/y" )` is applied to the column. Only "01/25/2010" appears in the column. The other values are missing; "30/09/2009" and "15/12/2013" are not valid m/d/y values.

Format Date-Time Values and Subset Data

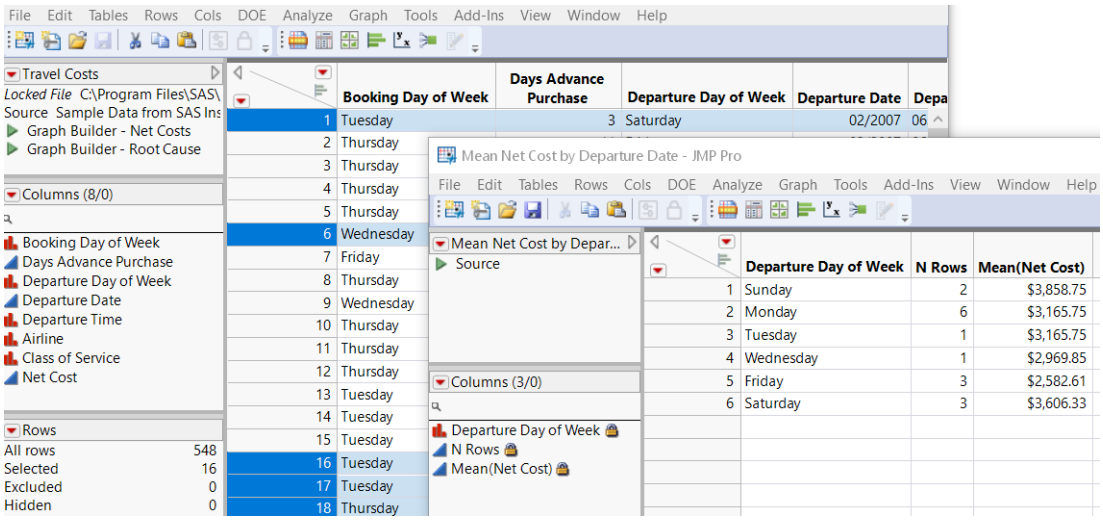
How can you work with dates in JMP? JMP provides a number of formats that you can use to make comparisons and then subset data based on the date.

Select Where Using Dates.jsl applies the Date MDY format to a column of departure dates and subsets the data. A summary table of mean net costs by departure date then appears (Figure 17.2).

```
/*  How can you work with dates in JSL? JMP provides a number of formats
that you can use to make comparisons and then subset data based on the date.
*/

hdt = Open( "$SAMPLE_DATA/Travel Costs.jmp" );
```

```
/* Apply the Date MDY format to Departure Date values and then select only
February dates. */
hdt << Select Where(
      (Date MDY( 02, 01, 2007 ) <= :Departure Date < Date MDY( 03, 1, 2007 ))
      );

/* Subset the selected rows into two tables: one table contains February
departure dates, the other contains all data for those departure dates. */
nt1 = hdt << Subset( Columns( :Departure Date ),
   Output Table Name( "February Departure Date" ) );
nt2 = hdt << Subset( Output Table Name( "February Data" ) );

/* Create a summary table, grouping mean cost by day of week that departure
took place. */
sumDt = nt2 << Summary(
   Group( :Departure Day of Week ),
   Mean( :Net Cost ),
   Output Table Name( "Mean Net Cost by Departure Date" )
);
```

Figure 17.2 The Original Table and the Final Summary Table

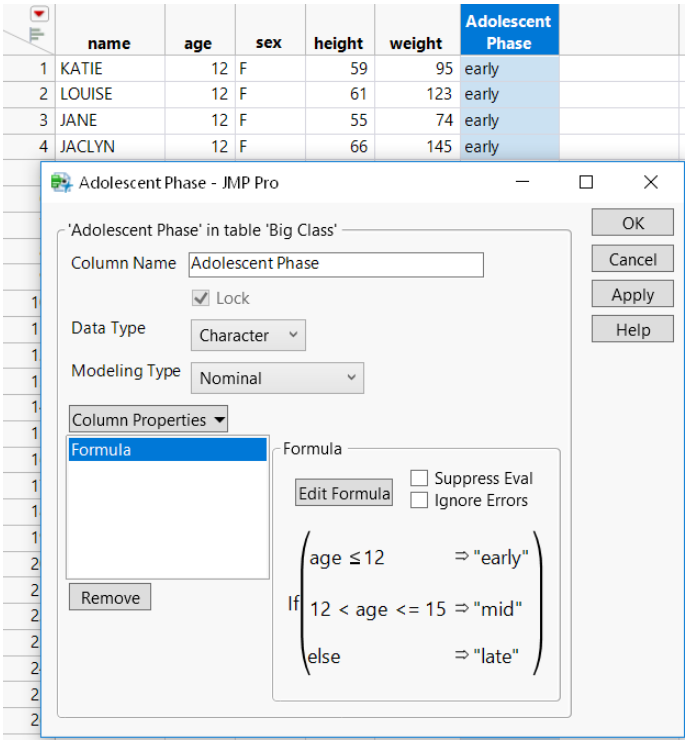


Create a Formula Column

How do you create a formula column that combines conditional expressions with value comparisons? Create a Formula Column.jsl shows how to create a new formula column that evaluates ages in Big Class.jmp and returns the result in the new column (Figure 17.3).

```
/* Scenario:

How do you create a formula column that combines conditional expressions
with value comparisons? This script shows how to create a new formula
column that evaluates ages in Big Class.jmp and returns the conditional
result in the new column. */

dt = Open( "$SAMPLE_DATA/Big Class.jmp" );

/* Create a new character column for the formula.
Insert "early" in the new column if the age is less than or
equal to 12. Insert "mid" if the age is less than or equal to
15 but greater than 12. For ages greater than 15, insert "late".
*/

dt << New Column( "Adolescent Phase",
   Character,
   Formula(
      If( :age <= 12, "early",
         12 < :age <= 15, "mid",
         "late"
         )
      )
);
```

Figure 17.3 Conditional Expression in Formula



Extract Values from an Analysis into a Report

How do you capture specific results of an analysis into a custom report using JSL?

The JMP platforms in the Analyze and Graph menus contain two objects known as the analysis and report layers. Messages are sent to the analysis layer that generate the desired results.

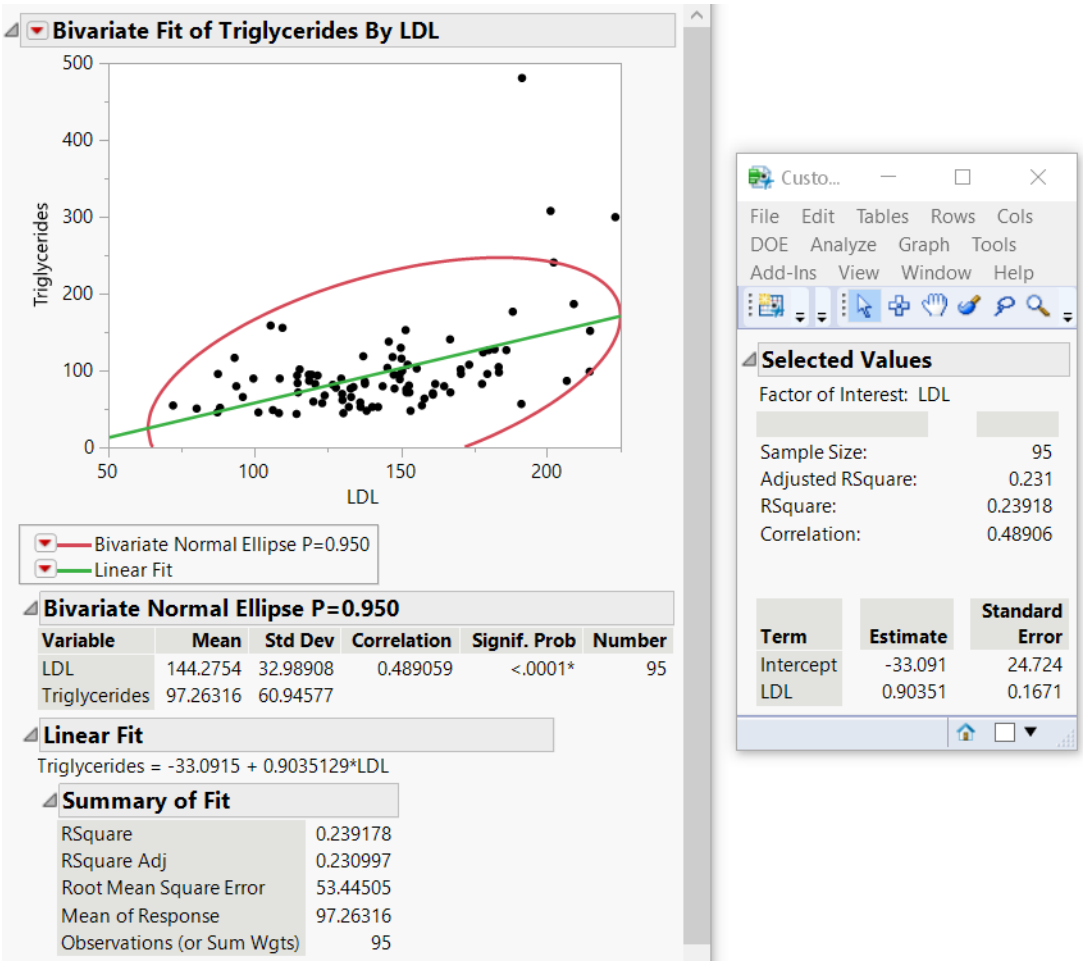
Extract Values from Reports.jsl performs a Bivariate analysis and shows results such as the sample size, RSquare, and Correlation in a new report window. Figure 17.4 shows the Bivariate report along with the customized report, though the Bivariate report window closes after the script is finished.

```
/* Scenario: How do you capture specific results of an analysis
in a report using JSL?

The JMP platforms in the Analyze and Graph menus contain two objects
known as the analysis and report layers.

Messages are sent to the analysis layer that generate the desired results.

This script performs a Bivariate analysis and shows results
such as the sample size, RSquare, and Correlation in a new report window.
*/

sd = Open( "$SAMPLE_DATA/Lipid Data.jmp" );

biv = Bivariate(        // biv is the analysis layer.
  Y( :Triglycerides ),
  X( :LDL ),
  Density Ellipse( 0.95, {Line Color( {213, 72, 87} )} ),
  Fit Line( {Line Color( {57, 177, 67} )} ),
);

/* Make sure the second Outline Box (called "Correlation")
in the Bivariate report is open. You can then see which content
is extracted into the Custom report. */
report(biv) [Outline Box( 2 )] << Close( 0 );

reportbiv = biv << Report;  // reportbiv is the report layer.

/* The density ellipse is generated first.
Extract the correlation coefficient. */
corrvalue = reportbiv[Outline Box( 2 )][Number Col Box( 3 )] << Get( 1 );

/* ...followed by Fit Line
Extract the numeric values from the Summary of Fit report
and place them in a matrix. */
sumfit = reportbiv[Outline Box( 4 )][Number Col Box( 1 )] << Get as Matrix;

// Extract the values of RSquare and AdjRSquare as one by one matrices.
rsquare = sumfit[1];
adjrsq = sumfit[2];
avg = sumfit[4];
samplesize = sumfit[5];

/* Extract the first column of the Parameter.
Estimates report as two objects. */
term = reportbiv[Outline Box( 7 )][String Col Box( 1 )] << Get();
```

```
// Clone the report layer as a String Col Box.
cloneterm = reportbiv[Outline Box( 7 )][String Col Box( 1 )] << Clone Box;

// Extract the Parameter Estimates values as a matrix.
est = reportbiv[Outline Box( 7 )][Number Col Box( 1 )] << Get as Matrix;

// Extract the Standard Error values as a matrix.
stde = reportbiv[Outline Box( 7 )][Number Col Box( 2 )] << Get as Matrix;

dvalues = [];
dvalues = samplesize |/ adjrsq |/ rsquare |/ corrvalue;
sfactor = term[2];

dlg = New Window( "Custom Report",
   Outline Box( "Selected Values",
      /* The Lineup box defines a two-column layout, each of which contains
      a Text Box. */
      Lineup Box( N Col( 2 ),
      Text Box( "Factor of Interest: " ),
      Text Box( sfactor ), ),
      tb = Table Box(
         /* Display an empty string in the first column
            and the text in the second column. */
         String Col Box( "  ",
            {"Sample Size: ", "Adjusted RSquare: ", "RSquare: ",
            "Correlation:"}
         ),
         // Insert a 30 pixel x 30 pixel spacer between the columns.
         Spacer Box( Size( 30, 30 ) ),
         ,
         /* Display an empty string in the first column
         and the dvalues in the second column. */
         Number Col Box( "  ", dvalues )
      ),
      // Insert a 1 x 30 spacer.
      Spacer Box( Size( 0, 30 ) ),
      ,
      Table Box(
         /* Display the cloned String Col Box followed by a spacer.
            Then insert the Parameter Estimates and Standard Error values. */
         CloneTerm,
         Spacer Box( Size( 10, 0 ) ),
         ,
         Number Col Box( "Estimate", est ),
         Spacer Box( Size( 10, 0 ) ),
```

```
            ,
            Number Col Box( "Standard Error", stde )
        )
    )
);
tb << Set Shade Headings( 0 ); // Turn off shaded table headings.
tb << Set Heading Column Borders( 0 ); // Turn off table column borders.
Close( sd ); // Close the data table.
```

Figure 17.4 Customized Report from the Bivariate Analysis



Create an Interactive Program

How do you gather numeric input from the users, perform a calculation on that input, and show the results in a new window?

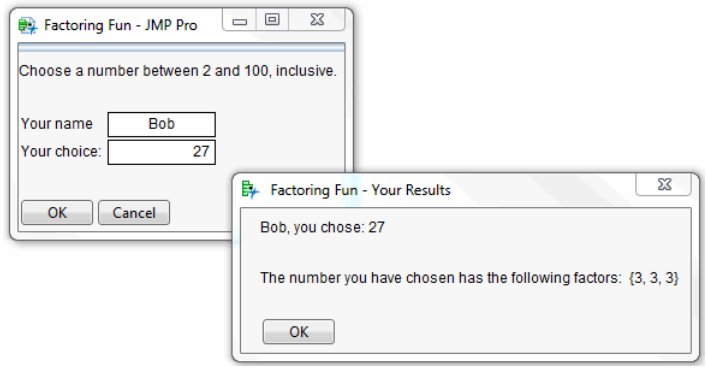
Prime Numbers.jsl asks the user to enter a number and then factors the number or confirms it as a prime number (Figure 17.5). This script is a good example of aligning several types of display boxes, concatenating text, and working with conditional functions.

```
/* How do you gather numeric input from the users, perform a
calculation on that input, and show the results in a new window?

This script demonstrates how to create an interactive program that
asks the user to enter a number and then factors the number or confirms
it as a prime number. */

// Ask the user to enter a name and number.
nw = New Window( "Factoring Fun",

V List Box(
   Text Box( "Choose a number between 2 and 100, inclusive. " ),
   Spacer Box( Size( 25, 25 ) )
    ),
    V List Box(
        Lineup Box(
            2,
            Text Box( "Your name   " ),
            uname = Text Edit Box( "< name > ", << Justify Text( Center ) ),
            Text Box( "Your choice " ),
            uprime = Number Edit Box( 2 )
        ),
        Spacer Box( Size( 25, 25 ) ),
        H List Box(
            Button Box( "OK",
         // Unload responses.
                username = uname << Get Text;
                fromUser0 = uprime << Get;

        // Test input for out of range condition.
                If( fromUser0 <= 1 | fromUser0 > 100,
           // Send message to user that input value is out of range.
                    nw2 = New Window( " Factoring Fun: Message for "
                      || username,
                        <<Modal,
                        Text Box(
                            "The number you chose, " || Char( fromUser0 ) ||
```

```
            " is not between 2 and 100, inclusive.  Please try
            again.  "
        ),
        Button Box( "OK" )

    ),
      /* Else the number is within range.
      Test for a prime number. If not prime, factor it.
      Create a vector which holds the prime numbers
      within specified range. */
    primes = [2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37,
    41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97];
    // Count the number of primes in the vector.
    p# = N Row( primes );

    isprime = 0; // Set flag.
    // Make a copy of the value for processing.
    fromuser1 = fromuser0;
    factors = {}; // Initialize list.

     /* Process the value by checking for prime then
     factoring if needed. */
    While( isPrime == 0,

        // Compare value to vector of prime numbers.
        If( Any( fromuser0 == primes ),
      // If found, place value in factor list.
            Insert Into( factors, fromUser0 );
            isPrime = 1 // Set condition to exit While loop.
            ;
        ); // End For loop.

        If( isprime == 0,
            For( q = 1, q <= p#, q++,
                If( Mod( fromuser0, primes[q] ) == 0,
                    fromUser0 = fromUser0 / primes[q];
                    Insert Into( factors, primes[q] );
                    q = p# + 1 // End if-then loop.
                    ;
                ); // End If loop.
            ); // End For loop.
        );  // End If/Then loop.

    ); // End while loop.

    cfUser0 = Char( fromUser1 );
```

```
                        nf = N Items( factors );
                        If( nf >= 2,
                            fmsg = "The number you have chosen has the following
                             factors: ",
                           fmsg = "The number you have chosen is a prime number: "
                        );
                // Show message to user about results.
                        nw3 = New Window( " Factoring Fun - Your Results",
                            <<Modal,
                            Text Box( username || ", you chose: " || cfUser0 ),
                            Spacer Box( Size( 25, 25 ) ),
                            Text Box( fmsg || " " || Char( factors ) ),
                            Spacer Box( Size( 25, 25 ) ),
                            Button Box( "OK" )
                        );
                    );
                ), // End the main OK button script.
                // Close the window and the program.
                Button Box( "Cancel", nw << Close Window )
            )
        )
    );
```

Figure 17.5 Factor Numbers Interactively



Appendix A

Compatibility Notes

Changes in Scripting in JMP 16

This appendix provides information about changes that affect compatibility in JSL scripts. Please update your scripts accordingly so that the scripts run correctly in the latest version of JMP.

Compatibility Issues in JMP 16.1

Legend Position

`Position()` has been phased out so that only `Legend Position()` is output in saved scripts. Previously saved scripts that include both the `Position()` and `Legend Position()` commands generate an error even though the script runs fine.

For example, change the following code:

```
{Legend Position( {3, [-4, -1], 5, [0, 1, 2]} ),
Position( {-4, -1, 0, 1, 2} )}
```

It should read as follows to avoid the error:

```
{Legend Position( {3, [-4, -1], 5, [0, 1, 2]} ),
Legend Position( {-4, -1, 0, 1, 2} )}
```

Compatibility Issues in JMP 16.0

Copy File(), Move File(), Rename File(), and Delete File()

In JMP 15.0, copying a file, moving a file, renaming a file, and deleting a file that didn't exist returned 0. Now, these actions throw an error.

Row Selection() and Case Sensitivity

`Match Case`, `Match Case()`, `Match Case(1)` and a missing `Match Case` do a case sensitive match in `Row Selection()`. `Match Case(0)` does a case insensitive match. Previously, `Row Selection()` matched case insensitively if `Match Case` was not supplied.

Target Level Column Property

You can use the new Target Level column property to specify the level of interest in binary logistic regression models. The default level is High. The main effects will be in the slope of the logistic probability curve, the signs of the linear model estimates, the profiler shape, and the ROC target. Be aware that some scripts that assume the target level in Fit Model or Fit Y by X logistic regression models should be checked.

Opening Multiple Data Tables

`Open()` returns a list of data tables if multiple data tables are created (for example, when importing PDF, HDF5, or Excel files). Previously, the `Open()` command returned a single table, even if multiple tables were created in the process.

Format(), Informat(), and Variations

The results of `Format()`, `Informat()` and their variations are written to the log if the format is invalid. Previously, they returned `m/d/y` when the format was invalid.

Graph Background Color Preference in Contour Plots

The Graph Background Color preference no longer affects the Contour Plot legend background color. Previously, the background color was applied to both the graph background and the legend background.

Set Auto Stretching and Get Auto Stretching for Display Boxes are Deprecated

Use `Set Stretch` and `Get Stretch` instead.

Appendix B

Efficient Scripts

Tips for Writing Scripts That Run Faster

This appendix provides ideas for writing scripts that run faster and more efficiently. For example, it is possible to control how much time is allocated to finding an optimal design. You can collapse outlines in a report that are not necessary. Or you can write static matrices, which is faster than developing them dynamically.

Contents

Using Matrices 943
Design of Experiments 944
Scripting Data Tables 944
Scripting Graphs and Display Boxes 945
Scripting Numbers, Strings, Arrays, and Lists 946
Programming Tips 946
Troubleshooting 946

Using Matrices

- The matrix concatenation operators, || and |/, add columns or rows to a matrix dynamically. However, the script runs faster if you define a matrix in a static fashion instead.

```
// faster example
t1 = Tick Seconds();
A = J( 10, 10000, . );
For( i = 1, i <= 10000, i++,
   a2 = J( 10, 1, Random Gamma( 1, 1 ) );
   // define a2 to be a 10 by 1 vector of random gamma values.
        A[0, i] = a2;
       /* populate the ith column of the matrix A to have the
         values in a2.  The first argument, 0, means "all rows". */
);
t2 = Tick Seconds();
Show( t2 - t1 );
   t2 - t1 = 0.0166666666627862;

// slower example
t3 = Tick Seconds();
B = J( 10, 0, . );
For( i = 1, i <= 10000, i++,
   a2 = J( 10, 1, Random Gamma( 1, 1 ) );
   B = B || a2;
);
t4 = Tick Seconds();
Show( t4 - t3 );
   t4 - t3 = 3.1333333333605;
```

- If you know the size of a matrix that you're going to need, it's typically better to allocate the matrix in its entirety first and then populate it. In the example below, the J() function is doing the allocation of the 100000x1 matrix all at the same time.

```
Delete Symbols( a1, a2 );
// faster example
t1 = Tick Seconds();
a1 = J( 100000, 1, Random Uniform() );
t2 = Tick Seconds();
Show( t2 - t1 );
Delete Symbols( a1 );
   t2 - t1 = 0;

// slower example
t3 = Tick Seconds();
a2 = J( 0, 0 );
```

```
For( i = 1, i <= 100000, i++,
   a2 |/= Random Uniform()
);
t4 = Tick Seconds();
Show( t4 - t3 );
   t4 - t3 = 5.05000000004657;
```

- Use matrices rather than a data table to store numbers.

Design of Experiments

In DOE, there are options such as `Number of Starts` and `Design Search Time` that enable you to control how much time is allocated to finding an optimal design.

```
// Number of Starts example
DOE(
   Custom Design,
   Add Factor( Continuous, -1, 1, "X1", 0 ),
   Add Factor( Continuous, -1, 1, "X2", 0 ),
   Add Factor( Continuous, -1, 1, "X3", 0 ),
   Number of Starts( 1000 ),
   Make Design
);
// Design Search Time example
d = DOE( Screening Design );
dt2 = Open( "$SAMPLE_DATA/Design Experiment/Weld Factors.jmp" );
d << Load Factors;
Close( dt2 );
d << {Set Random Seed( 12345 ), Screening Type( 1 ),
Suppress Cotter Designs, Design Search Time( 8 ),
Number of Column Starts( 50 ), Set Sample Size( 12 ), Make Design};
```

Scripting Data Tables

- Displaying a large data table can cause scripts to run slower. To speed things up, you can make a data table invisible, especially if you need to reference numbers in a report that will be used for future calculations or used in another platform. In this case, there isn't a need to see the current data table.

  ```
  dt = Open( "$SAMPLE_DATA/Big Class.jmp", invisible);
  ```

- Drawing objects is slow. Suppose that you have a graph with a lot of data. The platform does complicated calculations or the data table contains a formula. Adding a row can require a lot of processing. Use `Begin Data Update()` so you only redraw the graph once

instead of for each change. `Begin Data Update()` saves all update messages until `End Data Update()` is reached.

```
dt = Open( "MyData.jmp" );
dt << Distribution(
   Column( :"N=1"n, :"N=5"n), :"N=10"n )
);
Wait(1);
dt << Begin Data Update;
dt << Add Rows( 2000 );
dt << End Data Update;
```

- The use of functions might help speed things up, for example, using the `Summarize()` function rather than creating a Summary table. Maintaining a data table can be too much overhead. See "Store Summary Statistics in Global Variables".
- Data tables are great for keeping tabular data.
  - Keep data table strings under 23 characters. Longer strings have additional overhead.
  - Numeric values use less memory than character values.

Scripting Graphs and Display Boxes

- When displaying large numbers in graphs, you can set a preference to use Fast Markers.

  ```
  Preferences( Fast Marker Threshold( 25000 ) ); // 50000 is the default.
  ```

- Collapse some of the outline nodes in a report to speed up the script.

  ```
  SendToReport(
       Dispatch(
          {"Fit Life by X - Hours BY Arrhenius Celsius (Temp) Regression "},
          "Scatterplot",
          OutlineBox,
          {Close( 1 )}
       ),
  ```

- Construct your display boxes outside `New Window()` and then include them by reference in `New Window()`. This method saves the drawing until the end.

  ```
  tb = Text Box( "Select:" );
  lb = List Box( {"One", "Two", "Three"} );
  New Window( "test", tb, lb );
  ```

- The user might not need to see a report if the numbers will be used in another platform or for a future calculation. You can hide the report.

  ```
  dt= Open( "$SAMPLE_DATA/Big Class.jmp" );
  obj = Bivariate( Y( :Weight ), X( :Height ) );
  obj << Show Window( 0 ); // hide the window
  ```

```
obj << Fit Line( {Confid Curves Fit( 1 )} );
```

Submit the following script to see the report window. Notice that the confidence curves were fit even though the window was not shown.

```
obj << Show Window( 1 );
```

Scripting Numbers, Strings, Arrays, and Lists

- Use a faster algorithm. Replacing an O(n^2) algorithm with an O(n) algorithm is faster than anything else. See https://en.wikipedia.org/wiki/Big_O_notation.
- Avoid concatenating long strings. 100s of characters is OK; 1,000,000s of characters will be slow if you build the string in 1,000,000s of operations.
- Use numeric arrays (matrices, not lists).
  - Use numeric arrays with operators that operate on the entire array.
  - Avoid writing loops to manipulate numeric arrays.
- Use associative arrays for lookups.

Programming Tips

- Avoid `Break()`, `Continue()`, and `Return()` functions for speed. Using them might make it easier for you to read the script.
- Functions versus expressions:
  - User-defined functions have overhead for calling them but might make it easier for you to read the script.
  - Expressions have no overhead for calling but might make it more difficult for you to read the script.

Troubleshooting

Use the JSL Debugger to discover where the script is expending the most time. See “Debug or Profile Scripts”.

Appendix C

References

Morgan, J. (2010). "Expression Handling Functions: Part I." *JMPer Cable* 26:15–19. Accessed January 2, 2018. https://www.jmp.com/content/dam/jmp/documents/en/newsletters/jmper-cable/26_winter_2010.pdf.

Appendix D

Glossary

Terms, Concepts, and Placeholders

| In syntax summaries, | means "or" and separates possible choices. Usually choices separated by | are mutually exclusive. In other words, you have to pick one and cannot list several.

argument An argument is something specified inside the parentheses of a JSL function, message, and so forth. Big Class.jmp is the argument in `Open("Big Class.jmp")`.

You can often infer the meaning by the argument's position. For example, the values *200* and *100* in `size(200, 100)` are implicit arguments. The first value is always interpreted as the width; the second value is always interpreted as the height. See also *named argument*.

Boolean A Boolean value is a yes/no value, something that is on or off, shown or hidden, true or false, 1 or 0, yes or no. A value listed as *being a Boolean value* is one that evaluates to true or false (or missing).

col In syntax summaries, a placeholder for any reference to a data table column. For example, `Column("age")`.

command A generic description for a JSL statement that performs an action. The *Scripting Guide* prefers the more specific terms *function* and *message* when they are applicable.

current data table The current data table is the data table that `Current Data Table()` either returns or is assigned.

current row The current row for scripting is defined to be zero (no row) by default. You can set a current row with `Row()` or `For Each Row`, and so forth.

database Although the term is much more general, for JMP's purposes, the word "database" describes any external data source (such as SQL) accessed through ODBC with JSL's `Open Database` command.

Datafeed A Datafeed is a method to read real-time data continuously, such as from a laboratory measurement device connected to a serial port.

db In syntax summaries, a placeholder for any reference to a display box. For example, `report(Bivariate[1])`.

dt In syntax summaries, a placeholder for any reference to a data table. For example, `Current Data Table()` or `Data Table("Big Class.jmp")`.

eliding operator An eliding operator is one that causes arguments on either side to combine and evaluate differently than if the statement were evaluated strictly left to right. For example, `12<a<13` is a range check to test whether *a* is between 12 and 13: JMP reads the whole expression before evaluating. If < did not elide, the expression would be evaluated left to right as `(12<a)<13`. In other words, it would check whether the result of the comparison (1 or 0, for false or true) is below 13, which of course would always yield 1 for true. The << operator (for `object<<message`, which is equivalent to `Send(object, message)` is another example of an eliding operator.

function A function takes an argument or series of arguments inside parentheses after the function name. For example, the infix operator + has a function equivalent `Add()`. The statements `3 + 4` and `Add(3, 4)` are equivalent. All JSL's operators have function equivalents, but not all functions have operator equivalents. For example, `Sqrt(a)` can be represented only by the function. Also see the `Function` operator for storing a function under a name.

global variable A global variable is a name to hold values that exists for the remainder of a session. Globals can contain many types of values, including numbers, strings, lists, or references to objects. They are called globals because they can be referred to almost anywhere, not just in some specific context.

head The head of a JSL expression is the expression without any of its arguments. For example, the head of the expression `Assign( x, 100.1 )` is `Assign()`.

infix operator An infix operator takes one argument on each side, such as + in arithmetic, `3 + 4` , or the = in an assignment, `a=7`.

L-value Something that can be the destination of an assignment. In this manual, L-value describes an expression that normally returns its current value but that can alternatively receive an assignment to set its value. For example, you would ordinarily use a function such as `Row()` to get the current row number and assign it to something else. For example, `x=Row()`. However, since `Row` is an L-value, you can also place it on the left side of an assignment to set its value. For example, `Row()=10`.

list A list is a multiple-item data type entered in special brace { } notation or with the `List` function. Lists enable scripts to work with many things at once, often in the place of a single thing.

matrix A matrix is a JMP data type for a rectangular array of rows and columns of number. In JSL, matrices are entered in bracket `[ ]` notation or with the `Matrix` function.

message A message is a JSL statement that is directed to an *object*, which knows how to execute the message.

metadata In JMP data tables, metadata are data about the data, such as the source of the data, comments about each variable, scripts for working with the data, and so on.

mousedown An event generated by pressing down the mouse button. See "Handle()" and "Mousetrap()".

mouseup An event generated by releasing the mouse button. See "Handle()" and "Mousetrap()".

name A name is a reference to a JSL object. For example, when you assign the numeric value 3 to a global variable in the statement `a=3`, "a" is a name.

namespace A namespace is a collection of unique names and corresponding values. Namespaces are useful for avoiding name collisions between different scripts.

named argument A named argument is an optional argument that you select from a predetermined set and explicitly define. For example, `HexToNumber(hextext, <Base(number)>)` has a single named argument, `Base`. This is the fixed name that must directly be in the `HexToNumber` argument.

ODBC database The Microsoft standard for Open DataBase Connectivity. JSL supports access to any ODBC-enabled data source through the `Open Database` command.

obj In syntax summaries, a placeholder for any reference to an analysis platform. For example, `Bivariate[1]`.

object An object is a dynamic entity in JMP, such as a data table, a data column, a platform results window, a graph, and so forth. Most objects can receive messages telling them to act on themselves in some way.

operator Usually operator refers to a one- or two-character symbol such as + for addition or <= for less than or equal to.

POSIX POSIX is an acronym for Portable Operating System Interface and is a registered trademark of the IEEE. POSIX pathnames enable you to use one syntax for paths for any operating system, instead of having to use a different syntax for each.

postfix operator A postfix operator takes on argument on its left side (before the operator), such as `a++` for postincrement or `a--` for postdecrement.

pre-evaluated statistics Statistics that are calculated once and used as constants thereafter.

prefix operator A prefix operator takes one argument on its right side (after the operator), such as `!a` for negation.

reference A way to address a scriptable object in order to send it messages. For example,. `column("age")` or `Current Data Table()` or `Bivariate[1]`. Typically a reference is stored in a *global variable* for convenience.

row state A data element type to store any combination of the following attributes for data rows: excluded, hidden, labeled, selected, color, marker, hue, shade.

scalar A simple non-matrix numeric value.

scoping operator A scoping operator forces a name to be interpreted as a particular type of data element, for example the `:` operator in `:name` forces *`name`* to be resolved as a column; the `::` operator in `::name` forces *`name`* to be resolved as a global variable. In the rare case that you do want to share specific variables among all opened projects, use the `:::` root operator.

toggle Omitting the Boolean argument for a row state command toggles the setting. If the option is off, the message turns it on. If the option is on, the message turns it off. Sending such a command repeatedly flips back and forth between on and off. If you include the Boolean argument, the command sets an absolute on or off state, and sending the command repeatedly has no further effect. For all other messages, omitting the Boolean argument enables the option.

vector A matrix with only one column or row.

Appendix E

Technology License Notices

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.

  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

Version 16

JSL Syntax Reference

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. *JMP® 16 JSL Syntax Reference*. Cary, NC: SAS Institute Inc.

JMP® 16 JSL Syntax Reference

Copyright © 2020–2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

August 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Get the Most from JMP

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

https://www.jmp.com/getstarted

Contents

JSL Syntax Reference

1 Learn about JMP 9

Documentation and Additional Resources

Formatting Conventions in JMP Documentation 11
JMP Help 12
JMP Documentation Library 12
Additional Resources for Learning JMP 18
JMP Tutorials 18
Sample Data Tables 18
Learn about Statistical and JSL Terms 19
Learn JMP Tips and Tricks 19
JMP Tooltips 19
JMP User Community 20
Free Online Statistical Thinking Course 20
JMP New User Welcome Kit 20
Statistics Knowledge Portal 20
JMP Training 20
JMP Books by Users 21
The JMP Starter Window 21
JMP Technical Support 21

2 JSL Functions 23

Summary of Functions, Operators, and Messages

Assignment Functions 25
Character Functions 29
Character Pattern Functions 44
Comment Functions 56
Comparison Functions 57
Conditional and Logical Functions 62
Constant Functions 75
Date and Time Functions 76
Discrete Probability Functions 85
Display Functions 91
Expression Functions 127

File Functions ... 129
Financial Functions ... 151
Graphics Functions ... 156
HTTP Functions ... 176
List Functions ... 176
MATLAB Integration Functions ... 183
MATLAB JSL Function Interfaces ... 183
Matrix Functions ... 190
Numeric Functions ... 214
Optimization Functions ... 217
Probability Functions ... 221
Programming Functions ... 249
Python Integration Functions ... 268
R Integration Functions ... 274
Random Functions ... 279
Row Functions ... 290
Row State Functions ... 296
SAS Integration Functions ... 299
SQL Functions ... 318
Statistical Functions ... 320
Transcendental Functions ... 335
Trigonometric Functions ... 341
Utility Functions ... 344

3 JSL Messages ... 377
Summary of Messages for Objects and Display Boxes

Alpha Shape ... 380
Associative Arrays ... 380
Classes ... 381
Data Tables ... 383
Columns ... 414
Rows ... 423
Data Filter ... 424
Data Feed (Windows Only) ... 427
Display Boxes ... 430
All Display Boxes ... 430
Axis Boxes ... 440
Border Boxes ... 445
Data Browser Boxes ... 446
Data Filter Source Boxes ... 446

Frame Boxes . . . 446
Display 3D Boxes . . . 448
Excerpt Boxes . . . 449
Filter Col Selector . . . 449
Global Boxes . . . 449
Hier Boxes . . . 449
Matrix Boxes . . . 449
Nom Axis Boxes . . . 450
Number Col Boxes . . . 451
Number Col Edit Boxes . . . 454
Number Edit Box . . . 454
Outline Boxes . . . 455
Panel Boxes . . . 456
Plot Col Boxes . . . 456
Slider Boxes and Range Slider Boxes . . . 457
String Col Boxes . . . 458
Tab Boxes . . . 459
Table Boxes . . . 460
Text Boxes . . . 463
Tree Node and Tree Box . . . 464
Triangulation . . . 466
Windows . . . 468
Dynamic Link Libraries (DLLs) . . . 470
HTML 5 . . . 472
Web Report . . . 472
Images . . . 473
JMP Applications . . . 476
JMP App . . . 477
JMP App Module . . . 478
JMP App Module Instance . . . 479
MATLAB . . . 479
Namespaces . . . 483
Platforms . . . 485
Bubble Plot . . . 490
DOE . . . 491
Partition . . . 491
Response Screening . . . 492
Tabulate . . . 492
Python Integration Messages . . . 493
R Integration Messages . . . 497

SAS Integration Messages 500
Metadata Server Objects 500
SAS Server Objects 501
Stored Processes 510
SAS Results 516
Schedule 519
Segments 520
Sockets 520
SQL 523
Other Objects 527
Zip Archives 527
Journals 528

A SQL Functions Available for JMP Queries 529

Numeric SQL Functions 531
Date-Time SQL Functions 532
String SQL Functions 535
System SQL Functions 536
Aggregate SQL Functions 536

B References 539

C Technology License Notices 541

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation . . . 11
JMP Help . . . 12
JMP Documentation Library . . . 12
Additional Resources for Learning JMP . . . 18
JMP Tutorials . . . 18
Sample Data Tables . . . 18
Learn about Statistical and JSL Terms . . . 19
Learn JMP Tips and Tricks . . . 19
JMP Tooltips . . . 19
JMP User Community . . . 20
Free Online Statistical Thinking Course . . . 20
JMP New User Welcome Kit . . . 20
Statistics Knowledge Portal . . . 20
JMP Training . . . 20
JMP Books by Users . . . 21
The JMP Starter Window . . . 21
JMP Technical Support . . . 21

Formatting Conventions in JMP Documentation

These conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit https://www.jmp.com/software/pro.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from https://community.jmp.com.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Model Screening • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Legacy Control Charts • Pareto Plot • Diagram • Manage Spec Limits • OC Curves

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "JMP Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "JMP Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "JMP New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\16\Samples\Data

On macOS: \Library\Application Support\JMP\16\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit https://jmp.com/tools.

Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > Initial JMP Starter Window**.

JMP Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at https://www.jmp.com/support, including the technical support phone number.

Chapter 2

JSL Functions

Summary of Functions, Operators, and Messages

This topic provides abbreviated descriptions for many of JMP's functions, operators, and general object messages. For complete information about functions, see the JMP Scripting Index. In JMP, select **Help > Scripting Index**.

For information about platform messages, see the *Scripting Guide*.

Contents

Assignment Functions . . . 25
Character Functions . . . 28
Character Pattern Functions . . . 43
Comment Functions . . . 53
Comparison Functions . . . 54
Conditional and Logical Functions . . . 59
Constant Functions . . . 71
Date and Time Functions . . . 72
Discrete Probability Functions . . . 80
Display Functions . . . 86
Expression Functions . . . 120
File Functions . . . 122
Financial Functions . . . 144
Graphics Functions . . . 149
HTTP Functions . . . 168
List Functions . . . 169
MATLAB Integration Functions . . . 175
    MATLAB JSL Function Interfaces . . . 175
Matrix Functions . . . 182
Numeric Functions . . . 206
Optimization Functions . . . 210
Probability Functions . . . 213
Programming Functions . . . 243
Python Integration Functions . . . 260
R Integration Functions . . . 266
Random Functions . . . 271
Row Functions . . . 282
Row State Functions . . . 288
SAS Integration Functions . . . 291
SQL Functions . . . 309
Statistical Functions . . . 312
Transcendental Functions . . . 326
Trigonometric Functions . . . 332
Utility Functions . . . 335

Assignment Functions

JSL also provides operators for in-place arithmetic, or *assignment operators*. These operations are all done in place, meaning that the result of the operation is assigned to the first argument. The most basic assignment operator is the = operator (or the equivalent function `Assign`). For example, if *a* is 3 and you do `a+=4`, then *a* becomes 7.

The first argument to an assignment function must be something capable of being assigned (an *L-value*). You cannot do something like `3+=4`, because 3 is just a value and cannot be reassigned. However, you can do something like `a+=4`, because *a* is a variable whose value you can set.

`Add To(`*a*`, `*b*`)`

a+=*b*

Description

Adds *a* and *b* and places the sum into *a*.

Returns

The sum.

Arguments

a Must be a variable.

b Can be a variable, a list, a number, or a matrix.

Notes

The first argument must be a variable, because its value must be able to accept a value change. A number as the first argument produces an error.

For `Add To()`: Only two arguments are permitted. If one or no argument is specified, `Add To()` returns a missing value. Any arguments after the first two are ignored.

For `a+=b`: More than two arguments can be strung together. JMP evaluates pairs from right to left, and each sum is placed in the left-hand variable. All arguments except the last must be a variable.

Example

```
a+=b+=c
```

JMP adds *b* and *c* and places the sum into *b*. Then JMP adds *a* and *b* and places the sum into *a*.

`Assign(a, b)`

a=b

Description

Places the value of *b* into *a*.

Returns

The new value of *a*.

Arguments

a Must be a variable.

b Can be a variable, number, or matrix.

Notes

a must be a variable, because it must be able to accept a value change. A number as the first argument produces an error. If *b* is some sort of expression, it's evaluated first and the result is placed into *a*.

`Divide To(a, b)`

a/=b

Description

Divides *a* by *b* and places the result into *a*.

Returns

The quotient.

Arguments

a Must be a variable.

b Can be a variable, number, or matrix.

`Multiply To(a, b)`

a=b*

Description

Multiplies *a* and *b* and places the product into *a*.

Returns

The product.

Arguments

a Must be a variable.

b Can be a variable, number, or matrix.

Notes

The first argument must be a variable, because its value must be able to accept a value change. A number as the first argument produces an error.

For `Multiply To()`: Only two arguments are permitted. If one or no argument is specified, `Multiply To()` returns a missing value. Any arguments after the first two are ignored.

For *a*`*=`*b*: More than two arguments can be strung together. JMP evaluates pairs from right to left, and each product is placed in the left-hand variable. All arguments except the last must be a variable.

Example

```
a*=b*=c
```

JMP multiplies *b* and *c* and places the product into *b*. Then JMP multiplies *a* and *b* and places the product into *a*.

`PostDecrement(`*a*`)`

a`--`

Description

Post-decrement. Subtracts 1 from *a* and places the difference into *a*.

Returns

a-1

Argument

a Must be a variable.

Notes

If *a*`--` or `Post Decrement(`*a*`)` is inside another expression, the expression is evaluated first, and then the decrement operation is performed. This expression is mostly used for loop control.

`Post Increment(`*a*`)`

a`++`

Description

Post-increment. Adds 1 to *a* and places the sum into *a*.

Returns

a+1

Argument

a Must be a variable.

Notes

If `a++` or `Post Increment(a)` is inside another expression, the expression is evaluated first, and then the increment operation is performed. Mostly used for loop control.

Subtract To(*a*, *b*)

a-=*b*

Description

Subtracts *b* from *a* and places the difference into *a*.

Returns

The difference.

Arguments

a Must be a variable.

b Can be a variable, number, or matrix.

Notes

The first argument must be a variable, because its value must be able to accept a value change. A number as the first argument produces an error.

For SubtractTo(): Only two arguments are permitted. If fewer than two or more than two arguments is specified, SubtractTo() returns a missing value.

For *a*-=*b*: More than two arguments can be strung together. JMP evaluates pairs from right to left, and each difference is placed in the left-hand variable. All arguments except the last must be a variable.

Example

```
a-=b-=c
```

JMP subtracts *c* from *b* and places the difference into *b*. Then JMP subtracts *b* from *a* and places the difference into *a*.

Character Functions

Most character functions take character arguments and quoted return character strings, although some take numeric arguments or return numeric data. Arguments that are literal character strings must be enclosed in quotation marks.

BLOB To Char(*blob*, <Encoding="*enc*">)

Description

Reinterpret binary data as a quoted Unicode string.

Returns

A quoted string.

Required Argument

blob A binary large object.

Optional Argument

Encoding A quoted string that specifies an encoding. The default encoding for the character string is "utf-8". "utf-16le", "utf-16be", "us-ascii", "iso-8859-1", "ascii~hex", "shift-jis", and "euc-jp" are also supported.

Notes

The optional argument "ascii~hex" is intended to make conversions of blobs containing normal ASCII data simpler when the data might contain CR, LF, or TAB characters (for example) and those characters do not need any special attention.

BLOB To Matrix(*blob*, *type*, *bytes*, *endian*, <*nCols*>)

Description

Creates a matrix by converting each byte in the blob to numbers.

Returns

A matrix that represents the blob.

Required Arguments

blob A blob or reference to a blob.

type A quoted string that contains the named type of number. The options are "int", "uint", or "float".

bytes Byte size of the data in the blob. Options are 1, 2, 4, or 8.

endian The quoted endian-ness of your system: "Big" (the first byte is most significant), "Little" (the first byte is the least significant), or "Native" (the machine's native format).

Optional Argument

<*nCols*> The number of columns in the matrix. The default value is 1.

Char(*x*, <*width*>, <*decimal*>, < <<Use Locale(*Boolean*)>)

Description

Converts an expression or numeric value into a quoted character string.

Returns

A quoted string.

Required Argument

x an expression or a numeric value. An expression must be quoted with Expr(). Otherwise, its evaluated value is converted to a quoted string.

Optional Arguments

width A number that sets the maximum number of characters in the quoted string.

decimal A number that sets the maximum number of places after the decimal that is included in the quoted string.

Use Locale(*Boolean*) Preserves locale-specific numeric formatting.

Example

```
Char( Pi(), 10, 4)
   "3.1416"

Char( Pi(), 3, 4)
   "3.1"
```

Notes

The *width* argument overrides the *decimal* argument.

Char To BLOB(*string*, <encoding="*enc*">)

Description

Converts a quoted string of characters into a binary (blob).

Returns

A binary object.

Required Argument

string A quoted string or a reference to a string.

Optional Argument

encoding A quoted string that specifies an encoding. The default encoding for the blob is "utf-8". "utf-16le", "utf-16be", "us-ascii", "iso-8859-1", "ascii~hex", "shift-jis", and "euc-jp" are also supported.

Notes

Converting BLOBs into printable format escapes \ (in addition to ~ " ! and characters outside of the printable ASCII range) into hex notation (~5C for the backslash character).

```
x = Char To BLOB( "abc\def\!n" );
y = BLOB To Char( x, encoding = "ASCII~HEX" );
If(
   y == "abc~5Cdef~0A", "JMP 12.2 and later behavior",
   y == "abc\def~0A", "Pre-JMP 12.2 behavior"
);
   "JMP 12.2 and later behavior" // output
```

Char To Hex(*value*, <*integer*|encoding="*enc*">)

Hex(*value*, <*integer*|encoding="*enc*"|Base(*number*)|Pad To(*number*)>)

Description

Returns the hexadecimal (or other base number system) text corresponding to the given value and encoding, which can be a number a quoted string or a blob. If the value is a number, IEEE 754 64-bit encoding is used unless one of the optional arguments, *integer*, or *Base*, is provided.

Required Argument

value Any number, quoted string, or blob.

Optional Arguments

`integer` A quoted switch that causes the value to be interpreted as an integer.

`encoding` A quoted string that specifies an encoding. The default encoding is `"utf-8"`. `"utf-16le"`, `"utf-16be"`, `"us-ascii"`, `"iso-8859-1"`, `"ascii~hex"`, `"shift-jis"`, and `"euc-jp"` are also supported.

`Base(`*`number`*`)` An integer value between 2 and 36 inclusive. If base is specified, the function returns the text corresponding to the specified number in that base number system instead of hexadecimal.

`Pad To(`*`number`*`)` A value to specify the padded width of the hex output.

`Collapse Whitespace(`*`text`*`)`

Description

Trims leading and trailing whitespace and replaces interior whitespace with a single space. That is, if more than one white space character is present, the `Collapse Whitespace` function replaces the two spaces with one space.

Returns

A quoted string.

Required Argument

`text` A quoted string.

`Concat(`*`a`*`, `*`b`*`)`

`Concat(`*`A`*`, `*`B`*`)`

a||*b*

A||*B*

Description

For quoted strings: Appends the string *b* to the string *a*. Neither argument is changed.

For lists: Appends the list *b* to the list *a*. Neither argument is changed.

For matrices: Horizontal concatenation of two matrices, A and B.

Returns

For quoted strings: A quoted string composed of the string *a* directly followed by the string *b*.

For lists: A list composed of the list *a* directly followed by the list *b*.

For matrices: A matrix.

Arguments

Two or more quoted strings, quoted string variables, lists, or matrices.

Example

```
a = "Hello"; b = " "; c = "World"; a || b || c;
```

```
   "Hello World"
d = {"apples", "bananas"}; e = {"peaches", "pears"}; Concat( d, e );
   {"apples", "bananas", "peaches", "pears"}
A = [1 2 3]; B = [4 5 6]; Concat( A, B );
   [1 2 3 4 5 6]
```

Notes

More than two arguments can be strung together. Each additional quoted string is appended to the end, in left to right order. Each additional matrix is appended in left to right order.

Concat Items

See "Concat Items({string1, string2, ...}, <delimiter>})".

Concat To(*a*, *b*)

Concat To(*a*, *b*)

a||=*b*

A||=*B*

Description

For quoted strings: Appends the string *b* to the string *a* and places the new concatenated string into *a*.

For matrices: Appends the matrix *b* to the matrix *a* and places the new concatenated matrix into *a*.

For lists: Appends the list *b* to the list and places the new concatenated list into *a*.

Returns

For quoted strings: A string composed of the string *a* directly followed by the string *b*.

For matrices: A matrix.

For lists: A list composed of the list *a* directly followed by the list *b*.

Arguments

Two or more quoted strings, quoted string variables, matrices, or lists. The first variable must be a variable whose value can be changed.

Notes

More than two arguments can be strung together. Each additional quoted string, matrix, or list is appended to the end, in left to right order.

Example

```
a = "Hello"; b = " "; c = "World"; Concat To( a, b, c ); Show( a );
   a = "Hello World"
A = [1 2 3]; B = [4 5 6]; Concat To( A, B ); Show( A );
   A = [1 2 3 4 5 6];
```

```
d = {"apples", "bananas"}; e = {"peaches", "pears"}; Concat to(d,e); Show( d
);
d = {"apples", "bananas", "peaches", "pears"};
```

Contains(*whole*, *part*, <*start*>)

Description

Determines whether *part* is contained within *whole*.

Returns

If *part* is found: For lists, quoted strings, and namespaces, the numeric position where the first occurrence of *part* is located. For associative arrays, 1.

If *part* is not found, 0 is returned in all cases.

Required Arguments

whole A quoted string, list, namespace, or associative array.

part For a quoted string or namespace, a string that can be part of the string *whole*. For a list, an item that can be an item in the list *whole*. For an associative array, a key that can be one of the keys in the map *whole*.

Optional Argument

start A numeric argument that specifies a starting point within *whole*. If *start* is negative, contains searches *whole* for *part* backwards, beginning with the position specified by the length of *whole* – *start*. Note that *start* is meaningless for associative arrays and is ignored.

Example

```
nameList={"Katie", "Louise", "Jane", "Jaclyn"};
r = Contains(nameList, "Katie");
```

The example returns a 1 because “Katie” is the first item in the list.

Contains Item(*x*, <*item* | {*list*} | *pattern*>, <*delimiter*>)

Description

Identifies multiple responses by searching for the specified item, list, pattern, or delimiter. The function can be used on columns with the Multiple Response modeling type or column property.

Returns

Returns a Boolean that indicates whether the word (*item*), one of a list of words (*list*), or pattern (*pattern*) matches one of the words in the text represented by *x*. Words are delimited by the characters in the optional quoted delimiter (*delimiter*) string. A comma is the default delimiter. Blanks are trimmed from the ends of each extracted word from the input text string (*x*).

Example

The following example searches for *"pots"* followed by a comma and then outputs the result.

```
x = "Franklin Garden Supply is a leading online store featuring garden decor,
   statues, pots, shovels, benches, and much more.";
b = Contains Item( x, "pots", "," );
If( b,
   Write( "The specified items were found."  ), Write( "No match." )
);
   The specified items were found.
```

Ends With(*string*, *substring*)

Description

Determines whether the quoted `substring` appears at the end of `string`.

Returns

1 if `quoted string` ends with a quoted `substring`, otherwise 0.

Required Arguments

`string` A quoted string or a quoted string variable. Can also be a list.

`substring` A quoted string or a quoted string variable. Can also be a list.

Equivalent Expression

`Right(`*`string`*`, Length(`*`substring`*`)) == substring`

Hex(*value*, <*integer*|encoding="*enc*"|Base(*number*)|Pad To(*number*)>)

See "Char To Hex(value, <integer|encoding="enc">)".

Hex To BLOB(*string*)

Description

Converts the quoted hexadecimal *`string`* (including whitespace characters) to a blob (binary large object).

Example

```
Hex To BLOB( "4A4D50" );
   Char To BLOB("JMP", "ascii~hex")
```

Hex To Char(*string*, <*encoding*>)

Description

Converts the quoted hexadecimal *`string`* to its character equivalent.

Example

Hex To Char("30") results in "0".

Notes

The default quoted encoding for character string is "utf-8". "utf-16le", "utf-16be", "us-ascii", "iso-8859-1", "ascii~hex", "shift-jis", and "euc-jp" are also supported.

`Hex To Number(string, <Base(number)>)`

Description

Returns the number corresponding to the hexadecimal (or other base number system) text.

Required Argument

string A quoted hexadecimal string.

Optional Argument

Base(*number*) An integer between 2 and 36 inclusive. If base is specified, the text is treated as a quoted string representing the number in that base.

Example

```
Hex To Number( "80" );
  128
```

Notes

- 16-digit hexadecimal numbers are converted as IEEE 754 64-bit floating point numbers. Otherwise, the input is treated as a hexadecimal integer.
- Whitespace between bytes (or pairs of digits) and in the middle of bytes is permitted (for example, `FF 1919` and `F F1919`).

`Insert`

See "Insert(source, item, <position>)".

`Insert Into`

See "Insert Into(source, item, <position>)".

`Item(n|[first last], string, <delimiter>, <Unmatched(result string)>, <Include Boundary Delimiters(Boolean)>)`

Description

Returns the *n*th item or the span from the first to last item of the quoted *string* according to the quoted string *delimiters* given. If you include a third argument, any and all characters in that argument are taken to be delimiters.

Required Arguments

n The position of the word being extracted.

[*first last*] A matrix that defines the beginning and end word range to return.

string The quoted string that is evaluated.

Optional Arguments

delimiter The character used as a boundary. If *delimiter* is absent, an ASCII space is used. If *delimiter* is the quoted empty string, each character is treated as a separate word.

Unmatched(*result string*) The quoted string to print if no match is found.

Include Boundary Delimiters(*Boolean*) Determines how delimiters on the front and back of string are treated. The default value is false, which means that the delimiters on these boundaries are ignored. If the value is true, delimiters on these boundaries produce an empty element (similar to consecutive delimiters within a string).

Example

Consecutive delimiters are treated as though they have a word between them. In this example, the delimiters are a comma and a space.

```
Item( 4,"the  quick, brown fox", ", " ); // quick is preceded by two spaces
```

The expression is processed as follows:

the<delim[space]><word2><delim[space]>quick<delim[comma]><word 4><delim[space]>brown<delim[space]>fox

Because word4 is empty, this expression returns a quoted empty string.

Item() is the same as Word() except that Item() treats each delimiter character as a separate delimiter, and Word() treats several adjacent delimiters as a single delimiter.

```
Word( 4,"the  quick, brown fox", ", " ); // quick is preceded by two spaces
```

This expression is processed as follows:

the<delim[2 spaces]>quick<delim[comma + space]>brown<delim[space]>fox

It returns "fox".

Left(*string*, *n*, <*filler*>)

Left({*list*}, *n*, <*filler*>)

Description

Returns a truncated or padded version of the original quoted *string* or *list*. The result contains the left *n* characters or list items, padded with any *filler* on the right if the length of *string* is less than *n*.

Length(*string*)

Description

Returns the length of the given quoted string (in characters), list (in items), associative array (in number of keys), BLOB (in bytes), matrix (in elements), namespace (in number of functions and variables), or class (in number of methods, functions, and variables).

Lowercase(*string*)

Description

Converts any uppercase character found in quoted *string* to the equivalent lowercase character.

Matrix to BLOB(*matrix*, *type*, *bytesEach*, *endian(value)*)

Description

Makes a BLOB from a matrix by converting the matrix elements to 1-byte, 2-byte, or 4-byte signed or unsigned integers or 4-byte or 8-byte floating point numbers.

Required Arguments

matrix The matrix.

type The quoted type of BLOB: `int`, `uint`, or `float`.

bytesEach The number of bytes in each int, uint, or float. Integers can be 1, 2, or 4 bytes each, and floats can be 4 or 8 bytes each.

endian(value) The quoted endian-ness of your system: `"Big"` (the first byte is most significant), `"Little"` (the first byte is the least significant), or `"Native"` (the machine's native format).

Munger(*string*, *start position*, *find|length*, *<replacement string>*)

Description

Computes new quoted character strings from the quoted *string* by inserting or deleting characters. It can also produce substrings, calculate indexes, and perform other tasks depending on how you specify its arguments.

Required Arguments

start position A numeric expression that specifies the starting position to search in the quoted string. If the *start position* is greater than the position of the first instance of the find argument, the first instance is disregarded. If the *start position* is greater than the search string's length, `Munger()` uses the string's length as the *start position*.

find|length Specifies the string or number of characters to find.

Optional Argument

replacement string The quoted replacement string. If *replacement string* is omitted, a substring between start position, and position and length, is returned.

Num(*string*)

Description

Converts a quoted character string into a number.

`Regex(source, pattern, (<replacement string>, <format, "GLOBALREPLACE", "IGNORECASE">>)`

Description

Searches for the quoted *pattern* within the quoted *source* string.

Returns

The matched text as a quoted string or numeric missing if there was no match.

Required Arguments

source A quoted string.

pattern A quoted regular expression.

replacement string The replacement string.

Optional Arguments

format A backreference to the capturing group. The default is \0, which is the entire matched quoted string. \n returns the n^{th} match.

"IGNORECASE" The search is case sensitive, unless you specify "IGNORECASE".

"GLOBALREPLACE" Applies the regular expression to the quoted *source* string repeatedly until all matches are found.

`Remove`

See "Remove(source, position, <n>)".

`Remove From`

See "Remove From(source, position, <n>)".

`Repeat(source, a)`

`Repeat(matrix, a, b)`

Description

Returns a copy of *source* concatenated with itself *a* times. Or returns a matrix composed of *a* row repeats and *b* column repeats. The *source* can be text, a matrix, or a list.

`Reverse`

See "Reverse(source)".

`Reverse Into`

See "Reverse Into(source)".

Right(*string*, *n*, <*filler*>)

Right({*list*}, *n*, <*filler*>)

Description

Returns a truncated or padded version of the original quoted *string* or *list*. The result contains the right *n* characters or list items, padded with any optional *filler* on the left if the length of *string* is less than *n*.

Shift

See "Shift(source, <n>)".

Shift Into

See "Shift Into(source, <n>)".

Starts With(*string*, *substring*)

Description

Determines whether the quoted `substring` appears at the start of the quoted *string*.

Returns

1 if *string* starts with *substring*, otherwise 0.

Arguments

string A quoted string or a reference to one. Can also be a list.

substring A quoted string or a reference to one. Can also be a list.

Equivalent Expression

`Left(string, Length("substring")) = = "substring"`

Substitute

See "Substitute(string, "substring", "replacementString", ...)".

Substitute Into

See "Substitute Into(string, substring, replacementString, ...)".

Substr(*string*, *start*, *length*)

Description

Extracts the characters that are the portion of the first argument beginning at the position given by the second argument and ending based on the number of characters specified in the third argument. The first argument can be a character column or value, or an

expression evaluating to same. The starting argument and the length argument can be numbers or expressions that evaluate to numbers.

Example

This example extracts the first name:

```
Substr( "Katie Layman", 1, 5 );
```

The function starts at position 1, reads through position 5, and ignores the remaining characters, which yields "Katie."

`Text Score(`*`text column`*`, `*`text-to-number`*`, <`*`weighting`*`>, <{`*`support vectors`*`}>, <`*`text explorer setup`*`>)`

Description

Used to create scoring formulas in Text Explorer. Not supported for use with the Stem for Combining option.

Returns

Returns a vector of scores.

Required Arguments

`text column` The data table column.

`text-to-number` An associative array that maps lowercase words to numbers.

Optional Arguments

`weighting` The quoted `"Count"`, `"Binary"`, `"Ternary"`, `"LogCount"`, `"LCA"`, or an array of inverse document frequency weights for TFLogIDF. `"Count"` is the default value. The default value is `"Count"`.

`support vectors` A list of vectors that are used in the text scoring. The number and length of the vectors depends on the *`weighting`* argument.

`text explorer setup` An expression that contains a block of Text Explorer setup information.

`Titlecase(`*`string`*`)`

Description

Converts the quoted *`string`* to title case; each word in the string has an initial uppercase character and subsequent lowercase characters.

Returns

A quoted string.

Argument

`string` A quoted string.

Example

The following function capitalizes the name:

```
Titlecase( "veronica layman ")
```

```
"Veronica Layman"
```

`Trim(string,<"Left"|"Right"|"Both">)`

`Trim Whitespace(string,<"Left"|"Right"|"Both">)`

Description

Removes leading and trailing whitespace from the specified string.

Returns

A quoted string.

Required Argument

string A quoted string.

Optional Argument

`"Left"|"Right"|"Both"` A quoted string that specifies whether whitespace is removed from the left, the right, or both ends of the string. If omitted, whitespace is removed from both ends.

Example

For example, the following command returns `"John"`:

```
Trim( "  John    ", Both )
   "John"
```

`Uppercase(string)`

Description

Converts any lowercase character found in the quoted *string* to the equivalent uppercase character.

`Word(n|[first last], string, <delimiter>, <Unmatched(result string)>)`

Description

Returns the *n*th item of the string, where words are substrings separated by any number of any characters in the *delimiter* argument.

Required Arguments

n The position of the word being extracted.

`[first last]` A matrix that defines the beginning and end word range to return.

string The quoted string that is evaluated.

Optional Arguments

delimiter The character used as a boundary. If *delimiter* is absent, an ASCII space is used. If *delimiter* is the empty quoted string, each character is treated as a separate word.

`Unmatched(result string)` The quoted string to print if no match is found.

Examples

This example returns the last name:

```
Word( 2, "Katie Layman" );
   "Layman"
```

See Also

See "Item(n|[first last], string, <delimiter>, <Unmatched(result string)>, <Include Boundary Delimiters(Boolean)>)" for examples of how `Word()` differs from `Item()`.

Words

See "Words(string, <delimiter>)".

XPath Query(*xml*, *xpath_expression*)

Description

Runs an XPath expression on an XML document.

Returns

A list.

Required Arguments

xml A valid XML document.

xpath_expression A quoted XPath 1.0 expression.

Example

Suppose that you created a report of test results in JMP and exported important details to an XML document. The test results are enclosed in `<result>` tags.

The following example stores the XML document in a variable. The `XPath Query` expression parses the XML to find the text nodes inside the `<result>` tags. The results are returned in a list.

```
rpt =
"\[<?xml version="1.0" encoding="utf-8"?>
<JMP><report><title>Production Report</title>
<result>November 21st: Pass</result>
<result>November 22nd: Fail</result>
<note>Tests ran at 3:00 a.m.</note></report>
</JMP> ]\";
results = XPath Query( rpt, "//result/text()" );
   {"November 21st: Pass", "November 22nd: Fail"}
```

Character Pattern Functions

Pat Abort()

Description

Constructs a quoted pattern that immediately stops the pattern match. The matcher does not back up and retry any alternatives. Conditional assignments are *not* made. Immediate assignments that were already made are kept.

Returns

0 when a match is stopped.

Argument

none

Pat Altern(*pattern1*, <*pattern2*, ...>)

Description

Constructs a quoted pattern that matches any one of the pattern arguments.

Returns

A quoted pattern.

Argument

One or more patterns.

Pat Any(*string*)

Description

Constructs a quoted pattern that matches a single character in the argument.

Returns

A quoted pattern.

Argument

pattern A quoted string.

Pat Arb()

Description

Constructs a quoted pattern that matches an arbitrary quoted string. It initially matches the null quoted string. It then matches one additional character each time the pattern matcher backs into it.

Returns

A quoted pattern.

Argument

none

Example

```
p = "the beginning" + Pat Arb() >? stuffInTheMiddle + "the end";
Pat Match( "in the beginning of the story, and not near the end, there are
   three bears", p );
Show( stuffInTheMiddle );
   stuffInTheMiddle = " of the story, and not near "
```

Pat Arb No(*pattern*)

Description

Constructs a quoted pattern that matches zero or more copies of *pattern*.

Returns

A quoted pattern.

Argument

pattern A quoted pattern to match against.

Example

```
adjectives = "large" | "medium" | "small" | "warm" | "cold" | "hot" | "sweet";
rc = Pat Match( "I would like a medium hot, sweet tea please",
                Pat Arbno( adjectives | Pat Any(", ") ) >> adj +
                ("tea" | "coffee" | "milk") );
Show( rc, adj );
   rc = 1;
   adj = " medium hot, sweet ";
```

Pat At(*varName*)

Description

Constructs a quoted pattern that matches the null quoted string and stores the current position in the source string into the specified JSL variable (*varName*). The assignment is immediate, and the variable can be used with `expr()` to affect the remainder of the match.

Returns

A quoted pattern.

Argument

varName The name of a variable to store the result in.

Example

```
p = ":" + Pat At( listStart ) + Expr(
   If( listStart == 1,
      Pat Immediate( Pat Len( 3 ), early ),
      Pat Immediate( Pat Len( 2 ), late )
   )
);
```

```
early = "";
late = "";
Pat Match( ":123456789", p );
Show( early, late );
early = "";
late = "";
Pat Match( "   :123456789", p );
Show( early, late );
```

First this is produced:

```
early = "123"
late = ""
```

and later this:

```
early = ""
late = "12"
```

Pat Break(*string*)

Description

Constructs a quoted pattern that matches zero or more characters that are not in its argument; it stops or breaks on a character in its argument. It fails if a character in its argument is not found (in particular, it fails to match if it finds the end of the quoted source string without finding a break character).

Returns

A quoted pattern.

Argument

string A quoted string.

Pat Concat(*pattern1*, *pattern2* <*pattern3*, ...>)

Pattern1 + *Pattern2* + ...

Description

Constructs a quoted pattern that matches each pattern argument in turn.

Returns

A quoted pattern.

Argument

Two or more quoted patterns.

Pat Conditional(*pattern*, *varName*)

Description

Saves the result of the quoted pattern match, if it succeeds, to a variable named as the second argument (*varName*) after the match is finished.

Returns

A quoted pattern.

Arguments

pattern A quoted pattern to match against.

varName The name of a variable to store the result in.

Example

```
type = "undefined";
rc = Pat Match(
   "green apples",
   Pat Conditional( "red" | "green", type ) + " apples"
);
Show( rc, type );
   rc = 1;
   type = "green";
```

Pat Fail()

Description

Constructs a quoted pattern that fails whenever the matcher attempts to move forward through it. The matcher backs up and tries different alternatives. If and when there are no alternatives left, the match fails and `Pat Match` returns 0.

Returns

0 when a match fails.

Argument

none

Pat Fence()

Description

Constructs a pattern that succeeds and matches the quoted null string when the matcher moves forward through it, but fails when the matcher tries to back up through it. It is a one-way trap door that can be used to optimize some matches.

Returns

1 when the match succeeds, 0 otherwise.

Argument

none

Pat Immediate(*pattern*, *varName*)

Description

Saves the result of the pattern match to a variable named as the second argument (*varName*) immediately.

Returns

A quoted pattern.

Arguments

pattern A quoted pattern to match against.

varName The name of a variable to store the result in.

Example

```
type = "undefined";
rc = Pat Match(
   "green apples",
   ("red" | "green") >> type + " pears"
);
Show( rc, type );
   rc = 0
   type = "green"
```

Even though the match failed, the immediate assignment was made.

Pat Len(*int*)

Description

Constructs a quoted pattern that matches *n* characters.

Returns

A quoted pattern.

Argument

int An integer that specifies the number of characters.

Pat Look Ahead(*pattern*, *Boolean*)

Description

A zero-width pattern match after the current position.

Arguments

pattern The quoted pattern.

Boolean 0 (the default) indicates a match. 1 indicates a negative match or non-match.

Pat Look Behind(*pattern*, *Boolean*)

Description

A zero-width quoted pattern match before the current position.

Arguments

pattern The quoted pattern.

Boolean 0 (the default) indicates a match. 1 indicates a negative match or non-match.

Pat Match(*source text*, *pattern*, <*replacement string*>, <"NULL">, <"ANCHOR">, <"MATCHCASE">, <"FULLSCAN">)

Description

Pat Match executes the quoted *pattern* against the *source text*. The pattern must be constructed first, either inline or by assigning it to a JSL variable elsewhere.

Returns

1 if the pattern is found, 0 otherwise.

Required Arguments

source text A quoted string or quoted string variable that contains the text to be searched.

pattern A quoted pattern or pattern variable that contains the text to be searched for.

Optional Arguments

replacement string A quoted string that defines text to replace the pattern in the source text.

"NULL" A placeholder for the third argument if ANCHOR, MATCHCASE, or FULLSCAN are necessary *and* there is no replacement text.

"ANCHOR" Starts the pattern match at the beginning of the quoted string. The following match fails because the pattern, "cream", is not found at the beginning of the string:

```
Pat Match( "coffee with cream and sugar", "cream", NULL, ANCHOR );
```

"MATCHCASE" Optional command to consider capitalization in the match. By default, Pat Match() is case insensitive.

"FULLSCAN" Optional command to force Pat Match to try all alternatives, which uses more memory as the match expands. By default, Pat Match() does not use FULLSCAN, and makes some assumptions that allow the recursion to stop and the match to succeed.

Pat Not Any(*string*)

Description

Constructs a pattern that matches a single character that is not in the argument.

Returns

A quoted pattern.

Argument

string A quoted string.

Pat Pos(*int*)

Description

Constructs patterns that match the quoted null string if the current position is *int* from the left end of the string, and fail otherwise.

Returns

A quoted pattern.

Argument

int An integer that specifies the position in a quoted string.

Pat R Pos(*int*)

Description

Constructs patterns that match the quoted null string if the current position is *int* from the right end of the string, and fails otherwise.

Returns

A quoted pattern.

Argument

int An integer that specifies the position in a quoted string.

Pat R Tab(*int*)

Description

Constructs a quoted pattern that matches up to position *n* from the end of the quoted string. It can match 0 or more characters. It fails if it would have to move backwards or beyond the end of the string.

Returns

A quoted pattern.

Argument

int An integer that specifies a position in a quoted string.

Pat Regex(*string*)

Description

Constructs a quoted pattern that matches the regular expression in the quoted *string* argument.

Returns

A quoted pattern.

Argument

string A quoted string.

Pat Rem()

Description

Constructs a quoted pattern that matches the remainder of the quoted string. It is equivalent to `Pat R Tab(0)`.

Returns

A quoted pattern.

Argument

none

Pat Repeat(*pattern*, *minimum*, *maximum*, <"GREEDY"|"RELUCTANT">)

Description

Matches the quoted *pattern* between *minimum* and *maximum* times.

Returns

A quoted pattern.

Required Arguments

pattern A pattern to match against.

minimum An integer that must be smaller than maximum.

maximum An integer that must be greater than minimum.

Optional Argument

"GREEDY"|"RELUCTANT" If GREEDY is specified, it tries the maximum first and works back to the minimum. If RELUCTANT is specified, it tries the minimum first and works up to the maximum.

Notes

- `Pat Arbno(p)` is the same as `Pat Repeat(p, 0, infinity, RELUCTANT)`
- `Pat Repeat(p)` is the same as `Pat Repeat(p, 1, infinity, GREEDY)`
- `Pat Repeat(p, n)` is the same as `Pat Repeat(p, n, infinity, GREEDY)`
- `Pat Repeat(p, n, m)` is the same as `Pat Repeat(p, n, m, GREEDY)`

Pat Span(*string*)

Description

Constructs a pattern that matches one or more (not zero) occurrences of characters in its argument. It is greedy; it always matches the longest possible quoted string. It fails rather than matching zero characters.

Returns

A quoted pattern.

Argument

string A quoted string.

Pat String(*string*)

Description

Constructs a pattern that matches its quoted *string* argument.

Returns

A quoted pattern.

Argument

string A quoted string.

Pat Succeed()

Description

Constructs a pattern that always succeeds, even when the matcher backs into it. It matches the quoted null string.

Returns

1 when the match succeeds.

Argument

none

Pat Tab(*int*)

Description

Constructs a pattern that matches forward to position *int* in the quoted source string. It can match 0 or more characters. It fails if it would have to move backwards or beyond the end of the string.

Returns

A pattern.

Argument

int An integer that specifies a position in a quoted string.

Pat Test(*expr*)

Description

Constructs a pattern that succeeds and matches the quoted null string if *expr* is not zero and fails otherwise.

Returns

A quoted pattern.

Argument

expr An expression.

Notes

Usually the argument is wrapped with `expr()` because the test needs to be made on the current value of variables set by `Pat Immediate`, `Pat Conditional`, and `Pat At`. Without `expr`, the test is based on values that were known when the pattern was constructed, which means the test always succeeds or always fails at pattern execution time, which is probably not what you want.

Example

```
nCats = 0;
whichCat = 3;
string = "catch a catnapping cat in a catsup factory";
rc = Pat Match(
   string,
   "cat" + Pat Test(
      Expr(
         nCats = nCats + 1;
         nCats == whichCat;
      )
   ),
   "dog"
);
Show( rc, string, nCats );
   rc = 1
   string = "catch a catnapping dog in a catsup factory"
   nCats = 3
```

`Regex Match(`*source*`, `*pattern*`, <`*replacement string*`>|<"MATCHCASE">, <"NULL">)`

Description

Executes the pattern match in quoted *pattern* against the quoted *source* string.

Returns

A pattern.

Required Arguments

source A quoted string.

pattern A quoted pattern.

Optional Arguments

replacement string The quoted string that specifies the text to replace the source with.

"MATCHCASE" The search is case insensitive unless you specify MATCHCASE.

"NULL" Indicates that the expression contains MATCHCASE but you don't want to specify a replacement.

Examples

```
Regex Match(
   "person=Fred id=77 friend= favorite=tea", // source
   "(\w+)=(\S*) (\w+)=(\S*) (\w+)=(\S*) (\w+)=(\S*)" // pattern
);
   {"person=Fred id=77 friend= favorite=tea", "person", "Fred", "id", "77",
      "friend", "", "favorite", "tea"}

// case-insensitive, no replacement
Regex Match( "beliEve", "([aeiou])(.*?)(\1)" );
   {"eliE", "e", "li", "E"}
// case-sensitive, no replacement
Regex Match( "beliEve", "([aeiou])(.*?)(\1)", NULL, MATCHCASE );
   {"eliEve", "e", "liEv", "e"}
```

Comment Functions

// comment

Description

Comments to end of line.

Notes

Everything after the // is ignored when running the script.

/* comment */

Description

A comment that can appear in the middle of a line of script.

Notes

Anything between the beginning tag /* and the end tag */ is ignored when running the script. This comment style can be used almost anywhere, even inside lists of arguments. If you place a comment inside a double-quoted string, the comment is treated merely as part of the string and not a comment. You cannot place comments in the middle of operators.

Examples

```
+/*comment*/=
:/*comment*/name
```

are invalid and produce errors. The first comment interrupts += and the second interrupts :name.

```
sums = {(a+b /*comment*/), /*comment*/ (c^2)}
```

is valid JSL; the comments are both ignored.

`//!`

Description

If placed on the first line of a script, this comment line causes the script to be run when opened in JMP without opening into the script editor window.

Notes

You can over-ride this comment when opening the file. Select **File > Open**. Hold the Ctrl key while you select the JSL file and click **Open**. Or right-click the file in the Home Window Recent Files list and select **Edit Script**. The script opens into a script window instead of being executed.

`/*debug step*/`

`/*debug run*/`

Description

If placed on the first line of a script, the script is opened in the debugger when it is run.

Notes

All letters must be lowercase. There must be one space between debug and step or run, and there must be no other spaces present. Only one of these lines can be used, and it must be the first line of the script; a first line that is blank followed by this comment negates the debug command.

Comparison Functions

The comparison operators (<, <=, >, >=) work for numbers, quoted strings, and matrices. For matrices, they produce a matrix of results. If you compare mixed arguments, such as strings with numbers or matrices, the result is a missing value. Comparisons involving lists are not allowed and also return missing values.

The equality operators (== and !=) work for numbers, quoted strings, matrices, and lists. For matrices, they produce a matrix of results; for lists, they produce a single result. If you *test equality* of mixed results (for example. strings with numbers or matrices) the result is 0 or unequal.

Range check operators let you check whether something falls between two specified values:

```
a = 1;
Show( 1 <= a < 3 );
b = 2;
Show( 2 < b <= 3 );
   1 <= a < 3 = 1;
   2 < b <= 3 = 0;
```

Expressions with comparison operators are evaluated all at once, not in sequence

All the comparison operators are *eliding operators*. That means JMP treats arguments joined by comparison operators as one big clause, as opposed to the way most expressions are evaluated one operator at a time. Evaluating as a single clause produces different results than the more usual method of evaluating in pieces. For example, the following two statements are different:

```
12 < a < 13;
(12 < a) < 13;
```

The first statement checks whether *a* is between 12 and 13, because all three arguments and both operators are read and evaluated together. The second statement uses parentheses to regroup the operations explicitly to evaluate from left to right, which would be the normal way to evaluate most expressions. Thus it first checks whether 12 is less than *a*, returning 1 if true or 0 if false. Then it checks whether the result is less than 13, which is always true because 0 and 1 are both less than 13.

All the comparison operators are elided when they are used in matched pairs or in the unmatched pairs <... <= and <=... <. What this means is that if you want a comparison statement to be evaluated one comparison operator at a time, you should use parentheses ( ) to control the order of operations explicitly.

`Equal(a, b, ...)`

`a==b==...`

Description

Compares all the listed values and tests if they are all equal to each other.

Returns

1 (true) if all arguments evaluate to the same value.

0 (false) otherwise.

Arguments

Two or more variables, references, matrices, or numbers.

Notes

If more than two arguments are specified, a 1 is returned only if all arguments are exactly the same. This is typically used in conditional statements and to control loops.

The comparison is case-sensitive for quoted string comparisons.

`Greater(a, b, ...)`

`a>b>...`

Description

Compares all the list values and tests if, in each pair, the left value is greater than the right.

Returns

1 (true) if *a* evaluates strictly greater than *b* (and *b* evaluates strictly greater than *c*, and so on).

0 (false) otherwise.

Arguments

Two or more variables, references, matrices, or numbers.

Notes

If more than two arguments are specified, a 1 is returned only if each argument is greater than the one that follows it. This is typically used in conditional statements and to control loops.

`Greater`, `Less`, `Greater Or Equal`, and `Less Or Equal` can also be strung together (as in `0 < x <= 5`). If you do not group with parentheses, JMP evaluates each pair left to right. You can also use parentheses to explicitly tell JMP how to evaluate the expression.

`Greater or Equal(`*a*`, `*b*`, ...)`

a>=*b*>=...

Description

Compares all the list values and tests if, in each pair, the left value is greater than or equal to the right.

Returns

1 (true) if *a* evaluates strictly greater than or equal to *b* (and *b* evaluates strictly greater than or equal to *c*, and so on).

0 (false) otherwise.

Arguments

Two or more variables, references, matrices, or numbers.

Notes

If more than two arguments are specified, a 1 is returned only if each argument is greater than or equal to the one that follows it. This is typically used in conditional statements and to control loops.

`Greater`, `Less`, `Greater Or Equal`, and `Less Or Equal` can also be strung together (as in `0 < x <= 5`). If you do not group with parentheses, JMP evaluates each pair left to right. You can also use parentheses to explicitly tell JMP how to evaluate the expression.

`Is Missing(`*expr*`)`

Description

Returns 1 if the expression yields a missing value and 0 otherwise.

Less(*a*, *b*, ...)

a<*b*<...

Description

Compares all the list values and tests if, in each pair, the left value is less than the right.

Returns

1 (true) if *a* evaluates strictly less than *b* (and *b* evaluates strictly less than *c*, and so on).

0 (false) otherwise.

Arguments

Two or more variables, references, matrices, or numbers.

Notes

If more than two arguments are specified, a 1 is returned only if each argument is less than the one that follows it. This is typically used in conditional statements and to control loops.

`Greater`, `Less`, `Greater Or Equal`, and `Less Or Equal` can also be strung together (as in `0 < x <= 5`). If you do not group with parentheses, JMP evaluates each pair left to right. You can also use parentheses to explicitly tell JMP how to evaluate the expression.

Less Less Equal(*a*, *b*, *c*, ...)

a<*b*<=*c*<=...

Description

Range check, exclusive below and inclusive above.

Returns

1 (true) if *b* is greater than *a* and less than or equal to *c*.

0 (false) otherwise.

Arguments

a, *b*, *c* Variables, references, matrices, or numbers.

Notes

Returns 1 when two conditions are met: the first argument is less than the second argument, and each remaining argument is less than or equal to its argument on the right. This is typically used in conditional statements and to control loops.

Less or Equal(*a*, *b*, ...)

a<=*b*<=...

Description

Compares all the list values and tests if, in each pair, the left value is less than or equal to the right.

Returns

1 (true) if *a* evaluates strictly less than or equal to *b* (and *b* evaluates strictly less than or equal to *c*, and so on).

0 (false) otherwise.

Arguments

Two or more variables, references, matrices, or numbers.

Notes

If more than two arguments are specified, a 1 is returned only if each argument is less than or equal to the one that follows it. This is typically used in conditional statements and to control loops.

`Greater`, `Less`, `Greater Or Equal`, and `Less Or Equal` can also be strung together (as in `0 < x <= 5`). If you do not group with parentheses, JMP evaluates each pair left to right. You can also use parentheses to explicitly tell JMP how to evaluate the expression.

`Less Equal Less(a, b, c, ...)`

`a<=b<c<...`

Description

A range check, inclusive below and exclusive above.

Returns

1 (true) if *b* is greater than or equal to *a* and less than *c*.

0 (false) otherwise.

Arguments

a, *b*, *c* Variables, references, matrices, or numbers.

Notes

Returns 1 when two conditions are met: the first argument is less than or equal to the second argument, and each remaining argument is less than its argument on the right. This is typically used in conditional statements and to control loops.

`Not Equal(a, b)`

`a!=b`

Description

Compares *a* and *b* and tests if they are equal.

Returns

0 (false) if *a* and *b* evaluate to the same value.

1 (true) otherwise.

Argument

a, b Any variable or number.

Notes

Mostly used for conditional statements and loop control.

Conditional and Logical Functions

And(*a, b*)

a&*b*

Description

The logical And.

Returns

1 (true) if both *a* and *b* are true.

0 (false) if either *a* or *b* is false or if both *a* and *b* are false.

Missing if either *a* or *b* is a missing value or if both *a* and *b* are missing values.

Arguments

Two or more variables or expressions.

Notes

More than two arguments can be strung together. a&b returns 1 (true) only if all arguments evaluate to true.

AndMZ(*a, b*)

a:&*b*

Description

Returns the logical AND of all arguments. Missing values are treated as zeroes.

Returns

1 (true) if both *a* and *b* are true.

0 (false) if either *a* or *b* is false or if both *a* and *b* are false.

0 (false) if either *a* or *b* is a missing value or if both *a* and *b* are missing values.

Arguments

Two or more variables or expressions.

Notes

More than two arguments can be strung together. *a*:&*b* returns 1 (true) only if all arguments evaluate to true.

Break()

Description

Stops execution of a loop completely and continues to the statement following the loop.

Notes

`Break()` works with `For` and `While` loops and also with `For Each Row`.

Choose(*expr*, *r1*, *r2*, *r3*, ..., *resultElse*)

Description

Evaluates *expr*. If the value of *expr* is 1, *r1* is returned; if 2, the value of *r2* is returned, and so on. If no matches are found, the last argument (*resultElse*) is returned.

Returns

The value whose index in the list of arguments matches *expr*, or the value of the last argument.

Arguments

expr An expression or a value.

r1, *r2*, *r3*, ... An expression or a value.

resultElse The argument that is returned when no matches are found.

Continue()

Description

Ends the current iteration of a loop and begins the loop at the next iteration.

Notes

`Continue()` works with `For` and `While` loops, and also with `For Each Row`.

Filter Each(*names*, *container*, *locals*, *body*)

Description

Iterates over a container, which can be a list, an associative array, or a matrix, and returns a subset of the values in the container based on the evaluation of a Boolean expression at each iteration. The value, key, or element, as well as the index number, are all available at each iteration.

- For associative array containers, the key and value can be accessed using a two-item list.
- For matrix containers, a linear index is provided by default, but a two-item list can be used to access the row and column indices.

These symbols are provided within the body of the loop only. A list of local variables can also be provided. If local variables are defined, they are initialized after the first iteration symbols are set.

Returns

A subset from the original container. The return object has the same type as the original container.

Arguments

names Specification of loop control variable names, specified as a list. The form of the list is determined by the type of container. All of the names are optional, so you need to specify them only if you need to refer to them in the *locals* or *body* arguments. If you do not specify any names, the first argument should be an empty list, which can be specified as {} or List().

For a *list* container, the names list contains a name for each value in the list and a name for the index of each value in the list.

For an *associative array* container, the names list contains a two-item list of names and a name for the index of each item in the associative array.

For a *matrix* container, the names list contains a name for each element in the matrix and a second argument for the index of each element in the matrix. The second argument can be specified as a single name or as a two-item list of names that represent symbols for the row and column indices.

When specifying multiple containers using the Across() keyword, the names list contains a list of names that refer to values in each container. The number of names in the first item in the names list must match the number of containers specified in the Across() keyword.

container A list, associative array, or a matrix. The container can be defined in the argument or it can be a reference to a previously defined object.

This argument can also use the Across() keyword, which enables you to use the function across multiple containers. The multiple containers can be specified as separate arguments or as items in a list. The Across() keyword has an optional Count() argument that enables you to specify how containers of different sizes are handled. The available Count() options are: "Shortest", "Longest", "Enforce Equal", and N, where N is a number. If you specify a number, the function iterates through all containers exactly N times; note that the function loops back to the start of containers that have fewer than N items.

locals A list of variables that are local to the function. This is equivalent to other Local variable initializations in JSL. The initialization of the local variables occurs after the first loop control variables have been set, but the local variables do not get initialized again after that.

body Any number of valid JSL expressions, glued together if there are more than one. The result of the JSL expressions should be a Boolean value. If the result of the expression is true, the container value at the current iteration is included in the result; otherwise, the

container value at the current iteration is not included in the result. You can use the `Continue()` function as an equivalent way to return false and skip to the next iteration. You can also use `Break()` function to stop iteration through the loop and proceed to the next expression that follows the loop. See "Break and Continue".

Example

```
values = Filter Each( {x}, {10, 20, 30}, x > 15 );
Show( values );
  values = {20, 30};
```

For(*init*, *while*, *increment*, *body*)

Description

Repeats the statement(s) in the *body* as long as the *while* condition is true. *init* and *increment* control iterations.

Returns

Null.

Arguments

init Initialization of loop control counter.

while Condition for loop to continue or end. As long as the conditional statement while is true, the loop is iterated one more time. As soon as while is false, the loop is exited.

increment Increments (or decrements) the loop counter after while is evaluated every *time* the loop is executed.

body Any number of valid JSL expressions, glued together if there are more than one.

Example

```
mysum = 0; myprod = 1;
For( i = 1, i <= 10, i++, mysum += i; myprod *= i; );
Show( mysum, myprod );
  mysum = 55;
  myprod = 3628800;
```

For Each(*names*, *container*, *locals*, *body*)

Description

Iterates over a container, which can be a list, an associative array, or a matrix, and provides the value, key, or element at each iteration. The index number is also available at each iteration. For associative array containers, the key and value can be accessed using a two-item list. For matrix containers, a linear index is provided by default, but a two-item list can be used to access the row and column indices. These symbols are provided within the body of the loop only. A list of local variables can also be provided. If local variables are defined, they are initialized after the first iteration symbols are set.

Arguments

names Specification of loop control variable names, specified as a list. The form of the list is determined by the type of container. All of the names are optional, so you need to specify them only if you need to refer to them in the *locals* or *body* arguments. If you do not specify any names, the first argument should be an empty list, which can be specified as {} or List().

For a *list* container, the names list contains a name for each value in the list and a name for the index of each value in the list.

For an *associative array* container, the names list contains a two-item list of names and a name for the index of each item in the associative array.

For a *matrix* container, the names list contains a name for each element in the matrix and a second argument for the index of each element in the matrix. The second argument can be specified as a single name or as a two-item list of names that represent symbols for the row and column indices.

When specifying multiple containers using the Across() keyword, the names list contains a list of names that refer to values in each container. The number of names in the first item in the names list must match the number of containers specified in the Across() keyword.

container A list, associative array, or a matrix. The container can be defined in the argument or it can be a reference to a previously defined object.

This argument can also use the Across() keyword, which enables you to use the function across multiple containers. The multiple containers can be specified as separate arguments or as items in a list. The Across() keyword has an optional Count() argument that enables you to specify how containers of different sizes are handled. The available Count() options are: "Shortest", "Longest", "Enforce Equal", and N, where N is a number. If you specify a number, the function iterates through all containers exactly N times; note that the function loops back to the start of containers that have fewer than N items.

locals A list of variables that are local to the function. This is equivalent to other Local variable initializations in JSL. The initialization of the local variables occurs after the first loop control variables have been set, but the local variables do not get initialized again after that.

body Any number of valid JSL expressions, glued together if there are more than one. You can use the Continue() function to skip to the next iteration. You can also use Break() function to stop iteration through the loop and proceed to the next expression that follows the loop. See "Break and Continue".

Example

```
For Each( {value, index}, {10, 20, 30}, Show( value, index ) );
  value = 10;
  index = 1;
  value = 20;
  index = 2;
```

```
value = 30;
index = 3;
```

For Each Row(<*dt*>, *script*)

Description

Repeats the *script* on each row of the data table.

Returns

Null.

Required Argument

script Any valid JSL expressions.

Optional Argument

dt Positional argument that is a reference to a data table. If this argument is not in the form of an assignment, then it is considered a data table expression.

Example

The following example creates data table references and then iterates over each row in Big Class.jmp. If the value of age in a row is greater than 15, the age is printed to the log.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
For Each Row( dt, If( :age > 15, Show( :age ) ) );
```

If(*condition1*, *result1*, <*condition2*, *result2*>, ..., <*elseResult*>)

Description

Evaluates the first of each pair of arguments and returns the evaluation of the *result* expression associated with the first *condition* argument that evaluates to a nonzero result. The *condition* arguments are evaluated in order. If all of the *condition* arguments evaluate to zero, the optional *elseResult* is evaluated and the result is returned. If no *elseResult* is specified, and none of the conditions are true, a missing value is returned. If all of the *condition* arguments evaluate to missing, a missing value is returned.

IfMax(*expr1*, *result1*, *expr2*, *result2*, ... <*all missing results*>)

Description

Evaluates the first of each pair of arguments and returns the evaluation of the result expression (the second of each pair) associated with the maximum of the expressions. If more than one expression is the maximum, the first maximum is returned. If all expressions are missing and a final result is not specified, missing is returned. If all expressions are missing and a final result is specified, that final result is returned. The test expressions must evaluate to numeric values, but the result expressions can be anything.

Returns

The result expression associated with the maximum of the expressions

IfMin(*expr1*, *result1*, *expr2*, *result2*, ... <*all missing results*>)

Description

Evaluates the first of each pair of arguments and returns the evaluation of the result expression (the second of each pair) associated with the minimum of the expressions. If more than one expression is the minimum, the first minimum is returned. If all expressions are missing and a final result is not specified, missing is returned. If all expressions are missing and a final result is specified, that final result is returned. The test expressions must evaluate to numeric values, but the result expressions can be anything.

Returns

The result expression associated with the minimum of the expressions

IfMZ(*condition1*, *result1*, <*condition2*, *result2*>, ..., <*elseResult*>)

Description

Evaluates the first of each pair of arguments and returns the evaluation of the *result* expression associated with the *condition1* argument that evaluates to a nonzero result. The *condition* arguments are evaluated in order. If all of the *condition* arguments evaluate to zero or missing, the optional *elseResult* is evaluated and the result is returned. If no *elseResult* is specified, and none of the conditions are true, a missing value is returned.

Notes

The test arguments are evaluated in order until the first nonzero result. If all test results return zero or missing, the *elseExpr* argument is evaluated.

IfMZ() is equivalent to If() where missing values for evaluated *condition* arguments are treated as zero.

Interpolate(*x*|*xmatrix*|*xlist*, *x1*, *y1*, *x2*, *y2*, ...)

Interpolate(*x*|*xmatrix*|*xlist*, *xmatrix*, *ymatrix*)

Interpolate({*x*, *y*}, *xvector*, *yvector*, *zmatrix*)

Description

Performs linear interpolation for continuous data. There are many ways to specify the function.

In the simplest cases where the first argument is a single numeric value, the *y* value corresponding to a given *x* value between two sets of points or by matrices *xmatrix* and *ymatrix*. The *x* values, (*x1*, *x2*, ...) or *xmatrix*, must be in ascending order.

If the first argument is a matrix or list of numeric values, the resulting matrix or list is a set of interpolated values.

You can perform *bilinear interpolation* using the four-argument case. Here, the first argument is a list of two points, the second and third arguments are vectors that define the grid of x and y values, and the fourth argument is a matrix of data points. The function then finds the interpolated *z* value within the appropriate quadrant of the *zmatrix*.

Returns

The interpolated value or values. In the three-argument cases, the return object type matches the type of the first argument. In the four-argument case, the return object is a number.

`Is Associative Array(name)`

Description

Returns 1 if the evaluated argument is an associative array, or 0 otherwise.

`Is Empty(global)`

`Is Empty(dt)`

`Is Empty(col)`

Description

Returns 1 if the *global* variable, data table, or data column is undefined or holds the `Empty()` value, or 0 otherwise.

`Is Expr(x)`

Description

Returns 1 if the evaluated argument is an expression, or 0 otherwise.

`Is List`

See "Is List(x)".

`Is Name(x)`

Description

Returns 1 if the evaluated argument is a name, or 0 otherwise.

`Is Namespace(namespace)`

Description

Returns 1 if the namespace argument is a namespace; returns 0 otherwise.

`Is Number(x)`

Description

Returns 1 if the evaluated argument is a number or missing numeric value, or 0 otherwise.

`Is Scriptable(x)`

Description

Returns 1 if the evaluated argument is a scriptable object, or 0 otherwise.

`Is String(x)`

Description

Returns 1 if the evaluated argument is a quoted string, or 0 otherwise.

`Match(x, value1, result1, value2, result2, ..., resultElse)`

Description

If *a* is equal to *value1*, then *result1* is returned. If *a* is equal to *value2*, *result2* is returned, and so on.

Notes

The `Match()` function explicitly checks to see if the compare expression *x* is missing and if the value of *value1* is missing, then it returns the value of `result1`; otherwise it continues to compare the expression *x* to each *valueN* value in each *valueN/resultN* pair, ignoring any missing values. If the expression *x* is equal to any of the *valueN* value, then the corresponding *resultN* value is returned. If no matching *valueN* value is found, then the *resultElse* value is returned.

`MatchMZ(x, value1, expr1, value2, expr2, ..., exprElse)`

Description

Evaluates and returns the *exprN* argument that equals *x* or evaluates and returns the *exprElse* argument if no value equals *x*.

Notes

The `MatchMZ()` function works the same as the `Match()` function except that missing values are treated as 0.

`Not(a)`

`!a`

Description

The logical Not.

Returns

0 (false) if *a*>0.

1 (true) if *a*<=0.

Missing value if *a* is missing.

Argument

a Any variable or number. The variable must have a numeric or matrix value.

Notes

Mostly used for conditional statements and loop control.

`Or(a, b)`

`a|b`

Description

The logical Or.

Returns

1 (true) if either of or both *a* and *b* are true.

0 (false) otherwise.

Missing if either are missing.

Arguments

a, *b* Any variable or number.

Notes

Mostly used for conditional statements and loop control.

`OrMZ(a, b)`

`a:|b`

Description

Returns the logical OR of all arguments with missing values treated as zeroes: 1 if any arguments are nonzero and 0 otherwise.

Returns

1 (true) if either of or both *a* and *b* are true.

0 (false) otherwise.

Arguments

a, *b* Any variable or number.

Notes

– Mostly used for conditional statements and loop control.

– `Or()` returns missing if any evaluated argument is missing. `OrMZ()` returns 0 if any evaluated argument is missing.

`Return(<expr1>, <expr2>, ..., <exprN>)`

Description

Returns an expression value from a user-defined function.

Example

This example returns the evaluation of both expressions in the `Return()` function as a list. The `Return()` function can have more than one argument. If only one is present, then the value of the expression is returned. If more than one is present, then the values of all the expressions is returned in a list.

```
f = Function( {a, b},
  Return( a - b, a + b )
);
{lo, hi} = f( 10, 1 );
Show( lo, hi );
Show( f( 7, 15 ) );
   lo = 9;
   hi = 11;
   f(7, 15) = {-8, 22};
```

Notes

`Return()` not enclosed by a function, method, or recursive function call causes an error.

`Step(x0, x1, y1, x2, y2, ...)`

`Step(x0, [x1, x2, ...], [y1, y2, ...])`

Description

Returns the *y* argument corresponding to the largest *x* argument that is less than or equal to *x0*. The *x* points must be specified in ascending order.

`Stop()`

Description

Immediately stops a script that is running.

`Transform Each(names, container, <Output(type)>, <locals>, body)`

Description

Iterates over a container, which can be a list, an associative array, or a matrix, and updates each of the values in the container based on the evaluation of a JSL expression at each iteration. The value, key, or element, as well as the index number, are all available at each iteration. For associative array containers, the key and value can be accessed using a

two-item list. For matrix containers, a linear index is provided by default, but a two-item list can be used to access the row and column indices. These symbols are provided within the body of the loop only. A list of local variables can also be provided. If local variables are defined, they are initialized after the first iteration symbols are set.

Returns

An updated version of the original container. The return object has the same type as the original container, unless the `Output()` keyword is used to designate a different type for the returned container.

Arguments

`names` Specification of loop control variable names, specified as a list. The form of the list is determined by the type of container. All of the names are optional, so you need to specify them only if you need to refer to them in the *`locals`* or *`body`* arguments. If you do not specify any names, the first argument should be an empty list, which can be specified as `{}` or `List()`.

For a *list* container, the names list contains a name for each value in the list and a name for the index of each value in the list.

For an *associative array* container, the names list contains a two-item list of names and a name for the index of each item in the associative array.

For a *matrix* container, the names list contains a name for each element in the matrix and a second argument for the index of each element in the matrix. The second argument can be specified as a single name or as a two-item list of names that represent symbols for the row and column indices.

When specifying multiple containers using the `Across()` keyword, the names list contains a list of names that refer to values in each container. The number of names in the first item in the names list must match the number of containers specified in the `Across()` keyword.

`container` A list, associative array, or a matrix. The container can be defined in the argument or it can be a reference to a previously defined object.

This argument can also use the `Across()` keyword, which enables you to use the function across multiple containers. The multiple containers can be specified as separate arguments or as items in a list. The `Across()` keyword has an optional `Count()` argument that enables you to specify how containers of different sizes are handled. The available `Count()` options are: `"Shortest"`, `"Longest"`, `"Enforce Equal"`, and `N`, where `N` is a number. If you specify a number, the function iterates through all containers exactly N times; note that the function loops back to the start of containers that have fewer than N items.

`Output(type)` Specifies a type for the output. This can be `"List"`, `"Matrix"`, or `"Associative Array"`. By default, the output type matches the type of the input container.

`locals` A list of variables that are local to the function. This is equivalent to other Local variable initializations in JSL. The initialization of the local variables occurs after the

first loop control variables have been set, but the local variables do not get initialized again after that.

body Any number of valid JSL expressions, glued together if there are more than one. The result of the JSL expressions at each iteration is used in the output container. You can use the `Continue()` function to return no value for an iteration and skip to the next iteration. You can also use `Break()` function to stop iteration through the loop and proceed to the next expression that follows the loop. See "Break and Continue".

Example

```
values = Transform Each( {x}, {10, 20}, x + 10 );
Show( values );
   values = {20, 30};
```

While(*expr*, *body*)

Description

Repeatedly tests the *expr* condition and executes the *body* until the *expr* condition is no longer true.

Zero Or Missing(*expr*)

Description

Returns 1 if *expr* yields a missing value or zero, 0 otherwise.

Constant Functions

JMP provides functions for two useful constant functions.

Note: These functions do not take an argument, but the parentheses are required.

e()

Description

Returns the constant *e*, which is 2.7182818284590451...

Pi()

Description

Returns the constant π, which is 3.1415926535897931...

Date and Time Functions

Datetime values are handled internally as numbers of seconds since midnight, January 1, 1904.

The expression `x=01Jan1904` sets *x* to zero, since the indicated date is the base date or "zero date" in JMP. If you examine the values of dates, they should be appropriately large numbers (for example, `5oct1998` is 2990390400).

Abbrev Date(*date*)

Description

Converts the provided *date* to a quoted string.

Returns

A quoted string representation of the date.

Argument

date Can be the number of seconds since the base date (midnight, January 1, 1904), or any date-time operator.

Example

```
Abbrev Date( 29Feb2004 );
```

02/29/2004

As Date(*x*)

Description

Formats the number or expression *x* so that it shows as a date or duration when displayed in a text window. Values that represent one year or more are returned as dates. Values that represent less than a year are returned as durations.

Returns

A date that is calculated from the number or expression provided.

Argument

x Number or expression.

Date Difference(*datetime1*, *datetime2*, *interval name*, <*alignment*>)

Description

Returns the difference in intervals of two date-time values.

Returns

A number.

Required Arguments

datetime1, *datetime2* Date-time values.

interval name A quoted string that contains a date-time interval, such as "month", "day", or "hour".

Optional Arguments

alignment A quoted string:

- "Start" includes full or partial intervals.
- "Actual" counts only whole intervals.
- "Fractional" returns fractional differences using averages for "year", "quarter", and "month" intervals.

Date DMY(*day*, *month*, *year*)

Description

Constructs a date value from the arguments.

Returns

The specified date, expressed as the number of seconds since midnight, 1 January 1904.

Arguments

day The numeric day of month, 1-31. Note that there is no error-checking, so you can enter February 31.

month The numeric month, 1-12.

year The year.

Date Increment(*datetime*, *interval name*, <*increment*>, <*alignment*>)

Description

Adds 1 or more intervals to a starting datetime value.

Returns

Returns the new datetime value.

Required Arguments

datetime The starting datetime value.

interval name A quoted string that contains the name of a datetime interval. "year", "quarter", "month", "week", "day", "hour", "minute", and "second" are supported.

Optional Arguments

increment A number that specifies the number of intervals. The default value is 1.

alignment A quoted string that contains a keyword:

- "Start" truncates the date to the nearest interval prior to adding the increment. For example, it removes the time and outputs the date. "start" is the default value.
- "Actual" retains the full input datetime value.

– "Fractional" allows fractional incremental values using averages for the duration of "Year", "Quarter", and "Month" intervals.

Date MDY(*month*, *day*, *year*)

Description

Constructs a date value from the arguments.

Returns

The specified date, expressed as the number of seconds since midnight, 1 January 1904.

Arguments

month The numeric month, 1-12.

day The numeric day of month, 1-31. Note that there is no error-checking, so you can enter February 31.

year The year.

Day(*datetime*)

Description

Determines the day of the month supplied by the *datetime* argument.

Returns

Returns an integer representation for the day of the month of the date supplied.

Arguments

datetime The number of seconds since midnight, 1 January 1904. This can also be an expression.

Example

```
d1 = Date DMY( 12, 2, 2003 );
   3127852800

Day( 3127852800 );
   12
Day( d1 );
   12
```

Day Of Week(*datetime*)

Description

Determines the day of the week supplied by the *datetime* argument.

Returns

Returns an integer representation for the day of the week of the date supplied. Weeks are Sunday–Saturday.

Arguments

datetime The number of seconds since midnight, 1 January 1904. This can also be an expression.

Day Of Year(*datetime*)

Description

Determines the day of the year supplied by the *datetime* argument.

Returns

Returns an integer representation for the day of the year of the date supplied.

Arguments

datetime The number of seconds since midnight, 1 January 1904. This can also be an expression.

Format(*x*, formatString, *width*|<*width*, *decimal places*>, <"Use Thousands Separator">)

Format(*x*, "Best", <*width*>, <"Use Thousands Separator">)

Format(*x*, ("Fixed Dec"|"Percent"), *width*|<*width*, *decimal places*>, <"Use Thousands Separator">)

Format(*x*, "Pvalue", <*width*>)

Format(*x*, ("Scientific"|"Engineering"|"Engineering SI"), <*width*>|<*width*, *decimal places*>)

Format(*x*,"Precision", *width*|<*width*, *decimal places*>, <"Use Thousands Separator">, <"Keep Trailing Zeroes">, <"Keep All Whole Digits">)

Format(*x*, "Currency", <*currency code*>, <*width*>|<*width*, *decimal places*>, <"Use Thousands Separator">, < <<Use Locale(*Boolean*) >)

Format(*x*, *datetime*, <*width*>)

Format(*x*, ("Latitude DDD"|"Latitude DDM"|"Latitude DMS"|"Longitude DDD"|"Longitude DDM"|"Longitude DDM"), *width*|<*width*, *decimal places*>, ("PUN"|"DIR"|"PUNDIR"))

Format(*x*, "Custom", Formula(), <*width*>)

Description

Converts the value *x* into the quoted *format* that you specify in the subsequent arguments.

Returns

Returns the text that corresponds to the number in the specified format.

Arguments

See *Using JMP* for more information about the arguments. The arguments are also shown in a data table column's Column Info window.

Examples

```
Format( x, "Fixed Dec", 10, 2, "Use Thousands Separator");
Format( x, "Currency", "EUR", 20, <<Use Locale(0)); // ignores computer locale
Format( x, "m/d/y", 10 );
Format( x, "Precision", 10, 2, "Keep trailing zeroes", "Keep All Whole Digits"
  );
Format( x, "Latitude DDD", "PUNDIR"); // "PUN" for punctuation, "DIR" for
  direction, PUNDIR for both
Format( x, "Custom", Formula( Abs( value ) ), 15 );
```

Notes

- You must always precede the number of decimal places with the width.
- If the date format is unknown, an error is written to the log.

Format Date(*x*, *datetime*, <*width*>)

Description

Converts the value of *x* into the quoted *datetime* that you specify in the second argument. Format choices are those shown in the data table Column Info window.

Returns

Returns the number in the specified format.

Arguments

See *Using JMP* for more information about the arguments.

Example

```
Format Date( Today(), "yyyQq" );
   "04/03/2020"
```

Hour(*datetime*, <"12"|"24">)

Description

Determines the hour supplied by the *datetime* argument.

Returns

Returns an integer representation for the hour part of the date-time value supplied.

Arguments

`datetime` The number of seconds since midnight, 1 January 1904. This can also be an expression.

`"12"|"24"` Changes the mode to 12 hours (with am and pm). The default is 24-hour mode.

HP Time()

Description

Returns a high precision time value (in microseconds). This function is only useful relative to another `HP Time()` value. The time value represents the number of microseconds since the start of the JMP session.

Notes

For less precise time values use `Tick Seconds()`.

In Days(*n*)

Description

Returns the number of seconds per *n* days. Divide by this function to express seconds as days.

Informat(*string*, *format*)

Parse Date(*string*, *format*)

Description

Parses a quoted *string* of a given quoted *format* and returns a date/time value. The value is expressed as if surrounded by the `As Date()` function, returning the date in "*ddMonyyyy*" format.

Example

```
Informat( "07152000", "MMDDYYYY" );
```

15Jul2000

Notes

- To see the format options, open the Column Info window on a data table column, select a date/time value for the format, and view the Input Format list.
- If the date format is unknown, an error is written to the log.

See Also

"As Date(x)"

In Hours(*n*)

Description

Returns the number of seconds per *n* hours. Divide by this function to express seconds as hours.

In Minutes(*n*)

Description

Returns the number of seconds per *n* minutes. Divide by this function to express seconds as minutes.

In Weeks(*n*)

Description

Returns the number of seconds per *n* weeks. Divide by this function to express seconds as weeks.

In Years(*n*)

Description

Returns the number of seconds per *n* years. Divide by this function to express seconds as years.

ISO Year(*datetime*)

Description

Returns the ISO year of the *datetime* supplied. ISO years correspond to ISO weeks; they begin on the Monday of the first week that contains at least four days.

Long Date(*date*)

Description

Returns a locale-specific quoted string representation for the *date* supplied, formatted like "Sunday, February 29, 2004" or "Wednesday, November 9, 2011".

MDYHMS(*date*)

Description

Returns a quoted string representation for the *date* supplied, formatted like "2/29/04 00:02:20".

Minute(*datetime*)

Description

Determines the minute supplied by the *datetime* argument, 0-59.

Returns

Returns an integer representation for the minute part of the date-time value supplied.

Month(*date*)

Description

Returns an integer representation for the month of the *date* that is supplied.

Nth Day of Week in the Month(*datetime*)

Description

Determines the day of the week of the *datetime* argument and how many instances of that day of the week have occurred in the month of the *datetime* argument.

Returns

Returns an integer that represents the number of instances of the day of the week of the *datetime* argument that have occurred in the month.

Parse Date()

See "Informat(string, format)".

Quarter(*datetime*)

Description

Returns the annual quarter of a *datetime* value as an integer 1-4.

Second(*datetime*)

Description

Determines the second supplied by the *datetime* argument.

Returns

Returns an integer representation for the second part of the date-time value supplied.

Argument

datetime Number of seconds since midnight, 1 January 1904. This can also be an expression.

Short Date(*date*)

Description

Returns a quoted string representation for the *date* supplied in the format MM/DD/YYYY.

Tick Seconds()

Description

Measures the time taken for a script to run, measured down to the 60th of a second.

Notes

For higher time value resolution (for example, microseconds) use the `HP Time()` function.

`Time Of Day(datetime)`

Description

Returns an integer representation for the time of day of the *datetime* supplied.

`Today()`

Description

Returns the current date and time expressed as the number of seconds since midnight, 1 January 1904. No arguments are available, but the parentheses are still necessary.

`Week Of Year(date, <ruleN>)`

Description

Returns the week of the year that contains a date-time value. Three rules determine when the first week of the year begins.

- With rule 1 (the default), weeks start on Sunday, with the first Sunday of the year being week 2. Week 1 is a partial week or empty.
- With rule 2, the first Sunday begins with week 1, with previous days being week 0.
- With rule 3, the ISO-8601 week number is returned. Weeks start on Monday. Week 1 is the first week of the year with four days in that year. It is possible for the first or last three days of the year to belong to the neighboring year's week number.

`Year(date)`

Description

Returns an integer representation for the year of *date*.

Discrete Probability Functions

`Beta Binomial Distribution(k, p, n, delta)`

Description

Returns the cumulative distribution function (cdf) of the beta binomial distribution. This is the probability that a beta binomially distributed random variable is less than or equal to *k*. The cdf is calculated as the summation of the beta binomial pmf for values of X from 0 to *k*.

Arguments

k The count of interest. *k* must be an integer.

p The probability of success for each trial, which must be between 0 and 1.

n The number of trials, which must be greater than 1.

delta The overdispersion parameter, which must be between Maximum[-*p*/(*n*-*p*-1), -(1-*p*)/(*n*-2+*p*)] and 1. When the overdispersion parameter is zero, the distribution reduces to Binomial(*n*, *p*).

Beta Binomial Probability(*k*, *p*, *n*, *delta*)

Description

Returns the probability mass function (pmf) of the beta binomial distribution. This is the probability that a beta binomially distributed random variable is equal to *k*. The pmf is parameterized as follows:

$$P(X = k; p, n, \delta) = \binom{n}{k} \frac{\Gamma\left(\frac{1}{\delta} - 1\right)\Gamma\left[k + p\left(\frac{1}{\delta} - 1\right)\right]\Gamma\left[n - k + (1-p)\left(\frac{1}{\delta} - 1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta} - 1\right)\right]\Gamma\left[(1-p)\left(\frac{1}{\delta} - 1\right)\right]\Gamma\left(n + \frac{1}{\delta} - 1\right)}$$

Arguments

k The count of interest. *k* must be an integer.

p The probability of success for each trial, which must be between 0 and 1.

n The number of trials, which must be greater than 1.

delta The overdispersion parameter δ, which must be between Maximum[-*p*/(*n*-*p*-1), -(1-*p*)/(*n*-2+*p*)] and 1. When the overdispersion parameter is zero, the distribution reduces to Binomial(*n*, *p*).

Notes

The beta binomial distribution results from assuming that X|π follows a Binomial(n,π) distribution and π follows a Beta(p(1-δ)/δ,(1-p)(1-δ)/δ) distribution. It is useful when the data are a combination of several Binomial distributions that each have different probabilities of success.

Beta Binomial Quantile(*p*, *n*, *delta*, *cumprob*)

Description

Returns the smallest integer quantile for which the cumulative probability of the Beta Binomial(p, n, delta) distribution is larger than or equal to *cumprob*.

Arguments

p The probability of success for each trial. *p* must be between 0 and 1.

n The number of trials, which must be greater than 1.

delta The overdispersion parameter δ, which must be between Maximum[-*p*/(*n*-*p*-1), -(1-*p*)/(*n*-2+*p*)] and 1. When the overdispersion parameter is zero, the distribution reduces to Binomial(*n*, *p*).

cumprob The cumulative probability of the quantile desired. *cumprob* must be between 0 and 1.

Binomial Distribution(*p*, *n*, *k*)

Description

Returns the cumulative distribution function (cdf) of the binomial distribution. This is the probability that a binomially distributed random variable is less than or equal to *k*. The cdf is calculated as the summation of the binomial pmf for values of X from 0 to *k*.

Arguments

p The probability of success for each trial. *p* must be between 0 and 1.

n The number of trials.

k The number of successes, which must be less than or equal to *n*.

Binomial Probability(*p*, *n*, *k*)

Description

Returns the probability mass function (pmf) of the binomial distribution. This is the probability that a binomially distributed variable is equal to *k*. The pmf is parameterized as follows:

$$P(X = k;p, n) = \binom{n}{k} p^k (1-p)^{n-k}$$

Arguments

p The probability of success for each trial. *p* must be between 0 and 1.

n The number of trials.

k The number of successes, which must be less than or equal to *n*.

Binomial Quantile(*p*, *n*, *cumprob*)

Description

Returns the smallest integer quantile for which the cumulative probability of the Binomial(p, n) distribution is larger than or equal to *cumprob*.

Arguments

p The probability of success for each trial. *p* must be between 0 and 1.

n The number of trials.

cumprob The cumulative probability of the quantile desired. *cumprob* must be between 0 and 1.

Gamma Poisson Distribution(*k*, *lambda*, *sigma*)

Description

Returns the cumulative distribution function (cdf) of the gamma-Poisson distribution. This is the probability that a gamma-Poisson distributed random variable is less than or equal to *k*. The cdf is calculated as the summation of the gamma-Poisson pmf for values of X from 0 to *k*.

Arguments

k The count of interest. *k* must be an integer.

lambda The shape parameter λ, which much be greater than 0. This is the mean of the distribution.

sigma The overdispersion parameter σ, which must be greater than or equal to 1. When the overdispersion parameter is 1, the distribution reduces to a Poisson(λ) distribution.

Gamma Poisson Probability(*k*, *lambda*, *sigma*)

Description

Returns the probability mass function (pmf) of the gamma-Poisson distribution. This is the probability that a gamma-Poisson distributed random variable is equal to *k*. The pmf is parameterized as follows:

$$P(X = k;\lambda, \sigma) = \frac{\Gamma\left(k + \frac{\lambda}{\sigma - 1}\right)}{\Gamma(k+1)\Gamma\left(\frac{\lambda}{\sigma - 1}\right)}\left(\frac{\sigma - 1}{\sigma}\right)^{k}\left(\sigma^{-\left(\frac{\lambda}{\sigma - 1}\right)}\right)$$

where Γ(·) is the Gamma function.

Arguments

k The count of interest. *k* must be an integer.

lambda The shape parameter λ, which much be greater than 0. This is the mean of the distribution.

sigma The overdispersion parameter σ, which must be greater than or equal to 1. When the overdispersion parameter is 1, the distribution reduces to a Poisson(λ) distribution.

Notes

The gamma Poisson distribution results from assuming that X|μ follows a Poisson(μ) distribution and μ follows a Gamma(λ/(σ-1),σ-1) distribution. It is useful when the data are a combination of several Poisson(μ) distributions that each have different values of μ.

Gamma Poisson Quantile(*lambda*, *sigma*, *cumprob*)

Description

Returns the smallest integer quantile for which the cumulative probability of the Gamma Poisson(lambda, sigma) distribution is larger than or equal to *cumprob*.

Arguments

lambda The shape parameter λ, which much be greater than 0. This is the mean of the distribution.

sigma The overdispersion parameter σ, which must be greater than or equal to 1. When the overdispersion parameter is 1, the distribution reduces to a Poisson(λ) distribution.

cumprob The cumulative probability of the quantile desired. *cumprob* must be between 0 and 1.

Hypergeometric Distribution(*N*, *K*, *n*, *x*, <*r*>)

Description

Returns the cumulative distribution function (cdf) of the hypergeometric distribution. This is the probability that a hypergeometrically distributed random variable is less than or equal to *x*. The cdf is calculated as the summation of the hypergeometric pmf for values of X from 0 to *x*.

Required Arguments

N The population size.

k The number of items in the category of interest.

n The sample size.

x The count of interest, which must be less than or equal to *n* and *k*.

Optional Argument

r The odds ratio.

Hypergeometric Probability(*N*, *k*, *n*, *x*, <*r*>)

Description

Returns the probability mass function (pmf) of the hypergeometric distribution. This is the probability that a hypergeometrically distributed random variable is equal to *x*. The pmf is parameterized as follows:

$$P(X = x;N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, n - x \le N - k$$

Required Arguments

N The population size.

k The number of items in the category of interest.

n The sample size.

x The count of interest, which must be less than or equal to *n* and *k*.

Optional Argument

r The odds ratio.

Neg Binomial Distribution(*p*, *n*, *k*)

Description

Returns the cumulative distribution function (cdf) of the negative binomial distribution. This is the probability that a negative binomially distributed random variable is less than or equal to *k*. The cdf is calculated as the summation of the negative binomial pmf for values of X from 0 to *k*.

Arguments

p The probability of success for each trial. *p* must be between 0 and 1.

n The number of successes.

k The number of failures before the n^{th} success.

Neg Binomial Probability(*p*, *n*, *k*)

Description

Returns the probability mass function (pmf) of the negative binomial distribution. This is the probability that a negative binomially distributed random variable is equal to *k*. The pmf is parameterized as follows:

$$P(X = k; p, n) = \binom{n + k - 1}{k} p^n (1 - p)^k$$

Arguments

p The probability of success for each trial. *p* must be between 0 and 1.

n The number of successes.

k The number of failures before the n^{th} success.

Notes

The return value of the pmf is the probability of observing the n^{th} success after *k* failures have occurred.

Poisson Distribution(*lambda*, *k*)

Description

Returns the cumulative distribution function (cdf) of the Poisson distribution. This is the probability that a Poisson distributed random variable with mean *lambda* is less than or

equal to *k*. The cdf is calculated as the summation of the Poisson pmf for values of X from 0 to *k*.

Arguments

k The number of events in a given time interval. *k* must be an integer.

lambda The shape parameter λ, which much be greater than 0. This is the mean of the distribution.

`Poisson Probability(lambda, k)`

Description

Returns the probability mass function (pmf) of the Poisson distribution. This is the probability that a Poisson distributed random variable with mean *lambda* is equal to *k*. The pmf is parameterized as follows:

$$P(X = k;\lambda) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$

Arguments

k The number of events in a given time interval. *k* must be an integer.

lambda The shape parameter λ, which must be greater than 0. This is the mean of the distribution.

`Poisson Quantile(lambda, cumprob)`

Description

Returns the smallest integer quantile for which the cumulative probability of the Poisson(lambda) distribution is larger than or equal to *cumprob*.

Arguments

lambda The shape parameter λ, which must be greater than 0. This is the mean of the distribution.

cumprob The cumulative probability of the quantile desired. *cumprob* must be between 0 and 1.

Display Functions

`Alpha Shape(Triangulation())`

Description

Returns the alpha shape for the given triangulation.

Border Box(<Left(*pix*)>, <Right(*pix*)>, <Top(*pix*)>, <Bottom(*pix*)>, <Sides(*Boolean*)>, *db*)

Description

Constructs a bordered display box that contains another display box. Optional arguments (Left, Right, Top, Bottom) add space between the border box and what it contains. The other optional argument (Sides) draws borders around the border box on any single side or combination of sides; draws the border in black or the highlight color; makes the background transparent or white or erases the background of a display box that contains it.

Returns

The display box.

Required Argument

db A display box object (for example, a text box or another border box).

Optional Arguments

Left(*pix*) An integer that measures pixels.

Right(*pix*) An integer that measures pixels.

Top(*pix*) An integer that measures pixels.

Bottom(*pix*) An integer that measures pixels.

Sides(*pix*) An integer that maps to settings for the display box.

Notes

The formula for deriving the integer for Sides is: 1*top + 2*left + 4*bottom + 8*right + 16*highlightcolor + 32*whitebackground + 64*erase. Thus, if you want to just draw a black border on the top and bottom, 1+4 = 5. If you want that same box with a white background, 5+32 = 37.

Box Plot Seg(<*data*>, <*frequency*>, <*weight*>, <Vertical(*Boolean*)>)

Description

Returns a display seg that represents a box plot based on the passed x and y values.

Returns

The display box (a box plot).

Optional Arguments

data The data values within the box plot.

frequency The frequency values within the box plot.

weight The weights for observations on continuous Ys.

Vertical(*Boolean*) A vertical (1) or horizontal (0) box plot.

Example

```
win = New Window( "Box Plot Seg Example",
Graph Box(
```

```
      Frame Size( 40, 180 ),
      Y Scale( 0, 100 ),
      Box Plot Seg(
         [20, 30, 40], // data
         [1, 1, 3], // frequencies
         [1, 1, 1], // weights
         1 // vertical
         )
      )
   );
```

Busy Light(< <<Automatic(*Boolean*)>, <Size(*x*, *y*)>, < <<Disable>)

Description

Creates a rotating image to indicate a busy process.

Returns

A rotating image.

Optional Argument and Messages

<<Automatic(*Boolean*) Rotates the image.

Size(*x*, *y*) Specifies the size of the image in pixels.

<<Disable Hides the image.

Example

```
win = New Window( "Example",
   blb = Busy Light( <<Automatic( 1 ), Size( 50, 50 ) ) );
```

Button Box(*title*, *script*, < *<<Set Icon(path)*>, < <<Set Icon Location(left|right)>

Description

Constructs a button with the text *title* that executes *script* when clicked.

Returns

The display box (button box).

Required Arguments

title A quoted string or a string variable.

script A quoted string or a reference to a string that specifies a valid JSL script.

Optional Messages

<<Set Icon("*path*") Displays the image in the quoted pathname on the button. Most common graphic formats are supported, such as GIF, JPG, PNG, BMP, TIF. Since the title argument is optional, you can create a button with only a text title, with only an icon, or with both a text title and an icon. In the last case, the icon is placed next to the text title.

<<Set Icon Location("left"|"right") Allows the position of the icon on a button to be either left or right of the text.

Example

The following example shows a simple button box. When the user clicks the button box, "Pressed" is printed to the log.

```
win = New Window( "Simple Example",
   ex = Button Box( "Press me" )
);
ex << Set Script( Print( "Pressed" ) );
```

Notes

Line-break characters are ignored in button boxes.

Calendar Box(*title*, < <<Date>, < <<Min Date>, < <<Max Date>, < <<Show Time>)

Description

Constructs a pop-up calendar with selectable days and time.

Returns

The display box (calendar box).

Required Argument

title A quoted string or a string variable.

Optional Messages

<<Date The currently selected date.

<<Min Date The earliest date that can be selected.

<<Max Date The latest date that can be selected.

<<Show Time The time that can be specified.

Example

The following example creates a calendar with October 5, 1989 initially selected. The minimum date and maximum date are specified, so the user can select only dates in that range.

```
New Window( "Calendar Box Example", cal = Calendar Box() );
date = Date MDY (10, 5, 1989);
cal << Date( date );
cal << Show Time( 0 ); // omit the time

/* earliest date that can be selected is 60 days before 10/5/1989
"start" truncates the value so the time is not included */
cal << Min Date( Date Increment(date, "Day", -60, "start" ) );

// latest date that can be selected is 60 days after 10/5/1989
cal << Max Date( Date Increment(date, "Day",  60, "start" ) );
```

```
cal << Set Function( Function( {this}, Print( Abbrev Date(this << Get Date())
  ) ) ); // print the abbreviated date to the log
```

Cell Plot(*Y(column(s))*, <*X(column)*>)

Description

Displays each value in a cell graph.

Check Box({*list*}, <*script*>, < <<Get(*n*)>, < <<Set(*n*, *Boolean*)>, < <<Get Selected>, < <<Enable Item(*n*, *Boolean*)>, < <<Item Enabled(*check box item*)>)

Description

Constructs a display box to show one or more check boxes.

Returns

The display box (Check Box).

Required Argument

list A list of quoted strings or a reference to a list of strings.

Optional Argument

script An optional JSL script.

Optional Messages

<<Get(*n*) Returns 1 if the check box item specified by *n* is selected, or 0 otherwise.

<<Set(*n*, *Boolean*) Sets the check box item specified by *n* as either selected (1) or cleared (0).

<<Get Selected Returns a list of quoted strings that contain the names of the check box items that are selected.

<<Enable Item(*n*, *Boolean*) Sets the check box item specified by *n* as either enabled (1) or disabled (0). The state of a disabled check box cannot be changed.

<<Item Enabled(*check box item*) Returns 0 or 1 depending on whether the specific check box item is enabled.

Example

Create three check boxes labeled “one”, “two”, and “three”. The first check box is selected.

```
New Window( "Example", Check Box( {"one", "two", "three"}, <<Set( 1, 1 ) ) );
```

Col Box(*title*, *display boxes*)

Description

Returns a column box made up of the specified display boxes.

Arguments

title The quoted title of the column.

display boxes Display boxes that hold content within the column box.

Example

```
win = New Window( "Example",
   exx = 1;
   exy = 4;
   exz = 8;
   Table Box(
      String Col Box( "strings", {"x", "y", "z"} ),
      Col Box(
         "boxes",
         Slider Box( 0, 10, exx, Show( exx ) ),
         Slider Box( 0, 10, exy, Show( exy ) ),
         Slider Box( 0, 10, exz, Show( exz ) )
      )
   );
);
```

`Col List Box(<Data Table(`*name*`)>, <"All"|"Character"|"Numeric">, <width(`*pixels*`)>, <"Grouped">, <MaxSelected(`*n*`)>, <nLines(`*n*`)>, <MaxItems(`*n*`)>, <MinItems(`*n*`)>, <On Change(`*expr*`)>, < <<Modeling Type("Any"|"Continuous"|"Ordinal"|"Nominal"|Multiple Response"|"Unstructured Text"|"Vector"|"None"|"Row State") >, < <<Set Data Type(Any|Numeric|Character)>, <`*script*`>)`

Description

Constructs a display box to show a list box that allows selection of data table columns.

Returns

The display box (Col List Box).

Optional Arguments

`Data Table(`*name*`)` The quoted name of the data table.

`"All" | "Character" | "Numeric"` Adds all columns of the current data table into the list. Omitting `"All"` results in an empty col list box with the "optional" label. To display "optional character", specify `"Character"`. To display "optional numeric", specify `"Numeric"`.

`width(`*pixels*`)` Sets the width of the list box to `pixels`. `pixels` is a number that measures pixels.

`"Grouped"` Displays grouped columns in the box.

`MaxSelected(`*n*`)` Sets whether only one item can be selected. For *n*>1, *n* is ignored.

`nLines(`*n*`)` Sets the length of the list box to n number of lines. *n* is an integer.

script A script.

`MaxItems(`*n*`)` A number that allows only *n* columns to be added to the list.

`MinItems(`*n*`)` A number that only requires at least *n* columns for the list. If *n*=2, the top two slots in the col list box an initial display of "required numeric" (or whatever you set the data type to be).

On Change(*expr*) Evaluates the expression when the selection in the list changes. Dragging between two column list boxes that have this argument results in both expressions being evaluated. The expression for the target being dragged is evaluated first, then the expression for the source is evaluated.

Optional Messages

<<Set Tips({*Tip text 1*, *Tip text 2*, ...}) Quoted strings that set tool tips for items in the list box. A quoted null string or an empty list results in no tips. A list shorter that the list of items in the list box will use the last tip text for the remaining items in the list and the list box.

<<Set Tip(*Tip text*) A quoted string that overrides any tool tips set using Set Tips() function. If there is a tip set for the box, you cannot set tips for each individual item.

Using Set Tip() with no arguments clears the list box tip and allows the individual item tool tips to be displayed.

Notes

- The MaxSelected(*n*) argument only affects whether one or more than one item can be selected. It does not enforce a limit greater than 1.
- Specialty modeling types can be used only in a role (determined by the platform) that explicitly accepts columns of the same type.

Col Span Box(*title*, *display box args*)

Description

Creates spanned columns headers inside a table box. The top column header spans two child column headers.

Returns

The display box (a Col Span Box).

Arguments

title The title that appears in the box.

display box args Display boxes.

Example

```
win = New Window( "Col Span Box",
   <<Modal,
   Table Box(
      Col Span Box(
         "Confidence Limits",
         neb = Number Col Edit Box( "Upper limits", [0, 0] ),
         Number Col Edit Box( "Lower limits", [0, 0] )
      )
   )
);
```

Combo Box(*{items <(tip string)>, ...}*, *<script>*)

Description

Constructs a display box to show a drop-down list.

Returns

The display box (Combo Box).

Arguments

items The items that the user can select.

tip string A quoted string that specifies tooltip text.

script An optional JSL script.

Context Box(*display box*, ...)

Description

Defines a scoped evaluation context. Each Context Box is executed independently of each other.

Returns

A display box.

Arguments

Any number of display boxes.

Contour Seg(Triangulation, [levels], <zColor([*colors*]|{*colors*}, <"Cycle Colors"|"Interpolate Colors">)>, <"Fill"|"Fill Between"|"Fill Below"|"Fill Above">, <Transparency([]|*t*)>)

Description

Returns a display seg that represents contours of a Triangulation.

Required Arguments

Triangulation The columns to include in the Triangulation.

[levels] A matrix of values that control the contour levels that are drawn.

Optional Arguments

zColor([*colors*]|{colors} Colors for each level, specified as a list or matrix.

"Cycle Colors"|"Interpolate Colors Cycle Colors" alternates the colors (for example, red, green, red, green). With Interpolate Colors, the first contour is red, and the last is green. The contours between smoothly blend the colors.

"Fill Below"|"Fill Between"|"Fill Above"|"Fill" "Fill Below" fills the region below the lines. "Fill Between" fills only the middle region. "Fill Above" fills the region above the lines.. "Fill" works like "Fill Above", but the default is to display lines if no Fill options are specified.

Transparency([]|*t*) The transparency specified as a number or matrix.

Example

```
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
tri = Triangulation( X( :X, :Y ), Y( :POP ) );
{xx, yy} = tri << Get Points();
win = New Window( "Contour Seg Example",
   g = Graph Box(
      X Scale( Min( xx ) - .1, Max( xx ) + .1 ),
      Y Scale( Min( yy ) - .1, Max( yy ) + .1 ),
      Contour Seg(
         tri,
         [0, 400, 1000, 2000, 9000],
         zColor( 5 + [64 32 0 16 48] ),
         Transparency( [1, 1, 1, 1, 1] )
      )
   )
);
```

Notes

The triangulation is computed using the Xs, and the Y is a continuous variable defined at each position. The [levels] in this case defines values of POP that are drawn as lines, one line per level. If any Fill argument is specified, then the filled regions are [level1, level2], [level2, level3], ..., [level-n].

Current Journal()

Description

Gets the display box at the top of the current (topmost) journal.

Returns

Returns a reference to the display box at the top of the current journal.

Current Report()

Description

Gets the display box at the top of the current (topmost) report window.

Returns

Returns a reference to the display box at the top of the current report window.

Current Window()

Description

Returns a reference to the current window.

Data Filter Context Box(*display box*)

Description

Returns a display box that defines the extent of the local data filters that a display tree contains. Data filters and Data Filter Context Boxes can be arranged in a hierarchy and shared among platforms or boxes that the Data Filter Context Boxes contain.

Data Filter Source Box(*display box*)

Description

Defines which graph is the "source" of the selection filter. Selected rows in reports that are within the Data Filter Source box are included for analysis in the other reports that are within a common Data Filter Context Box.

Data Table Box(*data table*)

Description

Returns a table box the represents the specified data table.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Example", Data Table Box( dt ) );
```

Data Table Col Box(*column*)

Description

Returns a column box that corresponds to the specified data table column.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Example",
   Table Box( Data Table Col Box( :name ), Data Table Col Box( :height ) )
);
```

Dialog(*contents*)

Description

`Dialog` is deprecated. Use `New Window()` with the `Modal` argument instead.

Excerpt Box(*report*, *subscripts*)

Description

Returns a display box containing the excerpt designated by the report held at number *report* and the list of display subscripts *subscripts*. The subscripts reflect the current state of the report, after previous excerpts have been removed.

Expr As Picture(*Expr*(...), <Width(*pixels*)>)

Description

Converts expr() to a picture as it would appear in the Formula Editor.

Returns

Reference to the picture.

Required Argument

Expr(...) Place any valid JSL expression that can be displayed as a picture inside expr().

Optional Argument

Width(*pixels*) Sets the width of the box to pix. pix is a number that measures pixels.

Filter Col Selector(<Data Table(*name*)>, <Width(*pixels*)>, <nLines(*n*)>, <*script*>, <OnChange(*expr*)>)

Description

Returns a display box that contains a list of items. Control allows column filtering.

Get Project(*title*|*index*|*display box*|*window*)

Description

Returns a single project.

Examples

The following examples show how to get the window title of various projects.

```
Get Project( 1 ) << Get Window Title;
// first open project
Get Project( "My Project" ) << Get Window Title;
// first project named “My Project”
Get Project( display box ) << Get Window Title;
// parent project of the specified display box
```

Get Project List()

Description

Returns a list of all open projects.

Example

```
Get Project List() << Get Window Title;
// list of the titles of all open projects
```

Get Window(<Project(*title*|*index*|*display box*|*window*)>, <Type(*string*)>, *title*|*index*|*display box*)

Description

Returns a reference to a specific open window by title, index, or display box. When run in a project, Get Window() returns windows from the current project.

Optional Arguments

Project Specifies the title (a quoted string), index, display box, or window from another project.

Type(*string*) To limit the search to particular types of windows, use the Type() argument and one of these quoted strings: "Data Tables", "Journals", "Reports", or "Dialogs".

Examples

The following examples show how to get the window title of various windows.

```
Get Window( 1 ) << Get Window Title;
// first window in the current project

Get Window( "Big Class" ) << Get Window Title;
// Big Class window in the current project

Get Window( ob ) << Get Window Title;
// parent window of specified display box in the current project

Get Window( Project(), 1 ) << Get Window Title;
// first window, no project (global scope)

Get Window( Project( myProject ), "Big Class" << Get Window Title;
// Big Class window in the specified project
```

Get Window List(<Project(*title*|*index*|*display box*|*window*),><Type(*string*)>)

Description

Returns a list of currently open windows. By default, Get Window List() returns references to open windows in the current project. You can return an open window list from something other than the current project by using the Project() argument. To limit the search to particular types of windows, use the Type () argument and one of these quoted strings: "Data Tables", "Journals", "Reports", or "Dialogs".

Optional Arguments

Project Specifies the title, index, display box, or window from another project.

Type To limit the search to particular types of windows, use the Type () argument and one of these quoted strings: "Data Tables", "Journals", "Reports", or "Dialogs".

Examples

```
Get Window List() << Get Window Title;
```

```
// list of open windows in the current project
Get Window List( Type( "Reports" ) ) << Get Window Title;
// list of the titles of open reports in the current project
Get Window List( Project( 0 ), Type( "Reports" ) ); // positional arguments
// list of the titles of open reports outside of a project
Get Window List( 2 );
// second window list
```

Global Box(*global*)

Description

Constructs a box for editing *global* value directly.

Graph()

See "Graph Box(properties, script)".

Graph 3D Box(*properties*)

Description

Constructs a display box with 3-D content.

Returns

The display box.

Arguments

properties Properties can include: Frame Size(x, y), Xname("title"), Yname("title"), Zname("title").

Notes

This display box constructor is experimental.

Graph Box(*properties*, *script*)

Graph(*properties*, *script*)

Description

Constructs a graph with axes.

Returns

The display box (Graph Box).

Arguments

properties Named property arguments: Title("title"), XScale(low, high), YScale(low, high), FrameSize(h, v), XName("x"), YName("y"), SuppressAxes.

script Any script to be run on the graph box.

H Center Box(<*child box*>)

Returns a display box that contains the optional child display box argument. The box is centered in the horizontal space defined by the maximum size of that child display box and all of the other siblings of the center box.

H List Box(<Align("center"|"bottom")>, display box, ...)

Description

Creates a display box that contains other display boxes and displays them horizontally.

Arguments

`Align("center"|"bottom")` Specify center or bottom alignment of the contents in the list box. The contents are bottom aligned by default.

display box Any number of display box arguments can be contained in the list box.

H Scroll Box(<Size(*h*)>, *display box*)

Description

Returns a display box that positions a larger child box using a horizontal scroll bar.

Arguments

`Size(`*h*`)` The horizontal length of the scroll bar.

Notes

The flexible argument is deprecated. Use `Set Stretch` instead. See "V Scroll Box(<Size(v)>, display box)" for an example.

H Sheet Box(<<Hold(*report*), *display boxes*)

Description

Returns a display box that arranges the display boxes provided by the arguments in a horizontal layout. The `<<Hold()` message tells the sheet to own the report(s) that is excerpted.

H Splitter Box(<Size(*h*, *v*)>, *display box*, ...)

Description

Returns a display box that arranges the display boxes provided by the arguments in a horizontal layout (or panel). The splitter enables the user to interactively resize the panel.

Required Argument

display box Any number of display box arguments can be contained in the splitter box.

Optional Argument

`Size(h, v)` Specifies the size of the splitter box in pixels. This size is for the outer splitter box. Inner display boxes are proportionately sized according to the width and height of the outer splitter box.

Optional Messages

`<<Size(n)` Specifies the proportions of the last panel. `<<Size(.25)` resizes the last panel to 25% the splitter box height (or width, for vertical splitter boxes).

`<<Set Sizes({n, n})` Specifies the proportions of each panel. `db<<Set Sizes({.75, .25})` sizes the first panel to 75% and the second panel to 25% of the splitter box height (or width, for vertical splitter boxes).

`<<Close Panel(n, <Boolean>)` Closes the panel that you specify. `<<Close Panel(2)` closes the second panel. With three or more panels, you must include the second Boolean value. That value indicates which panel expands to fill the space left by the closed panel.

- `<<Close Panel(2,0)` closes the second panel; the following sibling takes the extra space.
- `<<Close Panel(2,1)` closes the second panel; the preceding sibling takes the extra space.

`<<Open Panel(n, <Boolean>)` Opens the panel that you specify. With three or more panels, you must include the second Boolean value. Works similar to `<<Close Panel` described above. The panels are initially opened. You use `<<Open Panel` only after using `<<Close Panel`.

`<<Get Sizes()` Returns the proportions of each panel as a list.

Hier Box(text, Hier Box(...), ...)

Description

Constructs a node of a tree (similar to Diagram output) containing text. `Hier Box` can contain additional `Hier Boxes`, allowing you to create a tree. The text can be a quoted string or a `Text Edit Box`.

Hist Seg([*data*], <[*freq column*]>, <[*weight column*]>, <Vertical(*Boolean*)>, <Row States()>

Description

Returns a histogram seg.

Required Argument

data The data in matrix format.

Optional Arguments

freq column The frequency column in matrix format.

weight column The weight column in matrix format.

`Vertical(Boolean)` Displays the histogram vertically by default (or if set to 1). Display the histogram horizontally by setting the value to 0.

`Row States` Specifies a data table reference or row states.

`Icon Box(name)`

Description

Constructs a display box containing an icon, where the argument is a name such as Popup, Locked, Labeled, Sub, Excluded, Hidden, Continuous, Nominal, or Ordinal. The argument can also be a path to an image.

Argument

name A quoted string that is the name of a JMP icon or the path to an icon.

Example

`Icon Box( "Nominal" )` constructs a display box that contains the Nominal icon.

`Icon Box( "$SAMPLE_IMAGES/pi.gif" )` inserts the pi.gif sample image.

Notes

- Some icons are used on both Windows and macOS. Other icons are platform specific.
- Consider installing the Built-In JMP Icons add-in. The add-in lets you view icons interactively and see what they look like in different contexts. Download the add-in from https://community.jmp.com/t5/JMP-Add-Ins/Built-In-JMP-Icons/ta-p/42251.

`If Box(Boolean, display boxes)`

Description

Constructs a display box whose contents are conditionally displayed.

Arguments

Boolean 1 displays the display boxes inside the `If Box`. 0 does not display them.

display boxes Any display box tree.

`If Seg(<State(Boolean)>)`

Description

Returns a display seg that shows or hides display seg children.

Arguments

`State(Boolean)` Determines whether the display seg children are shown (1) or hidden (0).

Example

```
lines = [30 20 80 70, 10 90 90 10, 40 20 60 30];
win = New Window( "Lines Seg Example",
   g = Graph Box( If Seg( true, <<Append( Lines Seg( lines ) ) ) )
```

```
);
```

`Journal Box(string)`

Description

Constructs a display box that displays the quoted string. We recommend that you do not generate the journal text by hand.

`Line Seg(x, y, <Row States(dt|dt, [rows]|dt, {{rows}, ...}|{row states})>, <Sizes(s)>)`

Description

Creates a display seg of connected line segments for the given *x* and *y* values. The optional third argument enables row state assignments from either a data table or independently.

`Lines Seg([x1 y1 x2 y2, ...])`

Description

Returns a display seg with a sequence of line segments for the given *x* and *y* values.

`Lineup Box(<nCol(n)>, <Spacing(pixels, <vspace>), display boxes, ...)`

Description

Constructs a display box to show an alignment of boxes in *n* columns.

`List Box({item, ...}, <Width(pixels)>, <maxSelected(n)>, <nLines(n)>, <script>)`

Description

Creates a display box to show a list box of selection items (quoted strings). The argument can be a list of two-item lists containing the item name and a quoted string that specifies the modeling type or sorting order. Item names are case sensitive by default. The icon appears next to the corresponding item in the list box.

`Marker Seg(x, y, <Row States(dt|dt, [rows]|dt, {{rows}, ...}|{row states})>, <Sizes(s)>)`

Description

Creates a display seg with markers for all of the x and y values. The optional third argument enables row state assignments from either a data table or independently.

Matrix Box(*x*)

Matrix Box(*matrix*, < <<Column Names(*col1*, *col2*, ...)>, < <<Row Names(*row1*, *row2*, ...)>>)

Description

Displays the *matrix* given in the usual array form. Column and row names are quoted strings.

Mouse Box(*display box arguments*)

Description

Returns a box that can make JSL callbacks for dragging and dropping, marking, or clicking and tracking mouse actions.

Arguments

display box arguments Specifies the object that the user interacts with, such as a Text Box or Button Box. See the Scripting Index in the Help menu.

Move to Project(<Source(*project*)|Destination(*project*)>, <Windows({*list of windows to move*})>)

Description

Moves one or more windows into a project or out of a project, or between projects.

Arguments

Source(*project*) The project containing the windows that you want to move.

Destination(*project*) The project to which you want to move the windows.

Windows({*list of windows to move*}) A list of windows to move to the project. If omitted, all windows will be moved. Note that the data table and all of its dependent reports will be moved. However, you need to specify only the data table name or report name in the windows argument to move it.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
report = dt << Run Script( "Bivariate" );
project = New Project();
// move the report and data table to a new project
Move to Project( Destination( project ), Windows( {report} ) );
```

New Image()

New Image(*width*, *height*)

New Image(*filepath*)

New Image(Open(*url*))

New Image(*picture*)

New Image(*matrix*)

New Image(rgb|r|rgba, {*matrix*, ...})

Description

Creates a new image which can then be edited using JSL. Valid formats are png, bmp, jpeg, jpg, tiff, tif, and gif.

Returns

An image.

Arguments

All argument sets are optional, but all arguments within each set are required.

width, height Sets the width and height of the image in pixels.

filepath The quoted filepath to an image.

Open(*url*) Opens the image at the specified URL path.

picture A JSL picture object.

matrix The image as a matrix of JSL color pixels.

rgb|r|rgba, {*matrix*, ...} Specifies the channels ("rgb", "r", or "rgba") and provides a matrix of values (0.0-1.0) for each channel. Examples:

```
New Image( "r", [r matrix] );
New Image( "rgb", {[r matrix], [g matrix], [b matrix]} );
New Image( "rgba", {[r matrix], [g matrix], [b matrix], [a matrix]} );
```

New Project(*arguments*)

Description

Creates a project using the specified script.

Arguments

<<Add Bookmarks({<File(*path*)>, <Folder(*path*, Expanded(*Boolean*)>)>, <Group(*name*, <Expanded(*Boolean*)>, {*contents*}>) Creates bookmarks for frequently used files in the project. The argument is a list of bookmark items, each of which is specified using File(), Folder(), or Group(). Group() accepts File(), Folder(), and Group() as children.

<<Reset Layout Sets the project to use the default layout.

<<Run Script Specifies the data tables and reports that appear in the project.

<<Save(*<path>*) Saves the project. Include a quoted path and file name to save the project to a specific location. Save As is an alias.

<<Set Bookmarks({<File(*path*)>, <Folder(*path*, Expanded(*Boolean*)>)>, <Group(*name*, <Expanded(*Boolean*)>, {*contents*}>) Sets the bookmarks for the project. The argument is a list of bookmark items, each of which is specified using File(), Folder(), or Group(). Group() accepts File(), Folder(), and Group() as children.

<<Set Layout Sets the window layout of the project.

<<Show Bookmarks Shows or hides the bookmarks.

<<Show Log Shows or hides the log.

<<Show Window List Shows or hides the Window List.

Example

The following example creates a project from BigClass.jmp and two reports.

```
project = New Project();
project << Run Script(
   dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
   dt << Run Script( "Bivariate" );
   dt << Run Script( "Distribution" );
);
```

New Window(*title*, *<arguments>*, *display box*)

Description

Makes a new window with the required quoted title and a display box tree.

Optional Arguments

<<Script(*<script>*) Creates a new script window. The optional quoted string *script* is placed inside the script window.

<<Journal Creates an empty journal.

<<Size Window(*x*, *y*) Creates a new window of the specified height and width.

<<Modal Makes the new window a modal window, which prevents any other actions in JMP until the window is closed. If you do not include an **OK** or **Cancel** button, one is added automatically for you. **Note:** If used, this argument must be the second argument, directly after the window title. Available modal window arguments are:

– <<On Open(*expr*) runs expr when the window is created.

Note: In data tables, On Open (or OnOpen) scripts that execute other programs are never run. Set the Evaluate OnOpen Scripts preference to control when the script is run.

– <<On Close(*expr*) runs expr when the window is closed. Returns 0 if the window fails to close.

- – `<<On Validate(expr)` runs `expr` when the **OK** button is pressed. If it returns True, the window is closed otherwise the window remains open.
- – `<<Return Result` changes the window's return value when it closes to match that of the deprecated `Dialog()` function.

`Show Toolbars(Boolean)` Shows or hides the toolbar. The default value is 1. (Windows only.)

`Show Menu(Boolean)` Shows or hides the menu bar. The default value is 1. (Windows only.)

`Suppress AutoHide(Boolean)` Suppresses or uses the auto-hide feature for menus and toolbars. The default value is 1. (Windows only).

Notes

`Dialog()` is deprecated. Use `New Window()` with the `Modal` argument instead.

`Number Col Box(title, numbers)`

Description

Creates a column named *title* (a quoted string) with numeric entries given in list or matrix form.

`Number Col Edit Box(title, {numbers}|[numbers])`

Description

Creates a column named *title* (a quoted string) with numeric entries given in list or matrix form. The numbers can be edited.

`Number Edit Box(initial value, <width>)`

Description

Creates an editable number box that initially contains the *initial value* argument.

Returns

The display box object.

Argument

initial value Any number to use as the initial value. If you use a date or time format, a date and time selector window is created.

<width> Sets the width of the box in characters.

`Outline Box(title, display box, ...)`

Description

Creates a new outline named *title* (a quoted string) containing the listed display boxes.

Page Break Box()

Description

Creates a display box that forces a page break when the window is printed.

Panel Box(*title*, *display box*)

Description

Creates a display box labeled with the quoted string *title* that contains the listed display boxes.

Picture Box(Open(*picture*), *format*)

Description

Creates a display box that contains a graphics picture object.

Returns

A reference to the display box.

Argument

Open(*picture*) Opens the directory that contains the picture.

format Specifies the graphic file format. Specifying the format opens the picture in JMP. If you omit this argument, the picture opens in the default graphics program. Valid formats are the quoted strings "png", "bmp", "jpeg", "jpg", "tiff", "tif", and "gif".

Example

```
New Window( "Example",
   Picture Box( Open( "$SAMPLE_IMAGES/pi.gif", gif ) ) );
```

Platform(*data table*, *script*)

Description

Evaluates the specified script in the context of the specified data table.

Returns

The resulting display box for embedding in a display tree.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
win = New Window( "Platform example",
   H List Box(
      Platform(
         dt,
         Bubble Plot(
            X( :weight ),
            Y( :height ),
```

```
            Sizes( :age ),
            Title Position( 0, 0 )
         )
      ),
      Platform(
         dt,
         Bubble Plot(
            X( :weight ),
            Y( :age ),
            Sizes( :height ),
            Title Position( 0, 0 )
         )
      )
   )
);
```

Plot Col Box(*title*, *numbers*)

Description

Returns a display box labeled with the quoted string *title* to graph the *numbers*. The numbers can be either a list or a matrix.

Poly Seg(*x values*, *y values*)

Description

Returns a display seg that represents a polygon with vertices based on the *x values* and *y values*.

Example

```
x = [10, 50, 90];
y = [10, 90, 10];
win = New Window( "Poly Seg Example",
g = Graph Box( Poly Seg( x, y ) ) );
frame = g[FrameBox( 1 )];
seg = (frame << Find Seg( "Poly Seg" ));
```

Popup Box({*command1*, *script1*, *command2*, *script2*, ...})

Description

Creates a red triangle menu. The single argument is an expression yielding a list of an even number of items alternating between the quoted command string and the expression that you want evaluated when the command is selected. If the command is an empty string, a separator line is inserted.

Notes

Pressing ALT and right-clicking the red triangle menu opens a window with check boxes for the commands.

`Radio Box({item, ...}, <script>)`

Description

Constructs a display box to show a set of radio buttons. The items are quoted strings. The optional script is run every time a radio button is selected.

`Range Slider Box(minValue, maxValue, lowVariable, highVariable, script)`

Description

`Range Slider Box()` returns a display box that shows a range slider control that ranges from *minValue* to *maxValue*. As the two sliders' positions change, their values are placed into *lowVariable* and *highVariable*, and the script is run.

Returns

The display box (Range Slider Box).

Arguments

minValue, *maxValue* Numbers that set the minimum and maximum value the slider represents.

lowVariable The variable whose value is set and changed by the lower slider.

highVariable The variable whose value is set and changed by the upper slider.

script Any valid JSL commands that are run as the slider is moved.

`Report(obj)`

Description

Returns the display tree of a platform *obj*. This can also be sent as a message to a platform: `obj<<Report.`

`Scene Box(x size, y size)`

Description

Creates an *x*-sized by *y*-sized scene box for 3-D graphics.

`Scene Display List`

Description

Returns a display list for 3-D graphics.

Example

```
ex = Scene Display List();
```

```
ex << Color( .9, .9, .9 );
ex << Text( center, middle, .3, "Hello World" );
exScene = Scene Box( 600, 600 );
exScene << Background Color( 0 );
exScene << Show Arcball( always );
New Window( "See HelloWorld.jsl in sample scripts", exScene );
exScene << Perspective( 45, .2, 20 );
exScene << Translate( 0.0, 0.0, -4.5 );
exScene << Arcball( ex, 1.5 );
exScene << Update;
```

Script Box(<*script*>, <*language*>, <*width*>, <*height*>)

Description

Constructs an editable box that contains the quoted string *script*. The editable box is a script window and can both be edited and run as JSL.

Optional Arguments

script A quoted string that appears in the script box.

language A quoted string that provides syntax highlighting for the specified language. Valid languages are "JSL", "Text", "SAS", "SAS Output", "SASLog", "R", "MATLAB", "JavaScript", "C", "SQL", "Python", "JSON", and "XML".

width An integer that sets the width of the script box.

height An integer that sets the height of the script box.

Example

```
// JSON
New Window( "JSON",
   Script Box(
      "{\!"a\!":1,\!"b\!":\!"test\!"}",
      "JSON"
   )
);
```

Scroll Box(<Size(*h*,*v*)>, *display box*, ...)

Description

Creates a display box that positions a larger child box using scroll bars.

Returns

A reference to the scroll box object.

Required Argument

display box Any number of display box arguments can be contained in the scroll box.

Optional Argument

Size(*h*,*v*) The *h* and *v* arguments specify the size of the box in pixels.

Notes

You can send a scroll box object a message to set the background color:

```
<<Set Background Color( {R, G, B} | <color> )
```

The `Flexible` argument is deprecated. Use the `Set Stretch` message instead. See "V Scroll Box(<Size(v)>, `display box`)" for an example.

You can set the Boolean flags for horizontal (h) and vertical (v) scrolling to enable (1) or disable (0) the scroll bars. If scrolling is disabled in a given direction, the Scroll Box will behave as a regular container in that direction.

```
<<Set Scrollers (h, v)
```

To return the flags for scrolling, use the following message:

```
<<Get Scrollers
```

To set the horizontal (h) and vertical (v) positions (in pixels) for the scrollers on the scroll bar:

```
<<Set Scroll Position (h,v)
```

To return the flags for scroll position, use the following message:

```
<<Get Scroll Position
```

To return the maximum positions for horizontal and vertical scrolling, use the following message:

```
<<Get Scroll Extents
```

Example

The following example shows a window containing a scroll box with the specified settings.

```
win = New Window( "Example",
   sb = Scroll Box(
      Size( 150, 75 ),
      List Box(
         {"First Item", "Second Item",
         "Third Item", "Fourth Item",
         "Fifth Item"},
         width( 200 ),
         max selected( 2 ),
         nLines( 6 )
      )
   )
);
win << Set Window Size( 300, 200 );
sb << Set Scrollers( 1, 1 ); // enable both scroll bars
// position the scrollers on the scroll bar
sb << Set Scroll Position( 0, 20 );
```

Shape Seg({Path(*path*), ...}, <Row States(*dt*|*dt*,[*rows*]|dt,{{*rows*}, ...}|{*row states*})>)

Description

Returns a display seg with a collection of shapes.

Required Argument

`Path` Specifies the path with an Nx3 matrix or with a text representation. A path matrix has three columns for x, y, and flags for each point in the path. The flag values are 0 for control, 1 for move, 2 for a line segment, 3 for a cubic Bézier segment, and are negative if the point also closes the path. Path text supports SVG syntax.

Optional Arguments

`Row States` Specifies a data table reference and optionally rows or the actual row state.

Example

```
win = New Window( "Shape Seg Example",
   Graph Box(
      Shape Seg(
         {Path( [10 10 1, 10 70 0, 70 70 0, 70 10 -3] ),
         Path( "M20,20 C20,60 60,60 60,20 Z" )},
         Row States( {Selected State( 1 ), Color State( "red" )} )
      )
   )
);
```

Sheet Box(<<Hold(*rpt*), *display box*, ...)

Description

Returns a display box that can organize other display boxes vertically or horizontally.

Sheet Panel Box(*title*, *display box*)

Description

Specifies whether the Sheet Box should be horizontal or vertical.

Slider Box(*minValue*, *maxValue*, *variable*, *script*, <Set Width(*n*)>, <Rescale Slider(*min*, *max*)>)

Description

Creates an interactive slider control.

Returns

The display box (Slider Box).

Required Arguments

minValue, *maxValue* Numbers that set the minimum and maximum value the slider represents.

variable The variable whose value is set and changed by the slider box.

script Valid JSL commands that are run as the slider box is moved.

Optional Arguments

Set Width(*n*) Specifies the width of the slider box in pixels.

Rescale Slider(*minValue*, *maxValue*) Resets the maximum and minimum values for the slider box.

Notes

You can send Set Width and Rescale Slider as messages to a slider object. For example:

```
ex = .6;
New Window( "Example", mybox = Slider Box( 0, 1, ex, Show( ex ) ) );
mybox << Set Width( 200 ) << Rescale Slider( 0, 5 );
```

Spacer Box(<Size(*h*, *v*)>, <Color(*color*)>)

Description

Creates a display box that can be used to maintain space between other display boxes, or to fill a cell in a LineUp Box.

Returns

A reference to the display box.

Optional Arguments

Size(*h*, *v*) The *h* and *v* arguments specify the size of the box in pixels.

Color(*color*) Sets the color of the box to the JSL color argument.

Spin Box(*script*)

Description

Returns a display box that shows a button with up and down controls.

Argument

script Invoked with an argument that indicates the direction of the arrow clicked. Negative is down, and positive is up. A value of 1 indicates a single click, while larger values indicate a repeating action.

Example

```
win = New Window( "Example",
   Lineup Box(
      2,
      nb = Number Edit Box( 3 ),
      sb = Spin Box( Function( {value}, nb << Increment( value ) ) )
   )
);
nb << Set Increment( 1 );
```

Splitter Box(<Size(*x*, *y*)>, *display box*, ...)

Description

Returns a display box that can organize other display boxes horizontally or vertically with interactive control of sizes. Child sizes are specified as a proportion of the width or height of the splitter box. The optional *Size* argument is used only for the top-most splitter box. Lower level display boxes are sized like any other child box.

Use `H Splitter Box()` or `V Splitter Box()`.

String Col Box(*title*, {*string*, ...})

Description

Creates a column in the data table containing the quoted *string* items. The column is named after the quoted *title*.

String Col Edit Box(*title*, {*string*, ...})

Description

Creates a column in the data table containing the quoted *string* items. The string boxes are editable. The column is named after the quoted *title*.

Notes

To retrieve the data, use the following message:

```
data = obj << Get;
```

Tab Box(Tab Page Box(Title(*page title 1*), <*options*>, *contents of page 1*), Tab Page Box(Title(*page title 2*), <*options*>, *contents of page 2*), ...);

Description

(Previously called `Tab List Box`.) Creates a tabbed window pane. The arguments are an even number of items alternating between the name of a tab page and the contents of the tab page. The page titles must be quoted.

Arguments

`Tab Page Box` Returns a display box that can be used in a tab box or as a stand-alone container with a title.

- `Title(page title #)` specifies the title of page 1.
- Options include <<`Closeable`(*Boolean*) (specifies whether the box can be closed), <<`Tip`(*string*) (a quoted string that specifies a tooltip), and <<`Icon`(*string*) (a quoted string that specifies the icon).
- *contents of page #* is a quoted string that specifies the text on the tab.

Example

```
New Window( "Example",
```

```
   Tab Box(
      t1 = Tab Page Box( Title( "alpha" ), Panel Box( "panel", Text Box( "text"
   ) ) ),
      t2 = Tab Page Box( Title( "beta" ), Popup Box( {"x", ex = 1, "y", ex = 2}
   ) ),
   )
);
```

Notes

Certain messages that you can send to `Tab Page Box` have been renamed:

- `Set Title` is called `Title`.
- `Set Tip` is called `Tip`.
- `Set Icon` is called `Icon`.
- `Set Closeable` is called `Closeable`.

Tab List Box(*title*, *tabExpr1*, ...)

See "Tab Box(Tab Page Box(Title(page title 1), <options>, contents of page 1), Tab Page Box(Title(page title 2), <options>, contents of page 2), ...);".

Tab Page Box(Title(*string*), <*options*>, *contents*)

Description

Returns a display box that can be used in a tab box or as a stand-alone container with a title.

Required Argument

`Title` A quoted string that specifies the title on the tab.

Optional Messages

`<<Tip(`*string*`)` A quoted string that specifies a tooltip.

`<<Closeable(`*Boolean*`)` Specifies whether the page can be closed.

`<<Icon(`*string*`)` A quoted string that specifies the icon.

`<<Set Font(font name, <size>, <"bold"|"italic"|"underline"|strikeout">, <angle>` Specifies the font properties.

`<<Set Font Name("`*font name*`")` Specifies the name of the font.

`<<Set Font Scale(f)` Specifies the scale factor for the font. The scale factor is applied to the size that is determined by the base font and point size.

`<<Set Font Size(n)` Specifies the font size in pixels.

`<<Set Font Style("plain"|"bold")` Specifies the font style.

Notes

Certain messages that you can send to `Tab Page Box` have been renamed:

- `Set Title` is called `Title`.

- `Set Tip` is called `Tip`.
- `Set Icon` is called `Icon`.
- `Set Closeable` is called `Closeable`.

`Table Box(`*`display box, ...`*`)`

Description

Creates a report table with the *`display boxes`* listed as columns.

`Text Box(`*`text, <arguments>`*`)`

Description

Constructs a box that contains the quoted string *`text`*.

Arguments

`<<Justify Text(`*`position`*`)` Justifies the text left, center, or right as specified in quotes.

`<<Set Wrap(`*`pixels`*`)` Sets the point at which text wraps.

`Text Edit Box(`*`text, <arguments>`*`)`

Description

Constructs an editable box that contains the quoted string *`text`*.

Arguments

`<<Password Style(`*`Boolean`*`)` Displays asterisks in the box rather than the password.

`<<Set Script` Runs the specified script after the text is edited.

`<<Set Width(`*`pixels`*`)` Sets the point at which text wraps.

`This Project()`

Description

Gets the current project when a JSL script is run from that project.

Example

The following example gets the window title of the current project.

```
project = New Project();
project << Save( "$DOCUMENTS/Test Project.jmpprj" );
project << Run Script(
   New Window( "Project Title",
      Text Box(This Project() << Get Window Title())
   );
);
```

Tree Box(<{*rootnodes*}>, <Size(*width*, *height*)>, <MultiSelect(*Boolean*)>)

Description

Constructs a box to show a hierarchical tree composed of tree nodes.

Arguments

{*rootnodes*} Specifies the names for the root nodes created by `Tree Node()` which the box contains.

`Size(`*width*, *height*`)` Specifies the width and height (in pixels) of the box.

`MultiSelect(`*Boolean*`)` Indicates that more than one item in the tree can be selected.

Tree Node(<*data*>)

Description

Creates a node for display in a Tree Box display. Tree Node is used for both parent and child nodes.

Notes

If you send a root node that contains one or more nodes with the `Set Node Select Script` defining a collapse message, then macOS runs the script twice. Windows doesn't run the script. This behavior on macOS doesn't just affect increments. Any script runs twice. It will print to the log twice, create a column twice, try to delete something twice, and so on.

Triangulation(<*dt*>, X(*col1*, *col1*), <Y(*col*)>)

Description

Returns an object containing the Delaunay triangulation of the given point set. The optional Y will be averaged for duplicate points, and all points in the output will be unique.

Examples

```
tri = Triangulation(
  X( [0 0 1 1], [0 1 0 1] ),
  Y( [0 1 2 3] )
);
dt = Open( "$SAMPLE_DATA/Cities.jmp" );
tri = Triangulation( X( :X, :Y ), Y( :POP ) );
```

V Center Box(*display box...*)

Returns a display box that contains the child display box argument. The box is centered in the vertical space defined by the maximum size of that child display box and all of the other siblings of the center box.

V List Box(<Align("Center"|"Right")>, *display box, ...*)

Description

Creates a display box that contains other display boxes and displays them vertically.

Arguments

`Align` Specify right or center alignment of the contents in the list box. The contents are center aligned by default.

`display box` Any number of display box arguments can be contained in the list box.

V Scroll Box(<Size(*v*)>, *display box*)

Description

Returns a display box that places a scroll bar on the bottom and right if the contents are bigger than the size of the scroll box.

Arguments

`Size(`*`v`*`)` The vertical length of the scroll bar.

`display box` Any number of display box arguments can be contained in the scroll box.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
New Window( "stretchable",
   H Splitter Box(
      Size( 400, 200 ),
      Scroll Box(
         Size( 200, 200 ),
         dt <<Run Script( "Distribution" ),
         <<Set Stretch( "Window", "Window" )
      ),
      Scroll Box(
         Size( 200, 200 ),
         dt <<Run Script( "Bivariate" ),
         <<Set Stretch( "Window", "Window" )
      ),
      <<Set Stretch( "Window", "Window" )
   )
);
```

Notes

The flexible argument is deprecated. Use `Set Stretch` instead.

V Sheet Box(<<Hold(*report*), *display box, ...*)

Description

Returns a display box that arranges the display boxes provided by the arguments in a vertical layout. The <<Hold() message tells the sheet to own the report(s) that is excerpted.

V Splitter Box(<Size(*h*,*v*)>, *display box...*, <*arguments*>)

Description

Returns a display box that arranges the display boxes provided by the arguments in a vertical layout (or panel). The splitter enables the user to interactively resize the panel.

Arguments

Size(*v*) The vertical length of the scroll bar.

display box Any number of display box arguments can be contained in the splitter box.

Optional Arguments

For more information about the optional arguments, see "H Splitter Box(<Size(h, v)>, display box, ...)".

Web Browser Box(*url*)

Description

Creates a display box that contains a web page. Available only on Windows in Internet Explorer.

Returns

A reference to the web browser box object.

Argument

url A quoted string containing the URL to the web page to display.

Example

The following example creates a splitter box with the web browser box on the left and the bubble plot on the right.

```
dt = Open( "$SAMPLE_DATA/PopAgeGroup.jmp" );
New Window( "Example",
  H Splitter Box(
     Size( 800, 300 ),
     wb = Web Browser Box( "http://www.jmp.com" ),
     dt << Run Script( "Bubble Plot Region" )
  )
);
wb << Set Stretch( "Window", "Window" ); // auto stretch horizontally and
  vertically
wb << Set Max Size( 10000, 10000 ); // maximum size in pixels
```

Notes

The <a href> target "_blank" opens the web page in a new Internet Explorer window. The <a href> target "_new" opens the web page in the active Internet Explorer tab.

`Window(<string|int>)`

Returns

Either a list of references to all open windows, or a reference to an explicitly named window.

Arguments

string A quoted string containing the name of a specific open window.

int The number of a specific open window.

Notes

If no argument is provided, a list of all open windows is returned.

If the argument (either a window name or number) does not exist, an empty list is returned.

`Wrap List Box(display box, ...)`

Description

Creates a list box that contains other display boxes and displays them horizontally, but wraps them when printing.

Arguments

display box Any number of display box arguments can be contained in the list box.

Expression Functions

`Arg(expr, i)`

`Arg Expr(expr, i)`

Description

Finds the argument numbered by *i* within the given expression.

Returns

The *i*th argument within the expression *expr*.

`Empty()` if that argument does not exist or is not specified.

Arguments

expr An expression.

i An integer that indicates which argument to return.

Notes

`Arg Expr()` is deprecated. Use `Arg()` instead.

`Eval Expr(`*expr*`)`

Description

Evaluates any expressions within *expr*, but leaves the outer expression unevaluated.

Returns

An expression with all the expressions inside *expr* evaluated.

Argument

expr A valid expression.

`Expr(`*expr, i*`)`

Description

Returns the argument unevaluated (expression-quoting).

Returns

The argument, unevaluated.

`Empty()` if that argument does not exist or is not specified.

Argument

expr An expression.

i An integer that indicates which argument to return.

`Extract Expr(`*expr*`, `*pattern*`)`

Description

Find *expr* matching *pattern*.

Returns

A pattern that matches the specified *pattern*.

Arguments

expr An expression.

pattern Any pattern.

`Head(`*exprArg*`)`

`Head Expr(`*exprArg*`)`

Description

Returns the head of the evaluated expression, without its arguments.

Notes

`Head Expr()` is deprecated. Use `Head()` instead.

`Head Name(expr)`

`Head Name Expr(expr)`

Description

Returns the head of the evaluated expression as a quoted string.

Notes

`Head Name Expr()` is deprecated. Use `Head Name()` instead.

`N Arg(exprArg)`

Description

Returns the number of arguments of the evaluated expression head.

`N Arg Expr(exprArg)`

Description

Returns the number of arguments of the expression head.

Notes

`N Arg Expr()` is deprecated. Use `N Arg()` instead.

`Name Expr(x)`

Description

Returns the unevaluated expression of *x* rather than the evaluation of *x*.

File Functions

`Close(<dt|query>, <noSave|Save(path)>)`

Description

Closes a data table, query, or JSON file. If no arguments are specified, the current file is closed. If the file has been changed, it is automatically saved. All dependent windows are also closed (for example, report windows that are based on the data table).

Returns

Void.

Arguments

dt An optional reference to a data table, query, or JSON file.

`nosave|Save(path)` An optional switch to either save the file to the specified quoted path before closing or to close the file without saving it.

`Close All(<Project(`*`title`*`|`*`index`*`|`*`display box`*`|`*`window`*`)>, `*`type`*`, <"Invisible"|"Private">, <NoSave|Save>)`

Description

Closes all open resources of *type*.

Required Argument

type A named argument that defines the type of resources that you want to close. The allowable types are: `Data Tables`, `Reports`, and `Journals`.

Optional Argument

`Project(`*`title`*`|`*`index`*`|`*`display box`*`|`*`window`*`)` Closes the specified project window.

`"Invisible"` Closes all invisible data tables.

`"Private"` Closes all private data tables.

`NoSave or Save` Saves the specified types of windows before closing or to close without saving.

`Close Database Connection(`*`db connection handle`*`)`

Description

Closes a database connection returned from Create Database Connection.

`Close Log(`*`Boolean`*`)`

Description

Closes the log window.

`Convert File Path(`*`path`*`, <"Absolute"|"Relative">, <"POSIX"|"Windows">, <Base(`*`path`*`)>, <`*`Search`*`>)`

Description

Converts a file path according to the arguments.

Returns

The converted path.

Required Argument

path A quoted pathname that can be either Windows or POSIX.

Optional Arguments

`"Absolute"|"Relative"` Specifies whether the returned pathname is in absolute or relative terms. The default value is `Absolute`.

`"POSIX"|"Windows"` Specifies whether the returned pathname is in Windows or POSIX style. The default is `POSIX`.

`Base(`*`path`*`)` Specifies the quoted base pathname, useful if `Relative` is specified. The default is the default directory. See "Set Default Directory(path)".

`Search` Searches through the specified directories for the specified string. In the example below, JMP searches $SAMPLE_DATA/ and $SAMPLE_DATA/Time Series/ for Air.jmp. If Air.jmp is in both directories, the first instance is returned. Without the search option, `Convert File Path()` uses the default directory instead.

Example

```
Set File Search Path(
   {Convert File Path( "$SAMPLE_DATA/" ),
   Convert File Path( "$SAMPLE_DATA/Time Series/" )}
);
Convert File Path( "Air.jmp", Search );
   Get File Search Path() = {"/C:/Program Files/SAS/JMPPRO/16/Samples/Data/",
      "/C:/Program Files/SAS/JMPPRO/16/Samples/Data/Time Series/"};
   Convert File Path("Air.jmp", search) = "/C:/Program
      Files/SAS/JMPPRO/16/Samples/Data/Time Series/Air.jmp";
```

Copy Directory(*from path*, *to path*, <Recursive(*Boolean*)>)

Description

Copies files from one directory to another, optionally copying subdirectories. The directory name is created in the *to path* and should not be part of the *to path* argument.

Returns

Returns 1 if the directory is copied; otherwise, returns 0.

Required Arguments

from path Specifies the directory containing the files to copy to the new directory. *from path* is a quoted string.

to path Specifies the path where the new directory should be created and to which the files are copied. *to path* is a quoted string.

Optional Argument

`Recursive(`*Boolean*`)>` Specifies whether to copy the *from path* subdirectory structure to the *to path*.

Notes

`Copy Directory(`*from path*, *to path*, *Boolean*`)` is deprecated.

Copy File(*from path*, *to path*)

Description

Copies one file to a new file using the same or a different name.

Returns

Returns 1 if the file is copied; otherwise, returns 0.

Arguments

from path Specifies the path and file name to copy to the new file. *from path* is a quoted string.

to path Specifies the path and file name for the new file. *from path* is a quoted string.

`Create Database Connection(`*dataSourceName*`|"Connect Dialog",<DriverPrompt(`*Boolean*`)>);`

Description

Creates a connection to the specified database or prompts the user to provide database log in information.

Returns

A handle to the database connection.

Required Argument

dataSourceName The quoted server connection string that contains information such as the data source name and user name.

Optional Arguments

`"Connect Dialog"` Opens the Select Data Source window, from which the user selects the database.

`Driver Prompt(`*Boolean*`)` Enables the ODBC driver to prompt for the connection information if necessary.

Examples

Specify the data source name, user name, and password:

```
Create Database Connection( "dsn=Books;UID=johnsmith;password=Christmas" );
```

Request that the ODBC driver prompt the user to enter connection information, because the quoted connection string does not specify the password:

```
Create Database Connection( "dsn=Books;UID=johnsmith", Driver Prompt( 1 ) );
```

Enable the user to select the data source, specify `"Connect Dialog"`:

```
Create Database Connection( "Connect Dialog" );
```

`Create Directory(`*path*`)`

Description

Creates a new directory at the specified path location.

Returns

Returns 1 if the directory is created; otherwise, returns 0.

Arguments

path Specifies the quoted path where the new directory should be located.

Creation Date(*path*)

Description

Returns the creation date for the specified file or directory.

Returns

Creation date.

Arguments

path Specifies the quoted directory or path and file name for the query.

Delete Directory(*path*, <Allow Undo(*Boolean*)>)

Description

Deletes the specified directory and its contents and any subdirectories.

Returns

Returns 1 if the directory is deleted; otherwise, returns 0.

Arguments

path Specifies the path and directory for deletion.

Allow Undo(*Boolean*) Allows undo operations, for example, moving to the Recycle Bin or Trash Can.

Delete File(*path*, <Allow Undo(*Boolean*)>)

Description

Deletes the specified file.

Returns

Returns 1 if the file is deleted; otherwise, returns 0.

Required Argument

path Specifies the quoted path and file name for deletion.

Optional Argument

Allow Undo(*Boolean*) Allows undo operations, for example, moving to the Recycle Bin or Trash Can.

Directory Exists(*path*)

Description

Verifies the specified directory exists.

Returns

Returns 1 if the directory exists; otherwise returns 0.

Arguments

path Specifies the quoted path and directory for verification.

File Exists(*path*)

Description

Verifies the specified file name exists at the specified path.

Returns

Returns 1 if the file exists; otherwise returns 0.

Arguments

path Specifies the quoted path and file name for verification.

File Size(*path*)

Description

Determines the size of the file within the specified path.

Arguments

path Specifies the quoted path and file name.

Example

```
File Size( "$SAMPLE_DATA/Big Class.jmp" );
  13142
```

Files In Directory(*path*, <Recursive(*Boolean*)>)

Description

Returns a list of filenames in the *path* given.

Returns

List of filenames. If Recursive(*Boolean*) is not specified, directory names are included in the list.

Required Argument

path A valid quoted pathname.

Optional Argument

Recursive(*Boolean*) A keyword that causes all folders in the path (and all folders that they contain, and so on) to be searched for files.

Notes

Files In Directory(*path*, "Recursive"|*Boolean*) is deprecated.

Find All(<Project(*title*|*index*|*display box*|*window*)>, *type*, <"Invisible"|"Private">)

Description

Finds all open files of the specified type.

Required Argument

type A named argument that defines the type of resources that you want to close. The allowable types are: `Data Tables`, `Reports`, and `Journals`.

Optional Argument

`Project(`*title*`|index|`*display box*`|`*window*`)` Closes the specified project window.

`"Invisible"` Closes all invisible data tables.

`"Private"` Closes all private data tables.

Example

The following example finds all open data tables:

```
exdt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
exdt2 = Open( "$SAMPLE_DATA/Animals.jmp" );
windows = Find All( Data Tables );
For( i = 1, i <= N Items( windows ), i++,
  Write( Char( windows[i] << Get Window Title ) || "\!N" )
);
  Big Class
  Animals
```

Get Default Directory()

Description

Retrieves the user's default directory. This path is used for subsequent relative paths.

If the default directory was set using `Set Default Directory()`, JMP returns the specified path as long as `Get Default Directory()` and `Set Default Directory()` are in the same script.

See "Set Default Directory(path)".

Returns

The absolute pathname as a quoted string.

Arguments

none

Notes

`Get Default Directory()` also gets the path of an active saved scripting window.

Get Excel Worksheets(*absolute path*)

Description

Retrieves a list of worksheets that are in the specified Microsoft Excel workbook. If no worksheets are found, an empty list is returned. The path is a quoted string.

Notes

The function supports .xlsx and Excel 1997 or later workbooks.

Get File Search Path()

Description

Retrieves the current list of directories to search for opening files.

This list is configured using the `Set File Search Path()` function. See "Set File Search Path(path|{list of paths})".

Returns

A list of pathnames as quoted strings.

Get Path Variable(*name*)

Description

Retrieves the value of name, a path variable.

Returns

The absolute pathname as a quoted string.

Argument

`name` A quoted string that contains a path variable. (Examples: SAMPLE_DATA and SAMPLE_SCRIPTS)

Google Sheet Export(*dt*, *email*, *spreadsheet URL|spreadsheet ID|new spreadsheet name*, *sheet name*)

Description

Exports a data table to a Google sheet.

Returns

"1" if the export is successful.

Arguments

`dt` The data table.

`email` The quoted Google email address. @gmail.com is unnecessary.

`spreadsheet URL|ID` The quoted spreadsheet's URL or ID (the string that precedes "spreadsheets/d/").

`new spreadsheet name` The quoted name of the new spreadsheet that you are creating.

`sheet name` The quoted name of the sheet (or tab) within the spreadsheet.

Notes

- JMP features such as formulas and List Check column properties are not supported in Google Sheets.
- If the spreadsheet is empty, look in the JMP log for error messages. On Windows, select **View > Log**. On macOS, select **Window > Log**.

See Also

See *Using JMP* for details about security, country restrictions, and more.

`Google Sheet Import(`*email*`, `*spreadsheet URL*`|`*spreadsheet ID*`, <`*sheet name*`|{`*sheet name*`, `*sheet name*`}, `*Google Sheet Settings*`>)`

Description

Imports sheets from a Google Sheet.

Returns

A data table (or the first data table imported if several sheets are imported at once).

Required Arguments

email The quoted Google email address. @gmail.com is unnecessary.

spreadsheet URL|*ID* The quoted spreadsheet's URL or ID (which precedes "spreadsheets/d/").

Optional Arguments

sheet name The quoted name of the sheet or sheets that you want to import.

Google Sheet Settings The settings that describe how the data is imported.

See Also

See *Using JMP* for details about security, country restrictions, and more.

`Is Directory(`*path*`)`

Description

Returns 1 if the quoted path argument is a directory and 0 otherwise.

`Is Directory Writable(`*path*`)`

Description

Returns 1 if the directory specified in the quoted path is writable and 0 otherwise.

`Is File(`*path*`)`

Description

Returns 1 if the quoted path is a file and 0 otherwise.

`Is File Writable(`*path*`)`

Description

Returns 1 if the file specified in the quoted path is writable and 0 otherwise.

JSON To Data Table(*JSON string*, (<Private(*Boolean*)>|<Invisible(*Boolean*)>), <Guess(Stack(*Boolean*)|"Tall"|"Wide"|"Huge"|"Pandas")>)

Description

Converts a quoted JSON string to a data table.

Returns

A data table reference. The parsing of an empty value, "" string, missing value, or any other invalid value returns an empty data table.

Required Argument

JSON string The quoted JSON string.

Optional Arguments

Private(*Boolean*) Hides the data table completely. Specify this argument if the user doesn't need to interact with the data table.

Invisible(*Boolean*) Hides the data table from view but shows it in the JMP Home Window.

Guess(Stack(*Boolean*)|"Tall"|"Wide"|"Huge"|"Pandas") Stack applies to nodes that repeat within a parent node that is creating rows. By default, extra values are stored in a single table cell separated by commas. If the value is 1, repeating values are stacked in extra rows. Be careful stacking data. It can cause non-obvious data errors.

"Tall" imports the data in a tall data table. Select this option when the JSON file contains many rows. This option is the default setting.

"Wide" imports the data in a wide data table. Select this option when the JSON file contains many columns.

"Huge" imports the data in a tall and wide data table.

"Pandas" imports data in the pandas format.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt = JSON to Data Table(
   "[ { \!"name\!": \!"KATIE\!", \!"age\!": 12, \!"sex\!": \!"F\!",
   \!"height\!": 59, \!"weight\!": 95 }, { \!"name\!": \!"LOUISE\!",
   \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!": 61, \!"weight\!": 123 }, {
   \!"name\!": \!"JANE\!", \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!":
   55, \!"weight\!": 74 } ]"
);
```

Notes

Stacking the data can cause non-obvious data errors. Values that are supposed to be in the same row might not be. Here's an example of stacking data:

```
d =
"\[
{
       "toys": [{
```

```
                "Wheels": 2,
                "Color": "Red"
        },
        {
                "Wheels": 4,
                "Size": "Large"
        }]
}
]\";

JSON to Data Table( d, Guess( Stack( 1 ) ) );
```

Figure 2.1 Example of Stacking Data that Shows Errors

	Wheels	Color	Size
1	2	Red	Large
2	4		

The first toy should be red and have 2 wheels. The second toy should be large and have 4 wheels.

JSON To List(*JSON string*)

Description

Converts quoted JSON string to a JSL list that represents the structure specified by the JSON data. The parsing of an empty value, "" string, missing value, or any other invalid value returns {}.

Example

```
l = JSON To List(
   "[ { \!"name\!": \!"KATIE\!", \!"age\!": 12, \!"sex\!": \!"F\!",
   \!"height\!": 59, \!"weight\!": 95 }, { \!"name\!": \!"LOUISE\!",
   \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!": 61, \!"weight\!": 123 }, {
   \!"name\!": \!"JANE\!", \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!":
   55, \!"weight\!": 74 } ]"
);
Show( l );
```

JSON Literal(*JSON string*)

Description

Returns a valid JSON Boolean or null constant based on the specification of the parameter. The parsing of an empty value, "" string, missing value, or any other invalid value returns Empty().

Last Modification Date(*path*)

Description

Returns the last modification date of the path.

Returns

Last modification date.

Arguments

path The quoted directory or file name.

Load Text File(*path*, <Charset(character set), Force("Throw"|"Alert"|"Silent")>)>, <Line Separator(character)>, <"XMLParse">|<"SASODSXML">|<"JSON">|<BLOB(<arguments>)>)>)

Description

Reads the text file at *path* into a JSL variable.

Returns

A quoted string.

Required Argument

path A quoted pathname that points to a text file. The path can be a URL.

Optional Arguments

`Charset("charater set")` Specifies the character set. Valid character sets are `"Best Guess"`, `"utf-8"`, `"utf-16"`, `"us-ascii"`,`" windows-1252"`, `"x-max-roman"`, `"x-mac-japanese"`, `"shift-jis"`, `"euc-jp"`, `"utf-16be"`, and `"gb2312"`.

`Force("Throw"|"Alert"|"Silent")` Specifies what happens if the character set cannot be detected.

`Line Separator(`*character*`)` Specifies the quoted end-of-line character. For example, `"\!n"` specifies a line feed character. `"\!t"` specifies a tab.

`"XMLParse"` Converts an XML file into JSL.

`"SASODSXML"` Parses the text file as SAS ODS default XML.

`"JSON"` Converts JSON into an expression tree.

`BLOB(<`*arguments*`>)` Returns data from the file as a blob rather than a quoted string. The following optional arguments are for reading parts of the file:

- `ReadOffsetFromBegin(`*n*`)` specifies the zero-based offset to begin reading from the beginning of the file.
- `ReadOffsetFromEnd(`*n*`)` specifies the zero-based offset to begin reading from the end of the file.
- `ReadLength(`*n*`)` specifies the number of bytes to read from the file, either from the beginning of the file or from one of the offset values.

– `Base64Compressed(Boolean)` specifies how the blob is converted to a printable representation. 0, the default and recommended setting, uses JMP's ASCII~HEX representation. 1 means the blob is compressed and converted to base 64 when printed.

`Move Directory(from path, to path)`

Description

Moves a directory and its contents (including subdirectories) from the quoted path to another quoted path.

Returns

Returns 1 if the directory is moved; otherwise returns 0.

Arguments

`from path` Specifies the path and directory for relocation.

`to path` Specifies the destination path and directory.

`Move File(from path, to path)`

Description

Moves a file from the quoted path to another quoted path with the same or different file name.

Returns

Returns 1 if the file is moved. Otherwise, it returns 0.

Arguments

`from path` Specifies the quoted path and file name for relocation.

`to path` Specifies the quoted destination path and file name.

Notes

On Windows, when you move a file to a folder that does not exist, Windows creates the folder and returns 1. On macOS, the folder is not created, and an error is returned.

`Open(path, <arguments>)`

Description

Opens the data table or other JMP file or object created from a file named by the *`path`*. If no path is specified, the Open window appears. Also opens JSON and HDF5 files. See the examples in the JMP Scripting Index in the Help menu for more information about which arguments apply to specific file types.

Required Argument

`path` The quoted path to the file that you want to open.

Optional Arguments

`Add to Recent Files(Boolean)` Determines whether the file is added to the Recent Files list in the Home Window.

`Charset(character set)` The available character set options for importing text files are `"Best Guess"`, `"utf-8"`, `"utf-16"`, `"us-ascii"`,`" windows-1252"`, `"x-max-roman"`, `"x-mac-japanese"`, `"shift-jis"`, `"euc-jp"`, `"utf-16be"`, and `"gb2312"`.

`Column Names Start(n)|Column Names are on Line(n)` Specifies the line number that column names start in the imported text file. If the text file uses returns between cells, column names could be on multiple lines.

`Columns(colName=colType(colWidth),...)` Specifies the columns by name in the text file to import into a data table where:

– `colName`: Specifies the column name used in the imported text file.

– `colType("Character"|"Numeric")`: Specifies whether the specified column contains character or numeric data.

– `colWidth(n)`: Specifies the integer width of the specified column.

`Columns(<arguments>)` For ESRI shapefiles (.shp), this argument and its settings indicate the following:

– `Shape=numeric(n)`: Specifies the column number in the imported ESRI shapefile that contains the shape number.

– `Part=numeric(n)`: Specifies the column number in the imported ESRI shapefile that contains the part number.

– `X=numeric(n)`: Specifies the column number in the imported ESRI shapefile that contains the decimal degree for the longitude (range of ±180°).

– `Y=numeric(n)`: Specifies the column number in the imported ESRI shapefile that contains the decimal degree for the latitude (range of ±90°).

`"Column Names Only"` Returns a list of column names.

`Compress Allow List Check(Boolean)` Specifies that JMP can compress data table created from the imported text file.

`Compress Character Columns(Boolean)` Specifies that JMP should compress data table columns that contain character data from the imported text file.

`Compress Numeric Columns(Boolean)` Specifies that JMP should compress data table columns that contain numeric data from the imported text file.

`Concatenate Worksheets(Boolean)` Specifies that JMP should combine the imported Excel worksheets into one data table.

`Create Concatenation Column(Boolean)` Specifies that JMP should combine columns from an imported Excel file into one column.

`Data Starts(n)|Data Starts on Line(n)` Specifies the line number where data starts in the imported text file.

`Debug JSL(Boolean)` Opens the specified JSL script in the Debugger instead of opening it.

`End Of Field("Tab"|"Space"|"Comma"|"Semicolon"|"Other"|"None")` Specifies the quoted character used to delimit the end of a field in the imported text file. To specify multiple characters, separate each character designation by a comma. If you use `"Other"`, designate the delimiter with `EOF Other()` argument.

`End Of Line("CRLF"|"CR"|"LF"|"Semicolon"|"Other")` Specifies the quoted character used to delimit the end of a line in the imported text file. To specify multiple characters, separate each character designation by a comma. If you use `"Other"`, designate the delimiter with `EOL Other()` argument.

`EOF Other(char)` If the imported text file uses an end-of-field character other than the one specified by `End of Field()`, this argument specifies the character used.

`EOL Other(char)` If the imported text file uses an end of line character other than the one specified by `End of Line()`, this argument specifies the character used.

`"Excel Wizard"` Opens Microsoft Excel worksheets in the Excel Import Wizard. If you omit this argument, the worksheets open directly as a data table.

`File Type(type)` An optional quoted string that specifies the type of file that you are opening (for example, "text", "journal", "sas", "script", "png", and "jmp"). This can be useful if your file does not have a file extension, the file extension of the file does not match the contents of the file, or you want to import a JSL BLOB. If you do not specify this string, the file opens in the default program for the file extension.

Note: The path argument should be used for a zip archive. The extension (.zip) is not required. See "Zip Archives" for the messages that you can send to a zip archive. The basic functionality is to get a list of files in the zip archive, to read a file in the zip archive into either a quoted string or a blob, and to write files into the zip archive. Note that reading a zip archive temporarily puts the contents into memory. Reading very large zip archives can cause errors.

`"Force Refresh"` Closes the specified JMP (.jrn, .jsl, .jrp, or .jmpappsource) file without saving and tries to reopen the file from disk. This argument deletes any changes made since the last time the file was opened.

`HTML Table(n, <ColumnNames(n)>, <DataStarts(n)>)` To import a table from an HTML web page, use the URL as the filepath. The optional *n* argument specifies which table number, *n*, on the web page to open. If you omit the value, only the first table on the page is imported. The optional `ColumnNames(n)` specifies the row that contains column names. The optional `DataStarts(n)` specifies the row on which the data begins.

Tip: If the table you are importing contains images, they are first imported as text. To load the images in your JMP data table, run the automatically generated table script named Load Pictures. A new expression column containing the images is created. See *Using JMP* for more information about expression columns.

`Ignore Columns(col1, ...)` Specifies the quoted column names in the JMP data table or other JMP file that should not be included in the data table.

`"Invisible"` Opens the file as invisible. This quoted keyword applies to these files: data table, JMP file, external, text, Excel, SAS, ESRI shapefile, and HTML. The data table appears only in the JMP Home Window and the Window menu.

`Labels(Boolean)` Specifies the imported text file contains labels or column headers as the first line in the file. The default value is `True`.

`Lines to Read(n)` Specifies the number of lines in the text file to import. JMP starts counting lines after column names are read.

`Number of Columns(n)` Specifies the number of columns contained in the imported text file.

`Run JSL(Boolean)` Runs the specified JSL file instead of opening it. Include a Boolean argument or an expression that contains a Boolean value. If the script begins with `//!`, which automatically runs the script, include the Boolean value `(0)` to open the script instead.

`Password(password)` Specifies the quoted password for a password-protected SAS file to avoid entering it manually. The password is not encrypted. (Password-protected Microsoft Excel files cannot be imported.)

`"PDF Wizard"` Opens a PDF file in the PDF Import Wizard, where you can control how data is imported.

`"Private"` Opens the table as invisible and without showing it in the JMP Home Window or Window menu. For example, you might create a private data table to hold temporary data that the user does not need to see. This quoted keyword applies to the following files: data table, JMP file, external, text, Excel, SAS, ESRI shapefile, or HTML. Creating a private data table speeds the process of getting to the data; it does not save the computer from allocating the memory necessary to hold the data table data.

`Quarantine Action("Allow Scripts"|"Block Scripts"|"Do Not Open"|"Show Dialog")` Specifies whether scripts run when you open downloaded data tables. Also provides an option to display a window that prompts users to examine or open the data table. If they click Examine, the scripts are disabled.

`Scan Whole File(Boolean)` Specifies how long JMP scans the text file to determine data types for the columns. This is a Boolean value. The default value is true; the entire file is scanned until the data type is determined. To import large files, consider setting the value to false, which scans the file for five seconds.

`Select Columns(col1, ...)` Specifies the quoted column names in the JMP data table or other JMP file that should be included in the data table.

"Strip Quotes"|Strip Enclosing Quotes(*Boolean*) If the fields in the text file are quoted, setting this to True removes the quotes, and setting it to False does not remove the quotes. The default value is True.

Table Contains Column Headers(*Boolean*) Indicates the imported text file contains labels or column headers as the first line in the file. The default value is True.

"Text Wizard" Opens the text file in the text import window, where you can select import options. Otherwise, the Text Data Files options in the JMP preferences apply, and the text file is automatically imported as a data table.

Treat Empty Columns as Numeric(*Boolean*) Specifies that JMP should import text file columns of missing data as numeric rather than character. Possible missing value indicators are a period, a Unicode dot, NaN, or a blank quoted string. The default value is False.

Use Apostrophe as Quotation Mark(*Boolean*) Declares apostrophes as quotation marks in importing text files. This option is not recommended unless your data comes from a nonstandard source that places apostrophes around data fields rather than quotation marks. The default value is False.

Use Labels for Var Names(*Boolean*) For SAS data sets, this option specifies to use SAS labels as JMP columns names. The default value is False.

Use for All Sheets(*Boolean*) Indicates that JMP should use the Worksheets settings for all worksheets in the Excel file to be opened as a data table.

Worksheet Settings(*Boolean*, <*options*>) specifies options for importing an Excel file into a JMP data table. Available options are:

- Has Column Headers(*Boolean*) specifies that the Excel file has column headers in the first row.
- Number of Rows in Headers(*n*) specifies the number of rows in the Excel file used as column headers.
- Headers Start on Row(*n*) specifies the row number in the Excel file where the column headers begin. Default is row 1.
- Data Starts on Row(*n*) specifies the row number in the Excel file where the data begins.
- Data Starts on Column(*n*) specifies the column number in the Excel file where the data begins.
- Data Ends on Row(*n*) specifies the row number in the Excel file where data ends.
- Data Ends on Column(*n*) specifies the column number in the Excel file where the data ends.
- Replicated Spanned Rows(*Boolean*) specifies that the Excel file contains spanned columns should be imported into JMP as spanned columns.
- Suppress Hidden Rows(*Boolean*) indicates that JMP should not import rows hidden in the Excel file.

– `Suppress Hidden Columns(`*`Boolean`*`)` indicates that JMP should not import columns hidden in the Excel file.

– `Treat as Hierarchy(`*`Boolean`*`)` indicates that JMP should treat multiple column headers (merged cells) as hierarchies when importing an Excel file. If `True`, the Excel file opens with the merged columns stacked (Tables > Stacked).

`Worksheets("Sheet Name"|{"Sheet Name", "Sheet Name",...}|`*`n`*`)` Opens the specified Excel file worksheet by name, all worksheets in a list of names, or the indexed number of the worksheet. If the worksheets are not specified, all worksheets in the spreadsheet open as separate data tables.

Note: You can import .xls worksheets from a web site by specifying worksheets arguments. .xlsx worksheets cannot be imported from a web site using `Open()`.

`"XML Wizard"` Opens XML files in the XML Import Wizard. If you omit this argument, the XML file opens as a data table.

```
Year Rule | Two Digit Year Rule
  ("1900-1999"|"1910-2009"|"1920-2019"|"1930-2029"|"1940-2039"|
  "1950-2049"|"1960-2059"|"1970-2069"|"1980-2079"|"1990-2089"|"2000-2099")
```

Indicates the year format used in the imported text file. For example, if the earliest date is 1979, use "1970-2069".

`Open Database(dataSourceName|"Connect Dialog", "SELECT ..."|"SQLFILE..."|tableName, <"Invisible"|"Private">, <outputTableName>)`

Description

Opens the database indicated by *`dataSourceName`* (or opened by the user) with the *`SELECT`* , *`SQLFILE`*, or *`tableName`* arguments.

Returns

A data table named *`outputTableName`*.

Required Argument

`dataSourceName``|"Connect Dialog"` dataSourceName specifies the name of the data source. `"Connect Dialog"` shows a window from which the user selects the data source.

Optional Arguments

`"Invisible"` Creates an invisible data table that hides the table from view but lists it in the JMP Home Window and Window menu.

`"Private"` Hides the data table completely. Creating a private data table speeds the process of getting to the data; it does not save the computer from allocating the memory necessary to hold the data table data.

`outputTableName` The name of the data table in JMP.

Example

```
Open Database(
  "DSN=dBASE Files;DBQ=C:/Program Files/SAS/JMPPRO/16/Samples/Import Data/;",
```

```
    // SQL statement
    "SELECT HEIGHT, WEIGHT FROM Bigclass", // selected columns
    "hw" // name of the output data table
);
```

Parse JSON(JSON string)

Description

Converts JSON string into an associative array or list based on the structure of the JSON data. Convert the result to a list if the parsed JSON object contains more than one member. The parsing of an empty value, "" quoted string, missing value, or any other invalid value returns Empty().

Example

The following example converts JSON into a list:

```
j = Parse JSON(
    "[ { \!"name\!": \!"KATIE\!", \!"age\!": 12, \!"sex\!": \!"F\!",
    \!"height\!": 59, \!"weight\!": 95 }, { \!"name\!": \!"LOUISE\!",
    \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!": 61, \!"weight\!": 123 }, {
    \!"name\!": \!"JANE\!", \!"age\!": 12, \!"sex\!": \!"F\!", \!"height\!":
    55, \!"weight\!": 74 } ]"
);
Show( j );
    j = {["age" => 12, "height" => 59, "name" => "KATIE", "sex" => "F", "weight"
       => 95], ["age" => 12, "height" => 61, "name" => "LOUISE", "sex" => "F",
       "weight" => 123], ["age" => 12, "height" => 55, "name" => "JANE", "sex"
       => "F", "weight" => 74]};
```

Pick Directory(<"Prompt">, <*path*>, <Show Files(*Boolean*)>)

Description

Prompts the user for a directory, returning the directory path as a quoted string.

Returns

The path for the directory that the user selects.

Optional Arguments

"Prompt" The quoted string appears at the top of the Browse window (Windows) or the Finder window (macOS).

path The quoted string specifies the initial directory that appears in the Pick Directory window.

Show Files(*Boolean*) 1 shows files in the Pick Directory window. 0 hide the files. The default is 0.

Pick File(<"Prompt">, <*initial directory*>, <{*filter list*}>, <*first filter*>, <Save Flag(*Boolean*)>, <default file>), <"Multiple">)

Description

Prompts the user to select one or more files in the Open window.

Returns

The path of the file that the user selects.

Optional Arguments

`"Prompt"` A quoted string. If provided, that string appears at the top of the Open window.

`initial directory` A quoted string that is a valid filepath to a folder. If provided, it specifies where the Open window begins. If not provided, or if it's an empty string, the JMP Default Directory is used.

`filter list` A list of quoted strings that define the filetypes to show in the Open window. See the following example for syntax.

`first filter` An integer that specifies which of the filters in the filter list to use initially. If you use an integer that is too large or small for the list (for example, 4 for a list of 3), the first filter in the list is used.

`Save Flag(`*`Boolean`*`)` A Boolean that specifies whether the Open window or Save window is used. 0 lets the user select a file to open in JMP. 1 lets the user save a new, empty file of the selected type in the selected folder. The default value is 0.

`default file` The quoted name of the file that appears in the window by default.

`"Multiple"` The quoted name that lets the user select multiple files if the `Save Flag` value is 0.

Notes

Although all arguments are optional, they are also positional. For example, you cannot specify a filter list without also specifying the caption and the initial directory.

The buffer size in the computer's physical memory affects the number of files the user can open.

Example

The following script assigns *Select JMP File* as the window title; shows the JMP Samples/Data directory; shows *JMP Files* and *All Files* in the File name list and selects *JMP Files*; displays the Open window; and shows the sample data file name Hollywood Movies.jmp.

```
Pick File(
  "Select JMP File",
  "$SAMPLE_DATA",
  {"JMP Files|jmp;jsl;jrn", "All Files|*"},
  1,
  0,
  "Hollywood Movies.jmp"
```

```
);
```

Rename Directory(*old path name*, *new directory name*)

Description

Renames a directory without moving or copying it.

Returns

Returns 1 if the directory is renamed; otherwise, returns 0.

Arguments

old path name A quoted string that specifies the path and old directory name.

new directory name A quoted string that specifies the new directory name.

Notes

When you specify the *new directory Name*, include only the directory name, not the entire path.

Rename File(*old path name*, *new file name*)

Description

Renames a file without moving or copying it.

Returns

Returns 1 if the file is renamed; otherwise, returns 0.

Arguments

old path name Specifies the path and old file name.

new file name Specifies the new file name.

Notes

When you specify the *new name*, include only the file name, not the entire path.

Save Text File(*path*, *text*|BLOB, <Mode("Replace"|"Append")>)

Description

Returns the path name of the created file. *text* is quoted string.

Set Default Directory(*path*)

Description

Sets the default directory, which is used for resolving relative paths.

See Also

"Get Default Directory()"

Set File Search Path(*path*|{*list of paths*})

Description

Sets the current list of directories to search for opening files. Using {"."} as the path configures JMP to use the current directory.

Example

```
Set File Search Path( {"C:/JMP/13/source", "C:/Program
  Files/SAS/JMPPRO/16/Samples"} );
```

See Also

"Get File Search Path()".

Set Path Variable(*name*, <*path*>)

Description

Sets the path stored in the variable.

Required Argument

name The quoted name of the variable.

Optional Argument

path The path.

TripleS Import(*path*, <*arguments*>)

Description

Imports the specified Triple-S Survey (SSS) file. The SSS format consists of a pair of files: .xml or .sss, and a .csv, .dat, or .asc file. Both sets of files must have the same root name and be in the same folder.

Required Argument

path A quoted string that contains the full path to the .xml or .sss file.

Optional Arguments

"Invisible" A quoted keyword that hides the table from view. The data table appears only in the JMP Home Window and the Window menu. Hidden data tables remain in memory until they are explicitly closed, reducing the amount of memory that is available to JMP. To explicitly close the hidden data table, call Close(*dt*), where *dt* is the data table reference returned by TripleS Import.

Use Labels for Imported Column Names(*Boolean*) Converts the label names to column headings. The default value is 1.

Example

```
dt = TripleS Import( "C:/Data/airlines.sss", "Invisible", Use Labels for
  Imported Column Names( 0 ) );
```

Financial Functions

Double Declining Balance(*cost*, *salvage*, *life*, *period*, <*factor*>)

Description

Returns the depreciation of an asset for a specified period of time. The function uses the double-declining balance method or some other depreciation factor.

Required Arguments

cost The initial cost.

salvage The value at the end of the depreciation.

life The number of periods in the depreciation cycle.

period The length of the period, in the same units as *life*.

Optional Argument

factor The rate at which the balance declines. The default value is 2.

Notes

This function is equivalent to the Excel function DDB.

Future Value(*rate*, *nper*, *pmt*, <*pv*>, <*Type(Boolean)*>)

Description

Returns the future value of an investment that is based on periodic, constant payments and a constant interest rate.

Required Arguments

rate The interest rate.

nper The number of periods.

pmt The constant payment.

Optional Arguments

pv The present value. The default value is 0.

Type(*Boolean*) 0 specifies end-of-period payments. 1 specifies beginning-of-period payments. The default value is 0.

Notes

This function is equivalent to the Excel function FV.

Interest Payment(*rate*, *per*, *nper*, *pv*, <*fv*>, <*Type(Boolean)*>)

Description

Returns the interest payment for a given period for an investment that is based on periodic, constant payments and a constant interest rate.

Required Arguments

rate The interest rate.

per The period for which you want the interest.

nper The total number of periods.

pv The present value.

Optional Arguments

fv The future value. The default value is 0.

Type(Boolean) 0 specifies end-of-period payments, and 1 specifies beginning-of-period payments. The default value is 0.

Notes

This function is equivalent to the Excel function IPMT.

Interest Rate(*nper*, *pmt*, *pv*, <*fv*>, <*Type(Boolean)*>, <*guess*>)

Description

Returns the interest rate per period of an annuity.

Required Arguments

nper The total number of periods.

pmt The constant payment.

pv The present value.

Optional Arguments

fv The future value. The default value is 0.

Type(Boolean) 0 specifies end-of-period payments, and 1 specifies beginning-of-period payments. The default value is 0.

guess Your estimate of the rate The default value is 0.1 (10%).

Notes

This function is equivalent to the Excel function RATE.

Internal Rate of Return(*values*, <*guess*>)

Internal Rate of Return(*guess*, *value1*, *value2*, ...)

Description

Returns the internal rate of return for a series of cash flows in the *values* argument.

Arguments

values A one-dimensional matrix of values. If the second form of the function is used, list each value separately.

guess What you think is near the result. The default value is 0.1 (10%). This argument is required if you specify values. It is required if you specify individual values.

Notes

This function is equivalent to the Excel function IRR.

`Modified Internal Rate of Return(`*values*`, `*finance rate*`, `*reinvest rate*`)`

`Modified Internal Rate of Return(`*finance rate*`, `*reinvest rate*`, `*value1*`, `*value2*`, ...)`

Description

Returns the modified internal rate of return for a series of periodic cash flows. The cost of investment and the interest received on reinvested cash is included.

Arguments

values A one-dimensional matrix of values. If the second form of the function is used, list each value separately.

finance rate The interest rate that you pay on the money in the cash flows.

reinvest rate The interest rate that you receive on the cash flows when you reinvest them.

Notes

This function is equivalent to the Excel function MIRR.

`Net Present Value(`*rate*`, `*values*`)`

`Net Present Value(`*rate*`, `*value1*`, `*value2*`, ...)`

Description

Returns the net present value of an investment by using a discount rate and a series of future payments (negative values) and income (positive values).

Arguments

rate The discount rate.

values A one-dimensional matrix of values. If the second form of the function is used, list each value separately.

Notes

This function is equivalent to the Excel function NPV.

`Number of Periods(`*rate*`, `*pmt*`, `*pv*`, <`*fv*`>, <`*type(0|1)*`>)`

Description

Returns the number of periods for an investment that is based on periodic, constant payments and a constant interest rate.

Required Arguments

rate The interest rate.

pmt The constant payment.

pv The present value.

Optional Arguments

fv The future value. The default value is 0.

type(0|1) 0 specifies end-of-period payments, and 1 specifies beginning-of-period payments. The default value is 0.

Notes

This function is equivalent to the Excel function NPER.

`Payment(`*rate*`, `*nper*`, `*pv*`, <`*fv*`>, <`*type(0|1)*`>)`

Description

Returns the payment for a loan that is based on constant payments and a constant interest rate.

Required Arguments

rate The interest rate.

nper The total number of periods.

pv The present value.

Optional Arguments

fv The future value. The default value is 0.

type(0|1) 0 specifies end-of-period payments, and 1 specifies beginning-of-period payments. The default value is 0.

Notes

This function is equivalent to the Excel function PMT.

`Present Value(`*rate*`, `*nper*`, `*pmt*`, <`*fv*`>, <`*type(0|1)*`>)`

Description

Returns the present value of an investment.

Arguments

rate The interest rate per period.

nper The total number of periods.

pmt The constant payment.

fv The future value. The default value is 0.

type(0|1) 0 specifies end-of-period payments, and 1 specifies beginning-of-period payments. The default value is 0.

Notes

This function is equivalent to the Excel function PV.

Principal Payment(*rate*, *per*, *nper*, *pv*, <*fv*>, <*type(0|1)*>)

Description

Returns the payment on the principal for a given period for an investment that is based on periodic, constant payments and a constant interest rate.

Required Arguments

rate The interest rate per period.

per The period for which you want the interest.

nper The total number of periods.

pv The present value.

Optional Arguments

fv The future value. The default value is 0.

type(0|1) An optional switch. 0 specifies end-of-period payments, and 1 specifies beginning-of-period payments. The default value is 0.

Notes

This function is equivalent to the Excel function PPMT.

Straight Line Depreciation(*cost*, *salvage*, *life*)

Description

Returns the straight-line depreciation of an asset for one period.

Arguments

cost The initial cost of the asset.

salvage The value at the end of the depreciation.

life The number of periods in the depreciation cycle.

Notes

This function is equivalent to the Excel function SLN.

Sum Of Years Digits Depreciation(*cost*, *salvage*, *life*, *per*)

Description

Returns the sum-of-years' digits depreciation of an asset for a specified period.

Arguments

cost The initial cost of the asset.

salvage The value at the end of the depreciation.

life The number of periods in the depreciation cycle.

per The length of the period, in the same units as life.

Notes

This function is equivalent to the Excel function SYD.

Graphics Functions

Add Color Theme({*name*, <*flags*>, {*color*}, <{*position*}>)

Description

Creates a custom color theme that you can apply to components such as markers, data table rows, and treemaps. Add the color theme to the JMP Preferences by including `Add Color Theme(...)` inside `Preferences()`.

Returns

Null.

Required Arguments

`name` The quoted name of the color theme.

`color` Lists of RGB values. These values define the blocks in categorical color themes and the gradients in continuous color themes. Each list of RGB values corresponds to a slider in the preferences Color Themes window.

`position` An optional list of numbers between 0 and 1 with one position per color. Each position corresponds to a slider in the preferences Color Themes window. If you omit the position, the sliders are evenly spaced.

Optional Argument

`flags` A flag for the Continuous or Categorical color theme list and category of color.

```
Continuous, <Continuous>, Sequential
Continuous, <Continuous>, Diverging
Continuous, <Continuous>, Chromatic
Categorical, <Continuous>, Sequential
Categorical, <Continuous>, Diverging
Categorical, Qualitative
Categorical, <Continuous>, Chromatic
```

With the default JMP color themes, Sequential colors transition from left to right or right to left. Diverging colors are lighter in the middle. Chromatic colors consist of blocks or gradients of bright color. All categories except for Qualitative can be both continuous and categorical.

If you omit the flag, the color is shown in the Continuous, Sequential and Categorical, Sequential categories.

Examples

The following example creates a continuous color theme named Blue to Purple. The color is in the Diverging category. RGB values are defined in the four lists.

```
Add Color Theme(
  {"Blue to Purple", {"Continuous", "Diverging"}, {{0, 0, 255},
  {57, 108, 244}, "white", {128, 0, 100}}} );
```

Notes

Any style except for Qualitative can be Continuous and Categorical at the same time. For example, the Cool to Warm Diverging theme is in the Continuous and Categorical theme lists. In JMP, select **Preferences > Graphs** to see examples.

To delete a color theme, use `Remove Color Theme()`.

See Also

"Remove Color Theme(name|{name, <flags>, {color, ...}, <{position, ...}>}>)"

`Arc(x1, y1, x2, y2, startangle, endangle)`

Description

Inscribes an arc in the rectangle described by the arguments.

Returns

Null.

Arguments

x1, *y1* The point at the top left of the rectangle

x2, *y2* The point at the bottom right of the rectangle

startangle, *endangle* The starting and ending angle in degrees, where 0 degrees is 12 o'clock and the arc or slice is drawn clockwise from *startangle* to *endangle*.

`Arrow(<pixellength>, {x1, y1}, {x2, y2})`

Description

Draws an arrow from the first point to the second point. The optional first argument specifies the length of the arrow's head lines (in pixels).

Returns

Null.

Required Arguments

{*x1*, *y1*}, {*x2*, *y2*} Two lists of two numbers that each specify a point in the graph.

Optional Argument

pixellength The length of the arrowhead in pixels.

Notes

The two points can also be enclosed in square brackets: `Arrow(<pixellength>, [x1, x2], [y1, y2])`.

`Back Color(name)`

Description

Sets the color used for filling the graph's background.

Returns

Null.

Argument

`name` A quoted color name or a color index (such as `"red"` or `"3"` for the color red).

`Char To Path(path)`

Description

Converts a path specification from a quoted string to a matrix.

Returns

A matrix.

Arguments

`path` A quoted string that contains the path specification.

`Circle({x, y}, radius|PixelRadius(n), <...>, <"Fill">)`

Description

Draws a circle centered at {x, y} with the specified radius.

Returns

Null.

Required Arguments

`{x, y}` A number that describes a point in the graph

`radius` The length of the circle's radius in relation to the vertical axis. If the vertical axis is resized, the circle is also resized.

`PixelRadius(n)` The length of the circle's radius in pixels. If the vertical axis is resized, the circle is *not* resized.

Optional Argument

`"Fill"` A positional argument indicating that all circles defined in the function are filled with the current fill color. If `"Fill"` is omitted, the circle is empty.

Notes

The center point and the radius can be placed in any order. You can also add additional center point and radius arguments and draw more than one circle in one statement. One point and several radii results in a bull's-eye. Adding another point still draws all previous circles, and then adds an additional circle with the last radius specified. This means that this code:

```
Graph Box(circle({20, 30}, 5, {50, 50}, 15))
```

results in three circles, not two. First, a circle with radius 5 is drawn at 20, 30. Second, a circle with radius 5 is drawn at 50, 50. Third, a circle with radius 15 is drawn at 50, 50.

Color To HLS(*color*)

Description

Converts the *color* argument (including any JMP color) to a list of HLS values.

Returns

A list of the hue, lightness, and saturation components of *color*. The values range between 0 and 1.

Argument

color A number from the JMP color index.

Example

The output from ColorToHLS() can either be assigned to a single list variable or to a list of three scalar variables:

```
hls = Color To HLS( 8 );
{h, l, s} = Color To HLS( 8 );
Show( hls, h, l, s );
   hls = {0.778005464480874, 0.509803921568627, 0.976};
   h = 0.778005464480874;
   l = 0.509803921568627;
   s = 0.976;
```

Color To RGB(*color*)

Description

Converts the *color* argument (including any JMP color) to a list of RGB values.

Returns

A list of the red, green, and blue components of *color*. The values range between 0 and 1.

Argument

color A number from the JMP color index.

Example

The output from ColorToRGB() can either be assigned to a single list variable or to a list of three scalar variables:

```
rgb = Color To RGB( 8 );
{r, g, b} = Color To RGB( 8 );
Show( rgb, r, g, b );
   rgb = {0.670588235294118, 0.0313725490196078, 0.988235294117647};
   r = 0.670588235294118;
   g = 0.0313725490196078;
   b = 0.988235294117647;
```

Contour(*xVector*, *yVector*, *zGridMatrix*, *zContour*, <*messages*>)

Description

Draws contours given a grid of values.

Returns

None.

Required Arguments

xVector The *n* values that describe *zGridMatrix*.

yVector The *m* values that describe *zGridMatrix*.

zGridMatrix An *nxm* matrix of values on some surface.

zContour The values for the contour lines.

Optional Messages

<<zColor(color, option) The colors to use for the contour lines.

<<"Fill"|"Fill Between"|"Fill Below"|"Fill Above" Specifies where the contour fill appears.

<<Transparency(vector) Specifies the transparency level of the fill in a vector.

Contour Function(*zExpr*, *xName*, *yName*, (*z*|*zMatrix*), < <<xGrid(*min*, *max*, *incr*)>, < <<yGrid(*min*, *max*, *incr*)>, < <<zColor(*color*, *option*)>, < <<*zLabeled*>, < <<"Filled">, < <<"FillBetween">, < <<"Ternary">, < <<Transparency(*number*|*matrix*))

Description

Draws sets of contour lines of the expression, a function of the two symbols. The *z* argument can be a single value or an index or matrix of values.

Returns

None.

Arguments

zExpr Any expression (for example, Sine(y)+Cosine(x)).

xName, yName The values to use in the expression.

z|zMatrix A z-value or a matrix of z-values.

Optional Messages

<<xGrid, <<yGrid Defines a box, beyond which the contour lines are not drawn.

<<zColor Defines the color in which to draw the contour lines. The argument can be either a scalar or a matrix but must evaluate to numeric.

<<zLabeled Labels the contours.

<<"Filled" Fills the contour levels using the current fill color.

<<"FillBetween" Fills only between adjacent contours using the current fill color. For nz contours specified, this option fills nz-1 regions for the intervals between the nz values. Using this option is recommended over using the <<Filled option.

<<"Ternary" Clips lines to be within the ternary coordinate system inside ternary plots.

<<Transparency(*number*|*matrix*) Sets the transparency level of the fill. A vector of numbers between 0 and 1 are sequenced through and cycled for the z contours. This option should be used only in conjunction with the <<FillBetween option.

Drag Line(*xMatrix*, *yMatrix*, <*dragScript*>, <*mouseUpScript*>)

Description

Draws line segments between draggable vertices at the coordinates given by the matrix arguments.

Returns

None.

Required Arguments

xMatrix A matrix of *x*-coordinates.

yMatrix A matrix of *y*-coordinates.

Optional Arguments

dragScript Any valid JSL script. The script is run at drag.

mouseUpScript Any valid JSL script. The script is run at mouseup.

Drag Marker(*xMatrix*, *yMatrix*, <*dragScript*>, <*mouseUpScript*>)

Description

Draws draggable markers at the coordinates given by the matrix arguments.

Returns

None.

Arguments

xMatrix A matrix of *x*-coordinates.

yMatrix A matrix of *y*-coordinates.

dragScript Any valid JSL script. The script is run at drag.

mouseUpScript Any valid JSL script. The script is run at mouseup.

Drag Polygon(*xMatrix*, *yMatrix*, <*dragScript*>, <*mouseupScript*>)

Description

Draws a filled polygon with draggable vertices at the coordinates given by the matrix arguments.

Returns

None.

Required Arguments

xMatrix A matrix of *x*-coordinates.

yMatrix A matrix of *y*-coordinates.

Optional Arguments

dragScript Any valid JSL script. The script is run at drag.

mouseUpScript Any valid JSL script. The script is run at mouseup.

Drag Rect(*xMatrix*, *yMatrix*, <*dragScript*>, <*mouseupScript*>)

Description

Draws a filled rectangle with draggable vertices at the first two coordinates given by the matrix arguments.

Returns

None.

Required Arguments

xMatrix A matrix of two *x*-coordinates.

yMatrix A matrix of two *y*-coordinates.

Optional Arguments

dragScript Any valid JSL script. The script is run at drag.

mouseUpScript Any valid JSL script. The script is run at mouseup.

Notes

xMatrix and *yMatrix* should each contain exactly two values. The resulting coordinate pairs should follow the rules for drawing a `Rect()`. The first point (given by the first value in *xMatrix* and the first value in *yMatrix*) must describe the top, left point in the rectangle. The second point (given by the second value in *xMatrix* and the second value in *yMatrix*) must describe the bottom, right point in the rectangle.

Drag Text(*xMatrix*, *yMatrix*, *string*, <*dragScript*>, <*mouseUpScript*>)

Description

Draws the *string* (or all the items if a list is specified) at the coordinates given by the matrix arguments.

Returns

None.

Arguments

xMatrix A matrix of *x*-coordinates.

yMatrix A matrix of *y*-coordinates.

string A quoted string to be drawn in the graph.

dragScript Any valid JSL script. The script is run at drag.

mouseUpScript Any valid JSL script. The script is run at mouseup.

Fill Color(*name*|*color*|*rgbList*)

Description

Sets the color used for filling solid areas.

Returns

None.

Arguments

name|color|rgbList The quoted color name, index, or list of RGB values.

`Fill Pattern(`*name|mask|image*`)`

Description

Sets the pattern for filled areas.

Arguments

name|mask|image The quoted color name, matrix of values between 0 and 1, or image path.

`Get Color Theme Detail(`*name*`)`

Description

Returns a script for the specified color theme.

Example

The following example returns the script for the JMP default color theme:

```
Get Color Theme Details( "JMP Default" );
  {"JMP Default", 9221, {{213, 72, 87}, {57, 177, 67}, {64, 111, 223}...}}
```

`Get Color Theme Names(<`*kind*`>)`

Description

Returns a list of all color theme names or color themes of the specified kind. The kinds include `"Continuous"`, `"Categorical"`, `"Sequential"`, `"Diverging"`, `"Qualitative"`, or `"Chromatic"`.

Example

The following example returns all color themes:

```
Get Color Theme Names();
  {"Green to Black to Red", "Green to White to Red", "White to Black"...}
```

The following example returns the diverging color themes:

```
Get Color Theme Names( "diverging" );
  {"Green to Black to Red", "Green to White to Red", "Blue to Gray to Red"...}
```

`Gradient Function(`*zExpr*`, `*xName*`, `*yName*`, [`*zLow*`, `*zHigh*`], zColor([`*colorLow*`, `*colorHigh*`]), < <<xGrid(`*min*`, `*max*`, `*incr*`)>, < <<yGrid(`*min*`, `*max*`, `*incr*`)> < <<Transparency(`*alpha|vector*`))`

Description

Fills a set of rectangles on a grid according to a color determined by the expression value as it crosses a range corresponding to a range of colors.

Required Arguments

zExpr Any expression (for example, Sine(y)+Cosine(x)).

xName, *yName* The values to use in the expression.

zLow, *zHigh* A matrix of low and high values (2 to 10).

zColor([colorLow, colorHigh) The two colors that are blended together (4 is green, 6 is orange).

Optional Messages

<<*xGrid*, <<*yGrid* Specifies a box, beyond which the gradients are not drawn.

<<*zColor* The color in which to draw the contour lines. The argument can be either a scalar or a matrix but must evaluate to numeric.

<<Transparency(*number*|*matrix*) The transparency level of the fill. A vector of numbers between 0 and 1 are sequenced through and cycled for the z contours.

Example

```
Gradient Function(Log(a * a + b * b),
a, b, [2 10],
zColor([4, 6]));
```

H Line(*x1*, *x2*, *y*)

H Line(*y*)

Description

Draws a horizontal line at *y* across the graph. If you specify start and end points on the *x*-axis (*x1* and *x2*), the line is drawn horizontally at *y* from *x1* to *x2*. You can also draw multiple lines by using a matrix of values in the *y* argument.

H Size()

Description

Returns the horizontal size of the graphics frame in pixels.

Handle(*xPos*, *yPos*, *dragScript*, *mouseUpScript*)

Description

Places draggable marker at coordinates given by *xPos*, *yPos*. The first script is executed at drag and the second at mouseup.

Heat Color(*x*, < <<*theme*>)

Heat Color(*x*)

Description

Returns the JMP color that corresponds to *n* in the color "*theme*".

Returns

An integer that is a JMP color.

Required Argument

x A value between 0 and 1.

Optional Message

theme A quoted color theme that is supported by Cell Plot. The default value is "Blue to Gray to Red".

`HLS Color(h, l, s)`

`HLS Color({h, l, s})`

Description

Converts hue, lightness, and saturation values into a JMP color number.

Returns

An integer that is a JMP color number.

Arguments

Hue, lightness, and saturation, or a list containing the three HLS values. All values should be between 0 and 1.

`In Path(x, y, (pathMatrix|pathText)`

Description

Determines it the point described by *x* and *y* falls in *path*.

Returns

True (1) if the point (x, y) is in the given path, False(0) otherwise.

Arguments

x and *y* The coordinates of a point.

pathMatrix|*pathText* Either a matrix or a quoted string describing a path.

`In Polygon(x, y, xMatrix, yMatrix)`

`In Polygon(x, y, xyPolygon)`

Description

Returns 1 or 0, indicating whether the point (*x*, *y*) is inside the polygon that is defined by the *xMatrix* and *yMatrix* vector arguments.

xMatrix, *yMatrix* can also be combined into a two-column matrix (*xyPolygon*), allowing you to use three arguments instead of four. Also, *x* and *y* can be conformable vectors, and then a vector of 0s and 1s are returned based on whether each (*x*, *y*) pair is inside the polygon.

Level Color(*i*)

Level Color(*i*, *n*)

Level Color(*i*, *n*, <*theme*>)

Level Color(*i*, <*theme*>)

Description

Assigns a JMP color to categorical data in a graphic.

Returns

An integer that is a JMP color.

Required Arguments

i An integer that is greater than or equal to 1 and less than or equal to the number of categories specified by n.

n The number of categories.

Optional Argument

theme A quoted color theme from the Value Color list of the Column Properties window. If not specified, the JMP Default color theme is applied.

Notes

When the second argument is a quoted character string and not *n*, then the second argument determines the color theme.

Line({*x1*, *y1*}, {*x2*, *y2*}, ...), <<ValueSpace(*Boolean*)

Line([*x1*, *x2*, ...], [*y1*, *y2*, ...]), <<ValueSpace(*Boolean*)

Description

Draws a line between points.

Arguments

{*x1*, *y1*}, {*x2*, *y2*}|[*x1*, *x2*, ...], [*y1*, *y2*, ...] Can be any number of lists of two points, separated by commas or a matrix of *x* values and a matrix of *y* values.

<<ValueSpace(*Boolean*) Draws lines that follow the projection when the line represents a movement of the underlying data, such as a bubble trail in a bubble plot. The Boolean value can be a constant or an expression.

Line Style(*name*|*number*)

Description

Sets the line style used to draw the graph.

Argument

n Can be either a quoted style name or the style's number:

– 0 or "Solid"

- 1 or "Dotted"
- 2 or "Dashed"
- 3 or "DashDot"
- 4 or "DashDotDot"

Marker(<*rowState*>, {*x1*, *y1*}, {*x2*, *y2*}, ...)

Marker(<*rowState*>, [*x1*, *x2*, ...], [*y1*, *y2*, ...])

Description

Draws one or more markers at the points described either by lists or matrices. The optional *rowState* argument sets the type of marker.

Marker Size(*n*)

Description

Sets the size used for markers.

Mousetrap(*dragScript*, *mouseUpscript*)

Description

Captures click coordinates to update graph properties. The first script is executed at drag and the second script at mouseup.

New Heat Image(*matrix*, <*colorTheme*>)

Description

Creates a heat map image based on a matrix and color theme or gradient.

Arguments

matrix A matrix of values. The function creates a heat map that has the same dimensions as the input matrix.

colorTheme A color theme or set of gradient properties that define the mapping of values to colors. The example below shows the use of a custom gradient.

Example

```
width = 256;
height = 256;
wstride = 16;
hstride = 16;
b = J( hstride, wstride, Random Normal() );
a = Transform Each( {z, {row, col}},
  J( height, width, 0 ),
  b[Floor( (row - 1) / hstride ) + 1,
  Floor( (col - 1) / wstride ) + 1]
```

```
);
E1 = New Heat Image(
   a,
   gradient(
      {Color Theme( "Blue To Gray To Orange" ),
      Scale Type( "Standard Deviation" ),
      Show Missing Color( "On" ),
      Reverse Gradient( 1 )}
   )
);
New Window( "test", Lineup Box( E1 ) );
```

Normal Contour(*prob*, *meanMatrix*, *stdMatrix*, *corrMatrix*, <*colorsMatrix*>, <*fill=x*>)

Description

Draws normal probability contours for *k* populations and two variables.

Required Arguments

prob A scalar or matrix of probabilities.

meanMatrix A matrix of means of size *k* by 2.

stdMatrix A matrix of standard deviations of size *k* by 2.

corrMatrix A matrix of correlations of size *k* by 1.

Optional Arguments

colorsMatrix Specifies the color(s) for the k contour(s). The colors must be specified as JSL colors (either JSL color integer values or return values of JSL Color functions such as RGB Color or HLS Color).

fill=x Specifies the amount of transparency for the contour fill color.

Oval(*x1*, *y1*, *x2*, *y2*, <*Fill(Boolean)*>)

Oval({*x1*, *y1*}, {*x2*, *y2*}, <*Fill(Boolean)*>)

Description

Draws an oval inside the rectangle whose diagonal has the coordinates (*x1, y1*) and (*x2, y2*). *Fill* is Boolean. If *Fill* is 0, the oval is empty. If *Fill* is nonzero, the oval is filled with the current fill color. The default value for *Fill* is 0.

Path((*pathMatrix*|*pathText*), <*Fill(Boolean)*>)

Description

Draws a stroke along the given path. If a fill is specified, the interior of the path is filled with the current fill color.

Required Arguments

pathMatrix An Nx3 matrix. If you don't specify the path matrix, specify the path text.

pathText A quoted string that contains SVG code.

Optional Arguments

Fill(Boolean) Specifies whether a line is drawn (0) or the path is filled (1). The default value is 0.

Notes

A path matrix has three columns, for x and y, and a flag. The flag value for each point can be 0 for control, 1 for move, 2 for line segment, 3 for cubic Bézier segment, and any negative value to close the path.

Path To Char(*path*)

Description

Converts a path specification from a matrix to a quoted string.

Returns

A quoted string.

Argument

path An Nx3 path matrix.

Notes

A path matrix has three columns, for x and y, and a flag. The flag value for each point can be 0 for control, 1 for move, 2 for line segment, 3 for cubic Bézier segment, and any negative value to close the path.

Pen Color(*n*)

Description

Sets the color used for the pen.

Pen Size(*n*)

Description

Sets the thickness of the pen in pixels.

Pick Color(<*window title*>, <*name*|*index*|*RGBlist*>)

Description

Creates a color picker, which enables the user to select a color to apply to graphs. The operating system color picker lets users select a predefined color or create their own color. You can also specify a default color in your script. If you omit the default color, Black is selected.

Returns

The color that the user selected in the operating system's color picker.

Arguments

window title The quoted title of the color picker window.

name The name of the default JMP color.

index The number of the default JMP color.

RGBlist The RGB values of the default color.

Pie(*left*, *top*, *right*, *bottom*, *startAngle*, *endAngle*)

Description

Draws a filled pie slice. The two points describe a rectangle, within which is a virtual oval. Only the slice described by the start and end angles is drawn.

Pixel Line To(*h*, *v*)

Description

Draws a one-pixel-wide line from the current pixel location to the location given in pixel coordinates. Set the current pixel location using the `Pixel Origin()` and `Pixel Move To()` functions.

Pixel Move To(*h*, *v*)

Description

Moves the current pixel location to a new location given in pixel coordinates.

Pixel Origin(x, *y*)

Description

Sets the origin, in graph coordinates, for subsequent `Pixel Line To()` or `Pixel Move To()` functions.

Polygon({*x1*, *y1*}, {*x2*, *y2*}, ...)

Polygon(*xMatrix*, *yMatrix*)

Description

Draws a filled polygon defined by the listed points.

`Polygon Area({x1, y1}, {x2, y2}, ...)`

`Polygon Area(xMatrix, yMatrix)`

Description

Calculates the area of the specified polygon.

Examples

```
area = Polygon Area( {0, 0}, {0, 10}, {10, 10}, {10, 0} );
area = Polygon Area( [10 20 30], [10 30 20] );
```

`Polygon Centroid({x1, y1}, {x2, y2}, ...)`

`Polygon Centroid(xMatrix, yMatrix)`

Description

Calculates the centroid of the specified polygon.

Examples

```
{cx, cy} = Polygon Centroid( {0, 0}, {0, 10}, {10, 10}, {10, 0} );
centroid = Polygon Centroid( [10 20 30], [10 30 20] );
```

`Pixel Path(h, v, (path matrix|path text), <fill=Boolean)>, <scale=n>, <orient={n, n})`

Description

Draws a stroke along the given pixel-based path if the fill is 0, or paints the interior of the path if the fill is not 0.

Required Arguments

h, *v* Specifies the horizontal and vertical position.

path matrix Contains three columns for *h*, *v*, and flags for each point in the path. The flag values are 0 for control, 1 for move, 2 for line segment, 3 for cubic Bézier segment, and are negative if the point also closes the path. Each flag (one for each point) can be 0, 1, 2 or 3, (or negative, if it closes the shape).

If you don't specify the path matrix, specify the path text.

path text Supports SVG syntax. The path is scaled (`scale`) and translated about its origin according to the optional parameters (`fill` and `orient`), with the orientation specified in the axis space.

Optional Arguments

`fill(Boolean)` 0 fills the shape. 1 draws an empty shape.

`orient` Rotates the object. For example, an orientation of `{Sqrt( 2 ), Sqrt( 2 )}` draws the shape on a 45-degree angle.

`scale` Sizes the object. For example, a scale of 1.0 draws the shape as it is defined; a scale of 2.0 draws it twice as big; and a scale of 0.5 draws it half as big.

`Pixel Text(<properties>, {h, v}, text, ...)`

Description

Moves to the {*h, v*} pixel position and draws text that the quoted *text* argument specifies.

Required Arguments

h The horizontal pixel position.

v The vertical pixel position.

`text` The quoted string to display at the start and end of the pixels.

Optional Arguments

`center justified` Center justifies the text.

`right justified` Right justifies the text.

`erased` Omits pixels from the edges of the rectangle of text. The text is drawn over the graph content in the background color, erasing the graph within the rectangle to make the text more readable.

`boxed` Displays a box around the text.

`counterclockwise` Rotates the text counterclockwise.

`clockwise` Rotates the text clockwise.

`Rect(left, top, right, bottom, <Fill(Boolean)>)`

`Rect({left, top}, {right, bottom}, <Fill(Boolean)>)`

Description

Draws a rectangle whose diagonal has the coordinates (*left*, *top*) and (*right*, *bottom*). *Fill* is Boolean. If *Fill* is 0, the rectangle is empty. If *Fill* is nonzero, the rectangle is filled with the current fill color. The default value for *Fill* is 0.

`Remove Color Theme(name|{name, <flags>, {color, ...}, <{position, ...}>}>)`

Description

Removes a custom color theme from the global list, either by name or by the full color theme object.

Required Arguments

name The quoted name of the color theme.

color The RGB values for the color.

Optional Arguments

flags A number that represents metadata such as whether the theme is continuous or categorical. Run `Get Color Theme Details("name")` on the color theme and use the flag that is returned.

position A number between 0 and 1. There is one number per color that indicates where on the gradient that color is, where 0 is the beginning and 1 is the end.

Example

```
Remove Color Theme( {"Yellow Blue", 0, {{255, 255, 0}, {0, 0, 255}}, {0.0,
  1.0}} );
```

`RGB Color(r, g, b)`

`RGB Color({r, g, b})`

Description

Converts red, green, and blue values into a JMP color number.

Returns

An integer that is a JMP color number.

Arguments

Red, green, and blue, or a list containing the three RGB values. All values should be between 0 and 1.

`Text(<properties>, ({x, y}|{left, bottom, right, top}), text)`

Description

Draws the quoted string *text* at the given point, either the x and y axes or the left, bottom, right, and top axes.

Properties can be any of several named arguments: `Center Justified`, `Right Justified`, `Erased`, `Boxed`, `Counterclockwise`, `Position`, and named arguments. The position, named arguments, and quoted strings can be added in any order. The position and named arguments apply to all the strings.

`Text Color(n)`

Description

Sets the color for text strings.

`Text Font(fontName, <size>, <bold italic underline strikeout>, <angle>)`

Description

Sets the font for text strings. Use without arguments to get the current font properties. Angle is in degrees clockwise.

Quote the font properties. To apply multiple properties, specify them together, as in `"bold italic"`.

`Text Size(n)`

Description

Sets the font size in points for text strings.

`Transparency(alpha)`

Description

Sets the transparency of the current drawing, with *alpha* between 0 and 1 where 0 is clear (no drawing) and 1 is completely opaque (the default).

Notes

Not all operating systems support transparency.

`V Line(x, <y1, y2>)`

Description

Draws a vertical line at *x* across the graph. If you specify start and end points on the *y*-axis (*y1* and *y2*), the line is drawn vertically at *x* from *y1* to *y2*. You can also draw multiple lines by using a matrix of values in the *x* argument.

`V Size()`

Description

Returns the vertical size of the graphics frame in pixels

`X Function(zExpr, yName, <Min(min), Max(max), Fill(Boolean), Inc(upper bound))`

Description

Draws a plot of the function as the *yName* is varied over the *y*-axis of the graph.

`X Origin()`

Description

Returns the *x*-value for the left edge of the graphics frame.

`X Range()`

Description

Returns the distance from the left to right edge of the display box. For example, `X Origin() + X Range()` is the right edge.

`X Scale(xMin, xMax)`

Description

Sets the range for the horizontal scale. The default value for *xMin* is 0. The default value for *xMax* is 100.

XY Function(*x(t)*, *y(t)*, *t*, Min(*min*), Max(*max*), (Inc(*bound*)|Steps(*min*))

Description

Combines an expression of *x(t)* and *y(t)* to draw an x-y curve for the specified range of parameter *t*. Each time *t* it is varied between Min and Max, the *x* and *y* expressions are evaluated using the current value of *t*.

Note: Either *Inc()* or *Steps()* is needed if the default granularity misses details.

Y Function(*yExpr*, *xName*, <Min(*min*), Max(*max*), Fill(*Boolean*), Inc(*bound*)>)

Description

Draws a plot of the function as the symbol is varied over the *x*-axis of the graph.

Y Origin()

Description

Returns the *y*-value for the bottom edge of the graphics frame.

Y Range()

Description

Returns the distance from the bottom to top edges of a display box. For example, Y Origin() + Y Range() is the top edge.

Y Scale(*yMin*, *yMax*)

Description

Sets the range for the vertical scale. If you do not specify a scale, it defaults to 0, 100.

HTTP Functions

Decode 64 Blob(*string*)

Description

Decodes the quoted string using Base-64 encoding.

Encode 64 Blob(*string*)

Description

Encodes the quoted string using Base-64 encoding.

List Functions

As List(*matrix*)

Description

Converts a matrix into a list. Multi-column matrices are converted to a list of row lists.

Returns

A list.

Argument

matrix Any matrix.

Concat Items({*string1*, *string2*, ...}, <*delimiter*>})

Description

Converts a list of quoted string expressions into one string, with each item separated by a delimiter. The delimiter is a blank if unspecified.

Returns

The concatenated quoted string.

Arguments

string Any quoted string.

delimiter An optional quoted string that is placed between each item. The delimiter can be more than one character long.

Example

```
str1 = "one";
str2 = "two";
str3 = "three";

comb = Concat Items({str1, str2, str3});
   "one two three"
comb = Concat Items({str1, str2, str3}, " : ");
   "one : two : three"
del = ",";
comb = Concat Items({str1, str2, str3}, del);
   "one,two,three"
```

Eval List({*list*})

Description

Evaluates expressions inside *list*.

Returns

A list of the evaluated expressions.

Arguments

list A list of valid JSL expressions.

`Insert(source, item, <position>)`

`Insert(source, key, value)`

Description

Inserts a new *item* into the *source* at the given *position*. If *position* is not given, *item* is added to the end.

For an associative array: Adds the *key* into the *source* associative array and assigns *value* to it. If the *key* exists in *source* already, its value is overwritten with the new *value*.

Required Arguments

source A quoted string, list, vector, expression, or associative array.

item|key Any value to be placed within *source*. For an associative array, *key* might or might not be present in *source*.

value A value to assign to the *key*.

Optional Arguments

position Optional numeric value that specifies the position in *source* to place the *item* into.

`Insert Into(source, item, <position>)`

`Insert Into(source, key, value)`

Description

Inserts a new item into the *source* at the given position in place. The *source* must be an L-value.

Arguments

source A variable that contains a quoted string, list, vector, display box, expression, or associative array.

item|key Any value to be placed within source. For an associative array, *key* might be present in *source*.

position Optional numeric value that specifies the position in *source* to place the *item* into.

value A value to assign to the *key*.

`Is List(x)`

Description

Returns 1 if the evaluated argument is a list, or 0 otherwise.

Items(*string*, <*delimiter*>, <Include Boundary Delimiters(*Boolean*)>)

Description

Returns a list of (possibly empty) quoted sub-strings separated by exactly one of any of the characters specified in the *delimiter* argument.

Arguments

`string` The quoted string that is evaluated.

`Delimiter` (Optional) The character used as a boundary. If *delimiter* is absent, an ASCII space is used. If *delimiter* is the empty quoted string, each character is treated as a separate word.

`Include Boundary Delimiters(Boolean)` (Optional) Returns the empty strings between the boundary and the delimiter.

Example

```
Items( "http://www.jmp.com", ":/." );
   {"http", "", "", "www", "jmp", "com"}
Items(",toy,", ",");
   {"toy"}
Items(",toy,", ",", Include Boundary Delimiters( 1 ));
   {"", "toy", ""}
   /* There is no text between the boundary (the beginning of the quoted string)
      and the comma delimiter, so you get an empty string. The same priniciple
      applies to the delimiter at the end of the string. */
```

List(a, b, c, ...)

{a, b, c, ...}

Description

Constructs a list from a set of items.

N Items(source)

Description

Determines the number of elements in the *source* specified.

Returns

For a list or display box, the number of items in the list or display box is returned. For an associative array, the number of keys is returned. For a matrix, the number of elements in the matrix is returned. For a namespace, the number of functions and variables in the namespace is returned. For a class object, the number of methods, functions, and variables is returned.

Arguments

`source` A list, associative array, matrix, display box, or namespace.

`Remove(source, position, <n>)`

`Remove(source, {items})`

`Remove(source, key)`

Description

Deletes the *n* item(s), starting from the indicated *position*. If *n* is omitted, the item at *position* is deleted. If *position* and *n* are omitted, the item at the end is removed. For an associative array: Deletes the *key* and its value.

Returns

A copy of the *source* with the items deleted.

Arguments

`source` A quoted string, list, vector, expression, or associative array.

`position or key` An integer (or list of integers) that points to a specific item (or items) in the list or expression.

`n` (Optional) An integer that specifies how many items to remove.

`Remove From(source, position, <n>)`

`Remove From(source, key)`

Description

Deletes the *n* item(s) in place, starting from the indicated *position*. If *n* is omitted, the item at *position* is deleted. If *position* and *n* are omitted, the item at the end is removed. For an associative array: Deletes the *key* and its value. The *source* must be an L-value.

Returns

The original *source* with the items deleted.

Arguments

`source` A quoted string, list, vector, expression, display box, or associative array.

`position or key` An integer (or list of integers) that points to a specific item (or items) in the list or expression.

`n` (Optional) An integer that specifies how many items to remove.

`Reverse(source)`

Description

Reverse order of elements or terms in the *source*.

Argument

`source` A quoted string, list, matrix, or expression.

Reverse Into(source)

Description

Reverses the order of elements or terms in *source* in place.

Argument

source A quoted string, list, matrix, display box, or expression.

Shift(source, <n>)

Description

Shifts an item or *n* items from the front to the back of the *source*.

Arguments

source A quoted string, list, matrix, or expression.

n (Optional) An integer that specifies the number of items to shift. Positive values shift items from the beginning of the *source* to the end. Negative values shift items from the end of the *source* to the beginning. The default value is 1.

Shift Into(source, <n>)

Description

Shifts items in place.

Arguments

source A quoted string, list, matrix, display box, or expression.

n (Optional) An integer that specifies the number of items to shift. Positive values shift items from the beginning of the *source* to the end. Negative values shift items from the end of the *source* to the beginning. The default value is 1.

Sort List({list}|expr)

Description

Sort the elements or terms of *list* or *expr*.

Sort List Into({list}|expr)

Description

Sort the elements or terms of *list* or *expr* in place.

`Substitute(`*`string`*`, "substring", "replacementString", ...)`

`Substitute({list}, listItem, replacementItem, ...)`

`Substitute(Expr(sourceExpr), Expr(findExpr), Expr(replacementExpr), ...)`

Description

This is a search and replace function. It searches for a specific portion (second argument) of the source (first argument), and replaces it (third argument).

If a quoted string, finds all matches to *`substring`* in the source *`string`*, and replaces them with the *`replacementString`*.

If a list, finds all matches to *`listItem`* in the source *`list`*, and replaces them with the *`replacementItem`*.

If an expression, finds all matches to the *`findExpr`* in the *`sourceExpr`*, and replaces them with the *`replacementExpr`*. Note that all expressions must be enclosed within an `Expr()` function.

Arguments

`string, list, sourceExpr` A quoted string, list, matrix, or expression in which to perform the substitution.

`substring, listItem, findExpr` A quoted string, list item, or expression to be found in the source string, list, or expression.

`replacementString, replacementItem, replacementExpr` A quoted string, list item, or expression to replace the found string, list item, or expression.

`Substitute Into(`*`string`*`, substring, replacementString, ...)`

`Substitute Into(list, listItem, replacementItem, ...)`

`Substitute Into(Expr(sourceExpr), Expr(findExpr), Expr(replacementExpr), ...)`

Description

This is a search and replace function, identical to `Substitute()` except in place. It searches for a specific portion (second argument) of the source (first argument), and replaces it (third argument). The first argument must be an L-value.

If a quoted string, finds all matches to *`substring`* in the source *`string`*, and replaces them with the *`replacementString`*.

If a list, finds all matches to *`listItem`* in the source *`list`*, and replaces them with the *`replacementItem`*.

If an expression, finds all matches to the *`findExpr`* in the *`sourceExpr`*, and replaces them with the *`replacementExpr`*. Note that all expressions must be enclosed within an `Expr()` function.

Arguments

`string, list, sourceExpr` A quoted string, list, matrix, or expression in which to perform the substitution.

`substring, listItem, findExpr` A quoted string, list item, or expression to be found in the source string, list, or expression.

`replacementString, replacementItem, replacementExpr` A quoted string, list item, or expression to replace the found string, list item, or expression.

`Words(string, <delimiter>)`

Description

Extracts the words from the quoted *string* according to the delimiters given. The default delimiter is ASCII whitespace. If you include a second argument, any and all characters in that argument are considered delimiters. If *delim* is an empty string, each character is treated as a separate word.

Examples

```
Words( "the quick brown fox" );
   {"the","quick","brown","fox"}
Words( "Doe, Jane P.",", ." );
   {"Doe","Jane","P"}
```

MATLAB Integration Functions

JMP provides the following interfaces to access MATLAB. The basic execution model is to first initialize the MATLAB connection, perform the required MATLAB operations, and then terminate the MATLAB connection. In most cases, these functions return 0 if the MATLAB operation was successful or an error code if it was not. If the MATLAB operation is not successful, a message is written to the Log window. The single exception to this is `MATLAB Get( )`, which returns a value.

MATLAB JSL Function Interfaces

`MATLAB Connect( <named arguments> )`

Description

Initializes the MATLAB integration interfaces and returns an active MATLAB integration interface connection as a scriptable object.

Returns

MATLAB scriptable object.

Named Arguments

`Echo(Boolean)` Sends the MATLAB source lines to the JMP log. The default value is true.

`MATLAB Control( <named arguments> )`

Description

Sends control operations to signal MATLAB with external events such as source line echoing.

Returns

None.

Arguments

None.

Named Arguments

`Echo(Boolean)` Global. Echo MATLAB source lines to the JMP log.

`Visible(Boolean)` Global. Determine whether to show or hide the active MATLAB workspace.

`MATLAB Execute( { list of inputs }, { list of outputs }, mCode, <named arguments> )`

Description

Submits the MATLAB code to the active global MATLAB connection given a list of inputs. Upon completion, retrieves a list of outputs.

Returns

0 if successful, otherwise nonzero.

Arguments

`{ list of inputs }` Positional, name list. List of JMP variable names to send to MATLAB as inputs.

`{ list of outputs }` Positional, name list. List of JMP variable names to retrieve from MATLAB as outputs.

`mCode` Positional, quoted string. The MATLAB code to submit.

Named Arguments

`Expand(Boolean)` Perform an Eval Insert on the MATLAB code prior to submission.

`Echo(Boolean)` Echo MATLAB source lines to the JMP log. Default is true.

Example

The following example sends the JMP variables x and y to MATLAB, executes the MATLAB statement *z = x * y*, and then gets the MATLAB variable z and returns it to JMP.

```
MATLAB Init();
x = [1 2 3];
```

```
y = [4 5 6];
MATLAB Execute( {x, y}, {z}, "z = x * y;" );
Show( z );
```

MATLAB Get(name)

Description

Gets named variable from MATLAB to JMP.

Returns

Value of named variable.

Arguments

`name` Positional. The name of a JMP variable to be sent to MATLAB.

Example

Suppose that a matrix named `qbx` and a structure named `df` are present in your MATLAB connection.

```
// get the MATLAB variable qbx and placed it into a JMP variable qbx
qbx = MATLAB Get( qbx );

/* get the MATLAB variable df and placed it into a JMP data table
referenced by df */
df = MATLAB Get( df );
```

Table 2.1 shows what JMP data types can be exchanged with MATLAB using the `MATLAB Get( )` function. Getting lists from MATLAB recursively examines each element of the list and sends each base MATLAB data type. Nested lists are supported.

Table 2.1 JMP and MATLAB Equivalent Data Types for `MATLAB Get( )`

MATLAB Data Type	JMP Data Type
Double	Numeric
Logical	Numeric (0 | 1)
String	String
Integer	Numeric
Date/Time	Numeric
Structure	Data Table
Matrix	Numeric Matrix
Numeric Vector	Numeric Matrix

Table 2.1 JMP and MATLAB Equivalent Data Types for `MATLAB Get( )` *(Continued)*

MATLAB Data Type	JMP Data Type
String Vector	List of Strings
Graph	Picture object

`MATLAB Get Graphics( format )`

Description

Get the last graphic object written to the MATLAB graph display window in a specific graphic format. The graphic object can be returned in several graphic formats.

Returns

JMP Picture object.

Argument

`format` Positional. The format the MATLAB graph display window contents are to be converted to. Valid formats are `png`, `bmp`, `jpeg`, `jpg`, `tiff`, `tif`, and `gif`.

`MATLAB Get Version`

Description

Returns the version number of MATLAB being used with the JMP MATLAB interfaces.

`MATLAB Init( <named arguments> )`

Description

Initializes the MATLAB integration interfaces.

Returns

Return code.

Named Arguments

`Echo(Boolean)` Sends MATLAB source lines to the JMP log. This option is global. The default value is true.

`MATLAB Is Connected()`

Description

Determines whether a MATLAB connection is active.

Returns

1 if connected, otherwise 0.

MATLAB JMP Name To MATLAB Name(name)

Description

Maps a JMP variable name to its corresponding MATLAB variable name using MATLAB variable name naming rules.

Returns

A quoted string, mapped MATLAB variable name.

Argument

`name` Positional. The name of a JMP variable to be sent to MATLAB.

MATLAB Send(name, <named arguments>)

Description

Sends the named variable from JMP to MATLAB.

Returns

0 if successful, otherwise nonzero.

Arguments

`name` Positional. The name of a JMP variable to be sent to MATLAB.

Named Arguments

The following optional arguments apply to data tables only:

`Selected(Boolean)` Send selected rows from the referenced data table to MATLAB.

`Excluded(Boolean)` Send only excluded rows from the referenced data table to MATLAB.

`Labeled(Boolean)` Send only labeled rows from the referenced data table to MATLAB.

`Hidden(Boolean)` Send only hidden rows from the referenced data table to MATLAB.

`Colored(Boolean)` Send only colored rows from the referenced data table to MATLAB.

`Markered(Boolean)` Send only markered rows from the referenced data table to MATLAB.

`Row States(Boolean, <named arguments>)` Send row states from referenced data table to MATLAB by adding an additional data column named “RowState”. Create multiple selections by adding together individual settings. The row state consists of individual settings with the following values:

- Selected = 1
- Excluded = 2
- Labeled = 4
- Hidden = 8
- Colored = 16
- Markered = 32

The following optional, named Row States arguments are supported:

`Colors(Boolean)` Send row colors. Adds additional data column named "RowStateColor".

`Markers(Boolean)` Send row markers. Adds additional data column named "RowStateMarker".

Example

```
// create a matrix, assign it to X, and send the matrix to MATLAB
X = [1 2 3];
ml = MATLAB Send( X );

/* open a data table, assign a reference to it to dt, and send the
data table along with its current row states to MATLAB */
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
ml = MATLAB Send( dt, Row States( 1 ) );
```

Table 2.2 shows what JMP data types can be exchanged with MATLAB using the `MATLAB Send( )` function. Sending lists to MATLAB recursively examines each element of the list and sends each base JMP data type. Nested lists are supported.

Table 2.2 JMP and MATLAB Equivalent Data Types for `MATLAB Send( )`

MATLAB Data Type	JMP Data Type
Double	Numeric
String	String
Double Matrix	Matrix
Structure	Data Table

Example

```
MATLAB Init( );
X = 1;
MATLAB Send( X );
S = "Report Title";
MATLAB Send( S );
M = [1 2 3, 4 5 6, 7 8 9];
MATLAB Send( M );
MATLAB Submit( "
X
S
M
" );
MATLAB Term( );
```

MATLAB Send File(filename, <MATLAB Name(name)>)

Description

Sends a data file to MATLAB.

Arguments

`filename` Specifies a quoted string that identifies the pathname to the file to be sent to MATLAB.

`MATLAB Name` Enables you to change the name of the file sent to MATLAB.

MATLAB Submit File('pathname', <named arguments>)

Description

Submits statements to MATLAB using a file pointed to by pathname.

Returns

0 if successful, otherwise nonzero.

Arguments

`Pathname` Positional, quoted string. Pathname to file containing MATLAB source lines to be executed.

Named Arguments

`Expand(Boolean)` Perform an Eval Insert on the MATLAB code prior to submission.

`Echo(Boolean)` Echo MATLAB source lines to the JMP log. Default is true.

MATLAB Submit(mCode, <named arguments>)

Description

Submits the MATLAB code to the active global MATLAB connection.

Returns

0 if successful, otherwise nonzero.

Arguments

`mCode` Positional, quoted string. The MATLAB code to submit.

Named Arguments

`Expand(Boolean)` Perform an Eval Insert on the MATLAB code prior to submission.

`Echo(Boolean)` Echo MATLAB source lines to the JMP log. Default is true.

Example

The following example creates two vectors of random points and plots them as x and y variables.

```
MATLAB Init();
mc = MATLAB Submit("/[
   x = rand(5);
   fprintf('%f/n', x);
```

```
    y = rand(5);
    fprintf('%f/n', x);
    z = plot(x, y);
]/" );
```

`MATLAB Term();`

Description

Terminates the currently active MATLAB integration interface.

Returns

1 if an active MATLAB connection exists, otherwise returns 0.

Arguments

None.

Matrix Functions

`All(A, ...)`

Returns

1 if all matrix arguments are nonzero; 0 otherwise.

`Any(A, ...)`

Returns

1 if one or more elements of one or more matrices are nonzero; 0 otherwise.

`B Spline Coef(x, Internal Knot Grid, <degree=3>, <Knot End Points=min(x)||max(x)>)`

Description

Finds the matrix of B-spline coefficients for the data in the *x* argument.

Returns

The matrix of B-spline basis coefficients. This matrix can be used as a design matrix in a linear model. The first column of the matrix contains an intercept term.

Arguments

`x` A row or column vector that contains the data.

`Internal Knot Grid` Either a single number that designates the number of desired knot points based on percentiles of *x* or a vector of values that designate the internal knot points. The number of internal knots must be greater than zero and less than or equal to the number of unique elements in *x* minus two.

`degree` A number that indicates the degree of the B-splines. Defaults to 3.

`Knot End Points` A 2x1 matrix that designates the lower and upper knot points. If this argument is not specified, the default lower and upper knot points are the minimum and maximum values of *x*, respectively.

Notes

This function is used in column formulas created by the Functional Data Explorer platform.

`CDF(Y)`

Description

Returns values of the empirical cumulative probability distribution function for *Y*, which can be a vector or a list. Cumulative probability is the proportion of data values less than or equal to the value of *QuantVec*.

Syntax

```
{QuantVec, CumProbVec} = CDF(YVec)
```

`Chol Update(L, V, C)`

Description

If L is the Cholesky root of an *n*x*n* matrix A, then after calling `cholUpdate` L is replaced with the Cholesky root of A+V*C*V' where C is an *m*x*m* symmetric matrix and V is an n*m matrix.

`Cholesky(A)`

Description

Finds the lower Cholesky root (L) of a positive semi-definitive matrix, L*L' = A.

Returns

L (the Cholesky root).

Arguments

`A` a symmetric matrix.

`Correlation(matrix, < <<"Pairwise">, < <<"Shrink">, < <<Freq(vector)>, < <<Weight(vector)>)`

Description

Calculates the correlation matrix of the data in the *matrix* argument.

Returns

The correlation matrix for the specified matrix.

Argument

`matrix` A matrix that contains the data. If the data has *m* rows and *n* columns, the result is an m-by-m matrix.

`"Pairwise"` Uses the pairwise method for missing values rather than the row-wise method.

`"Shrink"` Performs the Schafer-Strimmer shrinkage estimate.

`<<Freq(vector)` A vector that specifies frequencies for the rows of the matrix argument.

`<<Weight(vector)` A vector that specifies weights for the rows of the matrix argument.

Notes

By default, rows are discarded if they contain missing values. If the "Pairwise" option is specified, all pairs of nonmissing values are used in the correlation matrix calculation.

This function uses multithreading if available, so it is recommended for large problems with many rows.

When a column is constant, the correlations for it are 0, and the diagonal element is also 0.

`Covariance(matrix, < <<"Pairwise">, < <<"Shrink">, < <<Freq(vector)>, < <<Weight(vector)>)`

Description

Calculates the covariance matrix of the data in the *matrix* argument.

Returns

The covariance matrix for the specified matrix.

Argument

`matrix` A matrix that contains the data. If the data has *m* rows and *n* columns, the result is an m-by-n matrix.

`"Pairwise"` Uses the pairwise method for missing values rather than the row-wise method.

`"Shrink"` Performs the Schafer-Strimmer shrinkage estimate.

`<<Freq(vector)` A vector that specifies frequencies for the rows of the matrix argument.

`<<Weight(vector)` A vector that specifies weights for the rows of the matrix argument.

Notes

By default, rows are discarded if they contain missing values. If the "Pairwise" option is specified, all pairs of nonmissing values are used in the covariance matrix calculation.

This function uses multithreading if available, so it is recommended for large problems with many rows.

Design(vector, < levelsList | <<levels, <<ElseMissing >)

Description

Creates a design matrix that contains a column of 1s and 0s for each unique value of a *vector* of values.

Returns

A design matrix with a column of 1s and 0s for each unique value of the argument or a list that contains the design matrix and a list of levels.

Argument

`vector` A vector.

`levelsList` An optional list or matrix argument that specifies the levels in the returned matrix.

`<<levels` An optional argument that changes the return value to a list that contains the design matrix and a list of the levels.

`<<ElseMissing` An optional argument that changes the handling of values in the *vector* argument that do not appear in the *levelsList* argument. If this argument is specified, missing values are placed in the design matrix. Otherwise, 0s are placed in the design matrix.

Notes

Missing values in the *levelsList* argument are not ignored. For example:

```
Show( Design ( ., [. 0 1] ),
Design( 0, [. 0 1] ),
Design( 1, [. 0 1] ),
Design( [0 0 1 . 1], [. 0 1] ),
Design( {0, 0, 1, ., 1}, [. 0 1] ) );
   Design(., [. 0 1]) = [1 0 0];
   Design(0, [. 0 1]) = [0 1 0];
   Design(1, [. 0 1]) = [0 0 1];
   Design([0 0 1 . 1], [. 0 1]) =
   [  0 1 0,
      0 1 0,
      0 0 1,
      1 0 0,
      0 0 1];
   Design({0, 0, 1, ., 1}, [. 0 1]) =
   [  0 1 0,
      0 1 0,
      0 0 1,
      1 0 0,
      0 0 1];
```

`Design Last(vector, < levelsList, <<ElseMissing >)`

Description

Creates a design matrix that contains a column of 1s and 0s for all but the last of the unique values of the argument. The last level is coded as a row of 0s.

Returns

A full-rank design matrix or a list that contains the design matrix and a list of levels.

Arguments

`vector` A vector.

`levelsList` An optional list or matrix argument that specifies the levels in the returned matrix. If this argument is specified, the last level in this list or matrix is treated as the last level in the design matrix. Otherwise, the last level is defined as the largest value in the *`vector`* argument.

`<<ElseMissing` An optional argument that changes the handling of values in the *`vector`* argument that do not appear in the *`levelsList`* argument. If this argument is specified, missing values are placed in the design matrix. Otherwise, 0s are placed in the design matrix.

`Design Nom(vector, < levelsList | <<levels, <<ElseMissing >)`

`DesignF(vector, < levelsList | <<levels, <<ElseMissing >)`

Description

Creates a design matrix that contains a column of 1s and 0s for all but the last of the unique values of the argument. The last level is coded as a row of -1s.

Returns

A full-rank design matrix or a list that contains the design matrix and a list of levels.

Argument

`vector` A vector.

`levelsList` An optional list or matrix argument that specifies the levels in the returned matrix. If this argument is specified, the last level in this list or matrix is treated as the last level in the design matrix. Otherwise, the last level is defined as the largest value in the *`vector`* argument.

`<<levels` An optional argument that changes the return value to a list that contains the design matrix and a list of levels.

`<<ElseMissing` An optional argument that changes the handling of values in the *`vector`* argument that do not appear in the *`levelsList`* argument. If this argument is specified, missing values are placed in the design matrix. Otherwise, 0s are placed in the design matrix.

Notes

Missing values in the *`levelsList`* argument are not ignored. For example:

```
Show( Design Nom( ., [. 0 1] ),
Design Nom( 0, [. 0 1] ),
Design Nom( 1, [. 0 1] ),
Design Nom( [0 0 1 . 1], [. 0 1] ),
Design Nom( {0, 0, 1, ., 1}, [. 0 1] ) );
  Design Nom(., [. 0 1]) = [1 0];
  Design Nom(0, [. 0 1]) = [0 1];
  Design Nom(1, [. 0 1]) = [-1 -1];
  Design Nom([0 0 1 . 1], [. 0 1]) = [0 1, 0 1, -1 -1, 1 0, -1 -1];
  Design Nom({0, 0, 1, ., 1}, [. 0 1]) = [0 1, 0 1, -1 -1, 1 0, -1 -1];
```

Design Ord(vector, < levelsList | <<levels, <<ElseMissing >)

Description

Creates a design matrix that contains a column for all but the last of the unique values of the argument. The first level is coded as a row of 0s. Each subsequent (nth) level in the *levelsList* argument is coded as a row of (n-1) 1s and the rest 0s.

Returns

A full-rank design matrix or a list that contains the design matrix and a list of levels.

Argument

`vector` A vector.

`levelsList` An optional list or matrix argument that specifies the levels in the returned matrix.

`<<levels` An optional argument that changes the return value to a list that contains the design matrix and a list of levels.

`<<ElseMissing` An optional argument that changes the handling of values in the *vector* argument that do not appear in the *levelsList* argument. If this argument is specified, missing values are placed in the design matrix. Otherwise, 0s are placed in the design matrix.

Notes

Missing values in the *levelsList* argument are not ignored. For example:

```
Show( Design Ord( ., [. 0 1] ),
Design Ord( 0, [. 0 1] ),
Design Ord( 1, [. 0 1] ),
Design Ord( [0 0 1 . 1], [. 0 1] ),
Design Ord( {0, 0, 1, ., 1}, [. 0 1] ) );
  Design Ord(., [. 0 1]) = [0 0];
  Design Ord(0, [. 0 1]) = [1 0];
  Design Ord(1, [. 0 1]) = [1 1];
  Design Ord([0 0 1 . 1], [. 0 1]) = [1 0, 1 0, 1 1, 0 0, 1 1];
  Design Ord({0, 0, 1, ., 1}, [. 0 1]) = [1 0, 1 0, 1 1, 0 0, 1 1];
```

Det(A)

Description

Determinant of a square matrix.

Returns

The determinant.

Argument

A A square matrix.

Diag(A, <B>)

Description

Creates a diagonal matrix from a square matrix or a vector. If two matrices are provided, concatenates the matrices diagonally.

Returns

The matrix.

Argument

A a matrix or a vector.

Direct Product(A, B)

Description

Direct (Kronecker) product of square matrices or scalars A[i,j]*B.

Returns

The product.

Arguments

A, B Square matrices or scalars.

Distance(x1, x2, <scales>, <powers>)

Description

Produces a matrix of distances between rows of x1 and rows of x2.

Returns

A matrix.

Arguments

x1, x2 Two matrices.

scales Optional argument to customize the scaling of the matrix.

powers Optional argument to customize the powers of the matrix.

`E Div(A, B)`

`A:/B`

Returns

A matrix that contains the element-by-element division of two matrices.

Arguments

`A, B` Two matrices. Both matrix elements must have the same dimensions.

`E Max(A, B)`

Returns

A matrix that contains the element-by-element maximum values of two or more matrices or scalar arguments.

Arguments

`A, B` Two or more matrices or scalars. All matrix elements must have the same dimensions.

`E Min(A, B)`

Returns

A matrix that contains the element-by-element minimum values of two or more matrices or scalar arguments.

Arguments

`A, B` Two or more matrices or scalars. All matrix elements must have the same dimensions.

`E Mult(A, B)`

`A:*B`

Description

Element-by-element multiplication of two matrices.

Returns

A matrix that contains the element-by-element multiplication of two or more matrices or scalar arguments.

Arguments

`A, B` Two or more matrices or scalars. All matrix elements must have the same dimensions.

Eigen(A)

Description

Eigenvalue decomposition.

Returns

A list {M, E} such that E * Diag(M) * E = A'.

Argument

A A symmetric matrix.

Estimate Bartlett Factor Score(dataRow, ManMeans, LatMeans, S, A)

Description

Estimates factor scores from a structural equation model (SEM) using Bartlett's method.

Returns

A row vector of estimated factor scores based on the structural equation model.

Arguments

dataRow A row vector of data values.

ManMeans A vector of model-implied manifest variable means.

LatMeans A vector of model-implied latent variable means.

S The symmetric S matrix from a structural equation model.

A The rectangular A matrix from a structural equation model.

Estimate Factor Score(dataRow, Covariance, ManMeans, LatMeans)

Description

Estimates factor scores from a structural equation model (SEM).

Returns

A row vector of estimated factor scores based on the structural equation model.

Arguments

dataRow A row vector of data values.

Covariance A model-implied variance-covariance matrix.

ManMeans A vector of model-implied manifest variable means.

LatMeans A vector of model-implied latent variable means.

Note

This function is used in the Save Factor Scores option in an Structural Equation Model report.

Fourier Basis Coef(x, Number Pairs, <Period=max(x)-min(x)+1>)

Description

Finds the matrix of Fourier basis coefficients for the data in the *x* argument.

Returns

The matrix of Fourier basis coefficients. This can be used as a design matrix in a linear model. The first column of the matrix contains an intercept term. The remaining columns contain pairs of basis coefficients, where pair *i* is defined as the `sin()` and `cos()` of *i* * (2 * π / Period) * x.

Arguments

`x` A row or column vector that contains the data.

`Number Pairs` The number of `sin()` and `cos()` pairs for the Fourier basis.

`Period` The period for trigonometric functions that make up the Fourier basis.

Notes

This function is used in column formulas created by the Functional Data Explorer platform.

G Inverse(A)

Description

Generalized (Moore-Penrose) matrix inverse.

H Direct Product(A, B)

Description

Horizontal direct product of two square matrices of the same dimension or scalars.

Hough Line Transform(matrix, <NAngle(number)>, <NRadius(number)>)

Description

Takes a matrix of intensities and transforms it in a way that is useful for finding streaks in the matrix. Produces a matrix containing the Hough Line Transform with angles as columns and radiuses as rows.

Argument

`matrix` A matrix that can be derived from the intensities of an image, but is more likely from a semiconductor wafer that may have defects across in a streak due to planarization machines.

`NAngle(number)` Enter the number of the angle to obtain a different sized transform. The default is 180 degrees.

`NRadius(number)` Enter the number of the radius to obtain a different sized transform. The default is sqrt(NRow*nRow+nCol*nCol).

Identity(n)

Description

Creates an *n*-by-*n* identity matrix with ones on the diagonal and zeros elsewhere.

Returns

The matrix.

Argument

`n` An integer.

Index(i, j, <increment>)

i::j

Description

Creates a column matrix whose values range from *i* to *j*.

Returns

The matrix.

Arguments

`i, j` Integers that define the range: *i* is the beginning of the range, *j* is the end.

`increment` Optional argument to change the default increment, which is +1.

Inv()

See "Inverse(A)".

Inv Update(A, X, 1|-1)

Description

Efficiently update an **X**′**X** matrix.

Arguments

`A` The matrix to be updated.

`X` One or more rows to be added to or deleted from the matrix *A*.

`1|-1` The third argument controls whether the row or rows defined in the second argument, *X*, are added to or deleted from the matrix *A*. `1` means to add the row or rows and `-1` means to delete the row or rows.

Inverse(A)

Inv(A)

Description

Returns the matrix inverse. The matrix must be square non-singular.

Is Matrix(x)

Description

Returns 1 if the evaluated argument is a matrix, or 0 otherwise.

J(nrows, <ncols>, <value>)

Description

Creates a matrix of identical values.

Returns

The matrix.

Arguments

`nrows` Number of rows in matrix. If `ncols` is not specified, `nrows` is also used as `ncols`.

`ncols` Number of columns in matrix.

`value` The value used to populate the matrix. If `value` is not specified, 1 is used.

KDTable(matrix)

Description

Returns a table to efficiently look up near neighbors.

Returns

A KDTable object.

Argument

`matrix` A matrix of k-dimensional points. The number of dimensions or points is not limited. Each column in the matrix represents a dimension to the data, and each row represents a data point.

Messages

`<<Distance between rows(row1, row2)` Returns the distance between two the two specified rows in the KDTable. The distance applies to removed and inserted rows as well.

`<<K nearest rows(stop, <position>)` Returns a matrix. Position is a point that is described as a row vector for the coordinate of a row, or as the number of a row. If *`position`* is not specified, returns the *n* nearest rows and distances to all rows. If *`position`* is specified, returns the *n* nearest rows and distances to either a point or a row. *`Stop`* is either *n* or {*n*, *limit*}. The limit parameter limits the number of rows that will be found. It can be specified one of two ways: a number, like 5, means return the 5 nearest rows. A list, like {5,10}, means return up to 5 nearest rows, stopping when the distance of 10 is exceeded. In the second case, the last row may have a distance greater than 10. Since the command continues until it finds the closest row beyond the stop radius, this point is also returned. This can be especially useful if there are no rows within the radius.

<<Remove rows(number | vector) Remove either the row specified by *number* or the rows specified by *vector*. Returns the number of rows that were removed. Rows that were already removed are ignored.

<<Insert rows(number | vector) Re-insert either the row specified by *number* or the rows specified by *vector*. Returns the number of rows that were inserted. Rows that were already inserted are ignored.

Notes

When rows are removed or inserted, the row indices do not change. You can remove and re-insert only rows that are in the KDTable object. If you need different rows, construct a new KDTable.

`Least Squares Solve(y, X, < <<noIntercept, <<weights(OptionalWeightVector), <<method("Sweep"|"GInv")>)`

Description

Computes least squares regression estimates for the assumed model y = X * beta + error.

Returns

A list that contains the matrix `Beta=Inverse(X'X)X'y` and the estimated variance matrix of Beta.

Optional Named Arguments

<<noIntercept Specifies a no-intercept model.

<<weights(optional weight vector) Specifies a vector of weights to perform weighted least squares.

<<method("Sweep" | "GInv") Specifies the method for solving the normal equations. The default Sweep method is more computationally efficient, but you can also specify the generalized inverse ("GInv") method, which is more numerically stable.

`Linear Regression(y, X, < <<noIntercept, <<printToLog, <<weight(OptionalWeightVector), <<freq(OptionalFreqVector)>)`

Description

Fits a linear regression for the assumed model y = X * beta + error.

Returns

A list that contains a vector of the estimates, a vector of the standard error, and a list of diagnostics. The list of diagnostics contains vectors of the *t* statistics and *p*-values for the estimates, as well as the R-Square and adjusted R-Square values for the regression fit.

Optional Named Arguments

<<noIntercept Excludes the intercept.

<<printToLog Prints a summary of the fit to the log.

<<weight(vector) Specifies a vector of weights to perform weighted least squares.

<<freq(vector) Specifies a vector of frequencies for each row of y and X.

Example

```
n = 10;
x = J( n, 1, Random Normal() );
y = 1 + x * 3 + J( n, 1, Random Normal() );
{Estimates, Std_Error, Diagnostics} = Linear Regression( y, x, <<printToLog );
As Table( y || x );
Bivariate( Y( :Col1 ), X( :Col2 ), Fit Line( 1 ) );
```

`Loc(A)`

`Loc(A, item)`

Description

Returns a matrix of subscript positions where *A* is nonzero and nonmissing. For the two-argument function, Loc returns a matrix of positions where *item* is found within *A*. If the first argument is a list, the second argument is required.

Argument

`A` a matrix or a list

`item` the item to be found within the matrix or list *A*

`Loc Max(A)`

Description

Returns the position of the maximum element in a matrix.

Returns

An integer that is the specified position.

Argument

`A` a matrix

`Loc Min(A)`

Description

Returns the position of the minimum element in a matrix.

Returns

An integer that is the specified position.

Argument

`A` a matrix

`Loc NonMissing(matrix, ..., {list}, ...)`

Description

Returns indices of nonmissing rows in matrices or lists. In lists, the function can also return indices of nonempty characters.

Returns

The new matrix or list.

`Loc Sorted(A, B)`

Description

Returns a column vector of subscript positions where the values of *A* have values less than or equal to the values in *B* based on a binary search. *A* must be a matrix sorted in ascending order without missing values.

Returns

The new matrix, which has the same dimensions as *B*. If a value in *B* is less than the first value in *A*, the returned subscript position for that value is 1.

Argument

`A, B` matrices

`Matrix({{x11, ..., x1m}, {x21, ..., 2m}, {...}, {xn1, ..., xnm}})`

`Matrix({x1, ..., xn})`

`Matrix(n, m)`

Description

Constructs an *n*-by-*m* matrix. The following specification methods are available:

- If you specify a list of *n* lists that each contain *m* row values, the matrix is formed by vertically concatenating the evaluated lists. The list items must evaluate to numeric values or row vectors, and the dimensions of the items must be conformable.
- If you specify a single list of *n* items, the return value is an *n*-by-1 column vector. The items of the evaluated list must evaluate to numeric values.
- If you specify two integer arguments, the return value is a matrix of zeros that contains *n* rows and *m* columns.

Examples

```
Matrix({{1, 2, 3}, {4, [5 6]}, {7, 8, 9}});
  [1 2 3, 4 5 6, 7 8 9]
Matrix({{[1 2 3], 4, 5, 6, 7, 8, 9}});
  [1 2 3 4 5 6 7 8 9]
Matrix({2,3+7});
  [2, 10]
Matrix(2,3);
  [0 0 0, 0 0 0]
```

Matrix Mult(A, B)

C=A*B, ...

Description

Matrix multiplication.

Arguments

A, B, ... Two or more matrices, which must be conformable (all matrices after the first one listed must have the same number of rows as the number of columns in the first matrix).

Notes

Matrix Mult() allows only two arguments, while using the * operator enables you to multiply several matrices.

Matrix Rank(A)

Description

Returns the rank of the matrix A.

Mode({list} or matrix)

Description

Selects the most frequent item from a numeric or character list or a numeric matrix. In the event of a tie, the lower value is selected. If multiple arguments are specified, a combination of numeric values and quoted character strings is acceptable.

Arguments

Specify either a list or a matrix.

Multivariate Normal Impute(yVec, meanYvec, symCovMat, colMin, colMax)

Description

Imputes missing values in yVec based on the mean and covariance.

Arguments

yVec The vector of responses.

meanYvec The vector of response means.

symCovMat A symmetric matrix containing the response covariances. If the covariance matrix is not specified, then JMP imputes with means.

colMin A vector of column minimums. Provides lower bounds for the imputations.

colMax A vector of column maximums. Provides upper bounds for the imputations.

`NChooseK Matrix(n, k)`

Description

Returns a matrix of *n* things taken *k* at a time (*n* select *k*).

`N Col(x)`

`N Cols(x)`

Description

Returns the number of columns in either a data table or a matrix.

Argument

x Can be a data table or a matrix.

`Ortho(A, <Centered(0)>, <Scaled(1)>)`

Description

Orthonormalizes the columns of matrix *A* using the Gram Schmidt method. `Centered(0)` makes the columns to sum to zero. `Scaled(1)` makes them unit length.

`Ortho Poly(vector, order)`

Description

Returns orthogonal polynomials for a *vector* of indices representing spacings up to the *order* given.

`P Spline Coef(x, Internal Knot Grid, <degree=3>)`

Description

Finds the matrix of penalized basis spline (P-spline) coefficients for the data in the *x* argument.

Returns

The matrix of P-spline basis coefficients, which is the truncated power basis of the specified *degree*. The truncated power basis of degree p with knots k_1 through k_K is defined as follows:

$$1, x, x^2, \ldots, x^p, (x-k_1)_+^p, \ldots, (x-k_K)_+^p$$

where $(x - k_1)_+$ is the positive part of $(x - k_1)$ and is set to zero for negative values of $(x - k_1)$.

Arguments

x A row or column vector that contains the data.

`Internal Knot Grid` Either a single number that designates the number of desired knot points based on percentiles of *x* or a vector of values that designate the internal knot points.

`degree` A number that indicates the degree of the P-splines. Defaults to 3.

Notes

This function is used in column formulas created by the Functional Data Explorer platform.

`Parallel Assign({thread_local_var = global_var, ...}, matrix[a, b] = expression using a and b)`

Description

Uses multiple threads to assign values to the matrix. Enables you to take advantage of multiple cores on a computer. The function has two arguments.

- The first argument is a list of assignment statements that copies global variables into each thread's local variable list.
- The second argument is an assignment expression with a left-hand-side that is a matrix with one or two prototype indexes and a right hand side that can be any JSL expression using those indexes and the local variables from the list (and in JMP global: variables).

Example

The following example provides read access to the global namespace.

```
a = 42;
x = [1 2 3, 4 5 6, 7 8 9];
Show( Parallel Assign( {}, x[i, j] = global:a ) );
Show( x );
   Parallel Assign({}, x[i,j] = global:a) = 1;
   x =
   [  42 42 42,
      42 42 42,
      42 42 42];
```

`Print Matrix(M, <named arguments>)`

Description

Returns a quoted string that contains a well-formatted matrix. You can use the function, for example, to print the matrix to the log.

Argument

`M` A matrix.

Optional Named Arguments

`<<ignore locale(Boolean)` Set to false (0) to use the decimal separator for your locale. Set to true (1) to always use a period (`.`) as a separator. The default value is false (0).

`<<decimal digits(n)` An integer that specifies the number of digits after the decimal separator to print.

`<<style("style name")` Use one of three available styles: Parseable is a reformatted JSL matrix expression. Latex is formatted for LaTex. If you specify Other, you must define the following three arguments.

`<<separate("character")` Define the separator for concatenated entries.

`<<line begin("character")` Define the beginning line character.

`<<line end("character")` Define the ending line character.

`QR(A)`

Description

Returns the QR decomposition of *A*. Typical usage is `{Q, R} = QR(A)`.

`Rank Index(vector)`

`Rank(vector)`

Description

Returns a vector of indices that, used as a subscript to the original *vector*, sorts the vector by rank. Excludes missing values. Lists of numbers or quoted strings are supported in addition to matrices.

`Ranking(vector)`

Description

Returns a vector of ranks of the values of *vector*, low to high as 1 to *n*, ties arbitrary. Lists of numbers or quoted strings are supported in addition to matrices.

`Ranking Tie(vector)`

Description

Returns a vector of ranks of the values of *vector*, but ranks for ties are averaged. Lists of numbers or quoted strings are supported in addition to matrices.

`Scoring Impute(rowWithMissing, VMat, colMeanVec, colStdDevVec)`

Description

Provides streaming functionality for the Automated Data Imputation (ADI) algorithm.

Returns

Returns the row vector with the missing values imputed using the standard least squares estimation.

Arguments

`rowwithMissing` A row vector that contains missing values.

`VMat` A loading matrix that is produced by the ADI algorithm.

`colMeanVec` A vector of the column means ignoring missing cells.

`colStdDevVec` a vector of the column standard deviations ignoring missing cells.

`Shape(A, nrow, <ncol>, < <<bycol>)`

Description

Reshapes the matrix *A* across rows to the specified dimensions. Each value from the matrix *A* is placed into the reshaped matrix. By default, the values are placed row-by-row.

Returns

The reshaped matrix.

Arguments

`A` A matrix.

`nrow` The number of rows that the new matrix should have.

`ncol` (Optional) The number of columns the new matrix should have.

`<<bycol` (Optional) Specifies that the values be placed into the reshaped matrix column-by-column, instead of row-by-row.

Notes

If *ncol* is not specified, the number of columns is whatever is necessary to fit all of the original values of the matrix into the reshaped matrix.

If a missing value is specified for *nrow*, the number of rows is whatever is necessary to fit all of the original values of the matrix into the reshaped matrix.

If the new matrix is smaller than the original matrix, the extra values are discarded.

If the new matrix is larger than the original matrix, the values are repeated to fill the new matrix.

Examples

```
a = Matrix({ {1, 2, 3}, {4, 5, 6}, {7, 8, 9} });
   [  1 2 3,
      4 5 6,
      7 8 9]

Shape(a, 2);
   [  1 2 3 4 5,
      6 7 8 9 1]

Shape(a, 2, 2);
   [  1 2,
      3 4]

Shape(a, 4, 4);
```

```
[  1 2 3 4,
   5 6 7 8,
   9 1 2 3,
   4 5 6 7]

Shape(a, 4, 4, <<bycol);
[  1 5 9 4,
   2 6 1 5,
   3 7 2 6,
   4 8 3 7]
```

Solve(A, b)

Description

Solves a linear system. In other words, `x=inverse(A)*b`.

Sort Ascending(source)

Description

Returns a copy of a list or matrix *source* with the items in ascending order.

Sort Descending(source)

Description

Returns a copy of a list or matrix *source* with the items in descending order.

Sparse SVD(X, <nSingularValues=min(nRow, nCol)>,<tolerance=1e-10>)

Description

Computes the singular value decomposition of matrix X using the implicitly restarted, partially reorthogonalized Lanczos method for sparse matrices.

Returns

Returns a list (U, M, V) such that U*diag(M)*V is equal to x.

Spline Coef(x, y, lambda)

Description

Returns a five column matrix of the form `knots||a||b||c||d` where `knots` is the unique values in *x*.

`x` is a vector of regressor variables, `y` is the vector of response variables, and `lambda` is the smoothing argument. Larger values for `lambda` result in smoother splines.

Spline Eval(x, coef)

Description

Evaluates the spline predictions using the `coef` matrix in the same form as returned by `SplineCoef()`, in other words, `knots||a||b||c||d`. The `x` argument can be a scalar or a matrix of values to predict. The number of columns of `coef` can be any number greater than 1 and each is used for the next higher power. The powers of `x` are centered at the knot values. For example, the calculation for `coef` of `knots||a||b||c||d` is `j` is such that `knots[j]` is the largest knot smaller than `x`.

`xx = x-knots[j]` is the centered *x* value:

```
result = a[j] + xx * (b[j] + xx * (c[j] + xx * d[j]))
```

The following line is equivalent:

```
result = a[j] + b[j] * xx + c[j] * xx ^ 2 + d[j] * xx ^ 3
```

Spline Smooth(x, y, lambda)

Description

Returns the smoothed predicted values from a spline fit.

`x` is a vector of regressor variables, `y` is the vector of response variables, and `lambda` is the smoothing argument. Larger values for `lambda` result in smoother splines.

SVD(A)

Description

Singular value decomposition.

Sweep(A, <indices>)

Description

Sweeps, or inverts a matrix a partition at a time.

Trace(A)

Description

The trace, or the sum of the diagonal elements of a square matrix.

Transpose(A)

Description

Transposes the rows and columns of the matrix *A*.

Returns

The transposed matrix.

Arguments

A A matrix.

Equivalent Expression

A‘

`V Concat(A, B, ...)`

Description

Vertical concatenation of two or more matrices.

Returns

The new matrix.

Arguments

Two or more matrices.

`V Concat To(A, B, ...)`

Description

Vertical concatenation in place. This is an assignment operator.

Returns

The new matrix.

Arguments

Two or more matrices.

`V Max(matrix)`

Description

Returns a row vector containing the maximum of each column of *matrix*.

`V Mean(matrix)`

Description

Returns a row vector containing the mean of each column of *matrix*.

`V Median(matrix)`

Description

Returns a row vector containing the median of each column of *matrix*.

`V Min(matrix)`

Description

Returns a row vector containing the minimum of each column of *matrix*.

V Quantile(matrix, p)

Description

Returns a row vector containing the p^{th} quantile of each column of *matrix*.

V Standardize(matrix)

Description

Returns a matrix column-standardized to mean = 0 and standard deviation = 1.

V Std(matrix)

Description

Returns a row vector containing the standard deviations of each column of *matrix*.

V Sum(matrix)

Description

Returns a row vector containing the sum of each column of *matrix*.

Varimax(matrix, <norm=1>)

Description

Performs a varimax rotation.

Returns

A list that contains the rotated matrix and the orthogonal rotation matrix.

Arguments

matrix A matrix to be rotated.

norm Specify 1 to perform a normalized rotation, and specify 0 to perform a non-normalized rotation. The default value is 1.

Vec Diag(A)

Description

Creates a vector from the diagonals of a square matrix *A*.

Returns

The new matrix.

Arguments

A A square matrix.

Notes

Using a matrix that is not square results in an error.

`Vec Quadratic(symmetric matrix, rectangular matrix)`

Description

Constructs an *n*-by-*m* matrix. Used in calculation of hat values.

Returns

The new matrix.

Arguments

Two matrices. The first must be symmetric.

Equivalent Expression

```
Vec Diag(X*Sym*X‘)
```

`VPTree(matrix)`

Description

Returns a table that is used for efficiently looking up nearest neighbors. The vantage-point tree algorithm is particularly useful for wide data problems.

Returns

A VPTree object.

Argument

`matrix` A matrix of *k*-dimensional points. The number of dimensions or points is not limited. Each column in the matrix represents a dimension to the data, and each row represents a data point.

Numeric Functions

`Abs(n)`

Description

Calculates the absolute value of *n*.

Returns

Returns a positive number of the same magnitude as the value of *n*.

Argument

`n` Any number.

`Ceiling(n)`

Description

If *n* is not an integer, rounds *n* to the next highest integer.

Returns

Returns the smallest integer greater than or equal to *n*.

Argument

`n` Any number.

`Derivative(expr, {name, ...}, ...)`

Description

Calculates the derivative of the *expr* expression with respect to *name*.

Returns

Returns the derivative.

Arguments

`expr` Any expression. Indirect arguments (for example, `Name Expr`, `Expr`, `Eval`) are supported.

`name` Can be a single variable or a list of variables.

Notes

Adding an additional variable (`Derivative(expr, name, name2)`) takes the second derivative.

`Floor(n)`

Description

If *n* is not an integer, rounds *n* to the next lowest integer.

Returns

Returns the largest integer less than or equal to *n*.

Argument

`n` Any number.

Examples

```
Floor( 2.7 );
  2
Floor( -.5 );
  -1
```

`Integrate(expr, varname, lowLimit, upLimit, <<Tolerance(1e-10), <<StoreInfo({list}), <<StartingValue(val))`

Description

Integrates an expression with respect to a scalar value, using the adaptive quadrature method from Gander and Gautschi (2000).

Arguments

`expr` an expression that defines the integrand.

`varname` the name of the variable of integration. If this variable contains a value, that value specifies a starting value that is used as a typical value to improve the accuracy of the integral.

`lowLimit` specifies the lower limit of integration. To specify negative infinity as the lower limit of integration, set this to missing.

`upLimit` specifies the upper limit of integration. To specify positive infinity as the upper limit of integration, set this to missing.

`StoreInfo` saves diagnostics of the numerical integration routine to the argument of StoreInfo().

`StartingValue` specifies a starting value that is used as a typical value to improve the accuracy of the integral.

`Invert Expr(expr, name)`

Description

Attempts to unfold *expr* around *name*.

`Mod()`

See "Modulo(number, divisor)"

`Modulo(number, divisor)`

`Mod(number, divisor)`

Description

Returns the remainder when *number* is divided by *divisor*.

Examples

```
Modulo( 6, 5 );
  1
```

`Normal Integrate(muVector, sigmaMatrix, expr, x, nStrata, nSim)`

Description

Returns the result of radial-spherical integration for smooth functions of multivariate, normally distributed variables.

Arguments

`muVector` A vector.

`sigmaMatrix` A matrix.

`expr` An expression in terms of the variable *x*.

`x` The variable used in the expression *expr*.

`nStrata` Number of strata.

`nSim` Number of simulations.

`Num Deriv(f(x,...), <parnum=1>)`

Description

Returns the numerical derivative of the *f(x,...)* function with respect to one of its arguments. You can specify that argument as the second argument in the *Num Deriv* function. If no second argument is specified, the derivative is taken with respect to the function's first argument. The derivative is evaluated using numeric values specified in the *f(x,...)* function expression.

Notes

The `Num Deriv()` function might appear not to produce the correct results as seen here:

```
x = 3;
n = Num Deriv( 3 * x ^ 2 );
    // 9.00000000001455
```

The preceding usage is not correct. The function was designed to be used in the Nonlinear platform to differentiate functions for which it does not know the analytic derivatives. The proper usage takes the following form:

```
x = 3;
f = Function( {x}, 3 * x ^ 2 );
n = Num Deriv( f( x ), 1 );
   // 18.000029999854
```

`Num Deriv2(f(x,...))`

Description

Returns the numerical second derivative of the *f(x,...)* function with respect to *x*. The derivative is evaluated using numeric values specified in the *f(x,...)* function expression.

`Round(n, places)`

Description

Rounds *n* to number of decimal *places* given.

`Simplify Expr(expr(expression))`

`Simplify Expr(nameExpr(global))`

Description

Algebraically simplifies an expression

Optimization Functions

Constrained Maximize(expr, {x1(low1, up1), x2(low2, up2), ...}, messages)

Description

Finds the values for the *x* arguments, specified as a list, that maximize the *expr* expression with optional linear constraints. You must either specify lower and upper bounds in parentheses for each argument or with the optional `Set Variable Limit()` message. The *x* arguments can be scalar values or vectors.

In the following messages, `A` is a matrix of coefficients. `x = [x1, x2, ...]` is the vector of arguments. `b` is a vector that forms the right side of the expression.

Messages

`<<Less than EQ({A, b})` Sets the constraint to less than or equal to the specified values `(A*x <= b)`.

`<<Greater Than EQ({A, b})` Sets the constraint to greater than or equal to the specified values `(A*x >= b)`.

`<<Equal To({A, b})` Sets the constraint as equal to the specified values `(A*x = b)`.

`<<Starting Values([x1Start, x2Start, ...])` Specifies a starting point.

`<<Max Iter(int)` An integer that specifies the maximum number of iterations to be performed.

`<<Tolerance(p)` *p* sets the tolerance for the convergence criterion. The default tolerance is 10^{-5}.

`<<Show Details("true")` Returns a list with the final values for (objective value, number of iterations, gradient, and Hessian). Shows the step-by-step results of the optimizer in the log.

`<<SetVariableLimit({low,high})` Specifies vectors for the lower and upper limits for the optimization variables.

Constrained Minimize(expr, {x1(low1, up1), x2(low2, up2), ...}, messages)

Description

Finds the values for the *x* arguments, specified as a list, that minimize the *expr* expression with optional linear constraints. You must either specify lower and upper bounds in parentheses for each argument or with the optional `Set Variable Limit()` message. The *x* arguments can be scalar values or vectors.

In the following messages, `A` is a matrix of coefficients. `x = [x1, x2, ...]` is the vector of arguments. `b` is a vector that forms the right side of the expression.

Messages

`<<Less than EQ({A, b})` Sets the constraint to less than or equal to the specified values `(A*x <= b)`.

<<Greater Than EQ({A, b}) Sets the constraint to greater than or equal to the specified values (A*x >= b).

<<Equal To({A, b}) Sets the constraint as equal to the specified values (A*x = b).

<<Starting Values([x1Start, x2Start, ...]) Specifies a starting point.

<<Max Iter(int) An integer that specifies the maximum number of iterations to be performed.

<<Tolerance(p) *p* sets the tolerance for the convergence criterion. The default tolerance is 10^{-5}.

<<Show Details("true") Returns a list with the final values for (objective value, number of iterations, gradient, and Hessian). Shows the step-by-step results of the optimizer in the log.

<<SetVariableLimit({low,high}) Specifies vectors for the lower and upper limits for the optimization variables.

Desirability(yVector, desireVector, y)

Description

Fits a function to go through the three points, suitable for defining the desirability of a set of response variables (*y*s). *yVector* and *desireVector* are matrices with three values, corresponding to the three points defining the desirability function. The actual function depends on whether the desire values are in the shape of a larger-is-better, smaller-is-better, target, or antitarget.

Returns

The desirability function.

Arguments

yVector Three input values.

desireVector the corresponding three desirability values.

y the value of which to calculate the desirability.

LPSolve(A, b, c, L, U, neq, nle, nge, <slackVars(Boolean)>)

Description

Returns a list containing the decision variables (and slack variables if applicable) in the first list item and the optimal objective function value (if one exists) in the second list item.

Arguments

A A matrix of constraint coefficients.

b A matrix that is a column of right hand side values of the constraints.

c A vector of cost coefficients of the objective function.

L, U Matrices of lower and upper bounds for the variables.

neq The number of equality constraints.

`nle` The number of less than or equal inequalities.

`nge` The number of greater than or equal inequalities.

`slackVars(Boolean)` (Optional) Determines whether the slack variables are returned in addition to the decision variables. The default value is `0`.

Notes

The constraints must be listed as equalities first, less than or equal inequalities next, and greater than or equal inequalities last.

Maximize(expr, {x1(low1, up1), x2(low2, up2), ...}, messages)

Description

Finds the values for the *x* arguments, specified as a list, that maximize the expression *expr*. You can specify lower and upper bounds in parentheses for each argument. Additional arguments for the function enable you to set the maximum number of iterations, tolerance for convergence, and view more details about the optimization. The Newton-Raphson method is used when an analytical derivative is found for the Hessian. Otherwise, the Symmetric-Rank One method (SR1), a quasi-Newton method, is used.

Messages

`<<Max Iter(int)` An integer that specifies the maximum number of iterations to be performed. The default maximum number of iterations is 250.

`<<Tolerance(p)` *p* sets the tolerance for the convergence criterion. The default tolerance is 10^{-8}.

`<<Details("both" | "displaySteps" | "returnDetails")` Specifies what output is returned. If "displaySteps" is specified, step-by-step results of the optimization appear in the Log window. If "returnDetails" is specified, the function returns a list that contains the final values for the objective value, number of iterations, gradient, and Hessian. Specify "both" to get the return value and the results in the Log.

`<<Gradient(exprList)` Specifies a list of expressions that define the analytical gradient that is used for the optimization. Each expression in the list represents a derivative of the expression *expr*.

`<<Hessian(exprList)` Specifies a list of expressions that define the analytical Hessian that is used for the optimization. Each expression in the list represents the upper triangular portion of the Hessian matrix in row-major order.

`<<Method(NR | SR1)` Specifies either the Newton-Raphson (NR) method or the Symmetric-Rank One (SR1) method for the optimization method.

`<<UseNumericDeriv("true")` Specifies that the optimization use a numeric approximation.

`Minimize(expr, {x1(low1, up1), x2(low2, up2), ...}, messages)`

Description

Finds the values for the *x* arguments, specified as a list, that minimize the expression *expr*. You can specify lower and upper bounds in parentheses for each argument. Additional arguments for the function enable you to set the maximum number of iterations, tolerance for convergence, and view more details about the optimization. The Newton-Raphson method is used when an analytical derivative is found for the Hessian. Otherwise, the Symmetric-Rank One method (SR1), a quasi-Newton method, is used.

Messages

`<<Max Iter(int)` An integer that specifies the maximum number of iterations to be performed. The default maximum number of iterations is 250.

`<<Tolerance(p)` *p* sets the tolerance for the convergence criterion. The default tolerance is 10^{-8}.

`<<Details("both" | "displaySteps" | "returnDetails")` Specifies what output is returned. If "displaySteps" is specified, step-by-step results of the optimization appear in the Log window. If "returnDetails" is specified, the function returns a list that contains the final values for the objective value, number of iterations, gradient, and Hessian. Specify "both" to get the return value and the results in the Log.

`<<Gradient(exprList)` Specifies a list of expressions that define the analytical gradient that is used for the optimization. Each expression in the list represents a derivative of the expression *expr*.

`<<Hessian(exprList)` Specifies a list of expressions that define the analytical Hessian that is used for the optimization. Each expression in the list represents the upper triangular portion of the Hessian matrix in row-major order.

`<<Method(NR | SR1)` Specifies either the Newton-Raphson (NR) method or the Symmetric-Rank One (SR1) method for the optimization method.

`<<UseNumericDeriv("true")` Specifies that the optimization use a numeric approximation.

Probability Functions

`Beta Density(x, alpha, beta, <theta=0>, <sigma=1>)`

Description

Returns the probability density function (pdf) evaluated at *x* of the beta distribution. The pdf is parameterized as follows:

$$f(x) = \frac{1}{B(\alpha, \beta)\sigma^{\alpha+\beta-1}}(x-\theta)^{\alpha-1}(\theta+\sigma-x)^{\beta-1}$$

where B(·) is the Beta function.

Arguments

`x` A quantile at which the pdf is evaluated. *x* must be between *theta* and *theta* + *sigma*.

`alpha, beta` Shape parameters α and β, which must both be greater than 0.

`theta` Optional threshold parameter θ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

Notes

The beta distribution is useful for modeling the probabilistic behavior of random variables that are constrained to fall in the interval [0, 1], such as proportions.

`Beta Distribution(x, alpha, beta, <theta=0>, <sigma=1>)`

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the beta distribution. The cdf uses the same parameterization as the `Beta Density()` function.

Arguments

`x` A quantile at which the cdf is evaluated. *x* must be between *theta* and *theta* + *sigma*.

`alpha, beta` Shape parameters α and β, which must both be greater than 0.

`theta` Optional threshold parameter θ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

`Beta Quantile(p, alpha, beta, <theta=0>, <sigma=1>)`

Description

Returns the p^{th} quantile from a beta distribution with shape arguments *alpha* and *beta*. The quantile function does not have a closed form equation.

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`alpha, beta` Shape parameters α and β, which must both be greater than 0.

`theta` Optional threshold parameter θ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

`Cauchy Density(q, <center=0>, <scale=1>)`

Description

Returns the probability density function (pdf) evaluated at *q* of a Cauchy distribution. The pdf is parameterized as follows:

$$f(q) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{q-\mu}{\sigma}\right)^2}$$

Arguments

`q` A quantile at which the pdf is evaluated.

`center` Optional location parameter μ. The default is 0.

`scale` Optional scale parameter, σ, which must be greater than 0. The default is 1.

Cauchy Distribution(q, <center=0>, <scale=1>)

Description

Returns the cumulative distribution function (cdf) probability that a Cauchy distributed random variable is less than *q*. The cdf is parameterized as follows:

$$F(q) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right)$$

Arguments

`q` A quantile at which the cdf is evaluated.

`center` Optional location parameter μ. The default is 0.

`scale` Optional scale parameter, σ, which must be greater than 0. The default is 1.

Cauchy Quantile(p, <center=0>, <scale=1>)

Description

Returns the p^{th} quantile from a Cauchy distribution. The p^{th} quantile is the value for which the probability is *p* that a random value would be less than or equal to *p*. The quantile function is parameterized as follows:

$$F^{-1}(p) = \sigma\tan\left[\pi\left(p+\frac{1}{2}\right)\right] + \mu$$

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`center` Optional location parameter μ. The default is 0.

`scale` Optional scale parameter σ, which must be greater than 0. The default is 1.

ChiSquare Density(q, df, <nc=0>)

Description

Returns the probability density function (pdf) evaluated at *q* of the chi-square distribution. The pdf is parameterized as follows:

$$f(q) = \exp(-\lambda/2) \sum_{r=0}^{\infty} \frac{(\lambda/2)^r}{r!} f_{n+2r}(q)$$

where $f_{n+2r}(q)$ is the density of a central chi-square distribution with $n+2r$ degrees of freedom.

Arguments

q A quantile at which the pdf is evaluated. *q* must be greater than or equal to 0.

df The degrees of freedom *n*, which must be greater than 0.

nc Optional noncentrality parameter λ, which must be nonnegative. The default is 0.

ChiSquare Distribution(q, df, <nc=0>)

Description

Returns cumulative distribution function at quantile *x* for chi-square with *df* degrees of freedom centered at *nc*. The cdf is parameterized as follows:

$$F(q) = \exp(-\lambda/2) \sum_{r=0}^{\infty} \frac{(\lambda/2)^r}{r!} F_{n+2r}(q) \quad ,$$

where $F_{n+2r}(q)$ is the cumulative distribution of a central chi-square distribution with $n+2r$ degrees of freedom.

Arguments

q A quantile at which the cdf is evaluated. *q* must be greater than or equal to 0.

df The degrees of freedom *n*, must be greater than 0.

nc The optional noncentrality parameter λ, must be nonnegative. The default is 0.

ChiSquare Log CDistribution(x, df, <nc=0>)

Description

Returns the log of (1 - value), where value is the cumulative distribution function evaluated at *x* of the chi-square distribution with *df* degrees of freedom and noncentrality parameter *nc*.

ChiSquare Log Density(x, df, <nc=0>)

Description

Returns the log of the value of the probability density function evaluated at *x* of the chi-square distribution with *df* degrees of freedom and noncentrality parameter *nc*.

ChiSquare Log Distribution(x, df, <nc=0>)

Description

Returns the log of the value of the cumulative distribution function evaluated at quantile *x* of the chi-square distribution with *df* degrees of freedom and noncentrality parameter *nc*.

ChiSquare Noncentrality(x, df, prob)

Description

Returns the chi-square distribution noncentrality parameter *nc* that satisfies the following:

prob = ChiSquare Distribution(*x, df, nc*)

Arguments

x A quantile at which the cdf is evaluated.

df The degrees of freedom *n,* which must be greater than 0.

prob The probability of the quantile desired; *prob* must be between 0 and 1.

ChiSquare Quantile(p, df, <nc=0>)

Description

Returns the p^{th} quantile from a chi-square distribution with *df* degrees of freedom, centered at *nc*. The quantile function does not have a closed form equation.

Arguments

p The probability of the quantile desired. *p* must be between 0 and 1.

df The degrees of freedom *n,* which must be greater than 0.

nc Optional noncentrality parameter λ, which must be nonnegative. The default is 0.

Dunnett P Value(q, nTrt, dfe, <lambdaVec=.>)

Description

Returns the *p*-value from Dunnett's multiple comparisons test.

Arguments

q A number that is the test statistic.

nTrt The number of treatments being compared to the control treatment.

dfe The error degrees of freedom.

`lambdaVec` A vector of parameters. If *`lambdaVec`* is missing (`.`), each of the parameters is set to 1/Sqrt(2).

`Dunnett Quantile(1-alpha, nTrt, dfe, <lambdaVec=.>)`

Description

Returns the quantile used in Dunnett's multiple comparisons test.

Arguments

`1-alpha` A number that is the confidence level.

`nTrt` The number of treatments being compared to the control treatment.

`dfe` The error degrees of freedom.

`lambdaVec` A vector of parameters. If *`lambdaVec`* is missing (`.`), each of the parameters is set to 1/Sqrt(2).

`Exp Density(x, <theta=1>)`

Description

Returns the probability density function (pdf) evaluated at *x* of the exponential distribution. The pdf is parameterized as follows:

$$f(x) = \frac{1}{\theta}\exp(-x/\theta)$$

Arguments

`x` A quantile at which the pdf is evaluated. *x* must be greater than or equal to 0.

`theta` Optional scale parameter θ, which must be greater than 0. The default is 1.

`Exp Distribution(x, <theta=1>)`

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the exponential distribution. The cdf is parameterized as follows:

$$F(x) = 1 - \exp(-x/\theta)$$

Arguments

`x` A quantile at which the cdf is evaluated. *x* must be greater than or equal to 0.

`theta` Optional scale parameter θ, which must be greater than 0. The default is 1.

Exp Quantile(p, <theta=1>)

Description

Returns the p^{th} quantile from an exponential distribution with scale parameter *theta*. The quantile function is parameterized as follows:

$$F^{-1}(p) = -\theta\log(1-p)$$

Arguments

p The probability of the quantile desired. *p* must be between 0 and 1.

theta Optional scale parameter θ, which must be greater than 0. The default is 1.

F Density(x, dfnum, dfden, <nc=0>)

Description

Returns the probability density function (pdf) evaluated at *x* for the F distribution with numerator and denominator degrees of freedom *dfnum* and *dfden*, with optional noncentrality parameter *nc*.

$$f(x) = \exp(-\lambda/2)\sum_{r=0}^{\infty}\frac{(\lambda/2)^r}{\mathrm{B}\left(\frac{v_2}{2},\frac{v_1}{2}+r\right)r!}\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}+r}\left(1+\frac{v_1}{v_2}x\right)^{-\left(\frac{v_1+v_2}{2}+r\right)}x^{\frac{v_1}{2}-1+r}$$

where B(·) is the Beta function.

Arguments

x A quantile at which the pdf is evaluated. *x* must be greater than 0.

dfnum The degrees of freedom, v_1, of the chi-square distribution in the numerator of the *F*-distribution. *dfnum* must be greater than 0.

dfden The degrees of freedom, v_2, of the chi-square distribution in the denominator of the *F*-distribution. *dfden* must be greater than 0.

nc Optional noncentrality parameter λ, which must be nonnegative. The default is 0.

F Distribution(x, dfnum, dfden, <nc=0>)

Description

Returns the cumulative distribution function (cdf) evaluated at *x* for the F distribution with numerator and denominator degrees of freedom *dfnum* and *dfden* and noncentrality parameter *nc*.

F Log CDistribution(x, dfnum, dfden, <nc=0>)

Description

Returns the log of (1 - value), where value is the cumulative distribution function evaluated at *x* of the F distribution with numerator and denominator degrees of freedom *dfnum* and *dfden*, with optional noncentrality parameter *nc*.

F Log Density(x, dfnum, dfden, <nc=0>)

Description

Returns the log of the value of the probability density function (pdf) evaluated at *x* for the F distribution with numerator and denominator degrees of freedom *dfnum* and *dfden*, with optional noncentrality parameter *nc*.

F Log Distribution(x, dfnum, dfden, <nc=0>)

Description

Returns the log of the value of the cumulative distribution function (cdf) evaluated at *x* for the F distribution with numerator and denominator degrees of freedom *dfnum* and *dfden* and noncentrality parameter *nc*.

F Noncentrality(x, dfnum, dfden, prob)

Description

Returns the F distribution noncentrality parameter *nc* that satisfies the following:

prob = F Distribution(*x, dfnum, dfden, nc*)

See Also

"F Distribution(x, dfnum, dfden, <nc=0>)"

F Power(alpha, dfh, dfm, d, n)

Description

Returns the power from a given situation involving an *F* test or a *t* test.

Arguments

`alpha` The significance level of the test. *alpha* must be between 0 and 1.

`dfh` The hypothesis degrees of freedom. *dfh* must be greater than 0.

`dfm` The degrees of freedom in the whole model. *dfm* must be greater than 0.

`d` The squared effect size, defined as Δ^2/σ^2. In this equation, σ^2 is the error variance and Δ^2 is defined as follows:

$\Delta^2 = (\bar{x} - \mu)^2$ for a one-sample *t* test

$$\Delta^2 = \frac{(\bar{x}_1 - \bar{x}_2)^2}{4} \text{ for a two-sample } t \text{ test}$$

$$\Delta^2 = \sqrt{\sum_{i=1}^{k} \frac{(\bar{x}_i - \bar{x})^2}{k}} \text{ for a } k\text{-sample } F \text{ test}$$

n The total number of observations. *n* must be greater than *dfm*.

F Quantile(x, dfnum, dfden, <nc=0>)

Description

Returns the p^{th} quantile from the F distribution with numerator and denominator degrees of freedom *dfnum* and *dfden* and noncentrality parameter *nc*.

F Sample Size(alpha, dfh, dfm, d, power)

Description

Returns the sample size from a given situation involving an *F* test or a *t* test.

Arguments

alpha The significance level of the test. *alpha* must be between 0 and 1.

dfh The hypothesis degrees of freedom. *dfh* must be greater than 0.

dfm The degrees of freedom in the whole model. *dfm* must be greater than 0.

d The squared effect size, defined as Δ^2/σ^2. In this equation, σ^2 is the error variance and Δ^2 is defined as follows:

$$\Delta^2 = (\bar{x} - \mu)^2 \text{ for a one-sample } t \text{ test}$$

$$\Delta^2 = \frac{(\bar{x}_1 - \bar{x}_2)^2}{4} \text{ for a two-sample } t \text{ test}$$

$$\Delta^2 = \sqrt{\sum_{i=1}^{k} \frac{(\bar{x}_i - \bar{x})^2}{k}} \text{ for a } k\text{-sample } F \text{ test}$$

power The desired power for the test.

Frechet Density(x, mu, sigma)

Description

Returns the probability density function (pdf) evaluated at *x* of the Fréchet distribution. The pdf is parameterized as follows:

$$f(x) = \exp\left[-\exp\left(-\frac{\log(x)-\mu}{\sigma}\right)\right]\exp\left(-\frac{\log(x)-\mu}{\sigma}\right)\frac{1}{x\sigma}$$

Arguments

`x` A quantile at which the pdf is evaluated. *x* must be greater than 0.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`Frechet Distribution(x, mu, sigma)`

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the Fréchet distribution. The cdf is parameterized as follows:

$$F(x) = \exp\left[-\exp\left(-\frac{\log(x)-\mu}{\sigma}\right)\right]$$

Arguments

`x` A quantile at which the cdf is evaluated. *x* must be greater than 0.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`Frechet Quantile(p, mu, sigma)`

Description

Returns the p^{th} quantile from a Fréchet distribution with location *mu* and scale *sigma*. The quantile function is parameterized as follows:

$$F^{-1}(p) = \exp[-\sigma\log\{-\log(p)\} + \mu]$$

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`Gamma Density(x, <alpha=1>, <scale=1>, <threshold=0>)`

Description

Returns the probability density function (pdf) evaluated at *x* of the Gamma distribution. The pdf is parameterized as follows:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}(x-\theta)^{\alpha-1}\exp(-(x-\theta)/\beta)$$

Arguments

`x` A quantile at which the pdf is evaluated. *x* must be greater than θ.

`alpha` Optional shape parameter α, which must be greater than 0. The default is 1.

`scale` Optional scale parameter β, which must be greater than 0. The default is 1.

`threshold` Optional threshold parameter θ. The default is 0.

`Gamma Distribution(x, <alpha=1>, <scale=1>, <threshold=0>)`

`IGamma(x, <alpha=1, scale=1, threshold=0>)`

Description

Returns the cumulative distribution function (cdf) evaluated at quantile *x* for the gamma distribution with parameters *alpha*, *scale*, and *threshold*.

`Gamma Log CDistribution(x, <alpha=1>, <scale=1>, <threshold=0>)`

Description

Same as `Log(1 - Gamma Distribution(x, alpha))` except that it has a much greater range.

`Gamma Log Density(x, <alpha=1>, <scale=1>, <threshold=0>)`

Description

Same as `Log(Gamma Density(x, alpha))` except that it has a much greater range.

`Gamma Log Distribution(x, <alpha=1>, <scale=1>, <threshold=0>)`

Description

Same as `Log(Gamma Distribution(x, alpha))` except that it has a much greater range.

`Gamma Quantile(p, <alpha=1>, <scale=1>, threshold>)`

Description

Returns the *p*th quantile from the gamma distribution with the *alpha*, *scale*, and *threshold* parameters given.

GenGamma Density(x, mu, sigma, lambda)

Description

Returns the probability density function (pdf) evaluated at *x* of an extended generalized gamma probability distribution. The pdf is parameterized as follows:

$$f(x) = \begin{cases} \frac{|\lambda|}{x\sigma}\phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda \neq 0 \\ \frac{1}{x\sigma}\phi_{nor}(\omega) & \text{if } \lambda = 0 \end{cases}$$

where $\omega = [\log(x) - \mu]/\sigma$. Note that the following is the pdf for the standardized log-gamma variable with shape parameter $\kappa > 0$:

$$\phi_{lg}(z;\kappa) = \frac{1}{\Gamma(\kappa)}\exp[\kappa z - \exp(z)]$$

Note that $\phi_{nor}(\cdot)$ is the standard normal pdf.

Arguments

`x` A quantile at which the pdf is evaluated. *x* must be greater than 0.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ.

GenGamma Distribution(x, mu, sigma, lambda)

Description

Returns the cumulative distribution function (cdf) of the extended generalized gamma distribution. The cdf is parameterized as follows:

$$F(x) = \begin{cases} \Phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda > 0 \\ \Phi_{nor}(\omega) & \text{if } \lambda = 0 \\ 1 - \Phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda < 0 \end{cases}$$

where $\omega = [\log(x) - \mu]/\sigma$. Note that the following is the cdf for the standardized log-gamma variable with shape parameter $\kappa > 0$:

$$\Phi_{lg}(z;\kappa) = \Gamma_I[\exp(z);\kappa]$$

where $\Gamma_I[\cdot]$ denotes the incomplete gamma function. Note that $\Phi_{nor}(\cdot)$ is the standard normal cdf.

Arguments

`x` A quantile at which the cdf is evaluated. *x* must be greater than 0.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ.

`GenGamma Quantile(p, mu, sigma, lambda)`

Description

Returns the p^{th} quantile from an extended generalized gamma distribution with parameters *mu*, *sigma*, and *lambda*. The quantile function does not have a closed form equation.

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ.

`GLog Density(x, mu, sigma, lambda)`

Description

Returns the probability density function (pdf) evaluated at *x* of a generalized logarithmic distribution. The pdf is parameterized as follows:

$$f(x) = \phi\left\{\frac{1}{\sigma}\left[\log\left(\frac{x+\sqrt{x^2+\lambda^2}}{2}\right)-\mu\right]\right\}\frac{x+\sqrt{x^2+\lambda^2}}{\sigma(x^2+\lambda^2+x\sqrt{x^2+\lambda^2})}$$

where $\phi(\cdot)$ is the standard normal pdf.

Arguments

`x` A quantile at which the pdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ, which must be greater than 0.

Notes

When the shape parameter is equal to zero, the distribution reduces to a Lognormal(μ, σ).

GLog Distribution(x, mu, sigma, lambda)

Description

Returns the probability that a generalized logarithmically distribution random variable is less than *x*. The cdf is parameterized as follows:

$$F(x) = \Phi\left\{\frac{1}{\sigma}\left[\log\left(\frac{x+\sqrt{x^2+\lambda^2}}{2}\right)-\mu\right]\right\}$$

where $\Phi(\cdot)$ is the standard normal cdf.

Arguments

`x` A quantile at which the cdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ, which must be greater than 0.

GLog Quantile(p, mu, sigma, lambda)

Description

Returns the p^{th} quantile from a generalized logarithmic distribution.

IGamma()

See "Gamma Distribution(x, <alpha=1>, <scale=1>, <threshold=0>)".

Johnson Sb Density(q, gamma, delta, theta, sigma)

Description

Returns the probability density function (pdf) evaluated at *q* of a Johnson Sb distribution. The pdf is parameterized as follows:

$$f(q) = \phi\left[\gamma+\delta\ln\left(\frac{q-\theta}{\sigma-(q-\theta)}\right)\right]\left(\frac{\delta\sigma}{(q-\theta)(\sigma-(q-\theta))}\right)$$

where $\phi(\cdot)$ is the standard normal pdf.

Arguments

`q` A quantile at which the pdf is evaluated. *q* must be in the interval *theta* to *theta* + *sigma*.

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Scale parameter σ, which must be greater than 0.

Johnson Sb Distribution(q, gamma, delta, theta, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *q* of a Johnson Sb distribution. The pdf is parameterized as follows:

$$F(q) = \Phi\left[\gamma + \delta \ln\left(\frac{q-\theta}{\sigma-(q-\theta)}\right)\right]$$

where $\Phi(\cdot)$ is the standard normal cdf.

Arguments

q A quantile at which the cdf is evaluated. *q* must be in the interval *theta* to *theta* + *sigma*.

gamma Shape parameter γ.

delta Shape parameter δ, which must be greater than 0.

theta Location parameter θ.

sigma Scale parameter σ, which must be greater than 0.

Johnson Sb Quantile(p, gamma, delta, theta, sigma)

Description

Returns the p^{th} quantile from a Johnson Sb distribution.

Arguments

p The probability of the quantile desired. *p* must be between 0 and 1.

gamma Shape parameter γ.

delta Shape parameter δ, which must be greater than 0.

theta Location parameter θ.

sigma Scale parameter σ, which must be greater than 0.

Johnson Sl Density(x, gamma, delta, theta, sigma)

Description

Returns the probability density function (pdf) evaluated at *x* of a Johnson Sl distribution. The pdf is parameterized as follows:

$$f(x) = \frac{\delta}{|x-\theta|}\phi\left[\gamma + \delta \ln\left(\frac{x-\theta}{\sigma}\right)\right]$$

where $\phi(\cdot)$ is the standard normal pdf.

Arguments

x A quantile at which the pdf is evaluated. *x* must be greater than *theta* if *sigma* is 1 and less than *theta* if *sigma* is -1.

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Parameter σ that indicates if the distribution is skewed positively or negatively. `sigma` must be equal to either +1 (skewed positively) or -1 (skewed negatively).

Johnson Sl Distribution(q, gamma, delta, theta, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *q* of a Johnson Sl distribution.

$$F(x) = \begin{cases} \Phi\left[\gamma + \delta \ln\left(\frac{x-\theta}{\sigma}\right)\right], & \sigma = 1 \\ 1 - \Phi\left[\gamma + \delta \ln\left(\frac{x-\theta}{\sigma}\right)\right], & \sigma = -1 \end{cases}$$

where $\Phi(\cdot)$ is the standard normal cdf.

Arguments

`q` A quantile at which the cdf is evaluated. *q* must be greater than *theta* if *sigma* is 1 and less than *theta* if *sigma* is -1.

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Parameter σ that defines if the distribution is skewed positively or negatively. `Sigma` must be equal to either +1 (skewed positively) or -1 (skewed negatively).

Johnson Sl Quantile(p, gamma, delta, theta, sigma)

Description

Returns the p^{th} quantile from a Johnson Sl distribution.

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Parameter σ that defines if the distribution is skewed positively or negatively. `Sigma` must be equal to either +1 (skewed positively) or -1 (skewed negatively).

Johnson Su Density(x, gamma, delta, theta, sigma)

Description

Returns the probability density function (pdf) evaluated at *x* of a Johnson Su distribution. The pdf is parameterized as follows:

$$f(x) = \frac{\delta}{\sigma}\left[1 + \left(\frac{x-\theta}{\sigma}\right)^2\right]^{-1/2} \phi\left[\gamma + \delta \sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)\right]$$

where $\phi(\cdot)$ is the standard normal pdf.

Arguments

`x` A quantile at which the pdf is evaluated.

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Scale parameter σ, which must be greater than 0.

Johnson Su Distribution(q, gamma, delta, theta, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *q* of a Johnson Su distribution. The cdf is parameterized as follows:

$$F(x) = \Phi\left[\gamma + \delta \sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)\right]$$

where $\Phi(\cdot)$ is the standard normal cdf.

Arguments

`q` A quantile at which the cdf is evaluated.

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Scale parameter σ, which must be greater than 0.

Johnson Su Quantile(p, gamma, delta, theta, sigma)

Description

Returns the p^{th} quantile from a Johnson Su distribution.

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Scale parameter σ, which must be greater than 0.

LEV Density(x, mu, sigma)

Description

Returns the probability density function (pdf) evaluated at *x* of the largest extreme value distribution with location *mu* and scale *sigma*. The pdf is parameterized as follows:

$$f(x) = \frac{1}{\sigma}\exp\left[-\frac{x-\mu}{\sigma} - \exp\left(-\frac{x-\mu}{\sigma}\right)\right]$$

Arguments

`x` A quantile at which the pdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

LEV Distribution(x, mu, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the largest extreme value distribution with location *mu* and scale *sigma*. The cdf is parameterized as follows:

$$F(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]$$

Arguments

`x` A quantile at which the cdf is evaluated. *x* must be greater than *sigma*.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

LEV Quantile(p, mu, sigma)

Description

Returns the p^{th} quantile from a largest extreme value distribution with location *mu* and scale *sigma*. The quantile function is parameterized as follows:

$$F^{-1}(p) = -\sigma\log(-\log(p)) + \mu$$

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

LogGenGamma Density(x, mu, sigma, lambda)

Description

Returns the probability density function (pdf) evaluated at *x* of a log generalized gamma probability distribution with parameters *mu*, *sigma*, and *lambda*. The pdf is parameterized as follows:

$$f(x) = \begin{cases} \frac{|\lambda|}{\sigma}\phi_{lg}[\lambda\omega + \log(\lambda^{-2});\lambda^{-2}] & \text{if } \lambda \neq 0 \\ \frac{1}{\sigma}\phi_{nor}(\omega) & \text{if } \lambda = 0 \end{cases}$$

where $\omega = [x - \mu]/\sigma$. Note that the following is the pdf for the log-gamma variable with shape parameter $\kappa > 0$:

$$\phi_{lg}(z;\kappa) = \frac{1}{\Gamma(\kappa)}\exp[\kappa z - \exp(z)]$$

Note that $\phi_{nor}(\cdot)$ is the standard normal pdf.

Arguments

`x` A quantile at which the pdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ.

LogGenGamma Distribution(x, mu, sigma, lambda)

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the log generalized gamma distributed random variable (with parameters *mu*, *sigma*, and *lambda*) The cdf is parameterized as follows:

$$F(x) = \begin{cases} \Phi_{lg}[\lambda\omega + \log(\lambda^{-2});\lambda^{-2}] & \text{if } \lambda > 0 \\ \Phi_{nor}(\omega) & \text{if } \lambda = 0 \\ 1 - \Phi_{lg}[\lambda\omega + \log(\lambda^{-2}); \quad \lambda^{-2}] & \text{if } \lambda < 0 \end{cases}$$

where $\omega = [x - \mu]/\sigma$. Note that the following is the cdf for the log-gamma variable with shape parameter $\kappa > 0$:

$$\Phi_{lg}(z;\kappa) = \Gamma_I[\exp(z);\kappa]$$

where $\Gamma_I[\cdot]$ denotes the incomplete gamma function. Note that $\Phi_{nor}(\cdot)$ is the standard normal cdf.

Arguments

`x` A quantile at which the cdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ.

`LogGenGamma Quantile(p, mu, sigma, lambda)`

Description

Returns the p^{th} quantile from a log generalized gamma distribution.

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ.

`Logistic Density(x, mu, sigma)`

Description

Returns the probability density function (pdf) evaluated at *x* of a logistic distribution with location *mu* and scale *sigma*. The pdf is parameterized as follows:

$$f(x) = \frac{1}{\sigma}\frac{\exp\left(-\frac{x-\mu}{\sigma}\right)}{\left[1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right]^2}$$

Arguments

`x` A quantile at which the pdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`Logistic Distribution(x, mu, sigma)`

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the logistic distribution with location *mu* and scale *sigma*. The cdf is parameterized as follows:

$$F(x) = \frac{1}{\left[1 + \exp\left(-\frac{x-\mu}{\sigma}\right)\right]}$$

Arguments

`x` A quantile at which the cdf is evaluated. *x* must be greater than σ.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`Logistic Quantile(p, mu, sigma)`

Description

Returns the p^{th} quantile from a logistic distribution with location *mu* and scale *sigma*. The quantile function is parameterized as follows:

$$F^{-1}(p) = -\sigma\log\left(\frac{1}{p} - 1\right) + \mu$$

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`Loglogistic Density(x, mu, sigma)`

Description

Returns the probability density function (pdf) evaluated at *x* of a loglogistic distribution with location *mu* and scale *sigma*. The pdf is parameterized as follows:

$$f(x) = \frac{1}{x\sigma}\frac{\exp\left(\frac{\log(x)-\mu}{\sigma}\right)}{\left[1 + \exp\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^2}$$

Arguments

`x` A quantile at which the pdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

Loglogistic Distribution(x, mu, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the loglogistic distribution with location *mu* and scale *sigma*. The cdf is parameterized as follows:

$$F(x) = \frac{1}{1 + \exp\left(-\frac{\log(x) - \mu}{\sigma}\right)}$$

Arguments

x A quantile at which the cdf is evaluated.

mu The location parameter μ.

sigma The scale parameter σ, which must be greater than 0.

Loglogistic Quantile(p, mu, sigma)

Description

Returns the p^{th} quantile from a loglogistic distribution with location *mu* and scale *sigma*. The quantile function is parameterized as follows:

$$F^{-1}(p) = \exp\left[-\sigma\log\left(\frac{1}{p} - 1\right) + \mu\right]$$

Arguments

p The probability of the quantile desired. *p* must be between 0 and 1.

mu The location parameter μ.

sigma The scale parameter σ, which must be greater than 0.

Lognormal Density(x, mu, sigma)

Description

Returns the probability density function (pdf) evaluated at *x* of a lognormal distribution with location *mu* and scale *sigma*. The pdf is parameterized as follows:

$$f(x) = \frac{1}{x}\phi\left[\frac{\log(x) - \mu}{\sigma}\right]$$

where $\phi(\cdot)$ is the standard normal pdf.

Arguments

x A quantile at which the pdf is evaluated. *x* must be greater than or equal to 0.

mu The location parameter μ.

sigma The scale parameter σ, which must be greater than 0.

Lognormal Distribution(x, mu, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of a lognormal distribution with location *mu* and scale *sigma*. The cdf is parameterized as follows:

$$F(x) = \Phi\left[\frac{\log(x) - \mu}{\sigma}\right]$$

where $\Phi(\cdot)$ is the standard normal cdf.

Arguments

x A quantile at which the pdf is evaluated. *x* must be greater than or equal to 0.

mu The location parameter μ.

sigma The scale parameter σ, which must be greater than 0.

Lognormal Quantile(x, mu, sigma)

Description

Returns the p^{th} quantile of a lognormal distribution with location *mu* and scale *sigma*.

Normal Biv Distribution(x, y, r, <mu1>, <s1>, <mu2>, <s2>)

Description

Computes the probability that an observation (X, Y) is less than or equal to (*x, y*) with correlation coefficient *r* where X is individually normally distributed with mean *mu1* and standard deviation *s1* and Y is individually normally distributed with mean *mu2* and standard deviation *s2*. If *mu1*, *s1*, *mu2*, and *s2* are not given, the function assumes the standard normal bivariate distribution with *mu1*=0, *s1*=1, *mu2*=0, and *s2*=1.

Normal Density(x, <mean=0>, <stddev=1>)

Description

Returns the probability density function (pdf) evaluated at *x* for the normal distribution with *mean* and *stddev*. The pdf is parameterized as follows:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Arguments

x A quantile at which the pdf is evaluated.

mu Optional location parameter μ. The default is 0.

sigma Optional scale parameter σ, which must be greater than 0. The default is 1.

Notes

The normal distribution is bell shaped and symmetrical.

`Normal Distribution(x, <mean=0>, <stddev=1>)`

Description

Returns the cumulative distribution function (cdf) evaluated at *x* for the normal distribution with *mean* and *stddev*. The cdf is parameterized as follows:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Note that $\Phi(\cdot)$ is the standard normal cdf, defined as follows:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{t^2}{2}\right) dt$$

Arguments

`x` A quantile at which the pdf is evaluated.

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

`Normal Log CDistribution(x, <mean=0>, <std dev=1>)`

Description

Returns 1 - log (value) of the distribution function at quantile *x* for the normal distribution.

`Normal Log Density(x, <mean=0>, <stddev=1>)`

Description

Returns the log of the value of the density function at quantile *x* for the normal distribution with *mean* and *stddev*. The default *mean* is 0. The default *stddev* is 1.

`Normal Log Distribution(x, <mean=0>, <std dev=1>)`

Description

Returns the log of the value of the distribution function at quantile *x* for the normal distribution.

Normal Mixture Density(q, mean, stdev, probability)

Description

Returns the density at *q* of a normal mixture distribution with group means *mean*, group standard deviations *stdev*, and group probabilities *probability*. The *mean*, *stdev*, and *probability* arguments are all vectors of the same size.

Normal Mixture Distribution(q, mean, stdev, probability)

Description

Returns the probability that a normal mixture distributed variable with group means *mean*, group standard deviations *stdev*, and group probabilities *probability* is less than *q*. The *mean*, *stdev*, and *probability* arguments are all vectors of the same size.

Normal Mixture Quantile(p, mean, stdev, probability)

Description

Returns the p^{th} quantile, the values for which the probability is *p* that a random value would be lower. The *mean*, *stdev*, and *probability* arguments are all vectors of the same size.

Normal Quantile(p, <mean=0>, <stddev=1>)

Probit(p, <mean=0>, <stddev=1>)

Description

Returns the p^{th} quantile from the normal distribution with *mean* and *stddev*. The default *mean* is 0. the default *stddev* is 1.

Probit()

See "Normal Quantile(p, <mean=0>, <stddev=1>)".

SEV Density(x, mu, sigma)

Description

Returns the probability density function (pdf) evaluated at *x* of the smallest extreme distribution with location *mu* and scale *sigma*. The pdf is parameterized as follows:

$$f(x) = \frac{1}{\sigma}\exp\left[\frac{x-\mu}{\sigma} - \exp\left(\frac{x-\mu}{\sigma}\right)\right]$$

Arguments

x A quantile at which the pdf is evaluated.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

SEV Distribution(x, mu, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the smallest extreme distribution with location *mu* and scale *sigma*. The cdf is parameterized as follows:

$$F(x) = 1 - \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right]$$

Arguments

`x` A quantile at which the cdf is evaluated. *x* must be greater than σ.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

SEV Quantile(p, mu, sigma)

Description

Returns the p^{th} quantile of the smallest extreme distribution with location *mu* and scale *sigma*. The quantile function is parameterized as follows:

$$F^{-1}(p) = \sigma\log[-\log(1-p)] + \mu$$

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

SHASH Density(x, gamma, delta, theta, sigma)

Description

Returns the probability density function (pdf) evaluated at *x* of a sinh-arcsinh (SHASH) distribution. The pdf is parameterized as follows:

$$f(x) = \frac{\delta\cosh(w)}{\sqrt{\sigma^2 + (x-\theta)^2}}\phi[\sinh(w)]$$

where

$\phi(\cdot)$ is the standard normal pdf

$$w = \gamma + \delta \sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)$$

Arguments

x A quantile at which the pdf is evaluated.

gamma The shape parameter γ.

delta The shape parameter δ, which must be greater than 0.

theta The location parameter θ.

sigma The scale parameter σ, which must be greater than 0.

SHASH Distribution(x, gamma, delta, theta, sigma)

Description

Returns the cumulative distribution function (cdf) evaluated at *x* of the sinh-arcsinh (SHASH) distribution. The cdf is parameterized as follows:

$$F(x) = \Phi\left[\sinh\left(\gamma + \delta \sinh^{-1}\left(\frac{x-\theta}{\sigma}\right)\right)\right]$$

where $\Phi(\cdot)$ is the standard normal cdf.

Arguments

x A quantile at which the cdf is evaluated.

gamma The shape parameter γ.

delta The shape parameter δ, which must be greater than 0.

theta The location parameter θ.

sigma The scale parameter σ, which must be greater than 0.

SHASH Quantile(p, gamma, delta, theta, sigma)

Description

Returns the p^{th} quantile from a sinh-arcsinh (SHASH) distribution (with parameters *gamma*, *delta*, *theta*, and *sigma*).

Arguments

p The probability of the quantile desired. *p* must be between 0 and 1.

gamma The shape parameter γ.

delta The shape parameter δ, which must be greater than 0.

theta The location parameter θ.

sigma The scale parameter σ, which must be greater than 0.

Students t Density()

See "t Density(x, df, <nc=0>)".

Students t Distribution()

See "t Distribution(q, df, <nc=0>)".

Students t Quantile()

See "t Quantile(p, df, <nc=0>)".

t Density(x, df, <nc=0>)

Students t Density(x, df, <nc=0>)

Description

Returns the probability density function (pdf) evaluated at *x* of the Student's *t* distribution with degrees of freedom *df*. The pdf is parameterized as follows:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \left[1 + \frac{x^2}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)}$$

Arguments

x A quantile at which the pdf is evaluated.

df The degrees of freedom ν, which must be greater than or equal to 1.

nc The optional noncentrality parameter λ, which must be nonnegative. The default is 0.

t Distribution(q, df, <nc=0>)

Students t Distribution(q, df, <nc=0>)

Description

Returns the probability that a Student's *t* distributed random variable is less than *q*. *nc* defaults to 0.

t Log CDistribution(x, df, <nc=0>)

Description

Returns 1 - log (value) of the normal distribution function at quantile *x* for the *t* distribution.

t Log Density(x, df, <nc=0>)

Description

Returns the log of the value of the density function at quantile *x* for the *t* distribution.

t Log Distribution(x, df, <nc=0>)

Description

Returns the log of the value of the distribution function at quantile *x* for the *t* distribution.

t Noncentrality(x, df, prob)

Description

Returns the *t* distribution noncentrality parameter *nc* that satisfies the following:

prob = T Distribution(*x, df, nc*)

t Quantile(p, df, <nc=0>)

Students t Quantile(p, df, <nc=0>)

Description

Returns the p^{th} quantile from the Student's *t* distribution with degrees of freedom *df*. *nc* defaults to 0.

Tukey HSD P Value(q, n, dfe)

Description

Returns the *p*-value from Tukey's HSD multiple comparisons test.

Arguments

q The test statistic. The test statistic that is specified is Tukey's adjusted critical value, which is the quantile of Tukey's studentized range distribution divided by the square root of 2.

n The number of groups in the study.

dfe The error degrees of freedom, based on the total study sample.

Tukey HSD Quantile(1-alpha, n, dfe)

Description

Returns the quantile used in Tukey's HSD multiple comparisons test. The quantile that is returned is Tukey's adjusted critical value, which is the quantile of Tukey's studentized range distribution divided by the square root of 2.

Arguments

1-alpha The confidence level.

`n` The number of groups in the study.

`dfe` The error degrees of freedom, based on the total study sample.

`Weibull Density(x, shape, <scale=1>, <threshold=0>)`

Description

Returns the probability density function (pdf) evaluated at *x* of the Weibull distribution. The pdf is parameterized as follows:

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x-\theta}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{x-\theta}{\alpha}\right)^{\beta}\right]$$

Arguments

`x` A quantile the pdf is evaluated at. *x* must be greater than *threshold*.

`shape` Shape parameter β, which must be greater than 0.

`scale` Optional scale parameter α, which must be greater than 0. The default is 1.

`threshold` Optional threshold parameter θ. The default is 0.

`Weibull Distribution(x, shape, <scale=1>, <threshold=0>)`

Description

Returns the cumulative distribution function (cdf) at *x* of the Weibull distribution. The cdf is parameterized as follows:

$$F(x) = 1 - \exp\left[-\left(\frac{x-\theta}{\alpha}\right)^{\beta}\right]$$

Arguments

`x` A quantile at which the pdf is evaluated. *x* must be greater than *threshold*.

`shape` Shape parameter β, which must be greater than 0.

`scale` Optional scale parameter α, which must be greater than 0. The default is 1.

`threshold` Optional threshold parameter θ. The default is 0.

`Weibull Quantile(p, shape, <scale=1>, <threshold=0>)`

Description

Returns the p^{th} quantile from the Weibull distribution with the parameters given. The quantile function is calculated as follows:

$$F^{-1}(p) = \alpha[\ln(1-p)]^{\frac{1}{\beta}} + \theta$$

Arguments

`p` The probability of the quantile desired. *p* must be between 0 and 1.

`shape` Shape parameter β, which must be greater than 0.

`scale` Optional scale parameter α, which must be greater than 0. The default is 1.

`threshold` Optional threshold parameter θ. The default is 0.

Programming Functions

As Boolean(x)

Description

Evaluates a JSL expression and returns a JSL Boolean value for use with JSON data.

Example

```
x = 45;
b = As Boolean( x > 2 );
Show( b );
   b = true;
```

As Column(name)

As Column(dt, name)

:name

dt:name

Description

This scoping operator forces *name* to be evaluated as a data table column in the current data table (or the table given by the optional data table reference argument, *dt*) rather than as a global variable.

Arguments

`name` Variable name.

`dt` The data table reference

Notes

`:name` refers to a column name in the current data table. You can also specify which data table to refer to by use `dt:name`.

As Constant(expr)

Description

Evaluates an expression once to create a value that does not change after it is computed.

Returns

The result of the evaluation.

Argument

`expr` Any JSL expression.

Notes

A few platforms that can save prediction columns to a data table use `As Constant()`. The function is wrapped around the part of the formula that is constant across all rows. The argument is evaluated for the first row and then the result is used without re-evaluation for subsequent rows.

`As Global(name)`

`::name`

Description

This scoping operator forces *`name`* to be evaluated as a global variable rather than as a data table column.

Arguments

`name` Variable name.

`As List(matrix)`

See "As List(`matrix`)".

`As Name(`*`string`*`)`

Description

Evaluates argument as a quoted string and changes it into a name.

Returns

A name.

`As Namespace(name)`

Description

Accesses the specified namespace. An error is thrown if no such namespace exists.

Returns

The namespace.

Arguments

`name` Unquoted name of a defined namespace.

`As Scoped(namespace, variable)`

`namespace:variable`

Description

Accesses the specified *variable* within the specified *namespace*.

Returns

The value of the variable, or an error the scoped variable is not found.

Arguments

`namespace` The name of a defined namespace.

`variable` A variable defined within *namespace*.

`Associative Array({key, value}, ...)`

`Associative Array(keys, values)`

Description

Creates an associative array (also known as a dictionary or hash map).

Returns

An associative array object.

Arguments

Either list of key-value pairs; or a list, matrix, or data table column that contains keys followed by a list, matrix, or data table column, respectively, that contains the corresponding values.

`Class Exists(class)`

Description

Returns a value indicating whether a class definition represented by the class reference is a defined class.

Returns

0 or 1.

Argument

`class` A quoted string representation of the name of a defined class or reference to an instantiated class object.

`Clear Globals(<name>, <name>, ...)`

Description

Clears the values for all global symbols. Symbols in any scope other than global are not affected. If one or more names are specified, only those global symbols are cleared.

Returns

Null.

Optional Arguments

name Any global variable name(s).

See Also

"Clear Symbols(<name>, <name>, ...)"

`Clear Log()`

Description

Empties the log.

`Clear Symbols(<name>, <name>, ...)`

Description

Clear the values for all symbols in any and all scopes. If one or more names are specified, only those symbols are cleared.

Returns

Null.

Optional Arguments

name Any global variable name(s).

See Also

"Clear Globals(<name>, <name>, ...)"

`Close Log()`

Description

Closes the log.

`Define Class("class name", <Base Class( "base class name", <"base class name", ...> ),> <Show( All(Boolean) ) | Show( <Members(Boolean),> <Methods(Boolean),> <Functions(Boolean)> ),> <Assignment Statements>)`

Description

Defines a new class object.

Example

```
Define Class(
   "aa",
   _init_ = Method( {} ); x = 1; m1 = Method( {a, b}, a * b )
);
```

Delete Classes(<Force(Boolean)>, < <class>, ...>)

Description

Deletes all currently defined classes.

Optional Arguments

`Force(Boolean)` Deletes the class or classes even if they are in use.

`class` Specifies the classes to delete. You can specify more than one class. This argument can be a quoted string representation of the name of a defined class or a reference to an instantiated class object.

Delete Globals(<name>, <name>, ...)

Description

Deletes all global symbols, except global symbols that are locked. Symbols in any scope other than global are not affected. If one or more names are specified, only those global symbols are cleared.

Optional Arguments

`name` Any global variable name(s).

See Also

"Delete Symbols(<name>, <name>, ...)"

Delete Namespaces(<Force(Boolean expression)>, < <namespace reference>, ...>)

Delete(<Force(Boolean expression)>, < <namespace reference>, ...>)

Description

Deletes all currently defined namespaces or one or more specific namespaces.

Optional Arguments

`Force(Boolean expression)` Deletes the namespace even if it's in use.

`namespace reference` Specifies the namespaces to delete. You can specify more than one namespace reference.

Notes

- When you delete a namespace that contains locked namespaces, an error appears in the log. Use the `Force()` argument to delete the locked namespaces.
- With no arguments, `Delete Namespaces()` ignores locked namespaces.

Delete Symbols(<name>, <name>, ...)

Description

Deletes all symbols in any and all scopes. If one or more names are specified, only those symbols are deleted.

Optional Arguments

`name` Any global variable name(s).

See Also

"Delete Globals(<name>, <name>, ...)"

Eval(expr)

Description

Evaluates `expr`, and then evaluates the result of `expr` (unquoting).

Returns

The result of the evaluation.

Argument

`expr` Any JSL expression.

Eval Insert(*string*, <startDel>, <endDel>, < <<Use Locale(1) >)

Description

Allows for multiple substitutions.

Returns

The result.

Arguments

`string` A quoted string with embedded expressions.

`startDel` Optional starting delimiter. The default value is ^.

`endDel` optional ending delimited. The default value is the starting delimiter.

`Use Locale(1)` Optional argument that preserves locale-specific numeric formatting.

Eval Insert Into(*string*, <startDel>, <endDel>)

Description

Allows for multiple substitutions in place. The same operation as in `Eval Insert` is performed, and the result is placed into the quoted *`string`*.

Returns

The result.

Arguments

`string` A quoted string variable that contains a string with embedded expressions.

`startDel` Optional starting delimiter. The default value is ^.

`endDel` optional ending delimited. The default value is the starting delimiter.

Eval List

See "Eval List({list})".

Exit(<NoSave>)

Quit(<NoSave>)

Description

Exits JMP.

Returns

Void.

Arguments

`NoSave` Optional, named command; exits JMP without prompting to save any open files. This command is *not* case-sensitive, and spaces are optional.

First(expr, <expr>, ...)

Description

Evaluates all expressions provided as arguments.

Returns

Only the result of the first evaluated expression.

Arguments

`expr` Any valid JSL expression.

Function({arguments}, <{local variables}>, <Return(<expr>)>, script)

Description

Stores the body *`script`* with *`arguments`* as local variables.

Returns

The function as defined. If the `Return()` argument is specified, the expression is returned.

When called later, it returns the result of the *`script`* given the specified *`arguments`*.

Arguments

`{arguments}` A list of arguments to pass into the function. You can specify some arguments as optional or required.

`{local variables}` A list of variables that are local to the function. You can declare local variables in three ways:

```
{var1, var2}
{var1=0, var1="a string"}
{Default Local}
```

The last option declares that all unscoped variables used in the function are local to the function.

`Return(expr)` This optional argument returns an expression from an user defined function. If a null expression is used, a period, ".", is returned.

`script` Any valid JSL script.

Get Class Names(< <class>, ...>)

Description

Gets a set of names to all classes or the set of specific class references.

Arguments

`class` A quoted string representation of the name of a defined class or a reference to an instantiated class object.

Returns

A list of class names as determined by the arguments to the function.

Get Classes(< <class>, ...>)

Description

Gets a set of references to all classes or the set of specific class references.

Arguments

`class` A quoted string representation of the name of a defined class or a reference to an instantiated class object.

Returns

A list of class references as determined by the arguments to the function.

Get Environment Variable("variable")

Description

Retrieves the value of an operating system environment variable.

Returns

A quoted string that contains the value of the specified environment variable. If the specified variable is not found, an empty string is returned.

Arguments

`"variable"` A quoted string that contains the name of an environment variable.

Notes

On macOS, environment variable names are case-sensitive. On Windows, the names are case-insensitive.

Get Locale Setting()

Gets a local setting such as a decimal setting.

Get Log(<n>)

Description

Returns a list of lines from the log.

Returns

A list of quoted strings. Each string contains one line from the log.

Argument

n Optional, integer. If no argument is specified, all the lines are returned. If a positive number is specified, the first *n* lines are returned. If a negative number is specified, the last *n* lines are returned. If *n*=0, no lines are returned (an empty list). If the log is empty, an empty list is returned.

Get Namespace Names(< <namespace reference>,...>)

Description

Returns a list of the names of all currently defined namespaces.

Example

```
nsaa = New Namespace(
   "aa",
   {
      x = 1
   }
);
nsbb = New Namespace(
   "bb",
   {
      y = 1
   }
);
lns = Get Namespace Names();
Show( lns );
nsaa << Delete;
nsbb << Delete;
```

Get Namespaces(< <namespace reference>,...>)

Description

Returns a list of currently defined namespaces.

Example

```
nsaa = New Namespace(
   "aa",
   {
      x = 1
   }
```

```
);
nsbb = New Namespace(
   "bb",
   {
      y = 1
   }
);
lns = Get Namespaces();
```

Include("pathname", <named arguments>)

Description

Opens the script file identified by the quoted string *pathname*, parses the script in it, and executes it.

Returns

Whatever the included script returns. If you use the <<Parse Only option, Include returns the contents of the script.

Named Arguments

<<Parse Only Parses the script but does not execute the script.

<<New Context Causes the included script to be run its own unique namespace. When the parent and included scripts use the global namespace, include <<Names Default to Here along with <<New Context.

<<Allow Include File Recursion Lets the included script include itself.

Notes

If a trailing space is included in the path name, the space is ignored on Windows. On macOS, the script fails.

Include File List()

Description

Returns a list of files that are included at the point of execution.

Is Class(class)

Description

Returns a value that indicates whether the class reference is a class object.

Argument

A class reference to an instantiated class object.

Returns

Returns a zero or a 1.

Is Log Open()

Description

Returns result if log window is open.

Length

See "Length(string)".

List

See "List(a, b, c, ...)".

Local({name=value, ...}, script)

Description

Resolves *names* to local expressions.

Local Here(expression)

Description

Creates a local Here namespace block. Use this function to prevent name collisions when multiple scripts are executed from the same root namespace (for example, when a script executes two button scripts that have the same variables). The argument can be any valid JSL expression.

Lock Namespaces(<*string*>,|< {*string*}, ...>)

Description

Locks all variables or specified named variables in this namespace and prevents variables from being added, changed, or removed.

Example

```
ns = New Namespace(
   "aaa"
);
ns << Lock Namespaces;
Try( ns << Delete Namespaces, Show( exception_msg ) );
Delete Namespaces();
Try( Delete Namespaces( "aaa" ), Show( exception_msg ) );
```

Lock Globals(name1, name2, ...)

Description

Locks one or more global variables to prevent it or them from being changed.

Lock Symbols(<name>, <name>, ...)

Description

Locks the specified symbols, which prevents them from being modified or cleared. If no symbols are provided, all global symbols are locked. If no symbols are provided and the script has the *Names Default To Here* mode turned on, then all local symbols are locked.

Log Capture(expr)

Description

Evaluates the expr, captures the output that would normally be sent to the log, and instead returns it.

Returns

A quoted string that contains the log output.

Argument

Any valid JSL expression.

Notes

No output appears in the log.

Method({arg1 = val1, ...}, script)

Description

Creates a method within a class. Note that methods use local scoping for all variables that are not explicitly scoped, with the exception of class member variables.

Arguments

`{ arg1 = val1, … }` The set of expected arguments and optional initialization expressions to be passed to the method when called.

`script` Any valid JSL script.

N Items

See "N Items(source)".

Names Default To Here(Boolean)

Description

Determines where unresolved names are stored, either as a global or local (if *Boolean* is 0) or in the `Here` scope (if *Boolean* is 1).

Namespace(name)

Description

Returns a reference to the named namespace (*name*).

Argument

`Name` A namespace name quoted string or a reference to a namespace.

Namespace Exists(name)

Description

Returns 1 if a namespace with the specified *name* exists; otherwise, returns 0.

New Namespace(<"name">, <{expr, ...}>)

Description

Creates a new namespace with the specified name. If a name is not provided, an anonymous name is provided.

Returns

A reference to the namespace.

Arguments

`name` An optional, quoted string that contains the name of the new namespace.

`{list of expressions}` An optional list of expressions within the namespace.

Open Log(<Boolean>)

Description

Opens the log. Include the Boolean argument to make the window active, even if it is already open.

New Object("class name"(constructor arguments))
New Object(class name(constructor arguments))
New Object(class reference(constructor arguments))

Description

Creates an instance object of a class.

Arguments

`"class name"` Name of the class to be instantiated.

`class name` Unquoted name of the class to be instantiated.

`class reference` Reference to an existing class object that will be used to instantiate a new object of the same class.

`constructor arguments` Set of arguments to be passed to the _init_ constructor.

Example

```
Define Class(
   "complex",
   real = 0; imag = 0;
   _init_ = Method( {a, b}, real = a; imag = b; );
   Add = Method( {y}, complex( real + y:real, imag + y:imag ) );
   Sub = Method( {y}, complex( real - y:real, imag - y:imag ) );
   Mul = Method( {y},
      complex( real * y:real - imag * y:imag, imag * y:real + real * y:imag )
   );
   Div = Method( {y},
      t = complex( 0, 0 );
      mag2 = y:Magsq();
      t:real = real * y:real + imag * y:imag;
      t:imag = imag * y:real + real * y:imag;
      t:real = t:real / mag2;
      t:imag = t:imag / mag2;
      t;
   );
   Magsq = Method( {}, real * real + imag * imag );
   Mag = Method( {}, Sqrt( real * real + imag * imag ) );
   To String = Method( {}, Char( real ) || " + " || Char( imag ) || "i" )
);
cl = New Object( complex( 1, 2 ) );
```

Parameter({name=value, ...}, model expression)

Description

Defines formula parameters for models for the Nonlinear platform.

Parse(*string*)

Description

Converts a quoted character string into a JSL expression.

Print(expr, expr, ...)

Description

Prints the values of the specified *expressions* to the log.

Quit()

See “Exit(<NoSave>)”.

Recurse(function)

Description

Makes a recursive call of the defining *function*.

Save Log(pathname)

Description

Writes the contents of the log to the specified file location.

Send(obj, message)

obj << message

Description

Sends a *message* to a platform *object*.

Set Environment Variable("variable", <"value">)

Description

Sets the environment variable to the value specified. If the "value" argument is missing or is an empty quoted string, then the environment variable is deleted from the JMP process environment variable table.

Show(expr, expr, ...)

Description

Prints the name and value of each *expression* to the log.

Show Classes(< <class>,...>)

Description

Shows the contents of user-defined classes in the log. You can specify more than one class. If you do not specify an argument, all user-defined classes are shown in the log.

Example

```
Define Class(
   "aa",
   _init_ = Method( {} ); x = 1; m1 = Method( {a, b}, a * b )
);
Define Class(
   "bb",
   _init_ = Method( {} ); y = 1; m2 = Method( {a, b}, a / b )
);
Show Classes();
   // Class aa
```

```
_init_ = Method( {} );
m1 = Method( {a, b}, a * b );
x = 1;

// Class bb

_init_ = Method( {} );
m2 = Method( {a, b}, a / b );
y = 1;
```

Show Globals()

Description

Shows the values for all global symbols. Symbols in any scope other than global are not shown.

See Also

"Show Symbols()"

Show Namespaces(< <namespace reference>,...>)

Description

Shows the contents of all user-defined namespaces, both named and anonymous. You can specify zero or more namespaces.

Show Symbols()

Description

Shows the values for all symbols in any and all scopes.

See Also

"Show Globals()"

Sort List

See "Sort List({list}|expr)".

Sort List Into

See "Sort List Into({list}|expr)".

Throw("text")

Description

Returns a Throw. If you include *text*, throwing stores *text* in a global *exception_msg*. If *text* begins with "!" and is inside a Try() expression, throwing creates an error message

about where the exception was caught. "!" stops the script even if the `Throw()` is caught by the second argument of `Try()`.

`Try(expr1, expr2)`

Description

Evaluates *expr1*. If the evaluation returns a `Throw`, execution stops, and nothing is returned. *expr2* is evaluated next to return the result.

Examples

```
Try( Sqrt( "s" ), "invalid" );
   "invalid"

Try( Sqrt( "s" ), exception_msg );
   {"Cannot convert argument to a number [or matrix]"(1, 2, "Sqrt",
      Sqrt/*###*/("s"))}
```

Notes

`Expr2` can be a quoted character string or the global exception message (`exception_msg`) that contains more information about the error returned.

`Type(x)`

Description

Returns a quoted string that names the type of object *x* is. The list of possible types is: Unknown, List, DisplayBox, Picture, Column, TableVar, Table, Empty, Pattern, Date, Integer, Number, String, Name, Matrix, RowState, Expression, Associative Array, BLOB.

`Unlock Symbols(name1, name2, ...)`

`Unlock Globals(name1, name2, ...)`

Description

Unlocks the specified symbols that were locked with a `Lock Symbols()` or `Lock Globals()` command.

`Wait(n)`

Description

Pauses *n* seconds before continuing the script. The default setting is 3 seconds. Specifying `Wait(0)` enables one cycle of message processing. For example, you can use this function to allow a button press in the UI. The shortest duration that actually allows JMP to pause is `n = 0.01`. The longest duration you can specify without prompting a JMP dialog is `n = 60*60*4`.

Notes

You can use `Wait(n)` if you want something to stay on the screen long enough to see it, if you need a platform to finish launching before scripting it, or if you need to press buttons in the UI while the script runs.

`Watch(all | name1, ...)`

Description

Shows variables from global, here, and local namespaces and their values in a window. If "all" is provided as the argument, all globals are placed into the window.

Notes

- New globals are not added to the window list.
- Watching associative arrays that have been modified using messages is not supported.

`Wild()`

Description

Only used with `Extract Expr()` for expression matching to denote a wildcard position that matches any expression.

`Wild List()`

Description

Only used with `Extract Expr()` for expression matching to denote a series of wildcard arguments that match any expression.

`Write("text")`

Description

Prints *`text`* to the log without surrounding quotation marks.

Python Integration Functions

`Python Connect(<Echo(Boolean),> <Path(path),> <Use Python Version(`*`string`*`),> <Python System Path(list)>)`

Description

Initializes the Python integration interfaces and returns an active Python integration interface connection as a scriptable object.

Returns

A Python scriptable object.

Optional Named Arguments

`Echo(Boolean)` Global argument. Prints the Python source lines to the JMP log. The default value is true.

`Path` Specifies the path to the Python DLL or shared library.

`Use Python Version(`*`string`*`)` Specifies which version of Python should be used for JMP-to-Python processing.

`Python System Path` Specifies a JSL list of paths that define a Python sys path set on macOS.

`Python Control(<named arguments>)`

Description

Sends control operations to signal Python with external events, such as source line echoing.

Returns

Returns 0 if the call succeeded and 1 if an error occurred.

Optional Named Arguments

`Interactive(Boolean)` Enables interactive mode in the Python matplotlib package. Determines whether the graphics window is released or closed when graphics rendering is complete.

`Echo(Boolean)` Global argument. Prints the Python source lines to the JMP log. The default value is true.

`Python Disconnect`

Description

Terminates the Python interfaces.

`Python Execute({list of inputs}, {list of outputs}, Python_Code, named_arguments)`

Description

Submits Python code to the active global Python integration interface connection given a list of inputs. On completion, returns a list of outputs.

Returns

Returns 0 if successful and 1otherwise.

Positional Arguments

`{list of inputs}` A list of JMP variable names to be sent to Python as inputs.

`{list of outputs}` A list of JMP variable names to be retrieved from Python as outputs.

Python_Code The Python code to submit.

Named Arguments

See "Python Submit(Python_Code, <named_arguments>)".

Example

This example initiates the Python connection, sends a character variable, a numeric variable, and a set of matrices to Python. Python is then instructed to perform a set of matrix operations on the sent matrices. The Python Execute() function then get the set of matrices created by the matrix operations and gets the values of the character and numeric variables that was originally sent. Upon completion of the data retrieval, the Python connection is closed.

```
Python Init();
a = "abcdef";
d = 3.141;
v = [9 8 7, 6 5 4, 3 2 1];
m = [1 2 3, 4 5 6, 7 8 9];
ml = Python Execute(
   {v, m, a, d},
   {x1, x2, y1, y2, z1, z2, a, d},
   "\[
import numpy as np
x1 = np.multiply(v, m) # matrix product
print('x1=', x1)
x2 = np.divide(v, m) # matrix division
print('x2=', x2)
y1 = np.dot(v, m) # dot product of v and m
print('y1=', y1)
y2 = np.dot(m, v) # dot product of m and v
print('y2=', y2)
z1 = np.inner(v, m) # inner product of v and m
print('z1=', z1)
z2 = np.inner(m, v) # inner product of m and v
print('z2=', z2)
]\"
);
Show( v, m, ml, x1, x2, y1, y2, z1, z2, a, d );
Python Term();
   x1= [[  9.  16.  21.]
    [ 24.  25.  24.]
    [ 21.  16.   9.]]
   x2= [[ 9.          4.          2.33333333]
    [ 1.5         1.          0.66666667]
    [ 0.42857143  0.25        0.11111111]]
   ...
```

Python Get(name)

Description

Gets a named variable from Python to JMP.

Returns

Returns the value of the named variable.

Argument

`name` The name of the Python variable to be sent to JMP. The argument can represent any of the following Python data types: numeric, quoted string, matrix, list, or data frame.

Example

```
Python Init(); // initiate the Python connection

qbx = "The right stuff";

// send the qbx variable and sample data table "Animals.jmp" to Python
Python Send( qbx );

dt = Open( "$SAMPLE_DATA/Animals.jmp" );
Python Send( dt );
Close( dt, nosave );

// get the Python variable qbx and place it into a JMP variable qbx
qbx = Python Get( qbx );

/* get the Python variable dt and place it into a JMP data table
referenced by df */
df = Python Get( dt );

Python Term();

Show( qbx );
df << New Data View;
Wait( 10 );
Close( df, nosave );
Python Term();
   qbx = "The right stuff";
   0
```

Python Get Graphics(format)

Description

Gets the last graphics object written to the Python graph display window in the specified graphics format. The graphics object can be returned in several different graphic formats.

Returns

Returns a JMP picture object.

Argument

`format` The format that the Python graph display window contents are to be converted to. Valid formats are `png`, `bmp`, `jpeg`, `jpg`, `tiff`, `tif`, and `gif`.

`Python Get Version`

Description

Returns the version number of Python being used with the JMP Python interfaces.

`Python Init(<Echo(Boolean),> <Path(path),> <Use Python Version(`*`string`*`),> <Python System Path({list})>`

Description

Initializes the Python integration interfaces.

Returns

Returns 0 if operation is successful and 1 if not successful.

Optional Named Arguments

`Echo(Boolean)` Global argument. Prints the Python source lines to the JMP log. The default value is true.

`Path` Specifies the path to the Python DLL or shared library.

`Use Python Version(`*`string`*`)` Specifies which version of Python should be used for JMP-to-Python processing.

`Python System Path` Specifies a JSL list of paths that define a Python sys path set on macOS.

`Python Is Connected`

Description

Determines whether a Python integration interface connection is currently connected to Python.

Returns

Returns 1 if connected and 0 otherwise.

`Python JMP Name to Python Name(name)`

Description

Maps a JMP variable name to its corresponding Python variable name using Python variable name naming rules.

Returns

A quoted string, the mapped Python name.

Argument

`name` The name of the JMP variable to be sent to Python.

`Python Send(name)`

Description

Sends a named variable from JMP to Python.

Returns

Returns 0 if successful.

Argument

`name` The name of the JMP variable to be sent to Python.

`Python Send File(filename, <, Python Name(name)>)`

Description

Sends a data file to Python. The filename argument is a quoted string that specifies a pathname to the file to be sent to Python.

`Python Submit(Python_Code, <named_arguments>)`

Description

Submits Python code to the active global Python integration interface connection.

Returns

Returns 0 if successful and non-zero otherwise.

Named Arguments

`Python_Code` The Python code to submit. Statements can be a quoted string value or a list of string values.

`Expand(Boolean)` (Optional) Performs an `Eval Insert()` on the Python code before submission.

`Echo(Boolean)` (Optional) Prints the Python source lines to the JMP log.

Example

```
Python Init(); // initiate the Python connection
commands =
"
friends = ['john', 'pat', 'gary', 'michael']
print(friends)
for i, name in enumerate(friends):
    print( \!"iteration {iteration} is {name}\!".format(iteration=i,
  name=name))
";
```

```
Python Submit( commands );
Python Term();
   ['john', 'pat', 'gary', 'michael']
   iteration 0 is john
   iteration 1 is pat
   iteration 2 is gary
   iteration 3 is michael

   0
```

Python Submit File(path)

Description

Submits statements to Python using the file specified in the path name.

Argument

`path` The path to the file that contains the Python source lines to be executed.

Python Term

Description

Terminates the currently active Python integration interface.

Returns

Returns 0 if successful and 1 otherwise.

R Integration Functions

R Connect(<named_arguments>)

Description

Returns the current R connection object. If there is no connection to R, it intializes the R integration interfaces and returns an active R integration interface connection as a scriptable object.

Returns

R scriptable object.

Arguments

`Echo(Boolean)` (Optional) Sends all source lines to the JMP log. This option is global. The default value is `true`.

R Control(Interrupt|Async(Boolean)|Echo(Boolean))

Description

Changes the control options for R.

R Execute({ list of inputs }, { list of outputs }, "rCode", <named_arguments>)

Description

Submit the specified R code to the active global R connection given a list of inputs. On completion, the outputs are returned into the specified list.

Returns

0 if successful; nonzero otherwise.

Arguments

`{ list of inputs }` A list of JMP variable names to be sent to R as inputs.

`{ list of outputs }` A list of JMP variable names to contain the outputs returned from R.

`rCode` A quoted string that contains the R code to submit.

`Expand(Boolean)` An optional, Boolean, named argument. Performs an `Eval Insert()` on the R code before submitting to R.

`Echo(Boolean)` An optional, Boolean, named argument. Sends all source lines to the JMP log. This option is global. The default value is `true`.

Example

Send the JMP variables x and y to R, execute the R statement z <- x * y, and then get the R variable z and return it to JMP.

```
x = [1 2 3];
y = [4 5 6];
rc = R Execute( {x, y}, {z}, "z <- x * y" );
```

R Get(variable_name)

Description

Gets the named variable from R to JMP.

Returns

The value of the named variable.

Argument

`name` Required. The name of an R variable whose value to return to JMP.

Example

Assume that a matrix named qbx and a data frame named df are present in your R connection.

```
// get the R variable qbx and placed it into a JMP variable qbx
qbx = R Get( qbx );

// get the R variable df and placed it into a JMP data table referenced by df
df = R Get( df );
```

R Get Graphics("format")

Description

Gets the last graphics object written to the R graph display window in the specified format.

Returns

A JMP picture object.

Argument

`format` Required. Specifies the graphics format to be used. Valid formats are `png`, `bmp`, `jpeg`, `jpg`, `tiff`, `tif`, and `gif`.

R Get Version

Description

Returns the version number of R being used with JMP R interfaces.

R Init(named_arguments)

Description

Initializes the R session.

Returns

0 if the initialization is successful; any nonzero value otherwise.

Argument

`Echo(Boolean)` (Optional) Sends all source lines to the JMP log. This option is global. The default value is `true`.

R Is Connected()

Description

Determines whether a connection to R exists.

Returns

1 if connected; 0 otherwise.

Arguments

None.

R JMP Name to R Name(name)

Description

Maps the specified JMP variable name to the corresponding R variable name using R naming rules. Requires an active connection to R.

Argument

`name` The name of a JMP variable to be sent to R.

Returns

A quoted string that contains the R name.

`R Send( name, <R Name( name )>)`

Description

Sends named variables from JMP to R.

Returns

0 if the send is successful; any nonzero value otherwise.

Arguments

`name` required. The name of a JMP variable to be sent to R.

`R Name(name)` (Optional) You can give the variable that you send to R a different name. For example

```
R Send(Here:x, R Name("localx"))
```

For data tables only:

`Selected(Boolean)` optional, named, Boolean. Send only selected rows from the referenced data table to R.

`Excluded(Boolean)` optional, named, Boolean. Send only excluded rows from the referenced data table to R.

`Labeled(Boolean)` optional, named, Boolean. Send only labeled rows from the referenced data table to R.

`Hidden(Boolean)` optional, named, Boolean. Send only hidden rows from the referenced data table to R.

`Colored(Boolean)` optional, named, Boolean. Send only colored rows from the referenced data table to R.

`Markered(Boolean)` optional, named, Boolean. Send only markered rows from the referenced data table to R.

`Row States(Boolean, <named arguments>)` optional, named. Includes a Boolean argument and optional named arguments. Send row state information from the referenced data table to R by adding an additional data column named “RowState”. Multiple row states are created by adding together individual settings. Here are the individual values:

- Selected = 1
- Excluded = 2
- Hidden = 4
- Labeled = 8
- Colored = 16

– Markered = 32

Here are the named arguments for the `Row States()` argument:

`Colors(Boolean)` optional, named, Boolean. Sends row colors. Adds additional data column named "RowStateColor".

`Markers(Boolean)` optional, named, Boolean. Sends row markers. Adds additional data column named "RowStateMarker".

Examples

Create a matrix, assign it to X, and send the matrix to R:

```
X = [1 2 3];
rc = R Send( X );
```

Open a data table, assign a reference to it (*dt*), and send the data table, along with its current row states, to R:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
rc = R Send( dt, Row States(1) );
```

R Send File("pathname", <R Name("name")>)

Description

Sends the specified data file from JMP to R.

Returns

0 if the send is successful; any nonzero value otherwise.

Arguments

`pathname` required. A quoted string that contains a pathname for a file.

`R Name(name)` (Optional) You can give the data file that you send to R a different name.

R Submit("rCode", <named_arguments>)

Description

Submits the specified R code to the active global R connection.

Returns

0 if successful; nonzero otherwise.

Arguments

`rCode` A required, quoted string that contains the R code to submit.

`Expand(Boolean)` An optional, Boolean, named argument. Performs an `Eval Insert()` on the R code before submitting to R.

`Echo(Boolean)` An optional, Boolean, named argument. Sends all source lines to the JMP log. This option is global. The default value is `true`.

`Async(Boolean)` An optional, Boolean, named argument. If set to true (1), the submit can be canceled either by pressing the ESCAPE key, or by using this message to an R connection: `rconn<<Control( Interrupt( 1 ) )`. False (0) is the default value.

Example

```
rc = R Submit("\[
  x <- rnorm(5)
  print(x)
  y <- rnorm(5)
  print(y)
  z = plot(x, y)
]\" );
```

`R Submit File( "pathname" )`

Description

Submits statements to R using a file pointed in the specified *pathname*.

Returns

0 if successful; nonzero otherwise.

Argument

`Pathname` A quoted string that contains the pathname to the file that contains the R code to be executed.

`R Term()`

Description

Terminates the currently active R integration interface.

Returns

Returns 0 if the termination is successful and -1 otherwise.

Arguments

None

Random Functions

`Col Shuffle(<By var,...>)`

Description

Creates a random ordering of the row numbers of the current data table when used in a column formula.

Note: This function is generally used in a column formula.

Returns

A random integer between 1 and the number of rows in the current data table.

Argument

`By var` (Optional) A By variable enables you to randomly order the rows within the groups of the By variable values.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "Shuffle", Numeric, Continuous, Set Formula( Col Shuffle() )
  );
```

This example creates a column formula that shuffles the order of the row numbers (1 to 40) each time the formula is evaluated. Each number appears only once.

Make Validation Formula(rates, <<Stratification Columns(cols), <<Grouped Columns(cols), <<Cutpoint Column(col))

Description

Generates a validation column. This function is primarily used by the Make Validation Column utility.

Arguments

`rates` Vector of three rates that specify the training, validation, and test rates, respectively.

`<<Stratification Columns` Assigns one or more stratification columns.

`<<Grouped Columns` Assigns one or more grouping columns.

`<<Cutpoint Column` Assigns a numeric cutpoint column.

Predictive and Specialized Modeling

Random Beta(alpha, beta, <theta=0>, <sigma=1>)

Description

Returns a random number from a beta distribution with two shape parameters, *alpha* and *beta*, and optional parameters *theta* and *sigma*.

Arguments

`alpha, beta` Shape parameters α and β, which must both be greater than 0.

`theta` Optional threshold parameter θ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

Random Beta Binomial(n, p, <delta=0>)

Description

Returns a random number from a beta binomial distribution for *n* trials with probability *p* and overdispersion parameter *delta*.

Arguments

n The number of trials, which must be greater than or equal to 2. If the specified *n* is not an integer, the non-integer part is truncated.

p The probability of success for each trial, which must be between 0 and 1.

delta The overdispersion parameter δ, which must be between Maximum[-*p*/(*n*-*p*-1), -(1-*p*)/(*n*-2+*p*)] and 1. The default is 0.

Random Binomial(n, p)

Description

Returns a random number from a binomial distribution with *n* trials and probability *p* of the event of interest occurring.

Arguments

p The probability of success for each trial, which must be between 0 and 1.

n The number of trials.

Random Category(probA, resultA, probB, resultB, <..., ...,> resultElse)

Description

Returns one of the specified result expressions at random, chosen from pairs of probability and result expressions. A random uniform number is generated and compared to the *prob* arguments to determine which *result* argument is returned.

Arguments

probA Numeric value between 0 and 1 that represents the probability of the corresponding result expression being returned.

resultA Expression that corresponds to *probA*.

resultElse Expression that is returned if no previous result expression has been returned.

Random Cauchy()

Description

Returns a random number from a Cauchy distribution with a median of zero.

Random ChiSquare(df, <nc=0>)

Description

Returns a random number from a chi-square distribution with given *df* (degrees of freedom) and optional noncentrality parameter.

Arguments

df The degrees of freedom *n*, which must be greater than 0.

nc Optional noncentrality parameter λ, which must be nonnegative. The default is 0.

Random Exp()

Description

Returns a random number from an exponential distribution with scale parameter equal to 1. Equivalent to the negative log of `Random Uniform`.

Random F(dfnum, dfden, <noncentral=0>)

Description

Returns a random number from an F distribution with a given *dfnum*, *dfden*, and optional noncentrality parameter.

Arguments

`dfnum` The degrees of freedom, v_1, of the chi-square distribution in the numerator of the *F*-distribution. *dfnum* must be greater than 0.

`dfden` The degrees of freedom, v_2, of the chi-square distribution in the denominator of the *F*-distribution. *dfden* must be greater than 0.

`noncentral` Optional noncentrality parameter λ, which must be nonnegative. The default is 0.

Random Frechet(<mu=0>, <sigma=1>)

Description

Returns a random number from a Fréchet distribution with the location *mu* and scale *sigma*.

Arguments

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

Random Gamma(alpha, <scale=1>)

Description

Returns a random numbers from a gamma distribution for given *alpha* and optional *scale*.

Arguments

`alpha` The shape parameter α, which must be greater than 0.

`scale` Optional scale parameter β, which must be greater than 0. The default is 1.

Random Gamma Poisson(lambda, <sigma=1>)

Description

Returns a random number from a gamma Poisson distribution with parameters *lambda* and *sigma*.

Arguments

`lambda` The shape parameter λ, which much be greater than 0.

`sigma` Optional overdispersion parameter σ, which must be greater than or equal to 1. The default is 1. When the overdispersion parameter is 1, the distribution reduces to a Poisson(λ) distribution.

`Random GenGamma(<mu=0>, <sigma=1>, <lambda=0>)`

Description

Returns a random number from an extended generalized gamma distribution with parameters *mu*, *sigma*, and *lambda*.

Arguments

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

`lambda` Optional shape parameter λ. The default is 0.

`Random Geometric(p)`

Description

Returns a random number from the geometric distribution with probability *p* that a specific event occurs at any one trial.

`Random GLog(mu, sigma, lambda)`

Description

Returns a random number from a generalized logarithmic distribution with parameters *mu*, *sigma*, and *lambda*.

Arguments

`mu` The location parameter μ.

`sigma` The scale parameter σ, which must be greater than 0.

`lambda` A shape parameter λ, which must be greater than 0.

`Random Index(n, k)`

Description

Returns a *k* by 1 matrix of random integers between 1 and *n* with no duplicates.

`Random Integer(n)`

`Random Integer(k, n)`

Description

Returns a random integer from 1 to *n* or from *k* to *n*.

Random Johnson Sb(gamma, delta, theta, sigma)

Description

Returns a random number from a Johnson Sb distribution with parameters *gamma*, *delta*, *theta*, and *sigma*.

Arguments

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Scale parameter σ, which must be greater than 0.

Random Johnson Sl(gamma, delta, theta, <sigma=1>)

Description

Returns a random number from a Johnson Sl distribution with parameters *gamma*, *delta*, *theta*, and optional *sigma*.

Arguments

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Optional parameter σ that indicates if the distribution is skewed positively or negatively. `sigma` must be equal to either +1 (skewed positively) or -1 (skewed negatively). The default is +1.

Random Johnson Su(gamma, delta, theta, sigma)

Description

Returns a random number from a Johnson Su distribution with parameters *gamma*, *delta*, *theta*, and *sigma*.

Arguments

`gamma` Shape parameter γ.

`delta` Shape parameter δ, which must be greater than 0.

`theta` Location parameter θ.

`sigma` Scale parameter σ, which must be greater than 0.

Random LEV(<mu=0>, <sigma=1>)

Description

Returns a random number from an LEV distribution with the location *mu* and scale *sigma*.

Arguments

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

Random LogGenGamma(<mu=0>, <sigma=1>, <lambda=0>)

Description

Returns a random number from a log generalized gamma distribution with parameters *mu*, *sigma*, and *lambda*.

Arguments

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

`lambda` Optional shape parameter λ. The default is 0.

Random Logistic(<mu=0>, <sigma=1>)

Description

Returns a random number from a logistic distribution with location *mu* and scale *sigma*.

Arguments

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

Random Loglogistic(<mu=0>, <sigma=1>)

Description

Returns a random number from a loglogistic distribution with location *mu* and scale *sigma*.

Arguments

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

Random Lognormal(<mu=0>, <sigma=1>)

Description

Returns a random number from a lognormal distribution with location *mu* and scale *sigma*.

Arguments

`mu` Optional location parameter μ. The default is 0.

`sigma` Optional scale parameter σ, which must be greater than 0. The default is 1.

Random Multivariate Normal(mean, covar, <nrows=1>)

Description

Returns a random vector from a multivariate normal distribution with mean vector *mean* and covariance matrix *covar*. To generate multiple vectors, specify an integer greater than 1 for the *nrows* argument. When *nrows* is greater than 1, the return value is a matrix. The number of columns in the random vector or matrix is equal to the number of rows in the *covar* argument.

Arguments

mean Mean vector for the multivariate normal distribution.

covar Covariance matrix for the multivariate normal distribution. This matrix must be a symmetric square matrix that contains the same number of columns as the mean vector.

nrows Optional argument that specifies the number of random vectors returned. The default number of rows is 1.

Random Negative Binomial(n, p)

Description

Returns a random number from a negative binomial distribution for *n* successes with probability of success *p*.

Random Normal(<mu=0>, <sigma=1>)

Description

Returns a random number from a normal distribution with mean *mu* and standard deviation *sigma*.

Arguments

mu Optional location parameter μ. The default is 0.

sigma Optional scale parameter σ, which must be greater than 0. The default is 1.

Random Normal Mixture(meanvec, sdvec, probabvec)

Description

Returns a random number from a normal mixture distribution with the specified arguments.

Arguments

meanvec A vector that contains group means.

sdvec A vector that contains the group standard deviations.

probvec A vector that contains the group probabilities.

Random Poisson(lambda)

Description

Returns a random number from a Poisson distribution with shape parameter *lambda*.

Arguments

lambda The shape parameter λ, which much be greater than 0.

Random Reset(seed)

Description

Restarts the random number sequences with *seed*.

Notes

You can use the **Random Reset** function to set a random seed that enables you to reproduce a set of random numbers. The random number algorithms in JMP are sometimes refined, and therefore the results might not be consistent across different versions of JMP.

Random Seed State(<seed state>)

Description

Retrieves or restores the random seed state to or from a BLOB object.

Random SEV(<mu=0>, <sigma=1>)

Description

Returns a random number from an SEV distribution with the specified location *mu* and scale *sigma*.

Arguments

mu Optional location parameter μ. The default is 0.

sigma Optional scale parameter σ, which must be greater than 0. The default is 1.

Random SHASH(gamma, delta, theta, sigma)

Description

Returns a random number from a sinh-arcsinh (SHASH) distribution with parameters *gamma*, *delta*, *theta*, and *sigma*.

Arguments

gamma The shape parameter γ.

delta The shape parameter δ, which must be greater than 0.

theta The location parameter θ.

sigma The scale parameter σ, which must be greater than 0.

`Random Shuffle(matrix)`

Description

Returns the matrix with the elements shuffled into a random order.

`Random t(df, <noncentral=0>)`

Description

Returns a random number from a t distribution with the specified *df* (degrees of freedom). The noncentrality argument may be negative or positive. The default value of *noncentral* is 0.

`Random Triangular(min, mode, max)`

`Random Triangular(mode, max)`

`Random Triangular(mode)`

Description

Generates a random number from a triangular distribution between 0 and 1 with the *mode* that you specify. The triangular distribution is typically used for populations that have a small number of data.

Arguments

`min` Specifies the lower limit of the triangular distribution. The default value is 0.

`mode` Specifies the mode of the triangular distribution.

`max` Species the upper limit of the triangular distribution. The default value is 1.

Notes

If you specify only the mode, the minimum value is 0, and the maximum value is 1. If you specify the mode and maximum value, the minimum value is 0 by default.

`Random Uniform()`

`Random Uniform(x)`

`Random Uniform(min, max)`

Description

Generates a random number from a uniform distribution between 0 and 1. `Random Uniform(x)` generates a number between 0 and x. `Random Uniform (min, max)` generates a number between *min* and *max*. The result is an approximately even distribution.

Random Weibull(shape, <scale=1>)

Description

Returns a random number from a Weibull distribution with parameters *shape* and optional *scale*.

Arguments

`shape` Shape parameter β, which must be greater than 0.

`scale` Optional scale parameter α, which must be greater than 0. The default is 1.

Resample Freq(<rate=1, <column>>)

Description

Generates a frequency count for sampling with replacement. If no arguments are specified, the function generates a 100% resample.

Note: This function is generally used in a column formula.

Arguments

`rate` (Optional) Specifies the rate of resampling. The default value is 1. A negative value specifies that fractional frequencies are allowed.

`column` (Optional) If you specify *column*, you must also specify *rate*. The sample size is calculated by the rate multiplied by the sum of the specified column. If rate is negative, then the sample size is the negative of the rate multiplied by the sum of the specified column. If you do not specify a column, the generated frequencies sum to the number of rows.

Example

To ensure that the numbers in the frequency column match each time you run the script, use `As Constant()`. `As Constant()` evaluates an expression to create a constant value that does not change after it has been computed.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dc = dt << New Column( "column",
   Formula(
      As Constant(
         Random Reset( 123 );
         0;
      ) + Resample Freq()
   )
);
dc << Eval Formula;
```

Notes

– A typical use of this function generates a column with many 1s, some 0s, some 2s, and so forth, corresponding to which rows were randomly assigned any of *n* randomly selected rows.

– A typical use of this with an *existing* frequency column produces a new frequency column whose values are similar to the old frequency column (have the same expected value); however, the values vary somewhat due to random selection at the rates corresponding to the old frequency column.

Row Functions

```
As Table(matrix, <matrix 2, ...>, < <<invisible >, < <<private >, <
<<Column Names({list}) >)
```

Description

Creates a new data tables from the *matrix*.

Returns

The new data table.

Argument

`matrix` Any matrix.

`<<invisible` Creates an invisible data table that hides the table from view but lists it in the JMP Home Window and Window menu.

`<<private` Hides the table completely. Creating a private data table speeds the process of getting to the data; it does not save the computer from allocating the memory necessary to hold the data table data.

`<<Column Names(list)` The list specified column names for the data. The argument is a list of quoted column names.

```
Col Stored Value(<dt>, col, <row>)
```

Description

Returns the data values stored in the column and disregards values assigned through column properties (such as Missing Value Codes).

Arguments

`dt` Optional reference to a data table. If this value is not supplied, the current data table is used.

`col` Name of the column.

`row` (Optional) Row name or number. If this value is not specified, the current row is used.

Example

Suppose that the Missing Value Codes column property is assigned to the `x1` column to treat "999" as a missing value. Another column includes a formula that calculates the

mean. To use the value "999" instead of a missing value to calculate the mean, use `Col Stored Value()` in the formula:

```
Mean( Col Stored Value( :x1 ), :x2, :x3 )
```

`Column(<dt>, "name", "formatted")`

`Column(<dt>, n)`

Description

Gets a reference to the data table column.

Arguments

`dt` Optional reference to a data table. If this is not supplied, the current data table is used.

`name` A quoted string that is the name of the column.

`formatted` A quoted string that returns the formatted string of the cell value.

`n` The column number.

`Column Name(n)`

Description

Determines the name of the column specified by number.

Returns

The name of the n^{th} column as an expression (not a quoted string).

Argument

`n` The number of a column.

`Count(from, to, step, times)`

Description

Used for column formulas. Creates row by row the values beginning with the *from* value and ending with the *to* value. The number of *steps* specifies the number of values in the list between and including the *from* and *to* values. Each value determined by the first three arguments of the count function occurs consecutively the number of *times* that you specify. When the *to* value is reached, `count` starts over at the *from* value. If the *from* and *to* arguments are data table column names, `count` takes the values from the first row only. Values in subsequent rows are ignored.

Returns

The last value.

Arguments

`from` Number, column reference, or expression. `Count` starts counting with this value.

`to` Number, column reference, or expression. `Count` stops counting with this value.

`step` Number or expression. Specifies the number of steps to use to count between *from* and *to*, inclusive.

`times` Number or expression. Specifies the number of times each value is repeated before the next step.

Examples

/* the rows in the column named *colname* are filled with the series 0, 3, 6, 0,

```
... until all rows are filled */
For Each Row(:colname[row()]=count(0, 6, 3, 1))

/* the rows in the column named colname are filled with the series 0, 0, 3,
   3, 6, 6, 0, ... until all rows are filled */
For Each Row(:colname[row()]=count(0, 6, 3, 2))
```

Notes

`Count()` is dependent on `Row()`, and is therefore mainly useful in column formulas.

Current Data Table(<dt>)

Description

Without an argument, gets the current (topmost) data table. With an argument, sets the current data table.

Returns

Reference to the current data table.

Argument

`dt` Optional name of or reference to a data table.

Notes

Private tables cannot be made current with `Current Data Table()`.

Data Table(n)

Data Table("name")

Get Data Table(<project(title|index|box|window),> name|index)

Description

Gets reference to the *n*th open data table or the table with the given *name* in a global variable.

Returns

Reference to the specified data table.

Argument

`n` Number of a data table.

`name` Quoted string, name of a data table.

`Dif(col, n)`

Description

Calculates the difference of the value of the column *col* in the current row and the value *n* rows previous to the current row.

Returns

The difference.

Arguments

`col` A column name (for example, `:age`).

`n` A number.

`Dim(<dt|matrix>)`

Description

Returns a row vector with the dimensions of the current data table, a specified data table, or a matrix. The dimensions are the number of rows and the number of columns and are listed in that order.

Arguments

`dt` A data table.

`matrix` A matrix.

Notes

If no argument is specified, the dimensions of the current data table are returned.

`Get Data Table List(<Project(title|index|box|window>)`

Description

Returns a list of all open data tables.

Notes

Use `Project(0)` to specify no project when running the expression in a project.

`Lag(col, n)`

Description

Returns for each row the value of the column *n* rows previous.

`N Row(dt); NRow(matrix)`

`N Rows(dt); NRows(matrix)`

Description

Returns the number of rows in the data table given by *dt* or in the *matrix.*

N Table()

Description

Returns the number of open data tables. Private tables are not included.

New Column("name", <"data type">, <"modeling type">, <Width(n)>, Format("format", width, precision), <Formula()>, <Set Values>, <Like(column reference)>, <actions>)

Description

Adds a new column named *"name"* after the last column in *dt*. Unless otherwise specified, columns are numeric, continuous, and 12 characters wide.

Returns

A column reference.

Notes

Can also be used as a message: `dt<<New Column`.

The `Like()` argument copies the data type, modeling type, format, formula, and other properties from the reference column into the new column.

See Also

"dt<<New Column(name, <data type>, <modeling type>, <Format(format, width)>, <Formula()>, <Set Values({..., ..., })>, <Set Property(properties)>)"

New Table("name", <visibility("invisible" | "private" | "visible")>, <actions>)

Description

Creates a new data table with the specified *name*.

Arguments

`name` A quoted string that contains the name of the new table.

`visibility` Optional quoted keyword. `invisible` hides the data table from view but lists it in the JMP Home Window and Window menu. `private` hides the table completely. `visible` shows the data table. "`visible`" is the default value.

Note: Creating a private data table speeds the process of getting to the data; it does not save the computer from allocating the memory necessary to hold the data table data.

`actions` Optional argument that can define the new table.

```
Row()
```

```
Row() = y
```

Description

Returns or sets the current row number. No argument is expected.

```
Sequence(from, to, <step size>, <repeat times>)
```

Description

Produces an arithmetic sequence of numbers across the rows in a data table. The *step size* and *repeat times* arguments are optional, and the default value for both is 1.

```
Subscribe to Data Table List(<subscriber name|"">,
<OnOpen(function)|OnClose(function)|OnRename(function)>)
```

Description

Subscribes to the data table list. You will be notified when a new data table has been added, closed, or renamed.

```
Subscript(a, b, c)
```

```
list[i]
```

```
matrix[b, c]
```

Description

Subscripts for lists extract the *i*th item from the *list*, or the *b*th row and the *c*th column from a *matrix*.

```
Suppress Formula Eval(Boolean)
```

Description

Turns off automatic calculation of formulas for all data tables.

```
Unsubscribe to Data Table List(<subscriber name>,
<"OnOpen"|"OnClose"|"All">)
```

Description

Removes a subscription to the data table list that has been added through `Subscribe to Data Table List()`.

Row State Functions

As Row State(i)

Description

Converts *i* into a row state value.

Returns

A row state from the *i* given.

Argument

`i` an integer

Color Of(rowstate)

Description

Returns or sets the color index.

Returns

The color index of *rowstate*.

Argument

`rowstate` a row state argument

Example

Set the color of the fifth row to red.

```
Color Of( Rowstate( 5) ) = 3
```

Color State(i)

Description

Returns a row state with the color index of *i*.

Returns

A row state.

Argument

`i` index for a JMP color

Combine States(rowstate, rowstate, ...)

Description

Generates a row state combination from two or more row state arguments.

Returns

A single numeric representation of the combined row states.

Arguments

`rowstate` Two or more row states.

`Excluded(rowstate)`

Description

Returns or sets an excluded index.

Returns

The excluded attribute, 0 or 1.

Argument

`rowstate` One or more row states.

`Excluded State(num)`

Description

Returns a row state for exclusion from the *num* given.

`Hidden(rowstate)`

Description

Returns or sets the hidden index.

`Hidden State(num)`

Description

Returns a row state for hiding from the *num* given.

`Hue State(num)`

Description

Returns a hue state from the *num* given.

`Labeled(rowstate)`

Description

Returns or sets the labeled index.

`Labeled State(num)`

Description

Returns a labeled state from the *num* given.

Marker Of(rowstate)

Description

Returns or sets the marker index of a row state.

Marker State(num)

Description

Returns a marker state from the *num* given.

Row State(<dt,> <n>)

Description

Returns the row state changed from the initial condition of the active row or the *n*th row.

Arguments

dt Optional positional argument: a reference to a data table. If this argument is not in the form of an assignment, then it is considered a data table expression.

n The row number.

Example

The following example creates the data table references and then returns the row state of row 1 in Big Class.jmp:

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt2 = Open( "$SAMPLE_DATA/San Francisco Crime.jmp" );
Row State( dt1, 1 );
```

Selected(rowstate)

Description

Returns or sets the selected index.

Selected State(num)

Description

Returns a selected state from the *num* given.

Shade State(num)

Description

The Shade State function assigns 5 shade levels to a color or hue.

SAS Integration Functions

As C Expr(x)

Description

Returns a C programming language representation of the expression.

Returns

A quoted string.

As JavaScript Expr(x)

Description

Returns a JavaScript representation of the expression.

Returns

A quoted string.

As JSON Expr(x)

Description

Returns a JSON (JavaScript Object Notation) representation of the expression.

Returns

A quoted string.

As Python Expr(x)

Description

Returns a Python representation of the expression.

Returns

A quoted string.

As SAS Expr(x)

Description

Converts an expression to a version that is more suitable for SAS DATA step. The code must be wrapped in a PROC DS2 call. Use `Expr(...)` for literal expressions. Use `NameExpr(name)` for expressions stored in a variable. Otherwise, the expression returns the expression to convert.

Returns

A quoted string.

Current Metadata Connection()

Description

Returns the active SAS metadata server connection, if any, as a scriptable object.

Current SAS Connection()

Description

Gets the active global SAS server connection, if any, as a scriptable object.

Get SAS Version Preference()

Description

Returns the SAS version selected in the SAS Integration page of the Preferences as a quoted string.

JMP6 SAS Compatibility Mode(Boolean)

Description

Setting this to 1 (true) causes SAS operators to operate in a mode compatible with JMP 6 capabilities.

Meta Connect(<"machine", port>, <"authDomain">, <"username">, <"password">, <named arguments>)

Meta Connect(<Profile("profile name")>, <Password("password")>, <named arguments>)

Meta Connect(<Environment("environment name")>, <named arguments>)

Description

Connects to a SAS Metadata Server. If no arguments are specified, an empty connection window appears. If some arguments are specified, a window partially filled in with the argument values appears. If all arguments are specified, the connection is made and no window appears.

Returns

1 if connection is successful, 0 if not.

Arguments

machine (Optional) A quoted string that contains the DNS name of the machine.

port Required if machine is specified. A quoted string or integer that contains the port on which the metadata server listens.

authDomain (Optional) A quoted string that contains the authentication domain for the credentials supplied. Not necessary unless *username* and *password* are included.

username (Optional) A quoted string that contains the user name for the connection.

`password` (Optional) A quoted string that contains the password for the connection.

Optional Named Arguments

`Profile("profile_name")` A quoted string that contains the name of the metadata server connection profile from which connection information should be retrieved.

`Environment("environment_name")` A quoted string that contains the name of the WIP environment from which connection information should be retrieved.

`Password("password")` A quoted string that contains the password for the specified profile name.

`CheckPreferenceOnly(0|1)` If specified, `Meta Connect` returns the status of the **I want to connect to a SAS Metadata Server** option in the SAS Integration page of JMP Preferences. If that box is checked, **Meta Connect** returns 1; if not, 0.

`Repository(`*`string`*`)` Takes a quoted string that contains the name of the repository to which to connect.

`ProfileLookup(0|1)` If *`machine`* and *`port`* are specified rather than a profile name, and `ProfileLookup` is specified, an attempt is made to find a metadata server connection profile with a machine name and port matching those provided. If one is found, other connection information (such as authentication domain, user name, and password) is obtained from that profile.

`Prompt(Always|Never|IfNeeded)` Takes one of the keywords `Always` (always prompt before attempting to connect), `Never` (never prompt, just fail), or `IfNeeded` (the default; prompt if connection with the given arguments fails).

`SASVersion("<version number>" <,Strict>)` Attempts to change the SAS version preference to the specified value before making the metadata server connection. If the SAS version is already locked to a different version than the one specified, the SASVersion argument will fail. By default, if the SAS version cannot be set, JMP will try the metadata server connection. However, if you include `Strict` as the second argument, the inability to change the SAS version will be treated as an error, and JSL processing will stop. If you do not include `Strict`, the SAS version argument is treated as a hint and will set the version preference if it can. JMP will still try to connect if the version cannot be set. The order you put these arguments in can make a difference. The attempt to change the SAS Version is made immediately when that argument is encountered. That can affect the validity of other arguments, particularly for `MetaConnect`. Valid values for `SASVersion` are “9.3” and “9.4”. Note: Using the `SASVersion` argument has the same effect as changing the SAS Server Version on the SAS Integration Preferences page.

Notes

– If no arguments are included and if no profile is saved, the Connect to SAS Metadata Server window appears.

– If you connect to a physical workspace server, there is no metadata server involved, so metadata security is never applied. You must connect to a SAS Metadata Server and

then connect to a logical workspace server. Then metadata security is enforced on the metadata-defined libraries you access.

`Meta Create Profile("profile", <named arguments>)`

Description

Creates a metadata server connection profile and adds it to the current user's set of saved metadata server connection profiles.

Returns

1 if *`profile`* was successfully created, otherwise 0.

Arguments

`profile` A quoted string that contains the name of the created profile. If a profile by the given name already exists, `MetaCreateProfile` fails unless *`Replace`* is specified.

Optional Named Arguments

`HostName("name")` A quoted string that contains the name of the host computer running the SAS Metadata Server that this profile will connect to.

`Port(n)` The port number (*n*) that the SAS Metadata Server is listening for connections on.

`AuthenticationDomain("domain") | AuthDomain("domain")` A quoted string that sets the authentication domain to use for the connection.

`Description("desc") | Desc("desc")` A quoted string that sets a description for this profile.

`Password("password")` A quoted string that contains the password to store in this profile.

`Replace(0|1)` If *`name`* matches a profile that already exists, `Replace` must be specified for the existing profile to be replaced by the one provided. The default value is False (0).

`UserName("username")` A quoted string that contains the user name that this profile uses to connect to the SAS Metadata Server.

`UseSingleSignOn(0|1)` If specified, this profile attempts to use Single Sign-On (currently also known as Integrated Windows Authentication) to connect to the SAS Metadata Server. This option is valid only for connecting to SAS 9.3 or higher Metadata Servers. If `UseSingleSignOn` is True(1), *`UserName`* and *`Password`* cannot be specified. The default value is False (0).

`Meta Delete Profile("name")`

Description

Deletes the named metadata server connection profile from the current user's set of saved metadata server connection profiles

Returns

1 if profile was successfully deleted, otherwise 0.

Argument

`name` A quoted string that contains the name of the profile to delete.

`Meta Disconnect()`

Description

Disconnect the current SAS Metadata Server connection, if any.

Returns

Void.

`Meta Get Environments()`

Description

Returns a list of the SAS Environments that are defined in the SAS Environments definition file, which is configured in the SAS Preferences.

`Meta Get Repositories()`

Description

Gets a list of the repositories available on the current SAS Metadata Server connection.

Returns

A list of repository names as quoted strings.

`Meta Get Servers()`

Description

Get a list of the SAS Servers that are registered in the SAS Metadata Repository to which the session is currently connected.

Returns

A list of server names as quoted strings.

`Meta Get Stored Process("path")`

Description

Get a stored process object from the currently connected SAS Metadata Repository.

Returns

Stored Process scriptable object.

Arguments

`path` Quoted string that is the path to the stored process in metadata, starting at the BIP Tree.

Meta Is Connected()

Description

Determines whether a current connection to a SAS Metadata Server exists.

Returns

1 if a connection exists; 0 otherwise.

Arguments

None.

Meta Set Repository("repositoryName")

Description

Set the SAS Metadata Repository to use for metadata searches.

Returns

1 if setting the repository was successful, 0 otherwise.

Arguments

`repositoryName` A quoted string that contains the name of the repository to make current.

SAS Assign Lib Refs("libref", "path", <"engine">, <"engine options">)

Description

Assign a SAS libref on the active global SAS server connection.

Returns

1 if successful, 0 otherwise.

Arguments

`libref` A quoted string that contains a library reference (8-character maximum) to assign.

`path` A quoted string that contains the full path on the SAS server to the library being assigned.

`engine` Optional, quoted string that contains the engine for the SAS server to use when accessing members of this library.

`engine options` Optional, quoted string that contains the options needed for the engine being used.

SAS Connect(<"machine_name">, <"port">, <named_arguments>)

Description

Connect to a local, remote, or logical SAS server.

Returns

SAS Server scriptable object.

Arguments

`machine_name` (Optional) A quoted string that can contain a physical machine name or the name of a metadata-defined (logical) server. In the first case, the port must be provided. In the second case, a port must *not* be provided. If neither *`name`* nor *`port`* are included, and JMP is running on Windows, a connection to SAS on the local machine (via COM) is attempted, and all named arguments are ignored.

`port` (Optional) A quoted string or integer. If *`name`* is a physical machine name, this is the port on that machine to connect to. If *`name`* is a metadata-defined (logical) server, *`port`* must *not* be included.

Optional Named Arguments

`UserName("name")` A quoted string that contains the user name for the connection.

`Password("password")` A quoted string that contains the password for the connection.

`ReplaceGlobalConnection(0|1)` A Boolean. The default value is `True`. If `True`, and a successful SAS server connection is made, this connection replaces the active SAS connection that becomes the target of other global SAS JSL function calls. If `False`, the global SAS connection is not changed, and the returned SASServer scriptable object should be used to send messages to this server connection.

`ShowDialog(0|1)` A Boolean. The default value is `False`. If `True`, other arguments (except `ReplaceGlobalConnection`) are ignored and the SAS Server Connection window appears. This provides the JSL programmer a way to open the SAS Connect window.

`Prompt(Always|Never|IfNeeded)` A keyword. `Always` means always prompt before attempting to connect. `Never` means never prompt even if the connection attempt fails (just fail and send an error message to the log), and `IfNeeded` (the default value) means prompt if the attempt to connect with the given arguments fails (or is not possible with the information given).

`ConnectLibraries(0|1)` A Boolean. Defaults to the SAS Integration Preference setting governing whether to automatically connect metadata-defined libraries when connecting to a SAS server. If true, all metadata-defined libraries are connected at SAS server connection time, which can be time-consuming. If false, metadata-defined libraries are not connected. To connect specific libraries later, use the `SAS Connect Libref` global function or `Connect Libref` message to a SAS server object.

`SASVersion("<version number>" <,Strict>)` Attempts to change the SAS version preference to the specified value before making the metadata server connection. If the SAS version is already locked to a different version than the one specified, the SASVersion argument will fail. By default, if the SAS version cannot be set, the metadata server connection will still be tried. However, if you include `Strict` as the second argument, the inability to change the SAS version will be treated as an error and JSL processing will stop. If you do not include `Strict`, the SAS version argument is treated as a hint and will set the version preference if it can, but if it cannot it will still try to connect. The order you put these arguments in can make a difference. The attempt to

change the SAS Version is made immediately when that argument is encountered. That can affect the validity of other arguments, particularly for MetaConnect. Valid values for `SASVersion` are "9.3" and "9.4". Note: Using the `SASVersion` argument has the same effect as changing the SAS Server Version on the SAS Integration Preferences page.

Example

```
// prompt for login credentials
Meta Connect( "dev.company.com", 28561 );

sas = SAS Connect( "SASApp" );

// dump some libraries and data sets to the JMP log
Show( sas << Get Librefs() );
Show( sas << Get Data Sets( "Chocolate Enterprises 2017" ));

sas << Import Data( "Chocolate Enterprises 2017", "Products" );

/* The preceding lines produce the following text in the JMP log and import
   the data set: */
sas << Get Librefs:{"BOOKS", "CHOCOENT", "EGSAMP", "TEST", "TST92", "MAPS",
   "SASDATA", "SASHELP", "SASUSER", "SGFDATA", "TEMPDATA", "V6LIB", "WORK"}
sas << Get Data Sets("Chocolate Enterprises 2008"):{"CHOC_DATA",
   "CHOC_SURVEY", "CUSTOMERS", "ORDER_DETAIL", "PRODUCTS", "SALES_ANALYSIS",
   "SALES_SUMMARY"}
```

`Get Librefs` returns the short library names, not the longer logical names. However, you can use either one, and metadata security will still be applied for metadata-defined libraries.

SAS Connect Lib Refs(libref)

Description

Connects a SAS libref on the active SAS server connection.

Returns

1 if successful and 0 otherwise.

SAS Deassign Lib Refs("libref")

Description

De-assign a SAS libref on the active global SAS server connection.

Returns

1 if successful; 0 otherwise.

Arguments

`libref` A quoted string that contains the library reference to de-assign.

SAS Disconnect()

Description

Disconnect the active global SAS connection, if any.

Returns

1 if a SAS connection exists and was successfully disconnected, 0 otherwise.

Arguments

None.

SAS Export Data(dt, "library", "dataset", <named_arguments>)

Description

Exports a JMP data table to a SAS data set in a library on the active global SAS server connection.

Returns

1 if the data table was exported successfully; 0 otherwise.

Arguments

`dt` data table or a reference to a data table.

`"library"` the library to which to export the data table.

`"dataset"` the name of the new SAS data set.

Optional Named Arguments

`Columns(list)|Columns(col1, col2, ...)` A list of columns or a comma-separated list of columns.

`Password("password")` A quoted string that contains the password to serve as the READ, WRITE, and ALTER password for the exported SAS data set. If the exported data set is replacing an existing data set with an ALTER password, this password is used as the ALTER password for overwriting the data set. If *`Password`* is specified, values for *`ReadPassword`*, *`WritePassword`*, and *`AlterPassword`* are ignored.

`ReadPassword("password")` A quoted string that contains the password to serve as the READ password for the exported SAS data set.

`WritePassword("password")` A quoted string that contains the password to serve as the WRITE password for the exported SAS data set.

`AlterPassword("password")` A quoted string that contains the password to serve as the ALTER password for the exported SAS data set. If the exported data set is replacing an existing data set with an ALTER password, this password is used as the ALTER password for overwriting the data set.

`PreserveSASColumnNames(0|1)` A Boolean. If true *and* the JMP data table originally came from SAS, the original SAS column names are used in the exported SAS data set. The default value is `False`.

`PreserveSASFormats(0|1)` A Boolean. If true *and* the JMP data table originally came from SAS, the original SAS formats and informats are applied to the columns in the exported SAS data set. The default value is `True`.

`ReplaceExisting(0|1)` A Boolean. If true, an existing SAS data set with the specified name in the specified library is replaced by the exported SAS data set. If false, a data set with the specified name already exists in the specified library; the export is stopped. The default value is false.

`SaveJMPMetadata(0|1)` Includes SAS 9.4 Extended Attributed to store JMP metadata (such as table script and column properties). Default is 0 (disabled).

`HonorExcludedRows(0|1)` A Boolean. If true, any rows in the JMP data table that are marked as excluded are not exported. The default value is false.

Notes

Information about the export is sent to the log.

SAS Get Data Sets("libref")

Description

Returns a list of the data sets defined in a SAS library.

Returns

A list of quoted strings.

Arguments

`libref` A quoted string that contains the SAS libref or friendly library name associated with the library for which the list of defined SAS data sets is returned.

SAS Get File("source", "dest", "encoding")

Description

Get a file from the active global SAS server connection. JMP creates a FILENAME statement (with an encoding, if specified) and uses it to read the file on the SAS server.

Returns

1 if successful, 0 otherwise.

Arguments

`source` A quoted string that contains the full path of file on the server to be downloaded to the client machine.

`dest` A quoted string that contains the full path on the client machine for where to put the copy of the file downloaded from the server.

`encoding` A quoted string that contains the encoding used in the file (for example, `"utf-8"`). The server must support the specified encoding.

SAS Get File Names("fileref")

Description

Get a list of filenames found in the given fileref on the active global SAS server connection.

Returns

A list of quoted strings.

Arguments

`fileref` A quoted string that contains the name of the fileref from which to retrieve filenames.

SAS Get File Names In Path("path")

Description

Get a list of filenames found in the given path on the active global SAS server connection.

Returns

A list of quoted strings.

Arguments

`path` A quoted string that contains the directory path on the server from which to retrieve filenames.

SAS Get File Refs()

Description

Get a list of the currently defined SAS filerefs on the active global SAS server connection.

Returns

List of two lists. The first list is a list of quoted strings of fileref names. The second is a corresponding list of strings of physical names.

SAS Get Lib Refs(<named arguments>)

Description

Get a list of the currently defined SAS librefs on the current global SAS server connection.

Returns

A list of quoted strings.

Named Arguments

`Friendly Names(0|1)` Optional, Boolean. If `True`, then for any libraries that have friendly names (metadata-defined libraries), the friendly name is returned rather than the 8-character libref.

SAS Get Log()

Description

Retrieve the SAS Log from the active global SAS server connection.

Returns

A quoted string.

SAS Get Output()

Description

Retrieve the listing output from the last submission of SAS code to the current global SAS server connection.

Returns

A quoted string.

SAS Get Results()

Description

Retrieve the results of the previous SAS Submit as a scriptable object, which allows significant flexibility in what to do with the results.

Returns

A SAS Results Scriptable object.

SAS Get Var Names(*string*, <"dataset">, <password("password")>)

Description

Retrieves the variable names contained in the specified data set on the current global SAS server connection.

Returns

A list of quoted strings.

Arguments

`string` A quoted string that contains one of the following:

- The name of the SAS Library containing the SAS data set to be imported. In that case, the *`dataset`* name argument is required.
- The full member name of the SAS data set to be imported, in the form "libname.membername".
- The SAS Folders tree path to a logical SAS data table to be imported. This option requires a connection to a SAS 9.3 or higher Metadata Server.

`dataset` (Optional) A quoted string that contains the name of the data set from which to retrieve variable names.

`password("password")` (Optional) A quoted string that contains the read password for the data set. If this is not provided and the data set has a read password, the user is prompted to enter it.

`SAS Import Data(`*`string`*`, <"dataset">, <named arguments>)`

Description

Import a SAS data set from the active global SAS server connection into a JMP table.

Returns

JMP Data Table object.

Arguments

`string` A quoted string that contains one of the following:

- The name of the SAS Library containing the SAS data set to be imported. In that case, the *"dataset"* name argument is required. The name can be a friendly metadata library name or a SAS 8-character library name.
- The full member name of the SAS data set to be imported, in the form "libname.membername".
- The SAS Folders tree path to a logical SAS data table to be imported. This option requires a connection to a SAS 9.3 or higher Metadata Server.

`dataset` (Optional) A quoted string that contains the name of the data set.

Optional Named Arguments

`Columns("list")|Columns(col1, col2, ...)` A quoted string list or multiple strings that contain the names of columns to include in the import.

`ConvertCustomFormats(0|1)` The default value is True (1). If True and custom formats are found in the SAS data set being imported, an attempt is made to convert the SAS custom formats to JMP value labels for those columns.

`Invisible(0|1)` The default value is False (0). If true, the JMP data table is hidden from view. The data table appears only in the JMP Home Window and the Window menu. Hidden data tables remain in memory until they are explicitly closed, reducing the amount of memory that is available to JMP. To explicitly close the hidden data table, call `Close(dt)`, where *dt* is the data table reference returned by SASImportData.

`Where("filter")` A quoted string that contains the filter to use when importing data, as in `Where("salary<50000")`.

`Password("password")` A quoted string that contains the read password for the data set. If this is not provided and the data set has a read password, the user is prompted to enter it.

`UseLabelsForVarNames(0|1)` If `True`, the labels from the SAS data set become the column names in the resulting JMP table. If `False`, the variable names from the SAS data set become the column names in the JMP table. The default value is `False`.

`RestoreJMPMetadata(0|1)` Includes SAS 9.4 Extended Attributed to store JMP metadata. Default is 0 (disabled).

`Sample(named arguments)` optional, named. Allows a random sample of the SAS data set to be imported into JMP. If both Where and Sample are specified, the WHERE clause is used to filter the SAS data set first, and then a random sample of the resulting rows is taken based on the arguments supplied to Sample. Note that Sample uses PROC SURVEYSELECT on the SAS server, which is available only if the SAS/STAT package is licensed and installed on that server. The documentation for PROC SURVEYSELECT might be helpful in understanding how sampling is performed. By default (if no arguments are supplied), a 5% simple random sample is taken. Here are the available arguments (all optional):

- `Simple | Unrestricted`: If `Simple` is specified, sampling is performed without replacement. If `Unrestricted` is specified, sampling is performed with replacement. These two options are mutually exclusive and only one can be specified.
- `SampleSize(int) | N(int)`: Total number of rows for the sample, or number of rows per strata level for stratified sampling
- `SampleRate(number) | Rate(number) | Percent(number)`: Specifies the sampling rate. For stratified sampling, the rate is applied to each strata level. Note that the supplied value is assumed to be a percentage, so `SampleRate(3.5)` means a 3.5% sampling rate.
- `Strata({col1, col2, ...}) | Strata(col1, col2, ...)`: Perform stratified random sampling using the column names supplied as Strata variables.
- `NMin(int)`: Minimum number of rows (either overall or per strata level for stratified sampling) to return. Only applies to rate-based sampling.
- `NMax(int)`: Maximum number of rows (either overall or per strata level for stratified sampling) to return. Only applies to rate-based sampling.
- `Seed(int)`: Number to use as the seed for sampling. Useful for replicating the same sample. By default, the seed is a random number based on time of day. See PROC SURVEYSELECT documentation.
- `OutputHits(0|1)`: Boolean; the default value is false. When doing Unrestricted sampling, if the same row of the input data set is selected more than once, by default that row still appears only once in the resulting JMP data table, with the NumberHits column indicating the number of times that the row was selected. Setting OutputHits to true causes an input row that is selected multiple times to appear multiple times in the resulting JMP data table.
- `SelectAll(0|1)`: Boolean, the default value is true. If `SelectAll` is true, PROC SURVEYSELECT selects all stratum rows whenever the stratum sample size exceeds the total number of rows in the stratum. If `SelectAll` is false and PROC SURVEYSELECT finds a case where the stratum sample size exceeds the total number

of rows in a given stratum, an error results and sampling fails. `SelectAll` only applies to Simple random sampling.

`SQLTableVariable(0|1)` If `True`, an SQL table variable is created in the resulting JMP table that shows the SQL that was submitted to SAS to obtain the data. If `False`, the SQL table variable is not created. The default value is `True`. If an SQL table variable is created and the data set required a read password, the password is masked.

SAS Import Query("sqlquery", <named arguments>)

Description

Execute the requested SQL query on the current global SAS server connection, importing the results into a JMP data table.

Returns

JMP Data Table object.

Arguments

`sqlquery` A quoted string that contains the SQL query to perform and from which to import the result.

Optional Named Arguments

`ConvertCustomFormats(0|1)` The default value is true. If true and custom formats are found in the SAS data set being imported, an attempt is made to convert the SAS custom formats to JMP value labels for those columns.

`Invisible(0|1)` The default value is false. If true, the JMP data table is hidden from view. The data table appears only in the JMP Home Window and the Window menu. Hidden data tables remain in memory until they are explicitly closed, reducing the amount of memory that is available to JMP. To explicitly close the hidden data table, call `Close(dt)`, where *`dt`* is the data table reference returned by `SAS Import Query`.

`UseLabelsForVarNames(0|1)` The default value is true. If `True`, the labels from the SAS data set become the column names in the resulting JMP table. If `False`, the variable names from the SAS data set become the column names in the JMP table.

`RestoreJMPMetadata(0|1)` Includes SAS 9.4 Extended Attributes to store JMP metadata. Default is 0 (disabled).

`SQLTableVariable(0|1)` The default value is true. If `True`, an SQL table variable is created in the resulting JMP table that shows the SQL that was submitted to SAS to obtain the data. If `False`, the SQL table variable is not created. If an SQL table variable is created and the data set required a read password, the password is masked.

SAS Is Connected()

Description

Discovers whether there is an active global SAS server connection.

Returns

1 if an active global SAS connection exists, 0 otherwise.

SAS Is Local Server Available()

Description

Returns True if a local SAS Server is available; otherwise, returns False.

SAS Load Text File("path")

Description

Download the file specified in path from the active global SAS server connection and retrieve its contents as a quoted string.

Returns

A quoted string.

Arguments

"path" A quoted string that contains the full path on the server of the file to download and retrieve the contents as a string.

SAS Name("name")

SAS Name({list of names})

Description

Converts JMP variable names to SAS variable names by changing special characters and blanks to underscores and various other transformations to produce a valid SAS name.

Returns

A quoted string that contains one or more valid SAS names, separated by spaces.

Argument

"name" A quoted string that represents a JMP variable name; or a list of quoted JMP variable names.

SAS Open For Var Names("path")

Description

Opens a SAS data set only to obtain the names of its variables, returning those names as a list of quoted strings.

Returns

A list of variable names in the file.

Argument

path A quoted string that is a pathname of a SAS data set.

SAS Send File("source", "dest", "encoding")

Description

Send a file from the client machine to the active global SAS server connection. JMP creates a FILENAME statement (with an encoding, if specified) and uses it to save the file on the SAS server.

Returns

1 if successful, 0 otherwise.

Arguments

`source` A quoted string that contains the full path of the file on the client machine to be uploaded to the server.

`dest` A quoted string that contains the full path on the server that receives the file uploaded from the client machine.

`encoding` A quoted string that contains the encoding used in the file (for example, "utf-8"). The server must support the specified encoding.

SAS Submit("sasCode", <named arguments>)

Description

Submit some SAS code to the active global SAS server connection.

Returns

1 if successful, 0 otherwise.

Arguments

`sasCode` A quoted string that contains the SAS code to submit.

Optional Named Arguments

`Async(0|1)` A Boolean. If True (1), the submit occurs asynchronously (in the background). Use the `Get Submit Status()` message on the SAS Server Scriptable Object to determine the status of the submit. The default value is False (0).

`ConvertCustomFormats(0|1)` A Boolean. When SAS data sets generated by submitted SAS code are imported into JMP after the submit completes (see `Open Output Datasets`), the value of `ConvertCustomFormats` determines whether an attempt is made to convert any custom formats found on columns in the SAS data to JMP value labels. The default value is True (1).

`DeclareMacros(var1, var2, ...)` JSL variable names. Provides a simple way to pass the values of JSL variables to SAS as macro variables. Each JSL variable specified should evaluate to a quoted string or numeric value. Fully qualified JSL variables names, only the variable name is sent to SAS. For example, `namespace:variable_name` becomes `variable_name` in SAS.

`GetSASLog(<Boolean|OnError>, <JMPLog|Window>)` A Boolean. If no arguments are supplied, the SAS Log is retrieved and displayed in the location indicated in SAS Integration Preferences. The first argument to `GetSASLog` can be either a Boolean value

or the keyword OnError. If a Boolean value is supplied, true means display the SAS Log, and false means not to display it. OnError instructs JMP to only show the SAS Log if an error occurred in the submit. The second argument to GetSASLog tells JMP where to display the SAS Log. If JMPLog is specified, the SAS Log is appended to the JMP Log. If Window is specified, the SAS Log is opened in a separate window.

`GraphicsDevice(`*`string`*`) or GDevice(`*`string`*`)` A quoted string that specifies a value for the GDEVICE SAS option to be used for graphics generated by the submitted SAS code. The value must be a valid SAS graphics device. The default value is determined in Preferences.

`Interactive(0|1)` JMP includes the `QUIT` statement in the generated wrapper code. Interactive PROCs work even if JMP is generating the ODS wrapper. On every SUBMIT, specify the argument that is part of an interactive sequence. Otherwise, QUIT will be generated in both the prologue-generated and epilogue-generated code.

`NoOutputWindow` A Boolean. If `True`, the SAS Output window containing the listing output from the submission does not appear. The default value is `False`.

`ODS(0|1)` If true, additional SAS code is submitted causing ODS results to be generated for the submitted SAS code. The default value is determined in Preferences.

`ODSFormat(`*`string`*`)` A quoted string that determines the format of generated ODS results. Valid values are “HTML”, “RTF”, and “PDF”. The default value is determined in Preferences.

`ODSGraphics(0|1)` If true, ODS statistical graphics are generated for the submitted SAS code. Setting `ODSGraphics` to true causes ODS to also be set to true. The default value is determined in Preferences.

`ODSStyle(`*`string`*`)` A quoted string that specifies the ODS Style to use when generating ODS results. *`String`* must be a valid SAS Style. The default value is determined in Preferences.

`ODSStyleSheet(path)` A quoted string that specifies a local CSS style sheet to use when formatting generated ODS results. *`Path`* must be a path to a CSS file valid for the client machine (the machine running JMP). The default value is determined in Preferences.

`OnSubmitComplete(script)` A quoted string that specifies a JSL script that should be run when the submit completes. This is especially useful for asynchronous submits. If script is the name of a defined JSL function, that function is executed, with the SAS Server scriptable object passed as the first argument.

`OpenODSResults(0|1)` If true, ODS results that are generated by the submitted SAS code (due to ODS being true) are automatically opened after the submit completes. The default value is True (1).

`OpenOutputDatasets(<All|None|dataset1, dataset2, ...>)` JMP detects when submitted SAS code creates new SAS data sets. `OpenOutputDatasets` (which can be abbreviated `OutData`) determines what, if anything, is done with those data sets with the SAS Submit completes. If `All` is specified, all data sets generated by the SAS code are imported into JMP when the SAS Submit completes. If `None` is specified, none of the

generated data sets are imported. If there are specific data sets known to be generated by the submitted SAS code that you want to be imported into JMP when the SAS submit completes, you can alternative provide their names, and only the requested data sets are imported. The default value is determined in Preferences.

`Title(`*`string`*`)` A quoted string that specifies the window title to use for the window that displays ODS output from the submit.

`SAS Submit File("filename", <named arguments>)`

Description

Submit a SAS code file to the active global SAS server connection.

Returns

1 if successful; 0 otherwise.

Arguments

`filename` A quoted string that contains the name of file containing SAS code to submit.

Named Arguments

Same as for SAS Submit.

SQL Functions

Note: Database table names that contain the characters `$# -+/%()&|;?` are not supported.

`As SQL Expr(x, <style>)`

Description

Converts an expression to code that you can use in an SQL Select statement. Use `Expr(...)` for literal expressions. Use `NameExpr(name)` for expressions stored in a variable. Otherwise, the expression returns the expression to convert.

Returns

A quoted string that contains the expression converted to valid SQL syntax for use in an SQL Select statement.

```
New SQL Query(Connection
("ODBC:connection_string")|("SAS:connection_string"),
Select(Column("column", "t1")), From(Table("table", <Schema("schema")>,
```

```
<Alias("t1")>)), <Options(JMP 12 Compatible(1)|JMP 13 Compatible(1)|Run on
Open(1)))>
```

```
New SQL
Query(Connection("ODBC:connection_string;")|("SAS:connection_string;"),
Custom("SELECT col1, col2, col3 FROM table;")), <Options(JMP 12
Compatible(1)|JMP 13 Compatible(1)|Run on Open(1))>
```

Creates an SQL Query object for the specified connection, columns, data table, or for the custom SQL query.

Returns

A data table that contains the queried data. The data table includes the quoted SQL query string and table scripts for modifying and updating the query.

Arguments

`Connection` The quoted string for an ODBC or SAS connection.

`Select` The column that you want to select and its alias.

`From` The table that is queried and the optional schema and column alias.

`Custom` An SQL statement that selects columns from the specified table.

`Version` The minimum JMP version required to open the query. If this condition is not met, a message regarding compatibility is written to the log, and the query does not open.

`Options` Boolean. `JMP 12 Compatible` is included in generated scripts when you select the Query Builder preference to create a JMP 12 compatible option or select the corresponding Query Builder red triangle menu option. The option enables JMP 12 users to run a JMP 13 query that might contain compatibility issues. Include `Run on Open(1)` to run the query when opened rather than opening the query in edit mode.

Example

```
New SQL Query(
   Connection(
      "ODBC Connection String..."
   ),
   QueryName( "g6_Movies" ),
   Select( Column( "ItemNo", "t1" ), Column( "LengthMins", "t1" ), Column(
   "Genre", "t1" ) ),
   From( Table( "g6_Movies", Schema( "SQBTest" ), Alias( "t1" ) ) )
) << Run Background( On Run Complete( dt = queryResult ) );

Show( dt );
```

Notes

- Query Builder creates a symbol called `queryResult` in the context of an `On Run Complete()` script. This is a reference to the data table imported by the query. `queryResult` enables you to assign a global variable to the table for later use.
- `New SQL Query()` always closes the ODBC connection after performing the query.

- After you run `New SQL Query()`, the database connection that you make appears in the Database > Open Table window as an available connection. But the connection you see in the Database > Open Table window is due to JMP keeping track of how to connect to that database; no ODBC connection is left open.
- There is not a command to close an ODBC connection established with `New SQL Query()` because `New SQL Query()` always closes the connection after performing the query.
- After you run `New SQL Query()`, the database connection appears in the Database > Open Table window as an available connection. JMP is keeping track of how to connect to that database; the ODBC connection was not left open.

```
Query(<<dt1|Table(dt1, alias1)>, ..., <dtN, aliasN)>>, <private |
invisible>, <scalar>, sqlStatement )
```

Description

Performs a SQL query on selected data tables.

Returns

The result of the query, either a data table or a single value.

Arguments

`dt1, dtN` (Optional) A variable that has been assigned to the data table.

`Table` (Optional) Passes a reference to the data table.

`alias1, aliasN` Specifies the alias of the database table.

`private` (Optional) Avoids showing the resulting data table. Using a private data table speeds the process of getting to the data; it does not save the computer from allocating the memory necessary to hold the data table data.

`invisible` (Optional) Hides the resulting data table from view. The data table appears only in the JMP Home Window and the Window menu. Hidden data tables remain in memory until they are explicitly closed, reducing the amount of memory that is available to JMP. To explicitly close the hidden data table, call `Close(dt)`, where *`dt`* is the data table reference.

`scalar` (Optional) Indicates that the query returns a single value.

`sqlStatement` Required. The SQL statement, most likely a SELECT statement. The statement must be the last argument.

Example

The following example selects all data for students who are older than 14 years of age.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
result = Query( Table( dt, "t1" ), "SELECT * FROM t1 WHERE age > 14;" );
```

See Also

Appendix A, “SQL Functions Available for JMP Queries”

Statistical Functions

Arc Finder(X(col), Y(col), Group(lot, wafer))

Description

Finds arcs in the point data and creates a new column that identifies the arcs.

Example

```
dt = Open( "$SAMPLE_DATA/Wafer Stacked.jmp" );
Arc Finder(
   Group( :Lot, :Wafer ),
   X( :X_Die ),
   Y( :Y_Die ),
   Min Distance( 12 ), // minimum distance among 3 points to seed an arc
   Min Radius( 15 ), // minimum radius of the acceptable arc
   Max Radius( 2000 ), // maximum radius of acceptable arc
   Max Radius Error( 2 ), // how close a point needs to be added
   Min Arc Points( 5 ), // how many points to define an arc
   Number of Searches( 500 ), // how many random probes of data
   Max Number Arcs( 3 ) // number of arcs searched for
);
dt << Color or Mark by Column( :Arc Number );
dt << Graph Builder(
   Size( 1539, 921 ),
   Variables( X( :X_Die ), Y( :Y_Die ), Wrap( :Lot_Wafer Label ), Color( :Arc
   Number ) ),
   Elements( Points( X, Y, Legend( 6 ) ) )
);
```

Notes

- The function is scaled for data that have a range of 30 to 50 units.
- The function is suitable only for data that are subset to the interesting defect points.
- It is not suitable when the density of points is high.

ARIMA Forecast(column, length, model, estimates, from, to)

Description

Determines the forecasted values for the specified rows of the specified column using the specified model and estimates.

Returns

A vector of forecasted values for *column* within the range defined by *from* and *to*.

Arguments

`column` A data table column.

`length` Number of rows within the column to use.

`model` Messages for Time Series model options.

`estimates` A list of named values that matches the messages sent to `ARIMA Forecast()`. If you perform an ARIMA Forecast and save the script, the estimates are part of the script.

`from, to` Define the range of values. Typically, *`from`* is between 1 and *`to`*, inclusive. If *`from`* is less than or equal to 0, and if *`from`* is less than or equal to *`to`*, the results include filtered predictions.

`Best Partition(xindices, yindices, <<Ordered, <<Continuous Y, <<Continuous X)`

Description

Experimental function to determine the optimal grouping.

Returns

A list.

Arguments

`xindices, yindices` Same-dimension matrices.

`Col Cumulative Sum(name, <By var, ...>)`

`Cumulative Sum(name)`

Description

Returns the cumulative sum for the current row. `Col Cumulative Sum` supports By columns, which do not need to be sorted.

Arguments

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

`Col Maximum(name, <By var, ...>)`

`Col Max(name)`

Description

Calculates the maximum value across all rows of the specified column. The result is internally cached to speed up multiple evaluations.

Returns

The maximum value that appears in the column.

Arguments

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Col Mean(name, <By var, ...>)`

Description

Calculates the mean across all rows of the specified column. The result is internally cached to speed up multiple evaluations.

Returns

The mean of the column.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Col Median(name, <By var, ...>)`

Description

Calculates the median across all rows of the specified column. The ordering is cached internally to speed up multiple evaluations.

Returns

The median of the column.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Col Minimum(name, <By var, ...>)`

`Col Min(name)`

Description

Calculates the minimum value across all rows of the specified column. The result is internally cached to speed up multiple evaluations.

Returns

The minimum value that appears in the column.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Col Moving Average(name, options, <By var, ...>)`

`Moving Average(name, options)`

Description

Returns the moving average over a given interval based at the current row. `Col Moving Average` supports By columns.

Arguments

`name` A column name.

`Weighting(1|0|n)` Required positional argument. Determines how the values are weighted. 1 indicates uniform weighting. 0 indicates incremental weighting (a ramp or triangle). Any other number is the parameter for an exponential moving average (EWMA or EMA).

`Before(1|0|n)` Positional argument. Controls the size of the range (or window) by including the specified number of items before the current item in the average (in addition to the current item). The default value, -1, means all of the preceding items.

`After(1|0|n)` Positional argument. Controls the size of the range (or window) by including the specified number of items after the current item in the average (in addition to the current item). The default value, 0, means no following items.

`Partial Window is Missing` Boolean positional argument. Controls how missing values are treated. By default, missing values are ignored. 0 computes the average of partial windows.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Examples

```
// equal weighting of a five-item lagging range
Col Moving Average( x, 1, 4 );

// ramp weighting of all preceding items
Col Moving Average( x, 0 );

// triangle weighting of a five-item centered range
Col Moving Average( x, 0, 2, 2 );

// exponential weighting of all preceding items
Col Moving Average( x, 0.25 );
```

Col N Missing(name, <By var, ...>)

Description

Calculates the number of missing values across all rows of the specified column. The result is internally cached to speed up multiple evaluations.

Returns

The number of missing values in the column.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

Col Number(name, <By var, ...>)

Description

Calculates the number of nonmissing values across all rows of the specified column. The result is internally cached to speed up multiple evaluations.

Returns

The number of nonmissing values in the column.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Col Quantile(name, p, <ByVar>)`

Description

Calculates the specified quantile *p* across all rows of the specified column. The result is internally cached to speed up multiple evaluations.

Returns

The value of the quantile.

Argument

`name` A column name.

`p` A specified quantile *p* between 0 and 1.

`ByVar` (Optional) A By group.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Col Quantile( :height, .5 );
   63
```

63 is the 50th percentile, or the median, of all rows in the height column.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Col Rank(column, <ByVar, ...>, <<tie("average"|"arbitrary"|"row"|"minimum"))`

Description

Ranks each row's value, from 1 for the lowest value to the number of columns for the highest value. Ties are broken arbitrarily.

Arguments

`column` The column to be ranked.

`ByVar` (Optional) A By variable to compute statistics across groups of rows.

`<<tie` Determines how the tie is broken. A tie occurs when the values being ranked are the same. For the data `[33 55 77 55]`, `33` has rank 1 and 77 has rank 4, and the question is how to assign ranking for the `55`s. `average` reports the average of the possible rankings, 2.5, for both 55s. `arbitrary` matches JMP 12 behavior by assigning the possible rankings in an unspecified order, which could be 2 and 3 or 3 and 2. `row` assigns the ranks in the order that they originally appear. (The first 55 would be 2 and the second 55 would be 3.)

`minimum` gives both values the lowest possible rank, 2.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

Col Simple Exponential Smoothing(column, alpha, <ByVar>)

Description

Returns the simple exponential smoothing prediction for the current row using smoothing weight *alpha*.

Arguments

`column` The column of time series observations.

`alpha` The smoothing weight.

`ByVar` (Optional) A By variable to compute predictions across groups of rows. By variables do not need to be presorted.

Notes

The predicted value for row *t* is given by the following:

```
Predicted[t] = alpha * Observed[t-1] + (1-alpha) * Predicted[t-1]
```

By definition, `Predicted[1] = Observed[1]`.

Col Standardize(name,<By var, ...>)

Description

Calculates the column mean divided by the standard deviation across all rows of the specified column.

Returns

The standardized mean.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. If a By variable is specified, the values are standardized against the mean and standard deviation of their corresponding By variable group.

Notes

Standardizing centers the variable by its sample standard deviation. Thus, the following commands are equivalent:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << New Column( "stdht", Formula( Col Standardize( height ) ) );
dt << New Column( "stdht2",
  Formula( (height - Col Mean( height )) / Col Std Dev( height ) )
);
```

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

Col Std Dev(name,<By var, ...>)

Description

Calculates the standard deviation across rows in a column. The result is internally cached to speed up multiple evaluations.

Returns

The standard deviation.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Col Sum(name,<By var, ...>)`

Description

Calculates the sum across rows in a column. Calculating all missing values (`Col Sum(.,.)`) returns missing. The result is internally cached to speed up multiple evaluations.

Returns

The sum.

Argument

`name` A column name.

`By var` (Optional) A By variable to compute statistics across groups of rows. Use the By variable in a column formula or in a `For Each Row()` function.

Notes

If a data value is assigned by a column property (such as Missing Value Codes), use `Col Stored Value()` to base the calculation on the value stored in the column instead.

See Also

"Col Stored Value(<dt>, col, <row>)"

`Fit Censored(Distribution("name"), YLow(vector) | Y(Vector), <YHigh(vector)>, <Weight(vector)>, <X(matrix)>, <Z(matrix)>, <HoldParm(vector)>, <Use random sample to compute initial values(percent)>, <Use first N observations to compute initial values(nobs)>)`

Description

Fits a distribution using censored data.

Returns

A list that contains parameter estimates, the covariance matrix, the log-likelihood, the AICc, the BIC, and a convergence message. See *Fitting Linear Models*.

Arguments

`Distribution("name")` The quoted name of the distribution to fit.

`YLow(vector) | Y(Vector)` If you do not have censoring, then use `Y` and an array of your data, and do not specify `YHigh`. If you do have censoring, then specify `YLow` and `YHigh` as the lower and upper censoring values, respectively.

Optional Arguments

`YHigh(vector)` A vector that contains the upper censoring values. Specify this only if you have censoring and also specify YLow.

`Weight(vector)` A vector that contains the weight values.

`X(matrix)` The regression design matrix for location.

`Z(matrix)` The regression design matrix for scale.

`HoldParm(vector)` An array of specified parameters. The parameters should be nonmissing where they are to be held fixed, and missing where the are to be estimated. This is primarily used to test hypotheses that certain parameters are zero or some other specific value.

`Use random sample to compute initial values(percent)` A percent of the observations to be used in the computation of the initial values. Specify this if the data vector is large.

`Use first N observations to compute initial values(nobs)` A number of observations at the start of the data vector to be used in the computation of the initial values. Specify this if the data vector is large.

Fit Circle(Xvec, Yvec)

Description

Fits a circle that best goes through three or more points using a least squares approach. If only three points are specified, a direct solution can be found, and the sum of squared errors is zero.

Returns

A list that contains the X and Y coordinates of the center point of the circle, the length of the radius, and the sum of squared errors.

Arguments

`Xvec` Vector of X coordinates of three or more points.

`Yvec` Vector of Y coordinates of three or more points.

Syntax

```
{Xcenter, yCenter, radius, SSE} = Fit Circle(Xvec, Yvec)
```

Hier Clust(x)

Description

Returns the clustering history for a hierarchical clustering using Ward's method (without standardizing data).

Argument

`x` A data matrix.

IRT Ability(Q1, <Q2, Q3, ... Qn,> parmMatrix)

Description

Returns scores for the latent variable in an item response theory model with n binary items and a matrix of known parameters. The parameter matrix should contain as many rows as there are parameters in the model and as many columns as there are items in the analysis.

Arguments

`Q1, Q2, ..., Qn` A set of n binary items.

`parmMatrix` A matrix of parameters from an item response theory model.

`KDE(vector, <named arguments>)`

Description

Returns a kernel density estimator with automatic bandwidth selection.

Argument

`vector` A vector.

Optional Named Arguments

`<<weights` Must be a vector of the same length as vector, and can contain any nonnegative real numbers. `Weights` represents frequencies, counts, or similar concepts.

`<<bandwidth(n)` A nonnegative real number. Enter a value of 0 to use the bandwidth selection argument.

`<<bandwidth scale(n)` A positive real number.

`<<bandwidth selection(n)` *n* must be 0, 1, 2, or 3, corresponding to Sheather and Jones, Normal Reference, Silverman rule of thumb, or Oversmoother, respectively.

`<<kernel(n)` *n* must be 0, 1, 2, 3, or 4, corresponding to Gaussian, Epanechnikov, Biweight, Triangular, or Rectangular, respectively.

`LenthPSE(x)`

Description

Returns Lenth's pseudo-standard error of the values within a vector.

Argument

`x` A vector.

`Max()`

See "Maximum(var1, var2, ...)".

`Maximum(var1, var2, ...)`

`Max(var1, var2, ...)`

Description

Returns the maximum value of the arguments or of the values within a single matrix or list argument. If multiple arguments are specified, they must be all numeric values or all quoted strings.

`Mean(var1, var2, ...)`

Description

Returns the arithmetic mean of the arguments or of the values within a single matrix or list argument.

`Median(var1, var2, ...)`

Description

Returns the median of the arguments or of the values within a single matrix or list argument.

`Min()`

See "Minimum(var1, var2, ...)".

`Minimum(var1, var2, ...)`

`Min(var1, var2, ...)`

Description

Returns the minimum value of the arguments or of the values within a single matrix argument. If multiple arguments are specified, they must be either all numeric values or all quoted strings.

`N Missing(expression)`

Description

Rowwise number of missing values in variables specified.

`Number(var1, var2, ...)`

Description

Rowwise number of nonmissing values in variables specified.

`Product(i=initialValue, limitValue, bodyExpr)`

Description

Multiplies the results of *bodyExpr* over all *i* until the *limitValue* and returns a single product.

`Quantile(p, arguments)`

Description

Returns the p^{th} quantile of the arguments. The first argument can be a scalar or a matrix of values between 0 and 1. The remaining arguments can also be specified as values within a single matrix or list argument.

`Range(var1, var2, ...)`

Description

Returns the minimum and maximum values of the arguments. The result is returned as a two-element row vector that contains the minimum and the maximum.

`Robust PCA(X, <Lambda(2/sqrt(max(nrow, ncol)))>, <tolerance=1e-10>, <maxit(75)>, <Center(1)>, <Scale(1)>)`

Description

Performs a sequence of singular value decompositions and thresholding steps to decompose the data matrix into a low-rank matrix and a sparse matrix of residuals.

Returns

`A` The low-rank matrix estimation.

`E` The sparse matrix of residuals.

`S` A vector of singular values.

Arguments

`X` A data matrix.

`Lambda` Specifies a value greater than 0 that determines the sparsity of the matrix of residuals. For larger values of Lambda, the matrix of residuals is more sparse.

`tolerance` The convergence criterion.

`maxit` The maximum number of SVD iterations.

`Center` Centers the data prior to performing the SVD iterations.

`Scale` Scales the data prior to performing the SVD iterations

`Std Dev(var1, var2, ...)`

Description

Rowwise standard deviation of the variables specified.

`Sum(var1, var2, ...)`

Description

Rowwise sum of the variables specified. Calculating all missing values (`Sum(.,.)`returns missing.

`SSQ(x1, ...)`

Description

Returns the sum of squares of all elements. Takes numbers, matrices, or lists as arguments and returns a scalar number. Skips missing values.

`Summarize(<dt>, <by>, <count>, <sum>, <mean>, <min>, <max>, <stddev>, <corr>, <quantile>, <first>)`

Description

Gathers summary statistics for a data table and stores them in global variables.

Returns

None.

Arguments

`dt` Optional positional argument: a reference to a data table. If this argument is not in the form of an assignment, then it is considered a data table expression.

All other arguments are optional and can be included in any order. Typically, each argument is assigned to a variable so you can display or manipulate the values further.

`name=By(col | list | Eval)` Using a BY variable changes the output from single values for each statistic to a list of values for each group in the BY variable.

`Summarize YByX(X(<x columns>, Y (<y columns>), Group(<grouping columns>), Freq(<freq column>), Weight(<weight column>))`

Description

Calculates all Fit Y by X combinations on large-scale data sets.

Returns

A data table of *p*-values and LogWorth values for each Y and X combination.

Arguments

`X(col)` The factor columns used in the fit model.

`Y(col)` The response columns used in the fit model.

`Group(gcol)` The group of columns used in the fit model.

`Freq(col)` The frequency (for each row) column used in the fit model.

`Weight(col)` The importance (or influence) column used in the fit model.

Notes

Performs the same function as the Response Screening platform.

`Summation(init, limitvalue, body)`

Description

Summation sums the results of the *body* statement(s) over all *i* to return a single value.

`Tolerance Limit(1-alpha, p, n)`

Description

Constructs a *1-alpha* confidence interval to contain proportion *p* of the means with sample size *n*.

Transcendental Functions

`Arrhenius(n)`

Description

Converts the temperature *n* to the value of explanatory variable in Arrhenius model.

Returns

11604.5181215503/(n+273.15)

Argument

n Temperature in Celsius.

Notes

This is frequently used as a transformation.

`Arrhenius Inv(n)`

Description

The inverse of the Arrhenius function. Converts the value *n* to the temperature in Celsius.

Returns

(11604.5181215503/n)-273.15

Argument

n The value of the converted explanatory variable in Arrhenius model.

Notes

This is frequently used as a transformation.

`Beta(a, b)`

Description

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Returns

Returns the beta function.

Arguments

`a, b` numbers

`Cytometry Logicle(x, T, W, M, A)`

Description

Computes a cytometry logicle transformation. For more information about the logicle transformation, see Moore and Parks (2012).

`Cytometry Logicle Inverse(y, T, W, M, A)`

Description

Computes the inverse cytometry logicle transformation. For more information about the logicle transformation, see Moore and Parks (2012).

`Digamma(n)`

Description

The derivative of the log of the gamma function (`LGamma`).

Returns

The digamma function evaluated at *n*.

Argument

`n` A number

`Exp(a)`

Description

Raises e to the power *a*.

Returns

e^a.

Argument

`a` A number

Equivalent Expression

`e()^a`

`ExpM1(x)`

Description

Returns a more accurate calculation of *Exp(x)-1* when *x* is very small.

`Factorial(n)`

Description

Multiplies all numbers 1 through *n*, inclusive

Returns

The product.

Arguments

`n` Any integer

Notes

One and only one argument must be specified.

`FFT({list}, <named arguments>)`

Description

Conducts a Fast Fourier Transformation (FFT) on a list of matrices.

Returns

The function takes one matrix, or a list of matrices for complex numbers. The returned value is a list of two matrices with the same dimensions as the first argument.

Argument

`List` A list of one or two matrices. If one is provided, it is considered to be the real part. If two are provided, the first is the real part and the second is the imaginary part. Both matrices must have the same dimensions, and both must have more than one row.

Optional Named Arguments

`<<inverse(Boolean)` If true (1), an inverse FFT is conducted.

`<<multivariate(Boolean)` If true (1), a multivariate FFT is conducted. If false(0), a spatial FFT is conducted.

`<<scale(number)` Multiplies the return values by the specified *number*.

`Fit Transform To Normal(Distribution("name"), Y(vector), <Freq(vector))`

Description

Fits a transformation to normality for a vector of data. This includes the Johnson Sl, Johnson Sb, Johnson Su, and GLog distributions.

Returns

A list that contains parameter estimates, the covariance matrix, the log-likelihood, AICc, a convergence message, and the transformed values. See *Fitting Linear Models*.

`Gamma(t, <limit>)`

Description

The gamma function of *x*, or for each row if *x* is a column:

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$$

Returns

The gamma.

Arguments

`t` a number or a column

`limit` optional limit. The default is infinity.

Notes

`Gamma(t, limit)` is the same integral as `Gamma(t)` but with the limit of integration that is defined instead of infinity.

`LGamma(t)`

Description

Returns the log gamma function for *t*, which is the natural log of gamma.

`Ln(n)`

Description

Returns the natural logarithm (base *e* logarithm) of *n*.

`Log(n, <base>)`

Description

Returns the natural logarithm (base *e* logarithm) of *n*. An optional second argument lets you specify a different base. For example, `Log(n,3)` for the base 3 logarithm of *n*. The Log argument can be any numeric expression. The expression `Log(e())` evaluates as 1, and `Log(32,2)` is 5.

`Log10(n)`

Description

Returns the common (base 10) logarithm of *n*.

`Log1P(n)`

Description

Same as Log(1 + *x*), except that it is more accurate when *x* is very small.

`Logist(x)`

Description

Returns 1/(1+Exp(-x)), which converts a number in the domain -∞...+∞ into range 0...1. The function is useful in logistic regression.

`Logist Percent(p)`

Description

Similar to the `Logist()` function but with the result scaled from 0 to 100.

`Logit(p)`

Description

Returns `log(p/(1-p))`.

`Logit Percent(p)`

Description

Similar to the `Logit()` function with the argument 0 to 100 rather than 0 to 1.

`N Choose K(n, k)`

Description

This function returns the number of n things taken k at a time ("n choose k") and is computed in the standard way using factorials, as $n!/(k!(n-k)!)$. For example, `NChooseK(5,2)` evaluates as 10.

Notes

This is implemented internally in JMP using lGamma functions. The result is not always an integer.

`Power(a, <b>)`

`a^b`

Description

Raises *a* to the power of *b*.

Returns

The product of *a* multiplied by itself *b* times.

Arguments

a Can be a variable, number, or matrix.

b (Optional) Can be a variable or a number.

Notes

For `Power()`, the second argument (*b*) is optional, and the default value is 2. `Power(a)` returns a^2.

`Root(n, <r>)`

Description

Returns the *r*th root of *n*, where *r* defaults to 2 for square root.

`SbInv(z, gamma, delta, theta, sigma)`

Description

Returns a transformation of a standard normal variable to a double bounded Johnson variable.

`SbTrans(x, gamma, delta, theta, sigma)`

Description

Returns a transformation of a double bounded Johnson variable to a standard normal variable.

`Scheffe Cubic(x1, x2)`

Description

Returns `x1*x2*(x1-x2)`. This function supports notation for cubic mixture models.

`SHASHInv(z, gamma, delta, theta, sigma)`

Description

Returns a transformation of a standard normal variable to a sinh-arcsinh (SHASH) variable. The transformation is calculated as σ*`sinh((arcsinh(z)-`γ`)/`δ`)+`θ.

`SHASHTrans(x, gamma, delta, theta, sigma)`

Description

Returns a transformation of a sinh-arcsinh (SHASH) variable to a standard normal variable. The transformation is calculated as `sinh(`γ+δ*`arcsinh((x-`θ`)/`σ`))`.

`SlInv(z, gamma, delta, theta, sigma)`

Description

Returns a transformation of a standard normal variable to a Johnson Sl variable.

`SlTrans(x, gamma, delta, theta, sigma)`

Description

Returns a transformation of a Johnson Sl variable to a standard normal variable.

`Sqrt(n)`

Description

Returns the square root of *n*.

`Squash(expr)`

Description

An efficient computation of the function 1/ [1 + exp(*expr*)].

`Squish(expr)`

Description

Equivalent to `Squash(-expr)`, or $1/(1+e^{-\text{expr}})$.

`SuInv(z, gamma, delta, theta, sigma)`

Description

Returns a transformation of a standard normal variable to an unbounded Johnson variable.

`SuTrans(x, gamma, delta, theta, sigma)`

Description

Returns a transformation of an unbounded Johnson variable to a standard normal variable.

`Trigamma()`

Description

Returns the trigamma function evaluated at *n*. The trigamma function is the derivative of the digamma function.

Trigonometric Functions

JMP's trigonometric functions expect all angle arguments in radians.

`ArcCosH(x)`

Description

Inverse hyperbolic cosine.

Returns

The inverse hyperbolic cosine of *x*.

Argument

x Any number, numeric variable, or numeric expression.

`ArcCosine(x)`

`ArCos(x)`

Description

Inverse cosine.

Returns

The inverse cosine of *x*, an angle in radians.

Argument

x Any number, numeric variable, or numeric expression.

`ArcSine(x)`

`ArSin(x)`

Description

Inverse sine.

Returns

The inverse sine of *x*, an angle in radians.

Argument

x Any number, numeric variable, or numeric expression.

`ArcSinH(x)`

Description

Inverse hyperbolic sine.

Returns

The inverse hyperbolic sine of *x*.

Argument

x Any number, numeric variable, or numeric expression.

`ArcTangent(x1, <x2=1>)`

`ArcTan(x1 <x2=1>)`

`ATan(x1 <x2=1>)`

Description

Inverse tangent.

Returns

The inverse trigonometric tangent of *x1/x2*, where the result is in the range -Pi()/2, Pi()/2.

Argument

`x1` Any number, numeric variable, or numeric expression.

`x2=1` Specifies *atan2*.

`ArcTanH(x)`

Description

Inverse hyperbolic tangent.

Returns

The inverse hyperbolic tangent of *x*.

Argument

`x` Any number, numeric variable, or numeric expression.

`CosH(x)`

Description

Hyperbolic cosine.

Returns

The hyperbolic cosine of *x*.

Argument

`x` Any number, numeric variable, or numeric expression.

`Cosine(x)`

`Cos(x)`

Description

Cosine.

Returns

The cosine of *x*.

Argument

`x` Any number, numeric variable, or numeric expression. The angle in radians.

`Sine(expr)`

`Sin(expr)`

Description

Returns the sine.

`SinH(expr)`

Description

Returns the hyperbolic sine.

`Tangent(expr)`

`Tan(expr)`

Description

Returns the tangent of an argument given in radians.

`TanH(expr)`

Description

Returns the hyperbolic tangent of its argument.

Utility Functions

`Add(a, b, ...)`

`a+b+...`

Description

Adds the values of the listed arguments. No arguments are changed.

Returns

The sum.

Arguments

For `Add()`, a comma-separated list of variables, numbers, or matrices.

For `a+b`, any number of variables, numbers, or matrices.

Notes

- Any number of arguments is permitted. If no argument is specified, `Add()` returns 0.
- `Add()` returns missing if any arguments are missing. To ignore missing values, use `Sum()`.

See Also

the *Scripting Guide*

See "Sum(var1, var2, ...)"

Beep()

Description

Produces an alert sound.

Returns

Null.

BLOB MD5(blob)

Description

Converts the *blob* argument into a 16-byte blob.

Notes

The 16-byte blob is the MD5 checksum, or the hash, of the source blob.

BLOB Peek(blob, offset, length)

Description

Creates a new blob from a subrange of bytes of the *blob* argument.

Returns

A blob object.

Arguments

`blob` a binary large object.

`offset` An integer that specifies how many bytes into the blob to begin construction. The first byte is at offset 0, the second byte at offset 1.

`length` An integer that specifies how many bytes to copy into the new blob, starting at the offset.

Build Information()

Description

Returns the build date and time, whether it's a release or debug build, and the product name in a quoted comma-delimited string.

Caption({*h*, *v*}, "text", <Delayed(seconds)>, <Font(font)>, <Font Size(size)>, <Text Color("color")>, <Back Color("color")>, <Spoken(Boolean)

Description

Displays a caption window at the location described by *{h, v}* that displays *text*. The caption can be delayed before being displayed by *seconds*, or can be spoken. You can also specify the font type, size, and color and background color.

Returns

Null.

Arguments

`{h, v}` a list with two values. *h* is the horizontal displacement from the top left corner of the monitor in pixels. *v* is the vertical displacement from the top left corner in pixels.

`text` A quoted string or a reference to a string that is to be displayed in the caption.

`Delayed(seconds)` *seconds* is optional delay before displaying the caption. Setting this option causes this caption and all subsequent captions to be delayed by the specified number of seconds.

`Font(font)` Specify the font type.

`Font Size(size)` Specify the font size.

`Text Color("color")` Specify the color of text.

`Back Color("color")` Specify the background color.

`Spoken(Boolean)` Causes *text* to be spoken as well as displayed. The current setting (on or off) remains in effect until switched by another `Caption` statement that includes a `Spoken` setting.

Datafeed()

See "Open Datafeed()".

Debug Break()

When the JSL Debugger is open, this function stops a JSL script from executing at that point in the script. This function is useful for tracking in the debugger under user-specified conditions. If the JSL Debugger is not running, this function does not execute.

Decode64 BLOB(*string*)

Description

Decodes a printable quoted string of base 64 text into a blob.

Returns

A blob.

Arguments

`string` A quoted base-64 encoded string.

Example

```
Decode64 BLOB( "dGhlIHF1aWNrIGJyb3duIGZveA==" );
   Char To BLOB( "the quick brown fox", "ascii~hex" )
```

Decode64 Double(*string*)

Description

Creates a floating point number from a quoted base-64 encoded string.

Returns

A floating point number.

Arguments

`string` A quoted base-64 encoded string.

Disable JMP Live URL(url)

Description

Specifies the URLs that users cannot publish to. This preference must be added to jmpStartAdmin.jsl.

See the Common Tasks chapter in the *Scripting Guide* for details about the location of jmpStartAdmin.jsl.

If a URL appears in both the `Disable JMP Server()` and `Enable JMP Server()` lists, publishing to the URL is disabled.

Example

```
Disable JMP Live URL( "https://public.jmp.com" )
```

Divide(a, b)

Divide(x)

a/b

Description

Divides *a* by *b*. If only one argument is given (`divide(x)`), divides 1 by *x*.

Returns

The quotient of *a*/*b*; or the reciprocal of *x* (1/*x*) if only one argument is provided.

Arguments

`a, b, x` Can be a variable, number, or matrix.

Notes

If both arguments are matrices, it does matrix division.

Empty()

Description

Used in the formula editor for making empty boxes.

Returns

Missing.

Arguments

None.

Enable JMP Live URL(url)

Description

Specifies the URLs that users can publish to. This preference must be added to jmpStartAdmin.jsl.

See the Common Tasks chapter in the *Scripting Guide* for details about the location of jmpStartAdmin.jsl.

If a URL appears in both the `Enable JMP Server()` and `Disable JMP Server()` lists, publishing to the URL is disabled.

Example

```
Enable JMP Live URL( "https://public.jmp.com" )
```

Encode64 BLOB(x)

Description

Encodes a blob into a printable quoted string of base 64 text.

Returns

A quoted base-64 encoded string.

Example

```
Encode64 BLOB( Char To BLOB( "the quick brown fox" ) );
   "dGhlIHF1aWNrIGJyb3duIGZveA=="
```

Encode64 Double(n)

Description

Creates a quoted base-64 encoded string from a floating point number.

Returns

A quoted base-64 encoded string.

Arguments

`n` A floating point number.

Faure Quasi Random Sequence(nDim, nRow)

Description

Generates a sequence of spacefilling quasi random numbers using the Faure sequence.

Get Addin("id")

Description

Retrieves a registered add-in by *id*.

Returns

A scriptable object for the add-in. Returns empty if no add-in with the specified ID was found.

Argument

"id" The ID of an installed add-in.

Get Addins()

Returns

A list of all registered add-ins.

Get Addr Info("address", <port>)

Description

Converts a name to its numeric address.

Returns

A list of quoted strings. The first element is the command (Get Addr Info). The second is the results (for example, "ok" if the command was successful). The third is a list of strings of information. Included in that information is the address that corresponds to the name that was supplied.

Arguments

address A quoted string that specifies the name (for example, "www.sas.com").

port The port of the address.

Get Clipboard()

Description

Returns text from the computer's clipboard. If the content is not text, the result is null.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Copy Table Script( "Distribution" );
s = Get Clipboard();
nw = New Window( "Script", Script Box( s ) );
```

`Get Name Info("address", <port>)`

Description

Converts a numeric address to its name.

Returns

A list of quoted strings. The first element is the command (`GetNameInfo`). The second is the results (for example, "ok" if the command was successful). The third is a list of strings of information. Included in that information is the port name that corresponds to the address that was supplied.

Arguments

`address` A quoted string that specifies the numeric address (for example, `"149.173.5.120"`).

`port` The port of the address.

`Get Platform Preferences(<platform <(option, ...)> ... >)`

`Get Platform Preference(<platform <(option, ...)> ... >)`

Description

Returns the preferences for the specified platforms.

Returns

A list of platform preferences.

Argument

`platform` (Optional) Specifies the platform name. If not specified, all platform preferences are returned. You can specify one or more preferences for a platform.

`option` (Optional) Specifies the preference value. If not specified, all platform preference values are returned.

Notes

Table 2.3 describes the syntax for getting platform preferences.

Table 2.3 Get Platform Preferences() Syntax

Syntax	Description
`Get Platform Preferences( )`	Returns the current option values for all platform preferences.
`Get Platform Preferences( Platform )`	Returns the current option values for the specified platform preferences.

Table 2.3 Get Platform Preferences() Syntax *(Continued)*

Syntax	Description
`Get Platform Preferences( Platform( Option ) )`	Returns the current option values for the specified platform.
`Get Platform Preferences( <<Changed )`	Returns the current option values that have changed for all platforms.
`Get Platform Preferences( Platform( <<Changed ) )`	Returns the current option values that have changed for all platform preferences.
`Get Platform Preferences( Platform( Option ( <<Changed ) ) )`	Returns the current option values that have changed for the specified platform.

Examples

Suppose that the user modified several platform preferences through the JMP Platforms window or a script.

```
Platform Preferences(
  Distribution( Set Bin Width( 2 ), Horizontal Layout( 1 ) ),
  Model Dialog( Keep Dialog Open( 1 ) ),
  Graph Builder( Legend Position( "Bottom" ) )
);
```

To return all of the modified platform preferences, use `Get Platform Preferences( <<Changed )`:

```
Get Platform Preferences( <<Changed );
  Platform Preferences(
    Distribution( Horizontal Layout( 1 ), Set Bin Width( 2, <<On ) ),
    Graph Builder( Legend Position( "Bottom", <<On ) ),
    Model Dialog( Keep dialog open( 1 ) )
  );
```

Get Preferences(<preference_name>)

Get Preference(<preference_name>)

Description

Returns the settings for the specified preferences.

Returns

A list of preference settings.

Argument

`preference_name` (Optional) If no preference is specified, all preferences are returned. Otherwise, the settings for the specified preference are returned.

Notes

The preferences for the following areas are not accessible in JSL: Text Data Files, Windows Specific, Mac OS Settings, Fonts, Communications, Script Editor, and JMP Updates. For more information about getting platform preferences, see "Get Platform Preferences(<platform <(option, ...)> ... >)".

`Glue(expr1, expr2, ...)`

`expr1; expr2`

Description

Evaluates each argument in turn.

Returns

The result of the last argument evaluated.

Arguments

One or more valid JSL expressions.

Notes

A semicolon placed between expressions is the more commonly used form for `Glue()`.

`Gzip Compress(blob)`

Description

Compresses a blog of data into a gzip blob.

Example

```
Gzip Compress(
   Char To BLOB(
      "random data does not usually compress well and may get larger"
   )
);
   Char To BLOB(

      "~1F~8B~08~00~00~00~00~00~00~0A~0D~CA~C1~0D~00~21~08~04~C0V~B6~B5~CDA~FC
      ~80~5C~00c~EC^~E7=~C9)~E1~106~21~A1~85~19~8DU~8Bf~07_~F8~9FZ~85~ADfx~13~
      CE~83~A1~0Dc~0E~CD~0B~94*~16~1E=~00~00~00",
      "ascii~hex"
```

`Gzip Uncompress(blob)`

Description

Uncompresses a blob of gzip data into a blob.

Example

```
Gzip Uncompress( /*typically this data might come from Gzip Compress() but
  might also come from a .gz file using Load Text File() with the blob
  option*/
  Char To BLOB(

  "~1F~8B~08~00~00~00~00~00~00~0A~0D~CA~C1~0D~00~21~08~04~C0V~B6~B5~CDA~FC~8
  0~5C~00c~EC^~E7=~C9)~E1~106~21~A1~85~19~8DU~8Bf~07_~F8~9FZ~85~ADfx~13~CE~8
  3~A1~0Dc~0E~CD~0B~94*~16~1E=~00~00~00",
     "ascii~hex"
  )
);
  Char To BLOB(
     "random data does not usually compress well and may get larger",
     "ascii~hex"
  )
```

Host Is("argument")

Description

Determines whether the host environment is the specified OS.

Returns

True (1) if the current host environment matches the argument, false (0) otherwise.

Argument

`Argument Windows"` or "`Mac`" tests for the specified operating system.

Is Alt Key()

Description

Returns 1 if the Alt key is being pressed, or 0 otherwise.

Notes

On a macOS, `Is Alt Key()` tests for the Option key.

Is Command Key()

Description

Returns 1 if the Command key is being pressed, or 0 otherwise.

Is Context Key()

Description

Returns 1 if the Context key is being pressed, or 0 otherwise.

Is Control Key()

Description

Returns 1 if the Control key is being pressed, or 0 otherwise.

Notes

On a macOS, `Is Control Key()` tests for the Command key.

Is Option Key()

Description

Returns 1 if the Option key is being pressed, or 0 otherwise.

Is Shift Key()

Description

Returns 1 if the Shift key is being pressed, or 0 otherwise.

JMP Product Name()

Description

Returns either `"Standard"`, `"Pro"`, or `"Student"`, depending on which version of JMP is licensed.

JMP Version()

Description

Returns the version number of JMP that you are running.

Returns

release.revision<.fix>

Arguments

none

Load DLL("path" <,AutoDeclare(Boolean | Quiet | Verbose)>)

Load DLL("path" <, Quiet | Verbose>)

Description

Loads the DLL in the specified *path*.

Arguments

`path` A pathname that specifies where to load the DLL.

`AutoDeclare(Boolean | Quiet | Verbose)` Optional argument. `AutoDeclare(1)` and `AutoDeclare(Verbose)` write verbose messages to the log. `AutoDeclare(Quiet)`

turns off log window messages. If you omit this option, verbose messages are written to the log.

`Quiet | Verbose` Optional argument. When you use `Declare Function()`, this option turns off log window messaging (`Quiet`) or turns on log window messaging (`Verbose`).

See Also

Once a DLL is loaded, you send the DLL object messages to interact with it. See "Dynamic Link Libraries (DLLs)" for more information about these messages.

`Mail("address"|"addresses", "subject", "message", <"attachment filepath" | {"attachment 1 filepath", "attachment 2 filepath", ...}>)`

Description

(Windows) Sends e-mail (using MAPI) to the *`address`* with the specified *`subject`* and *`message`* texts. Sends one or more attachments specified by the optional *`attachment`* argument. The attachment argument can evaluate to a quoted string or list of strings.

(macOS) Creates an e-mail in the user's Mail application. The user must click **Send** in the e-mail. In Microsoft Outlook, you must manually add attachments to the e-mail.

Examples

To send an email with multiple attachments on Windows:

```
Mail(
   "yourname@company.com",
   "New data and script",
   "Today's updated data table and script are attached.",
   {"$DOCUMENTS/wd.jsl", "$DOCUMENTS/survey.jmp"}
);
```

or:

```
list = {"$DOCUMENTS/wd.jsl", "$DOCUMENTS/survey.jmp"};
Mail(
   "yourname@company.com",
   "New data and script",
   "Today's updated data table and script are attached.",
   list
);
```

To send an email to multiple recipients:

```
Mail(
   {"hername@company.com", "hisname@company.com"},
   "Database updates",
   "Today's sales database contains the numbers from last month."
);
```

Notes

On macOS, `Mail()` works on Yosemite and later operating systems.

Main Menu(*string*, <*string*>)

Description

Execute the command found on JMP's menu named by the quoted *string*.

Arguments

`string` The internal path name as shown in the menu editor for items. For example, "NEW" is the internal name for the New subcommand in the File menu.

`string` (Optional) The name of the window to send the command to.

Examples

`Main Menu()` accepts either a full path or a partial path. If a partial name is used, and there are other menu items with the same name, the first menu item found is executed. JMP searches the top-level menu (File, Tables, DOE, and so on) first for the partial name and then searches inside each of those menus in order.

```
Main Menu( "File:New:Data Table" ); // full path
Main Menu( "Data Table" ); // partial path
```

Minus(a)

-a

Description

Reverses the sign of *a*.

Returns

-*a* if *a* is positive (`a=3; -a=-3; Minus(a)=-3`).

a if *a* is negative (`a=-3; -a=3; Minus(a)=3`).

0 if *a* is 0 (`a=0; -a=0; Minus(a)=0`).

Missing if *a* is missing.

Argument

`a` Can be variable or a number. A variable must contain a number or a matrix.

Multiple File Import(arguments)

Description

Imports one or more files into a data table. You can create this JSL by selecting Save Script to Script Window from the Multiple File Import window.

Returns

Creates a Multiple File Import Object. The object accepts messages to set a folder, filter files, and specify import options.

Arguments

`<<Set Folder` Specifies the folder that contains the files you want to import.

<<Set Name Filter (Optional) Specifies the file name or extension of the files. The name filter uses * to represent zero-or-more characters and ? to represent exactly one character. * and ? also match a period. The default setting is *.*, or all files. Because you can specify files using a semicolon-delimited list of filters, file names that include semicolons or | must be specified using ? or *.

<<Set Name Enable(Boolean) Enables the name filter. The setting is off by default.

<<Set Size Filter (Optional) Filters the file list by file size. Specify the sizes by kB (kilobytes, or 1000 bytes) in a list. The default values are based on the size range of the files in the file list.

<<Set Size Enable(Boolean) (Optional) Enables the size filter. The default setting is off.

<<Set Date Filter (Optional) Filters the file list by date and time. Specify the date and time in a list in seconds. The default values are based on the date and time range of the files in the file list.

<<Set Date Enable(Boolean) (Optional) Enables the date filter. The default setting is off.

<<Set Add File Name Column(Boolean) (Optional) Includes a column that contains the imported file name. The default setting is off.

<<Set Add File Size Column(Boolean) (Optional) Includes a column that contains the size of the imported file. The default setting is off.

<<Set Add File Date(Boolean) (Optional) Includes a column that contains the time and date stamp of the imported file. The default setting is off.

<<Set Import Mode(Row Per File|Row Per Line|CSV Data) (Optional) Specifies the format of the file that is imported: whole file on one row, one line on one row, and CSV. CSV Data is the default setting.

<<Set Charset(Best Guess|utf-8|utf-16|us~ascii|windows-1252|x-mac-roman|x-mac-japanese|shift-jis|euc-jp|utf-16be|gb2312) (Optional) The character set in the imported file. The character set specified in the General preferences (Open Text File Charset) is set by default.

<<Set Stack Mode(Stack Similar|TablePerFile) (Optional) Specifies how the files are combined. Stack Similar is the default setting. (When JMP detects that the files have the same columns, the files are concatenated into a single data table.)

<<Set CSV Has Headers(Boolean) (Optional) Specifies whether the CSV file contains a header row. The setting is on by default.

<<Set CSV Allow Numeric(Boolean) (Optional) Sets the data type to numeric. The setting is on by default.

<<Set CSV First Header Line(*n*) (Optional) Specifies the header row number. 1 is the default setting.

<<Set CSV Number of Header Lines(*n*) (Optional) Specifies the number of header rows. 1 is the default setting.

<<Set CSV First Data Line(n) (Optional) Specifies the first line that contains data. 2 is the default setting.

<<Set CSV EOF Comma(Boolean) (Optional) Specifies a comma delimiter. The setting is on by default.

<<Set CSV EOF Tab(Boolean) (Optional) Specifies a tab delimiter.

<<Set CSV EOF Space(Boolean) (Optional) Specifies a space delimiter.

<<Set CSV EOF Spaces(Boolean) (Optional) Specifies spaces as the delimiter.

<<Set CSV EOF Other("") (Optional) Specifies a custom delimiter.

<<Set CSV EOF CRLF(Boolean) (Optional) Specifies carriage return and line feed end-of-line characters. The setting is on by default.

<<Set CSV EOF CR(Boolean) (Optional) Specifies a carriage return end-of-line character. The setting is on by default.

<<Set CSV EOF LF(Boolean) (Optional) Specifies a line feed end-of-line character.

<<Set CSV Semicolon(Boolean) (Optional) Specifies a semicolon end-of-line character. The setting is off by default.

<<Set CSV EOL Other("") (Optional) Specifies a custom end-of-line character.

<<Set CSV Quote("") (Optional) Specifies the character used as a quote. The default setting is \!", a double quotation mark.

<<Set CSV Escape("") (Optional) Specifies the escape sequence such as a backlash instead of doubling the quotation mark.

<<Import Data Imports the data.

Example

```
mfi = Multiple File Import(
  <<Set Folder( "$SAMPLE_IMPORT_DATA" ),
  <<Set Name Filter( "UN*.csv" ), // import files with this name
  <<Set Name Enable( 1 ) // display the file name in a column
)
<<Import Data();
```

Multiply(a, b, ...)

a*b*...

Description

Multiplies all values. No arguments are changed.

Returns

The product.

Arguments

Any number of variables, numbers, or matrices.

Notes

Any number of arguments is permitted. If no argument is specified, Multiply() returns 1.

Name(*string*)

Description

A name is something to call an item.

- If the name begins with an alphabetic character or underscore, and continues with alphanumeric characters, whitespace, Unicode mathematical symbols and certain punctuation (apostrophes, percentage signs, periods, backslashes, and underscores, then you can use the name directly in scripts.
- You can use names that do not follow these rules with the Name() keyword.

Examples

("name"n) is preferred. Name(*string*) is deprecated.

```
"my-var-name"n = "hello";
Length( "my-var-name"n )
   5
```

New OAuth2 Token(user(yourgoogleaccount@gmail.com), client ID(*string*), client secret(*string*), refresh token(*string*), token URL(*string*))

Description

Creates an OAuth2 token for securely accessing data across different web APIs.

Arguments (Required)

`user` The user name, email, or personal identifier for the account being accessed.

`client ID` The public identifier that acts as an API key.

`client secret` The private identifier that corresponds to Client ID.

`refresh token` A token used to get access tokens.

`token URL` The URL that access tokens are received from. Unique to every service and accessible through their API or OAuth page.

Arguments (Authorization Code Grant)

`redirect URL( "https://app.getpostman.com/oauth2/callback" )` The URL that an access code is sent back to. Unless your company or the service provides one, we recommended that you create a free Postman account and use this redirect.

`client secret( "1aB893cdDeFf2D" )` The private identifier that corresponds to Client ID.

`request auth( ... )` Extra parameters indicating that you'll use an Authorization Code flow. Requires `Auth URL()`. Some services require scope. You can add custom fields with Fields.

`scope( "spreadsheets email docs" )` A space-separated list of scopes. Unique to every service, and accessible through their API or OAuth page. Only usable in `Request Auth()`.

`auth URL( "https://www.example.com/oauth2/v1/authorize" )` The URL for requesting authorization. Unique to every service, and accessible through their API or OAuth page. Only usable in `Request Auth()`.

Arguments (Implicit Grant)

`redirect URL( "https://app.getpostman.com/oauth2/callback" )` The URL that an access code is sent back to. Unless your company or the service provides one, it's recommended to create a free Postman account and use this redirect.

Arguments (Resource Owner)

`password( "wordpass123" )` The password that corresponds to the username.

`client secret( "1aB893cdDeFf2D" )` The private identifier that corresponds to Client ID.

Arguments (Custom Data)

`fields( fields )` Custom fields that are equivalent to HTTP Request's Form(fields(fields)). Can be specified both in `New OAuth2 Token()` and in `Request Auth()`. Only necessary if the service requires information that is not defined in the OAuth 2.0 standard.

`headers( headers )` Custom headers that are equivalent to `HTTP Request's Headers( headers )`. Can only be specified in `New OAuth2 Token()`. Only necessary if the service requires information not defined in the OAuth 2.0 standard.

Example

```
token = New OAuth 2 Token (
  User( "yourgoogleaccount@gmail.com" ),
  Refresh Token( "1a2b3c4e5F" ),
  Token URL( "https://www.example.com/oauth2/token" ),
  Client ID( "12ab" ),
  Client Secret( "3456dEfG" )
);
```

See Also

See your API documentation for more information about how to get values such as the client secret and token URL.

Open Datafeed()

Datafeed()

Description

Creates a Datafeed object and window.

Returns

A reference to the Datafeed object.

Arguments

No arguments are required. You usually set up the basic operation of the data feed within the `Open Datafeed()` command, however.

`Open Help("Help"|"Statistics Index"|"Scripting Index", ...)`

Description

Opens the specified help window.

`Parse XML(`*`string`*`, On Element("tagname", Start Tag(expr), End Tag(expr), Text))`

Description

Parses an XML expression using the `On Element` expressions for the specified XML tags.

Example

```
XMLData = "
<Book name='Foods'>All you want to know
       <Chapter num='1'>Fruit
             <kind>Apple</kind>
             <kind>Cherry</kind>
             <ps>I love dessert!</ps>
       </Chapter>
       <Chapter num='2'>Bread
             <kind>Wheat</kind>
             <kind>Corn</kind>
             <ps>I love sandwiches!</ps>
       </Chapter>
       <Chapter num='3'>Veggy
             <kind>Squash</kind>
             <kind>Cabbage</kind>
             <ps>I love anything else!</ps>
       </Chapter>
        and more.
</Book> ";

// variables are initialized so text can be concatenated
title = "";
subtitle = "";
chap = "";
chapnum = "";
ps = "";

Parse XML( XMLData,
      On Element( "Book",
            // capture the name attribute during the start of the Book
            Start Tag( title = XML Attr( "name" ) ),
           /* this book has split the subtitle and needs to join the text;
              the joined text will be used by endTag.
              Text(...) supplies the JSL.*/
```

```
                Text( subtitle = subtitle || " -- " || Trim( XML Text() ) ),
                  /* after endTag processes the variables, set them back
                  to their initial state, just in case there is a second book
                  to process in the same XML. */
                endTag( Write( "\!n", title, " ", subtitle ); title = ""; subtitle
        = ""; )
            ),
            On Element( "Chapter",
                  // capture the chapter number during the start of the Chapter
                  Start Tag( chapnum = XML Attr( "num" ) ),
                /* the chapter text is joined together, newlines
                  and extra space is trimmed, and a single space is used to
                  separate the separated texts. The <kind> tag is ignored by
                  this ParseXML specification. The <kind> text is processed
                  by this Text(...) because it wasn't consumed by any other
                  On Element. */
                  Text( chap = chap || Trim( XML Text() ) || " " ),
                /* after endTag processes the variables, set them back to
                  their initial state, because there is another chapter
                  that needs to start with a clean slate.*/
                endTag( Write( "\!n", chapnum, " ", chap, " ps: ", ps ); chapnum =
        ""; chap = ""; ps = ""; )
            ),
            On Element( "ps", End Tag( ps = XML Text() ) )
    );
      1 Fruit Apple Cherry  ps: I love dessert!
      2 Bread Wheat Corn  ps: I love sandwiches!
      3 Veggy Squash Cabbage  ps: I love anything else!
      Foods  -- All you want to know -- and more.
```

`Platform Preferences(platform(option(value)), ...)`

`Platform Preference(platform(option(value)), ...)`

`Set Platform Preferences(platform(option(value)), ...)`

`Set Platform Preference(platform(option(value)), ...)`

Description

Sets and resets values for platform options and turns the options on and off.

Arguments

`platform` Specifies the platform of the preference.

`option` Specifies the preference name.

`value` Specifies the preference value.

Notes

Table 2.4 describes the syntax for setting platform preferences.

Table 2.4 Platform Preferences() Syntax

Syntax	Description

`Platform Preferences( <<Default )`
`Platform Preferences( <<Factory Default )`
`Platform Preferences( Default )`
	Resets all platform preferences to their default values.

`Platform Preferences( Platform( <<Default ) )`
`Platform Preferences( Platform( <<Factory Default ) )`
`Platform Preferences( Platform( Default ) )`
	Resets the specified platform preferences to their default values.

`Platform Preferences( Platform( option ( <<Default ) ) )` `Platform Preferences( Platform( option ( <<Factory Default ) ) )`	Resets the specified platform option to its default value.
`Platform Preferences( Platform( option( value, <<On ) ) )`	Sets the value of the specified platform option and turns it on.
`Platform Preferences( Platform( option( value, <<Off ) ) )`	Sets the value of the specified platform option and turns it off.

Example

The following expression selects (or turns on) Set Bin Width in the Distribution platform preferences and sets the value to 2:

```
Platform Preferences( Distribution( Set Bin Width( 2 ) ) );
```

The following expression changes the Set Bin Width value and turns the option off:

```
Platform Preferences( Distribution( Set Bin Width( 2, <<Off ) ) );
```

The following expression resets the default Set Bin Width value and deselects the preference:

```
Platform Preferences( Distribution( Set Bin Width( <<Default ) ) );
```

Polytope Uniform Random(samples, A, b, L, U, neq, nle, nge, <nwarm=200>, <nstride=25>)

Description

Generates random uniform points over a convex polytope.

Arguments

`Samples` The number of random points to be generated.

`A` The constraint coefficient matrix.

`B` The right hand side values of constraints.

`L, U` The lower and upper bounds for the variables.

`neq` The number of equality constraints.

`nle` The number of less than or equal inequalities.

`nge` The number of greater than or equal inequalities.

`nwarm` (Optional) The number of warm-up repetitions before points are written to the output matrix.

`nstride` (Optional) The number of repetitions between each point that is written to the output matrix.

Notes

The constraints must be listed as equalities first, less than or equal inequalities next, and greater than or equal inequalities last.

```
Preferences(pref1(value1), ...)
Preference(pref1(value1), ...)
Pref(pref1(value1), ...)
Prefs(pref1(value1), ...)
Set Preferences(pref1(value1), ...)
Set Preference(pref1(value1), ...)
```

Description

Sets preferences for JMP.

Arguments

`Add Files Opened by Scripts to the Recent Files List(Boolean)` Determines whether a file that is opened by a script is added to the Home Window's Recent Files list.

`Analysis Destination(window)` Specifies where to route new analyses.

`Annotation Font("font", size, "style")` Font choice for annotations in reports.

`Axis Font("font", size, "style")` Font choice for axis labels.

`Axis Title Font("font", size, "style")` Font choice for axis titles.

`Background Color( {R, G, B} | <color> )` Sets the background color for windows.

`Calculator Boxing(Boolean)` Turns on boxing to show hierarchy of expressions.

`Conditional Formatting Rules` Creates rules for conditionally formatting text in reports. See "Examples".

`Data Table Font("font", size, "style")` Font choice for data tables.

`Data Table Title on Output(Boolean)` Titles reports with name of data table.

`Date Title on Output(Boolean)` Titles reports with current date.

`Evaluate OnOpen Scripts("always"|"never"|"prompt")` Determines whether an `On Open` table script is run after the user opens the data table. By default, the user is prompted. Their choice is remembered each time they open the data table in the current JMP session. Scripts that execute other programs are never run.

`Excel Has Labels(Boolean)` When on, forces JMP to interpret the first row of data as column headings.

`Excel Selection(Boolean)` When on, the user is prompted for which non-blank Excel worksheets should be imported from an Excel workbook.

`File Location Settings(<Directory Type>("<path>"<,"initial directory">))` Valid directory types are:

`Data Files Directory` Sets the default location for data files.

`Help Files Directory` Sets the default location for help files.

`Installation Directory` By default, this location is set to the JMP installation folder on Windows:

"C:/Program Files/SAS/JMP/16" or "C:/Program Files/SAS/JMPPro/16"

`License File Path` Sets the default location for JMP license file.

`Preferences File Directory` Sets the default location for the preferences settings file.

`Save As Directory` Sets the default location for Save As file operations.

`Foreground Color(color)` Sets the foreground color for windows.

`Formula Font("font", size, "style")` Font choice for the formula editor.

`Graph Background Color(color)` Sets the color for the background area inside the graph frame.

`Graph Marker Size(size)` Default size for drawing markers.

`Heading Font("font", size, "style")` Font choice for table column headings in reports.

`Initial JMP Starter Window(Boolean)` Specifies whether the JMP Starter window is shown at launch.

`Initial Splash Window(Boolean)` Enables you to show or suppress the initial splash screen.

`Maximum JMP Call Depth(size)` Sets the default for the maximum call depth (or stack size) for JMP in which JSL built-in functions, user-defined functions, or `Recurse()` function calls can be made. By default, the maximum call depth is set to 256KB.

Each thread that JMP creates has a 2MB stack by default. Increasing the maximum call depth can cause a physical runtime stack overflow, so incrementally increase this preference in small amounts until you find the best value that works for your JSL script.

`Marker Font("font", size, "style")` Font choice for markers used in plots.

`Monospaced Font("font", size, "style")` Font choice for monospaced text.

`ODBC Suppress Internal Quoting(Boolean)` Prevents internal quoting in SQL statements that contain table and variable names with mixed case and spaces.

`Outline Connecting Lines(Boolean)` Draws lines between titles for same-level outline nodes.

`Print Settings(option(value), ... )` Changes print options on the Page Setup window:

`Margins( <n>, <n>, <n>, <n>)` sets the left, top, right, and bottom margins. Margins are in inches.

`Margins(<n>)` sets all margins to the same value in inches.

`Orientation("portrait" | "landscape")` changes the page's print orientation.

`Headers(<"char">, <"char">, <"char">)` specifies text that appears in the left, middle, and right header.

`Headers(<"char">)` specifies the only text in the header.

`Footers(<"char">, <"char">, <"char">)` specifies text that appears in the left, middle, and right footer.

`Footers(<"char">)` specifies the only text in the footer.

`Scale(<n>)` decreases or increases the percentage at which the content prints.

`Show Explanations(Boolean)` Some analyses have optional text that explains the output.

`Show Menu Tips(Boolean)` Turns menu tips on or off.

`Show Status Bar(Boolean)` Turns display of the status bar on or off.

`Small Font("font", size, "style")` Font choice for small text.

`Text Font("font", size, "style")` Font choice for general text in reports.

`Thin Postscript Lines(Boolean)` macOS only. Specifies that line widths drawn to a Postscript printer be narrower than otherwise.

`Title Font("font", size, "style")` Font choice for titles. Arguments are name of font (for example, `"Times"`), size in points, and style (`"bold"`, `"plain"`, `"underline"`, `"italic"`).

`Use Triple-S Labels as Headings(Boolean)` When on, this argument forces JMP to interpret label names as column headings. Example: `Pref(Name("Use Triple-S Labels as Headings")(0) )`; turns off the preference.

Examples

The following expressions reset all preferences to their default values.

```
Preferences( "Default" );
Preferences( "Factory Default" );
```

The following script creates conditions for formatting text in reports.

```
Preferences(
    Conditional Formatting Rules(
```

```
            RuleSet(
               RuleName( "Warning" ),

               // if the value is not equal to 0, format the text as 80% gray
               NotEqualTo( Value( 0 ), Format( TextAlpha( 0.8 ) ) )
             )
        )
    );
```

Notes

The preferences for the following areas are not accessible in JSL: Text Data Files, Windows Specific, Mac OS Settings, Fonts, Communications, Script Editor, and JMP Updates.

See Also

"Platform Preferences(platform(option(value)), ...)"`Register Addin("unique_id", "home_folder", <named_arguments>)`

Description

Register a JMP Add-In and load the add-in if it registers successfully.

Returns

If successful, returns a scriptable object representing the registered add-in. If unsuccessful, returns Empty.

Arguments

`unique_id` A quoted string that contains the unique identifier for the add-in. The string can contain up to 64 characters. The string must begin with a letter and contain only letters, numbers, periods, and underscores. Reverse-DNS names are recommended to increase the likelihood of uniqueness.

`home_folder` A quoted string that contains the filepath for the folder containing the add-in files. The filepath must conform to the valid pathname requirement for the host operating system.

`DisplayName( "name" )` An optional, quoted string that contains a name that can be displayed in the JMP user interface wherever add-in names are displayed, instead of the unique ID.

`JMPVersion("version")` An optional string that contains a specific version of JMP. The default value is "All", which enables the add-in to be loaded and run in any version of JMP that supports add-ins. "Current" restricts the use of the add-in to only the current version. Any quoted version number (for example, "7" or "9") restricts the add-in to a single specific version of JMP.

`LoadsAtStartup(Boolean)` An optional Boolean. The default value is True (1), which causes the add-in to be loaded when JMP is started. If the value is False (0), the add-in is not loaded automatically.

`LoadNow(Boolean)` Loads the add-in immediately.

Notes

If a file named addin.def is found in the specified home folder, values from that file are used for any optional arguments that are not included in the `Register Addin()` function.

Example

In the following example, the first argument is the unique identifier. The second argument identifies where the add-in is installed. The third argument is the name that appears where add-in names are displayed (for example, the **View > Add-Ins** menu on Windows).

```
Register Addin("com.company.lee.dan.MyAddIn","$DOCUMENTS/myaddin",
  displayname( "Calculator Addin" ));
```

The second argument becomes the `$ADDIN_HOME` path variable definition. When you refer to the add-in scripts, be sure to include a trailing slash after the path variable.

```
Include("$ADDIN_HOME(com.jmp.jperk.texttocols)/texttocols.jsl");
```

Revert Menu()

Description

Resets your JMP menus to factory defaults.

Run Program(Executable("path/filename.exe"), Options({"/a", "/b", "..."}), Read Function(expression), Write Function(expression), Parameter(expression))

Description

Runs the external program specified by the `Executable` argument, with the command line arguments specified by the `Options` argument.

Results

Returns either a quoted string, a blob, or a `Run Program` object as controlled by the `Read Function` argument.

Arguments

`Executable` The path to the executable. On macOS, type the full path to the executable.

`Options` Command line arguments for the executable.

`Read Function` If `Read Function( "text" )` is specified, a quoted string is returned. If `Read Function( "blob" )` is specified, a blob is returned. The script waits until the external program closes its stdout. `Run Program` then returns all data that the external program has written to its stdout as a quoted string or a blob.

If `Read Function` is not specified, a `Run Program` object is returned.

`Write Function` Optional argument that accepts a function as its value; it does not accept `"text"` or `"blob"`.

`Parameter` Optional argument to read and write the expression in `Read Function`.

Notes:

- Use global variables when `Run Program()` is inside a function.
- The `Run Program` object, which is returned if `Read Function` is not specified, accepts the following messages to read data from the external program's stdout:
  - `<<Read`: reads any available data as a quoted string. If no data is available, an empty string is returned.
  - `<<Can Read`: returns `true` if there is data available to be read.
  - `<<Is ReadEOF`: returns `true` when the external program has completed and all its data has been read.

  You can use these messages to poll for data and process the data as it is produced by the external program.
- A `Run Program` object accepts the following messages to write data to the external program's `stdin`:
  - `<<Write( "text" )`: sends data to the external program's stdin.
  - `<<Can Write`: returns `true` if the external program will accept data immediately; otherwise, calling `<<Write` causes your script to block.
  - `<<WriteEOF`: signals to the external program that you are done sending data to it.
- Instead of sending messages to the returned `Run Program` object, you can specify the `Read Function` argument as an inline function. `RP` is the `Run Program` object.

```
RP = Run Program(
   Executable( ... ),
   Read Function(
      Function( {RP},
          <your code here>
         RP << Read
      )
   )
);
```

The `Parameter(optParm)` argument is optional in `Read Function`. If specified, the functions defined for `Read Function` and `Write Function` can receive a second argument, which is the value of `optParm`.

Examples

The following script is an example of the `Write Function` argument. `RP` is the `Run Program` object. In this context, it accepts the `<<Write` and `<<WriteEOF` messages.

```
RP = Run Program(
   Executable( ... ),
   Write Function(
      Function( {RP},
         <your code here>
```

```
            RP << Write( "Program finished." )
        )
    )
);
```

The following script shows an example of `Parameter(optParm)` argument:

```
RP = Run Program(
   Executable( ... ),
   Parameter( x ),
   Read Function( Function( {RP, optParm},... ) )
);
```

Within the `Read Function`, `optParm` contains the value of *x*. Do not attempt to access the `optParm` argument in your function if you have not specified a `Parameter` argument.

Schedule(n, script)

Description

Queues an event to run the *script* after *n* seconds.

Set Clipboard(*string*)

Description

Evaluates the quoted *string* argument looking for a character result, and then places the string on the clipboard.

Example

```
Set Clipboard( "copy me" );
```

SetJVMOption(Version("<version number>"))

Description

Sets the Java Runtime Environment (JRE) version that you want JMP to use (rather than the version installed with JMP). This script must be run before JMP connects to the JRE.

Argument

`version` (Windows only) In the Windows registry, there are two requirements for the JavaSoft/Java Runtime Environment key: the key must include a quoted string called "RuntimeLib" that points to a valid jvm.dll. And the Java Runtime Environment key must include a key named after the quoted JVM version number.

Set Platform Preference()

Set Platform Preferences()

See "Platform Preferences(platform(option(value)), ...)".

`Set Preference()`

`Set Preferences()`

See "Preferences(pref1(value1), ...)".

`Set Toolbar Visibility( "toolbar name" | default | all, window type | all, "true" | "false" )`

Description

On Windows, shows or hides a toolbar based on the window type or for all windows.

Arguments

`toolbar name | default | all` The internal name of the toolbar (see the **View > Toolbars** list in JMP), the default toolbar for the specified window type, or all toolbars. Include quotes around `"toolbar name"`.

`window type | all` Data table, script, report, journal, or all windows.

`true | false` A quoted string that shows or hides the toolbar.

`Shortest Edit Script( A, B )`

`Shortest Edit Script( strings( A, B, <matrix( 0|1 )>, <limit( number )> ) )`

`Shortest Edit Script( lines( A, B, <matrix( 0|1 )>, <limit( number )>, <separators( characters )>, <ignore( characters )|ignore white space( )> ) )`

`Shortest Edit Script( sequences( nA, nB, Function( {iA, iB}, adata[iA] == bdata[ib] ) ) )`

Description

Compares two quoted strings, lines, or sequences.

Returns

Returns a list or a matrix of edit commands. The simplest form returns a list. quoted `strings` and `lines` return a matrix (if set to 1) or a list. `sequences` returns a matrix.

There are three possible commands: common data in both quoted strings, delete data from the first string, and keep data from the second string.

Optional Arguments

`matrix` Indicates whether the returned value is a matrix.

`limit` Stops the evaluation when the edit list exceeds the specified number of inserted or deleted items. Two quoted random strings have a lot of common characters in a lot of distinct sections. The function runs for a long time trying to find a best match. `limit` stops the function sooner.

Optional Lines Arguments

`matrix` Indicates whether the returned value is a matrix.

`limit` Stops the evaluation when the edit list exceeds the specified number of inserted or deleted items. Two quoted random strings have a lot of common characters in a lot of distinct sections. The function runs for a long time trying to find a best match. `limit` stops the function sooner.

`separators` A character that separates words.

`ignore` Ignores the specified spaces or characters in a line.

`ignore white space` Ignores white space in a line.

Optional Sequences Argument

`Function` A user-defined function.

Examples

The following example compares two quoted strings with three common sequences of characters between them.

```
Shortest Edit Script( "abcdef", "abdezgh" );
```
```
{{"Common", "ab"}, {"Remove", "c"}, {"Common", "de"}, {"Insert", "zgh"},
   {"Remove", "f"}}
```

The following example examines each line in quoted string aa and bb:

```
aa = "this is
a test of
shortest
edit script
lines with several words";

bb = "this is
a test 2 of
shortest
edit ?., script
lineswithseveral words";

Shortest Edit Script( lines( aa, bb, separators( "\!n" ),

  // quote and newline separators
  ignore( "?., " ) ) ); // ignore these characters and spaces
{{"Common", "this is // lines in aa and bb contain "this is"
"}, {"Remove", "a test of // only on line 2 of aa
"}, {"Insert", "a test 2 of // only on line 2 of bb
"},
{"Common", "shortest
edit script
lines with several words"}}
// lines in aa and bb contain "shortest", "edit script", and "lines with
   several words"
```

Show Addin Builder Dialog()

Description

Opens a window in which you can make custom add-ins.

Show Addins Dialog()

Description

Opens the Add-In Status window (**View > Add-Ins**).

Arguments

None.

Show Commands()

Description

Lists scriptable objects and operators. Arguments are All, DisplayBoxes, Scriptables, Scriptable Objects, StatTerms, Translations.

Show Preferences(<"all">)

Description

Shows current preferences. If no argument is specified, preferences that have been changed are shown. If "all" is given as the argument, all preferences are shown.

Show Properties(object)

Description

Shows the messages that the given *object* can interpret, along with some basic syntax information.

Sobol Quasi Random Sequence(nDim, nRow)

Description

Generates a sequence of space-filling quasi random numbers using the Sobol sequence in up to 4000 dimensions.

Socket(<STREAM | DGRAM>)

Description

Creates a socket.

Returns

The socket that was created.

Arguments

STREAM | DGRAM Optional argument to specify whether the socket is a stream or datagram socket. If no argument is supplied, a stream socket is created.

`Speak(text, <wait(Boolean)>)`

Description

Calls system's speech facilities to read aloud the text. If Wait is turned on, script execution pauses until speaking is done.

`Status Msg("message")`

Description

Writes the quoted message string to the status bar.

`Subtract(a, b)`

`a-b-...`

Description

Subtracts the values of the listed arguments, left to right. No arguments are changed.

Returns

The difference.

Arguments

Two or more variables, numbers, or matrices.

Notes

Two or more arguments are permitted.

`Unregister Addin("unique_id")`

Description

Unregisters (removes) a previously registered add-in.

Argument

`unique_id` A quoted string that contains the unique identifier for the add-in to unregister.

`Web(`*`string`*`, <JMP Window>)`

Description

Opens the URL stored in the quoted *`string`* in the default web browser.

The `http://` prefix in the URL is optional.

Examples

```
url = "www.jmp.com"; // open the URL in the default web browser
Web( url );

Web( "www.jmp.com" ); // open the URL in the default web browser

Web( "www.jmp.com", JMP Window ); // open the URL in the JMP browser window
```

XML Attr("attr name")

Description

Extracts the quoted string value of an XML argument in the context of evaluating a `Parse XML` command.

XML Decode("xml")

Description

Decodes symbols in XML to ordinary text. For example, `&` becomes &, and `<` becomes <.

Argument

`xml` A quoted string containing XML.

XML Encode("text")

Description

Prepares text for embedding in XML. For example, & becomes `&`, and < becomes `<`.

Argument

`xml` A quoted string containing plain text.

XML Text()

Description

Extracts the quoted string text of the body of an XML tag in the context of evaluating a `Parse XML` command.

Chapter 3

JSL Messages

Summary of Messages for Objects and Display Boxes

This topic provides abbreviated descriptions for many of JMP's general object messages. For complete information about object messages, see the JMP Scripting Index. In JMP, select **Help > Scripting Index**.

For information about platform messages, see *Scripting Guide*.

Contents

Alpha Shape . . . 380
Associative Arrays . . . 380
Classes . . . 381
Data Tables . . . 383
Columns . . . 412
Rows . . . 420
Data Filter . . . 421
Data Feed (Windows Only) . . . 424
Display Boxes . . . 427
All Display Boxes . . . 427
Axis Boxes . . . 437
Border Boxes . . . 442
Data Browser Boxes . . . 443
Data Filter Source Boxes . . . 443
Frame Boxes . . . 443
Display 3D Boxes . . . 445
Excerpt Boxes . . . 445
Filter Col Selector . . . 446
Global Boxes . . . 446
Hier Boxes . . . 446
Matrix Boxes . . . 446
Nom Axis Boxes . . . 447
Number Col Boxes . . . 447
Number Col Edit Boxes . . . 450
Number Edit Box . . . 451
Outline Boxes . . . 451
Panel Boxes . . . 452
Plot Col Boxes . . . 452
Slider Boxes and Range Slider Boxes . . . 453
String Col Boxes . . . 454
Tab Boxes . . . 455
Table Boxes . . . 456
Text Boxes . . . 459
Tree Node and Tree Box . . . 461
Triangulation . . . 463
Windows . . . 464
Dynamic Link Libraries (DLLs) . . . 467
HTML 5 . . . 469
Web Report . . . 469

Images 470
JMP Applications 473
JMP App 473
JMP App Module 475
JMP App Module Instance 475
MATLAB 476
Namespaces 479
Platforms 481
Bubble Plot 487
DOE 487
Partition 488
Response Screening 488
Tabulate 489
Python Integration Messages 490
R Integration Messages 493
SAS Integration Messages 496
Metadata Server Objects 496
SAS Server Objects 497
Stored Processes 506
SAS Results 512
Schedule 514
Segments 515
Sockets 515
SQL 519
Other Objects 523
Zip Archives 523
Journals 523

Alpha Shape

For the following messages, ashape stands for an alpha shape or a reference to one.

---

ashape <<Get Alpha

Returns the current alpha value.

---

ashape <<Set Alpha(*alpha*)

Sets the current *alpha* value and recomputes the triangulation.

---

ashape <<Get Tri Alpha

Returns the alpha values for each triangle.

Associative Arrays

For the following messages, map stands for an associative array or a reference to one.

---

map<<First

Returns the first key within *map*, or Empty() if *map* has no keys. Note that keys are returned in lexicographical order.

---

map<<Get Contents

Returns a list of all key-value pairs within *map*.

---

map<<Get Keys

Returns a list of all the keys within *map*.

---

map<<Get Default Value()

Returns the implicit value of all absent keys, or Empty() if none has been set.

---

map<<Get Value(*key*)

Returns the value for the *key* within *map*.

`map<<Get Values(<{keyList}>)`

If no argument is provided, a list of all values within *map* is returned.

If a list of keys is provided, a list of the values corresponding to only those keys is returned.

`map<<Insert(key, value)`

Inserts the *key* into *map* and assigns *value* to it. If *key* already exists in *map*, its value is replaced by the new *value* given. This message is equivalent to the function `Insert Into`.

`map<<Next(key)`

Returns the key following the given *key* within the *map*, or `Empty()` if *map* has no keys. Note that keys are returned in lexicographical order.

`map<<Remove(key)`

Removes the *key* and *value* from *map*. This message is equivalent to the function `Remove From`.

`map<<Set Default Value(v)`

Sets the implicit value of all absent keys. Any key added without a value is assigned this value by default.

Classes

`obj<<Clone`

Returns a reference to a new class object that is a copy of the *obj* class object.

`obj<<Contains(quoted string)`

Returns 1 if the *obj* class object contains the specified quoted string expression, and 0 otherwise.

`obj<<Delete Class`

Deletes the *obj* class object.

obj<<Equal(*classref*)

Returns 1 if the *classref* class object is equal to the *obj* class object, and 0 otherwise.

obj<<First

Returns the quoted string representation of the name of the first member (item) in the *obj* class object. The members (items) in the class object are sorted in alphabetical order.

obj<<Get Contents

Returns a list of members (items) in the *obj* class object. Each element in the list is a two-item list that contains a key and an associated value.

obj<<Get Keys

Returns a list of keys within the *obj* class. Each key is a quoted string representation of the name of a member (item) in the *obj* class object.

obj<<Get Name

Returns a quoted string representation of the name of the *obj* class object.

obj<<Get Value(*key quoted string*)

Returns the value of the specified member (item) within the *obj* class object. The quoted *key quoted string* argument specifies the key to the member (item).

obj<<Get Values

Returns a list of values of the members (items) in the *obj* class object. Each element in the list is the expression that represents the value of each member (item) in the class.

obj<<Insert(*quoted string*, *value*)

Inserts a member (item) into the *obj* class object. The quoted *string* argument is the name of the member (item), and the *value* argument is the expression value of the member (item).

obj<<Lock Class(<*quoted string*|*{quoted stringList}*>)

Locks the *obj* class object, or locks specific members (items) within the *obj* class object. When a class object is locked, members (items) cannot be added, changed, or removed. The quoted *string* or quoted *stringlist* arguments specify a member (item) to lock. You can also specify a list of quoted strings to lock multiple members (items).

obj<<N Items

Returns the number of members (items) in the *obj* class object.

obj<<Next(*quoted string*)

Returns the quoted *string* representation of the name of the member (item) in the *obj* class object that follows the member (item) specified by the quoted *string*. The members (items) in the class object are sorted in alphabetical order.

obj<<Remove(<*string*|*{stringList}*>)

Removes the member (item) specified by the quoted *string* or quoted *stringlist* from the *obj* class. You can remove multiple members (items) using a list of quoted strings.

obj<<Show Contents

Shows the contents of the *obj* class object in the log window.

obj<<Unlock Class(<*quoted string*|*{stringList}*>)

Unlocks the *obj* class object, or unlocks specific members (items) within the *obj* class object. When a class object is unlocked, members (items) can be added, changed, or removed. The quoted *string* or *quoted stringlist* specify a member (item) to unlock. You can also specify a list of quoted strings to unlock multiple members (items).

Data Tables

dt<<Add Column Properties(*property argument*, ...)

Adds the specified properties (such as Value Order and Missing Value Codes) to the selected column.

dt<<Add Multiple Columns(*column prefix*, *n*, <"Before First"|"After Last"|After(*column*)>, "Character"|"Numeric"|"Row State", <Field Width(*n*)>)

Description

Adds *n* columns to *dt* at the position indicated.

Required Arguments

column prefix The prefix to add to the new columns names.

n The number of columns to add.

Character A new character column.

`Numeric` A new numeric column.

`Row State` A new row state column.

Optional Arguments

`Before First` Adds the columns before the first column.

`After Last` Adds the columns after the last column.

`After(`*column*`)` Adds the columns after the specified column.

`Field Width(`*n*`)` Specifies the width of the columns.

Notes

If you omit arguments, or the arguments are incorrectly specified, the Add Multiple Columns window appears.

`dt<<Add Rows(<`*n*`>, <"At Start"|"At End"|After(`*row number*`)>|{`*column name=value pairs*`})`

Description

Adds rows at the start, at the end, or after a specified row in the data table. The message can also add rows based on the specified column name and value pairs. Those rows are added to the end of the data table.

Notes

If you omit arguments, or the arguments are incorrectly specified, the Add Rows window appears.

`dt<<Add Scripts to Table(`*script*`, ...)`

`dt<<Add Properties to Table(`*script*`, ...)`

Adds the specified scripts to the data table.

`dt<<Anonymize(<Columns(`*column list(s)*`)>, <Output Table Name(`*name*`)>);`

Removes unique identifiers from data, some column properties, and table scripts. Applies to a data table or the specified list of columns. The new data table has the name that is specified by the quoted *name* argument.

`dt<<Begin Data Update`

Holds off display updating to allow for quick updating of data table cells. Use `End Data Update` in conjunction with this command to turn display updating back on.

Notes

`Begin Data Update` does not affect the data refresh due to some other table manipulations. For example, when you delete or add columns, the data table is updated and then the data update begins.

dt<<Clear Column Selection

Deselects all selected columns.

dt<<Clear Edit Lock(<"Modify Cells">, <"Add Rows">, <"Add Columns">, <"Delete Rows">, <"Delete Columns">)

Description

Allows the specified data table operation again.

Notes

If no arguments are specified, all locks are cleared.

dt<<Clear Row States

Cancels any row states in effect.

dt<<Clear Select

Turns off the current selection.

dt<<Clone Formula Column(*column*, *n*, Substitute Column Reference(*column1*, *{list}*))

Creates *n* new formula columns, substituting references to *column1* with columns from the *list* into the formula from the original *column*.

dt<<Close Data Grid(*Boolean*)

If true, closes the data table grid.

dt<<Close Side Panels(*Boolean*)

If true, closes the side panel in a data table.

dt<<Color or Mark by Column(*column*, *<named arguments>*)

dt<<Color by Column(*column*, *<named arguments>*)

dt<<Marker By Column(*column*, *<named arguments>*);

Description

Assigns colors or markers according to the values of a data table column. If no optional arguments are provided, colors are assigned according to the default color theme.

Required Argument

`column` The column to color or mark.

Optional Named Arguments

`Color(n)` Uses the specified JMP color.

`Add Marker(Boolean)` Shows or hides the marker in the data table.

`Color Theme(color theme)` Uses the specified quoted *color theme*.

`Marker Theme(marker theme)` Uses the specified quoted *marker theme:* `"Standard"`, `"Hollow"`, `"Paired"`, `"Classic"`, or `"Alphanumeric"`.

`Continuous Scale|Continuous Scale(Boolean)` Assigns colors in a chromatic sequential fashion based on the values in the highlighted column.

`Reverse Scale|Reverse` Reverses the color scheme in use.

`Excluded Rows(Boolean)` If true, applies the row states to excluded columns.

`"Make Window with Legend"` Creates a separate window with a legend.

`dt<<Color Rows by Row State`

Colors the rows in the data table grid using the color assignments by row states. Send the message again to turn off the row colors.

`dt<<Combine Columns(Delimiter("delim"), Columns(column1, column2, etc.), Column Name(quoted string))`

Combines several columns into a single column. Each source columns' values are separated by the delimiter specified as the quoted *delim* argument.

Examples

```
dt = Open( "$SAMPLE_DATA/Consumer Preferences.jmp" );
dt << Combine Columns(
  Delimiter( "," ),
  Columns(
     :Brush After Waking Up,
     :Brush After Meal,
     :Brush Before Sleep,
     :Brush Another Time
  ),
  Column Name( "When to Brush" )
);
```

`dt<<Compress File When Saved(Boolean)`

Compresses the file when the data table is saved.

`dt<<Compress Selected Columns({column1, ...})`

Compresses the listed columns into the most compact form that is possible. Columns with character data are compressed to 1 byte if there are fewer than 255 levels. Columns with numeric data are compressed to 1 byte if the integers are between -127 and 127.

`dt<<Concatenate(dt2|Data Table(name)|Multiple Data Table(name) arguments, (<"Private"|"Invisible">), <Output Table Name(name)>|"Append to First Table">, <"Keep Formulas">, <"Create Source Column">)`

Description

Creates a new table (*name*) from the rows of *dt* and *dt2*. By default, Concatenate creates a new data table and appends the rows of each data table that is specified.

Returns

A reference to the concatenated data table.

Required Arguments

`dt2|Data Table(name)|Multiple Data Table(name)` A data table reference or the names of the data table or data tables that you would like to combine.

Optional Arguments

`"Private"` A quoted keyword that opens the data table without displaying it in a data table window.

`"Invisible"` A quoted keyword that hides the data table. Use this argument to keep the data table hidden but use it in a subsequent expression. The data table is displayed in the Home Window's Window List and the Window > Unhide list.

`Output Table name(name)` The name of the final data table. If you do not enter a name, JMP names the data table Untitled # (for example, Untitled 1).

`"Append to First Table"` Appends rows to the first data table reference or data table name in the first argument. This option is an alternative to creating a new data table.

`"Keep Formulas"` Includes formulas in the final data table.

`"Create Source Column"` Adds a column called Source Table to the new data table.

Notes

`"Private"` and `"Invisible"` only apply if *not* using `"Append to First Table"`.

`dt<<Copy Column Properties`

Copies all of the column properties for the selected columns into a list of separate lists of properties. Optionally, you can specify a list of source columns instead of preselecting them in the data table.

Example

```
dt = Open( "$SAMPLE_DATA/Tiretread.jmp" );
dt << Select Columns( :MODULUS, :ELONG );
dt << Copy Column Properties;
New Window( "Script", Script Box( "//Try paste here" ) );
```

or

```
dt = Open( "$SAMPLE_DATA/Tiretread.jmp" );
dt << Copy Column Properties( {:MODULUS, :ELONG} );
```

```
New Window( "Script", Script Box( "//Try paste here" ) );
```

dt<<Copy Selected Properties

Description

Copies the selected table properties to the clipboard.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Properties( {"Distribution", "Oneway"} );
proplist = dt << Copy Selected Properties();
New Window( "Script", Script Box( "//Try pasting here" ) );
```

dt<<Copy Table Script("No Data")

Copies the script to recreate the data table onto the clipboard so that it can be pasted somewhere else. Add the "No Data" argument to omit the data.

dt<<Copy Table Scripts

dt<<Copy Selected Properties

Description

Copies the selected scripts to the clipboard.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Properties( {"Distribution", "Oneway"} );
proplist = dt << Copy Table Scripts();
New Window( "Script", Script Box( "//Try pasting here" ) );
```

dt<<Data Filter(<Location(*x*, *y*), <"Close Outline">, <"Local">, <Inverse(*Boolean*)>, <Show Columns Selector(*Boolean*)>, <Title(*quoted string*)>, <Save and Restore Current Row States(*Boolean*)>, <Conditional(*Boolean*)>, <Auto Clear(*Boolean*)>, <Group By AND(*Boolean*)>, <Show Histograms and Bars(*Boolean*)>, <Count Excluded Rows(*Boolean*)>, <Mode(...)>, <Add Filter((*cols*(...), <Where(...)>, <Display(...)>, <Select Missing(*cols*)>, <Order By Count(*cols*)>)>, <Favorites(...)>, <Animation(...)>)

Constructs a data filter. If no arguments are specified, the Add Filter Columns window appears.

Optional Arguments

Location(*x*, *y*) Moves the data filter window to the specified location. *x* and *y* are measured in pixels. 0,0 is the top left of the monitor.

"Close Outline" Closes the data filter outline.

"Local" Enables the filter to be embedded in reports to filter one or more platforms without affecting other reports.

Inverse(*Boolean*) Selects all but the specified rows for all filters.

Show Columns Selector(*Boolean*) If true, a column list is shown that adds a new column to the filter.

Title(*quoted string*) The title that is displayed on the outline.

Save and Restore Current Row States(*Boolean*) Restores your current row states when the Data Filter window is closed.

Conditional(*Boolean*) Limits the categories displayed for the selected filter column.

Auto Clear(*Boolean*) If you have more than one nominal or ordinal column selected in the Data Filter, this option clears any other selections before making a new selection.

Group By AND(*Boolean*) Enables you to create a filter group, specify OR, and add one or ore filters to create second filter group. If you specify Grouped By And, the behavior is reversed and grouped by AND instead.

Show Histograms and Bars(*Boolean*) Shows or hides the histogram and bars in the data filter.

Count Excluded Rows(*Boolean*) Shows or hides the number of excluded rows.

Mode The three modes of filtering: Select(*Boolean*) shows or hides the selected rows in the data table in a highlighted state; Show(*Boolean*) shows or includes the unselected rows and shows the Hide icon; Include(*Boolean*) shows or includes the unselected rows and shows the Exclude icon.

The global data filter default is Select(), Show(0), and Include(0), The local data filter default is Show(1), Include(1), Select() is not a valid option.

Add Filter Creates the data filter. Arguments include Columns(), Where(), Display(), Select Missing(*cols*), and Order By Count(*cols*). Columns() takes one or more column names separated by commas. You can add one or more Where clauses to define the filter.

Where Defines a Where clause by which the data is filtered.

Display(*column*, *size*, *display type*) Sets how the specified categorical column levels are displayed in the filter. The arguments are Blocks Display, List Display, Single Category Display, Check Box Display, Radio Box Display. In categorical columns, you can include the Find(Set Text(*quoted string*)) argument to include and initialize the search field. Display can also be included for a continuous column and can contain a size argument.

Select Missing Cols(*cols*) Selects missing values in continuous columns.

Order by Count(*cols*) For a categorical column, this option sorts the values in decreasing order by count.

Favorites Saves the current data filter criteria as a favorite.

Animation Cycles through the sorted values of the specified column, selecting and deselecting rows. Optional arguments include Animate Column(*col*), Animate

Rate(*number*), and "Forward"|"Backward"|"Bounce"highlights values from first to last. "Backward" highlights values from last to first. "Bounce" highlights forward and then backward repeatedly.

---

dt<<Data View(*<named arguments>*)

Description

Duplicates the data table in a new window. If you specify one of the following quoted arguments, the new data table includes only the corresponding rows.

Returns

A reference to the data view.

Optional Named Arguments

Excluded The new data table includes only the rows that are marked as excluded in the original data table.

Labeled|Labelled The new data table includes only the rows that are marked as labeled in the original data table.

Hidden The new data table includes only the rows that are marked as hidden in the original data table.

Selected The new data table includes only the rows that are selected in the original data table.

---

dt<<Delete Columns(*column1*, *column2*, ...)

dt<<Delete Column

Description

Deletes one or more columns from the data table *dt*. Specify which column or columns to delete. Without an argument, deletes the selected columns, if any.

---

dt<<Delete Rows(*<n>*)

dt<<Delete Rows(*{n, o, p, ...}*)

dt<<Delete Rows(*{n::q}*)

dt<<Delete Rows([n, o, p])

dt<<Delete Row(*preceding arguments*)

Description

Deletes the currently selected rows or rows specified. Returns the number of rows that were deleted.

dt<<Delete Scripts(*table script name*|*{table script1, table script2, ...}*)

Description

Deletes the specified data table script or scripts with the quoted name or names, or deletes a list of data table scripts.

Notes

In JMP versions prior to 14, use `Delete Table Property` to delete a table script.

dt<<Delete Table Property(*name*|*{property1, property2, ...}*)

Deletes a table property (for example, a script or variable) with the quoted *name*.

dt<<Delete Table Variable(*name*)

Deletes a table variable with the quoted *name*.

dt<<Disable Undo(*Boolean*)

If true, disables undo operations in the data table.

dt<<End Data Update

Resumes display updating after a `Begin Data Update` message. These commands are used for quick updates of the data table when many changes have to be made. Speed is gained by turning off display updating.

dt<<Exclude

dt<<Unexclude

Toggles selected rows in *dt* from excluded to unexcluded or *vice versa*.

dt<<Get All Columns As Matrix

Returns the values from all columns of *dt* in a matrix. Character columns are numbered according to the levels, starting at 1.

dt<<Get As Matrix(*<{list of columns by name}>*|*<{list of columns by number}>*, *<column range>*)

Returns values from the numeric columns of *dt* in a matrix. The default output is all numeric columns.

Examples

```
dt1 = Open( "$SAMPLE_DATA/Big Class.jmp" );
```

```
cols = dt1 << Get As Matrix(); // returns all numeric columns
Show( cols );
   cols =
   [ 12 59 95,
      12 61 123,
      12 55 74,...]
colnums = dt1 << Get As Matrix( {4, 5} ); // returns columns four and five
Show( colnums );
   colnums = [ 59 95, 61 123, 55 74, 66 145, 52 64, 60 84, 61 128, ...]

dt2 = Open( "$SAMPLE_DATA/Probe.jmp" );
colrange = dt2 << Get As Matrix( 10::22); // returns columns 10 through 22
Show( colrange );
   colrange =
   [ -0.08818069845438 0.711340010166168 1.85904002189636 0.396923005580902
      4.50656986236572 7.86504983901978 1.53891003131866 -2.76178002357483
      0.0711032971739769 5.75577020645142 -3.62023997306824 -0.971698999404907
      -0.0525696985423565, ...]
```

dt<<Get As Report

Returns the data table as a report. If rows and columns are selected in the data table, only those rows and columns are in the report.

Example

The following script returns `Big Class.jmp` as a report and displays it and a distribution in one window.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dtRpt = dt << Get As Report;
distRpt = V List Box(
   dt << Distribution(
      Continuous Distribution( Column( :weight ) ),
      Nominal Distribution( Column( :age ) )
   )
);
New Window( "Example", H List Box( dtRpt, distRpt ) );
```

dt<<Get Cell Height

Returns the data table cell height in pixels.

dt<<Get Column Names(*quoted string, <modeling type>, <data type>*)

Description

Returns a list of column names in a data table. The quoted string returns a list of quoted strings rather than a list of column references.

Required Argument

quoted string Returns a list of quoted strings rather than a list of column references.

Optional Arguments

modeling type The quoted modeling type. The options are "Continuous", "Ordinal", "Nominal", "Multiple Response", "Unstructured Text", "None", and "Vector".

data type The quoted data type. The options are "Numeric", "Character", "Row State", and "Expression".

Notes

The data types and the modeling types get only the specified types of columns. More than one of each type can be specified.

---

dt<<Get Column Reference(*{list of column names}|[matrix of column numbers]*)

dt<<Get Column References(*{list of column names}|[matrix of column numbers]*)

Returns the column references of the quoted strings in the list or matrix. If no list or matrix is used, JMP returns a list of references to all columns.

---

dt<<Get Display Width

Returns the column display width in pixels.

---

dt<<Get Edit Lock

Returns a list of the disallowed operations on the data table (if cells cannot be edited; rows cannot be added or deleted; and columns cannot be added or deleted).

---

dt<<Get Excluded Columns

Returns the currently excluded columns in the data table.

---

dt<<Get Excluded Rows

Returns the rows that are excluded in the data table.

---

dt<<Get Header Height

Returns the column header's display height (in pixels).

---

dt<<Get Hidden Columns

Returns the columns that are hidden in the data table.

dt<<Get Hidden Rows

Returns the currently hidden rows in the data table.

dt<<Get Journal

Returns a quoted string that contains journal source for the display box.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( Y( :weight ), X( :height ) );
rbiv = biv << Report;
Print( rbiv << Get Journal );
```

dt<<Get Label Columns

dt<<Get Labeled Columns

dt<<Get Labelled Columns

Returns the currently labeled columns in the data table.

Example

In PopAgeGroup.jmp, the Country and Year columns are labeled. The following script returns a list of the labeled column names.

```
dt = Open( "$SAMPLE_DATA/PopAgeGroup.jmp" );
dt << Get Labeled Columns;
   {:Country, :Year}
```

dt<<Get Labeled Rows

dt<<Get Labelled Rows

Returns the currently labeled rows in the data table.

dt<<Get Name

Returns the name of the data table.

dt<<Get Path

Returns the absolute path for the JMP data table. Note that this function is not for imported data that is not saved yet.

dt<<Get Property(*name*)

Returns the script from the quoted property *name*.

dt<<Get Row Change Function

Returns the expression that is evaluated when a row is selected.

dt<<Get Row ID Width

Returns the row ID display width in pixels.

dt<<Get Row States

Returns a vector containing the row state for every row in the data table or data filter.

dt<<Get Rows Where(*where clause*)

Returns the rows in the data table that match the specified Where criteria. Here are some examples:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Get Rows Where( :sex == "M" );
dt << Get Rows Where( :sex == "M" & :age < 15 );
```

dt<<Get Script(<*script name*>)

Returns the script specified by the quoted *script name*. If the *script name* is omitted, `Get Script` returns a text representation of the data table and all scripts in the table.

dt<<Get Script Group(<*group name*>)

Description

Returns the list of table scripts in the quoted *group name*. If no group name is specified, a list of all table scripts in all groups is returned.

dt<<Get Script Group Names

Description

Returns the list of names of table script groups.

dt<<Get Scroll Locked Columns

Returns a list of columns that are locked from scrolling.

dt<<Get Selected Columns(<*quoted string*>)

Description

Returns a list of selected columns as column references. Include the quoted *string* argument to return the selected column names as a list of strings.

dt<<Get Selected Properties(*<{list of properties}>*)

Description

Returns the selected table properties in a list.

Optional Argument

list of properties Specifies the properties to get.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Select Properties( {2, 4} );
proplist = dt << Get Selected Properties();
// returns the second and fourth table scripts and highlights them
// in the data table
```

dt<<Get Selected Rows()

Returns the selected rows.

dt<<Get Table Script Names()

Returns a list of the names of all the scripts and properties in the data table.

dt<<Get Table Variable(*name*)

Returns the value of the quoted *name* variable.

dt<<Get Table Variable Names

Returns a list of the names of all the variables in the data table.

dt<<Go To Row(*n*)

Locates and selects row number *n* in *dt*.

dt<<Group Columns(*{column1, column2, ...}*)

dt<<Group Columns(*group name*, *column*, *n*)

dt<<Group Columns(*first column*, *n*)

Description

Groups the columns under the specified quoted *group name*. You can provide either a list of columns to group, or a column name and the number of columns to group. In the latter case, the number *n* specifies to group the column given with the *n*-1 columns that follow.

dt<<Group Scripts({*script1, script2, ...*})

Description

Groups a list of table scripts in the data table.

dt<<Hide

dt<<Unhide

Toggles selected rows in *dt* from hidden to unhidden or *vice versa*.

dt<<Hide and Exclude

Hides the selected rows from graphs and excludes them from contributing to calculations.

dt<<Invert Column Selection(<{*list of columns*}>)

Description

Selects any column that is currently deselected and deselects any column that is currently selected. If the *list of columns* is specified, the columns that are not in the list are selected.

dt<<Invert Row Selection

Selects any row that is currently deselected and deselects any row currently selected.

dt<<Is Dirty

Returns 1 if the table has been modified from its saved state. Otherwise, returns 0.

dt<<Join(With(*Data Table(name)*), (<"Private">|<"Invisible">), Select(*columns*), Select With(*columns*), (By Matching Columns(*column1=column2, ...*)|"Cartesian"|"By Row Number"), <"Merge Same Name Columns">, <"Match Flag">, <Copy Formula(*Boolean*)>, <Suppress Formula Evaluation(*Boolean*)>, <"Update">, <Drop Multiples(*Boolean*, *Boolean*)>, <Include Non Matches(*Boolean*, *Boolean*)>, <"Preserve Main Table Order">, <Output Table Name(*name*))>

Description

Combines data tables *dt* and *Data Table* side to side.

Returns

A data table.

Required Arguments

With(*Data Table(name)*) Specifies the data table to join with the active table.

"Private" Specifies a quoted keyword that opens the data table without displaying it in a data table window.

"Invisible" Specifies a quoted keyword that hides the data table. Use this argument to keep the data table hidden but use it in a subsequent expression. The data table is displayed in the Home Window's Window List and the Window > Unhide list.

Select(*columns*) Selects the data table to join with the active table.

Select With(*columns*)

By Matching Columns(*column1=column2*) Selects columns in both tables whose values and data types match.

"Cartesian" Joins two tables using a Cartesian fashion, where it forms a new data table consisting of all possible combinations of the rows from two original data tables. JMP crosses the data in the first table with the data in the second to display all combinations of the values in each set.

"By Row Number" Joins the two tables side by side.

"Merge Same Name Columns" Data from the second table replaces the data of the same name columns in the original table. Note that missing values in the first table are replaced by nonmissing values in the second.

"Match Flag" Determines whether the Match Flag column is created when you are matching by column.

Copy Formula(*Boolean*) Includes formulas from the main table and/or the second table in the output columns.

Suppress Formula Evaluation(*Boolean*) Prevents JMP from evaluating columns' formulas during the creation of the new table.

"Update" Column data from the second table change the data of the same name columns in the original table. The results are displayed in a new data table. Note the following: JMP does not replace data with missing values; the output table uses the same columns as the original table. Thus, when you use "Update", Select Columns, the "Update" option is available only when joining by row number or by matching columns.

Drop Multiples(*Boolean*, *Boolean*) Specifies that you want the new table to contain only one row for each name. Applies only when matching by columns.

Include Non Matches(*Boolean*, *Boolean*) Includes non-matching columns in the main table and new data table. Applies only when matching by columns.

"Preserve Main Table Order" Maintains the order of the original data table in the joined table, instead of sorting by the matching columns.

Output Table Name(*name*) Specifies the name of the joined table. If you do not specify a name, JMP names the data table Untitled # (for example, Untitled 1).

---

dt<<Journal

Makes a journal from the data table. Only the data grid is included, not notes, variables, or scripts.

Note

Journals that are created in JMP 14 or later might contain compressed matrix data for large matrices. If you have JSL scripts that open journals and extract data from them, you might need to use the `Get Journal` message (which does not compress the matrices) rather than saving the journal to disk with the `Journal` message.

See Also

See "dt<<Get Journal"

`dt<<Journal Link(<Save(<path>)|Embed()>, <Button Name(<name>)>))`

Adds a link to the data table in the current journal. If a journal does not exist, a new one is created.

Optional Arguments

`path` Specifies the quoted *path* where the table is saved. If omitted, the data table should already have a disk location (previously saved or loaded), otherwise the journal link is incomplete and will not reload the table.

`Embed` Embeds a JSL script to recreate the data table.

`Button Name(name)` Specifies the name that is displayed on the button. The *name* argument is quoted. If the button name is not specified, the button is named after the data table.

`dt<<Label`

`dt<<Unlabel`

Toggles selected rows in *dt* from labeled to unlabeled or *vice versa*.

`dt<<Last Modified`

Returns the date on which the data table was last saved.

`dt<<Layout`

`Layout` is deprecated and will be removed in a future release. Use `Journal` instead.

`dt<<Lock Data Table`

Description

Locks the data table so that data and column properties cannot be added or changed.

See Also

"dt <<Set Edit Lock(<"Modify Cells">, <"Add Rows">, <"Add Columns">, <"Delete Rows">, <"Delete Columns">)"

dt<<Make Indicator Columns(<Append Column Name(*Boolean*)>, <Include Missing(*Boolean*)>)

Creates indicator columns of 0 and 1 values for the specified categorical columns.

Example

The following example creates indicator columns for the sex column. Append Column Name creates columns named sex_F and sex_M. Otherwise, the columns are named after each level (F and M). Include Missing includes missing values.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Make Indicator Columns(
   Columns( :sex ),
   Append Column Name( 1 ),
   Include Missing( 1 )
);
```

dt<<Make RowState Handler

Creates a row state handler function. The argument of the function holds the rows whose row states get changed.

dt<<Make SAS Data Step

Returns the data table as a SAS Data Step.

dt<<Make SAS Data Step Window

Returns the data table as a SAS Data Step and places it in a SAS script window.

dt<<Make Validation Column

Creates a column that is used to divide the data into training and validation sets.

dt<<Marker by Column(*column*)

Description

Assigns markers according to the values of the specified data table *column*.

See Also

"dt<<Marker By Column(column, <named arguments>);"

dt<<Markers(*n*)

Assigns marker *n* to the selected rows.

dt<<Maximize Display

Deprecated. Use Optimize Display instead.

Forces the data table to remeasure all of its columns and zoom to the best-sized window.

dt<<Move Script Group(*group name*, "To First"|"To Last"|After(table *script name*)|After(*group*))

Description

Rearranges the table script groups that are specified by the quoted *group name*.

dt<<Move Selected Column(*name(s)*, "To First"|"To Last"|After(*name*))

dt<<Move Selected Columns(*name(s)*, "To First"|"To Last"|After(*name*))

Description

Moves the selected column or columns in the data table to the specified position. The *name* argument is quoted.

Example

The following example moves the age column to the last column in Big Class.jmp:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Go To( :age );
dt << Move Selected Columns( "To Last" );
```

You may also use a list to specify the column names.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
list = {"name", "sex"};
dt << Move Selected Columns( list, "To Last" );
```

dt<<Move Rows("At Start"|"At End"|After(*n*))

Moves the selected rows in the data table to the specified position. *n* represents a row number.

dt<<New Column(*name*, *<data type>*, *<modeling type>*, <Format(*format*, *width*)>, <Formula()>, <Set Values(*{..., ..., }*)>, <Set Property*(properties)*>)

Description

Adds a new column titled with the quoted name after the last column in *dt*. Unless otherwise specified, columns are numeric, continuous, and 12 characters wide.

Returns

A column reference.

Required Argument

`name` The name of the new column.

Optional Arguments

`data type` A quoted string that specifies the data type. The options are `"Numeric"`, `"Character"`, `"Row State"`, or `"Expression"`.

`modeling type` A quoted string that describes the modeling type (`"Continuous"`, `"Nominal"`, `"Ordinal"`, `"Multiple Response"`, `"Unstructured Text"`, `"None"`, or `"Vector"`).

`Format(`*`format`*`, `*`width`*`)>` Sets the format type and column width.

`Set Values({})` Specifies the data in the column.

`Formula` Specifies the column formula.

`Set Property(`*`properties`*`)` Specifies any messages that data table columns support. Action arguments are found in the Column Properties menu in the New Column window. `Axis` and `Link Reference` are action argument.

See Also

For examples of setting other numeric format properties, see "col<<Format(<width>, <decimal places>, <"Use Thousands Separator">)".

dt<<New Data Box()

Makes a data table view in a display box tree. Useful for displaying the data table and report in one window. A data browser box is created when you send the `New Data Box` message to the data table object.

Example

The following script creates a data table view and report in one window. The data table is placed in a data browser box. The width of that box is set to 800 pixels. Because auto stretch is turned off, the data table view remains 800 pixels wide even if you stretch the right border of the window.

```
dtA = Open( "$SAMPLE_DATA/Semiconductor Capability.jmp", invisible );
nw = New Window( "Example",
  H List Box(
     V List Box( dtbox = dtA << New Data Box() ),
     dtA << Distribution(
        Continuous Distribution( Column( :NPN1 ) ),
        Continuous Distribution( Column( :PNP1 ) )
     )
  )
);
dtbox << Set Stretch( "Off", "Off" ) << Set Width( 800 );
```

dt<<New Data View

Opens a duplicate of the data table. The second data table is identical to and linked to the original data table, so that any changes made in one are reflected in the other. Closing either data table also closes the other and all references to the data tables are deleted.

This can be useful to show an invisible data table.

dt<<New Script(*name*, *script*)

dt<<Set Property(*name*, *script*)

Creates a new table property (also called a *table script*) using the quoted *name* that stores the specified *script*.

Use New Script() or Set Property() rather than the deprecated New Property() and New Table Property().

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Set Property( "Bivariate Example", Bivariate( Y( :weight ), X( :height
  ), Fit Line ) );
```

dt<<New Table Variable(*name*, *number*)

dt<<Set Table Variable(*name*, *number*)

Creates a new table variable with the quoted *name* and the *number*.

dt<<Next Selected

Scrolls data table down to show the next selected row that is not already in view.

dt<<Optimize Display

Forces the data table to remeasure all of its columns and zoom to the best-sized window.

dt<<Original Order

Restores saved order of columns in *dt*.

dt<<Paste Column Properties

Pastes multiple lists of column properties to multiple columns. Optionally, you can specify a list of target columns instead of selecting them in the data table.

Example

```
dt = Open( "$SAMPLE_DATA/Tiretread.jmp" );
dt << Copy Column Properties( {:MODULUS, :ELONG} );
dt2 = New Table( "test it",
```

```
    New Column( "T1", numeric, continuous ),
    New Column( "T2", numeric, continuous ),
    New Column( "T3", numeric, continuous ),
    Add Rows( 10 )
);
dt2 << Paste Column Properties( {:T1, :T3} );
// pastes the column properties from MODULUS and ELONG to T1 and T3
```

`dt<<Predictor Screening`

`dt<<Screen Predictors`

Description

Used to identify strong predictors.

Example

```
dt = Open( "$SAMPLE_DATA/Boston Housing.jmp" );
obj = dt << Predictor Screening(
  Y( :mvalue ),
  X( :crim, :indus, :chas, :nox, :rooms, :age, :distance, :radial )
);
```

`dt<<Previous Selected`

Scrolls data table up to show the previous selected row that is not already in view.

`dt<<Print Window(<"Show Dialog">)`

Prints the window. If the optional named argument `"Show Dialog"` is specified, the print window is displayed. Otherwise, the window is printed to the default printer using the current settings, and no print window is displayed.

`dt<<Rename Script Group(`*`old name`*`, `*`new name`*`)`

Description

Renames the table script group.

Example

```
dt << Rename Script Group( "Maps", "Street Maps" );
```

`dt<<Reorder By Data Type`

Reorders columns in *dt*, row state first, then character, then numeric.

`dt<<Reorder By Modeling Type`

Reorders columns in *dt* to continuous, then ordinal, then nominal.

`dt<<Reorder By Name`

Reorders columns in *dt* to alphanumeric order by name.

`dt<<Rerun Formulas`

Recalculates all formula-based data table variables. Recalculations are performed in the proper dependency order.

`dt<<Reverse Order`

Reverses columns in *dt* from current order.

`dt<<Revert`

Reverts to the most recently saved version of *dt*.

`dt<<Row Selection(Select Where(condition), <current selection("Extend"|"Restrict"|"Clear")> <Dialog("Keep Dialog Open")>)`

Description

Selects all rows that meet the specified condition.

Required Argument

`Select Where(condition)` Specifies the condition by which the rows are selected.

Optional Arguments

`current selection("Extend"|"Restrict"|"Clear")` Extends, restricts, or clears the existing selections. `Clear` is the default value.

`Dialog("Keep Dialog Open")` Shows the dialog so that the user can edit the options.

`dt<<Run Formulas`

Performs all pending formula evaluations, including evaluations that are pending as a result of evaluating other formulas.

`dt<<Run Script(`*name*`)`

Finds the table property with the quoted name and runs it as a JSL script.

`dt<<Save(`*path*`)`

`dt<<Save As(`*path*`)`

Description

Saves the table in the specified quoted *path*.

`dt<<Save Database(connection information, table name, <"Replace">)`

Saves the data table to the database named using the quoted connection information and quoted table name. The `"Replace"` option replaces the existing database with the current database.

`dt<<Save Script to Script Window`

Saves a script to reproduce the data table in a script editor window. Appends the script to any script that currently appears in the script editor.

`dt<<Select All Rows`

Selects all rows in the data table.

`dt<<Select Columns(<column1>, <column2>,...|"All")`

Selects the specified columns (or all columns) in the data table.

`dt<<Select Duplicate Rows`

Description

Selects the second and subsequent duplicate rows. If columns are selected, duplicate values are found in the rows of those columns. The duplicate values are case sensitive.

`dt<<Select Excluded`

Selects only those rows in the data table that are currently excluded.

`dt<<Select Hidden`

Selects only those rows in the data table that are currently hidden.

`dt<<Select Labeled`

Selects only those rows in the data table that are currently labeled.

`dt<<Select Randomly(n|p|Sample Size(n)|Sampling Rate(p))`

Randomly selects the given percentage *p* of the rows in the data table, or the number of rows *n*. If you specify an argument with a keyword, numbers between 0 and 1 represent a percentage, and numbers larger than 1 represent a number of rows.

`dt<<Select Rows([row1, row2, ...])`

Selects the rows given in the list of row numbers.

dt<<Select Script Group(*<group name|{group1, group2, ...}>*)

Selects the table script group specified as a quoted *group name* or a list of quoted strings. If no argument is provided, all groups are selected.

dt<<Select Where(*condition*, <Current Selection("Extend"|"Restrict"|"Clear")>)

Description

Selects the rows in *dt* where the condition evaluates as true.

dt<<Set Dirty(*Boolean*)

Marks the data table as changed, even if no changes have been made.

dt <<Set Edit Lock(<"Modify Cells">, <"Add Rows">, <"Add Columns">, <"Delete Rows">, <"Delete Columns">)

Description

Prevents cells from being modified; rows from being added or deleted; and columns from being added or deleted.

dt <<Set Cell Height(*n*)

Sets the cell height to the specified number of pixels.

dt<<Set Header Height(*n*)

Sets the column header's height to the specified number of pixels.

dt<<Set Label Columns(*column1, columns2, ...*)

Assigns the specified columns as label columns.

dt<<Set Matrix(*[matrix]*)

Inserts the specified matrix into a data table, adding new columns and rows as necessary.

dt<<Set Name(*name*)

Description

Specifies a name for the table. The *name* argument is quoted.

Returns

The data table name as a quoted string.

Notes

A change was made to the `Set Name` message so that now the new table name is returned as a quoted string. In previous releases, `Set Name` returned a scriptable data table object. As a result of this change, JMP scripts might need to be updated for the desired result to be returned. For example, rewrite the following script:

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" ) << Set Name( "Test" );
```

Separate the messages so that *dt* represents the data table instead of "Test":

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
dt << Set Name( "Test" );
```

The result is the same as in previous releases but will run successfully in both earlier and newer versions of JMP.

`dt<<Set Property(`*name*`, `*script*`)`

See "dt<<New Script(name, script)".

`dt<<Set Label Columns(`*column(s)*`, ...)`

`dt<<Set Label Columns`

Turns on the Label attribute for the specified columns. If no columns are listed, it turns the Label attribute off.

`dt<<Set Row ID Width(`*n*`)`

Sets the row ID display width to the specified number of pixels. If *n* is set to zero, the row ID display width automatically resizes.

`dt<<Set Row States(`*[matrix]*`)`

Sets the row states for all rows in the data table.

`dt<<Set Scroll Lock Columns(`*column name, ...*`)`

Locks scrolling for the columns specified as quotes quoted strings. If no columns are listed, unlocks scrolling.

`dt<<Set Table Variable(`*name*`, `*value*`)`

See "dt<<New Table Variable(name, number)".

```
dt<<Sort(<"Private">|<"Invisible">, <"Replace Table">, By(columns),
Order("Descending" | "Ascending"), <Output Table Name(name))>
```

Description

Creates a new table (named after the quoted name) by rearranging the rows of *dt* according to the values of one or more *columns*.

Returns

A reference to the sorted table.

```
dt<<Split(Split(columns), Split By(column), <Group(column)>,
<"Private">|<"Invisible">, <Remaining Columns("Keep All"|"Drop All"|
Keep(columns)|Drop(columns))>, <Copy Formula(Boolean)>, <Suppress
Formula Evaluation(Boolean)>, <Sort by Column Property>, <Output
Table (name)>)
```

Description

Unstacks multiple rows for each `Split` column into multiple columns as identified by the `Split by` column. The `Split` and `Split by` arguments are required.

Returns

A reference to the split data table.

Required Arguments

`Split(columns)` The column to split.

`Split By(column)` The column to split by.

Optional Arguments

`Group` Splits data within the specified groups.

`Remaining Columns("Keep All"|"Drop All"| Keep(columns)|Drop(columns))` Specifies what to do with the remaining columns in the resulting table. `Keep All` is the default setting.

Note: `Keep All` includes all columns in the output data table. However, the *values* of every column are not included. Because multiple rows are collapsed to a single row in the output data table, some values of the kept columns are dropped.

`Copy Formula(Boolean)` Includes column formulas from the source table in the resulting table.

`Suppress Formula Evaluation(Boolean)` Stops any copied formulas from being evaluated. True is the default setting.

`Sort by Column Property` Sorts the order of the output columns by the sort column property that is defined for the Split by column.

`Output Table(name)` Generates the output to the specified table name.

---

dt<<Stack(<"Private">|<"Invisible">, Columns(*columns*), <Source Label Column(*quoted string*)>, <Stacked Data Column(*quoted string*)>, <Copy Formula(*Boolean*)>, <Suppress Formula Evaluation(*Boolean*)>, <Drop All Other Columns(*Boolean*)|Name(*non-stacked columns*)(Keep(*column1, column2, ...*))|Name(*non-stacked columns*) (Drop(*column1, column2, ...*))>, <Output Table(*name*)>), <Number of Series(*n*)>, <"Contiguous">

Description

Creates a new table by combining the values from several columns in *dt* into one column.

Returns

A reference to the stacked data table.

---

dt<<Subscribe("keyname"(<"*client*">), On Delete Columns(<*function*>|<*script*>)|On Add Columns(<*function*>|<*script*>)|On Add Rows(<*function*>|<*script*>)|On Delete Rows(<*function*>|<*script*>)|On Rename Column(<*function*>|<*script*>)|On Close(<*function*>|<*script*>)|On Save(<*function*>|<*script*>)|On Rename(<*function*>|<*script*>))

Description

Subscribes to a data table to get messages regarding changes in the data table.

Returns

The keyname.

Arguments

"keyname"(<*client*>) Specifies the subscription name so that it can be referenced. The quoted client triggers a close confirmation when a close is attempted on the data table, warning that other open windows depend on the data table.

On Delete Columns(<*function*>|<*script*>) Returns the keyname when columns are deleted.

On Add Columns(<*function*>|<*script*>) Returns the keyname when columns are added.

On Add Rows(<*function*>|<*script*>) Returns the keyname when rows are added.

On Delete Rows(<*function*>|<*script*>) Returns the keyname when rows are deleted.

On Rename Column(<*function*>|<*script*>) Returns the keyname when columns are renamed.

On Close(<*function*>|<*script*>) Returns the keyname when the data table is closed. Takes one argument, a function. The function requires only one argument, the data table name.

On Save(<*function*>|<*script*>) Returns the keyname when the data table is saved.

`On Rename(<function>|<script>)` Returns the keyname when a rename is attempted on the data table. The function can be either the name of a previously defined function or the function itself.

Notes

Each subscription option remains in effect until you unsubscribe.

```
dt<<Subset(<"Private">|<"Invisible">, <"Selected Columns">,
<Columns(column list)>, <"Selected Rows">, <Rows([number, number,
...])>, <By(column list)>, <Sampling Rate(fraction)>, <Sample
Size(integer)>, <Stratify(column list)>, <Link to Original Data
Table(Boolean)>, <Copy Formula(Boolean)>, <Suppress Formula
Evaluation(Boolean)>, <"Keep by Columns">)
```

Description

Creates a new table from the rows and columns that you specify in *dt*.

Returns

A reference to the subset data table.

```
dt<<Summary(<"Private">|<"Invisible">, <Group(column)>,
<Subgroup(column)>, <N(column)>, <Mean(column)>, <Std Dev(column)>,
<Min(column)>, <Max(column)>, <Range(column)>, <Sum(column)>,
<CV(column)>, <Freq(column)>, <Weight(column)>, "Include Marginal
Statistics", <Link to Original Data Table(Boolean)>, <Statistics
Column Name Format(Stat(column)|Column|Stat of Column|Column Stat)>)
```

Description

Creates a new table of summary statistics for the column that you specify, according to groups and subgroups. `Statistics Column Name Format` values are quoted.

Returns

A reference to the summary data table.

```
dt<<Suppress Formula Eval(Boolean)
```

Turns off automatic calculation of formulas for data table *dt*.

```
dt<<Text to Columns(Delimiters(<separator>, <"Tab">, <"Newline">),
Columns(column1, column2...))
```

Makes a set of text columns or indicator columns from a delimited text column. `"newline"` includes the three forms: \r, \n, and \r\n. The separator is quoted.

Example

```
dt = Open( "$SAMPLE_DATA/Consumer Preferences.jmp" );
dt << Text To Columns(
```

```
    delimiter( "," ),
    columns( :Brush Delimited )
);
```

`dt<<Transpose(Columns(`*`columns`*`), Rows([`*`matrix`*`]), Output Table Name(`*`name`*`))`

Description

Creates a new table (named after the quoted *name*) from the rows and columns that you specify.

Returns

A reference to the transposed data table.

`dt<<Ungroup Columns({`*`column1, column2, ...`*`})`

Ungroups the columns defined in the list argument.

`dt<<Ungroup Scripts(`*`Name of Script Group`*`|{`*`script1, script2,...`*`})`

Description

Removes the specified table scripts or group from the group. The *Name of Script Group* argument is quoted.

`dt<<Unsubscribe(`*`keyname`*`, "On Delete Columns"|"On Add Columns"|"On Add Rows"|"On Delete Rows"|"On Close"|"On Col Rename"|"All")`

Releases any previous subscriptions to the data table *dt*. The *keyname* argument is quoted.

`dt<<Update from Database`

Updates the data in the table *dt* with data reimported from the database.

Columns

`col<<Add Column Properties(`*`name`*`, `*`expression`*`)`

Adds the quoted column property *name* with the *expression* given. You can add any standard column property by name or a user-specified property.

`col<<Add From Row States`

Updates a row state column with any currently used row state changes that are not the default state.

`col<<Add To Row States`

Copies all row state values in a column that are not the default state to the currently used row state in the data table.

`col<<Color Cells(color)`

Description

Colors the cells of the column within the data table grid. Use any quoted named color or 0 to clear the color.

`col<<Color Cell by Value(Boolean)`

Description

Colors the cells of the column in the data table grid using the value color property.

`col<<Copy Column Properties`

Copies the column properties into the buffer.

`col<<Copy From Row States`

Copies all row state values currently used in the data table to a column.

`col<<Copy to Row States`

Copies all row state values in the column to the currently used row state in the data table.

`col<<Data Type(type, <Format(format quoted string)>, <Input Format(format quoted string)>, <width>)`

`col<<Set Data Type(type, <Format(format quoted string)>, <Input Format(format quoted string)>, <width>)`

Description

Sets the *data type* to *col*.

Required Argument

`type` Specifies the `"Numeric"`, `"Character"`, `"Row State"`, or `"Expression"` data type.

Optional Arguments

`Format(format quoted string)` Specifies the way the data are displayed, such as `h:m` for hours and minutes. The *format quoted string* argument is quoted.

`Input Format(format quoted string)` Specifies the way the data are input. The *format quoted string* argument is quoted.

`width` (Optional for numeric data) Specifies 1, 2, or 4 (the number of bytes in the column).

`col<<Delete Formula`

Deletes the formula from a column.

`col<<Delete Property(`*`name`*`)`

`col<<Delete Column Property(`*`name`*`)`

Deletes the quoted property name from a column.

`col<<Eval Formula`

Forces the formula to evaluate (perhaps again). If formula suppression is enabled, the evaluation is not performed.

`col<<Exclude(`*`Boolean`*`)`

Turns the excluded or unexcluded state on, depending on the Boolean argument.

`col<<Format(<`*`width`*`>, <`*`decimal places`*`>, <"Use Thousands Separator">)`

`col<<Format("Best", <`*`width`*`>, <"Use Thousands Separator">)`

`col<<Format(("Fixed Dec"|"Percent"), <`*`width`*`>, <`*`decimal places`*`>, <"Use Thousands Separator">)`

`col<<Format("Pvalue", <width>)`

`col<<Format(("Scientific"|"Engineering"|"Engineering SI"), <`*`width`*`>, <`*`decimal places`*`>)`

`col<<Format("Precision", <`*`width`*`>, <`*`decimal places`*`>, <"Use Thousands Separator">, <"Keep Trailing Zeroes">, <"Keep All Whole Digits">)`

`col<<Format("Currency", <"Currency Code">, <`*`width`*`>, <`*`decimal places`*`>, <"Use Thousands Separator">)`

`col<<Format("Datetime", <`*`width`*`>, <`*`input format`*`>)`

`col<<Format(("Latitude DDD"|"Latitude DDM"|"Latitude DMS"|"Longitude DDD"|"Longitude DDM"|"Longitude DDM"), <`*`width`*`>, <`*`decimal places`*`>, ("PUN"|"DIR"|"PUNDIR"))`

`col<<Format("Custom", Formula(...), <`*`width`*`>, <`*`input format`*`>)`

Description

Sets the numeric display specified format.

Arguments

See *Using JMP* for more information about the arguments.

Examples

```
col<<Format( 10, 2, "Use thousands separator");
col<<Format( "Currency", "EUR", 20 );
col<<Format( "m/d/y", 10 );
col<<Format( "Precision", 10, 2, "Keep trailing zeroes", "Keep all whole
  digits" );
col<<Format( "Latitude DDD", "PUNDIR"); // "PUN" for punctuation, "DIR" for
  direction, PUNDIR for both
col<<Format( "Custom", Formula( Abs( value ) ), 15 );
```

`col<<Formula(`*expression*`)`

`col<<Set Formula(`*expression*`)`

Sets the formula for the variable and evaluates it.

`col<<Get Column Field Width`

Returns the field width used for displaying data in the column.

`col<<Get Data Type`

Returns the data type of *col*.

`col<<Get Data Type Length`

Returns the data type and length of the data column. Only the data type is returned if the data length is not fixed, as with character columns.

`col<<Get Format`

Returns the format of the column.

`col<<Get Formula`

Returns the formula.

`col<<Get Hidden`

Returns 1 if the column is hidden.

`col<<Get Input Format`

Returns the format used for input and storing of data for the column.

`col<<Get Labeled`

Returns 1 if the column is labeled.

`col<<Get List Check`

Returns the list check definition. If list check is not defined for the column, a message is sent to the log stating so.

`col<<Get Lock`

Returns the current Lock setting.

`col<<Get Modeling Type`

Returns the modeling type of the column.

`col<<Get Name`

Returns the name of the column.

`col<<Get Property("property name")`

Returns the specified property definition. If the specified property is not defined for the column, a message is sent to the log stating so.

`col<<Get Range Check`

Returns the range check definition. If range check is not defined for the column, a message is sent to the log stating so.

`col<<Get Role`

Returns the preselected role of *col*.

`col<<Get Script`

Returns the script to reproduce the column.

`col<<Get Scroll Locked`

Returns 1 if the column is scroll locked.

`col<<Get Selected`

Returns 1 if the column is selected, or 0 otherwise.

`col<<Get Stored Values`

Returns the values in the columns without considering the Missing Value Codes column property.

`col<<Get Value Labels`

Returns the value labels definition. If value labels is not defined for the column, a message is sent to the log stating so.

`col<<Get Use Value Labels`

Returns 1 if the value labels are set to be used for the column, or 0 otherwise.

`col<<Get Values`

Returns the values in the column.

`col<<Hide(Boolean)`

Turns the Hide attribute on or off according to the Boolean argument given.

`col<<Ignore Errors`

Ignores formula evaluation errors in a column, and sets the cell value to missing when a formula error occurs.

`col<<Input Format(`*`format`*`)`

Sets the quoted *`format`* used for input and storage for the column. The argument is the name of any JMP format (for example, `"ddmmyyyy"` for a date column).

`date_col<<Is Transformed On SAS Export`

Returns true if the data in the resulting SAS data set for the date column will be changed when it is exported to SAS.

`col<<Label(`*`Boolean`*`)`

Turns the Label attribute on or off according to the Boolean argument given.

`col<<Lock(`*`Boolean`*`)`

`col<<Set Lock(`*`Boolean`*`)`

Turns the Lock attribute on or off according to the Boolean argument given.

`col<<Preselect Role(role)`

Preselects the specified *role* for the column. Choices are `"Y"`, `"X"`, `"Weight"`, `"Freq"`, and `"None"`, or `"No Role"`.

`col<<Reset Transform`

Removes the cached data for the transform column. Accessing column data rebuilds the caches. Use this option to reduce memory or to allow recalculation if the formula depends on external information.

`col<<Set Display Width(n)`

Sets the column display width to the *n* in pixels. If *n* is set to zero, the column display width automatically resizes.

`col<<Set Each Value(n)`

Sets all the values in the column to *n*.

`col<<Set Excluded`

Excludes the column.

`col<<Set Field Width(n)`

Sets the field width for the column to *n*.

`col<<Set Hidden`

Hides the column.

`col<<Set Labeled`

`col<<Set Labelled`

Uses the column's data values for labels.

`col<<Set Modeling Type(type)`

Sets the modeling type for the variable. Choices are `"Continuous"`, `"Ordinal"`, `"Nominal"`, `"None"`, `"Row State"`, `"Unstructured"`, `"Multiple Response"`, or `"Vector"`.

`col<<Set Name(name)`

Sets the name for the column. The *name* argument is quoted.

`col<<Set Property(name, expression)`

Sets the quoted property name to the *expression* given. You can set any standard column property by name or a user-specified property.

Examples

The following example adds the Value Colors column property to the sex column, with pink for females and blue for males.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column( "sex" ) << Set Property( "Value Colors", {"F" = 78, "M" = 69} );
```

The following example adds a custom column property named *Date recorded* to the height column.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Column( "height" ) << Set Property( "Date recorded", 05Jan1990 );
```

See Also

Scripting Guide

Using JMP

`col<<Set Scroll Locked(Boolean)`

Turns the Scroll Lock attribute on or off according to the Boolean argument given.

`col<<Set Selected(Boolean)`

Sets the column to be selected or not selected.

`col<<Set Use for Marker`

`col<<Use for Marker`

Uses the values in the column as markers in graphs. Designed to use with expression columns and character columns that have IDs. In the Big Class Families.jmp sample data table, the picture column is specified to use as markers in graphs. Not supported in Bubble Plot.

`col<<Set Values([matrix] or {list})`

`col<<Values([matrix] or {list})`

Sets values for the matrix (for numeric variables) or list (for character variables).

`col<<Suppress Eval(Boolean)`

Turns off automatic calculation of formulas for the column.

col<<Use For Marker(*Boolean*)

Uses the values in the column as markers in graphs or turns off the option. Designed to use with expression columns and character columns that have IDs. In the Big Class Families.jmp sample data table, the picture column is specified to use as markers in graphs.

Rows

row<<Colors(*n*)

Assigns the color *n* to the selected rows.

row<<Exclude(*Boolean*)

row<<Unexclude(*Boolean*)

Turns the excluded or unexcluded state on for the selected rows according to the Boolean argument given. Omit the argument to toggle the row state.

row<<Hide(*Boolean*)

row<<Unhide(*Boolean*)

Turns the Hide attribute on or off according to the Boolean argument given. Omit the argument to toggle the row state.

row<<Hide and Exclude

Shows or hides the selected rows from appearing on graphs, and excludes or unexcludes them from contributing to calculations.

row<<Label(*Boolean*)

row<<Unlabel(*Boolean*)

Turns the Label attribute on or off according to the Boolean argument given. Omit the argument to toggle the row state.

row<<Markers(*marker*)

Assigns the quoted "*marker*" to the selected rows.

row<<Next Selected

Causes the next selected row in the data table to blink.

`row<<Previous Selected`

Causes the previous selected row in the data table to blink.

`row<<Row Editor`

Opens the Row Editor window for the selected rows.

Data Filter

`dtf<<Add Favorites(`*`name`*`)`

Description

Associates the current filter selection with the quoted *"name"* and saves it in the Favorites list.

Returns

The favorite as a quoted string.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
df = dt << Data Filter(
   Add Filter(
      Columns( :age, :sex, :height, :weight ),
      Where( :sex == "F" ),
      Where( :height >= 55 & :height <= 65 )
   ),
   Mode( Select (1) )
);
Wait( 1 ); // for demonstration purposes
fav1 = df << Add Favorites( "Female Average Ht" );
```

`dtf<<Add Filter(Columns(`*`column1`*`, <`*`column2`*`>), <Where(`*`clause`*`)>)`

Add one or more filter columns in a new OR group.

`dtf<<Auto Clear(`*`Boolean`*`)`

Clears all currently selected rows before setting a new selection.

`dtf<<Clear`

Clears the currently selected rows.

`dtf<<Close`

Closes the data filter window.

`dtf<<Columns(column1, column2, ...)`

Sets the columns to use in the data filter.

`dtf<<Data Table Window`

Shows the data table that the data filter window is using.

`dtf<<Delete All`

Removes all filters that are set.

`dtf<<Delete(column1, column2, ...)`

Removes the specified columns from the data filter.

`dtf<<Display(column, <Size(x, y)>, "Blocks Display"|"List Display"|"Single Category Display"|"Checkbox Display")`

Sets how the specified categorical column levels are displayed in the filter.

`dtf<<Get Script`

Returns the data filter script as text in the log.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
df = dt << Data Filter(
     Add Filter( Columns( :age, :sex ), Where( :age == 12 ) )
);
txt = df << Get Script;
Show( txt );
```

`dtf<<Local Data Filter`

Embeds the data filter in the specified window. See *Scripting Guide* for more information about local data filters.

`dtf<<Location(x, y)`

Moves the data filter window to the specified location. *x* and *y* are measured in pixels. 0,0 is the top left of the monitor.

dtf<<Make Filter Change Handler(*function*)

Creates a data filter handler to handle notification that the filter has changed. The number of rows filtered is returned in the argument to the function.

Example

```
dt = Open( "$SAMPLE_DATA/PopAgeGroupSubset.jmp" );
dist = Distribution( Automatic Recalc( 1 ), Continuous Distribution( Column(
   :POP ) ) );
filter = dist << Local Data Filter( Add Filter( Columns( :Region ) ) );
f = Function( {a}, Print( a ) );
rs = filter << Make Filter Change Handler( f );
```

dtf<<Make Subset

Creates a new subset data table that contains the rows that are selected in the data filter.

dtf<<Match(Filter Columns(*column1*, column2, ...), Where(*clause*)

Sets the filter conditions for each column. The *Where clause* is used for all the columns listed. To use different Where clauses for different columns, send the Match message separately for each column.

dtf<<Mode(Select(*Boolean*)|Show(*Boolean*)|Include(*Boolean*))

Sets the action, or mode, that is used when rows are selected using the data filter.

dtf<<Save and Restore Current Row States

Saves the current row states for the data table, and then restores those states when the data filter is closed.

dtf<<Show Columns Selector(*Boolean*)

Displays or hides the column selector after completing a filter.

dtf<<To Clipboard

Creates a Where clause from the current state of the data filter and places it on the clipboard, where it can be pasted elsewhere.

dtf<<To Data Table

Creates a Where clause from the current state of the data filter and saves it as a property to the data table.

`dtf<<To Journal`

Creates a Where clause from the current state of the data filter and appends it to the current journal. If there is no current journal, a new journal is opened and the Where clause is added to it.

`dtf<<To Row State Column`

Creates a row state column whose formula is the Where clause.

`dtf<<To Script Window`

Creates a Where clause from the current state of the data filter and appends it to the current script window. If there is no current script window, a new script window is opened and the Where clause is added to it.

`dtf<<Use Floating Window(`*`Boolean`*`)`

Sets whether the data filter window floats on top of its associated data table or behaves as a normal window.

`dtf<<Where(`*`clause`*`)`

Sets a condition for selecting rows.

Data Feed (Windows Only)

`feed<<Close`

Closes the data feed object and its window.

`feed<<Connect(`*`port settings`*`)`

Sets up port settings for the connection to the device.

`feed<<Disconnect`

Disconnects the device from the data feed queue but leaves the data feed object active.

`feed<<EOL("CR", "LF", "CRLF")`

Sets the line ending value used as a separator when parsing incoming lines of data. The value is also used as the terminator in outgoing lines of data.

– `"CR"`: ASCII character 13 (carriage return)

- "LF": ASCII character 10 (line feed)
- "CRLF": Uses both CR and LF in sequence.

`feed<<Get Line`

Returns and removes one line from the data feed queue.

`feed<<Get Lines`

Returns as a list and removes all lines from the data feed queue.

`feed<<Print Queue`

Prints the internal queue of messages to the log window.

`feed<<Queue Line(`*`quoted string`*`)`

Sends one quoted *`string`* (or *line*) to the end of the data feed queue. `Queue Line` is primarily useful for testing your script without requiring it to be attached to a device. You can essentially simulate the data coming from the device to make sure the rest of your code handles the values properly when it's really attached to a working device.

`feed<<Restart`

Restarts processing queued lines.

`feed<<Set Script(`*`script`*`)`

Assigns the *`script`* that is run each time a line of data is received.

`feed<<Stop`

Stops processing queued lines.

`feed<<Write(`*`quoted string`*`)`

Description

Sends a quoted *`string`* to the data feed device.

Example

```
exfeed = Open Datafeed(
   Connect( Port( "com1" ), Baud rate( 4800 ), Parity( "even" ), DataBits( 8 )
   ),
   Set Script(
      ex = exfeed << Get Line;
      Show( ex );
```

```
    )
);
exfeed << Write( "Ready" );
/* Example - send a message to external device over the serial port to trigger
   data messages. This can be used to send control messages to a sensor or
   other attached device.*/
```

feed<<Write Line(*quoted string*)

Description

Sends a quoted *string* to the data feed device. If EOL has been set for the data feed, the quoted strings are terminated by the specified EOL value. If EOL has not been set, the line is terminated with CRLF.

Example

```
exfeed = Open Datafeed(
   Connect( Port( "com1" ), Baud rate( 4800 ), Parity( "even" ), DataBits( 8 )
   ),
   Set Script(
      ex = exfeed << Get Line;
      Show( ex );
   )
);
exfeed << Write Line( "Ready" );
/* Example - send a message to external device over the serial port to trigger
   data messages. This can be used to send control messages to a sensor or
   other attached device.*/
```

feed<<Write Lines({*quoted string1, quoted string2, quoted string3*})

Description

Sends a list of quoted *strings* to the data feed device. If EOL has been set for the data feed, the quoted strings are terminated by the specified EOL value. If EOL has not been set, the line is terminated with CRLF.

Example

```
exfeed = Open Datafeed(
   Connect( Port( "com1" ), Baud rate( 4800 ), Parity( "even" ), DataBits( 8 )
   ),
   Set Script(
      ex = exfeed << Get Line;
      Show( ex );
   )
);
exfeed << Write Lines( {"Ready", "Set", "Go"} );
```

```
/* Example - send a message to external device over the serial port to trigger
   data messages. This can be used to send control messages to a sensor or
   other attached device.*/
```

Display Boxes

For additional examples, see the JMP Scripting Index.

All Display Boxes

---

db<<Add Text Annotation(Text(*quoted string*), Text Box(<*x1, y1, x2, y2*>))

Draws a text annotation box at the specified pixel location that contains the quoted *string*. The Text Box argument controls where the text annotation box is drawn in the window, from the upper left corner to the lower right corner.

Note that *x1*, *y1*, *x2*, and *y2* are not graph axis values but the specific pixel locations in the window. Exactly where the text box appears depends on the user's window size, display resolution, and so on.

---

db<<Append(*db2*)

Adds *db2* as the last child of the *db*.

---

db<<Child

Returns the child of the box.

---

db<<Class Name

Returns the name of the display class for the box.

---

db<<Clone Box

Makes a new copy of the display box.

---

db<<Close Window

Closes the containing window.

---

db<<Copy Picture

Puts a picture of the box on the clipboard.

db<<Delete

Deletes the display box.

db<<Enable(*Boolean*)

Controls the ability to interact with the display box. 0 disables the display box. 1 enables the display box.

db<<Get HTML

Returns a quoted string containing HTML source for the box.

db<<Get Journal

Returns a quoted string containing journal source for the box.

db<<Get Menu Item State(*index*)

Returns the popup menu item state of the index menu item. The state can be normal (0), checked (1), or disabled (-1).

db<<Get Menu Items

Description

Returns the menu items used for popup menu when the button is clicked. Menu items are returned in a list.

See Also

For submenus see "db<<Get Submenu(index)".

db<<Get Menu Script

Returns the menu script attached to the calling object.

db<<Get Page Setup()

Returns the page setup settings.

Example

The example below creates a new window and returns the page setup configuration.

```
w = New Window( "Window",
   Text Box( "Page Setup Test" )
);
w << Get Page Setup();
```

The results of the message:

```
{Margins( {0.75, 0.75, 0.75, 0.75} ), Scale( 1 ), Portrait( 1 ),
Paper Size( "Letter" )}
```

`db<<Get Picture( <Scale(n)> )`

Captures *db* as a picture object. The `Scale(n)` argument is a factor of the original picture size. For example, `Scale(2)` makes the picture object twice as large.

`db<<Get RTF`

Returns a quoted string containing RTF source for the box.

`db<<Get Script`

Returns the script for recreating the display box.

`db<<Get Size`

Returns either { *x, y* } or { *h, v* } in pixels:

```
xy = DisplayBox << Get Size;
```

Returns x and y in pixels:

```
{ x, y } = DisplayBox << Get Size;
```

`db<<Get Submenu(index)`

Returns the number of submenu items under the given menu item.

Example

The example below creates a menu containing *"A"*, *"B"*, and *"C"* with *"A"* having a submenu *"A1"* and *"A2"* and *"B"* having a submenu *"B1"*, *"B2"*, and *"B3"*. `<<Get Submenu(inc)` returns the number of submenu items under each indexed menu item.

```
New Window( "Title",
obj = Outline Box( "title" ) );
submenus = { };
obj << Set Menu Script(
   {"A", "", "A1", Print( "A1" ), "A2", Print( "A2" ),
   "B", "", "B1", Print( "B1" ), "B2", Print( "B2" ), "B3", Print( "B3" ),
   "C", Print( "C" )}
);
obj << Set Submenu( 1, 2 ); // menu A with 2 items in submenu A1 and A2
obj << Set Submenu( 4, 3 ); // menu B with 3 items in submenu B1, B2, and B3
For( inc = 1, inc <= N Items( Words( obj << Get Menu Script, "," ) )/2, inc++,
   Insert Into( submenus, obj << Get Submenu( inc ) );
);
submenus;
   {2, 0, 0, 3, 0, 0, 0, 0}
```

The log output indicates that index(1) contains two submenu items and index(3) contains three submenu items.

`db<<Get Text`

Returns a quoted string containing the text of the box.

`db<<Horizontal Alignment(`*`position`*`)`

Aligns a child display box inside the display parent box according to the specified *`position`*. The default value is `"Left"`, or you can specify `"Center"`, or `"Right"`.

Example

```
New Window( "Example",
   Outline Box( "Parent display box",
      Button Box( "OK", <<Horizontal Alignment( "Center" ) )
   )
);
```

`db<<Inval`

Invalidates the display box area in the window. The window is updated the next time the operating system has an opportunity to update windows (for example, when the user resizes the display box).

Notes

Consider including the message `<<Update Window` rather than including `Wait(0)`. The problem with using `Wait(`*`n`*`)` is knowing how large *`n`* should be.

Many display box messages, such as `<<Set Text`, automatically mark the box as invalid, so the `<<Inval` message is usually unnecessary. Some interactive scripts that use sliders with JSL callbacks might need `<<Update Window` to keep various parts of the display synchronized with the slider.

`db<<Is Enabled`

Returns the enabled state of the control. The message is supported in `Busy Light Box()`, `Button Box()`, `Calendar Box()`, `Check Box()`, `Col List Box()`, `Combo Box()`, `Completion Box()`, `Filter Col Selector()`, `gtext()`, `List Box()`, `Number Edit Box()`, `Popup Box()`, `Radio Box()`, `Range Slider Box()`, `Slider Box()`, `Spin Box()`, `Text Edit Box()`, `Tree Box()`, `Tree Map Box()`, and `Tree Map Seg()`.

`db<<Journal`

Appends the box to the journal.

db<<Journal Window

Appends the containing window of the display box to the journal; compare with Journal.

db<<Move Window(*x*, *y*)

Moves the window to the (*x*, *y*) location on your screen.

db<<Page Break

Inserts a page break before the box.

db<<Parent

Returns the parent of this display box.

db<<Prepend(*db2*)

Add *db2* to the display tree before *db*.

db<<Prev Sib

Returns the previous sibling of the display box.

db<<Reshow

Invalidates the display box's area in the window and immediate removes invalid areas from the window.

db<<Save Capture(<*path*>, <*format*>, <Add Sibling(*n*)>)

Saves the display box as a graphic to the specified quoted *path* in the specified quoted *format*. The optional Add Sibling argument adds the number of sibling display boxes to include in the capture. The default value is 1, which captures only the specified display box. Note that the specified portion of the report is not guaranteed to be scrolled into view or unobstructed by other windows. If the display box is not visible, the saved graphic will not contain the contents that you expect.

If you omit the path, you are prompted to name and save the file when running the path.

db<<Save HTML(<*path*>, <*format*>)

Saves the HTML source and folder of graphics to the quoted *path* and in the quoted *format*. If you omit the *path* argument, you are prompted to name and save the file when running the script.

db<<Save Interactive HTML(*<path>*, "Is Static")

Saves the display box as a web page (that includes interactive HTML features) in the quoted *path*. Non-JMP users can then explore the data. Note that the data is embedded in the web page.

Arguments

path A optional quoted *path* that specifies the location where the web page will be saved.

"Is Static" Omits the data from the web page and saves a static version of the web page.

Examples

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( y( weight ), x( height ) );
rbiv = (biv << Report);
rbiv << Save Interactive HTML( "$DOCUMENTS/MyInteractiveHTML.htm" );
```

db<<Save Journal(*<path>*)

Saves the journal source for the box in the quoted *path*. If you omit the argument, you are prompted to name and specify the graphic type.

db<<Save MSWord(*<path>*)

(Windows Only) Saves the display box as a Microsoft Word document in the quoted *path*. If you omit the *path* argument, you are prompted to name and save the file when running the script.

db<<Save PDF(*<path>*, <Show Page Setup(*Boolean*)>, <Portrait(*Boolean*)>)

Description

Saves a PDF of the display box in the quoted *path*.

Optional Arguments

path Saves the file in the quoted *path*. If you omit the argument, you are prompted to name and save the file when running the script.

Show Page Setup(*Boolean*) (Windows only) Displays the Page Setup window, where you can specify page orientation, headers and footers, margins, page scale, and paper size.

Portrait(*Boolean*) Displays the content in portrait or landscape orientation.

Notes

The PDF file contains headers and footers. Use Save Picture to omit these components.

`db<<Save Picture(<path>, <format>)`

Description

Saves a picture of the display box in the quoted *path* and with the specified quoted format.

Notes

- If you omit the quoted *path* argument, you are prompted to name and save the file when running the script.
- Valid file formats include "PDF", "PNG", "GIF", "JPG" or "JPEG", "EPS", "SVG", and "EMF".
- On Windows, the Windows Specific preferences determine the resolution (or DPI), or you can run the following script:

  `Pref( Save Image DPI( number ) );`
- On macOS, the operating system determines the DPI.
- Use `Save Picture` to export a report as a PDF file with no headers or footers. Use `Save PDF` to include these components.

`db<<Save Presentation(<path>, <Template(path)>, <Insert("Begin"|"End"|n)|Replace("Begin"|"End"|n)|Append>, <Outline Titles(title location)>, <format>)`

Saves display boxes in a Microsoft PowerPoint presentation. You can open the file in any presentation software program.

Optional Arguments

`path` Saves the file in the quoted *path*. You must include the .pptx extension in the filename. If you omit the *path* argument, you are prompted to name and save the file when running the script.

`Template(path)` Specifies the quoted *path* of a custom PowerPoint template. Without this argument, JMP uses the default template located in the `pptx` folder of the installation directory.

Include a simple table in your template, or a default table format is applied to report tables. For an example on Windows, see `/pptx/JMPExportTemplate.pptx` in the JMP installation folder.

`Insert` Determines where the slides are inserted in an existing presentation.

– *n* inserts the slides as the *n*th slide number.

– "Begin" inserts the slides at the beginning of the presentation.

– "End" inserts the slides at the end of the presentation.

`Replace` Determines which slides are replaced in an existing presentation. The arguments are *n*, "Begin", and "End" as described for `Insert`.

`Append` The slides are inserted at the end of an existing presentation.

`Outline Titles` The location of the outline title and any parent outline titles on the slide. By default, the immediate parent outline title appears as a slide title above the slide content, with any parent outline titles positioned in the bottom left corner of the slide.

- `"None"` omits the slide title above the graphic and the outline titles.
- `"Hide"` omits the outline titles.
- `"TopLeft"`, `"TopRight"`, `"BottomLeft"`, `"BottomRight"` determine the position of any of the parent outline titles on the slide.

`format` The format of the embedded graphics. Options are `"Native"`, `"EMF"`, `"PNG"`, `"JPG"`, `"BMP"`, `"GIF"`, `"TIF"`. On Windows, the native format is EMF. On macOS, the native format is PDF. See "Notes" for compatibility issues. Without this argument, JMP applies the "Image Format for PowerPoint" General preference.

Notes

Windows does not support the native PDF graphics produced on macOS. macOS does not support the native EMF graphics produced on Windows. For cross-platform compatibility, specify `"PNG"`, `"JPG"`, `"GIF"`, or `"TIF"`.

If no arguments are provided, the user is prompted to name and save the file.

`db<<Save RTF(<path>, format)`

Saves the file in the specified `quoted path` and with the quoted `format`. If you omit the `path` argument, you are prompted to name and save the file when running the script.

`db<<Save Text(<path>, format)`

Saves a file containing the text of the box in the quoted `path` and with the specified quoted `format`. If you omit the `path` argument, you are prompted to name and save the file when running the script.

`db<<Scroll Window(Display Box|relative-vertical-pixels|relative-horizontal-pixels, relative-vertical-pixels|{absolute-vertical-pixels, absolute-horizontal-pixels})`

Scrolls the containing window.

`db<<Select`

`db<<Deselect`

Selects (highlights) or deselects the box.

db<<Set Menu Item State(*index*, 0|1|-1)

Sets the popup menu item at *index* to be normal (0), selected (1), or disabled (-1).

db<<Set Page Setup<Margins(*left*, *right*, *top*, *bottom*)>, <Scale(*s*)>,<Portrait(*Boolean*)>, <Paper Size(*paper size*)>)

db<<Set Page Setup<Margins({*left*, *right*, *top*, *bottom*})>, <Scale(*s*)>,<Portrait(*Boolean*)>, <Paper Size(*paper size*)>)

Sets the page settings. Margins are set in inches. Scale variable *s* is a number in the range of 10 (for 1000%) to 0.2 (for 20%) with the default as 1 (for 100%). If Portrait is True the page is oriented for portrait, otherwise the page is landscape. Paper Size is a quoted string specifying the paper size, for example, "Letter" or "Legal".

Example

The example below creates a new window and configures the page setup.

```
w = New Window( "Window",
   Text Box( "Page Setup Test" )
);
w << Set page setup(
   margins( 1, 1, 1, 1 ),
   scale( 1 ),
   portrait( 1 ),
   paper size( "Letter" )
);
```

db<<Set Print Headers(*left header*, *center header*, *right header*)

Description

Sets the left, center, and right header for print output.

Example

```
w = New Window( "Window", Text Box( "Header Example" ) );
w << Set Print Headers(
   "Today is: &d;", // left
   "&wt;", // center
   "Page &pn; of &pc;" // right
);
w << Print Window;
```

db<<Set Print Footers(*left footer*, *center footer*, *right footer*)

Description

Sets the left, center, and right footer for print output.

Example

```
w = New Window( "Window", Text Box( "Footer Example" ) );
```

```
w << Set Print Footers(
   "Today is: &d;", // left
   "&wt;", // center
   "Page &pn; of &pc;" // right
);
w << Print Window;
```

db<<Set Stretch(x,y)

Description

Sets the horizontal and vertical stretching of display boxes.

Arguments

"Window" The box stretches to fill the space provided by the window, The sizing depends on the minimum and maximum properties for the box.

"Fill" The box stretches to fill the available space in its container.

"Off" The box does not stretch.

"Neutral" Many boxes start in the neutral state. A leaf box (a box that has no children) that is neutral resolves to "Off". Most container boxes aggregate the stretching behavior (and minimum and maximum) of their child boxes.

db<<Set Submenu (*index*, *submenu count*)

Description

Sets the submenu items for the item (specified by index number) by specifying the number of items in the submenu.

Example

The example below creates a menu containing “A”, “B”, and “C” with “A” having a submenu “A1” and “A2” and “B” having a submenu “B1”, “B2”, and “B3”.

```
New Window( "title", ob = Outline Box( "title" ) );
ob << Set Menu Script(
   {"A", "", "A1", Print( "A1" ), "A2", Print( "A2" ),
   "B", "", "B1", Print( "B1" ), "B2", Print( "B2" ), "B3", Print( "B3" ),
   "C", Print( "C" )}
);
ob << Set Submenu(1, 2); // menu A with 2 items in submenu A1 and A2
ob << Set Submenu(4, 3); // menu B with 3 items in submenu B1, B2, and B3
```

db<<Set Report Title(*title*)

Sets a new *title*. The *title* is quoted.

Show Properties(*db*)

Shows the messages a given display box can interpret.

db<<Sib

Returns the sibling of the display box.

db<<Sib Append(*db2*)

Appends a display as a sibling to this one. The argument must evaluate to a display box owner or reference.

db<<Size Window(*x*, *y*)

Resizes the containing window.

db<<Update Window

Updates the window that holds the display box (and possibly other windows as well, depending on the operating system) if there are invalidated regions. Previously invalidated box areas are redrawn with their new content.

Notes

In some interactive JSL scripts that combine sliders with JSL callbacks, you might need to use <<Update Window to keep parts of the display synchronized with the slider.

db<<Zoom Window

Resizes the window to be large enough to show all of its contents.

Axis Boxes

Axis Box<<Axis Settings(*<named arguments>*)

Opens the Axis Specification window or specifies axis settings, such as tick marks and axis labels.

If no arguments are included, the axis specification window appears.

Otherwise, specify named arguments for each axis.

- Specify the Y axis as Axis Box(1).
- Specify the X axis as Axis Box(2).

Optional Named Arguments

All Axes

```
Scale("Linear"|"Log"|"Power"|"Geodesic"|"Geodesic US"|"Custom
  Scale"|"Normal Probability|Weibull Probability|Frechet
  Probability|Logistic Probability|Exponential Probability|Gamma
```

Probability|Beta Probability|Mixture of 2 Normals Probabilities|Mixture of 3 Normals Probabilities) Specifies the scale of the axis. If the type is Custom Scale, this message expects two additional named arguments: Scale to Internal(*expr*) and Scale to External(*expr*).

Min(*n*) Changes the minimum value on the axis.

Max(*n*) Changes the maximum value on the axis.

Reverse Order(*Boolean*) Reverses the axes by reversing the minimum and maximum values.

Inc(*n*) Shows the numbers at the specified increments.

Set Font(*font*) Specifies the quoted *font* that is applied to the numbers. The JMP Font preferences determine the default font.

Set Font Size(*points*) Specifies the size of the font that is applied to the numbers. The JMP Font preferences determine the default font.

Set Font Style("Strikeout"|"Underline") Specifies the quoted style that is applied to the numbers.

Automatic Font Size(*Boolean*) JMP attempts to decrease the font size (down to a certain minimum) if all of the labels cannot fit at the default size. If 0, the font size is not decreased.

Automatic Tick Marks(*Boolean*) Turns on tick marks only if one or more labels are hidden (due to insufficient space).

Label Orientation("Automatic"|"Horizontal"|"Vertical"|"Perpendicular"|"Parallel"|"Angled") Rotates the axis label. The default value is "Automatic", which is based on the width of the labels.

Lower Frame(*Boolean*) Shows a frame below the labels. The default value is off.

Value Labels Displays the label that you specify instead of the data value.

Inside Ticks(*Boolean*) Shows tick marks inside or outside of the axis.

Add Ref Line({Label Row Nesting(*n*), *begin range*, <*end range*>, <"Solid"|"Dotted"|"Dashed"|"DashDot"|"DashDotDot">, <*color*>, <*label*>, <width*(n)*>, <*opacity*(%)>}) Defines the reference line range, line pattern, color, label, width, and opacity. A solid, black, 1-pixel line is the default setting. Label Row Nesting(*n*) specifies the number of nested rows on the axis. The *color* and *label* arguments are quoted.

Categorical Axes

Wrap Lines(*n*) Wraps long labels across multiple lines (*n*).

Numeric Axes

Format(*arguments*) Specifies the format of the numeric axis data. See the Format list in a numeric column's column properties for arguments. If you specify the a datetime

format, also include the `Interval` argument: `"Numeric"`, `"Year"`, `"Quarter"`, `"Month"`, `"Week"`, `"Day"`, `"Hour"`, `"Minute"`, or `"Second"`.

`Minor Ticks(`*`number`*`)` Specifies the *number* of minor tick marks between major tick marks.

`Tick Offset(`*`number`*`)` Specifies the starting point of the tick marks.

`Major Ticks(`*`Boolean`*`)` Shows or hides a major tick mark between each number.

`Minor Ticks(`*`Boolean`*`)` Shows or hides a minor tick mark between each number.

`Show Major Grid(`*`Boolean`*`)` Shows or hides a grid line at each major tick mark.

`Show Minor Grid(`*`Boolean`*`)` Shows or hides a grid line at each minor tick mark.

`Major Grid Line Color(`*`color`*`)` Sets the color for the major grid (if enabled) using the quoted *color*.

`Minor Grid Line Color(`*`color`*`)` Specifies the color of the grid line at each minor tick mark.

Example

The following example creates a bivariate plot and defines basic settings for the X and Y axes.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = dt << Bivariate( X( :height ), Y( :weight ), FitLine );
rbiv = biv << Report;
xaxis = rbiv[Axis Box( 2 )];
yaxis = rbiv[Axis Box( 1 )];
xaxis << Axis Settings( Show Major Grid( 1 ) );
yaxis << Axis Settings( Decimal( 10, 3 ) );
```

`Axis Box<<Add Axis Label(`*`quoted string`*`)`

Adds an axis label with the specified quoted string.

`Axis Box<<Add Ref Line(`*`number`*`, `*`linestyle`*`, <`*`color`*`>, <`*`label`*`>, <`*`width`*`>)`

Adds a reference line at *number* in the specified *linestyle* (`"Solid"|"Dashed"|"Double"`), *color* (using the quoted *color*), quoted *label*, and *width* (in pixels).

`Axis Box<<Decimal(`*`width`*`, `*`decimal places`*`)`

Changes the numeric format for axis values.

`Axis Box<<Format(`*`name`*`)`

Changes to the numeric format given by the quoted *name*.

`Axis Box<<Get Inc(`*n*`)`

Gets the increment value of the axis.

`Axis Box<<Inc(`*n*`)`

Sets the increment between ticks.

`Axis Box<<Interval(`*format*`)`

Specifies the units used for `Inc()` with date/time formats: `"Numeric"`, `"Year"`, `"Quarter"`, `"Month"`, `"Week"`, `"Day"`, `"Hour"`, `"Minute"`, or `"Second"`.

`Axis Box<<Label Orientation(`*format*`)`

Rotates the axis label to one of the following formats: `"Automatic"` (the default setting based on the width of the label), `"Horizonal"`, `"Vertical"`, `"Perpendicular"`, `"Parallel"`, and `"Angled"`.

`Axis Box<<Major Grid Line Color(`*color*`)`

Sets the color for the major grid (if enabled) using the quoted *color*.

`Axis Box<<Max(`*maximum*`)`

Changes the maximum value on the axis.

`Axis Box<<Minor Grid Line Color(`*color*`)`

Sets the color for the minor grid (if enabled) using the quoted *color*.

`Axis Box<<Min(`*minimum*`)`

Changes the minimum value on the axis.

`Axis Box<<Minor Ticks(`*number*`)`

Specifies the *number* of minor tick marks between major tick marks.

`Axis Box<<Remove Axis Label`

Removes any label added with `Add Axis Label`.

Axis Box<<Reverse Scale(*Boolean*)

Reverses the normal scale direction so that the highest value is on the left or bottom (that is, closest to the origin).

Axis Box<<Revert Axis

Restores the axis' original settings (from time of creation).

Axis Box<<Scale(*type*)

Changes the scale of the axis to *type* ("Linear"|"Log"|"Exp Prob"|"Weibull Prob"|"Logistic Prob"|"Frechet Prob"|"Normal"|"Cube Root"|"Johnson Su Scale"|"Geodesic"|"Geodesic US"|"Custom Scale"|"Power"|"Gamma Prob"|"Beta Prob"|"Mixture of 2 Normals Prob"|"Mixture of 3 Normals Prob").

If the type is Custom Scale, this message expects two additional named arguments: Scale to Internal(expr) and Scale to External(expr).

Axis Box<<Tick Font(*name*, *<size>*, *<style|style style...>*, *<angle>*)

Sets the font name (quoted), size, and quoted properties for tick marks. To specify more than one style, include a space between each style and place them in quotes.

Axis Box<<Show Labels(*Boolean*)

Shows or hides labels for the axis values.

Axis Box<<Show Major Grid(*Boolean*)

Adds or removes grid lines at the major tick values.

Axis Box<<Show Major Ticks(*Boolean*)

Shows or hides major tick marks.

Axis Box<<Show Minor Grid(*Boolean*)

Adds or removes grid lines at the minor tick values.

Axis Box<<Show Minor Ticks(*Boolean*)

Shows or hides minor tick marks.

Axis Box<<Tick Label List(*<i>*, *{text1,text2, ...}*,*<{n1, n2, ...}>*)

Sets the values and positions of the axis tick labels.

Note: Major tick increments are automatically set to 1.0 if the tick labels are not specified.

Required Arguments

`{text1,text2, ...}` Specifies the quoted string titles for your labels.

Optional Arguments

i Specifies the label row index. Leaving it out clears any existing label rows and creates one new one as specified. Including it allows you to override any particular label row; using an index higher than the current number of label rows adds a new label row on to the end.

`{n1, n2, ...}` Specifies the values corresponding to each label. If the value list is omitted, the labels will be on integer increments starting with 1.

Border Boxes

Note: Border boxes support only one display box argument.

`Border Box<<Set Background Color({r, g, b}|<color>)`

Sets the background color for a border box. Specify an optional quoted list of RGB values or *color*. For example:

```
border box<<Set Background Color("red");
```

or

```
border box<<Set Background Color( {255, 192, 3} );
```

`Border Box<<Set Color({r, g, b}|<color>)`

Sets the border color for a border box. Specify a list of RGB values or a quoted `color`. For example:

```
border box<<Set Color("red");
```

`Border Box<<Get Color`

Gets the border color for a border box.

`Border Box<<Set Style(style)`

Sets the border style for a border box. Specify the *style* as one of the following numbers or keywords: `0` (`"Solid"`), `1` (`"Dotted"`), `2` (`"Dashed"`), `3` (`"DashDot"`), or `4` (`"DashDotDot"`). For example:

```
border box<<Set Style("Dotted");
```

Border Box<<Get Style

Gets the border style for a border box.

Data Browser Boxes

dbb<<Set Data Table(*<data table>*)

Sets the data table for the data browser box.

Data Filter Source Boxes

dfsb<<Set Row States(*dt*, *rs*)

Sets the row states for the given data table within the filter. Selections made in this row state will not be linked with the data table, but will be included in the reports linked to the selection filter.

Frame Boxes

Frame Box<<Add Graphics Script(*<order>*,*<description>*, *<script>*)

Description

Adds a script to draw graphics in the frame box.

Optional Arguments

order Specifies the order in which the graphics elements are drawn. The value can be the keyword "Back" or "Forward" or an integer that specifies the drawing order for a number of graphics element. 1 means the object is drawn first.

description A quoted string that appears in the Customize Graph window next to the graphics script. The *description* argument is quoted.

script A JSL script.

Example

In the following example, the graphics script draws the line first and then draws the other graphics elements: the grid lines, references lines, and markers that create the bivariate plot. Without the 1 order argument, the line is drawn last and covers up the markers.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
obj = dt << Bivariate( Y( :weight ), X( :height ) );
Report( obj )[FrameBox( 1 )] <<
Add Graphics Script(
  1, // draws the line first
  Description( "Pen Script" ),
```

```
    Pen Color( "red" );
    Pen Size( 5 );
    Y Function( 60 + 120 / 2 * (1 + Sine( (2 * Pi() * (x - 50)) / 22.5 )), x );
);
```

`Frame Box<<Append Seg`

Adds a display seg to the specified Frame Box.

`Frame Box<<Background Color({`*`RGB values`*`}|<`*`color`*`>)`

Changes the background color. Specify a list of RGB values or a quoted *color*.

`Frame Box<<Child Seg`

Returns the display seg child of the Frame Box.

`Frame Box<<Edit Graphics Script`

Brings up a dialog box to view, edit, or delete the current graphics scripts.

`Frame Box<<Find Seg`

Returns a display seg with the specified argument (for example, the name of a seg).

`Frame Box<<Frame Size(`*`x, y`*`)`

Resets the size of the frame, in pixel units.

`Frame Box<<Make Table of Graphs Like This`

Creates a data table of graphs.

`Frame Box<<Marker Size(`*`size`*`)`

Changes the marker size. The values are 0 (dot), 1 (small), 2 (medium), and so on.

`Frame Box<<Row Colors(`*`color`*`)`

`Frame Box<<Row Markers(`*`marker`*`)`

`Frame Box<<Row Exclude(`*`Boolean`*`)`

`Frame Box<<Row Hide(`*`Boolean`*`)`

`Frame Box<<Row Label(`*`Boolean`*`)`

Forwards commands to the data table associated with the report, so that the row states of selected rows can be manipulated. For `Row Exclude`, `Row Hide`, and `Row Label`, omitting

the argument toggles the option. If the option is off, the message turns it on. If the option is on, the message turns it off.

`frame box<<Set Background Fill(`*`Boolean`*`)`

Enables or disables filling the background with the background color. Use this option when you want to paste a graph and make the background transparent.

Example

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
biv = Bivariate( y( weight ), x( height ) );
rbiv = biv << Report;
framebox = rbiv[Frame Box( 1 )];
// set background color
framebox << Background Color( "red" );
// for demonstration purposes: wait to see the color change
Wait( 1 );
// turn off background fill color
framebox << Set Background Fill( 0 );
```

`frame box<<X Axis(<Min(`*`minimum`*`)>, <Max(`*`maximum`*`)>, <Inc(`*`n`*`)>, <`*`named arguments`*`>)`

Scales the X coordinate system.

`frame box<<Y Axis(<`*`Min(min)`*`>, <`*`Max(max)`*`>, <Inc(`*`n`*`)>, <`*`named arguments`*`>)`

Scales the Y coordinate system.

Display 3D Boxes

`Graph 3D Box()`

Sends display commands to the 3D plot.

Excerpt Boxes

`Excerpt Box(`*`rptnum`*`, `*`lstSubscripts`*`)`

Returns a display box containing the excerpt designated by the report held at number `rptnum` and the list of display subscripts `lstSubscripts`. The subscripts reflect the current state of the report after previous excerpts have been removed.

Filter Col Selector

`Filter Col Selector(<Data Table(name)>, <width(pixels)>, <nLines(n)>, <script>, <onChange(expr)>)`

Returns a display box that contains a list of items. The control supports column filtering.

Global Boxes

`Global Box(value)`

Creates a display box that shows the value of a global variable.

Hier Boxes

`Hier Box(title, Hier Box(...), Hier Box(...), ...)`

Returns a display box with the *title* (quoted) that contains a hierarchy of quoted strings.

Matrix Boxes

`Matrix Box<<Get`

Returns the matrix contents.

`Matrix Box<<Make Into Data Table(<Invisible(Boolean)|Private(Boolean)>)`

Description

Turns the matrix into a new data table. `Invisible(1)` hides the data table from view. An invisible data table can be open from the JMP Home Window or the Window menu. `Private(1)` opens the data table without displaying it in a data table window. A private data table is generally for scripts that want better control of the data table by not exposing it to general use.

Returns

A reference to the new data table.

Matrix Box<<Set Format(*<width>*, *<decimal places>*, *<*"Use Thousands Separator">)

Description

Sets the numeric format for matrix elements.

Arguments

A number of other formats can be set on matrix boxes. See "Number Col Box<<Set Format(<width>|<width, decimal places>, <"Use Thousands Separator">)" for more information about the syntax.

Matrix Box<<Sort(*column number*, *ascending*)

Sorts the rows of the matrix based on the column number specified by column_num. The default sort order is ascending.

If column number is 0, the sort is removed.

ascending is a Boolean value. If ascending is "True", the sort is performed in ascending order. If ascending is "False", the sort is in descending order.

Nom Axis Boxes

Nom Axis Box<<Divider Lines(*Boolean*)

Adds or removes divider lines between labels in the axis box.

Nom Axis Box<<Lower Frame(*Boolean*)

Adds or removes a lower frame around the axis.

Nom Axis Box<<Rotated Tick Labels(*Boolean*)

Rotates or unrotates the labels at each tick value.

Number Col Boxes

Number Col Box<<Add Element(*item*)

Adds the *item* to the Number Col Box. *item* can be a single number, a list of numbers, or a matrix.

`Number Col Box<<Bootstrap(`*`nsample`*`, Random Seed(`*`number`*`), Fractional Weights(`*`Boolean`*`), Split Selected Column(`*`Boolean`*`), Discard Stacked Table if Split Works(`*`Boolean`*`)`

Description

Bootstraps the analysis, repeating it many times with different resampling weights and collecting tables as selected.

Arguments

`nsample` Sets the number of times that you want to resample the data and compute the statistics. A larger number results in more precise estimates of the statistics' properties. By default, the number of bootstrap samples is set to 2,500.

`Random Seed(`*`number`*`)` Sets a random seed that you can re-enter in subsequent runs of the bootstrap analysis to duplicate your current results. By default, no seed is set.

`Fractional Weights(`*`Boolean`*`)` Performs a Bayesian bootstrap analysis. In each bootstrap iteration, each observation is assigned a weight that is calculated as described in *Basic Analysis*. The weighted observations are used in computing the statistics of interest. By default, the fractional weights option is not selected and a simple bootstrap analysis is conducted.

`Split Selected Column(`*`Boolean`*`)` Places bootstrap results for each statistic in the column that you selected for bootstrapping into a separate column in the Bootstrap Results table. Each row of the Bootstrap Results table (other than the first) corresponds to a single bootstrap sample.

If you exclude this option, a Stacked Bootstrap Results table appears. For each bootstrap iteration, this table contains results for the entire report table that contains the column that you selected for bootstrapping. Results for each row of the report table appear as rows in the Stacked Bootstrap Results table. Each column in the report table defines a column in the Stacked Bootstrap Results table.

`Discard Stacked Table if Split Works(`*`Boolean`*`)` (Applicable only if the Split Selected Column option is included.) Determines the number of results tables produced by Bootstrap. If the `Discard Stacked Table if Split Works` option is not selected, then two Bootstrap tables are shown. The Stacked Bootstrap Results table, which contains bootstrap results for each row of the table containing the column that you selected for bootstrapping, gives bootstrap results for every statistic in the report, where each column is defined by a statistic. The unstacked Bootstrap Results table, which is obtained by splitting the stacked table, provides results only for the column that is selected in the original report.

`Number Col Box<<Get`

`Number Col Box<<Get(`*`i`*`)`

Gets the values in a list, or the *i*th value.

Number Col Box<<Get As Matrix

Gets the values in a matrix, specifically a column vector.

Number Col Box<<Get Format

Returns the current format.

Number Col Box<<Get Heading

Returns the column heading text.

Number Col Box<<Remove Element(*row number*)

Removes an element from the column at the specified position.

Number Col Box<<Set Format(*<width>*|*<width, decimal places>*, *<*"Use Thousands Separator">)

Number Col Box<<Set Format("Best", *<width>*, <"Use Thousands Separator">)

Number Col Box<<Set Format(("Fixed Dec"|"Percent"), *<width>*|*<width, decimal places>*, <"Use Thousands Separator">)

Number Col Box<<Set Format("Pvalue", *<width>*)

Number Col Box<<Set Format(("Scientific"|"Engineering"|"Engineering SI"), *<width>*|*<width, decimal places>*)

Number Col Box<<Set Format("Precision", *<width>*|*<width, decimal places>*, <"Use Thousands Separator">, <"Keep Trailing Zeroes">, <"Keep All Whole Digits">)

Number Col Box<<Set Format("Currency", *<currency code>*, *<width>*|*<width, decimal places>*, <"Use Thousands Separator">)

Number Col Box<<Set Format(*datetime*, *<width>*, *<input format>*)

Number Col Box<<Set Format(("Latitude DDD"|"Latitude DDM"|"Latitude DMS"|"Longitude DDD"|"Longitude DDM"|"Longitude DDM"), *<width>*|*<width, decimal places>*, ("PUN"|"DIR"|"PUNDIR"))

Number Col Box<<Set Format("Custom", Formula(...), *<width>*, *<input format>*)

Description

Sets the column format.

Arguments

Using JMP describes the arguments. Note that `Matrix Box()`, `Number Col Box()`, `Number Col Edit Box()`, `Number Edit Box()` have the same `Set Format` syntax.

Examples

```
<<Set Format( 10, 2, "Use thousands separator");
<<Set Format( "Currency", "EUR", 20, );
<<Set Format( "m/d/y", 10 );
<<Set Format( "Precision", 10, 2, "Keep trailing zeroes", "Keep all whole
  digits" );
<<Set Format( "Latitude DDD", "PUNDIR"); // "PUN" for punctuation, "DIR" for
  direction, PUNDIR for both
<<Set Format( "Custom", Formula( Abs( value ) ), 15 );
```

Notes

- For a list of currency codes, see *Scripting Guide*. The currency code is based on the locale if the code is omitted.
- If you don't specify the format, set the decimal places to greater than 100 for datetime values and to 97 for *p*-values.
- You must always precede the number of decimal places with the width.
- Options can be defined in a list or a variable, or they can be in a `Function()` that is evaluated.

```
ncbFunc = Function({}, {"Fixed", 12, 5});
```

`number col box<<Set Heading(`*quoted string*`)`

Changes the column heading text.

Number Col Edit Boxes

`Number Col Edit Box<<Set Format(`*<width>*`, `*<decimal places>*`, <"Use Thousands Separator">|`*<other options>*`)`

Description

Sets the column format.

Arguments

A number of other formats can be set on number col edit boxes.

See Also

"Number Col Box<<Set Format(<width>|<width, decimal places>, <"Use Thousands Separator">)"

Number Col Edit Box<<Remove Element(*x position*, *y position*, *i*)

Removes an element from the column at the specified position.

Number Edit Box

Number Edit Box<<Set Format(*<width>*, *<decimal places>*, <"Use Thousands Separator">|*<other options>*)

Description

Sets the column format.

Arguments

A number of other formats can be set on number edit boxes.

See Also

"Number Col Box<<Set Format(<width>|<width, decimal places>, <"Use Thousands Separator">)"

Outline Boxes

Outline Box<<Close(*Boolean*)

Closes the outline box.

Outline Box<<Close All Below

Closes all the node's child nodes.

Outline Box<<Close All Like This

Closes all nodes similar to this outline box.

Outline Box<<Close Where No Outlines

Closes all nodes that do not have children.

Outline Box<<Get Title

Gets the title of the outline box.

Outline Box<<Horizontal(*Boolean*)

Horizontally aligns the node's children.

`Outline Box<<Open All Below`

Opens all the node's child nodes.

`Outline Box<<Open All Like This`

Opens all nodes similar to this outline box.

`Outline Box<<Set Menu Script(`*{quoted string1, script1, quoted string2, script2, ...}*`)`

Adds an entry to the menu when the red triangle on an outline box is selected.

`Outline Box<<Set Title(`*title*`)`

Specifies the quoted *title* of the outline box.

Panel Boxes

`Panel Box<<Get Title`

Gets the title of the panel box.

`Panel Box<<Set Title(`*title*`)`

Specifies the *title* (quoted) of the panel box.

Plot Col Boxes

`Plot Col Box<<Get As Matrix`

Gets the values in a matrix, specifically a column vector.

`Plot Col Box<<Get Labels`

Gets the labels for each row.

`Plot Col Box<<Remove Element(`*row number*`)`

Removes an element from the column at the specified position.

`Plot Col Box<<Set Labels(`*{list}*`)`

Sets the labels for each row.

`Plot Col Box<<Set Scale(`*minimum*`, `*maximum*`, <"format">, <`*width*`>, <`*decimal places*`>, <"Use Thousands Separator">)`

Description

Specifies the minimum and maximum values of the horizontal axis.

Arguments

The format arguments specify a format that is used when the first two arguments are 0 and 1. A number of formats can be set on plot col boxes. See "Number Col Box<<Set Format(<width>|<width, decimal places>, <"Use Thousands Separator">)" for more information about the syntax.

`Plot Col Box<<Set Values(`*[matrix]* or *{list}*`)`

Sets values for the matrix (for numeric variables) or list (for character variables).

Slider Boxes and Range Slider Boxes

`Slider Box<<Get(`*<index>*`)`

`Range Slider Box<<Get Lower(`*<index>*`)`

`Range Slider Box<<Get Upper(`*<index>*`)`

Returns the current value of the slider.

`Slider Box<<Get Max()`

Returns the maximum value possible for the range slider and slider.

`Slider Box<<Get Min()`

Returns the minimum value possible for the range slider and slider.

`Slider Box<<Get Var`

`Range Slider Box<<Get Lower Var`

`Range Slider Box<<Get Upper Var`

Returns the variable name associated with the slider.

Slider Box<<Set(*float*, <*index*>, <Run Script(*Boolean)*>)

Range Slider Box<<Set Lower(*float*, <*index*>, <Run Script(*Boolean*)>)

Range Slider Box<<Set Upper(*float*, <*index*>, <Run Script(*Boolean*)>)

Sets the value of the slider. Run Script(*Boolean*) controls whether an on-change script runs after the Set, Set Lower, or Set Upper message.

Slider Box<<Set Max(*float*, <*index*>)

Sets the maximum value possible for the range slider and slider.

Slider Box<<Set Min(*float*, <*index*>)

Sets the minimum value possible for the range slider and slider.

Slider Box<<Set Script(<*script*>)

Sets a script to be run when the range sliders and slider is updated.

Slider Box<<Set Var(*slider variable*)

Range Slider Box<<Set Lower Var(*slider variable*)

Range Slider Box<<Set Upper Var(*slider variable*)

Sets the variable name associated with the slider.

String Col Boxes

String Col Box<<Add Element(*item*)

Adds the item to the String Col Box. Item can be a single quoted string or a list of quoted strings.

String Col Box<<Get

String Col Box<<Get(*i*)

Gets the values in a list or the *i*th value.

String Col Box<<Get Heading

Returns the column heading text.

String Col Box<<Remove Element(*row number*)

Removes an element from the column at the specified position.

String Col Box<<Set Allow Text Search(*Boolean*)

Description

In table boxes with selectable rows, allows a quoted string column that has focus to respond to keyboard input to change the selected row.

Example

```
// Run the example.
// Select K2.
// Type the letter g. Notice the last row is selected.
// Type the letters ki. Notice the third row is selected.
New Window( "Mountains",
   tb = Table Box(
      sb =
      String Col Box( "Mountain",
         {"K2", "Delphi", "Kilimanjaro",
         "Grand Teton"}
      ),
      Number Col Box( "Elevation (meters)",
         {8611, 681, 5895, 4199}
      ),
      Plot Col Box( "", {8611, 681, 5895, 4199} )
   )
);
tb << Set Selectable Rows( 1 );
sb << Set Allow Text Search( 1 );
```

String Col Box<<Set Heading(*title*)

Changes the column heading specified in the quoted *title*.

String Col Box<<Set Justify(*justification*)

Specifies the alignment of the contents in the String Col Box to "Right", "Left", or "Center".

Tab Boxes

Tab Box<<Get Tab Margin()

Returns a list of the current margins in pixels for the tab box in this order: Left, Top, Right, and Bottom.

`Tab Box<<Set Style("Tab"|"Combo" |"Outline"|"Vertical Spread"|"Horizontal Spread"|"Minimize Size")`

Changes the appearance of the tab box from a tab to a combo box or outline node.

`"Vertical Spread"` and `"Horizontal Spread"` change the orientation of the tab title.

`"Minimize Size"` bases the tab style on the width of the tab title.

`Tab Box<<Set Tab Margin(`*n*`|{...})`

Sets the tab margin for the tab box. If a single number is specified, all four margins are set to that number of pixels. If a list of two numbers is specified, the left and right margins are set to the first number, and the top and bottom margins are set to the second number. If a list of four numbers is specified, the margins are set in this order: `{left, top, right, bottom}`.

`Tab Box<<Show Tabs(`*Boolean*`)`

Shows or hides the tabs for tab boxes. If you hide the tabs, you need to provide another way to select and show tabs. For example, a list box that contains a list of references to the tabs. The default value is 1.

Table Boxes

`Table Box<<Bootstrap(`*nsample*`, Random Seed(`*number*`), Fractional Weights(`*Boolean*`), Split Selected Column(`*Boolean*`), Discard Stacked Table if Split Works(`*Boolean*`)`

Description

Bootstraps the analysis, repeating it many times with different resampling weights and collecting tables as selected.

See Also

"Number Col Boxes"

`Table Box<<Get`

Gets the entries of the table in list form.

`Table Box<<Get As Matrix(<"Visible">)`

Gets the numeric entries of the table in matrix form. `"Visible"` means that only visible columns will be included.

Table Box<<Get Click Sort

Returns 1 if the table can be sorted by clicking a column header and 0 otherwise.

Table Box<<Get Locked Columns

Returns the number of columns that cannot be dragged with the cursor or have any columns dropped before them.

Table Box<<Get Row Change Function

Returns the expression that is evaluated when a row is selected.

Table Box<<Get Selectable Rows

Returns True if the table box currently allows row selection.

Table Box<<Get Selected Row Color

Returns the index number of the background color of the selected rows in the table box.

Table Box<<Make Combined Data Table

Returns a reference to the data table. Same as `Make Data Table`, but also searches the report for report tables with the same columns and combines all of these into the new data table.

Table Box<<Make Data Table(*name*)

Returns a reference to the data table. Turns the table entries into a new data table with the quoted *name* argument.

Table Box<<Reorder Columns(*from column index*, *to column index*)

Puts the column specified with *from column index* in the place of the column specified with *to column index*. The indexes are 0-based. For example, indicate the first column with "0", and indicate the second column with "1".

Table Box<<Set Cell Changed Function(Function({*this*, *col box*, *row*},<*script*>))

Description

Sets a function that is called whenever the user edits a cell in a column in a table.

Example

This example prints the new values for the changed cell to the log.

```
New Window( "Mountains",
   tb = Table Box(
      quoted string Col Edit Box(
         "Mountain",
         {"K2", "Delphi", "Kilimanjaro",
         "Grand Teton"}
      ),
      Number Col Edit Box(
         "Elevation (meters)",
         {8611, 681, 5895, 4199}
      ),
      Plot Col Box( "", {8611, 681, 5895, 4199} )
   )
);
tb <<
Set Cell Changed Function(
   Function( {this, col, row},
      Print(
         (col << Get Heading) || ": row:" ||
         Char( 3 ) || " is now " ||
         Char( col << Get( row ) )
      )
   )
);
```

Table Box<<Set Click Sort(*Boolean*)

Specifies if the table can be sorted by clicking a column header.

Table Box<<Set Column Borders(*Boolean*)

Draws a line on each side of the column.

Table Box<<Set Heading Column Borders(*Boolean*)

Draws a line on each side of the column headings.

Table Box<<Set Locked Columns(*n*)

Locks the first *n* columns. You cannot drag the locked columns or drag columns before them.

Table Box<<Set Row Borders(*Boolean*)

Draws a line above and below each row.

Table Box<<Set Row Change Function(*function*)

Sets the expression that is evaluated when a row is selected.

Table Box<<Set Selectable Rows(*Boolean*)

Makes the rows of the table box selectable or not.

Table Box<<Set Selected Row Color(*color*)

If the rows of the table box are selectable (`Set Selectable Rows(True)`), sets the background color (specified in the quoted *color* argument) for the selected rows.

Table Box<<Set Shade Cells(*Boolean*)

Shades the background of every cell in the table.

Table Box<<Set Shade Alternate Rows(*Boolean*)

Shades the background of every other row in the table.

Table Box<<Set Shade Heading(*Boolean*)

Shades the background in column headings.

Table Box<<Set Underline Headings(*Boolean*)

Draws a line underneath the column headings.

Table Box<<Sort By Column(<*column number*|*column title*>, <Ascending(*Boolean*)>

Sorts all rows based on the values in the specific column number or quoted column title. The default order sorting is descending.

Text Boxes

Text Box<<Font Color(*n*)

Sets the color for `Text` quoted strings.

Text Box<<Get Hidden State

Returns the current state of a text box.

Text Box<<Get Text

Returns the quoted string content of the box.

Text Box<<Get Tip

Returns the tooltip for the text box (or a text edit box).

Text Box<<Markup

Returns text formatted with the specified HTML tags. The HTML must be well-formed; make sure you close nested tags correctly.

The following example returns text formatted in bold, italic, and underlined.

```
w = New Window( "Formatted Text",
   Text Box( "This is <b>bold</b> text. This is <b><i>bold italic</i></b>
   text. This is <u>underlined</u> text.",
   <<Markup) );
```

Text Box<<Rotate Text(*direction*)

Rotates the text 90 degrees "Left" or "Right", or returns it to horizontal.

Text Box<<Set Font(*name*, *<size>*, *<style|style style...>*, *<angle>*)

Sets the font specified in the quoted *name* argument and the properties for text quoted strings. To specify more than one style, include a space between each *style* and place them in quotes.

Text Box<<Set Font Size(*n*)

Sets the font size in points for text quoted strings.

Text Box<<Set Script(*script*)

Associate a script with a text box. The script executes when the user presses Enter (or the text edit box otherwise loses focus).

Text Box<<Set Text(*quoted string*)

Sets the text in the box as specified in the quoted *string* argument.

Text Box<<Set Tip(*quoted string*)

Sets the tooltip for the text box (or a text edit box) as specified in the quoted *string* argument.

`Text Box<<Set Wrap(n)`

Set the wrap point, in pixels, in pixels (*n*).

Tree Node and Tree Box

For the following messages, `node` stands for a tree node or a reference to one and `root` stands for a tree box or a reference to one.

Caution: If you send a root node that contains one or more nodes with the `Set Node Select Script` defining a collapse message, then macOS runs the script twice. Windows doesn't run the script. This behavior on macOS doesn't just affect increments. Any script runs twice. It will print to the log twice, create a column twice, try to delete something twice, and so on.

`node<<Append(<node>)`

Inserts a referenced tree node after this node's children.

`node<<Collapse`

Closes the node. The behavior is not guaranteed if the node has a collapsed parent.

`node<<Expand`

Opens the node. The behavior is not guaranteed if the node has a collapsed parent.

`node<<Get Dimmed(<node>)`

Gets the option to dim text (decrease the opacity) for the node.

`node<<Get Font Style(<node>)`

Gets the font style for the node.

`node<<Get Tip`

Returns the tooltip for the node.

`node<<Prepend(<node>)`

Inserts a tree node before this node's children.

`node<<Remove`

Removes the given tree node and all its children from the tree display box.

`node<<Set Dimmed(`*`Boolean`*`)`

Sets the option to dim text (decrease the opacity)) for the node.

`node<<Set Font Style("Plain"|"Bold")`

Specifies the font style for the node.

`node<<Set Selected(`*`<node>`*`)`

Selects the node. The behavior is not guaranteed if the node has a collapsed parent.

`node<<Set Tip(`*`tooltip`*`)`

Sets a tooltip for the node. The *`tooltip`* argument is quoted.

`root<<Collapse(`*`<node>`*`)`

Collapses the given tree node.

`root<<Expand(`*`<node>`*`)`

Expands the given tree node.

`root<<Get Selected(`*`<node>`*`)`

Gets the currently selected tree node.

- In a single-item tree, the currently selected tree node or `Empty` is returned.
- Table 3.1 shows the results for a `Tree Box()` that contains the `MultiSelect` argument.

Table 3.1 Multi-Select Tree Results

Items Selected in Tree	Result
no items selected	empty
single item selected	list of one tree node
multiple items selected	list of selected tree nodes

`root<<Is Multiselect`

Returns 1 for a `MultiSelect` tree and 0 for a single-select tree.

`root<<Set Selected(node|{nodes}, <Boolean>)`

Selects the given tree node in the tree display box. In list of tree nodes, all nodes in the list are selected for `MultiSelect` trees. Otherwise, the first node in the list is selected. Specify the Boolean argument to indicate whether the node or nodes should be selected or unselected. The default value is 1, which selects the nodes.

Notes:

- On Windows, the `Set Selected` message expands all nodes between the selected node or nodes and the root of the tree; items that are selected deep within the tree are shown. The expansion state does not change for nodes that were previously selected.
- On macOS, the `Set Selected` message does not change the tree expansion state.

Triangulation

For the following messages, `tri` stands for a triangulation or a reference to one.

`tri<<Get N Points`

Returns the number of unique points in the triangulation.

`tri<<Get Points`

Returns the coordinates of the unique points in the triangulation.

`tri<<Get Y`

Returns the Y values of the unique points in the triangulation.

`tri<<Get N Hull Points`

Returns the number of points on the boundary of the triangulation.

`tri<<Get Hull Points`

Returns the indices of the points on the boundary of the triangulation.

`tri<<Get N Hull Edges`

Returns the number of edges on the boundary of the triangulation.

`tri<<Get Hull Edges`

Returns the indices of the edges on the boundary of the triangulation.

`tri<<Get Hull Path`

Returns the boundary of the triangulation as a path.

`tri<<Get N Triangles`

Returns the number of triangles.

`tri<<Get Triangles`

Returns the indices of the triangles in the form of an Nx3 matrix.

`tri<<Get N Edges`

Returns the number of edges in the triangulation.

`tri<<Get Edges`

Returns the indices of the edges in the form of an *N*x2 matrix.

`tri<<Subset(`*`{indices}`*`)`

Returns a triangulation resulting from the given subset of points.

`tri<<Peel`

Peel the boundary layer of a triangulation, returning a new triangulation.

Windows

`window<<Bring Window to Front`

Brings the window to the front.

`window<<Close Window(`*`<nosave>`*`)`

Closes the window. If the optional argument nosave is specified, the window (journal, report, and so forth) is closed without saving or prompting.

`window<<Get Content Size`

Returns the size of the window's contents.

`window<<Get Window Icon`

Returns the name of the window's icon.

`window<<Get Window Position`

Returns the position of the window.

`window<<Get Window Size`

Returns the size of the window.

`window<<Get Window Title`

Returns the title of the window.

`window<<Inval`

Invalidate the display box. The window updates either when the `<<Update Window` message is sent or when the operating system has time for the update. See "window<<Reshow" for another method.

`window<<Maximize Display`

Maximizes the window. Deprecated.

`window<<Maximize Window(`*`Boolean`*`)`

Maximizes the window. Deprecated.

`window<<Minimize Window(`*`Boolean`*`)`

Minimizes the window.

`window<<Move Window(`*`x`*`, `*`y`*`)`

Moves the window to the specified position.

`window<<On Close(`*`script`*`)`

Runs the script when the window is closed.

`window<<Pad Window(`*`Boolean`*`)`

Turns padding around a window's contents on (`1`) or off (`0`). The default value is off.

`window<<Print Window`

Prints the window to the default printer. Note that the Print window is not opened and user input is not required.

`window<<Reshow`

Invalidates the display box and updates the window with the new content. See `<<Inval` and `<<Update Window` messages if more control over timing of the update is required.

`window<<Save Window to Report(`*`pathname`*`, <Embed Data(`*`Boolean`*`)>)`

Saves the current report window to a JMP report file (.jrp).

`window<<Set Main Window`

Sets the specified window as the default window that appears when JMP is run.

`window<<Set Window Icon(`*`icon name`*`)`

Sets the window's icon as specified in the quoted *`icon name`* argument.

`window<<Set Window Size(`*`x`*`, `*`y`*`)`

Resizes the window.

`window<<Show Window(`*`Boolean`*`)`

1 shows the window (only if the window is not currently open). 0 hides the window. If the window is also minimized (on Windows) or docked (on macOS), showing the window restores it to the normal state and brings it to the front.

`window<<Size Window(`*`x`*`, `*`y`*`)`

Resizes the window.

`window<<Update Window`

Updates or refreshes the window holding the display box if there are invalidated regions. See "window<<Inval" and "window<<Reshow" for additional methods.

`window<<Window Class Name`

Returns the name of the window class for the display box. Valid responses include: DataTable, FormulaEditor, Starter, Journal, Launcher, Report, Dialog, DialogWithMenu, ModalDialog, FindReplace, User, Generic, ToolWindow, FindReplace, AppBuilder, and Debugger.

`window<<Zoom Window`

Resizes the window to be large enough to show all of its contents.

Dynamic Link Libraries (DLLs)

dll object<<Call DLL(*function name*, *signature*, *arguments*)

Calls the specified function in the DLL with the specified signature and arguments.

dll object<<Declare Function(*name*, *<named arguments>*)

Description

Declares the return type and argument types for the specified function so that it can be successfully invoked. You can use one of the named arguments for Convention: "STDCALL" or "PASCAL", or "CDECL". The *type* argument for Returns takes the same named arguments as Arg. The *name* argument is quoted.

Optional Named Arguments

Alias(*name*) Specifies a quoted *name* that you can include if you don't like the name encoded in the DLL.

Arg(*type*, *<description>*, *<access mode>*, *<array>*) Arg can appear multiple times, once for each argument to be sent to the function.

type is one of these keywords that specifies the argument type: "Int8", "UInt8", "Int16", "UInt16", "Int32", "UInt32", "Int64", "UInt64", "Float", "Double", "Ansiquoted string", "Unicodequoted string", "Struct", "IntPtr", "UIntPtr", or "ObjPtr".

description is a quoted string that describes the argument for reference.

access mode is an optional keyword that specifies how the argument is passed. "input" specifies that the argument is passed by value. "output" specifies that the argument is passed by address with the initial value undefined. "update" specifies that the argument is passed by reference and the value of the JSL variable is set as the initial value. The default value is "input".

array is an optional keyword. It is valid only if the *type* is specified as "Double" and the access mode is specified as either "input" or "update". Specifies that the exported function expects an array of doubles.

Convention(*calling convention*) Specifies the calling convention: "STDCALL" or "PASCAL", or "CDECL". The default value is "STDCALL". STDCALL and PASCAL are equivalent.

MaxArgs(*n*) Specifies the maximum number of arguments that can be supplied.

MinArgs(*n*) Specifies the minimum number of arguments that can be supplied.

Returns(*type*) Specifies the data type that the function returns: "Int8", "UInt8", "Int16", "UInt16", "Int32", "UInt32", "Int64", "UInt64", "Float", "Double", "Ansiquoted string", "Unicodequoted string", "Struct", "IntPtr", "UIntPtr", or "ObjPtr".

`StackOrder(`*`order`*`)` Specifies the order in which arguments are placed on the stack when calling the function. Valid values are `"L2R"` (left-to-right) and `"R2L"` (right-to-left). The default value is `"R2L"`.

`StackPop(`*`pop`*`)` Specifies how the exported function expects the stack to be cleared after the function returns. Valid values are `"CALLER"` and `"CALLEE"`. The default value is `"CALLEE"`.

`StructArg(`*`Arg(...)`*`, `*`<Arg(...)>`*`, ..., `*`<access mode>`*`, `*`<pack mode>`*`, `*`<description>`*`)` Can appear multiple times. If an exported DLL function requires that a structure argument be passed in as an argument, use `StructArg` to declare the structure members. The `Arg` arguments use the same syntax as for `Arg` arguments to `Declare Function` (one for each structure member), an `access mode` indicator and a `pack mode` indicator.

`access mode` is an optional keyword that indicates whether the struct argument should be passed by value (`input`) or by reference (`update`).

`pack mode` is an optional integer that determines how the structure is packed. Valid values are 1, 2, 4, 8, and 16. The default value is 8.

`description` is an optional, quoted string that contains a description of the structure for reference.

`dll object<<Get Declaration JSL`

Sends the declaration JSL from the DLL object to log.

`dll object<<Load DLL(`*`path`*`, <AutoDeclare(`*`Boolean`*`|"Quiet"|"Verbose")>)`

`dll object<<Load DLL(`*`path`*`, <"Quiet"|"Verbose">)`

Description

Loads the DLL from the specified path.

Required Argument

`path` A quoted path that specifies where to load the DLL.

Optional Named Arguments

`AutoDeclare(`*`Boolean`*`|"Quiet"|"Verbose")` `AutoDeclare(1)` and `AutoDeclare("Verbose")` write verbose messages to the log. `AutoDeclare("Quiet")` turns off log window messages. If you omit this option, verbose messages are written to the log.

`Quiet|Verbose` When you use `Declare Function`, this option turns off log window messaging (`"Quiet"`) or turns on log window messaging (`"Verbose"`).

`dll object<<Show Functions`

Sends the declared functions for the DLL object to the log.

`dll object<<Unload DLL`

Unloads the DLL.

HTML 5

Web Report

`webreport<<Publish(<Add Image(...)>, <Add Report(...)>, <Add Reports(...)>, <Public(`*Boolean*`)>, <Index(...), <User Name(...)>, <Password(...)|`*password function*`)>, <Prompt("IfNeeded"|"Always"|"Never")>, <URL(...)>, <Publish Data(`*Boolean*`)>, Replace(<`*id*`>, Prompt("IfNeeded"|"Always"|"Never"))>)`

Description

Publishes the web report to the JMP server.

Returns

On success, the URL of the published report is returned.

Optional Arguments

`Add Image` Inserts an image at the top of the index page. Valid formats are `"png"`, `"bmp"`, `"jpeg"`, `"jpg"`, `"tiff"`, and `"tif"`. `Title` and `Description` are optional. `Title` appears above the image. `Description` appears below the image. Use `File(`*filepath*`)` or just a quoted string. Here is an example:

```
webrpt << Add Image( File( "C:\Users\Public\JMP\Projects\WebJMP\atlas.jpg" ), Title( "Atlas" ), Description( "Holding up the world as always." ) );
```

`Add Report` Adds a report to publish within the web report.

`Public(`*Boolean*`)` Specifies whether the public has access to the report. By default, the report is private.

`Index` The name of the index page for multiple reports. You can also specify the description.

`User Name` Specifies the user name registered on the JMP server.

`Password` Specifies the user's password. You can also define a password function.

`Prompt` Displays a window in which the user types the server URL, user name, and password.

`URL` The location that you are publishing to.

`Publish Data(`*Boolean*`)` Includes the data in the HTML. Reports contains static rather than interactive images. In a public report, you might not want to share the data.

`Replace` Replaces the report. Get the URL from the address field in the browser where the page is displayed.

Images

The Scripting Index provides examples for processing images. In JMP, select **Help > Scripting Index** to view this interactive resource.

Additional resources are available from the JMP File Exchange at https://community.jmp.com/community/file-exchange.

`img<<Crop(Left(`*pix*`), Right(`*pix*`), Top(`*pix*`), Bottom(`*pix*`))`

Creates a new image from an existing image to the specified dimensions (in pixels).

`img<<Filter(`*name*`, <`*n*`>)`

Filters the image based on the specified algorithm. Filtering is useful for cleaning up noise in the image.

Note: All of the JMP image filters are supported at the operating system level. Images that are processed on Windows might differ from images processed on macOS.

Argument

`name` Specifies the quoted name of a JMP image filter. The following filters are available:

- `"Despeckle"` removes defects (that is, speckles) from a scanned or captured image (for example, scratches, dust, etc.).
- `"Edge"` identifies pixels in an image where the brightness changes sharply and darkens pixels with no sharp change. Edge detection is used to detect changes in surface, depth, material, and lighting.
- `"Enhance"` reduces the contrast between pixels in a noisy image.
- `"Median"` reduces noise (that is, the random variation) and smooths an image by comparing each pixel's brightness with its neighbors' and, if the value is very different, replaces it with the average of the neighbors' values.
- `"Negate"` creates the negative of the color or gray-scale image by changing each pixel color to its complementary color.
- `"Normalize"` changes a color image's pixels to use the full range of the file format's number system. Normalization will make the image's colors more intense.
- `"Sharpen"` reduces blur by sharpening edges of an image.
- `"Contrast"`, *n* brightens or darkens an image. A higher number (>0.0) brightens an image; a lower number (<0.0) darkens an image.

- "Gamma", *n* corrects the image visual display (brightness and intensity) to account for differences in monitor hardware. A higher number (> 1.0) lightens the image; a lower number (< 1.0) darkens the image.
- "Reduce Noise", *n* reduces the random variation (or noise) that occurs with higher ISO sensitivity or longer exposure times.
- "Gaussian Blur", radius, sigma reduces image noise and detail creating a smoother image. Radius is equal to the blur radius around each pixel and sigma is the standard deviation of the Gaussian distribution. Gaussian blur is commonly used when resizing or performing edge detection.

`img<<Flip Both`

Flips the image from left to right and top to bottom.

`img<<Flip Horizontal`

Flips the image from left to right.

`img<<Flip Vertical`

Flips the image from top to bottom.

`img<<Get EXIF`

Returns EXIF data from the image (such as the shutter speed and aperture value) in an associative array.

`img<<Get N Frames`

Description

Returns the number of frames in a multi-frame TIF or animated GIF file, where the number of frames begins with frame 0.

Example

The following example places a four-frame TIF file in a new window and shows the image that is in the first frame.

```
img = New Image( "$DOWNLOADS/Multiframe.tif" );
nframes = img << Get N Frames(); // return 4
img << Set Current Frame( 1 ); // show image 1
win = New Window( "Multi-Frame TIFF", img );
```

`img<<Get Size`

`img<<Size`

Returns a list containing the width and height (in pixels) of the image.

`img<<Rotate(`*degrees*`)`

Rotates the image by the specified number of degrees.

`img<<Save Image(`*path*`)`

Saves the image to the quoted *path*.

`img<<Scale(`*scale*`|`*xscale*`, `*yscale*`)`

Resizes the image by the specified dimensions. Provide one argument to resize both the width and height. Provide two arguments to resize the width and height separately.

Examples

```
img = New Image( "$SAMPLE_IMAGES/tile.jpg" );
xs = 2;
img << Scale( xs );
New Window( "Tilex 2", img );

img = New Image( "$SAMPLE_IMAGES/tile.jpg" );
img << Scale( 2, 0.5 ); // scale image width by 2 and height by 1/2
New Window( "Tile squished", img );
```

Notes

Using `Scale` is an alternative to getting the size of the image, multiplying by the scale factor, and then setting the size.

`img<<Set Current Frame`

Description

Sets the frame that shows in a multi-frame TIFF or animated GIF file. Specify 0 through the number of frames minus 1. For example, with four frames, you can specify frame 0 through frame 3.

See Also

"img<<Get N Frames"

`img<<Set Size(`*width*`, `*height*`)`

Resizes the image to the specified dimensions (in pixels). To scale the image proportionally, specify a width and height that correspond to the aspect ratio in the original image.

`img<<Transparency(`*fraction*`)`

Sets the transparency for the image where the fraction is between 0.0 (full transparency) to 1.0 (no transparency).

JMP Applications

The JMP Application Builder and JMP Dashboard Builder use the same infrastructure to design and execute applications and dashboards. Because a dashboard is a special form of an application, this section uses the term *application* to describe how both dashboard and application objects use scripting.

See the Scripting Index in the JMP Help menu for examples.

JMP App

The `JMP App` object is the main controller for JMP applications built by Application Builder or Dashboard Builder. Scripts both inside and outside of a JMP application can use a `JMP App` object.

A JMP application can have one of three states: initial (with no editor, and the application is not running), running, or editing. A `JMP App` object only exists in one state at a time; if you are editing a JMP Application and choose to run it, a copy of the JMP application is created before it is run.

`app<<Combine Windows({`*reports or data tables*`})`

Description

Combines the given list of platform reports or data tables into a new module. The application should be in the initial state when this message is sent.

`app<<Debug`

Invokes the JSL Debugger on a JMP application. The application script will run first. The Debugger then breaks as each module is created, invoking the module scripts. In the Debugger, set breakpoints to debug the scripts that are associated with the application or modules.

`app<<Edit`

Starts Application Builder on a JMP application that is in the initial state. There is no editor, and the application is not running.

`app<<Get Modules`

Gets the list of modules associated with an application. In Application Builder, each module corresponds to a tab in the workspace, which describes the layout and behavior for one type of window in the application.

`app<<Get Namespace`

The `JMP App()` object automatically creates an anonymous namespace for the variables created within the application script. Use this message to get a handle to this namespace to inspect or modify variables. There is a default symbol in this namespace called `thisApplication`, which holds a reference to the application itself.

`app<<Get Windows`

Gets a list of all windows created as instances of JMP app modules. Some modules might create more than one instance. All windows might not exist at the same time, so the number of windows might vary and might differ from the number of modules.

`app<<Open File(`*`path`*`)`

Resets the state of an existing application from a the .jmpapp or .jmappsource file. *`path`* is quoted.

`app<<Relaunch Analysis`

Creates a new copy of a running application and runs the new instance.

`app<<Run`

Runs the application. The application script runs first, and depending on settings, one or more JMP app module instance objects might be created automatically.

`app<<Save Script to Add-In`

`app<<Save Script to Data Table`

`app<<Save Script to Journal`

`app<<Save Script to Script Window`

Saves the script for the application to the given destination. An application script consists of a `JMP App()` object that contains the definition for the application. Scripts saved to an add-in, data table, or journal include a `Run` message to run the application. A script saved to the script window includes an `Edit` message to open Application Builder.

JMP App Module

A JMP application module is a definition of the display box layout and behavior for a single component in a JMP application or dashboard. Depending on the module type, the component might represent a window in the application or just part of a window.

module<<Create Instance

Use `Create Instance` within a `JMP App()` or `JMP App Module()` script to create an instance of a JMP app module. By default, one instance of each JMP application module is created when an application is run. For more complex applications with multiple windows, such as a launcher and report combination, it might be necessary to change the default settings and control how the module instance is created.

module<<Get Instance

Returns a handle to the application that owns a module.

JMP App Module Instance

The JMP application module instance is a running realization of a JMP app module, a window on the screen, or a collection of display box elements that can be inserted into another window.

inst<<Create Objects

Appears in the default template for a JMP app module script. This message controls the point at which the display and window for a module instance are created. The message appears in the script so that the script writer can choose to do certain setup before the objects are created. One example of this setup is for a parameterized application.

inst<<Get Box

Returns a handle to the top-most display box associated with a module instance. This might be useful to issue display or window commands, such as the `Save to PDF` or `Close Window` messages.

inst<<Get Namespace

Like `JMP App()`, each JMP application module instance also creates an anonymous namespace for all variables created in the module script. The namespace also includes all the variables that represent the display boxes in the module. This namespace contains a default symbol named `thisModuleInstance` that refers to itself.

`inst<<Get User Data`

`inst<<Set User Data`

Stores and retrieves a JSL value in the JMP application module instance. The value could be a number, quoted string, list, associative array, or other JSL type that is returned with the `Type()` function.

MATLAB

The MATLAB interfaces are scriptable using a MATLAB connection object. Use the MATLAB Connect() JSL function to obtain a scriptable MATLAB connection object.

`mlconn<<Control(<Echo(`*`Boolean`*`)>, <Visible(`*`Boolean`*`)>)`

Controls the execution of MATLAB.

Returns

None.

Optional Global Arguments

`Echo(`*`Boolean`*`)` Echo MATLAB source lines to the JMP Log window.

`Visible(`*`Boolean`*`)` Determine whether to show or hide the active MATLAB workspace.

`mlconn<<Disconnect()`

Description

Disconnects this MATLAB integration interface connection.

`mlconn<<Execute({`*`list of inputs`*`}, {`*`list of outputs`*`}, `*`mCode`*`, <Expand(`*`Boolean`*`)>, <Echo(`*`Boolean`*`)>)`

Submits MATLAB code to the active global MATLAB integration interface connection given a list of inputs and upon completion a list of outputs are retrieved.

Returns

0 if successful, otherwise nonzero.

Required Arguments

`{`*`list of inputs`*`}` Positional, name list. List of JMP variable names to be sent to MATLAB as inputs.

`{`*`list of outputs`*`}` Positional, name list. List of JMP variable names to be retrieved from MATLAB as outputs.

`mCode` Positional, quoted string. The MATLAB code to submit.

Optional Named Arguments

Expand(*Boolean*) Performs an Eval Insert on the MATLAB code prior to submission.

Echo(*Boolean*) Echos MATLAB source lines to the JMP Log window. Default is true.

---

mlconn<<Get Graphics(*format*)

Gets the last graphic object written to the MATLAB graph display window. The graphic object can be returned in several graphic formats.

Returns

JMP Picture object.

Optional Argument

format Positional. The quoted format the MATLAB graph display window contents are to be converted to. Valid formats are "png", "bmp", "jpeg", "jpg", "tiff", and "tif".

---

mlconn<<Get Version()

Gets the current version of the installed MATLAB.

Returns

Matrix, returns a vector of length 3 containing the MATLAB version number.

---

mlconn<<Get(*name*)

Description

Gets a named variable from MATLAB to JMP.

Returns

Value of named variable.

Required Argument

name The name of a JMP variable to be retrieved from MATLAB.

---

mlconn<<Is Connected()

Description

Determines whether connection is active.

Returns

1 if connected, otherwise 0.

---

mlconn<<JMP Name To MATLAB Name(*jmp name*)

Description

Maps a JMP variable name to its corresponding MATLAB variable name using MATLAB variable name naming rules.

Returns

A quoted string, a mapped MATLAB name.

Required Argument

`jmp name` Positional. The name of a JMP variable to be sent to MATLAB.

`mlconn<<Send(name, <named arguments>)`

Description

Sends the named variable from JMP to MATLAB.

Returns

0 if successful, otherwise nonzero.

Required Argument

`name` Positional. The name of a JMP variable to be sent to MATLAB.

Named Arguments

The following arguments are for data tables only:

`Selected(Boolean)` Send selected rows from the referenced data table to MATLAB.

`Excluded(Boolean)` Send only excluded rows from the referenced data table to MATLAB.

`Labeled(Boolean)` Send only labeled rows from the referenced data table to MATLAB.

`Hidden(Boolean)` Send only hidden rows from the referenced data table to MATLAB.

`Colored(Boolean)` Send only colored rows from the referenced data table to MATLAB.

`Markered(Boolean)` Send only markered rows from the referenced data table to MATLAB.

`Row States (Boolean, <named arguments>)` Send row states from referenced data table to MATLAB by adding an additional data column named "RowState". Create multiple selections by adding together individual settings. The row state consists of individual settings with the following values:

- Selected = 1
- Excluded = 2
- Labeled = 4
- Hidden = 8
- Colored = 16
- Markered = 32

Row State Optional Named Arguments

The following optional, named Row States arguments are supported:

`Colors(Boolean)` Send row colors. Adds additional data column named "RowStateColor".

Markers(*Boolean*) Send row markers. Adds additional data column named "RowStateMarker".

---

mlconn<<Submit(*mCode*, *<named arguments>*)

Description

Submits MATLAB code to the active global MATLAB integration interface connection.

Returns

0 if successful, otherwise nonzero.

Required Argument

mCode Positional quoted string. The MATLAB code to submit.

Named Arguments

Expand(*Boolean*) Perform an Eval Insert on the MATLAB code prior to submission.

Echo(*Boolean*) Echo MATLAB source lines to the JMP log. The default is true.

---

mlconn<<Submit File(*path*)

Description

Submits statements to MATLAB using a quoted *path*.

Returns

0 if successful, otherwise nonzero.

Arguments

path Positional quoted string. The path to a file containing the MATLAB source lines to be executed.

Namespaces

---

ns<<Contains(*string*)

Returns 1 or 0, depending on whether the specified quoted string exists within the namespace.

---

ns<<Delete Namespace

Removes this namespace from the internal global list.

To delete variables in the namespace, use the Remove(*variable name*) message.

---

ns<<First

Returns a quoted string that contains the first variable name used within the namespace.

`ns<<Get Contents`

Returns a list of key-value pairs, which are each enclosed in a list. Each key is a quoted string that contains a variable name, and each value is the unevaluated expression that the variable contains.

`ns<<Get Keys`

Returns a list of variable names.

`ns<<Get Name`

Returns the name of this namespace.

`ns<<Get Value(`*`variable name`*`);`

Returns the unevaluated expression that the quoted *`variable name`* contains in this namespace.

`ns<<Get Values`

Returns a list of unevaluated expressions that each variable in the namespace contains.

`ns<<Get Values({`*`variable name1, variable name2, ...`*` });`

Returns a list of unevaluated expressions that each quoted variable in the namespace specified in the list argument contains. If a requested variable name is not found, an error is returned.

`ns<<Insert(`*`variable name`*`, `*`expr`*`);`

Inserts into this namespace a quoted variable named *`variable name`* that holds the expression *`expr`*.

`ns<<Lock Namespace(<`*`variable name, ...`*`>)`

Locks all specified variables in the namespace and prevents quotes variables from being added or removed. If no variables are specified, all variables in the namespace are locked.

`ns<<N Items`

Returns the number of variables contained in the namespace.

`ns<<New Namespace(name, <{list of expressions}>)`

Creates a namespace where all functions and variables created are defined only within the optional quoted *name* argument.

`ns<<Next(variable name)l`

Returns the name of the variable that follows the specified quoted variable.

`ns<<Remove(variable name, ...)`

Removes the specified quoted variable or list of variables.

`ns<<Show Contents`

Shows the contents of a namespace in the log.

`ns<<Unlock Namespace(variable name, ...);`

Unlocks the specified quoted variables in the namespace. If no variables are specified, all variables are unlocked.

Platforms

`obj<<Action`

Evaluates expressions. Useful for quoted stringing together multiple platforms interrupted by user input.

`obj<<Automatic Recalc`

Redoes the analysis automatically for exclude and data changes. If automatic recalc is on, you should use `Wait(0)` commands to let the triggers take effect and do the recalculation.

Note: Not supported on all platforms.

`obj<<Bring Window To Front`

Brings the current window to the front.

`obj<<Close Window`

Closes window identified by *obj*, typically a platform surface.

`obj<<Column Switcher(`*`default column, {column1, column2, ...}`*`)`

Adds a control panel to a platform for switching variables.

`obj<<Copy ByGroup Script`

Creates a script to produce this analysis containing By variables and place it on the clipboard.

`obj<<Copy Script`

Creates a script to produce this analysis and place it on the clipboard.

`obj<<Data Table Window`

Makes the associated data table window active (front-most).

`obj<<Get Data Table`

Returns a reference to the data table.

`obj<<Get Script`

Returns script to reproduce the analysis as an expression in the log.

`obj<<Get Script With Data Table`

Creates a script to reproduce the analysis, specifically referencing the source data table, and returns it as an expression in the log.

`obj<<Get Timing`

Times the launch of the platform and returns it in the log.

`obj<<Get Web Support`

Returns the score for the display tree that is about to be saved as interactive HTML. Possible values are -1 (unsupported), 0 (supported), and 1 (supported). If the score does not equal -1, interactive HTML is supported and the `Save Interactive HTML` message can be used.

`obj<<Get Window Position`

Gets the position of the window. Returns an ordered pair.

obj<<Get Window Size

Gets the window size, in pixels. Returns an ordered pair.

obj<<Ignore Platform Preferences(*Boolean*)

Ignores the current settings of the platform's preferences. The message is ignored when sent to the platform after creation.

obj<<Journal Window

Appends the contents of the window to the journal.

obj<<Local Data Filter

Filters data to specific groups or ranges, but stays local to the platform.

obj<<Maximize Window

Maximizes the window. Equivalent to pushing the maximize button in the corner of the window. This message takes an optional Boolean argument:

```
// maximize the window:
obj<<Maximize Window(1)
// restore the window:
obj<<Maximize Window(0)
```

obj<<Minimize Window

Minimizes the window. Equivalent to pushing the minimize button in the corner of the window. This message takes an optional Boolean argument:

```
// minimize the window
obj<<Minimize Window( 1 )
// restore the window
obj<<Minimize Window( 0 )
```

obj<<Move Window(*x*, *y*)

Moves the window to the (*x*, *y*) location on your screen.

obj<<Print Window

Sends the selected window to the printer.

obj<<Redo Analysis

Reruns the analysis with the same options.

`obj<<Redo ByGroup Analysis`

Reruns the same analysis involving By groups.

`obj<<Relaunch Analysis`

Returns to the launch window for this analysis.

`obj<<Relaunch ByGroup`

Returns to the launch window for this analysis involving By groups.

`obj<<Remove Column Switcher`

Removes all Column Switchers that were added to the platform.

`obj<<Remove Local Data Filter`

Removes all Local Data Filters that were added to the platform.

`obj<<Report`

`Report(`*`obj`*`)`

Returns a display box reference for the report in the platform window.

`obj<<Report View`

Determines the level of detail visible in a platform report. Full shows all detail and Summary shows only select content, dependent upon the specific platform. For customized behavior, use the `Set Summary Behavior` message with display boxes.

`obj<<Save ByGroup Script to Data Table(<`*`name`*`>, <Append Suffix(`*`Boolean`*`)>, <Prompt(`*`Boolean`*`)>, <Replace(`*`Boolean`*`)>)`

Description

Creates a table script to produce the analysis involving By variables and saves it as a table script in the data table.

Optional Arguments

`name` The name of the script. *`name`* is quoted. If omitted, the platform names the script. For example, in Tabulate, the script is named "Tabulate". In Bivariate, the script might be named "Bivariate of height by weight" to reflect the platform and column names.

`Append Suffix(`*`Boolean`*`)` If true, appends a numerical suffix to the script name. This suffix differentiates the script from an existing script with the same name.

`Prompt(`*`Boolean`*`)` If true, prompts the user to specify a script name.

Replace(*Boolean*) If true, replaces an existing script with the same name.

`obj<<Save ByGroup Script to Journal`

Creates a table script to produce the analysis involving By variables and adds a button to the journal containing this script.

`obj<<Save ByGroup Script to Script Window`

Creates a script to produce the analysis involving By variables and appends it to the current Script window.

`obj<<Save Script for All Objects`

Saves script to reproduce all analyses found within the object's window in the Script Journal window.

`obj<<Save Script for All Objects to Data Table`

Saves a script for all report objects to the current data table. The script is named after the platform unless you specify the script name in quotes.

```
obj << Save Script for All Objects To Data Table("My Script")
```

`obj<<Save Script to Data Table`

Saves script to reproduce analysis as a property in the associated data table.

`obj<<Save Script to Journal`

Creates a script to produce the analysis and adds a button to the journal containing this script.

`obj<<Save Script to Report`

Saves script to reproduce analysis as a text box at the top of the report.

`obj<<Save Script to Script Window`

Saves a script to reproduce analysis in the Script Journal.

`obj<<Scroll Window(`*x*`, `*y*`)`

`obj<<Scroll Window({`*x*`, `*y*`})`

Scrolls the window *x* pixels to the right and *y* pixels down from the current position. Negative coordinates go left and up. If the coordinates are a list in braces { }, they are

absolute coordinates. The window scrolls to the point *x* pixels from the left and *y* pixels from the top.

`obj<<SendToReport`

Used with the Dispatch function to customize the appearance of a report.

`obj<<SendToByGroup`

Sends messages to open platforms or turn on platform features to each level of a by-group.

`obj<<Show Window(`*Boolean*`)`

1 shows the window (brings it to the front). 0 hides the window. If the window is also minimized (on Windows) or docked (on macOS), showing the window restores it to the normal state and brings it to the front.

`obj<<Size Window(`*x*`, `*y*`)`

Resizes the window to *x* pixels wide by *y* pixels high.

`obj<<Title(`*new title*`)`

Sets the quoted title of the platform.

`obj<<Top Report`

Returns a reference to the top display box in the report. Useful for By groups or other cases when several platform reports are in one window.

`obj<<View Web XML`

Returns the XML used to create the interactive HTML report. The XML code appears in the log.

`obj<<Zoom Window`

Resizes the window to be large enough to show all of its contents.

Bubble Plot

---

bp<<Set
Shape("Circle"|"Triangle"|"Square"|"Diamond"|"Arrow"|"Custom")

Description

Sets the shape for the bubble. When specifying a custom shape, use the `Set Custom Path` message to specify the path for the bubbles.

See Also

"bp<<Set Custom Path(path matrix|path text)".

---

bp<<Set Custom Path(*path matrix*|*path text*)

Sets a path for custom bubbles.

Arguments

path matrix An *N*x3 matrix.

path text A quoted string that contains SVG code.

See Also

"bp<<Set Shape("Circle"|"Triangle"|"Square"|"Diamond"|"Arrow"|"Custom")".

DOE

---

obj<<Get Prediction Variances

Description

Returns a vector of the MC variances used for the FDS plot.

Example

```
dt = Open( "$SAMPLE_DATA/Design Experiment/Bounce Data.jmp" );
d = DOE( Evaluate Design, X( :Silica, :Sulfur, :Silane ), Y( :Stretch ) );
d << Get Prediction Variances;
```

---

obj<<Set Number of FDS Points()

Description

Enables you to specify the number of runs used to generate an FDS plot.

Example

```
dt = Open( "$SAMPLE_DATA/Design Experiment/Bounce Data.jmp" );
d = DOE( Evaluate Design, X( :Silica, :Sulfur, :Silane ), Y( :Stretch ) );
d << Set Number of FDS Points( 20000 );
```

Partition

`obj<<Initial Splits(`*`condition`*`, `*`{left}`*`, `*`{right}`*`)`

Description

Describes the splits that are performed.

Example

```
dt = Open( "$SAMPLE_DATA/Car Poll.jmp" );
obj = Partition(
  Y( :country ),
  X( :sex, :marital status, :age, :type, :size ),
  Method( "Decision Tree" ),
  Initial Splits( :size == {"Large"}, {}, {:size == {"Medium"}} )
);
```

Notes

The condition is for the left side and is either `[name compareoperator value]` or `[name == list of values]`. The left is an empty list if the right has splits. Omit the right side if there are no splits. The left and right continue recursively in this format.

Response Screening

`obj<<Get PValues`

Returns a reference to a PValues table.

`obj<<Save PValues`

Stores the *p*-values in an output data table.

`obj<<Save Compare Means`

Stores the means comparisons in an output data table.

`obj<<Save Mean`

Stores the means in an output data table.

`obj<<Save Outlier Indicator`

Saves Outlier Indicator for each fit.

`obj<<Save Std Residuals`

Saves the residual formula for each fit.

obj<<Select Columns

Select columns in the original table corresponding to selected rows in this table.

Tabulate

obj<<Display Column Width(*<Data Column(Column Table(n),<column name path>), Row Label(Row Table(n), <column name path>)>, <width>*)

Returns or sets the display pixel width of a column in a Tabulate table.

Required Argument

Row Label Use Row Table and heading for columns in the row labels area.

Optional Arguments

Data Column Use Column Table and column references to define columns in the main body of the table.

column name path Specifies the Column Table or Row Table (both quoted), and the series of column headings that traces the path of the column. **Note:** Column Table or Row Table can be omitted if the table referenced is the first table.

width Specifies the pixel width of the column.

Examples

```
dt = Open( "$SAMPLE_DATA/Car Poll.jmp" );
obj = dt << Tabulate(
  Add Table(
     Column Table(
        Grouping Columns( :sex, :marital status ),
        Analysis Columns( :age ),
        Statistics( Sum, "% of Total" )
     ),
     Row Table( Grouping Columns( :type ) ),
     Row Table( Grouping Columns( :country, :size ) )
  )
);
Wait( 3 ); // for demonstration purposes
obj << Display Column Width( Row Label( Row Table( 2 ), "country" ), 150 );
Wait( 3 ); // for demonstration purposes
obj << Display Column Width(
  Data Column(
     Column Table( 1 ),
     "sex",
     "Female",
     "marital status",
     "Married",
     "age",
```

```
        "Sum"
    ),
    150
);
```

Python Integration Messages

The Python interfaces are also scriptable using a Python connection object. A scriptable Python connection object can be obtained using the `Python Connect()` function. See "Python Connect(<Echo(Boolean),> <Path(path),> <Use Python Version(string),> <Python System Path(list)>)".

`pythconn<<Control(<Interactive(`*`Boolean`*`)>|<Echo(`*`Boolean`*`)>)`

Description

Controls the execution of Python.

Optional Named Arguments

`Interactive(`*`Boolean`*`)` Enables interactive mode in the Python matplotlib package. Determines whether the graphics window is released or closed when graphics rendering is complete.

`Echo(`*`Boolean`*`)` Global argument. Prints the Python source lines to the JMP log. The default value is true.

`pythconn<<Disconnect`

Disconnects the Python integration interface connection.

`pythconn<<Execute(`*`{list of inputs}, {list of outputs}, Python code, <named arguments>`*`)`

Description

Submits Python code to the active global Python integration interface connection given a list of inputs. On completion, returns a list of outputs.

Returns

Returns 0 if successful and 1 otherwise. The results are returned using the list of outputs. Given each element of the JMP output list, the corresponding Python variable value is returned.

Positional Arguments

`{list of inputs}` A list of JMP variable names to be sent to Python as inputs.

`{list of outputs}` A list of JMP variable names to be retrieved from Python as outputs.

Python code The quoted Python code to submit.

Optional Named Arguments

See "pythconn<<Submit(Python code, <Expand(Boolean)>, <Echo(Boolean)>)".

---

pythconn<<Get(*name*)

Description

Gets a named variable from Python.

Returns

Returns the value of the named variable.

Argument

name The name of the JMP variable to be received from Python. The argument can represent any of the following Python data types: numeric, quoted string, matrix, list, or data frame.

---

pythconn<<Get Graphics(*format*)

Description

Gets the last graphics object written to the Python graph display window in the specified graphics format. The graphics object can be returned in several different graphic formats.

Returns

Returns a JMP picture object.

Argument

format The format that the Python graph display window contents are to be converted to. Valid formats are PNG, BMP, JPEG, JPG, TIFF, and TIF.

---

pythconn<<Get Version

Description

Gets the current version of the Python installation.

Returns

Returns a list of length 5 that contains the five components of the version number: major, minor, micro, releaselevel, and serial. The releaselevel value is a quoted string.

---

pythconn<<Is Connected

Description

Determines if the connection is active.

Returns

Returns 1 if connected and 0 otherwise.

pythconn<<JMP Name to Python Name(*name*)

Description

Maps a JMP variable name to its corresponding Python variable name using Python variable name naming rules.

Argument

name The name of the JMP variable to be sent to Python. Some variable names allowed by JMP are not allowed by Python. When you send using these variables from JMP to Python (the Send message), their names get changed. Use JMP Name to Python Name to determine what the variable name was changed to.

pythconn<<Send(*name*)

Description

Sends a named variable from JMP to Python.

Returns

Returns 0 if successful and non-zero otherwise.

Argument

name The name of the JMP variable to be sent to Python.

pythconn<<Submit(*Python code*, <Expand*(Boolean)*>, <Echo*(Boolean)*>)

Description

Submits Python code to the active global Python integration interface connection.

Returns

Returns 0 if successful and 1 otherwise.

Required Arguments

Python code The quoted Python code to submit.

Optional Arguments

Expand(*Boolean*) Performs an Eval Insert() on the Python code before submission.

Echo(*Boolean*) Prints the Python source lines to the JMP log. The default value is true.

pythconn<<Submit File(*path*)

Description

Submits statements to Python using the quoted path name.

Argument

path The quoted path to the file that contains the Python source lines to be executed.

R Integration Messages

The R interfaces are also scriptable using an R connection object. A scriptable R connection object can be obtained using the `R Connect()` function.

`rconn<<Control(Interrupt|Async(`*`Boolean`*`)|Echo(`*`Boolean`*`))`

Changes the control options for R. If `Async` is set to true (`1`) for `R Submit()`, this message immediately stops the execution of the R code that was submitted.

`rconn<<Disconnect()`

Disconnects this R connection.

`rconn<<Is Connected()`

Returns 1if the R connection is active, 0 otherwise.

`rconn<<Send File(`*`name`*`, <named arguments>)`

Send the specified JMP variable to R.

Returns

0 if successful, nonzero otherwise.

Required Argument

`name` A quoted string contains the name of a JMP variable to send to R.

Optional Named Arguments for Data Tables

`Selected(`*`Boolean`*`)` If true, sends only the selected rows from the referenced data table to R.

`Excluded(`*`Boolean`*`)` If true, sends only the excluded rows from the referenced data table to R.

`Labeled(`*`Boolean`*`)` If true, sends only labeled rows from the referenced data table to R.

`Hidden(`*`Boolean`*`)` If true, sends only hidden rows from the referenced data table to R.

`Colored(`*`Boolean`*`)` If true, sends only colored rows from the referenced data table to R.

`Markered(`*`Boolean`*`)` If true, sends only markered rows from the referenced data table to R.

`Row States(`*`Boolean`*`, <named arguments>)` Includes a Boolean argument and optional named arguments. Sends row state information from the referenced data table to R by adding an additional data column named “RowState”. The row state value consists of individual settings with the values shown in Table 3.2.

Table 3.2 Row States

Multiple row states are created by adding together individual settings.	Selected = 1 Excluded = 2 Labeled = 4 Hidden = 8 Colored = 16 Markered = 32
Arguments	`Colors(`*`Boolean`*`)` (Optional) If true, sends row colors and adds an additional data column named "RowStateColor". `Markers(`*`Boolean`*`)` (Optional) If true, sends row markers and adds an additional data column named "RowStateMarker".

`rconn<<Send(`*`name, <R Name(name)>`*`)`

Sends the quoted JMP data file to R. The *`name`* argument can represent any of the following data types: numeric, quoted string, matrix, list, or data table.

`rconn<<Get(`*`name`*`)`

Returns data from R. The *`name`* argument can represent any of the following data types: numeric, quoted string, matrix, list, or data table.

Returns

The value of the specified variable.

Arguments

`name` Specifies the quoted name of a JMP variable to be retrieved from R.

`rconn<<Get Graphics(`*`type`*`)`

Gets the last graphics object written to the R graph display window. The graphics object can be returned in different graphic formats.

Returns

A JMP picture object.

Required Argument

`type` The format the R graph display window contents are converted to. Valid formats are `"png"`, `"bmp"`, `"jpeg"`, `"jpg"`, `"tiff"`, and `"tif"`.

`rconn<<Submit(R code, Expand(Boolean), Echo(Boolean))`

Submits the quoted R code.

Returns

0 if successful, nonzero otherwise.

Required Argument

`code` Specifies the quoted R code to submit.

Optional Named Arguments

`Expand(Boolean)` Performs an `Eval Insert()` on the R code before submitting the code.

`Echo(Boolean)` Echoes the R source lines to the JMP log. The default value is true.

`Rconn<<Submit File(path)`

Submits statements to R using the file in the quoted *path*.

Arguments

`path` Specifies the quoted path to the file that contains R code to be executed.

`rconn<<Execute({list of inputs}, {list of outputs}, R code, <named arguments>)`

Submits the quoted R code to the R connection using the list of inputs. Upon completion, a list of outputs is returned.

Returns

0 if successful, nonzero otherwise.

Required Arguments

R code Specifies the quoted R code to submit.

`{list of inputs}` List of JMP variable names to be sent to R as inputs.

`{list of outputs}` List of JMP variable names to be retrieved from R as outputs.

Optional Named Arguments

See "rconn<<Submit(R code, Expand(Boolean), Echo(Boolean))".

`rconn<<Control(<Echo(Boolean)>)`

Controls the execution of R.

Returns

Void.

Optional Named Argument

`Echo(Boolean)` Echoes the R source lines to the JMP log.

`rconn<<Get Version()`

Gets the current version of R that is installed.

Returns

A vector of length 3 containing the R version number.

`rconn<<JMP Name To R Name(`*name*`)`

Maps a quoted JMP Name to its corresponding R Name using R variable name naming rules.

Returns

A quoted string that contains the quoted R name.

Arguments

name A quoted string that specifies the name of a JMP variable to be sent to R.

SAS Integration Messages

Metadata Server Objects

`metaserver<<Disconnect()`

Description

Disconnects the metadata server.

Returns

Void.

`metaserver<<Get Display Name()`

Description

Gets the display name of the metadata server.

Returns

A quoted string.

`metaserver<<Get Host Name()`

Description

Gets the host (machine) name of the metadata server.

Returns

A quoted string.

`metaserver<<Get Port()`

Description

Gets the port used for the metadata server connection.

Returns

An integer.

`metaserver<<Get User Identity()`

Description

Gets the identify of the connected user as defined in metadata.

Returns

A quoted string.

`metaserver<<Get User Name()`

Description

Gets the user name (login ID) that was used for the metadata server connection.

Returns

A quoted string.

SAS Server Objects

`sasconn<<Assign Libref(`*`libref`*`, `*`path`*`, `*`engine`*`, `*`engine options`*`)`

Description

Assign a SAS libref on this SAS server connection.

Returns

Void.

Arguments

See “SAS Assign Lib Refs("libref", "path", <"engine">, <"engine options">)”.

`sasconn<<Cancel Submit()`

Description

Cancels the currently running SAS Submit for this server that is presumably running asynchronously.

Returns

1 if a running submit was found and canceled; 0 otherwise.

`sasconn<<Clear Log History()`

Description

Clears the SAS Log history for this server.

Returns

Void.

`sasconn<<Clear Output History()`

Clears the SAS Output history for this server.

`sasconn<<Connect(<User Name(`*name*`)>, <Password(`*password*`)>, <Prompt("Always"|"Never"|"IfNeeded")>)`

Description

Attempt to reconnect a SAS server connection object that has become disconnected.

Returns

1 if the connection was successful, 0 otherwise.

Optional Named Arguments

`User Name(`*name*`)` Specifies the quoted user name for the connection.

`Password(`*password*`)` Specifies the quoted password for the connection.

`Prompt` A quoted keyword. `"Always"` means always prompt before attempt to connect. `"Never"` means never prompt even if the connection attempt fails (just fail with an error message going to the log), and `"IfNeeded"` (the default) means prompt if the attempt to connect with the given arguments fails (or is not possible with the information given).

`sasconn<<Deassign Libref(`*libref*`)`

Description

De-assign the quoted SAS libref on this SAS server connection.

Returns

Void.

Arguments

`libref` Specifies the quoted library reference.

`sasconn<<Disconnect()`

Description

Disconnect this SAS server connection.

Returns

Void.

sasconn<Does Module Exist(*module name*)

Description

Determines whether the specified SAS module exists in the SAS installation represented by the SAS connection. This can be helpful in determining whether certain SAS products are installed. The SAS DATA Step function MODEXIST is used to determine module existence.

Returns

1 if the specified module is found to exist, 0 if it does not exist.

Argument

`module name` Specifies the quoted SAS module, the existence of which should be checked. Do not include any extension.

sasconn<<Export Data(*dt*, *library*, *dataset*, *<named arguments>*)

Description

Exports a JMP data table to the specified SAS data set in the specified library on the active SAS server connection.

Returns

1 if the data table was exported successfully; 0 otherwise.

Optional Named Arguments

See "SAS Export Data(dt, "library", "dataset", <named_arguments>)".

sasconn<<Get Data Sets(*libref*)

Description

Returns a list of the data sets defined in a SAS library on this SAS server connection.

Returns

A list of quoted strings.

Arguments

libref Specifies the quoted SAS libref or friendly library name associated with the library for which the list of defined SAS data sets will be returned.

sasconn<<Get Error Count()

Description

Gets the count of the number of errors encountered in the previous SAS Submit.

Returns

An integer.

`sasconn<<Get File(source, dest)`

Description

Downloads a file from this SAS server connection.

Returns

Void.

Arguments

See "SAS Get File("source", "dest", "encoding")".

`sasconn<<Get File Names(fileref)`

Description

Gets a list of filenames found in the quoted *fileref* on this SAS server connection.

Returns

A list of quoted strings.

Arguments

fileref A quoted string that contains the name of *fileref* from which to retrieve filenames.

`sasconn<<Get File Names In Path(path)`

Description

Gets a list of filenames found in the quoted *path* on the current SAS server connection.

Returns

A list of quoted strings.

Arguments

path The quoted directory path on the server from which to retrieve filenames.

`sasconn<<Get File Refs()`

Description

Gets a list of the currently defined SAS filerefs on this SAS server connection.

Returns

A list of quoted strings.

`sasconn<<Get Librefs(<named arguments>)`

Description

Gets a list of the currently defined SAS librefs on this SAS server connection.

Returns

A list of quoted strings.

Optional Named Arguments

See "SAS Get Lib Refs(<named arguments>)".

`sasconn<<Get Log()`

Description

Retrieves the SAS Log from the last SAS Submit from this SAS server connection.

Returns

A quoted string.

`sasconn<<Get Option Name()`

Description

Queries SAS for the value of a SAS option variable.

Returns

A quoted string.

Example

The following script iterates through the define variables and prints out the values:

```
option_names = sasconn << Get Option Names();
For(i=1, i <= N Items(option_names), i++,
   option_value = sasconn << Get Option Value (option_names[i]);
   output = option_names[i] || "=" || char(option_value) || "/!n";
   Write(output);
);
```

`sasconn<<Get Output()`

Description

Retrieves the listing output from the last submission of SAS code to this SASServer object.

Returns

A quoted string.

`sasconn<<Get Results()`

Description

Retrieves the results of the previous SAS Submit as a scriptable object, which allows significant flexibility in what to do with the results.

Returns

A SAS Results Scriptable Object.

`sasconn<<Get Submit Status()`

Description

Gets the current status of a SAS Submit for this server that is presumably running asynchronously.

Returns

1 if the submit has not started; 2 if the submit is running; 3 if the submit has been canceled; 10 if the submit has completed successfully; 11 if the submit has completed with errors.

`sasconn<<Get Var Info(`*`libref`*`, `*`dataset`*`, <Password(`*`password`*`)>)`

Description

Returns information about the variables the specified SAS data set.

Required Arguments

`libref` The library name.

`dataset` The quoted name of the data set from which to retrieve variable names.

Optional Argument

`Password(`*`password`*`)` The quoted *`password`* for the connection.

`sasconn<<Get Var Names(`*`libref`*`, `*`dataset`*`, <`*`named arguments`*`>)`

Description

Retrieves the variable names contained in the specified data set on this SAS server connection.

Returns

A list of quoted strings.

Arguments

See "SAS Get Var Names(`string`, <"dataset">, <password("password")>)".

`sasconn<<Get Version(<"Long">)`

Description

Returns the SAS version as a quoted string such as "9.3" or "9.4".

Returns

A quoted string that contains the SAS version.

Optional Argument

`Long` A quoted keyword that specifies to return the long SAS version, which corresponds to the SYSVLONG SAS macro (for example, "9.02.02M0P01152009").

sasconn<<Get Work Folder()

Description

Returns the full path of the folder corresponding to the WORK library for this server.

Returns

A quoted string that specifies the work folder path.

sasconn<<Import Data(*library*, *dataset*, *<named arguments>*)

Description

Imports a SAS data set from this SAS server connection into a JMP table.

Returns

A JMP Data Table object.

Arguments

See "SAS Import Data(string, <"dataset">, <named arguments>)".

sasconn<<Import Query(*sqlquery*, *<named arguments>*)

Description

Executes the requested SQL query on this SAS server connection, importing the results into a JMP data table.

Returns

A JMP data table object.

Arguments

See "SAS Import Query("sqlquery", <named arguments>)".

sasconn<<Is Connected()

Description

Determines whether this SAS Server object is currently connected to SAS.

Returns

1 if *sasconn* is connect, 0 otherwise.

sasconn<<Is Product Available(*product name*)

Description

Determines whether the quoted SAS product is both licensed and installed in the session represented by the SAS connection. The SAS DATA Step functions SYSPROD and MODEXIST are used to determine the licensed and installed status of the product.

Returns

1 if the specified product is licensed, 0 if the product is not licensed, or -1 if the specified product is not recognized by SAS. This function throws an exception if the requested product is not one for which JMP knows how to check the installed status.

Required Argument

`product name` The quoted SAS product for which licensing should be checked. The product name can be specified with or without the "SAS/" prefix.

`sasconn<<Is Product Licensed(`*`product name`*`)`

Description

Determines whether the quoted SAS product is licensed in the session represented by the SAS connection. The SAS DATA Step function SYSPROD is used to determine the licensing status of the product.

Returns

1 if the specified product is licensed, 0 if the product is not licensed, or -1 if the specified product is not recognized by SAS.

Required Argument

`product name` The quoted SAS product for which licensing should be checked. The product name can be specified with or without the "SAS/" prefix.

`sasconn<<Kill Session(<`*`n`*`>)`

Description

If no argument is provided, the SAS connection is immediately terminated.

Returns

Void.

Arguments

`n` An optional number. The system waits n seconds for a normal shut-down before immediately terminating the SAS connection.

`sasconn<<Load Text File(`*`path`*`, <named arguments>)`

Description

Downloads the file specified in the quoted *`path`* from the active SAS server connection and retrieve its contents as a quoted string.

Returns

A quoted string.

Arguments

See "SAS Load Text File("path")".

sasconn<<Open Log Window()

Description

Opens (or brings to the front) the SAS Log window for this server.

Returns

Void.

sasconn<<Open Output Window()

Description

Opens (or brings to the front) the SAS Output window for this server.

Returns

Void.

sasconn<<Open SAS Results()

Description

Opens the results from the previous SAS Submit. Intended to be used with asynchronous SAS submits or the use of the OnSubmitComplete option to SAS Submit to give the JSL author a way to conditionally open the results of a submit.

Returns

Void.

sasconn<<Open Submit Results()

Description

Opens all the results from the last SAS Submit command.

Returns

Void.

sasconn<<Send File(*source*, *dest*)

Description

Uploads a file to this SAS server connection.

Returns

Void.

Arguments

See “SAS Send File("source", "dest", "encoding")”.

`sasconn<<Submit(sas code, <named arguments>)`

Description

Submits quoted SAS code to this SAS server connection.

Returns

Void.

Arguments

See "SAS Submit("sasCode", <named arguments>)".

`sasconn<<Submit File(filename, <named arguments>)`

Description

Submits a SAS code file to this SAS server connection.

Returns

Void.

Arguments

See "SAS Submit File("filename", <named arguments>)".

Stored Processes

`stp<<Begin Run(<named arguments>)`

Description

Starts this stored process executing in the background. This message is paired with `End Run`, which should also be called at some point after `Begin Run` to wait for the stored process to complete.

Returns

`-1` = execution failed.

`1` = not started.

`2` = running.

`3` = canceled.

`10` = completed successfully.

`11` = completed with errors.

Optional Named Arguments

Same as `Run`, except `AutoOpenResults` and `NoAlerts` are not supported. They are available on `EndRun`.

`AutoResume(<filename>)` If specified with no argument, it specifies that the stored process results should be auto-opened when the stored process completes. If a quoted

filename is specified, filename is opened rather than all results of the stored process being auto-opened.

`AutoResumeScript(`*`script`*`)` Specifies that after stored process execution completes, the quoted script should be evaluated. If the script is a function taking at least one argument, the function is evaluated with the scriptable stored process object passed as the first (and only) argument. `AutoResume` and `AutoResumeScript` are mutually exclusive.

`stp<<Delete Results(`*`<named arguments>`*`)`

Description

Deletes all results from the execution of this stored process.

Returns

`1` if deletion is successful, `0` otherwise (error message to JMP log).

Optional Named Arguments

`NoAlerts(`*`Boolean`*`)` If `True`, the user is not prompted for confirmation before the attempt is made to delete results.

`DeleteDirectory(`*`Boolean`*`)` If true, deletes the directory containing the stored process results along with the result files themselves. The default value is true.

`stp<<Edit Param Values()`

Description

Opens the stored process window for interactively setting parameter values.

Returns

1 if the user clicks OK to dismiss the window, 0 if the user clicks Cancel.

`stp<<End Run(`*`<named arguments>`*`)`

Description

Waits a specified amount of time (or forever) for a stored process started with `Begin Run` to complete. If the stored process is complete, retrieves the results, and opens them.

Returns

`-1` = execution failed.

`1` = not started.

`2` = running.

`3` = canceled.

`10` = completed successfully.

`11` = completed with errors.

Optional Named Arguments

AutoOpenResults(*Boolean*) Optional, Boolean. If True, results are automatically opened if the stored process completes in the time specified by *MaxWait*. If False, results are not automatically opened, and can be manually opened via the object returned by the Get Results message. Default is True.

MaxWait(*milliseconds*) An integer that specifies the maximum amount of time in milliseconds to wait for the stored process to complete. If MaxWait is not specified, End Run waits forever for the stored process to complete.

NoAlerts(*Boolean*) If True, error messages are sent to the JMP log rather than message boxes. The default value is False.

---

stp<<Get Metadata Id()

Description

Returns the metadata ID of the stored process.

Returns

A quoted string.

---

stp<<Get Metadata Path()

Description

Returns the full metadata path of the stored process.

Returns

A quoted string.

Stp<<Get Name()

Description

Returns the name of the stored process.

Returns

A quoted string.

---

stp<<Get Param Enum Labels(*name*)

Description

Gets the enumeration labels specified by the quoted *name* for a parameter.

Returns

A list of quoted strings.

Arguments

name Specifies the quoted name of the parameter whose enumeration labels to retrieve.

`stp<<Get Param Enum Values(`*name*`)`

Description

Gets the possible enumerated values for a parameter.

Returns

A list of quoted strings.

Arguments

`name` Specifies the quoted name of the parameter whose possible enumerated values to retrieve.

`stp<<Get Param Names(<`*named arguments*`>)`

Description

Gets a list of parameter names for this stored process of specific types.

Returns

A list of quoted strings.

Optional Named Arguments

`Visible(`*Boolean*`)` If true, gets only visible parameters. If `False`, gets only non-visible parameters. If not specified, gets both visible and non-visible parameters.

`Modifiable(`*Boolean*`)` If true, gets only modifiable parameters. If `False`, gets only non-modifiable parameters. If not specified, gets both modifiable and non-modifiable parameters.

`Required(`*Boolean*`)` If true, gets only required parameters. If `False`, gets only non-required parameters. If not specified, gets both required and non-required parameters.

`Expert(`*Boolean*`)` If true, gets only expert parameters. If `False`, gets only non-expert parameters. If not specified, gets both expert and non-expert parameters.

`stp<<Get Param Value(`*name*`)`

Description

Gets the current value of the specified parameter.

Returns

A quoted string.

Required Argument

name Specifies the name of the parameter whose value to retrieve.

`stp<<Get Results()`

Description

Gets the results generated by the execution of this stored process as a scriptable object.

Returns

A SAS Results scriptable object.

`stp<<Get Status()`

Description

Gets the execution status of the stored process.

Returns

-1 = execution failed.

1 = not started.

2 = running.

3 = canceled.

10 = completed successfully.

11 = completed with errors.

`stp<<Get Status Message()`

Description

Gets the message associated with the failure of the stored process, if any.

Returns

A quoted string.

`stp<<Reset Param Values()`

Description

Resets all parameter values to their metadata-defined default values.

Returns

Void.

`stp<<Run(<named arguments>)`

Description

Executes this stored process object in the foreground.

Returns

-1 = execution failed.

1 = not started.

2 = running.

3 = canceled.

10 = completed successfully.

11 = completed with errors.

Optional Named Arguments

AutoOpenResults(*Boolean*) If True, results are automatically opened when the stored process completes. If False, results are not auto-opened, and can be manually opened via the object returned by the GetResults message. The default value is True.

UserName(*username*) Specifies the quoted *user name* under which to run the stored process.

Password(*password*) Specifies the quoted *password* for *UserName*.

AuthDomain(*authDomain*) Specifies the quoted authentication domain of the credentials (*username*, *password*) given.

ODSDest(*dest*) Specifies the quoted ODS destination ("HTML", "PDF", "tagsets.SASReport12") for any ODS-generated results from the stored process. This requires the stored process SAS code to call %STPBEGIN. The default value is "HTML".

GraphicsDevice(*device*) Specifies the quoted SAS graphics device to use when generating graphics in ODS results. This requires the stored process SAS code to call %STPBEGIN. The default value is "GIF".

ODSStyle(*style name*) Specifies the quoted ODS style to apply to the results. This requires the stored process SAS code to call %STPBEGIN. There is no default value.

ODSStyleSheet(*path*) Specifies the quoted *path* to a CSS file on the client machine that is to be applied to generated ODS results. This requires the stored process SAS code to call %STPBEGIN. There is no default value.

NoAlerts(*Boolean*) If True, error messages are sent to the JMP log rather than message boxes. The default value is False.

---

stp<<Set Param Value(*name*, *value*)

Description

Sets the value of the specified stored process parameter to the specified value.

Returns

1 if successful, 0 otherwise (value can violate the parameter's constraints).

Arguments

name Specifies the quoted name of the parameter whose value to set.

value Specifies the quoted string that you want to set the parameter to.

---

stp<<Set Results Directory(*path*)

Description

Sets the quoted *path* on the client machine where stored process results are placed.

Returns

A quoted string.

Arguments

`path` Specifies the full quoted path of the directory where results of the stored process execution should be placed. The directory must exist or be creatable. If the results directory is not set, a temporary location appropriate for the operating system will be used, and that directory can be retrieved from the stored process Results scriptable object after the stored process executes.

SAS Results

`results<<Delete All Result Files()`

Description

Deletes all files created by the SAS Submit or Stored Process execution. Note that any result files that are still in use are not deleted.

Returns

1 if the deletion was successful; 0 if some of the files could not be deleted.

`results<<Get Directory()`

Description

Gets the directory where the results generated by the stored process or SAS submit are located.

Returns

A quoted string.

`results<<Get Log()`

Description

Get the SAS Log from the execution of the stored process or SAS submit.

Returns

A quoted string.

`results<<Get Main Result File Name(<Fullpath(Boolean)>)`

Description

Gets the full path of the main result file generated by the stored process or SAS submit.

Returns

A quoted string.

Optional Named Argument

`Fullpath(Boolean)` If true, the main result filename is returned as a full path. The default value is false.

`results<<Get Output()`

Description

Gets the SAS Listing output from the execution of the stored process or SAS submit.

Returns

A quoted string.

`results<<Get Output Datasets()`

Description

Get a list of output data set generated by the SAS Submit that created this SAS Results object.

Returns

A list of data set names in the form "libname.membername".

`results<<Get Result File Info(<Mimetype(`*`mime-type`*`)>, <Fullpath(`*`Boolean`*`)>)`

Description

Get information about result files that were generated by the execution of the stored process or SAS submit.

Returns

List of two lists of quoted strings. The first list is filenames, and the second list is the MIME-type of the corresponding file from the first list.

Optional Arguments

`Mimetype(`*`mime_type`*`)` Restricts the set of files for which information is returned to only those files with the specified quoted MIME-type. If not specified, information about all generated files is returned.

`Fullpath(Boolean)` If true, the filename returned for each result file is returned as a full path; if false, only the name of the file is returned. The default value is false.

`results<<Make JMP Report()`

Description

Parses the ODS XML results and creates a JMP report.

Returns

The display box for the report.

results<<Open All Results()

Description

Opens all results generated by the execution of the stored process or SAS submit.

Returns

Void.

results<<Open Result File(*filename*, <Run Script(*Boolean*)>)

Description

Attempts to open the result file with the given name.

Returns

JMP Data Table if one was opened.

Required Argument

filename Specifies the quoted name of the file from the generated results to open. filename should just be the name of the file, not the full path. If *filename* is a filename with no extension, both JMP data tables and JSL scripts in the results are searched for a match, and if both exist, both are opened.

Optional Argument

Run Script(*Boolean*) If true, and if *filename* is a JSL script, the script is executed. If false, *filename* is just opened, even if it is a JSL script.

results<<Run Script(*filename*)

Description

Looks for the JSL file in the results with the given *filename* and runs it if it finds it.

Returns

Void.

Argument

"filename" Specifies the quoted name of the JSL file from the generated results to open. The *filename* argument should just be the name of the file, not the full path, and it does not need to include the .jsl extension.

Schedule

See also "Schedule(n, script)".

sch<<Clear Schedule()

Cancels all scheduled events.

sch<<Close()

Closes the scheduler.

sch<<Restart()

Restarts the scheduler after it was stopped from running all scheduled events.

sch<<Show Schedule()

Shows a list of all scheduled events.

sch<<Stop()

Stops the scheduler from running all scheduled events.

Segments

Pie Seg(<*style*>, {*x*, *y*}, *radius*, [*values*])

Description

Creates a pie seg at the specified origin, with the specified radius, based on given values.

Required Arguments

{*x*, *y*} Specifies the x and y coordinates at which the pie seg is displayed.

radius Specifies the radius.

values Specifies the values specified in matrix format.

Optional Argument

style A quoted string that specifies the style: "Pie" (traditional pie chart with each slice sized by the Summary Statistic), "Ring" (each variable or level of a stratifying variable is represented by a concentric ring), or "Coxcomb" (the central angles for all slices are equal).

Sockets

skt<<Accept(<*callback*, *timeout*>)

Description

Tells the server socket to accept a connection and return a new connected socket.

Returns

A list of up to four items. The first is a quoted string that echoes the command (`"accept"`). The second is a quoted string, either `"ok"` or an error. The third is a quoted string that specifies the name of the machine that just connected. The fourth is a reference to the socket that you can send more messages.

Optional Arguments

`callback` Specifies the name of a function to receive the data.

`timeout` If you use a *`callback`*, *`timeout`* specifies how long the function should wait for an answer. For a server socket, 0 is an acceptable value because a server should not shut down because no one has connected to it recently.

`skt<<bind(`*`localhost`*`, `*`port`*`)`

Description

Associates a port on the local machine with the socket.

Returns

A list of two quoted strings. The first string is the command name (“bind”) and the second is “ok” if successful or an error.

Required Arguments

`localhost` Specifies the quoted local machine. You cannot bind to another machine.

`port` Specifies the port that should be used.

`skt<<Close()`

Description

Closes a socket.

Returns

A list of two quoted strings. The first string is the command name (“close”) and the second is “ok” if successful.

`skt<<Connect(`*`socketname`*`, `*`port`*`)`

Description

Connects to a listening socket.

Returns

A list of two quoted strings. The first string is the command name (“connect”) and the second is “ok” for a successful connection or an error sent back by the other socket.

Arguments

`socketname` Specifies the name of the other socket. If you are connecting to a web server, this is the web address (the name is preferred to the IP address).

port Specifies the port of the other socket to connect through.

`skt<<GetPeerName()`

Description

Retrieves the address and port of the socket at the other end of the connection.

Returns

A list of four quoted strings. The first echoes the command (`"getpeername"`). The second is either "ok" or an error. The third and fourth are the address and the port.

`skt<<Get Sock Name()`

Description

Retrieves the address and port of the socket at this end of the connection.

Returns

A list of four quoted strings. The first echoes the command (`"getsockname"`). The second is either "ok" or an error. The third and fourth are the address and the port.

`skt<<ioctl(FIONBIO, Boolean)`

Description

Controls the socket's blocking behavior.

Returns

A list of two quoted strings. The first string is the command name ("ioctl") and the second is "ok" if successful or an error.

Arguments

`FIONBIO, 1 FIONBIO` means Non-Blocking I/O. If true, turns on the behavior and the argument.

`skt<<Listen()`

Description

Tells the server socket to listen for connections.

Returns

A list of two quoted strings. The first echoes the command (`"listen"`) and the second is `"ok"` or an error message.

`skt<<recv(n, <callback, timeout>)`

`skt<<recvfrom(n, <callback, timeout>)`

Description

Receives either a stream message (`recv`) or a datagram message (`recvfrom`) from the other socket. If the two optional arguments are used, the data is not received immediately. Instead, the data is received when the function *callback* is called.

Returns

A list of three quoted strings. The first string is the command name ("recv" or "recvto"). The second is "ok" if successful or an error message if not. The third string is the data that was received. If a callback function is used, a fourth element is the socket that was used in the original `recv` or `recvfrom` message.

Required Argument

n Specifies the number of bytes to receive from the other socket.

Optional Arguments

callback Specifies the name of a function to receive the data.

timeout If you use a *callback*, *timeout* specifies how long the function should wait for an answer.

`skt<<Send(stream)`

`skt<<SendTo(dgram)`

Description

Sends the data in the argument to the other socket. `Send` sends a stream and `sendto` sends a datagram.

Returns

A list of three quoted strings. The first string is the command name ("send" or "sendto"). The second is "ok" if successful or an error message if not. The third string is any portion of the stream that could not be sent, or empty if all the data was sent correctly.

Arguments

stream Specifies the command to send to the other socket.

dgram Specifies the command to send to the other socket.

Notes

Either argument might need to contain binary data. JMP represents non-printable ASCII characters with a tilde (~) followed by the hexadecimal number. For example,

```
skt<<send(("GET / HTTP/1.0~0d~0a~0d~0a");
```

sends a "get request" to an HTTP server.

SQL

obj<<Custom SQL(*sql*)

Description

Changes the query to a custom SQL query and sets the SQL.

Required Argument

`sql` The quoted SQL query.

obj<<Generate SQL

Returns the SQL that the query generates when you run it.

obj<<Modify

Opens the query in Query Builder.

obj<<PostQueryScript(*script as text*)

Sets a JSL script that runs after the query finishes executing. *script as text* is quoted JSL code.

obj<<Query Name(<*new name*>)

Gets (without the *new name* argument) or sets (with the *new name* argument) the name of the query. The name of the query is used as the name of the data table that results from running the query.

obj<<Run(<"Private"|"Invisible">, <Update Table(*table*)>, <OnRunComplete(*script*)>, <OnRunCanceled(*script*)>, <OnError(*script*)>)

Description

Runs the SQL query in the background or foreground depending on the Query Builder preference "Run queries in the background when possible".

Returns

Null (if the query runs in the background) or a data table (if the query runs in the foreground).

Optional Named Arguments

`"Private"` A quoted keyword that opens the data table that the query produces without displaying it in a data table window. `"Private"` is available only if `OnRunComplete` is included in the script.

"Invisible" A quoted keyword that hides the data table that the query produces. Use this argument to keep the query result hidden but use it in a subsequent query. The data table is displayed in the Home Window's Window List and the Window > Unhide list.

Update Table Updates the specified data table. Runs the query in the foreground.

OnRunComplete Specifies a script to run after the query is complete. To get the resulting data table, include OnRunComplete. The OnRunComplete script needs to be defined in the global namespace, as indicated by the double colons in this example:

```
Names Default To Here( 1 );
::onComplete = Function( {dt},
  {default local},
  Write(
     "\!NQuery is complete!  Result name: \!"",
     dt << Get Name,
     "\!", Number of rows: ",
     N Rows( dt )
  )
);

query = Include( "rentals_fam_romcom.jmpquery" );
query << Run Background( On Run Complete( ::onComplete ) );
```

OnRunCanceled Specifies a script to run after the user cancels the query.

OnError Specifies a script to run if an error occurs.

Notes

If you want the data table that results from the background query, use the OnRunComplete optional argument. You can include a script that runs when the query completes and then assigns a data table reference to the resulting data table. Or you might pass the name of a function that accepts a data table as its first argument. That function is called when the query completes.

Examples

The following example opens a query that you previously saved from Query Builder. The query opens privately, that is, without opening Query Builder. The query runs, and the resulting data table opens.

```
query = Open( "c:/My Data/Movies.jmpquery", "Private");
dt = query << Run();
```

You can include a .jmpquery file in a script and run the query in the background using the <<Run Background message.

```
query = Include( "C:/Queries/movies.jmpquery");
query <<Run Background();
```

The following example queries the database, opens the resulting data table, and prints the number of data table rows to the log.

```
confirmation = Function( {dtResult},
  Write( "\!NNumber of rows in query result: ", N Rows( dtResult ) )
);
query = New SQL Query(
  Connection(
     "ODBC:DSN=SQL
  Databases;APP=MYAPP;TrustedConnection=yes;WSID=D79255;DATABASE=SQB;"
  ),
  QueryName( "movies_to_update" ),
  Select( Column( "YearMade", "t1" ), Column( "Rating", "t1" ) ),
  From( Table( "g6_Movies", Schema( "SQB" ), Alias( "t1" ) ) ),

);
query << Run( OnRunComplete( confirmation ) );
```

`Run Background(<OnRunComplete(`*script*`), <"Private"|"Invisible">>, <OnRunCanceled(`*script*`)>, <OnError(`*script*`)>)`

Description

Runs the SQL query in the background. The running query is not displayed.

Returns

Null (or the data table object, if `OnRunComplete` is included).

Optional Named Arguments

`OnRunComplete(`*script*`)` Specifies a script to run after the query is complete. To get the resulting data table, include `OnRunComplete`. The `OnRunComplete` script needs to be defined in the global namespace, as indicated by the double colons in this example:

```
Names Default To Here( 1 );
::onComplete = Function( {dt},
  {default local},
  Write(
     "\!NQuery is complete!  Result name: \!"",
     dt << Get Name,
     "\!", Number of rows: ",
     N Rows( dt )
  )
);

query = Include( "rentals_fam_romcom.jmpquery" );
query << Run Background( On Run Complete( ::onComplete ) );
```

`"Private"` Does not open the resulting data table. Specify only with `OnRunComplete`. If you include `private` in a background query, JMP opens the data table as invisible instead.

`"Invisible"` Hides the data table. Use this argument to keep the query result hidden but use it in a subsequent query. The data table is displayed in the Home Window's Window List and the Window > Unhide list.

`OnRunCanceled` Specifies a script to run after the user cancels the query.

`OnError` Specifies a script to run if an error occurs.

Notes

All queries except for SAS queries run in the background based on the Query Builder preference "Run the queries in the background when possible", which is selected by default. For SAS queries, `Run Background()` is ignored.

You can include a .jmpquery file in a script and run the query in the background using the `Run Background` message.

```
query = Include( "C:/Queries/movies.jmpquery");
query <<Run Background();
```

`Run Foreground(<OnRunComplete(`*`script`*`), <"Private"|"Invisible">>, <OnRunCanceled(`*`script`*`)>, <OnError(`*`script`*`)>)`

Description

Runs the SQL query in the foreground.

Returns

A data table that opens when the query is finished.

See Also

"Run Background(<OnRunComplete(`script`), <"Private"|"Invisible">>, <OnRunCanceled(`script`)>, <OnError(`script`)>)"

`obj<<Save`

Saves the query to its associated file. The save fails if the query does not yet have an associated file.

`obj<<Save As(`*`path`*`, <Replace Existing(`*`Boolean`*`))`

Saves the query to the specified file. If the file already exists, the save fails unless `Replace Existing` is true.

Other Objects

Zip Archives

`list = za<<Dir`

Returns a list of member names

`data = za<<Read(`*`member name`*`, <Format("blob")>)`

Returns a quoted string that contains the entire quoted *`member name`*. A zip file consists of filenames, also called "member names".

Notes

For remote files, JMP copies the URL data to the local disk. When the zip archive is no longer accessible, the local data file is deleted.

`actualname = za<<Write(`*`member name`*`, `*`member data`*`, <"replace">)`

Writes a quoted string or quoted blob to a zip archive member file. If the quoted *`member name`* isn't in the current zip file, the returned `actualname` is the same as *`member name`*. This member name will be changed to prevent overwriting an existing member; the name actually used is returned. The quoted *`member data`* argument is the data to write into the zip file's member of that name. *`replace`* creates the file with a temporary name, deletes the old file, and renames the temporary file to the existing name.

Journals

`jnl<<Save HTML(<`*`path`*`>, <`*`format`*`>)`

Saves the journal as HTML.

Optional Arguments

`"path"` Specifies the quoted path for the saved HTML file (for example, `"c:/myFile.html"`).

`format` Specifies the quoted graphic file format. JPG, PNG, and TIFF formats are supported. The graphics are saved in a subdirectory named gfx.

`jnl<<Save RTF(<`*`path`*`>, <format>)`

Saves the journal as an RTF file.

Optional Arguments

`"path"` Specifies the quoted path for the saved RTF file (for example, `"c:/myFile.rtf"`).

"format" Specifies the quoted file format for the embedded graphics. "JPG", "PNG", and "EMF" formats are supported on Windows. All journals are saved as PDF files by default on macOS.

Notes

If no *path* or *format* are provided, you are prompted to name the file and specify the format on Windows. On macOS, you are prompted to name the file. The file is saved as a PDF file by default.

---

jnl<<Save PDF(<*path*>, <Show Page Setup(*Boolean*)>, <Portrait(*Boolean*)>)

Saves the journal as a PDF file.

Optional Arguments

"path" The quoted *path* for the saved PDF file (for example, "c:/myFile.pdf").

Show Page Setup If set to true, opens the Page Setup window to let the user change the margin, magnification level, and other page layout options.

Portrait Determines whether the page orientation is portrait or landscape. Overrides the user's selection in the Show Page Setup window.

Appendix A

SQL Functions Available for JMP Queries

The `Query()` JSL function performs a SQL query on selected tables and exports the data to a data table. The following example first assigns the `t1` alias to Big Class.jmp. `name`, `age` greater than 13, and `height` are then selected from the `t1` table.

```
dt = Open( "$SAMPLE_DATA/Big Class.jmp" );
Query( Table( dt, "t1" ),
       "SELECT t1.name, t1.age, t1.height FROM t1
       WHERE t1.age > 13" );
```

You can use SQL functions in a query. For example, `SELECT CURRENT_TIMESTAMP` returns the current UTC/GMT time stamp as a SQLite time string:

```
Query( Scalar, "SELECT CURRENT_TIMESTAMP;" );
```

This appendix lists the numeric, date-time, string, system SQL, and aggregate functions that you can use in SQL queries. “Yes” in the SQLite column indicates native SQLite functions. See the Online SQLite documentation at https://www.sqlite.org/lang.html.

Contents

Numeric SQL Functions 531
Date-Time SQL Functions 532
String SQL Functions 535
System SQL Functions 536
Aggregate SQL Functions 536

Numeric SQL Functions

The numeric SQL functions are described here.

Numeric Function	Native SQLite	Description
ABS(*number*)		Returns the absolute value of the specified number.
ACOS(*cosine*)		Returns the angle in radians for the specified cosine.
ASIN(*sin*)		Returns the angle in radians for the specified sine.
ATAN(*tangent*)		Returns the angle in radians for the specified tangent.
ATAN2(*x*, *y*)		Two-argument arctangent function.
CEILING(*number*) CEIL(*number*)		Returns the smallest integer larger than the specified number.
COS(*radians*)		Returns the cosine of the specified angle in radians.
COT(*radians*)		Returns the cotangent of the specified angle in radians.
DEGREES(*radians*)		Converts an angle in radians to an angle in degrees.
EXP(*number*)		Returns the constant *e* raised to the specified power.
FLOOR(*number*)		Returns the largest integer smaller than the specified number.
LN(*number*) LOG(*number*)		Returns the natural logarithm of the specified number.
LOG10(*number*)		Returns the common logarithm of the specified number.
MAX(*n1*, *n2*, ...)	Yes	Returns the largest of the specified numbers. A minimum of two numbers must be specified.
MIN(*n1*, *n2*, ...)	Yes	Returns the smallest of the specified numbers. A minimum of two numbers must be specified.
MOD(*dividend*, *divisor*)		Returns the remainder when *dividend* is divided by *divisor*. Floating-point values are truncated to integers before the modulus operation is performed.
PI()		Returns the value of the constant pi (π).
POWER(*number*, *power*) POW(*number*, *power*)		Raises *number* to the specified *power*.

Numeric Function	Native SQLite	Description
RADIANS(*degrees*)		Converts an angle in degrees to an angle in radians.
RANDOM() RANDOM(*max*) RANDOM(*min*, *max*)		Returns a random number. RANDOM() returns a number between 0 and 1. RANDOM(max) returns a number between 0 and *max*. RANDOM(min, max) returns a number between *min* and *max*. This function is equivalent to the Random Uniform() JSL function, and its seed can be controlled using the Random Reset() JSL function. RANDOM can be shortened to RAND.
RANDOMBLOB(*length*)	Yes	Returns an N-byte blob that contains pseudo-random bytes. See the SQLite Online documentation at https://www.sqlite.org/lang.html.
ROUND(*number*, *<precision>*)		Rounds *number* to the number of decimal places given by *precision*. The default value of *precision* is 0, and *precision* can be negative.
SIGN(*number*)		Returns 1 if *number* is positive, -1 if *number* is negative, or 0 if *number* is zero.
SIN(*radians*)		Returns the sin of the specified angle in radians.
SQRT(*number*)		Returns the square root of *number*.
TAN(*radians*)		Returns the tangent of the specified angle in radians.
TRUNCATE(*number*, *<precision>*)		Truncates *number* at the number of decimal places given by *precision*. The default value of *precision* is 0, and *precision* can be negative. TRUNCATE() can be shortened to TRUNC().

Date-Time SQL Functions

Using date-time functions in JMP queries is complicated by the fact that the SQL engine that handles JMP queries (SQLite) uses different formats for storing dates than JMP does. SQLite stores date-times as strings. However, JMP stores date-times as the number of seconds since January 1, 1904. When you have columns in your table that contain date-times, the conversions are handled automatically. However, when you use functions that return date-times, you might need to let JMP know when a conversion is required.

Consider the CURRENT_TIMESTAMP function. CURRENT_TIMESTAMP is a built-in SQLite function that returns the current UTC/GMT time stamp as a SQLite time string:

```
Query( Scalar, "SELECT CURRENT_TIMESTAMP;" );
```

returns:

```
"2016-02-16 15:44:42"
```

The string could perhaps be parsed as a date to return it as a JMP date. To prevent the need to do so, wrap the CURRENT_TIMESTAMP function in the JMPDATE() function:

```
Query( Scalar, "SELECT JMPDATE( CURRENT_TIMESTAMP );" );
```

returns:

```
3538482531
```

The string is an unformatted JMP date. However, if you pass a SQLite time string to another SQL date-time function, you do not need to use JMPDate(); the value will be converted to a JMP date automatically. Here is an example:

```
Query( Scalar, "SELECT EXTRACT('YEAR', CURRENT_TIMESTAMP);" );
```

Using native SQLite date-time functions (date(), time(), datetime(), julianday(), strftime()) in JMP queries is not recommended because JMP date-time values are not compatible with those functions.

Date-Time Function	Naive SQLite	Description
CURRENT_DATE	Yes	Returns the current date (UTC/GMT) as a SQLite time string.
CURRENT_TIME	Yes	Returns the current time (UTC/GMT) as a SQLite time string.
CURRENT_TIMESTAMP	Yes	Returns the current date and time (UTC/GMT) as a SQLite time string.
DATEDIFF(*date1*, *date2*, *interval*, <*alignment = "Start"*>)		Computes the difference between two dates in units specified by *interval*, based on *alignment*. This function works the same as the Date Difference() JSL function. Valid values for *interval* are: "Year", "Quarter", "Month", "Week", "Day", "Hour", "Minute" and "Second". Valid values for *alignment* are "Start", "Actual" and "Fractional". If *alignment* is not specified, "Start" is used.

Date-Time Function	Naive SQLite	Description	
EXTRACT(*datepart*, *datetime*, *<use_locale = 1>*)		Extracts a specific part of a date or date-time value. *Datetime* is a JMP date-time value or a SQLite time string. *Use_locale* is optional and applies only to date name parts such as "MonthName" and "DayName" and determines whether values from the current language or English are returned. The following values of *datepart* are supported:	
		"Year"	Returns the year as a number.
		"Month"	Returns the numeric month (1-12).
		"MonthName"	Returns the full name of the month in the current language (*use_locale = 1*) or English (*use_locale = 0*).
		"Mon", "MMM"	Returns the abbreviated name of the month.
		"Day"	Returns the day of the month (1-31).
		"DayName"	Returns the name of the day of the week.
		"DayOfWeek"	Returns the numeric day of the week (1-7).
		"DayOfYear"	Returns the numeric day of the year (1-366).
		"Quarter"	Returns the numeric quarter (1-4).
		"Hour"	Returns the hour (0-23).
		"Minute"	Returns the minute (0-59).
		"Second"	Returns the seconds, including any fractional part.
		"Date"	Returns just the date portion of a date-time value as a JMP date-time value.
		"Time"	Returns just the time portion of a date-time value as a JMP date-time value.
JMPDATE(*SQLite time string*)		Converts a SQLite time string to the equivalent JMP date-time value.	
NOW()		A synonym for TODAY().	
TODAY()		Returns the JMP date-time value of the current moment in *local* time, which matches the JMP Today() function.	

String SQL Functions

The string SQL functions are described here
.

Function	Native SQLite	Description
HEX(*binary*)	Yes	SQLite built-in function that converts a BLOB to a string of hexadecimal characters. Useful when paired with the RANDOMBLOB() function.
JLEFT(*string*, *len*, *<pad>*)		Like the JSL Left() function. Returns *len* characters from the beginning of *string*. If *pad* is specified and fewer than *len* characters are present in *string*, the result is padded with *pad* out to length *len*.
JRIGHT(*string*, *len*, *<pad>*)		Like the JSL Right() function. Returns *len* characters from the end of *string*. If *pad* is specified and fewer than *len* characters are present in *string*, the result is padded with *pad* at the front out to length *len*.
LENGTH(*string*)	Yes	SQLite equivalent of the ANSI standard CHAR_LENGTH() function. Returns the length of its string argument in characters.
LOCATE(*string1*, *string2*) POSITION(*string1*, *string2*)		Returns the (1-based) starting position of *string1* within *string2*, returning 0 if *string1* is not found within *string2*.
LOWER(*string*)		Returns a copy of *string* with all uppercase characters converted to lowercase.
LTRIM(*string*, *<trimchars>*)	Yes	Trims any characters contained in *trimchars* from the beginning of *string* and returns the result. If *trimchars* is omitted, spaces are trimmed.
PRINTF(*format*, *<arg1, ..., argN>*)	Yes	Allows constructing strings using placeholders and arguments. See the SQLite Online documentation at https://www.sqlite.org/lang.html.
REPLACE(*string*, *find*, *replace*)	Yes	Replaces all instances of *find* in *string* with *replace* and returns the result. If *replace* is numeric, it is converted to a string.
REVERSE(*string*)		Returns a copy of *string* with the order of the characters reversed.

Function	Native SQLite	Description
RTRIM(*string*, *<trimchars>*)	Yes	Trims any characters contained in *trimchars* from the end of *string* and returns the result. If *trimchars* is omitted, spaces are trimmed.
SPACE(*length*)		Returns a string consisting of *length* space characters.
SUBSTR(*string*, *start*, *<length>*)	Yes	Returns the substring of *string* starting at *start* (1-based) that is *length* characters long. If *length* is omitted, the substring starting at *start* and continuing to the end of *string* is returned.
TRIM(*string*, *<trimchars>*)		Trims any characters contained in *trimchars* from the end of *string* and returns the result. If *trimchars* is omitted, spaces are trimmed.
UPPER(*string*)		Returns a copy of *string* with all lowercase characters converted to uppercase.

System SQL Functions

The system SQL functions are described here.

Function	SQLite	Description
COALESCE(*arg1*,..., *argN*)	Yes	Returns the first argument passed to it that is non-NULL. Returns NULL if all arguments are NULL. Requires at least two arguments.
IFNULL(*arg1*, *arg2*)	Yes	Returns *arg1* if not NULL, otherwise *arg2*. Basically, IFNULL is a two-argument version of COALESCE().
NULLIF(*arg1*, *arg2*)	Yes	Returns *arg1* if *arg1* and *arg2* are different and returns NULL if the arguments are equal. Used when you have non-NULL values in your database that you want to treat as NULL.

Aggregate SQL Functions

When passing a single argument to an aggregate function, that argument can be preceded by the keyword DISTINCT, which filters out duplicate values.

For all aggregations other than COUNT(*), NULL and missing values are ignored.

Function	SQLite	Description
AVG(*num_expr*)		Computes the average of *num_expr* for the rows in the group. *Num_expr* must be numeric.
COUNT(*expr*) COUNT(*)		Counts the number of times *expr* is not NULL in the group. COUNT(*) returns the total number of rows in the group.
GROUP_CONCAT(*expr*, *<separator = ','>*)	Yes	Concatenates all non-NULL values of *expr* and returns them as a string. Numeric values of *expr* are converted to character. If *separator* is present, it is placed between the values. The default separator is a comma. DISTINCT can be used only with GROUP_CONCAT() if *separator* is not specified.
MAX(*expr*)		Returns the maximum value of *expr* in the group. *Expr* can be character or numeric.
MIN(*expr*)		Returns the minimum value of *expr* in the group. *Expr* can be character or numeric.
STDDEV_POP(*num_expr*)		Computes the population standard deviation of *num_expr* for the group.
STDDEV_SAMP(*num_expr*)		Computes the sample standard deviation of all *num_expr* for the group.
SUM(*num_expr*)		Returns the sum of *num_expr* for the group. If no non-NULL values are found, SUM() returns NULL.
TOTAL(*num_expr*)	Yes	Same as SUM(*num_expr*), except TOTAL() returns 0.0 if no non-NULL values are found.
VAR_POP(*num_expr*)		Computes the population variance of *num_expr* for the group.
VAR_SAMP(*num_expr*)		Computes the sample variance of *num_expr* for the group.

Appendix B

References

Gander, W., and Gautschi, W. (2000). "Adaptive Quadrature–Revisited." *BIT Numerical Mathematics* 40:84–101.

Moore, W. A., and Parks, D. R. (2012). "Update for the logicle data scale including operational code implementations." *Cytometry Part A* 81A:273–277. Accessed January 26, 2021. https://onlinelibrary.wiley.com/doi/full/10.1002/cyto.a.22030.

Appendix C

Technology License Notices

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software ia Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.